

2013

Election Data Visualisation

Long, Elena

<http://hdl.handle.net/10026.1/1589>

<http://dx.doi.org/10.24382/4332>

University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

COPYRIGHT STATEMENT

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's prior consent.

Election Data Visualisation

By

Elena Long

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Management
Plymouth Business School

December 2012

Election Data Visualisation

Elena Long

ABSTRACT

Visualisations of election data produced by the mass media, other organisations and even individuals are becoming increasingly available across a wide variety of platforms and in many different forms. As more data become available digitally and as improvements to computer hardware and software are made, these visualisations have become more ambitious in scope and more user-friendly. Research has shown that visualising data is an extremely powerful method of communicating information to specialists and non-specialists alike. This amounts to a democratisation of access to political and electoral data.

To some extent political science lags behind the progress that has been made in the field of data visualisation. Much of the academic output remains committed to the paper format and much of the data presentation is in the form of simple text and tables. In the digital and information age there is a danger that political science will fall behind.

This thesis reports on a number of case studies where efforts were made to visualise election data in order to clarify its structure and to present its meaning. The first case study demonstrates the value of data visualisation to the research process itself, facilitating the understanding of effects produced by different ways of estimating missing data. A second study sought to use visualisation to explain complex aspects of voting systems to the wider public.

Three further case studies demonstrate the value of collaboration between political scientists and others possessing a range of skills embracing data management, software engineering, broadcasting and graphic design. These studies also demonstrate some of the problems that are encountered when trying to distil complex data into a form that can be easily viewed and interpreted by non-expert users. More importantly, these studies suggest that when the skills balance is correct then visualisation is both viable and necessary for communicating information on elections.

List of Contents

Abstract.....	iii
Acknowledgements.....	xviii
Author’s declaration.....	xix
1 Introduction.....	1
1.1 Background.....	1
1.2 Dissertation origins.....	3
1.3 Aims and Structure	5
2 Visualising data.....	12
2.1 Introduction	12
2.2 Perception and Neurocognition	13
2.3 The evolution of information visualisation	27
2.4 Theories of information visualisation.....	31
2.5 Conclusions	41
3 Visualising politics.....	43
3.1 Introduction	43
3.2 Imaging politics	44
3.3 Technological change and political visualisation.....	47
3.4 Democratizing broadcasting	52
3.5 Electioneering and mobile phone technologies	53

3.6	Conclusions	56
4	Mapping data	57
4.1	Introduction	57
4.2	Mapping: optimising data visualisation.....	58
4.3	The visualisation of election results	61
4.3.1	Interactive maps	64
4.3.2	Choropleth Maps.....	66
4.3.3	Proportional or graduated symbols maps.....	68
4.3.4	Election Dashboards	70
4.3.5	Cartograms of Election Results	74
4.4	UK General Election 2010: Different views	79
4.4.1	The Guardian	79
4.4.2	The Telegraph.....	80
4.4.3	The Times	81
4.4.4	Sky News	82
4.4.5	Yahoo!.....	83
4.4.6	BBC.....	83
4.5	Themes within the visualisation of Election Results.....	87
4.6	Conclusions	91
5	Data visualisation within the research process	93
5.1	Introduction	93

5.2	Electoral research and mapping.....	95
5.3	Triangles and three-party competition.....	106
5.4	Handling missing data and computing averages	112
5.5	Conclusions	128
6	Explaining voting systems to the general public	130
6.1	Introduction	130
6.2	Electoral bias	134
6.2.1	Background.....	134
6.2.2	Structuring the presentation	135
6.2.3	Electoral Bias and its components.....	137
6.2.4	Examining the survey evidence: electoral bias.....	156
6.3	Supplementary Vote	164
6.3.1	Background.....	164
6.3.2	The presentation.....	165
6.4	Examining the survey evidence: supplementary vote	173
6.5	Conclusions	176
7	Accessing election data by mobile telephone	179
7.1	Introduction	179
7.2	Growth and development of mobile phone communications.....	181
7.3	Mobile search versus Internet search.....	183
7.4	Election Day 2009 European Parliament.....	184

7.5	The problems faced	187
7.6	Conclusions	191
8	Natural Interface to Election Data	193
8.1	Introduction	193
8.2	Designing a Web-Based Dialog System.....	195
8.3	Identifying Potential Users	197
8.4	Designing for the analysis of Election Data	198
8.5	Input and Output Interfaces	200
8.5.1	Input interface	201
8.5.2	Output Interface	203
8.6	Users' Requests	207
8.7	From enquiry to SQL Query translation.....	210
8.8	Help Instructions.....	212
8.9	Difficulties with Natural Language Enquiries.....	213
8.10	Interacting with electoral data	214
8.11	Drill Down Mapping	217
8.12	Conclusions	222
9	Political Science meets Broadcasters.....	225
9.1	Introduction	225
9.2	Poll tracker.....	227
9.3	Broadening the Sky News Elections Data Coverage.....	232

9.3.1	Historical data	232
9.3.2	Historical Cartograms	233
9.4	Interpreting general election manifestos at the 2010 general election	242
9.5	House of Commons projections	246
9.6	The Alternative Vote Referendum	251
9.7	Conclusions	261
10	Conclusions.....	263
10.1	Overview	263
10.2	Chapter summaries	264
10.3	Evaluating Outcomes.....	270
10.4	The future	275
11	List of References	277
12	Bound in Copies of Publications.....	297
12.1	Mobile Election	297
12.2	Natural Interface to Election Data	309
12.3	Election Data Visualization	321
12.4	Forecasting the 2010 General Election.....	339

List of Figures

Figure 2.1 Hierarchical structure of visual system	18
Figure 2.2 Two parallel streams of visual information processing.....	19
Figure 2.3 Focus of attention	22
Figure 2.4 Eye Movements: Yarbus, 1965	24
Figure 2.5 Eye Movements: 2008 illustration.....	24
Figure 2.6 A coxcomb chart.....	29
Figure 2.7 John Snow's Cholera Graphic.....	30
Figure 2.8 Minard's Graphic of Napoleon's Moscow Campaign of 1812	30
Figure 2.9 The visual variables described by Bertin:	32
Figure 2.10: Tukey's boxplot.....	35
Figure 2.11 Confusion between area and position inside a tree map.....	38
Figure 2.12 Voronoi treemap.....	39
Figure 2.13 Cone Tree visualisation	39
Figure 2.14 Perspective Wall.....	40
Figure 2.15 Hyperbolic tree	40
Figure 2.16 Screenshots from Tableau Software.....	41
Figure 3.1: Triangular graph of party share	44
Figure 3.2 Composite screenshot of General Election visualisations.....	47
Figure 3.3 Tweetminster	51
Figure 3.4 Facebook polling day	52
Figure 3.5 Conservative Party iPhone app.....	54
Figure 3.6 Labour Party iPhone app	55
Figure 3.7 Liberal Democrats iPhone app	55

Figure 4.1 USA Today and the 2004 Presidential Election	62
Figure 4.2: Mapping electoral data across the United States	63
Figure 4.3 The New York Times map	64
Figure 4.4 The Sky News US Election map	64
Figure 4.5: A choropleth map of United States	66
Figure 4.6: USA Today Presidential Election 2008.....	67
Figure 4.7: New York Times and the Presidential Race.....	68
Figure 4.8: Washington Post and County level data.....	68
Figure 4.9: Heat maps of US elections 2008	69
Figure 4.10: iDashboards of the 2008 US Presidential election	72
Figure 4.11: Yahoo US Election 2008 Dashboard.....	73
Figure 4.12: Grundy’s Map of the United States	75
Figure 4.13: Dorling’s centroids	76
Figure 4.14: Gastner & Newman map of US Presidential Election	77
Figure 4.15: Comparison of two images of US presidential election	77
Figure 4.16: The Guardian 2010 election map	80
Figure 4.17 The Telegraph 2010 election maps.....	81
Figure 4.18: The Times 2010 election map	82
Figure 4.19: Sky News 2010 constituency battleground	82
Figure 4.20: Yahoo 2010 election map and Commons projection	83
Figure 4.21: BBC 2010 election visualisation	84
Figure 4.22: Comparisons of mapping UK elections.....	85
Figure 5.1: Ward boundary maps and seats won by Liberal Democrats 1983-1994.....	96
Figure 5.2: Ward boundary maps and seats won by Liberal Democrats by 1994	97

Figure 5.3: Wards, polling districts, and polling stations in the London Borough of Brent, 2001.....	99
Figure 5.4: Percentage differences in predicted turnout when re-siting polling stations for European elections	99
Figure 5.5: Percentage differences in turnout when re-siting polling stations for local elections	100
Figure 5.6: Distribution of percentage vote shares for Respect.....	102
Figure 5.7: Distribution of percentage vote shares for Green.....	102
Figure 5.8: Distribution of percentage vote shares for UKIP	103
Figure 5.9: Distribution of percentage vote shares for BNP.....	103
Figure 5.10 Pattern of Green Party competition and 2004 London Assembly list vote	105
Figure 5.11 Pattern of Green Party competition and 2008 London Assembly list vote	106
Figure 5.12: Distribution of three-party vote shares, 2005 general election	108
Figure 5.13: Distribution of three-party vote shares:.....	109
Figure 5.14: The superposition $ABC+ACB+BAC+BCA+CAB+CBA$	110
Figure 5.15: screen grab1 evolution of votes 2005-2010 general election	111
Figure 5.16: screen grab2 evolution of votes 2005-2010 general election	111
Figure 5.17: Decline in eligible by-elections for forecast model.....	114
Figure 5.18: Declining percentage of by-elections used in forecast model.....	114
Figure 5.19: Old model estimates for Conservative share.....	114
Figure 5.20: Consequence of new data selection criteria	119
Figure 5.21: Estimating missing vote shares and model stability.....	119
Figure 5.22: Effect of using one month or three-month averages	121
Figure 5.23: Effect of different weighting schemes.....	122
Figure 5.24: Adjusting for size of total votes	122

Figure 5.25: Effect of controlling for ward marginality (cut point 20% majority).....	123
Figure 5.26: Effect of controlling for ward marginality (cut point 15% majority).....	124
Figure 5.27: Effect of using three cut points.....	124
Figure 5.28: Urban vs. Rural by-elections	125
Figure 5.29: Comparison of forecasts for Conservative national vote share.....	125
Figure 5.30: Comparing random Sub-samples of data	126
Figure 5.31: Revised model and National Equivalent Vote, 1993-2010	127
Figure 6.1: Introducing electoral bias in the UK context.....	137
Figure 6.2: Introducing the hypothetical election	138
Figure 6.3: The contribution of electoral inequalities to electoral bias	139
Figure 6.4: Differences in electorate size and the pattern of party wins.....	140
Figure 6.5: Declaring the national result.....	141
Figure 6.6: A summary of the election outcome.....	141
Figure 6.7: Variations in electoral turnout	143
Figure 6.8: Differences in turnout and winning party.....	143
Figure 6.9: Electoral bias produced by differences in turnout.....	144
Figure 6.10: Problems of explaining the impact of vote distribution	145
Figure 6.11: Variations in vote shares across constituencies.....	146
Figure 6.12: Different patterns of votes	146
Figure 6.13: Frequency of second placed parties at the 2010 general election	148
Figure 6.14: Ineffective votes cast for Conservative and Labour at the 2010 general election	149
Figure 6.15: Comparing ineffective votes at the 2010 general election	149
Figure 6.16: Converting votes into seats: the Conservative party in 2010	150
Figure 6.17: Converting votes into seats: the Liberal Democrat party in 2010.....	151

Figure 6.18: Comparing conversion of votes into seats, 2010 general election	152
Figure 6.19: Summary of bias position after 2010 general election	152
Figure 6.20: Boundary Commission impact	154
Figure 6.21: The effects of boundary changes on future electoral bias	155
Figure 6.22: Summarising the presentation on electoral bias	155
Figure 6.23: Age profile of electoral bias survey respondents	157
Figure 6.24: Knowledge of electoral bias before and after viewing presentation	158
Figure 6.25: Supplementary Vote explained in one minute	165
Figure 6.26: Completing a Supplementary Vote ballot paper	166
Figure 6.27: First vote counts in a Supplementary Vote election.....	167
Figure 6.28: Process of eliminating all except the top two candidates after first count	167
Figure 6.29: Allocating second vote preferences from ballots cast for bottom-placed candidate	168
Figure 6.30: Successful and Unsuccessful vote transfers	169
Figure 6.31: Recalculation of votes for top two candidates.....	169
Figure 6.32: Review of second votes on ballots cast for third-placed candidate.....	170
Figure 6.33: Successful and unsuccessful vote transfers from third-placed candidate.....	171
Figure 6.34: Declaring the winner in the Supplementary Vote election.....	172
Figure 6.35: Summarising the second vote strategy	172
Figure 6.36: Age profile of respondents to SV survey	173
Figure 6.37: Levels of satisfaction with executive summary of SV	175
Figure 6.38: Satisfaction with explanations of how SV determines a winner	176
Figure 7.1: World Mobile Phone subscriptions per 100.....	181
Figure 8.1: The Architecture of Web-based Dialogue System	196
Figure 8.2: Diagram of Web-based Dialogue System Components Dependencies.....	197

Figure 8.3: Table Output Interface.....	203
Figure 8.4: Chart Output Interface.....	204
Figure 8.5: Mapping output interface	205
Figure 8.6: Input and Output UI Interaction	205
Figure 8.7: Example of result of the 2010 general election.....	206
Figure 8.8: Screenshot of the Natural Language Enquiry Template	208
Figure 8.9: Screenshot of the Interface for Enquiry Creation.....	209
Figure 8.10: Example of feedback generated for user enquiry	209
Figure 8.11: Screenshot of Production Rules creation.....	212
Figure 8.12: Help Instructions for NITED.....	213
Figure 8.13: Table Output generated from User enquiry.....	215
Figure 8.14: Histogram and Mapping Output.....	216
Figure 8.15: Two views of UK General Election 2010 results.....	218
Figure 8.16: Step 2 of the DD map algorithm	219
Figure 8.17: Step 3 of the DD map algorithm	219
Figure 8.18: Merging Steps 2 & 3	220
Figure 8.19: Step 5 of the DD map algorithm	220
Figure 8.20: The completed DD map for Harrow East.....	220
Figure 8.21 Examples of DD-Mapping	221
Figure 9.1: Poll Tracker for Sky News	229
Figure 9.2: Poll Tracker using histograms to show single poll results	230
Figure 9.3: Poll Tracker and polls produced by separate polling companies	231
Figure 9.4: Raw Data for Historical Cartograms	237
Figure 9.5: Cartogram of the 1832 General Election.....	238
Figure 9.6: Cartogram of the 1880 General Election.....	239

Figure 9.7: Cartogram of the 2005 General Election.....	240
Figure 9.8: Comparing the 1992 and 1997 General Elections.....	241
Figure 9.9: Opening screen of the ‘Who Should I Vote For?’ application.....	243
Figure 9.10: Policy Choices and Level of Importance	244
Figure 9.11: Manifesto Options on Tax Policies	244
Figure 9.12: Outcome of User choices	246
Figure 9.13: State of the Parties showing Labour Majority.....	247
Figure 9.14: State of the Parties showing Hung Parliament	249
Figure 9.15: State of the Parties showing Conservative Majority	249
Figure 9.16: Comparing winners/non-winners under FPTP and AV	252
Figure 9.17: Elimination of Bottom Candidate in AV.....	253
Figure 9.18: Elimination of Other Candidates and Vote Transfers under AV	253
Figure 9.19: Example of data for AV application.....	256
Figure 9.20: Opening screen of the Alternative Vote Application	257
Figure 9.21: Selection of the 1992 General Election to test effects of AV	257
Figure 9.22: Allocation of Conservative Voters’ 2 nd preference votes.....	258
Figure 9.23: Experiment showing Conservative 2 nd votes transferring to Liberal Democrats	259
Figure 9.24: Experiment showing Liberal Democrat 2 nd votes transferring to Conservatives	259
Figure 9.25: Effect on 2010 House of Commons composition of vote transfers	260

List of Tables

Table 1: Satisfaction with slides relating to bias caused by electorate size:.....	158
Table 2: Satisfaction with slides relating to bias caused by turnout differences	159
Table 3: Satisfaction with slides relating to bias caused by vote distribution	159
Table 4: Summarising the bias position after the 2010 general election:	160
Table 5: Summarising abstention bias after the 2010 general election	161
Table 6: Summarising electorate size bias after the 2010 general election	161
Table 7: Summarising vote distribution bias after the 2010 general election.....	162
Table 8: Number of correct responses to quiz	162
Table 9: Quiz answers controlling for age and interest in elections	163
Table 10: Viewer assessments of knowledge both prior and after watching presentation	174

ACKNOWLEDGEMENTS

Although I appear as the sole author of this thesis, it is really a collective effort. I am indebted to many great people who have contributed to the work herein and supported me personally throughout my time in graduate school.

I gratefully acknowledge the financial assistance provided by the Economic and Social Research Council.

I would like to thank my supervisors, Professor Michael Thrasher and Dr Galina Borisyuk for giving me trust, ideas, and advice throughout the project and for giving me the opportunity to work and collaborate with many wonderful people at the Elections Centre.

A massive thank you goes to my late father, Vladimir Lovitskii, for his constant encouragement and ongoing support during the long journey.

AUTHOR'S DECLARATION

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without the prior agreement of the Graduate Committee.

This study was financed with the aid of a studentship from the Economic and Social research Council.

PUBLICATIONS:

Long, E., V. Lovitskii, M. Thrasher, and D. Traynor. 2009. "Mobile Election." *Information Science and Computing, Intelligent Processing* 9: 19-28.

Long, E., V. Lovitskii, and M. Thrasher. 2010. "Natural Interface to Election Data." Paper presented at the ITHEA, Sofia, Bulgaria.

Long, E., V. Lovitskii, and M. Thrasher. 2011. "Election Data Visualisation." Paper presented at the ITHEA, Sofia, Bulgaria.

Rallings, Colin, Michael Thrasher, Galina Borisjuk, and Elena Long. 2011. "Forecasting the 2010 General Election Using Aggregate Local Election Data." *Electoral Studies* 30: 269-77

WORD COUNT of the main body of the thesis: 58,356 (excluding all tables and figures).

Elena Long
19 December 2012

1 Introduction

1.1 Background

The advantages of presenting data in graphical rather than tabular form are well known. Human brains find it much easier to absorb information from a graphical representation than either plain text or numbers contained within the rows and columns of a table. There are many explanations for why this is the case (see Chapter 2) but this is only part of the story. There are some ways of displaying data that are better than others but implementing these approaches has proved difficult. While this thesis addresses some of the reasons why some graphical displays work while others do not it is mostly concerned about the practical aspects of implementing methods that deliver information in the desired format. It may be that there is a set of rules (a universal algorithm) that when followed will ensure success but we doubt it. Instead, progress in this field follows from trial and error – some methods work, some do not. Some brilliant innovations capture the imagination which other people then copy and improve upon. But what works for one era might not be suitable for another; the current generation has grown up with the Internet and prefers information on a screen than on the pages of a book. The type of graphical display of data that works for one type of audience might not work for another audience.

Visualising data involves passing data from a source through a transmitter, a medium, and towards a receiver. The data may be of constant form, for example, population statistics, health records or in our case, voting records. But everything else inside the equation is changing. As data becomes more widely available it is not only government-appointed

statisticians that transmit information. Academics, journalists, broadcasters arguably began this process but their numbers are now dwarfed by members of the public that bring new insight and techniques for handling data released into the public domain. Technological changes mean that traditional paper has been joined by telecommunication technologies, touch-screens and most recently interactivity between the user and inter-face. The conversation with data has become much broader. Government statisticians were once compiling data for other government departments, partly to justify expenditure and partly to enable future policy planning. When academics began to use data it was largely to speak to other academics and sometimes their students. The problem was that once the data conversation spread much further than this then it became too difficult to control. We did not need Darrell Huff (Huff 1954) to tell us that people lie with statistics because Mark Twain (1906) was attributing to Disraeli the phrase “Lies, damned lies and statistics” as one way in which people used numbers to strengthen their argument.

People that are familiar and comfortable with statistics are largely aware of the tricks that are used to present numbers in the way that satisfies the transmitter. But most people do not fall into this category. Alternatively, the person transmitting the information may have good intentions but simply does not understand the information in front of them. The way in which journalists and headline writers report the findings from opinion polls provides a good illustration of that; unfortunately it is not easy to tell whether the misleading headline comes from ignorance or is deliberate. The danger is, of course, that most of their readers do not know also.

1.2 Dissertation origins

This thesis began with the idea that there ought to be ways in which complex election data could be transmitted to various audiences in such a way that its content could be easily understood. When I joined what was then called the Local Government Chronicle Elections Centre (now simply The Elections Centre) at Plymouth University in 2008 I became part of a team that had as one of its objectives the desire to develop new methods for conveying electoral information to as many people as possible. This had begun in the 1980s when the Centre's Directors, Colin Rallings and Michael Thrasher began publication of their Local Elections Handbooks. At that point it was a huge challenge to collect, collate and produce a machine-readable of more than three thousand county council division results and then to publish them in hard copy. Twenty three years later the publishing of results continued but the technological changes in terms of computer hardware and software over that period had been immense. In 1985 the data were stored in a database located on the University main frame computer; there were no desktop computers. By 2008 even the smallest personal computer contained more processing power than the main-frame had. Computer software was also changing. The standard package used by social scientists for quantitative analysis was SPSS – Statistical Package for Social Scientists. Initially, the process of data input and writing instructions was done using punched cards. Later, remote terminals could be used to 'submit jobs'. Two decades later it is possible to run SPSS on a simple PC. The current version of SPSS will process complex data sets containing hundreds of thousands of cases and provide the skilled user with sophisticated graphical displays.

Yet, it is interesting to read through specialist journals in Political Science and other social sciences and then compare their output with journals in the sciences and engineering. Most social science journals, when they do contain articles that report data, still persist with using

tables of some form (Kastellec and Leonin 2007). There are exceptions, notably the American Political Science Review and the American Journal of Political Science but relatively few UK-based journals use graphics and even fewer use colour (although the main journal, Political Studies, has just started to do this). The gap between what is being published by scientific and social scientific journals is very large in terms of data visualisation.

The research team at Plymouth was unusual because it contained a variety of mathematicians, statisticians, a computer software specialist, a systems engineer with an expertise in database design and political scientists. It was also unusual in simultaneously talking to different audiences – academic specialists, newspaper and broadcast journalists, broadcast computing and graphic designers, private sector industries, especially in telecommunications, administrators, politicians and last but not least the general public. My personal role within this team began as a research assistant/data processor but gradually evolved into becoming more of a designer, a communicator between the different specialist activities that were being conducted within and then beyond the Centre. I worked on different projects, collaborating with different people and making a range of contributions within each team.

Within the Centre the aims were to explore the possibilities of data visualisation through different kinds of media and for different audiences. Beyond the Centre there was a larger network of people with a broader range of skills, including mobile phone technology developers, software engineers, website designers and broadcast interactive graphic developers and designers. In later chapters I describe how people brought these skills together to try and facilitate election data visualisation, sometimes succeeding but at other times failing. In fact, for every ‘success’ there were systems that never got developed and

even for some that did there was eventual failure. What appeared at first to be a straightforward project proved more difficult than anyone could predict.

I have talked about the way in which the development of different media has expanded the possibilities of presenting data. But in writing this thesis I have struggled with the problems of how to describe on paper the interactivity and dynamics of new methods of data visualisation. It is far easier, for example, to show someone on a mobile phone or on a computer how a particular software programme works than to describe it using printed text and screen-grabs. In many ways it would have been much easier to film this thesis than to write it! One further problem that arises with this dissertation is that some of the web-based graphics and software programmes described here no longer have ‘live’ links attached to them so that interested readers could make a note of the internet address and try it out for themselves. With hindsight it would have been sensible to record these programmes in action and then to archive these recordings together with the documentation that describes them. Of course, further improvements in computer hardware and more powerful software capabilities would make these programmes appear increasingly ‘clunky’ but for future researchers there is an opportunity missed when we fail to archive our work properly. Throughout the thesis the reader should assume that when no active link is given and when no secondary source citing the information is available then the original data are deemed ‘missing’.

1.3 Aims and Structure

The aim of this thesis is to discover better methods for communicating voting data (mostly but not entirely election results) to a range of different audiences (from expert political scientists to ordinary members of the public) and then to describe those methods so that others might build upon what has been discovered. Better methods mean using more visual

displays of data than are currently found in political science journals. It is not just political scientists that are interested in elections, however, and much of the thesis describes ways in which the Elections Centre has become more public-oriented in its data output. Some of these methods are passive; they do not require the user to do anything more than observe. Of course, it matters whether the observer is scanning pages or a computer screen – the medium helps to define the content to a certain extent. When graphics have to be published in black and white, for example, but computers can handle millions of colours that tends to constrain what you can do on the printed page compared with what is possible on a screen. Other methods offer more opportunities to the user and allow them to become engaged in manipulating the data. This is a challenge – how can we reduce the complexities of data so that non-expert users can interact with the information and take some kind of control over it. In our experience there has been a trade-off. Sometimes, it is necessary to simplify something so that it might be understood by non-expert users.

The structure of the dissertation is as follows. Chapter 2 describes the evolution of data visualisation methods and shows that human beings have been visualising information for centuries. We are all familiar with the phrase that a picture is worth many words but for many years research into the relationship between cognition and perception has been trying to understand why this might be the case. This has now become big business since data about our eye movements when we view images reveals a great deal about what we are thinking. What is most striking, however, is the rapid acceleration of new methods that are being created by faster computers with bigger storage capacity and faster processing capabilities. There is still a gap, however, between people's ability to create these graphical displays. Access to high-speed computers is not the problem but writing software that is accessible to a broad audience is a difficulty. It is not only access that is a problem. Even if the software

could be written it is still in the user's control to visualise the data and the selection of methods requires skill and knowledge about the data that is being manipulated. Voting data is a very good example of this; the user needs to know what they have before they can use it intelligently.

Presenting data that relates to the field of politics is the subject of Chapter 3. This begins with recording the criticisms raised by some political scientists about the failure to use a broad range of methods for describing data and insisting instead on using tabular data. The interaction between politics and technological change goes further than just the needs of political scientists. The Arab spring demonstrated how mobile phones and social media have become important to the communication of politics and political events. Media such as twitter, Facebook etc. have allowed the general public to participate in politics in ways that were not imagined twenty years ago (Artusi and Maurizzi 2011). Broadcast media have played a big part in broadening the ways in which information about politics is communicated to people and they have incorporated a range of social media into their programmes. Aware that more voters are using social media the political parties, not just in the UK, have started to try and talk to people using these means. What is still important, however, is not only the source of information but also its appearance which is determined by such things as word limits on twitter, the costs of downloading large files etc.

Chapter 4 focuses specifically upon the power of maps to convey large amounts of information in an effective way. It considers how maps have evolved from hand-drawn lines on parchment to computer-generated graphics that can transmit huge amounts of information in the shortest possible time. Cartographers have spent much time considering the best methods for communicating information but once again the advent of computers and computer software known broadly as Geographical Information Systems (GIS) has meant

that many more people can create maps and then populate these images with data. There is a large audience that can ‘read’ and understand maps – most people in the UK are familiar with the physical outline of the country and region in which they live and can even relate to views such as western Europe, The United States of America etc. This means that maps can be readily used by political scientists, newspapers and broadcasters to communicate information about election statistics, for example, although it remains the case that using GIS requires some training and experience to work properly. New ways of generating maps that convey political/electoral information are described. Some of these maps are static, presenting details of interest to the observer. More recently, however, interactivity has been introduced with either the on-screen commentator or the web-based application able to zoom in and out of the maps to shrink/enlarge the area of interest. The chapter closes with an examination of how the media used maps to display aspects of the 2010 general election.

Chapter 5 marks the opening chapter in the second part of the thesis that describes a series of research collaborations where data processing and visualisation feature prominently. In this part the focus moves away from observing recent developments achieved by others (Chapters 3 and 4) and instead through a series of case studies reports on work undertaken by the Elections Centre, either independently or in collaboration. The selection of these case studies for inclusion in this thesis was based on a number of criteria. First, we wanted a study that showed how data visualisation has become a very useful means of communicating details and aspect of data relationships within a team of researchers (Chapter 5). Second, we realised that the internet is becoming an increasingly efficient means for communication ideas and information to the general public. Political scientists are suddenly presented with new opportunities to publicise and disseminate their research outside of the specialised channels. In turn, the scope of political scientists can become more public-oriented, fulfilling a

perceived need for information amongst the public. In this vein, Chapter 6 presents two case studies that used YouTube to broadcast research. The first of these studies reports on advanced research undertaken by the Elections Centre on the operation of electoral bias which was then re-drafted for the YouTube medium. The second of these studies begins with a different problem – how can political scientists seek to answer questions that appear to be of concern to the general public. The specific problem was that in the autumn 2012 the electorate were being asked to participate in elections for new Police and Crime Commissioners using a method of voting known as the Supplementary Vote (SV). It became clear very quickly that most electors were unaware of how SV would work in practice and therefore the challenge became one of devising a video based on Microsoft's Powerpoint software that could 'educate' the public about SV. A third criterion for selecting the case studies was to show that modern techniques in data visualisation are multi-dimensional and require specialist teams to combine and collaborate. Chapters 7-8 report on collaborations with software engineers, while Chapter 9 presents a series of case studies involving collaboration between the Elections Centre and a national multi-platform media organisation.

Having established the reasons that informed the choice of case studies I now provide a short description of the content of each of these chapters. In Chapter 5, I describe how data visualisation directly informed the research process by reporting on work that I was involved with to develop a new model for forecasting national vote shares from local election results. This research was eventually published in *Electoral Studies* (see section on published papers).

One of the greatest strengths of data visualisation is that it facilitates understanding of the effects of changing data. This chapter demonstrates how a sequence of graphs (in this case line graphs plotting vote shares over time after controlling in different ways for missing data)

is used within a specialist research team to determine which method is the best one to use in forecasting models.

The following four Chapters are linked – how to communicate political data to the wider public. These chapters are also sequential, building from users as passive actors towards users as interactive agents. Chapter 6 shows how a new media, YouTube, can be used as the means for transmitting information about two complex aspects of voting systems. The first is a video that describes the operation of electoral bias in a first past the post voting system while the second is a more recent video that describes the operation of the Supplementary Vote system. In both cases the intended audience is the general public and we report on some feedback that we received on these experiments.

Chapters 7 and 8 report on efforts to give mobile telephone and Internet users access to election data that could be user-controlled using natural language interfaces - software that could ‘learn’ from people’s requests. After each user accessed the data the system would remember the syntax used to create a data request and then develop itself so that future requests made in the same way could be satisfied. These experiments demonstrate a sad truth – while political scientists may easily understand concepts like ‘constituency’, ‘vote share’ and ‘percentage change’ that is not the case generally. We conclude from these two experiments that the moment when non-experts can readily access complex-structured databases and use natural language to interrogate those databases is a long way from being realised.

Chapter 9 provides a more optimistic report about advances made when a research centre collaborates with a television broadcaster to create new ways of imagining electoral data and also new methods for user interactivity. The Centre’s directors, Colin Rallings and Michael

Thrasher have been involved with election broadcasting for four decades and have been closely involved with how broadcasters manage electoral data and present it to viewers. The development of the Internet opened up new opportunities with television stations become multi-platform, developing their own web-sites as well as broadcasting programmes. Thrasher's association with BSkyB broadcasting has been especially productive in the development of interactive web-based applications that have aimed to educate and engage the general public. This type of collaboration between elections database developers and broadcast/web designers is certainly one possibility for the future direction of data visualisation in the field of politics.

The conclusions are described in Chapter 10. The opening section of that chapter provides a short overview of the general topic before providing short summaries of the previous chapters. Finally, we provide some thoughts about the future of data visualisation and how it should become more embedded in political science and its research literature and, of course, more widely available for the general public.

2 Visualising data

2.1 Introduction

This chapter begins by discussing theories of perception and neurocognition in order to try and understand the power of pictorial presentation. Visualisation is a complex process that requires significant effort and thinking to produce efficient and effective outputs. One important advantage of visualisation is its use of a specificity of the human visual system to interpret and understand vast amounts of data in a small amount of time. Interpretation of data does not stop with locating individual values in the data, but also identifying hidden patterns, which would be hidden if not in a visual format. The third section reviews some of the most significant examples of data visualisation that have set the standards for graphic design. Many of those standards are the same today as they were when authors began to summarise data in visual formats. These early pioneers included the 18th century Swiss mathematician Johann Heinrich Lambert who was among the first to use graphs to display data and William Playfair, the Scottish engineer who is regarded as the inventor of statistical graphics at the turn of the 19th century. Many of these early developers worked by instinct and were not concerned with developing theories of information visualisation. In the fourth section, we examine some of the literature that tried to develop a theoretical development of data visualisation, identifying the key elements that should be incorporated within good design. We identify the impact of computerization in the field of data visualisation and

examine some important new developments in methods for presenting data. As we stated in Chapter 1, however, data visualisation requires a source, a medium but also a target.

2.2 Perception and Neurocognition

People often say “A picture is worth a thousand words”. Some even cite an old eastern proverb that refers instead to “ten thousand words”. We can also recall the nineteenth century novel, “Fathers and Sons” (1862) written by the great Russian novelist Ivan Turgenev, where the young nihilist Bazarov, who acclaims science to be above anything else, says: “A picture presents to me clearly the same that requires ten pages of written text” (English translation from the original Russian “Рисунок наглядно представит мне то, что в книге изложено на целых десяти страницах”, (Turgenev 2012). But what exactly does science know about the idea that people comprehend information quicker, easier, better if it is presented in graphical form rather than described in words?

Scientists based in the University of Pennsylvania and Princeton University suggest that the human retina would transmit data at roughly the rate of an Ethernet connection, i.e. at the rate of 6–13 Mbit per second (Koch, McLean et al. 2006). Does such incredible speed imply that people absorb visual information more efficiently than information presented through words? Can we say, for example, that people learn and remember visually presented information much better than when this information is only provided to them verbally or that information retention from visual communication is xx times more effective than through words?

There have been a number of experiments that try to demonstrate that visual communication is more powerful than verbal one. Some authors find no significant link between presentation style (verbal vs. multimedia) and recall unless individual preferences are taken into account (Butler and Mautz 1996). However, another source (Lester 2011) reports that according to

the psychologist Jerome Bruner people only remember 10% of what they hear and 20% of what they read, but about 80 percent of what they see. Although the multiple intelligence theory (Gardner 2011) states that there are eight different types of intelligence (tactile/kinesthetic, interpersonal, intrapersonal, verbal/linguistic, logical/mathematical, naturalistic, visual/spatial, and musical) and this defines what is the most effective way to learn new information. However, visual/spatial intelligence is the most dominant type of intelligence and many researchers and teachers believe that most of students learn best visually. For example, the Visual Teaching Alliance website <http://www.visualteachingalliance.com/> cites Martin Scorsese: “If one wants to reach younger people at an earlier age to shape their minds in a critical way, you really need to know how ideas and emotions are expressed visually”, and reports (as well-known facts) that “(1) approximately 65 percent of the population are visual learners; (2) the brain processes visual information 60,000 faster than text; (3) 90 percent of information that comes to the brain is visual”.

Why does it appear that we understand words/numbers and graphs differently? People interpret numbers and graphs differently because they are processed differently in the brain. Numbers and words are generally handled by the verbal linguistic system and graphs are handled by both the non-verbal linguistic system and the limbic system. The bit rate of the visual system is about 10 million bits per second (Koch, McLean et al. 2006) and the rate of reading is approximately 150-400 words per minute. To understand how this works, and provide a foundation for further reading, a very brief review of the relevant neuroscience seems in order.

Below we review some neuroanatomical details of the human brain. However, there are many other important aspects of cognition such as perception, memory, emotional modulation,

integration of signals from different sensory modalities etc. These cognitive concepts are essential and there are multiple theories discussing these cognitive primitives. For example, some psychological research on how we perceive verbal and visual information is based on Double-Coding Theory by Paivio (Paivio 1971) who asserted that the human perceptual system consists of two subsystems. There is, however, some disagreement among researchers on the neuroanatomical bases of processing of abstract and concrete verbal information. Followers of Dual-Coding theory support the assumption that processing of abstract words is confined to the left hemisphere, whereas concrete words are processed also by right-hemispheric brain areas. Theories such as 'psychophysical complementarity' (Shepard 1975; Shepard 1981; Shepard 1984) and 'psychological essentialism' (Medin and Ortony 1989; Averill 1993) believe that our visual perception is sensitive to the structural organization of human memory and that the accurate visual presentation of a body of information is best achieved by data visualisation techniques. Thus, a focused psychological approach to data visualisation must, first and foremost, concern itself with the cognitive description of information, and later with the result of perceptual processes upon the transference of this information (Lee and Vickers 1998). People have an exceptional recognition memory for pictures, and researchers have long established that pictures are superior to words in conveying meanings and patterns (Gorman 1961; Shepard 1967). Other psychologists have sought to measure the extent of visual recognition memory (Hartman 1961; Standing, Conezio et al. 1970). Hartman compared the effects of an audio presentation to a pictorial presentation (e.g. spoken words of the objects were compared with their respective pictures) and the pictorial channel was found to have the advantage - see also (Gorman 1961; Jenkins, Neale et al. 1967; Rohwer, Lynch et al. 1967; Shepard 1967).

First, we describe the main brain structures relevant to processing of verbal information and explain that the verbal processing is relatively slow process (150-400 words per minute) because it based on a sequential processing and information exchange between brain structures. After that we consider a structure of the visual system in the brain (human or primate) and show that although the anatomical structure is very complex, the processing of visual information is much faster because this happens in a massively parallel manner.

There are two main brain regions dealing with verbal symbolic information (words and numbers) (Bear, Connors et al. 2006). Wernicke's area provides syntactic processing of symbolic information. Patients with lesions in this area are able to speak words and operate by their vocabulary but they cannot organise their thoughts into syntactically correct sentences. Broca's area relates to semantic aspects of information and contains dictionaries. Patients with lesions in Broca's area have a problem in finding a proper word but they can somehow make a syntactical construction. There is a bundle of axons connecting Wernicke's area and Broca's area which provides a communication channel for mutual information exchange between syntactic and semantic centres of information processing. Multiple exchanges between these two centres results in a relatively slow processing of the verbal information.

Now we will describe some anatomical details of the visual system (Tovée 1996) and demonstrate that the organization of the visual system is much more complex than the structure of the verbal system. The visual system is distributed around the brain and is organised into hierarchical structures of centres for visual information processing on different levels of details. Using an analogy between signal processing in the brain and in artificial computational devices, it seems that the processing of verbal information is a sequential process with multiple exchanges between two main "processors". Processing of visual

information is organised as parallel information processing by many coupled centres (similar to organization of NVIDIA graphical processor) (Lindholm, Nickolls et al. 2008).

Results of research on the macaque monkey brain shows (Figure 2.1) that its visual system consists of 32 different regions with about 170 reciprocal connections between them (Felleman and Van Essen 1991). First of all, light arrives to the eyeball and reaches the Retinal Ganglion Cells (RGCs) which are located at the back of the ball. RGCs and other neurons of the retina transform the distribution of light on the retina to multiple short electrical pulses called 'spikes' with a duration of about 1 millisecond. The spikes generation by RCG depends on the boundaries between areas with different light intensity. Thus a simple detection of edges is performed by RGCs and coded by spikes. This information is transmitted by the optic nerve to the part of the thalamus called the Lateral Geniculate Nucleus (LGN) which is the first station of visual processing in the cerebral cortex. The LGN and the primary visual cortex V1 further process the edges of objects in the visual scene and start integration of edges into object representation. On this stage of processing visual objects are represented by orientation of small bars indicating the edges of objects and their parts (Hubel and Wiesel 1965). From V1 the coded visual information is distributed in parallel along multiple ascending pathways. Different areas of the visual system are highly specialized in particular types of information processing. For example, some higher cortical regions deal with object recognition. They receive highly processed information which is a result of integration of many particular details which have been considered and analyzed at previous stages of processing.

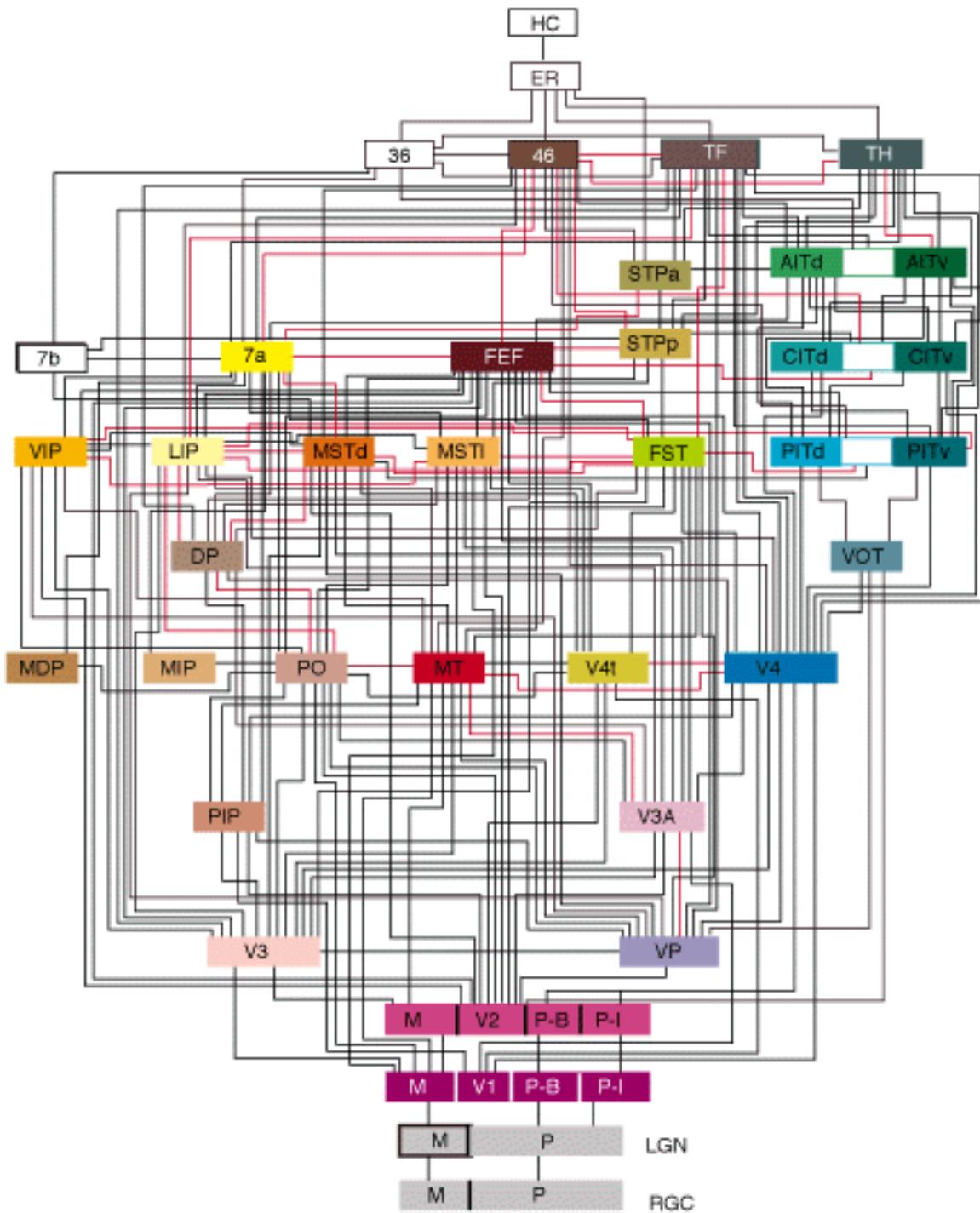


Figure 2.1 Hierarchical structure of visual system

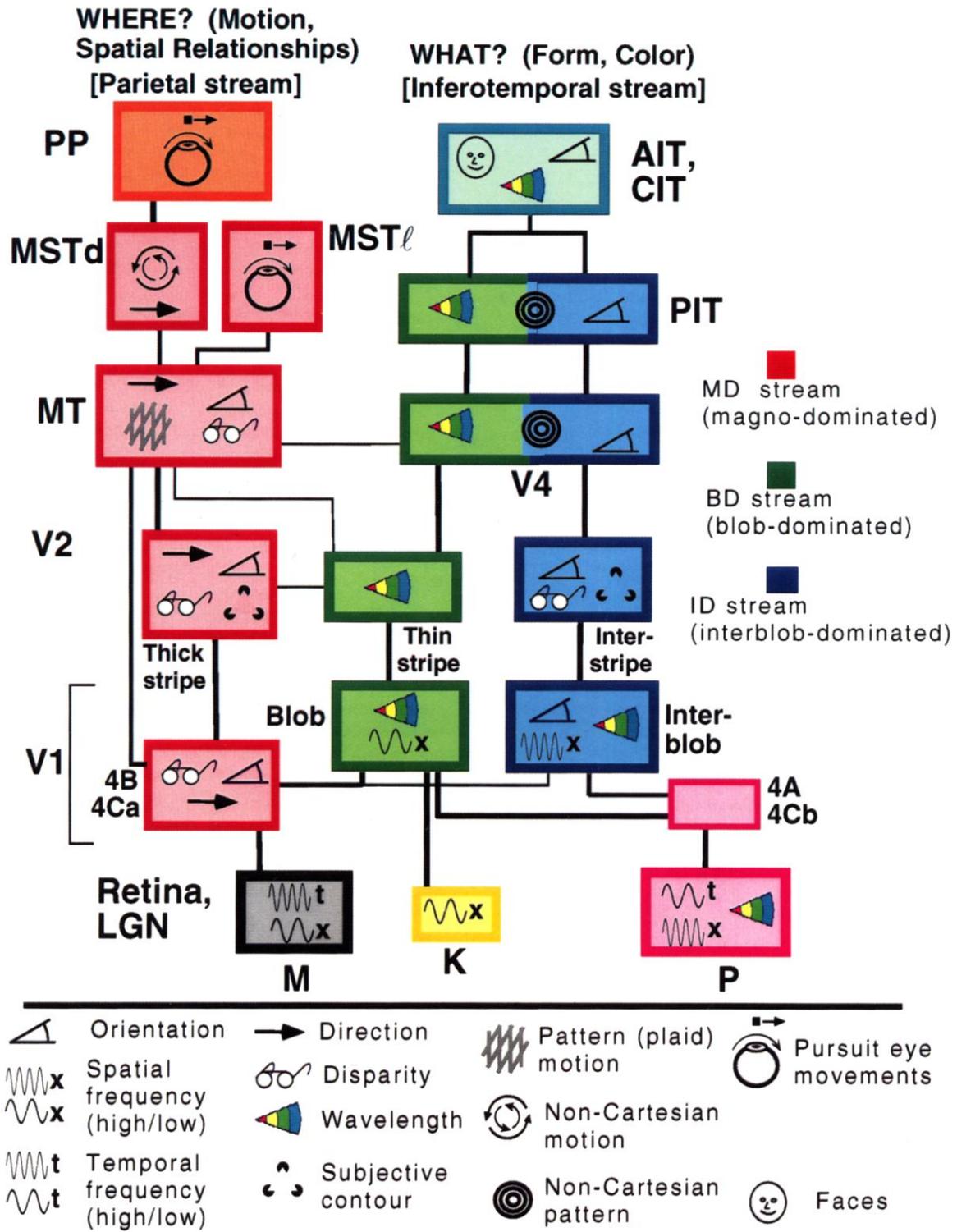


Figure 2.2 Two parallel streams of visual information processing

A diagram of hierarchical structures in Figure 2.1 is complex and difficult to understand. A simplification of this diagram is presented in Figure 2.2 (Van Essen and Galant 1994) where two parallel streams of visual information processing are considered. The ventral or ‘WHAT’ stream propagates information on what (i.e. what objects) is presented in the visual scene. The dorsal or ‘WHERE’ stream transmits the information on locations of objects in the visual scene. These two streams run in parallel from RCG, LGN and V1 to higher cortical areas resulting in recognition and eventual understanding of the visual information (Riesenhuber and Poggio 1999; Serre, Oliva et al. 2007).

Ware’s consideration of pre-attentive visual processing concludes that sometimes the time taken to process information from a visual cue is almost instantaneous (Ware 2004). Being very short lived (their lifespan being about 100msec), much of what we ‘see’ of visual representations is generally discarded before it reaches consciousness. Much of this rapid, unconscious processing involves representations in our conceptual short-term memory (Potter 1993) where small bits of information (such as individual words) are merged into more meaningful structures. However, additional processing stages are required before humans become aware of a particular stimulus and it survives in longer-term memory. Repetition blindness and attentional blink (Coltheart 1999) are some of the issues that cause failure in retaining visual information because of higher-level processing of rapidly presented sequences of visual stimuli, and are thus important to designers of visual analytic systems.

However, it is a challenging question how the brain extracts the visual object from a distributed representation of different features such as local contour details, short bar orientations, shapes, forms, colours, brightness, contrasts, movement direction etc. with a high speed and efficacy. There are several theories how these multiple features distributed between brain regions and locations (Figure 2.2) can be bound together to produce an

impression of the whole image which can be recognized and classified. The most prominent is the temporal correlation theory (von der Marlsburg 1981; Gray 1999; von der Marlsburg 1999). According to this theory the neuronal mechanism of feature binding is based on synchronization of neural activity which propagates from along subsequent cortical areas. The same principle of synchronization can be used to model the visual attention which is important part of cognitive processing.

Selective visual attention is a mechanism that allows a living system to select the most important part of the visual input and ignore other components (objects) of incoming visual information. This mechanism is extremely important because the nervous system has a limited processing capacity and it is extremely important to define and select the most important/significant object and process it in a short time. Focusing attention on the significant part of the visual scene provides a possibility to process the selected object more carefully with taking into account many details of the object (Chik, Borisyuk et al. 2009). An important question of selective attention is how to define the most significant object. There are different possibilities here: to select the brightest object, the most colourful object, the most segregated from the background, etc. Itti and colleagues (Itti, Koch et al. 1998) defined the theoretical bases of an object's 'saliency'. The idea of Itti and collaborators is to combine different features of each object and calculate a saliency map of the visual scene which can be used to navigate attention from one object to another. Using this approach they have developed a saliency-based attention system for rapid scene visual attention. This attention system scans the visual scene and finds the most salient object. Thus, attention is focused and the object is selected from a cluttered visual scene. Figure 2.3 demonstrates how the system works. The red can is the most 'salient' object in this complex environment and

the attention system selects this object (shown by the yellow circle) despite that the can is partially blocked by other objects (Itti and Koch 2001).



Figure 2.3 Focus of attention

As a result of evolution, primate (including human) brains have many highly developed specialised visual cortical areas that perform detailed visual processing. Another important characteristic of a primates' visual system is its ability to demonstrate advanced oculomotor behaviour – primates can use their eyes to identify/select different objects from complex visual scenes and pursue them in a dynamic environment. In fact, the brain combines a control of eye movement (saccades and gaze direction) with signals from attention system to reach an efficient processing of visual information, object recognition and visual scene understanding.

Many studies use eye movements to investigate cognitive processes such as visual search, scene perception, reading, typing etc. (see for example, (Rayner 1998; Stone, Miles et al.

2002; Hayhoe and Ballard 2005; Torralba, Oliva et al. 2006)). Several times per second the eye moves unconsciously in a step-like manner - the gaze 'jumps' to some object/particular area of visual scene, 'freezes' there for a few moments and jumps again to a new area. These very vast and random-looking eye movements are called saccades. Even when an observer looks calmly at fixed object his/her eyes make small but continuous jumps from side to side. It was shown that the direction of unconscious saccadic eye movements has direct connection to cognitive processes. The classic experiments of Russian psychologist Alfred Lukyanovich Yarbus (1914 -1986) over 50 years ago revealed that saccadic eye movements reflect cognitive processes (Yarbus 1967). Yarbus found that eye movements during observation of complex natural objects/scenes depend on the task the observer was asked to perform and therefore reflect cognitive processes. In one experiment, he asked several people to look at the same famous painting of Russian Ilya Repin, "The Unexpected". If the viewer does not receive a particular task, i.e. during a 'free' viewing, the gaze tends to jump from face to face (especially concentrated on the eyes and mouth of the characters). However, when asked to evaluate, for example, the family's financial situation, the trajectories of the gaze reveal that the eyes focus on areas of the image considered relevant to the question. The eye-tracking records clearly show that the subjects visually investigate the picture in a completely different way dependent upon the information they desire. Figure 2.4 shows the original black and white illustration from Yarbus's book – Repin's painting and the trajectories of the gaze during different tasks. Later, Archibald re-produced the original image but in colour (Archibald 2008) – Figure 2.5 presents the original image (a), the image plus eye movement trajectories during free examination (b), and finally the image plus trajectories when the observer was asked to estimate the material circumstances of the family (c).

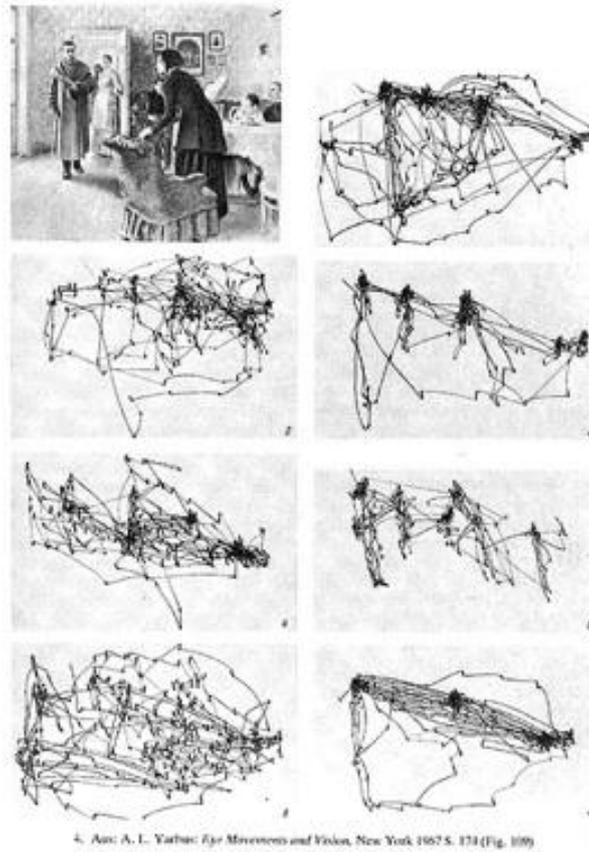


Figure 2.4 Eye Movements: Yarbus, 1965

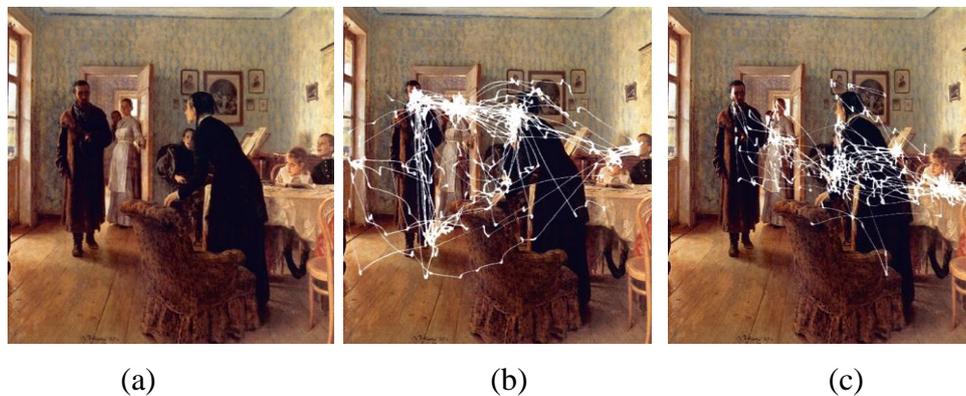


Figure 2.5 Eye Movements: 2008 illustration

Originally, Yarbus made his discovery using a small mirror attached to the eyeball with a rubber disc. The mirror reflects light directed toward the observer permitting a record to be made of tiny eye movements. In recent years, this apparatus was replaced by digital devices.

Eye tracking technologies are now used in many subject areas including scientific research (cognitive science, psycholinguistic, human-computer interaction), medicine (diagnostic of balance disorders of central origin) and commercial applications (assessment of web design and website usability, evaluation of advertising campaigns in terms of actual visual attention). The trajectories of eye movements are recorded and then statistically analysed to provide evidence about which features are the most eye-catching, which features cause confusion and which ones are ignored altogether and ultimately about the effectiveness of a given medium or product.

Results from psychology and neuroscience on perception of visual information reviewed here clearly show that issues of cognition, perception and visual psychology must also be taken into account when deciding how to contextualise, prioritise and present information and support its manipulation. These issues have an impact on the number of dimensions of data that can be usefully presented, spatial positioning and relationships, for example through the use of colour and tone, as well as any dynamic changes in these elements. According to (Ware 2005) the “power of a visualisation comes from the fact that it is possible to have a far more complex concept structure represented externally in a visual display than can be held in visual and verbal working memories” (Ware 2005, p28).

One of the key goals of data visualisation is to communicate information precisely, and to require little effort for the user to understand the data communicated. It is only logical to conclude, therefore, that the human visual system must necessarily be one of the constraints placed on graphics employed in data visualisation. Hence, all information is necessarily compelled by the characteristics of perceptual processing when it is given as a visual display; visualisation techniques that facilitate the flow of information can only be developed when these cognitive processes are fully understood.

There remain a variety of academic views about the method whereby humans understand graphics – for example contrast Bertin and Berg’s view (Bertin and Berg 2010) with that of Spence and Lewandowsky (Spence and Lewandowsky 1990). A theory that is notably relevant to the use and construction of statistical graphs is that of Cleveland (Cleveland 1993). His theory of graphical perception addresses how certain jobs associated with comprehending a graph are performed by the human perceptual system. The perceptual tasks the user can do most easily are those that need to be aligned with the information that is being presented; thus this makes it the goal of any data analysis task. This then provides rules for graph construction: elementary tasks should be used as high in the order as can be successfully achieved. These guidelines are used in a variety of graphs, such as pie charts, statistical maps with shading, bar charts and divided bar charts. Sometimes, the approach to graphical display is grounded in perception but in the case of Tufte (Tufte 1990) it was more about the aesthetic.

Regardless of the theoretical approaches essential to the sense-making process is the perception of patterns in objects. Abstract information, such as political information, could be coded visually and this would in turn create patterns that let the viewer explore and understand the given information; and this in turn can lead to insights that could never occur if the data was studied in any other way.

2.3 The evolution of information visualisation

Visualisation has its historical roots as a way to convey data and as an aid for thinking; from the first maps drawn in the 12th century by the Chinese. The invention of Cartesian coordinates in the 17th century and advances in the fields of mathematics paved the way for data graphics. Since the introduction of data graphics in the late 1700's (Tufte 1983; Cleveland 1993) visual representations of abstract information have been used to explore data and reveal patterns. Information visualisation has its origin with Lambert (1728-1777) and Playfair (1759-1823) who were the first to introduce graphics rather than simply presenting data in tabular form and can therefore be regarded as the inventors of modern graphics design (Tufte 1983; Tufte 2001).

Systematic visual representations replaced tables of numbers towards the end of the 18th century, after Playfair wrote the *Commercial and Political Atlas* in 1786 (Playfair 1786) and the *Statistical Breviary*, (Playfair 1801) which presented graphs and charts in a form that is easily understood by a modern reader. The statistical line graph, pie chart and bar chart; three of the four basic forms were invented by Playfair. Once Playfair had published these new graphical representations other writers began to contribute their own ideas about pictorial representation.

Joseph Priestly, (1733-1804) influenced by Playfair, was the first to create the concept of representing time geometrically. His revolutionary idea was the use of a grid with time on the horizontal axis; and the reigns of different monarchs represented by different length bars which granted instantaneous visual comparison. Similarly, the French physician Jacques Barbeau-Dubourg (1709–1779) and the Scottish philosopher Adam Ferguson (1723–1816) produced plots that followed a similar principle. In Dubourg's case it was a scroll produced

in 1753 that presented a complex timeline from the time of Creation to the present; Dubourg believed this period spanned 6,480 years. Ferguson published a timeline that begins at the time of the Great Flood (2344 BC—though indicating clearly that this was 1656 years after The Creation), and ranged across the births and deaths of all civilizations until 1780. James Playfair, a Scottish minister although no relation of William Playfair, published *A System of Chronology*, (Playfair 1784) in the style of Priestly time-bars.

During the 19th century, various forms of graphs, thematic maps and charts were developed. Florence Nightingale, for example, followed Playfair's examples in order to convey statistical information to a wide audience. Nightingale produced *Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army* (1859), after witnessing the poor conditions in Crimea for soldiers that were injured (Nightingale 1859). For a variety of reasons this was an important publication but it is interesting from our point of view because it featured several graphs which she described as the "Coxcombs"¹ (Figure 2.6) but which are more commonly known as polar area diagrams. The diagram is used to plot cyclical data (here monthly data are being compared) with each month taking 30 degrees of the circle. The size of the radius reports the size of the quantity of interest (In Nightingale's diagram it is mortality statistics) and the separate shadings represent cause of death (the proportions taking up by the grey shading clearly demonstrate that most soldiers were dying from diseases caught whilst in hospital rather than wounds inflicted on the battlefield).

¹The Coxcomb keeps angles constant and varies radius (proportional to square-root (frequency)).

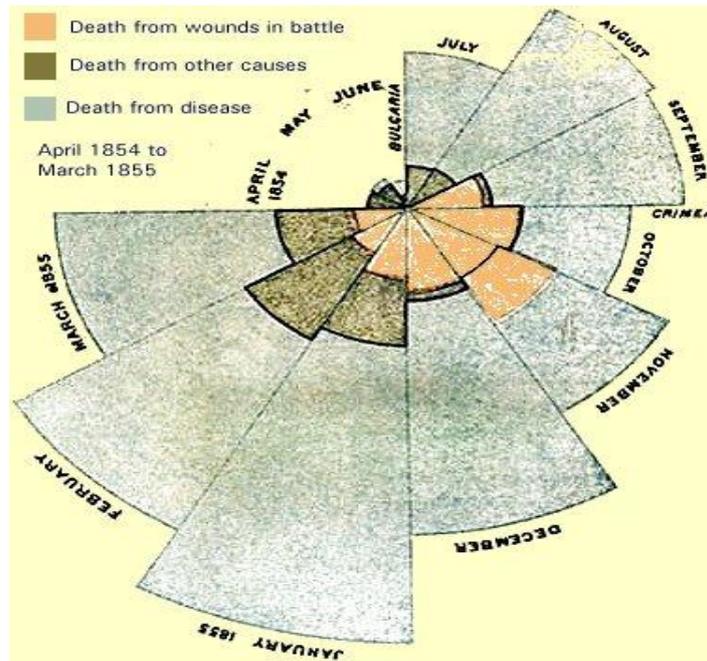


Figure 2.6 A coxcomb chart

The association between mapping information and health statistics was pioneered when Dr. John Snow plotted cholera deaths during an outbreak in central London in 1854 (Figure 2.7). At the time it was widely thought that diseases like cholera were caused by ‘foul air’ but Snow was sceptical of this theory. After talking with Soho residents Snow deduced that the outbreak was centred on a water pump located in Broad Street and the disease declined after the pump was taken out of service. Later, when he was recording the case Snow simply marked the location of water pumps with crosses and deaths with dots on a map of the area.



Figure 2.7 John Snow's Cholera Graphic

Shortly after the publication of Snow's map the French engineer, Charles Joseph Minard produced what is commonly regarded as one of the best statistical graphs ever published (Tufté 1983; Wainer 2000; Friendly 2002). This graphic portrays Napoleon's losses suffered during his invasion of Russia in 1812 (Figure 2.8).

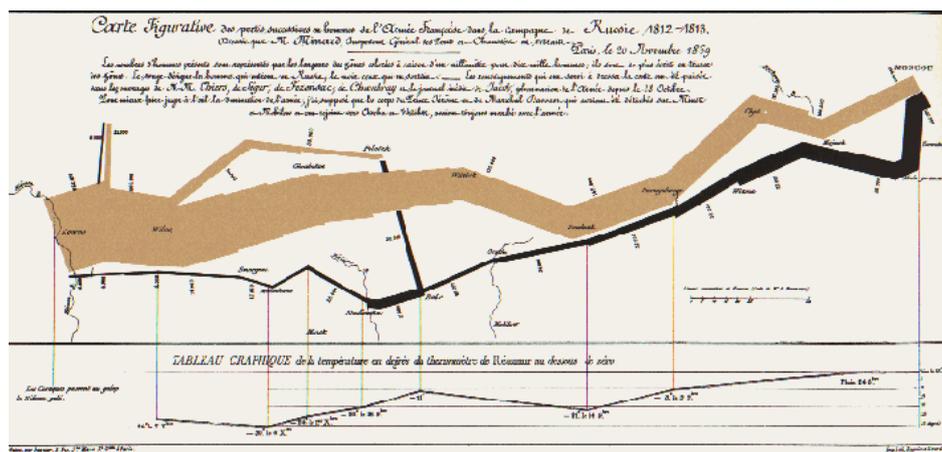


Figure 2.8 Minard's Graphic of Napoleon's Moscow Campaign of 1812

The brown and black lines show Napoleon's army on its advance (brown) into Russia and then its later retreat (black). The thickness of the line shows the army's size where it becomes clear that the army that arrived in Moscow was roughly a fifth the size of the one that began the march. The reader is able to follow the stylized geography because the map shows latitude and longitude and dates are shown at key points of the army's movements. A temperature chart shows what the army had to endure, particularly during its retreat. Although it is clear that Minard's map is brilliant at conveying information about the Russian campaign it is not entirely clear why it achieves what it does achieve. In the following section, therefore, we attempt to discover what lies behind good graphic design.

2.4 Theories of information visualisation

Until recently, the term *visualisation* meant "formation of mental *visual* images" (<http://www.merriam-webster.com/>). It has now come to mean "something more like a graphical representation of data" (Ware 2004). The concept of *information* has also undergone many transformations across time and become more discipline specific (Capurro and Hjørland 2003). The word *information* became extremely influential in all areas of society and fashionable in English and other languages after publication of Norbert Wiener's book on Cybernetics (Wiener 1948) and a paper "The Mathematical Theory of Communication" by Shannon and Weaver (Shannon and Weaver 1949). Although from an information-theoretical point of view *information* can be precisely defined and measured, psychologists and computer science researchers use the term "information visualisation" just to express a representational mode (instead of using verbal descriptions of subject-matter content) which is then employed to display, for example, objects, procedures, dynamics of systems, processes and events in a visual-spatial manner.

Jacques Bertin, is one of the leading theorists of information visualisation (Bertin 1983).

Bertin created a theoretical framework for analysing visualisations, with the identification of various data types (e.g. categories, maps, numbers and networks) as well as retinal variables for visual encoding such as position, shape, orientation, size and colour. Although his work was published before the creation of graphical user interfaces, Bertin appreciated the value of interaction.

Bertin's delineation of a set of fundamental graphic variables (size, location, colour, value, shape, texture/grain and orientation) is perhaps the most famous aspect of his work (Figure 2.9). These variables have been considered and then further developed by other writers as they tried to build the theoretical underpinning of information visualisation (See, for example, (Spiess 1970; Morrison 1974; MacEachren 1994)).

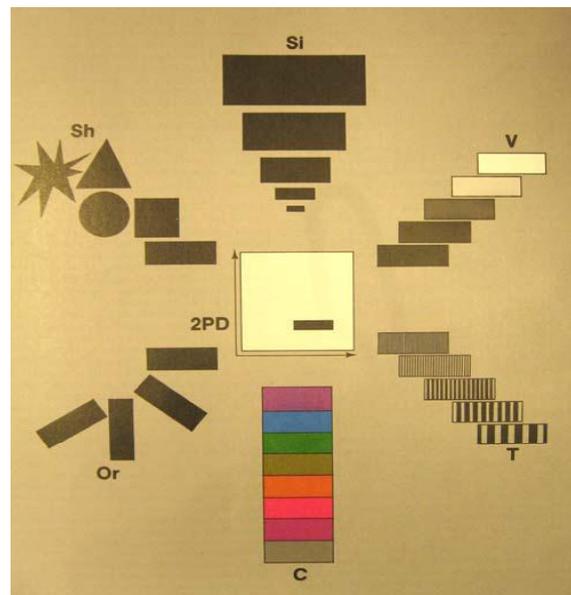


Figure 2.9 The visual variables described by Bertin:

Size (Si), value (V), texture (T), colour (C), orientation (Or), shape (Sh), and the 2D planar projection (2PD)

Subsequent writers further developed the foundations laid by Bertin. In 1983, Tufte published a theory of data graphics that emphasized maximization of useful information (Tufte 1983). Tufte, outlined a number of design guidelines (Tufte 1983; Tufte 1990; Tufte 2006) which included avoiding uninformative elements or so-called chart junk, maximising the ratio of data-ink to non-data-ink and the usage of small multiples displays for multi-dimensional data (see also (Ellis and Dix 2007)).

In 2004 Stephen Few wrote a comprehensive and practical guide to business graphics, - Show Me the Numbers: Designing Tables and Graphs to Enlighten (Few 2004), in which he broadened the principles of colour use developed earlier by Bertin and Tufte. Few outlined nine rules of graphics, which are:

Rule 1: If you want different objects of the same colour in a table or graph to look the same, make sure that the background (the colour that surrounds them) is consistent. (*A gradient background just adds confusion*).

Rule 2: If you want objects in a table or graph to be easily seen, use a background colour that contrasts sufficiently with the object.

Rule 3: Use colour only when needed to serve a particular communication goal.

Rule 4: Use different colours only when they correspond to differences of meaning in the data. (*Adding a different colour for each country to a bar chart adds nothing when you already have identified their names*).

Rule 5: Use soft, natural colours to display most information, and bright colours and/or dark colours to highlight information that requires greater attention. (*He gives an example of palette with eight soft natural and bright highlight colours*).

Rule 6: When using colour to encode a sequential range of quantitative values, stick with a single hue (or a small set of closely related hues) and vary intensity from pale colours for low values to increasingly darker and brighter colours for high values.

Rule 7: Non-data components of tables and graphs should be displayed just visibly enough to perform their role, but not more so, for excessive salience could cause them to distract attention from the data. (*The scales and borders should not visually overwhelm your data.*)

Rule 8: To guarantee that most people who are colourblind can distinguish groups of data that are colour coded, avoid using a combination of red and green in the same display. (*About 10% of men cannot tell red from green, and identify traffic lights by position only!*)

Rule 9: Avoid using visual effects in graphs. (*A plain bar chart is preferable to one with three-dimensional rods.*)

The power of graphic presentation and visual perception had two functions for Bertin and Tufte, serving both as a tool for discovery and to augment cognition. Visualisation theory has also been advanced by theorist-practitioners; for example the American mathematician John Tukey used graphical representations to examine data before conducting confirmatory analysis and used this to advocate more exploratory data analysis (Tukey 1977). A core principle of graphical display is to concentrate attention on what should be seen in the data. Tukey's boxplot displays individual observations in the extremes where they are most likely to be noticed, yet suppresses all points in the middle of a distribution (Figure 2.10).

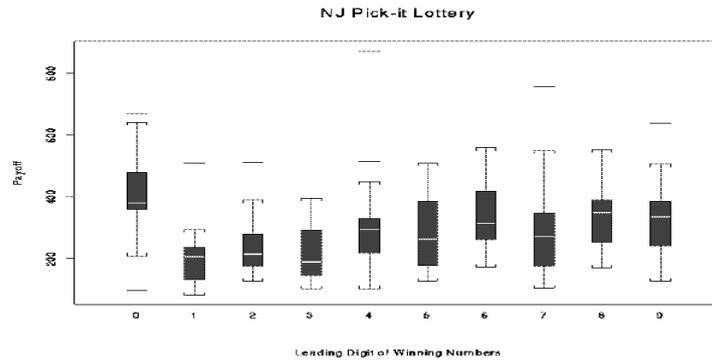


Figure 2.10: Tukey's boxplot

During the 1950s and 1960s computer processing of statistical data was developing with the creation of FORTRAN, the first high-level language for computing. By the late 1960s, widespread mainframe university computers offered the possibility to construct graphic forms by computer programs, some of which were new techniques in graphical display. High resolution and interactive statistical applications would allow users to generate their own graphics from templates incorporated in the software (Fowlkes 1969; Fishkeller, Friedman et al. 1974).

A set of experiments comparing the efficacy of varying visual encodings for conveying different data types were conducted by William Cleveland (Cleveland 1984; Cleveland and McGill 1984; Cleveland and McGill 1985; Cleveland 1993; Cleveland 1994). His conclusions display a rank-ordering of visual mappings for varied data types. For example, area is significantly less effective at demonstrating quantitative values than are either length or position. Another example shows that texture, colour and shape are better for communicating categorical attributes, but are less useful for conveying numerical data.

Taking this research further Mackinlay (Mackinlay 1986) extended and formalized this knowledge into a computational model for the automatic generation of static data graphics. Later, visualisation grew to encompass not only visual design principles, but interaction.

Early work focused on the use of interactivity, such as brushing techniques to select and highlight visualized data points (Becker and Cleveland 1987). Shneiderman (Shneiderman 1996) proposed a list of tasks that visualisations should support and a classification of the different data types that could be subject to visualisation. By excluding uninteresting items, focussing on items of interest, showing an interactive history, highlighting connections between details he discovered the interactive task of getting an overview of a data group. He identified the data types of different dimensions (one dimensional, two dimensional etc.) as well as time, trees, and networks. These ideas are known as Shneiderman mantra: “overview first, zoom and filter, then details-on-demand”. They have been used for many years as a leading methodology and a guiding theoretical framework by many software and web designers for data visualisation and exploration.

The first part of the Shneiderman mantra means that the data overview is first and the most important part of the visualisation process. This idea becomes even more important these days when the internet and cloud computing are actively used for data visualisation. The second part of the mantra stresses that software for data visualisation should include the appropriate tools for manipulation of data, such as zooming, and filtering abilities. The third mantra’s part means that a multi-layered structure of data visualisation should be applied. Data visualisation on some level (layer) of details provides answers to multiple questions and also can “show” answers even to question which have not been formulated yet. That leads to new questions on the next more detailed level and therefore leads to another even deeper layer of visual representation where answers can be discovered, etc. Thus, applying the mantra ideology one can imagine a fractal structure of multiple layers moving deeper and providing a better visualisation with more details which is also is combined with flexible instruments to travel between layers and it will lead to new discoveries from the data.

However, a different approach is taken by Ware (Ware 2004; Ware 2005) that starts with an entity-relationship separation of data, from which more elaborate data types can then be created.

Bertin's research has been used as a model for formalising the modern graphic constructions being created computationally, as the new field of information visualisation (Infovis) has developed. Card and Mackinlay, regarded as authors of one of major publications in the field of information visualisation (Card, Mackinlay et al. 1999), authored an earlier paper entitled "The Structure of the Information Visualisation Design Space" (Card and Mackinlay 1997) in which they expanded on Bertin's taxonomy (adding "connection" and "closure" to the list of retinal variables) and applied this framework to newly developed visualisation techniques. They explored ways in which computer-aided visual representations of data were beginning to split from the traditional semantic models backed up by Bertin's $x y z$ approach. By relying on increasingly indeterminate semantic mapping, which lacked the raw intuitiveness of uncomplicated infographics, modern visual representations of data became progressively less familiar. Card and Mackinlay identified a number of semantic mapping ambiguities in the area of "information landscapes and spaces" and large-scale treemaps (Figure 2.11).



Figure 2.11 Confusion between area and position inside a tree map

Described as “space-filling”, a treemap is a representation of hierarchical data (e.g. the structure of folders on a hard drive) where branches of the hierarchy are shown as successively embedded quadrilaterals whose areas correspond to the “size” of the node they represent (the area would be equivalent to the file size of the folder for example).

The Voronoi treemap (Figure 2.12) is a modern approach to the traditional treemap, and it attempts to solve some of the inherent problems of the earlier examples. Similarly to the network graph, the treemap shows spatial variances where the concrete positions are mostly superficial.

However, many of these public information visualisation designs do not translate for the new information consumer although some types of visualisation have successfully adopted a more public-friendly design: Many-Eyes², gapminder³, and Swivel⁴, are three good examples that have done this but for the most part visualisation design remains accessible only to specialists. Tableau⁵ (Figure 2.16) and Spotfire⁶ are end-user applications for complex visualisations, but they require installation, training and expertise and to date none of them allows for easy publishing to the web.



Figure 2.16 Screenshots from Tableau Software

2.5 Conclusions

Information visualisation will play a major role in the way we communicate data in a society that is becoming increasingly digital. The modern era is characterized by a scale of information production that has not only dwarfed that of previous eras in just a few decades, but also promises to expand further with new technological advancements in telecommunications, computerization and yet to be invented methods for engaging with

² <http://www.many-eyes.com>

³ <http://www.gapminder.org/downloads/applications/>

⁴ <http://www.swivel.com/>

⁵ <http://www.tableausoftware.com/>

⁶ <http://spotfire.tibco.com/demo/>

information. Visualisation has a major role to play in helping us to avoid information overload.

This review of the literature demonstrates a varied approach towards answering the questions – how do we perceive visual images and just how and why are our visual senses such a powerful means of assimilating information. Graphic design has attempted to distil the various elements of size, shape, texture, colour, relationship etc. into an overarching framework that accurately describes the composition of information visualisation.

Psychologists and others, especially those in the neurosciences, have demonstrated the advantages of pictorial displays over other types of information sources, but that is not the same as being able to optimize these for each and every occasion. Information visualisation remains both science and art – we can see what works and what does not work and although we can prescribe some do's and don'ts when it comes down to graphical design we are a long way from providing a template for good design.

Moreover, developments in computer hardware (processing speed, hard disc memory storage etc.) and software (artificial intelligence, interactivity) probably mean that new insights, methods and techniques for information visualisation will continue to be discovered and developed. This suggests that the gap between written text and visual display of information will continue to grow in the latter's favour. The ability for our brains to acquire data efficiently and effectively from even brief exposures to information graphics means that more energy and effort will be channelled in this direction by data providers.

3 Visualising politics

3.1 Introduction

The aim of this chapter is to demonstrate some of the ways in which social science disciplines such as political science and geography have approached the challenges of visualising empirical data when new technological developments create new opportunities that facilitate visualisation. Certainly, within political science there has been considerable criticism from some quarters about the relatively poor presentation of research findings, particularly focussing on the unfriendly nature of that presentation consisting largely of detailed tabulations of data rather than simplified graphical data displays.

This debate comprises the first section before the discussion moves away from academic disciplines and towards the broader social change brought about by the spread of new telecommunication technologies and social media. Changes in these areas permit new ways of engaging with political information – empowered citizens choosing their own ways of discovering about politics. In the third section we show how the political parties in the UK are responding. As people have engaged more with social media so too have the parties learned to adapt their campaigning styles, following voters into new territories by designing new mobile phone applications and web platforms in the never-ending search for votes.

3.2 *Imaging politics*

Naturally, there have been researchers that have directly addressed the problems of political data presentation as well as many more that have thought about the best means for visualising their data findings. A good example of innovation is the development of a graphical presentation of vote share in a three-party system. Upton pioneered this method (Upton 1976; Upton 1991) but (Dorling 1991) took it a stage further, developing a system that uses a single triangle to describe the three-party dynamic of modern parliamentary elections in the UK (Figure 3.1). One objective of Dorling's was to develop an alternative presentation of 'swing' a metric that is most appropriate for a two- and not a multi-party system. These triangles worked most effectively when presented in varying shades of party colours but lost most of their impact when grey-scales were used.

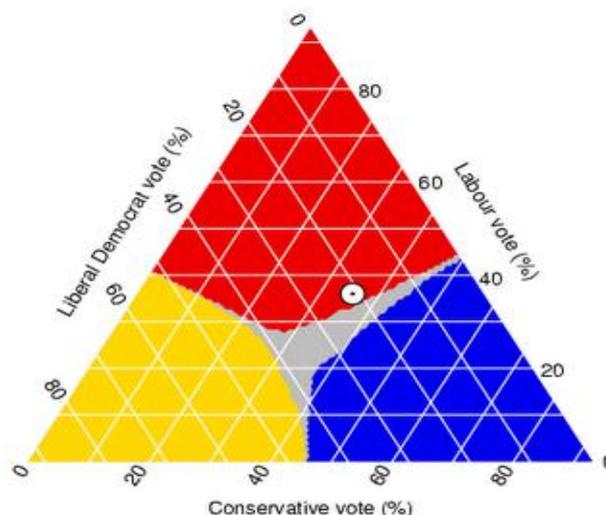


Figure 3.1: Triangular graph of party share

Data visualisation in political science is specific to representing quantitative data as numerical tables and graphs and should thus be set apart from other types of visualisation in this discipline (such as concept visualisation, strategy and workflow visualisation, metaphor visualisation and general information and knowledge visualisation, etc.).

Quantitative information can be presented in a succinct and simple form by using a graphical display. Compared to tabular displays of numeric information, graphs have several advantages specifically when the primary objective is to find the systemic structure that exists across the units of analysis. A graph displays quantitative information by encoding it as geometric construction, but human perception and understanding must be used to decode the given graph and comprehend its implications in regards to the context in which it appears. This process works precisely because the human visual processing system provides a very effective means for understanding complicated information (see Chapter 2).

Moreover, graphs effectively bypass some of the implicit (but important) assumptions that underlie the interpretation of sample statistics by showing all of the data, rather than just providing numerical summaries. Ultimately, graphs encourage interaction between the researcher and the data, because they highlight interesting and unusual features which lead to closer inspection and new insights (Atkinson 1985; Brown, Earnshaw et al. 1995; Bajaj 1999; Hansen and Johnson 2005; Spence 2007; Few 2009; Jacoby and Schneider 2010; Kelleher and Wagener 2011; Yau 2011).

However, social scientists seldom take full advantage of the information available in their statistical results, missing the chance to present their findings to the best effect (Jacoby 1997; Jacoby 1998; King, Tomz et al. 2000; Jacoby 2006).

Several groups of methods that have been employed for data visualisation can be identified within political science:

- Statistical graphics and infographics with extensive use of colour, form, size, shape and style to superimpose many quantitative variables in the same chart or diagram.
- Geographical information systems (GIS) to visualize geographically-linked data.

- Graph visualisation or network maps for representing relations between objects.

Data visualisation in political science increasingly has closer affinities with methodologies developed for geographical information systems since so many aspects of electoral systems and party voting are best described using this approach. In the words of one GIS specialist:

“Space is fundamental to perception and cognition because it provides a common ground for our senses as well as our actions. The constant mutual reinforcement of visual, auditory and tactile cues has developed our spatial cognition to an extent unmatched by any other domain. Perception, manipulation, and motion in space are largely subconscious activities, imposing little cognitive load, while offering intuitive inference patterns. Space has a strong inner coherence that proves useful for designing and combining metaphors.” (Kuhn 1996 p.3)

The value of visual-spatial cognition to information processing can be established by the usage of cartographic maps and information graphics. Both cartographic and information maps (the distinctions are made clearer in Chapter 4) are so common that almost everyone can read or at least gather basic information from them and the software is becoming easier to use and within the range of many more users (Maceachren and Kraak 1997; Monmonier 2007; Farmer 2009; Kraak and Ormeling 2010; Jarvis 2011).

Increasingly, as the application of mapping software becomes more widespread and user-friendly, electoral data is being mapped. For example, the recent General Election in UK brought up a broad range of thematic maps, cartograms, dashboards and information graphics (Figure 3.2) that are used effectively to avoid information overload.

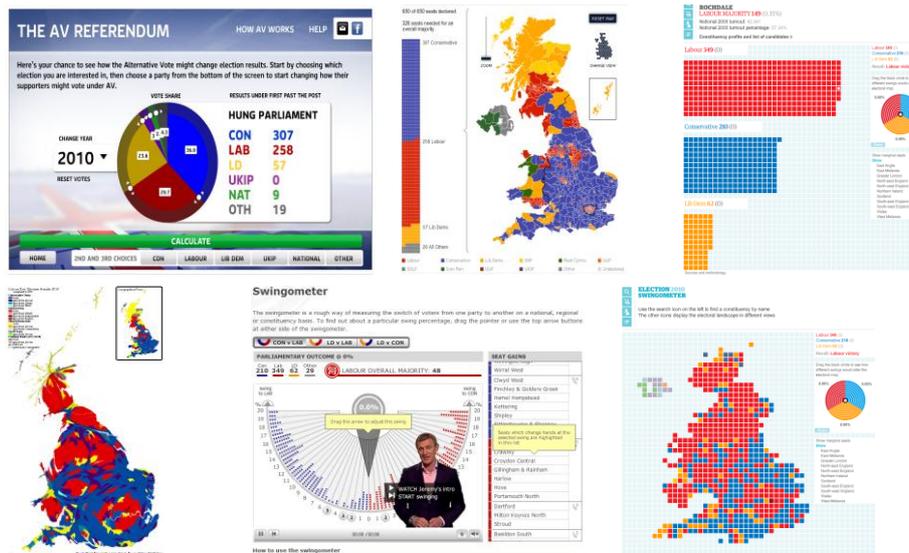


Figure 3.2 Composite screenshot of General Election visualisations

The goal of optimizing maps as vehicles of visual communication has motivated a great deal of cartographic research (Petchenik 1983; Harley and Woodward 1987; Monmonier 1992; Fairbairn, Andrienko et al. 2001; Perkins 2003). The role of maps as “form of abstract thinking” was emphasized by (Muehrcke 1980). The computer allows direct representation of multiple approaches to the same data, viewer interaction with maps, movement and change and the mixing of maps with text, other graphics and sound. Now, several publicly available GIS systems have appeared with Google Earth being the most famous (MacEachren 1998).

3.3 Technological change and political visualisation

Social networking sites such as YouTube, Tumblr, Flickr and Pinterest that are based on image exchange are also becoming more main-stream and this development runs parallel with the public’s ability to take pictures from mobile devices, for example, telephone, personal computers. Much of my research is concerned with optimisation of methods for communicating information to non-expert users. Similar to the Internet, mobile phones facilitate communication and rapid access to information. Compared to the Internet, however,

mobile phone diffusion has reached a larger proportion of the population in most countries, and thus the impact of this new medium is conceivably greater. While it is too early to determine the political effects of mobile phone dissemination, the political events in different countries suggest that mobile technology may come to play an important role in political participation and democracy.

The existence of the visual culture is not solely dependent on the images themselves; it is also based on the human urge to picture and visualise experience in a compulsory way, compulsory because pictures create and contest meanings, while at the same time they relate to other meanings in the public domain (Mirzoeff 1999). Since politics has become mediated, images play a vital role: people experience the political process through media visualisations whether that is through traditional means, for example, newspapers and broadcast news, or more contemporary methods such as media blogs (Bennett and Entman 2001; Louw 2005; Stanyer 2007). New technologies allow users to combine individual information and knowledge to make government processes more visible for the general public; graphics are often considered as some of the most valuable information (Bekkers and Moody 2011).

Two socio-technological developments are responsible for the rapid democratisation of the production and distribution of visual events. The first and most notable of these is the emergence of the internet (Web 2.0), while the second is the arrival of handheld devices capable of recording video and images. In comparison with the traditional media and large scale information systems that rely heavily on a costly infrastructure, the combination of the Web and handheld telecommunication devices have created a culture where pictures taken on these devices are uploaded immediately onto social, interactive networks, thereby making the dispersion of this information virtually instantaneous (Hermanns 2008). This means that it has become possible to have real time coverage of almost any event purely through user-

generated content. Not only does the web stress the value of users as co-producers of relevant content (e.g. information, contacts, experiences and knowledge), but it also shows that users can implement their visual and interactive potential to monitor government communications and services. Cases like the political engagements in Iran and China (2009) and later in Tunisia and Egypt (2011) where social media (Facebook and Twitter) serve as a medium for the development of a social movement is an apparent sign of the use of social media in a democracy. Furthermore, online communication surpasses geographical borders permitting the formation of transnational communities based on shared language, culture, or interests (Michaelsen 2011 p.14). The Internet becomes the valuable means of communication for social movements and groups challenging established power structures (van de Donk et al. 2004). In Western democracies, it was expected to reduce the democratic deficits of corporate-dominated media systems, influenced by consumer culture and linked with powerful elites. As for authoritarian systems, the Internet promised not only to undermine the state's control on information circulation but also to open up new communication channels for suppressed opposition groups and dissidents (Ferdinand 2000).

Mainstream media engage digital techniques particularly during election periods, with a range of strategies aimed at allowing users to interact with content and the data visualisation helps to explain the morass of election data (see Chapters 4 and 9 for a detailed consideration of this). The use of dynamic, real-time mapping and charting, for example, often facilitate the understanding of the state of play in modern coverage of elections.

With new applications of the mainstream media, elections become a time for innovation, and the Internet can lead to growing influence on online engagement (Foot and Schneider 2002; Wicks and Souley 2003; Souley and Wicks 2005; Verser and Wicks 2006; Super 2009). We have seen this with recent elections in both the United States and the UK. From Twitter to

Facebook, through the mobile phones applications and internet polling, technology gave political parties and mainstream media organisations powerful new ways to engage with both voters and viewers (Gibson, Margolis et al. 2003; Stanyer 2006; Gibson 2008; Gunter, Campbell et al. 2009; Gibson 2012). Newman (Newman 2010) argues that social and digital media increased political engagement and appear to have contributed to higher turnout – particularly amongst the 18–24 group (+7% on 2005⁷). Social media and internet activity provided new routes to transparency during this election, contributing to a new and more open political climate.

An example of how the new technologies were engaged in the election campaigns was the Tweetminster website and set of services which helped to make sense of all of the political conversations that were happening on Twitter. In the realm of social media, Twitter is of indelible concern to democratization for two reasons: it is public and it is virtually unstoppable. Twitter updates, or “tweets,” go out on two networks—the Internet and SMS, which is used for text messaging systems. Additionally, Twitter is a broadcast service, unlike Facebook, and can therefore allow for messages to be accessed by virtually anyone (Grossman 2009).

During the campaign, it maintained the definitive lists of all of the MPs and parliamentary candidates, monitored the level and type of political activity and described trends through links and especially data visualisation (Figure 3.3).

⁷ MORI study showed 44% turnout amongst 18–24 year olds. Turnout increase was highest in this group, which is most active in social networks (www.ipsos-mori.com/researchpublications/researcharchive/poll.aspx?oItemId=2613&view=wide).

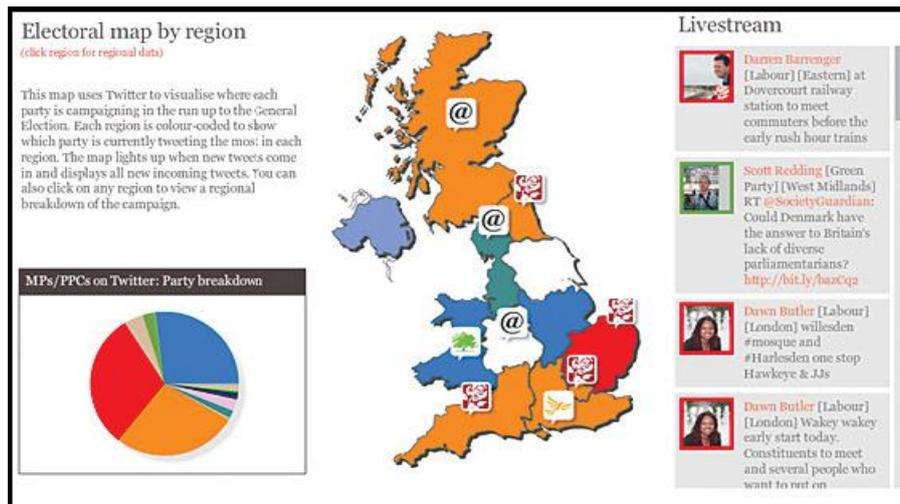


Figure 3.3 Tweetminster⁸

The image above (Figure 3.3), which appeared on the front page of the website, visualises which parties are more active on Twitter in particular regions of the UK at any particular time. In this case, during the second week of the campaign, it showed Liberal Democrat activity in the South West and Scotland. During the third and last debate between the three party leaders of the Conservative, Liberal Democrat and the Labour parties 154,342 tweets appeared containing various terms from the debate, with a frequency of arrival at 26.77 tweets per second from 33,095 different people.

Election 2010 produced higher levels of participation and creativity online than happened at previous elections (Gibson 2012). This is not surprising given the rapid growth of new social media tools. YouTube and Facebook first appeared in 2005, while Twitter had not yet appeared on the social scene. According to the Oxford Internet Survey (2009)⁹, participation in online social networks doubled in just two years to almost 50 per cent of UK internet users. Around 20 per cent of all time spent on the internet is now with these networks¹⁰.

⁸ <http://tweetminster.co.uk/>

⁹ Oxford Internet Survey: 49% of internet users had created or updated a social networking profile in previous 12 months (www.oii.ox.ac.uk/microsites/oxis).

¹⁰ Nielsen figures from 2009 say one in every six minutes in the UK (<http://blog.nielsen.com/nielsenwire/global/social-networking-new-global-footprint>).

Half a million people participated in a Facebook poll¹¹ on the eve of the campaign (Nick Clegg won) and 1.87 million people shared the fact that they had voted by the time the polls closed, using a real-time counter placed on every page of the UK site (see Figure 3.4).



Figure 3.4 Facebook polling day

The growing influence of new information and communication technologies, in particular mobile phone technology, on many aspects of life has been noted, but detailed analysis of possible effects on politics has begun only recently. While the possibilities of e-voting are still being explored (although electoral fraud and transparency of election counts remain concerns), the political influence of mobile phones can be noticed in the wider context of democracy, namely the building of networks, the provision of information and the mobilisation of activists. They are also being increasingly used to spread information about the progress of election counts despite their usage apparently being illegal.

3.4 Democratizing broadcasting

YouTube has become an important channel in which parties could broadcast messages to bypass the restrictions of mainstream media and the regulated television environment (Anstead and Chadwick 2009). The most popular video on the webcameronuk channel was

¹¹ <http://www.facebook.com/democracyuk>

Cameron's four minute speech, posted just three days before the General Election¹². Publicly available statistics on the video show that of 193,000 total views around 170,000 came before May 5th (a two day period). The second most viewed video on the YouTube ([webcameronuk 2010](http://www.youtube.com/watch?v=aQC1oiJdtbs)) tells a similar story. Posted on April 19th following the first leadership debate, 'David Cameron: What it takes to change a country' registered 168,000 views with around 160,000 coming in the first few days and around 99,000 of those coming from paid advertising online. A total of 21 videos were posted by webcameronuk during the official campaign timetable. The Labour Party had a moderate budget at the 2010 election so almost all of its YouTube views were not based on advertising. The most viewed post was the interactive video manifesto featuring an 'ordinary' cartoon family and was posted on April 11th¹³ and watched by 140,000 people. The second most viewed video¹⁴ on the channel was the video 'Brilliant Britain' that points out how the Conservatives had increased funds for spending on advertisements. A total of 36 videos were posted by Labour during the 2010 campaign. Viewing statistics for videos posted by the Liberal Democrats were always lower than those registered by the two main parties. The video that proved the greatest success was a Party Election Broadcast uploaded to YouTube¹⁵ which gained 40,000 views.

3.5 Electioneering and mobile phone technologies

The rapid growth of mobile phones since the 1990s, the emergence of text messaging, and the increasing interconnection between mobile phones and the Internet have made it possible for people to coordinate and organize political collective action to a degree never seen before (Gibson and Ward 1998; Gibson, Margolis et al. 2003; Gibson 2008). The political influence

¹² <http://www.youtube.com/watch?v=aQC1oiJdtbs>

¹³ Labour Manifesto 2010: A future fair for all: <http://www.youtube.com/watch?v=SCO-KwYpH0M>

¹⁴ Roughly 135,000 views

¹⁵ Liberal Democrats: Say goodbye to broken promises : <http://www.youtube.com/watch?v=jTLR8R9JXz4>

of mobile phones can be viewed in the building of networks, the spread of news and information and the mobilisation of activists.

This form of new media has applications for modern electioneering. During the General Election 2010, all the parties invested in mobile phone technology, and provided their supporters with means to connect online in different ways.

The Conservatives were the only party to develop web content optimised for mobile, which renders well and allows easy navigation of the key content (Figure 3.5).



Figure 3.5 Conservative Party iPhone app

News feed and the web stories from the party were available by way of a plainly formatted headlines page. A scrollwheel allowed users to explore the details of Conservative policy. The Conservative mobile phone application even contained a Swingometer. Tilting the iPhone made the country go progressively bluer until the Conservatives had a majority and could form a majority government (it was impossible to tilt in the opposite direction to obtain a Labour victory). The application main menu also included links that took users to Conservative channels on Twitter, Facebook and YouTube.

Labour also introduced an iPhone app (Figure 3.6), which follows a standard lay out, providing content and assets to mobilise grass roots supporters.



Figure 3.6 Labour Party iPhone app

Unlike the Conservative equivalent, Labour had devoted space to a consideration of the dangers of voting for other parties; after all they were the defending party and were concerned about the level of protest voting that might take place. Labour's Twitter feed was more closely integrated into their application than was the case for the Conservatives.

The Liberal Democrats adopted another approach with their own mobile phone application (Figure 3.7). The opening screen invites the user to pick three issues, and 'mix their video'. As soon as the third issue is selected, a video starts playing, featuring leader Nick Clegg, and tuned to playing the segments appropriate for the interests the user has expressed.

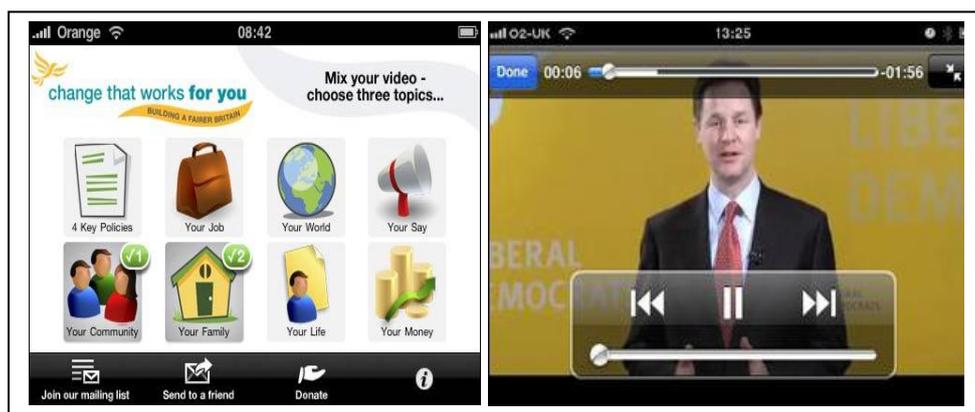


Figure 3.7 Liberal Democrats iPhone app

Other options available with the app included registration with the party's mailing list, forwarding the app to a friend, or making a financial donation to the party.

3.6 Conclusions

Despite the complaints from some political scientists there does not appear to have been a substantial change in the way that most quantitatively-based research in political science (perhaps less so in the case of geography) reports its research findings. Opening the pages of most of the discipline's empirically-leaning journals still reveals that describing political research comprises tables and then more tables. It is still rare to encounter maps, graphical displays of data findings.

The printed page lags further and further behind the capabilities of new technologies and the ways in which people now communicate with one another about political events. The enormous growth in social media such as Twitter and Facebook has seen those technologies turned towards the spread of political information. Whilst academics have not yet apparently altered their behaviour the pressures on the parties to be more imaginative in using these technologies are greater. If competitor parties build a new method for communicating with voters across social media then each party has to do the same. These kinds of developments began in the United States but were more visible at the recent 2010 general election in the UK.

These debates and trends are interesting but it is also interesting to note that as yet the pre-eminence of maps as conveyances of political and especially electoral information have yet to be seriously challenged. In the following chapter we focus on the ways in which mapping data became and remains easily the most accessible means for engaging with politics, especially electoral politics.

4 Mapping data

4.1 Introduction

This chapter is entirely concerned with mapping as the pre-eminent means of visualising political and especially electoral data. The previous chapter revealed the relative lack of progress in visualising data within the discipline of political science, particularly in comparison with the closely-related geography discipline. The development of new methods for mapping data, especially in the production of GIS, had helped to bring these technologies within the reach of many geographers whereas it was once a specialised skill that was reserved for cartographers. Maps rightly have a special place in the field of data visualisation for a number of reasons that will be outlined and discussed in this chapter.

The chapter begins by providing some of the general principles that lie behind the map as an example of visual display and then moving into a much greater examination of maps within the political/electoral setting. There are certain elements within an electoral process that maps have to address – the size and shape of electoral districts, the patterns of party competition, the distribution of votes and seats to name just a few. Technological developments in recent decades have resulted in a transformation of the mapping process. Two of these, interactive mapping and the growing use of dashboard graphical displays that are centred on mapped data are the subject of separate sections. One of the continuing

discussions within the map-making field is about the best form of map to use. Of course, the answer to this question depends on what is the map's purpose and who is the intended viewer. This consideration has particular importance for electoral data where electoral districts are different in terms of physical size and electoral weight (number of legislators elected from the district/state etc). There is a section, therefore, that considers a range of different map types and their fitness for purpose. The closing section provides a short case study of the ways in which media coverage of the 2010 general election employed a variety of mapping procedures to try and engage with viewers and readers.

4.2 Mapping: optimising data visualisation

Maps provide excellent means for describing similarities and differences across specific objects. Maps are a way of translating or converting spatial data - "cartographic visualisation can be claimed to solve many of the fundamental problems identified in studying spatial social distributions" (Dorling 1991). While statistical methods are used to explore patterns and relationships in the data, maps illustrate and support the analysis of spatial data in an easily comprehensible style. As a combination of abstraction and representation, cartography is an incomparable method of a data-handling and data-communicating. The interaction between maps and statistical texts (graphs and charts) is vital to the understanding of scientific phenomena (Cockerill 2003).

Many cartographic technologies are now available, from geographic information systems (GIS), which offer a broad range of analytical functions and integrate map components, to high-end cartographic information systems (CIS) for professional map and atlas makers (Anselin 2012). Over time, conceptual and technically-driven developments in computer graphics, computation and user interfaces have begun to transform the map into something

that is displayed within a visual toolbox to be used interactively for exploratory data analysis (including multiple representations such as statistical charts, three-dimensional plots and other visual aids to understanding that collectively may be regarded as geovisualisation (Hubbard, Kitchin et al. 2008).

Conventional maps are effective at representing geographical information such as locations, areas, distances and topological relationships by transforming the surface of the sphere using different map projections and have dominated the history of cartographic visualisation. The most common way of adopting conventional maps to describe socio-economic information is to build up thematic maps such as dot maps, proportional symbol maps and choropleth maps, which employ symbols, shadings and colours to represent socio-economic data overlaid on geographical features.

Digital maps are now transmitted over computer networks. The web very quickly became the spatial data delivery mechanism of choice (Masser 1997; Trainor 1997). The combination of maps and the Internet is a significant development, not only for improving distribution but also because it makes a more interactive form of mapping possible. Trends are also evident in the types of maps available through the Internet, many of which are becoming interactive, allowing users to use simple queries and zooming. Typically, query functions gave additional information on map objects, the coordinates of a selected point, as well as locator functions. Interactive visualisation methods can increase cognitive resources by providing an additional, external visual resource to the human memory. They may reduce the amount of searching and ease the recognition of patterns as well as enhance understanding of relationships in large amounts of data and information. In addition, they provide a medium that enables the user to have a representation of information that he or she can quickly and easily modify, restructure or consider from a different perspective.

The different media (maps, diagrams, pictures, text, video, animation and sound) applied in cartography are used to inform about spatial objects, their relationships, and spatial processes. This information process can be efficient only, if the cognitive aspects in media design are considered. A good map does not just instil data in its geographical context, but also serves as an instrument of generalisation and summarisation, helping a user to see insight. In the field of visual thinking the most important function media has is constructing mental models and transforming information into a greater context. It has to give insight into the subject's complete structure; it should support the creation of mental models of the subject's correlations and processes. Maps, abstract graphical presentations, and animations can convey this knowledge. Whilst an air photograph or satellite image may be data rich, a well designed topographic map may effectively present abstract information that reveals unseen patterns not directly visible in the landscape. Interaction with a thematic value-by-area map, such as a representation of election results combined with socio-demographic data, may draw attention to hidden patterns that were not evident in spatial representations of the data (Andrienko and Andrienko 2007; Andrienko, Andrienko et al. 2010).

Digitising maps provides new challenges that cartographers have not previously considered.

These include:

- Visual dynamism, animated and updated in real-time to data changes.
- Multimedia, the integration of text and image with animation, video, and audio.
- Interactivity, the opportunity for the user participate in the creation of the map by changing it.
- Network distribution and data gathering.
- 3D and virtual environments.

The degree of complexity of what and how the representation is achieved is substantially greater than with static maps.

When cartographers create thematic maps, they need to consider several important things. The most significant - is the intended audience, since this determines what items should be integrated on the thematic map as reference points in addition to the map's theme: "If we then make the obvious assumption that the content of a map is appropriate to its purpose, there yet remains the equally significant evaluation of the visual methods employed to convey that content" (Robinson 1952 p.17). A classification of representation forms and their potential application areas should be readily available to those exploring geospatial data (Fairbairn, Andrienko et al. 2001). Associated with the application of an appropriate representation method, it is important to ensure that the suitable level of data abstraction for that representation is displayed. Further, it is vital that, once displayed, users are also able to navigate and effectively assimilate all of the data that are available using interactive graphical tools.

4.3 The visualisation of election results

Maps that are used to visualise some aspects of the electoral process normally provide a broad overview of the varying democratic choices across different regions and social contexts. The most common maps of politics are univariate choropleth maps for instance, presenting party or candidate votes (in percentages) across different electoral districts as a single theme. The more complex multivariate thematic maps have more than one layer of data. Quite often election results are used in combination with a second socioeconomic data layer in order to suggest graphically potential correlations between socioeconomic variables and voting behaviour.

For this chapter particular attention will be focused upon the American presidential elections.

The outline shape of the United States lends itself to map visualisation projections

(rectangular) that does not apply so well to the UK (long and thin).

The United States started the trend of reporting election results on the Internet using visualisations. The image of blue and red states, after the 2000 Bush-Gore elections and the 2004 Bush – Kerry elections (Figure 4.1.) gained widespread use in mainstream political discussions, and has assisted the way Americans view politics and society in their country, indicating the efficacy of this visualisation.

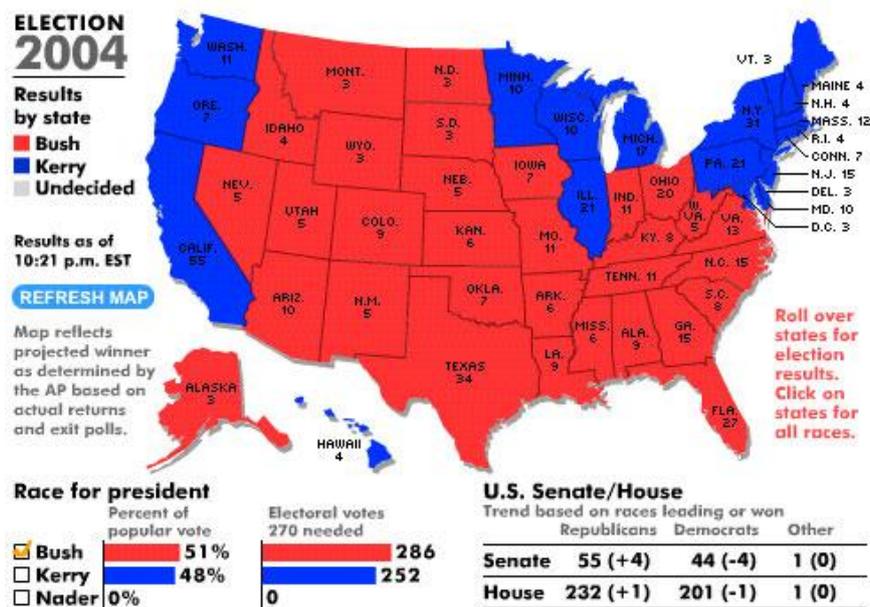


Figure 4.1 USA Today and the 2004 Presidential Election

Election results are typically represented by colouring each political unit according to the candidate that carried it. This approach is problematic because the visual weight of each political unit is based upon the physical area of the state and not the population which is what determines its actual weight within the electoral college. As a result, maps showing US presidential elections often are dominated by one colour, as one party tends to carry vast but sparsely-populated rural areas while the other party tends to carry small but largely-populated urban areas. The overall visual weight of each party is thus not proportional to the actual

election results – as later chapters show this is also a problem for the UK where urban constituencies are physically small and normally coloured red while rural seats are much larger and coloured blue, creating false impressions that most seats are won by the blue party. Though scaling the voting process to physical geographic space and spatial aggregation are inherent weaknesses of red and blue state mapping, the power of such maps in affecting political perceptions is conspicuous. Despite the fact that election results can be displayed at finer scales that better capture spatial variability, the problem of projecting a human activity onto a map representing geographic area remains and many different solutions have been attempted (Figure 4.2).

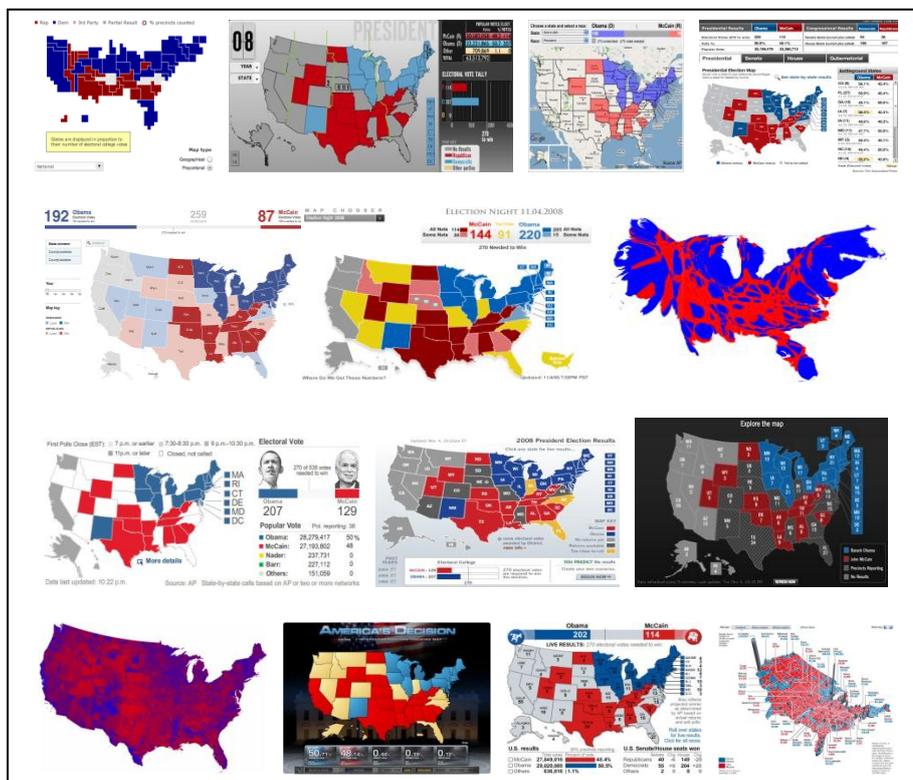


Figure 4.2: Mapping electoral data across the United States

Subsequent US elections have seen further progress being made in terms of compiling static maps that provide users with substantial political information. The 2008 election was the most polled presidential election ever. In these elections, the citizens had an opportunity to use the maps to find and measure political statistics (raw votes cast, percentages of votes cast

for parties, population by race, third-party voting) for any county in the US. Similarly, the maps allowed users to analyse how particular counties compare to the rest of the state.

4.3.1 Interactive maps

Interactive maps on the Internet present data most effectively when they invite action from the user. Showing relationships between data is easier when the user has the power to change the visuals. It is only with interaction that user could really find out about a data domain, explore it from different points of view, and manipulate the medium to accommodate any new questions that come to mind (Andrienko and Andrienko 1999; Jenny, Jenny et al. 2010). With online interactive maps, the interactive navigational techniques such as dragging, zooming, selecting, rotating, filtering and tracking helps to reveal relationships between data and content. Active navigation means the user can search for significant visual or spatial patterns which may provide important information, and allow the user to shift their cognitive process from reading and reasoning to the mental processes of visual and spatial cognition. Interaction techniques can be categorized based on the effects they have on the display. Navigation techniques focus on modifying the projection of the data onto the screen, using either manual or automated methods.

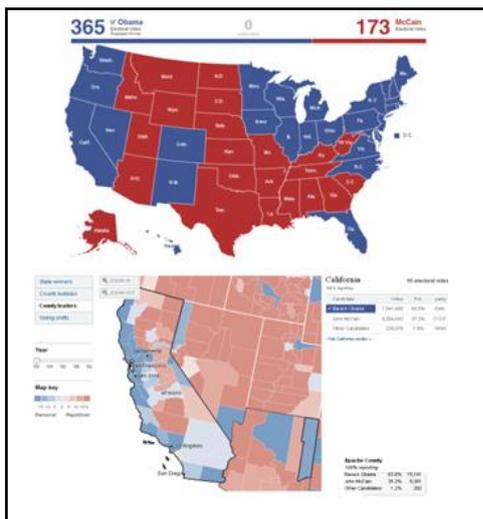


Figure 4.3 The New York Times map

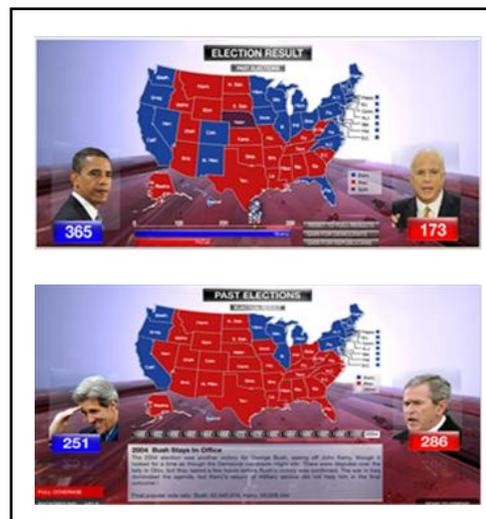


Figure 4.4 The Sky News US Election map

The New York Times map¹⁶ (Figure 4.3) allowed users to zoom in to individual states and counties, and compare 2008 results to results from the past four elections. Zooming is a well known view modification technique that is widely used in a number of applications. In dealing with large amounts of data, it is important to present the data in a highly compressed form to provide an overview of the data, but at the same time, allowing a variable display of the data at different resolutions. Zooming does not only mean displaying the data objects larger, but also that the data representation may automatically change to present more details on higher zoom levels. The objects may, for example, be represented as single pixels at a low zoom level, as icons at an intermediate zoom level, and as labelled objects at a high resolution.

Selection techniques provide users with the ability to isolate a subset of the displayed data for operations such as highlighting, filtering, and quantitative analysis. Selection can be done directly on the visualisation (direct manipulation) or via dialog boxes or other query mechanisms (indirect manipulation).

The Sky News US Election map¹⁷ (Figure 4.4) projected the outcome according to the latest opinion polling. Users could check the candidates' biography or choose the state to find out more about it: the make-up of the population, and how the state has been represented in the Senate and Congress. Another screen, the Past Elections section gave the users detailed election results since 1984.

¹⁶ <http://elections.nytimes.com/2008/president/whos-ahead/key-states/map.html>

¹⁷ <http://news.sky.com/skynews/Interactive-Graphics/US-Election-Map>

4.3.2 Choropleth Maps

While cartographers use many different methods of spatial representation of social phenomena a commonly used method is provided by the choropleth map (Figure 4.5). The choropleth is a map that portrays quantitative data as a colour and can show density, per cent, average value or quantity of an event within a geographic area. Sequential colours on these maps represent increasing or decreasing data values. Normally, each colour represents a range of values. Choropleth mapping is an essential instrument of graphical exploratory data analysis of geo-referenced statistics (constituencies, counties, nations, etc.), and colour variation is the symbolisation of choice (Fairbairn, Andrienko et al. 2001 p.25). Colour selection as a visual variable was initially developed from a cartographic perspective by Robinson in the 1950s and later by Bertin and Few (see Chapter 2). Such work has resulted in the standard application of three variables – hue, value and saturation of colour – in the representation of data (for example nominal or ordinal data). Humans can discern twelve different levels of hue and saturation; when combining size, brightness, and hue, the number of different levels increases to seventeen, and when subjects were asked to identify the location of a dot in a square, they were capable of differentiating about 24 different levels (Tegarden 1999).

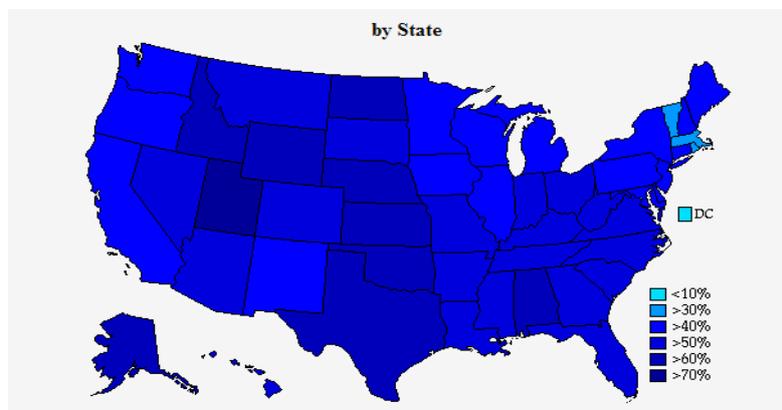


Figure 4.5: A choropleth map of United States

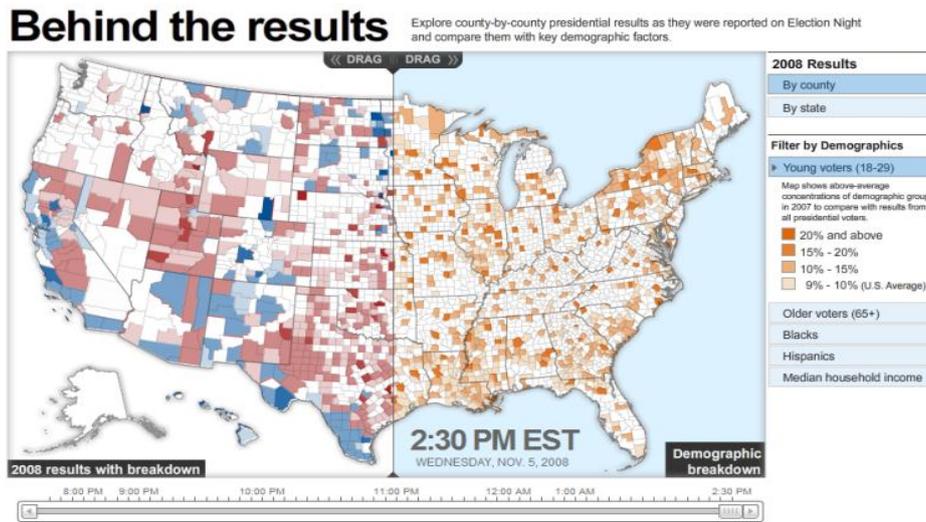


Figure 4.6: USA Today Presidential Election 2008

In the USA Today map¹⁸ above (Figure 4.6) viewers could scrutinise individual state detail from all fifty states. The map allowed users to compare the percentage of voters in a demographic category with how that demographic category voted. The demographic options included older voters (65+), younger voters (18-29), Blacks, Hispanics, and median household income. Interactive filtering is a combination of selection and view enhancement. In exploring large data sets, it is important to interactively divide the data set into segments and focus on interesting subsets. This can be done by a direct selection of the desired subset (*browsing*) or by a specification of properties of the desired subset (*querying*). Browsing is very difficult for very large data sets and querying often does not produce the desired results, but on the other hand, these tools offer many advantages over traditional controls. Therefore, a number of interactive selection techniques have been developed to improve interactive filtering in data exploration.

¹⁸ <http://www.usatoday.com/news/politics/election2008/results.htm>

4.3.3 Proportional or graduated symbols maps

Proportional or graduated symbols are the next type of map and represent data associated with point locations such as cities. Data is displayed on these maps with proportionally sized symbols to show differences in occurrences. Circles are most often used but squares and other geometric shapes are suitable and have been used. The commonest method for sizing these symbols is to make their areas proportional to the values to be depicted with mapping or drawing software.

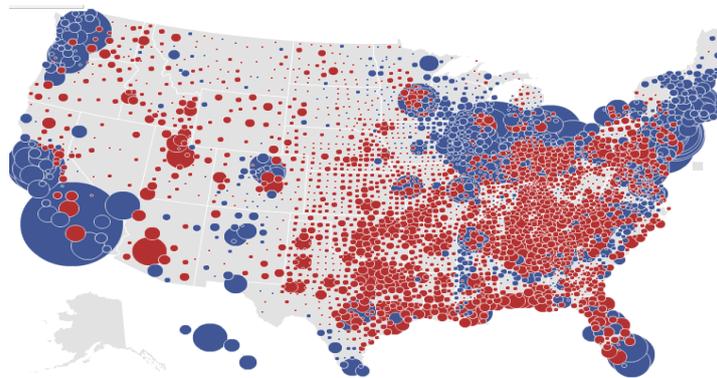


Figure 4.7: New York Times and the Presidential Race

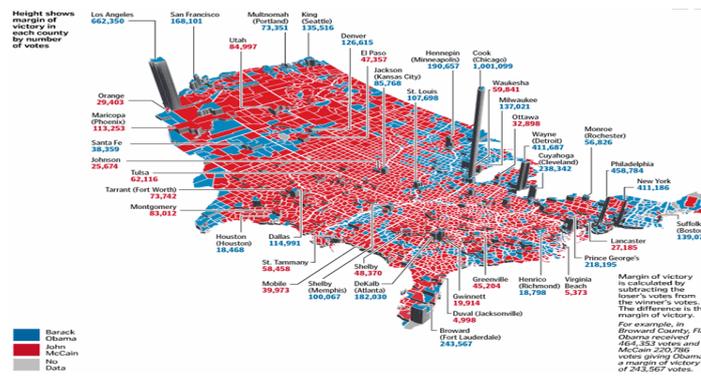


Figure 4.8: Washington Post and County level data

In the New York Times map (Figure 4.7) circle size was proportional to the amount each county's the leading candidate was ahead of his rival in the presidential race.

One of the more interesting versions of the proportional symbol maps was the Washington Post's map (Figure 4.8) broken down by county. The map used Flash software to give a full 360-degree view of election trends. The results of an entire campaign were summed up effectively in a single graphic. Height showed margin of victory in each county by number of votes. This is where maps tend to overlap with (or even become) infographics. The difference is in presentation and design. In the Washington Post's map, the information was directly tied to a particular area. The data, then, would be relevant only to the context in which it was collected.

On traditional print maps, the legend serves as a translator for the symbols used. Contextual windows eliminate the need for legends in many online maps. Instead of having to refer to an explanation of the symbol in the margins, the user simply clicks on a point to find out more about it.

Legends are still needed in certain cases. Heat maps, for example, display intensity by shade of colour, and users usually require a reference bar to make sense of the information (Figure 4.9)¹⁹.

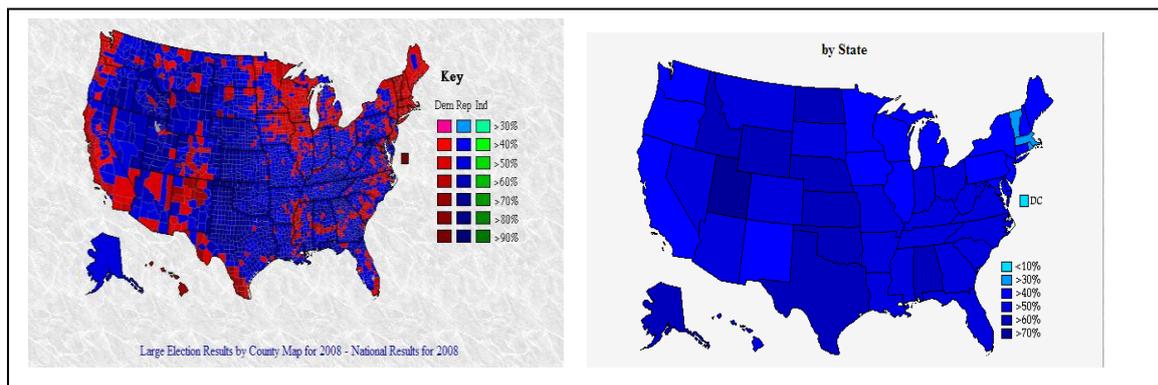


Figure 4.9: Heat maps of US elections 2008

¹⁹ Dave Leip's Atlas of U.S. Presidential Elections (<http://uselectionatlas.org/RESULTS/countymap>)

Interactive visualisation, like interactive graphics, allows the viewers to choose the direction and position from which they are viewing, what they are seeing and how it is depicted, lit, animated and so on. Whether a particular visualisation is good depends on how well it represents the visualized data, how effectively it employs human perception, and whether it makes the most crucial aspects of the data the most accessible.

Which is the most suitable thematic mapping technique will depend on the type of data being described and the specific purpose of the map. A number of factors must be taken into account if the data is to be represented effectively: apart from the accuracy of spatial and statistical data, map makers must also consider the aesthetic qualities of the map, the format in which it is to be presented and the intended audience.

4.3.4 Election Dashboards

A recent trend in visualising election data is the use of ‘dashboards’ (Few 2006) which have parallels with visual displays of data that would be found inside motor vehicles, aeroplanes etc. Data visualisation elements such as charts and graphs, gauges and maps are integrated in the summary view and the multi-dimensional view of a dashboard application not only helps in optimising the use of screen real estate, it also aids in quick understanding of data.

Dashboard users are easily able to identify the high and low values in the data and they also obtain an idea of the trend with which the data is changing.

An important feature of dashboards is that they are multi-layered applications, with the top layer summary view presenting a comprehensive summary of data. If the user needs additional information she/he could navigate to the next level to obtain additional data, and at the same time could manipulate the data for the purpose of analysis. The bottom layer is the

detailed view layer, where the user can see details at the smallest level of aggregation that applies to any given application.

Given the purpose that dashboards serve, according to Few (2006) they must be designed to support the following process of visual monitoring:

1. See the whole picture.
2. Focus in on the specific parts of information that need attention.
3. Rapidly explore additional information that is needed to take action.

Step three can be achieved through suitable links to additional information, however steps one and two require visual design that allows viewers to scan a vast amount of information quickly to get an overview, easily recognise items of interest, and then obtain enough information about those items to assess the potential need for a response.

In election dashboards used to visualise US elections there is normally a detailed breakdown of candidate support by race, age, state, gender, marital status and educational background.

An example shown in iDashboard²⁰ (Figure 4.10), also provided additional information which includes a map of state electoral votes and financial details about the political parties. It used colour-coded graphs to allow voters to identify and track candidate support. Users could apply the hover-and-click capability to obtain finer details, examining specific elements; in this case the candidates' support groups.

²⁰ <http://live.idashboards.com/president2008/?guestuser=pres2008>



Figure 4.10: iDashboards of the 2008 US Presidential election

Colour can be used in powerful ways to emphasize and encode data, or create a relationship between individual items on a dashboard, but it is sometimes over-used and misused. Some colours are hot and demand our attention while others are cooler and less visible. When colours in two different displays are the same users are tempted to relate them to one another. It was often assumed that the colours like red, yellow and green can be used to assign important meanings to data, but in doing so, the 10% of males and 1% of females who are colour blind will be excluded (Few 2006 p.75).

In the iDashboards above, the bar graph misuses colour in several ways but there is one problem that stands out most. What is the meaning of the separate colour for each bar (Figure 4.10, right hand side of screen grab)? The correct answer is that the colours mean nothing. Nevertheless, the users—whether consciously or unconsciously—searched for the nonexistent meaning of these differences. As a result the iDashboard suffers from useless decoration; visual flourishes that serve merely to distract rather than inform.

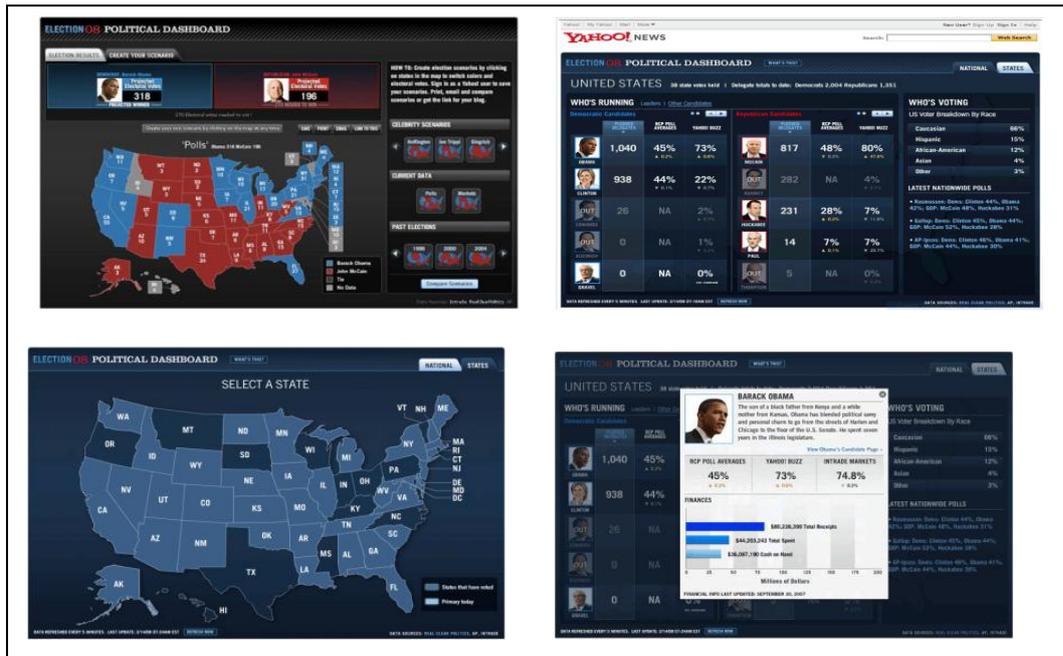


Figure 4.11: Yahoo US Election 2008 Dashboard

The Yahoo’s political dashboard²¹ (Figure 4.11) incorporating tabs, panes and pop-ups made this design intuitive to use. It provided a minute-by-minute account of the polls and alternative indices tracking the election. A comparison of electoral votes from the elections past to the 2008 president election helped to spot changes in the country. It was clear to see where the candidates were focusing their energy and how those choices could influence the turnout for them. In addition to showing the winners and losers by the usual metrics, there was the use of some non-traditional data sources. Next to each candidate’s picture were four basic statistics: recent polling results; ‘buzz’ which showed the relative popularity based on Yahoo search queries; prediction market, which showed the likelihood of a candidate to win based on real world and the total money rose for each candidate. Users could also click and see the data on a state-by-state basis, and see other demographic data.

²¹ //news.yahoo.com/election/2008/dashboard

Dashboards are appealing because they are presenting an extensive numbers of different values in a single combined view with the high-level summaries of the data. It quickly tells the user what is happening, but not why it is happening, which may or may not serve as the starting point for further investigation. Dashboards provide intuitive indicators, such as gauges and stoplights that are instantly understandable and it is easy to see why they have become more popular with each election.

Dashboards offer a solution to information overload, but only when they are appropriately designed. To achieve their potential and provide their purpose, dashboards must display a vast amount of information in a small space in a manner that communicates clearly and immediately. This requires design that takes into consideration and uses the power of visual perception and the human brain to sense and process several bits of information simultaneously. This can only be achieved when the visual design of dashboards is informed by an understanding of visual perception and human cognition—what works, what does not, and why. The fundamental challenge of dashboard design is to display all the required information on a single screen clearly and without distraction, and in a manner that can be quickly examined and understood.

4.3.5 Cartograms of Election Results

Cartograms are a well-known technique for showing geography-related statistical information, such as demographic and epidemiological data (Slingsby, Dykes et al. 2010; Sun and Li 2010). The basic idea is to alter a physical map by resizing its regions according to a statistical parameter in a way that users be able to understand quickly the displayed data and relate it to the original image. For example, traditional election maps misrepresent political support because the area of a geographic unit displayed is not equivalent to the number of

votes cast there. This flaw can be addressed through the creation of cartograms, maps in which geographic space is scaled by a value other than area.

Cartograms date back to at least the middle 18th century and have been used to portray a wide variety of data. One of the earliest examples of cartogram use by the American press appeared in the Washington Post in 1929, when state areas were scaled by population and by payment of federal taxes (Figure 4.12) to demonstrate the unfairness of each state having equal voting strength on tariff measures²².

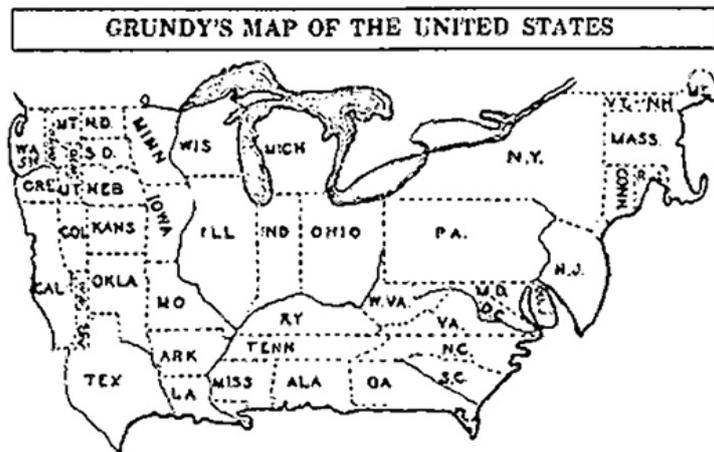


Figure 4.12: Grundy's Map of the United States

Monmonier (1977), introduced the idea that map could be distorted to give emphasis to a particular part of the picture (see also Kadman and Shlomi 1978; Lichtner 1983). Hägerstrand (1957) based his map distortion on a predicted distribution and distance decay of migrants in Sweden, a method that was later adopted by others (Sarkar and Brown 1994),

²² <http://indiemaps.com/blog/2008/12/early-cartograms/>



Figure 4.13: Dorling's centroids

Dorling, (1991) developed a novel approach using only centroids - a two dimensional, non-overlapping circle which are sized proportionally according to some variable such as population²³ (Figure 4.13). These bubbles are then allowed to expand, or contract, to attain the appropriate areal extent.

In what became a significant development in cartography Gastner and Newman (Gastner and Newman 2004) used gas diffusion laws to create a method of generating contiguous cartograms (Figure 4.14) which are easy to produce and give easily understood results even for complex mapping exercises. This has proved to be a widely used technique for cartographic mapping.

²³ Dorling Cartogram representations of the Percentage of the Male Population of Working Age in 1891 (ncgia.ucsb.edu)

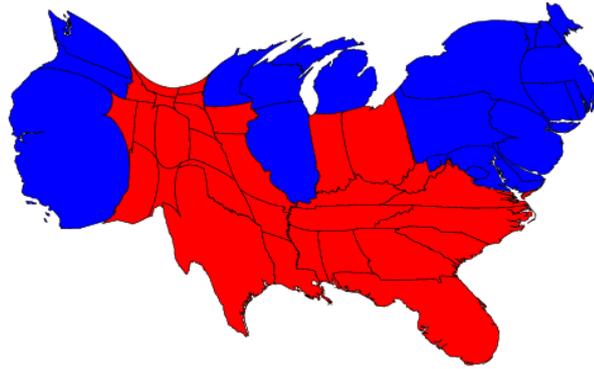


Figure 4.14: Gastner & Newman map of US Presidential Election

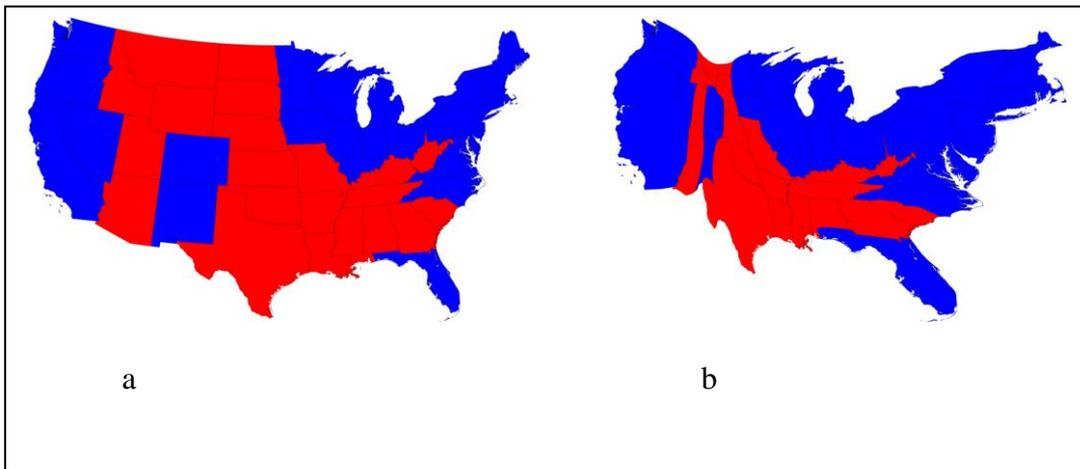


Figure 4.15: Comparison of two images of US presidential election

Election results are typically represented by colouring each political unit according to the candidate that won the popular vote (Figure 4.15, a). This approach is problematic because the visual weight of each political unit is based upon the area of the state and not the population of the state – compare, for example, Montana with New York. As a result, maps showing US presidential elections often are dominated by one colour, as one party tends to carry large but sparsely-populated areas while the other party tends to carry small but densely-populated urban areas. The overall visual weight of each party is thus not proportional to the actual election results.

The left map (Figure 4.15, a) gives the impression that the Republicans won the election, since there is rather more red on the map than there is blue. In fact, however, the reverse is true – the Democrats won by a substantial margin. In the map on the right (Figure 4.15, b), the size of each state is altered based on its population. The less populous western states are reduced in size, while the more densely populated states are increased. Though the shapes of some states are warped significantly, most are still recognizable, and the map gives a more accurate picture of the election results.

For a cartogram to be effective, users must be able to easily understand the displayed data and relate it to the original map. Recognition depends on preserving basic properties, such as shape, orientation, and contiguity. This is not always easy to achieve and it has been shown that the cartogram problem is unsolvable in the general case (Keim, North et al. 2002).

Although the cartogram is better than the choropleth approach, there is always a trade-off it appears regarding shape preservation, topology preservation, and visual equalization (Roth, Woodruff et al. 2010). Many cartogram algorithms often allow massive shape distortion, although many include limitations designed to minimise local and global shape error (Keim, Panse et al. 2002; Keim, Panse et al. 2005).

Topology preservation refers to the perpetuation of areal unit adjacencies from the original geography to the cartogram and may be calculated as the percentage of shared nodes on the original geography that are carried through into the transformed cartogram (Keim and Herrmann 1998). Visual or density equalization refers to the adjustment of the basemap according to a relevant variable (e.g., population in the context of mapping social data) to make enumeration units that are more thematically important more noticeable (Tobler 2004).

4.4 UK General Election 2010: Different views

No single party won an overall majority in 2010, for the first time in a UK general election since February 1974. The Conservatives won the most seats, 306, gaining 96 compared with notional 2005 general election results on the new constituency boundaries. Labour lost 90 seats, leaving them with 258, while the Liberal Democrats were down five on 57.

The 2010 UK election saw the internet playing a growing role according to Echo Research²⁴, which found that almost half of the population went online for information about the parties and candidates. The Internet mass media gave users an opportunity to find extensive information about general election results and the users used the maps to discover and measure political statistics. Here, we identify and briefly describe the main strengths and weaknesses of some web applications that were used to track the 2010 general election- The Guardian, Daily/Sunday Telegraph, The Times, Sky News, Yahoo and the BBC. These sites all designed around a map of the UK drew on different data sources, and used different visual techniques to display their information.

4.4.1 The Guardian

The Guardian newspaper has taken a particular interest in data visualisation, even to the point of inviting readers to undertake their own ‘mash-ups’ of data that can be downloaded from the paper’s website. The Guardian’s election map²⁵ (Figure 4.16) used simple blocks to construct a cartographic picture of Britain, although an alternative zoomable, geographical view was also available from the menu, as was an animated bar-chart view. Also available were pages describing each parliamentary constituency containing a brief profile, historic

²⁴ <http://www.echoresearch.com>

²⁵ <http://www.guardian.co.uk/politics/interactive/2010/apr/05/general-election-map-swingometer>

results and candidate information were accessible. The map included links to candidate websites and Twitter feeds. The three-way swingometer gave an opportunity for users to see how different configurations of party vote shares could affect the outcome of the election.

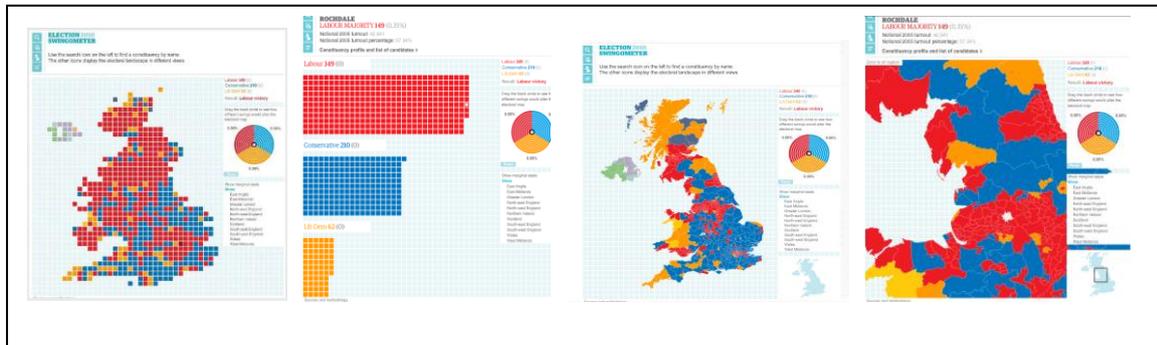


Figure 4.16: The Guardian 2010 election map

To the right of the map was a pie chart of concentric circles, showing the relationships between the three main parties. The centre of the circle could be grabbed on a mouse-click and pulled in, out and around each of the three coloured segments. Moving the circle towards or against each party would be instantly reflected on the map itself, whether in cartogram, geographical or bar-chart mode.

4.4.2 The Telegraph

The Telegraph used honeycomb map²⁶ – a variation on the cartogram (Figure 4.17) with each cell representing a constituency, the name of which would appear using mouseover action. Each hexagon drilled down to information on candidates, health, crime and education statistics for each constituency. In the main map view, there were links to the latest polling data which were displayed on the map, along with details of all seats that would change hands.

²⁶ <http://www.telegraph.co.uk/news/politics/2432632/UK-General-Election-2010-political-map.html>

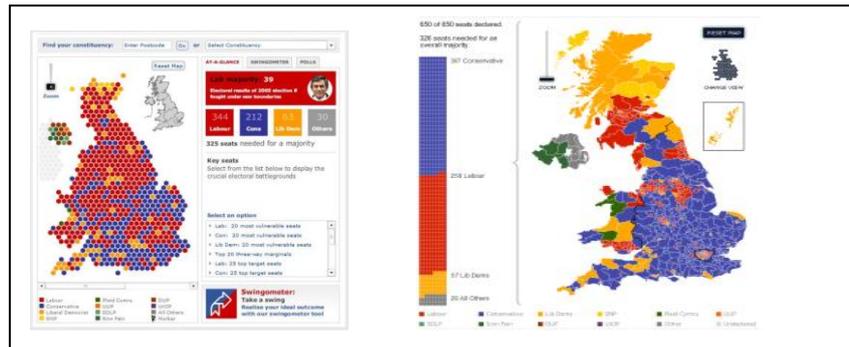


Figure 4.17 The Telegraph 2010 election maps

There was also a simple swingometer tool which let users see the impact on seat distributions of voting swings between the parties on the map. When a constituency was selected an arrow appeared over the cell and a pop-up window came into view with the results of the 2005 estimated election. Within the window there was also a pane dedicated to related articles and links to the current MP's voting record in the House of Commons, via the website 'theyworkforyou.com'.

4.4.3 The Times

At The Times (Figure 4.18), the approach was to use simple tools to make sense of the mass of election data but as can be seen the mapping employed used standard rather than cartographic designs. The Times Election Map²⁷ let users zoom in by region and by individual constituency, rapidly presenting a graphical and percentage breakdown of the selected seat as well as brief details of the area. Searching also could be done by a postcode. The Times presentation gave an opportunity for users to use the latest betting odds information to estimate probabilities of each party's chance of winning a given constituency.

²⁷ <http://generalelection2010.timesonline.co.uk>

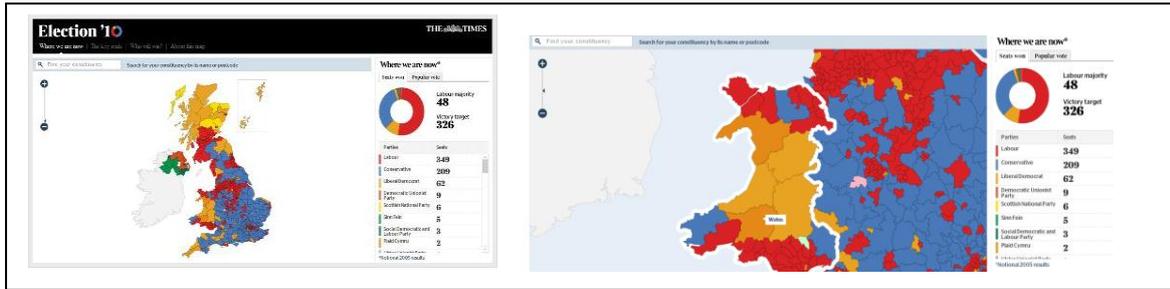


Figure 4.18: The Times 2010 election map

4.4.4 Sky News

The Sky News visualisation²⁸ (Figure 4.19), another that used standard mapping procedures, had a concise constituency and candidate information, with emphasis on expenses. Also it contained intensive use of social media to populate constituency news pages.

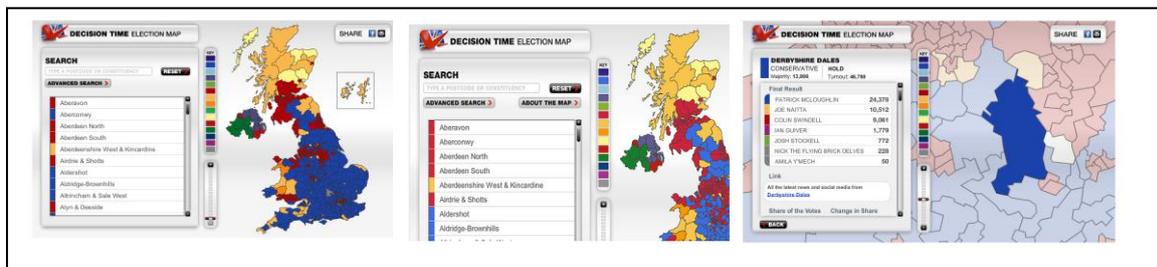


Figure 4.19: Sky News 2010 constituency battleground

Each of the individual constituency pages contained short profiles on the candidates; the political history of the area; hyperlinks to content from national and local news sources; the latest pictures, videos and tweets from candidates; links to social media sites including Flickr and Twitter reports, images, and video uploaded by Sky News' reporters in real time. Sky also had a separate set of iPad media, some of which overlapped with the website but others that were specific to the iPad application itself.

²⁸ <http://news.sky.com/skynews/Election/Map>

4.4.5 Yahoo!

Yahoo's interactive election map²⁹ (Figure 4.20) was lacking in colour in comparison with the traditional red, yellow and blue maps on other websites. However, it performed well with a considerable amount of functionality.

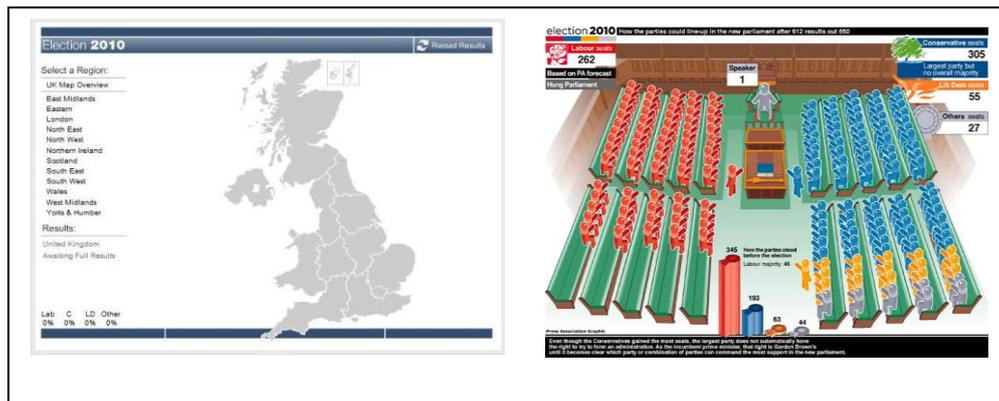


Figure 4.20: Yahoo 2010 election map and Commons projection

Users could select a region and an animated zoom was immediately offered a close up map of the region, over which they could hover to choose a specific constituency or select from the alphabetical list which appears alongside. On selecting a constituency, an information panel pops up with brief details around the status of the seat and the candidates, although no forecasts regarding potential results were available.

4.4.6 BBC

The BBC's Election seat calculator employed several tools used by the other applications reviewed here, and made them available in a single view³⁰ (Figure 4.21). The distribution of vote shares between the main parties could be adjusted by moving handles on the colour wheel and changing the size of each segment to see the impact on a honeycomb map, similar

²⁹ <http://uk.news.yahoo.com/elections/>

³⁰ http://news.bbc.co.uk/1/hi/uk_politics/election_2010/

to that used by Sky News and The Telegraph, and a bar chart much like that of The Guardian’s application.

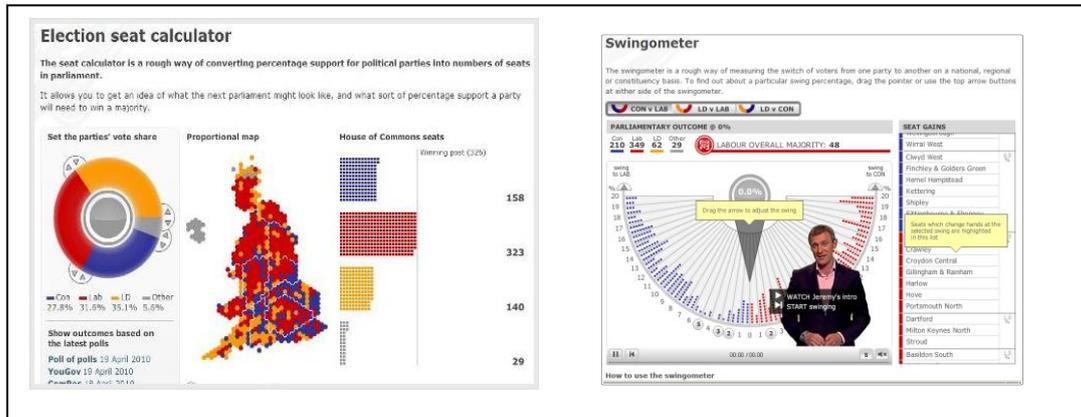


Figure 4.21: BBC 2010 election visualisation

The results of elections since 1974 and their mapping appearances were also available – note that Sky News prepared a set of maps that began with the 1832 general election and permitted users to compare two election outcomes alongside one another (see Chapter 9). The BBC site also contained a ‘Where they Stand’ feature which allowed manifestos to be compared by party and policy, whilst the ‘Seat Calculator’ showed how different national vote percentages would affect the number of seats won by each party and whether or not a single party would win an overall majority. The BBC visualisation had more extensive constituency descriptions than some of the others listed above, but in other respects these pages are surprisingly information-light by comparison to other sites with only limited electoral detail and candidate information available.

From the viewpoint of mapping UK electoral data, therefore, it is quite clear that different methods are being employed. The following three maps (Figure 4.22) summarise the different ways of mapping the British election results in comparison.

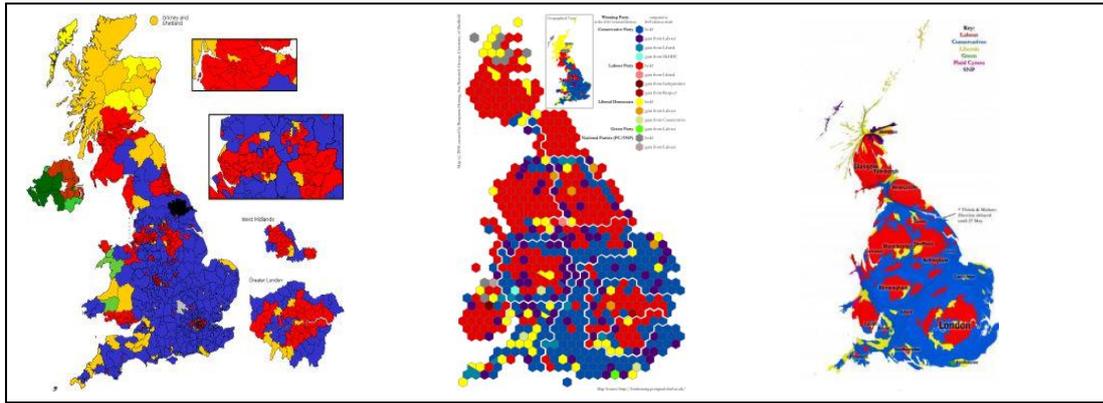


Figure 4.22: Comparisons of mapping UK elections

The left map uses a traditional projection (using the British National Grid), which corresponds to the geographical area and thus overstates the vote of rural areas (making blue much more dominant than the real results are).

The map in the middle uses hexagons to signify the constituencies, so that this map distorts land area in favour of a representation of seats in the British Parliament. The right map shows the election results on the gridded population cartogram, which resizes the results related to the population distribution. It shows, how many people are represented by the winning party, and each grid cell refers to the same geographical extent. The smaller a grid cell, the fewer people are living there, so that rural areas are those where many lines are close together, whereas conurbations literally bulge out of the map.

It has been established that the strength of cartograms is their ability to show geographically-related statistical information by depicting the chosen attribute of a geographic object as that object's area. Cartograms (maps) are not ideal for showing precise values but they are a good way of combining spatial and numerical meaning and presenting it visually. This being the case, precision is sometimes subordinated to the presentation of spatial features.

Construction methods for cartograms must be able to preserve the cues essential for recognition of region shape and automatically achieve the desired region area while maintaining correct map topology. A cartogram constructed using one mathematical algorithm will look considerably different from those generated using other algorithms. An algorithm which is perfect for creating a cartogram of the UK may not be appropriate for US. Even within one cartogram, use of a particular algorithm may produce a result that is good in some areas and unrecognizable in others. In other words, the choice of algorithm, the attribute being mapped, and the geography are the main factors that will determine the final effect.

Squares, circles, triangles and hexagons are commonly used to represent objects on cartograms because of their ability to express statistical information precisely. However, the fact that shapes are identical and simplified can make it difficult to identify the locations of objects. In an extremely simplified cartogram such as that shown in Figure 4.22 in the middle, it is important to add references such as labels or a normal map as a basic guide for locating objects. Where a large number of objects are being mapped, some areas may be nearly invisible or too similar to be distinguishable enough for coding or labelling. An effective way of adding further information is to add mouse-over text. By zooming in or zooming out, every object on the map can always be found quickly and accurately.

Given the difficulties of balancing distortion and accuracy, it is perhaps not surprising that there is no specific standard for evaluating cartograms. While innovations such as the diffusion cartogram (Figure 4.22 right) are welcomed for their fresh approach and visual impact, established forms such as those based on constraint-based algorithms are still in widespread use.

Comparisons between choropleth maps and cartograms have revealed that map patterns are not always communicated clearly to readers. Visual deceptions occur commonly in choropleth map reading. Meanwhile, the distortion of geographical information on cartograms usually leads to misunderstanding of adjacency relationships between areas. It is not possible to achieve a completely correct map, only a suitable map. Thus, it is suggested to present cartograms and conventional maps together in order to avoid misleading the reader and to enable them to construct their own hypotheses.

Each of the maps above is useful for itself depending on what users may want to know about the election outcome, as all three are telling a very different story of it.

What was surprising is how similar the maps produced by different media organizations were to one another. Both the BBC's and the Telegraph's election results maps change between geographical maps of Great Britain and Northern Ireland and hexagon-based cartograms that give each constituency the same area. Others take one side of the geographic/cartogram divide: the Times' map is resolutely geographical; the Guardian's rectangular-based cartogram, which though unlovely is at least different from the rest. The exception was Sky News where the main news website contained standard map projections but a separate application created a Flash-driven cartographic animation of every general election since 1832.

4.5 Themes within the visualisation of Election Results

Visualisation of election results is very challenging because of the complexity of the data. The data types need to be visualized include: candidate (name, party affiliation, and state or

constituency), electorate (party affiliation and vote), geographical distribution, balance of power, change in balance of power, margin of victory.

Maps are useful tools for making sense of compound data, though the choices of metrics, colour schemes, analytical methods, scale, and symbols are critical in presenting results that can be understood and evaluated. There are multiple ways of mapping the information in the election results to visual properties. A good mapping is able to convey facts to the user.

However, an inadequately designed visualisation may bewilder the user, distort the data, or even cause the user to draw false conclusions about the data. As Van Wijk (2005) points out, visualisation is subjective and it can lead some users to the wrong conclusions about the data being presented to them. Although most people are able to use the web to absorb information that process is somewhat passive – the user navigates through a pre-set options rather than being able to interrogate information independently. Lying behind such websites may be complex data sets but their potential is not realized and users may become frustrated that the questions that they want to ask, the relationships they want to see, have not been anticipated by the web-page designers.

There are several problems in creating and interpreting election maps that should be considered. To begin with - popular vote data is necessarily aggregated at several levels, such as counties, region or countries/states, which are then coloured to show election results, and the problem arises when these spatial units differ in size (area) and significance (as in the United States where electoral college votes are determined by population size). These maps give extra visual weight to larger areal units, whether by county or state.

Another problem relates to data classification and the trade-off between few or many categories. Election maps often use a two or three main class colour scheme (red and blue, or

red, blue and yellow in case of the UK), which results in a map that is easy to read but is highly generalised. Some maps use more classes, such as shades of the main colours to indicate the degree of election victory. These maps provide a more detailed picture, but have various problems associated with classification of data. The designer must decide how many classes to use and how to break the data into those classes; software programmes will often do this automatically but the outcome is not necessarily aesthetically pleasant. The look of a map can vary significantly depending on the classification choices. The choices of colour and shading likewise affect the map's appearance. Moreover, colour underscores how we interpret the world and can shift depending on context. Various cultures see colour differently. An earlier study of Saito, in 1996, compared colour preferences in Japan, China, and Indonesia. Saito asked 490 subjects (175 Japanese, 158 Chinese, and 157 Indonesian) to choose the three colours from a chart that they liked most and a further three that they liked the least (Saito 1996). He found differences between respondents according to their country of origin. Although each country's respondents significantly differed from each other in the selection of colours of certain hues, a strong preference for white was ubiquitous. The interpretation of colour, of course, can become culturally embedded and those cultural associations may subsequently affect the use of these colours in a political setting. In the UK, for example, the colours blue and red are strongly associated with the two main parties, Conservative and Labour. By contrast, in the United States, the more right-wing Republican party is depicted by the colour red while the centre-left party, Democrats, are denoted by a blue colour. In other European countries black is synonymous with conservatism, red with socialism, while brown is still sometimes associated with the Nazis (Gage 1999).

Finally, there are problems associated with human perception. Large areas of colour appear more saturated than small areas of the same colour. A combination of incompatible colours

and shades can result in contrast misperceptions. For example, an area shaded light red surrounded by areas shaded dark red will appear even lighter. Differing shades of red and blue compound this problem of perception.

Many graphical symbols are transient and tied to a local culture or application. On the other hand, in some cases, arbitrary representations can be almost universal. We have many standardized visualization techniques that work well and attempts to change them would be imprudent. In many applications, good design is standardised design.

In respect of the design of animated maps, problems arise from the speed of change of individual frames of an animated map on-screen, which limited the time to examine fine details. In other words, there are obvious cognitive and perceptual limits that must be understood and used in Web design. Stepping outside of these limits is likely to leave the user frustrated or unsure of what they have seen. Basic map-reading tasks, such as comparing colours on a map with those on a legend, become significantly more difficult when the map is constantly changing and thus ‘the compression of time as well as space in dynamic cartography poses new problems requiring the recasting, if not rethinking, or the principles of map generalization’ (Monmonier 1996 p.96).

Viewers’ perception and cognition (discussed in Chapter 2) all need to be taken into consideration with regard to how to present, contextualise and prioritise information. Matters to be taken into account include data dimensions presented, spatial positioning and relationships, colour and tone, and any dynamic changes taking place in these elements. For the interface of human-computer interaction, the designer needs to analyse the tasks to be supported, and address comprehensibility, navigability and the effectiveness of any solution, including its usability (Macdonald 2003).

4.6 Conclusions

This chapter has focussed almost exclusively on maps as powerful mechanisms for data visualisation. In particular, we have shown how the use of maps and cartograms is widespread in presenting election results both in the mass media and increasingly on the Internet. Maps are intuitive and shape our basic understanding of the spatial distribution of social phenomena. Cartographers are tasked with the job of ensuring these maps show a complete and unbiased view (or as close as possible) of the represented social data. Visual equalization using the cartogram, rather than traditional methods of map projection, has proved to be an essential method for providing a more representative (in the sense of giving equal coverage to equivalent electoral units) view of aggregated social data.

We have outlined various methods used to present political/electoral information on the Internet; many people depend on the Internet for real –time updates of election results which means that web design is critical in permitting both quick and easy access and clear and simple explanations when people reach the site. Unlike most news stories, where readers expect to read textual descriptions, view videos and images, for elections, users like to look more at summaries of data in the form of tables, charts, maps and other visualisations. This probably explains the ubiquity of dashboard designs which by using a configuration of separate elements permit users to become familiar with the data in a very short space of time.

The clear trend in mapping data is away from standardised presentation of data in a ‘take it or leave it’ style and towards a richer data experience (different levels and types of data lying behind map elements) and eventually some kind of user-interactivity. The problem with viewing the end-products as we have done in this chapter is that we are only partly aware of the negotiation between data being totally accurate and complete and being usable. It is

possible to populate general election applications, for example, with a huge amount of data but if the consequence is information overload then implementing the system becomes pointless – a confused and confusing website is one where only the most interested user will stay for any length of time. In later chapters, therefore, we examine in closer detail some of the mechanics in constructing these websites in order to understand more about the interaction between data providers, web designers and web users.

5 Data visualisation within the research process

5.1 Introduction

An important feature of data visualisation is to act as an aid to improve the research process. It has been shown earlier (Chapter 2) that visualization applications can enhance the capability to process and then communicate information that is either uni-dimensional or multi-dimensional and can provide superior representations to written text. These visual applications can provide insight into the similarities and differences within data utilised by researchers. A critical component of the research process involves developing, sharing, and comparing concepts and then applying those concepts to data to generate new knowledge. As research develops that those people collaborating with one another need to see the effects on data of what they are trying to do. This is very much like data-mining – progress is achieved through a series of ‘trial and error’ exercises. By detecting and displaying similarity and structure in the data, methods, perspectives, and analysis procedures used by researchers, we are able to produce visual depictions of the core concepts involved in a domain at several levels of abstraction. Because this activity takes place among research collaborators the results are not always intended for either other academics in the form of published papers or the wider public in terms of presenting data that goes into the public domain; data visualisation is simply assisting the identification of patterns in data and communicating those

patterns within the research team. This is an aspect of data visualisation that is rarely reported upon within the academic literature but one which we believe to be of great importance.

Subsequent chapters will report on cases where visualisation is used to communicate with the broader academic and general public communities but it would be wrong to ignore the research process itself.

The point of this chapter is to show from some actual examples how this process works and the important role played in it by data visualisation methods. These methods reflect the differing interests and approaches of two different groups: On the statistical side, researchers are interested in finding effective and precise ways of representing data, whether raw data, statistics or model analyses. On the Infovis side, they are interested in grabbing the readers' attention and telling them a story (see Chapter 6). Both approaches have their value and it would probably be best if both could be combined.

The first section is a short description of the role played specifically by mapping in helping to visualize election data. This is then followed by two cases studies where the focus is more about the evolution of ideas and thinking and the ways in which data visualisation helped that. In the first case study we report on the use of triangles (described earlier in Chapter 3) when talking about three party competition. The second case study shows how we used simple line graphs to display the consequences of missing data imputation and applying different versions of moving averages for the vote share forecasting model. As data analysts, we see a large and continuing role for traditional display tools such as line plots for example, however there is a place for thinking seriously about using design to make the graphic more attractive and interesting. As we shall see, the very features that make effective information visualization can be disadvantageous to statistical presentation of data—and vice-versa. Thus, both researchers and designers can benefit from understanding each other's perspective, with the

aim not being a single display that makes everyone happy but a set of different data views that serve different purposes (Gelman and Unwin 2011).

This section is the longest in the chapter because it attempts to show how the research process is improved through trial and error approaches which is simple a way of visualizing a whole series of ‘what-if’ experiments. It culminated with the paper by Rallings, Thrasher, Borisyuk and Long that was published in *Electoral Studies* (Rallings, Thrasher et al. 2011 see Appendix A). My contribution towards that paper was to conduct the trial and error experiments and to provide graphics that showed the effects of controlling for missing data using different procedures.

5.2 Electoral research and mapping

Over the past two decades staff in the Centre have collaborated with human geographers and in the process have seen the value of ‘seeing’ aggregate data patterns revealed through mapping processes. The first example came with a paper published in *Political Geography* and co-authored with Daniel Dorling, then of Bristol University (Dorling, Rallings et al. 1998). In fact, the idea began following a map that was published by the London Borough of Kingston that showed the distribution of seats at local elections beginning in 1964. This appeared to show a pattern in the growth of the Liberals (later the Alliance and more recently the Liberal Democrats.). The pattern suggested that following an initial victory or two the party would then expand its control of seats in a particular way with new seats being won that touched the boundaries of wards that had already been won by the party. Rallings and Thrasher supplied the voting data but Dorling took the analysis a stage further by displaying national maps that sought to demonstrate that the spread of the Liberals was similar to that of a disease – spreading through contact and association (Figure 5.1 and Figure 5.2). This is why the title included the word ‘epidemiology’.

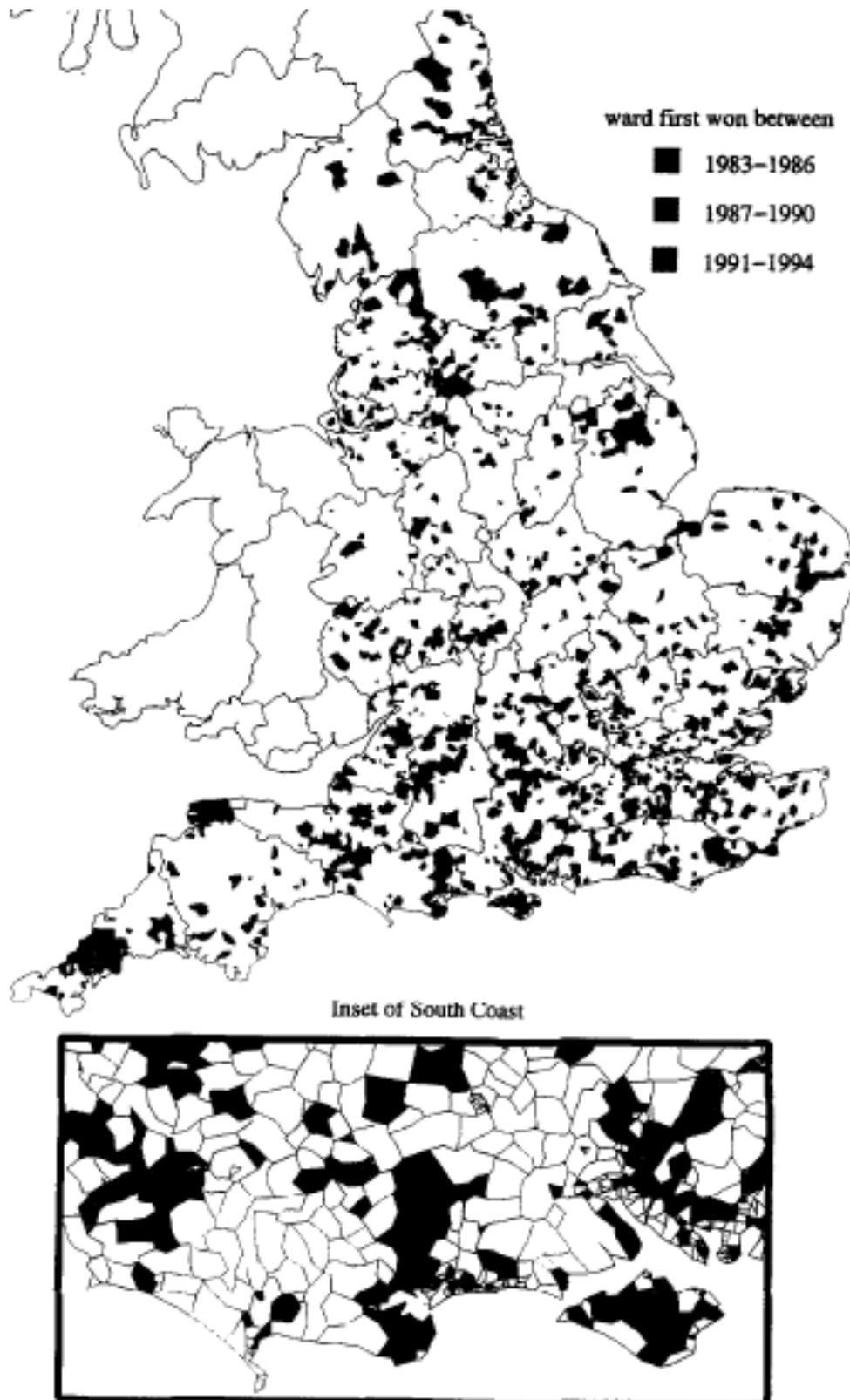


FIGURE 7. Liberal victories at local elections in England 1983-94 by ward.

Figure 5.1: Ward boundary maps and seats won by Liberal Democrats 1983-1994

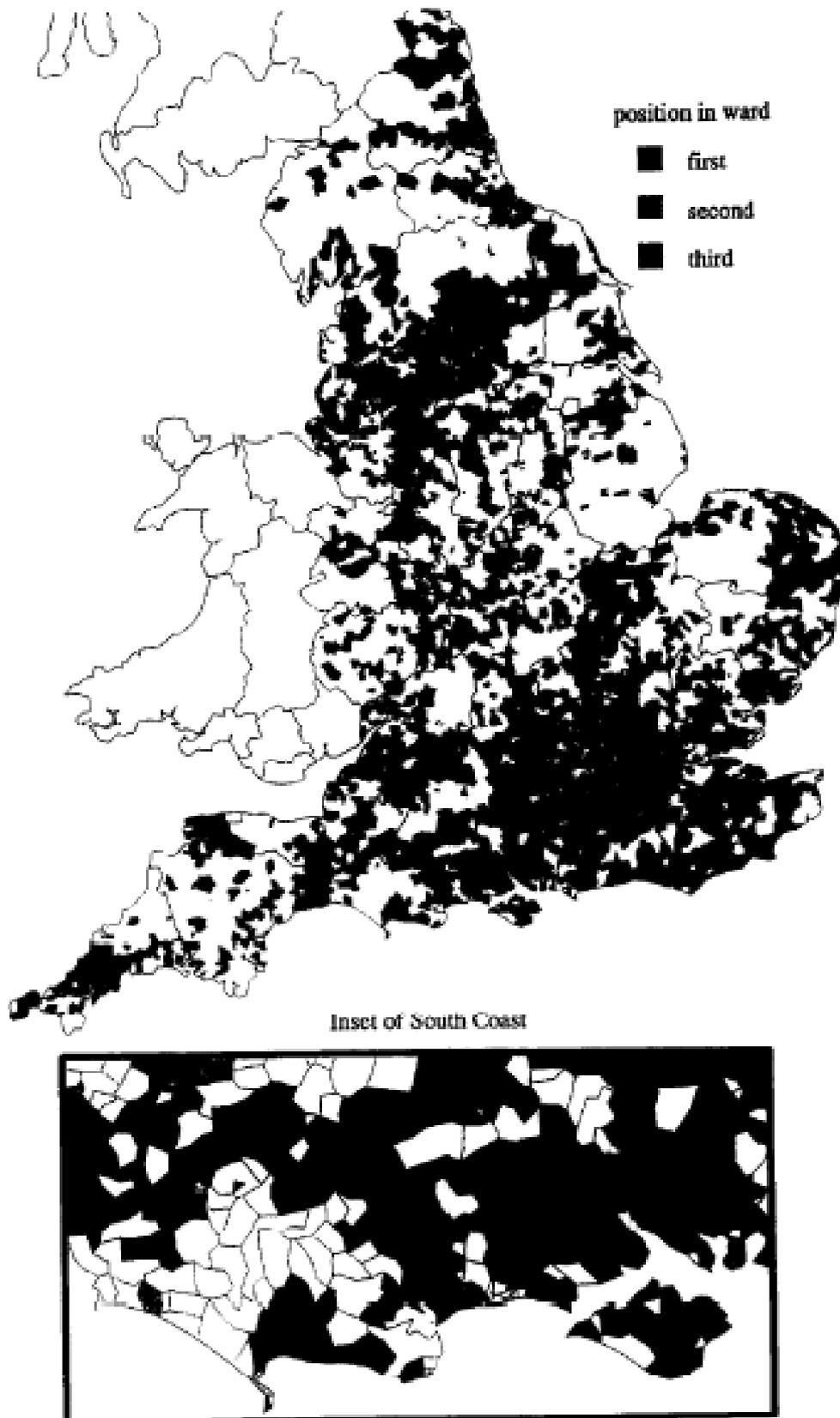


FIGURE 8. Liberal positions in local elections in England by 1994, by ward.

Figure 5.2: Ward boundary maps and seats won by Liberal Democrats by 1994

During this research Dorling would prepare maps for separate parts of the country and then share these with the Plymouth team. Being able to see the growth and pattern of the growth was vital to aid the understanding of the dynamics of Liberal Democrat success. Maps enabled the researchers to visualize the process more effectively than would have been possible using other types of methodology, for example that would include data in table formats.

Another Elections Centre collaboration, this time involving the geographer Scott Orford of Cardiff University again used mapping to display patterns in the data (Orford, Rallings et al. 2008; Orford, Rallings et al. 2009; Orford, Rallings et al. 2011). This research focused on election turnout but gathered at much lower levels of aggregation than normal. It is the case for turnout to general elections that turnout is only provided at the constituency level. For local elections the turnout is given at the ward level. But some local authority election officers retain details for polling stations. Rallings and Thrasher discovered that some election officers kept very good records, not only keeping turnout figures for different types of election but also having an archive over time. As with the previous collaboration with Dorling the research process began with Plymouth assembling the aggregate voting turnout data for each polling station area, including post code identifiers for each station. Orford then used GIS to map these stations (Figure 5.3) providing the researchers with the capacity to test different hypotheses about the relationships between geography (distance to polling station, nature of the terrain, proximity to road networks) and the level of turnout for three types of election (parliamentary, European and local elections) (Figure 5.4 and Figure 5.5). Once again, it is clear to see the advantages the maps bring to understanding where polling stations were located (often for administrative convenience) and where they might be better located if the aim was to improve turnout at low turnout elections for the European parliament and some local elections.

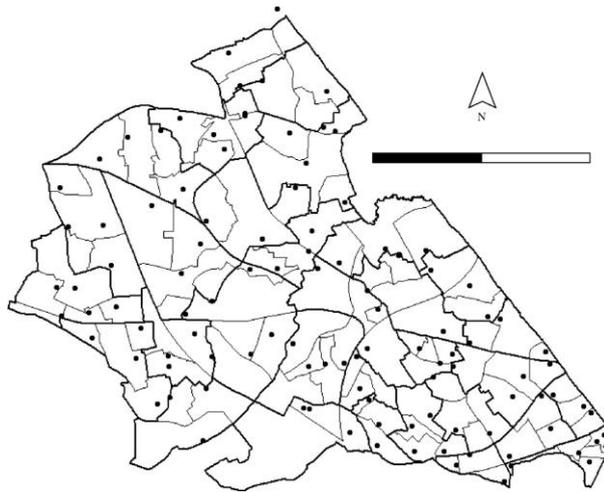


Figure 5.3: Wards, polling districts, and polling stations in the London Borough of Brent, 2001

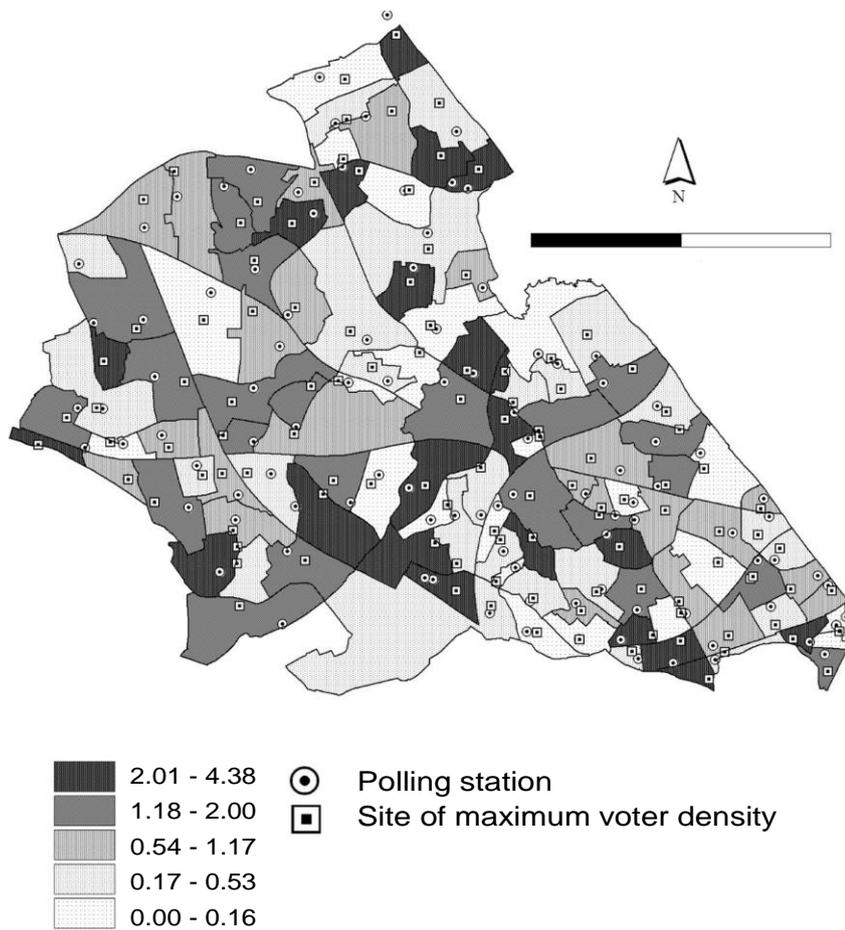


Figure 5.4: Percentage differences in predicted turnout when re-siting polling stations for European elections

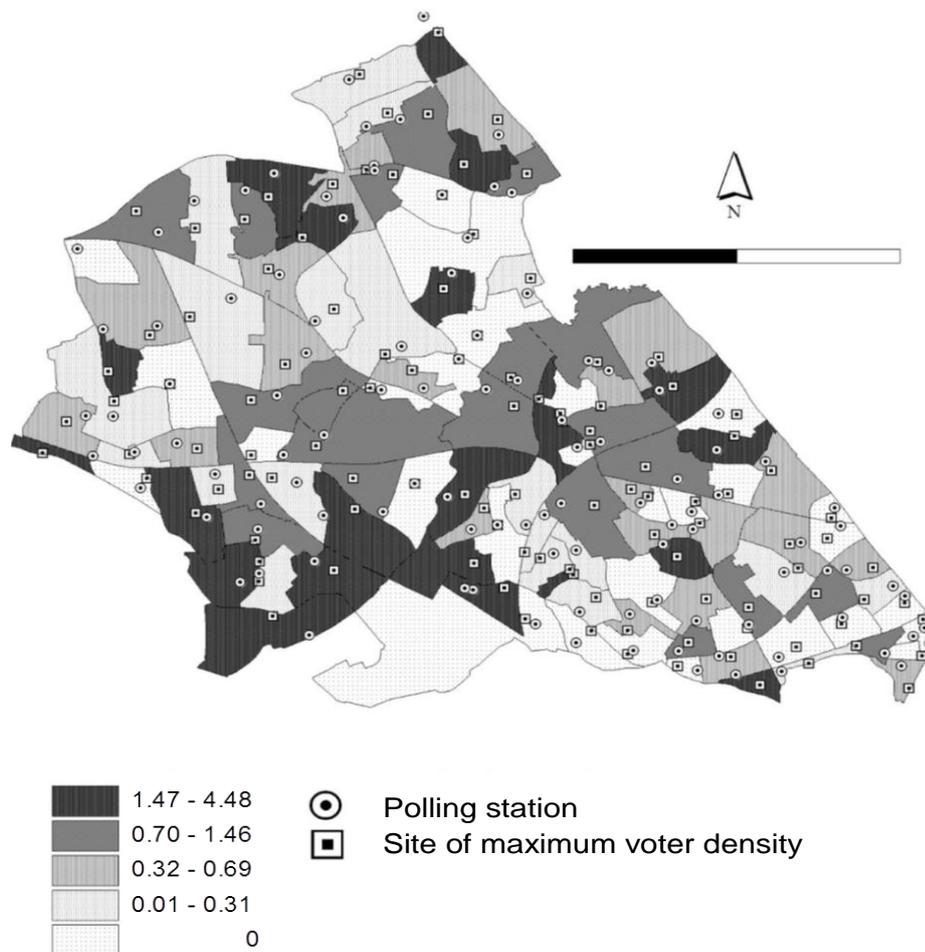


Figure 5.5: Percentage differences in turnout when re-siting polling stations for local elections

Following these collaboration the Elections Centre staff began to use GIS to examine some characteristics of patterns of voting at elections for the Greater London Authority. Here, the mapping process greatly assisted the visualisation of first, the geography of support for four small parties (Green, BNP, UKIP and Respect – see Figure 5.6 - Figure 5.9) (Borisjuk, Rallings et al. 2007) and second, whether small parties can use the knowledge gained from one type of electoral experience to help it win votes and seats at another type of election ((Thrasher, Borisjuk et al. 2012), Figure 5.11).

In the case of the first four maps relating to the 2004 elections the aim was to see whether there were any spatial patterns in the data. Although the elections were for the Greater London Authority the London Elects organisation that administered the elections using electronic counting provided information about party votes at the ward level. The Centre then used ward level GIS mapping coordinates to visualise the distribution of support for the Green party, Respect, BNP and UKIP. This had not been possible to do before since these are small parties and do not have candidates contesting local elections in every ward. But the Greater London Authority elections uses List voting and therefore everyone in London could now vote for these parties.

Mapping the data (Figure 5.6 - Figure 5.9) meant that the researchers could now see those areas that gave relatively strong support (dark shaded areas) to each one of these parties and where the support was weak (lighter shading). It was immediately clear that the anti-Iraq war Respect party which drew support from non-white Muslim populations, was strong in a group of wards slightly to the north east of central London while the anti-immigrant BNP (in wards dominated by white working class populations) were strongest in the wards that were immediately to the east of those wards.

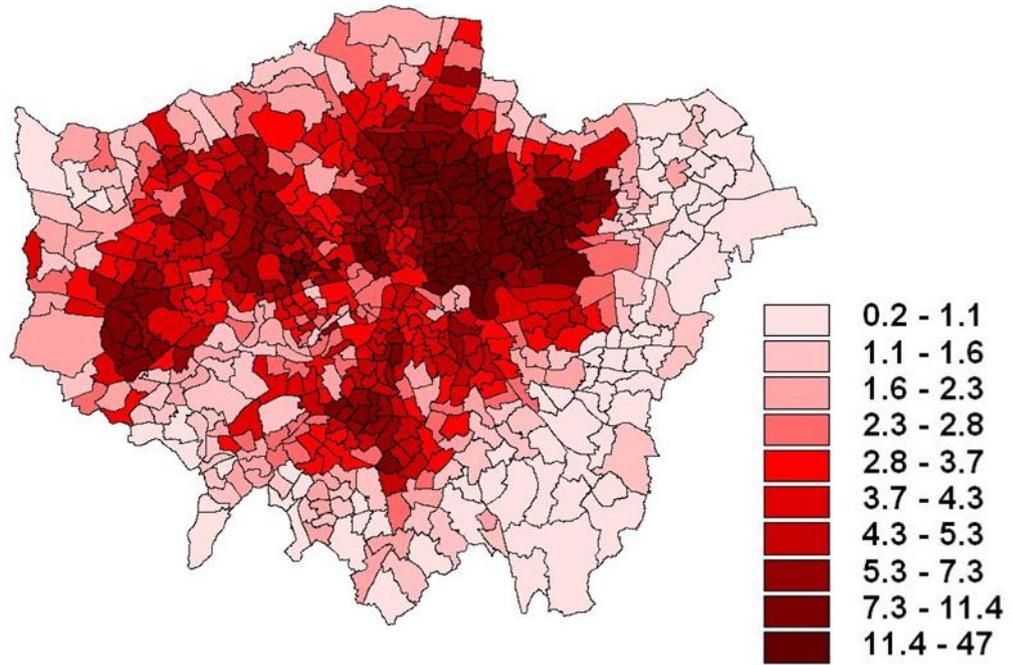


Figure 5.6: Distribution of percentage vote shares for Respect

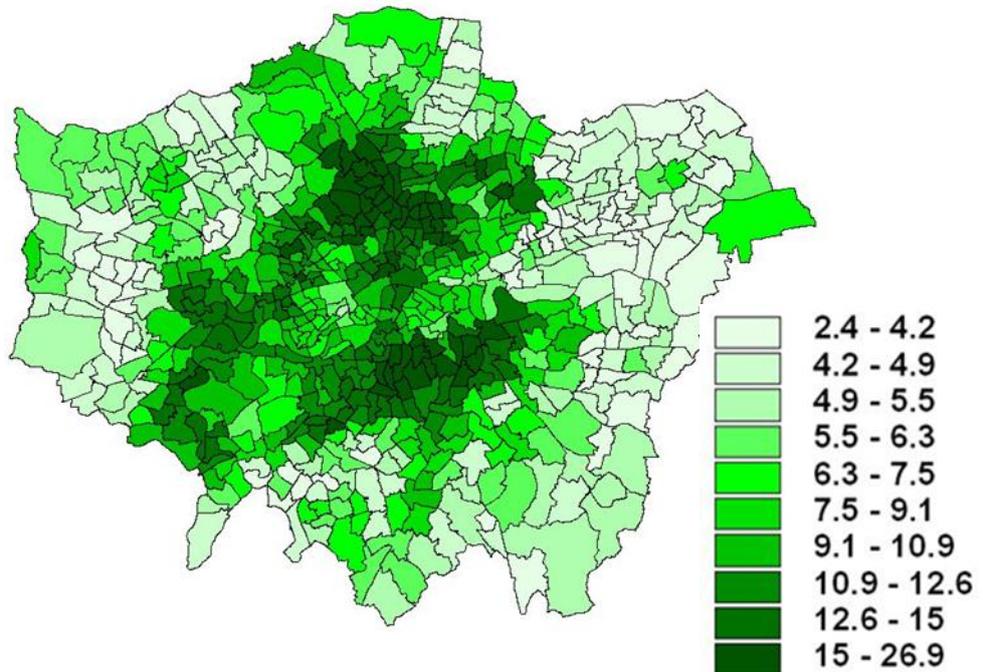


Figure 5.7: Distribution of percentage vote shares for Green

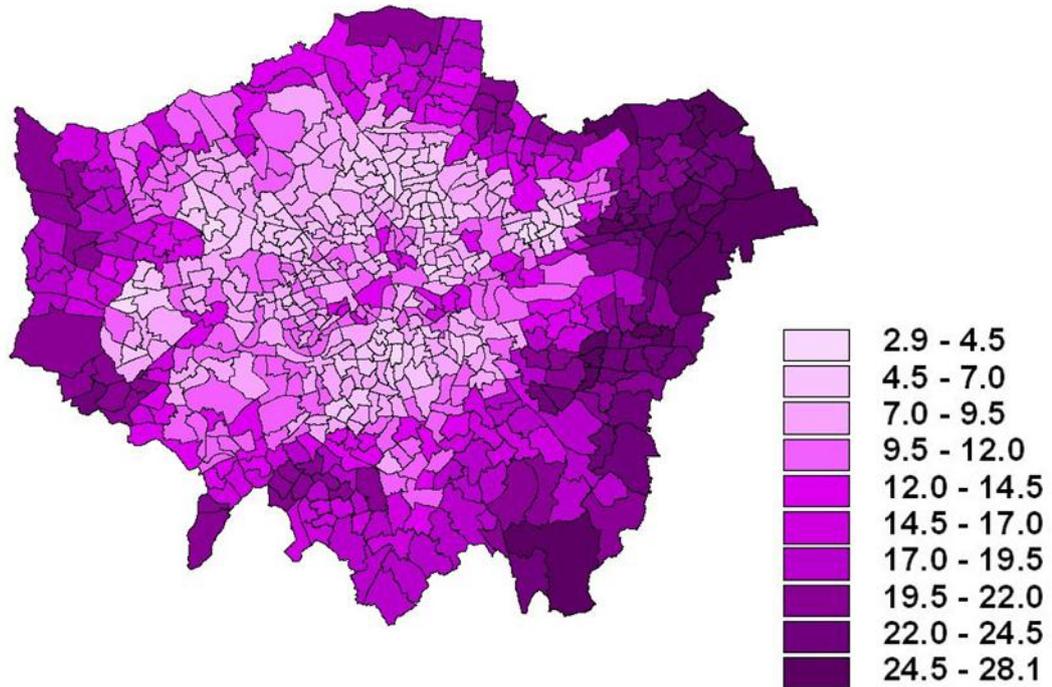


Figure 5.8: Distribution of percentage vote shares for UKIP

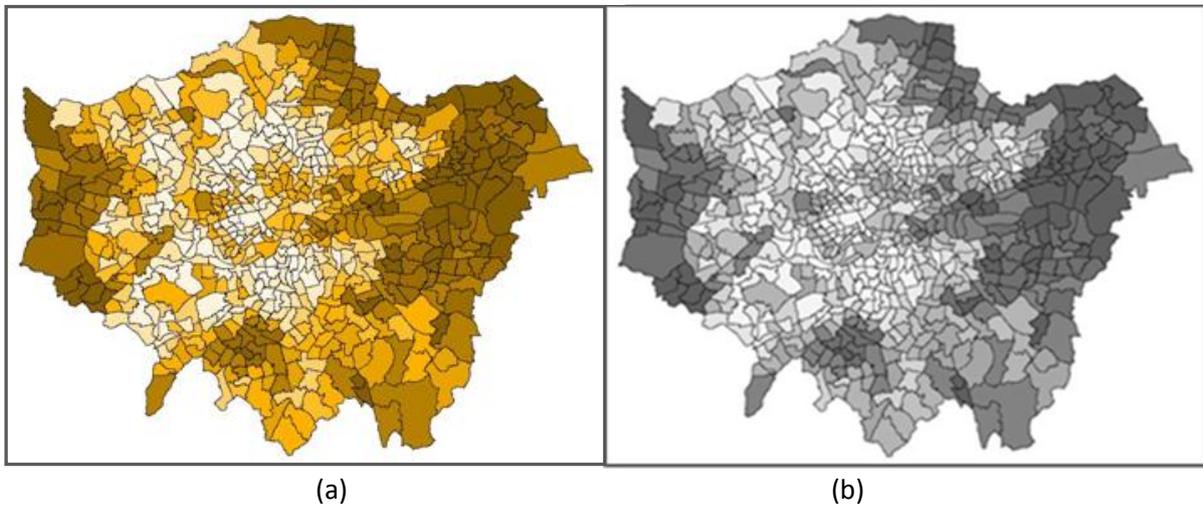


Figure 5.9: Distribution of percentage vote shares for BNP

Figure 5.9 is presented here in order to demonstrate how publishers, particularly the publishers of academic journals are lagging behind developments in computer software.

Figure 5.9a and Figure 5.9b are identical apart from the obvious fact that the first map has

picture and the other is presented in grey-scales. When the article was accepted for publication the publisher asked for maps to be re-submitted in the version on the right. The amount of information that is lost from this map by moving from a single colour to black and white is enormous. Some journals in political science are moving towards colour printing but there is still a great deal of limitation on the types of graphics that can be submitted which reduces the power of the authors to make their case.

Following the 2008 Greater London Authority elections the Centre conducted further research that took advantage of the fact the data for these elections was being provided at the ward level by London Elects. It was decided to undertake a comparative analysis of voter support for the Green party in local elections across the London boroughs and compare that with how the party did at the GLA elections. The research focused on whether or not the Green party could learn from GLA elections about the wards where it was attracting votes and then use that information to fight those same wards in local elections.

Because the paper was being sent to the same journal for consideration there was considerable discussion about the best method for overcoming the switch from colour to black and white. On this occasion the objective was not only to scale vote share for the Green party but also three configurations of whether or not a Green candidate had contested a ward at a previous and a subsequent election to the GLA election.

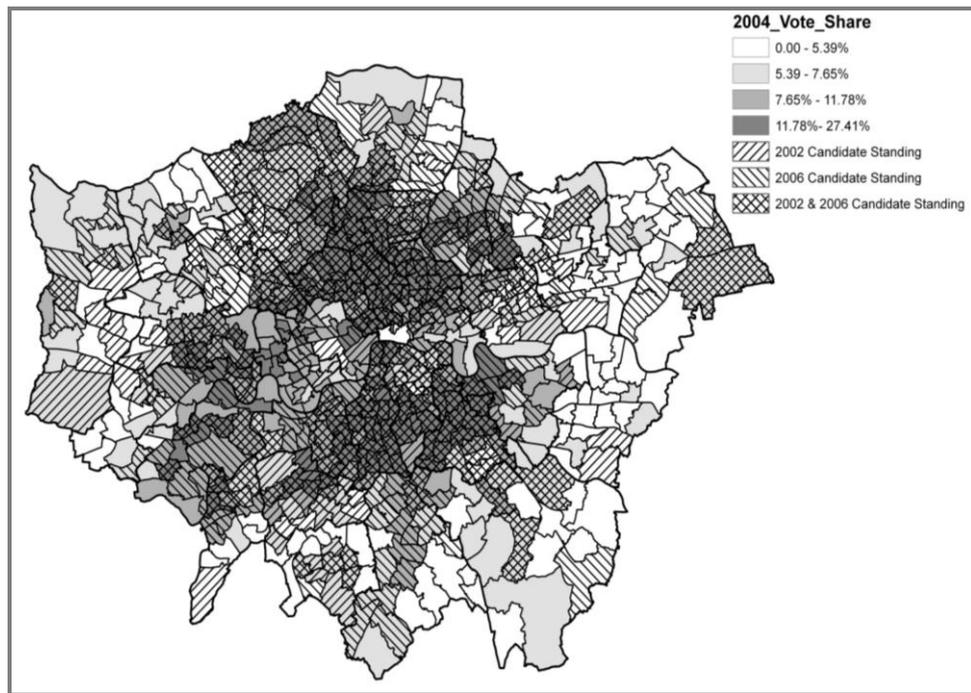


Figure 5.10 Pattern of Green Party competition and 2004 London Assembly list vote

Figure 5.10 and Figure 5.11, for example, use four shades for vote share categories which are then overlaid by cross-hatching that describes Green party competition for seats. The dark shading is noticeable in two parallel lines running west to east but the real purpose of the map and the paper is the pattern of party competition. Some of the detail is lost simply because the ward areas located in central London are smaller because of higher population densities. With hindsight it might have been better to have used cartograms rather than physical maps. This would have avoided the problem of seeing the detail in small areas but the compromise of shading and cross-hatching did result in better mapping than was used in the earlier article.

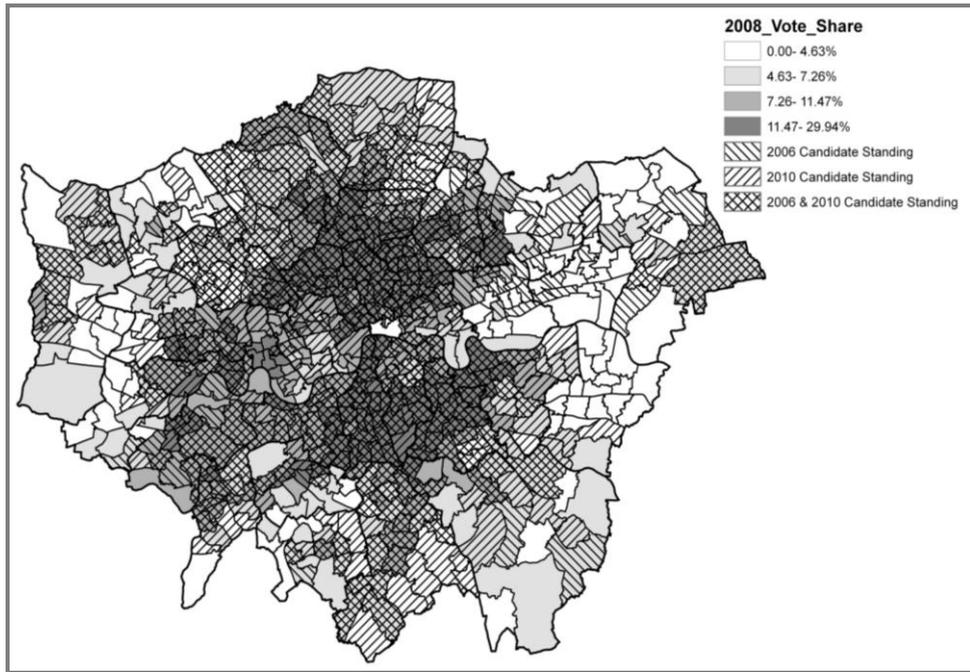


Figure 5.11 Pattern of Green Party competition and 2008 London Assembly list vote

5.3 Triangles and three-party competition

Although we strongly support maps when visualizing data there are still opportunities for using different approaches. This first case study examines the use of triangles when describing aspects of electoral and party competition. The approach adopts similar methods developed by Upton (Upton 1991) and by Dorling (Dorling, Pattie et al. 1993) that were discussed briefly in Chapter 3. The method that is particularly suitable for viewing the dynamics of three-party competition and was used in research that developed a new method for decomposing electoral bias (Borisjuk, Johnston et al. 2010).

Traditionally, a single straight line is used to describe some aspects of two-party competition whether it is vote share or the simple left-right continuum. In his *Economic Theory of Democracy* Anthony Downs (Downs 1957) used this method to show how two parties

competed for the centre ground where the maximum number of voters were likely to be located. There was also some description using the left-right dimension to show why third parties do not do well in first past the post voting. This presentation is not as useful when trying to display three-party competition, however. In Chapter 3 we showed how Upton (Upton 1976) began to use triangles to describe three parties and this method was adopted by others (Dorling, Pattie et al. 1993; Katz and King 1999). It was a method that the Elections Centre adopted as it began to try and develop a method for decomposing electoral bias for the three party case.

The method for decomposing electoral bias was formulated by Ralph Brookes in the late 1950s and then adopted for the UK by Ron Johnston and others (Johnston, Pattie et al. 2001). The problem was that Brookes' original formula was written for two parties and needed to be adapted to fit the three-party case. In recent times the UK has moved away from a two-party system where the combined vote and share of seats won by the Conservative and Labour parties was very large to more of a three-party system where the share of the national vote won by the third party, the Liberal Democrats since 1988, has been rising and also the party's share of seats.

In about 2007 Johnston began to collaborate with Plymouth University as they sought to develop a revised method. Our particular interest in this research project lay in the methods the team used to develop their ideas about the underlying nature of the voting data. Because of the three-party competition it was decided that triangular graphs would be the best way to show the distribution of the three-party vote.

If we imagine a hypothetical result in a constituency where the three main parties, Conservative, Labour and Liberal Democrat, all tie in votes – i.e. the candidates all get

exactly 33.3% of the vote share. This constituency would be located in the exact centre of the triangle. Constituencies would move around inside the triangle depending on the relative strengths of the three main parties. It was important to do this because the researchers needed to understand the shape of the vote distribution between the main parties after each general election. Figure 5.12 below shows the distribution after the 2005 general election. This clearly shows a large number of Labour/Conservative marginal seats (at the intersection marked with a dotted black line) but many fewer Labour/Liberal Democrat marginal seats. In the bottom third of the pyramid is a group of seats where the Labour vote is very low and the effective two-party competition is between Conservative and Liberal Democrat.

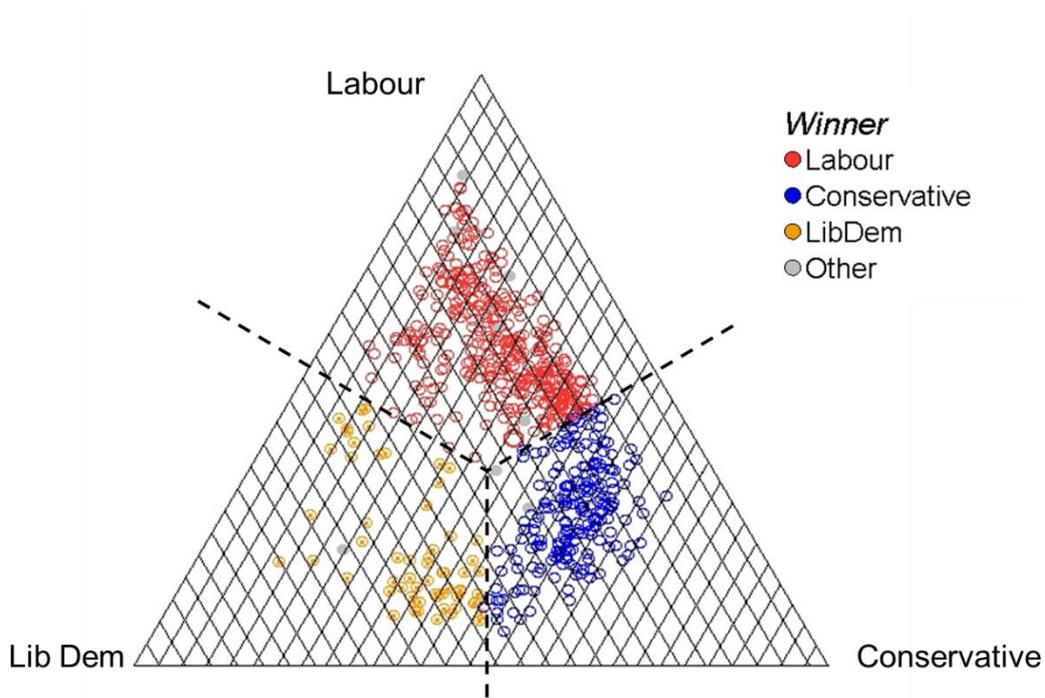
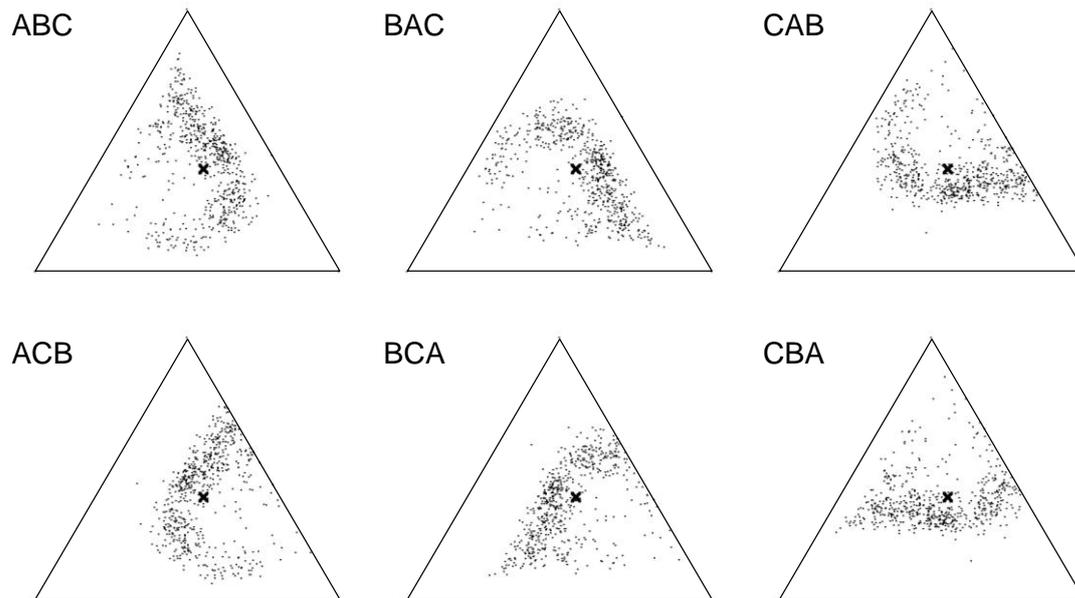


Figure 5.12: Distribution of three-party vote shares, 2005 general election

They also needed to see what happened to that distribution when they applied their new method for estimating bias. This involved running a number of simulations of what the election would have looked like if the original finishing order of the parties (e.g. ABC) had been different (e.g. ACB). The method required comparing the actual result with all possible

combinations of the parties' finishing order – ABC, ACB, BCA, BAC, CAB, and CBA). The resulting set of graphics is shown (Figure 5.13) below partly to demonstrate how the method works but also to show the deterioration in visualisation that occurs when there is the switch from colour to black and white – again the journal could not accept coloured graphics.



**Figure 5.13: Distribution of three-party vote shares:
ABC (actual), ACB, BAC, BCA, CAB, and CBA (notional elections)**

The final composite picture (the amalgamation of all six elections) is revealed in Figure 5.14. It is by from this display of the data that the researchers were able to summarise the distribution of overall bias between Conservative, Labour and Liberal Democrat parties for the 2005 general election.

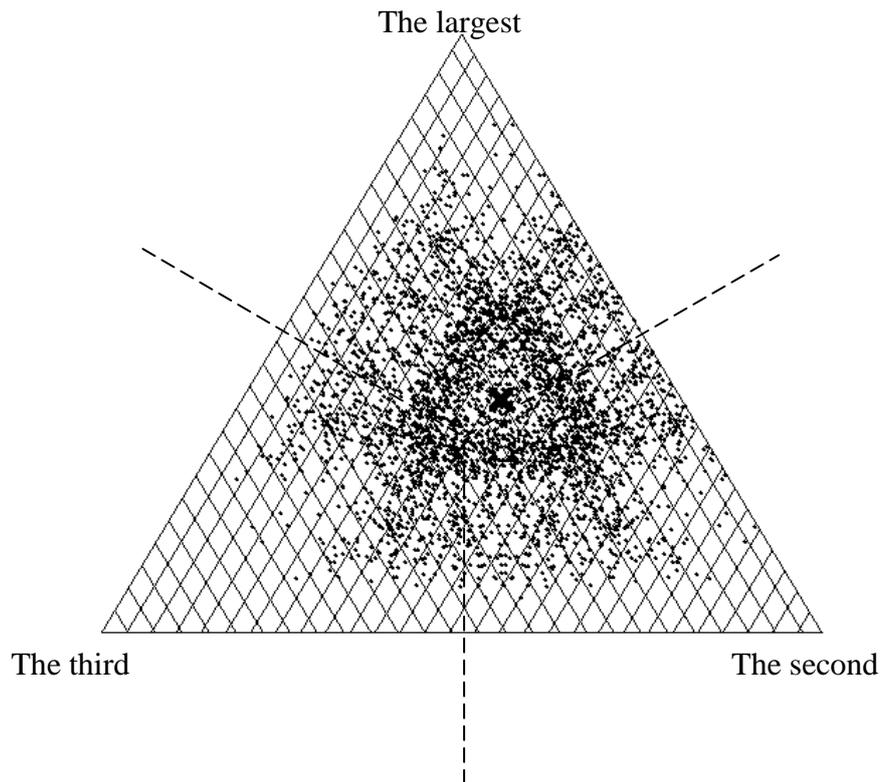


Figure 5.14: The superposition $ABC+ACB+BAC+BCA+CAB+CBA$

As stated earlier (Chapter 1) the paper format stands in the way of reporting data visualisations properly. The Elections Centre, for example, during this research decided that a better way of watching the effects of running different election simulations was to use Matlab software to create film-like animations that would show the changing dynamic of actual party competition for a series of general elections. The two figures below are simple screen-grabs of different parts of the evolution between the 2005 and 2010 general elections in this case. Each constituency is a point beginning with the 2005 general election which then ‘migrates’ into another area of the triangle dependent upon the direction of flow of vote. If the change in share affected only two parties, e.g. Labour and Conservative then the line would run along a certain vector but if there was also a change in vote share for the Liberal Democrats then the vector would take that into account also.

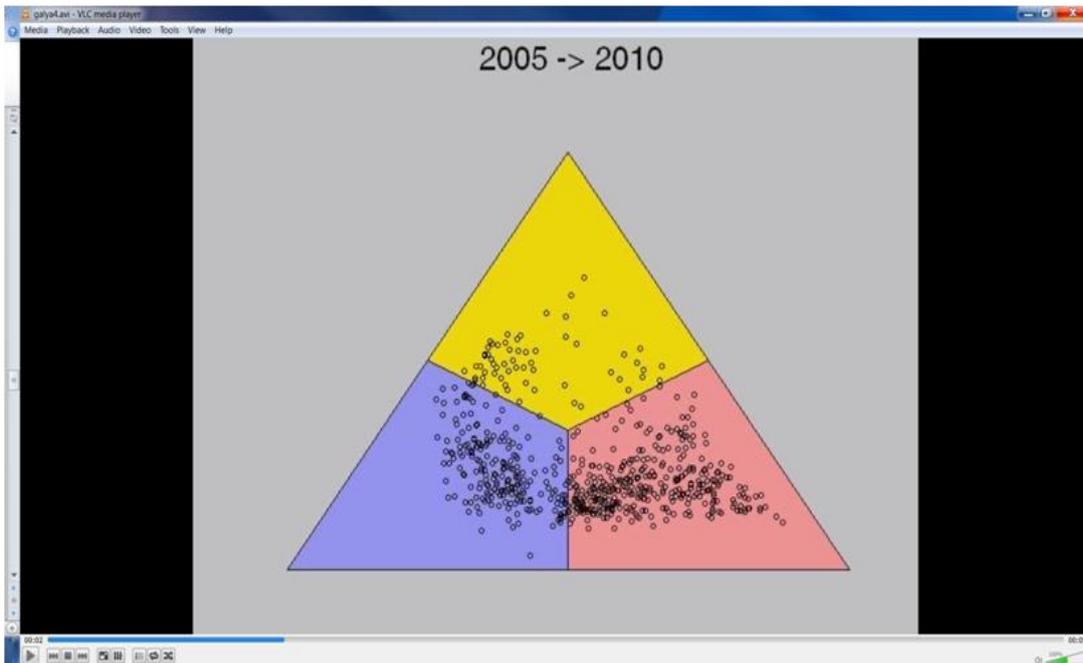


Figure 5.15: screen grab1 evolution of votes 2005-2010 general election

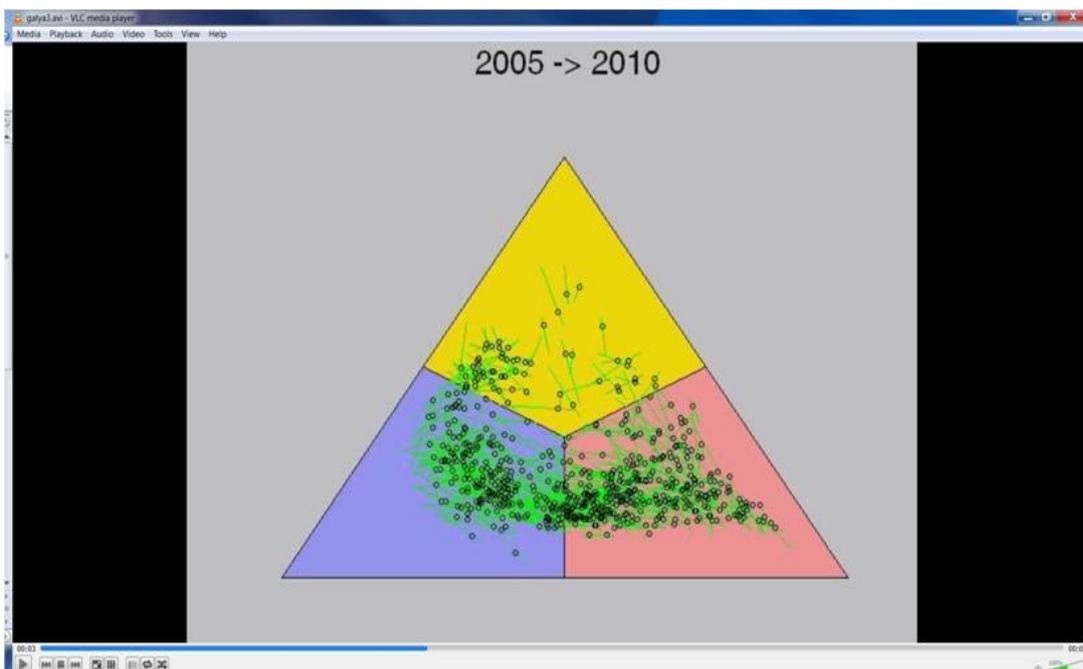


Figure 5.16: screen grab2 evolution of votes 2005-2010 general election

The first (Figure 5.15) shows the distribution of constituencies at the time of the 2005 general election (estimate seat shares because constituency boundaries changed during this period) while the second (Figure 5.16) shows the movement of each constituency in the three-party share space. The direction of each line is an indication of the flow of votes between parties during that election. Watching these animations provided us with a much clearer understanding about flow of votes that could be obtained from tabular data that would need to split the data into categories and would be difficult to read. The animations were in fact better than the static pyramids because with these the user had to move their eyes from one to another. Unfortunately, while this method was useful for visualizing electoral change it cannot be applied when submitting papers to academic journals unless those journals publish online versions that contain hyperlinks to data, animations, graphics etc. The technology is already available but the publishers appear unwilling or unable to use it.

5.4 Handling missing data and computing averages

For this second case study the focus is not about a relatively sophisticated method for visualizing data but instead it shows how traditional methods, in this case simple line graphs, are useful for trial and error experiments that seek to find the most appropriate method for summarizing the effects of different ways of calculation. This section focuses particularly on research undertaken by the Elections Centre in modifying its electoral forecasting model that uses the results from local council by-elections to estimate national vote shares (Rallings and Thrasher 1996; Rallings and Thrasher 1999; Rallings, Thrasher et al. 2011).

The basic by-election forecast model is:

Let *CONsh*, *LABsh*, and *LDsh* be the ward vote share at the by-election for Conservative, Labour and Liberal Democrats respectively while the parties' vote share at the previous ward election held in May would be *CONsh_{May}*, *LABsh_{May}*, and *LDsh_{May}*.

The national equivalent vote at the date of the relevant previous May local election is designated as *NEV.CON*, *NEV.LAB*, and *NEV.LD*.

The difference in share of vote between May and By-elections for enumerated parties might then be represented as follows:

$$\text{CONchange} = \text{CONsh} - \text{CONsh}_{\text{May}}$$

$$\text{LABchange} = \text{LABsh} - \text{LABsh}_{\text{May}}$$

$$\text{LDchange} = \text{LDsh} - \text{LDsh}_{\text{May}}$$

Finally, to estimate the current national equivalent vote we add to the previous national equivalent vote the difference between a party's by-election vote share and its vote share recorded at the May election.

Estimates of current NEV are calculated as follows,

$$\text{NEV.CON} + \text{CONchange}, \text{NEV.LAB} + \text{LABchange}, \text{NEV.LD} + \text{LDchange}.$$

This process is repeated for as many by-elections as fit our specified criteria and each party's current national equivalent vote is arrived at by averaging the results over a stated time period (Rallings, Thrasher et al. 2011).

From 2001 the rather dramatic change to the pattern of party competition had implications for modelling. In 2000 some 68% of by-elections featured candidates from all three main parties; this dropped to 62% in 2001 with a further fall to 55% in 2002. Although the proportion of three-party contests recovered from this low point it did not match the consistent levels seen throughout the 1990s when the electoral forecasting model was first developed. This, combined with fewer three-party contests in the main May elections led to model estimates being sourced from a declining level of data points – fewer and fewer by-elections could be included in the modelling because there were fewer cases where all three parties contested both the main May election and the subsequent by-election (Figure 5.17, Figure 5.18). As a result, the model estimates become less stable and oversensitive to

random variations (see, for example, model estimates for Conservative share of vote - Figure 5.19).

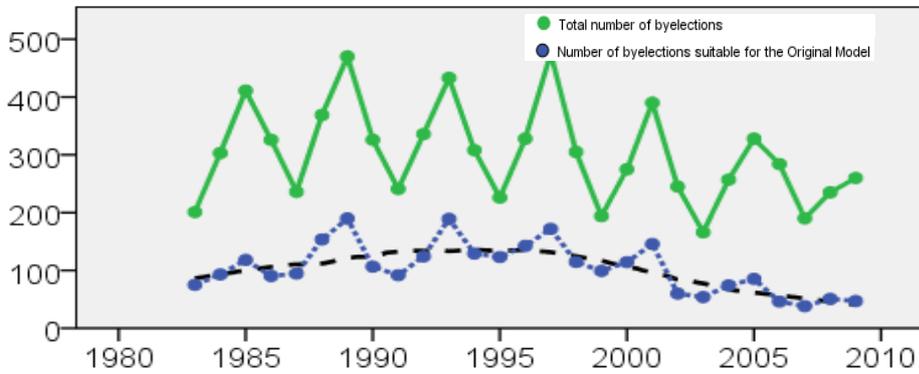


Figure 5.17: Decline in eligible by-elections for forecast model

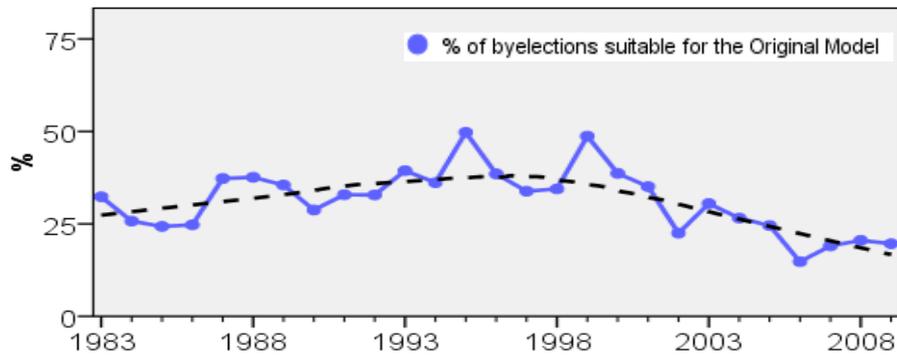


Figure 5.18: Declining percentage of by-elections used in forecast model

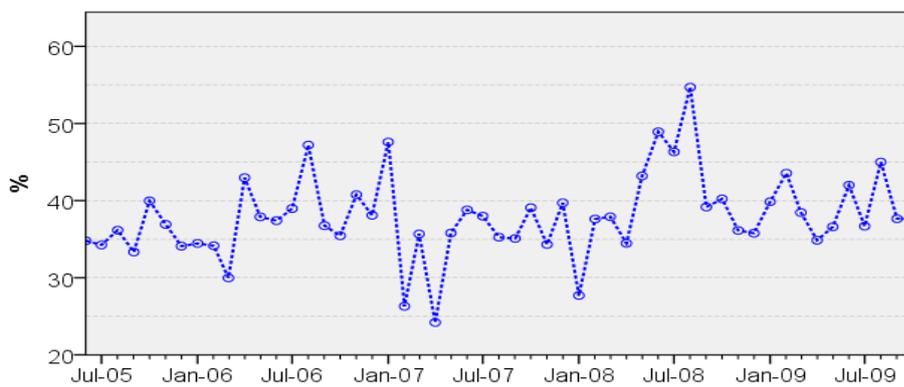


Figure 5.19: Old model estimates for Conservative share

The initial step was to investigate the changing structure of party competition and second to devise methods that might then compensate for missing values, permitting more cases to be used to estimate national support. A third aim was to determine the optimal time frame for averaging – it should be a trade-off between averages being responsive to new information and yet not over- sensitive to random variations.

A starting point was to examine for every case the structure of party competition at both the by-election and the previous May election. In more than 7,000 by-elections there was three-party competition at both the May and by-election in just under half - 3,425 cases. In a further 544 by-elections a Liberal Democrat candidate, present for the May contest was missing from the subsequent by-election. In another 250 cases and 77 cases respectively it was the Labour and Conservative candidate respectively that failed to contest the by-election. Sometimes, the process worked in the opposite direction with by-election vacancies attracting greater party competition than had the May equivalent. For example, in 648 cases where three main parties contested a by-election the Liberal Democrats had not challenged when the main May election was fought. In a further 255 and 155 cases it was Labour and the Conservative candidates respectively that are missing from the May election but are present at the by-election contest. In other examples the structure of party competition was partial but stable in the sense that perhaps only two of the three parties competed at both elections. In 457 cases, for example, only Conservative and Labour challenge one another with the Liberal Democrats absent on both occasions. In a further 331 examples the two protagonists are Conservative and Liberal Democrats while 104 cases are Labour versus Liberal Democrats only. The structure of party competition meant different responses in devising new procedures designed to include more of the cases that were available for devising model estimates.

The first examples consider cases where the pattern of party competition is more extensive at the main elections in May than it is for the subsequent by-election. Local voters in May could select from Conservative, Labour and Liberal Democrat candidates but the Liberal Democrats (in the example shown) decide to stand aside from the by-election contest .

Let all three main parties have candidates in May election but Liberal Democrats provide no candidate in the by-election:

May election → By-election
 con+lab+ld → con+lab

The imputation procedure implies the following notional by-election result:

$$\begin{aligned} LDsh^{new} &= \text{MINIMUM}_{\text{across district in May}}(LDsh) \\ CONsh^{new} &= CONsh - CONsh/(CONsh+LABsh) * LDsh^{new} \\ LABsh^{new} &= LABsh - LABsh/(CONsh+LABsh) * LDsh^{new} \end{aligned}$$

In order to preserve the actual *direction* of changes for all parties some restrictions are imposed:

If $CONsh > CONsh_{May}$ but $CONsh^{new} < CONsh_{May}$ then we set **CONchange = 0**

If $LABsh > LABsh_{May}$ but $LABsh^{new} < LABsh_{May}$ then **LABchange = 0**

Note: In above formulae superscript ‘new’ reflects notional by- or May election when any procedure of imputation is applied. Following the imputation procedure, notional election result (i.e. ‘new’) is then treated in the usual way for the purposes of estimating NEV.

Previously, the model ignored such cases but the new method estimates a notional by-election share for the Liberal Democrats which is equal to the minimum share value the party achieved across the relevant local authority at the May election. So, if the ward is in Plymouth then the model finds the ward with the lowest Liberal Democrat share and applies that share to the example ward. The theoretical basis for this assumption is that the likely

explanation for the Liberal Democrats to withdraw from the by-election is the expectation of receiving a low level of votes. The votes for the parties that did contest the by-election are then adjusted to take account of the estimated vote for the absent party. It is important that the process of normalising votes in this way does not distort the actual result, for example, transforming a positive change in vote share for one of the parties that did contest into a negative one. In order to prevent this from happening the share change for a party that both contests and increases its share is never allowed to fall below zero. Following these adjustments the by-election result is then treated in the usual way for the purposes of estimating national vote shares.

There are other occasions when the extent of party competition is even more fragmented; three main parties contest the May election but now **two** fail to contest the by-election. For example, both Labour and Liberal Democrats do not present by-election candidates to challenge the Conservatives.

Example 2: Imputation for missing vote share for both Labour and Liberal Democrats absent from by-election contest (con+lab+ld → con).

Notional By-election result:

$$LDsh^{new} = \text{MINIMUM}_{\text{across district in May}}(LDsh)$$

$$LABsh^{new} = \text{MINIMUM}_{\text{across district in May}}(LABsh)$$

$$CONsh^{new} = CONsh - LABsh^{new} - LDsh^{new}$$

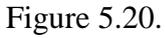
If $CONsh > CONsh_{May}$ but $CONsh^{new} < CONsh_{May}$ then we set **CONchange = 0**.

If $CONsh \leq CONsh_{May}$, then

$$LABchange = \text{MISSING}, LDchange = \text{MISSING}; CONchange = CONsh - CONsh_{May}$$

The problem here is that the change in Conservative vote share from May is likely to be inflated because of the absence of two of its competitors. In such cases we proceed with the method described above, now estimating by-election shares for both Labour and Liberal

Democrats based on the minimum values across the local authority at the previous May election. Following this procedure the Conservative by-election share is recalculated by subtracting from its actual share the estimated shares for both Labour and Liberal Democrats. Again, if the actual Conservative by-election share is an increase from its share in May but the process of estimating shares for the two missing parties transforms that to a decrease then change is limited to zero. Where the Conservative by-election share does actually decrease, despite the two missing parties but presumably because of the support for independents or other smaller parties, then no estimates are made for change in share for Labour and Liberal Democrats while the Conservatives are given the actual May to by-election change.

These two examples are sufficient to show the approach that was made to missing data either from the May election or the later by-election. The result of making such compensations is that a greater proportion of by-election cases may be included in modelling national equivalent vote estimates. The extent of that increase is shown in  Figure 5.20. The two lines show the proportion of by-elections used in estimating national vote shares. The broken line is the proportion that were usable under the strict criteria of three-party competition in both the May and by-elections and a large fraction of total votes cast for the main parties. The solid line is the proportion after compensating for incomplete three-party competition and allowing for missing values when one or other of the three main parties fails to compete at either election. The most dramatic difference occurs in the period after 2001 when there was the significant move away from three-party competition.

Of course, we needed to be careful that estimating missing vote shares did not ruin the model. We did this using line graphs where we could compare the consequences for the model at a given point in time when missing data was imputed using the method outlined above.

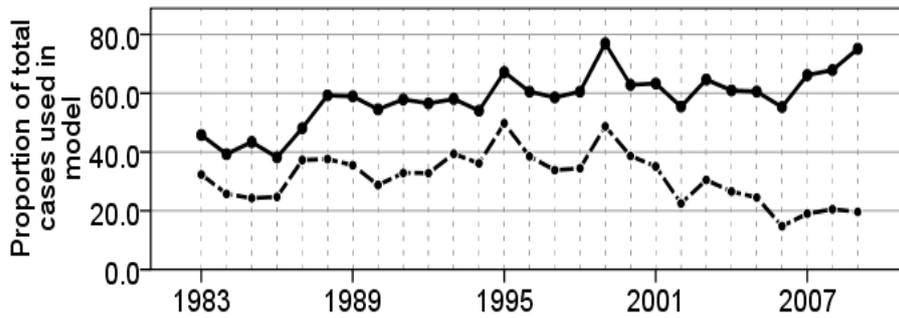


Figure 5.20: Consequence of new data selection criteria proportion of by-elections used by original and revised models

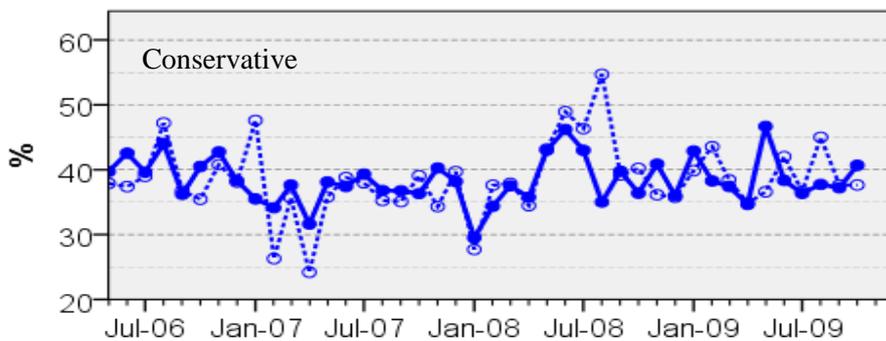


Figure 5.21: Estimating missing vote shares and model stability

Plotting each party's forecast vote share over a long period reveals that the effect of estimating votes for missing parties is to reduce the amount of volatility in the monthly estimates due to the increased number of cases used (Figure 5.21). In the period from 2006, for example, the Conservative three-month average varies between 23 and 55% (range 23 points) using the original method while the range is just 16 percentage points (30-46%) for the revised method. Similarly, Labour varies between 10 and 32% and between 20 and 30% for the old and new methods respectively while the Liberal Democrats were ranging between 15 to 43% and now lie between 21 to 42% (not shown here).

Another measure of volatility is the level of change of support from one month to the next. Fluctuations that are very pronounced are more likely to be of a random nature rather than reflecting real changes in the public mood. Compared with the original method the revised model estimates demonstrate greater stability in short-term support³¹. Visually speaking, this is apparent in the relative smoothness in the solid blue line compared with the dotted line but it also shows the old and new model estimates are more or less in agreement with one another and therefore the procedure for calculating missing data and increasing the number of cases in the modelling does not adversely affect the model's predictive power.

The final element in the revision process was to consider how best to create estimates for a given point in time. The graphs that are shown in the final part of this case study refer only to the Conservative party although identical graphs were also constructed for both Labour and the Liberal Democrats.

Providing estimates of national support based on by-elections from a single week would be ill advised since these can vary considerably, influenced by a range of local and national factors and also being affected by relatively small numbers of cases. A more robust approach is to use a broader time period (Figure 5.22). This has the effect of smoothing large fluctuations that may occur using weekly data.

³¹ In terms of Conservative support during the 2005 parliament the original method suggested a maximum increase of 12 percentage points and a maximal decline of minus 24 points (standard deviation = 4.5) while using the revised method the Conservative figures do not change by more than 5.6 points (std = 1.9) across a two-month period. The equivalent figures for Labour are +5 to -6 (std = 2.0) now changing between +4 to -3.5 (std = 1.4) and Liberal Democrats +19 to -15 (std = 4.7) using the old method to change between +4 and -10 (std = 2.2) on the revised model.

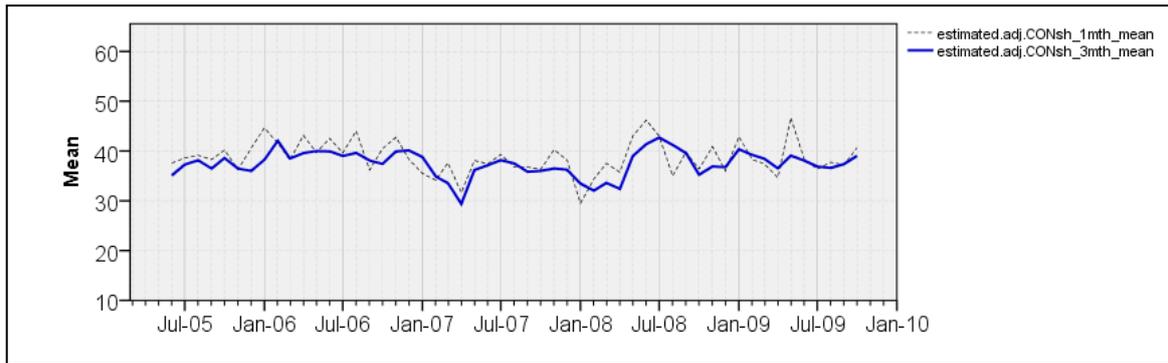


Figure 5.22: Effect of using one month or three-month averages

Closer examination of results and trends over a twenty-five year period suggested that a more reliable procedure is a weighted quarterly moving average³². Thus, each forecast is based not simply on the figures for a single month but also some information from the preceding two months (Figure 5.23). The weighting procedure takes into consideration the time elapsed from when each by-election occurs and the date of forecast, usually the last day of the month of interest. Thus, a by-election that happens on the last day of June has a bigger impact on June's averages compared to a by-election that held on April 1. There is a linear decrease in weights that reflects the days elapsed from the forecast date. In turn, the above June by-election will also have an impact on July and August 'averages' but its impact on the model estimate is reduced with time. In short, if the number of by-elections stays more or less the same across all months then the influence of this June by-election decreases linearly (it is highest for June, smaller for July, and smallest for August) before it is completely removed from the quarterly calculation. This experiment clearly shows the advantage of using the weighted three-month average calculation.

³² Our experience is that employing a three month average provides a better smoothing of the data than a two month average.

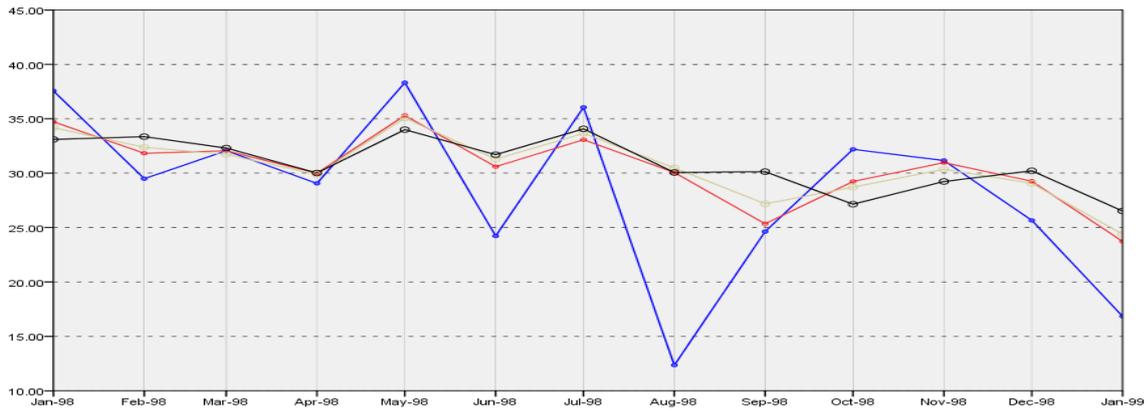


Figure 5.23: Effect of different weighting schemes

Note: blue line is monthly mean; light brown is three-month average with weights 3:2:1 (after taking into account number of by-elections in each month); black line – similar to the previous line but weights are 1:1:1; red line – average for 3 months with weights that account for number of days from a by-election to the month of interest.

Having established the advantage of first replacing missing data and second the strength of a three-month running average the research team tried to find other ways to improve the model and once again visualizing the data provided crucial information about the effects of these experiments.

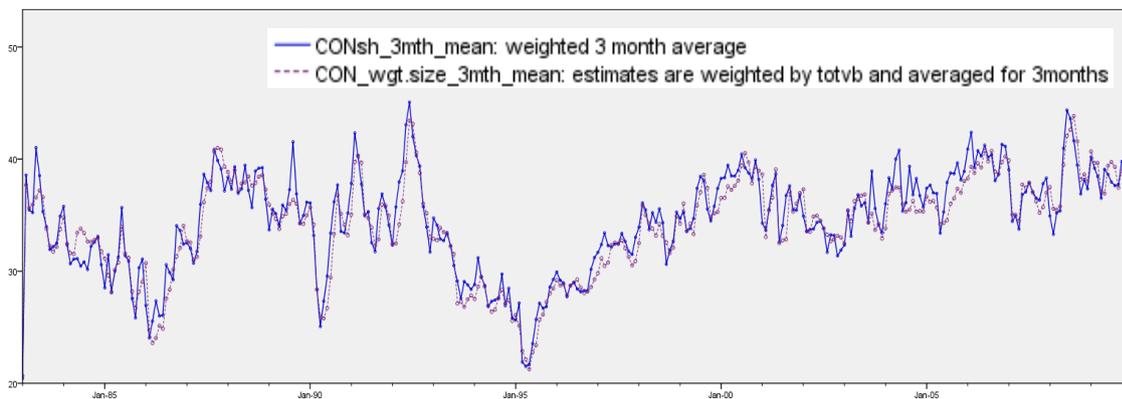


Figure 5.24: Adjusting for size of total votes

Figure 5.24, for example, showed the rather small effect that occurred when we tried to take account of ward electorate size. The thinking that lay behind this experiment was that a ward containing, say 10,000 voters would reveal more information about the state of electoral opinion than a ward with just 1,000 voters. There is no consistent and large effect according to the line graph and so this idea was abandoned.

Another line of thinking was that voters behave differently in by-elections if they believe the election is going to be close or not. To test this we split by-election wards into two categories. The first category could be described as marginal wards where the winning party's majority at the previous May election was below 20%. The second category is all other cases. The results of this is shown in Figure 5.25. This appears to show a higher level of Conservative support in the marginal wards than when the safer wards are examined.

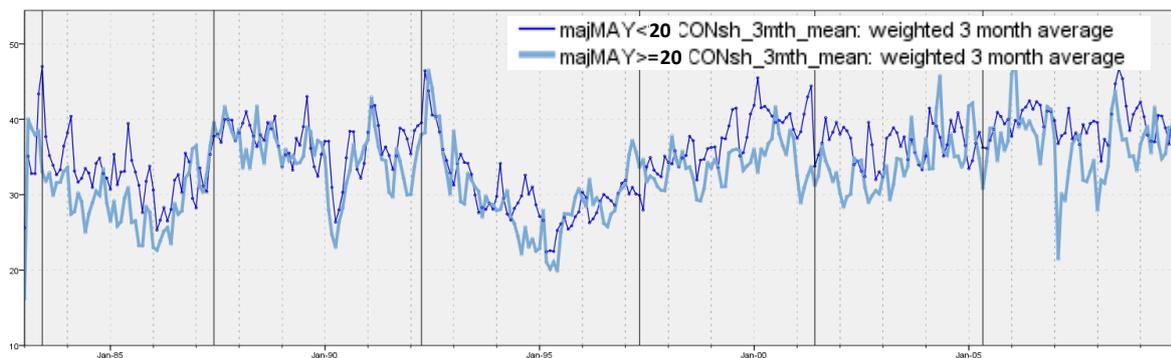


Figure 5.25: Effect of controlling for ward marginality (cut point 20% majority)

It was decided to explore the potential of this procedure in two other experiments. The first involved using a lower cut point of 15% (Figure 5.26) while the second approach created three categories of marginality instead of just two (Figure 5.27).

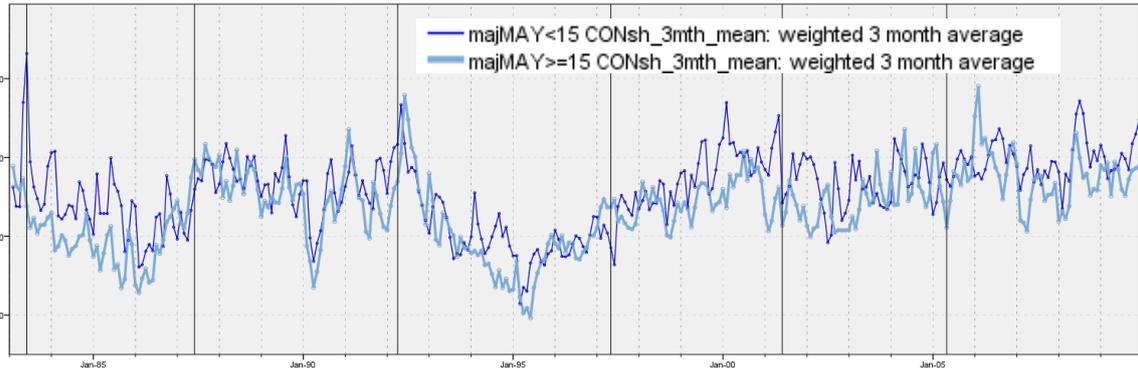


Figure 5.26: Effect of controlling for ward marginality (cut point 15% majority)

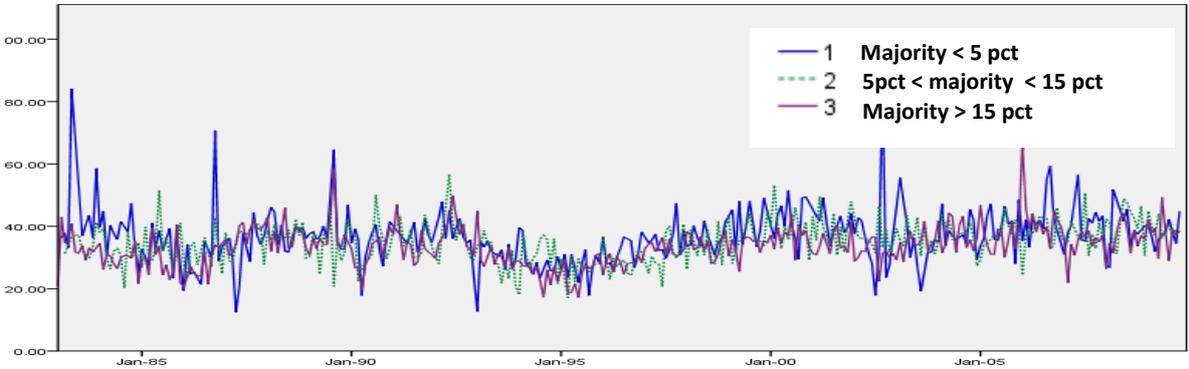


Figure 5.27: Effect of using three cut points

Although these experiments were interesting the visual differences were not sufficiently distinct to make the adjustment permanently. It was also felt that this might produce problems later if the distribution of by-elections suddenly meant that there were many more/fewer by-elections in particular categories.

Another line of thinking was to test for whether geography and the distribution of by-elections had any effect on the levels of support modelled for each party. If the Conservatives were doing well in the south but not in the north, for example, then a disproportionate number of by-elections in one part of the country might skew the forecasts. As a way of testing this we divided by-election wards into two categories – urban and rural. Figure 5.28 does suggest that there are differences with Conservative support being forecast

to be higher in rural areas compared with the pattern in urban wards. The most important point to consider was that the long-term trends were roughly the same – it was not the case that Conservative support would be rising in one type of ward at the same time that it was declining in another type. Again, the researchers felt that it was more practical to keep all of the by-elections together and not to sub-divide them into marginal/safe wards or urban/rural wards.



Figure 5.28: Urban vs. Rural by-elections

Figure 5.29 shows the final comparison between the three-month weighted averages for the original (dotted line) and revised (solid line) models after taking into account the best method for averaging and weighting by-election data. It is clear just by eye-balling the two lines that one is more consistent and smoother than the other and that the new procedures would create a model forecast that was more stable than what had come before.

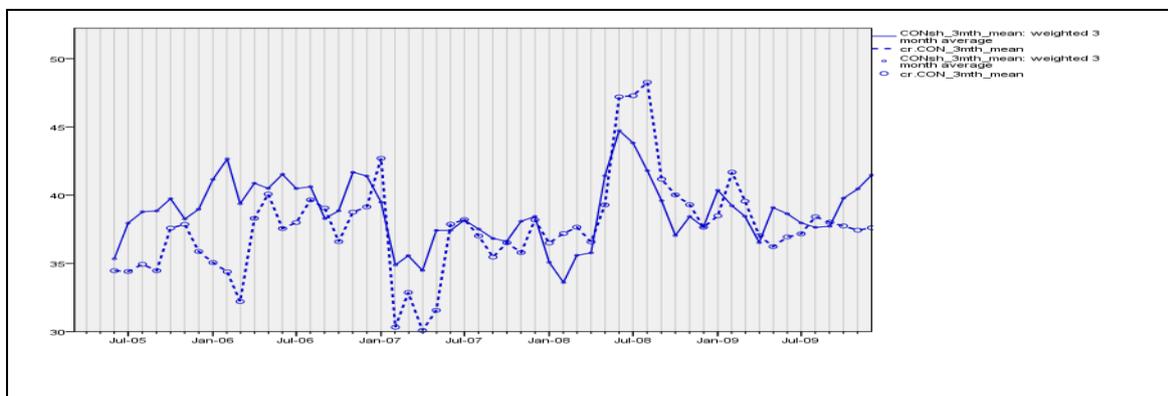


Figure 5.29: Comparison of forecasts for Conservative national vote share

A final procedure for measuring the effects of different data transformations is to create forecasts (using exactly the same weighting and average procedures) but based on random sub-samples of the dataset and then comparing these forecasts with one another (Figure 5.30). If the lines run parallel (or more or less parallel) that suggests the methodological changes are robust. If they run off in different directions and do not follow the same trends then that suggests there is something about the methodology that should be investigated further. As the graph shows the randomization procedure did not result in any divergences in the forecasts. Achieving the same level of confirmation using tabular data or some kind of statistical techniques might have achieved the same aim but it would not have been as quick and effective as using line graphs was.

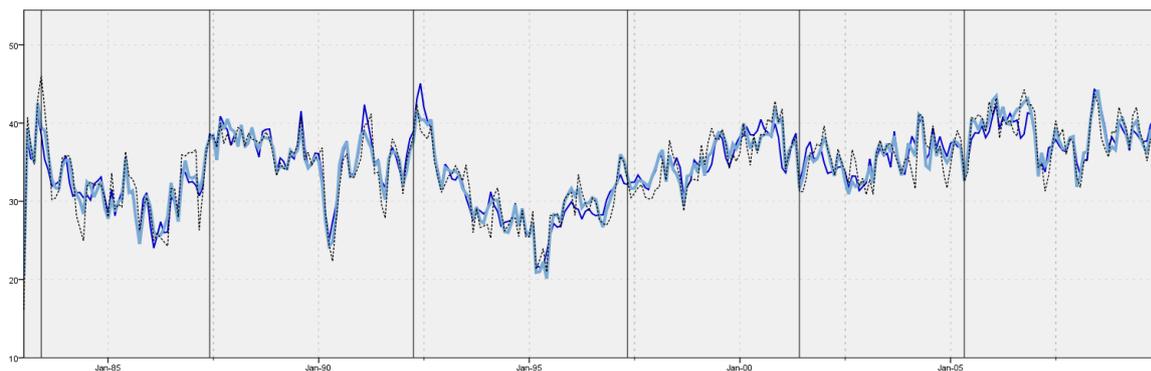


Figure 5.30: Comparing random Sub-samples of data

The first test of the revised model lay in its ability to forecast correctly the May elections using the April model estimate. The first iteration of the model was a reliable method for forecasting the national equivalent vote (NEV) for local elections and it would be a retrograde step if the revised method performed less well. Figure 5.31 shows for each election year since 1993 both the April by-election model vote share (comprising data from the April, March and February results) and the eventual May NEV for the three main parties. The solid line represents the by-election model share and the dotted line is NEV. Overall, the

methods for estimating missing data do not appear to impact negatively on the model forecasts.

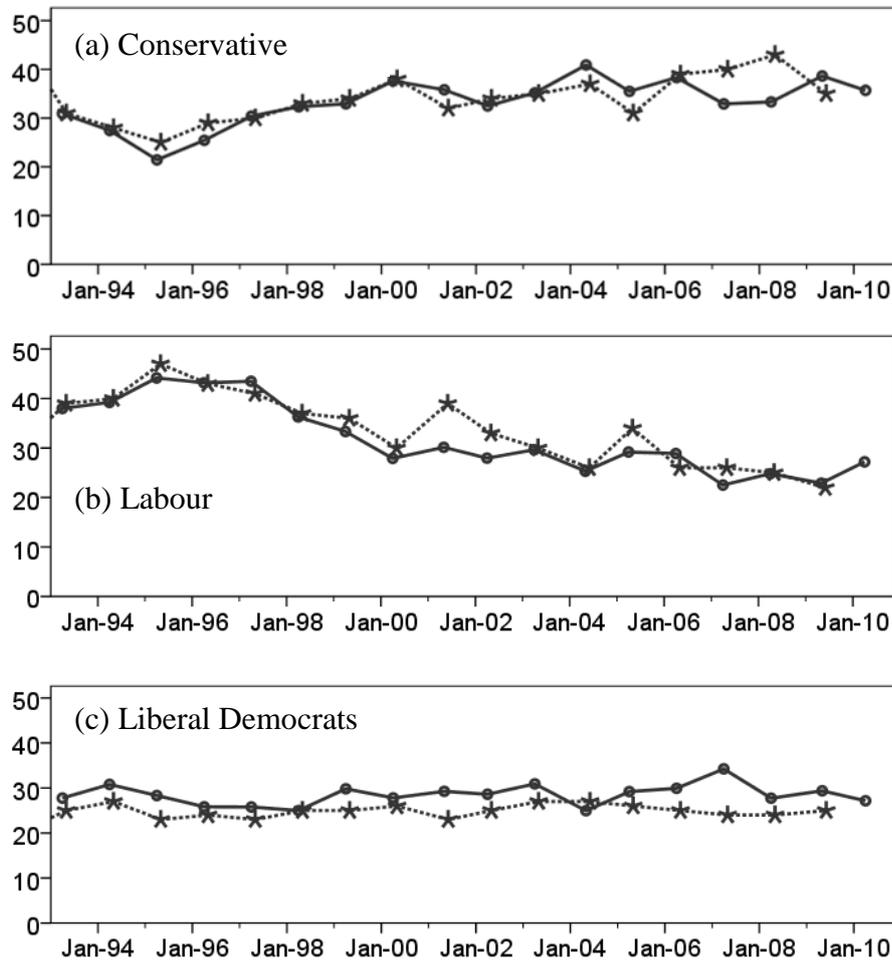


Figure 5.31: Revised model and National Equivalent Vote, 1993-2010

For the Conservatives the two curves are close together but there is no consistent pattern in terms of the forecast accuracy. A close examination of the pattern of support for Labour again shows that the two curves are close together with two clear exceptions, both of which are general election years (2001 and 2005). For the Liberal Democrats the pattern is the reverse of Labour's performance. The tendency is for the by-election model to give a higher estimate than Liberal Democrats' actual NEV. This feature is particularly noticeable in both 2003 and four years later in 2007, the peak of the local electoral cycle in terms of council

seats up for election. What these graphs were able to demonstrate, however, is that adjusting for missing data and therefore increasing the number of cases that could be included in the modelling had been successful.

5.5 Conclusions

In the first of our case studies the intention was to show how information visualisation is used and developed by researchers in different fields of political science and electoral geography. When preparing different research projects the Elections Centre staff cooperate with one another and also people from other parts of the University and other Universities. The collaborations also include not just political scientists but also geographers, computer scientists, mathematicians etc. Researchers that collaborate in the natural sciences (physics, chemistry, biology) would normally use graphical representations of their data but for most political scientists this is probably the exception. In fact, by not doing this more political scientists are ignoring a very valuable tool that could help them formulate new methods of working and then to trial those methods on actual data. We believe that such practices might be adopted by other political scientists that would not normally regard themselves as ‘quantitative’ but whose research interests might nevertheless be quantifiable, for example, in graphically identifying patterns in text or speech.

This chapter has shown that maps are very powerful tools for making the point in a very clear and precise way. Using GIS in research speeds up the process of understanding when there is a lot of information stored in the spatial data. Well-designed visual representations can replace cognitive calculations with simple presentation and improve understanding, memory, and decision making. By making data more accessible and appealing, visual representations

may also help engage more diverse audiences in exploration and analysis. The challenge is to create effective and engaging visualisations that are appropriate to the data statistical display.

The collaborations with Dorling and Orford, both geographers, that was described in the first part of the Chapter, used GIS to preview research findings and to measure progress in formulating the argument. Later the Elections Centre developed its own skills base in geographic mapping software so that it was able to incorporate this into its publications relating to voting for the Greater London Authority elections in both 2004 and 2008.

The two case studies independently show the power of graphical representations of data during the process of constructing a new method for measuring electoral bias and making adjustments in an election forecasting model. Graphical representations visualize the data and any changes to the data more easily than any other method. Visualisation helps people to understand what may be subtle differences in the data from adopting one way of measuring something to another. Triangles are an extremely powerful way of seeing the dynamics of three-party competition because it is immediately apparent what is the relative distribution of seats either for an actual election or in our case simulated elections. When the topic being investigated is a time series of by-election data the best and easiest way to view these data are with line graphs. Making adjustments in those data and then visualizing the effects of making those changes is again much simpler when line graphs are used.

6 Explaining voting systems to the general public

6.1 Introduction

The previous chapter demonstrated clearly how data visualisation in the form of graphical displays and mapping facilitates the research process and is a more efficient and effective means of communicating ideas within a research community than simple text-based methods. But while this has been largely ignored by researchers when finally reporting research findings the topic of this chapter is attracting considerably more attention as scientists and social scientists alike realise the importance of sharing their research with a general public that now has a greatly enhanced ability to access information in many different formats. This requires a separate range of skills because the target audience is now rather general, rather than specialised. Where some aspects of the data could be taken for granted when the visualisation is designed for internal use that is certainly not the case when the intended audience becomes the general public. In fact, the process becomes much more difficult because of that as this chapter highlights.

In this chapter, therefore, we examine two attempts that have been made to introduce some complex aspect of voting systems to the general public using PowerPoint visualisations. In the previous chapter we examined how using visualisation can assist the research process but when the objective is to bring information to a wider audience then that places different kinds

of difficulties. This chapter considers how the Elections Centre prepared videos from powerpoint presentations that were then placed on YouTube. The first subject was the decomposition of electoral bias while the second video gave a short introduction and explanation of the Supplementary Vote (SV) system that was used for the election of Police and Crime Commissioners in November 2012.

Compiling these videos presented a range of different challenges. The most important challenge was how to explain to a public audience how these complex procedures work. It was decided that in tackling the subject of electoral bias first we could learn from this how to make the second video work better. It was decided that we would invite viewers to undertake a short survey after the viewing of this video. Analysis of the survey responses might provide important information about what viewers had found easy to understand and what parts of the video had not worked very well.

The subject of the second talk, SV, was likely to be read by a much larger audience since we felt that people that were going to vote in the PCC elections might search on the internet for an explanation of how the voting system worked. It was important, therefore, to learn from the mistakes of the first video and compile a video that people would find interesting and informative.

Second, these videos were not intended to be interactive (see chapters 7-9) and therefore the viewer was assumed to be inactive. This meant that we had to take into account the amount of time that each video took to view. When it was recorded the Electoral Bias video was timed at almost forty minutes. The average time for most videos posted on YouTube is about a quarter of that and probably reflects most people's attention span. Another view was that only people with a particular interest in elections would bother to watch a video on such a

topic and therefore the video length was not really a problem. Although we discussed cutting the original video into four equal parts (see section 2 below) it was decided to retain it as a single video and then wait for the reactions from those people that had watched it.

The third element of this challenge was to provide something different to the videos on YouTube and iTunes U which mostly consist of specialists delivering lectures to students in a lecture theatre. With the proposed expansion of iTunes U and universities such as Harvard, Berkley and the London School of Economics moving towards the online curriculum one of the difficulties they will face is the nature of the subject delivery. Being physically inside the lecture theatre is a completely different experience to sitting in front of a computer screen viewing a lecture, even if the streaming is live. In creating these videos, therefore, we felt that it was important to think about how the public/students might engage with the subject better if they were not presented with the usual format of lecturer/audience. There are some good examples on YouTube of, for example, presentations of statistics and mathematics that have graphic presentations that are voiced over by the presenter but very few that deal with political science topics. These were used as models when creating our own presentations.

The chapter is divided into two main sections. The first considers the design and content of the PowerPoint presentation on electoral bias and the subsequent analysis of the survey responses provided by its viewers. The second section deals with the design and content of the video created to explain how SV works. There is a brief analysis of the survey data collected from people who watched that video. In the concluding section we consider what are the strengths and weaknesses of using this method of visualising data to explain what are complex voting systems to in-expert users.

The rationale for conducting surveys was to establish the utility of visualisation techniques and the PowerPoint presentation applications for political scientists that were interested in explaining the complex information to the general public. A growing trend in the field of public participation is to incorporate computers to enable participants to view and explore interactive maps (see chapter 4) and other multimedia information resources. These kinds of resources can be particularly useful in deliberations about complex political issues, where they are used to help explaining the morass of election data and the voting process.

The surveys described there were designed to be carried out by means of online questionnaires. They were designed to collect views and responses directly from people who watched the YouTube and iTunes U presentations. Face to face interviews were felt to be unnecessary (and perhaps even inappropriate) as the answers might be influenced by the interaction between interviewer and interviewee; another reason for choosing to conduct the questionnaire online was to avoid this potential for bias. With hindsight, given the rather poor response to the survey questions that were asked, a better tactic would have been to convene a focus group comprising members of the general public. Interviewing this group before and after they were shown the video might have better demonstrated whether any knowledge had been acquired after watching the video and, moreover, would have identified the strengths and weaknesses of the data visualisation process.

6.2 Electoral bias

6.2.1 Background

One of the most important research topics within the Elections Centre over the past five years has been to develop a new method for the decomposition of electoral bias. This was mentioned in the previous chapter where we considered how triangles had been used for visualising competition in a three-party system. It was decided at an early stage that explaining the use of these triangles to the general public would take too much time and might be the topic of a video in its own right. Instead, we approached this presentation in terms of why would the public want to view such a topic in the first place.

The answer to that question was the context of the 2010 general election and the subsequent decision to change the parliamentary constituency boundaries. In the run-up to the 2010 general election many people were confused about the reasons why the Conservatives needed a large lead over Labour in the popular vote in order to win an overall majority but the reverse situation (Labour lead over Conservatives) was asymmetrical. During discussions about the state of the opinion polls it was pointed out that the Conservatives needed a ten percentage point over Labour for Cameron to become Prime Minister but that Gordon Brown could remain in power if Labour's lead was just a few percentage points.

The conclusion that a number of people reached was that while it had been in government Labour had manipulated the parliamentary constituency boundaries to improve its own position. In fact, the result of the boundaries review before the 2010 general election showed the overall impact was that Labour would be defending a notional majority over all other parties of 48 at the next general election, compared with the 66 seat majority achieved in 2005. The Conservatives made a net gain of 12 seats from the new boundaries and Labour

had a net loss of seven seats. Despite this advantage for the Conservatives they were convinced that the re-drawing of boundaries had been hijacked by Labour (Borisjuk, Johnston et al. 2010). The subject of parliamentary reviews became an important part of the coalition agreement following the 2010 general election. As well as a national referendum on the Alternative Vote there would be a reduction in the size of the House of Commons from 650 to 600 seats and the procedures for re-drawing boundaries would also change (Johnston, Pattie et al. 2012).

The problem was that too many of the people involved in bringing about these changes did not understand precisely why there was an electoral bias towards Labour. They believed that the bias could only be produced by the inequalities between electorates. This is known as malapportionment. Since the boundary commissions were responsible for equalising electorates then they must have failed to do what was being asked of them. This view is mistaken and fails to take into account the different features of the electoral system that contribute towards bias. Correcting this misunderstanding became the goal of the Electoral Bias presentation.

6.2.2 Structuring the presentation

Given that the presentation was intended for the general public to view and not specialists on voting systems it was first decided to simplify it by removing some of the elements that contribute towards the manufacture of electoral bias. In the context of the UK there are six factors that may contribute towards the creation of bias. These are:

- National differences resulting from differences in the respective electoral quotas.
- Differences in electorate size between constituencies additional to those reflecting the varying national quotas.
- Differences between constituencies in their number of abstentions.

- Differences between constituencies in the number of votes obtained by ‘third parties’ (i.e., other than the Conservatives and Labour).
- The pattern of constituency victories for those ‘third parties’.
- Differences between the parties in the efficiency with which their votes are distributed across the constituencies (Rallings et al. 2008).

In recent general elections the most important of these elements (in terms of the size bias) have been electoral equality, voter turnout and vote distribution and so it was these three features that were included in the presentation.

The other important aspects of the presentation had to be the question of first, whether or not the boundary commissions had failed to carry out their reviews properly and second, what would be the likely impact of the next boundary review on bias. The view from within the Elections Centre was that because electoral equalisation was only a small part of the overall bias then the likely impact of the boundary review scheduled before the 2015 general election was always going to be small.

Having decided that the structure would be a short discussion about the meaning of electoral bias, three sections describing factors that contribute towards bias and the context of recent boundary reviews we wrote a short description that could be written quickly and easily by anyone accessing the video. This read as follows:

‘This presentation considers electoral bias, the factors that contribute towards that bias and examines in detail the 2010 UK general election.

Bias can be decomposed into different elements, three of which are examined in detail in this presentation. First, there is bias caused by unequal electorates or "malapportionment". Second, bias caused by differences in turnout or abstention. Third, bias caused by differences in vote distributions or "geography". Of course, in a real election all of these bias factors operate at the same time and so the 2010 UK general election is used to show this.’

6.2.3 Electoral Bias and its components

Explaining electoral bias

The first challenge was to explain the meaning of bias. Figure 6.1 shows a screen grab from this segment of the presentation. The screen began by showing the two bars on the left hand side of the screen. This graph shows the extent of the lead in the national vote that the Conservatives would require if they were to win a majority of 20 seats in the House of Commons. The vote shares are shown as Conservative, 39.7% and Labour 27.0%. Following the presentation of the two columns the percentage point gap (in this case 12.7) was added to emphasise how big that gap needed to be. The second stage of the animation then introduced another histogram featuring Conservatives (in blue) and Labour (in red) but this time the example was about a Labour majority in the House of Commons of 20 seats. This showed a scenario for the next general election where Labour would get 35.4% and the Conservatives would win 31.4% - a gap of 4% was then added for emphasis. This slide animation sets up the situation for the non-expert user and says that for an identically sized majority the Conservatives need a much bigger lead over Labour than Labour's lead over the Conservatives for the same kind of majority.

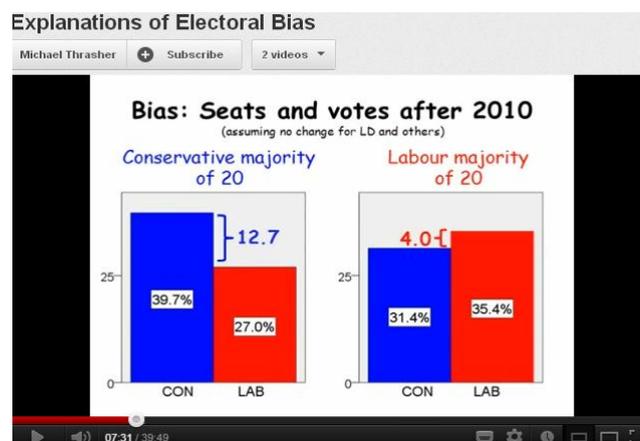


Figure 6.1: Introducing electoral bias in the UK context³³

³³ <http://www.youtube.com/watch?v=cLLgLDkBD00>

Having established what was the scale of electoral bias towards Labour and against the Conservatives it was time to introduce the hypothetical situation that would be used throughout most of the video. This needed to be extremely simple and easy to understand. Figure 6.2 sets up the basic story.



Figure 6.2: Introducing the hypothetical election

In this situation there is a country where there are only two parties, party A and party B. We discussed using three parties by adding party C into the list. This would assist people who knew that the UK now had three relatively large parties and not two parties. On the other hand all of the examples that could show how electoral bias works would work best using just two parties. We agreed that the two party explanation would be the simplest to use. The assumption, therefore, was that the two parties (A and B) fought an election where the result was a dead-heat in terms of the national vote with both parties winning 50%. However, party A would win 58% of the seats while party B would win 42% of the seats. Although both parties had equal vote shares they had unequal seat shares – the system was biased towards party A and biased against party B. To help with people’s understanding of the situation

even more we decided that in this hypothetical country there would be exactly 100 constituencies. In this way each constituency was worth exactly 1% of the total seats.

Bias and unequally sized constituencies.

The first example of decomposition of bias referred to malapportionment or the effect that is caused by having constituencies of unequal size. As Figure 6.3 shows the assumptions that lay behind this example was that half of the 100 constituencies had larger than average electorates while the remaining half were smaller than average electorates. The other two features were kept constant and it was assumed that turnout was 100% everywhere and that the vote split between winners (60%) and losers (40%) was exactly the same everywhere. It was important to do this because this is the only way to show the bias effect caused by malapportionment. When we composed this slide it was important to keep the text to a minimum in order to try and keep the presentation as short as possible.

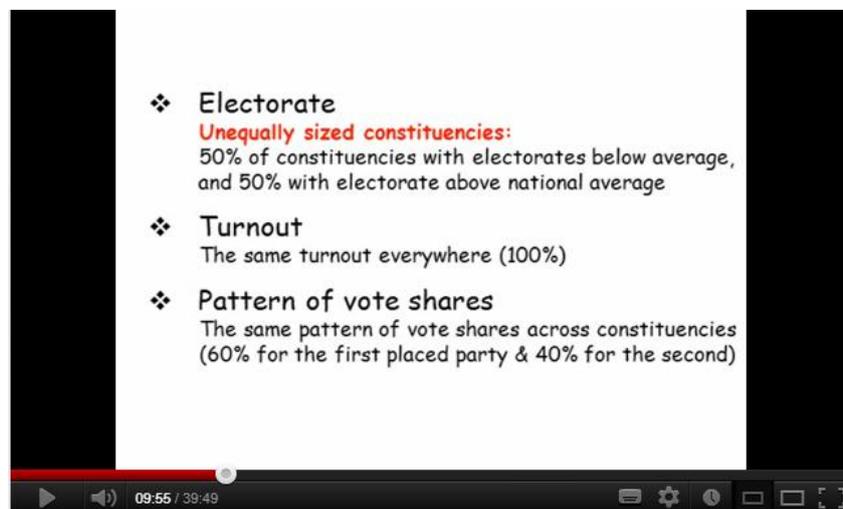


Figure 6.3: The contribution of electoral inequalities to electoral bias

The grid of 100 constituencies shows differences in the size of each according to the size of the electorate (Figure 6.4). The fifty rectangles shaded with a grey background are those with

below average electorates. The fifty rectangles shaded in yellow are the larger electorates and are slightly larger in size than the grey ones. This was intended to provide a visual cue of the differences between the two types of constituency.

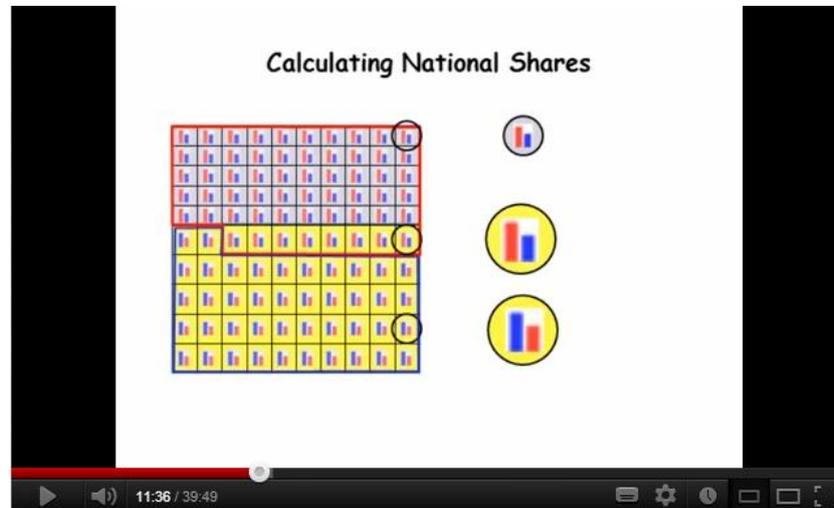


Figure 6.4: Differences in electorate size and the pattern of party wins

The first part of this animation showed a simple grid (grey and yellow shaded) and then the distribution of the party vote shares was then shown. The red party A wins all of the seats in the grey shaded/small electorate constituencies (the height of the red/blue histograms are all 60:40) while in the yellow shaded seats it is the blue party that wins 42 of these leaving the red party to win 8 seats. This is emphasised by a red border and a blue border that contains all the seats won by parties A and B respectively. The size of the circles is meant to give a visual cue that the two seats with background yellow are going to contribute a larger number of votes to the national share (because turnout is 100%) than the smaller grey shaded circle.

The narrative then asks the viewer to imagine that the remaining 97 seats declare their results (Figure 6.5). The animation then shows separate red and blue seats move across the screen and we are left with a review of the national picture (Figure 6.6).

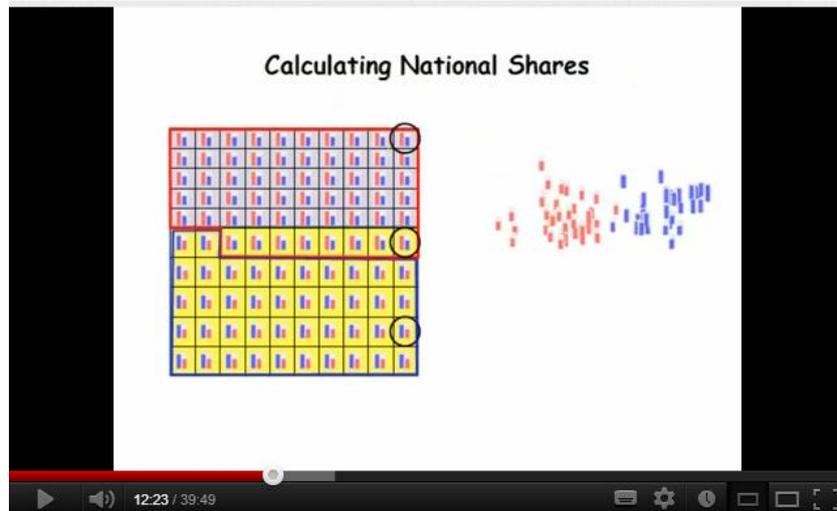


Figure 6.5: Declaring the national result

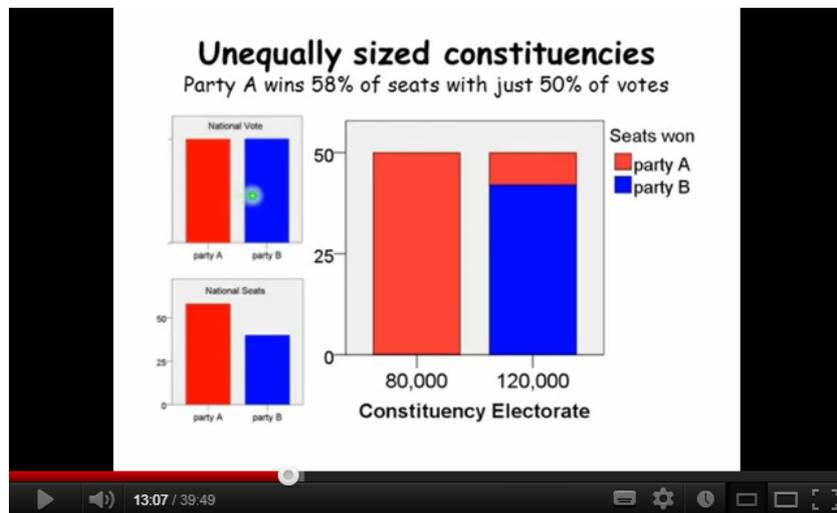


Figure 6.6: A summary of the election outcome

Figure 6.6 is a very important graphic because it presents a summary of the election outcome. The small histogram on the left shows the position in terms of seats with party A winning 58 (or 58%) and party B winning just 42 (42%) despite each party winning 50% share of the national vote. The explanation for this outcome appears on the right hand side histogram. This shows that in the constituencies where the electorate was just 80,000 the red party won all 50 of those seats but in the larger constituencies with electorates of 120,000 both the parties won seats there. The blue party won 42 of the 50 with the red party winning eight of

them. Simple arithmetic shows that the national vote for the red party would be $50 \times 80,000 + 8 \times 120,000$ giving a total of 496,000 votes. The national vote for the blue party is $42 \times 120,000$ which gives 504,000 votes. When percentages are rounded this gives 50% to each and yet the red party, because it won in the smaller seats, managed to win a greater share of seats than votes – the system is biased towards the red party and the bias is caused by malapportionment.

Bias caused by differences in turnout

As stated earlier there are three factors that lead to electoral bias, the first of which, electoral inequalities, was discussed above. When the presentation was being planned it was felt that each of the three segments should follow a similar structure to one another. In this way the viewer would be better able to follow the logic of the second and third sequences. In the second section, therefore, the slides have a similar appearance and format to those included in the opening section.

The sequence on electoral turnout, therefore, presents an image similar to that described in Figure 6.3 except that now electorates are all equal, the proportion of vote share won by the winning party is the same everywhere but turnout varies from 53% in half the constituencies and 80% in the other seats (Figure 6.7). This difference is visualised by having equally sized rectangles in the 10 by 10 grid but using grey and yellow shading to describe the differences in the numbers of people turning out to vote.

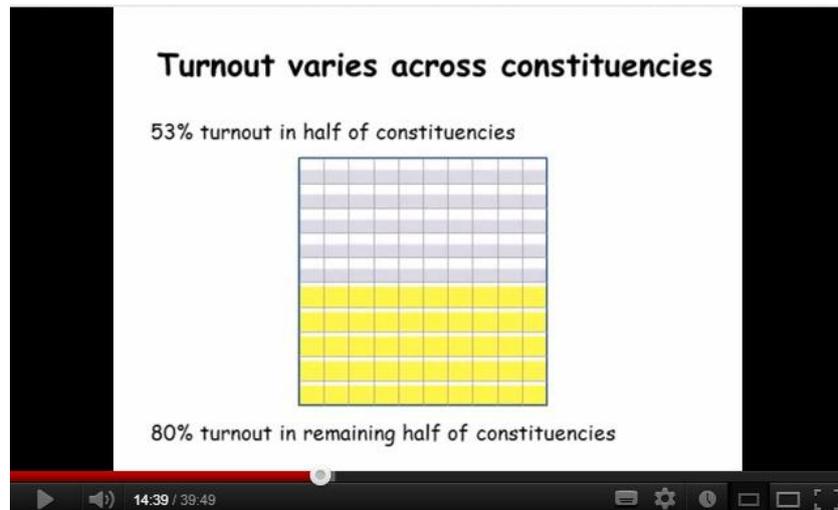


Figure 6.7: Variations in electoral turnout

In the same way as previously the election is then ‘run’ and results in the outcome shown in Figure 6.8. Again, the seats won by the red party A are bordered in red while those won by party B are surrounded by a blue border. Even the outcome is constructed to be the same – party A is winning 58% of the seats with party B winning the remaining 42% - despite winning an equivalent share of the national vote as party A.

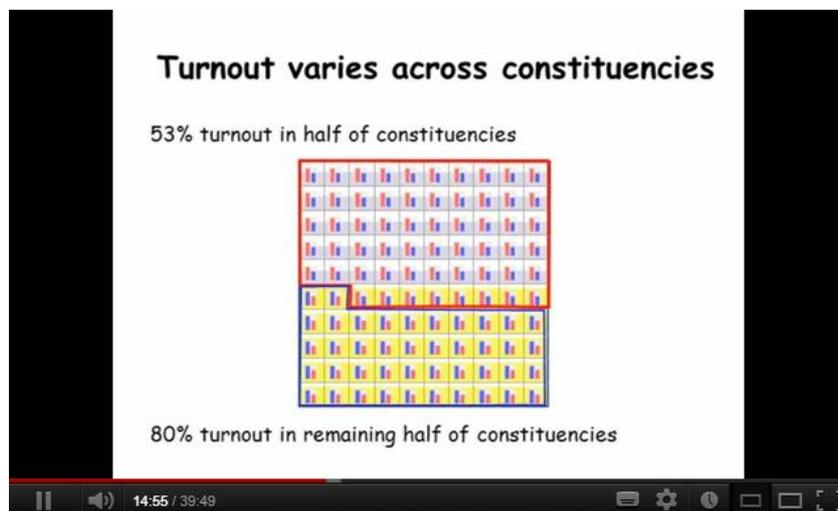


Figure 6.8: Differences in turnout and winning party

This sequence of slides follows the same format as earlier and finishes with a slide that provides the whole election (Figure 6.9). The two small histograms show that the two parties

have equal national votes share but unequal shares of seats. The largest histogram shows the frequencies of seats captured after dividing constituencies into low (53%) and high (80%) turnout areas. It is clear that the red party has concentrated most of its victories in low turnout areas (as does the Labour party) with only a small number of victories in high turnout areas. It is the blue party (and the Conservative party) that has won in the high turnout areas. Because of this there is electoral bias towards the red party that is contributed by the higher level of abstention in the seats that it won – it needs fewer votes to gain its victories than does the blue party.

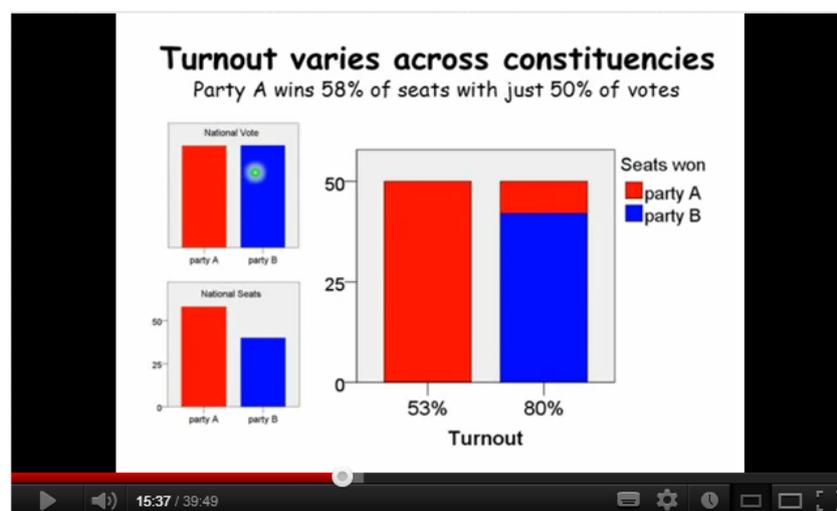


Figure 6.9: Electoral bias produced by differences in turnout

Although the original intention was to follow the same format for all three explanations of electoral bias it was felt that explaining the bias caused by vote distribution was more complicated than either electorate size or turnout. For this reason, Figure 6.10 was included as the explanation. During this part of the presentation there is a lot of narration that the listener has to follow but in order to minimise the text the lines read “surplus votes are...” and so on. There were other slides in the main powerpoint presentation that illustrated this better but it was felt that explaining these graphics would take too long for the YouTube version. This raises the problem of the trade-off between complexity and shortening the

explanation. Political scientists may not be the best people to make that judgement about what the general public can understand or what it cannot.



Figure 6.10: Problems of explaining the impact of vote distribution

Once this slide animation had finished the presentation could return to the style adopted for the previous explanations. Figure 6.11 follows the description of the assumptions that lie behind this element of the electoral bias – electorates are identical, turnout is assumed to be the same also but the size of the margin between the first and second placed parties is allowed to vary – the red party will win small but lose big (see Figure 6.10) but the blue party will have an inferior vote distribution – it will win big and lose small. For this part of the exercise in half the constituencies the vote shares will be 60/40 but in the other half it will be 64/36.

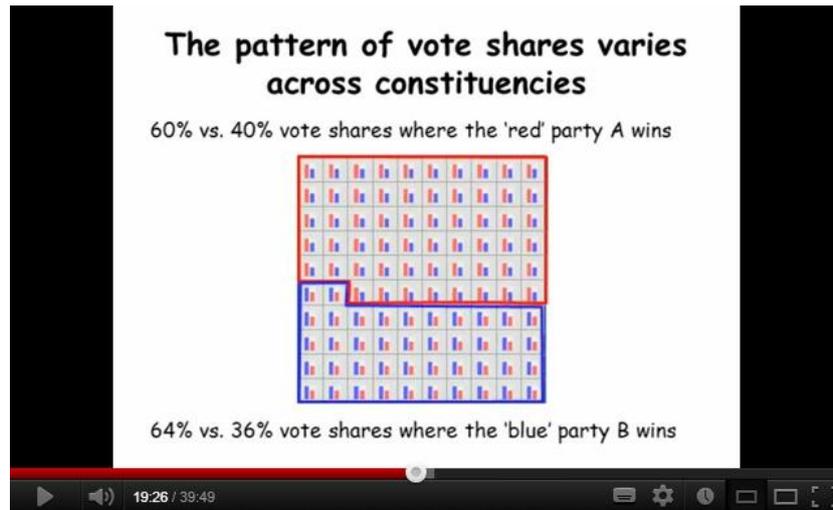


Figure 6.11: Variations in vote shares across constituencies

Almost half way through the presentation the expectation was that the viewer would be familiar with the graphical presentation – the red party winning in the top segment of seats with the blue party winning most of the remaining seats. Red and blue border lines denote the extent of red and blue party victories.

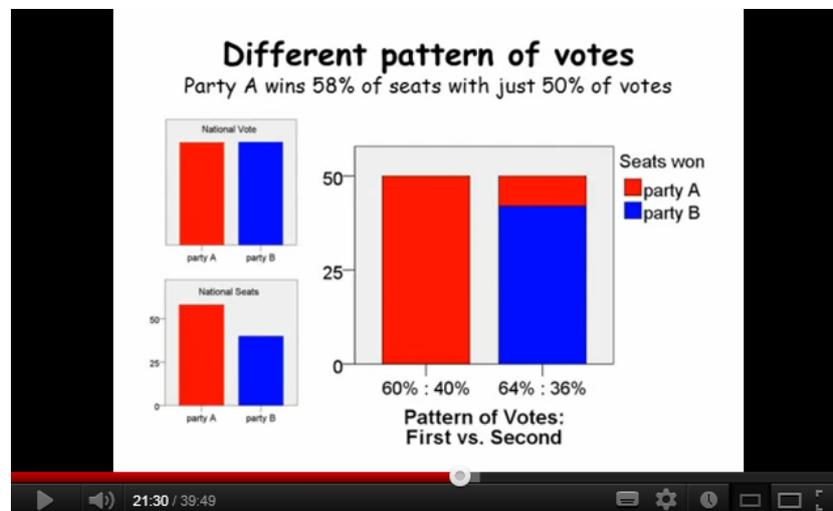


Figure 6.12: Different patterns of votes

With all three explanations the closing slide was virtually identical also. Figure 6.12 shows why, despite the fact that the two parties win equal vote share in terms of national support. The two histograms to the left show the distributions of votes and then seats while the largest

graph shows the red party winning the closer seats (majorities of 20%) and a handful of the other seats, most of which are won by the blue party with bigger majorities of 28 percentage points. Electoral bias favours the red party because it is following the advice of “win small and lose big” that produces a very efficient vote distribution compared to its competitor, the blue party.

With hindsight this would have been a good point to break the presentation. As the timer shows this slide is timed at 21 minutes and 30 seconds of a total time of 39 minutes and 49 seconds. This part one could have been viewed as a stand-alone presentation that explained electoral bias, how it occurs and why some parties are favoured by it while others are disadvantaged. The remaining part of the presentation is more concerned about the realities of UK electoral politics. So, a viewer could easily watch part one and not bother with the second part but of course the second part would not be understood if the first part had not been viewed at all. This raises problems in terms of long presentations and where breaks can be made that will not interrupt the flow too much. One solution that is used is to number sequences within the presentation, part one, part two etc. but then viewers might become discouraged if they see that the sequence contains many parts and will require a considerable time investment in order to reach the end of the presentation.

Applying electoral bias to the 2010 general election

Although the Conservative party has become famous for its complaints about the electoral bias that favours the Labour party it is in fact the Liberal Democrats that are most disadvantaged by the voting system. Figure 6.13 is intended to describe quickly that should be the case. The slide's title is intended to give the clue while the Table provides the evidence. When a party comes second in a first past the post system all it in fact is doing is adding votes but not converting those votes into seats. This may look good in terms of the

national vote share (24% at the 2010 general election for the Liberal Democrats) but it is not good in terms of gaining representation. The problem with this tabular presentation, however, is that it is not immediately obvious how the viewer should be interpreting the relationship between rows and columns. Of course the vocalised narrative is helpful but here it is also being assisted by the facility of lighting up part of the screen (near the figure 167 in the Liberal Democrat column). Reviewing this part of the presentation we later felt that a series of histograms might have proved more effective but this approach was originally rejected because it was felt that previous slides had used histograms extensively and that viewers would grow tired of seeing information always presented in the same manner – it is a good idea to keep graphical representation clean and simple but they should also be diverse where possible.

Coming second in a first-past-the-post voting system - not a good idea!

		2nd		
		Conservative	Labour	Lib Dem
1st	Conservative		137	167
	Labour	147		76
	Lib Dem	38	16	
	Total	185	153	243

Figure 6.13: Frequency of second placed parties at the 2010 general election

A well known method for disguising relationships between data is to manipulate vertical axes that describe quantities. Sometimes, however, this potential weakness can be turned into a strength. Figure 6.14 and Figure 6.15 show ineffective votes per seat won – for Conservative and Labour the data are contained and repeated in both graphics but in the second Figure the original scaling has been altered to allow the inclusion of the data for the Liberal Democrats.

In the original presentation this is a single slide containing animation. The image described in Figure 6.14 is the starting point for the slide and makes the case that in terms of ineffective votes there is not much difference between the Conservative and Labour parties. The slide then animates by simultaneously shrinking the red and blue columns and re-scaling the vertical axis which then allows us to bring in the much larger number of ineffective votes (surplus votes plus wasted votes) cast for the Liberal Democrats (Figure 6.15).



Figure 6.14: Ineffective votes cast for Conservative and Labour at the 2010 general election

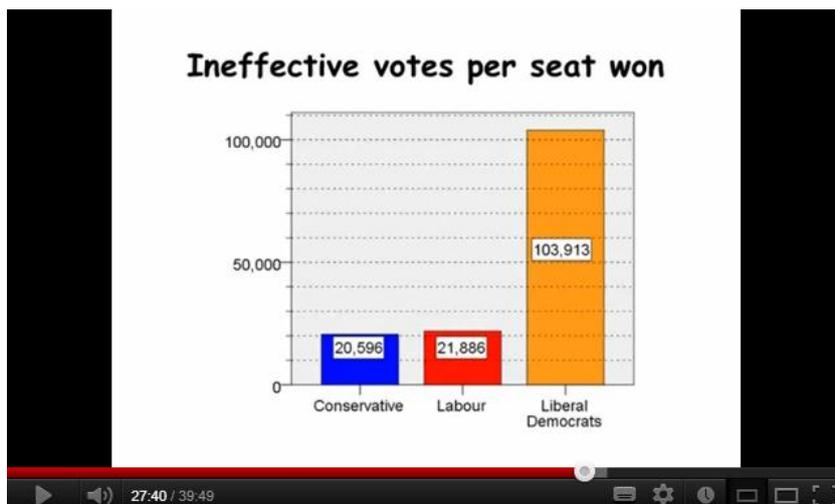


Figure 6.15: Comparing ineffective votes at the 2010 general election

The effect of re-scaling the vertical axis really does emphasise the huge differences between the two main parties and the third party. The numbers contained inside each histogram are not really required but it is a simple way of reinforcing the point.

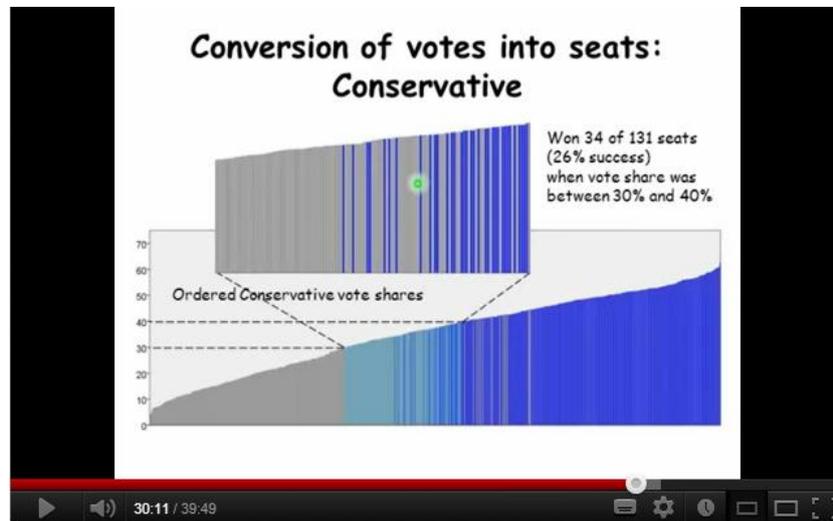


Figure 6.16: Converting votes into seats: the Conservative party in 2010

This efficiency of vote distribution is very important for the full explanation of electoral bias in the UK because it is currently the largest single cause of bias. It was decided, therefore, that it deserved a more extended treatment. We developed a method of visualising the ability of parties to convert votes into seats within the critical vote share range of 30-40%.

Figure 6.16 demonstrates this for the Conservative party at the 2010 general election. It works by rank ordering constituencies according to the Conservative vote share in every constituency – from smallest on the left hand of the horizontal scale to the highest on the opposite end. Apart from the title this is the only graph to appear at this stage with the narration describing the graph similar to the above description. Where the vertical line is coloured grey (mainly on the left hand side) that indicates a Conservative candidate stood but did not win. Where the vertical line is coloured in blue, however, that is an indication that the Conservative candidate won the seat. The presenter then begins to animate the

presentation, first by drawing horizontal lines that intercept the vertical axis at the 30% and 40% marks. This becomes the 'area of interest' and the vertical lines (constituencies) contained within this segment are highlighted. Next, two dotted lines emerge from this area and end with an exploded segment of the whole graph that features every result where the Conservative candidate obtained between 30-40% of the constituency vote. In this way the viewer can quickly observe that only a minority of these particular seats were won (denoted by a blue rather than grey line). Again, the narrator assists with this interpretation by using the Control key on the computer keyboard to highlight a particular section of the segment.

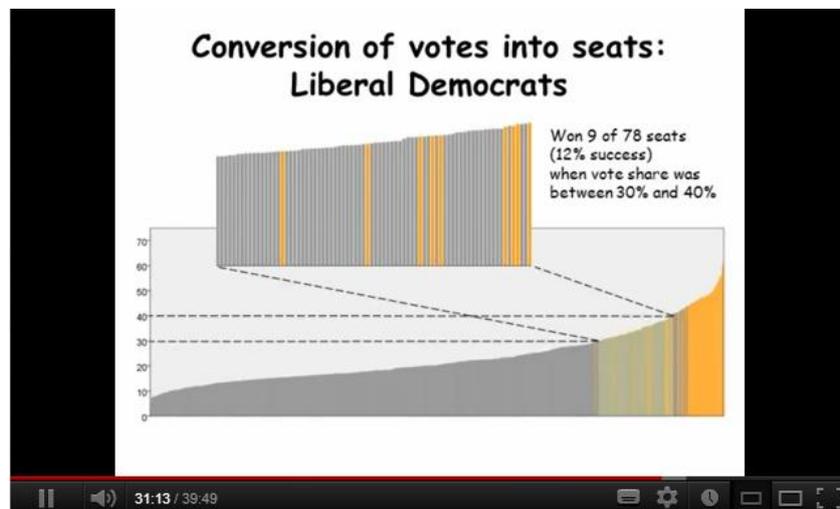


Figure 6.17: Converting votes into seats: the Liberal Democrat party in 2010

The point of showing the data in this format for both the Conservative and Labour (not shown here) is that it sets up Figure 6.17 which is the equivalent graph for the Liberal Democrats.

As noted earlier the Liberal Democrats took second place in a large number of parliamentary seats at the 2010 general election and that built up unhealthy votes to seats ration.

Figure 6.17 reinforces that point nicely. Once again the segment showing the seats where the party won between 30-40% of the vote is magnified and the viewer immediately notes the

very few yellow bars contained within that segment. The data on the actual number is then added for additional emphasis.

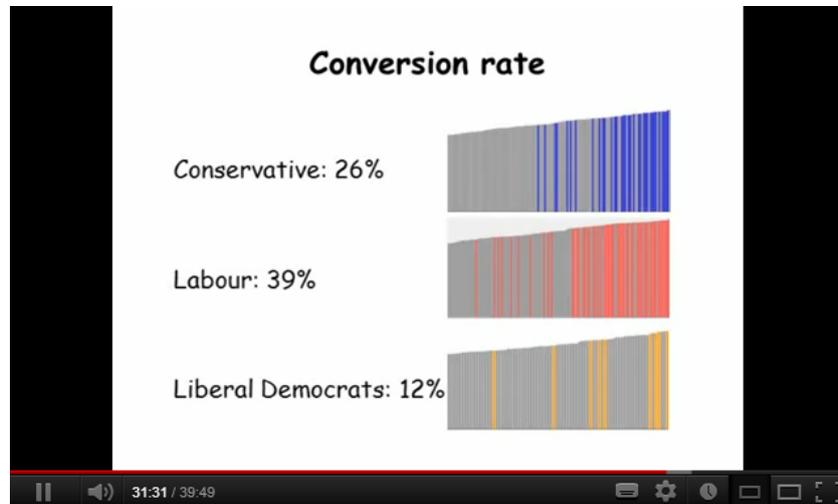


Figure 6.18: Comparing conversion of votes into seats, 2010 general election

Finally, Figure 6.18 summarises the previous slides by arranging all three segments together making comparison of the significant success rates achieved by the main parties in terms of converting votes into seats. This is a simple but highly effective graphic because it contains very few words, the viewer can absorb the differences easily and the images are easy to interpret.

	Labour	Conservative	Liberal Democrats
Total Bias	+63	+13	-76
Electorate	6	-7	1
Abstention	13	-11	-6
Geography	31	36	-74
Minor party	2	-2	-1
Net interaction	11	-3	4

Figure 6.19: Summary of bias position after 2010 general election

After about 33 minutes the video switches to a detailed examination of the distribution of bias following the 2010 general election and also digresses into a discussion about the post 2005 constituency boundary changes. With hindsight we think that this part of the presentation is too long, too detailed and does not convey information as well as the previous sections.

Some of the reasoning behind this view is visible in Figure 6.19. Although most political scientists would have little trouble in reading this Table for the general viewer there is a lot of information to absorb and then to appreciate the importance of the differences between the different numbers. For example, a seasoned reader of tabulated data would be able to read the top line and then realise that the line containing information about geography (vote distribution discussed earlier) accounts for most of the bias. Someone unused to reading data in this way would struggle to make this connection. The voice narration points this out of course but graphics would have worked far better.

A crucial part of the story about reviewing constituency boundaries after the 2010 general election is that the coalition government planned legislation that would speed up the process and remove some of the opportunities for the public to be consulted on the proposals. It also wanted to have a much stricter application of the electoral quota to be enforced – this reflected the Conservative thinking that inequalities in electorate size was the major cause of electoral bias and that it had not been removed by the previous boundary review. Figure 6.20 is intended to demonstrate that this is a false picture.

2005 actual and estimated

	Labour		Conservative		Liberal Democrats	
Total bias	+83	+75	-30	-21	-52	-52
Geography	41	41	-5	2	-46	-49
Electorate	11	4	-12	-6	-3	2
Abstention	16	17	-14	-14	-10	-9
Minor Party	3	3	-3	-3	-1	0
Net Interaction	13	11	3	0	7	4

**Figure 6.20: Boundary Commission impact
prior to 2005 general election**

The main point of this slide is the line marked “Electorate”; in this screen shot the small green highlighter can be seen and the voice narration is making the point. But as an exercise in data visualisation and an information facilitator it does not work. The point being made is the comparison between the distribution of bias both before and after the post 2005 boundary review and the fact that the malapportionment element (Electorate) has altered (a positive bias towards Labour reduces from 11 to 4 seats; a negative bias of 12 seats reduces to minus 6 seats for the Conservatives). In other words the Boundary Commissioners successfully achieved what they set out to do. The other part of the argument is that although they succeeded in reducing that part of the bias it did not have a major effect on the overall picture. The animation shown helps the explanation but there is no disguising the fact that for the general public there are too many numbers on this slide.

Removing 2010 electorate bias

	Labour	Conservative	Liberal Democrats
Total Bias	+57 +57	+20 +20	-77 -77
Geography	31	36	-74
Electorate	0 0	0 0	0 0
Abstention	13	-11	-6
Minor party	2	-2	-1
Net interaction	11	-3	4

Figure 6.21: The effects of boundary changes on future electoral bias

Figure 6.21 repeats the previous slide but up-dates the information for the likely impact of the post 2010 general election boundary changes. It is making the case that the level of electorate bias is small and even if it is removed there remain a net positive bias towards Labour of 57 seats, a smaller positive bias of 20 seats for the Conservatives and that the Liberal Democrats are massively disadvantaged.

- ### Summary
- Labour continues to benefit from +ve bias but the Conservatives now also benefit
 - Much of the change in bias stems from geography which now also favours Conservatives
 - Labour continues to benefit from abstention
 - Electorate bias is generally small but favours Labour
 - Next boundary review will NOT radically alter overall bias even if electorate size bias is removed and Commons is reduced to 600 seats

Figure 6.22: Summarising the presentation on electoral bias

Figure 6.22 is the final slide in the electoral bias presentation. It is clearly an attempt to summarise the argument not in terms of the overall presentation but simply the part that deals

with the 2010 general election and the next boundary review. It probably works as a piece of text although there are rather a lot of words and could have been edited to be shorter than it is. But it is not visual – it might have been better to have told the story with animated graphs based on the contents of Figure 6.21.

6.2.4 Examining the survey evidence: electoral bias

An essential part of this experiment was to obtain feedback from viewers about whether they had enjoyed the experience, learnt anything useful and how easily they had understood the content.

Although the number of views reached 1,500 by November 2012 the numbers participating in the online survey have proved rather disappointing. Only 86 people participated in our survey between July 7 and November 8, 2012 with only 70 of these (81.4%) completing the survey. A total of 70 usable surveys from 1,500 views represents a 5% response rate!

Superficially, this looks unimpressive but it should be stated that the Elections Centre now conducts its annual survey of local election candidates online and in both 2011 and 2012 the response rate has struggled to reach 15% of those sent letters inviting them to participate. It is well known that online response rates are rather poor and if the number of views had been considerably greater then a greater number of survey respondents would have resulted. It should also be mentioned that we attempted to prevent users from undertaking the survey multiple times by rejecting any data that revealed repeat ip addresses in the dataset. Although it is difficult to imagine why anyone would actually want to undertake a survey more than once!

The data show that among respondents the youngest is 15years old while the oldest respondent was 82years (Figure 6.23). The mean age is 39years (median age = 34)

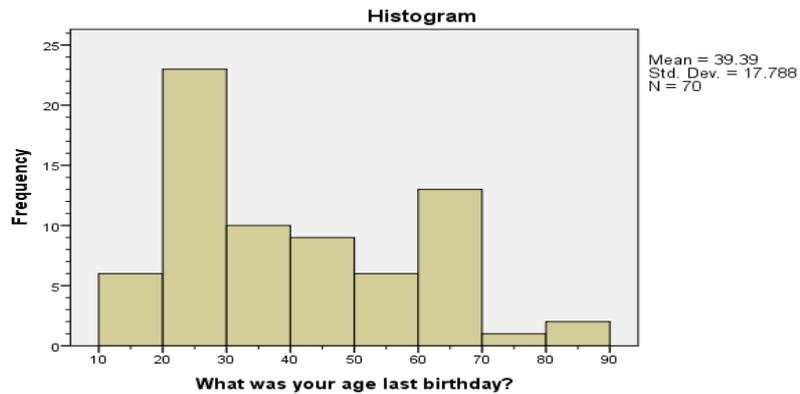


Figure 6.23: Age profile of electoral bias survey respondents

A total of 54 respondents gave information on about which party they voted for at the 2010 general election: 24% Conservative, 37% Labour, 26% Liberal Democrat, 13% Other).

Unsurprisingly, given that they have just watched a video presentation of about 40 minutes most of the respondents said that they are interested in elections: on a scale of 0-100 (most interested) the median value is 94.

It was interesting to discover that 44% (35 of 80) said that they had none or little knowledge about electoral bias before watching the presentation. It is encouraging to note that after watching the presentation only 8% (4 of 77) still had no or little; 25% (19 of 77) reported the same level of knowledge; 75% had improved their knowledge (Figure 6.24).

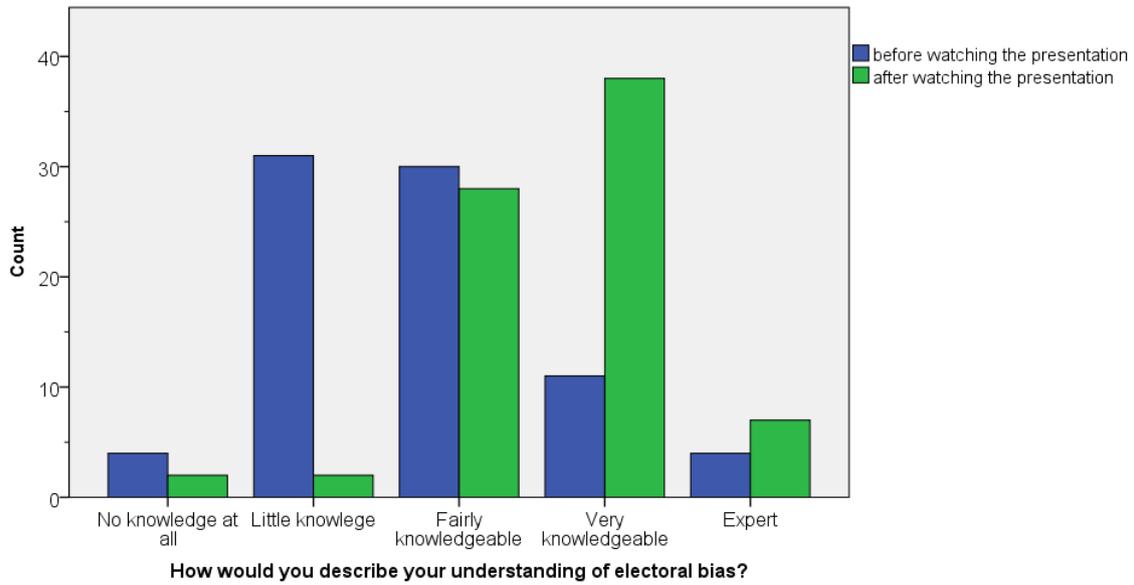


Figure 6.24: Knowledge of electoral bias before and after viewing presentation

For each of the three categories of bias there was a question that presented the respondent with two examples of slides in this particular sequence and a question attached. The question format was: “How satisfied are you that this series of slides informs you about [bias associated with Electorate size/Turnout/Vote distribution]?”

Table 1: Satisfaction with slides relating to bias caused by electorate size:

How satisfied are you ... slides about ...-Electorate size

		Frequency	%	Valid %	Cumulative %
Valid	Very satisfied	29	33.7	38.7	38.7
	Satisfied	42	48.8	56.0	94.7
	Neutral	3	3.5	4.0	98.7
	Dissatisfied	0	0.0	0.0	98.7
	Very Dissatisfied	1	1.2	1.3	100.0
	Total	75	87.2	100.0	
Missing	System	11	12.8		
Total		86	100.0		

The results for the slides on electorate size bias are very encouraging and show that 95% of our respondents were either satisfied or very satisfied with the level of explanation of bias provided by these slides (Table 1).

There is only a slightly lower level of satisfaction with the slides on turnout (Table 2) but that level does not carry over into viewers' assessment about the slides that were concerned with geography (Table 3).

Table 2: Satisfaction with slides relating to bias caused by turnout differences
How satisfied are you ... slides about ... Turnout

		Frequency	%	Valid %	Cumulative %
Valid	Very satisfied	27	31.4	37.0	37.0
	Satisfied	38	44.2	52.1	89.0
	Neutral	6	7.0	8.2	97.3
	Dissatisfied	1	1.2	1.4	98.6
	Very Dissatisfied	1	1.2	1.4	100.0
	Total	73	84.9	100.0	
Missing	System	13	15.1		
Total		86	100.0		

Table 3: Satisfaction with slides relating to bias caused by vote distribution
How satisfied are you ... slides about ... Geography

		Frequency	%	Valid %	Cumulative %
Valid	Very satisfied	28	32.6	43.1	100.0
	Satisfied	34	39.5	52.3	56.9
	Neutral	1	1.2	1.5	4.6
	Dissatisfied	1	1.2	1.5	3.1
	Very Dissatisfied	1	1.2	1.5	1.5
	Total	65	75.6	100.0	
Missing	System	21	24.4		
Total		86	100.0		

We had anticipated that since this element was the most complex bias component to explain it would require more imagination and graphics to get the point across. That was the reason why there was an additional slide – the explainer – that tried to summarise the argument in a short phrase – “win small, lose big”.

In order to test respondents’ appreciation of electoral bias and their ‘reading’ of the tables we inserted some quiz-type questions.

Table 4 shows the result from one of these questions which asked them to identify the party which was most disadvantaged by electoral bias. A comforting 81% gave the correct answer of Liberal Democrat but there were 12 respondents that felt the Conservatives alone were the most disadvantaged.

Table 4: Summarising the bias position after the 2010 general election:

Which party is the most disadvantaged by electoral bias after 2010 general election?

		Frequency	%	Valid %	Cumulative %
Valid	Con	12	14.0	17.1	17.1
	LD	57	66.3	81.4	98.6
	Con & LD	1	1.2	1.4	100.0
	Total	70	81.4	100.0	
Missing	System	16	18.6		
Total		86	100.0		

There was no such disagreement about the correct answer for the party that benefitted the most from the abstention or turnout bias – all but two people that answered this question provided the correct answer (Table 5).

Table 5: Summarising abstention bias after the 2010 general election

Which party benefits from abstention/turnout bias?

		Frequency	%	Valid %	Cumulative %
Valid	Con	1	1.2	1.4	1.4
	Lab	68	79.1	97.1	98.6
	LD	1	1.2	1.4	100.0
	Total	70	81.4	100.0	
Missing	System	16	18.6		
Total		86	100.0		

Table 6 and Table 7 show the responses to similar questions asked in Table 5 but this time relating to electorate size and vote distribution bias respectively. The question on electorate size bias (Table 6) was challenging with the answer being that both Labour and the Liberal Democrats were advantaged although the latter's advantage was only a single seat. Despite this some 70% got the answer right

Table 6: Summarising electorate size bias after the 2010 general election

Which two parties benefit from a positive bias relating to variation in electorate size?

		Frequency	%	Valid %	Cumulative %
Valid	Con	2	2.3	2.9	2.9
	Lab	2	2.3	2.9	5.7
	Con & Lab	17	19.8	24.3	30.0
	Lab & LD	49	57.0	70.0	100.0
	Total	70	81.4	100.0	
Missing	System	16	18.6		
Total		86	100.0		

Table 7 shows that seven in ten of the respondents answered correctly that the Conservatives benefitted the most from vote distribution bias after the last election. A quarter, however, did

answer that it was Labour – in fact Labour had been the main beneficiary of geography bias before the 2010 election but the Conservatives overtook them on this measure at the most recent election.

Table 7: Summarising vote distribution bias after the 2010 general election
Which party benefits most from the bias relating to vote distribution?

		Frequency	%	Valid %	Cumulative %
Valid	Con	49	57.0	71.0	71.0
	Lab	18	20.9	26.1	97.1
	LD	1	1.2	1.4	98.6
	Lab & LD	1	1.2	1.4	100.0
	Total	69	80.2	100.0	
Missing	System	17	19.8		
Total		86	100.0		

Overall, relating to the quiz-type questions between 70% and 97% of respondents answered correctly while 50% (35 of 70) answered all four questions correctly (Table 8). There were four people, however, who got three of four questions wrong.

Table 8: Number of correct responses to quiz

		Frequency	%	Valid %	Cumulative %
Valid	1	4	4.7	5.7	5.7
	2	14	16.3	20.0	25.7
	3	17	19.8	24.3	50.0
	4	35	40.7	50.0	100.0
	Total	70	81.4	100.0	
Missing	System	16	18.6		
Total		86	100.0		

We are reluctant to do much cross-tabulation of these data because the number of respondents is too small but we were interested to know whether there were any noticeable differences between the people who performed well on the quiz questions and those that performed less well (Table 9). Age does not appear to be a factor – the mean age of those answering all questions correctly was 38 years but those identifying two answers was 45 years but any linear relationship is upset by the fact that for three questions answered correctly the mean age falls to the youngest age. There is no difference either when it comes to the respondent’s overall interest in elections, although as we noted earlier there is a very high mean value for this variable with very little variance in the measurement.

Table 9: Quiz answers controlling for age and interest in elections

No. correct answers	What was your age last birthday?			On a scale running from zero to 100 how interested are you in elections?			
	Median	Mean	Std Error of Mean	Median	Mean	Std Error of Mean	Valid N
1	47.5	44.8	10.7	35.0	42.5	25.3	4
2	37.5	44.6	5.8	95.5	92.6	2.5	14
3	27.0	36.2	4.8	95.0	92.3	2.5	17
4	35.0	38.2	2.5	92.0	92.3	1.3	35

6.3 Supplementary Vote

6.3.1 Background

After posting the Electoral Bias video on You Tube and iTunes U we decided that the next opportunity to continue the experiment with visualising data for the general public would be the Police and Crime Commissioner (PCC) elections. These new officials would be elected by the Supplementary Vote (SV) which was first introduced for the London mayoral election in 2000, then used for other mayoral elections. There are explanations of how it works and its perceived weaknesses in the academic research literature (see Rallings and Thrasher 1999; Rallings et al. 2002) but before 2012 it was not really seen as necessary to explain it to the wider public. Because the PCC elections were national (except for Scotland, Northern Ireland and London) the Electoral Commission became responsible for supervising the conduct of these elections and it produced its own guide (<http://www.youtube.com/user/ElectoralCommission1>) which consisted of a 43 second cartoon animation. Although this provided an explanation of how a voter should complete a ballot paper it gave no assistance in terms of how the voting system came up with the winner. The Elections Centre believed that it could do better.

Based on the experience with the Bias video the aim was to produce a much shorter guide to SV – the guide timing was between 10-12 minutes. The style would be similar – a short introduction explaining the purpose of the video, a PowerPoint presentation of the main principles behind SV and then a final observation on film together with a request that viewers complete a short questionnaire (similar to the questionnaire attached to the bias video).

6.3.2 *The presentation*

The first slide (Figure 6.25) offers the viewer a one minute summary of SV; it was hoped that viewers would watch the whole presentation but experience tells us that some people get bored very quickly. This part of the presentation, lasting two minutes, would at least tell the viewer how voting in an SV election takes place.

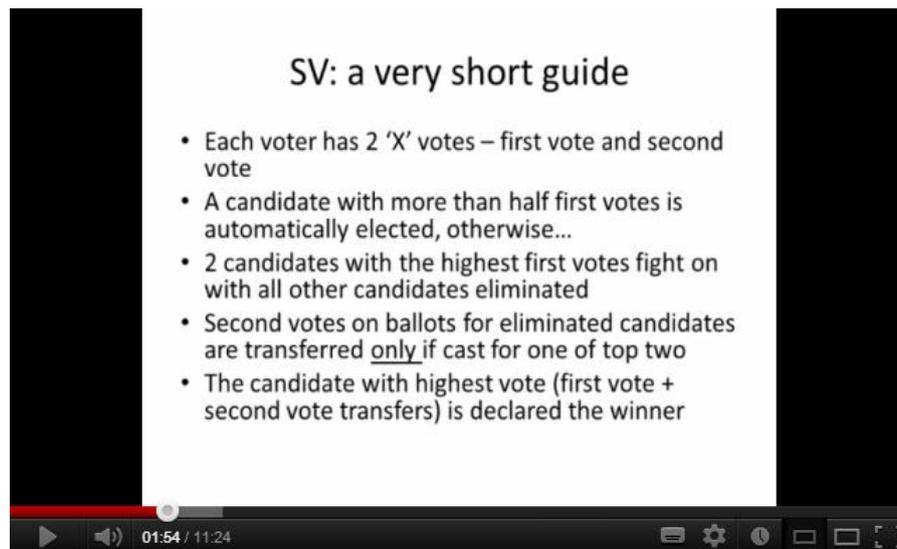


Figure 6.25: Supplementary Vote explained in one minute³⁴

Figure 6.26 presents a mock ballot paper. This lists four candidates with surname initials A, B, C, and D which correspond to the candidate's party names also. We also used four surnames that 'represent' England, Wales, Scotland and Ireland but we are not sure how many people noticed this attention to detail. We could not obtain a copy of an actual ballot paper before building the presentation but later we were able to compare our version with the official ballot paper and the two were quite similar.

³⁴ http://www.youtube.com/watch?v=6_givU4tv7E

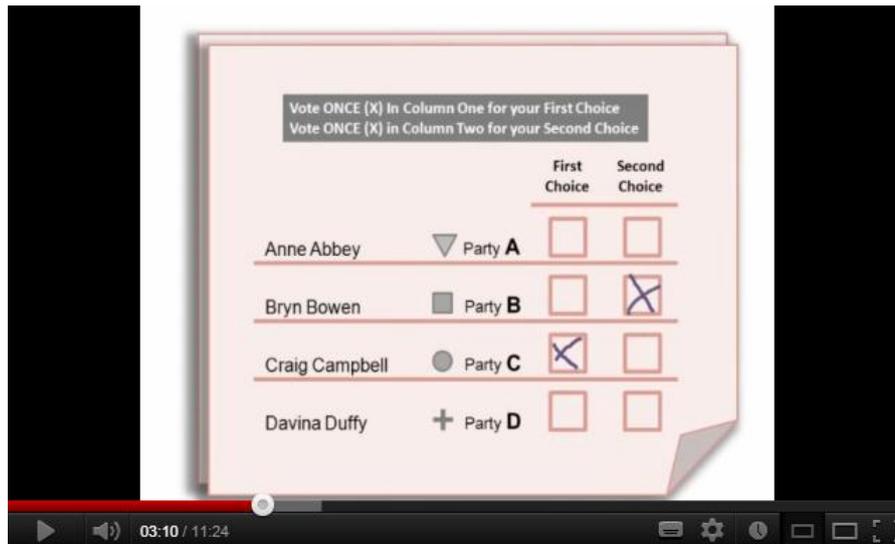


Figure 6.26: Completing a Supplementary Vote ballot paper

The screen shot shows a ballot where a voter has cast a first vote for Campbell and the second vote has gone to Bowen. This sequence includes a set of ballot papers, recording votes for different candidates. It includes also examples of ballots where a person has only used the first vote or cast two votes for the same candidate. This is perceived as one of the weak features of SV and later examples return to show what happens to second votes that are cast in this way.

At a point that is almost half- way through the presentation the viewer is ready to learn about what happens when the second votes are taken into account. Figure 6.27 reveals that Abbey leads after the first count with 35 first votes. Duffy narrowly wins the second place, pushing Bowen into third place by just one vote. The example is deliberately constructed that the sum of votes (35+24+16+25) equals 100 and therefore are also percentages. This is because in SV if a candidate win more than 50% of first votes then he or she is automatically elected.

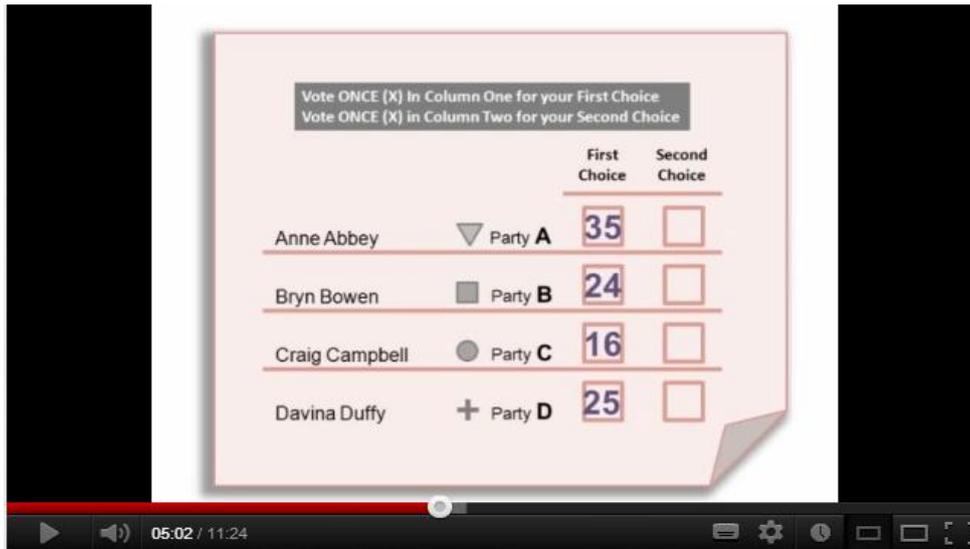


Figure 6.27: First vote counts in a Supplementary Vote election

The rules of SV state that in the event of no candidate winning more than half the first votes then the top two candidates remain in competition but all the other candidates are eliminated. Figure 6.28 then highlights Bowen and Campbell to show that these are the two that will be eliminated when only the top two candidates are allowed to go into the second round.

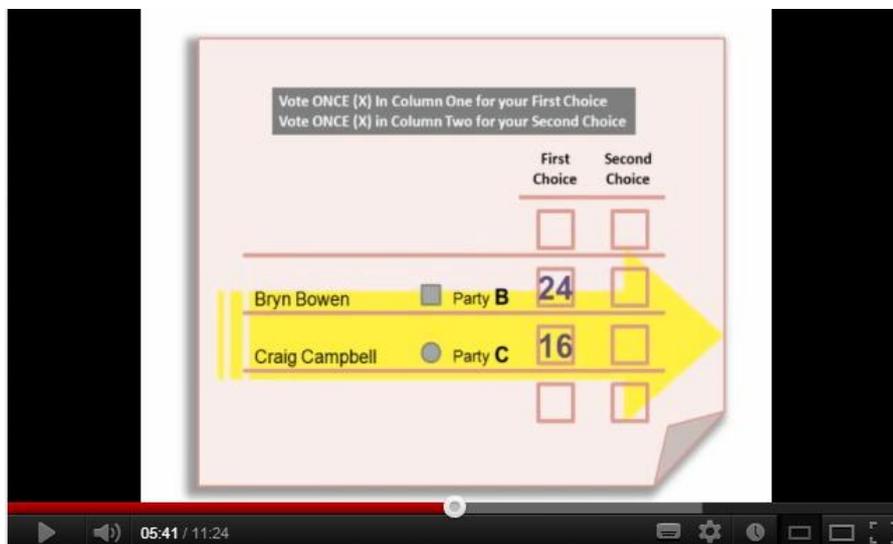


Figure 6.28: Process of eliminating all except the top two candidates after first count

The next Figure 6.29 copies the image that we used to demonstrate the different examples of completed ballot papers and then takes the viewer through the list. Because all of these

ballots contain ‘first’ votes for Campbell every first choice is cast for “C”. But there are different kinds of configurations shown to try and capture how varied are the actual ballots. A number of people have managed to cast a second vote that can be transferred (i.e. votes that are C/A or C/D). Others have not cast a second vote, some people have used both votes for Campbell, and a third group of people have cast a vote for (B)owen. None of these second votes will count and this is emphasised also in the voice over.

First Choice Party C: 16 Ballots
How were second votes cast?

	First Choice	Second Choice
1	C	C
2	C	A
3	C	
4	C	D
5	C	A
6	C	
7	C	D
8	C	B

	First Choice	Second Choice
9	C	B
10	C	C
11	C	A
12	C	D
13	C	D
14	C	B
15	C	
16	C	D

Figure 6.29: Allocating second vote preferences from ballots cast for bottom-placed candidate

Figure 6.30 below gives the visualisation of second votes and what happens at the second count stage. Sets of ballots are discussed; 1, 2, 3 are eliminated first because there is no second vote, then 4 and 5 because first and second votes are cast for Campbell, then 6, 7, and 8 because they contain second votes for candidate B who has also been eliminated at the same time as Campbell. Those votes that can be transferred are highlighted in sequence – yellow shading at votes for candidate A and darker shading for candidate D. An additional 3 votes have gone to A and 5 votes have gone to D.

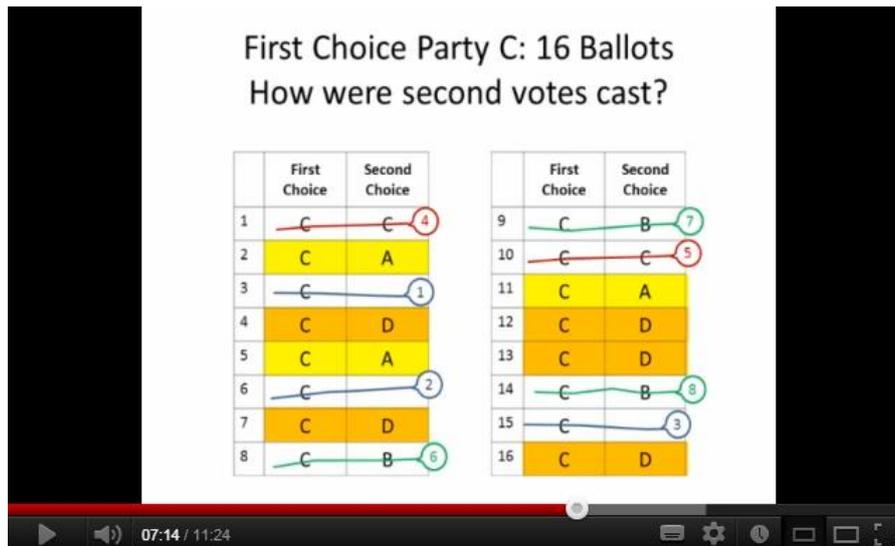


Figure 6.30: Successful and Unsuccessful vote transfers

The idea behind Figure 6.31 is to show that all of the second votes for eliminated candidate C have been taken into account. A number have been discarded but the dotted line indicates that some have been successfully transferred to either candidate A or D.



Figure 6.31: Recalculation of votes for top two candidates

By using four candidates for our example we are able to take the viewer through exactly the same process when considering what happens to the second votes cast in favour of the other

eliminated candidate, Bryn Bowen. Figure 6.32 reminds the viewer that two candidates remained and that candidate B will be treated in the same way as candidate C. Because more votes were cast for B and to save time the layout of second votes (Figure 6.33) is more condensed and the elimination/transfer sequence is done in reverse – votes to be transferred are highlighted first, totals to be transferred (five votes for D and 9 votes for A) are provided and finally, all remaining 10 votes are struck out because they cannot be transferred. This sequence of examining the second votes for the second eliminated candidate takes only one minute of the presentation but it is perhaps the most crucial part of the exercise. The viewer is now clear that the process of examining second ballots happens for all eliminated candidates no matter how many votes are being transferred.

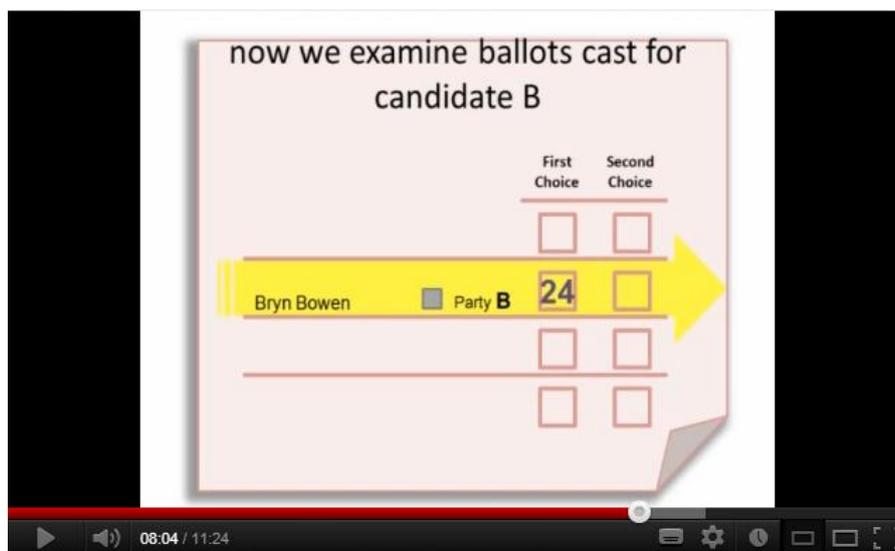


Figure 6.32: Review of second votes on ballots cast for third-placed candidate

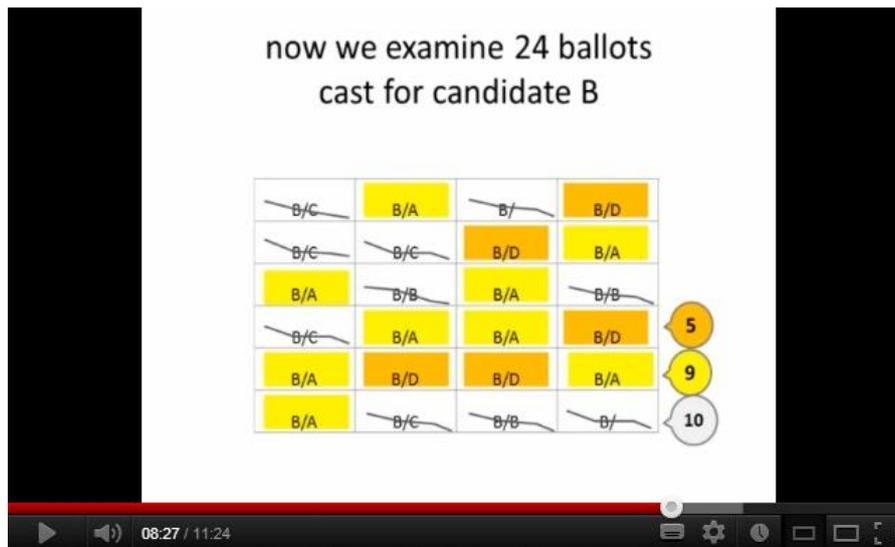


Figure 6.33: Successful and unsuccessful vote transfers from third-placed candidate

In the same way that the column of second votes cast for candidate C was emptied and votes transferred to either A or D then the same thing happens to the second votes cast for candidate B (see Figure 6.34). As the votes are transferred the blue and red bars rise – dotted lines across the bars show the high-water marks achieved when votes from C were transferred and then subsequently from candidate B. The same data is actually revealed three times on this slide – the bars rise, the numbers appear at the bottom and then finally the words above the winning candidate A appear. This example is designed also to show that because the total for candidate A is 47 she therefore falls short of the 50% mark. This was done because many people mistakenly believe that in order to win an SV election the winner requires an absolute majority of votes.

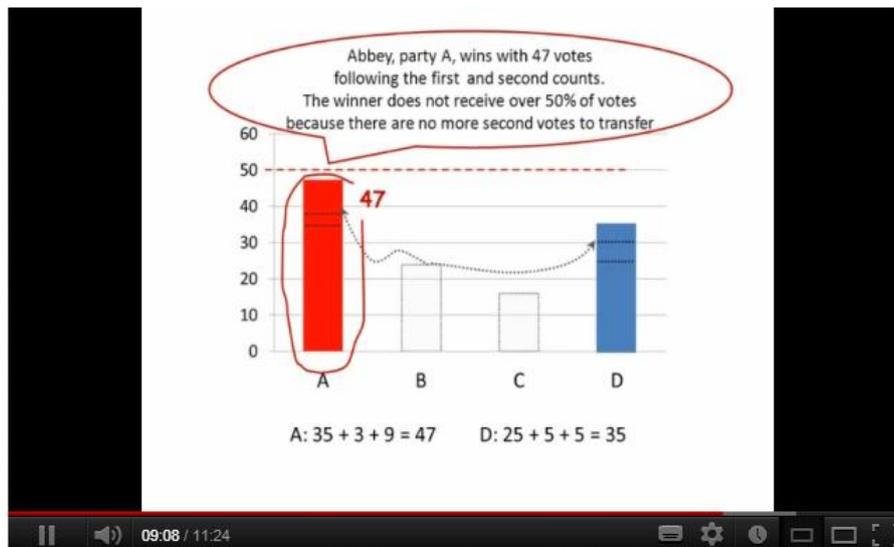


Figure 6.34: Declaring the winner in the Supplementary Vote election

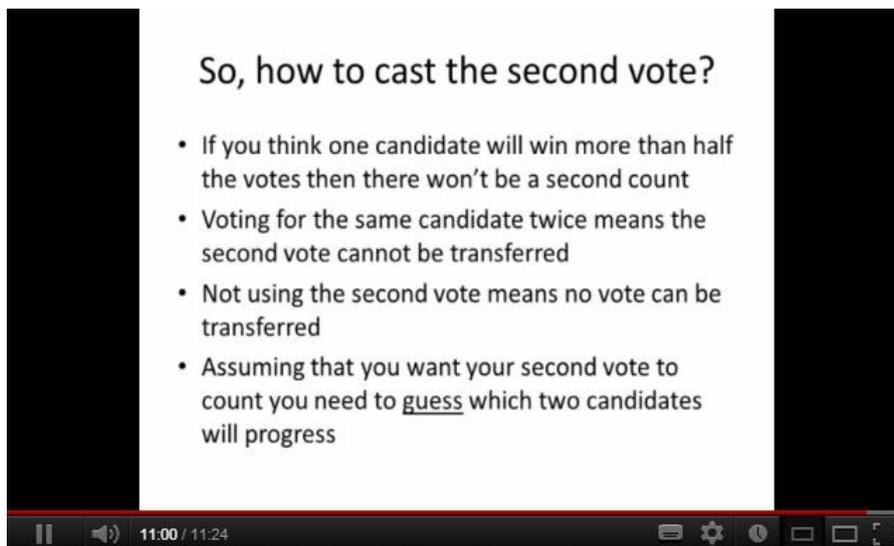


Figure 6.35: Summarising the second vote strategy

The final Figure 6.35 tries to repeat what the first slide does – provide a very quick review of SV. One of the main reasons for preparing this video was that SV is a peculiar voting system which requires voters to guess which two candidates are likely to remain in the race if no one candidate is able to win more than half the first votes cast. That is the final sentence on this slide. It was also the case that in our worked example that following the first count only a single vote separated candidate B and candidate D but this meant that Bowen rather than

Duffy was eliminated. A number of people who voted for candidate C gave second votes to B because they may have guessed (incorrectly) that B would be left to fight out the contest with A. We wonder how many viewers will have noticed such subtlety.

6.4 Examining the survey evidence: supplementary vote

In the same way that we conducted a survey of viewers that had watched the presentation on electoral bias we also invited those who visited the supplementary vote visualisation to report any comments they might have. At this point we should admit that we received a rather disappointing response – only 48 usable replies were received, an approximate response rate of 10% of all viewers. Nevertheless, we feel that the data do reveal some findings of interest.

Respondents were asked for their age and as Figure 6.36 shows the presentation attracted a rather large proportion of younger viewers with the largest group in the 20-30 years age category. By contrast, there were few respondents aged over 40 but this almost certainly is some function of YouTube’s viewing demographic.

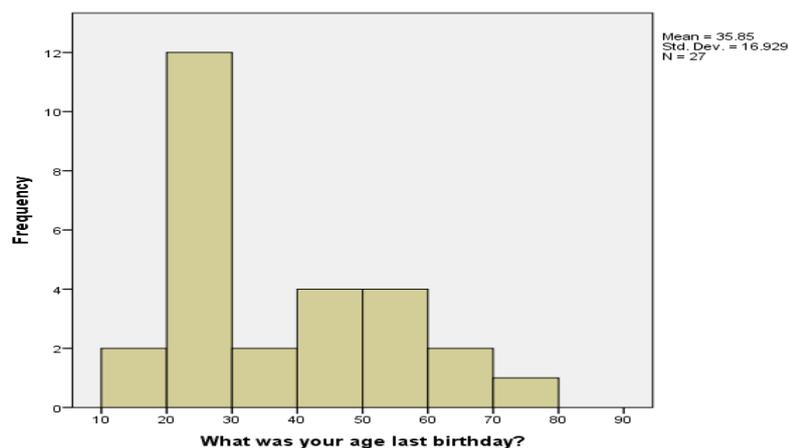


Figure 6.36: Age profile of respondents to SV survey

The survey asked people about their knowledge of Supplementary Vote both before and after watching the presentation. The categories ranged from having no knowledge at all to expert

viewer. Table 10 shows a cross tabulation of these two questions. Two thirds of those that had no prior knowledge of SV felt themselves to be ‘fairly knowledgeable’ after watching it while a further third claimed to be ‘very knowledgeable’. The effect of the presentation appeared to be linear in terms of its teaching capacity. Among those with a ‘little knowledge’ prior to viewing some 43% later thought that they were now ‘experts’. Reassuringly, no-one who completed the survey later thought that they had none or little knowledge of SV – at least we did not confuse anyone, therefore.

Table 10: Viewer assessments of knowledge both prior and after watching presentation

		Current understanding of the SV (after watching) (Row %)			
		No or Little knowledge	Fairly knowledgeable	Very knowledgeable	Expert
Understanding of SV before watching presentation	No knowledge at all	0	67	33	0
	Little knowledge	0	29	29	43
	Fairly knowledgeable	0	0	90	10
	Very knowledgeable	0	0	88	13
	Expert	0	0	0	100
Total		0	13	61	26

Conscious of the fact that the presentation on electoral bias lasted for almost forty minutes we took the precaution with the SV presentation of including a single slide (Figure 6.25) close to the beginning where the narration said that the explanation would take only a minute if the viewer was already feeling bored. We assume, however, that everyone who took the survey would have viewed the presentation to the end. Nevertheless, it is interesting to see reactions to this very short executive summary of SV. Figure 6.37 shows the set of responses to a question that asked about satisfaction with this quick snapshot of SV after controlling for a respondent’s prior knowledge of SV. It appears that those people that felt they had no knowledge at all about SV were not particularly satisfied with the ‘one slide’ approach – in

fact as many were very dissatisfied with it as were satisfied. This is certainly something that needs to be given full consideration when designing future presentations of this kind.

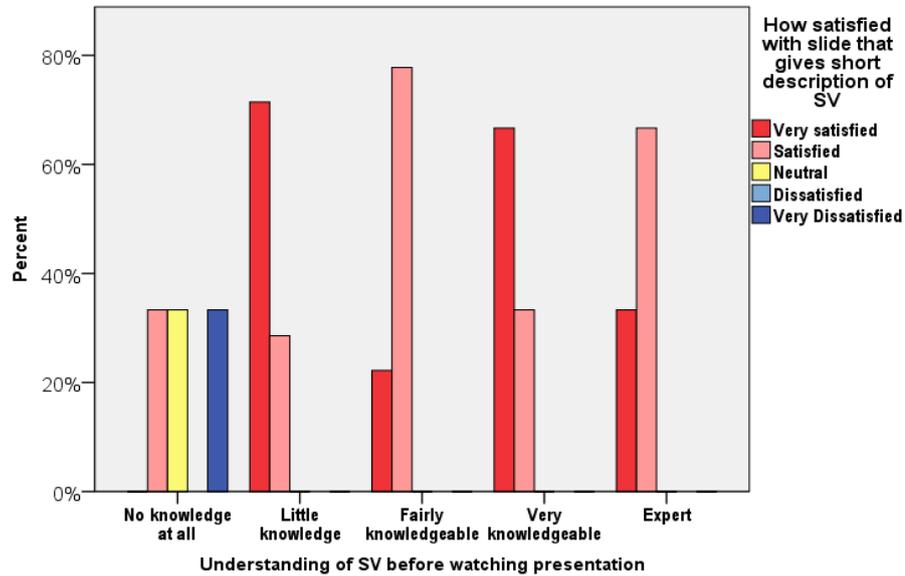


Figure 6.37: Levels of satisfaction with executive summary of SV

A considerable amount of discussion was taking place prior to the Police and Crime Commissioner elections about how the winning candidates would be determined under SV rules. Figure 6.34 showed a slide that explained how that would be done for our hypothetical example and was careful to demonstrate that it was not the case that each winning candidate had to obtain more than half the votes cast. The survey asked respondents about their satisfaction with this crucial slide and Figure 6.38 shows responses to that question by prior knowledge about SV. We can clearly see that no-one felt dissatisfaction at any level and only a small percentage of very knowledgeable respondents felt neutral about their level of satisfaction. Once again, these are encouraging findings and certainly convince us that visual presentations, even for complex subjects such as explaining the intricacies of SV, are an improvement over simple text.

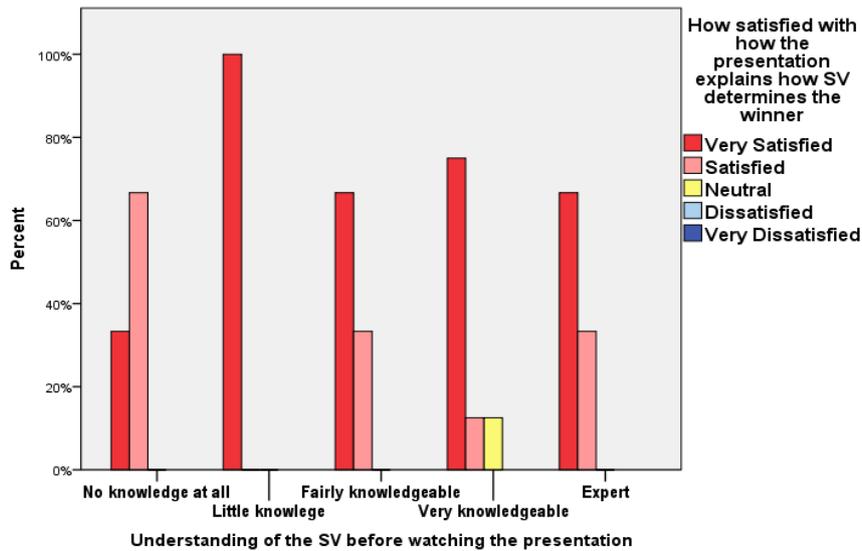


Figure 6.38: Satisfaction with explanations of how SV determines a winner

6.5 Conclusions

There is growing pressure on academics to share research findings with the general public; indeed there is pressure upon them to publish in journals that are online and free to view rather than maintaining the tradition of publishing papers in peer-reviewed journals owned and controlled by professional publishers. As part of this democratising process research scientists must adjust some of the criteria they use when publishing findings. Currently, very few journal articles receive a very wide readership and are almost intended as a form of communication between specialists. As more online journals are established and academics are encouraged to publish in them (in fact, compelled to publish in them if their research was funded from public funds) they will need to understand that the pool of readers may be much larger than before. However, online journals offer opportunities that text-based journals cannot, including embedding video, coloured charts and graphics that using programs such as Flash can animate, and attachments containing powerpoint or even the original data sets.

Accordingly, the aim of this chapter was to describe the process of compiling video presentations on some technical aspects of voting systems that could be viewed and understood by the general public. Two cases were examined in detail. The first study was a video of almost forty minutes duration that described electoral bias and three factors that contribute towards that bias. The second study was based on a video presentation of about 12 minutes that looked at the Supplementary Vote. Both videos are based on research being undertaken by research staff within the Elections Centre and both were made because it was felt that these were important topics that interested people that engaged with politics but might be puzzled by bias and/or the Supplementary Vote.

There was neither the time or expertise to make these presentations interactive such that users could manipulate any aspect of the talk – the viewer is passive. That places a considerable burden on the presentation makers – we can ask students to sit and watch/listen to a PowerPoint presentation but demanding that of the public is more difficult. The challenge, therefore, was to create something that informed without being boring.

Of course, we have only imperfect ways of measuring reaction. There are many more people that watch videos of cats and dogs on YouTube than visited our talks. We remain surprised that the much more technical and longer video on bias has received relatively more ‘hits’ than the one on SV (although only 15% voted in the PCC elections so perhaps it was always going to have less interest). But, what would be an acceptable number of views before we could say something was success or failure.

Apart from the problems of measuring reaction there is also the content to take into account although the two are going to be related. Videos on either YouTube or iTunes that feature a person looking at a camera and talking or worse, reading, are poor content. PowerPoint

slides that contain too much text and do not animate may be okay for specialist audiences but they will not keep most people's attention for very long. These videos avoided those kinds of mistakes and the animations do help the 'flow'.

Would this type of approach be helpful for political scientists that were interested in visualising their data? The short answer to that is yes but other factors are important. The software is not yet designed to a level where the non-expert can develop presentations of this level of content and animation. Few political scientists possess the skills to do this and those skills are not taught as part of doctoral programmes. Second, there is the problem of time. At forty minutes the bias video is extremely long. One solution might have been to divide it into separate parts but we risk losing part of the audience. Another solution would simply to make it shorter. With hindsight this presentation could have been shorter but not much shorter. The problem/concept is too complex to explain in ten minutes and so there has to be times when videos are longer and designed not the general public but for people with a particular interest in politics.

7 Accessing election data by mobile telephone

7.1 Introduction

Connecting potential users to election data poses some interesting problems. This chapter and the next report on some of the difficulties encountered when attempts are made to build an interactive system whereby inexperienced users can access complex databases. This chapter focuses upon an experiment that was conducted prior to the 2009 European parliament elections. The following chapter reports on the construction and development of a web-based search engine that was designed to answer people's queries in relation to information on the 2010 general election. Initial findings from these experiments were presented earlier (Long, Lovitskii et al. 2009; Long, Lovitskii et al. 2010; Long, Lovitskii et al. 2011).

Both chapters address issues about the user's search activity and how intelligent software is structured that can 'learn' from the syntax employed in searches for election data. Political scientists are familiar with such terms as 'percentage vote share', 'share change' and 'turnout' but the language which researchers use to describe these phenomena is not necessarily understood and shared by members of the public. What appears to be a

straightforward process of meeting requests for information expressed in terms of natural language – what was the result in the South West region – is not a simple process at all.

These two chapters build upon work that was conducted within the Elections Centre over a number of years. The collaboration between Vladimir Lovitskii and Michael Thrasher began with the attempt to develop intelligent software that could process requests to access election databases (Coles and Lovitskii 2000; Braithwaite, Lishman et al. 2007; Francis, Lishman et al. 2007; Johnston, Lovitskii et al. 2008; Lovitskii, Thrasher et al. 2008) using simple text-based applications on mobile telephones. Later approaches adopted a different approach using web-based applications that would process simple text commands inputted by the user to produce answers to those requests in the forms of text, maps and graphical displays (Long, Lovitskii et al. 2009; Long, Lovitskii et al. 2010; Long, Lovitskii et al. 2011)

The structure of this chapter is as follows. The opening section sets the context by providing a short description of the growth in the volume and sophistication of mobile phone technology. This is followed by a description of how the 2009 mobile phone text data mining access to European election data was created and implemented. Following the experiment we conducted a lengthy review of the lessons learnt from what was always intended to be a pilot project that would be rolled out a year before the expected general election in 2010. Because of the results that were gathered from this pilot it was decided to abandon the text-based/mobile phone approach and instead focus our new developments on a query search engine via a web-browser (this is discussed in the next chapter)

7.2 Growth and development of mobile phone communications

Wireless communication technology is spreading around the planet faster than any other communication technology. The mobile phone has been able to connect millions of the previously unconnected and ITU (International Telecommunication Union)³⁵ estimates that worldwide the number of mobile phone subscriptions per 100 inhabitants has reached over 80 per cent of the global population by end 2011. No other information and communication technology and service has been able to have the same impact in terms of subscribers, particularly in the developing world, in so little time. In the UK, nearly half of internet users accessed the internet on their mobile phone (46 per cent) and the mobile internet advertising market in the UK grew by 118 per cent between 2009 and 2010 (Figure 7.1).

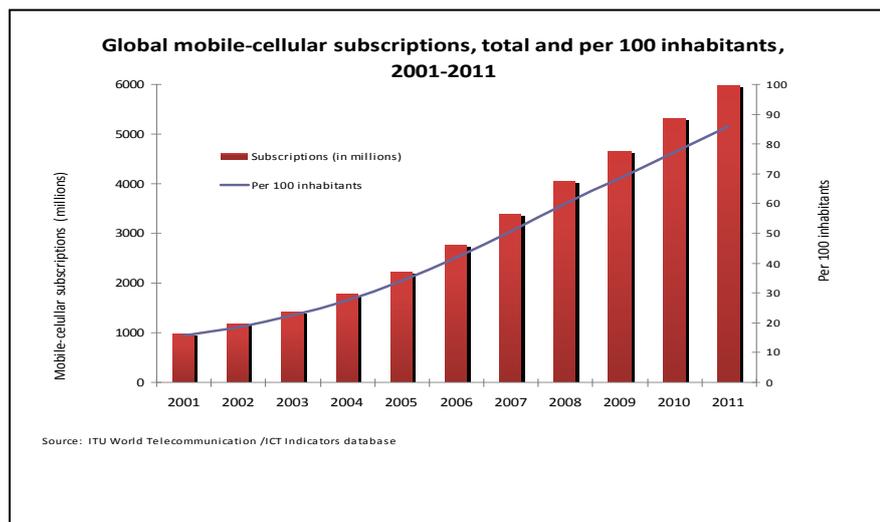


Figure 7.1: World Mobile Phone subscriptions per 100 inhabitants and mobile phone trend in the UK

Source: ITU, 2011 (<http://www.itu.int/ITU-D/ict/statistics/>)

Mobile phones offer possibilities the Internet cannot provide. Mobile phones add mobility and a spontaneity factor to potential political and social mobilisation, because users can react

³⁵ <http://www.itu.int/ITU-D/ict/facts/2011/index.html>

instantly and emotionally to events. Furthermore, personalisation is given through the typically person-to-person and social type of contact. Finally, the mobile phone is multimodal because it can transmit voice, images, and sounds, making it a tool for live transmission of events to be shared on the network (Castells 2007). In recent years there has been a gradual extension of the mobile phone into the political domain. Some examples include:

- Election pilots that allowed people to vote in the UK by text message May 2002
- The Italian government in 2004 sent a reminder message about upcoming European elections to all mobile phone subscribers.
- In 2004 Tony Blair became the first UK Prime Minister to use text message technology to answer questions submitted in advance by text message from members of the public as well as in real-time in a mobile phone chat-room, transmitted live from No.10 Downing Street.
- The Estonian Government allowed its citizens to vote using their mobile phones in the March 2011 general election.

Experts are predicting that in future elections, the use of mobile technology will become the standard and play an even larger role than in 2008 and 2012 US election where Barack Obama's use of text messaging was regarded as a successful way to reach out to the constituents. Like the Internet, mobile phones facilitate communication and rapid access to information. Compared to the Internet, however, mobile phone diffusion has reached a larger proportion of the population in most countries, and thus the impact of this new medium is possibly greater. In terms of accessibility, mobile phones can reach a public that other more traditional news channels cannot. For many people the mobile phone is never far from their side and using it to follow daily events can offer a greater sense of involvement than, for example, reading a newspaper. As more people use their mobile phones to access content, information and entertainment, a search function becomes a vital tool for finding what they are looking for quickly and easily.

7.3 *Mobile search versus Internet search*

Regarding the search process the computer desktop and mobile phone share some common issues, but also have some conspicuous differences. On a desktop website, it takes most visitors (70 percent) one week to act on a search in comparison to one hour to act on mobile phone, as people who are searching via mobile are ready to act and highly motivated to do so³⁶. Mobile searches are more precise than desktop searches, using up to twice as many keywords. However, as Nielsen's 2009³⁷ study on mobile usability pointed out, users' success rates when using mobile devices to access mobile sites averaged only 64 percent, in comparison to the 80 percent average success rate for users who access Web sites on a computer.

Mobile sites can take advantage of technology that automatically detects where users are to present local search results. When users set up their preferences or profile, personalized search results become even more relevant and valuable to them. Another fundamental difference is the type of information sought. Internet users often seek comprehensive, lengthy information whereas mobile users typically seek concise information, such as news headlines, sports scores, weather etc. all of which must be organized in a way that makes it fast and convenient to find exactly what they want.

These limitations describe the problems that need to be taken into account to provide an acceptable mobile search. In order for mobile search to reach its full potential, it has to be fast and convenient to use, regardless of whether the user has a smart phone with a QWERTY

³⁶ <http://torsionmobile.com/>

³⁷ Nielsen Norman Group Report: Usability of Mobile Websites & Applications: 237 Design Guidelines for Improving the User Experience of Mobile Sites and Apps.

keyboard or a feature phone. The ideal mobile search should provide users with immediate, relevant answers, with no distractions.

7.4 Election Day 2009 European Parliament

An experiment using mobile phone technologies was conducted for the 2009 European elections with a view to introducing a more extended version in time for the 2010 general election. The aim was to create intelligent systems that learn from the ways in which mobile phone users try to interrogate electoral data. The data covered the 2004 and 2009 European election results and 2009 nominations.

For this purpose it was decided that we would use the Question-Answering Mobile ENgine (QAMEN) (Burns, Fallon et al. 2005). QAMEN converts web pages to a simplified format compatible with handheld devices and it is based on the industry-standard SMS messaging technology and thus works with any mobile in any GSM network. The idea was that QAMEN would sit in front of a database containing details of electoral data relating to the 2004 and 2009 European elections and that users would send texts that would be passed through the question-answering engine with the answer being returned back to the user.

The experiment was tested after launching a working system a few days before EU 2009. It was decided that the specialist group of the Political Studies Association in the UK that focuses on elections would be the most suitable way to advertise the experiment. The group, Elections, Public Opinion and Parties (EPOP) agreed through its national convenor to facilitate the experiment. A notification was sent to all EPOP members via email list with a request for their participation:

“One of the Election Centre’s doctoral students, Elena Long, as part of her research is piloting an experiment using mobile phone technologies prior to this year’s European elections with a view to introducing a more extended version in time for next year’s general election. The aim is to create intelligent systems that learn from the ways in which mobile phone users try to interrogate electoral data. A parallel system running over the internet may be used to compare syntax structures between the different technologies – what we type into a search engine is different to what we may txt. The data only cover the 2004 and 2009 European election results and the 2009 nominations. We need willing participants for this experiment to function properly. Where better to start than with EPOP members who realise the value of participation and who may find European election results interesting. The maximum charge for a text message query is 10p but for most people with phone contracts and few friends the charge is effectively zero per text. The more people that participate the better, so please forward the web address and text number to others that may find the experiment interesting and potentially informative.

QAMEN (Question-Answering Mobile ENgine) is an SMS oriented engine which frees users to have an expensive mobile phone with a web browser. EPE- 2009 data are used by default; therefore “2004” is needed to be added to the SMS for the request about 2004 European Election. No commas or full stop are required in the texts.

Send SMS Request to +4477 8620 1900 “

A database containing information about the parties’ votes share, an electorate, the total vote cast and the turnout by a region was created. It also includes number of seats and the names of the elected Members of the European Parliament 2004 and candidates nominated for the 2009 elections.

By default any request is considered by QAMEN as a request for searching in the local knowledge base (effectively a hard drive located in the company's office, and/or the Internet. For the purpose of the experiment only the local knowledge base was used. QAMEN is based on industry-standard SMS messaging technology and works with the mobile phones in any GSM network therefore it is useful for people on the move and who are unable to access a personal computer.

A mobile web consists of web pages that are designed specifically for display on mobile devices. Due to their limited capabilities (relative to standard computers) mobile devices access and render web content using the specialized Wireless Application Protocol (WAP). WAP is the current global standard for mobile internet applications and browsing. It offers functionality similar to HTML web browsers, but is designed to accommodate small browsers with limited memory, as well as low bandwidth connections to the internet. However, most do not work well due to the volatile nature of the current mobile web—there are many outdated and dead links.

By contrast, QAMEN is an HTML compatible search engine. It works almost the same as computer browsers and can access HTML search pages from vast internet webpage databases in the same way that computer browsers do. Mobile question-answering differs from standard information retrieval methods. First, it needs to retrieve specific fact information rather than whole documents. Second, it should select from among the found facts the shortest and appropriate fact to meet the 160 characters requirement of an SMS text message. QAMEN appeared ideal for this purpose because what a user really wants is a precise answer to a question.

Furthermore, QAMEN does not return full search-engine-type listings with links to websites, as the regular Google or other search engine services on the web do. The interaction model is

simpler and makes it possible to enter a query without bringing up a page and navigating to the query form. The results come back in a very compact form - more suited for a small screen (Lovitskii, Thrasher et al. 2008).

7.5 The problems faced

In terms of our initial expectation, some of results provide a rather disappointing outcome. Some of these outcomes are now discussed.

First and foremost, the total number of mobile requests was 226, which was far from satisfactory considering that the target audience was election specialists. It shows the necessity of a longer better advertising campaign if we wanted to attain a greater response. Of course one possibility is that we used the wrong audience for the research, as EPOP members most likely already had all the information that they are interested in unlike the general public that probably knew very little about European elections – but then these people might not have wanted to learn anymore about the candidates that were standing.

A much more serious problem to arise within the experiment is that end users are generally unable to describe completely and unambiguously what it is they are looking for at the start of a search. Natural language is often ambiguous and is dependent on a great deal of world knowledge. However, most problems are shown not to be due to ambiguity but to excessive expectation of the capability of the system on the part of the user or the use of world knowledge that is not reflected in the system's knowledge base.

Before a request for information can be answered, it must be known what knowledge sources are available. If the answer to a question is not present in the data sources (for example, information about other elections that a candidate might have contested), no matter how well

we perform question processing, retrieval and extraction of the answer, we will not be able to provide a correct result.

In order to implement a working natural language system we restricted the European election experiment to cover only a limited subset of the vocabulary and syntax of a full natural language. This allows ambiguity to be reduced and processing time to be kept within reasonable bounds but it also meant that some types of queries would fail simply because the natural language system was restricted in its scope – it could process some but not all of the vocabulary that users might employ to describe ‘vote share’ for example.

As a rule, QAMEN does not require the use of upper case or punctuation. Even more, the system allowed users to omit certain search terms such as those associated with questions: ‘what’, ‘which’, ‘who’ and ‘how many’. This greatly helps to simplify a user-system dialog. The fact is that the search queries in majority cases was not spelt, punctuated, and capitalised correctly, but the main requirement for the system is to handle non-standard or poorly formed/structured (but, nevertheless, meaningful) user’s inquiries (Francis, Lishman et al. 2007). Sometimes, users mobile requests were ungrammatical. A long request such as “How many votes did the Conservative party win in the South East in 2004 European Elections?” could be replaced by “con share se 2009” and the system was able to respond to that.

Answer extraction depends on the complexity of the question, on the answer type provided by question processing, on the actual data where the answer is searched and on the question focus and context. The answers should be presented in as natural as possible way. In some cases, simple extraction is sufficient. For example, when the question classification indicates that the answer type is a name (of a member of parliament), a quantity (turnout, electorate or a vote share) an extraction of information is simple. For other cases, the presentation of the

answer may require the use of fusion techniques that combine the partial answers from multiple documents.

In theory the experiment allowed users to compare candidate nominations for 2004 and 2009. However, rather than typing a long text request such as: 'List of people who have been nominated second time', a user simply texted: "who again". This shows that the natural language interfaces led users to attribute more intelligence to the system than was warranted. This leads to unrealistic expectations of the capabilities of the system on the part of the user. Such expectations will make it difficult to learn the restrictions of the system if they attribute too much capability to it, and will lead to disappointment when the system fails to perform as expected.

What people usually do not realise is that every database driven application starts by knowing nothing. Building a knowledge service thus involves creating and populating database tables, and writing software to manipulate and present those tables. Such systems need to be taught everything from scratch, and at the end they know only the knowledge put in by the developers.

In the general case, under the knowledge base structure one should understand the regularity of data distribution in memory assuring the storage of various links between separate elements of stored information. At every moment knowledge base deals only with relatively small fragments of the external world. So, the corresponding structures are needed to integrate these fragments separated in time into the integral picture. The structures obtained as a result of integration should contain more information than had been used for its creation.

The organisation of knowledge storage should make allowance for such features of the human memory as: associability; ability to reflect similar features for different objects and

different features for similar objects; independence of the knowledge extraction time from the volume of knowledge being stored in this memory; associative relations weight variable; representation of the environment statistical properties (Lovitskii, Thrasher et al. 2008). More importantly, the knowledge cannot realistically be regarded as a static resource, to be accumulated and stored within a system. It is a formative, self-organising character, with the ability to change the organisation within which it is held. The local database of any mobile question-answering system, therefore, cannot be complete; we shall never be able to establish information completeness of the knowledge base. Furthermore, the system should be self – learning. When QAMEN could not find a suitable response in the local data base, it should search using the Internet and then not only send the reply to the user but also automatically extend the existing knowledge base.

The Screen Size Problem: The experiment highlighted the problems with the user interactions with small display devices as it remains a challenge due to the limited display capabilities and information overload. Search interfaces for small screen devices are associated with the within-page vertical scrolling or paging requirements for viewing content. Vertical scrolling describes the action of viewing the content outside the screen display area shown in a progressive manner, successively. By contrast, paging permits access to the next full screen's worth of content without any further action by the user (Jones, Marsden et al. 1999).

To make content available for displaying on small screen devices, it is not uncommon for long lists of search results to be divided into separate pages that contain a reduced number of results. Breaking the content up into smaller, more manageable chunks may be necessary for transmission requirements, as well as a means of aiding presentation. However, there is an associated cost with this approach; page-to-page navigation increases the amount of user

interaction and reading time. Both of these factors may have financial implications and as a consequence may impact on the way users use such services.

7.6 Conclusions

This chapter present an investigation into the effectiveness of using Question-Answering Mobile Engine (QAMEN) and a natural language search to obtain information about European Elections in the UK on users mobile phones. In terms of our initial expectations, some of the results from the experiment, such as complete lack of interest from EPOP members, was just the beginning of what must be regarded as a failed experiment. Nevertheless, even with a failed experiment there are opportunities to learn some important lessons.

At the outset, it is important to consider how questions about the electoral process can be adapted so that mobile phone technology becomes yet another tool of engagement. Wireless communication technology is spreading around the planet faster than any other communication technology to date. In order for mobile phone searches to become commonplace two things must happen: users must try and then regularly use the services, and companies in the mobile communication industries must generate profits from these searches. People are apparently comfortable in using their phones to participate in online voting for such programmes as The X Factor and Strictly Come Dancing but we suspect there is much more reluctance to use mobile phone to access information about the candidates standing for an election.

The experiment demonstrates that we need to understand a lot more about the differences between how users access content on their mobile phones versus personal computers and of

course tablets. When users tried to access information from their mobile phones they expect the same depth and breadth of content that they would find on their computers. Furthermore, as with all mobile applications, mobile search must contend with a variety of handset displays, input capabilities memory levels and operating system. The ideal mobile search solution should provide a consistently good user experience across all networks and devices.

From the experiment, it becomes clear that mobile search solves multiple end-user needs to get relevant information. Nearly all providers of mobile search agree that providing an appropriate result that immediately addresses the mobile subscribers' needs is the most effective approach. Yet, there are limits in the information available in many contexts, as well as our own language limits.

Additionally, two main influential factors need to be kept in mind in this discussion: the length limitation to 160 characters and the challenge of the interface (as long as mobile telephone keyboards are far from being comfortable for writing).

Overall, the experiment proved to be a disappointment, especially when setting it up took a lot of time and energy. With such a small response and users' expectations that the system would be able to understand all requests for information it proved impossible to execute the experiment in a way that we had originally hoped. It was decided, after this experience, to abandon plans to implement a similar system in time for the 2010 general election. Instead, we launched a computer-based system designed to run over the internet but one which was still based on an intelligent question-answering search engine.

8 Natural Interface to Election Data

8.1 Introduction

This chapter examines the creation, development and implementation of a system that was designed so that members of the public, with little or no experience of managing and interrogating databases, could independently issue questions and receive answers from an intelligent system. Having failed to implement a fully-functional scheme for phone text query-based system we then tried a different approach which we termed, ‘Natural Interface to Election Data’ or NITED.

The aim of this project was to bring the contents of a relational database, in this case one containing results and data relating to the 2005 and 2010 UK general elections, to the public’s use. The objective was to implement NITED in time for the 2010 general election and provide a facility that would enable people to write inquiries using natural language and to deliver data responses in the form of tables, graphical displays and crucially, a new mapping visualisation for three-party vote share.

The structure of the chapter is as follows. Following this introduction the section 2 considers some general points about the factors that need to be taken into account when designing a web-based dialog system that is intended to allow users to issue queries that use a natural language interface. This section highlights recurring problems with such systems – what kind

of users are likely to want to use such a system, how will they interact with it and what kind of responses would then want from such a website? When developing a web-based application that interrogates a database of election results there are considerations that have to be thought about in terms of field characteristics (how long are constituency names), inconsistencies in the data (naming of constituencies, for example,) ambiguities in the data (place names and surnames having text strings in common, for example, Wells is both a constituency name and a candidate surname). Unfortunately, as section 4 shows many of these problems are unforeseen and only surface during beta testing but solving them slows down implementation of the system. In section 5 we describe the process for designing input and output interfaces – the overall ‘look’ of the application. Given that one of our principal concerns is data visualisation this is an important element of NITED that required close attention. One of the objectives that we set ourselves when building NITED was how inexperienced users could build quite sophisticated database queries. Some progress was made in this respect, as section 6 shows, but this also shows that the complexity of electoral data meant that implementation of this part of the system was far from ideal.

Section 7 describes how for this experiment the data was originally stored in an Access database where SQL (standard query language) is normally used to extract data. One of the challenges, therefore, was in designing a system that allowed natural language queries to be translated into a SQL format and then to output the required data. We went to considerable lengths to provide users with a good ‘help’ facility in NITED but as section 8 shows, the more we tried to guess and answer some ‘frequently answered questions’ the bulkier the user-interface became with long and detailed instructions about how to use the system. Section 9 describes some of the difficulties that were encountered in trying to implement natural language enquiries. With hindsight, introducing this facility simply made the system too

cumbersome and difficult to use. The chapter finishes on an optimistic note, however!

While it is fair to say that the system we designed is deficient in many respects there is one feature that we introduced that is innovative and an advance on methods for visualizing party vote shares within each constituency (section 10). Although NITED in its existing form will never be fully implemented it is hoped that our ‘drill-down maps’ may become commonplace as examples of intuitive design.

8.2 Designing a Web-Based Dialog System

Dialog is the form of data processing involving an interaction between a computer system and a user who uses a keyboard and/or mouse to enter requests which the computer responds to. As users struggle to navigate on-line information there is need for automated question-answering systems. We need systems that allow a user to ask a question in everyday language and receive an answer. Current search engines, such as Google, Firefox etc, can return ranked lists of documents, but they do not deliver answers to the user.

To answer a question, a system must analyse the question, perhaps in the context of some ongoing interaction; it must find one or more answers by consulting on-line resources; and it must present the answer to the user in some appropriate form, perhaps associated with justification or supporting materials.

At first glance, the creation of Web-based Dialogue System (WDS) for NITED is a well defined problem because the developing of dialogue systems has a long history, but from our point of view there are several problems which need to be solved before developing this dialogue system. Figure 8.1 depicts the architecture of WDS NITED

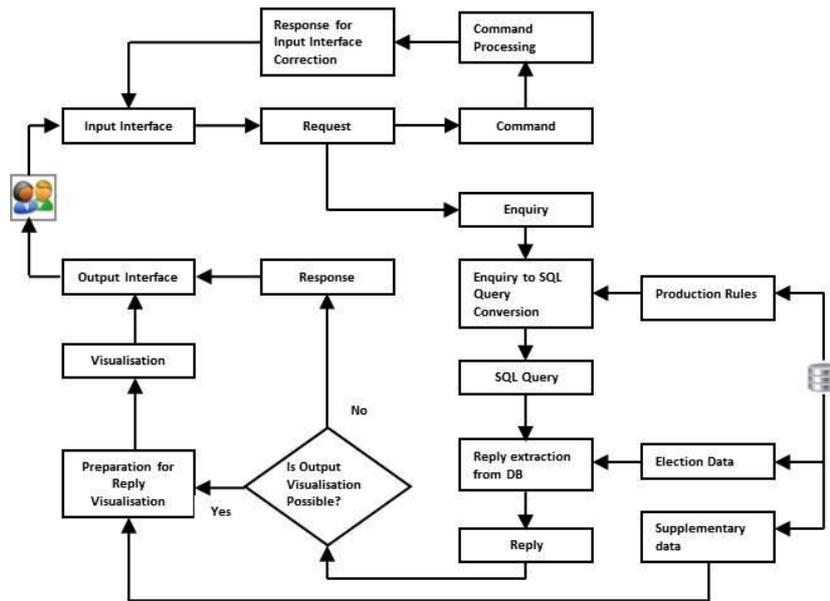


Figure 8.1: The Architecture of Web-based Dialogue System

NITED

There is the general set of dialogue system components, namely: user, enquiry, input and output interfaces, data, processor, and result. There are some dependencies between these components; the diagram of these is shown in Figure 8.2.

The entire WDS consists of two main parts: one part is responsible for request handling; the other provides the means for extracting the corresponding reply from the database and structuring a reply. Reply is always represented by a single table. Not every reply table can be visualized (if the requests asks for a list of names then table is the only response that can be made) and therefore the reply table should undergo a preliminary analysis to discover an appropriate form of response representation – table, graphical display or map or some combination of these.

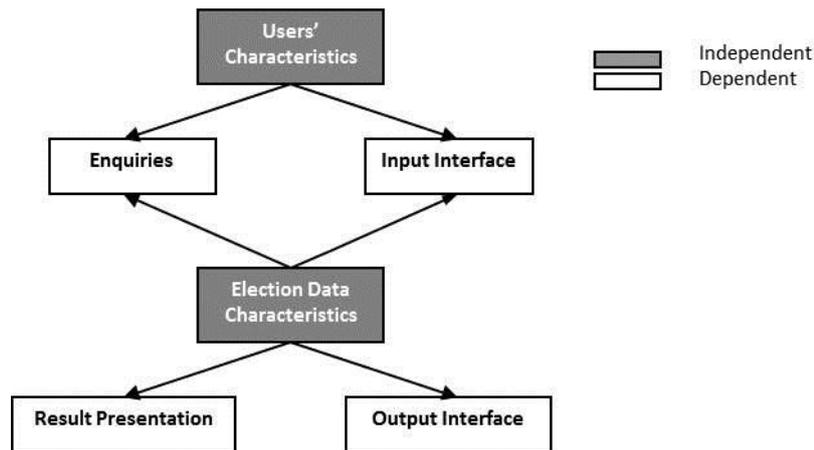


Figure 8.2: Diagram of Web-based Dialogue System Components Dependencies

Two independent components have a strong influence on the rest of the system architecture and that is why the characteristics of these independent components should be investigated first of all.

8.3 Identifying Potential Users

A web dialogue system should be aware of the diversity of the potential users. Before we can answer the question “*How do we make our user interfaces better?*” we must first address the question: “*Better for whom?*” One of the primary mantras of user interface design is “Know the user.” From Card and his fellow researchers perspective (p.67), “Successful designers are aware that other people learn, think, and solve problems in different ways.” (Card, Mackinlay et al. 1999).

It is important to note that when systems are being designed for usability, the first step should be to identify the target user population – in our case this was the general public. A computer system might considered to be in working order in a sense that it works without falling over, but it may not do what the users want to do in order to achieve their goals. Computer systems should be designed for the needs and capabilities of the users for whom they are intended

(Stone and Stone 2005). Consequently, the designer must understand the diversity of the users and their tasks. A design that is better for a technically skilled user might not be better for a non-technical politician, businessman or an artist. Because NITED is a web-based system, any kind of users could in theory have access to it.

Over a number of years and across a series of research programmes (Francis, Lishman et al. 2007; Long, Lovitskii et al. 2009) users' enquiries about electoral data were collected and stored. Direct observation of these enquiries demonstrated that most users are naturally inclined to achieve the desired result by using minimum effort. They do not want to type in a long enquiry such as, "*Who won the general election in the Middlesbrough South & Cleveland East constituency?*" but instead "*winner in Middlesbrough?*"

This generally means that the users want minimal input to achieve the desired output i.e.:

- The user does not like to use the keyboard.
- The user prefers select and click mouse instead of typing in an enquiry and then pressing the '*Enter*' key.
- If the Dialog System (DS) requires Multilevel Input Interfaces then users prefer the minimum number of levels, and the minimum number of Input Interfaces in each level. User fatigue sets in quickly.

In addition we need to take into account that many users may not be visually oriented. They may prefer textual or numerical data formats in scrolling lists rather than visual presentations. This matches user-preferences for textual instructions rather than a map.

8.4 Designing for the analysis of Election Data

Any election provides information which constitute some of the important variables in the description of results – the number of eligible electors who vote (turnout) and the distribution

of votes among the competing candidates/parties (vote and vote share). Other information such as the sex of candidates and the finishing position of candidates on the ballot is also requested. In a general election there are results for many constituencies to be considered. Voting figures can be aggregated, averaged, graphed and used to construct maps.

To design effective visualisations, the designer must first understand the data that will be used in visualisation. In general elections, for instance, we have the percentage turnout and the percentage share of the vote for each party in every constituency, which constitute the most obvious variables to begin with when designing a query-based system.

The designer must identify the sources of the data, the level of completeness of the data, and the type of the data. Good data visualisations are not only task dependent, but they also become election dependent – the data apply to one and only one election. In our case the Access database contained the separate results for the 2005 and 2010 general elections.

Let us consider some important characteristics of GE2005 and GE2010 databases and the challenges they provide:

- **Length.** The length of the text value for some fields is rather long (e.g. “*Knowsley North and Sefton East*”) and might cause problem for users having to type in such long string.
- **Meaning.** For one meaning different words (symbols) are used within the database e.g. “*Blackpool North and Fleetwood*” and “*Hackney North & Stoke Newington*”.
- **Ambiguity.** Many ambiguities exist within an elections database that pose problems for a natural language interface. For example, the words *Angus, Bath, Corby, ..., Wells* are values both in the fields *Constituency* and *Surname* in the database. In effect, the parsing of even a simple enquiry is quite complicated and requires a powerful knowledge base to interpret these requests correctly (Lovitskii and Wittamore 1997).
- **Similarity.** Sometimes a user simply makes a mistake and instead of typing in the intended constituency *Hereford* in the enquiry: “*Who won the election in Hereford*” the user instead enters *Hertford* (it is **wrong** but at the same time it is **right** from the NITED point of view because the text query consists of part of an existing constituency *Hertford & Stortford*).

Naturally, NITED cannot know the user has misspelt Hereford and will return an answer for the constituency *Hertford & Stortford*.

8.5 Input and Output Interfaces

The user interface (UI) is a very important consideration in the development of any web-based dialog system. Studies have been done on the feasibility of natural language interfacing (Hendrix 1978; Walts 1978; Copestake and K. 1989; Androutsopoulos, Ritchie et al. 1995; Androutsopoulos and Ritchie 2000) that has resulted in an abundance of systems being developed and implemented. These include, PRECISE (Popescu, Etzioni et al. 2003), STEP (Minock 2005), ORAKEL (Cimiano, Haase et al. 2007), Aqualog (Lopez, Pasin et al. 2005), GINSENG (Bernstein, Kaufmann et al. 2006). Because of these systems, natural language processing has significantly contributed to the field of human-computer interaction in terms of theoretical advances and practical applications.

The starting position is that the user interface for applications should be simple, intuitive, and user-friendly. This generally means that the application minimizes undesired outputs for the user. We tried to create a natural user interface to improve the efficiency, effectiveness, and naturalness of user-NITED interaction by representing, reasoning, and acting on models of the user, domain and tasks.

For NITED the main part of the NUI is a graphical interface, which accepts input via computer keyboard and mouse. The actions are usually performed through direct manipulation of the graphical control elements rather than by extensive text entry.

NITED's user interface has to be election domain specific i.e. contents and features of the election should shape and define the structures of Input and Output interfaces. The interface should be natural for any user, meaning that navigating around the site is intuitive.

Natural language, understood here to mean a tool that people use to communicate with each other, however, has characteristics that reduce the efficiency of textual information retrieval systems. These properties include linguistic variation and ambiguity. By linguistic variation we mean the possibility of using different words or expressions to communicate the same idea – is “turnout” the total number of people that voted or the percentage of the electorate that participated. Linguistic ambiguity is when a word or phrase allows for more than one interpretation – does ‘change’ mean change in vote share, difference in votes or when the seat changes hands (Vallez and Pedraza-Jimenez 2007). In general, the design of natural language interfaces to databases is regarded as a difficult problem since human interaction is often vague, ambiguous or highly contextualized (Androutsopoulos and Ritchie 2000).

The central question to be addressed is how to improve access to information for would-be users. Such users may not understand the structure of the database, may not know exactly what is in it and how data is stored there. The purpose the User Interface (UI) is to provide a means of:

- **Input** UI, allowing the user to manipulate an application, and
- **Output** UI, allowing the application to indicate the effects of users’ manipulation.

8.5.1 Input interface

We designed three different types of Input UI:

Command-Line User Interface allowed users to choose form “**GE2005**” – “**GE2010**” to activate the corresponding database. Only text field, keyboard and mouse are required for CLUI.

The next two Input UIs are Graphical User Interfaces (GUI). The convenience of the GUI made it the standard in human-computer interaction. The GUI generally provides users with

immediate, visual feedback about the effect of each action. There are several general requirements to designing a good GUI:

- Provide meaningful contrast between screen elements.
- Create groupings.
- Align screen elements and groups.
- Provide three dimensional representation.
- Use colours and graphics effectively and simply.

In respect of NITED there were two GUI's.

1. Graphical Query Input User Interface (GQUI) is a user interface based on graphics (checkboxes, clickable images and list boxes of Regions, Constituencies, Parties and Candidates) instead of text; it uses a mouse as well as a keyboard as an input device.

GQUI provides three different types of user's request:

- *Frequently Asked Questions (FAQ)*.
- *Natural Language Enquiry Template (NLET)* combines FAQ and a slot represented by each variable within square brackets, the value of which should be selected by the user from the corresponding list of values. For example, "In which constituency did [party] achieve its highest vote?" where the user supplies the party category.

2. Graphical Charts Input User Interface (GCIUI) provides the direct way to accomplish chart creation as a response to user's request restricted to three party (Conservative, Labour and Liberal Democrat) groups to see how they are oriented in the UK map of regional spaces. For example, a user may want to compare party performances in the South West of England, in Scotland or in Wales.

GCUI is a useful method for the analysis of election data. For example, NITED displays the results using interactive maps. Initially the user is shown the UK map with selectable regional areas. When a region is selected, NITED will direct the user to an analysis of that region's election results.

8.5.2 Output Interface

NITED constructed three different types of Output UI:

Table Output Interface (TOUI) is the general way to represent results to each user's enquiry. The default for a query that seeks election result is represented by a single table (Figure 8.3), in this case the response to a query that places elected women M.P.s in rank order of vote.



Figure 8.3: Table Output Interface

Charts Output Interface (COUI) allows the user to see election data represented by Histograms and Maps (Figure 8.4). COUI should meet some basic requirements:

- Charts should be designed to encourage the user to make comparisons between different pieces of data.

- COUI should provide views of the data at many levels of detail – spatial separations is an obvious example. This principle relates to the "Drill down" and "Level-Of-Detail" capabilities of visualisations. With these capabilities, NITED can allow a broad overview of the data to be given and, at the same time, allow the user to have access to the detailed data that underlies the overview.

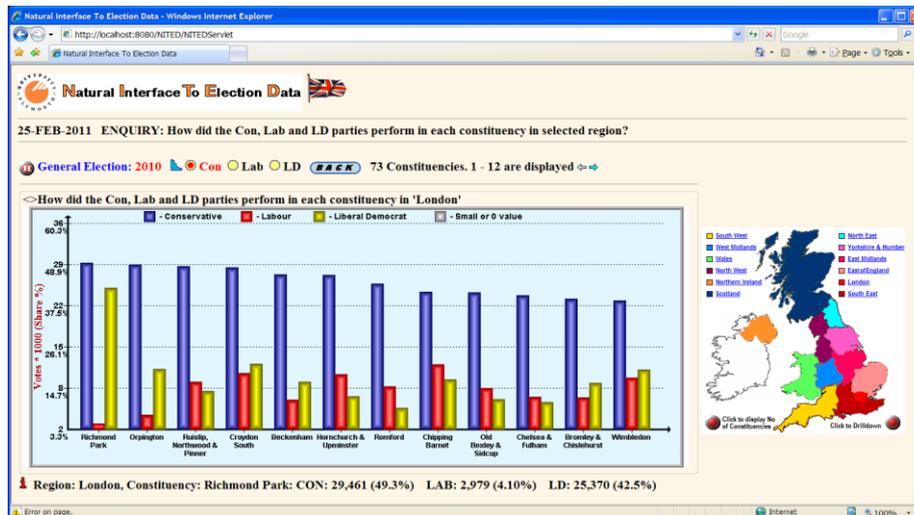


Figure 8.4: Chart Output Interface

Drill Down Maps (DD-maps) - Maps are an obvious form of representation for spatially-based data. For example, if a user would like to see how some region is performing in comparison to other regions in terms of party votes, the user could look at a map visualisation. A map-based representation can become a true multivariate representation (Figure 8.5). A more detailed description of this innovative mapping procedure is featured in 8.11

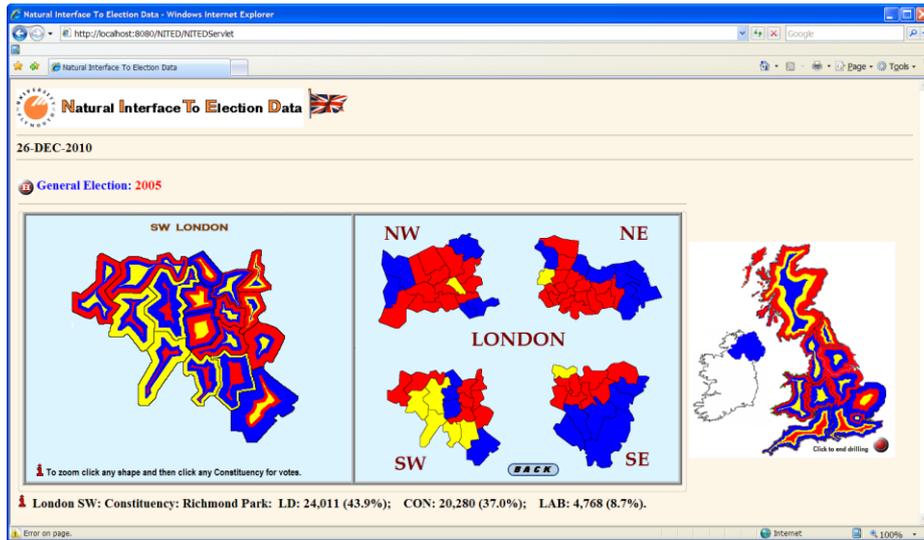


Figure 8.5: Mapping output interface

When the user has to move through various levels of data, a drill down style of navigation is commonly offered –national, country, region, constituency. The drill-down clearly groups information of magnitudes ranging anywhere from the entire country down to an individual constituency.

Input and Output UIs interactions are shown in Figure 8.6.

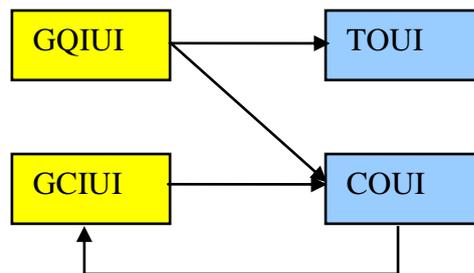


Figure 8.6: Input and Output UI Interaction

GQIUI always first of all produces the TOUI as its default operation. Additionally, NITED, as an intelligent system, analyses the result to discover whether it is logically appropriate to represent the output as a chart as well using COUI. If the answer is “YES” TOUI displays button “Chart” (see Figure 8.7) As a result of clicking this button, NITED substitutes TOUI

for COUI (see Figure 8.4). COUI represents the dynamic part of GCIUI and a user can click any histogram to see the result for the selected party. Within NITED charts were always preferred to Tables because users more easily recognize patterns in data.

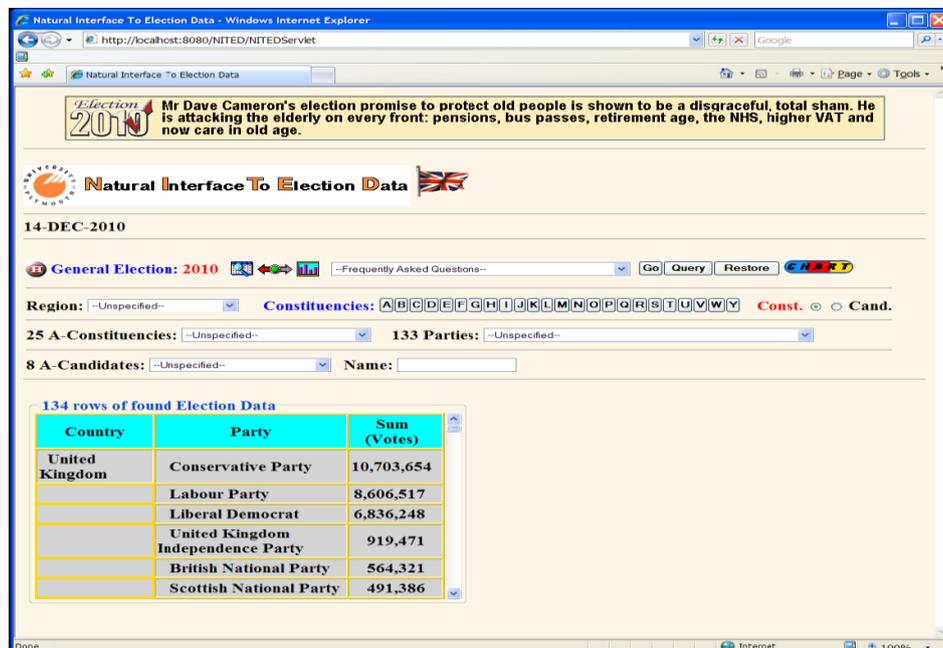


Figure 8.7: Example of result of the 2010 general election

The user interface should satisfy some general requirements:

- It should be consistent.
- Every action taken by a user should have an explicit response from NITED, i.e., the UI should provide meaningful feedback. Very often a user's natural language enquiry cannot be interpreted because the language lies outside the application domain. Therefore, the database access system has to be able to decide whether the enquiry is meaningful or not.
- The UI should provide an easy way for the user to undo their actions. This allows the user to correct their errors by stepping back.
- Usability is mainly a characteristic of the UI, but is also associated with the functionalities of the WDS and the process to design it. It describes how well a WDS can be used for its intended purpose by its target users. To improve usability of the WDS input interface the two levels of input interfaces are used (see Figure 8.6).

8.6 Users' Requests

In general, the design of natural language interfaces to databases is regarded as a difficult problem since language is often vague, ambiguous or highly contextualised (Androutsopoulos and Ritchie 2000). However, formal query languages are difficult to learn and utilise by non-computer-specialists. Graphical interfaces and form-based interfaces are easier to use but still require some learning by the end-user. In contrast, an ideal NULI would allow queries to be formulated in the user's native language. This means that an ideal NULI would be more suitable for casual users, since there would be no need for the user to spend time learning the system's communication language.

It is important to distinguish two different forms of request regarding the NITED: command and enquiry. Command is used by the user to change an input interface and/or create an enquiry. To extract a reply the natural language enquiry has to be converted to the corresponding SQL query (hereafter word query will have just one meaning SQL Query).

We distinguish two different types of commands:

- **Hidden Command (HC)** is used to change the input interface and/or create the user's enquiry. HC is generated by clicking the corresponding check box, radio button, clickable image or some item from list box.
- **Text Command** is represented by single words and is seldom used. Just two pairs of text commands are used: "GE2005"–"GE2010" to activate the corresponding database.

Enquiry can be represented by:

Natural Language Enquiry Template combines a list of values to be selected when required and were constructed as generalisations of what type of query users would require.

Examples of some Frequently Asked Questions (FAQ) that users could select from a menu-type list are shown below:

- What was the result in [constituency]?
- How many votes did [party] win in [constituency]?
- Which party won the election in [constituency]?
- Who won the election in [constituency]?

We constructed the initial set of FAQ but when users began to use the system new natural language enquiries were collected and stored by NITED and then added to FAQ. When the user selects an appropriate query with some descriptor in square brackets, then selects the corresponding values from the list and clicks “Go” - the result is displayed instantly (see Figure 8.8 which shows the response to the query, what was the largest number of constituency votes acquired by the Conservative party?).

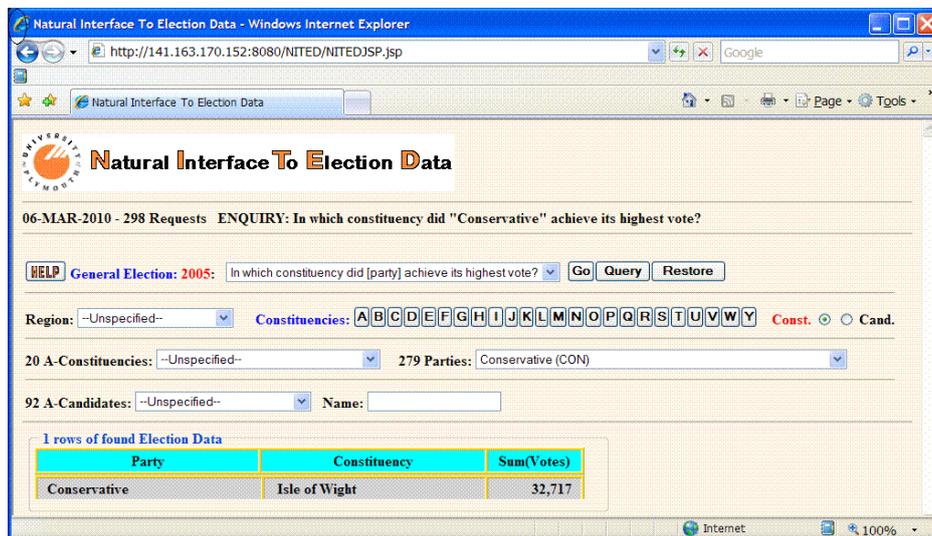


Figure 8.8: Screenshot of the Natural Language Enquiry Template

Created Enquiry (CE). The user can build his/her own enquiries using any combination of the descriptors; each of these descriptors represents a field within the database, e.g. [region,

constituency, party, etc.]. In the screenshot described in Figure 8.9 the enquiry has asked for women MPs in the south west of England.

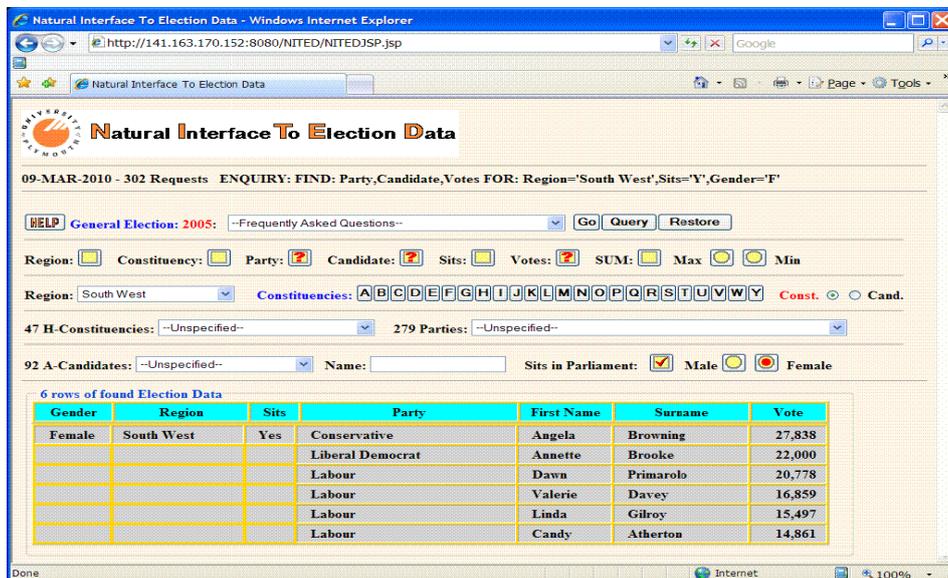


Figure 8.9: Screenshot of the Interface for Enquiry Creation

Every query set by a user in the process of creating an enquiry is then returned as explicit feedback from NITED as a form of confirmation so that the user can see that the data correspond with the terms of a specific query (see Figure 8.10). The aim here was to demonstrate to the user that the system is responding to the particular parameters set by the user – party and candidate name for women candidates that were elected in the south west.

ENQUIRY: FIND: Party,Candidate FOR: Region='South West',Sits='Y',Gender='F'

Figure 8.10: Example of feedback generated for user enquiry

Personalized Enquiry (PE). Personalization of an enquiry reflects the system’s capacity to store particular enquiries issued by certain users. For example, suppose the user saved the enquiry: “FIND: Party, Candidate FOR: Region=’South West’, Sits=’Y’, Gender=’F’” as a personalized enquiry. The next time when that user calls NITED and identifies himself/herself a personalized list of FAQ appears and the user will be able to select from the list. Before including the enquiry in the list of FAQ NITED translates it

into natural language: “*Show a list of parties and all the women elected in the South West region who now sit in Parliament*”.

8.7 From enquiry to SQL Query translation

Structured Query Language, or SQL, is the standard query language for accessing data held within a relational database. Despite the processes of enquiry to query conversion becoming well defined (Lovitskii and Wittamore 1997) the process remains quite complicated and depends on database complexity. For a complex database there would be a necessity to access data from many different tables within and join those tables together in a report in order for a response to happen. It is extremely important that an intelligent system can do this because non-technical users do not know how to join tables to get a more comprehensive view of the data. Sometimes a very simple question written as English is in fact a very complicated query for the database to output.

For example, a quite simple query: “*Who won the election in [constituency]?*” or “*How did [party] perform in [region]?*” is difficult to respond to unless the system understands “*who is who*”, the meaning of “*won*” and “*perform*” NITED cannot answer such questions without a certain level of prior understanding. To “explain” it to NITED the **Production Rules (PR)** need to be specified.

An example of the pre-condition production rules is described below:

- | | | | | |
|--|---|----------------------------------|---|------------------------------------|
| <Precondition> | ↦ | <Antecedent> | ⇒ | <Consequent> |
| 1. AD:Election2010 | | ↦ who | ⇒ | candidate; |
| 2. AD:Election2010 | | ↦ [candidate]:<win⊕won⊕highest > | ⇒ | [SQL]:<MAX(votes)>; |
| 3. AD:Election2010 &
database:MS Access | | ↦ votes | ⇒ | [Field]:<gcr_post_election_votes>; |

- | | | |
|----|---|--|
| 4. | AD:Election2010 &
database:MS Access | \mapsto candidate \Rightarrow [Field]:<can_first_name, can_last_name>; |
| 5. | AD:Election2010 &
database:Oracle | \mapsto [party]:<win \oplus won \oplus highest> \Rightarrow
[SQL]:<MAX(SUM(votes))>; |
| 6. | AD:Election2010 &
database:MS Access | \mapsto [party]:<win \oplus won \oplus highest > \Rightarrow [SQL]:<TOP1,
SUM(votes),DESC>; |
| 7. | AD:Election2010 &
database:MS Access | \mapsto perform \Rightarrow candidate,votes; |

where \oplus - denotes “*exclusive OR*”. **Precondition** consist of **class₁:value₁ [& class_i:value_i]**.

Antecedent might be represented by: (i) **single word** (e.g. *who, won, perform, etc.*), (ii) **sequence of words** (e.g. *highest vote, etc.*), or (iii) **pair - [context]:<value>**. Context means avoiding word ambiguity and thereby distinguishing the difference in meaning between “Candidate *winning an election*” and “Party *winning the election*”.

Like any web-application NITED is closed to everyone except the developer. But in reality the person who knows and understands the database – in this case the elections expert has to check FAQ, delete unused queries and add new ones in response to user requests. Of course, the expert developer writes the SQL syntax for the new FAQ. Because the FAQ and production rules files lie outside the database with respect to NITED the external tool is used to create and amend/up-date where necessary, the production rules (see Figure 8.11 where new production rules are being assembled).

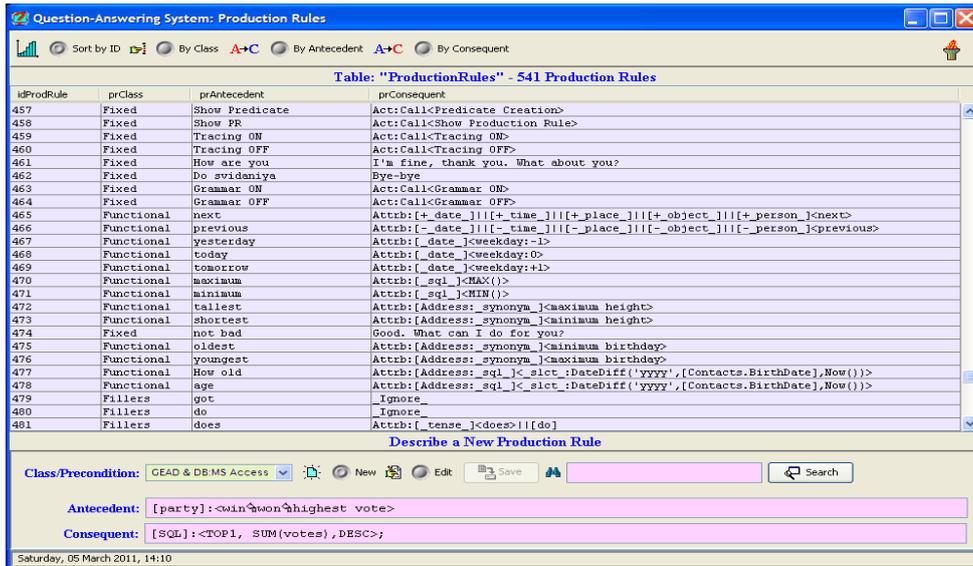


Figure 8.11: Screenshot of Production Rules creation

8.8 Help Instructions

Help Instructions in a web application are needed to assist users who may be inexperienced about electoral matters. Clear instruction has to take account of the type of users that would be interacting with NITED. As a minimum these instructions should be short enough but provide sufficient information about the screen function, the available options, how to get started and frequently used features. Step-by-step worked examples were shown in our help instructions.

The problem was the NITED is a complex database and we were trying to develop a natural language interface for it. Inevitably the set of instructions grew and became unmanageable (see Figure 8.12).

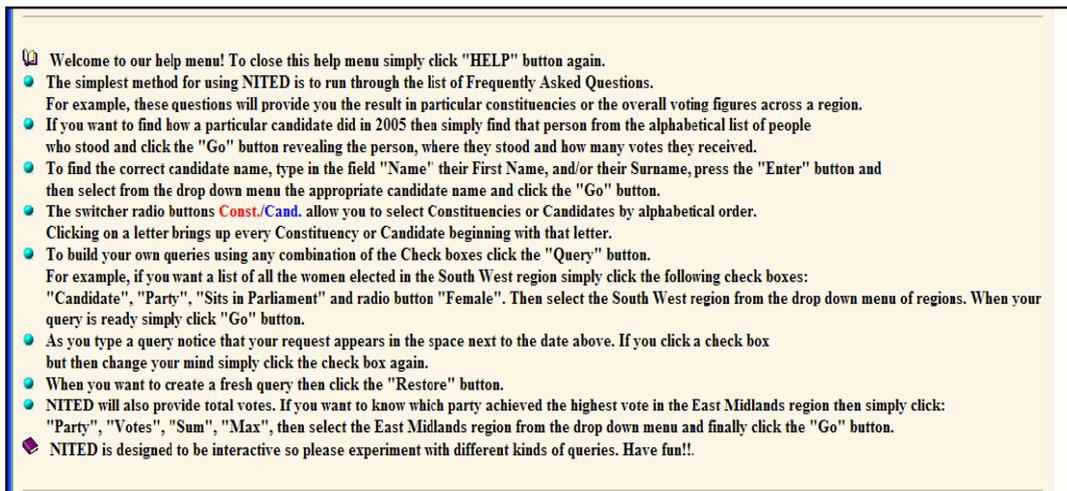


Figure 8.12: Help Instructions for NITED

8.9 Difficulties with Natural Language Enquiries

In theory a natural language enquiry provides end users with the ability to retrieve data from a complex database by asking questions using plain English rather than SQL. It attempts to use natural language processing to understand the nature of the question and then to search and return a subset of the data that contains the answer. But there are problems when using this approach.

Ambiguity management poses a significant problem in natural language systems, because processing all possible natural language meanings leads to exponential, time-consuming computations. The other fundamental problem is that understanding language often means being able to reason as a human being would reason, knowing what concepts a word or phrase stands for and how to link those concepts in a meaningful way, and this problem has proven to be intractable.

Fundamentally, the system's linguistic capabilities are not obvious to the user (Hendrix 1982; Tennant, Ross et al. 1983; Cohen 1991). Existing natural language interfaces can only cope

with limited subsets of natural language. Understandably, users find it difficult to understand (and remember) what kinds of questions the interface can and cannot answer.

Another problem is that the system misunderstands a user request, without the user becoming aware of the misinterpretation. In these cases, it may be hard for the user to understand that the system has actually answered a different question. To avoid such misunderstandings, NITED converted the SQL query back into natural language. The reconstructed natural language request is re-presented to the user but it is debatable how many NITED users actually read these converted SQL requests.

8.10 Interacting with electoral data

The output interface supports users' data exploration by means of three separate types of reporting - traditional tabular layout, bar charts and maps. Different media like maps, graphs and text show different aspects of electoral information and sometimes one type is an inappropriate method of reporting – for example, details about candidates are not easily or intuitively mapped. But where there are opportunities important information should be presented not only by one medium but by a suitable combination of different media. Tables work well when we wish to look up particular values or we need precise values (Figure 8.13).

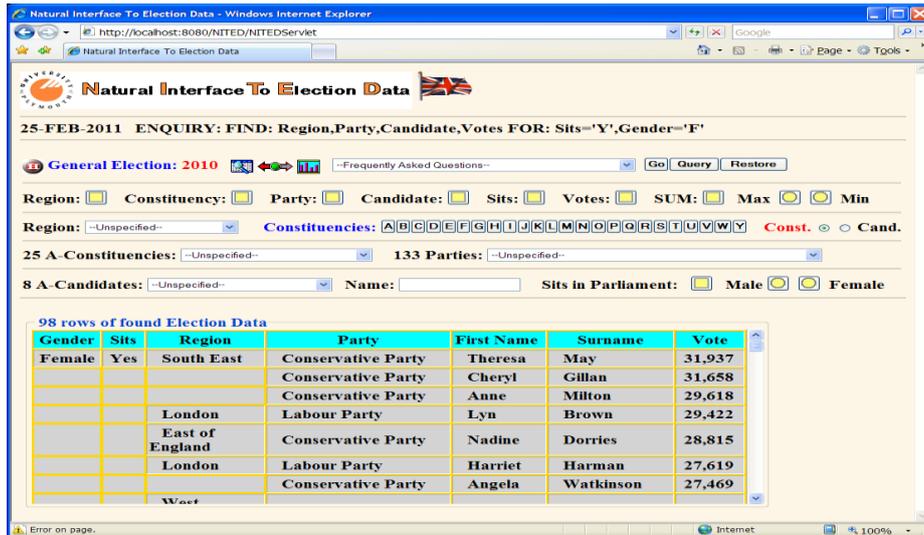


Figure 8.13: Table Output generated from User enquiry

Graphs, by comparison, make meaningful relationships between values visible by giving them size, shape, and colour and draw attention to an important trends, patterns, and exceptions in quantitative data (Figure 8.14). Moreover, graphs encourage interaction between the user and the data, because they highlight interesting features which lead to closer inspection and new insights. Graphics should provide views of the data at many levels of detail. When the user has to move through various levels of specificity of data, a drill down style of navigation is commonly offered. Visualisation should ensure that information is presented with as strong a contextual foundation as possible and offer the user an option to choose what information is displayed in order to satisfy his/her needs. The drill-down clearly groups information of magnitudes ranging anywhere from the “big picture” down to an individual constituency.

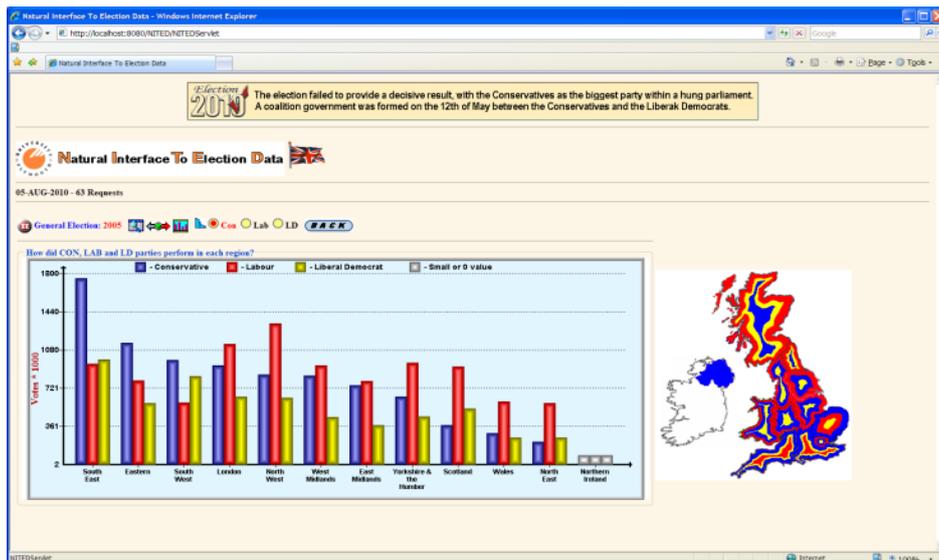


Figure 8.14: Histogram and Mapping Output generated from User enquiry

This project was concerned with not only the creation of new algorithms for the visualisation of election data, but also with an analysis that considers the relationship between the type of inquiry and the corresponding spatial layout of data on the user's screen. When visualisation was used, what users noticed first is the visual part not the interactive. Colours, shapes, regular patterns are strong influences on people's perception of it – the look and feel of the website. However, it is only with interaction that one can really reason about a data domain, explore it under different points of view, and manipulate the medium to accommodate that particular type of question.

Any visualisation requires a minimal set of criteria. Based on Kosara's work (Kosara 2007) the following criteria are a minimal set of requirements for any visualisation:

- It is based on (non-visual) data. The data to be visualised must come from outside the program, and the program must be able (at least in principle) to work on different data sets; in our case two Access relational databases with multiple tables.

- It produces an image. Obviously, visualisation has the goal of producing images from the data, and the visual must be the primary means of communicating the data. Other media can be part of a visualisation, but the visualisation must be able to stand on its own.
- The result is readable and recognizable. A visualisation must produce images that are readable by a viewer. Visualisation images must also be recognizable as such, and not appear to be something else.

Visualisations also have other properties like interaction but while these are certainly important, the above criteria appear to be sufficient to define the requirements of good information visualisation.

NITED, therefore, was able to assess the type of response to each query and then consider the optimum type of user output, whether that be a table, graph or map or some combinations of the three types. Despite the problems that were faced with the natural language interface and the system's ability to 'learn' users' language/syntax structures, it is in the area of visualisation that we think some real progress was made. Graphing some output provided a visually simple but information-rich method for conveying information but it is in the area of mapping data that NITED was innovative and we turn now to explain this particular feature of the experiment.

8.11 Drill Down Mapping

After the 2010 election was over, the country's voting patterns were mapped and broken down by constituency/region. Each result was directly linked to a particular geography, therefore. The interactive mapping enables the user to focus NITED on the level of required information in terms of level of aggregation.

A new visual encoding to represent election results on a choropleth map was developed for this purpose. As mentioned previously (Chapter 4), the mass media in the UK and United States have used three different ways of mapping election results: conventional maps (showing land area), the increasingly used constituency maps (showing one shape per constituency) and the gridded population cartogram - maps which resize the results related to electorate/population distribution.

The new visualisation used within NITED encodes uses conventional maps as a basemap. However, compared to a traditional visualisation of election results (Figure 8.15, a), it attempts to show all three parties' vote share in every areal unit (Figure 8.15, b). In this new map (which only displays shares for the three main parties although all parties in theory could be visualized) the areas are drawn inside the boundaries of each constituency, mirroring those boundaries and ensuring that each area reflects *the shares of the vote* inside each constituency. This technique helps to prevent the addition of visual weight to large area units, and gives non-expert users a chance quickly to identify the high and low values in the data, as well as acquire an idea of trends and patterns in the data.

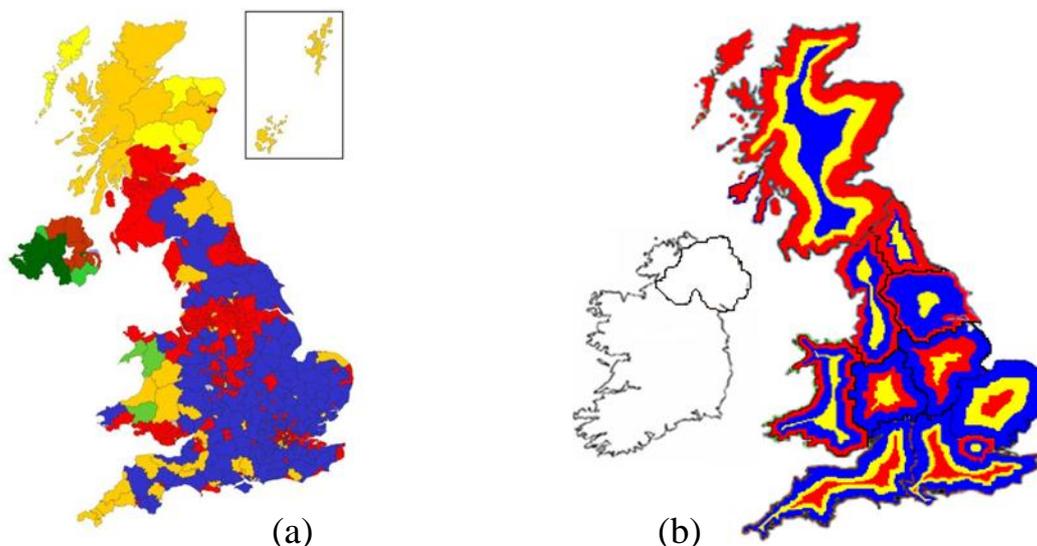


Figure 8.15: Two views of UK General Election 2010 results

The algorithm, which is activated when the user makes a request for constituency-level data, for the creation of a DD-Map is explained using the general election 2010 data for Harrow East, London:

GE 2010: Region: London, Constituency: Harrow East

Results: CON: 21,435 LAB: 18,032 LD: 6,850

Step 1. Calculate the percentage of votes relative to maximum vote:

CON: 21,435 – **100%**

LAB: 18,032 – **84%** = $18032 / 21435 * 100$

LD: 6,850 – **32%** = $6850 / 21435 * 100$ (**38%** = $6850 / 18032 * 100$)

Step 2. Extract the constituency of Harrow East from the London region map using Paint tool. The square of image corresponds to 21,435 (or 100%). We call this the main image (Figure 8.16).



Figure 8.16: Step 2 of the DD map algorithm

Step 3. Using Paint (see the screenshot of Paint on the next page) stretch image to 84% and change the colour to red (Figure 8.17). Now the square of image corresponds to 18,032 (or 84%).



Figure 8.17: Step 3 of the DD map algorithm

Step 4. Using Paint copy and paste the stretched image in the centre of the main image
(Figure 8.18).

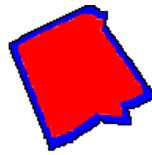


Figure 8.18: Merging Steps 2 & 3

Step 5. Using Paint stretch “blue” image to 32% (or “red” image to 38%) and change colour to yellow. Now the square of image corresponds to 6,850 (or 32% (38%)) - see Figure 8.19.



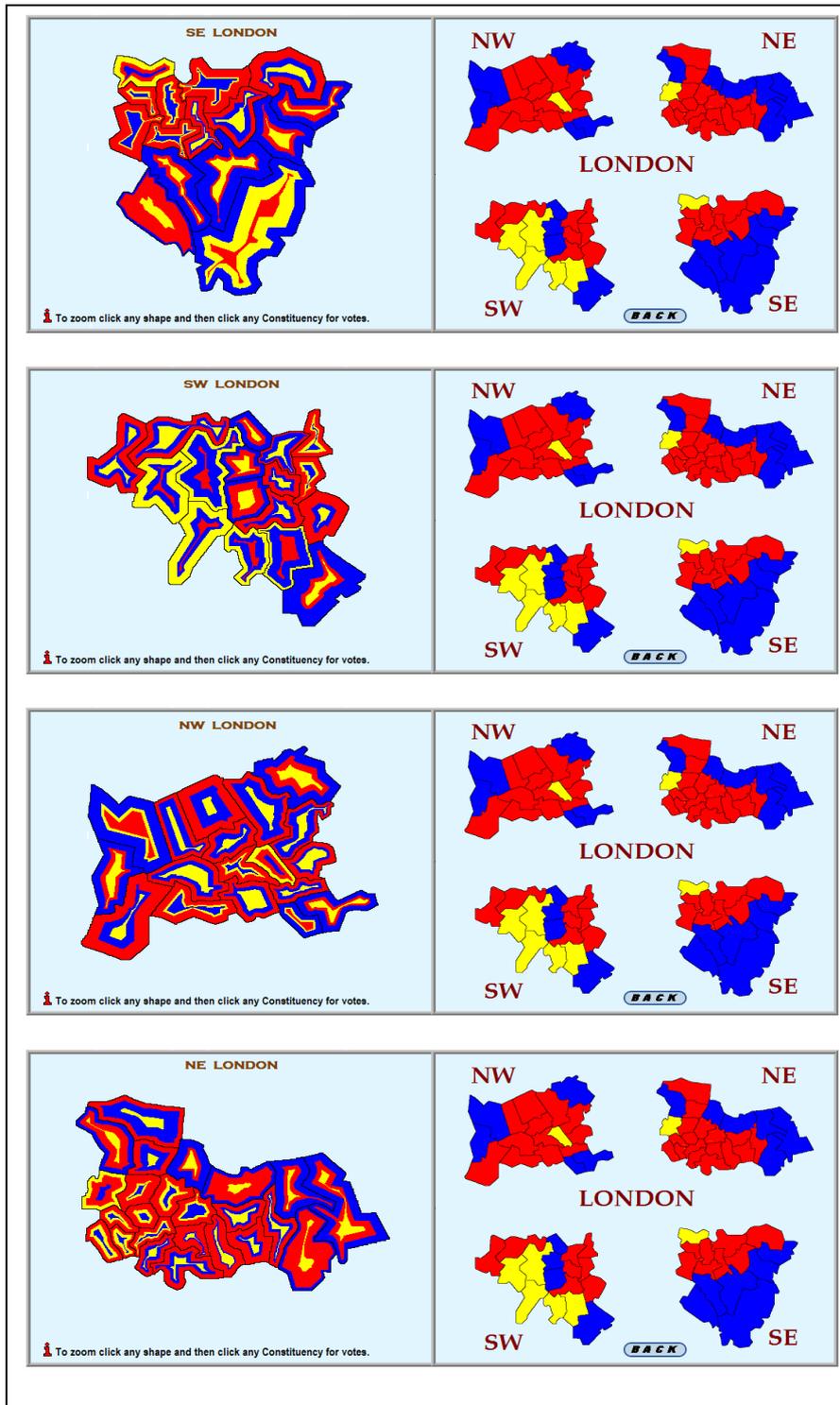
Figure 8.19: Step 5 of the DD map algorithm

Step 6. Using Paint copy and paste the stretched image in the centre of previous main image
(Figure 8.20).



Figure 8.20: The completed DD map for Harrow East

Studying this image the user would immediately realise that Harrow East was won by the Conservative party (outer boundary) with Labour in second-place with the Liberal Democrats placed a long way behind the two main parties challenging in this particular constituency.



**Figure 8.21 Examples of DD-Mapping
General Election 2005**

We feel that DD-Map offers a richer method of visualizing constituency/regional vote shares. The outer boundary is always coloured to reflect the dominant party and so at a glance it is clear which party captured the constituency. Furthermore, by installing the colour of the second-placed party adjacent to the boundary colour it becomes clear which party is the main challenger. Finally, the third colour and the space it occupies at the centre of the map gives a quick indication of whether the seat is a two or three-party battle.

Of course, a conventional set of web GIS tools might be used to produce DD-Map but these tools were unavailable to us for this experiment – it is the visualisation not the software that is most important. Buffering is one of the standard spatial analysis techniques in GIS, it works in Euclidean space and uses a two-dimensional algorithm. The width of the buffer can be specified in one of two ways:

- Fixed distance: Specify a constant distance to apply to the Input Features.
- From field: Choose the name of a numeric distance field from the specified feature class. Each feature will be buffered according to its associated value in the chosen field.

For our purpose, buffering polygon features with the negative buffer distances seemingly might be used, however, is it not clear how the buffer zone could be specified, as it should reflect the share of the vote inside each constituency.

8.12 Conclusions

This chapter has reported an experiment that aimed to create and deliver a Natural Interface to an Election Database (NITED). The objective was to create a web-based application whereby users could interrogate election results stored in a relational database using a combination of point and click mouse actions and critically, a natural language interface. In

theory the latter would allow users to specify their own queries using natural language which the system would then read, interpret and respond with a desired output. To augment this natural language interface we also installed a short menu of frequent queries that would work as a kind of menu from which users could select the menu item and then add some details of their own from drop-down boxes or click boxes.

For a number of reasons NITED was never fully implemented in terms of reaching a public audience although it did go through a form of Beta-testing. It is fair to say that the system encountered a number of problems which are outlined in this chapter. Here, we provide a short summary of those weaknesses.

When users encountered the graphical interface they required a long period of acclimatization in order to understand the basic layout and what the system might be capable of producing. In practical terms, many users would have not stepped past first base and would have clicked away from the application.

The system attempted to fuse different outputs (tables, graphs and maps) but the result was possibly too detailed for what most users would have required. The problem, of course, is that elections data is rather detailed and guessing beforehand what kinds of output any given users preferred would be impossible.

Perhaps the most serious problem affected the natural language interface and its ability to populate new production rules following user-driven queries. Compiling a list of synonyms for words like 'constituency', 'candidate' might appear to be straightforward but there is the additional problem of how words are structured as sentences or even sentence fragments. It was an ambitious plan to develop a system that would allow non-specialists to interact with it but it was a plan that failed to an important degree.

There were also general issues of design. Although there was considerable expertise within the Elections Centre to plan, design and implement NITED it lacked, with hindsight, some essential qualities of web design. When NITED, developed wholly by Elections Centre staff, is compared with some of the applications described in Chapter 10, it should become clear that great design is not simply about function but it is also about appearance. There are some important lessons to be learnt by election specialists from this experiment and maybe the most important is that web users require a simple and attractive interface where it is almost immediately obvious how an application functions and that within a short time the user should be able to interact with that system.

However, the experiment was not entirely a failure. In addition to acquiring new knowledge about what works and what does not in an election-centred application we also developed what we find to be an intuitively strong method for visualizing three-party shares using drill-down mapping. User responses to these map visualisations were positive and this is something that could be developed more fully using GIS software in the future.

9 Political Science meets Broadcasters

9.1 Introduction

This chapter examines collaborations between the Elections Centre, Plymouth University and Sky News broadcasting. Michael Thrasher has been the Sky News ‘on-air’ elections commentator since May 1989. Over the past ten years the Elections Centre has been active in helping to expand and improve the broadcaster’s web-based applications, some of which have subsequently been used in live broadcasting. The people working at Sky on these developments has changed over the years but for most of the developments described here it was Hugh Westbrook, whose job description of multi-platform digital producer helps to describe the increasing convergence between internet and broadcast activities. The Elections Centre input into this process has centred on supplying data but also in developing ideas about how interactivity can be introduced to the data query process.

The aim of this chapter is to show how for broadcasters such as Sky and the BBC there is a growing link between what is made available to viewers on their television screens and what is available for them to use interactively on the web. Data on audience viewing suggest that for younger age groups there is relatively less television viewing and more active engagement with web-based applications. This trend is likely to increase in the future therefore and broadcasters are likely to spend more resources in addressing this kind of viewing audience. A second aim of this chapter, however, is to demonstrate that sometimes there is a trade-off

between interactivity and usability. This is particularly the case in respect of some of the applications that have been developed as collaborations between Plymouth and Sky News. We show that for some applications the level of difficulty in understanding the nature of the interactive element was rather straightforward but for others of a more complex nature it proved too difficult to build an application.

This chapter is divided into five main sections. These sections follow specific applications and are as follows:

- Poll tracker – a system designed to allow web users to view the trends in party political support and to highlight individual polls. This was first implemented in 2008 but used data from 1983 onwards.
- Parliamentary history – a visualisation of the historical development of parliamentary elections from 1832 onwards using cartograms.
- An interactive guide that allowed users to indicate policy preferences which were compared with party manifestos. It was entitled “How do I Vote?”
- A ‘ready reckoner’ that converted either the user’s own estimation or poll ratings of likely 2010 general election national vote shares and then converted these to parliamentary seats applying the principle of uniform swing.
- An interactive application that permitted users to estimate the likely impact of the adoption of the Alternative Vote at the May 2011 referendum.

These sections are ordered in more or less chronological order but it happens to be the case that they are also ordered in terms of degree of difficulty and complexity. The first application, poll tracker, was a rather simple but effective way of visualising national polling data in the months leading up to the 2010 general election. The second and third examples were again provided before the 2010 general election and can be placed more towards the end of the visualisation scale and away from the interaction end of the continuum. The final two applications, however, were much more ‘hands-on’ and were intended so that users could do some election analysis of their own. However, the cost of fully engaging with the application

meant the user had to learn a great deal about how to operate the application and would need quite an extensive understanding of how the voting system worked before they could use it. We conclude that the introduction of Poll Tracker was regarded as a success while parliamentary maps, 'Who do I Vote' and the 'ready reckoner' were believed to have made a good contribution to broadcast coverage as far as Sky News was concerned. However the Alternative Vote application, which had been the most difficult and complex to implement as viewed as a failure.

9.2 Poll tracker

The aim of the poll tracker was to give a snapshot of how opinion polls were rating the main political parties since 1983. This starting figure was chosen simply because that is when the Election Centre's data holdings began. Once it was implemented I became responsible for obtaining each day's polling data and sending reports to developers at Sky News who would up-date the application. Because of security issues and firewalls it was impractical to automate up-date procedures although this would have been quicker.

The web application contained various help menus designed to explain how the process worked in practice. The general guidance read as follows:

'Our monthly chart shows you the monthly performance of all the parties at a glance, with the figure specially calculated by Sky's election expert Michael Thrasher. He looks at every poll, every month, and calculates the average for an overall result. Choose beginning and end dates over any period to see how trends have changed, and click on the buttons to see key events which have influenced public sentiment. If you want more detail, the recent polls button will

show you details of every single poll taken since 1993. You can get there by clicking the button or simply clicking on the poll tracker timeline.’

For any broadcaster or newspaper it is important that they have some kind of ‘branding’ that distinguishes their particular version apart from all the others. For the poll tracker this ‘exclusivity’ came from the method for calculating the weighted monthly average. Again this was explained to users who were interested in discovering how the numbers were being calculated:

‘Sky News calculates a monthly weighted average of voting intention from polls conducted by the country’s leading opinion survey companies. Pollsters’ methods vary, some continue to do face-to-face interviews while others conduct all their polls by telephone. In recent years there has been a growth in internet polling. Even among those that use the same method of surveying there are also differences in how the final voting intention figures are compiled. At least one pollster’s headline voting figures are based only on respondents that are certain to vote at a future general election. Some pollsters weight their answers according to past voting or some other variable. Because of these differences in conducting and compiling voting intention figures Sky News decided its monthly average should be weighted’.

In practice we had arrived at this method of performing the weighted monthly averages after running a whole series of experiments to assess the effect of using different methodologies. A number of these methods would have worked but one of the reasons for selecting this particular method was that it provided some ‘value added’ aspect – the procedure was too complex for users to do it themselves and was therefore different to any method that simply took averages.

Figure 9.1 shows the final screen version of the web application. Pressing the “info” button took users to the set of pages explaining how the figures were calculated. The logo located on the top left was part of the original design but this was subsequently altered to reflect Sky News’ branding for the 2010 general election.

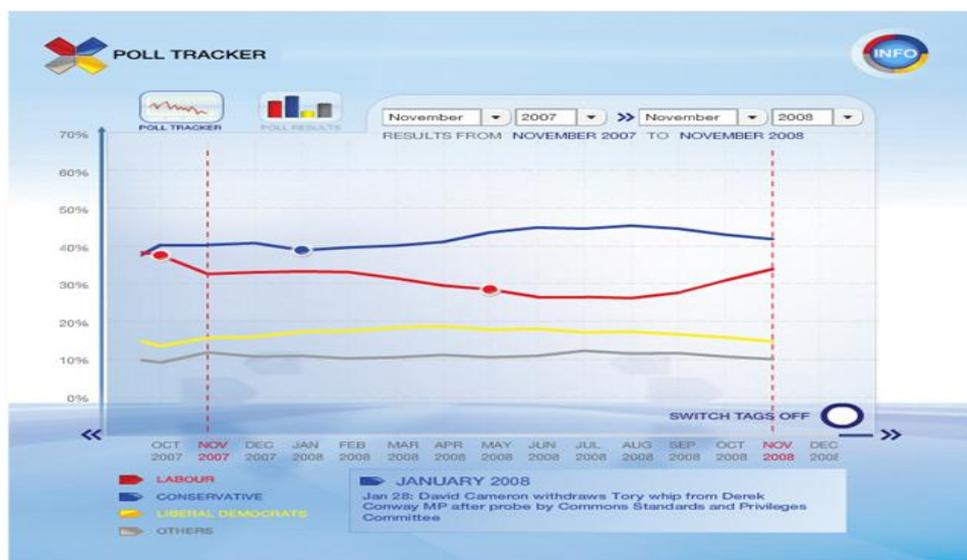


Figure 9.1: Poll Tracker for Sky News³⁸

Users could select between the line graph as shown above or histograms show later Figure 9.2. The timeline screen shown here covers the period between 2007-2008 but users could view the overall trend back to the mid 1980s. The timeline could be adjusted either by completing the drop-down boxes for start and finish times (the greater the time difference the more concentrated the time period) or simply by dragging the line graph either left or right according to whether the user wanted to move backwards or forwards in time. This method retained the scaling but would not allow the user to see the big picture in a single glance.

The line colours represented the three main parties (with Liberal Democrats in yellow and other parties combined into one single grey line). Since these were national polling figures

³⁸ <http://news.sky.com/skynews/Interactive-Graphics/Poll-Tracker>

all data excluded Northern Ireland because the polling companies do not survey there. At some critical time points there were also brief summaries that related to politics generally or, as is the case in the example chosen here, to a specific party (Conservative leader David Cameron suspended the whip from one of his own MPs). These short descriptions could be switched on or off by selecting the ‘Switch Tags Off’ button.

As well as the line graph feature the user could instead opt to choose to view specific polls that had been conducted by a particular polling company or those conducted on a particular date. By pressing the button next to the line graph button this facility was switched on or off as desired.



Figure 9.2: Poll Tracker using histograms to show single poll results

Figure 9.2 illustrates how this worked. Here, the user has selected a particular date during the election campaign where a number of pollsters have published polls. If a date with only a single poll appeared then that would tell the user there were no other polls on that date. If no date appear at all then that would indicate no published polls were available. For each poll chosen the histogram would show the percentage share for each of the four party categories.

When this application was being developed there was a great deal of discussion about the different polling methodologies that were being used by the separate polling companies. While the previous Figure shows the result for YouGov, for example, the following picture Figure 9.3 presents the data for ICM which was still publishing polls for The Guardian newspaper and using telephone polling rather than online surveys.



Figure 9.3: Poll Tracker and polls produced by separate polling companies

Poll tracker was a very simple and yet highly effective method for viewing polling results. It allowed users to view single polls, to compare separate polls published at about the same time and permitted users to select the time period for doing so. In its line graph form it presented a quick and easy visualisation of any trend that was (or was not) developing and also what the long-term trend had been over several decades. It was very simple to use and internally the broadcaster was pleased with the outcome, reporting that audience feedback was generally good. Unfortunately, despite this success it was discontinued immediately after the 2010 general election.

9.3 Broadening the Sky News Elections Data Coverage

As part of its broad approach to the forthcoming general election that Sky News began in early 2009 to plan its web-based presence and how this might be successfully incorporated into its broadcast output. Hugh Westbrook and Michael Thrasher held a number of meetings that would establish what could be planned and implemented within a two-year period.

Following this, the website should develop a comprehensive approach to general election statistics. This can be usefully divided into data that are historical and those that relate specifically to the next general election. As one of the co-authors of *British Electoral Facts* Thrasher (Rallings and Thrasher 2007) was keen to persuade Sky to host much of these data. From its point of view Sky's Hugh Westbrook was more interested in visual rather than tabular displays of data. It was a natural development therefore for Sky to push for some kind of historical mapping of UK general elections.

9.3.1 Historical data

It was proposed that following the success of Sky's coverage of the 2008 US Presidential election that something similar should be done for UK general elections. These data would be made available to Sky in electronic format but would need compiling into a database that could be used by both the website and broadcast computing.

During the planning phase it was proposed that there would be comprehensive coverage of each general election since modern elections began in 1832. Users would be offered the selection from UK (incorporating results from Ireland and Northern Ireland) or Great Britain (opinion polls cover only GB so this comparison became relevant to integration with Poll tracker).

These pages would show for the main parties, votes, vote share, seats, seats share and winning party. It could also show a portrait/photograph of eventual Prime Minister, e.g. Blair after 1997. Summary data could also be shown for each country – England, Scotland, Wales, Ireland, Northern Ireland. Potentially, summary data could also be shown at the regional level using Government Office Regions. However, these data exist only from 1992 because of changes in the definition of regions.

In addition to the overall voting figures the website might offer other kinds of information that users of all types would find interesting. These include, numbers of registered voters, numbers turning out to vote, post-war swings, the number of seats that changed hands at each election, the weather on polling day.

One of the key aspects of this proposed development from the broadcaster's point of view was that the data would be visual. Although Plymouth had not developed the required level of expertise in GIS to achieve this there were connections with geographers based at the University of Sheffield. Following an agreement with the research team led by Professor Daniel Dorling it was decided that Sheffield would produce a series of cartograms for each separate election since 1832 and that this would be animated to cycle through the elections. Maps – a team from the University of Sheffield will grant permission for Sky to use an animated cartogram. Each map would portray individual constituencies and the winning party at each election since 1832.

9.3.2 Historical Cartograms

The original planning for each separate election was ambitious. Some indication of the proposed scale of the development can be assessed by the following description of how the

data sets would be built. Data would be divided into “zones” that reflected different levels of aggregation.

Zone 1:

Year
Majority_Party
Overall_Majority
Total_Electorate
Turnout_percent
Total_Votes
Total_MPs
Total_Female_Members
Electorate_Uncontested_Seats
Percent_of_Total

Zone 2:

Con_MPs
Con_Members_Percent
Con_Votes
Con_Percent
Con_Female_Members
Con_Ethnic_Members

Zone 3:

Lab_MPs
Lab_Members_Percent
Lab_Votes
Lab_Percent
Lab_Female_Members
Lab_Ethnic_Members

Zone 4:

Lib_MPs
Lib_Members_Percent
Lib_Votes
Lib_Percent
Lib_Female_Members
Lib_Ethnic_Members

Zone 5:

Oth_MPs
Oth_Members_Percent
Oth_Votes
Oth_Percent
Oth_Female_Members
Oth_Ethnic_Members

Zone 6:

PrimeMinister_Name
PrimeMinister_Biography (maybe a button to bring up a small pop-up window)
Con_party_leader
Whig_liberal_party_leader
Labour_party_leader

Zone 7:

Constituency (headed first to declare)
Declaration time
Weather

As with Poll tracker it was determined that the main screen would also benefit from a button called ‘About Cartograms’ which would bring up a separate window explaining how this mapping procedure works (each constituency is scaled to be the same size and the standard physical outline of the UK is replaced by one that closely resembles it but is more concerned with more accurately representing the relative number of seats gained by each party) and other information about how the data were compiled and managed within the application.

After the initial planning it was decided that users might want to have the facility for comparing the outcome of two separate general elections. It was not felt that many users would want to compare maps showing other than the overall distribution of seats.

Accordingly it was decided to use the following data:

- Year
- Majority Party
- Overall Majority
- Con_MPs
- Lab_MPs
- Lib_MPs
- Oth_MPs

Gradually, during 2009 it became clear that the ambitious scale of the historical data development would mean problems in meeting deadlines. It was decided to abandon some of the “zones” outlined earlier, leaving the core data intact. This would mean that the only data that would sit behind the historical cartograms would be the seats won by the different parties and the outcome of the general election itself which would include the name of the new Prime Minister and the governing party’s majority in the House of Commons.

Figure 9.4 shows the structure of the data that was sent from Plymouth to the University of Sheffield and then on to Sky Broadcasting. Each data row records information for a single general election. These data would be used to build individual cartograms which would be stitched together using Flash Player software.

Earl Grey											
	A	BS	BT	BU	BV	BW	BX	BY	BZ	CA	CB
1	Year	LibMembers	LibMemb%	OthMembers	OthMemb%	Total	Majority Party	Overall Majority	party leader con	party leader whigliber	party leader lab
2	1832		441	67.0	42	6.4	658 Lib		225 Duke of Wellington	Earl Grey	
3	1835		385	58.5			658 Lib		113 Sir Robert Peel	Viscount Melbourne	
4	1837		344	52.3			658 Lib		29 Sir Robert Peel	Viscount Melbourne	
5	1841		271	41.2	20	3.0	658 Con		77 Sir Robert Peel	Viscount Melbourne	
6	1847		292	44.5	39	6.0	656 None		0 Earl of Derby	Lord John Russell	
7	1852		324	49.5			654 Con		7 Earl of Derby	Lord John Russell	
8	1857		377	57.6	13	2.0	654 Lib		100 Earl of Derby	Lord Palmerston	
9	1859		356	54.4			654 Lib		59 Earl of Derby	Lord Palmerston	
10	1865		369	56.1			658 Lib		81 Earl of Derby	Lord Palmerston	
11	1868		387	58.8			658 Lib		115 Benjamin Disraeli	William Ewart Gladstone	
12	1874		242	37.1	60	9.2	652 Con		49 Benjamin Disraeli	William Ewart Gladstone	
13	1880		352	54.0	63	9.7	652 Lib		51 Earl of Beaconsfield	Earl of Granville	
14	1885		319	47.6	102	15.2	670 None		0 Marquis of Salisbury	William Ewart Gladstone	
15	1886		192	28.7	85	12.6	670 Con		116 Marquis of Salisbury	William Ewart Gladstone	
16	1892		272	40.6	85	12.7	670 None		0 Marquis of Salisbury	William Ewart Gladstone	
17	1895		177	26.4	82	12.2	670 Con		153 Marquis of Salisbury	Earl of Roseberry	
18	1900		183	27.3	83	12.4	670 Con		135 Marquis of Salisbury	Sir Henry Cambell-Bannerman	
19	1906		399	59.6	86	12.8	670 Lib		129 Arthur Balfour	Sir Henry Cambell-Bannerman	
20	1910(J)		274	40.9	84	12.5	670 None		0 Arthur Balfour	Herbert Asquith	Arthur Henderson
21	1910(D)		272	40.6	85	12.7	670 None		0 Arthur Balfour	Herbert Asquith	George Barnes
22	1918		163	23.1	105	14.8	707 Coalition		283 Andrew Bonar Law	Herbert Asquith	William Adamson

Figure 9.4: Raw Data for Historical Cartograms

The completed application appeared in the first months of 2010. Figure 9.5 shows the outcome of the 1832 election won by the Liberals with a landslide majority. It is a classic presentation of map and tabular data that many people will be familiar with. The map retains a shaded outline of the physical boundaries of the UK and overlaid with circles, one for each House of Commons constituency. In this version of the map it was not possible to hover the mouse over a particular circle and then drill down into the data for that constituency. The tabulated data on the left hand side shows the frequencies for seats won by the main parties and ‘other parties’. An early decision was taken in terms of how the names of parties would be managed. Strictly speaking, these were seats won by the Liberal Party, for example, but the table records them as having been won by the Liberal Democrats, a party that would not come into existence until 1988. It was felt that it was more important to be consistent than to

be strictly accurate. Additionally, for those users that might be annoyed by this method of coping with name changes, there was information that lay behind the screen which would explain how the data were being managed.

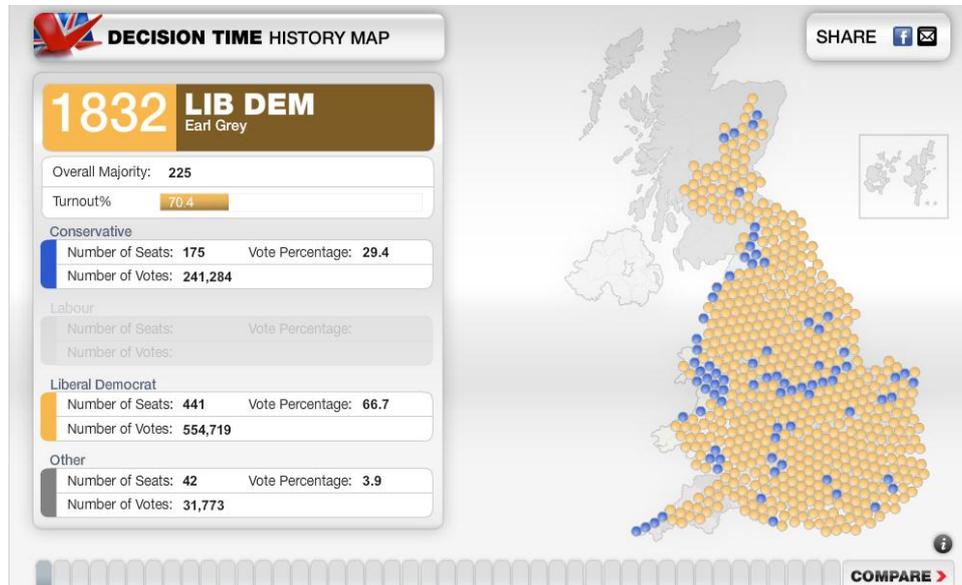


Figure 9.5: Cartogram of the 1832 General Election³⁹

One of the visual strengths of mapping electoral data is that users can very easily gauge patterns in the data. For example, Figure 9.6 shows how the map had changed by the end of the nineteenth century as the country began to divide into a Conservative south of England and Liberals dominant in other parts of the country. But this map also demonstrates a weakness in this method of viewing data. It is clear that the number of seats won by others at the 1880 election is 63 seats – but these others do not appear anywhere on the cartogram. Once again, this can be explained by the need to sacrifice historical accuracy in order to make a web-based application that works relatively seamlessly. At the time of the 1880 election there were still 101 constituencies that were allocated to Ireland, 63 of which were won by Home Rule candidates in 1880. But because Ireland was no longer part of UK general

³⁹ <http://interactive.news.sky.com/2010/election/History/>

elections after Ireland gained independent in the early 1920s the decision had been taken to ignore it for the whole series of historical maps going back to 1832.

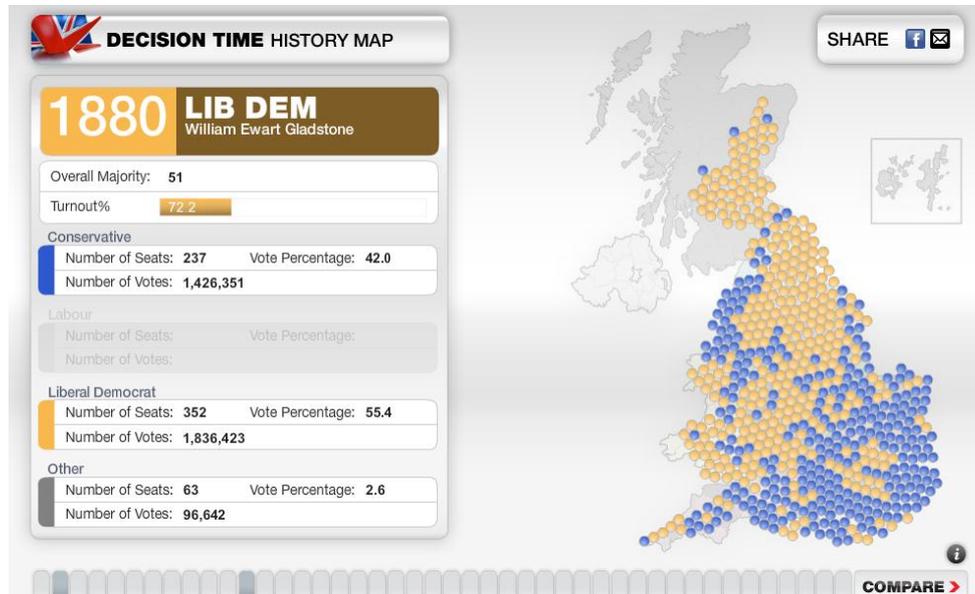


Figure 9.6: Cartogram of the 1880 General Election

It was felt, however, that while decisions like that regarding whether or not to include data from Ireland would upset the purists most users of the web application would use it to examine in more detail the distribution of seats at recent general elections. Figure 9.7, for example, shows the outcome of the 2005 general election. One of the interesting features of this map and table is that it shows very simply and clearly how Tony Blair's Labour government won many more seats than the Conservatives won even though the two parties were very close together in terms of vote shares. The map also clearly shows where Labour's support was becoming concentrated as it retreated from constituencies captured at the 1997 general election.

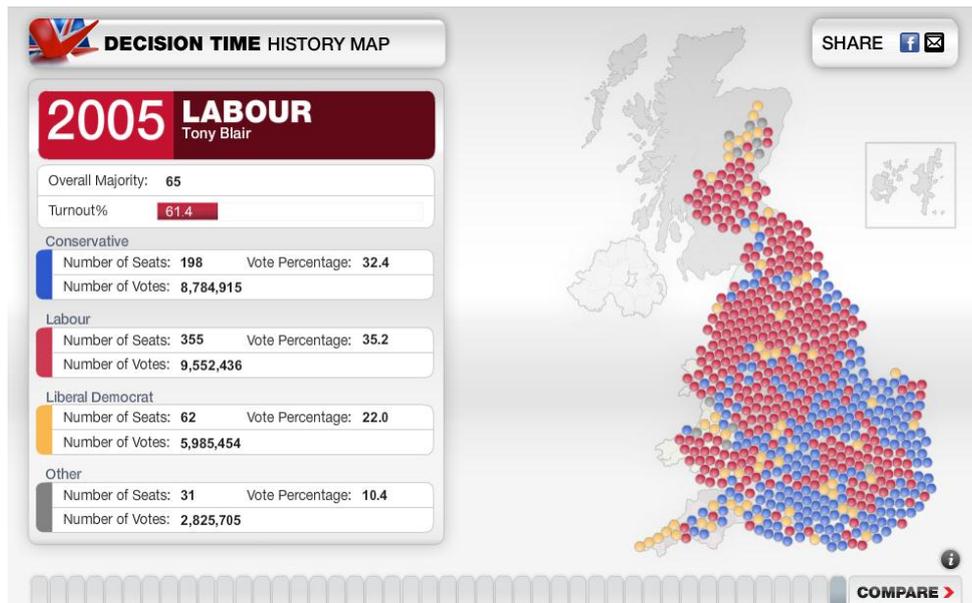


Figure 9.7: Cartogram of the 2005 General Election

It was always felt that the real strength of cartographic mapping is that because each constituency is equally scaled (as opposed to taking its physical boundaries which gave too much emphasis to rural seats and under-estimated the number of densely packed urban seats) that users would be able to visualise change in party fortunes quite easily. The Flash version of the cartograms cycled through the elections from 1832 in a way that showed the movement between blue and yellow and then later the emergence of Labour (red) in the 1920s as it began to replace the Liberals. It was decided to give users an even greater control by permitting them to compare the outcome of two election maps alongside one another.

Figure 9.8 proves a very good example of this. The application has grey-scaled bars at the bottom of the screen and when the mouse hovered over these a general election year (e.g. 1979) would appear. The left hand side (to the left of the button marked “Back” would select the election to be shown on the left hand part of the screen and vice versa for the right hand side. This screen grab shows the user has compared the 1992 and 1997 general elections. The two maps occupy the extreme left and right hand sides of the screen while the tabular

data, located in the middle of the screen, is left and right indented. The colour coding indicates the 1992 election was won by the Conservatives with a majority of 21 while the following election was won by Labour, led by Tony Blair who won a majority of 178 seats. Although it is possible to study the numbers closely (the Conservatives fell from 336 seats to just 165 seats and lost almost five million votes) the most immediate impression of change is given by the cartograms where a short flick of the eyes will see the map change from mostly blue to overwhelmingly red. A slightly longer viewing is required to spot the increase in Liberal Democrat seats, however. While their seats more than doubled, from 20 to 46 the yellow colour does not make that immediately obvious.

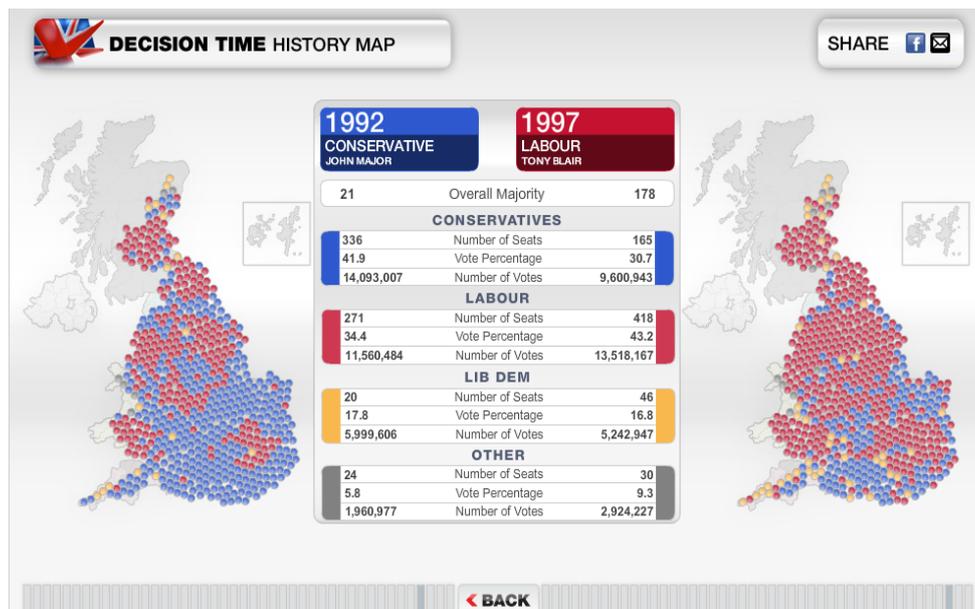


Figure 9.8: Comparing the 1992 and 1997 General Elections

9.4 Interpreting general election manifestos at the 2010 general election

It is well known that very few people read the manifestos produced by each political party prior to a general election. In fact, it has been shown that many people vote for parties despite them being in disagreement with the party's policy position – many Liberal Democrat voters, for example, are unaware that the party is very pro-European despite much of the electorate being Euro-sceptic. As part of the pre-election planning leading up to the 2010 general election it was decided to develop a web application that would allow the user to answer a series of questions about different aspects of policy and then dependent upon the answers the application would suggest which party was closest to the user's own view and perhaps which party to vote for. It was intended to be informative but fun at the same time.

It was agreed that only the three main party manifestos would be covered and that there would not be a separate application that would cover Scotland, Wales and Northern Ireland. Because the rules that apply in relation to broadcasting do not apply in the same way to internet-based material this was felt to be okay.

When the manifestos were published the original decision was about the number of categories that would be covered. After reading the three documents a total of ten separate categories were constructed. For each of these categories the user would see a policy summary taken from each manifesto although obviously these quotes were anonymised so the user could not identify which manifesto it had come from.

The overall look of the application is shown in Figure 9.9 with the application going under the title "who should I vote for". This was not an original formulation with similar kinds of exercise being conducted at previous elections and other websites were developing similar applications for the 2010 general election. The user was asked simply to identify the party he

or she would normally support at election time. After doing that the application would walk them through the procedures with on-screen instructions available.

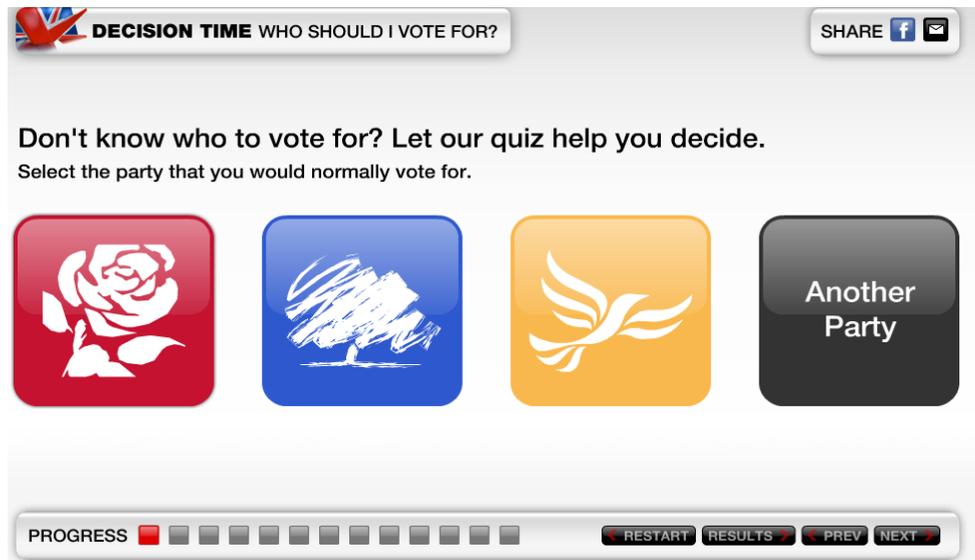


Figure 9.9: Opening screen of the ‘Who Should I Vote For?’ application⁴⁰

The initial procedure was to establish for each voter the policy hierarchy – what policies were important and which ones were less important (Figure 9.10). The algorithm that sat behind the application would weight each person’s response according to whether the policy was critical or not at all important etc. So, for example, a user that felt tax and spending to be critical would drag those into the box marked “critical” while if education was not at all important it could be dragged to that box. There was no forcing users to restrict the number of policies that could be placed in each box, so all policies might be ‘critical’, for example.

⁴⁰ <http://interactive.news.sky.com/2010/election/Manifesto/>

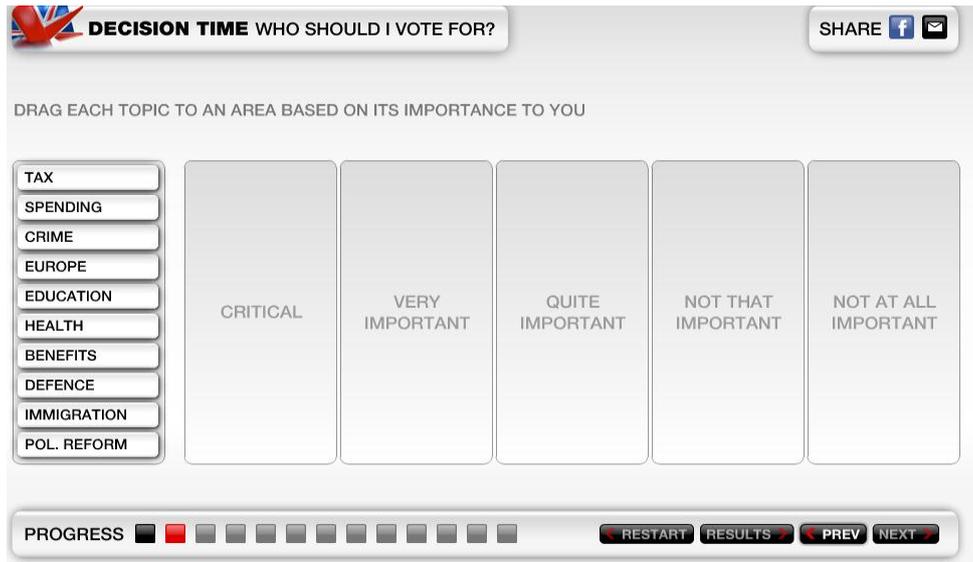


Figure 9.10: Policy Choices and Level of Importance

Once the allocation of importance had been given to each policy area the application would then cycle through the policy choices in the order as they appeared on the original screen shown in Figure 9.10. The example shown here in Figure 9.11 relates to the three party manifestos and a key statement relating to tax policy. When the user had chosen the statement that was closest to their own thinking on tax then they would vote for that choice by clicking the box to the right of the statement.

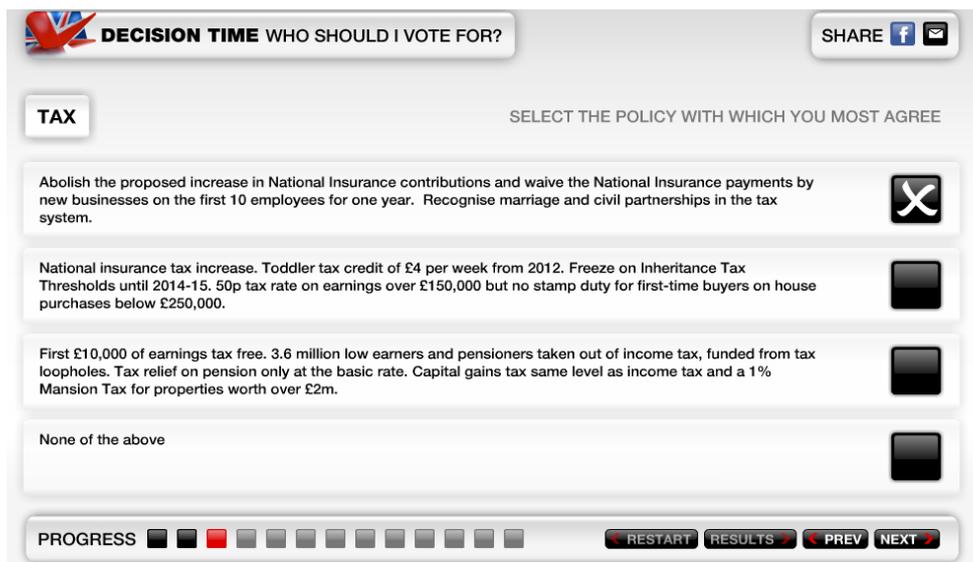


Figure 9.11: Manifesto Options on Tax Policies

This process would be repeated for the remaining policy options. At all times the user was offered the choice of “none of the above” in addition to summaries of the policy positions being preferred in the party manifestos. If the user changed their minds and wanted to go backwards then that option was given and equally they could move forward in the application without answering any of the remaining policy questions. One of the weaknesses of some online polling questionnaires, for example, is that users are forced into completing a screen of questions before being allowed to progress onto the next screen. Since this was meant to be informative and entertaining it was decided not to implement any restrictions on how much interaction was required on the user’s part.

The final screen shot in this sequence (Figure 9.12) delivers the verdict on how to vote. This particular user has selected mostly policies that were contained in the Labour manifesto but also some that were taken from the other parties’ manifestos. The user could identify at this point which of their selections agreed with Labour and which with the Conservatives or the Liberal Democrats. The size of each bar reflected the importance that the user had originally attached to the policy. A long bar would rank as ‘critical’ with the shortest bar allocated to the category ‘not at all important’. Since all of the parties made their manifestos as free to view pdf files a button was added to the application that would take them directly to the party manifesto website.

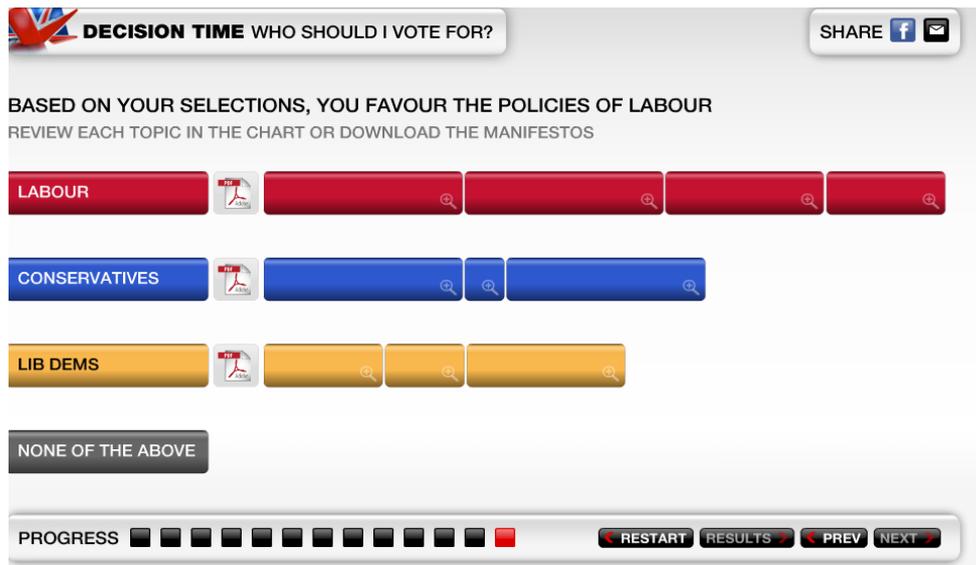


Figure 9.12: Outcome of User choices

Although this application worked well and its implementation was successful it was subsequently felt that some users might become bored by having to think about so many policy categories and to read summaries of the manifestos relating to each. But measured in terms of ‘ease of use’ the application was well designed and intuitive.

9.5 House of Commons projections

When the television viewing public first encountered election analysts there was a fascination with how it was possible to take a handful of results from early declaring parliamentary constituencies and use those data to make projections of the likely composition of the next House of Commons. Of course, there is nothing mysterious about applying change in vote share to each constituency and then re-calculating how many seats are won by the various parties but the myth about projecting seats has survived.

Prior to the 2010 general election it was decided that it was time for the ability to make House of Commons forecasts to be given to the general public (see a general discussion of

this in Chapter 4). The Elections Centre supplied a simple spreadsheet containing the estimated results for the 2005 general election. Estimated results were required because the 2010 general election was being fought on new parliamentary constituency boundaries. The spreadsheet contained vote shares for each constituency. The idea was that users would select a particular configuration of national vote shares and then information about what that meant for change in vote share would be calculated from that. For example, if it was projected that Labour would get 30% of the vote that would assume a six percentage-point drop in Labour share. In the projection calculation Labour's vote share would then be reduced in every single constituency (apart from those in Northern Ireland which Labour does not contest). The same process would be applied to the Conservative, Liberal Democrat and Nationalist (Plaid Cymru and Scottish National Party) shares. Following these calculations the new vote shares would be used to calculate winners in each constituency and then total seats would be provided. The process is reasonably straightforward, therefore, but the challenge lay in building an application that non-expert users could learn quickly and apply without too much trouble.

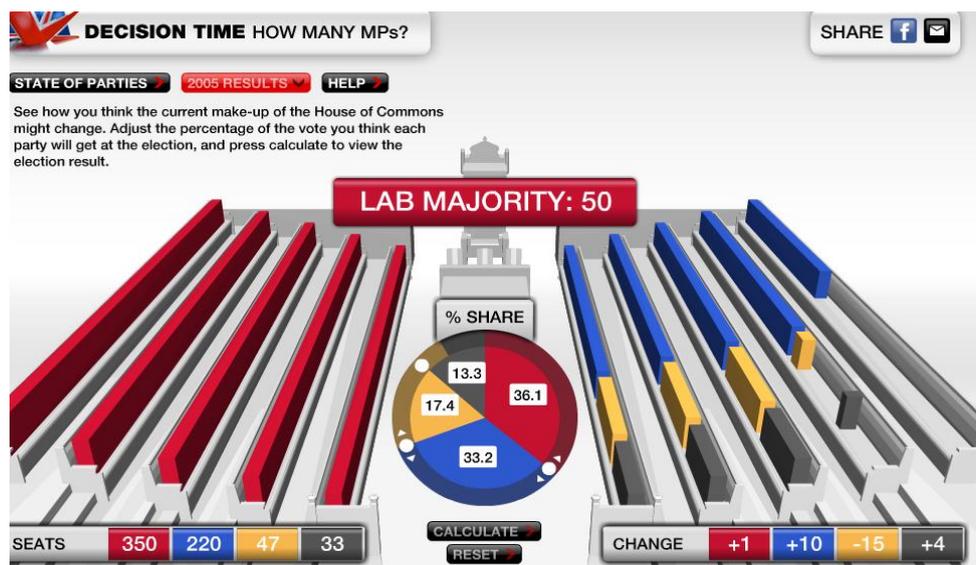


Figure 9.13: State of the Parties showing Labour Majority

Various prototypes were developed and rejected because the user would be required to enter data and for those numbers to sum to 100%. The original design was based on a Table that showed percentage national vote gained by each party and users were invited to plug in either their own estimate of 2010 vote share or the results from a recent public opinion poll. When we tested this it was immediately obvious that many people cannot count to 100. We then added a box to the screen that would progressively total the numbers as they were added. When the total passed 100 then the colour would change to try and alert the user that they had tried to add numbers that were too high. It was decided to abandon any method that involved users inputting their own numbers and expecting them to know what the total was and either add or subtract numbers until the total was 100.

Figure 9.13 clearly shows at the centre the solution that was finally made. Instead of allowing users to add/subtract numbers and expect this process to be zero-sum (for every increase there has to be an equal and opposite decrease) the pie chart principle does that automatically. The percentages are constrained within the circle. If someone adjusts the Labour vote downwards by six points then the Conservative vote would increase by the same amount.

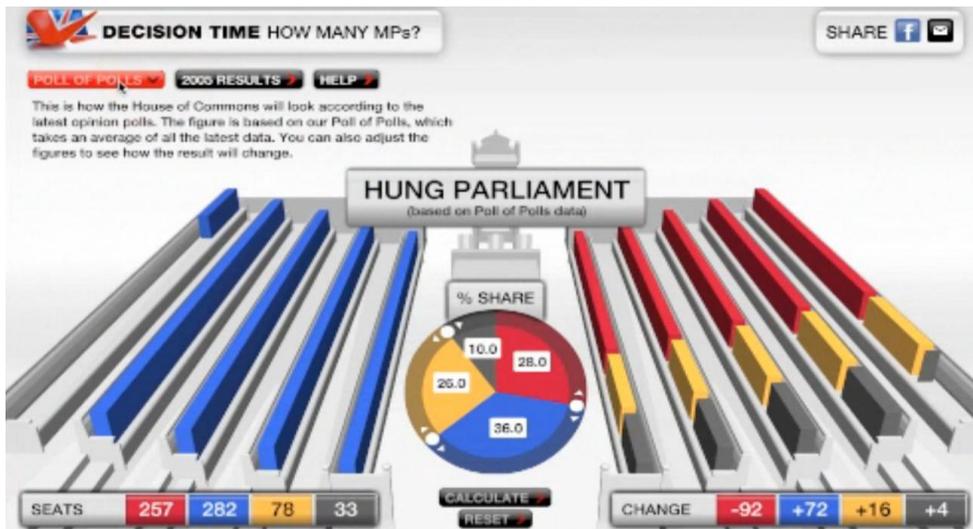


Figure 9.14: State of the Parties showing Hung Parliament

Figure 9.14 shows a slightly different configuration of vote share has been selected producing a hung parliament as the result. Note that the colour of the benches reflects the different outcomes. In Figure 9.13 the benches on the left, the government benches, are shaded red to denote a Labour majority of 50. In Figure 9.14 the left hand side benches are blue simply because the Conservatives would be the largest party in the hung parliament.

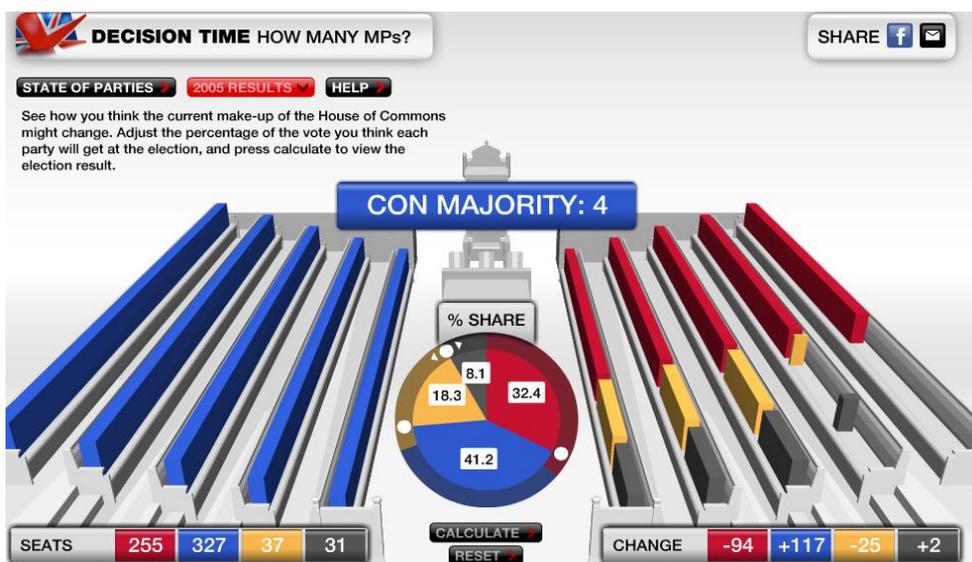


Figure 9.15: State of the Parties showing Conservative Majority

During the 2010 general election campaign there were many discussions about not only the possibility of a hung parliament but also why and how the electoral system appeared biased towards the Labour party. This was partly examined in Chapter 6 but with this graphic users were able to manipulate the vote shares to the point where they could see precisely what kind of lead the Conservatives would require in order to win a narrow majority in the House of Commons. Figure 9.15 shows the Conservatives have been assumed to win 41.2% of the national vote with Labour on just 32.4 – a Conservative lead of nine percentage points. Assuming uniform swing across every constituency this would have delivered an additional 117 seats for the Conservatives with 94 losses for Labour leaving the Conservatives with an overall majority of just four seats.

Various aspects of this web application borrowed from previous applications built by the collaboration of Plymouth and Sky News broadcasting. There were explanations and help buttons. Background shading was used to suggest that this was the House of Commons chamber. The party colours were those used in previous applications. Some graphical interfaces use “people” but the Sky designers preferred clean lines and edges to their graphics – the user understood implicitly that the length of the rectangle indicated the number of MPs. The major flaw in the design lay with the user’s ability to manipulate the vote shares. A simple example illustrates the problem. The small handles (white dots on the outer edge of the circle) were to be moved in one direction to increase/decrease party vote shares. But suppose the user wanted only to vary the vote shares of just Labour and the Liberal Democrats. Because these segments have no common boundary the user had first to adjust the Labour boundary (let’s assume it is a decrease) by grabbing the handle between Conservative and Labour or Others and Labour so that the Labour vote decrease by six points. But this would increase the Conservative/Other vote by six points also. The user

would then need to adjust another handle, this time giving the Liberal Democrats a six point increase (either from the conservatives or from other parties. A considerable amount of time and re-design went into trying to solve this rather small problem but no solution was found for this particular web application. But this example does show that very minor design issues (which most people would never notice) do pre-occupy those people responsible for delivering the final version.

9.6 The Alternative Vote Referendum

Shortly after the general election in 2010 and the formation of a Conservative/Liberal Democrat coalition it became clear that there was planned a number of constitutional reforms including changes in the House of Lords and a national referendum on a new voting system for the House of Commons, the Alternative Vote. Preliminary planning meetings relating to the possibilities of building an interactive graphic to explain the Alternative Vote (AV) method began in the autumn 2010. With the referendum being scheduled for the following May this did not leave much time to develop and test a prototype. We wrote a description of how AV works for the Sky design team (the slides were never intended for outside consumption and there was no need for good design!). Figure 9.16 - Figure 9.18 featured as part of a PowerPoint presentation that demonstrated the differences between a conventional election using first past the post (FPTP), simple plurality voting and an AV system.

Exeter result 2010 AND 'first count'		
Labour	19942	38.2
Conservative	17221	33.3
Lib Dem	10581	20.3
UKIP	1930	3.7
Liberal	1108	2.1
Green	792	1.5
BNP	673	1.3
Labour elected under FPTP; not yet under AV		

Figure 9.16: Comparing winners/non-winners under FPTP and AV

Figure 9.16 shows the 2010 election result in Exeter where the Labour candidate is elected under FPTP rules but not yet under AV. Figure 9.17 shows that the bottom-placed candidate, in this case the BNP candidate, would drop out of the contest following the first count. In this example, it is assumed that of the 673 BNP votes a total of 273 votes would have no other party indicated on the ballot papers and these would therefore be non-transferable. The remaining 400 votes would be divided between the remaining parties. One hundred votes would transfer to Labour with a further 100 going to the Conservatives and UKIP each. The Liberal Democrats would gain 50 votes, Liberal 30 votes and finally the Green party would gain 20 votes.

Labour	19942 +100	38.6
Conservative	17221 +100	33.3
Lib Dem	10581 +50	20.5
UKIP	1930 +100	3.9
Liberal	1108 +30	2.2
Green	792 +20	1.6
BNP	673	1.3
Non-transferable	273	

Figure 9.17: Elimination of Bottom Candidate in AV

This process would continue if no candidate passed 50% of the overall vote. Figure 9.18, therefore, shows the results following the successive elimination of the Green, Liberal, UKIP and finally Liberal Democrat candidates. Second and lower preferences have been transferred to candidates left in the race until finally the Conservative candidate passes the 50% mark – unsurprising since only two candidates are left.

Labour	19942 +100+150+200+200+3000	49.9 (45.2)
Conservative	17221 +100+100+70+1200+5000	50.1 (45.3)
Lib Dem	10581 +50+250+600+100	22.8
UKIP	1930 +100+10+30	4.0
Liberal	1108 +30+40	2.3
Green	792 +20	1.6
BNP	673	1.3
Non-transferable	273+262+278+570+3581 (4964)	

Figure 9.18: Elimination of Other Candidates and Vote Transfers under AV

Following this short presentation of how AV works it was immediately clear that although the voting system was relatively simple to explain it would be challenging for Sky's graphic design team to produce an application capable of offering the right permutations and to carry out the necessary calculations.

One solution would be to have a grid containing a list of parties (Con, Lab, LD, SNP, PC (or Nats), Green, BNP, UKIP) then the user would be allowed to make their own judgements about how, say Conservatives, would cast second preferences. In the case of Conservatives there would be a procedure where they can build the second preferences so that they sum to 100. A pie chart would do that – I might presume 60% of Conservatives would not cast a second preference vote; 20% would vote UKIP, 10% Lib Dem and 10% Labour. The completed pie chart would announce it was complete and then move the user on to Labour and repeat this process for all other parties.

As it is described here that is expecting a great deal from users, most of whom might want to know about the effects of AV on the bigger picture – would there be a different winner than under FPTP.

The design difficulty therefore was how to slim down the procedure for estimating how second preferences would be cast. One approach was to set the default so that the assumption was that 100% of each party's voters would **not** cast a second preference. In this way users could ignore some parties in favour of one or two parties only.

The other major problem lay with how AV determines the winner. From the software programming point of view this took several months to resolve. The AV method works according to a quota. This quota is equal to exactly half the total first preference votes cast plus one vote. When a candidate reaches the quota he or she is elected. Of course, a candidate can win after the opening round of voting by winning 50% +1 of all valid votes

cast. Otherwise, a transfer or successive transfers of votes from ballots cast for eliminated candidates is undertaken. The AV procedure needs to find for each constituency the named party (Conservative, Labour, Lib Dem, Nats. Green, BNP, UKIP etc) that is at the bottom of the pile.

Let's assume that in constituency X it is a Green candidate with 2000 votes. The user has previously indicated that 50% of Greens (effectively 1000 voters in this example) would not cast a second preference, 30% (600 votes) would go to Labour and the remaining 20% (400 votes) are transferred to the Lib Dems.

The votes in Constituency X are now re-calculated – if the leading party candidate receives 50%+1 or more of the votes (the original first preference vote quota) then he or she is now elected.

If no majority winner emerges then the programme loops, finds the next candidate on the list, let's say it is UKIP, and transfers these votes in the proportions already selected at the outset. Assume UKIP got 3000 votes and the pre-stated pattern was 30% no transfers, 50% Conservative, 10% Green and 10% Labour. This means for Constituency X that 30% + 10% (for the Greens) are not transferred (i.e. 1,200 votes), another 1500 votes move to the Tory column and the remaining 300 votes go to Labour. Re-calculate the constituency vote and see if there is a winner. (50% +1 > of first preference votes). If yes, stop. If no, carry on re-distributing votes.

Also, because of the way AV was going to be implemented if the referendum passed (it was compulsory to select all candidates with a preference vote) it does not follow that the winning candidate in order to be elected had to win an overall majority of the first preference vote. In a constituency where the leading party wins 40% of the votes but also voters from the

remaining parties in the contest do not cast any preference votes at all the winner would remain with the same 40% vote share.

Following a number of discussions it was decided that if the application was going to be ready and useful then more elections needed to be added (so that people could measure the effects of AV for recent elections) but that the number of lower order preference votes need to be restricted and the number of party categories (the Nationalists and UKIP were retain but the Greens and BNP now fell into ‘others’. The application was eventually designed to permit up to three preferences but strongly suggested that people only concern themselves with two preference votes (i.e. a first choice and a second choice.) Accordingly, a spreadsheet contain results from the last five general election was prepared – Figure 9.19 displays some data relating to the 2005 general election. These data would show the finishing position for each party in every constituency and therefore the order of elimination for non-winning candidates beginning from the bottom upwards.

1	year	id	name	con	lab	id	nats	ukip	othv	totv	elect
2	2005	1	Aberavon	3064	18077	4140	3545	0	1278	30104	51080
3	2005	2	Aberdeen North	3456	15557	8762	8168	0	691	36634	65714
4	2005	3	Aberdeen South	7134	15272	13924	4120	0	1171	41621	67012
5	2005	4	Aberdeenshire West & Kincardine	11814	5470	19285	4700	0	379	41648	65548
6	2005	5	Airdrie & Shotts	3271	19568	3792	5484	0	1043	33158	61955
7	2005	6	Aldershot	20572	9895	15238	0	1182	1254	48141	78553
8	2005	7	Aldridge - Brownhills	18744	13237	4862	0	1093	1620	39556	61761
9	2005	8	Altrincham & Sale West	20569	13410	9595	0	736	0	44310	67247
10	2005	9	Alyn & Deeside	8953	17331	6174	1320	918	800	35496	58939
11	2005	10	Amber Valley	16318	21593	6225	0	788	2467	47391	75376
12	2005	11	Angus	11250	6850	6660	12840	0	556	38156	63093
13	2005	15	Argyll & Bute	10150	9096	15786	6716	0	881	43229	67325
14	2005	16	Arundel & South Downs	24752	8482	13443	0	2700	313	49690	72535
15	2005	17	Ashfield	10220	20433	5829	0	0	5569	42051	73403
16	2005	18	Ashford	26651	13353	8308	0	1620	1753	51685	79493
17	2005	19	Ashton under Lyne	7259	21211	5108	0	768	2621	36967	72000
18	2005	20	Aylesbury	25253	9540	14187	0	2479	0	51459	82428
19	2005	21	Ayr, Carrick & Cumnock	10436	20433	6341	5932	365	1541	45048	73448
20	2005	22	Ayrshire Central	9482	19905	6881	4969	346	1288	42871	68643
21	2005	23	Ayrshire North & Arran	8121	19417	7264	7938	382	1083	44205	72986
22	2005	24	Banbury	26382	15585	10076	0	1241	2915	56199	87168
23	2005	25	Barnf & Buchan	7207	4476	4952	19044	442	1095	37216	65570
24	2005	26	Barking	4943	13826	3411	0	803	6123	29106	57658
25	2005	27	Barnsley Central	3813	17478	4746	0	0	2578	28615	60592
26	2005	28	Barnsley East & Mexborough	4853	20779	6654	0	0	740	33026	66941
27	2005	29	Barnsley West & Penistone	9058	20372	7422	0	0	0	36852	66985
28	2005	30	Barrow & Furness	11323	17360	6130	0	758	922	36493	61883

Figure 9.19: Example of data for AV application

It is clear from Figure 9.20 that some of the design features of the seat calculator were imported into the AV design. Because the graphic spoke about some of the smaller parties, however, it was important to separate them out of the “others” category. The fact that more

than one election was being examined it was important that for each election that the opening screen describe the outcome at that election. For example, Figure 9.21 shows the outcome of the 1992 general election in terms of national (Great Britain) vote share and the number of seats won in the House of Commons.

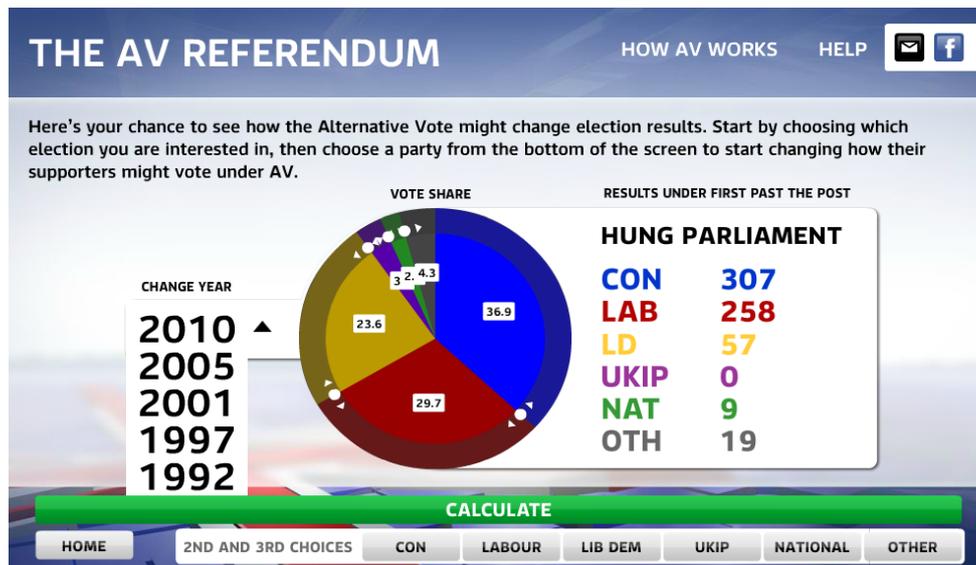


Figure 9.20: Opening screen of the Alternative Vote Application⁴¹

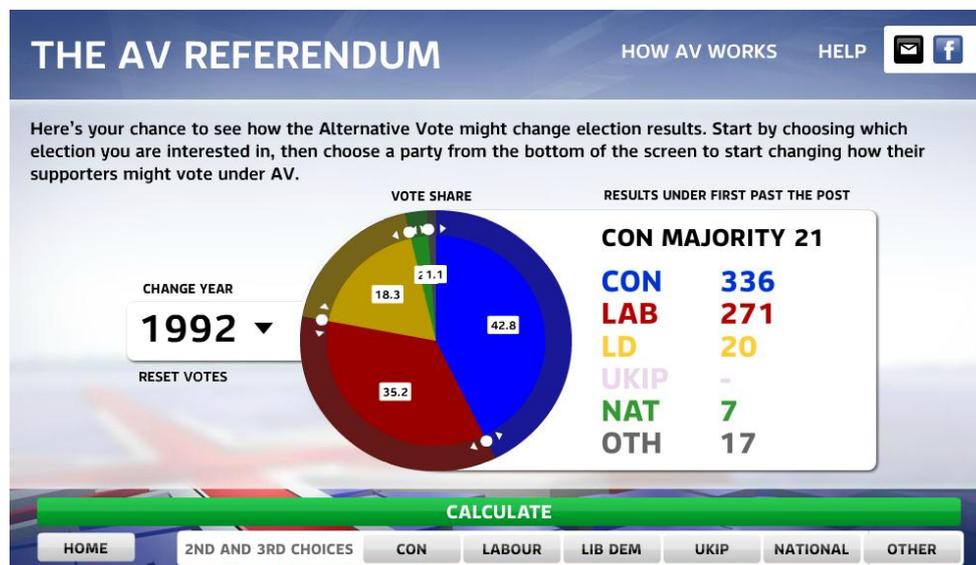


Figure 9.21: Selection of the 1992 General Election to test effects of AV

⁴¹ <http://news.sky.com/skynews/Interactive-Graphics/avcalculator>

At the bottom of each screen below the green bar marked “Calculate” was a list of parties. By clicking on one of these another screen would open (see Figure 9.22). This screen relates to the second choices of people whose first vote choice was for the Conservatives. In this example the user has opted for 43% of Conservative voters giving their second choice to the Labour candidate, 39% for the Liberal Democrats and 32% for UKIP. This is clearly more than 100% however as the number “114%” indicates. It was possible for this to be less than 100% however as the brief explanation makes clear. There is also the option to “Enable Third Choices” which would result in another screen relating to Conservative voters to open where the process of allocating third choice preferences among the other parties would continue. It was felt that only expert users would have the patience to go past the second preferences stage.

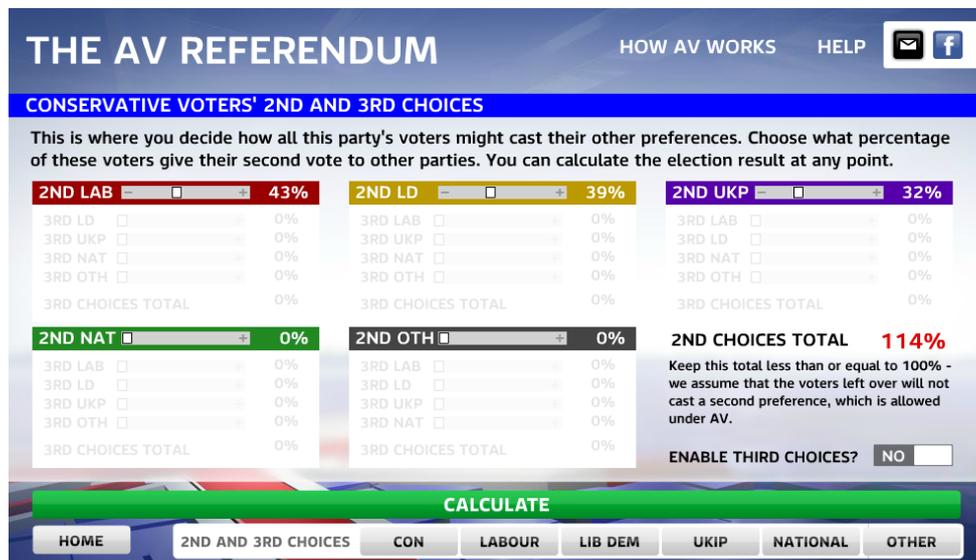


Figure 9.22: Allocation of Conservative Voters' 2nd preference votes

The AV application was a very powerful tool for imagining different scenarios. Figure 9.23, for example, illustrates this by showing Conservative voters disproportionality favouring Liberal Democrats (their new coalition partners) with 90% of second preferences.

Figure 9.24 returns the compliment with Liberal Democrat voters using second preference votes to favour the Conservatives.

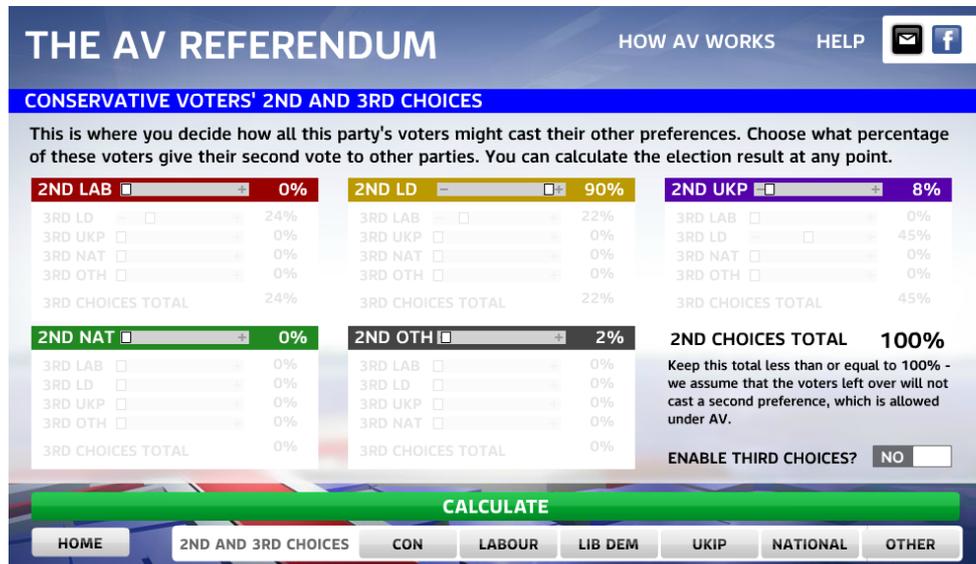


Figure 9.23: Experiment showing Conservative 2nd votes transferring to Liberal Democrats

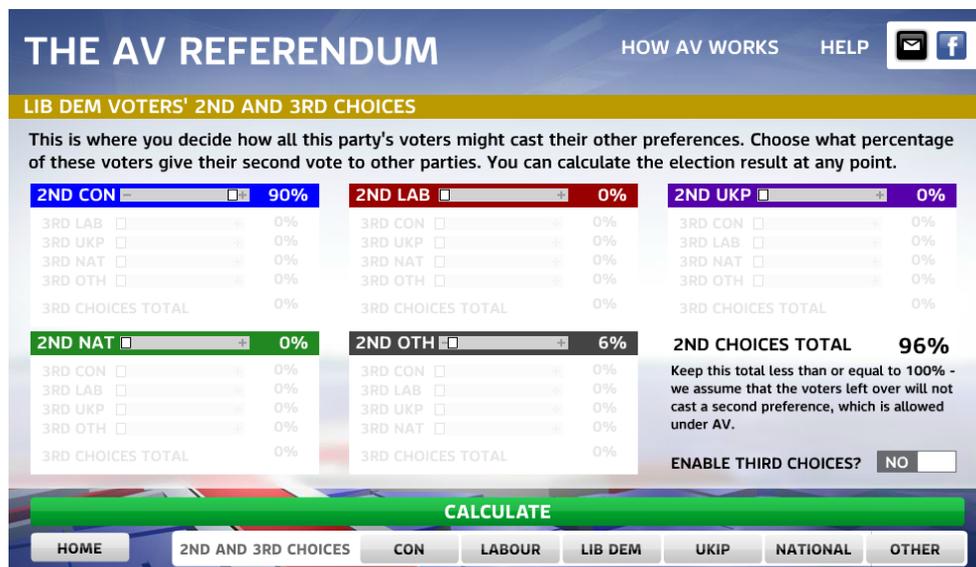


Figure 9.24: Experiment showing Liberal Democrat 2nd votes transferring to Conservatives

Of course, these are absurd examples but from the viewpoint of user interactivity this was a very powerful application. However, users required patience to step through the parties and

to make decisions about the casting of lower order preference votes. Of course, it was possible for users to ignore some of the categories. In Figure 9.25, for example, the user has merely adjusted second votes of Conservative and Liberal Democrat voters in the ways described in Figure 9.23 and Figure 9.24.

Pressing the calculate button would present the user with Figure 9.25 which displays two maps of the Britain (Northern Ireland is missing and is always assumed to give 18 seats into the ‘others column’). The result on the left is the actual 2010 result under first past the post. It was decided to use physical maps rather than cartograms because it was more important for people to be able to identify which constituencies would have changed hands if AV had been used and Conservative voters had switched to Liberal Democrats with their second preferences and vice versa. The right-hand map, therefore, simply shows those seats with difference winners under AV than under FPTP. According to this particular configuration, therefore, the AV system would have given the Conservatives an additional 84 seats, 32 extra seats for the Liberal Democrats but 114 fewer seats for Labour.

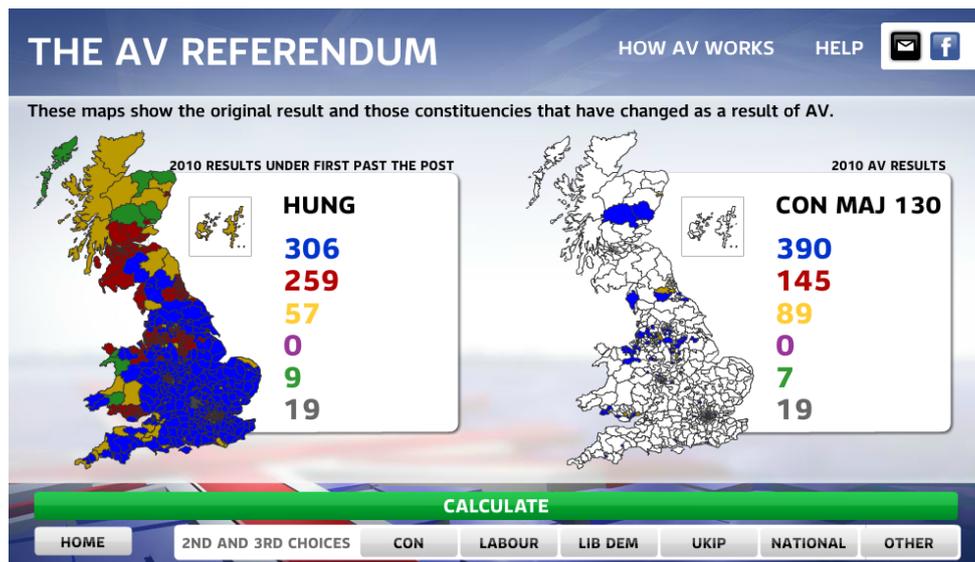


Figure 9.25: Effect on 2010 House of Commons composition of vote transfers

The AV application was the most complex of the web applications considered in this chapter and while it was a technical success in terms of programming it arguably failed as an example of web interactivity. Many of the designers that had worked on previous projects were involved in this one – the overall impression is similar to the general election applications but it was the effort required by the user to engage with the application that was the problem. Unless the user knows something about the AV system it is difficult to explain that quickly. There was an explanation that lay behind the button marked “How AV works” but this was quite a lengthy explanation using a video animation lasting about three minutes (of course some users might have accessed this application simply to gain an understanding of how voting under AV worked).

The most insurmountable problem, however, came with trying to understand the many configurations between parties and distribution of second choice, let alone third choices. Potentially, with six parties to choose from and a further five parties in each case to give second preference votes to a user would be involved in more than 30 decisions about how to cast second votes and in what proportion. Practically speaking, not many users would be prepared to do that.

9.7 Conclusions

This chapter has discussed the results of some collaboration between the Elections Centre, Plymouth University and Sky News Broadcasting. The aim of that collaboration was to build multi-platform interactive applications that would engage users interested in different aspect of British politics, particularly the electoral process. This is not a complete list of the activities but the applications examined here represent a comprehensive range of the type of applications that were developed.

Two main principles were involved. The data behind the applications was of a factual nature – opinion polls, historical election results, results of recent general elections, party manifestos and the Alternative Vote. This meant that the data could normally be supplied in Excel spreadsheet format or as text files in terms of instructions for using the application, explanations about the methodologies and in the case of the manifestos extracts or summaries of these. Second, although the Elections Centre was involved in the initial planning and was consulted about design mechanics it was not involved in the overall ‘look’. The appearance was left to the software and graphic designers at Sky Broadcasting to determine. In other words Plymouth supplied data and had input into the mechanics of operation which it shared with Sky but design, appearance and interactivity was left with Sky.

In fact, at the moment it could not have worked in any other way. There is insufficient programming and graphic design available within the Elections Centre to be able to create stand-alone applications that could match the sophistication of the final product. There is insufficient familiarity within Sky broadcasting of election data and electoral systems generally to be able to do the work completely in-house.

One of two events would need to happen before the Plymouth/Sky model would change. First, electoral data becomes freely available on the internet although it is difficult to see who would produce this on a regular and consistent basis. Second, software and graphic design becomes intelligent, allowing even novice users to programme interactive graphics. Of these two possibilities we think the first would happen before the second.

10 Conclusions

10.1 Overview

Our goal was to provide an overview of the state of information visualisation of political and especially electoral data and then to explore how visualisation might be used to improve engagement and understanding with those data. Accordingly, we made some attempts to develop a variety of methods that would, in theory, offer users additional way of obtaining, interpreting and understanding political information.

Information visualisation will continue to play a major role in the way we communicate data in a society that is becoming increasingly digital. There is now so much information available, in different forms and with varying levels of quality, from a wide range of sources. Some of these sources are officially sanctioned (government statistics), others are obtained from the media and to this list we should now add information supplied by ordinary members of the public as they interact with new forms of social media.

As information becomes more ubiquitous in its various forms (though primarily as digital content delivered via the internet), we, as a general public, are becoming more comfortable with our ability to navigate these information spaces as part of our everyday lives: we use Google to search the web; we look for audiovisual content on sites like YouTube and Flickr; we exchange information and ideas across complex networks of blogs and news aggregators.

These activities point to the development of a generalised information literacy deployed against the information deluge. The so-called Information Age is not simply about the imposing presence of information at the internet scale, but also about the ways that we respond to it socially and culturally.

Accordingly, people are becoming more willing to engage with information visualisation as a means to mediate some of this information exchange, not necessarily as academic specialists but as members of the interested public. Unfortunately, as parts of this dissertation demonstrate information visualisation design is not yet ready to engage with us on these terms. There is a big gap between imagining how the public might engage with electoral data and actually designing software and systems that can implement that vision.

10.2 Chapter summaries

Following the introductory chapter we provided a basic introduction to information visualisation in Chapter 2. Information visualisation, as a work activity, is interdisciplinary, requiring both technical expertise and creative ability. Some of the most celebrated examples of graphical display were shown in order to demonstrate how this information visualisation has evolved, each era building on the best examples produced by the previous one. Its development in the modern era is informed by work in a variety of domains, from neuroscience and understanding of how the brain functions to interpret images, from cognitive psychology the processes of perception and understanding, from art and design the sense of what constitutes an image that is both aesthetically pleasing and functional. Undoubtedly, one of the biggest contributions to this activity has been the growth of computers, in terms of hardware processing power and storage capacity, growth in new

telecommunication technologies and the associated growth in software engineering. This growth leads to ever more sophisticated ways of creating displays of complex data.

Chapter 3 looked at different methods and applications of political data visualisation. It began with a criticism of research that is published in the field of political science. Despite widespread changes in ways of displaying data that appear in scientific journals, and in the published and broadcast media, most journals in political science continue to report empirical findings in tabulated data formats. Despite criticisms from some leading political scientists that more effort should be paid to the presentation of research findings very little has been to address their concerns. The gap between how political scientists report research and how the general public has become accustomed to view political and electoral data is growing.

The emergence of the internet and the arrival of handheld devices capable of recording video and images have contributed to many changes in modern politics. In advanced post-industrial societies like Britain and the United States, growing numbers of citizens are turning to the internet for news, especially younger people. Election news is likely to be viewed on a larger variety of platforms than before. Increasing numbers of people in the UK also got election news via mobile phones and through interactive digital TV (the red button). However, these audiences are quite distinct. As long as the web continues to supplement rather than substitute for traditional media like television, the audience for richer audio and video content online appears to correlate with the more politically engaged.

The demand for interactivity might include the use of interactive issues guides which allow people to see how closely their stance on the issues matches those of the political parties. The audience is also likely to want more types of interactivity, such as links to external blogs as well as increased blogging by journalists; online forums where politicians can be questioned

directly; and perhaps the increased use of citizen journalists to report on local campaigns or local issues.

Chapter 4 focussed on the visualizing power of mapping, especially election maps and cartograms in order to explore both their strengths as data describers and sometimes their practical disadvantages. Electoral data is an ideal subject for mapping processes because this is an excellent way of uncovering spatial characteristics of voting patterns, for example. But comparisons between conventional mapping procedures, choropleth maps and cartograms have revealed that map patterns are not always communicated clearly to users. Visual deceptions occur commonly in choropleth map reading, especially when there is an inverse correlation in an electoral area's physical size and its voting population and weight in the legislature. Meanwhile, the distortion of geographical information on cartograms usually leads to misunderstanding of adjacency relationships between areas. Cartographers have spent considerable effort in designing methods to avoid these problems but not entirely successfully. It is not possible to achieve a completely correct map, only a suitable map. In an ideal world we should present cartograms and conventional maps together in order to avoid misleading the reader.

Advances in GIS have provided better opportunities for map construction to become an activity engaged in by the many rather than the few. The new technologies of Web-based GIS have helped to strengthen the analytical power of maps and to disseminate them across large numbers of users more quickly than ever before. This study has not been able to give any definitive answers on which maps are better than others for presenting election results. The modern trend is towards 'dashboard' technologies where web-based information contains a mix of mapping and other forms of data display. Media coverage of elections in the United States began this trend and as Chapter 4 demonstrated it has now migrated into the UK. It

appears that most newspapers and news broadcast channels have the capacity to generate map-based election content for their websites and broadcast output. Our analysis shows these organisations approached the map-building process in different ways during the 2010 general election campaign in Britain. All of these applications were interactive, however, further emphasising the gap between political science output that is both static and largely restricted to tabulated data reporting.

Chapter 5 comprised a series of short case studies that suggest that the study of politics would benefit considerably if there was more reliance on more visual displays of data. The Elections Centre has collaborated with human geographers over recent years and experienced at first hand the value of mapping data in order to detect patterns and trends that would probably have remained undetected using standard tables of frequencies, cross-tabulations and correlation statistics. Both the work undertaken on the pattern of support for the Liberal Democrats during the 1980s and 1990s and that investigating electoral participation in a range of different elections demonstrating the power of using visual displays when undertaking empirical research. This method worked so well that it was subsequently used in two further projects both of which considered some features of voting patterns for the Greater London Authority. Sadly, although both of these papers were subsequently published, in neither case were the maps reproduced in full-colour where the different distributions and configurations of party competition would have been seen more clearly.

The second part of Chapter 5 described how graphical displays of data can greatly assist the research process, enabling collaborators to see the effects on trend data of correcting for missing data and computing average values for different sections of time. The Centre is renowned for its election forecasting methods that contrary to most other forecasting models does not employ opinion poll data. The problem derived from local parties sudden reluctance

to contest by-election vacancies when there was little or no prospect of electoral success. The point of this case study was to demonstrate that working with graphical (line graphs primarily) data displays can help to highlight the effects of estimating missing data in ways that other, non-graphical methods, would not do so.

Chapters 6-9 report on different projects that all relied upon some form of data visualisation in order to communicate with non-expert users about aspects of the electoral process.

Chapter 6 provides an example at what may become more fashionable as online learning is undertaken by more universities in the coming decades. Most lecturers and teachers that engage in degree or higher-degree learning do so through the familiar lecture/seminar process. Although PowerPoint is frequently used as the platform of choice, we suspect that the existing skill-levels in utilising this software are not particularly high. We set out to record two presentations that would explain some technical aspects of the electoral process – election bias generation and the supplementary vote – where the target audience was not university undergraduates but instead the general public. While the experiment did not demand some interactive element, the audience was passive, not active, it did require the presentation to be suitably interesting that people would stay with it. YouTube is yet another innovation that provides new opportunities to capture data and transform it using skill and imagination in a way that broadens the potential audience for political science outputs. While there were problems with this experiment it does suggest that there is considerable scope for further expansion into this method for disseminating data.

The following two chapters, 7 and 8 respectively, are linked by the ambition of constructing a natural language interface that as originally envisaged would permit non-expert users to interact with complex relational databases. Chapter 7 reported a short-lived experiment to allow mobile phone users to send texts to a host number, requesting information about some

particular aspect of the European parliament elections. The proposal was that it would be possible to construct an interface that would receive a natural language enquiry, such as ‘how many votes did Conservative party get in south west’ and then convert that into a syntax that the computer database could correctly interpret and issue a text response containing the answer to the original question. Expressed in those terms it looks like a deceptively simple task to achieve. It was not. We encountered many problems along the way, mostly associated with the range of natural language enquiries but also including other problems relating to the small response to our request for participants and the limitations of the telephone technology, especially small screen and text character limits.

Chapter 8, therefore, was a second attempt that moved the idea to a different platform, the personal computer, and became much more ambitious in its focus. By switching from the low interest European elections to the relatively higher interest general election, we believed that more participants could be recruited into the experiment. Unfortunately, the outcome was largely unsuccessful and there are many lessons that can be learned. Quite simply, constructing a natural language interface that can work within the realms of election data processing prove to be beyond our abilities. It was not possible to consider the many permutations that natural language reveals for what political scientists assume to be straightforward concepts. The concept of ‘winning’ in natural language can translate as the name of the winning party, the winning candidate, the number of votes received, and vote share as far as a computer is concerned.

But the scheme did have some positive outcomes. We demonstrated some principles about election data reporting, distributing output as tables, graphical displays or some level of mapping. These principles were implemented within NITED. Tables were used not as the default but rather as the method of last resort. It was always preferred to display data as

histograms or maps. Although it was brave to imagine non-expert users could interrogate an elections database the reality was that we could not succeed in building an interface that was simple to understand and easy to use. The length of the Help menu grew as we encountered potential problems and misunderstandings that might stop users from accessing and manipulating the data. The real success, we believe, was in the mapping visualisation of three-party vote share. Our drill-down maps are a real innovation and provide an intuitively attractive way of displaying the relative strengths of three parties at different levels of aggregation.

If the stories of Chapters 7 and 8 were rather pessimistic in terms of the overall outcome it was a much better story regarding the collaboration between the Centre and Sky broadcasting described in Chapter 9. The collaboration brought together a Plymouth team that were experts in the field of electoral politics and the Sky-based team that comprised a range of people that complemented our own strengths in database construction and management but then added to that combination other people who were experts in design and production. From the initial development of poll tracker to the Alternative Vote application the collaboration proved successful and rewarding for those involved. Political scientists with a real interest in data visualisation suddenly had access to a group of people with the necessary skills to implement that vision. It was not entirely successful, however, with some applications proving more difficult to implement than others but that has to be seen as a part of the learning process.

10.3 Evaluating Outcomes

The opening chapters described some of the theories and practices that lie behind data visualisation methods. From the earliest attempts at visualising information using graphs,

tables and maps we have and continue to make rapid improvements in this field. The ability to generate interesting and informative visualisations is no longer restricted to a small group of experts but is now extended to anyone with access to a computer and the internet. The volume of freely available data that is becoming available means that there are fewer limits on what we can visualise and how we visualise it.

This thesis contains a number of case studies that attempted to create new ways of visualising election-related data to a wide audience. Chapter 5 demonstrated to us that visualisation is an extremely powerful means of seeing patterns in data and allowed us to trail different approaches to the problem of estimating missing data values and averaging over different time periods. The direct result was an improved forecasting model but the indirect outcome was a stronger awareness that telling the story through graphical and other displays was the way forward. If it worked for us (in the sense of the research team) then it should work for other audiences.

Political scientists, like other academics from across the disciplines, tend to be content to talk to one another. This partially explains why innovation in the way data findings are reported is slow and the voices that want things to change are often not heard. The other part of the explanation is that academics have traditionally published their findings in hard copy formats. Mostly, these books and journals continue to publish these static images in black and white. This is a lethal combination from the viewpoint of reaching out to a wider audience. Most empirical data is still reported as tables; graphical displays of data are rare and mapping is even rarer. Most academics are not trained to use visualisation software and most have only a basic knowledge of building a PowerPoint presentation. This is not necessarily their fault since so little value is placed on the ability to present data in an attractive and stimulating format.

So, in Chapter 6 we set out to see if we could do something that would be attractive and stimulating and most of all, easily understood by the interested public. Although the video presentations which are the subject of this chapter do not compare with the design elements in the applications described in Chapter 9, they are still interesting because they explain complex issues about elections and voting to a general audience. Very few people are prepared to read long text-based explanations of either electoral bias or the supplementary vote but many more than that are willing to sit down and watch a video about those subjects. What was learnt from the first experiment was that forty minutes is too long, that some parts of the explanation are difficult for non-experts to follow but that generally people understood more about the subject at the end than they did at the beginning! At the time of writing it has received just over 1,800 views. Consequently, the video on Supplementary Vote was shortened to just over eleven minutes, simplified the explanation, did not go into too much detail about the consequences of SV and yet still succeeded in getting a generally positive response from viewers. At the time of writing it has received just under 500 views. This may not seem like a lot but it is probably more people than would have read texts explaining electoral bias and the supplementary vote. Hans Rosling's visualisation of global life expectancy currently stands at more than five a half million views on YouTube (<http://www.youtube.com/watch?v=jbkSRLYSojo>) suggesting that when the production values are excellent and the presentation is relevant and interesting then visualisations can reach many people.

Our original PowerPoint presentations from which these videos were made are certainly better than most of those created by political scientists but they were not developed by people with skills in graphic design. PowerPoint software has improved considerably over the last decade but the interface is not particularly user-friendly and achieving animation is a rather

time-consuming exercise. This is one of the problems with data visualisation – most software can easily output tables and lists but it is not very good for going beyond that without the user having to train how to use the program to its fullest capability.

Another route to data exploration is described in Chapters 7 and 8. The aim was to design an interface (mobile phone text messaging in one case, a graphical-based system in the other) that would be natural from the user's viewpoint. If users could engage in a natural way and not forced to use language written by software developers then it should be possible to provide data answers to written questions. The question-answer approach was, with hindsight, over-ambitious. It might work in some areas – providing answers to simple questions but we could never develop a system that could do that for electoral data. Allowing users to employ their own language descriptions and syntax constructions proved too complex for us to manage the numerous combinations using a set of production rules. This is not to say that election databases cannot be interrogated in this manner in the future but we suspect the time is a long way off when someone using a computer keyboard or a phone keypad can write a question freely and expect an answer that is immediate (without further qualifications being requested) and accurate (this is what was expected and not some other answer to a question that was not asked).

Despite this, some important lessons were learned from these two experiments. If the list of Help Instructions is large then the application will not work. People expecting to interact with data do not have the time or patience to wade through a long list of how to use the system. Similarly, if the list of Frequently Asked Questions becomes large then the input interface is not working properly. If the system has to constrain language forms where the user must use a particular word to describe an object (vote, share etc) then it is no longer a natural language interface. On the positive side the system did function and allowed non-

expert users to interrogate a relational database that contained multi-level data (region, constituency, candidate) in order to obtain data that revealed itself in the form of tables, graphs and maps according to an algorithm that determined the optimum output for any particular item of data. Moreover, the example of mapping that reveals the distribution of vote shares across three parties is novel, simple and easy to understand.

Although these two case studies were sequential (NITED was developed after the mobile phone experiment failed) they were both developed in parallel with the collaboration that was taking place between the Elections Centre and Sky News. Apart from the objective to construct a natural language interface there are some similarities between these different cases. All involved the principle that data visualisation/information retrieval could be given to non-expert users. Both NITED and the Sky applications believed in the principle that if data could be visualised then it should be.

The real differences lay with the design elements and in the case of Sky News the overarching need to keep things simple. Their experience with both broadcast graphics and website interactive applications is that users could not be encouraged to spend more than several moments learning how a graphic was meant to work or how an application was interactive. Starting from that position, rather than the political scientists' desire to extract as much out of a system that it contained is, we think, the difference between making a visualisation that works and one that does not. This is especially the case with interactive graphics where it has to be intuitive from the viewing point and easy from the interactive point.

This is why poll tracker and most of the other applications worked but the application designed for the Alternative Vote did not. In the case of the latter this breached the rule that

instructions should not be too long and convoluted and also broke the rule that the mechanics of what the user was meant to do required too much explanation. To be fair, explaining and then running a simulation of the Alternative Vote, is not a simple matter and this raises the issue about subject complexity. Maybe there are some aspects of election data and voting systems that lie beyond the capabilities of even the best political scientists/broadcast producers/graphic designers.

10.4 The future

The field of information visualisation is on the cusp of a transition to mainstream relevance that could revolutionise the way citizens engage with political information. It now represents a widely-based means of communication that is open to many developers. As computerisation proceeds and as more information becomes available in digital format then this process of democratisation opens up huge possibilities. Graphical displays of data have grown rapidly over the past centuries but it is likely that this current century will see that growth increase still further.

Where does this leave political science? If non-expert users can interrogate databases, interact with online applications, control the flow of information in terms of content, speed and appearance then perhaps they have a limited role in the future. During the 2010 UK general election there were many website applications that permitted users to plug in either their own assessment of likely party votes or those predicted by the national opinion polls. Swingometers and forecasting models are now available to all.

The next generation of students of politics will be used to seeing their subject visualised in many different ways, across different formats (YouTube, desktops, interactive web

applications, Flickr etc) and in real-time. By comparison, powerpoint presentations that are basically text, journal articles and books that contained mostly words with a few tables, are going to look rather dated. Gary King and others have appealed to political scientists to present their empirical findings in more imaginative ways, using graphical displays, maps and to that list we should now add interactive applications also. The movement towards more online publishing for academic journals certainly offers greater possibilities than traditional paper formats. It is time that political scientists accepted the challenge.

11 List of References

- Ahlberg, C. and B. Shneiderman (1994). Visual information seeking: Tight coupling of dynamic query filters with starfield displays. *Human Factors in Computing Systems, Chi '94 Conference Proceedings - Celebrating Interdependence*. B. Adelson, S. Dumais and J. Olson: 313-317.
- Andrienko, G. L. and N. V. Andrienko (1999). "Interactive maps for visual data exploration." *International Journal of Geographical Information Science* 13(4): 355-374.
- Andrienko, G. and N. Andrienko (2007). Coordinated multiple views: a critical view. *CMV 2007: Fifth International Conference on Coordinated & Multiple Views in Exploratory Visualization, Proceedings*. G. Andrienko, J. C. Roberts and C. Weaver. Zurich, Switzerland 72-74.
- Andrienko, G., N. Andrienko, et al. (2010). "Space, time and visual analytics." *International Journal of Geographical Information Science* 24(10): 1577-1600.
- Androutsopoulos, I., G. Ritchie, et al. (1995). "Natural language interfaces to databases-an introduction." *Journal of Language Engineering*, 1(1): 29-81.
- Androutsopoulos, I. and G. Ritchie (2000). Database interfaces. *Handbook of Natural Language Processing*. R. Dale, H. Moisl and H. Somers, Marcel Dekker Inc.: 209-240.
- Anselin, L. (2012). "From SpaceStat to CyberGIS: Twenty Years of Spatial Data Analysis Software." *International Regional Science Review* 35(2): 131-157.

- Anstead, N. and A. Chadwick (2009). Parties, election campaigning, and the internet: toward a comparative institutional approach. In A Chadwick & PN Howard (eds), *The Handbook of Internet Politics*. Routledge: 56-71.
- Archibald, S. (2008). *Ways of Seeing: Alfred Yarbus's science of visual attention*. *Cabinet Magazine*. 30.
- Artusi, M. and A. Maurizzi (2011). "The involvement process through social media in 2010 UK elections: is it a real Facebook based participation?" *International Journal of Technology Marketing* 6(4): 305-315.
- Atkinson, A. C. (1985). *Plots, transformations, and regression : an introduction to graphical methods of diagnostic regression analysis*. Oxford, Clarendon Press.
- Averill, E. (1993). "Hidden kind classification." *Psychology of Learning and Motivation* 29: 437-467.
- Bajaj, C. (1999). *Data visualization techniques*. Chichester, John Wiley.
- Bear, M. F., B. W. Connors, et al. (2006). *Neuroscience: Exploring the brain*. New York, Lippincott Williams & Wilkins.
- Becker, R. and W. Cleveland (1987). "Brushing Scatterplots." *Technometrics* 29(2): 127-142.
- Bekkers, V. and R. Moody (2011). "Visual events and electronic government: What do pictures mean in digital government for citizen relations?" *Government Information Quarterly* 28(4): 457-465.
- Bennett, L. and R. Entman (2001). *Mediated politics: an Introduction*. *Mediated Politics*. L. Bennett and R. Entmann. Cambridge Cambridge University Press: 1-29.
- Bernstein, A., E. Kaufmann, et al. (2006). Ginseng: A guided input natural language search engine for querying ontologies. *5th International Semantic Web Conference*: 144-157.
- Bertin, J. (1983). *Semiology of graphics*. Madison, Wis., University of Wisconsin Press.

- Bertin, J. and W. J. Berg (2010). *Semiology of graphics : diagrams, networks, maps*, ESRI Press.
- Borisyuk, G., C. Rallings, et al. (2007). "Voter support for minor parties - Assessing the social and political context of voting at the 2004 European elections in greater London." *Party Politics* 13: 669-693.
- Borisyuk, G., R. Johnston, et al. (2010). "A method for measuring and decomposing electoral bias for the three-party case, illustrated by the British case." *Electoral Studies* 29(4): 733-745.
- Borisyuk, G., R. Johnston, et al. (2010). "Parliamentary Constituency Boundary Reviews and Electoral Bias: How Important Are Variations in Constituency Size?" *Parliamentary Affairs* 63(1): 4-21.
- Braithwaite, K., M. Lishman, et al. (2007). "Distinctive Features of Mobile Messages Processing." *International Journal "Information Theories & Applications"* 14(2): 154-160.
- Brown, J. R., R. Earnshaw, et al. (1995). "Visualization: using computer graphics to explore data and present information", Wiley.
- Burns, D., R. Fallon, et al. (2005). "Verbal Dialogue Versus Written Dialogue." *International Journal of Information Theories & Applications* 14(2): 369-377.
- Butler, J. and R. Mautz (1996). "Multimedia Presentations and Learning: a laboratory experiment." *Issues in Accounting Education* 11: 259-280.
- Capurro, R. and B. Hjørland (2003). *The Concept of Information. Annual Review of Information Science and Technology*. B. Cronin. 37: 343-411.
- Card, S., G. Robertson, et al. (1991). "The Information Visualizer, an Information Workspace." *CHI'91*: 181-188.

- Card, S. and J. Mackinlay (1997). "The Structure of the Information Visualisation Design Space". Proceedings of the 1997 IEEE Symposium on Information Visualization, IEEE Computer Society. : 92-99.
- Card, S., J. Mackinlay, et al. (1999). Readings in Information Visualization: Using Vision to Think. San Francisco, California, Morgan-Kaufmann.
- Castells, N. (2007). "Science and technology in society: From biotechnology to the internet." Contemporary Sociology-a Journal of Reviews 36(2): 156-158.
- Chik, D., R. Borisjuk, et al. (2009). "Selective attention model with spiking elements." Neural Networks 22: 890-900.
- Cimiano, P., P. Haase, et al. (2007). "Towards portable natural language interfaces to knowledge bases: The case of the ORAKEL system." Data and Knowledge Engineering, 65: 325-354.
- Cleveland, W. (1984). "Graphs in Scientific Publications." American Statistician 38: 270-280.
- Cleveland, W. and R. McGill (1984). "The many faces of a scatterplot." Journal of the American Statistical Association 79(388): 807-822.
- Cleveland, W. and R. McGill (1985). "Graphical Perception and Graphical Methods for Analyzing Scientific Data." Science 229: 828-833.
- Cleveland, W. (1993). "A Model for Studying Display Methods of Statistical Graphics." Journal of Computational and Graphical Statistics 3: 323-364.
- Cleveland, W. (1993). Visualizing Data, AT&T Bell Laboratories.
- Cleveland, W. (1994). The Elements of Graphing Data,. Murray Hill, NJ, AT&T Bell Laboratories.
- Cockerill, J. (2003). "Improving digital imaging." Plastic and Reconstructive Surgery 111(1): 503-504.

- Cohen, P. (1991). The Role of Natural Language in a Multimodal Interface. UIST '92 Proceedings of the 5th annual ACM symposium on User interface software and technology, ACM New York: 143 - 149
- Coles, T. and V. Lovitskii (2000). "Text Searching and Mining." Journal of Artificial Intelligence National Academy of Sciences of Ukraine 3: 488-496.
- Coltheart, V. (1999). *Fleeting memories: Cognition of brief visual stimuli*, MIT Press.
- Copestake., A. and J. K. (1989). "Natural language interfaces to databases." Knowledge Engineering Review. Special Issue on the Applications of Natural Language Processing Techniques 5(4): 225-249.
- Dorling, D. (1991). *The Visualization of Spatial Structure*, University of Newcastle upon Tyne.
- Dorling, D. F. L., C. Pattie, et al. (1993). "Measuring Electoral Change in Three-Party Systems: An Alternative to Swing." PS: Political Science and Politics 26(4): 737-741.
- Dorling, D., C. Rallings, et al. (1998). "The epidemiology of the Liberal Democrat vote." Political Geography 17(1): 45-70.
- Downs, A. (1957). *An economic theory of democracy*. New York, Harper & Row.
- Ellis, G. and A. Dix (2007). "A taxonomy of clutter reduction for information visualisation." Ieee Transactions on Visualization and Computer Graphics 13(6): 1216-1223.
- Fairbairn, D., G. Andrienko, et al. (2001). "Representation and its Relationship with Cartographic Visualization." Cartography and Geographic Information Science 28(1): 13-48.
- Farmer, E. (2009). "A primer of GIS: fundamental geographic and cartographic concepts." International Journal of Geographical Information Science 23(4): 548-549.
- Felleman, D. and D. Van Essen (1991). "Distributed hierarchical processing in the primate cerebral cortex." Cerebral Cortex 1(1): 1-47.

- Ferdinand, P. (2000). *The Internet, Democracy and Democratization*. London: Frank Cass Publishers.
- Few, S. (2004). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press.
- Few, S. (2006). *Information Dashboard Design*. Beijing, O'Reilly.
- Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*, Analytics Press.
- Fishkeller, M., J. Friedman, et al. (1974). PRIM-9: An interactive multidimensional data display and analysis system. *Proceedings of the Pacific ACM Regional Conference*.
- Foot, K. A. and S. M. Schneider (2002). "Online action in campaign 2000: An exploratory analysis of the US political web sphere." *Journal of Broadcasting & Electronic Media* 46(2): 222-244.
- Fowlkes, E. (1969). *User's manual for a system for interactive probability plotting on Graphic-2*, Bell Laboratories.
- Francis, G., M. Lishman, et al. (2007). "Instantaneous Database Access." *International Journal Information Theories & Applications* 14: 161-168.
- Friendly, M. (2002). "Visions and re-visions of Charles Joseph Minard." *Journal of Educational and Behavioral Statistics* 27(1): 31-52.
- Furnas, G. (1981). *The Fisheye View: A New Look at Structured Files*. San Francisco, Morgan Kaufmann.
- Gage, J. (1999). *Colour and meaning. Art, Science and Symbolism*. Thames and Hudson. London.
- Gardner, H. (2011). *Frames of Mind: The Theory of Multiple Intelligences*. New York, Basic Books.

- Gastner, M. and M. Newman (2004). "Diffusion-based method for producing density-equalizing maps." *Proceedings of the National Academy of Sciences*, 101(20): 7499-7504.
- Gelman, A., Unwin, A. (2011). *Infovis and Statistical Graphics: Different Goals, Different Looks*. *Journal of Statistical Computing and Graphics*.
- Gibson, R. and S. J. Ward (1998). "UK political parties and the Internet - "Politics as usual" in the new media?" *Harvard International Journal of Press-Politics* 3(3): 14-38.
- Gibson, R., M. Margolis, et al. (2003). "Election campaigning on the WWW in the USA and UK - A comparative analysis." *Party Politics* 9(1): 47-75.
- Gibson, R. (2008). "Internet politics: States, citizens, and new communication technologies." *Governance-an International Journal of Policy and Administration* 21(1): 155-156.
- Gibson, R. (2012). "From Brochureware to 'MyBo': An Overview of Online Elections and Campaigning." *Politics* 32(2): 77-84.
- Gorman, A. (1961). "Recognition memory for nouns as a function of abstractness and frequency " *Journal of Experimental Psychology* 61: 23-29.
- Gray, C. (1999). "Temporal correlation hypothesis of visual feature integration: is still alive and well." *Neuron* 24: 31-47.
- Grossman, L. (2009). "Iran's Protests: Why Twitter is the Medium of the Movement." *Time Magazine*.
- Gunter, B., V. Campbell, et al. (2009). "Blogs, news and credibility." *Aslib Proceedings* 61(2): 185-204.
- Hansen, C. and C. Johnson (2005). *The Visualization Handbook*. Oxford, Elsevier.
- Harley, J. and D. Woodward (1987). *The History of Cartography, Cartography in Prehistoric, Ancient, and Medieval Europe and the Mediterranean*, University of Chicago Press.

- Hartman, F. (1961). "Single and multiple channel communication: a review of research and a proposed model." *Audio Visual Communication Review* 9: 235-262.
- Hayhoe, M. and D. Ballard (2005). "Eye movements in natural behaviour." *Trends in Cognitive Sciences* 9: 188-194.
- Heer, J., S. Card, et al. (2005). Prefuse: a Toolkit for Interactive Information Visualization. CHI '05 Proceedings of the SIGCHI conference on Human factors in computing systems: 421-430.
- Hendrix, G. (1978). "Developing a Natural Language Interface to Complex Data." *ACM Trans. Database System* 3(2): 105-147.
- Hendrix, G. (1982). "Natural Language Interface (panel)." *Computational Linguistics* 8(2): 55-61.
- Hermanns, H. (2008). "Mobile Democracy: Mobile Phones as Democratic Tools." *Politics* 28(2): 74-82.
- Hubbard, P., R. Kitchin, et al. (2008). *Key texts in human geography*. London, SAGE.
- Hubel, D. H. and T. N. Wiesel (1965). "Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat." *Journal of Neurophysiology* 28: 229-289.
- Huff, D. (1954). *How to lie with statistics*. London, Gollancz.
- Igarashi, T. and K. Hinckley (2000). "Speed-Dependent Automatic Zooming for Browsing Large Documents." *UIST'00*: 139-148.
- Itti, L., C. Koch, et al. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Itti, L. and C. Koch (2001). "Computational Modelling of Visual Attention." *Nature Reviews Neuroscience* 2(3): 194-203.

- Jacoby, W. G. (1997). *Statistical graphics for univariate and bivariate data*. Thousand Oaks, Calif.; London, Sage.
- Jacoby, W. G. (1998). *Statistical graphics for visualizing multivariate data*. Thousand Oaks, Calif. ; London, Sage.
- Jacoby, W. G. (2006). "The Dot Plot: A Graphical Display for Labelled Quantitative Values." *Political Methodologist* 14(1): 6-14.
- Jacoby, W. and S. Schneider (2010). *Graphical displays for political science. The Visions in Methodology Conference*. University of Iowa.
- Jarvis, C. (2011). "Cartography: visualization of spatial data, 3rd edition." *International Journal of Geographical Information Science* 25(7): 1214-1215.
- Jenkins, J., D. Neale, et al. (1967). "Differential memory for picture and word stimuli." *Journal of Educational Psychology*, 58: 303-307.
- Jenny, H., B. Jenny, et al. (2010). "Interactive Design of 3D Maps with Progressive Projection." *Cartographic Journal* 47(3): 211-221.
- Johnston, L., V. Lovitskii, et al. (2008). *Personalized Question- Answering Mobile System*. International Book Series Information Science and Computing. Varna, Artificial Intelligence Research Institute. *Advanced Research in Artificial Intelligence*: 123-132.
- Johnston, R., C. Pattie, et al. (2001). *From Votes to Seats: The Operation of the UK Electoral System since 1945*. Manchester, Manchester University Press.
- Johnston, R., C. Pattie, et al. (2012). "Somewhat more disruptive than we had in mind': The Boundary Commission for England's 2011 Proposed Redistribution of Parliamentary Constituencies." *Political Quarterly* 83(1): 44-59.
- Jones, M., G. Marsden, et al. (1999). "Improving Web interaction on small displays." *Computer Networks-the International Journal of Computer and Telecommunications Networking* 31(11-16): 1129-1137.

- Kadman, N. and E. Shlomi (1978). "A Polyfocal Projection for Statistic Surfaces." *The Cartographic Journal* 15(1): 36-41.
- Kastellec, J. P. and E. L. Leonin (2007). "Using Graphs instead of Tables in Political Science." *Perspectives on Politics* 5: 755-771.
- Katz, J. N. and G. King (1999). "A statistical model for multiparty electoral data." *American Political Science Review* 93(1): 15-32.
- Keim, D. A. and A. Herrmann (1998). "The Gridfit algorithm: An efficient and effective approach to visualizing large amounts of spatial data." *Visualization '98, Proceedings*: 181-192.
- Keim, D. A., S. C. North, et al. (2002). *Efficient cartogram generation: A comparison*.
- Keim, D., C. Panse, et al. (2002). *Cartodraw: A fast algorithm for generating contiguous cartograms*. Information Visualization Research Group. Florham Park, AT&T Laboratories.
- Keim, D., C. Panse, et al. (2005). "Medial-Axis-Based Cartograms." *IEEE Computer Graphics and Applications* 25(3): 60-68.
- Kelleher, C. and T. Wagener (2011). "Ten guidelines for effective data visualization in scientific publications." *Environmental Modelling & Software* 26(6): 822-827.
- King, G., M. Tomz, et al. (2000). "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(2): 347-361.
- Koch, K., J., J. McLean, et al. (2006). "How Much the Eye Tells the Brain." *Current Biology* 16: 1428-1434.
- Kosara, R. (2007). *Visualisation Criticism — the missing link between information visualization and art*. The 11th International Conference on Information Visualisation IEEE CS Press.

- Kraak, M. J. and F. Ormeling (2010). *Cartography : visualization of geospatial data*. Harlow, Prentice Hall.
- Kuhn, W. (1996). *Handling Data Spatially: Spatializing User Interfaces* . 7th International Symposium on Spatial Data Handling, SDH'96, Delft, The Netherlands, IGU.
- Lamping, J. and R. Rao (1996). "The Hyperbolic Browser: A Focus + Context Technique for Visualizing Large Hierarchies." *Journal of Visual Languages and Computing* 7(1): 33-55.
- Lee, M. and D. Vickers (1998). *Psychological Approaches To Data Visualisation*, DSTO Electronics and Surveillance Research Laboratory, Salisbury, Australia.
- Lester, P. (2011). "Syntactic Theory of Visual Communication." California.
- Leung, Y. and M. Apperley (1994). "A Review and Taxonomy of Distortion-Oriented Presentation Techniques." *ACM Transactions on Computer-Human Interaction* 1(2): 126-160.
- Lichtner, W. (1983). "Computerunterstützte Verzerrung von Kartenbildern bei der Herstellung thematischer Karten." *Internationales Jahrbuch für Kartographie* XXIII: 83-96.
- Lindholm, E., J. Nickolls, et al. (2008). "NVIDIA Tesla: A unified graphics and computing architecture." *Micro, IEEE* 28(2): 39-55.
- Long, E., V. Lovitskii, et al. (2009). "Mobile Election." *Information Science and Computing, Intelligent Processing* 9: 19-28.
- Long, E., V. Lovitskii, et al. (2010). *Natural Interface to Election Data*. ITHEA, Sofia, Bulgaria.
- Long, E., V. Lovitskii, et al. (2011). *Election Data Visualization*. ITHEA. Sofia, Bulgaria.
- Lopez, V., M. Pasin, et al. (2005). *Aqualog: an ontology-portable question answering system for the semantic web.*, The European Semantic Web Conference (ESWC).

- Louw, E. (2005). *The Media and the Political Process*. Los Angeles, Sage.
- Lovitskii, V. and K. Wittamore (1997). DANIL: Databases Access using a Natural Interface Language. International Joint Conference on Knowledge-Dialogue-Solution: KDS-97, Yalta, Ukrain.
- Lovitskii, V., M. Thrasher, et al. (2008). "Automated Response To Query System." *International Journal of Information Theories & Applications* 15(2): 143-152.
- Macdonald, A. (2003). Modern-day map-making may be a way out of Web design's stasis *Eye Magazine*. Autumn.
- MacEachren, A. M. (1994). *Some truth with maps : a primer on symbolization and design*. Washington, D.C., Association of American Geographers.
- MacEachren, A. M. (1998). "Cartography, GIS and the World Wide Web." *Progress in Human Geography* 22(4): 575-585.
- Maceachren, A. M. and M. J. Kraak (1997). "Exploratory cartographic visualization: Advancing the agenda." *Computers & Geosciences* 23(4): 335-343.
- Mackinlay, J. (1986). "Automating the design of graphical presentations of relational information." *ACM Transactions on Graphics* 5(2): 110-141.
- Mackinlay, J., S. Card, et al. (1990). "Rapid, Controlled Movement through a Virtual 3d Workspace." *Computer Graphics* 42: 1971-1976.
- Mackinlay, J., G. Robertson, et al. (1991). "The Perspective Wall: Detail and Context Smoothly Integrated." *CHI91*: 173-179.
- Masser, I. (1997). "The national spatial data infrastructure in perspective." *GIS/LIS '97 Cincinnati*: 717-723.
- Medin, D. and A. Ortony (1989). *Psychological essentialism, Similarity and analogical reasoning*. S. Vosniadou and A. Ortony. New York, NY, Cambridge University Press. : 179-195.

- Michaelsen, M. (2011). *New Media vs. Old Politics*. Friedrich-Ebert-Stiftung (FES), Berlin.
- Minock, M. (2005). A phrasal approach to natural language interfaces over databases. *Proceedings of NLDB'05*.
- Mirzoeff, N. (1999). *An Introduction to Visual Culture*, New York, Routledge.
- Monmonier, M. (1977). "Nonlinear reprojection to reduce the congestion of symbols on thematic maps, ." *The Canadian Cartographer*, 14(1): 35-47.
- Monmonier, M. (1992). "Summary Graphics For Integrated Visualization In Dynamic Cartography." *Cartography and Geographic Information Systems* 19(1): 23-36.
- Monmonier, M. (1996). "Temporal generalization for dynamic maps. ." *Cartography and Geographic Information Systems*, 23(2): 96-98.
- Monmonier, M. (2007). "Cartography: the multidisciplinary pluralism of cartographic art, geospatial technology, and empirical scholarship." *Progress in Human Geography* 31(3): 371-379.
- Morrison, J. (1974). "A theoretical framework for cartographic generalization with the emphasis on the process of symbolization." *International Yearbook of Cartography* 14: 115-127.
- Muehrcke, P. (1980). *Maps in Geography. Maps in Modern Geography. Geographical Perspectives in the New Cartography*. L. Guelke. Monography 27, *Cartographica*, Vol.18, : 1-41.
- Nightingale, F. (1859). *Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army*. London.
- Orford, S., C. Rallings, et al. (2008). "Investigating differences in electoral turnout: the influence of ward-level context on participation in local and parliamentary elections in Britain." *Environment and Planning A* 40(5): 1250-1268.

- Orford, S., C. Rallings, et al. (2009). "Electoral salience and the costs of voting at national, sub-national and supra-national elections in the UK: a case study of Brent, UK." *Transactions of the Institute of British Geographers* 34(2): 195-214.
- Orford, S., C. Rallings, et al. (2011). "Changes in the probability of voter turnout when resiting polling stations: a case study in Brent, UK." *Environment and Planning C: Government and Policy* 29(1): 149-169.
- Paivio, A. (1971). *Imagery and verbal processes*. New York; London, Holt, Rinehart and Winston.
- Perkins, C. (2003). "Cartography: mapping theory." *Progress in Human Geography* 27(3): 341-351.
- Petchenik, B. (1983). *A Map Maker's Perspective on Map Design Research 1950-1980*. *Graphics Communication and Design in Contemporary Cartography*,. D. R. F. Taylor. New York, John Willey & Sons: 37-65.
- Plaisant, C., J. Grosjean, et al. (2002). "Spacetree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation." *InfoVis'02 Boston MA*: 57-64.
- Playfair, J. (1784). *A system of chronology*. Edinburgh.
- Playfair, W. (1786). *Commercial and Political Atlas*.
- Playfair, W. (1801). *Statistical Breviary*. London, Wallis.
- Popescu, A., O. Etzioni, et al. (2003). *Towards a theory of natural language interfaces to databases*. *Proceedings IUI'03*.
- Potter, M. (1993). "Very short-term conceptual memory." *Memory & Cognition* 21(2): 156-161.

- Rallings, C. and M. Thrasher (1996). *Forecasting Vote and Seat Share in Local Elections. British Elections & Parties Yearbook 1996*. D. M. Farrell, D. Broughton, D. Denver and J. Fisher. London, Frank Cass & Co.
- Rallings, C. and M. Thrasher (1999). "Local votes, national forecasts - using local government by-elections in Britain to estimate, party support." *International Journal of Forecasting* 15(2).
- Rallings, C. and M. Thrasher (2007). *British electoral facts, 1832-2006*. Aldershot ; Burlington, VT, Ashgate.
- Rallings, C., M. Thrasher, et al. (2011). "Forecasting the 2010 general election using aggregate local election data." *Electoral Studies* 30(2): 269-277.
- Rayner, K. (1998). "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin* 124: 374-422.
- Riesenhuber, M. and T. Poggio (1999). "Hierarchical models of object recognition in cortex." *Nature Neuroscience* 2(11): 1019-1025.
- Robertson, G., J. Mackinlay, et al. (1991). "Cone Trees: Animated 3D Visualizations of Hierarchical Information. ." *CHI'91*: 189-194.
- Robinson, A. (1952). *The Look of Maps*,. Madison, University of Wisconsin Press.
- Rohwer, W., S. Lynch, et al. (1967). "Pictorial and verbal factors in the efficient learning of paired associates." *Journal of Educational Psychology* 58: 278-284.
- Roth, R. E., A. W. Woodruff, et al. (2010). "Value-by-alpha Maps: An Alternative Technique to the Cartogram." *Cartographic Journal* 47(2): 130-140.
- Saito, M. (1996). "A Comparative-Study of Color Preferences in Japan, China and Indonesia, with Emphasis on the Preference for White." *Perceptual and Motor Skills* 83(1): 115-28.

- Sarkar, M. and M. Brown (1994). "Graphical fisheye views." *Communications, Association of Computer Machinery* 37(12): 73-84.
- Serre, T., A. Oliva, et al. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*.
- Shannon, C. and W. Weaver (1949). *The Mathematical Theory of Communication*. Urbana, University of Illinois Press.
- Shepard, R. (1967). "Recognition memory for words, sentences and pictures'." *Journal of Verbal Learning and Verbal Behaviour* 6: 156-163.
- Shepard, R. (1975). Form, formation, and transformation of internal representations, in *Information processing and cognition: The Loyola symposium*. e. (R. L. Solso. Potomac, MD, Erlbaum: 87-122.
- Shepard, R. (1981). Psychophysical complementarity. *Perceptual organization*. M. Kubovy and J. R. Pomerantz. Hillsdale, NJ, Erlbaum,: 279-341.
- Shepard, R. (1984). "Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming,." *Psychological Review* 91(4): 417-447.
- Shneiderman, B. (1996). *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. *Proc. IEEE Visual Languages*.
- Slingsby, A., J. Dykes, et al. (2010). "Rectangular Hierarchical Cartograms for Socio-Economic Data." *Journal of Maps*: 330-345.
- Souley, B. and R. H. Wicks (2005). "Tracking the 2004 presidential campaign Web sites - Similarities and differences." *American Behavioral Scientist* 49(4): 535-547.
- Spence, I. and J. Lewandowsky (1990). *Graphical Perception. Modern Methods of Data Analysis*. . J. Fox and J. Scott Long. Newbury Park, CA, Sage.

- Spence, R. (2007). *Information Visualization: Design for Interaction* Prentice Hall; 2nd edition (January 15, 2007)
- Spiess, E. (1970). "Characteristics of Combinations of Graphic Variables." *Mitteilungen Der Osterreichischen Geographischen Gesellschaft* 112(2-3): 324-338.
- Standing, L., J. Conezio, et al. (1970). "Perception and memory for pictures: single trial learning of 2,500 visual stimuli." *Psychonomic Science* 19: 73-74.
- Stanyer, J. (2006). "Online campaign communication and the phenomenon of blogging - An analysis of web logs during the 2005 British general election campaign." *Aslib Proceedings* 58(5): 404-415.
- Stanyer, J. (2007). *Modern Political Communication*. Cambridge, Polity Press.
- Stone, D. L. and D. Stone (2005). *User Interface Design And Evaluation*, Morgan Kaufmann.
- Stone, L. S., F. A. Miles, et al. (2002). "Linking eye movements and perception." *Journal of Vision* 3(11): 1-11.
- Sun, H. and Z. H. Li (2010). "Effectiveness of Cartogram for the Representation of Spatial Data." *Cartographic Journal* 47(1): 12-21.
- Super, B. (2009). "The Internet Election: Perspectives on the Web in Campaign 2004." *Political Studies Review* 7(2): 300-301.
- Tennant, H., K. Ross, et al. (1983). *Menu-Based Natural Language Understanding*. The 21st Annual Meeting of ACL, Cambridge, Massachusetts, Proceedings of the 21st Annual Meeting of ACL.
- Tegarden, D. (1999). *Business Information Visualization*. *Communications of AIS* 1(4).
- Thrasher, M., G. Borisyuk, et al. (2012). "Voting systems in parallel and the benefits for small parties: An examination of Green Party candidates in London elections " *Party Politics*.

- Tobler, W. (2004). "Thirty Five Years of Computer Cartograms,." *Annals of the Association of American Geographers*, 94: 58-73.
- Torralba, A., A. Oliva, et al. (2006). "Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search." *Psychological Review* 113(4): 766-786.
- Tovée, M. J. (1996). *An introduction to the visual system.* . Cambridge, Cambridge University Press.
- Trainor, T. (1997). *The Census Bureau's electronic data access and dissemination system: new challenges for cartographic applications.* ICA/ACI International Cartographic Conference (ICC 97),. Stockholm: 541-547.
- Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, Conn, Graphics Press.
- Tufte, E. R. (1990). *Envisioning information.* Cheshire, Conn. , Graphics Press.
- Tufte, E. R. (2006). *Beautiful evidence.* Cheshire, Graphics Press.
- Tukey, J. (1977). *Exploratory Data Analysis.* Reading, MA, Addison-Wesley.
- Turgenev, I. S. (2012). *Fathers and sons.* Oxford, Oxford University Press.
- Upton, G. (1976). "Diagrammatic Representation of Three Party Contests." *Political Studies* 24: 448-454.
- Upton, G. (1991). "Displaying Election Results." *Political Geography Quarterly* 10(3).
- Vallez, M. and R. Pedraza-Jimenez (2007) "Natural Language Processing in Textual Information Retrieval and Related Topics."
- Van de Donk, W., Loader, B., Nixon, P., Rucht, D. (2004): *Cyberprotest: New Media, Citizens and Social Movements.* London/New York: Routledge.

- Van Essen, D. and J. L. Galant (1994). "Neural mechanisms of form and motion processing in the primate visual system." *Neuron* 13(1): 1-10.
- Verser, R. and R. H. Wicks (2006). "Managing voter impressions: The use of images on presidential candidate Web sites during the 2000 campaign." *Journal of Communication* 56(1): 178-197.
- von der Marlsburg, C. (1981). *The correlation theory of brain function*. Gottingen, West Germany, Max Planck Institute for Biophysical Chemistry, Dept. of Neurobiology. Internal Report 83-2.
- von der Marlsburg, C. (1999). "The what and why of binding: The modeler's perspective." *Neuron* 25: 95-104.
- Wainer, H. (2000). *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot*. New Jersey, LEA.
- Walts, D. (1978). "An English Language Question Answering System for a Large Relational Database." *Communications of the ACM* 21(7): 526-539.
- Ware, C. (2004). *Information Visualization - Perception for Design*, Morgan Kaufmann.
- Ware, C. (2005). *Visual Queries: The Foundation of Visual Thinking*. Knowledge and Information Visualization, . S. Tergan and T. Keller. Berlin Heidelberg . Springer-Verlag: 27 – 35.
- Wicks, R. H. and B. Souley (2003). "Going negative: Candidate usage of Internet web sites during the 2000 presidential campaign." *Journalism & Mass Communication Quarterly* 80(1): 128-144.
- Wiener, N. (1948). *Cybernetics; or, Control and communication in the animal and the machine*. New York, J. Wiley.
- Wijk, J. (2005). "The value of visualization." *Proceedings of IEEE Visualization (VIS '05)*: 78-86.

Yarbus, L. (1967). Eye movements and vision. New York, Plenum Press.

Yau, N. (2011). Visualize this : the flowing data guide to design, visualization, and statistics.
Indianapolis, Wiley.

12 Bound in Copies of Publications

12.1 Mobile Election

Long, E., V. Lovitskii, M. Thrasher, and D. Traynor. 2009. "Mobile Election." *Information Science and Computing, Intelligent Processing 9*: 19-28.

MOBILE ELECTION

Elena Long, Vladimir Lovitskii, Michael Thrasher, David Traynor

Abstract: *Mobile phones have the potential of fostering political mobilisation. There is a significant political power in mobile technology. Like the Internet, mobile phones facilitate communication and rapid access to information. Compared to the Internet, however, mobile phone diffusion has reached a larger proportion of the population in most countries, and thus the impact of this new medium is conceivably greater. There are now more mobile phones in the UK than there are people (averaging at 121 mobile phones for every 100 people). In this paper, the attempt to use modern mobile technology to handle the General Election, is discussed. The pre-election advertising, election day issues, including the election news and results as they come in, and answering questions via text message regarding the results of current and/or previous general elections are considered.*

Keywords: *mobile text messages, mobile election, mobile advertising, question-answering system*

ACM Classification Keywords: *1.2 Artificial intelligence: 1.2.7 Natural Language Processing: Text analysis.*

Conference: *The paper is selected from Second International Conference "Intelligent Information and Engineering Systems" INFOS 2009, Varna, Bulgaria, June-July 2009*

Introduction

In our previous papers [1,2] the Question-Answering Mobile Engine (QAMEN) has been represented. QAMEN is based on industry-standard SMS messaging technology and thus works with any mobile in any GSM network. In this paper QAMEN will be restricted by General Elections issues. For clarity we will distinguish two versions of QAMEN for election: PC version – QAMEN-E_{PC} and Mobile version QAMEN-E_{MBL} and only the **Election** Application Domain (AD) will be considered.

While it is too early to determine the political effects of mobile phone diffusion, the political events in different countries suggest that mobile technology may come to play an important role in political participation and democracy. Text messaging has already played an important role in the 2008 United States campaign of Barack Obama. Obama's text message announcing his vice-presidential candidate selection of Joe Biden reached approximately 2.9 million Mobile Subscribers (MS), many of whom signed-up with the promise of receiving vital information via text messaging on their mobile phone.

We are definitely against blind copying the digital techniques used in the 2008 US presidential election. One disincentive is the short campaigning time of UK general elections. US presidential campaigns are two years long and so the candidates were able to test different digital media. But with UK general election campaigns only three and a half weeks long, UK parties will have to get it right, first time. Planners argue that use of digital media will come down to the demographics of swing voters. It depends on whether the parties believe that the next election will be won by retaining the loyalty of the over-50s or at the margin of first-time and under-25 voters, most of whom are mobile-literate.

Despite the differences, the growth of the importance of technology in politics cannot be ignored. UK politicians are still learning the new tone, style and dynamics needed for mobile campaigning. Politicians need to see it as an opportunity to express themselves, rather than a threat; i.e. politicians need to overcome their fear of the unfamiliar and embrace the opportunities it provides.

The Electoral Commission has warned that the UK electoral system continues to operate under 19th century structures and requires urgent reform before the next general election,. "The planning and running of elections need to be more robust and coordinated," said Sam Younger, chairman of the Electoral Commission. "We are still trying to run 21st century elections with 19th century structures, and the system is under severe strain." [3].

In the UK, legislation currently prevents electronic voting in general elections, but the technology has been well tested in 17 pilot projects during local and European elections. These trials were funded by the Office of the Deputy Prime Minister, which is also investing £12m into Core (co-ordinated online register of electors), a project that aims to modernise electoral rolls - a crucial step in enabling national e-voting systems.

The biggest e-voting trial conducted so far in the UK was in Sheffield, where 174,000 citizens were given the opportunity to vote using the internet, kiosks and mobile phones. Voters in the city were given the choice of voting using a traditional ballot paper, a mobile phone text message, a touch-tone telephone, a website or a touch-screen internet kiosk at a polling station.

The Electoral Commission wants e-voting technology to be proven before it will give the green light to an electronic general election, but proving the technology is difficult without a large-scale trial. The Office of the Deputy Prime Minister says there will be no large-scale e-voting projects until "issues of secrecy, security and technological penetration have been addressed". This makes it into a **vicious circle**. In this paper the way of breaking such a circle will be discussed. QAMEN-E_{MBL} will be represented as an election monitoring system. The purpose of this paper is to describe the main stages of mobile election and discuss the ways of mobile election monitoring.

Reading this paper will tell you the following:

- • Mobile election overview.
- • Pre-election advertising.
- • The Election Day.
- • Mobile Election Results.
- • QAMEN-E_{MBL} Versus QAMEN-E_{PC}.
- • Mobile Request Processing.

Mobile Election Overview

- The first local and mayoral electoral vote in the UK by text message took place on 23rd May 2002 [4].
- Korean campaigners sent text messages to encourage voting in 2002, while the Italian government in 2004 sent a reminder message about upcoming European elections to all mobile phone subscribers [5].
- Mr. Tony Blair became the first UK Prime Minister to use text message technology to talk directly to the people on 25th November 2004 [4], answering questions submitted in advance by text message from members of the public as well as in real-time in a mobile phone chat-room, transmitted live from No.10 Downing Street.
- The Spanish general election of 2004 occurred in the wake of an unprecedented terrorist attack, but its outcome reflects the potential that mobile phones have to provide the user with independent information and bring about voter mobilisation. Mobile technology contributed to the quality of democratic practices in the sense that SMS messages helped provide citizens with more information about the rapidly unfolding events related to the terrorist attacks, including the reaction by the government and the opposition party as well as the investigation and the protests during the "day of reflection" [6].
- The information about the first mobile phone election in the UK had been announced in April 2005 [7]. In the UK, the General Election always sparked a frenzied dash for information as voters tried to keep up to date with hot election issues. As with all major news stories, people want to be able to follow events as they unfold and the mobile phone is proving to be the ideal tool for this.
- On May 21, 2006 the Montenegro, a small country in the former Yugoslavia, saw the first instance of volunteer monitors using SMS, as their main election reporting tool. A Montenegrin NGO (Non-Governmental Organization) was the first organization in the world to use text messaging to meet all election day reporting requirements.

- Since then, mobile phones have been deployed in six elections in countries around the world, with volunteers systematically using text messaging in election monitoring. SMS monitoring is becoming a highly sophisticated rapid-report tool, used not just in a referendum election like in Montenegro, but in parliamentary elections with a plethora of candidates and parties and complex data reported via SMS. This was the case in Bahrain, a small country in the Middle East, where monitors reported individual election tallies in a series of five to forty concurrent SMS messages, using a sophisticated coding system, with near accuracy.
- In the two years since the first large-scale SMS monitoring in Montenegro, there have been rapid improvements in mobile services as competition in the wireless industry has increased worldwide, and there is growing interest and understanding on the part of NGOs that systematic election monitoring is not as difficult as it first may seem. As election monitoring via SMS becomes standardized and NGOs gain experience, there is no reason for mobile phones and SMS not to play a greater role in other areas of civic participation.
- Election in Sierra Leone [8] is lead by the National Election Watch (NEW), a coalition of over 200 NGOs in the country. NEW has monitors at 500 of the 6171 polling stations. Monitors report on whether there are any irregularities via SMS back to headquarters.
- The Estonian Government has passed a new bill that will allow its citizens to vote using their mobile phones in the next election [9]. The measure will come into effect for the 2011 election, and makes Estonia the first country in the world to approve such a method. In order to vote using their mobile phone, Estonians will be required to purchase a special chip for their handsets which will verify the voter's identity and authorise their vote on the system.

The future is bright for innovative ways in which mobile phones are used by citizens to participate and engage in their countries as the mobile revolution unfolds. In fact, the 2008 US presidential election saw the widest possible mix of offline and online media used to help candidates connect with some 200 million voters scattered across America's vast expanse. Experts are predicting that in future elections, the use of mobile technology will become the standard and play an even larger role than in recent US election where Barack Obama's use of text messaging is already being held as a successful way to reach out to the constituents.

Mobile campaigning needs a new political vocabulary, style and humour - far removed from the stultifying prose of traditional party literature. Mobile is all about personalized content; offering politicians a way of communicating with voters that is simply not possible with television or newspapers.

Pre-election Advertising

Election advertising is defined as any content that can reasonably be regarded as attempting to gain electoral success for any candidate or political party which seeks to increase their status or position. Election advertising is anything that can reasonably be regarded as influencing a voter in a particular direction, and it's a media-neutral definition. Election advertising requires advertiser identification, and spending limits apply. A message supporting someone else needs the beneficiary's written permission. The Election Advertising Campaign (EAC) will provide a unique perspective for both candidate or political party, and MS. EACs goal is to get more votes, more attention and more exposure for the General Election 2010 in the UK.

What can the General Election campaign, political parties, candidates and Registered Mobile Subscribers (RMS) expect from EAC?

- Increased voter turnout due to text message reminders.
- Votes by text could increase turnout, especially a potential record number of young voters. 95% of 16-24 year olds use text messaging regularly, each sending an average of 400 texts per month. Young Britons would be far more willing to vote in the General Elections if the government were to introduce voting by mobile phone. ICM Research found a huge 73% of 18 to 24-year-olds would have voted if they had the option of doing so by mobile phone. The opportunity of being able to cast a vote via text message would make people more likely to participate in an election.

- Political parties, for instance, might invite party members and ordinary citizens to participate in the pre-election candidate nomination process by mobile phone voting.
- Candidates will be able to use text messages to collect campaign contributions. Mobile search is the ideal opportunity for candidates to position themselves about key issues like jobs, schools, knife crime, etc.
- RMS will be able to receive text messages from their candidate or party of choice. This is another sign of how political outreach could try to keep up with the changing nature of personal communication.
- Every evening at 8pm a text summary of the key political events of the day and the latest election news will be sent to the RMS.
- Receiving a reminder for scheduled and special elections events could become the norm.
- Be the first to know with breaking news alerts.
- EAC keeps RMS abreast of the fast changing political scene in the final days before the General Election.
- RMS might be asked from time to time for whom they are likely to vote, to define the current situation of candidates popularity as a projection of the election results.

Registration as RMS is very simple:

- Type in your name, post code and password.
- Send this message to 5-digit short code. It should be memorable short code like ELECT (35328), or 2ergo (23746). It is important to note that these short codes should only be considered as examples.
- Confirmation about registration will be sent as a text message to the RMS mobile phone along with a Pin number (required for election day only) and notes about what kind of information will be regularly sent to the RMS mobile phone. All information, of course, needs to be suitable for the limited text messaging format of 160 characters.
- Simply send a text message CANCEL to 35328, or 23746 and you will be unsubscribed from all services.

For non-registered MS it would be enough to text NEWS to 35328, or 23746 for breaking news alerts, or text EVENTS to the same short codes for the nearest election events, or text RESULT for actual election results.

In order for EAC to succeed, a number of conditions have to be in place:

- The power of the message lies beyond 160 characters: One of the best tactics of wise political campaign managers has been the use of embedded links in each message that direct to a candidate's homepage.
- Political campaigns have utilized demographic information to target voters in particular cities, regions or with specific issues in mind. With location-based services becoming more commonplace on phones, the ability to deliver targeted messages will be one of the biggest enhancements to mobile messaging in the near future.
- QAMEN-E_{MBL} should be able to handle thousands of messages per minute to one mobile short code.
- The mobile provider (e.g. 2ergo for QAMEN-E_{MBL}) would give the highest priority to the election mobile short code.
- Note that each SMS message sent to QAMEN-E_{MBL} for this service will cost 10-15p (network charges vary).

An important consideration is the cost of a wide-scale program. An analysis of the last years Presidential Election in the United States,[10] allows the surmise that not only has text messaging surpassed actual calls as the most popular form of mobile communication but its cost-effectiveness in elections was astounding. For a political campaign, the acquisition cost of an "opt in" text user (a RMS in our case) is about 13 cents. By comparison, the acquisition cost of door-to-door canvassing is about \$2.50 and for phone calling, it's about \$1.00. The difference in actual cost-per-vote results is even more obvious: \$1.62 per vote for opt-in text lists compared to \$20-\$35 per vote from phone calls, leaflets and door-to-door visits.

However, the Sheffield trial did reveal some potential problems with e-voting. The biggest problem was cost. Offering voters access to secure and robust electronic voting channels more than doubled the cost of running the election, to £55 per voter, according to the Institute of Public Policy Research.

Mobile phone technology allows users to be directly accessible at all times and locations and that is why it would be more effective for political parties to use the power of mobile messaging, namely QAMEN-E_{MBL}, not only during the short campaigning time of UK general elections but the whole year.

The Election Day

RMS from around the country used the power of text messaging to cast their vote. Attendees cast their ballots by texting only the name of the candidate to the short code 2VOTE (28683), or ELECT, or 2ergo to take part in the election.

A note about security. RMS voting requires the entry of a password and pin. In a traditional voting system no formal identification document is required when voters present themselves at a polling station; and yet the same person may be required to present three forms of ID just to register at their local Blockbuster video store. Therefore QAMEN is a vast step forward in RMS security.

Arriving messages are passed to QAMEN-E_{MBL} and then the data is compiled in a database (DB) ready for analysis. It is amazing to see the numbers change on the screen as the SMS messages pour into the DB (see Figure 1). The RMS will be kept updated on the latest election news, latest announcements and sent reports on quantitative data such as real-time voter turnout and even on the actual election results. For non-registered MS it would be enough to text RESULT to 35328, or 23746 to get the latest election results anytime and anywhere.

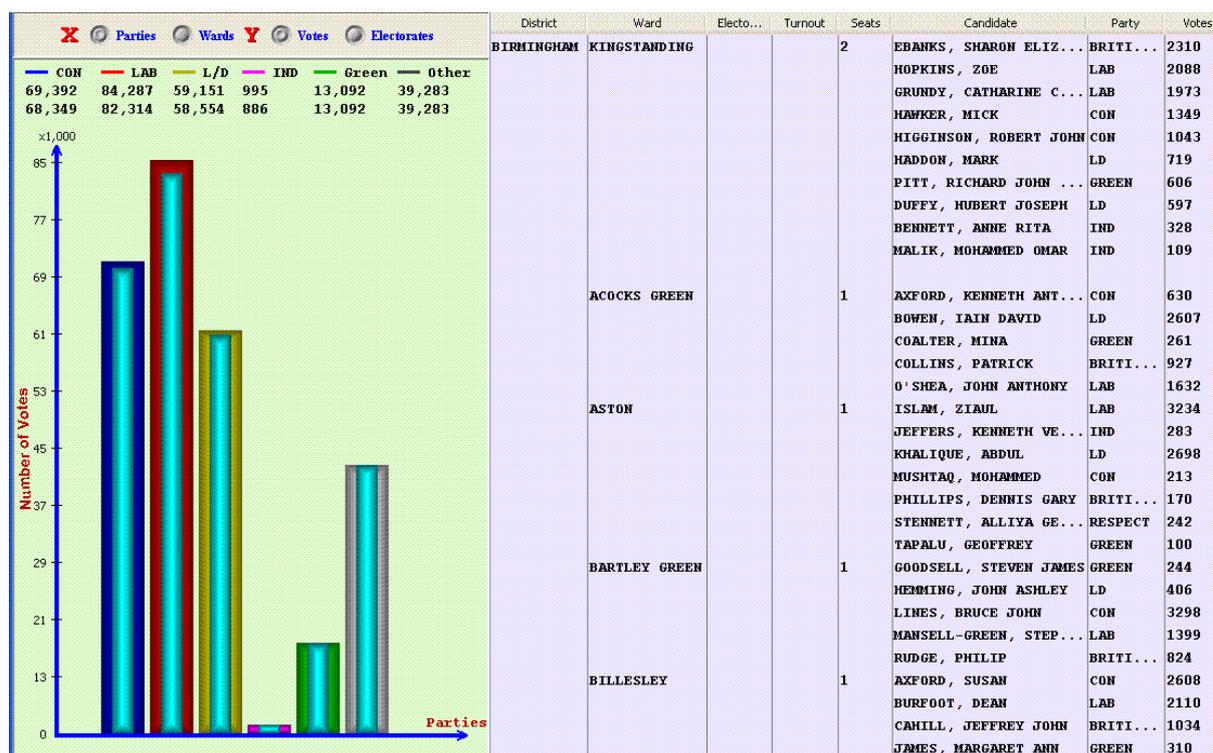


Figure 1. Current result of local election in Birmingham

Mobile Election Results

QAMEN-E_{MBL} is supposed to make nationwide election results available not only to RMS but to all mobile devices. MS can send any request regarding election results to 35328, or 23746 and QAMEN-E_{MBL} will reply via text message.

The visualization of election results can make it more intuitive and productive. People have always relied on visual tools such as maps, charts, and diagrams to better understand problems and solve them quicker. The mobility context and technical limitations such as a small screen size make it impossible to simply port visualization applications from desktop computers to mobile devices, but researchers are starting to address these challenges. Considerable effort is needed to understand how to design effective visualizations for mobile devices. Although many researchers have proposed specific techniques, no reports on the topic have yet provided a broad discussion of mobile visualization that could be useful to mobile application developers. In this paper only textual presentation of election results will be considered.

Mobile phones (and consequently QAMEN-E_{MBL}) have some limitations when compared to PCs (QAMEN-E_{PC}). These limitations are the problems that need to be taken into account when developing an acceptable mobile question-answering procedure.

QAMEN-E_{MBL} Versus QAMEN-E_{PC}

Compared to QAMEN-E_{PC}, QAMEN-E_{MBL} has many restrictions that have to be considered and overcome:

- Displays are very limited due to smaller screen size, the 160 characters SMS restriction and lower resolution.
- Input peripherals such as tiny keypads, micro joysticks, and rollers are often inadequate for complex tasks.
- Connectivity is slower, affecting interactivity when a significant quantity of data is stored on remote databases.
- QAMEN-E_{PC} might provide the powerful command prompt (see Figure 2).
- It is easy to use such a prompt to represent requests in clear, grammatical and correctly-spelt language. Often Mobile Request (MR) for QAMEN-EMBL will be ungrammatical. As a rule MSs:
 - Do not want to use upper case to type request like *“george bush, washington dc”* [11]. Or use dots to separate *“d”* and *“c”*.
 - Dropping *‘?’* at the end of MR.
 - Not using any punctuation at all.
 - Deletion of articles.
- The fact is that MR simply will not be spelt, punctuated, and capitalised correctly but the main requirement for QAMEN-EMBL is - to handle non-standard or poorly formed/structured (but, nevertheless, meaningful) user's MR.
- QAMEN-E_{PC} has no problems displaying the response in a convenient format for the user(see Figure 3) but for QAMEN-EMBL the small screen and 160 characters restrictions may cause some problems when it comes to displaying the response.

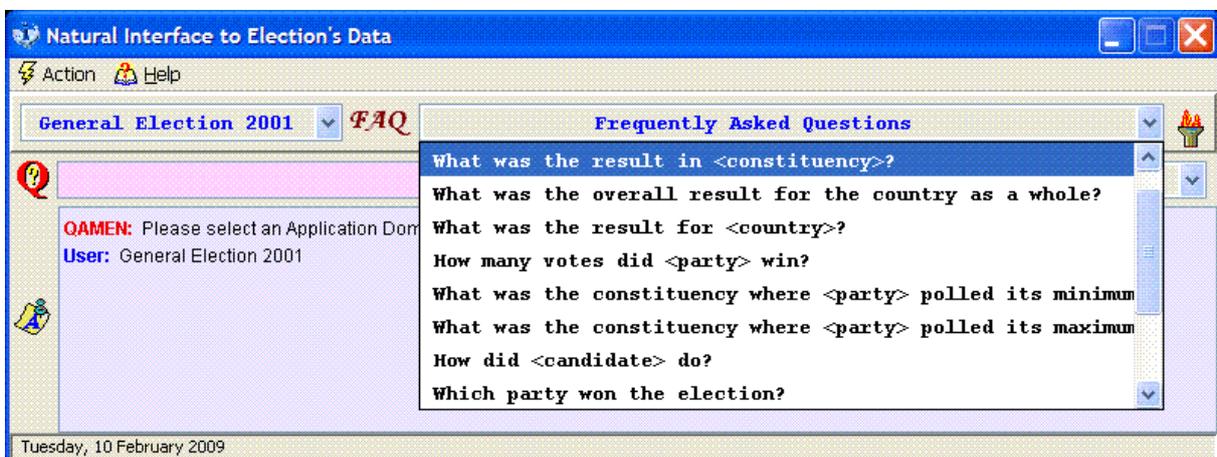


Figure 2. QAMEN-E_{PC}: Command prompt interface

The main conclusion from such comparison is QAMEN- E_{MBL} should be **more intelligent** than QAMEN- E_{PC} . More evidence for such conclusion comes from the fact that it is simply impossible to require the users to remember, for example, the exact name of the constituency in order to correctly ask what seems a very simple question: "Who won the election in Suffolk Central & Ipswich North in 2001?". It would be expected that the user instead of using the symbol '&' types in "**and**". In which case QAMEN- E_{PC} will not find the constituency in DB and will have to generate the clarification dialog:

QAMEN: Do you mean Suffolk Coastal, Suffolk South, or Suffolk West constituency?

User: No, I mean Suffolk Central.

QAMEN: Suffolk Central constituency does not exist but there is Suffolk Central & Ipswich North constituency.

User: It's exactly what I meant.

QAMEN: Thank you.

Theoretically QAMEN- E_{MBL} can do the same, but such dialog would not be acceptable due to time and money. But QAMEN- E_{MBL} is an intelligent system and in the result of similarity measurement [12] between MR *Suffolk Central and Ipswich North* and similar DB values namely: *Suffolk Central & Ipswich North*, *Suffolk Coastal*, *Suffolk South*, *Suffolk West* and *Ipswich* constituencies, QAMEN- E_{MBL} selects *Suffolk Central & Ipswich North* with great confidence.

In the case when user simply made a mistake and instead of typing in the desirable constituency *Hereford* in the MR: "Who won the election in Hereford" he/she entered *Hertford* (it's **wrong** but at the same time it's **right** from the QAMEN- E_{MBL} point of view because it has the right part of an existing constituency), QAMEN- E_{MBL} found an answer for the constituency *Hertford & Stortford*. When MS sees the response, he/she realises that MR was wrong and corrects it.

General Election 2001 **FAQ** **Frequently Asked Questions**

What was the result in <constituency>? **New** **Constituencies**

QAMEN: Please select an Application Domain.
User: General Election 2001
User: What was the result in Plymouth devonport?

General Election 2001

Constituency:	Candidates	Votes
Plymouth devonport		
Parties		
Lab	Jamieson, David	24,322
Con	Glen, John	11,289
LD	Baldry, Keith	4,513
UK Ind	Parker, Michael	958
Soc All	Staunton, Tony	334
Soc Lab	Hawkins, Rob	303

QAMEN: Any more questions?

Tuesday, 10 February 2009

Figure 3. QAMEN- E_{PC} : Question-Answering interface

Mobile Request Processing

The main purpose of MR processing is to understand *what was meant* rather than *what was said* and in the result of such, understanding and returning to MS only elections-relevant information. The mechanism of query phrasing is very simple: "eliminating the unnecessary until only the necessary remains" and has been discussed elsewhere [13]. Here we just remind ourselves of the main steps involved in MR processing.

- QAMEN-E_{MBL} takes the MR as a character sequence and converts the original MR to a *skeleton* by noisy (non-searchable) words elimination. In the result of such conversion MR will contain only **meaningful** words. Let's call word meaningful if it represents DB field descriptor or DB field value.
- AD election is represented by DB. DB **meaningful** fields (i.e. they don't represent primary or foreign keys) contain election data. Each meaningful fields has a list of descriptors. Between descriptors and meaningful fields exists an one-to-one attitude.
- The purpose of MR processing is to match MR meaningful words against the DB fields descriptors.
- To highlight the complexity of such matching, it is enough to consider quite a simple MR: "Who won an election in <constituency>?". Without knowing "who is who" and meaning of "won election" QAMEN-E_{MBL} cannot answer this question. To explain it to QAMEN-E_{MBL} the Production Rules (PR) need to be involved.

The subset of PR in format:

$$\langle \text{Precondition} \rangle \mapsto \langle \text{Antecedent} \rangle \Rightarrow \langle \text{Consequent} \rangle$$

is shown below.

1. AD:Election \mapsto who \Rightarrow candidate
2. AD:Election \mapsto [candidate]:<win \oplus won> \Rightarrow [SQL]:<MAX(votes)>
3. AD:Athletics \mapsto [runner]:<win \oplus won> \Rightarrow [SQL]:<MIN(time)>
4. AD:Athletics \mapsto [shooter]:<win \oplus won> \Rightarrow [SQL]:<MAX(distance)>
5. AD:Election \mapsto votes \Rightarrow [Field]:<CANDIDATE.VOTE>
6. AD:Election \mapsto candidate \Rightarrow [Field]:<CANDIDATE.[CANDIDATE NAME]>
7. AD:Election \mapsto party \Rightarrow [Field]:<CANDIDATE.PARTY>
8. AD:Election \mapsto [party]:<win \oplus won> \Rightarrow [SQL]:<TOP1, SUM(votes), DESC>

where \oplus - denotes "exclusive OR".

Precondition consist of **class₁:value₁ {& class_i:value_i}**.

Antecedent might be represented by: (i) **single word** (e.g. *who, won, August, seven, etc.*), (ii) **sequence of words** (e.g. *as soon as, create KB, How are you doing, etc.*), or (iii) **pair - [context]:<value>**.

Context allows one to avoid word ambiguity and thereby distinguish difference between "Candidate won an election" and "Party won an election".

Presentation of **Consequent** is similar to Antecedent structure except (iii). For *Consequent pair* represents **[descriptor]:<value>**.

For AD *Election* subset (1, 2, 5..8) of PR is used. PR 3 and 4 in fact show another meaning of the same word *won* but for a different AD.

- The final step of a phrased MR to SQL query conversion is quite complicated because it is necessary to access data from many different tables within an AD and join those tables together in SQL query. Result of conversion of MR "Which party won the election" to SQL-query using selected PR and executing of produced SQL query is shown in Figure 4.

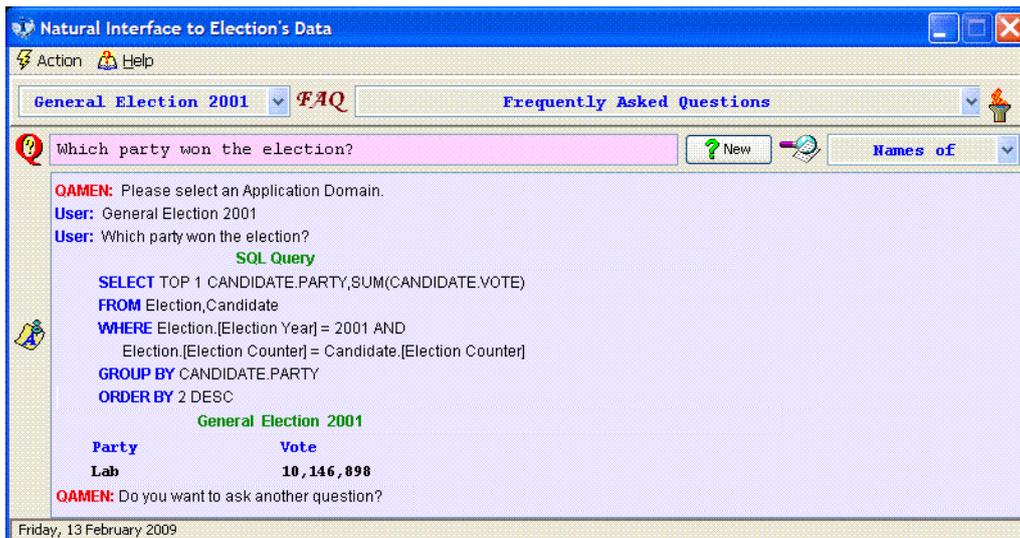


Figure 4. QAMEN-PC: Request Parsing: SQL query creation

Using QAMEN-EMBL for testing is quite expensive and that is why QAMEN-EWEB, which is oriented toward AD Election, was used. Because AD is predetermined, for QAMEN-EMBL (QAMEN-EWEB) it would be easy to understand what the MS meant by the entered MR. Instead of type in the MR "What was the result in Plymouth Devonport constituency in 2001 General Election?" it would be enough for MS to text in just "Plymouth Devonport 2001". By the way, exactly these three words represent the result of initial MR processing. After conversion to SQL query and running it result will be displayed in the mobile format (see Figure 5) i.e. QAMEN-EMBL (QAMEN-EWEB) is trying to minimize the space for response (compare with Figure 3).

Conclusion

Like any technology, mobile telecommunication can have a wide variety of effects on political behaviour and practices, and the fact that it has been around for such a short period of time makes it impossible to reach a general conclusion about its ultimate impact. Nevertheless, we hope that in future elections, the use of mobile technology will become the standard and play an even larger role, helping to make nationwide election campaigns available to all mobile users.

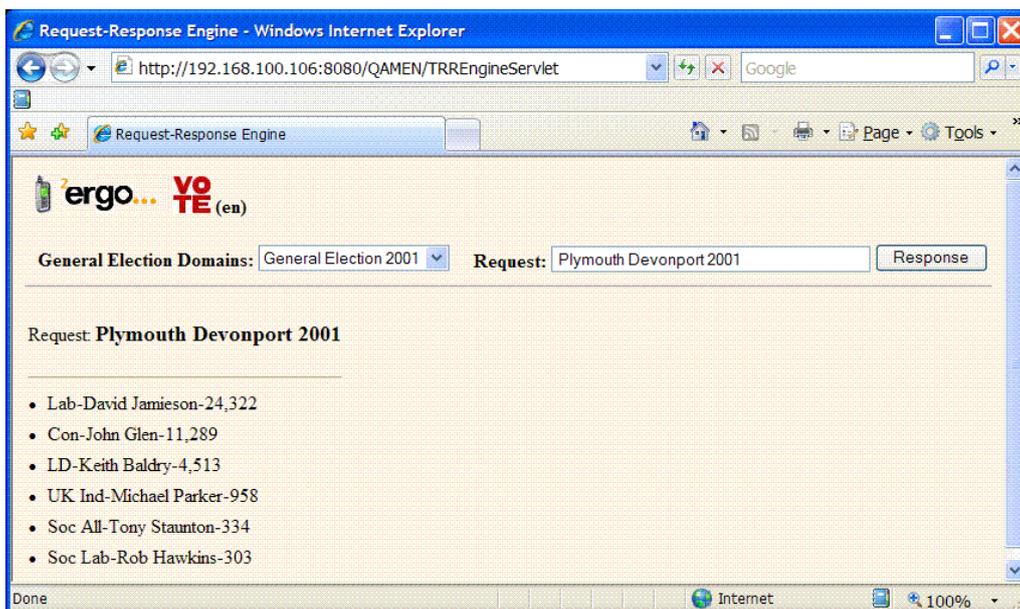


Figure 5. QAMEN-EWEB: Question-Answering interface

Without a doubt mobile phones have the potential to change certain aspects of political behaviour, including people's desire to participate in the political process. We believe that political parties should have their mobile campaign constantly, every year and not only during the three-and-a-half weeks of the election campaign every four or five years. Moreover, it would be very useful to have access to the results of any general election instantly via mobile phone. In our paper we tried to demonstrate that QAMEN-E_{MBL} can provide such services to mobile users.

Bibliography

- [1] Vladimir Lovitskii, Michael Thrasher, David Traynor, "Automated Response To Query System", *International Journal "Information Theories & Applications"*, Vol 15(2), 143-152, 2008.
- [2] Lee Johnston, Vladimir Lovitskii, Ian Price, Michael Thrasher, David Traynor, "Personalized Question-Answering Mobile System", International Book Series: "*Information Science and Computing*", book 2: "Advanced Research in Artificial Intelligence", 123-132, 2008.
- [3] http://news.bbc.co.uk/1/hi/uk_politics/7582671.stm
- [4] http://www.text.it/mediacentre/facts_figures.cfm
- [5] Heike Hermanns, "Mobile Democracy: Mobile Phones as Democratic Tools", *Journal "Politics"*, Vol 28(2), 74-82, 2008.
- [6] Sandra L. Suarez, "Mobile Democracy: Text Messages, Voter Turnout and the 2004 Spanish General Election", *Representation*, Vol 42(2), 117-128, 2006,
<http://www.temple.edu/polsci/suarez/documents/MobileDemocracy2006.pdf>
- [7] Sean Coughlan, "The first mobile phone election",
http://news.bbc.co.uk/1/hi/uk_politics/vote_2005/frontpage/4457723.stm
- [8] <http://mobileactive.org/texting-it-in>
- [9] <http://www.pcpro.co.uk/news/242835/election-to-be-decided-by-mobile-phone.html>
- [10] Jonathan Spalter, "How Mobile Technologies are Changing Elections",
http://www.huffingtonpost.com/jonathan-spalter/how-mobile-technologies-a_b_134996.html
- [11] www.smseducation.org
- [12] Ken Braithwaite, Mark Lishman, Vladimir Lovitskii, David Traynor, "Distinctive Features of Mobile Messages Processing", International Journal "*Information Theories & Applications*", Vol 14(2), 154-160, 2007.
- [13] T.Coles, V.A.Lovitskii, "Text Searching and Mining", *Journal of Artificial Intelligence*, National Academy of Sciences of Ukraine, Vol 3, 488-496, 2000.

Authors information

Elena Long – University of Plymouth, Plymouth, Devon, PL4 6DX, UK, e-mail: elena.long@plymouth.ac.uk

Vladimir Lovitskii – 2 Ergo Limited, 4th Floor, Digital World Centre, 1 Lowry Plaza, The Quays, Salford, Manchester, M50 3UB, UK, e-mail: vladimir.lovitskii@fsmail.net

Michael Thrasher – University of Plymouth, Plymouth, Devon, PL4 6DX, UK, e-mail: mthrasher@plymouth.ac.uk

David Traynor – 2 Ergo Limited, 4th Floor, Digital World Centre, 1 Lowry Plaza, The Quays, Salford, Manchester, M50 3UB, UK, e-mail: david.traynor@2ergo.com

1.1 Natural Interface to Election Data

Long, E., V. Lovitskii, and M. Thrasher. 2010. "Natural Interface to Election Data." Paper presented at the ITHEA, Sofia, Bulgaria.

Krassimir Markov, Vladimir Ryazanov,
Vitalii Velychko, Levon Aslanyan
(editors)

New Trends
in
Classification and Data Mining

I T H E A
SOFIA
2010

Krassimir Markov, Vladimir Ryazanov, Vitalii Velychko, Levon Aslanyan (ed.)
New Trends in Classification and Data Mining

ITHEA®

Sofia, Bulgaria, 2010

First edition

Recommended for publication by The Scientific Council of the Institute of Information Theories and Applications FOI ITHEA

This book maintains articles on actual problems of classification, data mining and forecasting as well as natural language processing:

- new approaches, models, algorithms and methods for classification, forecasting and clusterisation. Classification of non complete and noise data;
- discrete optimization in logic recognition algorithms construction, complexity, asymptotically optimal algorithms, mixed-integer problem of minimization of empirical risk, multi-objective linear integer programming problems;
- questions of complexity of some discrete optimization tasks and corresponding tasks of data analysis and pattern recognition;
- the algebraic approach for pattern recognition - problems of correct classification algorithms construction, logical correctors and resolvability of challenges of classification, construction of optimum algebraic correctors over sets of algorithms of computation of estimations, conditions of correct algorithms existence;
- regressions, restoring of dependences according to training sampling, parametrical approach for piecewise linear dependences restoration, and nonparametric regressions based on collective solution on set of tasks of recognition;
- multi-agent systems in knowledge discovery, collective evolutionary systems, advantages and disadvantages of synthetic data mining methods, intelligent search agent model realizing information extraction on ontological model of data mining methods;
- methods of search of logic regularities sets of classes and extraction of optimal subsets, construction of convex combination of associated predictors that minimizes mean error;
- algorithmic constructions in a model of recognizing the nearest neighbors in binary data sets, discrete isoperimetry problem solutions, logic-combinatorial scheme in high-throughput gene expression data;
- researches in area of neural network classifiers, and applications in finance field;
- text mining, automatic classification of scientific papers, information extraction from natural language texts, semantic text analysis, natural language processing.

It is represented that book articles will be interesting as experts in the field of classifying, data mining and forecasting, and to practical users from medicine, sociology, economy, chemistry, biology, and other areas.

General Sponsor: Consortium FOI Bulgaria (www.foibg.com).

Printed in Bulgaria

Copyright © 2010 All rights reserved

© 2010 ITHEA® – Publisher; Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org ; e-mail: info@foibg.com

© 2010 Krassimir Markov, Vladimir Ryazanov, Vitalii Velychko, Levon Aslanyan – Editors

© 2010 Ina Markova – Technical editor

© 2010 For all authors in the book.

® ITHEA is a registered trade mark of FOI-COMMERCE Co.

ISBN 978-954-16-0042-9

© Jusaautor, Sofia, 2010

NATURAL INTERFACE TO ELECTION DATA

Elena Long, Vladimir Lovitskii, Michael Thrasher

Abstract: *Modern technology has the facility to empower citizens by providing easy access to vital electoral information. The majority of such users simply want to use the information; they do not wish to become embroiled in technological details that provide that access; the technology is a means to an end and if allows to obscure the real purpose (to access information) it represents a cost not a benefit. Much of the potential benefit is therefore lost unless a simple and consistent interface can be provided which shields the user from the complexity of the underlying data system retrieval and should be natural enough to be used without training. Currently there are limited tools and information available online where end users can view and interrogate electoral data. The main purpose of our paper is to report upon developments that seek to provide an easy to use interface for users to obtain information regarding the results of general elections within the United Kingdom (UK).*

Keywords: *natural interface, natural language processing, database accessing, SQL-query, production rules*

ACM Classification Keywords: *I.2 Artificial intelligence: I.2.7 Natural Language Processing: Text analysis.*

Introduction

Following the rapid development of both computer and communications technologies, our society now has the potential to access vast amounts of information almost instantaneously on a world-wide basis. One of the major obstacles to achieve it is to realising the potential for wealth and knowledge creation that information represents is the means of simple access by naïve users to relevant information locked in possibly complex data structures. Individuals are not expected to know in detail what information is required, or where it might be found and certainly does not know about data structures. In respect of electoral data, for example, the citizen simply requires information that is relevant to his or her particular area of interest - and no more.

This paper represents results of our further research in the natural language interface creation to database (DB) [V.A.Lovitskii and K.Wittamore, 1997; Guy Francis *et al.*, 2007; Elena Long *et al.*, 2009]. The data source addressed here is the DB with the results of 2005 UK General Election. The result is our current vision of “*natural interface*” has been implemented (<http://141.163.170.152:8080/NITED/NITEDJSP.jsp>) as a Web application named **NITED (Natural Interface To Election Data)** where a user can see essential election data online. The aim of design is that the application must offer simple, intuitive and responsive user interfaces that allows users to achieve their objectives regarding information retrieval with minimum effort and time..

Despite the intuitive appeal of a natural language interface, some researchers have argued that a language like English has too many ambiguities to be useful for communicating with computers. Indeed, there is little experimental data supporting the efficacy of a natural language interface, and the few studies that have compared natural language interfaces to other styles of interface have been generally negative towards the former.

Indeed, two major obstacles lie in the way of achieving the ultimate goal of support for arbitrary natural language queries. First, automatically understanding natural language (both syntactically and semantically) remains an open research problem. Second, even if there were a perfect parser that could fully understand any arbitrary natural language query, translating the parsed natural language query into a correct formal query still remains an issue since this translation requires mapping the understanding of intent into a specific database schema.

Natural language is not only very often ambiguous but is dependent on a great deal of world knowledge. In order to implement a working natural language system one must usually restrict it to cover only a limited subset of the vocabulary and syntax of a full natural language. This allows ambiguity to be reduced and processing time to be kept within reasonable bounds. In order for it still to be considered a natural language interface, most of the positive traits of a general natural language interface would have to be maintained. To retain the properties of ease of use and ease of remembering, the limitations of the system must somehow be conveyed to the user without requiring them to learn the rules explicitly.

Natural language interfaces, if they are the only form of interaction, do not take advantage of the capabilities of the computer -- those strategies that work in human-human communication are probably not best suited to human-computer interactions, where the computer can display information many times faster than people can enter commands

The principal purpose of our paper is to offer the natural (versus natural language) user interface which makes it easy, efficient, and enjoyable to operate NITED in a way which produces the desired result. This generally means that the user is required to provide minimal input to achieve the desired output, and also that NITED minimizes undesired outputs or data clutter.

Reading this paper will tell you the following:

- Natural user interface.
- Natural user enquiry.
- Help instructions.
- Production rules.
- Natural enquiry to SQL query conversion.

Natural User Interface

The natural user interface (NUI) is a key to application usability. NUI is needed when interaction between users and NITED occurs. The goal of interaction between the user and the NITED at the NUI is effective operation and control of the NITED, and feedback from the NITED in desirable for the user format i.e. NUI provides a means of input, allowing the users to ask question, and output, allowing the NITED to reply on user's question.

The design of a NUI affects the amount of effort the user must expend to provide input and to interpret the output of the system, and how much effort is required to learn this. Usability is mainly a characteristic of the NUI, but is also associated with the functionalities of the product and the process to design it. It describes how well the NITED can be used for its intended purpose by its target users with efficiency, effectiveness, and satisfaction, also taking into account the requirements from its context of use. A key property of a good user interface is consistency.

There are three important aspects [http://en.wikipedia.org/wiki/User_interface] to be taken into account. First, the controls for different features should be presented in a consistent manner so that users can find the controls easily. For example, users find it very difficult to use software when some commands are available through menus, some through icons, and some through right-clicks. A good user interface might provide shortcuts or "synonyms" that provide parallel access to a feature, but users do not have to search multiple sources to find what they're looking for.

Second, the "principle of minimum astonishment" is crucial. Various features should work in similar ways. For example, some features in Adobe Acrobat are "select tool, then select text to which apply." Others are "select text, then apply action to selection."

Third, user interfaces should strive for minimum change version-to-version -- user interfaces must remain upward compatible. For example, the change from the menu bars of Microsoft Office 2003 to the "ribbon" of Microsoft Office 2007 is universally hated by established users, many of whom found it difficult to achieve what had become routinized tasks. The "ribbon" could easily have been "better" in the mid-1990's than the menu interface if writing on a blank slate, but once hundreds of millions of users are familiar with the old interface, the costs of change and adaptation far exceed the benefit of improvement. The vast majority of users viewed this forced change, without a backward-compatibility mode, as unfavorable; more than a few viewed it as verging on malevolence. Re-design should introduce change incrementally such that existing users are not alienated by a revised product.

Good user interface design is about setting and meeting user expectations because the best NUI from a programmer's point of view is not, as a rule, the best from a user's point of view.

We have tried to create a NUI to improve the efficiency, effectiveness, and naturalness of user-NITED interaction by representing, reasoning, and acting on models of the user, domain and tasks. The main part of NUI is a graphical interface, which accepts input via computer keyboard and mouse. The actions are usually performed through direct manipulation of the graphical control elements. The natural way to represent the output for election application domain (EAD) is a table. In the next section we will discuss in detail the input enquiry presentation.

Natural User Enquiry

- Over a number of years [Guy Francis *et al.*, 2007; Elena Long *et al.*, 2009] users' natural language enquiries (NLE) have been collected by us in a series of research programmes. Direct observation of users' NLE shows, unsurprisingly, that all users are lazy i.e. they want to achieve the desired result whilst expending minimum effort. They do not want to type in the long NLE such as "*How many votes did the Demanding Honesty in Politics and Whitehall candidate obtain in Dumfriesshire, Clydesdale and Tweeddale?*" This is the natural behaviour of human being in accordance with the **principle of simplicity**, or **Occam's razor principle** (*Occam's (or Ockham's) razor is a principle attributed to the 14th century logician and Franciscan friar; William of Occam. Ockham was the village in the English county of Surrey where he was born*). The principle states that "Everything should be made as simple as possible, but not simpler". Finding a balance between simplicity and sophistication at the input side has been discussed elsewhere [L.Huang *et al.*, 2001].

On the one hand, NLE provides end users with the ability to retrieve data from a DB by asking questions using plain English. But, on the other hand, there are several problems of using NLE:

- The end users are generally unable to describe completely and unambiguously what it is they are looking for at the start of a search. They need to refine their enquiry by giving feedback on the results of initial search e.g. "*I'm looking for a nice city in France for holiday*" (where *Nice* is a city in France but also an adjective in English). Similar ambiguities exist for the UK general election database. For example, the words *Angus, Bath, Corby, ..., Wells* are values of fields *Constituency* and *Surname* in the *General Election data 2005 DB* but are also common nouns and place names. Parsing of such simple NLE is quite complicated and requires powerful knowledge base from system [V.A.Lovitskii and K.Wittamore, 1997].

- It is simply impossible to require that users know the exact values in DB (e.g. name of constituency). For example, if user makes the enquiry: “*Who won the election in Suffolk Central & Ipswich North*”? but instead of using the symbol ‘&’ types in “**and**” NITED will not find the constituency in DB.
- In the case when user simply made a mistake and instead of typing in the desirable constituency *Hereford* in the NLE: “*Who won the election in Hereford*” user entered *Hertford* (it’s **wrong** but at the same time it’s **right** from the NITED point of view because it has the right part of an existing constituency *Hertford & Stortford*), NITED will find the answer for the constituency *Hertford & Stortford*. When user sees the response, he/she realises that constituency was wrong and simply corrects it.
- As a rule a user’s NLE cannot be interpreted by NITED without additional knowledge because the concepts involved in NLE are outside of the EAD. For example, in NLE “*How did the Conservative perform in South West?*” NITED should know the meaning of word “*perform*” regarding the election data, and in the NLE “*Which party won the Aberdeenshire West and Kincardine constituency?*” correctly interprets word “*won*”.
- In conclusion it would be sensible to underline the main problem which hinders the use of NLE the cognitive process of “*understanding*” is itself not understood. First, we must ask: “*What it means to understand a NLE?*” The usual answer to that question is to model its meaning. But this answer just generates another question: “*What does meaning mean?*” The meaning of a NLE depends not only on the things it describes, explicitly and implicitly, but also on both aspects of its causality: “*What caused it to be said*” and “*What result is intended by saying it*”. In other words, the meaning of a NLE depends not only on the sentence itself, but also on the context: **Who** is asking the question, and **How** the question is phrased.

In the result of NLE analysis we decided to distinguish two different types of NUE: (1) NLE Template (**NLET**) and (2) Natural Descriptors Enquiry (**NDE**). Such enquiries permit users to communicate with a DB in a natural way rather than through the medium of formal query languages. Obviously issues in these two NUE are related, and the knowledge needed to deal with them is represented as a set of Production Rules (PR). Let us consider these two types of NUE.

Natural Language Enquiry Template combines a list of values to be selected when required and generalization of users’ NLETs. Examples of some Frequently Asked Questions (**FAQ**) are shown below:

- What was the **result** in [constituency]?
- In which **constituency** did [party] achieve its **highest vote**?
- **Who won** the [constituency]?

Figure 1. Natural Language Enquiry Template

The initial set of FAQ has been created by export in *EAD* but in the result of activities new NUE have been collected by NITED, analysed, generalized, converted to the NLET. These have then either been added to FAQ, or substituted for the under-used NLET. When the user selects an appropriate NLET with some descriptor in square brackets, selects the corresponding values from the list and click button **Go** the result will be displayed instantly (see Figure 1).

The user can build his/her own enquiries using any combination of the descriptors, each of them represents the corresponding meaningful field of the Election DB (see Figure 2). The definition of “meaningful fields” depends on AD objectives. For the considered EAD is a list of descriptors: {*region, constituency, party, etc.*}. Between descriptors and meaningful fields exist one-to-one attitude. Such attitudes are represented by the production rules (see section below).

Let’s call enquiries using descriptors as a **Natural Descriptors Enquiries (NDE)**. For example, if user wants a list of all the women elected in the South West region simply click the following check boxes: "Party", "Candidate", "Votes", "Sits in Parliament" and radio button "Female". Then select the South West region from the drop down menu of regions. When NDE is ready the user should simply click "Go" button and NITED instantly displays the result (see Figure 2). As user clicks the check boxes and selects the radio buttons NDE appears in the space next to the date above. If user clicks a check box but then change his/her mind the check box should simply be clicked again.

09-MAR-2010 - 302 Requests ENQUIRY: FIND: Party,Candidate,Votes FOR: Region='South West',Sits='Y',Gender='F'

HELP General Election: 2005: --Frequently Asked Questions-- Go Query Restore

Region: Constituency: Party: Candidate: Sits: Votes: SUM: Max Min

Region: South West Constituencies: A B C D E F G H I J K L M N O P Q R S T U V W Y Const. Cand.

47 H-Constituencies: --Unspecified-- 279 Parties: --Unspecified--

92 A-Candidates: --Unspecified-- Name: Sits in Parliament: Male Female

6 rows of found Election Data

Gender	Region	Sits	Party	First Name	Surname	Vote
Female	South West	Yes	Conservative	Angela	Browning	27,838
			Liberal Democrat	Annette	Brooke	22,000
			Labour	Dawn	Primarolo	20,778
			Labour	Valerie	Davey	16,859
			Labour	Linda	Gilroy	15,497
			Labour	Candy	Atherton	14,861

Figure 2. Natural Descriptors Enquiry

Help Instructions

Help Instructions (HI) in a Web application context means on-screen help. HI are needed for system efficiency and users' satisfaction. Clear HI can significantly reduce the number of disappointed users. Producing clear instructions that really help people is difficult as evidenced by the low-quality instructions encountered in many web applications. If designing HI were easy, there would not be so many poor examples!

Good HI have to take into account the type of users who will presumably use the NITED:

- Users' computer literacy is the basic IT literacy.
- Users should not require a conceptual background before they can use the NITED.
- Users might be absolute beginners or moderately familiar with the subject but they should not be subject matter expert.

Requirements to Help Instructions:

- HI should be short enough but provide sufficient information about the screen function.
- Good HI does not mean that all options should be explained in detail.
- HI should include brief information that is at least sufficient to get started.
- The most frequently used features should be explained.
- Top-level tasks, without much detail about particular fields, should be described.
- Step-by-step worked examples that users can follow should be represented in the .HI.

We tried to meet all of these requirements in the HI for NITED (see Figure 3).

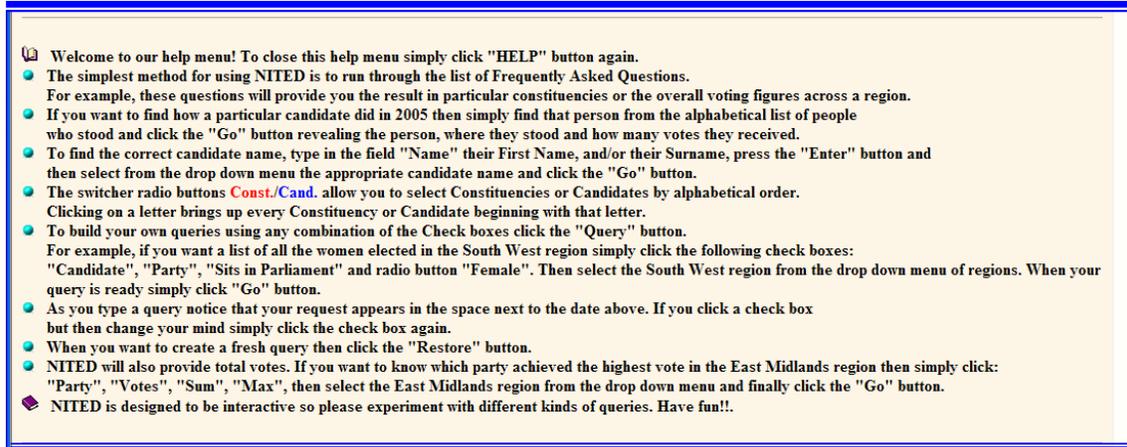


Figure 3. Help Instructions

Production Rules

At first glance, the NLET is an ideal way to communicate with EAD but in reality there are some problems, which need to be solved to provide lightness of communication. To highlight such problems is enough to consider quite a simple NLET: "Who won an election in [constituency]?" or "How did the [party] perform in [region]?". Without knowing "who is who" and meaning of "won" and "perform" NITED cannot answer such questions. To explain it to NITED the **Production Rules (PR)** need to be involved. Many researchers are investigating what information is needed and how the information needs to be represented in the PR. From our point of view the **Preconditioned PR (PPR)** should be used. The PPR is a quite powerful approach to solve this problem. The subset of PPR in format:

$$\langle \text{Precondition} \rangle \mapsto \langle \text{Antecedent} \rangle \Rightarrow \langle \text{Consequent} \rangle$$

is shown below.

- AD:Election2005 \mapsto who \Rightarrow candidate;
- AD:Election2005 \mapsto [candidate]:<win \oplus won \oplus highest > \Rightarrow [SQL]:<MAX(votes)>;
- AD:Athletics \mapsto [runner]:<win \oplus won> \Rightarrow [SQL]:<MIN(time)>;
- AD:Athletics \mapsto [shooter]:<win \oplus won> \Rightarrow [SQL]:<MAX(distance)>;
- AD:Election2005 & DB:MS Access \mapsto votes \Rightarrow [Field]:<gcr_post_election_votes>;
- AD:Election2005 & DB:MS Access \mapsto candidate \Rightarrow [Field]:<can_first_name, can_last_name>;
- AD:Election2005 & DB:Oracle \mapsto [party]:<win \oplus won \oplus highest> \Rightarrow [SQL]:<MAX(SUM(votes))>;
- AD:Election2005 & DB:MS Access \mapsto [party]:<win \oplus won \oplus highest > \Rightarrow [SQL]:<TOP1, SUM(votes),DESC>;
- AD:Election2005 & DB:MS Access \mapsto perform \Rightarrow candidate,votes;

where \oplus - denotes "exclusive OR". **Precondition** consist of **class₁:value₁ {& class₂:value₂}**. **Antecedent** might be represented by: (i) **single word** (e.g. *who, won, perform, etc.*), (ii) **sequence of words** (e.g. *as soon as,*

create KB, How are you doing, etc.), or (iii) **pair - [context]:<value>**. Context allows one to avoid word ambiguity and thereby distinguish difference between “Candidate won an election” and “Party won an election”. Presentation of **Consequent** is similar to Antecedent structure except (iii). For Consequent pair represents **[descriptor]:<value>**.

Region	Party	First Name	Surname	Vote
South West	Conservative	Christopher	Chope	28,208
		Angela	Browning	27,838
		Michael	Ancram	27,253
		James	Gray	26,282
		Robert	Key	25,961
		Adrian	Flook	25,191
		Charles	Cox	25,013
		Oliver	Letwin	24,763
		Andrew	Murrison	24,749
		Robert	Walter	23,714
		Geoffrey	Clifton-Brown	23,326

Figure 4. Reply to NLET after describing the word “perform” in the PPR

For EAD subset {1, 2, 5, 6, 8, 9} of PPR is used. PPR 3 and 4, in fact, show another meaning of the same word “won” but for a different AD. The PPR 7 shows the simplest way to cover the difference in SQL for different DB. Result of using selected PPR to reply to NLET “How did the [party] perform in [region]?” is shown on Figure 4.

Thus, NLET allows the user to “be lazy” but requires some effort to create the proper set of PPR.

Natural Enquiry to SQL Query Conversion

Two types of NUE have been considered. The NDE does not require great effort to be converted to the corresponding SQL query. Only NLET need some parsing. The mechanism of NLET parsing is very simple: “eliminating the unnecessary until only the necessary remains”. Several steps involved in NLET processing.

- NITED takes the NLET as a character sequence and converts the original NLET to a *skeleton* by noisy (non-searchable) words elimination. As a result of such conversion the NLET will contain only **meaningful** words: let’s call the word meaningful if it represents DB field descriptor or DB field value.
- EAD is represented by DB. DB **meaningful** fields (i.e. they don’t represent primary or foreign keys) contain election data. Each meaningful fields has a list of descriptors. Between descriptors and meaningful fields exists an one-to-one attitude.

- The purpose of NLET processing is to match NLET meaningful words against the DB fields descriptors.
- The final step of NLET to SQL query conversion is rather complicated because it is necessary to access data from many different tables within an EAD and join those tables together in SQL query.

Conclusion

NITED is designed through the Internet to make nationwide election results available to any user. We hope that NITED has the potential to change certain aspects of political behaviour, including people's desire to engage with the political process. Like any technology, systems like NITED can have a wide variety of effects on political behaviour and practices, but it is too soon yet to make general conclusions about its impact. Nevertheless, we intend that, following the 2010 UK General Elections in the NITED will play an important role, helping to make nationwide election results available to Web users.

Bibliography

- Guy Francis, Mark Lishman, Vladimir Lovitskii, Michael Thrasher, David Traynor, 2007. "Instantaneous Database Access", *International Journal "Information Theories & Applications"*, Vol 14(2), 161-168.
- L.Huang, T.Ulrich, M.Hemmje, E.Neuhold, 2001. "Adaptively Constructing the Query Interface for Meta Search Engines", Proc. of the Intelligent User Interface Conf.
- Elena Long, Vladimir Lovitskii, Michael Thrasher, David Traynor, 2009. "Mobile Election", International Book Series "Information Science and Computing", Book 9, Intelligent Processing, 19-28.
- V.A.Lovitskii and K.Wittamore, 1997. "DANIL: Databases Access using a Natural Interface Language", *Proc. of the International Joint Conference on Knowledge-Dialogue-Solution: KDS-97, Yalta (Ukraine)*, 282-288.

Authors information



*Elena Long – University of Plymouth, Plymouth, Devon, PL4 6DX, UK,
e-mail: elena.long@plymouth.ac.uk
Major Fields of Scientific Research: Political science*



*Vladimir Lovitskii – University of Plymouth, Plymouth, Devon, PL4 6DX, UK,
e-mail: vladimir.lovitskii@fsmail.net
Major Fields of Scientific Research: Artificial Intelligence*



*Michael Thrasher – University of Plymouth, Plymouth, Devon, PL4 6DX, UK
e-mail: mthrasher@plymouth.ac.uk
Major Fields of Scientific Research: Political science*

1.1 Election Data Visualization

Long, E., V. Lovitskii, and M. Thrasher. 2011. "Election Data Visualization." Paper presented at the ITHEA, Sofia, Bulgaria..

ELECTION DATA VISUALIZATION

Elena Long, Vladimir Lovitskii, Michael Thrasher

Abstract: *Data visualization has direct link to data interface, data capture, data analysis, and data presentation. At the present time there is still a huge gap between our ability to extract answers and our ability to present the information in meaningful ways. There is consensus that future breakthroughs will come from integrated solutions that allow end users to explore data using graphical metaphors - the goal is to unify data mining algorithms and visual human interfaces. The main purpose of our paper is to discuss one approach to that "breakthrough". The paper uses data from recent UK parliamentary elections to illustrate the approach.*

Keywords: *natural interface, data visualization, graphical interface*

ACM Classification Keywords: *1.2 Artificial intelligence: 1.2.7 Natural Language Processing: Text analysis.*

Introduction

This paper represents results of our further research in the natural language and graphical interface to database (DB) [V.A.Lovitskii and K.Wittamore, 1997; Guy Francis *et al.*, 2007; Elena Long *et al.*, 2009; Elena Long *et al.*, 2010]. The data source addressed here is a DB containing the results of the 2005 and 2010 UK General Elections. The result shows the current vision of our "natural interface" that has been implemented (<http://141.163.170.152:8080/NITED/NITEDJSP.jsp>) as a Web application. This is termed "NITED" (Natural Interface To Election Data) where a user can see and interact with UK General Election data online. The aim of design is that the application must offer **simple, intuitive and responsive user interfaces** that allow users to achieve their objectives regarding information retrieval with minimum effort and time.

S. K. Card, J. Mackinlay, and B. Shneiderman [1999] define **information visualization** (infovis) broadly: *infovis is the use of computers to interactively amplify cognition, using visual representations*. Therefore if we take this as our starting definition, systems must be computer-based, interactive, provide visual representations, and most importantly, **amplify cognition**. Our aim is that NITED meets these requirements.

Information visualization emerged and developed at the confluence of computer graphics, human-computer interaction, and databases [C. Ware, 2000; S. Card, 2003]. Continuing research in information visualization has produced a wide variety of techniques for displaying large multidimensional datasets on a computer screen and allowing users to identify complex patterns in multidimensional data. Here, we will take into account some distinctive features of election data and their potential for visualization.

People are increasingly relying on the Internet mass media as their primary source of visualizing election results. It is thus crucial for the media to develop the ability to provide accurate, informative, and user-friendly visualizations. However, this is not a straightforward task because of the complexity of the data. Because

elections visualization is such an important yet challenging task, it is useful to take into account its effectiveness. The effectiveness of the visualization not only of election results but also the visualization on Input User Interface is therefore very important, since it is the one visualization that people pay strong attention to. On the one hand, rather than have users struggle through tables of data, an effective visualization of the election results could potentially provide users with a more powerful understanding of the situation. On the other hand, a poorly-designed visualization may bring confusion rather than clarity. Worse still, such a poor visualization environment could persuade users to turn to competitor websites for their information requirements. From the user's standpoint some visualization may potentially lead them to draw incorrect conclusions about the data.

The principal purpose of our paper is to offer an Input and Output user interface which makes it easy, efficient, and enjoyable to operate NITED in a way which produces the desired result. This generally means that a user is required to provide minimal input to achieve the desired output. Essentially election data visualization is a graphical representation of data, and can be implemented in a variety of ways - from charts and graphs to more complex mapping of data that has the potential to optimise the flow of information to the end user. In this paper the process of output election data visualization will be discussed in details.

Reading this paper will tell you the following:

- User interface.
- Election interactive map.
- Drill down map.
- Drill down map creation.
- Drill down maps examples..

User Interface

The user interface (UI) for applications should be simple and intuitive for the end user - most users want to achieve the desired result whilst expending minimum effort. The end user should have a positive experience when using the application. This generally means that the user needs to provide minimal input to achieve the desired output, and crucially, that the application minimizes undesired outputs to the user. **NITED's UI has to be election domain specific** i.e. contents and features of **Election Application Domain (EAD)** should define the strictures of Input and Output user's interfaces. It should describes how well an application can be used for its intended purpose by its target users with efficiency, effectiveness, and satisfaction Some general requirements of UI are discussed in our earlier paper [Elena Long *et al.*, 2010]. The “golden rule” to UI creation is that *the UI* should be made as simple as possible, but not simpler. The proper balance between simplicity and sophistication at the input side is the main requirement to NITED UI.

UI provides a means of:

- **Input** UI, allowing the user to manipulate an application, and
- **Output** UI, allowing the application to indicate the effects of users' manipulation.

NITED uses three different kind of **Input UI**:

- **Command-Line User Interface (CLUI)** which accepts two pairs of text commands: "GE2005" – "GE2010" to activate the corresponding Data Base (DB) and "START" to activate slides show regarding the General Election 2010 and "STOP" to quit it. Only text field, keyboard and mouse are required for CLUI. This interface is hidden as a less frequently used.
- The next two Input UI are GUIs. The great convenience of the GUI made it as the standard in human-computer interaction. The GUI generally provides users with immediate, visual feedback about the effect of each action. There are several general requirements to GUI:
 - Provide meaningful contrast between screen elements.
 - Create groupings.
 - Align screen elements and groups.
 - Provide three dimensional representation
 - Use colors and graphics effectively and simply.

Let us consider two Input GUIs.

- **Graphical Query Input User Interface (GQUI)** is an interface to General Election DB. GUI is a user interface based on graphics (checkboxes, clickable images and list boxes of Regions, Constituencies, Parties and Candidates) instead of text; it uses a mouse as well as a keyboard as an input device. GQUI provides three different types of user's request:
 - **Frequently Asked Question (FAQ)**.
 - **Natural Language Enquiry Template (NLET)** combines FAQ and a slot represented by variable in the square brackets, the value of which should be selected by the user from the corresponding list of values. For example, "*In which constituency did [party] achieve its highest vote?*"
 - **Natural Descriptors Enquiry (NDE)**. Such enquiries permit users to communicate with a DB in a natural way rather than through the medium of formal query languages.

These three types of user's requests are described in greater detail in our paper [Elena Long *et al.*, 2010].

- **Graphical Charts Input User Interface (GCIUI)** provides the direct way to accomplish chart creation as a response to user's request represented only by FAQ which are classified into 3 parties (Conservative, Labour and Liberal Democrat) groups to see how they are oriented in the UK map of regional spaces, for example, party performances in the South West of England, in Scotland or in Wales. These three parties were selected initially because the election results among them and other parties are not compatible because of differential patterns of party competition (see Figure 1). GCIUI is a very useful method for the analysis of election data. For example, NITED displays the results using interactive maps. Initially the user is shown the UK map with selectable regional areas. When a region is selected, the NITED will direct the user to an analysis of its election results. Let us distinguish two parts of GCIUI: **fixed** and **dynamic**. Fixed GCIUI is based on a list box of FAQ, radio buttons group, map of Regions and clickable images of buttons. Dynamic GCIUI will be described below.

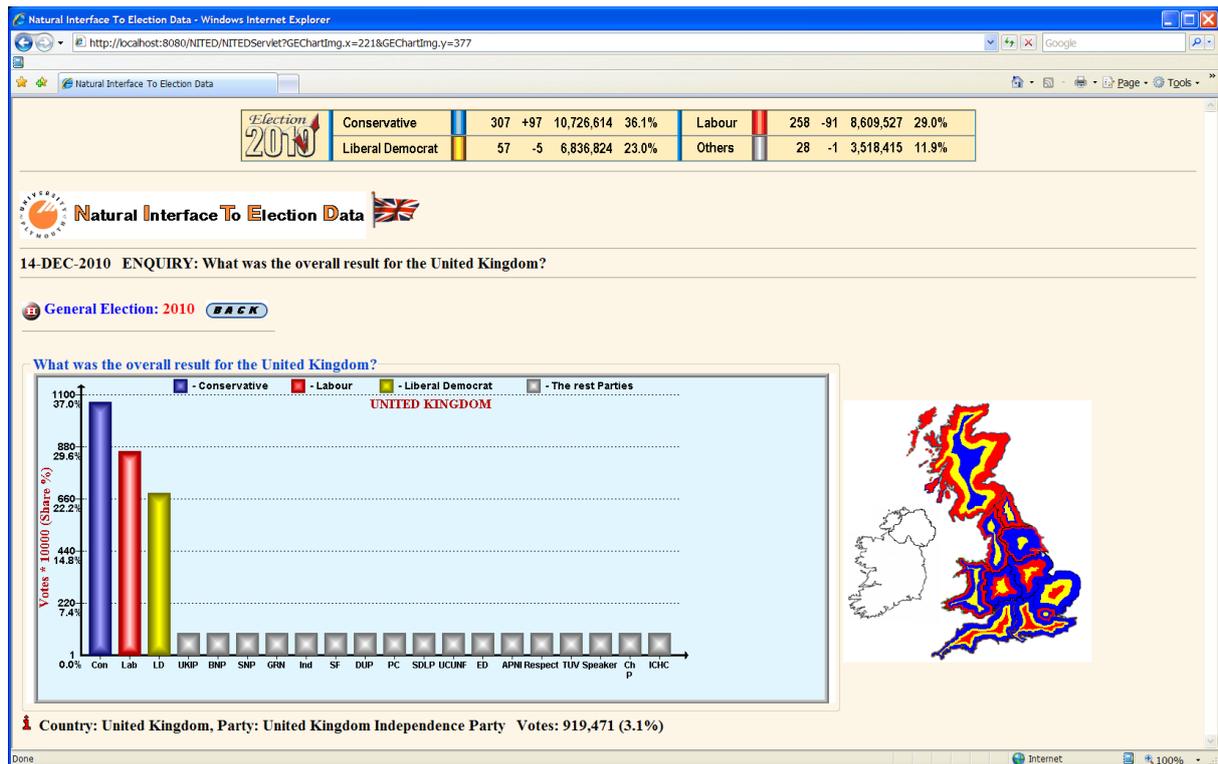


Figure 1. Chart result of the GE 2010 in the *United Kingdom*

NITED distinguishes two different types of **Output UI**:

- **Table Output (TOUI)** is the general and natural way to represent the result to user's request.
- **Chart Output (COUI)** allows the user to see the election data. COUI should satisfy some basic criteria in order to be useful for users:
 - Charts should be designed to encourage the user to make comparisons between discrete elements of data, for example, difference in vote share across constituencies/regions.
 - COUI should provide views of the data at many levels of detail. This principle relates to the "**Drill down**" and "**Level-Of-Detail**" capabilities of visualizations. With these capabilities, NITED can allow a broad overview of the data to be given and, at the same time, allow the user to have access to the detailed data that underlies the overview.
 - Maps are a potential "natural" representation for entities that can be analyzed geographically. For example, if a user would like to see how some region is performing in comparison to other regions in terms of party votes, the user could look at a map-based chart type of visualization. A map-based representation can become a true multivariate representation.

Input and Output UIs interactions is shown in Figure 2. GQIUI always first of all produces the TOUI. But NITED, as an intelligent system, analyses the result to discover whether it is logically appropriate to represent the output as a chart as well using COUI. If answer is YES TOUI displays button "*Chart*" (see Figure 3). As a result of clicking this button NITED substitutes TOUI for COUI (see Figure 1). COUI represents the **dynamic** part of GCIUI

and a user can click any histogram to see the result for the selected party. *Charts* should be preferred to *Tables* whenever it is important for users to quickly and easily recognize characteristics of and patterns in data.

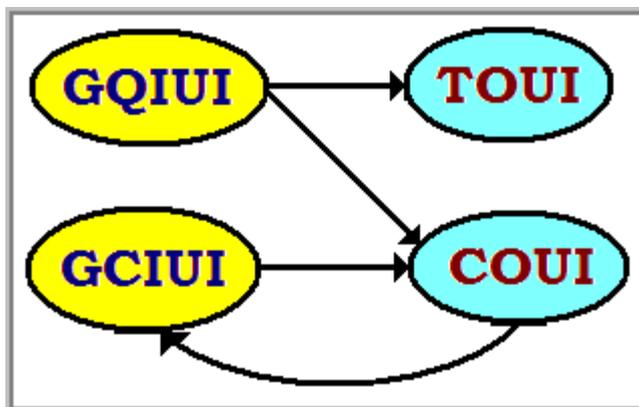


Figure 2. Input and Output UI Interaction

Mr Dave Cameron's election promise to protect old people is shown to be a disgraceful, total sham. He is attacking the elderly on every front: pensions, bus passes, retirement age, the NHS, higher VAT and now care in old age.

Natural Interface To Election Data

14-DEC-2010

General Election: 2010 --Frequently Asked Questions-- Go Query Restore **CHART**

Region: --Unspecified-- Constituencies: A B C D E F G H I J K L M N O P Q R S T U V W Y Const. Cand.

25 A-Constituencies: --Unspecified-- 133 Parties: --Unspecified--

8 A-Candidates: --Unspecified-- Name:

134 rows of found Election Data

Country	Party	Sum (Votes)
United Kingdom	Conservative Party	10,703,654
	Labour Party	8,606,517
	Liberal Democrat	6,836,248
	United Kingdom Independence Party	919,471
	British National Party	564,321
	Scottish National Party	491,386

Figure 3. Table result of the GE 2010 in the *United Kingdom*

Election Interactive Map

The expansion of Web technology over the past decade has created fresh opportunities for presenting data online. One of the most rapidly improving tools for interactive presentation is the map. Interactive maps on the Internet **present data most effectively when they invite action from the user**. Showing relationships between data is easier when the user has the power to change the visuals. Drilling down is fundamental to the interactive map. These actions enable the user to focus their browser on the amount of information that they are comfortable with.

More recent developments have included interactive data visualization which can quickly respond to new data or collect data about the user to better refine the visualization which take advantage of the power to quickly recalculate and display to find unexpected patterns, and proximity mapping that use the relationships between people, concepts, or words to determine proximity.

When a parliamentary election was over, the country's voting patterns were mapped and broken down by region. The election result was directly tied to a particular area. The interactive map enables the user to focus NITED on the amount of information that they are comfortable with. Each change in the behaviour of the NITED should be accompanied by a corresponding change in the appearance of the interface.

For user convenience, a Regions' legends are added at the top (see Figure 4) allowing the user to choose which region's data should be represented. The user can then view the results for a single constituency or the total votes and vote shares across each region.

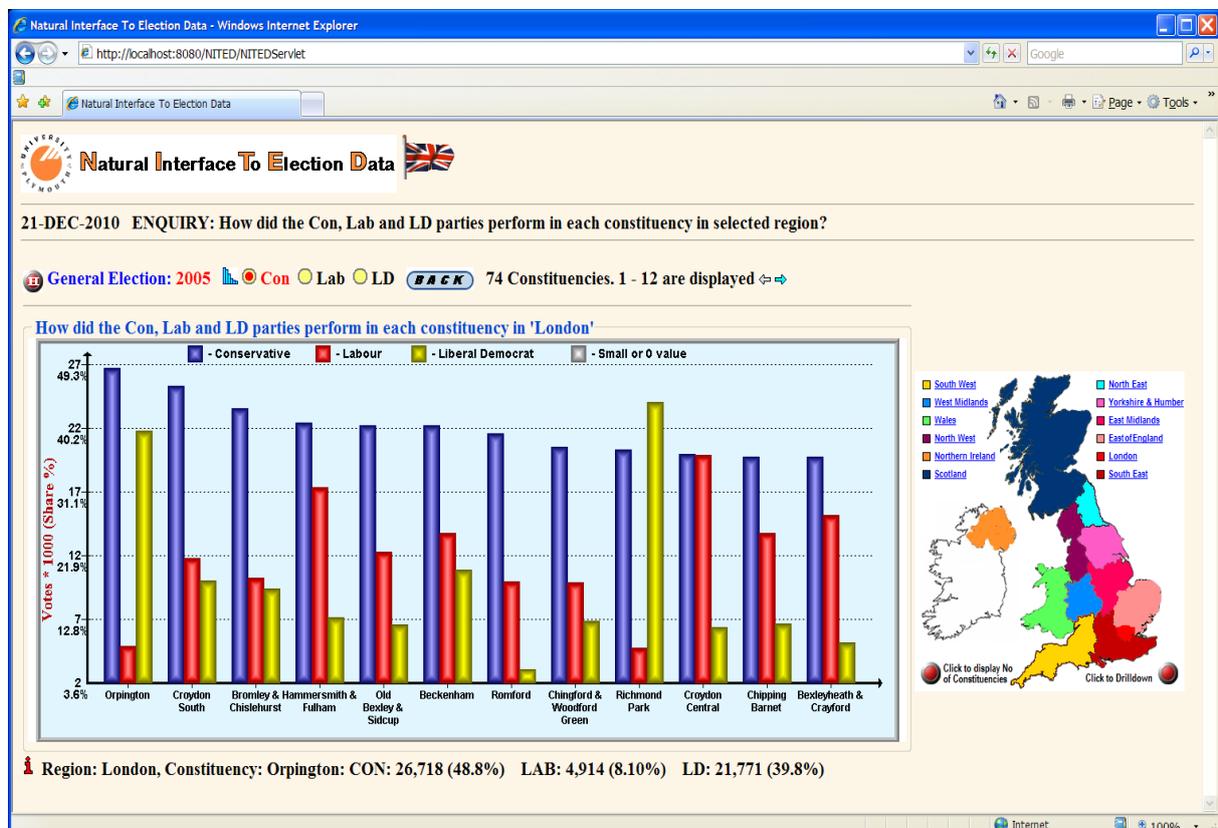


Figure 4. Constituency results of the GE 2005 in London

Drill Down Map (DD-Map)

When the user has to move through various levels of specificity of data, a drill down style of navigation is commonly offered. The drill-down clearly groups information of magnitudes ranging anywhere from the "big picture" down to an individual case.

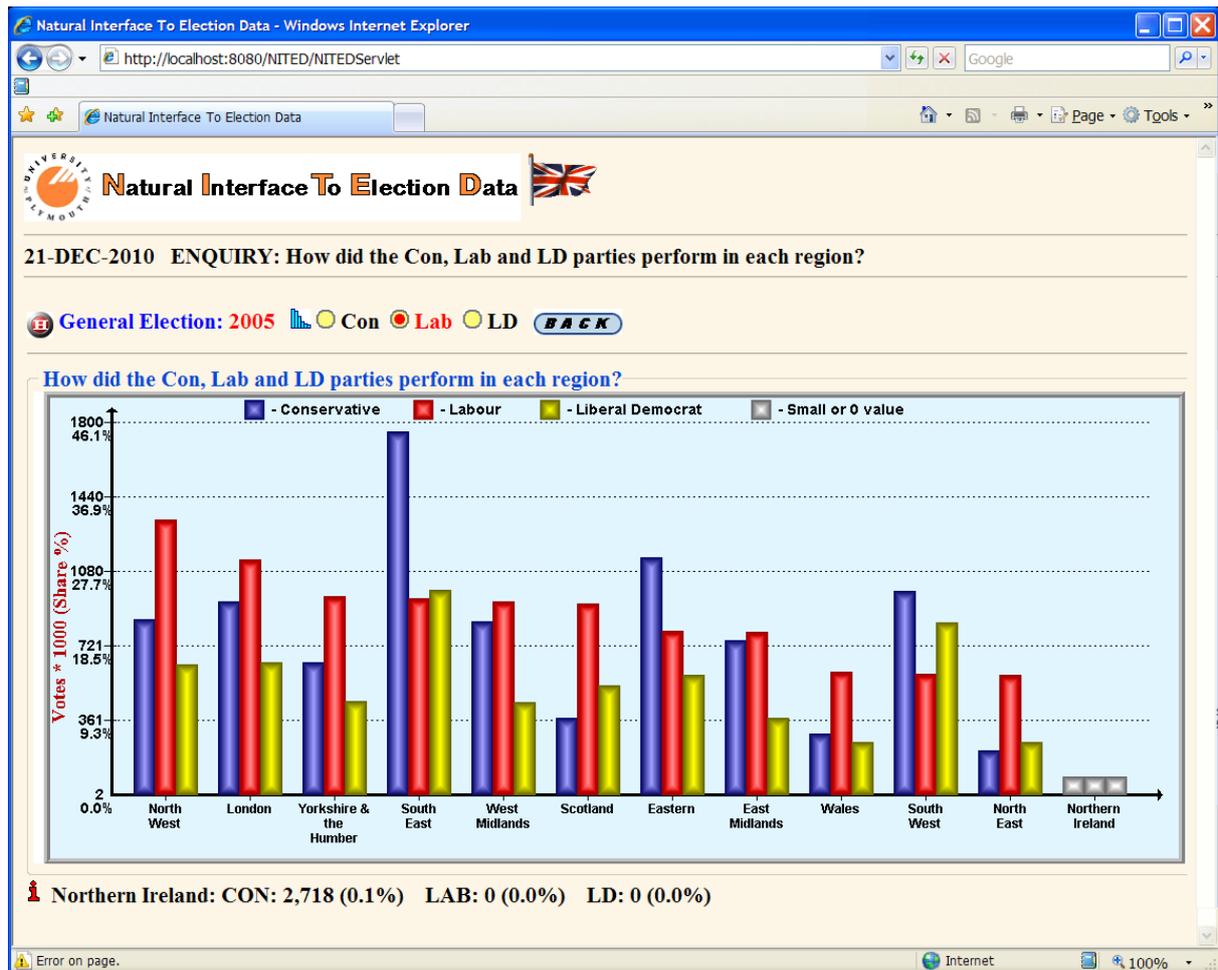


Figure 5. Regions results of the GE 2005

One Chart diagram (see Figure 5) shows that Conservatives "won" in three English regions and also in Northern Ireland (purely an artifact of party competition given that only the Conservatives of the three main parties contested seats in Northern Ireland at the 2010 general election) and that is why in the Regions' Map (see Figure 6) those regions are coloured blue (the standard colour of the Conservative party). Charts are preferred to Tables: a single chart conveys important features of the data more vividly and memorably than columns of data.

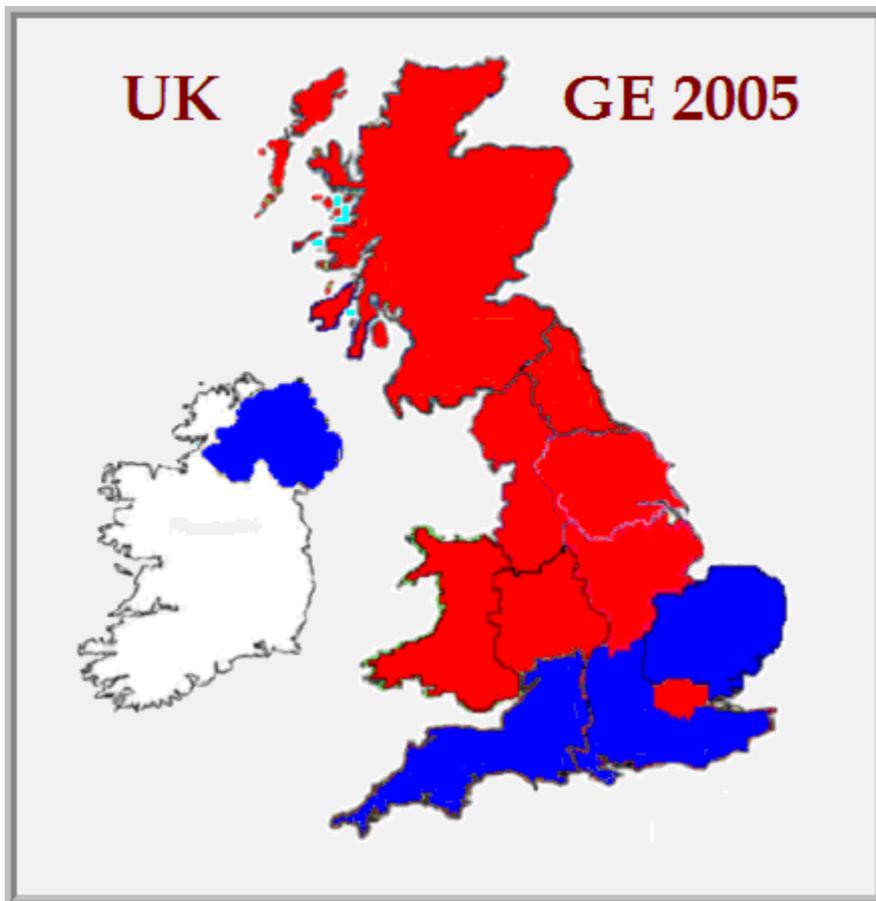


Figure 6. Results of the GE 2005 by Regions

The main purpose of using the map is to show thematically relevant information in an easily interpreted manner. **But, what is the value of such a map?** At one level of understanding the map shown in Figure 6 does not work. The regions are unequal in area giving the user a false impression of the mapped data distribution. People, not familiar with the UK voting system and the way in which votes are translated into seats, might assume that Labour won a clear majority of UK votes. However, the race was much closer. Labour won 35.2%, Conservatives - 32.3%, and Liberal Democrats won 22.0% of votes. Displaying simply the colour of a regional winner does not show the real distribution of votes between Labour, Conservative and Liberal Democrat in each region. For example,

East Midlands:	LAB: 785,943 (39.0%)	CON: 747,438 (37.1%)	LD: 372,041 (18.5%)
North West:	LAB: 1,327,668 (45.1%)	CON: 846,195 (28.7%)	LD: 629,250 (21.4%)
Scotland:	LAB: 922,397 (39.5%)	CON: 369,388 (15.8%)	LD: 528,076 (22.6%)

The attempt to produce a map depicting the results by voting percentage (i.e. each region's colour is a mix of red, blue and yellow colours in proportion to the results for that region) was not a total success [Stephen L. Sperry, 2006]. We decided instead to change colour of region in proportion of votes **change square size of region in proportion to the result** i.e. the initial size e.g. of the East Midlands region from map in Figure 6 equal 100%, or 785,943 votes. The size of the East Midlands region for Coservative party should be less and occupied **95.1%** = $747438 / 785943 * 100$ of initial Labour map, and size of Liberal Democrat map should be significantly less and occupied just **47.34%** of initial Labour map. When the sizes are changed and region maps are coloured accordingly they are placed one above another in descending order. Such map we called **Drill Down Map (DD-Map)**. The subsequent DD-Map displaying the result of the 2005 election by regions is shown in Figure 7.

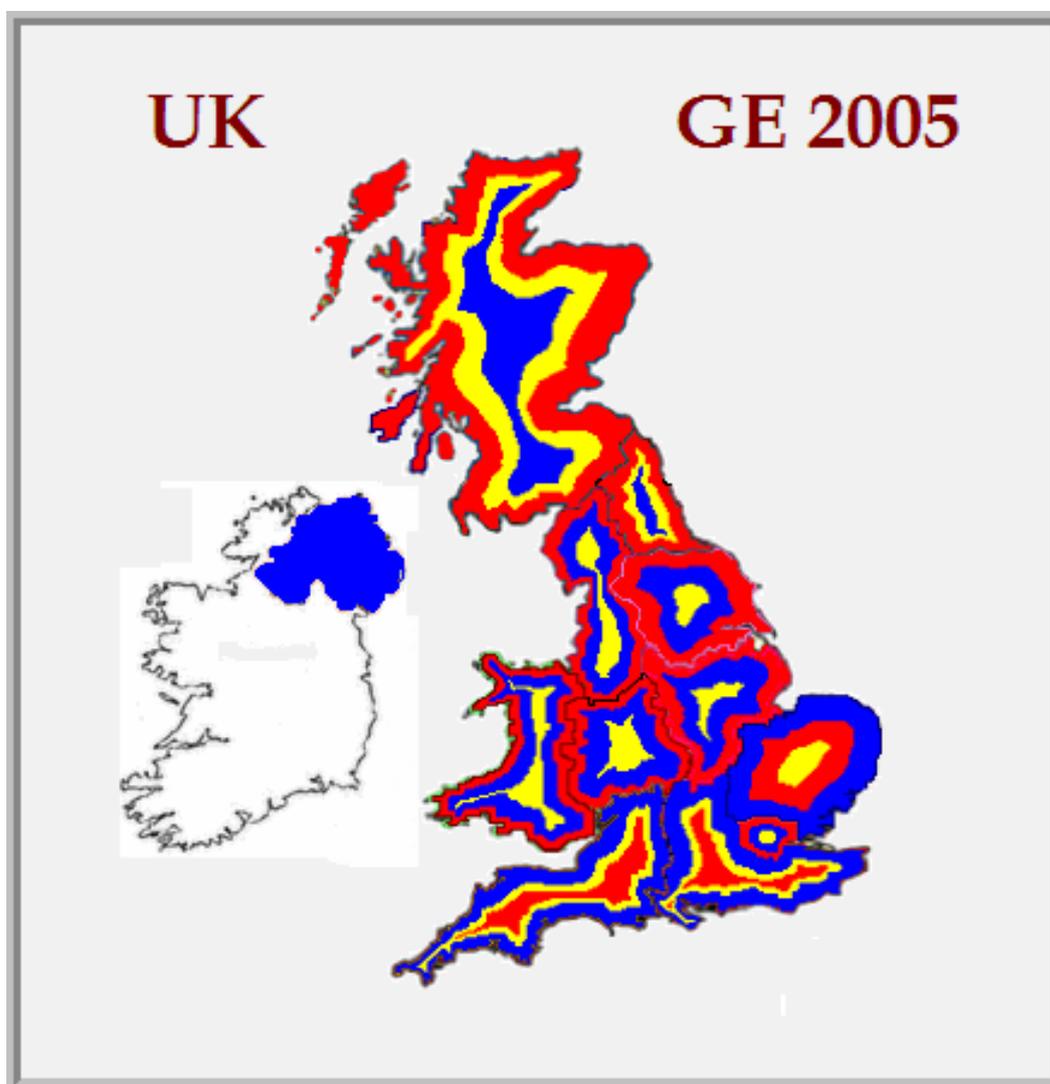


Figure 7. DD-Map of the GE 2005 by Regions

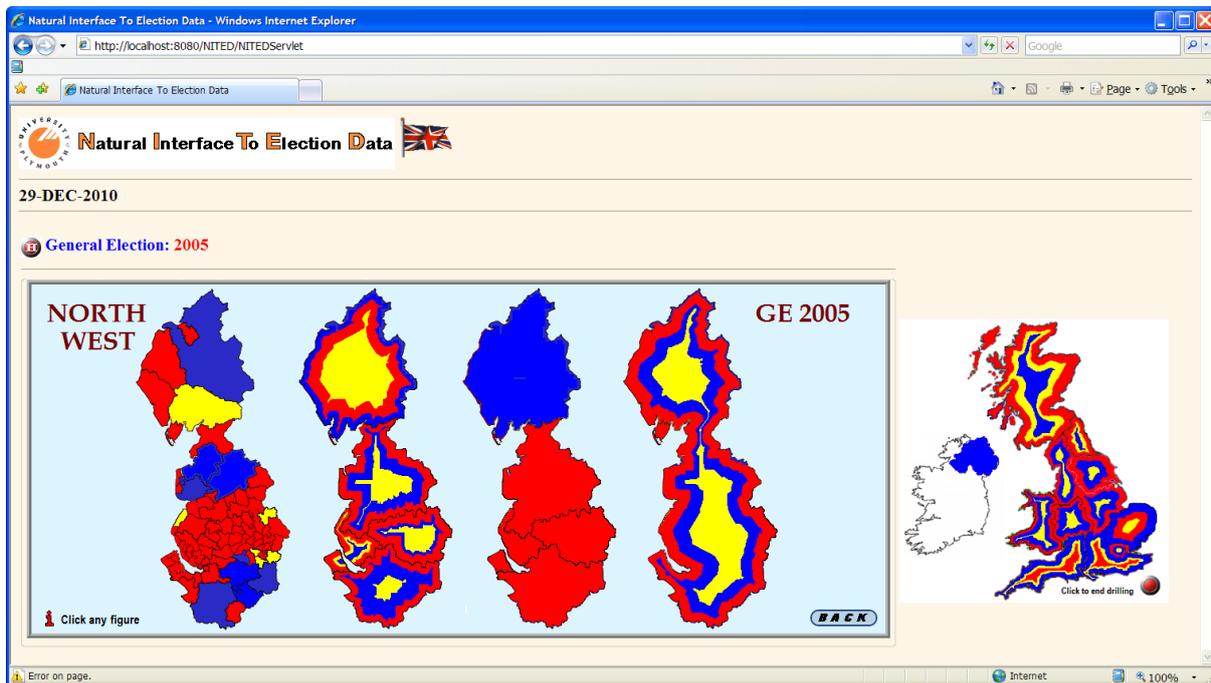


Figure 8. DD-Map of the GE 2005 for the North West region

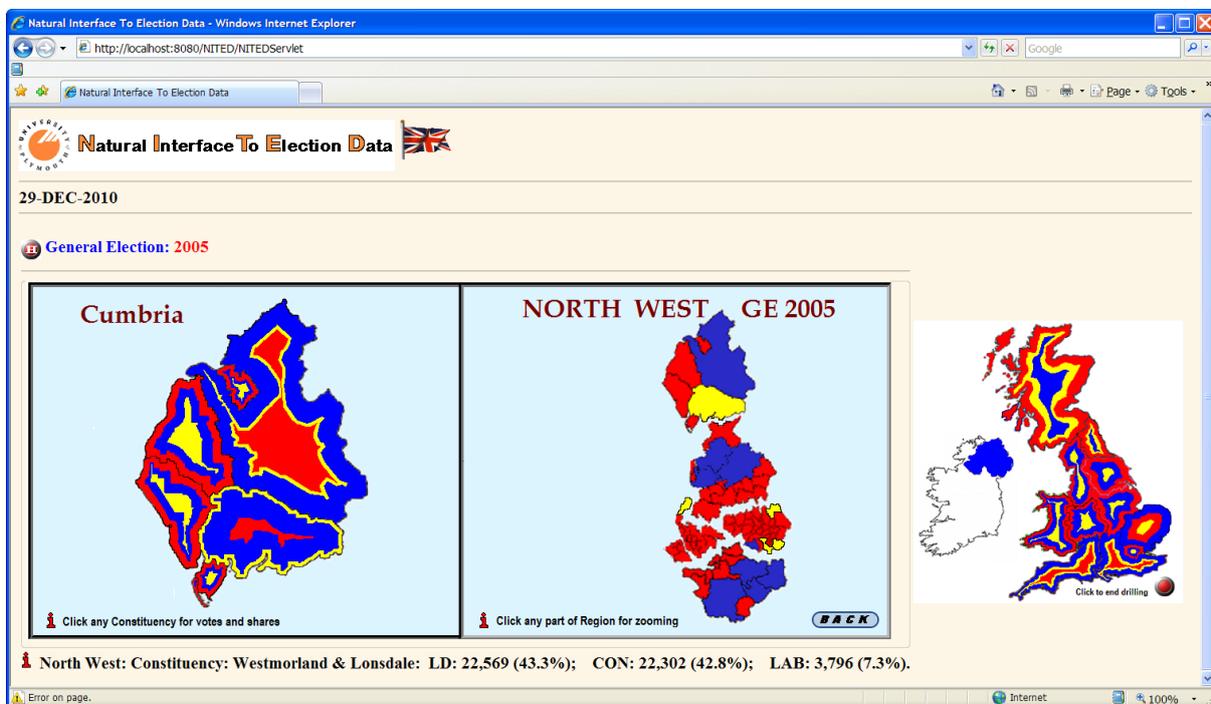


Figure 9. DD-Map of the GE 2005 for splitting of the North West region to Counties

As a result of clicking, for instance, the North West region DD-Maps of different levels will be displayed (see Figure 8). The North West region is sub-dividing into five Counties and then each County will be disaggregating to the different Constituencies. Clicking on any DD-Map of region (Figure 8) the next level of DD-Map will be displayed (see Figure 9). Finally, after clicking any specific constituency the result of election for that constituency will be shown.

Drill Down Map Creation

The well established Geographic Information computer-based System (GIS) for mapping cannot be used to produce a DD-Map. The explanation for this is very simple. **Buffering feature** usually creates two areas: one area that is **within** a specified distance to selected real world features and the other area that is **beyond**. The area that is within the specified distance is called the **buffer zone**. A buffer distance always has to be defined as a whole number (integer) or a decimal number (floating point value). This value is defined in map units (meters, feet, decimal degrees) according to the Coordinate Reference System (CRS) of the vector layer.

There are three main reasons why GIS cannot be used to create DD-Map:

1. The GIS produce the buffer zone around the given centre point (i.e. building, town, river etc.).
2. A buffer distance has to be defined precisely for buffer zone producing.
3. The direction of DD-Map creation is from Region (or Constituency) border to centre and distance cannot be calculated explicitly.

Let's describe the operation of our algorithm for creating a DD-Map using real data:

GE 2010: Region: London, Constituency: Harrow East

Results: **CON: 21,435** **LAB: 18,032** **LD: 6,850**

Step 1. Calculate the percentage of votes with regard to maximum vote:

CON: 21,435 – 100%

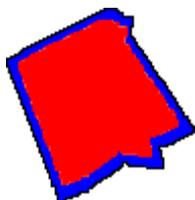
LAB: 18,032 – 84% = $18032 / 21435 * 100$



Step 2. Extract constituency from the Region London map using the Paint tool. Each Windows based computer has this tool along with Notepad, Word, Access etc. The square of image corresponds to 21,435 (or 100%). Let's call it **main image** and fill it with blue colour.



Step 3. Using the Paint (see Figure 10) stretch image to 84% and change the colour to red. Now the square of image corresponds to 18,032 (or 84%)



Step 4. Using the Paint, copy and paste the stretched image in the centre of the main image.



Step 5. Using the Paint stretch "blue" image to 32% (or "red" image to 38%) and change colour to yellow. Now the square of image corresponds to 6,850 (or 32% (38%))



Step 6. Using the Paint copy and paste the stretched image in the centre of previous main image and then place the mail image in Region's image.

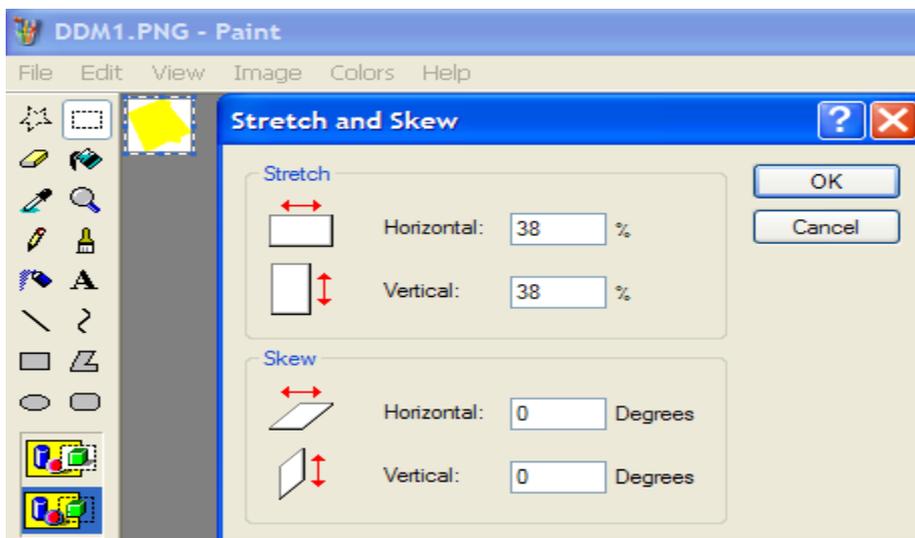


Figure 10. Using Paint to Stretch image

Drill Down Maps Examples

The next set of Figures are used to illustrate how the various mapping approaches can affect how we visualize the election results and what information the user can thereby perceive. This is especially true if you have no prior knowledge of the area being mapped; context is very important for visualization. Let us consider a drilling down of the London region for the 2005 parliamentary election.

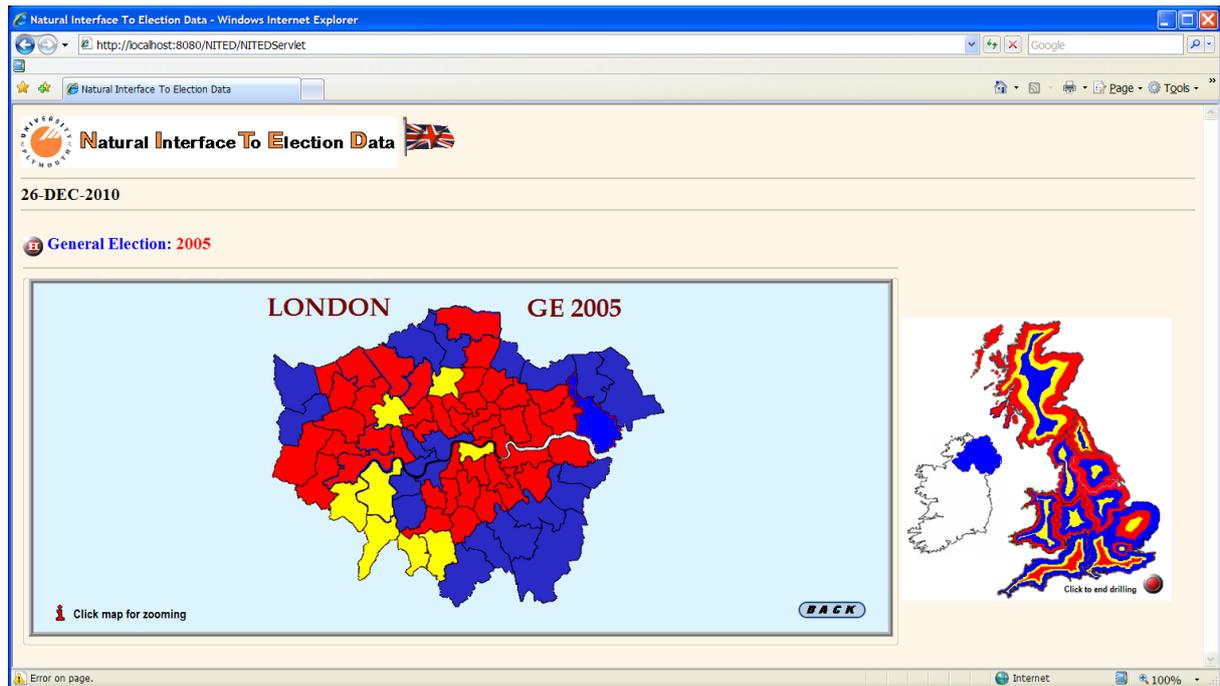


Figure 11. Constituencies Map of winner in London region

0-Level of London DD-Map (LDD-Map0) is shown in Figure 6 and indicates that Labour received the maximum votes. LDD-Map1 from Figure 7 shows that Labour did not win the absolute number of votes:

London: LAB: 1,135,687 (38.9%) CON: 931,966 (31.9%) LD: 638,533 (21.9%)

LDD-Map2 from Figure 11 shows the winner in each Constituency. For further drilling down the London region is sub-divided into four parts: North West (NW), North East (NE), South West (SW), and South East (SE). LDD-Map3 for each of this part is shown in Figures 12-15. To see LDD-Map3 for any particular part of the London region it is sufficient simply to click this part. Mouse click on any constituency of LDD-Map3 immediately display the votes in that constituency.

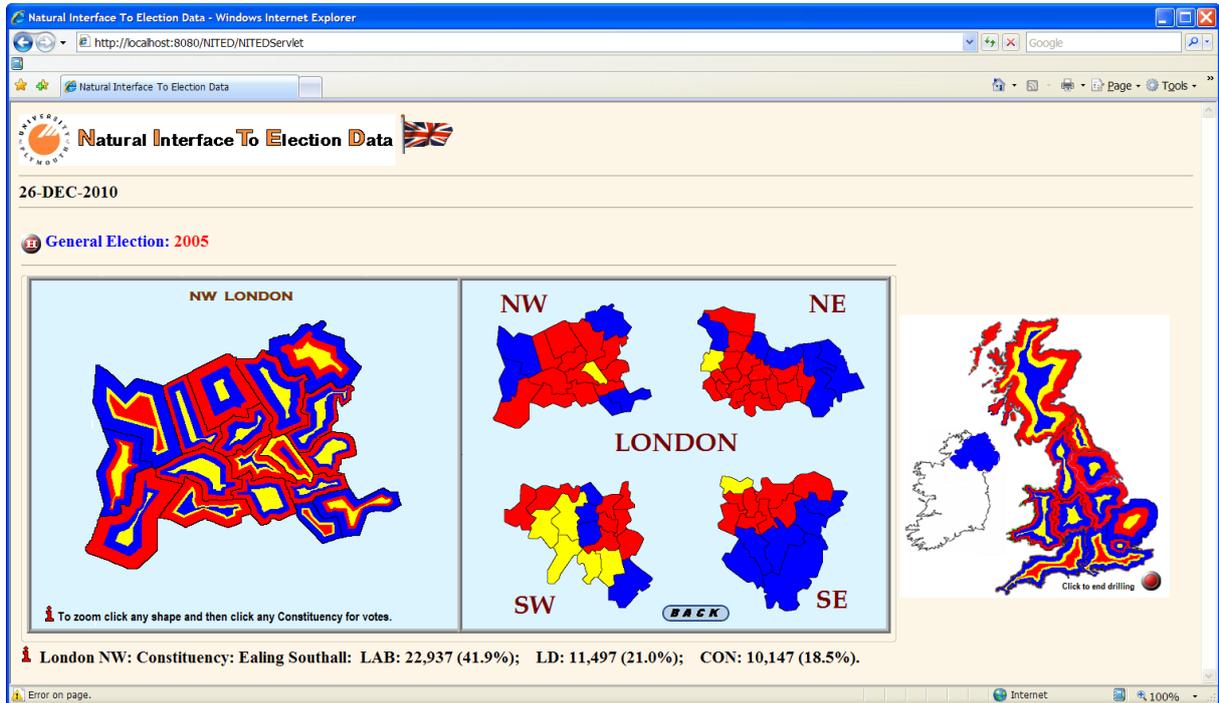


Figure 12. DD-Map of the GE 2005 for the NW part of London

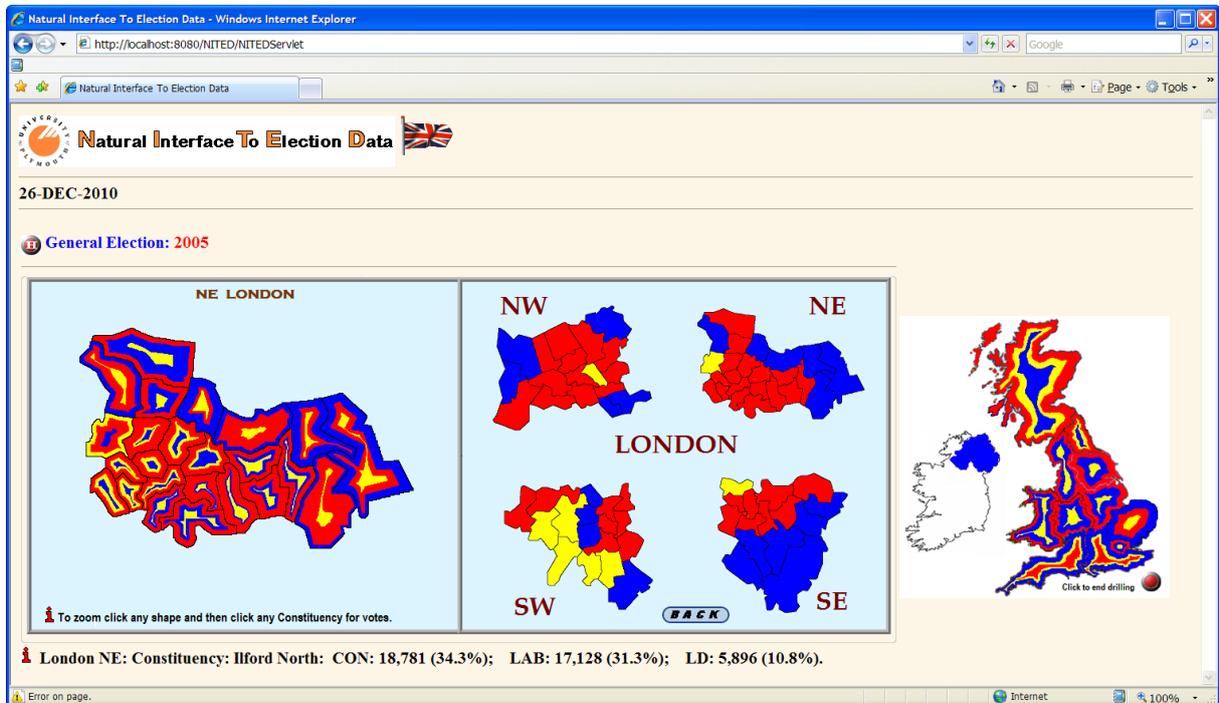


Figure 13. DD-Map of the GE 2005 for the NE part of London

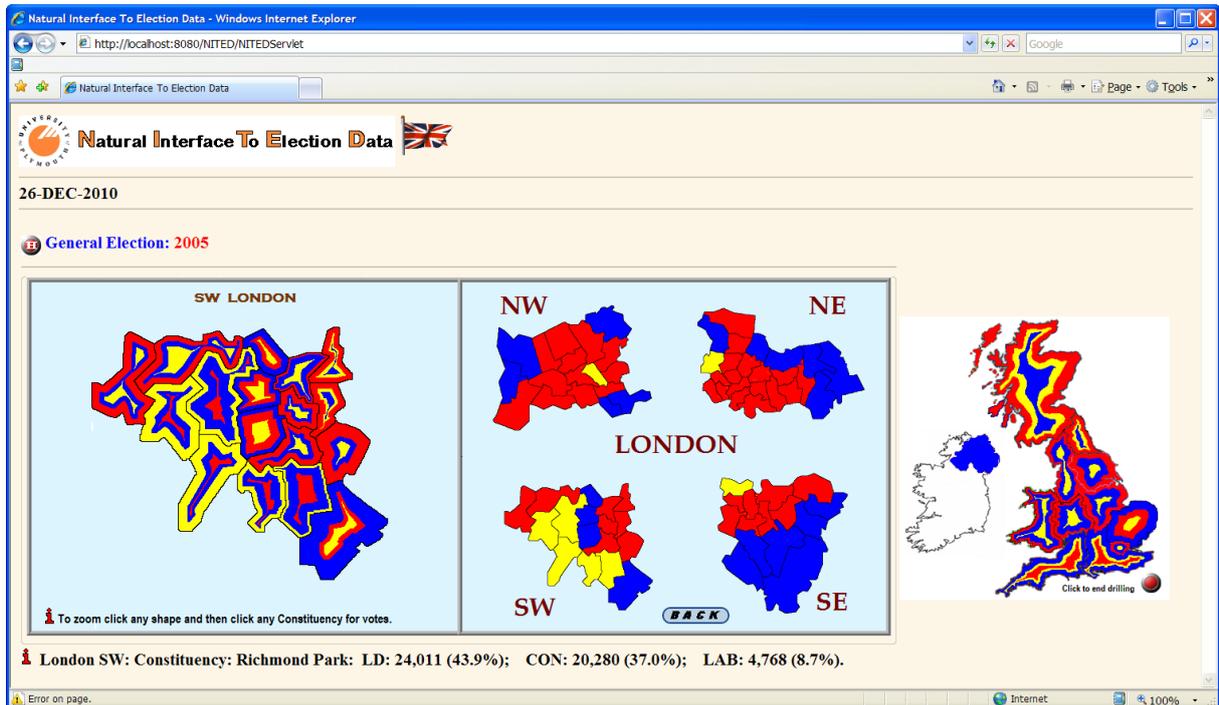


Figure 14. DD-Map of the GE 2005 for the SW part of London

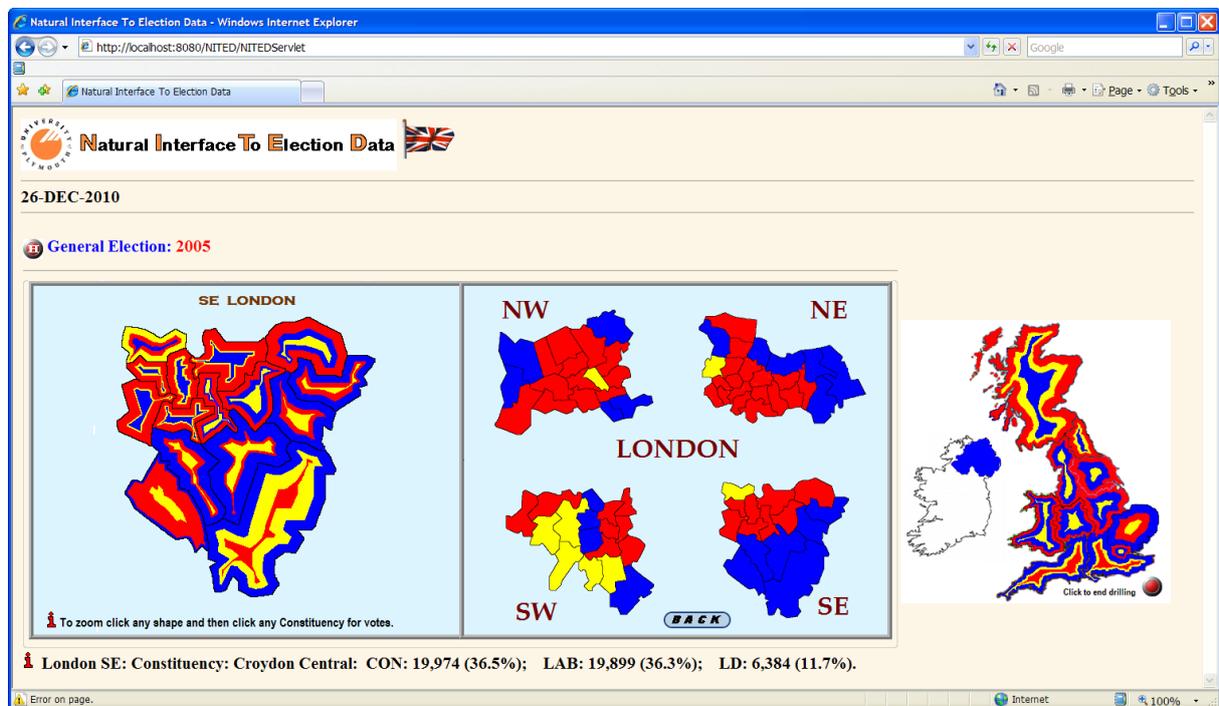


Figure 15. DD-Map of the GE 2005 for the SE part of London

Conclusion

We believe that Internet-based visualization of election results is and will continue to be very important because it facilitates user-interaction and engagement with the democratic process. NITED is designed through the Internet to make UK parliamentary election results available to any user, not simply those that possess relatively sophisticated knowledge about voting systems and their operation. Many users like to examine tables, charts, maps and DD-Maps to analyze election results. It is therefore very crucial that the Internet mass media provide visualizations that are accurate, user-friendly, and clear. We believe that NITED has the potential to encourage more people to engage with the political process.

Data, that is increasingly freely-available to end-users, lies at the heart of the modern world. Despite living in the information age, there is only so much information one person can handle at a given time. We are therefore fortunate that this is the digital information age, allowing computers to assist in translating vast quantities of data into a more usable form – away from tabular data and towards visualized data. Nevertheless, tables that come from querying databases are precisely what is used to create these visualizations. Furthermore, we have also designed some simple alternative visualization (DD-Maps) that are able to convey the same information much more clearly.

In future research we are planning to use visualization of election data for comparative analysis, thereby permitting users to visualize the dynamics of electoral change (see Fig.16).

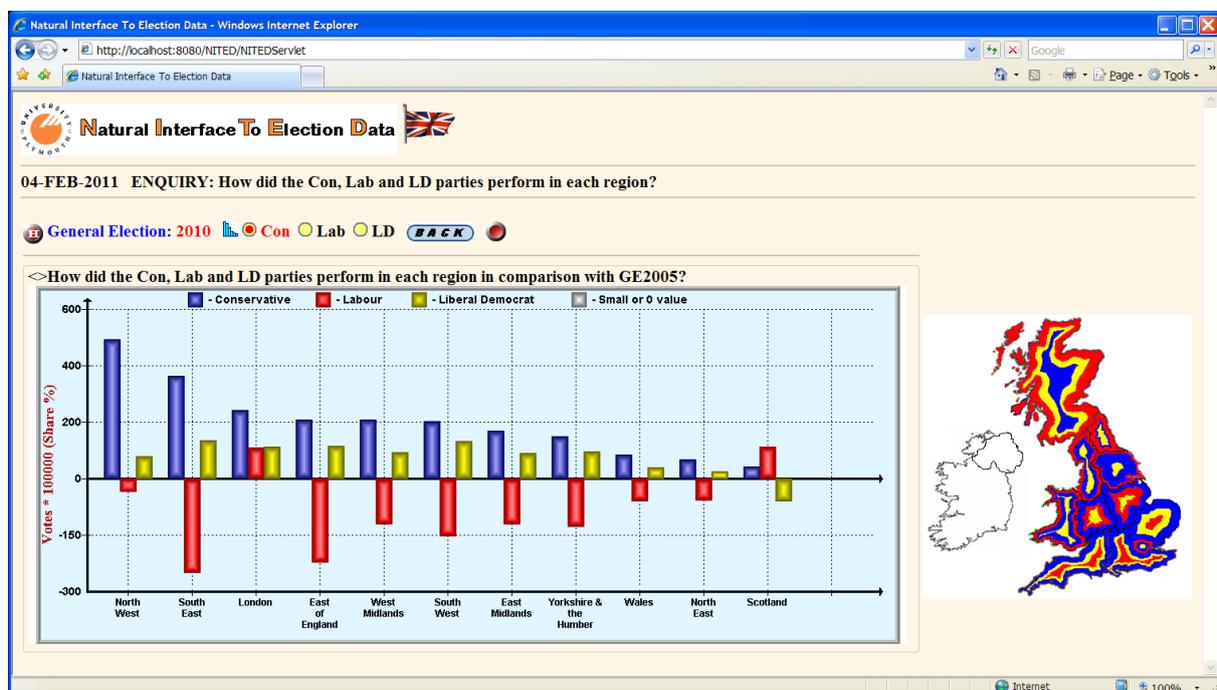


Figure 16. Comparison Analysis of GE2005/GE2010 results for Regions

Bibliography

- C. Ware, 2000. Information Visualization: Perception for Design. Morgan Kaufman, San Francisco, CA
- Guy Francis, Mark Lishman, Vladimir Lovitskii, Michael Thrasher, David Traynor, 2007. "Instantaneous Database Access", *International Journal "Information Theories & Applications"*, Vol 14(2), 161-168.
- Elena Long, Vladimir Lovitskii, Michael Thrasher, David Traynor, 2009. "Mobile Election", International Book Series "*Information Science and Computing*", Book 9, *Intelligent Processing*, 19-28.
- Elena Long, Vladimir Lovitskii, Michael Thrasher, 2010. "Natural Interface to Election Data, International Book "*New Trends in Classification and Data Mining*", ITHEA, Sofia, Bulgaria, 2010, 138-146.
- S. K. Card, J. Mackinlay, and B. Shneiderman, 1999. Readings in Information Visualization, Using Visualization to Think, chapter 1, pages 1–34. Morgan Kaufmann.
- S. Card, 2003. Information visualization. In J. A. Jacko and A. Sears, editors, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, chapter 28, pages 544–582. Lawrence Erlbaum Associates, Mahwah, NJ.
- Stephen L. Sperry, 2006, http://proceedings.esri.com/library/userconf/serug06/papers/rep_dem.pdf
- V.A.Lovitskii and K.Wittamore, 1997. "DANIL: Databases Access using a Natural Interface Language", *Proc. of the International Joint Conference on Knowledge-Dialogue-Solution: KDS-97*, Yalta (Ukraine), 282-288.

Authors information



Elena Long – University of Plymouth, Plymouth, Devon, PL4 6DX, UK,

e-mail: elena.long@plymouth.ac.uk

Major Fields of Scientific Research: Data visualization in Political science



Vladimir Lovitskii – University of Plymouth, Plymouth, Devon, PL4 6DX, UK,

e-mail: vladimir.lovitskii@fsmail.net.

Major Fields of Scientific Research: Artificial Intelligence



Michael Thrasher– University of Plymouth, Plymouth, Devon, PL4 6DX, UK,

e-mail: mthrasher@plymouth.ac.uk

Major Fields of Scientific Research: Electoral analysis

1.1 Forecasting the 2010 General Election

Rallings, Colin, Michael Thrasher, Galina Borisyuk, and Elena Long. 2011. "Forecasting the 2010 General Election Using Aggregate Local Election Data." *Electoral Studies* 30: 269-77



Forecasting the 2010 general election using aggregate local election data[☆]

Colin Rallings*, Michael Thrasher, Galina Borisyyuk, Elena Long

Department of Political Science, LGC Elections Centre, University of Plymouth, PL4 8AA, UK

ARTICLE INFO

Article history:

Received 17 September 2010

Accepted 22 September 2010

Keywords:

British election
Election forecasting
Political party

ABSTRACT

The paper presents a revised method for estimating national vote shares using aggregate data from local government by-elections. The model was originally developed to forecast the annual outcome of local elections but was adapted in time to provide an accurate forecast of Labour's landslide victory at the 1997 general election. However, over the past decade the changing pattern of party competition which has seen parties becoming more selective about which elections to contest has led to more elections being excluded from the modelling because they failed to meet the exacting criteria that all three major parties, Conservative, Labour and Liberal Democrats, had contested both the by-election and the previous main election, normally held in May. Relaxing these criteria, although increasing the number of available cases would adversely affect the forecast, over- or under-estimating party votes. Instead, the revised method overcomes the problem of differential competition by estimating vote shares for parties that contest one but not both elections. A further innovation is the calculation of a weighted moving quarterly average which takes account of the number of days elapsed between the by-election date and the date of forecast. Using the new method we provide estimates for likely party shares for the 2010 general election.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Unlike most other election forecasting models, the example described here is primarily designed not to forecast a national parliamentary election but instead to forecast national equivalent votes at annual local elections. Judged by this criterion it has proved successful. The model uses aggregate level data obtained from local council by-election results from the early 1980s onwards that take place in virtually every week of the year. It operates by calculating change in vote share across two elections, the

main election and the subsequent by-election, using cases that feature candidates from all three of the main parties, viz., Conservative, Labour and Liberal Democrats at both types of contest. The theoretical basis behind the approach is that, unlike their parliamentary equivalents that generate a media circus and become a vehicle for voters' protests, local council by-elections are relatively straightforward electoral events where people behave in a normal manner and where any idiosyncratic outcomes have a way of cancelling themselves out over the longer term.

From 1995, when we began using the model for forecasting the May local elections, it immediately demonstrated its value, so much so that we used it to forecast the 1997 general election and were pleased to see that it outperformed the national polling companies (Rallings and Thrasher, 1999). A retrospective look at the 1992 election, however, found that in common with the main polling companies the model forecast was a narrow Labour victory instead of an eight-point Conservative lead (Rallings and Thrasher, 1999).

[☆] International Symposium on *Methods and Models for Election Forecasting: The UK General Election of 2010*, organised by the Democracy, Citizens and Elections Research Network (DCERN) in association with Institute for Social Change and Politics and the Elections, Public Opinion and Parties (EPOP) Specialist Group of the UK Political Studies Association. University of Manchester, March 19–20th 2010.

* Corresponding author.

E-mail address: c.rallings@plymouth.ac.uk (C. Rallings).

Prior to the 2001 general election we encountered problems caused by an outbreak of foot and mouth disease which led some local authorities to restrict local by-election activity. This meant that from the end of February to the synchronous local/general election in June there were just 51 cases, many fewer than normal and none of which were held in the month prior to the general election itself. Nevertheless, our forecast for *The Sunday Times* published on June 2 used our analysis of split-ticket voting at the synchronous 1997 elections to generate figures from the model data (Rallings and Thrasher, 1998, 2001, 2003). The forecast read Labour 41% (42.0% actual), Conservative 32% (32.7%) and Liberal Democrats 20% (18.8%). Interestingly, three of the five companies conducting national polls over-estimated Labour by 3–5 points in their eve of poll surveys, continuing the pattern from the early 1990s.

Following the 2001 election, where recorded turnout fell below 60%, voter apathy appeared to spread to the local parties. It was not that there were many fewer by-elections than before but that the pattern of party competition started to change. Where three-party contests had once been commonplace they now became less so as one or other of the main parties failed to present candidates. Furthermore, in some areas more candidates from minor parties began to participate and secure significant electoral support, thereby making such cases unsuitable for national forecasting purposes. In short, an increasing number of cases were being excluded from the by-election modelling because the pattern of party competition at both the by-election and its May predecessor were incompatible with the task of estimating national vote shares. On May 1, 2005 our *Sunday Times* forecast was again adjusted on the basis of split-ticket voting (now 2001 as well as 1997 aggregate level data were available). The forecast was Labour to win a majority of 96 seats having polled 37% (36.1%), the Conservatives 34% (33.2%), and Liberal Democrats 21% (22.6%). The eve of poll findings from the polling companies proved as accurate.

During the recent parliament, however, the need to address the problem of declining case selection has become imperative. In the following section we outline the initial by-election model before reporting on our efforts to address the problems affecting the admission/exclusion of cases. Next, we introduce the revised model and then assess its utility by applying forecasts retrospectively both to the May local electoral cycle from the 1990s onwards. Finally, we use recent evidence to forecast the likely outcome of the 2010 general election. Although we are committed to the value of these data for forecasting since these are, “real votes in real ballot boxes” and constitute in Austin Ranney’s terms the ‘hardest’ data political scientists can get (Ranney, 1962) the model remains a work in progress. We are still testing whether the number and location of by-elections, to an extent affected by the wider electoral cycle, is a factor that influences forecasts and, if so, in what direction. We are also working with data that records the cause of the by-election vacancy since the circumstances (enforced resignation, retirement or death of incumbent for example) may affect the distribution of subsequent party support.

2. The original by-election model

We have been collecting local council by-election results from across Britain since the mid 1980s. Each year sees an average of 290 vacancies although this fluctuates with the broader electoral cycle. The forecast model requires information about both the by-election result and the outcome at the preceding May-election for each ward (the local electoral district). Clearly, assuming that the pattern of party competition is identical across elections it is straightforward to calculate change in vote share and swing for a given ward but forecasts, generalising from the particular, require a set of benchmark figures that are common across a range of wards. This comes in the guise of the ‘national equivalent vote’ (NEV) an estimate of how the country as a whole might have voted extrapolated from actual local election voting in any given year (Curtice and Payne, 1991; Rallings and Thrasher, 1993).

For any given ward election in May, therefore, we know both the distribution of party support in the ward and how that compares with the country as a whole. The original method used only by-elections which featured candidates from Conservative, Labour and Liberal Democrat parties at both the May election and the by-election. An additional caveat was that cases would be excluded from consideration where votes for other parties and Independents at either the May or by-election were greater than 10% of the total vote. The exception to this was cases where the intervention and support for other parties/independents was consistent across the two elections. Because of a non-uniform local electoral cycle it was important to note the particular year when a by-election ward had held its May election since this became an important part of the calculation. A worked example is provided in Table 1 while the method is described formally in the Appendix.

3. The revised by-election model

From 2001 the rather dramatic change to the pattern of party competition had implications for modelling. In 2000 some 68% of by-elections featured candidates from all three main parties; this dropped to 62% in 2001 with a further fall to 55% in 2002. Although the proportion of three-party contests recovered from this low point it did not match the

Table 1

Calculating the current national equivalent vote from the November, 1996 by-election result in Ixworth ward, St Edmundsbury Council.

Steps in the method	Conservative	Labour	Lib Dem
a) By-election share in ward	43.4	26.2	30.4
b) 1995 May vote in the ward	36.1	29.6	34.3
c) Change in vote share (a–b)	+7.3	–3.4	–3.9
d) 1995 National Equivalent Vote	25	47	23
e) Estimate of current NEV (d + c)	32.3	43.6	19.1

Of course, in a given ward it is possible that the change in a party’s vote share may be greater than its NEV for a particular year, leading to a current NEV estimate that is nonsensical. However, by averaging the estimates across all by-elections over a month/three month period, any extreme results are smoothed out. In essence, therefore, between the May and the subsequent by-election the model is calculating change in each party’s vote share, adding/subtracting that change to the NEV value for the relevant year and averaging across cases to estimate a current NEV for a given point in time.

consistent levels seen throughout the 1990s. This, combined with a retreat from three-party contests in the main May elections, especially in the English shires, led to model estimates being sourced from a declining base of data.

The initial response was to investigate thoroughly the changing structure of party competition and second to devise methods that might then compensate for missing values, permitting more cases to be used to estimate national support. A third aim was to determine the optimal time frame for averaging – it should be a trade-off between averages being responsive to new information and yet not over-sensitive to random variations.

A starting point was to examine for every case the structure of party competition at both the by-election and the previous May election. In more than 7000 by-elections there was three-party competition at both the May and by-election in 3425 cases. In a further 544 by-elections a Liberal Democrat candidate, present for the May contest was missing from the subsequent by-election. In another 250 and 77 cases it was the Labour and Conservative candidate respectively that missed the by-election. Of course, the process worked in the opposite direction with by-election vacancies attracting greater party competition than had the May equivalent. For example, in 648 cases where three main parties contested a by-election the Liberal Democrats had not challenged when the main May election was fought. In a further 255 and 155 cases it was Labour and the Conservative candidates respectively that are missing from the May election but are present at the by-election contest. In other examples the structure of party competition was partial but stable in the sense that perhaps only two of the three parties competed at both elections. In 457 cases, for example, only Conservative and Labour challenge one another with the Liberal Democrats absent on both occasions. In a further 331 examples the two protagonists are Conservative and Liberal Democrats while 104 cases are Labour versus Liberal Democrats only. The structure of party competition, therefore, dictated differing responses in devising new procedures designed to include more cases for devising model estimates.

The first examples consider cases where the pattern of party competition is more extensive at the main elections in May than it is for the subsequent by-election. Local voters in May could select from Conservative, Labour and Liberal Democrat candidates but the Liberal Democrats (in the example shown) decide to stand aside from the by-election contest ([Example 1, Appendix](#)). Previously, the model ignored such cases but a different method is now employed, providing a notional by-election share for the Liberal Democrats which is equal to the minimum share value the party achieved across the relevant local authority at the May election. The theoretical basis for this assumption is that the likely explanation for the Liberal Democrats to withdraw from the by-election is the expectation of receiving a low level of votes. The votes for the parties that did contest the by-election are then adjusted to take account of the estimated vote for the absent party. It is important that the process of normalising votes in this way does not distort the actual result, for example, transforming a positive change in vote share for one of the parties that did contest into a negative one. In order to prevent this

from happening the share change for a party that both contests and increases its share is never allowed to fall below zero. Following these adjustments the by-election result is then treated in the usual way for the purposes of estimating national vote shares.

There are other occasions when the extent of party competition is further fragmented; three main parties contest the May election but now **two** fail to contest the by-election. For example, both Labour and Liberal Democrats do not present by-election candidates to challenge the Conservatives ([Example 2 in Appendix](#)). The problem here is that the change in Conservative vote share from May is likely to be inflated because of the absence of two of its competitors. In such cases we proceed with the method described above, now estimating by-election shares for both Labour and Liberal Democrats based on the minimum values across the local authority at the previous May election. Following this procedure the Conservative by-election share is recalculated by subtracting from its actual share the estimated shares for both Labour and Liberal Democrats. Again, if the actual Conservative by-election share is an increase from its share in May but the process of estimating shares for the two missing parties transforms that to a decrease then change is limited to zero. Where the Conservative by-election share does actually decrease, despite the two missing parties but presumably because of the support for independents or other smaller parties, then no estimates are made for change in share for Labour and Liberal Democrats while the Conservatives are given the actual May to by-election change.

Another scenario is that now just two parties contest the May election but only one of these parties has a by-election candidate. An example is that Conservative and Labour challenge one another in May but that Labour does not contest the by-election ([Example 3, Appendix](#)). Since there is no Liberal Democrat standing at either election its change is regarded as missing data in the modelling. An estimate for Labour's by-election vote is made in the usual manner by assuming that its by-election vote would be equivalent to its share in its worst performing ward across the local authority at the previous May election. The adjusted Conservative by-election vote share then becomes its actual share minus the estimated Labour share with the caveat that the direction of change cannot be counter-intuitive: it is set to zero if an actual positive change becomes a negative one after adjustments and estimated vote shares are made.

The examples discussed so far have focused on procedures when the pattern of party competition at the by-election is less than at the previous May election. There are cases where the opposite situation applies and more parties challenge for the by-election vacancy. The first situation involves two main parties with candidates in May but these are subsequently joined by the missing party when the by-election is fought. We assume that Conservative and Labour are rivals in May but the Liberal Democrats put in a by-election appearance ([Example 4, Appendix](#)). In such a case an estimated May vote for the Liberal Democrats is based on the party's worst performing ward across the local authority. A restriction is imposed and this assumes that its by-election vote share is a positive

change in vote from the estimated May election; if not then change in share is limited to zero. For the two parties that contested both elections change is set to zero if after estimating a May vote for the Liberal Democrats a negative change in vote becomes a positive one.

A minor adjustment to the example given above is that two parties contest the by-election but only one of these two was present at the previous May election. In May the Conservatives compete against independents and/or minor parties only but are then subsequently challenged by a Labour by-election candidate (Example 5, Appendix). Since there is no Liberal Democrat at either election its change is treated as missing. Labour's estimated May vote share is that in its worst performing ward across the local authority. The Conservative share in May is recalculated to be its actual vote minus the estimated Labour share. The restriction is that Labour's by-election share cannot be less than its estimated May share and is set to zero if that would happen. Similarly, Conservative change is set to zero if the consequences of these adjustments are that the direction of change is altered.

A more extensive adjustment is required when only one of the main parties contests the seat in May but the by-election sees all three parties contest. One such situation would be where the Conservatives stood a candidate on each occasion its two rivals did not fight the May election (Example 6, Appendix). Estimates are calculated in the normal manner for each of Labour and Liberal Democrats that are equal to vote share in their weakest wards across the authority.

Finally, a more complex pattern of party competition occurs when two participate in the May election but a different pair of parties contest the by-election. For example, Conservative and Labour compete for the seat in May but then the by-election has no Labour candidates but local Liberal Democrats decide to contest (Example 7, Appendix). Step one estimates a May vote share for the Liberal Democrats calculated in the usual manner. Step two sees Labour's May vote reduced by subtracting its worst performing ward result while step three recalculates the share for the Conservatives. The normal restrictions are then applied.

The result of making such compensations is that a greater proportion of by-election cases may be included in modelling national equivalent vote estimates. The extent of that increase is shown in Fig. 1. The two curves show the proportion of by-elections used in estimating national vote shares. The broken line is the proportion that were usable under the strict criteria of three-party competition in both the May and by-elections and a large fraction of total votes cast for the main parties. The solid line is the proportion

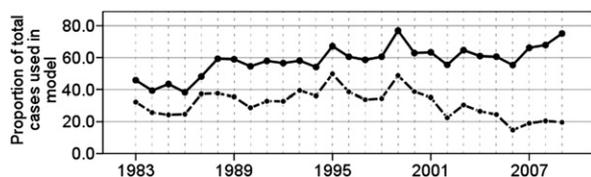


Fig. 1. Consequence of new data selection criteria: proportion of by-elections used by the original and the revised model.

after compensating for incomplete three-party competition and allowing for missing values when one or other of the three main parties fails to compete at either election. The most dramatic difference occurs in the period after 2001 when, as we reported earlier, there was a significant move away from the pattern of party competition that had evolved over an almost twenty-year period before then.

Of course, we need to be assured that the effects of estimating missing vote shares do not introduce inaccuracies to a model that enjoyed a proven track record of forecasting national support for the annual round of local elections. Plotting each party's forecast vote share over a long period reveals that the effect of estimating votes for missing parties is to reduce the amount of volatility in the monthly estimates due to the increased number of cases used. In the period from 2005, for example, the Conservative three month average varies between 18 and 44% (range 26 points) using the original method while the range is just 10 percentage points (19–32%) for the revised method. Similarly, Labour varies between 19 and 32% and between 22 and 30% for the old and new methods respectively while the Liberal Democrats were ranging between 19 and 49% and now lie between 22 and 34%. Another measure of volatility is the level of change of support from one month to the next. Fluctuations that are very pronounced are more likely to be of a random nature rather than reflecting real changes in the public mood. Compared with the original method the revised model estimates demonstrate greater stability in short-term support¹.

The final element in the revision process was to consider how best to create estimates for a given point in time. Providing estimates of national support based on by-elections from a single week would be ill advised since these can vary considerably, influenced by a range of local and national factors and also being affected by relatively small numbers of cases. A more robust approach is to use a broader time period. This has the effect of smoothing large fluctuations that may occur using weekly data. Closer examination of results and trends over a twenty-five year period suggested that a more reliable procedure is a weighted quarterly moving average². Thus, each forecast is based not simply on the figures for a single month but also some information from the preceding two months. The weighting procedure takes into consideration the time elapsed from when each by-election occurs and the date of forecast, usually the last day of the month of interest. Thus, a by-election that happens on the last day of June has a bigger impact on June's averages compared to a by-election that held on April 1. There is a linear decrease in

¹ In terms of Conservative support during the 2005 parliament the original method suggested a maximum increase of 12 percentage points and a maximal decline of minus 24 points (standard deviation = 4.5) while using the revised method the Conservative figures do not change by more than 5.6 points (std = 1.9) across a two month period. The equivalent figures for Labour are +5 to –6 (std = 2.0) now changing between +4 and –3.5 (std = 1.4) and Liberal Democrats +19 to –15 (std = 4.7) using the old method to change between +4 and –10 (std = 2.2) on the revised model.

² Our experience is that employing a three month average provides a better smoothing of the data than a two month average.

weights that reflects the days elapsed from the forecast date. In turn, the above June by-election will also have an impact on July and August 'averages' but its impact on the model estimate is reduced with time. In short, if the number of by-elections stays more or less the same across all months then the influence of this June by-election decreases linearly (it is highest for June, smaller for July, and smallest for August) before it is completely removed from the quarterly calculation.

4. Estimating national equivalent vote share using the revised model

The first test of the revised model lay in its ability to forecast correctly the May elections using the April model estimate. The first iteration of the model was a reliable method for forecasting the national equivalent vote for local elections and it would be a retrograde step if the revised method performed less well. Fig. 2a–c shows for each election year since 1993 both the April by-election model vote share (comprising data from the April, March and February results) and the eventual May NEV for the three main parties. The solid line represents the by-election model share and the dotted line is NEV. Overall, the methods for estimating missing data do not appear to impact negatively on the model forecasts.

For the Conservatives the two curves are close together but there is no consistent pattern in terms of the forecast accuracy. The largest gap over the period is in 2005 when the model over-estimated the Conservative NEV (note: this

is the party's local election performance rather than its general election vote share) and both 2007 and 2008 when it did somewhat better at the actual elections than had been forecast from the by-election data. A close examination of the pattern of support for Labour again shows that the two curves are close together with two clear exceptions, both of which are general election years (2001 and 2005). The 1997 general election does not repeat this gap but it should be noted that this period (1995–1997) represented the peak of Labour's local electoral performance and it is unlikely that the party could have improved further on its by-election results. The 2001, 2005 data are suggesting that Labour appears to raise its game (or its supporters take notice) where a general election contest arises. It is very rare for Labour to do better in by-elections than in the main May local elections; the one exception over this period is in 1997 but the gap is rather small. For the Liberal Democrats (Fig. 2c) the pattern is the reverse of Labour's performance. The tendency is for the by-election model to give a higher estimate than Liberal Democrats' actual NEV. This feature is particularly noticeable in both 2003 and four years later in 2007, the peak of the local electoral cycle in terms of council seats up for election. It does seem that the Liberal Democrats perform better in the by-election situation than at the national level when much of the country is voting in the main May elections. Of course, compared with the party's general election performance this gap (between by-election model forecast and general election vote) is greater still for reasons addressed earlier.

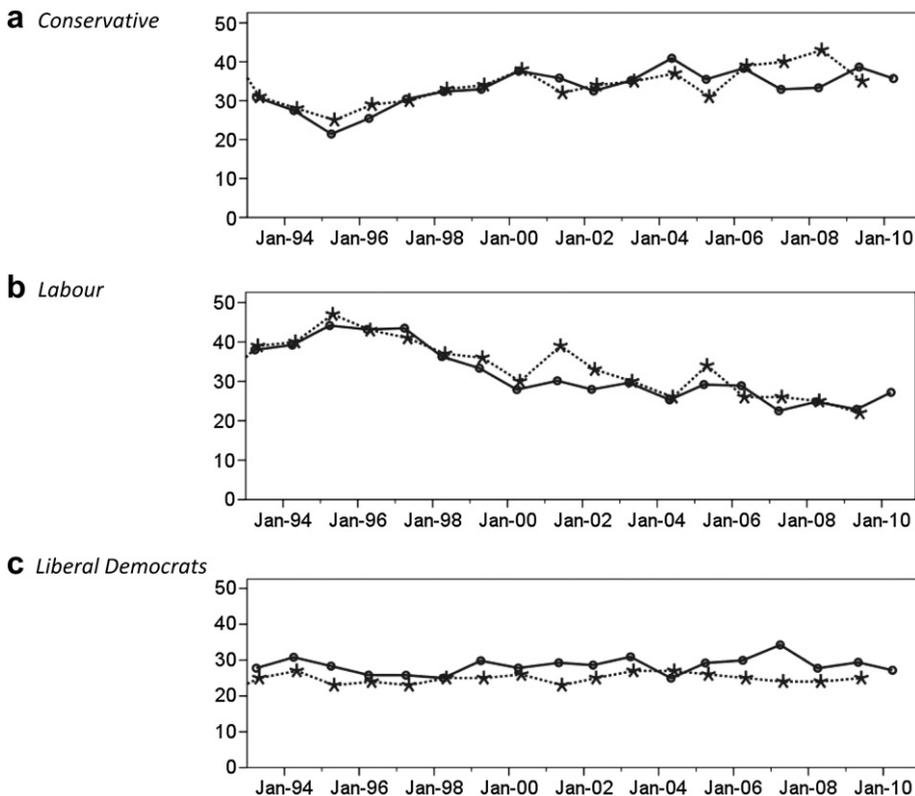


Fig. 2. The revised by-election model and NEV, 1993–2010. (a)Conservative, (b)Labour, (c)Liberal Democrats.

5. The trends in the by-election model and opinion polls since 2005

In this section we consider how the model has estimated support since the 2005 general election and compare that with the opinion polls using weighted moving quarterly averages. The intention, using Fig. 3a–d, is to show more clearly the relationship between model estimates and the polls. Fig. 3a tracks support for the Conservatives and shows that although both curves run close together, seldom being more than a few percentage points apart, there is no consistent pattern in the sense that one estimate is always above/below the other. In the immediate aftermath of the May 2005 election Conservative support rises gradually but then, following Brown’s hesitation about a snap election in autumn 2007 it accelerates towards a peak in midsummer 2008 according to the

by-election model but two months later in terms of poll rating. Since then the trend has been gradually downwards. In the run-up to the 2005 general election Labour’s estimated national vote using the by-election model was around 30% while the polls were somewhat higher, averaging in the high 30s (Fig. 3b). Following the general election the party receives a small boost in the polls but its performance in by-elections notably declines towards the mid 20s. Thereafter the two curves track one another whilst remaining between 6–12 percentage points apart – however people are responding to opinion surveys it appears that Labour is unable to translate such support into actual votes. Blair’s announcement to leave office appears to be the catalyst that turns the party’s fortunes around on both measures, noticeably so in terms of the poll ratings. Brown’s hesitation over the election date pops the ratings bubble until the economic crisis and the Prime Minister’s

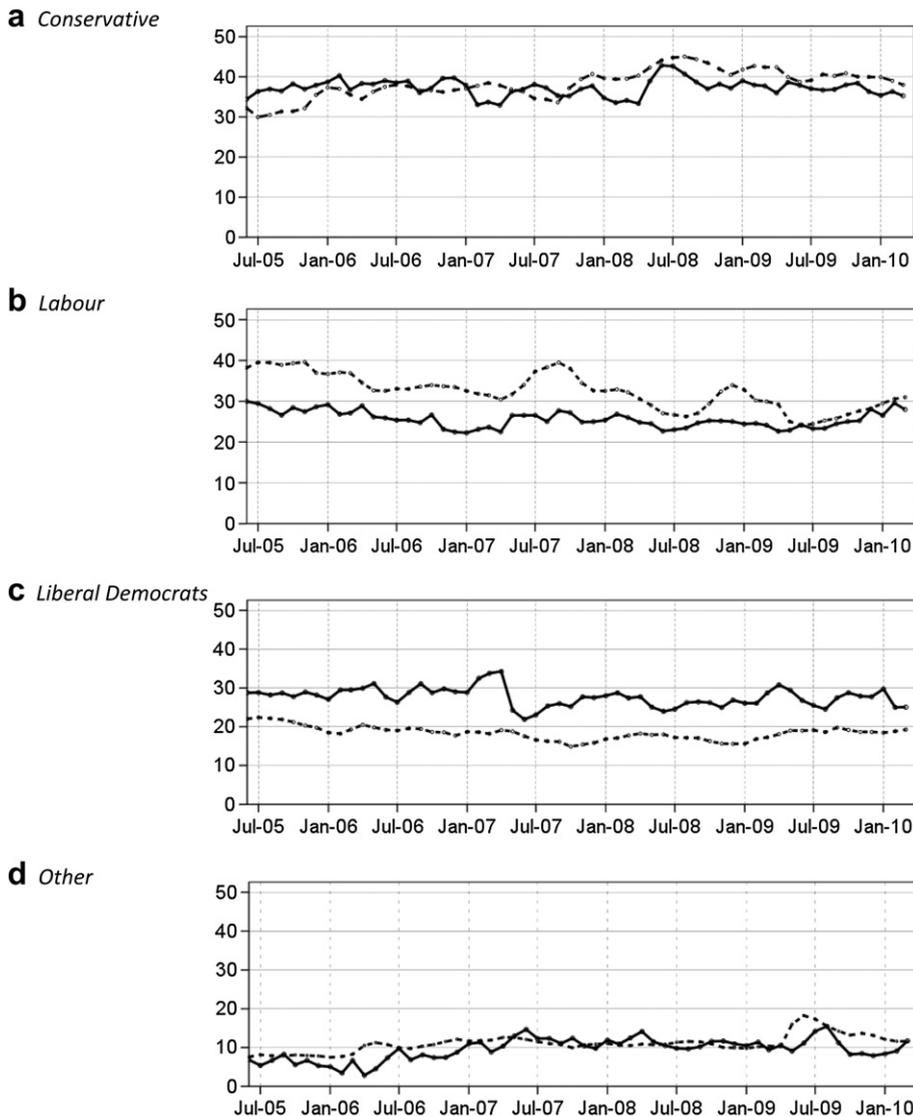


Fig. 3. The by-election model and opinion polls, 2005–2010. (a)Conservative, (b)Labour, (c)Liberal Democrats, (d)Other.

role in brokering international agreements contributes to another bounce in the polls for his party. In the meantime, Labour's by-election performance has remained lacklustre with any improvement barely registering on this measure. In June 2009 the two measures intersect for the first time this parliament when the weighted poll average reaches 24.1% and presumably hits the bedrock support for Labour; the by-election model had been estimating support of around that level for the previous year.

The post 2005 data for the Liberal Democrats confirm the pattern of the previous two decades. The party consistently performs better at local elections than its national poll rating suggests (Fig. 3c). For the eighteen months following the 2005 general election the party hovers on or around 30% but January 2007 sees a significant improvement in its fortunes, reaching a four-year high of 34% in April that year. But then there is a rather dramatic 10-point collapse in May and June followed by a recovery over the autumn and winter months. Polling suggests that party support has ranged over a few percentage points throughout the entire parliament, entering the run-in to the 2010 general election a point or two lower than for the previous election.

Finally, Fig. 3d shows the trend in support for other parties. The two measures are close together for most of the time although the influence of the approaching European elections in June 2009, which triggers a growth in support for such parties in the opinion polls is delayed slightly for the by-election model (a case of polls driving votes?) and is short-lived although as the general election nears the two lines may be intersecting.

Any student of parliamentary by-elections and their outcomes would know that they are not reliable guides to how people might vote at a general election but it does appear that aggregate local electoral data are picking up similar movements in electoral opinion to the individual level data acquired in national surveys. The votes cast there are real votes and voters have incurred some costs in performing these actions. There is no need to adjust them in terms of weighted past vote or the likelihood of actually voting but there is a need to take account of variance in the pattern of party competition. On some occasions the by-election model takes time to respond to issues that are immediately apparent in the polls but on other occasions they may be a better guide to the underlying trends. They provide a more reliable indicator of how local voters may behave than surveys but what do they portend for the 2010 general election.

6. Estimating the 2010 vote shares

At the outset we stated that the model is principally designed for a purpose different to the one outlined here. Judged solely on the basis of its ability to forecast national equivalent vote shares for the annual local election cycle it is a success, seldom being more than a couple of percentage points out for any single party if general election years are excluded. There is no doubt, however, that for the model to work to estimate parliamentary voting then certain adjustments may be required, dependent upon the general election context. In the past the main adjustment affects

support for the third party, the Liberal Democrats, both in terms of the level of protest voting that it has received and also the extent of split-ticket voting (either because they genuinely prefer to vote Liberal Democrat at local council elections or because of tactical voting reasons). The context of the 2010 general election, however, is that support for the two main parties has declined with the Liberal Democrats the principal beneficiary of this swing against the two-party system establishment. There are other, perhaps less prominent factors, that should also be taken into account when adapting this model to generate general election forecasts. One such is Labour's general weakness in apparently getting its vote out. Even allowing for some over-estimation of its support by the pollsters, Labour has under-achieved in the past, with the notable exception of the pre-1997 period.

The April NEV estimate is showing Conservatives 35.7%, Labour 27.2% and Liberal Democrats on the same figure, 27.2%. In preparing this general election forecast we assume, in a departure from the recent past, that no downward adjustment of the Liberal Democrat vote is necessary. Indeed, we are proposing on this occasion to add a percentage point to both its and Labour's support to take account of additional supporters for these parties entering the general election contest. This produces, after rounding, a national vote share forecast of Conservative 36%, with Labour and Liberal Democrats each on 28%. However, an additional adjustment is made to the seat forecast based on a more complete understanding of by-election voting in Conservative-target parliamentary constituencies. Two separate forecast models were created, one comprising only wards located in the top 150 Conservative-target parliamentary constituencies with all other wards entered into the second model. This demonstrated that the Conservatives were performing slightly better in their target seat areas vis a vis Labour. Accordingly, the seat projection does not assume national uniform swing but a rather higher swing from Labour to the Conservatives in their battleground seats. The seat forecast, therefore, is Conservatives 299 seats, Labour 237 seats, Liberal Democrats to win 83 seats while other parties win 31 seats.

Appendix

Estimation of current national equivalent vote shares

The model may be formally expressed as follows:

Let **CONsh**, **LABsh**, and **LDsh** be the ward vote share at the by-election for Conservative, Labour and Liberal Democrats respectively while the parties' vote share at the previous ward election held in May would be **CONsh_{May}**, **LABsh_{May}**, and **LDsh_{May}**. The national equivalent vote at the date of the relevant previous May local election is designated as **NEV.CON**, **NEV.LAB**, and **NEV.LD**.

The difference in share of vote between May and By-elections for enumerated parties might then be represented as follows:

$$\begin{aligned} \text{CONchange} &= \text{CONsh} - \text{CONsh}_{\text{May}} \\ \text{LABchange} &= \text{LABsh} - \text{LABsh}_{\text{May}} \\ \text{LDchange} &= \text{LDsh} - \text{LDsh}_{\text{May}} \end{aligned}$$

Finally, to estimate the current national equivalent vote we add to the previous national equivalent vote the difference between a party's by-election vote share and its vote share recorded at the May election.

Estimates of current NEV are calculated as follows,

NEV.CON + **CONchange**, **NEV.LAB** + **LABchange**, **NEV.LD** + **LDchange**.

This process is repeated for as many by-elections as for our specified criteria and each party's current national equivalent vote is arrived at by averaging the results over a stated time period.

Missing data imputation

Example 1: Imputation for missing vote share for Liberal Democrats absent from by-election contest.

Let all three main parties have candidates in May election but Liberal Democrats provide no candidate in the by-election:

May election → **By-election**

con + lab + ld → con + lab

The imputation procedure implies the following notional by-election result:

$$\begin{aligned} LDsh^{new} &= \text{MINIMUM}_{\text{across district in May}}(LDsh) \\ CONsh^{new} &= CONsh - CONsh / \\ & (CONsh + LABsh) \times LDsh^{new} \\ LABsh^{new} &= LABsh - LABsh / \\ & (CONsh + LABsh) \times LDsh^{new} \end{aligned}$$

In order to preserve the actual *direction* of changes for all parties some restrictions are imposed:

If $CONsh > CONsh_{May}$ but $CONsh^{new} < CONsh_{May}$ then we set **CONchange** = 0.

If $LABsh > LABsh_{May}$ but $LABsh^{new} < LABsh_{May}$ then **LABchange** = 0.

Note: In above formulae and everywhere else in the paper, superscript 'new' reflects notional by- or May election when any procedure of imputation is applied. Following the imputation procedure, notional election result (i.e. 'new') is then treated in the usual way for the purposes of estimating NEV.

Example 2: Imputation for missing vote share for both Labour and Liberal Democrats absent from by-election contest (con + lab + ld → con)

Notional By-election result:

$$\begin{aligned} LDsh^{new} &= \text{MINIMUM}_{\text{across district in May}}(LDsh) \\ LABsh^{new} &= \text{MINIMUM}_{\text{across district in May}}(LABsh) \\ CONsh^{new} &= CONsh - LABsh^{new} - LDsh^{new} \end{aligned}$$

If $CONsh > CONsh_{May}$ but $CONsh^{new} < CONsh_{May}$ then we set **CONchange** = 0.

If $CONsh \leq CONsh_{May}$, then

LABchange = **MISSING**, **LDchange** = **MISSING**;
CONchange = $CONsh - CONsh_{May}$

Example 3: Imputation for missing Labour by-election share in case when Liberal Democrats absent from both May and by-election contest (con + lab → con)

LDchange = **MISSING**

Notional By-election result:

$$\begin{aligned} LABsh^{new} &= \text{MINIMUM}_{\text{across district in May}}(LABsh) \\ CONsh^{new} &= CONsh - LABsh^{new} \end{aligned}$$

If $CONsh > CONsh_{May}$ but $CONsh^{new} < CONsh_{May}$ then we set **CONchange** = 0.

Example 4: Imputation for May vote share for Liberal Democrats (con + lab → con + lab + ld)

Notional May-election results:

$$\begin{aligned} LDsh_{May}^{new} &= \text{MINIMUM}_{\text{across district in May}}(LDsh) \\ CONsh_{May}^{new} &= CONsh_{May} - CONsh_{May} / \\ & (CONsh_{May} + LABsh_{May}) \times LDsh_{May}^{new} \\ LABsh_{May}^{new} &= LABsh_{May} - LABsh_{May} / \\ & (CONsh_{May} + LABsh_{May}) \times LDsh_{May}^{new} \end{aligned}$$

Restrictions:

If $LDsh < LDsh_{May}^{new}$ then **LDchange** = 0.

If $CONsh < CONsh_{May}$ but $CONsh^{new} > CONsh_{May}^{new}$ then we set **CONchange** = 0.

If $LABsh < LABsh_{May}$ but $LABsh^{new} > LABsh_{May}^{new}$ then we set **LABchange** = 0.

Example 5: Imputation for missing Labour May vote share in case when Liberal Democrats absent from both May and by-election contest (con → con + lab)

There is no LD candidate in May and by-elections. So, we cannot assess LD changes:

LDchange = **MISSING**

Notional May-election results:

$$\begin{aligned} LABsh_{May}^{new} &= \text{MINIMUM}_{\text{across district in May}}(LABsh) \\ CONsh_{May}^{new} &= CONsh_{May} - LABsh_{May}^{new} \end{aligned}$$

Restrictions:

We have to get at least a non-negative change for LAB, so

If $LABsh < LABsh_{May}^{new}$ then **LABchange** = 0.

If $CONsh < CONsh_{May}$ but $CONsh^{new} > CONsh_{May}^{new}$ then we set **CONchange** = 0.

Example 6: Imputation for missing May vote shares for both Labour and Liberal Democrats (con → con + lab + ld)

Notional May-election results:

$$\begin{aligned} LABsh_{May}^{new} &= \text{MINIMUM}_{\text{across district in May}}(LABsh) \\ LDsh_{May}^{new} &= \text{MINIMUM}_{\text{across district in May}}(LDsh) \\ CONsh_{May}^{new} &= CONsh_{May} - LABsh_{May}^{new} - LDsh_{May}^{new} \end{aligned}$$

Then usual procedure for calculation of changes and estimated shares is applied.

Restrictions:

We have to get at least a non-negative change for LAB and LD, so

If $LABsh < LABsh_{May}^{new}$ then **LABchange** = 0.

If $LDsh < LDsh_{May}^{new}$ then **LDchange** = 0.

If $CONsh < CONsh_{May}$ but $CONsh^{new} > CONsh_{May}^{new}$ then we set **CONchange** = 0.

Example 7: Imputation for missing vote shares for both May and by-elections for different parties (con + lab → con + ld)

Notional May-election results:

$$\begin{aligned} LDsh_{May}^{new} &= \text{MINIMUM}_{\text{across district in May}}(LDsh) \\ LABsh_{May}^{new} &= LABsh_{May} - \text{MINIMUM}_{\text{across district in May}}(LABsh) \\ CONsh_{May}^{new} &= CONsh_{May} - LDsh_{May}^{new} + LABsh_{May}^{new} \end{aligned}$$

Restrictions:

We have to get at least a non-negative change for LD:

if $LDsh < LDsh_{May}^{new}$ then **LDchange** = 0.

If $CONsh < CONsh_{May}$ but $CONsh^{new} > CONsh_{May}^{new}$

or $CONsh > CONsh_{May}$ but $CONsh^{new} < CONsh_{May}^{new}$

then we set **CONchange** = 0.

References

Curtice, J., Payne, C., 1991. "Local elections as national referendums in Great-Britain." *Electoral Studies* 10 (1), 3–17.

- Rallings, C., Thrasher, M., 1993. "Exploring uniformity and variability in local electoral outcomes – some evidence from English local elections 1985–1991." *Electoral Studies* 12, 366–384.
- Rallings, C., Thrasher, M., 1998. "Split-Ticket Voting at the 1997 British General and Local Elections: An Aggregate Analysis." In: Denver, D., Fisher, J., Cowley, P., Pattie, C. (Eds.), *British Elections & Parties Review*. Frank Cass Publishers, London.
- Rallings, C., Thrasher, M., 1999. "Local votes, national forecasts – using local government by-elections in Britain to estimate party support." *International Journal of Forecasting* 15, 153–162.
- Rallings, C., Thrasher, M., 2001. "Measuring the level and direction of split-ticket voting at the 1979 and 1997 British general and local elections: a survey-based analysis." *Political Studies* 49 (2), 323–330.
- Rallings, C., Thrasher, M., 2003. "Explaining split-ticket voting at the 1979 and 1997 general and local elections in England." *Political Studies* 51 (3), 558–572.
- Ranney, A., 1962. "The Utility and Limitations of Aggregate Data in the Study of Electoral Behaviour." In: Ranney, A. (Ed.), *Essays on the Behavioral Study of Politics*. University of Illinois Press, Urbana.