BallotMaps: Detecting Name Bias in Alphabetically Ordered Ballot Papers

Jo Wood, *Member, IEEE*, Donia Badawood, Jason Dykes, and Aidan Slingsby *Member, IEEE*

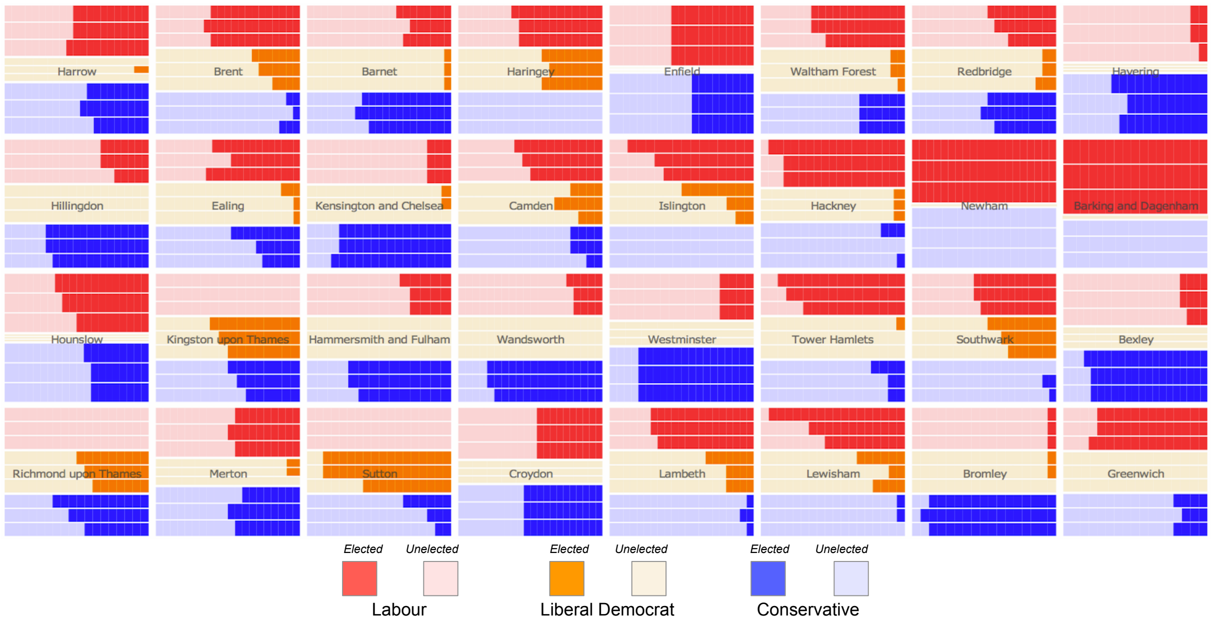

Fig. 1. BallotMap showing electoral success (or otherwise) of each candidate for the three main parties in wards (small rectangles) in each London borough (grid squares) in the 2010 local government elections. Vertical ordering of candidates within each borough is by ballot paper position within party (top row first, middle row second, bottom row third). Main parties with three candidates in a ward are shown. If no ballot ordering bias existed there would be no systematic structure to bar lengths. This ballotMap shows that more candidates get elected who are listed first within their party than do candidates who are second or third.

Abstract—The relationship between candidates' position on a ballot paper and vote rank is explored in the case of 5000 candidates for the UK 2010 local government elections in the Greater London area. This design study uses hierarchical spatially arranged graphics to represent two locations that affect candidates at very different scales: the geographical areas for which they seek election and the spatial location of their names on the ballot paper. This approach allows the effect of position bias to be assessed; that is, the degree to which the position of a candidate's name on the ballot paper influences the number of votes received by the candidate, and whether this varies geographically. Results show that position bias was significant enough to influence rank order of candidates, and in the case of many marginal electoral wards, to influence who was elected to government. Position bias was observed most strongly for Liberal Democrat candidates but present for all major political parties. Visual analysis of classification of candidate names by ethnicity suggests that this too had an effect on votes received by candidates, in some cases overcoming alphabetic name bias. The results found contradict some earlier research suggesting that alphabetic name bias was not sufficiently significant to affect electoral outcome and add new evidence for the geographic and ethnicity influences on voting behaviour. The visual approach proposed here can be applied to a wider range of electoral data and the patterns identified and hypotheses derived from them could have significant implications for the design of ballot papers and the conduct of fair elections.

Index Terms—Voting, election, bias, democracy, governance, treemaps, geovisualization, hierarchy, governance.

1 Introduction

There has long been a suspicion from candidates standing for election that the order in which candidate names appear on a ballot paper may, in part, influence the number of votes received (e.g. [18]). Despite a number of studies investigating the degree of this effect [24, 1, 13, 5, 20, 19], evidence appears inconclusive and sometimes contradictory. This paper considers how information visualization may be designed and applied to investigating the degree to which some form of name bias may exist in influencing votes received by candidates.

The overall aims of the work are twofold: to identify the degree to which the position of candidate name affects numbers of votes received; and to develop a data visualization design appropriate for exploring the spatial and non-spatial influences over candidate votes. Addressing these aims is important because conducting fair and neutral elections is an essential part of the democratic process. The aims

- *Jo Wood (jwo@city.ac.uk), Donia Badawood (Donia.Badawood.1@city.ac.uk), Jason Dykes (j.dykes@city.ac.uk), and Aidan Slingsby (sbbb717@city.ac.uk) are at the giCentre (http://gicentre.org/) in the School of Informatics, City University London.*

Manuscript received 31 March 2011; accepted 1 August 2011; posted online 23 October 2011; mailed on 14 October 2011.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

are timely because access to detailed digital electoral results and legislation is becoming increasingly easy (e.g. [4, 6, 15]) and is considered an important aspect of participatory democratic accountability (e.g. [7]. In her 2008 position paper addressing the grand challenges of information visualization, Tamara Munzner called for work in visualization that leads to *"total political transparency... through civilian oversight of data on voting records, campaign contributions... "* [14]. She stated that while such data are available in theory, they are in practice not understandable by the citizens meant to be empowered by them. This work, while not analysing voting records of politicians, makes accessible patterns in the voting behaviour that led to their selection and rejection. We assert that this is an important part of democratic accountability enabled though good visualization design.

This design study examines results of the Greater London local elections held on the 6th May 2010. These elections took place in parallel with the UK general election where members of Parliament and the national government were elected. Within the Greater London area, 6829 candidates were standing for 1842 positions across 614 electoral wards. These wards were in turn aggregated into 32 London boroughs each of which formed a local government responsible for administering local services such as transport and education. Unlike the national general election, each ward in the local elections studied here had three electable positions available and voters were permitted to vote for up to three candidates. All ballots used the 'first past the post' system, where the three candidates with the largest number of votes in each ward were elected. Most candidates stood as part of a political party slate, and their party affiliation was indicated on the ballot paper. All candidates for each ward were listed on the ballot paper from top to bottom in alphabetical order (see sample ballot paper in Fig. 2).

For the purposes of this study, we limited analysis to only those candidates from the three major political parties in the UK – the centre-left Labour Party, the centrist Liberal Democratic Party and the centre-right Conservative Party. Since these parties provided three candidates for almost all wards, this allowed us to investigate the effect of name ordering within parties as well as between parties without the need to account for an uneven distribution of candidates. Of the total of 6829 candidates, 5973 of them stood for one of the three main parties and 5025 of them were in wards where the main parties offered three candidates for election. This sample size of 5025 (74%) was sufficiently large to allow potential bias in voting behaviour (i.e. influences other than party preference, or individual candidate characteristics) to be identified.

We identified four research questions around which we conducted our visualization design and application:

1. To what extent does the position of a name on a ballot paper affect the number of votes received by a candidate within their party?

2. To what extent does the position of a name on a ballot paper affect the number of votes received by a candidate independently of their party?

3. How does any name bias vary geographically within the Greater London area?

4. Does the apparent ethnicity of candidates as indicated by their name contribute to any name bias?

The challenge in designing an appropriate visual means to explore these research questions was to deal with two separate scales of spatial pattern – that of the spatial arrangement of names on the ballot paper and the hierarchical geospatial arrangement of candidates and voters in the Greater London area. Together with political party and ethnicity suggested by candidate name these provided four sets of independent variables that may have an influence on number of votes received and who was elected to local government for each of the 5025 candidates in the sample. Given the spatial relationships and exploratory nature of this investigation into the extent of any such patterns, a means by which alternative visual arrangements of the data could be quickly constructed was necessary.

VOTE FOR NO MORE THAN THREE CANDIDATES

1	**AARON** **Lawrence Aaron** 17 Newington Road, London N1 6FG Liberal Democrats	Liberal Democrats	
2	**CHADWELL** **Gertrude Chadwell** 22 Some St, London N1 2AB UK Independence Party	UKIP	
3	**CROUSE** **Justin Crouse** (Address in constituency) The Labour Party Candidate	Labour	
4	**DEBOSE** **Joanne Debose** 16 Acer Avenue, London NW4 8XT Green Party	Green Party	
5	**HANDY** **William Handy** (Address in constituency) The Labour Party Candidate	Labour	
6	**HOOPER** **Malcolm Hooper** (Address in constituency) The Conservative Party Candidate	Conservatives	
7	**KOZLOWSKI** **Michael Kozlowski** (Address in constituency) The Conservative Party Candidate	Conservatives	
8	**NOOR** **Anjit Noor** (Address in constituency) The Labour Party Candidate	Labour	
9	**PFEIFFER** **Dale Pfeiffer** 103 Elephant Way, London NW1 8RH Liberal Democrats	Liberal Democrats	
10	**TALLY** **Deborah Tally** (Address in constituency) The Conservative Party Candidate	Conservatives	
11	**WHITFIELD** **Sarah Whitfield** 45 Kingham Place, London N1 6SL Liberal Democrats	Liberal Democrats	
12	**YILMAZ** **Shaquil Yilmaz** 4 Pocklington Walk, London N1 5DS Independent Candidate		

Fig. 2. Sample ballot paper from the London May 2010 local election showing alphabetic arrangement of fictitious candidates on the page. Each voter can vote for up to three candidates on the ballot paper (order of preference not indicated).

2 PREVIOUS WORK

The work conducted here was informed by previous studies on both voting behaviour specifically, and choice selection in general. Krosnick proposed the theory of recency and primacy when looking at how choices are made from a selection when presented sequentially in time (recency) and in space (primacy) [9]. The primacy effect is expected when there are number of choices listed vertically; this means that the first item in the list is the more likely to be chosen than others in the list. It suggests that an alphabetic ordering of candidate names on a ballot paper would favour those with names closer to the head of the alphabet. What is less clear is whether the tendency towards primacy is sufficiently strong to overcome other preferences expressed in the voting booth such as knowledge of the candidate's record, party preference or other judgements based on the name of the candidate.

Rallings and colleagues analysed English local election results including London borough elections between 1991 and 2006 [20, 19]. With regards to the positioning of names, both studies suggested a statistically significant name ordering effect and concluded that candidates listed first on the ballot paper within their party were more likely to get more votes than candidates listed second or third. The Electoral

Commission report [24] considered this research and concluded that while alphabetical discrimination may occur in multi-member elections, the strength of this effect was not sufficient to suggest it had a significant outcome on who was elected.

Several other studies around the world support the view that there may be an effect of name positioning on the ballot paper. For example, Lijphart and Pintor analysed the 1982 and 1986 senate elections in Spain and concluded that the candidate listed higher on the ballot paper enjoyed a vote advantage over the next candidate from the same party [11]. Studies in the USA supported this view such as the work of Koppell and Steen who assessed name-ordering effects by rotating names on a ballot paper so all candidates would occupy the first position once [8]. They concluded that a candidates performance was much stronger when listed first on the ballot paper than any other position. A similar experiment conducted by Ho and Imai who randomized names on a ballot came to similar conclusions [5].

On the other hand, some studies suggest that the position on the ballot paper has little or no effect on candidates' performance [1, 13]. There is also some doubt about the effect of positions other than first in a candidate list. Pack conducted an analysis of name position effect on election of candidates within the Liberal Democrat party [18]. He suggested that being either at the top or bottom of a list might confer a weak advantage over those listed in the middle when candidates were selected from a ballot paper with vertical ordering.

What is clear from previous work is that there is evidence for some degree of advantage to those listed first in a ballot, but neither the degree of that advantage, nor whether patterns exist for names in other positions on the ballot are known. We can find no studies that look at a possible geographic effect, a name ethnicity effect, nor any that use visualization to explore and communicate these patterns.

3 Exploring alphabetical name bias

Data on the votes received, borough and ward names, political parties and names of all candidates who stood in the May 6th 2010 London local elections were retrieved from the London Data Store [3]. The site provided by the Greater London Authority (GLA – London's regional government) with the stated aim to allow *"citizens to be able access the data that the GLA and other public sector organisations hold, and to use that data however they see fit – free of charge... Raw data often doesn't tell you anything until it has been presented in a meaningful way. We want to encourage the masses of technical talent that we have in London to transform rows of text and numbers into apps, websites or mobile products which people can actually find useful"* [4].

The data were cleaned, requiring the correction of 20 errors out of the 6829 candidate records. Errors were identified where vote tallies did not correspond with the records of elected candidates and where duplicate records appeared in the database. For these records, the correct names and vote tallies were found by consulting the relevant local government source via the web. Ballot paper position for each candidate was calculated by performing an alphabetical comparison with candidates sharing the same ward. The cleaned election data were georeferenced by identifying each ward centroid from Ordnance Survey's OpenData Boundary-Line dataset [17]. Together these data sources generated a set of nine primary variables for each candidate (see Table 1). Variables were divided into three groups: those relating to geospatial location; those relating to the candidate name; and those relating to the votes received by the candidate and their political party. From these data a further set of six secondary derived variables was calculated in order to assess the degree of alphabetical name bias (see Table 2).

The signed chi statistic [25] was calculated to give an indication as to the variation in votes acquired by candidates relating to issues other than party affiliation as

$$\chi = \frac{obs - exp}{\sqrt{exp}} \tag{1}$$

where the expected number of votes for each candidate was one third of the total party votes for their ward (each candidate in the sample stood in a ward with two other candidates from the same party) and

Table 1. Primary candidate variables extracted from the London Data Store and OS Boundary-Line datasets

Location	Name	Votes
Borough name	Candidate name	# candidate votes
Ward name	Alpha position in ballot (1-24)	Party
Easting		Was elected (y/n)
Northing		

Table 2. Secondary derived variables constructed for visual exploration in HiDE

Name	Votes	Combined
Alpha position in party (1-3)	# party votes	Signed chi
	% of party vote	Residual
	Vote order in party (1-3)	

the observed value was the actual number of votes received by the candidate. Thus positive values of χ indicate that the candidate received more than the expected number of votes if only political party was assumed to influence candidate choice, while negative values indicate fewer than expected votes were received.

The residual measure was designed to identify anomalies that did not show name ordering bias and was calculated as the difference between the percentage of party votes received by a candidate and that expected for an average candidate with the same 'alpha' (alphabetical) position with their party. Thus while the chi statistic assesses the degree of name order bias, the residual identifies candidates that have greater or fewer votes than predicted given their party affiliation having taken any name order bias into account.

Chi-squared analysis of the numbers of votes received by candidates in each of the alpha positions in their party revealed a significant ordering effect in excess of the 99% confidence level for within party variation and 95% level for received votes independent of party. However, a visual approach was investigated to see how this ordering effect may vary with respect to some of the other variables in our dataset.

3.1 Hierarchical Visualization

The design challenge was to provide a visual exploration environment to represent the interrelations between the 15 variables identified in Tables 1 and 2. Four of those variables had a spatial component (borough location, ward location, alpha position in ballot and alpha position in party) while the remaining 11 were non-spatial. Standard graphical techniques for examining relationships between variables, such as faceted scatterplots or scatterplot matrices, were not seen as sufficient for a number of reasons. Firstly, many of the variables we wished to explore were categorical (boroughs, wards, candidate names, party, was elected) or comprised a small number of values (alpha position within party, vote order within party). These do not lend themselves to depiction in plots more suited to continuous measurement data. Secondly we wished to be able to compare many variables simultaneously, requiring sufficient graphical space to use hue and brightness/saturation color components simultaneously. Point based symbolisation techniques are not well suited to discriminating use of color in this way.

While the Greater London area is largely urban, the density of households varies considerably within the region, and consequently the size of boroughs varies in order to approximately balance the number of constituents in each unit (see Fig. 3). Consequently, to aid visualization we used a spatial treemap layout [27] to produce a rectangular cartogram of the distribution of boroughs and the wards within them. Here, the graphical area of each borough is fixed (which roughly approximates its voting population and aids graphical comparison) rather than reflecting geographical area (which has little bearing on the voting behaviour). Additionally, we wished to reflect the spatial arrangement of candidate names on the ballot paper in our exploratory graphics. We achieved this by ordering representations of individual candidates vertically from top to bottom in the same order in which they appeared on the ballot paper. By combining with color to reflect the non-spatial variables we constructed a framework for the creation of *ballotMaps*.

Fig. 3. The 32 London boroughs in the study area. With the exception of 'City' (not part of the election), boroughs vary in area roughly to equalise voting population.

Fig. 1 shows an example ballotMap that reflects the geospatial distribution of wards and spatial distribution of names on the ballot paper. Each large square represents a borough positioned approximately in relation to its geographic location (northerly boroughs towards the top of the ballotMap, inner boroughs in the centre etc.) [27]. Each is divided into three horizontal rectangles showing the three main political parties symbolized by hue. The size of each is proportional to the number of candidates in each party standing for election. Small rectangles represent individual candidates either in a light color if not elected or darker shade if they were. Candidate rectangles are ordered vertically according to their position on the ballot paper within their party, so that candidates who were alphabetically first within their party appear in the top row, second in the middle row and third in the bottom row. Candidates are ordered from left to right according to electoral success.

This particular depiction shows some evidence of name bias that itself is geographically and party related. If electoral success were based only on party preference and candidate suitability there should be no relation with ballot paper position. We would therefore expect the horizontal length of each dark bar to be roughly similar for each party in each borough; any variation being random. But Fig. 1 clearly shows a trend towards top bars being longer than middle bars being longer than lower bars. This is particularly evident in the boroughs of Islington, Richmond upon Thames, Sutton and Lewisham where candidates listed first in their party are more likely to be elected than those second or third. Some boroughs show this effect largely for certain parties, such as the (blue) Conservatives in Ealing and the (orange) Liberal Democrats in Brent. A few boroughs appear to show no ordering effect, such as Bromley and Croydon and a few others where party preference dominates the distribution of elected councillors, such as Newham and Barking and Dagenham.

Given there are 15 possible variables to view and many thousands of combinations of layout, color, size and order for each, an environment for rapid exploration of design possibilities was required. To do this we used HiVE (Hierarchical Visualization Expression [22]) to encode the visual design options and used HiDE (Hierarchical Data Explorer [2])[1] to implement the encoded designs. This allowed rapid exploratory design options to be implemented, shared and evaluated. We regard both the ability to rapidly vary design options and the direct exploration of the data as important and complementary contributors to effective information visualization. The former, which has been termed 're-expression' in the cartographic context [21], allows the adaptation of visual encoding to match the research questions being investigated [23].

The design of Fig. 1 can be summarised in HiVE as:

```
sHier(/,$borough,$party,$alphaInParty,
                        $candidate,$isElected);
sOrder(/,[$x,$y],[NULL,HIER],[NULL,HIER],
                  [$success,NULL],[HIER,NULL]);
sSize(/,FX,$numCandidates,FX,FX,FX);
sColor(/,NULL,NULL,NULL,NULL,HIER);
sLayout(/,SF,SF,SF,SF,SF);
```

After using HiDE to explore many alternative representations of the data shown in Fig. 1, it became apparent that considering only who was or was not elected was not sufficient to identify the extent of possible name order bias. In particular, in wards and boroughs where party loyalty was strong (e.g. Conservative Bromley and Labour Barking and Dagenham), any possible ordering effect within parties was being hidden. To investigate these hidden effects, we used HiDE to construct ballotMaps comparing alpha order of candidates with their rank order by number of votes received regardless of whether or not they were elected. This allowed us to explore name ordering effects even in wards with strong party preferences. In HiDE this simply involved substituting the variable `$isElected` with `$partyAndVoteOrder` ('Vote order in party (1-3)' in Table 2). The results are shown in Fig. 4.

The name ordering effect can be seen more strongly here, with top row of most boroughs being significantly darker and more saturated than the bottom row. Unlike previous studies, this pattern also shows a significant difference between the vote order of candidates in second and third alpha position on the ballot. Geographical variation in this ordering effect is also evident. The southern boroughs tend to show a lighter lower row than the northern ones, and the effect is particularly strong for the Liberal Democrats (e.g. Richmond, Sutton, Lambeth). Towards the east in the traditionally working class region Barking and Dagenham, where few Liberal Democrat candidates stood for election, there is little ordering effect for the other two main parties. Because each party provided three candidates for election here, a strong party preference results in little opportunity to select candidates by ballot position. In contrast, the prosperous boroughs west and southwest of central London show a strong ordering effect for right of center Conservative candidates (e.g. Kensington and Chelsea, Hammersmith and Fulham, Richmond upon Thames) where there may be more tendency for voters to split their three votes between Conservative and Liberal Democrat candidates. Likewise, the less prosperous boroughs to the east of central London show a strong ordering effect for left of center Labour candidates (e.g. Lewisham, Tower Hamlets, Hackney, Waltham Forest) where votes may be inclined to split their selection between Labour and the centrist Liberal Democrat candidates. The proportion of candidates affected by name order is summarised in Fig. 5, indicating that on average, a candidate listed first in their party is 6.3 times as likely to get the most votes in their party than a candidate listed third. The effect is strongest for Liberal Democrat candidates; a candidate listed first in their party is 8.6 times more likely to get the most party votes than one listed third.

To identify the degree to which ballot position influences votes received, we constructed ballotMaps showing the signed chi values for each candidate. Fig. 6 shows this for each party. Assuming an expected value for each candidate of exactly one third of the total votes

[1] The source code to HiDE is available at code.google.com/p/viztweets. An application with the voting data and figures used here is available as an addendum to this paper.

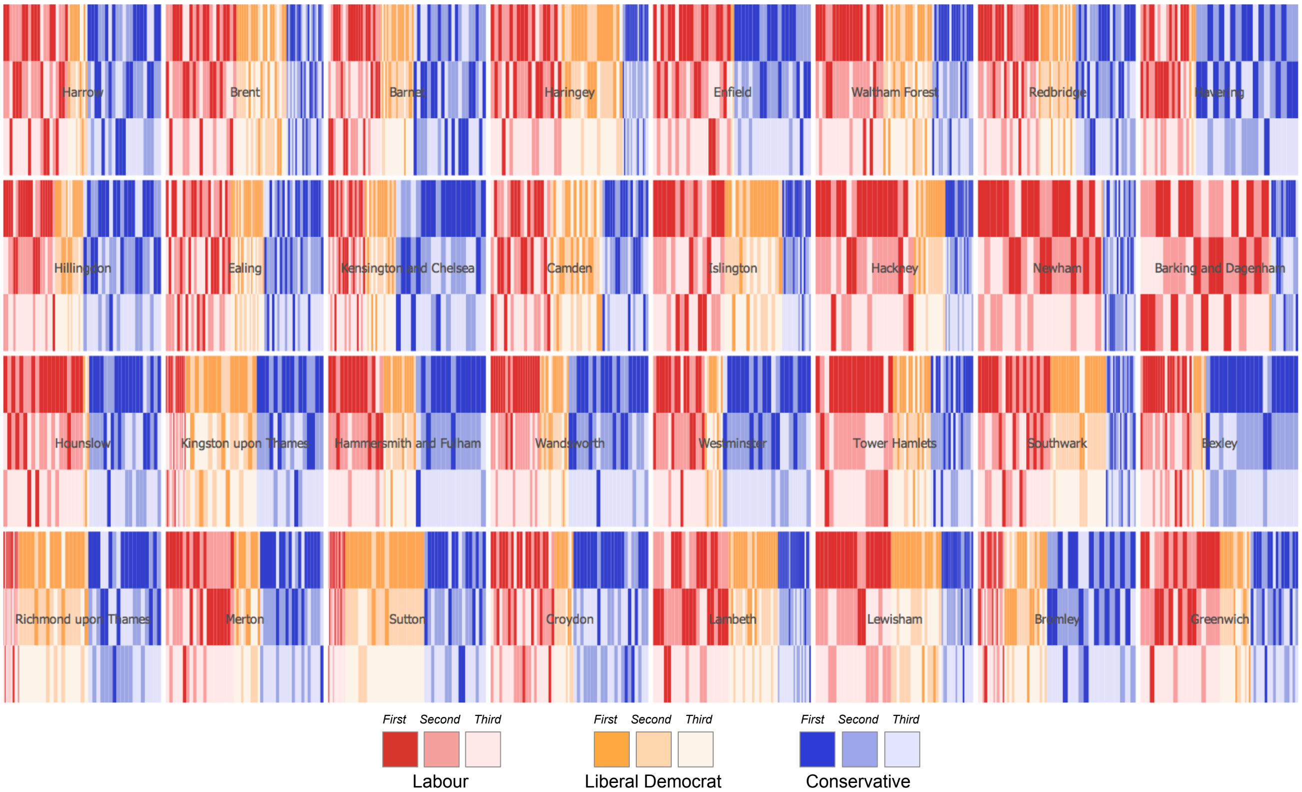

Fig. 4. Alpha position within party (vertical position) and voting rank within party for the three main parties in each ward (vertical bars) in each borough (grid squares). If no name order bias existed, dark and light cells would be randomly distributed in the top, middle and bottom thirds of each borough. Actual voting data show that darker cells (indicating a candidate most votes within their party) are more common in the upper third (listed first on the ballot paper within their party) and lighter cells (least votes within party) are more common in the lower third (listed third within their on the ballot paper).

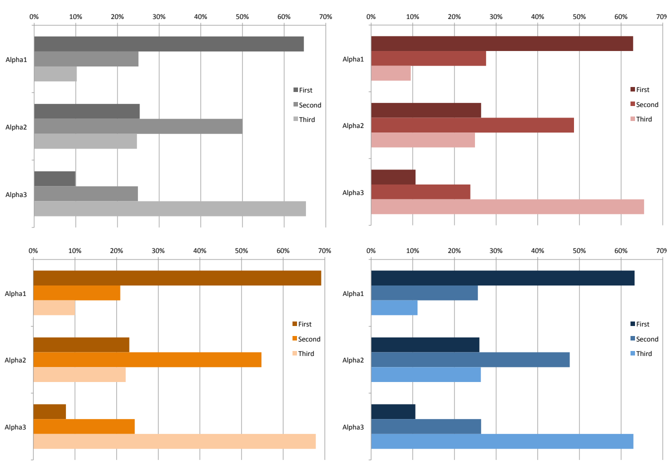

Fig. 5. Alpha position and vote order for, all candidates (gray); Labour candidates (red); Conservative (blue) and Liberal Democrat (orange). If no name order bias existed, all bars would be about the same length.

for their party in their ward, the chi ballotMap clearly shows the systematic ordering effect when values sorted graphically from top to bottom by order within party then order on ballot paper. If there was no ordering effect, we would expect a random distribution of purple and green cells. By breaking down the distributions by party, it is also evident that the strongest ordering effect is for Liberal Democrat candidates (the top and bottom thirds of the central column in Fig. 6 are generally darker green and darker purple than the top and bottom thirds of the left and right columns representing the other two parties). Labour candidates positioned first in their party show a slightly stronger ordering effect than Conservative candidates.

One of the benefits of the chi ballotMap is that it is also possible to identify anomalous candidates who do not appear to be subject to an alphabetical ordering effect. These appear as purple cells in the upper third of the ballotMap or green cells in the lower third. Through interactive query in HiDE we were able to quickly browse the names of these candidates, which suggest there may be an alternative source of name-related bias.

4 Other Sources of Name Bias

Interactive query of anomalous candidate names suggested there might be an association with the apparent ethnicity implied by the name. Initially we created tag clouds comprising all names of candidates positioned first in their party with a negative residual value - that is, those who received less than the average percentage party vote for candidates positioned first. In order to indicate whether this distribution of names was systematically different to those of all candidates in alpha1 position, we then compared this distribution with tag clouds of random selections drawn from the alpha1 sample. We borrowed from the process of graphical inference [26] to compare the observed values (alpha1 names with negative residuals) with a null hypothesis assuming no structure to anomalies (random samples from alpha1). While this indicated there might be some degree of ethnicity bias present, we wished to examine the structure of that bias in more detail.

To investigate possible ethnicity bias, we allocated each candidate to a class relating to the likely ethnic origin of their name using OnoMAP [12, 16]. This classification, evaluated for use in public health policy [10], compares given and family names to classify each pair into one of 16 possible OnoMAP ethnic groups. The distribution

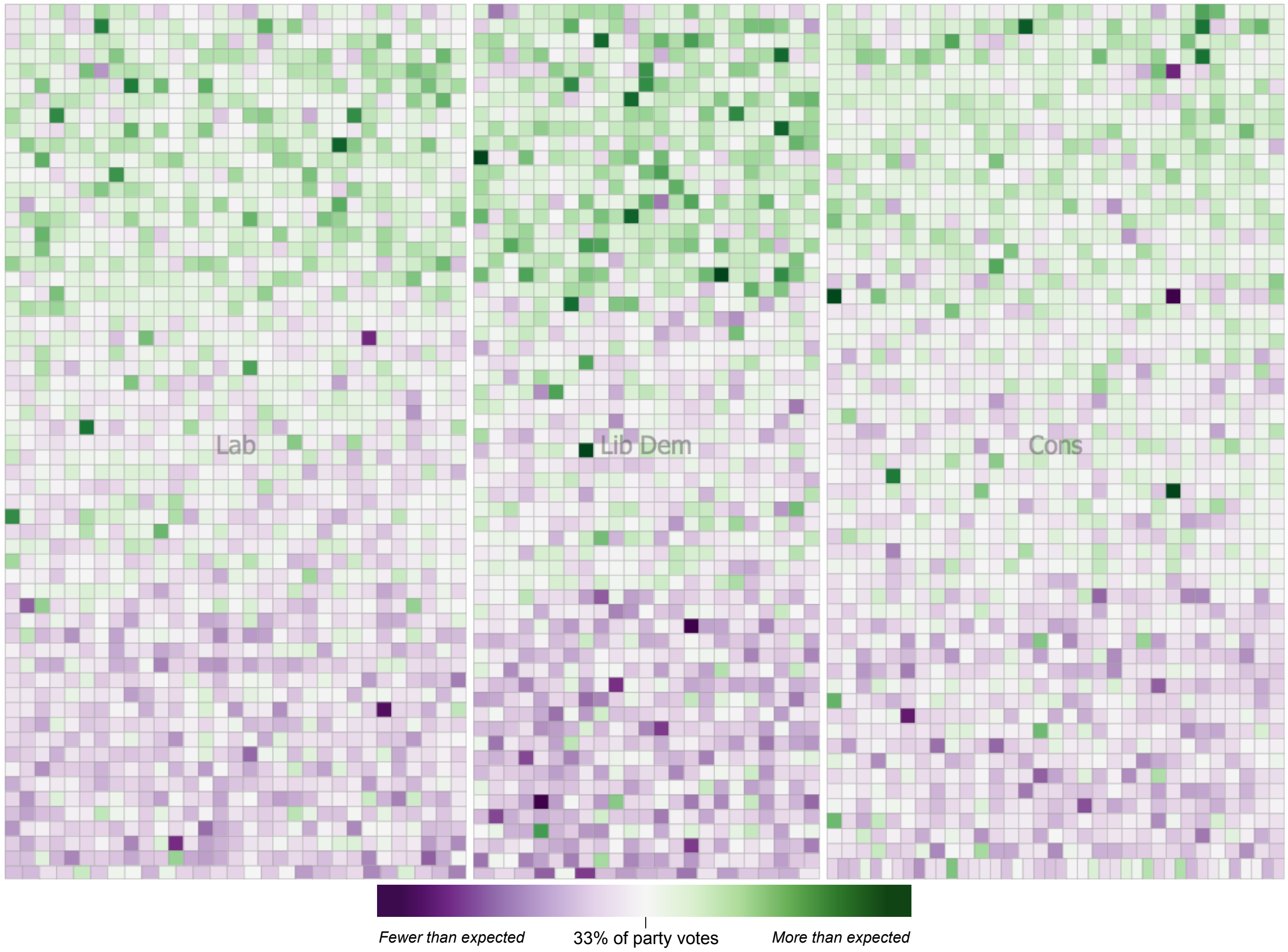

Fig. 6. Signed chi values for each candidate arranged by party (left to right) and ballot position within party (top to bottom, row by row). The top third represents candidates ordered first in their party, then ordered by absolute position on the ballot paper; the middle third represents candidate ordered second within their party etc. If no name order bias existed, purple and green cells would be randomly distributed within the ballotMap.

of candidates using this classification is shown in Table 3. While there may be some inaccuracy in classification and origin of name does not necessarily indicate ethnicity of candidate, we felt this was a valid process in this instance since we are attempting to measure the effect of the name itself on voter behaviour, not knowledge of the candidate directly.

Table 3. OnoMAP ethnicity categories of candidates

OnoMAP category	Number of candidates
African	58
Celtic	886
East Asian and Pacific	14
English	3018
European	124
Greek	31
Hispanic	22
International	7
Jewish and Armenian	27
Muslim	489
Nordic	8
Sikh	110
South Asian	204
Unclassified	27

The numbers of candidates in some of the OnoMAP groups were too small to draw significant conclusions, and there was also a question of the discriminating power of voters in being willing or able to distinguish between certain groups. We therefore chose to group all candidates into two broad groups – 'English or Celtic', comprising the OnoMAP 'English' and OnoMAP 'Celtic' groups of names that are likely to originate in the British Isles, and 'Other Name Origins', comprising all other name origin groups. Fig. 7 shows the chi values for all candidates broken down by these two super-groups.

The BallotMap shows that there are approximately similar numbers of candidates from both ethnic super-groups in all alpha positions, but that name ordering bias is much higher in the 'English or Celtic' group. The 'other name origins' group shows many more purple candidates with fewer votes than expected in the alpha1 position. This suggests that, for some candidates at least, a propensity not to select a candidate due to their non British Isles name origin outweighs a propensity to select them because they are positioned first within their party on the ballot paper.

To explore whether this effect had any geographical component, we constructed BallotMaps showing the chi values by ethnic super-group for each borough (see Fig. 8). The ballotMap shows that the ethnicity of candidates varies by geography, for example the western boroughs of Harrow, Brent, Ealing and Hounslow having a higher proportion of 'other name origin' candidates compared with southern boroughs of Richmond, Merton, Sutton, Bromley and Greenwich. All boroughs show a name bias in the 'English or Celtic' supergroup (upper thirds greener than lower thirds), but in the 'other name origins' group, the pattern is more varied. In many of the outer boroughs, the alpha1 can-

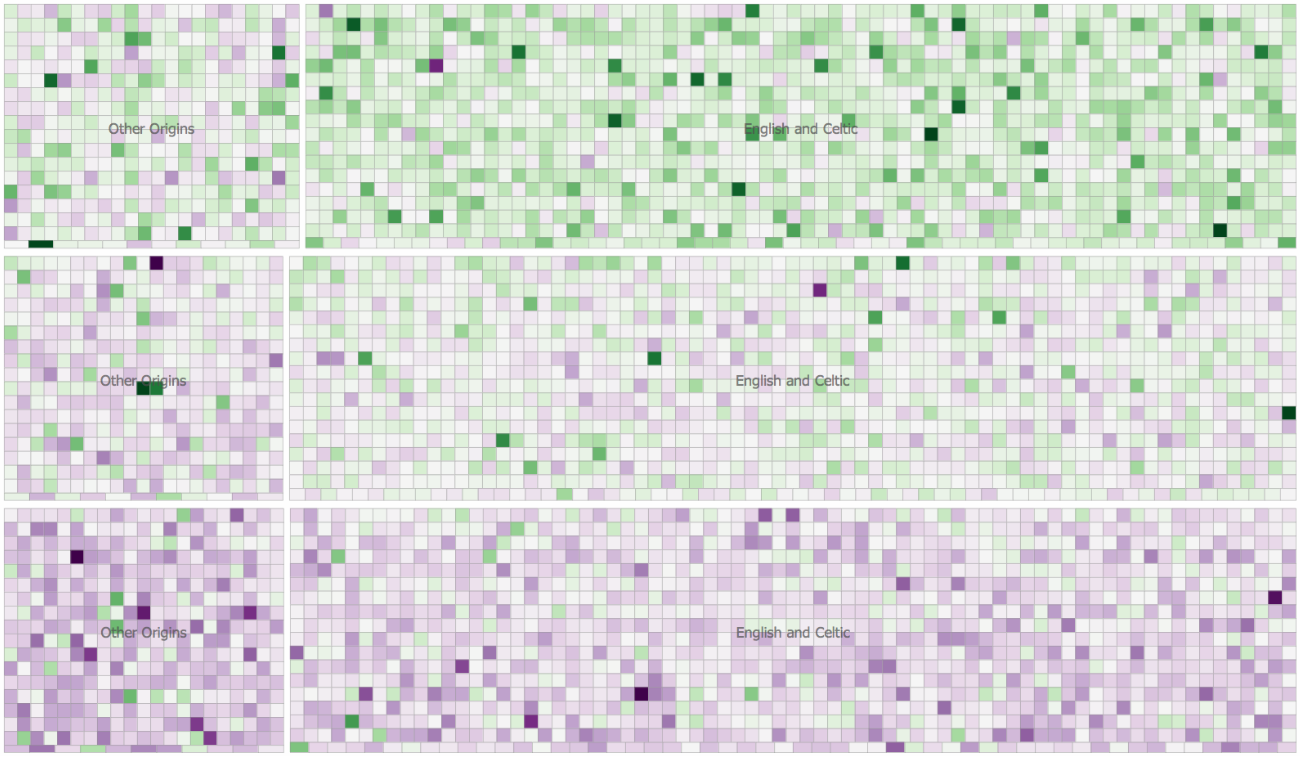

Fig. 7. Signed chi values for candidates arranged by binary classification of name origin and ballot position within party (top to bottom, row by row). The top third represents candidates ordered first in their party, then ordered by absolute position on the ballot paper; the middle third represents candidate ordered second within their party etc. Name order bias (tendency for green cells in the upper third and purple in the lower third) is stronger for 'English or Celtic' names than for other names where candidates listed first are not so likely to get more votes than expected.

didates show fewer than expected votes (purple cells in the top left of the 32 squares representing each borough), for example Brent, Harrow, Kingston, Sutton, Bromley and Greenwich. In contrast, some of the inner boroughs with higher numbers of candidates in the 'other name origins' super-group show a name ordering bias within this group that is similar to or stronger than that seen in the 'English or Celtic' group (e.g. Hackney, Tower Hamlets, Newham).

5 Conclusion

The aims of this study were twofold – to identify the degree to which candidate name influenced votes received in a multi-member election, and to consider appropriate information visualization design for analysing and communicating spatial and non-spatial data on democratic decision making. Our first aim was broken down into four research questions. We have shown that ballot position did indeed strongly influence the number of votes received by candidates in the most recent local government elections in London and that some of those who are currently representing London may have benefitted from this effect just as those who are not suffered from it. The effect was sufficiently strong to confer first positioned candidates a 6-times advantage over third positioned candidates in the same party. Visual exploration revealed for which parties (Liberal Democrats) and in which locations (southern outer boroughs) this effect was most strong. There is some evidence that the strength of this effect is sufficient to overcome voter preference for party, most likely in marginal seats where voters are prepared to allocate their three votes to more than one party. This affects the centrist Liberal Democrats most strongly because they are more likely to receive votes from voters willing to support one of the other two main parties (fewer voters willing to split their vote between the right of centre Conservatives and the left of centre Labour party).

Exceptions to name-ordering bias were identifiable though visual means and this led to the exploration of influence of apparent ethnicity of choice of candidates by voters. Here, we were able to identify where ethnicity bias overcame name ordering bias (outer boroughs) and where it reinforced name ordering bias (inner boroughs with higher proportion 'other name origin' candidates).

The results found here warrant further investigation, to see if these patterns are consistent over time and are reflected in other areas. The design of BallotMaps provides a framework for doing this. BallotMaps use consistent layout rules to reflect the micro-scale spatial arrangement of names on ballot papers and macro-scale geospatial arrangements of political wards. Unlike scatterplots for examining associations between variables the graphical space filling character of ballotMaps allow more effective use of colour symbolisation to represent additional variables. While this is also a characteristic of some other graphical statistics such as mosaic plots, the freedom to rearrange variables dynamically within the hiaearchy allows different layout options to be found more rapidly. Maintaining consistent use of layout and color symbolisation helps to navigate the large design space suggested by complex multi-variate public datasets. That design space can be rapidly explored using the HiVE/HiDE framework, and was found to be crucial in arriving at layout and symbolisation rules that answered our research questions directly.

It remains to be seen whether in doing so we have managed to "*transform rows of text and numbers into apps, websites or mobile products which people can actually find useful*" [4], but the resulting design and approach may lead to Londoners and others questioning both the alphabetical ordering that prevails on many ballot papers and their own voting behaviours, perhaps going some small way towards addressing Munzner's grand challenge of total political transparency.

Acknowledgments

The authors are grateful to Pablo Meteos, University College London, for providing access to OnoMAP for classification of candidate names. The development of HiDE was funded through the JISC VRE Rapid Innovation programme under the vizTweets award.

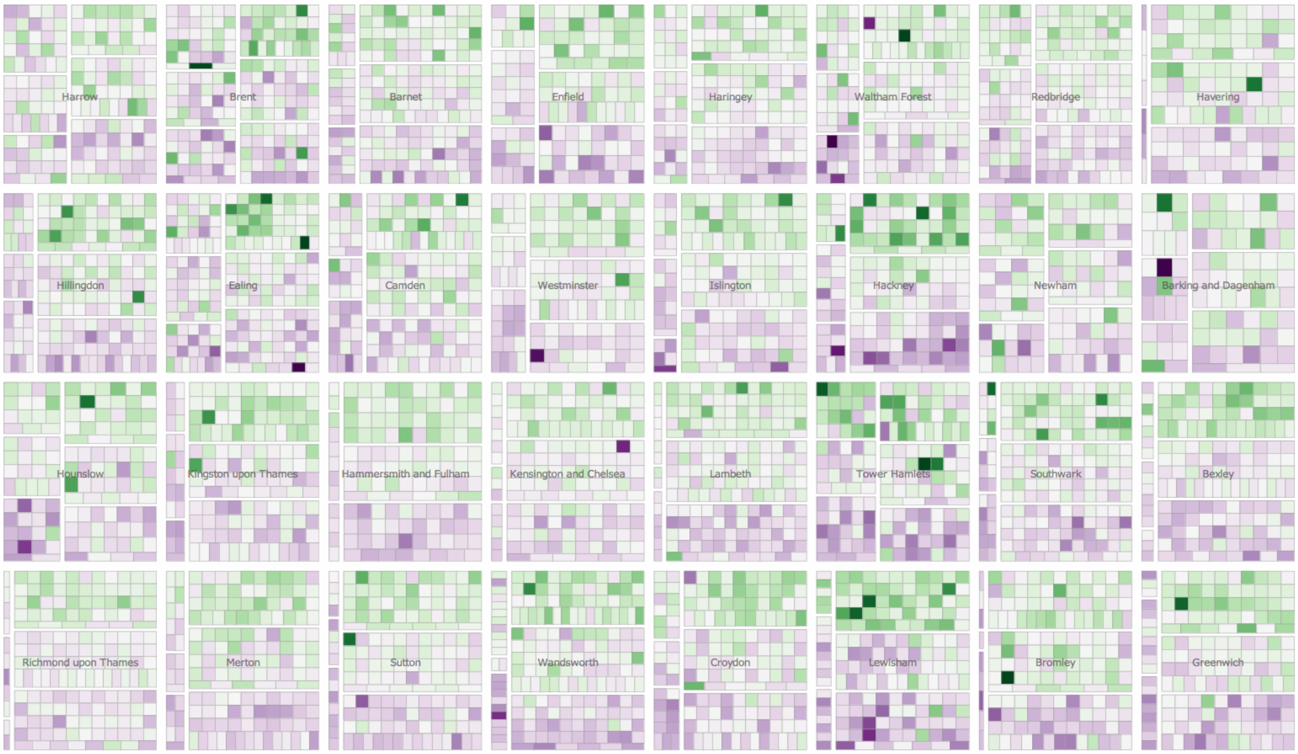

Fig. 8. Signed chi values for each candidate in each borough arranged by binary classification of name origin ('other name origins' left, 'English or Celtic' right) and ballot position within party (top to bottom). The degree of name order bias is indicated by the strength of separation of green (more votes than expected) and purple (fewer votes than expected) cells. This varies by borough and by ethnic origin of candidate names.

References

[1] R. Darcy. Position effects with party column ballots. *The Western Political Quarterly*, 39(4):648–662, 1986.

[2] giCentre. Hierarchical data explorer. `http://www.gicentre.org/hide`, 2011.

[3] GLA. Borough council election results. `http://data.london.gov.uk/datastore/package/borough-council-election-results-2010`, 2010.

[4] GLA. London data store. `http://data.london.gov.uk`, 2011.

[5] D. Ho and K. Imai. Estimating causal effects of ballot order from a randomized natural experiment. *Public Opinion Quarterly*, 72(2):216–240, 2008.

[6] IBM. Many bills: A visual bill explorer. `http://manybills.researchlabs.ibm.com`, 2011.

[7] P. Kinnaird, H. Rouzati, and X. Sun. Connect 2 congress. In *IEEE Information Visualization Poster Extended Abstracts*, Atlantic City, 2009. IEEE.

[8] J. G. Koppell and J. A. Steen. The effects of ballot position on election outcomes. *The Journal of Politics*, 66(1):267–81, 2004.

[9] J. Krosnick. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5:213–236, 1991.

[10] F. Lakha, D. Gorman, and P. Mateos. Name analysis to classify populations by ethnicity in public health: Validation of onoMAP in Scotland. *Public Health*, in press, 2011.

[11] A. Lijphart and R. L. Pintor. Alphabetic bias in partisan elections: Patterns of voting for the spanish senate, 1982 and 1986. *Electoral Studies*, 7(3):225–231, 1988.

[12] P. Mateos, R. Webber, and P. Longley. The cultural, ethnic and linguistic classification of populations and neighbourhoods using personal names. In *CASA Working Paper*, volume 116. University College London, 2007.

[13] J. Miller and J. Krosnick. The impact of candidate name order on election outcomes. *Public Opinion Quarterly*, 62(3):291–330, 1998.

[14] T. Munzner. Outward and inward grand challenges. In *VisWeek08 Panel: Grand Challenges for Information Visualization*, Columbus OH, 2008. IEEE.

[15] mySociety. They work for you. `http://www.theyworkforyou.com`, 2011.

[16] OnoMAP. OnoMAP. `http://www.onomap.org`, 2011.

[17] Ordnance Survey. Ordnance survey opendata. `https://www.ordnancesurvey.co.uk/opendatadownload/products.html`, 2010.

[18] M. Pack. Does the alphabet matter when it comes to Liberal Democrat internal elections? `http://bit.ly/libdemvoice`, 2010.

[19] C. Rallings, M. Thrasher, and G. Borisyuk. Ballot structure, parties and voters: Measuring effects at the local elections 2002-2006. *Proceedings, Elections, Opinion Polls and Parties Annual Conference*, Nottingham University, 2006.

[20] C. Rallings, M. Thrasher, and C. Gunter. Patterns of voting choice in multimember districts: the case of english local elections. *Electoral Studies*, 17(1):111–128, 1998.

[21] I. Shepherd. Putting time on the map: Dynamic displays in data visualization and GIS. In P. Fisher, editor, *Innovations in GIS*, volume 2, pages 169–187. Taylor & Francis, London, 1995.

[22] A. Slingsby, J. Dykes, and J. Wood. Configuring hierarchical layouts to address research questions. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):977–984, 2009.

[23] A. Slingsby, J. Dykes, and J. Wood. Tweeting visualizations for collaborative visual analysis. In *Poster Proceedings, IEEE Infovis*, Salt Lake City, 2010.

[24] The Electoral Commission. *Ballot Paper Design*. Electoral Commission, London, 2003.

[25] M. Visvalingam. The signed chi-score measure for the classification and mapping of polychotomous data. *Cartographic Journal*, 18(1):32–43, 1981.

[26] H. Wickham, D. Cook, H. Hofmann, and A. Buja. Graphical inference for infovis. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):973–979, 2010.

[27] J. Wood and J. Dykes. Spatially ordered treemaps. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1348–1355, 2008.