HEC MONTRÉAL

Communication of Uncertainty for Statistical Modelling of Election Outcomes

par

Digvijaysinh Jadeja

Léo Belzile
HEC Montréal
Directeur de recherche

Sciences de la gestion
(Spécialisation *User Experience in a Business Context*)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences
(M. Sc.)

May 2021

Abstract

Several studies have estimated voting outcomes for multiparty elections, but little attention has been paid to understanding their variability. From the perspective of the user, this uncertainty must be taken into account in decision-making and understanding election results. In this thesis, we attempt to address the uncertainty in the election outcomes of England for the United Kingdom (UK) general elections through the use of Bayesian multinomial regression incorporating demographic and historical data. We also introduce an offset to measure changes in outcomes caused by changes in voting intentions.

In the first chapter, we lay the foundation on UK's election dynamics by describing the UK's electoral system and its history and conducting an exploratory analysis of the relationship between demographic constituency information from the census and the election results. We also conduct a literature review of election outcome modelling techniques. The second chapter introduces the Bayesian framework for statistical modelling and the use of the STAN programming language. We also present specification techniques for multinomial regression involving categorical outcomes. The third chapter focuses on an assessment of the model with and without random effects and evaluation of its sensitivity to prior specification. In the last chapter, we focus on techniques to communicate election outcomes and the associated uncertainty that help improve interpretation by a broad audience.

Keywords: Uncertainty, Election outcomes, Multinomial regression, STAN, Bayesian statistics, Random effect, Hexagonal maps, Visualization, User experience

Contents

List of Tables

List of Figures

Acknowledgements

Its been an absolute privilege to complete my master's degree at HEC Montreal as it provided me with a valuable experience that helped me to improve my skills and gain knowledge in the relevant domains. It gave me exposure to various opportunities, which in turn refined my vision for research. I want to express my gratitude to all the professors and mention special thanks to my thesis supervisor Professor Léo Belzile whose guidance and invaluable insights helped me tackle roadblocks to complete my thesis. This thesis would not have been possible without the support of my friends and family.

Chapter 1

The UK Elections

1.1 Research Objective

Election forecasts in news outlets are more popular than ever. However, the media commonly overestimate the precision of polls, cherry-pick data, and then paint the election outcomes as a sure thing (Silver, 2017). Even with accurate models, there are widespread errors of interpretation and communication. Consumers may misinterpret forecasts because they do not factor in the dependencies between outcomes. In addition, people making the forecasts may not adequately convey the uncertainties involved. Thus from a user experience perspective, the predicted election outcomes must be communicated effectively in order to allow the users to gain correct insights. Moreover, effective communication becomes even more important when users are not aware of the local political process (Gupta et al., 2016). The communication of election results has two facets: the uncertainty attached to the predicted outcomes, and the visualization of these outcomes. In this thesis, we will consider both dimensions. Our main objective is to build a model that captures the uncertainty in the predicted outcomes and our secondary objective is to create a visualization framework for conveying these outcomes.

A number of methods involving complex regression techniques have been proposed

for predicting elections. Often, the resulting predictions are point estimates without attached uncertainties. From a user perspective, uncertainty and variability are really important in terms of decision-making because polling data need not reflect the outcomes on election day. For example, when new polls are incorporated into the model a wide range of factors could influence the election prediction. Techniques that simply provide deterministic changes in outcome do not capture the uncertainty of these potential changes in results, in contrast to probabilistic models.

Perceptions and cognitive experiences are not only influenced by the type of data provided for interpretation but are also impacted by the visualization of that information. In this thesis, we focus on results for England due to the UK's complex political dynamics. The objective of our research is to build a baseline predictive model using Bayesian multinomial logistic regression. Doing so allows us to predict a new posterior distribution of predicted election results for England while using an offset to account for changes in voting intentions that could be generalized to provide dynamic forecasts that account for the change in voting intention and poll uncertainty. Lastly, we study the best techniques for representing and visualizing predicted election outcomes and their attached uncertainties.

We use the demographic data from the UK 2011 census and the UK election results provided by the UK parliament website to model past election outcomes. This first chapter focuses on the UK election system. We also perform a basic exploratory data analysis to observe the relationship between election outcomes and demographic data and examine basic spatial visualization tools through examples. Chapter 2 explains the Bayesian modelling concept and its implementation in STAN, a probabilistic programming language. In the third chapter, we build a predictive model encapsulating uncertainty. The last chapter examines state-of-the-art research techniques to create a visualization framework explaining election outcomes and the uncertainty associated with them.

1.2 UK Electoral System

The legislative assembly of the United Kingdom (UK) comprises elected representatives from each of the 650 constituencies present in the UK. Of these constituencies, 533 are located in England, 59 in Scotland, 40 in Wales, and the last 18 are situated in Northern Ireland. In the UK, members of parliament are elected to represent their constituencies via the first-past-the-post system. In the latter, voters cast their vote for a single candidate of their choosing, and the candidate with the highest number of votes is elected to represent the constituency (Parliament, 2021).

According to the UK Office for National Statistics, the average electorate per constituency in England is approximately 72,200 voters. Most of the competition in the UK is among the four main parties, namely, the Conservative, Labour, Scottish National Party (SNP), and Liberal Democrats. The Conservatives stand for private property, a strong military, and preservation of traditional cultural values (Louth and Webb, 2019). The Labour Party promotes economic prosperity and social provisions. The Liberal Democrats can be said to follow the center-left school of thought. The SNP's defining goal has always been independence for Scotland; thus, their candidates do not stand in England or Wales constituencies. Due to the complex political dynamics in the UK, we will focus solely on England.

1.3 Election History

In the last 25 years, only the Conservatives and Labour have held power in Westminster. Labour won three successive elections in 1997, 2001, and 2005, and, over that same period, the Liberal Democrats improved their voter share. Figure 1.1 summarizes results of the past four elections in England (2010, 2015, 2017, and 2019). The 2010 elections resulted in a large swing in favour of the Conservatives, who won 306 seats across the UK but fell 20 seats short of a majority, resulting in a hung parliament. The Conservatives and Liberal Democrats went on to form a coalition government.

In the next election cycle (2015), polls and commentators predicted the outcome would be too close to call and would result in a second consecutive hung parliament. Polls underestimated the Conservatives, as the party unexpectedly won a majority (330 seats) in the UK, giving them a small majority of 12 seats and their first mandate in 23 years. The Liberal Democrats had their worst showing since the formation of their party, retaining just eight of their 57 seats. The next UK general elections occurred just two years later in 2017. According to the *Fixed-Term Parliaments Act of 2011*, parliament is automatically dissolved after five years, but there are specific cases in which early elections can take place:

- A motion of no confidence is passed against Her Majesty's Government
- A motion for a general election is agreed upon by two-thirds of the Members of Parliament

Elections were not due until 2020 but two-thirds of the House of Commons voted to hold a snap election. The Conservatives did well but failed to retain their majority, which led to the formation of a Conservative-led minority government with a confidence-and-supply agreement with the Democratic Unionist Party (DUP) of Northern Ireland.

The most recent UK general elections were held in 2019. The Conservative Party had faced prolonged parliamentary deadlock over Brexit while it governed in the minority with the support of the DUP. This deadlock led to another early election. The Conservatives racked up a net gain of 48 seats in a landslide victory, giving them a comfortable 80-seat majority. Many Conservative gains were made in constituencies long-held by Labour in the North of England. The Labour Party won 202 seats, its lowest number and proportion of seats since 1935. The Liberal Democrats improved their vote share to 11.6% but won only 11 seats, a net loss of one since the last election.

In view of this brief historical review, it is clear that UK elections are highly complex. The purpose of this work is to examine uncertain outcomes as seats that have huge majorities are not in play even if the average swing is 5000 votes from one election to the

next but closely contested seats change hands even with small swings.

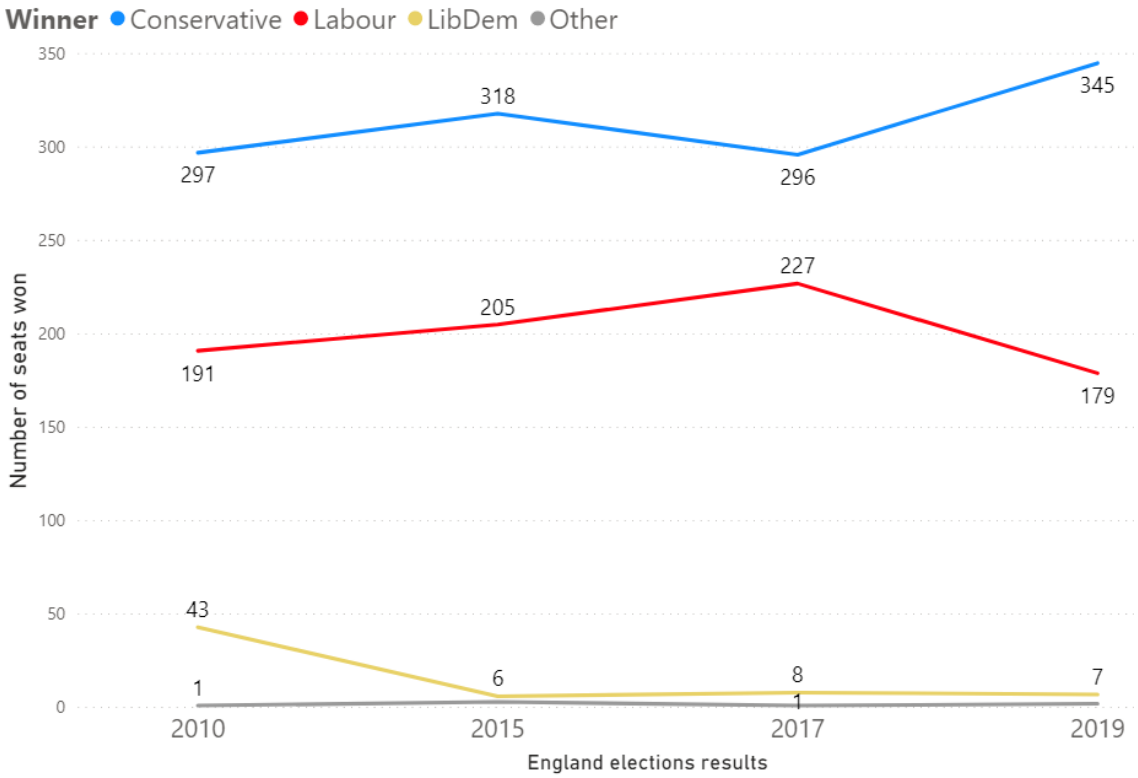


Figure 1.1 – Line chart of England election results for the UK 2010, 2015, 2017, and 2019 general elections

1.4 Exploratory Data Analysis

1.4.1 Data Sources

Our two primary data sources are both made available by the UK Parliament. One is the UK General election results dataset, and the second one is the 2011 Census dataset (aggregated at the constituency level). For the purpose of the model, we combined these data sources. Further, since the UK conducts a decadal census, and the results of the 2021 Census are not yet available, we assume that the census data is constant over time. Our modelling process must be able to account for multiple parties. Additionally, there are many third parties, which will be included together in an "other parties" category. For the purpose of this exercise, we will add their votes together, which will result in an overestimation of the level of support. Subsequently, we will analyze the election results for 2010-2019 based on the number of votes received in each constituency by the Conservatives, Labour, Liberal Democrats, and other parties and use demographic features as covariates. The boundaries of electoral districts were different prior to the year 2010 which is why we restrict our attention to the last four elections.

The data is organized and structured in wide format, with our target variables being the votes for the Conservatives, Labour, Liberal Democrats, and Other Parties. The list of covariates employed in our later model is given in Table 1.1.

Census Data	**Election Data**
Above age 65 (%)	Constituency
White (%)	Region
Christian (%)	Year
Unemployed (%)	Number of votes per party/constituencies
Employed (%)	
Student (%)	
House owned (%)	

Table 1.1 – List of variables used from demographic and election results data

1.4.2 Demographic Data

Table 1.1 lists the demographic variables that will be of prime interest in our model development. We first consider the percentages of white residents and christian residents in the constituencies. Further, the percentages of unemployed, employed, and students in a particular constituency provide us with information on the economic activity within that constituency. By examining the percentage of residents above age 65, we can identify voting patterns for a given age group. The last variable represents the percentage of individuals in a constituency who own their homes. To examine demographic voting patterns, we use bar plots and percentage stacked bar plots for the demographic variables with bucket sizes of 10 percentage points. We focus on 2017 election data for visualization purposes.

For the race variable (white), the data in Figure 1.2a indicates that the Conservatives hold more than half of the seats in constituencies that have more than 80% whites. Further, when the percentage of white voters declines, the winning rate for Labour increases. According to Figure 1.2b, the number of multi-cultural constituencies with less than 50% white voters is relatively low.

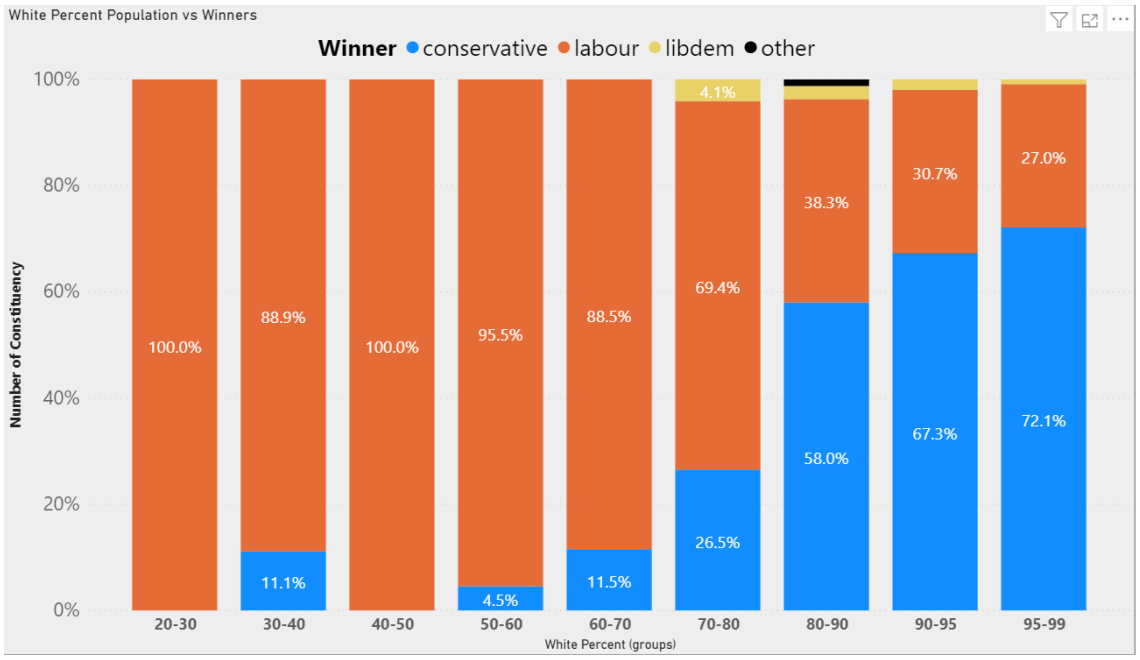
In Figure 1.3a, we observe that constituencies with higher shares of young voters lean towards Labour. In contrast, constituencies with higher shares of voters who are aged 65 and above lean towards the Conservatives. The winning percentage for Conservative candidates in the bin for 25–32 (%) whites is also high. However, in Figure 1.3b, we observe that the number of such constituencies is smaller. This voting pattern is consistent with survey data from YouGov (McDonnell and Curtis, 2017).

In Figure 1.4a, when the percentage of home owners in a constituency is less than 40%, Labour candidates having a winning rate exceeding 90%. Also, when the percentage of home owners is less than 50%, the Labour winning rate is low. Note, however,

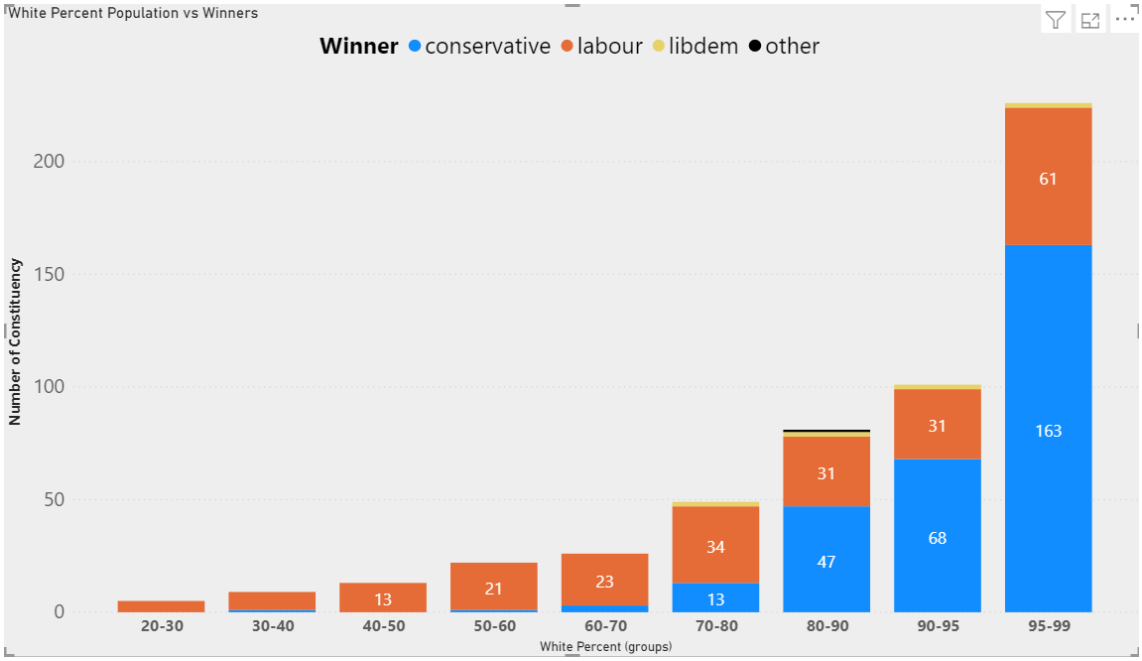
that in Figure 1.4b, the number of constituencies with the percentage of home owners less than 50% is not high.

The employment data in Figure 1.5b indicates that constituencies with relatively high rates of employment favor Conservatives. Seventy percent of constituencies with employment rates of 60–70% elected Conservatives; on the other hand, 80% of constituencies with employment rates between 50 and 60% elected Labour candidates. Finally, the percentage of Christians does reveal a voting pattern. Nevertheless, we do observe that constituencies that have at least 50% Christian voters usually have more than 45% percent of their seats go to the Conservative party.

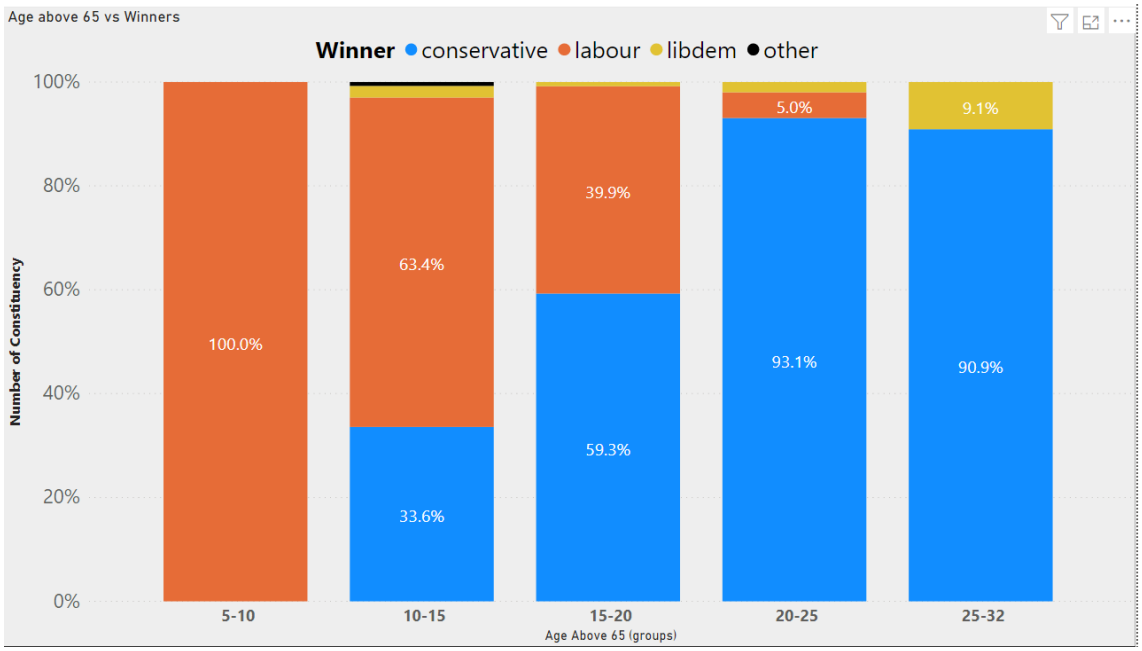
All the observation above clearly illustrates that voting patterns exist within certain demographic groups in England. Thus, this demographic data provides important input for a multinomial logistic regression model of election outcomes. We will also capture the uncertainty regarding the outcomes through the inclusion of the demographic features as well as random effects at the constituency level.



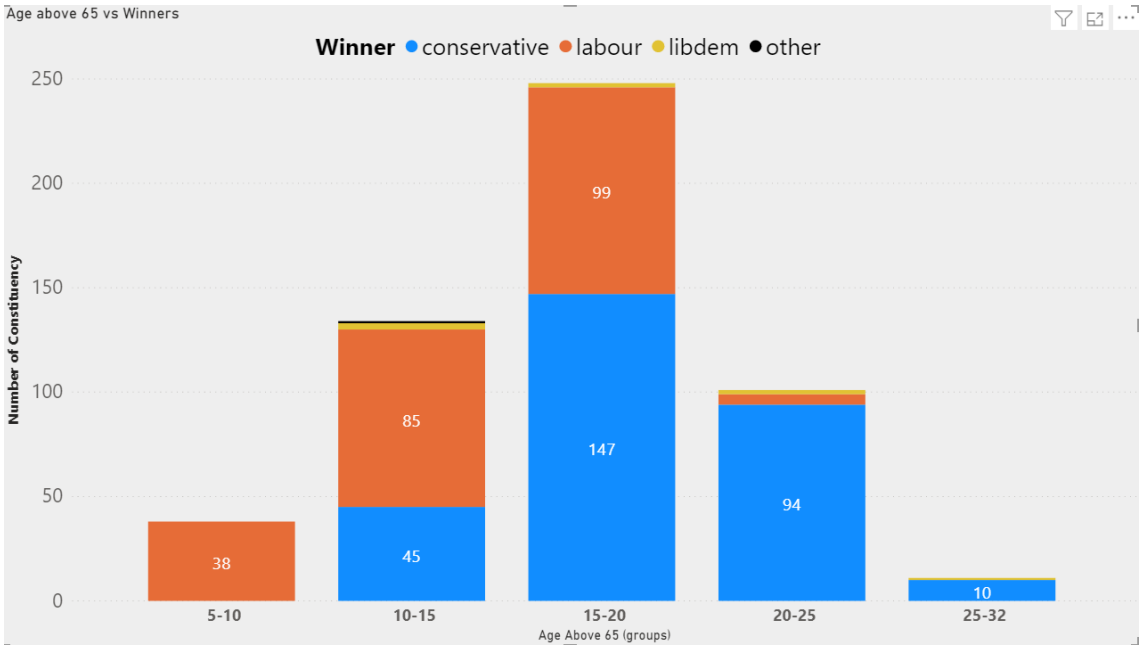
(a) Stacked percentage column chart of seats won by parties in the 2017 England election in constituencies grouped according to the percentage of white voters in the population



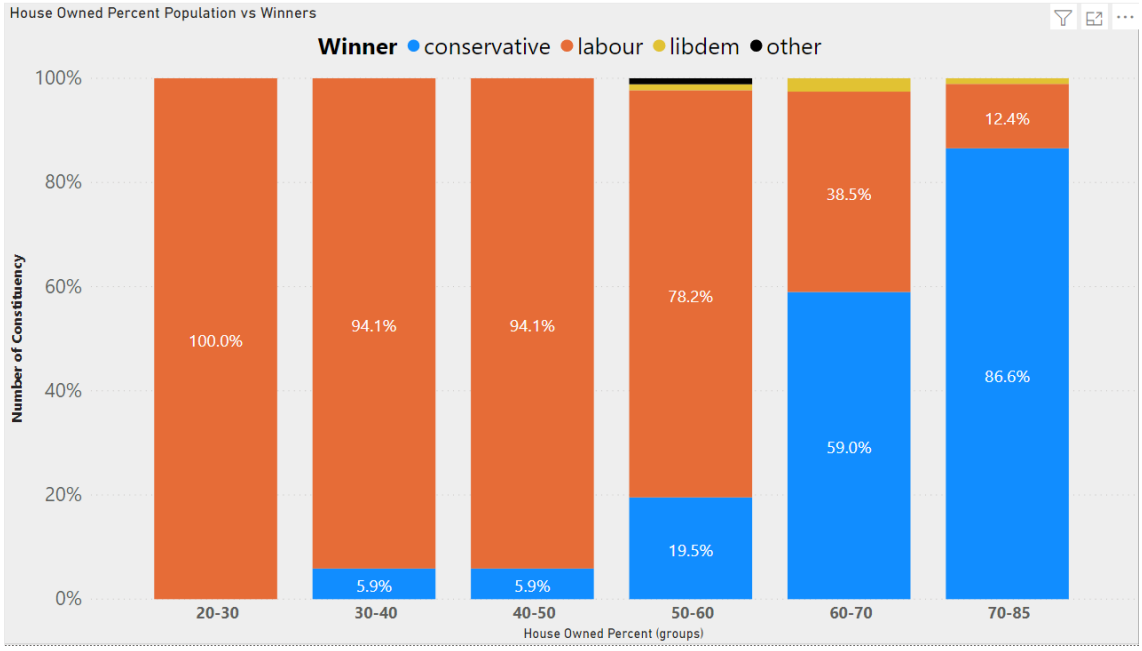
(b) Bar plot of seats won by parties in the 2017 England election in constituencies grouped according to the percentage of white voters in the population



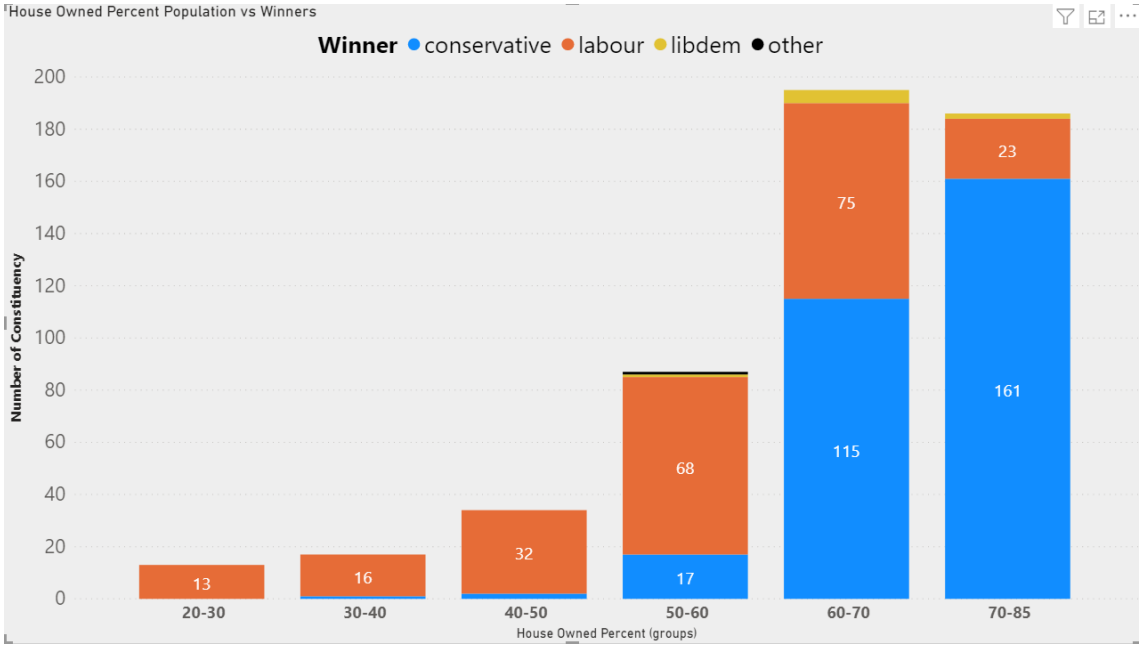
(a) Stacked percentage column chart of seats won by parties in the 2017 England election in constituencies grouped according to the percentage of voters aged 65 and above in the population



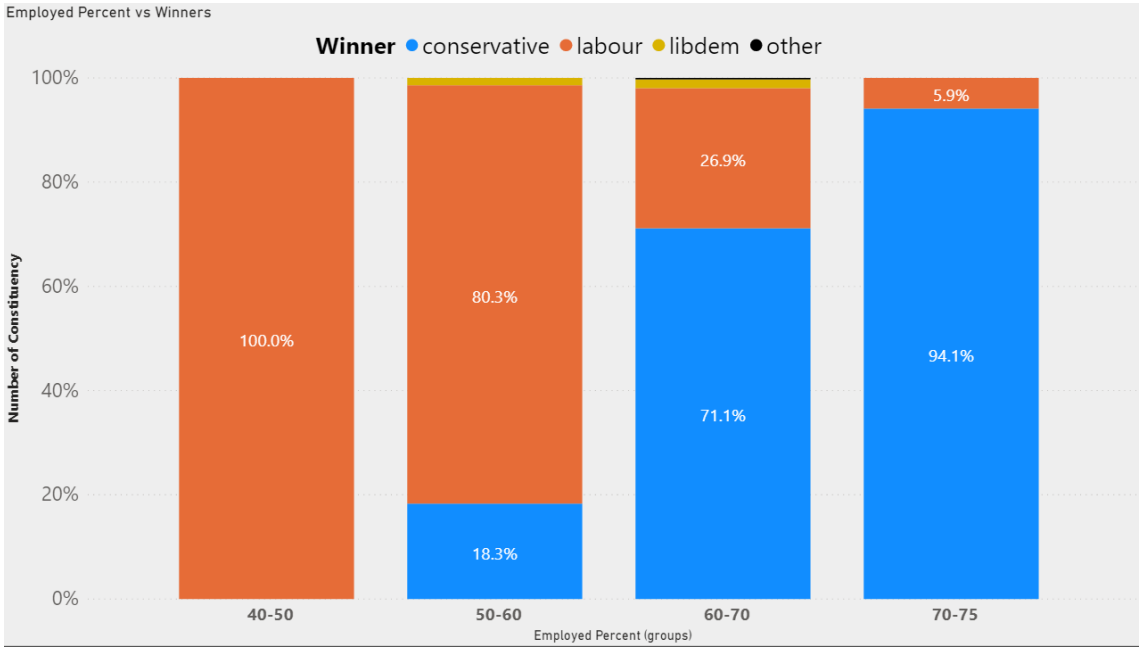
(b) Bar plot of seats won by parties in the 2017 England election in constituencies grouped according to the percentage of voters aged 65 and above in the population



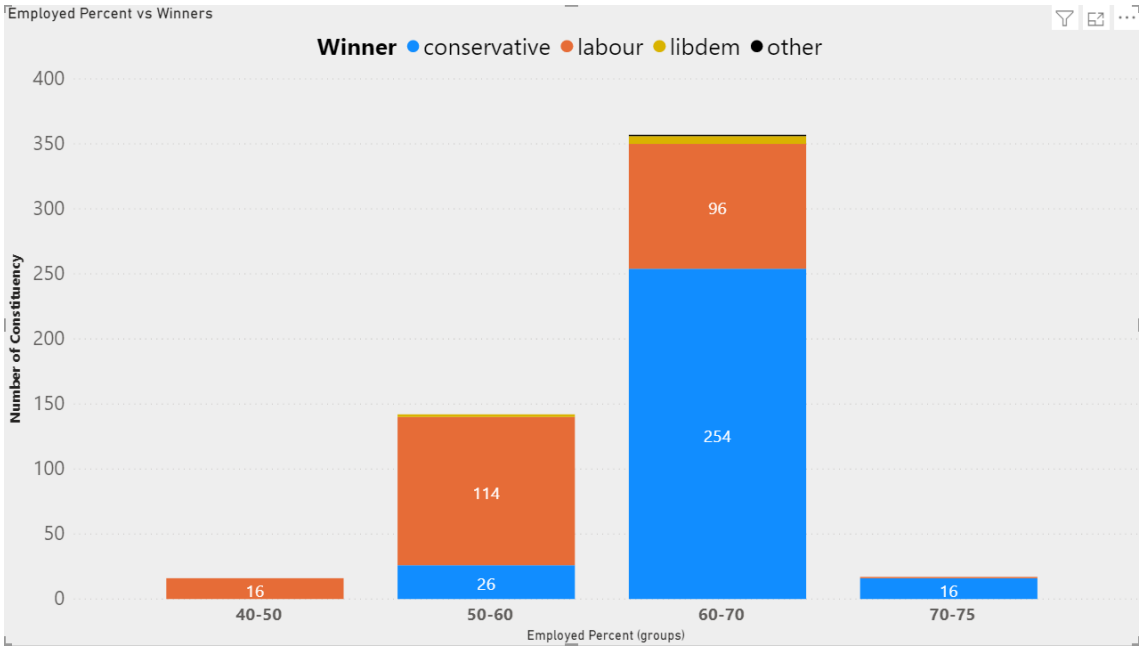
(a) Stacked percentage column chart of seats won by parties in the 2017 England election in constituencies grouped according to the percentage of voters who own their homes in the population



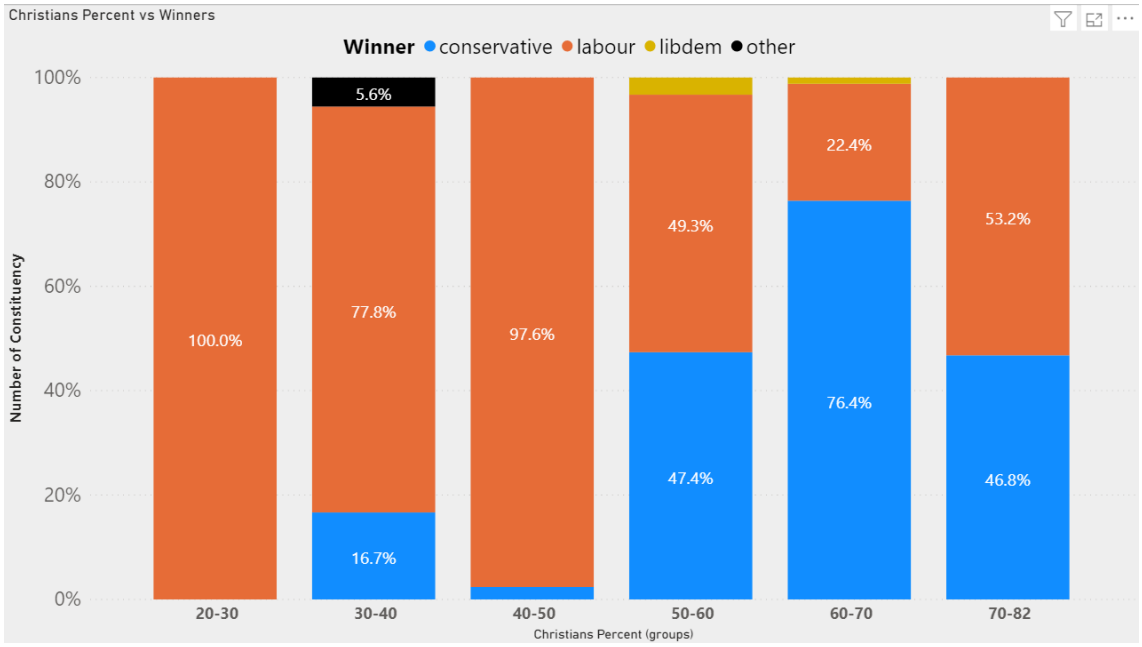
(b) Bar plot of seats won by parties in the 2017 England election in constituencies grouped according to the percentage of voters who own their homes in the population



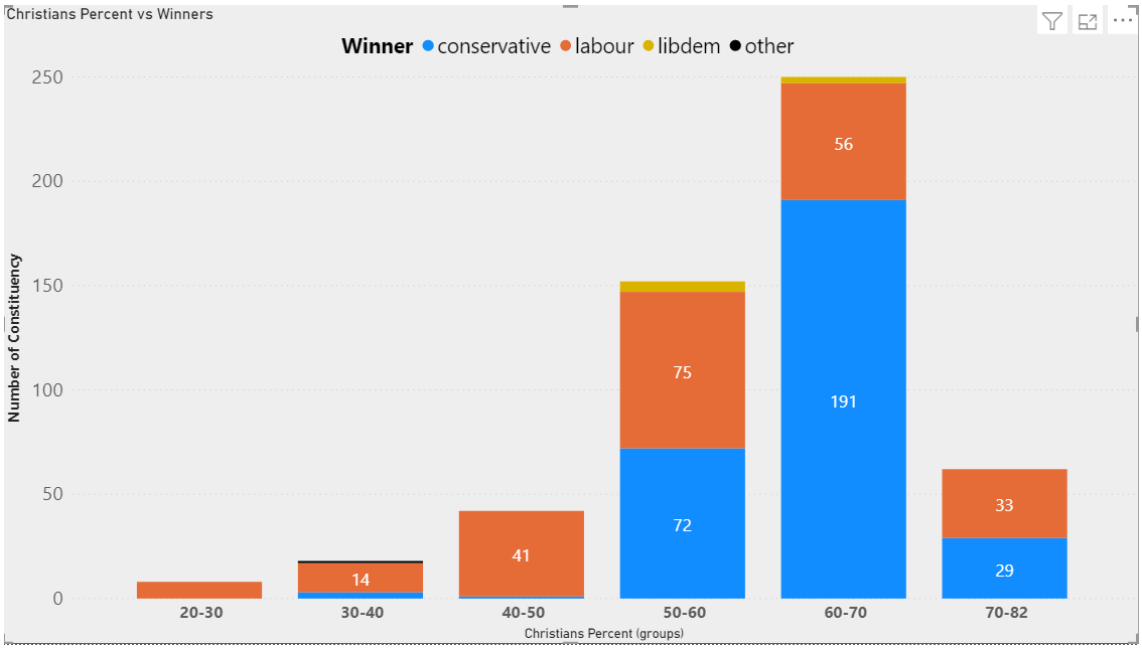
(a) Stacked percentage column chart of seats won by parties in the 2017 England election in constituencies grouped according to the percentage of employed voters in the population



(b) Bar plot of seats won by parties in the 2017 England election in constituencies grouped according to the percentage of employed voters in the population



(a) Stacked percentage column chart of seats won by parties in the 2017 England election in constituencies grouped according to the percentage of Christian voters in the population



(b) Bar plot of seats won by parties in the 2017 England election in constituencies grouped according to the percentage of Christian voters in the population

1.4.3 Graphical Representation of Geographic Data

There are multiple factors that affect the interpretation of data and information, one of which is data visualization. The latter is intended to convey complex information using only a few figures; written information and blocks of data in tables are not part of visuals. Thus, the visuals themselves are very important, and choosing the right type of visual can reduce the cognitive load that goes into interpreting the information. In addition, visual features should be presented in a clear hierarchy to help the user identify the order in which they should look for information. It should be easy to find answers to specific questions and to build a narrative out of an effective visualization, resulting in a positive overall user experience.

Generally speaking, spatial data consists of characteristics attached to physical locations within a geographical coordinate system. These data can also include specific attributes of the locations that could be represented by any type of variable, including nominal (e.g, a group of polygons colored according to the type of soil they represent), ordinal (e.g, the temperature may be categorized as "hot", "warm", "lukewarm", "chilly," or "cold," with a certain order), interval (e.g, pH values), and ratio (e.g, population density) variables.

To visualize this geospatial data for election results, we can use choropleth and hexagonal maps. Choropleths are thematic maps in which each regional unit is shaded according to the value of a particular variable for that region and whose size on the map matches its area. Figure 1.7, is an example of choropleth: we can see that the constituencies in the South East are barely visible on choropleth as they are densely populated regions with small areas. In hexagonal maps, regions are mapped on a hexagonal grid with hexagons of equal size. Therefore, all regions should be visualized clearly in hexagonal maps, but it is harder to relate location and constituencies because of the spatial distortion in them. One aspect is important: people vote, not the land and we review alternatives and benefits of different visuals in section 4.2.

The decline in seats experienced by the Liberal Democrats is clearly visible in both the choropleth and hexagonal maps in Figure 1.7 and Figure 1.8, respectively. However, if we wish to analyze the seat changes in urban and highly populated seats of London, Leeds, Liverpool, etc. then hexagonal maps are a better choice, as these small regions are more clearly visible on a hexagonal grid. The choice of visual depends on the purpose. If the real scale is important in the visualization process, then choropleths should be considered. However, if smaller regions are key, then hexagonal maps are a better option (Langton and Solymosi, 2021). In the last chapter, we will elaborate more upon the type of visual that should be utilized to improve user decision-making.

The hexagonal maps shown in Figure 1.8 facilitate the construction of a narrative concerning England's election results from 2010 to 2019. Firstly, in terms of the Liberal Democrats' fortunes, it is clear that the party has been rapidly losing ground, particularly in the southeast of England. Secondly, the clusters of red indicating Labour wins in the eastern region and London-based constituencies appear to be stable over time; therefore, these constituencies represent safe seats for Labour. Lastly, the maps show that the number of seats won by Labour in the northern regions decreased from 2010 to 2019.

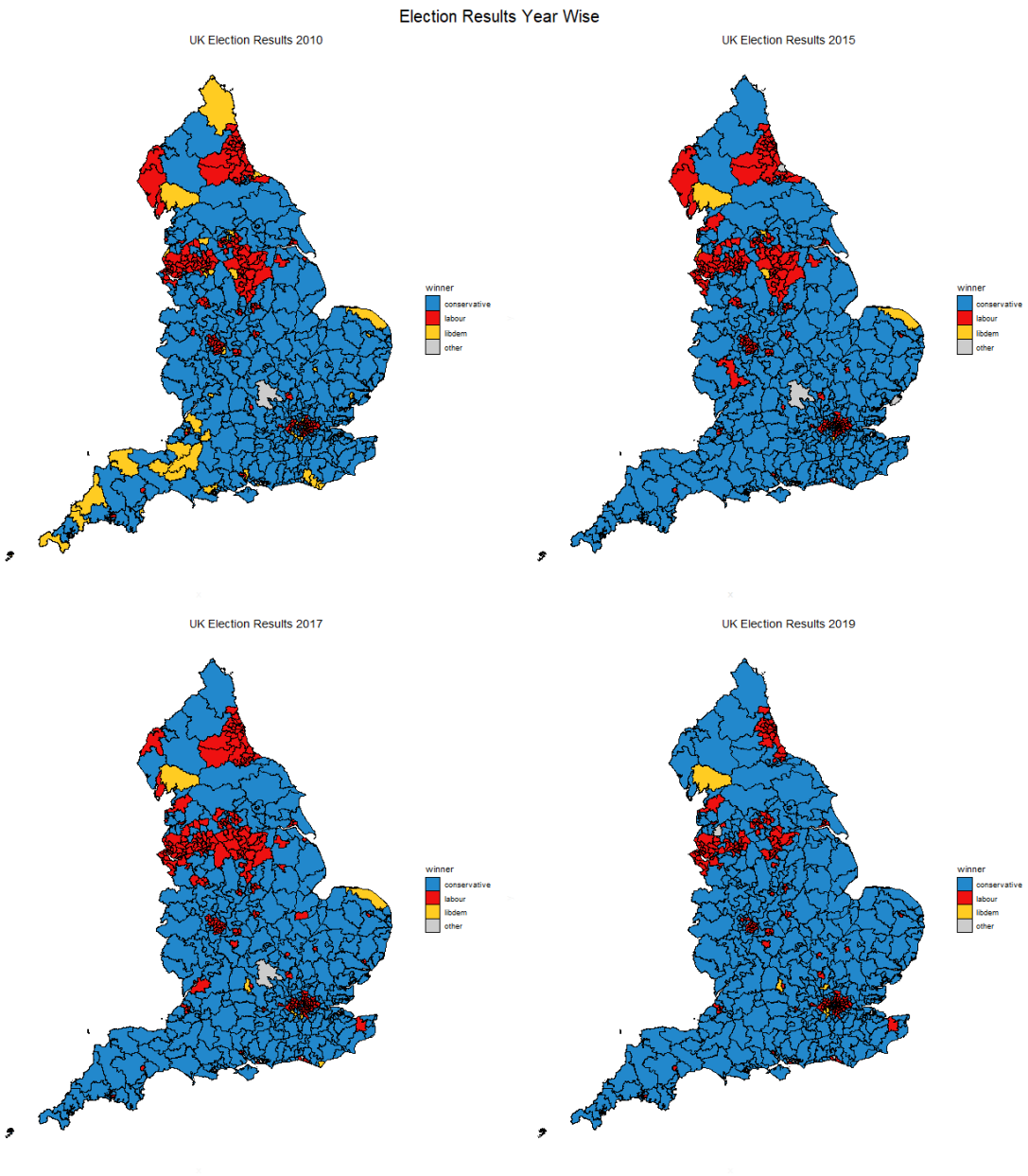


Figure 1.7 – Choropleth maps of the English general elections in 2010 (top left), 2015 (top right), 2017 (bottom left), and 2019 (bottom right)

Figure 1.8 – Hexagonal maps of the English general elections for 2010 (top left), 2015 (top right), 2017 (bottom left), and 2019 (bottom right)

1.5 Literature Review of Election Modelling Techniques

There are various aspects that researchers try to analyze in the context of an election, including voting intentions, turnout, and the election swing effect. We wish to model the UK Election for the 533 English constituencies. The usual modelling approach makes use of weighted pools, previous election results, and other factors, such as incumbencies and the presence of star candidates, including party leaders. This procedure relies heavily on the accuracy and weighting of opinion polls. New research using this framework is focused on better usage of polling data to improve the model's efficiency. The multilevel regression and post-stratification (MRP), uniform swing, and proportional swing models are some of the techniques that have been considered.

The YouGov model uses the MRP strategy for estimating voting intention. In MRP, non-representative samples are adjusted so that they are representative of the population, which can improve the quality of the analyses of opinions and survey responses. This method uses regression modelling to relate individual-level survey responses to various characteristics and then rebuilds the sample to match the population demographics by performing post-stratification (Kastellec et al., 2010). The steps that are involved in building an MRP model are as follows (Kennedy and Gelman., 2020):

- Take samples of survey data and examine key demographic characteristics of the samples
- Develop the post-stratification table and estimate population counts for the combinations of these demographic factors; prepare the table in such a way that it is consistent with the survey dataset
- Identify the quantity of interest and estimate it using multilevel regression modelling based on the demographics features available in the survey dataset.
- In the post-stratification stage, estimate the same quantity of interest in each cell in the post-stratification table and aggregate the results over the cells to determine the population-level estimate

The main advantage of an MRP strategy is that it can adjust the sample to make it more representative of the population (Park et al., 2004). The MRP model is also very useful in estimating state-level public opinion poll results based on small and medium-sized samples (Kastellec et al., 2010). Researchers use MRP when they have formalized a population, identified key variables that are believed to have an influence on the outcome variable of interest, and distinguished multiple differences between the sample and the population-based on these variables (Kennedy and Gelman., 2020). For an MRP strategy to produce valid results, the opinion of interest must be related to the other variables and/or the constituent level; it can fail if it attempts to measure subjective beliefs. As long as we are modelling very predictable beliefs and the over-represented and under-represented characteristics are included in the post-stratification frame, the MRP method performs well, but if these characteristics are not included in the post-stratification frame, the method will perform poorly (Hanretty, 2020).

The YouGov model consists of two steps (Lauderdale and Blumenau, 2019). First, a population model determines how people with different characteristics voted. These characteristics include age, gender, and level of educational qualifications. Second, a survey response model then examines how the individuals with this characteristic responded to a survey on voting intention. Using the combination of these two models, YouGov estimates the voting intentions of all types of individuals and predicts the constituency votes using these estimates.

MRP requires sample sizes, but most polls conducted do not have this level of precision and provide minimal regional and demographic breakdown; we nevertheless combine polls to reduce uncertainty and compare the average voting intention with past results using uniform swings or proportional swings. In the uniform swing model (Canadian Election Watch, 2019b), if a party goes up or down by X percentage points in one riding, then it will go up or down by X percentage points in every riding in that area. In other words, if the Liberal Democrats vote share increases from 13.4% to 26.8% in England, then a uniform swing would increase the Liberal Democrats voting share by 13.4% in ev-

ery riding in England. In the proportional swing model, if a party increases or decreases by fraction X in a region, then it increases or decreases by fraction X in each riding in that region. To illustrate this point, if the Liberal Democrats voting share increases from 13.4% to 26.8%, a proportional swing would double their share in each riding. When no party goes from very low to competitive (or vice versa) in any region, then the type of swing, uniform, or proportional matters very little. However, if some parties double their support, then it matters a great deal. A uniform swing is used in the Canadian Election Watch model (Canadian Election Watch, 2019a), and the proportional approach is used in the 338 model (Fournier, 2018) in Canadian elections.

The uniform swing model performed very well in terms of Canadian elections until 2011. In 2019, the Canadian Election Watch correctly predicted 298 of the 338 final winners. The Liberal party was projected to win 133 seats but received 157 seats, so the model underestimated the party's strength. In the 338 Canada model, proportional swings and demographic data are used to make regional adjustments. For example, if party A appears to have 56% of the votes in a riding and 6% across the province, then if the voting intentions regarding that party double across the province, party A should not be projected to win 102% of the votes in that riding (as in the straight proportional model). Demographic adjustments can include information on language spoken, age, population, employment, and income. All of the above data can be found in the Canadian census. Polls are weighted according to sample size and field date. This model accurately predicted the most likely outcome of the Canadian federal election in 2019 which was an LPC minority. Although the model underestimated the Conservative wins and overestimated the NDP wins; it correctly identified 299 of 338 winners. Hence, it is unclear which is better – the proportional swing or the uniform swing – and both have their disadvantages. From a user perspective, the 338 Canada model is excellent because it takes uncertainty into account, and produces calibrated forecasts, presenting them in a systematic and straightforward manner.

Another interesting election prediction technique, in which polling information is used

to estimate voting intentions, is used by the FiveThirtyEight General Election Forecast (Silver, 2016), which combines a regression model for the state of the economy with polling data. First, polls are collected and weighted based on sample size, date of collection, and pollster ratings. These adjusted polls are then combined with demographic and economic data. The main source of uncertainty in the model is the polling information, which includes national, state, and regional errors. The national errors are systematic; thus, a random number to represent the error is drawn and applied to every state equally. The magnitude of this estimated error is determined by factors such as the time until the election, the number of undecided voters, and the number of third-party voters. To account for regional error, the model simulations assign random variances to demographic groups and regions. Finally, state-specific errors are added, with larger errors assigned to states with smaller populations, as their demographic idiosyncrasies make polling more difficult.

Most of the work mentioned so far involved using polls to estimate election outcomes. However, Bloomberg uses a model that estimates county-level turnout rates using covariates based on demographics and historical information instead of actual election results (Therriault, 2020). The data sources included historical US election information, US census data, and election competitiveness information provided by FiveThirtyEight and Ballotpedia. Using the county's turnout rate as the outcome, Bloomberg employs different algorithms, such as random forest, extra-tree regression, and gradient-boosted decision trees. Finally, an ensemble model averages the algorithms' predictions to achieve the best results. When investigating the model features, it was discovered that historical turnout was the best feature, reflecting voting fidelity and incumbency effects. It was also observed that ridings with high levels of education, high incomes, and high homeownership rates have high voter turnouts, while ridings with more mobile residents have low voter turnouts. Although the response variable is different in our model, this model can help guide us in model assessment in terms of the impacts of features and which results are sensical and which are not.

Contrary to the MRP, uniform swing, and proportional swing strategies which pool the latest polls to predict election outcomes our model does not; we nevertheless show how such information could be incorporated using an offset term. But the model above does point out that demographic information is very important in modelling various aspects of elections, such as results and turnout rates and our exploratory data analysis indicated that the demographic data for the UK can be very helpful in predicting voting patterns. As such we plan to develop a baseline model incorporating demographic covariates and historical election data.

We also found that the 338 Canada model incorporates uncertainty adequately. Uncertainty plays a key role in the dynamic decision-making of a user regarding election outcomes. Point estimates of the election results are important, but they do not provide a range of plausible values that will help users understand how certain the results are. Therefore, users who make use of election forecasts often fail to understand what outcomes are plausible (Silver, 2016). Therefore, this thesis aims to adequately capture the uncertainly using a Bayesian model and communicate it in a manner that will assist in decision making. For example, if our projections predict that party A will win an average of 170 seats, then the uncertainty of this outcome will be made clearer by providing a range of seats and the likeliness of different scenarios.

In the last chapter, we will consider a simple scenario in which a term referred to as an offset is added to our multinomial logistic regression model to explain the changes in voting intentions compared to a baseline year. Instead of using a uniform or proportional swing, we will use this offset to swing the results. A swingometer can be constructed to measure the effect of changes in voting intentions (Wells, 2020). Once we have developed a baseline model, we will propose a mechanism and framework for visualizing spatial data and identify key considerations that should be kept in mind when displaying uncertainty for this type of data.

Chapter 2

Modelling Techniques

2.1 Probabilistic Models for Data

We consider first elements appearing in likelihood-based inference and Bayesian statistics. Let y denote the vector of observations and assume each data point y_i originates from a common distribution with density or mass function $f(y_i;\theta)$. Assuming that each observation is independent and identically distributed, the likelihood function $L(\theta \mid y) = \prod_{i=1}^{n} f(y_i \mid \theta)$ is formed by taking the product of individual contributions and is a function of the model parameter θ. The likelihood allows us to calculate how likely a sample is under the postulated distribution for a given value of the parameter θ of the model. For example, if we have n independent Bernoulli trials (each outcome is a head or a tail), and the parameter θ denotes the probability of getting a head, then the likelihood for k successes out of n trials is $L(\theta \mid k,n) = f(k \mid n,\theta) = \binom{n}{k}\theta^k(1-\theta)^{(n-k)}$.

In, frequentist statistics, the point estimates of parameters are typically found using maximum likelihood estimation. The maximum likelihood estimate is the point in the parameter space that maximizes the likelihood function, and it can be obtained by finding the root of the derivative of the likelihood function with respect to the parameter. Thus, for the above-mentioned example involving n independent Bernoulli trials, the maximum likelihood estimate of parameter θ is k/n.

The Bayesian approach to probability models involves a synthesis of evidence (likelihood) and personal or expert judgments (prior) regarding the data which are combined to obtain the posterior distribution using the Bayes' rule. The probability distribution through which one's beliefs about a certain parameter are expressed before taking some evidence into consideration is the prior and denoted by $p(\theta)$. It can be obtained from previous research or kept purposefully vague and weakly informative (Cohen, 2020). Choosing the prior technically involves choosing a family of probability distributions, then choosing a set of hyperparameter values that govern that probability distribution. For example, for coin flips, the beta distribution can provide a continuous distribution function for the probability of getting a head θ parametrized by hyperparameters α and β.

In Bayesian inference, the posterior distribution is obtained after updating our prior beliefs encoded in the prior $p(\theta)$ with the help of information we retrieved from the likelihood function $p(\mathbf{y} \mid \theta)$. If we want to find a point estimate of a parameter that incorporates information from a prior distribution then we use maximum a posteriori estimation. The point estimate can again be computed using numerical optimization methods. Thus, returning to the our example, if we use a beta distribution with hyperparameters α and β as our prior distribution, and our likelihood function is the binomial distribution for n independent Bernoulli trials with k success, then the maximum a posteriori estimate of parameter θ is the mode $(k+\alpha-1)/(n+\alpha+\beta-2)$. To obtain the final distribution for our parameters given that we observed the data, i.e., $p(\theta \mid \mathbf{y})$ we use Bayes' Theorem $p(\theta \mid \mathbf{y}) \propto p(\theta) \times p(\mathbf{y} \mid \theta)$. The proportionality concept implies that to assign a measure of one to the entire space and calculate the posterior probabilities, it is necessary to multiply or divide it by a normalization constant so that it integrates to one. The normalizing constant requires integrating the likelihood against the prior, but it is difficult to estimate that in complex models.

Bayesian Inference - Coin flip example

In a coin flip, we generally assume that the probability of landing a head is 0.5, which can be considered a prior belief. For Bayesian inference, we can encode this prior belief in $p(\theta)$ using a beta distribution $\mathsf{Beta}(k+\alpha, n-k+\beta)$, see Figure 2.1. The next step is to specify a probabilistic model for the data. We use the binomial distribution $\mathsf{Bin}(k:n,\theta)$, where θ is the probability of getting a head, and k is the number of heads in n trials, for the outcome; its mass function is

$$f(y \mid \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k} \tag{2.1}$$

If we combine the prior and the likelihood functions, the resulting posterior distribution is $\mathsf{Beta}(k+\alpha, n-k+\beta)$. The plot on the left in Figure 2.1 shows the likelihood, prior, and posterior distribution for one coin toss (n = 1) resulting in a head. The posterior density is shifted to the left relative to the likelihood function because the prior assigns mass everywhere, but the change is only marginal as we have only one coin toss. We can assess the impact of change in sample size on the posterior distribution with the same $\mathsf{Beta}(\alpha = 2, \beta = 2)$ prior. The plot on the right in Figure 2.1 shows the likelihood, prior, and posterior distribution for 27 coin tosses resulting in 18 heads, which clearly shows how the posterior distribution shifts towards the likelihood due to an increase in sample size. In this example, we calculated the posterior distribution analytically, but, for even slightly more complicated models, it is impossible to compute the posterior in closed form. Thus, we turn to the Markov chain Monte Carlo algorithm to approximate the posterior distribution.

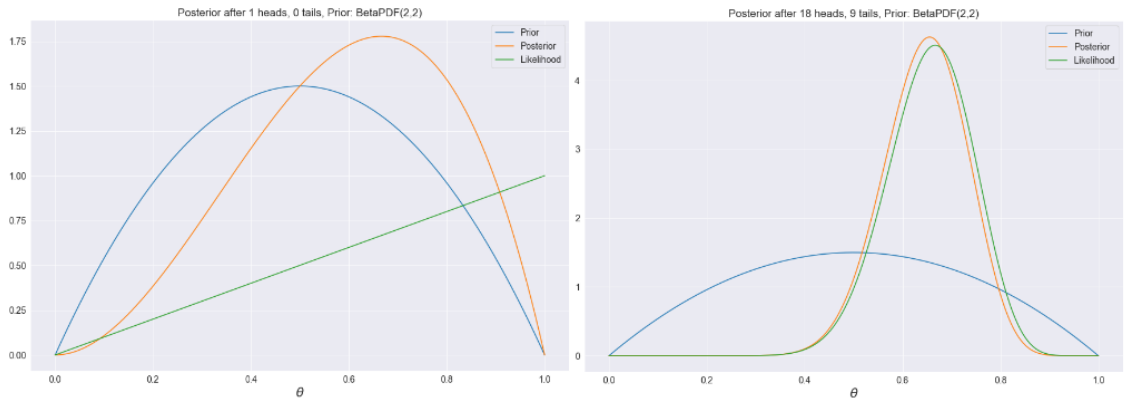


Figure 2.1 – Left: Prior, likelihood and posterior distribution for the probability of heads in one coin flip with one head and zero tails, Right: Prior, likelihood and posterior distribution for the probability of heads in 27 coin flips featuring 18 heads and nine tails

Markov Chain Monte Carlo

Oftentimes, one is interested in estimating a function of the unknown parameters of a distribution like the expectation or variance of a random variable. While numerical integration methods are competitive in low-dimensional problems, they are not well-suited to handle high-dimensional problems. An alternative is the use of Monte Carlo methods, which are a class of algorithms by which one draws samples from a distribution of interest and approximates the functional by taking of empirical averages of random samples drawn from that distribution; the law of large number guarantees the validity of this construction if the number of draws is large. Since designing a sampling algorithm for a complicated distribution is a hard task, Bayesian analysis often proceeds with Markov chain Monte Carlo (MCMC) algorithms. These are used to construct a Markov chain whose stationary (i.e., limiting distribution if the chain runs forever or long enough) is the posterior distribution of interest. Specifically, starting from an initial state θ^0, one defines a sequence of random vectors $\theta^1, \theta^2, \theta^3, \ldots$ which, if the algorithm is properly designed, will eventually approximate draws from the target distribution (Christensen et al., 2011). However, compared to the Monte Carlo method (which produces typically independent

draws), the samples provided by the Markov chain are auto-correlated.

In practice, it is difficult to assess whether a Markov chain has reached its stationary distribution. Theoretically, this requirement is that of geometric ergodicity (Roberts et al., 2004), but proving the latter is often tedious. One way to assess whether the chain is stable is to run multiple Markov chains, starting from different initial values, and compare the between-variance of the chains to the within-variance. A split $\widehat{R}$ (Hoffman and Gelman, 2014) statistic is used instead to assess the variation within a sequence of Markov chains which can is defined as (Vats and Knudson, 2018)

$$\hat{R} = \sqrt{\frac{\text{chain length} - 1}{\text{chain length}} + \frac{\text{between-chain variance}}{\text{within-chain variance}}} \tag{2.2}$$

We should suspect that geometric ergodicity is not being fulfilled when $\widehat{R}$ is not near the nominal value of one.

There are various algorithms that can be used to design appropriate transitions of Markov chains for generic distribution, like the Gibbs Sampler and the Metropolis–Hastings algorithm (Robert and Casella, 1999). Gibbs sampling is a method for constructing a Markov chain that is extremely useful when it is possible to identify the conditional distribution of each parameter. As part of this process, samples are obtained from each conditional distribution of each parameter given the others, in turn. In order to use the Gibbs sampler, we must presume that we know how to sample from full conditional distributions. The more general Metropolis–Hastings algorithm can be implemented to simulate the entire vector θ with every iteration of the algorithm: it is a type of accept-reject algorithm that relies on a candidate generating distribution, also known as a proposal distribution. Rather generic, the efficiency of the sampler depends largely on the proposal: good proposals ensure that the chains have low autocorrelation and explore efficiently the target once it is reached. More complex algorithms, use the gradient of the log-posterior to create proposals. One such algorithm is Hamilton Monte Carlo (HMC) method, which was primarily formulated in the late 1980s under the name Hybrid Monte Carlo to handle computations in lattice quantum chromodynamics (Duane et al., 1987). The method builds on a rich theoretical foundation that suits it especially well to problems with high

dimensions (Betancourt et al., 2014). STAN provides an implementation of the Hamiltonian Monte Carlo algorithm with no U-turn sampler and the probabilistic programming language allows one to easily define hierarchical models and draw samples from them.

2.2 STAN - Probabilistic Programming Language

To perform Bayesian inference using Markov chain Monte Carlo algorithms we can use STAN, which is a probabilistic programming language. Technically, STAN is a free and open-source C++ program that performs Bayesian inference or optimization for user-specified models and can be called with a command line in R, Python, MATLAB, and Julia. It has great promise in terms of fitting large and complex statistical models in many application areas (Gelman et al., 2015). The iterative process used in Bayesian modeling can be described in the following steps: (Farrell, 2020)

1. Model design
2. Prior selection
3. Sampling from the posterior distribution
4. Model convergence diagnostics

The first two steps in the above procedure are dealt with in the STAN code. We will outline the model specifications by presenting various blocks of STAN code.

Data In this block, the dimensions of the data and the variable names and types are specified.

Parameters The parameter block is used to assign names to the parameters while keeping in mind any mandatory restrictions and dimensions. Examples of parameters are the intercept and slope in a standard linear regression model.

Transformed Parameters This block is used for various reparameterization techniques and multiparameter constraints to help with identifiability in complex models.

Model This is the block in which the sampling statements for the likelihood function, priors, and hyperpriors are defined (a hyperprior is a prior distribution for the parameters found in the priors of the likelihood).

Generated Quantities We can simulate quantities to be predicted by the likelihood function using parameter draws. Other libraries, such as `rstanarm`, have functions that help run the generated quantities separately from the sampling exercise.

We then use the fit command to run our model after the latter is compiled in C++. The important information passed along to Stan fit code includes the data, the number of iterations for warm-up (these iterations will not be used for the posterior distribution later), the total number of iterations, how many chains to run, and the number of cores to be used (Gelman et al., 1992).

Convergence of Markov chains is diagnosed with the $\hat{R}$ value, the effective sample size, and the traceplots of the fitted parameters. We usually only need to worry if the effective sample size number is less than 1/100th or 1/1000th of the number of iterations (Farrell, 2020). Further, the $\hat{R}$ value near nominal value near 1 indicates convergence.

Basic Linear Model in STAN

The process above can be outlined via an example of a linear model. Consider the following data from the National Snow and Ice Data Center detailing the decline in sea ice extending over time in Table 2.1.
Our linear model equation is

$$y_n = \alpha + \beta x_n + \varepsilon_n, \quad \text{where} \quad \varepsilon_n \sim \mathsf{No}(0, \sigma).$$

Here, the response variable y is `extent_north`, x represents the year, β is the slope, α is the intercept, and ε are independent mean-zero error term for n years. Fitting the linear model using the least squares method gives estimates of the parameters in Table 2.2

	year	extent_north	extent_south
1	1979	12.33	11.70
2	1980	12.34	11.23
3	1981	12.13	11.44
4	1982	12.45	11.64
5	1983	12.33	11.39
6	1984	11.91	11.45

Table 2.1 – Data from the National Snow and Ice dataset

	Estimate	Std. Error	t value
Intercept	12.56	0.07	174.42
year	−0.05	0.04	−17.40

Table 2.2 – Summary statistics for the frequentist linear model for the decline in sea ice

To prepare the STAN code for this linear model, the three main blocks needed are the data, parameter, and model blocks which are specified in Listing 4. Flat priors are specified for this example. To fit the model using the STAN sampling function, we use 500 warm-up and 500 sampling iterations, and four Markov chains. After sampling from the Bayesian model, we can obtain estimates of mean and note that the intercepts and slopes of the modelling techniques are similar, as shown in Table 2.2 and Table 2.3.

	mean	sd	2.5%	97.5%	$\widehat{R}$
Intercept (α)	12.56	0.07	12.42	12.70	1.00
year (β)	−0.05	0.01	−0.06	−0.05	1.00
Std dev (σ)	0.23	0.03	0.18	0.29	1.00

Table 2.3 – Summary statistics for the Bayesian linear model for the decline in sea ice

The multiple estimates of the posterior draws are shown in Figure 2.2. This plot provides an understanding of the standard deviation of the draws captured during the sampling process. In Table 2.3, the $\hat{R}$ values are all 1.0, and in Figure 2.3 the chains shown in the traceplot are stationary, which indicates that the model has converged.

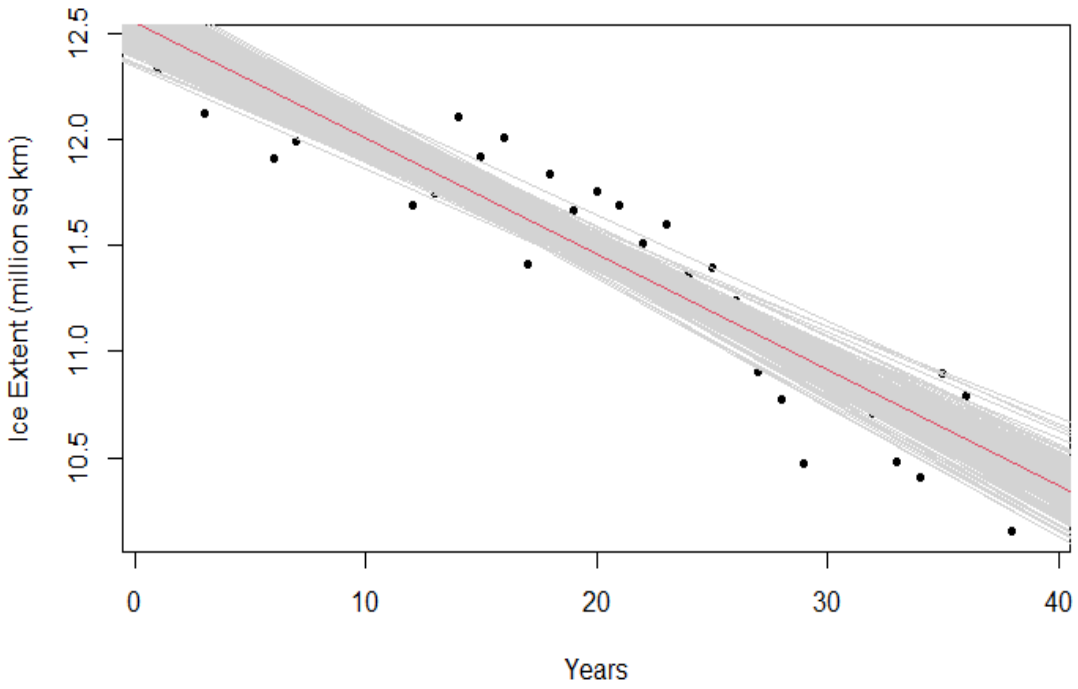


Figure 2.2 – Posterior draws line plot for change in sea ice extent in the Northern Hemisphere from year 1979 to 2017

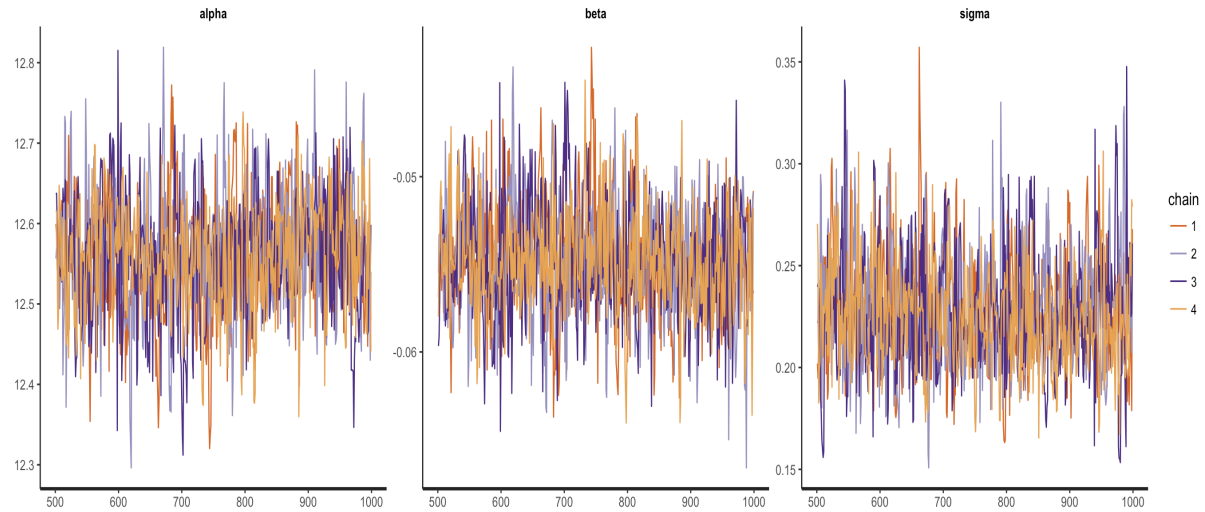


Figure 2.3 – Trace plot of the parameters for all four chains for the Bayesian linear model for the decline in sea ice

2.3 Regression Models for Multivariate Categorical Outcomes

The example used above involved a classical linear regression model, but election results are categorical responses. A lot of research has been done on the binary outcome variable for a two-party election. However, modelling multivariate outcomes is a bigger challenge. The multinomial response Y_{ri} for the r-th observation belonging to i-th category out of K categories for given total trails n_r can be modelled using a multinomial distribution (Rodríguez, 2007):

$$\Pr\{Y_{r1} = y_{r1}, \ldots, Y_{rK} = y_{rK}\} = \begin{pmatrix} n_r \\ y_{r1}, \ldots, y_{rK} \end{pmatrix} \theta_{r1}^{y_{r1}} \cdots \theta_{rK}^{y_{rK}}, \tag{2.3}$$

where, for individual data points, $n_r = \sum_i y_{ri} = 1$. However, for grouped data, $n_r = \sum_i y_{ri}$ denotes the total number of cases of the i-th observation. The log-odds η_{ri} can be mapped to the probability scale using

$$\theta_{ri} = \text{softmax}(\eta_{ri}) = \frac{\exp\{\eta_{ri}\}}{\sum_{i=1}^{K} \exp\{\eta_{ri}\}}, \tag{2.4}$$

and

$$\eta_{ri} = \alpha_i + x_r' \beta_i \tag{2.5}$$

where x_r is the vector of covariates associated with the r-th observation, α_i is a constant, and β_i is a vector of regression coefficients for $i = 1, 2, \ldots, K-1$.

In STAN, multi-logit regression can be performed for multiple outcomes for individual data points. For example, if we have R data points with K possible outcomes, one of which is Y, then the STAN code would be(Stan Development Team, 2020):

```
data {
  int K; // number of categories
  int R; // number of observations
  int D; // number of covariates
```

```
  int y[R]; // Outcome variable out of K categories
  matrix[R, D] x; // matrix of covariates
}
parameters {
  matrix[D, K] beta; ;  // regression coefficients
}
model {
  matrix[R, K] x_beta = x * beta;
  to_vector(beta) ~ normal(0, 5);

  for (r in 1:R)
    y[r] ~ categorical_logit(x_beta[r]');
}
```

In this example, the likelihood function is the categorical-logit sampling function and appears in the model block. In the context of this paper, Y is the winner out of K parties in R constituencies, and D is the number of covariates making up the individual response data. Similarly, if the response were to be the number of votes received by each party, the grouped data case arises, and the multinomial sampling statement shown below would be utilized.

```
model {
    for (r in 1:R){
    y[r] ~ multinomial_logit(x_beta[r]');
    }
}
```

Lastly, if the outcome Y is the proportion of votes per party, with these proportions summing to 1, then another likelihood function such as the Dirichlet distribution, could

be used as the sampling statement in the model block (Sennhenn-Reulen, 2018).

Nonidentifiability

Nonidentification in inference arises due to erroneous model specifications. It occurs when two sets of quantities have the same probability outcomes for different sets of parameters (Rannala, 2002). For example, if a constant is added to all the log-odds η_{ri}, then the same probability outcomes as before are generated. Generally, the model is identified with the help of suitable priors for convergence, but another way to avoid nonidentifiability is to add constraints to the parameters. One possibility is to use only $K-1$ parameters when modelling K categories. Thus, the last parameter would be excluded from the model and set to zero using $\eta_{rK} = 0$. This procedure can be carried out using the transformed parameters block. The corresponding STAN blocks would be (Stan Development Team, 2020):

```
parameters {
  matrix[K - 1, D] beta_raw;
}


transformed parameters {
  matrix[K, D] beta;
  beta = append_row(beta_raw, zeros);
}
```

Another way to avoid nonidentifiability is to impose the sum-to-zero constraint on the parameters using $\sum_{i=1}^{K} \eta_{ri} = 0$. This constraint avoids making one of the alternatives stand out, as in Feit (2020):

```
parameters {
    vector[K-1] beta_raw;
```

```
}
transformed parameters{
  vector[K] beta;
  beta = append_row(-sum(beta_raw), beta_raw);
}
```

The details provided above concerning Bayesian modelling and STAN are sufficient to move ahead with the experiment in order to evaluate the models' performance with different sets of covariates. The objective is to build a model that models election outcomes/votes in such a manner that an offset-based swingometer can be created (Wells, 2020) based on a Bayesian approach.

Chapter 3

Statistical Models for Elections

The main objective of this chapter is to build a multinomial logistic regression model that can capture the posterior distribution of election outcomes accurately. This model will be based on demographic information and the election results for England for the past four elections. Thus, the posterior distribution of the outcomes should be able to capture the uncertainty in these outcomes. The final step is to add an offset to this model so that it can predict new outcomes based on changes in voting intention relative to the baseline election year. Starting with the basic modeling techniques outlined in Chapter 2, we will build models with increasing complexity. In summary diagnostics of each model will help us to assess how well the model fits the data, and the accuracy of each model will be measured by comparing the outcomes predicted by the model and the original election outcomes. Once we are able to fit a complex model that can accurately model the outcomes, we will examine the sensitivity of the prior by checking whether the posterior distribution changes with changes in the prior.

3.1 Logistic Model for Categorical Response

Logistic regression is used for binary dependent variables. In this type of regression, the log odds of the dependent variable are modeled as a linear combination of the predictor variables. The odds ratio is the probability of event A divided by the probability of the

complement of event A, and the log odds ratio is the logarithm of this ratio.

However, the multiparty system of election results that we are examining yield categorical rather than binary outcomes. Thus, the categorical logit model must be used instead of logistic regression. This model is intuitively similar to logistic regression, usually with a reference category for identifiability. With the exception of the reference category, the odds ratios must be calculated for each category versus the reference category at different factor levels (Littler, 2018).

Further, a link function transforms the probabilities of the levels of a categorical response variable to a continuous scale for modelling purposes (Rodríguez, 2007). The softmax inverse link function will be used to accomplish this. In essence, it turns a vector of K real values into a vector of K real values that sum to one(Wood, 2020).

3.1.1 Modelling Approach

In this first model, the effect of time and repeated outcomes will be ignored and we will focus on the 2017 election data. We consider a generalized linear mixed model for Y_{ri}, the categorical response indicating a win for party i, $i \in \{1,\ldots,K\}$, in riding r in 2017, such that Y_{ri} takes the value 1 if party i wins riding r and 0 otherwise . The response is modelled using a multinomial distribution (eq. 2.3), where $n_r = \sum_i y_{ri} = 1$ since only one party i among K parties can win riding r. Further, θ_{ri} denotes the probability of party i winning riding r. We map the linear predictor η_{ri} to the probability scale using the softmax inverse link function (eq. 2.4),

$$\eta_{ri} = \alpha_i + \mathbf{x}_r \beta_i,$$

so that α_i is the intercept, $\mathbf{x}_r$ is a row vector of dummies for regions, and β_i represents the coefficients associated to $\mathbf{x}_r$. We specify vague normal priors for β_i, α_i with five as standard deviation such that $\beta_i \sim \mathsf{No}\left(0, 5^2\right)$, and $\alpha_i \sim \mathsf{No}\left(0, 5^2\right)$

To implement the model, we use the STAN code shown in Listing 1, which is divided into appropriate sections concerning the data, parameters, model, and generated quantities needed to quantify the posterior distribution of the outcomes. We ran four Markov chains

with 2000 iterations per chain. One way to measure the efficiency is via the effective sample size (ESS): An estimator of the population mean based on mB (m chains and B iterations) correlated draws has the same variance as the estimator could have obtained by sampling equivalent n_{eff} independent draws (Leinster, 2014). The effective sample size for all the covariates is above 5000, which is more than enough for the estimation of the parameters.

3.1.2 Region-wise Model Results

Table 3.1 gives the percentage of seats won per party by region for the 8000 posterior predicted elections results. These simulated results unsurprisingly represent the actual results of the 2017 Election data at the regional level.

When examining the results of the 8000 posterior predictions of the present model at the constituency level, we observe some mismatches between the predicted outcomes and the actual outcomes. For example, in 2017, in the riding of Ashfield, Labour won by a margin of 440 votes. Since it was a marginal victory, Labour should have roughly 4000 wins out of 8000 predictions for Ashfield in our posterior prediction. However, our posterior draws for Ashfield show only 34 wins for Labour out of 8000 elections, which does not make sense. This issue arose due to modelling by region only and it can arguably be fixed by adding other covariates.

3.1.3 Graphical Analysis of the Posterior Predictive Draws

From Figure 3.1 we can infer that the distribution of the number of seats won for both Conservative and Labour is symmetric and they have a similar standard deviation and interquartile range. In Figure 3.1, we observe that the posterior distribution is right-skewed towards zero. The actual number of seats won by Other Parties in England in the 2017 UK elections was also one. The posterior predictive distribution for the Liberal Democrats

	Region	Conservative	LibDem	Labour	Other
1	East Midlands	66.8	32.6	0.4	0.1
2	Eastern	86.0	12.2	1.5	0.1
3	London	28.9	67.06	3.9	0.1
4	North East	10.3	89.3	0.3	0.1
5	North West	26.7	71.9	1.2	0.01
6	South East	86.6	9.9	2.3	1.0
7	South West	85.5	12.7	1.5	0.1
8	West Midlands	59.2	40.5	0.1	0.1
9	Yorkshire and The Humber	31.4	68.2	0.2	0.1

Table 3.1 – Predicted percentages of seats per party in 8000 posterior draws categorized by region

shown in Figure 3.1 is centered at a mean of eight seats, which is the actual seats they won in England during the 2017 UK Elections.

Overall, the model does not have extreme outliers in the posterior distributions of any of the parties. For example, the Conservative winning 320 seats and Liberal Democrats winning 20 seats is not an implausible scenario. Note also that the means of the posterior distributions of seats won for the parties are representative of the real election results for 2017. However, this model does not include demographic covariates, which are essential for building a complex and accurate model capable of producing offset-based predictions.

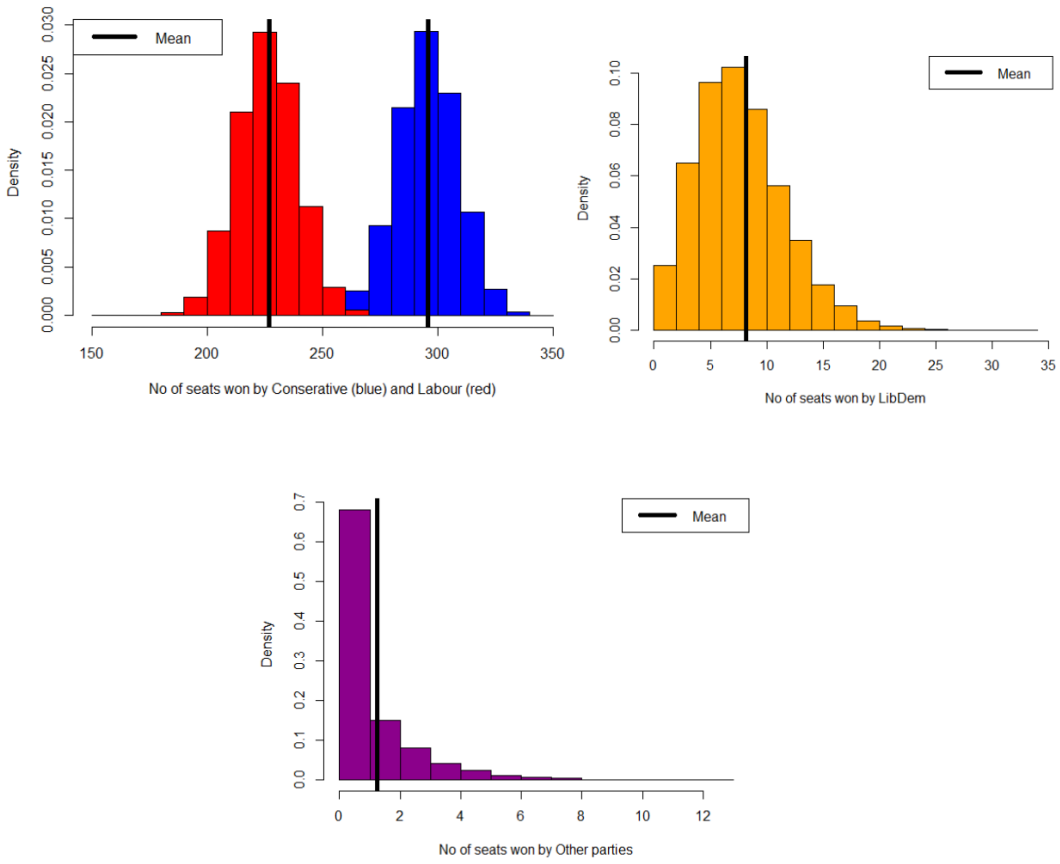


Figure 3.1 – Histogram of the posterior predictive distributions of the number of seats won by Conservative and Labour (top left), LibDem (top right) and Other parties (bottom) based on 8000 posterior predictive draws

3.2 Multinomial - Demographic Model

We can increase the model complexity by adding the demographic covariates shown in Table 1.1. Limiting the model to just demographics allows us the examine the accuracy of the predictions in the absence of the regional/constituency-specific random effects. Further, it allows us to examine other aspects of the model. For instance, it turns out that using the number of seats as the target variable is not effective, as it hides the margin of victory. For example, in the 2017 election, the constituency of Barnsley Central was a safe seat for the Labour party, as the margin of victory there was 15,546 votes, but in Ashfield, the Labour had a slim margin of victory of 440 votes. Therefore, a better target variable would be the vote share per party. Thus, we will change our target variable from the number of seats per party to the number of votes per party. The election results are available at the constituency level for each election year, but in this model, we remain focused on the 2017 election data. We will once more assess how accurately the model predicts the election outcomes compared to the real election results of 2017.

3.2.1 Modelling Approach

We consider a generalized linear mixed model for Y_{ri}, the total number of ballots cast for party i, $i \in \{1,\ldots,K\}$, in riding r in 2017. The response is modelled again using a multinomial distribution (2.3), where $n_r = \sum_i y_{ri}$ now denotes the total votes cast in the riding r, and θ_{ri} denotes the probability of party i winning riding r.

$$\theta_{ri} = \text{softmax}(\eta_{ri}) = \text{softmax}(\alpha_i + \mathbf{x}_r \beta_i) \tag{3.1}$$

where α_i is the intercept for party i, $\mathbf{x}_r$ is a row vector of time-invariant riding-specific demographic information from the census, and β_i represents the associated coefficient. The demographics data are all in percentages. Thus, we standardize the explanatory variables to have mean zero, unit variance, to facilitate the prior specification. To address unidentifiability, we add the constraint $\eta_{rK} = 0$. We again assign normal priors to the parameters β_i and α_i with $\beta_i \sim \text{No}\left(0, 5^2\right)$, and $\alpha_i \sim \text{No}\left(0, 5^2\right)$

In Listing 2, the first block of data contains the input data information needed for modelling. In STAN, the stand-alone generated quantities block can be used after model sampling to simulate new observations from the posterior distribution. For this type of operation, R would introduce complications; thus, we use the **cmdstanr** library, which allows simulations of generated quantities to be conducted separately.

3.2.2 Graphical Analysis of Posterior Predictive Draws

The variability in outcomes captured by the posterior distribution of the model shown in Figure 3.2 is very low compared to the posterior distribution of the model in Section 3.1. This difference is primarily since in the categorical model, the response variable was the seat outcome per riding, whereas, in the multinomial model, the response variable was the number of votes per party per riding. Moreover, the results here indicate that the Liberal Democrats and other parties have not won a single seat in the 8000 posterior draws. Three factors are responsible for this, namely, the model includes only one election year, no regional or constituency-specific indicators were used, and the Liberal Democratic margin of victory was relatively small in the 2017 election. In the elections held in England in 2017, the Conservatives won 297 seats, and Labour won 227 seats, whereas, in the posterior predictive draws, the Conservatives win an average of 312 seats and Labour win an average of 220 seats. Thus, a total of 72 ridings were predicted incorrectly, which is due to both the lack of Liberal Democratic victories as well as the tendency of the model to favor the Conservative party. Clearly, the demographic factors alone are insufficient for predicting election outcomes accurately.

This model also suffers from collinearity, as evidenced by the large variance inflation factors (VIFs) presented in Table 3.2. The VIF estimates the inflation in the variances of the regression coefficients, and while their values do not affect the predictions, they have an impact on the correlations in posterior samples of the mean coefficients. More importantly, however, this model does a poor job of predicting outcomes, as it does not include constituency-specific information such as historical voting records and incumben-

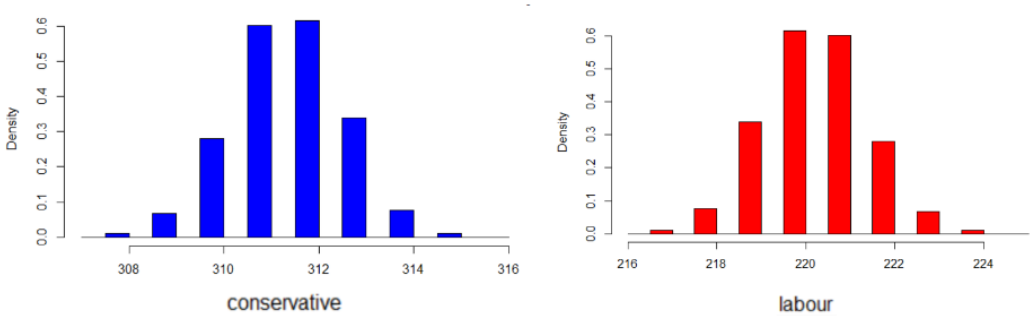


Figure 3.2 – Histogram of posterior predictive distribution of the number of seats won by the Conservative (left) and Labour (right) based on 8000 posterior predictive draws - Multinomial model (3.1)

cies. There is strong heterogeneity between ridings with similar demographic characteristics. As such, in the next section, we consider an extension that includes riding-specific random effects along with shared time effects. By accounting for the correlations between riding outcomes in successive elections using random effects, the collinearity issue vanishes.

Feature	VIF
Above age 65 (%)	58.0
White (%)	128.1
Christian (%)	137.3
Unemployed (%)	10.8
Employed (%)	85.7
Student (%)	4.7
House owned (%)	104.8

Table 3.2 – Variance inflation factor of the demographic variables

3.3 Multinomial Random Effects Model

3.3.1 Modelling Approach

Extending the Multinomial - Demographic Model (3.1), this section focuses on building a model that incorporates fixed effects for the election year, constituency-level demographic data, and random effects for constituencies. In a fixed-effects model, the only source of variability comes from the random sample we take to estimate our variables. However, in a mixed-effects model, we model the variances of the effects rather than their means.

A mixed-effects model can be described as:

$$\widehat{y}_{ij} = \alpha_{j[i]} + \beta x_i,$$

where β is the vector of fixed effects, and the random intercepts are captured by $\alpha_{j[i]}$ for group j of the i^{th} observation (Freeman, 2017).
We consider a multinomial generalized linear mixed-effects model for Y_{rit}, the total number of ballots cast for party i, $i \in \{1,\ldots,K\}$, in riding r in election year t with

$$\Pr\{Y_{r1t} = y_{r1t},\ldots,Y_{rKt} = y_{rKt}\} = \begin{pmatrix} n_{rt} \\ y_{r1t},\ldots,y_{rKt} \end{pmatrix} \theta_{r1t}^{y_{r1t}} \cdots \theta_{rKt}^{y_{rKt}},$$

where $n_{rt} = \sum_i y_{rit}$ denotes the total votes cast in riding r in election year t, θ_{rit} is the probability of party i winning riding r in election year t and

$$\theta_{rit} = \text{softmax}(\eta_{rit}) = \text{softmax}(\mathbf{x}_r\beta_i + \alpha_{r[i]} + \gamma_{it}),$$

In this model, $\mathbf{x}_r$ is a row vector of time-invariant demographic information from the census that is riding specific, γ_{it} represents the fixed-effect coefficients associated with the election year, and $\alpha_{r[i]}$ indicates the random effects that are riding specific and time invariant. As discussed in Section 2.3, to ensure identifiability, we add the constraint $\eta_{rKt} = 0$ for all ridings r and all election years t since $\sum_{i=1}^{K} \theta_{rit} = 1$.

Further, in the prior specifications, the demographic coefficients β_i and election year-based coefficients γ_{it} follow Normal priors, where $\beta_i \sim \text{No}\left(0, 5^2\right)$, and $\gamma_{it} \sim \text{No}\left(0, 5^2\right)$.

To account for the heterogeneity at the riding level, we include riding-specific random intercepts. Because the response is a multivariate vector, the random intercept for riding r is a $K-1$ vector; further, we assume the latter has a Multivariate Normal distribution with mean 0_{K-1} and covariance matrix Σ. Thus, $\alpha_{r_i} \sim \mathsf{No}_{K-1}(0_{K-1}, \Sigma)$. To complete the prior specification, we need to specify a hyperprior distribution on the covariance matrix Σ; we decompose the latter in terms of scale and correlation matrix into $\mathbf{S} \cdot \mathbf{C} \cdot \mathbf{S}$, where $\mathbf{S} = \text{diag}(\sigma_1, \ldots, \sigma_{K-1})$ is a diagonal matrix with standard deviation $\sigma_i \sim \mathsf{Ca}_+(0, 20)$ $(i = 1, \ldots, K-1)$, and Ca_+ denotes the truncated Cauchy distribution with location 0 and scale 20 on $\mathbb{R}_+$.

For the correlation matrix $\mathbf{C}$, we adopt the popular Lewandowski–Kurowicka–Joe (Lewandowski et al., 2009; Ramnath, 2016) prior with concentration parameter ν. The $\mathsf{LKJ}(\nu)$ prior density is proportional to the determinant of the correlation matrix, $p(\mathbf{C}) = |\mathbf{C}|^{\nu}$, with ν governing the strength of the correlation between entries: if $\nu = 1$, the prior is uniform over the set of all correlation matrices, whereas values of $\nu > 1$ (respectively $\nu 1$) lead to smaller (larger) correlations between entries.

We use STAN (Stan Development Team, 2020) to sample draws from the posterior distribution with four chains of size 3K, discarding 1000 warm-up iterations and keeping the last 2000 iterations. Visual diagnostics show that all have reached stationarity, and the parameters' $\hat{R}$ values range between 1.0 and 1.1. The effective sample size is above 2000 for all covariates, giving us large enough samples out of 8000 simulations to use in parameter estimation.

3.3.2 Model Outcomes

Figure 3.3 shows the bar plots of the results of 8000 draws in England for the Conservative and Labour parties. As with Multinomial - Demographic Model, this model also does not yield any victories for the Liberal Democrats and Other Parties for the 2017 election, although wins are included for these parties in the 2010, 2015, and 2019 posterior election draws. Therefore, this model performs better than the previous model in terms of capturing electoral wins for political parties winning few seats. In the 2017 English election results, the Conservatives won 297 seats, and Labour won 227 seats. The outcomes for this model show a mean of 284 seats for the Conservatives and 248 seats for Labour. Overall, 35 incorrect predictions were made, compared to the 72 incorrect predictions made by the multinomial demographic model. Thus, this model is more accurate compared to the previous models.

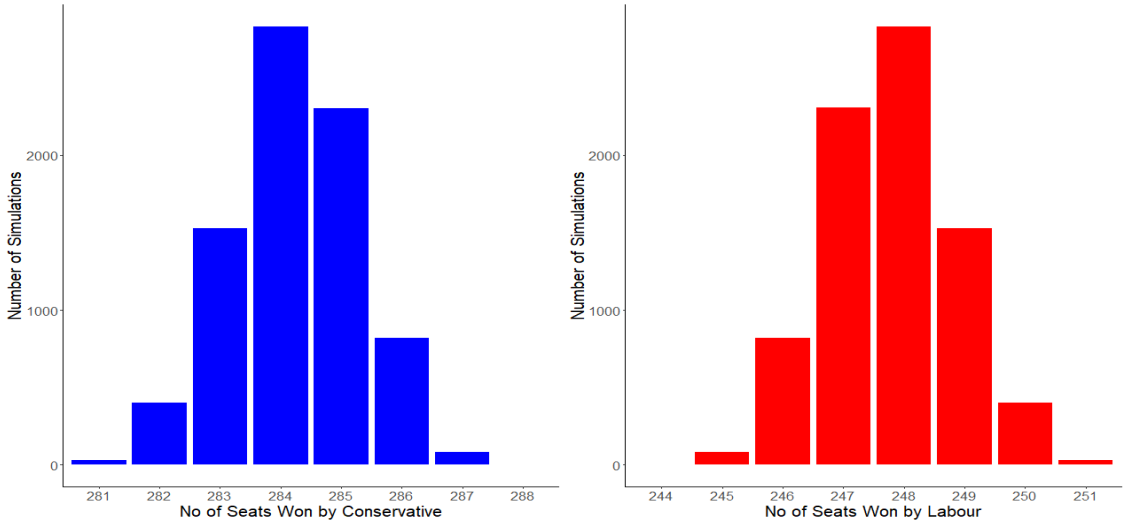


Figure 3.3 – Barplot of posterior predictive distribution of the number of seats won by the Conservative (left) and Labour (right) based on 8000 posterior predictive draws - Random effects model

3.3.3 Posterior Predictive Results

There are multiple ways to aggregate the posterior draws to examine the uncertainty in outcomes captured by the model at the constituency level. One way to aggregate these draws is by calculating the probabilities of winning. For example, in constituency X, if 80% of the posterior draws show wins for Conservatives, then the Conservative Party should be declared most likely the winner of constituency X. See Table 3.3 for an example of such aggregation.

	Constituency	Conservative	Labour	LibDem	Other	Winner
1	Bassetlaw	97.17	2.83	0.00	0.00	Conservative
2	Blackpool South	90.38	9.62	0.00	0.00	Conservative
3	Bosworth	0.01	0.00	99.99	0.00	LibDem

Table 3.3 – Aggregation of posterior predictive draws to represent the winner of the riding

Another method that can be used to aggregate the data is to calculate the margin of victory for every draw in a constituency and then present the mean margin of victory along with its relative standard deviation to account for the uncertainty in the margin of victory. An example of such aggregation is found in Table 3.4.

	Constituency	Predicted margin of victory	Relative standard deviation
1	Worcester	196	76.5%
2	Portsmouth South	172	74.2%
3	South Swindon	884	27.8%

Table 3.4 – Aggregation of posterior predictive draws to represent the margin of victory and its associated uncertainty

For now, we will use these two data aggregation methods to represent the uncertainty in the winning probability and margin of victory.

Posterior outcome draws vs Observed election outcomes

We will examine how much uncertainty our model captures in the winning probabilities from the posterior predictive draws by comparing them to the actual results. Accordingly, marginal seats should have winning probabilities close to 50% for both of the competitors

in that marginal election, meaning that the model captures the uncertainty appropriately in its posterior estimations. This section will use posterior draws aggregated according to the method used in Table 3.3. We will analyze the posterior draws year-wise and address the following questions:

- How many correct and incorrect results can be found in the posterior draws for the respective years?
- Does the uncertainty surrounding marginal seats influence the probabilities of the outcomes?
- What could be contributing to the mismatch of the posterior draws with the actual results?

Posterior predictive draws for the 2010 General election

Fifty results from our model differed from the actual 2010 election results, as seen in Table 3.5. Only eight out of these 50 constituencies had the probability of winning for the projected winner at other than 100%, which indicates some overfitting.

	Constituency	Conservative	Labour	LibDem	Other
1	Bassetlaw	97.17	2.83	0.00	0.00
2	Blackpool South	90.38	9.62	0.00	0.00
3	Bosworth	0.01	0.00	99.99	0.00
4	Bity of Shester	0.16	99.84	0.00	0.00
5	Darlington	85.80	14.20	0.00	0.00
6	Gedling	81.74	18.26	0.00	0.00
7	Ilford North	9.32	90.67	0.00	0.00
8	Wirral West	1.71	98.29	0.00	0.00

Table 3.5 – Winning probabilities in constituencies with predicted winners that differed from the actual winners of the 2010 elections

The majority of incorrect predictions for the 2010 elections were for constituencies that were won by one party by overwhelming margins in 2015, 2017, and 2019. Therefore, the results favored particular parties due to the random effect created by those con-

stituencies.

For example, in the ridings of Brighton Kemptown and Ealing Central Action, the Conservative party won the 2010 election. However, our model favored the Labour party for these ridings since the Labour party won these constituencies by large margins in the 2017 and 2019 elections and the random effects factored into our model. Another noteworthy finding is that 15 of the 50 incorrect predictions came when the actual winners were Conservative and Labour candidates, but the predicted winners were Liberal Democrats. Therefore, the model adjusted for the shared-time effects of the parties because the Liberal Democrats won more seats in 2010 compared to other elections, in which their vote share was inefficient.

Posterior predictive draws for the 2015 General election

Out of 532 constituencies, 29 results of the 2015 election for our model are different from the actual results of the 2015 election. Among these outcomes, six results have uncertainty in the outcomes, as presented in Table 3.6.

	Constituency	Conservative	Labour	Libdem	Other
1	Derby North	27.19	72.81	0.00	0.00
2	Dewsbury	96.15	3.85	0.00	0.00
3	Middlesbrough South and Cleveland East	76.71	23.29	0.00	0.00
4	Twickenham	0.95	0.00	99.05	0.00
5	Wolverhampton South West	88.81	11.19	0.00	0.00
6	Barnsley East	0.00	12.56	0.00	87.44

Table 3.6 – Winning probabilities in constituencies for which the predicted winners are different from the winners of the 2015 elections

Middlesbrough South and Cleveland East have 23% probabilities of a Labour victory based on the posterior predictive results, even though Labour comfortably won these constituencies during the 2015 general election. While the election history from 2010 to 2019 for this riding does not indicate that it was very competitive, the Conservative party

won this seat by a tremendous margin in the 2019 general election. Due to this one odd outcome, the random effects for the constituency steered the model in the Conservative's favor in the 2015 election posterior draws.

Similar patterns are observed in both the 2017 and 2019 posterior predictive draws. In some of the instances, the posterior draws do not change heavily across time and depend highly on the random effects. In some cases, we also observe that the model outcomes behave according to the demographic features. Thus, this type of model produces outcomes that are based on demographic features as well as random effects.

Findings From Comparisons of the Outcomes

We have the following model performance findings:

- The model can capture uncertainty for constituency-based random effects, and the uncertainty in those outcomes adjusts across time.
- The impact of the random effects on the outcome varies by election year.
- Some of the uncertainty is captured by covariates' fixed effects instead of the random effects.
- The observed variability in some ridings is captured due to peculiar results in a specific year rather than close results.

Prior Sensitivity

As mentioned in the model specification, we have the following hyperprior specifications for σ_i and correlation matrix $\mathbf{C}$ for the random effects.

- $\sigma_i \sim \mathsf{Ca}_+(0,\tau)$ $(i = 1,\ldots,K-1)$
- $\mathbf{C} \sim \mathsf{LKJ}(\nu)$

If ν=1, the density is uniform over the correlation matrix of each dimension, and, if ν> 1, the identity matrix is the mode of the prior, with a sharper peak in density for larger ν

values (Bürkner, 2017). By applying the LKJ prior to the model, the expected correlation between the parameters can be controlled. The half-Cauchy distribution $\mathsf{Ca}_+(0,\tau)$ has a peak located at 0, and its width is dictated by a positive scale parameter τ. Thus, a higher value of τ would provide a wider prior distribution for σ_i (Bois, 2019).

We can assess the sensitivity to prior specification by varying both of these priors to observe the changes in the posterior draws for the random effects and checking to see the specification is having any undue influence on the model outcomes and uncertainty. The specifications of priors for different models are shown in Table 3.7 and the results indicate little sensitivity. We observe in Figure 3.4 that the posterior draws of the random effects for the Bury North constituency in the various models hardly change regardless of the priors, and the outcomes of the posterior predictive draws do not change. Thus, our model makes robust estimates of parameters.

Model	Priors
1	$\sigma_i \sim \mathsf{Ca}_+(0,1)$
	$\mathbf{C} \sim \mathsf{LKJ}(1)$
2	$\sigma_i \sim \mathsf{Ca}_+(0,5))$
	$\mathbf{C} \sim \mathsf{LKJ}(1)$
3	$\sigma_i \sim \mathsf{Ca}_+(0,5)$
	$\mathbf{C} \sim \mathsf{LKJ}(2)$
4	$\sigma_i \sim \mathsf{Ca}_+(0,20)$
	$\mathbf{C} \sim \mathsf{LKJ}(1)$

Table 3.7 – Priors for different models

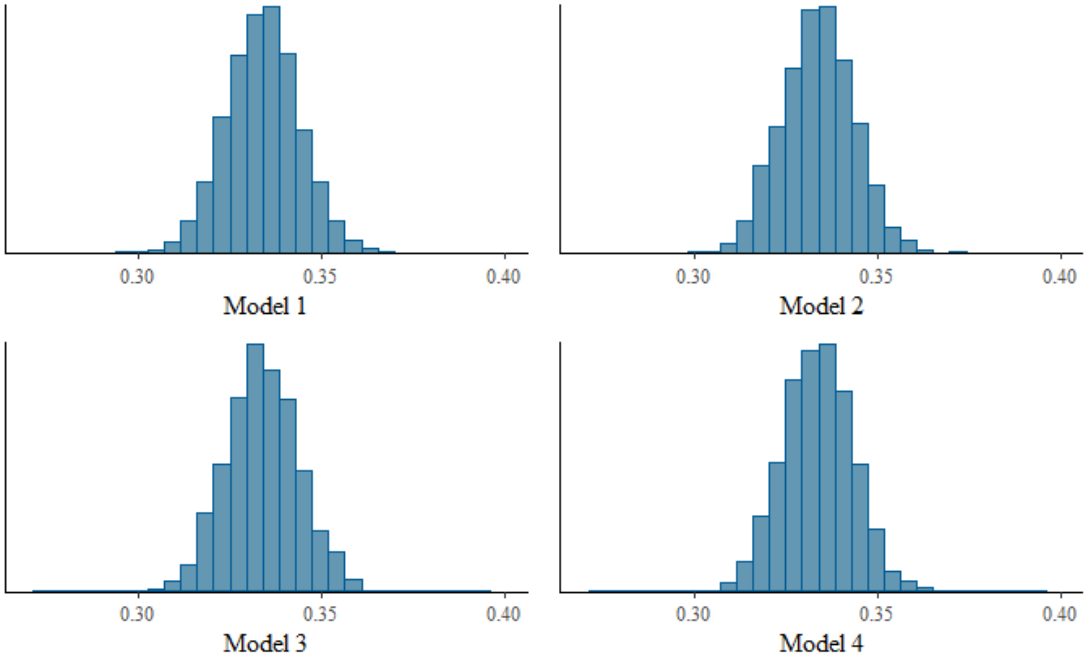


Figure 3.4 – Posterior draws of random effect for the Bury North constituency for Model 1[$\tau = 1, \nu = 1$] (top left), Model 2[$\tau = 5, \nu = 1$] (top right), Model 3[$\tau = 5, \nu = 2$] (bottom left), Model 4[$\tau = 20, \nu = 1$] (bottom right)

Mapping Changes in Voting Intentions Using an Offset

We consider a simple scenario in which an offset term is added to a multinomial logistic regression model. The purpose of the offset in the context of election modelling is to account for changes in voting intentions relative to the base election year. For simplicity, we only consider a single poll as opposed to, say, a polling average, but this could be easily generalized. Once the offset is added to the model, we get a new posterior predictive distribution of election outcomes.

Our multinomial mixed model incorporates a fixed effect for each election year. Thus, the relative overall strength for each year is incorporated into the intercepts of fixed effects for each year. Suppose that average probabilities of success are 0.4, 0.35, and 0.10 for parties A, B, and C and that the poll indicates a swing of 3 points from party A to party B, which corresponds to 0.37, 0.38, and 0.10 point estimates of the average probabilities with associated uncertainties. To incorporate this change using an offset, we must map the change on the probability scale to a change on the log-odds or odds scale. The easiest way to do so is by simulating new outcomes (using the poll's sample size and estimated voting intentions as values) and then mapping them back to the linear predictor scale. Specifically,

- Simulate B new multinomial draws with n trials each and probabilities specified by the new voting intentions, resulting in sample proportion of $p_i^{(b)}$ for each of the i parties, where $i = 1, \ldots K$.
- Calculate the log odds ratio of voting intention (relative to the overall average from the model p_0), say $o^{(b)} = \log(p^{(b)}/p_0)$.
- Add $o^{(b)}$ to the linear predictor of each constituency and draw new multinomial outcomes for a new posterior predictive distribution.

For example, if there were a change in voting intention relative to the baseline prediction, where the change includes a 5% increase for the Conservatives, a 7% decrease

for Labour, a 2% increase for the Liberal Democrats, and no change for the Other Parties. If we incorporate these changes by adjusting the offset in our model, we arrive at the changes in our posterior predictive outcomes shown in Table 3.8.

			Quantile			
	Mode	Min	25%	50%	75%	Max
Conservative seat changes	59	37	55	58	61	79
Labour seat changes	−64	−81	−65	−63	−60	−44
Libdem seat changes	5	0	3	5	5	10
Other seat changes	0	0	0	0	0	0

Table 3.8 – Changes in seats after adjusting for voting intention changes with a 5% increase for the Conservatives, 7% decrease for Labour, and 2% increase for the Liberal Democrats

The example above shows how a swing in intentions led to seat changes. We can assign uncertainty to those changes by assessing the variability in the outcomes. The voting intention changes indicate big swings in the Conservative and Labour's seat outcomes, but the two-point increase in the Liberal Democrat's voting intentions failed to translate into a big change in Labour's outcomes. Thus, it is necessary to embrace variation, which, in turn, requires accepting uncertainty (Vasishth and Gelman, 2019). In order to facilitate a proper cognitive interpretation of changes in outcomes, this uncertainty must be effectively communicated to users. The next chapter will focus on communicating election data and its uncertainty.

Chapter 4

Communication of Election Outcomes

A Bayesian model was used in the previous chapter to model election outcomes and capture the variability in the predictive outcomes. Moreover, the offset enabled us to generate new results reflecting the changes in voting intention relative to previous elections. Firstly, rather than providing only point estimates of election results, we have the distribution of the election results from which summaries such as mean and range of the number of seats can be easily extracted. The uncertainty in the outcomes can be notably determined by the standard deviation of the predicted outcomes. Secondly, when the offset for voting intention is adjusted, the model predicts a new outcome along with the original prediction, making it necessary to analyze both outcomes to evaluate the change in voting intention. It's clear that the information provided by the model is complex and must be visualized appropriately for effective assessment and interpretation.

Election results and prediction data are available in massive quantities from a variety of sources (news websites and individual models) as an election draws near. Often, these different sources use various data aggregation techniques to represent their prediction data. Further historic election information is available and maintained by governments on their websites and it is stored in primitive data formats following no specific framework. In addition, there are no recommendations as to how to visualize, and gather insights

from, this data. Meanwhile, researchers are trying to find ways in which this web-based data can be visualized in a precise and easy-to-interpret manner.

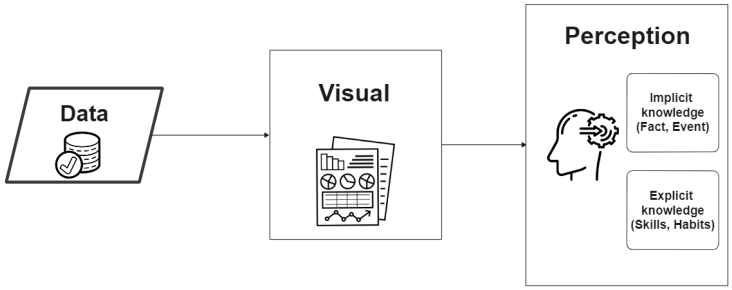
4.1 Visualization Framework

According to Mackinlay et al. (2007), the primary objective in data visualization is to gain insight into an information space (collected raw data). Further, visualization techniques follow various frameworks. Three such frameworks are addressed in Chen et al. (2008), which discusses data, information, and knowledge. Data consist of attributes of objects that are stored in raw form. Information concerning data is presented using various statistical measures and data attributes, helping to address questions as to “who”, “what”, “where” and “when” (Ackoff, 1989). Finally, knowledge provides domain-based assistance in addressing the "how" question.

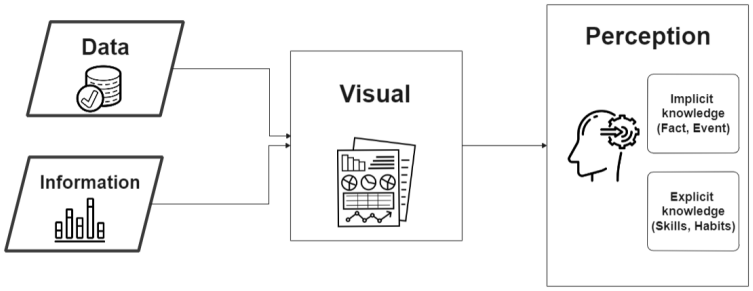
The first framework shown in Figure 4.1a is a search-based model, where data is visualized using a given graph, layout, and colour scheme. Users are meant to make sense of the visualized data via their perceptions and cognitive spaces. This framework can be employed when the user is equipped with enough prior knowledge to draw inferences (Chen et al., 2008). In the information-assisted visualization framework shown in Figure 4.1b, the main visualization is accompanied by supplemental information providing statistical insights (e.g mean, standard deviation, histogram of distribution), and aggregated attributes and traits, such as colour and hue, in order to guide the user. This framework is especially crucial for complex data with multiple attributes. The third framework presented in Figure 4.1c is a knowledge-assisted visualization model that provides domain-based guidance either implicitly in the visual design, or explicitly through the inclusion of domain-based information. In many instances, users lack domain knowledge, making it difficult for them to understand the visuals; thus, this framework can assist users by bridging the gap between visualization and inference. This last framework is also help-

ful in removing any prior cognitive biases that would interfere with making meaningful inferences.

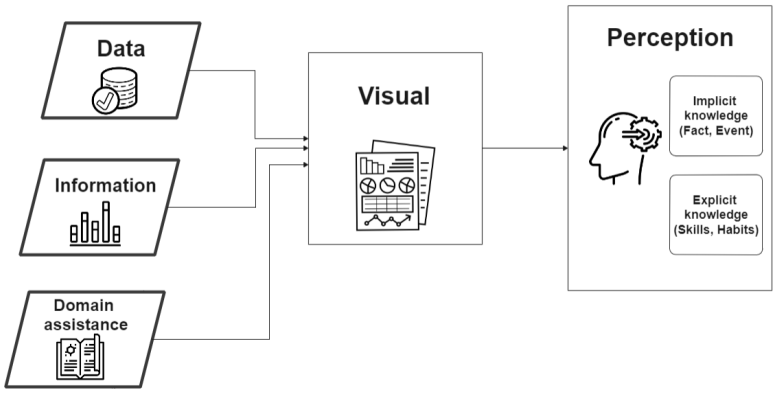
In our case, we certainly cannot adopt the first framework, as election forecasts do not contain the entire election narrative. Information assistance is required to understand the full narrative. For example, should a hexagonal map of election outcomes be created using this framework, it will not contain information beyond the basic visual, and its legend will only be able to explain election patterns in the regions covered by the map. It will be unable to address party voting shares, both overall and region-wise. Further, since the predicted outcomes are not point estimates and come with their associated uncertainty, no basic visual (univariate choropleth maps) will be able to explain the data fully. For example, such a visual will be unable to identify which regions have high uncertainty in terms of results, and which parties have higher uncertainties in terms of seats. Interpretation will become even more difficult when the attempts are made to compare the two results, i.e, the first one containing the original prediction, and the second one containing the prediction after changes in voting intentions have been considered. In this case, the first framework cannot address how many seats will swing, which party will have the maximum swing, and which swings will lead to close election results. Further, should two hexagonal maps be used (one for each result) absent statistical insights, it would be difficult to find subtle indications of swing. Thus, the first framework cannot be used to communicate the predicted election results



(a) First visualization framework – The search-based framework



(b) Second visualization framework – The information-assisted framework



(c) Third visualization framework – The knowledge-assisted framework

Figure 4.1 – Visualization frameworks created and adapted as per the designs provided by Chen et al. (2008)

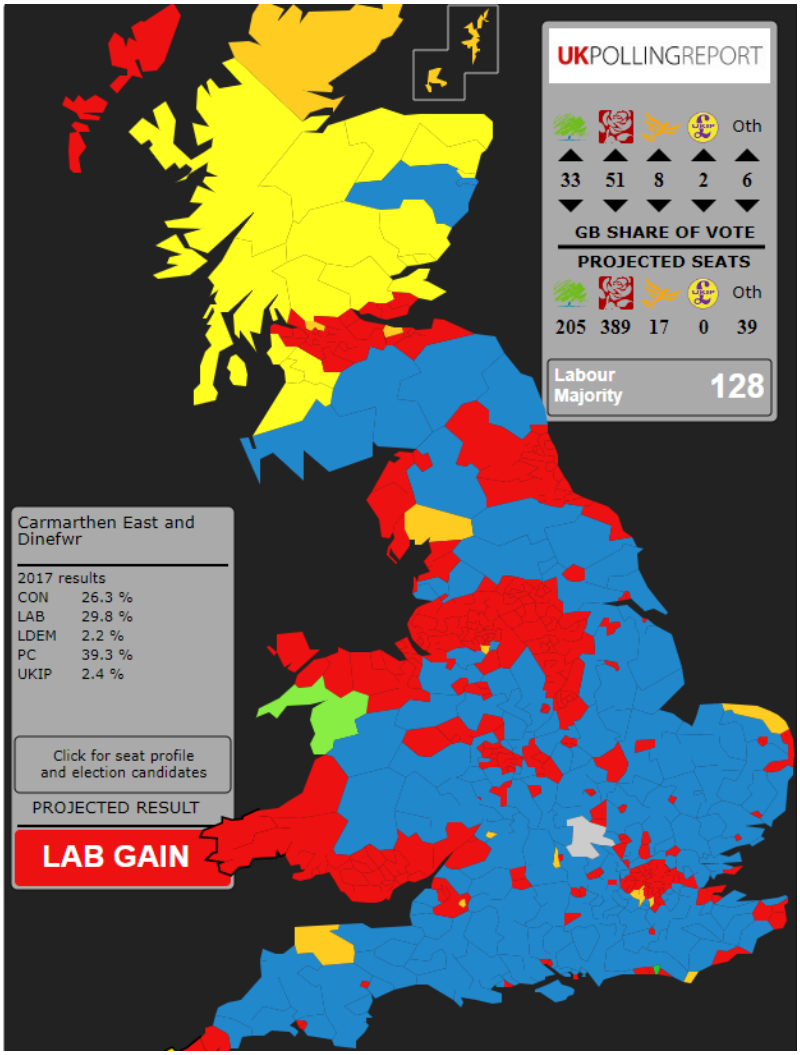


Figure 4.2 – Voting intention based swingometer provided by the UK Polling report and created by Wells (2020)

The second framework uses information to assist visualization. Summary tables of voting shares per party and shares of seats per party in addition to the visualization will help the user obtain a better picture. Further, we can include uncertainty by, incorporating the range in the share of seats going to a party, as shown in Figure 4.3. When interpreting two results due to shifts in voting intentions, we can provide a table of ranges in swings per party. In addition, when examining the changes in the results due to swings in voting intentions, we display results visually, as per Figure 4.4.

The second framework also allows us to provide visuals of multiples predictions resulting from multiple swings in voting intentions to help users build a narrative of the model outcomes. Further, the effectiveness can be maximized via interactivity. The visuals used in the UK polling report (Wells, 2020), shown in Figure 4.2, are a great example of how interactive swing visuals can be combined with information assistance. The report includes an interactive button that allows the user to increase or decrease the swings in voting intentions (i.e., operate the swingometer) and thus build a perception of the voting intention parameters that could lead to a marginal election. Adopting an information-assisted framework would allow us to address specific questions while helping a user with no background in UK elections develop a decent understanding of the potential election outcomes.

The third framework could be extremely helpful in providing quality insights, as it provides domain-based assistance. Domain-based assistance could be provided listing information about the previous election or a table listing changes in seats regarding the previous election. Domain-based assistance can also include displaying specific facts, such as seats needed to win a national election in the UK. Further, brief descriptions concerning the political system, nature of the parties, and Brexit could prove useful, particularly for users unfamiliar with the UK election system. Thus, domain-based assistance helps users interpret election outcomes within the proper context and, short explanations of data and modelling processes can be added. When our two outcomes are presented, it would be good to be able to describe how our model predicted the new outcomes based on changes in voting intentions and include the baseline year for the offset. Basically, domain-based assistance can vary according to specific user requirements. Overall, a domain-assisted visualization framework would be very beneficial in our work.

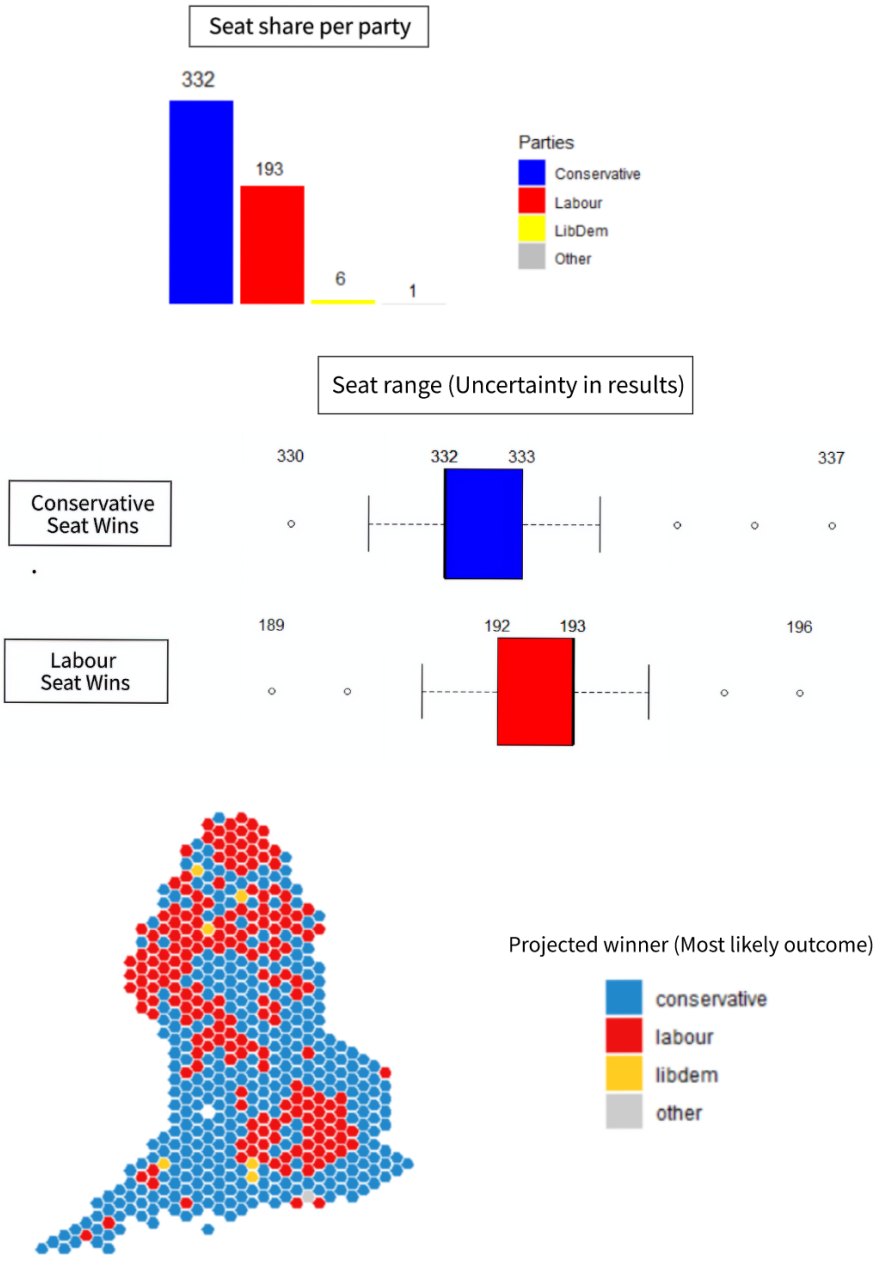


Figure 4.3 – Hexagonal map (bottom) of baseline predicted England election outcomes, along with statistical insights in form of seat share per party (top) and seat ranges depicting uncertainty (middle)

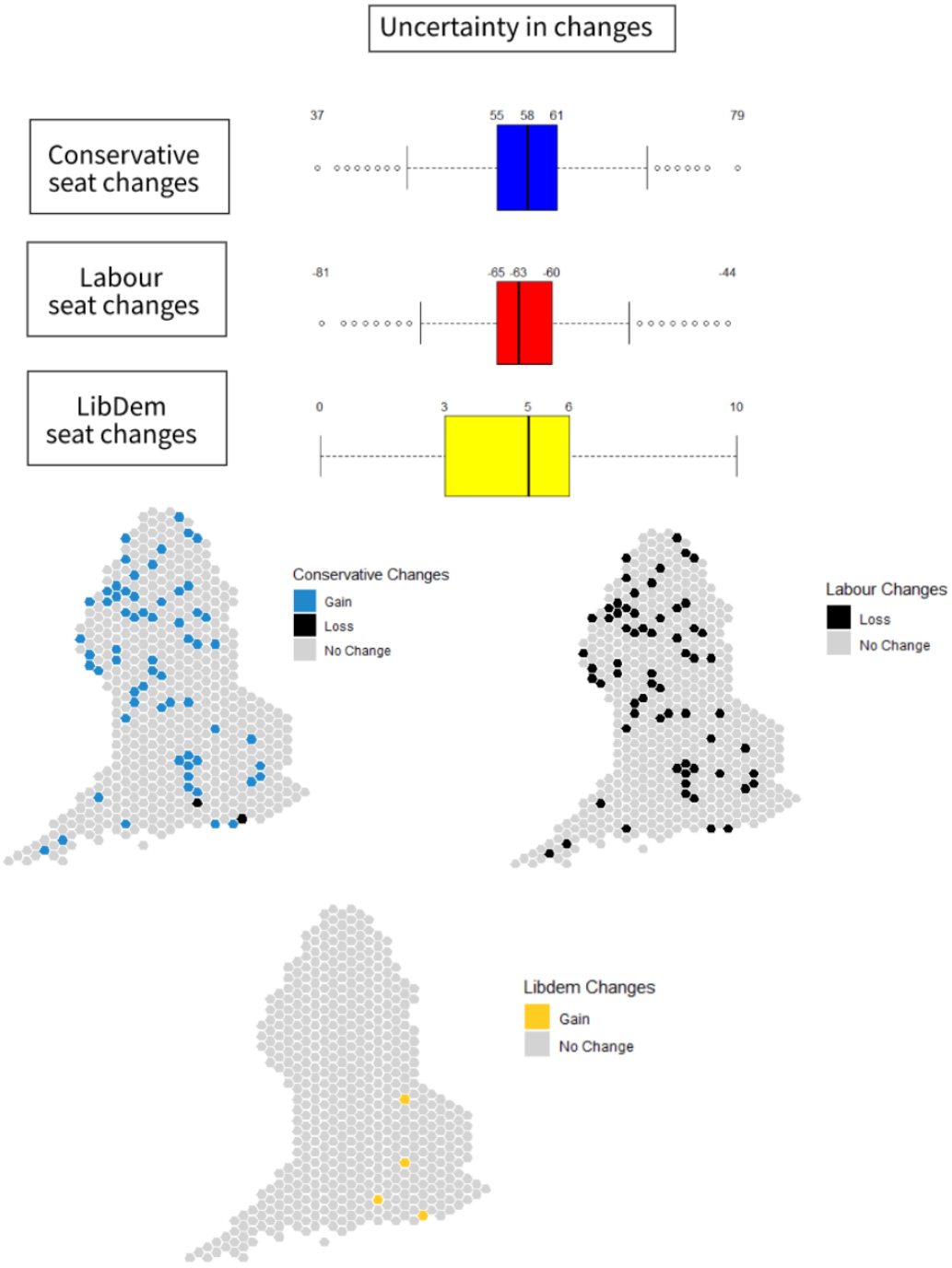


Figure 4.4 – Hexagonal map (bottom) of changes in England election outcomes when there are shifts in voting intentions, plus information assistance in form of statistical insights (top) into the uncertainty

4.2 Visualizing Election Outcomes

Hadi et al. (2020) developed an open-source tool for visualization of Canadian election data. The researchers noticed how difficult it was for users to make informed decisions without gathering, analyzing, and researching the outcomes and details of past elections. The number of candidates, party popularity over time, opinion or intention shifts, and trends in party support based on geographic or demographic data for any given election are presented in textual formats that are not only challenging to comprehend but also make it difficult to derive conclusions without detailed analyses. The absence of a visualization framework for representing this vast amount of raw data prevents meaningful insights.

The objective of this research was to design an interactive platform that could generate powerful graphs and visualizations of this data in a geospatial format to improve the non-expert users' comprehension of election data. This objective was achieved via choropleth maps with colour notations for complex information, such as the number of seats won, and votes and seats shared by any political party for any given election. These maps allow the user to explore this massive amount of election data in an efficient manner. Information such as the percentage of seats won or the number of votes won can be shown with complementary graphs, including bar graphs, and pie and donut charts. This combination of elements produces an infographic capable of providing an overall sense of the most and least prominent players in an election, along with some visual representations of the results.

Hadi et al. (2020) shed light on how election data can be illustrated with vertical bar graphs and line graphs. Trend analysis can also be done with line graphs featuring information such as the number of seats won, percentage of votes won, and percentage of seats won, as well as with regression lines for each party to identify the trends such as popularity. For instance, a line would make it easy to see that although the Liberals in Canada have won most of the elections throughout history, their popularity has been slowly de-

creasing in the long run. 60% of the surveyed participants of Hadi et al. (2020) knew little about Canadian elections, yet they were capable of comprehending and achieving a comparative understanding of this process based on the visualizations provided. Thus, when depicting UK election outcomes using an Information-assisted framework, we can use other displays, such as line charts, and bar graphs, alongside the original map-based visuals. Further, to provide additional clarity in the interpretation of the predicted model outcomes, we can present the original trends in election outcomes using line graphs.

In work by Gupta et al. (2016) on the importance of design and visual discovery for communicating complex insights from election data, it was found that a large number of skilled analysts and journalists are required for the national news outlets to do a good job in communicating national election results to audiences, but there was lack of such resources for some states in India to analyze state elections. Thus their research focuses on developing and deploying a set of tools or templates for creating meaningful, editable, generic visualizations that can be used to convey complex election data.

The premise of this research is that creating any visualization template capable of presenting complex information involves taking advantage of three major components, plus their subcomponents:

- Visual encoding, including the use of colour, sizes, areas, volumes, and angles
- Different types of data types, including scales, measurements, and attributes
- Relationships or correlations among the various entities

To take advantage of the above-mentioned points, generating an appropriate balance between graphical perception, design, and aesthetics is the key in build a successful infographic for complex data, such as election data. If we utilize these characteristics meticulously, even multi-faceted data, such as the data for the distribution of seats, voting shares, alliances formed among parties, voter turnouts, and state-level data can be articulated in

a single infographic without overwhelming the user or increasing their cognitive load.

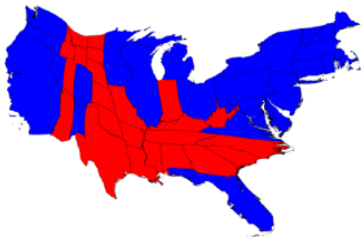
Many principles could have been used to develop the infographic. For example, it was suggested that various saturations of the same hues be used to depict the margin of victory on a map. It was also suggested that colour palettes that are commonly misperceived by individuals with colour vision deficiencies not be used. Thus, we should use colours with significant differences in shades and give the user a way to change the colours on an interactive map. A good visualization design requires an appropriate chart, which can have multiple attributes. For instance, a fresh change in behavioral patterns or uncertainty in elections can be demonstrated by using colour-coded shades on timelines. Thus, principles such as these should be kept in mind when creating infographics.

Choropleths are great to visualize quantitative geodata, but they tend to overemphasize large administrative units by assigning them stronger visual weights. A classic example of this phenomenon is a typical US election map, where central and western states receive bigger visual weights compared to smaller eastern states, producing area-size biases. Choropleths are also not good at visualizing intra-regional variations in data since it displays them uniformly across the shape of the original geographic unit, suggesting uniformity across space, which is not usually the case. The shortcomings discussed above have been debated in geographic domains, and alternatives have been considered (Skowronnek, 2015).

One such alternative, shown in Figure 4.5, involves cartograms. These are non-true maps or map-like visualizations that try to depict abstracted geographic space, which involves some degree of distortion of the actual geographic space. Non-contiguous cartograms have been suggested as an alternative to choropleths since they are not connected, leaving the user free to increase or decrease the sizes of entities. Dorling cartograms are depictions of geographic data using higher degrees of spatial abstraction. The entities are usually shown as circles positioned on the geographic centroid, where overlap with

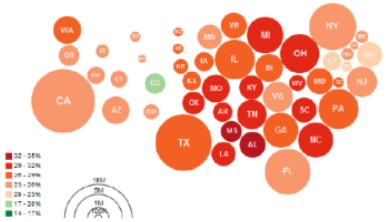
other regions is avoided. However, research suggests that these maps are not adequate for administrative areas of different sizes. Because we are using constituency-level data from constituencies with highly varying sizes, we would prefer not to use these maps to represent UK election outcomes.

The second alternative involves grid maps. These maps have recently gained in popularity based due to their adoption in the data journalism and data science domains. They are very simple, with each entity depicted by a simple polygon. Unlike cartograms, they rely on colour shades and hues, not size. The third option we have is spatial treemaps. The advantage that spatial treemaps have over choropleth maps is their ability to display hierarchical data using nested rectangular polygons. The challenge, however, is the recognizability of, and the correlations between, the entities within the maps. The extensive research done in this field in terms of visualization can be utilized for our case study. We have frequently used grid maps due to the simplicity of transmitting information to the user while taking advantage of colours and shades for visual emphasis or the distinction between the concentration or amount of an attribute conveyed by a polygon.



Cartogram

Non-contiguous cartogram



Dorling cartogram

Figure 4.5 – Cartograms by Skowronnek (2015), Top: Cartogram of the 2012 US presidential election results; Democrats (blue) won over Republicans (red). Bottom left: Non-contiguous cartogram of the United States showing each state resized in proportion to the relative influence of the individual voters who live there. Bottom right: Dorling cartogram of obesity levels in the United States, 2008

4.3 Communication of Uncertainty

Now that we have built a robust model that can capture uncertainty in a specific manner, we need to understand how to best visualize it using state-of-the-art techniques. Much research has been done on depicting uncertainty visually for geospatial data, but only some systemic user reviews have been used in empirical evaluations of such techniques. Kinkeldey et al. (2014) reviewed and analyzed the studies that focus on the usability of these techniques in communicating uncertainty.

In the Kinkeldey et al. (2014), the assessment was done of the three main ways of representing uncertainty: coincident/adjacent, intrinsic/extrinsic, and static/dynamic. In intrinsic techniques, uncertainty is represented by making changes in colour and hue (Aerts et al., 2003) on bivariate maps of the type shown in Figure 4.6, whereas, in extrinsic techniques, other objects such as error bars, and glyphs, are used to visualize the uncertainty (Alberti, 2014), as shown in Figure 4.7. In a coincident approach, the data and uncertainty are integrated into the same view; in contrast, in the adjacent approach, they remain in separate views (MacEachren, 1992). As the name suggests, the static approach includes no animations, while the dynamic view can include interactive and non-interactive animations for depicting uncertainty. The findings regarding these approaches have been developed in the lab, on the web, and in interviews and focus groups (Kinkeldey et al., 2014).

Further, most of these assessments were based on both objective and subjective tasks. Objective tasks are focused on value retrievals, where the data and uncertainty values have to be identified separately (Aerts et al., 2003). Subjective tasks are done to evaluate the ease of use of a particular visualization technique, preferences in visuals, and the intuitiveness of symbols. Findings are based on user performance, acceptance, user confidence, user preference, and intuitiveness.

The following recommendations were made in Kinkeldey et al. (2014) regarding the three main ways to represent uncertainty based on user preferences.

Intrinsic/Extrinsic Colour hue and values, as well as transparency, are recommended over colour saturation to represent uncertainty. Whitening (Hengl, 2003) did not appear to be effective. At the same time, changes in the resolution were recommended over transparency and changes in colour attributes. However, extrinsic displays, such as glyphs and grid-based techniques, were considered better than intrinsic techniques. Finally, in situations where only overviews were needed, the intrinsic approach was preferred, whereas

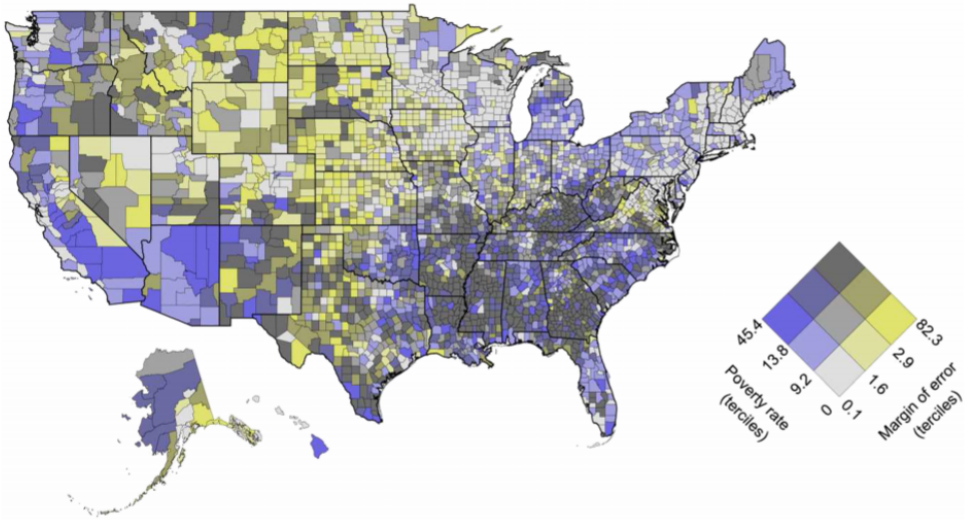


Figure 4.6 – Map of US showing poverty levels in 2015, along with grid representing the margin of error, from Lucchesi and Wikle (2017)

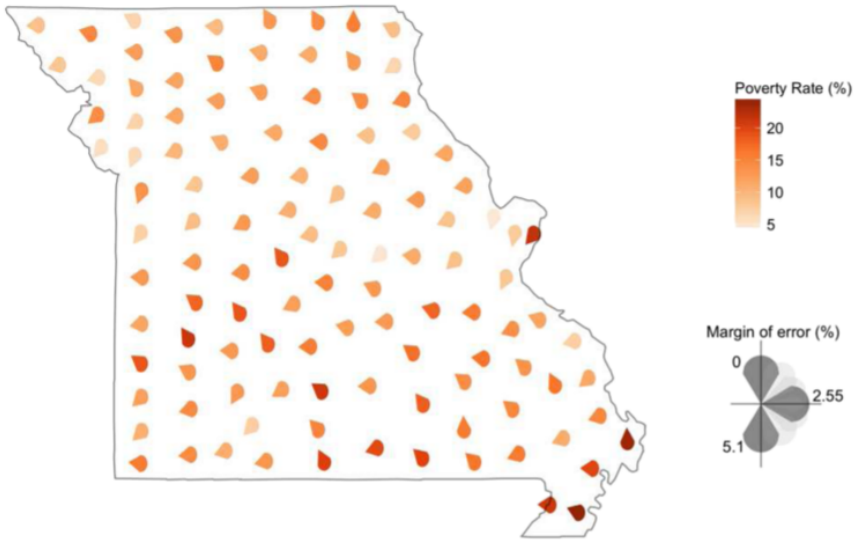


Figure 4.7 – Map of Missouri showing poverty levels in 2015, along with glyphs representing the margin of error, from Lucchesi and Wikle (2017)

extrinsic techniques were preferred for in-depth analyses of uncertainty.

Coincident/Adjacent For simple tasks, both maps were equally efficient, but adjacent views were not as helpful for complex ranking tasks involving the retrieval of more than one value. One study (Kubicek and Šašinka, 2011) suggested that response time was better for coincident maps. Thus, coincident maps were preferred unless they became too cluttered while trying to display uncertainty.

Static/Dynamic The dynamic approaches were difficult to assess, as numerous dynamic techniques exist. It appears that animated visuals are helpful, but only for those cases in which static methods were feasible; otherwise, no significant performance improvements were observed.

A general perception was that adding uncertainty in geo-spatial visualizations made the user feel overwhelmed due to the amount of information being displayed. However, research suggests that when adequate solutions are provided for visualizing uncertainty, users were more accepting (Aerts et al., 2003).

Although Kinkeldey et al. (2014) does not contain much information related to user confidence in uncertainty visualizations, the studies they surveyed suggested that when a technique was accurate to represent uncertainty and had a fast response for a task then user confidence was high. When user preferences were evaluated, it was found that user preference and performance were sometimes mismatched. However, the key observation was that user preference is a vital factor in user acceptance. Fuzziness was ranked high when assessing intuitiveness and light colours were preferred when indicating uncertainty, whereas darker colours were preferred when indicating certainty (MacEachren et al., 2012).

Thus, by taking the user perspective into account, we realize that the effectiveness of the various techniques in communicating uncertainty through maps changed across different domains. However, some essential insights were gained from the various studies done on this issue. Since our model captures uncertainties in election outcomes that must be analyzed to a certain depth, extrinsic techniques appear to be a good fit. However, since the data is at a constituency level, the extrinsic technique may result in congestion. Also, quick response times are not necessary, and, since only some outcomes represent uncertainty, there is no cluttering involved. Hence, a coincident approach could be used. There is also no need for animated maps, as statics maps can convey the relevant uncertainty parameters but for implementing a voting intention swingometer, animated maps should be used. By keeping these points in mind while designing the visualization, we can represent the uncertainty present in the random effect modelling technique well.

Given that we have information available at the constituency level for the estimates and their uncertainties, we mapped the margin of victory and relative standard deviation as per the Table 3.3, of the predicted election outcomes after accounting offset based changes in voting intention (+5% Conservative, - 7% Labour, +2% LibDem) as per section 3.3.3, for all constituencies using the library provided by Lucchesi and Kuhnert (2017), as shown in Figure 4.8. As mentioned above, any technique that results in a sense of confusion by cluttering up information or extrinsic objects should be avoided. The equivalent glyph structure diagram of Figure 4.7 contains 532 such structures which lead to congestion, and they should not be utilized. Thus, the bivariate map presented in Figure 4.8, is more effective. Furthermore, maintaining the above-mentioned considerations in mind, one may be able to adapt the uncertainty representation techniques to the specific context of the task at hand. All of the above research contributes to the construction of strong visual designs in order to communicate information. However, there is scope to perform user research assessment of uncertainty representation within a given domain.

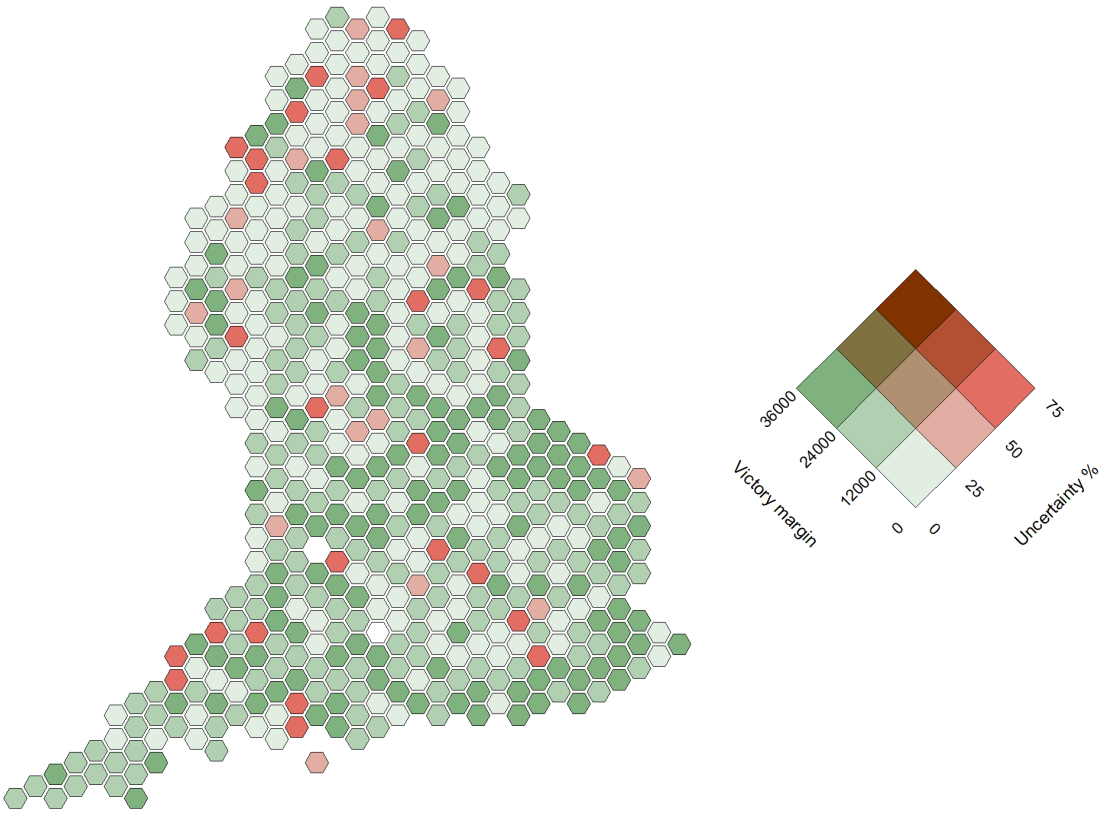


Figure 4.8 – Margins of victory and uncertainty (Coefficient of variation or Relative standard deviation) of the predicted election results after accounting for changes in voting intention (+5% Conservative, - 7% Labour, +2% LibDem) using the offset, presented on a hexagonal map with a bivariate legend

Conclusion

We developed a Bayesian model capable of capturing the uncertainty of the predicted election results for England. The idea was to model these uncertainties so that they help users in comprehending election outcomes. Adding random effects to the multinomial model helped alleviate any collinearity among covariates; and we can use it to create a predictive model using an offset-based swingometer, although the use of this mixed model leads to overfitting in some cases. Another important finding was that it was essential to deal with the identifiability issue when modelling multivariate outcomes. Some future challenges include the addition of constituency-specific random slopes for the covariates and time-based random effects. It would be interesting to analyze how these changes would affect uncertainty in the simulation outcomes and observe how they impact model overfitting.

Additionally, we examined visualization techniques and the ways in which they affected the representation of uncertainties. To begin with, we explored how adapting information-assistance and domain-assistance frameworks can improve cognitive interpretations of visual information. It is also evident that the use of different types of visuals to represent election results can help in communicating them to an audience with little to no prior knowledge of election results. When developing visuals of this type, it is necessary to take into consideration color, hue, and saturation. Choropleth is not the only type of geo-spatial visualization that can be used, and since we have data at the constituency level, grid maps are more suitable for representing outcomes to eliminate contingency. In addition, user experience research suggests that various representational strategies, in-

cluding intrinsic/extrinsic, coincident/concurrent, and static/dynamic, can be evaluated on the basis of various performance metrics and acceptance metrics to better determine which strategy would be appropriate for a given task. A large number of user studies have been conducted regarding communication techniques of uncertainty. However, they could be further analyzed under a domain-specific perspective. To evaluate user performance for a domain-specific application, A/B testing could be carried out on different representation techniques. Lastly, qualitative user research can be conducted on the use of interactive figures which enable users to experiment with different data filters and visual tools. By doing so, it is possible to perform a more comprehensive analysis for evaluating the effectiveness of visualization and representation techniques. This thesis exercise helped build an understanding of the basics of the research process and how to address questions using statistical techniques and tools, all of which went toward building a model for improving the user experience within a specific domain.

Bibliography

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16(1):3–9.

Aerts, J. C. J. H., Clarke, K. C., and Keuper, A. D. (2003). Testing popular visualization techniques for representing model uncertainty. *Cartography and Geographic Information Science*, 30(3):249–261.

Alberti, K. (2014). Web-based visualization of uncertain spatio-temporal data. Master's thesis, Utrecht University.

Betancourt, M., Byrne, S., Livingstone, S., and Girolami, M. (2014). The geometric foundations of Hamiltonian Monte Carlo. *Statistical Science*, 23.

Bois, J. (2019). Cauchy distribution. `https://distribution-explorer.github.io/continuous/cauchy.html#`.

Bürkner, P.-C. (2017). `brms`: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28.

Canadian Election Watch (2019a). Final 2019 projection. `http://cdnelectionwatch.blogspot.com/2019/10/final-2019-projection-o.html`.

Canadian Election Watch (2019b). Understanding projections i: Uniform vs. proportional swing. `http://cdnelectionwatch.blogspot.com/2015/10/understanding-projections-i-uniform-vs.html`.

Chen, M., Ebert, D., Hagen, H., Laramee, R. S., Van Liere, R., Ma, K.-L., Ribarsky, W., Scheuermann, G., and Silver, D. (2008). Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications*, 29(1):12–19.

Christensen, R., Johnson, W., Branscum, A., and Hanson, T. E. (2011). *Bayesian ideas and data analysis: an introduction for scientists and statisticians*. CRC press.

Cohen, D. (2020). Applied Bayesian statistics using STAN and R `https://www.mzes.uni-mannheim.de/socialsciencedatalab/article/applied-bayesian-statistics/#bayesian-fundamentals`.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics letters B*, 195(2):216–222.

Farrell, P. B. M. (2020). Identification of multinomial logit models in STAN. `http://eleafeit.com/post/mnl_identification_stan/`.

Feit, E. (2020). Getting started with Bayesian modelling in Stan. `http://eleafeit.com/post/mnl_identification_stan/`.

Fournier, P. J. (2018). Welcome to 338 Canada. `https://338canada.blogspot.com/2018/11/welcome-to-338canada.html#metho`.

Freeman, M. (2017). An introduction to hierarchical modeling. `http://mfviz.com/hierarchical-models/`.

Gelman, A., Lee, D., and Guo, J. (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, 40(5):530–543.

Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

Gupta, K., Sampat, S., Sharma, M., and Rajamanickam, V. (2016). Visualization of election data: Using interaction design and visual discovery for communicating complex insights. *JeDEM-eJournal of eDemocracy and Open Government*, 8(2):59–86.

Hadi, M. A., Fard, F. H., and Vrbik, I. (2020). Geo-spatial data visualization and critical metrics predictions for Canadian elections. In *2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–7. IEEE.

Hanretty, C. (2020). An introduction to multilevel regression and post-stratification for estimating constituency opinion. *Political Studies Review*, 18(4):630–645.

Hengl, T. (2003). Visualisation of uncertainty using the hsi colour model: computations with colours. In *Proceedings of the 7th International Conference on GeoComputation*. University of Southampton Southampton, United Kingdom.

Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Kastellec, J., Lax, J., and Phillips, J. (2010). *Estimating State Public Opinion With Multi-Level Regression and Poststratification using R*.

Kennedy, L. and Gelman., A. (2020). Know your population and know your model: Using model-based regression and poststratification to generalize findings beyond the observed sample. arxiv 1906.11323.

Kinkeldey, C., MacEachren, A. M., and Schiewe, J. (2014). How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4):372–386.

Kubicek, P. and Šašinka, (2011). Thematic uncertainty visualization usability—comparison of basic methods. *Annals of GIS*, 17:253–263.

Langton, S. H. and Solymosi, R. (2021). Cartograms, hexograms and regular grids: Minimising misrepresentation in spatial data visualisations. *Environment and Planning B: Urban Analytics and City Science*, 48(2):348–357.

Lauderdale, B. and Blumenau (2019). MRP methodology, tables and figures. `https://yougov.co.uk/topics/politics/articles-reports/2019/11/27/how-yougovs-2019-general-election-model-works`.

Leinster, T. (2014). Effective sample size. `https://golem.ph.utexas.edu/category/2014/12/effective_sample_size.html`.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989–2001.

Littler, S. (2018). Analysing categorical data using logistic regression models. `https://select-statistics.co.uk/blog/analysing-categorical-data-using-logistic-regression-models/`.

Louth, L. N. and Webb, P. D. (2019). National union of conservative and unionist associations, tories, tory party. *Encyclopædia Britannica*.

Lucchesi, L. and Kuhnert, P. (2017). `Vizumap`: An R package for visualizing uncertainty in spatial data. `https://lydialucchesi.github.io/Vizumap/`.

Lucchesi, L. and Wikle, C. (2017). Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation. *Stat*, 6.

MacEachren, A., Roth, R., O'Brien, J., Li, B., Swingley, D., and Gahegan, M. (2012). Visual semiotics uncertainty visualization: An empirical study. *Visualization and Computer Graphics, IEEE Transactions on*, 18:2496–2505.

MacEachren, A. M. (1992). Visualizing uncertain information. *Cartographic perspectives*, (13):10–19.

Mackinlay, J., Hanrahan, P., and Stolte, C. (2007). Show me: Automatic presentation for visual analysis. *IEEE transactions on visualization and computer graphics*, 13(6):1137–1144.

McDonnell, A. and Curtis, C. (2017). How Britain voted in the 2017 general election. `https://yougov.co.uk/topics/politics/articles-reports/2017/06/13/how-britain-voted-2017-general-election`.

Park, D. K., Gelman, A., and Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis, 375–385*.

Parliament, U. K. (2021). Voting systems in the UK. `https://www.parliament.uk/about/how/elections-and-voting/voting-systems/`.

Ramnath (2016). Using a LKJ prior in Stan. `http://stla.github.io/stlapblog/posts/StanLKJprior.html#`.

Rannala, B. (2002). Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Systematic biology*, 51(5):754–760.

Robert, C. P. and Casella, G. (1999). The metropolis–hastings algorithm. In *Monte Carlo Statistical Methods, 231–283*. Springer.

Roberts, G. O., Rosenthal, J. S., et al. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71.

Rodríguez, G. (2007). Lecture notes on generalized linear models. `https://data.princeton.edu/wws509/notes/`.

Sennhenn-Reulen, H. (2018). Bayesian regression for a Dirichlet distributed response using Stan. arxiv:1808.06399.

Silver, N. (2016). A user's guide to FiveThirtyEight's 2016 general election forecast. https://fivethirtyeight.com/features/a-users-guide-to-fivethirtyeights-2016-general-election-forecast/.

Silver, N. (2017). The media has a probability problem. https://fivethirtyeight.com/features/the-media-has-a-probability-problem/.

Skowronnek, A. (2015). Beyond choropleth maps: A review of techniques to visualize quantitative areal geodata. *Infovis Reading Group WS*, 16.

Stan Development Team (2020). Multinomial distribution. https://mc-stan.org/docs/2_26/functions-reference/multinomial-distribution.html.

Therriault, A. (2020). Bloomberg news 2020 election turnout model. https://www.bloomberg.com/graphics/2020-us-election-results/methodology.

Vasishth, S. and Gelman, A. (2019). How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. https://www.psyarxiv.com/zcf8s.

Vats, D. and Knudson, C. (2018). Revisiting the gelman-rubin diagnostic. *arXiv preprint arXiv:1812.09384*.

Wells, A. (2020). Swingometer map. http://ukpollingreport.co.uk/swingometer-map.

Wood, T. (2020). What is the softmax function? https://deepai.org/machine-learning-glossary-and-terms/softmax-layer.

Appendix A – STAN Code

Listing 1 – Categorical Model

```
data {
  int K; // number of parties
  int N; // number of observations
  int D; // number of covariates
  int<lower = 1, upper = K> y[N]; // integer for winner
  matrix[N, D] x; // matrix of covariates
}
parameters {
  matrix[D, K] beta; // regression coefficients
}
model {
  to_vector(beta) ~ normal(0, 5);
  for (i in 1:N)
    y[i] ~ categorical_logit((x[i,] * beta)'); // likelihood function
}
generated quantities {
  vector[N] winner;
  for(i in 1:N)
    winner[i]<-categorical_logit_rng((x[i,] * beta)');
}
```

Listing 2 – Multinomial Model

```
data {
  int K; // number of parties
  int N; // number of observations
  int D; // number of covariates
  int y[N, K]; // matrix of votes for K parties in constituency
  matrix[N, D] x; // matrix of covariates
}
parameters {
  matrix[K-1, D] beta_raw; // regression coefficients
}
transformed parameters{
  matrix[K,D] beta;
  for (l in 1:D) {
    beta[K,l] = 0.0;
    }
  for (k in 1:(K-1)) {
      for (l in 1:D) {
        beta[k,l] = beta_raw[k,l];
      }
  }
}
model {
  for (k in 1:(K-1)) {
    for (l in 1:D) {
    beta_raw[k,l] ~ normal(0,5);  // prior for coefficients
    }
}
  for (n in 1:N) {
    vector[K] logits;
    for (m in 1:K){
          logits[m] = x[n,] * transpose(beta[m,]);
      }
      y[n,] ~ multinomial(softmax(logits)); // likelihood function
    }
}
```

Listing 3 – Random effect model

```
data {
  int N; // number of observations
  int K; // number of parties / group for multinom
  int R; // number of ridings
  int D; // number of covariates
  int M; // number of elections years
  int y[N, K]; // matrix of votes for K parties in constituency
  matrix[N, D] x; // matrix of covariates
  int<lower=1, upper=M> Gy[N]; // integer for election year
  int<lower=1, upper=R> Gr[N]; // integer for constituency
  vector[K-1] nu;       // prior mean of random effect
}
parameters {
  matrix[K-1, D] beta;  // regression coefficients
  matrix[K-1, R-1] b_raw; // matrix of random effects
  cholesky_factor_corr[K-1] Lcorr; // Cholesky root of correlation for random effects
  vector<lower=0>[K-1] sigma; // standard deviations for random effects
}
transformed parameters{
  matrix[K-1, R] b;
  for(k in 1:(K-1)) { // sum-to-zero constraint
    b[k,] = append_col(b_raw[k,], -sum(b_raw[k,]));
  }
}
model {
  sigma ~ cauchy(0,20);        // prior for standard dev of b
  Lcorr ~ lkj_corr_cholesky(1); // prior for correlation matrix of b
  for(r in 1:(R-1)){
    b_raw[,r] ~ multi_normal_cholesky(nu, diag_pre_multiply(sigma, Lcorr));
  }
  //nu ~ normal(0,1);
  to_vector(beta) ~ normal(0,5);
  vector[K] logits;
  for (n in 1:N) {
    logits[K] = 0;
    for (k in 1:(K-1)){
        logits[k] = x[n,] * transpose(beta[k,]) + b[k, Gr[n]];
    }
    y[n,] ~ multinomial_logit(logits);
  }
```

```
}
```

Listing 4 – Linear model

```
data {
 int < lower = 1 > N; // Sample size
 vector[N] x; // Predictor
 vector[N] y; // Outcome
}

parameters {
 real alpha; // Intercept
 real beta; // Slope (regression coefficients)
 real < lower = 0 > sigma; // Error SD
}

model {
 y ~ normal(alpha + x * beta , sigma);
}
```