

Changing Demographics in the UK Labour Party's Vote

Author: Toby Staines

1 Introduction

1.1 Domain Overview and Motivation

The past three years have provided a succession of electoral shocks, both in the UK, and around the world. Most recent of these was the 2017 UK General Election. The election campaign began with The Conservative Party enjoying a 20% opinion poll lead[1], but ended with a very close result (Conservatives 42.4%, Labour 40%[2]). Although the Conservatives remained in government, the result was widely recognised as a success for Labour[3][4], relative to their expected performance.

Much of Labour's success has been credited to an increase in support from young people and the well-educated middle class, as opposed to Labour's traditional white, working class base[4]— a group whose declining support, especially in the wake of Jeremy Corbyn's election as Labour leader and the 2016 EU referendum result, was expected to be a cause of their failure[5]. The extent to which this is the case has significant implications for the future direction of both major UK political parties.

1.2 Aims

This report aims to investigate the relationship between the demographic makeup of parliamentary constituencies and the way they voted in 2017, focusing on the surprise success of The Labour Party. Specifically:

1. To what extent is the improvement in Labour's electoral performance related to the demographic makeup of the electorate?
2. Is this relationship the same across the country?
3. How much does this represent a change in the social makeup of Labour's support?

The EU referendum results were published at Local Authority level, making detailed comparison between the outcome of these two elections impossible, within the scope of this project. However, it will be informative to see if the trends identified correlate with previous examinations of the referendum result.

1.3 Data

Electoral data for the 2015 and 2017 UK general elections is provided by the UK Parliament[6][7], giving the number of votes cast for each party in each of the 632 constituencies in Britain. Northern Ireland is excluded from this investigation as Labour do not stand candidates there. There were no changes to the definitions of UK parliamentary constituencies between the 2015 and 2017 elections, making comparison of results straightforward.

The results of the 2011 UK census are used to understand the demographic nature of each constituency. These results are provided by The Office for National Statistics at Output Area level[8], as are mappings of Output Areas to parliamentary constituencies[9][10][11], enabling comparison of the census with election results.

2 Approach

2.1 Electoral Measures

The report uses two measures to assess the 2017 election result: swing from Conservative to Labour, and Labour's baseline support.

2.1.1 Swing

Analysing voting in UK general elections is more complicated than in referenda, due to their non-binary nature. The number of parties involved means there is not a single, continuous measure which describes the result. To constrain the scope of the report, it will look only at the change in vote share of the two major parties. This is a simplification, especially in areas where these parties are commonly not the two largest (e.g. Scotland and the South West of England). However, a reasonable method for measuring the success of the opposition is to compare them to the governing party, supported by the fact that the 2017 election was the most binary since 1970, with these two parties accounting for over 80% of the vote[7].

Swing is commonly used as a measure of change between elections; in this case the swing from Conservative to Labour is used as the measure of Labour's success in 2017. Swing is calculated by:

$$Swing = \frac{C_{Labour} - C_{Conservative}}{2}$$

Where:

$$C = 2017 \text{ percentage vote share} - 2015 \text{ percentage vote share}$$

2.1.2 Labour's Baseline

To understand the social makeup of Labour's support prior to the recent run of unexpected election results, census data is compared to the 2015 General Election result. As a measure of Labour's 'traditional' support base this is again a simplification, as 2015 was seen as a below par result for Labour, but it is the best baseline available to understand changes in voting behaviour over the past two years.

2.2 Analytical Tasks

1. Understand the geographic distribution of each electoral measure:
This will be done by visualising the measures for each constituency on a hexagon map of the UK, and by plotting the measures against each other on a scatter plot. The use of a hexagon map gives a clear view of each constituency, which is not the case on a standard choropleth map, and also scales them according to their political importance, each being represented by a single MP, regardless of size or population. This approach will allow a visual comparison of where Labour's existing vote is based, against where it has improved, giving an overall understanding of the results, and helping to answer questions 2 and 3.
2. Identify demographic variables which explain the observed election results, helping to answer question 1:
 - a) Examine the correlation between demographic variables and electoral measures, by means of a plot of each variable against each measure. A visual assessment will indicate possible demographic drivers of the measures, which will then be used as inputs to the models produced in later steps. Colouring the points by region may also point towards regional variations from national trends.

- b) Examine the correlation between demographic variables. Several of the census variables are likely to be related to each other. Identification of these relationships should allow some to be removed from consideration, leading to simpler, more explainable models. Correlation will be computed and then displayed in a matrix, allowing easy interpretation of the results.
3. Model the relationship between explanatory variables and results measures, and assess their geographic variation:
 The output from tasks 2a and 2b will dictate the selection of explanatory variables for each measure. Models will be built on these variables which compute an expected result for each constituency. The residuals between these results and the true value will then be mapped in a similar way to task 1, allowing a visual interpretation of how well these models explain the results across the country, helping to answer question 2, and how much each measure is affected by different variables, helping to answer question 3.
4. Investigate regional variation using geographically weighted models:
 The models created in task 3 assume uniform correlation between variables and measures across the country. The use of geographically weighted models will allow the exploration of how the correlations identified in task 2 vary over space.

3 Analytical Steps

1. Figure 2 displays the results of task 1, showing that there are distinct regional variations in both electoral measures, and that they do not seem to follow the same pattern, suggesting both that the factors affecting the measures may be different and that they may be different in different parts of the country.

Figure 1 shows that there is only a weak correlation between the two electoral measures. It also highlights some distinct regional patterns which are reflected in Figure 2. Regions are based on European Parliament regions. The North East has been combined with Yorkshire and The Humber as it had only 29 constituencies, they are adjacent, and the reduction in the number of regions makes visual interpretation of plots easier.

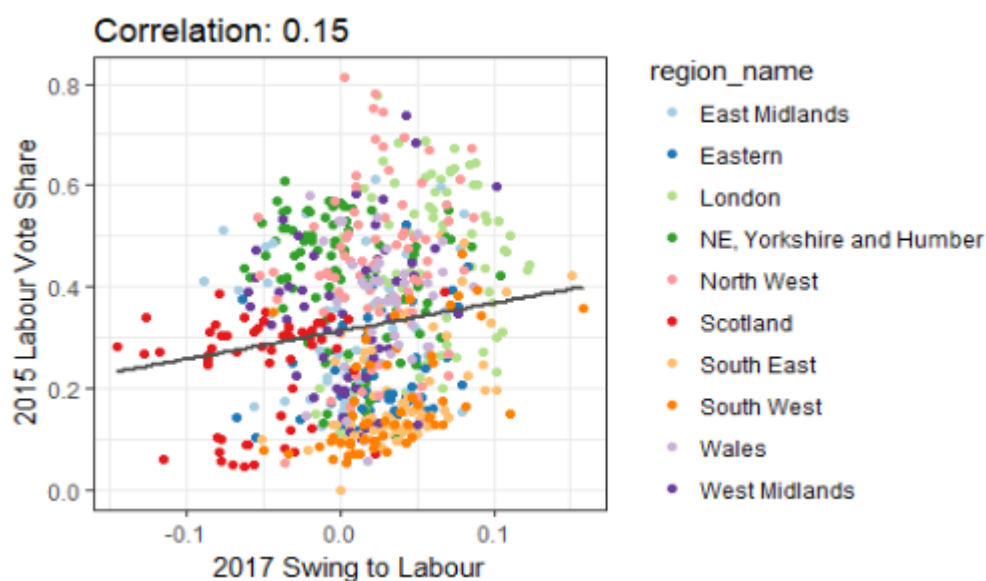


Figure 1: 2017 Swing to Labour vs 2015 Labour Vote Share. Only a weak correlation is observed between the two measures.

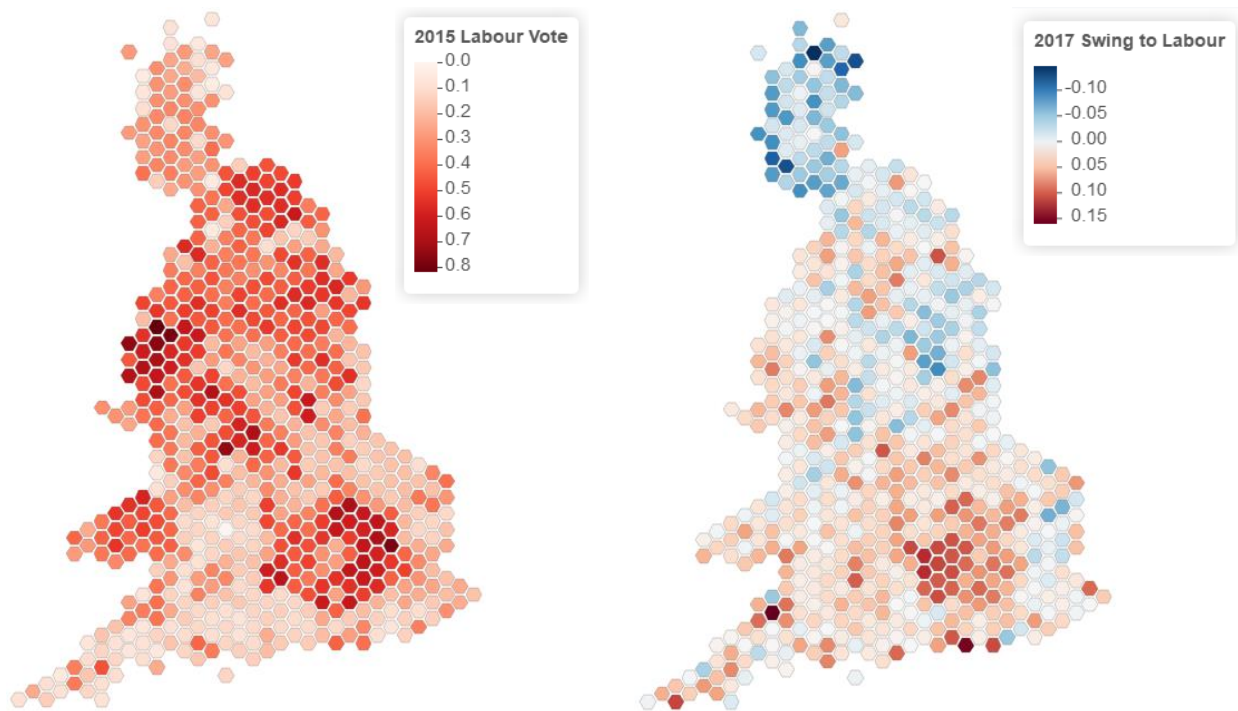


Figure 2: Left: The Labour Party's share of the vote in the 2015 General Election. Right: Swing from Conservative to Labour in the 2017 General Election.

Scotland shows moderate or weak support for Labour in 2015 and a swing towards Conservative; the South West shows low support for Labour in 2015, but a swing towards them; London and the North West show strong support for Labour in 2015 and further swing in that direction, and the North East, Yorkshire and The Humber shows strong support for Labour in 2015 but a swing towards Conservative in 2017.

2. Census variables were based on those used previously by Beecham et al.[12], with a slightly different focus. Variables were chosen which might be expected to identify the groups or areas generally accepted to support Labour – young, educated middle class, working class, and areas of higher social diversity, as described in Table 1. As much has been made of young voter turnout assisting Labour in 2017 the new variable 'Young Adults' was added, representing the proportion of the population aged 20 to 24, in addition to 'Younger Adults', representing the proportion aged 20 to 44. Proportion of people living in social housing is

Census Variable	Explanation
Young Adults	Young, educated, middle-class
Younger Adults	
Degree Educated Professionals	
White	Social diversity (lack of)
Christian	
English Speaking	
Single Ethnicity Household	
Social Housing	Poor or working class
No Car	
Not Good Health	

Table 1: Census variables selected for analysis and the group or area characteristic which they are expected to indicate.

used rather than home ownership, as this is seen as a more accurate indicator of a working-class area. Residency and car ownership are seen in their more traditional context as wealth indicators rather than urban/rural area indicators.

- a) Figures 4 and 5 illustrate the relationships between the electoral measures and 11 demographic variables derived from the 2011 census results. Comparison shows that some variables have a similar relationship to both electoral measures: those representing youth (Young Adults and Younger Adults) have a positive correlation, and those representing 'traditional' Britain (or a lack of social diversity) all have a negative correlation. However, Degree Educated and Professionals both show a slight negative correlation with Labour's baseline, but a definite positive correlation with swing. The poor or working-class indicators all show a strong positive correlation with the baseline, but little relationship with swing. Also notable in Figure 5, constituencies in Scotland consistently show a lower swing to Labour than the rest of the country, across all variables.
- b) Figure 3 shows the correlation between each of the census variables. Several pairs or groups of variables can be identified which are in the same conceptual category and are also strongly correlated:
- Young Adults and Younger Adults
 - Degree Educated and Professionals
 - Social Housing and No Car
 - White, Christian, English Speaking and Single Ethnicity Household.



Figure 3: Correlation matrix of demographic variables and electoral measures.

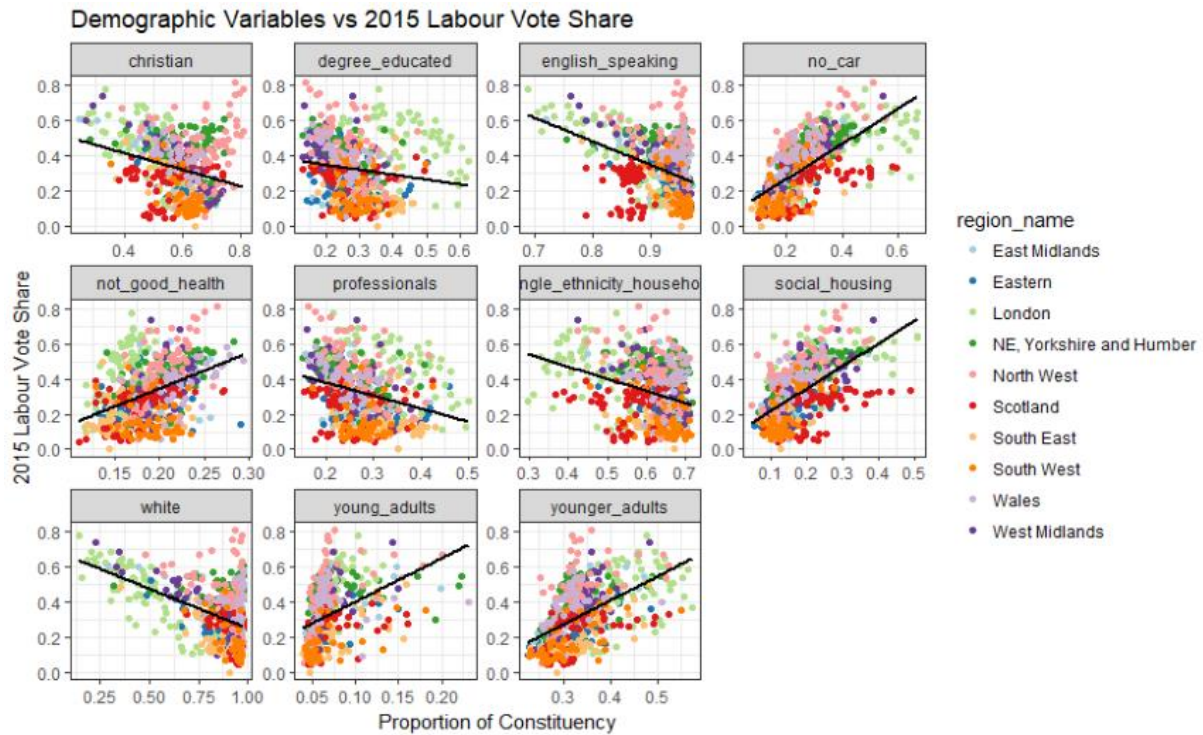


Figure 4: Demographic variables vs 2015 Labour vote share.

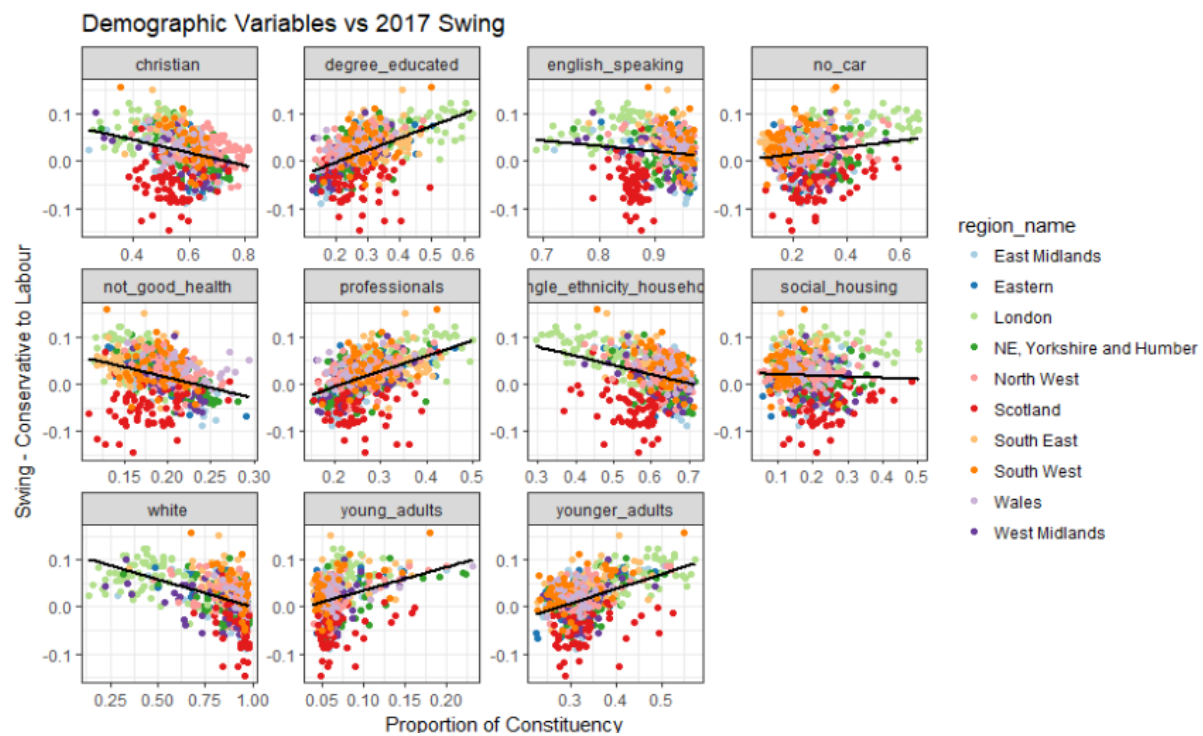


Figure 5: Demographic variables vs 2017 swing towards Labour.

As Young Adults and Younger Adults are definitively linked, both being derived from the same census question, and no significant difference was identified between them, Young Adults was discarded at this point. Younger Adults had shown a slightly stronger correlation to both measures.

- Results from 1 and 2 suggested that different models were required to explain each measure. Figure 6 shows the residuals from univariate global regression models based on

the strongest explanatory variable for each measure. For baseline support this is No Car, which explains 49% of the variation in results (using Adjusted R^2). For Swing in 2017 this is Degree Educated, although this explains only 28% of the variation. Residuals have been normalised to a scale of $[-1, 1]$ to assist with comparison, as the absolute values of swing tend to be smaller than those of vote share.

Much like previous analysis of the EU referendum[12], both models show a similar variation in Scotland, where they over-predict Labour success, indicating different factors influencing voting behaviour in Scotland relative to the rest of the country in both 2015 and 2017.

The No Car model under-predicts success in the North of England and most of London and over-predicts it in West London and the rest of the South, reflecting well the accepted understanding of Labour's traditional base. The Degree Educated model shows a reasonably accurate uniform result across England and Wales, supporting the idea that the increase in Labour's support in 2017 is due to different groups than those that make up their historic support. Removal of Scotland from the models increased Adjusted R^2 results (from 0.48 to 0.55, and from 0.28 to 0.38 respectively).

In an attempt to increase the explanatory power of the models, further variables were added, based on the results of task 2. The first multivariate model used the explanatory variables for 2015 support identified by task 2a: Younger Adults, White, English Speaking, Not Good Health, No Car, and Social Housing. This model had an Adjusted R^2 of 0.67, and analysis of Variance Inflation Factors (VIF) showed that both Younger Adults and No Car had VIFs greater than five. Although 10 is often accepted as a threshold for VIF [citation], some strong collinearity had been identified in Figure 3, so the model was simplified, first with the removal of Social Housing, and then by replacing Social Housing and removing No Car. Although No Car was the strongest individual indicator it was strongly correlated with both Social Housing and Younger Adults. Its removal reduced the VIF for the remaining variables to below five. All three attempted models had an Adjusted R^2 of 0.67, meaning they describe the variation in the results equally well.

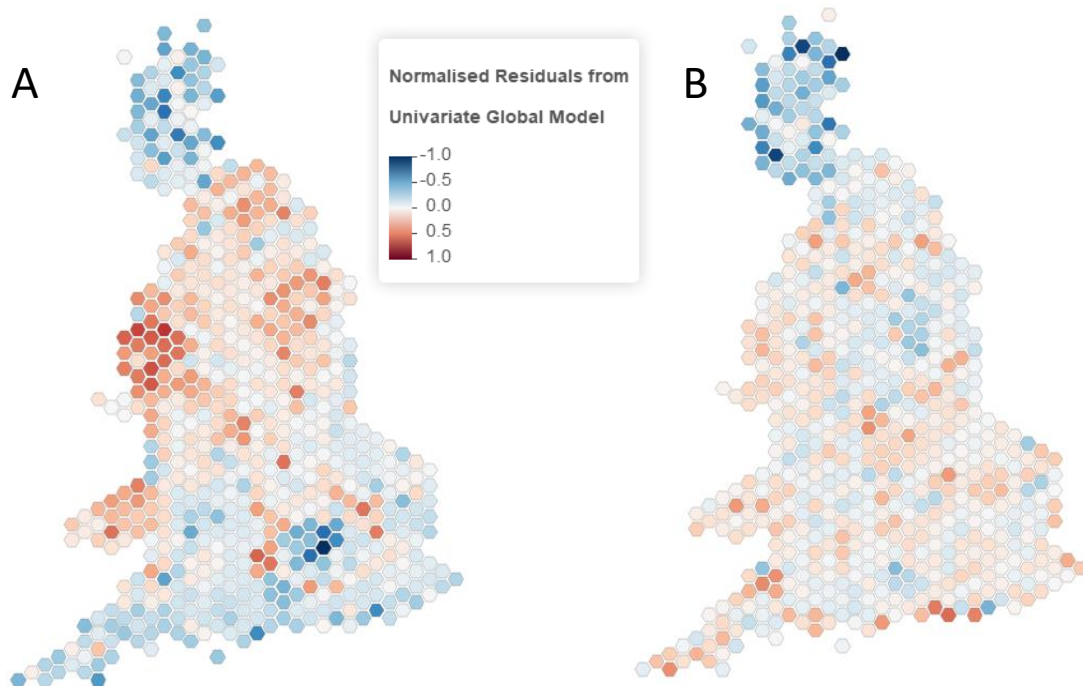


Figure 6: Residuals from global univariate regression models, normalised to a range of $[-1, 1]$. A) No Car vs 2015 Labour vote share. B) Degree Educated vs 2017 swing towards Labour

To model swing, four variables were selected: Younger Adults, Degree Educated, Professionals, and White. Degree Educated and Professionals were known to be strongly colinear, and showed VIFs of 15.2 and 12.7 respectively. Since Degree Educated has the stronger relationship to swing, Professionals was removed for the next iteration. The Adjusted R^2 was reduced from 0.38 to 0.35, but the VIFs for the remaining variables were all less than 2.5.

Both national multivariate models show a similar geographic pattern to their equivalent univariate models, but with a slightly increased degree of accuracy (improving from 0.48 to 0.67 for 2015 and from 0.28 to 0.35 for 2017 swing). Removing Scotland again increased Adjusted R^2 , from 0.67 to 0.7, and 0.35 to 0.46 respectively.

4. Use of the gwss package in R produced the maps shown in Figure 7, highlighting how the correlation between census variables and election measures varies across Britain, showing some interesting results, which were not easily observable in earlier stages of the analysis. A bandwidth of 50 was chosen, following experimentation with larger and smaller values. 25 was too granular and 75 produced results so smoothed as to provide little insight.

4 Summary of Findings

4.1 To what extent is the improvement in Labour's electoral performance related to the demographic makeup of the electorate?

The Labour Party's improved results in the 2017 General Election, as measured by swing from Conservative to Labour, do show correlation to the demographic makeup of parliamentary

Geographically Weighted Correlation to 2015 Vote Share and 2017 Swing to Labour

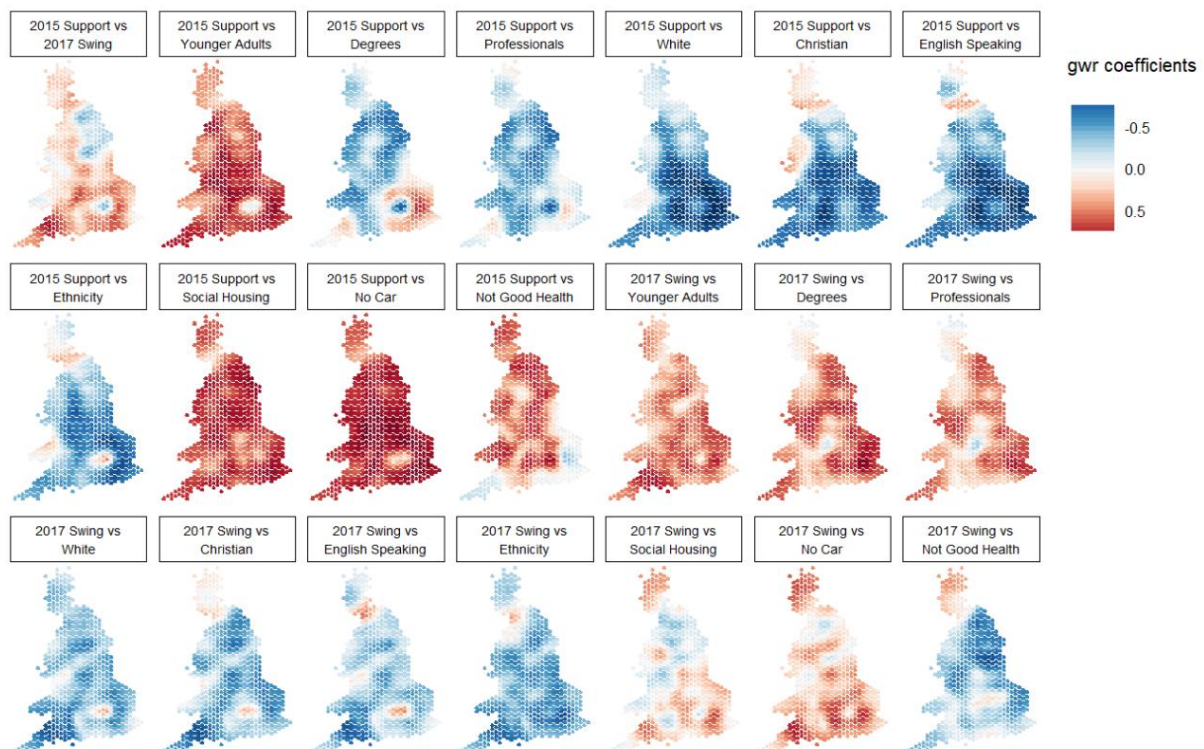


Figure 7: The first map shows the geographically weighted correlation between 2015 Labour vote share and 2017 swing towards Labour. Each subsequent map shows the geographically weighted correlation between one demographic variable and one election measure.

constituencies, although not as strongly as does their baseline support, as measured by their vote share in the 2015 General Election.

The most significant indicators for the strength of this swing are those representing the young, educated, middle-class (proportion of residents aged 20 to 44, with degree level qualifications, or in professional level employment), individually explaining 20%, 28% and 23% of variation in results. This is in contrast to the strongest indicators for 2015 vote share, which were those linked to poorer or working-class communities (proportion of people living in social housing, not owning a car, or with poor health (36%, 49% and 17%)).

Social diversity was the constant across both 2015 results and 2017 swing, with a negative correlation in both cases, though this category was not the most significant in either case.

4.2 Is this relationship the same across the country?

Residuals from global models showed, in particular, that there is a difference between Scotland and the rest of Great Britain. Geographically weighted models showed that indicators of baseline support tend to be fairly global in nature (with the exception in some cases of Scotland); correlation is generally in the same direction and with similar strength across England and Wales, particularly diversity and working-class indicators.

Correlation with swing is generally weaker, and shows greater regional variation for several measures. However, the strongest indicators do show a global pattern.

4.3 How much does this represent a change in the social makeup of Labour's support?

At a national level, none of the indicators showed a switch from positive to negative correlation from 2015 result to 2017 swing, indicating that Labour did not lose significant support among groups which already supported it. The switch from negative to positive correlation for Degree Educated and Professionals could indicate a change in the demographic makeup of Labour's support, but it is difficult to give a quantitative response to question 3. Further analysis of the relative volume of votes contributed by these indicators and examination of results over a longer period would help to answer this question in more detail but were beyond of the scope of this analysis.

5 Reflection

5.1 Effectiveness and Validity of Approach

One must be careful, with this type of analysis, not to attempt to draw conclusions about individual voting behaviour. While we observe relationships at the constituency level, we do not know how the groups within each constituency voted, or what their relative turnout was, and we cannot say that a certain group voted in a certain way. The analysis also does not attempt to capture specific local issues which may have affected voting behaviour, and cannot account for demographic shifts which may have occurred since the 2011 census. Such connections might be more easily made with an opinion poll, where there is a direct link between a person's demographic, views, and vote. However, these are expensive, and rarely conducted on a scale sufficient to uncover the regional variation which this report investigates. The strength of this analysis is that it includes every person and every vote, so its scope is much broader than opinion poll analysis.

The choice of electoral measures was reasonable, and provided useful results, but, as highlighted in section 2.1, the measures are limited. Scotland was noted as an area where swing between Conservative and Labour may be too simplistic, because neither was the largest party there.

Scotland also held its own referendum on membership of the UK in 2014; a very significant and specific political event. Sure enough, Scotland did exhibit different patterns from the rest of the country, and a specific Scottish analysis including the Scottish National Party would likely provide a better understanding of these relationships.

The choice of hexagon maps for display of geographic relationships worked well, allowing clear visual interpretation of results, although the geographic distortion means that the identity of specific locations is not always immediately obvious. Their use with the global regression models was particularly helpful in answering question 2.

Hexagon maps were a good compliment to scatter plots, which gave a high-level view of correlation, but were difficult to interpret in detail due to the large number of data points and the relatively high number of regions, making it hard to distinguish patterns in the crowded centre of plots. Further analysis of individual regions would be interesting, but was not feasible as part of this work.

The geographically weighted models for correlation showed interesting regional variations, but understanding how these differing correlations combine into an overall picture is not straightforward, and it is especially difficult to quantify how they might affect the two separate election measures in combination. Further work using partition based clustering may shed some light on this and assist with arriving at a more quantifiable answer to question 3.

5.2 Implications for the Domain

In the multi-party, first past the post, electoral system of the UK, understanding how voting patterns are affected at local level is important. Leading up to elections there is much discussion of national polls, and how a change in national vote share might be reflected in parliamentary seats. This inference is notoriously difficult, and as several recent elections have shown, relatively small opinion polls do not always generalise well across the country. As such, an understanding of how a party performed against demographic and regional variables could be very valuable. However, the analysis is inherently reactive; it may provide insight into past results, but that does not imply that it could be used to make predictions about future elections, as the gap between the sampling and the future event is too great.

The results showed a similar pattern of correlation for the Degree Educated variable with swing towards Labour as was noted[12] for voting Remain in the 2016 EU referendum, although the relationship was not as strong. This supports the view that part of Labour's success was a reaction to the referendum result, with dissatisfied professional Conservative, Remain voters switching to Labour, whilst the predicted opposite response amongst working-class Leave voting Labour supporters did not materialise. However, it has also been argued that both results are a reflection of a more fundamental underlying change in voter behaviour[13].

If Labour can maintain this pattern of widening their support base to new demographics whilst maintaining existing support it will obviously be beneficial for them. Analysis of current electoral margins at a constituency level, against demographic indicators might be an interesting avenue to explore, in order to understand how much of an impact continuation of these patterns could have in future.

5.3 Generalisability of Approach

The techniques used in this analysis are highly generalisable to other domains where complex spatial multivariate data is common. Beyond politics, similar approaches could be applied to investigating spatial variation in the answers to a wide variety of social questions, for example comparisons of any of wealth or income equality, health, educational outcomes, or environmental change.

6 Bibliography

- [1] J. Holder, N. Kommenda, C. Barr, and S. Clarke, "How did Theresa May's election gamble fail? | Politics | The Guardian." [Online]. Available: <https://www.theguardian.com/politics/ng-interactive/2017/jun/09/theresa-may-election-gamble-fail-conservatives-majority-polls>. [Accessed: 01-Dec-2017].
- [2] "UK election 2017: full results | Politics | The Guardian." [Online]. Available: <https://www.theguardian.com/politics/ng-interactive/2017/jun/08/live-uk-election-results-in-full-2017>. [Accessed: 01-Dec-2017].
- [3] P. Walker and J. Elgot, "Corbyn defies doubters as Labour gains seats | Politics | The Guardian." [Online]. Available: <https://www.theguardian.com/politics/2017/jun/09/jeremy-corbyn-labour-defies-doubters-gain-seats-election-2017>. [Accessed: 31-Dec-2017].
- [4] H. Lewis, "What drove Labour's success? A tough line on immigration, and an appeal to the middle class." [Online]. Available: <https://www.newstatesman.com/politics/june2017/2017/06/what-drove-labours-success-tough-line-immigration-and-appeal-middle-class>. [Accessed: 31-Dec-2017].
- [5] K. McCann, "Labour now third most popular party among working class voters, poll reveals." [Online]. Available: <http://www.telegraph.co.uk/news/2017/02/13/labour-now-third-popular-party-among-working-class-voters-poll/>. [Accessed: 31-Dec-2017].
- [6] UK Parliament, "General Election 2015 Results - data.parliament.uk." [Online]. Available: <http://www.data.parliament.uk/dataset/general-election-2015>. [Accessed: 13-Dec-2017].
- [7] C. Baker *et al.*, "General Election 2017: full results and analysis." [Online]. Available: <http://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-7979>. [Accessed: 13-Dec-2017].
- [8] O. for N. S. UK Government, "Bulk data downloads - Office for National Statistics." [Online]. Available: <https://www.ons.gov.uk/census/2011census/2011censusdata/bulkdata/bulkdatadownloads>. [Accessed: 01-Dec-2017].
- [9] O. for N. S. UK Government, "Output Area to Ward to Local Authority District (December 2016) Lookup in England and Wales | Open Geography portal," 2016. [Online]. Available: <http://geoportal.statistics.gov.uk/datasets/output-area-to-ward-to-local-authority-district-december-2016-lookup-in-england-and-wales>. [Accessed: 11-Dec-2017].
- [10] O. for N. S. UK Government, "Ward to Westminster Parliamentary Constituency to Local Authority District (December 2016) Lookup in the United Kingdom | Open Geography portal." [Online]. Available: http://geoportal1-ons.opendata.arcgis.com/datasets/f52c48ea5cf2494b824d35f4be88f2ec_0/data. [Accessed: 11-Dec-2017].
- [11] The Scottish Government, "Data Zone and Intermediate Zone 2011 Lookups." [Online]. Available: <http://www.gov.scot/Topics/Statistics/sns/SNSRef/DZ2011Lookups>. [Accessed: 13-Dec-2017].
- [12] R. Beecham, A. Slingsby, and B. Chris, "Locally-varying explanations behind the United Kingdom's vote to leave the European Union," 2016.
- [13] W. Jennings and G. Stoker, "Tilting Towards the Cosmopolitan Axis? Political Change in England and the 2017 General Election," *Polit. Q.*, vol. 88, no. 3, pp. 359–369, 2017.

