Modelling age-related patterns in UK electoral data 2015-2019.

Eleanor Bill

Enfranchising 16-17-year olds has long been a debate in modern politics. They are allowed to vote in all Scottish elections and are demanding a voice in response to concerns about the effect of the climate crisis on their future, and stasis from politicians. What effect would enfranchising 16-17-year olds have had on the recent 2019 general election? Would it have swung the vote much either way? In this analysis I explore age-related patterns in UK general election data using visual analytics and model the hypothetical effect of introducing these new voters to the voting pool. I found that enfranchising 16-17-year olds could have affected the election – the results of the modelling I produced saw the Conservative majority reduced by 22 seats.

1 Problem Statement

This analysis project looks at electoral data from the 2015, 2017 and 2019 general elections, alongside 2016 EU referendum data and 2019 EU election data. The datasets I have sourced give a standardised constituency-level view of election data and demographic information associated with those constituencies. After initially reviewing the data, I decided to re-model the results of the general election in 2019 based on also allowing those aged 16-17 to vote and visualise the results. In terms of research questions, I am looking into whether allowing 16-17-year olds to vote in UK general elections would have any significant effect on results, particularly in the case of the most recent 2019 general election.

2 State of the Art

There are many relevant studies on visual analytics being used in electoral data analysis. Badawood and Wood (2012) [1] used visual analytics to investigate the effect of candidate position on ballot papers in the London 2010 local council elections. They used clustered bar chart and spatial tree maps (HiDE) to analyse the order of placement of names of candidates and whether the name positioning acts to bias voters towards those whose names are at the top of the ballot paper. They rejected the null hypothesis of no correlation and found that in more marginal boroughs the effect is stronger- significant enough to result in a change of distribution of elected councillors and in boroughs with large number of marginal wards, likely to have an impact on party political balance of power.

Stoffel et. al. (2012) [2] visualised the results of political elections in a similar way as to this analysis project. They specifically focused on voting patterns for non-majority party preference (who came second in elections). This they hoped would address issues with visualising election results in a meaningful way across parties, elections and location. They used bi-polar colourmaps to represent both majority and minority party – in each constituency the majority party was represented by a constituency-shaped polygon in the majority party colour, and the minority party (who got the second greatest number of votes) was represented by a second, smaller polygon within the first. The scaling amount of the secondary polygon was inversely proportional to the majority of the majority party. They hypothesised that it was a good way to visualise campaign tactics- to get out canvassers in the places where majority is slimmer. It also enables us to see outliers more clearly. I would have liked to have done something similar in this project but lacked the technical ability and tools to do so.

3 Properties of the Data

The 2015 GE and 2016 referendum results were sourced from a Kaggle dataset, which was in turn scraped from two Wikipedia pages. These results are at constituency level. The results are identifiable by constituency and region, and contain number of votes per political party, as well as a record of who won the seat, who came second, majority, turnout, and Remain/Leave percentage in 2016.

The 2019 European Parliament election results data was sourced from Chris Hanretty, a politics professor at Royal Holloway. As EU election results are gathered at local authority level instead of constituency level, the constituency level results are an estimate only.

The age demographic data I sourced from Lord Ashcroft's 2019 General Election survey, which was based on a sample of 13,000 people.

I also used another couple of datasets as lookups between constituency and local authority codes, which at the fields across which the datasets are linked.

The separate datasets are joined on constituency code.

All together these data represent five different elections, 604 different constituencies (650 in the UK, missing 18 in NI), the voting patterns of the entirety of the UK aside from Northern Ireland (as __ didn't include NI in the dataset). The data contains vote share and turnout figures and demographic data on a constituency level.

The quality of the data was generally comprehensive aside from the inclusion of Northern Ireland figures in the constituency level data- sometimes it was included and sometimes it wasn't. I also had to clean some commas formatted into the numeric data and replace some valid nulls with 0s. In my modelling I had to use linear regression to estimate voting patterns of 16-17-year olds based on trends in the data.

After initial exploratory analysis I decided to look specifically into age related data based on Lord Ashcroft's preliminary analysis of the 2019 election results (Fig. 1).

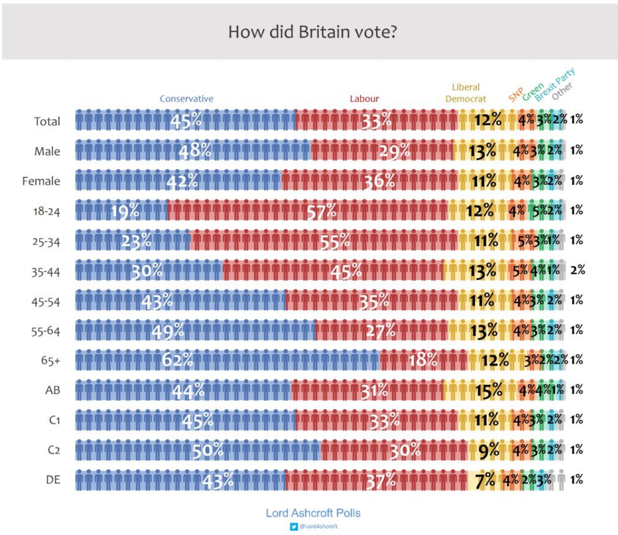

Fig. 1. 2019 general election demographic voting patterns [3].

This visual analysis and some of the conclusions I have seen drawn about the election initially suggest that the vote is deeply polarised across age lines (and equally or less so across gender and race lines), with the younger voters more likely to vote Labour, and the older generations more likely to vote Conservative. I would like to further validate this and look at it in terms of the 2016 referendum vote.

4 Analysis

4.1 Approach

My approach to conducting this analysis is first to bring all the datasets together into one large dataset on a constituency level.

I plan to pre-process the data – looking for outliers, nulls and other gaps in the data. I will use some visualisation in this step – attempting to visualise outliers, nulls etc. to see where I need to edit or narrow down the data, possibly impute values.

I will then do some exploratory analysis and data visualisation with the dataset. At first, I wasn't entirely sure what to base this project around aside from election data, so looking further into the data allowed me to narrow down the focus based on my perception on what looked interesting or important to look into. This is a vital part of the analysis process, which cannot be conducted by a computer and requires human input.

After deciding to focus on age-related voting patterns I narrowed down the data and performed more preliminary data analysis and visualisation, which I will describe more in the process section. After exploring the data in this way, I used modelling to predict voting patterns of 16-17-year olds and visualised the hypothetical, revised election results using another choropleth map. This visualisation enabled me to draw conclusions, ultimately, about the potential effect of enfranchising 16-17-year olds in future elections. The diagram below maps out my analysis approach.

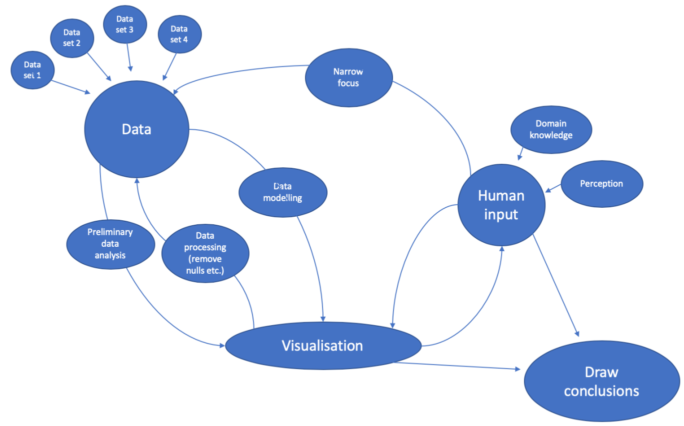

Fig. 2. A diagram of my analysis process/approach.

4.2 Process

I first brought all my datasets together using joins, across constituency code inner joins. I had to do some processing of the age dataset to bin the ages together (they had individual constituency level population numbers by age which had to be aggregated) before doing this. I also had to feature engineer these ages as percentages to compare them against other populations in the region.

As part of my exploratory data analysis I generated choropleth maps for each election (Fig. 4). I wanted to see if there were any voting patterns immediately apparent when visualising the election results in this way. Aside from the region differences between Scotland and England/Wales, the most notable factor visually is the contrast in voting patterns between cities (London, Manchester, Sheffield, Liverpool, Newcastle) and more rural areas.

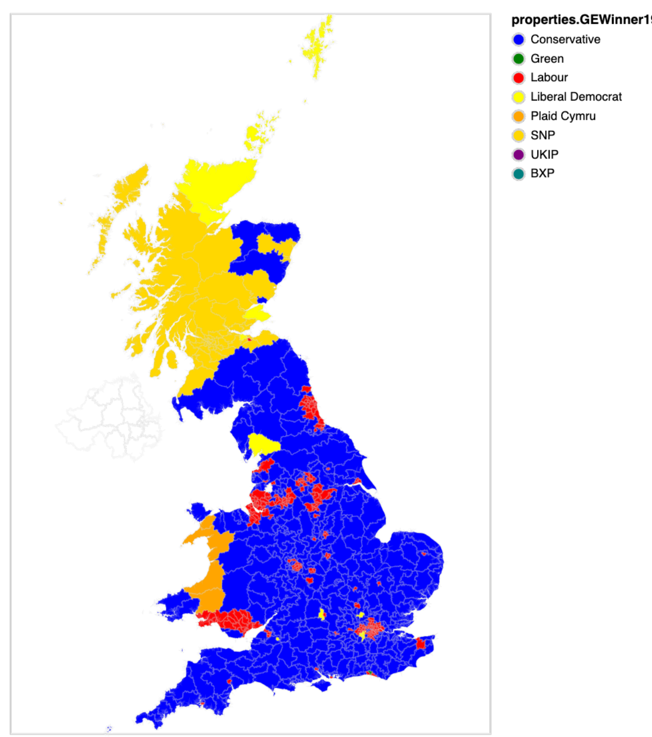

Fig. 3. Choropleth visualisation of the 2019 general election by party.

To follow, I mapped different demographic aspects across England and Wales, including age demographics (Fig. 4). Visually, I can see that higher percentages of younger people positively correlate with more Labour areas (most notably cities), while higher populations of over 65s negatively correlate with Labour-voting areas (most notably in Cornwall and the North-East.

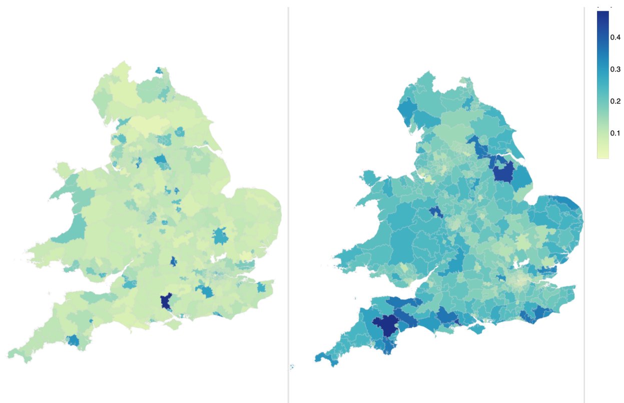

Fig. 4. Choropleth maps show age group density across regions – on the left 18-24, on the right 65+.

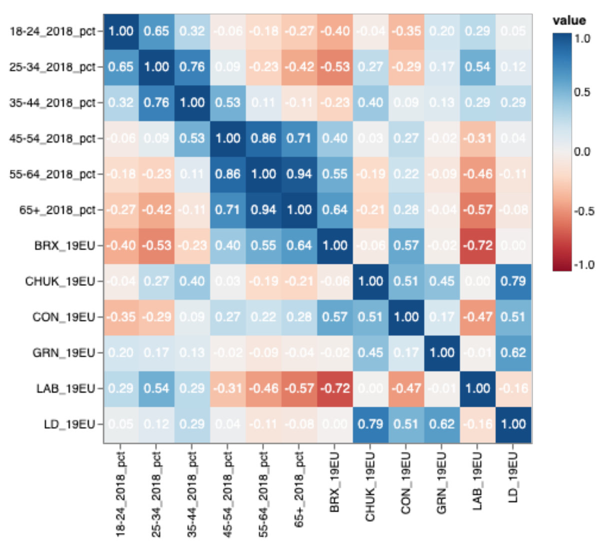

Fig. 5. Correlation heat map displays correlations between age groups and voting patterns in the 2019 EU elections.

I looked into these correlations using the 2019 EU election results, which show constituencies with a greater population of younger voters were more likely to vote Labour overall and to a lesser extent Green, while constituencies more heavily populated with older voters were more likely to vote Conservative or for the Brexit Party. These conclusions were drawn from a correlation heat map (Fig. 5), in particular the bottom left corner of this heat map which shows clear opposing patterns in terms of which age demographics correlate with which types of parties (left or right wing).

This pattern holds up across elections - I generated scatter plots based on demographic voting records for age groups across elections. The 18-24 group varies significantly from the 65+ group in terms of voting preference in the 2016 EU referendum, with older voters more likely to have backed

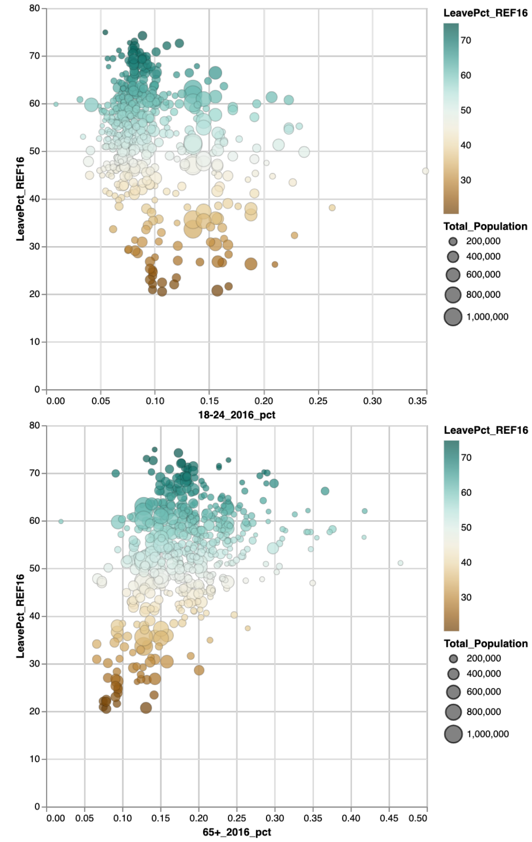

Fig. 6. 18-24 correlation with the Leave vote percentage (left) vs. 65+ (right).

Leave. The scatter plots between these two ages show gradually shifting voting patterns.

By 2050 it is estimated that one in four people will be 65 or over (up from one in five in 2018) [4]. Taking the age to voting preference correlation into account, this could impact the UK elections increasingly in the coming years. To counteract this, could enfranchising 16-17-year olds help restore balance? Can this be explored using visual analytics?

To model this scenario, I used linear regression and existing demographic voting data (from Lord Ashcroft's voter surveying) to predict how 16-17-year olds would vote, and also how they would turn-out in an election. I visualised these using line graphs, with lines of best fit based on linear regression. These were able to predict voting likelihood for 16-17-year olds for each party (Fig. 8) and overall turnout. Figure 8 clearly shows the opposing patterns in voting between the Conservative and Labour parties.

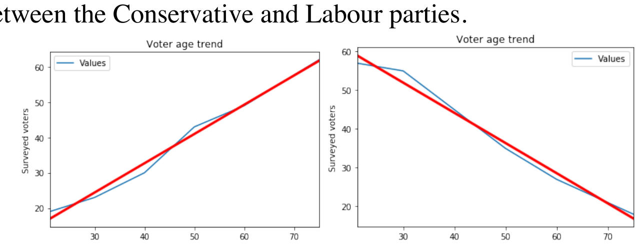

Fig. 8. Modelling voter age trends for the Conservative (left) and Labour (right) parties.

I applied these figures to the existing data and updated the voting numbers, to give each constituency a revised winner. The large, tabular nature of these results (considering the number of constituencies) meant it was hard for a human to perceive the change and effect. To improve my ability to understand the results of my modelling I mapped the revised results using a choropleth graph to compare directly with the actual results (Fig. 9). There is a greater area of red on the map, specifically in the North East and between Manchester and Sheffield – more Labour constituencies. There is also a small band of red in the South West, but no visible effect on Wales or the South East. Although these changes are visible, they are not wholly well-represented, improvements could be made to this visual analysis by using an election map that represents constituencies using equal-sized hexagons, instead of counting upon their actual size. This is becoming more common in political-related visual analysis, especially in countries such as the UK and USA, where a combination of population density and size of constituency (or state) mean direct geographical mapping of election results doesn't allow the human viewer to correctly perceive proportional representation of voters.

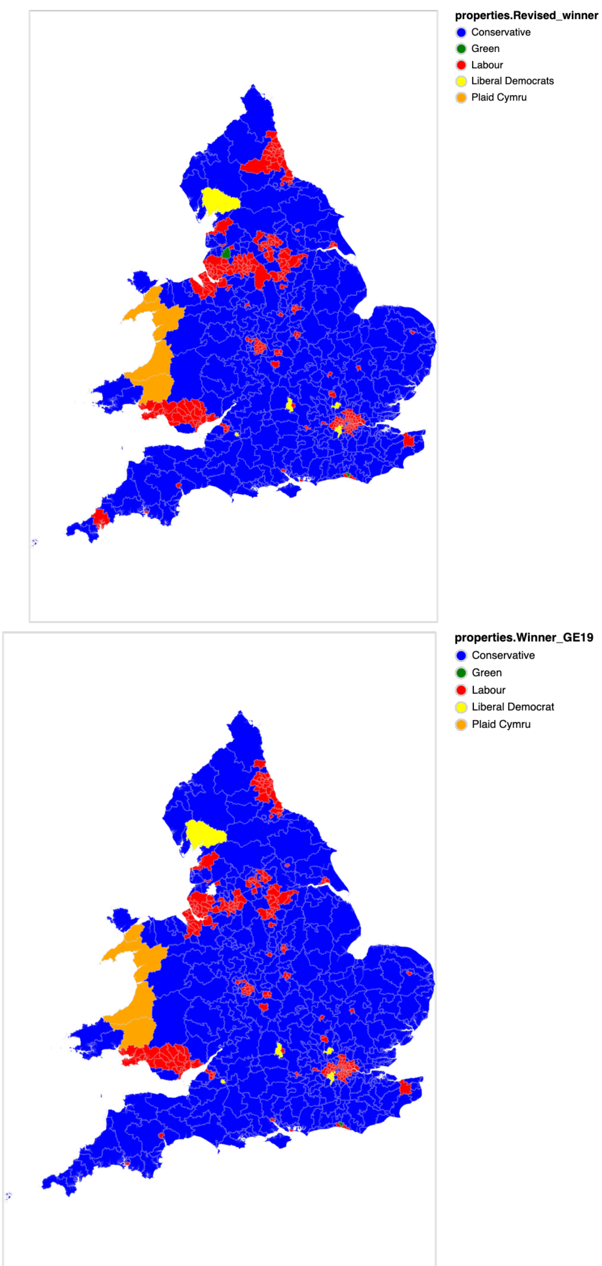

Fig. 9. The modelled effect of introducing 16-17-year olds to the voting pool for the 2019 general election.

4.3 Results

The resulting remodelling election data gives us a visually changed electoral map in contrast to the actual results. However, the small size of the affected constituencies means the impact is understated visually- this could be improved by using a hexagonal electoral map. In the remodelled election results, the Conservatives lose 22 seats, 21 to Labour and one (Chorley) to the Green Party. This could have decreased their majority significantly to 344, though wouldn't have affected the overall result: a Conservative majority government. This would be interesting to model for other, tighter elections, most notably the 2017 general election. The other visual analysis of voting patterns across demographic divides show clearly that age is a very good predictor of voting intention: with older generations more likely to vote for right-wing political parties versus younger generations being more likely to vote for more left-wing parties. While not necessarily a new revelation, this information is important to communicate and understand how it may affect future elections as the population (and therefore electorate) ages.

5 Critical reflection

As mentioned above, I would like to have visualised the modelling results using a hexagon-based electoral map but didn't figure out how to do so in Python. I would also like to have applied the visualization used by Stoffel et. al. (2012), with the minority party also visualized on the map.

I think this would be an interesting analytical method to apply to other UK elections, especially those closer in terms of results, to see if enfranchising 16-17-year olds could affect the conclusion of the election.

In terms of the modelling, I think the outcome can only be described as a loosely modelled result, as the demographic data about specific party voting intentions (vs. the constituency-level correlation data) was drawn from Lord Ashcroft's survey data, which was based on 13,000 voters. While 13,000 is a significant number of people for a survey, this population may not fully represent the 47m-strong electorate.

The modelling also doesn't take into account the concept that 16-17-year olds may not follow the trend of the rest of the electorate in terms of turnout and voting preference, and of course this can change election-to-election based on party policies and whether different parties engage this demographic.

Ultimately, I do think the analysis proves a sufficiently conclusive and contrasting pattern in voting across age demographics, and therefore I did answer my research questions. This analysis is useful to the domain –the more we understand voting patterns and who is more likely to be an undecided voter, or who is less likely to turn out and vote, the more we can tailor our campaign targeting. I think this provides a strong case to add enfranchising 16-17-year olds to the platforms of progressive parties in order to rebalance the effect of an ageing electorate, and to engage younger people with politics sooner in their lives.

Table of word counts

Problem statement	115/250
State of the art	311/500
Properties of the data	396/500
Analysis: Approach	237/500
Analysis: Process	781/1500
Analysis: Results	170/200
Critical reflection	305/500

References

[1] Badawood, D., Wood, J. 'Effects of Candidate Position on Ballot Papers: Exploratory Visualisation of Voter Choice in London Local Council Elections 2010', 2012

[2] Stoffel, F., Janetzko, H., Mansmann, F., 'Proportions in Categorical and Geographic Data: Visualizing the Results of Political Elections', 2012.

[3] Lord Ashcroft, 2019, 'How Britain voted and why: My 2019 general election post-vote poll' Link: https://lordashcroftpolls.com/2019/12/how-britain-voted-and-why-my-2019-general-election-post-vote-poll/#more-16379

[4] Office for National Statistics, 'Overview of the UK population: August 2019: The UK's population is ageing.' Link: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/august2019#the-uks-population-is-ageing

Datasets:

2015, 2016, 2017 data:
- https://www.kaggle.com/tiagotvv/uk-constituency-results/data#

2019 European election data:
- https://www.markpack.org.uk/159014/european-parliament-election-results-2019-broken-down-by-westminster-constituency/

Demographic data, census data, lookup from previous lab exercise.

Lookup dataset:
- https://data.gov.uk/dataset/a4f260f2-bfe8-4af7-84b2-c345dd4c7090/ward-to-westminster-parliamentary-constituency-to-local-authority-district-december-2018-lookup-in-the-united-kingdom

2019 GE data:
- https://docs.google.com/spreadsheets/d/1vuLS04XbYRNji_qIJ6JsFh7y6eW9vA7KMTn89u_cZl8/edit