Space Matters

Investigating Geographic Variability of Voting Patterns in English Parliamentary Constituencies Using Spatial Statistics

Mie Arnau Martinez (MA)

Student reg. no.: 201806701

Spatial Analytics | Aarhus University

Course Instructor: Adéla Sobotkova

Sofie Nødskov Ditmer (SD)

Student reg. no.: 201805308

Spatial Analytics | Aarhus University

Course Instructor: Adéla Sobotkova

Date: June 10, 2021

While we consider this paper a joint effort, author initials have been assigned to each section, with the first mentioned author being the primary contributor. Nonetheless, we wish to emphasize that both authors contributed equally.

Github repository:

https://github.com/sofieditmer/SpatialAnalyticsExamProject

Abstract

Does geographical location influence the voting behavior displayed by the electorate? In other words, does space matter when it comes to electoral processes? This paper examines the influence of spatial dynamics on the 2019 United Kingdom general election with a focus on the parliamentary constituencies of England. The degree of spatial clustering is assessed by performing spatial autocorrelation tests in which the Moran's I statistics is used as a measure of spatial dependence. Furthermore, a geographically weighted regression (GWR) analysis is conducted to assess the influence of demographic variables on voting behavior. The spatial analysis demonstrates that spatial autocorrelation is indeed present, suggesting that neighboring constituencies tend to display similar voting preferences compared to a random distribution. Moreover, the GWR analysis yields a substantial increase in performance compared to a global regression analysis, implying that by taking spatial location into account, a more nuanced understanding of the voting behavior is achieved.

Keywords: *spatial autocorrelation, geographically weighted regression (GWR), modifiable areal unit problem (MAUP), Moran's I*

Contents

1 Introduction

Are voting patterns determined by spatial location? The first law of geography states that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). In other words, various phenomena appear to display a geographical dependency, which has also been demonstrated for electoral processes (Lysek et al., 2020; Mansley and Demšar, 2015). The motivation underlying this project was to understand the spatial dynamics underlying voting patterns displayed by parliamentary constituencies in the 2019 United Kingdom (UK) general election. In particular, this paper examines how voting patterns are spatially distributed across English parliamentary constituencies in the 2019 general election. Emphasis was placed on assessing the difference between voting patterns associated with the Conservative and the Labour parties from a spatial perspective, quantifying the degree of spatial dependency, and investigating the spatial relationship between various socioeconomic variables and voting behaviors.

1.1 Problems and Background (MA)

In 2019, Brits had a choice to make. Who was going to lead their country and negotiate their withdrawal from the European Union?

Traditionally, two parties have dominated the UK parliament: The Conservative Party (CP) and the Labour Party (LP; Gallagher and Mitchell, 2005). CP is placed at the center-right position of the political spectrum, and the party advocates for minimizing governmental influence and increasing citizen independence (Viereck et al., 2021). LP represents the center-left wing and acts as a socialist democratic party formed to support the working class. The party emphasizes the benefits of governmental influence, redistribution of wealth through taxation, and upholding the national health service (Webb, 2020). In recent years, the Liberal Party has gained increasing support from the LP voter-base, however, this paper will focus on the voting patterns related to the former two main parties.

In previous elections, a political and ideological polarization became increasingly evident in the UK. While CP has drawn significant support from rural areas, LP has attracted votes from urban areas (Mamonova and Franquesa, 2020; Rodden, 2019). This tendency is also seen in other countries like the United States, France and Poland (Mamonova and Franquesa, 2020), which leads to an intriguing question: Is there something intrinsic in the living environment that biases you to vote in a certain way?

Differences in voting patterns can partially be attributed to demographic variables. Studies have shown that students and young voters in general tend to display a preference for left-winged parties, and since young people predominantly live in the city this will inevitably affect the distribution of voting patterns (Barton, 2020).

Similarly, immigrants and people of other ethnic origin tend to live in the city adding to the vote share for left-winged parties in urban areas and increasing population heterogeneity (Pew Research Center, 2020). Being continuously confronted with people of different backgrounds and nationalities might increase the support for left-wing parties by mitigating nationalistic affiliation. Contrarily, traditional values associated with right-wing parties are more prominent in rural areas with a more homogeneous population due to less ethnic diversity (Garner and Bhattacharyya, 2011).

Studies have also shown that voters' decisions are influenced by their income (Winkler, 2019), and depend on the party's policy on redistribution of wealth. People with low income favor high taxation and social support services. Contrarily, people with high income prefer parties that want to reduce taxes and governmental involvement (Huber and Stanig, 2009).

A central component of British culture is their 'public houses' (pubs) that play a significant role in creating a sense of community and social network. Mount and Cabras (2016) examined the effect of pubs on social coherence in rural areas of Northern England and found a positive impact of pubs in promoting social engagement and involvement among rural residents. In this paper, pubs per inhabitant is used as a proxy for social engagement to study the effect on voting patterns.

In this paper, we explore whether voting patterns across English parliamentary constituencies depend on the spatial location. In particular, we study the UK general election results from 2019 in England across constituencies and relate these to socio-demographic variables like age, ethnic diversity and income. Moreover, the influence of pubs per inhabitant and population density are examined.

1.2 Related Work (SD)

To gain a thorough insight into the underpinnings of political elections one must acknowledge that electoral processes are influenced by spatial dynamics (Lysek et al., 2020; Mansley and Demšar, 2015).

Using GIS software and spatial analytical methods including Geographically Weighted Regression (GWR), Lysek et al. (2020) identified clear geographical divisions of the

Czech Republic characterized by low education levels, high unemployment rates and other socioeconomic variables.

Similarly, Mansley and Demšar (2015) examined voter turnout in the London mayoral election in 2012. Specifically, the authors explored how voter turnout was linked to sociodemographic variables and how this relationship varied across the 625 wards of London. This relationship was mapped with GWR allowing the authors to disaggregate local variation in voting patterns.

The spatial analyses conducted by Lysek et al. (2020) and Mansley and Demšar (2015) demonstrated how analyzing voting patterns using spatial statistics can inform our understanding of the influence of contextual dynamics on voting patterns.

2 Methods

This section includes a brief overview of the software framework underlying the current project, a detailed description of data acquisition and preprocessing procedures, as well as a run-through of the main spatial analysis method employed.

2.1 Software Framework (SD)

The code developed for this project has been tested on both a 2013 MacBook Pro, 16 Gb RAM, that currently runs on macOS Big Sur (v. 11.2.3) as well as a 2020 Lenovo IdeaPad S340-14IIL, 8 Gb RAM, that runs on the Window 10 operating system. On both machines, the desktop version of R is 4.0.4 while the version RStudio is 1.2.5033 and 1.3.959 respectively. An overview of the packages needed to reproduce the contents of this project are listed in Table 2.1.

2.2 Data Acquisition and Processing

This section includes a brief overview of the provenance of the data sources used in this paper as well as details of particular data manipulation procedures performed to prepare for further spatial analysis. To gain insight into the preprocessing procedures applied to the spatial data, the reader is referred to the preprocessing script available in the GitHub repository (see metadata in Section 7).

Package	Version	Acknowledgement
tidyverse	1.3.0	Wickham et al. (2019)
sf	0.9.7	Pebesma (2018)
tmap	3.3.1	Tennekes (2018
rgdal	1.5.23	Bivand, Keitt, & Rowlingson (2021)
spatialEco	1.3.7	Evans (2021)
pacman	0.5.1	Rinker & Kurkiewicz (2017)
spdep	1.1.5	Bivand & Wong (2018)
maptools	1.0.2	Bivand & Lewin-Koh (2020)
cartogram	0.2.2	Jeworutzki (2020)
spgwr	0.6.34	Bivand & Yu (2020)

Table 2.1: Packages that need to be installed in RStudio in order to reproduce the contents of this project

2.2.1 General Election Data (MA)

The 2019 election results were obtained from the House of Commons Library (2020a). The dataset contains results of the UK general election represented by constituency from 1918 to 2019. For each parliamentary constituency, information on the number of total votes and the number of votes for each political party, electorate and turnout are included. The vote shares for each major political party, CP and LP, were calculated manually. Since the data included information for all countries within UK, the data was filtered to only encompass England as part of the data preprocessing.

2.2.2 Spatial Constituency Vector Data (SD)

The spatial constituency vector boundaries of the parliamentary constituencies were obtained from the Office for National Statistics (ONS; 2018) in a shapefile format that stores the digital vector boundaries for the parliamentary constituencies as of December 2016. Relying on the `st_transform` method available in the `sf` package (Pebesma, 2018), the coordinate reference system (CRS) of the spatial data was converted to EPSG:27700; the projected coordinate reference system of Great Britain. The constituency of Chorley was excluded from analysis as part of the preprocessing because of missing data values.

2.2.3 Point Data for English Pubs (MA)

The point data of all pubs in England was acquired through Kaggle (Tatman, 2017). The dataset originates from the Food Hygiene Ratings published by the Food Standards Agency (FSA) and from the ONS Postcode Directory licensed under the Open Government License

(Office for National Statistics, 2020). Since the spatial point data was missing longitude and latitude coordinates for 72 pubs, an attempt was made to recover these from the easting and northing coordinates. However, as these coordinates proved to be inaccurate it was decided to merge the spatial pub data with postcode geodata derived directly from the ONS Postcode Directory to acquire the longitude and latitude for the remaining pubs.

The preprocessed spatial pub data was then transformed into an `sf` object and assigned the CRS EPSG;27700. To ensure that only pubs located in English constituencies were included in the analysis, the `st_intersection` method available in the `sf` package (Pebesma, 2018) was employed. To calculate the number of pubs per inhabitant, the `point.in.poly` function available in `spatialEco` (Evans, 2021) was used to group the pubs according to which constituency they are located in. Next, we calculated how many pubs were located in each constituency and this was then divided by inhabitants in the current constituency.

2.2.4 Demographic Data (SD)

The demography data includes age, ethnicity and average income. The age data was acquired from the Office for National Statistics (2020) and contains population estimates of age in England and Wales from mid 2019. The age data includes the distribution of ages for inhabitants in each constituency and from this, the average age for each constituency was calculated. Data on ethnicity was retrieved from the House of Commons Library (2020b). The percentage of white population in each constituency was used as a proxy for ethnic homogeneity. Lastly, data on average income for constituencies in the years 2018-2019 was retrieved from the non-ministerial government department Her Majesty's (HM) Revenue and Customs (2013).

2.2.5 Classifying Urban and Rural Constituencies (MA)

The classification of rural and urban English parliamentary constituencies was obtained from ONS (Office for National Statistics, 2016, 2). It includes 6 subdivisions; major urban, large urban, other urban, significant rural, rural-50, and rural-75. These subdivisions have been formalized by the Rural Evidence Research Centre (RERC) at Birkbeck College. For more information about the classification scheme see Office for National Statistics, 2016 (2). The classifications have been defined based on settlement population. While settlements with populations of over 10,000 are considered urban, settlements with populations below 10,000 are considered rural.

We combined the subdivisions to comprise a binary classification. This was used to define the parliamentary constituencies as either urban or rural. The classification data also contains population counts for each constituency which we use to calculate population density. First, we computed the area of each constituency using the `st_area` function from the `sf` package (Pebesma, 2018). Then, we divided population count by area.

2.3 Spatial Analytical Methods

The following section includes an overview of the spatial analytical methods employed in this paper. While sections 2.3.1.-2.3.3 describe the methods used when quantifying spatial autocorrelation sections 2.3.4-2.3.5 outline the spatial regression analyses. The associated scripts containing the code for each analysis can be found in the GitHub repository (see 7.1 Metadata).

2.3.1 Cartograms (SD, MA)

When mapping parliamentary constituencies, larger constituencies with a low population density carry more visual weight compared to smaller constituencies with a high population density. To correct for this visual imbalance, cartograms were created to modify the size of each constituency to be proportional to a specified variable. This means that constituencies with a low population density are contracted in size while high density urban constituencies are expanded, while still retaining roughly the same location of the constituencies relative to one another. To produce the cartograms, the `cartogram_cont` function available in the `cartogram` package (Jeworutzki, 2020) was used.

2.3.2 Testing for Spatial Autocorrelation: Moran's I (SD, MA)

To test for spatial autocorrelation among constituencies the Moran's I statistic was computed using Monte Carlo simulations to estimate the significance of the spatial dependency. In this regard, we relied on the `moran.mc` method available in the `spdep` package (Bivand and Wong, 2018).

To ensure that the results obtained were not depending on the definitions of neighborhoods, different neighborhood classifications were implemented (Fig. 2.1). Consequently, we employed a contiguity-based neighborhood definition which classifies neighbors according to adjacency, as well as a distance-based neighborhood definition that classifies neighbors based on a specified proximity (Gimond, 2021).

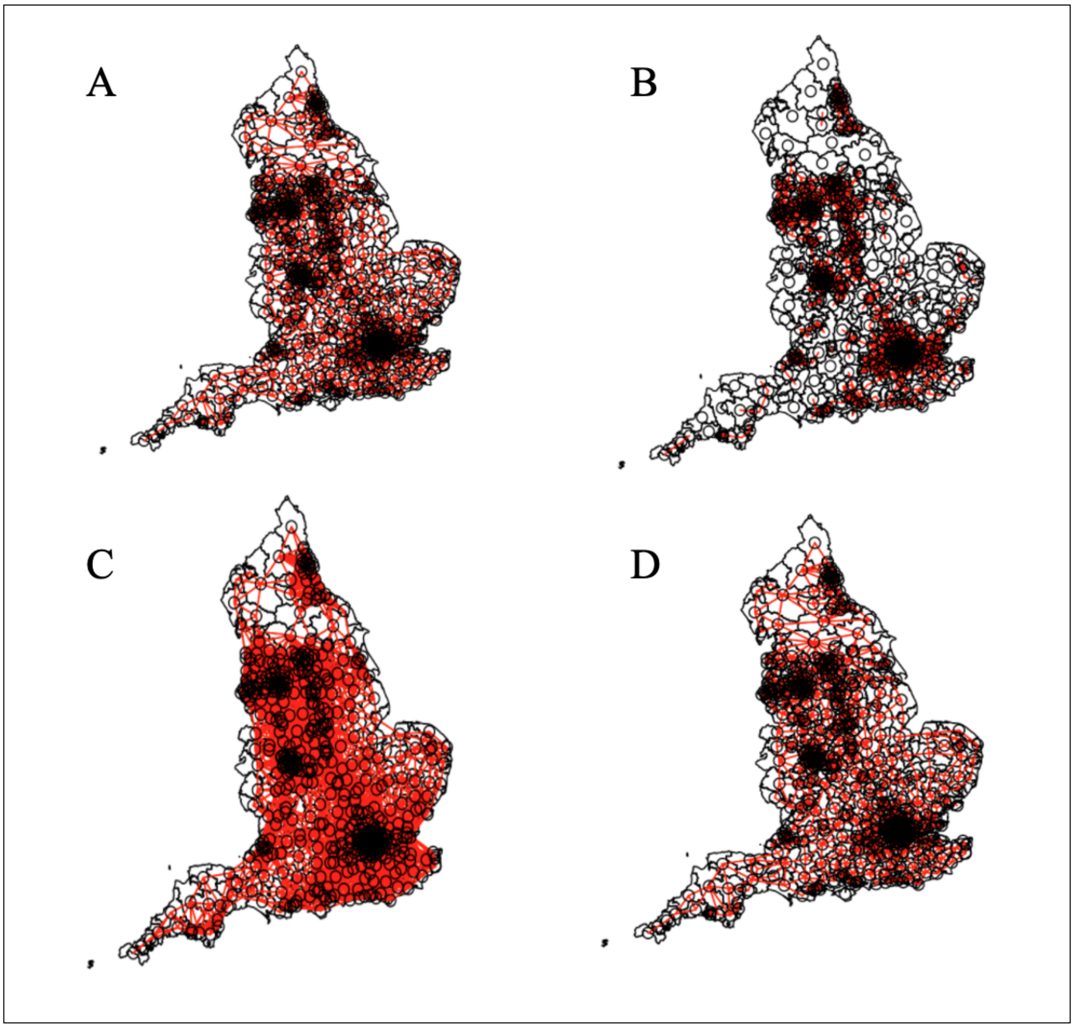

Figure 2.1: Neighboring constituencies defined by adjacency (A), distance band of 20 km (B), distance band of 50 km (C) and k-nearest neighbors k=3 (D)

2.3.3 Addressing the Modifiable Areal Unit Problem (MAUP) (SD, MA)

To address the Modifiable Areal Unit Problem (MAUP) associated with using aggregated data, an alternative aggregation scheme was employed to assess the influence of the spatial boundaries on the degree of spatial autocorrelation (Fig. 2.2). Hence, the constituencies were aggregated into counties and vote shares were averaged for constituencies belonging to the same county.

For this purpose, the aggregate method for sf objects was employed. The newly generated boundaries were simplified using the `st_simplify` method from the `sf` package (Pebesma, 2018) to increase processing speed. Moreover, neighborhoods for the aggregated constituencies were defined relying on the adjacency neighborhood definition using the `poly2nb` method (Bivand and Wong, 2018).

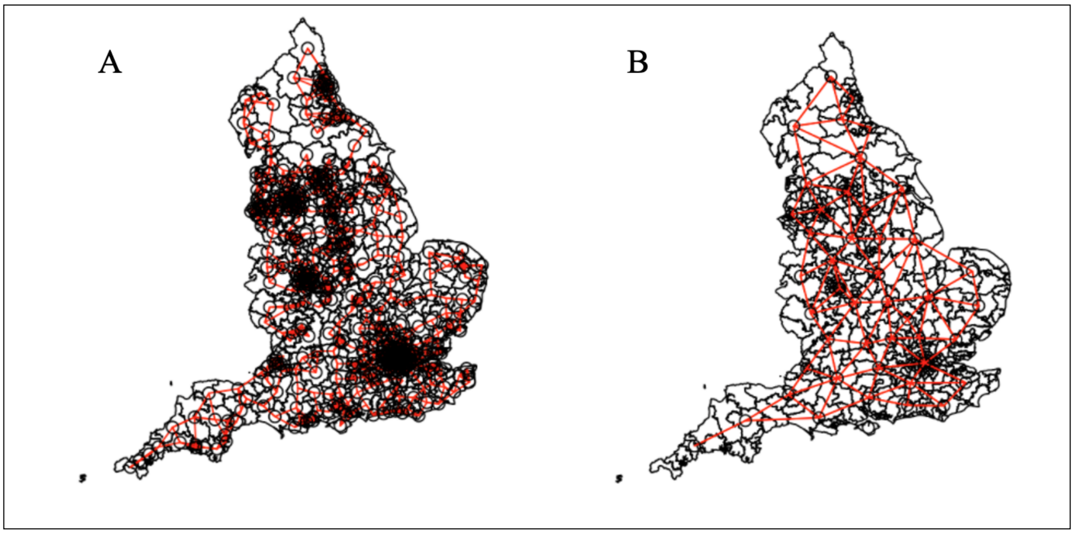

Figure 2.2: Neighborhoods defined by adjacency for constituencies (A) and for counties (B)

2.3.4 Global Regression Analysis (MA, SD)

To examine the relationship between CP and LP vote shares and demographic variables, a global linear regression model was fitted to the election data (Equation 1). This model takes all observations into consideration and weights all data equally.

$$vote\ share \sim age + income + ethnicity + populationdensity + pubs/inhabitant \tag{1}$$

As the fit of the global model was expected to vary across constituencies, Moran's I with Monte Carlo simulation was calculated on the model residuals to test for spatial autocorrelation.

2.3.5 Geographically Weighted Regression (GWR) Analysis (MA, SD)

GWR is a spatial regression that takes geographical location into account. The model allows coefficients for each variable to vary according to space instead of estimating one global coefficient. This means that the model estimates depend on the location of the observations. For this project, it made it possible to discern the importance of the predictors in different constituencies in relation to how these affect voting patterns.

When performing a GWR analysis, the specified bandwidth determines how neighboring constituencies are defined. This study implemented an adaptive bandwidth. The `gwr.sel` method from the `spgwr` package (Bivand and Yu, 2020) was employed, performing cross validation to estimate the optimal bandwidth.

Furthermore, a Gaussian kernel function was applied in the GWR which allows all locations to contribute to the model at focal point but ensures that near locations are weighted higher than distant locations.

The spatial model is fitted to the data using the `gwr` function from the `spgwr` package (Bivand and Yu, 2020). To assess the validity of the model, the model residuals are examined. Here, an autocorrelation test was performed using Moran's I with Monte Carlo simulation.

3 Empirical Results

The following sections include an overview of the main results related to the spatial autocorrelation analysis across different neighborhood definitions and aggregation schemes as well as the spatial regression analysis.

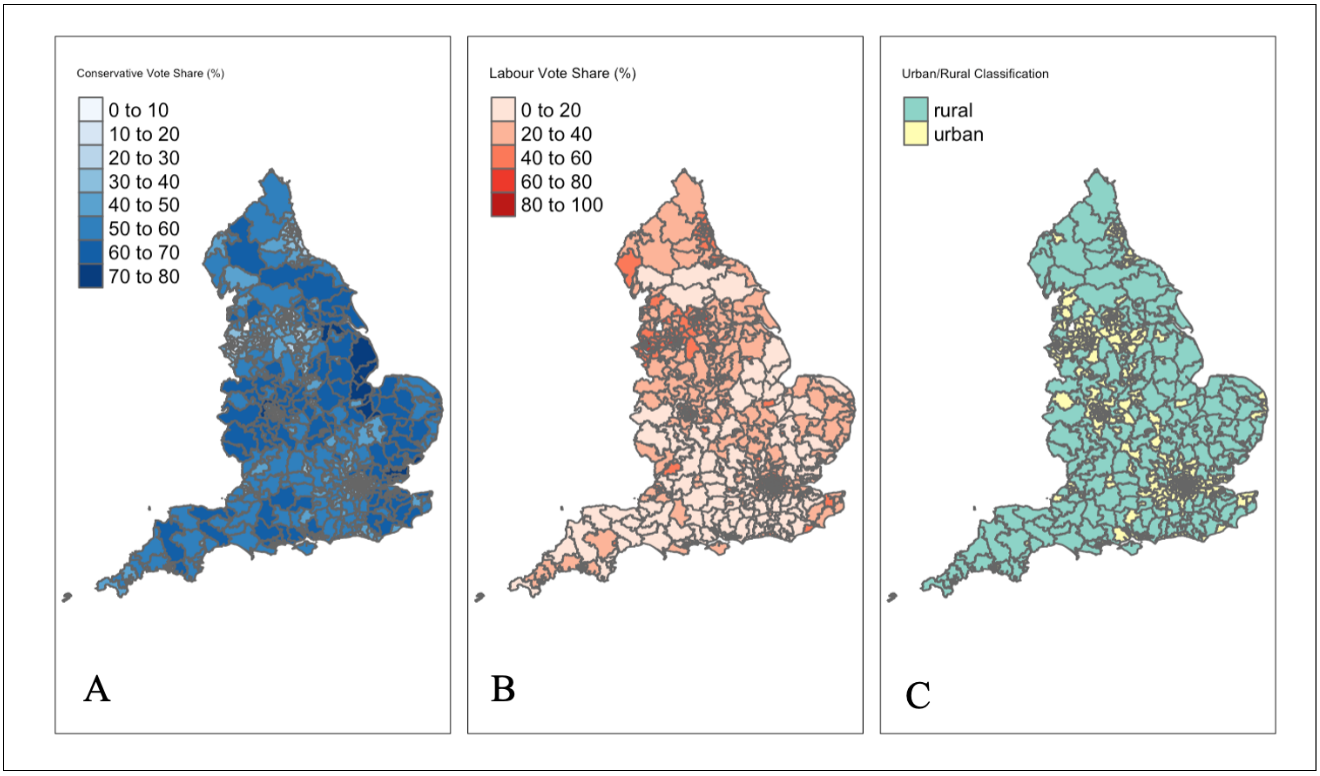

Figure 3.1: Maps showing the conservative vote share for each constituency (A), the labour vote share for each constituency (B) and each constituency classified as either rural or urban (C)

3.1 Spatial Autocorrelation Analysis (SD, MA)

Assessing political polarization in relation to rural and urban constituencies it appears that there is a spatial dependency between favouring CP (Fig. 3.1A) and living in rural areas (Fig. 3.1C). Similarly, urban areas seem to display a tendency of favouring the LP (Fig. 3.1B and Fig. 3.1C).

To correct the imbalance in visual weight in Fig. 3.1, several cartograms were produced in which the area of each constituency was scaled to a specified variable. The area of the constituencies was modified to be proportional to the total number of eligible voters (Fig. 3.2A), the CP vote share (Fig. 3.2B), and the LP vote share (Fig. 3.2C). When assessing the cartograms in which urban areas have higher visual representation, it becomes evident that urban constituencies are more prone to vote for LP compared to rural areas. This is exemplified by Liverpool marked in yellow in Fig 3.2.

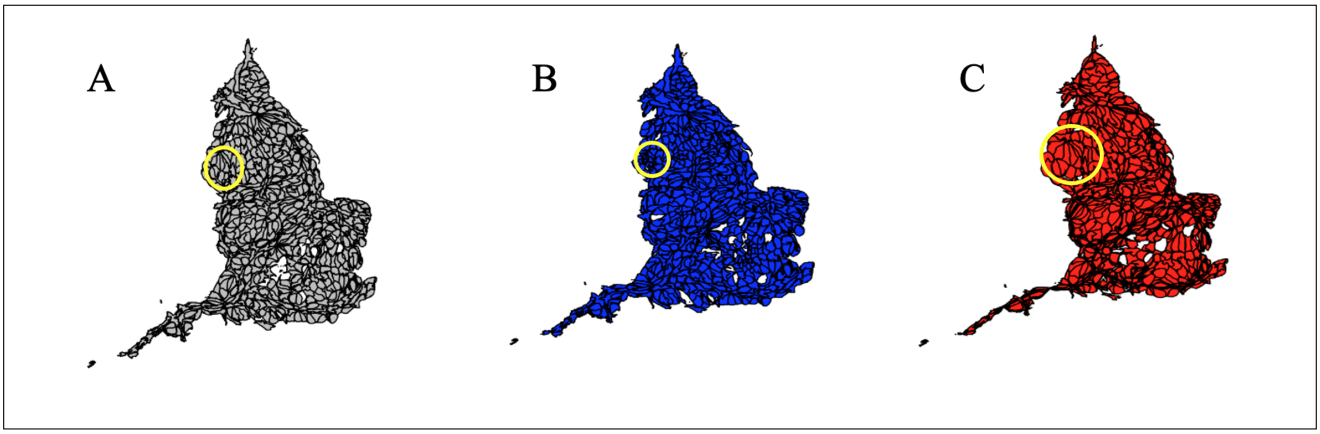

Figure 3.2: Cartograms displaying constituency area scaled to electorate (A), Conservative vote share (B), and Labour vote share (C).

When quantifying the degree of spatial autocorrelation in regard to CP and LP vote shares, Moran's I was found to be positive and the p-value was statistically significant across all neighborhood definitions as well as aggregation schemes (Table 3.1). The highest degree of spatial clustering was observed for CP vote share aggregated by constituency using the Queen adjacency neighborhood definition, Moran's I: .66, $p < .001$, while the lowest degree of spatial clustering was observed for CP vote share aggregated by county using Queen adjacency neighborhood definition, Moran's I: .20, $p < .02$.

Conservative Vote Share			
Moran's I	**p-value**	**Neighborhood Definition**	**Aggregation Scheme**
0.66	0.001	Queen adjacency	Constituencies
0.41	0.001	Distance-based (20 km.)	Constituencies
0.23	0.001	Distance-based (50 km.)	Constituencies
0.65	0.001	K-nearest neighbors (k = 3)	Constituencies
0.20	0.02	Queen adjacency	Counties
Labour Vote Share			
0.62	0.001	Queen adjacency	Constituencies
0.38	0.001	Distance-based (20 km.)	Constituencies
0.23	0.001	Distance-based (50 km.)	Constituencies
0.58	0.001	K-nearest neighbors (k = 3)	Constituencies
0.39	0.001	Queen adjacency	Counties

Table 3.1: Results of spatial autocorrelation

3.2 Spatial Regression Analysis

3.2.1 Global Regression (MA, SD)

When assessing the global regression model for the CP, population density, pubs per inhabitants, average age and average income were found to significantly predict CP vote share: $F(5, 525) = 141$, $p < .0001$, adjusted $R2 = .57$. However, the percentage of white population did not significantly predict the percentage of votes, $= .11$ (SE $= .31$), $t = .37$, $p > 0.05$. Moreover, the adjusted R2 indicates that 57% of the variance present in the election data can be explained by the model.

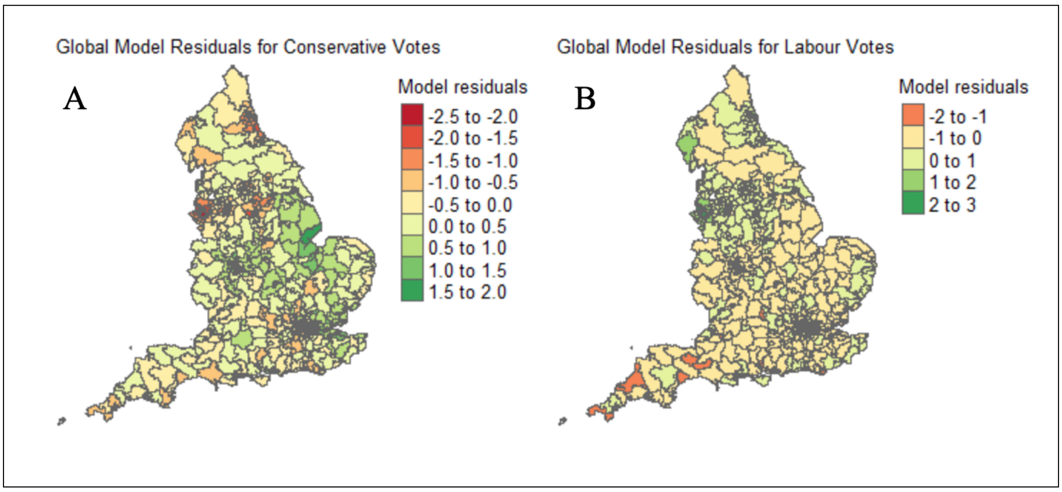

Figure 3.3: Global models residuals for Conservative vote share (A) and Labour vote share (B)

When assessing the global regression model for LP, population density, pubs per inhabitants, average age and average income were found to significantly predict vote share:

$F(5, 525) = 204.9$, $p < .0001$, adjusted R2 = .66. As with the previous model, the percentage of white population did not significantly predict the vote share: = -.52 (SE = .28), t = -1.87, $p > .05$. The adjusted R2 indicates that 66% of the variance can be explained by this model.

The results obtained from the Moran's I test for spatial autocorrelation indicate that the model residuals have spatial clustering for both CP (Fig 3.3A), Moran's I = .60, p = .001 as well as LP (Fig 3.3B, Moran's I = .52, p = .001.

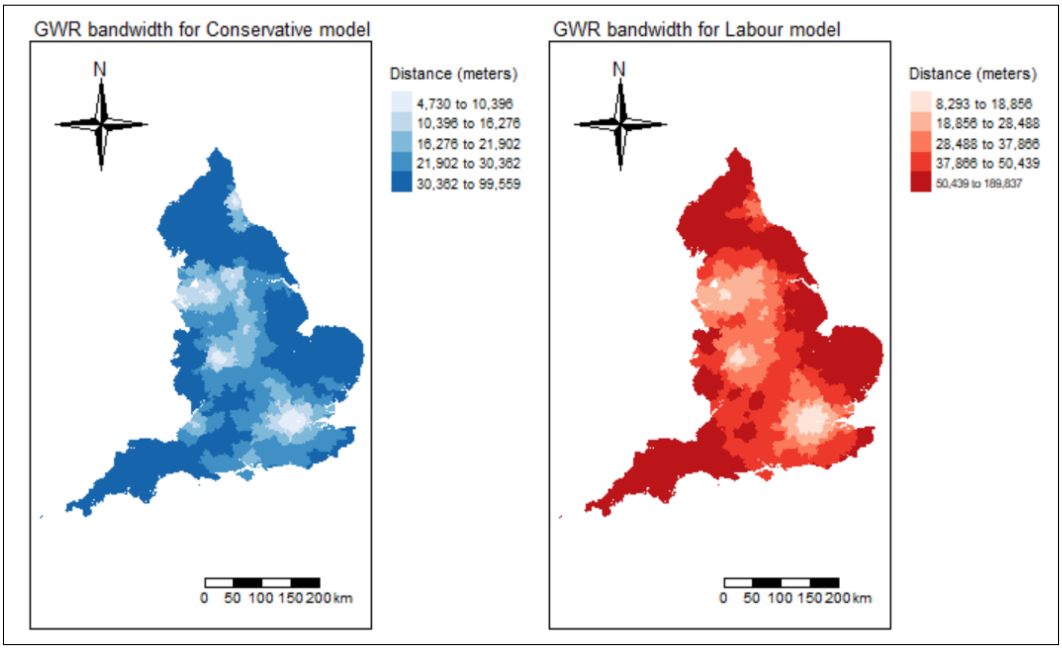

Figure 3.4: Bandwidth distance values across parliamentary constituencies

3.2.2 Geographically Weighted Regression (GWR) (MA, SD)

The adaptive bandwidth was estimated to 7 neighbors for the GWR model for CP and 18 for the GWR model for LP. For clarification on how this affected the bandwidth distance see Fig. 3.4.

As expected, Fig. 3.4 illustrates how dense, urban areas like London and Manchester are assigned smaller bandwidths compared to rural areas. This trend is evident for both CP as well as LP. However, it should be noted that the color scales for the two maps differ.

The GWR model of CP vote share obtained an adjusted R2 of 91%. Similarly, the model acquired an AIC value of 404.35. In comparison, the GWR model for LP vote share obtained an adjusted R2 of 87%, and an AIC value of 498.41.

Moreover, spatial autocorrelation was quantified for model residuals for both the GWR model on CP vote share (Fig. 3.5A) and LP vote share (Fig. 3.5B). For the CP model, we obtained a Moran's I = .13, $p < .001$. For the LP model, the autocorrelation test found

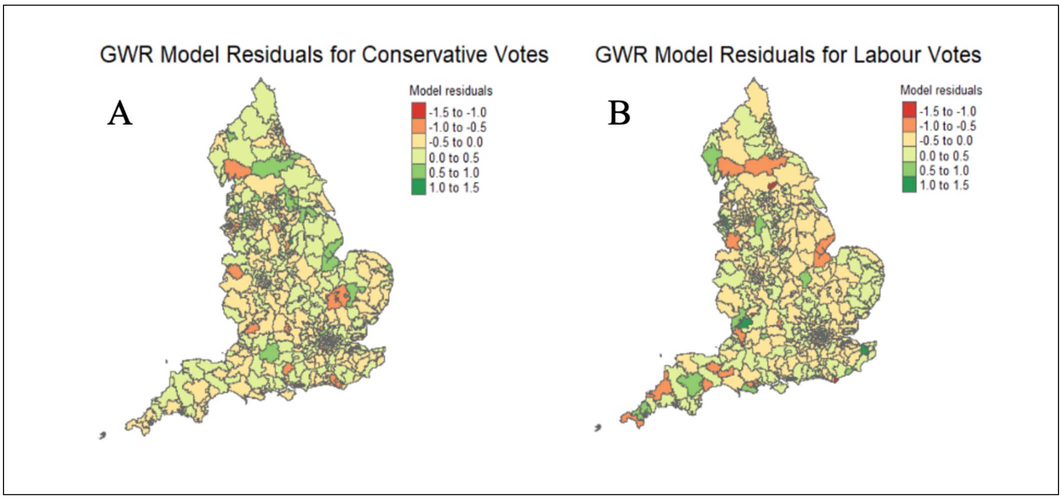

Figure 3.5: Model residuals of the GWR Conservative model (A) and the GWR Labour model (B)

Moran's I = .10, p <.001

The GWR model coefficients as well as the global model estimates are reported in Table 3.2. When assessing these values for both GWR models, it is evident that the coefficients across predictors display great local variation. Furthermore, it seems that the directionality of the effect of each predictor varies substantially depending on geographical location.

For instance, pubs per inhabitants has a negative global effect on CP vote share (Table 3.2). However, when assessing the GWR model, the effect of pubs is found to vary quite substantially across constituencies. Fig. 3.6 illustrates the influence of pubs per inhabitants on voting patterns across constituencies and how this variation differs for CP and LP vote shares.

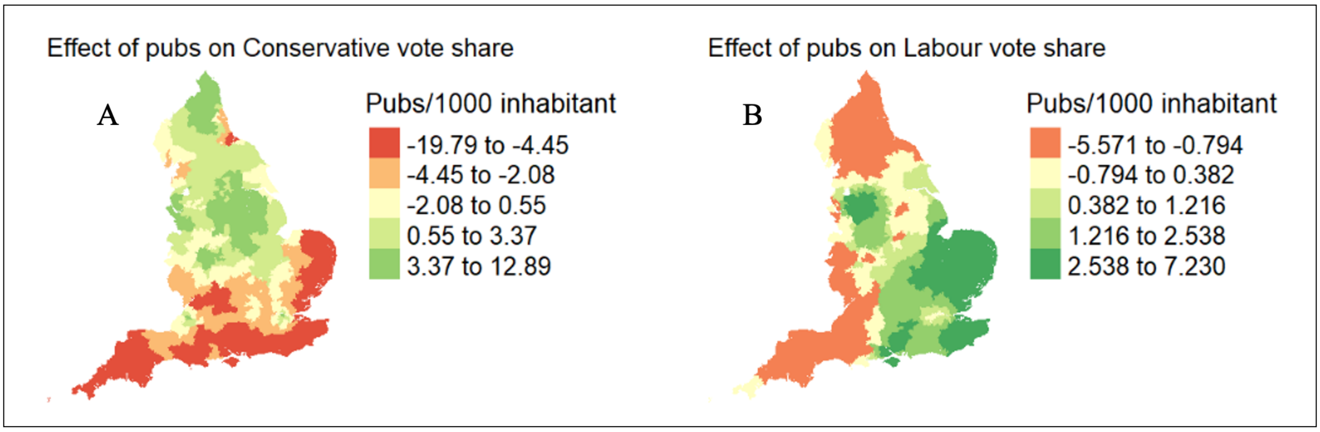

Figure 3.6: The effect of pubs per inhabitants (in 1000) for the Conservative party (A) and the Labour party (B).

In comparison, effect of average age on LP vote share was relatively robust across models. Here, an increase in average age appears to be related to an increase in LP vote share across

Coefficients for Conservative model				
Predictor	Min	Median	Max	Global
Population density*	-31.17	-5.33	7.81	-7.32
Pubs per thousand inhabitants*	-1.25	-0.05	0.82	-0.26
Percent white population	-16.37	1.25	23.93	0.11
Average age*	-0.07	0.12	0.41	0.12
Average income*	-0.0002	-0.000004	0.00009	0.00002
Coefficients for Labour model				
Population density*	0.79	5.72	19.38	7.85
Pubs per thousand inhabitants*	-0.32	0.04	0.41	0.17
Percent white population	-3.41	-1.71	5.12	-0.52
Average age*	-0.21	-0.11	-0.04	-0.11
Average income*	-0.00005	-0.00002	0.00003	-0.00004

Table 3.2: Global and GWR model coefficients for the Conservative party (A) and the Labour party (B). * indicates $p < .05$.

constituencies. Similarly, average age has a positive correlation with CP vote share across the majority of the constituencies (Table 3.2). In this regard, it should be noted that although directionality appears more stable for the effect of age, the variation across constituencies is still present. For map visualizations of all model coefficients see Appendix 8.1 and 8.2.

4 Critical Evaluation

The following section evaluates the results in the light of the data sources and research premises. This evaluation includes a discussion of the representativeness, reliability, and generalizability of the observed spatial patterns of voting behavior.

4.1 Space Matters (SD, MA)

Quantifying the degree to which neighboring constituencies display similar voting preferences, a positive Moran's I value and a significant p-value were observed across adjacency- and distance-based neighborhood definitions as well as different aggregation schemes. These results suggest that spatial autocorrelation was indeed present for the 2019 UK election data. In other words, neighboring constituencies appear highly similar in regard to voting preferences compared to a random distribution. However, differences in the degree of spatial clustering was observed for election data aggregated by constituency as compared to aggregated by county. While a high degree of spatial dependency was present for vote shares aggregated by constituencies, indicated by a large, positive Moran's I, a lower degree of spatial dependency was observed for constituencies aggregated by county. These results suggest that spatial resolution influences the degree to which constituencies display similar voting behavior, emphasizing the importance of addressing the Modifiable Areal Unit Problem (MAUP) and exhibiting caution when interpreting the results of aggregated data. One could argue that increasing the spatial resolution would have yielded more reliable results, however, the availability of election data of this kind was unfortunately found to be highly limited.

Nevertheless, the use of different neighborhood definitions and aggregation schemes contribute to an increase in the reliability and robustness of the observed spatial patterns. Moreover, it ensures that ecological fallacy, i.e., assuming that one statistical relationship holds for all aggregation schemes, is avoided. Given that spatial autocorrelation was found to be present and significant across different neighborhood definitions and aggregations, it seems reasonable to assert that the voting patterns are spatially dependent.

4.2 Letting the Data Determine the Variability (MA, SD)

With a geographically weighted regression approach, we built a model that was able to explain roughly 91% of the variation in CP vote share using a set of demographic and socioeconomic variables (a marked increase from 57% obtained by the global model). With GWR, the AIC decreased from 1067 to 404, indicating a substantial increase in model quality. Similarly, the GWR model for LP vote shares was able to explain 87% of the variance (an increase from 66% obtained by the global model). Likewise, the AIC value decreased from 945 to 498 for the GWR model. These results indicate that when employing a model that takes space into account, the data can be accounted for to a much higher extent.

Furthermore, when examining model residuals, we see that spatial clustering is markedly reduced by the GWR models compared to the global regressions. This suggests that the GWR models capture spatial dependency in the data that was present in the global model residuals. However, there still seems to be some spatial aspect in our data that is not explained by the GWR models as spatial clustering for the residuals is still significant. While this indicates that weighting based on location improves the model, additional predictors should be included to increase model quality.

In addition, all predictors were found to vary substantially across space (Table 3.2). When assessing the influence of pubs on voting patterns, it appeared that the effect was biggest in rural areas and close to zero in urban areas (Fig. 3.6 and Fig. 3.1C). However, the directionality of the effect did not seem to depend on the area classification. More importantly, we found that pubs indeed seemed to affect voting patterns and the size of the effect depends on the spatial location, e.g., in the Southern parts of England more pubs tend to result in fewer votes for CP (Fig. 3.6). While pubs are an inherent part of British culture, one could argue that it might not constitute the most optimal proxy for socialization. Therefore, it might be beneficial to combine pub data with other types of cultural institutions (e.g., libraries).

Lastly, when examining the voting patterns for each party in the global models, we found an inverse relationship between all predictors. However, when assessing the voting patterns in the two GWR models, the inverse relationship between the predictors for CP and LP was still present but found to vary across constituencies. This suggests that taking space into account provides a more nuanced understanding of how the demographic predictors influence voting behavior.

4.3 Assessing Data and Model Quality (SD, MA)

When considering the quality of the data sources used in the current study, it seems relevant to note that the nature of the data provenance was of high priority. Hence, all data sources were either obtained from the Office for National Statistics or the House of Commons Library; Governmental maintained institutions.

In relation to the quality of the spatial regression models, one could argue that including additional cultural predictors of voting behavior such as religion could potentially increase model performance. In relation to this matter two points can be made. Firstly, the availability of such data on a national level was found to be highly limited, and secondly, it was deemed essential to avoid increasing the complexity of the model to an unnecessary extent. However, adding additional cultural variables to the spatial analysis could be an interesting avenue for further research.

But is the analysis generalizable? While urban and rural political polarization is a widespread phenomenon, the observed relationship with political orientation only holds for Western democracies. For instance, India and most of Southeast Asia tend to support communist parties (Zagoria, 1971). Hence, one should be aware that the results of this analysis would not be transferable if compared with election results outside the Western world. On a similar note, this election was marked by being at the backdrop of Brexit. Consequently, political and economic policies unrelated to the EU withdrawal agreement might have been of low priority in this particular election, which could obscure the generalizability of the results. Performing the spatial analysis on previous elections would be an interesting avenue for further research.

5 Conclusion

In 2019, the Conservative Party were re-elected for the UK parliament with a large majority of the votes (44 %) giving them 364 seats. This paper examined the spatial dependencies underlying the UK general election in 2019 in English parliamentary constituencies. The results revealed a spatial clustering of percentage of vote shares supporting the notion that geography should indeed be taken into account when investigating voting patterns. Furthermore, the spatial analysis displayed a tendency for rural electorates to support the Conservative Party and urban electorates to favor the Labour Party. Moreover, the influence of demographic predictors were found to vary substantially across constituencies for both the Conservative Party and the Labour Party. This means that space matters to how environmental features influence our vote preferences. All things considered, space does seem to matter when it comes to electoral processes.

6 References

Barton, C. (2020). GE2019: How did demographics affect the result? https://commonslibrary.parliament.uk/ge2019-how-did-demographics-affect-the-result/

Bivand, R. and Yu, D. (2020). spgwr: Geographically Weighted Regression. R package version 0.6-34. https://CRAN.R-project.org/package=spgwr

Bivand, R., Keitt, T., and Rowlingson, B. (2021). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.5-23. https://CRAN.R-project.org/package=rgdal

Bivand, R. Lewin-Koh, N. (2020). maptools: Tools for Handling Spatial Objects. R package version 1.0-2. https://CRAN.R-project.org/package=maptools

Bivand, R., Wong, D. (2018) Comparing implementations of global and local indicators of spatial association TEST, 27(3), 716-748. doi: https://doi.org/10.1007/s11749-018-0599-x

EPSG:27700. Epsg.Io. https://epsg.io/27700

Evans, J. S. (2021). spatialEco. R package version 1.3-6, doi:https://github.com/jeffreyevans/spatialEco

Gallagher, M., Mitchell, P. (2005). The Politics of Electoral Systems. OUP Oxford.

Garner, S. and Bhattacharyya, G. (2011) Poverty, ethnicity and place. York: JRF. Gimond, M. (2021). Intro to GIS and Spatial Analysis. https://mgimond.github.io/Spatial/index.html

Her Majesty's (HM) Revenue and Customs (2013). Income and tax by Parliamentary constituency, confidence intervals. Retrieved from: https://www.gov.uk/government/statistics/income-and-tax-by-parliamentary-constituency-confidence-intervals

House of Commons Library (2020a). General election results from 1918-2019. Retrieved from: https://commonslibrary.parliament.uk/research-briefings/cbp-8647/

House of Commons Library (2020b). Constituency data: ethnicity. Retrieved from: https://commonslibrary.parliament.uk/constituency-statistics-ethnicity/

Huber, J. D., Stanig, P. (2009). Individual income and voting for redistribution across democracies. 29.

Jeworutzki, S. (2020). cartogram: Create Cartograms with R. R package version 0.2.2. https://CRAN.R-project.org/package=cartogram

Lysek, J., Pánek, J., Lebeda, T. (2020). Who are the voters and where are they? Using spatial statistics to analyse voting patterns in the parliamentary elections of the Czech Republic. Journal of Maps, 1-6.

Mamonova, N., Franquesa, J. (2020). Right-Wing Populism in Rural Europe. Introduction to the Special Issue. Sociologia Ruralis, 60(4), 702–709.https://doi.org/10.1111/soru.12306

Mansley, E., Demšar, U. (2015). Space matters: Geographic variability of electoral turnout determinants in the 2012 London mayoral election. Electoral Studies, 40, 322–334. https://doi.org/10.1016/j.electstud.2015.10.003

Mount, M., Cabras, I. (2016). Community Cohesion and Village Pubs in Northern England: An Econometric Study. Regional Studies, 50(7), 1203–1216. https://doi.org/10.1080/00343404.2014.989150

Office for National Statistics (2016) (1). Rural/urban local authority (LA) classification (England). Retrieved from: https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2001rural

Office for National Statistics (2016) (2). Rural and Urban Statistics Guidance Notes. https://www.ons.gov.uk/file?uri=/methodology/geography/geographicalproducts/ruralurbanclassifications/

Office for National Statistics (2018). Westminster Parliamentary Constituencies (December 2016) Full Clipped Boundaries in Great Britain. Retrieved from: https://data.gov.uk/dataset/4fd0633a-1c26-4619-b905-15c89ff7c6a0/westminster-parliamentary-constituencies-december-2016-full-clipped-boundaries-in-great-britain

Office for National Statistics (2020). Open Postcode Geo. Retrieved from: https://www.getthedata.com/open-postcode-geo

Office for National Statistics (2020). Parliamentary constituency population estimates (Experimental Statistics). Retrieved from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/parliamentaryconstituencymidyearpopulationestimates

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10(1), 439-446, https://doi.org/10.32614/RJ-2018-009

Pew Research Center (2020). In Changing U.S. Electorate, Race and

Education Remain Stark Dividing Lines. Pew Research Center - U.S. Politics Policy. https://www.pewresearch.org/politics/2020/06/02/in-changing-u-s-electorate-race-and-education-remain-stark-dividing-lines/

Rinker, T. W. and Kurkiewicz, D. (2017). pacman: Package Management for R. version 0.5.0. Buffalo, New York. http://github.com/trinker/pacman

Rodden, J. A. (2019). Why Cities Lose: The Deep Roots of the Urban-Rural Political Divide. Hachette UK.

Tatman, R. (2017). Every Pub in England (Version 1). Retrieved from https://www.kaggle.com/rtatman/every-pub-in-england/metadata

Tennekes, M. (2018). tmap: Thematic Maps in R. Journal of Statistical Software, 84(6), 1-39. doi: https://doi.org/10.18637/jss.v084.i06

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography, 46(sup1), 234-240.

Viereck, P., Dagger, R., Minogue, K., and Ball, T. (2021, May 17). Conservatism. Encyclopedia Britannica. https://www.britannica.com/topic/conservatism

Webb, P. D. (2020, April 7). Labour Party. Encyclopedia Britannica. https://www.britannica.com/topic/Labour-Party-political-party

Wickham et al. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Winkler, H. (2019). The effect of income inequality on political polarization: Evidence from European regions, 2002–2014. Economics Politics, 31(2), 137–162. https://doi.org/10.1111/ecpo.12129

Zagoria, D. S. (1971). The Ecology of Peasant Communism in India. The American Political Science Review, 65(1), 144–160. https://doi.org/10.2307/1955050

7 Metadata

7.1 Software metadata

Nr	Software metadata description	
S1	Current software version	R 4.0.4; RStudio 1.2.5033
S2	Permanent link to your code in your Github repository	https://github.com/sofieditmer/SpatialAnalyticsExamProject
S3	Legal Software License	MIT (see https://github.com/sofieditmer/SpatialAnalyticsExamProject/blob/main/LICENSE)
S4	Computing platform / Operating System	The code has been tested on MacOS Big Sur (v. 11.2.3) and Windows 10
S6	If available Link to software documentation for special software	https://cran.r-project.org/web/packages/spgwr/spgwr.pdf
S6	Support email for questions	201805308@post.au.dk; 201806701@post.au.dk

7.2 Data metadata

Nr	Metadata description	License
D1	Data on income	Open Government License v3.0
D2	General election data	Open Parliament License
D3	Data on ethnicity	Open Parliament License
D4	Urban/rural classification data	Open Government License v3.0
D5	Spatial constituency boundary data	Open Government License v3.0
D6	Postcode Geodata	Open Government License v3.0
D7	Population data	Open Government License v3.0

Table 7.1: For information on Open Government License v3.0 see: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/; For information on Open Parliament License see: https://www.parliament.uk/site-information/copyright-parliament/open-parliament-licence/

8 Appendices

8.1 Appendix A - Conservative GWR model coefficients

Figure 8.1: Conservative model coefficients for percentage of white population (A), pubs/1000 inhabitants, population density (C), average age (D), and average income (E)

8.2 Appendix B - Labour GWR model coefficients

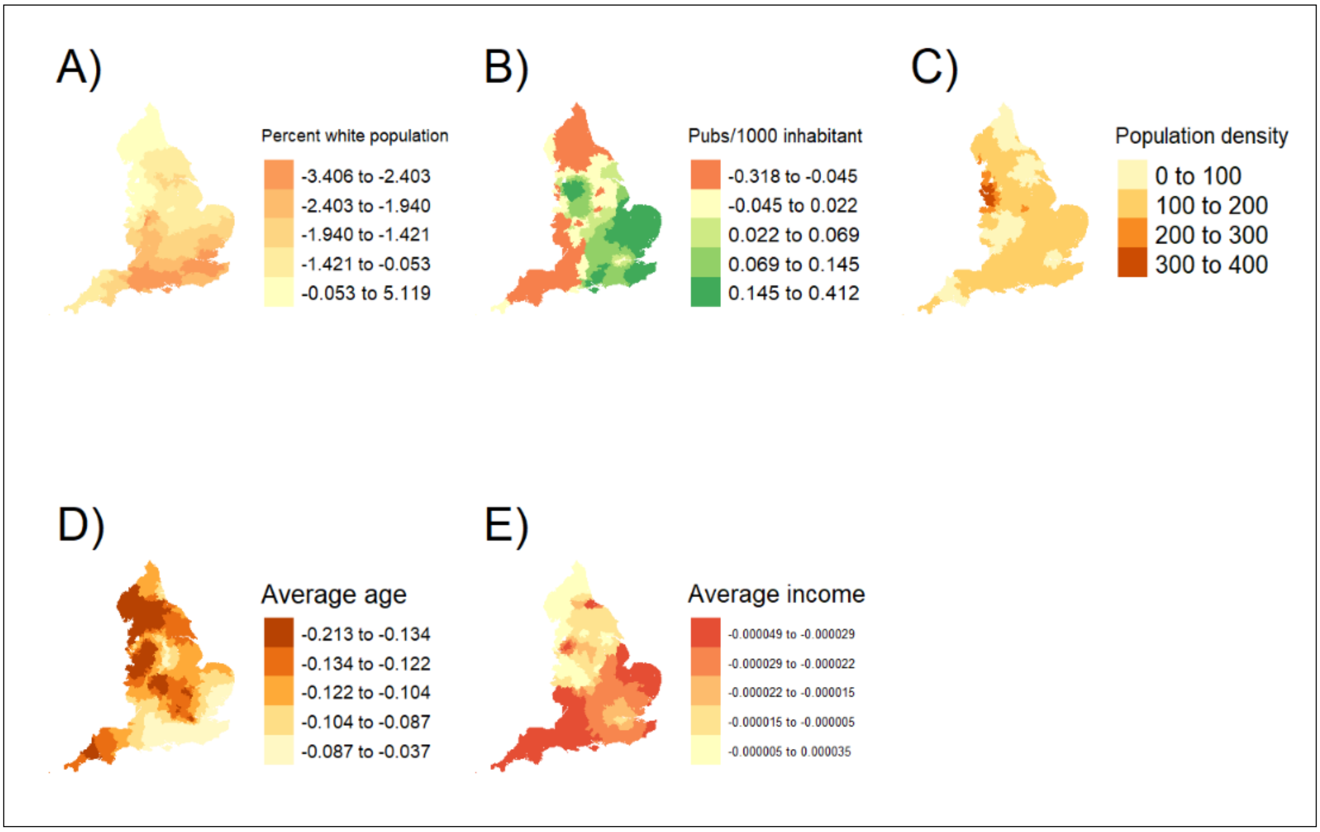

Figure 8.2: Labour model coefficients for percentage of white population (A), pubs/1000 inhabitants, population density (C), average age (D), and average income (E)