Spatially varying explanations behind the UK's vote to leave the EU

Roger Beecham*[1], Aidan Slingsby†[1], Chris Brunsdon‡[2] and Rob Radburn§[3]

[1]giCentre, City, University of London
[2]National Centre for Geocomputation, Maynooth University
[3]Leicestershire County Council

January 19, 2017

Summary

Explanations behind area-based (Local Authority-level) voting preference in the 2016 EU referendum are explored using aggregate-level data. Variables describing the *economic competitiveness* of Local Authorities most strongly discriminate variation in the vote. To a lesser extent this is the case for variables linked to 'metropolitan vs. traditional' *values* – and relationships here tend to diverge spatially. Local Authorities in Scotland and London are distinct in their strong support for remaining in the EU. Whilst the difference for London can largely be accounted for using area-level data, this is less true of Scotland.

KEYWORDS: EU referendum, regression, geographically-weighted statistics, area-based analysis.

1 Introduction

On 23rd June 2016 the United Kingdom held a referendum, for the second time, on its political association with Europe. The outcome of the first, on 5th June 1975, was emphatic, with 67% voting for continued membership of the then European Community (EC). The result of the most recent vote – this time on the European Union (EU) – was more fractured. Whilst the overall outcome was a slight preference for leaving the EU (51.9%), this result varied greatly across the country. Of the 11 regions of Great Britain (GB), only Scotland and London voted in favour of Remain (62% and 60% respectively). At a Local Authority (LA) level, this was the case for just 30% of 378 LAs in GB.

Political Scientists have argued that the spatially-fragmented vote is symptomatic of widening social divisions in the UK linked to structural change (Goodwin and Heath, 2016b). LAs recording the

*roger.beecham@city.ac.uk
†a.slingsby@city.ac.uk
‡christopher.brunsdon@nuim.ie
§robert.radburn@leics.gov.uk

strongest preference for Leave have been described as 'left-behind' places, characterised by chronic low skills and socially conservative values, whilst those associated strongly with Remain with more affluent, highly-educated and diverse populations (Goodwin and Heath, 2016a).

Our analysis combines referendum results data, aggregated to Local Authority (LA) level, with demographic and socio-economic indicators as measured by the 2011 Census. The aim is to identify the most discriminating and generalisable socio-economic variables that account for spatial differences in LA-level voting preference. We are particularly interested in whether these explanatory variables hold equally well across the country. Analysis is structured around the following research questions:

- What factors explain area-based differences in voting preference in the 2016 EU referendum?
- Do these factors explain differences in area-based voting equally well across Great Britain?

2 Data and methods

Results data are published at LA-level by the Electoral Commission [1]. Population characteristics, as measured by 2011 Census, were selected based on the media discourse around *place-based* histories: the varying responses to, and experiences of, de-industrialisation (Cox, 2016) (Table 1).

Table 1: *Proposed 2011 Census variables for explaining Local Authority share of Leave vote.*

variable	justification/theory
degree-educated *professional occupations* *younger adults*	post-industrialisation / knowledge-economy
English speaking *single-ethnicity* *not good health* *white British/Irish* *Christian*	diversity / values
own home *don't own car* *private transport to work*	metropolitan / urban-rural / outcomes

We built multivariate regression models to investigate possible explanatory variables. Variable selection and regularisation was performed using the least absolute shrinkage and selection operation (LASSO) (Tibshirani, 1996). LASSO minimises the sum of squared differences between the observed outcome and the model (e.g. OLS) whilst also penalising against the absolute sum of regression coefficients. This latter constraint – penalising the sum of regression coefficients – enables variable selection as regression coefficients are shrunk to 0 and therefore less important variables are excluded from the model.

Since our research questions are concerned with spatial variation in factors explaining LA-level voting preference, we investigate how well the models hold for different regions of the country. We do this by attending to the spatial distribution of residuals from our global models and using

[1] `http://www.electoralcommission.org.uk`

geographically-weighted summary statistics (Brunsdon et al., 2002). Specifically, we investigate spatially varying relationships by clustering LAs on their geographically-weighted correlation coefficients and identifying groups of LAs exhibiting distinct associations between candidate explanatory variables and the Leave vote.

3 Analysis

3.1 Characterising LA-level variation in voting preference

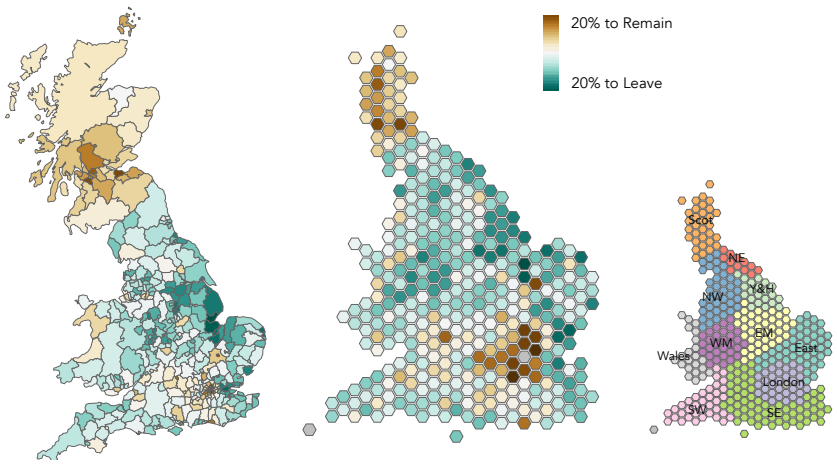

Figure 1: Maps displaying share of vote in favour of Leave (green) and Remain (brown). Results data are displayed using a conventional choropleth map and a hexagonal cartogram (cartogram published by ESRI[2]).

Fig. 1 displays voting preference by LA: intensity of green represents strength of vote in favour of Leave; intensity of brown, strength of vote in favour of Remain. The maps expose a very obvious contrast between most of England and Wales (in favour of Leave) and Scotland and London (in favour of Remain).

3.2 Exploring candidate explanatory variables

The scatterplots in Fig. 2 allow factors driving this LA-level variation in voting preference to explored. Variables associated with post-industrialisation and the knowledge-economy – *degree-educated* and *professionals* – most obviously discriminate differences in voting preference. To a lesser extent are those concerned with *diversity*, *values* and *outcomes*. There is evidence of regional specificity in relationships. Across all variables, LAs in London (purple) and Scotland (orange) can be separated.

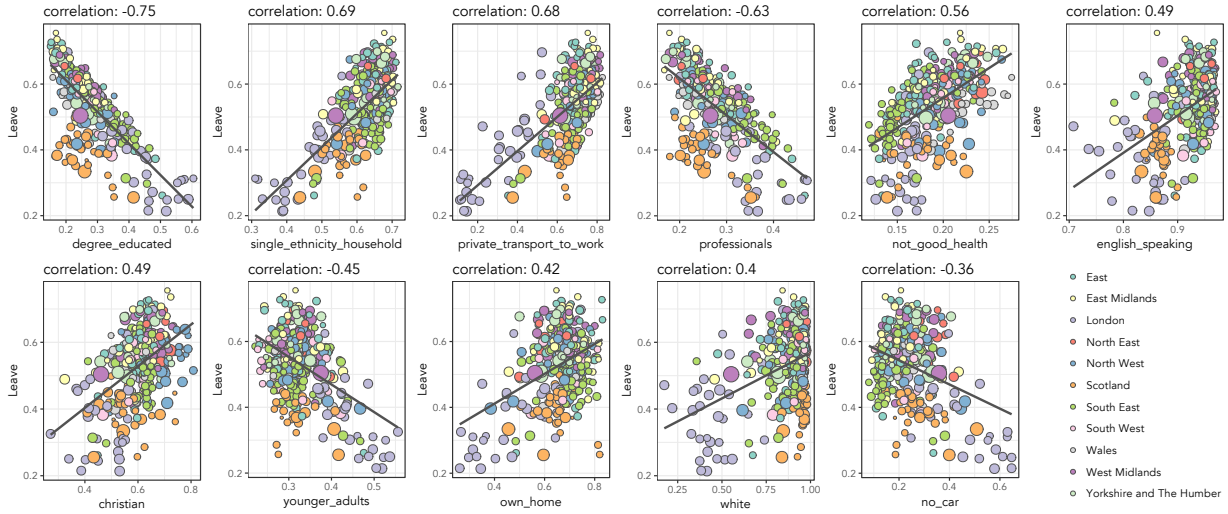

Figure 2: Scatterplots of candidate 2011 Census variables against the outcome variable – % of LA vote in favour of Leave. Points are sized according by size of electorate and coloured by region.

3.3 Developing explanatory models

We consider these relationships more formally through linear regression. Model outputs are presented in Fig. 3. The *degree-educated* variable is most obviously associated with our outcome and itself explains 57% of the variation in voting preference between LAs in GB. This increases to 73% when Scotland and London are excluded from the model. Notice the strong negative residuals for Scotland: our univariate model consistently overestimates the Leave vote given levels of education there. That the sign in residuals reverses at the Scottish border suggests that, rather than some coherent spatial process, Scotland is *categorically* different.

Building a multivariate model, using the variables enumerated in Table 1, and also Scotland and London as dummy variables, we can substantially improve model fit: we more fully explain LA-level variation in the vote. Note that holding *degree-education*, *car ownership* and *Christian* constant, our model estimates that a LA's share of Leave vote reduces by 16% due to it being in Scotland. The equivalent (LASSO) regression coefficient for London is extremely small and in the opposite direction: that is, once we control for the same set of Census variables, we expect a 1% increase in the Leave vote due to a LA being in the London region. This is an important observation. We have identified London and Scotland as special cases in their preference for Remain. Model III suggests that for LAs in London this difference dissappears after controlling for the the socio-economic characteristics (and concepts) identified in Table 1; for LAs in Scotland, these socio-economic circumstances alone do not account for the difference in voting behaviour.

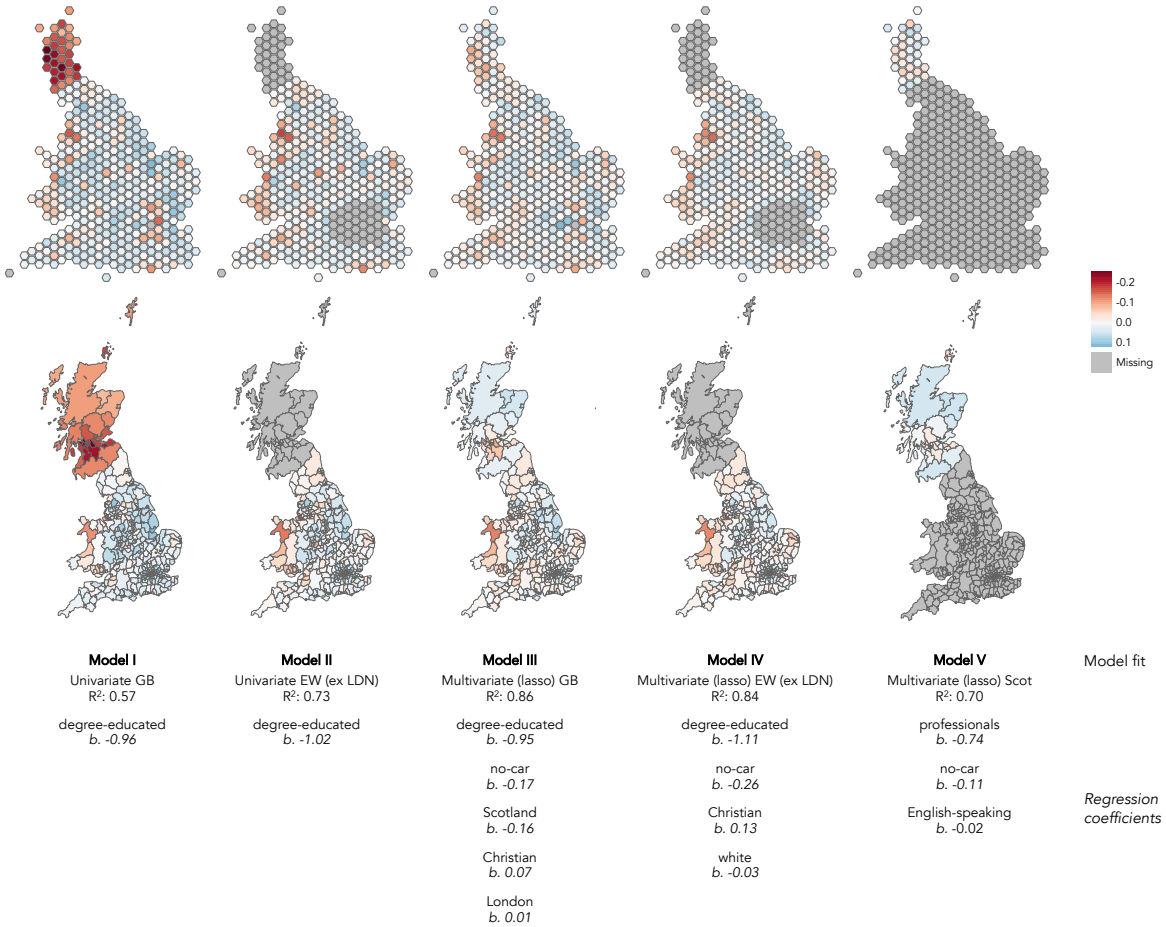

Figure 3: Output from linear models where the *outcome* is level of support for Leave and the *explanatory variables* are those listed in Table 1. The LASSO method was used for regularisation and variable selection (Tibshirani, 1996).

Figure 4: Geographically-weighted correlation coefficients for Census variables against LA share of Leave vote.

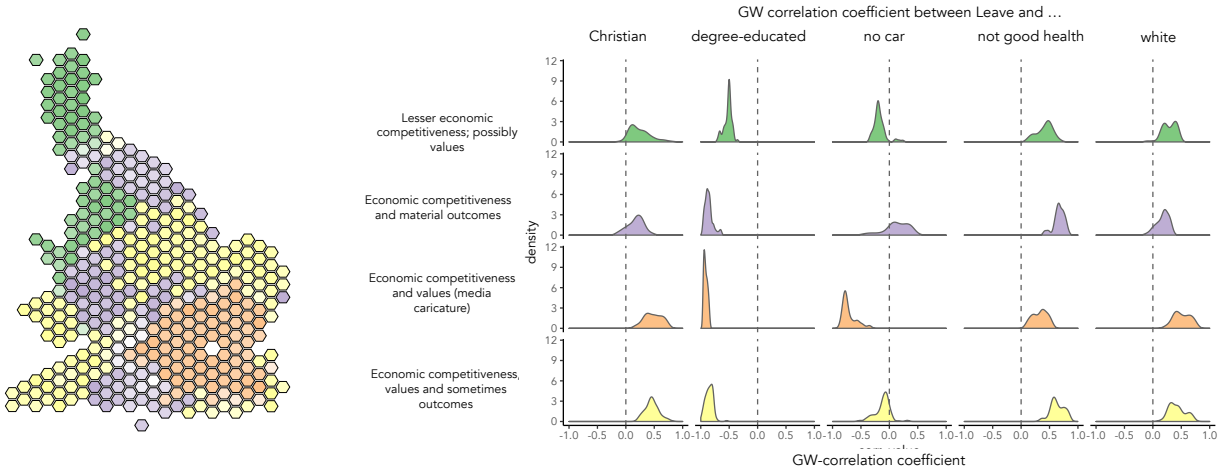

Figure 5: LAs are clustered according to local (geographically-weighted) associations between voting preference and selected variables from 2011 Census. Left: LAs are coloured according to cluster membership and similarity to cluster centre. Right: density plots of variables on which clusters are defined.

3.4 Exploring spatially-varying explanations

It is conceivable that certain relationships with the Leave vote vary spatially in ways not accounted for by global models. Most obviously, the *no car* variable was assumed to represent 'metropolitan' living, but in rural locations it might instead relate to lack of material outcome. This, and the fact that residuals for the multivariate model still exhibit spatial autocorrelation (Moran's I for Model III is 0.53) is justification for investigating locally-varying relationships further.

Fig. 4 displays geographically-weighted correlation coefficients for each population variable against share of Leave vote. We introduce here a further explanatory variable: the number of EU-born residents in a LA expressed as a proportion of the total LA population. The maps confirm that *degree-educated* and *professionals* are strongly negatively correlated with the Leave vote. Although the strength of relationship varies, the direction remains the same. *Private transport to work* and *not good health* are both associated positively with Leave. A cursory glance at Fig. 4 reveals London's unique context. The *no car* variable is very strongly negatively associated with Leave around London; elsewhere this variable is less discriminating and in fact is positively associated with Leave in the North East. A similar pattern, though in the opposite direction, is present in the *own home* variable. The association with variables related to diversity is interesting. *Christians* has a strong and positive relationship with the Leave vote in the East and the Scottish borders and to a lesser extent the South West. This spatial pattern is also true of the *English speaking* and *white* variables. Notice also the pattern in *EU-born*. For most of the country this variable, which might be a proxy for 'diversity', is negatively associated with Leave. For parts of East Anglia and Lincolnshire, where it is argued that population change due to recent EU migration has been more keenly felt (Asthana, 2016), the relationship is reversed.

One means of studying these patterns of relationship more systematically is by clustering LAs on their geographically-weighted correlation coefficients. In Fig. 5 we report on a cluster solution that tries to identify groups of LAs sharing similar combinations of relationship. Cluster memberships are represented spatially and for the variables on which groups are defined. The cluster labels are an attempt to characterise these distributions.

4 Conclusion

We present an initial exploration of themes, presented widely in the UK media (Ford, 2016; Cox, 2016), that might explain the geography of EU referendum voting preference. The most convincing variables are those relating to 'post-industrialisation' and the 'knowledge-economy' – the *degree-educated* and *professionals* variables. Variables linked to 'traditional values' (e.g. *Christian*), 'metropolitan living' and 'material outcomes' are also discriminating and help to further explain variations in voting preference. When these more minor variables are considered, relationships appear to be more spatially-specific. That voting preference can be explained less well in Scotland is an important observation. There may be some unique context that informs attitudes towards membership of the EU for Scottish LAs.

As an aggregate-level analysis, the usual caveats around the *ecological fallacy* (Rose, 1973) apply and it is possible that we overstate the importance of social demographics (measured at the area-level) at the expense of attitudinal explanations (Kaufmann, 2016). Given the democratic *milieu* of the UK, however, the political identities of *places* clearly matters. That the UK continues to experience widening geographic inequalities in outcome (Dorling, 2010), and that links have been made with this and EU voting preference, is further justification for the area-based focus.

5 Biography

Roger Beecham is Research Fellow in Geographic Data Science at the giCentre, City University of London. His work spans several technical and applied areas: spatial data analysis, information visualization, transport planning, economic and social geography and crime science.

Aidan Slingsby is Lecturer in Visual and Analytic Computing at the giCentre, City University of London. He designs, applies and evaluates software for supporting communication and analysis of spatiotemporal data in a variety of application areas.

Chris Brunsdon is Professor of Geocomputation and Director of the National Centre for Geocomputation at Maynooth University, Ireland. As well as developing new approaches in spatial statistics, he has interests in analysing social data relating to crime, health, migration and other related topics.

Rob Radburn is Service Development Team Leader, Business Intelligence Group, Leicestershire County Council. He is responsible for extracting and integrating the Council's proprietary data for analysis. Rob is currently one of 27 Tableau Zen Masters[3].

References

Asthana, A. (2016). Immigration and the EU referendum: the angry, frustrated voice of the British public. [Online; posted 20-June-2016].

Brunsdon, C., Fortheringham, M., and Charlton, M. (2002). Geographically weighted summary statistics: a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems*, 26:501–524.

Cox, E. (2016). Leaving the North behind led to Brexit. Here's what has to happen next. [Online; posted 24-June-2016].

Dorling, D. (2010). Persistent North-South divides. In Coe, N. M. and Jones, A., editors, *The Economic Geography of the UK*, chapter 2, pages 12–28. Sage, London.

Ford, R. (2016). Older 'left-behind' voters turned against a political class with values opposed to theirs. [Online; posted 25-June-2016].

[3] `http://www.tableau.com/ZenMasters`

Goodwin, M. and Heath, O. (2016a). Brexit vote explained: poverty, low skills and lack of opportunities. [Online].

Goodwin, M. and Heath, O. (2016b). The 2016 Referendum, Brexit and the Left Behind: An Aggregate-level Analysis of the Result. *The Political Quarterly*, 87:323–332.

Kaufmann (2016). The 'Left Behind'?: reconciling individual and aggregate UK Independence Party voting. In *American Political Science Association meetings*, Philadelphia, USA.

Rose, D. (1973). National and local forces in state politics: The implications of multi-level policy analysis. *American Political Science Review*, 67:1162–1173.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:267–288.