VISUALISATION

Brexit: surname diversity and voting patterns

Mario Cortina Borja, **Julian Stander** and **Luciana Dalla Valle** examine associations between the results of the EU referendum and the prevalence of unique surnames across the UK

An interesting by-product of the UK's referendum on membership of the EU (page 4) has been the wide variety of excellent data analyses and visualisations to explain and add context to the results (see bit.ly/29W7Glx, for example). However, one of the few aspects that has not been analysed is how surname diversity in districts relates to referendum voting patterns.

Surname distributions are increasingly used in geography, for example, to characterise cultural regions.[1] There are, however, few studies analysing the associations between surname distributions and voting patterns.[2] This is what we set out to do here, using data on surnames and locations from the 2001 UK electoral register and the results of the EU referendum (bit.ly/29W8tCR).

The data sets

Readers may be wondering why our surname data is 15 years old, but there is a good reason. Prior to 2001, the UK electoral register contained the names and addresses of *all* people who were registered to vote in UK elections. In 2001 a change in the law allowed voters to opt out of the publicly available version of the electoral register. This had an effect on the completeness of the data set as the level of opt out for 2007 was estimated at 30%.[3]

The 2001 electoral register is therefore the last version of the electoral register before opting out was possible and is, to our knowledge, the best source of names and locations publicly available. For this article, we used the "enhanced" register, which was supplemented with details of people not registered to vote. An analysis of this data set appeared in *Significance* back in 2008.[4] The enhanced register contains surnames (but no personal names) of 45.6 million people who were resident in the UK in October 2001, sorted according to 434 administrative districts.

The analyses in this article, however, are based on the 372 (out of 399) voting areas that we could directly match to one of these 434 administrative districts. Unfortunately, we were unable to link voting areas in Northern Ireland and Cornwall, for example.

Generally, as the percentage of unique surnames increases, the percentage of Leave voters decreases

Hapax legomena

Although surname diversity has been used as a proxy for ethnic heterogeneity by some authors,[5] we do not advocate this approach here, restricting our attention to surname diversity itself.

We summarised surname frequencies using measures first developed to study vocabulary richness, as distributions of vocabularies and collections of surnames have many similar features. Such statistics are studied by Tweedie and Baayen, for example.[6]

Based on the results from the 2008 *Significance* article, we chose to restrict

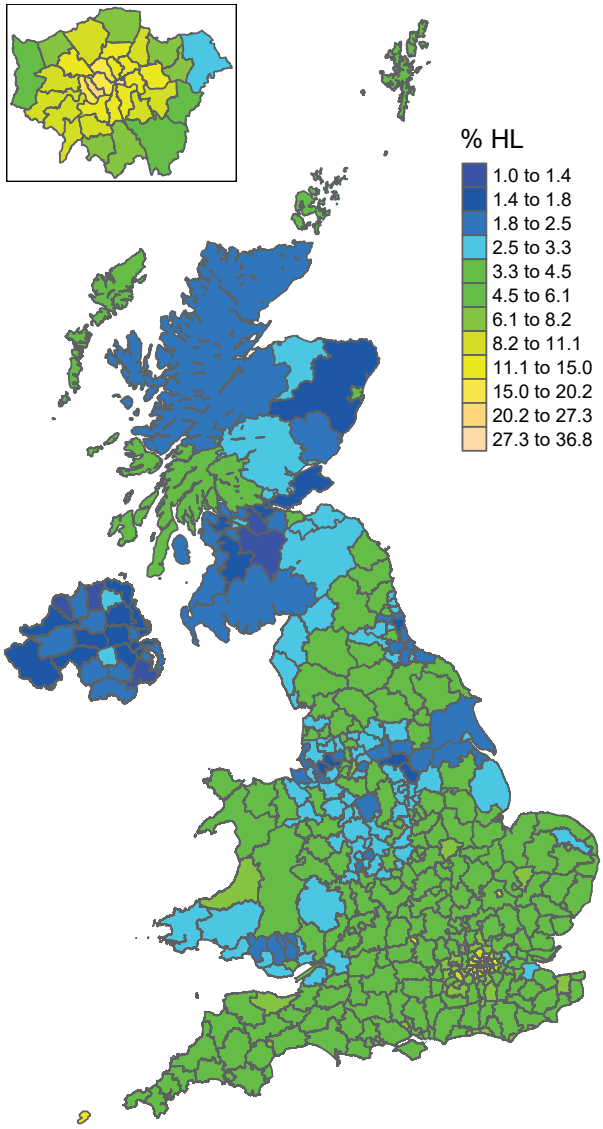

FIGURE 1 Map of the percentage of people with unique surnames in UK administrative districts. Greater London districts are shown separately in the top left corner

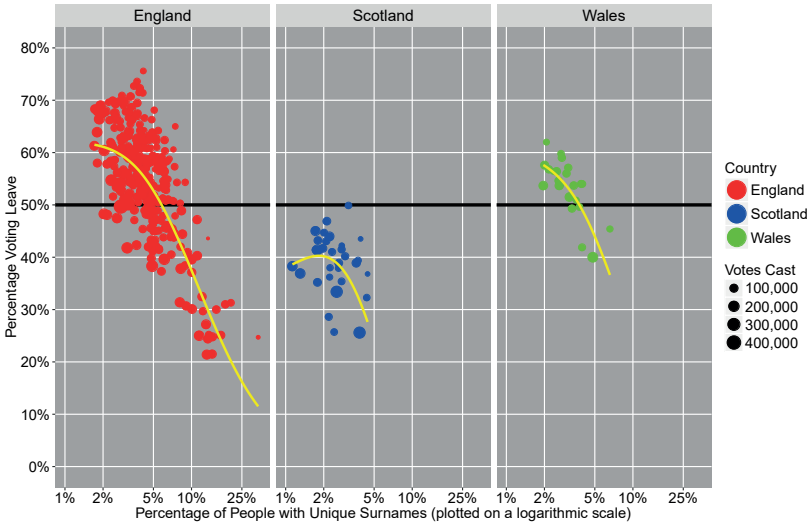

FIGURE 2 The association of the percentage voting to leave the EU with the percentage of people with unique surnames for England, Scotland and Wales

our attention to the percentage of people with surnames that occur uniquely in a district; we call this variable *HL*, after the term *hapax legomena*, which is used in vocabulary distributions to denote words appearing uniquely in a text. Higher values of *HL* indicate an increased number of unique surnames and therefore a higher surname diversity in the population.

Results

Figure 1 maps *HL* for the 434 UK administrative districts (as per the October 2001 electoral register). The map shows, for example, that Northern Ireland, South Wales, parts of Yorkshire and Humberside, and much of Scotland stand out for having more districts with lower percentages of people with unique surnames; conversely, the Scilly Isles, Oxford, Cambridge, and most of Central London are the most heterogeneous districts in the UK.

In Figure 2, we plot the percentage voting to leave the EU against the percentage of people with unique surnames for each of the 372 matching administrative districts in England, Scotland and Wales.

The fact that Scotland favoured Remain, while large parts of England and Wales supported Leave, can clearly be seen. We modelled the percentage of Leave voters as a smooth function of the percentage of people with unique surnames using a different logistic regression model for each country. The yellow lines represent fitted values from the models. We observe that generally, as the percentage of unique surnames increases – indicating higher surname heterogeneity within a district – the percentage of Leave voters decreases. This effect is less strong in Scotland than in the other two countries, possibly due to the generally strong Remain feeling there.

Strategic implications

Our analyses follow the approach of Cheshire and Longley,[7] who emphasised the importance of analysing historic and contemporary surname databases to study "population characteristics, and the long and short term dynamics that characterise population change".

We have seen that the percentage of voters opting for Leave shows some dependence on measures of surname diversity, such as the percentage of people with surnames occurring only once. This suggests that if political strategists on both the Remain and Leave sides had considered surname diversity, they may have had a better idea of where to concentrate their campaign efforts.

This conclusion must be regarded tentatively, however, as our analyses have several drawbacks. First, the surname data used to calculate the diversity data is 15 years old and so there may be a bias towards underestimating surname variability – especially perhaps in areas that may have had higher levels of immigration since 2001. However, a comparison between the top 10 surnames from the 1881 census and the 2001 electoral register showed few changes,[4] and a permanency of the top UK surnames across centuries was also found by Tucker, indicating that any bias coming from underestimating surname variability may be small.[8] Secondly, the matching between previous administrative districts and 2016 voting areas is not precise and may therefore be a source of additional bias.

Nevertheless, our results show how looking at surname distributions can add useful and data-driven insights to political analyses and thinking. We therefore suggest that further use should be made of them in future policy-making. ■

For an extended version of this article, with additional analyses, see significancemagazine.com/surnames

Mario Cortina Borja is chairman of the *Significance* editorial board, and professor of biostatistics in the Population Policy and Practice Programme, Institute of Child Health, University College London

Julian Stander is associate professor (reader) in mathematics and statistics in the School of Computing, Electronics and Mathematics, Plymouth University

Luciana Dalla Valle is a lecturer in statistics in the School of Computing, Electronics and Mathematics, Plymouth University

References

1. Mateos, P. (2014) *Names, Ethnicity and Populations: Tracing Identity in Space*. Berlin: Springer.

2. Cantú, F. (2013) Identifying irregularities in Mexican local elections. *American Journal of Political Science*, **58**, 936–951.

3. Equifax (2007) ER voter opt-out hits record levels. *Data Strategy*, **3**(4), 6.

4. McElduff F., Mateos P., Wade A. and Cortina Borja M. (2008) What's in a name? The frequency and geographic distributions of UK surnames. *Significance*, **5**, 189–192.

5. Lauderdale, D. S. and Kestenbaum B. (2000) Asian American ethnic identification by surname. *Population Research and Policy Review*, **19**, 283–300.

6. Tweedie, F. J. and Baayen, R. H. (1998) How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities*, **32**, 323–352.

7. Cheshire, J. and Longley, P. (2011) Spatial concentrations of surnames in Great Britain. *Procedia – Social and Behavioral Sciences*, **21**, 279–286.

8. Tucker, D. K. (2004) The forenames and surnames from the GB 1998 Electoral Roll compared with those from the UK 1881 Census. *Nomina*, **27**, 5–40.