What’s in a name? A Spatial Analysis of Surnames.

James Cheshire
UCL Department of Geography, james.cheshire@ucl.ac.uk

“It may be thought by some that the investigation of the distribution of names is an idle amusement, productive of no utility of man. I have come to think, however...that it is a matter of much importance to the antiquarian, the historian, the ethnologist and also to the more practical politician.” Henry Guppy 1890.

1. Introduction

A sense of regional identity is important to many in Great Britain, and surnames are inherently regional phenomena. They often originate in specific places, are inherited from fathers, and only move with the people who possess them. Early research showed that in Britain around the year 1000 people's names were unique and that this was not far from the truth. People are significantly less likely to have unusual names today and many are likely to have long histories and complex couplet histories. There is much evidence for the existence of cultural and ancestral heartlands in Britain and surnames provide insights into these. However, much of the research into the nature of these regions has focused on single events, serendipitous datasets, or specific case studies, without regard to robust measurement and comprehensive coverage.

2. The Study

This study seeks to unearth the spatial patterns of surnames in Britain by using the most comprehensive data available. As can be seen in Figure 2 surnames are inherently spatial phenomena. This may seem surprising given the mobility of contemporary society. Undertaking spatial analysis on the millions of surnames in Britain (and beyond) enables us to see patterns that house additional insights into the population. These are often overlooked within population databases because surnames are seen as little more than a means of identification. This poster shows a selection of family name analysis conducted by UCL Department of Geography.

3. Data

The principal analysis was undertaken using the British 2001 Enhanced Electoral Register (EER). It contains the names and addresses of residents aged 17 or over who eligible to vote. To include non-adults, the EER also contains people sourced from commercial surveys and crime recording databases. The EER contains the names of 45.6 million people. The majority of registered names in Britain (7.45 million of them) have less than 10 occurrences with very few names exceeding 100,000 occurrences. This highly skewed distribution has been represented in Figure 1. The surname data are obtained from a mixture of datasets, conducted through a subtle colour change from north to south England. Most interesting are the similarities between many of the large urban areas. The result presents a way the intrinsic diversity to be made between the populations of these areas. For example south and west London are more similar to Manchester whilst north east London has more in common with Leicester. Figure 4 illustrates the analysis at European level.

4. Methods

Creating a distance matrix is the most effective way of comparing the surnames found in different areas. Every area is listed as a row and column and the matrix populated with a distance measure that compares each area (all figures they are wards with all others. For figure 1 however the mapping was approximately 10500 (the number of wards) by 10500, producing a matrix of approximately 10500 x 3. Multidimensional scaling (MDS) was selected to do this as it maintains the concept of distance between areas. Each of the three dimensions can be thought of as a location in 3D space with areas further apart being more different. Two of the dimensions are plotted in Figure 4 and 5 represent all 3 dimensions by treating each as a red, green and blue value that can be combined to produce the colour used in the maps. Each area has therefore been assigned a colour that directly corresponds to its MDS coordinates. The areas given similar colours will therefor have similar surname compositions and vice versa.

Figure 2 (above right): Surname density plots of 9 example surnames which demonstrate their uneven distribution from east to west (British National Grid). It is clear that Welsh (Wel) and Cornish (Cor) names are most frequent in the west, whilst English (Eng) and Scottish (Sct) names have more even distributions. This highly skewed distribution has been represented in Figure 1. The surname data are obtained from a mixture of datasets, conducted through a subtle colour change from north to south England. Most interesting are the similarities between many of the large urban areas.

Figure 3(right): Heat showing two of the three MDS dimensions from the Laser Distance matrix. The data have been partitioned into the 9 Government Office Regions (GORs). Hexagonal binning has been used with the number of points per bin shown by the scale bar.

Figure 4 (below): The surname regions of Great Britain fit with similar colours to show those of the same surnames. The opposite is true for areas of different colours. The spatial unit of analysis is the Census Area Statistics Ward. There are approx. 15500 of these in Great Britain.

5. Results

It is clear from Figures 3 and 4 that surname regions exist in Great Britain. The differences between Scotland, England and Wales are clear, although the transition from England to Scotland is far more abrupt than from England to Wales. Distance decay is evident through a subtle colour change from north to south England. Most interesting are the differences between many of the large urban areas. The result presents a way that Belgium, Switzerland and Luxembourg subdivide along linguistic lines. The surname regions of Great Britain.

Figure 5 (Right): The surname regions of Europe. Much of this map is unsurprising but there are some interesting patterns. For example the south coast of Spain appears to differ from the rest of the country, with some parts showing a closer resemblance to the UK. The contrast between Norway and Sweden is also interesting as is the distinction between France, Switzerland and Luxembourg subdivide along linguistic lines.

6. Applications

Spatial analysis of surnames has many applications, Geneticians, for example, are keen to know about surname regions because of the link between surname and chromosomes (both are patrimonial). One can also infer the degree of diversity in a population. Areas with many surnames relative to their population size are likely to have had more intermarriage and experience greater social, cultural and genetic diversity than those areas with fewer surnames. In addition to the broad scale observations of Figures 3 and 5 demonstrates the utility of surnames for fine scale analysis in London. Populations can be subised into ethnicities beyond those listed in censuses. Each group may have important spatial characteristics that are lost in the aggregations of groups listed in official government statistics.

Further Reading


Acknowledgements

Thanks go to Pablo Mateos, Paul Longley, Alex Singleton, Muhammad Adnan H.A., Mohammad Naderi, Iman Shirvani, East Midlands, and UCL Department of Geography. This poster was produced using Wordle (wordle.net). UCL corporate colours have been used throughout. This study was undertaken in collaboration with ESRI UK and ESRI US and supported by the ESRI Development Center.

www.namesnetwork.org.uk
www.publicprofiler.org/worldnames
www.ordnancemap.org
spatialanalysis.co.uk
twitter: @spatialanalysis

This poster... seeks to provide insight into the geography of surnames. The intention was to create a simple layout combined with eye-catching images to draw attention and clearly display information. The image below was produced using Wordle (wordle.net). UCL corporate colours have been used throughout. The words used to produce the image were taken from the EER, and were selected to reflect the areas of the population.

www.ordnancemap.org
spatialanalysis.co.uk
twitter: @spatialanalysis