Combining historic interpretations of the Great Britain population with contemporary spatial analysis: The case of surnames

3 authors, including:

Pablo Mateos
Ciesas Occidente
71 PUBLICATIONS 1,177 CITATIONS

Some of the authors of this publication are also working on these related projects:

Names and ethnicity - Onolytics View project

Central America-North America Migration Dialogue View project

Combining Historic Interpretations of the Great Britain Population with Contemporary Spatial Analysis: the Case of Surnames.

James Cheshire, Paul Longley, Pablo Mateos

Department of Geography, University College London

james.cheshire@ucl.ac.uk

Abstract

This study applies contemporary geospatial analysis to the 1881 Census of Great Britain to establish, through a process of inductive generalisation, the surname regions in 19th century Britain. All surnames in the census were geo-referenced to the 662 Registration Districts listed. The districts' surname compositions were compared using the Lasker Distance calculation. The resulting similarity values, assigned to each district, were grouped according to Ward's hierarchical clustering. By mapping the cluster allocations a clear regionality, characterised by a high spatial autocorrelation value, emerges that closely matches the historical attempts to establish the surname regions in 19th Century Britain. As one of the first studies of this kind to utilise a complete population register for 1881, with results aligned to historical interpretations, one can think of the maps presented here as a baseline for the regionality of surnames in Great Britain, against which the effects of more recent migrations can be compared.

1. Introduction

The contemporary explosion in data availability has not been limited to present- day sources; historical datasets are also being digitised and make richer and generalisable spatial analysis more possible than ever before. This study takes one such historical dataset- the 1881 Census of Great Britain- and applies contemporary geospatial analysis to establish, through a process of inductive generalisation, the surname regions in the 19th century. The results of this analysis are used alongside literature published around the time of the census to develop a baseline regionality of surnames in Great Britain, against which the effects of more recent migrations might be compared.

The basis of surname formation, combined its hereditary nature in patronymical societies such as Great Britain, render it, in most cases, inherently spatial. Toponyms, that is to say surnames derived from places or geographical features, for example "Rivers" or "Cheshire", are the most obvious manifestations of the geographical influences on surname acquisition. Other naming conventions also point to the origin of the name. There is a dominance of patronymic names in Wales, for example "Johns" meaning son of "John", and a prevalence of tribal names, such as "MacBain", in Scotland. The surnames of other regions may be characterised by a small number of first and last syllables for example Cornish names are often identified as containing "*Tre, Pol and Pen*" [1].

This association between locality and surname formation is tempered by those names derived from characteristics common throughout the country, and are therefore likely to have been coined in multiple locations. In this category are names referring to common occupations, such as "Smith", or some nicknames, such as "Fox" and "Little". In the latter category there remains some geographical influence: a relatively high number of Norfolk surnames, for example, are taken from people's characters in pageants (such as "Pope").

The regionality of surnames is in large part an effect of the different naming conventions from the many invaders and migrants that came to make up the British population prior to surname formalisation. The influence of these groups has been determined by much painstaking work from the likes of Lower [2][3], Guppy [4] and Baring-Gould [5]. These early formalisations of the nomenclature and etymology of surnames were explicit about the importance of location in naming. This was taken to a sub-regional level with, for example, a contribution to the Transactions of the Royal Philological Society on "The Races of Lancashire as Indicated by the Local Names and Dialects of the County" [6]. Lower [2] was keen to stress that the subject of the local distribution of surnames is one that deserves more attention- a statement that remains true today [7].

Presented here is one of the first attempts to create an impression of the regionality inherent within the 1881 Great Britain population, through a series of aggregation and spatial analysis techniques not possible before the digitisation of the 1881 Census. This data source (courtesy of the UK Data Archive) has been geo-referenced to the 662 Registration District boundaries digitised by the Great Britain Historical Geographical Information System project (downloaded from, http://edina.ac.uk/ukborders/). It forms one of the earliest nationally representative population registers of Great Britain. Previous analysis has been constrained by the practicalities of handling hand-written census returns.

2. Methodology

Within spatial science and geography there has hitherto been little interest in characterising regions by their surname composition. The subject has received most attention from geneticists who have analysed contemporary datasets, such as marriage records and telephone directories, to distinguish the regional characteristics of an area from its surnames (see [8] for a recent review). Surname studies within genetics and more widely in human biology are based on the principle that to the extent that two individuals with the same surname are ultimately to share the same lineage, isonymy indicates biological relatedness [9]. Isonymy can be thought of as the likelihood of encountering identical surnames in the ancestors of a couple [10]. The hereditary nature of surnames and their tendency to remain highly concentrated in their areas of origin are the two traits most utilised by geneticists and population biologists to make inferences about the genetics of a population.

The Lasker Distance, derived from a coefficient of isonymy (Equation 1), is a measure that determines the similarity in surname structure between two geographical areas [8]. The coefficient of isonymy compares the probability isonymy occurring between members of two geographic areas [11].

It is calculated as:

$$R_i = \frac{\sum(S_{i_1}S_{i2})}{2\sum S_{i1}\sum S_{i2}}$$

(Equation 1)

where S_{i1} is the number of occurrences of the ith surname from area 1 and S_{i2} is the number of occurrences of the same surname from area 2 [9]. In order to compare the surname characteristics of each location relative to all the others in a country, the Lasker coefficient (R_i) values are converted to a distance measure. The Lasker Distance (Equation 2) measures the isonymic distance between each pair of locations. This metric produces a distance matrix between all pairs of locations in the study area, in which distances are symmetric about a zero principal diagonal. The Lasker Distance is calculated as:

$$L_{ij} = -\ln(2R_i)$$

(Equation 2)

where L is the Lasker distance and i and j are two separate areas. One can think of the Lasker Distance as a measure of similarity, or difference, between two areas in surname space [12]. The greater the Lasker Distance the less similar the composition of surnames in the two geographic areas. Table 1 is a subset of the resulting Lasker Distance matrix. From this, one can see that Yeovil has a more similar surname composition to Aberayron than it does to Aberdeen.

The Lasker Distance values in the matrix were partitioned into 15 groups using Ward's Hierachical Clustering [13]. Linking the cluster allocations assigned to each registration district to the 1881 Census boundaries enables mapping of the spatial distributions of these clusters. The resulting map is shown in Figure 1.

Table 1. Subset of the Lasker Distance Matrix.

District	Yarm.	Yeovil	York
Aberayron	6.38954	6.28992	6.43836
Aberdeen	6.35615	7.01935	6.21322
Abergavenny	6.41289	6.36175	6.56671

3. Discussion

The most notable aspect of the mapped results is the high spatial autocorrelation of the resulting cluster allocations for 15 clusters (Moran's I Index= 0.44, over 99% confidence (calculated in ArcGIS 9.3)). This is not surprising because, as we have already established, surnames are not spatially random phenomena.

Of more interest is the alignment of the resulting surnames with the regions proposed by early authors and derived mainly from historical and anecdotal evidence. In the opening pages of his book *Homes of Family Names in Britain,* Guppy [4] outlines his impression of surname regions. Based on his description, these have been approximately drawn and overlaid on Figure 1. One can see a close resemblance between the Ward's clustering results (dividing Britain into 15 clusters), with transitions occurring along the majority of Guppy's borders. This result is interesting because Yeoman were the only group studied in Guppy's work on account of their "stationary habits and purity of extraction" [14]. By using a full population register, this study suggests that "stationary habits" were common throughout the population, and not just Yeoman as Guppy believed. The agreement between Guppy's interpretation of British surname regions and those of this study is reassuring and demonstrates the utility of inductive generalisation in this context.

A second example is that of the Danelaw Line (see Figure 2). The Danelaw Line is used to mark the southern extent of Danish rule in England during the 9th and 10th centuries [15]. To the north of this line, it is likely that there was some integration of naming practices between the Danish and native populations within Danelaw. Evidence suggests that the spread of Danish names south of Danelaw only took place through the land owning elite and would therefore have had a minor influence on the broader population [16]. The strong correspondence between Danelaw and the composition of British surnames, illustrated by Figure 2, is important because it suggests for at least 700 years, Great Britain's surname geography remained unchanged.

In contrast, Lower [17] was concerned that the "locomotive character of the present age "was doing much to "fuse all provincial peculiarities and distinctions". Both he and Guppy [4] encourage "competent observers in various parts of the kingdom to record the *habitats* (original emphasis) of particular names ere the opportunity now existing shall have

passed away" [17]. The digitised 1881 Census has recreated the opportunity to produce a baseline Great British population with all its provincial distinctions that can be compared with more recent surname records to determine the effects of migration. Longley et al. [18] have, for example, identified Cornish migration to Middlesbrough through the use of surnames, whilst we have unearthed significant anomalies in the town of Corby's contemporary surname structure created by Scottish immigration throughout the 19th Century.
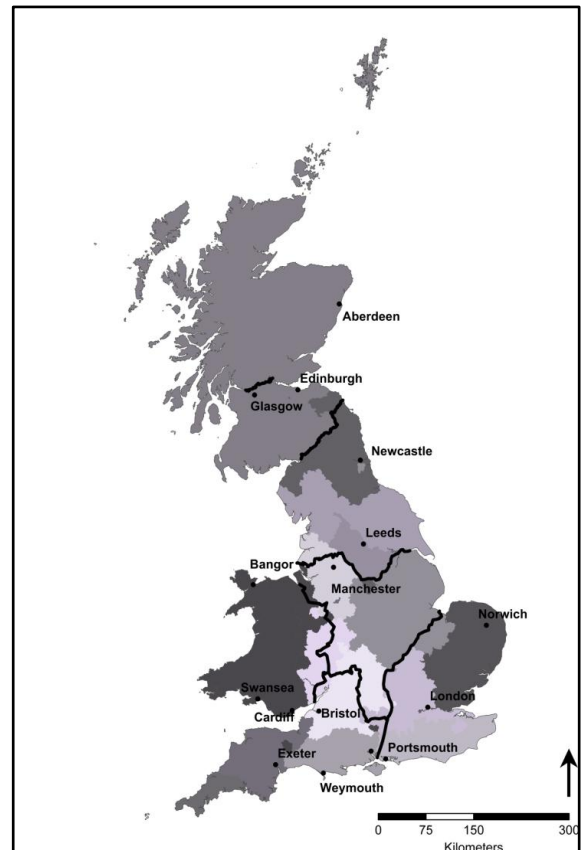

Figure 1. A map illustrating the surname regions of Great Britain in 1881. Each of the 15 cluster allocations has been assigned a unique colour. The black lines are digitised versions based on Guppy's (1890) description of suggested surname boundaries. Full colour available online: http://jamescheshire.co.uk/surnames.

4. Conclusions

The methodology presented here takes full advantage of the digitised version of the 1881 Census by undertaking quantitative spatial analysis to determine the surname regions of Great Britain. The regions highlighted in this process give an inductive impression of "what was" by analysing the most complete record of the population. This approach contrasts with previous surname research that bases inferences on the regionality of British surnames from a sample of names augmented by historical and contextual knowledge. The synthesis of approaches provides a powerful insight into the regionality of British surnames and the population they represented in 1881. The combination of data, methodological framework and availability of historical texts has provided the opportunity to create a baseline surname population that is likely to have remained unchanged since the Norman invasion, creating the bedrock over which subsequent migrants have deposited their own names. This knowledge will help geo-genealogists, cultural geographers and those with a passing interest in their heritage to constrain their searches to specific geographical regions.
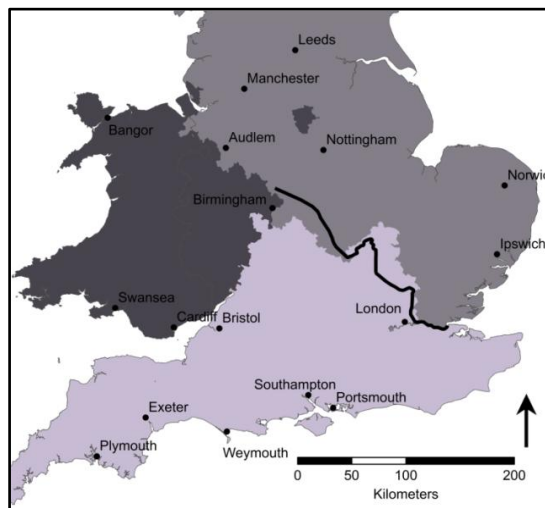

Figure 2. England and Wales mapped based on 3 cluster allocations. The black line represents the southern extent of Danish rule in the 9th Century. It demonstrates a clear correspondence between the north/ south split in British surname regions and the Danelaw line. Full colour available online: http://iamescheshire.co.uk/surnames.

5. References

[1] J. Thorold. *The Wreath of Heraldry.* H. George, Bath. Date Unknown. P50.

[2] M.A. Lower. *Patronymica Britannica*. J.R. Smith, London. 1860.

[3] M.A. Lower. *English Surnames*. J.R. Smith, London. 1842.

[4] H. Guppy. *Homes of Family Names in Britain.* Harrison and Sons, London. 1890.

[5] S. Baring-Gould *Family Names and their Story.* Seeley and Co. Limited, London. 1910.

[6] J. Davies. On the races of Lancashire, as indicated by Local Names and the Dialect of the County. *Transactions of the Royal Philological Society.* 1855. 13: 210:284

[7] W. Zelinsky. Along the Frontiers of Name Geography. *Professional Geographer.* 1997. 49, 4: 465-466.

[8] S. Colantonio, G. Lasker, B. Kaplan, V. Fuster. Use of Surname models in Human Population Biology: A Review of Recent Developments. *Human Biology.* 2003. 785-787.

[9] G. Lasker. *Surnames and Genetic Structure.* Cambridge University Press, Cambridge. 1985.

[10] G. Lasker. The occurrence of identical (isonymous) surnames in various relationships in pedigrees: A preliminary analysis of the relation of surname combinations to inbreeding. *American Journal of Human Genetics.* 1968. 20:250–257.

[11] W. Fox., G. Lasker. The Distribution of Surname Frequencies. *International Statistical Review.* 1983. 81-87.

[12] A. Rodriguez-Larralde., C. Scapoli., M. Beretta, C. Nesti, E. Mamolini, I. Barrai. "Isonymy and the genetic structure of Switzerland. II. Isolation by distance." *Annals of Human Biology.* 1998. 6: 533-540.

[13] J. Ward. "Hierachical Grouping to Optimize an Objective Function". *Journal of the American Statistical Association.* 1963. 58, 301:236-244

[14] H. Guppy. *Homes of Family Names in Britain.* Harrison and Sons, London. 1890. Page 2.

[15] H. Darby. *A New Historical Geography of England.* Cambridge University Press, Cambridge. 1973.

[16] S. Keynes. 'The Vikings in England c.790-1016'. In Sawyer, P.(ed). The Oxford Illustrated History of the Vikings. 48-82. 1997.

[17] M. A. Lower. Patronymica Britannica. J.R. Smith, London. 1860. Page xxvii.

[18] P.A. Longley, R. Webber, D. Lloyd. The Quantitative Analysis of Family Names: Historic Migration and the Present Day Neighbourhood Structure of Middlesbrough, United Kingdom. *Annals of the Association of American Geographers.* 2007. 97, 1: 31-48.