Surnames as a quantitative evidence resource for the Social Sciences
Daryl A. Lloyd, Richard Webber and Paul A. Longley

Centre for Advanced Spatial Analysis, University College London,
1-19 Torrington Place, Gower Street,
London WC1E 7HB
Tel: +44 (020) 7679 4272   Fax: +44 (020) 7679 4293
Email: d.lloyd@ucl.ac.uk, richardwebber@blueyonder.co.uk, plongley@geog.ucl.ac.uk
URL: www.casa.ucl.ac.uk

1. Introduction

Almost everyone in Britain possesses a surname. In England it is thought that the most of the population gained their surnames somewhere between 1100 and 1300 AD (McKinley, 1975), whilst in the rest of the United Kingdom and Northern Ireland, this took place between 1250 and 1400 AD. The main purpose of surnames was to distinguish between individuals with shared common names such as Paul, Richard and Daryl.

2. GIS and the geographical origins of surnames

Most surnames give some indication of the geographical origin of a person's ancestors (Reaney and Wilson, 1997). In the north-east of England, for instance, in an area falling between the Humber and the Tyne, it was common for people to be named after their father. These names, such as Thompson, Davidson and Donaldson, are called *patronyms*. In south-west England and South Wales, on the other hand, rather than place 'son' at the end of the father's name, the letter 's' simply sufficed, producing Williams, Roberts and Davies.

Other types of surname origins, including *metonyms*, the assignment of names based on occupation and *toponyms*, where the surname follows location of origin, also regularly display distinctive spatial patterns (Fellows-Jenson, 2002).
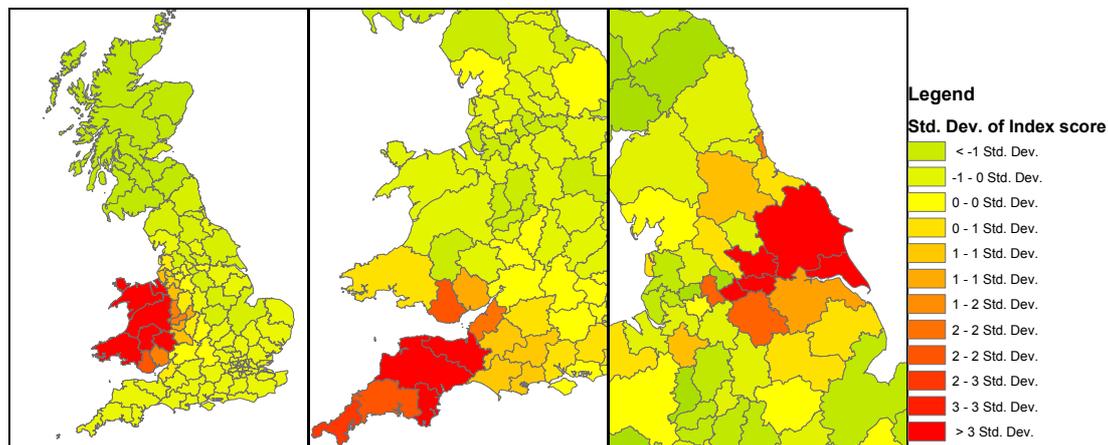

Figure 1. The distribution of the surnames of the authors, Lloyd, Webber and Longley, in 1998

By using GIS various statistical measures of dispersal can be derived and studied to allow understanding of these spatial patterns. Example applications are provided in Section 5, and cover a wide range of possible topics, including analysing the geographical origins of surnames, or for looking at movement of names between different points in time. Figure 1 demonstrates an

example of this, showing the distributions of the authors' surnames in 1998, showing the different spreads and concentrations of Lloyds, Webbers and Longleys throughout Britain.

3. Databases of surnames
An ESRC-funded project at CASA is using inventories of names from a number of Anglophone countries to create a database that will facilitate the geographical analysis of surname patterns both within Britain and internationally, between English speaking countries. This project is developing important substantive findings, and will also provide an important resource for use by other GIS researchers.

For Britain, inventories are available both from the 1998 Electoral Register and from the 1881 Census of Population. Contemporary inventories are available for: the United States of America; Canada; Australia and New Zealand. For each of these countries data are available on the distribution of names by levels of spatial resolution which are roughly comparable.

Each of the names in the database is classified by its origins – in terms of both country and type, be it patronym, metonym, regional, and so on. This allows names of similar linguistic or geographical origin to be grouped together in order to facilitate aggregate analysis. As more names are extracted from a range of databases, this classification becomes larger and more significant, and is becoming a highly useful data source in its own right.

A website dedicated to informing possible users of the data has been set up, providing metadata on each data source, alongside the main classification and indicating the local or regional origins of most names.

4. Statistical measures of surname distribution
The 1998 British data are georeferenced to unit postcodes, allowing aggregation to any level of postal units. However, the 1881 data are based on a parish level, the boundaries for which do not exist digitally. (They would also take such a significant amount of time to produce, and are suitable for only a limited range of end uses.) However, by assigning 1881 parishes to current postcode areas (such as PL. EX, TQ) we are now able to analyse changes over time in both the concentration and the geographic distribution of names and classes of name.

From these analyses it has been possible to show that certain names (and classes of name) are much more concentrated geographically than others and that certain parts of the country (such as the Stoke, Halifax and Dudley postal areas) have much higher proportions of residents with indigenous local names than do others (such as Oxford, Portsmouth or Kingston)

This analysis has thrown up a number of interesting methodological issues, such as for example how one can compare the relative concentrations of names of very different national frequencies and how one can make legitimate comparisons between postal areas of different sizes and populations in respect of the degree to which their names are concentrated.

As with many other GIS datasets, the statistical analysis of surnames in areal units (be they census- or postcode-based) suffers from the modifiable areal unit problem. This is compounded when taking into account both the variability of a surname and the base population. To compare areal units, indices of surnames are produced, where a score of 100 indicates the number of persons with any given surname based on the overall ratio of population to surname count. Areas which have a surname count which is higher than expected will score greater than 100, whilst

areas which are more barren score lower. However, very common names, such as Smith, rarely deviate significantly from 100, whilst names that are much rarer and are highly regionised, such as Midgley, have a higher variability and are not unusual in scoring in excess of 1000 on the index. If, for a name such as this, two areal units are merged together, one of which contains an above average index score, and one of which contains a below average index, then the significance of the 'pocket' of Midgleys will be lowered. Therefore the index is a function of both variability of the name and the population of the areal units into which they fall.

5. Applications of a GIS-able surnames database

Traditionally the study of surnames has been the specialist preserve of genealogists and local historians. With access to our database we believe that the range of applications will increase. For example the database has been used to examine differential patterns of migration from Britain to Australia; to examine the patterns of Cornish and Irish migration into Cumbria and to identify local areas, which on account of their high proportions of people with local provenance, are likely to have high levels of social capital and which might possibly therefore warrant different policing styles. The relationship between surname diversity, health and economic vitality is a field worthy of further research whilst the database has also proved useful in applications such as the analysis of gambling and the reasons for different levels of segregation of the Asian community in different parts of Britain.

Migration studies are facilitated by the ability to compare distributions in 1881 and 1998. For instance, did the peaks for highly regionised names that existed in 1881 subside by 1998, under the cumulative effects of migration? Previous work has attempted to identify sharp local difference in the name pattern of adjacent areas, in order to identify natural regional divides – some of which (such as the Lake District) represent environmental barriers, others (such as the upper Trent) of which have come about through social movement, forming 'cultural' rather than physical divides.

6. Consolidation

This paper outlines the construction of a new GIS-able digital infrastructure for the analysis of surnames in Britain and elsewhere in the English-speaking world. It identifies some of the wide range of possible applications for geographers, historians, genealogists and migration specialists, as well private industry. Furthermore, it is something for which everyone will have a reaction, be they lay-person or GIS professional. Everyone has their own name, and all are interested in where it came from and what has happened to it.

7. Acknowledgements

8. References

Fellows-Jenson, G. (2002), Variations in naming practice in areas of Viking settlement in the British Isles. In Postles, D. (ed.), *Introduction to names, society and regional identity* (London: Leopards Head Press).

McKinley, R.A. (1975), Norfolk and Suffolk surnames in the Middle Ages, Phillimore, London.

Reaney, P.H. and Wilson, R.M. (1997), A dictionary of English surnames 3rd ed. (Oxford: OUP).

Biography

Daryl Lloyd is completing a doctorate at University College London on town centre definition. His research interests are around based urban and quantitative socio-economic geography, as well as the geographies of literature. He desperately plays down his Welsh roots and hopes that Lancashire will win the County Championship this year! Richard Webber is Visiting Professor at the Centre for Advanced Spatial Analysis (UCL), where Paul Longley is Professor of Geographic Information Science.