

## Surnames and the search for regions

**K. Schürer**

The author has previously served as Secretary to *LPS* and is currently Chairman. He is Director of both the UK Data Archive and Economic and Social Data Service and is also Professor of History at the University of Essex.

### Introduction

The parish has undeniably been an important point of reference for past populations.<sup>1</sup> A place to worship, a place to wed, a centre of jurisdiction, and for those unfortunate enough to fall upon hard times, a place of settlement, welfare and charity. Arguably the parish was less important in the lives of urban populations than rural<sup>2</sup> and equally diminished in importance from the eighteenth century as the pace of urbanisation quickened, yet this should not detract from the presence of the parish in the lives of many. The parish has also been the focus of observation for many articles published in this journal. Hardly an issue passes by without one or more articles examining some aspect of the demographic past of a particular parish. This, of course, is a reflection, perhaps more so than anything, of the administrative processes that led to the creation of the primary sources through which historical demography is chiefly explored, namely parish registers and to a lesser extent census enumerators' books. However, although the parish is a very practical, and in many cases, convenient unit of study, and although it was a geographical focus to many important activities within an individual's life, to what extent was the parish subordinate to a wider region when it comes to attempting to understand how populations in the past interacted?

Arguments to study past populations in a regional rather than parochial context are not new. In particular Marshall has long argued for the need to 'find out how contemporaries formed their allegiances to particular districts'.<sup>3</sup> Marshall is in many respects concerned with social and economic activities rather than demographic analysis *per se* but it is realistically impossible to separate the timing of demographic events from the socio-economic contexts within which they occur, which in turn are strongly influenced by place. Such sentiments are clearly echoed in Woods's advocacy of what he has termed 'spatial demography'<sup>4</sup>

In recent years Pythian-Adams, the former Head of the Department of English Local History at the University of Leicester, has argued for a greater emphasis to be placed on the identity of cultural regions in the past. In particular Pythian-Adams has called upon historians to move away from single parish or community-based research to a more regionally-orientated approach, stating that 'there may exist regionalized social systems that, over significant stretches of time in pre-modern periods at least, involve personal interactions more within broadly recognizable territories than between such territories'.<sup>5</sup> To enable the reconstruction of such territories Pythian-Adams has called for the 'measurement and understanding of such systems and their inter-connectedness in terms of actual people living on the ground in the past'.<sup>6</sup> In short, Pythian-Adams has urged historians to use the 'cultural province' as the territorial basis of their studies.<sup>7</sup> In a similar vein, stressing both the importance of regional identity in the past and the need to study territories with which individuals would themselves have identified, Hey has stressed the importance of examining regions through the notion of 'countries'.<sup>8</sup> This, he shows, was a term often used in the nineteenth century to describe 'a district to which people felt that they belonged, one which could evoke sentimental feelings amongst those who had moved away'.<sup>9</sup> To this he adds 'the word "country" had no precise meaning, but it is nevertheless a useful concept for local historians, for it conveys a sense of a local society much wider than that of a town or a rural parish, but usually smaller than that of a county, a sense of the district which people felt was inhabited by their relations, friends and fellow workers and which had a character all of its own'.<sup>10</sup>

However, whatever they are called – regions, countries or cultural provinces – identifying the geographical territories 'to which people felt that they belonged' is no easy task. The problem is heightened by the fact that the cultural territories by which individuals identified themselves may have differed according to a person's class or situation. Pythian-Adams has argued that such regions of 'human activity' should be bounded 'preferably in more or less coincidental physical and cultural terms'. In seeking to find a geographical entity that matches this requirement, the answer, he suggests, may lie in the country's drainage system: 'the great centrally-focused river-drainage basins on the one hand or, on the other, those de-centralized but localized groups of broadly parallel or slightly convergent rivers that are delimited inland in each case by the same watershed line, and which share an identifiable stretch of coastline at their points of outlet'.<sup>11</sup> An alternative approach to the study of historic regions has been to focus on a particular region and examine a range of different sources such as diaries, journals and account books in an attempt to reconstruct patterns

of regionally-based activity.<sup>12</sup> However, due to the idiosyncrasies of the sources used, this approach cannot easily be replicated for comparative purposes for other parts of the country. In other words, it is neither practical nor possible to undertake such studies at a national level. In order to overcome this problem it is necessary to turn to an historical source which is available in a standard form across the entire country.

In an attempt to overcome this problem, Hey has recently utilised a sample of surnames covering England and Wales drawn from the Registrar-General's quarterly indexes of deaths for the period 1842–6.<sup>13</sup> The frequencies of surnames and the extent to which they change over time have previously been analysed to provide surrogate measures of the stability of local populations,<sup>14</sup> yet Hey's study concentrates on the geographic distribution of individuals with the same surname. Hey's study is not the first to use surname distributions to examine patterns of regional identity and behaviour,<sup>15</sup> yet it provides an important attempt to undertake a systematic historical survey of the entire country.<sup>16</sup> However, Hey's study is not without its limitations. One drawback is the source itself, which only allows analysis to be undertaken at the level of the Registration District. More problematic is the general approach taken. The analysis is essentially based on the examination of a series of maps for individual surnames, each showing a particular regional distribution or concentration. In his article a total of thirteen surname maps are presented and discussed. Although it is clear from this work that the study of historic surname distributions 'is a useful tool for determining the nature and boundaries of local societies'<sup>17</sup> and that for the surnames presented there is a distinct tendency toward geographical concentration, it remains unclear as to whether there were distinct zones of similar surnames that might be labelled regions. This article is an attempt to expand the work of Hey by examining the geographical distribution of surnames in order to determine regional divisions in the past.

## **Sources**

Obviously there are many historical sources that record the surnames of individuals. However in order to study the geographical distribution of a particular surname a source is required that is universal both in terms of its spatial and demographic coverage. In order to observe where people with a given surname are at a particular point in time, arguably a cross-sectional source is preferred over one compiled over a period of time. Thus an obvious candidate source with which to examine the geographical distribution of surnames is the census. Of all available historical sources the nineteenth-century census

enumerators' books (CEBs) are the nearest it is possible to get to a complete coverage of the population.<sup>18</sup> Fortunately, it is also the case that the returns for one historical census, that of 1881, are available for research in computerised form.<sup>19</sup> This article is therefore based entirely on an examination of the surnames extracted from the 1881 census.

The 1881 CEBs for England and Wales were originally transcribed and converted into computerised form by an army of volunteers from the Federation of Family History Societies and the Genealogical Society of Utah.<sup>20</sup> The resulting data were then subsequently subjected to further processing by the Department of History at the University of Essex.<sup>21</sup> This processing exercise not only standardised and classified much of the information captured by the census, it also, most importantly for this article, linked the administrative geography of the 1881 census to a parish-based Geographical Information System (GIS) so that the information recorded within the census can be mapped.<sup>22</sup> As anyone who has tried to trace individuals across successive censuses will know, the recording of surnames in the nineteenth-century CEBs can be inconsistent, for a variety of reasons, and in this respect the 1881 census is no different. The surnames contained within the 1881 database used for this article also suffer from the possibility of transcription or data entry mistakes. That said, there is nothing to indicate that the surnames recorded in the database do not represent a statistically representative cross-section of the surname pool of England and Wales at this moment in time.

Prior to discussing the analyses presented in this article it is important to note some basic characteristics of the surname pool. First, the size of the surname pool in England and Wales at this time was considerably larger than might have been expected or guessed at. In all, the database used in this research contains a total of 396,776 unique surname strings.<sup>23</sup> Obviously this number conceals a number of both transcription and typing mistakes, as is indicated by the fact that the database contains 158,876 surnames with a frequency of one. For this reason, the analyses presented in this article, unless otherwise stated, are based on only those surnames with a total frequency of 25 or more, of which there are 41,203. This equates to an average of one new surname per 630 persons across the whole population.<sup>24</sup> However, even concentrating on just those surnames with a frequency of 25 or over, the frequency distribution of these surnames was very far from even. Basically a large number of people shared a relatively small number of surnames, while, conversely, a large number of surnames were attributed to a small number of people. This exponential-type relationship is indeed common to most populations, both historic and modern.<sup>25</sup> The details

are given in Table 1, which illustrates that a fifth of the population shared just under 60 surnames, a half of the population were accounted for by some 600 surnames, while the top 10,000 surnames covered 90 per cent of the population. Conversely, ten per cent of the population, those with the rarest surnames, jointly accounted for some 30,000 surnames, more if those with frequencies of less than 25 are also considered.

One further characteristic of the data source needs consideration. Surnames, certainly by the nineteenth century, are inherited from one's father and women usually take their husband's surname upon marriage. Thus in the 1881 there are 5.4 million women, some 21 per cent of the population who no longer have the surname they were born with. What might the impact on the surname pool have been if women had retained their maiden name? Although it is impossible to measure this with accuracy, at a national level the impact was probably minimal since the chances are that as many men named Smith, for example, married women names Jones, and *vice versa*, thus cancelling one another out.

### **Linguistic regions?**

So, what can a database covering some 26 million people with some 41,000 different surnames reveal about regional diversity? In order to explore the geography of surnames in mid to late nineteenth-century England and Wales, it is appropriate to start by investigating the density of local surname pools across the country, or in other words the average number of persons per surname. This simple measure is shown for the parishes of England and Wales in Figure 1. Immediately, regional patterns begin to emerge with quite large variations in the density of surnames across the country. These are not just mirrors of urban and rural differences, indeed, the geography of surname density completely cuts across urban-rural divides. Not surprisingly Wales stands out as having a low surname density (or a high number of people per surname) but the same is also true of the south-west tip of Cornwall, the Kent and East Sussex Weald, a large grouping of parishes running across the east Midlands and East Anglia, taking in a large part of Cambridgeshire, west Essex, Bedfordshire, north Hertfordshire and part of Buckinghamshire. Further north other areas of low surname density are found in north Durham and an area consisting of south Lancashire and the south-western parts of the West Riding of Yorkshire. This last area stands in stark contrast to a fairly large area of high surname density running from south Durham through North Yorkshire to Nottinghamshire, Lincolnshire and the Wash. At first glance it looks as if the south Lancashire area of low surname density is separated by the area of high density in Yorkshire by the Pennines, but

on closer inspection the reality is not as straightforward as this with part of the low density area cutting across the hills in a belt approximately between Barnsley and Sheffield.

To a large extent the regional diversity shown in Figure 1 will reflect different customs relating to the origin of surnames. This is perhaps most clearly illustrated in the case of Wales (although note differences in south Wales, especially Monmouthshire) where surnames were adopted relatively later and very often consisted of a narrow set of elided patronyms with a genitival *-s*, such as Davies, Williams, Jones.<sup>26</sup> Equally, the low density of surnames in south Lancashire and the West Riding of Yorkshire is most probably associated with a propensity towards toponymic originating surnames.<sup>27</sup> However, two points of observation must be made. First, although some overlaps can be seen, the map of surname densities in 1881 does not follow physical geography. Second, surname densities when mapped at parish level do not correspond to the geography of administrative counties, a unit which is often used to aggregate and plot demographic, social and economic information.<sup>28</sup>

Plotting broad typologies or categories of surnames can also illustrate striking regional contrasts, as shown in Figures 2 and 3. The first of these plots the geographical distribution, by parish, of the percentage of the population enumerated in the 1881 census with patronymic and metronymic surnames (those ending in *-son*, for example, Johnson, Richardson, Moulson). Such surnames have been claimed to be of Scandinavian origin, although Reaney claims that they are more widely distributed and originated later than generally thought.<sup>29</sup> Whatever the origin, the geographical distribution is striking. Such surnames were heavily concentrated in the counties of Cumberland, Northumberland, Lancashire north of the Ribble, Durham, the North and East Ridings of Yorkshire and eastern Lincolnshire. Conversely, the areas of low surname density in south Lancashire and the western part of the West Riding of Yorkshire identified in relation to Figure 1, seem to correspond with relatively low proportions of patronymic and metronymic surnames. Overall, patronymic and metronymic surnames are relatively rare in proportional terms south of a V-shape formed by a line running from the mouths of the Mersey to the Thames and another running from the Wash to the Severn. It is particularly interesting to note that this nineteenth-century distribution of patronymic and metronymic surnames is very similar to that depicted by the Lay Subsidy Rolls some 500 to 600 years earlier.<sup>30</sup> The processes of industrialisation and migration, even over half a millennium, had not fundamentally changed where families with this particular form of surname were located.

Turning to a different form of patronym, those surnames ending with a genitival –s, Figure 3 again illustrates how this form of surname has a distinct geographical distribution. Such names are predictably very heavily concentrated in Wales, with the proportion of the population with such surnames declining quite sharply moving eastwards. Yet it is also true that very few incidences of such surnames are found north and east of a line running diagonally from the Mersey and the Thames. Thus, comparing Figures 2 and 3, it can be seen that there is minimal overlap between surnames of the –son and –s forms.

Moving to surnames of a different typology, those which are formed from occupational titles, it is instructive to examine the geographical distribution of three occupational surnames – Fuller, Tucker and Walker. Each of these surnames essentially refers to the same occupation, someone who works in the preparation of textiles scouring or beating the cloth as a means finishing or cleansing the fabric. Yet the terms are dialectic and traditionally thought to have distinct regional roots: Tucker mainly being used in the west country, Fuller in the east of the country and Walker in the Midlands and north. Turning to the 1881 census the first point that needs to be made in relation to these three names is that they are of unequal distribution in terms of absolute frequency. Walker was a very common surname nationally, being ranked 18<sup>th</sup> in England and Wales as a whole, with a total of 83,001 individuals with the surname. By contrast Tucker accounts for less than a quarter of this number, there being 16,430 Tuckers in 1881, with the surname being ranked 194<sup>th</sup> overall, while there were only 12,042 Fullers, making it ranked 303<sup>rd</sup>. Despite these differences in frequency, turning to the geographical distribution of the three surnames, although there is a degree of dislocation, Figure 4 largely confirms the general dialectic divide for these occupational titles outlined above. Walkers, despite the national importance of the surname predominate in the area north of the Wash-Severn line. Fullers occur in the south and east and Tuckers in the west. Those parishes in which only Fullers are found are located around the South Downs, western Suffolk and south west Norfolk, while those parishes where only Tuckers are found are situated in north and south Devon, Somerset, west Dorset and scattered across Cornwall. It is also interesting to note that with the exception of two areas, one south of the Wash and the other in the Weald, where Walkers and Fullers both occur, there is not much mixing of the surnames in what might be considered to be transition zones. Again, as with patronymic and metronymic surnames, there is evidence to suggest that this broad regional distribution of the three surnames is similar to that of the early fourteenth century.<sup>31</sup>

Taken together, these four examples, to which others could be added, would seem to indicate that cultural linguistic and dialectal variations might, predictably, have strong regional associations. However, what is more surprising is the fact that these variations had not broken down significantly by the late nineteenth century. Can this be taken as evidence of the strength and durability of regional cultures?

### **Economic regions?**

Taking a different methodological approach it is possible to examine the extent to which the surnames in a particular place overlap or correspond with those of another place. This is an approach, using surname distributions, that has previously been undertaken by human biologists interested in gene frequency distributions and assortative mating for genes of polymorphic systems.<sup>32</sup> In studying levels of in-breeding within populations human biologists have examined differences in surname distribution to measure the degree of biological kinship between communities by observing the frequency of shared surnames.<sup>33</sup> Taking a similar approach, it is possible to measure the proportional correspondence in terms of the shared surname pool between not only one place and another but between a particular place and all others in the country.<sup>34</sup> This is illustrated in Figures 5 and 6 which plot the degree of overlap between the surnames in each parish in England and Wales with those enumerated, respectively, in Lancaster and York. For both of these figures only those surnames ranked nationally in the top 10,000 are considered. In the case of Lancaster (Figure 5) there is a high degree of correspondence with the parishes close by. But correspondence is also high with parishes extending through the North and East Ridings of Yorkshire, Lincolnshire and also parts of north Wales. Conversely, the degree of correspondence is relatively weak with much of the southern part of the county and the West Riding of Yorkshire. It would appear that there are echoes here of the regional pattern displayed in patronymic and metronymic surnames. Turning to the white rose city of York (Figure 6), the parishes with the highest degree of surname correspondence are located in the North and East Ridings of the county, but relatively high levels of correspondence are also displayed by parishes in Cumberland, and to a lesser degree Lincolnshire, Nottinghamshire, Derbyshire and Leicestershire. Equally, most parishes in the west of the country (except Cumbria) record relatively low levels of surname correspondence with York. It may be cold comfort to Yorkshire loyalists to learn that the North and East Ridings of 'God's own county' have more in common, using this measure, with north Lancashire and Cumberland

than they do with the West Riding, which, in turn, seems bound at the hip with south Lancashire.<sup>35</sup>

The fact that the surname correspondence between places is not a straightforward function of distance is clearly illustrated in the case of Ely. Figure 7 shows how the surnames in the East Anglian parishes to the east of Ely share a high degree of correspondence with the city, yet this is not matched by parishes to the west. The links between East Anglia and Ely are vividly illustrated when rarer surnames, those with a national ranking over 10,000 are considered (Figure 8). It is also important to note that when considering the correspondence of surnames between places, the relationship is not necessary reciprocal. Unlike geographical distances in which the distance between X and Y is the same as between Y and X, the same is not true with surname pools. Thus if two places, X and Y, share twenty surnames in common, if X has a population of 1,000 and Y a population of 500, then it could be argued that the degree of correspondence between Y and X is twice that between X and Y. The fact that the geographical distribution of correspondence need not be reciprocal can again be illustrated by the case of Ely. While Figures 7 and 8 both measure the extent to which the pool of surnames in each parish of England and Wales overlap with the surnames of those living in Ely, Figure 9 turns the relationship around and measures, for those surnames nationally ranked over 10,000, the level of correspondence with the surnames found in Ely with other parishes in England and Wales. While East Anglian parishes are quite well represented, by comparison London stands out as having a high degree of correspondence, as does too, but to a lesser degree, the urban and manufacturing centres of the Black Country. Putting both sides of the Ely coin together it seems plausible to suggest that while individuals from the parishes of west Norfolk and Suffolk in particular were drawn to Ely, those from Ely were more likely to be attracted to the golden pavements of London and the black smoke of Birmingham.

Turning attention to London, surname correspondence reveals, in quite dramatic fashion, the importance of the capital, which by 1881 was not only the most populous urban settlement in England and Wales but also the world. Aggregating the various parishes of metropolitan London together, in 1881 there was not a single parish in the whole of England and Wales that was home to a surname ranked in the top 10,000 that was not also present in London. At this level every parish in the country reached out to London. Such a finding would seem to support the estimate by Wrigley that in the century 1650 to 1750 of all those born and surviving childhood, one in six of the country's population, maybe even more, lived at some stage of their life in London.<sup>36</sup> Even for rarer

surnames, those ranked over 10,000, Figure 10 shows that many of these surnames were also found in London. Yet here, a potential north-south divide begins to emerge, with the level of correspondence falling quite sharply northwards of a line running approximately from the Mersey to the Wash, but with Cornwall and parts of Wales, in this sense, being 'northern'.<sup>37</sup>

### **The importance of distance?**

The previous section examined the extent to which the pool of surnames in one place correspond or overlap with those of other places. In addition, it is possible, of course, to measure geographical distances between places: how far one place is from another. Putting these two pieces of information together, this section examines the degree to which individuals with the same surname were separated by distance.

Using the parish-based GIS for the 1881 census, 'as the crow flies' distances can be calculated from the centroid point of a parish to those of all other parishes. Using this measure it is then possible to calculate a mean distance separating all individuals with the same surname. This was done in the following way. For a given surname, starting with the place which records the highest frequency of the surname, the number of individuals with the surname is termed the 'total of separated persons'. Then the distance between this place and the place with the next highest frequency is multiplied by the number of individuals in the second place. This is called the 'total separation distance' and the number of individuals with the surname in the second place is added to the total of separated persons. Then the distance between the first place and the place with the third highest frequency is multiplied by the number of individuals in the third place. This is added to the total separation distance and the number of individuals with the surname in the third place is added to the total of separated persons. This is repeated for all places in which the surname occurs. Then the distance between the second place and the third place is multiplied by the number of individuals in the third place, and added to the total separation distance and the number of individuals with the surname in the third place is added to the total of separated persons. This sequence is repeated until all pairs of places have been considered. The mean separation distance then equals the total separation distance divided by the total of separated persons. Thus if a given surname occurs in four places, A to D, and A has 100 occurrences of the surname, B 50, C 20 and D 10, and the distances between the four places are A-B 15 km, A-C 25 km, A-D 30 km, B-C 10km, B-D 7 km, C-D 5 km, the mean separation distance would be calculated as follows:

$$\begin{aligned}
&= ((50 \times 15) + (20 \times 25) + (10 \times 30) + (20 \times 10) + (10 \times 7) + (10 \times 5)) / (100 + 50 + 20 + 10 + 20 + 10 + 10) \\
&= 1870 / 220 \\
&= 8.5 \text{ km.}
\end{aligned}$$

Taking those surnames with a national frequency of 25 or more, the mean separation distance across all surnames was 183.9 kilometres. It may have been expected that a relationship might exist between size of the surname group and mean separation distance for that surname, the two being positively correlated, but as Figure 11 demonstrates, this is not the case. Indeed, interestingly a standard distribution-type relationship exists in which the mean, mode and modal separation distances all occur close to one another, with most high frequency surnames falling around the mean separation distance point and with lower frequency surnames having mean separation distances both below and above the average.

Feeding these mean separation distances for each surname back into the places in which the surname occurs, a mean separation for each place was calculated. Thus if a given place had a population of 1,000, consisting of three surnames with frequencies of 500, 300 and 200, with mean separation distances respectively of 150, 100 and 50 kilometres, the mean separation distance for the place would be calculated as:  $((500 \times 150) + (300 \times 100) + (200 \times 50)) / 1,000 = 115$  kilometres.

Figure 12 plots these mean separation distances for parishes. This measure reveals a number of fairly distinct regional divisions. Amongst those regions with the lowest separation distances (with the darkest shading in the map) south Lancashire and the southern parts of the West Riding of Yorkshire again stand out as a common area. A second belt of low separation distances takes in Cheshire, north Staffordshire and north-east Derbyshire, while a third joins much of Sussex with south-west Kent. Four main belts of 'middling' separation distances can be identified: a group including Middlesex, Surrey, south Hertfordshire, south Buckinghamshire and east Berkshire; a large belt consisting of Shropshire, south Staffordshire, south Derbyshire, stretching over to east Nottinghamshire and Leicestershire; Lincolnshire being joined by Cambridgeshire; lastly Essex, being separated from its East Anglian neighbours of Suffolk and Norfolk. Three main areas of high separation distances include Cornwall, Devon, Somerset and Dorset, north Wales and Northumberland coupled with west Cumberland. In some respects these areas of high separation distance are a result of being furthest from London, which as been mentioned, was the home to surnames drawn from all over the country, but clearly

urbanisation and distance from London are not the only factors in influencing the geographical pattern shown in Figure 12.

### **A patchwork quilt?**

The final approach taken is to attempt to group together parishes with similar pools of surnames. This was done by applying a statistical technique termed cluster analysis.<sup>38</sup> The results are shown in Figures 13 and 14. The first of these maps the outcome of a cluster analysis on Registration Districts considering just the top 1,000 surnames ranked nationally. Again, one of the features to arise from this exercise is a general diagonal divide running from the mouth of the Mersey to the Thames, evident in other surname distributions presented previously, and perhaps not that dissimilar from the line demarcating the southern limit of Danelaw in the ninth century. An enlarged East Anglia stands out as a distinct cluster, including Norfolk, Suffolk, south Lincolnshire (Holland), much of Cambridgeshire, Huntingdonshire, Bedfordshire, and north Hertfordshire and Essex, while a third stretches from Glamorganshire, westward across Monmouthshire into parts of Gloucestershire, Wiltshire, Somerset and Devon, and over to Oxfordshire and Berkshire. This exercise does not, however, form south Lancashire and the southern West Riding of Yorkshire into a discrete cluster, these being grouped with parts of London, Durham and the south coast in a residual cluster.

Figure 13 maps the resulting clusters considering instead the surnames ranked from 2,000 to 7,500. East Anglia stands out again, yet this time does not extend as far north and west, and instead extends down in Kent. Two main midland clusters stand out. The east midlands group comprises south Lincolnshire, Nottinghamshire, Leicestershire, Rutland and parts of Northamptonshire, Huntingdonshire and Derbyshire, while the west midlands group includes Cheshire and parts of several counties to the south - Shropshire, Staffordshire and Worcestershire. A northern cluster brings together the counties of Cumbria, and parts of Northumberland, the North Riding of Yorkshire and Durham. Lastly, Cornwall and Devon joint to form a distinct South West cluster.

### **Conclusions**

This article set out as a tentative and speculative exploration to identify historical cultural regions via an examination of the surnames of individuals as records in the 1881 CEBs. It would be naïve to claim that a clear set of regional boundaries have been determined, however, a number of regional groupings have identified.

It is also clear that the search for distinct regional divisions may prove illusive, if only because regions will be nested and overlapping depending on the point of focus adopted.<sup>39</sup> Yet, equally, this is just the start of the journey. Further analysis is required to both examine the stability of surname geographies over time<sup>40</sup> and compare the evidence offered by the surname distributions to other factors such as physical divides, geographies of religion,<sup>41</sup> demographic variation,<sup>42</sup> agricultural regions<sup>43</sup> and political territories.<sup>44</sup> Equally, more robust and appropriate statistical procedures can be applied to the data.<sup>45</sup> One can only hope that the search to identify historic regions does not turn into a quest for the Holy Grail.

## Endnotes

1. S. Hindle, 'A sense of place? Becoming and belonging in the rural parish, 1550-1650', in A. Shepard and P. Withington, eds *Communities in early modern England*, (Manchester, 2000), 96-114. See also the discussion in K. D. M. Snell, 'English rural societies and geographical marital endogamy, 1700-1837' *Economic History Review*, (2002), 55, 262-98 (especially pp.268-70).
2. However, various researchers have noted the importance of local networks in urban settings. In his study of seventeenth-century Southwark, for example, Boulton noted how many 'Boroughside householders may have possessed geographically restricted social horizons, living out much of their lives within a local social system'. J. Boulton, *Neighbourhood and society: a London suburb in the seventeenth century*, (Cambridge, 1987), 291. For the nineteenth century see R. Dennis and S. Daniels, "'Community" and the social geography of Victorian cities', in M. Drake, eds, *Time, family and community. Perspectives on family and community*, (Oxford, 1994), 201-24. Reprinted from *Urban History Yearbook*, (1981), 7-23.
3. J. D. Marshall, 'Communities, societies, regions and local history. Perceptions of locality in High and Low Furness', *The Local Historian*, (1996), 26, 36-47 (cited here at p.39). The case for studying regions is set out in his two articles 'Why study regions? (1)', *Journal of Regional and Local Studies*, (1985), 5, 15-27 and 'Why study regions? (2): some historical considerations', *Journal of Regional and Local Studies*, (1986), 6, 1-12. See also the compelling arguments in his *The tyranny of the discrete. A discussion of the problems of local history in England*, (Aldershot, 1997).
4. R. I. Woods, 'Spatial demography', in J. I. Clarke, ed., *Geography and population: approaches and applications*, (Oxford, 1984), 43-50 (see especially p. 43 where the term is defined). For an example of a study in historical demography that very much emphasises the importance of place see E. Garrett, A. Reid, K. Schürer and S. Szreter, *Changing family size in England and Wales. Place, class and demography, 1891-1911*, (Cambridge, 2001). Interestingly, King has also argued that historians of the old and new poor laws should concentrate 'their energies on identifying the boundaries of .....regional welfare systems'. S. King, *Poverty and welfare in England, 1700-1850. A regional perspective*, (Manchester, 2000), 251.

5. C. Phythian-Adams, 'Local history and societal history', *LPS*, (1993), **51**, 30–45 (cited here at p.31); see also C. Phythian-Adams, *Rethinking local history*, (Leicester, 1987). For further discussions of the importance of place see also K. Schürer, 'Regional identity and populations in the past', in D. Postles, *Naming, society and regional identity*, (Oxford, 2002), 201-27.
6. Phythian-Adams, 'Local history and societal history', 31.
7. See, in particular, C. Phythian-Adams, 'Introduction: an agenda for English local history', in C. Phythian-Adams, ed. *Societies, culture and kinship, 1560–1850. Cultural provinces in English local history*, (Leicester, 1993), 1-23.
8. D. Hey, 'The local history of family names', *The Local Historian*, (1997), **27**, i–xx.
9. Hey, 'The local history of family names', ii.
10. Hey, 'The local history of family names', ii.
11. Phythian-Adams, 'Introduction'. The quote and those of the previous sentence are cited at p.10. See also the maps (Figure I.1 to I.4) produced on the four un-numbered pages preceding p.1.
12. For example, J. D. Marshall, 'Communities, societies, regions and local history'. See also the approaches taken in M. Carter, 'Town or urban society? St Ives in Huntingdonshire, 1630-1740' and E. Lord, 'Communities of common interest: the social landscape of south-east Surrey, 1750-1850', both in Phythian-Adams, *Societies, cultures and kinship*, respectively pages 77-130 and 131-99.
13. Hey, 'The local history of family names'. See also the use of marriage registers in K. D. M. Snell, 'English rural societies and geographical marital endogamy, 1700-1837' *Economic History Review*, (2002), **55**, 262-98 and Carter, 'Town or urban society?' and for the use of carter routes from trade directories see A. Everitt, 'Town and country in Victorian Leicestershire: the role of the village carrier', in A. Everitt, ed., *Perspectives in English urban history*, (London, 1973), 213-40. However, using either parish registers or trade directories at a national level would pose huge data preparation problems.

14. Examples include: E. F. Buckatzsch, 'The constancy of local populations and migration in England before 1800', *Population Studies*, (1951), **5**, 62-9; N. Evans, 'The descent of dissenters in the Chiltern Hundreds', in M. Spufford, ed. *The world of rural dissenters, 1520-1725*, (Cambridge, 1995), 288-308; W. G. Hoskins, *The Midland peasant*, (London, 1957), 195-6; E. Lord, 'The distribution and stability of surnames in south-east Surrey – 1664-1851', in D. Hooke and D. Postles, eds *Names, time and place. Essays in memory of Richard McKinley*, (Oxford, 2003), 181-92; S. A. Peyton, 'Village populations in Tudor Lay Subsidy Rolls', *English Historical Review*, (1915), **30**, 234-50; D. Postles, *The surnames of Leicestershire and Rutland*, (English Surname Series VII, Oxford, 1998), 333-48; S. and D. Postles, 'Surnames and stability: a detailed case study', in Hooke and Postles, *Names, time and place*, 193-207; P. Spufford, 'The comparative mobility and immobility of Lollard descendants in early modern England', in Spufford, *The world of rural dissenters*, 309-31; R. Watson, 'A study of surname distribution in a group of Cambridgeshire parishes 1538-1840', *LPS*, (1975), **15**, 23-32.
15. For the use of surnames to study regional migration patterns, see P. McClure, 'Patterns of migration in the late middle ages: the evidence of English place-name surnames', *Economic History Review* (2nd series), (1979), **32**, 167-82 and R. A. McKinley, 'The distribution of surnames derived from the names of some Yorkshire towns', in F. G. Emmison and R. Stephens, eds *Tribute to an antiquary. Essays presented to Marc Fitch by some of his friends*, (London, 1976), 165-75. For an early example of the analysis of surname distribution see H. B. Guppy, *The homes of family names in Great Britain*, (London, 1890).
16. For an analysis of modern surname distribution see R. R. Sokal, R. M. Harding, G. W. Lasker, C. G. N. Mascie-Taylor, 'A spatial analysis of 100 surnames in England and Wales', *Annals of Human Biology*, (1992), **19**, 445-76. See also G. W. Lasker and C. G. N. Mascie-Taylor, *Atlas of British surnames*, (Detroit, 1990) and the 'Appendix: map and diagrams of 100 surnames in England and Wales' to G. W. Lasker, *Surnames and genetic structure*, (Cambridge, 1985), 90-141.
17. Hey, 'The local history of family names', xix.
18. For background information on the nineteenth-century censuses see E. Higgs, *A clearer sense of the census. The Victorian censuses and historical research*, (London, 1996); E. Higgs, *Making sense of the census: the manuscript returns of the census 1801-1901*, (London, 1989); D. Mills and K. Schürer, eds *Local communities in the Victorian census enumerators' books*, (Oxford, 1996).

19. K. Schürer and M. Woollard, *1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enriched Version)* [computer file]. Genealogical Society of Utah, Federation of Family History Societies [original data producers]. Colchester, Essex: UK Data Archive [distributor], 2000. SN:4177. The 1901 census for England and Wales is also available online at <http://www.census.pro.gov.uk> however this is of limited value for historical research in its present form. Data are also available for Scotland (see K. Schürer and M. Woollard, *1881 Census for Scotland (Enriched Version)* [computer file]. Genealogical Society of Utah, Federation of Family History Societies [original data producers]. Colchester, Essex: UK Data Archive [distributor], 2000. SN:4178) but for pragmatic reasons the analyses for this article are confined to England and Wales only.

20. For details of this project see M. Woollard, 'Creating a Machine-Readable Version of the 1881 Census of England and Wales', in C. Harvey and J. Press, eds *Databases in Historical Research*, (London, 1996), 98–101.

21. Details of the processing work are provided in K. Schürer and M. Woollard, *The national sample of the 1881 census of Great Britain. A user guide and workbook*, (Colchester, 2001).

22. The 1881 GIS was created in collaboration with the Great British Historical GIS project formally based at the Department of Geography, Queen Mary College, London and now located at Portsmouth University.

23. A minimal amount of standardisation was undertaken on the surname strings (for example all punctuation and spaces were removed, so 'O' Conner' became Oconner', Mc and Mac's were all changed to Mc, and for multi-barrelled surnames only the last surname was taken), however, no attempt was made to group surnames, thus 'Smith', 'Smyth' and 'Smythe' all remain as distinct surnames.

24. However, this figure seems rather high in comparison to evidence from the civil registers. Analysing the registers of births for England and Wales for the first quarter of 1851, together with the registers of deaths for the first quarter of 1853, the Registrar General found there to be 32,818 different surnames out of a total of 275,405 persons registered, giving an overall rate of 8.4 persons per surname (see Registrar General, 'Family nomenclature in England and Wales', *Sixteenth Annual Report* (London, 1856), xvii–xxviii, cited at p.xvii). Even using the total number of surname strings in the 1881 census (396,776) only gives a rate of 65.5 persons per surname.

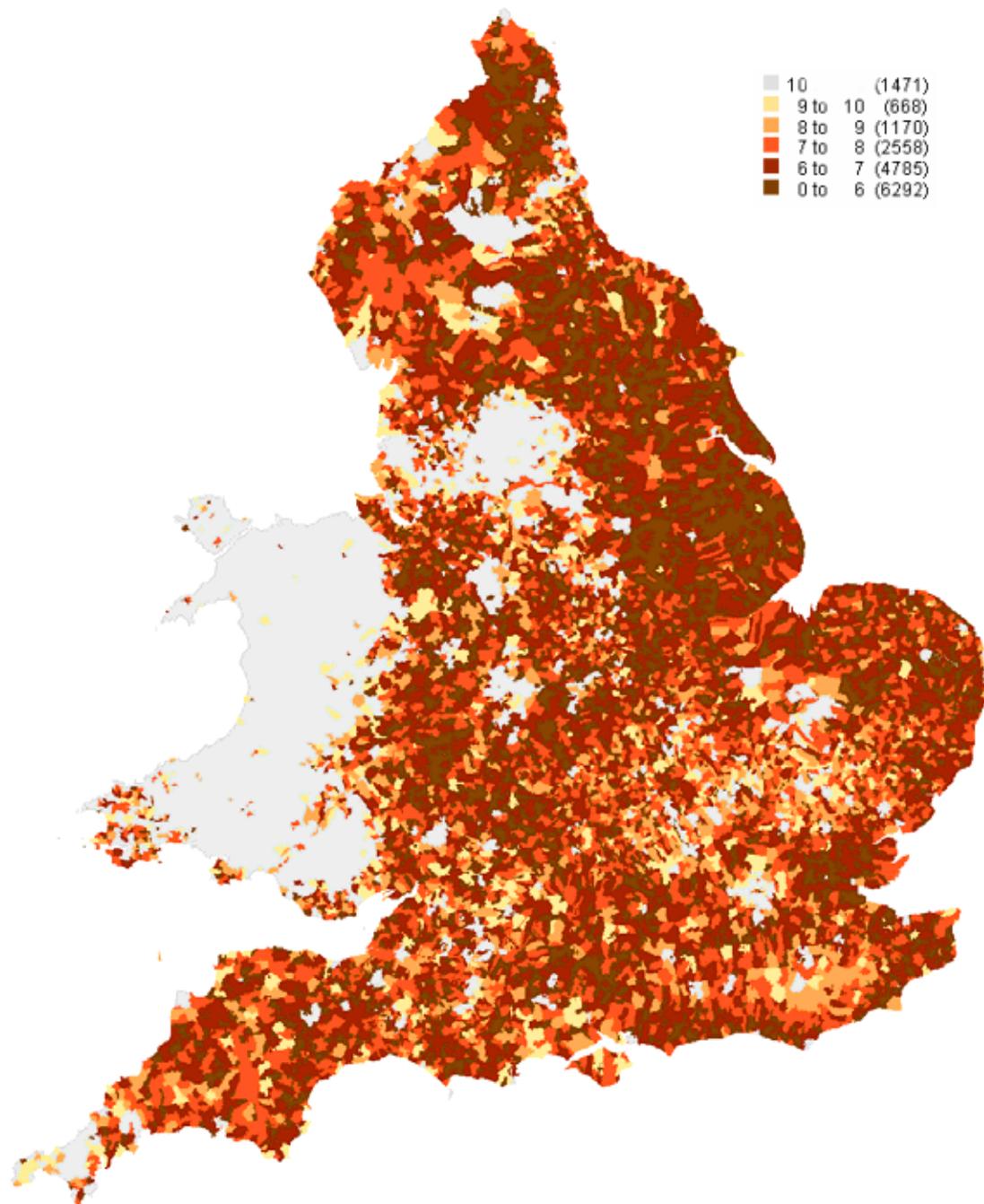
25. J. Randerson, 'No shame in a name', *New Scientist* (2002), **173**, 19.
26. J. and S. Rowlands, *The surnames of Wales*, (Birmingham, 1996). It is important to note that Welsh surnames also have a particular form of patronymic surnames, the ap or ab, from the Welsh word for son, mab. This gave rise to names such as David ap Hugh which became David Pugh, or Thomas ab Evan which became Thomas Bevan.
27. Toponymic is used here to indicate surnames specifically originating from a placename as distinct from topographical surnames originating from features in the landscape (Hill, Wood, etc). For the relative importance of toponymic surname in Lancashire see R. A. McKinley, *The surnames of Lancashire*, (English Surname Series IV, London, 1981) who shows that such surnames were particularly numerous in Salford Hundred in the south east of the county (pp.5, 78-110, 442-53). It is also the case that Lancashire has the highest proportion of 'local' surnames (toponymic and topographical) of the counties for which the late thirteenth and early fourteenth-century Lay Subsidy Rolls were analysed by Reaney. See P. H. Reaney, *The origin of English surnames*, (London, 1967), 22. For the toponymic surnames in the West Riding of Yorkshire see G. Redmonds, *Yorkshire West Riding*, (English Surname Series I, Chichester, 1973) 59-60, 63-5.
28. Marshall appropriately referred to the county as 'a fine repository of convenient statistics'. J. D. Marshall, 'Discussion article: regions, regionalism and regional scholarship', *Journal of Regional and Local Studies*, (1990), **10**, 68.
29. Reaney, *Origin*, 86. The high density areas of patronymic and metronymic surnames shown in Figure 2 show a high degree of correspondance with the areas of pre-880 Danish settlement shown on the map (based on A. H. Smith, *English place-name elements*, 1-2, (English Place-Name Society XXV-XXVI, Cambridge, 1956), Map 10) given in G. Fellows-Jenson, 'Variations in naming practice in areas of Viking settlement in the British Isles', in Postles, *Naming, society and regional identity*, 124-42 (p.124). See also the discussion in Postles, *Surnames of Leicestershire and Rutland*, 18-9.
30. See the map, based on counties, presented in Postles, *Surnames of Leicestershire and Rutland*, 17.
31. See Table 5.11 in D. Postles, *The surnames of Devon*, (English Surname Series VI, Oxford, 1995), 140.

32. For overviews see G. W. Lasker, 'Using surnames to analyse population structure', in Postles, *Naming, society and regional identity*, 5-23 and Lasker, *Surnames and genetic structure*, 6-11.
33. Examples are provided in G. W. Lasker, 'A coefficient of relationship by isonymy: a method for estimating the genetic relationship between populations', *Human Biology*, (1977), **49**, 489-93 and G. W. Lasker, 'Increments through migration to the coefficient of relationship between communities estimated by isonymy', *Human Biology*, (1978), **50**, 235-40. The measure has also been applied to historical data, although these studies have tended to be rather limited in geographical coverage. Parish registers have been used to study the changing isonymy over time between a group of Oxfordshire parishes (see C. F. Küchemann, G. W. Lasker and D. I. Smith, 'Historical changes in the coefficient or relationship by isonymy among the populations of the Otmoor villages', *Human Biology*, (1979), **51**, 63-77; G. W. Lasker, 'Evidence from surnames on the population structure of villages surrounding the Otmoor', in G. A. Harrison, ed. *The human biology of the English village*, (Oxford, 1995), 137-43) and the correspondence between a group of 37 parishes in east Kent has been studied using Marriage Duty Act returns for 1705 (see D. Souden and G. W. Lasker, 'Biological inter-relationships between parishes in east Kent: an analysis of Marriage Duty Act returns for 1705', *LPS*, (1978), **21**, 303-11). Of particular note is a project on Fylingdales, North Yorkshire, which uses nineteenth-century census surname and occupation data to explore differences between social groups using multidimensional scaling techniques (see M. T. Smith and B. L. Hudson, 'Isonymic relationships in the parish of Fylingdales, North Yorkshire in 1851', *Annals of Human Biology*, (1984), **11**, 141-8; M. T. Smith, B. L. Smith and W. R. Williams, 'Changing isonymic relationships in Fylingdales parish, North Yorkshire, 1841-1881', *Annals of Human Biology*, (1984), **11**, 449-57; M. Smith, 'The inference of genetic structure and the micro-evolutionary process from the distribution and changing pattern of surnames', in Postles, *Naming, society and regional identity*, 27-45).
34. A similar approach, yet on a local scale, is taken in examining the relative changing hinterlands of St Ives and Huntingdon between the mid-seventeenth and early-eighteenth centuries in Carter, 'Town or urban society?', 123-5.
35. It is interesting to note the comments in D. Neave, 'The identity of the East Riding of Yorkshire', in E. Royle, ed. *Issues of regional identity. In honour of John Marshall*, (Manchester, 1998), 184-200.

36. E. A. Wrigley, 'A simple model of London's importance in a changing English society and economy, 1650-1750' chapter 6 of *People, cities and wealth. The transformation of traditional society*, (Oxford, 1987), 133-56 (cited at p.137). Reprinted from *Past and Present*, (1967), 37, 44-70.
37. H. M. Jewell, *The north-south divide: the origins of Northern consciousness in England*, (Manchester, 1994).
38. Cluster analysis is a statistical operation that identifies hierarchical clusters (or similar groups) of observations within a dataset. The routine used in this exercise was the Statistical Analysis Software (SAS) FASTCLUS Procedure. This was chosen as it is designed especially for very large datasets. The procedure performs a disjoint cluster analysis on the basis of Euclidean distances computed, in this case, from an array of binary variables  $s_1$ - $s_{7500}$  for each Registration District set to 1 if the surname is present in that district and 0 if it is not, where  $s_1$  equals the surname ranked first in terms of frequency nationally (Smith),  $s_2$  the second ranked surname (Jones) and so on.
39. This point is made suggested in part in King, *Poverty and welfare*, 4-5.
40. See, for example, D. Hey, 'Stable families in Tudor and Stuart England', in Hooke and Postles, *Names, time and place*, 165-80 and D. Hey, 'Locating the home of a family name', appendix to P. H. Reaney and R. M. Wilson, *A dictionary of English surnames*, (Oxford, 1997, revised third edition), 511-20.
41. See K. D. M. Snell and P. S. Ell, *Rival Jerusalems: the geography of Victorian religion*, (Cambridge, 2000).
42. See R. Woods and N. Shelton, *An atlas of Victorian mortality*, (Liverpool, 1997)
43. See J. Thirsk, *England's agricultural regions and agrarian history, 1500-1750*, (Basingstoke, 1987) and A. Kussmaul, *A general view of the rural economy of England, 1538-1840*, (Cambridge, 1990).
44. See H. Pelling, *Social geography of British elections, 1835-1910*, (London, 1967).

45. Such as improved cluster analyses and applying the wombling technique.  
See R. R. Sokal *et al*, 'A spatial analysis of 100 surnames'.

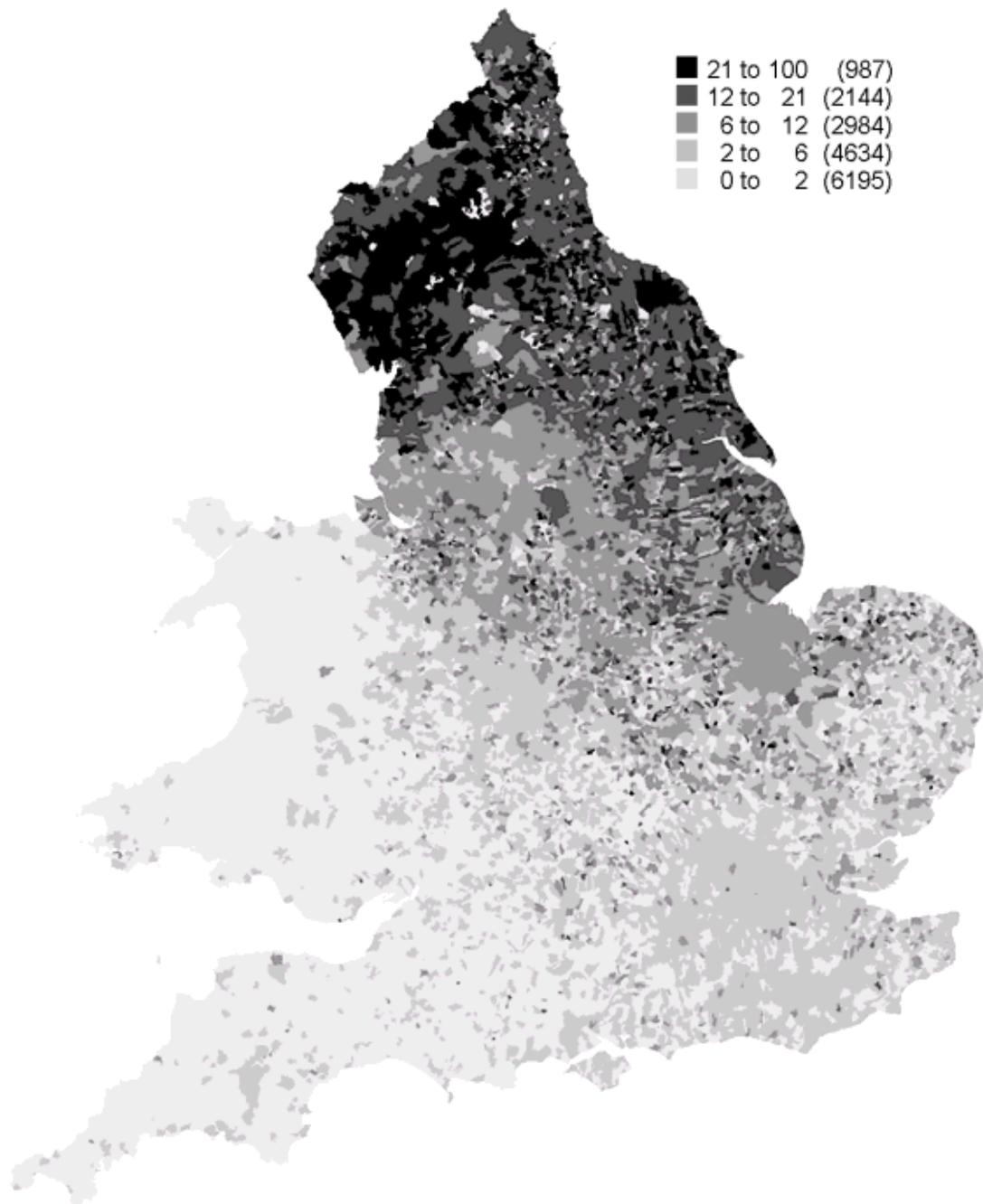
**Figure 1** Surname density, by parish, 1881.



**Note:** The shading on the map denotes the number of persons per surname, with the lighter areas having the lowest surname density and the darker areas the highest. The grey shading equates to 10 or more persons per surname.

**Source:** 1881 CEBs

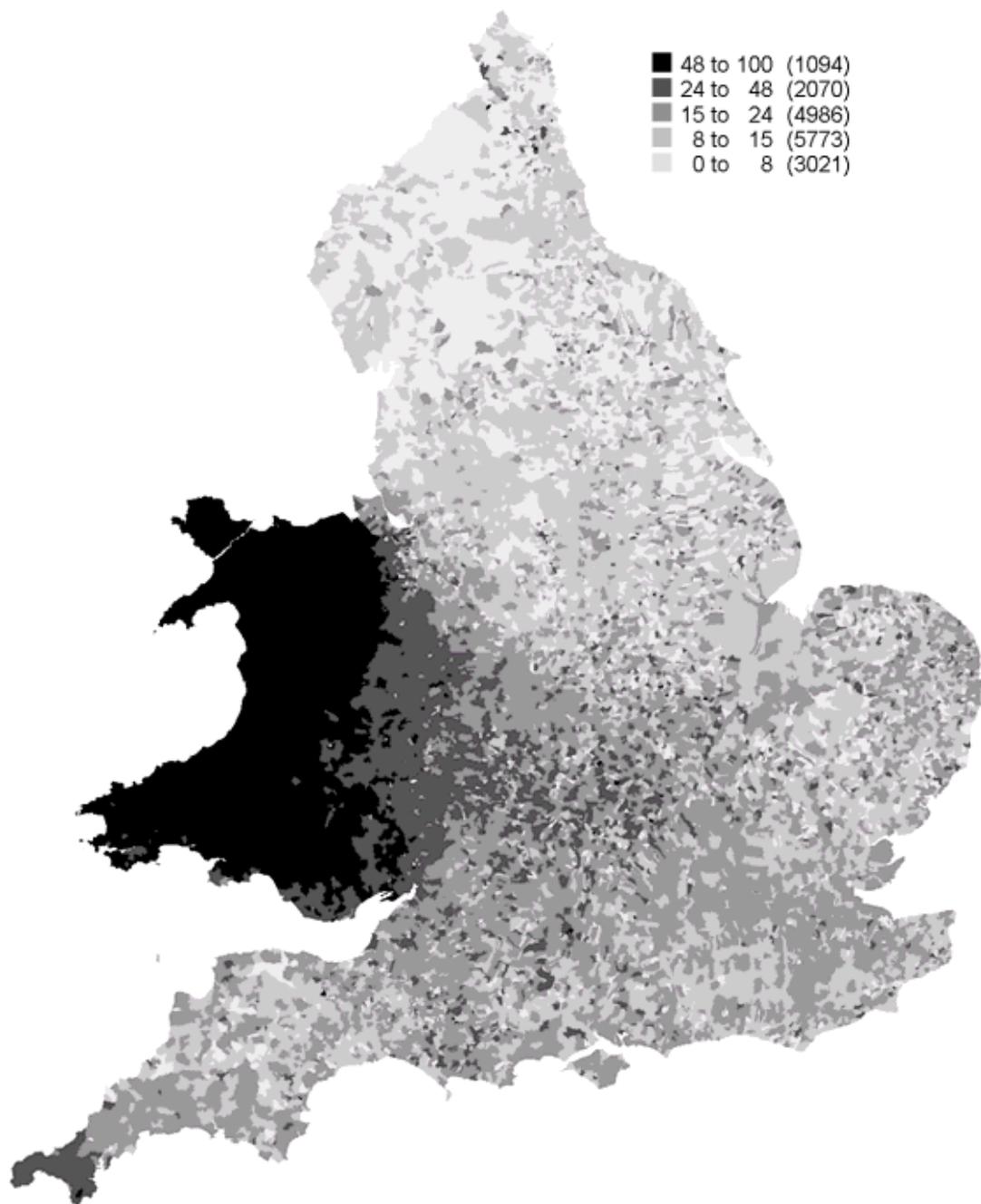
**Figure 2** Distribution of patronymic and metronymic surnames by parish, 1881.



**Notes:** The key denotes the percentage of the population in each parish with patronymic and metronymic surnames ending in -son. The number in brackets indicates the number of parishes within the given category.

**Source:** 1881 CEBs

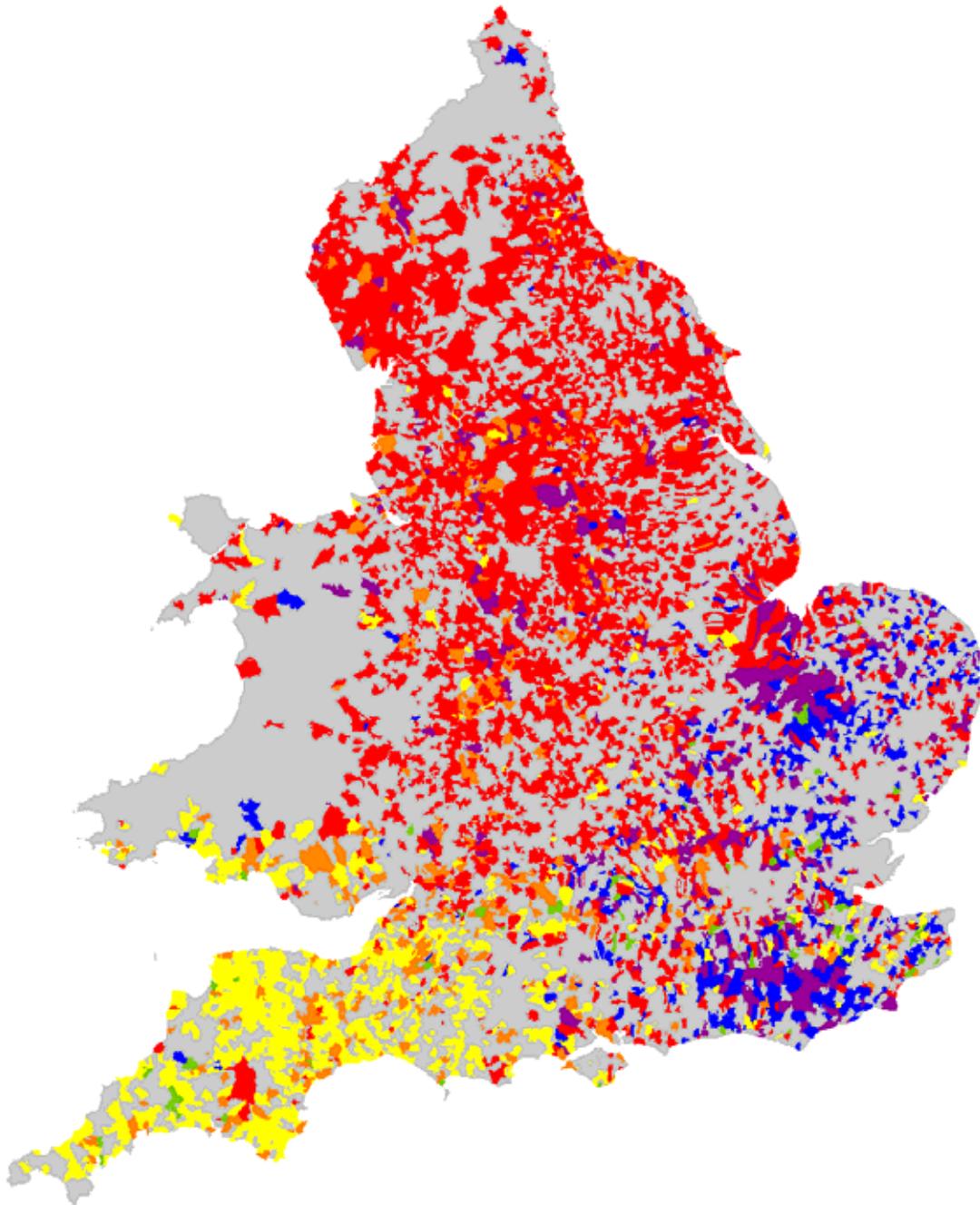
**Figure 3** Distribution of patronymic and metronymic surnames ending with a genitival –s, by parish, 1881.



**Notes:** The key denotes the percentage of the population in each parish with patronymic and metronymic surnames ending with a genitival –s. The number in brackets indicates the number of parishes within the given category.

**Source:** 1881 CEBs

**Figure 4** Distribution of Tuckers, Fullers and Walkers, by parish, 1881.



**Note:** The shading on the map is colour coded as follows: Blue = parishes in which only Fullers occur; Red = parishes in which only Walkers occur; Yellow = parishes in which only Tuckers occur; Purple = parishes in which Walkers and Fullers occur; Orange = parishes in which Walkers and Tuckers occur; Green = parishes in which Tuckers and Fullers occur; Grey = parishes in which either Walkers, Tuckers and Fullers all occur or non occur.

**Source:** 1881 CEBs

**Figure 5** The proportion of surnames ranked <10,000 matching those present in Lancaster, 1881.



**Note:** For this map, and those in Figures 6 to 9, the darker the shading the greater the degree of correspondence in the surname pool, and *vice versa*.

**Source:** 1881 CEBs.

**Figure 6** The proportion of surnames ranked <10,000 matching those present in York, 1881.



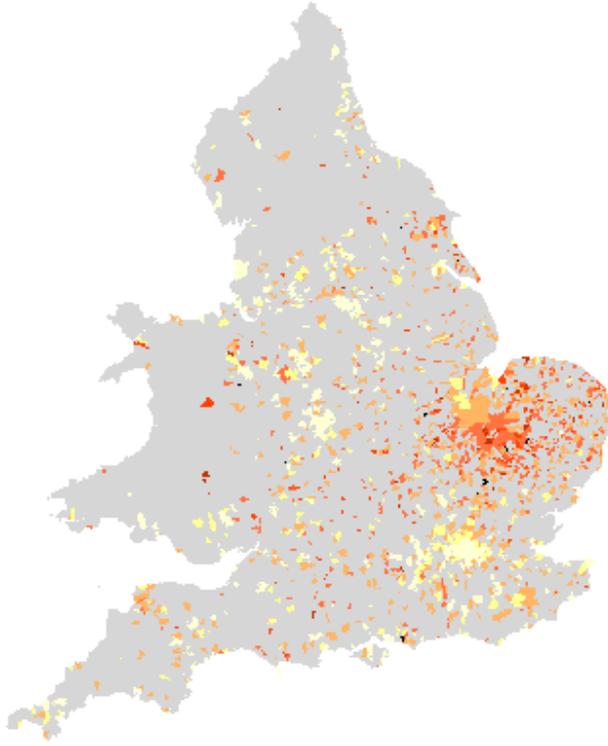
**Source:** 1881 CEBs.

**Figure 7** The proportion of surnames ranked <10,000 matching those present in Ely, 1881.



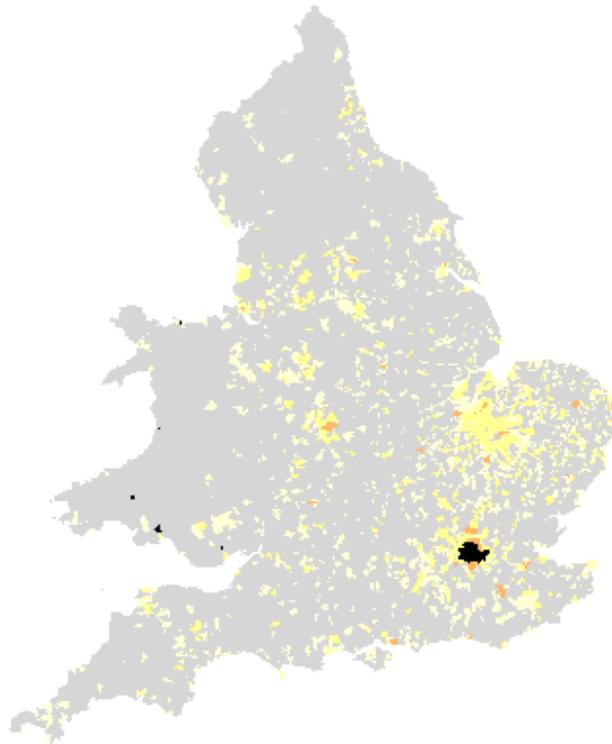
**Source:** 1881 CEBs.

**Figure 8** The proportion of surnames ranked >10,000 matching those present in Ely, 1881.



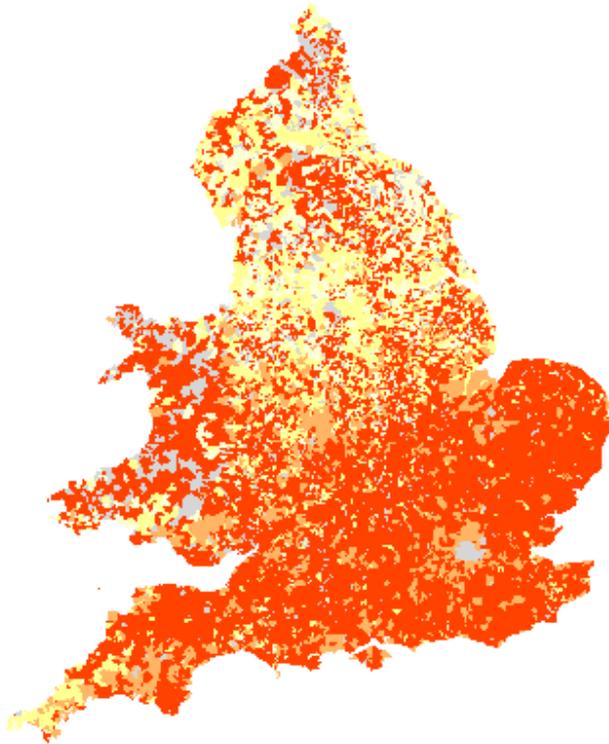
**Source:** 1881 CEBs.

**Figure 9** The proportion of surnames ranked >10,000 found in Ely matching those in other parishes, 1881.



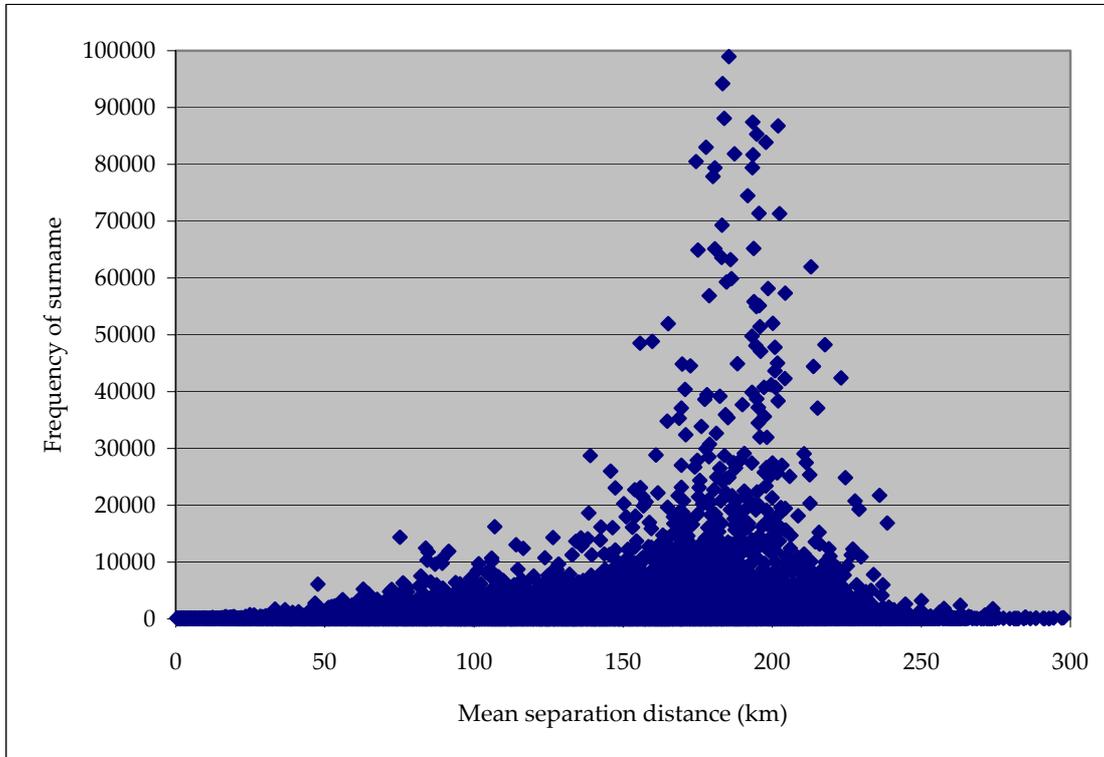
**Source:** 1881 CEBs.

**Figure 10** The proportion of surnames ranked >10,000 matching those present in London, 1881.



**Source:** 1881 CEBs.

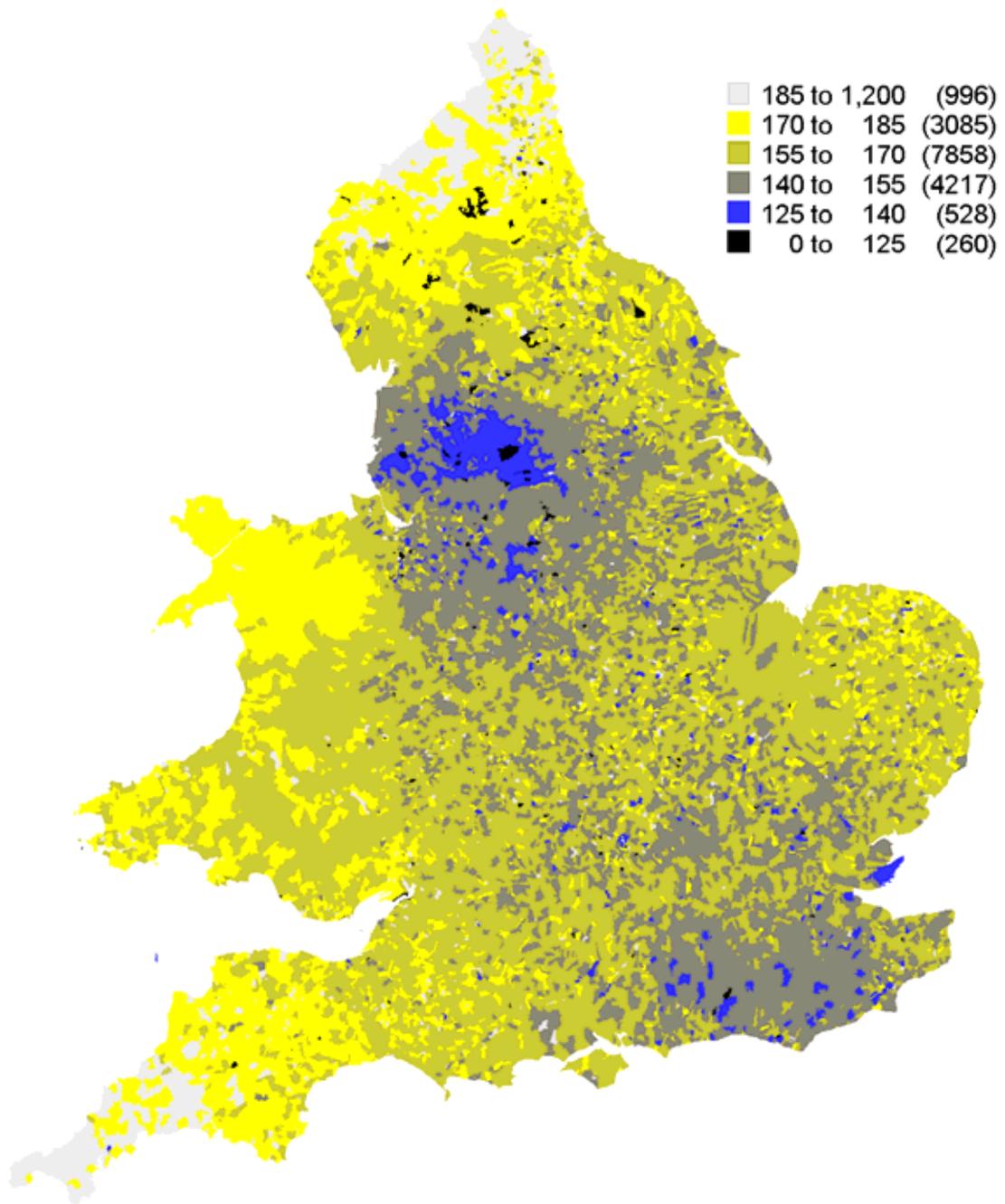
**Figure 11** Mean separation distance by frequency of surname, 1881.



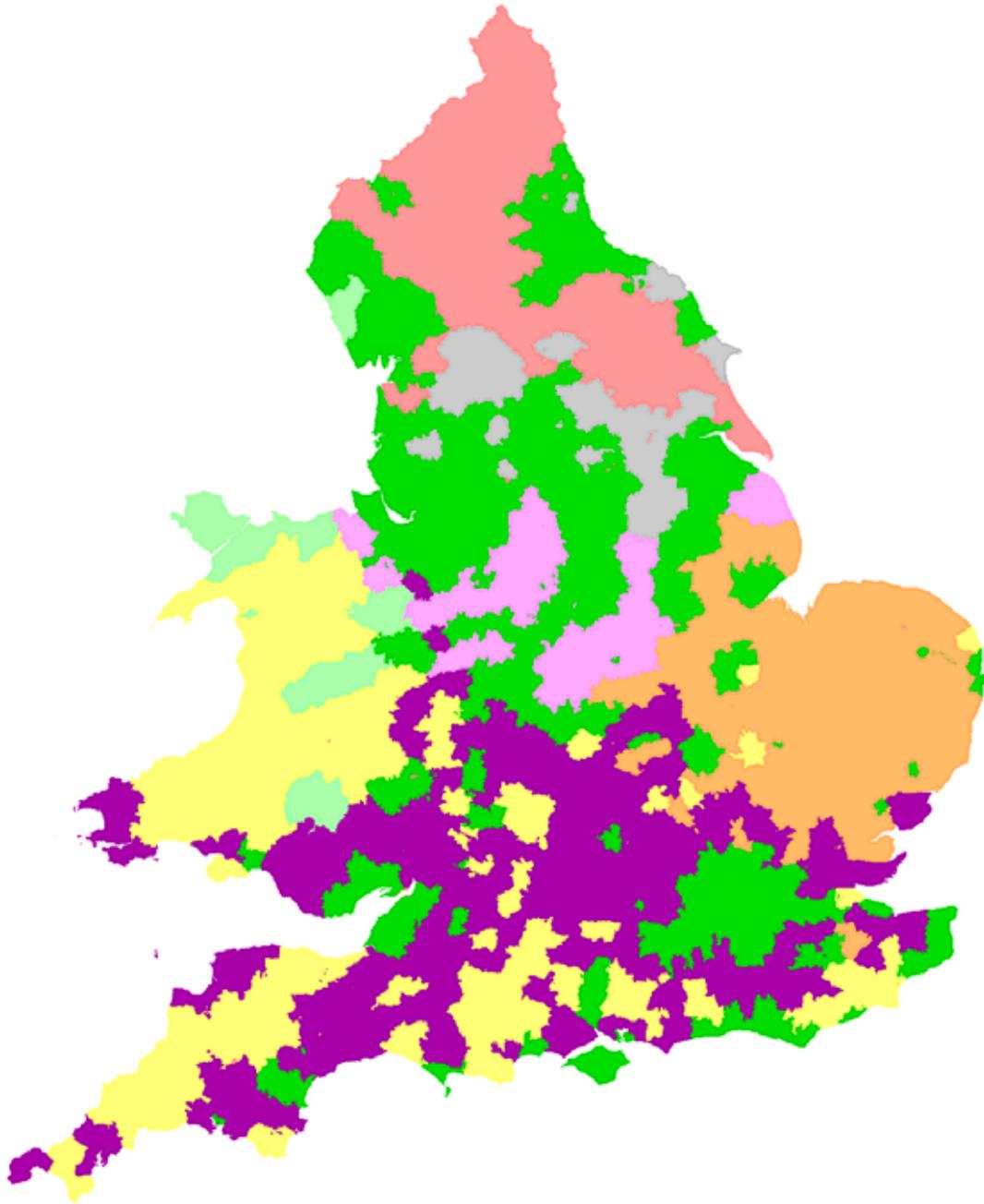
**Note:** The ten surnames with a total national frequency of over 100,000 are omitted for clarity. The mean separation distance for these ten surnames ranges between 150 and 206 kilometres.

**Source:** 1881 CEBs.

**Figure 12** Mean separation distance of surnames, by parish, 1881.

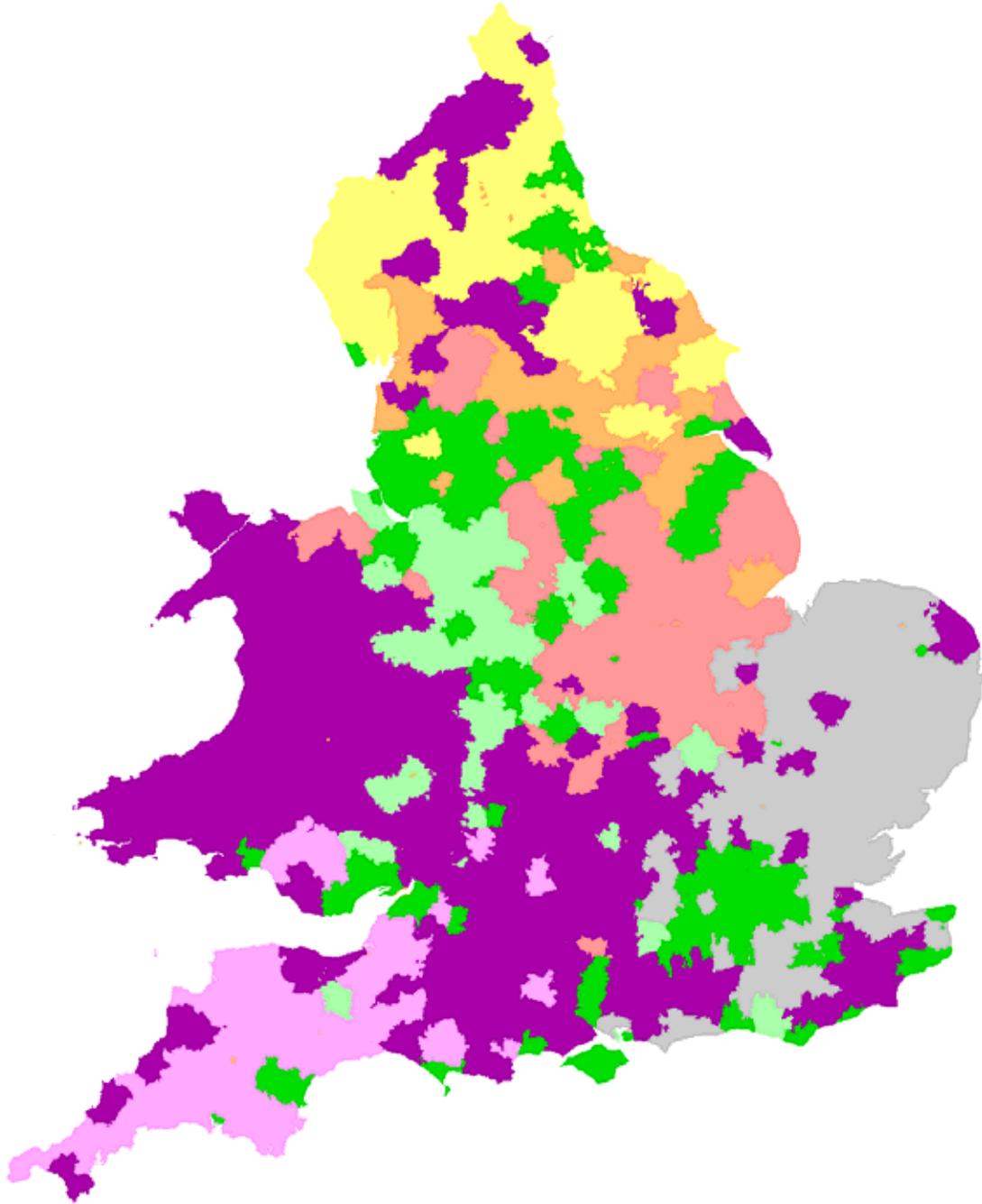


**Figure 13** Results of cluster analysis on surnames ranked <1000, 1881.



**Source:** 1881 CEBs.

**Figure 14** Results of cluster analysis on surnames ranked between 2,000 and 7,500, 1881.



**Source:** 1881 CEBs.

**Table 1** Distribution of the population by number of surnames, 1881.

Accumulated percentage of the population	Number of surnames accounting for percentage population
10	17
20	58
30	144
40	306
50	601
60	1,141
70	2,173
80	4,359
90	10,009
95	17,385

**Source:** 1881 CEBs.

