INVESTIGATING POPULATION MOBILITY IN MID NINETEENTH CENTURY ENGLAND AND WALES

N. Spencer and D.A. Gatley

Neil Spencer is Senior Lecturer in the Department of Statistics, Accounting and Management Systems at the University of Hertfordshire. David Gatley is in the School of Humanities and Social Sciences at Staffordshire University

Introduction

In this paper we describe a technique for analysing the relationship between migration, population, age structure and occupational structure. The analysis is based on materials compiled by David Gatley, whose work on computerising the 1861 census was described in an earlier issue of this journal.[1]

The analysis is based on census data compiled and computerised at the level of the registration district.[2] As a unit of analysis with which to study the population geography of England in the nineteenth century, registration districts are not perfect. Many contained several diverse and distinct communities, involved in different occupations and having different patterns of migratory behaviour. Amalgamating the data from these sub-groups into a single registration district necessarily involves some loss of information. However, with 635 districts covering England and Wales, the level of detail available is still appreciable. Registration districts also have two advantages over other units of analysis. First, they give us national coverage, enabling us to examine migration on a national basis while retaining a degree of detail not available with a division of the countries into areas such as counties. Second, the degree of information available at this level in the 1861 Census is particularly detailed.

To undertake our work we extracted several strands of data from the census. The first data strand involved records of the proportion of males in each registration district employed in six occupational sectors: manufacturing, agriculture, service occupations, mining, transport and the armed forces. The grouping of six occupational sectors was carried out in accordance with the classification schemes of Booth and Armstrong, but we have deviated slightly from them in a few instances.[3] First, we have included clerks and dealers along with public service workers in the service sector. Second, dockyard artificers have been reclassified under manufacturing. Third, we have excluded both general labourers and construction workers from our analysis, because general labourers tended to move from industry to industry, and many construction workers were also migratory workers. Their inclusion, therefore, is likely to have biased the results. Finally, the armed forces have been placed into their own sector.

The second data strand contains the age profile of the population in each registration district, using age bands 0–14 years, 15–29 years, 30–44 years, 45–59 years, and 60 years and over. The third strand looks at the change in the size of the population of each registration district since the 1851 census. These second and third strands give us signals as to the shifts in population around the country in the mid nineteenth century, while the first strand enables us to look for reasons for the shifts that involve different occupations being located in varying concentrations in different parts of the country.

There are factors that, because of lack of space and the need to simplify our work, we have not been able to address. Notably, this analysis does not consider the employment and migration of women.

Cluster analysis

The technique that we apply to the 1861 registration district data is known as *cluster analysis*. The object of cluster analysis, as used here, is to identify cases that have similar profiles over the variables in a data set and place them together into groups or 'clusters'. In this work, cluster analysis enables us to group together registration districts that are similar with regard to the variables under consideration although they may be very far apart geographically. This grouping of the districts is useful from a practical point of view, as with 635 registration districts an analysis of each district individually is not really feasible. Discussion of the characteristics of registration districts in terms of clusters is very helpful in understanding the population geography of the country as a whole.

In this paper, we use a hierarchical type of cluster analysis. This starts by treating all the registration districts as distinct (that is, we imagine there are 635 groups, each containing one district). The two districts that are 'nearest' to each other, where by 'nearest' we mean having the most similar profiles over the variables in the data set, are then amalgamated to form a 'cluster' of two, which is then treated as a single group. After amalgamating the two most similar districts, therefore, we have 634 groups, 633 containing one district, and one containing two.

The process of amalgamation of the two closest groups is then repeated many times. There is a choice of methods (linkages) to use for forming the clusters in hierarchical cluster analysis. One frequently used type of linkage is Ward's method and it is this which we employ here. There are several text books that discuss cluster analysis. Prominent among these are books by Everitt, Manly and Chatfield and Collins.[4] A good example of an application of cluster analysis in a historical setting is given by Power and Campbell.[5] Any standard statistical software can be used to carry out the cluster analysis used in this paper. The authors used the Statistical Package for the Social Sciences (SPSS) but others (for example Minitab, SAS, or S-Plus) could have been used equally effectively.

One key aim of cluster analysis is to group the cases (in our example, the registration districts) into a small number of 'natural clusters', in which each cluster

Table 1 Mean number of males aged 20 years and over per thousand employed in various sectors: eight occupational clusters

	Number of districts	Manufacturing	Agriculture	Service occupations	Mining	Transport	Armed forces	% population change, 1851–1861
Total	635	289	240	136	59	64	22	+11.9
Cluster A	250	140	573	89	12	21	7	-0.9
Cluster B	149	167	406	111	50	43	14	+6.9
Cluster C	86	431	199	101	76	33	5	+13.0
Cluster D	39	177	215	79	337	38	4	+15.0
Cluster E	63	329	67	238	9	85	18	+16.2
Cluster F	12	226	132	142	8	62	227	+19.4
Cluster G	30	254	118	155	31	220	14	+24.4
Cluster H	6	127	101	83	6	43	497	+62.2

Note: This table should be read as follows. In the registration districts in Cluster A, an average of 140 out of every 1,000 males aged 20 years and over were employed in manufacturing, 573 in agriculture, 89 in service occupations, 12 in mining, 21 in transport and 7 in the armed forces.

includes registration districts which are similar to one another with respect to the variables being considered, but registration districts in different clusters are dissimilar. An important issue, therefore, is that of deciding how many 'natural clusters' exist in the data set. A review of many mathematical criteria for helping to make this decision is given by Milligan and Cooper.[6] Here we prefer to use an intuitive method developed by the authors of this paper.[7]

We carried out three cluster analyses, each using one of the strands of data mentioned above. The 'natural clusters' were identified and the characteristics of each such cluster examined. After carrying out the cluster analysis for each of the three strands of data, each registration district can be identified with an occupational cluster, an age profile cluster and a population change cluster. To examine population change in relation to occupation and age structure, the interrelationships between the occupational clusters, the age profile clusters, and the population change clusters are examined.

Clustering with regard to male occupational structure

A hierarchical cluster analysis was carried out using data for the six occupational sectors. Eight 'natural clusters' were identified. Table 1 shows the mean number of males aged 20 years and over employed in each occupational sector per thousand males aged 20 years and over, both nationally and for each of the occupational clusters, ordered according to their mean change in population between 1851 and 1861.

In cluster A are districts with a much higher than average number of males aged 20 and over employed in agriculture (we have called this the 'high agriculture' cluster).[8] Cluster B is similar except that agriculture dominates to a slightly lesser extent ('agriculture' cluster). Cluster C contains districts such as Coventry and Rochdale, employing a large proportion of people in manufacturing ('manufacturing' cluster). In cluster D are mining communities ('mining' cluster). Cluster E contains districts which are service centres with a noticeably larger than average number of males aged 20 and over working in service occupations ('services' cluster). Typically these districts are in London or are regional centres such as Derby and Worcester. Cluster F districts have a high proportion of men in the armed forces ('armed forces' cluster). These districts, such as Colchester and Plymouth are regional centres as well as military areas, and contrast with districts in cluster H such as Farnborough and Farnham which are dominated to an even greater extent by armed forces and are not regional centres. We label cluster H 'major armed forces'. In cluster G we have high numbers of people employed in transport ('transport' cluster). These districts are mainly ports such as Gravesend and Liverpool.

Clustering with regards to age profile

For each registration district, the proportion of the population in each age band (defined earlier) was calculated. A cluster analysis was carried out for the age bands using standardised data. Five 'natural clusters' were identified. In Table 2, the mean number of people per thousand in each of the age bands is shown for each of these clusters, ordered according to their mean population change. The mean number of deaths per thousand in 1861 and the mean sex ratio (males/females) are also shown.

Cluster 1 has a large proportion in the 60 years and over age group and to a lesser extent in the 45–59 year age group ('high 45 and over' cluster). This is associated with relatively low mortality and low or negative population growth, possibly caused by younger people leaving these (mainly rural) districts. Cluster 2 has a higher than average proportion of people in the 0–14 year age band ('high 0–14' cluster). It may be that these districts have seen a birth rate higher than the national average in the years preceding the 1861 census. Districts in cluster 3 are mainly regional centres. The proportion of people aged 45 years and over is high, and the proportion of people aged 0–14 years is low ('low 0–14, high 45 and over' cluster). In cluster 4, there is a low proportion aged 45 years and over ('high 0–44, low 45 plus' cluster). London districts and large towns in the north of England are prevalent in this cluster. Possibly

Table 2 Average percentages of population in each age band: five age profile clusters

	Number of districts	0-14 years	15–29 years	30–44 years	45–59 years	60 years and over	Crude death rate per thousand	% population change, 1851–1861
Total	635	36	27	19	12	7	21.68	+11.9
Cluster 1	331	36	24	17	13	10	19.07	-0.7
Cluster 2	148	37	26	18	12	8	20.70	+10.5
Cluster 3	43	32	27	19	13	9	21.82	+11.9
Cluster 4	92	37	28	19	11	5	24.06	+26.3
Cluster 5	21	30	33	21	11	6	22.17	+29.1

Note: This table should be read as follows. In the registration districts in Cluster 1, an average of 36 per cent of people were aged 0–14 years, 24 per cent of people aged 15–29 years, 17 per cent aged 30–44 years, 13 per cent aged 45–59 years and 10 per cent aged 60 years and over.

young adults have moved into these areas to work and started families, meaning that older people will form a smaller proportion of the total population in the districts. Their relatively high death rates may also be associated with the low proportion of elderly. Cluster 5 has a large proportion in the 15–29 year and 30–44 year age bands, and a low proportion aged under 15 years ('low 0–14, high 15–44' cluster). These districts are mainly in central London (people may have moved in to work but not yet started families) or are associated with military bases (where military personnel provide the reason for the unbalanced age profile).

Clustering with regard to population change

The percentage change in population from the previous census in 1851 to the 1861 census was used to cluster the registration districts with regard to population change. We concluded that seven 'natural clusters' exist. The mean percentage change in population between the 1851 and 1861 censuses is shown in Table 3 for each of the population change clusters.

We see that cluster I has, on average, a drop of almost six per cent in population from 1851 to 1861, possibly because of people moving away in search of work. In the largest cluster (cluster II), the population hardly changed from 1851 to 1861. In cluster III the population increased by almost 10 per cent on

Table 3 Population change clusters

	Number of districts	Average % population change, 1851–1861
Total	635	+11.9
Cluster I	126	-5.8
Cluster II	247	+0.5
Cluster III	125	+9.6
Cluster IV	67	+19.3
Cluster V	52	+32.5
Cluster VI	17	+66.7
Cluster VII	1	+161.5

average. A 19 per cent average increase in population is noted for districts in cluster IV, and even larger increases are observed in cluster V (32 per cent on average) and cluster VI (67 per cent on average). The largest increase in population occurs in cluster VII containing just one district, Farnham, where a military base had a large impact. It is likely that the increases in population for clusters IV, V and VI were caused by an influx of people looking for work. Differential death rates for the clusters are not responsible for the different levels of population change.

Relationships between clusters

Cross-tabulations of the occupational clusters and age profile clusters, of the occupational clusters and population change clusters, and of the age profile clusters and population change clusters are shown in Tables 4–6. For each comparison, these tables compare the actual numbers of registration districts in each cell and the numbers which would be expected if the two dimensions were independent. So, for example, the top left-hand cell of Table 4 (+80) shows that there are 80 more districts in this cell that we should expect if age profile and occupational profile were independent.

From Table 4 we see that the two agricultural clusters have more districts than expected in the first age profile cluster, where the proportion of people aged 45 and over is large. Younger people have probably left these rural (agricultural) districts, leaving older members of those communities. The 'manufacturing' and 'mining' clusters show similar patterns in that age profile clusters 1 and 3 are under-represented and age profile clusters 2 and 4 are

Table 4 Relationship between age profile clusters and occupational clusters

Occupational clusters (with descriptions)	Age profile clusters (with descriptions)					
	Cluster 1 (high 45 and over)	Cluster 2 (high 0–14)	Cluster 3 (low 0–14, high 45 and	Cluster 4 (high 0–44, low 45 and	Cluster 5 (low 0–14, high 15–44)	Number of districts in each row
Cluster A (high agriculture)	+80	–21	–14	–36	–8	250
Cluster B (agriculture)	+16	+7	+2	–21	–5	149
Cluster C (manufacturing)	–30	+14	–3	+22	–3	86
Cluster D (mining)	–13	+9	–3	+8	–1	39
Cluster E (services)	–31	–4	+13	+14	+8	63
Cluster F (armed forces)	–6	–3	+5	–1	+5	12
Cluster G (transport)	–13	–1	0	+15	–1	30
Cluster H (major armed forces)	–3	–1	0	–1	+6	6
Number of districts in each column	331	148	43	92	21	635

Notes: For interpretation, see text.

more highly represented than would be expected. It can be hypothesised that this effect has been brought about by the workforce in the manufacturing and mining industries starting families once they have migrated to these areas and found work.

The ‘services’ cluster has a deficit of districts in the ‘high 45 and over’ cluster, made up for by having more districts than expected in the age profile clusters 3 and 4. In age profile cluster 3 we may be observing older people in regional centres staying put while the younger generation migrate away (producing a low birth rate and smaller proportions in the 0–14 age group). In age profile cluster 4 we may be seeing an influx of people into the districts to work in the manufacturing, service or transport sectors causing the ‘high 0–44, low 45 and over’ pattern to be over-represented. With such small numbers of districts in the armed forces occupational clusters, it is difficult to discern any pattern

Table 5 Relationship between population change clusters and occupational clusters

Occupational clusters (with descriptions)	Population change clusters (with average percentage change)							
	Cluster I (5.8 % decrease)	Cluster II (0.5 % increase)	Cluster III (9.6 % increase)	Cluster IV (19.3 % increase)	Cluster V (32.5 % increase)	Cluster VI (66.7 % increase)	Cluster VII (161.5 % increase)	Number of districts in each row
Cluster A (high agriculture)	+29	+53	–31	–23	–20	–7	0	250
Cluster B (agriculture)	–6	–5	+14	+7	–7	–3	0	149
Cluster C (manufacturing)	–7	–16	+11	+6	+6	+1	0	86
Cluster D (mining)	–4	–9	+8	–1	+5	+1	0	39
Cluster E (services)	–5	–12	–2	+8	+7	+3	0	63
Cluster F (armed forces)	–2	–3	0	+3	+3	0	0	12
Cluster G (transport)	–5	–6	+2	0	+5	+4	0	30
Cluster H (major armed forces)	–1	–2	–1	0	+3	+1	+1	6
Number of districts in each column	126	247	125	67	52	17	1	635

Notes: For interpretation, see text.

with regards to the age profile cluster that we can claim to be meaningful.

From Table 5, we see that the prevalence of agriculture is clearly related to low or negative population change, probably largely as a result of migration from the rural areas to more industrialised areas.

Districts in the 'agriculture' and 'manufacturing' clusters show similar patterns, with more districts than would be expected in population change clusters III and IV. There is also a tendency for the manufacturing districts to have more representation in the higher population change clusters as well (perhaps due to migration and higher birth rates due to the presence of recently arrived

Table 6 Relationship between population change clusters and age profile clusters

Age profile clusters (with descriptions)	Population change clusters (with mean percentage change)							
	Cluster I (5.8 % decrease)	Cluster II (0.5 % increase)	Cluster III (9.6 % increase)	Cluster IV (19.3 % increase)	Cluster V (32.5 % increase)	Cluster VI (66.7 % increase)	Cluster VII (161.5 % increase)	Number of districts in each row
Cluster 1 (high 45 and over)	+44	+56	–33	–31	–27	–9	–1	331
Cluster 2 (high 0–14)	–22	–19	+34	+13	–3	–3	0	148
Cluster 3 (low 0–14, high 45 and over)	–6	–7	+9	+2	+1	0	0	43
Cluster 4 (high 0–44, low 45 and over)	–15	–27	–5	+13	+24	+10	0	92
Cluster 5 (low 0–14, high 15–44)	–1	–4	–4	+2	+4	+2	+1	21
Number of districts in each column	126	247	125	67	52	17	1	635

Notes: For interpretation, see text.

workers starting families). The 'services' and 'transport' clusters and both armed forces clusters are associated with population increases, again probably as a result of migration and higher birth rates due to recently arrived workers having children.

In Table 6 we see a very clear relationship between the age profile clusters and population change clusters. Age profile cluster 1, with a high proportion of people aged 45 and over, is clearly composed disproportionately of districts where population change was negative or zero. Districts in age profile clusters 4 and 5 (containing high proportions of young people of working age) are associated with substantial increases in population. Age profile clusters 2 and 3 are associated with more moderate population increases.

Conclusions

The use of cluster analysis in this work has enabled us to examine the characteristics of all 635 registration districts existing in the 1861 census. A detailed discussion of the individual characteristics of all the districts would prove unwieldy and confusing, but by clustering similar districts together, a simplified pattern emerges. This pattern is then much more amenable to interpretation.

We have carried out a cluster analysis of the registration districts using three different strands of data. This resulted in three different ways of clustering the districts. By then looking at the three separate two-way tables that could be produced, we were able to draw conclusions about the relationships that exist between the three strands of data.

In the patterning of Tables 4, 5 and 6, some of the well-known relationships between occupation, age structure and migration in mid-nineteenth century England and Wales can be discerned.[9] We see the movement of people in their younger working years away from the traditional occupations associated with agriculture towards jobs in towns and cities associated with service industries, manufacturing and transport. This movement of people has led to registration districts having age distributions biased towards surfeits and/or deficits of people in certain age bands, notably the 15–44 age range. Also, although England and Wales experienced a 11·9 per cent rise in population between the 1851 and 1861 censuses, the movement of people has meant that this increase was far from evenly distributed. Some areas saw increases far in excess of the 11·9 per cent as people moved in, and others saw very little change in the size of their population or even experienced population decline.

The methods used in this paper can be used to try and detect patterns in any large database. The census records in England and Wales are a rich source of information on the population and it would be of interest to look at strands of data other than those discussed in this paper. Analyses of the other Victorian and more modern censuses using these methods may yield results that can be compared and contrasted with those shown in this paper. Patterns of change may also be revealed. In addition, the resulting 'clusters' of similar registration districts could be mapped, revealing aspects of population geography. Finally, it is not only the census that yields large databases to which the methods shown here can be applied. The Great Britain Historical Database contains many data sets (including part of the 1861 census) that could be analysed using these methods.[10]

Acknowledgement

The authors would like to acknowledge the helpful comments and suggestions made by members of the *Local Population Studies* Editorial Board.

NOTES

1. D.A. Gatley, 'Computerising the 1861 Census and vital registration statistics', *Local Population Studies*, **58** (1997), 37–47.
2. The data are obtainable on disk from the Local Population History Book Club by contacting Dr

Peter Franklin, 46 Fountain Street, Accrington, Lancashire BB5 0QP.

3. C. Booth, 'Occupations of the people of the United Kingdom', *Journal of the Royal Statistical Society*, **49** (1886), 351–71; W.A. Armstrong, 'The use of information about occupations', in E.A. Wrigley ed., *Nineteenth-century society: essays in the use of quantitative methods for the study of social data*, (Cambridge, 1972), 226–52.
4. B.S. Everitt, *Cluster analysis*, 3rd edn (London, 1993); B.F.J. Manly, *Multivariate statistical methods :a primer*, (London, 1986); C. Chatfield and A.J. Collins, *Introduction to multivariate analysis*, (London, 1980).
5. J.P. Power and B.M.S. Campbell, 'Cluster analysis and the classification of medieval demesne-farming systems', *Transactions of the Institute of British Geographers*, new series, **17** (1992), 227–45. Other examples in settings accessible to the non-expert include F.J. Lapointe and P. Legendre, 'A classification of pure malt scotch whiskeys', *Journal of the Royal Statistical Society, Series C*, **43** (1994), 237–57; D. Owen, 'The spatial and socio-economic patterns of minority ethnic groups in Great Britain', *Scottish Geographical Magazine*, **111** (1995), 27–35; M. Silver, 'Scales of measurement and cluster-analysis – an application concerning market segments in the babyfood market', *Journal of the Royal Statistical Society, Series D*, **44** (1995), 101–12; and G. Sumner, 'Daily precipitation patterns over Wales – towards a detailed precipitation climatology', *Transactions of the Institute of British Geographers*, new series, **21** (1996), 157–76.
6. G.W. Milligan and M.C. Cooper, 'An examination of procedures for determining the number of clusters in a dataset', *Psychometrika*, **50** (1985), 159–79.
7. N.H. Spencer and D.A. Gatley, 'A graphical aid for determining the number of natural clusters in a dataset'. Unpublished typescript, available from the authors.
8. The names we have given the clusters here and in following sections should be regarded as tentative names to be used as labels rather than as descriptions of the essential natures of the clusters.
9. D. Friedlander and R.J. Roshier, 'A study of internal migration in England and Wales: part I', *Population Studies*, **19** (1966), 239–79; R. Lawton, 'Population changes in England and Wales in the later nineteenth century: an analysis of trends by registration district', *Transactions of the Institute of British Geographers*, **44** (1968), 55–74; E.G. Ravenstein, 'The laws of migration', *Journal of the Royal Statistical Society*, **48** (1885), 167–227; E.G. Ravenstein, 'The laws of migration', *Journal of the Royal Statistical Society*, **52** (1889), 214–301.
10. See http://hds.essex.ac.uk/gbh.asp.