

**The Quantitative Analysis of Family Names:
Historic Migration and the Present Day Neighbourhood Structure of
Middlesbrough, United Kingdom**

Paul A. Longley*, Richard Webber*, Daryl Lloyd**

*Department of Geography, University College London, 21 Bedford Way, London
WC1H 0AP, United Kingdom. plongley@geog.ucl.ac.uk

** Cabinet Office, 70 Whitehall, London SW1A 2AS, United Kingdom.
daryl.lloyd@cabinet-office.x.gsi.gov.uk

Abstract

This paper describes the way in which a unique new quantitative data resource and evidence base has been used to relate historic measures of U.K. migration flows to the contemporary socio-economic patterning of neighbourhoods. The resource enables generalised analysis of the regional origins of British and Irish people from their family names, and makes it possible to relate the current regional distribution of names in the US, Great Britain and other English speaking countries to equivalent information from the Great Britain Census of 1881. Illustrative applications may be viewed at www.spatial-literacy.org. In this paper we develop a number of indices of the historic origins of English and Irish family names, as a prelude to detailed micro scale analysis of late twentieth century surname patterns. We illustrate the usefulness of these various indices through case study analysis of Middlesbrough and East Cleveland, an area of the U.K. which attracted large numbers of economic migrants during its rapid nineteenth century industrialisation. We use our quantitative evidence of the historical distributions of different family names in order to characterise the social mobility of descendants of Scottish, Irish and Cornish migrants, and to evaluate the practice of ascribing family names to particular localities in historical GIS. The case study thus illustrates the ways in which our data resource may be used to substantiate existing thinking about historic migration and residential structure, as well as to generate and investigate new hypotheses that might guide future work.

Keywords: *family names, migration, neighbourhood profiling*

1. Introduction

This paper takes as its starting point the observation that recent work on historical migration patterns and their contemporary outcomes has refocused away from approaches that are “mechanistic and quantitative” towards approaches that are deemed more adept at identifying processes “of social and cultural change affecting both individuals and communities” (Pooley and Turnbull 1998, 327). Our own concerns echo those of Pooley and Turnbull (1998, 330) that, in emphasising the structural context of society and the social and cultural processes of change, “too many studies in the past have .. overstated the noticeable and atypical aspects of migration and mobility, and have neglected the everyday and commonplace dimensions of population movement”. Moreover, we observe that studies of contemporary residential differentiation provide only blunt tools for generalising about social structure. Case study representations may not be securely anchored in any rich and generalised understanding of past or present circumstances: data pertaining to the past are often piecemeal and conventional archive sources require painstaking, labour-intensive attention; and quantitative analysis of the present is often restricted to a very limited range of datasets (such as population censuses) that tell us frustratingly little about what is actually going on in contemporary society. Yet generalised representation of structure and change in society requires that the individuals and their activities have known and prespecified chances of inclusion.

Our own previous work in Great Britain (G.B.) has focused more upon the present than the past. In various papers (e.g. Longley and Harris 1999; Webber and Longley 2003; Ashby and Longley 2005) we have focused upon the application of enhanced

geodemographic indicators to problems of public service delivery. The motivation for the research reported here is the realisation that it is now possible to link these rich, spatially disaggregate representations of the present to rich generalised representations of nineteenth century society. This makes it possible to take a long term, generalisable perspective upon historic migration patterns, and to undertake detailed geographical reinterpretations of social and residential mobility in the present day.

We believe that we have achieved such linkage through the quantitative analysis of British and Irish family names, using digital records for every G.B. individual in 1881 and every enfranchised G.B. adult in 1998. We see our primary contribution as creating a generalised evidence base for historical GIS, a demonstration of which is available at www.spatial-literacy.org and which is available in its entirety for academic usage from the U.K. Data Archive (www.essex.ac.uk). We believe that there is a range of research applications of this resource, only some of which can be addressed in a single academic journal article such as this. In the case study that we develop here, we illustrate the use of our dataset to address a range of issues of generality that may be framed in terms of Ravenstein's (1885) classic Laws of Migration (see Sui 2004 and associated articles in a forum discussion for a discussion of the remit of "laws" in geography). These include Ravenstein's assertions that most migrants moved in a series of steps (creating a wave-like movement of migrants into urban centres), and that every migratory current has a counter-current. In so doing, our objective is not so much to revive the somewhat discredited notion of a grand theory which "binds the various traditions together into the immense tapestry of modernisation" (Zelinsky 1993, 218), but rather to provide

consistent, generalised context for study at a range of spatial scales, and through GIS to make the sub-discipline of historical geography more relevant to the study of contemporary migration, labour markets and social mobility.

We believe that our data resource can be used to create evidence-based benchmarks for historical GIS and urban geography across a range of scales. It also opens up prospects for a more data led approach to some areas of historical geography, that achieves generality in similar ways to much current research in GIS and spatial analysis. We illustrate these ideas here in a case study of Middlesbrough, in which we use evidence to suggest a range of hypotheses about the historic and present day social structure of Middlesbrough and East Cleveland, U.K.. By extension, this approach is applicable at other scales from the regional to the international, in any Anglophone country covered in our dataset. Taken together, this work provides valuable context for the study of the nature and extent of local, regional, national and international residential mobility, the motivations for migration and the links between mobility and socioeconomic circumstances.

2. Family names as quantitative indicators of historic migration and social mobility

There has long been recognition of the relationship between Anglophone family names (or surnames) and places (e.g. Schürer 2004). For example, the Dictionary of American Family Names (Hanks 2002) provides detailed information not just on the country of origin or each of the most frequently occurring names in the United States, but for many family names it also provides information on the locality or region of the emigrant

country in which the names originated. The most obvious reason for the association between family names and localities is that it was often the practice, when family names were first adopted in the late Middle Ages, for people to be assigned family names on the basis of the locality from which they originated. Names of this sort are known as toponyms, and people with names such as Illingworth or Cheney provide examples of English toponyms that are still significantly over-represented in communities close to the places from which they originate (e.g. see Longley et al. 2005, 8-11). Family names such as Prager and Bhopal, now increasingly common in Britain, provide evidence that the practice of naming people by place of origin is not confined to Great Britain. Variation across Great Britain in the terms once given to topographic features also contributes to local variations in surname frequencies. For example family names ending in –bottom are most common in the North and in –thorp in the East Midlands, as can be seen in Figure 1, whilst family names ending in –combe and –cott are still much more common in the South West than in other regions of the country. By contrast the practice of selecting placenames as family names was much less common in Ireland, Scotland and Wales.

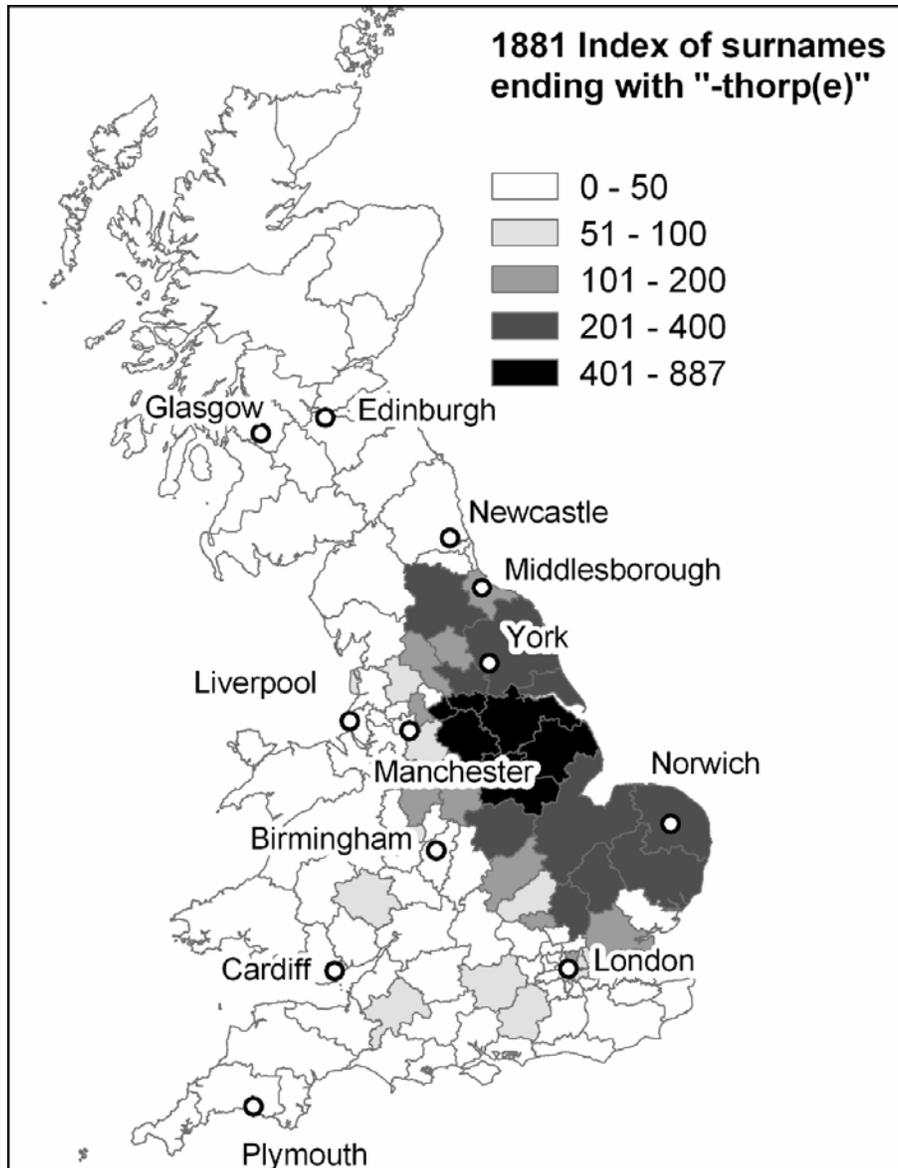


Figure 1: Frequency of persons with names ending in –thorp in 1881, postal area index values (Great Britain average = 100)

Other family names are derived from occupational groupings. These are known as metonyms (Reaney and Wilson 1997). In Britain, since some of these occupations were specific to individual regions, many of these names are highly regionalised. Regional differences in the terms used to describe different occupations also contribute to regional variations in surname frequency. For example the occupation described as a “Walker” in

the North of England is referred to as a “Fuller” in the South East and as a “Tucker” in the South West (Schürer 2004; Longley et al. 2005, 131-2). A third important class of family names contains those such as Jones and Johnson that derive from a parental name, which can be male or female. Such family names are known as patronyms. Patronyms are particularly common in Wales and the South West, where it was common for an “s” to be added to a first name, as in Richards, Nicholls. In the North East of England, where patronyms are also common, a more common practice was to add the suffix –son, as in Richardson or Nicholson.

Immigration from Scotland, Ireland, Wales and Cornwall has contributed in Britain as in the United States to many other categories of Anglophone surname while a new series of commonly occurring family names, of which Goldstein, Patel and Hussain are examples, has been added more recently as a result of the immigration of Jews, Asians and Muslims. Indeed of all immigrants groups only those from the Caribbean have not brought with them a distinctive set of family names. Given the tendency of these immigrant groups to settle in particular localities, whether in Great Britain, the United States or Ireland, their family names are often even more geographically concentrated than names of Anglo-Saxon, Norman or continental European origin.

Surname specialists have taken an interest in the regionality of family names for over seventy years. However, historical geographers have usually only been able to piece together regional or national pictures by assembling the results of sometimes disparate locality studies (e.g. Dodgson and Butlin 1978). Family names have previously been

used to monitor migration behaviour and labour markets by Long (2005), in a longitudinal analysis of 1851 and 1881 Census data that seeks to establish the economic efficiency of migration flows. This latter analysis is, however, inevitably partial and piecemeal: Long restricts his analysis to a 2% sample of 1851 Census records and his apparent over-all 41% success rate in matching members of this sample to the 1881 Census likely conceals a number of sources of bias – not least because many individuals with common family names create multiple or ambiguous matches and are necessarily excluded from the analysis. While such work contributes to the important task of systematising our understanding of migration flows, it falls well short of providing universal analysis of any aspects of population change. More generally, Pooley and Turnbull's (1998, 319) work suggests that “research based on a sample of longitudinal migration histories is not a substitute for aggregate studies of net and gross migration flows between administrative units derived from census and vital registration data”. However, such study can be extremely labour intensive and areal analysis is often frustrated by changes in administrative geographies (Gregory and Southall 1998). Taken together, this has meant that our ability to use analysis of the past to understand the present is frustrated by the incompleteness of our geographic representations of the past, and the absence of clear correspondence tables between historic and current geographic boundaries.

It is in this context that our research seeks to establish connections between two hitherto largely unrelated literatures and to contribute to each. First, it offers a contribution to our systematic understanding the historical geography of inter-regional migration patterns,

through a Great Britain wide classification and mapping of all common family names in 1881. Second, though extension and linkage of this classification to the 1998 Great Britain Register of Electors, we investigate the present day socio economic situation of migrants, and thus make a systematic contribution to the study of social mobility and neighbourhood patterning (Johnston 2000; Harris, Webber, and Sleight 2005). This is achieved by using GIS to classify and link rich individual level population datasets pertaining to 1881 and 1998. As such, our primary motivation is not to develop new methodological or technological insights but rather, through data integration, to demonstrate the richness and relevance of GIS to migration and social geography, past and present. We develop and illustrate these ideas using a case study, based on Middlesbrough.

The data resource that we have created and used in our case study opens up a number of additional potentially fruitful directions for analysis. In addition to the national frequency of any given name, both in Great Britain and in other countries of the Anglophone diaspora¹, the attributes of our G.B. dataset include a number of measures of geographical distribution such as the identity of the postcode area² and of the postcode district (defined below) which has the highest concentration of each family name in 1998; the identity of the postcode area which had the highest concentration of each family name in 1881; the level of concentration of each name in its historic heartland at both dates; and the relative change in the frequency of occurrence of each name between 1881 and 1998. Information is also held on the frequency of the name in the US, Australia and New Zealand and on the state, statistical sub division and province

that has the highest concentrations of each name. We have also assigned the each of these family names to one of a set of 224 different name types, such as patronyms ending in –son, toponyms ending in –combe or metonyms relating to occupations. This classification of names also includes a number of classes of name “imported from abroad” based on country or culture of origin. These attributes may be viewed at www.spatial-literacy.org – where it is also possible to view the geographic distribution of 25,630 individual family names in 1881 and 1998.

3. Past migration to Middlesbrough and its contemporary social structure: data sources

Our intention is to illustrate here how our national database of name distributions can contribute to a generalised understanding of inter-regional migration patterns, and also help us to understand the level of social and economic mobility (as manifest in contemporary residential structure) that has occurred over the last 120 years. We do this by developing an historical analysis of contemporary name distributions that is also classified according to a leading contemporary neighbourhood (geodemographic) typology. The names database is used to identify the areas of the British Isles from which Middlesbrough and East Cleveland derived its expanding nineteenth century population and to examine the detailed residential status of the current generations that are descended from these different immigrant communities.

Unlike many other northern industrial towns, Middlesbrough was no more than a fishing village in 1831, with 383 residents enumerated in the census of that year. With the

development of a coal port and in 1840 of a steel works, the population of the town grew to 5,709 in 1841 and to 56,000 in 1881. During this time it attracted a disproportionate number of migrants from the remoter corners of the British Isles. In these particular circumstances it seems reasonable to assume that migrants to Middlesbrough, unlike those that moved to London, were not assimilated into any existing, invariant or immutable “host” culture or social structure (see Feldman 2001). 1881 thus provides an appropriate baseline against which to measure changes in Middlesbrough’s residential structure. We concur with historians such as Renton (2005) and Feldman (2001) that migration studies may be used to investigate the long term outcomes of social and political relations upon the opportunities realised by migrants, and suggest that our database provides a comprehensive and fully inclusive framework within which to analyse outcomes. As such it provides a detailed backcloth against which hitherto unconnected local case studies may be viewed.

In generalising about the geographic origins and long term social mobility of economic migrants, we use GIS to create and link a number of data sources, the three most important of which are as follows. The first geographically referenced data source that we have created contains information relating to each of the 25,630 family names for which more than a hundred occurrences appear on the 1998 versions of electoral registers in Great Britain³. The raw information on which this database is built is supplied by the information services company Experian (Nottingham, U.K.), who arrange for the annual data capture of the names and addresses of all electors entitled to vote in British elections. As such, it includes resident citizens of the Irish Republic and

of the (then) 11 continental European Union nations who were entitled to vote in some elections. These 25,630 names account for 37,246,881 registered electors and range from Smith, the most numerous with 514,898 occurrences, to 147 different names with exactly one hundred and one occurrences. These 25,630 names can be ascribed to the four levels of U.K. post-(Zip-)code geography: 121 Postcode Areas (henceforth Level 1 areas, such as Tees-side, TS: each covers an area with an average of approximately 0.5 million individuals); 2,827 Postcode Districts (Level 2 areas); 9,487 Postcode Sectors (Level 3 areas) and approximately 1.7 million Unit Postcodes (Level 4 areas). Unit Postcodes are the U.K. equivalent of Zip+4 codes in the US, and are broadly similar in size to Canadian postcodes. This geography is illustrated in Figure 2.

For the Level 1 postcode areas it is meaningful to compare the counts of each name with the total number of electors in each postcode area who have names that occur more than (the threshold of) a hundred times nationally. This has been done both for the 1998 electoral register and for the 1881 census. Because postcodes did not exist at the time of the 1881 census counts at the level of parish have been accumulated to the 1998 postcode area in which the largest proportion of each 1881 parish falls. By comparing family name frequencies with population counts it is possible to identify whether any name is more or less common in each Level 1 postcode area, relative to the total number of electors (or in the case of 1881, total enumerated population), than it is across the country as a whole, and by how much. Such measures are expressed in the form of an “index” value, where a value of 100 indicates a proportional level of frequency of a name in a postcode area equal to its national average frequency. Thus a value of 200

indicates twice as many occurrences of a name in a postcode area as one would expect on the basis of a purely uniform national distribution. It is a reasonable working assumption that the greatest single concentration of any surname remains in or very close to the postcode area in which it originated. Thus, by identifying for each name the Level 1 postcode area with the highest index value it is possible to locate the region of the country from which the name is most likely to have originated. This can be considered as the name's "heartland".

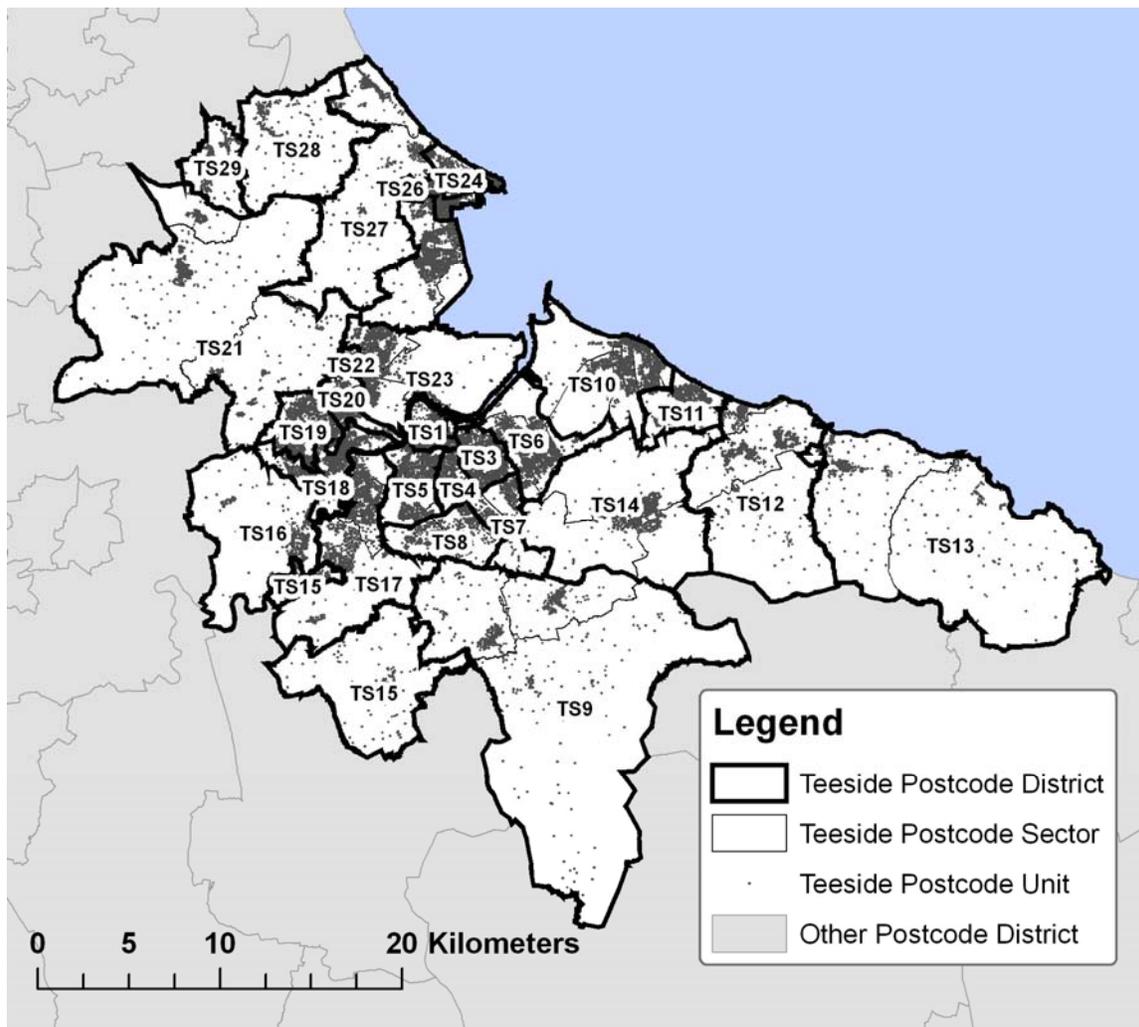


Figure 2: Teeside Postal Geography

The U.K. 1881 Census of Population⁴ is one in a series of compulsory national surveys which has been carried out at least every 10 years (except in 1941) since 1801. Census statistics are available only in summary form until the original household forms are released on the 100 year anniversary of data collection. The 1881 Census is unique, in that data from the individual enumerator sheets of the 1881 Census have been captured digitally. Unlike the 1998 U.K. Electoral Register, the 1881 Census summary file contains information on all residents (adults and children, electors and non-electors). Comparison of the numbers of occurrences of a name on the 1881 Census with the number on the 1998 Electoral Register provides useful information, both at local and national level, on the relative change in the frequency of particular family names and in their geographical concentration.

Whereas this national database provides statistics for each of these 25,630 names at the level of the postcode area, for purposes of our case study we extracted from the 2003 Electoral Roll the names and addresses of all registered electors in the two parliamentary constituencies of Middlesbrough and of Middlesbrough South and Cleveland East⁵ (henceforth referred to as the “study area”). The area covered by these two constituencies is shown in Figure 3. The 2003 Electoral Register, supplied to us by Experian, contains information on the names and addresses of 147,110 electors⁶ in the study area. 12,073 different names are recorded on these registers, 8,261 of which are also found on the list of national names containing more than one hundred electors nationwide. These 8,261 names account for 137,619 (93.5%) of the 147,110 electors in

the study area. The five most common names in the study area are Smith (2,104 occurrences), Brown (1,197), Robinson (1,147), Wilson (1,104) and Thompson (964). By contrast, for example, the files contained no occurrences of the name Hodges (a name associated with the Newport (NP) postcode area), despite there being 10,230 occurrences of Hodges nationally. Nor were there any occurrences of the names Davenport or Rimmer even though there are over 7,000 G.B. electors with each name nationally.



Figure 3: The study area: Middlesbrough, and Middlesbrough South and Cleveland East

The third database used in this study is a national classification of unit postcodes (Level 4), known as the U.K. “Mosaic” classification, which is a proprietary geodemographic product of Experian (see Thrall 2002 and Harris, Webber, and Sleight 2005 for discussion and critique of geodemographics in general and Webber 2004a for an account of the building of Mosaic in particular⁷). Equivalent classifications have been built in the US, in Australia, in New Zealand, in the Republic of Ireland and in 17 other countries

around the world. Other systems, such as Prizm and Tapestry (in the US) and ACORN (in the U.K.) are available (Thrall 2002). The 2003 U.K. Mosaic classification uses a set of nearly 400 different statistical items known about a Unit Postcode (Level 4) or its immediately surrounding area in order to identify sets of postcodes which are broadly similar across a far wider range of demographic and socio-economic characteristics than census data alone. It identifies a set of 61 distinct types of residential neighbourhood which are known as “Mosaic types”. These 61 classifications are then ordered into a coarser set of 11 “Mosaic groups”. These types and groups are devised principally for use by commercial organisations to target their recruitment of new customers (Sleight 2004) and to optimise the management of existing customer relationships. Lately they have been increasingly used for the development of local strategies for policing, health, education and party political campaigning (e.g. Ashby and Longley 2005, Webber 2006). Commercial users of the system characterise each type by a convenient shorthand label, such as “Cultural Leadership” and “Bungalow Retirement” as well as more detailed “pen portraits” of their defining characteristics (details of these “pen portraits” can be downloaded from www.business-strategies.co.uk). Some public sector users simply characterise each type by a three digit code, such as A01 where the leading character denotes the group and the subsequent two digits the full type. Some academic users use the labels, some the codes.

Evidence from market research surveys demonstrates that when the consumption of products and services is cross tabulated by Mosaic, it provides considerably more discriminatory power than do individual person or household level classifiers such as

age, income or social class (e.g. Brown et al. 2001; Webber 2004b). Additionally, because the categories are engineered to be homogeneous, inferences drawn from analyses based on them are less vulnerable than many cross classifications to the criticism that they could be subject to ecological fallacy. For purposes of this study, we appended Mosaic type and Mosaic group codes to each of the 137,619 records on the Electoral Registers of Middlesbrough and of Middlesbrough South and Cleveland East, for each of the names that surpassed our arbitrary threshold of a hundred occurrences nationwide.

4. Analysis of past migration using the current mix of names in the study area

In order to create indicators of the life experiences of the descendants of Middlesbrough's historic immigrant groups, we have developed a classification of family names in order to assign membership to one of four groups. The first of these groups comprises "long distance" migrants of Asian, Irish, Scottish and Cornish descent⁸. Welsh names were not included in this classification since the data suggested there were fewer such names in the study area than in other areas of comparable size and distance from Wales. In other words Middlesbrough has never been a popular destination for Welsh migrants. The three other groups correspond to three bands of "shorter distance" migrants: people descended from families local to Middlesbrough; people descended from families from other parts of North East England; and people descended from families from other areas of Great Britain excluding Scotland, Wales and Cornwall. Our motivation for classifying the past and present residents of Middlesbrough in this way was to identify the extent of migration from different

sources, as both a means of characterising Middlesbrough's historic cultural mix and as a way of understanding its contemporary social structure. As such, our analysis can be seen as an investigation to ascertain the nature and extent of linkages between social and spatial mobility, as manifest in contemporary neighbourhood patterning.

Whilst Asian, Irish and Scottish names are relatively easy to define solely on the basis of linguistic structure⁹, this is to a lesser extent the case with Cornish names, a significant number of which can only be defined by examining their concentration in Cornwall¹⁰. The method used in our broader study was to identify for each name the postcode area in which it has the highest concentration. The Level 1 Postcode Area that covers most of Cornwall is TR (Truro). From those names which were more common in Truro (TR: Level 1) postcode area than any other postcode area, we developed index values based upon national incidences of each name type, and then defined as "Cornish" those names whose index in postcode area TR exceeded 2000. In other words to be considered Cornish, we applied a rule that any name had to be at least twenty times more common in the Truro postcode area than in Great Britain as a whole¹¹.

In addition to these four cultural groups we also sought to identify a number of types of English name. One group of English names which we were keen to identify were names peculiar to Middlesbrough: names which, because they are still hardly found anywhere else in the country in any concentration, are likely to have been native to the study area prior to the onset of industrialisation in the nineteenth century. This group was defined in a manner not dissimilar to the method used to define Cornish names, this being that the

names should be at least twenty times more common in the study area than they are in the country as a whole. This condition was satisfied by a set of 168 names containing 42,518 (or 0.11%) Great Britain electors. These names accounted for 3.67% of all study area electors. Each of these names has relatively few occurrences elsewhere in the North East region. Profitt, Postgate, Cairnes, Limon, Mendum and Harbisher are the six names on this list which have the highest representation in the study area compared to the country as a whole.

The second group of English names was defined in such a way as to identify residents whose ancestors are likely to have originated from areas of North East England other than the immediate vicinity of Middlesbrough and who, unlike the previous group, are likely to have migrated a relatively short distance from their ancestral homelands to live and work in Middlesbrough. This group of names is defined using a different set of criteria. To qualify as a member of this group a resident's name has to be relatively common, in this case with a minimum of 20,000 occurrences nationwide, as well as to have an index within the study area of 175 or more. A third requirement is that the name must also not fall within any of the ethnic categories previously defined. This last criterion removes two names from this list, Hussain and Quinn. Together, these selection criteria identified 27 names. The ones on this list with the highest index values for the study area are Hodgson, Robson, Stephenson, Hutchinson and Atkinson. Eighteen of these 27 names are patronyms ending in -son, including the eight with the highest index values. This provides evidence of the extent to which specific naming conventions as well as specific names can be associated with particular regions within the country. It is

interesting to note that the name Hodgson, which leads this list, has the same meaning, son of Hodge or Roger, as Hodges. Earlier we have seen that Hodges is the name with the largest number of occurrences nationwide which is not present at all in the study area.

In the third group of English names we seek to identify residents whose family names originate from a “heartland” region of Britain far away from Teesside other than Scotland, Cornwall or Wales. To qualify for membership, we stipulated that names had to be relatively common at a national level, that is to have a minimum of 20,000 occurrences in 1998, but to have an index in the study area of 66 or lower. These people, one would suppose, are particularly likely to be descendants of people originating from regions of the country which had plentiful employment opportunities during the late nineteenth century or which, if they did not, were more conveniently placed in relation to other employment magnets than they were to Middlesbrough. Names in this list with the lowest index are: Barton, Sutton and Walsh, all of which are centred in the North West; Griffin, which is centred on Dudley (DY) in the Midlands; and Wheeler and Miles, both of which are centred on the Swindon (SN) postcode area in the South West. There are reasonable grounds to suppose that study area residents with these names are relatively likely to be recent migrants to the area or descendants of families who arrived more recently than the economic migrants from the Celtic periphery.

The results of our analysis are shown in Table 1: the top row shows the number of electors within the study area and within Great Britain whose names have more than one

hundred occurrences nationwide. The other rows of Table 1 show the corresponding number and percentage of these electors whose names fall into the various cultural groupings whose current life experiences we are interested in comparing. As can be seen from the “Index” column, electors with Irish names are very overrepresented in Middlesbrough, their level being some 33% above the national average. The proportion of electors with Asian names is broadly in line with the national average whilst the proportions with Scottish and Cornish names are just over one half the national average rates. The disparity between the apparent concentrations of Scottish and Cornish names as compared with Asian and Irish names in the final column of this table arises because the Great Britain register includes Scotland and Cornwall, not Ireland or South Asia. Thus the index value identifies that, relative to the rest of Great Britain, Irish family names are concentrated in Middlesbrough and Asian names are not: however, the index value for Cornwall suggests the local concentration of Cornish names relative to surrounding areas in the North East. The Postcode Area (Level 1) that includes the study area is Tees-side, and the adjacent postcode areas of York (YO), Durham (DH) and Darlington (DL) all show lower concentration of both Scottish and Cornish names.

	Study Area Index	Study area	G.B.	Study area %	G.B. %	Index
All names (>100 in G.B.)	Any	137,619	37,246,881	100.00	100.00	100
Asian	Any	3,377	958,797	2.45	2.57	95
Irish	Any	7,019	1,429,599	5.10	3.84	133
Scottish	Any	2,157	1,142,092	1.57	3.07	51
Cornish	Any	95	41,401	0.07	0.11	62
Middlesbrough	over 2000	5,046	42,518	3.67	0.11	3212
North Eastern (G.B. > 20k)	over 175	2,429	1,287,073	1.77	3.46	51
Other English (G.B. > 20k)	under 66	11,206	1,334,699	8.14	3.58	227

Table 1: Study area electors by type of name (source: 1998 Electoral Register)

Ethnic Group	Study Area	Study Area %	G.B. %	Index
White: British (inc. White Scottish)	173,614	94.03	88.2	107
White: Irish	877	0.47	1.21	39
White: Other White	1,435	0.78	2.49	31
Asian or Asian British	6,220	3.37	4.08	83
Black or Black British	479	0.26	2.01	13
Chinese or Other Ethnic Group	658	0.36	0.83	43
Mixed	1,359	0.74	1.18	62
All People	184,642	100.00	100.00	100.00

Table 2: Population by ethnic group in the study area in 2001 (source: 2001 Census of Population)

Comparing Tables 1 and 2 it is evident that a slightly smaller proportion of names on the 1998 Electoral Roll appear to be of Asian origin compared with the proportion of people giving Asian as their ethnic status on the 2001 Census form. By contrast the proportion of people in the study area who have Irish names is ten times greater than the proportion of people who describe themselves as Irish in the Census. The respective figure for Great Britain (calculated from our dataset) is three times. An implication of this analysis is therefore that the study area is a place where a disproportionate number of the descendants of Irish immigrants no longer describe themselves as Irish. Moreover, it suggests that the self assignment of individuals to cultural identities, in Middlesbrough at least, provides a very unreliable guide to the population size and geography of descendants of a group which may have immigrated over a hundred years ago. Therefore it is only by studying the location of electors with Irish names that we can identify what has happened to the descendants of the town's original Irish immigrants.

5. Analysis of contemporary neighbourhood structure by name group

Having defined these seven mutually exclusive populations, we turned our attention to the present day distribution of these groups in present day Middlesbrough. Ever since the work of the Chicago School in the 1920s, it has been a fundamental tenet of the residential differentiation literature that the dynamics of immigration are manifest in residential differentiation and urban social patterning (e.g. see Johnston 2000), with the implication that migrants “filter” upwards to successively more desirable neighbourhoods. Middlesbrough is no longer characterised by recurring waves of immigration, and the lack of identification of Irish descendents with their ancestry suggests an absence of cultural and ethnic divisions that has also been noted elsewhere (e.g. Renton 2005). In such circumstances, we might anticipate that the different migrant groups of a society in which social mobility was the norm would be distributed uniformly across the study area. Conversely, if residents belonging to certain name groups were more likely to find themselves living in certain types of neighbourhood than others (Figure 4 and Table 3), the opposite might be true. Our dataset of family names offers a novel way of re-engaging with empirical assessment of the “filtering” of individuals through different urban social areas (e.g. see Saunders 2001), theories which are also central to the development of generalised historical GIS (Gregory and Southall 1998) and are also undergoing some rehabilitation in mainstream urban studies.

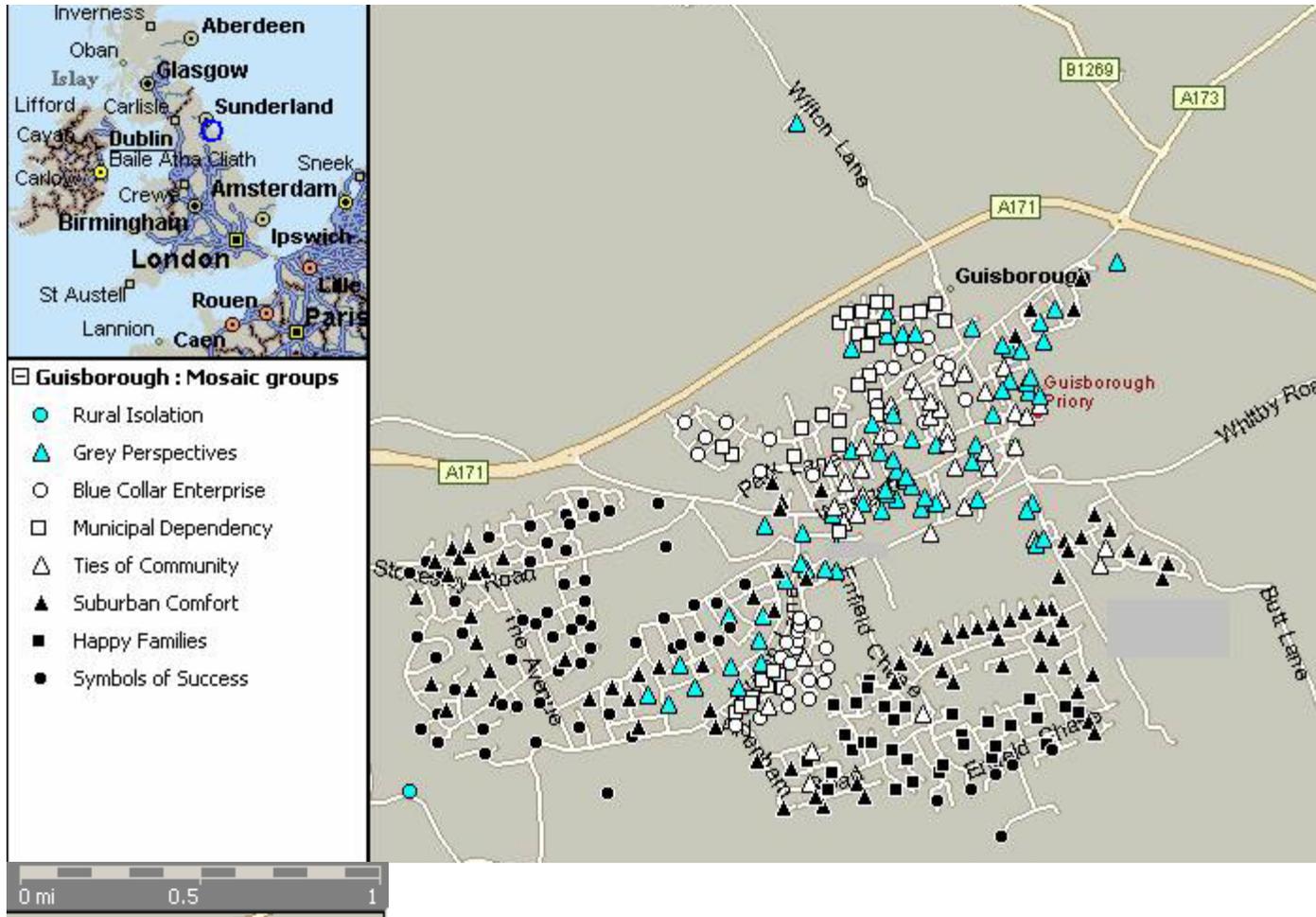


Figure 4: Illustration of postcodes by type of neighbourhood in the Guisborough part of the study area

Type of neighbourhood	Type of name							
	All	Asian	Irish	Scottish	Cornish	Middles brough names	North Eastern	Other English
A Symbols of Success ¹	100	65	85	116	39	54	131	97
B Happy Families ²	100	54	93	106	75	104	92	104
C Suburban Comfort ³	100	90	89	101	94	87	111	103
D Ties of Community ⁴	100	246	94	91	131	107	92	97
E Urban Intelligence ⁵	100	329	104	71	218	48	114	72
F Welfare Borderline ⁶	100	98	135	95	85	107	108	92
G Municipal Dependency ⁷	100	26	126	110	96	121	95	98
H Blue Collar Enterprise ⁸	100	28	87	95	156	111	82	113
I Twilight Subsistence ⁹	100	25	90	89	67	90	108	101
J Grey Perspectives ¹⁰	100	42	75	93	50	63	122	104
K Rural Isolation ¹¹	100	0	58	69	41	66	72	101

Table 3: Relationship between name type and neighbourhood type, study area, 1998, index values (study area average = 100).

Using the Experian Electoral Register database, we therefore coded each elector according to the current geodemographic (in this study, Mosaic) type and group of his or her Unit Postcode (Level 4). Some Middlesbrough examples of these neighbourhoods are shown in Figure 5. As would be expected, the study area shows a significant over-representation of the Mosaic groups Municipal Dependency and Blue Collar Owners and, compared with the country as a whole, relatively few electors resident in neighbourhoods characterised by Urban Intelligence. By comparing the proportion of electors in each Mosaic type belonging to each name type with the total number of electors, we have been able to create a table of index values indicating the extent to which different types of name are concentrated in different categories of neighbourhood. These index values are shown in Table 3.



Figure 5: Example of different types of neighbourhood in the study area : clockwise (from top left), group B (Happy Families) in Guisborough, group J (Grey Perspectives) in Saltburn by the Sea, group D (Ties of Community) in Middlesbrough and group G (Municipal Dependency) in Middlesbrough

Of the eleven Mosaic groups, by far the most prestigious in conventional socioeconomic terms is what commercial users of the system would call Symbols of Success (Group A). In these neighbourhoods of higher status households we find particularly high proportions of people with English names from regions other than the North East. Houses in such neighbourhoods are particularly likely to have residents with names such as Butler, Gardiner or Parker, typically Southern or Midland names which were poorly represented in the Middlesbrough area a hundred years ago. Residents are also relatively more likely to be named McDonald or Christie than Murphy or O'Connor. By contrast

these are not the neighbourhoods in which people have old standing Middlesbrough names. Asian names are also under-represented, although less so that would probably have been the case ten years ago.

By contrast the least prestigious of the Mosaic neighbourhood groups are Ties of Community, Welfare Borderline and Municipal Dependency. Asian names and, to a degree, Cornish names are disproportionately represented in Ties of Community (though for different reasons as will become clear below). The descendants of the Irish migrants are more likely to be found in the areas of inner city and peripheral council (public) housing (Welfare Borderline and Municipal Dependency), which they occupy to a much greater degree than the Scots. Whilst Asian names are numerous among inner city council estates (Welfare Borderline) relatively few Asians occupy tenancies on the deprived peripheral council states. People with local names are particularly likely to live in areas classified by Mosaic as of “Municipal Dependency”.

Other patterns are the tendency for people with Cornish names or names typical of the North East to live in the better and increasingly privatised low rise council estates (Blue Collar Enterprise) and for the people with names from elsewhere in Britain to live in areas of Grey Perspectives, such as in Saltburn by the Sea (Figure 5). None of the seven types of name is well represented in areas of Rural Isolation.

The correspondence between name type and geography is equally evident when we examine the distribution of name types across the study area (Table 4). Asian names are

highly concentrated in the central districts of Middlesbrough whilst names of Irish and Scottish origin are much more evenly diffused. Nevertheless there is clear evidence that the suburbanisation of the Irish is stronger in areas of overspill council housing than it is in areas of high status or young family housing. Neither the Irish nor the Scots have moved out in significant numbers into the rural, small town or retirement neighbourhoods in the eastern parts of the study area.

Location	Postcode Sector	Asian	Irish	Scottish	Cornish	North East names	Non Local English names
Middlesbrough Central	TS1 1	61	128	190	0	78	114
Middlesbrough Central	TS1 2	745	83	63	26	81	101
Middlesbrough Central	TS1 3	896	101	71	71	60	82
Middlesbrough Central	TS1 4	338	108	87	51	88	103
Middlesbrough Central	TS1 5	62	137	107	132	109	86
Middlesbrough : Port Clarence	TS2 1	13	136	148	160	99	55
Middlesbrough : Ormesby	TS3 0	18	134	113	107	86	109
Middlesbrough : North Ormesby	TS3 6	103	141	114	0	103	115
Middlesbrough : Ormesby	TS3 7	17	143	126	30	99	80
Middlesbrough : Ormesby	TS3 8	28	137	121	53	103	96
Middlesbrough : Ormesby	TS3 9	22	130	116	58	106	88
Middlesbrough : Marton Road	TS4 2	315	133	89	104	93	80
Middlesbrough : Marton Road	TS4 3	35	151	96	37	101	122
Middlesbrough : Acklam	TS5 4	53	137	123	23	99	102
Middlesbrough : Acklam	TS5 5	256	97	105	52	91	103
Middlesbrough : Acklam	TS5 6	185	140	115	27	93	81
Middlesbrough : Acklam	TS5 7	125	114	120	76	98	119
Middlesbrough : Acklam	TS5 8	97	102	93	58	109	106
Middlesbrough : Nunthorpe	TS7 0	32	92	104	4	104	135
Middlesbrough : Marton	TS7 8	49	100	114	63	100	98
Middlesbrough : Coulby Newham	TS8 0	23	121	112	0	105	118
Middlesbrough : Hemlington	TS8 9	20	95	113	33	95	127
Saltburn	TS12 1	15	49	91	138	110	121
Brotton	TS12 2	16	31	81	260	116	72
Lingdale	TS12 3	5	54	48	808	101	98
Loftus	TS13 4	6	42	68	411	115	72
Guisborough	TS14 6	10	51	69	145	96	98
Guisborough	TS14 7	12	43	80	90	109	77
Guisborough	TS14 8	12	67	108	29	101	108

Table 4: Concentration of name types by postcode sector (Index values, study area average = 100)

By contrast, Cornish migrants, who came to work the iron ore mines rather than to labour in the shipyards and steel works, originally settled in the eastern part of the study area, where the majority of their descendants can still be found in Lingdale, Brotton and Loftus. Very few of the descendants of the Cornish migrants have moved from the coastal mining communities to Middlesbrough (Figure 6). Likewise very few of the descendants of Irish or Scottish immigrants have moved from Middlesbrough into the surrounding towns and villages. Differences in the mix of housing types between the west and east of the study area illustrate why it is that Cornish names should be so uncommon in areas of Welfare Borderline and Municipal Dependency (which mostly occur in Middlesbrough) and more common in Ties of Community (which characterise many of the ex mining settlements).

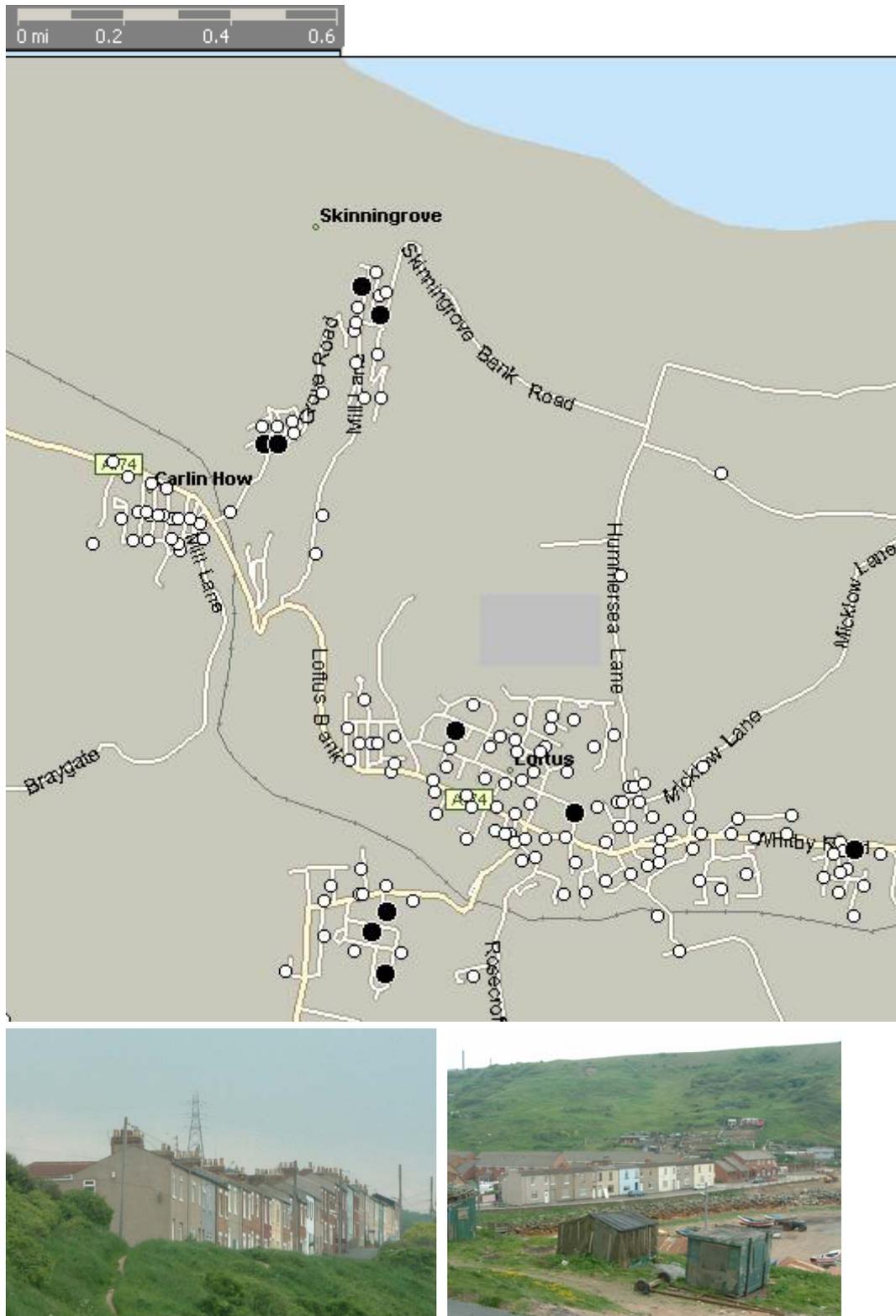


Figure 6: (A) Current locations of postcodes (black dots: other postcodes are white dots) in East Cleveland with rare Cornish names, including (B) Skinningrove

The differences in the geodemographic profiles within the study area are reflected in the distribution of English name types. Names common in the North East are much more prevalent in the eastern part of the study area, such as Saltburn, Brotton and Loftus, than they are in Middlesbrough, reflecting the pre-existence of significant populations in these areas prior to 1851. By contrast English names common in other parts of the country but not in the North East are more common in the southern suburbs of Middlesbrough, such as Hemlington and Nunthorpe, and in the retirement coastal town of Saltburn, places in which professional classes moving from the South would feel more comfortable. They are less common in the rural and mining communities and in the older inner areas of Middlesbrough.

This case study evidence illustrates how family name analysis can be used to create distance-based analyses of migrant flows, for example by relating the density of particular types of name at postcode area level with the distance of these postcode areas' centroids from the family names' heartland. The decay functions that result from these analyses can be set against expected norms, such as Ravenstein's "law" of stepped migration, or anticipated rates of filtering through neighbourhoods of different status. In both instances, the size, detail and coverage of our dataset make it possible to conduct analyses that are both more detailed and more sophisticated than hitherto. We are able to ascribe additional individual characteristics to short, intermediate and long distance migrants, and to view migration flows as a response to the unique economic conditions of regional and local economies as well as responses to prevailing national and

international economic conditions. The results of our case study support Pooley and Turnbull's (1998) observation that the distances moved to smaller places may well have been as great as those associated with the highest order settlements of the urban hierarchy. In a similar vein, the use of surname analysis alongside geodemographic profiling makes it possible to develop critical perspectives upon social mobility that extend our understanding of filtering and social mobility (see Saunders 2001)

6. Data mining and hypothesis generation with respect to immigration and residential differentiation

The preceding section has identified some of the core ways in which our data resource may be deployed in support of prevailing theories of migration and residential structure. By extension, these data and methods also offer many other opportunities for data mining and hypothesis generation. Our primary objective in this paper has been to illustrate the research benefits of the linkage of hitherto unrelated national datasets into a robust quantitative resource which can be used to improve and systematise our understanding of migrant flows in Great Britain and Ireland at a range of spatial scales, and to understand some of the outcomes of migration in terms of social and neighbourhood segregation. In creating this resource, it is not our intention to renew the somewhat discredited quest for a meta theory that comprehensively accounts for either migration flows or residential structure, but rather to systematise our understanding of the magnitude of population flows at different scales, and in different time periods. This is perhaps in the softer sense of the use of "laws" in geography, discussed in this journal (e.g. Sui 2004). Viewed from this perspective, there are a number of respects in which

this resource may be used, first, to reappraise the domains of existing laws and hypotheses in social investigation and, second, to generate new hypotheses.

With respect to the former, relevant aspects of Ravenstein's (1885) laws of migration include the effects of distance upon migration between settlements of different sizes, and the relative contributions of wave-like migration and more localised circulations of population in accounting for net migration rates in regions and localities. Relevant aspects of theory pertaining to residential location and mobility centre upon the extent and feasibility of "filtering" through the housing market that takes place over the long term following the decision to migrate. These concerns may be addressed using our data resource with generality and inclusiveness, as demonstrated in our case study.

Our data resource may also be mined in order to suggest or substantiate more specific hypotheses with regard to migration, the functioning of labour markets and the inter-generational extent of residential mobility.

Elsewhere, one of us (Webber 2005) has begun to illustrate the range of hypotheses that rudimentary analysis of our dataset has thrown up, both with respect to Middlesbrough and more generally. A strength of this approach is that it not only provides rich individual level temporal profiles for case study settlements, such as Middlesbrough, but also enables these to be viewed against a backcloth of all of Great Britain and Ireland. This enables us to: anticipate the geographic *origins* of the various migrant groups that migrated to Middlesbrough, both with regard to specific areas and to general distance

bands; hypothesise about the *timing* of various migration flows and the (direct and indirect) *routes* taken by migrants to the town; describe the detailed *social and economic geography* of the area in two time slices separated by nearly 120 years; over the same period, suggest ways in which Middlesbrough's *labour market* has functioned relative to the regional and national economy; and advance hypotheses about likely levels of *residential filtering* and *social mobility* in the town.

Aspects of these topics are discussed in detail in Webber (2005). Here we summarise them, principally by way of establishing the merits of the GIS-based resource that we have created. First, the methodology for identifying geographic concentrations of individual family names has been used to suggest the likely magnitude of migration from different regions and counties of Great Britain and Ireland. Using the indexing method described above, we have been able to establish whether names that are disproportionately concentrated in Middlesbrough are also disproportionately concentrated in the various counties of Ireland and G.B. postal areas (Level 1). Our data mining was guided by previous historical research on the industrial origins of Middlesbrough migrants (Renton 2005), work on the demise of the Cornish tin mining industry (HES 2005) and more general research on the nature and extent of emigration from Ireland (Baines 1985). This guided us to identify the U.K. postcode areas and Irish counties from which a disproportionate number of migrants were probably sourced, and to suggest distinct channels along which local and regional communities are likely to have migrated. By extension, this has allowed us to separate the importance of such channelled movement from all other movements, such as the more general circulatory

flows and more idiosyncratic motivations for migration described by Pooley and Turnbull (1998, 34-35).

In the case of Middlesbrough, we have gone as far as to identify the geographic concentrations of individual, county specific, Irish names relative to Irish names in general. This has enabled us to hypothesise about the relative importance of channelled versus more general (diffuse and probably indirect) migration to the town. Analysis of the geographic concentrations of specific Irish names in each English postal area also makes it possible to hypothesise about the shared characteristics of Irish immigrant destinations, as well as the specifics of flows between particular origin-destination pairs. More speculatively, analysis of such concentrations may also be used to suggest chronologies of inter-regional migratory patterns. Middlesbrough, for instance, shares common Irish surname characteristics (and hence probable ancestry) with Clydeside, Tyneside and South Wales, but not with other Irish surname concentrations in London: this might emphasise the importance of shared historic industrial structure, and a possible role for Clydeside as a staging post for some migrants destined for Middlesbrough. In addition to providing perspective upon the relative importance of wave-like migration versus more general circulatory movements, discussed above, this line of analysis also provides a means of examining the degree of common genetic inheritance of different parts of the country. This is particularly relevant to the understanding of contemporary Asian and African immigration into Britain.

The dataset may also be used to uncover industrial and labour market linkages, as through data mining of Cornish names. Using the historic names profile of what is now the Truro (Level 1) postal area, it is possible to detect local concentrations of Cornish names, such as that which occurs in 1881 to the south of Middlesbrough. This kind of analysis makes it possible to quantify the magnitudes of past industrial change and linkage (e.g. the collapse of the Cornish tin mining industry and development of Middlesbrough's iron industry), and to put these into national context.

Other analysis has been couched explicitly in the temporal domain. The 1881 Census was carried out at a time following the Irish Famine of 1845-8, which although not a time of crisis emigration nevertheless saw continued scattering of the Irish population. Cousens (1960) has observed that much of Irish emigration during and in the immediate aftermath of the Famine did not come from the most seriously affected areas in the west and south-west. Baines (1985, 31) suggests a complex explanation for this, embracing labour market, land tenure and cultural factors. Whatever the causes, the observed net effect is that in 1998 there were 1,179,565 people with identifiably Irish family names in our Republic of Ireland file and 1,414,885 in Great Britain. Our database makes it possible to extend this research by considering the ratios of occurrences of individual Irish names the 1998 Irish and G.B. files. From this data led perspective, low ratios suggest late migrants and vice versa. Viewed in the context of international research, this also provides a means of ascertaining the extent to which local and regional migration from Ireland was to parts of the world other than Great Britain.

Our remaining hypotheses consider the outcomes of migration in terms of contemporary residential structure. Here, urban theory ever since the work of Burgess and the Chicago School has been predicated upon the “filtering” of migrants through the built environment of the growing city, while later work has advanced much less optimistic hypotheses about the structural inequalities that characterise capitalist societies. Our surname database establishes a genetic linkage of present day populations to historic migrant groups: in our case study we have differentiated between those who have remained local to Middlesbrough ever since the use of family names became commonplace in the twelfth century, those who originate from other areas of North East England, those that originate in most of the rest of England, and those who originate from the most distant parts of the British Isles (specifically the Scottish, the Cornish and the Irish).

The matching of the 1998 Electoral Roll with the 2003 Mosaic geodemographic system makes it possible for us to measure how well these different migrant groups have fared. Our empirical analysis suggests that Irish and Scottish names are generally associated with Middlesbrough neighbourhoods that are low status in terms of the U.K. classification. That is to say, they are names associated with low status economic migrants, who appear to have remained in such neighbourhoods because of the restrictive social mobility of Great Britain. Similar outcomes characterise the socio-spatial differentiation of Cornish immigrants, and mining of our database reveals an interesting trend: the more Cornish the name, the more concentrated the spatial distribution of its representatives within the study area; and, conversely, Middlesbrough

representatives of names that are Cornish to a lesser extent (i.e. they are disproportionately represented in Cornwall but not extremely so) are less likely to be settled in the hard core Cornish areas. This may be indicative of differentiation within a migrant group – it may be that those migrants drawn from a wider area within Cornwall were less likely to be skilled in mining.

Leaving these mass migrants aside, neighbourhood analysis suggests an inverse relationship between social success and distance of migration – those with traditional Middlesbrough names have fared least well, those from the rest of the North East have attained or retained intermediate status and those that have moved furthest have attained or retained the greatest status. This is of relevance in thinking about the functioning of the labour market: the high residential status of those with “outside North East” English names suggests that they have moved to the area in response to the needs of the national labour market, while those with more local names are more likely to be participants in the local, less skilled, labour market. Indeed it may well be the case that the more enterprising people with local names are disproportionately likely to have migrated out from the region.

Each of these hypotheses has been suggested through selective mining of our database, and the hypotheses so generated are advanced here in a spirit of humility rather than conviction. Yet we do believe that they are suggestive of the enormous potential of GIS-based linkage of data pertaining to the past and present, and bear testimony to the value of the particular resource that we have created.

We are aware, however, of a number of caveats to the use of our family names database. First, mining of the database itself suggests limitations to the nomenclature of family names. For example, most names with the prefix “O”, as for example in O’Sullivan, are far more common in Ireland than in Britain, many Irish emigrants having dropped the “O”. For this reason there is no support for the proposition that names which are relatively more common in Ireland than in Britain are indicative of more recent emigrants. Second, the relinquishing of maiden names upon marriage means that our database implies that females and males were equally mobile, despite evidence that, for example, female domestic service workers were more mobile than males, at least at local and regional scales. Pooley and Turnbull’s (1998) extensive analysis of migration and mobility in Britain since the 18th Century concludes that gender accounts for rather little of the variability in the structure and process of migration, although the motivation of women to move was often quite different to that of men.

8. Conclusions

The conclusions of this study can be considered both in terms of substantive findings and in terms of analysis methods.

Key substantive findings of our illustrative case study can be summarised as follows. The ethnicity question in the 2001 U.K. Census of Population significantly underestimates the proportion of the population of Middlesbrough that are of Scottish or Irish descent. The genetic make up of the city and its surroundings are much more mixed

than would appear from conventional population statistics. This may help explain the broader antagonism to regional devolution in the North East. Whilst the region has fewer people of recent Irish descent than other regions of Great Britain, it has a larger proportion than other parts of Britain who are descended from early Irish immigrants. In terms of the social status of the neighbourhoods in which they live, the descendants of both Irish and the Scottish communities have achieved a fair level of integration, though neither live in as prestigious neighbourhoods as the descendants of economic migrants from other parts of England. The descendants of Scottish migrants have generally prospered to a slightly greater degree than the descendants of the Irish. Although both communities have expanded geographically from the inner city heartlands, neither community has migrated outside the city to its surrounding commuter settlements, retirement resorts of rural villages. By contrast, the Cornish migrants who came to work in the iron ore mines rather than in the steel works or chemical plants, settled in very distinct communities outside the city of Middlesbrough. Their descendants are still highly localised in these communities and very few of them have moved to live in Middlesbrough.

From the broader analysis described by Webber (2005), it is evident that the Irish migrants to Middlesbrough came from different regions of Ireland to those who migrated to London and that the Middlesbrough Irish are likely to have greater genetic affinity with the Irish in Scotland and South Wales than with those who migrated to London. As with the Cornish, residents descended from Irish and Scottish immigrant communities who migrated en masse in earlier waves of emigration appear to have been significantly

less successful than residents descended from more recent waves of migration. It is descendants from these early immigrant waves that are particularly concentrated in the more deprived overspill council (public housing) estates, characterised by low levels of education, income and car ownership and high levels of single parents, people without a job and young offenders. Among those with English names, people from families who lived in the Middlesbrough area prior to its rapid nineteenth century industrialisation appear to have been less successful than those from families who have traditionally lived elsewhere in the North East. Residents whose families originally came from other regions of the country have proved to be the most successful in terms of the neighbourhoods they live in.

Middlesbrough, its wealthier southern dormitory suburbs and the Cleveland coast have very different genetic mixtures, with the Cleveland coast containing a much higher proportion of indigenous North East families, inner Middlesbrough much higher proportions of residents of Celtic origin and many fewer originating from the South of England, with the dormitory suburbs having by far the largest proportion of people from other regions of England.

In terms of analysis methods, we believe that this research has created a quantitative data resource that has the following key implications for the development and study of “historical GIS” (Gregory and Southall 1998). First, analysis of names provides a rich, disaggregate and generalised picture of past and present demographic conditions in Great Britain: this picture is further enriched if the classification of commonly held

names extends beyond those of Anglo Saxon origin to include Celtic and Asian names. Second, the concept of the surname “heartland” is helpful in defining both Scottish and Cornish names, and we have found postcode areas centred upon historically important transport intersections to provide a useful organising framework (Raper, Rhind, and Shepherd, 1992). Third, measurement of the relative frequency of names in different countries, as for instance in Great Britain and the Republic of Ireland, is a useful means of defining the culture of origin of different names.

Our fourth methodological conclusion is that measures of the degree of concentration of names within their “heartlands” provide a useful tool for differentiating between names which are truly regional in origin from names which reflect naming conventions across a wider geographical area and which, therefore, are less effective proxies for migrants from a very specific locality. Related to this, identification of the region (Level 1 postcode area) with the highest concentration of names from another country, such as Ireland, can be helpful in tracing the spatio-temporal detail of historic migration flows. However, the measurement of the relative frequency of a name in two countries as an indication of the period of migration of that name seems is not effective in this case because of the manner in which certain types of surname have been altered in the host country.

In addition to the creation and dissemination of this database, extensions to this research (already underway) open up the possibility of using this approach to demonstrate the still broader contribution that GIS and quantitative analysis can make to the understanding of

social processes, particularly in relation to migration and to the delineation of regional identity, within Great Britain, the US, Australia and New Zealand. In our future work, we hope to demonstrate how the analysis of the geographic distribution of names associated with particular regions or sub-regions of origin is helpful in identifying and explaining important migration paths between population importing and population exporting localities; in demonstrating the inertial effect of distance on the migration of populations; in quantifying the extent to which local areas are likely to have higher or lower levels of genetic kinship; and in identifying natural boundaries between areas of differing cultural traditions. As such, this work creates not only a broad yet intricate canvas for historical GIS but also quantitative insight into contemporary social structure and mobility. Here, we have illustrated our ideas with reference to a case study of Middlesbrough, but in principle researchers might use the resource that we have created¹² as the evidence base for regional or locality studies of any part of the United Kingdom, the Irish Republic, the US, Australia or New Zealand.

9. Acknowledgements

This research was funded under Economic and Social Research Council (ESRC) Grant RES-000-22-0400. The authors are very grateful to Kevin Schürer of the ESRC Data Archive and to Experian for the supply of historical and contemporary data, to Alex Singleton for help with the artwork, and to the anonymous referees who helpfully and constructively commented on the first version of this manuscript.

10. References

Ashby, D. I. and P. A. Longley. 2005. Geocomputation, geodemographics and resource allocation for local policing. *Transactions in GIS* 9: 53-72

Baines, D. 1985. *Migration in a mature economy: emigration and internal migration in England and Wales, 1861-1900*. Cambridge, UK: Cambridge University Press.

Brown, P. J. B., P. G. McCulloch, E. M. I. Williams, and D. C. Ashurst. 2001. Control sample design using a geodemographic discriminator: an application of Super Profiles. *Journal of Geographical Systems*, 3: 107-35

Champion A. G., A. E. Green, D. W. Owen, D. J. Ellin, and M. G. Coombes. 1987. *Changing Places: Britain's Demographic, Economic and Social Complexion*. London: Edward Arnold.

Cousens, S. M. 1960. The regional pattern of emigration during the Great Irish Famine. *Transactions of the Institute of British Geographers* 28, 119-34

Dodgson, R., and R. A. Butlin (eds) 1978. *An Historical Geography of England and Wales*. London: Academic Press.

Feldman, D. 2001. Migration. In *The Cambridge Urban History of Britain*, ed. M. Daunton, Chapter 6, 220-260. Cambridge, U.K.: Cambridge University Press.

Gregory, I. and H. Southall 1998 Putting the past in its place: the Great Britain Historical GIS. In *Innovations in GIS 5*, ed. S. Carver, 210-21. London: Taylor and Francis.

Hanks, P. (ed.). 2003. *Dictionary of American Family Names*. New York: Oxford University Press.

Harris R., P. Sleight, and R. Webber. 2005 *Geodemographics, GIS and Neighbourhood Targeting*. Wiley: Chichester.

HES. 2005. *Cornish Mining*. <http://www.cornish-mining.org.uk/comcult/migrat.htm>, accessed 8 November 2005

Johnston, R. J. 2000. Urban studies. In *The Dictionary of Human Geography (fourth edition)*, eds. R. J. Johnston, D. Gregory, G. Pratt, and M. Watts, 870-1. Oxford: Blackwell.

Long, J. 2005. Rural-Urban Migration and Socioeconomic Mobility in Victorian Britain. *The Journal of Economic History*, 65(1): 1-20.

Longley, P. A. and R. J. Harris 1999. Towards a new digital data infrastructure for urban analysis and modelling. *Environment and Planning B* 26: 855-78.

Longley, P. A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind. 2005. *Geographic Information Systems and Science (second edition)*. Chichester: Wiley.

Pooley, C. and J. Turnbull 1998. *Migration and mobility in Britain since the 18th century*. London: UCL Press.

Raper, J. F., D. W. Rhind, and J. R. Shepherd. 1992. *Postcodes: the New Geography*. Harlow: Longman.

Ravenstein, E. G. 1885. The laws of migration. *Journal of the Royal Statistical Society* 48: 167-227.

Reaney, P. H. and R. M. Wilson (eds). 1997. *Dictionary of English Surnames*. Oxford: Oxford University Press.

Renton, D. K. 2005. [Migration to the North East: an unfinished story](http://www.dkrenton.co.uk/research/nehostility.htm).
www.dkrenton.co.uk/research/nehostility.htm (accessed 8 November 2005)

Saunders, P. 2001. Urban ecology. In *Handbook of Urban Studies*, ed. R. Paddison, 31-51. London: Sage.

Schürer, K. 2004. Surnames and the search for regions. *Local Population Studies* 72: 50-76.

Sleight, P. 2004. *Targeting Customers: How to use Geodemographic and Lifestyle Data in Your Business (3rd Edition)*. Henley-on-Thames: WARC.

Sui, D. Z. 2004. Tobler's First Law of Geography: A big idea for a small world? *Annals of the Association of American Geographers* 94: 269-77.

Thrall, G. I. 2002. *Business Geography and New Real Estate Market Analysis*. New York: Oxford University Press.

Webber, R. 2004a. Designing geodemographic classifications to meet contemporary business needs. *Journal of Interactive Marketing* 5(3), January/March.

— 2004b. The relative power of geodemographics vis a vis person and household level demographic variables as discriminators of consumer behaviour. *CASA Working Paper* 84, http://www.casa.ucl.ac.uk/publications/full_list.htm, accessed 23 June 2005.

— 2005. Neighbourhood Segregation and Social Mobility among the descendants of Middlesbrough's 19th century Celtic Immigrants. *CASA Working Paper* 90, http://www.casa.ucl.ac.uk/publications/full_list.htm, accessed 23 June 2005.

— 2006. The 2005. British General Election campaign: how the parties applied segmentation to their Direct Marketing strategies. *Journal of Direct, Data and Digital Marketing Practice*, Volume 7(3), January 2006, in press.

Webber, R. and P. A. Longley. 2003. Geodemographic analysis of similarity and proximity: their roles in the understanding of the geography of need. In *Advanced Spatial Analysis*, eds P. A. Longley and M. Batty. Redlands CA, ESRI Press, 233-66.

Zelinsky, W. 1993. Classics in human geography revisited: author's response. *Progress in Human Geography* 17: 217-9.

Correspondence: Department of Geography, University College London, 21 Bedford Way, London WC1H 0AP, United Kingdom. plongley@geog.ucl.ac.uk

List of Figures

Figure 1: Frequency of persons with names ending in –thorp in 1881, postal area index values (Great Britain average = 100)

Figure 2: The United Kingdom postal geography: Postcode Area (Level 1), with its constituent Postcode Districts (Level 2), Postcode Sectors (Level 3) and Unit Postcodes (Level 4)

Figure 3: The study area: Middlesbrough, and Middlesbrough South and Cleveland East

Figure 4: Illustration of postcodes by type of neighbourhood in the Guisborough part of the study area

Figure 5: Example of different types of neighbourhood in the study area : clockwise (from top left), group B (Happy Families) in Guisborough, group J (Grey Perspectives)

in Saltburn by the Sea, group D (Ties of Community) in Middlesbrough and group G (Municipal Dependency) in Middlesbrough

Figure 6: (A) Current locations of postcodes (black dots: other postcodes are white dots) in East Cleveland with rare Cornish names, including (B) Skinningrove

List of Tables

Table 1: Study area electors by type of name

Table 2: Population by ethnic group in the study area in 2001 (source: 2001 Census of Population)

Table 3: Relationship between name type and neighbourhood type, study area, 1998, index values (study area average = 100)

Table 4: Concentration of name types by postcode sector (Index values, study area average = 100)

Notes to Table 4 (see Harris, Sleight, and Webber 2005 for full details):

¹ very high levels of household income, typically areas of pleasant, detached houses in the better areas of town, with an older population of professionals and senior managers, mostly with children of university age or beyond.

² modern, owner occupied housing, typically built in the last thirty years by large developers, at fairly high residential densities. Households in these sorts of area are typically climbing career ladders in large organisations, are first time or second time mortgagees and have one or two children, typically of school or pre school age.

³ areas of inter war, semi detached houses, where older people have now paid off a large proportion of their mortgages. These are areas of higher than average incomes, but contain many middle rather than senior managers.

⁴ areas of older terraced housing, much of which was originally built to house a late Victorian manufacturing labour force. Located in older, inner city areas, close to sources of industrial employment, many of these neighbourhoods have been taken over by Asians. However the type also occurs in older mining communities in East Cleveland.

⁵ areas containing a mix of well educated, single and transient populations, typically including students and young professionals, many of whom live in small but smart rented flats. These areas also contain large old houses, many of which are suitable for subdivision, suitable for first generation migrants.

⁶ inner city areas of social housing, many of which result from the clearance of older terraces and which often take the form of mid rise or high rise flats. This group has a high proportion of households with only a single person resident.

⁷ large estates of low rise social housing, mostly on the periphery of large cities, which have been abandoned by the more enterprising skilled manual workers and where, as a result, few residents have exercised the right to buy. These areas typically have large proportions of single parent families, people who are sick or unemployed, who do not have access to a car and who are troubled by vandalism and delinquency.

⁸ found typically among the better and smaller council estates where former tenants have espoused more middle class lifestyles and where people are sufficiently enamoured by the quality of their estate to have exercised their right to buy homes previously rented from the council.

⁹ high proportions of old people reliant on state benefits. They mostly live in council accommodation.

¹⁰ high proportions of old people but, by contrast with Low Income Elders, these tend to be better off, enjoying reasonable health and life expectancy as well as active leisure pursuits funded by occupational pension schemes.

¹¹ genuinely rural in character, not just commuting countryside. Though only a minority of the workforce is now engaged in agriculture, these postcodes are the last bastions of Britain's rural way of life and places where neighbours are most likely to look out for each other.

¹ Specifically in the USA, Canada, Ireland, Australia and New Zealand

² Current U.K. postcode (zipcode) areas were delineated by the Royal Mail in the 1970s and, with few exceptions, are centred upon major transport route intersections. As such they present something akin to a tessellation of functional regions (see Champion et al. 1987: 9). Our choice of study area was motivated by the desire to investigate the effects of the industrial development of Middlesbrough and East Cleveland, and for reasons also of data acquisition excludes the adjacent constituency of Redcar.

³ No information is available for electors in Northern Ireland. The threshold of 100 occurrences is an arbitrary one adopted, in practice, for the entire Great Britain population in 1881 and for all adults eligible to vote in 1998.

⁴ Thanks are due to Kevin Schürer, Director of the ESRC Archive at the University of Essex.

⁵ More specifically the study area consists of a set of postcode sectors which are entirely contained within or fall predominantly within the two constituencies.

⁶ Elector registration on the Great Britain Electoral Roll is a legal requirement, although electors are allowed to opt out of their names appearing in the version of the register that is available at the individual level in the public domain.

⁷ A full description of the U.K. Mosaic types can be found on the Website of Experian Business Strategies <http://www.business-strategies.co.uk/Content.asp?ArticleID=566>

⁸ An important factor in the growth of Middlesbrough in the 1850s was the development of the iron ore industry. This coincided with the decline of the tin and copper mines of West Cornwall with the result that many redundant Cornish miners moved to Teesside during this period: hence the high concentration of Cornish names today, notwithstanding the distance of Middlesbrough from Cornwall.

⁹ Names originating in Scotland are identified partly on the basis of linguistic structure, i.e. names starting with Mac- or ending in -ie. Confusions can be resolved by calculating the proportion of occurrences of the name which are currently resident in Scotland. A file containing the frequencies of different names in the Republic of Ireland enables us to compare the relative frequencies of names in Britain and the Republic of Ireland, which also assists in the allocation of names of uncertain provenance. Asian names can also be identified in part from their linguistic structure, and in part from studying their geographical distribution. Another valuable source of information for identifying names with origins outside England is the use of counts of association between the name and the first names contained against that name on the electoral roll. The reliability of the identification process is indicated by the strength of the correlation (0.957) between the degree of under or over representation of Asian names defined in this way across each of the 61 Mosaic geodemographic clusters and the degree of under or over representation of persons born in South Asia.

¹⁰ A number of Cornish names can be defined using linguistic analysis. Examples are names with the prefix “Tre”, “Bos” and “Pol”. Tremayne and Tregonning are examples. However there are many other Cornish names, such as Laity, which are toponymic but which have no apparent Celtic linguistic root. See http://www.casa.ucl.ac.uk/family_names/papers/20_cornish_migration.pdf

¹¹ The indices created in this paper are described in Webber (2005) and are quite different from location quotients. In general our dataset provides a more than adequate number of cases for the kinds of analysis that we have undertaken

¹² Available from the U.K. Data Archive at www.essex.ac.uk