1
Introducing Geodemographics

Learning Objectives

In this chapter we will:

- Define geodemographics as the analysis of people by where they live.
- Explore why it is a useful framework for public- and private-sector decision making.
- Offer initial explanations and worked examples of how geodemographics 'works'.
- Introduce Tobler's 'first law of geography' and the concept of spatial autocorrelation.
- Present two of the specially commissioned case studies that appear throughout this book. The first is authored by Martin Callingham, formerly the group manager of market research for Whitbread plc and is about using the geodemographic approach to model price sensitivity in the restaurant market. The second is by Keith Dugmore of Demographic Decisions Ltd and chair of the Demographics User Group, and is about using geodemographics in the public sector.

Geodemographics, GIS and Neighbourhood Targeting Richard Harris, Peter Sleight and Richard Webber
 ISBNs: 0-470-86413-3 (HB); 0-470-86414-1 (PB)

Introduction

Geodemographics is the 'analysis of people by where they live' (Sleight, 1997, p. 16). It is the suggestion that *where* you are, says something about *who* you are; that knowing where someone lives provides useful information about how that person lives. To quote some product advertising, it is the possibility that 'we know who you are, because we know where you live.' Figure 1.1 illustrates this link between people and places. It is a simple idea – one that has shown itself to be of commercial value and the catalyst of a rapidly growing and globalizing industry.

The purpose of this book is to consider the relevance, strengths and limitations of the geodemographic idea for public- and private-sector decision making. We provide an introduction to and overview of the methods, theory and classification techniques that provide the foundations of neighbourhood analysis and commercial geodemographic products. We give examples of using geodemographic analysis effectively to target resources and offer guidance to best practice that draws upon our contributors' experiences of working within the geodemographic industry.

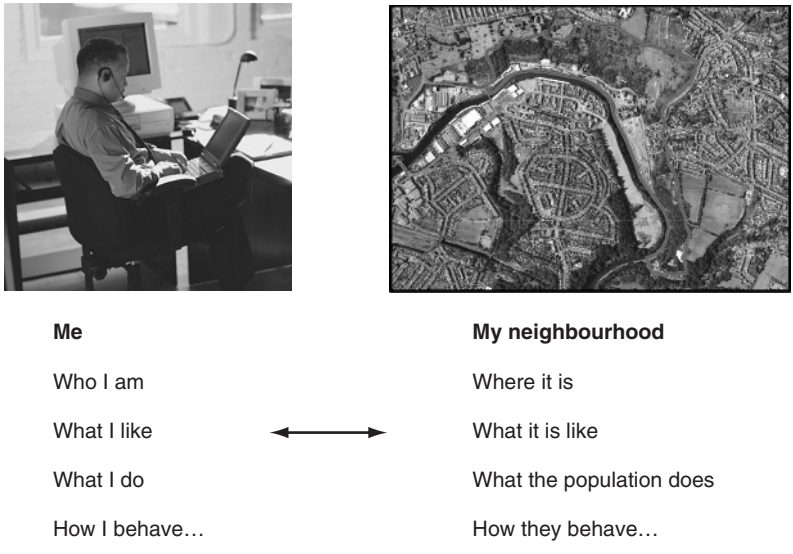

Figure 1.1 Geodemographics is 'the analysis of people by where they live', linking people to places

Within the book, particular focus is given to linking geodemographics with the theories and products of geographical information science, notably geographic information systems (GIS). Our aim throughout is to provide reader-friendly theory and moderate statistical explanation, supported by relevant case studies, short vignettes and applied 'how to' sections that will appeal to an international and professional audience at work in business and service planning, and to students of marketing, geography or other spatial, social science. Along the way we include some formulae and mathematical notation as often these provide the most succinct and accurate way of describing a particular calculation or procedure. However, we know such notation will not suit everybody so we also take time to explain, in words, what the symbols summarize. In this chapter we introduce the core principles and ideas that underpin the use of geodemographic information for neighbourhood profiling and set out the agenda for the rest of the book.

1.1 The use of geodemographics

We are neither the first to take an interest in geodemographics; nor, we hope, will we to be the last! Over a decade ago Brown (1991, p. 221) commented that,

> [g]eodemographics has come into use as a shorthand label for both the development and the application of area typologies [neighbourhood classifications] that have proved to be powerful discriminators of consumer behaviours and aids to 'market analysis'.

The 'proof' is found in the increased value of the geodemographic market. In Britain this was estimated at a value of £25 million in 1992 (Sleight, 1997, p. 15, citing Mitchell, 1992). By 1995 the same market was valued at £54 million. In 1998, *Directions Magazine* (www.directionsmag.com) reported its 'conservative estimate' of 20,000 companies in the USA and Canada using commercial neighbourhood classifications as part of their marketing information. Weiss (2000) reports that US marketers spend an estimated $300 million annually on clustering techniques (see Chapter 7), profiling the behaviour of the nation's 100 million households: 'cluster-based marketing has gone mainstream and is now used by corporate, nonprofit, and political groups alike to target their audiences' (p. 4). As we shall show in Chapters 3, 7 and 9, the market has continued to evolve, the most recent stimulants being the release of twenty-first-century census data and the emergence

of extensive 'data warehouses' associated with a growing trade in consumer-oriented data.

It is also over 10 years ago that Leventhal (1993, p. 223) recognized the potential of geodemographics to inform strategic marketing, planning and communications, presenting examples of its application to the Market Research Society under three main headings:

- **Survey design** ('samples may be stratified or selected using geodemographics, and many large-scale surveys take advantage of this facility').
- **Retail planning** ('knowledge of the types of people living in a catchment area can be a key ingredient in understanding store performance and the same information can help in deciding on store location').
- **Direct marketing** ('the selection of prospects [prospective customers] can be improved by using geodemographics, whether for direct mail, "door-to-door" distribution or sales calls').

Those three themes correspond to the market research, market analysis and direct marketing streams identified by Curry (1993, p. 200) and summarized by Figure 1.2. Curry also outlines a fourth role for geodemographics in advertising and media analysis (as do Sleight, 1997, pp. 117–21 and Webber, 1985). Each author recognizes the value of geodemographics as an important business tool that offers a type of analysis which is both understandable and operational within an applied, decision-making environment.

Figure 1.2 offers a marketing perspective on geodemographic applications. We shall also consider the role of neighbourhood analysis in public-sector policy and planning, avoiding the impression that geodemographics is a solely commercial affair, driven by the needs of market analysis and consumer profiling. Although these commercial elements certainly are important aspects of the field, Chapter 2 reveals that the origins and evolution of neighbourhood classification have an academic pedigree in urban geography, urban sociology and in urban planning. Batey and Brown (1995, p. 78) record that,

> [t]he pragmatic approach of the marketing analyst has much in common with that of the urban planner. In both cases an area classification system is required to provide up-to-date information that is actionable, and the test of a good system is whether it works in practice. It is perhaps not surprising, therefore, that classifications generated for use in public policy-making ultimately found their way into the private sector.

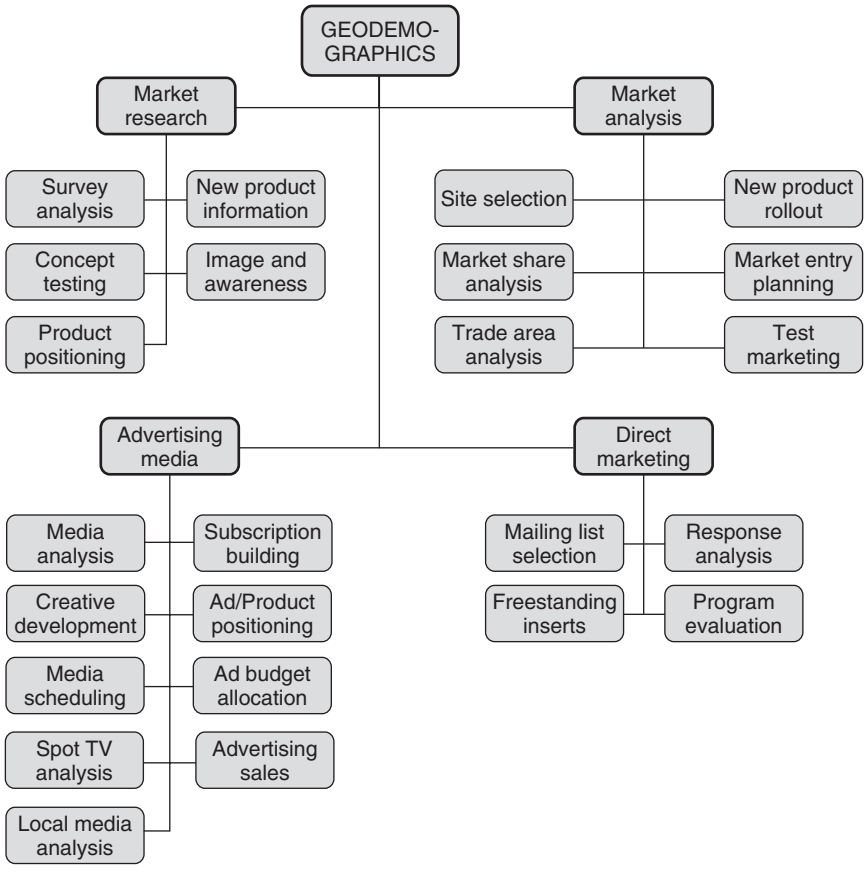

Figure 1.2 Some commercial applications of geodemographics. Source: adapted from Curry (1993, p. 200)

1.2 A simple application: opening a coffee shop in Atlantic City

A traditional, American Monopoly board (Figure 1.3) has 40 locations to land on, of which 22 are streets in Atlantic City. These streets are grouped together by eight distinct colour bands that indicate the cost of acquisition, development and rental value of the land (in the early twentieth century). Each street belongs to one and only one of the eight colour groups. Therefore, Mediterranean Avenue is either in the brown group or it isn't (it is); Boardwalk is either in the green group or it isn't (it is not, it is in the dark blue group). No street is only part in a group – each is either

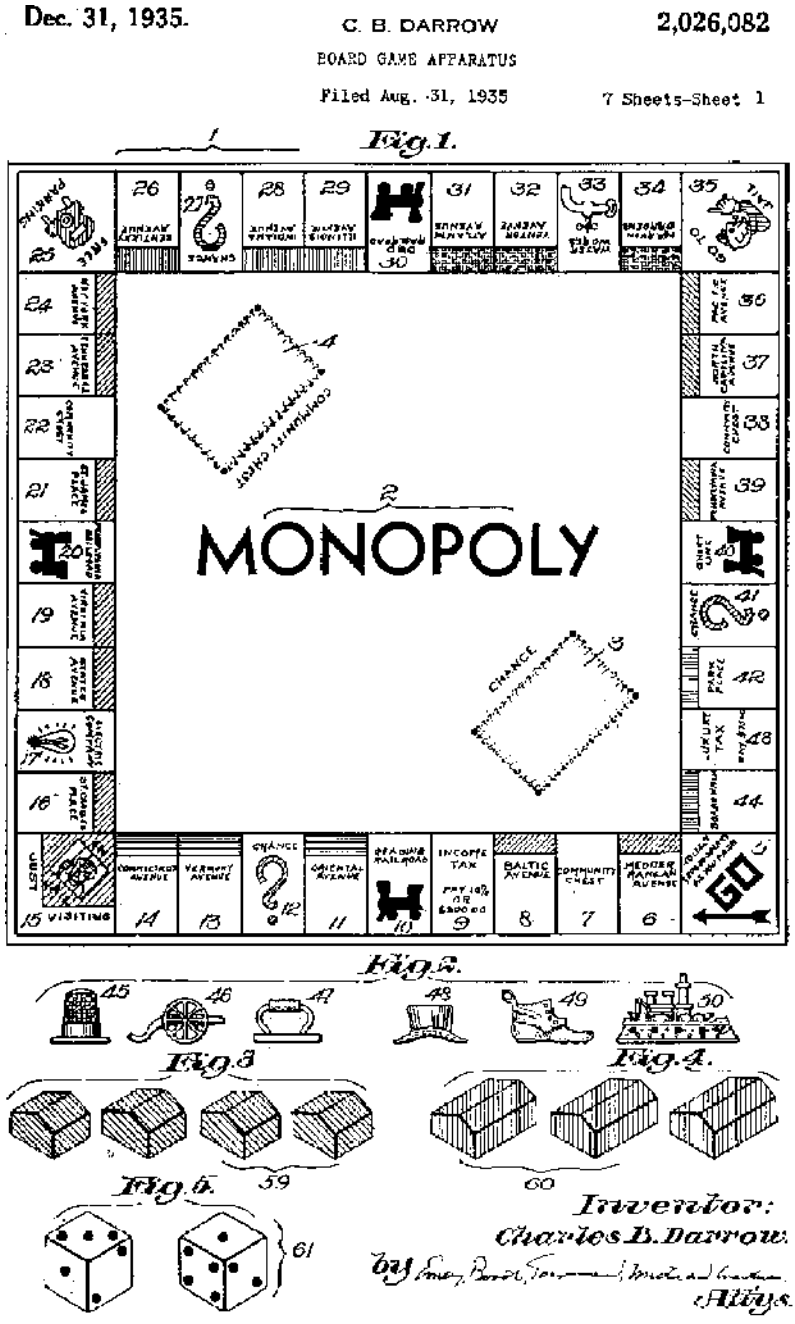

Figure 1.3 An early geodemographic classification. Source: US Patent 2,026,082 (United States Patent and Trademark Office)

entirely in else it is entirely out (this is known as Boolean logic where a proposition is either true or false). Furthermore, if a location does not belong to the one group then it must belong to another. If necessary, locations can be placed in a ninth, residual category of 'Other' (see Table 1.1).

Table 1.1 Segmentation of sample data by street and by neighbourhood

Group	Code	Description	Streets	Outlet here?	Customers per street	% Sample per group
Brown	A	Low rental value, construction cost and low maintenance	Mediterranean Ave.	No	10	1.34
			Baltic Ave.	No	0	
Light blue	B		Oriental Ave.	No	0	4.30
			Vermont Ave.	No	13	
			Connecticut Ave.	Yes	19	
			St. Charles Place	Yes	81	
Pink	C		States Ave.	Yes	94	23.52
			Virginia Ave.	No	0	
			St. James Place	Yes	136	
Orange	D		Tennessee Ave.	No	0	18.28
			New York Ave.	No	0	
			Kentucky Ave.	Yes	124	
Red	E		Indiana Ave.	Yes	153	37.23
			Illinois Ave.	No	0	
			Atlantic Ave.	Yes	58	
Yellow	F		Ventnor Ave.	No	0	9.41
			Marvin Gardens	No	12	
			Pacific Ave.	No	0	
Green	G		North Carolina Ave.	No	34	4.57
			Penn Ave.	No	0	
Dark blue	H	High rental value, construction cost and high maintenance	Park Place	No	9	1.21
			Boardwalk	No	0	
Other	Z	Other types of location, including utilities and railroad stations	Reading Railroad	No	0	
			Municipal Jail	No	0	
			Electric Company	No	1	
			Water Works	No	0	
			Penn Railroad	No	0	0.13
			Parking Lot	No	0	
			B & O Railroad	No	0	
			Shortline Railroad	No	0	
			Internal Revenue Service	No		
				Total	744	100

Travelling clockwise around the board from 'Go', the potential revenues from the properties increase but so too do the maintenance costs. In summary, the board is a simple, area-level classification of Atlantic City that uses economic value as the variable that differentiates between types of street and permits them to be grouped together on a like-with-like basis. The classification is both mutually exclusive and collectively exhaustive (Curry, 1993) – every location is classified as belonging entirely to one and only one neighbourhood type (albeit 'other'). A classification tree is shown in Figure 1.4.

Now we have our neighbourhood classification, what can we do with it? To answer the question, imagine you are the owner of *Caffeine-II-Goad*, a small coffee chain operating in Atlantic City. You have interviewed a sample of customers visiting your stores, finding out where they have travelled from to collect their coffee. That data is shown in Table 1.1, together with information about where your existing outlets are located. You find that the highest percentage of your sample is from streets of the red type (37.23%). Assuming the sample you have taken is representative of the customers you have not sampled – a critical assumption! – then, as proprietor, make the following decision: where will you open your next outlet?

Your answer:

Your reasons:

The site we have chosen is Illinois Avenue. The rationale is that this street is of the same neighbourhood type as both Kentucky and Indiana avenues. On this basis, we feel more able to attract customers to our brand of store if we open on Illinois than if we open on, say, Baltic. Of course, the neighbourhood classification cannot guarantee that we will

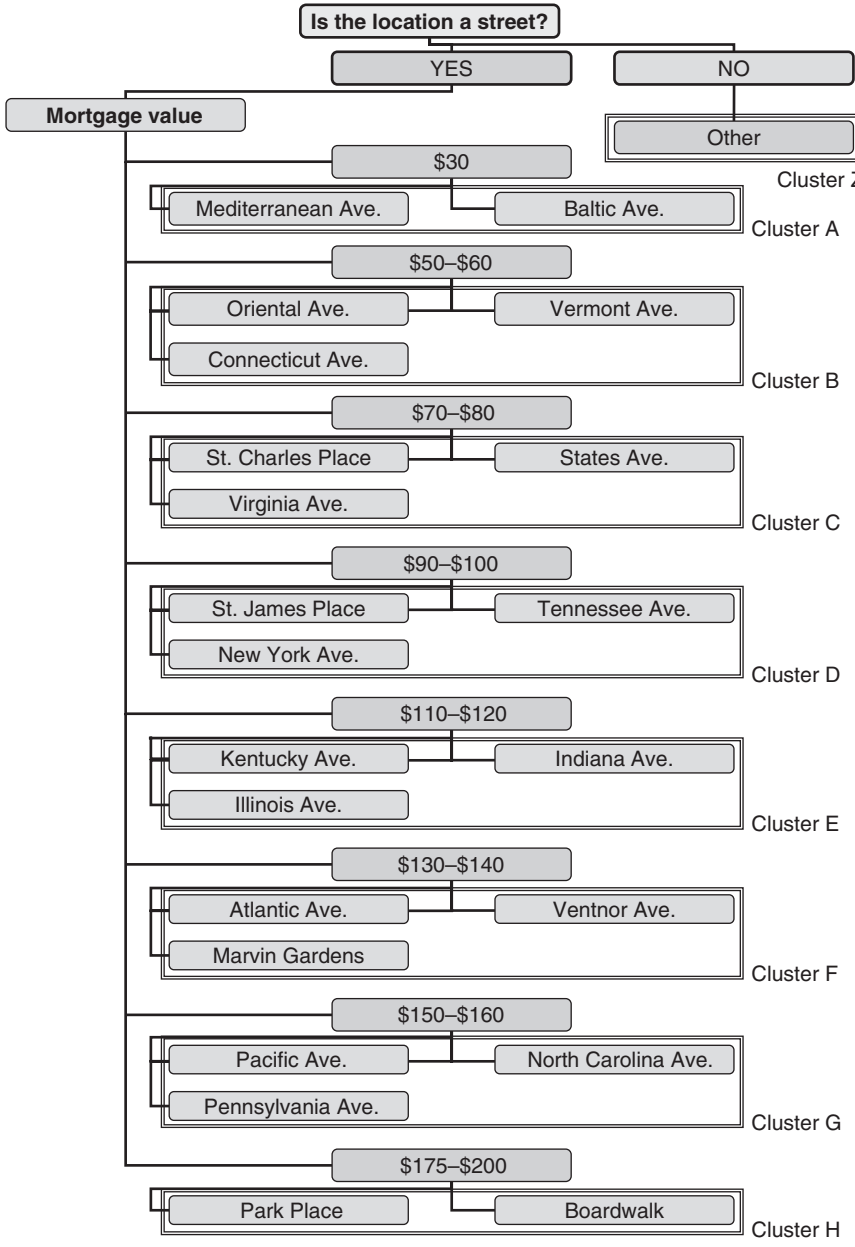

Figure 1.4 Classification tree for neighbourhood classification of Atlantic City

attract more customers but it has helped us to 'play the odds' and identify where potential customers might visit. Our strategy is not dissimilar to that of completing a colour set in Monopoly. However, it is not the only way of playing the game! In practice the decision is much more complex, requiring consideration of the number of people in each street (market size), the location of competitors' outlets (competition), distance from our existing outlets (catchments) and, not least, our business objectives (strategy). Nevertheless, the neighbourhood classification can help to inform the decision-making processes and help focus on a set of possibilities.

1.3 Another application: guiding neighbourhood regeneration funding

Imagine that instead of indicating the mortgage values of streets in Atlantic City, the monies shown in Figure 1.4 now represent the average weekly household incomes of people living in the different neighbourhoods (the values are not realistic but for the purposes of our discussion this does not matter). The question to answer now is, given that the city has received funding to promote regeneration of the most deprived neighbourhoods, where will you, as a city planner, target the funds?

Your answer:

Your reasons:

The streets we have chosen are Mediterranean and Baltic avenues. Our reasoning seems obvious – these streets appear to contain the lowest income households! What is less clear is why we chose only these two streets and did not include also, for example, Oriental, Vermont and

Connecticut Avenues. The answer is we chose the 10% most-deprived streets (the top decile, ranking the streets by average income from most to least deprived), making the assumption that deprivation correlates with average household income.

Again, in practice things are more complicated than our explanation reveals. First, the 10% threshold is an arbitrary cut-off and has the effect of segmenting the streets and consequently the resident households into two groups (or cluster types): those that are deemed deprived and those that are not; alternatively, those areas that will receive funding and those that will not. This approach could be regarded as 'heavy handed' – might those streets near the 10% threshold receive some of the funding, albeit a lower proportion?

Second, classifying 'deprived' streets is not the same thing as classifying deprived households – they exist at different geographic scales. A critical issue is how similar households are to the average for the street; how similar are they to each other? In areas of socio-economic diversity an average value (with no knowledge of the variance around that average) can give misleading results. It is possible to find pockets of deprivation in areas that otherwise appear affluent and these pockets risk being overlooked. Unfortunately we often only have the average or sum total values for an area (this is particularly true for national census data), with little knowledge of the within-area variations and population diversity (Harris and Longley, 2004).

Third, household income is not necessarily a robust indicator of deprivation. We ought to make corrections for the number of people living in the household and perhaps also allow for regional housing costs (e.g. the fact that London is more expensive to live in than Liverpool), the age of the population (retired persons would tend to have lower incomes, does this mean they are necessarily deprived?), the number of dependents (children are expensive to raise!) and so forth. Furthermore, although income may reasonably be regarded as an important factor in determining social inequalities (Hall and Pfeiffer, 2000), it is not the only measure of a person's standard of living or their level of integration within (or exclusion from) society. Deprivation indicators usually take a more rounded view, taking a selection of variables to create a multivariate profile or score of a neighbourhood's position in relation to others in the country (Lee, Murie and Gordon, 1995; ODPM, 2004). In the same way, commercial geodemographic classifications do not classify neighbourhoods based on a single variable such as mortgage value (in Figure 1.4) but take a range of data to better emphasize the differences between neighbourhood types.

1.4 Using geodemographics for retail targeting

Our example of using geodemographic classification in Atlantic City can be generalized to a retail company that takes its client list and sorts it into different types of consumer, making its judgement by where the client lives. The sorting first begins by linking the address of each client to a predetermined classification of the type of area that address is found in. As a consequence, the clients are segmented into groups not actually on the basis of their own, individual characteristics but according to some sort of social average for the area in which they live – by the type of area in which they reside (this distinction is important and one we return to). The area type is defined by the classification used to sort the consumers into groups. Such a classification would normally be purchased from a third-party data vendor (see Chapter 3 for examples). A 'look-up' file then allows the retail company to determine in which type of neighbourhood each of its customers lives.

As we discuss in Chapter 6, the neighbourhood classification is produced by the data vendor as a statistical amalgam of small-area (often census) data for the N mapping units of the region concerned (usually a country or nation). There are, for example, $N = 175{,}434$ small output areas in England and Wales following the 2001 Census. There are approximately $N = 8.5$ million blocks recorded in the 2000 US Census and there are $N = 38{,}366$ meshblocks covering New Zealand for its 2001 Census. The geodemographic classification is produced by grouping the N areas into a much smaller number of k classes, on a like-with-like basis. Commonly k is in the range from about 10^1 (10 clusters or area types) to 10^2 (100 clusters or area types), depending on the level of granularity required for the analysis. For example, the 1991 Census-based SuperProfiles classification of the UK has $k = 160$, $k = 40$ and $k = 10$ available to choose from, allowing the user some choice in the precision of geodemographic analysis desired. That hierarchy of clusters is shown in Figure 1.5.

The retail company completes its analysis by comparing the proportion of its clients in each of the k classes with the corresponding proportions for all consumers within the company's catchment area (or some other suitable measure). This comparison of the observed distribution by neighbourhood type with an otherwise expected baseline distribution allows the retail company to infer useful information about its core customers and market to them accordingly (see Chapter 5 and also Birkin, 1995).

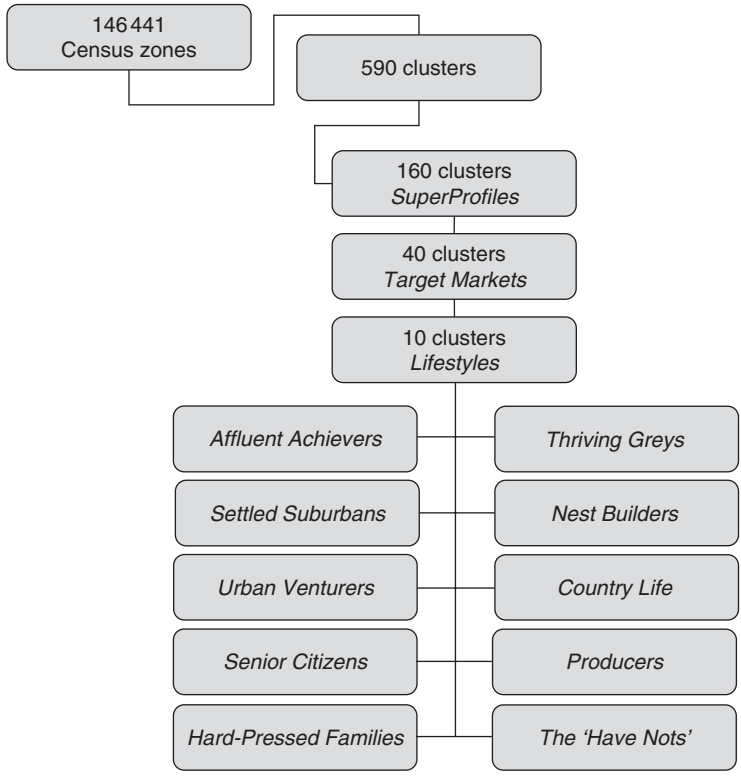

Figure 1.5 The post-1991 Census, SuperProfiles classification hierarchy. Source: based on Batey and Brown (1995, p. 94)

An interesting and somewhat contentious real-world application of the SuperProfiles classification has been by the Higher Education Funding Council for England (HEFCE). It has used geodemographics as part of its programme of widening participation in higher education by increasing the recruitment and retention of students from under-represented social groups. HEFCE (1998, Annex A, paragraph 17; and 1999, paragraphs 24 and 22) make clear that:

> Our preferred method of measuring the socio-economic background of students is through the use of the postcode to identify the neighbourhoods of students' homes. We have found that the Super Profiles geodemographic classifier is able to distinguish between neighbourhood types with markedly different higher education participation rates and, currently, this classifier is our first choice for funding allocations. [para 17]

> [T]he additional funding will be applied to young full-time undergraduate entrants to HE [Higher Education] from lower than average participating geodemographic groups in each institution. [para 24]
>
> Although we recognise that there are some flaws in the use of geodemographic data, we believe this is the best measure currently available for allocating funds in respect of students from disadvantaged backgrounds. Therefore, we will use this measure until any better and more precise measures are developed in the future. [para 22]

At the time of writing, an education-specific version of the UK Mosaic geodemographic classification (see Chapter 3) was being developed in conjunction with the UK's Universities and Colleges Admissions Service (UCAS) (Farr and Singleton, 2004).

1.5 How it works: a short theory of geodemographics

In seeking to define the nature of geodemographics, one of us (Harris, 2003, p. 225) has suggested it is 'the analysis of socio-economic and behavioural data about people, to investigate the geographical patterns that structure and are structured by the forms and functions of settlements.' Taking time to consider what that long-winded statement actually means sheds light on why neighbourhood classifications may be useful, predictive tools for analysis and decision taking.

First, the statement suggests that identifying geographical patterns or trends within societies is an important step towards understanding the processes and phenomena that gave rise to those patterns in the first place. Where appropriate, those processes or geographical events can then be managed. Admittedly, it is a simplistic 'theory' and the link between pattern and process is often much more ambiguous. For example: two or more completely different processes can result in similar outcomes (geographical patterns); particular socio-economic processes do not operate in isolation from others present at the same time and in the same place, so disentangling the effects of any one from the others may be difficult, perhaps even pointless; any one outcome may be unique to a particular time and place and, in such a case, ought not be generalized; and, finally, since order can arise from chaos, simple notions of cause and effect may not be tenable (cf. Flake, 1998).

Such caveats are important and increasingly central to geographical and other disciplinary understandings of how settlements and societies

form and function (see, for example, Longley and Batty, 2003). However, Webber (1975) suggests that whereas most social scientific research is undertaken in order to investigate existing hypotheses (the deductive approach), one of the key purposes of a geodemographic classification is to generate new ideas and insights that can then become subject to further debate and analysis (an inductive approach; cf. Chapter 9, Section 9.2.2). In this regard, geodemographics is regarded as a data exploration tool, not a statistical method of hypothesis confirmation or rejection. While we fully recognize and will explore a number of limits to the geodemographic method, at the same time we echo the sentiments of an editorial written in response to an article by one of us, in which had been questioned the merits of the geodemographic approach (Harris, 1999). The editor wrote:

> The elegant simplicity of the insights that can still be gained into the overall profile of one's customers by a simple analysis of postcodes is one which continues to impress this practitioner. [. . .] Mapping and geodemographics continue to offer powerful benefits (Whitehead, 1999, pp. 109–10).

If we consider again an analysis of neighbourhood deprivation that is intended to guide the allocation of finances to local regeneration schemes, it is self-evident that targeting resources to the right communities requires that the locations of those communities first should accurately be determined (Policy Action Team 18, 2000). Cross-referencing those locations with, for example, an analysis of house prices might give an indication of the sorts of socio-economic processes (in this case, the effects of housing markets) that cause the geographical differentiation of communities and which lead to social exclusion. The fact that various socio-economic forces can 'filter' people to live in certain 'types' of neighbourhood is the basis for how neighbourhood analysis predicts the likely population characteristics to be found in any given area. Hence, pattern reveals process, and process invites prediction – of, for example, changing consumer markets (Miles, Anderson and Meethan, 2001) or of the growth of urban sprawl (Besussi and Chin, 2003).

However, the relationship between places and people is neither one way nor solely the consequence of external factors. When people speak of 'their neighbourhood' or 'their community' they do so in a way that suggests an attachment to place. Harris (2003) implies that there is an interrelationship between people and places – the link illustrated by Figure 1.1. Therefore, the physical, social and economic properties of settlements in some way reflects the character, choices, preferences, ideals, affluence, consumer lifestyles (and so forth) of past and present populations living in those settlements but also are a consequence of governmental policies, for

example in respect of planning controls and social housing initiatives. Because a place usually pre-dates the residents, so the relationship is two way: the style and character of the settlement 'draws in' certain population groups, perhaps by choice, perhaps by necessity; those residents then shape further evolution of the area. Longley and Batty (1996, p. 76) write that:

> [t]he behaviour of individuals in [geographic] space together contribute to the development of places over time and these place effects in turn condition subsequent spatial [geographical] behaviour.

The interrelationship suggests that measures of the physical, social and economic properties of settlements can yield useful information about the characteristics, preferences and lifestyle choices of the populations resident within those settlements, because people and places are dependent on each other.

These theoretical ideas are summed up by the adage 'birds of a feather flock together'. This, according to Flowerdew and Leventhal (1998) is the basic tenet of geodemographics. In fact, birds of a feather may not just flock together but also increasingly become alike. This is because very few of us (the birds) live in complete isolation from the rest of the society (even if there are times when we wish otherwise!). It is likely that many of our behaviours, choices, aspirations and ideals are influenced by those with whom we interact in our everyday lives (and vice versa) and to assume otherwise is known as the atomistic fallacy (see Chapter 9). Despite the emergence of cyberspace and the popularity of online chat rooms or other forms of communication, it remains reasonable to suppose that geographical distance and location impart constraints on who we meet and when. Weiss (2000, p. 25) argues that there is value in classifying populations at a neighbourhood level and it relates to a 'core truth . . . you are like your neighbors.'

The geographical effects of 'birds' flocking together are expressed by Waldo Tobler's often quoted first law of geography: 'Everything is related to everything else, but near things are more related than those far apart' (Tobler, 1970; see also the forum discussing the law in the *Annals of the Association of American Geographers* **94**(2)). This 'law' is an expression of what spatial statisticians refer to as spatial autocorrelation (Cliff and Ord, 1973). This type of autocorrelation is present in a dataset if it can be shown that particular attributes of the population (such as low-income pensioners, lone-parent households, eat-out regularly couples or sports-car-owning adults) display a non-uniform and non-random patterning but, instead, are clustered into particular localities.

Geodemographic methods assume (positive) spatial autocorrelation when residents of the same neighbourhood are taken to share, in broad terms at least, some common socio-economic and/or behavioural characteristics. This stance permits the consumer profile of a non-sampled person living in a certain neighbourhood to be inferred from data about other persons living in the same neighbourhood – the assumption being that proximity is related to similarity. However, Tobler's first law does need to be modified when looked at in a geodemographic context. The geodemographic methods that are the subject of this book assume not only that proximate populations are related but so also are populations living in the same 'class' of neighbourhood. In other words, near and far things are related – by neighbourhood type.

The use of the word 'neighbourhood' is problematic here because, as we explore in later chapters, it can have formal and informal, administrative and sociological meaning (Martin, 1998b). Moreover, from an 'internal' perspective, Weiss (2000) identifies that while geodemographic literature has often interchanged the words cluster and neighbourhood – for reasons that if not already apparent will become so in Chapter 6 – such synonymy is no longer always appropriate. The reason is that classifications have been built at increasingly fine resolutions. The evolution in the UK, for example, has been first electoral wards, then census small areas, then even smaller unit postcodes and, finally, households and individuals. Intuitively, at some scale between wards and households, the importance of neighbourhood has disappeared – but at which scale?

The answer, we suggest, is none. Unless we believe that people live in a social vacuum, that physical proximity to others has no effect on our own behaviours and that everyone has an entirely unconstrained choice about where they live, then geography remains important and so the presence of 'neighbourhood effects' will be captured even if we collect data at an individual or household scale. Geography matters and is usually there somewhere within the data, not least because apparently individual or household classifications usually incorporate areal data to improve the robustness of those geodemographic models.

These considerations partly explain why we prefer the term neighbourhood over cluster type and use it more frequently (though not exclusively) throughout this book. It is also because the use of the word neighbourhood helps identify the origins of geodemographics in social research (Chapter 2). Finally and most pragmatically, the majority of this book *is* about classifying small areas ('neighbourhoods' in a strictly technical sense) and about using geography as the basis for modelling, even inferring people's demographic, socio-economic and behavioural

characteristics. We do not ignore individual or household classifications but the focus is more geographical.

It may also seem ornithological! Returning to birds of a feather, neighbourhood classifications use multivariate (many variable) clustering techniques to model geographies of 'the flocks'. They simplify a complex geographic reality to make the basis and process of decision making easier, faster and more intelligible to stakeholders such as boards of directors or front-line operatives, including store managers and divisional superintendents. Like any other geographic model, geodemographics can both be criticized and praised for the level of abstraction they impart. This is especially true of multivariate techniques that numerically summarize often complex spatial patterning within datasets and consequently create an uneasy, operational tension of trying to express that complexity using a simple and understandable, 'everyday language'. The tension appears most acute when geodemographic vendors attempt to sum up each type of neighbourhood produced by their classification techniques using evocative vocabulary. Unsurprisingly it is these verbal labels or 'pen portraits' that have attracted most criticism within some academic literature (notably Goss, 1995; Curry, 1998). In any case, there are conceptual, theoretical and practical limits to what any area classification or dataset can usefully reveal about the character of populations living within neighbourhoods (we discuss some of these in Chapter 8). Accordingly, there are times when neighbourhood analysis is appropriate and times when it is not. Discerning between these two instances is important and by means of this book we hope to facilitate such a judgement.

One important use of neighbourhood classification is when analysts encounter the task of predicting the likely behaviours and needs of existing or potential new customers or clients in the absence of direct information about those customers, clients or consumers. If the scenario sounds unlikely then consider that for various reasons which include privacy laws and the regulation of personal data, specific personal data about individuals is usually less easily obtainable (and then more expensive to purchase) than aggregate, summary information about groups of people. In some cases, obtaining individual information may actually be illegal! Even supposing that individual-level data were available to the analyst, it is rare for such a source dataset to be fully relevant, comprehensive and complete for the entire population to be analysed (see Chapter 9 for a discussion of these issues in the context of 'lifestyle databases'). Geodemographic information can therefore fill holes in our knowledge base. In the absence of more specific information, knowledge about neighbourhoods provides a useful (and perhaps only) step towards

knowledge about people. Combining this knowledge with what is revealed from loyalty card, EPOS (electronic point-of-sale) or other sources of consumer/population data is the first step towards, for example, building predictive models of consumer behaviour or to targeting accurately community funding.

Case study: modelling price sensitivity and geodemographic categories in the restaurant market

Martin Callingham, professor in geography, Birkbeck College, University of London (formerly group manager of market research for Whitbread plc)

This case study is about how geographically to tailor prices within a market. Specifically the work was done in the restaurant market for a chain of about 280 pub restaurants called *Brewers Fayre* that was geographically dispersed about the UK. The sites selected to build these restaurants were semi-rural and the product offering was a low price but substantial, traditional English, full restaurant meal. The chain was very successful (a brand leader) and this appeared to offer potential for increasing the overall prices. However, there was concern that such an increase would negatively impact upon outlets located in poorer areas and this in turn suggested differential pricing should be introduced. To make this introduction, two questions needed to be addressed:

- How to identify the range of prices to be used.
- How to allocate each restaurant to a price band.

To answer these questions a highly novel approach was adopted and used to attribute a price-sensitivity index to ACORN classes of neighbourhood type (ACORN is a commercial geodemographic classification). From existing knowledge of the ACORN mix within each of the restaurants' catchment area it would then be possible to model the collective price sensitivity around each restaurant. However, an important methodological issue had first to be addressed. Specifically, what should be optimized by varying the price? Is it the number of trips to the outlet or the profit that accrues from the trips? Clearly it would be possible to lower prices and increase total trip numbers but this is of no value to the company or its shareholders if the effect is to lower net cash profit. The intent, therefore, was for the restaurants to be more profitable.

The method adopted was to construct a market research survey of 1200 people stratified by neighbourhood type (200 per ACORN category) and where each person was located within the catchment area of an established Brewers Fayre restaurant. To enable the efficient use of interviewers, 30 (about 10%) restaurants were selected at random from the chain list, using an interval sampling method. Interviews, which were conducted face to face, broadly were quota controlled by ACORN category within each catchment area and interviewers were given a postal address list to help them achieve this.

The price-sensitivity questions used were a modification of the method developed by Van Westendorf. This is an unusual method that allows individuals to select prices at their personal conceptual breakpoints. Specifically there are four breakpoints: two 'outside' prices – the price which is so low that the quality of the product would be in doubt and the price at which the product is too expensive – and two 'intermediate' prices at which the product seems either on the cheap side or is becoming too expensive. Respondents were asked to select a price for a typical main meal at a Brewers Faye restaurant from a choice of possible prices covering the entire sensible range (and more!) and broken into a non-obvious interval to avoid conceptual 'rounding up': in this case the prices differed by seven pence units. The meal was described to interviewees using a show card.

The resulting data are normally analysed by a cumulative process that suggests the proportion of the sample that is prepared to pay a particular price for a particular meal. However, in this study a more insightful modelling method was preferred. This was based on asking two additional questions about behaviour at two of the prices given – specifically, the breakpoint prices identified by each respondent (and therefore differing from individual to individual). At the two outside prices, where the product either is too cheap or too expensive, no sale would be expected so the frequency of visiting a restaurant would be zero. However, at the two intermediate prices, sales would be expected, so the interviewee's likely frequency of visiting a restaurant at these two prices was asked. Generally, though not always, the frequency was higher for the lower of the two prices.

The price sensitivity of each individual was then modelled using three linear equations: one for the prices between 'too cheap' and 'on the cheap side'; one for prices between 'on the cheap side' and 'on the expensive side'; and one for prices between 'on the expensive side' and 'too expensive'. Subsequently, the three equations were used to impute an expected visiting frequency of each individual at any specified price. The

derived frequency of visiting a restaurant for each respondent was calculated within a fixed range of prices: in this case for 25p differences from £3.50 to £8.50 giving 21 different prices in all (the current price for a main meal was actually £5.40).

Next, the individual frequencies were grouped by ACORN category. This permitted the mean frequency of visiting (and its variance) to be calculated by neighbourhood type, giving a grid of six ACORN categories by 21 imputed frequencies. At each position in the grid the cash profit for a trip to a Brewers Fayre was calculated by applying financial analysis rules to determine profitability. (Profit is generally taken at the outlet level to be the sales less the costs of achieving them but how costs are attributed to sales is quite complicated. Some costs per sale, such as the cost of capital, get less as the sales increase since there are more sales to spread the 'fixed costs' over. Some costs go through steps changes as sales increase – the need to put on more staff, for example – and some costs, such as ingredients, stay very similar per meal sold.)

The resulting (second) grid of net cash profit was then separately multiplied by the ACORN profile of each of the restaurant catchments where the profile expressed the actual number of people of each neighbourhood type that were to be found within each catchment. This gave the modelled net cash profit that could be expected for each restaurant at each of the 21 price points and thereby suggested the most effective profit-generating price for each specific restaurant. A distribution of these 'most effective' prices was made and from this three price bands were selected to be applied across the chain. Each restaurant was allocated to a price band according to the geodemographic profile of its catchment and according to the modelled price sensitivity of that particular socio-economic mix.

Case study: using geodemographics in the public sector

Keith Dugmore, Demographic Decisions Ltd and chair of The Demographics User Group

The development of geodemographics in the public sector of the UK really has been rather odd! Taking the core of the public services to be central government departments, local authorities and health authorities, interest in area classifications based on Census statistics began to develop as early as the 1960s and blossomed in the 1970s. There were several strands. In local

government, the Greater London Council (GLC) was very active, producing classifications of boroughs and then electoral wards. Liverpool Council developed a finer scale Enumeration District (ED) level classification focused on deprivation. In central government, the Office of Population Censuses and Surveys published a national classification of local authority districts, while in the mid-1970s work in this field was pushed further forward at the Centre for Environmental Studies (CES), developing nationwide classifications down to ward level. This early history of geodemographics is discussed in more detail in the following chapter.

Meanwhile, in another part of 'the forest' during the late 1970s, area classifications began to be applied for target marketing. This topic is covered in more detail in Chapter 3 of this book. Suffice to say that starting with ACORN, which provided an ED level classification for the whole of Great Britain, there were two vital innovations. First, the statistical clusters were popularized using names, descriptions and photographs. More importantly, the EDs were matched to the Central Postcode Directory to produce a postcode to geodemographic code look-up table – a means to link a person's property address to their neighbourhood type.

It was this that led to a great leap forward, also enabling area classifications to be linked to other datasets and thus giving birth to geodemographic analysis, rather than just standalone area classifications. Analysis took three main forms and opened up new territory for users in thousands of commercial companies. It also sets the scene for considering geodemographic developments within the public services. These forms are:

- Coding and analysing sample surveys. By taking records of respondents coded with a postal unit identifier and attaching a geodemographic code to these it became possible to profile sample surveys. The most celebrated early example was the contrast between the profiles of *Daily Telegraph* and *Guardian* readers exposed by the Target Group Index survey (a consumer research survey – see Chapter 3).
- Coding and profiling customer files, seeking variations between different segments of the data for such topics as consumption of goods and services, and customer profitability.
- Area analysis, profiling non-standard areas such as postal sectors and store catchments (see Chapter 5). When combined with profiles from surveys or customer files, this opened up the potential for estimating and comparing market potential in different parts of the country.

Such analyses have proven to be of great value to commercial companies. Suppliers have vied to produce new geodemographic classifications and

their use has expanded in the last two decades to become second nature in almost all large and medium-sized companies which sell in the consumer marketplace.

Meanwhile, following the arrival of the 1981 and then 1991 Censuses, the public services continued their keen interest in area classifications. But how much geodemographic analysis, linking datasets, was done? The answer is: remarkably little! Taking the three forms of analysis done by commercial companies, there are several examples but these are sporadic.

Some sample surveys have been coded. The Home Office's British Crime Survey was first coded with ACORN back in the 1980s and this led to some fascinating analysis of fear of crime in different types of neighbourhood. More recently, the ONS' Expenditure and Food Survey has been coded with both ACORN and Mosaic but this is another rare exception among the very large number of government surveys.

Turning to the coding of customers or administrative files, progress has been very patchy. Following the report of the Korner Committee on Health Service information (The Korner Committee, 1984), health authorities began during the 1980s systematically to postcode patients' files and this led to some interest in geodemographic coding. Some police forces have also shown interest (although this has often evaporated when the pioneer responsible was promoted). In general, few local authorities have sought to code files and it is significant that the addition of postcode analysis to the ubiquitous SASPAC Census software in 1991 was little used.

As for area analysis, the interest in producing area classifications has continued unabated, with both general and specific (for example health) classifications being produced for a variety of standard administrative areas. In some cases, such as the Index of Multiple Deprivation 2000 and the updated Index of 2004 (DETR, 1998; ODPM, 2004), non-census data sources have also been included but the classifications themselves usually have been treated as free-standing, rather than integrated with survey or administrative data to estimate, for example, the incidence of smoking in small areas.

There appears to be no single reason for this slow progress but the factors probably include the traditional cultural divide and mutual incomprehension between private and public sectors in Britain, manifested in this case by a reluctance to adopt techniques that are used by marketers and also the fact that the commercial geodemographic classifications have not been available free of charge. The result has been that the use of geodemographics has not become everyday in most public-service organizations.

However, this looks set to change. From a policy point of view, governments have become more interested in targeting their resources, both to areas and individuals, and this seems likely to continue. The policy of neighbourhood renewal in the UK has triggered a demand of better data for small areas and the development of the various neighbourhood statistics services (see Chapter 3, Section 3.2). This will increase the use of multiple datasets for small areas and should encourage more creative integration of sources.

Turning to the supply side, the most dramatic change is that the ONS is producing an area classification down to the finest census output area (OA) level for the first time. This is for the whole of the United Kingdom and, when allied with the postcode to OA look-up directory and digital boundaries, which are freely available as part of the Census Access project, will provide an immensely valuable resource. The door is now open to much more adventurous exploration of geodemographic analyses by public services. The most obvious starting point would be for the new classification to be appended to all government surveys and for profiles to be produced as standard products for a wide range of topics such as crime, health and education. Both these and profiles derived from coded administrative files – such as the Inland Revenue – could be included as part of the Neighbourhood Statistics website. The availability of such information from central government would then have every chance of encouraging similar developments among local government, heath authorities and police services.

1.6 Where next? An overview of the following chapters

In this chapter we have introduced geodemographics as the analysis of people by where they live, identified how neighbourhood classifications might be used for both public- and private-sector decision making and offered some ideas on how neighbourhood-based analysis can be a useful way of making sense of geographical information.

In the next chapter we look at the precursors to the present geodemographic industry, in particular Charles Booth's studies of poverty within London and the Chicago School of urban sociologists. Our gaze is more than one of academic curiosity but views also the origins of the geodemographic method, including its strength and weaknesses as an

analytical approach. Here we introduce the risk of ecological fallacy, something that has been described as the stick by which geodemographics is beaten. In Chapter 3 we continue our study of the evolution of geodemographics exploring how it came to be commercialized and the 'big players' that have emerged within the sector.

In Chapter 4 we shift our attention sideways to look at another field where analysing geographical information is paramount. Here we outline the principles of geographical information systems (GIS) but do so from the perspective of a geodemographic user. We also make connections between GIS and geodemographics but go on, in Chapter 5, to suggest that GIS are not always necessarily the best software to undertake typical sorts of geodemographic analysis such as neighbourhood profiling or catchment analysis. Therefore, in Chapter 5 we also compare and contrast GIS with what have been labelled geodemographic information systems (GDIS). We describe the sorts of functionality a user might expect from a GDIS and demonstrate the types of analysis that may be carried out within it.

Chapter 6 is about building a neighbourhood classification and considers in detail the data sourcing, validation and grouping strategies employed by commercial vendors, as well as the sorts of issues they encounter when constructing a neighbourhood typology. Chapter 7 considers some of the differences in the construction and geographies of neighbourhood classifications that have been built around the world.

Chapter 8 adopts a more critical perspective and asks – 'but does geodemographics work?' Here some of the shortfalls of the geodemographic method are addressed, notably the problems of diversity within neighbourhood types, modifiable units and whether geographical data require explicitly geographical ways of analysing them. The critique is counterbalanced by the ultimately pragmatic consideration 'well, it has worked for many!' and suggests why. Examples of where geodemographics has proved useful are given.

Chapter 9 brings with it a change of scale. Whereas all the preceding chapters are about top-down, area classifications, here the application of segmentation techniques to individual or household data are considered, together with 'bottom-up' geolifestyle methods. Such approaches require that special consideration be given to personal data protection issues; issues of data uncertainty must also be addressed. Finally in what we have referred to as a postscript, we identify the 'three Is' that gave rise to geodemographics and continue to shape its future.

1.7 Conclusion

Geodemographics is the analysis of people based on a statistical classification of the area in which they live. The classification aims to capture the important socio-economic 'dimensions' of, and differences between, neighbourhoods. The geodemographic approach has been found by many to be a useful aid for guiding decision making and the management of geographical information.

Neighbourhood classifications usually are produced by grouping together a large number of usually administrative units into a much smaller number of groups, clusters or neighbourhood types on a like-with-like basis. A common choice of data to define the similarity, or otherwise, of neighbourhoods are national census statistics. Such classifications have been the bedrock of a rapidly growing industry that has its origins in urban geography and sociology. Present applications include survey design, retail planning, direct marketing and media analysis, as well as other strategic marketing, planning and decision taking in both the public and private sectors. Neighbourhood classifications can be used to look for geographical patterns in various socio-economic, behavioural, attitudinal or consumer datasets.

The usefulness of neighbourhood classifications derives from the idea that knowing where someone lives provides useful information about how someone lives. A simple theory of geodemographics is there is an inter-relationship between people and places, and also between individuals and the people they regularly meet. The adage, 'birds of a feather flock together' and Tobler's 'first law of geography' go some way to explaining why neighbourhood classifications can usefully be applied to extract information about people from information about places. When direct knowledge about potential customers, clients or consumers is not consistent, neighbourhood analysis provides an inferential tool linking what is known to what is not.

Summary

- Geodemographics has been described as the analysis of people by where they live.
- It has been widely used to inform strategic marketing and planning.
- Neighbourhood classifications usually are produced by grouping a large number of administrative units into a smaller number of clusters, on a like-with-like basis.

- An assumption is that 'birds of a feather flock together' such that populations living in the same neighbourhood type share broad socio-economic and consumer characteristics.
- Multivariate classification techniques simplify a complex geographic reality to make the basis and process of decision making easier, faster and more intelligible to stakeholders.
- Geodemographics is better for exploratory analysis than for hypothesis testing.

Further Reading

- Longley, P. and Clarke, G. (eds) (1995) *GIS for Business and Service Planning*, GeoInformation International, Cambridge, Chapters 5–6.
- Sleight, P. (2004) *Targeting Customers: How to Use Geodemographic and Lifestyle Data in Your Business*, World Advertising Research Center, Henley-on-Thames.
- Weiss, M. (2000) *The Clustered World*, Little, Brown, New York.
- Our website: www.geodemographics.info.
- The Geodemographics Knowledge Base: www.geodemographics.org.uk – a comprehensive list of websites for people interested in the application of geodemographics and geospatial analysis, produced by the Census and Geodemographics Group of The Market Research Society.