CONCEPTS OF SPACE AND GEOGRAPHICAL DATA

A C GATRELL

Space is defined here as a relation on a set of objects. Different types of space may be created, depending upon how this relation is defined; Euclidean distance is perhaps the simplest example. This chapter begins by reviewing such concepts of space and then considering what is meant by the term 'spatial object'. Different categories of such objects are discussed and the varying types of attributes or data that can be associated with those objects are examined. Having looked at standard concepts of distance some consideration is given to other measures of spatial proximity, such as time distance. The need to visualize spaces defined by different types of relation is addressed and examples are presented of map transformations (such as cartograms).

INTRODUCTION

> *The true potential value of Geographical Information Systems lies in their ability to analyse spatial data using the techniques of spatial analysis.* (Goodchild 1988: 76)

Although something of an over-simplification, it is worth making a distinction between two ways in which GIS can be used. The first is to use GIS to ask descriptive questions, such as spatial queries of the form: 'locate and display all settlements with a population greater than 5000'. These are questions of the 'what' or 'where' type and are, of course, genuinely important in planning and resource management. Embodied in the quotation, however, is a second category of use: the need to ask analytical questions, perhaps to construct models and perform predictions. These may involve questions prefixed with 'why' or 'what if'. It is essential, therefore, that we endow a GIS with capabilities for spatial analysis (see Openshaw 1991 in this volume), 'that set of analytical methods which requires access both to the attributes of the objects under study and to their locational information' (Goodchild 1988: 68).

Given that spatial analysis and GIS are thus inextricably linked there is a need to investigate more fully what concepts of space are appropriate and what types of spatial objects and data must be handled. This is undertaken here by first seeking a general, and very powerful, definition of 'space' and then constructing a typology of 'spaces'. Different types of spatial object are then considered before examination of different types of spatial relationship. Examples of visual representations that arise from such spatial relationships are offered later.

THE WORLD IS FULL OF SPACES

Space is taken to mean 'a relation defined on a set of objects'. This immediately requires definition of what is meant by 'relation' and 'object': much of this chapter is concerned with uncovering the meanings of these words. To be more concrete, consider an example (Fig. 9.1), where the objects of primary interest are landmarks in Manhattan and the relation that binds these landmarks together is the 'distance' between them. (It will soon become clear why there are quotation marks around the word distance.)

It may be assumed that the objects are of interest (otherwise they would never have become

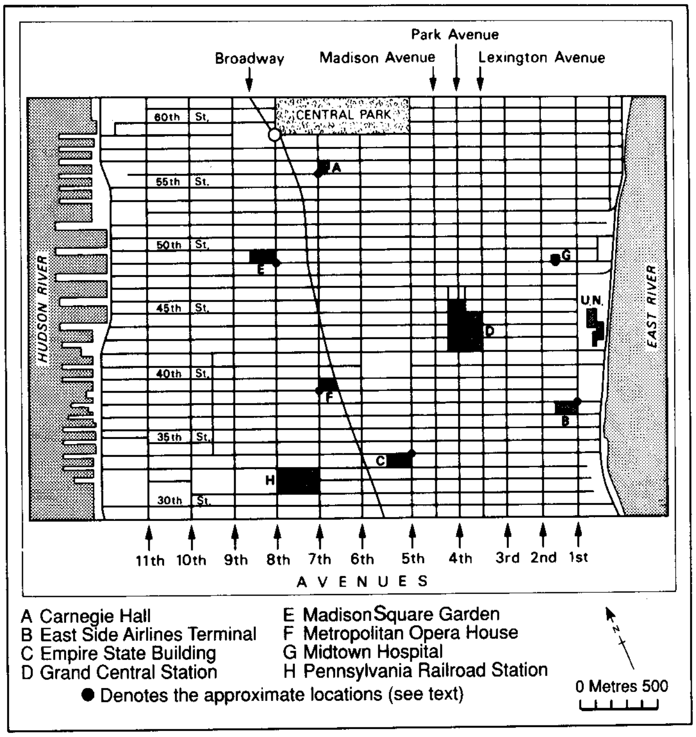

Fig. 9.1 A set of landmarks in Manhattan, New York City, USA.

part of a GIS!) and, therefore, that they have attributes or properties associated with them; numbers of people employed, or rateable value, for instance. Indeed, spatial objects such as these may be defined as a set of spatial locations together with a set of properties characterizing those locations (Smith *et al.* 1987). The term 'object' has a different meaning in the field of object orientation where it simply means any feature of interest (including spatial features, as defined above, non-spatial features and attribute data; see Maguire, Worboys and Hearnshaw 1990).

Thus, 'space' embraces the following: a 'set' of objects (what Peuquet, 1988, prefers to call 'entities' and others refer to as 'features'), to which may be attached associated attributes ('properties' in Peuquet's terminology), together with a relation, or relations, defined on that set. A straight line drawn between pairs of objects is one such relation. As is made clear below, however, this is only one way of conceptualizing 'distance', and there are in turn several other ways of looking at spatial relationships.

Metric spaces

The concept of metric space can be illustrated by examining the landmarks in Manhattan a little more closely and assuming, for the sake of argument, that they occupy point locations rather than physical areas. Thus there is a set of locations whose coordinates can be denoted as $\{x_i, y_i\}$ and distance can be measured conventionally between members of this set as:

$$d(i,j) = \sqrt{[(x_i - x_j)^2 + (y_i - y_j)^2]} \qquad [9.1]$$

This is a formal definition of Euclidean distance learnt at school as part of Pythagoras' theorem. At the scale considered here (embracing a relatively small part of New York City) complications due to the curvature of the earth's surface can reasonably be ignored, but at wider spatial scales they would need to be taken into account (see Maling 1991 in this volume).

Euclidean distance is an example of what mathematicians call a 'metric'. A formal definition of this is offered elsewhere (Gatrell 1983: 25–6) but, briefly, two properties are important; first, such distances are symmetric: $d(i,j) = d(j,i)$; second, they obey the so-called triangle inequality:

$$d(i,k) \leq d(i,j) + d(j,k) \qquad [9.2]$$

For example, the distance from the Empire State Building to East Side Airline Terminal is no greater than the sum of the distances from the Empire State Building to the Metropolitan Opera House and East Side Airline Terminal to the Metropolitan Opera House.

This Euclidean view of the world is firmly embedded in most of what happens under the name of GIS. It is, of course, entirely appropriate in land management applications, in survey work and in the kind of large-scale digital mapping undertaken by the utilities (see Mahoney 1991 in this volume). In other words, for most descriptive applications of GIS it will suffice, as indeed it will for much spatial analysis. However, it is quite clear that Manhattan is not a flat, featureless plain but that space there is structured by the road network. A Euclidean view of Manhattan is inappropriate because 'real' distances are not, in most cases, as the crow flies. Instead, city blocks are traversed. This is familiar to anyone giving or receiving directions in a 'grid-based' urban environment; 'go two blocks north and three blocks east'.

An alternative method of measuring distance, which takes this into account is called 'Manhattan' or 'taxicab' distance:

$$d(i,j) = |x_i - x_j| + |y_i - y_j| \qquad [9.3]$$

Three points can be made about this measure of distance. First, as with Euclidean distance, it too has the properties of a metric; distances are still symmetric and the triangle inequality still holds. Secondly, if questions are asked about spatial proximity, the results obtained in terms of the relative orderings of distances may well differ from those given by the Euclidean metric (Gatrell 1983: 25–9). Thirdly, Manhattan distances are not coordinate invariant; if the axes are reoriented, different measurements of distance between pairs of points can be obtained. This suggests that the Manhattan metric may only be of real value in grid-like cities in which the axes follow the street pattern.

The differences between Euclidean and Manhattan metrics are not purely of academic interest. Different results will be obtained from GIS functions depending on which of these two metrics is used. The following two examples illustrate this. First, if Thiessen polygons are constructed around points (which might be a first simple attempt to define catchment or trade areas) in Euclidean terms this can be done by finding bisectors of straight lines joining pairs of points (Fig. 9.2). This is a common function in GIS. However, if this is carried out in Manhattan space then quite irregular boundaries are obtained (von Hohenbalken and West 1984). Within each polygon are all locations lying closer to the central point than to any other, but this 'closeness' is defined in an unconventional way (Fig. 9.2).

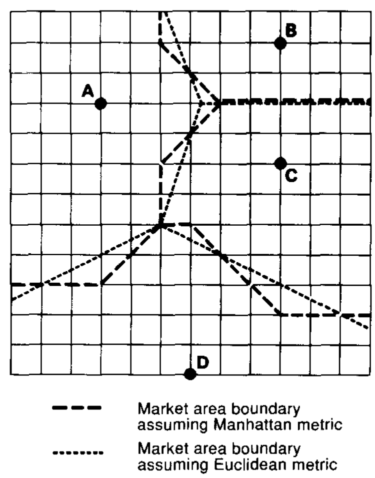

Fig. 9.2 Thiessen polygons using Euclidean and Manhattan metrics.

As a second example, suppose it is important to find an optimal location for a new supermarket or public facility. Using a very simple example, suppose that the 'demand nodes' at which the service is required are equally weighted and exist at only a handful of point locations in Manhattan space (Fig. 9.3). It is easily verified that, if there is an odd number of demand nodes (Fig. 9.3a), then there is a unique optimal location (that which minimizes the sum of distances to demand nodes); this may be located, as here, at one of those nodes. However, if there is an even number of demand nodes (Fig. 9.3b) there is no unique optimum; rather, there is a region, within which any location is equally optimal. In Manhattan space, therefore, it is possible to detect some 'uncertainty' in location and this is simply a property of space itself.

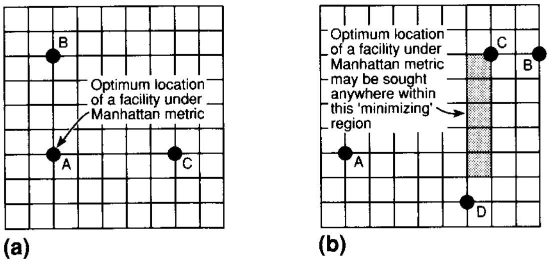

Fig. 9.3 Optimal locations using the Manhattan metric.

Armstrong (1988) has looked empirically at how different definitions of distance influence the results of such location (and location-allocation) models, using an example from the southeastern United States. Having defined a set of 26 demand nodes and 10 facilities to be allocated he then computes distances along the road network, where this is digitized at different scales. He compares these estimates with Euclidean and Manhattan distances and demonstrates how the results change quite dramatically as different measures of spatial separation are adopted. There needs to be wider recognition of this variation among users of location-allocation models within a GIS framework.

Non-metric spaces

So far, only metric spaces have been considered. Yet non-metric spaces will be encountered frequently in some applications of GIS, even if the formal properties of such spaces are unlikely ever to be defined. For instance, some GIS permit weights or impedances to be attached to the route links in transport networks. These might be used to model one-way streets, for instance, or to restrict movement along narrow roads and encourage use of major roads as is done in sophisticated shortest path algorithms (Dunn and Newton 1989). One of the implications of such work is that a new model of space is defined, one which is unlikely to be metric. In general, distances will not be symmetric, nor will the triangle inequality hold.

Thus far it has been assumed that spaces in which objects have locations are characterized by a pair of coordinates $\{x_i, y_i\}$. There are two remarks to be made about this, one of which requires an extended discussion. The first comment is that, because of inherent uncertainty, the locations of spatial objects may be endowed with a spurious accuracy. For instance, soil or land use boundaries may be represented by strings of locational coordinates but these lines are inherently fuzzy (see Chrisman 1991 in this volume). While some attempts have been made to introduce notions from the theory of fuzzy sets (Robinson, Miller and Klesh 1988; Burrough 1989), no work has been done on how fuzzy spaces may be incorporated within a proprietary GIS. This is an important area of research since our everyday spatial concepts (for example, 'near', 'far', 'close to'), which may be used to phrase spatial queries, are inherently vague and uncertain (Mark, Svorou and Zubin 1988; see also Frank and Mark 1991 in this volume).

The second remark to be made about locational referencing is that it is not always necessary to have accurate coordinate information. In the case of how to use a GIS to navigate from one node to another on a road network (see White 1991 in this volume), is it really essential to know all the locational coordinates of the links on the network to perform this task adequately? Consider the well-known example of the London underground map (Fig. 9.4). This contains all the useful information needed to get from i to j on the network. Yet the locations are, to some extent, arbitrary and the relation that defines this non-metric space is simply one of connectedness; whether stations are adjacent on the network. The visual representation allows us to make spatial deductions about the ease of travel between stations which are not adjacent.

Many will be familiar with this map and with the fact that it is commonly used in textbooks to introduce the concept of topology (e.g. see Abler, Adams and Gould 1971). The map is a visual representation of a topological space. Note, as an aside, that even in fully vector-based GIS, with all locations represented by $\{x_i, y_i\}$ pairs, people speak of the topology of the link–node structure. This refers to the ability of the GIS to store (among other things) what lies to the left and right of a line segment. Thus it is possible to refer to the topology of a Euclidean representation of space. To repeat, a topological space is one in which there is some arbitrariness in the positioning of locations and arcs and where the only relation that matters is contiguity.

Some GIS adopt a data structure that is a hybrid of Euclidean and topological space. A good example is the TIGER system (Topologically Integrated Geographic Encoding Referencing) used by the US Census agency (Broome 1986). Here, a

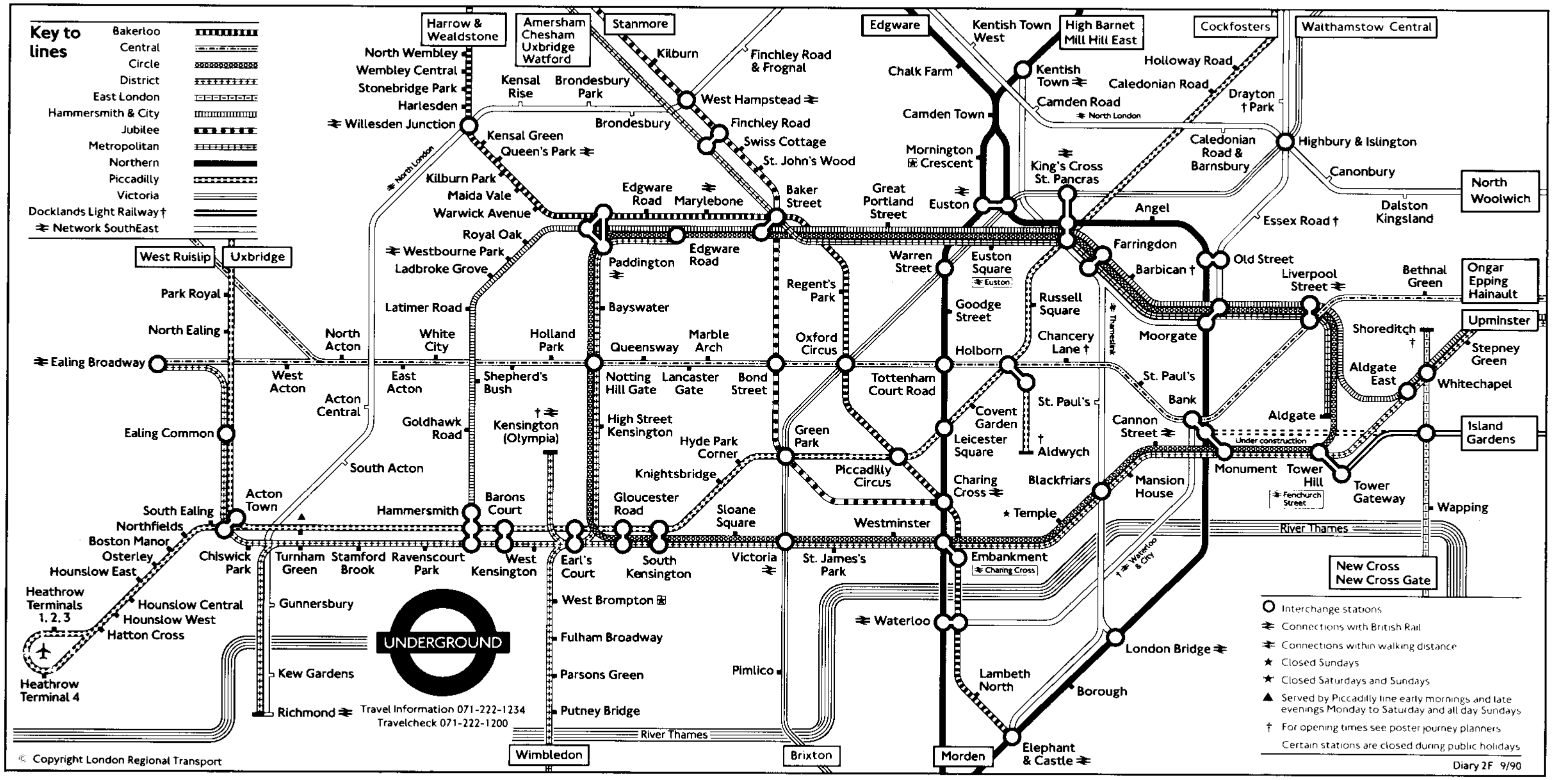

Fig. 9.4 A topological map of the London, England underground (Registered user no. 91/1341).

street name is stored, together with locational coordinates for the start and end nodes of that street. The blocks to the left and right are also stored, in the form of property identifiers for the first and last dwellings in those blocks. This is in contrast to structures adopted elsewhere, notably by Pinpoint Analysis Ltd in Britain. In the late 1980s this company started the task of digitizing all properties in Britain and attaching full Ordnance Survey grid references to them.

Carving up space

When digitizing a map in vector mode it is possible to capture a locational coordinate anywhere on that map if necessary. To any point on the earth's surface may be attached a locational reference of some form, be it a pair of Cartesian coordinates at large scales or latitude and longitude at smaller scales (see Maling 1991 in this volume). In essence, therefore, a continuous view of space can be adopted. Some attributes, notably physical ones such as air temperature and pressure, are, in principle, observable at any location. However, in many applications of GIS, data are collected for areal units and this gives a discrete representation of space. It is necessary to say something about different ways of producing this discretization.

Remote sensing deals with discrete space, the units typically comprising squares or 'pixels' whose dimensions depend upon the available satellite technology. This 'raster' view of the world has been mimicked by some GIS, particularly those which are targeted at microcomputers; IDRISI (Eastman 1988) and pMAP (Berry 1988) are examples. Much has been written about the merits and disadvantages of such raster-based systems, about algorithms for converting from raster to vector and vice versa, about storage of raster data, and so on (see Egenhofer and Herring 1991 in this volume). Suffice it to say here that while a raster-based view of the world hardly matches our daily experience, such systems do offer computational advantages. For instance, apart from grid squares lying at the edge of the map each cell has a fixed number of neighbours and all cells are of constant size and shape. These properties do make many types of spatial analysis much easier. The overlay of maps and the use of various logical operators is very simple, compared with the substantial amounts of processing required in the vector overlay of polygons. Irregular polygonal zones, as used widely for administrative (notably, census) purposes, create numerous problems because of the varying sizes, shapes and neighbours (see Flowerdew 1991 in this volume).

Problems of storing raster data have been much researched. If blocks of pixels cover large areas of homogeneous territory (e.g. 'forest') then, rather than storing data about that property for each pixel it can be stored for the first element and then the number of pixels that follow can be counted, given some ordering scheme for the pixels. This compressed storage is known as 'run-length encoding' (Goodchild and Grandfield 1983). Other researchers have examined different schemes for producing one-dimensional ordering of two-dimensional rasters; some of these are reviewed and illustrated in Goodchild and Mark (1987).

An alternative to such raster schemes is provided by the 'quadtree' (and its equivalent in three dimensions, the 'octree'). Here, a map is divided into four quadrants. Each is then further subdivided until the minimum pixel size is reached (Fig. 9.5). As the illustration makes clear, large areas of contiguous and identical pixels do not require subdivision and may be stored at a higher level of the hierarchical tree. Samet (1984, 1990) has provided a full review of such storage methods. They are not merely of academic interest, since some GIS, notably SPANS, are built around just such data structures.

TYPOLOGY OF SPATIAL OBJECTS

'One of the classic characterizations of data in this field is as 'points, lines and areas. I think that's a pretty primitive model' (Tobler 1988: 51).

Having reviewed different types of space and space-dividing strategies, it is now appropriate to unravel further the definition of space ('a relation defined on a set of objects') and say something about those objects. The discussion has, inevitably, already touched on this subject.

Tobler's stricture addresses the conventional wisdom of classifying spatial objects, as set out in Robinson *et al.* (1984) and Haggett (1965), and used in spatial analysis for many years (Unwin 1981; see also Maguire and Dangermond 1991 in this volume). Points are objects represented by a single pair of locational coordinates. Lines are sets of

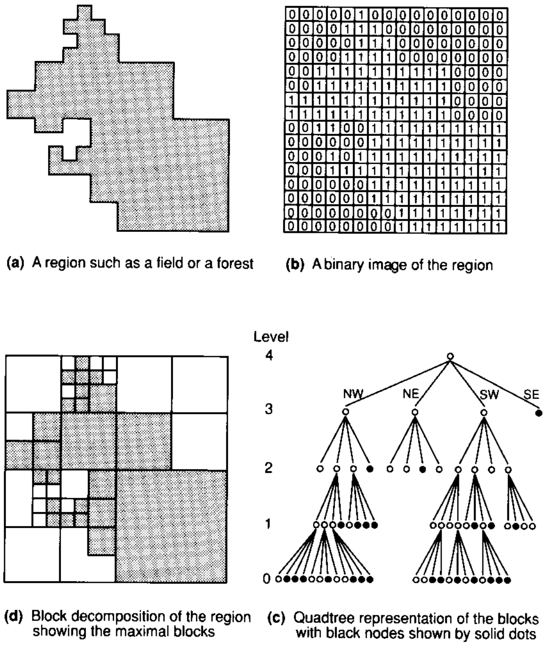

Fig. 9.5 The quadtree data structure (*Source*: Gahegan M, Hogg J (1986) *A pilot GIS based on linear quadtrees and a relational database for regional analysis*. Working Paper 464, Uni. of Leeds, p. 19).

ordered and connected points; they may represent inherently linear features such as roads or streams, or they might denote boundaries. Areas (zones or polygons) are collections of line segments that close to form discrete units. Classically, these objects are thought of as having a length dimension of 0, 1 and 2, respectively (see DCDSTF, 1988, for further definition of such objects). For reasons that will become clear shortly, these are referred to as 'topological dimensions'.

This trilogy covers the set of objects embedded in a space of two dimensions. However, points can exist in three-dimensional space (a small ore-body in a large geological structure, for instance). In three dimensions, lines will no longer necessarily exist in a plane but might trace convoluted paths through volumes. The implication of this is that it is necessary to distinguish between two types of 'dimension': the topological dimension of the object, and the dimension of the space within which it is embedded.

Surfaces have also been considered as a class of spatial object (Unwin 1981), but they are rather different features in that they are defined by data; terrain or temperature surfaces cannot exist without height or temperature data. Within a GIS, surfaces can be represented in a variety of ways. Commonly, their cartographic representation is as a set of contours (1-D objects embedded in a 2-D space). Alternatively, input data used to create a surface after interpolation may be spot-heights (0-D objects in a 2-D space). Finally, a surface may be represented as a Triangulated Irregular Network (TIN), wherein 2-D objects (triangles) are, strictly speaking, embedded in a 3-D space.

Fractals

Until about 20 years ago definitions of the 'topological dimensionality' of objects were not considered problematic. In 1967, however, the French mathematician Mandelbrot published an article that drew upon earlier work by Lewis Richardson and which popularized the concept of a line of fractional dimension (Mandelbrot 1967). Richardson had demonstrated empirically that the measured length of an irregular line depends upon the scale at which it is measured; specifically, its length increases as the scale increases, since more detail is revealed. This can be verified readily by walking a pair of dividers along the line and then repeating this with the dividers set to a narrower width. In contrast, a perfectly straight line has a constant length regardless of scale. If the logarithm of length is plotted against the logarithm of the sampling interval, an approximately linear relationship is observed (Fig. 9.6). For a measured straight line this relationship will appear as a horizontal line. The slope of the line may be denoted as (1-D), suggesting that for a perfectly smooth line on the map D = 1. The more irregular the line on the map (and hence the steeper the slope of the line on the graph) the higher the value of D. At the limit, a line can be imagined so irregular that it in effect 'fills' a piece of paper; such lines have the property D = 2. Since D can take on values between 1 and 2 Mandelbrot referred to it as a measure of fractional dimensionality and gave the name 'fractal' to an irregular line.

The same principles hold for surfaces. A perfectly flat, featureless plain has D = 2 while a solid cube has D = 3. Surfaces of varying

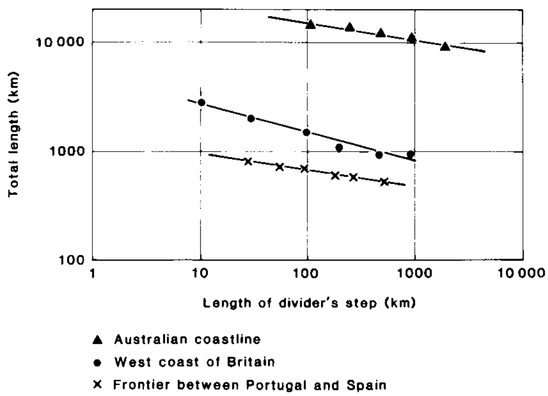

Fig. 9.6 Relationship between length of a line and sampling interval (*Source*: Modified from Maling D (1989) *Measurements from Maps*. Pergamon Press, Cambridge, p. 286).

irregularity have fractional dimension between 2 and 3. It is possible to simulate surfaces of various degrees of roughness (see Goodchild and Mark 1987) and it has been argued that they provide 'norms for the interpretation of geomorphological observations' (Goodchild and Mark 1987: 271).

One of the key properties of fractals is that they are 'self-similar'. This means that any part of the object, when enlarged, is indistinguishable from the object as a whole. Practically, what this means is that an entire coastline, for instance, has the same 'look', and is generated by the same process, as any small part of that coastline. Needless to say, geomorphologists find this hard to believe, since different coastal processes operate at different spatial scales. Despite this caveat, cartographers have found fractals useful, especially in studies of line generalization and line enhancement (see Muller 1991 in this volume).

More broadly, scale determines how to characterize the topological, as much as the fractional, dimensionality of spatial objects. For instance, at a small scale, dealing with data for an entire region or country, settlements can be stored as point objects, with whatever attributes are appropriate. However, at larger scales, of course, such objects must be defined as area objects, the boundaries of which are those of the built-up area. At a yet more detailed scale these area objects are, in fact, collections of buildings (stored at that scale as point objects), while at scales of perhaps 1 : 1500 or greater those buildings are areal objects and the perimeters and areas are stored as attribute fields.

The dimensionality of objects is not as simple as might first be supposed!

Transforming objects

One of the functions that a GIS should support is the ability to create new objects from quite simple or primitive ones. An obvious example is where one area map (e.g. land use) is overlaid with another (e.g. geology) to create, via 'polygon overlay', new areal objects. Several further examples of this are discussed below.

First, the discussion of Euclidean and Manhattan spaces has shown how to create a Thiessen polygon. Thiessen polygons are simply areal objects defined on a set of point objects. All locations within such a polygon lie closer to the point used to define that polygon than to any other such point. Thiessen polygons (also called Voronoi polygons) are useful in the geometric analysis of point patterns (Boots and Getis 1988) and are particularly useful in creating artificial polygons when true boundaries have not been digitized. For example, the primary data collection units in the British Population Census are Enumeration Districts (EDs) and these are not generally available in digital form. However, the census does provide point references for each ED and these may be used to define 'pseudo-EDs' (Fig. 9.7).

This procedure can be examined in reverse. Given a set of areal boundaries is it possible to replace the area definition by a point? This is trivial, in that any point 'seed' can be placed anywhere within an areal unit. However, a GIS is frequently required to define a geographical centre or 'centroid' of such an areal unit (see Monmonier 1982: 116–18 for algorithms) and this is useful if there is a need to measure distances between areas, as is required in location-allocation decisions. One well-known problem is that the centroid can lie outside the areal unit if the latter has a very convoluted shape. In such cases, interactive map editing is required to re-position it.

Surfaces can be regarded as transformations of lower dimensional objects. For instance, a set of spot-heights can be defined and interpolation methods used to create a Digital Terrain Model (DTM); the kind of representation depends upon the density and spacing of sample points. Alternatively, a set of point locations can be used to construct a TIN; algorithms may be used to 'thread'

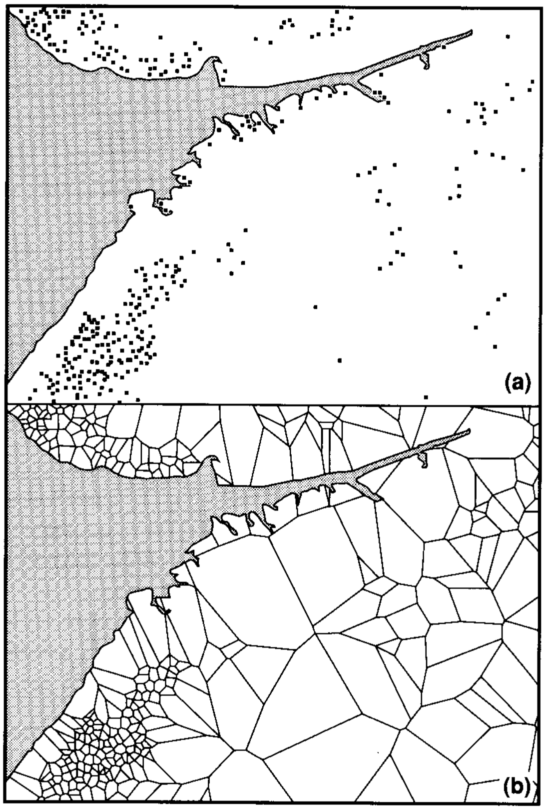

Fig. 9.7 A set of point objects (a) and associated Thiessen polygons (b) for Enumeration Districts in part of Lancashire, England.

contours through the facets of such a network. (Research on DTMs within GIS is discussed by Weibel and Heller 1991 in this volume.)

The creation of complex objects from simple points, lines and areas is an important area of research. It is particularly important in the translation of digital maps into a GIS framework. National mapping agencies such as the Ordnance Survey in Britain (see Sowton 1991 in this volume) and USGS in the United States (see Starr 1991 in this volume) are heavily involved in digital mapping, attaching feature codes to objects such as fences, edges of buildings, and so on. Yet, if there is a need to define land parcels for planning enquiries, methods are required to allow definition of these 'higher order' objects. For example, in the case of the object 'university', at a large scale the digital representation would code buildings, squares, playing fields, car parking areas, and so on, as separate features. There is a need to structure the data in such a way as to collect these objects together to define 'university'. It is to this problem that contemporary research on 'object-oriented' databases is addressed (see Healey 1991 in this volume).

A GIS must, therefore, allow the user to create complex objects from basic building blocks. The above examples show how objects of different dimension are created, but a GIS must also permit the definition of more complex objects of the same dimension. For instance, in the case of a set of line segments representing a street network, a shortest path through that network is not itself a primitive object but may need to be defined as a new object, to which attributes (such as traffic flows) should be attached.

Objects for spatial analysis

The preceding discussion demonstrated that spatial objects have locations and properties (attributes). It is now necessary to say something about the ways in which GIS can assist in analysing the geometry of the locational patterns and in analysing the attribute data. There is no need to dwell on this at length since the subject of spatial analysis is covered elsewhere (see Openshaw 1991 in this volume).

Attribute data may be analysed by statistical methods and models that do not necessarily require reference to location. To compute simple descriptive statistics (e.g. means and variances) does not require knowledge of locations. The methods used for such analysis depend on the level of measurement of the data, be this nominal (categorical), ordinal (ranked) or interval/ratio (continuous). Full discussion of such methods and data types are available in introductory texts (e.g. see Ebdon 1985). Relatively few GIS permit much in the way of statistical analysis and modelling. Efforts to devise transparent links between such systems and conventional statistical packages such as MINITAB, SPSS(X) and GLIM are, therefore, to be encouraged (Kehris 1989).

Equally, the vast battery of techniques used by geographers for many years to describe map pattern have yet to find wide adoption within GIS. For instance, while all GIS handle point objects and some allow the construction of Thiessen polygons, it is hard to find many proprietary systems that allow quantitative analysis and modelling of point data; for instance, the analysis of order neighbour relations or the second-order analysis of point patterns (Diggle 1983).

One important tool in spatial analysis that serves to fuse the geometric and attribute properties of spatial objects is that of spatial autocorrelation. Excellent introductions to this topic are available elsewhere (Goodchild 1986; Odland 1988). Suffice it to say here that spatial autocorrelation refers to the patterning of objects and attributes on the map. For instance, if a variable is measured over a set of areal units and it is observed that high values tend to cluster in one region and low values elsewhere the map is said to exhibit positive spatial autocorrelation. The absence of autocorrelation suggests that attributes are arranged randomly across the map. If high and low values tend to alternate in periodic fashion this is referred to as negative spatial autocorrelation. Thus if there is autocorrelation on the map there is a clear relationship between 'spatial similarity' (i.e. proximity) and attribute similarity (Hubert, Golledge and Costanzo 1981).

Measures of autocorrelation have been devised for point and area objects (and can be adapted for line features also). Measures are available for nominal, ordinal and interval/ratio data. Goodchild (1986) has provided simple program listings to allow implementation of such statistics, while Griffith (1987) shows how to implement them within MINITAB. However, these very useful descriptions of map pattern are still not widely available within proprietary GIS. They should be, since they offer a powerful summary description of the map and a quantitative assessment of whether what is displayed there is any more than would be expected on a chance basis.

Reference has already been made to the concept of 'fuzziness' in spatial data. As has already been mentioned, uncertainty often exists about both location and attribute(s). The attachment of coordinate values to spatial objects often gives a false sense of precision, while there will also frequently be uncertainty associated with attribute data. For example, population data attached to areal units are derived usually from national censuses. There is uncertainty associated with such headcounts, not to mention the fact that, were the data wholly error-free at the time of collection, they are error-laden soon after they are released, since demographic change occurs continuously. There is still a need for research on map and attribute error and how to attach 'uncertainty scores' (however these might be defined!) to such data.

TYPOLOGY OF SPATIAL RELATIONS

Having examined different concepts of space and said something about different types of spatial object and spatial data the discussion now examines more closely some alternative definitions of spatial relationship. To date, emphasis has been on spatial relations, defined largely in terms of some measure of physical distance, although this assumption was relaxed in the brief examination of non-metric spaces. Here, other ways of conceiving spatial proximity are investigated.

Before doing so, recall that space is defined as a relation on a set of objects. This may be a relation existing between one set and another set of objects. (To give a simple example, this might be a 'preference relation' between a set of GIS users and a set of software products!) This is potentially a very powerful way of conceptualizing GIS functions. Figure 9.8 shows in schematic form how some operations in GIS (only a handful are illustrated) can be regarded as relations between different sets of spatial object (Unwin 1981; Goodchild 1985). For instance, 'point-in-polygon' operations relate points to areas. A study of rail routes transporting hazardous cargoes across particular areal units uses a line–area relation. Polygon overlay is an area–area relation, and so on. For simplicity, only half of this table is shown suggesting that it is symmetrical. As a generalization this is an adequate representation, although detailed examination shows that there are differences between some relations. For example, determining if a point is in an area (area-point) is different from creating a set of areas (Thiessen polygons) from a set of points (point-areas).

	Points	Lines	Areas
Points	• is a neighbour of • is allocated to	• is near to • lies on	• is a centroid of • is within
Lines		• crosses • joins	• intersects • is a boundary of
Areas			• is overlain by • is adjacent to

Fig. 9.8 Relations between classes of spatial object.

It is clear, then, that there is a whole variety of 'spatial relations'. In this section, however, attention is restricted to point–point relations: Euclidean and Manhattan distances were two earlier examples.

In our everyday worlds, rather than the artificial worlds constructed in GIS, spatial separation is experienced less in terms of physical distance and more in terms of time, cost, or 'perception' (Gatrell 1983). What use has been made of these other ways of regarding proximity or separation within a GIS framework? The simple answer is – not much!

Time-distance

Spatial separation, in terms of travel time, can be examined at a variety of spatial scales: global, national or local. A set of spatial objects may be defined, typically point locations such as towns at a national scale or nodes on a road network at a local scale, and a matrix of travel times between all pairs of places constructed. It is possible to look at either a single row or column of the matrix, or to analyse the structure of the entire matrix. But before carrying this out it should be made clear that there is no single relation defined on any such set of places. While physical distances remain constant, travel times themselves are time dependent, varying on an hourly basis within urban areas and perhaps over longer time scales globally (as improvements in transport technology work their way through the system). Moreover, the relation certainly does not define a metric space; time–distances will not, in general, be symmetrical, nor will the triangle inequality necessarily hold.

Notwithstanding these problems there is still scope for exploring the usefulness of such relations in GIS. Consider an application in which there is a requirement to describe and display the travel times from one centre or node on a network to all other locations in a spatial database. There is a need for GIS that will interpolate between such locations to produce an 'isochrone map', comprising contours which represent lines of equal travel time (Fig. 9.9a). The scale of such a map is still in physical distance units, however, and there should be the ability within GIS to transform or distort the underlying base map so that the isochrones become concentric circles (Fig. 9.9b) and the geometry of the transport network (and other spatial objects on the map) is modified (Muller 1978; Clark 1977). Some software systems currently available do indeed offer the option of defining catchment or trade areas based on travel time. For purposes of visualization it would be useful, however, to have more capabilities for producing distorted maps (what Muehrcke, 1978, has called 'linear cartograms').

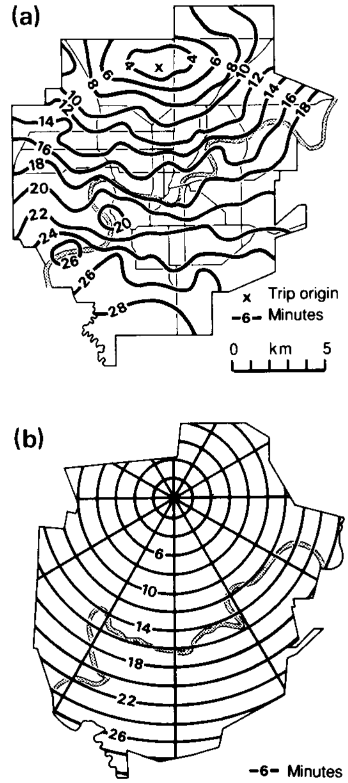

Fig. 9.9 Time–distance in Edmonton, Canada: (a) isochrones in geographical space; (b) isochrones in time–space.

It is very difficult to use all the information contained in the travel time matrix, rather than only one row or column, because the relation is not metric. (Given a matrix of physical distances between a set of places, it is quite feasible, using trilateration methods from surveying, to produce a map of correct locations; e.g. see Tobler 1982.) Fortunately, methods are available for producing a metric space from non-metric data; such procedures go under the general name of 'multi-dimensional scaling' (MDS).

There is insufficient space here to deal at length with such map transformation methods; a full treatment is available in Gatrell (1983). Suffice it to say that MDS seeks to provide a spatial configuration of a set of objects such that the distances between objects in that configuration match as closely as possible the input data (such as travel times). In principle, this configuration or 'map' might have three or more dimensions; in practice, two will often suffice for the kinds of application geographers deal with. Many algorithms are available to perform MDS, the most widely known being ALSCAL, part of the SPSS(X) package. However, no attempt has been made to marry this method with GIS software.

Given the importance of travel time in human behaviour, in assessing planning needs, and so on, it seems worth contemplating such a marriage, particularly in view of the importance attached to visualization. As an example, consider planning proposals for a bypass in the town of State College, Pennsylvania (Fig. 9.10a). In this study (Weir 1975), 42 nodes on the road network were defined and peak-hour travel times estimated among all pairs. MDS was then used to generate from this matrix a configuration of nodes in 'time–space' (Fig. 9.10b). Various options for the bypass were considered, each producing a different travel time matrix. One such option, for a northern bypass, generated a configuration that may be compared with the *status quo* (using rigorous methods for performing overlays of point patterns). The visual evidence for the impact of such a proposal is striking, revealing how locations in the north-west are pulled east, while those in the south-east are pulled north. This kind of approach allows transport planners to see the potential impacts of alternative proposals. There is really no reason why such methods could not be adopted within GIS, with other contextual map information being added if required.

Other measures of spatial proximity

Muller (1982, 1984) has modified the MDS approach to map transformations. He seeks configurations of points such that the distances in the configuration match the input data (travel time, cost, etc.), but depart from the MDS model by allowing more general definitions of distance in the transformed space. His algorithm fits models to the input data of the form:

$$\mathrm{d}(i,j) = (a|\mathrm{d}x|^p + b|\mathrm{d}y|^p)^r \qquad [9.4]$$

where dx and dy are differences in the x and y directions and a, b, p and r are parameters to be estimated. (Compare this with the formula for Euclidean distance, where a and b are 1, $p = 2$ and $r = 1/2$). For example, looking at the costs of sending parcels among a set of 14 Canadian cities, he obtains:

$$\mathrm{d}(i,j) = (4.33\mathrm{d}x^{0.56} + 3.62\mathrm{d}y^{0.56})^{0.39} \qquad [9.5]$$

Note that as $p < 1$ the space is a non-metric one. Muller shows how to represent such functions graphically, along with the resulting configurations. Such transformed maps do offer the possibilities of new visual perspectives, though whether they will ever become standard in proprietary GIS products is debatable.

All this work can be regarded as 'vector based', inasmuch as (x,y) locations of places are sought, albeit in transformed spaces. In the case of raster-based approaches, the obvious measure of spatial separation is based on Pythagorean distance. However, Berry (1988) has shown how to generalize this to incorporate, within a PC-based GIS (pMAP), more sophisticated measures of separation. For example, weights can be assigned to particular grid squares, according to how easy it is to traverse particular types of terrain (woodland versus open land, for instance). These weights can be based on notional time or cost estimates. Berry also shows how barriers can be placed between grid cells, either restricting movement or denying it altogether. Indeed, any GIS should be capable of allowing the user to attach weights or 'impedances' on to particular links. These might reflect different classes or capacities of road, the existence of one-way streets, and so on. Such facilities greatly enhance the usefulness of such systems in network modelling.

One important aspect of such network modelling is, of course, to deal with flows through the network. There is a vast literature on spatial interaction modelling (see Senior 1979 for an introduction), which does not, as yet, seem to have found its way into many proprietary GIS. There is insufficient space to deal with such models here, but they are mentioned because there is a very real sense in which the volume of such interaction serves as a measure of 'proximity'. Two places, between which considerable movement of traffic takes place, are 'closer' in an 'interaction space' than are places between which little interaction occurs.

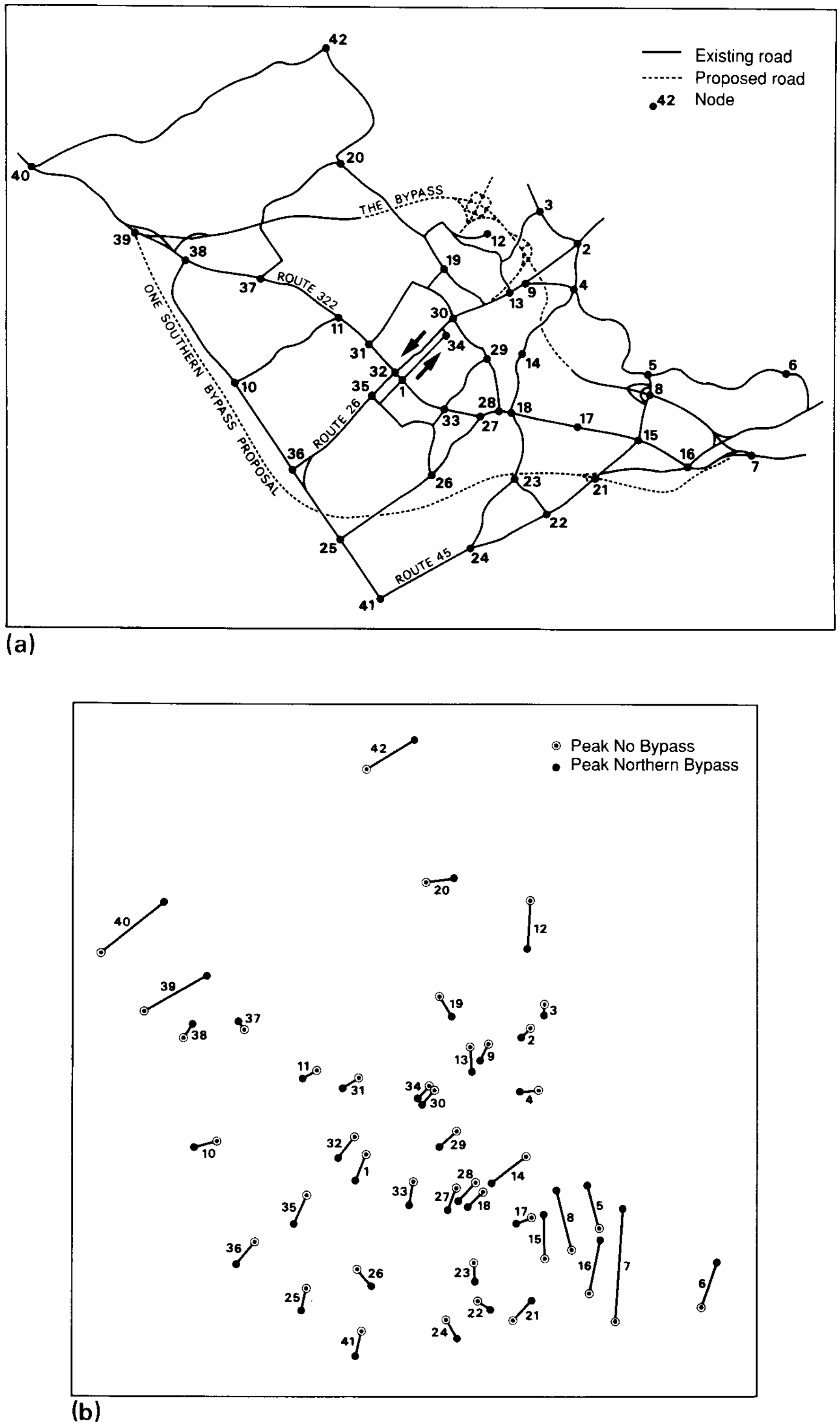

Fig. 9.10 State College, Pennsylvania, USA in (a) geographical space and (b) time-space.

Cartograms

Reference has been made earlier to the importance of visualization in GIS. Using MDS, it is possible to manipulate data on spatial relationships to produce transformed maps. Production of such transformations using area objects with associated map attributes is now considered. The maps that are produced are known as 'area cartograms' (Muehrcke 1978) or 'cartograms' for short.

Consider a map of local government administrative units ('wards') in Lancaster, England (Fig. 9.11a), the boundaries of which are in digital form in a GIS. Given data for each areal unit on total population, it is possible to re-draw the map so that the new areas are in proportion to population size and contiguity and adjacency maintained. It is possible, though tedious, to produce a graphical solution (in effect, numerous possible solutions) by hand, using nothing more than graph paper (e.g. see Monmonier 1977). However, efforts have been made over many years to automate such map production (Tobler 1973; Dougenik, Chrisman and Niemeyer 1985; Selvin, Merrill and Sacks 1988).

Dougenik *et al.* have suggested one method, which was employed to produce a transformed map of Lancaster (Fig. 9.11b). Their method operates on all points comprising the boundaries of areas, computing a vector displacement for each point and then checking to see if the new areal units correspond better to the 'target' data (such as the population count for each area). The displacement is a function of the difference between actual and 'desired' area, together with the distance from that point to the centroid of the polygon. The procedure is then repeated in an iterative fashion. However, the method can give rise to topological problems and checks for these must be built in.

One issue that has not been given sufficient attention in cartogram production has been the need to devise ways of associating point objects with the transformed map. For instance, data on the locations of disease cases may be available for display on a population cartogram. Selvin *et al.* (1988) have offered a solution to this problem, by devising an alternative algorithm for cartogram generation.

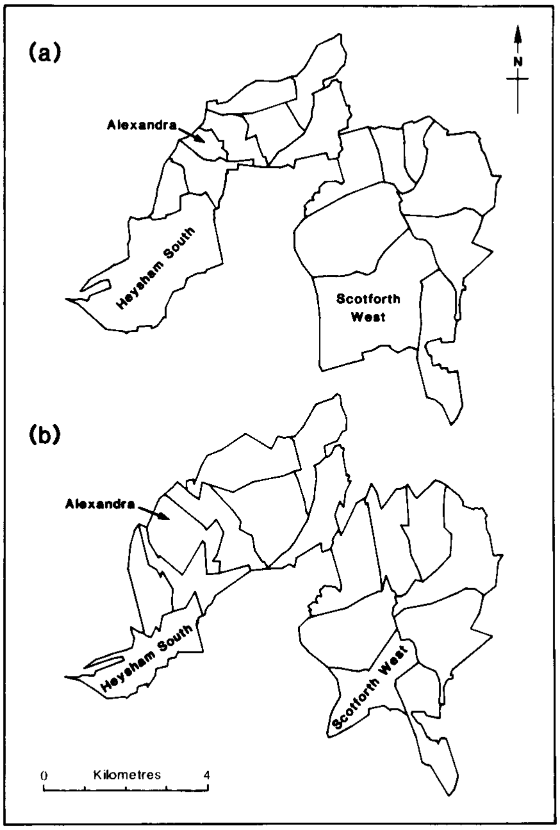

Fig. 9.11 Electoral wards in Lancaster, England urban area in (a) geographical space and (b) as a cartogram.

CONCLUSIONS

It should be clear from the above that space is not as simple as might first be imagined. Different 'spaces' arise as the definition of the relation on a set of objects changes. While Euclidean distance has, quite properly, long been accorded a primary place within GIS, there is scope for encouraging more imaginative and flexible approaches, whereby other measures of spatial separation are incorporated. Of particular value is likely to be the greater use of travel time, both as a source of data for analysis and modelling, and as a means of visualizing spatial relationships. Visualization in GIS would also profit from the wider incorporation of options to produce cartograms. These are particularly valuable in epidemiological work.

These new ways of looking at space and spatial relations within a GIS context are unlikely to prove of value to those more interested in land management and in digital mapping, in which a Euclidean perspective rightly reigns supreme. For those interested in the analytical side of GIS, or for those seeking novel descriptions, however, such views may be appealing and offer fresh insights into geographical patterns; this should be the ultimate aim of research that uses GIS technology.

REFERENCES

Abler R, Adams J, Gould P R (1971) *Spatial Organization: the geographer's view of the world.* Prentice-Hall, Englewood Cliffs New Jersey

Armstrong M P (1988) Distance imprecision and error in spatial decision support systems. In: Aangeenbrug R T, Schiffman Y M (eds.) *International Geographic Information Systems (IGIS) Symposium, Arlington, Virginia*, Vol 2. NASA, Washington DC, pp. 23–34

Berry J K (1988) Computer-based map analysis: characterizing proximity and connectivity. In: Aangeenbrug R T, Schiffman Y M (eds.) *International Geographic Information Systems (IGIS) Symposium, Arlington, Virginia*, Vol 2. NASA, Washington DC, pp. 11–22

Boots B, Getis A (1988) Point pattern analysis. *Sage Scientific Geography Series, Number 8.* Sage Publications, London

Broome F (1986) Mapping from a topologically encoded database: the US Bureau of the Census example. In: Blakemore M J (ed.) *Proceedings of AUTOCARTO London.* Royal Institution of Chartered Surveyors, London, pp. 402–11

Burrough P A (1989) Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science* **40**: 477–92

Chrisman N R (1991) The error component in spatial data. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications.* Longman, London, pp. 165–74, Vol 1

Clark J W (1977) Time–distance transformations of transportation networks. *Geographical Analysis* **9**: 195–205

DCDSTF (Digital Cartographic Data Standards Task Force) (1988) The proposed standard for digital cartographic data. Volume 1. *The American Cartographer* **15**: 9–140

Diggle P J (1983) *Statistical Analysis of Spatial Point Patterns.* Academic Press, London

Dougenik J A, Chrisman N R, Niemeyer D R (1985) An algorithm to construct continuous area cartograms. *Professional Geographer* **37**: 75–81

Dunn C, Newton D (1989) Notes on shortest path algorithms for GIS. *North West Regional Research Laboratory Research Report 3.* North West Regional Research Laboratory, Lancaster

Eastman R (1988) *IDRISI: a grid-based geographic analysis system.* Graduate School of Geography Clark University, Massachusetts

Ebdon D (1985) *Statistics in Geography: a practical approach*, 2nd edn. Basil Blackwell, Oxford

Egenhofer M J, Herring J R (1991) High-level spatial data structures for GIS. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications.* Longman, London, pp. 227–37, Vol 1

Flowerdew R (1991) Spatial data integration In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications.* Longman, London, pp. 375–87, Vol 1

Frank A U, Mark D M (1991) Language issues for GIS. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications.* Longman, London, pp. 147–63, Vol 1

Gatrell A C (1983) *Distance and Space: a geographical perspective.* Oxford University Press, Oxford

Goodchild M F (1985) Geographical information systems in undergraduate geography: a contemporary dilemma. *The Operational Geographer* **8**: 34–8

Goodchild M F (1986) Spatial autocorrelation. *CATMOG 47* GeoAbstracts, Norwich

Goodchild M F (1988) Towards an enumeration and classification of GIS functions. In: Aangeenbrug R T, Schiffman Y M (eds.) *International Geographic Information Systems (IGIS) Symposium, Arlington, Virginia*, Vol 2. NASA, Washington DC, pp. 65–77

Goodchild M F, Grandfield A W (1983) Optimizing raster storage: an evaluation of four alternatives. In: *Proceedings of AUTOCARTO6 (2).* ASPRS, Falls Church, pp. 400–7

Goodchild M F, Mark D M (1987) The fractal nature of geographic phenomena. *Annals of the Association of American Geographers* **77**: 265–78

Griffith D A (1987) Spatial autocorrelation: a primer. *Association of American Geographers, Resource Publications in Geography.* Association of American Geographers, Washington

Haggett P (1965) *Lotational Analysis in Human Geology.* Edward Arnold, London

Healey R G (1991) Database management systems. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications.* Longman, London, pp. 251–67, Vol 1

Hubert L, Golledge R G, Costanzo C M (1981) Generalized procedures for evaluating spatial autocorrelation. *Geographical Analysis* **13**: 224–33

Kehris E (1989) Interfacing ARC/INFO with GLIM: a progress report. *North West Regional Research Laboratory Research Report 5.* North West Regional Research Laboratory, Lancaster

Maguire D J, Dangermond J (1991) The functionality of GIS. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications.* Longman, London, pp. 319–35, Vol 1

Maguire D J, Worboys M F, Hearnshaw H M (1990) An introduction to object-oriented Geographical Information Systems. *Mapping Awareness* **4** (2): 36–9

Mahoney R P (1991) GIS and utilities. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications*. Longman, London, pp. 101–14, Vol 2
Maling D H (1991) Coordinate systems and map projections for GIS. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications*. Longman, London, pp. 135–46, Vol 1
Mandelbrot B (1967) How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* **156**: 636–8
Mark D M, Svorou S, Zubin D (1988) Spatial terms and spatial concepts: geographic, cognitive, and linguistic perspectives. In: Aangeenbrug R T, Schiffman Y M (eds.) *International Geographic Information Systems (IGIS) Symposium, Arlington, Virginia*. NASA, Washington DC, pp. 101–12
Monmonier M S (1977) Maps, distortion and meaning. *Association of American Geographers Resource Paper in Geography*, **75–4**. Association of American Geographers, Washington
Monmonier M S (1982) *Computer-Assisted Cartography: principles and prospects*. Prentice-Hall, Englewood Cliffs New Jersey
Muehrcke P (1978) *Map Use: reading, analysis and interpretation*. J P Publications, Madison Wisconsin
Muller J-C (1978) The mapping of travel time in Edmonton, Alberta. *Canadian Geographer* **22**: 195–210
Muller J-C (1982) Non-Euclidean geographic spaces: mapping functional distances. *Geographical Analysis* **14**: 189–203
Muller J-C (1984) Canada's elastic space: a portrayal of route and cost distances. *Canadian Geographer* **28**: 46–62
Muller J-C (1991) Generalization of spatial databases. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications*. Longman, London, pp. 457–75, Vol 1

Odland J (1988) Spatial autocorrelation. *Sage Scientific Geography Series Number 9*. Sage Publications, London
Openshaw S (1991) Developing appropriate spatial analysis methods for GIS. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications*. Longman, London, pp. 389–402, Vol 1

Peuquet D J (1988) Representations of geographic space: toward a conceptual synthesis. *Annals of the Association of American Geographers* **78**: 375–94

Robinson A H, Sale R D, Morrison J L, Muehrcke P C (1984) *Elements of Cartography*, 2nd edn. John Wiley, Chichester
Robinson V B, Miller R, Klesh L (1988) Issues in the use of expert systems to manage uncertainty in geographic information systems. In: Aangeenbrug R T, Schiffman Y M (eds.) *International Geographic Information Systems (IGIS) Symposium, Arlington, Virginia*, Vol 2. NASA, Washington DC, pp. 89–100

Samet H (1984) The quadtree and related hierarchical data structures. *ACM Computing Surveys* **16**: 187–260
Samet H (1990) *The Design and Analysis of Spatial Data Structures*. Addison-Wesley, Reading Massachusetts
Selvin S, Merrill D W, Sacks S (1988) Transformations of maps to investigate clusters of disease. *Social Science and Medicine* **26**: 215–21
Senior M L (1979) From gravity modelling to entropy maximizing: a pedagogic guide. *Progress in Human Geography* **3**: 179–210
Smith T R, Menon S, Star J L, Estes J E (1987) Requirements and principles for the implementation and construction of large-scale geographical information systems. *International Journal of Geographical Information Systems* **1**: 13–32
Sowton M (1991) Development of GIS-related activities at the Ordnance Survey. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications*. Longman, London, pp. 23–38, Vol 2
Starr L E, Anderson K E (1991) A USGS perspective on GIS. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications*. Longman, London, pp. 11–22, Vol 2

Tobler W R (1973) A continuous transformation useful for redistricting. *Annals of the New York Academy of Sciences* **219**: 215–20
Tobler W R (1982) Surveying multidimensional measurement. In: Golledge R G, Rayner J N (eds.) *Proximity and Preference: problems in the multidimensional analysis of large data sets*. University of Minnesota Press, Minneapolis, pp. 3–4
Tobler W R (1988) Geographic information systems research agenda: the scientific community perspective. In: Aangeenbrug R T, Schiffman Y M (eds.) *International Geographic Information Systems (IGIS) Symposium, Arlington, Virginia*, Vol 1. NASA, Washington DC, pp. 49–52

Unwin D J (1981) *Introductory Spatial Analysis*. Methuen, London

von Hohenbalken B and West D S (1984) Manhattan versus Euclid: market areas computed and compared. *Regional Science and Urban Economics* **14**: 19–35

Weibel R, Heller M (1991) Digital terrain modelling. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications*. Longman, London, pp. 269–97, Vol 1
Weir S (1975) *Getting around town: modifications in a local travel time space caused by expressway construction*. Unpublished MSc thesis. Department of Geography, Pennsylvania State University
White M (1991) Car navigation systems. In: Maguire D J, Goodchild M F, Rhind D W (eds.) *Geographical Information Systems: principles and applications*. Longman, London, pp. 115–25, Vol 2