2 The Nature of Geographic Data

LEARNING OBJECTIVES

This chapter elaborates on the *spatial is special* theme by examining the nature of geographic data. It sets out the distinguishing characteristics of geographic data and suggests a range of guiding principles for working with them. Many geographic data are correctly thought of as sample observations, selected from the larger universe of possible observations that could be made. This chapter describes the main principles that govern scientific sampling and the principles that are invoked to infer information about the gaps between samples. It also discusses the relevance of these principles when using crowd-sourced data and other Big Data sources. When devising spatial sample designs, it is important to be aware of the nature of spatial variation, and here we learn how this is formalized and measured as spatial autocorrelation. Another key property of geographic information is the level of detail that is apparent at particular scales of analysis. The concept of fractals provides a solid theoretical foundation for understanding scale when building geographic representations.

After studying this chapter you will understand:

- How Tobler's First Law of Geography is formalized through the concept of spatial autocorrelation.
- The relationship between scale and the level of geographic detail in a representation.
- The principles of building representations around geographic samples.
- How the properties of smoothness and continuous variation can be used to characterize geographic data.
- How fractals can be used to measure and simulate apparently irregular geographic phenomena.

2.1 Introduction

In Chapter 1 we identified the central motivation for geographic information (GI) science as the development of representations, not only of how the world *looks*, but also how it *works*. In this chapter we develop a fuller understanding of the ways in which we think about the *nature of spatial variation*. We do this by asserting three principles:

1. Proximity effects are key to representing and understanding spatial variation and to joining incomplete representations of unique places.
2. Issues of geographic scale and level of detail are key to building appropriate representations of the world.
3. Different measures of the world *covary*, and understanding the nature of covariation can help us to predict.

Implicit in all this is one further principle that we will develop in Chapters 3 and 5: because we rarely if ever have complete information about the world, we must selectively sample from it and relate what we know to what we are less sure about. GI science is about representing spatial and temporal phenomena in the observable world, and because the observable world is complicated, this task is difficult, error prone, and in the terminology of Chapter 5, *uncertain*. The observable world provides an intriguing laboratory in which to examine phenomena, but is one in which it can be impossible to control for variation in all characteristics—be they relevant to landscape evolution, consumer behavior, urban growth, or whatever. As a consequence, in the terminology of Section 1.3, generalized *laws* governing spatial distributions and temporal dynamics are therefore most unlikely to work perfectly.

We choose to describe the three points listed above as principles rather than laws because, like most applications of GI science, this chapter is grounded in empirical generalization about the real world. A more elevated discussion of the way that these principles build into fundamental laws of GI science has been published by Goodchild.

2.2 The Fundamental Problem

It is instructive to reflect for a moment on how we might describe our lives to date to a friend or colleague. A human lifetime is infinitesimally small compared with the geographic extent and history of the world, but as we move to finer spatial and temporal scales, it nevertheless remains very intricate in detail. Viewed in aggregate, human behavior in space appears structured when we aggregate the outcomes of day-to-day (often repetitive) decisions about where to go, what to do, or how much time to spend doing it. Over the longer term, structure also arises out of (one-off) decisions about where to live, how to achieve career objectives, and how to balance work, leisure, and family pursuits. But just as giving clear narrative accounts of our lives requires clear recall of the key decisions that have very much made us what we are, so the GI scientist must be competent in discarding (or not feeling troubled to measure) the inessentials while retaining the salient characteristics of the observable world.

GI scientists distinguish between *controlled* and *uncontrolled* variation over time. Controlled variation in our lives oscillates around a steady state (daily, weekly) pattern, whereas uncontrolled variation (career changes, residential moves) does not. The same might be said, respectively, of seasonal change in climate and the phenomenon of global warming. When relating our own daily regimens and life histories, or indeed any short- or long-term *time series* of events, we are usually mindful of the contexts in which our decisions (to go to work, to change jobs, to marry) are made; "the past is the key to the present" aptly summarizes the effect of temporal context on our actions. The day-to-day operational context to our activities is very much determined by where we live and work. The longer-term strategic context may well be provided by where we were born, grew up, or went to college.

Our behavior in geographic space often reflects past patterns of behavior.

The relationship between consecutive events in *time* can be formalized in the concept of *temporal autocorrelation*. The analysis of time series data is in some senses straightforward because the direction of causality is only one way: past events are sequentially related to the present and to the future. This chapter (and book) is principally concerned with spatial, rather than temporal, autocorrelation. Spatial autocorrelation shares some similarities with its temporal counterpart. Yet time moves in one direction only (forward), making temporal autocorrelation one dimensional, whereas spatial events can potentially have consequences anywhere in two-dimensional or even three-dimensional space.

Explanation in time need only look to the past, but explanation in space must look in all directions simultaneously.

Just as the expectation is that temporal autocorrelation will be strongest between events that happen at about the same time as one another, so we expect spatial autocorrelation to be stronger for occurrences that are located close to one another. This guiding principle is often elevated to the status of a First Law of Geography, attributed to Waldo Tobler that *all places are similar, but nearby places are more similar than distant places*.

Tobler's First Law of Geography: Everything is related to everything else, but near things are more related than distant things.

Assessment of spatial autocorrelation can be informed by knowledge of the degree and nature of *spatial heterogeneity*—the tendency of geographic places and regions to be different from each other. Everyone would recognize the extreme difference of landscapes between such regions as the Antarctic, the Nile Delta, the Sahara Desert, or the Amazon Basin, and many would recognize the more subtle differences between the Central Valley of California, the Northern Plain of China, and the valley of the Ganges in India. As with change over time, variation in space may be controlled or uncontrolled: the spatial variation in some processes simply oscillates about an average (controlled variation), whereas other processes vary ever more the longer they are observed (uncontrolled variation). For example, controlled variation characterizes the operational environment of GI applications in utility management, or the tactical environment of retail promotions, whereas longer-term processes such as global warming or deforestation may lead to uncontrolled variation.

As a general rule, spatial data exhibit an increasing range of values, hence increased heterogeneity, with increased distance. In this chapter we focus on the ways in which phenomena vary across space and on the general nature of geographic variation. Later, in Section 13.2.1, we will return to the techniques for measuring spatial heterogeneity. This requires us to move beyond thinking of GI as abstracted only from the continuous spatial distributions implied by Tobler's Law and from sequences of events over continuous time. Some events, such as the daily rhythm of the journey to work, are clearly incremental extensions of past practice, whereas others, such as residential relocation, constitute sudden breaks with the past. Similarly, landscapes of gently undulating terrain are best thought of as smooth and continuous, whereas others (such as the landscapes developed about fault systems or mountain ranges) are best conceived as discretely bounded, jagged, and irregular. Smoothness and irregularity turn out to be among the most important distinguishing characteristics of geographic data.

Some geographic phenomena vary smoothly across space, whereas others can exhibit extreme irregularity, in violation of Tobler's Law.

It is highly likely that a representation of the real world that is suitable for predicting future change will need to incorporate information on how two or more factors *covary*. For example, planners seeking to justify improvements to a city's public transit system might wish to point out how house prices increase with proximity to existing rail stops. It is highly likely that patterns of spatial autocorrelation in one variable will, to a greater or lesser extent, be mirrored in another. However, although this is helpful in building representations of the real world, we will see in Section 14.5.1 that the property of spatial autocorrelation can frustrate our attempts to build inferential statistical models of the covariation of geographic phenomena.

Spatial autocorrelation helps us to build representations but frustrates our efforts to predict.

The nature of geographic variation, the scale at which uncontrolled variation occurs, and the way in which different geographic phenomena covary all help us to understand the form and functioning of the real world. These principles are of practical importance and guide us toward answering questions such as: What is an appropriate scale or level of detail at which I should measure the relevant and observable characteristics of the world? How do I design my spatial sample? How do I generalize from my sample measurements? And what formal methods and techniques can I use to relate key spatial events and outcomes to one another?

Each of these questions is a facet of the fundamental problem of the analysis of GI, that is, of selecting what to measure and record, using the scales of measurement described in Box 2.1. The Tobler Law

Technical Box 2.1

Types of Attributes

The simplest type of attribute, termed *nominal*, is one that serves only to identify or distinguish one entity from another. Place-names are a good example, as are names of houses, or the numbers on a driver's license—each serves only to identify the particular instance of a class of entities and to distinguish it from other members of the same class. Nominal attributes include numbers, letters, and even colors. Even though a nominal attribute can be numeric, it makes no sense to apply arithmetic operations to it: adding two nominal attributes, such as two drivers' license numbers, creates nonsense.

Attributes are *ordinal* if their values have a natural order. For example, Canada rates its agricultural land by classes of soil quality, with Class 1 being the best, Class 2 not so good, and so on. Adding or taking the ratios of such numbers makes little sense because 2 is not twice as much of anything as 1, but at least ordinal attributes have inherent order. Averaging makes no sense either, but the *median*, or the value such that half of the attributes are higher ranked and half are lower ranked, is an effective substitute for the average for ordinal data as it gives a useful central value.

Attributes are *interval* if the differences between values make sense. The scale of Celsius temperature is interval because it makes sense to say that 30 and 20 are as different as 20 and 10.

Attributes are *ratio* if the ratios between values make sense. Weight is ratio because it makes sense to say that a person of 100 kg is twice as heavy as a person of 50 kg, but Celsius temperature is only interval because 20 is not twice as hot as 10 (and this argument applies to all scales that are based on arbitrary zero points, including longitude).

It is sometimes necessary to deal with GI that fall into categories beyond these four. For example, data can be directional or *cyclic*, including flow direction on a map, or compass direction, or longitude, or month of the year. The special problem here is that the number following 359 degrees is 0. Averaging two directions such as 359 and 1 yields 180, so the average of two directions close to north can appear to be south. Because the cyclic case sometimes occurs in GI, and few designers of GI software have made special arrangements for them, it is important to be alert to the problems that may arise.

amounts to a succinct definition of spatial autocorrelation. A prior understanding of the nature of the spatial autocorrelation that characterizes a GI application helps us *deduce* how best to collect and assemble data for a representation and also how best to develop inferences between events and occurrences. The concept of geographic *scale* or level of detail will be fundamental to observed measures of the likely strength and nature of autocorrelation in any given application. Together, the scale and spatial structure of a particular application suggest ways in which we should *sample* geographic reality and should *weight* sample observations to build our representation. We will return to the key concepts of scale, sampling, and weighting throughout much of this book and further discuss how representations are conceived and built from samples in Chapter 3. If data are collected for one purpose and reused for another, secondary,

Technical Box 2.2

Types of Spatial Objects

Geographic objects are classified according to their *topological dimension*, which provides a measure of the way they fill space. For present purposes we assume that these dimensions are restricted to *integer* (whole number) values, though later (Section 2.8) we relax this constraint and consider geographic objects of noninteger (fractional, or *fractal*) dimension.

Geometric objects can be used to represent occurrences of recognizable features in space (sometimes described as *natural* objects), or they may be used to summarize spatial distributions of recognizable features (creating what are sometimes described as *artificial* objects).

A *point* has neither length nor breadth nor depth, and hence it is said to be of dimension 0. Points may be used to indicate spatial occurrences or events and their spatial patterning. *Point pattern analysis* is used to identify whether occurrences or events are interrelated—as in analyzing the incidence of crime or in identifying whether patterns of disease infection might be related to environmental or social factors (see Section 13.3.3). The *centroid* of an area object is an artificial point reference, which is located to provide a summary measure of the location of the object (see Section 14.2.1).

Lines have length, but not breadth or depth and hence are of dimension 1. They are used to represent linear entities such as roads, pipelines, and cables, which frequently build together into networks. They can also be used to measure distances between spatial objects, as in the measurement of intercentroid distance. To reduce the burden of data capture and storage, lines are often held in GI systems in *generalized* form (see Section 3.8).

Area objects have the two dimensions of length and breadth, but not depth. They may be used to represent natural objects, such as agricultural fields, but are also commonly used to represent artificial aggregations, such as census tracts (see below). Areas may bound linear features and enclose points, and GI systems can be used to identify whether a given area encloses a given point (Section 13.2.3).

Volume objects have length, breadth, and depth and hence are of dimension 3. They are used to represent natural objects such as river basins or artificial phenomena such as the population potential of shopping centers or the density of resident populations (Section 13.3.5).

Time is often considered to be the fourth dimension of spatial objects, although GI science remains poorly adapted to the modeling of temporal change.

Lower-dimension objects can be derived from those of higher dimension but not vice versa.

Certain phenomena may be represented in a GI database as either natural or artificial spatial objects—a human individual may be represented as a point or as part of a population represented as living within a census tract, for example. As will be discussed in Chapter 3, the chosen way of representing objects in space defines not only the apparent nature of geographic variation, but also the ways in which geographic variation may be analyzed. Some objects, such as agricultural fields or digita zero l terrain models, are represented in their natural, observable state. Others are transformed from one spatial object class to another, which is what happens if population data are aggregated from individual points to census tract areas for reasons of confidentiality, for example. Some high-order representations are created by interpolation between lower-order objects, as in the creation of digital elevation models (DEMs) from spot height data (see Section 13.3.6).

The classification of spatial phenomena into object types is dependent fundamentally on scale. For example, on a less-detailed map of the world, New York is represented as a zero-dimensional point. On a more detailed map such as a road atlas, it will be represented as a two-dimensional area. Yet if we visit the city, it is very much experienced as a three-dimensional entity, and virtual reality systems seek to represent it as such (see Section 10.3.1).

purpose, it is important that the scale, sampling, and weighting of the data are appropriate to the secondary application.

2.3 Spatial Autocorrelation and Scale

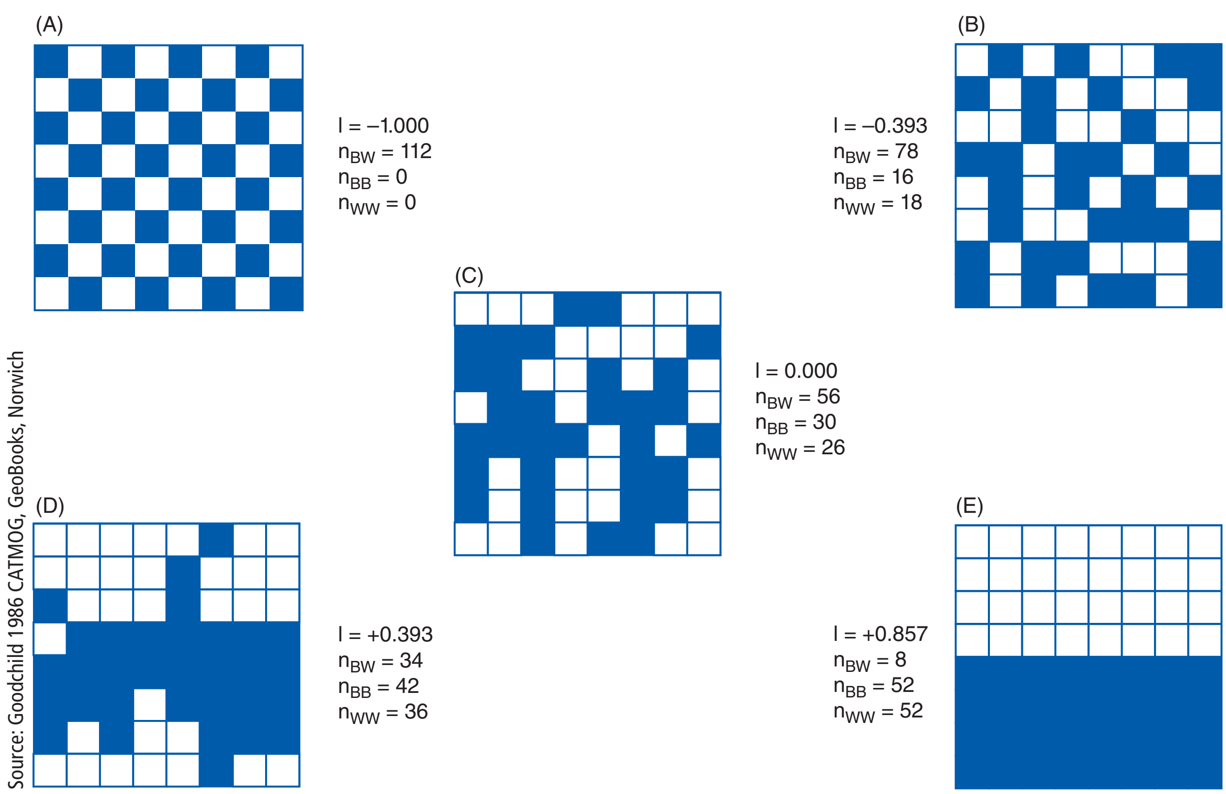
Objects existing in space are described by locational (spatial) descriptors and are conventionally classified using the taxonomy shown in Box 2.2. Spatial autocorrelation measures attempt to deal simultaneously with similarities in the location of spatial objects (Box 2.2) and their attributes (Box 2.1). If features that are similar in location are also similar in attributes, then the pattern as a whole is said to exhibit *positive spatial autocorrelation*. Conversely, *negative spatial autocorrelation* is said to exist when features that are close together in space tend to be more dissimilar in attributes than features that are further apart (in opposition to Tobler's Law). Zero autocorrelation occurs when attributes are independent of location.

Figure 2.1 presents some simple field representations of a geographic variable in 64 cells that can each take one of two values, coded blue and white. Each of the five illustrations contains the same set of attributes, 32 white cells and 32 blue cells, yet the spatial arrangements are very different. Figure 2.1A presents the familiar chess board and illustrates extreme negative spatial autocorrelation between neighboring cells. Figure 2.1E presents the opposite extreme of positive autocorrelation, when blue and white cells cluster together in homogeneous regions. The other illustrations show arrangements that exhibit intermediate levels of autocorrelation. Figure 2.1C corresponds to spatial independence, or no autocorrelation; Figure 2.1B shows a relatively dispersed arrangement; and Figure 2.1D shows a relatively clustered one.

Spatial autocorrelation is determined by similarities in both position and attributes.

The patterns shown in Figure 2.1 are examples of a particular case of spatial autocorrelation. In terms of the measurement scales described in Box 2.1, the attribute data are *nominal* (blue and white simply

Figure 2.1 Field arrangements of blue and white cells exhibiting: (A) extreme negative spatial autocorrelation, (B) a dispersed arrangement, (C) spatial independence, (D) spatial clustering, and (E) extreme positive spatial autocorrelation. The values of the I statistic are calculated using the equation in Section 2.6.



identify two different possibilities, with no implied order and no possibility of difference or ratio). The figure gives no clue to the true dimensions of the area being represented. In other cases, similarities in attribute values may be more precisely measured on higher-order measurement scales, enabling continuous measures of spatial variation (see Section 5.3.2.2 and Box 5.4 for a discussion of precision).

The way in which we define what we mean by *neighboring* in investigating spatial arrangements may be more or less sophisticated. In considering the various arrangements shown in Figure 2.1, we have only considered the relationship between the attributes of a cell and those of its four *immediate* neighbors. But we could include a cell's four diagonal neighbors in the comparison (see Figure 2.2), and more generally there is no reason why we should not interpret Tobler's Law in terms of a gradual, incremental attenuating effect of distance as we traverse successive cells.

We began this chapter by considering a time series analysis of events that are highly, even perfectly, repetitive in the short term. Activity patterns often exhibit strong positive temporal autocorrelation (where you were at this time last week or this time yesterday is likely to affect where you are now), but only if measures are made at the same time every day—that is, at the temporal scale of the daily interval. If, say, sample measurements were taken every 17 hours, measures of the temporal autocorrelation of your activity patterns would likely be much lower. Similarly, if the measures of the blue/white property were made at intervals that did not coincide with the dimensions

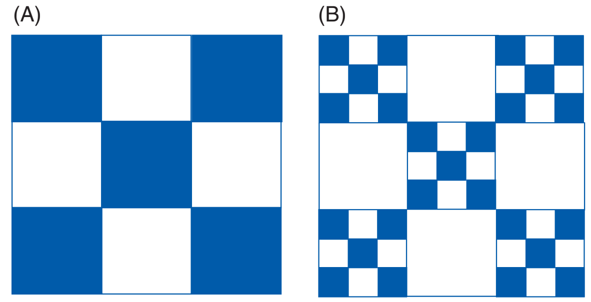


Figure 2.2 A Sierpinski carpet at two levels of resolution: (A) coarse scale, and (B) finer scale. In general, measures of spatial and temporal autocorrelation are scale dependent (see Box 2.7; this point is explored in connection with the chessboard example in Section 13.2.5).

of the squares of the chess boards in Figure 2.1, then the spatial autocorrelation measures would be different. Thus the issue of *sampling interval* is of direct importance in the measurement of spatial autocorrelation because spatial events and occurrences may or may not accommodate spatial structure.

Scale (see Box 2.3) is often integral to the trade-off between the level of spatial resolution and the degree of attribute detail that can be stored in a given application—as in the trade-off between spatial and spectral resolution in remote sensing (see Section 8.2.1). The scale at which data from different sources are usually made available is discussed in Chapter 8.

A particular instance of the importance of scale is illustrated using a mosaic of squares in Figure 2.2. Figure 2.2A is a coarse-scale representation of attributes in nine squares and a pattern of negative spatial autocorrelation. However, the pattern is

Technical Box 2.3

The Many Meanings of Scale

The concept of *scale* is fundamental to GI science, but unfortunately the word has acquired too many meanings in the course of time. Because they are to some extent contradictory, it is best to use other terms that have a clearer meaning where appropriate.

Scale is in the details. Many scientists use scale in the sense of spatial resolution, or the level of spatial detail in data. Data are fine scaled (or are at a fine level of granularity) if they include records of small objects and coarse-scaled (coarse-grained) if they do not.

Scale is about extent. Scientists also use scale to talk about the geographic extent or scope of a project: a large-scale project covers a large area, and a small-scale project covers a small area. Scale can also refer to other aspects of the project's scope, including the cost or the number of people involved.

The scale of a map. Geographic data are often obtained from maps and often displayed in map form. Cartographers use the term *scale* to refer to a map's *representative fraction* (the ratio of distance on the map to distance on the ground; see Section 3.7). Unfortunately, this leads to confusion (and often bemusement) over the meaning of *large* and *small* with respect to scale. To a cartographer a large scale corresponds to a large representative fraction, in other words to plenty of geographic detail. This is exactly the opposite of what an average scientist understands by a large-scale study. In this book we have tried to avoid this problem by using the terms *coarse* and *fine* instead.

© Paulo Ferreira/iStockphoto

Figure 2.3 Individual rocks may resemble the forms of larger structures, such as rock outcrops or eroded coastlines.

self-replicating at finer scales, and in Figure 2.2B, a finer-scale representation reveals that the smallest blue cells replicate the pattern of the whole area in a recursive manner. The pattern of spatial autocorrelation at the coarser scale is replicated at the finer scale, and the overall pattern is said to exhibit the property of *self-similarity*. Self-similar structure is characteristic of natural as well as social systems. For example, a rock may resemble the physical form of the mountain or coastline from which it was broken (Figure 2.3), small coastal features may resemble larger bays and inlets in structure and form, and neighborhoods may be of similar population size and each offer similar ranges of retail facilities across a metropolitan area. Self-similarity is a core concept of fractals, a topic introduced in Section 2.8.

2.4 Spatial Sampling

The quest to generalize about the myriad complexity of the real world requires us to abstract, or sample, events and occurrences from the universe of eligible elements of interest, which is known as the *sample frame*. A spatial sampling frame might be defined by the rectangle formed by four pairs of coordinates, or by the combined extent of a set of natural or artificial objects (see Box 2.2). We can think of spatial sampling as the process of selecting points from within this bounding rectangle or mosaic of objects. The process of sampling requires us to select some points or objects and to discard the others. The procedure of selecting some elements rather than others from a sample frame can very much determine the quality of the representation that is built using them.

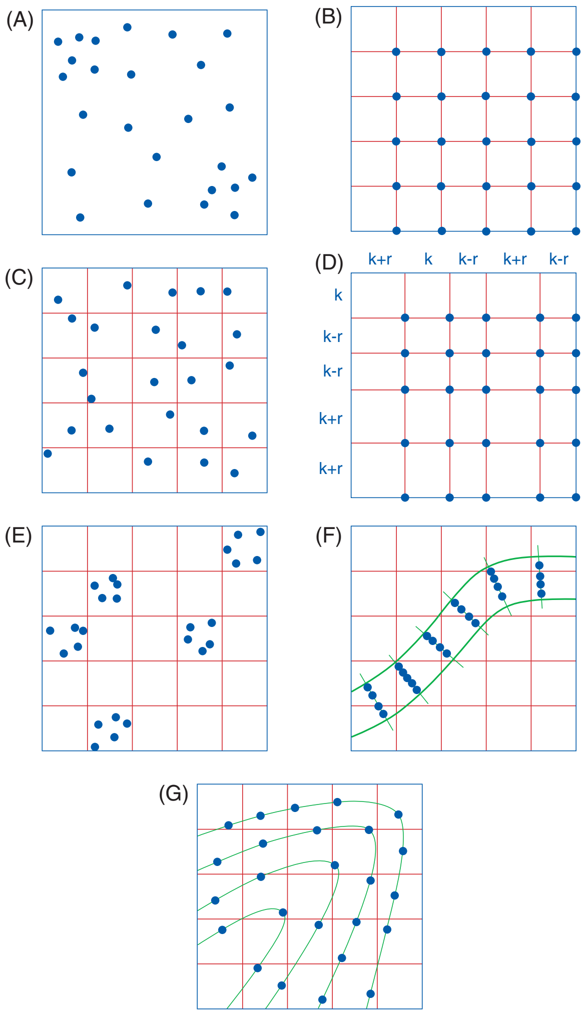
Scientific sampling requires that each element in the sample frame have a known and prespecified chance of selection. In some important senses, we can think of any geographic representation as a kind of sample, in that the elements of reality that are retained are abstracted from the observable world in accordance with some overall design. This is the case in remote sensing, for example (see Sections 3.6.1 and 8.2.1), in which each pixel value takes a spatially averaged reflectance value calculated at the spatial resolution characteristic of the sensor. In many situations, we will need to consciously select some observations, and not others, to create a generalizable abstraction. This is because, as a general rule, the resources available to any given project do not stretch to measuring every single one of the elements (soil profiles, migrating animals, shoppers) that we know to make up our population of interest. And even if resources were available, science tells us that this would be wasteful because procedures of *statistical inference* allow us to infer from samples to the populations from which they were drawn. We will return to the process of statistical inference in Section 14.5. Here, we will confine ourselves to the question, how do we ensure a good sample?

Geographic data are only as good as the sampling scheme used to create them.

Classical statistics often emphasizes the importance of randomness in sound sample design. The purest form, simple random sampling, is well known: each element in the sample frame is assigned a unique number, and a prespecified number of elements is selected using a random number generator. In the case of a spatial sample from continuous space, x, y coordinate pairs might be randomly sampled within the range of x and y values (see Section 4.8 for information on coordinate systems). Because each randomly selected element has a known and prespecified probability of selection, it is possible to make robust and defensible generalizations about the population from which the sample was drawn. A spatially random sample is shown in Figure 2.4A.

Random sampling is integral to probability theory, and this enables us to use the distribution of values in our sample to tell us something about the likely distribution of values in the parent population from which the sample was drawn. However, sheer bad luck can mean that randomly drawn elements are disproportionately concentrated among some parts of the population at the expense of others, particularly when the size of our sample is small relative to the population. For example, a survey of household incomes might happen to select households with unusually low incomes. Systematic spatial sampling aims to circumvent this problem and ensure greater evenness of coverage across the sample frame. This is achieved by identifying a regular sampling interval *k* (equal to the reciprocal of the sampling fraction *N/n*, where *n* is the required sample size and *N* is the size of the population) and proceeding to select every *k*th element. In spatial terms, the sampling interval of spatially systematic samples maps into a regularly spaced grid, as shown in Figure 2.4B.

Figure 2.4 Spatial sample designs: (A) simple random sampling, (B) stratified sampling, (C) stratified random sampling, (D) stratified sampling with random variation in grid spacing, (E) clustered sampling, (F) transect sampling, and (G) contour sampling.



The advantage this offers over simple random sampling may be two-edged, however, if the sampling interval and the spatial structure of the study area coincide, that is, if the sample frame exhibits *periodicity*. A sample survey of urban land use along streets originally surveyed under the U.S. Public Land Survey System (PLSS; Section 4.6) would be ill-advised to take a sampling interval of one mile, for example, for this was the interval at which blocks within townships were originally laid out, and urban structure is still likely to be repetitive about this original design. In such instances, there may be a consequent failure to detect the true extent of heterogeneity of population attributes. For example, it is extremely unlikely that the attributes of street intersection locations would be representative of land uses elsewhere in a block structure. A number of hybrid sample designs have been devised to get around the vulnerability of spatially systematic sample designs to periodicity and the danger that simple random sampling may generate freak samples. These include stratified random sampling to ensure evenness of coverage (Figure 2.4C) and periodic random changes in the grid width of a spatially systematic sample (Figure 2.5D), perhaps subject to minimum spacing intervals.

In certain circumstances, it may be more efficient to restrict measurement to a specified range of sites—because of the prohibitive costs of transport over large areas, for example. Clustered sample designs, such as that shown in Figure 2.4E, may be used to generalize about attributes if the cluster presents a microcosm of surrounding conditions. In fact, this provides a legitimate use of a comprehensive study of one area to say something about conditions beyond it—as long as the study area is known to be representative of the broader study region. For example, political opinion polls are often taken in shopping centers, where shoppers can be deemed broadly representative of the population at large. However, instances where they provide a comprehensive detailed picture of spatial structure are likely to be the exception rather than the rule, and in practice increased sample sizes are often used to mitigate this fact.

Use of either simple random or spatially systematic sampling presumes that each observation is of equal importance, and hence of equal weight, in building a representation. As such, these sample designs are suitable for circumstances in which spatial structure is weak or nonexistent, or where (as in circumstances fully described by Tobler's Law) the attenuating effect of distance is constant in all directions. They are also suitable in circumstances where spatial structure is unknown. Yet in most practical applications, spatial structure is (to some extent at least) known, even if it cannot be wholly explained by Tobler's Law. These circumstances make it both more efficient and necessary to devise application-specific sample designs. This makes for improved

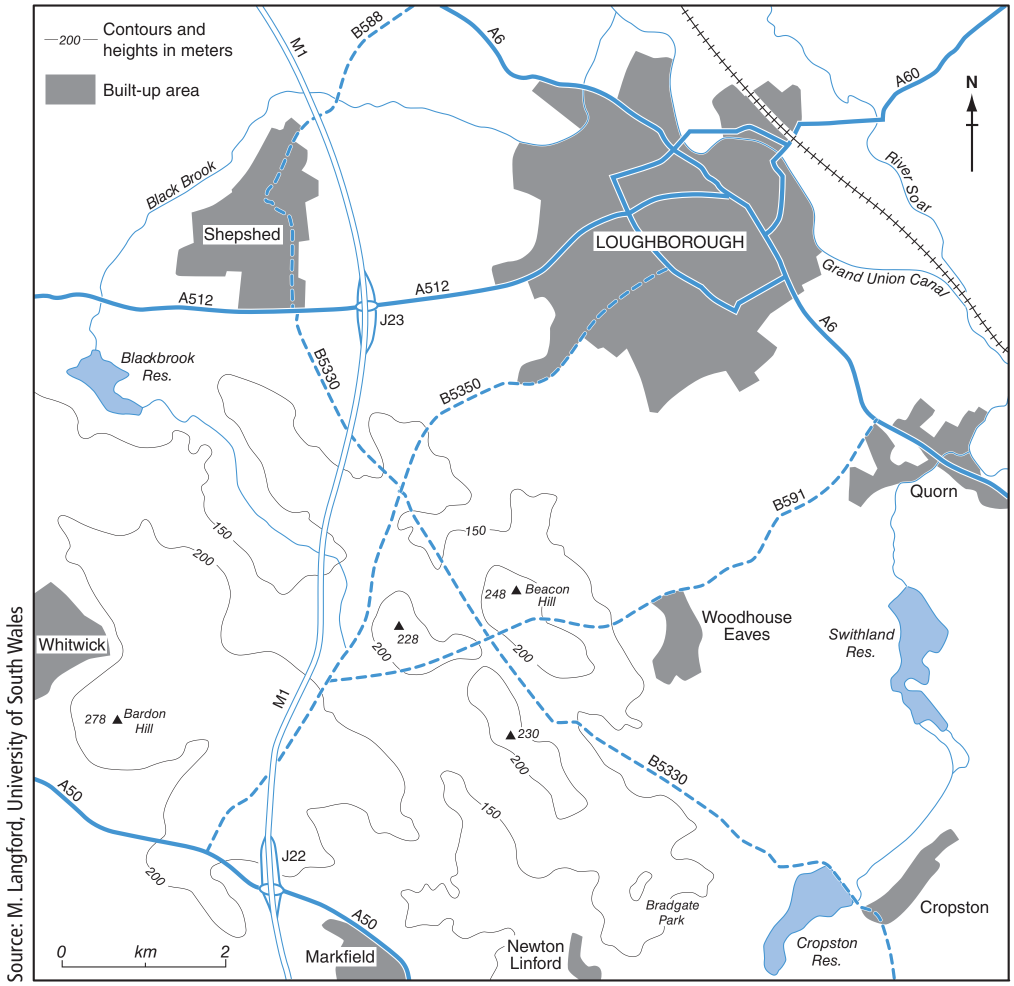


Figure 2.5 An example of physical terrain in which differential sampling would be advisable to construct a representation of elevation.

quality of representation, with minimum resource costs of collecting data. Relevant sample designs include sampling along a transect, such as a soil profile (Figure 2.4F), or along a contour line (Figure 2.4G).

Consider the area of Leicestershire, UK, illustrated in Figure 2.5. It depicts a landscape in which the hilly relief of an upland area falls away sharply toward a river's floodplain. In identifying the sample spot heights that we might measure and hold in a GI database to create a representation of this area, we would be well advised to sample a disproportionate number of observations in the upland part of the study area where the local variability of heights is greatest. In a socioeconomic context, consider the task of ascertaining the total repair cost of bringing all housing in a city up to a specified standard. (Such applications are common, for example, in forming bids for federal or central government funding.) If spatial data are available relating to the time period during which different neighborhoods were developed, these would provide a useful guide to effective use of sampling resources. Newer houses are all likely to be in more or less the same condition, whereas the repair costs of the older houses are likely to be much more heterogeneous and dependent on the different levels of attention that the occupants have lavished on them. As a general rule, older neighborhoods warrant a higher sampling interval than the newer ones, but other considerations may also be accommodated into the sampling design as well—such as construction type (duplex versus apartment, etc.) and local geology (as an indicator of risk of subsidence).

In any application where the events or phenomena that we are studying are spatially heterogeneous, we will require a large sample to capture the full variability of attribute values at all possible locations. Other parts of the study area may be much more homogeneous in attributes, and a sparser sampling interval may thus be more appropriate. Both simple random and systematic sample designs (and their variants) may be adapted to allow a differential sampling interval over a given study area (see Section 5.3.2.1 for more on this issue with respect to sampling vegetation cover). Thus it may be sensible to partition the sample frame into subareas, based on our knowledge of spatial structure—specifically our knowledge of the likely variability of the attributes that we are measuring. Other application-specific special circumstances include the following:

- Whether source data are ubiquitous or must be specially collected
- The resources available for any survey undertaking
- The accessibility of all parts of the study area to field observation (still difficult even in the era of ubiquitous availability of Global Positioning System receivers; see Section 4.9)

Stratified sampling designs accommodate the unequal abundance of different phenomena on the Earth's surface.

It is very important to be aware that this discussion of sampling is appropriate to problems where there is a large hypothetical population of evenly distributed locations (elements, in the terminology of sampling theory, or atoms of information in the terminology of Section 3.4) and that each has a known and prespecified probability of selection. Random selection of elements plays a part in each of the sample designs illustrated in Figure 2.4, albeit that the probability of selecting an element may be greater for clearly defined subpopulations that lie along a contour line or across a soil transect, for example. In circumstances where spatial structure is either weak or is explicitly incorporated through clear definition of subpopulations, standard statistical theory provides a robust framework for inferring the attributes of the population from those of the sample.

But reality is somewhat messier. In most practical applications, the population of elements (animals, glacial features, voters) may not be large, and its distribution across space may be far from random and independent. In these circumstances, conventional wisdom suggests a number of rules of thumb to compensate for the likely increase in error in estimating the true population value—as in clustered sampling, where slightly more than doubling the sample size is usually taken to accommodate the effects of spatial autocorrelation within a spatial cluster. However, the existence of spatial autocorrelation may fundamentally undermine the inferential framework and invalidate the process of generalizing from samples to populations. We examine this issue in more detail in our discussion of inference and hypothesis testing in Section 14.5.1.

2.5 Sampling and VGI

Thus far, our discussion has assumed that we have the luxury of collecting our own data for our own particular purpose. The reality of analysis in a data-rich world that is increasingly dominated by Big and Open Data (Section 17.4) is that more and more of the data that we use are collected by other parties, often for very different purposes. These parties may variously be nonexpert, self-selecting, or self-interested. The almost inevitable consequence is bias in the nature of the data that are available, and the source and effects of such bias may be far from obvious. Even if properly detected, sampling theory tells us that seeking to accommodate the effects of bias by reweighting under- and overrepresented observations is a very dangerous practice. For example, if few successful interviews of resident opinions are collected from addresses in gated communities, for reasons of access, then it is not acceptable to assign high weight to the few successful responses. In such cases the metadata of the data set are crucially important in establishing their provenance for the particular investigation that we may wish to undertake (see Section 10.2).

Still more insidious problems can characterize *volunteered geographic information* (VGI; see Section 1.5.6, and Box 3.5). At its best, it may be possible to invoke automated procedures that are robust and open to scrutiny to ensure that the data provided are unbiased in terms of their content, coverage, and collection procedures. But in the absence of knowledge about the proficiency and knowledge of every volunteer, the devil may be in the details. The wider "Wikification of GI systems" entails few safeguards about the reliability of ways in which the properties of places are represented in social media or community Web services.

The content, coverage, and collection procedures of VGI are often ill-defined

In the broad sense, VGI can be defined to include, for example, data that we "volunteer" to supermarket chains, telephone companies, Internet search engines, or financial services providers about our purchasing behavior and intentions to receive discounts or other incentives. Such data differ from "pure" VGI in that they are not supplied for purely altruistic purposes, but if the broader market niche of the businesses that collect such data is known, it may be possible to establish generalizations from such samples to the broader populations from which they are drawn.

Here are some questions to ask about the generalizability of VGI, some of which have been discussed in relation to the emergence of "citizen science" by researchers such as Muki Haklay and Dan Sui (Box 2.4):

1. Are all places equally accessible to volunteers, or is access to some areas difficult (e.g., physically remote areas) or constrained (e.g., gated communities)?
2. Are volunteers more likely to provide data on places or properties that interest them than those that do not? For example, social class bias among

OpenStreetMap volunteers appears to have led to a lack of coverage of low-status, peripheral housing developments in the early stages of the endeavor.

3. Where volunteers supply information about themselves, is this representative of other nonvolunteers? Retail-store loyalty card data, for example, may not be representative of those who do not patronize the store chain, or those who choose not to participate in loyalty programs.
4. Are volunteers at liberty to collect locationally sensitive data? Some governments are resistant to free use of social media or remain sensitive to record taking near military installations.
5. Should volunteers supply data that are open to malevolent use? This choice may appear straightforward in some circumstances (e.g., not revealing the locations of rare bird nesting sites to possible egg collectors) but in other instances is less clear-cut (e.g., mapping the geography of mobility-impaired individuals will help in developing a neighborhood fire evacuation strategy, but the information may also be of interest to criminals).
6. Is it socially acceptable to make available *any* observations of uniquely identifiable individuals—especially if they might be undertaking behavior that might be construed as socially unacceptable)?
7. Is the information date-stamped, as an indicator of its provenance and current reliability?

2.6 Distance Decay

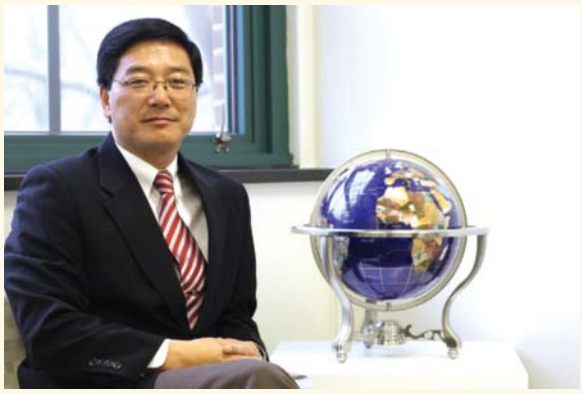
In selectively abstracting, or sampling, reality, judgment is required to fill in the gaps between the observations that make up a representation. This requires us to specify the likely attenuating effect of distance between the sample observations based on our understanding of the nature of geographic

Biographical Box 2.4

Dan Sui, Geographer and GI Scientist

Daniel Sui (隋殿志 or 老隋 in Chinese; Figure 2.6) is Distinguished Professor of Social and Behavioral Sciences at The Ohio State University, where he has also served as Chair of Geography, as a Director of the Institute of Population Research, and as Director of the Center for Urban and Regional Analysis. Dan undertook his undergraduate and master's degrees at Peking University and completed his PhD at the University of Georgia in 1993.

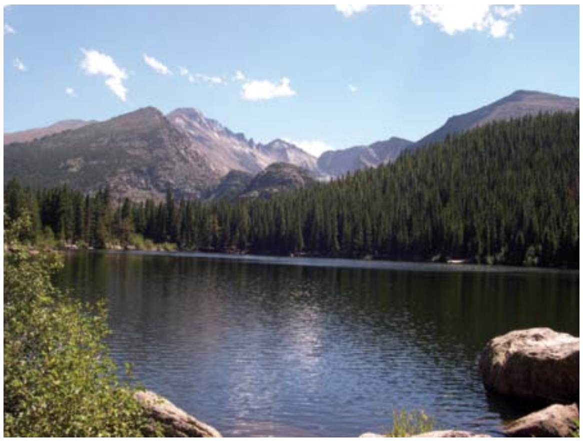
Dan's early work focused on new approaches to using GI in urban, environment, and health-related studies. In the early 2000s, he also worked on integrating GI systems with the mainstream of academic media studies to probe the challenging social, legal, and ethical issues related the emerging geographies of today's information society. Most recently, he has conducted research into the use of volunteered geographic information (VGI) and social media for crowdsourcing geographic knowledge. His 2012 coedited book provides a comprehensive primer on the phenomenon of VGI, seeing it as part of a profound transformation in the ways in which geographic data, information, and knowledge are produced and circulated (see Section 1.2). VGI contributes to the Big Data deluge (Section 1.4), and the way that it is handled is thus an important priority for data science (Section 1.5.4). Dan's work also makes clear that the uncertainties inherent in the use of VGI for geographic knowledge production need to be systematically evaluated if its contribution to science is to be both convincing and enduring.



Courtesy: Dan Sui

Figure 2.6 Dan Sui, geographer and GI scientist.

The supply and use of VGI is part of a much wider trend in the practice of science. Dan says: "After 50 years of development, both the social and technological environments for GI have changed dramatically. What happens in the next five years will be critical for the next phase of their development. A post–GI systems age needs to fully embrace the value of our emerging open culture—including Open Data, open software, open hardware, open standards, open research collaboration, open publication, open funding, and open education/learning. This offers us the best opportunities for responding to the challenges of Big Data, for better understanding our changing planet, and for reaping the benefits of VGI through citizen science. New technology-driven, application-led, science-inspired, and education-focused opportunities will propel GI systems to a new level of excellence, but we must also be mindful of the academic, legal, political, and environmental barriers for the development of open GI systems."



© Rachel Turk/iStockphoto

Figure 2.7 We require different ways of interpolating between points, as well as different sample designs, for representing mountains and forested hillsides.

data (Figure 2.7). That is to say, we need to make an informed judgment about an appropriate *interpolation* function and how to *weight* adjacent observations. A literal interpretation of Tobler's Law implies a continuous, smooth, attenuating effect of distance on the attribute values of adjacent or contiguous spatial objects, or incremental variation in attribute values as we traverse a surface. The polluting effect of a chemical spillage decreases in a predictable fashion with distance from the point source, aircraft noise decreases on a linear trend with distance from the flight path, and the number of visits from localities (suitably normalized by population) to a national park decreases at a regular rate as we traverse the counties that adjoin it. This section focuses on principles and introduces some of the functions that are used to describe attenuation effects over distance, or the nature of geographic variation. Section 13.3.6 discusses ways in which the principles of distance decay are embodied in techniques of spatial interpolation.

The precise nature of the function used to represent the effects of distance is likely to vary between applications, and Figure 2.8 illustrates several hypothetical types. In mathematical terms, if i is a point for which we have a recorded measure of an attribute and j is a point with no recorded measurement, we use b as a parameter that determines the rate at which the weight w_{ij} assigned to point j declines with distance from i. A small b value produces a slower decrease than a large one. In most applications, the choice of distance attenuation function is the outcome of past experience, the fit of a particular application data set, and convention. Figure 2.8A presents the simple case of linear distance decay, given by the expression:

$$w_{ij} = a - bd_{ij}$$

for $d_{ij} < a/b$. This function might, for example, reflect the noise levels experienced across a transect perpendicular to an aircraft flight path. Figure 2.8B presents a negative power distance decay function, given by the expression:

$$w_{ij} = d_{ij}^{-b}$$

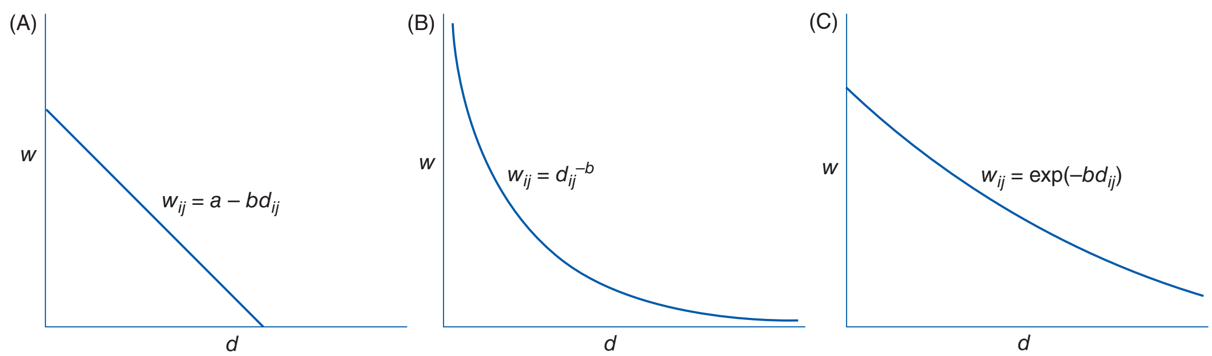
which some researchers have used to describe the decline in the density of resident population with distance from historic central business district (CBD) areas. Figure 2.8C illustrates a negative exponential statistical fit, given by the expression:

$$w_{ij} = e^{-bd_{ij}}$$

conventionally used in human geography to represent the decrease in retail store patronage with distance from it (e here denotes an exponential term, approximately equal to 2.71828, sometimes written *exp*).

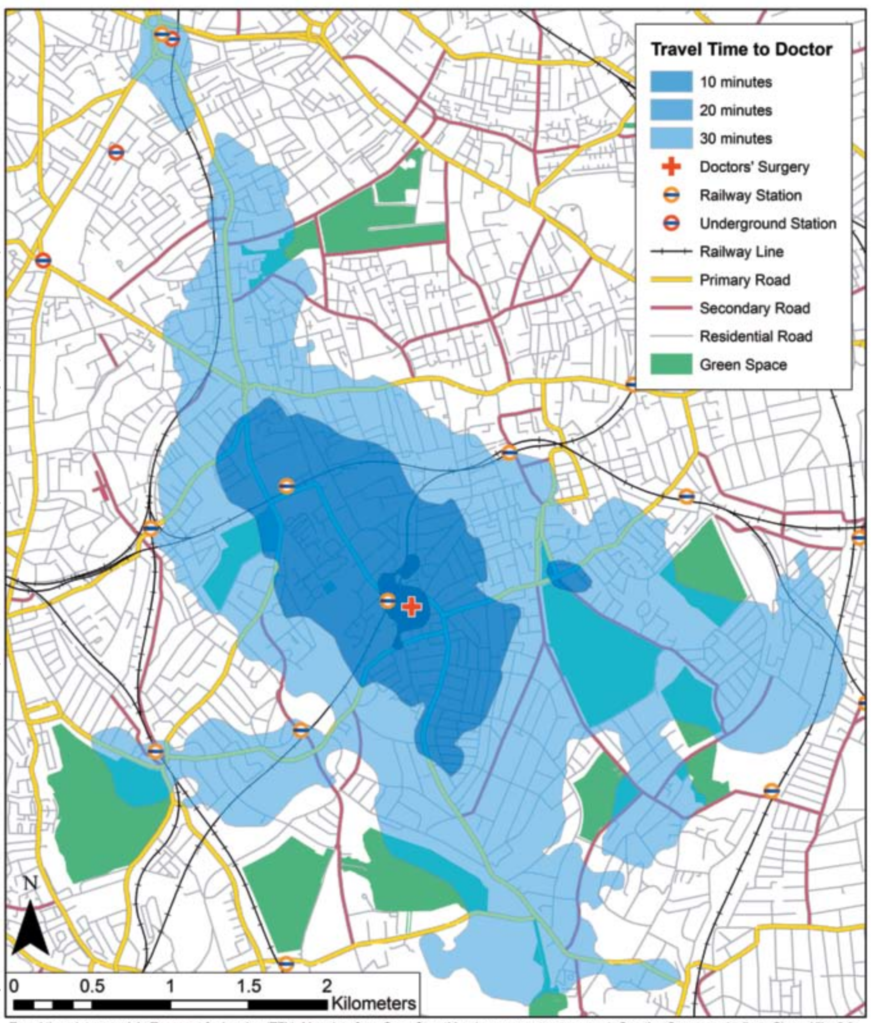
Each of the attenuation functions illustrated in Figure 2.8 is idealized in that the effects of distance are presumed to be regular, continuous, and *isotropic* (uniform in every direction). This notion of smooth and continuous variation underpins many of the representational traditions in cartography, as in the creation of *isopleth* (or isoline) maps (see Box 2.5). It is

Figure 2.8 The attenuating effect of distance: (A) linear distance decay, $w_{ij} = a - bd_{ij}$; (B) negative power distance decay, $w_{ij} = d_{ij}^{-b}$; and (C) negative exponential distance decay, $w_{ij} = e^{-bd_{ij}}$.



also consistent with our experience of high-school math, which is redolent of a world in which variation is continuous and best represented by interpolating smooth curves between everything. Yet even casual observation of the real world tells us that geographic variation is often far from smooth and continuous. The Earth's surface and geology, for example, are discontinuous at cliffs and fault lines, whereas the socioeconomic patterning of city neighborhoods can be similarly characterized by abrupt changes. Some illustrative issues pertaining to the catchment of a doctor's surgery general practice are apparent from Figure 2.9. Naïve GI analysis might assume that a map of 10-, 20-, and 30-minute travel times to see the doctor depicts a series of equidistant concentric circles. On this basis we might assume an isotropic linear distance decay function (Figure 2.8A), although the GI representation of the accessibility of the general practice is affected by

- Mode of transport—the 10-minute travel time buffer (Section 13.3.2) in effect measures walk time, whereas bus, car, and rail modes increase the extent of the 20- and 30-minute buffers.
- The quality and capacity of road and rail infrastructure, congestion issues at different times of day, and access constraints (e.g., to railway stations or the effects of road traffic calming in residential areas and one-way streets).
- Availability of public transport—by road or rail.
- Informal rights of way and the navigability of public open space (note that the public parks in Figure 2.9 appear to impede accessibility rather than enhance it).
- Physical barriers to movement, such as rivers.



Travel time data copyright Transport for London (TFL). Map data from Open StreetMap (www.openstreetmap.org), Creative Commons Attribute-Share Alike 2.0.
Map author: Daniel Lewis, UCL Department of Geography, May 2009

Courtesy: Daniel Lewis; travel time data © Transport for London; map data courtesy Open Streetmap.

Figure 2.9 Map showing 10-, 20-, and 30-minute travel times to a doctor's surgery in South London.

Technical Box 2.5

Isopleth and Choropleth Maps

Isopleth maps are used to visualize phenomena that vary continuously over space. An *isoline* connects points with equal attribute values, such as contour lines (equal height above sea level), *isohyets* (points of equal precipitation), *isochrones* (points of equal travel time), or *isodapanes* (points of equal transport cost). Figure 2.10 illustrates the procedures that are used to create a surface about a set of point measurements (Figure 2.10A), such as might be collected from rain gauges across a study region (see Section 13.3.6 for more technical detail on the process of spatial interpolation). A parsimonious number of user-defined values is identified to define the contour intervals (Figure 2.10B). A standard GI system operation is to interpolate a contour between point observations of greater and lesser value (Figure 2.10C) using standard procedures of distance decay, and the other contours are then interpolated using the same procedure (Figure 2.10D). Hue or shading can be added to improve user interpretability (Figure 2.10E).

Choropleth maps are constructed from values describing the properties of nonoverlapping areas, such as counties or census tracts. Each area is colored, shaded, or cross-hatched to symbolize the value of a specific variable, as in the 2011 UK Census data shown in Figure 2.11. Two types of variables can be used, termed *spatially extensive* and *spatially intensive*. Spatially extensive variables are those whose values are true only of entire areas, such as total population or total number of children under 5 years of age. Spatially intensive variables are those that could potentially be true of every part of an area, if the area were homogeneous; examples include densities, rates, or proportions. Conceptually, in terms of the field-object distinction, which we will introduce in the next chapter, a spatially intensive variable is a field, averaged over each area, whereas a spatially extensive variable is a field of density whose values are summed or integrated to obtain each area's value.

Source: After Kraak and Ormeling 2003: *Cartography: Visualization of Geospatial Data*, p134

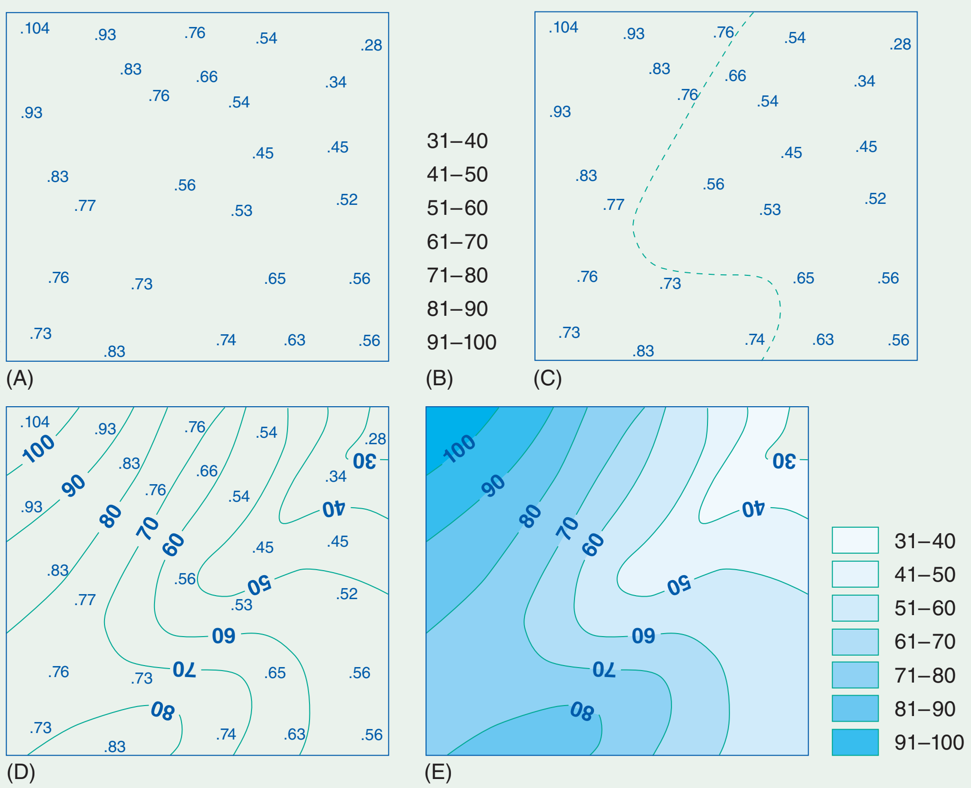
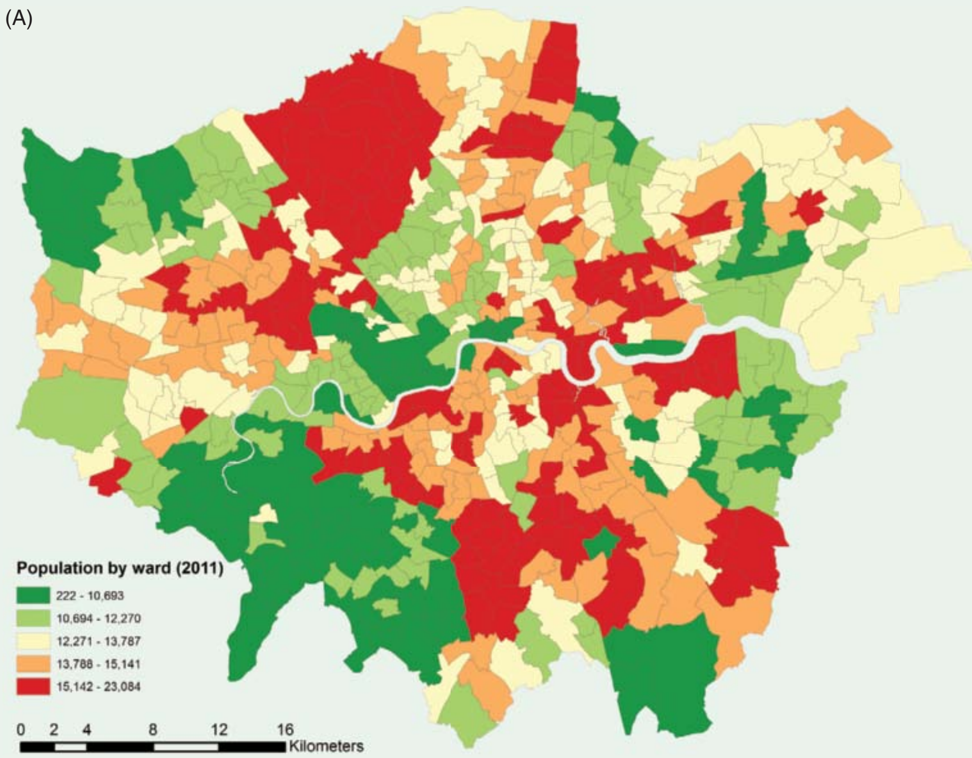


Figure 2.10 The creation of isopleth maps: (A) point attribute values, (B) user-defined classes, (C) interpolation of class boundary between points, (D) addition and labeling of other class boundaries, and (E) use of hue to enhance perception of trends.

Source: 2011 UK Census of Population. Images courtesy: Alistair Leak



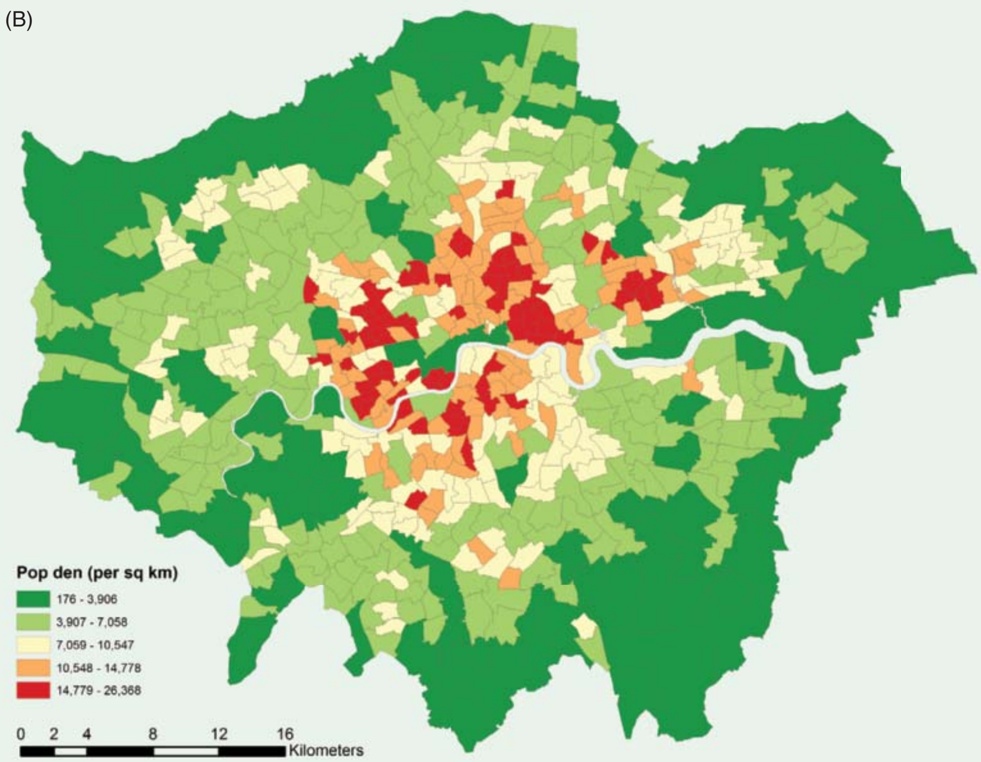


Figure 2.11 Choropleth maps of (A) a spatially extensive variable, total population, and (B) a related but spatially intensive variable, population density. Many cartographers would argue that (A) is misleading and that spatially extensive variables should always be converted to spatially intensive form (as densities, ratios, or proportions) before being displayed as choropleth maps.

In practice, the number of patients who use the general practice is also likely to depend on patient characteristics, as measured by

- Health care needs, their match with medical practice specialisms, and other socioeconomic characteristics
- The availability of medical services from other local providers.
- A demand constraint, which requires that the probabilities of any individual attending each of the available practices sum to 1 (unless people opt out of the health-care system)

2.7 Measuring Distance Effects as Spatial Autocorrelation

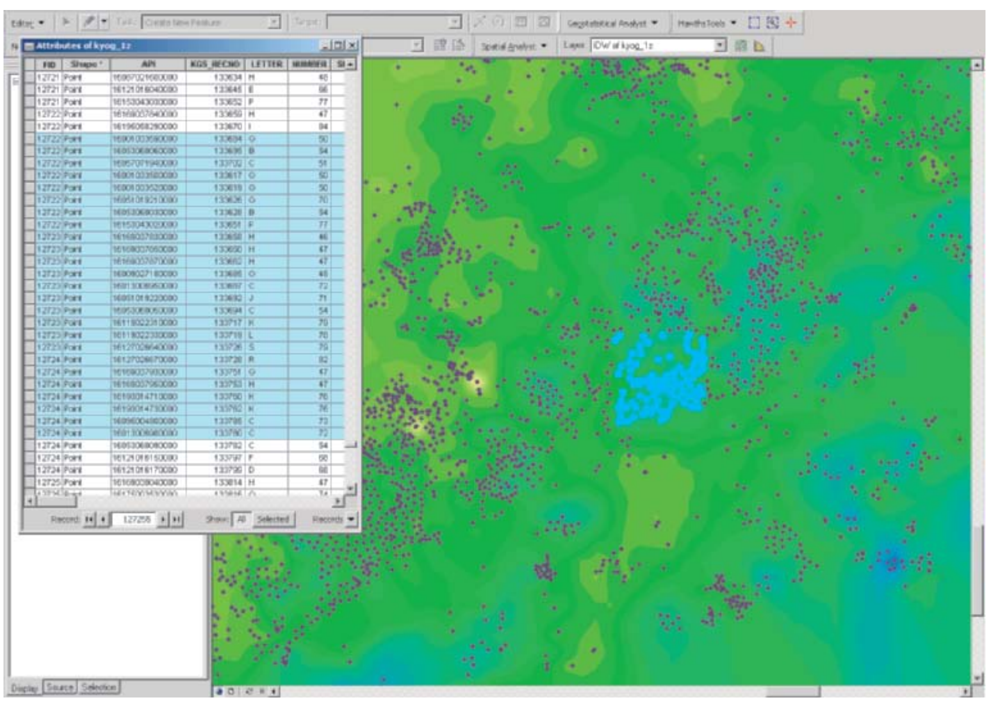
Understanding spatial structure is key to building understanding of other real-world structures because it helps us to *deduce* a good sampling strategy and to use an appropriate means of interpolating between sampled points that is fit for purpose. Knowledge of the actual or likely nature of spatial autocorrelation is thus important deductive understanding. However, in many applications we do not understand enough about geographic variability, distance effects, and spatial structure to make reliable deductions. A further branch of spatial analysis thus emphasizes the *measurement* of spatial autocorrelation as an end in itself. This amounts to a more *inductive* approach to developing an understanding of the nature of a geographic data set.

Induction reasons from data to build up understanding, whereas deduction begins with theory and principle as a basis for looking at data.

In Section 2.3 we saw that spatial autocorrelation measures the extent to which similarities in position match similarities in attributes. Methods of measuring spatial autocorrelation depend on the types of objects used as the basis of a representation, and as we saw in Section 2.2, the scale of attribute measurement is important too. Interpretation depends on how the objects relate to our conceptualization of the phenomena they represent.

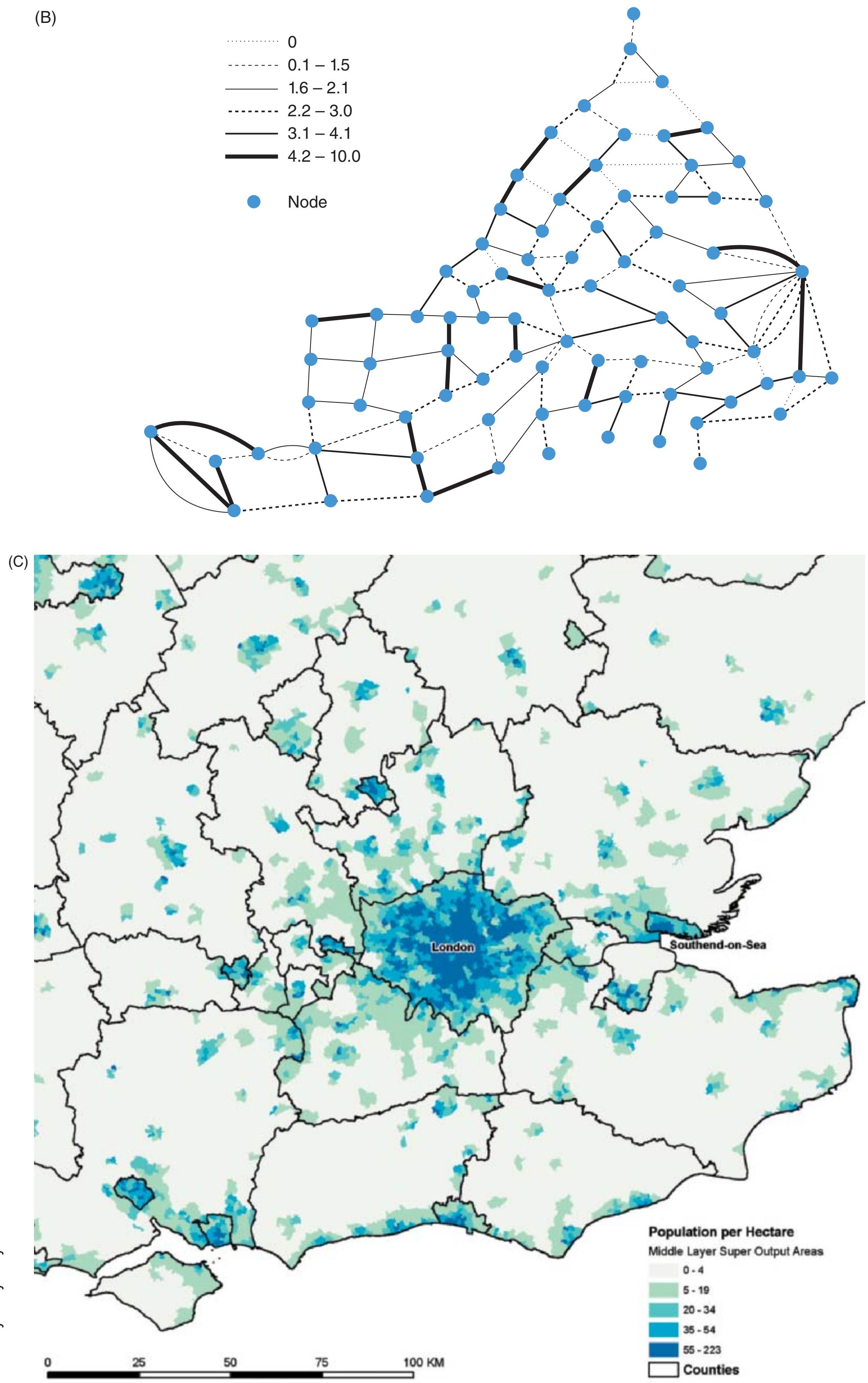
Figure 2.12 shows examples of each of the four object types described in Box 2.2, with associated attributes, chosen to represent situations in which a scientist might wish to measure spatial autocorrelation. The point data in Figure 2.12A comprise data on wellbores over an area of 30 km^2 and together provide information on the depth of an aquifer beneath the surface (the blue shading identifies those within a given threshold). We would expect values to exhibit strong spatial autocorrelation, with departures from this indicative of changes in bedrock structure or form. The line data in Figure 2.12B present numbers of accidents for links of road over a lengthy survey

(A)



Courtesy: Esri Inc.

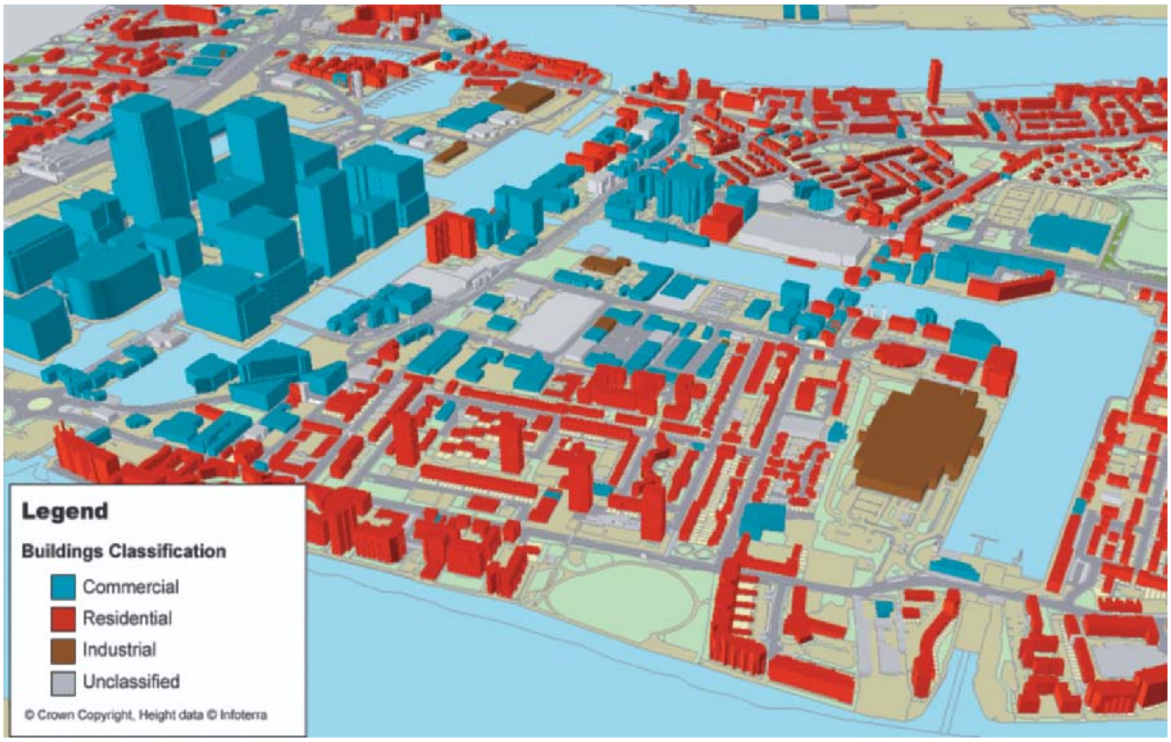
Figure 2.12 Situations in which a scientist might want to measure spatial autocorrelation: (A) point data (wells with attributes stored in a spreadsheet; linear extent of image 0.6 km); (*continued*)



Courtesy: Daryl Lloyd

Figure 2.12 (*continued*) (B) line data (accident rates in the southwestern Ontario provincial highway network); (C) area data (percentage of population that are old age pensioners in southeast England);

(D)



Courtesy: Duncan Smith.

Figure 2.12 (*continued*) (D) volume data (elevation and volume of buildings in London: east to the top of the image; image extent 1.6 km north-south, 1.5 km east-west).

period in the southwestern Ontario (Canada) provincial highway network. Low spatial autocorrelation in these statistics implies that local causative factors (such as badly laid out junctions) account for most accidents, whereas strong spatial autocorrelation would imply a more regional scale of variation, implying a link between accident rates and lifestyles, climate, or population density. The area data in Figure 2.12C illustrate the settlement structure of southeast England and invite analysis of the effectiveness of land use zoning (green belt) policies, for example. The volume data in Figure 2.12D allow some measure of the spatial autocorrelation of high-rise structures to be made as part of a study of the relationship between the built form and economic function of London, UK. The way that spatial autocorrelation statistic might actually be calculated for the data used to construct Figure 2.12C is described in Box 2.6.

Spatial autocorrelation measures tell us about the interrelatedness of phenomena across space, one attribute at a time. Its measurement is key to formalizing and understanding many geographic problems, and it is central to locational analysis in geography. Another important facet to the nature of geographic

Technical Box 2.6

Measuring Similarity between Neighbors

In the simple example shown in Figure 2.13, we compare neighboring values of spatial attributes by defining a weights matrix **W** in which each element w_{ij} measures the locational similarity of i and j (i identifies the row and j the column of the matrix). We use a simple measure of contiguity, coding $w_{ij} = 1$ if regions i and j are contiguous and $w_{ij} = 0$ otherwise. w_{ii} is set equal to 0 for all i. This is shown in Table 2.1.

The weights matrix provides a simple way of representing similarities between location and attribute values in a region of contiguous areal objects. Autocorrelation is identified by the presence of neighboring cells or zones that take the same (binary) attribute value. More sophisticated measures of w_{ij} might include a decreasing function (such as one of those shown in Figure 2.7) of the straight-line distance between points at the centers of zones, or the lengths

Table 2.1 The weights matrix **W** derived from the zoning system shown in Figure 2.12

	1	2	3	4	5	6	7	8
1	0	1	1	1	0	0	0	0
2	1	0	1	0	0	1	1	0
3	1	1	0	1	1	1	0	0
4	1	0	1	0	1	0	0	0
5	0	0	1	1	0	1	0	1
6	0	1	1	0	1	0	1	1
7	0	1	0	0	0	1	0	1
8	0	0	0	0	1	1	1	0

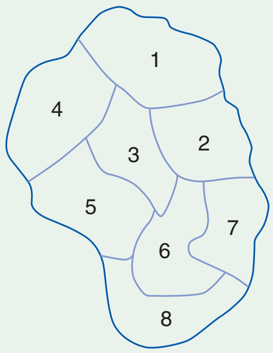


Figure 2.13 A simple mosaic of zones.

of common boundaries (although in this case weight would be an increasing function of length, rather than the decreasing function of distance implied by *b* in Figure 2.8). A range of different spatial metrics may also be used, such as existence of linkage by air, or a decreasing function of travel time by air, road, or rail, or the strength of linkages between individuals or firms on some (nonspatial) network.

The weights matrix makes it possible to develop measures of spatial autocorrelation using attributes measured on the nominal, ordinal, interval, or ratio scales (see Box 2.1) and the dimensioned classes of spatial objects shown in Box 2.2. Any measure of spatial autocorrelation seeks to compare a set of locational similarities w_{ij} (contained in a weights matrix) with a corresponding set of attribute similarities c_{ij}, combining them into a single index in the form of a cross product:

$$\sum_i \sum_j c_{ij} w_{ij}$$

This expression is the total obtained by multiplying every cell in the **W** matrix with its corresponding entry in the **C** matrix and summing.

There are different ways of measuring the attribute similarities, c_{ij}, depending on whether they are measured on the nominal, ordinal, interval, or ratio scale (see Box 2.1). For nominal data, the usual approach is to set c_{ij} to 1 if *i* and *j* take the same attribute value, and zero otherwise. For ordinal data, similarity is usually based on comparing the ranks of *i* and *j*. For interval and ratio data, the attribute of interest is denoted z_i, and the product $(z_i - \bar{z})(z_j - \bar{z})$ is calculated, where $\bar{z}$ denotes the average of the *z*'s.

One of the most widely used spatial autocorrelation statistics for the case of area objects and interval scale attributes is the Moran Index. This is positive when nearby areas tend to be similar in attributes, negative when they tend to be more dissimilar than one might expect, and approximately zero when attribute values are arranged randomly and independently in space. It is given by the expression:

$$I = \frac{n \sum_i \sum_j w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\sum_i \sum_j w_{ij} \sum_i (z_i - \bar{z})^2}$$

where *n* is the number of areal objects in the set. This brief exposition is provided at this point to emphasize the way in which spatial autocorrelation measures are able to accommodate attributes scaled as nominal, ordinal, interval, and ratio data and to illustrate that there is flexibility in the nature of contiguity (or adjacency) relations that may be specified. Further techniques for measuring spatial autocorrelation are reviewed in connection with spatial interpolation in Section 13.3.6.

data is the tendency for relationships to exist between different phenomena at the same location. The interrelatedness of the various properties of a location that together constitute place (Section 1.3) is an important aspect of the nature of geographic data and is key to understanding how the world works (Section 1.3). But it is also a property that defies conventional statistical analysis because most such methods assume zero spatial autocorrelation of sampled observations—in direct contradiction to Tobler's Law.

2.8 Taming Geographic Monsters

Thus far in our discussion of the nature of geographic data, we have assumed that spatial variation is smooth and continuous, apart from when we encounter abrupt truncations and discrete shifts at boundaries. However, much spatial variation does not appear to possess these properties, but rather is jagged and apparently irregular. The processes that give rise to the form of a mountain range produce features that are spatially autocorrelated (for example, the highest peaks tend to be clustered), yet it would be wholly inappropriate to represent a mountainscape using smooth interpolation between peaks or between valley troughs.

Jagged irregularity turns out to be a property that is also often observed across a range of scales, and detailed irregularity may resemble coarse irregularity in shape, structure, and form. We commented on this in Section 2.3 when we suggested that a rock broken off a mountain may, for reasons of lithology, represent the mountain in form; this property has been termed *self-similarity*. Urban geographers also recognize that cities and city systems are also self-similar in organization across a range of scales, and Batty and Longley have discussed the ways in which this echoes many of the earlier ideas of Walter Christaller's Central Place Theory. It is unlikely that idealized smooth curves and conventional mathematical functions will provide useful representations for self-similar, irregular spatial structures: at what scale, if any, does it become meaningful to approximate the San Andreas Fault system by a continuous curve? Urban geographers, for example, have long sought to represent the apparent decline in population density with distance from historic central business districts (CBDs) as a continuous curve (Figure 2.8B), yet the three-dimensional profiles of cities (Figure 2.12D) are clearly characterized by urban canyons between irregularly spaced high-rise buildings. Each of these phenomena is characterized by spatial trends (the largest faults, the largest mountains, or the largest skyscrapers each tend to be close to one another), but they are not contiguous and smoothly joined, and the kinds of surface functions shown in Figure 2.8 present inappropriate generalizations of their structure.

For many years, such features were considered geometrical monsters that defied intuition. More recently, however, a more general geometry of the irregular, termed *fractal geometry* by Benoît Mandelbrot, has come to provide a more appropriate and general means of summarizing the structure and character of spatial objects. Fractals can be thought of as geometric objects that are, literally, between Euclidean dimensions, as described in Box 2.7.

Technical Box 2.7

The Strange Story of the Lengths of Geographic Objects

How long is the coastline of Maine (Figure 2.14)? (Benoît Mandelbrot, a French mathematician of Polish origin, posed a similar question in 1967 with regard to the coastline of Great Britain.) Consider the stretch of coastline shown in Figure 2.14A. With dividers set to measure 100-km intervals, we would take approximately 4.4 swings and record a length of 340 km (Figure 2.14B).

If we then halved the divider span to measure 50-km swings, we would take approximately 7.1 swings, and the measured length would increase to 355 km (Figure 2.14C). If we halved the divider span once again to measure 25-km swings, we would take approximately 16.6 swings, and the measured length would increase still further to 415 km (Figure 2.14D).

And so on until the divider span was so small that it picked up all of the detail on this particular representation of the coastline. But that would not be the end of the story.

If we were to resort instead to field measurement, using a tape measure or the distance measuring instruments (DMIs) used by highway departments, the length would increase still further, as we picked up wiggles in the coast that even the most detailed maps do not seek to represent. If we were to use dividers, or even microscopic measuring devices, to measure every last grain of sand or earth particle, our recorded length measurement would stretch toward infinity, without apparent limit.

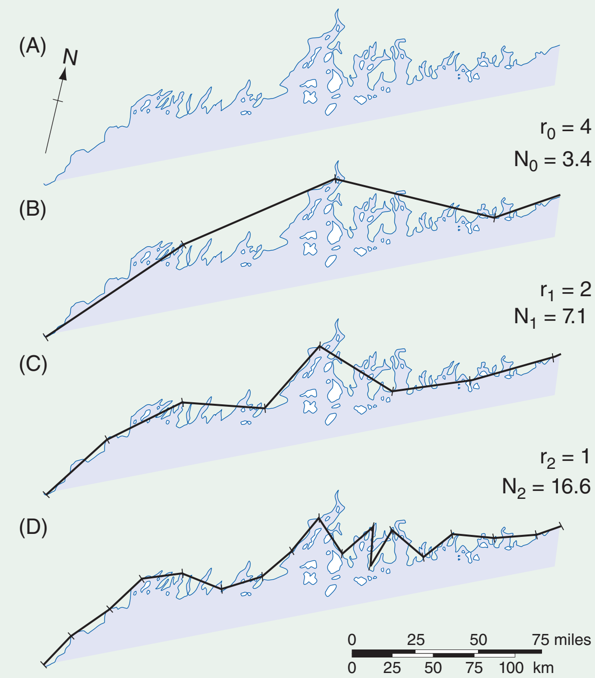


Figure 2.14 The coastline of Maine, at three levels of recursion: (A) the base curve of the coastline, (B) approximation using 100-km steps, (C) 50-km step approximation, and (D) 25-km step approximation.

In a self-similar object, each part has the same structure as the whole.

Ideas from fractal geometry are important, and for many phenomena a measurement of fractal dimension is as important as measures of spatial autocorrelation, or of medians and modes in standard statistics. An important application of fractal concepts is discussed in Section 14.3.1, and we return again to the issue of length estimation in Section 13.3.1. Ascertaining the fractal dimension of an object involves identifying the scaling relation between its length or extent and the yardstick (or level of detail) that is used to measure it. Regression analysis, which can be thought of as putting a best fit line through a scatter of points, provides one (of many) means of establishing this relationship. If we return to the Maine coastline example in Figure 2.14, we might obtain scale-dependent coast-length estimates (L) of 14.6 (4×3.4), 14.1 (2×7.1), and 16.6 (1×16.6) units for the step lengths (r) used in Figures 2.14B, 2.14C, and 2.14D, respectively. (It is arbitrary whether the steps are measured in kilometers or miles.)

If we then plot the natural log of L (on the y-axis) against the natural log of r for these and other pairs of values, we will build up the scatterplot shown in Figure 2.15. If the points lie more or less on a straight line and we fit a trend (regression) line through it, the value of the slope (b) parameter is equal to ($1 - D$), where

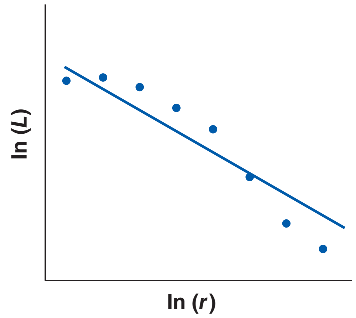


Figure 2.15 The relationship between recorded length (L) and step length (r).

D is the fractal dimension of the line. This method for analyzing the nature of geographic lines was originally developed by Lewis Fry Richardson (Box 2.8).

2.9 Induction and Deduction and How It All Comes Together

The continual message of this chapter is that spatial is special—that geographic data have a unique nature. Tobler's Law presents an elementary general rule about spatial structure and provides a starting point for the measurement and simulation of spatially autocorrelated structures. This in turn assists us in devising

Biographical Box 2.8

Lewis Fry Richardson

Lewis Fry Richardson (1881–1953; Figure 2.16) was one of the founding fathers of the ideas of scaling and fractals. He was brought up a Quaker and, after earning a degree at Cambridge University, went to work for the Meteorological Office, but his pacifist beliefs forced him to leave in 1920 when the Meteorological Office was militarized under the Air Ministry. His early work on how atmospheric turbulence is related at different scales established his scientific reputation. Later he became interested in the causes of war and human conflict, and in order to pursue one of his investigations, he found that he needed a rigorous way of defining the length of a boundary between two states. Unfortunately, published lengths tended to vary dramatically, a specific instance being the difference between the lengths of the Spanish–Portuguese border as stated by Spain and by Portugal. Richardson developed a method of walking a pair of dividers along a mapped line and analyzed the relationship between the length estimate and the setting of the dividers, finding remarkable predictability. In the 1960s Benoît Mandelbrot's concept of fractals finally provided the theoretical framework needed to understand this result.

Source: Wikipedia

Figure 2.16 Lewis Fry Richardson.

appropriate spatial sampling schemes and creating improved representations, which tell us still more about the real world and how we might represent it.

Spatial data provide the foundations for operational and strategic applications of GI, and these foundations must be developed creatively, yet rigorously, if they are to support the spatial analysis superstructure that we wish to erect on them. This entails much more than technical competence with software. An understanding of the nature of geographic data allows us to use induction (reasoning from observations) and deduction (reasoning from principles and theory) alongside each other to develop effective spatial representations that are safe to use.

Questions for Further Study

1. Many jurisdictions tout the number of miles of shoreline in their community—for example, Ottawa County, Ohio, claims 107 miles of Lake Erie shoreline. What does this mean, and how could you make it more meaningful?

2. With reference to Figure 2.11, list the design considerations that should be incorporated into GI software to measure accessibility of (a) a neighborhood medical center to wheelchair-bound pedestrians, (b) a grocery store to high-income customers, and (c) all residential buildings in a small town from a single fire service station.

3. The apparatus of inference was developed by statisticians because they wanted to be able to reason from the results of experiments involving small samples to make conclusions about the results of much larger, hypothetical experiments—for example, in using samples to test the effects of drugs. Summarize the problems inherent in using this apparatus for geographic data, in your own words.

4. "Can geometry deliver what the Greek root of its name [geo-] seemed to promise—truthful measurement, not only of cultivated fields along the Nile River but also of untamed Earth?" Discuss this challenge posed by Mandelbrot, offered in his 2012 autobiography, *The Fractalist: Memoir of a Scientific Maverick* (New York: Pantheon Books, p. xii).

Further Reading

Batty, M. and Longley, P. A. 1994. *Fractal Cities: A Geometry of Form and Function*. London: Academic Press. Available for download (free of charge) from www.fractalcities.org.

De Smith, M., Goodchild, M. F., and Longley, P. A. 2009. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools.* Chapter 5: Data Exploration and Spatial Statistics. Leicester: Troubador. See also www.spatialanalysisonline.com.

Goodchild, M. F. 2004. The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers* 94(2): 300–304.

Haklay, M. 2010. *Interacting with Geospatial Technologies*. Chichester: Wiley.

Harvey, F. J., Tobler, W., Chrisman, N. R., Frank, A. U., Sui, D., and Goodchild, M. F. 2012. *Are There Fundamental Principles in Geographic Information Science?* www.createspace.com.

Mandelbrot, B. B. 1984. *The Fractal Geometry of Nature*. San Francisco: Freeman.