3

Geographic Information Systems

Michael F. Goodchild

In Chapter 1 Anne Godlewska explored the idea of the map and showed how the development of printed representations of the world changed our view of our surroundings. The role of maps in structuring our thinking about the world is often subtle but nevertheless real, and recent work by cartographers has helped to draw attention to the power of maps to influence the relationship between people and their environment (Wood 1992). Although we often think of maps as pictorial, and therefore neutral, representations of the world, the choice of features, the ways in which they are differentiated, and the way in which the curved surface of the earth is distorted to fit onto a flat sheet of paper are all representational choices that can powerfully influence our understanding of the world. Even the custom of placing north at the top of a map has been seen as sending a subtle cultural message.

Over the past three decades, as part of a more general trend toward the use of digital technology for information handling, a significant change has occurred in the nature of geographic information and its role in society. Unlike numbers and text, maps and images have created major problems for digital storage and processing because of their relative complexity and high density of information. The so-called "geographic information technologies" that have appeared over the past twenty years and grown to form a major new area of application in electronic data processing are the result of significant research developments in hardware and software. The new technologies are likely to be every bit as influential as the map has been on our thinking

about the world. They broaden our perspective by offering new capabilities that are less constrained, but at the same time they impose new filters that may be just as subtle as the ones associated with paper maps.

Limitations of Traditional Mapmaking

The earliest symbolic representations of the world around us made use of a variety of materials, from scratches on rocks to stick models used to navigate the Pacific (Harley and Woodward 1987). But the emergence of paper and printing created a medium that came to dominate mapmaking, at least until the development of aerial photography and digital imagery in the twentieth century.

With simple tools, it is possible to make representations on paper of a wide range of geographic phenomena and features and thus to communicate knowledge of them to the person reading the map. It is possible to print text using movable type or templates, to draw lines of constant width, and to fill areas with uniform color. Additional techniques allow areas to be filled with uniform, pre-printed patterns, such as dots or cross-hatching. But despite their apparent flexibility, these techniques are in reality quite limiting (Goodchild 1988). It is difficult, for example, to vary line width continuously in order to represent changes in the width of a road or river or to indicate varying uncertainty about the position of a boundary line or contour. It is hard to fill with continuously changing color or pattern in order to represent gradients in the densities of species at the edge of a forest or in rainfall in the lee of a mountain range. It is also difficult for the map reader to interpret continuous variation correctly. As a result, map representations tend to emphasize the abrupt or crisp aspects of geographic variation and to suppress smooth change or fuzziness.

Nowhere is this sense of technical limitation stronger than in the conventions devised to depict topographic elevation. The familiar isoline or contour technique has been standard for many decades, but its limitations are clear in the need for instruction in map reading in schools and in the use of supplementary techniques such as color or hachuring to help the map reader interpret the pattern of contours correctly. Isolines evolved not only because of their efficiency in communicating information about topographic elevation, but also because they were one of the few methods available within the constraints of cartographic technique; they could be drawn with a pen. They give no information at all about the shape of the surface between contoured elevations or about the cartographer's uncertainty regarding elevation

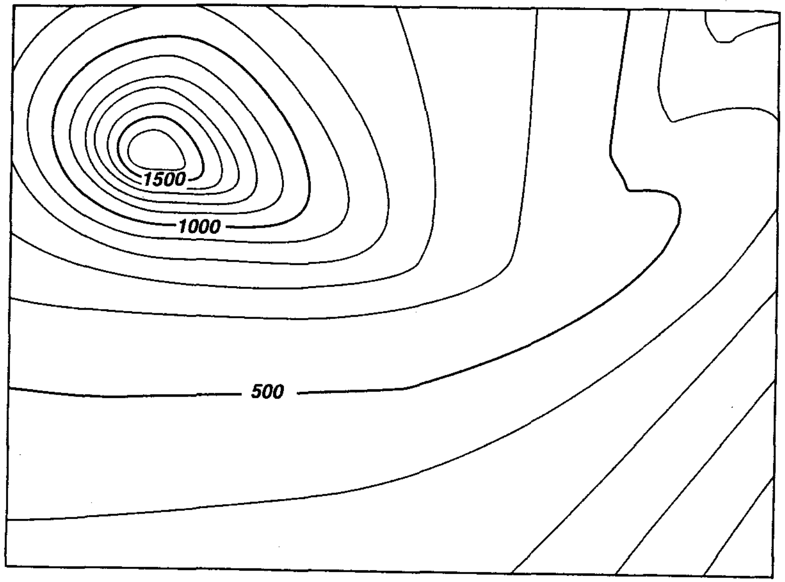

Figure 3.1. Part of a contour representation of a surface. Mapmaking technique requires that contours be of uniform width, and hence cannot represent either varying uncertainty in contour position or the form of the surface between contours.

measurements and the consequent impact of that uncertainty on contour positions (Figure 3.1).

If the objective of the map is to communicate the knowledge of the mapmaker to the map reader, then one can reasonably ask questions about the efficiency of the communication. In the case of contours, the information available to a mapmaker following conventional topographic mapping practice includes direct measurements of elevation at selected points (spot heights), plus the ability to estimate elevation photogrammetrically at any point. The map reader is provided with estimates of elevation of known accuracy at points lying on contours, plus the ability to estimate elevation at other points with much less accuracy. For much of the map, elevation could be anywhere between the elevations of the adjacent contours, so the quality of estimates depends directly on the contour interval.

In short, because only a small fraction of the information available to the mapmaker is actually communicated, the efficiency of map communication can be disappointingly low. Similar points can be made about the efficiency of soil maps in communicating the knowledge of field surveys by soil scientists or about choropleth maps as communicators of knowledge about population density. In all of these examples, the information available to the map reader is only a small fraction of that available to the mapmaker, at least in part because of the constraints imposed by the technical limitations of cartographic technique. Because the real world is a complex and often frightening place, however, a map that simplifies the world and communicates only a small fraction of its real complexity may actually provide the map reader with subtle reassurance.

Mark and Frank (1991) and others have recently introduced principles of cognitive science and linguistics into discussions of the nature of geographic information and mapmaking. These scholars suggest that our ability to learn and reason about geographic information is limited by the structures inherent in language, structures that are strongly biased toward a view of the world in terms of discrete, crisp objects and sharp change. Such patterns are the geographic expression of a more general propensity to use discrete categories and to attach labels. The English language, for example, is rich in terms describing spatial relationships between objects (terms such as "within," "over," "across," "outside"), but fuzziness, uncertainty, and continuous change tend to be avoided in everyday human discourse and are less well served by English vocabulary.

These arguments suggest that the use of contours to depict elevation, or of crisply defined areas to depict soil variation, reflects human cognitive preference as much as the limitations of cartographic technique. In this view, the filtering imposed by the communication channel is considered not an unfortunate loss of information, but rather a desirable stage of generalization, a way of improving the usefulness of cartographic information to the map reader. A map that showed the world as it really is, in all its complexity, fuzziness, and spatially continuous change, would be less useful than the conventional product because its contents would fail to fit with our cognitive processes. In short, we find it difficult to learn about the world or to describe it to others in terms of grayness or fuzziness; we also have difficulty reasoning with such information.

This tension between an enlightened, scientific view of maps as imperfect communication channels and a cognitive view based on an analysis of the limitations of human reasoning is central to debates over the role of new technologies. On the one hand, computer technology has the potential to remove the constraints of cartographic technique, opening the prospect of a brave new world in which the scientific knowledge of the soil scientist, for example, will be fully available to others through novel forms of representation. On the other hand, computer technologies designed to reflect our limited abilities for spatial reasoning can make maps easier to read, and can move map design into closer compatibility with human intuition. Clearly there is room for both viewpoints, depending on what one is willing to assume about the context of application. In the remainder of this chapter I look more specifically at some of the new ideas that have emerged with the development of computer-based tools for mapping and geographic information handling and consider their potential impact on our understanding of the world.

The New Geographic Information Technologies

Geographic Information Systems

The term geographic information system was coined in the 1960s. In Canada, Roger Tomlinson was leading the effort to build a computer-based system to handle the mass of geographic data created by the Canada Land Inventory (Tomlinson, Calkins, and Marble, 1976). A land inventory must focus on measures of land area, such as the acreage available for growing arable crops, measures that are difficult to obtain from maps by the conventional methods of dots and counting, or tracing around areas using a mechanical device called a planimeter, both of which provide estimates that are crude at best. To make matters worse, land-use planning requires the accurate estimation of areas of land having multiple characteristics, which would have to be obtained from the various maps in the Canada Land Inventory through the labor-intensive preparation of transparent overlays. In planning forest use, for example, one would want to know the capability of land not only for forestry, but also for recreation and agriculture. A digital computer offered a potentially cost-effective and much more accurate alternative. The term geographic information system was coined to describe the system's emphasis on the geographic dimension of data, and its ability to store, retrieve, and analyze a wide variety of data types through simple manipulations (see Maguire 1991, for a detailed discussion of definitions of GIS).

From these early beginnings, geographic information systems have grown to become a significant area of electronic data processing. Recent figures place the total annual U.S. expenditure on GIS software at around $450 million (Daratech 1994), and expenditures on associated hardware, data collection, maintenance, and analysis are far higher. Many of the research breakthroughs associated with GIS occurred

during the development of the early systems, particularly Tomlinson's Canada Geographic Information System (Tomlinson, Calkins, and Marble, 1976), and the products of the Harvard Laboratory for Computer Graphics and Spatial Analysis, which flourished in Cambridge during the 1970s. Commercial products incorporating these developments began to appear in the late 1970s, and by the late 1980s GIS research had spawned a series of conferences, magazines, societies, and university and college programs. The field continues to expand, with software sales rising by 20 percent or more annually, with no end in sight. GIS has been described as "simultaneously the telescope, the microscope, the computer, and the xerox machine of regional analysis and synthesis" (Abler 1988:137) and as "the biggest step forward in the handling of geographic information since the invention of the map" (Department of the Environment 1987:8).

A modern GIS such as ARC/INFO, a product of Environmental Systems Research Institute of Redlands, California, or Geographical Resources Analysis Support (GRASS), developed by the U.S. Army Construction Engineering Research Laboratory of Champaign-Urbana, Illinois, has facilities to store digital representations of a wide range of geographic features and types of geographic variation. Besides geometric representations, a GIS can also store a range of attributes of each feature (in other words information that is known about each feature and serves to distinguish it from others) and also a range of relationships between features, such as connectivity, adjacency, or proximity. Capabilities exist to transform data from one projection to another, to input data from a variety of different systems and devices, and to display data in cartographic form. In addition, the true power of GIS lies in its ability to analyze data by measuring areas, overlaying different data sets, or carrying out a range of standard types of spatial analysis or modeling. A GIS is a tool for conducting geographical analysis, just as a statistical package is a tool for conducting statistical analysis; both provide researchers with the means to implement a body of well-defined analytical methods.[1]

At its simplest, a GIS database provides a digital representation of the contents of one or more maps. Because features shown on maps are invariably crisp, with sharp, well-defined boundaries, creating digital representations has been possible by using combinations of points, lines, and areas. Points are represented by pairs of coordinates, lines by ordered sequences of points assumed to be connected by straight lines, and areas by ordered sequences of points forming the vertices of polygons. The term polyline is often used to describe this common approach to line representation, through the obvious analogy to polygon. Thus GIS data structures not only preserve the essential crispness of mapped geographic features but abstract them further by insisting on straight line segments where reality may exhibit smooth curves. Although polylines and polygons are appropriate ways of representing city streets or land parcels, in the case of a river, for example, the conventional GIS representation of its plan form shows little understanding of the effects of hydrologic processes, and in this sense one might argue that GIS data structures are less effective representations than maps, because a cartographer is able to draw a smooth curve. Similarly, in the case of land cover maps, GIS designers have typically preserved the sharp transitions between types that are conventional on land cover maps (for example, between forest and open grassland, or between conifer forest and deciduous forest), rather than search for representations that can portray spatially continuous transitions.

Remote Sensing

The earliest aerial photographs date from the last century, but the advent of earth-orbiting satellites in the 1960s opened an entirely new range of technological possibilities. Digital satellite images of earth made it possible to create databases of comprehensive global coverage, without concern for political boundaries. The military applications of remote sensing fueled much of its development, but civilian applications followed, enabling a new perspective on mapping.

Unlike maps, remotely sensed images are composed of rectangular picture elements, or pixels, each carrying a signal representing the radiation from that area of the earth's surface in a particular spectral range. Instruments such as the Thematic Mapper (TM), mounted on the Landsat satellite, return images with a high level of spatial resolution (approximately 30 meters in the case of TM) and overfly any area at regular intervals (nineteen days in the case of Landsat).

Every image from space is constructed from thousands of measurements of radiation intensity. Color and texture changes continuously, and it is not until the image is classified by assigning each pixel to one of a limited number of classes that it begins to resemble a map. Further processing is necessary, such as the filtering of "salt and pepper," or instances of isolated pixels surrounded by pixels of another class, before the world as seen from space begins to resemble the more familiar map representation.

When not inhibited by cloud cover, remote sensing is capable of identifying a wide range of geographic variables with varying degrees of certainty. In recent years data derived from remote sensing have been used as input to analyses and modeling carried out in GIS, where they can

be merged with information on features and characteristics not identifiable from space, such as demographic and economic variables, political boundaries, or land ownership. The problems of integrating GIS and remote sensing technologies (and reconciling inherently different perspectives on the world) include accuracy, incompatibility of definitions, and institutional issues (Star 1991).

Global Positioning Systems

In the past five years, remote sensing and GIS have been joined by a third digital geographic technology, now known as the global positioning system (GPS) (Leick 1990). Like remote sensing, its development was funded in large part by the military (Smith 1992 discusses the military role in the development of GIS). The system consists of a constellation of earth-orbiting satellites, each emitting precisely timed signals, which can be used by a ground receiver to compute position on the earth's surface. In military use, a simple GPS receiver can determine position to 32 meters with 90 percent reliability, but for civilian use the signals are corrupted and the 90 percent reliability distance is approximately 100 meters. Receiver technology has now advanced to the point where hand-held receivers the size of a pocket calculator are available for a few hundred dollars, and GPS boards are available for installation in personal computers.

Various techniques have been devised to improve the accuracy of GPS and to circumvent the corruption of the signal. Differential GPS requires use of two receivers, one at a fixed and known location and one roving, communicating by a radio frequency signal. With such systems, positional accuracies of 1 meter or better are achievable, and even higher accuracies are possible if one is willing to collect data at the unknown location for an extended time.

GPS is having a revolutionary effect on the profession of surveying. Traditionally, positions are established by taking measurements with respect to monuments of known position, and each nation maintains a hierarchy of such monuments, including a small number the positions of which are known with great accuracy. GPS allows position on the earth's surface to be determined directly, and its use is leading to the identification of numerous errors in previously determined positions, particularly in remote areas. For scientists working in comparatively featureless areas such as Antarctica, GPS is providing the first reliable way of fixing observations in space.

Like remote sensing, GPS has been integrated with GIS as a source of measured coordinates. Again like remote sensing, its use raises numerous issues of accuracy, professional practice, and terminology.

Taken together, the three new digital geographic information technologies—GIS, remote sensing, and GPS—amount to a massive change in the availability, reliability, and value of geographic information to society. This change has enormous implications for how we represent and understand the world.

New Ideas of Geographic Representation

All three of the geographic information technologies have opened new possibilities for geographic representation. As we have already seen, the representations available to the traditional cartographer, working with pen and ink, are limited, although one can argue that the effects of these constraints are, nevertheless, consistent with patterns of human reasoning. In the digital environment no such constraints exist, and representations are limited only by the imaginations of their creators. In this section, I review some of the key developments of the past three decades.

Features and Attributes

Maps can be designed to show the locations of features, and a number of techniques have been devised for differentiating features or displaying their attributes. One can replace points by symbols denoting, for example, towers or fire hydrants; one can color line symbols or code their characteristics through patterns of dots and dashes to distinguish roads from railroads, for example; and one can fill area symbols with cross hatching or color to denote land cover class or political status. But in each case there are strict limits to the number of attributes that can be displayed simultaneously for a given object. By combining color fill, cross hatching, and labels, it may be possible to show as many as five independent attributes of area objects, but this is, of course, far less than the number of attributes that may be available for certain types of areas. For example, hundreds of attributes may be available for each reporting zone used by the census, including average family income and incidence of a certain disease, but only a very small number can be shown simultaneously on the same map. The clever facial expressions shown in Figure 3.2 are perhaps the most ingenious method for displaying many attributes simultaneously within the constraints of mapping technology.

GIS offers a much richer environment. One commonly used technique is to allow the user to point to a specific feature on the screen, such as a census reporting zone, with a mouse or other pointing device. The database is then searched for attributes of the feature (for

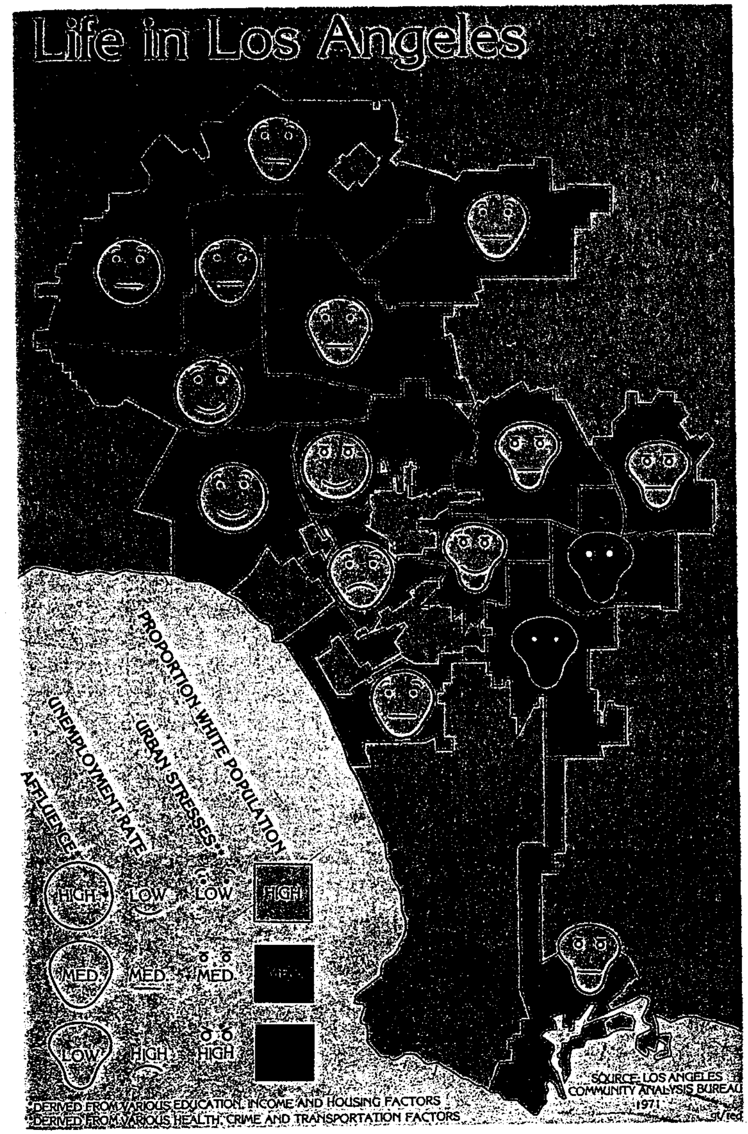

Figure 3.2. Facial expressions send many different messages. In this technique, developed by Chernoff (1973), caricatures of faces are used to display many more attributes than can normally be shown simultaneously on a map. Source: Muehrcke and Muehrcke (1992).

example, the age characteristics of the population over sixty-five in a county, or information about the owner of a parcel of land), and these are displayed in an unused part of the screen. In contrast to a printed map, this gives the display a dynamic aspect, allowing the user to change the visible information on demand. As a result, a far greater range of information can be made available than would be possible in the world of conventional mapping. Through this technique, the user can feel empowered and less dependent on the mapmaker's agenda. If taken too far, however, this new access to geographic detail can cross an invisible line to become an invasion of privacy. For example, it is possible to take data sets already available to the public, such as telephone directories and street maps, and create simple GIS applications that provide names and telephone numbers of house occupants from a map in response to a click of a mouse.

Hierarchical Access

Display of attributes on demand is one way of overcoming the space limitations of two-dimensional maps. Geographic reality has the interesting property that the more closely one looks at the surface of the earth, the more detail one sees (some teachers of geography have even been known to impress this fact upon their students by requiring them to map the intricacies of a 0.25 square meter plot of campus lawn). It is sometimes possible to predict this unraveling of detail, at least for certain cases of natural geographic objects such as shorelines, with some degree of accuracy. This property is so inconsistent with traditional thinking in geometry that it led Benoit Mandelbrot (1982) to propose a general theory of what he called fractals, or objects that display this kind of behavior. Of course map representations do not have this property; the detail that can be included for a given map scale is always limited, because it is difficult to represent objects if they would be smaller than approximately 0.5 millimeters across on the map, however large they might be on the surface of the earth. The Hitachi Corporation has demonstrated the potential of current etching technology by creating an image of the earth on a 5-centimeter wafer of silicon, with detail on the individual streets of London (at this scale a street is about ten atoms across), but such technical marvels remain quite useless.

Unlike a map, a spatial database has no scale, there being no distances inside a digital store to compare to distances on the earth's surface. Thus many GIS displays make use of a magnifying glass metaphor, by providing the user viewing a display at a given scale with an icon that allows any area of the map to be selectively enlarged.

Already-visible features can be shown in greater detail with more attributes, and new features that were invisible at the original scale of the display can be added. Of course this process is limited by the level of detail available in the database, but it provides a powerful new way of looking at geographic variation.

Mapmakers have met the need for views of the world at different scales by publishing multiple map series. Thus topographic maps of the United States are available at 1:24,000, 1:100,000 (for the conterminous United States), 1:250,000, and 1:2,000,000.[2] The definitions used and sets of features shown are different for each series, and the processes of mapmaking are largely independent of each other. As a result, many multi-scale GIS databases preserve this independence, providing no logical linkages among maps of the same area at different scales. For example, although digital representations of the railroad network are available at 1:100,000 and 1:2,000,000, the contents are so different that it is sometimes hard to tell that one is looking at the same area. Although much effort has gone into developing methods for automatically creating small-scale maps from large-scale maps and into "fusing" the information contained in both, it has proven exceedingly difficult to capture the complex and domain-specific processes of generalization used by cartographers into programs for digital computers (Buttenfield and McMaster 1991).

Instead, current GISs contain an increasing number of concepts of hierarchy, or explicit relationships between representations at different scales. The quadtree (Samet 1990a 1990b), for example, provides a method of representing variation at multiple scales that has no precedent or analog in traditional mapping. Figure 3.3 illustrates the construction of a simple quadtree representation of the spatial variation of a variable such as land cover class. If the entire area to be mapped contains more than one class, it is subdivided into four equal quadrants. Each is then examined to see if it is homogeneous with respect to land cover class, and if not, it is subdivided again. The process continues until the map consists of a series of quadrants of varying size, each of which is either homogeneous with respect to land cover class or is of the smallest size available given the spatial resolution of the data. Figure 3.3 shows the quadtree representation as a map and as a tree structure.

The quadtree is widely used as an economical representation that adapts the size of its basic geographic units to the degree of local variability in the mapped variable. It is used in the SPANS GIS (Tydac Technologies, Ottawa, Canada), in Samet's QUILT, and as a method of compressing images so that they can be transmitted over networks in

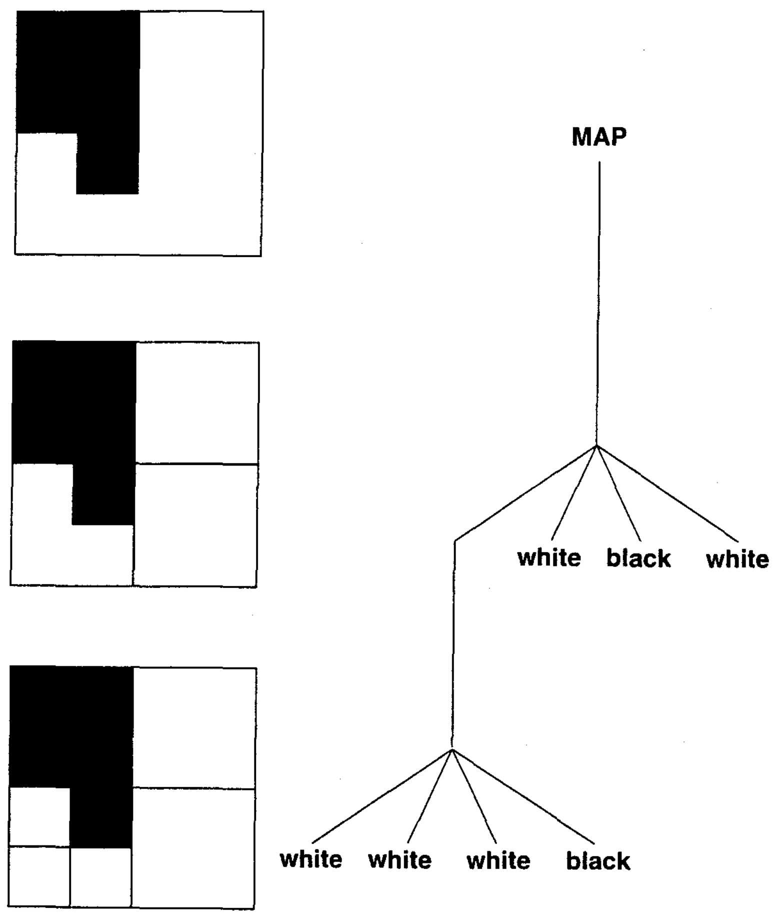

Figure 3.3. An example of a quadtree representation, as applied to a simple two-class (black and white) map. The map as a whole forms the root (top) of the tree and is first divided into four equal quadrants. If any quadrant includes more than one class, it is divided again into four, and the process continues until every subdivision contains only one class. Complex patterns require more levels in the tree than simple patterns, and the number of levels of the tree may also be determined by the basic cell size or resolution of the data.

less time. The quadtree concept has been generalized to the curved surface of the earth (Goodchild and Yang 1992), and similar tree-like structures have been devised for complex linear objects and for indexing discrete features (van Oosterom 1993). Recently there has been much interest in the concept of wavelets (Chui 1992), which can be seen as a generalization of the quadtree in which variation within each quadrant is described by a simple wave function. In all of these cases, a hierarchical tree is used to create a representation with information at all levels of spatial resolution, a concept that is radically different from the uniform resolution of traditional maps or earth images.

Hierarchical approaches to the structuring of geographic data are radically different from the pictorial perspectives provided by maps and images of uniform scale. They give us greater opportunity to explore interesting areas in more detail, or to step back and place areas in their regional context. Hierarchical structures give us access to the important local details that may not be visible at a given scale, and they emphasize the nature of generalized maps as approximations to the truth.

Scan Orders

Remotely sensed images convey information through assemblages of picture elements or pixels, and it is conventional to order such images by row from the top left corner, by analogy to the convention of many, but not all, written languages. This approach is echoed in the data structures of so-called raster GISs, which similarly represent spatial variation by dividing space into rectangular or square cells. In contrast, a vector GIS represents variation by describing the geometric form and attributes of discrete geographic features, either points, lines, or areas. For example, three approaches are commonly used to represent topographic elevation in GIS. The Digital Elevation Model (DEM) uses a square array or raster of point elevations. Digital Line Graph (DLG) representations, often obtained from topographic maps, capture topography in the vector format of digitized contour lines. Finally, Triangulated Irregular Networks (TINs), a second vector option, cover the surface with a mesh of irregular triangles, and assume that the surface varies linearly within each triangle.

Because vector databases prescribe no specific order for the features they contain, it is possible to think of them as scanning space in random order, by contrast to the systematic row-by-row scan of raster databases. Perhaps there is an analogy between vector databases and the action of the eye in scanning a work of art, as it views the image in a sequence that is subtly controlled by the artist. Conversely, a row-by-row scan is analogous to the way images are transmitted by television, or constructed by a computer display.

Early in the development of GIS, it became clear that one might achieve certain objectives by scanning space in orders that had no analogy in written text. In the Canada Geographic Information System (CGIS), data derived from each map sheet were stored on tape. The time required to find the data corresponding to a given map sheet could be substantial if it required winding through long lengths of tape or waiting for new tapes to be mounted. The system's designers argued that one could anticipate a certain pattern of requests in many applications, specifically that the next map needed was likely to be spatially adjacent to the one currently being processed. This principle seemed likely to be valid over a wide range of processes, from basic data editing to query and display. It followed that the optimum storage sequence would be one in which map sheets that were adjacent in space were most likely to be adjacent in storage.

Of course it is impossible to find a linear ordering of a two dimensional space that preserves spatial adjacencies, because a map sheet can be adjacent to at most two map sheets in a linear order but is mostly adjacent to four in space (defining adjacency in the sense of shared edges). The order devised by Guy Morton (1966) and illustrated in Figure 3.4 was selected as the best possible option and implemented in CGIS. It proves to be intimately related to the quadtree (Samet 1990a, 1990b), and Mark (1990) has compared it systematically to other scan orders using a variety of statistics. By coincidence, Morton is also the name of the southwesternmost county in Kansas.

The human eye is a remarkably effective processor of two-dimensional information, but other senses and methods of communication are no better than linear ones. Our administrative tools, such as filing cabinets, lists, and tables, are also largely linear. The development of the Morton scan order is just one example of how GIS research is stimulating new thinking about the relationship between linear and spatial approaches to the organization of information, and the importance of efficient methods for conversion between them.

Time

Because a map is static, cartographic representations clearly favor those aspects of geographic variation that are similarly static or change only slowly through time. Estes and Mooneyhan (1994) have recently commented on the age of much U.S. topographic mapping, which includes many sheets that have not been updated for more than fifteen years. Certain dimensions of topographic maps clearly deteriorate more rap-

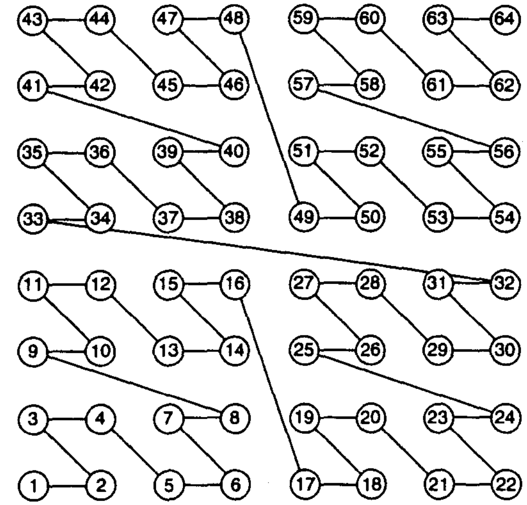

Figure 3.4. The Morton order, devised to ensure that areas close together in space are also stored close to each other in the database.

idly than others: although the physical form of the surface may remain more or less constant, cultural features are much more transitory.

Recently, there has been much interest in adding the temporal element to geographic databases. Remotely sensed images are snapshots taken at regular and frequent intervals, and sophisticated methods have been developed in GIS for detecting change between images, while allowing for the effects of seasons and sun angle. It seems that many of the functions currently performed by topographic maps could equally well be performed by suitably corrected air photos, and in recent years the U.S. Geological Survey has developed the necessary methods for producing digital orthophoto quadrangles (DOQs), which are air photos corrected to vertical perspective, with a pixel size of 1 meter. Because DOQs can be produced for less than one-tenth the cost of a standard topographic quadrangle, frequent updates in areas of rapid geographic change are cost effective. Relative to standard topographic maps they lack only the identification and interpretation of features. With increasing use of satellite images and DOQs, the orthographic but uninterpreted view from space is likely to become the basis of much of our perspective on geography in the future.

Langran (1992) discusses alternative approaches to the representation of temporal change. In the simplest case, time provides a series of snapshots, or layers of information in the GIS database, with no explicit logical relationships between the contents of adjacent time slices. In the second case, which is equally easily handled within the design of current GIS, features persist in fixed locations through time, but their attributes change in longitudinal series. Much census data follows this model, in that the decennial census provides a regular update on the contents of a series of fixed reporting zones. The third case is characterized by features with persistent identity, but with positions and attributes that change through time. Change of position may be described by movement parameters of speed and direction or by knowledge of the feature's position at fixed points in time. This model has been used to represent space-time behavior of individuals (Goodchild, Klinkenberg, and Janelle 1993) or populations of animals in zoological applications of GIS, but the functions and data model concepts necessary to support it fully do not yet exist.

In the most problematic case, features have identity at points in time, but identities fail to persist as features move, dissolve, coalesce, and split. Although it may be possible to count the number of clouds on a single image, it is clearly not possible to model clouds as discrete, well-defined features through time. Thus the only available approach to handling this type of data is to fall back on the simple time slice model, the first case discussed earlier. Households show analogous characteristics, as they emerge through marriage and other household formation processes, persist through time, and then dissolve through death, divorce, or other forms of fragmentation.

Although none of these approaches is in any way original as a model of time-dependent information, the treatment of time within GIS represents something of a conceptual breakthrough, given the static nature of traditional maps and the prevalence of the map metaphor within early GIS architectures. It remains to be seen whether the development of better methods for representing time in spatial databases will lead to a rethinking of associated methods of data collection. With new storage and processing technology, is it still appropriate to collect census data every ten years, or should we move to some more continuous scheme, perhaps based on a smaller sample?

Exploratory Spatial Data Analysis

By the late 1970s, the computer had become a widely used tool for statistical analysis, allowing the analyst to avoid the tedious calculations that had characterized early applications of such techniques as factor analysis. As the tool stimulated new ways of thinking, new methods of analysis were developed that capitalized on its power.

Nowhere was this more apparent than in the development of exploratory data analysis (EDA), a set of techniques pioneered by Tukey (1977) and others. Whereas conventional statistical analysis stressed the formulation of hypotheses, followed by formal testing, EDA offered tools for exploration of data as part of the process of hypothesis generation, together with new methods for visualizing samples from interesting perspectives. Today, the EDA paradigm is widely accepted, along with related approaches that engage the computational power of the digital environment.

The advent of GIS has raised the prospect of similar opportunities in spatial data analysis. Because maps are tedious and costly to produce, much traditional analysis of spatial data has tended to ignore the spatial component, treating reporting zones as if they were independent samples from a statistical population and adopting statistical techniques based on the assumption of independence.[3] The consequences have been documented repeatedly (see, for example, Openshaw 1983; Haining 1990; Anselin 1989).

Exploratory spatial data analysis (ESDA) applies the EDA approach to spatial analysis (see, for example, MacDougall 1992). These exploratory methods include tools to display multiple perspectives on data in the form of maps, time series, tables, and charts; active visual linkages between perspectives, such that pointing to a data element on a scatterplot will simultaneously highlight the same element on a map; hierarchical linkages between and among data that allow analysis at multiple scales; and animation. Fotheringham and Rogerson (1994) provide a review of recent GIS-based efforts to find novel approaches to spatial data analysis.

For decades, geographers have argued over whether the purpose of their research should be to search for general laws that apply everywhere on the earth's surface, or for understanding of the particular characteristics of places. The former paradigm is strongly associated with the scientific, quantitative philosophy of geography that flourished in the 1960s, while the latter is linked to an older focus on regional exploration and description and is currently echoed in postmodern thought. If laws are the same everywhere, then selecting a study area for research is analogous to selecting a sample and can be assumed to have minimal effect on the outcome. Most of the tools of statistical research that were widely adopted by geographers during the quantitative revolution of the late 1950s and 1960s were aspatial, ignoring the geographic locations of the cases being analyzed. But GIS tools capture geographic locations, giving the user the choice between search for general laws and exploration of geographic uniqueness. They allow us to look for exception to theories or to ask, for example, whether theories that seem to work for suburban areas also work for rural areas.

Uncertainty

Although techniques for expressing ignorance and uncertainty were commonly used in early maps (for example, the blank areas or "terra incognita" in the interior of Africa), modern mapping preserves few if any ways of communicating lack of knowledge or of warning the map reader of possible inaccuracy. The border between the Yemen and Saudi Arabia is undefined and is commonly shown as dashed; intermittent streams are marked with the same dashed symbol to indicate uncertainty over the presence of water. Dashes are also commonly used to indicate planned features or features under construction. With these few exceptions, maps continue to reassure us that the information they present is certain.

In reality, the quality of the information shown on maps may be uneven and uncertain. Soil maps, for example, show areas of uniform class separated by thin lines representing sharp transition, although it is clear that transitions are often far from sharp, and areas are often far from homogeneous. Such information may be available to the informed reader in the legend, information that often accompanies the map, or in separately published data-quality statements. But when the map is digitized into a GIS database, such qualifications are unlikely to be readily available to the data's eventual user.

To be fair, many maps were never intended to be repositories of scientifically objective measurements. Contours drawn on topographic maps are intended to convey an impression of the form of the surface, not to capture precise elevations. But the unwary user of a GIS database may well believe that the response received to a query about elevation at a point is indeed a scientific measurement, with an accuracy implied by the degree of detail reported by the system. In such strange ways does the GIS confer accuracy and authority on the data that it contains.

Recent thinking on the subject of data quality in GIS has converged on visualization as the appropriate method for communicating a meaningful sense of qualification to the user. Colors can be diluted toward gray to indicate lack of confidence in attributes. Contour lines in a computer display can be broadened or blurred to reflect the positional uncertainty that is a logical consequence of uncertainty in knowledge of elevation. Boundaries between classes on soil or land cover maps can be blurred to indicate transition zones, and randomly shaped inclusions can be placed within patches to indicate knowledge

of heterogeneity. The important message that such random inclusions are not the truth, but one possible version of the truth, can be conveyed by displaying a series of possible maps on the screen, or by animation. In effect, research is reverting to the days when mapmakers routinely portrayed uncertainty by leaving areas blank, but a GIS can convey uncertainty in a more rigorous and useful fashion.

Conclusion

New tools have often provoked new thoughts, and recent digital geographic information technologies are no exception. Traditional technologies used to make maps impose constraints on how their users see the world; those constraints may coincide with an innate human desire for a simpler, more ordered environment. The newer digital technologies are not so constrained, and so open up a rich new set of options for representing the world. As a result, new kinds of thinking have emerged through the interplay among geographic information technologies, their developers, and their users.

Even the most elegant cartographic techniques allow only a small number of attributes to be displayed for each feature shown on a map, and map scale imposes limits on the sizes of features that can be shown. Because a digital database has no scale, it imposes no limits on the density of information. A digital store does not require that large amounts of space be devoted to empty desert, nor does it require simplifying the complexity of dense urban areas, or limiting the amount of information that can be associated with a single feature. Maps tend to emphasize the horizontal relationships among features of comparable extent and to equate extent with importance. A geographical database can give access to the relationships between large and small features and can draw attention to the importance of the small and particular as well as the large and general.

Maps are inherently two-dimensional and static, whereas digital technologies can capture the three-dimensional and time-dependent aspects of geographic variation. A two-dimensional map cannot show vertical differentiation in the atmosphere, or the weather effects that result, or the behavior of a three-dimensional plume of polluting hydrocarbons underneath a leaking underground fuel tank. The ability to handle three-dimensional data in digital geographic databases will encourage us to investigate the subsurface environment and to think about the atmosphere as a three-dimensional system. Adding the temporal dimension to databases will encourage us to think in terms of continuous change in the geographic distribution of human populations, rather than the periodic time slices provided by the traditional census.

GIS and the new techniques of exploratory spatial data analysis allow us to examine information in its geographical context and to ask how conditions change from one area to another. New methods of spatial analysis, such as the technique developed by Openshaw and his colleagues to search for clusters of disease, emphasize the importance of exploration, and the need to combine the intuitive power of the human eye to search for patterns with the rigor of statistical analysis (Openshaw et al. 1987).

Finally, the simple action of analyzing large amounts of geographic data of comparatively low accuracy in very precise computing machines has forced us to think hard about the role of maps as repositories of scientific measurements. We clearly do not know the world as accurately as we thought we did. This realization is forcing us to rediscover earlier methods of communicating uncertainty and to invent new ones in order to caution the users of geographic data technologies against ascribing excessive accuracy to their results. Once again, the world is proving to be not as simple as our maps had led us to believe. The impact on our map-based systems of land ownership, regulation, and taxation remains to be seen.

The ability to store the contents of maps in computers will continue to suggest new ways of thinking about the world for years to come. In this sense GIS has only just begun to influence the myriad ways in which society makes use of geographic information. We have yet to see how the ability to display digital road maps in automobiles will influence road map design or the way people explore and navigate in strange environments. Will the widespread adoption of multimedia computers in American homes lead to a richer knowledge of the rest of the world or one that is controlled even more than at present by simple stereotypes? The ideas that have emerged from GIS are currently limited to the research community and the agencies that have had the resources to acquire the technology. But we are now entering an era of mass marketing of GIS, using simpler tools and aiming at broader applications. The impacts of these tools will make fascinating topics of study for geographers (Pickles 1995).

The emphasis in this chapter has almost inevitably been on the novel technical ideas inherent in GIS, because these have dominated the field over the past three decades as researchers have worked to make GIS technically feasible and cost effective. In the future, the ideas that emerge from the technology are likely to have a stronger social flavor, as we begin to see the impact of GIS on society at large. Much has already been written on the power of information to change society, and a small part of this literature is specifically concerned with geographic data (Obermeyer and Pinto 1994; King and Kraemer 1993). In the future, we can expect a strong tie to national efforts to coordinate the creation and use of geographic information, particularly the National Spatial Data Infrastructure (National Research Council 1993), to which the Clinton administration appears strongly committed.

But these future prospects aside, the GIS tool and its relatives have already provoked their share of new thoughts and new ways of thinking about the geographical world. Some of these will be enlightening, while others may be as constraining as pencil and paper. After all, to the person with nothing but a hammer, everything sooner or later begins to look like a nail. Similarly, an overly simple GIS can encourage a restricted view of a world in which places are reduced to points, lines, and areas. Only by working to improve the tool can we ensure that its effects continue to be stimulating.

Notes

The National Center for Geographic Information and Analysis is supported by the National Science Foundation, under Cooperative Agreement SBR 8810917.

1. Introductions to geographic information systems are provided by Star and Estes (1990), Burrough (1986), Laurini and Thompson (1992), and many others. Maguire, Goodchild, and Rhind (1991) provide a comprehensive review of the technology.

2. These numbers refer to the ratio between distances on the map and distances on the earth's surface. For example, one inch on a 1:250,000 map equals 250,000 inches on the ground.

3. Suppose, for example, you were studying the relationship between precipitation and elevation in Vermont, and your reporting units are towns. Traditional statistical methods, such as calculating the correlation between precipitation and elevation, assume that observations are independent, that is, the precipitation and elevation in one town are unrelated to those for other towns. Yet we know that a town's elevation and precipitation levels are, in fact, more closely related to levels in nearby neighboring towns than to those in far away towns. The towns should therefore not be treated as independent observations.

References

Abler, R. F. 1988. Awards, Rewards, and Excellence: Keeping Geography Alive and Well. *The Professional Geographer* 40(2):135–140.

Anselin, L. 1989. *What Is Special about Spatial Data? Alternative Perspectives on Spatial Data Analysis*. Technical Paper 89–4. Santa Barbara, Calif.: National Center for Geographic Information and Analysis.

Burrough, P. A. 1986. *Principles of Geographical Information Systems for Land Resources Assessment.* Oxford: Clarendon.

Buttenfield, B. P., and R. B. McMaster, eds. 1991. *Map Generalization: Making Rules for Knowledge Representation.* London: Longman.

Chernoff, H. 1973. The Use of Faces to Represent Points in k-dimensional Space Graphically. *Journal of the American Statistical Association* 70:548–554.

Chui, C. K. 1992. *An Introduction to Wavelets.* Boston: Academic Press.

Daratech. 1994. New GIS Market Study. *Geodetical Info Magazine* (May):36–37.

Department of the Environment. 1987. *Handling Geographic Information: Report of the Committee of Enquiry Chaired by Lord Chorley.* London: Her Majesty's Stationery Office.

Estes, J. E., and D. W. Mooneyhan. 1994. Of Maps and Myths. *Photogrammetric Engineering and Remote Sensing* 60(5):517–524.

Fotheringham, A. S., and P. A. Rogerson, eds. 1994. *Spatial Analysis and GIS.* London: Taylor and Francis.

Goodchild, M. F. 1988. Stepping Over the Line: Technological Constraints and the New Cartography. *The American Cartographer* 15(3):311–319.

Goodchild, M. F., and S. R. Yang. 1992. A Hierarchical Spatial Data Structure for Global Geographic Information Systems. *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing* 54(1):31–44.

Goodchild, M. F., B. Klinkenberg, and D. G. Janelle. 1993. A Factorial Model of Aggregate Spatio-temporal Behavior: Application to the Diurnal Cycle. *Geographical Analysis* 25(4): 277–294.

Haining, R. P. 1990. *Spatial Data Analysis in the Social and Environmental Sciences.* New York: Cambridge University Press.

Harley, J. B., and D. Woodward, eds. 1987. *The History of Cartography.* Chicago: University of Chicago Press.

King, J. L., and K. L. Kraemer. 1993. Models, Facts, and the Policy Process: The Political Ecology of Estimated Truth. In *Environmental Modeling with GIS,* eds. M. F. Goodchild, B. O. Parks, and L. T. Steyaert, 353–360. New York: Oxford University Press.

Langran, G. 1992. *Time in Geographic Information Systems.* London: Taylor and Francis.

Laurini, R., and D. Thompson. 1992. *Fundamentals of Spatial Information Systems.* San Diego: Academic Press.

Leick, A. 1990. *GPS Satellite Surveying.* New York: Wiley.

MacDougall, E. B. 1992. Exploratory Analysis, Dynamic Statistical Visualization, and Geographic Information Systems. *Cartography and Geographic Information Systems* 19(4):237–246.

Maguire, D. J. 1991. An Overview and Definition of GIS. In *Geographical Information Systems: Principles and Applications,* eds. D. J. Maguire, M. F. Goodchild, D. W. Rhind, 9–20. London: Longman.

Maguire, D. J., M. F. Goodchild, and D. W. Rhind, eds. 1991. *Geographical Information Systems: Principles and Applications.* London: Longman.

Mandelbrot, B. B. 1982. *The Fractal Geometry of Nature.* San Francisco: Freeman.

Mark, D. M. 1990. Neighbor-based Properties of Some Orderings of Two-dimensional Space. *Geographical Analysis* 22(2):145–157.

Mark, D. M., and A. U. Frank, eds. 1991. *Cognitive and Linguistic Aspects of Geographic Space.* Dordrecht: Kluwer.

Morton, G. M. 1966. *A Computer Oriented Geodetic Data Base and New Technique in File Sequencing.* Ottawa: IBM Canada.

Muehrcke, P. C., and J. O. Muehrcke. 1992. *Map Use: Reading, Analysis, and Interpretation.* Madison: JP Publications.

National Research Council. 1993. *Toward a Coordinated Spatial Data Infrastructure for the Nation.* Washington, D.C.: National Academies Press.

Obermeyer, N. J., and J. K. Pinto. 1994. *Managing Geographic Information Systems.* New York: Guilford.

Openshaw, S. 1983. *The Modifiable Areal Unit Problem.* Norwich: Geobooks.

Openshaw, S., M. Charlton, C. Wymer, and C. Craft. 1987. A Mark I Geographical Analysis Machine for the Automated Analysis of Past Data Sets. *International Journal of Geographical Information Systems* 1:335–358.

Pickles, J. 1995. *Ground Truth: The Social Implications of Geographical Information Systems.* New York: Guilford.

Samet, H. 1990a. *Applications of Spatial Data Structures: Computer Graphics, Image Processing, and GIS.* Reading, Mass.: Addison-Wesley.

———. 1990b. *The Design and Analysis of Spatial Data Structures.* Reading, Mass.: Addison-Wesley.

Smith, N. 1992. History and Philosophy of Geography: Real Wars, Theory Wars. *Progress in Human Geography* 16(2):257–271.

Star, J. L., ed. 1991. *The Integration of Remote Sensing and Geographic Information Systems.* Bethesda: American Society for Photogrammetry and Remote Sensing.

Star, J. L., and J. E. Estes. 1990. *Geographic Information Systems: An Introduction.* Englewood Cliffs, N.J.: Prentice-Hall.

Tomlinson, R. F., H. W. Calkins, and D. F. Marble. 1976. *Computer Handling of Geographical Data.* Paris: UNESCO.

Tukey, J. W. 1977. *Exploratory Data Analysis.* Reading, Mass.: Addison-Wesley.

Van Oosterom, P.J.M. 1993. *Reactive Data Structures for Geographic Information Systems.* New York: Oxford University Press.

Wood, D., with J. Fels. 1992. *The Power of Maps.* New York: Guilford Press.