Chapter 1

INTRODUCTION

Not only is it easy to lie with maps, it's essential. To portray meaningful relationships for a complex, three-dimensional world on a flat sheet of paper or a video screen, a map must distort reality. As a scale model, the map must use symbols that almost always are proportionally much bigger or thicker than the features they represent. To avoid hiding critical information in a fog of detail, the map must offer a selective, incomplete view of reality. There's no escape from the cartographic paradox: to present a useful and truthful picture, an accurate map must tell white lies.

Because most map users willingly tolerate white lies on maps, it's not difficult for maps also to tell more serious lies. Map users generally are a trusting lot: they understand the need to distort geometry and suppress features, and they believe the cartographer really does know where to draw the line, figuratively as well as literally. As with many things beyond their full understanding, they readily entrust mapmaking to a priesthood of technically competent designers and drafters working for government agencies and commercial firms. Yet cartographers are not licensed, and many mapmakers competent in commercial art or the use of computer workstations have never studied cartography. Map users seldom, if ever, question these authorities, and they often fail to appreciate the map's power as a tool of deliberate falsification or subtle propaganda.

Because of personal computers and electronic publishing, map users can now easily lie to themselves—and be unaware of it. Before the personal computer, folk cartography consisted largely of hand-drawn maps giving directions. The direction giver had full control over pencil and paper and usually had no difficulty transferring routes, landmarks, and other relevant recollections from mind to map. The computer allows programmers, marketing experts, and other anonymous middlemen without cartographic savvy to strongly influence the look of the map and gives modern-day folk maps the crisp type, uniform symbols, and verisimilitude of maps from the cartographic priesthood. Yet software developers commonly have made it easy for the lay cartographer to select an inappropriate projection or a misleading set of symbols. Because of advances in low-cost computer graphics, inadvertent yet serious cartographic lies can appear respectable and accurate.

The potential for cartographic mischief extends well beyond the deliberate suppression used by some cartographer-politicians and the electronic blunders made by the cartographically ignorant. If any single caveat can alert map users to their unhealthy but widespread naïveté, it is that *a single map is but one of an indefinitely large number of maps that might be produced for the same situation or from the same data*. The italics reflect an academic lifetime of browbeating undergraduates with this obvious but readily ignored warning. How easy it is to forget, and how revealing to recall, that map authors can experiment freely with features, measurements, area of coverage, and symbols and can pick the map that best presents their case or supports their unconscious bias. Map users must be aware that cartographic license is enormously broad.

The purpose of this book is to promote a healthy skepticism about maps, not to foster either cynicism or deliberate dishonesty. In showing how to lie with maps, I want to make readers aware that maps, like speeches and paintings, are authored collections of information and also are subject to distortions arising from ignorance, greed, ideological blindness, or malice.

Examining the misuses of maps also provides an interesting introduction to the nature of maps and their range of appropriate uses. Chapter 2 considers as potential sources of distortion the map's main elements: scale, projection, and symbolization. Chapter 3 further pursues the effects of scale by examining the various white lies cartographers justify as necessary generalization, and chapter 4 looks at common blunders resulting from the mapmaker's ignorance or oversight. Chapter 5 treats the seductive use of symbols in advertising maps, and chapter 6 explores exaggeration and sup-

pression in maps prepared for development plans and environmental impact statements. Chapters 7 and 8 examine distorted maps used by governments as political propaganda and as "disinformation" for military opponents. The next two chapters are particularly relevant to users of mapping software and electronic publishing: chapter 9 addresses distortion and self-deception in statistical maps made from census data and other quantitative information, and chapter 10 looks at how a careless or Machiavellian choice of colors can confuse or mislead the map viewer. Chapter 11 concludes by noting maps' dual and sometimes conflicting roles and by recommending a skeptical assessment of the map author's motives.

A book about how to lie with maps can be more useful than a book about how to lie with words. After all, everyone is familiar with verbal lies, nefarious as well as white, and is wary about how words can be manipulated. Our schools teach their pupils to be cautious consumers who read the fine print and between the lines, and the public has a guarded respect for advertising, law, marketing, politics, public relations, writing, and other occupations requiring skill in verbal manipulation. Yet education in the use of maps and diagrams is spotty and limited, and many otherwise educated people are graphically and cartographically illiterate. Maps, like numbers, are often arcane images accorded undue respect and credibility. This book's principal goal is to dispel this cartographic mystique and promote a more informed use of maps based upon an understanding and appreciation of their flexibility as a medium of communication.

The book's insights can be especially useful for those who might more effectively use maps in their work or as citizens fighting environmental deterioration or social ills. The informed skeptic becomes a perceptive map author, better able to describe locational characters and explain geographic relationships as well as better equipped to recognize and counter the self-serving arguments of biased or dishonest mapmakers.

Where a deep mistrust of maps reflects either ignorance of how maps work or a bad personal experience with maps, this book can help overcome an unhealthy skepticism called *cartophobia*. Maps need be no more threatening or less reliable than words, and rejecting or avoiding or ignoring maps is akin to the mindless fears of illiterates who regard books as evil or dangerous. This book's revelations about how maps *must* be white lies but may *sometimes* become real lies should provide the same sort of reassuring knowledge that allows humans to control and exploit fire and electricity.

Chapter 2

Elements of the Map

Maps have three basic attributes: scale, projection, and symbolization. Each element is a source of distortion. As a group, they describe the essence of the map's possibilities and limitations. No one can use maps or make maps safely and effectively without understanding map scales, map projections, and map symbols.

Scale

Most maps are smaller than the reality they represent, and map scales tell us how much smaller. Maps can state their scale in three ways: as a ratio, as a short sentence, and as a simple graph. Figure 2.1 shows some typical statements of map scale.

Ratio scales relate one unit of distance on the map to a specific distance on the ground. The units must be the same, so that a ratio of 1:10,000 means that a 1-inch line on the map represents a 10,000-inch stretch of road—or that 1 centimeter represents 10,000 centimeters or 1 foot stands for 10,000 feet. As long as they are the same, the units don't matter and need not be stated; the ratio scale is a dimensionless number. By convention, the part of the ratio to the left of the colon is always 1.

Some maps state the ratio scale as a fraction, but both forms have the same meaning. Whether the mapmaker uses 1:24,000 or 1/24,000 is solely a matter of style.

Fractional statements help the user compare map scales. A scale of 1/10,000 (or 1:10,000) is larger than a scale of 1/250,000 (or 1:250,000) because 1/10,000 is a larger fraction than 1/250,000. Recall that small fractions have big denomi-

Ratio Scales	Verbal Scales
1:9,600	One inch represents 800 feet.
1:24,000	One inch represents 2,000 feet.
1:50,000	One centimeter represents 500 meters.
1:250,000	One inch represents (approximately) 4 miles.
1:2,000,000	One inch represents (approximately) 32 miles, one centimeter represents 20 kilometers.

Graphic Scales

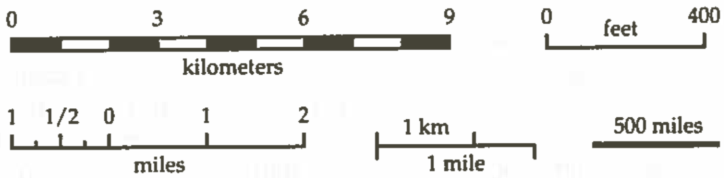


FIGURE 2.1. Types of map scales.

nators and big fractions have small denominators, or that half (1/2) a pie is more than a quarter (1/4) of the pie. In general, "large-scale" maps have scales of 1:24,000 or larger, whereas "small-scale" maps have scales of 1:500,000 or smaller. But these distinctions are relative: in a city planning office where the largest map scale is 1:50,000, "small-scale" might refer to maps at 1:24,000 or smaller and "large-scale" to maps at 1:4,800 or larger.

Large-scale maps tend to be more detailed than small-scale maps. Consider two maps, one at 1:10,000 and the other at 1:10,000,000. A 1-inch line at 1:10,000 represents 10,000 inches, which is 833 1/3 feet, or roughly 0.16 miles. At this scale a square measuring 1 inch on each side represents an area of .025 mi^2, or roughly 16 acres. In contrast, at 1:10,000,000 the 1-inch line on the map represents almost 158 miles, and the square inch would represent an area slightly over 24,900 mi^2, or nearly 16 million acres. In this example the square inch on the large-scale map could show features on the ground in far greater detail than the square inch on the small-scale map. Both maps would have to suppress some details, but the designer of the 1:10,000,000-scale map must be far more selective than the cartographer producing the 1:10,000-scale map. In the sense that all maps tell white lies about the planet, small-

scale maps have a smaller capacity for truth than large-scale maps.

Verbal statements such as "one inch represents one mile" relate units convenient for measuring distances on the map to units commonly used for estimating and thinking about distances on the ground. For most users this simple sentence is more meaningful than the corresponding ratio scale of 1:63,360, or its close approximation, 1:62,500. British map users commonly identify various map series with adjective phrases such as "inch to the mile" or "four miles to the inch" (a close approximation for 1:250,000).

Sometimes a mapmaker might say "equals" instead of "represents." Although technically absurd, "equals" in these cases might more kindly be considered a shorthand for "is the equivalent of." Yet the skeptic rightly warns of cartographic seduction, for "one inch equals one mile" not only robs the user of a subtle reminder that the map is merely a symbolic model but also falsely suggests that the mapped image is reality. As later chapters show, this delusion can be dangerous.

Metric units make verbal scales less necessary. Persons familiar with centimeters and kilometers have little need for sentences to tell them that at 1:100,000, one centimeter represents one kilometer, or that at 1:25,000 four centimeters represent one kilometer. In Europe, where metric units are standard, round-number map scales of 1:10,000, 1:25,000, 1:50,000, and 1:100,000 are common. In the United States, where the metric system's most prominent inroads have been in the liquor and drug businesses, large-scale maps typically represent reality at scales of 1:9,600 ("one inch represents 800 feet"), 1:24,000 ("one inch represents 2,000 feet"), and 1:62,500 ("one inch represents [slightly less than] one mile").

Graphic scales are not only the most helpful means of communicating map scale but also the safest. An alternative to blind trust in the user's sense of distance and skill in mental arithmetic, the simple bar scale typically portrays a series of conveniently rounded distances appropriate to the map's function and the area covered. Graphic scales are particularly safe when a newspaper or magazine publisher might reduce or enlarge the map without consulting the mapmaker. For example, a five-inch-wide map labeled "1:50,000" would have a scale less than 1:80,000 if reduced to fit a newspaper column three inches wide, whereas a scale bar representing a half-mile would shrink along with the map's other symbols and distances. Ratio and verbal scales are useless on video maps, since television screens and thus the map scales vary widely and unpredictably.

Map Projections

Map projections, which transform the curved, three-dimensional surface of the planet into a flat, two-dimensional plane, can greatly distort map scale. Although the globe can be a true scale model of the earth, with a constant scale at all points and in all directions, the flat map stretches some distances and shortens others, so that scale varies from point to point. Moreover, scale at a point tends to vary with direction as well.

The world map projection in figure 2.2 illustrates the often severe scale differences found on maps portraying large areas. In this instance map scale is constant along the equator and the meridians, shown as straight lines perpendicular to the equator and running from the North Pole to the South Pole. (If the terms *parallel*, *meridian*, *latitude*, and *longitude* seem puzzling, the quick review of basic world geography found in the Appendix might be helpful.) Because the meridians have the same scale as the equator, each meridian (if we assume the earth is a *perfect* sphere) is half the length of the equator. Because scale is constant along the meridians, the map preserves the even spacing of parallels separated by 30° of latitude. But on this map all parallels are the same length, even though on the earth or a globe parallels decrease in length from the equator to the poles. Moreover, the map projection has stretched the poles from points with no length to lines as long as the equator. North-south scale is constant, but east-west scale increases to twice the north-south scale at 60° N and 60° S, and to infinity at the poles.

Ratio scales commonly describe a world map's capacity for detail. But the scale is strictly valid for just a few lines on the map—in the case of figure 2.2, only for the equator and the meridians. Most world maps don't warn that using the scale ratio to convert distances between map symbols to distances between real places almost always yields an erroneous result. Figure 2.2, for instance, would greatly inflate the distance

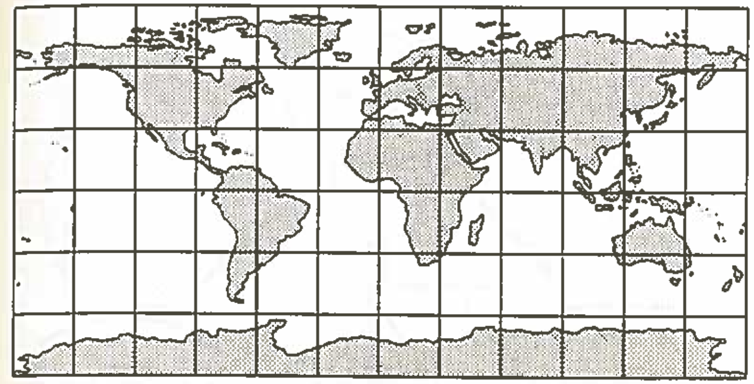


FIGURE 2.2. Equatorial cylindrical projection with true meridians.

between Chicago and Stockholm, which are far apart and both well north of the equator. Cartographers wisely avoid decorating world maps with graphic scales, which might encourage this type of abuse. In contrast, scale distortion of distance usually is negligible on large-scale maps, where the area covered is comparatively small.

Figure 2.3 helps explain the meaning and limitations of ratio scales on world maps by treating map projection as a two-stage process. Stage one shrinks the earth to a globe, for which the ratio scale is valid everywhere and in all directions. Stage two projects symbols from the globe onto a flattenable surface, such as a plane, a cone, or a cylinder, which is attached to the globe at a point or at one or two *standard lines*. On flat maps, the scale usually is constant only along these standard lines. In figure 2.2, a type of cylindrical projection called the *plane chart*, the equator is a standard line and the meridians show true scale as well.

In general, scale distortion increases with distance from the standard line. The common *developable surfaces*—plane, cone, and cylinder—allow the mapmaker to minimize distortion by centering the projection in or near the region featured on the map. World maps commonly use a cylindrical projection, centered on the equator. Figure 2.4 shows that a *secant* cylindrical projection, which cuts through the globe, yields two standard lines, whereas a *tangent* cylindrical projection, which merely touches the globe, has only one. Average distortion is less for a secant projection because the average place is closer

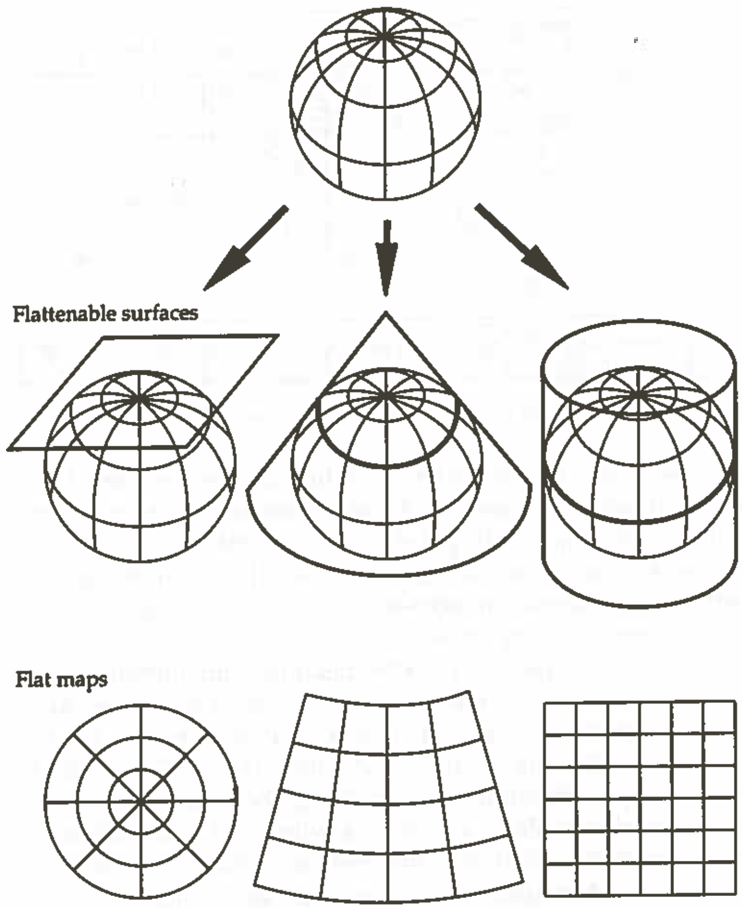


FIGURE 2.3. Developable surfaces in the second stage of map projection.

to one of the two standard lines. Conic projections are well suited to large mid-latitude areas, such as North America, Europe, and the Soviet Union, and secant conic projections offer less average distortion than tangent conic projections. *Azimuthal* projections, which use the plane as their developable surface, are used most commonly for maps of polar regions.

For each developable surface, the mapmaker can choose among a variety of projections, each with a unique pattern of distortion. Some projections, called *equivalent* or *equal-area*, allow

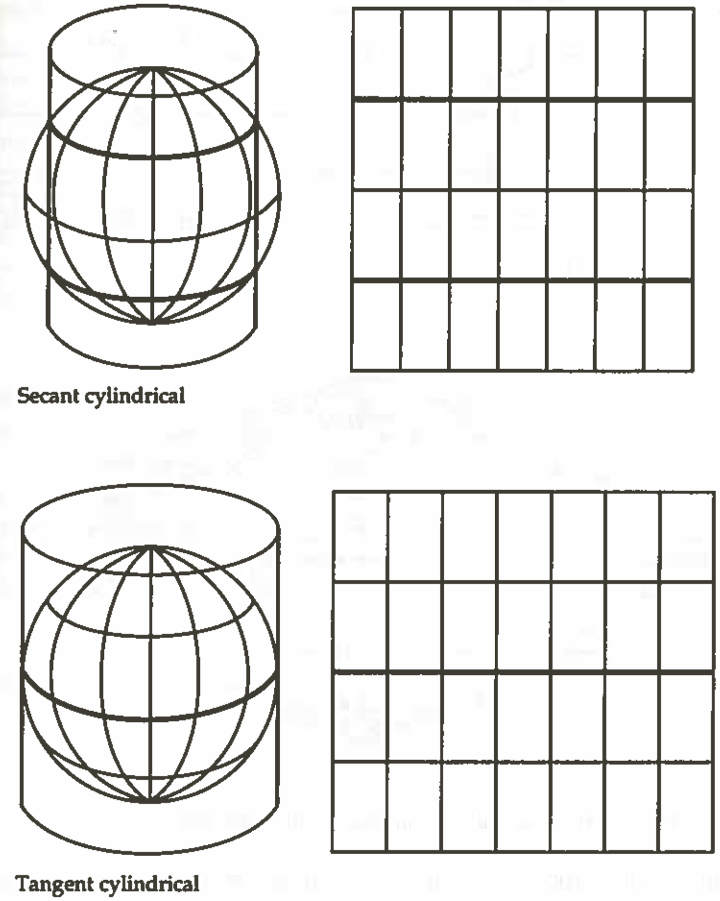
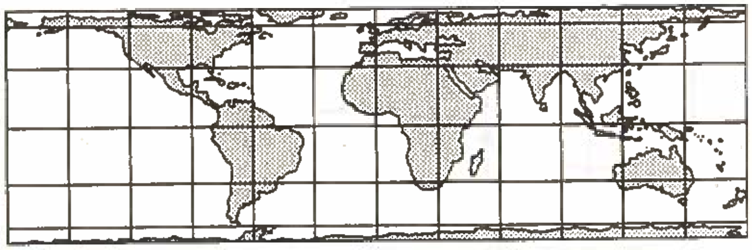
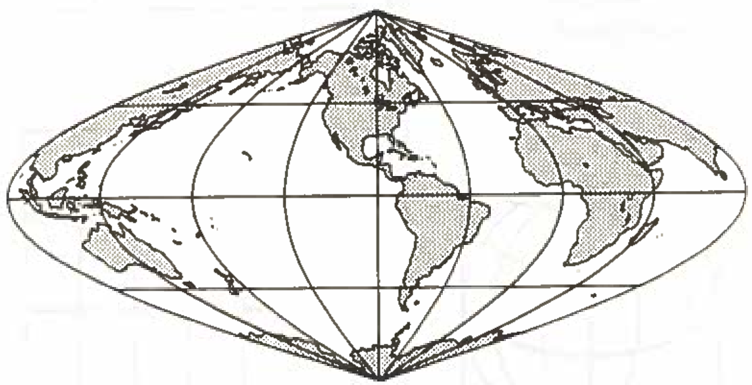


FIGURE 2.4. Secant (above) and tangent (below) cylindrical projections.

the mapmaker to preserve areal relationships. Thus if South America is eight times larger than Greenland on the globe, it will also be eight times larger on an equal-area projection. Figure 2.5 shows two ways to reduce the areal distortion of the plane chart (fig. 2.2). The cylindrical equal-area projection at the top compensates for the severe poleward exaggeration by reducing the separation of the parallels as distance from the equator increases. In contrast, the sinusoidal projection below maintains true scale along the equator, all other parallels, and the central meridian and at the same time pulls the



Cylindrical equal-area projection



Sinusoidal projection

FIGURE 2.5. Two varieties of equal-area cylindrical projection.

meridians inward, toward the poles, compensating for the areal exaggeration that would otherwise occur. Distortion is least pronounced in a cross-shaped zone along the equator and the central meridian and most severe between these axes toward the edge of the projection. Despite the highly distorted shapes in these "corners," the areas of continents, countries, and belts between adjoining parallels are in correct proportion.

Reduced distortion around the central meridian suggests that a sinusoidal projection "centered" on a meridian through, say, Kansas might yield a decent equal-area representation of North America, whereas a sinusoidal projection with a straight-line central meridian passing between Warsaw and Moscow would afford a suitable companion view of the Eurasian land mass. In the early 1920s, University of Chicago geography

professor J. Paul Goode extended this notion of a zoned world map and devised the composite projection in figure 2.6. Goode's Interrupted Homolosine Equal-Area projection has six lobes, which join along the equator. To avoid severe pinching of the meridians toward the poles, Goode divided each lobe into two zones at about 40°—an equatorial zone based on the sinusoidal projection and a poleward zone in which the equal-area Mollweide projection portrays high-latitude areas with less east-west compression. Goode's projection mollifies the trade-off of more distorted shapes for true relative areas by giving up continuous oceans for less severely distorted land masses. If interrupted over the land to minimize distortion of the oceans, Goode's projection can be equally adept in serving studies of fisheries and other marine elements.

No flat map can match the globe in preserving areas, angles, gross shapes, distances, and directions, and any map projection is a compromise solution. Yet Goode's projection is a particularly worthy compromise when the mapmaker uses dot symbols to portray the worldwide density pattern of population, hogs, wheat, or other dryland variables. On a dot-distribution map with one dot representing 500,000 swine, for example, the spacing of these dots represents relative density. Important hog-producing regions, such as the American Midwest and northern Europe, have many closely spaced dots, whereas hog-poor regions such as India and Australia have few. But a projection that distorts area might show contrasting densities for two regions of equal size on the globe and with similar

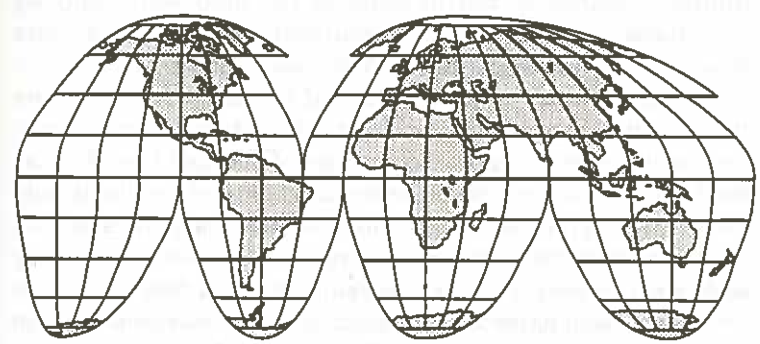


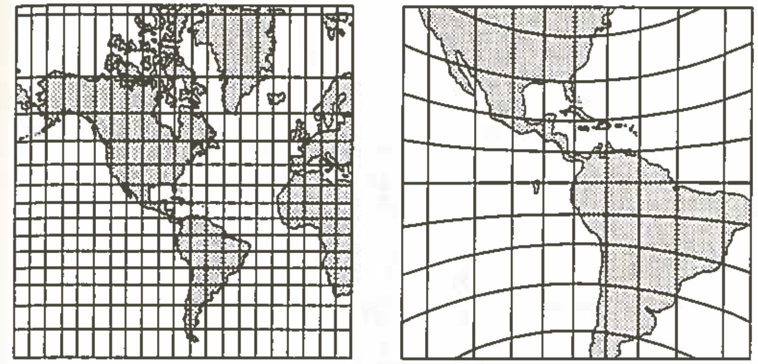
FIGURE 2.6. Goode's Homolosine Equal-Area projection.

levels of hog production; if both regions had 40 dots representing 20 million swine, the region occupying 2 cm² of the map would have a greater spacing between dots and appear less intensively involved in raising pigs than the region occupying only 1 cm². Projections that are not equal-area encourage such spurious inferences. Equivalence is also important when the map user might compare the sizes of countries or the areas covered by various map categories.

As equal-area projections preserve areas, *conformal* projections preserve local angles. That is, on a conformal projection the angle between any two intersecting lines will be the same on both globe and flat map. By compressing three-dimensional physical features onto a two-dimensional surface, a conformal projection can noticeably distort the shapes of long features, but within a small neighborhood of the point of intersection, scale will be the same in all directions and shape will be correct. Thus tiny circles on the globe remain tiny circles on a conformal map. As with all projections, though, scale still varies from place to place, and tiny circles identical in size on the globe can vary markedly in size on a conformal projection covering a large region. Although all projections distort the shapes of continents and other large territories, in general a conformal projection offers a less distorted picture of gross shape than a projection that is not conformal.

Perhaps the most striking trade-off in map projection is between conformality and equivalence. Although some projections distort both angles and areas, no projection can be both conformal and equivalent. Not only are these properties mutually exclusive, but in parts of the map well removed from the standard line(s) conformal maps severely exaggerate area and equal-area maps severely distort shape.

Two conformal projections useful in navigation illustrate how badly a map can distort area. The Mercator projection, on the left side of figure 2.7, renders Greenland as large as South America, whereas a globe would show Greenland only about one-eighth as large. North-south scale increases so sharply toward the poles that the poles themselves lie at infinity and never appear on an equatorially centered Mercator map. The right side of figure 2.7 reveals an even more severe distortion of area on the gnomonic projection, which cannot portray even half the globe.



Mercator projection   Gnomonic projection

FIGURE 2.7. Straight lines on an equatorially based Mercator projection (left) are rhumb lines, which show constant geographic direction, whereas straight lines on a gnomonic projection (right) are great circles, which show the shortest route between two points.

Why, then, are these projections used at all? Although two of the worst possible perspectives for general-purpose base maps and wall maps, these maps are of enormous value to a navigator with a straightedge. On the Mercator map, for instance, a straight line is a *rhumb line* or *loxodrome*, which shows an easily followed route of constant bearing. A navigator at A can draw a straight line to B, measure with a protractor the angle between this rhumb line and the meridian, and use this bearing and a corrected compass to sail or fly from A to B. On the gnomonic map, in contrast, a straight line represents a *great circle* and shows the shortest course from A to B. An efficient navigator would identify a few intermediate points on this great-circle route, transfer these course-adjustment points from the gnomonic map to the Mercator map, mark a chain of rhumb lines between successive intermediate points, measure each rhumb line's bearing, and proceed from A to B along a compromise course of easily followed segments that collectively approximate a shortest-distance route.

Map projections distort five geographic relationships: areas, angles, gross shapes, distances, and directions. Although some projections preserve local angles but not areas, others preserve areas but not local angles. All distort large shapes noticeably (but some distort continental shapes more than others), and

FIGURE 2.8. Oblique azimuthal equidistant projection centered on Chicago, Illinois, just east of the meridian at 90° W.

all distort at least some distances and some directions. Yet as the Mercator and gnomonic maps demonstrate, the mapmaker often can tailor the projection to serve a specific need. For instance, the oblique azimuthal *equidistant* projection in figure 2.8 shows true distance and directional relationships for shortest-distance great-circle routes converging on Chicago, Illinois. Although highly useful for someone concerned with relative proximity to Chicago, this projection is of no use for distance comparisons not involving Chicago. Moreover, its poor portrayal of the shapes and relative areas of continents, especially when extended to a full-world map, limits its value as a general-purpose reference map. With an interactive computer graphics system and good mapping software, of course, map users can become their own highly versatile mapmakers and tailor projections to many unique needs.

Among the more highly tailored map projections are *cartograms*, which portray such relative measures as travel time,

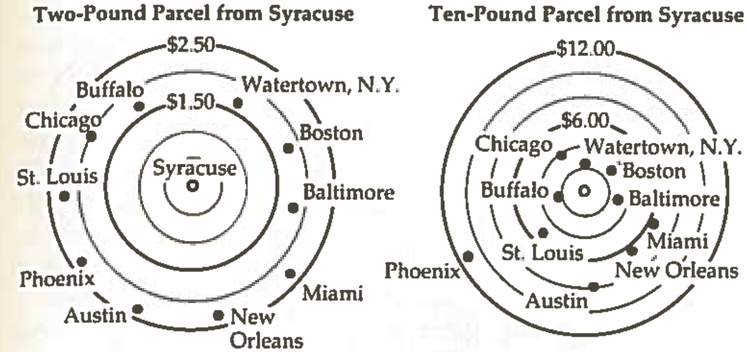


FIGURE 2.9. Distance cartograms showing relative spaces based on parcel-post rates from or to Syracuse, New York.

transport cost, and population size. Although a more conventional map might address these with tailored symbols and a standard projection, the geometry and layout of the cartogram make a strong visual statement of distance or area relationships. The distance *cartograms* in figure 2.9, for example, provide a dramatic comparison of two postal rates, which define different transport-cost spaces for their focal point, Syracuse, New York. Note that the rate for a two-pound parcel mailed to Watertown, New York, is a little more than half the rate from Syracuse to Phoenix, Arizona, whereas the corresponding rates for a ten-pound parcel more nearly reflect Watertown's relative proximity (only seventy miles north of Syracuse). These schematic maps omit boundaries and other traditional frame-of-reference features, which are less relevant here than the names of the destinations shown.

Coastlines and some national boundaries are more useful in figure 2.10, an *area cartogram*, which even includes a pseudogrid to create the visual impression of "the world on a torus." This projection is a *demographic base map*, on which the relative sizes of areal units represent population, not land area. Note that the map portrays India almost thirty times larger than Canada because the Indian population is about thirty times larger than the Canadian population, even though Canada's 3.8 million mi^2 area is much larger than India's 1.2 million mi^2. The cartogram has merged some countries with

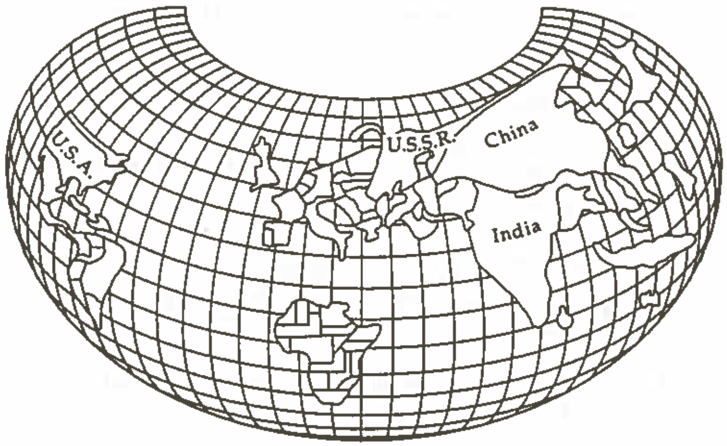


FIGURE 2.10. "World on a Torus" demographic base map is an area cartogram based on the populations of major countries.

smaller populations, demonstrating the mapmaker's political insensitivity in sacrificing nationalism for clarity. Yet traditionalist cartographers who scorn cartograms as foolish, inaccurate cartoons ignore the power of map distortions to address a wide array of communication and analytical needs.

Map Symbols

Graphic symbols complement map scale and projection by making visible the features, places, and other locational information represented on the map. By describing and differentiating features and places, map symbols serve as a graphic code for storing and retrieving data in a two-dimensional geographic framework. This code can be simple and straightforward, as on a route map drawn to show a new neighbor how to find the local elementary school; a few simple lines, labels, and Xs representing selected streets and landmarks should do. Labels such as "Elm St." and "Fire Dept." tie the map to reality and make a key or legend unnecessary.

When the purpose of the map is specific and straightforward, selection of map features also serves to suppress unimportant information. But sheet maps and atlas maps mass-produced by government mapping agencies and commercial map publishers must address a wide variety of questions, and the map's symbols must tell the user what's relevant and what isn't. Without the mapmaker present to explain unfamiliar details, these maps need a symbolic code based on an understanding of graphic logic and the limitations of visual perception. A haphazard choice of symbols, adequate for the labels and little pictures of way-finding maps and other folk cartography, can fail miserably on general-purpose maps rich in information.

Some maps, such as geologic maps and weather charts, have complex but standardized symbologies that organize an enormous amount of data meaningful only to those who understand the field and its cartographic conventions. Although as arcane to most people as a foreign language or mathematics, these maps also benefit from symbols designed according to principles of logic and communication.

Appreciating the logic of map symbols begins with understanding the three geometric categories of map symbols and the six visual variables shown in figure 2.11. Symbols on flat maps are either point symbols, line symbols, or area symbols. Road maps and most other general-purpose maps use combinations of all three: point symbols to mark the locations of landmarks and villages, line symbols to show the lengths and shapes of rivers and roads, and area symbols to depict the form and size of state parks and major cities. In contrast, *statistical maps*, which portray numerical data, commonly rely upon a single type of symbol, such as dots denoting 10,000 people or graytones representing election results by county.

Maps need contrasting symbols to portray geographic differences. As figure 2.11 illustrates, map symbols can differ in size, shape, graytone value, texture, orientation, and hue—that is, color differences as between blue, green, and red (pl. 1). Each of these six visual variables excels in portraying one kind of geographic difference. Shape, texture, and hue are effective in showing qualitative differences, as among land uses or dominant religions. For quantitative differences, size is more suited to showing variation in amount or count, such as the number of television viewers by market area, whereas graytone value is preferred for portraying differences in rate or intensity, such as the proportion of the viewing audience watching the

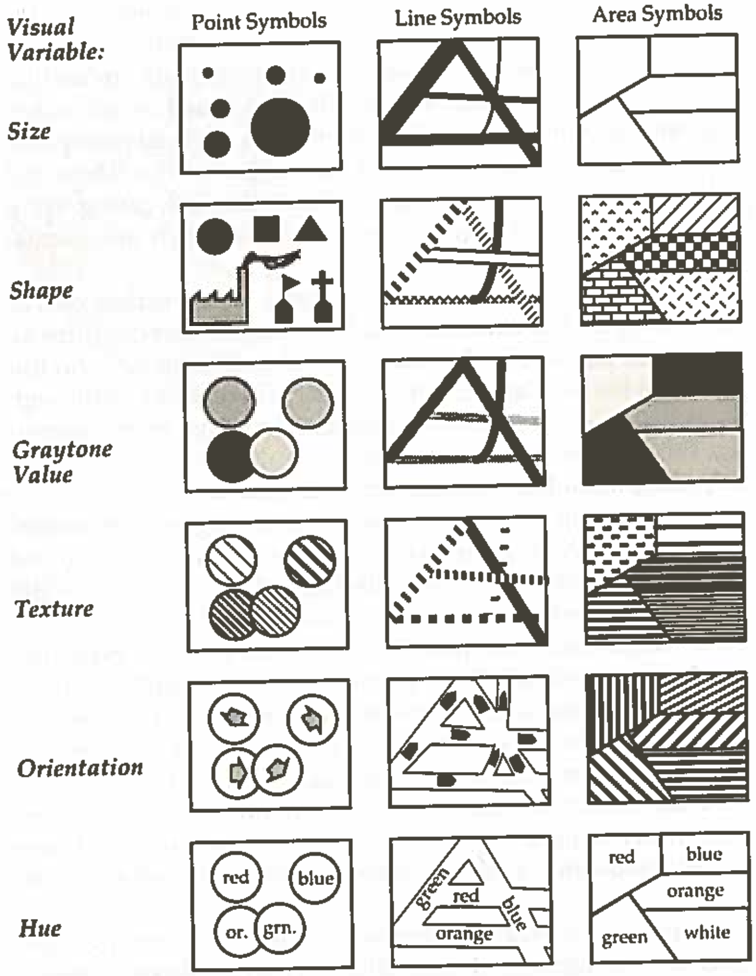


FIGURE 2.11. The six principal visual variables.

seventh game of the World Series. Symbols varying in orientation are useful mostly for representing winds, migration streams, troop movements, and other directional occurrences.

Some visual variables are unsuitable for small point symbols and thin line symbols that provide insufficient contrast with background. Hue, for instance, is more effective in showing differences in kind for area symbols than for tiny point symbols, such as the dots on a dot-distribution map. Graytone value, which usually works well in portraying percentages

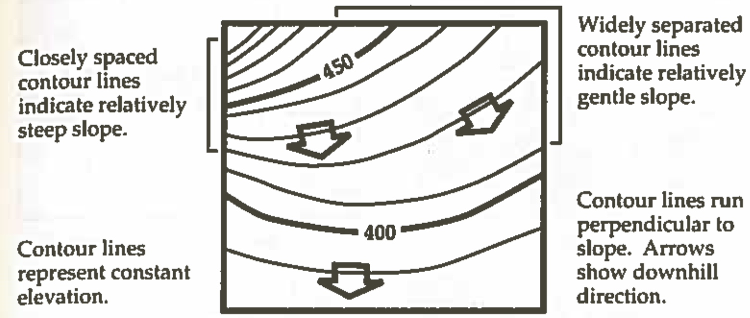


FIGURE 2.12. Elevation contours use two visual variables: spacing (texture) portrays steepness, and contour orientation is perpendicular to the direction of slope.

and rates for area symbols, is visually less effective with point and line symbols, which tend to be thinner than area symbols. Point symbols commonly rely on shape to show differences in kind and on size to show differences in amount. Line symbols usually use hue or texture to distinguish rivers from railways and town boundaries from dirt roads. Size is useful in representing magnitude for links in a network: a thick line readily suggests greater capacity or heavier traffic than a thin line implies. Area symbols usually are large enough to reveal differences in hue, graytone, and pattern, but a detail inset, with a larger scale, might be needed to show very small yet important areal units.

Some symbols combine two visual variables. For example, the elevation contours on a topographic map involve both orientation and spacing, an element of pattern. As figure 2.12 demonstrates, a contour line's direction indicates the local direction of slope because the land slopes downward perpendicular to the trend of the contour line. And the spacing of the contour lines shows the relative tilt of the land because close contours mark steep slopes and separated contours indicate gentle slopes. Similarly, the spread of dots on a dot-distribution map may show the relative sizes of hog-producing regions, whereas the spacing or clustering of these dots reveals the relative intensity and geographic concentration of production.

A poor match between the data and the visual variable can frustrate or confuse the map user. Among the worst offenders are novice mapmakers seduced by the brilliant colors of com-

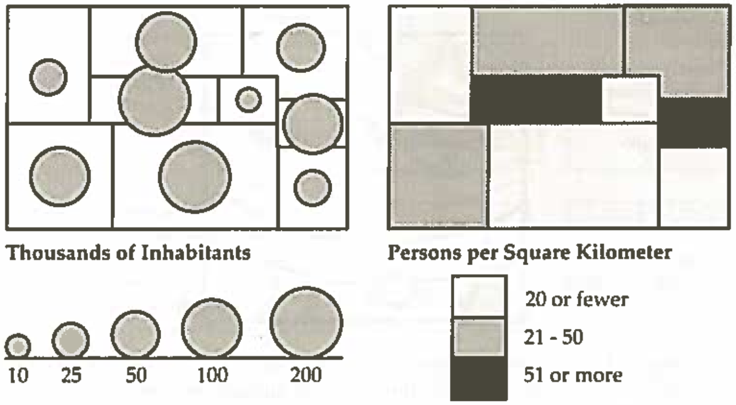


FIGURE 2.13. Graduated point symbols (left) and graytone area symbols (right) offer straightforward portrayals of population size and population density.

puter graphics systems into using reds, blues, greens, yellows, and oranges to portray quantitative differences. Contrasting hues, however visually dramatic, are not an appropriate substitute for a logical series of easily ordered graytones. Except among physicists and professional "colorists," who understand the relation between hue and wavelength of light, map users cannot easily and consistently organize colors into an ordered sequence. And those with imperfect color vision might not even distinguish reds from greens. Yet most map users can readily sort five or six graytones evenly spaced between light gray and black; decoding is simple when darker means more and lighter means less. A legend might make a bad map useful, but it can't make it efficient.

Area symbols are not the only ones useful for portraying numerical data for states, counties, and other areal units. If the map must emphasize magnitudes such as the number of inhabitants rather than intensities such as the number of persons per square mile, point symbols varying in size are more appropriate than area symbols varying in graytone. The two areal-unit maps in figure 2.13 illustrate the different graphic strategies required for portraying population size and population density. The map on the left uses *graduated point symbols* positioned near the center of each area; the size of the point symbol represents population size. At its right a *choropleth map* uses graytone symbols that fill the areal units; the relative

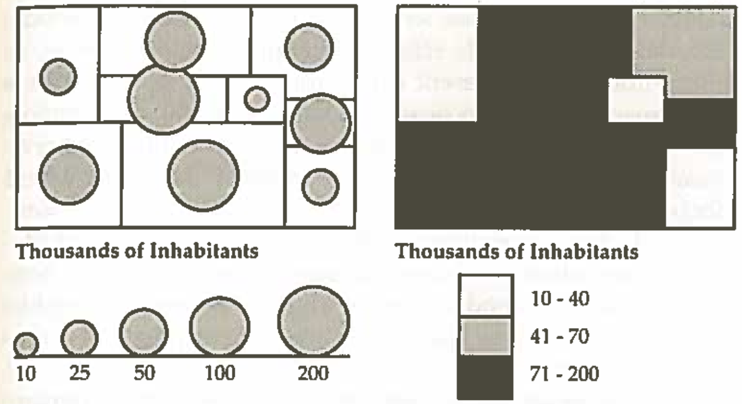


FIGURE 2.14. Map with graduated point symbols (left) using symbol size to portray magnitude demonstrates an appropriate choice of visual variable. Map with graytone area symbols (right) is ill suited to portray magnitude.

darkness of the symbol shows the concentration of population on the land.

Because the visual variables match the measures portrayed, these maps are straightforward and revealing. At the left, big point symbols represent large populations, which occur in both large and small areas, and small point symbols represent small populations. On the choropleth map to the right, a dark symbol indicates many people occupying a relatively small area, whereas a light symbol represents either relatively few people in a small area or many people spread rather thinly across a large area.

Figure 2.14 illustrates the danger of an inappropriate match between measurement and symbol. Both maps portray population size, but the choropleth map at the right is misleading because its area symbols suggest intensity, not magnitude. Note, for instance, that the dark graytone representing a large county with a large but relatively sparsely distributed population also represents a small county with an equally large but much more densely concentrated population. In contrast, the map at the left provides not only a more direct symbolic representation of population size but a clearer picture of area boundaries and area size. The map user should beware of spurious choropleth maps based on magnitude yet suggesting density or concentration.

Form and color make some map symbols easy to decode. Pictorial point symbols effectively exploit familiar forms, as when little tents represent campgrounds and tiny buildings with crosses on top indicate churches. Alphabetic symbols also use form to promote decoding, as with common abbreviations ("PO" for post office), place-names ("Baltimore"), and labels describing the type of feature ("Southern Pacific Railway"). Color conventions allow map symbols to exploit idealized associations of lakes and streams with a bright, nonmurky blue and wooded areas with a wholesome, springlike green. Weather maps take advantage of perceptions of red as warm and blue as cold.

Color codes often rely more on convention than on perception, as with land-use maps, where red commonly represents retail sales and blue stands for manufacturing. Physical-political reference maps found in atlases and on schoolroom walls reinforce the convention of *hypsometric tints*, a series of color-coded elevation symbols ranging from greens to yellows to browns. Although highly useful for those who know the code, elevation tints invite misinterpretation among the unwary. The greens used to represent lowlands, for instance, might suggest lush vegetation, whereas the browns representing highlands can connote barren land—despite the many lowland deserts and highland forests throughout the world. Like map projections, map symbols can lead naive users to wrong conclusions.

Chapter 3

Map Generalization: Little White Lies and Lots of Them

A good map tells a multitude of little white lies; it suppresses truth to help the user see what needs to be seen. Reality is three-dimensional, rich in detail, and far too factual to allow a complete yet uncluttered two-dimensional graphic scale model. Indeed, a map that did not generalize would be useless. But the value of a map depends on how well its generalized geometry and generalized content reflect a chosen aspect of reality.

Geometry

Clarity demands geometric generalization because map symbols usually occupy proportionately more space on the map than the features they represent occupy on the ground. For instance, a line 1/50 inch wide representing a road on a 1:100,000-scale map is the graphic equivalent of a corridor 167 feet wide. If a road's actual right-of-way was only 40 feet wide, say, a 1/50-inch-wide line symbol would claim excess territory at scales smaller than 1:24,000. At 1:100,000, this road symbol would crowd out sidewalks, houses, lesser roads, and other features. And at still smaller scales more important features might eliminate the road itself. These more important features could include national, state, or county boundaries, which have no width whatever on the ground.

Point, line, and area symbols require different kinds of generalization. For instance, cartographers recognize the five fundamental processes of geometric line generalization described in figure 3.1. First, of course, is the *selection* of complete features for the map. Selection is a positive term that implies the suppression, or nonselection, of most features. Ideally the map author approaches selection with goals to be satisfied by

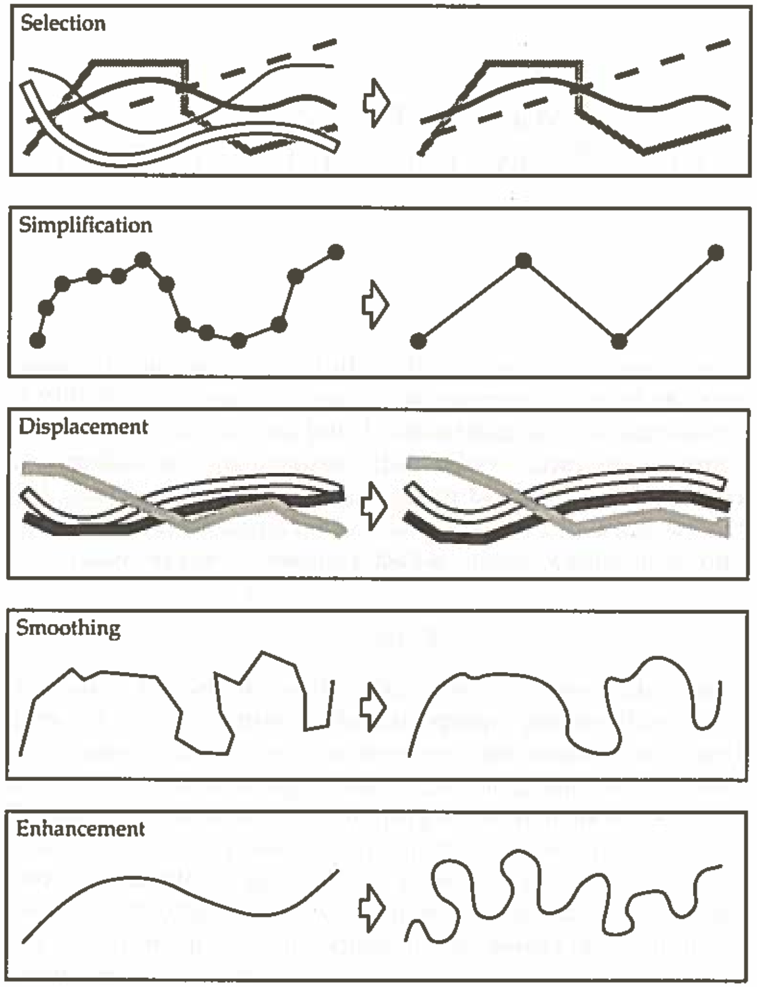


FIGURE 3.1. Elementary geometric operations in the generalization of line features.

a well-chosen subset of all possible features that might be mapped and by map symbols chosen to distinguish unlike features and provide a sense of graphic hierarchy. Features selected to support the specific theme for the map usually require more prominent symbols than background features, chosen to give a geographic frame of reference. Selecting background details that are effective in relating new informa-

tion on the map to the viewer's geographic savvy and existing "mental map" often requires more insight and attention than selecting the map's main features. In the holistic process of planning a map, feature selection is the prime link between generalization and overall design.

The four remaining generalization processes in figure 3.1 alter the appearance and spatial position of linear map features represented by a series of points stored in the computer as a list of two-dimensional (X, Y) coordinates. Although the growing use of computers to generalize maps led to the isolation of these four generalization operations, traditional cartographers perform essentially the same operations by hand but with less structure, less formal awareness, and less consistency. *Simplification*, which reduces detail and angularity by eliminating points from the list, is particularly useful if excessive detail was "captured" in developing a cartographic data file, or if data developed for display at one scale are to be displayed at a smaller scale. *Displacement* avoids graphic interference by shifting apart features that otherwise would overlap or coalesce. A substantial reduction in scale, say, from 1:25,000 to 1:1,000,000, usually results in an incomprehensibly congested collection of map symbols that calls for eliminating some features and displacing others. *Smoothing*, which also diminishes detail and angularity, might displace some points and add others to the list. A prime objective of smoothing is to avoid a series of abruptly joined straight line segments. *Enhancement* adds detail to give map symbols a more realistic appearance. Lines representing streams, for instance, might be given typical meander loops, whereas shorelines might be made to look more coastlike. Enhanced map symbols are more readily interpreted as well as more aesthetic.

Point features and map labels require a somewhat different set of generalization operators. Figure 3.2 illustrates that, as with linear features, selection and displacement avoid graphic interference when too many close symbols might overlap or coalesce. When displacement moves a label ambiguously far from the feature it names, *graphic association* with a tie line or a numeric code might be needed to link the label with its symbol. *Abbreviation* is another strategy for generalizing labels on congested small-scale maps. *Aggregation* is useful where many equivalent features might overwhelm the map if ac-

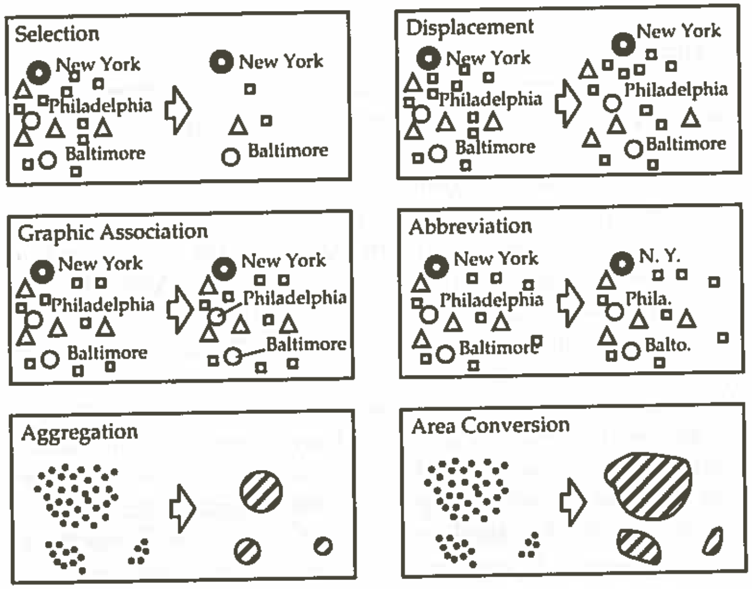


FIGURE 3.2. Elementary geometric operations in the generalization of point features and map labels.

corded separate symbols. In assigning a single symbol to several point features, as when one dot represents twenty reported tornadoes, aggregation usually requires the symbol either to portray the "center of mass" of the individual symbols it replaces or to reflect the largest of several discrete clusters.

Where scale reduction is severe, as from 1:100,000 to 1:20,000,000, *area conversion* is useful for shifting the map viewer's attention from individual occurrences of equivalent features to zones of relative concentration. For example, instead of showing individual tornadoes, the map might define a belt in which tornadoes are comparatively common. In highlighting zones of concentration or higher density, area conversion replaces all point symbols with one or more area symbols. Several density levels, perhaps labeled "severe," "moderate," and "rare," might provide a richer, less generalized geographic pattern.

Area features, as figure 3.3 demonstrates, require the largest set of generalization operators because area boundaries are

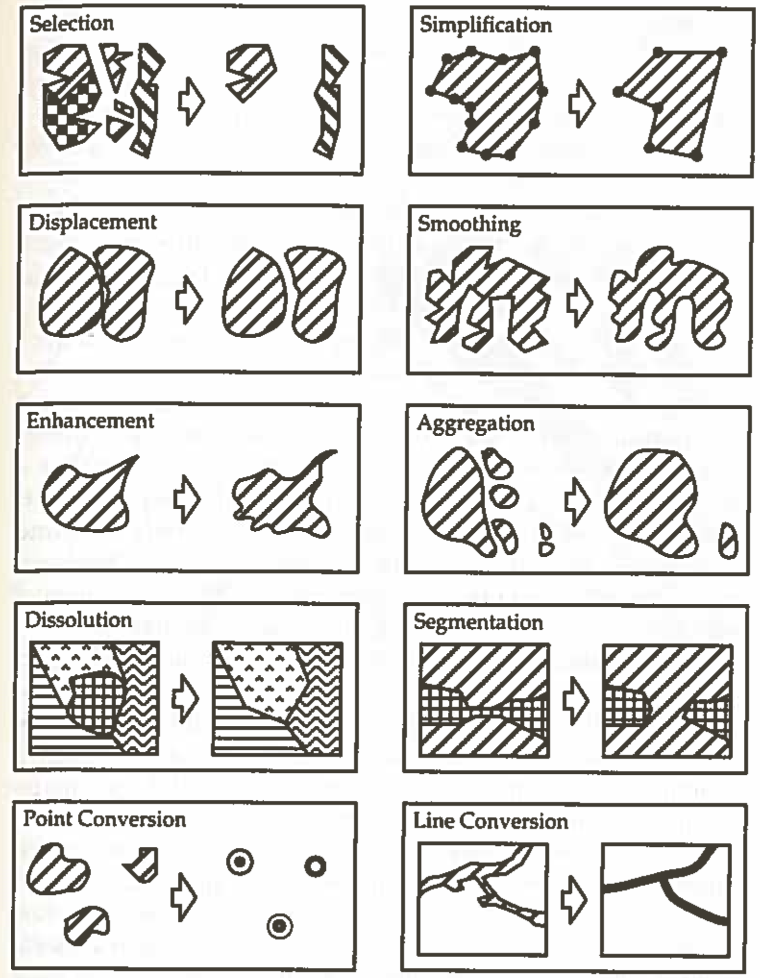


FIGURE 3.3. Elementary geometric operations in the generalization of area features.

subject to aggregation and point conversion and all five elements of line generalization as well as to several operators unique to areas. Selection is particularly important when area features must share the map with numerous linear and point features. A standardized minimum mapping size can direct the selection of area features and promote consistency among the numerous sheets of a map series. For example, 1:24,000-scale topographic maps exclude woodlands smaller than one acre unless important as landmarks or shelterbelts. Soil scientists use a less precise but equally pragmatic size threshold—the head of a pencil—to eliminate tiny, insignificant areas on soils maps.

Aggregation might override selection when a patch otherwise too small to include is either combined with one or more small, similar areas nearby or merged into a larger neighbor. On soils maps and land-use maps, which assign all land to some category, aggregation of two close but separated area features might require the *dissolution* or *segmentation* of the intervening area. A land-use map might, for example, show transportation land only for railroad yards, highway interchanges, and service areas where the right-of-way satisfies a minimum-width threshold. Simplification, displacement, smoothing, and enhancement are needed not only to refine the level of detail and to avoid graphic interference between area boundaries and other line symbols, but also to reconstruct boundaries disrupted by aggregation and segmentation.

Generalization often accommodates a substantial reduction in scale by converting area features to linear or point features. Line conversion is common on small-scale reference maps that represent all but the widest rivers with a single readily recognized line symbol of uniform width. Highway maps also help the map user by focusing not on width of right-of-way but on connectivity and orientation. In treating more compact area features as point locations, point conversion highlights large, sprawling cities such as London and Los Angeles on small-scale atlas maps and focuses the traveler's attention on highway interchanges on intermediate-scale road maps. Linear and point conversion are often necessary because an area symbol at scale would be too tiny or too thin for reliable and efficient visual identification.

Comparing two or more maps showing the same area at substantially different scales is a good way to appreciate the need for geometric generalization. Consider, for instance, the two maps in figure 3.4. The rectangles represent the same area extracted from maps published at scales of 1:24,000 and 1:250,000; enlargement of the small-scale excerpt to roughly the same size as its more detailed counterpart reveals the need for

FIGURE 3.4. Area near Northumberland, Pennsylvania, as portrayed on topographic maps at 1:24,000 (left) and 1:250,000, enlarged to roughly 1:24,000 for comparison (right).

considerable generalization at 1:250,000. The substantially fewer features shown at 1:250,000 demonstrate how feature selection helps the mapmaker avoid clutter. Note that the smaller-scale map omits most of the streets, all labels in this area, all individual buildings, and the island in the middle of the river. The railroad and highway that cross the river are smoother and farther apart, allowing space for the bridge symbols added at 1:250,000. Because the 1:24,000-scale map in a sense portrays the same area in a space over a hundred times larger, it can show many more features in much greater detail.

How precisely are symbols positioned on maps? The U.S. Office of Management and Budget addresses this concern with the National Map Accuracy Standards, honored by the U.S. Geological Survey and other federal mapping agencies. To receive the endorsement "This map complies with the National Map Accuracy Standards," a map at a scale of 1:20,000 or smaller must be checked for symbols that deviate from their correct positions by more than 1/50 inch. This tolerance reflects the limitations of surveying and mapping equipment and human hand-eye coordination. Yet only 90 percent of the points tested must meet the tolerance, and the 10 percent that don't can deviate substantially from their correct positions. Whether a failing point deviates from its true position by 2/50 inch or 20/50 inch doesn't matter—if 90 percent of the points checked meet the tolerance, the map sheet passes.

The National Map Accuracy Standards tolerate geometric generalization. Checkers test only "well-defined points" that are readily identified on the ground or on aerial photographs, easily plotted on a map, and conveniently checked for horizontal accuracy; these include survey markers, roads and railway intersections, corners of large buildings, and centers of small buildings. Guidelines encourage checkers to ignore features that might have been displaced to avoid overlap or to provide a minimum clearance between symbols exaggerated in size to ensure visibility. In areas where features are clustered, maps tend to be less accurate than in more open areas. Thus Pennsylvania villages, with comparatively narrow streets and no front yards, would yield less accurate maps than, say, Colorado villages, with wide streets, spacious front yards, and big lots. But as long as 90 percent of a sample of well-defined points not needing displacement meet the tolerance, the map sheet passes.

Maps that meet the standards show only *planimetric* distance, that is, distance measured in a plane. As figure 3.5 shows, a planimetric map compresses the three-dimensional land surface onto a two-dimensional sheet by projecting each point perpendicularly onto a horizontal plane. For two points at different elevations, the map distance between their "planimetrically accurate" positions underestimates both overland distance across the land surface and straight-line distance in three dimensions. Yet this portrayal of planimetric distance is a geometric generalization essential for large-scale flat maps.

The user should be wary, though, of the caveat "approximately positioned" or the warning "This map may not meet the National Map Accuracy Standards." In most cases such maps have been compiled from unrectified aerial photographs,

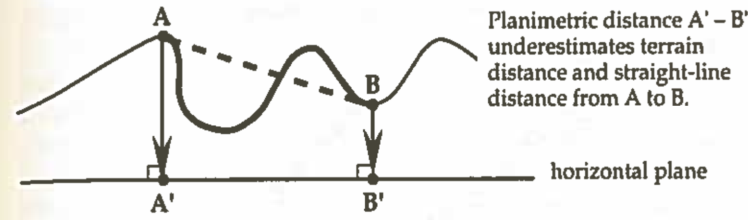


FIGURE 3.5. Planimetric map generalizes distance by the perpendicular projection of all positions onto a horizontal plane.

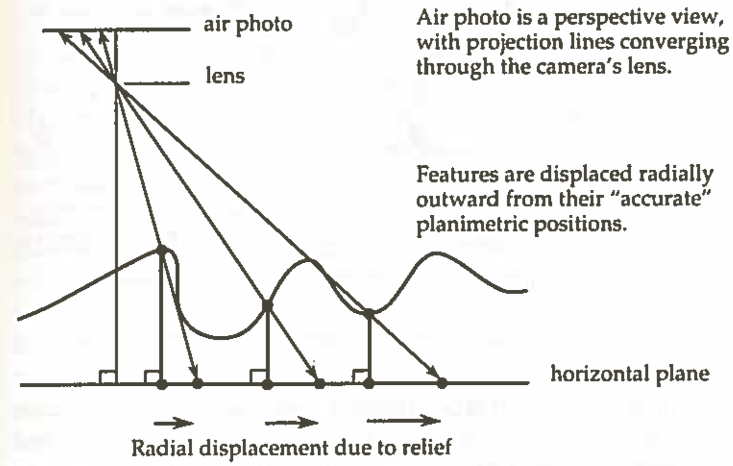


FIGURE 3.6. A vertical aerial photograph (and any map with symbols traced directly from an air photo) is a perspective view with points displaced radially from their planimetric positions.

on which horizontal error tends to be particularly great for rugged, hilly areas. Figure 3.6 shows the difference between the air photo's perspective view of the terrain and the planimetric map's representation of distances in a horizontal plane. Because lines of sight converge through the camera's lens, the air photo displaces most points on the land surface from their planimetric positions. Note that displacement is radially outward from the center of the photo, that displacement is greater for points well above the horizontal plane than for lower points, and that displacement tends to be greater near the edges than near the center. Cartographers call this effect

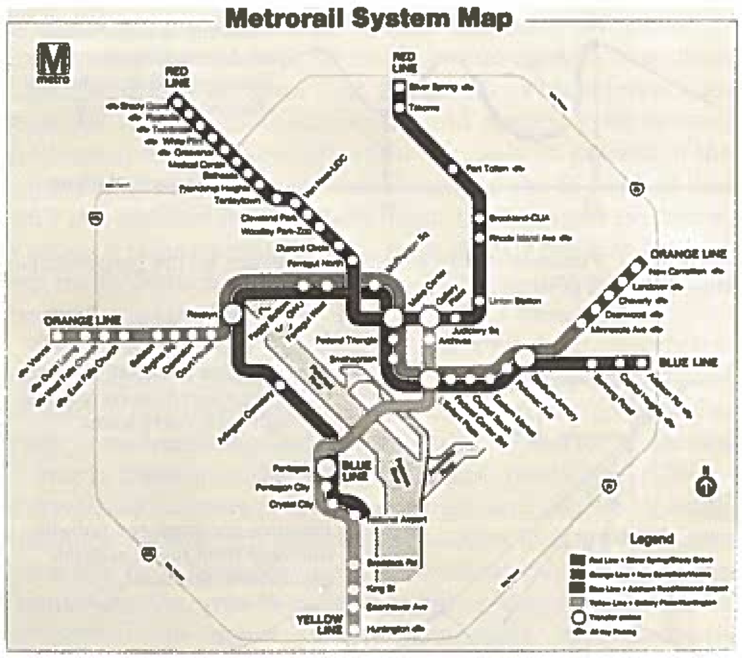


FIGURE 3.7. Linear cartogram of the Washington, D.C., Metro system.

"radial displacement due to relief," or simply *relief displacement*. An exception is the *orthophoto*, an air-photo image electronically stretched to remove relief displacement. An *orthophotomap*, produced from orthophotos, is a planimetrically accurate photo-image map.

For some maps, though, geometric accuracy is less important than linkages, adjacency, and relative position. Among the more effective highly generalized maps are the linear cartograms portraying subway and rapid transit systems. As in figure 3.7, scale is relatively large for the inner city, where the routes converge and connect; stops in the central business district might be only four or five blocks apart, and a larger scale is needed here to accommodate more route lines and station names. In contrast, toward the fringes of the city, where stations are perhaps a mile or more apart, scale can be smaller because mapped features are less dense. Contrasting colors usually differentiate the various lines; the Washington, D.C., Metro system, in fact, calls its routes the Blue Line, the

Red Line, and so forth, to enhance the effectiveness of its map. By sacrificing geometric accuracy, these schematic maps are particularly efficient in addressing the subway rider's basic questions: Where am I on the system? Where is my destination? Do I need to change trains? If so, where and to what line? In which direction do I need to go? What is the name of the station at the end of the line? How many stops do I ride before I get off? Function dictates form, and a map more "accurate" in the usual sense would not work as well.

Content

As geometric generalization seeks graphic clarity by avoiding overlapping symbols, content generalization promotes clarity of purpose or meaning by filtering out details irrelevant to the map's function or theme. Content generalization has only two essential elements, selection and classification. Selection, which serves geometric generalization by suppressing some information, promotes content generalization by choosing only relevant features. Classification, in contrast, makes the map helpfully informative as well as usable by recognizing similarities among the features chosen so that a single type of symbol can represent a group of similar features. Although all map features are in some sense unique, usually each feature cannot have a unique symbol. Even though some maps approach uniqueness by naming individual streets or numbering lots, these maps also use very few types of line symbols, to emphasize similarities among roads and property boundaries as groups. Indeed, the graphic vocabulary of most maps is limited to a small set of standardized, contrasting symbols.

Occasionally the "template effect" of standardized symbols will misinform the map user by grouping functionally different features. Standard symbols, designed for ready, unambiguous recognition and proportioned for a particular scale, are common in cartography and promote efficiency in both map production and map use. Traditional cartographers use plastic drawing templates to trace in ink the outlines of highway shields and other symbols not easily rendered freehand. Drafters can cut area and point symbols from printed sheets and stick them onto the map and can apply dashed, dotted, or parallel lines from rolls of specially printed flexible tape. Electronic publishing systems allow the mapmaker not only to choose from a menu of point, line, and area symbols provided with the software but also to design and store new forms, readily duplicated and added where needed. Consistent symbols also benefit users of the U.S. Geological Survey's series of thousands of large-scale topographic map sheets, all sharing a single graphic vocabulary. On highway maps, the key (or "legend") usually presents the complete set of symbols so that while examining the map, at least, the reader encounters no surprises. Difficulties arise, though, when a standard symbol must represent functionally dissimilar elements. Although a small typeset annotation next to the feature sometimes flags an important exception, for instance, a section of highway "under construction," mapmakers frequently omit useful warnings.

Generalized highway interchanges are a prime example of how information obscured by the template effect can mislead or inconvenience a trusting map user. The left panel of figure 3.8 is a detailed view of the interchange near Rochester, New York, between highways 104 and 590, as portrayed at 1:9,600 on a state transportation department map. Note that a motorist traveling from the east (that is, from the right) on N.Y. 104 cannot easily turn north (toward the top of the map) onto N.Y. 590. The upper right portion of the left-hand map shows that the necessary connecting lanes from N.Y. 104 were started but not completed. In contrast, the right panel shows how various commercial map publishers portray this interchange on their small-scale statewide highway maps. Two diamond-shaped interchange symbols suggest separate and equivalent connections with the eastward and westward portions of N.Y. 104. Yet the large-scale map clearly indicates that a driver expecting an easy connection from N.Y. 104 westbound onto N.Y. 590 northbound must travel to the next exit west or south and then double back. Until the road builders complete their planned connecting lanes, such discrepancies between reality and art will frustrate motorists who assume all little diamonds represent full interchanges.

Effective classification and selection often depend on a mixture of informed intuition and a good working definition. This is particularly true for geologic maps and soils maps, commonly prepared by several field scientists working in widely

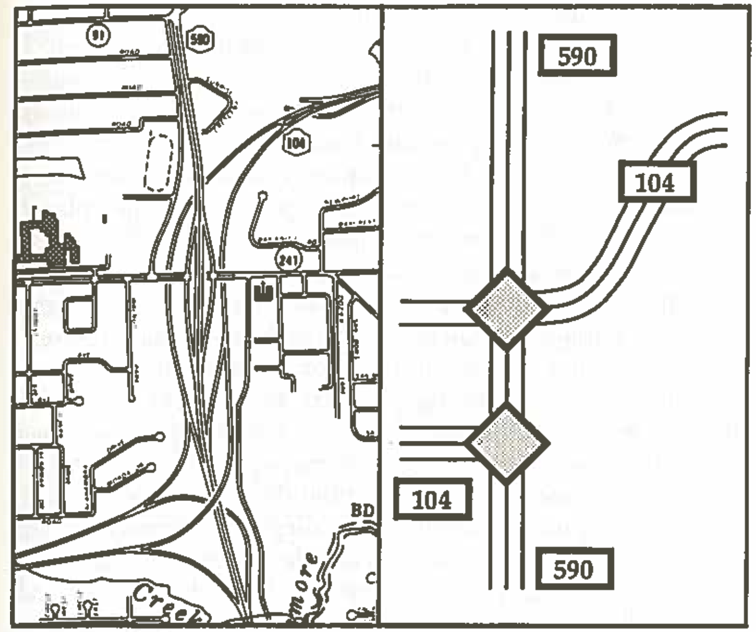


FIGURE 3.8. Highway interchange near Rochester, New York, as portrayed on a detailed transportation planning map (left) and on several commercial road maps (right).

separated places. A detailed description is necessary if two people mapping areas a hundred miles apart must identify and draw boundaries for different parts of the same feature. These descriptions should also address the mapping category's internal homogeneity and the sharpness of its "contacts" with neighboring units. In soils mapping, for instance, small patches of soil B might lie within an area labeled as soil A. This practice is accepted because these enclaves of soil B are too small to be shown separately, and because the soil scientist cannot be aware of all such enclaves. Soil mapping, after all, is slow, tedious work that requires taking samples below the surface with a drill or auger and occasionally digging a pit to examine the soil's vertical profile. Map accuracy thus depends upon the field scientist's understanding of the effects of terrain and geology (if known) on soil development as well as on expertise in selecting sample points and intuition in plotting boundaries.

That crisp, definitive lines on soils maps mark inherently fuzzy boundaries is unfortunate. More appalling, though, is the uncritical use in computerized geographic information systems of soil boundaries plotted on "unrectified" aerial photos subject to the relief-displacement error described in figure 3.6. Like quoting a public figure out of context, extracting soils data from a photomap invites misinterpretation. When placed in a database with more precise information, these data readily acquire a false aura of accuracy.

Computers generally play a positive role in map analysis and map display, the GIGO effect (garbage in, garbage out) notwithstanding. Particularly promising is the ability of computers to generalize the geometry and content of maps so that one or two geographic databases might support a broad range of display scales. Large-scale maps presenting a detailed portrayal of a small area could exploit the richness of the data, whereas computer-generalized smaller-scale displays could present a smaller selection of available features, suitably displaced to avoid graphic interference. Both the content and scale of the map can be tailored to the particular needs of individual users.

Computer-generalized maps of land use and land cover illustrate how a single database can yield radically different cartographic pictures of a landscape. The three maps in figure 3.9 show a rectangular region of approximately 700 mi^2 (1,800 km^2) that includes the city of Harrisburg, Pennsylvania, above and slightly to the right of center. A computer program generalized these maps from a large, more detailed database that represents much smaller patches of land and describes land cover with a more refined set of categories. The generalization program used different sets of weights or priorities to produce the three patterns in figure 3.9. The map at the upper left differs from the other two maps because the computer was told to emphasize urban and built-up land. This map makes some small built-up areas more visible by reducing the size of area symbols representing other land covers. In contrast, the map at the upper right reflects a high visual preference for agricultural land. A more complex set of criteria guided generalization for the display at the lower left: forest land is dominant overall, but urban land dominates agricultural land. In addition, for this lower map the computer dissolved water

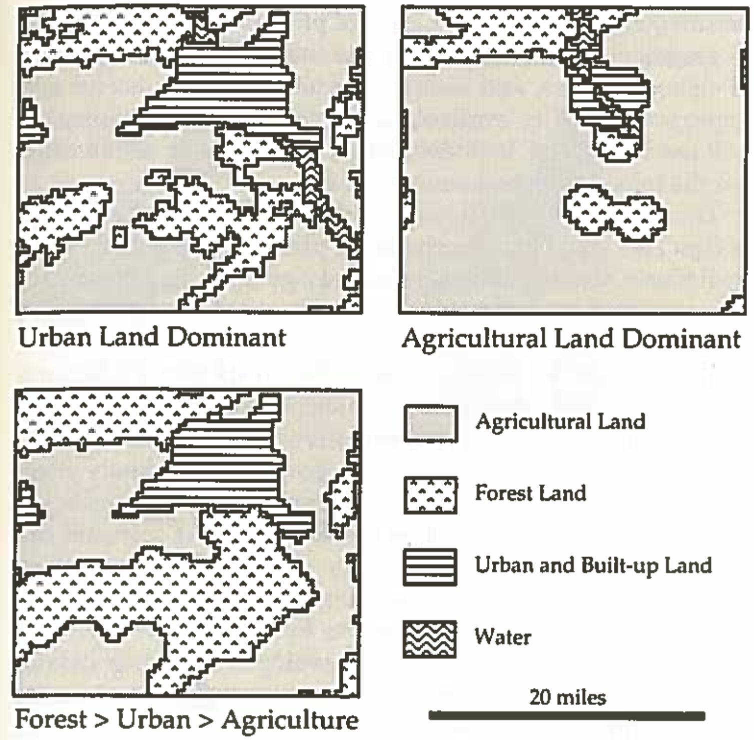


FIGURE 3.9. Land-use and land-cover maps generalized by computer from more detailed data according to three different sets of display priorities.

areas, which were discontinuous because of variations in the width of the river. These differences in emphasis might meet the respective needs and biases of demographers, agronomists, and foresters.

Generalized maps almost always reflect judgments about the relative importance of mappable features and details. The systematic bias demonstrated by these generalized land-cover maps is not exclusive to computer-generated maps; manual cartographers have similar goals and biases, however vaguely defined and unevenly applied. Through the consistent application of explicit specifications, the computer offers the possibility of a better map. Yet whether the map's title or description reveals these biases is an important clue to the integrity of the mapmaker or publisher. Automated mapping allows experimentation with different sets of priorities. Hence computer generalization should make the cartographer more aware of choices, values, and biases. But just because a useful and appropriate tool is available does not mean the mapmaker will use it. Indeed, laziness and lack of curiosity all too often are the most important source of bias.

The choropleth map (introduced as the right-hand elements of figs. 2.13 and 2.14) is perhaps the prime example of this bias by default. Choropleth maps portray geographic patterns for regions composed of areal units such as states, counties, and voting precincts. Usually two to six graytone symbols, on a scale from light to dark, represent two to six nonoverlapping categories for an intensity index such as population density or the percentage of the adult population voting in the last election. The breaks between these categories can markedly affect the mapped pattern, and the cautious map author tests the effects of different sets of class breaks. Mapping software can unwittingly encourage laziness by presenting a map based upon a "default" classification scheme that might, for instance, divide the range of data values into five equal intervals. As a marketing strategy, the software developer uses such default specifications to make the product more attractive by helping the first-time or prospective user experience success. Too commonly, though, the naive or noncritical user accepts this arbitrary display as the standard solution, not merely as a starting point, and ignores the invitation of the program's pull-down menus to explore other approaches to data classification.

Different sets of categories can lead to radically different interpretations. The two maps in figure 3.10, for example, offer very different impressions of the spatial pattern of homes in the northeastern United States still lacking telephones in 1960. Both maps have three classes, portrayed with a graded sequence of graytone area symbols that imply "low," "medium," and "high" rates of phonelessness. Both sets of categories use round-number breaks, which mapmakers for some mysterious reason tend to favor. The map at the left shows a single state, Virginia, in its high, most deficient class, and a single state, Connecticut, in its low, most well-connected class. The casual viewer might attribute these extremes to Virginia's higher proportion of disadvantaged blacks and to Connecticut's af-

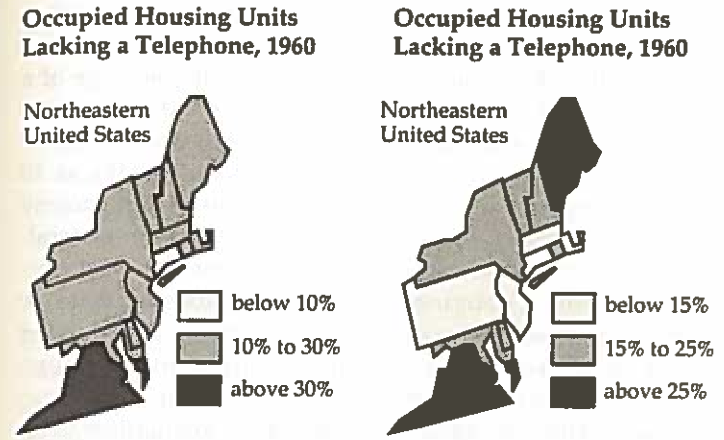


FIGURE 3.10. Different sets of class breaks applied to the same data yield different-looking choropleth maps.

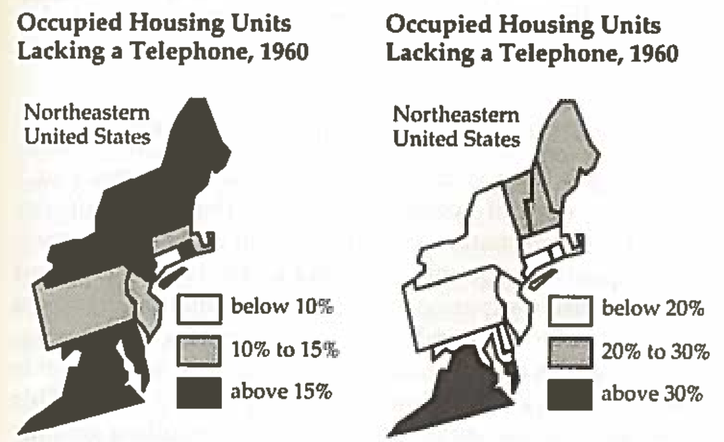


FIGURE 3.11. Class breaks can be manipulated to yield choropleth maps supporting politically divergent interpretations.

fluent suburbs and regard the remaining states as homogeneously "average." In contrast, the map at the right portrays a more balanced distribution of states among the three groups and suggests a different interpretation. Both states in the high category have substantial dispersed rural populations, and all four in the low category are highly urban and industrialized.

Moreover, a smaller middle group suggests less overall homogeneity.

Machiavellian bias can easily manipulate the message of a choropleth map. Figure 3.11, for example, presents two cartographic treatments with substantially different political interpretations. The map on the left uses rounded breaks at 10 percent and 15 percent, forcing most states into its high, poorly connected category and suggesting a Northeast with generally poor communications. Perhaps the government is ineffective in regulating a gouging telecommunications industry or in eradicating poverty. Its counterpart on the right uses rounded breaks at 20 percent and 30 percent to paint a rosier picture, with only one state in the high group and eight in the low, well-served category. Perhaps government regulation is effective, industry benign, and poverty rare.

The four maps in figures 3.10 and 3.11 hold two lessons for the skeptical map reader. First, a single choropleth map presents only one of many possible views of a geographic variable. And second, the white lies of map generalization might also mask the real lies of the political propagandist.

Intuition and Ethics in Map Generalization

Small-scale generalized maps often are authored views of a landscape or a set of spatial data. Like the author of any scholarly work or artistic creation based on reality, the conscientious map author not only examines a variety of sources but relies on extensive experience with the information or region portrayed. Intuition and induction guide the choice of features, graphic hierarchy, and abstraction of detail. The map is as it is because the map author "knows" how it should look. This knowledge, of course, might be faulty, or the resulting graphic interpretation might differ significantly from that of another competent observer. As is often the case, two views might both be valid.

Chapter 4

Blunders That Mislead

Some maps fail because of the mapmaker's ignorance or oversight. The range of blunders affecting maps includes graphic scales that invite users to estimate distances from world maps, maps based on incompatible sources, misspelled place-names, and graytone symbols changed by poor printing or poor planning. By definition a blunder is not a lie, but the informed map user must be aware of cartographic fallibility, and even of a bit of mischief.

Cartographic Carelessness

Mapmakers are human, and they make mistakes. Although poor training and sloppy design account for some errors, most cartographic blunders reflect a combination of inattention and inadequate editing. If the mapmaker is rushed, if the employer views willingness to work for minimal wages as more important than skill in doing the job, or if no one checks and rechecks the work, missing or misplaced features and misspelled labels are inevitable.

Large-scale base maps have surprisingly few errors. A costly but efficient bureaucratic structure at government mapping agencies usually guarantees a highly accurate product. Several layers of fact checking and editing support technicians or contractors selected for skill and concern with quality. Making topographic maps is a somewhat tedious, multistep manufacturing process, and using outside contractors for compilation or drafting requires a strong commitment to quality control buttressed by the bureaucrat's inherent fear of embarrassment. Blunders occasionally slip through, but these are rare.

Errors are more common on derivative maps—that is, maps compiled from other maps—than on basic maps compiled from air photos and other primary data. Artists lacking cartographic training and an appreciation of geographic details draw most tourist maps and news maps, and poorly paid drafting technicians produce most American street maps. Omission and garbling are particularly likely when information is transferred manually from one map to another. Getting all the appropriate information from the large-scale base map onto a small-scale derivative map is not an easy chore. Several base maps might be needed, the compiler might not have a clear idea of what is necessary, or several compilers might work on the same map. Using another derivative map as a source can save time, but only at the risk of incorporating someone else's errors.

Map blunders make amusing anecdotes, and the press helps keep cartographers conscientious by reporting the more outrageous ones. In the early 1960s, for instance, the American Automobile Association "lost Seattle," as the Associated Press reported the accidental omission of the country's twenty-third largest city from the AAA's United States road map. Embarrassed, the AAA confessed that "it just fell through the editing crack" and ordered an expensive recall and reprinting.

Equally mortified was the Canadian government tourist office that omitted Ottawa from an airline map in a brochure prepared to attract British tourists. The official explanation that Ottawa had not been a major international point of entry and that the map was compiled before initiation of direct New York–Ottawa air service didn't diminish the irritation of Ottawa residents. The map included Calgary, Regina, and Winnipeg, and as an executive of the city's Capital Visitors and Convention Bureau noted, "Ottawa should be shown in any case, even if the only point of entry was by two-man kayak."

Faulty map reading almost led to an international incident in 1988, when the Manila press reported the Malaysian annexation of the Turtle Islands. News maps showing the Malaysian encroachment supported three days of media hysteria and saber rattling. These maps were later traced to the erroneous reading of an American navigation chart by a Philippine naval officer who mistook a line representing the recommended deepwater route for ships passing the Turtle Islands for the boundary of Malaysia's newly declared exclusive economic zone.

Although map blunders provoking outrage between minor powers make amusing anecdotes, inaccurate maps in a war zone can be deadly. The American Civil War illustrates the effect on both sides of wildly conflicting topographic maps and inadequate numbers of trained topographic engineers and geographers. In 1862, for instance, the Union army planned a swift defeat of the Confederates by capturing their capital, Richmond. But unexpected obstacles slowed the northern army's advance after General McClellan's staff based battle plans on inaccurate maps. A lack of good maps also plagued Confederate forces, who were unaware of strategic advantages that would have allowed them to overwhelm McClellan's retreating army.

Modern warfare is particularly vulnerable to bad maps, as the 1983 invasion of Grenada by United States troops and their Caribbean allies demonstrates. The only cartographic intelligence distributed to troops carrying out this politically convenient rescue of American medical students consisted of hastily printed copies of a few obsolete British maps and a tourist map with a military grid added. An air attack destroyed a mental hospital not marked on the maps. Another air strike, ordered by a field commander using one set of grid coordinates but carried out by planes using a map with another grid, wounded eighteen soldiers, one fatally.

Journalists' and social scientists' accounts of the invasion added further cartographic insult. In addition to misplaced symbols and misspelled place-names, one group of coauthors (or their free-lance illustrator) distorted the size and relative position of the Grenadian state's two smaller members, the islands of Carriacou and Petite Martinique. As the lower right part of figure 4.1 illustrates, the book's regional locator map made these islands much smaller than their true relative sizes and moved them much closer to the main island. These errors probably originated in careless compilation from an official map sheet on which an inset showed the smaller islands at a smaller scale than the main island. Since an earlier derived map had the same errors, the authors apparently based their map on a faulty source, which lax editorial checking obviously failed to detect.

The news media have their own cartographic glitches. Errors on news maps reflect both the minimal cartographic

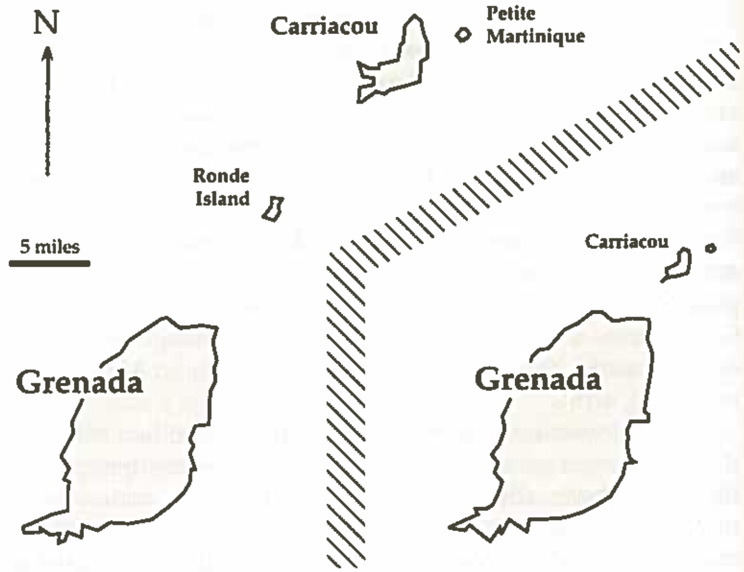
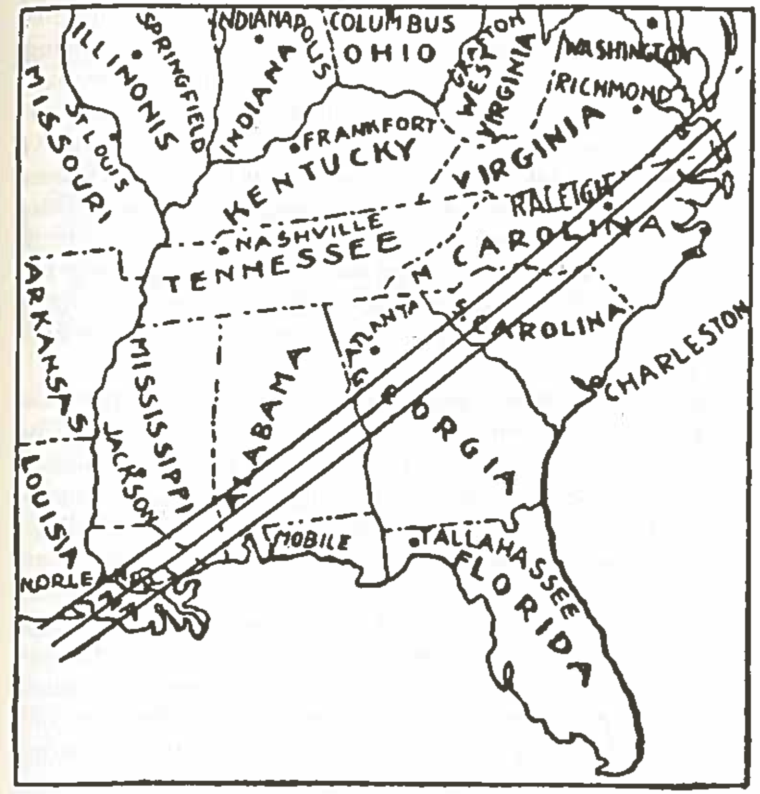


FIGURE 4.1. Reconstruction of a journalistic misrepresentation (right) of the size and location of Grenada's smaller islands, as presented in a book published shortly after the 1983 invasion. The left and upper portion of the figure portrays Carriacou and Petite Martinique correctly.

knowledge of most newspaper artists and the high-pressure, deadline-driven environment in which news is gathered, processed, and published. The use of graphics in newspapers began to grow in the late nineteenth century, when photoengraving enabled news publishers to use more line drawings and photographs and feature syndicates arose to serve smaller papers unable to hire their own artists. Unfortunately, technological advances that made news maps less expensive and more common also allowed poorly skilled, geographically illiterate artists to make publishable maps rapidly. Many news maps distributed about 1900 had the crude boundaries and careless spelling exemplified by figure 4.2, a map accompanying a syndicated story about a coming eclipse; note the extra *n* in the label for Illinois. Hastily drawn news maps have also annexed Michigan's northern peninsula to Wisconsin, Virginia's eastern counties on the Delmarva Peninsula into Maryland, and both North Korea and South Korea (and part of China) to the Soviet Union.



PATH OF THE TOTAL ECLIPSE.

FIGURE 4.2. On 28 May 1900 the *Cortland* (N.Y.) *Evening Standard* printed this hand-drawn map, received from a feature syndicate with the name Illinois misspelled.

Computer graphics, a more recent impetus for journalistic cartography, makes it easier for reporters and editors to alter decent-looking base maps and inadvertently eliminate features and misplace symbols and labels. Typical examples include adding Finnish territory to the Soviet Union on the map decorating a *New York Times* article on Canadian–Soviet relations and switching the labels identifying New Hampshire and Vermont on a Knight-Ridder Graphics Network map of areas in the United States affected by drought.

Perhaps more annoying are blunders on road maps and street maps, particularly when the place you are trying to find is missing, misplaced, misindexed, mislabeled, or badly misshaped. That these errors are not more common is surprising, though. Publishers of street and highway maps must manage a complex, constantly changing database and produce a low-cost, enormously detailed product for largely unappreciative consumers in a highly competitive market. Perhaps because oil companies distributed free road maps for several decades until the early 1970s, and because state and local tourist councils perpetuate the free travel map, the American map buyer has little appreciation of the well-designed, highly accurate maps that the European map user is conditioned to respect, demand, and pay for.

Blunders on street maps reflect how the maps are made. Basic data for established parts of an area can be found on large-scale topographic maps published by the Geological Survey. These maps are in the public domain and can be copied freely, but their publication scale of 1:24,000 allows room for few street names, and many map sheets are ten years or more out of date. Mapmakers thus turn to maps maintained by city and county engineering and highway departments for street names as well as for new streets and other changes. Personnel responsible for copying street alignments, typesetting, and type placement are sometimes inexperienced or inattentive, and editing is not always thorough. A file of customers' complaints can be a useful though not fully reliable source of corrections for the next edition. Some publishers send drafts of their maps to planning and engineering departments for comment and unrealistically depend on overworked civil servants for additional editing.

Unfortunately, many official city maps show the rights-of-way of approved streets that were never cleared, graded, or paved. Often a planned but otherwise fictitious street persists on the city engineer's map until it is officially deleted, as when a developer petitions the planning board to build across the right-of-way or a homeowner attempts to buy the adjoining strip of land. Street-map publishers who compile only from municipal maps are likely to pick up a number of "paper streets" like Garden Street and Pinnacle Street in figure 4.3, found on an official map for Syracuse, New York. Geological Survey and New York State quadrangle maps of the area don't show these streets, but the mapmaker might conveniently

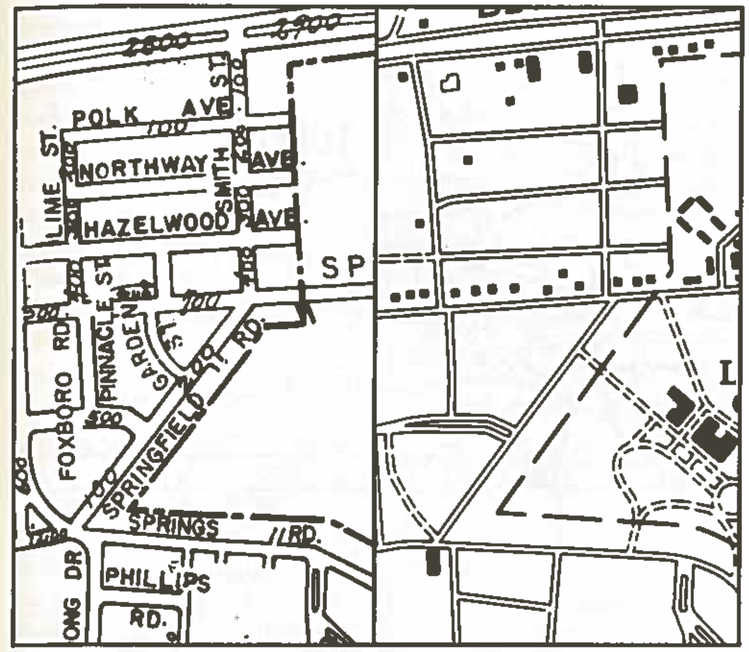


FIGURE 4.3. Two paper streets in Syracuse, New York, as shown on a municipal street map (left panel; see Garden Street and Pinnacle Street just to the left of the center of the map) and the corresponding area as represented on a New York State highway planning map. (In the left panel a tiny arrow points to block-long Pinnacle Street, which is parallel to Smith Street.)

assume that the "official" map provided by local government is more likely to be accurate than a map from Washington or the state capital. A slim profit margin usually precludes field checking, and up-to-date large-scale air photos are expensive to purchase and seldom convenient to consult. Because a mapmaker in a distant city cannot readily determine whether the feature is an obsolete paper street or a recently opened thoroughfare, commercial street maps occasionally pick up phantom streets.

Deliberate Blunders

Although none dare talk about it, publishers of street maps also turn to each other for street names and changes. The

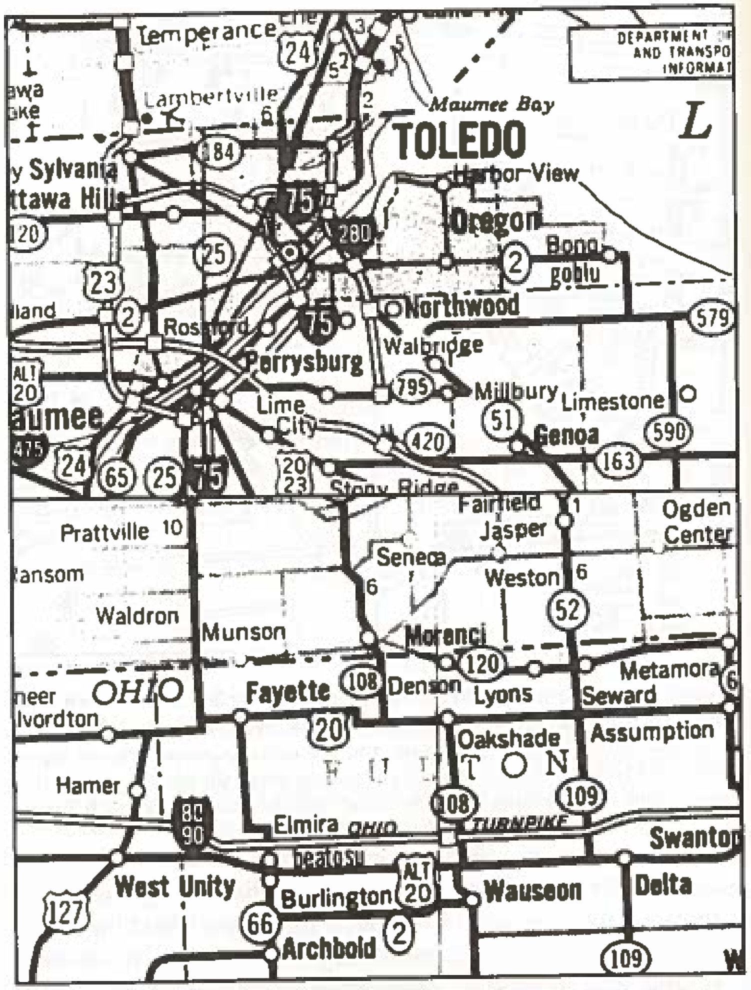


FIGURE 4.4. Fictitious towns "goblu" (above, to the right, below Bono) and "beatosu" (toward the bottom, above Burlington) on the 1979 Michigan highway map reflect an unknown mapmaker's support for the University of Michigan football team (the Blue) over its traditional rival, Ohio State University (OSU).

euphemism for this type of compilation is "editing the competition," but the legal term is copyright infringement—if you crib from a single source and get caught. To be able to demonstrate copyright infringement in court, and possibly to enjoy a

cash settlement by catching a careless competitor in the act, map publishers have been known to deliberately falsify their maps by adding "trap streets." As deterrents to the theft of copyright-protected information, trap streets are usually placed subtly, in out-of-way locations unlikely to confuse or antagonize map users. Map publishers are understandably reluctant to talk about this dubious practice of deliberate falsification.

Map drafters having fun are another source of cartographic fiction. Michigan's state highway map for 1979, for instance, included two fictitious towns reflecting the traditional football rivalry between the University of Michigan and Ohio State University. As figure 4.4 shows, the cartographic culprit clearly was a not only a Michigan fan but a loyal citizen of Michigan, perpetrating place-name pollution only in neighboring portions of Ohio—and perhaps thinking the editor would be less careful in checking out-of-state features. Toledo's new eastern suburb "goblu" is a slightly compacted version of the familiar cheer of fans rooting for the Michigan Blue (a nickname based on the primary school color), and the new town "beatosu," north of Burlington and south of the Ohio Turnpike, reflects the Blue's principal annual gridiron goal, defeating OSU.

A more personal example of creative cartography is Mount Richard, which in the early 1970s suddenly appeared on the continental divide on a county map prepared in Boulder, Colorado. Believed to be the work of Richard Ciacci, a draftsman in the public works department, Mount Richard was not discovered for two years. Such pranks raise questions about the extent of yet-undetected mischief by mapmakers reaching for geographic immortality.

Distorted Graytones: Not Getting What the Map Author Saw

Printing can radically alter the appearance of a map, and failure to plan for the distorting effects of reproduction can yield thousands of printed maps that look quite different from the original artwork. Graytones can be particularly fragile during two steps in map reproduction: the photographic transfer of the map image from a drawing or computer plot onto a photographic negative (or in some cases directly onto a press plate) and the transfer of the inked image from the printing plate onto the paper. Overexposure or underdevelopment of the photograph or an underinking printing press yields a faint image, from which fine type or fine area patterns might have disappeared. Underexposure or overdevelopment of the photograph or an overinking press might fill in the corners and tightly closed loops of small type and can noticeably darken some graytone area symbols. Usually the culprit is an overinking printing press, which fattens image elements through a malfunction called *ink spread*. Photographic exposure and development can be controlled, and if necessary the artwork can be reshot. But at least some overinking is common for a significant part of most printing runs, and in many cases significant overinking occurs throughout the run.

Fine dot screens used as area symbols are vulnerable to ink spread, and the map designer who underestimates the effect of ink spread on screens with dots spaced 120 or more to the inch (47 or more per centimeter) risks medium grays that turn out black and choropleth maps on which low becomes high and high becomes low. Figure 4.5 uses greatly enlarged views of two hypothetical graytone area symbols to illustrate this effect on a relatively fine 150-line screen (59 lines per centimeter), with dots spaced 0.007 inch (0.169 mm) apart. In this simulation a small amount of ink spread that increases the radius of each dot by 0.001 inch (0.025 mm) raises the black area of a 20 percent screen to 49 percent and that of an 80 percent screen to 96 percent. Screens for original graytones of less than 50 percent black grow darker because ink added around the edges of tiny black dots makes the dots larger, in some cases causing them to coalesce. Screens for original graytones of more than 50 percent black grow darker because ink added on the inside edges of tiny clear dots on a black background makes the clear dots smaller, in some cases completely filling them in. The cautious mapmaker adjusts screen texture to printing quality. The same amount of ink spread applied to moderately coarse, 65-line screens (26 lines per centimeter) would increase the 20 percent screen to only 32 percent black and the 80 percent screen to only 89 percent black.

The map user should be particularly wary of choropleth maps with both fine and coarse dot screens. Before examining the mapped patterns for clusters of dark, high-value areas and

light, low-value areas, the user should inspect the key for possible inversion of part of the graytone sequence. Ink spread might, for example, have reproduced a 40 percent fine graytone as darker than a comparatively coarse 60 percent graytone. Overinking might also have aggregated the two or three highest categories into a single group represented by solid black area symbols. Ink spread can produce particularly troublesome distortions of visual contrast and graphic logic on maps printed in color.

Laser printers and other remote electronic displays are another source of graphic noise. Pattern description codes that produce one symbol on the map designer's computer might yield a very different area symbol when the map is printed or displayed on the other side of the room or, if transmitted in compact "object code" over a telecommunications network, on the other side of the continent or world. A graphics

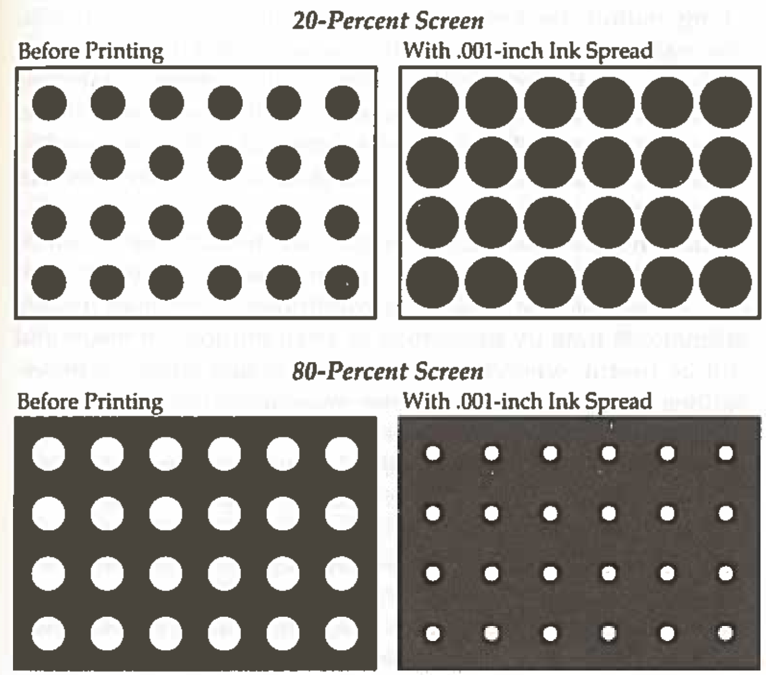


FIGURE 4.5. Enlarged diagram showing the effects of 0.001 inch of ink spread on 20 percent (above) and 80 percent (below) 150-line graytone dot screens.

workstation with a moderate-resolution monitor might display a graytone area symbol at 36 dots per inch (14 dots per centimeter), for instance, whereas a distant laser printer programmed to take full advantage of its higher resolution might render the same area symbol with a much finer, 150-line dot screen. A map author unable to review a laser-printer image before it is printed could be surprised and embarrassed by the effect of ink spread on a graphically unstable fine-dot screen.

Temporal Inconsistency: What a Difference a Day (or Year or Decade) Makes

Maps are like milk: their information is perishable, and it is wise to check the date. But even when the map author provides one, the date might reflect the time of publication, not the time for which the information was gathered. And when the map was compiled from more than one source or through a long, tedious field survey, the information itself might be so temporally variable as to require not a single date but a range of dates. Particularly troublesome is the carefully dated or current-situation map that shows obsolete features or omits more recent ones. These errors might be few and not readily apparent; a map that is 99.9 percent accurate easily deceives most users.

Inaccurately dated or temporally inconsistent maps can be a particular hazard when the information portrayed is volatile. A map of past geological conditions might inaccurately estimate its date by thousands or even millions of years and still be useful, whereas the temperature and pressure observations used to prepare weather maps must be synchronized to within an hour or less. Moreover, maps forecasting weather patterns must state accurately the date and time for the forecast.

Historians in particular should be skeptical of dates on maps. Medieval maps, for instance, can cover a much broader range of time than a single year-date suggests. As historian of cartography David Woodward has noted, medieval *mappaemundi* (world maps), instead of providing an accurate or perceived image of the earth at an instant of time, often "consist of historical aggregations or cumulative inventories of events that occur in space." For instance, the famous Hereford map, named for the British cathedral that owns it, was compiled

about 1290 from a variety of sources. Its place-names present an asynchronous geography ranging from the fourth-century Roman Empire to contemporary thirteenth-century England.

Conscientious users of modern maps read whatever fine print the mapmaker provides. A large-scale topographic map released in year N with a publication date of year N – 2, for instance, might be based on air photos taken in year N – 3 or N – 4, and might have been field checked in year N – 2 or N – 3. But field checking might not detect all significant changes, and maps of areas undergoing rapid urban development often are appallingly obsolete. Moreover, derived maps without fine print can be the cartographic equivalent of snake oil. Because of publication delays and slow revision, "new" derived maps may well be ten years or more out of date. Or they may be only four years out of date in some areas, and ten years or more in others.

A temporally accurate map need not focus on a single date or include only features that currently exist or once existed. Planners, for instance, need maps to record their previous decisions about future projects so that new decisions do not conflict with old ones. Fairfax County, Virginia, learned that inaccurate planning maps can have costly consequences. Because of inadequate mapping, a developer was authorized to start building a subdivision in the path of a planned limited-access highway. Buying the seventeen affected lots cost county taxpayers $1.5 million. The same highway also required the county to buy and raze five new homes in another subdivision. After these two embarrassing incidents, county officials set up a special mapping office.

Temporal consistency is also troublesome for users of statistical maps produced from census and survey data. Choropleth maps and other data maps sometimes portray a ratio or other index that compares data collected for different instants or periods of time. Usually the temporal incompatibility is minor, as when the Bureau of the Census computes per capita income by dividing an area's total income for the calendar year 1989 by its population as enumerated on 1 April 1990—getting an accurate count of the population and honest and reliable estimates of personal income is far more problematic. One must beware, though, of indexes that relate mid-decade survey data and beginning-of-decade census tabulations, for

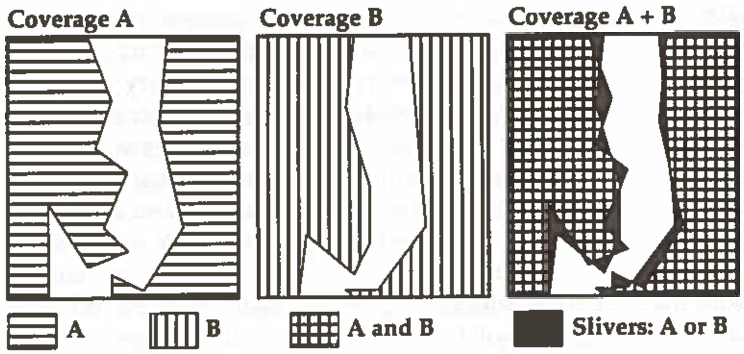


FIGURE 4.6. Spurious sliver polygons result from the overlay of two inaccurately digitized coverages A and B.

example, by dividing 1995 income totals by 1990 populations. Migration can significantly distort ratios based on asynchronous data, particularly for small areal units such as suburban towns, which can grow by 500 percent or more in ten years.

International data based on inconsistent definitions as well as asynchronous censuses or surveys can yield highly questionable maps. Worldwide maps of poverty, occupational categories, and the proportion of the population living in urban areas are inherently imprecise because of significant international differences in the relevant definitions. World maps based on statistical data are particularly suspect—whatever validity they have arises not only from knowing adjustment by scholars or United Nations officials but from broad, very general categories that tend to mask spurious differences within the groups of more and less developed nations.

Maps based on electronic data files can be highly erroneous, especially when several sources contributed the data and the user or compiler lacked the time or interest to verify their accuracy. Obviously inaccurate data from careless, profit-driven firms have unpleasantly surprised purchasers of street network information, and even data from reliable vendors occasionally have infuriating errors. Without careful editing, streets are more easily omitted or misplaced in a computer database than on a paper street map. Transfer from paper map to electronic format invites many inconsistencies, especially when many sets of features, or "coverages," can be related in a geographic information system. Roads and boundaries entered from adjoining map sheets often fail to align at the common edge, and features extracted from different maps of the same area might be misaligned, perhaps because their source maps have different projections. Overlays of spatially or temporally incompatible data—for example, the overlay of two closely related coverages collected separately— can yield slivers and other spurious polygons, as the example in figure 4.6 demonstrates. Moreover, the software used to process the information might also be flawed. Although software errors can be blatant, the possibility of subtle programming errors that could lead to disastrous decisions should encourage the user to carefully explore unfamiliar data and software. "Garbage in, garbage out" is a useful warning, but sometimes you can't tell the data are garbage until they have been used for a while.

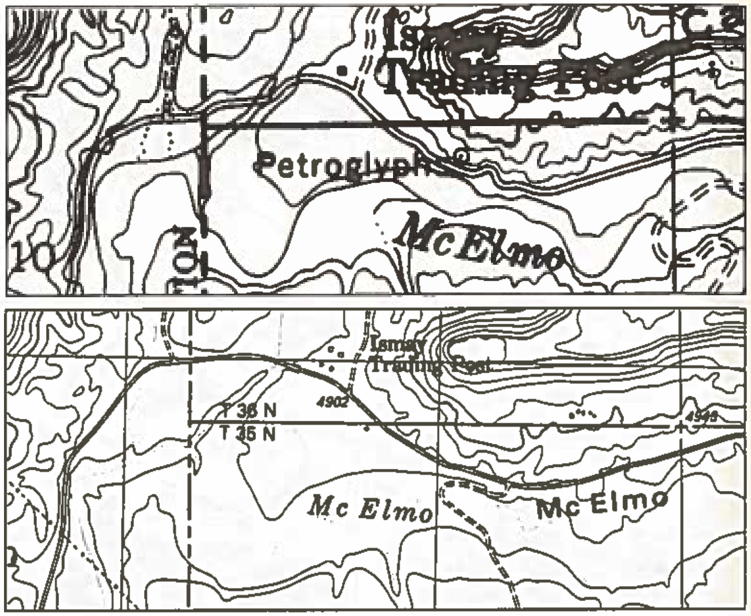


FIGURE 9.8. A late 1950s topographic map (above, enlarged from 1:62,500 to 1:24,000) pointed out the petroglyphs just southeast of the Ismay Trading Post, but more recent 7.5-minute maps (below) omit this and numerous other archaeological sites.

impossible to assess, accuracy and completeness clearly can have unintended consequences.

National mapping organizations are easy to criticize for much the same reason maps must lie: the enormous number of choices in selecting features, assigning symbols, and setting the scale of a map series. There's the added responsibility, though, of keeping tens of thousands of large-scale maps current through periodic review and revision. Because of limited resources, even a range of map series cannot satisfy the diverse requirements of map users. And while agencies cope with political pressures for greater cost-recovery, more effective federal-state-local coordination, and increased privatization, technological advances create ever more promising yet costly choices as well as a growing need to reassess priorities and revise standards. To be adequately informed, the map user must be at least vaguely aware of how cartographic bureaucracies work, what they value, and how values and biases affect their products.

Chapter 10

DATA MAPS: MAKING NONSENSE OF THE CENSUS

A single set of numerical data, say, for the states of the United States, can yield markedly dissimilar maps. By manipulating breaks between categories of a choropleth map, for instance, a mapmaker can often create two distinctly different spatial patterns. A single map is thus just one of many maps that might be prepared from the same information, and the map author who fails to look carefully at the data and explore cartographic alternatives easily overlooks interesting spatial trends or regional groupings. Moreover, because of powerful personal computers and "user-friendly" mapping software, map authorship is perhaps too easy, and unintentional cartographic self-deception is inevitable. How many software users know that using area-shading symbols with magnitude data produces misleading maps? How many of these instant mapmakers are aware that size differences among areal units such as counties and census tracts can radically distort map comparisons? In addition to the ill-conceived charts of hacker-cartographers, wary map users must watch out for statistical maps carefully contrived to prove the points of self-promoting scientists, manipulating politicians, misleading advertisers, and other propagandists.

This chapter uses several simple hypothetical examples to examine the effects of areal aggregation and data classification on mapped patterns. Read it carefully and look closely at the maps and diagrams, and this excursion into cartographic data analysis should be richly rewarding rather than technically tedious. Anyone interested in public-policy analysis, marketing, social science, or disease control needs to know how maps based on census data can yield useful information as well as flagrant distortions.

Aggregation, Homogeneity, and Areal Units

Most quantitative maps display data collected by areas such as counties, states, and countries. When displayed on a map, presented on a statistical plot, or analyzed using correlation coefficients or other measures, geographic data produce results that reflect the type of areal unit. Because different areal aggregations of the data might yield substantially different patterns or relationships, the analyst should qualify any description or interpretation by stating the type of geographic unit used. Noting that values generally increase from north to south "at the county-unit level" warns the reader (and the mapmaker as well!) that a different trend might arise with state-level data, for instance.

Areal aggregation can have a striking effect on the mapped patterns of rates and ratios. A ratio such as the average number of television sets per household might, for example, produce radically different maps when the data are aggregated separately by counties and by the towns that make up these counties. The three town-level maps in figure 10.1 are spatially

Number of Televisions

1,000	100	50	100	50	100	50
200	100	200	100	200	100	200
100	200	100	4,000	100	200	100
200	400	200	400	200	400	3,000

Number of Households

2,000	200	100	200	100	200	100
200	100	200	100	200	100	200
100	200	100	4,000	100	200	100
100	200	100	200	100	200	1,500

Televisions per Household

0.5	0.5	0.5	0.5	0.5	0.5	0.5
1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0
2.0	2.0	2.0	2.0	2.0	2.0	2.0

FIGURE 10.1. Town-unit number tables showing number of televisions (top left), number of households (top right), and average number of televisions per household (bottom) for twenty-eight hypothetical towns.

Number of Televisions

2,300	5,700	4,150

Number of Households

3,100	5,500	2,600

Televisions per Household

0.74	1.04	1.60

FIGURE 10.2. County-unit number tables of number of televisions (left), number of households (middle), and average number of televisions per household (right) for a three-county aggregation of the twenty-eight hypothetical towns in figure 10.1.

ordered number tables, without graphic symbols, so that we can see how rate calculations depend on what boundaries are used and how they are drawn. The upper left-hand map shows the number of televisions in each of twenty-eight towns, the upper right-hand map represents the number of households, and the lower map portrays the television-ownership rate. Note the straightforward top-to-bottom pattern of the rates: low in the upper tier of towns, average in the two middle tiers, and high in the lower tier. Note also that three towns in the upper left, lower right, and just below the center of the region have relatively high numbers of households. These variations in household density underlie the markedly different left-to-right trend in television-ownership rates in figure 10.2, based on the same data aggregated by county.

Spatial pattern at the town-unit level of aggregation depends on how somewhat arbitrary political boundaries group towns into counties. Figure 10.3 uses two additional aggregations of these twenty-eight towns to demonstrate the possible effect of historical accident. The upper row of maps shows an alternative aggregation of towns into three horizontal counties that reflect the town-level top-to-bottom trend. In contrast, the lower series of maps shows an equally plausible aggregation into four counties, three based on the concentrations of households and one comprising the balance of the region. The television-ownership map for this lower set isolates what might be more urban counties from a single much larger, more rural county with an average of slightly more than one television per household. Graytone area symbols would yield very different choropleth maps for the three sets of rates shown in the right-hand maps of figures 10.2 and 10.3.

Another example illustrates how areal aggregation can affect geographic pattern. Whereas figure 10.3 demonstrates that different aggregations of towns into counties can yield markedly different county-level patterns, figure 10.4 illustrates how a single aggregation can produce the same county-level pattern from markedly different town-level patterns. Note that the town-level maps in figure 10.4 reflect a pattern of television-ownership rates very different from that in figure 10.1. Note in particular the progression of rates from a tier of low-ownership towns across the bottom of the region to a peak of much higher rates at the upper right. Yet when aggregated according to the county boundaries in figure 10.2, these data will yield similar county-unit rates. Comparing this trio of spatial number tables with those in figure 10.1 demonstrates the importance of stating clearly the data units used and of not assuming that a trend apparent at one level of aggregation exists at other levels as well.

The counties in these examples obviously are not homogeneous. But can we assume homogeneity even within the towns? What spatial variations in the distribution and density of these 11,200 households lie hidden in the network of town boundaries? Figure 10.5 presents one of many plausible point patterns that could produce the aggregated town-level counts and rates in figure 10.1. Three types of point symbols represent

Number of Televisions

1,450
5,900
4,800

Number of Households

2,900
5,900
2,400

Televisions per Household

0.5
1.0
2.0

Number of Televisions **Number of Households** **Televisions per Household**

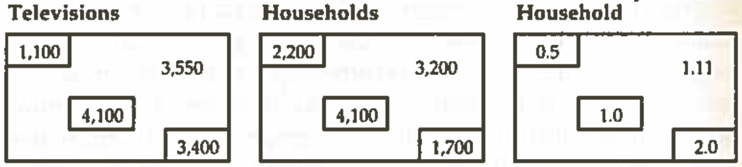


FIGURE 10.3. County-unit number tables based on other aggregations of the twenty-eight towns into counties.

Number of Televisions

190	285	200	350	350	210	890
455	450	1,085	960	895	520	1,260
355	315	525	480	595	360	700
130	120	80	100	80	110	100

Number of Households

100	150	100	200	100	50	100
350	300	700	600	500	200	300
500	450	700	600	700	400	500
650	600	400	500	400	550	500

Televisions per Household

1.90	1.90	2.00	1.75	3.50	4.20	8.90
1.30	1.50	1.55	1.60	1.79	2.60	4.20
0.71	0.70	0.75	0.80	0.85	0.90	1.40
0.20	0.20	0.20	0.20	0.20	0.20	0.20

FIGURE 10.4. Patterns of the number of televisions, number of households, and the television-ownership rate radically different from those in figure 10.1 could yield county-unit patterns identical to those in figure 10.2.

groups of 10, 100, and 500 households. Each symbol represents a group of households owning an average of 0, 1, or 2 televisions. The small, ten-household symbols represent rural residences, which lack TV receivers for religious reasons, lack of cable service, or a deep commitment to reading. Because of rough terrain, swamps, park or forest land, and undeveloped federal land, large parts of the region are uninhabited. Of the six large villages, with 400 or more households, two have two-TV households on the average, two have one-TV households, and two have video-free households. Although figure 10.5 contains elements of both the top-to-bottom town-level trend in figure 10.1 and the left-to-right county-unit trend in figure 10.2, its pattern of television ownership is more similar to the lower right of figure 10.3, where county boundaries segregate three large population clusters from the balance of the region. Yet even here the differences are striking, again demonstrating how the configuration of areal units can hide interesting spatial detail and present a biased view of a variable's geography.

Aggregation's effects become even more serious if the careless analyst or naive reader leaps from a pattern based on areal

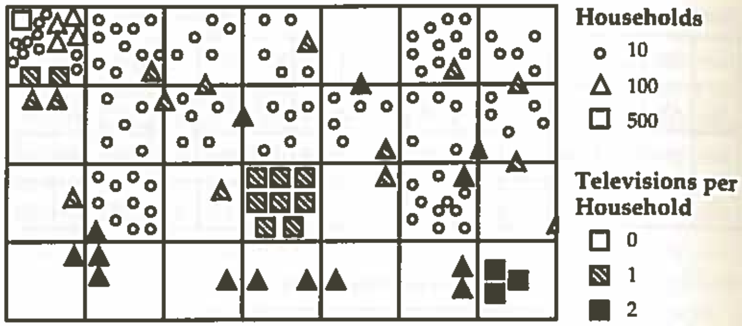


FIGURE 10.5. Detailed map of television ownership for villages and rural households illustrates one possible spatial structure that could yield the town-unit and county-unit maps in figures 10.1 and 10.2.

units to conclusions based on individual households. Consider, for instance, the large village toward the lower right-hand corner of figure 10.5. The average television-ownership rate here of 2.0 need not mean that each of the village's 1,700 houses has two TVs. Some households might have none while others have three or four or five. One or two residents might even be compulsive collectors, so that more than half the homes have one or none.

If households collecting old television sets seems farfetched, consider average household income, an index used frequently by social scientists and marketing analysts. Because of one or two innovative, unscrupulously manipulative, or otherwise successful residents, a small village might have an enormous *mean* household income. More a statistical quirk than a realistic reflection of overall local prosperity, this high average income might mask the employment of most villagers as household servants, gardeners, or security guards. Because nondisclosure rules prohibit a more precise publication of individual incomes, aggregated census data are the most refined information available. They provide an average for the place but say little about individual residents.

Are areally aggregated data bad? Surely not. In many cases, particularly in public policy analysis, towns and counties are the truly relevant units for which state and federal governments allocate funds and measure performance. And even more highly aggregated data can be useful, for instance, when governors and senators want to compare their states with the

other forty-nine. Local officials and social scientists concerned with differences among neighborhoods readily acknowledge the value of geographic aggregation. Moreover, nondisclosure regulations needed to ensure cooperation with censuses and surveys require aggregation, and areally aggregated data are better than no data at all. Thus persons who depend upon local-area data encourage the Bureau of the Census to modify boundaries to preserve the homogeneity of *census tracts* and other reporting areas. And when tract data are not adequate, they sometimes pay for new aggregations of the data to more meaningful areal units.

What else can the conscientious analyst do? Very little aside from the obvious: know the area and the data, experiment with data for a variety of levels of aggregation, and carefully qualify all conclusions.

And what should the skeptical map user do? Look for and compare maps with different levels of detail, and be wary of cartographic manipulators who choose the level of aggregation that best proves their point.

Aggregation, Classification, and Outliers

Choropleth mapping further aggregates the data by grouping all areas with a range of data values into a single category represented by a single symbol. This type of aggregation addresses the difficulty of displaying more than six or seven visually distinct graytones in a consistent light-to-dark sequence. Often the mapmaker prefers only four or five categories, especially when the area symbols available do not afford an unambiguous graded series. (For aesthetic reasons or to avoid confusion with interior lakes or areas without data, black and white are not good graytone symbols for choropleth maps.)

But classification introduces the risk of a mapped pattern that distorts spatial trends. Arbitrary selection of breaks between categories might mask a clear coherent trend with a needlessly fragmented map or oversimplify a meaningfully intricate pattern with an excessively smoothed view. Figure 10.6 illustrates the influence of class breaks on the appearance of choropleth maps of the town-level television-ownership rates in figure 10.4. Note that the map on the left presents a

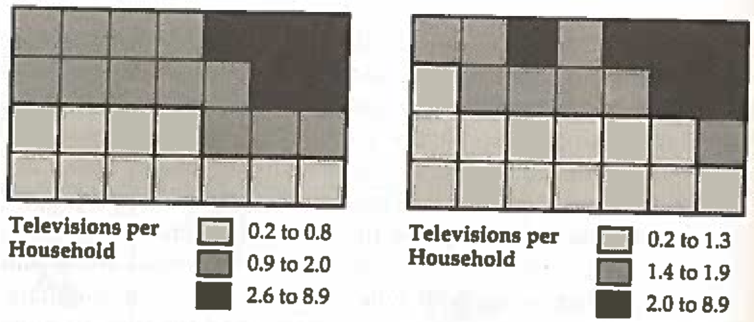


FIGURE 10.6. Different sets of categories yield different three-category choropleth maps for the data in figure 10.4.

clear, straightforward, readily remembered upward trend toward a peak at the upper right of the region, whereas the map at the right offers a more fractured view of the same data.

Classification raises many questions. Which map, if either, is right? Or if "right" sounds too dogmatic, which provides a better representation of the data? Don't both maps hide much variation in the broad third category, represented by the darkest symbol? Shouldn't the seven towns with rates of 0.2 occupy a category by themselves? Is a difference of, say, 0.1 at the lower end of the overall range of data values more important than a similar difference at the upper end? Can a three-class map provide even a remotely adequate solution?

But how many map authors bother to ask these questions? Because choropleth mapping is readily available through personal computers, so that the map viewer is often also the mapmaker, some instruction in map authorship is warranted.

Software vendors usually provide a few options for "automatic" classification, and naive mapmakers often settle for one of the easier options. Sometimes the computer program even provides a map instantly, without offering a choice of classification strategies. Called a "default option," this automatic choice of class breaks is a good marketing ploy because it gives the hesitant prospective purchaser an immediate success.

But does the default give you a good map? Figure 10.7 shows four-category mapped patterns produced by two common default classing options for the same town-level television-ownership data used in figure 10.6. The *equal-intervals*

scheme, on the left, divides the range (8.7) between the lowest and highest data values (from 0.2 to 8.9) into four equal parts (each spanning 2.175 units). Note, though, that this classification assigns most of the region to a single category and that the third category (from 4.6 to 6.7) is empty. Of possible use when data values are uniformly distributed across the range, the only consistent asset of equal-interval classification is ease of calculation.

In contrast, the *quartile* scheme, on the right, ranks the data values and then divides them so that all categories have the same number of areal units. Of course, only an approximately equal balance is possible when the number of areas is not a multiple of four or when a tie thwarts an equal allocation (as occurs here at the upper left, where the highest category receives both of the towns with rates of 1.9). Although the map pattern is more visually balanced, the upper category is broad and highly heterogeneous, and the break between the second and third categories falls between two very close values (1.3 and 1.4). Yet the map based on these four quartile categories does have meaning for the viewer interested in the locations of towns in the highest and lowest quarters of the data values. Called *quintiles* for five categories and *quantiles* more generally, this rank-and-balance approach can accommodate any number of classes.

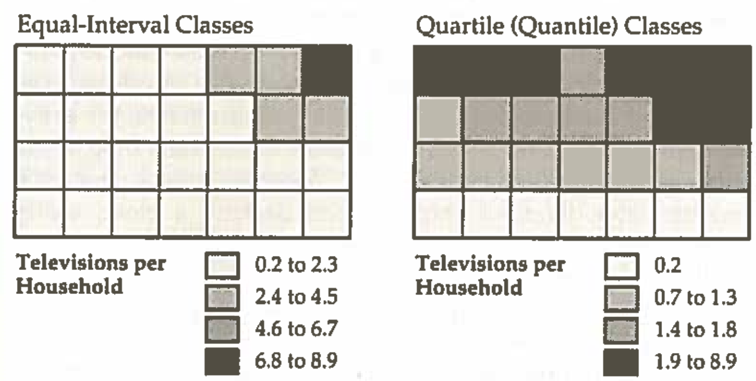


FIGURE 10.7. Two common classing schemes used as "defaults" by choropleth mapping software yield radically different four-category patterns for the data in figure 10.4.

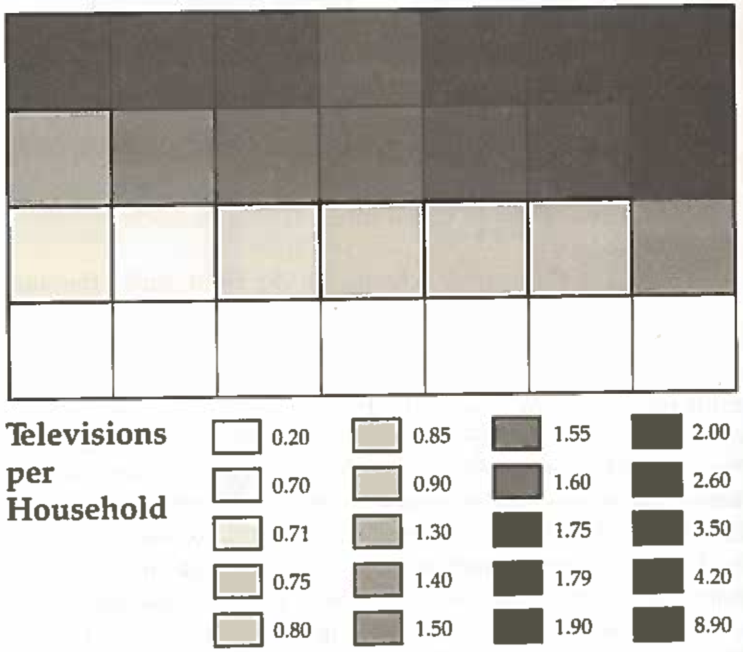


FIGURE 10.8. Continuous-tone, nonclassed choropleth map for the data in figure 10.4.

At least one mapping program offers the option of a "no-class" or "classless" choropleth map, on which each unique data value (perhaps up to fifty of them) receives a unique graytone. In principle this might seem a good way to sidestep the need to set class breaks. But as figure 10.8 illustrates, the graytones might not form a well-ordered series, and the map key is either abbreviated or cumbersome. Moreover, assigning each unique value its own category can destroy a clear, easily remembered picture of a strong, meaningful spatial trend. This ideal solution might not be so ideal after all.

Eschewing defaults and panaceas, the astute map author begins by asking two basic questions: How are the data distributed throughout their range? And what, if any, class breaks might have particular meaning to the map viewer? The answer to this second question depends on the data and on whether the map author deems useful a comparison with the national

or regional average. On state-level maps, for instance, a break at the United States average would allow governors and senators to compare their constituents' or their own performance with that of the rest of the nation. Of course the map key would have to identify this break to make it truly meaningful.

After addressing the question of meaningful breaks, the conscientious map author might then plot a *number line* similar to that in figure 10.9. A horizontal scale with tick marks and labels represents the range of the data. Each dot represents a data value, and identical values plot at the same position along the scale, one above the other. The resulting graph readily reveals natural breaks, if any occur, and distinct clusters of homogeneous data values, which the classification ought not subdivide. Number lines allow the map author to visualize the distribution of data values and to choose an appropriate number of categories and appropriate positions for class breaks. Computer algorithms can also search the data distribution for an optimum set of breaks, but in many cases the computer-determined optimum is not significantly better than a visually identified suboptimal grouping. Rounded breaks and a more balanced allocation of places among categories can be important secondary factors in choropleth mapping.

Extremely high or extremely low values isolated from the rest of the distribution can confound both human cartographers and sophisticated mapping software. Should these *outliers* be grouped with markedly more homogeneous clusters higher or lower on the number line? Should each be accorded its own category? Can two or three widely separated data values at either end of the distribution be grouped into a single highly heterogeneous category? Or should each outlier be treated as its own category, with its own symbol, at the risk of

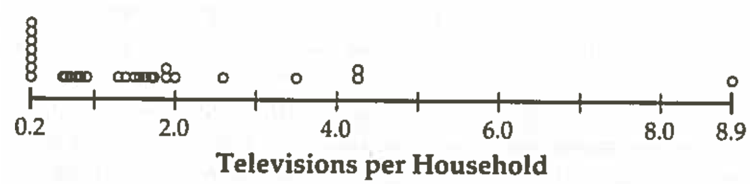


FIGURE 10.9. Number line for the town-level television-ownership rates in figure 10.4.

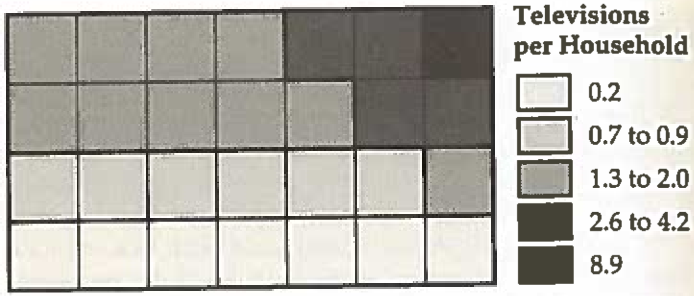


FIGURE 10.10. Choropleth map based on the number line in figure 10.9 and the character of the data.

reducing graphic differentiation among graytones? Or might the map author treat outliers as outcasts—errors or deviants that "don't belong"—and either omit them or give them a special symbol?

No simple, standard solution addresses all outliers. The map author should know the data, know whether these deviant values are real or improbable, and know whether a large difference between outliers really matters. Also important is the relation of outliers to the theme of the map and the interests of map viewers. For the television-ownership data in figure 10.9, an average of 8.9 TVs per household surely is not only exceptional but probably significantly higher than its neighboring values at 4.2. If not an error, it deserves special treatment in a category of its own. The next four lower values, 4.2 (twice), 3.5, and 2.6, might then constitute a single category; all are above the more plausible rate of 2.0, and yet 4.2 TVs per household is not improbable, especially in an affluent area.

Other breaks seem warranted between 0.9 and 1.3, a gap that includes the inherently meaningful rate of one TV per household, and between 0.2 and 0.7, to separate the seven videophobic towns at the lower end of the distribution. The resulting five-category map in figure 10.10 provides not only an honest, meaningful representation of the data values and their statistical distribution, but a straightforward portrayal of the spatial trend as well. An arbitrary classification, such as a computer program's default categories, is unlikely to do as well, even with six or more categories.

Classification, Correlation, and Visual Perception

Choropleth maps readily distort geographic relationships between two distributions. Hastily selected or deliberately manipulated categories can diminish the visual similarity of two essentially identical trends or force an apparent similarity between two very different patterns.

Consider as a case in point figure 10.11, a spatial data table and number line for the mean number of children per household, which has a strong town-level relationship to television ownership. Although the range of data values is not as broad for this index of family size, the highest values are at the upper right and the lowest values occur across the bottom of the region. Towns toward the right and toward the top of the region generally have more children in the home than do towns toward the bottom or left edge of the map. That the pair of maps in figure 10.12 shows identical spatial patterns for children and televisions is thus not surprising.

Statistical analysts commonly depict correlation with a two-dimensional scatterplot, with data values for one variable measured along the vertical axis and those for the other scaled along the horizontal axis. A dot represents each place, and the density and orientation of the point cloud indicates the strength and direction of the correlation. Figure 10.13 is a pair of scatterplots, both showing the strong positive association between the household rates for children and TVs. The perpendicular lines extending from the scales of the left-hand

2.25	2.30	2.31	2.35	2.52	2.41	3.20
2.04	2.15	2.18	2.20	2.27	2.42	2.53
1.62	1.53	1.63	1.63	1.71	1.92	2.11
0.37	0.39	0.31	0.36	0.49	0.44	0.30

Average Number of Children per Household

FIGURE 10.11. Spatial data table and number line for average number of children per household.

scatterplot into the scatter of points represent the class breaks in figure 10.12. These two sets of four lines each divide the scatterplot into an irregular five-by-five grid. Because all dots on the left-hand scatterplot lie within one of the five diagonal cells, the two five-category maps in figure 10.12 have identical patterns, enhancing the impression of a strong correlation.

Figure 10.13's right-hand scatterplot adds some cartographic skulduggery. As before, the perpendicular lines from

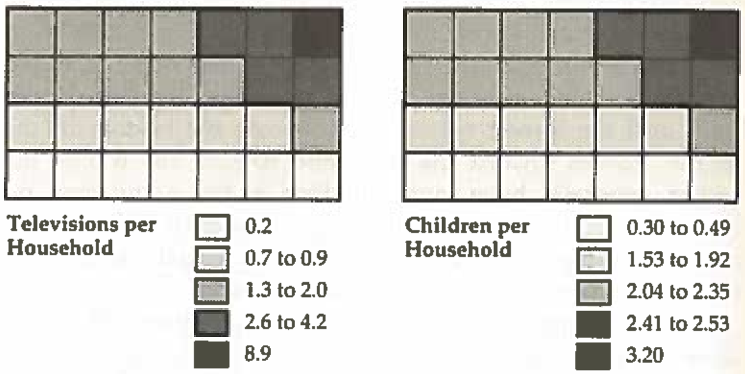


FIGURE 10.12. Choropleth maps with identical patterns for television ownership rate and average number of children per household.

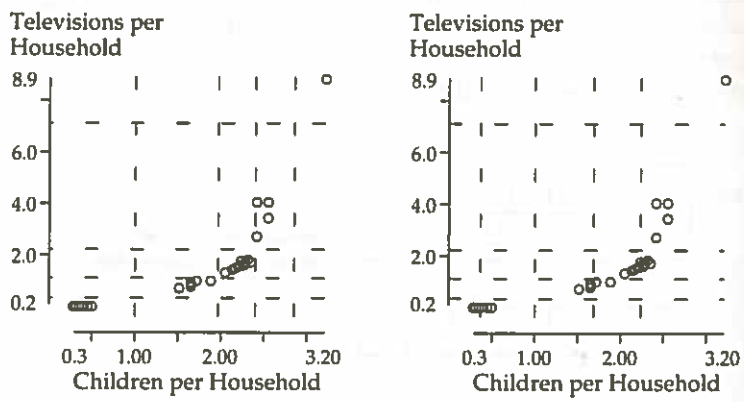


FIGURE 10.13. Scatterplots for the town-level television-ownership rate and average number of children per household. Additional lines on the left-hand scatterplot represent class breaks for the pair of maps in figure 10.12. Additional lines on the right-hand scatterplot show breaks used in figure 10.14.

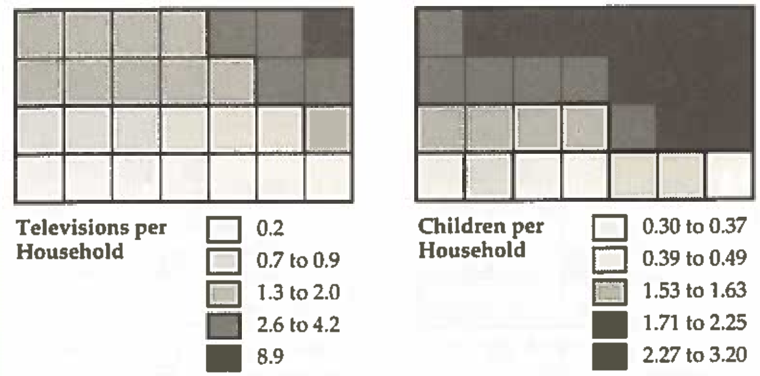


FIGURE 10.14. Distinctly different choropleth maps suggest minimal correlation between television ownership and family size.

the scales into the point cloud represent class breaks and form a five-by-five grid. But note that this configuration of breaks places all but four dots in an off-diagonal cell so that few towns will belong to the same category on both maps. Figure 10.14 demonstrates the resulting dissimilarity in map pattern and suggests a mediocre correlation at best. Similar tactics might make a weak relationship appear strong, especially if the maps are identical for the highest category, with the darkest symbol. Indeed, the spatial correspondence of the darkest, most eye-catching symbols strongly influences judgments of map similarity by naive map viewers. Some will even regard as similar two maps with roughly equal amounts of the darkest symbol—even if the high areas are in different parts of the region! Different area symbols for the two maps and different numbers of categories are other ways of tricking the map viewer or deluding oneself.

Another visual distortion might lie in the base map the data are plotted on. Not all sets of areal units are as uniform and visually equivalent as the square towns in the preceding examples. Figure 10.15 demonstrates this point with a deceptively similar-looking pair of maps based on the numerical data and class breaks of the visually dissimilar maps in figure 10.14. These twenty-eight towns vary markedly in size, and similarity is high because the largest towns belong to the same category. Towns not in the same category on both maps are smaller and less visually influential.

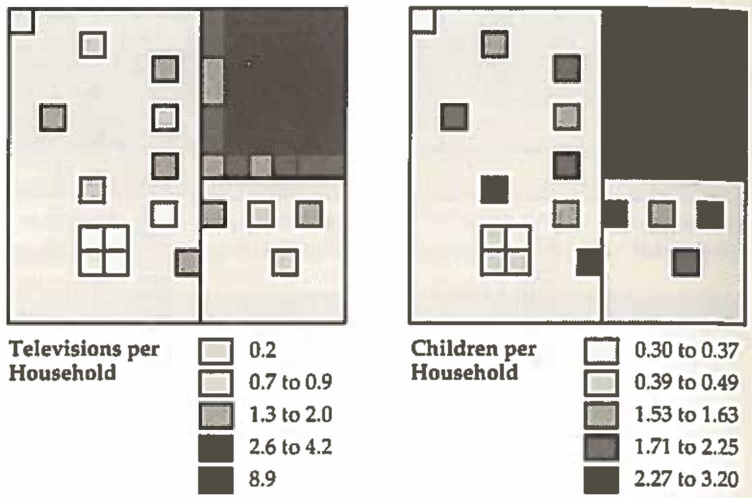


FIGURE 10.15. Similarity among large areas can distort visual estimates of correlation by masking significant dissimilarity among small areas. Numerical data and mapping categories are identical to those for the more obviously dissimilar pair of maps in figure 10.14.

Although this example is contrived, it is not atypical. Wards, census tracts, congressional districts, and other areal units designed to have similar populations often vary widely in area because of variations in population density. Disparities are even worse on county-unit maps, where populous metropolitan counties often are much smaller than rural counties with few inhabitants. The careful map viewer never judges numerical correlation by the similarity in map pattern alone and is especially cautious when some data areas are much bigger than others.

To avoid estimates of correlation biased by the size of areal units, the astute analyst will inspect the more egalitarian scatterplot, on which identical dots represent each area. As figure 10.16 illustrates, the density and orientation of the point cloud reflect the strength and direction of the correlation. If a straight line provides a good generalization of the point cloud, the correlation is called linear and the scatter of points around the line indicates the strength of the *linear correlation*. Positive relationships slope upward to the right, negative relationships slope downward to the right, and a point cloud without a dis-

cernible relationship has no apparent slope. Weak correlations have a wide, barely coherent scatter about the trend line, whereas for strong linear correlations most points are near or on the line. Not all correlations are linear, though; a strong *curvilinear correlation* has a marked curved trend, which a curved line fits better than a straight line.

Statisticians use a single number, the *correlation coefficient*, to measure the strength and direction of a linear correlation. Represented by the symbol r, the correlation coefficient shows the direction of the relationship by its sign and the strength of the relationship by its absolute value. The coefficient ranges from +1.00 to –1.00; r would be .9 or higher for a strong positive correlation, –.9 or lower for a strong negative correlation, and close to zero for an indeterminate or very weak correlation. (As a rule of thumb, squaring r yields the proportion of one variable's variation accounted for by the other variable. Thus, if r is –.6, the correlation is negative and one variable might be said to "explain" 36 percent of the other variable. A correlation coefficient measures only association, not causation, which depends upon logic and supporting evidence.)

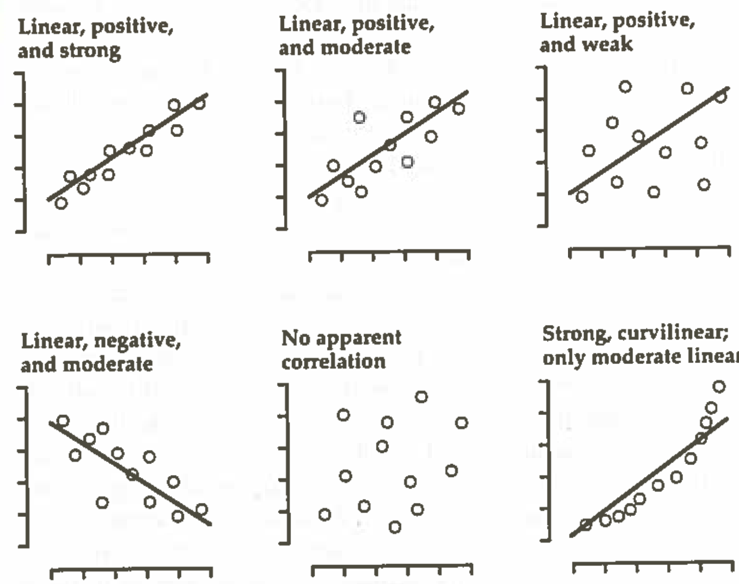


FIGURE 10.16. Scatterplots and trend lines for various types of correlation.

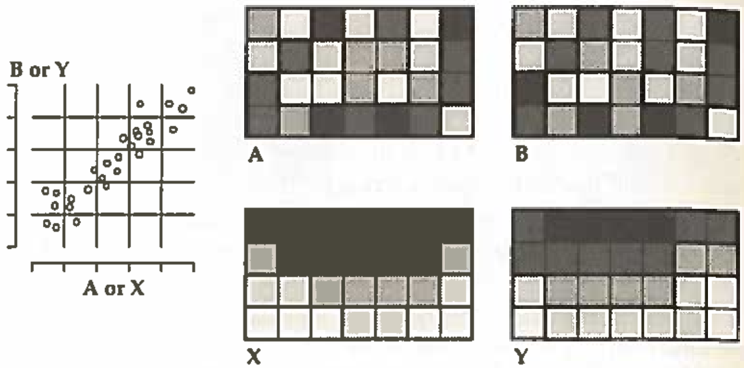


FIGURE 10.17. Two pairs of variables with identical scatterplots, correlation coefficients ($r = .93$), and class breaks, yet distinctly different map patterns.

Maps, scatterplots, and correlation coefficients are complementary, and the analyst interested in correlation relies on all three. The correlation coefficient, which provides a concise comparison for a pair of variables, measures only linear correlation. Yet a scatterplot quickly reveals a strong curvilinear relationship, with a mediocre value of *r*. Scatterplots also show outliers, which can greatly bias the calculation of *r*. But reliance upon visual estimation makes scatterplots poor for comparing strengths of relationships. Moreover, scatterplots and correlation coefficients tell us nothing about the locations of places, whereas maps, which present spatial trends, can offer unreliable estimates of correlation.

Maps also show a different kind of correlation, a *geographic correlation* distinct from the statistical correlation of the scatterplot and correlation coefficient. Statistical correlation is aspatial and reveals nothing about spatial trends. Figure 10.17 demonstrates this difference with two map pairs distinct in spatial pattern yet identical in scatterplot and correlation coefficient. Variables A and B, which share a comparatively chaotic, fragmented pattern, clearly differ in geographic correlation from variables X and Y, which have a distinct common trend with higher values toward the top of the region and lower values toward the bottom. Although not identical, the maps for X and Y suggest the influence of a third, underlying geographic factor, such as latitude, ethnicity, soil fertility, or proximity to a major source of pollution. Despite the problems posed by areal aggregation, the analyst of geographic data

who explores correlation without also checking for spatial pattern is either ignorant, careless, or callous. And the nonskeptical reader is easily misled.

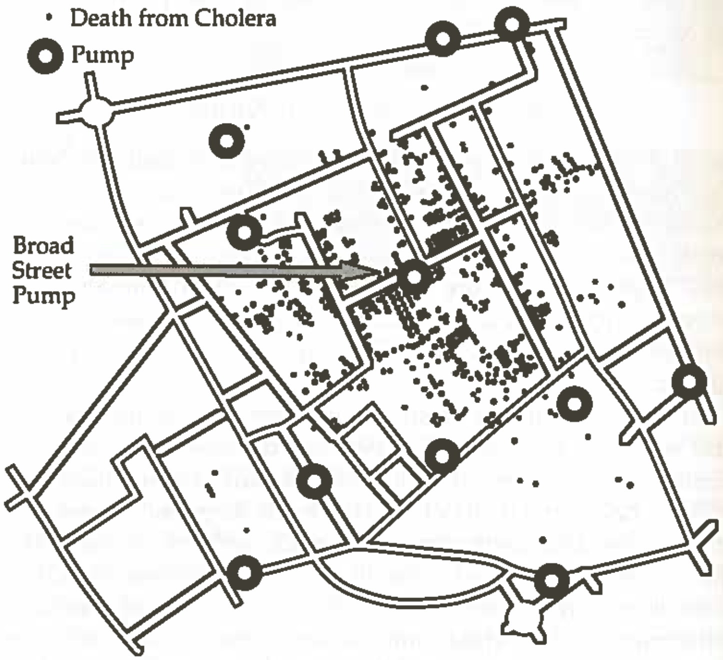
Places, Time, and Small Numbers

Areal data can yield particularly questionable patterns when choropleth maps show rates based on infrequent events, such as deaths from a rare type of cancer. Yet disease maps based on small numbers are a common tool of the epidemiologist, who uses mapping to explore the possible effects on human health of radon-rich soils, incinerators, and chemical waste dumps. But one question arises whenever the map shows a trend or cluster: Is the pattern real?

The problem is one of small numbers. Pandemics are rare, and seldom is the association between disease and an environmental cause so overwhelming that the link is easily identified and unchallenged. Clusters of deaths or diagnosed cases usually are few and unspectacularly small, perhaps no more than three deaths in a town or two in the same neighborhood. Epidemiologists map these cases both as points, to get a sense of patterning, and by areal units, to adjust for spatial differences in the number of people at risk. After all, an area with half the region's cases is not remarkable if it has half the region's population. But what is the significance of a small area with two or three cases and a rate several times above the national or regional rate? Could this pattern have arisen by chance? Would one or two fewer cases make the area no longer a "hot spot"? If one more case were to occur elsewhere, would this other area also have a high rate? To what extent does the pattern of high rates reflect arbitrary boundaries, drawn in the last century to promote efficient government or thirty years ago to expedite delivery of mail? Might another partitioning of the region yield a markedly different pattern? Might another level of aggregation—larger units or smaller units—alter the pattern? Is the mapping method inflating the significance of some clusters? And is it possibly hiding others?

Consider, for example, the maps in figure 10.18. At the top is John Snow's famous map showing cholera deaths clustered around the Broad Street Pump. A physician working in London during the cholera epidemic of 1854, Snow suspected drinking water as the source of infection. At that time homes

Snow's Dot Map



Areal Aggregations and Density Symbols

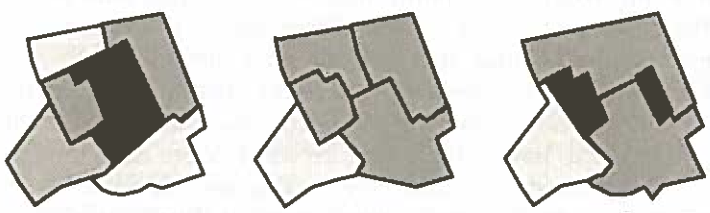


FIGURE 10.18. A reconstruction of John Snow's famous dot map of cholera (above) and three choropleth maps (below) produced by different areal aggregations of this part of London.

did not have running water, and people carried buckets from a nearby pump. Snow's map provided important evidence for the waterborne transmission of cholera; when authorities removed the pump's handle, new cases in this part of the city plummeted.

But what might have happened had Snow not worked with point data? The three maps at the bottom of figure 10.18 show how various schemes of areal aggregation might have diluted the Broad Street cluster. If addresses are available, as on most death certificates, aggregation to census tracts or other areal units larger than the city block increases the risk of missing intense, highly local clusters.

Aggregation involves not only areal units but also time, disease classification, and demography. One solution to the question of significance is to get more data by collecting information over a longer time span. Adding together several years of data, or even several decades, dampens the effect of chance occurrences but risks involving a wider range of causal agents. Aggregation over time might, for instance, mask important temporal trends, dilute the impact of new or abated environmental contaminants, or incorporate difficult-to-measure effects of population mobility. Likewise, combining several disease categories or the mortality of diverse demographic groups promotes stability and significance by increasing the number of cases and broadening the set of causes.

Clearly one map is not sufficient, although one good map can signal the need for a more detailed investigation. It is then up to a variety of scientific researchers to explore further the effects of geography and environment by examining employment and residential histories, characteristics of residence and neighborhood, and hereditary factors; by carefully studying maps at various levels of spatial, temporal, and demographic aggregation; through computer simulation to test the stability of known clusters and automated pattern recognition to identify new ones; and through related clinical and laboratory studies. Although maps can indeed lie, they can also hold vital clues for the medical detective.

Indexes, Rates, and Rates of Change

Another danger of one-map solutions is a set of measurements that presents an unduly positive or negative view. Often the map author has a single theme in mind and has several variables to choose from. Usually some variables are markedly more optimistic in tone or pattern than others, and the name of the index can cast a favorable or an unfavorable impression in the map title. "Labor Force Participation," for instance, sounds optimistic, whereas "Job Losses" clearly is a pessimist's term. An appropriately brazen title offers a good way to overstate economic health or industrial illness.

If the picture is bleaker or brighter than suits your politics, try a rate of change rather than a mere rate. After all, minor downturns often interrupt a run of good years, and depressions do not last forever. If unemployment is high now but a bit lower than a year, six months, or a month ago, the optimist in power would want a map showing a significant number of areas with declining unemployment. Conversely, the pessimist who is out of power will want a map depicting conditions at least as bad as before the current scoundrels took over. A time interval that begins when proportionately fewer people were out of work will make the opposition party's point, especially if unemployment has become worse in large, visually prominent, mostly rural regions.

A useful index for the optimist is one with relatively low values, such as the unemployment rate, if conditions have improved, or an index with comparatively high values, such as employment level, if conditions are worse. Thus a drop of one percentage point from a base of 4 percent unemployment yields an impressive 25 percent improvement! Yet a substantial increase in the unemployment rate from 4 to 6 percent can be viewed more optimistically as a drop in labor force participation from 96 to 94 percent—a mere 2 percent drop in employment.

Point symbols and counts, rather than rates, can be useful too. If the economy has been improving in all regions, the current government might want a map with graduated circles or bars showing actual counts beneath the title "Employment Gains." If the country is in a widespread recession, the opposition would use similar point symbols with the title "New Job Losses."

The cartographic propagandist is also sensitive to spatial patterns. Favorable symbols should be large and prominent, and unfavorable ones small and indistinct. Thus the optimist might present the unemployment data in figure 10.19 with the map at the lower left, to focus attention on improved conditions in larger areas, whereas the pessimist would prefer the map at the lower right, to emphasize the much greater number

Area	Labor Force (000s)	Unemployment (000s)			Unemployment Rate			Percentage Change
		t_1	t_2	Change	t_1	t_2	Change	
1	3,000	120	180	+60	4.0%	6.0%	+2.0%	+50.0%
2	16,000	640	800	+160	4.0%	5.0%	+1.0%	+25.0%
3	2,500	125	113	-12	5.0%	4.5%	-0.5%	-9.6%
4	800	56	48	-8	7.0%	6.0%	-1.0%	-14.3%
5	500	40	35	-5	8.0%	7.0%	-1.0%	-12.5%

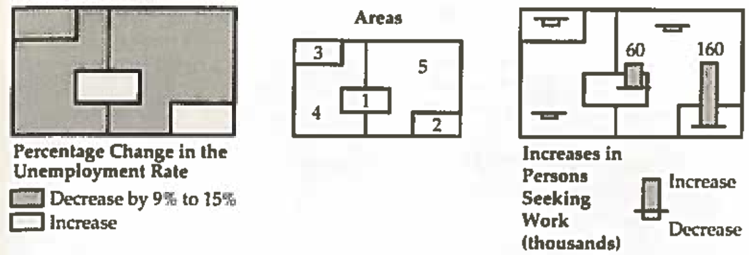


FIGURE 10.19. Unemployment data (top) for a hypothetical region (bottom center) yield different maps, supporting an optimistic (bottom left) and a pessimistic (bottom right) view of recent temporal trends.

of unemployed persons in more urban areas. Note as well how the titles and keys in these examples reinforce cartographic manipulation.

Labor economists, who commonly adjust unemployment data for seasonal effects, discourage some manipulation of time intervals. After all, more people are seeking work in early summer, when many high-school and college graduates enter the labor force for the first time. And more people find at least temporary work in November and December, the peak shopping season. Local seasonal effects, such as tourism and the temporary hiring of cannery workers in agricultural areas, also require seasonal adjustment.

Mortality, fertility, and other phenomena that do not affect all segments of the population equally also require adjustment. Figure 10.20, a comparison of the age-adjusted death rate with the crude death rate, illustrates the advantage of mapping demographically adjusted rates. The map at the left is a simple rate, which does not consider such age differences as a relatively young population in Alaska and older populations in Arkansas and Maine. When the rates portrayed in the right-hand map are adjusted for age differences, Alaska and some southeastern states emerge as high-rate areas whereas the

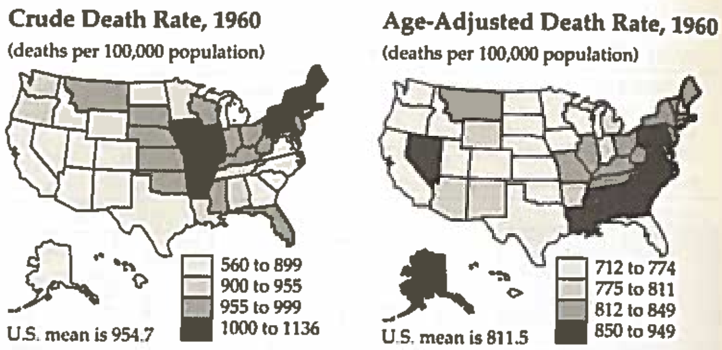


FIGURE 10.20. Maps of the crude death rate (left) and the age-adjusted death rate (right) can present markedly different geographic patterns of mortality.

Northeast and Midwest slip to a lower category. Age-adjustment allows the map at the right to reveal the effects of relatively good health care and a higher socioeconomic status in the New England, Middle Atlantic, and North Central states, widespread poverty and less accessible health care in the South of the 1960s, and the effects of accidents and isolation in Alaska.

When a single variable might yield many different maps, which one is right? Or is this the key question? Should there be just one map? Should not the viewer be given several maps, or perhaps the opportunity to experiment with symbolization through a computer workstation? If unable to trust the presenter's honesty and thoroughness, the skeptical viewer must question the representativeness of a single graphic. Guard against not only the cartographic manipulator, but also the careless map author unaware of the effects of aggregation and classification.

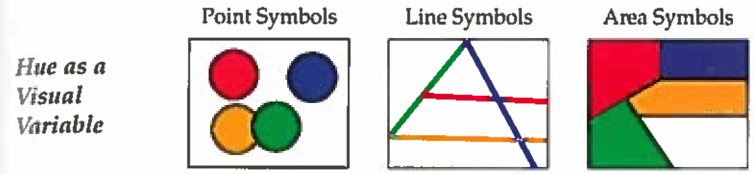


PLATE 1. For area symbols in particular, hue is often more forceful than the other five principal visual variables (fig. 2.11).

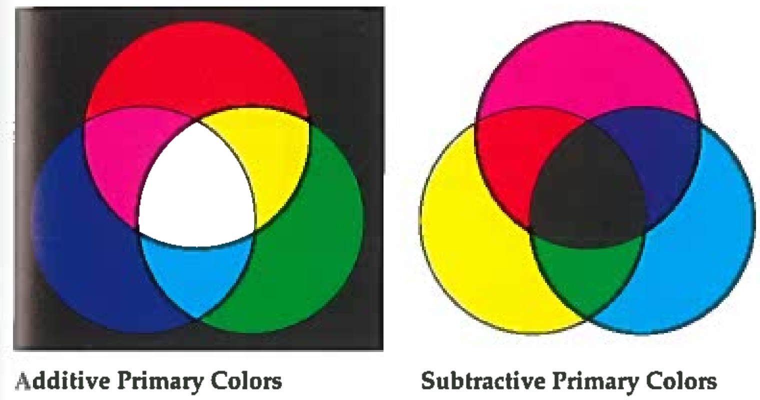


PLATE 2. Primary colors yield other hues when combined as either beams of light (left) or patches of dye (right).

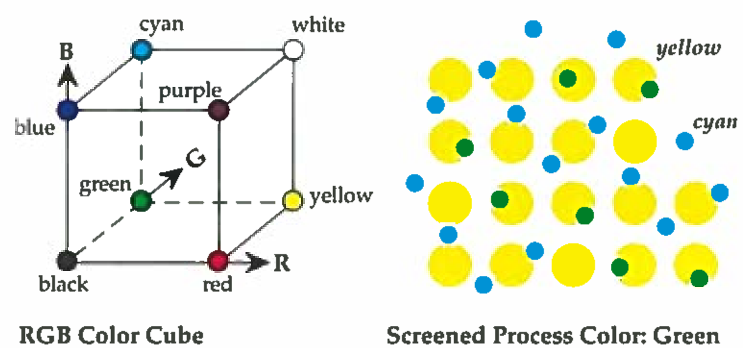


PLATE 3. RGB color cube (left) and greatly enlarged representation of overprinted screens of colored dots used in process printing to produce green (right).

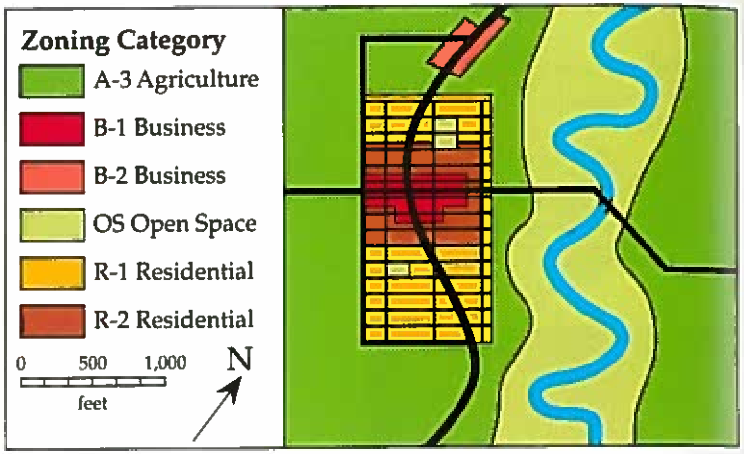


PLATE 4. Contrasting hues efficiently describe qualitative differences on zoning map shown in monochrome in figure 6.1.

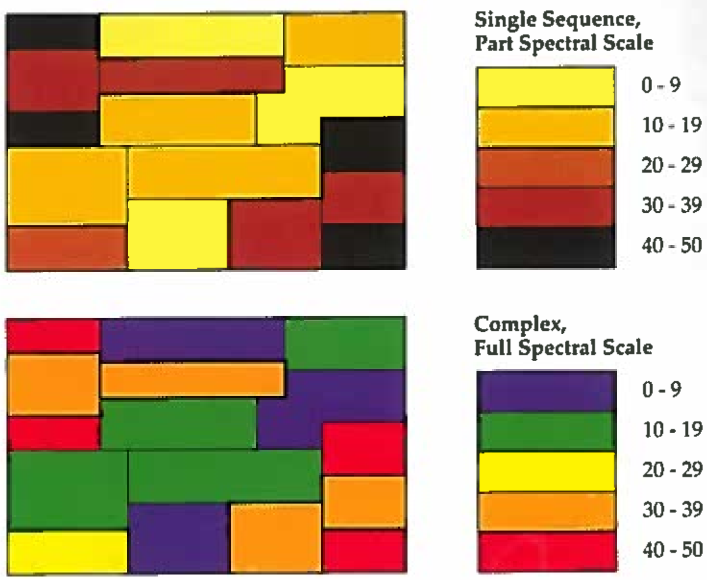


PLATE 5. Limited set of hues (top) is more easily grasped than an illogical, complex sequence of spectral hues (bottom).

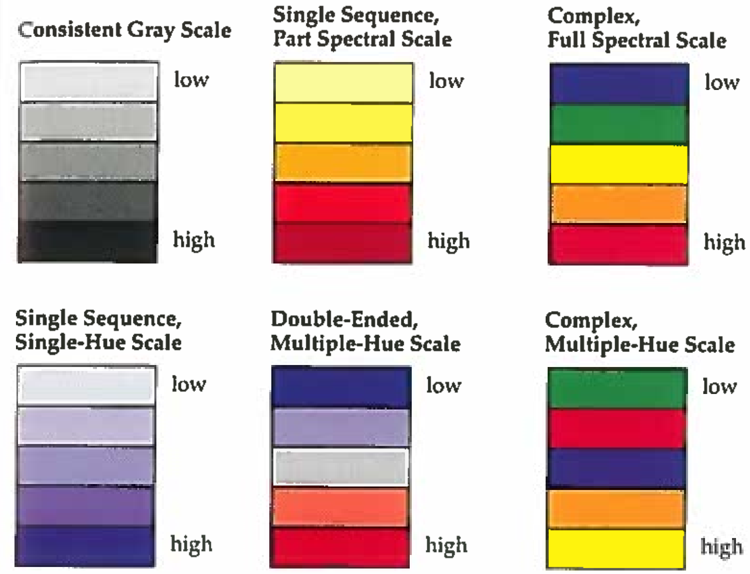


PLATE 6. Some color sequences found on choropleth maps.

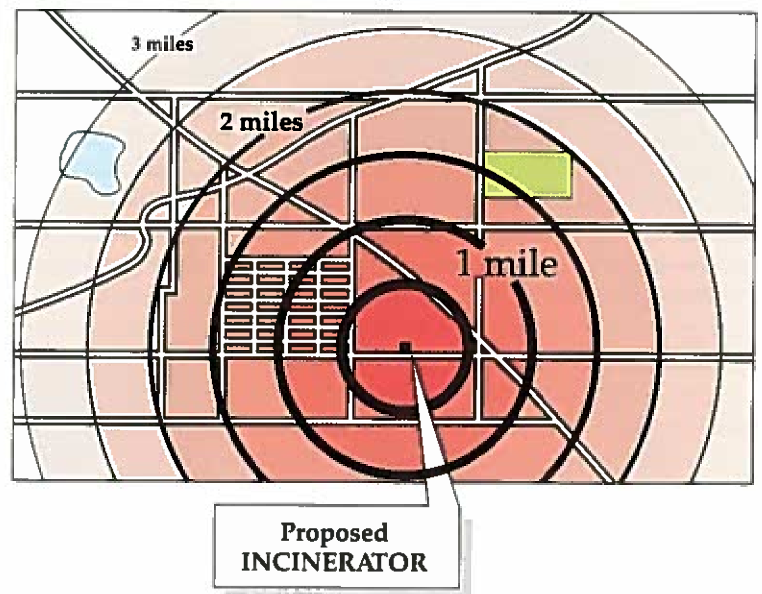


PLATE 7. Red area symbols connoting increased danger near the site of a proposed incinerator strengthen the message of a monochrome environmental propaganda map (fig. 7.19).

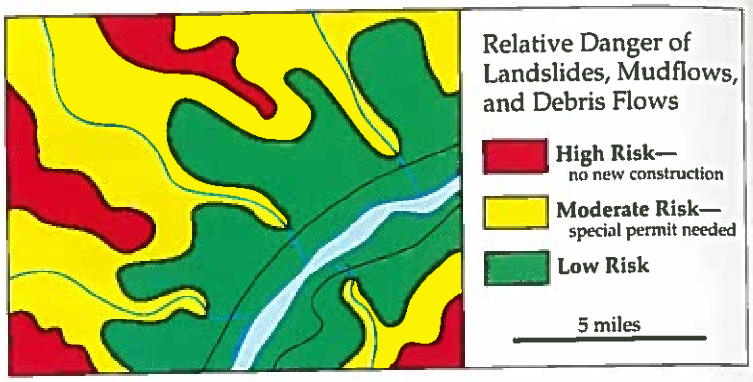


PLATE 8. The map viewer reminded of the graphic metaphor can readily decode a sequence of three traffic-light colors portraying degree of environmental risk.

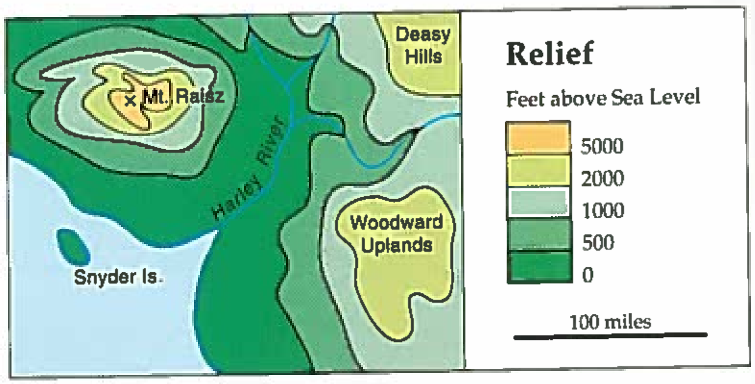


PLATE 9. Although hypsometric tints are widely used to portray relief with color-coded elevation categories, map viewers must be aware that lowland areas shown in green might well be dry and barren.

PLATE 10. Squares at the center of these three boxes are identical, but because of simultaneous contrast the gray center appears darker when surrounded by a more brilliant color.

Chapter 11

Color
Attraction and Distraction

Color is a cartographic quagmire. Color symbols can make a map visually attractive as well as fulfill the need for contrast on road maps, geological maps, and other maps with many categories. Yet the complexity and seductiveness of color overwhelm many mapmakers, and countless maps in computer graphics demonstrations, business presentations, and daily newspapers reveal a widespread ignorance of how color can help or hurt a map. Persons unaware of the appropriate use of color in cartography are easily impressed and might accept as useful a poor map that merely looks pretty.

Technological change accounts for much of the misuse of color on maps. Before the 1980s color printing was expensive and seldom used thoughtlessly, and color maps were comparatively rare. Advances since 1980 in electronic computing and graphic arts have encouraged a fuller use—and abuse—of color. Inexpensive color monitors, color printers, and slide generators have made color effortlessly available to the amateur mapmaker, and run-of-press color lithography encourages a similar misuse by cartographically illiterate commercial artists, responsible for most news illustration. Moreover, many viewers and readers expect maps with richly contrasting hues, even when black-and-white or more subdued symbols might be more readily and reliably decoded. This chapter briefly explains the nature of color and examines how graphic logic, visual perception, and cultural preferences affect the use of color on maps.

The Phenomenon of Color

As a biophysical phenomenon, color is a sensory response to electromagnetic radiation in a narrow part of the wavelength