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. A map can state its scale in three ways: as a ratio, as a short sentence, or 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 denominators and big fractions have small denominators, or that half (1/2) of 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 smallest 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.

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

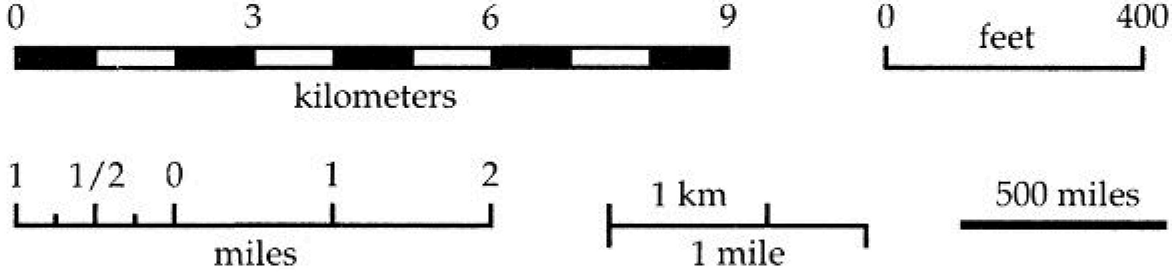


Figure 2.1. Types of map scales.

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 0.025 square miles, 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 square miles, 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 used to 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 eight hundred feet”), 1:24,000 (“one inch represents two thousand 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 for maps that might be reduced or enlarged for publication, or by users. For example, a 5-inch-wide map labeled “1:50,000” would have a scale less than 1:80,000 if reduced to fit a newspaper column or a mobile-device screen that is 3 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 digital maps, since screens and thus the map scales vary widely and unpredictably.

Web-based maps and similar interactive applications occasion a fourth type of map scale: the zoom slider that moves up or down to indicate relative distance above the surface, or the interactive plus and minus buttons that produce the same effect. Zooming out yields a broader geographic scope with a smaller scale and less detail, and zooming in provides a narrower view with greater detail.

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. Indeed, the bar scale sometimes included in the lower right of an online map can be blatantly misleading when the user zooms out to show an entire continent. 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, which are 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-geographic concepts 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.

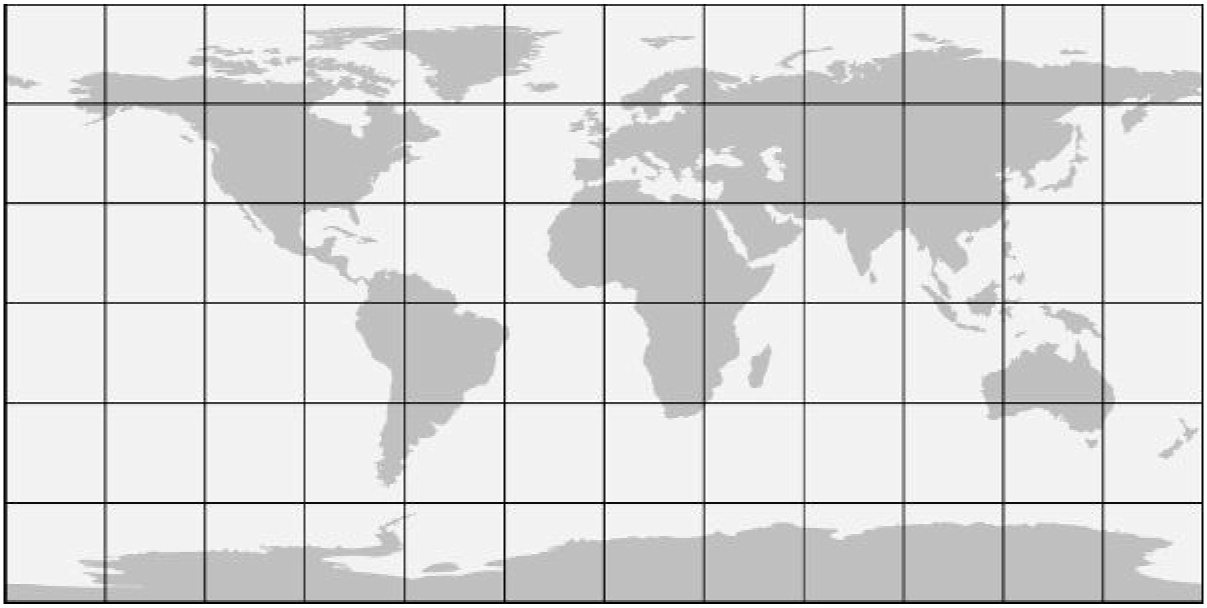


Figure 2.2. Equatorial cylindrical projection with true meridians.

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, it is valid 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 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.

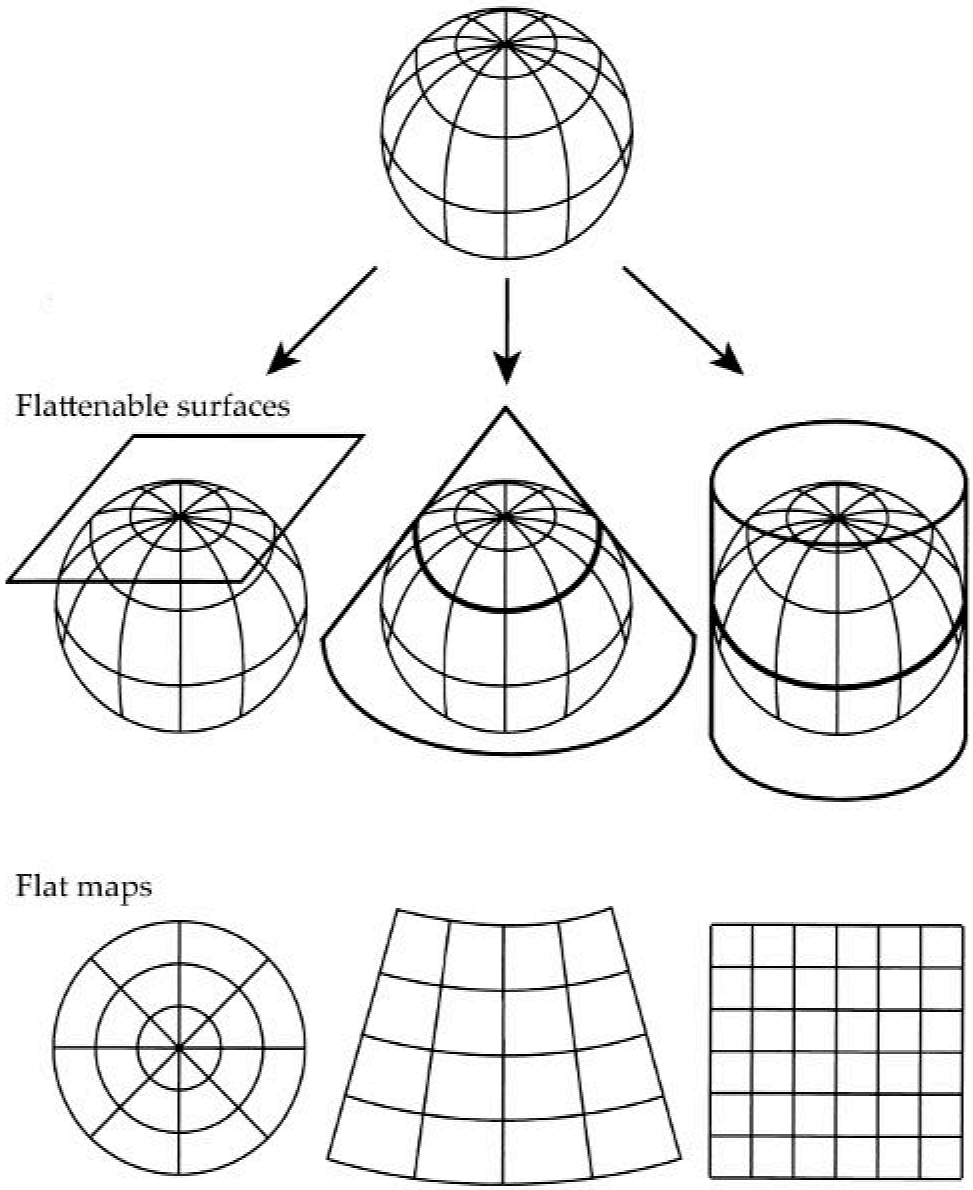
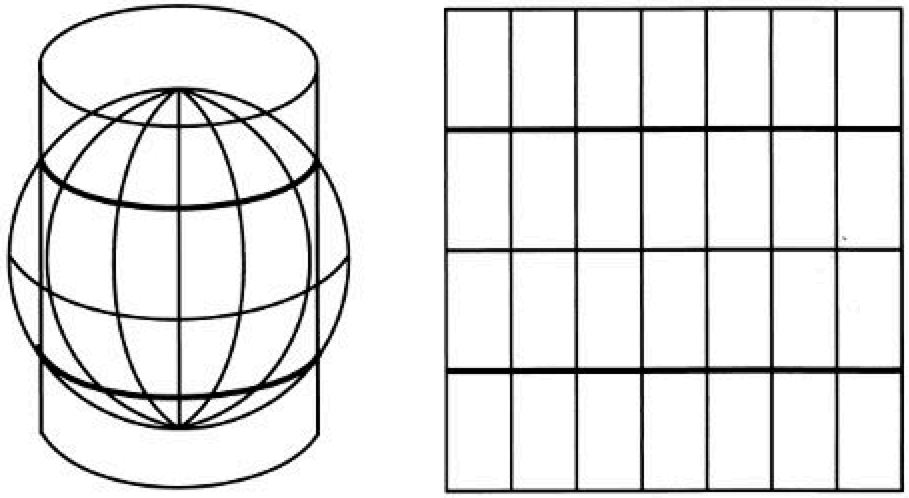
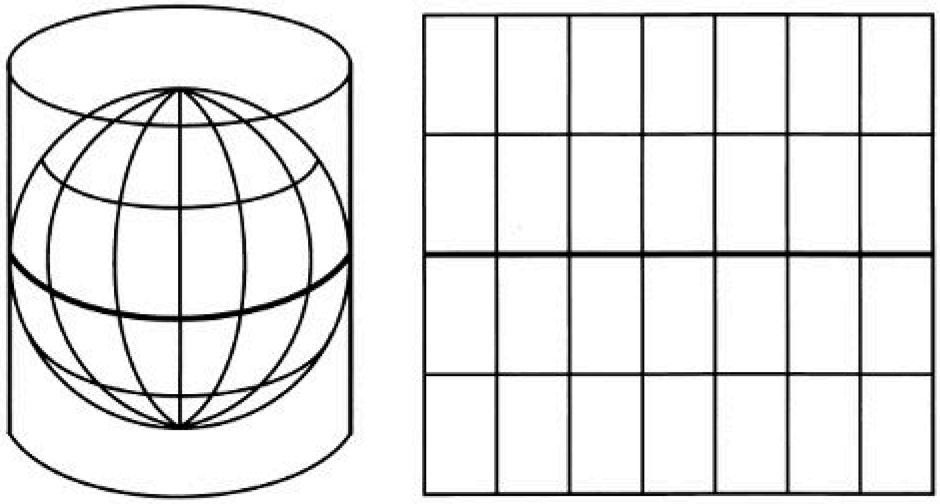


Figure 2.3. Developable surfaces in the second stage of map projection.

Figure 2.3 helps explain the meaning and limitations of ratio scales in 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 a *plane chart*, the equator is a standard line and the meridians show true scale as well.



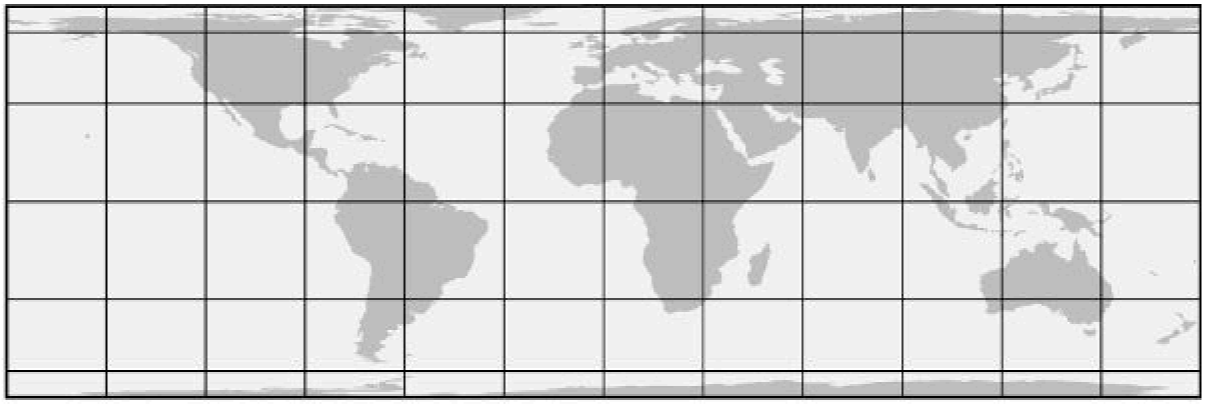
Secant cylindrical



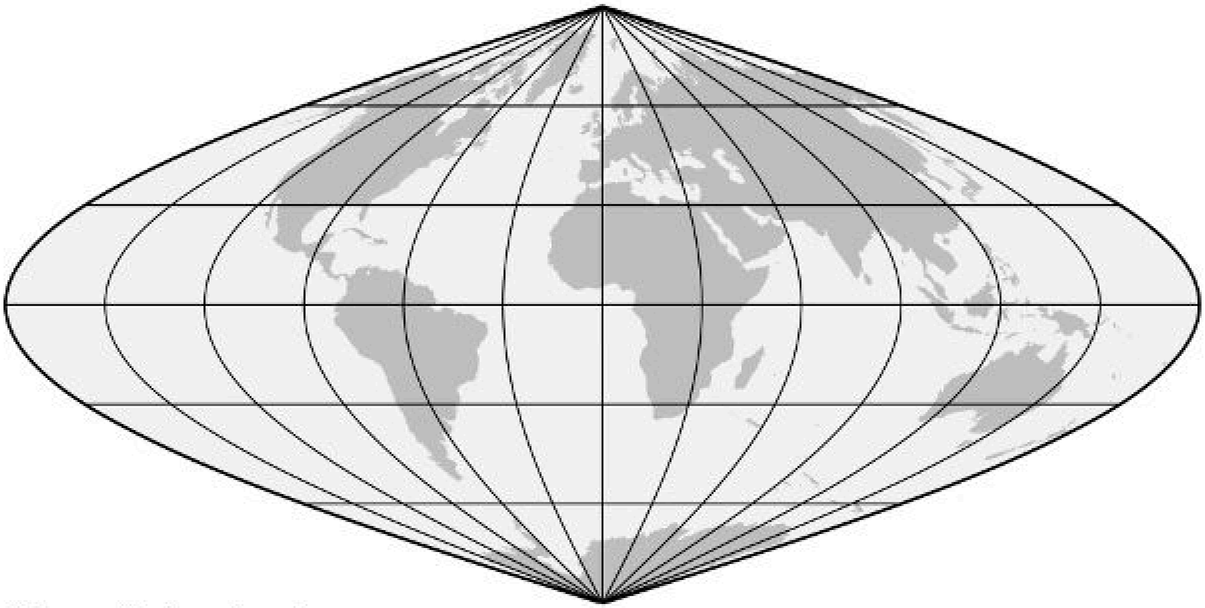
Tangent cylindrical

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

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 to one of the two standard lines. Conic projections are well suited to large mid-latitude areas, such as North America, Europe, and Russia, 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.



Cylindrical equal-area projection



Sinusoidal projection

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

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

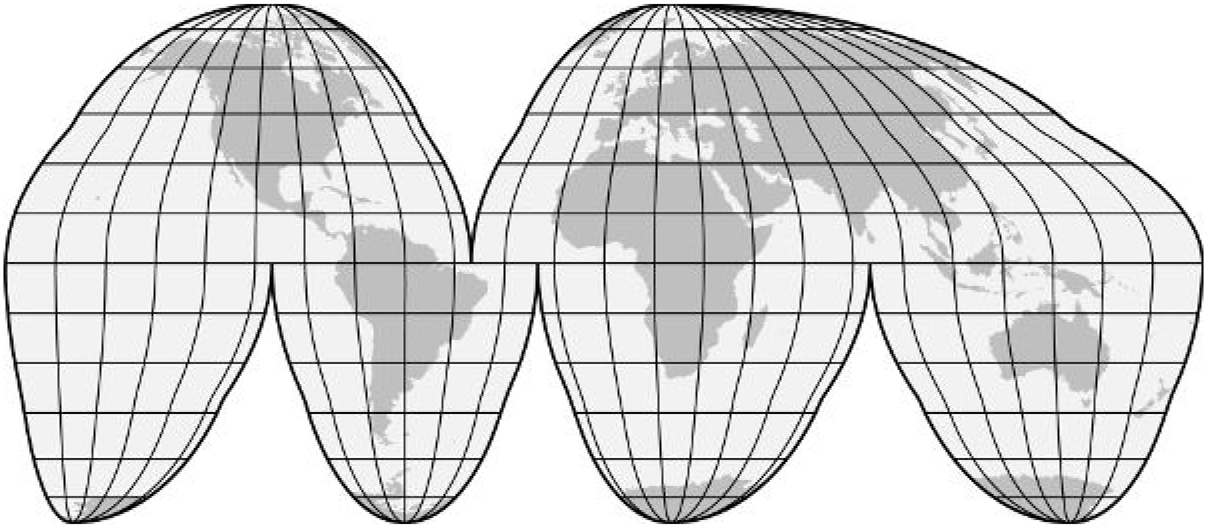


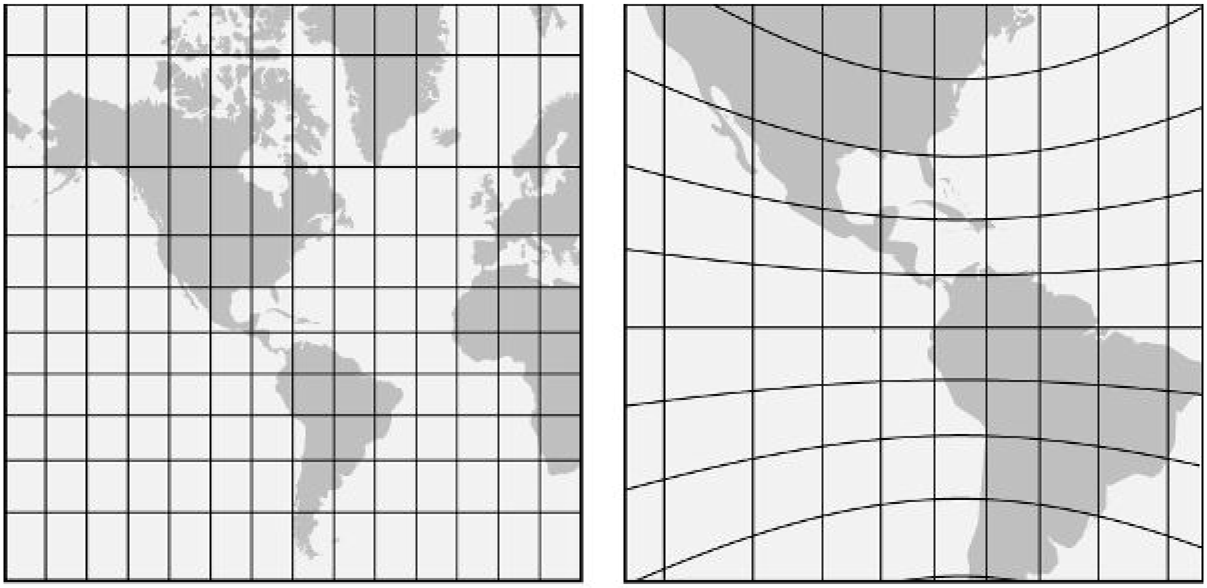
Figure 2.6. Goode's Homolosine Equal-Area projection.

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 landmasses. If interrupted over the land to minimize distortion of the oceans, Goode's projection can be equally adept at 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 compromised 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 five hundred thousand swine, for example, the spacing of these dots represents relative density. Important hog-producing regions, such as the American Midwest and southeastern China, 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 levels of hog production; if both regions have forty dots representing twenty million swine, the region occupying 2 square centimeters of the map would have a greater spacing between dots and appear less intensively involved in raising pigs than the region occupying only 1 square centimeter. 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 a globe and a 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.



Mercator projection

Gnomonic projection

Figure 2.7. Straight lines on an equatorially centered 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.

Perhaps the most striking trade-off in map projection is between conformality and areal 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 as 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.

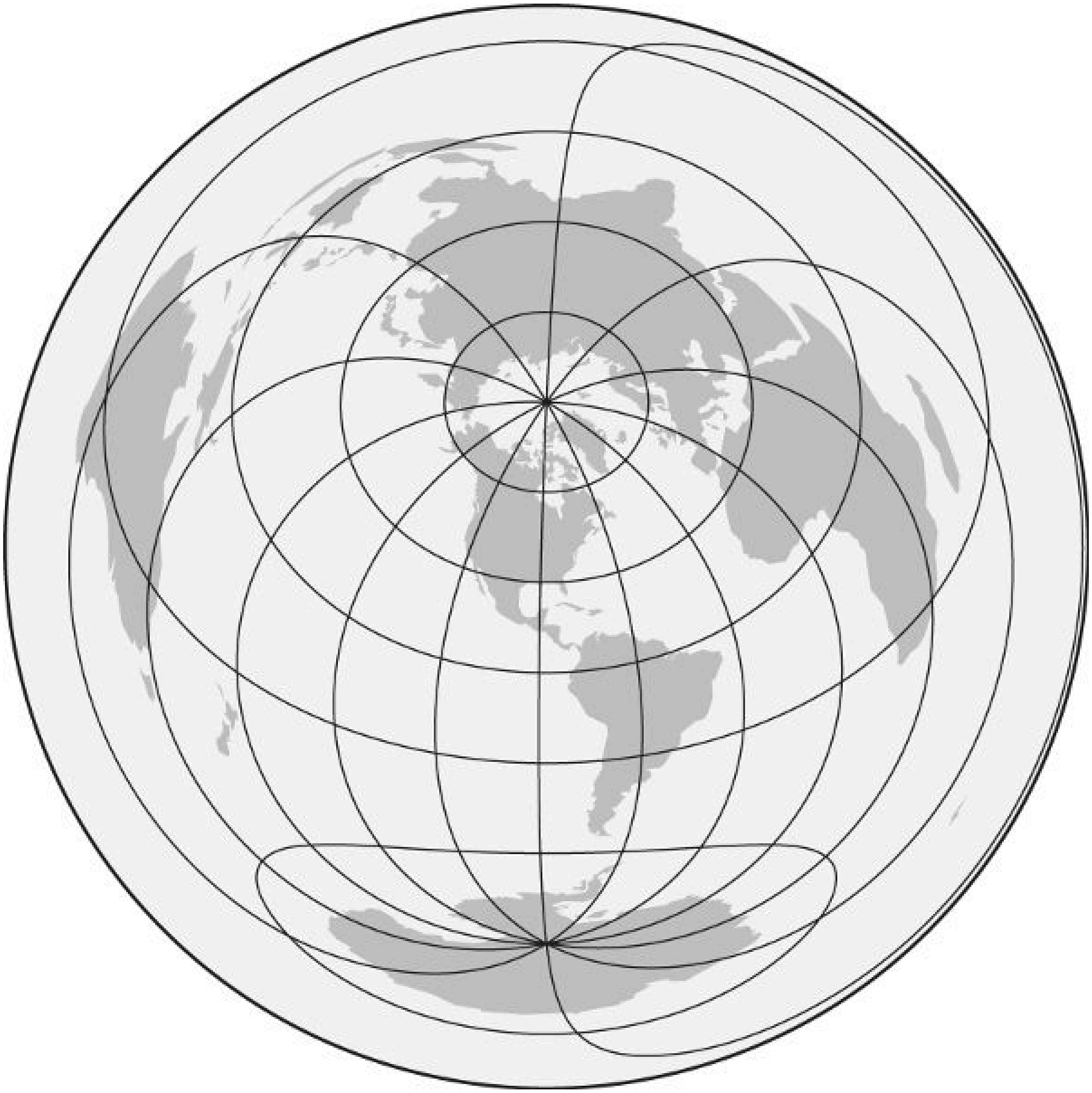


Figure 2.8. Oblique azimuthal equidistant projection centered on Chicago, Illinois, just east of the meridian at 90° W.

Why, then, are these projections used at all? Although presenting 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 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 graphics system and good mapping software, of course, map users can become their own highly versatile mapmakers and tailor projections to many unique needs. For instance, an azimuthal equidistant map centered on North Korea might better inform a discussion of possible missile attacks by a rogue nation than concentric circles drawn thoughtlessly on an off-the-shelf world map.

Among the more highly tailored map projections are *cartograms*, which portray such relative measures as travel time, 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 2-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 10-pound parcel more nearly reflect Watertown's relative proximity (only 70 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.

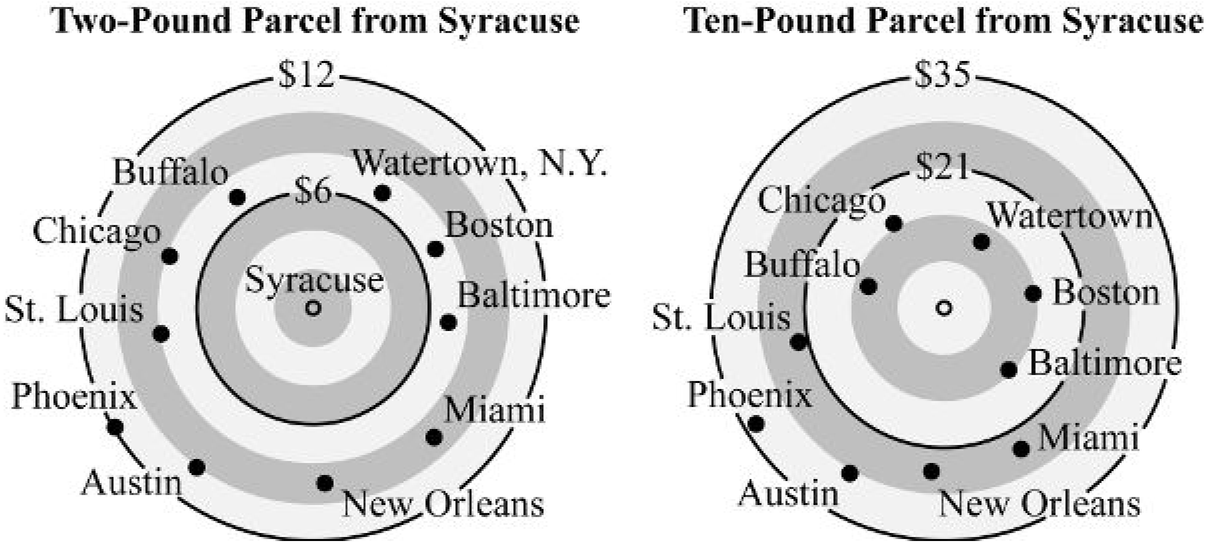


Figure 2.9. Distance cartograms showing relative spaces based on parcel-post rates from or to Syracuse, New York.

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 as vastly larger than Canada because the Indian population is more than thirty-five times larger than the Canadian population, even though Canada's area of 3.8 million square miles is markedly larger than India's 1.2 million square miles. The cartogram has merged some countries with 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 communicational and analytical needs.

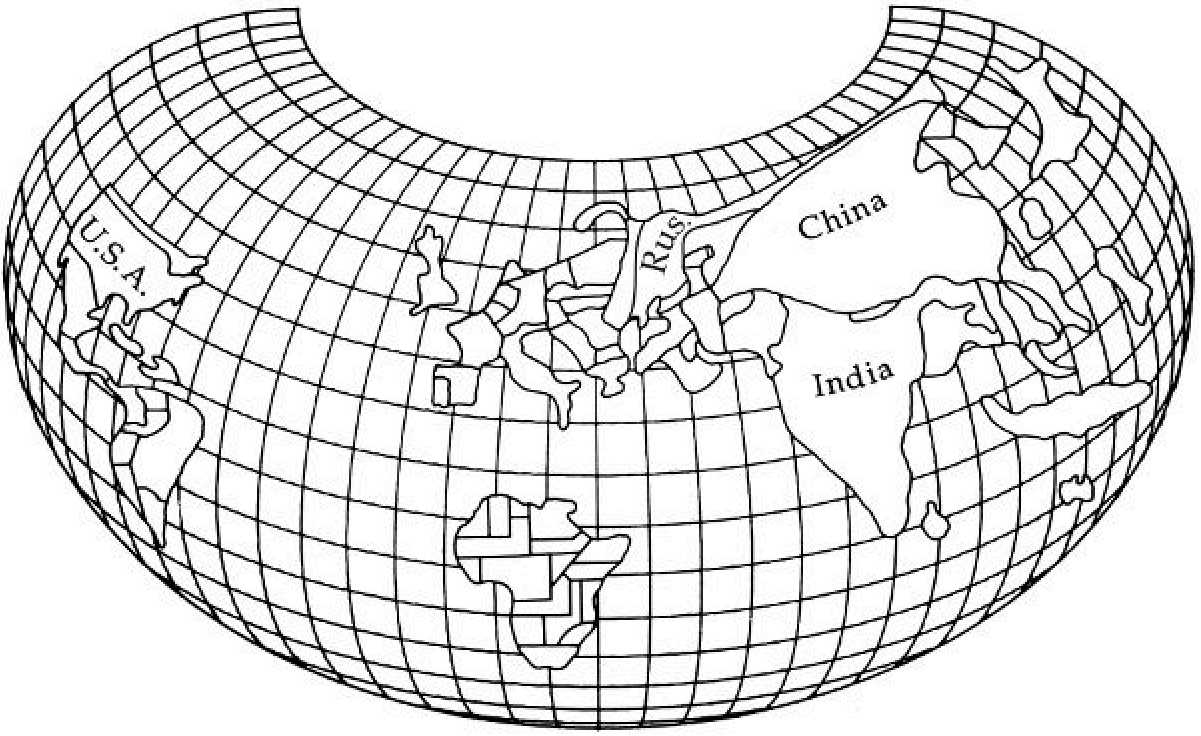


Figure 2.10. "World on a Torus" demographic base map is an area cartogram based on the populations of major countries.

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 maps mass-produced by government mapping agencies and commercial map publishers must address a wide variety of questions, and the maps' 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 higher 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. By contrast, *statistical maps*, which portray numerical data, commonly rely on a single type of symbol, such as dots that each denote ten thousand 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 (see chapter 5). 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 between types of land use 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 seventh game of the World Series. Symbols varying in orientation are useful mostly for representing winds, migration streams, troop movements, and other directional occurrences.

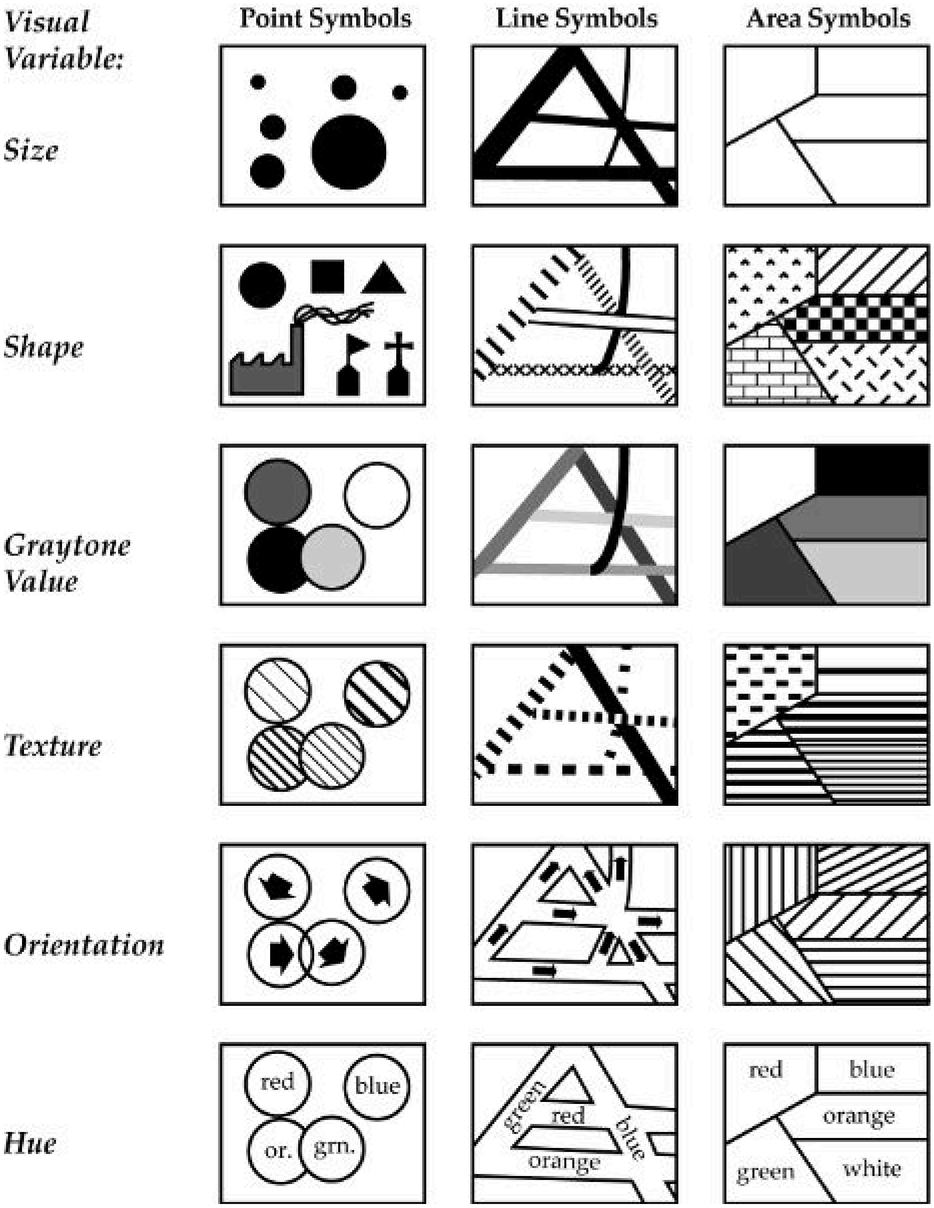
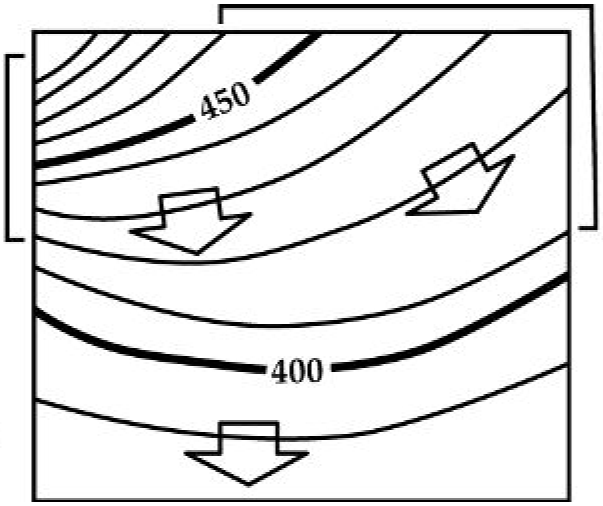


Figure 2.11. The six principal visual variables.



Closely spaced contour lines indicate relatively steep slope.

Widely separated contour lines indicate relatively gentle slope.

Contour lines represent constant elevation.

Contour lines run perpendicular to slope. Arrows show downhill direction.

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

Some visual variables are unsuitable for small point symbols and thin line symbols that provide insufficient contrast with a background. Hue, for instance, is more effective at 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 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 electronic displays and color printers 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 darkness of the symbol shows the concentration of population on the land.

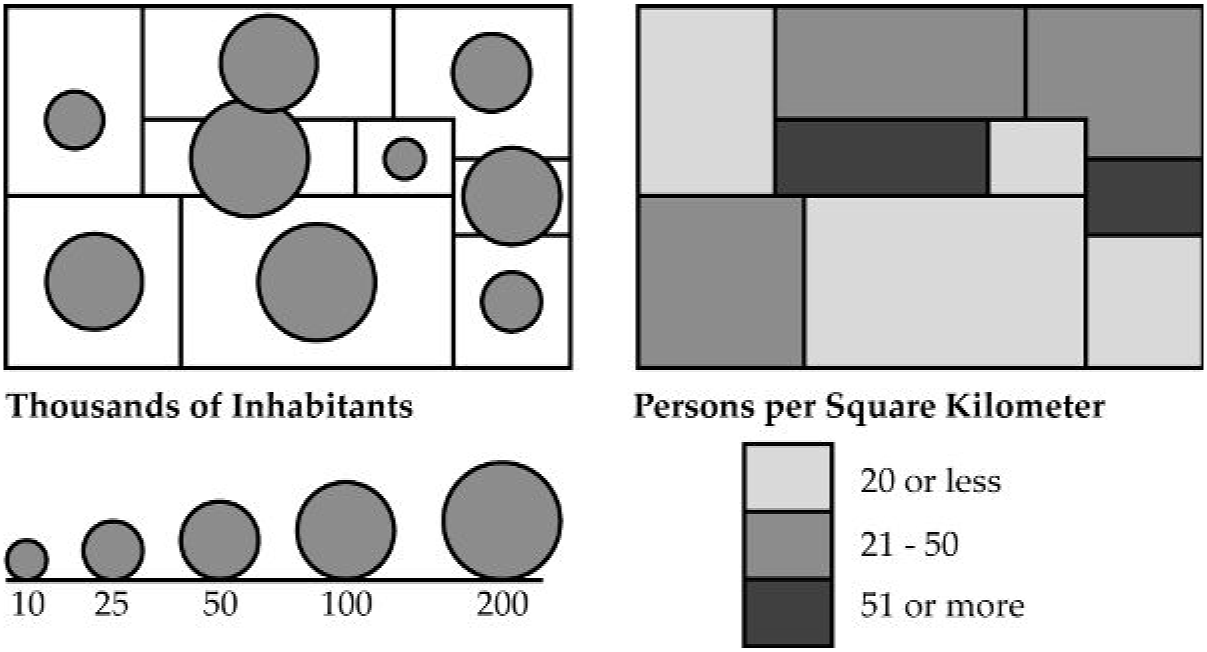


Figure 2.13. Graduated point symbols (left) and graytone area symbols (right) offer straightforward portrayals of population size and population density.

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 (“Union Pacific Railroad”). 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. Similarly, dashed lines might connote uncertainty in the location of a geologic fault line, while mildly transparent area symbols avoid graphic conflict with useful line symbols otherwise omitted.

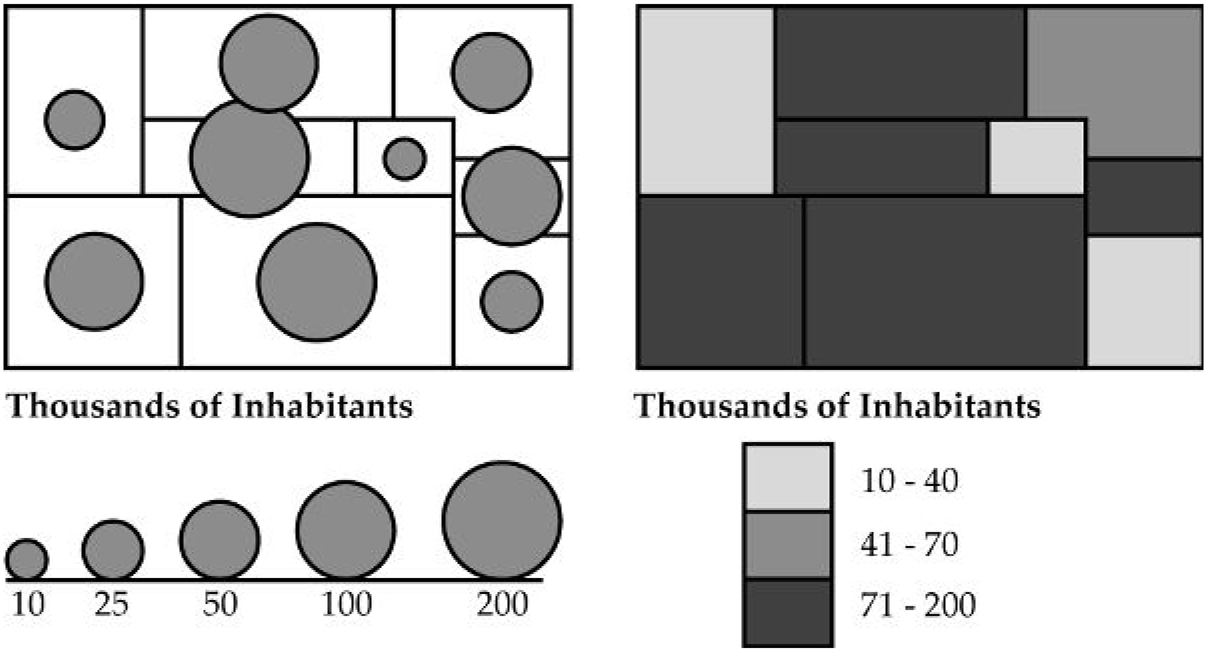


Figure 2.14. The map with graduated point symbols (left) using symbol size to portray magnitude demonstrates an appropriate choice of visual variable. The map with graytone area symbols (right) is ill-suited to portray magnitude.

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 whatsoever 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 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 provide a geographic frame of reference. Selecting background details that are effective in relating new information 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.

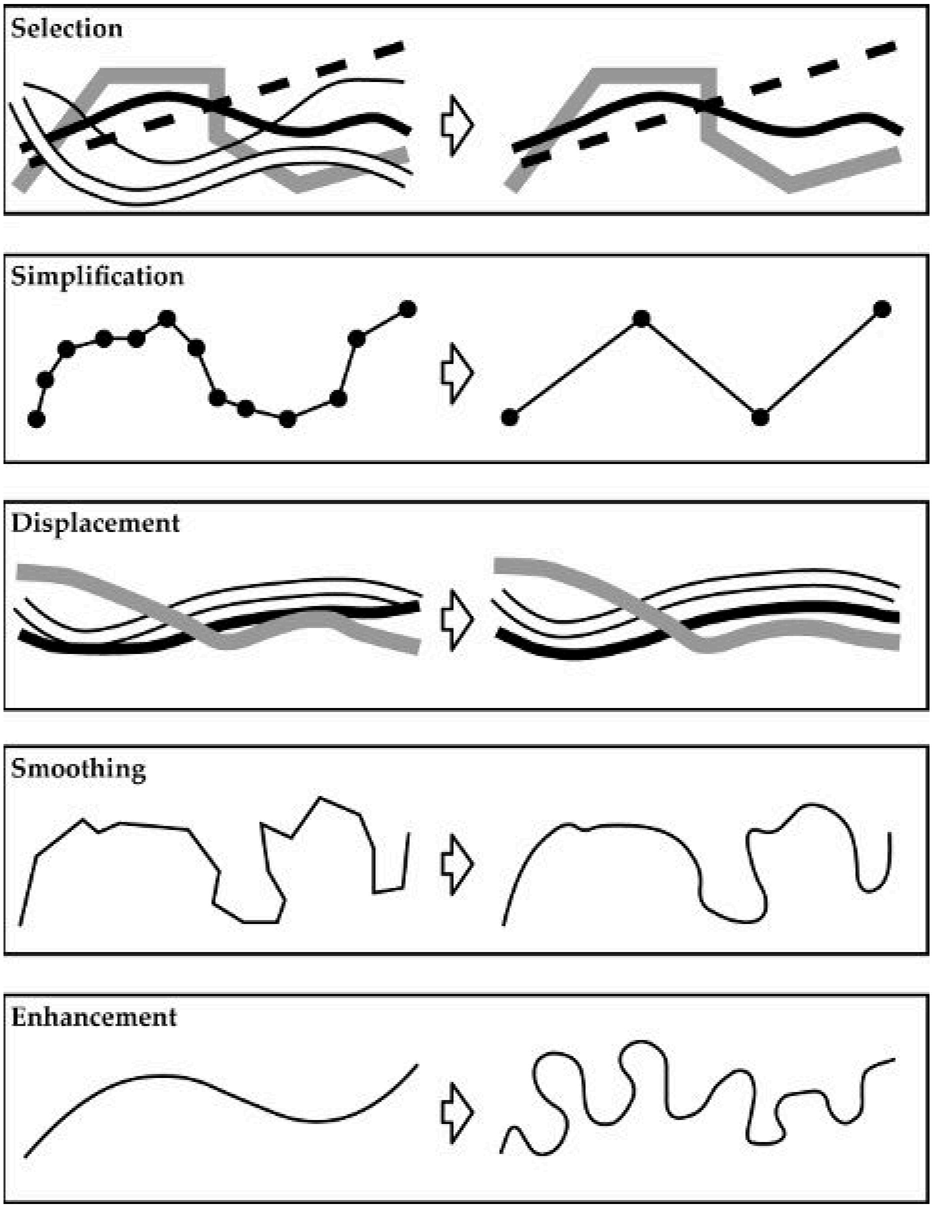


Figure 3.1. Elementary geometric operations in the generalization of line features.

The four remaining generalization processes in figure 3.1 alter the appearance and spatial position of linear map features that are represented by a series of points stored electronically as a list of two-dimensional (X, Y) coordinates. Although the growing use of software to generalize maps has led to the isolation of these four generalization operations, traditional cartographers perform essentially the same operations manually, at a graphics workstation or with pen and ink, but typically do so with less structure, less formal awareness, and less consistency than a computer algorithm. *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 coast-like. Enhanced map symbols are more readily interpreted as well as more aesthetic.

Point features and map labels require a slightly different set of generalization operators. Figure 3.2 illustrates that with point features, 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 *leader 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 accorded separate symbols. In assigning a single symbol to several point features, as when one dot represents twenty reported tornadoes, aggregation usually requires the symbol to either portray the "center of mass" of the individual symbols it replaces or reflect the largest of several discrete clusters.

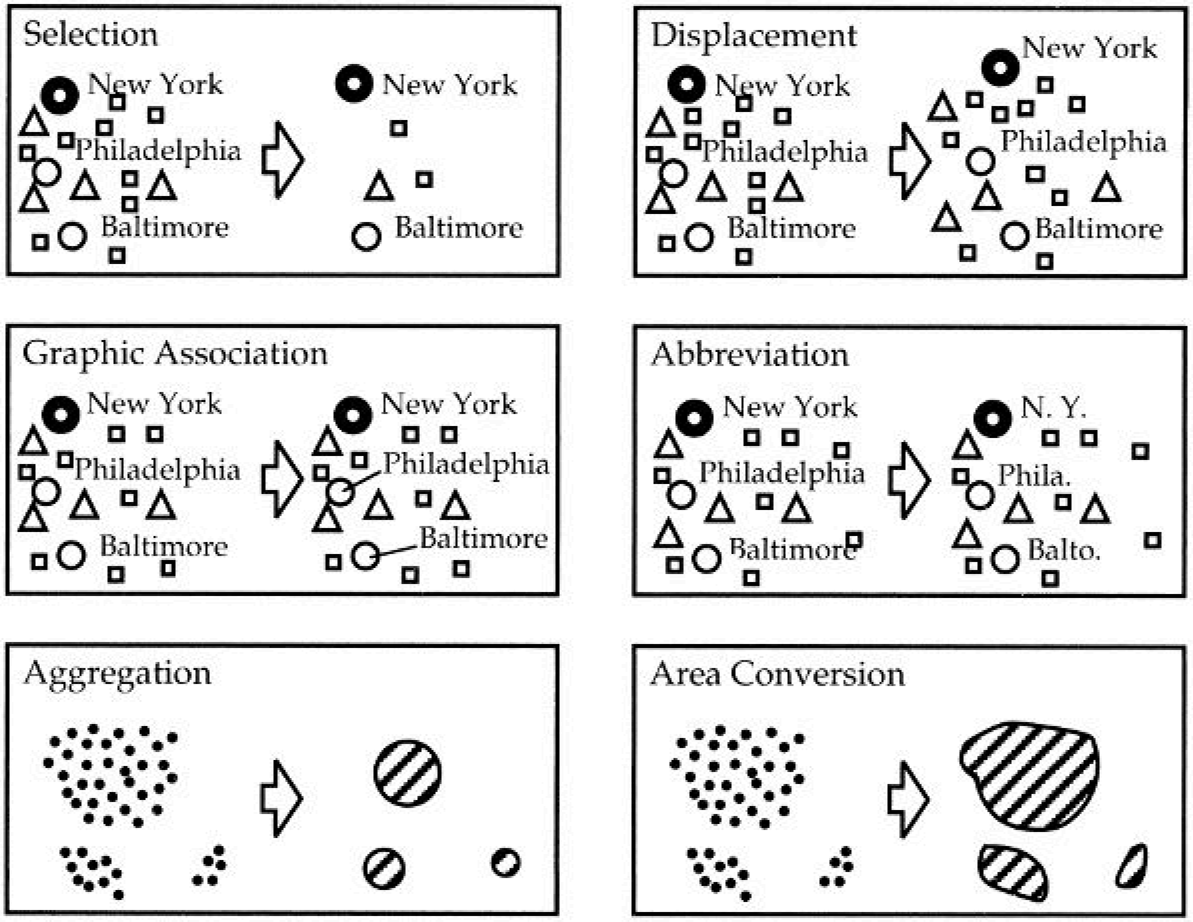


Figure 3.2. Elementary geometric operations in the generalization of point features and map labels.

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.

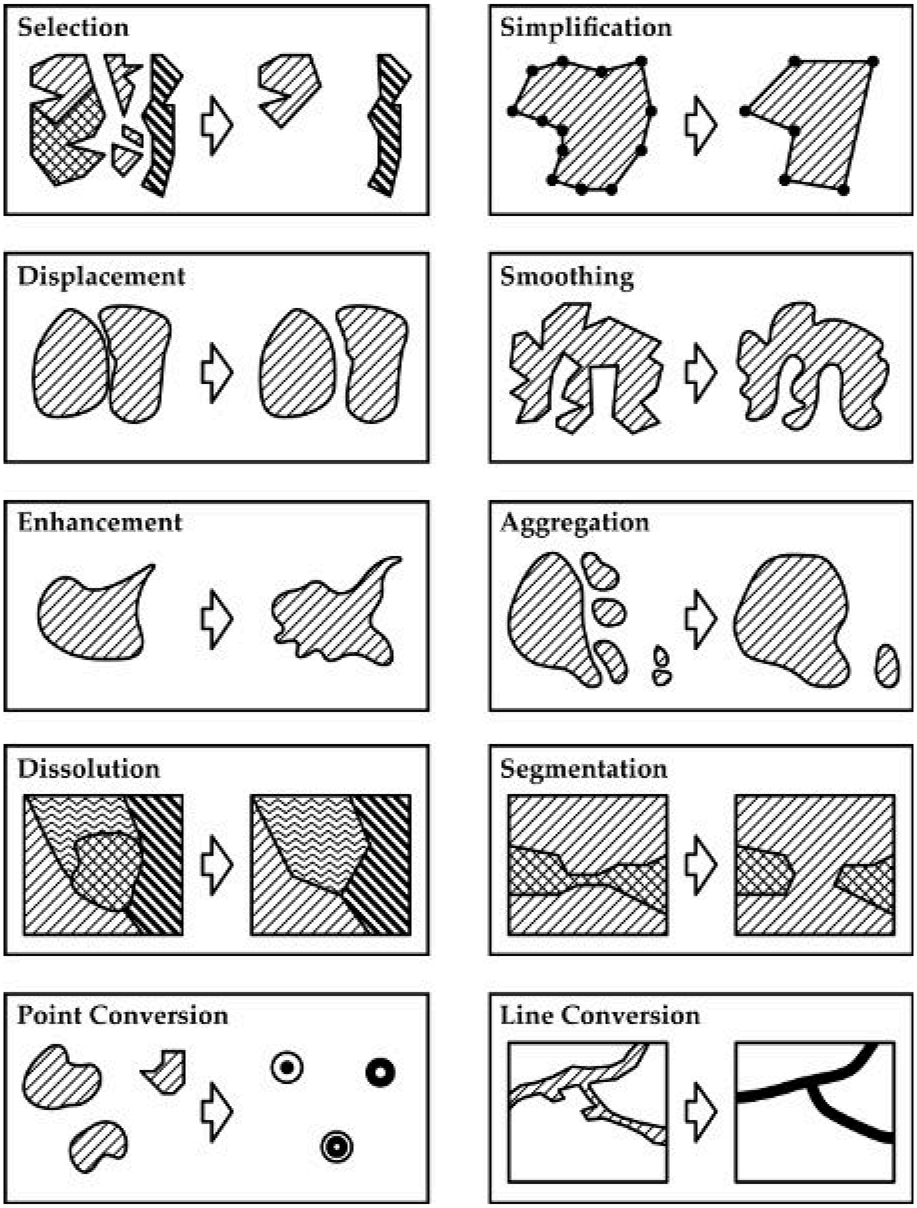


Figure 3.3. Elementary geometric operations in the generalization of area features.

Area features, as figure 3.3 demonstrates, require the largest set of generalization operators because area boundaries are subject to aggregation, 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 they are important as landmarks or shelterbelts.

Soil scientists have used 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 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 and all individual buildings in this area, and the island in the middle of the river. The railroad and the 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.

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

How precisely are symbols positioned on maps? The US Office of Management and Budget addressed this concern with the National Map Accuracy Standards, honored by the United States Geological Survey and other federal mapping agencies. To receive the endorsement that “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, are easily plotted on a map, and can be 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.

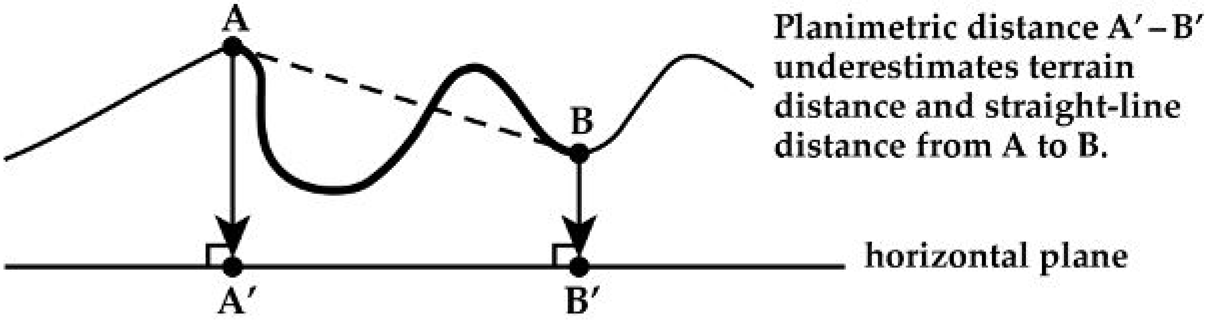


Figure 3.5. Planimetric map generalizes distance by the perpendicular projection of all positions onto a horizontal plane.

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, 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, is greater for points well above the horizontal plane than for lower points, and tends to be greater near the edges than near the center. Cartographers call this effect "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. Chapter 12 examines a wider range of image maps.

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 plate 1, showing the Washington, DC, Metro system, 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. By 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 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.

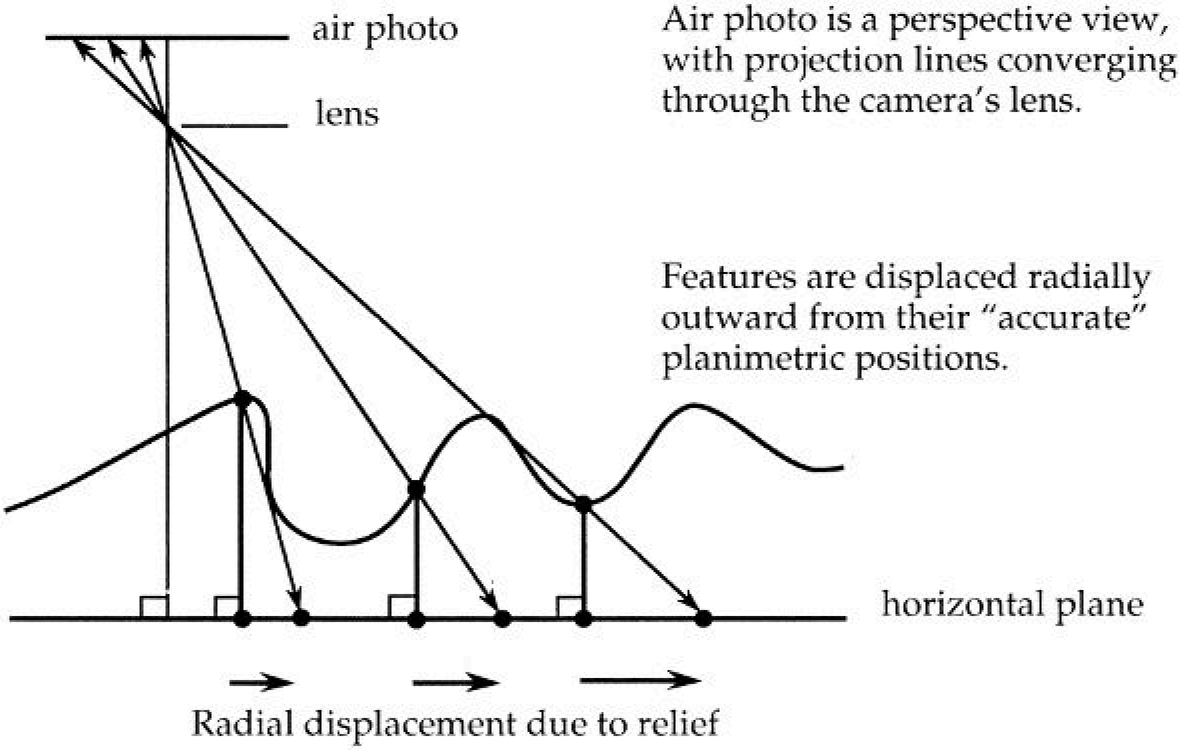


Figure 3.6. A vertical aerial photograph (and any map with symbols traced directly from an air photo) is a perspective view with points displaced radially from their planimetric positions.

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 used plastic drawing templates to trace in ink the outlines of highway shields and other symbols not easily rendered freehand. Drafters would cut area and point symbols from printed sheets and stick them onto the map and apply dashed, dotted, or parallel lines from rolls of specially printed flexible tape. Currently, graphics software allows the mapmaker to not only choose from a menu of standardized point, line, and area symbols provided with the software but also design and store new forms, readily duplicated and added where needed. Consistent symbols also benefit users of the United States 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, at least while examining the map, 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 that is "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 3.7 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 NY 104 cannot easily turn north (toward the top of the map) onto NY 590. The upper-right portion of the left-hand map shows that the necessary connecting lanes from NY 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 NY 104. Yet the large-scale map clearly indicates that a driver expecting an easy connection from NY 104 westbound onto NY 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 separated places. A detailed description is necessary if two people mapping areas 100 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 “inclusions” of soil B are too small to be shown separately and because the soil scientist cannot be aware of all such inclusions. Soils 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 on the field scientist’s understanding of the effects of terrain and geology (if known) on soil development, expertise in selecting sample points, and intuition in plotting boundaries.

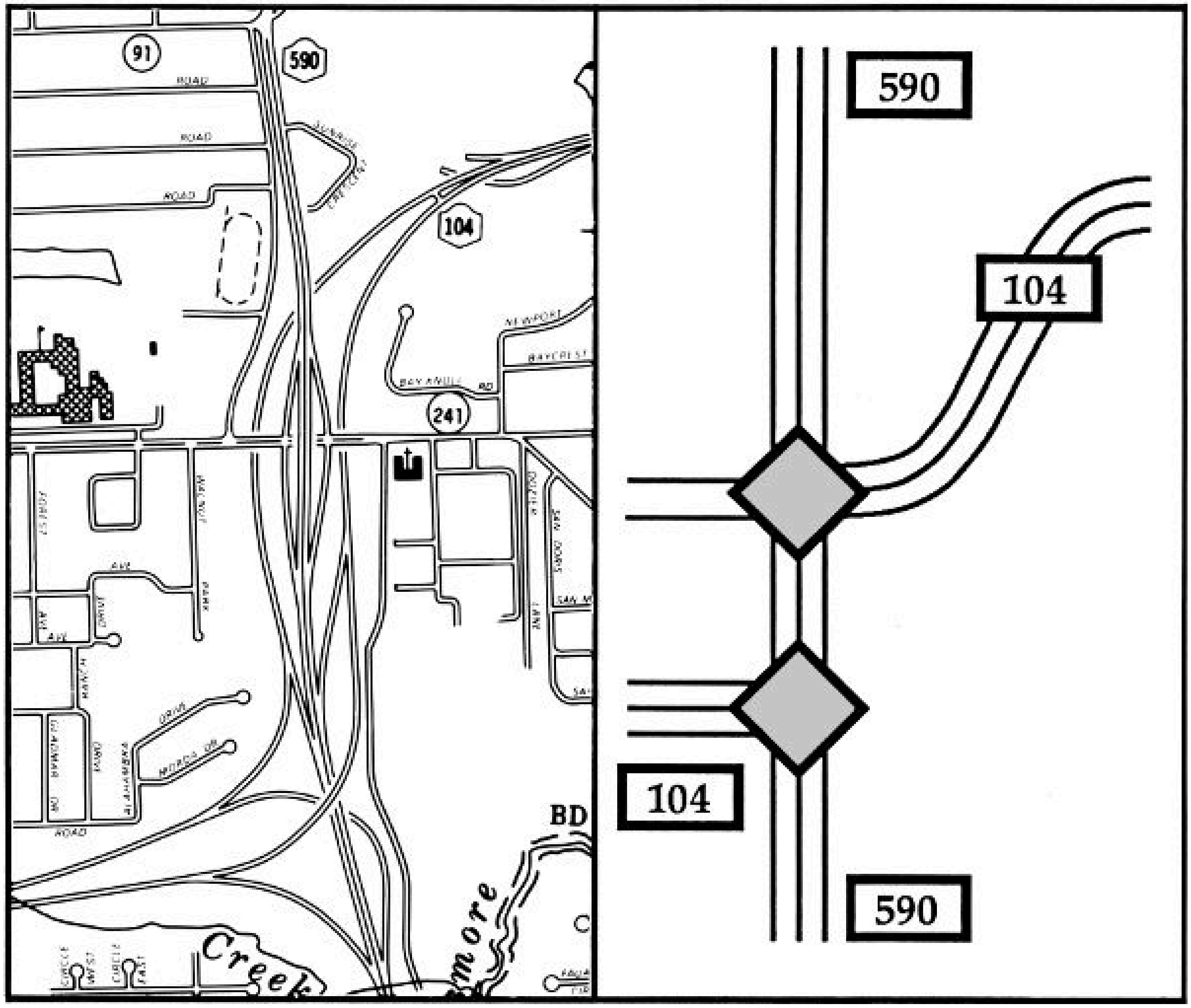


Figure 3.7. Highway interchange near Rochester, New York, as portrayed on a detailed transportation-planning map (left) and on several commercial road maps (right).

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, which are subject to the relief-displacement error shown 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.

Graphics software generally plays a positive role in map analysis and map display, the notion of “garbage in, garbage out” notwithstanding. Particularly useful is the ability of such software to generalize the geometry and content of maps so that one or two geographic databases can support a broad range of display scales. Large-scale maps presenting a detailed portrayal of a small area can exploit the richness of the data, whereas software-generalized smaller-scale displays can 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.

Software-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 3.8 show a rectangular region of approximately 700 square miles (1,800 square kilometers) that includes the city of Harrisburg, Pennsylvania, above and slightly to the right of center. Software 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 3.8. 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 software dissolved water 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.

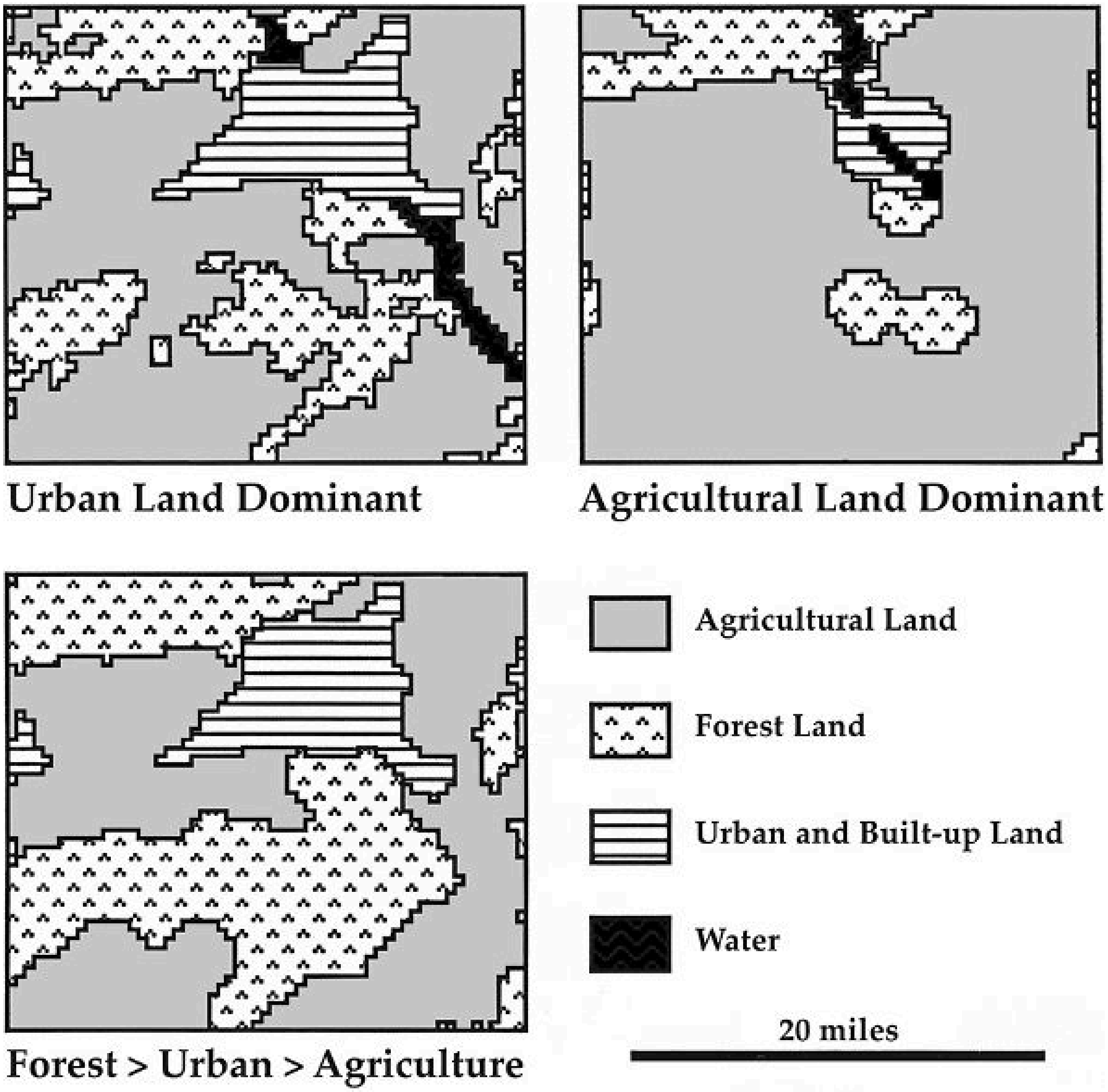


Figure 3.8. Land-use and land-cover maps generalized by computer from more detailed data according to three different sets of display priorities.

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 software-generated maps; manual cartographers have similar goals and biases, however vaguely defined and unevenly applied. Through the consistent application of explicit specifications, the generalization algorithm 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 software generalization should make the cartographer more aware of the multitude 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 sources of bias.

The choropleth map (introduced as the right-hand elements of figures 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 on 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 3.9, for example, offer very different impressions of the spatial pattern of homes still lacking telephones in the northeastern United States in 1960, when minimal economic security demanded a landline. 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 low-income African Americans and to Connecticut’s affluent 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.

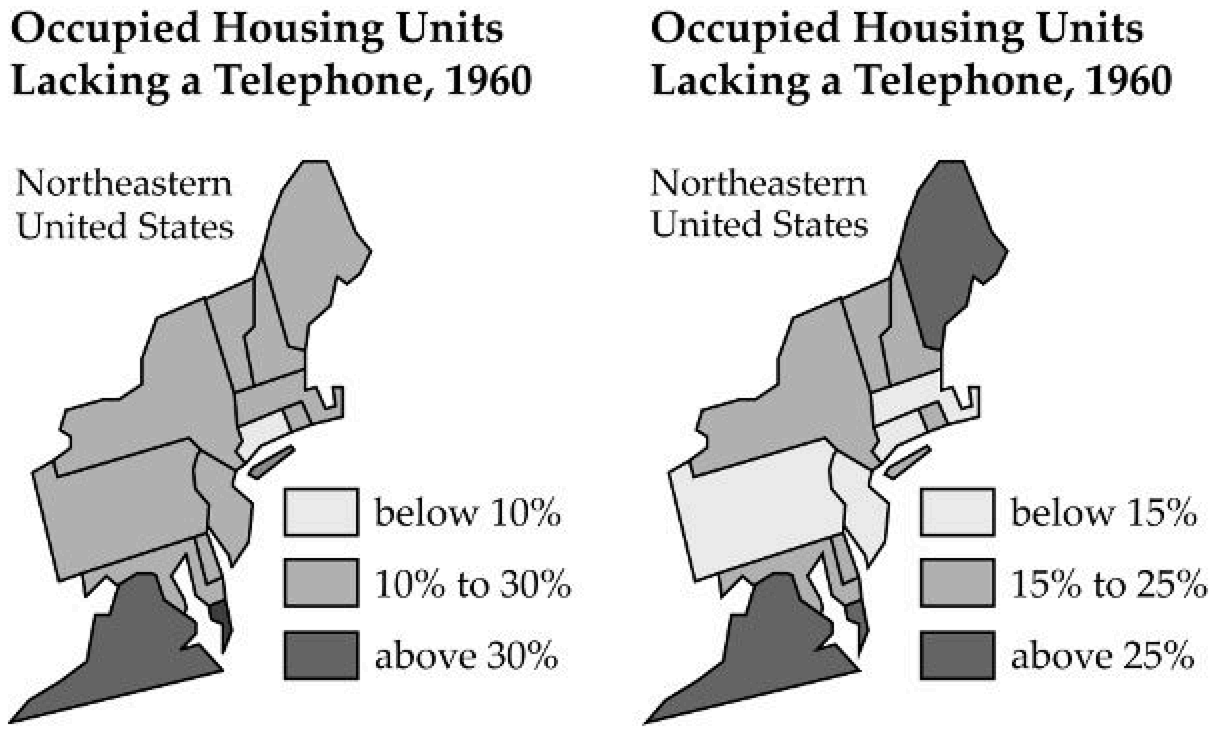


Figure 3.9. Different sets of class breaks applied to the same data yield different-looking choropleth maps.

Machiavellian bias can easily manipulate the message of a choropleth map. Figure 3.10, 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 3.9 and 3.10 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.

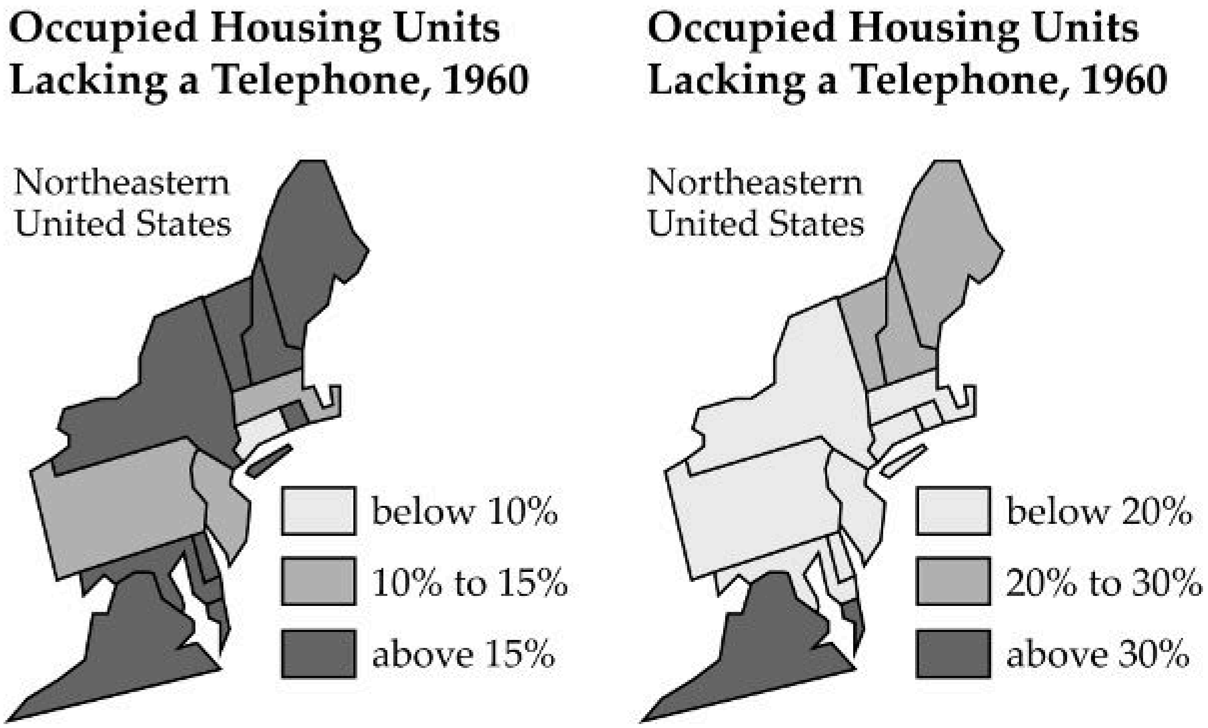


Figure 3.10. Class breaks can be manipulated to yield choropleth maps supporting politically divergent interpretations.

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 also 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.