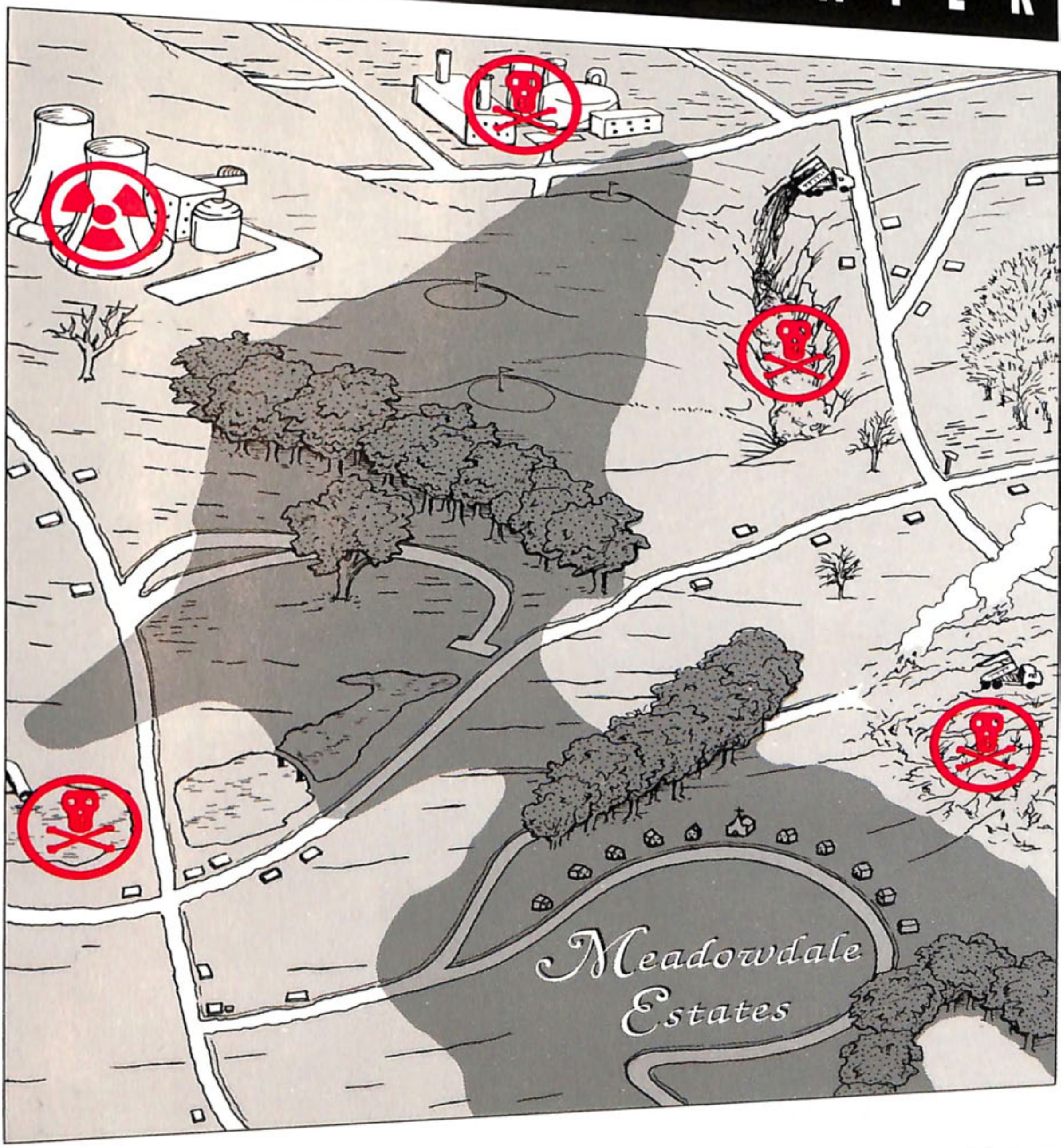
HOW to LIE with MAPS

MARK MONMONIER



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

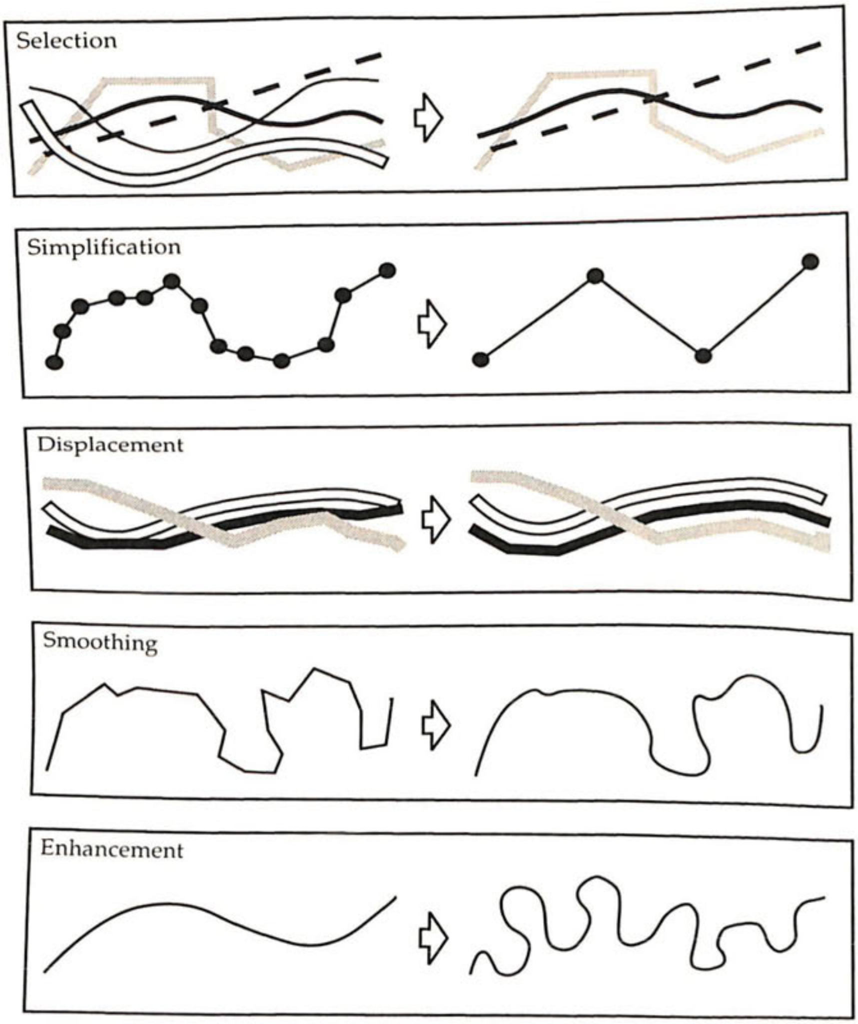


FIGURE 3.1. Elementary geometric operations in the generalization of line features.

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 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.

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-

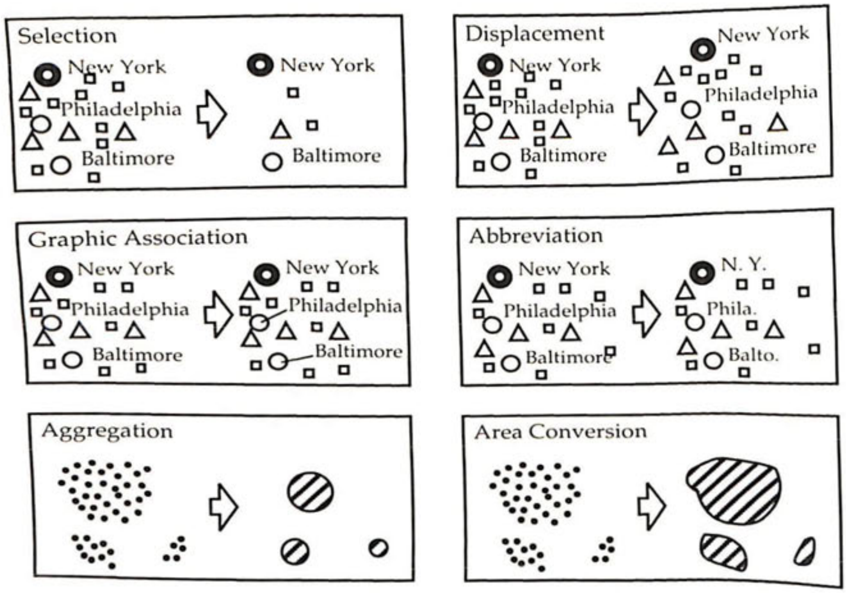


FIGURE 3.2. Elementary geometric operations in the generalization of point features and map labels.

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

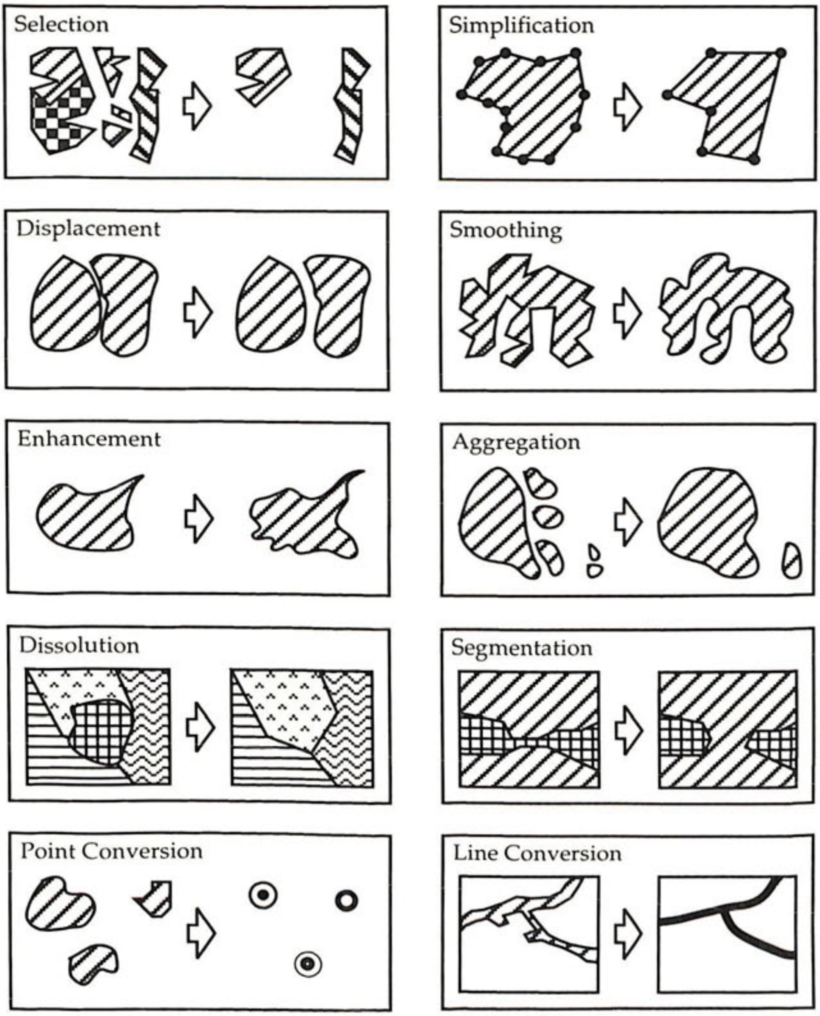


FIGURE 3.3. Elementary geometric operations in the generalization of area features.

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,

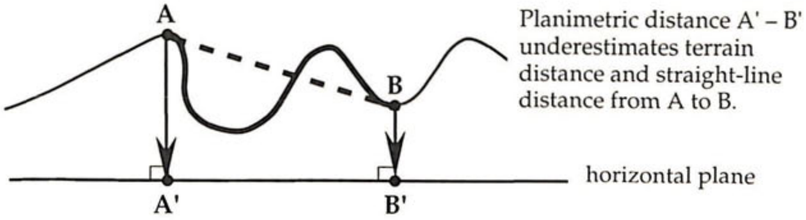


FIGURE 3.5. Planimetric map generalizes distance by the perpendicular projection of all positions onto a horizontal plane.

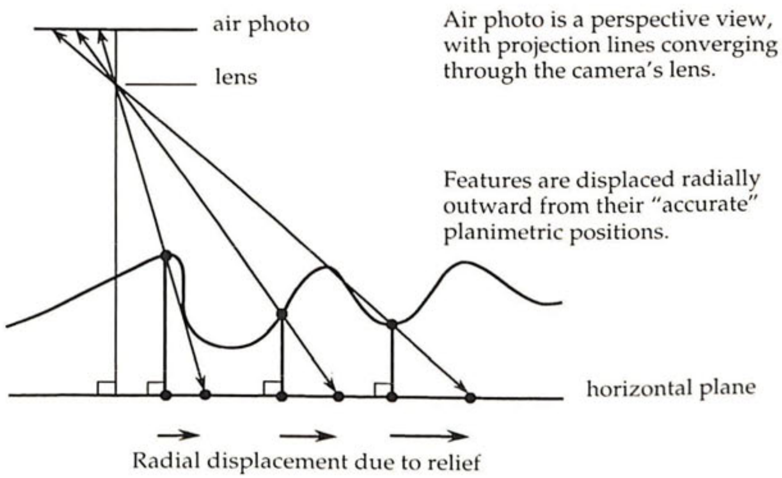


FIGURE 3.6. A vertical aerial photograph (and any map with symbols traced directly from an air photo) is a perspective view with points displaced radially from their planimetric positions.

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

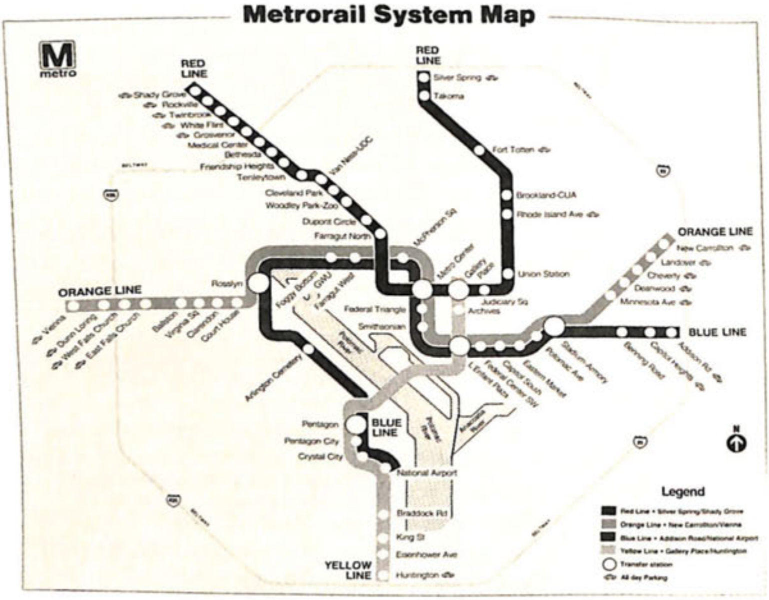


FIGURE 3.7. Linear cartogram of the Washington, D.C., Metro system.

"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

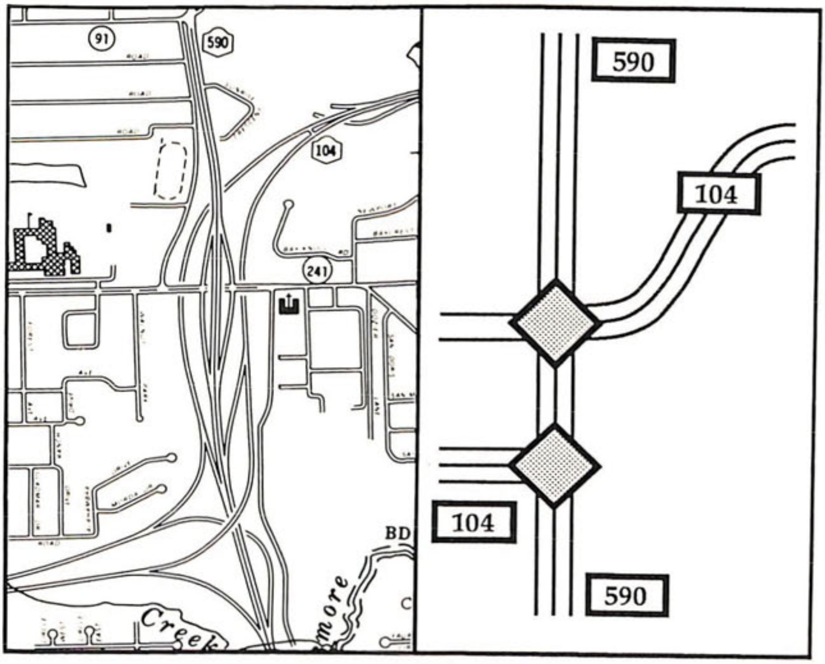


FIGURE 3.8. Highway interchange near Rochester, New York, as portrayed on a detailed transportation planning map (left) and on several commercial road maps (right).

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

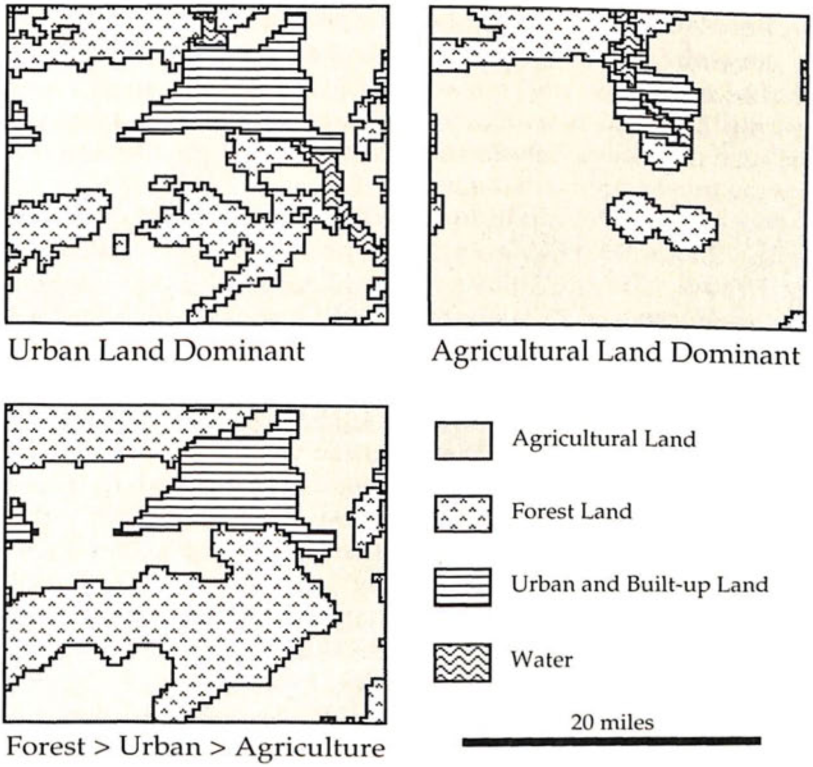


FIGURE 3.9. Land-use and land-cover maps generalized by computer from more detailed data according to three different sets of display priorities.

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-

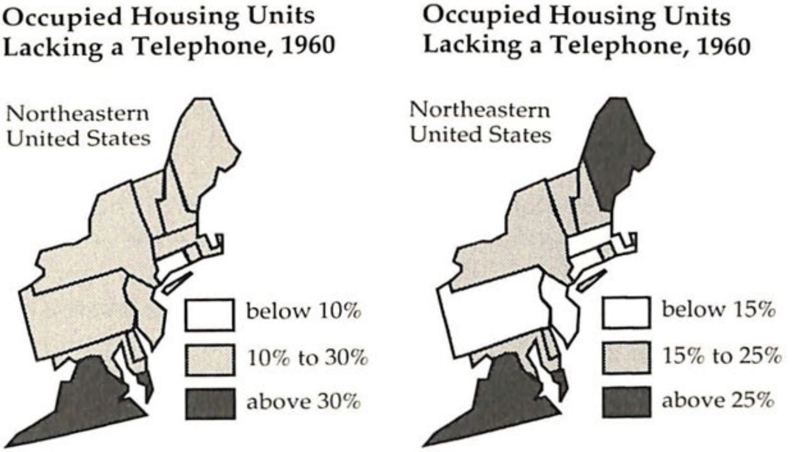


FIGURE 3.10. Different sets of class breaks applied to the same data yield different-looking choropleth maps.

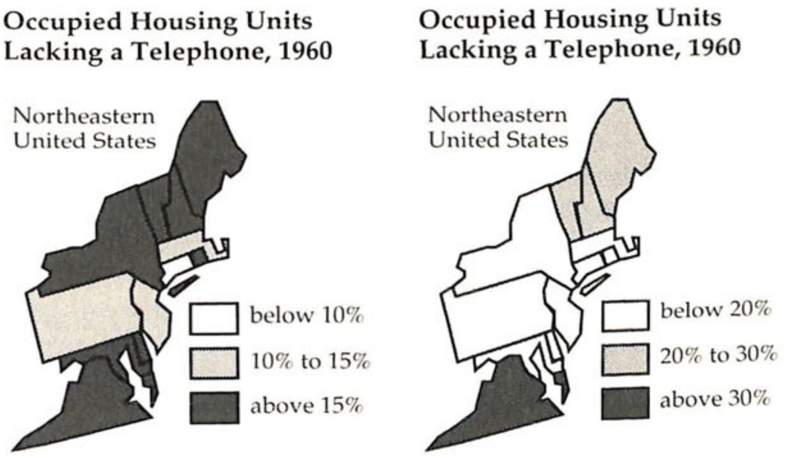


FIGURE 3.11. Class breaks can be manipulated to yield choropleth maps supporting politically divergent interpretations.

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.