Lying with Maps

Mark Monmonier

Abstract. Darrell Huff's *How to Lie with Statistics* was the inspiration for *How to Lie with Maps*, in which the author showed that geometric distortion and graphic generalization of data are unavoidable elements of cartographic representation. New examples of how ill-conceived or deliberately contrived statistical maps can greatly distort geographic reality demonstrate that lying with maps is a special case of lying with statistics. Issues addressed include the effects of map scale on geometry and feature selection, the importance of using a symbolization metaphor appropriate to the data and the power of data classification to either reveal meaningful spatial trends or promote misleading interpretations.

Key words and phrases: Classification, deception, generalization, maps, statistical graphics.

1. INTRODUCTION

I never met Darrell Huff, but his insightful little book *How to Lie with Statistics* was a favorite long before I appropriated the first four words of its title for *How to Lie with Maps*, published in 1991. I don't recall when I first became aware of Huff's book—the oldest of two copies in my library is the 25th printing—but its title was irresistible. Equally intriguing were Huff's straightforward examples, all served up in good humor, of how an unscrupulous or naive statistician could manipulate numbers and graphs to spin a questionable if not downright misleading interpretation of a correlation or time series. In the mid 1980s, when I taught a course titled Information Graphics, *How to Lie with Statistics* provided an engaging supplemental reading.

Huff's approach was as much an inspiration as his title. I already had the kernel of *How to Lie with Maps* in my comparatively obscure *Maps, Distortion, and Meaning*, published in 1977 by the Association of American Geographers as a "Resource Paper" for the Commission on College Geography. Information theory and communication models provided a conceptual framework for an illustrated excursion into the roles of map scale, projection, symbolization, and classification

Mark Monmonier is Distinguished Professor of Geography, Maxwell School of Citizenship and Public Affairs, Syracuse University, Syracuse, New York 13244-1020, USA (e-mail: mon2ier@maxwell.syr. edu).

in cartographic generalizations of geographic data—hardly light material. Written with upper-division college students in mind, *Maps, Distortion, and Meaning* supplemented its 51 letter-size pages of academic prose and real-world examples with a bibliography listing 92 books and articles. By contrast, the first edition of *How to Lie with Maps* gleefully indulged in contrived Huffian examples and blithely ignored the scholarly record—a deficiency rectified five years later when the University of Chicago Press commissioned an expanded edition that added 72 relevant references, chapters on multimedia and national mapping programs, and four pages of color illustrations.

Huff's footsteps offered an easy trek through the forest of popular academic publishing. In addition to providing the conceptual model for an exposé of representational sleight of hand, *How to Lie with Statistics* attracted the benevolent eye of reviewers like John Swan (1992), who situated my book "in the fine tradition of Darrell Huff's *How to Lie with Statistics*," and Scott Kruse (1992), who opined that "what Huff did for statistics, Monmonier has done for cartography." Quoting favorable reviews might sound boorishly vain, but these excerpts demonstrate that Huff's book was not only well-known but an exemplar worth imitating.

Lying with maps is, of course, a lot different from lying with statistics. Most maps are massive reductions of the reality they represent, and clarity demands that much of that reality be suppressed. The mapmaker who tries to tell the whole truth in a single map typically produces a confusing display, especially if the

area is large and the phenomenon at least moderately complex. Map users understand this and trust the map-maker to select relevant facts and highlight what's important, even if the map must grossly distort the earth's geometry as well as lump together dissimilar features. When combined with the public's naive acceptance of maps as objective representations, cartographic generalization becomes an open invitation to both deliberate and unintentional prevarication.

At the risk of stretching the notion of lying, I'm convinced that inadvertent fabrication is far more common these days than intentional deceit. Moreover, because most maps now are customized, one-of-a-kind graphics that never make it into print or onto the Internet, prevaricating mapmakers often lie more to themselves than to an audience. Blame technology—a conspiracy between user-friendly mapping software (or not-so-user-friendly geographic information systems) and high-resolution laser printers that can render crisp type and convincing symbols with little effort or thought. There's a warning here I'm sure Darrell Huff would applaud: watch out for the well-intended mapmaker who doesn't understand cartographic principles yet blindly trusts the equally naive software developer determined to give the buyer an immediate success experience—default settings are some of the worst offenders. Because lying with maps is so easy in our information-rich world, infrequent mapmakers need to understand the pitfalls of map generalization and map readers need to become informed skeptics.

As this essay suggests, maps can lie in diverse ways. Among the topics discussed here are the effects of map scale on geometry and feature selection, the importance of using a symbolization metaphor appropriate to the data and the power of data classification to reveal meaningful spatial trends or promote misleading interpretations.

2. SELECTIVE TRUTH

An understanding of how maps distort reality requires an appreciation of *scale*, defined simply as the ratio of map distance to ground distance. For example, a map at 1:24,000, the scale of the U.S. Geological Survey's most detailed topographic maps, uses a one-inch line to represent a road or stream 24,000 inches (or 2,000 feet) long. Ratio scales are often reported as fractions, which account for distinctions between "large-scale" and "small-scale." Thus a quadrangle map showing a small portion of a county at 1/24,000 is very much a large-scale map when compared, for instance, to an atlas map showing the whole world at 1/75,000,000—a markedly smaller fraction. (Planners and engineers sometimes confuse scale and *geographic scope*, the size of the area represented. It might seem counterintuitive that small-scale maps can cover vast regions while large-scale maps are much more narrowly focused, but when the issue is scale, not scope, "large" means comparatively detailed whereas "small" means highly generalized.)

Mapmakers can report a map's scale as a ratio or fraction, state it verbally using specific distance units—"one inch represents two miles" is more user friendly than 1:126,720—or provide a scale bar illustrating one or more representative distances. Bar scales, also called graphic scales, are ideal for large-scale maps because they promote direct estimates of distance, without requiring the user to locate or envision a ruler. What's more, a graphic scale remains true when you use a photocopier to compress a larger map onto letter-size paper. Not so with ratio or verbal scales.

However helpful they might be on large-scale maps, bar scales should never appear on maps of the world, a continent, or a large country, all of which are drastically distorted in some fashion when coastlines and other features are transferred from a spherical earth to a flat map. Because of the stretching and compression involved in flattening the globe, the distance represented by a one-inch line can vary enormously across a world map, and scale can fluctuate significantly along, say, a six-inch line. Because map scale varies not only from point to point but also with direction, a bar scale on a small-scale map invites grossly inaccurate estimates. Fortunately for hikers and city planners, earth curvature is not problematic for the small areas shown on large-scale maps; use an appropriate map projection, and scale distortion is negligible.

What's not negligible on most large-scale maps is the generalization required when map symbols with a finite width represent political boundaries, streams, streets and railroads. Legibility requires line symbols not much thinner than 0.02 inch. At 1:24,000, for instance, a 1/50-inch line represents a corridor 40 feet wide, appreciably broader than the average residential street, rural road or single-track railway but usually not troublesome if the mapmaker foregoes a detailed treatment of driveways, property lines, rivulets and rail yards. At 1:100,000 and 1:250,000, which cartographers typically consider "intermediate" scales, symbolic corridors 166.7 and 416.7 feet wide, respectively, make graphic congestion ever more likely unless the mapmaker weeds out less significant features, simplifies complex curves and displaces otherwise overlapping symbols.
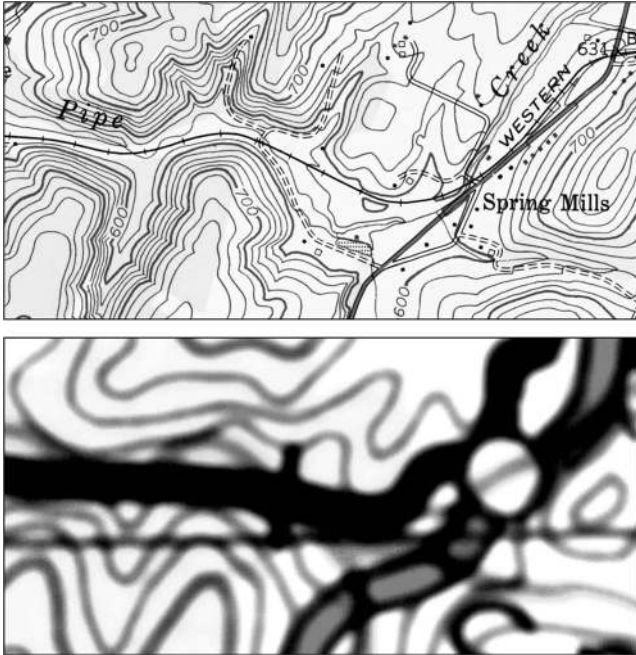
FIG. 1. *Juxtaposition of map excerpts at* 1:24,000 *(above) and* 1:250,000, *enlarged to* 1:24,000 *(below), illustrate some of the effects of scale on cartographic generalization. Both images show the same area, in and around Spring Mills, Maryland.*

FIG. 2. *Crude birth rates, 2000, by state, based on equal-intervals cut-points and plotted on a visibility base map.*

Figure 1 illustrates the effect of cartographic generalization on the U.S. Geological Survey's treatment of Spring Mills, Maryland (south of Westminster) at scales of 1:24,000 and 1:250,000. Both excerpts cover the same area, but the upper panel is a same-size black-and-white excerpt from the larger-scale, 1:24,000 map, whereas the lower panel shows the corresponding portion of the 1:250,000 map enlarged to 1:24,000 to reveal the impact of noticeably wider symbolic corridors. At the smaller scale the hamlet of Spring Mills becomes an open circle, rather than a cluster of buildings, and the railroad and main highway are moved apart for clarity. Mapmakers compiling intermediate-scale maps typically select features from existing large-scale maps. When the difference between scales is substantial, as it is here, few features survive the cut, and those that do are usually smoothed or displaced.

"White lies" like these are unavoidable if maps are to tell the truth without burying it in meaningless details. In a similar vein mapmakers use tiny picnic-bench symbols to locate public parks and small, highly simplified single-engine airplanes to represent airports. These icons work because they're readily decoded, even without a map key. Legends and labels also help, especially for small-scale reference maps, on which mere points or circles substitute for complex city boundaries.
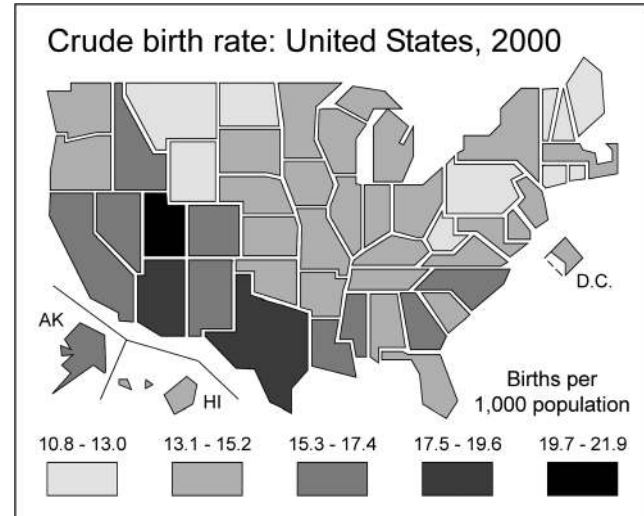
A geometric distortion especially useful in portraying statistical data for the United States is the "visibility base map" (Figure 2), which replaces the contorted outlines of Maine and Massachusetts with simplified five- and thirteen-point polygons, instantly recognizable because of their relative location and characteristic shape. Although simplified polygons can lighten the computational burden of real-time cartographic animation, the prime goal is to help viewers of small, column-width choropleth maps see and decode the otherwise obscure area symbols representing rates or other statistics for small states like Delaware and Rhode Island. (*Choropleth map* is the cartographic term for a map based on established areal units, like states or census tracts, grouped into categories, each represented by a specific color or graytone.) While purists might object to the visibility map's caricatured shapes and grossly generalized coastlines, this type of simplification is no more outrageous than summarizing a spatially complex entity like California or New York with a statewide average.

3. CUT-POINTS AND FRIVOLOUS FILLS

Statistical data like the spatial series of birth rates in Figures 2 and 3 are easily distorted when mapmakers succumb to a software vendor's sense of what works without probing the data to discover what's meaningful. Whenever mapping software serves up an instant, no-thought, default classification for a choropleth map, the usual result is five categories based on either equal-intervals or quantile classing. The method of grouping is almost always more problematic than the number of
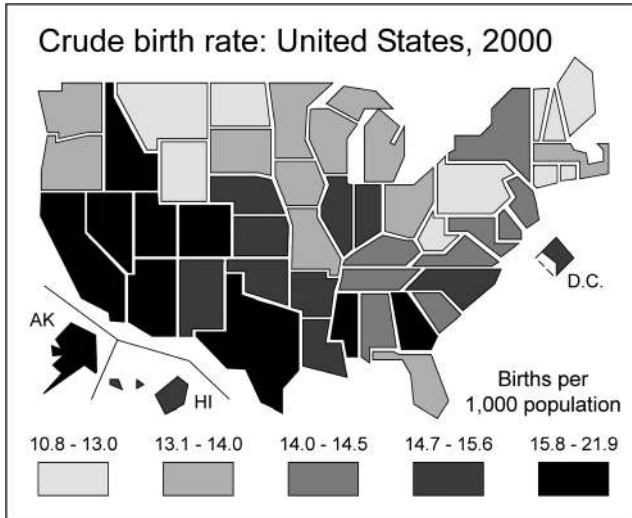
FIG. 3. *Crude birth rates, 2000, by state, based on quantile cut-points and plotted on a visibility base map.*

groups: unless the data contain fewer or only slightly more highly distinct clusters, five categories seems a reasonable compromise between a less informative two-, three- or four-category map and a comparatively busy map on which six or more symbols are less easily differentiated. Equal-intervals cut-points, computed by dividing the full range of data values into intervals of equal length, are computationally simpler than quantiles, which requires sorting the data and apportioning an equal number of places to each category. One must also make adjustments to avoid placing identical values in different categories. Because of these adjustments, Figure's 3 categories vary in size from 9 to 11.

Figures 2 and 3 offer distinctly different portraits of crude birth rates in the United States for the millennial year. My hunch is that the equal-intervals display (Figure 2), which recognizes well-above-average birth rates in Utah (21.9) and Texas (17.8), comes closer to getting it right than the quantile map (Figure 3), which lumps together states with rates between 15.8 and 21.9. Even so, viewers of the latter display might appreciate categories based on commonsense notions like lowest fifth and highest fifth.

If the map author is at all concerned with full disclosure, a number line (univariate scatterplot-histogram)
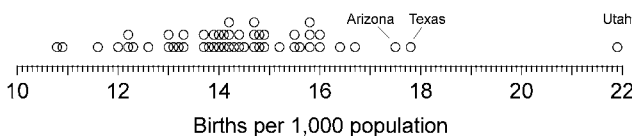

FIG. 4. *Number line describes variation in the data for Figures 2 and 3.*

like Figure 4 is a must. This simple graphic quickly reveals pitfalls like the possible assignment of Arizona and Texas (17.5 and 17.8, resp.) to separate categories. Mapmakers who plot a number line are less likely to miss potentially significant groupings of data values, but there's no guarantee that the data will form distinct categories neatly separated by readily apparent "natural breaks." Although algorithmic strategies for finding natural breaks have been around for over three decades (Jenks and Caspall, 1971), classifications that minimize within-group variance are not necessarily revealing. Even so, a programmed natural-breaks solution is arguably better than a quantile scheme certain to ignore Utah's exceptionally high birth rate or an equal-interval solution that might separate close outliers like Texas and Arizona.

Optimization algorithms and standardized schemes like equal-intervals and quantiles are prone to miss cut-points like the national average, which helps viewers compare individual states to the country as a whole. And for maps describing rates of change, programmed solutions readily overlook the intuitively obvious cut-point at zero, which separates gains from losses.

Although the large number of potentially meaningful cut-points precludes their use in a printed article or in an atlas intended for a national audience, a dynamic map included with exploratory data analysis software or available over the Internet could let users manipulate cut-points interactively. A software vendor interested in informed analysis as well as openness would, I hope, supplement moveable cut-points with a number line so that viewers could readily recognize outliers and clumpiness in the data as well as appreciate the value of looking at and presenting more than one map.

The ability to explore data interactively can be an invitation to buttress specious arguments with biased maps. For example, a polemicist out to demonstrate that American fertility is dangerously low might devise a map like Figure 5, which assigns nearly three-quarters of the states to its lowest category. Similarly, a demagogue arguing that birth rates are too high would no doubt prefer Figure 6, which paints much of the country an ominous black. Extreme views like these are useful reminders that maps are readily manipulated.

Another hazard of mapping software is the ease with which naive users can create convincing choropleth maps with "count" variables like resident population or number of births. Although Figure 7 might look convincing, close inspection reveals nothing more than a
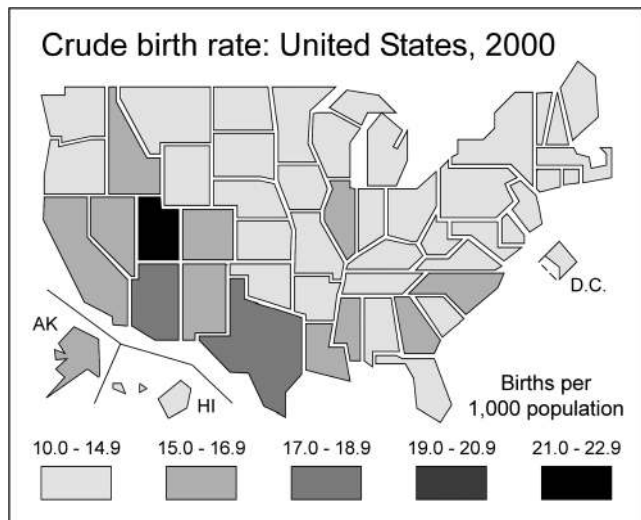
FIG. 5. *Crude birth rates, 2000, by state, categorized to suggest dangerously low rates overall.*

FIG. 7. *The darker-is-more-intense metaphor of choropleth maps offers a potentially misleading view of numbers of births.*

pale shadow of population—states with more people, not surprisingly, register more births, whereas those with the smallest populations are in the lowest category. If you want to explore geographic differences in fertility, it's far more sensible to look at birth rates as well as the total fertility index and other more sensitive fertility measures used in demography (Srinivasan, 1998). A map focusing on number of births, rather than a rate, has little meaning outside an education or marketing campaign pitched at obstetricians, new mothers or toy manufacturers.

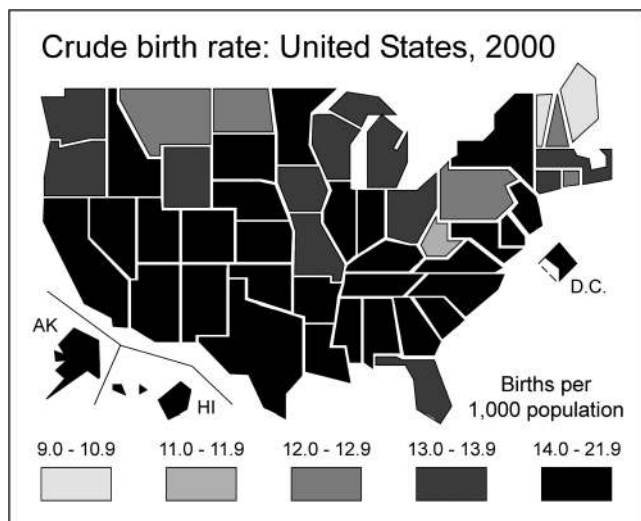Whenever a map of count data makes sense, perhaps to place a map of rates in perspective, graphic

FIG. 6. *Crude birth rates, 2000, by state, categorized to suggest dangerously high rates overall.*
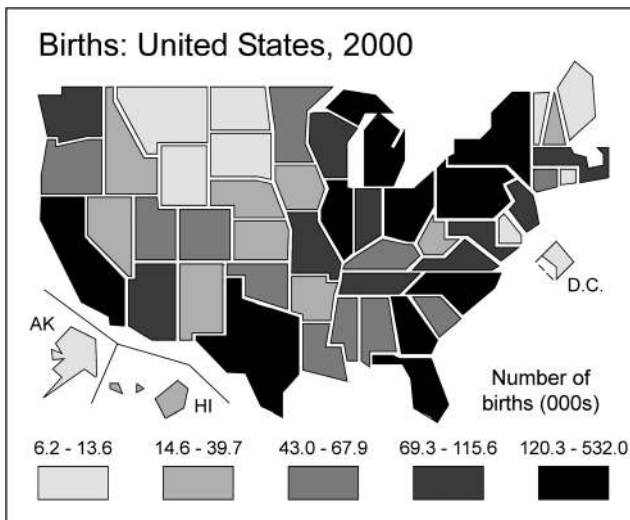
theory condemns using a choropleth map because its ink (or toner) metaphor is misleading. Graytone area symbols, whereby darker suggests "denser" or "more intense" while lighter implies "more dispersed" or "less intense," are wholly inappropriate for count data, which are much better served by symbols that vary in size to portray differences in magnitude (Bertin, 1983). In other words, while rate data mesh nicely with the choropleth map's darker-means-more rule, count data require bigger-means-more coding.

Although college courses on map design emphasize this fundamental distinction between intensity data and count (magnitude) data, developers of geographic information systems and other mapping software show little interest in preventing misuse of their products. No warning pops up when a user asks for a choropleth map of count data, training manuals invoke choropleth maps of count data to illustrate commands and settings, and alternative symbols like squares or circles that vary with magnitude are either absent or awkwardly implemented. One developer—I won't name names—not only requires users to digitize center points of states but also scales the graduated symbols by height rather than area, a fallacious strategy famously ridiculed by Huff's pair of caricatured blast furnaces, scaled by height to compare steel capacity added during the 1930s and 1940s (Huff, 1954, page 71). Map viewers see these differences in height, but differences in area are more prominent if not overwhelming.

Several remedies are indicated: improved software manuals, more savvy users, metadata (data about data) that can alert the software to incompatible symbols
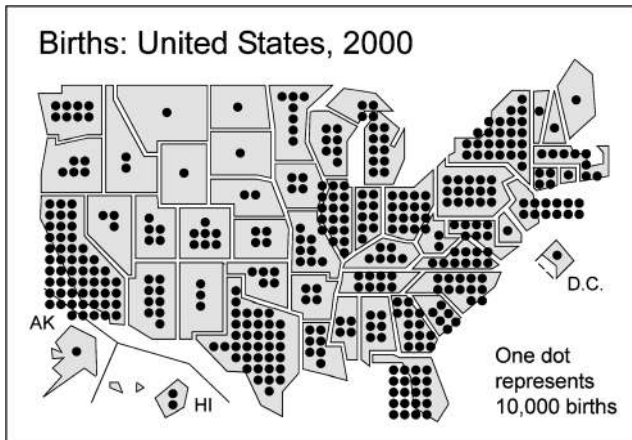
FIG. 8.   *The bigger-means-more metaphor of this dot-array map affords a more appropriate treatment of the count data in Figure* 7.

and sophisticated display algorithms that automate dot-array symbols like those in Figure 8. I like the dot array because a state's dots are not only countable but collectively constitute a magnitude symbol that visually sorts out promising and poor locations for a diaper factory. Although dot arrays are easily constructed with illustration software like Adobe Illustrator and Macromedia Freehand, describing the process in C++ would be a daunting undertaking if the programmer had to include quirky local solutions like rotating the dot array to fit South Carolina or extending it into the ocean to accommodate New Jersey.

Equally reckless is the software industry's insistence in promoting choropleth maps with widely varied hues. Although a spectral sequence from blue up through green, yellow, orange and red might make sense to fans of the *USA Today* weather chart, color maps that lack a temperature chart's emotive hues and conveniently nested bands can be difficult to decode. While software developers and map authors might argue that all the information needed to read a multi-hue map is right there, in the legend, forcing the conscientious user to look back and forth between map and key is hardly as helpful as relying on the straightforward darker-is-more metaphor. Color can be a thicket for map authors, and because color artwork is not an option for this essay, I won't go into it here aside from noting that color is convenient for maps on which a second visual variable portrays reliability (MacEachren, Brewer and Pickle, 1998)—the cartographic equivalent of error bars.

4. MAPS AND BIVARIATE CORRELATION

Just as cut-points can be manipulated to suggest that birth rates are dangerously low or high overall, pairs

of choropleth maps can purposely heighten or suppress perceptions of bivariate association. Figure 9 offers a telling example. The map at the top describes state-level rates of population change between the 1960 and 1970 census enumerations, and the two lower maps show rates of net-migration over the same period. I call the upper map the referent because the data for the two lower maps were categorized to maximize and minimize visual similarity with this particular five-class categorization (Monmonier, 1977, pages 32–33).

This three-map display originated with a comparatively innocent attempt to find cut-points that enhance the visual similarity of two maps. An iterative algorithm generated a large number of trial maps for the classed variable, evaluated each map's assumed visual similarity to the referent and saved the cut-points if the new trail was more similar than the previous best match (Monmonier, 1976). My assumption that area alone, rather than shape or location, affects a state's contribution to visual similarity is admittedly simplistic, but it seems reasonable that a pair of maps with matching graytones for Texas will look more similar on average than a pair of maps with matching graytones for Rhode Island. Although trial-and-error optimization might unreasonably inflate the visual similarity of two weakly or mildly associated variables, I chose as my classed variable the net-migration rate for the 1960s, which has a logical, highly positive ($r = 0.93$) relationship with population change, insofar as states with net losses or low rates of increase were plagued by net out-migration, while those that surged forward did so largely because many more people moved in than moved out. The result was the map at the lower left, which looks a great deal like the referent at the top.

Since I developed *Maps, Distortion, and Meaning* shortly after describing the process in an article titled "Modifying objective functions and constraints for maximizing visual correspondence of choroplethic maps," it's not surprising that this coincidence inspired a wicked thought: Why not *minimize* correspondence visually by saving the cut-points with the *worst* assumed similarity? Altering a few lines of computer code yielded the map at the lower right, which looks most unlike the referent, largely because three of its five categories have only one member while a vast category ranging from −11.82 to 50.48 captures a lion's share of the states. Word to the wary: if you see a choropleth map with one huge category and several very small ones, be suspicious.
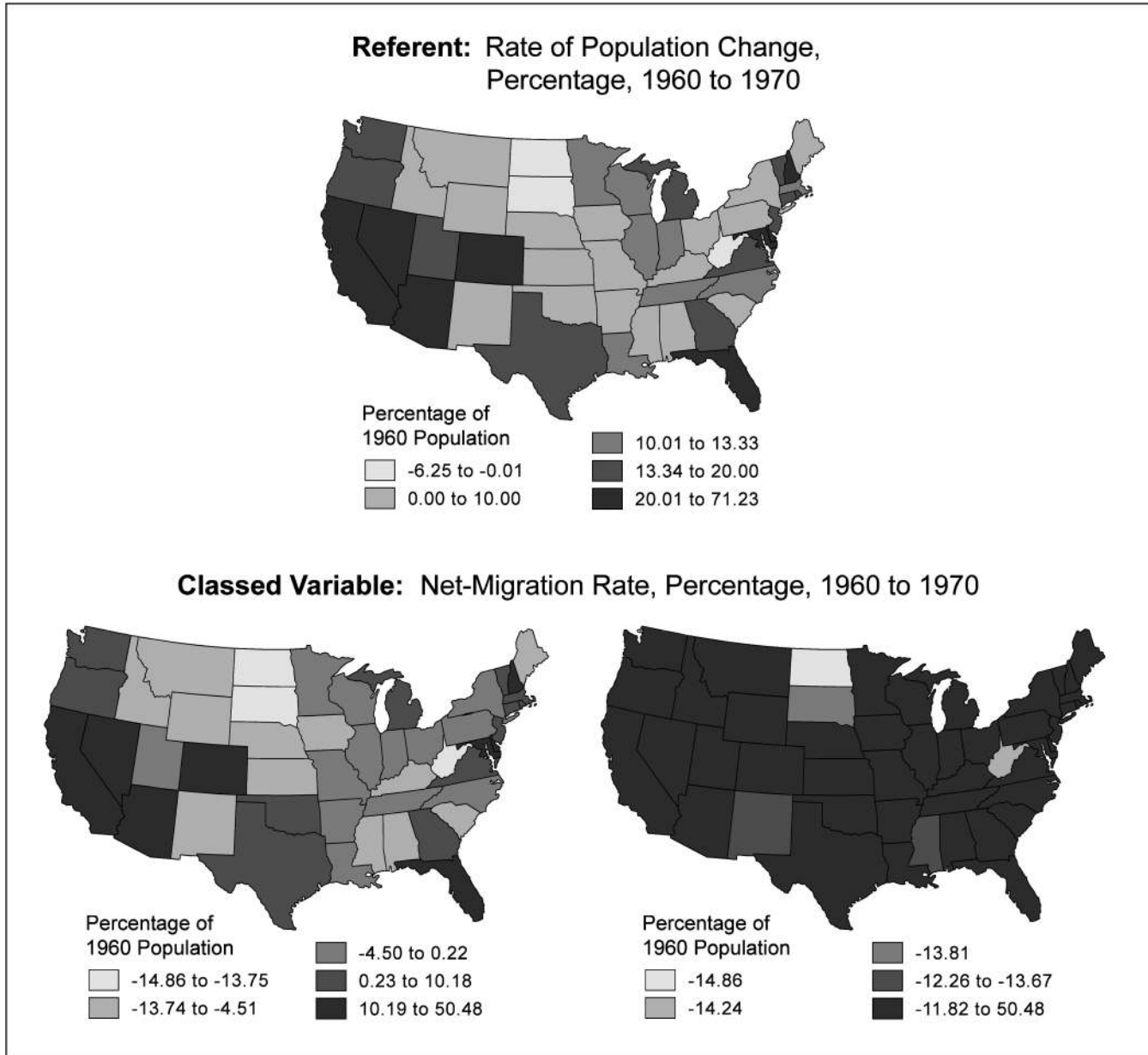
Referent: Rate of Population Change, Percentage, 1960 to 1970

Percentage of
1960 Population

☐ -6.25 to -0.01   ☐ 10.01 to 13.33
☐ 0.00 to 10.00   ☐ 13.34 to 20.00
                  ☐ 20.01 to 71.23

Classed Variable: Net-Migration Rate, Percentage, 1960 to 1970

Percentage of
1960 Population

☐ -14.86 to -13.75   ☐ -4.50 to 0.22
☐ -13.74 to -4.51   ☐ 0.23 to 10.18
                    ☐ 10.19 to 50.48

Percentage of
1960 Population

☐ -14.86   ☐ -13.81
☐ -14.24   ☐ -12.26 to -13.67
           ☐ -11.82 to 50.48

FIG. 9. *The two lower maps are different representations of the same data. An optimization algorithm found cut-points intended to yield displays that look very similar* (lower left) *and very dissimilar* (lower right) *to the map at the top. Cut-points for the upper map include* 0.0, *which separates gains from losses, and* 13.3, *the national rate.*

5. CONCLUDING COMMENT

As Darrell Huff eloquently demonstrated a half century ago, consumers of statistical analyses and data graphics must be informed skeptics. This plea is equally relevant to map users, who need to appreciate the perils and limitations of cartographic simplification as well its power and utility. Because abstract representations of data can distort almost as readily as they can reveal, analytical tools are also rhetorical instruments fully capable of "lying" in the hands of malevolent, naive, or sloppily expedient authors. Huff's engaging little book performed a vital public service by calling attention to the power of analytical tools for self-deception as well as mass trickery.

REFERENCES

BERTIN, J. (1983). *Semiology of Graphics*. Univ. Wisconsin Press, Madison.

HUFF, D. (1954). *How to Lie with Statistics*. Norton, New York.

JENKS, G. F. and CASPALL, F. C. (1971). Error on choroplethic maps: Definition, measurement, reduction. *Annals of the Association of American Geographers* **61** 217–244.

KRUSE, S. (1992). Review of *How to Lie with Maps. Whole Earth Review* **74** 109.

MACEACHREN, A. M., BREWER, C. A. and PICKLE, L. W. (1998). Visualizing georeferenced data: Representing reliability of health statistics. *Environment and Planning A* **30** 1547–1561.

MONMONIER, M. (1976). Modifying objective functions and constraints for maximizing visual correspondence of choroplethic maps. *Canadian Cartographer* **13** 21–34.

MONMONIER, M. (1977). *Maps, Distortion, and Meaning*. Commission on College Geography Resource Paper 75-4. Association of American Geographers, Washington.

MONMONIER, M. (1991). *How to Lie with Maps*. Univ. Chicago Press, Chicago.

MONMONIER, M. (1996). *How to Lie with Maps*, 2nd ed. Univ. Chicago Press, Chicago.

SWAN, J. (1992). Review of *How to Lie with Maps. J. Information Ethics* **1** 86–89.

SRINIVASAN, K. (1998). *Basic Demographic Techniques and Applications*. Sage, Thousand Oaks, CA.