COGNITIVE RESEARCH FOR THE DESIGN OF STATISTICAL RATE MAPS

Linda W. Pickle, National Cancer Institute, and Douglas J. Herrmann, Indiana State University
Linda W. Pickle, NCI, 6130 Executive Blvd., EPN 313, MS 7344, Bethesda, MD 20892

Key Words: cartography, mortality, data visualization

Abstract

Maps have the potential to display the geographic patterns of millions of statistical data points, something impossible using a tabular display. A poorly designed map, however, can fail to convey important underlying features in the data or can even distort their true geographic patterns. From 1990 to 1999, the National Center for Health Statistics conducted an interdisciplinary research program in the human cognitive processes of statistical map reading in an effort to improve map design and thus to communicate geographic statistics more effectively. This paper summarizes the results of this research and provides guidance to others who wish to map statistical rate data.

1. Introduction

There is no graphical alternative to a map when georeferenced statistics for a large number of small areas are involved. With rapid improvements in computing power and mapping software over the past 30 years, mapmaking has become automated and, most recently, available to everyone on their desktop computers (Herrmann 1999). These technological advances launched the field of medical geography and led to the production of over 50 disease atlases around the world (Walter and Birnie 1991). For more than a century, cartographers have designed statistical maps (Howe 1989). Now, mapping software is so user-friendly that even those with no training in cartography can quickly produce maps of statistical data. This is a double-edged sword -- the subject matter expert is now the one who designs the data map, so that knowledge of the data can be brought to bear on the map design, such as in selecting meaningful data ranges. However, inexperience can lead to bad map designs which can fail to convey important underlying features in the data or can even distort their true geographic patterns (Herrmann 1998). The design of maps of statistical rates is unlike that of road maps or maps of demographic characteristics. Positional accuracy is most important in the former, and the latter are typically maps of observed counts with no associated (or important) variance. When rates are to be mapped, it is important to remember that the reader will compare rates in different geographic areas, even if this is not the primary intent of the map-maker. Thus the map must present rates that are comparable, such as directly age-adjusted rates (Pickle and White 1995) and must, in some way, indicate the different degrees of confidence in rates based on different size populations.

We found when attempting to design a recently-published atlas (Pickle 1996) that even the experts disagree on what is the preferred design for a particular map. Because of this, we initiated a research program to better understand the human cognitive processes of statistical map reading (Sirken 1994). Our goal was to improve the design of maps, specifically statistical rate maps, and thus to communicate geographic statistics more effectively. Our research indicates that the reading of any statistical map is achieved through the performance of four cognitive stages, each of which concerns a series of psychological processes that differs between stages. Using this model to segregate map reading tasks, we conducted cognitive studies of basic map styles and specific map elements to determine the impact of their design on the reader's understanding of the mapped data. In this paper we review this and related research and provide guidelines for statistical rate map design. We begin with a description of a cognitive model of map perception, then follow with summaries of related research on basic map styles, color schemes, legend design and indication of unreliable data.

2. Processing stages of statistical map reading

The variety of visual representation schemes is considerable. Maps of the same data constructed using different methods are often seen as reflecting different underlying data (Pickle 1994). However, the accuracy and efficiency of map reading varies not only with qualities of the map, but also characteristics of the task and the reader. These factors and the variables they encompass are summarized in Table 1. Until recently, map reading has been studied very little by cognitive psychologists. Consequently, there were no cognitive models of map perception (although see Kosslyn, 1985 and Simkin and Hastie 1987). We have investigated the cognition underlying map reading and interpretation, on our own and with cognitive psychologists in academe. The results of these investigations build upon initial assumptions about map cognition drawn from the fields of cartography, statistics, and psychology, and have led to a stage model of how people read statistical maps.

Any map-reading task can be broken down into cognitive stages, each of which makes its own peculiar demands on cognition. Our research indicates that the reading of any statistical map is achieved through the performance of four cognitive stages:

(1) map orientation,
(2) legend comprehension,
(3) integration of the map and the map legend, and
(4) discernment of spatial patterns and relationships.

Each of these stages concerns a series of psychological processes, e.g., perception, memory, and problem solving. For map orientation, the visual characteristics of different areas on a map must be resolved into patterns of figure and background and the reader needs to determine what geographic areas are represented. To understand the legend, the map reader must be able to understand the relationship between the rate and its symbolization (i.e., lines, symbols, or area shading); to identify how many rate categories, if any, are being used, as well as to remember the ordering of the categories; and to distinguish among the various symbols or colors used. Integrating the map with the legend necessitates the detection of visual similarities between the legend and map. For example, a reader may be able to easily discriminate among 10 shades of gray on an ordered legend. However, he may be unable to mentally match similar adjacent shadings on the body of the map to the legend. Finally, to see patterns and relationships on a single map or between maps, one must compare visual properties of a map, making use of prior knowledge about the geographic units depicted on it (Herrmann and Pickle 1996; Pickle and Herrmann 1994, 1995).

There is empirical support for such a cognitive model from open-ended interviews and timed experiments (Beu 1989; Jobe and Beu 1991; Kerwin and Herrmann 1992; Hastie 1996). First, map readers distinguished the different stages in describing how they answered questions from a map. Second, the time required to complete a task without interruption (3.5 seconds) was very close to the sum of times required to read a legend and orient themselves to the map (1.64 seconds) and then to discern the rate information from the map (1.9 seconds) (Hastie 1996).

Furthermore, the pattern of accuracies and response times changes with task complexity (Hastie 1996). For example, map readers were faster at identifying the rate for a geographic unit with a multi-hued scale than with a scale of graduated hue or gray, presumably because of the ease of matching a distinct color on the map to its legend counterpart. Similarly, detection of map clusters was found to vary by type of color scales (Lewandowsky 1993, 1999). With more complex tasks, response time may not accurately reveal the effects of visual map attributes on perception time, because these effects may be masked by the longer time spent in higher level cognitive reasoning, for example, averaging the rates of two or more geographic units (Antes and Chang 1990).

Finally, persons in from different disciplines employ different cognitive processes when reading maps and differ in the kind of map they prefer to use (Maher 1992; White 1995). For example, statisticians verbalize what they read from a map less than epidemiologists; additionally statisticians and geographers prefer more complex maps than epidemiologists.

Table 1. Factors that affect the accuracy and efficiency of statistical rate map reading.

Map Attributes (Map Complexity):
- Geographical unit size and number,
- Number of variables represented (e.g., rate, reliability of rates),
- Number and method of class intervals (if any),
- Style of rate representation: Lines, symbols, or shaded areas,
- Legend: Color scheme (e.g., monochrome, multi-hued, double-ended),
- Visual discriminability of symbolization (e.g., among levels of gray),
- Layering (e.g., hatching over symbols or colors),
- Orientation (e.g., north at top of page),
- Labeling,

Map-Reading Tasks:
- Point estimation: Estimation of a statistic for a particular geographic area,
- Pattern detection: Comparing statistical values with geographic characteristics (e.g., identifying regional patterns or data clusters),
- Pattern comparison: Comparing patterns in one map to another

Map-Reader Characteristics:
- Knowledge/expertise: Statistical and cartographic expertise and experience
- Personal Involvement: Level of interest, relevance of map information to reader goals.

3. Comparing basic map styles

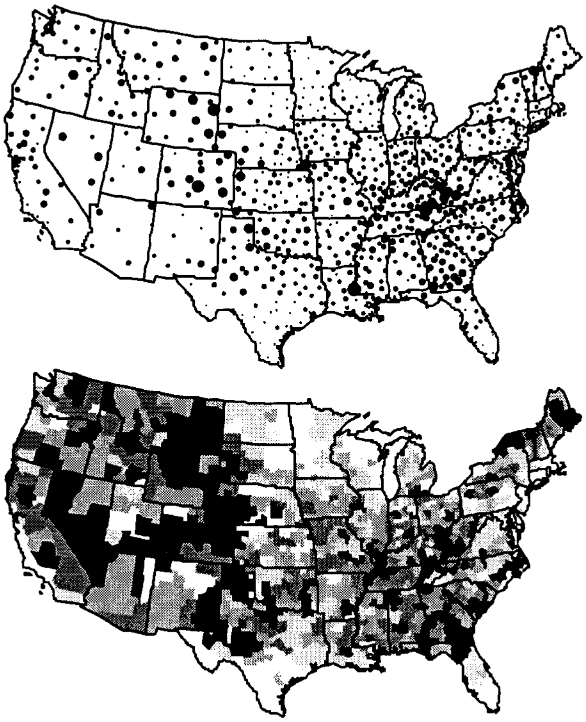
Our initial studies compared preferences for each of a number of basic styles for mapping statistical rates, including unclassed (proportional), classed and smoothed maps of various designs. In unclassed maps, a visual characteristic of the map, such as the darkness or saturation of a color, increases proportionally to the mapped rate, whereas in classed maps, the rates are classified into discrete ranges, each represented by a single visual characteristic. Smoothed maps have had some generalizing algorithm applied to the actual data, such as a moving average; these maps are useful for clarifying geographic patterns but are not suitable for reading a rate in a particular area. For each style, we represented the mapped data by lines of equal rates (isopleth maps), symbols, or shaded geographic areas (choropleth maps). The maps in Figure 1 illustrate how different styles can convey different map patterns, in part because of differences in symbolization (dot vs. shaded area) and in rate classification (none vs. five categories).

Subsequent cognitive studies tested map readers' performance on one or more of the map-reading tasks (Table 1). When the correct answers were known (e.g., point estimation, pattern comparison), response accuracy was used to rank the designs. For more subjective cluster identification, consistency of response across readers was used. Reader preferences were solicited in most studies.

Our first study was a brief experiment comparing numerous map styles followed by a focus group discussion. Epidemiologists liked the choropleth (classed, area shaded) maps best and used them most accurately (Pickle 1994). People with some cartographic training preferred the more complex map designs, but they were equally accurate on all types. In a follow-up study, monochrome classed choropleth maps led to the most consistent identification of map clusters, compared to dot density, pie chart, and double-ended choropleth maps (Lewandowsky 1993). These results led us to use the classed choropleth map style for our mortality rate atlas.

The remainder of this paper will focus on design elements for choropleth maps of rates, although the theoretical approach presented here may be extended to other kinds of maps for rates or counts.

Figure 1. Chronic obstructive pulmonary disease mortality rates mapped using proportional symbol (top) and 5-class choropleth (bottom)styles



Note: darker shade and largest dot represent highest rates

4. Choosing a color scale

Conventional cartographic wisdom recommends the use of a multi-hue color scheme for qualitative data and a sequential color scheme for quantitative data (Dent 1993). A sequential color scheme is a progression of a visual characteristic (e.g., a light to dark sequence of a single hue) corresponding to a progression of the mapped values. Two cognitive experiments showed that very distinct colors (e.g., a rainbow palette) were best for reading a single rate off the map, but that a color gradient was best for cluster recognition (Hastie 1996; Lewandowsky 1993), consistent with color recommendations of cartographers (Dent 1993). The superior accuracy of multi-hue schemes for reading a rate may be because of their greater discriminability than single hues (Pickle 1994; Pickle and Herrmann 1994). Conversely, similar monochrome shades may aid visual identification of areas with similar rates (Dent 1993; Pickle 1994).

We decided to use a double-ended color scheme for the age-adjusted rate map in the Atlas—this is a combination of a color gradient for each of two hues (one for high rates, one for low rates), with the most saturated or darkest shades used for the most extreme rates. The cognitive burden on the reader is less with the double-ended scheme than with a sequential scheme because the reader need only remember and distinguish among half as many shades of the same hue. Lewandowsky's finding that cluster identification was more difficult with a double-ended scheme may have been confounded by his use of a colored hatch pattern rather than a solid fill and by the use of blue to represent high rates in some maps.

We collaborated with other researchers to refine the color recommendations and to further evaluate the performance of double-ended versus single hue schemes. The motivation for these studies was our desire for a single map design that would be good for answering rate readout, clustering, and regional rate questions.

Carswell examined whether a double-ended color scale can be used as effectively as a sequential gray scale for rate readout and cluster identification, and whether the color conventions of "darker for more" and "warmer for more" matter in devising double-ended scales (Carswell 1995). She tested state maps using gray and double-ended scales with all pairwise combinations of red, blue, and yellow. For some maps one color represented high rates while in others it represented low rates. Color schemes including yellow had differences of lightness as well as hue, whereas the reds and blues were of equal saturation and lightness. Study subjects identified the most important cluster of high and of low rates and were then timed in a rate readout test. For each task, the best double-ended color scale was better than the gray scale although, as expected, the gray scale was only slightly worse for cluster identification. The best color scales were those varying both in hue and lightness (i.e., including yellow). Results showed the advantage of the use of color in reducing rate readout errors and also pointed to the importance of following cartographic conventions. More errors occurred using double-ended color schemes where color convention was violated (e.g., using red for low rates or blue for high rates) or in conflict (darker blue and warm yellow - which is high?).

For the next study of color, Brewer and MacEachren tested very carefully chosen color scales for usability on both seven-category full page and five-category quarter page size maps (Brewer 1997). Double-ended, sequential, spectral, and gray scales were tested. Pairs of hues for the double-ended scales were chosen to avoid confusion by color-blind readers, and to avoid simultaneous contrast ("surround effect"), where a color's appearance changes depending on adjacent colors. Lightness and saturation changed in roughly equal steps from top to bottom or from the center to the extremes; these changes were the same for both hues in the pair. The spectral scale was constructed of highly saturated hues with a similar dark to light to dark progression (bottom to top) as the double-ended scales.

Tasks included rate readout, comparison of perceived regional rates, and perceived map clustering. There were few significant differences among the color scales but the gray scale was found difficult to use. This study showed that it is possible to choose hue pairs that avoid (a) confusion by color blind readers and (b) simultaneous contrast. Also, there were few differences in results by map size, probably because fewer rate classes were used for the quarter page maps. The gray scale was perceived as less pleasant and more difficult to use than color scales. The carefully chosen spectral scale performed as well as double-ended scales.

5. Legend design

We conducted an in-house study of various legend designs using a seven-category choropleth map (Pickle, In press). Epidemiologists and statisticians answered rate readout and rate comparison questions and ranked 8 different legends for ease of use. These legends included horizontal and vertical orientation, fixed box size and boxes whose size was proportional to the cutpoint rates, singly or doubly labeled (left and right), and tick marks for cutpoints or range labels. Only one subject made one error in 16 questions, so all of these were usable designs. One epidemiologist summed up the consensus of the group by saying "We could learn to use any of these legends given enough time, but we don't want to spend that time. We would rather get on to what is more interesting to us, i.e., reading the map, so we want easy to understand legends." Consistent with this comment was the ranking of the standard vertically-oriented, fixed-box legend as the most preferred.

6. Mapping reliability

Our final series of experiments examined methods of displaying two variables on the same map, specifically rates and indications of reliability/unreliability. Based on earlier experience with cancer maps (Pickle 1987), we suspected that readers are unable to discern trends or clusters of similar rates if many of the unreliable rate areas are blanked out or grouped together regardless of the level of the rate. Lewandowsky and Behrens, using a spectral color scale with quintile rate categories, tested the following methods of indicating that a rate was unreliable: the color saturation was cut to 40%, the rate color was overlaid with a white hatch pattern, or its shading was removed altogether (Lewandowsky 1995). The control map had no reliability indication at all.

Undergraduates, epidemiologists and geographers were asked to draw around the most important high and low rate clusters on the map and to comment on their certainty of their response. Prior to the test, study subjects were given either very explicit or very vague instructions as to the meaning and importance of rate reliability in drawing conclusions. The location and degree of scatter of the centroids were similar for the cross-hatch and saturation methods but there was greater variability when only reliable areas were shown, indicating disagreement about the cluster locations using this method. Only the professionals were sensitive to the method of indicating unreliability and were less confident of their answers when they were given explicit instructions about unreliable data.

Lewandowsky's results made it clear that we had to highlight unreliable rates in some way in the Atlas. MacEachren and Brewer conducted the next study which compared three more sophisticated ways of doing that (MacEachren 1998). The study design was similar to that of the color study described in Section 4. The color scales used were: a double-ended purple-green, a spectral scale, and a red-yellow sequential scale. Tasks included rate readout, regional rate estimation and cluster identification. Methods tested were color saturation reduction, a double white-black diagonal hatching, and separation of the rate and reliability information into two separate maps. For the latter method, the reader can use just one map or may need to integrate the information from both maps, depending on the task. This study showed that reliability can be indicated with minimal impact, i.e., there were few errors at all for the rate readout task, and whether reliability was indicated made no difference in estimating regional rates or identifying clusters. Map pairs and hatching performed better than the color change method for comparison of regional rates and selecting the most reliable region. The perceived ease of use and pleasantness were best for maps without reliability shown at all and for map pairs and hatching.

7. Summary

Every map imposes a cognitive processing burden on every map reader. The map designer's job is to minimize this burden while conveying the underlying statistical data fairly and accurately. People can only correctly interpret data from a map if they understand how the map represents data. Some maps can be misinterpreted because features are ignored or misunderstood, while others are too complex for most readers to comprehend. For example, people can overlook the fact that a map is presenting data from a skewed or platykurtic distribution when rates are categorized into quintiles. By paying attention to the values associated with each legend category, the map reader will acquire a processing burden - that of trying to figure out the relationship of the values of the mapped statistic to the geographic distribution. Alternatively, if the map reader does not notice the unequal class intervals that are used when quintiles represent a skewed data set, the map reader does not incur a burden. Such a map reader will not recognize that the upper or lower 20 % of data represents an unusually wider set of values. This map reader will not understand the map fully and furthermore will not understand this lack of understanding.

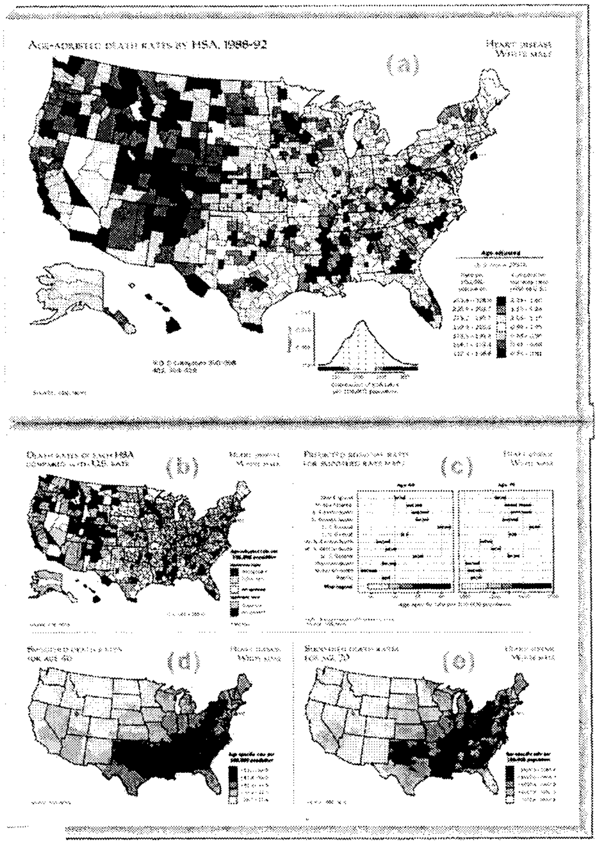
Likewise, people cannot process correctly when they are given a lot of information. It is relatively easy to develop graphic tasks that few (or no) humans can perform. Every single layer on a map asks the reader to think in terms of at least three dimensions: the statistic being mapped plus the two geographic dimensions of the map, in addition to keeping in mind the number and shape of the geographic units and the distribution of the statistic as it relates to the legend categories.

One alternative for complex maps is to display multiple variables simultaneously, i.e., overlaid on top of each other on a single geographic map using graphic styles that are transparent and do not interfere with each other (Lewandowsky 1995). For example, many GIS maps employ such multilayering of symbols. Another alternative is to use dynamic queries in real time on a computer so that different levels of different variables can be displayed sequentially at a rapid rate - resembling perhaps an animation effect. A third alternative is to display several maps, each displaying a variable of interest, arranged so that the correspondence between physical locations on the separate maps can be discerned readily. We know little about how these different kinds of multilayer designs perform with map readers. The knowledge that we gained through the use of cognitive experimentation can facilitate the design of maps, but more complex map designs, such as computer-based representations, remain to be tested. As yet, not enough research has been done to understand all of the possible interactions between map variables.

In conclusion, we feel that we learned a great deal about map design from this series of cognitive experiments and we applied this knowledge to the design of our new atlas (Pickle 1996). A double-page layout as shown below was included for each sex, race, and cause. Included are a classed choropleth map of age-adjusted death rates (Fig. 2a) using a double-ended color scheme with the warmer color for high rates and hatching over areas with unreliable rates; a rate significance map (Fig. 2b); and smoothed age-specific maps (Fig 2d,e). The latter permit the reader to identify differences in geographic patterns by age that would be masked in the age-adjusted rate map. A regional rate graph (Fig. 2c) was included because estimating regional rates was found in the experiments to be more difficult than rate readout or cluster identification. The page layout, labeling, and color schemes are consistent throughout the book.

The Atlas has been well received, winning several design awards to date. The results of our cognitive research helped to elucidate why maps are or are not accurately understood. By better understanding the

Figure 2. Sample of Atlas two-page layout (see color on p.13, http:\\www.cdc.gov\nchswww\data\atlasmet.pdf)



cognitive processes that underlie the reading of a statistical map, we believe that we have shown that it is possible to improve map design (Sirken 1993).

References

Antes J. R. and Chang K. (1990). An empirical analysis of the design principles for quantitative and qualitative area symbols. Cartography and Geographic Information Systems, 17, 271-277.

Beu D. H., Mingay D. J., White A. A. (1989). Cognitive experiments in data presentation. Proceedings of the Statistical Graphics Section of the ASA, Pp. 30-35.

Brewer C.A., MacEachren A.M., Pickle L.W., Herrmann D. (1997) Mapping mortality: Evaluating color schemes for choropleth maps. Annals of the Association of American Geographers, 87(3):411-38.

Carswell C.M., Kinslow H.S., Pickle L.W., Herrmann D. (1995) Using color to represent magnitude in statistical maps: The case for double-ended scales. In: Pickle LW, Herrmann DJ, eds. Cognitive Aspects of Statistical Mapping. Working Paper Series Report no. 18. Hyattsville, Maryland: NCHS. Pp. 201-228.

Dent B. D. (1993). Cartography: Thematic Map Design, 3rd edition. Dubuque, Iowa: Wm. C. Brown Publishers.

Hastie R., Hammerle O., Kerwin J., Croner C., Herrmann, D. J. (1996). Human performance reading statistical maps. J of Experimental Psychology: Applied, 2, 3-16.

Herrmann D.J. and Pickle L.W. (1996) A cognitive subtask model of statistical map reading. Visual Cognition 3(2):165-190.

Herrmann D.J. (1998) What people can and do get from visual displays: A review of Tufte's *Visual Explanations*. JASA, 93, 398-399.

Herrmann D. (1999). Potential contributions of the CASM movement beyond questionnaire design: Cognitive technology and survey methodology. In: M..Sirken, D. Herrmann, S. Schecter, N. Schwarz, J. Tanur, R. Tourangeau, R. (Eds.) Cognition and Survey Research. Boston: Wiley.

Howe G. M. (1989). Historical evolution of disease mapping in general and specifically of cancer mapping. In P. Boyle, C. S. Muir, & E. Grundmann (Eds.), Cancer mapping. New York: Springer-Verlag.

Jobe J. B., and Beu D. H. (1991). Cognitive factors in interpreting statistical graphs. Unpublished manuscript. Hyattsville, MD: NCHS.

Kerwin J. and Herrmann D. (1992). Benefits of pretesting statistical maps with cognitive interviews. Unpublished manuscript. Hyattsville, MD: NCHS.

Kosslyn S. N. (1985). Graphics and human information processing. JASA, 80, 499-512.

Lewandowsky S., Herrmann D., Behrens J. T., Li S., Pickle L.W., Jobe J. B. (1993). Perception of clusters in statistical maps. Applied Cog Psych, 7, 533-551.

Lewandowsky S., Behrens J.T., Pickle L.W., et al. (1995) Perception of clusters in mortality maps: Representing magnitude and statistical reliability. In: Pickle LW, Herrmann DJ, eds. Cognitive Aspects of Statistical Mapping. Working Paper Series Report no. 18. Hyattsville, Maryland: NCHS. Pp. 107-132.

Lewandowsky S. (1999). Statistical graphs and maps: Higher level cognitive processes. In M. Sirken, D. Herrmann, S. Schecter, N. Schwarz, J. Tanur, R. Tourangeau (Eds.) Cognition and Survey Research. Boston: Wiley.

MacEachren A.M., Brewer C.A., Pickle L.W. (1998) Visualizing georeferenced data: Representing reliability of health statistics. Environment & Planning A.

Maher R. (1992). Statistical map perception as a function of expertise for congressional staffers, epidemiologists, and statisticians. Unpublished rept, Hyattsville, MD: NCHS.

Pickle L. W. and Herrmann D. J. (1994). The process of reading statistical maps: The effect of color. Statistical Computing and Graphics Newsletter. 5, 12-15.

Pickle L.W., and Herrmann D.J., eds. (1995) *Cognitive Aspects of Statistical Mapping*. Working Paper Series Report no. 18. Hyattsville, MD: NCHS. Pp. 201-228.

Pickle L. W. and White A.A. (1995) Effects of the choice of age-adjustment method on maps of death rates. Statistics in Medicine, 14:615-627.

Pickle L. W., Mason T. J., Howard N., Hoover R., Fraumeni J. F., Jr. (1987). Atlas of U.S. cancer mortality among whites: 1950-1980. DHHS (NIH) Publ. No. 87-2900. Washington, DC: U.S. Government Printing Office.

Pickle L.W., Herrmann D., Kerwin J., Croner C.M., White A.A. (1994) The impact of statistical graphic design on interpretation of disease rate maps. Proceedings of the Statistical Graphics Section of the 1993 Annual Meeting of the ASA, p.111-116.

Pickle L.W., Mungiole M., Jones G.K., White A.A. (1996) *Atlas of United States Mortality*. Hyattsville, MD: National Center for Health Statistics.

Pickle L.W., Herrmann D.J., Wilson B.F. (In press) A legendary study of statistical map reading: The cognitive effectiveness of statistical map legends. Journal of Computational and Graphical Statistics.

Simkin D., and Hastie R. (1987). An information-processing analysis of graph perception. JASA, 82, 454-465.

Sirken M., Herrmann D., White A. (1993). The cognitive aspects of designing statistical maps. Proceedings of the Survey Research Section of the ASA, p. 586-591.

Walter S. D. and Birnie S. E. (1991) Mapping mortality and morbidity paterns: An international comparison. Int J Epidemiol 20:678-89.

White A.A., Pickle L.W., Herrmann D.J., Croner C.M., Wilson B.F. (1995) Map design preferences associated with professional discipline. Proceedings of the Statistical Graphics Section of the 1994 Annual Meeting of the ASA, p. 54-59.