package prints the necessary “Content-type” line to inform the world that it is producing an HTML document; all other fields need to be filled in by the script. The program starts by calling the `&ReadParse` routine. If the `submit` field of the form has a value, that means that the program is being called from a form; otherwise, it’s being called directly, and it simply needs to print the necessary HTML to display the form. To keep this example simple, we’ll just have one fill-in-the-blank field, and a set of three checkboxes, but any of the more complex form elements can be easily accommodated with this scheme. Similarly, the results are just printed back to the browser, but the full range of perl’s capabilities could be used to do something more interesting with this information.

Further Resources

This article doesn’t begin to show you the wide variety of packages and scripts that are available to extend perl’s capabilities. As always, the first place to start is the CPAN archives (`http://www.perl.com/perl/CPAN/CPAN.html`). Another source of scripts is the Metronet archive at `http://www.metronet.com/perlinfo/scripts`.

Phil Spector
UC at Berkeley
spector@stat.berkeley.edu

TOPICS IN INFORMATION VISUALIZATION

Emphasizing Statistical Summaries and Showing Spatial Context with Micromaps

By Daniel B. Carr and Suzanne M. Pierson

1. Introduction

This article concerns redesigning a choropleth map. During my (Dan) fellowship at the Bureau of Labor Statistics (BLS), the staff showed me press releases with maps similar to the map in Figure 1. They did not like the map and asked me to develop new and innovative methods for displaying the data. My thoughts turned quickly to micromaps. In response to Tony Olsen’s guiding query, I had previously proposed micromaps for linking row-labeled plots to ecoregion maps (Olsen, Carr, Courbois, and Pierson 1996). Sue Pierson’s first implementations and Pip Courbois’ variations demonstrated to the team that micromaps do more than provide links. Micromap sequences directly reveal spatial patterns. Below is our rationale for evolving from a traditional choropleth map to a new and powerful template that links micromaps and row-labeled plots.

CONTINUED ON THE NEXT PAGE

```
sub doform{
# The action field in the form below is the location of this script.

print &PrintHeader;
print <<EOF;
<html>
<title> Sample Form </title>
<body>
<h3 align=center><b> Welcome to the Sample Form! </b></h3>
<form method="post" action="http://wherever/cgi-bin/sampleform.pl">
Fill in the blank:<br>
My name is <input name="name" size=40>
<p>Choose your favorite color:<br>
  <input name="check" type=checkbox value="Blue"> Blue<br>
  <input name="check" type=checkbox value="Green"> Green<br>
  <input name="check" type=checkbox value="Red"> Red<br>
<p>
<input type=submit name="submit" value="submit">
</form>
</body> </html>
EOF
}
```

2. Visual and Representational Problems

The choropleth map in Figure 1 has several visual and representational problems. Visually, the map is reminiscent of pen plotter era. Representing values by line density and crosshatched patterns is predicated upon notions of reproduction ease and cost, not upon notions of aesthetic communication. That is, lines copy better than half-toned gray and less expensively than color. Questions about state grouping reveal the groups to be administrative divisions that have little bearing on analysis. The state grouping can be dropped. However, the problems extend beyond unaesthetic appearance and irrelevant grouping.

When evaluating Figure 1, consider the story to be presented. The story focuses on sample-based estimates of unemployment rates with associated uncertainty estimates, counts of unemployed, and spatial indices. Directly to the point, The Power of Maps by Wood (1992) contains an intriguing chapter entitled “Every Map Shows This ... But Not That.” If we look though statistical eyes at what is and is not represented in Figure 1, we see cartographic bias and representational problems.

The cartographic bias in Figure 1 reflects representational choice and relative emphasis. The cartographic literature (for example, see Bertin 1983, MacEachren 1994) provides systematic treatment of ways to represent variables on maps. Cartographic choices implicitly assume that the best representation, position along a common scale, is devoted to the two spatial coordinates. This leaves second best choices to show data values and other information. The systematic treatment fails to show scatterplot alternatives favoring statistics, e.g., melanoma rates on the y axis, cloud-free days on the x axis, latitude encoded as circle area and longitude encoded as circle color. Carr, Littlefield, Nicholson, and Littlefield (1987) provide an early example of a balanced representation. They use position along a common scale for data values, spatial coordinates, and time while linking subsets across panels with color. Our redesign of Figure 1 also uses position along a common scale for both data values and spatial coordinates, improving representation of the statistical information.

The traditional cartographic choice plays out in terms of emphasis. Figure 1 emphasizes state boundaries. The often-used Albers projection preserves the relative areas of the continental U.S. states. A large number of vertices is devoted to boundary representation and substantial graphic space is dedicated to representing state area. In contrast, detail associated with unemployment rates is limited to a few class boundaries. The unemployment rates appear as a caricature because the conversion to class intervals adds noise.

Class interval options are caricature options for statistical distributions. Common choices include equal size intervals and gap-based intervals covering the range. Even for percentage point options, the default is often based on the number of regions. Carr and Olsen (1995a and 1995b) argue for class intervals based on percentage points of a cumulative distribution function chosen for interpretation purposes, such as the percent of people involved. Carr and Olsen also propose a visual summary of the distribution that can appear in a small legend. In the current case, only fifty-one estimates are to be represented. We take the radical approach of directly showing all estimates.

The representational problem is more than caricaturizing the unemployment rates. The map does not show estimate uncertainties. While MacEachren (1994) describes methods for representing uncertainty on maps, they are rarely used. (A notable exception is Pickle, Mungiole, Jones, and White 1996). Beyond failing to provide details about estimate precision, the omission of estimate uncertainties is serious on two counts.

First, the presence of uncertainties suggests that sound statistical methodology produces the estimates. The world is awash in convenience-based guesstimates. The public needs clues to decide if data is statistically sound. Omission of clues helps politics and sales compete on an equal footing with science. Frederick Mosteller (*) says, “It is easy to lie with statistics, but easier to lie without them.” A corollary is: it is easy to lie with confidence intervals, but it is much easier to lie without them.

A second consideration is that the public needs to be educated about uncertainty. People may not like the probabilities of weather forecasters, but over time the reporting convention has become familiar. The failure to show confidence intervals for estimates is a missed chance to educate the public.

The most important design task is to represent the statistical summary. The spatial component of the summary is important, but secondary. The new design should reflect this priority.

3. Dot Plots, Visual Simplicity and Grouping Considerations

The variable of interest is the state unemployment rate. Labeled dot plots (Cleveland 1985, 1993) provide a good way to show such estimates. Unfortunately, dot plots have been slow to appear in government publications. The efforts of Carr, Valliant, and Rope (1996) are intended to help remedy the situation by providing

Figure 1: A standard choropleth map.

JAVA-based network tools and examples using government data. Our effort extends the scope of the examples. The design goals for the current example include 1) adding information to ease interpretation and 2) simplifying visual appearance to facilitate communication with the public. This section addresses the second task, simplifying visual appearance.

For improved visual interpretation, Cleveland (1985) advocates presenting dots in sorted order. Carr (1994) and Carr and Olsen (1996) echo this advice and note that sorting improves plot simplicity by reducing the visual path between dots.

One can further simplify plot appearance. Important visual simplification techniques include grouping and layering. For example, see Kosslyn (1994) for a discussion about grouping and Tufte (1983 and 1990) for discussions of small multiples and layering. Graphics software has not successfully automated thoughtful grouping and layering, so some thought about the current example is instructive.

A list with fifty-one lines can be visually intimidating (see Carr and Olsen 1996). By analogy, a fifty-one line paragraph may visually intimidate many people. One can lose one's place. Breaking a long paragraph into short paragraphs helps the reader with visual tracking. Similarly, breaking a list of names into groups helps the reader in visual tracking. A further benefit of creating smaller perceptual groups is that readers can easily spot names at group edges. Spotting an interesting name draws the reader into the graphic. In other words, small perceptual groups increase the number of interest-based entry points.

How do we divide fifty-one states into visually effective groups? (Here we include the District of Columbia as a state, but it is often preferable to treat D.C. as a city.) One choice would be to partition the states using large jumps in the sorted rates. This has merit, but can get awkward when many states have similar values. Our approach starts with regular partitioning into groups of five. (For many applications Kosslyn (1994) recommends groups of four or less.) For vertical grouping we find that groups of five facilitate counting and still allow quick label and value matching by relative position. For example, one can easily match the third label in a group of five labels to the third dot in a group of five dots. This works even when the labels and dots are separated by nearly a page width. Consequently, grouping obviates the need for the horizontal dots that Cleveland (1985 and 1993) used to assist in visual tracking. The partitioning produces ten groups of five and one group of one.

The first layer of grouping produces eleven groups. Eleven groups are too many to put into a single simple-appearing perceptual unit. The groups need to be grouped. As shown in Figure 2, we create an additional information layer using three larger groupings with a 5-1-5 pattern. This creates symmetry and calls attention to the median.

The basic layout for Figure 2 is an eleven row (5-1-5 pattern) by four column matrix. The state names appear in the second column and unemployment rates with 95 percent confidence intervals appear in the third column. (The confidence intervals are model-based and subject to refinement. The BLS has not extended the confidence interval calculations to the seasonally adjusted estimates it often shows.) The confidence intervals detract a bit from the goal of visual simplicity, but the above discussion motivates their inclusion.

Representation of confidence intervals is a design challenge. Typical error bars draw visual attention to the least precise estimates (Carr 1994). Further, error bars detract from the visual flow in following the estimates. In Figure 2, the interval endpoints appear as small gray dots. Connecting adjacent endpoints with a thin black line reinforces desirable vertical flow and brackets the group of estimates. The pinch points call attention to the most precise estimates. In other examples, an estimate dot can overplot the confidence interval dots. In isolated cases, the slope of confidence interval lines from above or below can suggest the size of the hidden interval. In the new design, even those new to confidence intervals may surmise a connection between the California pinch point and the large number unemployed. Perhaps better approaches will emerge, but the approach in Figure 2 has considerable merit.

The next steps are to show the remaining secondary information. The information includes spatial positions and estimated numbers of people unemployed. The final step is to link everything together with the state names.

4. Micromap Design

Associated with each unemployment rate is a spatial position, the state location. In the current example, as in many studies, the exact spatial position and precise boundaries are not important. All that is needed is a map caricature showing the general position and neighborhood relationships. As demonstrated in the first column of Figure 2, a micromap for each group of five can show the state locations. A full page map is not required.

The design of small maps requires attention because distinguishing hues in small regions can be difficult. The map caricature needs to enlarge small regions while retaining enough features to provide region recognition.

The task is not as simple as it might seem. After several independent attempts, we chose to modify a state visibility map developed over a decade ago by Monmonier and illustrated in *Mapping it Out* (Monmonier 1993).

Figure 2 shows ten micromaps on a standard page in portrait orientation. Making the micromaps much smaller will complicate color perception for relatively small states like Rhode Island. The example is pretty close to the minimum size limit. Observe that Illinois has the median rate and appears in black in the two middle maps. This avoids the need for an eleventh micromap.

5. Related Data

Carr, Valliant and Rope (1996) argue that graphics should provide metadata to facilitate proper interpretation. The current example is static, so we cannot exploit web-based access to the Bureau of Labor Statistics Handbook of Methods or to other information. However, the design readily accommodates one or two additional columns. Figure 2 shows the number of unemployed in the fourth column. While unemployment rates are useful for comparison purposes, the number of unemployed shows the importance of the rates in terms of human lives.

The confidence intervals for the counts were not available for this article. As an approximation, the rate intervals could be scaled by the population size. We chose to focus attention on the rates and to omit confidence intervals for the secondary information. The elegance of the confidence interval representation depends on sorting. Adding confidence intervals to the count estimates competes with the goal of visual simplicity.

6. Region Labels and Linking

In Figure 2, the state labels link the information. The relative vertical position of a label and a dot within a group of five is an adequate positional link. The colored dot beside the state name may not increase the matching speed over the positional link, but may remove doubt about a correct match. The colored dot in the label is an important link to a particular state. Some may assume that everyone knows state positions and question the need for a color link. However, many informed U.S. citizens may hesitate when labeling all the states on a map. Further, altered boundaries in the map caricature may slow recognition for some states. We conjecture that color links to the map increase matching speed and sometimes provide an educational tool. The color links go both ways. Some people use maps rather than names to find their state and the corresponding estimate.

With the holiday season as motivation, we seriously considered using colored names in the label column and dropping the dots. This drops one visual element and simplifies plot appearance. However, Monmonier (1993) recommends against colored labels because they need to be large enough to carry color and are difficult to read. Switching to a bold font in Figure 2 works fine for carrying color. We concur that changing colors makes reading a little harder. Further, the colorful names draw visual attention away from other columns. Still, we think the problems are minor and that colored names might be used on occasion to add variety.

7. Interpretation and Comparison

Figure 2 provides much more statistical information than Figure 1. Consider four questions.

1. What is the unemployment rate for California?
2. Is the unemployment rate higher for California or Alaska?
3. What are the confidence bounds for the California estimate?
4. On the average how many were unemployed in California?

For the first question, Figure 2 provides a more precise determination of the estimated rate. For perceptual accuracy of extraction (Cleveland and McGill 1984), dot plots with grid lines are a hard graphic to beat. For question two, Figure 2 provides a complete ranking while Figure 1 only provides a ranking of equivalence classes. Figure 1 does not provide answers for questions three and four.

The traditional choropleth map does not do as well as the linked micromap row-labeled plot template in regard to the above questions. This motivates the search for tasks in which the large choropleth map has performance advantages. Two tasks seem evident, finding the value of a position-known state and locating the values of neighboring states. The micromap template requires a scan of small maps or a list of fifty-one names before linking to a value. The scan is a slow process. An additional alphabetic-label position-link column (not illustrated here) may speed locating a given state by name. The position link following a name can be little black dots in a 5-1-5 pattern with the location dot highlighted by color (or shape). For Hawaii, the third dot would be red and link to red in the third vertical group. The real memory and search intensive task in using micromaps is to find values for a state's neighbors. With micromaps one can quickly observe if neighboring states have similar rankings, but that is not the same as finding all the values. Traditional choropleth maps have a few advantages.

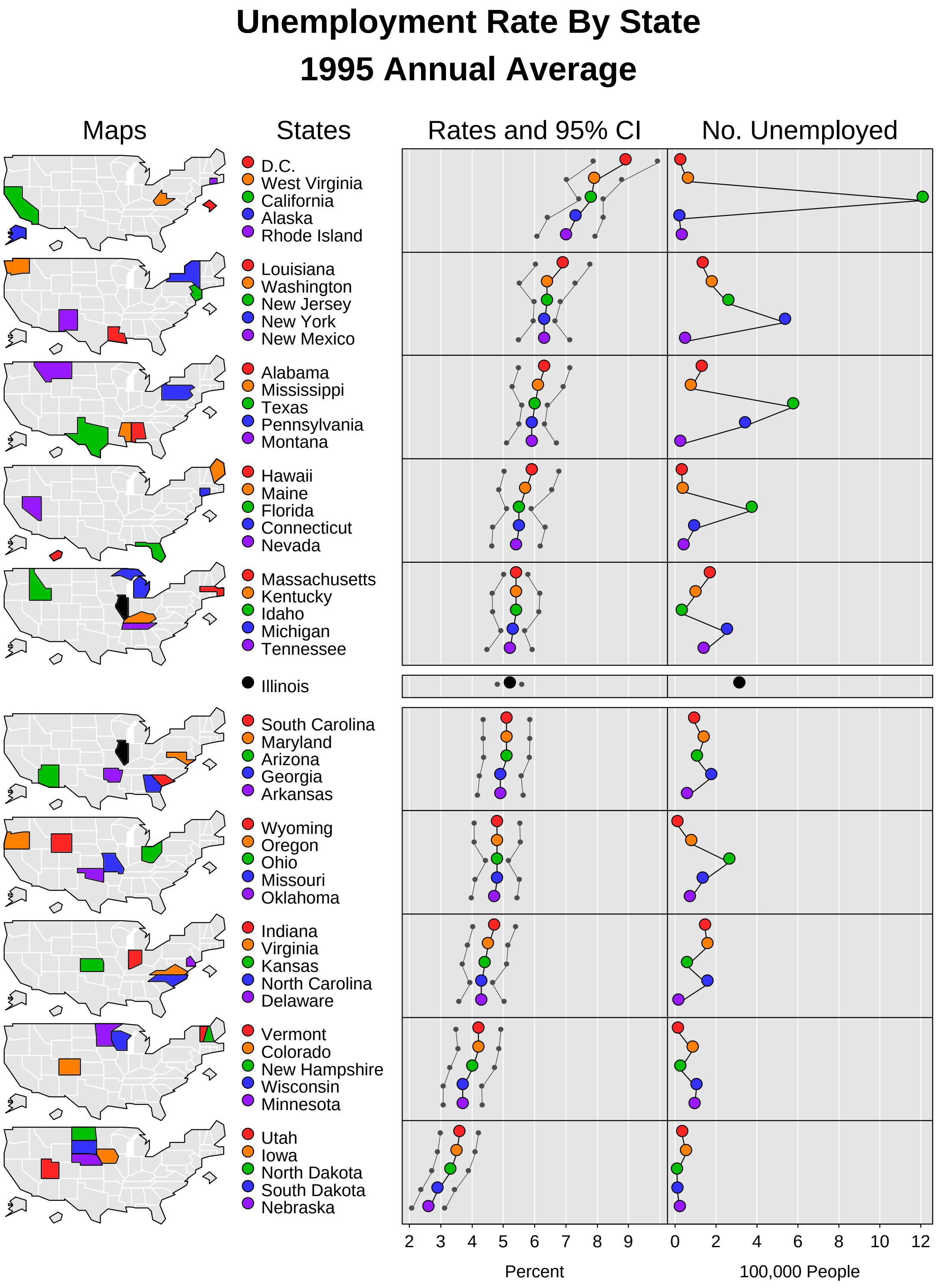

Figure 2: Improved display with micromaps.

Cartographers may also suggest that a single choropleth map provides better global spatial pattern perception than a sequence of micromaps. Since integrating information while scanning across all the micromaps is nontrivial, the claim is likely true. However, many important tasks involve local pattern perception.

For tasks involving local pattern perception, micromap sequences may be very competitive to a single choropleth map. While the colors in Figure 2 may be distracting, ten micromaps provide the rough equivalent of ten class intervals. For sound perceptual reasons, typical choropleth maps show six or fewer class intervals. This complicates direct comparison in cognitive testing. Different class intervals bring out different patterns, so there may be no clear winner.

The micromap patterns can be quite suggestive. Figure 2 shows many small groups and raises questions about economic similarities. As two of several examples, Vermont and New Hampshire form a pair while Virginia and North Carolina form another. The two bottom micromaps show a larger group of states in the upper Midwest and Northern Plains. It is not hard to integrate patterns across two small juxtaposed maps. While beyond the domain of local pattern perception, the micromap sequence wins hands down when it comes to ranking. The cyclic colors in a rough spectral order break the ties within the ten micromaps. The combination of multiple maps and color provides a complete ranking without reference to the other columns.

Variations on micromaps may strengthen global spatial pattern perception or at least bring out additional patterns. A darker shade of gray can distinguish all states above median and provide another layer of information in the top five micromaps. The darker gray region appears the same in all five maps, except for the overriding hue-linked states that provide compositional detail for the high unemployment region. New patterns emerge. For example Appalachian states have above average rates. A similar approach accentuates the low unemployment region in the bottom five micromaps.

8. Closing Remarks

Viewing graphics as puzzles to be solved is often instructive. How do the pieces fit together and what do they mean? We conjecture that the template illustrated by Figure 2 is a puzzle accessible to many. Learning to read a dot plot is easy. Linking by color and position is simple. Determining that the visual islands are Alaska, Hawaii, and D.C. should be manageable by those not familiar with the U. S. A deep understanding of confidence intervals goes beyond the graphic, but readers new to statistics may surmise that the big dots for rate estimates are not exactly THE TRUTH. For those with a little background, the figure is pretty close to being self-explanatory.

Figure 2 has educational merit beyond showing unemployment rates. One can learn about the positions of the states. The figure can prepare people to answer the question, to what side does the median belong? More importantly, the figure can prepare people to learn about uncertainty. To the thoughtful, the figure suggests that rank orderings are a bit arbitrary in the presence of uncertainty. The figure also suggests that rate magnitude and rate importance are distinct concepts. The template of linked micromaps and row-labeled plots extends to many other spatial contexts. For example, one can show county data within a state. When fifty or fewer counties will be displayed, a layout similar to Figure 2 will likely suffice. The main challenge would be to develop a county within state visibility map. Some states are not easy.

Many variations of the template are possible besides those mentioned above. Some may prefer a different set of hues. For example, one can pick a set designed for the color blind. A good reference concerning color choice and mapping is Brewer (1994). Other variations may slightly improve perceptual accuracy of extraction for dot plots. For example translucent dots in the right two panels will help keep the grid lines visible. A small, similar hue dot inside each big dot may help locate the dot center precisely without being too distracting. The possibilities are numerous.

The data and Splus source code used to generate Figure 2 are available through anonymous ftp (`galaxy.gmu.edu`). The newsletter software directory seems to change periodically. The current path for this article is `pub/dcarr/newsletter/micromap`. For Splus users, the matrix layout tools should be of interest by themselves and a technical report with documentation details should be available in the same time frame as this article.

I (Dan) continue to seek design challenges and opportunities for collaboration. Also, I appreciate gentle comments about potential improvements. Please contact me at the address below.

Acknowledgments

Conceptual work behind this paper was supported by the EPA under cooperative agreement No. CR8280820-01-0. Specific work on this application was supported by the BLS. The article has not been subject to review by either the EPA or the BLS, does not necessarily reflect the view of the agencies, and no official endorsement should be inferred. Special thanks go to Tony Olsen for the design challenges, Pip Courbois for experiments in micromap variations, and to Rick Valliant and Dan Rope for coordinating the BLS data and graphics access.

References

Bertin, J.B. (1983), *Semiology of Graphics Diagrams Networks Maps*, Translated by Berg, W.J., London, UK: The University of Wisconsin Press.

Brewer, C.A. (1994), “Color Use Guidelines for Mapping and Visualization,” in *Visualization in Modern Cartography*, 123–147, eds. MacEachren, A.M., and Taylor, D.R.F., Oxford, UK: Pergamon/Elsevier Science.

Carr, D.B., (1994), “Converting Plots to Tables,” Technical Report No. 101, Center for Computational Statistics, George Mason University, Fairfax, VA 22030.

Carr, D.B., Littlefield, R.J., Nicholson, W.L., and Littlefield, J.S. (1987), “Scatterplot Matrix Techniques For Large N,” *Journal of the American Statistical Association*, 82, 424–436.

Carr, D.B., and Olsen, A.R. (1995a), “Parallel Coordinate Plots For Representing Distribution Summaries in Map Legends,” *Proceedings 1 of the 17th International Cartography Association Conference, 10th General Assembly of the ICA*, 733–742.

Carr, D.B., and Olsen, A.R. (1995b), “Parallel Coordinate Variants of CDF and Quantile Plots,” *Statistical Computing and Statistical Graphics Newsletter*, Vol. 6, No. 1 13–18.

Carr, D.B., and Olsen, A.R. (1996), “Simplifying Visual Appearance By Sorting: An Example Using 159 AVHRR Classes,” *Statistical Computing and Statistical Graphics Newsletter*, Vol. 7, No. 1, 10–16.

Carr, D.B., Valliant, R., and Rope, D. (1996), “Plot Interpretation and Information Webs: A Time-Series Example From the Bureau of Labor Statistics,” *Statistical Computing and Statistical Graphics Newsletter*, Vol. 7, No. 2 19–26.

Cleveland, W.S., and McGill, R. (1984), “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods,” *Journal of he American Statistical Association*, 79, 531–554.

Cleveland, W.S. (1985), *The Elements of Graphing Data*, Summit, NJ: Hobart Press.

Cleveland, W.S. (1993), Visualizing Data, Summit NJ: Hobart Press.

Kosslyn, S.M. (1994), *Elements of Graph Design*, New York, NY: W.H. Freeman and Company.

MacEachren, A.M. (1994), *Some Truth with Maps: A Primer on Symbolization &Design*, Washington D. C.: Association of American Geographers

Monmonier, M. (1993), *Mapping It Out*, Chicago, IL: The University of Chicago Press.

Olsen, A. R., Carr, D.B., Courbois, J.P., and Pierson S.M. (1996), “Presentation of Data in Linked Attribute and Geographic Space,” Poster presentation, ASA Annual Meeting, Chicago, Il.

Mosteller, F. (*), Personal communication—The quote is not from a publication but likely originated in a speech.

Pickle, L.W., Mungiole, M., Jones, G.K., and White A.A. (1996), *Atlas of United States Mortality*, Hyattsville, MD: Public Health Service Pub. No. 97-1015.

Tufte, E.R. (1983), *The Visual Display of Quantitative Information*, Cheshire, CT: Graphics Press.

Tufte, E.R. (1990), *Envisioning Information*, Cheshire, CT: Graphics Press.

Woods, D. (1992), *The Power of Maps*, New York, NY; The Guilford Press.

Daniel B. Carr
George Mason University
dcarr@voxel.galaxy.gmu.edu

Susanne M. Pierson
Anteon Corporation
spierson@heart.cor.epa.gov