Comment on W.S. Cleveland, A Model for Studying Display Methods of Statistical Graphics

Leland Wilkinson
SYSTAT, Inc. and Northwestern University

Revised version published in *Journal of Computational and Graphical Statistics, 2*, 1993, 355-360.

The social psychologist Susan Fiske has shown that commonsense notions about human behavior are as often wrong as right. For example, the popular maxim "opposites attract" is generally false. Numerous experiments by another psychologist, Amos Tversky, have demonstrated that we -- novices and Bayesian statisticians alike -- are poor judges of quantity and chance. Finally, perceptual experiments, including some by Bill Cleveland himself, have shown that statisticians and other humans succumb to visual illusions when viewing statistical graphs.

Martin Gardner, Richard Feynman, and the Amazing Randi have all shown how easy it is for scientists to fool themselves. Because we all think we are expert psychologists, we are at greatest risk when we study ourselves and our perceptions. Bill Cleveland has done a service to statisticians by grounding discussions about graphics in experimentation. His ideas on graphical elements, based on a series of experiments (Cleveland, 1985), have influenced statistics packages (e.g. S^+, STATA, SYSTAT) and have inspired further experimentation in graphical perception (see Spence and Lewandowsky, 1990).

An experimental viewpoint should not diminish the work of graphic designers and others who have creative instincts for good displays. Bertin and Tufte, for example, have shown that effective graphs need not be dull or lack style. We must be suspicious of any design prescription that is not supported by experimental results, however. Good design does not always lead to effective statistical graphics.

Most of Cleveland's early experiments concerned graphical *elements* -- lines, angles, areas, volumes, colors. This paper applies his thinking to graphical *composites* -- reference grids, plotting symbols, and aspect ratios. Although its title is "A Model for Studying Display Methods of Statistical Graphics," it is really more concerned with a variety of approaches to evaluating the use of these composites. While I cannot disagree with most of his conclusions about good usage, I find the model itself somewhat restrictive. Cleveland wishes to stay grounded in perceptual psychology, but the topics he discusses also involve areas of higher cognition.

Cognitive psychologists such as Pinker (1990), Simken and Hastie (1987), and Kosslyn (1989) have proposed process models for graphical perceptual processing. Statisticians, on the other hand, like to think of the meaning of a graph as predefined: construct a graph properly and its meaning will be self-evident. A schematic (box) plot, for example, is a summary of the first few letter values of a batch. Viewers don't always see it that way, however. They may read meaning into the width of the box even when it is constant in every box they see.

A process model of graphical information processing covers all the stages in the graphical communication event, from the statistician who has information to communicate, to the viewer who makes a judgment about that information. These stages are:

1. The quantitative/qualitative information
2. The retinal image
3. The decomposition in the visual cortex
4. Integration and transformation via temporary storage in short term memory and schemas accessed in long term memory

Each of these stages in the process model has implications for some of the points covered in Cleveland's paper. Here are a few.

1. Quantitative/qualitative information. Cleveland mentions that "There are two types of information in the data region of a graph - quantitative and categorical." Actually, quantitative and categorical information is part of what the statistician *intends* to convey to the viewer. The actual information in the data region is a more complex arrangement of texture, form, edge, and other features. Sometimes categorical information is mapped onto quantitative features, such as angles or line lengths and quantitative information is mapped onto categorical features such as numerals.

This distinction between the information the statistician intends to convey and the actual organization of the graph is important because most of the formal arguments about graphs involve this stage. The statistician must select data features to highlight before constructing the graph. A single graph cannot always reveal everything about the data. For example, Figure 1 shows data from a study of neural firings in a cat's retina (Levine et al., 1987).. The cat figure parallels Cleveland's Figure 1. The lower graph in the cat figure uses Cleveland's median absolute slopes procedure to set the physical scale optimally for local segments in the plot. As Cleveland has noted, however, the slopes procedure does not specify which frequency to highlight in the series. The upper cat plot is scaled to reveal a low frequency (approximately 2.5 second) wave in the firings. This frequency component was caused by the cat's respiration, which affected blood oxygen and moderated the firings.

Figure 2 shows how information can be emphasized by the addition of a feature. Cleveland's Figure 2 contains a dot plot of agricultural data. Adding connecting lines highlights two crossovers in the Grand Rapids and University Farm plots. This highlighting can interfere with the perception of other aspects of the graph, however.

A well constructed graph from a formal point of view may nevertheless be misperceived due to cognitive processes in the later stages. Let's review how these come into play.

2. The retinal image. The initial image of the perceived graph is on the retina. This image differs in important ways from the physical image because lighting, viewing position, and other factors can create different retinal images of the same graph. Black and white graphs with high contrast are often preferable to ornate colored ones because these produce more constant retinal images under different lighting and viewing conditions.

3. The visual cortex. Retinal images are transformed in the visual pathway by decomposing them into features such as orientation and texture. Cleveland's discussion about core-cue symbols is based on this stage of visual processing. The operations at this level are highly parallel and organized to extract spatial frequency, orientation, and other features needed to construct complex visual scenes. I believe Cleveland is appropriate in grounding his discussion in the experimental literature on texture perception and discrimination. Cleveland's selection of core-cue symbols is based on studies of *pairwise* discriminability, however. There are likely configural effects when more than two of these symbols are used in a plot. A global recommendation may not apply in all cases.

While we're in the business of recommending symbols, however, I propose an alternative set. The symbols Cleveland offers are discriminable mainly because they stimulate different sets of feature detectors in the visual cortex, namely verticals, horizontals, and diagonals. Why not use a set based on the elements which stimulate these feature detectors more exclusively? The set 'o','|','-','x' is used in Figure 3. Compare the results with Cleveland's Figure 5. This set, while not shown in Cleveland's Table 1, performs as well as any in texture experiments.

Incidentally, some of Cleveland's results in earlier studies are due to the operation of vertical, horizontal, and diagonal feature detectors in the visual cortex. Lines near verticals, horizontals, and 45 degree diagonals tend to be resolved toward these canonical positions. This is one reason there are visual biases near these orientations. These biases become even more pronounced as information is stored in long term memory.

4. Integration and transformation through schemas. Making a judgment about a graph involves integrating the features detected in the visual cortex by making use of a short term memory store -- sometimes verbal, sometimes iconic -- and schemas residing in long term memory. While the visual cortical operations are highly parallel, the operations at this stage are both parallel and serial. Short term memory allows temporary (less than half a minute) storage of information in order to perform serial operations. For example, what Cleveland calls "table look-up" is a process of short term memory access which occurs at this stage. The viewer stores perceived information temporarily in order to construct higher-order comparisons such as scale references. Because of limitations in this store, only a few (five to ten) distinct pieces of information can be stored simultaneously. This is why higher order interaction plots are difficult, if not impossible, to interpret. The number of comparisons required to understand interactions increase exponentially. Cleveland's gridding task is a good example in which processing at this stage depends on props (such as grid lines) that help in the temporary storage of information for subsequent use in higher order comparisons such as the ranking of minima among curves.

Information is stored and accessed in short term memory via schemas available in long term memory. This is where some fascinating biases enter into the judgment of statistical graphs. For example, Barbara Tversky and others have found that features in maps can distort Euclidean distance judgments. Two towns divided by a river, for example, are judged to be farther apart than two towns separated by the same distance on uniform terrain. These distortions have nothing to do with the visual illusions which distort area, angle, and hue perceptions which Cleveland and others have demonstrated in statistical maps and other graphs. They have more to do with what Amos Tversky has called "framing," in which schemas assembled from a lifetime of experiences serve as templates for higher order judgments. Rivers are difficult to cross. Bar charts look like stacks of building blocks, so they have bases on the ground (at zero).

Many of the arguments concerning realism in graphs (e.g. Becker and Cleveland, 1991) hinge on the effects of "meaning" on the perception of quantity. Realism is not an unalloyed blessing in graphics because it can invoke schemas which are useful for decoding real world images but wholly inappropriate for quantitative graphical displays. Chernoff argued this point for perception of real vs cartoon faces and the psychologists Haber and Biederman have done the same for real versus artificial scenes.

In conclusion, Cleveland has provided some useful guidelines for selecting display methods based on controlled experimentation. By placing Cleveland's model in the context of the more general information processing model favored by most psychologists today, we can help to understand *how* and *why* distortions occur in the perception of graphs.

References

Becker, R., and Cleveland, W.S. (1991). Take a broader view of scientific vixualization. *Pixel*, *2*, 42-44.

Cleveland, W.S. (1985). *The Elements of Graphing Data*. Monterey, CA: Wadsworth.

Kosslyn, S.M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3*, 185-226.

Levine, M.W., Frishman, L.J., and Enroth-Cugell, C. (1987). Interactions between the rod and the cone pathways in the cat retina. *Vision Research, 27*, 1093-1104.

Pinker, S. (1990). A theory of graph comprehension. In R. Friedle, (Ed.), Artificial Intelligence and the Future of Testing. Norwood, NJ: Ablex, 73-126.

Simkin, D., and Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association, 82*, 454-465.

Spence, I., and Lewandowsky, S. (1990). Graphical percption. In *Modern Methods of Data Analysis*. Beverly Hills, CA: Sage Publications.

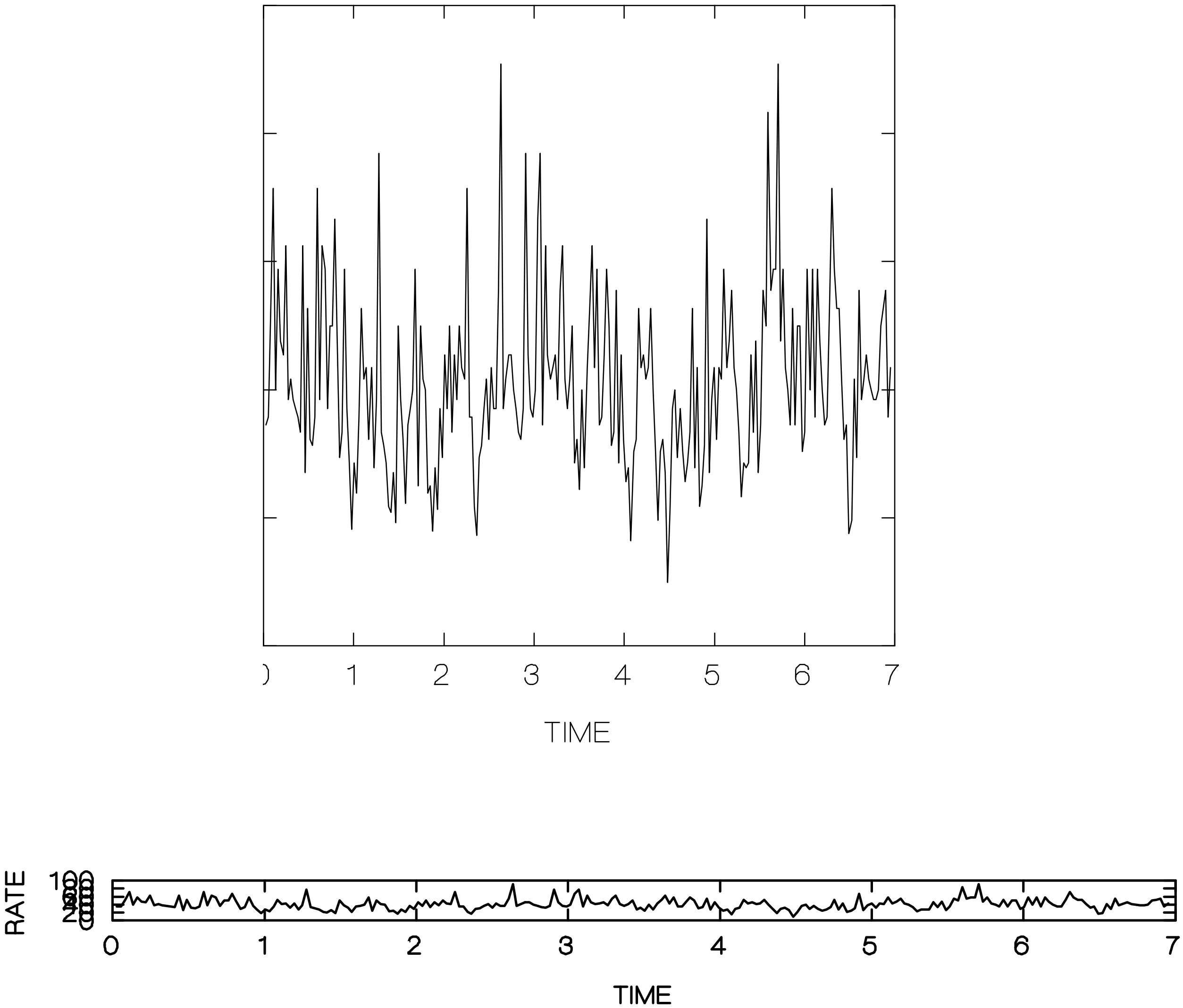

Figure 1: Neural firing data from Levine et al.(1987)

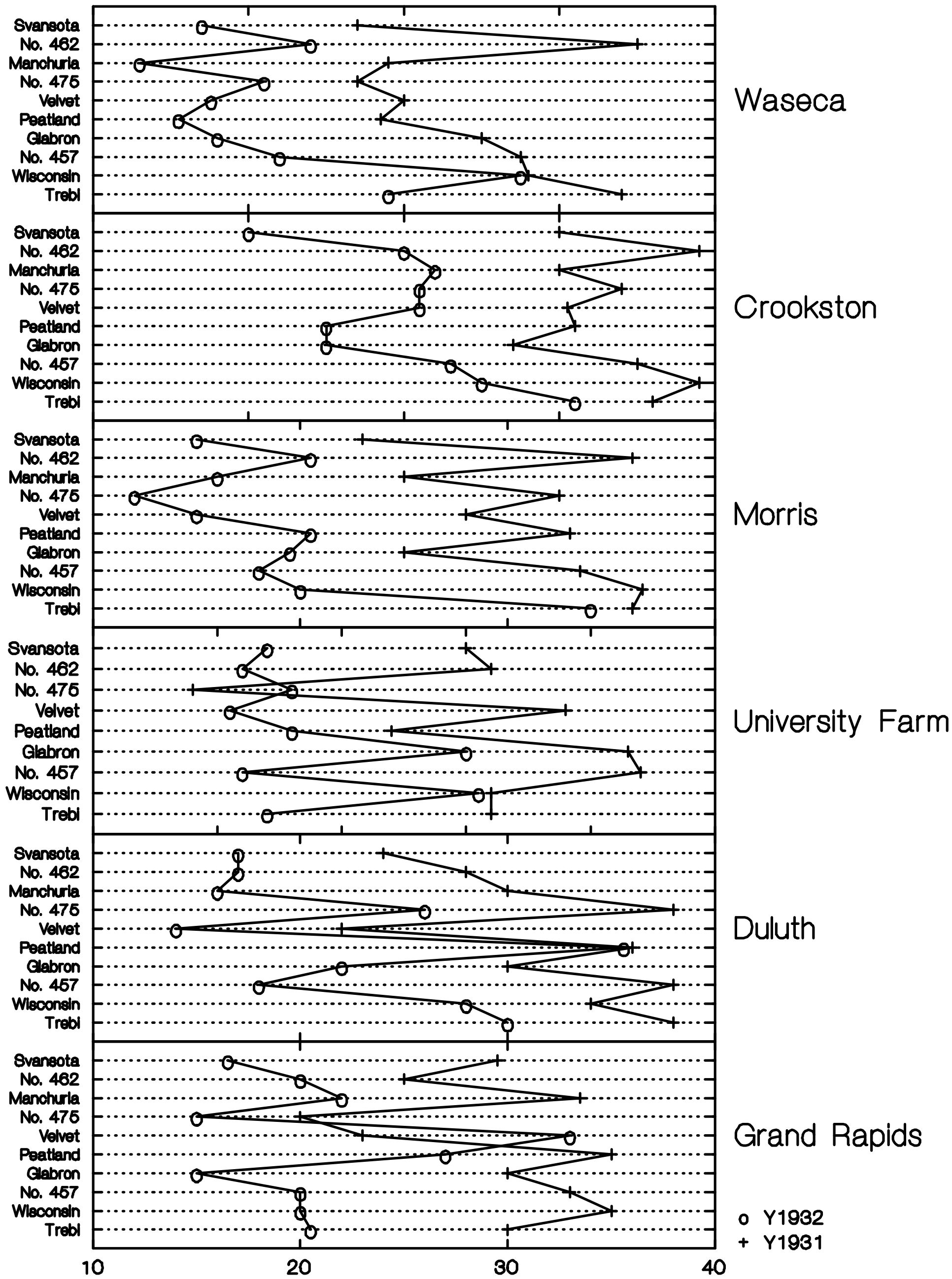

Figure 2: Barley yields with line enhancement

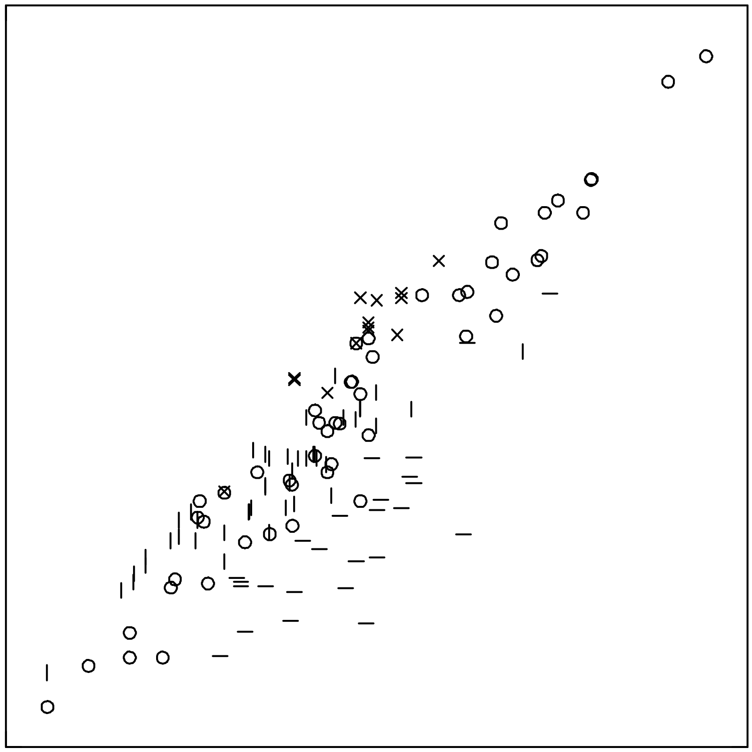

Figure 3: Core-cue symbols