cially welcome are the emphases on clear thinking as the basis of good writing, on repeated revision, and on writing for the ease of the reader. In this last regard, Pechenik advises writers, "Minimize turbulence. Always remind the reader of what has come before, and help the reader anticipate what is coming next" (p. 197).

Although few teachers of college biology courses might choose to assign students an entire book on writing, most faculty members whose biology teaching includes writing assignments can productively draw on this book, both for giving students guidance and for evaluating students' work. In addition, the book could serve well as a text in writing courses for biology students. I recommend it highly.

BARBARA GASTEL
Department of Journalism
Texas A&M University
College Station, TX 77843-4111

RIGHT BETWEEN THE EYES

Visualizing Data. William S. Cleveland. Hobart Press, Summit, NJ, 1993. 360 pp., illus. $40.00 (ISBN 0-9634884-0-6 cloth).

Clear and convincing communication of quantitative information is mandatory in scientific papers and presentations, and in classroom lectures and discussions. All of us surely can recall blank stares from our students as we scribbled graphs on the blackboard, comments from puzzled reviewers about confusing figures, or ineffective slides and overheads that left our audience unable to grasp the main points of a seminar. Several books published in the last decade have illustrated good design principles for data presentation (e.g., Cleveland 1985, Tufte 1983, 1990). Most of the graphic types recommended by these authors are now available in many commercial statistical packages, although graphics that maximize "interocular (between the eyes) impact" (Tukey 1993) still are used rarely by biologists (Ellison 1993). Concomitant with this renaissance in techniques for graphical display, William S. Cleveland and his colleagues at AT&T Bell Laboratories have been pioneering the use of interactive graphical methods for data analysis (Becker et al. 1988). These two lines of inquiry are brought together beautifully, both intellectually and aesthetically, in Cleveland's newest book, *Visualizing Data.*

The central theme of this book is that "[V]isualization is critical to data analysis. It provides a front line of attack, revealing intricate structure in data that cannot be absorbed in any other way. We discover unimagined effects, and we challenge imagined ones" (p. 1). Through detailed analysis of datasets ranging from agronomic to intergalactic, Cleveland illustrates how the two components of data visualization—graphing data and fitting mathematical functions to the data—can illuminate information and reveal unexpected patterns.

Cleveland's goal in writing this book is to effect a paradigm shift in the way that scientists work with data. We are all familiar with classical probabilistic inference in data analysis, the modern foundations of which are attributable to R. A. Fisher. A random perusal of scientific journals, however, would reveal a slavish attention to what Cleveland considers rote data analysis: carrying out probabilistic analyses using canned statistical packages, usually with little or no checking of assumptions, and where P-values take primacy over meaningful interpretation of the data based on knowledge of the system under study. Cleveland proposes data visualization as an alternative paradigm for guiding data analysis. Visualization "stresses a penetrating look at the structure of data," where inferences drawn from visualization are guided by detailed knowledge of the subject under study. Although Cleveland illustrates many examples where visualization replaces probabilistic inference in data analysis, he presents other examples where it is used principally to check the assumptions of classical statistical analysis. The latter is particularly important, because the most commonly used statistical techniques in the biological literature—parametric regression and analysis of variance (ANOVA)—depend strongly on assumptions about the structure of the population from which the data were derived.

Four classes of datasets are used to illustrate data visualization tools and techniques: univariate, bivariate, trivariate, and multivariate data. The bulk of these tools are presented in the context of univariate data (Chapter 2), where measurements are of a single quantitative variable that may be broken up into discrete categorical groups. Univariate data often are visualized with histograms, but Cleveland illustrates clearly and effectively that this old favorite is a poor choice for illustrating these data when compared with alternatives such as quantile plots, quantile-quantile plots, mean difference plots, box plots, dot plots, and normal probability plots. These plots also are used to examine fitted functions, their residuals, relationships between model estimates and residuals, and effects of data transformation. In these contexts, Cleveland introduces residual-fit spread plots and spread-location plots. The power of these techniques, which can be summarized as fitting sample estimators (e.g., means) to the data, subtracting them to yield residuals, and then graphing simultaneously the fit, the residuals, and their relationship, is highlighted by the complete absence of probabilistic inference anywhere in the second chapter. Yet, I came away from this chapter with a clearer understanding and appreciation of the power of univariate analysis than I have ever gotten from a more standard statistics text.

Bivariate (paired measurements of two quantitative variables) and higher-order data require several additional visualization tools, which are developed in Chapters 3 and 4. The most important of these tools is robust curve-fitting. Classically, we have used linear least-squares regression to fit lines to bivariate data. Cleveland uses *loess*, for *lo*cal re*gress*ion, to illustrate the limitations of traditional regression analysis. He introduces the bisquare, a robust estimation method for dealing with ill-behaved datasets, and he applies bisquare to both classical regression and loess. Analysis of bivariate data proceeds through iterative curve fit-

ting and residual analysis until underlying patterns are isolated. Two graphical display techniques to better view and present analytical results are emphasized: curve banking, where the aspect ratio of a graph (its height divided by its width) is adjusted to optimize perception of relationships indicated by fitted curves, and jittering, where a small amount of random noise is added to the data before graphing so that overlapping points are revealed. Dot plots and box plots are used to illustrate classified quantitative data, such as might be analyzed traditionally using one-way ANOVA.

Again, Cleveland does not resort to probabilistic inference, preferring instead to let the data tell the story. For example, in Chapter 3, Cleveland compares the results of his analysis with those of Lia et al. (1987), who reported a linear relationship between retinal area in cats and their ratio of central ganglion cell density to peripheral ganglion cell density (C/P). Cleveland quotes Lia et al. (1987): "[T]he linear relationship between the C/P ratio and retinal area is highly significant (slope = 0.107 ± 0.010; $P < 0.001$)" as an example of "the ritual of science," where the quotation of P-values is "a sprinkling of the holy waters in an effort to sanctify the data analysis and turn consumers of the results into true believers" (p. 177). However, Cleveland shows that appropriate use of visualization tools reveals that not only does the proposed line not fit the data well, but also that the data did not conform to necessary assumptions of linear regression. In fact, a quadratic polynomial relationship fitted the data significantly better.

We are limited by the dimensionality of paper to two-dimensional representations of trivariate and higher-order data. Cleveland shows that coplots and scatterplot matrices are excellent methods to work with higher-order data, especially when combined with the techniques introduced for univariate and bivariate visualization. He also shows how color and shading can be used to illuminate patterns in multivariate data, as opposed to the more common use of colors: to dazzle an audience with day-glo. In studies where multiple factors are examined simultaneously, as in multiway factorial designs, these techniques for display and analysis would be invaluable.

All of these visualization techniques are used most effectively when they can be manipulated directly by the user in real time. As statistical software for visualization has become more available (e.g., Becker et al. 1988) and as computer hardware has become faster, graphics that can be directly manipulated in real time have become a reality. Although it is difficult to illustrate effectively interactive direct manipulation graphics on the printed page, Cleveland has communicated the ideas and richness of these methods in a beautifully produced, well-written book, which Hobart Press has published at exceptionally low cost. In addition, all of the datasets used in the book are electronically archived in statlib, and are retrievable as either ASCII or S datasets over the Internet.[1]

Because it is impossible in a short, unillustrated review to do justice to the importance and utility of *Visualizing Data*, I recommend strongly that you read this book. Anyone with a familiarity with basic statistical techniques and least-squares methods of fitting regression lines to data should have no trouble with the material presented, and access to the datasets means that one could work through all of the examples in the book with virtually any statistical package. *Visualizing Data* should be required reading for every scientist and always should be kept in easy reach.

AARON M. ELLISON
Department of Biological Sciences
Mount Holyoke College
South Hadley, MA 01075

References cited

Becker, R. A., J. M. Chambers, and A. R. Wilks. 1988. *The New S Language*. Wadsworth & Brooks/Cole Advanced Books and Software, Pacific Grove, CA.

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

Ellison, A. M. 1993. Exploratory data analysis and graphic display. Pages 14–45 in S. M. Scheiner and J. Gurevitch, eds. *Design and Analysis of Ecological Experiments*. Chapman and Hall, New York.

Lia, B., R. W. Williams, and L. M. Chalupa. 1987. Formation of retinal ganglion cell topography during prenatal development. *Science* 236: 848–851.

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

_____. 1990. *Envisioning Information*. Graphics Press, Cheshire, CT.

Tukey, J. W. 1993. Graphic comparisons of several linked aspects: alternatives and suggested principles (with discussion). *Journal of Computational and Graphical Statistics* 2: 1–48.

[1]statlib is accessible via E-mail to statlib@lib.stat.cmu.edu (send the command SEND INDEX to get started); via ftp to lib.stat.cmu.edu (128.2.241.142) (login as userid: statlib and send userid as password); or via gopher to the University of Minnesota.

A NATIONAL APPROACH TO DOING SCIENCE

Styles of Scientific Thought: The German Genetics Community 1900–1933. Jonathan Harwood. The University of Chicago Press, Chicago, IL, 1993. 423 pp., illus. $65.00 (ISBN 0-226-31881-8 cloth), $23.95 (ISBN 0-226-31882-6 paper).

Assigning national traits to any group is risky and usually is based on polemics and anecdotes rather than on facts. Jonathan Harwood has done a remarkable job in his attempt to discover national differences between German and American workers in their approach to genetic research. By choosing a period during which international exchanges were still infrequent (and the field young), the survey is of necessity restricted to a small sample of more or less prominent scientists, a caveat to keep in mind when interpreting some of the data. The book, by the author's own admission (p. xvii), is almost Germanic in detail (more than 900 references) and thus should satisfy most historians of science. It may also be of interest to the premolecular generation of geneticists; most of the German scientists (and some Americans) mentioned cannot be found in modern genetics textbooks, R. Goldschmidt and C. Stern perhaps being the exceptions.

Harwood finds a major differ-