

WILLIAM S. CLEVELAND

**Discussion and comments. Approche graphique
en analyse des données**

Journal de la société française de statistique, tome 141, n° 4 (2000),
p. 43-44

<http://www.numdam.org/item?id=JSFS_2000__141_4_43_0>

© Société française de statistique, 2000, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

DISCUSSION AND COMMENTS

Approche graphique en analyse des données

William S. CLEVELAND¹

Jean-Paul Valois has provided an exceptionally informed overview of fundamental currents of thinking about graphical data display. It goes well beyond a mere encyclopedia of ideas, and synthesizes new viewpoints in an interesting way. The following section titles correspond to those of his article.

SOME HISTORICAL MILESTONES

John Tukey had a profound effect on data display, partly by inventing compelling methods, such as boxplots, which have become standards, and partly by isolating and illustrating the role graphs play in the analysis of data. His influence enlisted many in the study of data display in the 1970s and 1980s, particularly statisticians, reversing the general lack of interest in developing new graphical methods.

Tukey had help, though — the computer. Electronic computation took a task that required a graphic designer for its ultimate execution as a communication device, and put it in the hands of the statistician. This replaced manuals reciting standards for graphic design, prepared by those who made pictures but did not analyze data, with writings about new ideas for display, prepared by those who faced the task of analyzing data. This greatly increased the invention of new methods.

THE ROLE OF GRAPHS AND CHARTS

When we study a graph, and draw conclusions about the data, we are building a model for the data. The model might not be described by explicit mathematical structures and explicit probability distributions. Instead, we often derive a fuzzy mathematical model, such as “ Y appears to be an increasing convex function of X plus variation that does not depend on X .” Often the model is incomplete, not describing all of the variation in the data. Graphs are powerful tools for such model building, either explicit mathematical models or fuzzy models, because graphs provide us with an

1. Bell Labs, Murray Hill; e-mail : wsc@bell-labs.com

DISCUSSION AND COMMENTS

ability to compare an extremely wide range of alternative models, far wider than we could consider by formulating an explicit large super-model, and using formal hypothesis tests to pick a sub-model.

TYPOGRAPHY

A wholly new aspect of data, the ubiquity of very large databases, together with vast increases in the capabilities of computer graphics, have changed fundamentally both the needs and the possibilities for data display. An effective way to display a very large database is to expand the virtual screen real estate by allowing the graph to cover multiple pages. Today, we can routinely construct displays with hundreds of pages and study them with a document viewer. And each page can consist of tens of panels so that altogether we can have thousands of sub-displays. The thought of so many panels and pages might seem intimidating at first, but with the structure of trellis (lattice) display, which organizes the components of the display, and a well-designed viewer, much can be learned. Multi-page display is typically far more effective for very large databases than trying to force an immense amount of information into a single page.

THE IMPACT OF A GRAPH

Why do certain informative patterns emerge in some displays and not others? More study of graphical perception would surely lead to more insight and better methods. But in carrying out such study, we are likely better to study the elementary particles first, and then later, the large molecules. A whole graph is an immensely large molecule because the mental processing we employ to reach conclusions about the data is so intricate and poorly understood; it involves an elaborate set of perceptual and cognitive tasks that interact in a complex way. In contrast, the judgment of the orientations of two or more line segments, which we carry out to visually decode slope, is more akin to an elementary particle.

An acute intuition about issues of graphical perception is a prerequisite to effective study. And acute intuition is most likely to arise from extensive empirical experimentation with many sets of data, trying alternative methods of display. The intuition as a theory, even a vague undeveloped one, is vital as a guide to determine important aspects to study, to hypothesize, and to design informative experiments about the hypotheses. As in many fields, good experiments arise from good theory.