The Collected Works of John W. Tukey (W. S. Cleveland, ed.) **5**. Wadsworth, Pacific Grove, Calif.

TUKEY, J. W. and TUKEY, P. A. (1985). Computer graphics and exploratory data analysis: An introduction. *Proc. Sixth Ann. Conf. and Exposition of the National Computer Graphics Assoc.* 773–785.

TUKEY, J. W. and WILK, M. B. (1965). Data analysis and statistics: Techniques and approaches. In *Proc. Symp. Information Processing in Sight Sensory Systems* (P. W. Nye, ed.) 7–27. California Institute of Technology, Pasadena, Calif.

YULE, G. U. (1927). On the method of investigating periodicity in disturbed series, with special reference to Wolfer's sunspot numbers. *Philos. Trans. Roy. Soc. London Ser. A* **226** 267–298.

Comment

John W. Tukey

The overall impact of this paper is both substantial and sound. Thus my comments will have to focus on recommended changes in flavoring or on possibilities for the future.

1. THE MUD CAN BE DULL

The Murray Hill tradition in data analysis has long included aspects of "plant your feet firmly on the ground, even if they do sink deep in the mud." The limitation of scatterplot matrices to original coordinates is a case in point. The discussion of (brain weight)/(body weight)$^{2/3}$ in Section 2.1 is another case in point. A scatter of "log brain weight MINUS ⅔ log body weight" against log body weight would be a useful supplement to the scatter of Figure 2, in part because it would offer an expanded vertical scale. In Figure 11, where "abrasion is stated to be the intended response," an additional row and an additional column for "abrasion loss residual" and "tensile strength" would greatly clear up the situation—perhaps leaving brushing the task of finding still subtler behavior.

2. HIGHLIGHTING MAY BE INESCAPABLE—BUT IS STILL INADEQUATE

Paper representations of screens with highlighted points are rather weak and wan—and highlighted screens may be somewhat so. Particularly for paper versions, we ought to further enhance the contrast between emphasized and background points. Two easy ways to do this are: a) median +'s or x's for emphasis, with dimensions at least 3 times those of background circles, or b) filled circles for emphasis and little dots for background. This sort of improvement is needed for alternagraphic emphasis as well as for brushed emphasis. (Compare with the last paragraph of Section 2.1 in Becker, Cleveland and Wilks.)

3. DO PANELS MAKE UP A TABLE OR A GRAPH?

To Becker, Cleveland and Wilks, the answer seems to be "clearly, a table!" because they number rows of panels from above down (and put the vertical coordinate first!). For some of us, the answer seems to be "clearly, a graph" so we would number rows of panels from below upward, and put the horizontal coordinate first.

Whichever side you take—if one is to write in text about panel numbers, panel rows and panel columns, in the pictures, they should be labeled clearly enough (e.g. (1, −) and (−, 3) *or* (−, 1) and (3, −)) so it would be really hard for the reader/viewer to miss the point.

If you do adhere to the graphic paradigm, the distinctive diagonal of your scatter-plot matrix will run NE-SW and not NW-SE.

4. RECTANGLES, ANYONE?

It would have been helpful if the account of brushing had said—if it is true, as I think—that brushes are rectangular because rectangles can be computed faster. How much faster? Are we near the present boundaries when we include brushing? Or could we afford other brush shapes?

5. COGNOSTICS, ANYONE?

The paragraph in Section 2.4 on the scatterplot matrix assumes that scrolling is our only remedy when p is too large. Another approach would be to use cognostics (e.g., Tukey and Tukey, 1985) to help us rearrange our variables so that the initial view shows the most interesting k of them. Instead of simple scrolling, then, we might hold the first $k - 2$ (or $k - 3$, or $k - 1$) fixed and scroll the other 2 (or 3 or 1).

John W. Tukey is Senior Research Statistician and Donner Professor of Science, Emeritus, at Princeton University. His mailing address is 408 Fine Hall, Washington Road, Princeton, New Jersey 08544.

6. JITTERING OR TEXTURING?

Many of the jitterings suggested by Becker, Cleveland and Wilks are quite effective. But the jittering of the categorical variable in Figures 16 and 17 can be appreciably improved. A form of limited randomized (Tukey, Tukey and Veitch, in preparation) called texturing has appreciable advantages.

7. A GOOD POINT, BUT NOT TO BE OVERWORKED

The second paragraph of Section 2.5 in Becker, Cleveland and Wilks opens with "we often judge the slopes of line segments to determine the rate of change." The rescaling in their Figure 21 shows how we can do fairly well while preserving the original coordinates. If slopes are the main issue, however, using the full rectangle for two panels, one above another, showing year to year *changes* in sunspot number in one and sunspot numbers *themselves* in the other, would give more visual impact?

8. WHICH AXES?

The discussion under "Control" in Section 2.6 is too three-dimensional for adequate insight and visualization. Axes of rotations in k dimensions are $(k-2)$-dimensional manifolds. As a consequence rotations are better described as "from one axis toward another axis" (really "from one coordinate toward another coordinate") with a tacit understanding that $k-2$ orthogonal coordinates remain unchanged.

9. PERSPECTIVE

I was very pleased that "Enhancements" (in Section 2.6) did not argue for the use of calculated perspective, something whose utility I doubt. I am, however, confused about "the rules of perspective" in paragraph 2 of this subsection.

10. LIMITATIONS

The discussion in this subsection of Section 2.6 could well be taken further. We can easily judge near-linearity and direction of curvature for the two front variables. But no three-dimensional scheme, including rotation, lets us judge near-linearity of curves that involve the back variable. Not enough thought has been given as to how far we can avoid this difficulty—and to how to do this.

It would have helped also if we had been reminded that there are really three clusters in this iris data, and that projection pursuit comes close to resolving the two that naive scatterplot matrices do not.

11. DREIBEINS

The second paragraph of Section 2.6 on "advanced strategies" regrets the loss of "feet on the ground" when we rotate coordinates. It does notice that keeping track of the linear transformation can help us "to infer meaning." It omits, however, the visual presentation of a "dreibein," consisting of labeled projections of all original coordinate axes, emanating from a common point. This does a lot to explain the present coordinate system.

We need to think much harder about how to express this sort of information visually.

More attention needs to be given to the combination of multiple views and rotation. (The last paragraph of Section 2.6 discusses the combination of brushing and rotation—this is good, but the triple combination can be even better.) Two or more views of both point cloud and dreibein may be of considerable value. Doing this may make us want to be able to modify two different rotations (exactly or nearly) simultaneously.

12. SPECIAL ARITHMETIC

Becker, Cleveland and Wilks wisely emphasize the usefulness of integer arithmetic. The overall process of data analysis would probably benefit from the availability of middle decimal point arithmetic, in which much less rescaling would be needed.

13. SCALING FOR DATA DISPLAY

The discussion of "precomputation" assumes scaling in terms of "max MINUS min." Presumably the rescaling should include a translation to keep the location—as well as the separation—of "max" and "min" constant. My own preference would be to fix upper and lower eighths of the frequency distribution, rather than max and min.

14. SCATTERPLOT MATRICES

In addition to the need for numbering rows and columns (see Point 3 above), the scatterplot matrices use of numbers in the corners of the "empty" diagonal cells to specify the range covered for each variable could have been mentioned in the text. More important would be a more effective visual display of the scales involved. This could be placed on the margins of the whole matrix or we could put the information in the "empty" cells (in each of a variety of ways).

15. MULTIPLE SCATTERPLOT MATRICES

Becker, Cleveland and Wilks emphasizes scatters of raw variables. Without eliminating this display from use, we could supplement it by such displays of other scatter matrices as: a) simple residual scatters, with

$y_{.x}$ scattered against x where y = rows and x = columns, b) partial residual scatters, with $y_{.\mathrm{REST}}$ (where REST is all but y and x) scattered against x and c) full residual scatters, with $y_{.\mathrm{ALL}}$ (where ALL is all but y) scattered against x.

16. SUMMARY

The paper of Becker, Cleveland and Wilks has gone a long way from graphic archeology ("you can see it, if you know how to look!") toward graphic impact ("you can't miss it"). But we need to go further (Points 1, 2, 7, 11 and 15). The proposed styles of graphic presentation could have been improved in a few other cases (Points 3, 5, 6, 13 and 14). There are a variety of points where the text could have been clarified (Points 4, 8, 9, 10, 13 and 14). In general, however, one can only praise the paper of Becker, Cleveland and Wilks.

ACKNOWLEDGMENT

This research was supported by the Army Research Office under Contract DAAL03-86-K-0073.

Comment

Peter J. Huber

The authors deserve to be congratulated for a competent overview of an area of statistics that is becoming more important (and accessible) every year.

They try hard to give a sober, no-nonsense account of the current state of the art. There is a core of simple-minded, extremely useful techniques—foremost among them methods for identification and labeling. But there is also a halo of experimental techniques, and the cautious statement: "Far more experimentation is needed with these advanced strategies" (Section 2.6) ought to be translated into plain English as: "We could not make sense out of those strategies, but perhaps somebody else will." The lunatic fringe techniques are useful to hatch new ideas, but only few of them will survive.

I believe that newcomers to high interaction graphics—the buzzword "dynamic" is semantically inaccurate, by the way—are still attracted mostly for the wrong reasons, namely by the video game glamor of fancy techniques. Reflection on the intrinsic, practical value of a technique comes only afterward, when the glamor has rubbed off. For example, when we gained direct, hands-on access to decent computer graphics in 1978, Stuetzle and I began to experiment first with the most exotic techniques—like interactively controlled sharpening (see Tukey and Tukey, 1981), combined with kinematic graphics in three dimensions. It took a long while for me to realize that I never would be able to interpret the curious shapes I saw in those sharpened scatterplots.

Peter J. Huber is Professor of Statistics, Harvard University, Cambridge, Massachusetts 02138.

The following are comments on points where I either disagree with the authors or would put the emphasis differently.

It is important to avoid gimmicks. Built-in side effects can become extremely annoying when a technique is used in a context not anticipated by its designer. For this reason rescaling after a deletion ought not be automatic (Section 2.2), but should require a separate user request.

In Section 2.4, I suspect that the authors' judgment has been colored by accidental features of their implementations (cf. the remarks of Huber, 1987, Section 4). I fail to understand why roping in a region by drawing a line around it should be intrinsically slower than, and inferior to, brushing the interior of that region. Roping is relatively elastic with regard to timing considerations, while the response to brushing can become unacceptably slow in the case of very large scatterplots.

Undeleting is trickier than the authors make it appear (end of Section 2.4). The problem is to *selectively* undelete a few points. It is aggravating if you have to undelete everything and then to start the deletion process from scratch. A more convenient solution, using alternagraphics, is due to Thoma: alternate with a view showing the deleted points only. If you delete a point in the alternative view, it gets visible in the original view, and vice versa. Incidentally, this is one of many examples where alternagraphic switching is better done by the user hitting a button, rather than by an automatic timer.

In Section 2.6, the authors say that "it is entirely reasonable to implement all three ... rotation-control methods ...." I would substitute "feasible" for