Comment: Deja View

Howard Wainer

I was distracted during my reading of this paper by a high-spirited giggle that wafted in over my shoulder. This was surprising because the door to my office was closed and I thought I was alone. As I turned to try to discover the source of this disturbance I caught a glimpse of the high spirit whose presence had been heralded earlier. It was that old schoolmaster H. G. Funkhouser, whose history of graphics, published half a century ago (1937), has provided a jumping-off point for modern chroniclers. In that treatise Funkhouser was forced to describe graphics that, for technical reasons, could not be reproduced. In doing so he frequently (but regretfully) spent the thousand or so words that each picture required. Mark Twain described women's use of profanity as "They know the words, but not the music." Similarly a verbal description of a graphic may contain the facts of the matter, but not the nuance.

The reasons for his posthumous return to my office became clearer. "*Le plus c'est change, le plus c'est le meme chose*" he mumbled. The current authors (Becker, Cleveland and Wilks) faced the same problem; trying to convey both the fact and nuance of dynamic displays within a static medium. That they did not fail is to their credit, but the result is a combination of 19th Century female profanity and the bipedal walking characteristics of Samuel Johnson's dog. They have done a good job of conveying the general idea and many of the details, but the magic is missing. The pale approximation of these exciting new methods that is conveyed on these pages seems to be a prime example of what Ron Thisted has called the "*Instamatic in Yosemite*" syndrome, in which the frustrated describer can only sputter "You had to be there." It may be that we are approaching the limits of what a static medium can effectively communicate in this area.

Over the years I have formed the opinion that the principal goal of our Bell Labs colleagues at scientific meetings is to foment frustration among the rest of us. Time and time again they report methods and procedures that require machinery, budgets and expertise among support personnel that outstrip what the rest of us can practically contemplate. Yet they blithely go on telling us how they spin and mask and brush, while we look at BMD output. I often laugh as a naive assistant professor enthusiastically responds to one of these all too familiar presentations.

> "What kind of equipment do I need to be able to do this? asked naive assistant professor.
> "We used a BLNFY-87 (pronounced 'Blinfree' = Bell Labs Not For You) in combination with an RB-1," responded Bell Labs colleagues.
> "Can other equipment be programmed to do this?" naive assistant professor hoped.
> "I imagine so, although we haven't tried," tempted Bell Labs colleagues.

What naive assistant professor didn't learn, and what Bell Labs colleagues never mentioned, was the secret ingredient in their work; the RB-1 {one Rick Becker}. Without one of these, the task is made either much more difficult, or impossible.

So much for my envious grousing. Let me get on with it. We should all be grateful to our Bell Labs colleagues for their path-breaking work in this area. They have done a lot of hard thinking to arrive at a set of methods that appear to be enormously helpful. In accomplishing this I assume that a great deal of time was spent in blind alleys pursuing what seemed in prospect like a good idea but, after a lot of work, turned out not to work very well. The polished product we read about here can be used and, perhaps in a less formal communication, we can find out about the less fruitful paths. I assume that anyone interested in pursuing this area will have the good sense to bend an elbow with one of these authors and find out what else they did.

Much to my surprise, I am less envious now than I have been in previous situations. The principal reason for this is that much of what they are describing is available to the rest of us now (or will be shortly) for less than the annual budget of Saudi Arabia.

MacSpin (Donoho, Donoho and Gasko, 1985) costs under $200 and runs on a MacIntosh. It allows one to view scatter plots three dimensions at a time with rotation around any of the three screen axes. It's facility for the identification of data points is, if anything, *better* than what Becker, Cleveland and Wilks describe. In addition to being able to go in either direction ("Where is '*dog*'?" or "What is this point?"), it can also provide three levels of detail about each data point. Automation is easy, and, through its judicious use, so too is alternagraphics. Software

Howard Wainer is Principal Research Scientist in the Division of Statistical and Psychometric Research, Educational Testing Service, Princeton, New Jersey 08541.

development is dynamic, especially in dynamic display, and so I expect that Andrew Donoho will almost surely have implemented even more tools by the time this commentary finds its way into print.

The DataDesk (Velleman and Velleman, 1986) is a statistical analysis program that also contains some rudimentary dynamic capability. It can produce spinning plots; indeed several windows with a three-dimensional plot in each of them (although only one spins at a time).

These two programs, used in concert, provide a powerful tool for the data analyst at a price that we all can manage. Moreover, being commercially available programs with wide usage means that they have been debugged in a way that more narrowly circulated experimental computer packages are not. The appearance of dynamic display capability in MacSpin and later in DataDesk portends well for the future. Who will buy a data analysis program without such accoutrement? Thus, other software developers will be forced to add these capacities to their programs and we will be the richer for it.

Let me conclude by expressing my gratitude to our Bell Labs colleagues in general, and to Becker, Cleveland and Wilks in particular, for their continuing research into data analytic tools that more fully utilize the computing power now available and the human information processing ability inbedded in our visual system. Their imagination and sweat has provided us with the knowledge of a battery of methods that every salt-worthy data analyst would want to have close at hand. Simultaneously, a second set of talented folks are working hard to make these tools available for the rest of us. To both groups I give my heartfelt thanks, and ask that they stop wasting their time reading this and get back to work—I have a data set that I've been looking at, and I think I'm missing something.

ADDITIONAL REFERENCES

FUNKHOUSER, H. G. (1937). Historical development of the graphical representation of statistical data. *Osiris* **3** 269–404.

VELLEMAN, P. and VELLEMAN, A. (1986). *The DataDesk Handbook.* Data Description, Inc., Ithaca, N.Y.

Comment

Edward R. Tufte

Even though we navigate daily through a perceptual world of three spatial dimensions and reason occasionally about still higher dimensional arenas with mathematical and statistical ease, the world portrayed by our information displays is caught up in the two-dimensional poverty of endless flatlands of paper and video screen. Escaping this flatland is the major task of envisioning information—for all the interesting worlds (imaginary, human, physical, biological) we seek to understand are inevitably and happily multivariate worlds. Not flatlands.

Such escapes grow more difficult as ties of data to the familiar spatial world weaken and as the number of data dimensions increases. But the history of information displays and statistical graphics—indeed the history of communication devices in general—is nothing but a progress of methods for enhancing the density, richness, efficiency, complexity and dimensionality of communication. Methods for escaping flatland include layering and separation, micro/macro readings, contours, perspective, narratives, multiplying of images, use of color and dynamic graphics (Tufte, 1983, 1988).

The visual display tasks involved in dynamic graphics for data analysis are very nearly identical with the flatland portrayal of any dynamic physical system. It is, after all, data moving.

For example, when Galileo first looked through his telescope in 1610, he was confronted with displaying the dynamics of sunspots. In "The Starry Messenger" and "Three Letters on Sunspots," Galileo reported his observations in a large collection of small multiples sequenced on time, recording complex data of moving sunspots on a rotating sun observed from an orbiting and rotating earth (Figure 1).

Through some 370 years of astronomical research, sunspot records have evolved into data-rich time series. The Maunder butterfly diagram records the distribution of sunspots in latitude only moving over time, sacrificing area for time (Maunder, 1904) (Figure 2).

The modern version partially recovers area in reporting an enormous volume of information (Figure 3).

Edward R. Tufte is Professor of Political Science and Statistics, Lecturer in Law, and Senior Critic in the Program in Graphic Design, Yale University, 3532 Yale Station, New Haven, Connecticut 06520.