climatic and faunal—on a large time chart. The relatively crisp picture contributed by the geophysical evidence was nicely corroborated by evidence from pollen and micromammal (rodents, rabbits and the like) data, both of which are excellent climatic indicators. But enthusiasm for adding a data set just because it existed quickly turned the chart into scatter of points, all but obscuring correlative events.

A greater selection of data is required, with emphasis given to faunal groups that contain many species and have tight environmental requirements, such as the bovids, which include antelopes, wildebeest, and their relatives. The bovid data from Africa already show bursts of speciation at the 2.4 and 0.9 million year events, and possibly around 5 million years too. The Lamont meeting has initiated the systematic assessment of other data sets that might match the quality of the bovids, the result of which will reveal just how synchronous the pulses of extinctions and speciations really are.

—ROGER LEWIN

References

1. E. S. Vrba, in *Living Fossils*, N. Eldredge and S. M. Stanley, Eds. (Springer-Verlag, New York, 1984); R. Lewin, *Science* **223**, 383 (1984).
2. R. A. Kerr, *Science* **224**, 141 (1984)
3. N. J. Shackleton *et al.*, *Nature (London)* **307**, 620.

The Proper Display of Data

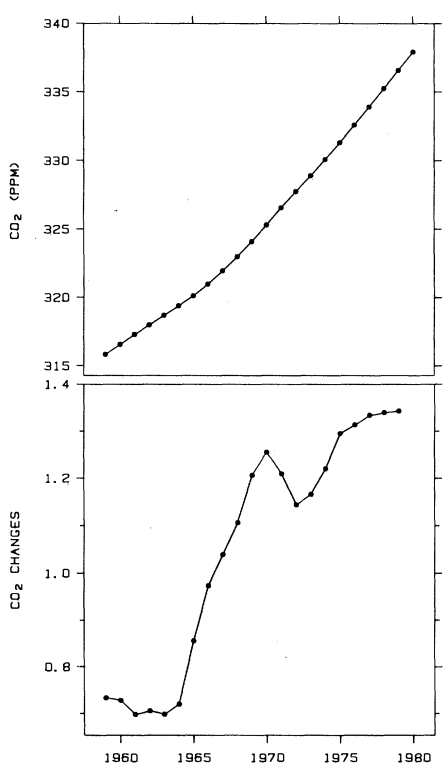
Two statisticians looked at graphs in scientific publications and they suggest means of improvement

A few years ago, William Cleveland of AT&T Bell Laboratories began looking through scientific journals to see how researchers use graphs. There had been a revolution in graphic methods in their own field of statistics as investigators began developing new ways of analyzing and presenting data and Cleveland thought that similar developments might have occurred in other areas of science. "I figured, 'Look, there's this incredible talent in science. Surely, there's all kinds of amazing techniques invented that those of us in statistics don't know about. Let's find out about them and bring them into our tool kit,' " he reasoned.

But, Cleveland remarks, "It didn't happen. I found just the opposite was true. I saw all kinds of errors and abuse [in the graphical display of data]."

In order to determine how common graphs are in scientific publications, Cleveland and Marylyn McGill of MEM Research in New Providence, New Jersey, surveyed 57 journals from 14 disciplines. Among these, the publication with the most graphs was the *Journal of Geophysical Research*, which devoted a third of its space to graphs. The median amount of space devoted to graphs was 6 percent and the journal with the fewest graphs in the group was the *Journal of Social Psychology*.

Then Cleveland and Robert McGill of AT&T Bell Laboratories chose to study *Science* in particular and looked through every issue of volume 207 (January to March 1980) carefully examining each of the 377 graphs there. They found generally familiar formats; furthermore, they report that 30 percent of the graphs contained at least one error. Less extensive analysis of other journals led them to conclude that the graph problem is widespread. But the tedious task of surveying journals had one beneficial result. It led Cleveland and McGill to ask what it is that makes some graphs better than others in displaying data and to devise guidelines that scientists can use to make their use of graphs far more effective. (Edward R. Tufte of Yale University did a similar analysis of graphs in the mass media and published his results in his book *The Visual Display of Quantitative Information* (Graphics Press, Santa Monica, California, 1983).



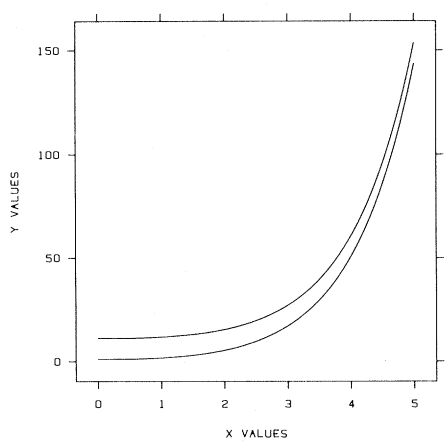
Slope is hard to judge

The visual impression from the top panel is that the rate of change of atmospheric CO_2 is constant from 1967 to 1980. But in the bottom panel, where the yearly changes are graphed, it can be seen that there is a dip in the rate of change around 1970.

After sifting through the journals, McGill and Cleveland began their research on how people visually decode the quantitative information in graphs. "The main thing we did was to sit and look at graphs and think very hard about what exactly are the elements that contain quantitative information," Cleveland explains. Then we got people and had them look at graphs, asking them questions that made them focus on very basic perceptual tasks, such as judging slopes of lines or areas." Although Cleveland and McGill tested subjects with a variety of backgrounds—scientists, high school students, university students, housewives—they learned that the different groups performed more or less the same.

Among the most difficult graphs to read are those that involve estimations of area. For example, a graph may show the amount of coal mined in different areas of England, with circles at each spot on the map of England where coal is mined and with the circles drawn so that their areas are proportional to the amount of coal mined at each place. Such a graph is very difficult to read, Cleveland and McGill find, and although scientists do not often use graphs that require estimates of area, such graphs are common in newspapers and magazines.

Slope is also difficult to judge, but scientists frequently require their readers to estimate it. Any time that a variable is graphed and readers need to know its rate of change, they must estimate



Vertical distances can be deceptive

The vertical distances between the curves are equal, although most people would guess that the curves become closer together in going from left to right.

slope. One example is a graph of smoothed yearly average carbon dioxide measurements from Mauna Loa, Hawaii (see illustration). The impression that this graph gives is that the carbon dioxide concentrations increase constantly from 1959 to 1965 and then start to increase more rapidly from 1967 to 1980. But when the yearly changes in carbon dioxide concentrations are graphed against time, it becomes clear that there is a drop in the rate of increase in the early 1970's.

Another difficult task is to judge the vertical distance between curves. "The curves look much closer together when they are steep," Cleveland explains (see illustration). But, he notes, "these things have remedies." All that is required is for researchers to take the aspect of the data that they want to convey and show it separately.

In addition to suggesting that scientists avoid asking readers to do such things as judge slopes or vertical distances between curves, Cleveland also suggests that they reconsider their use of error bars in their graphs to portray plus and minus one standard error of the statistic. A standard error of a statistic, Cleveland points out, "has value only insofar as it conveys information about confidence intervals. The standard error by itself conveys little." He notes that in the ideal case when the sample's distribution is normal and the sample size is not small, plus or minus one standard error of the mean is a only 66 percent confidence interval, which is not the most interesting interval. Researchers generally are interested in intervals with high levels of confidence, such as 95 or 99 percent. "Does anyone care about a 66 percent confidence interval? Are confidence intervals thought about at all when error bars are put on graphs?" Cleveland asks.

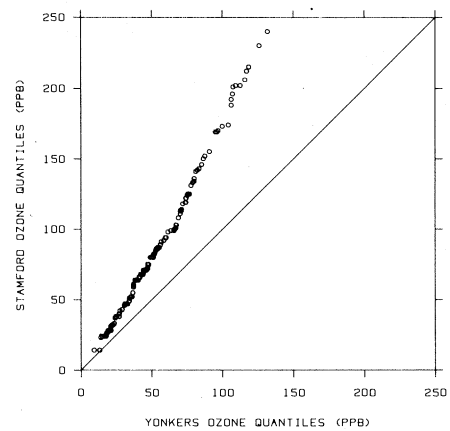
Cleveland speculates that the convention of putting one standard error bar on graphs arose "as a knee-jerk reaction" to the numerical convention for describing sample-to-sample variation. If investigators want to communicate such variation in situations where the data are normally distributed, it is reasonable to give the mean and one standard error, letting the reader calculate the confidence intervals. But when this convention is naïvely translated to graphs, Cleveland says, "we are locked into what is shown by the error bars. It is hard to visually multiply the bars by some constant to get a desired visual confidence interval."

A better way to express confidence intervals in graphs, Cleveland suggests, would be to draw error bars showing plus and minus 0.67 times the standard error—a 50 percent confidence interval—and plus and minus 1.96 times the standard error—a 95 percent confidence interval. But, he remarks, "the important thing is to show confidence intervals and not standard errors."

Still another way for scientists to improve their graphical analysis presentation of data would be for them to learn some of the new methods of displaying data that were invented by statisticians. "Graphical methods in statistics deserve to be much more widely disseminated," Cleveland says. For example, there is the "empirical quantile-quantile plot," devised by Martin Wilk of Statistics Canada and Ran Gnanadesikan of Bell Communications Research. The purpose of this plot is to compare distributions of two sets of measurements (see illustration).

One place the empirical quantile-quantile plot might be used is in a drug study in which one group of animals gets active drug and another group gets a placebo. Some aspect of the animals—say blood pressure—is measured. The graphical method permits a comparison of all the aspects of the distribution of the two sets of measurements in a way that not only is powerful but is, Cleveland stresses, "terribly simple."

The idea is to take the median of one of the groups of animals' blood pressures and plot it against the median blood pressure of the other group. Then plot the 75th percentile of one group against the 75th percentile of the other, then the 25th percentiles of one against the other and so on. In the end, the investigator will pick out a variety of different percentiles and, in each case, plot one group's value against that of the other group.



An empirical quantile-quantile plot

Since the plot goes through the origin and has slope 1.6, it can be concluded that each Stamford quantile is about 60 percent larger than the corresponding Yonkers quantile. Because the plot is not parallel to the line y = x, *an ordinary* t*-test could not be used to test significance.*

The way the resulting graph behaves can reveal a great deal about the data. If the points lie along the line $x = y$, then the two populations are identical. If they systematically depart from that line, additional information comes to light. For example, if the points lie along a straight line through the origin with slope 1.2, then the high values in one group are always 20 percent greater than the high values in the other group, the low values in that group are 20 percent greater and, in fact, the whole distribution of values in that group are 20 percent greater than in the other group.

Many statistical procedures, including the *t*-test, as it is standardly used, are valid only when the points of the empirical quantile-quantile plot are parallel to the line $y = x$. "But," says Cleveland, "the sobering thing is that many times scientists employ these tests when their data [if plotted in a quantile-quantile graph] are not parallel to that line."

The empirical quantile-quantile method has the two properties that Cleveland and McGill believe are the reasons for using graphs in the first place—it illustrates the behavior of the data and it provides information that can be used in deciding what sort of statistical analyses are appropriate. Not every graphical display can be that good, but, at least, Cleveland urges, scientists should not ask their audience to put up with inaccurate graphs or graphs whose data are nearly impossible to read.—**Gina Kolata**

Additional Reading

1. W. S. Cleveland and R. McGill, "Graphical perception: Theory, experiments, and the application to the development of graphical methods," *J. Am. Stat. Assoc.* (September 1984).
2. W. S. Cleveland, *The Elements of Graphing Data* (Wadsworth, Belmont, Calif., in press).