evils and the advances in scientific and engineering thinking are presented as linked, but how science thereby made itself alien to modern culture is not clear to me.

The third chapter is a polemic against "the corporate and military power structure of the United States" for misleading physicists into building the atomic bomb. The case, argued by others on both sides, is not so clear-cut. In this chapter, Schwartz ventures into economics and a bit of social welfare, but, again, how this explains the alienation of science from modern culture escapes me.

The fourth chapter concerns the development of molecular biology as a focused field in the decades on either side of 1946. Here the author has managed to skip over chemical gradients and Gibbs free energy by claiming that physics does not apply much to biology in his sentence, "The secret of life, so to speak, is revealed not through the equations of the laws of physics but through the depiction of a finite sequence of events." In contrast to Schwartz, I think the sequence of events has an energy drive; otherwise the sequence dies down. Later he claims that intelligence is not inherited, using the ridiculous argument that whoever is raised in Hungary learns Hungarian. Thus, his sentence "You cannot predict the language a child will speak from knowledge of its DNA" might be politically correct, but there is a big hole in the logic connecting this fact to the question of inherited intelligence.

In the fifth chapter, the author discusses particle physics in terms of opaque mathematics, unimaginative theory and very expensive equipment. The aim is not to bring physics to life for the reader, apparently, but rather to knock theoretical physics as currently sterile.

The preface says that the sixth chapter addresses the question of why the general public is afraid of science. The chapter certainly lacks much in the way of solutions to the author's concerns. He bewails this and that part of science and the general lack of connections between science and society, but fails to offer helpful advice. I am not sure who the book is aimed at—perhaps those who dream of an earlier and simpler world. So many of the arguments strike me as illogical and convoluted that I cannot imagine a serious scientist would approve.—*Lucian B. Platt, Geology, Bryn Mawr College*

Computer Software

Systat: Number Crunching on Cereal Statistics

Systat. Systat, Inc., 1800 Sherman Avenue, Evanston, IL 60201; 708-864-5670. Available for the Apple Macintosh and for computers running the MS-DOS operating system or Microsoft Windows. $895.

Statistics is often presented as the cod-liver oil of a scientific education. It is supposed to be good for you, but no one is expected to enjoy it. This reputation for disagreeable dullness could not be more unfortunate, for statistics addresses a central question of the entire scientific enterprise: How can we know what is true?

The Cosmic Background Explorer (COBE) satellite was launched in 1989 with the primary mission of searching for a slight anisotropy—a directional bias—in the microwave afterglow of the Big Bang. The satellite performed that mission brilliantly, but it could not return a simple yes-or-no answer. It sent back some six megabytes of data per day, which were subjected to an elaborate statistical examination. Only after three years of analysis could the COBE team announce that they had detected some cosmic hot spots and cool spots, which differ in temperature by a few millionths of a degree Kelvin. The conclusion of the experiment relied as much on statistical analysis as on aerospace engineering.

Today no one would undertake any serious statistical work without the aid of a computer. The machine offers more than just liberation from the drudgery of doing voluminous calculations. It has made possible a new approach to statistics, generally known by the slogan "exploratory data analysis." Traditional methods of analysis work on the oracle principle: The investigator asks a question—"How closely are these variables correlated?" "Is this difference in outcomes significant?"—and the oracle of statistics returns an answer. The new exploratory methods offer not only answers but also help in formulating the right questions. They emphasize graphics and other techniques for seeing patterns in data—patterns that can then be examined more closely by the conventional, oracular, tools.

Systat is one of several software packages that facilitate exploratory analysis. The program and the company were created a decade ago by Leland Wilkinson, who studied statistics under L. Rowell Huesmann at Yale University. For this review I tested the Macintosh edition of Systat, which is currently in version 5.2. The editions for MS-DOS and for Microsoft Windows are at version 5.0. Systat, Inc., also publishes a smaller and less-expensive program called Fastat as well as a student edition, Mystat, that is included in several statistics textbooks.

The program comes in a large box with four manuals (*Getting Started, Data, Statistics* and *Graphics*) and a fistful of disks. The manuals, which amount to some 1,800 pages, would be an impressive product even without the software. They endeavor to explain not only the Systat program but also the underlying techniques of statistical and graphical analysis. The *Graphics* manual is particularly noteworthy, offering sound advice on how to present data accurately or at least honestly.

For a test of Systat I chose a data set distributed by the American Statistical Association: a compilation of nutritional information on breakfast cereals. (The data set is available through StatLib, an Internet archive maintained at Carnegie Mellon University.) Importing the data into Systat would have been effortless but for one annoying awkwardness. Systat limits character variables to 12 letters, which means that "Oatmeal Raisin Crisp" must be abbreviated to something barely intelligible, such as OatRaisinCrp. Worse, when the 12-letter limit is exceeded inadvertently, the program's handling of the error is anything but graceful. I had to reimport the data several times before I got it right.

The Systat manuals advocate a particular approach to the initial examination of a data set, namely constructing a scatterplot matrix, or SPLOM. The matrix shows the pairwise dependencies of the variables in the data set. Figure 1 is a SPLOM for five of the variables describing the nutritional value of cereals. Each of the 25 panels is a separate graph; for example, the panel in the upper right corner records caloric value (on the vertical scale) as a function of sugar content (the horizontal scale). The graphs are tiny, but they are adequate for spotting patterns and correlations. For example, it appears that calories are positively correlated with all the other variables, and especially with carbo-

hydrates and sugars. This intuition is confirmed by a numerical analysis, which shows that the Pearson correlation coefficient between calories and the sum of carbohydrates and sugars is 0.789.

Some methods of interactive data analysis are difficult to illustrate in static diagrams on the printed page. Systat provides tools for "data brushing," or selecting and identifying specific data points in a graph. Clicking on a point in one panel of the SPLOM highlights the corresponding point in all the other panels and also brings the corresponding record to the top of the data-editor window. Such poking around in the data set turns up a few interesting facts, including a hint that two cereals—Puffed Rice and Puffed Wheat—may have a somewhat unusual status within this group. They lie in the lower left corner of almost all the plots; in other words, they are low-calorie, low-fat, low-sugar, low-everything cereals.

The analysis of the cereal data could be pursued in many directions. I decided to look for clusters—to see if the cereals fall naturally into groups or categories. Figure 2 shows the result of one such analysis for a subset of the cereals. I standardized the nutritional data, so that all the variables would have the same mean and standard deviation, then transposed the data matrix and calculated a new matrix in which similarities between cereals were expressed as Euclidean distances in the five-dimensional space formed by the five nutritional variables. Then a procedure called nonmetric multidimensional scaling, or MDS, created a two-dimensional map based on these pairwise distances. The map shows that most of the cereals form a fairly tight cluster, but there are four outliers: the two Puffed cereals and two kinds of bran.

Producing these analyses was not without difficulty or unpleasant incident. Although the data set is not very large (15 variables for each of 76 cereals), Systat repeatedly ran out of memory. The recommended memory partition for the Macintosh edition is 1.8 megabytes, but I found that even four megabytes was inadequate. What's more, when memory was exhausted, the program died a sudden death, allowing no opportunity to save work in progress. There were also lesser glitches, such as printing failures and spurious error messages.

Another area of dissatisfaction is the graphics format. In high-resolution mode, labels are restricted to a few hand-made fonts, which are much inferior to the Postscript fonts recognized by virtually every other Macintosh program. Overall graphic quality is probably adequate for publication in many journals, but it is not up to the standards of a magazine such as *American Scientist*.

In addition to these failings of usability and presentation, there is one serious omission from the statistical tools provided in Systat. In the past decade the computer has transformed statistics not only by fostering a more exploratory approach but also by making possible new analytic techniques that rely on massive, brute-force computations. The new techniques are known generically as resampling methods; the best-known example is the "bootstrap" procedure devised by Bradley Efron of Stanford University. Resampling methods would fit perfectly into Systat's statistical toolkit, but the current version of the program offers no direct support for them.

Even with its failings and limitations, however, Systat is impressive software with an equally impressive set of manuals. It raises statistics up from the unappetizing level of cod-liver oil, beyond All Bran and Shredded Wheat, perhaps to the exalted status of Cheerios or Froot Loops.—*Brian Hayes*

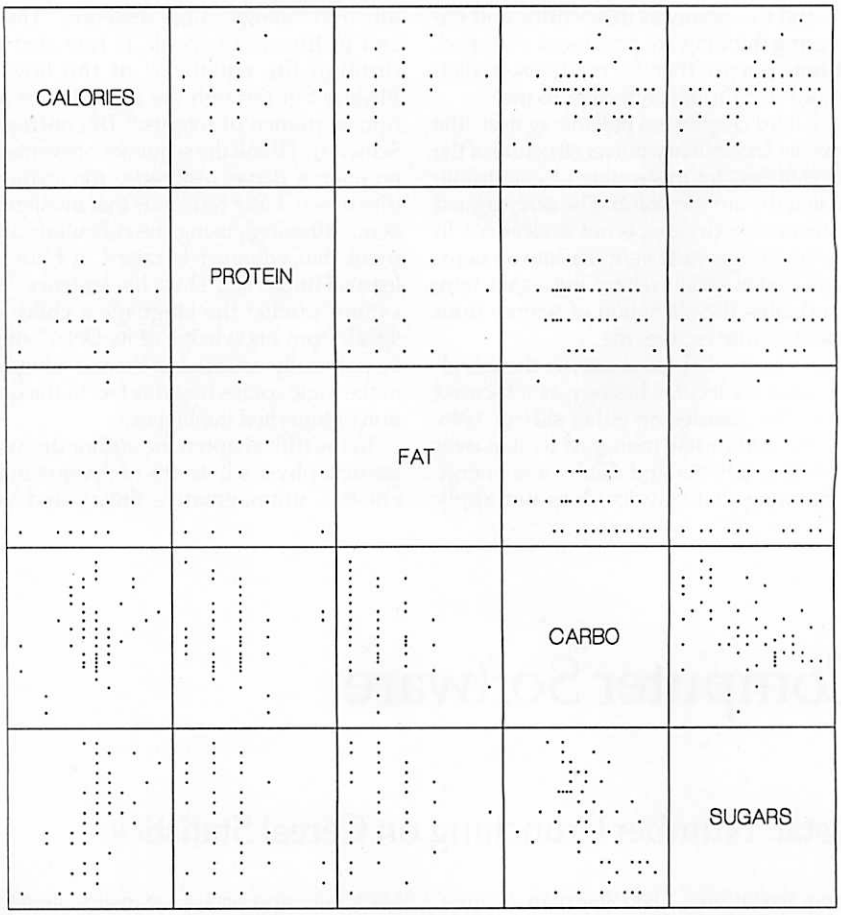

Figure 1. Scatter-plot matrix (SPLOM) for nutritional data on 76 breakfast cereals.

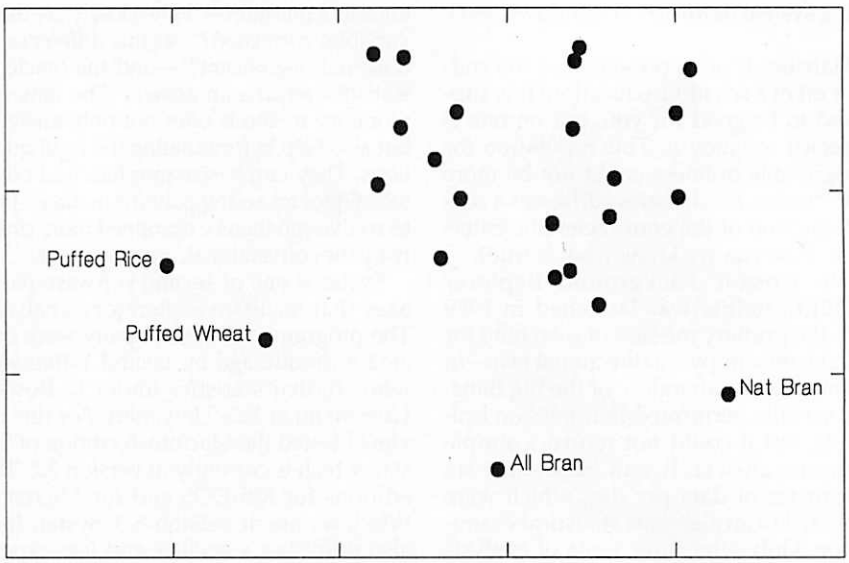

Figure 2. A subset of the cereals mapped on the Euclidean plane.