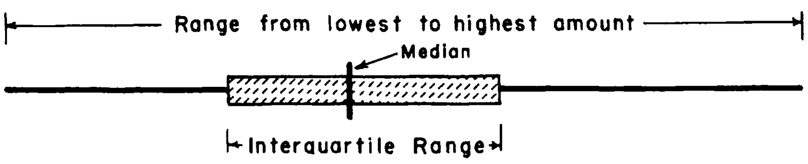
6 *Data-Ink Maximization and Graphical Design*

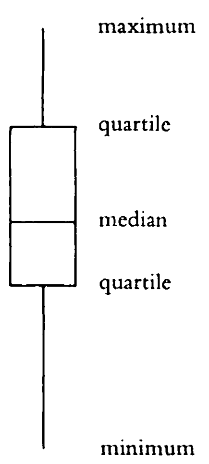
So far the principles of maximizing data-ink and erasing have helped to generate a series of choices in the process of graphical revision. This is an important result, but can the ideas reach beyond the details and particularities of editing? Is it possible to do what a theory of graphics is supposed to do, that is, to derive new graphical forms? In this chapter the principles are applied to many graphical designs, basic and advanced, including box plots, bar charts, histograms, and scatterplots. New designs result.

Redesign of the Box Plot

Mary Eleanor Spear's "range bar"



and John Tukey's "box plot"



Mary Eleanor Spear, *Charting Statistics* (New York, 1952), p. 166; and John W. Tukey, *Exploratory Data Analysis* (Reading, Massachusetts, 1977).

can be mostly erased without loss of information:

The revised design, a *quartile plot,* shows the same five numbers. It is easy to draw by hand or computer and, most importantly, can replace the conventional scatterplot frame. The straightedge need only be placed on the paper once to draw the quartile plot, compared to six separate placings for the box plot. An alternative is

but this design will not work effectively to frame a scatterplot. Nor does it look very good.

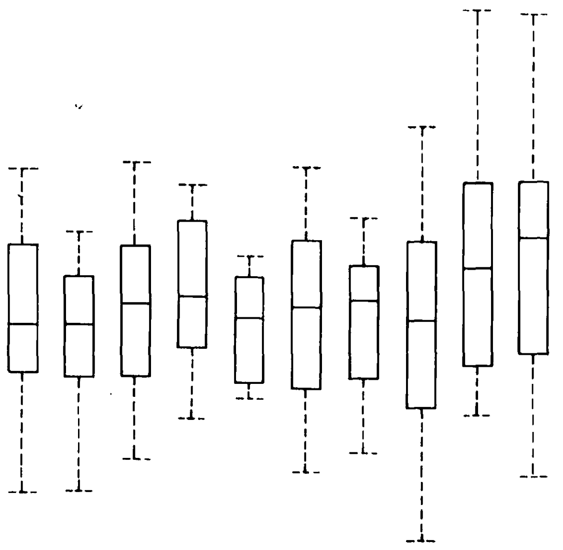
Perhaps special emphasis should be given to the middle half of the distribution, however, as in the box plot. This can be done by changing line weights

or, even better, by offsetting the middle half:

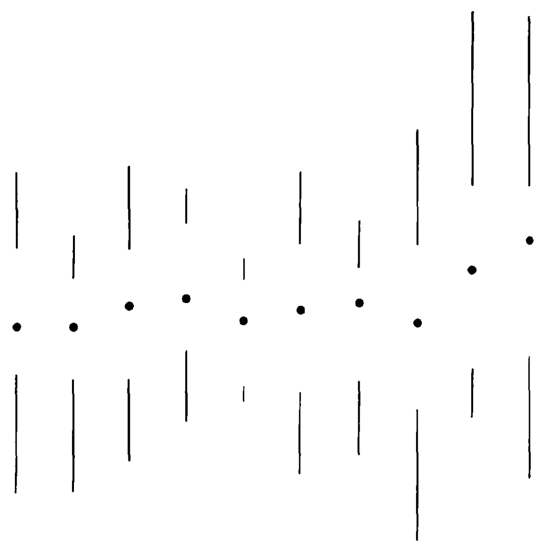
This latter design is the preferred form of the quartile plot. It uses the ink effectively and looks good.

In these revisions of the box plot, the principle of maximizing data-ink has suggested a variety of designs, but the choice of the best overall arrangement naturally also rests on statistical and aesthetic criteria—in other words, the procedure is one of *reasonable* data-ink maximizing.

The same logic applies to many similar designs, such as this "parallel schematic plot." The original required 80 separate placings of the straightedge, 50 horizontals and 30 verticals:



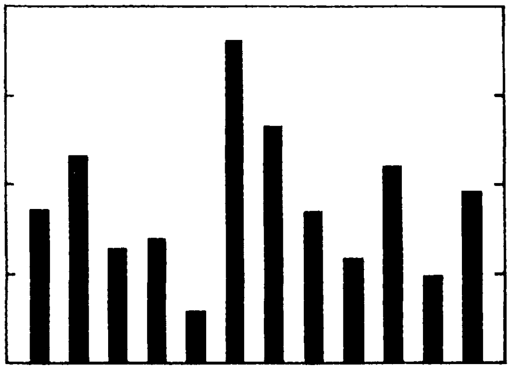
An erased version requires only 10 verticals to show the same information:



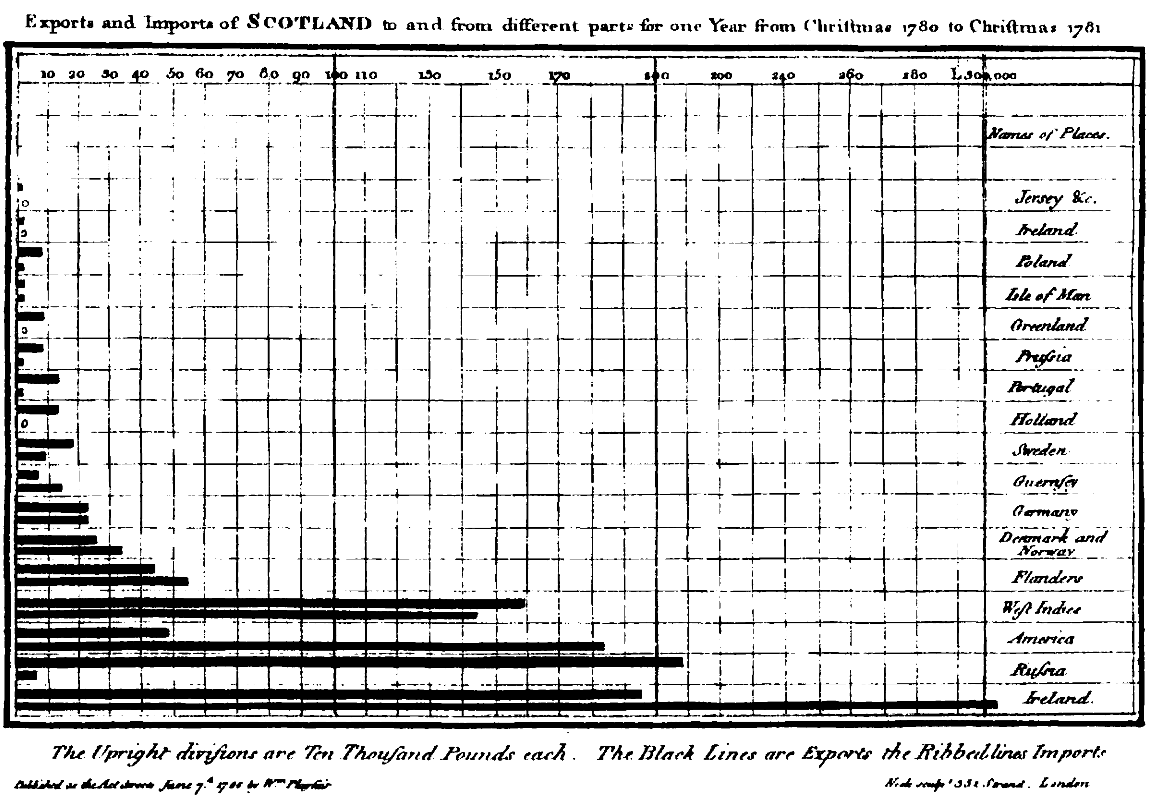
The large reduction in the amount of drawing is relevant for the use of such designs in informal, exploratory data analysis, where the research worker's time should be devoted to matters other than drawing lines.

Redesign of the Bar Chart/Histogram

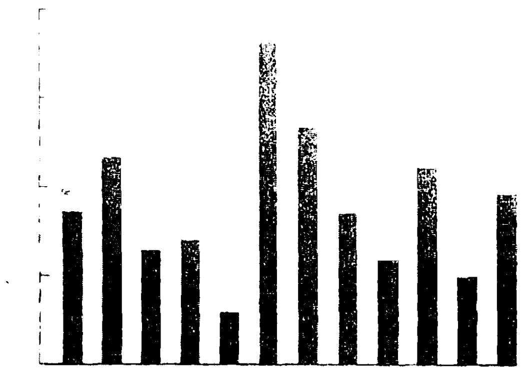
Here is the standard model bar chart, with the design endorsed by the practices and the style sheets of many statistical and scientific publications:



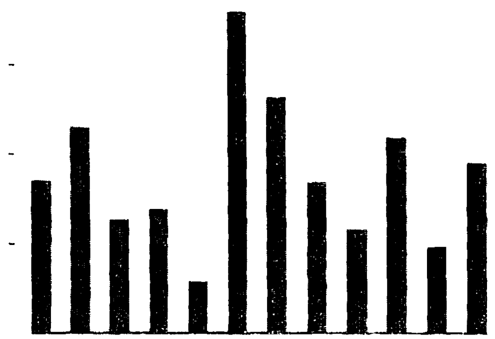
Its architecture differs little from Playfair's original design:



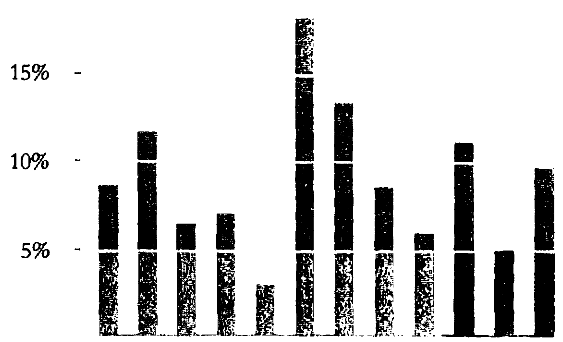
The box can be erased:



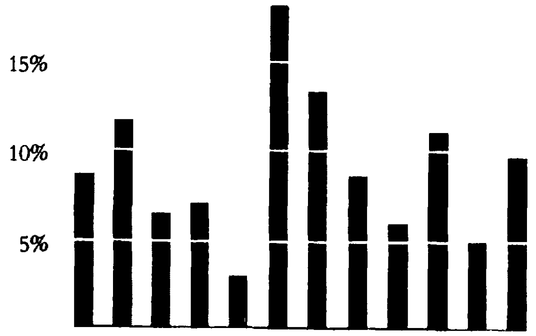
And the vertical axis, except for the ticks:



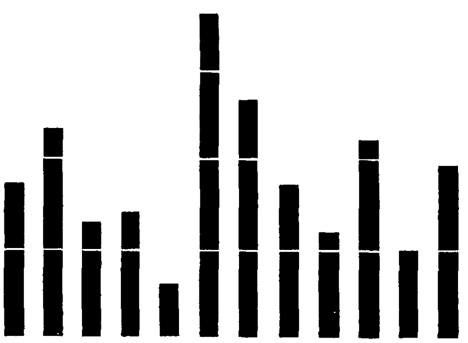
Even part of the data measures can be erased, making a *white grid*, which shows the coordinate lines more precisely than ticks alone:



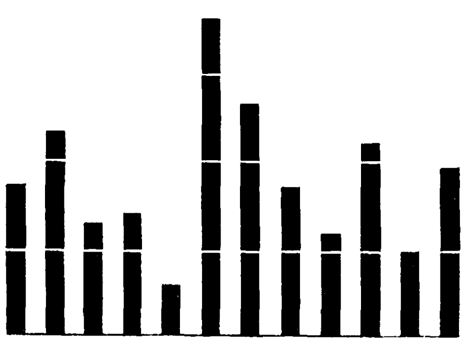
The white grid eliminates the tick marks, since the numerical labels on the vertical are tied directly to the white lines:



Although the intersection of the thicker bar with the thinner baseline creates an attractive visual effect (but also the optical illusion of gray dots at the intersections), the baseline can be erased since the bars define the end-point at the bottom:

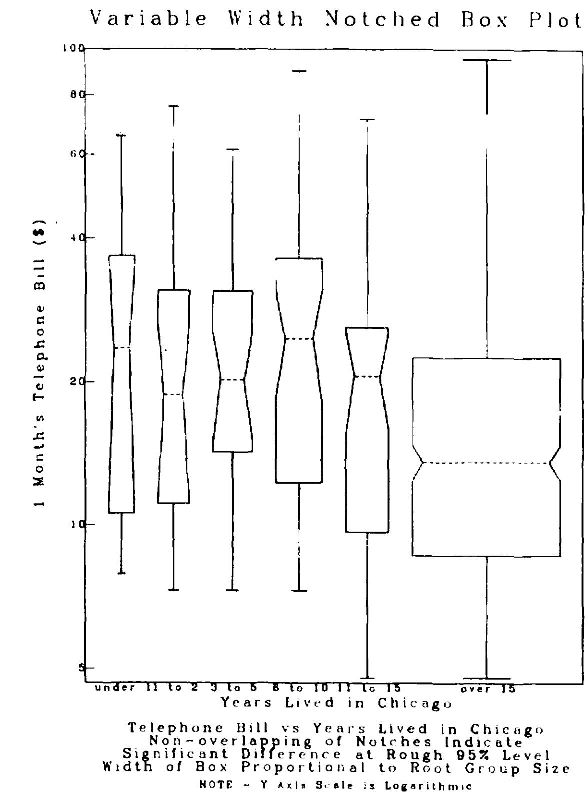


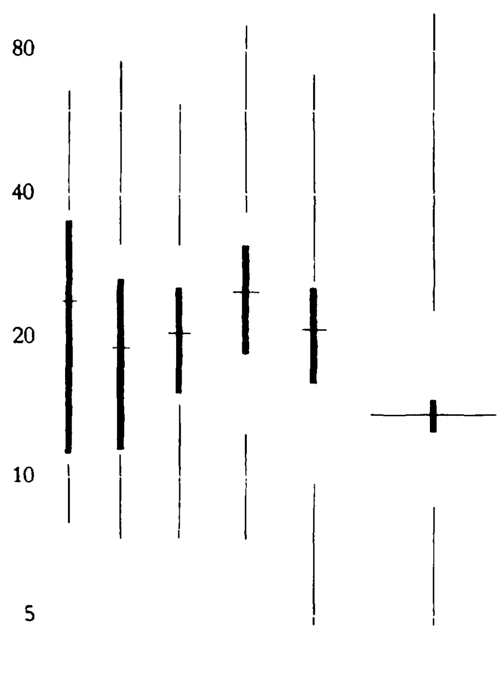
Still, a thin baseline looks good:



Erasing and data-ink maximizing have induced changes in the plain old bar chart. The techniques—no frame, no vertical axis, no ticks, and the white grid—apply to other designs:

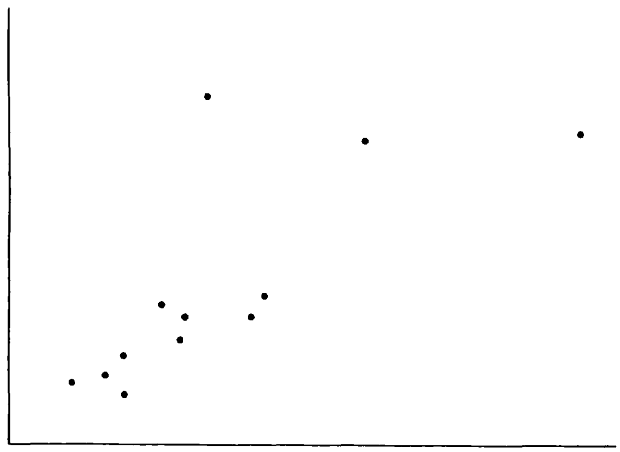
Robert McGill, John W. Tukey, and Wayne A. Larsen, "Variations of Box Plots," *American Statistician*, 32 (1978), 12–16.





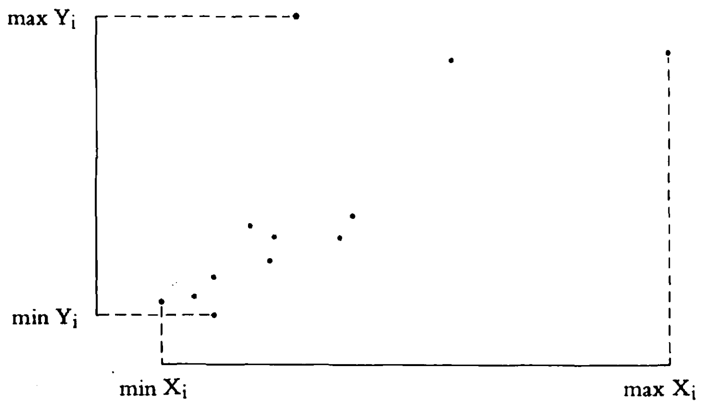
Redesign of the Scatterplot

Consider the standard bivariate scatterplot:

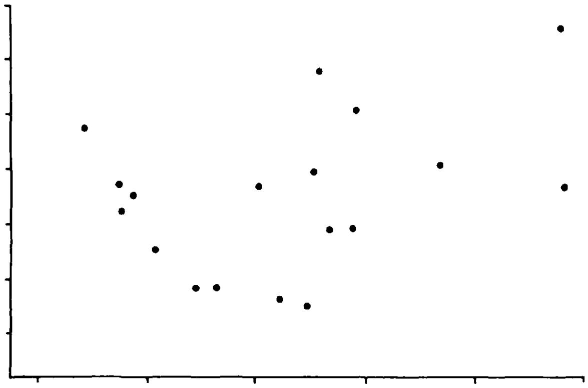


A useful fact, brought to notice by the maximization and erasing principles, is that the frame of a graphic can become an effective data-communicating element simply by erasing part of it. The frame lines should extend only to the measured limits of the data rather than, as is customary, to some arbitrary point like the next round number marking off the grid and grid ticks of the plot. That part of the frame exceeding the limits of the observed data is trimmed off:

The result, a *range-frame*, explicitly shows the maximum and minimum of both variables plotted (along with the range), information available only by extrapolation and visual estimation in the conventional design. The data-ink ratio has increased: some non-data-ink has been erased, and the remainder of the frame, now carrying information, has gone over to the side of data-ink.



Nothing but the tails of the frame need change:

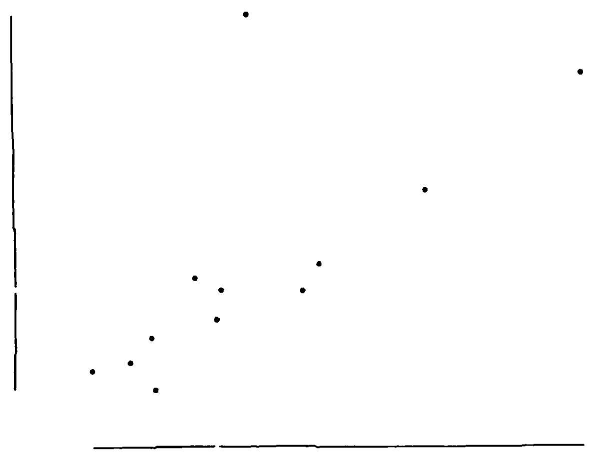


Conventional Scatterplot

Range-Frame

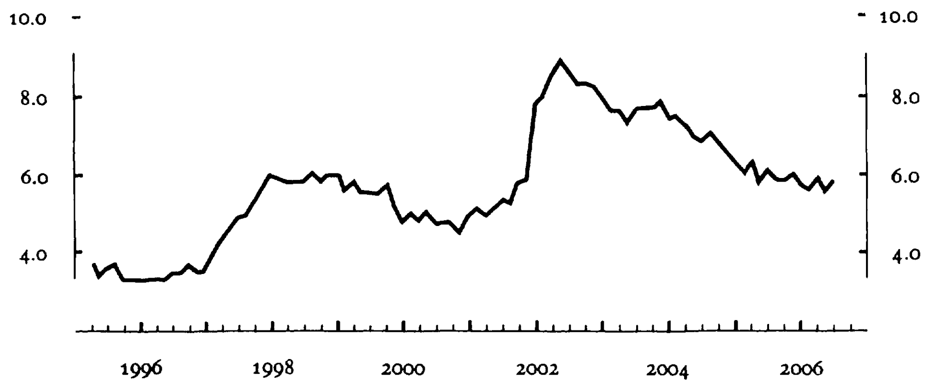
A range-frame does not require any viewing or decoding instructions; it is not a graphical puzzle and most viewers can easily tell what is going on. Since it is more informative about the data in a clear and precise manner, the range-frame should replace the non-data-bearing frame in many graphical applications.

A small shift in the remaining ink turns each range-frame into a quartile plot:



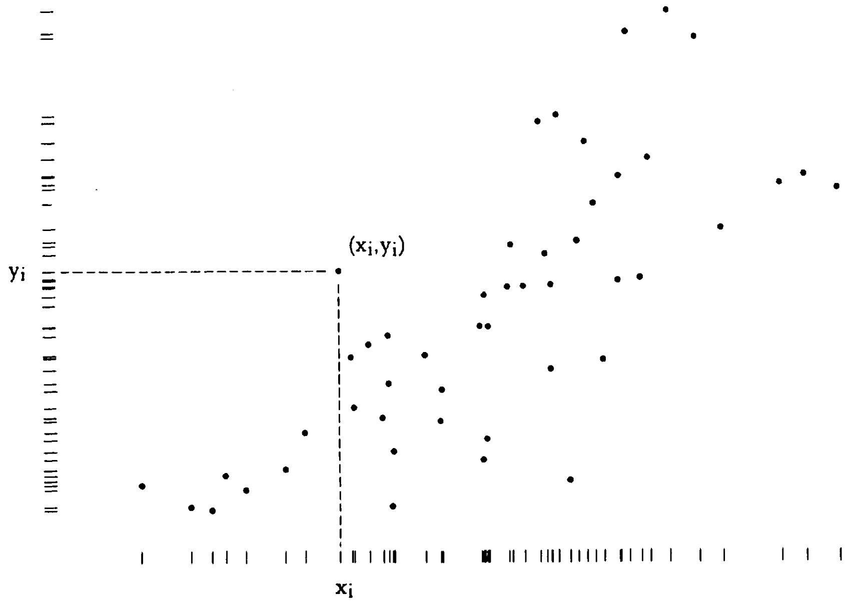
Erasing and editing has led to the display of ten extra numbers (the minimum, maximum, two quartiles, and the median for both variables). The design is useful for analytical and exploratory data analysis, as well as for published graphics where summary characterizations of the marginal distributions have interest. The design is nearly always better than the conventionally framed scatterplot.

Range-frames can also present ranges along a single dimension. Here the historical high and low are shown in the vertical frame. This is an excellent practice and should be used widely in all sorts of displays, both scientific and unscientific:



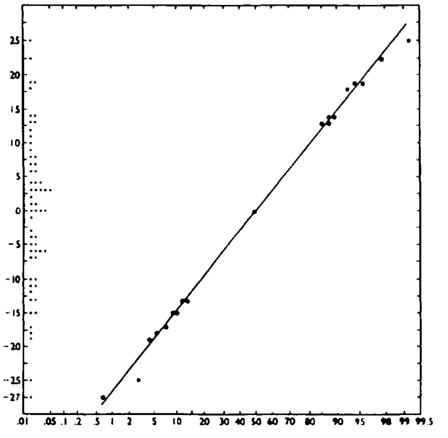
Finally, the entire frame can be turned into data by framing the bivariate scatter with the marginal distribution of each variable. The *dot-dash-plot* results.[1]

[1] The terminology follows tradition, for scatterplots were once called "dot diagrams"—for example, in R. A. Fisher's *Statistical Methods for Research Workers* (Edinburgh, 1925).



The dot-dash-plot combines the two fundamental graphical designs used in statistical analysis, the marginal frequency distribution and the bivariate distribution. Dot-dash-plots make routine what good data analysts do already—plotting marginal and joint distributions together.

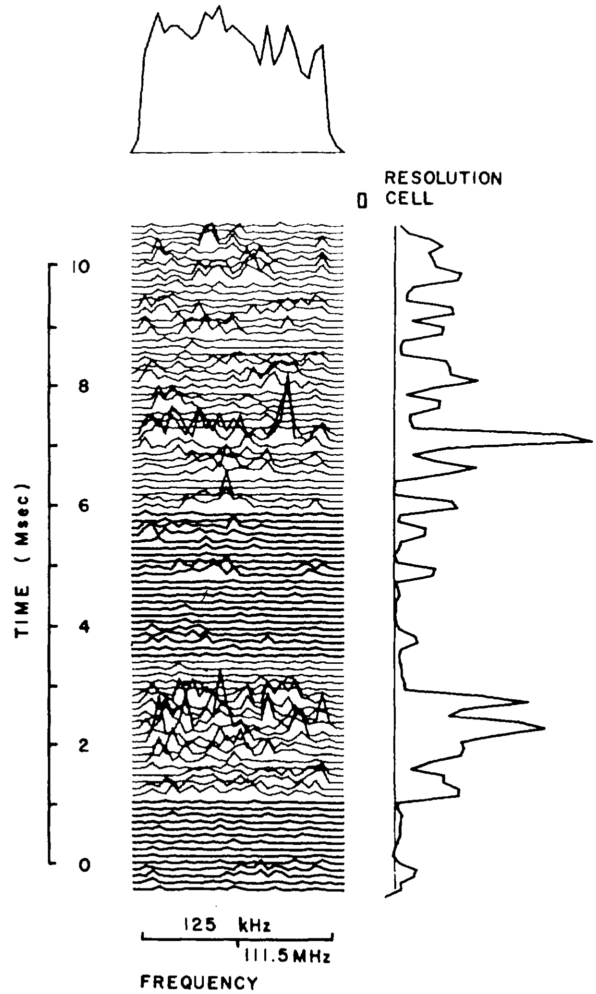
An empirical cumulative distribution of residuals on a normal grid shows the outer 18 terms plus the 30th term, with all 60 points plotted in the marginal distribution:



Cuthbert Daniel, *Applications of Statistics to Industrial Experimentation* (New York, 1976), p. 155.

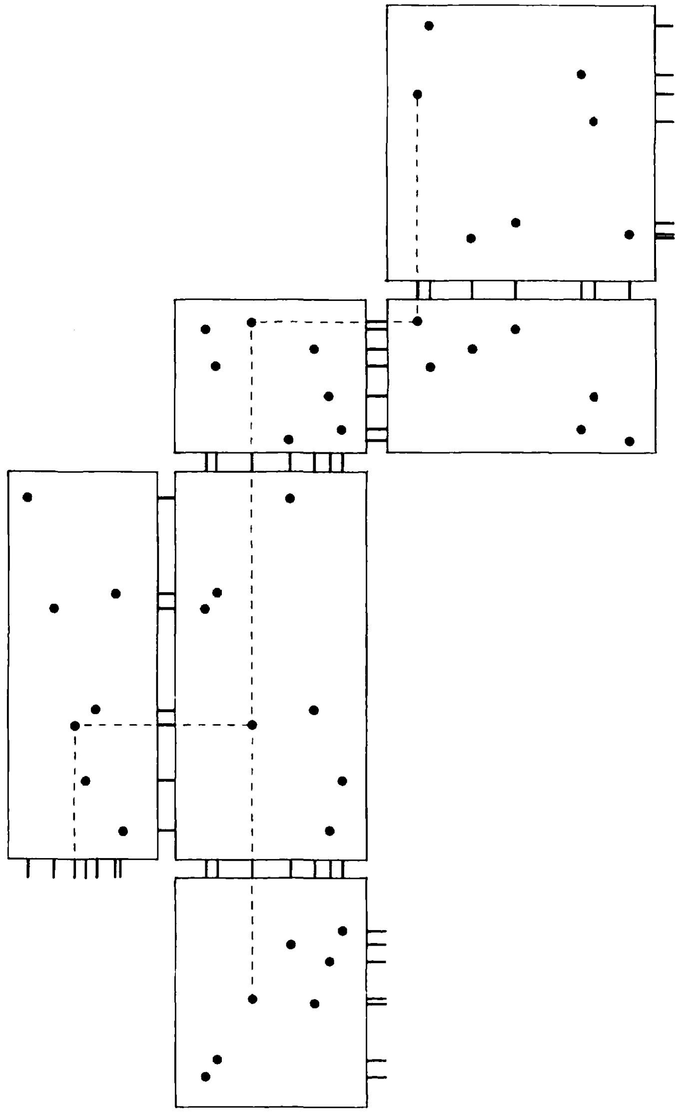
Similarly, this data-rich graphic of signals from pulsars shows both marginal distributions:

Timothy H. Hankins and Barney J. Rickett, "Pulsar Signal Processing," in Berni Alder, et al., eds., *Methods in Computational Physics, Volume 14: Radio Astronomy* (New York, 1975), p. 108.



Narrowband spectra of individual subpulses. Each point of the intensity $I_q(t)$ plotted on the right is the sum of the distribution of intensities across the receiver bandwidth shown in the center. At the top is plotted the spectrum averaged over the pulse. In the limit of many thousands of pulses this would show the receiver bandpass shape.

The fringe of dashes in the dot-dash-plot can connect a series of bivariate scatters in a *rugplot* (since it resembles a set of fringed rugs—and covers the statistical ground):



Reflecting the one-dimensional projections from each scatter, the dashes encourage the eye to notice how each plot filters and translates the data through the scatter from one adjacent plot to the next. Sometimes it is useful to think of each bivariate scatter as the imperfect empirical representation of an underlying curve that transforms one variable into another. In the rugplot, the sequence of variables can wander off as appropriate. The quantitative history of a single observation can be traced through a series of one- and two-dimensional contexts.

Conclusion

The first part of a theory of data graphics is in place. The idea, as described in the previous three chapters, is that most of a graphic's ink should vary in response to data variation. The theory has something to say about a great variety of graphics—workaday scientific charts, the unique drawings of Roger Hayward, the exemplars of graphical handbooks, newspaper displays, computer graphics, standard statistical graphics, and the recent inventions of Chernoff and Tukey.

The observed increases in efficiency, in how much of the graphic's ink carries information, are sometimes quite large. In several cases, the data-ink ratio increased from .1 or .2 to nearly 1.0. The transformed designs are less cluttered and can be shrunk down more readily than the originals.

But, are the transformed designs *better*?

(1) They are necessarily better within the principles of the theory, for more information per unit of space and per unit of ink is displayed. And this is significant; indeed, the history of devices for communicating information is written in terms of increases in efficiency of communication and production.

(2) Graphics are almost always going to improve as they go through editing, revision, and testing against different design options. The principles of maximizing data-ink and erasing generate graphical alternatives and also suggest a direction in which revisions should move.

(3) Then there is the audience: will those looking at the new designs be confused? Some of the designs are self-explanatory, as in the case of the range-frame. The dot-dash-plot is more difficult, although it still shows all the standard information found in the scatterplot. Nothing is lost to those puzzled by the frame of dashes, and something is gained by those who do understand. Moreover, it is a frequent mistake in thinking about statistical graphics to underestimate the audience. Instead, why not assume that if you understand it, most other readers will, too? Graphics should be as intelligent and sophisticated as the accompanying text.

(4) Some of the new designs may appear odd, but this is probably because we have not seen them before. The conventional designs for statistical graphics have been viewed thousands of times by nearly every reader of this book; on the other hand, the range-frame, the dot-dash-plot, the white grid, the quartile plot, the rugplot, and the half-face just a few times. With use, the new designs will come to look just as reasonable as the old.

Maximizing data ink (within reason) is but a single dimension of a complex and multivariate design task. The principle helps conduct experiments in graphical design. Some of those experiments will succeed. There remain, however, many other considerations in the design of statistical graphics—not only of efficiency, but also of complexity, structure, density, and even beauty.