1   *Graphical Excellence*

Excellence in statistical graphics consists of complex ideas communicated with clarity, precision, and efficiency. Graphical displays should

- show the data

- induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else

- avoid distorting what the data have to say

- present many numbers in a small space

- make large data sets coherent

- encourage the eye to compare different pieces of data

- reveal the data at several levels of detail, from a broad overview to the fine structure

- serve a reasonably clear purpose: description, exploration, tabulation, or decoration

- be closely integrated with the statistical and verbal descriptions of a data set.

Graphics *reveal* data. Indeed graphics can be more precise and revealing than conventional statistical computations. Consider Anscombe's quartet: all four of these data sets are described by exactly the same linear model (at least until the residuals are examined).

| I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

$N = 11$
mean of X's $= 9.0$
mean of Y's $= 7.5$
equation of regression line: $Y = 3 + 0.5X$
standard error of estimate of slope $= 0.118$
$t = 4.24$
sum of squares $X - \overline{X} = 110.0$
regression sum of squares $= 27.50$
residual sum of squares of Y $= 13.75$
correlation coefficient $= .82$
$r^2 = .67$

And yet how they differ, as the graphical display of the data makes vividly clear:
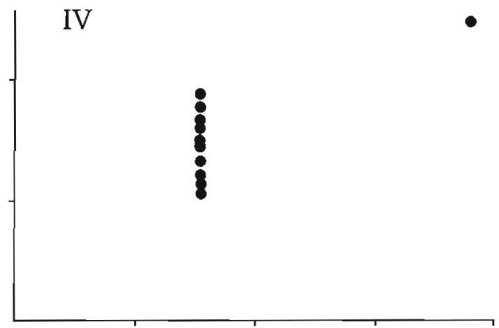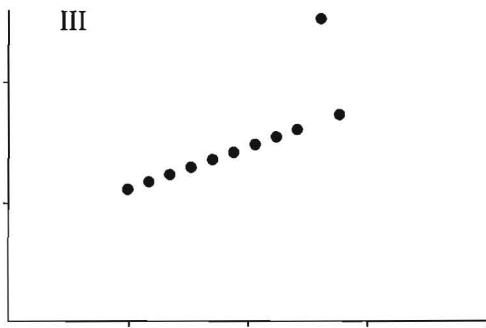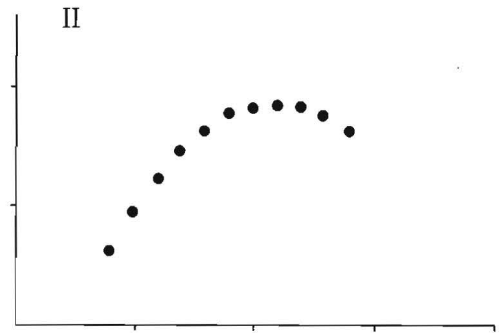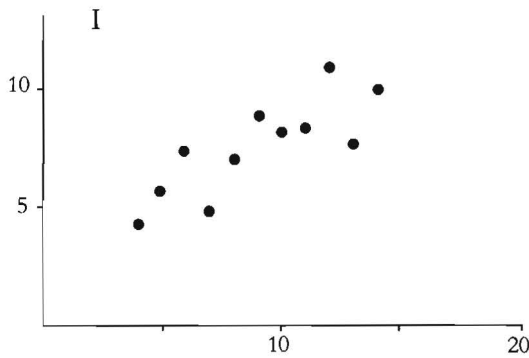
And likewise a graphic easily reveals point A, a wildshot observation that will dominate standard statistical calculations. Note that point A hides in the marginal distribution but shows up as clearly exceptional in the bivariate scatter.
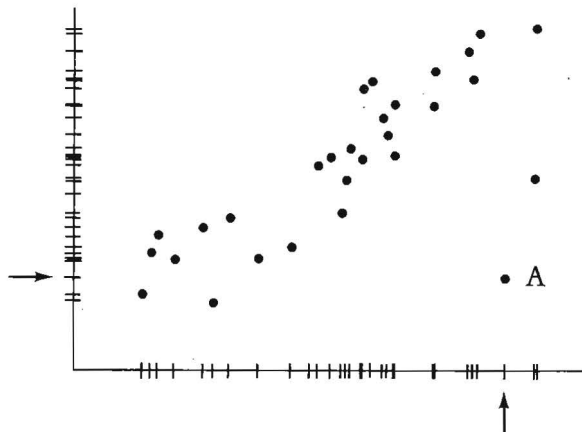
Of course, statistical graphics, just like statistical calculations, are only as good as what goes into them. An ill-specified or preposterous model or a puny data set cannot be rescued by a graphic (or by calculation), no matter how clever or fancy. A silly theory means a silly graphic:

Edward R. Dewey and Edwin F. Dakin, *Cycles: The Science of Prediction* (New York, 1947), p. 144.
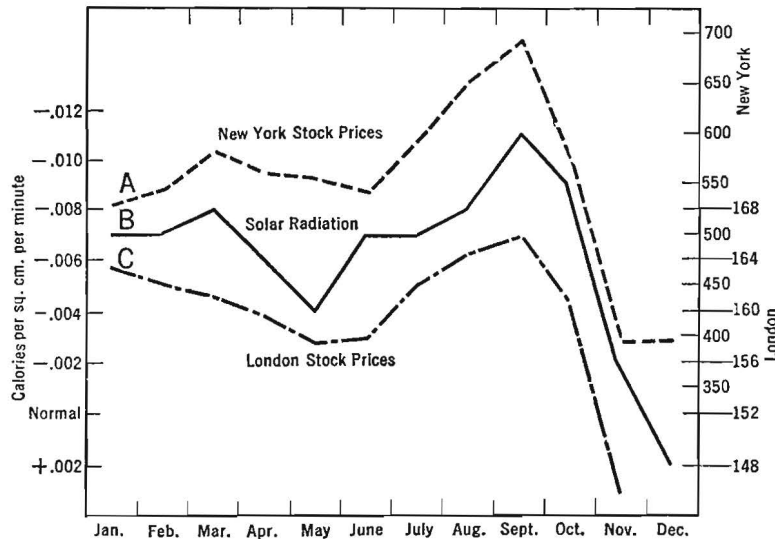
SOLAR RADIATION AND STOCK PRICES

A. New York stock prices (Barron's average). B. Solar Radiation, inverted, and C. London stock prices, all by months, 1929 (after Garcia-Mata and Shaffner).

Let us turn to the practice of graphical excellence, the efficient communication of complex quantitative ideas. Excellence, nearly always of a multivariate sort, is illustrated here for fundamental graphical designs: data maps, time-series, space-time narrative designs, and relational graphics. These examples serve several purposes, providing a set of high-quality graphics that can be discussed (and sometimes even redrawn) in constructing a theory of data graphics, helping to demonstrate a descriptive terminology, and telling in brief about the history of graphical development. Most of all, we will be able to see just how good statistical graphics can be.