*The Quality Magazine* **7** #4 (1998), 64-68.

# Good Graphs for Better Business

W. S. Cleveland & N.I. Fisher

Are we on track with our target for market share? Is our installation process capable of meeting the industry benchmark? What things are causing most of the unhappiness with our staff?

These are typical of the management problems that require timely and accurate information. They are also problems for which effective use of graphs can make a big difference.

The good news is that graphical capabilities are now readily available in statistical, spreadsheet, and word-processing packages. The bad news is that much of this graphical capability produces graphs that can hide or seriously distort crucial information contained in the data.

For a long time, how to make a good graph was largely a matter of opinion. However, the last 20 years have seen the development of a set of principles for sound graphical construction, based on solid scientific research and experimentation. Good entry points to these principles are provided in the References. In this article, we look at a couple of common examples of applying these principles.

It is helpful to think about the whole process of graphing, shown schematically in Figure 1. The crucial question is: *How does the choice of graph affect the information as perceived by the recipient of the graph?*

To see how graphs can conceal, or reveal, information, consider the humble pie chart. Figure 2 shows data on the contributions to enterprise profits of a particular product in various regions R1, R2, … around Australia (labels modified from original, but retaining the same ordering, which was alphabetical). What information is this supposed to be purveying? Certainly the caption doesn't enlighten us. If we wanted the actual percentage share for each region, we should simply use a table: tables are intended to provide precise numerical data, whereas the purpose of graphs is to reveal pattern.

For a more elaborate example, we turn to another popular graphical display, the divided bar chart for trend data. Figure 3 shows data on market share of whitegoods sales for an eighteen month period, based on monthly industry surveys. What can we glean from this? Total sales aren't changing. Manufacturer 1 has the biggest market share. Is there nothing else?

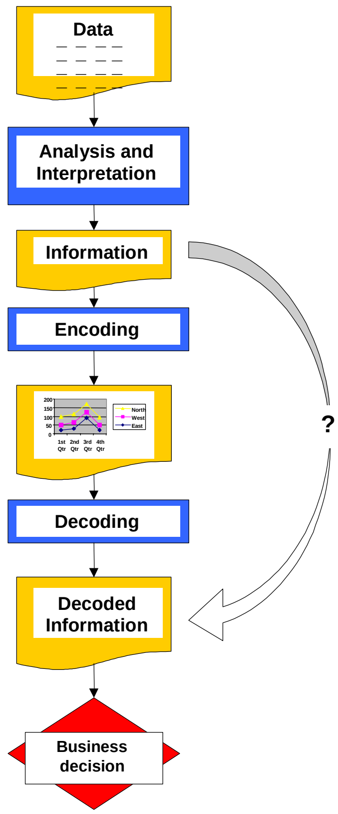


**Figure 1.** The graphical process involves extraction of information from data, a decision about which *patterns* are to be displayed, and then selection of a type of graph that will reveal this pattern to the user, without distortion.

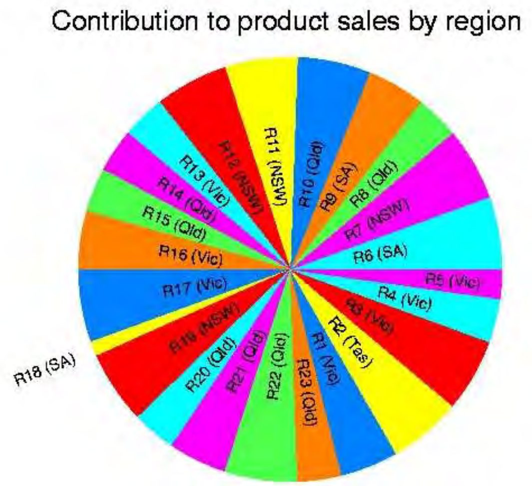


**Figure 2**. Pie chart, showing the relative contributions to the profits of an enterprise from various Divisions around Australia.

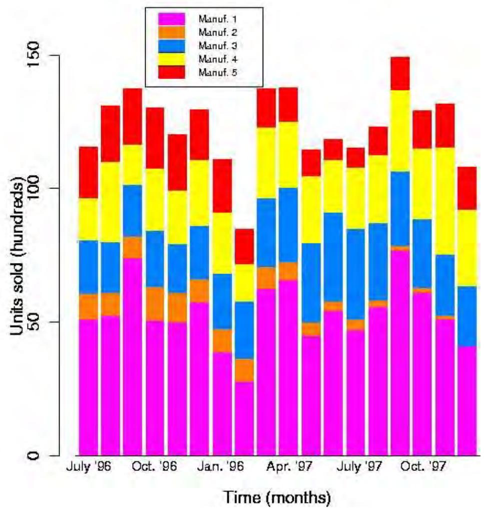


**Figure 3**. Divided bar chart, showing monthly sales of different brands of whitegoods over an 18-month period.

Returning to Figure 2, no obvious patterns emerge. So what's happened in the graphical process? Perhaps the *Analysis and Interpretation* step wasn't carried out. So, let's try a different plot that shows things more simply, a *dotplot*: see Figure 4.

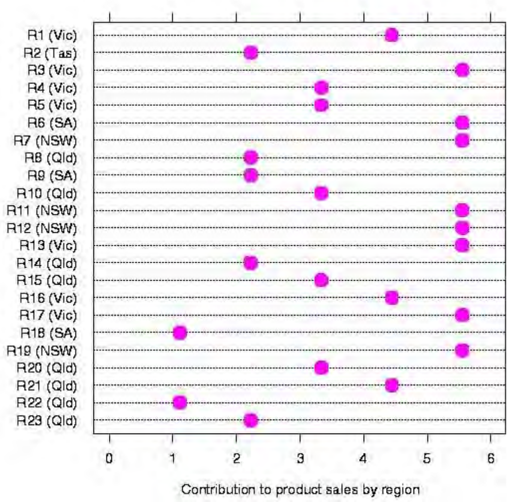


**Figure 4**. A dotplot of the data used in Figure 2, showing the relative contributions to enterprise profits from its various Divisions around  Australia.  The discrete nature of the data is immediately evident.

A simple pattern emerges immediately: there are only five different levels of contribution; some rounding of the raw data has been performed. Why wasn't this evident in the pie chart?

The answer to this is provided by one of the fundamental tenets of graphing. Detection of *pattern* with this type of data is best done when each measurement is plotted as a distance from a common baseline. The baseline in Figure 4 is the left vertical axis, and we're simply comparing horizontal lengths that are vertically aligned.

On the other hand, in the pie chart, we're trying to compare *angles*, and very small angles at that. This sort of comparison is known to be imprecise. Similar problems occur when the data are graphed in other colourful or picturesque ways, such as when their sizes are represented by 3-dimensional solid volumes.

However, we haven't finished with this data set yet. At least one more step should be taken: re-order the data so that they plot from largest to smallest. The final result is shown in Figure 5.

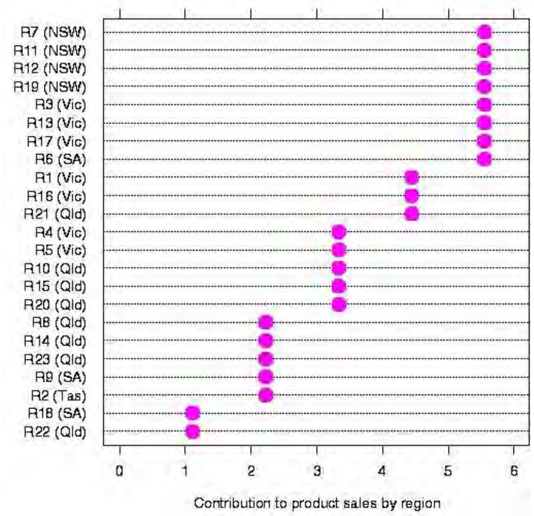


**Figure 5**. The display from Figure 4 has been modified, so that the data plot from largest to smallest. A further pattern emerges: different States tend to contribute differently to enterprise profits.

A (potentially) more important pattern has now emerged: different States are not contributing equally to group profits.  This information may be vital in helping management to identify a major improvement opportunity. We create one more graph to bring this out: see Figure 6.

What can we now say in defence of Figure 2 in the light of Figures 5 and 6? Really, only that it was produced from a spreadsheet at the press of a button. But judged by a standard of how well it conveys information, it has failed. This is typical of pie charts.

Now let's re-visit the market share data plotted in Figure 3. What aspects of this graph might be hindering or preventing us from seeing important pattern?

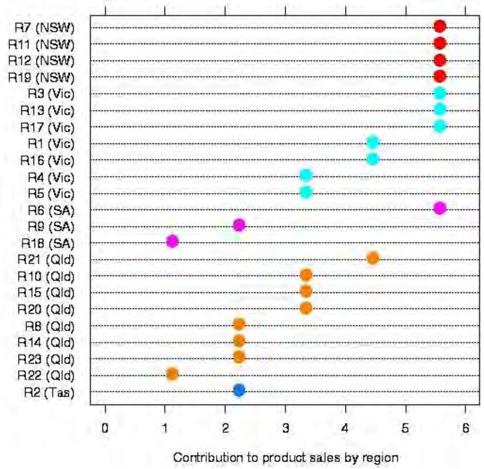


**Figure 6**. The contributions to group profit by different regions are plotted by State. The clear differences between States are evident.

We can get some idea of what's happening with the goods sold by Manufacturer 1: probably not much. We see this pattern because we're effectively comparing aligned (vertical) lengths: the 18 values of monthly sales for this Manufacturer are all measured up from a common baseline (the horizontal axis).

However, this isn't the case for any of the other variables. For example, the individual values for the second manufacturer are measured upwards from where the corresponding values for the first manufacturer stop: the lengths are *not* aligned. So the first step is to re-plot the data in order that, for each variable, we are comparing aligned lengths. See Figure 7.

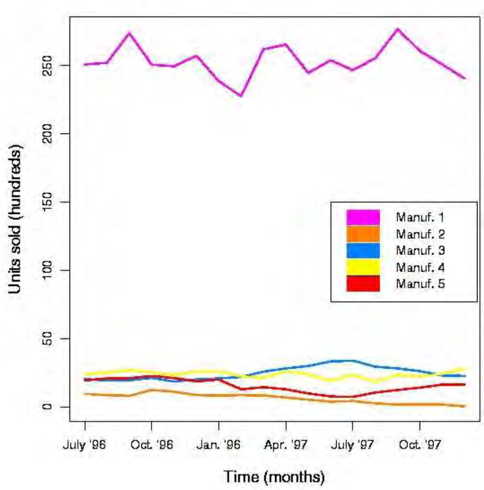


**Figure 7**. The monthly sales data from Figure 3 have been replotted so that sales patterns for each manufacturer can be seen without distortion. However, the curve for the dominant manufacturer is compressing patterns in the other curves.

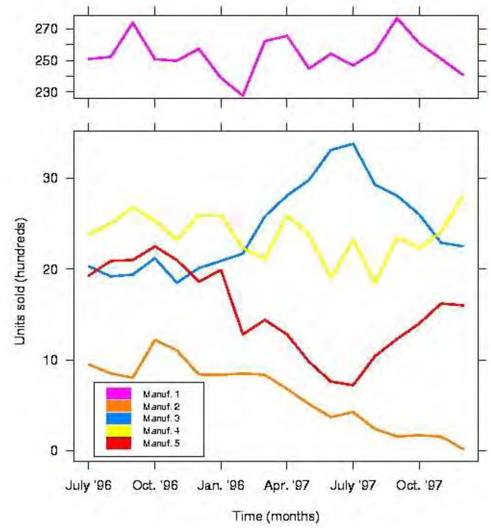


**Figure 8**. The monthly sales data from Figure 7 have been replotted so that the dominant curve is displayed separately, with a false origin, and the other curves that are measured on a much smaller scale can then be plotted using a better aspect ratio. This reveals more information about individual and comparative trends in the curves.

We have made some progress. Manufacturer 1 is more easily studied, and there is little evidence of anything other than random fluctuations around an average monthly sales volume of about 25,000 units. However, possible patterns in the other curves are difficult to detect because the scale of the axes is adjusted to allow all the data to be plotted on the same graph. The next step is to display the dominant curve separately, and to use a better *aspect ratio* when plotting the other variables. See Figure 8.

Now some very interesting patterns emerge. It is evident that sales for Manufacturer 2 have been in decline for most of this period. Not much is happening with Manufacturer 4, who is averaging about 2500 units. However, there is something happening with Manufacturers 3 and 5. From months 6 to 18, sales for Manufacturer 3 rose significantly and then declined. This appears to have been at the expense of Manufacturer 5, whose sales declined significantly and then recovered. If this comparison is really of interest, it should be studied separately. The difference is plotted in Figure 9.

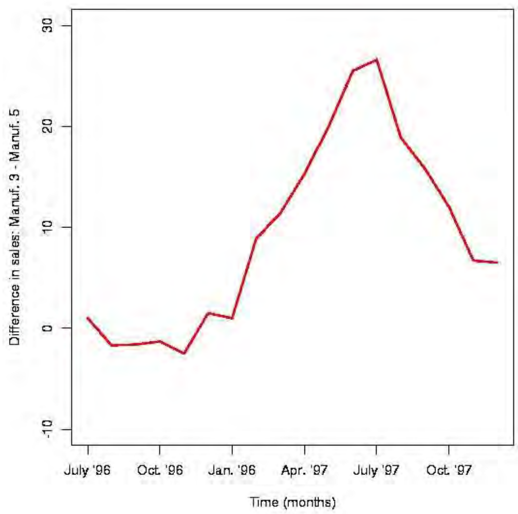


**Figure 9**. This graph shows the difference between sales of Manufacturers 5 and 3. Over the period January – July 1997 there was a marked increase in sales in favour of Manufacture 5; after July, this advantage declined steadily to the end of the year.

One plausible explanation would be a short but intense marketing campaign conducted during this period; there may be others. The main point is that appropriate graphs have elicited very interesting patterns in the data that may well be worthy of further exploration. The divided bar chart has done poorly; this is typically the case.

There is much to be learnt about constructing good graphs, to do with effective use of colour, choice of aspect ratio, displaying large volumes of data, use of false origin, and so on. These issues are discussed at length in the references.

To summarise, what basic messages can we draw from these simple examples? There are several:

- the process of effective graph construction begins with simple analyses to see what sorts of patterns, that is, *information*, are present.
- for many graphs, pattern detection is far more acute when the data are measured from a common baseline, so that we are comparing *aligned lengths*
- re-arranging the values in a dot plot so that they are in decreasing order provides greatly enhanced pattern recognition
- graph construction is an iterative process
- sometimes more than one graph is needed to show all the interesting patterns in the best way
- two very commonly-used displays, pie charts and divided bar charts, typically do a poor job of revealing pattern

Can you afford *not* to be using graphs in the way information is reported to you, or the way you are reporting it? How else can vital patterns be revealed and presented, and so provide effective input to decision-making at all levels of your enterprise?

**References**

Cleveland, William S. (1993), *Visualizing Data.* Hobart Press, Summit, New Jersey.

Cleveland, William S. (1994), *The Elements of Graphing Data.* Second edition. Hobart Press, Summit, New Jersey.

Tufte, Edward R. (1983). *The Visual Display of Quantitative Information.* Graphics Press, Cheshire, Connecticut.

Bill Cleveland is a member of the Statistics & Data Mining Research Department of Bell Labs, Murray Hill, NJ, USA.

Nick Fisher is a member of the Organisational Performance Measurement and Data Mining research groups in CSIRO.