What will this chapter tell me?

As I got a bit older I used to love exploring. At school they would teach you about maps and how important it was to know where you were going and what you were doing. I used to have a more relaxed view of exploration and there is a little bit of a theme of me wandering off to whatever looked most exciting at the time. I got lost at a holiday camp once when I was about 3 or 4. I remember nothing about this, but apparently my parents were frantically running around trying to find me while I was happily entertaining myself (probably by throwing myself head first out of a tree or something). My older brother, who was supposed to be watching me, got a bit of flak for that, but he was probably working out equations to bend time and space at the time. He did that a lot when he was 7. The careless explorer in me hasn't really gone away: in new cities I tend to just wander off and hope for the best, usually get lost and so far haven't managed to die (although I tested my luck once by wandering through part of New Orleans where apparently tourists get mugged a lot – it seemed fine to me). When exploring data you can't afford not to have a map; to explore data in the way that the 6-year-old me used to explore the world is to spin around 8000 times while drunk and then run along the edge of a cliff. Wright (2003) quotes Rosenthal, who said that researchers should 'make friends with their data'. This wasn't meant to imply that people who use statistics may as well befriend their data because the data are the only friend they'll have; instead Rosenthal meant that researchers often rush their analysis. Wright makes the analogy of a fine wine: you should savour the bouquet and delicate flavours to truly enjoy the experience. That's perhaps overstating the joys of data analysis, but rushing your analysis is, I suppose, a bit like gulping down a bottle of wine: the outcome is messy and incoherent. To negotiate your way around your data you need a map. Maps of data are called graphs, and it is into this tranquil and tropical ocean that we now dive (with a compass and ample supply of oxygen, obviously).

What makes a good graph?

I want to begin by talking about some general issues when presenting data. SPSS and other packages make it very easy to produce snazzy-looking graphs (see Section 4.9), and you may find yourself losing consciousness at the excitement of colouring your graph bright pink (really, it's amazing how excited my psychology students get at the prospect of bright pink graphs – personally I'm not a fan of pink). Much as pink graphs might send a twinge of delight down your spine, I want to urge you to remember why you're drawing the graph – it's not to make yourself (or others) purr with delight at the pinkness of your graph; it's to present information (dull, perhaps, but true).

Tufte (2001) wrote an excellent book about how data should be presented. He points out that graphs should do the following, among other things:

Show the data.

Induce the reader to think about the data being presented (rather than some other aspect of the graph, like how pink it is).

Avoid distorting the data.

Present many numbers with minimum ink.

Make large data sets (assuming you have one) coherent.

Encourage the reader to compare different pieces of data.

Reveal the underlying message of the data.

However, graphs often don't do these things (see Wainer, 1984, for some examples). Let's look at an example of a bad graph. When searching around for the worst example of a graph that I have ever seen, it turned out that I didn't need to look any further than myself – it's in the first edition of this book (Field, 2000). Overexcited by SPSS's ability to add pointless fluff to graphs (like 3-D effects, fill effects and so on – Tufte calls these **chartjunk**), I literally went into some weird orgasmic state and produced an absolute abomination (I'm surprised Tufte didn't kill himself just so he could turn in his grave at the sight of it). The only consolation was that because the book was published in black and white, it's not bloody pink! The graph is reproduced in Figure 4.2. What's wrong with this graph?

The bars have a 3-D effect: Never use 3-D plots for a graph plotting two variables because it obscures the data.[1] In particular, 3-D effects make it hard to see the values of the bars: in Figure 4.2, for example, the 3-D effect makes the error bars almost impossible to read.

Patterns: The bars also have patterns, which, although very pretty, distract the eye from what matters (namely the data). These are completely unnecessary.

Cylindrical bars: Were my data so sewage-like that I wanted to put them in silos? The cylinder effect muddies the data and distracts the eye from what is important.

Badly labelled *y*-axis: 'Number' of what? Delusions? Fish? Cabbage-eating sea lizards from the eighth dimension? Idiots who don't know how to draw graphs?

Now, take a look at the alternative version of this graph (Figure 4.3). Can you see what improvements have been made?

A 2-D plot: The completely unnecessary third dimension is gone, making it much easier to compare the values across therapies and thoughts/behaviours.

The *y*-axis has a more informative label: We now know that it was the number of obsessive thoughts or actions per day that was being measured.

Distractions: There are fewer distractions like patterns, cylindrical bars and the like.

---

[1] If you do 3-D plots when you're plotting only two variables then a bearded statistician will come to your house, lock you in a room and make you write I μυστ νοτ δο 3–Δ γραπησ 75,172 times on the blackboard. Really, they will.

Minimum ink: I've got rid of superfluous ink by getting rid of the axis lines and by using lines on the bars rather than grid lines to indicate values on the *y*-axis. Tufte would be pleased.