Bench philosophy (14): Preparing figures

Effective scientific illustrations

We all know that effective communication is an essential skill in science. The ability to write in such a way that the reader is able, with the minimum possible effort, to grasp clearly what you are trying to say is a skill we rightly spend many hours perfecting. Sadly, much less attention is often paid to visual literacy – the art of preparing scientific figures.

A well-prepared figure can communicate large, complex datasets in a way that exploits the special ability of the brain to grasp complex ideas and relationships visually. As an editor of an international peer-reviewed journal, I come across interesting, well-written papers, which would have a greater impact if more attention were paid to the figures. Like good writing, preparing good figures is an art but one that can be improved by paying attention to certain guidelines.

Tell the story in pictures

The first guideline, particularly when it comes to preparing a paper for publication, is to tell the story with pictures. Aim to allow the reader to grasp the outline of your findings simply by looking over the figures. Most readers "scan" a new paper by just glancing at the figures, so it is often the figures that determine whether the reader will decide to read the paper in detail. Indeed, in my lab we draft the whole paper around the figures, letting them tell the story and drive the structure of the paper. This approach also helps decide what information should be included in the figure. Visual designer Edward Tufte (*www.edwardtufte.*

com/tufte/) is an excellent source of ideas on visual presentation of data. He has introduced the concept of chart-junk: the unhelpful tendency for unwanted lines, shading, etc. to clutter a figure. To help avoid this, he advises doing everything you can to reduce the ink:data ratio.

Consider figure 1A, which is a typical default diagram from Microsoft Excel. Excel is an excellent package with many virtues but preparing figures is not one of its best, especially if the user accepts the default settings without question. The horizontal lines in figure 1A are not only unnecessary, their 3D appearance introduces a false perspective that can cause the first impressions of the data to be misleading. The actual value for column 1 is 44 but it aligns with a value around 40 on the lines behind. Which is the real value? Secondly, does the reader need to read the information from the chart to the precision implied by the lines?

If the actual values of the data are the focus rather than trends in the data, the numbers should be given in the text instead of a figure (which they should be in any case if only a small dataset is given). What about the 3D appearance of the columns – is that really needed? It doesn't tell us anything, so lets get rid of it. As for colour, well you've probably got the idea by now. So now we have our revised figure 1B. Not as flashy perhaps but the reader's brain has fewer distractions and easily grasps a more accurate impression.

But there is still more work to do to minimise wasted ink. The horizontal axis is completely redundant and the vertical axis could perhaps contain the actual values for the columns. Furthermore, using the minimum line weight possible for the axes gives higher
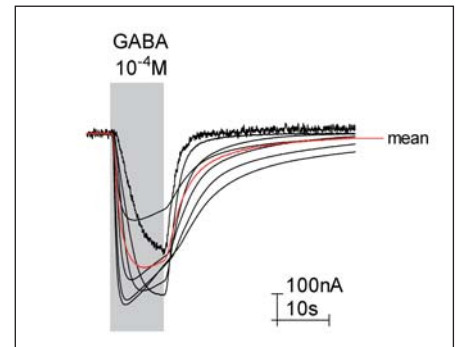
prominence to the data, as does reducing the line weight of the column borders. Our new version of the figure (figure 1C) is now much simpler with more data available to the reader than in the original.

You could go one step further and eliminate the axes altogether, printing the actual values in the columns or at the top of each column but you risk aggravating the conservative element in the scientific community! Incidentally, giving two or three figures in the form of a bar chart is rarely justifiable: a simple statement in the text is usually much better.

Sometimes more is more

Remember, it is the ink:data ratio we want to minimise, not the amount of data. That is why figure 1C is still unsatisfactory, despite the removal of junk. Add more information to the figure if it will help the reader understand what you are saying. For instance, if figure 1 were a comparison of a new drug treatment with controls, why not include data from comparable studies on other drugs? This will help the reader put your findings into context (depending on your findings, you may be tempted not to do this – good figure preparation is a test of your honesty!). We can apply this to our original figure to obtain figure 1D.

Raw or processed data?

Transparency is a vital quality in presenting scientific data. It distinguishes the scientist from the salesman. Authors in-
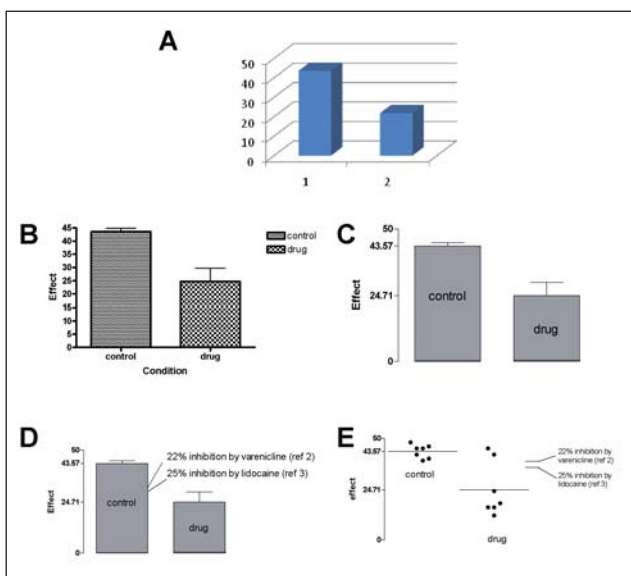

Figure 1

stinctively opt for presenting means plus/minus standard error of the mean but that is not always the best option. Why not present the raw data? Dot plots are an excellent way of presenting scattered data, providing the points are not too numerous, and can reveal a lot of information that is often hidden in means p/m sem. In figure 1E, the data in figures 1A-E are presented as raw data points. While the means-based plots in figure 1A-E reveal a clear difference between the values, plotting the raw data reveals that there is also an important difference in the way the points are distributed. Now we can clearly see that the differences between the two groups reflect strongly on some individuals and have a weak effect on others. Plotting the data as means concealed an important feature of the data, which might otherwise have led to new experiments.

There are many cases where inventive ways of presenting raw data can provide high-density information to the reader in a simple, clear and effective manner. For instance, consider the common practice of presenting a mean finding alongside a "typical" result: figure 2 is a representative electrophysiological response of a GABA receptor that illustrates a statement in the text that "responses were 352 ± 130 nA in amplitude". Instead of the common practice of presenting one representative trace, this alternative figure shows all the responses in
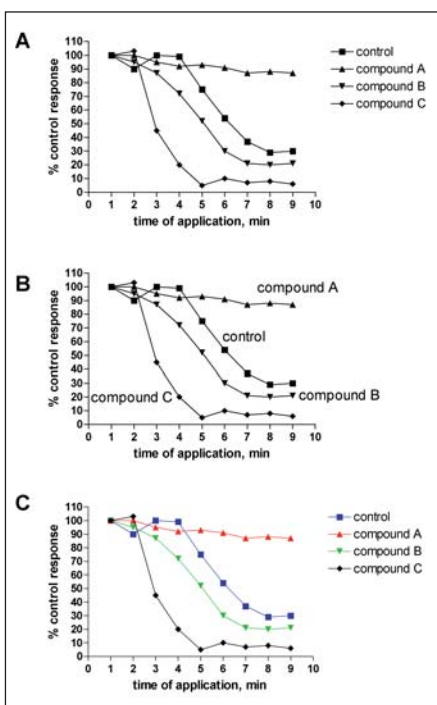
Figure 3

the study, with the mean trace clearly highlighted. The traces are easy to see, the timecourses of all responses and the spread of the data can be easily grasped; considerably more information is available for critical examination by the reader and there is less scope for careful selection of the "typical" result by the author. Incidentally, in this figure the traditional bar representing drug application is replaced with a box behind the traces that enables the reader to see how the traces correspond with the exact points of drug on and off.

Some datasets are too large to be presented in these ways and in many cases it is the summary of the data that becomes the focus of interest. Here, mean $\pm$ sem may indeed be appropriate, but again, there are alternatives with even greater affordances. A neglected but powerful alternative to the simple bar plot is the box-and-whisker plot. The interquartile range of the data (25% to the 75% range of values) is represented as a rectangle, the total range by a whisker and outliers (values more than 1.5 standard deviations from the mean) are excluded from the rectangle but indicated as points. A line marks the median (or other measure of centrality, if preferred).

Label the figure, not the legend

Data sets like that illustrated in figure 3 are often seen in papers, with the key only provided in the legend. The reader has to look at the figure, then scan through the legend for the key to the datasets, then go back to the figure, all the while having to keep in mind what the symbols stand for. This places an unnecessary strain on the mind, which should be reserved for analysing the data in the figure. This situation can be improved by placing the key on the figure itself (fig 3A) but the reader still needs to scan between the data and the legend – so why not label the figure itself (fig 3B)? Many figures do this but then defeat the object by using abbreviations instead of words, thereby thwarting the purpose of marking the figure directly.

Using colours or shading to distinguish columns in a bar chart must be done with care. It can result in charts becoming so complicated that too many attention shifts from key to figure make the chart almost unreadable, leaving the mind with too little "RAM" for analysing the data. Again, this may be overcome by labelling the figure directly. Where a key is preferred, however, colour (if you can afford it – many journals
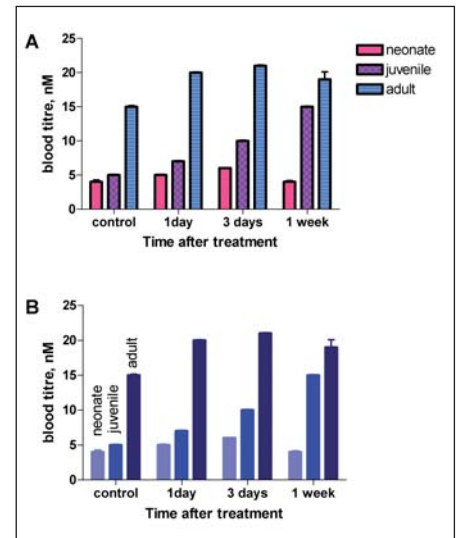
Figure 4

charge for this) can be used effectively to "bind" the key to the chart (figure 3C).

Edward Tufte offers the following simple advice when it comes to colours and shading: choose the minimum difference that clearly distinguishes components of the figure. Compare the two charts in figure 4: the use of closely matched colours adds to the unity of the figure but jarring colours have the opposite effect. In addition, in keeping with the ink:data ratio principle, the author should choose between hatching or colour to distinguish the columns.

In short

Telling the story through figures, reducing ink:data ratio, using as much raw data as possible, direct labelling and minimum effective distance: preparing good scientific figures is an art that can be built upon by following these simple guidelines. No doubt there are other rules that should also be observed. The examples I have used only cover a limited type of the great variety of figures used in research papers and reports but the reader can extrapolate these principles with a little imagination. But perhaps I should now take my own and cut my ink:data ratio by boiling it down to this: be creative, do the readers' work for them.

STEVEN D. BUCKINGHAM