



Statistical literacy guide

How to read charts

Last updated: September 2007

Author: Paul Bolton

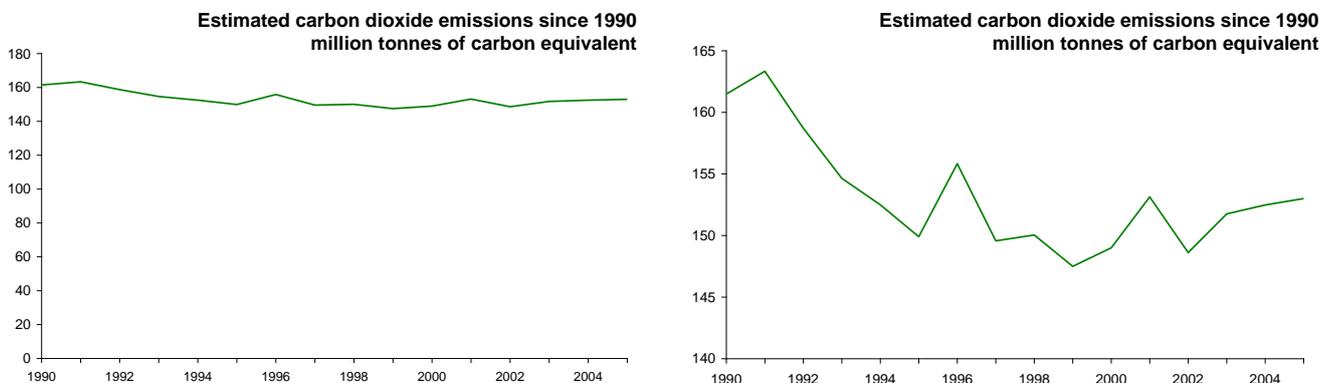
A well designed chart clearly illustrates the patterns in its underlying data and is straightforward to read. A simple chart, say one with a single variable over time, should make basic patterns clear. A more complex one should perform fundamentally the same function, even if it has three or more variables, uses a variety of colours/variables or is made up of multiple elements. In simple time series charts these patterns include the direction of change, variability of trends, size of change and approximate values at any one time. In other types of chart these patterns might include the contribution of different parts to the whole, the scale of differences between groups or areas, relative position of different elements etc. Such patterns should be clear to anyone with at least basic numeracy skills. The understanding is based on the broad impression that such charts give as well as conscious analysis of individual elements of the chart.

This guide is not, therefore, aimed at such charts. Instead, listed below are some common examples of charts where, either through design/type of chart, or the type of data chosen, basic patterns are not always clear to all readers. Some suggestions are given to help the reader better understand the data in these cases. These are summarised at the end of this note.

Shortened value axis

A feature of many charts, especially in the press and on television, is a shortened value axis. This is where **instead of starting at zero the axis starts at a greater value**. The result is that the variations shown in the chart are magnified. So, for instance, a chart of recent FSTE 100 values appears to show a series of dramatic crashes and upswings, whereas in fact the changes shown are only a few percentage points from peak to trough.

The charts below illustrate the same process. The one of the left has the full value (vertical) axis; the one on the right is shortened. It is clear the second gives much greater emphasis to a series that varies from its first value by less than 8%.



The first step to better understand the data is to spot a shortened value axis. In some cases it will be indicated by a zig-zag (\approx) at the foot of the axis. In others the only

way is to **carefully read the axis** values. The next step is to work out the relative importance of changes shown. This may involve concentrating on the actual changes and/or calculating approximate percentage changes. In the second chart above this could be done by reading off the values and calculating the decline up to 1995 was less than 10% of the 1990 level and there has been little change since.

Charts that give changes only

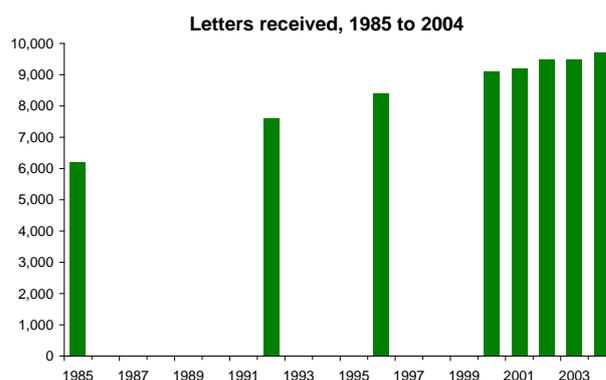
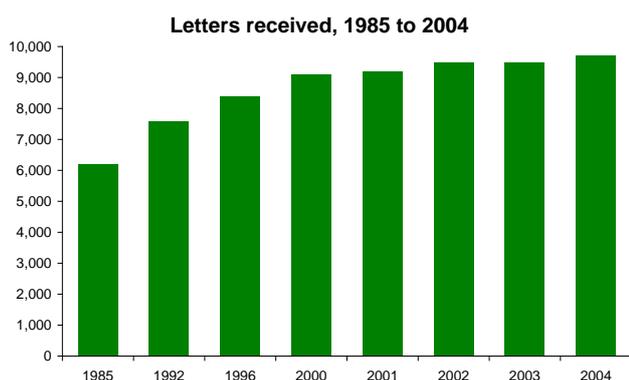
A number of different charts give rise to broadly similar issues. Charts that only look at percentage changes give a partial picture of trends and overemphasise proportionate changes –as with a shortened value axis. In addition, they separate trends into their shortest time period thus concentrate reading at the elementary level (one point to the next, rather than points further apart). This may be useful if the aim of the chart is to do with trends in the rate of change, but as percentage changes are compounded it can be very difficult to work out the overall change over the period shown, or even whether the overall change was positive or negative (if there are increases and decreases). The **underlying data** would be needed in most cases to draw any conclusions about trends in absolute values.

Index charts

Charts that look at **index values** are similar, although more can be concluded about overall changes. The chart title and/or axis should identify that the series is an index rather than actual values. These charts may compare one or more indices over time and concentrate on relative values. If the reader looks at values for individual years they show how the series compares to its base value, not its actual value at that time. If two or more series are used it is important not to conclude anything about the values of different series from the index chart alone. If the series A line is above the series B line it means that A has increased more than B since the base year, not (necessarily) that A is greater than B. The guide in this series on [index numbers](#) gives more background.

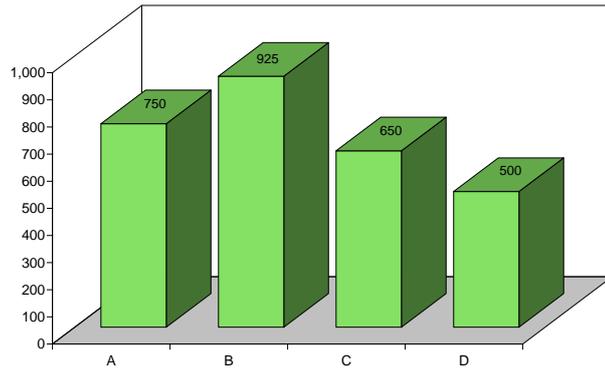
Compressed category axis

Time series charts that present data from irregular dates as if they were regular effectively **compress the category (horizontal) axis**. The charts below illustrate the impact this can have on the pattern shown by the chart. The example on the left condenses the categories and appears to show a levelling off in recent years. The chart on the right shows how it would look if it were presented with appropriate gaps between years –a fairly constant increase in numbers over time. It also shows the 'holes' in the data. Again this needs to be identified by **carefully reading both axes**. It will not always be possible to picture the 'true' pattern after this has been identified; especially if a number of gaps have been 'condensed' (as below). But identifying this does indicate that the pattern shown may not be the whole picture. In such cases it may be necessary to **get the underlying data** where possible, or estimate its approximate values where not.



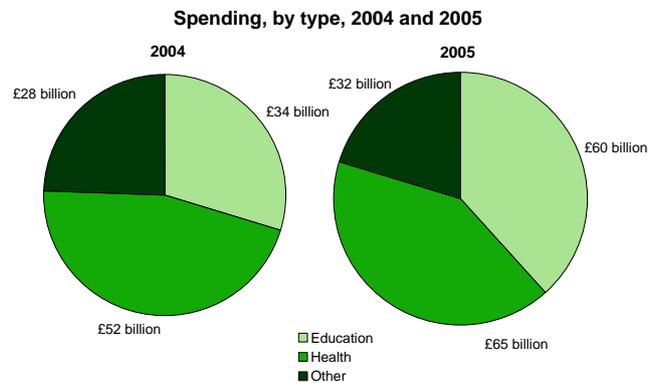
3-D charts

The chart illustrates one of the main problems with 3-D charts; it can be difficult to tell what the underlying values are. The point of a chart is to illustrate patterns rather than to give precise values, but in this example it is difficult to judge within 10% of the actual values. Moreover, the **3-D effects distract from the underlying patterns of the data**. There is no easy way around this problem, but if the chart has gridlines these can be used to by the reader to give a better indication of values.



Multiple pie charts

Pie charts are commonly used to look at contribution of different elements to the whole. They are also used to compare these shares for different groups, areas or time periods. The example opposite compares two different time periods. Here the relative size of two areas (health and other) were smaller in 2005 than in 2004. This is despite the fact that both their absolute values increased. Their share fell because the size of education increased by a greater proportion. Pie charts are better at focussing on proportionate contributions from a small number of elements and may give the wrong impression in a case such as this.

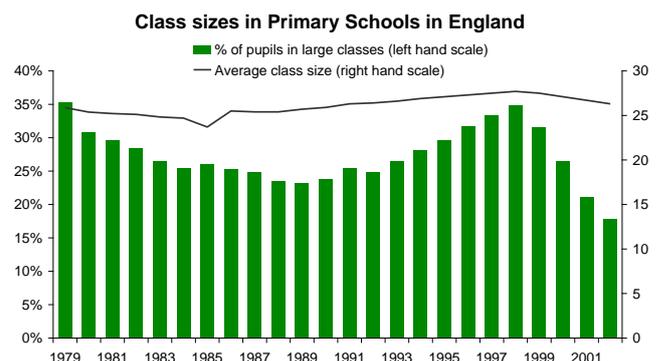


In some charts the size of the second pie may be changed in proportion to the difference in the total. Even in these cases the pattern of the underlying data may not be clear as people are generally less good at perceiving differences in area than in alternatives such as length or position on a line.

Again the key to better understanding the data is to identify this problem by **reading the values**. If only percentage figures are given a total may be added to the chart, if not, there is no way of spotting this and the chart is misleading. If you know the values or total then some assessment of relative changes can be made, but this is not always straightforward.

Charts with two value axes

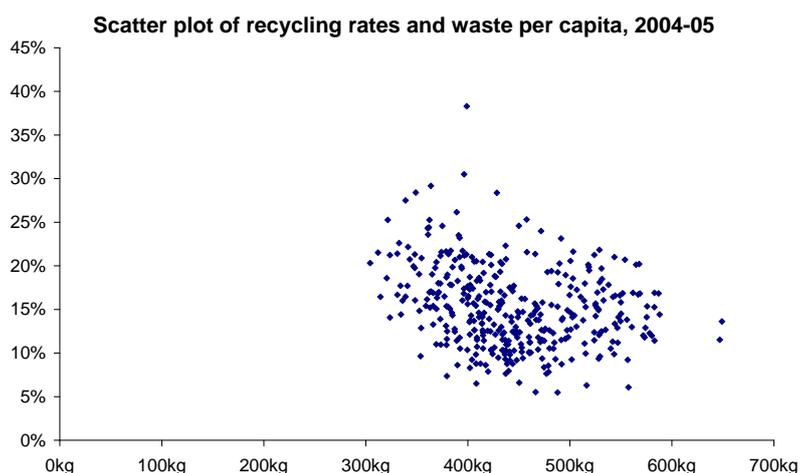
Charts with more than one series let the reader compare and contrast changes. Two or more separate charts can make the series difficult to compare, but it is not always possible to present these data on a single axis, for instance when one is on absolute values and the other is a percentage figure. The example opposite shows such a case. The absolute data is plotted on the right hand axis, the



percentage figures on the left hand one. The key to interpreting such charts is to first to **identify which series is plotted on which axes** (legend keys and clearly distinguishable series help in this case) and second to **look at relative patterns rather than absolute values**. In this example it is of no importance that in all but one year the line was above the column values. It would be very easy to change the axis scales to reverse this. The main message in this chart is that the percentage series is more volatile and since 1998 has fallen dramatically while the absolute series has fallen only marginally.

Scatter plots

Scatter plots are normally used to look at the **relationship between two variables** to help identify whether and how they are associated. They can also help identify 'outliers' from a general pattern. A visual display like this can only give a **partial picture** of the relationship, unless the association is very strong (all the dots form a line) or very weak (dots distributed evenly). In practice virtually all scatter plots will be somewhere in between and a purely visual interpretation of the data can be like reading tea leaves.

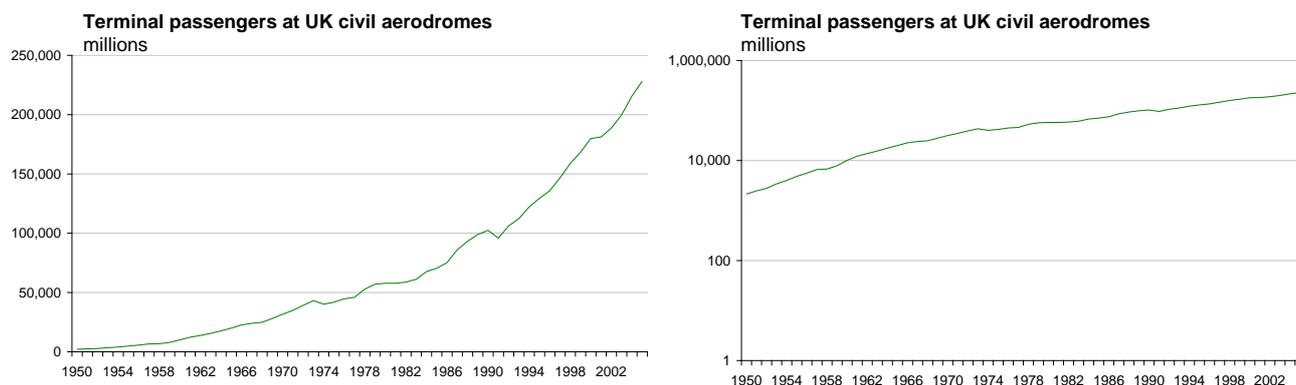


The example above might appear to show a negative association (one indicator increases when the other decreases) to some readers, but little or no association to others. The only way to tell definitively is to **read any accompanying text** to see whether the relevant regression statistics have been calculated. These should include the slope of the line of best fit, whether the value of this is significantly different from zero and the r-squared value. These tell the reader how changes in one variable affect the other and the strength of any association. In the example above a 10kg increase in waste per capita was associated with recycling rates that were 0.2% higher. While the slope of the line of best fit was significantly less than zero this simple model does not explain much of the pattern as only 7% of the variation in one was explained by variation in the other. The guides in this series on [regression](#) and [confidence intervals and statistical significance](#) give more background.

Logarithmic scales

Occasionally charts appear that have a value axis that seems to miss values- instead of, say, 10, 20, 30 etc. they progress 1, 10, 100, 1,000 etc. These are known as logarithmic scales. These can be useful in interpreting **proportionate changes** in series that show long-term growth or decline and are also used in analysis of various natural phenomena. In such charts a conventional or linear axis would overemphasise the proportional impact of changes in an increasing series and underemphasise those in a decreasing series.

The two charts below illustrate the same data, but the one on the left uses a conventional value axis, the one on the right a logarithmic axis.



The chart with the conventional axis clearly shows the growth of aviation and the fact that *absolute* increases in passengers were larger in the second half of the period. It is difficult to draw any conclusions about rates of change during this period.

In a chart with a logarithmic scale a series that has a constant rate of change (whether this is an annual doubling or an increase of 1% a year) will be shown as a straight line. If the rate of change is increasing then the slope of line will increase or bend towards the vertical. If the rate of change declines the slope will flatten or bend towards the horizontal. A fall in the absolute value will still be shown as a downwards slope. Therefore we can see in the second chart that the rate of increase in air passengers was fairly constant up to the early 1970s. Since then it has increased at a slower, but still noticeably steady, rate.

It is more difficult to get an accurate impression of absolute values in charts with logarithmic scales. Their value lies in their presentation of proportionate changes. Again the key to understanding the data is to **read the axis** and hence **spot the logarithmic scale**. You should then concentrate on proportionate changes rather than absolute changes/values.

Summary

- Look for any unusual features in the chart by closely studying its constituent elements:
 - Title, legend, value axis, category axis, units, data values
- Pay particular attention to gaps in axes/uncommon scales
- Any unusual feature(s) may mean you have to adapt the broad impression you initially get from the visual elements of the chart and limit the type of inferences you can draw
- Read any accompanying text before making any firm conclusions
- If all else fails try to get the underlying data