The effective display of information

We are constantly bombarded by graphs in newspapers, magazines and journals. There is an assumption by those who use these graphs that the meaning is transparent. In practice, this assumption may be a long way from reality. Students need to be able to read, understand and interpret graphs and draw conclusions about the information presented in them.

There has been a great deal of discussion of late about the impact of an ageing population on Australia's economy. As well as talk of the increasing demands on health care, many people are concerned about outliving their retirement nest egg. Financial advisers frequently use information on predicted changes to the population to encourage people to invest for their futures. As an example, the following graph appeared in a financial planning brochure in 2006 as part of an article titled "Thinking about your future".

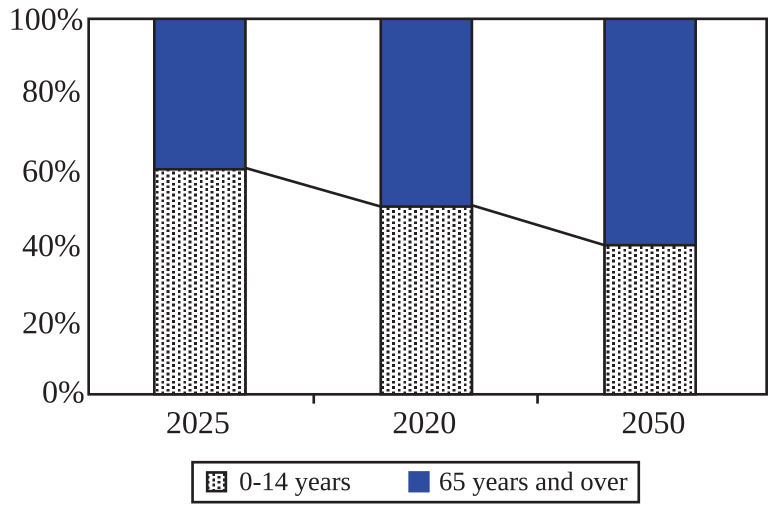


Table 1 Australia's population, prediction by age group
Australian Social trends, 2006, 4102.0, Australian Bureau of Statistics,
http://www.abs.gov.au/ausstats sourced 21 July 2006.

The text of the article included the following: "What is more concerning is the balance between young and old people. As Australia's population is ageing, the proportion of younger people is shrinking." The graph that was labelled Table 1 appeared to add weight to this statement. Yet if you consider what the graph is displaying carefully, there is at least one obvious error on the graph. The use of "years" on the horizontal scale is meant to produce some sense of a trend. Yet the years on the horizontal scale are not in sequence[1]. Moreover, there does not appear to be a suitable scale on the horizontal axis, as the gap between each column does not represent the same number of years. Having equal spacing on the time axis is important if you want to gain a sense of the rate at which change is happening.

The more that you look at the graph the more other questions come to mind. Percentages are useful in making comparisons, as long as you know what the percentages refer to. The vertical axis goes from 0% to 100%. This is normally quite comforting as it means that you haven't left something out or counted things twice. Unfortunately, when you examine the legend of the graph, quite a number of people do appear to be omitted. As the intent of the graph is to compare the young and the old, people between the ages of 15 and 64 years are not represented on the graph. Yet this creates a problem when trying to make sense of the

[1] When the quoted source of the data is checked, the year 2025 should have been 2005. The typographical error does not influence the other problems evident in the graph.

data. Currently two-thirds of Australia's population is aged 15–64 years. This group, which encompasses almost all of the working population, is projected to decline to 59% in 2051. You cannot tell this from the graph (Table 1), which at first sight appears to suggest that about 60% of the population could be over 65 years by 2051.

The intent of using the graph (Table 1) was no doubt to convince the reader that the proportion of people over 65 years will increase while those under 15 will decrease. Although this is likely to be true, the design of the graph does not represent the data in a way that supports an understanding of trends. That is, you cannot accurately gain a sense of how rapidly this change will take place, or the magnitude of the change. Without knowing what 100% refers to on the graph, it is not possible to determine what changes in the proportions might mean.

Choosing and using graphs

Graphs are commonly used to persuade or convince. Even before the statement "a picture is worth a thousand words" became a cliché, graphs have been considered a highly effective way to visually communicate information. The construction and interpretation of graphs is also one of the most substantial examples of numeracy across the curriculum. Yet graphs are also subject to frequent misuse and abuse.

By the time they enter high school, students have learnt about picture graphs, column graphs, line graphs, divided bar graphs and sector graphs, also known as pie charts. This is quite an impressive array of data displays. Different types of graphs are used for different purposes and audiences. In CAFS the types of graphs that students need to be most familiar with include line graphs, bar graphs and divided area graphics, such as pie charts. The ability to accurately read and interpret data presented in graphs is particularly important in Stage 6 CAFS. Graphs are essential for presenting much of the data collected for the Independent Research Project. Many of the higher order multiple choice questions will require students to not only interpret information presented in a graph but to also draw conclusions based on this information.

When to connect the dots

Graphs often link together two pieces of information. This can be achieved by plotting points that represent the junction of the two pieces of information. The two pieces of information are sometimes referred to as two variables, or bivariate data. If the purpose of a graph is to look at the way the data changes, the dots are joined to create a line graph. Line graphs are very effective ways of presenting information. For example, a line graph is commonly used to show how two things are related. When points have been plotted it is very tempting to join them up. However, it is not always a good idea to join the dots. Say you plot the batting averages of a cricket team against the order in which the batsmen play. Although you might expect to see a relationship between batting averages and order of play, the dots should not be joined with a line. Line graphs should only be used where meaning can be attached to the points on the line between plotted points. As there can be no batsman between the third and fourth batsmen, there should not be a line connecting these two points.

The most important initial criteria that should guide students in choosing a graph type is the type of information (variable) being displayed. That is, what type of data is it, and, in particular, does it measure a continuous quantity? As students learn to distinguish between types of variables used in graphs, they also learn to identify misrepresentation of data in graphs.

Are pie charts always a good choice?

Sector graphs, also known as pie charts, are very common in annual reports of organisations. The frequency with which they are used suggests that they must be easy to interpret. This can be true if it is clear who gets the biggest piece of the pie. Pie charts are often used more for visual appeal than as the best choice of graph.

The main limitation of pie charts is that the visual basis of comparison used, that is, how you read the graph, is the size of the angles in each sector. It was recognised over 100 years ago that people have difficulties making judgements about the size of angles. Modern research into neuropsychology tells us that humans overestimate acute angles (angles less than 90°) and underestimate obtuse angles (angles greater than 90°). This in turn impacts on our capacity to interpret information based on visual estimates of angle size. Look at the basic pie chart in Figure 1. Can you tell who gets the biggest piece of the pie?

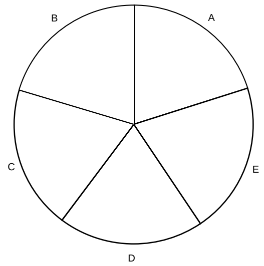


Figure 1 A sector graph representing numbers with angles

Each piece of pie is different, yet it is extremely difficult to visually compare the size of the sectors. Many sector graphs in annual reports also contain numerical data which shifts the interpretation of the data away from estimating angle sizes.

If the same information is presented in a column graph (Figure 2), the comparison between categories is easier because it is based on comparing heights using a common baseline.

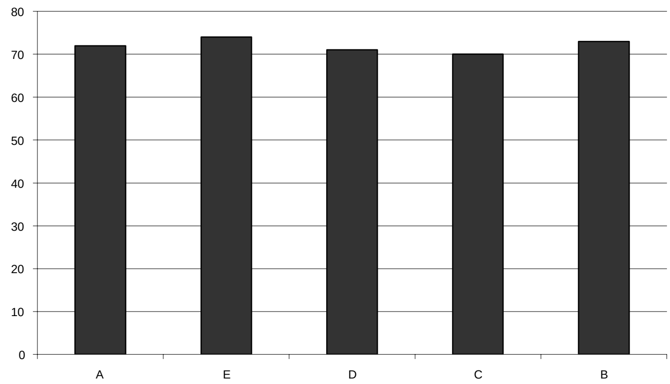


Figure 2 A column graph with the same data

Comparing position along a common scale is the basis of many effective graphs including histograms, column graphs and scatterplots. Graphs that can be read by judging position along a scale are much easier to interpret correctly than those that require you to compare areas or angles. If the parts of the total that you want to display using a pie chart are close, a pie chart is not a good choice of graph.

Selecting the right type of graph
Stem-and-leaf plots (Figure 3) are the most recently developed ways of displaying data.

1	8 9
2	3 5 5 6 6 6 8 8 9 9 9
3	0 0 1 1 1 2 3 3 4 4 4 5 5 6 7

Figure 3 Noon temperatures (°C) in February

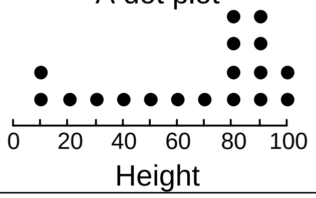
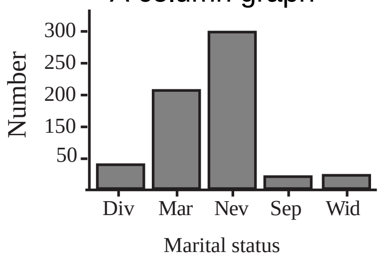
Stem-and-leaf plots resemble column graphs, with the added advantage that you can see all of the scores. If you had a lot of scores to graph you would use a column graph or a histogram rather than a stem-and-leaf plot.

Following some simple guiding principles can reduce the risk of misrepresenting data. Before you construct any display of data, decisions have to be made about appropriate ways of presenting the data.

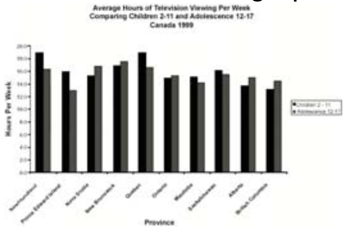
Some graphs are clearly appropriate for displaying some data forms and not others. Data fall into two main types, quantitative or categorical. Quantitative data generally tells you 'how much' (e.g. how many litres, what temperature, how high) whereas categorical data tells you 'what kind' (e.g. gender, colour, make of car). Quantitative data can be further subdivided into discrete and continuous quantities. Continuous data always has another value between any two measures. For example, height is a continuous measure whereas the number of children in a family is a discrete measure.

A simple way to determine which type of graph is appropriate to use is to first determine what type of data you have. If you measured the heights of all of the students in your class you would have continuous data in one variable, height. When you count how many items are in a category, you create discrete quantitative data. That is, frequency or how many of each thing, is the main example of discrete quantitative data we encounter.

Representing data
Once you have determined what type of data you have, say continuous or categorical data, which graph types are best suited to graphing the different data types? The following table provides a simplified summary of a variety of graph types and when you might use each type. This is a simplified list as graph types are like text types; new ones continue to be formed and not every example can be allocated to a unique type.

Graph types	**Example**	**When do you use it?**
A dot plot A histogram (no gaps between columns) Stem-and-leaf plot A box-and-whisker plot	A dot plot 	Displaying continuous data in one variable
A column graph Divided bar graph Sector graph	A column graph 	Displaying categorical data in one variable
Line graph Travel graph (distance and time) Scatter plot	A line graph 	Displaying continuous data in two variables

<table>
<tr>
<td>Split graphs:

back-to-back stem-and-leaf plots,
double column graphs,
multiple box-and-whisker plots</td>
<td>A double column graph

</td>
<td>Displaying one continuous and one categorical variable</td>
</tr>
<tr>
<td>Two-way tables</td>
<td>A two-way table
<table>
<tr><td rowspan="2">Grey Hair</td><td colspan="2">Age</td><td rowspan="2">Total</td></tr>
<tr><td>Below 40</td><td>40 or older</td></tr>
<tr><td>No</td><td>40</td><td>5</td><td>45</td></tr>
<tr><td>Yes</td><td>20</td><td>35</td><td>55</td></tr>
<tr><td>Total</td><td>60</td><td>40</td><td>100</td></tr>
</table></td>
<td>Displaying categorical data in two variables</td>
</tr>
</table>

A dot plot, a box plot and a histogram are all well suited to displaying continuous data (in one variable) because the underlying scale is continuous. A column graph is best for categorical data or a divided bar graph or sector graph could also be used. The height of the columns or the proportion of the divided bar graph refers to the number (frequency) in each category.

The method used to graph a continuous variable against a categorical variable is usually to split a single continuous variable graph into two or more plots. We do this, for example, with back-to-back stem-and-leaf plots and with side-by-side common scale box-and-whisker plots.

Male		Female
8	15	4 4 5 5 6 8
9 8 5	16	0 2 2 3 5 5 6
		9 9 9
9 9 8 7 7 5 5 4	17	2 5 5
4 4 1		
5 4 4 3 1 1	18	0 1

It may appear odd to include a table in a summary of graph types but sometimes the best type of graph is a well-prepared table. This is certainly true of representing two categories of information from one group.

Beware of broken scales
One of the major problems newspapers and journals face in using graphs is limited print space. This has led to the common use of a "broken scale" characterised by a zig-zag line on the scale. If the purpose of the graph is to provide comparisons and the broken scale is used with some values but not others, accurate comparisons are not possible.

A scatter plot is an effective way of graphing a continuous variable against another continuous variable. For example, height against age. A scatter plot does not assume a linear relationship between the two variables. Consequently, a scatter plot provides a useful method of looking for relationships in the data. It is possible to fit a line to a scatter plot when the data suggests that it could be modelled by a linear relationship.

As mentioned earlier, not all graphs will fit easily into this table as new types are still being formed and many other names are used for similar graphs. Some graphs also attempt to link three or more pieces of information together (Figure 4).

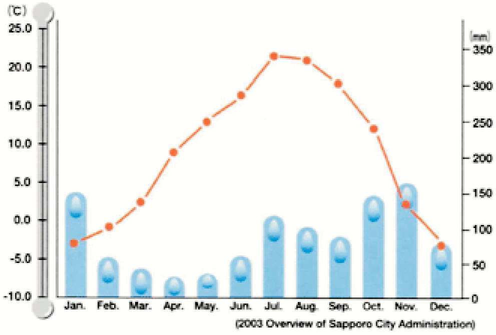


Figure 4 Rainfall and temperature in Sapporo

The graph in Figure 4 uses two graph types (line graph and column graph) combined. It is a good idea to use two visually distinct graph types for this combined graph as the scales and the units used are quite different. One scale is measuring temperature and the other is measuring rainfall. Just as with written texts it is the purpose that determines the text type, so too the selection of graph type depends upon the purpose of effective communication.

If the purpose of a graph is to represent the data in a way that illuminates rather than distorts the information, then students need to be able to do more than label the axes or add a title to a graph. To interpret a graph, students need to appreciate the use of a scale. One of the most common causes of a graph being misleading is the use of an inaccurate scale. Understanding that the type of data influences the choice of a graph is important to developing numerate learners.

To strengthen students' numeracy within CAFS, when you are using a graph in a teaching activity, spend a couple of minutes discussing the choice of graph to represent the information. Why is it a good or poor choice of graph? What makes the graph easy to read or difficult to understand? How would the information look if a different type of graph had been used? Was the choice of scale a good choice? Is the graph in any way misleading?

Thank you to Peter Gould, Manager Mathematics for his contribution to this article.