Working Mathematically with Data

Heather McMaster and Michael Mitchelmore

We live in the information age. In a recent radio national programme it was reported that more information has been produced and stored in the last five years than in the rest of the history of humankind. Data are also more readily accessible than ever before. It is therefore of paramount importance that students have the skills to be able to collect or select the data they require for a particular purpose, and then to interpret and communicate the data effectively in both words and pictures. The Data strand of the syllabus equips students with the ability to work with data and to evaluate critically the use of data by others.

When the new 7-10 syllabus was introduced in NSW schools 2004, it was not surprising that differences from the old syllabus were greatest within the Data strand. Many skills that were previously taught in high school are now being taught in primary school (Stage 3). At the MANSW conference we learnt that by the time year 7 students in 2006 reach year 11, there will be new HSC mathematics courses (Board of Studies presentation). These courses, like the new 7 – 10 courses, could be expected to place a greater emphasis on statistical literacy.

Within each strand of the NSW K-10 syllabus, students are now required to "work mathematically". "Working mathematically" includes the processes of questioning, reflecting, reasoning, applying strategies and communicating. The "working mathematically" approach to teaching has resulted in some important new emphases in the Data strand. Through working mathematically, students learn to make sense of data and use data appropriately and effectively for a particular purpose.

In our workshop at the 2005 MANSW conference, new emphases in the Data strand were explored in relation to outcome DS4.1 "Constructs, reads and interprets graphs, tables, charts and statistical information" (Board of Studies, 2002). To achieve this outcome, students need to know more than just ways of displaying data and reading those displays. The following dot points for outcome DS4.1 were discussed:

- choose appropriate forms to display data *(Communicating)*
- compare strengths and weaknesses of different forms of data display *(Reasoning, Communicating)*
- generate questions from information displayed in graphs *(Questioning)*
- analyse graphical displays to recognise features that may cause a misleading representation *(Communicating, Reasoning)*

.

Choose appropriate forms to display data

In Stage 3, students draw and interpret picture graphs, divided bar and sector graphs, column graphs and line graphs and learn to "determine what type of graph is the best one to display a set of data" (Board of Studies, 2002). In Stage 4, students learn to recognise data as being quantitative (either discrete or continuous) or categorical. Based on this understanding, they are again required to "choose appropriate forms to display data".

In our workshop, participants took part in an activity from *Working Mathematically: Data Stage 4* (McMaster and Mitchelmore, 2005) that required them to solve a problem graphically. Given the thicknesses of two types of paper (newspaper and photocopy paper), they were asked to decide which type would be better to fold and chock under the leg of a desk to prevent it from wobbling. The resulting graph is in Figure 1. In this context, a solid line graph was not an appropriate form of display because of the discrete nature of the data. The activity also illustrated the importance of both mathematical reasoning and practical considerations when modelling a situation. Theoretically, folding up a sheet of photocopy paper was the answer, but in practice, this was not found to be the best solution to the problem.

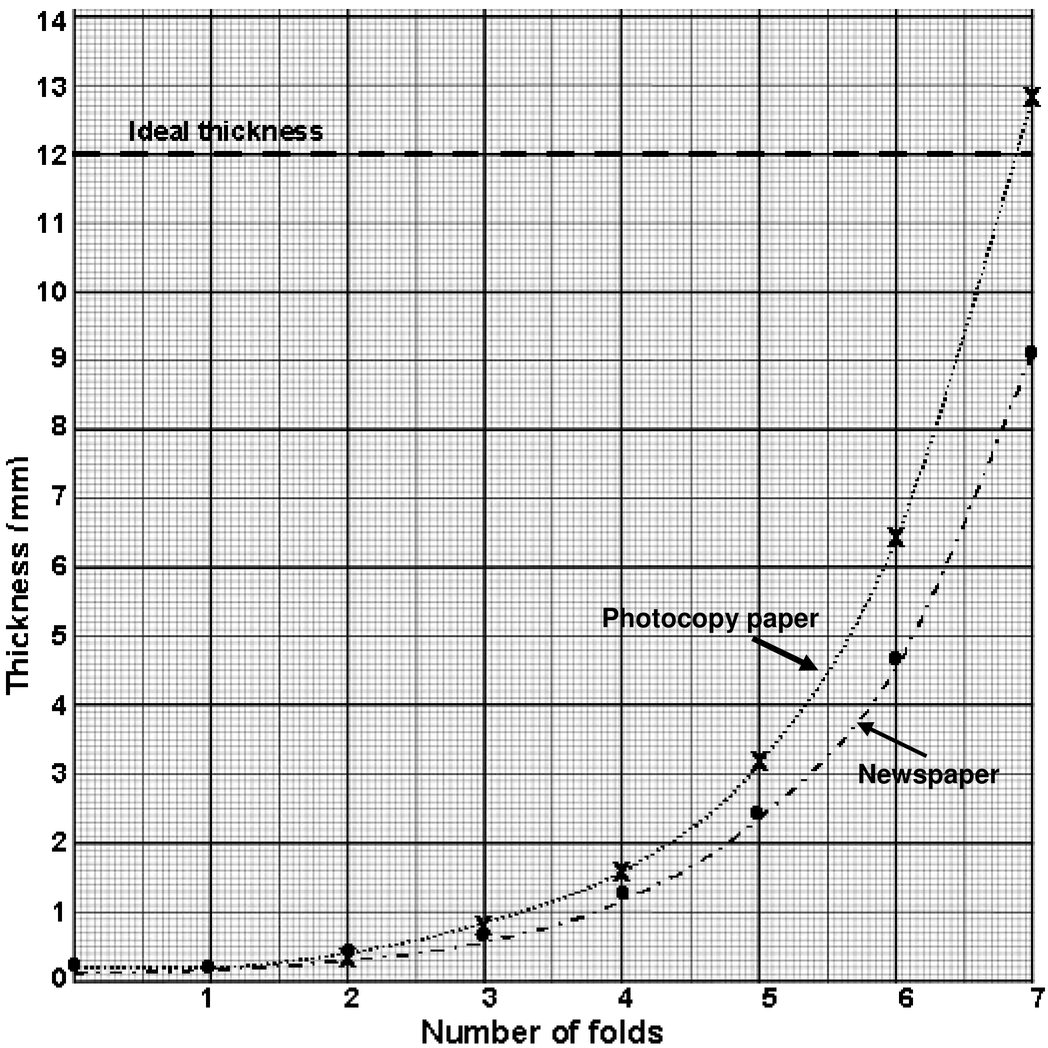

Figure 1. Graphical representation of a paper-folding investigation

Compare strengths and weaknesses of different forms of data display

When choosing an appropriate form to display data, students are not necessarily “right” or “wrong”. There is always more than one appropriate way of displaying data but some ways are better than others, depending on the purpose of the display.

In Stage 3, “advantages and disadvantages of different representations of the same data should be explicitly taught” (Board of Studies, 2002). In Stage 4, students learn about some additional types of graphs, namely graphs that are used to represent frequency distributions and correlations. Again, students are required to “compare strengths and weaknesses of different forms of data display” (Board of Studies, 2002).

This year, a teacher reported to us the results of a task she set for a group of 10 students who had completed Stage 3. They were asked to investigate the relationship (if any) between the span of their hands and the length of their feet. After measuring themselves, they were required to display the data, then write statements about the relationship they found. Their graph is shown in Figure 2.

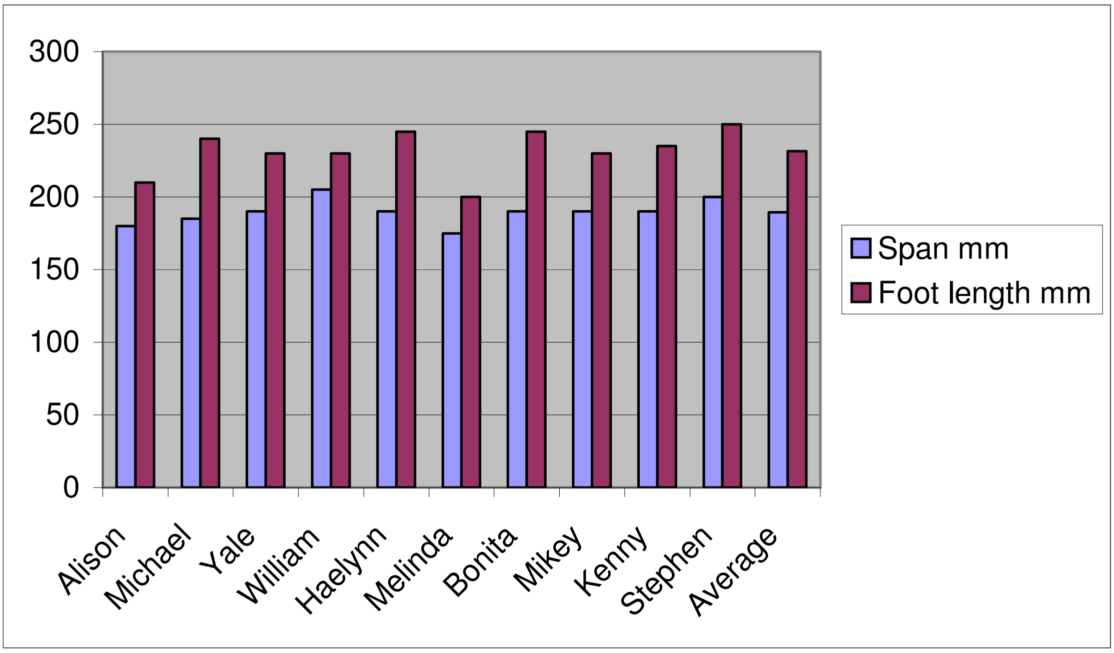

Figure 2. A column graph of students’ spans and foot lengths.

The students obtained this graph by simply putting their data into an Excel spreadsheet, highlighting the table, clicking on the Chart Wizard icon, then choosing all the default options. Having used Excel to calculate averages at the bottom of their table before graphing it, these values were automatically transferred to their graph.

From their display, the students made the following statements:

1) Yale, Haelynn, Bonita and Mikey all have the same span
2) Melinda is the only person to have the average foot length
3) Melinda's foot length is less than William's span
4) 7 people are over the span average
5) Most people have the same difference in their span and foot length
6) Michael, Yale Haelynn, Bonita, Mikey and Kenny have around the same span
7) The shortest span is Melinda's
8) The biggest foot length is Stephen's

As you can see, their statements were all about the individual sets of data or individual cases with the exception of their fifth statement: "Most people have the same difference in their span and foot length". Had they reflected more on the purpose of their investigation, they may have chosen to display their data differently. With a different display, additional information about the relationship could have been revealed.

For example, by ordering categories (in this case, ordering people according to their foot length) their graph would look like the one in Figure 3.

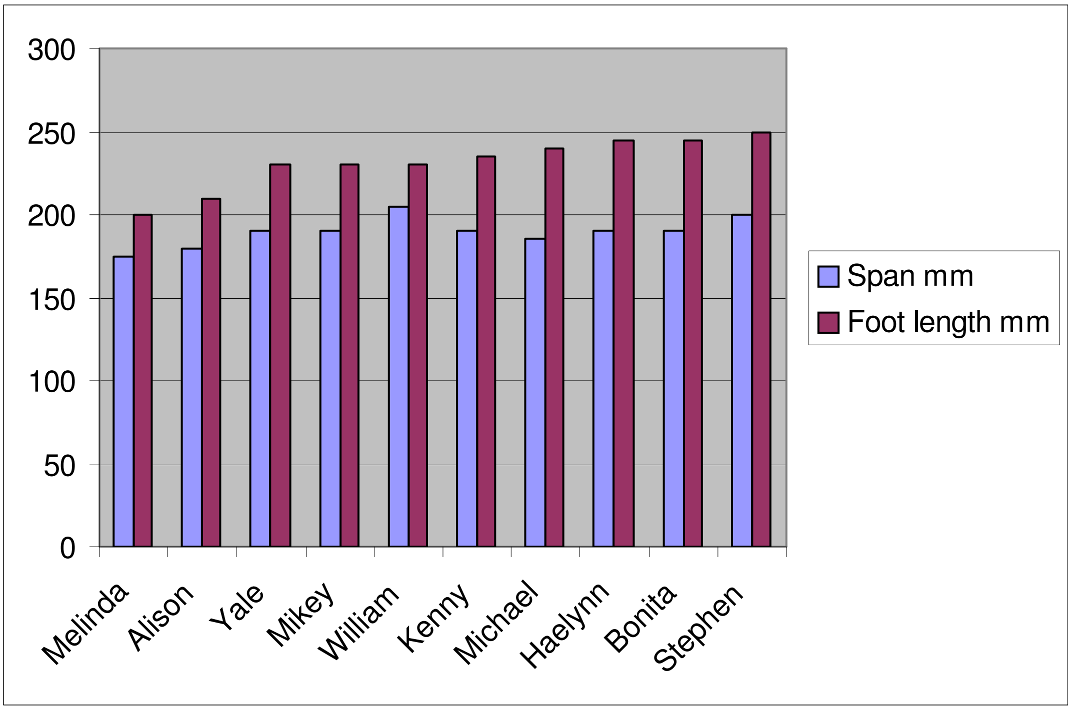

Figure 3. An ordered column graph of students' spans and foot lengths.

If they had then decided to change from using columns to plotting points, their display would have looked like the one in Figure 4. All this would require is the selection of “XY (Scatter)” as the graph type in Excel, rather than “Column”. Plotting points is perhaps a clearer way of showing that in their group, those with long feet tended to have long spans (ie. there is a positive correlation between the data sets).

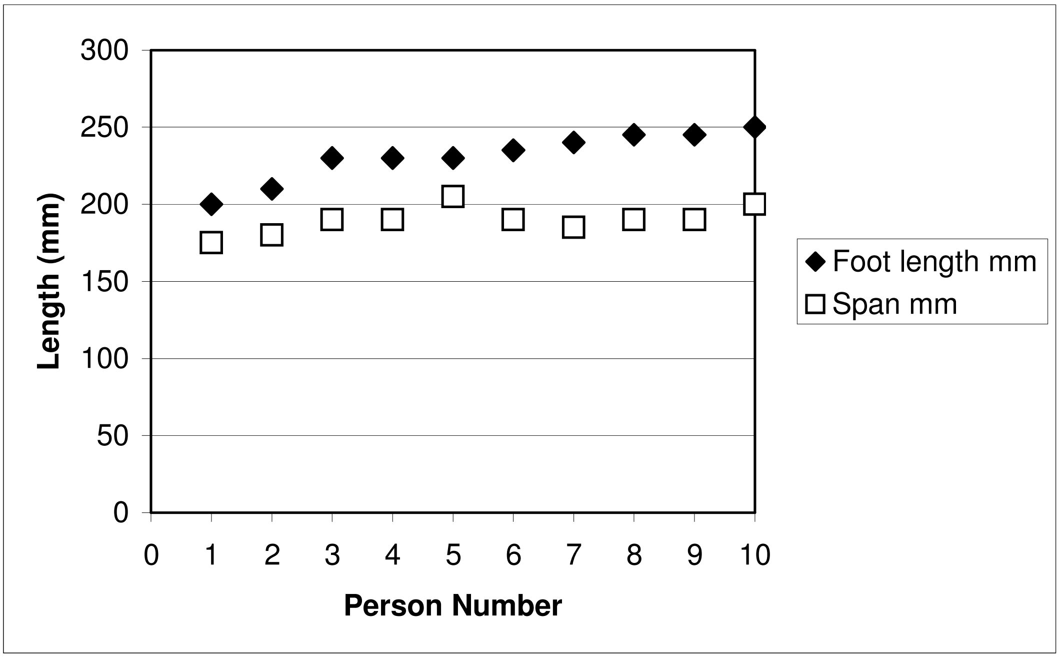

Figure 4. An ordered plot of students’ spans and foot lengths.

After plotting ordered points, the next step could be to make a scatter diagram (a Stage 4 outcome) as shown in Figure 5. This diagram is made in Excel by highlighting only the columns of foot length and span data (not the column of people’s names) then selecting the graph type “XY (Scatter)”. The names of the students could be added later as labels.

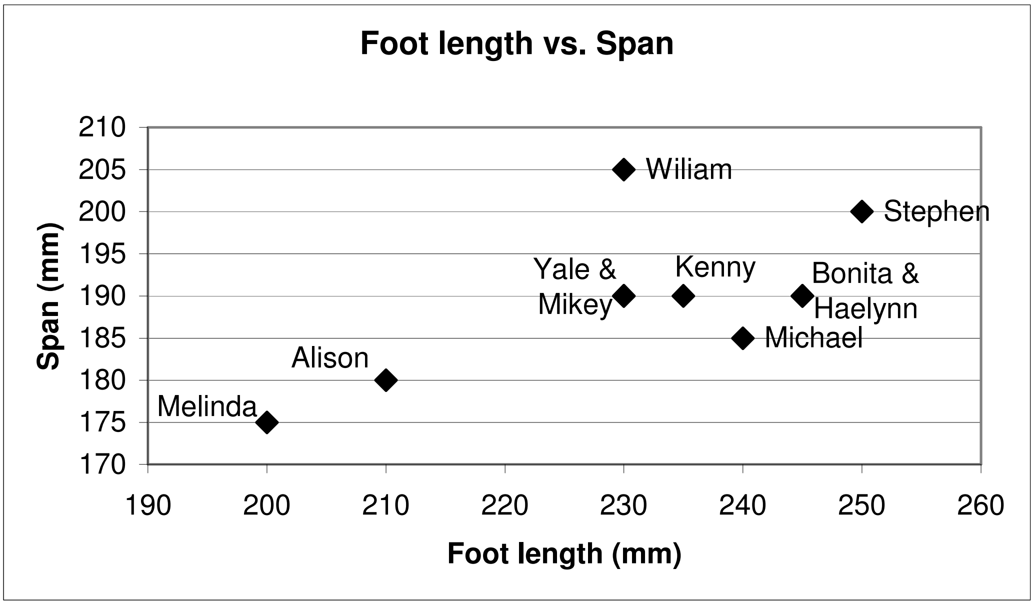

Figure 5. A scatter diagram of students’ spans and foot lengths.

This process of changing representations of data to help in its interpretation is called “transnumerative thinking” (Wild and Pfannkuch, 1999). It is a critical component of student’s statistical reasoning. If the students had drawn a scatter diagram of their data, they may have questioned the outlier (William) and checked his measurements. The correlation between foot length and span can then be made more obvious by drawing a line-of-best fit. This can be done in Excel by highlighting the data series, then selecting “Add Trendline” from the pop-up menu.

Generate questions from information displayed in graphs

Questioning is one of the five working mathematically processes and a vital aspect of statistical literacy. Students should be encouraged to become “statistical detectives”. There are many types of questions that can be asked about data displayed in a graph:

- Context questions eg. Where do the data come from? Are the data reliable?
- Prediction questions eg. Based on the data we have, what would we expect…..?
- “What” questions eg. What is the data telling us about…..?
- “How“ questions eg. How is…... related to…..?
- “Why” questions eg. Why is the graph that shape? Why is that case an outlier?

Great insights and scientific breakthroughs have come through asking questions like these.

Students will naturally question data if they know the data are real or interesting or important to them eg. data from CensusAtSchool provided by the Australian Bureau of Statistics (ABS). The ABS website (http://www.abs.gov.au) also has a large amount of free up-to-date data about Australia and its people. Publications and resources available include *Year Book Australia* and *Australian Social Trends* along with statistics on migration, education, health, the environment and the labour force.

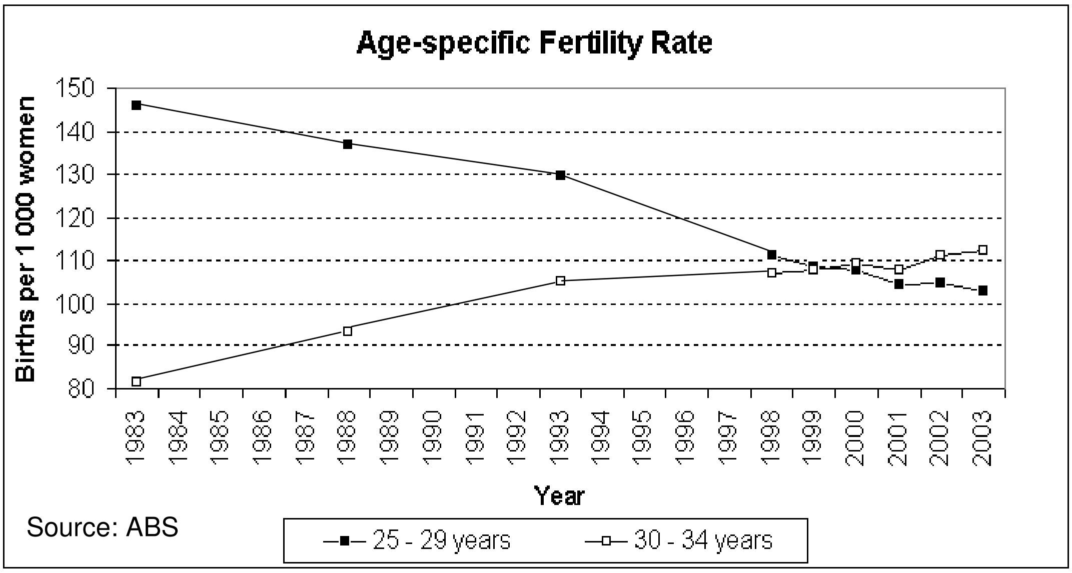

Figure 6. Age specific fertility rates in Australia, 1983 - 2003

Some ABS data are displayed in Figure 6. The negative correlation between the data sets could provoke some interesting questions about predictions and social trends.

The data in Figure 7 come from the United Nations Food and Agriculture Organization (FAO), an organisation that also supplies extensive up-to-date data sets free-of-charge on the internet (http://www.fao.org). The correlation between the area of wheat sown and the quantity of wheat produced in Australia is positive. Students might question why this is so, or why one data set has a steeper long-term trend than the other. If they plot the last 10 years of data as a scatter diagram (Figure 8), they might question why the year 2002 is an outlier.

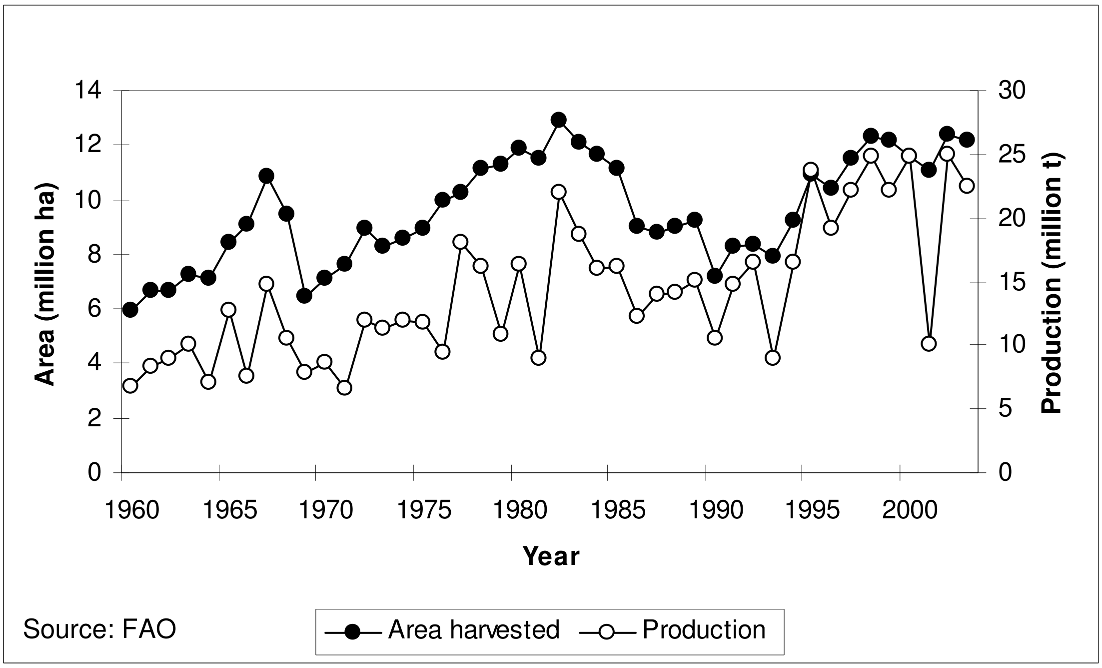

Figure 7. Area and production of wheat in Australia, 1961 - 2004

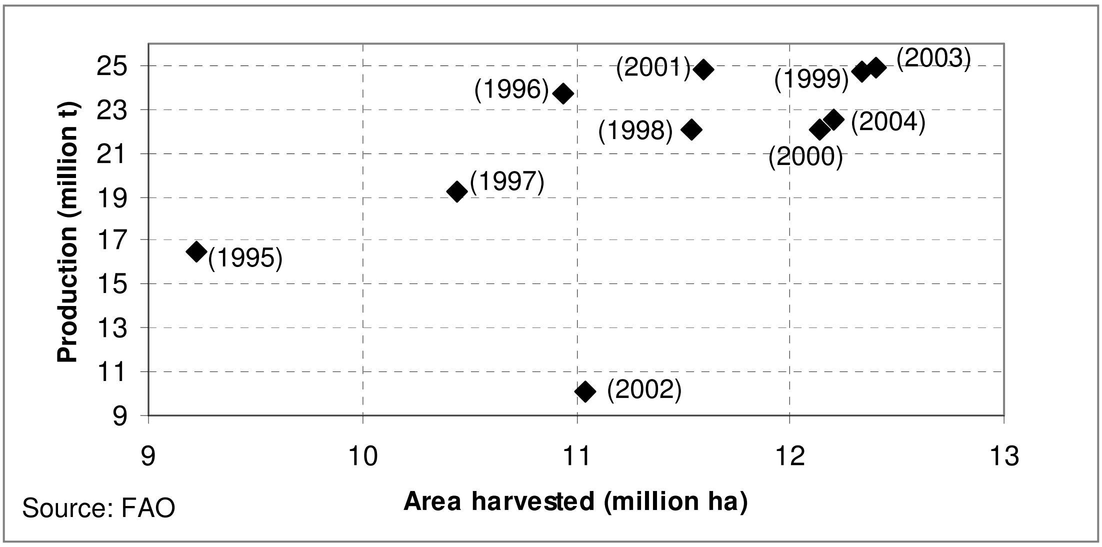

Figure 8. Area of wheat harvested vs. wheat production Australia, 1995 - 2004

Figure 9 comes from the internet site http://www.venganza.org. Students could question whether this data is reliable and what it is saying. They could also reason why, in the context of a debate on the greenhouse effect, they think someone would want to use this graph.

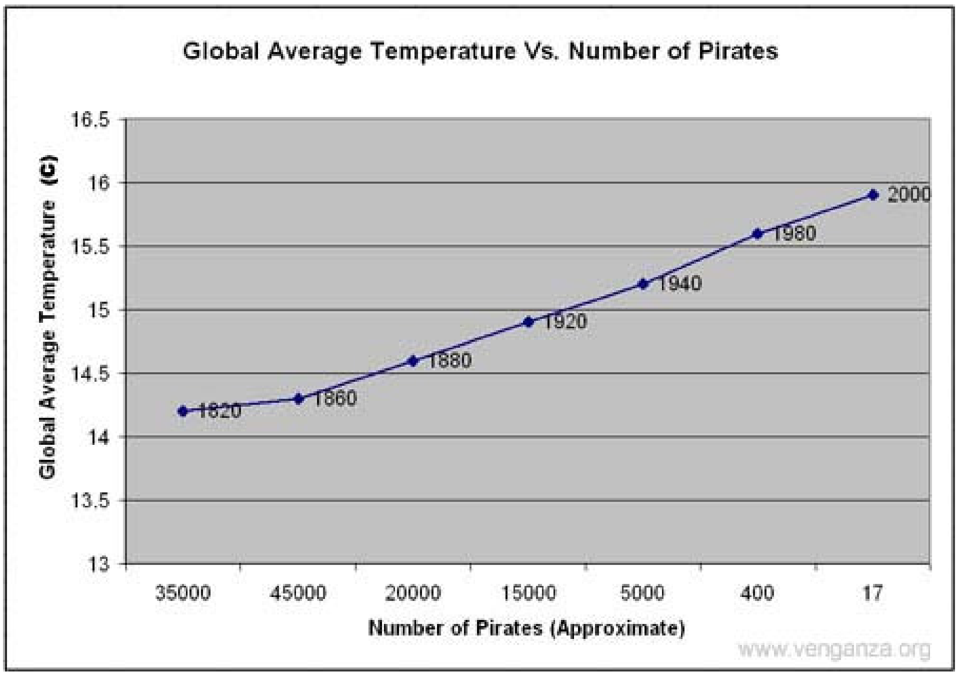

Figure 9. A graph from http://www.venganza.org

Analyse graphical displays to recognise features that may cause a misleading representation

Graphs can be used to win an argument, promote an agenda or sell something. The graphs used for these purposes may have accurate data and be statistically legitimate (unlike the graph in Figure 9) but be misleading. Students need to be aware of the devices that are used so they are not deceived.

The common forms of misleading representations are dimensional effects (eg. using 2 or 3-dimensional pictures instead of columns), selective labelling on axes (eg. non-zero starts, missing segments of an axis or irregular scales), scaling effects, selective data (eg. the choice of a beginning date when showing a trend in a time series) and various visual effects such as the choice of patterns and colours. Students should be careful reading pie charts (ie. sector graphs), particularly when they are 3-dimensional. Angles of sectors are more difficult to compare visually than lengths and they are changed when the graph is drawn with a 3-dimensional effect. 3-dimensional pie charts and exploded pie charts are now commonly used because they can be drawn with Excel.

Misleading graphs are frequently found in advertisements and articles in the media that put a certain spin on the data. Newspapers and magazines are therefore a rich source of “working mathematically” resources for your students. Teachers can have students collect some graphs and discuss their misleading features. Talented students can have some fun designing their own misleading representations of data.

Conclusion

Statistical literacy is becoming increasingly important in our society. It is a significant component of almost every field of tertiary study (business and economics, social sciences, medicine and public health, environmental studies, law, journalism etc.).

The working mathematically processes of questioning, reflecting, reasoning, applying strategies and communicating are essential in the Data strand. The learning in this strand, more than any other, needs to be based on real life contexts and data that interest students. Collecting, analysing and representing data that is of no interest to them or which serves no particular purpose, can be extremely tedious and boring. Interesting real life data, on the other hand, stimulates questioning and reflecting. To collect data and analyse it appropriately, students need to reason and apply strategies, keeping in mind the purpose of their investigation. They then need to communicate their data effectively through graphical representations. This is what working mathematically is all about.

References

Board of Studies (2002). *Mathematics Years 7-10 Syllabus.* Sydney: Author.

McMaster, H., & Mitchelmore, M. (2005). *Working Mathematically: Data Stage 4* (two parts). Sydney: Workingmaths.

Wild, C.J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223-248.