calculations can be checked by measuring a plumb-line from the window-sill to the ground below the window.

Concluding comment

This approach demonstrates the following basic principles of pedagogy. The first two of these I distinctly remember from my own teacher education days in 1951.

1. Go from concrete to abstract. Avoid starting with definitions.
2. Go from particular to general. Here, we started with a few special angles before going on to general acute angles and the need for finding the values of the tangent of such angles by construction or by using a calculator or other means.
3. Immerse students in the context of any new concept before explicating its technicalities and intricacies and mathematical jargon. The above approach illustrates the truth that students can be using the tangent function before they have even heard of the term!
4. The lesson introducing a new concept should be one that results in favourable reactions from the students. The Year 10 students shown in Figure 1 dramatised their delight with their first experience of anything trigonometric. Herman Tay who ran their first lesson (while a pre-service student) reported that the boys were 'enthusiastic', 'excited', and 'thrilled'. He went on to report that 'the whole class was awash with enthusiasm, once one group attached the drinking straw to the set-square,' and, 'We were oblivious of the bells at lunch break; nobody was anxious to leave the classroom.'

Cyril Quinlan
Australian Catholic University
c.quinlan@mary.acu.edu.au

SIMPLE for

HELEN CHICK

Most students start their statistical experiences in primary school with simple data handling techniques such as tallies and bar charts. These situations often involve single variable data, so that a typical activity might involve producing a frequency graph showing the favourite football team of students in the class.

Data analysis becomes much more interesting when the data set involves multiple variables. This is because relationships among the variables can be explored. Data exploration now might involve making comparisons and determining the existence of associations. Of course, this complexity in the data brings with it challenges in dealing with the data, to produce the representations and calculations that help identify those relationships and contrasts. In teaching we sometimes leave the study of multivariate data until quite late in schooling because some of the techniques for dealing with such data are deemed too complicated. There are, however, some simple strategies that make such data analysis accessible to younger students. These techniques are probably familiar to us as teachers, especially if we use spreadsheets, and yet often we do not highlight them for our students.

To illustrate this, we will look at the work of some Year 7 students who were asked to consider the data set in Figure 1. The idea for this data set arose from the work of Watson and her colleagues (e.g., Watson, Collis, Callingham & Moritz, 1995). We note that, of course, it is usually better if students collect their own data about a topic of interest, but in this case I wanted to be sure that the data set was not too large and that there were relationships evident.

STRATEGIES dealing with data

There are four variables in the data set, as indicated by the four columns. Two are numerical variables (the number of hours of exercise per week and the number of fast food meals consumed each week), and two are categorical (favourite activity and name). The 'name' variable is an interesting one: it has 16 categories with only one entry in each! Many students would agree that name has nothing to do with the other variables and yet students often want to hang on to this variable in their representations. 'Name', in fact, gives rise to a fifth variable, 'gender', which may well exhibit a relationship with the other variables. The issue of producing new variables from old is an important one in data analysis although we will not examine this explicitly here; nor will we consider whether or not there are any relationships in the data involving gender.

If you take a close look at the data set, you may notice some trends. One is that the people who eat lots of fast food do not seem to do much exercise; another is that the people who have more active favourite activities do more exercise during the week. In such a small data set these relationships can be seen simply by scanning the data as presented in the table. If the data set is larger, however, such scanning may not be possible. Even if you do observe some trends, how do you really convince yourself and then someone else that these trends are there? These questions highlight one of the important aspects of statistics: data analysis is about finding messages in data and then conveying those messages to others in an effective way.

The power of sorting

To highlight these issues, consider the first relationship mentioned before: that the people who have high fast food consumption do not exercise very much. When this trend was pointed out to the Year 7 students and they were asked to draw a graph or something similar to show the trend, many of the students produced graphs similar to that shown in Figure 2. By scanning the whole graph, in a similar way to scanning the whole

Name	Number of hours exercise per week	Favourite activity	Fast foods meals eaten each week
Georgia	5	sport	0
Mary	0	watching TV	4
Kate	1	musical instrument	3
Cathy	1	watching TV	2
Jack	1	computer games	0
Brian	0	computer games	2
David	3	sport	1
Paul	2	musical instrument	2
Alice	2	sport	1
Nathan	1	watching TV	3
Olivia	3	musical instrument	0
Frank	1	computer games	4
Liam	4	sport	1
Erin	3	musical instrument	1
Harry	2	watching TV	3
Isabel	3	sport	1

Figure 1. The data set.

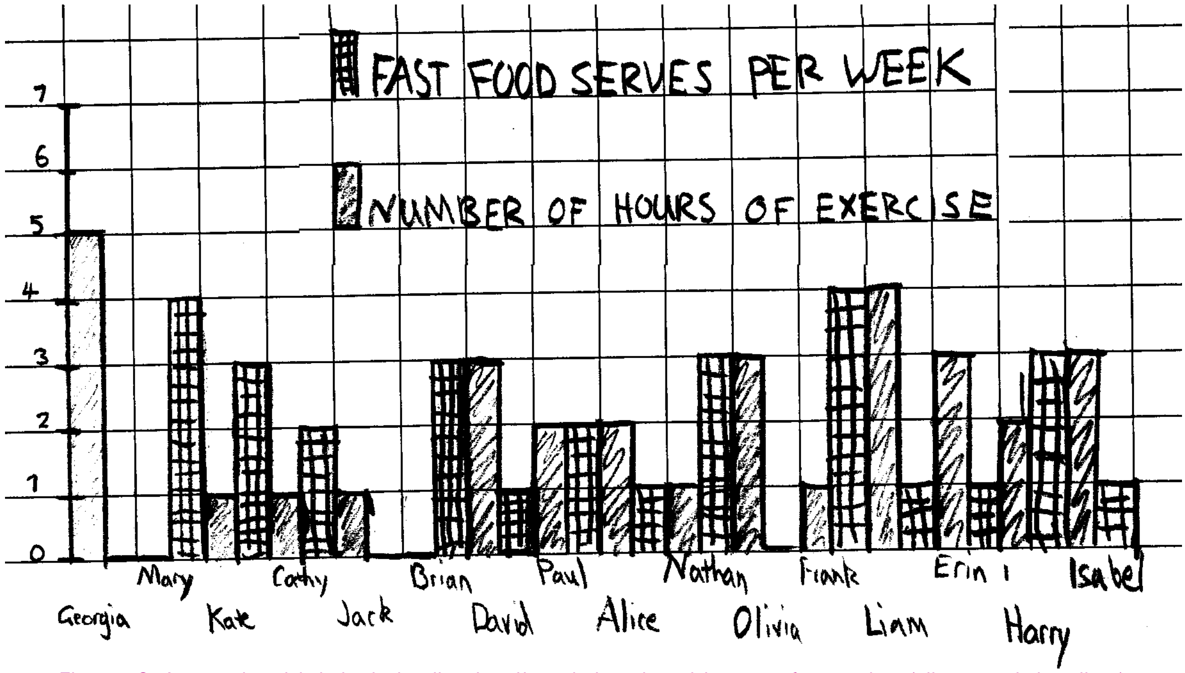

Figure 2. A graph which just duplicates the data about hours of exercise (diagonal shading) and fast food (square shading), in the same order as in the table.

table, it is possible to see the trend, but it is hard work. Many of the students recognised this when discussing their graphs. You have to compare each person's hours of exercise with his or her fast food consumption, and keep track of whether one is high when the other is low, and whether or not this continues across the whole graph. You also have to keep track of how common the exceptions are, because although we might accept a few contradictory values, we certainly do not want too many.

Figure 3 shows what a difference can be achieved if the data are sorted first, using one of the variables. In this case the sorted data are in order of increasing number of hours of exercise, but other than this, the approach to graphing the data is exactly the same as in Figure 2. It is now much easier to see the trend: the dark bars showing hours of exercise increase from left to right, while at the same time the lighter lines, showing fast food consumption, tend to decrease. Among the 70 or so Year 7 students who were asked to graph this data set no-one actually produced a representation like Figure 3. We as teachers may take the idea of 'sorting' for granted because of its simplicity, and yet it is a powerful technique for discovering and displaying trends in data, and we should take opportunities to identify this approach more explicitly for students.

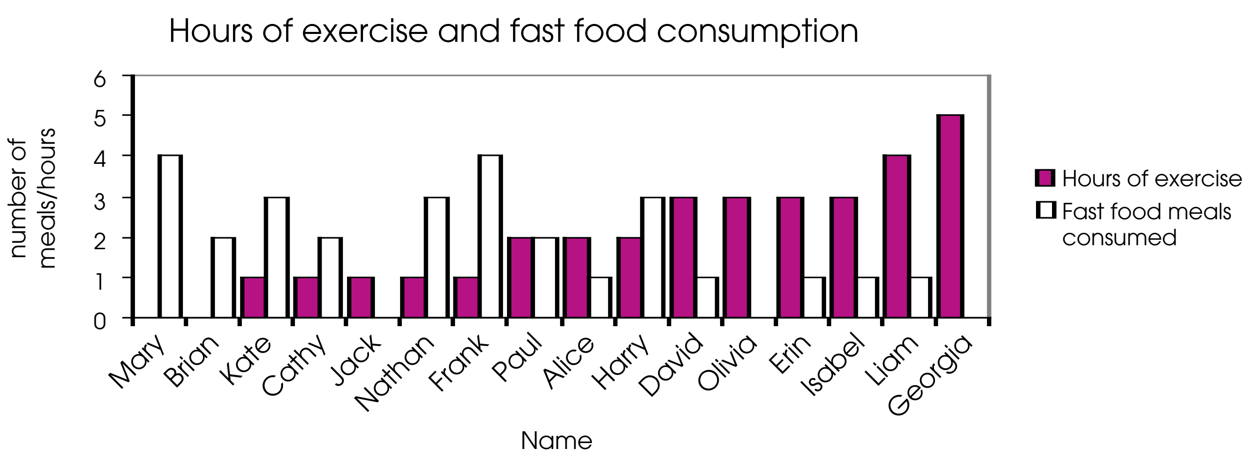

Figure 3. Graphing all of the hours of exercise and fast food consumption data, but after sorting on number of hours of exercise per week.

The power of the scatter graph

Another technique that works well for showing relationships between numerical variables is, of course, the scatter graph. Many curriculum statements leave this strategy until late in schooling, yet it is a simple approach that is very effective and easily understood by students. It has the advantage of requiring no rearranging of the data in advance of plotting points on the graph, because it is the structure of the graph itself that allows trends to be revealed. Figure 4 shows a scatter graph produced by one of the Year 7 students. The relationship between exercise and fast food consumption is clearly shown by the way the values start in the top left corner and tend down to the bottom right.

The Year 7 student who produced the graph in Figure 4 colour-coded all the points so that each point could be associated with the name of the corresponding person in the data set. This retention of the identities of the data occurred in various kinds of representation by other students as well. It seems that students like to retain all details of the data for as long as possible, perhaps reluctant to compress or omit data despite the fact that doing so might make the message in the data clearer.

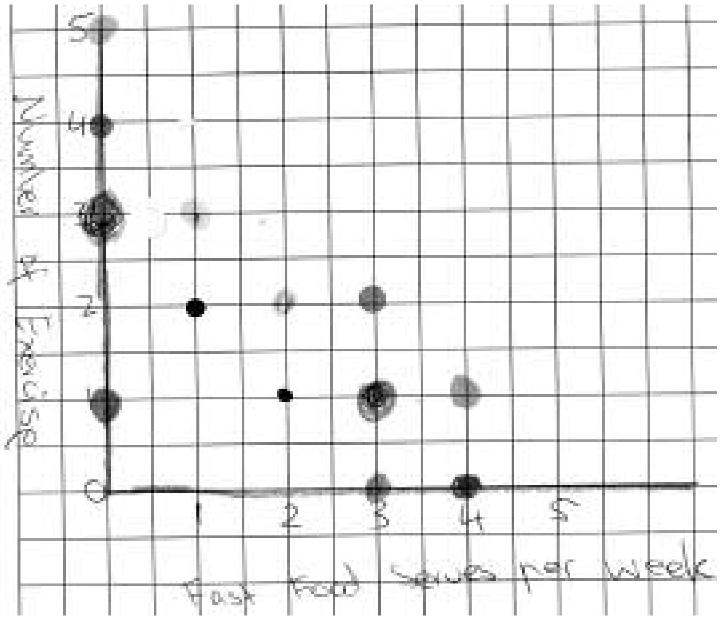

Figure 4. A scatter graph illustrating the association between hours of exercise and fast food.

The power of grouping

The second relationship evident in the data set is that people who have more active favourite activities do more exercise during the week. Here one of the challenges with data exploration and representation is that one variable is numerical (hours of exercise) and the other is categorical (favourite activity), which makes it difficult to use a scatter graph. Many of the students produced representations similar to those in Figure 2: unordered, and with the added complication that it is hard to show a categorical variable in a bar graph. In these representations it was very difficult to see the claimed relationship.

In contrast, some students realised that the categories in the 'favourite activity' variable allowed them to group the data, and then the hours of exercise could be shown for the people in each group. The effectiveness of such a grouping strategy is evident in Figure 5. Here a 'by eye' visual comparison across the groups makes it evident that the sports players and the musicians exercise a lot, whereas the computer games users and television watchers are rather lethargic.

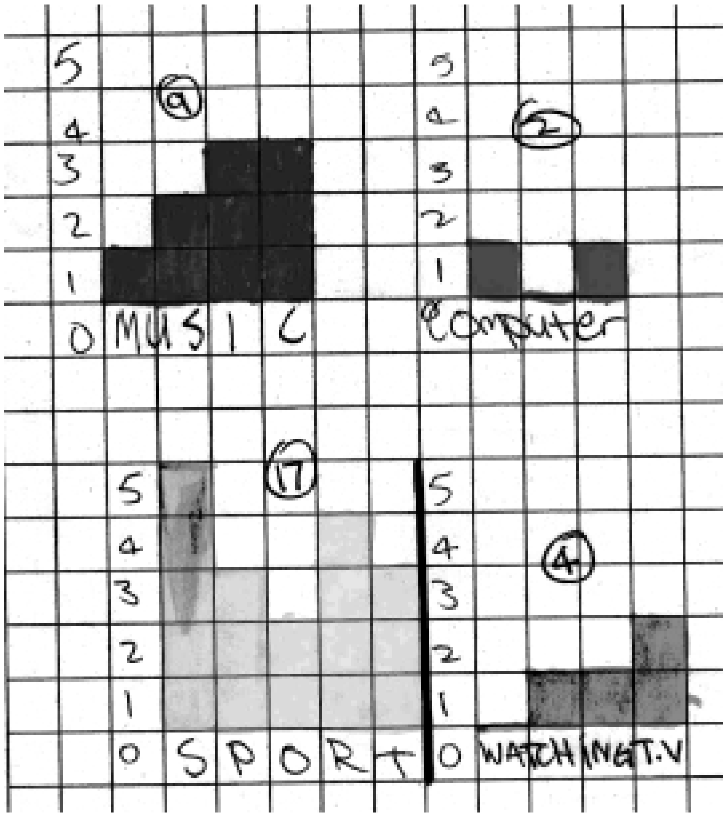

Figure 5. Data grouped by favourite activity, with sets of bar graphs showing the number of hours of exercise for each student in the activity category.

The power of the mean

What Figure 5 does not clearly take into account, of course, is that there are different numbers of people in each of the groups. Few students seem to appreciate the power of the mean for dealing with different group sizes and enabling comparisons across groups (cf. Watson & Moritz, 1999). Only four of the Year 7 students calculated means for the four groups, as shown in Figure 6. One student went one step further and sorted the mean values into decreasing order in a table to highlight further the relationship between hours of exercise and favourite activity. The use of the mean allows us to quantify the differences visible in graphs like in Figure 5. This level of response requires both grouping of data and then compression of data through computation. It may be that students' reluctance to 'lose' data by calculating the mean inhibits their use of it.

None of the Year 7 students considered doing box-and-whisker plots, but this is not surprising considering the small sizes of the favourite activity groups. We will not discuss box-and-whisker plots further here, except to highlight that they are another powerful yet simple way of making comparisons among groups, with the added advantage of showing not only a measure of central tendency (the median) but also an indication of the range of data in the set.

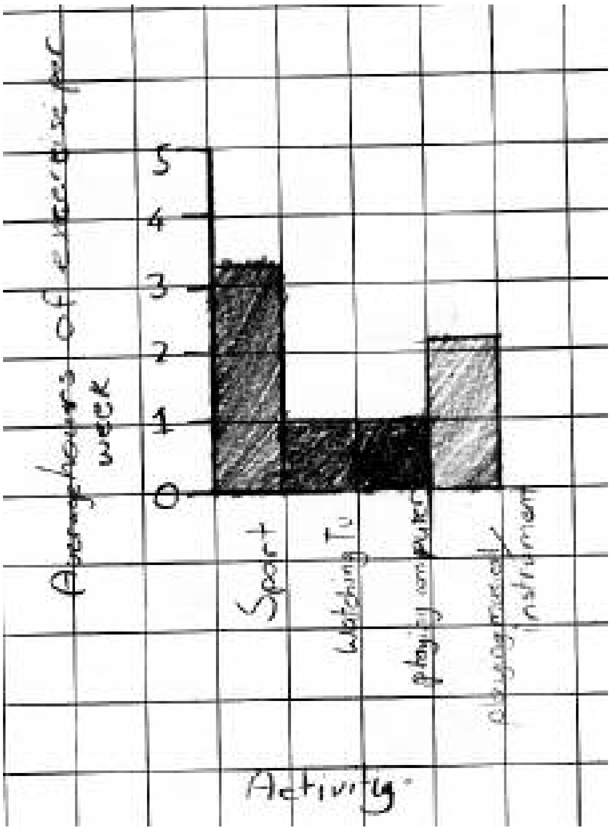

Figure 6. A student's table of average hours of exercise for each of the favourite activities.

Conclusions

None of the techniques—sorting, scatter graphs, grouping, or calculating means—that have been highlighted here are particularly sophisticated, and yet their simplicity is often more than adequate for displaying the trends in data or for making comparisons in a convincing way. These are strategies that are easy to introduce to students and that allow students to grapple with the complexities of multivariate data. In particular, we should highlight the mean as a statistic that allows us to make comparisons across groups.

The discussion here also highlights an even more important issue. We need to help students to understand that the purpose and power of statistics is for answering questions using data and that answering questions also means convincing others of the validity of the answers found. The techniques described here, and the more sophisticated ones learned later in students' statistical education, allow us to find answers in data, and provide evidence for others of the trends that we observe. If students do not appreciate this purpose, then there is no motivation to carry out data exploration or to go through the data representation process with the intention of conveying a convincing message.

References

Watson, J. M., Collis, K. F., Callingham, R. A. & Moritz, J. B. (1995). A model for assessing higher order thinking in statistics. *Educational Research and Evaluation, 1*, 247–275.

Watson, J. M. & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*, 145–168.

Helen Chick
University of Melbourne
h.chick@unimelb.edu.au