Statistical literacy in action

Should all graphs start at zero?

Jane Watson
University of Tasmania
<jane.watson@utas.edu.au>

Using statistical literacy skills to determine appropriate scales to be used on graphs is an essential part of numeracy. Using several meaningful contexts, this article explains very clearly when it is appropriate and inappropriate to begin the scale of a graph at zero.

As statistics has become a component of the mathematics curriculum (e.g., Australian Education Council, 1994; Australian Curriculum, Reporting and Assessment Authority [ACARA], 2015), there has been growing recognition of the importance of statistical literacy. Beyond calculating means and drawing bar charts,

> "statistical literacy is the meeting point of the data and chance curriculum and the everyday world, where encounters involve unrehearsed contexts and spontaneous decision-making based on the ability to apply statistical tools, general contextual knowledge, and critical literacy skill" (Watson, 2006, p. 11).

This article arose from three Grade 6 experiences either explicitly or implicitly suggesting that all graphs should start at zero: a text book, a state curriculum document, and a teacher's comment.

Graphs are a visual presentation of the everyday world and they are often met in the media with little explanation but with the expectation that the reader will understand the point being made. Understanding the context and the construction of the graph and being able to read the information presented critically allows one to make decisions about the message embedded in the graph. This aspect of statistical literacy is recognised within the *General Capabilities* in the *Australian Curriculum* (ACARA, 2013) where under the Numeracy Capability, "Interpreting statistical information" is one of the elements.

> This element involves students gaining familiarity with the way statistical information is represented through solving problems in authentic contexts that involve collecting, recording, displaying, comparing and evaluating the effectiveness of data displays of various types (p. 37).

Examples of misleading graphs have long been provided in books on misleading statistics (e.g., Huff, 1954; Jaffe & Spirer, 1987). These often show graphs that mislead because they exaggerate small differences by presenting data on an axis whose scale does not begin at zero. Some authors, for example Wainer (2005), however, go on to point out how producing graphs that do begin at zero, when it does not fit the story being told, can also be misleading and inappropriate. This article calls for a balance in presenting views on graphing, focussing discussion on the importance the context of the data being displayed and the inference being drawn.

In terms of the issue of graphs "starting at zero," it is certainly true that the *Australian Curriculum: Mathematics* (ACARA, 2015) has children creating graphs in the early years, representing counts of categorical variables, which start at zero as shown in Figure 1. As such, graphs represent frequencies for variables. They can be created with blocks as students report data and a variable with no blocks has a value of zero. In cases like this it is important to begin the vertical scale at zero and label appropriately for the frequency. Students hence may be told, “Begin your scale at zero.” Sometimes however, this becomes a universal instruction for years to come: “Every graph you draw begins at zero.” Similar to edicts like “multiplication makes bigger,” however, eventually situations occur where it is not appropriate.

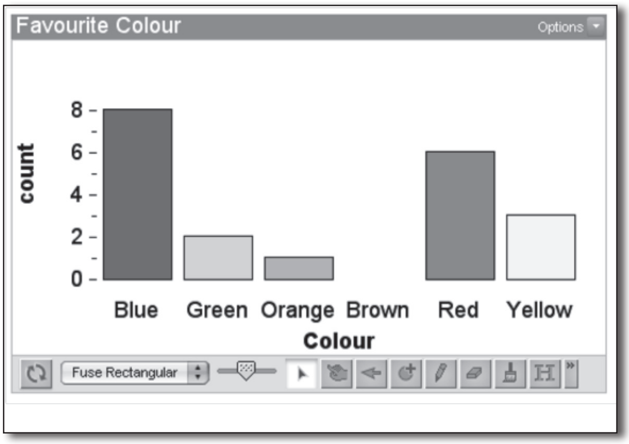

Figure 1. Bar graph with vertical scale starting at zero.

As data sets become more complex, the context to which the data relate and the variation present in the data may demand that to tell the story of the data clearly and realistically, the graph scales do not begin at zero. Instead they are truncated to display the variation realistically. To appreciate why this is so, students need to explore data sets and create graphs to tell stories, not just to plot points on a graph.

Consider two examples. First, students may be measuring the heights of some middle school students to consider the distribution of heights and the typical height of the group. They might be asking about their class or predicting the typical height of middle school students in Australia. They are likely to plot their class data as part of the investigation. As examples, in Figure 2 the data are plotted in two ways for the heights of 30 middle school students collected from the Australian Bureau of Statistics (ABS) *CensusAtSchool* website (ABS, n.d.). In the top plot, the horizontal axis starts at 0; in the bottom one it starts at 140 cm. From the context of the question, would anyone expect a value of height for the students to be near zero? What part of the story does the blank area of the graph convey?

More importantly, it is virtually impossible to judge the variation among the heights of the 30 students in the top plot. It is likely that the objective of the lesson is related to the variation, centre and range seen in the heights and the expectation of the typical height for the students. In the bottom plot, it is possible to talk about the clusters, gaps, and potential outliers in the data, whether the data are evenly distributed, and what the typical value or values might be. The gap, which may or may not be important, is not visible in the top plot.

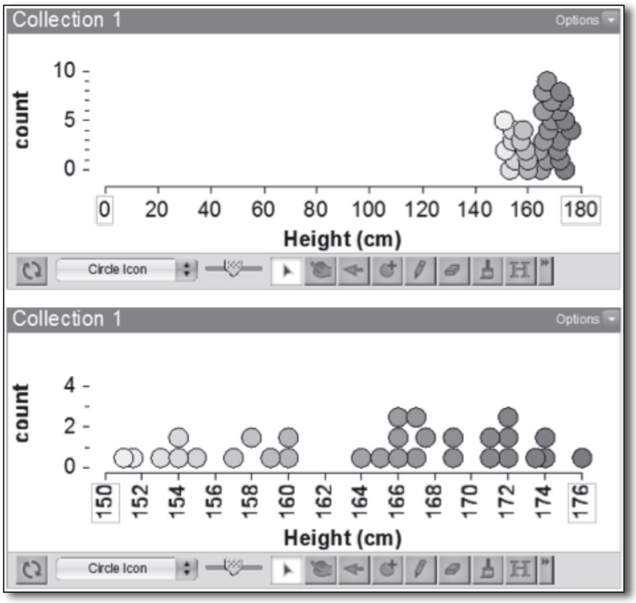

Figure 2. The heights of 30 middle school students plotted on two different scales.

The second example is based on the winning times of the Melbourne Cup over the years of the race. The data suggest presentation in a time series scatterplot because the years are discrete, which students are likely to meet by Year 6 in some Australian states. The Year of the race is placed on the horizontal axis and the winning Time on the vertical axis. In this context, what would a value for the variable of Year of 0 mean? The context for representation does demand knowing when the Melbourne Cup began. Although commentators may talk about “the 150th Melbourne Cup,” it would not be meaningful to relabel the graph starting with 1860 as “0,” so that 1861 is “1” and 2015 is “155.” Or what would the graph look like with a horizontal scale of (0, 2020)? Part of knowing the history of the Melbourne Cup is knowing that the first race was in 1861. Accepting the common use of time series plots, what about plotting the winning times of the races?

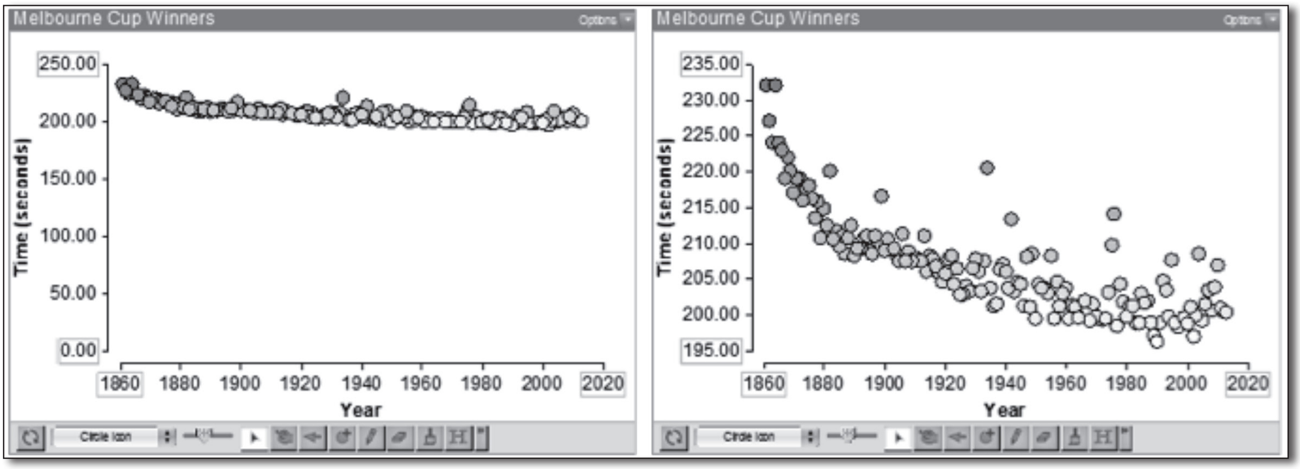

Figure 3. The winning times for the Melbourne Cup plotted on different scales.

Again two plots are shown in Figure 3, the one on the left with the vertical axis starting at 0 seconds and the one on the right starting at 195 seconds. As with the heights of middle school students, the point of creating a graph is surely to illustrate how the times have changed relative to one another and the general trend to see faster times over the years. The variation in elapsed time relative to the total time to run the race is very little but the interest is in the change between 1861 and now. In the context of the race, no one would expect the times to be near zero and the interesting story of the times is lost in the plot on the left.

In the early days of students using technology, Ben-Zvi (2000) studied two boys' use of software to plot the data for the winning times of the 100 m race at the Olympic Games. The boys created several plots with different scales similar to those in Figure 3, as well as others, and wrote different conclusions from what they saw, ranging from "little change in times" to "considerable improvement". Their discussion illustrates the power of graphs to tell a story and the influence that starting the scale at zero can have. Important outcomes of the primary and middle school years include the intuitions about what plots are appropriate for the context of and variation in the data being studied. Although the boys considered many situations where the vertical axis changed, and choose different ranges for Year, they never proposed starting the horizontal axis at zero.

The choice of scales as shown in Figure 3 can become controversial when one is tempted to use the type of scale on the right to mislead the viewer in to believing the variation is greater than is realistic in the context of the data. What becomes important for statistical literacy is that the context where data are presented sometimes raises the possibility of purposely choosing the scale of a graph to convey a misleading message.

A variety of examples exists in the media. The graph in Figure 4 is from the *Numeracy in the News* website and originally appeared in the *Hobart Mercury* (Basic Parliamentary Salaries, 1993). The graph was meant to illustrate how badly paid Tasmanian politicians were in comparison to politicians in other States and Federally. Examples such as this sometimes appear in curriculum documents or textbooks to show how graphs can be made to be misleading. What is important is the context of the graph and the comparison that is being made within it. In this case the relative difference in politicians' incomes is important in relation to their total incomes, not just the difference between the lowest total income and the highest total income of \$68 000 – \$47 000 = \$21 000. Seen as a percentage of the range shown on the graph of \$40 000 to \$68 000, it appears that Tasmanian politicians only earned about a quarter of what the highest paid earned (working it out in percentage terms the Tasmanian bar is only 25% of the highest bar in the graph: $\frac{7000}{28000}=0.25$).

If, instead of the graph in Figure 4, a graph with the vertical axis starting at 0 were created, as in Figure 5, it would be seen that the Tasmanian income was in fact about two-thirds of federal parliamentarians' income ($\frac{47000}{68000}=0.69$). In this situation, and in similar contexts, it is certainly true that the scale on the graph should start at zero. The lengths of the bars accurately represent the relationship of the salaries. In the media and

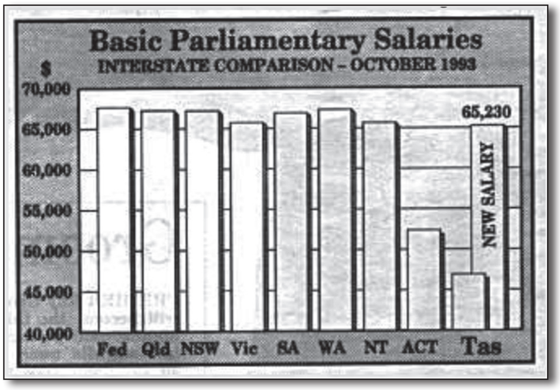

Figure 4. Comparing parliamentary salaries in 1993.

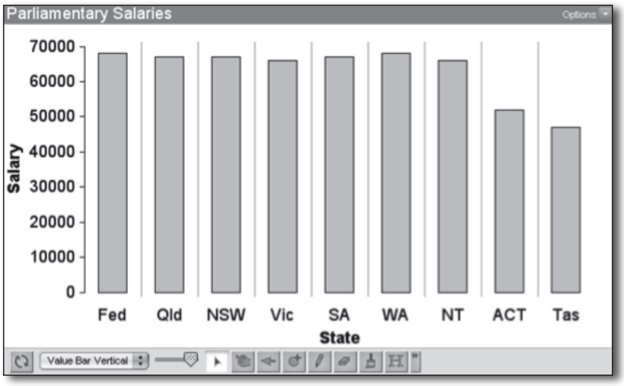

Figure 5. Parliamentary salaries replotted with a vertical scale beginning at 0.

in politics making a meaningful comparison may not be the aim of the writer. Unfortunately examples such as these can be used to reinforce the simplistic rule, "Every graph should begin at zero." Comparing the contexts of the Melbourne Cup winning times (Figure 3, right) and the incomes of parliamentarians (Figure 4), race commentators are not heard to talk of the race time for 2014 being 85.2% of the winning time in 1861. What interest does this have for race-goers? Of more interest is whether it was 2.6 seconds faster than last year (and why) or 1.4 seconds slower than the fastest time ever. Context determines how the data represented are interpreted.

Part of thinking statistically when doing a statistical investigation or being statistically literate when encountering graphs from the media or elsewhere, is knowing the context from which the data came and knowing the story that is being told about the variation in the data. There is no rule that fits every situation when drawing a graph and labelling the axes. Hence students require the General Capability element of interpreting statistical information (ACARA, 2013), as reported at the beginning of this article.

Although the *Australian Curriculum: Mathematics* (ACARA, 2015) does not specifically mention the issue of zero as a starting point in creating graphs, the description of the Reasoning proficiency in Year 4 includes "communicating information using graphical displays and evaluating the appropriateness of different displays" (p. 36). As well in Year 6 one of the elaborations for the descriptor "Interpret secondary data presented in digital media and elsewhere" (ACSMP148) says in part "identifying potentially misleading data representations in the media, such as graphs with broken axes or non-linear scales, graphics not drawn to scale" (p. 53). These statements point in the right direction but everyone (teacher, student, teacher educator) needs to be aware of the finer points of graph construction and context, not just using potentially misleading graphs as a reason to make up a rule for every context.

The message in this article can be related to the development of statistical literacy as a three-tiered hierarchy applied to graphs (Watson, 2006). For Tier 1, students require a basic understanding of graph types and their properties. For Tier 2, students need to be able to make meaning of a particular type of graph when it is presented within an authentic context. For Tier 3, it is necessary to apply critical thinking skills to judge whether the claim represented in the graph is accurate or misleading in the context. In some cases this requires students to have the confidence to challenge statements that cannot be justified. Tier 3 reasoning is an example of another General Capability (ACARA, 2013), that of "Critical and creative thinking" (p. 66).

Becoming statistically literate in unrehearsed contexts involving spontaneous decision-making about graphing, as well as other topics in the statistics curriculum, does not develop during a particular year of the curriculum. It develops over many years and requires many experiences giving opportunities and contexts in which to make judgments.

Acknowledgements

The author thanks Bruce Duncan for helpful discussions on this topic and reviewers for further suggestions. Figures 1, 2, 3, and 5 were created

with the software *TinkerPlots* (Konold & Miller, 2011).

References

Australian Bureau of Statistics. (n.d.). *CensusAtSchool Australia.* Canberra: Author. Retrieved from http://www.cas.abs.gov.au/cgi-local/cassampler.pl

Australian Curriculum, Assessment and Reporting Authority. (2013). *General capabilities in the Australian Curriculum*. Sydney.

Australian Curriculum, Assessment and Reporting Authority. (2015). *Australian Curriculum: Mathematics, Version 7.4*. Sydney.

Australian Education Council. (1991). *A national statement on mathematics for Australian schools*. Melbourne.

Basic Parliamentary Salaries. (1993, November 11). *The Mercury* (Hobart, Tasmania), p. 2. Retrieved from www.mercurynie.com.au/mathguys/mercury.htm

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1 & 2), 127–155.

Huff, D. (1954). *How to lie with statistics.* London: Penguin Books.

Jaffe, A. J. & Spirer, H. F. (1987). *Misused statistics: Straight talk for twisted numbers*. New York: Marcel Dekker.

Konold, C. & Miller, C. D. (2011). *TinkerPlots: Dynamic data exploration* [computer software, Version 2.0]. Emeryville, CA: Key Curriculum Press.

Wainer, H. (2005). *Graphic discovery: A trout in the milk and other visual adventures*. Princeton, NJ: University Press.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.