Seeing is Believing?

Kate Richards, Neville Davies, Gemma Parkinson, Dominic Martignetti
Royal Statistical Society Centre for Statistical Education (RSSCSE)
Plymouth University
Corresponding author: Kate Richards, e-mail katharine.richards@plymouth.ac.uk

'Internal Migration in England & Wales, year ending June 2011' (Office for National Statistics, ONS), 'Your Olympic Athlete Body Match' (British Broadcasting Corporation, BBC) and 'Gapminder World' provide some examples of data visualisation (DV) that not only present information and tell stories but also motivate people to want to find out more. However, can we 'believe' every picture and representation we see and get trustworthy information from them? There are many examples of DV that distort and misrepresent the data they are presenting. These can be unintentional but sometimes may be deliberate to make a point not evidenced by the data. In this paper we look at examples of poor and misleading graphs and charts. Some extremely bad examples are taken from the world of business and finance while others are generated from data captured from learners who participate in the International *CensusAtSchool* Project. We will show how examples of DV can be used to create engaging learning resources for school-aged learners to demonstrate and encourage good practice when presenting data.

Keywords: Data visualisation, learning and teaching resources.

1. Introduction

In newspapers, magazines, journals or information leaflets there are many examples of DV. This is also true for online articles such as those published by, for example, the BBC, The Guardian, The Economist magazine and Heat (a gossip magazine). Building societies, banks, companies and businesses often use traditional charts and graphs to display information to attract or inform new clients.

Much has been written regarding the proper construction of graphs and charts. Tufte (1983) states: 'Excellence in statistical graphics consists of complex ideas communicated with clarity, precision and efficiency'. Key guidelines for proper presentation of graphical information in order to promote unbiased decision making have been established (Bertin, 1983). Bertin's system comprises seven visual variables: position, form, orientation, colour, texture, value and size, combined with selective, associative, ordered and quantitative visual variables. Arunachalam et al. (2002) point out that graphs are often designed to support decision making and conclude, amongst other things, that improperly designed graphs can radically alter viewers' choices. In many cases the visual inference the constructor intended people to get from the DV is distorted owing to poor design.

2. Misleading DV

Amer (2005) demonstrated that visual illusion in graphs may bias decision makers even when they are prepared according to generally accepted preparation guidelines. Figure 1 shows the sales figures of a product for two-monthly periods in the year 2001. The main problem with this DV is that it is three-dimensional. Without horizontal grid lines the sales for Nov-Dec look equal to those for May-Jun. However, in figure 2 it is clear that this is not the case. This optical illusion is called the "Poggendorff illusion" (Finemann, 1996) and causes underestimation or overestimation of the values displayed in graphs. Figures 3 and 4 demonstrate, using height and age variables for children who took part in *CensusAtSchool* in 2011/12, that a three-dimensional graph can mislead by, in this case, not showing points of intersection.

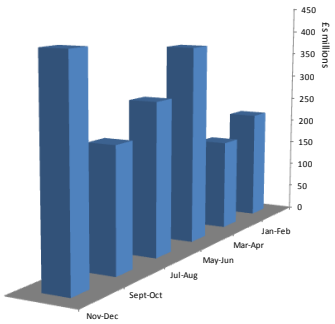

Figure 1 Sales of a product during 2001 without gridlines

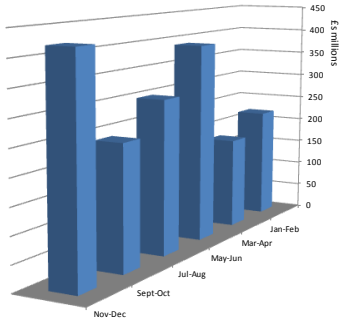

Figure 2 Sales of a product during 2001 with gridlines

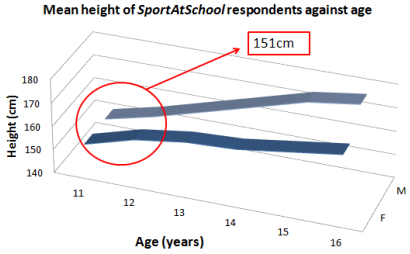

Figure 3 Mean height of children against age *CensusAtSchool* 2011/12

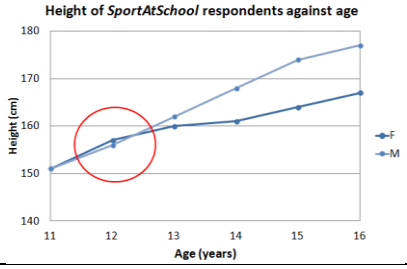

Figure 4 Mean height of children against age *CensusAtSchool* 2011/12

Another optical illusion that can mislead is one of contrast (Robinson, 1998; Block and Yuker, 1989). The judgement about the length of the boxes in figure 5 are affected by the appearance of the lengths of the whiskers. These are longer in boxplot b) than c). These result in the centre box section appearing shorter in b) than in c) when, in fact, they are the same size; see figure 6.

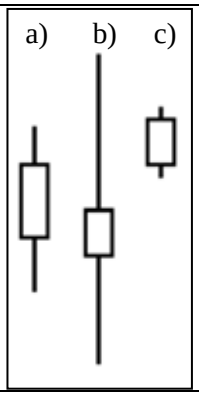

Figure 5 Boxplots with no gridlines

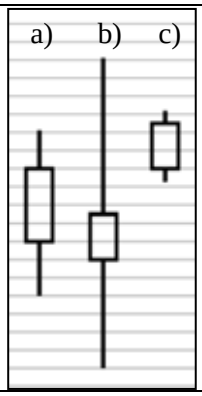

Figure 6 Boxplots with gridlines

The representation of data is often distorted by suppressing the origin (Freeman et al., 2008). An article on the Rangers Football Club website in January 2013 (www.rangers.co.uk/news/headlines/item/3147-seasons-to-be-cheerful) was written to show readers that Rangers' average annual attendance had increased in 2012-2013. Figure 7 shows a Club employee holding a bar chart with each bar representing the average annual attendance for three consecutive seasons. The vertical axis for this bar chart starts at approximately 37,500. This, along with the angle the chart is held at, exaggerates the relative differences in the heights of the bars. When the origin is not suppressed, see figure 8, it is very difficult to see any differences in the heights of the bars. Kozac (2011) addresses this issue.

Figure 7 Rangers football club attendance with suppressed origin

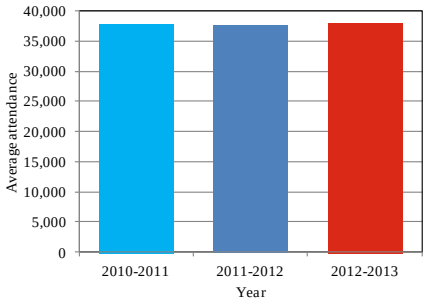

Figure 8 Rangers football club attendance, vertical axis starting at zero

There is also confusion between when to plot a bar chart rather than a histogram. Even the BBC have made this mistake in their Bitesize video on bar charts where they use a histogram to display the discrete variable, the number of goals scored each day (www.bbc.co.uk/learningzone/clips/an-introduction-to-bar-charts/2941.html).

Often the aspect ratio, the ratio of height to width for a plot, is not considered when graphs and charts are inserted into text. Invariably they are an accident of the space available. The extract in figure 9 is from an article in the Economist (www.economist.com/blogs/buttonwood/2012/10/markets-and-economy) 'Industrial Unrest' and is an example of a graph being stretched vertically to fit the space. This distorts the message by exaggerating the downward gradient.

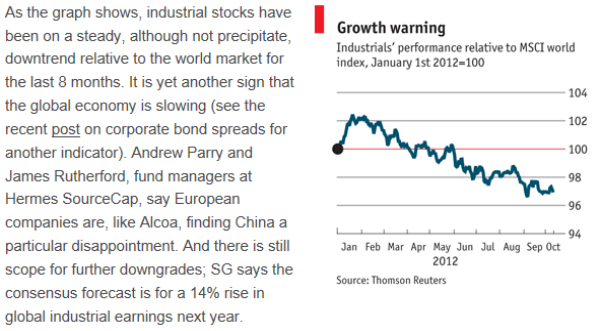
As the graph shows, industrial stocks have been on a steady, although not precipitate, downtrend relative to the world market for the last 8 months. It is yet another sign that the global economy is slowing (see the recent post on corporate bond spreads for another indicator). Andrew Parry and James Rutherford, fund managers at Hermes SourceCap, say European companies are, like Alcoa, finding China a particular disappointment. And there is still scope for further downgrades; SG says the consensus forecast is for a 14% rise in global industrial earnings next year.

Figure 9 Extract from Industrial Unrest The Economist, 11 October 2011

We now present three examples of very bad pie charts. In figures 10 and 11 a UK building society shows their customers the assets their cash and growth funds will be invested in. Figure 10 is nonsensical as the whole pie is one colour, and it has a key with seven types of allocations in spite of the fact that it is totally a cash investment. Figure 11 is a three-dimensional pie chart that exaggerates the sector for overseas equities and again has a key with allocations not included in the chart. Also due to the colours in the key, it is difficult to distinguish between Gilts and Overseas Equities.

The choice of colour when designing charts and graphs is important as is demonstrated by the third pie chart example in figure 12. The Lloyds TSB bank offer customers 'Money Manager' (www.lloydstsb.com/internet_banking/money_manager.asp). This displays their spending as a doughnut pie chart. It is difficult to compare the size of the sectors in this type of chart and a person who is red/green colour blind (deuteranopia), see figure 13, or who is completely colour blind (achromatopsia), see figure 14, will not be able to distinguish between some sectors.

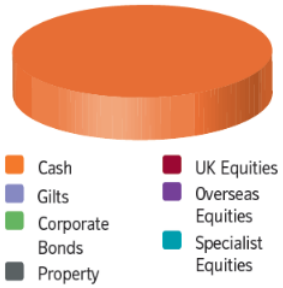

Figure 10 Cash fund asset allocation for a UK building society

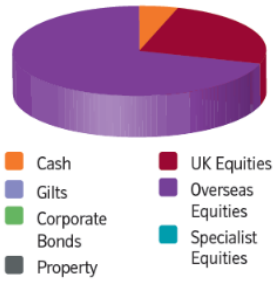

Figure 11 Growth fund asset allocation for UK building society

Figure 12 Lloyds TSB Money Manager full colour

Figure 13 Lloyds TSB Money Manager red/green colour blind filter

Figure 14 Lloyds TSB Money Manager full colour blind filter

Huff (1991) uses the image in figure 15 to demonstrate how increases in frequencies can be deliberately exaggerated. The moneybag on the left represents the average weekly salary of carpenters in a fictional country. The moneybag on the right represents the average weekly salary of carpenters in America. The amount of salary in the larger bag is twice that of the smaller, yet as the height and width of the larger bag have doubled the bag is four times the area of the smaller bag.

Examples like these can be used to challenge learners to identify why the representations are misleading, to display the data correctly and to make the correct conclusions. Two examples of this type of resource, 'Bad Bar Charts' and 'Bad Pie Charts', can be found on the *WinAtSchool* website (www.winatschool.org.uk/?page_id=282). Further examples, videos, PowerPoint presentations and lesson plans can be found on the Internet for Classrooms website (www.internet4classrooms.com/grade_level_help/misleading_graphs_math_sixth_6th_grade.htm)

Figure 15 Huff's money bags for a fictional country (left) and America (right)

Examples of poor DV are becoming increasingly common. Many examples can be seen on the websites such as Junk Charts (http://junkcharts.typepad.com), The Top Ten Worst Graphs (www.biostat.wisc.edu/~kbroman/topten_worstgraphs) and Eval Blog (http://evalblog.com/2012/01/23/tragic-graphic-the-wall-street-journal-lies-with-statistics).

3. Interactive DV

There are some excellent examples of interactive DV that can be used in the classroom to engage learners. However, in order to develop the initial visual interest into class lessons or investigative projects many DV tools need specific class-level activities/resources, lesson plans and ideas to be developed.

The Gapminder quiz uses the animated Gapminder World graph (www.gapminder.org/GapminderMedia/wp-uploads/teachers/Gapminder%20quiz.pdf) to enable users to answer eight multiple choice questions. Each question has a link to an animated graph of the variable being considered against time similar to figures 16 and 17. There is also some background information under the heading 'If student asks'. This activity is a very useful starter for a main lesson of global investigations.

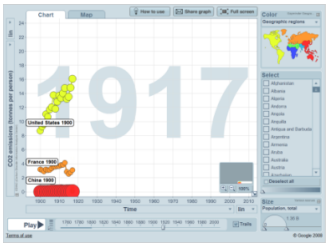

Figure 16 Gapminder World CO_2 emissions against time 1930 -1917

Figure 17 Gapminder World CO_2 emissions against time 1930 -2011

Users can compare their own height and weight with those of Olympians using the BBC Athlete Body Match (www.bbc.co.uk/news/uk-19050139). Heights and weights can be entered and are ploted as coordinates on a scatter graph of height against weight for hundreds of Olympians (figures 18 and 19). This is an appealing way to introduce scatter graphs to learners as long as they are not sensitive about their height or weight.

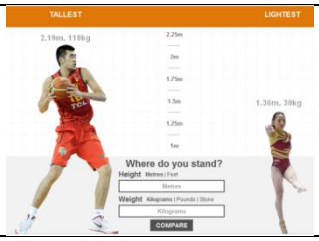

Figure 18 BBC Athlete Body Match

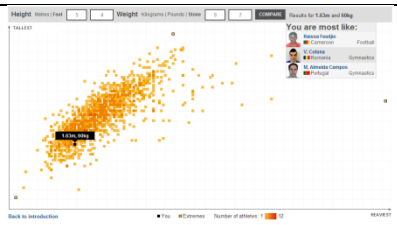

Figure 19 Athlete Body height against weight for Olympians

Multi-dimensional data on topics such as pensions and savings, sexually transmitted infections and poverty using the web-ready visualisation tools, together with a number of learning and teaching resources are available on the SMART Centre website (www.dur.ac.uk/smart.centre1/cceamaster).

The RSSCSE has created an online quiz, a survey, card activities, worksheets, teachers' notes and investigations based on the interactive DVs and output from the UK national Census 2011 (www.censusatschool.org.uk). An example of a question in the online quiz is shown in figure 20. This invites learners to select from four options, the area that has the highest percentage of over-90 year olds. Once the selection is made the solution is displayed with a link to the interactive population pyramids, figure 21, on the ONS website where the answer can be confirmed (www.ons.gov.uk/ons/interactive/vp2-2011-census-comparator/index.html).

Another question queries which area has the highest percentage of White British people, with the four choices of London, Birmingham, Bristol or Leicester. Learners are directed to the ONS Ethnicity Data Map (www.ons.gov.uk/ons/interactive/census-map-2-1---ethnicity/index.html) to find the answer. Data maps can carry a huge amount of data in a small place (Tufte, 1983). However, they show geographical regions rather than population size which can be misleading as a larger region could suggest a larger population. The ONS have addressed this by selecting areas of approximately the same population size.

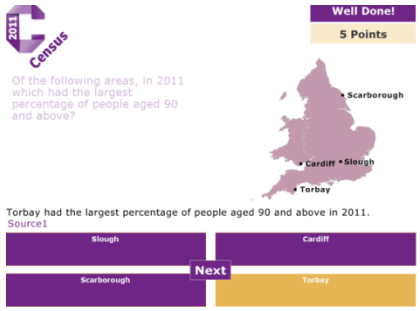

Figure 20 Question from RSSCSE online quiz

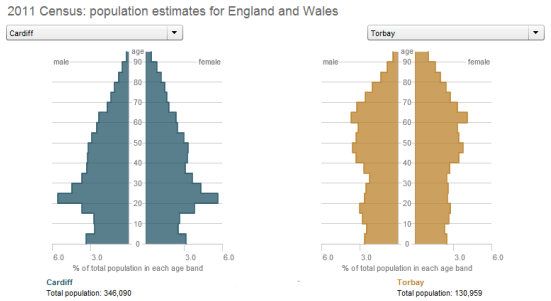

Figure 21 Population pyramids for Cardiff and Torbay ONS Census 2011output

4. Conclusion

Learners need to be taught to present data well, to be able to identify poor or misleading charts and graphs and to interpret well designed DV to ensure they get trustworthy information. One way to do this is to create engaging resources using good and bad dynamic and static graphs and charts. Before DV can be used in the classroom for effective learning, exciting and fun activities need to be designed that lead to real and meaningful investigations so that learners can feel they have discovered something new and relevant.

5. References

Amer, T. S. (2005) "Bias Due to Visual Illusion in the Graphical Presentation of Accounting Information," *Journal of Information Systems*,19 (Spring), 1 – 18.

Arunachalam, V., Pie, B. K. W. and Steinbart, P. J. (2002) "Impression management with graphs: Effects on choices," *Journal of Information Systems*, 16 (Fall), 183–202.

Amer, T. S. (2008) "*The Effect of Visual Illusions on the Graphical Display of Information: Evidence of Bias and Mitigation*" Working Paper Series 08-05.

Bertin, J. (1983) *The Semiology of Graphics*, University of Wisconsin Press, Madison, WI.

Block, J. R. and Yuker, H.E. (1989) *Can You Believe Your Eyes?* Gardner Press, New York.

Fineman, M. (1996) *The Nature of Visual Illusion.* New York: Dover

Freeman, J. V., Walters, S. J. and Campbell, M. J. (2008) *How to Display Data*. BMJ Books, Blackwell Publishing.

Huff, D. (1991) *How to lie with statistics*, Penguin Books, London.

Kozak, M. (2011) "When Should Zero be included on a Scale Showing Magnitude?" *Teaching Statistics Journal*, 33, Number 2, 53-58.

Robinson, J. O. (1998) *The Psychology of Visual Illusion*, Dover Publications, Inc. Mineola, NY.

Tufte, E. R. (1983) *The Visual Display of Quantitative Information*, Graphics Press, Cheshire.