Statistical Graphics Recommendations for the ASA/NCTM Annual Poster Competition and Project Competition

Richard M. Heiberger* Naomi B. Robbins † Jürgen Symanzik ‡

Abstract
Following the 2013 Statistics Poster Competition, accessible at `http://www.amstat.org/education/posterprojects/2013posters.cfm`, the three authors of this article commented on the the graphical contents of the winning posters. Our comments (Symanzik, Robbins, and Heiberger, 2014) appeared in *The Statistics Teacher Network*, the on-line journal published jointly by the ASA and the NCTM. Based on our review of the 2013 posters and our review of the online documentation for poster designers (`http://www.amstat.org/education/posterprojects/`), we are proposing guidelines for statistical graphics that we would like the competition organizers to consider posting on their website. Our guidelines will include many examples of graphs, both good and bad, with clear description of what features of the graphs are responsible for our evaluation. We are moving in the direction of the booklet for participants that the *Poster Competition* has been contemplating for years. We will do so in a way that is properly diplomatic. For example, only good examples from the participants will be included. All bad examples will be by us and clearly labeled as not from the participants.

Key Words: Poster Competition, Graphs.

1. Background

The *American Statistical Association* (ASA) runs an annual *Poster Competition and Project Competition* (`http://www.amstat.org/education/posterprojects`) under the direction of the *ASA/National Council of Teachers of Mathematics (NCTM) Joint Committee*. The *Project Competition* was initiated in 1987. The *Poster Competition* was initiated by Lorraine Denby of the *ASA Section on Statistical Graphics* in 1989 and it was first held in the spring of 1990 with Jerry Moreno as chair (see `http://www.amstat.org/education/posterprojects/history.cfm` for further details). In recent years, the posters have been judged with no input from the *ASA Section on Statistical Graphics*. Unfortunately, some of the winning posters in the most recent years contained several sub-optimal graphs.

In response, the *ASA Section on Statistical Graphics* formed a special task force in the fall of 2012 (the three authors of this article with Symanzik as the convener) to review graphical content and, in close collaboration with the *ASA/NCTM Joint Committee on the Curriculum in Statistics and Probability*, make recommendations on the structure of the *Poster Competition* and criteria for judging the graphs. Our comments on the 2013 posters were published in Symanzik, Robbins, and Heiberger (2014).

In this article, we first describe our observations from the 2013 posters in Section 2. We continue with bad graphs from the media or constructed by ourselves in Section 3. We outline the plans of our task force for the next two years in Section 4 and we provide a summary of the discussion during the JSM session in Section 5. We finish with our take-home messages from this article in Section 6.

*Department of Statistics, Temple University, Speakman Hall, Philadelphia, PA 19122
†NBR Graphs, 11 Christine Court, Wayne, NJ 07470
‡Department of Mathematics and Statistics, Utah State University, Logan, UT 84322-3900

2. Observations from the 2013 Posters

The winners of the 2013 ASA *Poster Competition* are presented at `http://magazine.amstat.org/blog/2013/08/01/poster-and-project`. In each of four age groups (Kindergarten to Grade 3, Grades 4–6, Grades 7–9, and Grades 10–12), typically three winners were awarded. Sometimes, there were additional posters that deserved honorable mention.

This section features several of the student graphs from the winning and honorable mention posters, along with comments on the positive graphical choices that each graph shows. In the following section, we will also show several other graphs, some that have been published and some that we constructed (and specifically not from the competition) that we use to comment on negative features that are sometimes used in graphical displays.

A variety of excellent graphs were used in the winning posters, including box plots, scatterplots with or without fitted summary lines, bar charts, and maps. There were also the ubiquitous pie charts, some of which were drawn nicely.

Box Plots

The graph in Figure 1 shows the distribution of a response variable for two groups. Both boxes are on the same scale. They are shown with different colors representing the group values. The response y-axis has a sensible descriptive label and properly spaced tick labels. The graph has an informative main title.

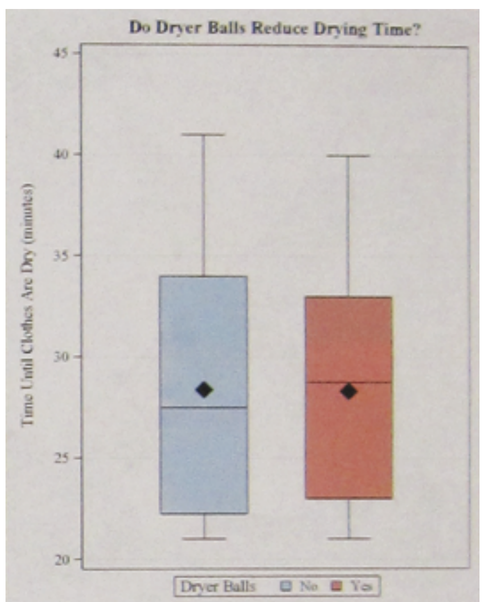

Figure 1: Box plot with two groups.

Scatterplots

The graph in Figure 2 is a scatterplot with a response variable, a predictor variable, and two groups. The plot shows appropriate labels and tick labels on the axes. The groups are plotted with different colors and plotting symbols. The use of two visual characteristics is helpful for readers with normal color vision and imperative for readers with deficient color vision. In addition, the use of distinct plotting symbols for the groups protect the usability of the graph when it is (inevitably) photocopied in black and white. The graph has an informative main title.

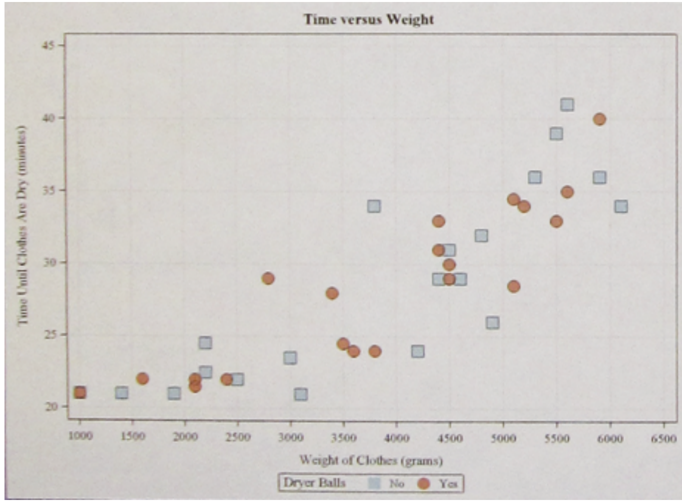

Figure 2: Scatterplot with two groups.

Fitted Summary Lines

The graph in Figure 3 is a scatterplot with one response variable and one predictor variable. The graph includes a fitted summary line. In this graph both axes measure percentages. The graph has been plotted with an aspect ratio of 1, that is, the same number of inches represents 10% on both axes. Again, there are an informative main title, axis labels, and tick labels.

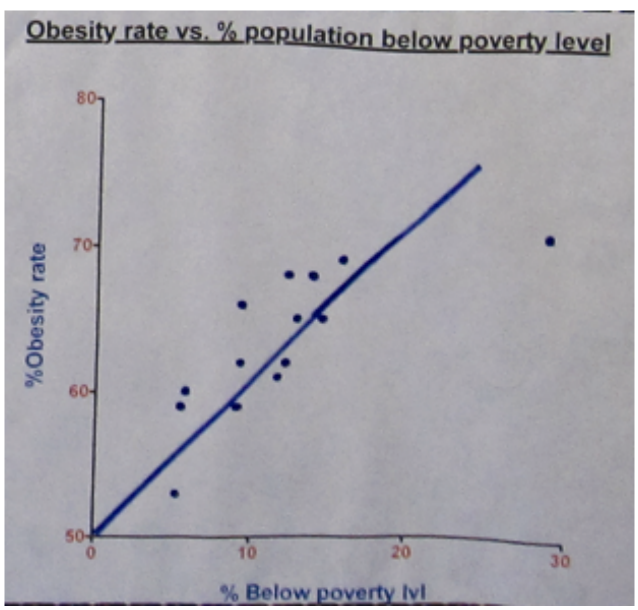

Figure 3: Scatterplot with a fitted summary line.

Maps

The graph in Figure 4 is a map showing the location in geographical coordinates of regions associated with values of the categorical response value. This is data on the North Sea oil fields. The sea is critical for this data and that is where the emphasis is placed in the graph. The land areas are identified as many of the readers will have more familiarity with the political boundaries on the land.

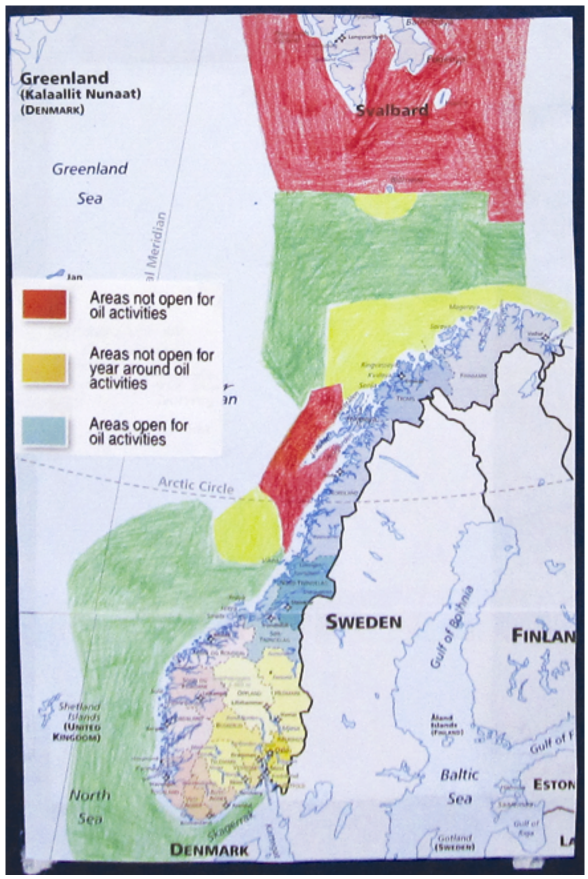

Figure 4: Geographical data displayed on a map.

Pie Charts (Done Nicely)

The graph in Figure 5 is a pie chart. The graphical area displayed is a proper circle aligned with the plotting surface. There is a sequential color choice appropriate for a set of sequential labels. There is one dimension in the data and it is represented by one dimension (the angle) in the graph. There is a main title. The pie slices are labeled with their category names and with their associated percentages. The pie slices are ordered in a meaningful way, here, from the largest percentage to the smallest percentage in a clockwise direction starting at the top ("noon") position.

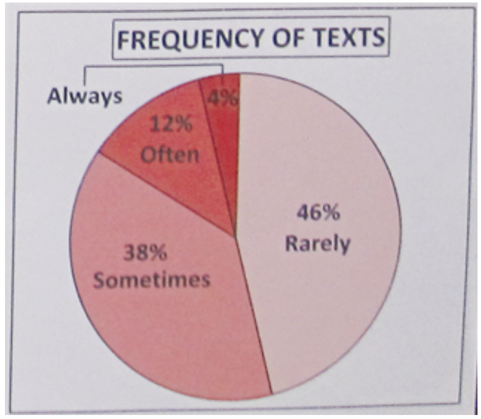

Figure 5: Pie chart done well.

Bar Charts (Done Nicely)

The graph in Figure 6 is a bar chart showing two response variables plotted in parallel for levels of a categorical variable. Two colors are used, one for each response variable. The levels of the categorical variable are sorted by the length of the bars for one of the two variables (which likely is considered to be the more important one by the creators of this graph). This is good, because alphabetical sorting is usually not helpful to the reader. There is a main title and informative axis labels and tick labels.

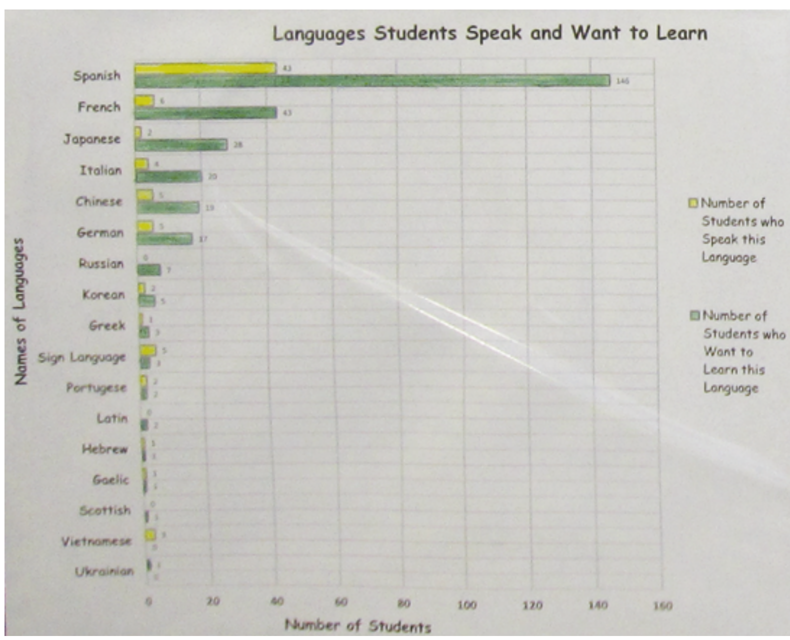

Figure 6: Bar chart done well.

Figures 1 to 6 in this section have been adapted from `http://www.amstat.org/educationposterprojects/2013posters.cfm` (see also Amstat News, August 2013, pp. 26–33, `http://magazine.amstat.org/wp-content/uploads/2013an/August2013.pdf`). Photos are courtesy of the American Statistical Association.

3. Graphs Needing Lots of Extra Work

The next set of graphs begins with one that we constructed to represent some terrible things that could be done. Other graphs in this section are published graphs that need some refinement in order to be effective in their communication goal.

Bar Charts Needing Improvement — This is Our Example, NOT a Student Example

The graph in Figure 7 shows a three-dimensional bar chart, a graphical disaster that occurs very commonly. The pseudo-third dimension thoroughly misleads readers. Before going on, please look at the graph and determine from the graph the heights of the bars.

Some of the specific design errors in this graph are:

- Use of three-dimensional bars for a one-dimensional number at each group level.
- Perspective further distorts graph.
- Lack of a common baseline.
- Gradient colors are confusing — they inappropriately suggest to the the reader that this is a stacked bar chart.
- Excel's "Series 1" label is unnecessary and takes up real estate. The label is actually an attempt at a legend. The colored glyph is intended to indicate a single value, but not a range of values where such color gradients frequently can be found.
- Too many y-axis tick labels.
- No main title and no labels for the two axes.

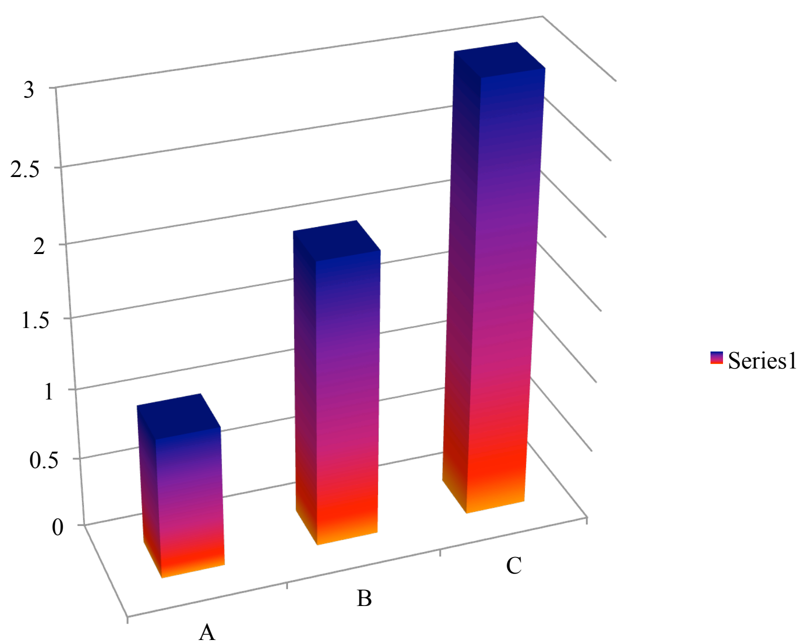

Figure 7: Three-dimensional bar chart: Can you determine how tall the bars are?

Improved Figure

The graph in Figure 8 shows the same data as Figure 7. Here we show one dimension, no false perspective, and a single color for a single bar; hence there is no need for a legend. In addition, we now have a title, labels for both axes, and light-colored grid lines behind the bars. In this presentation, the y-values are unambiguously 1, 2, and 3.

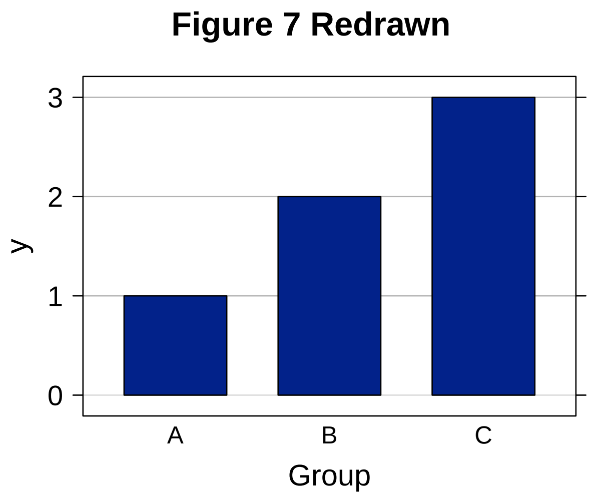

Figure 8: Improved bar chart: The bars have heights 1, 2, and 3.

Truncated Bar Chart (A Misleading Graph)

The graph in Figure 9 shows a truncated bar chart for two categories, with the baseline at zero not shown. It is very clear from the content of the graph that the base is not zero. The effect of this misleading presentation is that 7,066,000 people appear almost three times as large as 6,000,000 people.

Figure 9: Obamacare enrollment from Fox News, retrieved from Wyatt's March 31, 2014 blog: 7,066,000 people appear almost three times as large as 6,000,000 people.

Fixing a Graph

The graph in Figure 10 shows our correctly drawn bar chart that is based on the same numbers that form the basis for the misleading graph in Figure 9. In our bar chart, we see the proper inference that the ratio 7,066,000 to 6,000,000 (≈1.2) is fairly close to 1. Also, our corrected graph doesn't show any of the *chart junk* that is visible in the background of the misleading bar chart in Figure 9. We overlay our accurate plot on the misleading plot in the right panel of Figure 10. We can now see the extent of the distortion of information that follows from displaying a truncated graph.

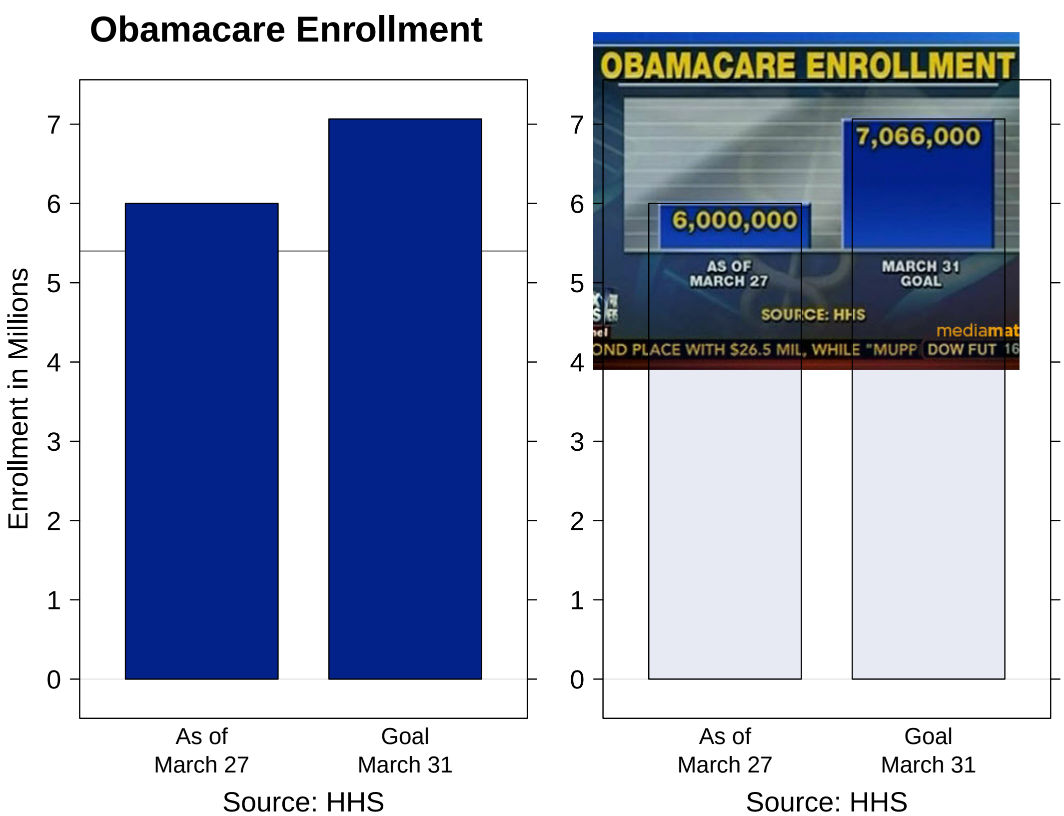

Figure 10: Comparison of misleading graph with accurate graph. The left panel shows an accurately drawn bar chart, with origin at 0, of the same data. We now see that the bar representing 7,066,000 people now appears only slightly larger than the bar for 6,000,000 people. A light horizontal grid line in the left panel shows the location of the inappropriate origin at 5.4 million in the misleading graph. In the right panel, the misleading graph is overlaid with the accurate plot. We can now see that the misleading graph cut off the bottom 80% of the full plot with origin at 0. We can also see how the excessive "chart junk" in the misleading graph occupies more area than the informative section of the graph.

How to Read a Circle

The graphs in Figure 11 show two bubble plots. In the one on top, the values of the response variable determine the diameters and in the one at the bottom, the values of the response variable determine the areas of the circles. The human eyes and brain compare and interpret areas, so of these two alternatives, the area representation at the bottom is visually better for most people.

Neither of these two bubble plot representations of the data is optimal. In both cases, there is only one dimension to the data, so a one-dimensional graph would better represent the information. We show one-dimensional bar charts for both of these bubble plots in Figure 12.

In both of the graphs in Figure 11, the labels of the groups are in minuscule type so the information on what is being compared is almost completely hidden from the reader.

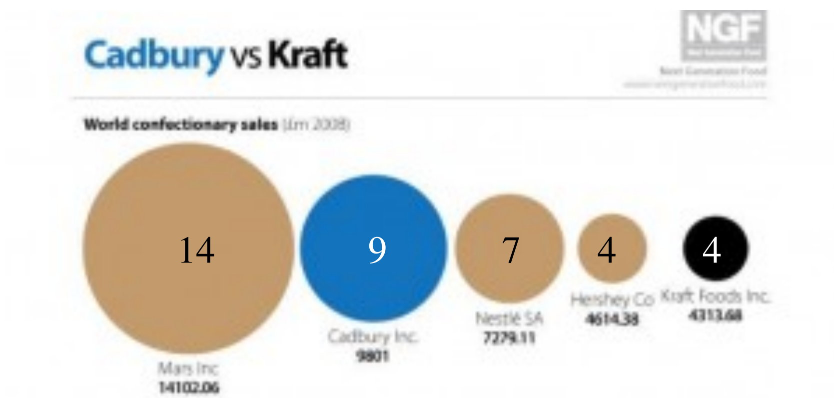

Figure 2. Value encoded by diameter. Source: http://www.nextgenerationfood.com/news/cadbury-sale-agreed/

Figure 3. Value encoded by area. Source: http://www.guardian.co.uk/news/datablog/2010/oct/08/nobel-peace-prize-winners-list-liu-xiaobo#zoomed-picture

Figure 11: Two bubble plots. The one on top uses diameter to represent the values of the response variable and the one at the bottom uses area to represent the values of the response variable. Retrieved from Robbins' February 28, 2012 blog.

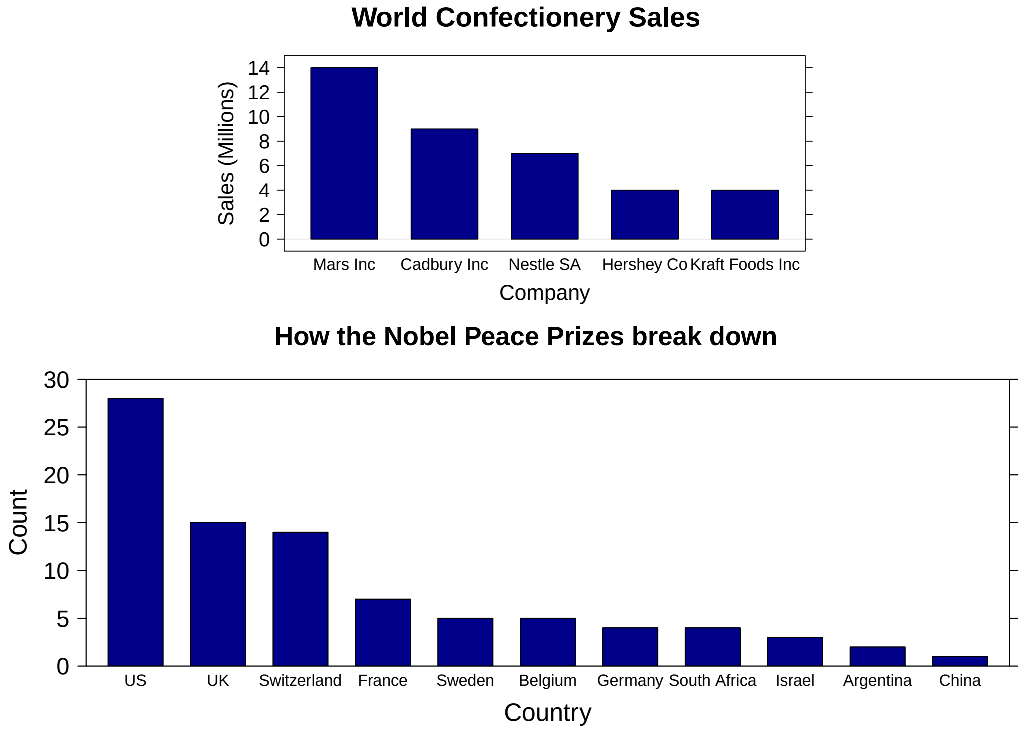

Figure 12: Reconstruction of the graphs from Figure 11 as bar charts: Bars are better suited for one-dimensional data.

4. Future Plans of the Task Force

Our special task force plans to work on the following tasks during the next two years.

Reassessment of the 2014 Winning Posters

In August 2014, the winning posters of the *2014 Poster Competition* were posted at `http://magazine.amstat.org/blog/2014/08/01/poster-winners/`. Our first impression is very positive. We still need to conduct a thorough review of these posters, similar to our review of the 2013 posters. We plan to finalize our review of these posters in the fall of 2014 and make our observations available through the *ASA Poster Competition and Project Competition* web site, similar to our observations from the 2013 posters that were accessible at `http://www.amstat.org/education/posterprojects/2013Posters_Observations.pdf`, prior to the publication of the slightly revised version in *The Statistics Teacher Network*.

Preparation of a "Good Graph Design" Webinar

The ASA web pages host many *K–12 Statistics Education Webinars*, accessible at `http://www.amstat.org/education/webinars/`. These are recorded web-based seminars that deal with statistics education topics from Kindergarten to Grade 12. The majority of the webinars last about 50 to 60 minutes. We have been invited by Rebecca Nichols, the *ASA Director of Education*, to prepare a new webinar on good statistical graphs. Our target date for this webinar is Spring 2015.

Preparation of a Booklet for Future Contestants and Future Judges

According to Jerry Moreno, the first chair of the *Poster Competition*, the organizers of this competition have contemplated creating a booklet of graph design principles. Unfortunately, such a booklet has not been prepared so far. We plan to prepare such a booklet and will elaborate on the types of examples shown in this article and in the forthcoming webinar. Many excellent recommendations on how to create good graphs exist in books and articles, such as Robbins (2013), Su (2008), Tufte (2001), and Wainer (2000). A variety of blogs (by Camoes, Peltier, and Robbins, for example) are dedicated to the creation of good graphs. All of these are focused primarily on adults/professionals. Our focus has to be that this booklet will be meaningful for students from Kindergarten to Grade 12 — and also for their teachers who may not necessarily be trained in statistics, even less in the field of statistical graphics. Therefore, we will focus on delineating *Good Practice* and explain why what we recommend is good practice. We will also illustrate *Bad Practice* and explain why we believe it to be bad. Similar to this article, the good examples will be directly taken from the winning posters while the bad examples will mimic some of the not-so-good examples from the posters, but will be created by us. While a first version of this booklet should be available for the judges of the *2015 Poster Competition*, we expect that it will require a lot of discussions and fine-tuning before the booklet can be made widely available. We hope to have this booklet finalized for the *2016 Poster Competition*. Feedback on the booklet will be highly welcome!

5. Discussion at the JSM Session

During the presentation of our work at the JSM in Boston in August 2014, we tried to initiate a brief discussion with the audience participating in our session. Unfortunately, the session format with seven 15-minute presentations provided time for only a very brief discussion. Apparently, there is more than just one possible answer for each of these questions.

Questions for the Audience

- What is the right standard for evaluating the ASA posters? Should it be the same for early grades as for later grades? — The current *Poster Judging Rubric* is posted at `http://www.amstat.org/education/posterprojects/pdfs/PosterJudgingRubric.pdf`. The *ASA/NCTM Joint Committee* has started discussions on revising the rubric, but no changes have been made so far.

- What restrictions might be placed on the contestants? Would you exclude (or strongly warn against) three-dimensional graphs? — A previous webinar titled "Working with K–12 Students to Create a Statistics Poster" was held by John Gabrosek and Neal Rogness in December 2007. On page 33 of their Powerpoint slides, the authors addressed "Common Pitfalls to Avoid" and only listed "3-D items are not securely attached", but they did not completely ban the use of three-dimensional graphs in general. Their video and slides can be accessed at `http://www.amstat.org/education/posterprojects/PosterCompetitionWebinar12-3-07.wmv` and at `http://www.amstat.org/education/posterprojects/PosterCompetitionWebinarSlides12-03-07.ppt`, respectively.

- Should hand-drawn or computer-drawn graphs be used? — Some opposing perspectives are given in the subsection below.

Hand-drawn or Computer-drawn Graphs

Time did not allow for a detailed discussion of this question at the JSM. The answers given below are our comments, and not those of the audience.

Advantages of Hand-drawn Graphs:

- Helps to think about the best way to present the data rather than what the software menu offers.
- Avoids temptation to add pseudo three-dimensional effects and other decorations (such as *"chart junk"*) that distort the data.
- Allows all students to participate — not just those with computer access.
- Provides a better feel for the data for some participants. Hand-drawing requires the contestants to look at each graphical component carefully before drawing the graph.

Advantages of Using Good Graphics Software for Designing and Drawing Graphs:

- Let the computer do the mechanics and leave more time and energy for the data analysis and graph design steps.
- Well-designed graphics software will provide good graphics examples.
- More likely to be drawn to scale.
- Provides a better feel for the data for some participants. Computer access to the dataset simplifies the ability of the analyst to consider rearranged organization of the data, hence get a better sense about the relation of the variables. The analyst is less locked in to the first graph that is drawn.

Note that the fourth item begins identically for both positions. The second sentences show the reasoning behind the differences. Apparently, no definite answer can be given for this question — and this is apparent in the winning posters: some contain excellent hand-drawn graphs while other contain excellent computer-drawn graphs. The majority of the graphs from the winning posters that are included in this article are computer-drawn, but this is pure coincidence and does not imply that we generally prefer computer-drawn graphs.

6. Take-Home Messages

We want to conclude this article with a few take-home messages.

- Make your collaborators at schools and the teachers of your kids aware of the *ASA Poster Competition and Project Competition*. Regional competitions exist and contact persons are listed at `http://magazine.amstat.org/blog/2014/08/01/poster-winners/`. Unfortunately, regional competitions are held in only a few US states by surprisingly few ASA Chapters and other groups. Although only a few regional competitions exist, students outside areas with regional competitions can still participate by sending their posters directly to the ASA office. There are several hundred participants each year who participate outside regional competitions. In our opinion, many more K–12 students and schools should participate in these competitions and many more ASA Chapters would benefit by cosponsoring regional competitions.

- Software (in particular Microsoft Excel) does not create good graphs by default — we (the graph designers) must know and apply the principles of good graphs. Just using the default settings provided by the software often does not lead to a good graph.

- Many sources and recommendations for good graphs are available. Some excellent suggestions how to create good graphs are provided in books and articles, such as Robbins (2013), Su (2008), Tufte (2001), and Wainer (2000). Several blogs (by Camoes, Peltier, and Robbins, for example) are dedicated to the creation of good graphs.

- We would like to obtain your feedback!!! Please write to the task force (all three of us) in case of any questions, comments, or suggestions:

  - Jürgen Symanzik `juergen.symanzik@usu.edu`
  - Naomi B. Robbins `naomi@nbr-graphs.com`
  - Richard M. Heiberger `rmh@temple.edu`

REFERENCES

Our 2013 Observations

Symanzik, J., Robbins, N. B., Heiberger, R. M. (2014). "Observations from the winners of the 2013 Statistics Poster Competition — Praise and future improvements." *The Statistics Teacher Network*, 83:2–5. `http://www.amstat.org/education/stn/pdfs/STN83.pdf`

Books and Articles

Robbins, N. B. (2013). *Creating More Effective Graphs.* Chart House: Ramsey, NJ. (reprinted from Wiley 2005)

Su, Y-S. (2008). "It's easy to produce chartjunk using Microsoft Excel 2007, but hard to make good graphs." *Computational Statistics and Data Analysis* 52:4594-4601.

Tufte, E. R. (2001). *The Visual Display of Quantitative Information, 2nd edition.* Graphics Press: Cheshire, CT. (first edition 1983)

Wainer, H. (2000). *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot.* Psychology Press: London, UK. (reprinted from Copernicus 1997)

Blogs

Camoes, J. "ExcelCharts," `http://www.excelcharts.com/blog/`

Peltier, J. "Peltier Tech Blog — Peltier Tech Excel Charts and Programming Blog," `http://peltiertech.com/WordPress/`

Robbins, N. B. "Effective Graphs," `http://www.forbes.com/sites/naomirobbins/`

Robbins, N. B. (February 28, 2012). "Misleading Graphs: Displaying a Change in One Variable Using Area or Volume," Forbes.com LLC. `http://www.forbes.com/sites/naomirobbins/2012/02/28/misleading-graphs-displaying-a-change-in-one-variable-using-area-or-volume/`

Wyatt, S. (March 31, 2014). "Dishonest Fox Charts: Obamacare Enrollment Edition," Media Matters for America. `http://mediamatters.org/blog/2014/03/31/dishonest-fox-charts-obamacare-enrollment-editi/198679`