InfoVis and Statistical Graphics: Comment

Paul MURRELL

1. IT IS GOOD TO TALK

The main points of Gelman and Unwin's discussion article are uncontroversial. Yes, statistical graphics and infographics have different goals. Yes, statistical graphics people have an unfortunate tendency to make boring ugly graphs, while infographics people have an unfortunate tendency to trample on the principles of clear communication. Yes, both sides could learn from each other.

And a good way for both sides to learn from each other is to have a discussion like this.

But a discussion relies on two things: *both* sides have to *talk* to each other and *both* sides have to *listen* to each other. Perhaps the issue is that not all of these things have been happening.

1.1 I KNOW YOU CAN HEAR ME

There is ample evidence that the statistical graphics community is talking.

In the peer-reviewed literature, there are numerous famous and accessible publications, including the work of Cleveland and McGill (Cleveland and McGill 1985; Cleveland 1985, 1994; Cleveland 1993), Robbins (Robbins 2005), and Tufte (Tufte 2001, 2010). These publications have been very successful at conveying important ideas for data display, as evidenced by the fact that concepts such as the "data-ink ratio" and "pie charts are the work of the devil" (people are poor at judging angles) are widely known and widely reported.

There are also many examples of comment and criticism from the statistical graphics community in less rigid forums. This discussion provides one example and this is not the first such example to be published in this journal (Gelman 2011). There are also many examples of members of the statistical graphics community casting a critical eye over graphical displays in various blogs and online forums (Gelman 2012; Theus 2012).

The only possible problem is that the statistical graphics community is communicating their message too much or too often. It cannot be claimed that the statistical graphics community is hiding their message under a bushel.

Paul Murrell is Senior Lecturer (E-mail: *paul@stat.auckland.ac.nz*), Department of Statistics, The University of Auckland, Auckland, New Zealand.

Journal of Computational and Graphical Statistics, Volume 22, Number 1, Pages 33–37
DOI: 10.1080/10618600.2012.751875

1.2 READ MY LIPS

Not only is the statistical graphics community sending out signals, but the message is also straightforward and clear.

The basic principles that have been established for clearly and efficiently presenting data are relatively few and relatively simple. So few and so simple in fact that they can be summarized in a short list:

1. Display data values using position or length.
2. Use horizontal lengths in preference to vertical lengths.
3. Watch your data-ink ratio.
4. Think very carefully before using color to represent data values.
5. Do *not* use areas to represent data values.
6. *Please* do not use angles or slopes to represent data values.
7. *Please*, *please* do not use volumes to represent data values.

This list provides all of the knowledge that is necessary to avoid the worst mistakes when producing data visualizations. The list can be easily learned and is easy to understand and, as mentioned in the previous section, the points on the list are constantly being broadcast far and wide.

1.3 ARE YOU LISTENING TO ME?

So if there is a simple message and the message is being sent, why is the message being ignored?

Why is it that it is astoundingly easy to find egregious violations of the basic rules of data display any day of the week?

Partly, as suggested by Gelman and Unwin, this may be because of conflicting goals. Sometimes good communication is sacrificed for a greater emphasis on grabbing attention. But is it also the case that, as Gelman and Unwin speculate, the infographics community needs to listen more?

1.4 I'M ALL EARS

As well as being good talkers, there is ample evidence that the statistical graphics community is prepared to listen. Although there are again conflicting goals and we may be prepared to sacrifice sophisticated cool for clarity of display, the statistical graphics community is prepared to include aesthetic features where possible.

Within the R project, this is evidenced by the attempts of some graphics package authors to provide more pleasant defaults for plots. Example of this are the `lattice` package (Sarkar 2008) and `ggplot2` (Wickham 2009).

There has also been a lot of work done to ensure that the aesthetic features of a graph can be controlled and that more artistic graphical components can be included in plots that

are drawn with R. Examples of this are the ability to control details such as line endings (Murrell 2004), the ability to include external images in plots (Murrell 2009b, 2011), and support for drawing curves and diagrams (Murrell 2009a).

1.5 THAT ALL SOUNDS VERY NICE

The statistical graphics community is listening, but what are we listening to?

What is the message that is being sent from the infographics community?

One of the best sources of good graphic design information that the statistical graphics community has been aware of for a long time is the work of Edward Tufte. This is the source of the often-quoted "data-ink" ratio.

The problem is, at the risk of over-simplification, there is frustratingly little else. Tufte has produced several books, all of which contain numerous examples of great and good graphical displays, but much of the message is anecdotal or demonstration by example. It is a great pleasure to read Tufte's books, but it is very hard to come away with general principles or simple rules that can be applied to your own work. It is difficult to absorb Tufte's genius.

And this problem is not just limited to Tufte's work. Most of the other publications that I have seen that attempt to address the more aesthetic features of graphs also rely heavily on anecdote (Wainer 2005).

The examples mentioned so far come from individuals who span both the statistical graphics and infographics communities. What about the infographics community itself? Many books have been published recently on data visualization, particularly for use on the web, but they are also heavy on beautiful examples and light on simple lists of easily understood and easily applied guidelines.

The statistical graphics community is listening, but all we can hear is white-noise. We need the infographics community to talk to us in words of one syllable.

1.6 ARE YOU HIDING SOMETHING FROM ME?

Where is the simple list of infographics do's and don'ts?

There is a suspicion that more could be done here. For example, it is only relatively recently that the statistical graphics community has become aware of the benefits of working in color spaces that have rational dimensions and useful psychophysical properties. Within the R project, this knowledge has been hungrily devoured and used to produce packages such as `colorspace` (Zeileis, Hornik, and Murrell 2009; Ihaka et al. 2011) and `RColorBrewer` (Neuwirth 2007). These packages do not turn a statistical graphics person into a designer, but it is a significant step that when a statistical graphics person is trying to select a set of colors she is no longer wandering aimless in the RGB wilderness.

Are there similar useful summary guidelines for fonts that could keep the statistical graphics community away from the worst typographical atrocities? What about simple (design) guidelines on the use of white space in layout?

1.7 TELL ME WHAT YOU REALLY THINK

Perhaps infographic design is just too hard to crystallize into a simple bullet-point list. If there is not a simple list of guidelines, is there anything else that the infographics community could usefully say to the statistical graphics community?

Yes there is. Recall that there are plenty of examples of efforts by the statistical graphics community to provide criticism of plots and graphs. One of the ways that the statistical graphics community is talking to other communities is by taking examples of poor graphical displays and suggesting improvements.

Would the infographics community be prepared to return the favor?

Where are the forums that take a boring and ugly statistical graphic and try to point out how to jazz it up a little?

In the absence of a simple general rule of infographics design that could be used to improve all statistical graphics forevermore, it would still be useful to get concrete examples of how to improve upon poorly designed statistical plots. Perhaps the infographics community would be prepared to educate the statistical graphics community one plot at a time.

1.8 STATISTICIAN TEACH THYSELF

The argument could be made that infographic design is hard. You cannot expect to become a designer just by reading a simple bullet-point list. It could be argued that the only way to get better is to take the time to educate yourself properly.

Why are we not prepared to send every person who is ever going to draw a graph on an infographic design course?

For a start, before they go on a design course, everyone who is ever going to draw a plot should really go on a course in statistical graphics. Unfortunately, as Gelman and Unwin lament, there is a lack of perceived importance of statistical graphics within statistics. If we struggle to include statistical graphics within the standard statistics curriculum, the chances of adding a design course must be lower still.

The statistical graphics community needs something relatively simple and easily consumable. We cannot expect everyone to train in infographic design, but hopefully there are some rules of thumb (match your belt and shoes?) that would make everyone just a little more infographically hip.

1.9 DO NOT TAKE THIS THE WRONG WAY

Gelman and Unwin, both coming from the statistical graphics camp, are at pains to emphasize that they are not just trying to have a moan about infographics. This sets a civilized tone for the debate, which will hopefully encourage a useful discussion.

I also come from the statistical graphics perspective. I have indulged in a little more criticism, but I also want to emphasize a fundamental respect for the other side. While I have tried to suggest that it would be nice if the infographics community could pay us more attention, my main complaint is not that infographic is doing the wrong thing; I like and envy the smart, sophisticated look that they can create. My complaint is that they will not tell me how to do what they are doing!

2. IN CONCLUSION

The main message of this contribution to the discussion is a plea to the infographics community. It would be nice if you could listen to us more, but that is really up to you.

What you could *really* do for us is this:

1. If you have a bullet list hidden away somewhere, please put us out of our misery and let us have it.
2. If not, then do what we do and have a go back at us. Pick some boring statistical graphics and rip them to shreds; tell us what we are doing wrong and how to get better.

Thanks in advance!

REFERENCES

Cleveland, W. S. (1993), *Visualizing Data*, Summit, NJ: Hobart Press. [33]

——— (1985, 1994), *The Elements of Graphing Data*, Summit, NJ: Hobart Press. [33]

Cleveland, W. S., and McGill, R. (1985), "Graphical Perception and Graphical Methods for Analyzing Scientific Data," *Science*, 229, 828–833. [33]

Gelman, A. (2011), "Why Tables Are Really Much Better Than Graphs," *Journal of Computational and Graphical Statistics*, 20, 3–7. [33]

——— (2012), "Statistical Modeling, Causal Inference, and Social Science" [online]. Available at *http://andrewgelman.com/category/statistical-graphics/*. [33]

Ihaka, R., Murrell, P., Hornik, K., and Zeileis, A. (2011), "Colorspace: Color Space Manipulation," R package version 1.1-0, R Development Core Team. [35]

Murrell, P. (2004), "Fonts, Lines, and Transparency in R Graphics," *R News*, 4, 5–9. [35]

——— (2009a), "Drawing Diagrams With R," *The R Journal*, 1, 15–21. [35]

——— (2009b), "Importing Vector Graphics: The GrImport Package for R," *Journal of Statistical Software*, 30, 1–37. [35]

——— (2011), "Raster Images in R Graphics," *The R Journal*, 3, 48–54. [35]

Neuwirth, E. (2007), "RColorBrewer: ColorBrewer Palettes," R package version 1.0-2, R Development Core Team. [35]

Robbins, N. (2005), *Creating More Effective Graphs*, New York: Wiley. [33]

Sarkar, D. (2008), *Lattice: Multivariate Data Visualization With R*, New York: Springer. [34]

Theus, M. (2012), "Statistical Graphics and More," available at *http://www.theusrus.de/blog/*. [33]

Tufte, E. R. (2001), *The Visual Display of Quantitative Information* (2nd ed.), Cheshire, CT: Graphics Press. [33]

——— (2010), *Beautiful Evidence*, Cheshire, CT: Graphics Press. [33]

Wainer, H. (2005), *Graphic Discovery*, Princeton, NJ: Princeton University Press. [35]

Wickham, H. (2009), *ggplot: Elegant Graphics for Data Analysis*, New York: Springer. [34]

Zeileis, A., Hornik, K., and Murrell, P. (2009), "Escaping RGBland: Selecting Colors for Statistical Graphics," *Computational Statistics & Data Analysis*, 53, 3259–3270. [35]