Exploring All of Your Options

Data Visualization

Written by: Jeffrey A. Shaffer
February 21, 2013
Jeffrey A. Shaffer is the Vice President of Information Technology and Analytics at Unifund. Mr. Shaffer joined Unifund in 1996 and has been instrumental in the creation and development of the complex systems, analytics and business intelligence platform at Unifund. Mr. Shaffer holds a BM and MM degree from the University of Cincinnati and an MBA from Xavier University where he was the winner of the 2006 Graduate Student Scholarly Project in Research. Mr. Shaffer has attended the Harvard Business School's Executive Education Program, is a Certified Manager of Quality and Organizational Excellence through the American Society for Quality, a Certified Project Management Professional through the Project Management Institute and has completed Six Sigma Green Belt and Black Belt training with the Xavier Consulting Group. Mr. Shaffer is also Adjunct Assistant Professor at the University of Cincinnati in the Carl H. Lindner College of Business teaching Data Visualization in the Graduate Course series for Data Analytics. He is also a regular speaker at business intelligence conferences and symposiums on the topic of data visualization, writes for the data visualization blog at MakingDataMeaningful.com for LÜCRUM, Inc. and was a finalist in the 2011 Tableau Interactive Visualization Competition.

For more from Mr. Shaffer visit www.MakingDataMeaningful.com or visit his blog at www.HighVizAbility.com.
Exploring All of Your Options: Data Visualization

I recently read a blog post by Naomi Robbins from Forbes entitled *Line Charts are Not Always the Best Way to Show Time Series*. For those of you who are not familiar with Naomi Robbins you should add her to your reading list because she has some terrific posts regarding data visualization practices ([http://blogs.forbes.com/naomirobbins/](http://blogs.forbes.com/naomirobbins/)). This was an interesting post regarding her recent internet usage on a trip using a mifi for internet access. She writes:

“Many blogs and books suggest line charts or area charts similar to this one for time series but they don’t work in this case since interpolation does not make sense.”

It is true about blogs and books suggesting line charts for time series data. In fact, when teaching data visualization at the University of Cincinnati I always reinforce to my students that time series data is best as a line chart. This is because we, as readers, typically understand time when plotted on the x-axis and we typically want to see a trend over time. This is the biggest advantage of a line chart as it shows trend over time better than any other chart type.

Time series data is in two categories. In statistics we refer to this as continuous and discrete. Time series data that is continuous is where we have an observation at every point in time, for example an electrocardiogram at a hospital monitoring a patient’s vital signs. Time series data that is discrete is where the observation is at spaced intervals, frequently evenly spaced intervals, for example monthly sales revenue. One might conclude from this that line charts would work better for the continuous data and a bar chart would be better for the discrete data, however, even discrete time series trends are typically displayed better as a line chart. Again, this is because the biggest advantage of a line chart is that it shows a trend over time very well.

The example used by Naomi Robbins is her daily internet usage for what appears to be one monthly billing cycle, from December 7th to January 6th. I agree with her that the original area graph provided is not the best way to visualize this data. However, in this case a simple line chart does work well, even for the discrete time series data. Consider her redesign vs. a simple line chart.
In this case the line chart shows a clear trend for the period as well as the zero values. Although the zero values are not ideal, they are not distracting from the overall message here. The trend is clear as well, we have an increase up through December 9\textsuperscript{th} followed by a decline, with very light use on December 14\textsuperscript{th}. I would not consider this data as the exception to break the rule of line charts and further, since many users would have data over a the full billing cycle, a line would be better for covering a full month of days rather than displaying 30 bars each month. I’m sure the application developer would rather avoid using special rules for showing different data in different ways.

There are situations with certain datasets where a bar chart might be better than a line chart. One example would be fragmented dates where there are large gaps of time or unequal time intervals. However, even in these cases a line chart can be adapted to fit the needs of the data. Stephen Few has a terrific example of this on his blog in a redesign of Time Magazine vs. Newsweek (http://www.perceptuledge.com/example14.php). In this example he uses the lines spread out over their appropriate interval and connects unknown points with dotted lines.

Another example is when there are lots of zero values in the data, and if you plot it as a line chart then there would simply be a line at zero along the x-axis for the majority of the data. In this case, it would be better to just plot the point that has the data and not all of the zeros. Consider this example below which one of my students created (as part of a dashboard) showing rainfall in inches over a 4-year period of time on Halloween.

Why is a line chart difficult in this example? The primary reason is because there is no trend to show here, and 3 of the 4 values are zero. In this case plotting that single point would be better. A bar chart would have worked too or a simple dot for the point. In this case I chose a lollipop chart for the single point of data.

A few design points to mention as well. I adjusted the scale because in the original the .75 inches was plotted as a very large number when in fact it is not that much rain for the day and was really just light rain during trick or treating. Another redesign point to mention here is that there is no reason to list 9 points on the y-axis to give the reader the ability to approximate 1 point of actual data, so in this case it’s better to simply plot that data point and remove the y-axis completely.
Naomi continues her discussion in another blog post entitled *How to Position Y-Axis Labels in Graphs*, and in this post discusses the various placements of the y-axis labels. She offers four comparisons for the data usage chart and explains why the y-axis title is positioned vertically. See her examples here on her blog: [http://blogs-images.forbes.com/naomirobbins/files/2013/02/YaxisPositions.jpg](http://blogs-images.forbes.com/naomirobbins/files/2013/02/YaxisPositions.jpg).

She offers some comparisons of the chart explaining that the vertical placement of the y-axis title is better than some other alternatives and based on the alternatives she offered I would agree. However, I don’t think we’ve considered all of the available options. There are simple alternatives that would avoid this problem and when creating data visualizations we should not feel like we are boxed in or stuck with making bad choices. I will use her bar chart design to make some comparisons and try to demonstrate a number of different options. It appears that Naomi was keeping the overall size of her chart the same, so I will do the same and keep each chart’s overall footprint exactly the same.

In the first comparison below I am comparing her design to the design I used above. Simply adding a subtitle at the top denoting that the units are “in gigabytes” eliminates the need for the y-axis title completely. Another option is to add “in GB” after Usage in the main title.

![Comparison Chart 1](http://blogs-images.forbes.com/naomirobbins/files/2013/02/YaxisPositions.jpg)

In this next redesign, I chose to drop the word “usage” from the y-axis title.

![Comparison Chart 2](http://blogs-images.forbes.com/naomirobbins/files/2013/02/YaxisPositions.jpg)

You could also provide a very detailed description in the y-axis title if necessary and still keep it vertical by breaking it up on more than one line, although in this case I don’t think this is necessary. Also a better title for the chart might help.
Another interesting point is that Naomi chose to remove the redundant information on the y-axis, the label “GB” over and over again, yet on the x-axis this remains with “Dec”. Each date has “Dec” being used one after the other. In her four-chart redesign this causes an issue because when using the day and month together the x-axis labels will auto rotate in Excel when the space is limited, which should be avoided. Therefore she contends that rotating the y-axis label is better. However, using any of the options above would avoid this issue. Even if it could not be avoided we still have other alternatives.

There are even options as we cross over months, for example if this data happened to start on December 28th instead of December 7th.
After having walked through all of these steps, I think a very simple line chart is the best way to visualize this data. Even with the discrete dates over a very short period of time and the two days of zero value, the simple line chart serves the purpose of displaying the trend to the reader.

Finally, I would like to point out some finer design details that you may or may not have noticed. If you look back through the charts that I presented you will notice that I avoided black text. This was done purposefully. When black text is used it’s as if you are starting a music concert at the loudest point. The only place to go from there requires another use of emphasis, for example, bold, italics or a larger font size. You will have much more control over the reader’s attention if you widen your range. In this case, by using gray as the default it will give you the ability to drive attention in more ways. By changing your default font color to a medium to dark gray (along with gridlines, axis lines, titles, etc.) then you will have more options when you want to drive attention to a specific area of your visualization or maybe a single data point.
In this example, the only emphasis added to the single data label is by using a black font color contrasting the rest of the visualization that is using gray. The black font color here stands out almost as much as the title that is in a much larger font size and bold.

![Internet Usage from Dec. 7 to Jan. 6](chart1.png)

This final example is taken from a Site Statistics chart in WordPress. This is a great example of where they used a design very similar to Naomi’s redesign. In this case, a line chart would have produced zero values for 12 of the 15 days and the technique used in this case, the same technique as Naomi’s recommendation, achieved a much better solution. Also notice the very clean graph with many of the same features described above, as well as a band in gray for the weekend days. This graph is very well done.

![Site Stats Chart](chart2.png)

I hope that you find these tips helpful and happy charting!
LÚCRUM, Incorporated 2013

Contact Us Today:
Erin Macdonald | Marketing Associate
emacdonald@lucruminc.com
513.241.5949 x234

LÚCRUM, Inc.
7755 Montgomery Road, Suite 160
Cincinnati, Ohio 45236

LÚCRUM creates long-term relationships with its clients to build Business Intelligence and Analytics Solutions that increase profitability, drive operational efficiency and maximize decision-making capabilities.

Visit our website:
www.lucruminc.com
Like us on Facebook:
www.facebook.com/LucrumInc
Follow us on Twitter:
@LucrumInc
Connect on LinkedIn.