
The Dos and Don'ts of Communicating with Data: The double y-axis

Friday, November 1, 2019

Author(s)

Samantha Tyner

This post is a version of another posted on the author's [personal blog](https://sctyner.github.io/personal-blog) (<https://sctyner.github.io/redoing-graphs.html>).

On a [previous blog post](https://sctyner.github.io/redoing-graphs.html) (<https://sctyner.github.io/redoing-graphs.html>) I explain in detail how to reconstruct a graph to make clear the information it aims to convey. This post expands on that post by providing further insights into why a double y-axis can be a misleading form of data visualization.

Don't: the double y-axis

The double y-axis is a common data visualization *faux pas*. This visual monster frequently rears its ugly head when the author is trying to make a statement about the relationship between changes in two variables over time. At the top of this page, for example, there is a graph showing that arcade revenue and the number of computer science PhDs awarded are highly correlated. There is,

however, no evidence that arcade revenue and computer science degrees are related. This graphic was made to warn people of the dangers of these types of “[spurious correlations](https://www.tylervigen.com/spurious-correlations)” (<https://www.tylervigen.com/spurious-correlations>).

"Total revenue generated by arcades correlates with computer science doctorates awarded in US"
Data sources: U.S. Census Bureau and National Science Foundation | Image source:
<https://www.tylervigen.com/spurious-correlations>

When seeing the data in the same visual space, however, the brain immediately begins making connections. We can come up with all sorts of hypothetical justifications that arcade revenue and computer science PhDs could be related. After all, if we picture the stereotypical arcade game player and doctor of computer science, they will probably look pretty similar: nerdy, male, enthusiastic about something very few people can relate to, wears glasses because his eyes are strained from looking at screens, etc. This is the danger of the double y-axis: our minds are so powerful that they can create connections and relationships where none exist. The double y-axis just adds fuel to the fire.

A typical example

An example of the double y-axis “in the wild” can be found in [this tweet](https://twitter.com/NWSDesMoines/status/1097357137158828032) (<https://twitter.com/NWSDesMoines/status/1097357137158828032>) from the National Weather Service of Des Moines, Iowa. The plot shows daily snowfall on the left for the 2018-19 winter season, where the scale ranges from 0-6 inches. On the right, the cumulative snowfall for that season is shown, ranging from 0-45 inches. The resulting combined plot of daily snowfall and cumulative snowfall makes the daily values seem far more extreme than they actually are. At first glance, the viewer can easily interpret the graph to mean that there are daily snowfall values of over 30 inches! Although daily snowfall this extreme is not [impossible](https://weather.com/science/weather-explainers/news/monthly-seasonal-daily-snowfall-records-united-states) (<https://weather.com/science/weather-explainers/news/monthly-seasonal-daily-snowfall-records-united-states>) in the U.S., the maximum daily snowfall shown barely exceeds 5 inches.

Do: save your reader’s time

A revised version of the weather service’s plot is shown [here](https://sctyner.github.io/figure/source/2019-02-18-redoing-graphs/facetgeoms-2.png) (<https://sctyner.github.io/figure/source/2019-02-18-redoing-graphs/facetgeoms-2.png>). There are several reasons this plot represents the data better:

- **Units take up the same amount of space:** one inch of daily snowfall takes up the same visual space as one inch of cumulative snowfall
- **Data of the same scale are grouped together:** the top panel shows the cumulative snowfall for the 2018-2019 season and the average cumulative snowfall from 1981-2010, while the bottom panel shows the corresponding daily measurements.
- **Color corresponds to shared groups:** where the first graph has three colors, the revised plot only needs two colors, dark blue for the 2018-19 season and light blue for the 30-year average

As a result, the revised plot requires less mental effort from the viewer: there are fewer axes and colors to process, and there is no deciphering which line goes with which axis. The goal of data visualization is to communicate quickly and effectively with your reader. Do not make them work harder than absolutely necessary; [be gone dual y-axis!](http://www.storytellingwithdata.com/blog/2016/2/1/be-gone-dual-y-axis)

(<http://www.storytellingwithdata.com/blog/2016/2/1/be-gone-dual-y-axis>)

If you learned from this post and what to know *how to* reconstruct these graphs, see the [original post](https://sctyner.github.io/redoing-graphs.html) (<https://sctyner.github.io/redoing-graphs.html>) on my blog.

Author: Samantha Tyner, PhD

Dr. Tyner is a 2019-20 STPF Fellow with the [Office of Survey Methods Research](https://www.bls.gov/osmr/) (<https://www.bls.gov/osmr/>) at the [Bureau of Labor Statistics](https://www.bls.gov/) (<https://www.bls.gov/>). She is an applied statistician with interests in data science, data visualization, forensic science, machine learning, text mining, and network analysis. You can follow her on Twitter at [@sctyner](https://twitter.com/sctyner) (<https://twitter.com/sctyner>).

Source URL:

<https://www.aaaspolicyfellowships.org/blog/dos-and-donts-communicating-data-double-y-axis>

List of links present in page

- <https://sctyner.github.io/redoing-graphs.html>
- <https://www.tylervigen.com/spurious-correlations>
- <https://twitter.com/NWSDesMoines/status/1097357137158828032>
- <https://weather.com/science/weather-explainers/news/monthly-seasonal-daily-snowfall-records-united-states>
- <https://sctyner.github.io/figure/source/2019-02-18-redoing-graphs/facetgeoms-2.png>
- <http://www.storytellingwithdata.com/blog/2016/2/1/be-gone-dual-y-axis>
- <https://www.bls.gov/osmr/>
- <https://www.bls.gov/>
- <https://twitter.com/sctyner>