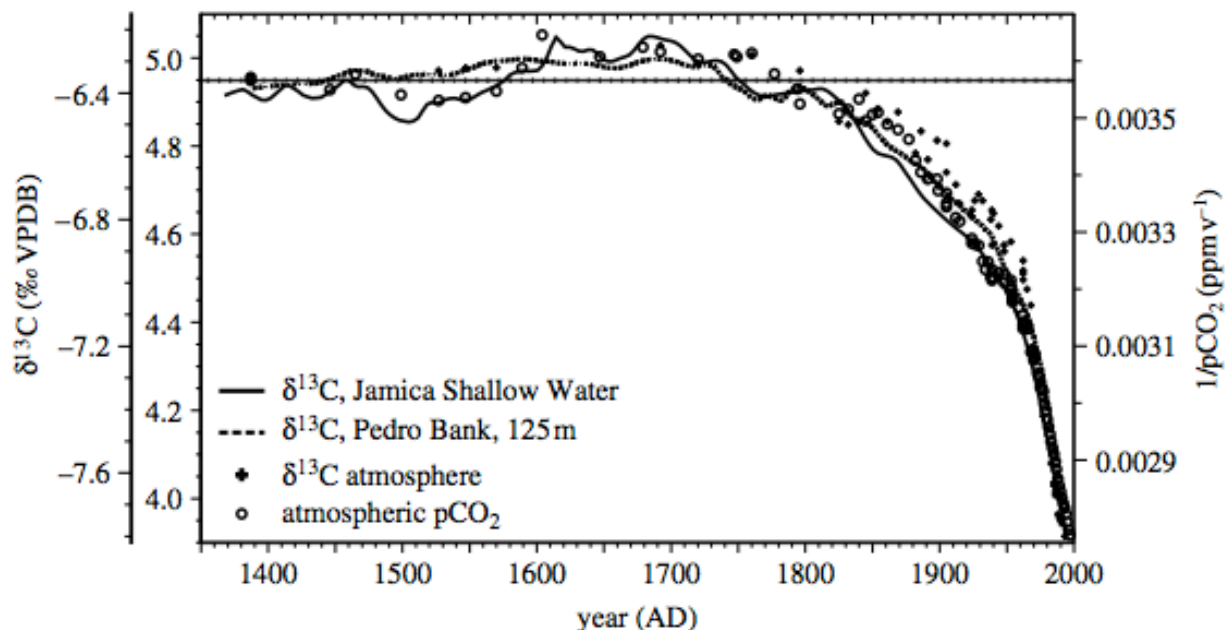


OmniGraphSketcher Best Practices: Stacked Y-Axes

One of the most common requests we get for OmniGraphSketcher is to support multiple y-scales overlaid on the same graph, to make it possible to plot several different types of related data on the same chart. Some people call this approach "dual y-axes." For a while now we've been debating whether or not this is a good idea.

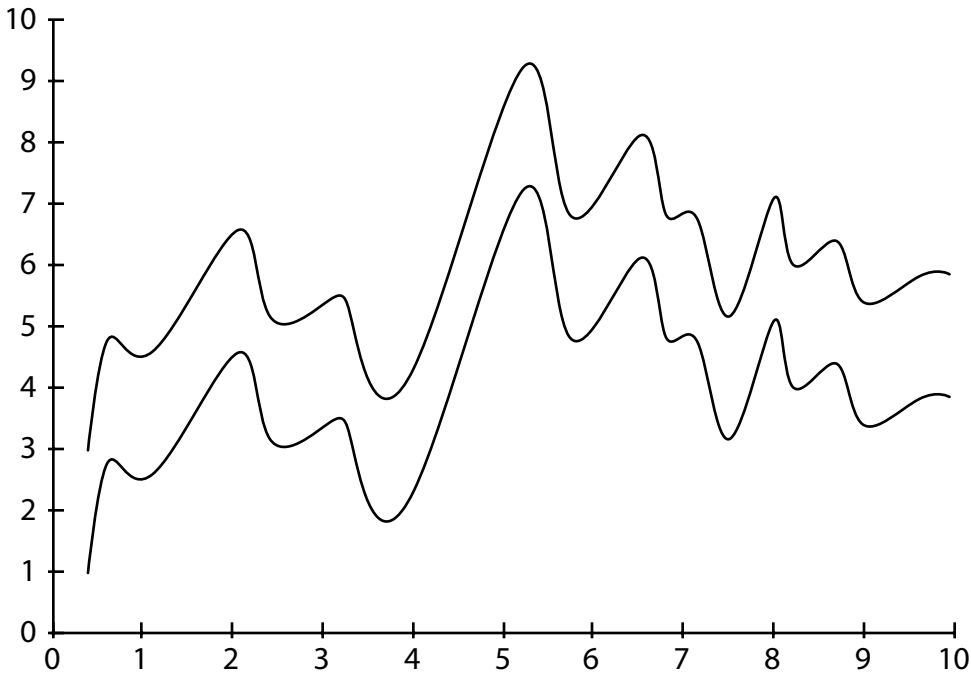
The best argument I've heard for dual axes is that it allows you make extra-efficient use of space, which can be especially important when space is at a premium (for example, in scientific papers with limited length). If you just have to fit all sorts of data into a very small area, it might be that the best way to do it is to overlay all of the data onto the same chart space, using multiple scales and legends so that the reader can decode it later. Here is a fairly extreme example, showing four data series with three different scales ([source](#)):



The trouble with dual charts like this is that they're hard to read. The dual axes "duel" for your attention. You have to keep reminding yourself which scale applies to which data series. Color-coding can help, but there's always going to be a cognitive disconnect when you look at a data point on the left side that actually corresponds to a separate axis scale on the right side. Moreover, it's harder to see the individual shape of each distinct data series. Overall, the approach is analogous to writing a letter with multiple layers of handwriting (each layer written in a different direction). It maximally conserves space but makes it much harder for the reader to decode.

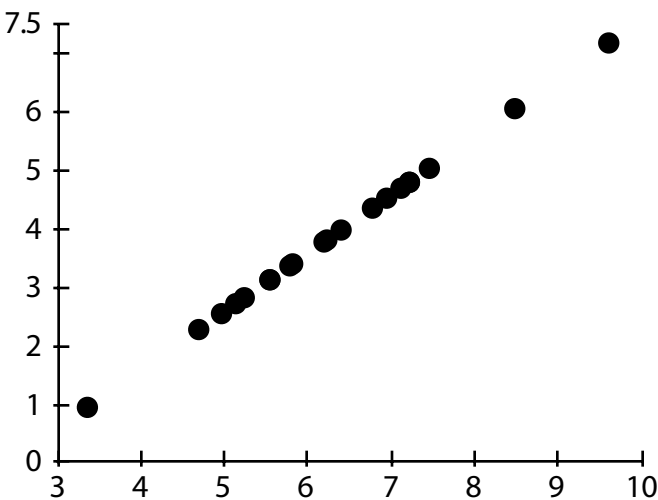
The other argument we've heard in support of dual y-axes is that putting two data series really close to each other lets you compare the data more easily. This seems to make sense intuitively; certainly, the graph above makes the point that the data series are highly corre-

lated. But... exactly HOW correlated? Humans aren't actually that good at comparing data series that are right next to each other. How correlated are these two data series?



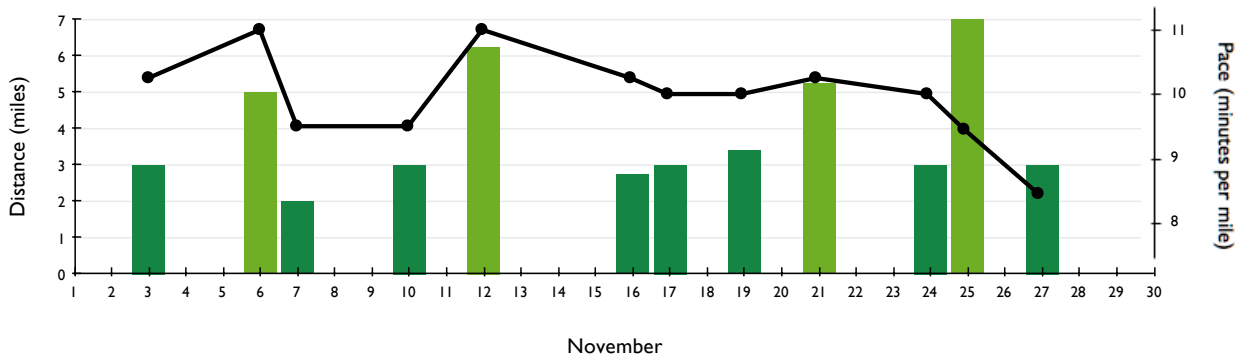
It turns out that they're exactly the same (just shifted vertically). Shouldn't that have been obvious right away?

If your goal is to demonstrate correlation, it is best to actually plot the data series against each other. For the extreme case above, it becomes immediately obvious that the data are perfectly correlated:



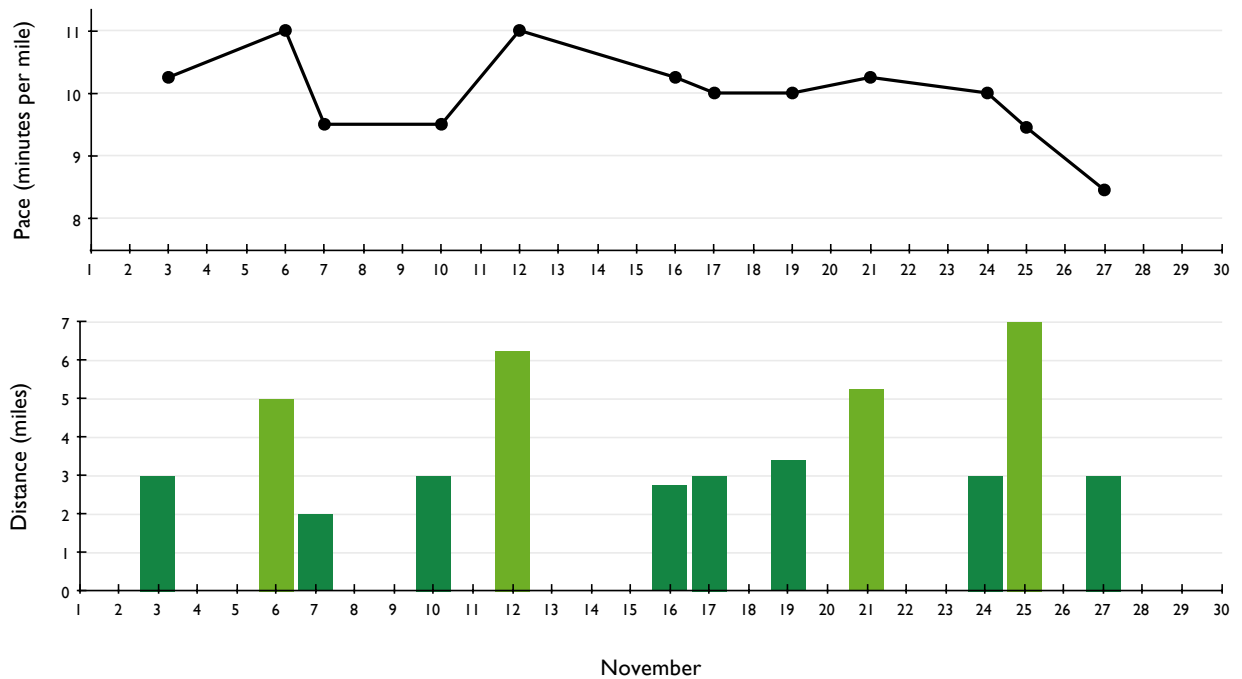
But let's stop theorizing and look at a real-life example. Conveniently, Linda has been working with the OmniGraphSketcher team on how to best present her fitness-tracking data. She recorded the distance and time for each of her runs in the month of November. She's inter-

ested in seeing how her running pace compares with the distance she runs. For many people, the first instinct might be to use dual y-axes to compare these data series. Here is a mockup of how that would look if you could do it in OmniGraphSketcher:



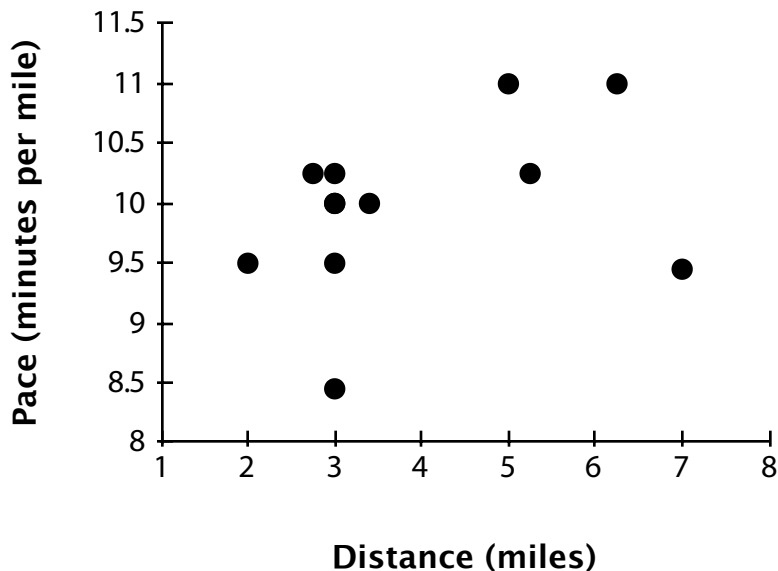
My first reaction is that the graph looks "impressive" because there's a lot going on. Next, I see that there are two data series and two scales. Looking a bit harder, I notice that the distance and pace data are roughly correlated; the pace tends to be slower when the run distances are longer, except near the end of the month. Then I step back again and think, the graph seems cluttered. When I want to know what the data values are (what was Linda's pace on Nov. 6?), I have to figure out which scale to look at and whether I can use the grid lines. Should grid lines be present at all?

In other words, I'm impressed at first, but disappointed when I try to actually *use* the graph. Let's try the alternative: stacking the two data series in separate but aligned graphs. This is the approach that Edward Tufte calls "small multiples." (In this case, it's probably more accurate to just call it "multiples.")



This configuration takes more space. But several aspects of the data immediately become more clear. First, it's obvious that I have two separate data series. Each individual data series is easier to follow -- I can immediately notice (1) how Linda's pace improved over the month and (2) her pattern of jogging distances. Reading the scales becomes trivial, and the grid lines are always there to help. And it actually seems easier to compare the two series to each other, since I can see them individually so much better. Since the x-axes are aligned, it's still easy to notice, for example, how the long runs on November 6 and 12 correspond with slower paces. Overall, the presentation leaves itself "room to breathe."

But if the real goal is to correlate these two data series, we should also plot them against each other. To do that, we ignore the x-value (day of the month) and plot the distance against the pace for each run. It looks like this:



Sure enough, longer runs tend to have a slower pace. But this graph also makes it clear that the correlation is not very strong. Moreover, the outliers really stand out, such as the 7-mile run at a fast 9.5-minute-mile pace.

Conclusion? Overlaying data with disparate scales on top of each other saves space, but makes the data significantly harder to interpret. Instead of "dueling" y-axes, use a set of multiple graphs, aligned along their common axis. Graph Sketcher does not have a built-in option for this, but you can Copy As Image with each graph and paste into OmniGraffle or another program with layout capabilities.

When space is a serious issue, overlaid y-axes become tempting, but situations that really have severe space limits are becoming rare. In a scientific paper, for example, you can display a representative data set and refer the reader to a website for the comprehensive visualization.

But if you have a graph with overlaid scales that you think more effectively compares data than its separated-multiples counterpart, please let us know -- we do still have an open mind about it.