REDESIGNS

By this chapter, your data visualization toolbox contains much more than it did when you began this book. We've seen dozens of graphs, many of which may have been new to you. As you develop your own eye for data visualization, you'll find places where these new graph types may be especially useful.

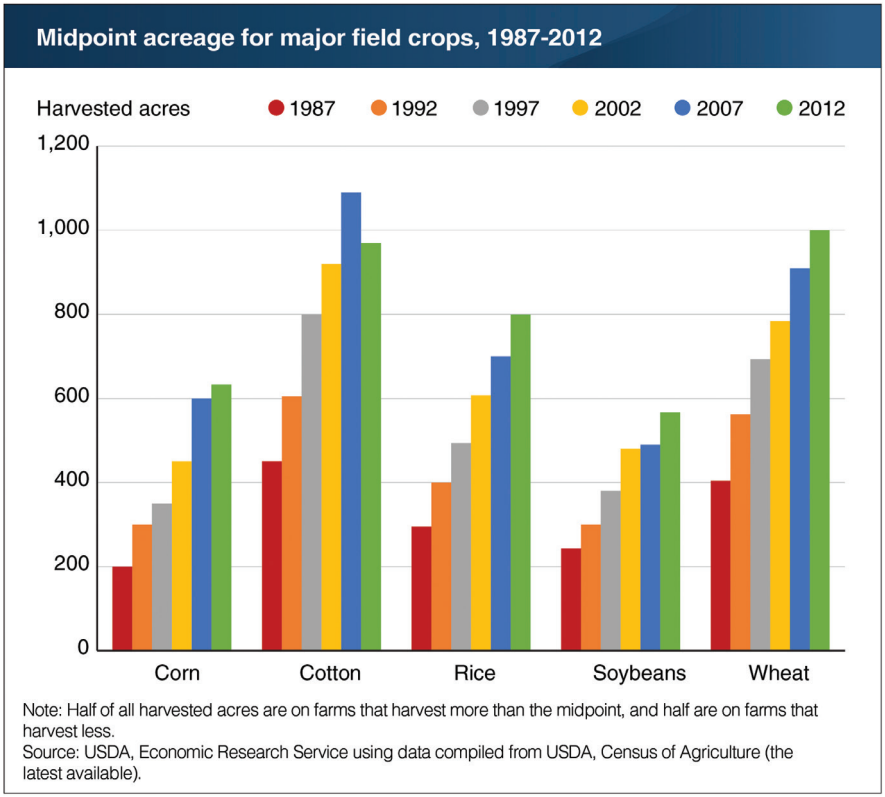
In this chapter, we'll cover a handful of data visualization redesigns. The graphs I choose to redesign here are not all especially *bad* graphs. Some are simply chosen because I believe there are more effective ways to plot the data. My goal is not to criticize these chart creators or their efforts but to demonstrate how the lessons we have learned can be applied to making data visualizations cleaner, clearer, and more effective.

The changes made here are by no means the only ways to modify these graphs, but each redesign follows the guidelines discussed throughout this book. In general, there is no "right" or "wrong" approach, just different ways of making improvements. As you develop an eye for better data visualization design, you will develop your own aesthetic and preferences.

PAIRED BAR CHART: ACREAGE FOR MAJOR FIELD CROPS

Take a moment and examine this bar chart from the U.S. Department of Agriculture that shows the number of harvested acres for five major crops in the United States for six different years. What do you see first?

My guess is you saw what I first saw: The acreage for all five crops increased over time. Your second observation, which quickly follows the first, is that cotton acreage (the second



Basic bar chart from the U.S. Department of Agriculture.

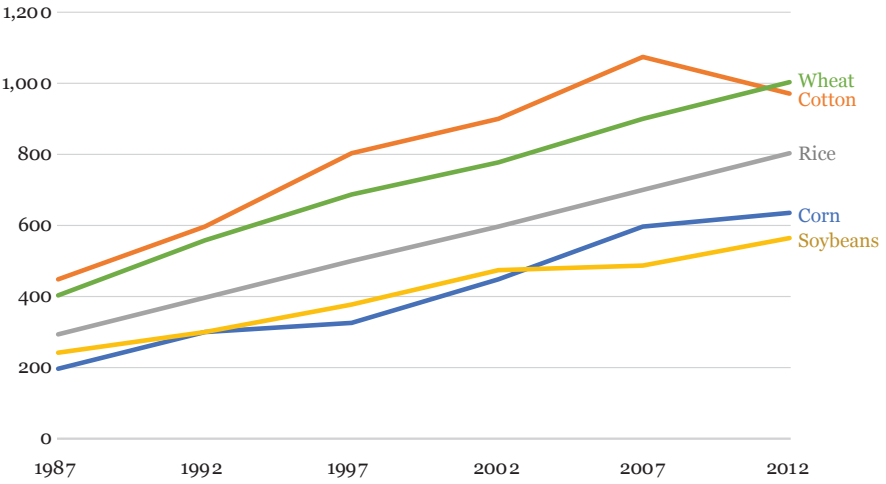
group) fell in the last year. Unlike the other groups, the last bar for cotton (the green bar) is shorter than the bar for the preceding year. But it doesn't jump out at you because there is so much ink and color in the graph.

If the goal with this chart is to show relative trends in acreage among five crops, a bar chart is a poor choice. The paired bar chart is good at showing exact values, but the relative trends are not clear or immediately evident.

We could redesign this as a simple line chart.

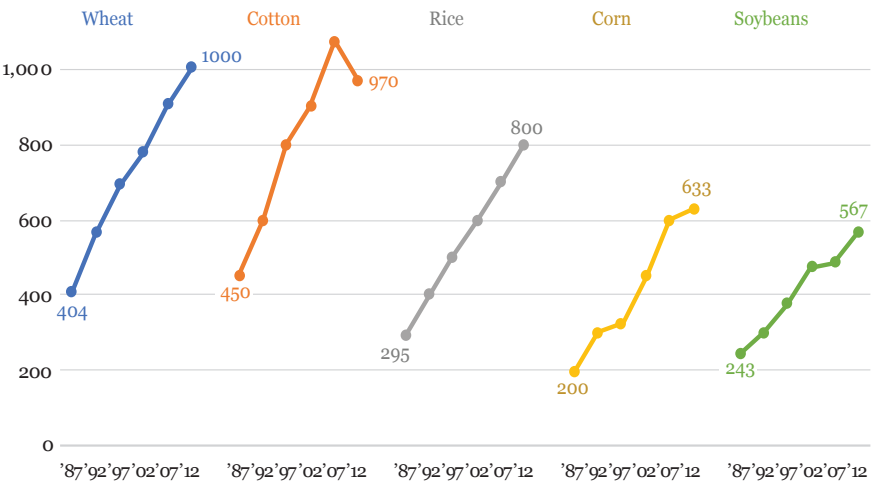
Here, the drop in acreage for cotton is very clear, as are the relative sizes of the five crops. In the bar chart, I couldn't see immediately that rice acreage sits right in the middle of the five crops, but here I can see that right away. I didn't use a legend here, as might be the default approach, but instead added the labels at the end of each line, using color to link them with the lines.

Midpoint acreage for major field crops, 1987-2012
(Midpoint acreages more than doubled for all five major field crops)



Source: U.S. Department of Agriculture

Midpoint acreage for major field crops, 1987-2012
(Midpoint acreages more than doubled for all five major field crops)



Source: U.S. Department of Agriculture

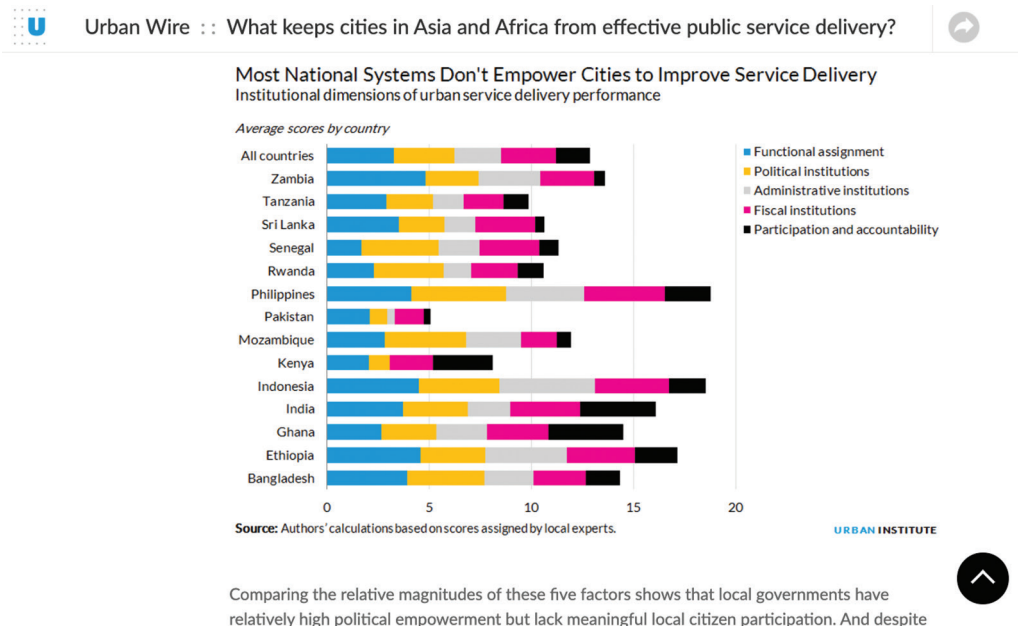
Two ways to redesign the USDA bar chart: A line chart or a cycle chart.

Another approach is a cycle chart. Instead of putting the lines together, this cycle chart is essentially a small multiples line chart where each crop gets its own panel. The advantage is that there's more space for the graph and it's perhaps a little more engaging because it's different. The disadvantage is that relative patterns are slightly less clear than in the line chart.

STACKED BAR CHART: SERVICE DELIVERY

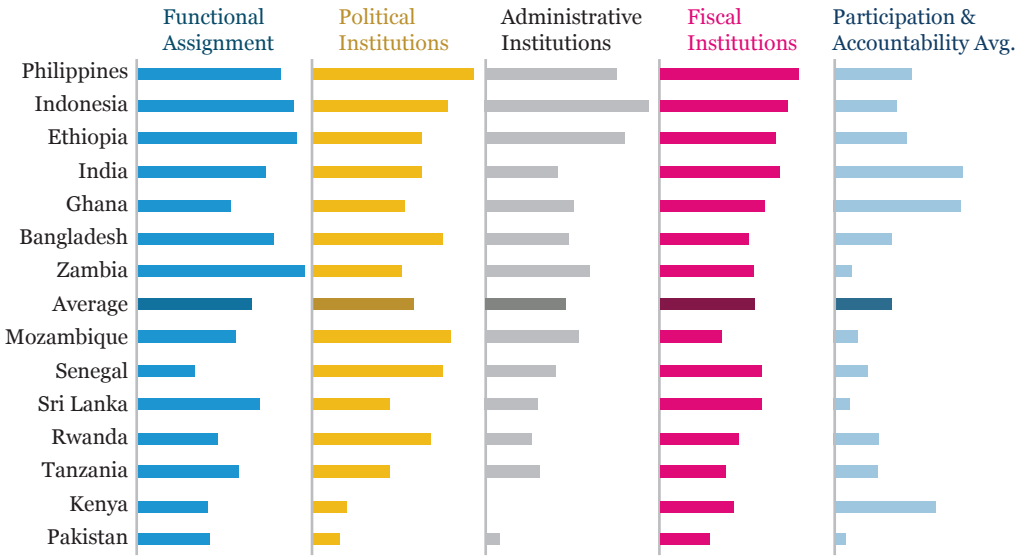
Let's go back to page 14 in Chapter 1 and consider the perceptual rankings diagram. At the very top are graphs positioned along common scales—the bar chart or line chart with a single horizontal axis, for example. One step below are those graphs that are not positioned along common scales graphs. It is slightly harder to accurately assess the values in these.

This graph contains data from both sections of the ranking diagram. We can clearly discern the differences between the values of the blue series (Functional assignment) because they all sit on the same vertical baseline. We are not as well equipped, however, to similarly



We can barely see that the value for Political Institutions is larger for *All countries* than *Zambia*.
Source: Roth and Malik, 2016

Most national systems don't empower cities to improve service delivery
(Institutional dimensions of urban service delivery performance, Average scores by country)



Source: Roth and Malik, 2016

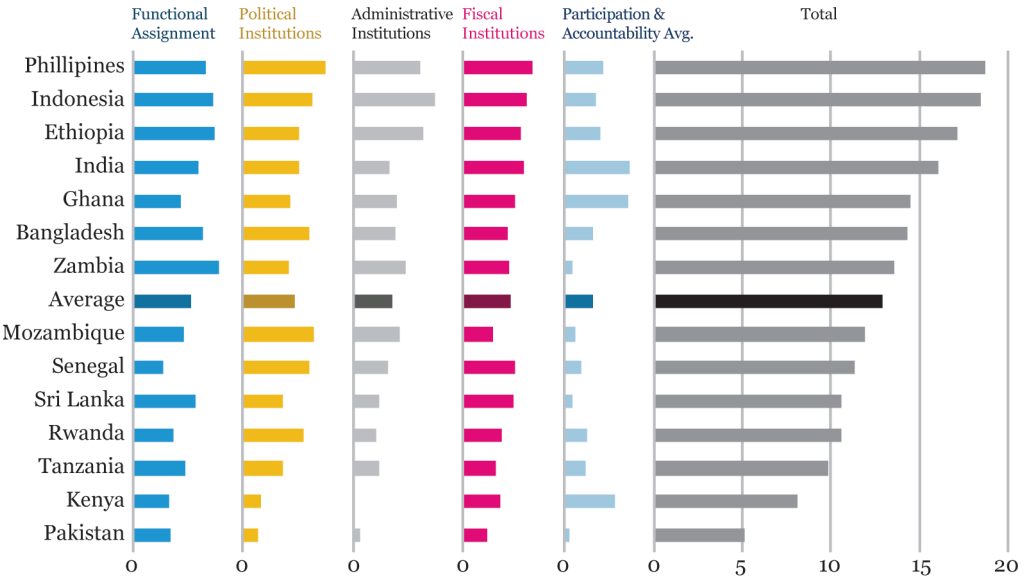
One way to redesign the stacked bar chart is to break them up and use a small multiples approach.

assess the values for the other series, because they don't share the same baseline. You can test this yourself: Is the value for Political Institutions (the yellow series) larger for "All Countries" or Zambia (the first two series)?

Instead of packing all of the data onto a single chart, we can break it into five separate charts. In this case, each series is given its own vertical baseline, so it's easier to make comparisons across countries within each series. The important point with making a graph like this is that the horizontal space for each series is the same. If we shrank the space for "Fiscal Institutions," for example, it might look like those values are larger than others.

This approach, however, doesn't tell you much about the *overall* values between countries. There's nothing wrong with adding a series for the overall *Total*, again, as long as we use the same horizontal spacing. In other words, the space between each gridline is the same. This approach works even better in cases where the values sum to the same total or to 100 percent, because the total length will be the same for all of the bars and thus a *Total* segment is unnecessary.

Most national systems don't empower cities to improve service delivery
(Institutional dimensions of urban service delivery performance, Average scores by country)



Source: Roth and Malik, 2016

When breaking up stacked bar charts, it is sometimes important to include the totals.

LINE CHART: THE SOCIAL SECURITY TRUSTEES

Each year, the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds report on the current and projected status of the U.S. Social Security program. The Trustees are responsible for estimating the current and future financial picture of the program to communicate to the public and policymakers the challenges the program faces. The Social Security Technical Panel is an independent expert panel responsible for reviewing the work of the Trustees, including the methodological details, economic and demographic assumptions, and the Trustees communication efforts.

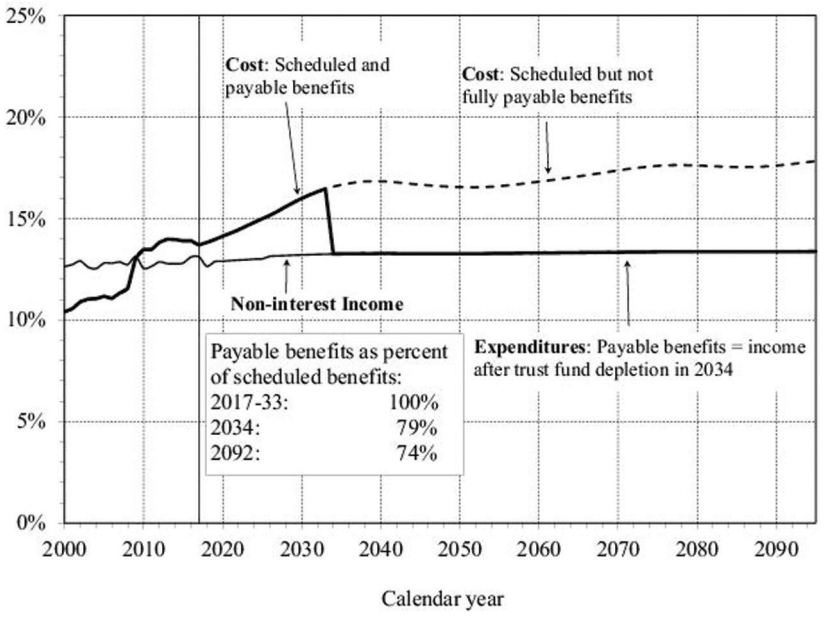
The 2019 Technical Panel placed an emphasis on this latter category: "The Panel believes that trust in public institutions is enhanced by greater understanding . . . In this context, we believe it is paramount for the Trustees to communicate clearly and effectively with the general public about its finances." The panel emphasized clear, plain language, a focus on the core message, and better data visualizations in the Trustee's work.

Let's look, then, at two of their data visualizations.

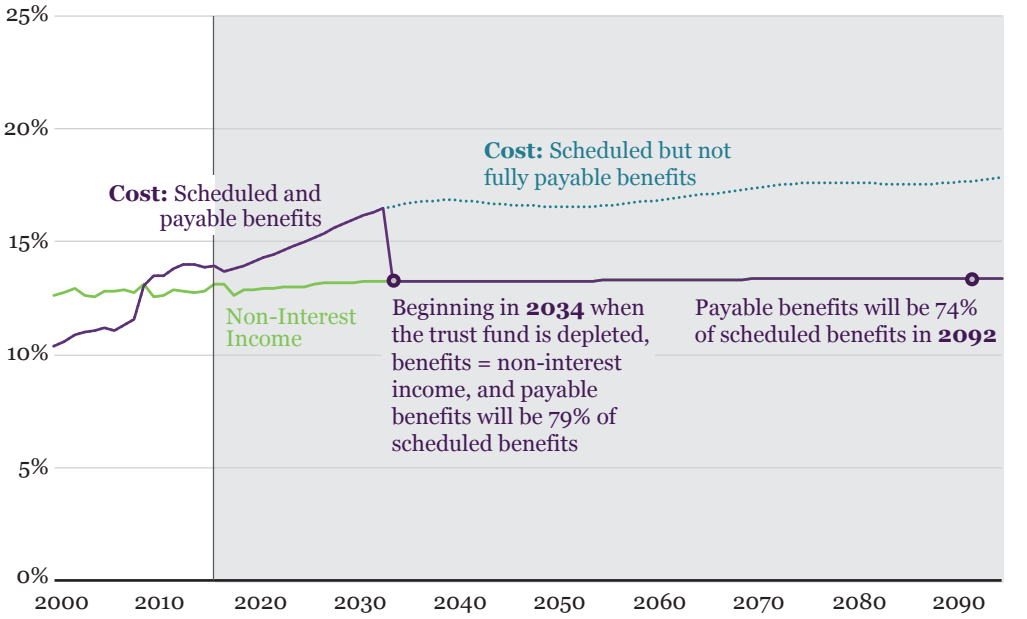
A CLEANUP

The first example is a relatively simple clean-up rather than a wholesale redesign. This line chart—which has appeared in virtually every Trustees Report—shows the time series of the basic finances of the Social Security system. System income (taxes paid into the system) are set next to system costs (benefits paid to beneficiaries) for a short historical period (here from 2000 to 2018) and out in the longer projection period (here from 2018 through 2092). Two sets of costs are shown: one that shows how many benefits are *scheduled* to be paid (the dashed line) and one that shows how many benefits can *actually* be paid (the solid bold line).

The existing graph has annotation and labels to help the reader better understand the content and the concepts. A small table near the bottom of the graph lists benefit shares in



The Social Security Administration (2019) has published this graph showing the basic finances of the Social Security system for many years.



Source: Social Security Administration, 2019

Some basic cleanup and annotation improves the clarity of the Social Security finances chart.

specific years. But there is also a lot of ink used on extraneous details: horizontal and vertical gridlines and tick marks for every percentage point and year.

Let's take a simple approach to redesigning this graph by removing some of these extraneous details and markers. Here, I've removed the vertical gridlines and all of the tick marks. I deleted the small table and instead directly labeled the years those numbers referenced. I used some slight color here—which is consistent with black-and-white printing—and added a gray box to the projection period (after 2018) to draw attention to the imbalance.

A BETTER DOT PLOT

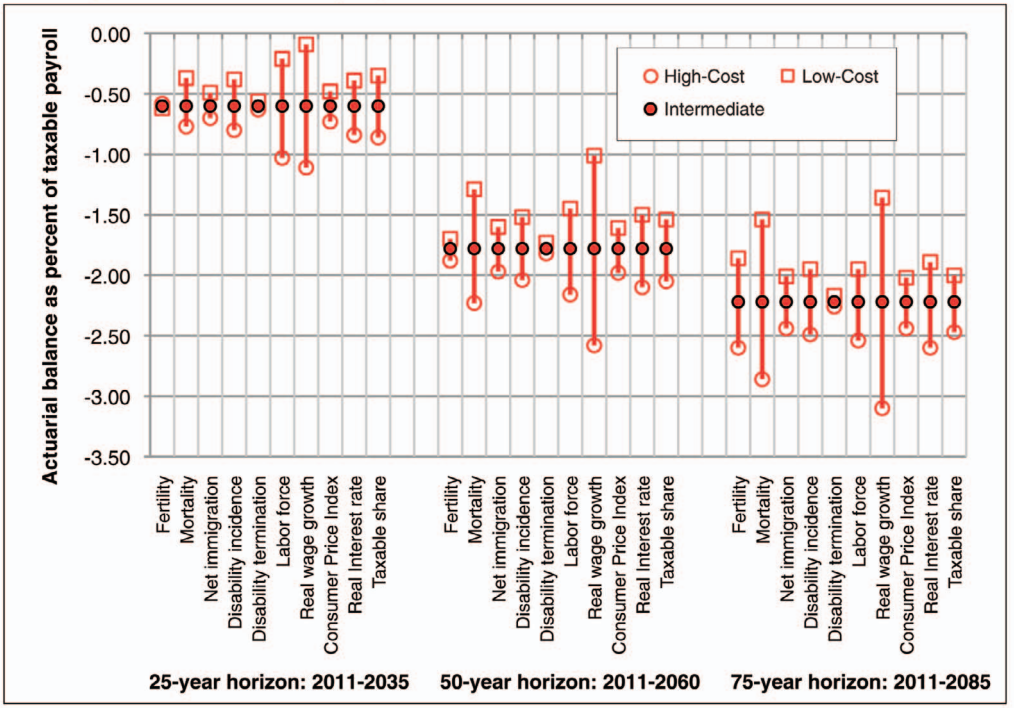
The next graph appeared in the 2011 Technical Panel Report. It shows the sensitivity of different assumptions of the Social Security model. Most of the six Technical Panel Reports published since 1999 include this information as a series of tables. But in 2011, the Panel

presented these estimates in what amounts to a dot plot or a simplified box-and-whisker plot. Here, instead of two dots with a connecting line, the chart has an "intermediate" estimate in the middle and a "low-cost" and "high-cost" options sit on either side.

Notice the different shapes, vertical and horizontal gridlines, and rotated axis labels. Using our basic guidelines from Chapter 2—showing the data, reducing the clutter, integrating the text and the graph, use more graphs, and start with gray—we can improve this visualization to make it clearer and easier to read.

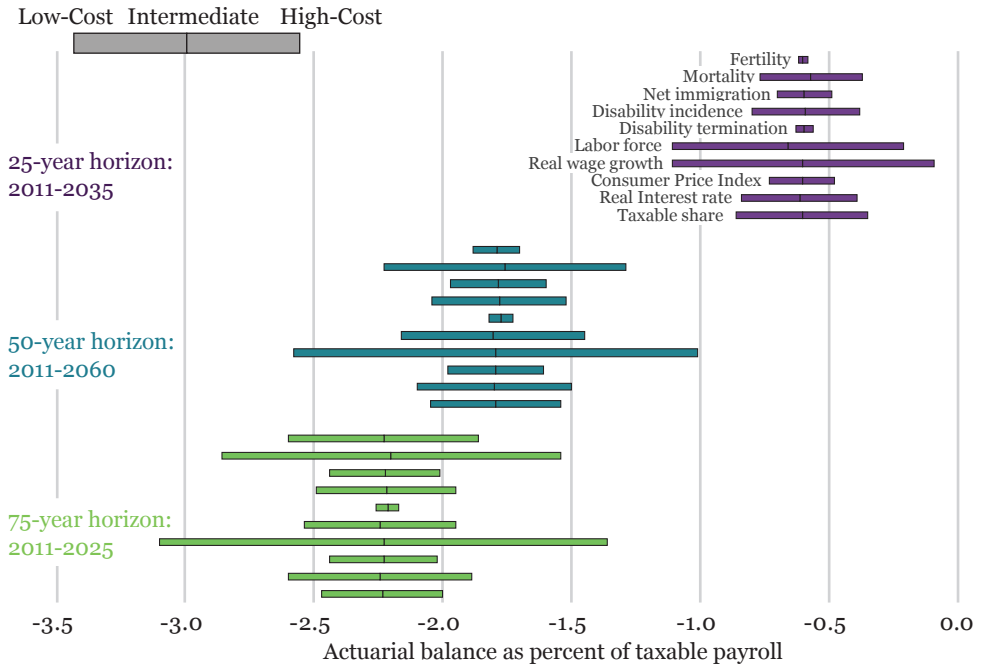
A simple start is to rotate the entire visualization. Now we don't need to turn our heads to read the labels, and we can place what was formerly along the vertical axis along the horizontal axis. Instead of dots—which works just fine—I converted them to boxes, which

Figure 4. Sensitivity of Summarized Actuarial Balance to Range of Assumptions: 25-, 50-, and 75-Year Horizons (as a Percent of Taxable Payroll)[4]



Source: 2011 Trustees Report, Appendix D; additional estimates provided by Office of the Chief Actuary, Social Security Administration.

Gridlines, rotated text, and general clutter make this graph hard to read.
Source: 2011 Technical Panel Report on Assumptions and Methods



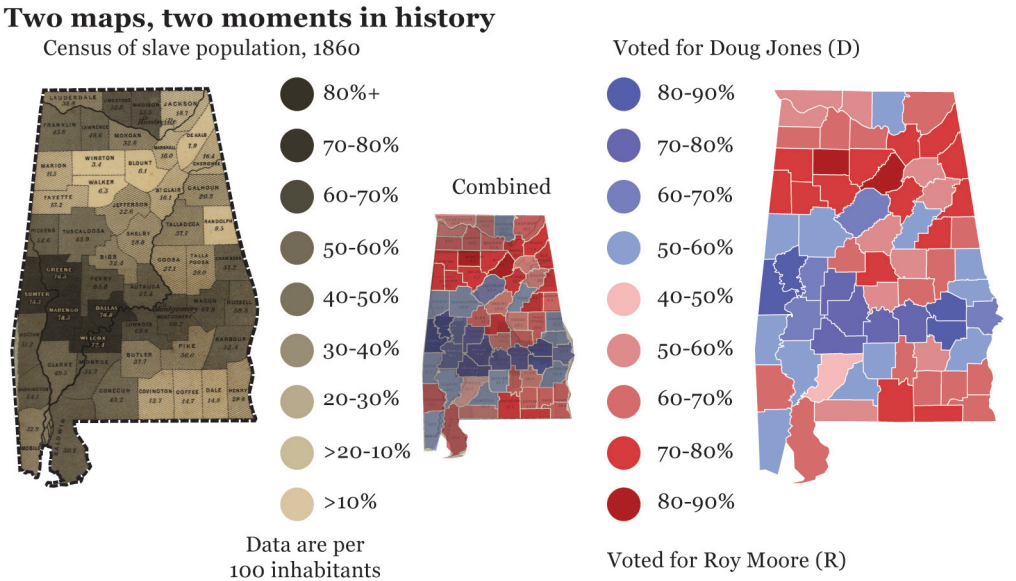
Source: Social Security Administration, 2011

Rearranging the plot and removing some of the clutter makes the graph easier to read.

reduces the extra lines and dots in the graph. By using a box, one end can encode the low-cost value; one end, the high-cost value; and a middle marker, the intermediate value. We can also place the labels for each metric right next to the first set of boxes though we could also repeat them if we thought it was necessary.

CHOROPLETH MAP: ALABAMA SLAVERY AND SENATE ELECTIONS

In late 2017, there was a runoff election for a U.S. Senate seat in Alabama. In a tense and competitive election, journalist Sarah Slobin (then at Quartz) wrote a story about the relationship between voting behavior and the distribution of enslaved people in 1860. Slobin wrote that, "[W]hile correlation is not causation, there is a startling visual parallel when you zoom in to Alabama . . . and compare it to how Alabama just voted this week."



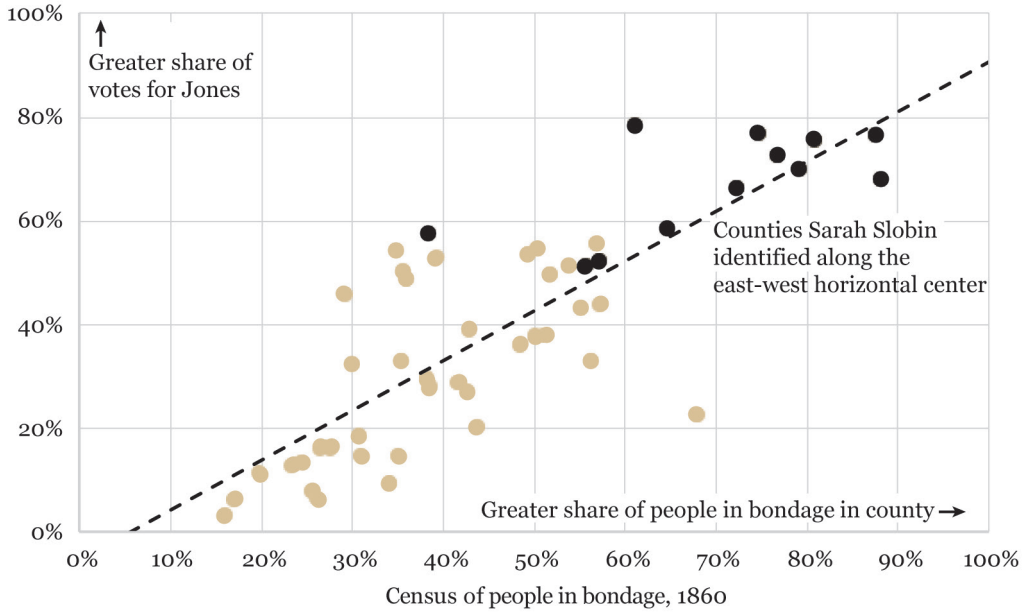
Using an 1860 map from the US Bureau of the Census and voting results in the 2017 Alabama Senate election, we can see similar bands of darker colors running through the middle of the state.

Source: Author's rendering based on original chart from Quartz. Map from the US Bureau of the Census; voting data provided courtesy of Sarah Slobin

Using voting data from that election, there is a clear horizontal band of dark blue (representing more Democratic votes) in the map on the right. The map on the left, from an 1860 map published by the U.S. Bureau of the Census, shows a streak of darker colors representing counties with a larger proportion of people held in bondage. Slobin writes: "If you focus on the 'black belt' moving horizontally across both maps, you can see that in areas with a history of slavery, the vote went to [the Democratic candidate] Jones."

Though visually striking, this pairing forces the reader to jump between the two maps to see the similarity in the bands. Can we take a different approach?

The share of the vote going to the Democrat in Alabama's sixty-seven counties ranged from 16.1 percent to 88.1 percent. These roughly (though not perfectly) overlap with fifty-one counties I could identify in the Census Bureau map, which range from 3.1 percent to 78.3 percent. (There are some data issues we will ignore here as counties have changed, merged, or broken up between 1860 and 2017).



Source: Voting data provided courtesy of Sarah Slobin

An alternative—or addition—to two maps is to use a scatterplot.

Plotting the two variables in a scatterplot lets us more easily see the positive relationship with the (darker) circles in the top-right part of the graph marking those twelve counties along the east-west corridor.

Slobin's original maps are visually striking. For a news story, they may well be the best way to show the data. It's easy to see the basic band pattern running through each map. By comparison, the scatterplot may take some additional explanation for a casual reader who may be less familiar with this graph type. We could publish *both* graphs to pair the visually-striking maps—to draw readers in—with the more technical scatterplot for those who want to see the detailed comparison. If I were publishing this in a peer-reviewed academic journal, I would lean toward the scatterplot because it clearly shows the association between the two series.

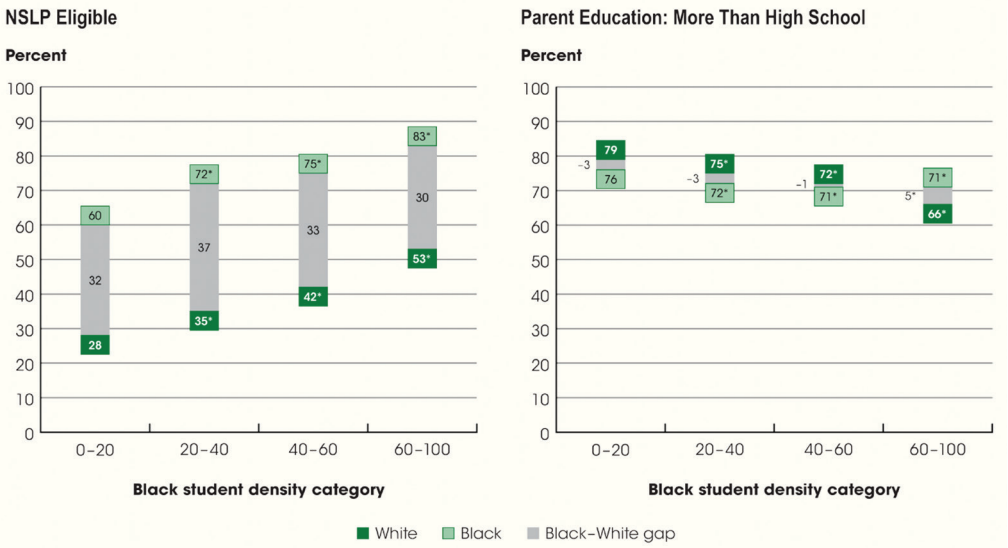
DOT PLOT: THE NATIONAL SCHOOL LUNCH PROGRAM

In reviewing a report from the National Assessment of Educational Progress (NAEP) on the achievement gaps between Black and white students, I came across this chart, which shows

differences in school achievement scores for Black and white students, arranged by scores of Black students.

Let's focus on the bars on the far-left side of the graph. There are three numbers here: 28 percent, 32 percent, and 60 percent. The numbers in the green boxes show test scores for white (60 percent) and Black (28 percent) students, and the middle number shows the gap between the two groups (32 percent). But the green boxes make it appear as if the 28 percent represents a range of numbers, from, say, 22 percent to 28 percent. By using rectangles instead of points or markers, it resembles a stacked chart rather than the dot plot, which was likely the intention.

Figure 8. Percentage of Black and White students who were National School Lunch Program (NSLP) eligible and percentage who had a parent with more than a high school education, by Black student density category: 2011

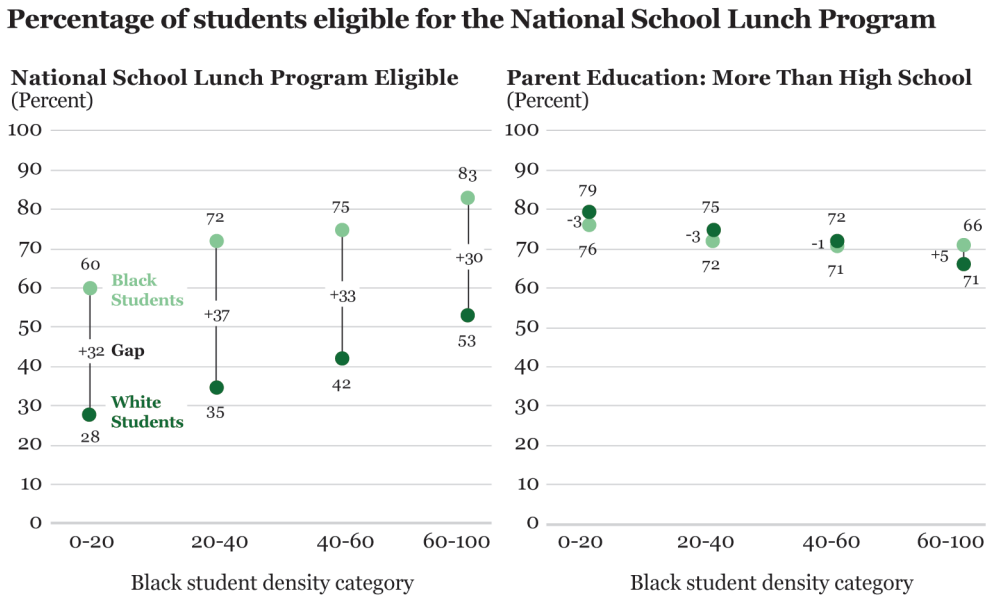


* Significantly different ($p < .05$) from the 0 percent to 20 percent density category.

NOTE: The measures displayed in this figure are percentages of students within each Black student density category.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Mathematics Grade 8 Assessment.

This graph from the National Center for Education Statistics shows the percentage of students eligible for the National School Lunch Program.



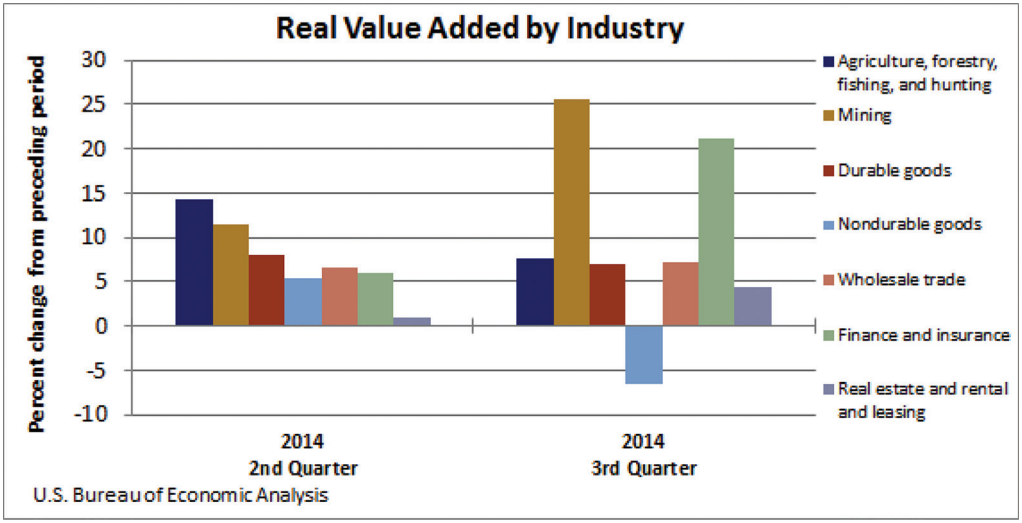
Changing shapes and removing some of the clutter makes the graph from the National Center for Education Statistics easier to read.

As an alternative, let's make it a true (vertical) dot plot. We can replace the green boxes with green circles and connect them with a gray, vertical line. Now we perceive the green circles as specific points rather than ranges or a stacked set of values.

You might also notice that I deleted the legend and labeled the three series directly on the left chart. I didn't repeat the labeling in the chart on the right for two reasons: First, because the gaps are smaller, there is less space for the labels. And second, the reader doesn't need to be reminded of the definition of each dot and line at every single occurrence on the page.

DOT PLOT: GDP GROWTH IN THE UNITED STATES

Every quarter, the U.S. Bureau of Economic Analysis (BEA)—the federal agency responsible for producing some of the most important measures of the U.S. economy—releases their



This bar chart from the Bureau of Economic Analysis's quarterly report does not match what's written in the text.

report about changes in gross domestic product (GDP). And each quarter, they publish a press release on changes over time and in specific industries.

The graph above is from the third quarter of 2014 press release. It shows the "Real Value Added"—a measure of each industry's contribution to GDP—for major industries in the country. Given what you've learned so far, there are a variety of things you might change to make this graph more effective. You might directly label the bars, rotate the vertical axis legend and place it near the title, and lighten some of the gridlines.

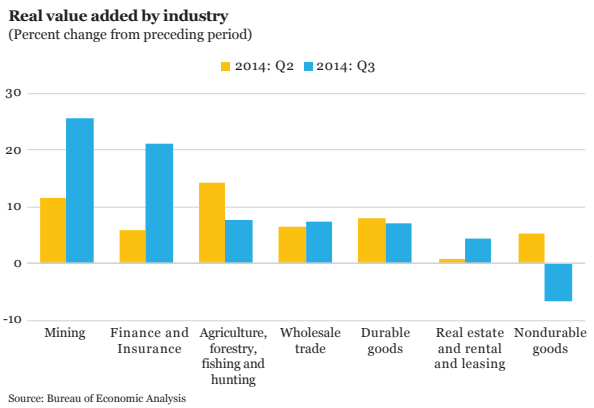
More importantly, let's take a look at what this graph is *supposed* to show. Here are the six bullet points that surround the Real Value Added (RVA) by Industry graph in the BEA document:

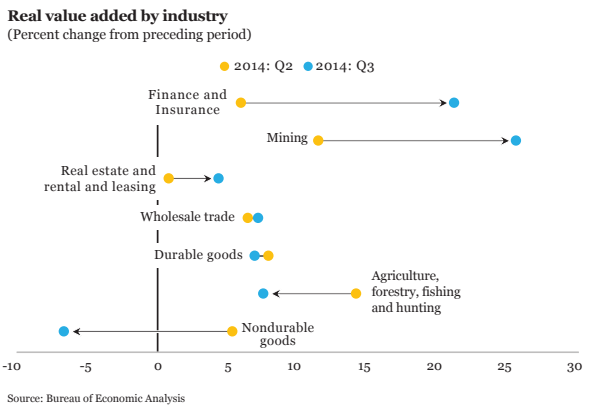
- Finance and insurance real value added—a measure of an industry's contribution to GDP—increased 21.2 percent in the third quarter, after increasing 6.0 percent in the second quarter.

- Mining increased 25.6 percent, after increasing 11.5 percent. This was the largest increase since the fourth quarter of 2008.
- Real estate and rental and leasing increased 4.4 percent, after increasing 0.9 percent.
- Real value added for manufacturing increased 0.5 percent, after increasing 6.8 percent. Durable-goods increased 7.0 percent following an increase of 8.0 percent, while nondurable-goods decreased 6.6 percent, after increasing 5.4 percent.
- Agriculture, forestry, fishing, and hunting increased 7.6 percent after increasing 14.2 percent.
- Wholesale trade continued to show strong growth, increasing 7.3 percent, after increasing 6.5 percent.

What do you notice about how these points are arranged? Each one details how RVA changed between the first and last period for each industry. The structure of the graph, however, is formatted to compare *across* industries *within* each period.

A better approach for the graph would match the text and show the inverse: the change within each industry across periods. A paired bar chart and dot plot are two effective ways to do this. In the paired bar chart, I've sorted the data according to the most recent period (2014:Q3) to subtly guide the reader to the best-performing industries. In the dot plot, the data are sorted based on the *change* between the two periods—the largest positive changes are at the top of the graph and the largest declines at the bottom.





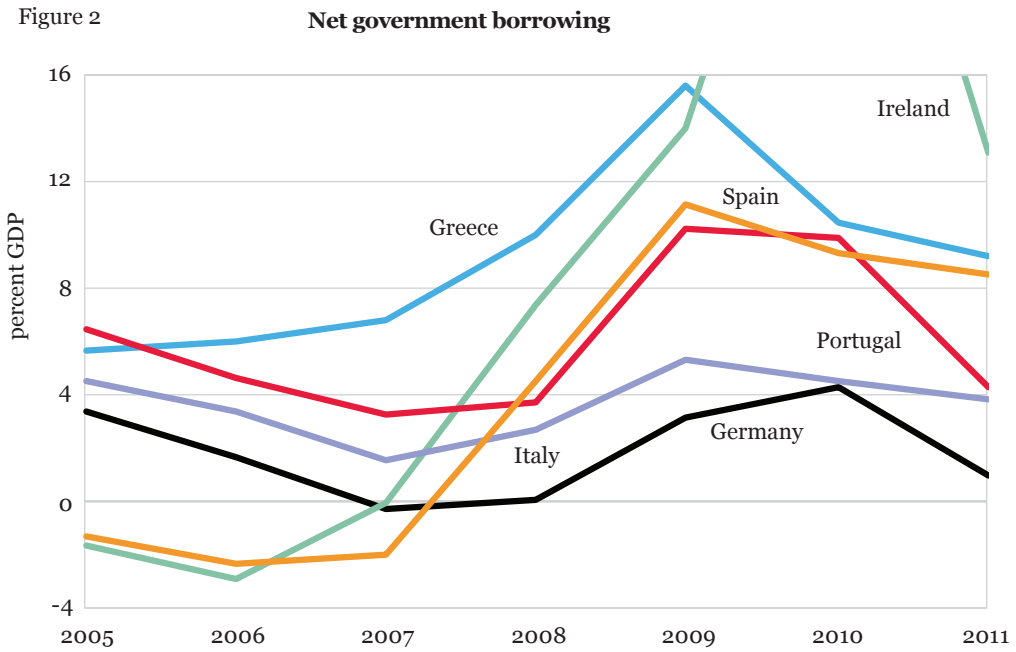
Two options to redesign the original BEA graph to match the organization of the press release.

In either case, these two graphs better illustrate the takeaways from the text. They are sorted by industry so that now, for example, you can more easily see that, "Mining increased 25.6 percent, after increasing 11.5 percent." Your data visualizations should not be intended to break up long sections of text or to provide a "visual break." They are there to support your argument. Integrate them with your writing for a seamless reading experience.

LINE CHART: NET GOVERNMENT BORROWING

As I mentioned in Chapter 5, I'm a big fan of line charts. They clearly show changes over time, and everyone knows how to read them. Consider, however, this line chart in a 2012 Economic Policy Paper from the Federal Reserve Bank of Minneapolis.

Figure 2



This line chart, originally published by the Federal Reserve Bank of Minneapolis, simply cuts off the data for Ireland.

Source: Author's rendering based on original chart from Arellano, Conesa, and Kehoe (2012).

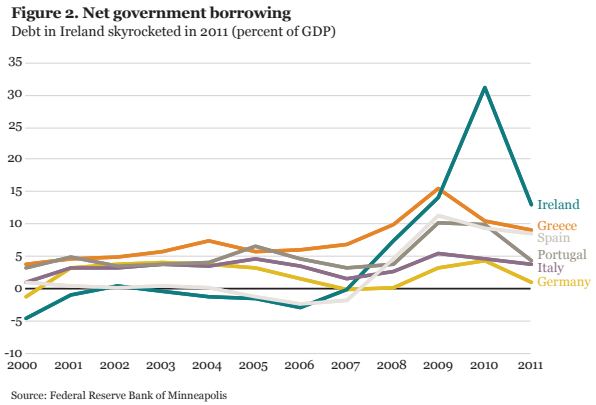
Notice anything strange about this graph? Anything aside from the title split into three parts ("Figure 2" in the top-left; "Net government borrowing" centered over the graph; and "percent GDP" along the vertical axis)? Anything besides the equal-weighted gridlines even though zero is not at the bottom? Or the mix of pastel and bright colors?

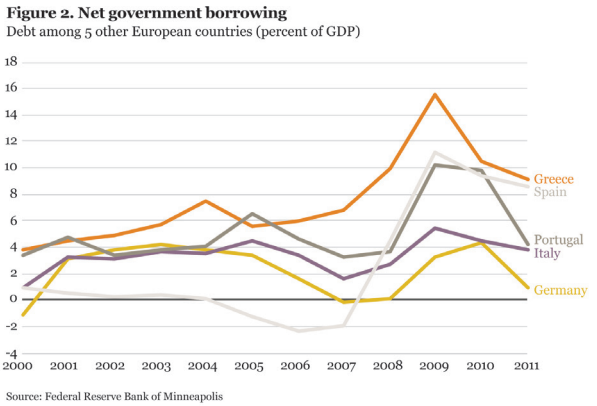
How about how the line for Ireland shoots off the top of the graph? You must have a very good reason not to show all of the data on the graph, and putting the value in the footnote does not count!

The chart creator here faced a problem: Ireland had a debt spike in 2011, far more than the other countries. To show all the data in one chart would scrunch up and lose the detail among the other countries.

But there is a better solution: Use two graphs. One that includes Ireland with a vertical axis that ranges from −10 percent to 35 percent, and another with a vertical axis from −4 percent to 18 percent that shows the detail among the other countries. I could leave these as equal-sized charts, or even make the second one smaller in a sort of zoom-out/zoom-in comparison. In either case, I use the subtitle to explain that one graph includes Ireland and the other does not.

Do not be afraid to use more than one graph, if that's what it takes to clearly communicate your argument. We are now a digital-first society—using more space only requires more computer memory, not more paper.





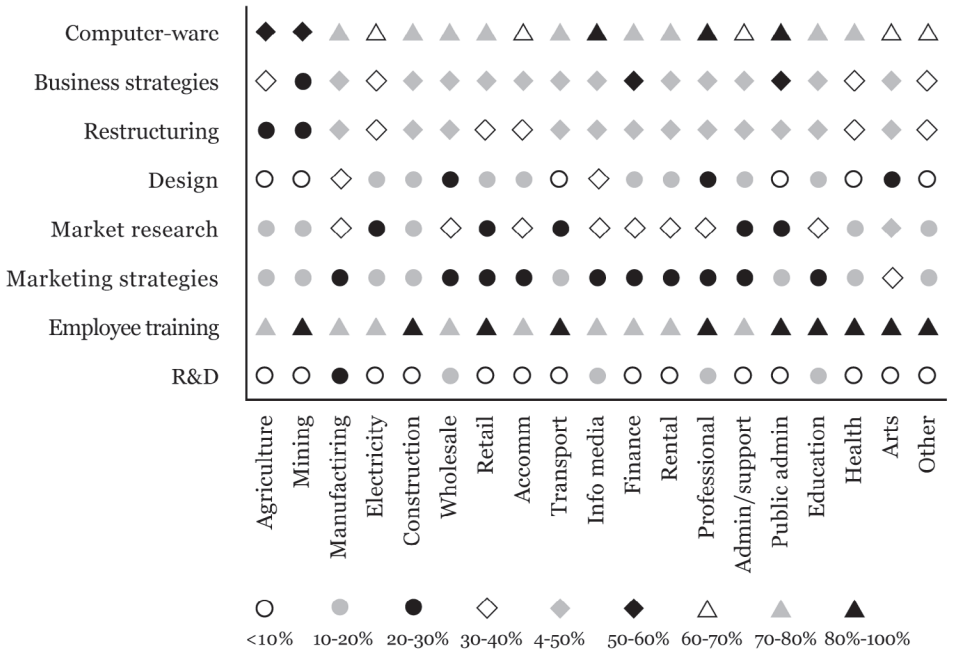
Instead of packing all of the information on a single chart, try breaking them up and create a "zoom in" view.

TABLE: FIRM ENGAGEMENT

As we saw in Chapter 11, there are many ways to make our tables more visual. We can add color, icons, bars, or other elements to highlight the important values for our reader instead of asking them to sift through all the data values.

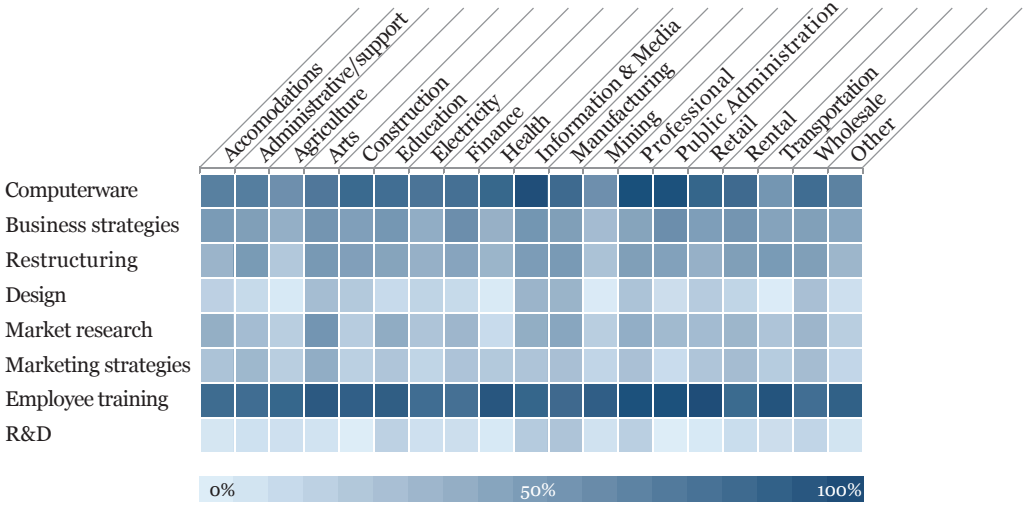
This table uses different shapes and shades of gray to show the share of firms that engage in different business activities like design and market research. As the reader, we must understand which shapes correspond to which percentages and then figure out the different

Figure 1: Proportion of firm-years engaging in each intangible acitivty, by industry



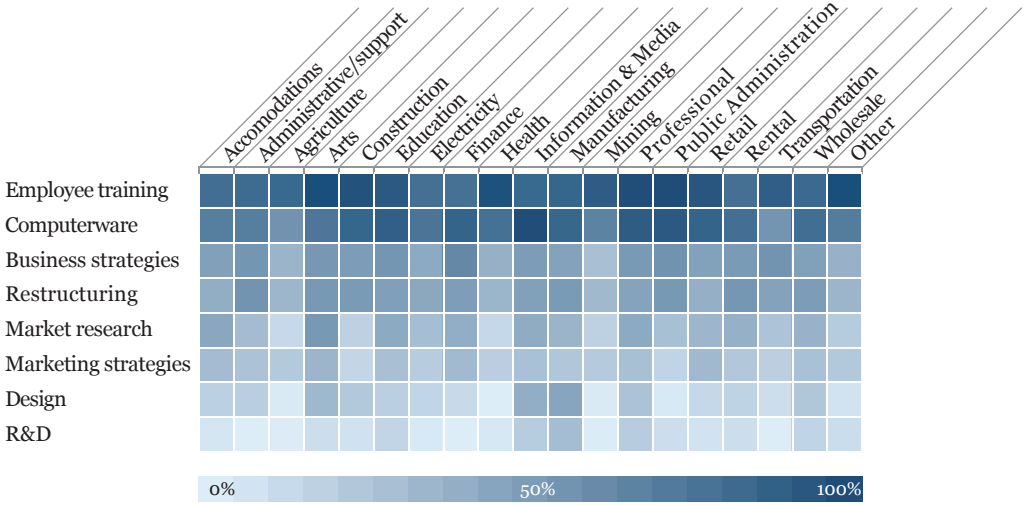
Source: Author's rendering of original chart by Chappell and Jaffe, 2018
Note: Data based on a visual inspection of the original graphic.

Author's rendering of an original chart by Chappell and Jaffe (2018), which could be improved by changing how the values are displayed.



Source: Chappell and Jaffe, 2018
Note: Data based on a visual inspection of the original graphic.

A heatmap is one alternative to the Chappel and Jaffe (2018) chart.



Source: Chappell and Jaffe, 2018
Note: Data based on a visual inspection of the original graphic.

This heatmap alternative to the Chappel and Jaffe (2018) chart sorts the data.

shading styles. Of course, triangles don't necessarily mean "more" of something than circles, so the rank-ordering of the values is hard to interpret.

Instead, what if we use a monochromatic color ramp moving from a light blue for the lower percentages to darker blues for the higher values? In this heatmap approach, it's much easier to see that there is a lot of time spent on *Employee training*, the dark blue row towards the bottom of the table.

We can take this a small step forward and sort the data, which will naturally focus the reader's visual attention. The longer labels either require us to use rotated labels, as I've done here, or perhaps rotate the entire graph and change the spacing or cell size so all of the text can fit.

CONCLUSION

With more graphs in your data visualization toolbox to choose from, and having seen more graphs and best practices, I'm confident that you're ready to improve upon your own graphs. Finding and redesigning even the simplest graph—I find that mining the academic peer-reviewed literature a good place to start—can help you refine your skills and develop your own data visualization aesthetic. Like any other skill, practice makes better.

Two important caveats. First, if you critique a graph publicly, keep in mind that someone made that graph and that even your well-intentioned efforts to redesign it may not be appreciated. The chart creator may have had time pressures, software limitations, or organizational demands of which you are not aware. Reaching out to the person who created the original graph may be worth your effort. Second, try to identify the central goal of the chart and the possible challenges of the data series. This will help lead you to the best chart type for the task at hand.