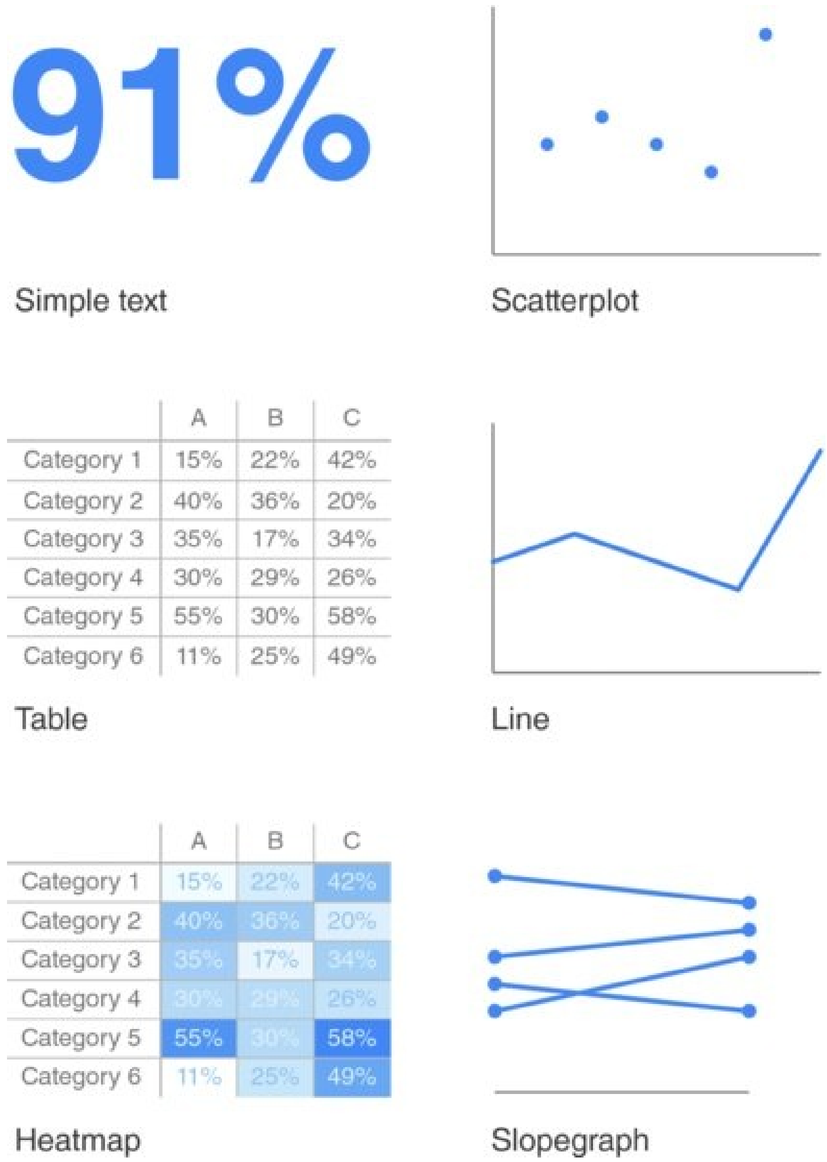
chapter 2
choosing an effective visual

There are many different graphs and other types of visual displays of information, but a handful will work for the majority of your needs. When I look back over the 150+ visuals that I created for workshops and consulting projects in the past year, there were only a dozen different types of visuals that I used (Figure 2.1). These are the visuals we'll focus on in this chapter.



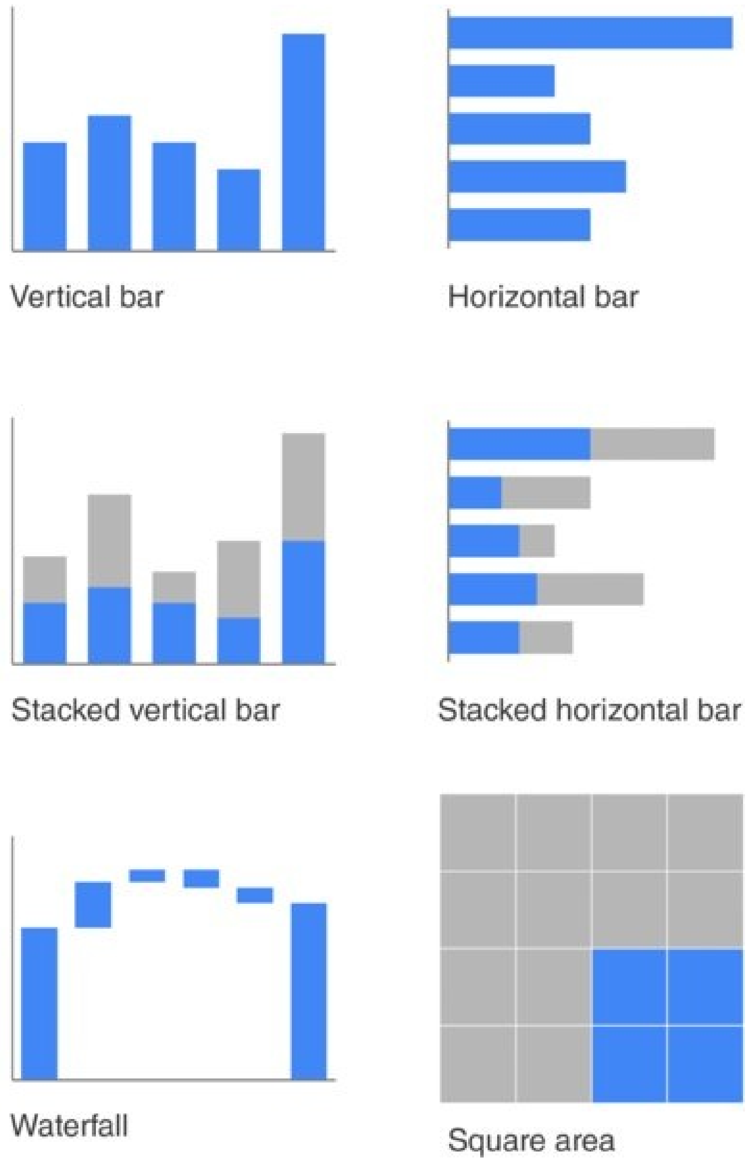


Figure 2.1 The visuals I use most

Simple text

When you have just a number or two to share, simple text can be a great way to communicate. Think about solely using the number—making it as prominent as possible—and a few supporting words to clearly make your point. Beyond potentially being misleading, putting one or only a couple of numbers in a table or graph simply causes the numbers to lose some of their oomph. When you have a number or two that you want to communicate, think about using the numbers themselves.

To illustrate this concept, let’s consider the following example. A graph similar to Figure 2.2 accompanied an April 2014 Pew Research Center report on stay-at-home moms.

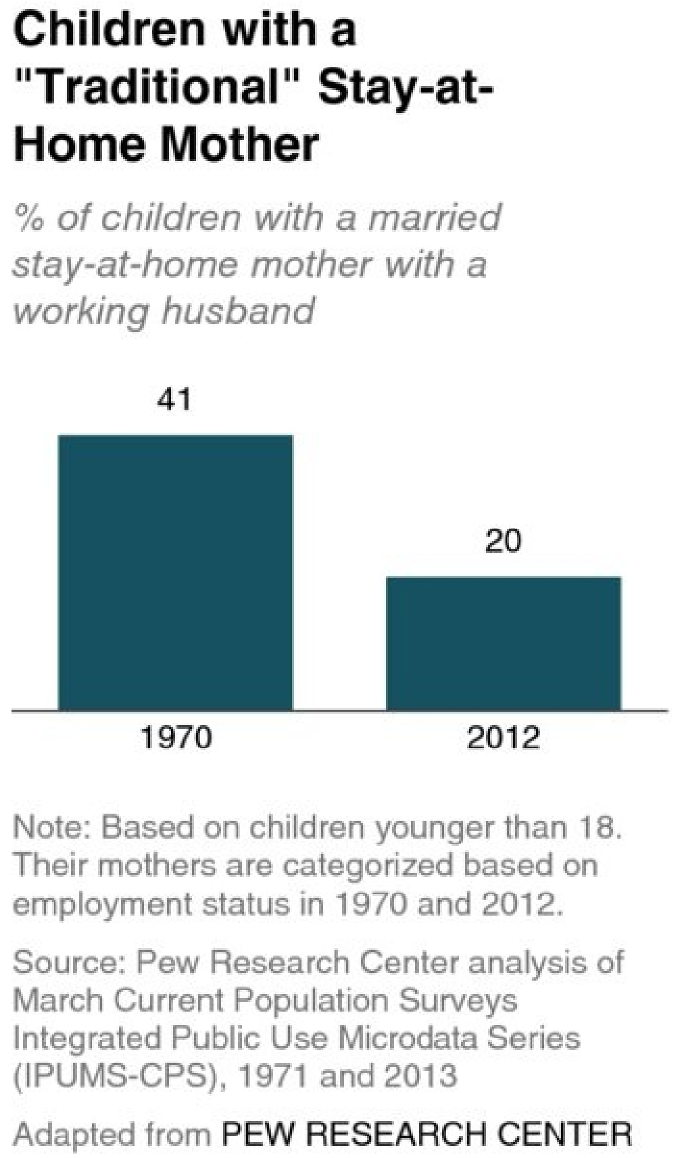


Figure 2.2 Stay-at-home moms original graph

The fact that you have some numbers does not mean that you need a graph! In Figure 2.2, quite a lot of text and space are used for a grand total of *two* numbers. The graph doesn't do much to aid in the interpretation of the numbers (and with the positioning of the data labels outside of the bars, it can even skew your perception of relative height such that 20 is less than half of 41 doesn't really come across visually).

In this case, a simple sentence would suffice: *20% of children had a traditional stay-at-home mom in 2012, compared to 41% in 1970.*

Alternatively, in a presentation or report, your visual could look something like Figure 2.3.

Figure 2.3 Stay-at-home moms simple text makeover

As a side note, one consideration in this specific example might be whether you want to show an entirely different metric. For example, you could reframe in terms of the percent change: "The number of children having a traditional stay-at-home mom decreased more than 50% between 1970 and 2012." I advise caution, however, any time you reduce from multiple numbers down to a single one—think about what context may be lost in doing so. In this case, I find that the actual magnitude of the numbers (20% and 41%) is helpful in interpreting and understanding the change.

When you have just a number or two that you want to communicate: *use the numbers directly.*

When you have more data that you want to show, generally a table or graph is the way to go. One thing to understand is that people interact differently with these two types of visuals. Let's discuss each in detail and look at some specific varieties and use cases.

Tables

Tables interact with our verbal system, which means that we *read* them. When I have a table in front of me, I typically have my index finger out: I'm reading across rows and down columns or I’m comparing values. Tables are great for just that—communicating to a mixed audience whose members will each look for their particular row of interest. If you need to communicate multiple different units of measure, this is also typically easier with a table than a graph.

Tables in live presentations

Using a table in a live presentation is rarely a good idea. As your audience reads it, you lose their ears and attention to make your point verbally. When you find yourself using a table in a presentation or report, ask yourself: what is the point you are trying to make? Odds are that there will be a better way to pull out and visualize the piece or pieces of interest. In the event that you feel you’re losing too much by doing this, consider whether including the full table in the appendix and a link or reference to it will meet your audience’s needs.

One thing to keep in mind with a table is that you want the design to fade into the background, letting the data take center stage. Don’t let heavy borders or shading compete for attention. Instead, think of using light borders or simply white space to set apart elements of the table.

Take a look at the example tables in Figure 2.4. As you do, note how the data stands out more than the structural components of the table in the second and third iterations (light borders, minimal borders).

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

Light borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

Figure 2.4 Table borders

Borders should be used to improve the legibility of your table. Think about pushing them to the background by making them grey, or getting rid of them altogether. The data should be what stands out, not the borders.

Recommended reading

For more on table design, check out Stephen Few’s book, *Show Me the Numbers*. There is an entire chapter dedicated to the design of tables, with discussion on the structural components of tables and best practices in table design.

Next, let’s shift our focus to a special case of tables: the heatmap.

Heatmap

One approach for mixing the detail you can include in a table while also making use of visual cues is via a heatmap. A heatmap is a way to visualize data in tabular format, where in place of (or in addition to) the numbers, you leverage colored cells that convey the relative magnitude of the numbers.

Consider Figure 2.5, which shows some generic data in a table and also a heatmap.

Table

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

LOW-**HIGH**

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Figure 2.5 Two views of the same data

In the table in Figure 2.5, you are left to read the data. I find myself scanning across rows and down columns to get a sense of what I'm looking at, where numbers are higher or lower, and mentally stack rank the categories presented in the table.

To reduce this mental processing, we can use **color saturation** to provide visual cues, helping our eyes and brains more quickly target the potential points of interest. In the second iteration of the table on the right entitled "Heatmap," the higher saturation of blue, the higher the number. This makes the process of picking out the tails of the spectrum—the lowest number (11%) and highest number (58%)—an easier and faster process than it was in the original table where we didn't have any visual cues to help direct our attention.

Graphing applications (like Excel) typically have conditional formatting functionality built in that allows you to apply formatting like that shown in Figure 2.5 with ease. Be sure when you leverage this to always include a legend to help the reader interpret the data (in this case, the LOW-HIGH subtitle on the heatmap with color corresponding to the conditional formatting color serves this purpose).

Next, let's shift our discussion to the visuals we tend to think of first when it comes to communicating with data: graphs.

Graphs

While tables interact with our verbal system, graphs interact with our visual system, which is faster at processing information. This means that a well-designed graph will typically get the information across more quickly than a well-designed table. As I mentioned at the onset of this chapter, there are a plethora of graph types out there. The good news is that a handful of them will meet most of your everyday needs.

The types of graphs I frequently use fall into four categories: points, lines, bars, and area. We will examine these more closely and discuss the subtypes that I find myself using on a regular basis, with specific use cases and examples for each.

Chart or graph?

Some draw a distinction between charts and graphs. Typically, "chart" is the broader category, with "graphs" being one of the subtypes (other chart types include maps and diagrams). I don't tend to draw this distinction, since nearly all of the charts I deal with on a regular basis are graphs. Throughout this book, I use the words chart and graph interchangeably.

Points

Scatterplot

Scatterplots can be useful for showing the relationship between two things, because they allow you to encode data simultaneously on a horizontal *x*-axis and vertical *y*-axis to see whether and what relationship exists. They tend to be more frequently used in scientific fields (and perhaps, because of this, are sometimes viewed as complicated to understand by those less familiar with them). Though infrequent, there are use cases for scatterplots in the business world as well.

For example, let's say that we manage a bus fleet and want to understand the relationship between miles driven and cost per mile. The scatterplot may look something like Figure 2.6.

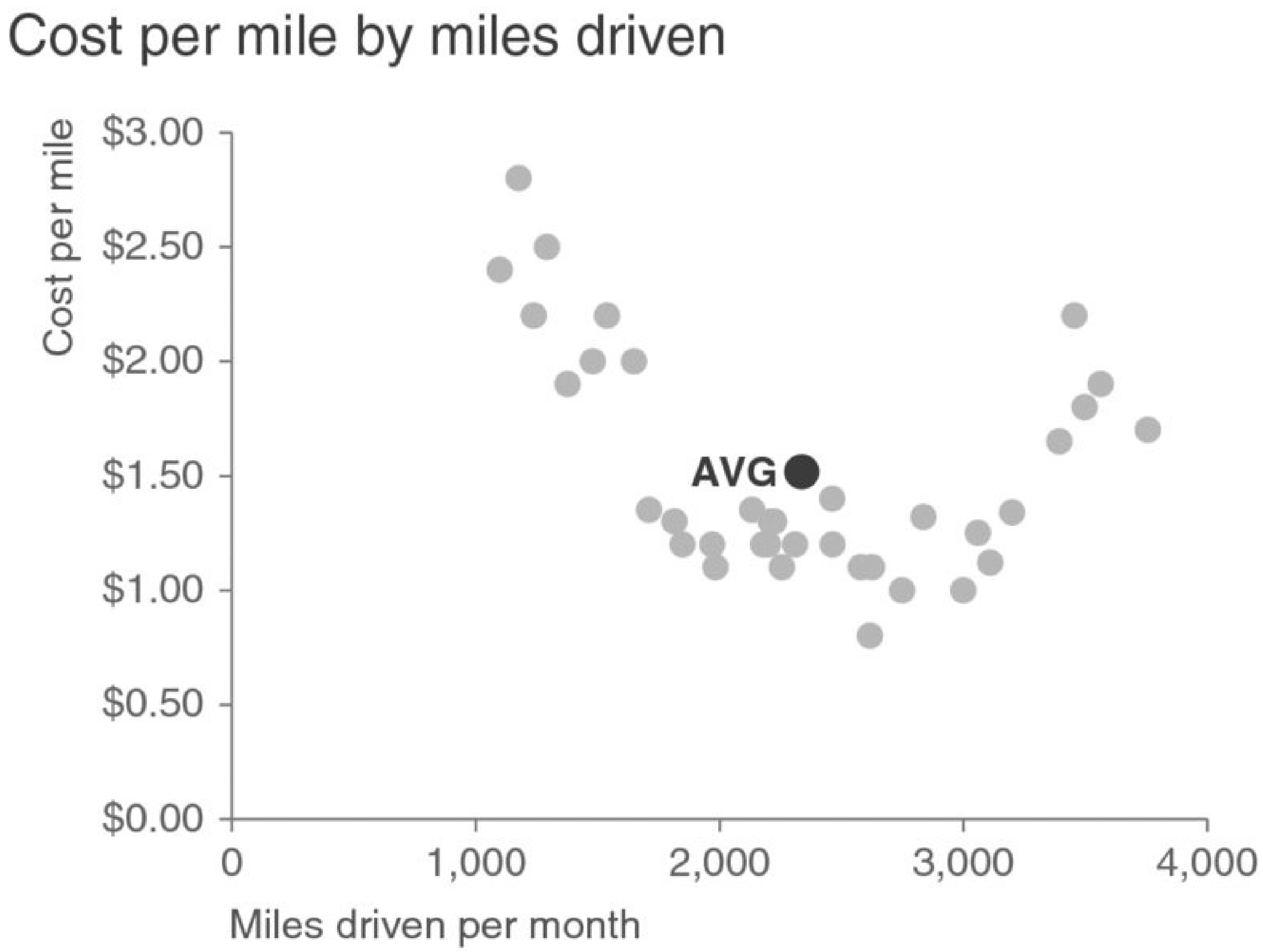


Figure 2.6 Scatterplot

If we want to focus primarily on those cases where cost per mile is above average, a slightly modified scatterplot designed to draw our eye there more quickly might look something like what is shown in Figure 2.7.

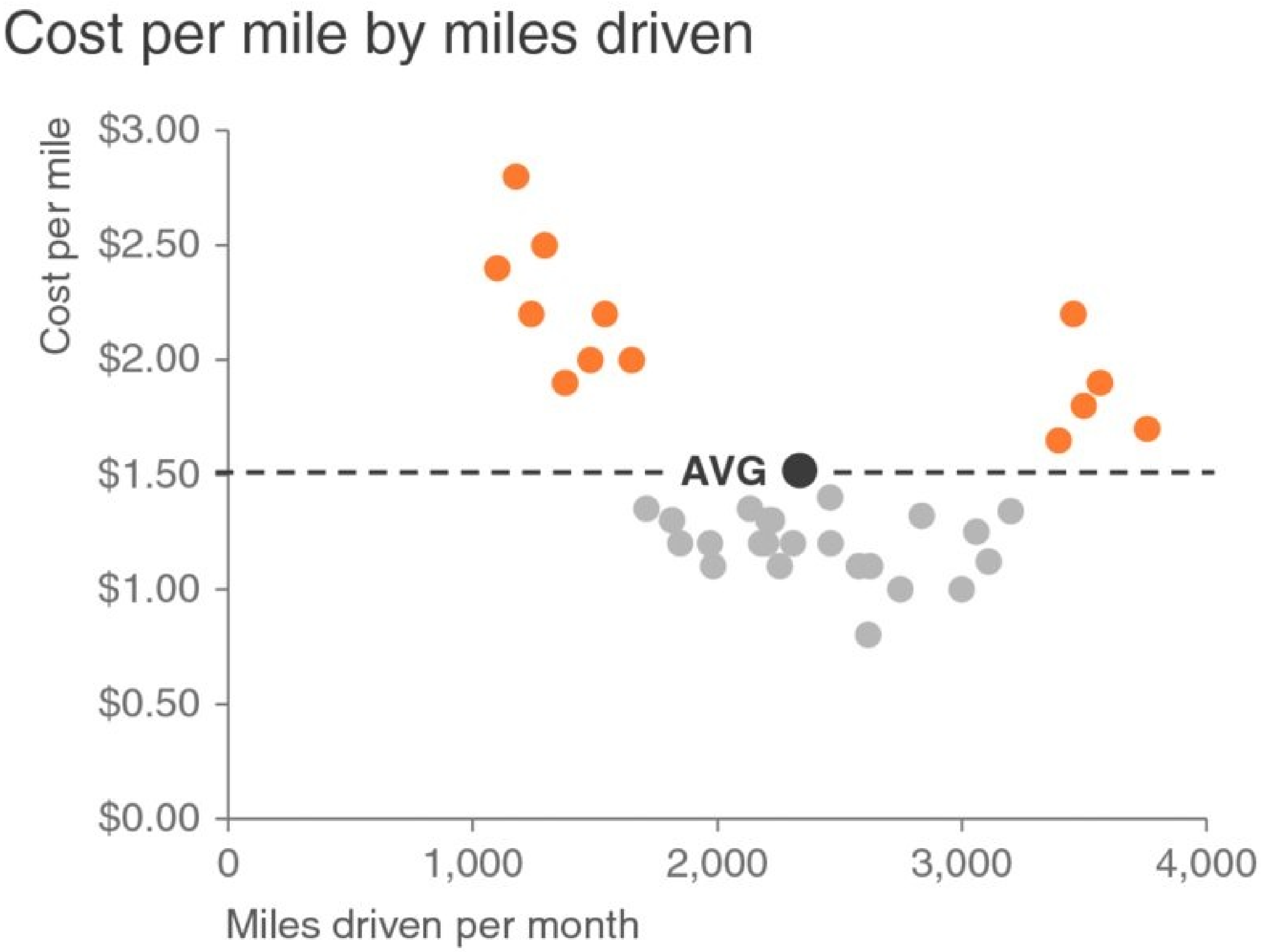


Figure 2.7 Modified scatterplot

We can use Figure 2.7 to make observations such as cost per mile is higher than average when less than about 1,700 miles or more than about 3,300 miles were driven for the sample observed. We'll talk more about the design choices made here and reasons for them in upcoming chapters.

Lines

Line graphs are most commonly used to plot continuous data. Because the points are physically connected via the line, it implies a connection between the points that may not make sense for categorical data (a set of data that is sorted or divided into different categories). Often, our continuous data is in some unit of time: days, months, quarters, or years.

Within the line graph category, there are two types of charts that I frequently find myself using: the standard line graph and the slopegraph.

Line graph

The line graph can show a single series of data, two series of data, or multiple series, as illustrated in Figure 2.8.

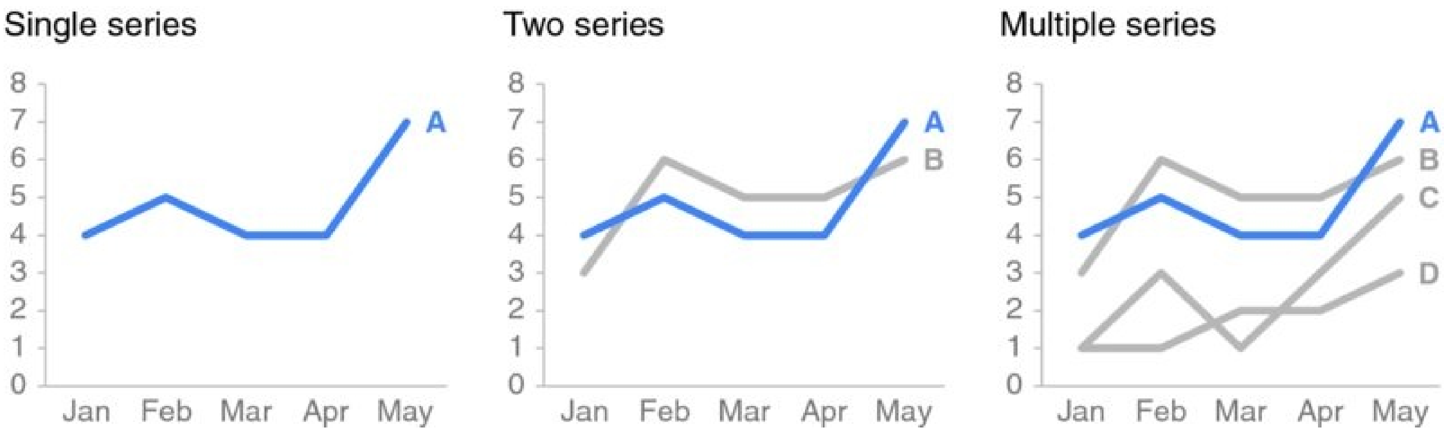


Figure 2.8 Line graphs

Showing average within a range in a line graph

In some cases, the line in your line graph may represent a summary statistic, like the average, or the point estimate of a forecast. If you also want to give a sense of the range (or confidence level, depending on the situation), you can do that directly on the graph by also visualizing this range. For example, the graph in Figure 2.9 shows the minimum, average, and maximum wait times at passport control for an airport over a 13-month period.

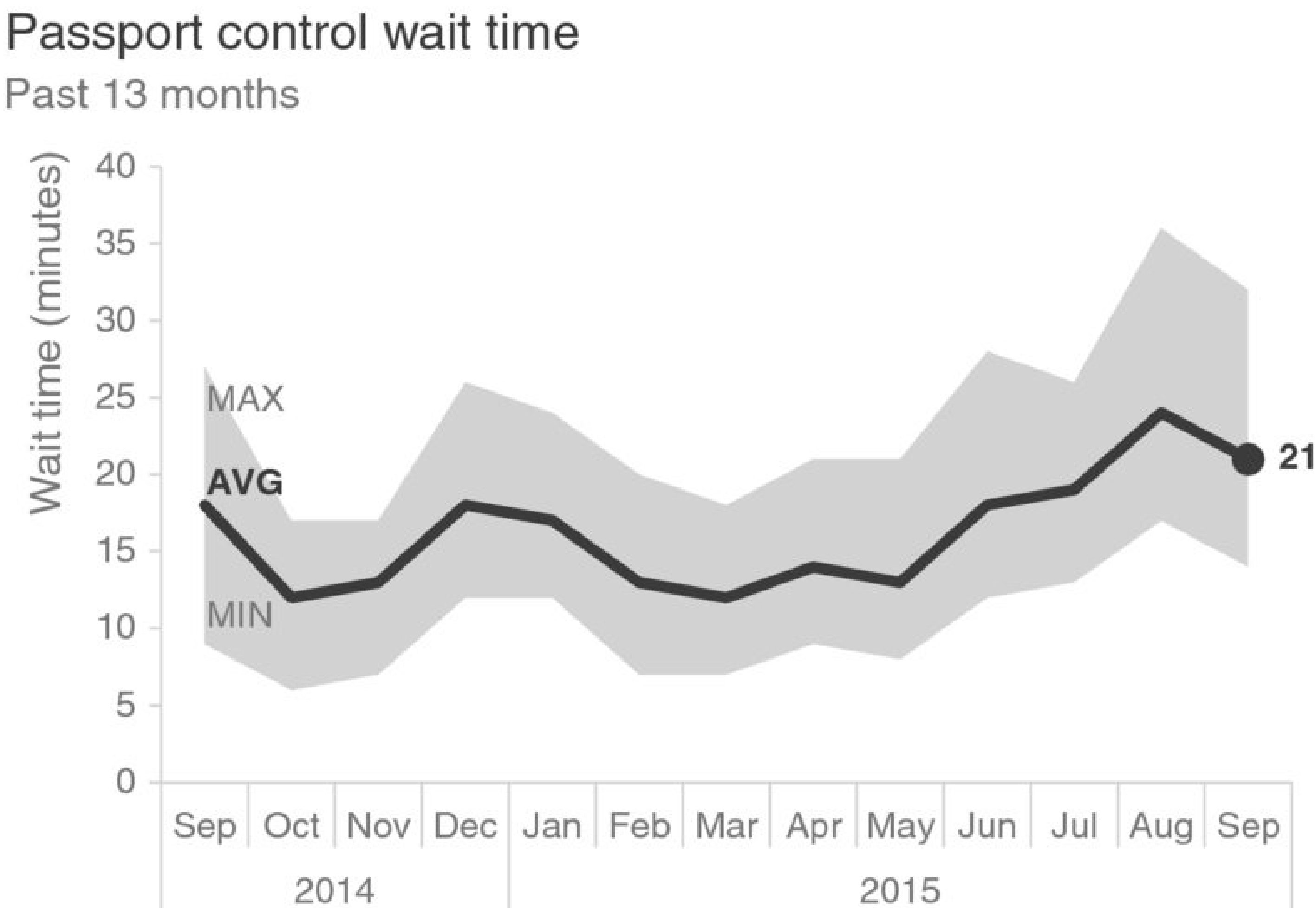


Figure 2.9 Showing average within a range in a line graph

Note that when you're graphing time on the horizontal *x*-axis of a line graph, the data plotted must be in consistent intervals. I recently saw a graph where the units on the *x*-axis were decades from 1900 forward (1910, 1920, 1930, etc.) and then switched to yearly after 2010 (2011, 2012, 2013, 2014). This meant that the distance between the decade points and annual points looked the same. This is a misleading way to show the data. Be consistent in the time points you plot.

Slopegraph

Slopegraphs can be useful when you have two time periods or points of comparison and want to quickly show relative increases and decreases or differences across various categories between the two data points.

The best way to explain the value of and use case for slopegraphs is through a specific example. Imagine that you are analyzing and communicating data from a recent employee feedback survey. To show the relative change in survey categories from 2014 to 2015, the slopegraph might look something like Figure 2.10.

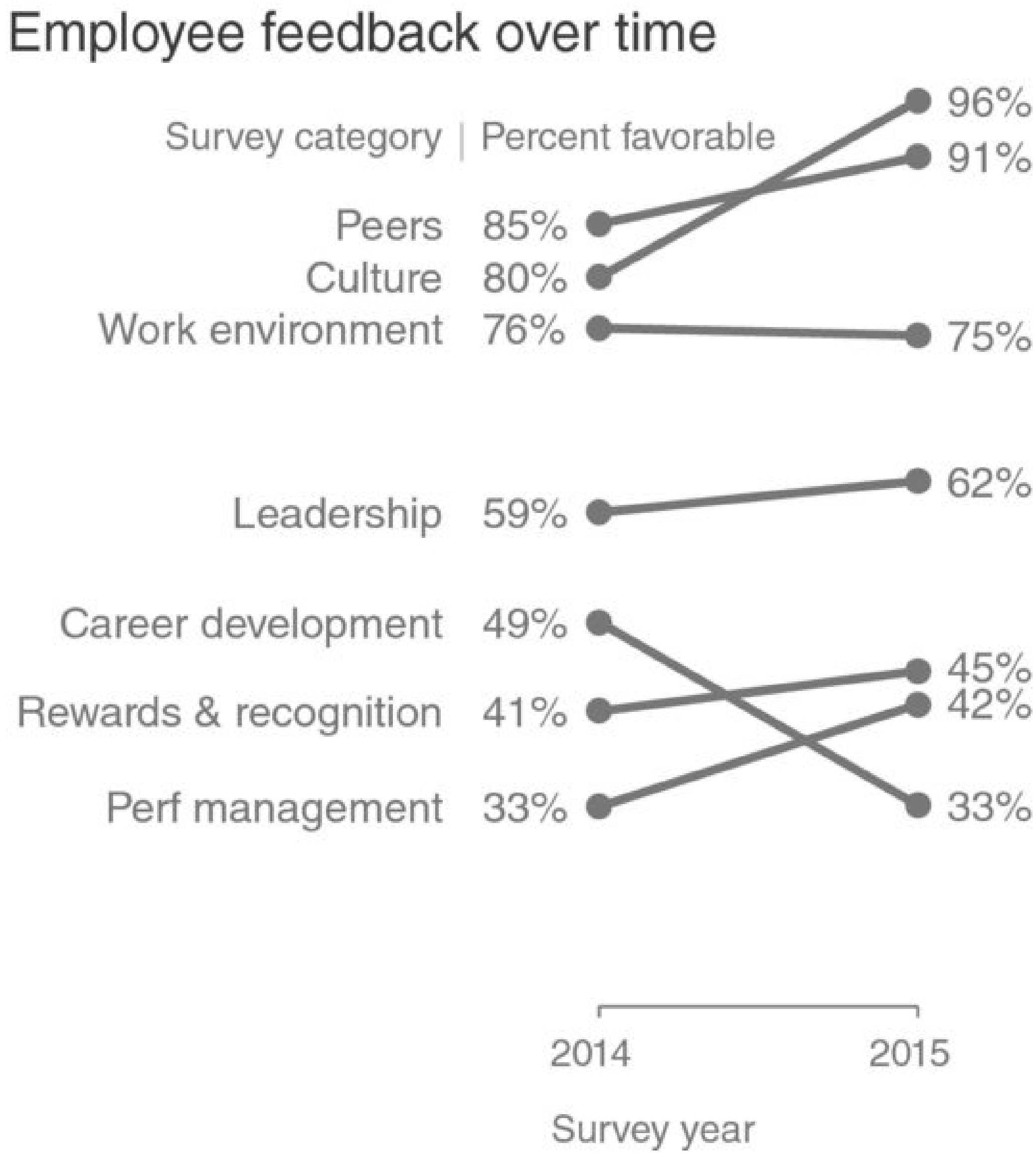


Figure 2.10 Slopegraph

Slopegraphs pack in a lot of information. In addition to the absolute values (the points), the lines that connect them give you the visual increase or decrease in rate of change (via the slope or direction) without ever having to explain that's what they are doing, or what exactly a "rate of change" is—rather, it's intuitive.

Slopegraph template

Slopegraphs can take a bit of patience to set up because they often aren't one of the standard graphs included in graphing applications. An Excel template with an example slopegraph and instructions for customized use can be downloaded here: storytellingwithdata.com/slopegraph-template.

Whether a slopegraph will work in your specific situation depends on the data itself. If many of the lines are overlapping, a slopegraph may not work, though in some cases you can still emphasize a single series at a time with success. For example, we can draw attention to the single category that decreased over time from the preceding example.

In Figure 2.11, our attention is drawn immediately to the decrease in “Career development,” while the rest of the data is preserved for context without competing for attention. We will talk about the strategy behind this when we discuss preattentive attributes in Chapter 4.

Figure 2.11 Modified slopegraph

While lines work well to show data over time, bars tend to be my go-to graph type for plotting categorical data, where information is organized into groups.

Bars

Sometimes bar charts are avoided because they are common. This is a mistake. Rather, bar charts should be leveraged *because they are common*, as this means less of a learning curve for your audience. Instead of using their brain power to try to understand how to read the graph, your audience spends it figuring out what information to take away from the visual.

Bar charts are easy for our eyes to read. Our eyes compare the end points of the bars, so it is easy to see quickly which category is the biggest, which is the smallest, and also the incremental difference between categories. Note that, because of how our eyes compare the relative end points of the bars, it is important that bar charts always have a zero baseline (where the *x*-axis crosses the *y*-axis at zero), otherwise you get a false visual comparison.

Consider Figure 2.12 from Fox News.

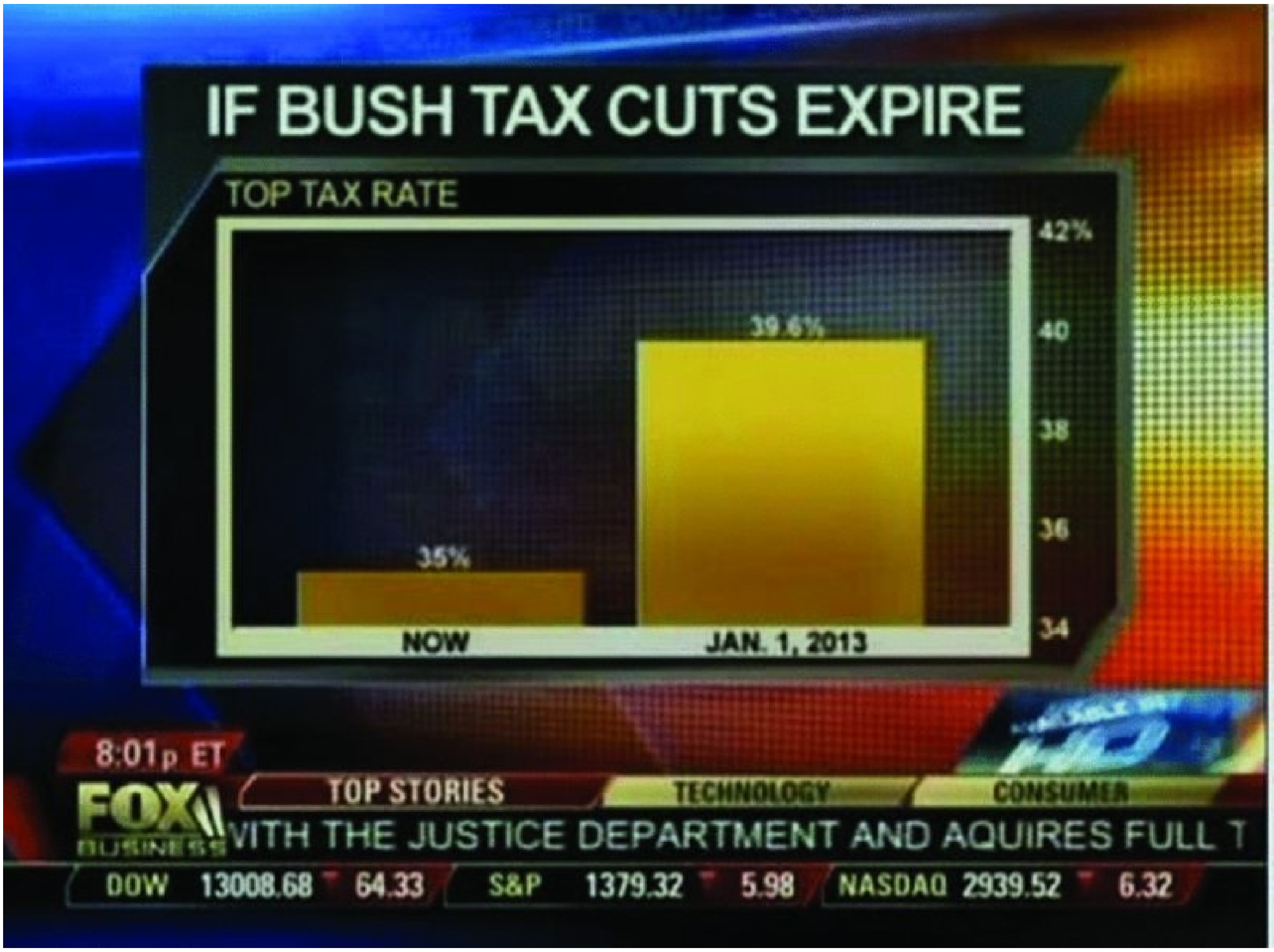


Figure 2.12 Fox News bar chart

For this example, let's imagine we are back in the fall of 2012. We are wondering what will happen if the Bush tax cuts expire. On the left-hand side, we have what the top tax rate is currently, 35%, and on the right-hand side what it will be as of January 1, at 39.6%.

When you look at this graph, how does it make you feel about the potential expiration of the tax cuts? Perhaps worried about the huge increase? Let's take a closer look.

Note that the bottom number on the vertical axis (shown at the far right) is not zero, but rather 34. This means that the bars, in theory, should continue down through the bottom of the page. In fact, the way this is graphed, the visual increase is 460% (the heights of the bars are 35 – 34 = 1 and 39.6 – 34 = 5.6, so (5.6 – 1) / 1 = 460%). If we graph the bars with a zero baseline so that the heights are accurately represented (35 and 39.6), we get an actual visual increase of 13% ((39.6 – 35) / 35). Let's look at a side-by-side comparison in Figure 2.13.

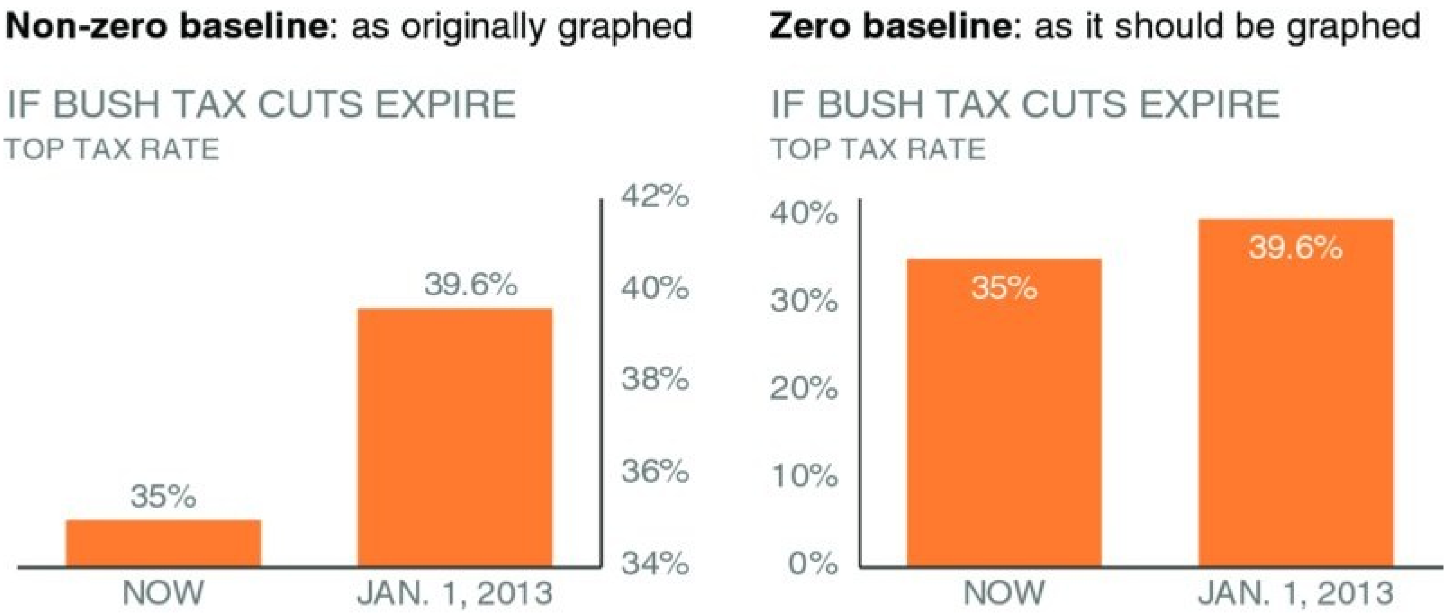


Figure 2.13 Bar charts must have a zero baseline

In Figure 2.13, what looked like a huge increase on the left is reduced considerably when plotted appropriately. Perhaps the tax increase isn't so worrisome, or at least not as severe as originally depicted. Because of the way our eyes compare the relative end points of the bars, it's important to have the context of the entire bar there in order to make an accurate comparison.

You'll note that a couple of other design changes were made in the remake of this visual as well. The *y*-axis labels that were placed on the right-hand side of the original visual were moved to the left (so we see how to interpret the data before we get to the actual data). The data labels that were originally outside of the bars were pulled inside to reduce clutter. If I were plotting this data outside of this specific lesson, I might omit the *y*-axis entirely and show only the data labels within the bars to reduce redundant information. However, in this case, I preserved the axis to make it clear that it begins at zero.

Graph axis vs. data labels

When graphing data, a common decision to make is whether to preserve the axis labels or eliminate the axis and instead label the data points directly. In making this decision, consider the level of specificity needed. If you want your audience to focus on big-picture trends, think about preserving the axis but deemphasizing it by making it grey. If the specific numerical values are important, it may be better to label the data points directly. In this latter case, it's usually best to omit the axis to avoid the inclusion of redundant information. Always consider how you want your audience to use the visual and construct it accordingly.

The rule we've illustrated here is that *bar charts must have a zero baseline*. Note that this rule does not apply to line graphs. With line graphs, since the focus is on the relative position in space (rather than the length from the baseline or axis), you can get away with a nonzero baseline. Still, you should approach with caution—make it clear to your audience that you are using a nonzero baseline and take context into account so you don't overzoom and make minor changes or differences appear significant.

Ethics and data visualization

But what if changing the scale on a bar chart or otherwise manipulating the data better reinforces the point you want to make? Misleading in this manner by inaccurately visualizing data is not OK. Beyond ethical concerns, it is risky territory. All it takes is one discerning audience member to notice the issue (for example, the *y*-axis of a bar chart beginning at something other than zero) and your entire argument will be thrown out the window, along with your credibility.

While we're considering lengths of bars, let's also spend a moment on the *width* of bars. There's no hard-and-fast rule here, but in general the bars should be wider than the white space between the bars. You don't want the bars to be so wide, however, that your audience wants to compare areas instead of lengths. Consider the following "Goldilocks" of bar charts: too thin, too thick, and just right.

We've discussed some best practices when it comes to bar charts in general. Next let's take a look at some different varieties. Having a number of bar charts at your disposal gives you flexibility when facing different data visualization challenges. We'll look at the ones I think you should be familiar with here.

Vertical bar chart

The plain vanilla bar chart is the vertical bar chart, or column chart. Like line graphs, vertical bar charts can be single series, two series, or multiple series. Note that as you add more series of data, it becomes more difficult to focus on one at a time and pull out insight, so use multiple series bar charts with caution. Be aware also that there is visual grouping that happens as a result of the spacing in bar charts having more than one data series. This makes the relative order of the categorization important. Consider what you want your audience to be able to compare, and structure your categorization hierarchy to make that as easy as possible.

Stacked vertical bar chart

Use cases for stacked vertical bar charts are more limited. They are meant to allow you to compare totals across categories and also see the subcomponent pieces within a given category. This can quickly become visually overwhelming, however—especially given the varied default color schemes in most graphing applications (more to come on that). It is hard to compare the subcomponents across the various categories once you get beyond the bottom series (the one directly next to the *x*-axis) because you no longer have a consistent baseline to use to compare. This makes it a harder comparison for our eyes to make, as illustrated in Figure 2.16.

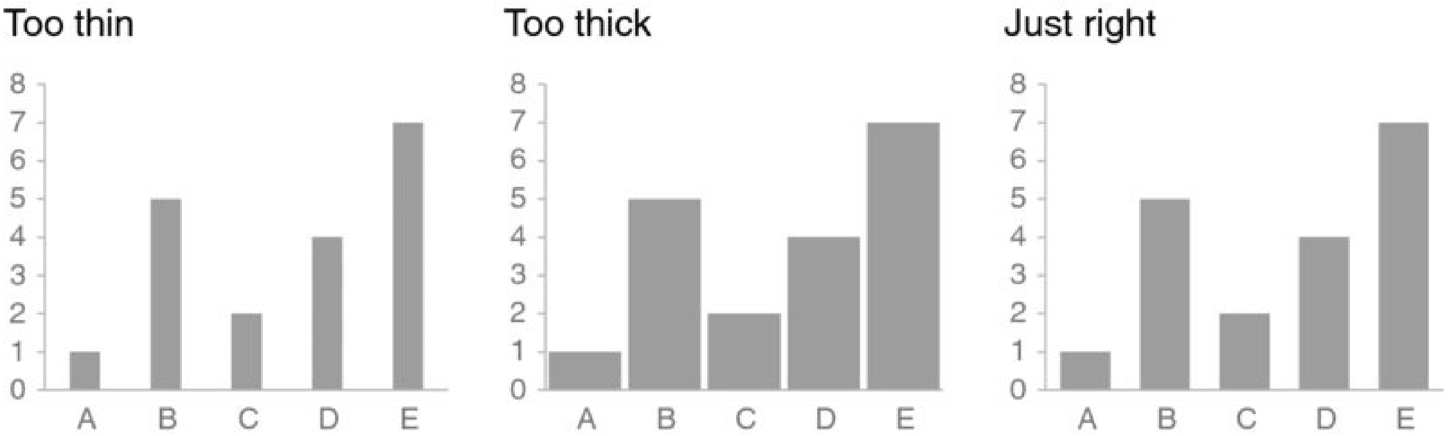


Figure 2.14 Bar width

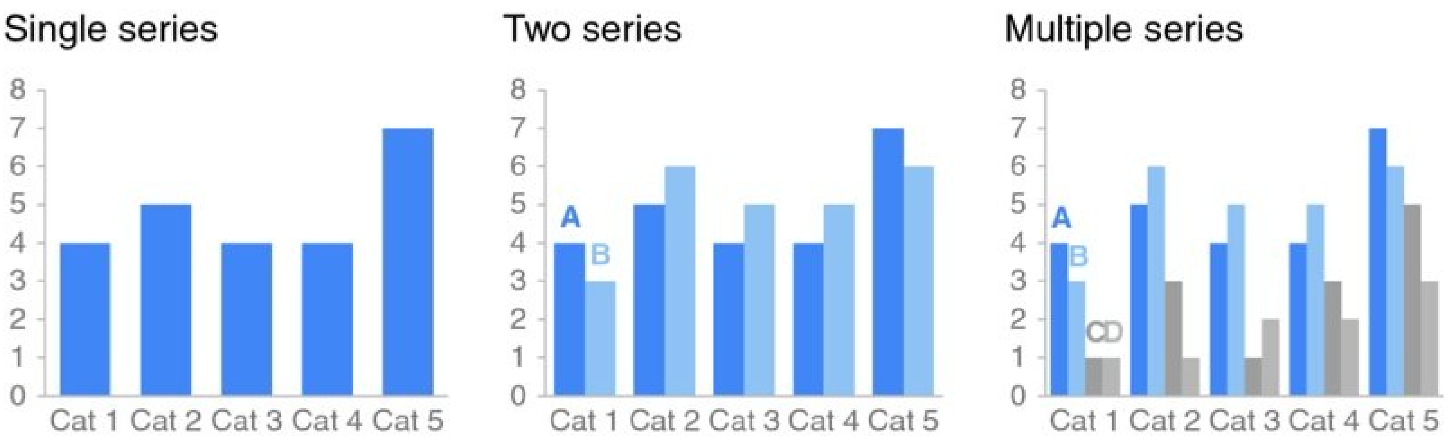


Figure 2.15 Bar charts

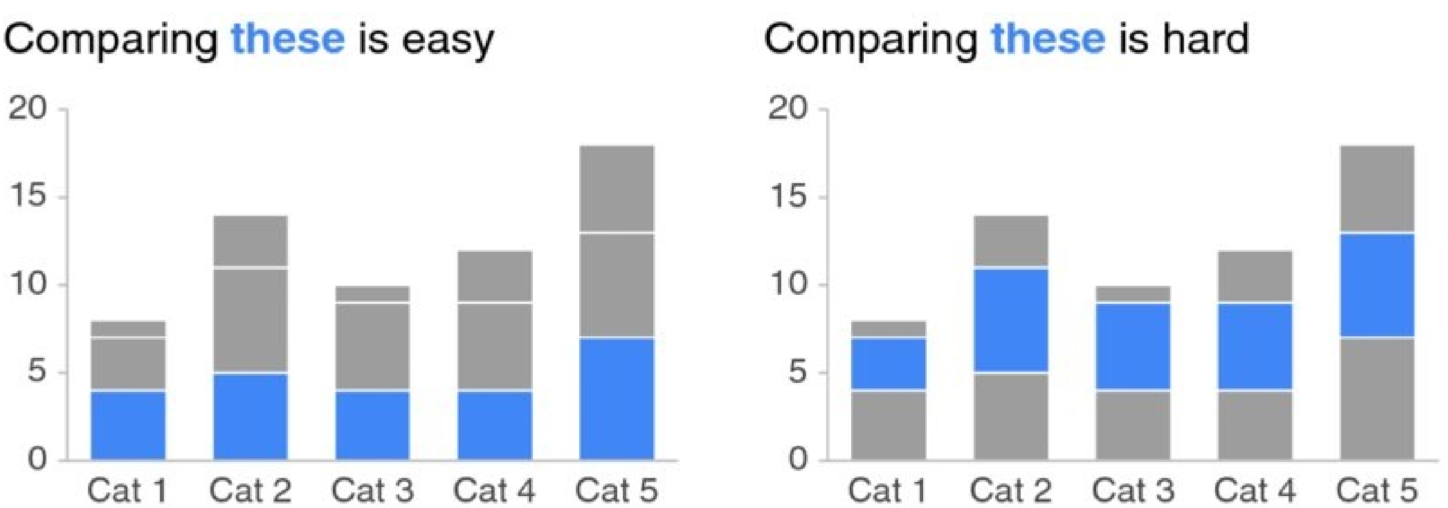


Figure 2.16 Comparing series with stacked bar charts

The stacked vertical bar chart can be structured as absolute numbers (where you plot the numbers directly, as shown in Figure 2.16), or with each column summing to 100% (where you plot the percent of total for each vertical segment; we'll look at a specific example of this in Chapter 9). Which you choose depends on what you are trying to communicate to your audience. When you use the 100% stacked bar, think about whether it makes sense to also include the absolute numbers for each category total (either in an unobtrusive way in the graph directly, or possibly in a footnote), which may aid in the interpretation of the data.

Waterfall chart

The waterfall chart can be used to pull apart the pieces of a stacked bar chart to focus on one at a time, or to show a starting point, increases and decreases, and the resulting ending point.

The best way to illustrate the use case for a waterfall chart is through a specific example. Imagine that you are an HR business partner and want to understand and communicate how employee headcount has changed over the past year for the client group you support.

A waterfall chart showing this breakdown might look something like Figure 2.17.

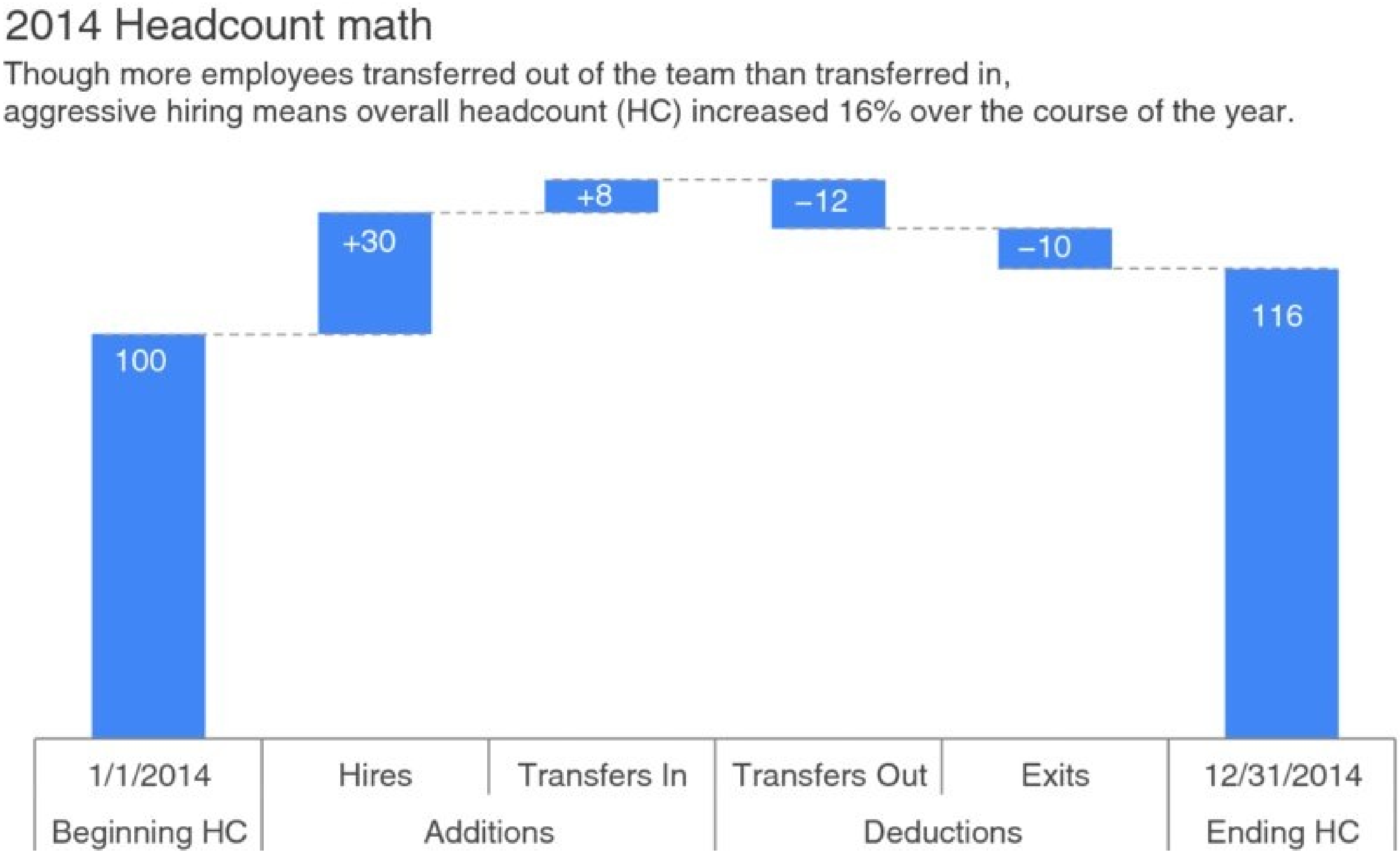


Figure 2.17 Waterfall chart

On the left-hand side, we see what the employee headcount for the given team was at the beginning of the year. As we move to the right, first we encounter the incremental additions: new hires and employees transferring into the team from other parts of the organization. This is followed by the deductions: transfers out of the team to other parts of the organization and attrition. The final column represents employee headcount at the end of the year, after the additions and deductions have been applied to the beginning of year headcount.

Brute-force waterfall charts

If your graphing application doesn't have waterfall chart functionality built in, fret not. The secret is to leverage the stacked bar chart and make the first series (the one that appears closest to the *x*-axis) invisible. It takes a bit of math to set up correctly, but it works great. A blog post on this topic, along with an example Excel version of the above chart and instructions on how to set one up for your own purposes can be downloaded at storytellingwithdata.com/waterfall-chart.

Horizontal bar chart

If I had to pick a single go-to graph for categorical data, it would be the horizontal bar chart, which flips the vertical version on its side. Why? Because it is *extremely easy to read*. The horizontal bar chart is especially useful if your category names are long, as the text is written from left to right, as most audiences read, making your graph legible for your audience. Also, because of the way we typically process information—starting at top left and making z's with our eyes across the screen or page—the structure of the horizontal bar chart is such that our eyes hit the category names before the actual data. This means by the time we get to the data, we already know what it represents (instead of the darting back and forth our eyes do between the data and category names with vertical bar charts).

Like the vertical bar chart, the horizontal bar chart can be single series, two series, or multiple series (Figure 2.18).

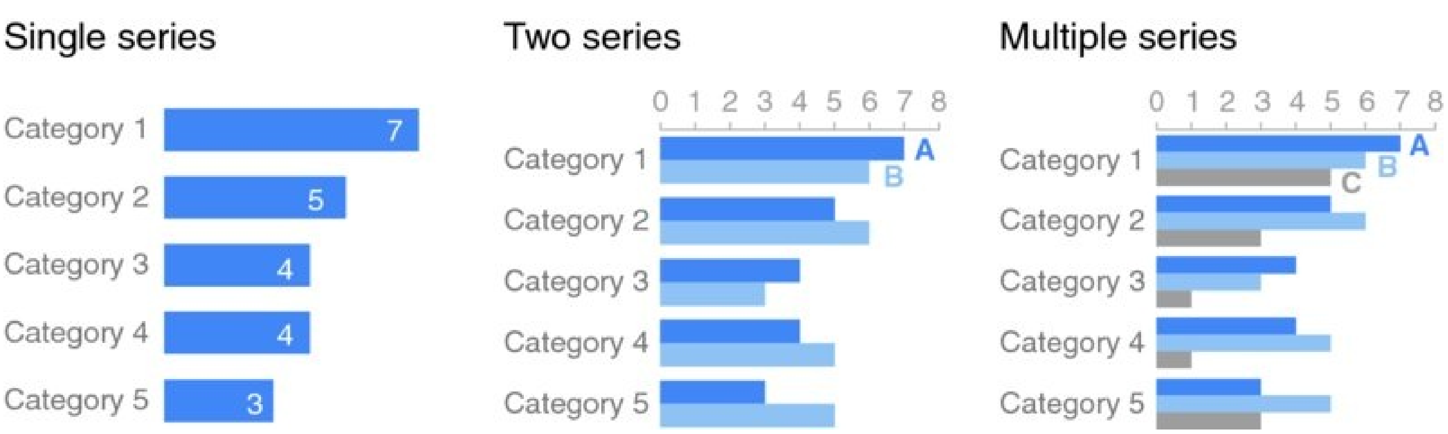


Figure 2.18 Horizontal bar charts

The logical ordering of categories

When designing any graph showing categorical data, be thoughtful about how your categories are ordered. If there is a natural ordering to your categories, it may make sense to leverage that. For example, if your categories are age groups—0–10 years old, 11–20 years old, and so on—keep the categories in numerical order. If, however, there isn't a natural ordering in your categories that makes sense to leverage, think about what ordering of your data will make the most sense. Being thoughtful here can mean providing a construct for your audience, easing the interpretation process.

Your audience (without other visual cues) will typically look at your visual starting at the top left and zigzagging in "z" shapes. This means they will encounter the top of your graph first. If the biggest category is the most important, think about putting that first and ordering the rest of the categories in decreasing numerical order. Or if the smallest is most important, put that at the top and order by ascending data values.

For a specific example about the logical ordering of data, check out case study 3 in Chapter 9.

Stacked horizontal bar chart

Similar to the stacked vertical bar chart, stacked horizontal bar charts can be used to show the totals across different categories but also give a sense of the subcomponent pieces. They can be structured to show either absolute values or sum to 100%.

I find this latter approach can work well for visualizing portions of a whole on a scale from negative to positive, because you get a consistent baseline on both the far left and the far right, allowing for easy comparison of the left-most pieces as well as the right-most pieces. For example, this approach can work well for visualizing survey data collected along a Likert scale (a scale commonly used in surveys that typically ranges from Strongly Disagree to Strongly Agree), as shown in Figure 2.19.

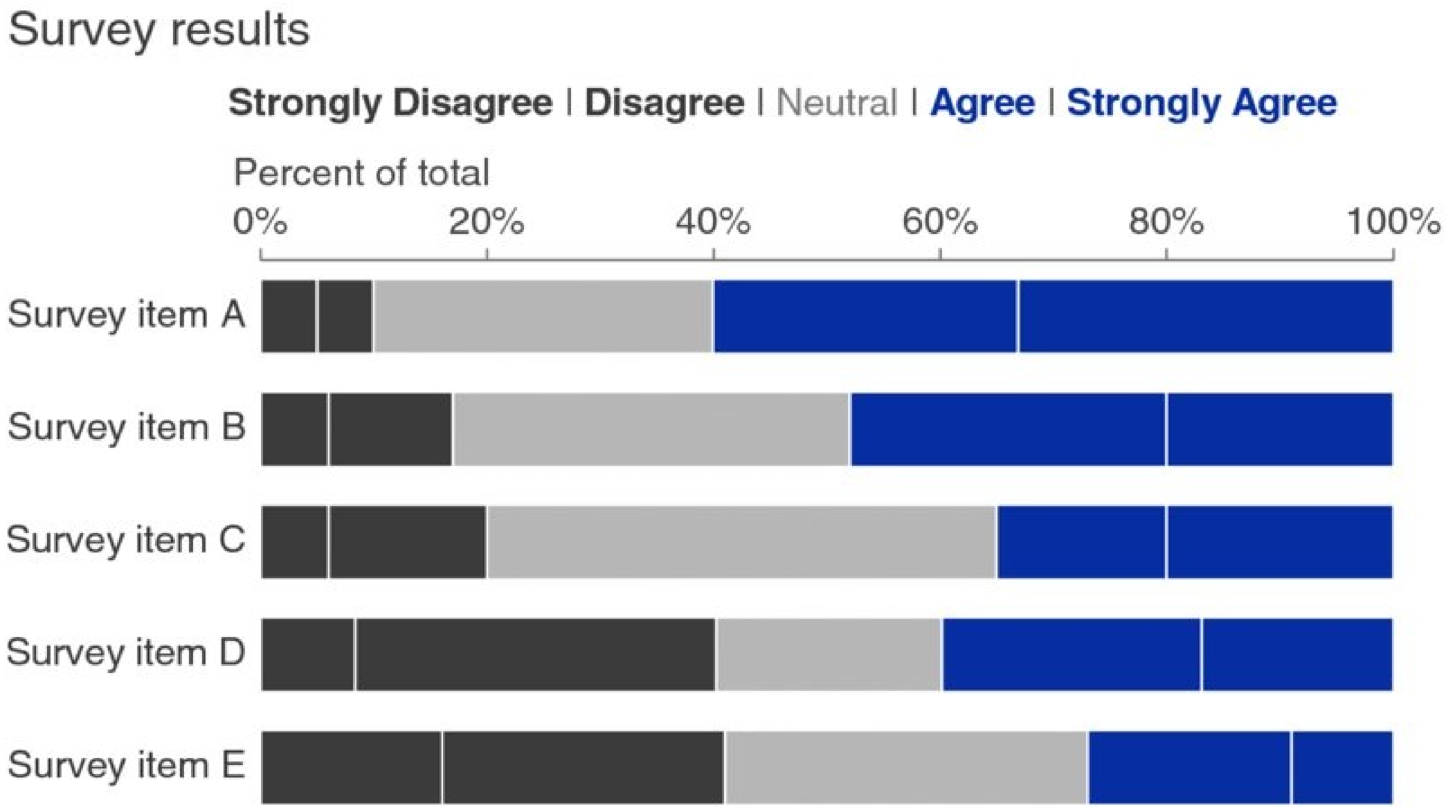


Figure 2.19 100% stacked horizontal bar chart

Area

I avoid most area graphs. Humans' eyes don't do a great job of attributing quantitative value to two-dimensional space, which can render area graphs harder to read than some of the other types of visual displays we've discussed. For this reason, I typically avoid them, with one exception—when I need to visualize numbers of vastly different magnitudes. The second dimension you get using a square for this (which has both height and width, compared to a bar that has only height *or* width) allows this to be done in a more compact way than possible with a single dimension, as shown in Figure 2.20.

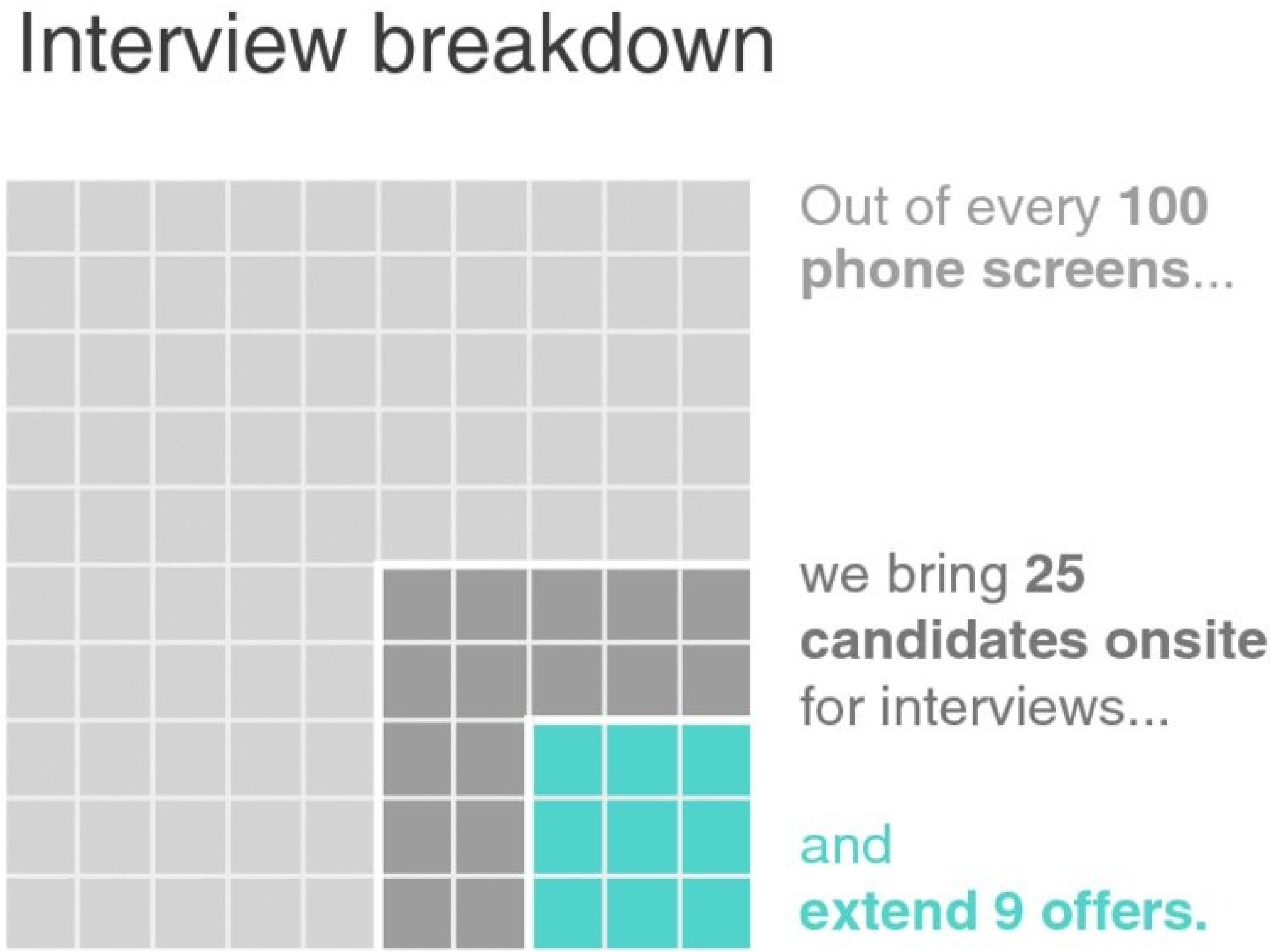


Figure 2.20 Square area graph

Other types of graphs

What I've covered up to this point are the types of graphs I find myself commonly using. This is certainly not an exhaustive list. However, they should meet the majority of your everyday needs. Mastering the basics is imperative before exploring novel types of data visualization.

There are many other types of graphs out there. When it comes to selecting a graph, first and foremost, choose a graph type that will enable you to clearly get your message across to your audience. With less familiar types of visuals, you will likely need to take extra care in making them accessible and understandable.

Infographics

Infographic is a term that is frequently misused. An infographic is simply a graphical representation of information or data. Visuals coined infographic run the gamut from fluffy to informative. On the inadequate end of the spectrum, they often include elements like garish, oversized numbers and cartoonish graphics. These designs have a certain visual appeal and can seduce the reader. On second glance, however, they appear shallow and leave a discerning audience dissatisfied. Here, the description of "information graphic"—though often used—is not appropriate. On the other end of the spectrum are infographics that live up to their name and actually inform. There are many good examples in the area of data journalism (for example, the *New York Times* and *National Geographic*).

There are critical questions information designers must be able to answer before they begin the design process. These are the same questions we've discussed when it comes to understanding the context for storytelling with data. Who is your audience? What do you need them to know or do? It is only after the answers to these questions can be succinctly articulated that an effective method of display that will best aid the message can be chosen. Good data visualization—infographic or otherwise—is not simply a collection of facts on a given topic; good data visualization tells a story.

To be avoided

We've discussed the visuals that I use most commonly to communicate data in a business setting. There are also some specific graph types and elements that you should avoid: pie charts, donut charts, 3D, and secondary *y*-axes. Let's discuss each of these.

Pie charts are evil

I have a well-documented disdain for pie charts. In short, they are evil. To understand how I arrived at this conclusion, let's look at an example.

The pie chart shown in Figure 2.21 (based on a real example) shows market share across four suppliers: A, B, C, and D. If I asked you to make a simple observation—which supplier is the largest based on this visual—what would you say?

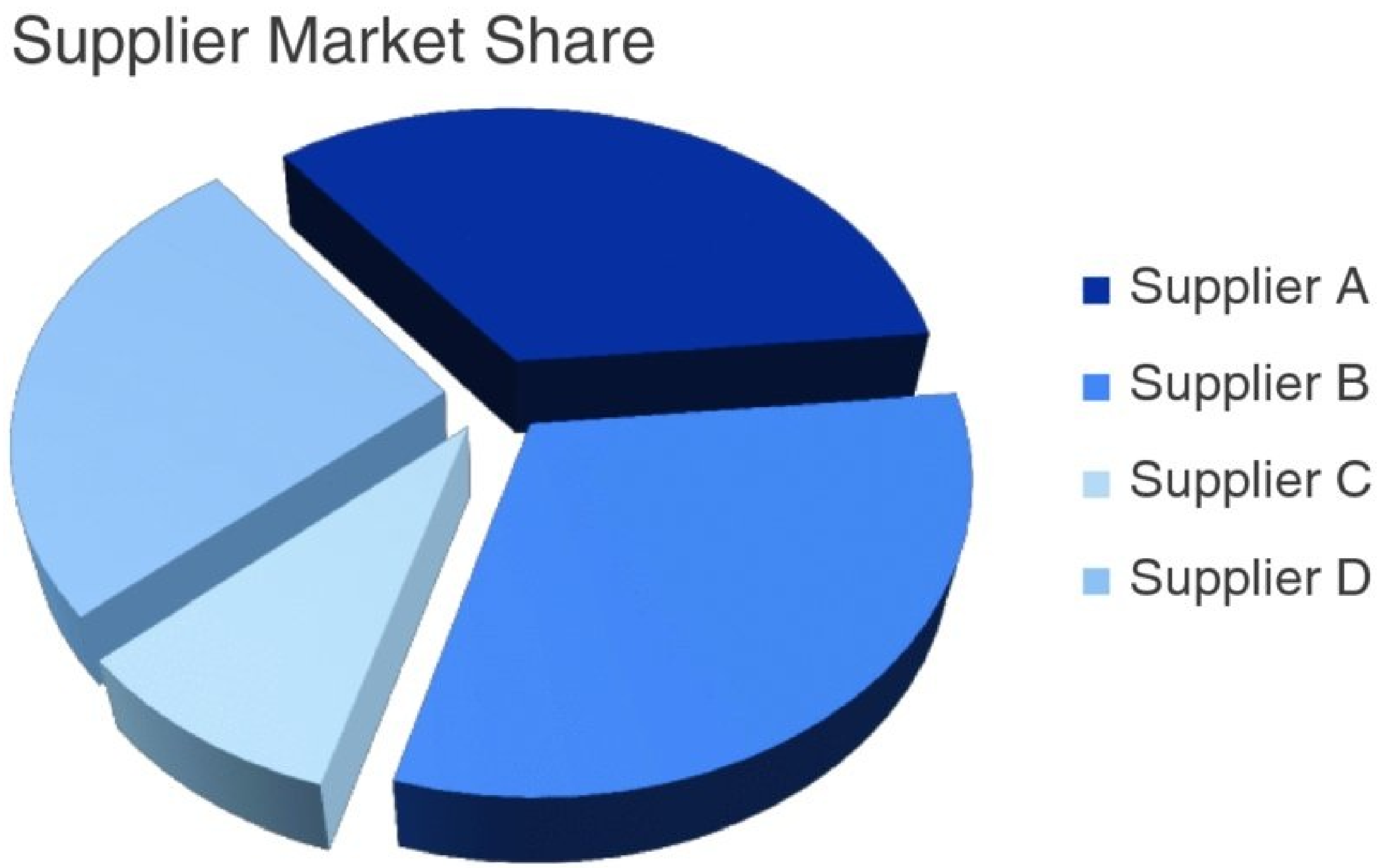


Figure 2.21 Pie chart

Most people will agree that “Supplier B,” rendered in medium blue at the bottom right, appears to be the largest. If you had to estimate what proportion supplier B makes up of the overall market, what percent might you estimate?

35%?

40%?

Perhaps you can tell by my leading questioning that something fishy is going on here. Take a look at what happens when we add the numbers to the pie segments, as shown in Figure 2.22.

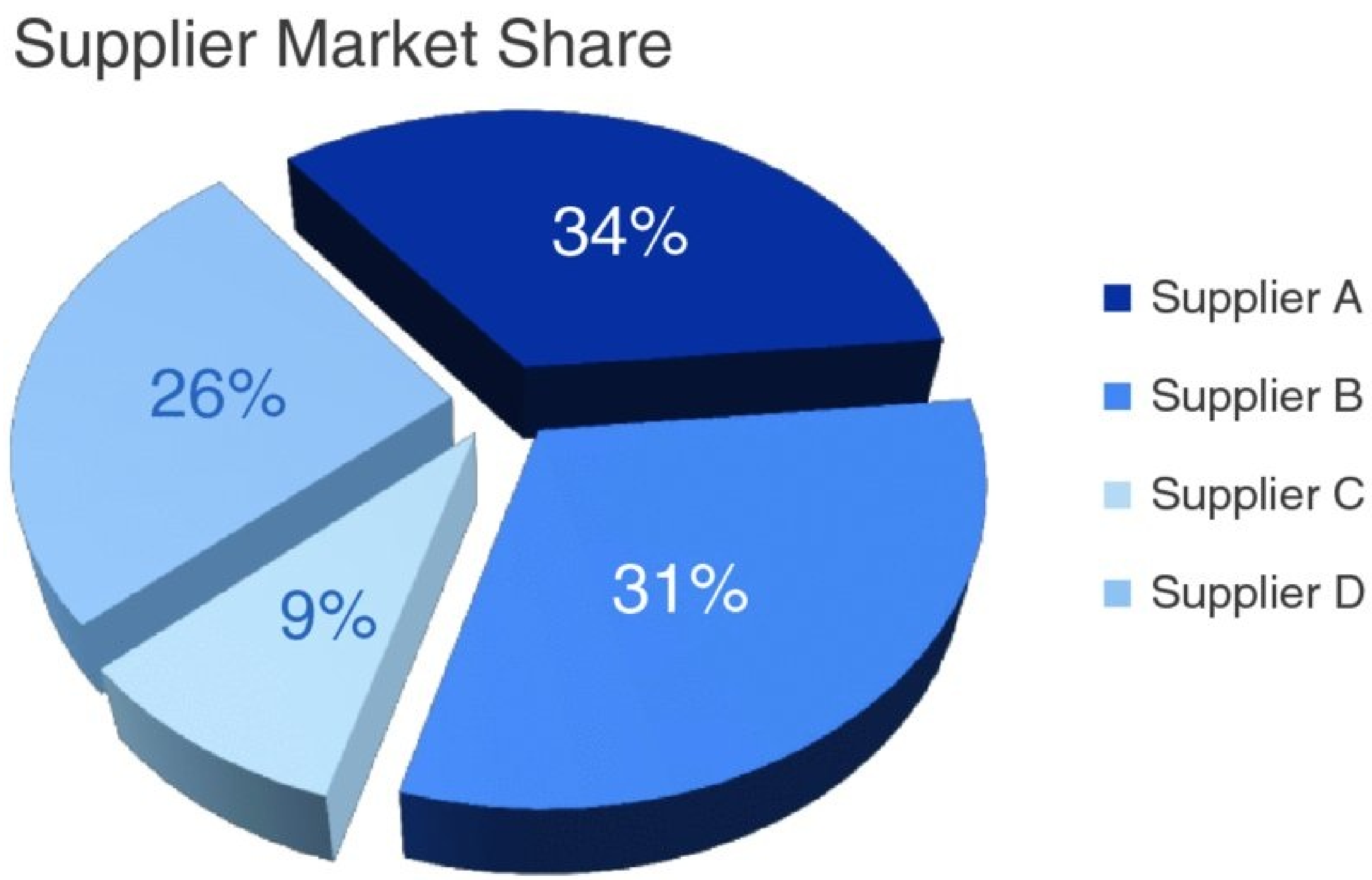


Figure 2.22 Pie chart with labeled segments

"Supplier B"—which *looks* largest, at 31%—is actually smaller than "Supplier A" above it, which looks smaller.

Let's discuss a couple of issues that pose a challenge for accurately interpreting this data. The first thing that catches your eye (and suspicion, if you're a discerning chart reader) is the 3D and strange perspective that's been applied to the graph, tilting the pie and making the pieces at the top appear farther away and thus smaller than they actually are, while the pieces at the bottom appear closer and thus bigger than they actually are. We'll talk more about 3D soon, but for now I'll articulate a relevant data visualization rule: *don't use 3D!* It does nothing good, and can actually do a whole lot of harm, as we see here with the way it skews the visual perception of the numbers.

Even when we strip away the 3D and flatten the pie, interpretation challenges remain. The human eye isn't good at ascribing quantitative value to two-dimensional space. Said more simply: *pie charts are hard for people to read*. When segments are close in size, it's difficult (if not impossible) to tell which is bigger. When they aren't close in size, the best you can do is determine that one is bigger than the other, but you can't judge by how much. To get over this, you can add data labels as has been done here. But I'd still argue the visual isn't worth the space it takes up.

What should you do instead? One approach is to replace the pie chart with a horizontal bar chart, as illustrated in Figure 2.23, organized from greatest to least or vice versa (unless there is some natural ordering to the categories that makes sense to leverage, as mentioned earlier). Remember, with bar charts, our eyes compare the end points. Because they are aligned at a common baseline, it is easy to assess relative size. This makes it straightforward to see not only which segment is the largest, for example, but also *how incrementally larger* it is than the other segments.

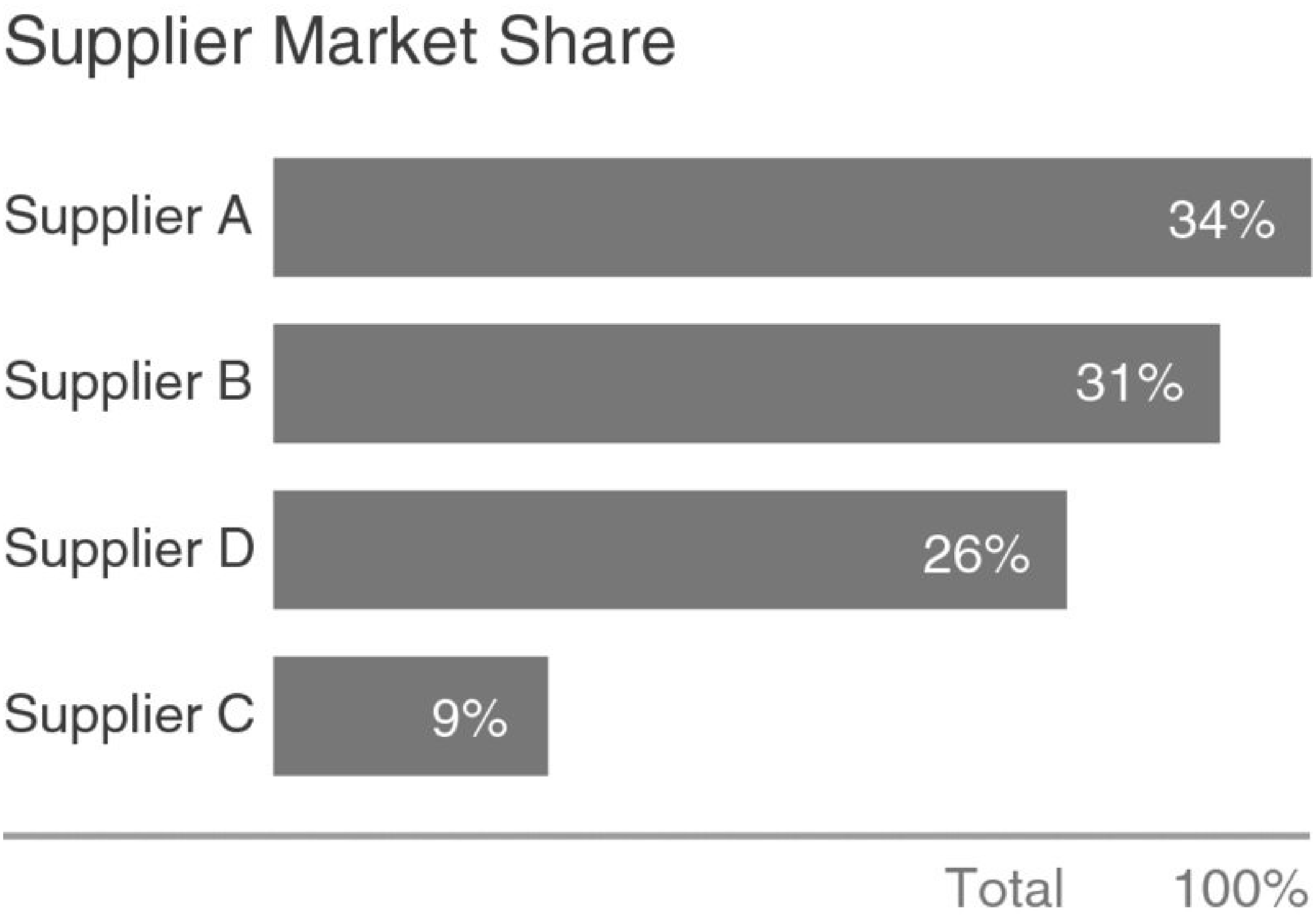


Figure 2.23 An alternative to the pie chart

One might argue that you lose something in the transition from pie to bar. The unique thing you get with a pie chart is the concept of there being a whole and, thus, parts of a whole. But if the visual is difficult to read, is it worth it? In Figure 2.23, I've tried to address this by showing that the pieces sum to 100%. It isn't a perfect solution, but something to consider. For more alternatives to pie charts, check out case study 5 in Chapter 9.

If you find yourself using a pie chart, pause and ask yourself: *why?* If you're able to answer this question, you've probably put enough thought into it to use the pie chart, but it certainly shouldn't be the first type of graph that you reach for, given some of the difficulties in visual interpretation we've discussed here.

While we're on the topic of pie charts, let's look quickly at another "dessert visual" to avoid: the donut chart.

With pies, we are asking our audience to compare angles and areas. With a donut chart, we are asking our audience to compare one arc length to another arc length (for example, in Figure 2.24, the length of *arc A* compared to *arc B*). How confident do you feel in your eyes' ability to ascribe quantitative value to an arc length?

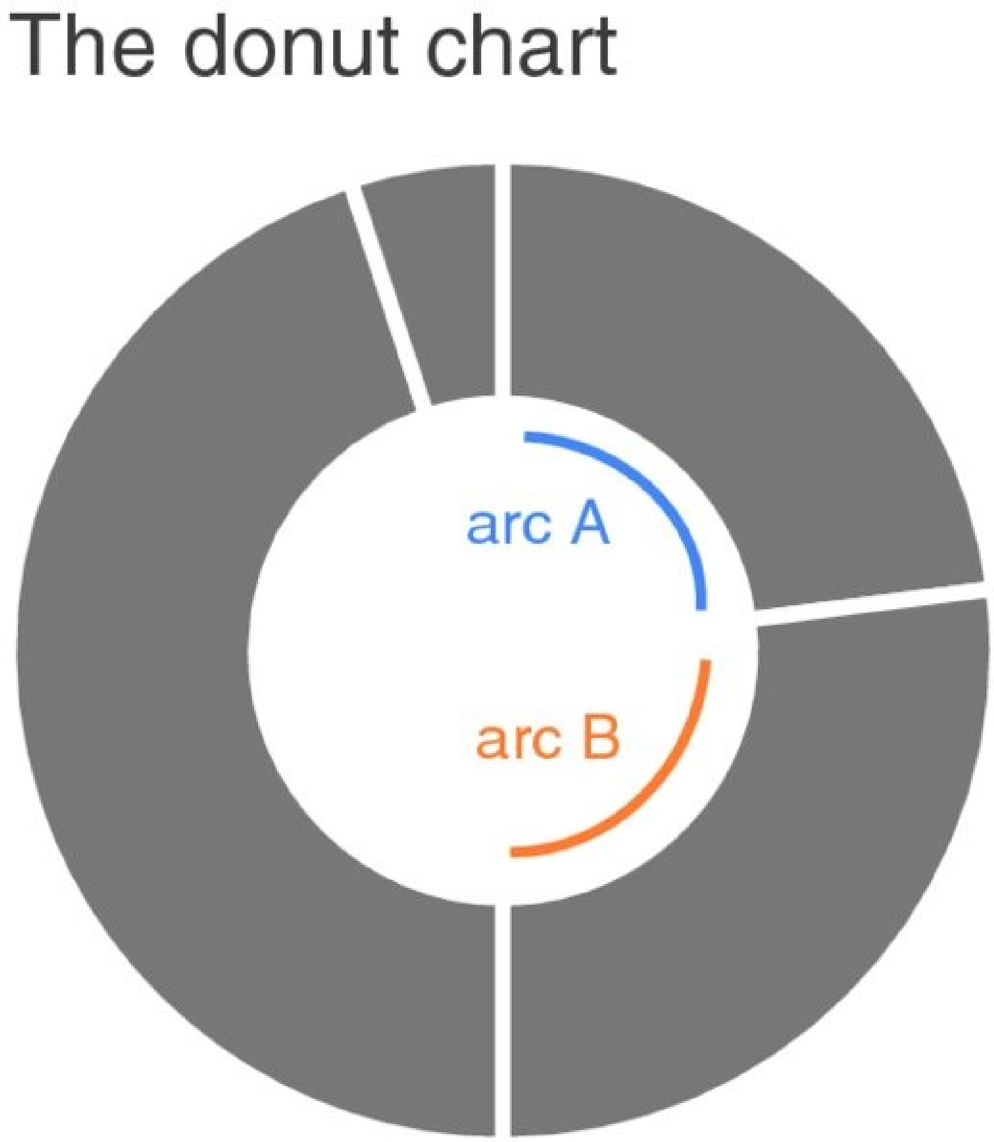


Figure 2.24 Donut chart

Not very? That's what I thought. Don't use donut charts.

Never use 3D

One of the golden rules of data visualization goes like this: never use 3D. Repeat after me: never use 3D. The only exception is if you are actually *plotting a third dimension* (and even then, things get really tricky really quickly, so take care when doing this)—and you should never use 3D to plot a single dimension. As we saw in the pie chart example previously, 3D skews our numbers, making them difficult or impossible to interpret or compare.

Adding 3D to graphs introduces unnecessary chart elements like side and floor panels. Even worse than these distractions, graphing applications do some pretty strange things when it comes to plotting values in 3D. For example, in a 3D bar chart, you might think that your graphing application plots the front of the bar or perhaps the back of the bar. Unfortunately, it's often even less straightforward than that. In Excel, for example, the bar height is determined by an invisible tangent plane intersecting the corresponding height on the *y*-axis. This gives rise to graphs like the one shown in Figure 2.25.

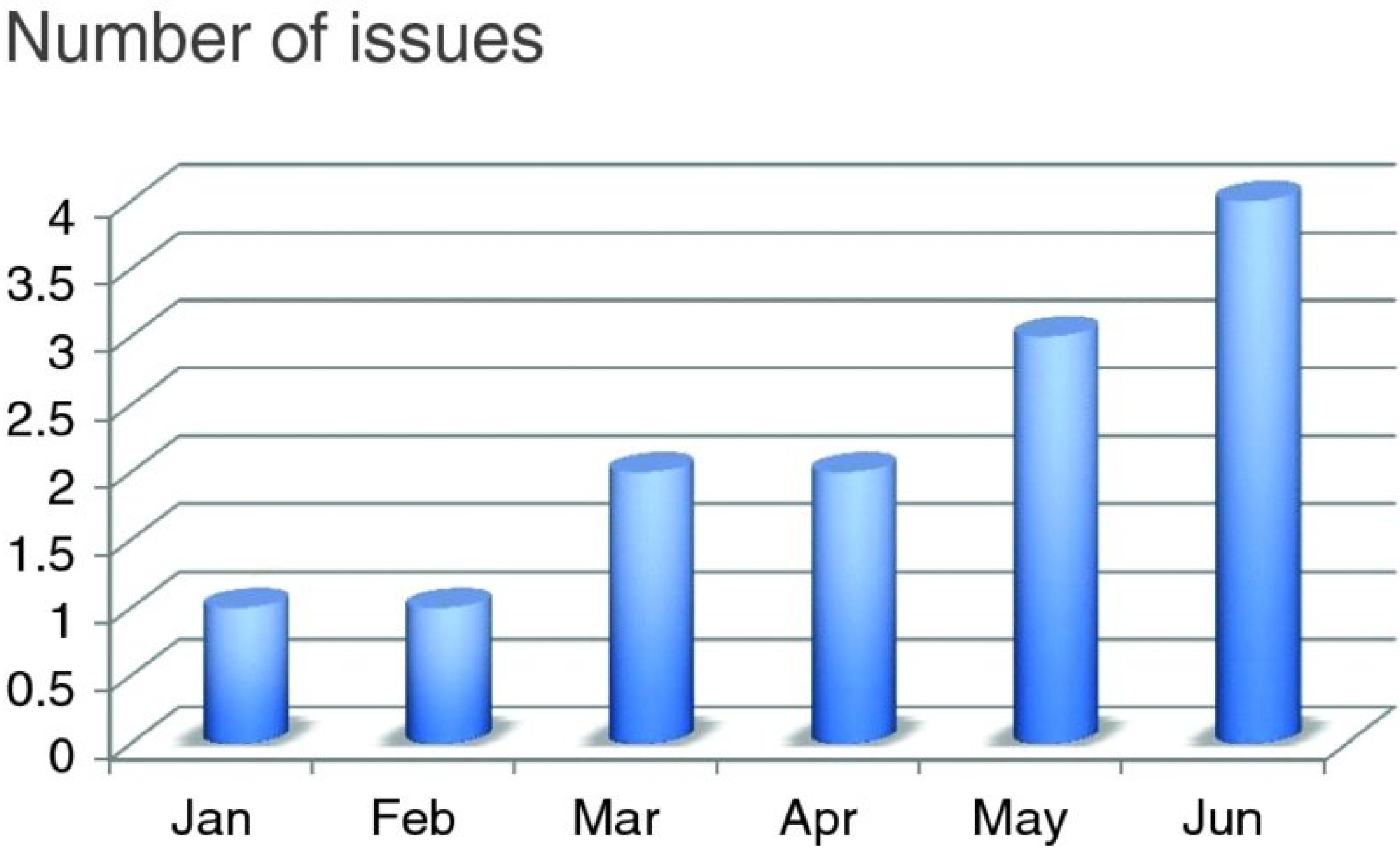


Figure 2.25 3D column chart

Judging by Figure 2.25, how many issues were there in January and February? I've plotted a single issue for each of these months. However, the way I read the chart, if I compare the bar height to the gridlines and follow it leftward to the *y*-axis, I'd estimate visually a value of maybe 0.8. This is simply bad data visualization. Don't use 3D.

Secondary *y*-axis: generally not a good idea

Sometimes it's useful to be able to plot data that is in entirely different units against the same *x*-axis. This often gives rise to the secondary *y*-axis: another vertical axis on the right-hand side of the graph. Consider the example shown in Figure 2.26.

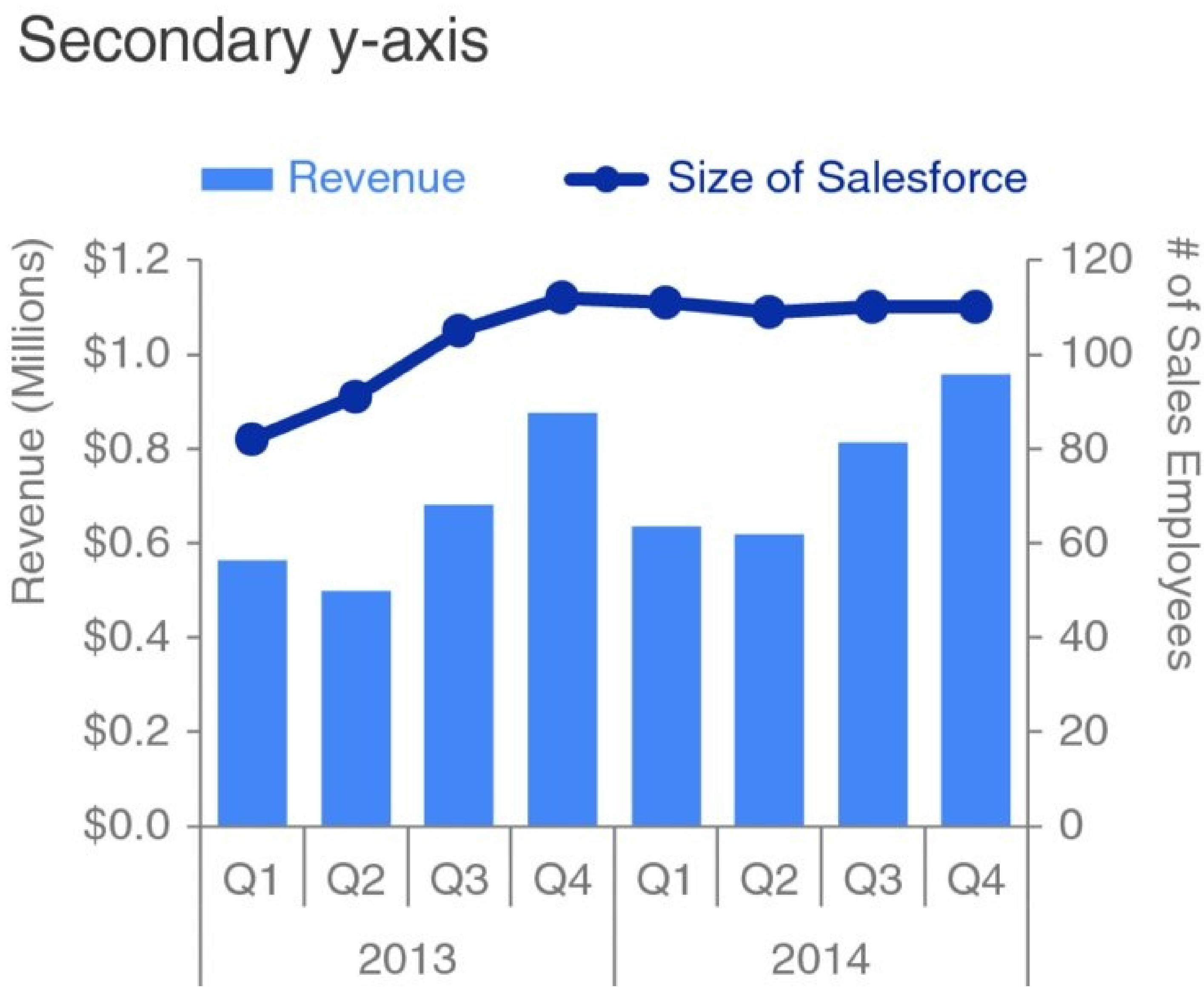


Figure 2.26 Secondary *y*-axis

When interpreting Figure 2.26, it takes some time and reading to understand which data should be read against which axis. Because of this, you should avoid the use of a secondary or right-hand *y*-axis. Instead, think about whether one of the following approaches will meet your needs:

1. Don't show the second *y*-axis. Instead, label the data points that belong on this axis directly.
2. Pull the graphs apart vertically and have a separate *y*-axis for each (both along the left) but leverage the same *x*-axis across both.

Figure 2.27 illustrates these options.

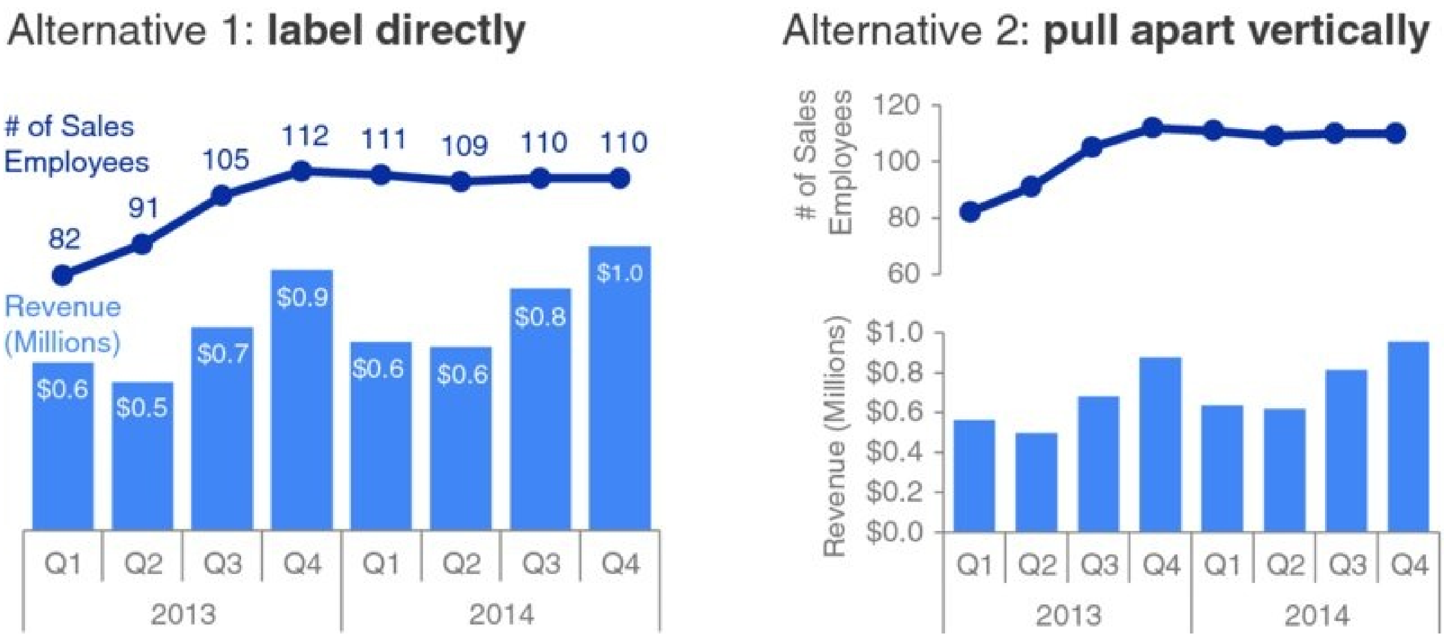


Figure 2.27 Strategies for avoiding a secondary *y*-axis

A third potential option not shown here is to link the axis to the data to be read against it through the use of color. For example, in the original graph depicted in Figure 2.26, I could write the left *y*-axis title "Revenue" in blue and keep the revenue bars blue while at the same time writing the right *y*-axis title "# of Sales Employees" in orange and making the line graph orange to tie these together visually. I don't recommend this approach because color can typically be used more strategically. We'll spend a lot more time discussing color in Chapter 4.

It is also worth noting that when you display two datasets against the same axis, it can imply a relationship that may or may not exist. This is something to be aware of when determining whether this is an appropriate approach in the first place.

When you're facing a secondary *y*-axis challenge and considering which alternative shown in Figure 2.27 will better meet your needs, think about the level of specificity you need. Alternative 1, where each data point is labeled explicitly, puts more attention on the specific numbers. Alternative 2, where the axes are shown at the left, puts more focus on the overarching trends. In general, avoid a secondary *y*-axis and instead employ one of these alternate approaches.

In closing

In this chapter, we've explored the types of visual displays I find myself using most. There will be use cases for other types of visuals, but what we've covered here should meet the majority of everyday needs.

In many cases, there isn't a single correct visual display; rather, often there are different types of visuals that could meet a given need. Drawing from the previous chapter on context, most important is to have that need clearly articulated: *What do you need your audience to know?* Then choose a visual display that will enable you to make this clear.

If you're wondering *What is the right graph for my situation?*, the answer is always the same: whatever will be easiest for your audience to read. There is an easy way to test this, which is to create your visual and show it to a friend or colleague. Have them articulate the following as they process the information: where they focus, what they see, what observations they make, what questions they have. This will help you assess whether your visual is hitting the mark, or in the case where it isn't, help you know where to concentrate your changes.

You now know the second lesson of storytelling with data: how to **choose an appropriate visual display**.