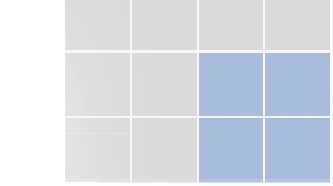
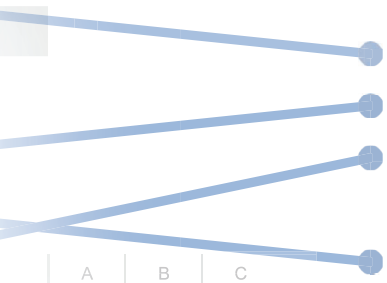
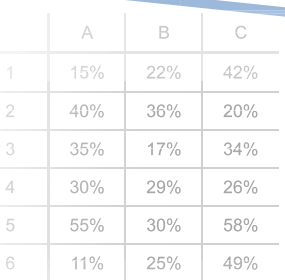


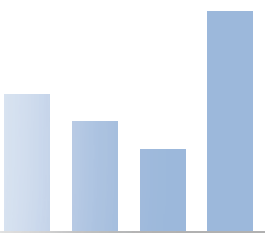
cole nussbaumer knaflic



storytelling with data



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



	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%

a data visualization guide for business professionals

WILEY

chapter one

the importance of context

This may sound counterintuitive, but success in data visualization does not start with data visualization. Rather, before you begin down the path of creating a data visualization or communication, attention and time should be paid to understanding the **context** for the need to communicate. In this chapter, we will focus on understanding the important components of context and discuss some strategies to help set you up for success when it comes to communicating visually with data.

Exploratory vs. explanatory analysis

Before we get into the specifics of context, there is one important distinction to draw, between *exploratory* and *explanatory* analysis. Exploratory analysis is what you do to understand the data and figure out what might be noteworthy or interesting to highlight to others. When we do exploratory analysis, it's like hunting for pearls in oysters.

We might have to open 100 oysters (test 100 different hypotheses or look at the data in 100 different ways) to find perhaps two pearls. When we're at the point of communicating our analysis to our audience, we really want to be in the *explanatory* space, meaning you have a specific thing you want to explain, a specific story you want to tell—probably about those two pearls.

Too often, people err and think it's OK to show exploratory analysis (simply present the data, all 100 oysters) when they should be showing explanatory (taking the time to turn the data into information that can be consumed by an audience: the two pearls). It is an understandable mistake. After undertaking an entire analysis, it can be tempting to want to show your audience *everything*, as evidence of all of the work you did and the robustness of the analysis. Resist this urge. You are making your audience reopen all of the oysters! Concentrate on the pearls, the information your audience needs to know.

Here, we focus on **explanatory** analysis and communication.

Recommended reading

For those interested in learning more about *exploratory* analysis, check out Nathan Yau's book, *Data Points.* Yau focuses on data visualization as a medium, rather than a tool, and spends a good portion of the book discussing the data itself and strategies for exploring and analyzing it.

Who, what, and how

When it comes to explanatory analysis, there are a few things to think about and be extremely clear on before visualizing any data or creating content. First, *To whom are you communicating?* It is important to have a good understanding of who your audience is and how they perceive you. This can help you to identify common ground that will

help you ensure they hear your message. Second, *What do you want your audience to know or do?* You should be clear how you want your audience to act and take into account how you will communicate to them and the overall tone that you want to set for your communication.

It's only after you can concisely answer these first two questions that you're ready to move forward with the third: *How can you use data to help make your point?*

Let's look at the context of who, what, and how in a little more detail.

Who

Your audience

The more specific you can be about who your audience is, the better position you will be in for successful communication. Avoid general audiences, such as "internal and external stakeholders" or "anyone who might be interested"—by trying to communicate to too many different people with disparate needs at once, you put yourself in a position where you can't communicate to any one of them as effectively as you could if you narrowed your target audience. Sometimes this means creating different communications for different audiences. Identifying the decision maker is one way of narrowing your audience. The more you know about your audience, the better positioned you'll be to understand how to resonate with them and form a communication that will meet their needs and yours.

You

It's also helpful to think about the relationship that you have with your audience and how you expect that they will perceive you. Will you be encountering each other for the first time through this communication, or do you have an established relationship? Do they already trust you as an expert, or do you need to work to establish credibility? These are important considerations when it comes to

determining how to structure your communication and whether and when to use data, and may impact the order and flow of the overall story you aim to tell.

Recommended reading

In Nancy Duarte's book *Resonate*, she recommends thinking of your audience as the hero and outlines specific strategies for getting to know your audience, segmenting your audience, and creating common ground. A free multimedia version of *Resonate* is available at duarte.com.

What

Action

What do you need your audience to know or do? This is the point where you think through how to make what you communicate relevant for your audience and form a clear understanding of why they should care about what you say. You should always want your audience to know or do something. If you can't concisely articulate that, you should revisit whether you need to communicate in the first place.

This can be an uncomfortable space for many. Often, this discomfort seems to be driven by the belief that the audience knows better than the presenter and therefore should choose whether and how to act on the information presented. This assumption is false. If you are the one analyzing and communicating the data, *you* likely know it best—*you* are a subject matter expert. This puts you in a unique position to interpret the data and help lead people to understanding and action. In general, those communicating with data need to take a more confident stance when it comes to making specific observations and recommendations based on their analysis. This will feel outside of your comfort zone if you haven't been routinely doing it.

Start doing it now—it will get easier with time. And know that even if you highlight or recommend the wrong thing, it prompts the right sort of conversation focused on action.

When it really isn't appropriate to recommend an action explicitly, encourage discussion toward one. Suggesting possible next steps can be a great way to get the conversation going because it gives your audience something to react to rather than starting with a blank slate. If you simply present data, it's easy for your audience to say, "Oh, that's interesting," and move on to the next thing. But if you ask for action, your audience has to make a decision whether to comply or not. This elicits a more productive reaction from your audience, which can lead to a more productive conversation—one that might never have been started if you hadn't recommended the action in the first place.

Prompting action

Here are some action words to help act as thought starters as you determine what you are asking of your audience:

accept | agree | begin | believe | change | collaborate | commence | create | defend | desire | differentiate | do | empathize | empower | encourage | engage | establish | examine | facilitate | familiarize | form | implement | include | influence | invest | invigorate | know | learn | like | persuade | plan | promote | pursue | recommend | receive | remember | report | respond | secure | support | simplify | start | try | understand | validate

Mechanism

How will you communicate to your audience? The method you will use to communicate to your audience has implications on a number of factors, including the amount of control you will have over how the audience takes in the information and the level of detail that

needs to be explicit. We can think of the communication mechanism along a continuum, with live presentation at the left and a written document or email at the right, as shown in Figure 1.1. Consider the level of control you have over how the information is consumed as well as the amount of detail needed at either end of the spectrum.

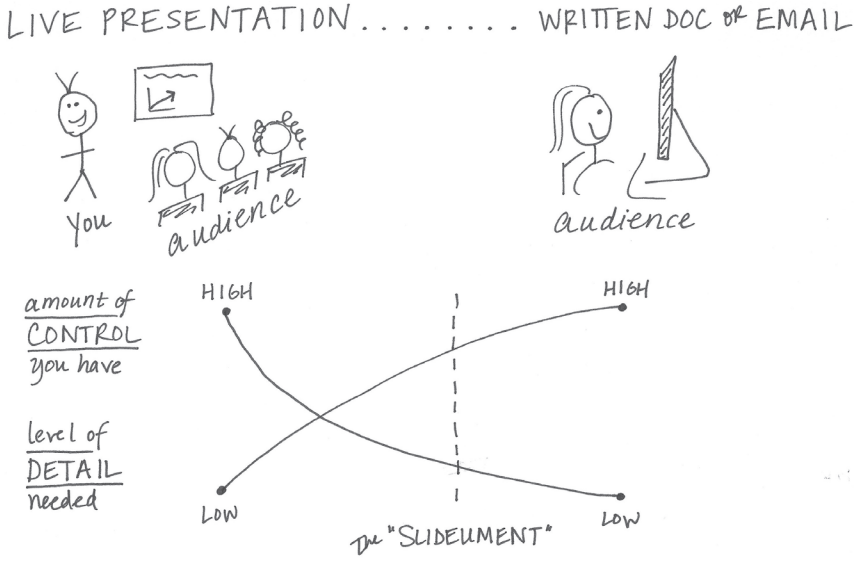


FIGURE 1.1 Communication mechanism continuum

At the left, with a **live presentation**, you (the presenter) are in full control. You determine what the audience sees and when they see it. You can respond to visual cues to speed up, slow down, or go into a particular point in more or less detail. Not all of the detail needs to be directly in the communication (the presentation or slide deck), because you, the subject matter expert, are there to answer any questions that arise over the course of the presentation and should be able and prepared to do so irrespective of whether that detail is in the presentation itself.

For live presentations, practice makes perfect

Do not use your slides as your teleprompter! If you find yourself reading each slide out loud during a presentation, you are using them as one. This creates a painful audience experience. You have to know your content to give a good presentation and this means practice, practice, and more practice! Keep your slides sparse, and only put things on them that help reinforce what you will say. Your slides can remind you of the next topic, but shouldn't act as your speaking notes.

Here are a few tips for getting comfortable with your material as you prepare for your presentation:

- Write out speaking notes with the important points you want to make with each slide.
- Practice what you want to say out loud to yourself: this ignites a different part of the brain to help you remember your talking points. It also forces you to articulate the transitions between slides that sometimes trip up presenters.
- Give a mock presentation to a friend or colleague.

At the right side of the spectrum, with a **written document or email**, you (the creator of the document or email) have less control. In this case, the audience is in control of how they consume the information. The level of detail that is needed here is typically higher because you aren't there to see and respond to your audience's cues. Rather, the document will need to directly address more of the potential questions.

In an ideal world, the work product for the two sides of this continuum would be totally different—sparse slides for a live presentation (since you're there to explain anything in more detail as needed), and

denser documents when the audience is left to consume on their own. But in reality—due to time and other constraints—it is often the same product that is created to try to meet both of these needs. This gives rise to the **slideument**, a single document that's meant to solve both of these needs. This poses some challenges because of the diverse needs it is meant to satisfy, but we'll look at strategies for addressing and overcoming these challenges later in the book.

At this point at the onset of the communication process, it is important to identify the primary communication vehicle you'll be leveraging: live presentation, written document, or something else. Considerations on how much control you'll have over how your audience consumes the information and the level of detail needed will become very important once you start to generate content.

Tone

What tone do you want your communication to set? Another important consideration is the tone you want your communication to convey to your audience. Are you celebrating a success? Trying to light a fire to drive action? Is the topic lighthearted or serious? The tone you desire for your communication will have implications on the design choices that we will discuss in future chapters. For now, think about and specify the general tone that you want to establish when you set out on the data visualization path.

How

Finally—and only after we can clearly articulate who our audience is and what we need them to know or do—we can turn to the data and ask the question: *What data is available that will help make my point?* Data becomes supporting evidence of the story you will build and tell. We'll discuss much more on how to present this data visually in subsequent chapters.

Ignore the nonsupporting data?

You might assume that showing only the data that backs up your point and ignoring the rest will make for a stronger case. I do not recommend this. Beyond being misleading by painting a one-sided story, this is very risky. A discerning audience will poke holes in a story that doesn't hold up or data that shows one aspect but ignores the rest. The right amount of context and supporting and opposing data will vary depending on the situation, the level of trust you have with your audience, and other factors.

Who, what, and how: illustrated by example

Let's consider a specific example to illustrate these concepts. Imagine you are a fourth grade science teacher. You just wrapped up an experimental pilot summer learning program on science that was aimed at giving kids exposure to the unpopular subject. You surveyed the children at the onset and end of the program to understand whether and how perceptions toward science changed. You believe the data shows a great success story. You would like to continue to offer the summer learning program on science going forward.

Let's start with the *who* by identifying our audience. There are a number of different potential audiences who might be interested in this information: parents of students who participated in the program, parents of prospective future participants, the future potential participants themselves, other teachers who might be interested in doing something similar, or the budget committee that controls the funding you need to continue the program. You can imagine how the story you would tell to each of these audiences might differ. The emphasis might change. The call to action would be different for the different groups. The data you would show (or the decision to show data at all) could be different for the various audiences. You can imagine how, if we crafted a single communication meant to address

all of these disparate audiences' needs, it would likely not exactly meet any single audience's need. This illustrates the importance of identifying a *specific* audience and crafting a communication with that specific audience in mind.

Let's assume in this case the audience we want to communicate to is the budget committee, which controls the funding we need to continue the program.

Now that we have answered the question of *who*, the *what* becomes easier to identify and articulate. If we're addressing the budget committee, a likely focus would be to demonstrate the success of the program and ask for a specific funding amount to continue to offer it. After identifying who our audience is and what we need from them, next we can think about the data we have available that will act as evidence of the story we want to tell. We can leverage the data collected via survey at the onset and end of the program to illustrate the increase in positive perceptions of science before and after the pilot summer learning program.

This won't be the last time we'll consider this example. Let's recap who we have identified as our audience, what we need them to know and do, and the data that will help us make our case:

Who: The budget committee that can approve funding for continuation of the summer learning program.

What: The summer learning program on science was a success; please approve budget of $X to continue.

How: Illustrate success with data collected through the survey conducted before and after the pilot program.

Consulting for context: questions to ask

Often, the communication or deliverable you are creating is at the request of someone else: a client, a stakeholder, or your boss. This means you may not have all of the context and might need to consult

with the requester to fully understand the situation. There is sometimes additional context in the head of this requester that they may assume is known or not think to say out loud. Following are some questions you can use as you work to tease out this information. If you're on the requesting side of the communication and asking your support team to build a communication, think about answering these questions for them up front:

- What background information is relevant or essential?
- Who is the audience or decision maker? What do we know about them?
- What biases does our audience have that might make them supportive of or resistant to our message?
- What data is available that would strengthen our case? Is our audience familiar with this data, or is it new?
- Where are the risks: what factors could weaken our case and do we need to proactively address them?
- What would a successful outcome look like?
- If you only had a limited amount of time or a single sentence to tell your audience what they need to know, what would you say?

In particular, I find that these last two questions can lead to insightful conversation. Knowing what the desired outcome is before you start preparing the communication is critical for structuring it well. Putting a significant constraint on the message (a short amount of time or a single sentence) can help you to boil the overall communication down to the single, most important message. To that end, there are a couple of concepts I recommend knowing and employing: the 3-minute story and the Big Idea.

The 3-minute story & Big Idea

The idea behind each of these concepts is that you are able to boil the "so-what" down to a paragraph and, ultimately, to a single, concise statement. You have to really know your stuff—know what the most important pieces are as well as what *isn't* essential in the

most stripped-down version. While it sounds easy, being concise is often more challenging than being verbose. Mathematician and philosopher Blaise Pascal recognized this in his native French, with a statement that translates roughly to "I would have written a shorter letter, but I did not have the time" (a sentiment often attributed to Mark Twain).

3-minute story

The 3-minute story is exactly that: if you had only three minutes to tell your audience what they need to know, what would you say? This is a great way to ensure you are clear on and can articulate the story you want to tell. Being able to do this removes you from dependence on your slides or visuals for a presentation. This is useful in the situation where your boss asks you what you're working on or if you find yourself in an elevator with one of your stakeholders and want to give her the quick rundown. Or if your half-hour on the agenda gets shortened to ten minutes, or to five. If you know exactly what it is you want to communicate, you can make it fit the time slot you're given, even if it isn't the one for which you are prepared.

Big Idea

The Big Idea boils the so-what down even further: to a single sentence. This is a concept that Nancy Duarte discusses in her book, *Resonate* (2010). She says the Big Idea has three components:

1. It must articulate your unique point of view;
2. It must convey what's at stake; and
3. It must be a complete sentence.

Let's consider an illustrative 3-minute story and Big Idea, leveraging the summer learning program on science example that was introduced previously.

3-minute story: *A group of us in the science department were brainstorming about how to resolve an ongoing issue we have with incoming fourth-graders. It seems that when kids get to their first science class, they come in with this attitude that it's going to be difficult and they aren't going to like it. It takes a good amount of time at the beginning of the school year to get beyond that. So we thought, what if we try to give kids exposure to science sooner? Can we influence their perception? We piloted a learning program last summer aimed at doing just that. We invited elementary school students and ended up with a large group of second- and third-graders. Our goal was to give them earlier exposure to science in hopes of forming positive perception. To test whether we were successful, we surveyed the students before and after the program. We found that, going into the program, the biggest segment of students, 40%, felt just "OK" about science, whereas after the program, most of these shifted into positive perceptions, with nearly 70% of total students expressing some level of interest toward science. We feel that this demonstrates the success of the program and that we should not only continue to offer it, but also to expand our reach with it going forward.*

Big Idea: *The pilot summer learning program was successful at improving students' perceptions of science and, because of this success, we recommend continuing to offer it going forward; please approve our budget for this program.*

When you've articulated your story this clearly and concisely, creating content for your communication becomes much easier. Let's shift gears now and discuss a specific strategy when it comes to planning content: storyboarding.

Storyboarding

Storyboarding is perhaps the single most important thing you can do up front to ensure the communication you craft is on point. The storyboard establishes a structure for your communication. It is a visual outline of the content you plan to create. It can be subject to

change as you work through the details, but establishing a structure early on will set you up for success. When you can (and as makes sense), get acceptance from your client or stakeholder at this step. It will help ensure that what you're planning is in line with the need.

When it comes to storyboarding, the biggest piece of advice I have is this: don't start with presentation software. It is too easy to go into slide-generating mode without thinking about how the pieces fit together and end up with a massive presentation deck that says nothing effectively. Additionally, as we start creating content via our computer, something happens that causes us to form an attachment to it. This attachment can be such that, even if we know what we've created isn't exactly on the mark or should be changed or eliminated, we are sometimes resistant to doing so because of the work we've already put in to get it to where it is.

Avoid this unnecessary attachment (and work!) by starting low tech. Use a whiteboard, Post-it notes, or plain paper. It's much easier to put a line through an idea on a piece of paper or recycle a Post-it note without feeling the same sense of loss as when you cut something you've spent time creating with your computer. I like using Post-it notes when I storyboard because you can rearrange (and add and remove) the pieces easily to explore different narrative flows.

If we storyboard our communication for the summer learning program on science, it might look something like Figure 1.2.

Note that in this example storyboard, the Big Idea is at the end, in the recommendation. Perhaps we'd want to consider leading with that to ensure that our audience doesn't miss the main point and to help set up why we are communicating to them and why they should care in the first place. We'll discuss additional considerations related to the narrative order and flow in Chapter 7.

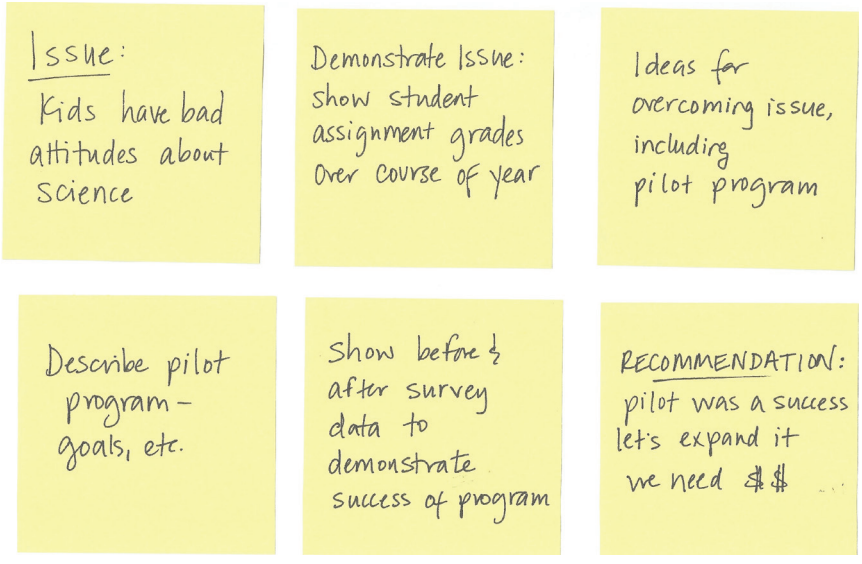


FIGURE 1.2 Example storyboard

In closing

When it comes to explanatory analysis, being able to concisely articulate exactly who you want to communicate to and what you want to convey before you start to build content reduces iterations and helps ensure that the communication you build meets the intended purpose. Understanding and employing concepts like the 3-minute story, the Big Idea, and storyboarding will enable you to clearly and succinctly tell your story and identify the desired flow.

While pausing before actually building the communication might feel like it's a step that slows you down, in fact it helps ensure that you have a solid understanding of what you want to do before you start creating content, which will save you time down the road.

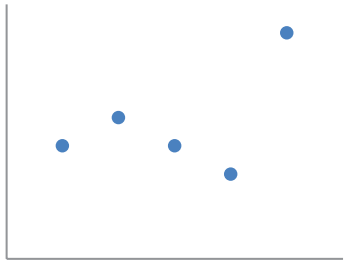
With that, consider your first lesson learned. You now **understand the importance of context**.

chapter two

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.

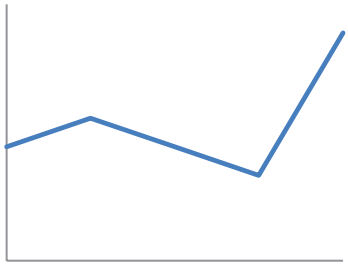
Simple text



Scatterplot

	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%

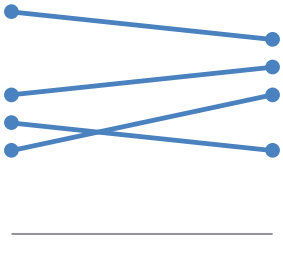
Table



Line

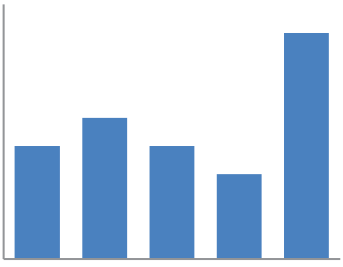
	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

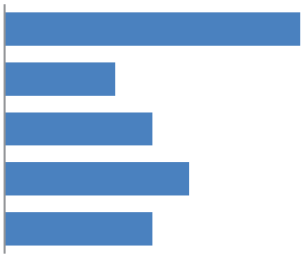


Slopegraph

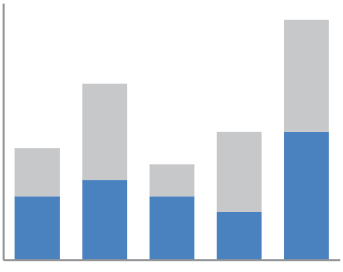
FIGURE 2.1 The visuals I use most



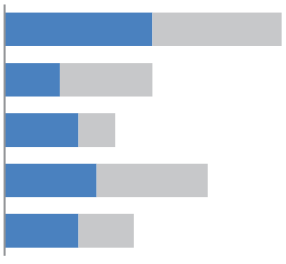
Vertical bar



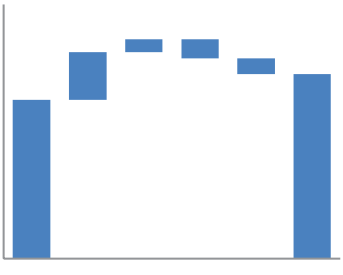
Horizontal bar



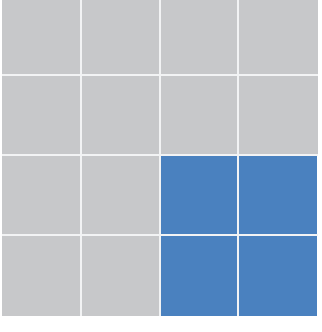
Stacked vertical bar



Stacked horizontal bar



Waterfall



Square area

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.

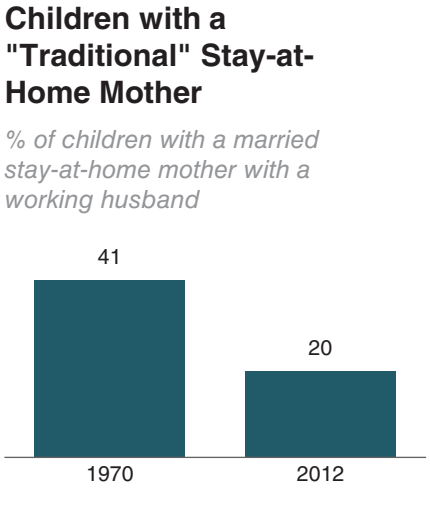


FIGURE 2.2 Stay-at-home moms original graph

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.

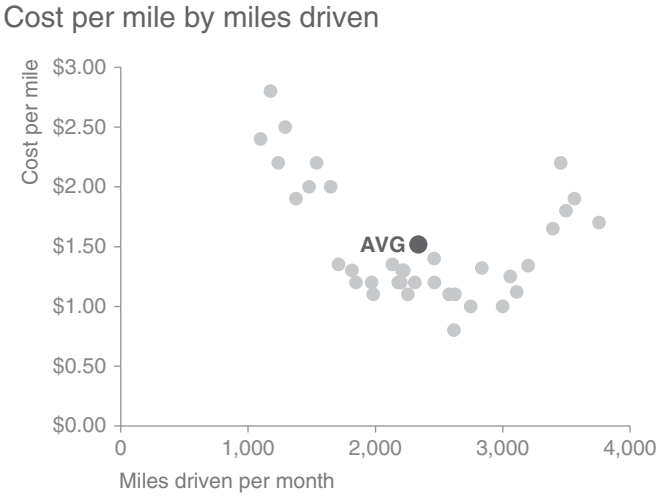


FIGURE 2.6 Scatterplot

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.

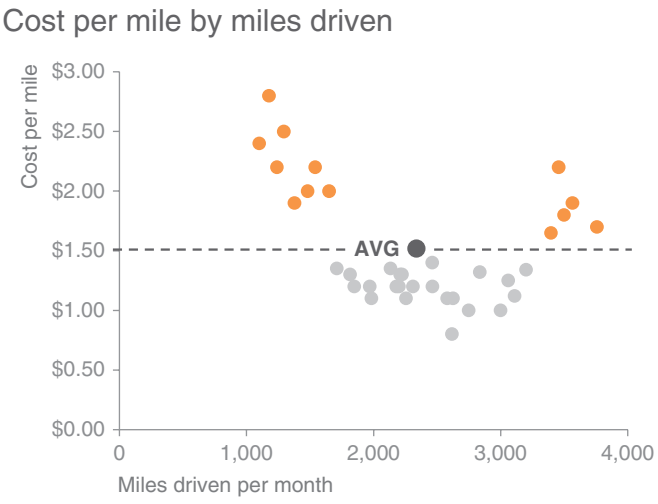


FIGURE 2.7 Modified scatterplot

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.

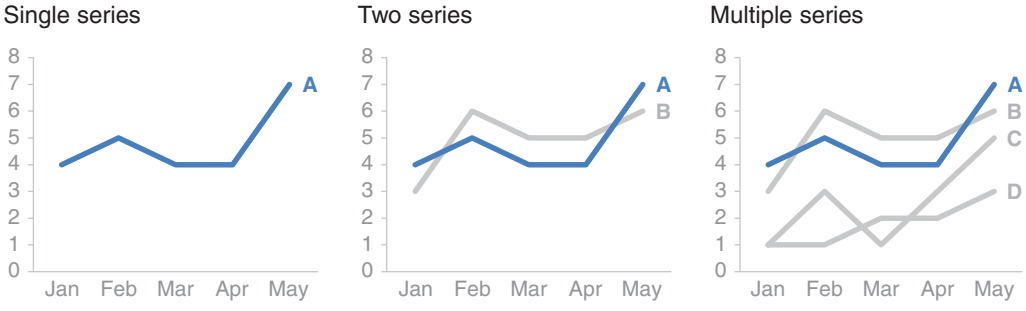


FIGURE 2.8 Line graphs

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.

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.

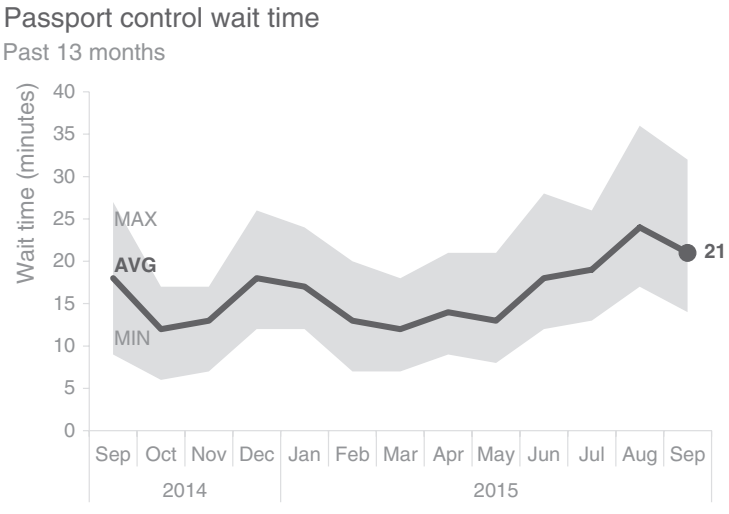


FIGURE 2.9 Showing average within a range in a line graph

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.

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.

FIGURE 2.10 Slopegraph

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.

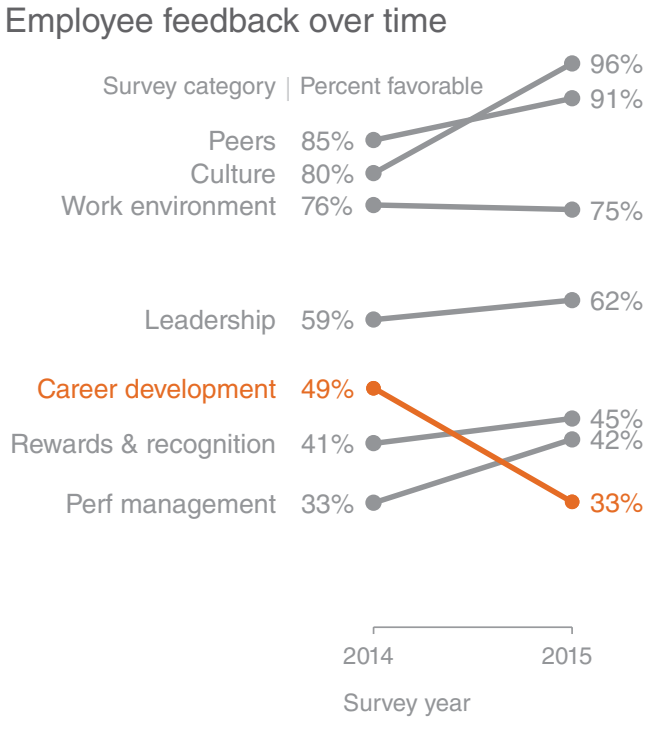


FIGURE 2.11 Modified slopegraph

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.

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.

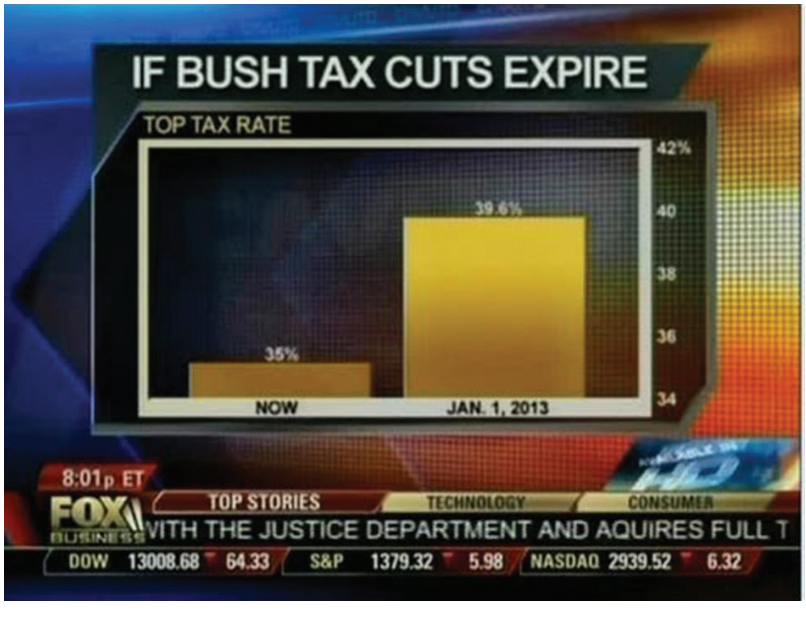


FIGURE 2.12 Fox News bar chart

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.

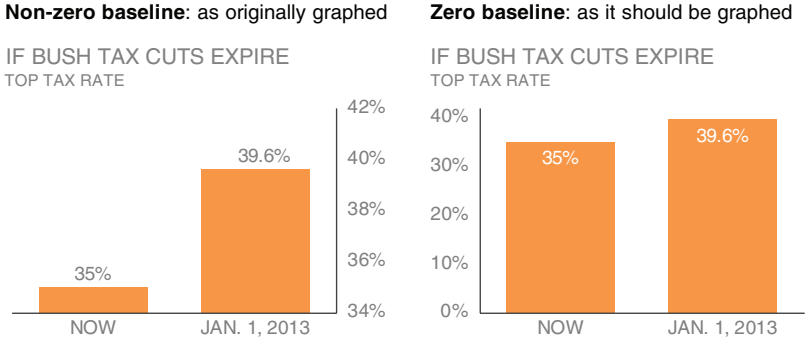


FIGURE 2.13 Bar charts must have a zero baseline

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.

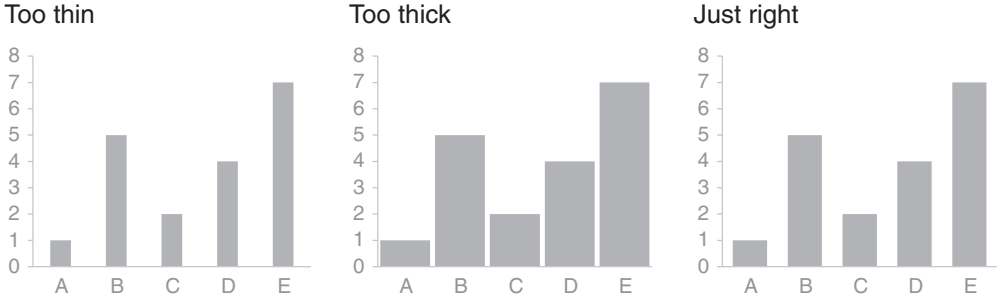


FIGURE 2.14 Bar width

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.

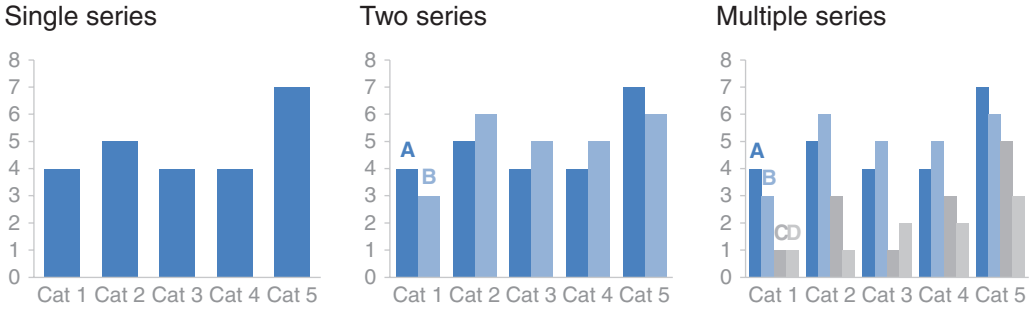


FIGURE 2.15 Bar charts

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.

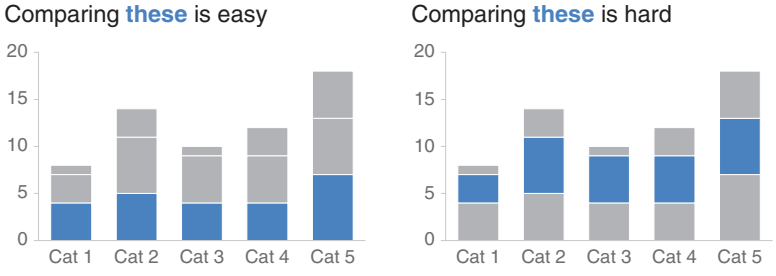


FIGURE 2.16 Comparing series with stacked bar charts

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.

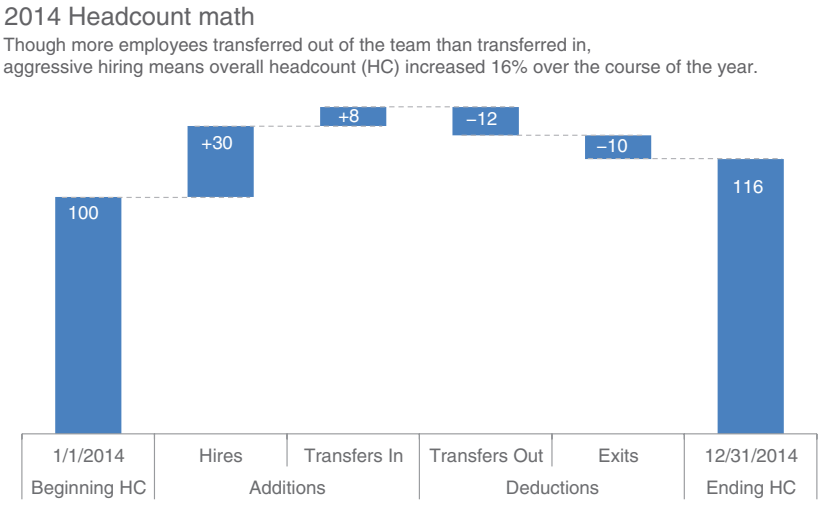


FIGURE 2.17 Waterfall chart

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).

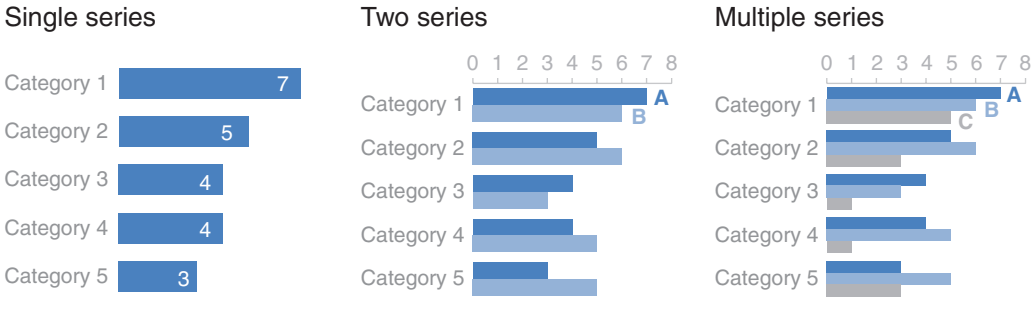


FIGURE 2.18 Horizontal bar charts

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.

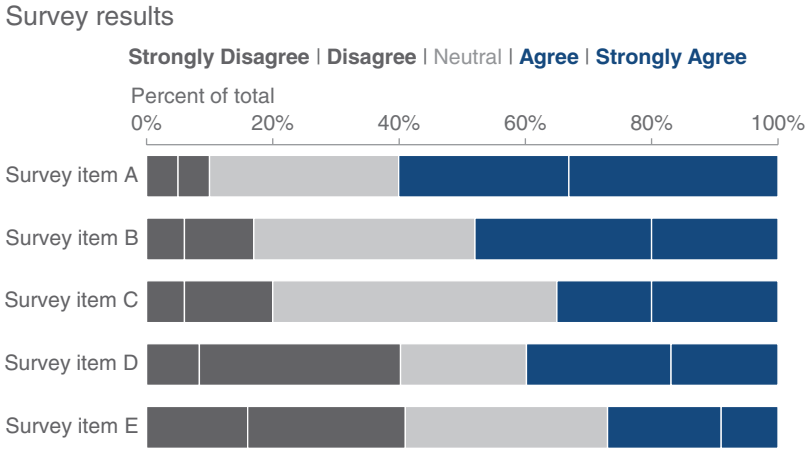


FIGURE 2.19 100% stacked horizontal bar chart

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?

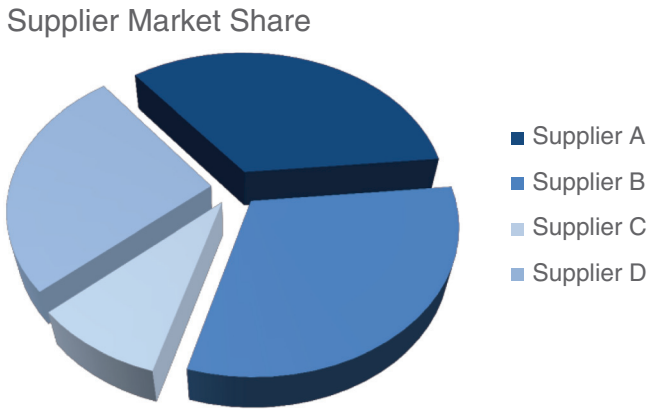


FIGURE 2.21 Pie chart

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.

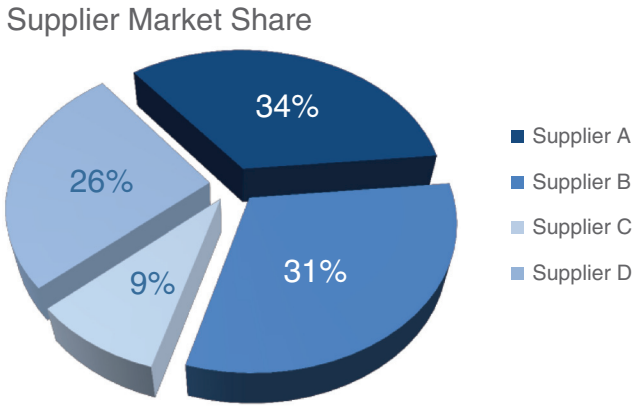


FIGURE 2.22 Pie chart with labeled segments

"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.

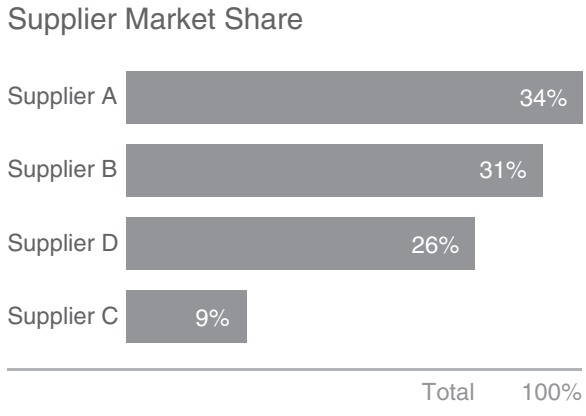


FIGURE 2.23 An alternative to the pie chart

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.

The donut chart

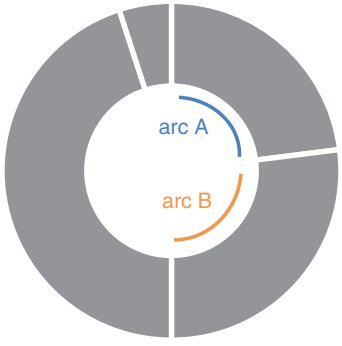


FIGURE 2.24 Donut chart

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?

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.

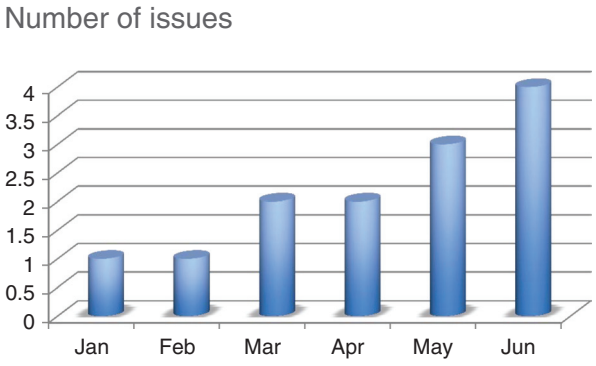


FIGURE 2.25 3D column chart

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.

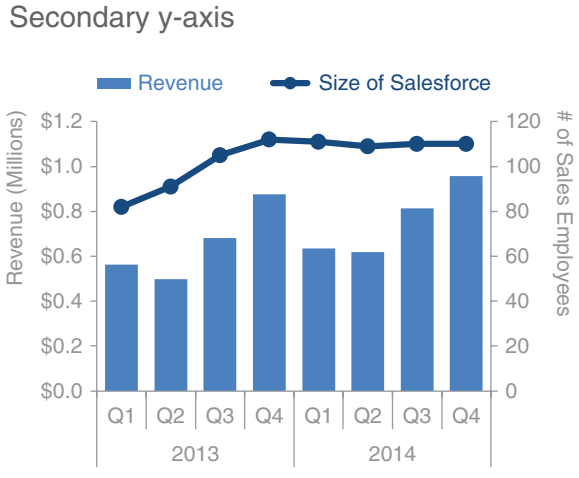


FIGURE 2.26 Secondary *y*-axis

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.

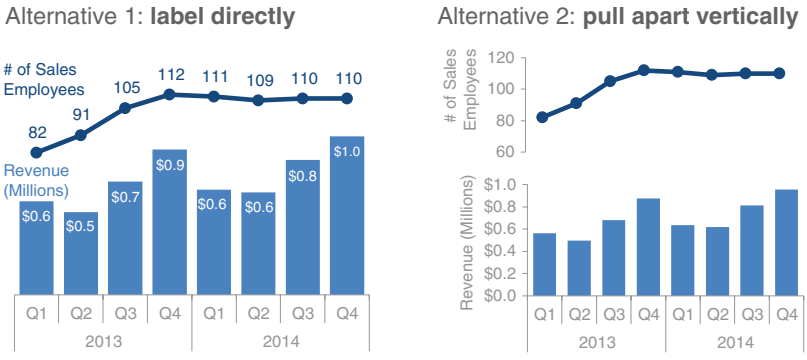


FIGURE 2.27 Strategies for avoiding a secondary *y*-axis

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**.

chapter three

clutter is your enemy!

Picture a blank page or a blank screen: every single element you add to that page or screen takes up cognitive load on the part of your audience—in other words, takes them brain power to process. Therefore, we want to take a discerning look at the visual elements that we allow into our communications. In general, identify anything that isn't adding informative value—or isn't adding *enough* informative value to make up for its presence—and remove those things. Identifying and eliminating such clutter is the focus of this chapter.

Cognitive load

You have felt the burden of cognitive load before. Perhaps you were sitting in a conference room as the person leading the meeting was flipping through their projected slides and they paused on one that looked overwhelmingly busy and complicated. Yikes, did you say "ugh" out loud, or was that just in your head? Or maybe you were reading through a report or the newspaper, and a graph caught your eye just long enough for you to think, "this looks interesting

but I have no idea what I'm meant to get out of it"—and rather than spend more time to decipher it, you turned the page.

In both of these instances, what you've experienced is excessive or extraneous cognitive load.

We experience cognitive load *anytime* we take in information. Cognitive load can be thought of as the mental effort that's required to learn new information. When we ask a computer to do work, we are relying on the computer's processing power. When we ask our audience to do work, we are leveraging their mental processing power. This is cognitive load. Humans' brains have a finite amount of this mental processing power. As designers of information, we want to be smart about how we use our audience's brain power. The preceding examples point to extraneous cognitive load: processing that takes up mental resources but doesn't help the audience understand the information. This is something we want to avoid.

The data-ink or signal-to-noise ratio

A number of concepts have been introduced over time in an effort to explain and help provide guidance for reducing the cognitive load we push to our audience through our visual communications. In his book *The Visual Display of Quantitative Information*, Edward Tufte refers to maximizing the data-ink ratio, saying "the larger the share of a graphic's ink devoted to data, the better (other relevant matters being equal)." This can also be referred to as maximizing the signal-to-noise ratio (see Nancy Duarte's book *Resonate*), where the signal is the information we want to communicate, and the noise are those elements that either don't add to, or in some cases detract from, the message we are trying to impart to our audience.

What matters most when it comes to our visual communications is the *perceived* cognitive load on the part of our audience: how hard they believe they are going to have to work to get the information out of your communication. This is a decision they likely reach without giving it much (if any) conscious thought, and yet it can make the difference between getting your message across or not.

In general, think about minimizing the perceived cognitive load (to the extent that is reasonable and still allows you to get the information across) for your audience.

Clutter

One culprit that can contribute to excessive or extraneous cognitive load is something I refer to simply as **clutter**. These are visual elements that take up space but don't increase understanding. We'll take a more specific look at exactly what elements can be considered clutter soon, but in the meantime I want to talk generally about why clutter is a bad thing.

There is a simple reason we should aim to reduce clutter: because it makes our visuals appear more complicated than necessary.

Perhaps without explicitly recognizing it, the presence of clutter in our visual communications can cause a less-than-ideal—or worse—uncomfortable user experience for our audience (this is that "ugh" moment I referred to at the beginning of this chapter). Clutter can make something feel more complicated than it actually is. When our visuals feel complicated, we run the risk of our audience deciding they don't want to take the time to understand what we're showing, at which point we've lost our ability to communicate with them. This is not a good thing.

Gestalt principles of visual perception

When it comes to identifying which elements in our visuals are signal (the information we want to communicate) and which might be noise (clutter), consider the **Gestalt Principles of Visual Perception**. The Gestalt School of Psychology set out in the early 1900s to understand how individuals perceive order in the world around them. What they came away with are the principles of visual perception still accepted today that define how people interact with and create order out of visual stimuli.

We'll discuss six principles here: proximity, similarity, enclosure, closure, continuity, and connection. For each, I'll show an example of the principle applied to a table or graph.

Proximity

We tend to think of objects that are physically close together as belonging to part of a group. The proximity principle is demonstrated in Figure 3.1: you naturally see the dots as three distinct groups because of their relative proximity to each other.

FIGURE 3.1 Gestalt principle of proximity

We can leverage this way that people see in table design. In Figure 3.2, simply by virtue of differentiating the spacing between the dots, your eyes are drawn either down the columns in the first case or across the rows in the second case.

FIGURE 3.2 You see columns and rows, simply due to dot spacing

Similarity

Objects that are of similar color, shape, size, or orientation are perceived as related or belonging to part of a group. In Figure 3.3, you naturally associate the blue circles together on the left or the grey squares together on the right.

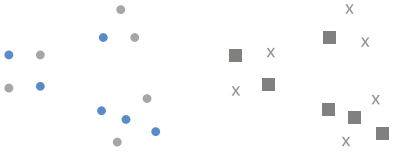


FIGURE 3.3 Gestalt principle of similarity

This can be leveraged in tables to help draw our audience's eyes in the direction we want them to focus. In Figure 3.4, the similarity of color is a cue for our eyes to read across the rows (rather than down the columns). This eliminates the need for additional elements such as borders to help direct our attention.

FIGURE 3.4 You see rows due to similarity of color

Enclosure

We think of objects that are physically enclosed together as belonging to part of a group. It doesn't take a very strong enclosure to do this: light background shading is often enough, as demonstrated in Figure 3.5.

FIGURE 3.5 Gestalt principle of enclosure

One way we can leverage the enclosure principle is to draw a visual distinction within our data, as is done in the graph in Figure 3.6.

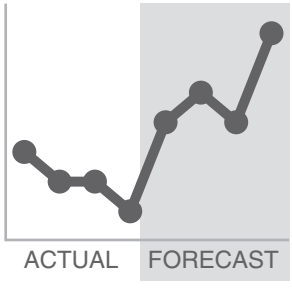


FIGURE 3.6 The shaded area separates the forecast from actual data

Closure

The closure concept says that people like things to be simple and to fit in the constructs that are already in our heads. Because of this, people tend to perceive a set of individual elements as a single, recognizable shape when they can—when parts of a whole are missing, our eyes fill in the gap. For example, the elements in Figure 3.7 will tend to be perceived as a circle first and only after that as individual elements.

FIGURE 3.7 Gestalt principle of closure

It is common for graphing applications (for example, Excel) to have default settings that include elements like chart borders and background shading. The closure principle tells us that these are unnecessary—we can remove them and our graph still appears as a cohesive entity. Bonus: when we take away those unnecessary elements, our data stands out more, as shown in Figure 3.8.

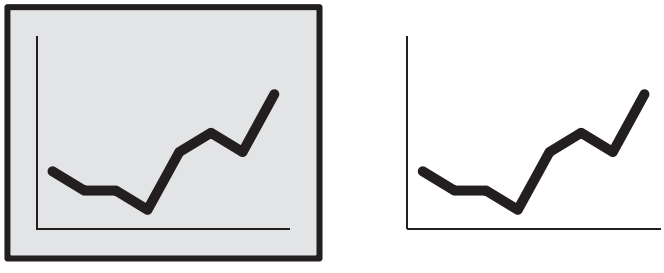


FIGURE 3.8 The graph still appears complete without the border and background shading

Continuity

The principle of continuity is similar to closure: when looking at objects, our eyes seek the smoothest path and naturally create continuity in what we see even where it may not explicitly exist. By way of example, in Figure 3.9, if I take the objects (1) and pull them apart, most people will expect to see what is shown next (2), whereas it could as easily be what is shown after that (3).

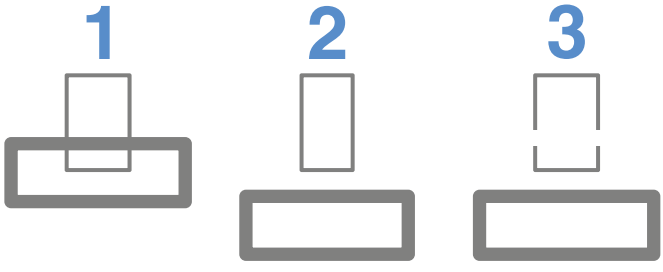


FIGURE 3.9 Gestalt principle of continuity

In the application of this principle, I've removed the vertical *y*-axis line from the graph in Figure 3.10 altogether. Your eyes actually still see that the bars are lined up at the same point because of the consistent white space (the smoothest path) between the labels on the left and the data on the right. As we saw with the closure principle in application, stripping away unnecessary elements allows our data to stand out more.

FIGURE 3.10 Graph with *y*-axis line removed

Connection

The final Gestalt principle we'll focus on is connection. We tend to think of objects that are physically connected as part of a group. The connective property typically has a stronger associative value than similar color, size, or shape. Note when looking at Figure 3.11, your eyes probably pair the shapes connected by lines (rather than similar color, size, or shape): that's the connection principle in action. The connective property *isn't* typically stronger than enclosure, but you can impact this relationship through thickness and darkness of lines to create the desired visual hierarchy (we'll talk more about visual hierarchy when we discuss preattentive attributes in Chapter 4).

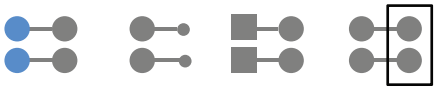


FIGURE 3.11 Gestalt principle of connection

One way that we frequently leverage the connection principle is in line graphs, to help our eyes see order in the data, as shown in Figure 3.12.

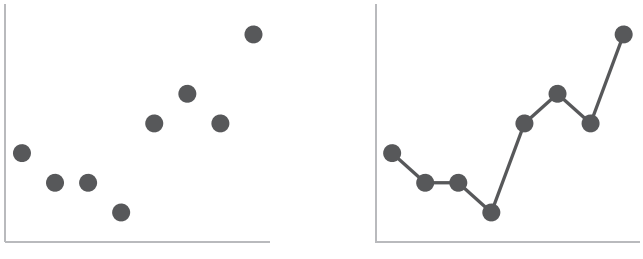


FIGURE 3.12 Lines connect the dots

As you have learned from this brief overview, the Gestalt principles help us understand how people see, which we can use to identify unnecessary elements and ease the processing of our visual communications. We aren't done with them yet. At the end of this chapter, we'll discuss how we can apply some of these principles to a real-world example.

But first, let's shift our focus to a couple of other types of visual clutter.

Lack of visual order

When design is thoughtful, it fades into the background so that your audience doesn't even notice it. When it's not, however, your audience feels the burden. Let's look at an example to understand the impact visual order—and lack thereof—can have on our visual communications.

Take a moment to study Figure 3.13, which summarizes survey feedback about factors considered by nonprofits in vendor selection. Note specifically any observations you may have regarding the arrangement of elements on the page.

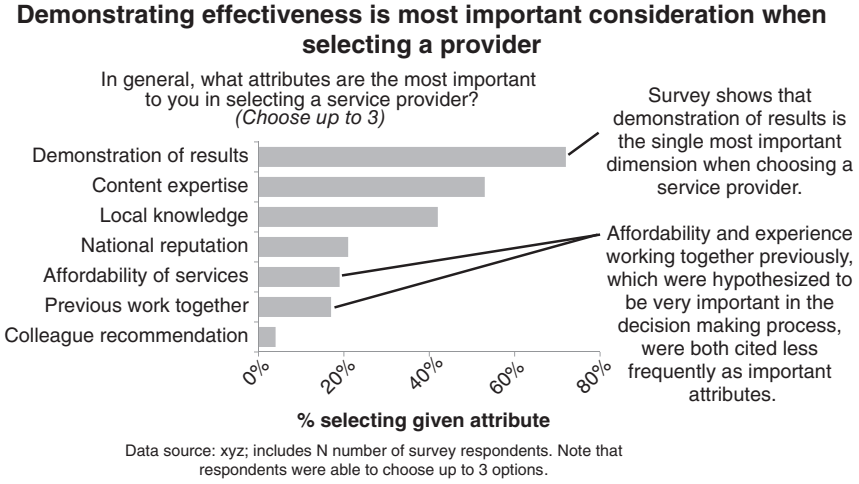


FIGURE 3.13 Summary of survey feedback

As you look over the information, you might be thinking, "this looks pretty good." I'll concede: it's not horrible. On the positive side, the takeaway is clearly outlined, the graph is well ordered and labeled, and key observations are articulated and tied visually to where we're meant to look in the graph. But when it comes to the overall design of the page and placement of elements, I'd have to disagree with any

praise. To me, the aggregate visual feels disorganized and uncomfortable to look at, as if the various components were haphazardly put there without regard for the structure of the overall page.

We can improve this visual markedly by making some relatively minor changes. Take a look at Figure 3.14. The content is exactly the same; only the placement and formatting of elements have been modified.

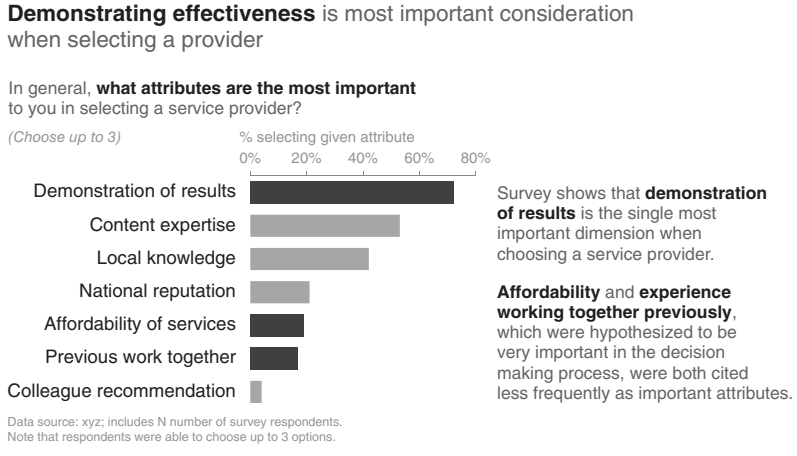


FIGURE 3.14 Revamped summary of survey feedback

Compared to the original visual, the second iteration feels somehow easier. There is order. It is evident that conscious thought was paid to the overarching design and arrangement of components. Specifically, the latter version has been designed with greater attention to alignment and white space. Let's look at each of these in detail.

Alignment

The single change having the biggest impact in the preceding before-and-after example was the shift from center-aligned to left-justified text. In the original version, each block of text on the page is center-aligned. This does not create clean lines either on the left or on the right, which can make even a thoughtful layout appear

sloppy. I tend to avoid center-aligned text for this reason. The decision of whether to left- or right-justify your text should be made in context of the other elements on the page. In general, the goal is to create clean lines (both horizontally and vertically) of elements and white space.

Presentation software tips for aligning elements

To help ensure that your elements line up when you are placing them on a page within your presentation software, turn on the rulers or gridlines that are built into most programs. This will allow you to precisely align your elements to create a cleaner look and feel. The table functionality built into most presentation applications can also be used as a makeshift brute-force method: create a table to give yourself guidelines for the placement of discrete elements. When you have everything lined up exactly like you want it, remove the table or make the table's borders invisible so that all that is left is your perfectly arranged page.

Without other visual cues, your audience will typically start at the top left of the page or screen and will move their eyes in a "z" shape (or multiple "z" shapes, depending on the layout) across the page or screen as they take in information. Because of this, when it comes to tables and graphs, I like to upper-left-most justify the text (title, axis titles, legend). This means the audience will hit the details that tell them how to read the table or graph before they get to the data itself.

As part of our discussion on alignment, let's spend a bit of time on **diagonal components.** In the previous example, the original version (Figure 3.13) had diagonal lines connecting the takeaways to the data and diagonally oriented *x*-axis labels; the former were removed and the latter changed to horizontal orientation in the makeover (Figure 3.14). Generally, diagonal elements such as lines and text should be

avoided. They look messy and, in the case of text, are harder to read than their horizontal counterparts. When it comes to the orientation of text, one study (Wigdor & Balakrishnan, 2005) found that the reading of rotated text 45 degrees in either direction was, on average, 52% slower than reading normally oriented text (text rotated 90 degrees in either direction was 205% slower on average). It is best to avoid diagonal elements on the page.

White space

I've never quite understood this phenomenon, but for some reason, people tend to fear white space on a page. I use "white space" to refer to blank space on the page. If your pages are blue, for example, this would be "blue space"—I'm not sure why they would be blue, but the use of color is a conversation we will have later. Perhaps you've heard this feedback before: "there is still some space left on that page, so let's add something there," or worse, "there is still some space left on that page, so let's add more data." No! Never add data just for the sake of adding data—only add data with a thoughtful and specific purpose in mind!

We need to get more comfortable with white space.

White space in visual communication is as important as *pauses* in public speaking. Perhaps you have sat through a presentation that lacked pauses. It feels something like this: *there is a speaker up in front of you and possibly due to nerves or perhaps because they're trying to get through more material than they should in the allotted time they are speaking a mile a minute and you're wondering how they're even able to breathe you'd like to ask a question but the speaker has already moved on to the next topic and still hasn't paused long enough for you to be able to raise your question.* This is an uncomfortable experience for the audience, similar to the discomfort you may have felt reading through the preceding run-on, unpunctuated sentence.

Now imagine the effect if that same presenter were to make a single bold statement: "Death to pie charts!"

And then pause for a full 15 seconds to let that statement resonate.

Go ahead—say it out loud and then count to 15 slowly.

That's a dramatic pause.

And it got your attention, didn't it?

That is the same powerful effect that white space used strategically can have on our visual communications. The lack of it—like the lack of pauses in a spoken presentation—is simply uncomfortable for our audience. Audience discomfort in response to the design of our visual communications is something we should aim to avoid. White

space can be used strategically to draw attention to the parts of the page that are *not* white space.

When it comes to preserving white space, here are some minimal guidelines. Margins should remain free of text and visuals. Resist the urge to stretch visuals to take up the available space; instead, appropriately size your visuals to their content. Beyond these guidelines, think about how you can use white space strategically for emphasis, as was illustrated with the dramatic pause earlier. If there is one thing that is really important, think about making that *the only thing on the page*. In some cases, this could be a single sentence or even a single number. We'll talk further about using white space strategically and look at an example when we discuss aesthetics in Chapter 5.

Non-strategic use of contrast

Clear contrast can be a signal to our audience, helping them understand where to focus their attention. We will further explore this idea in greater detail in later chapters. The *lack of clear contrast*, on the other hand, can be a form of visual clutter. When discussing the critical value of contrast, there is an analogy I often borrow from Colin Ware (*Information Visualization: Perception for Design*, 2004), who said it's easy to spot a hawk in a sky full of pigeons, but as the variety of birds increases, that hawk becomes harder and harder to pick out. This highlights the importance of the strategic use of contrast in visual design: the more things we make different, the lesser the degree to which any of them stand out. To explain this another way, if there is something really important we want our audience to know or see (the hawk), we should make that *the one thing* that is very different from the rest.

Let's look at an example to further illustrate this concept.

Imagine you work for a U.S. retailer and want to understand how your customers feel about various dimensions of their shopping experience in your store compared to your competitors. You have conducted a survey to collect this information and are now trying

to understand what it tells you. You have created a weighted performance index to summarize each category of interest (the higher the index, the better the performance, and vice versa). Figure 3.15 shows the weighted performance index across categories for your company and five competitors.

Study it for a moment and make note of your thought process as you take in the information.

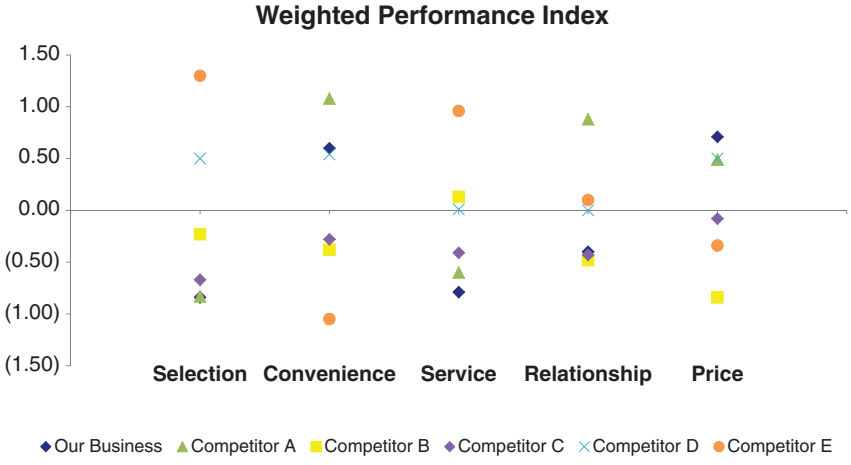


FIGURE 3.15 Original graph

If you had to describe Figure 3.15 in a single word, what would that word be? Words like *busy, confusing,* and perhaps *exhausting* come to mind. There is a lot going on in this graph. So many things are competing for our attention that it is hard to know where to look.

Let's review exactly what we're looking at. As I mentioned, the data graphed is a weighted performance index. You don't need to worry about the details of how this is calculated, but rather understand that this is a summary performance metric that we'd like to compare across various categories (shown across the horizontal *x*-axis: Selection, Convenience, Service, Relationship, and Price) for "Our Business" (depicted by the blue diamond) compared to a number

of competitors (the other colored shapes). A higher index represents better performance, and a lower index means lower performance.

Taking in this information is a slow process, with a lot of back and forth between the legend at the bottom and the data in the graph to decipher what is being conveyed. Even if we are very patient and really want to get information out of this visual, it is nearly impossible because "Our Business" (the blue diamond) is sometimes obscured by other data points, making it so we can't even see the comparison that is most important to make!

This is a case where lack of contrast (as well as some other design issues) makes the information much harder to interpret than it need be.

Consider Figure 3.16, where we use contrast more strategically.

FIGURE 3.16 Revamped graph, using contrast strategically

In the revised graph, I've made a number of changes. First, I chose a horizontal bar chart to depict the information. In doing so, I rescaled all the numbers to be on a positive scale—in the original scatterplot, there were some negative values that complicated the visualization challenge. This change works here since we're more interested in relative differences than absolute values. In this remake, the categories that were previously along the horizontal *x*-axis now run down the vertical *y*-axis. Within each category, the length of the bar shows the summary metric across "Our business" (blue) and the various competitors (grey), with longer bars representing better performance. The decision not to show the actual *x*-axis scale in this case was a deliberate one, which forces the audience to focus on relative differences rather than get caught up in the minutiae of the specific numbers.

With this design, it is easy to see two things quickly:

1. We can let our eyes scan across the blue bars to get a relative sense of how "Our business" is doing across the various categories: we score high on Price and Convenience and lower on Relationship, possibly because we're struggling when it comes to Service and Selection, as evidenced by low scores in these areas.
2. Within a given category, we can compare the blue bar to the grey bars to see how our business is faring relative to competitors: winning compared to the competition on Price, losing on Service and Selection.

Competitors are distinguished from each other based on the order in which they appear (Competitor A always appears directly after the blue bar, Competitor B after that, and so on), which is outlined in the legend at the left. If it were important to be able to quickly identify each competitor, this design doesn't immediately allow for that. But if that is a second- or third-order comparison in terms of priority and isn't the most critical thing, this approach can work well. In the makeover, I've also organized the categories in order of decreasing weighted performance index for "Our business," which provides a construct for our audience to use as they take in the information,

and added a summary metric (relative rank) so it's easy to know quickly how "Our business" ranks in each category in relation to our competition.

Note here how the effective use of contrast (and some other thoughtful design choices) makes it a much faster, easier, and just more comfortable-feeling process to get the information we're after than it was in the original graph.

When redundant details shouldn't be considered clutter

I've seen cases where the title of the visual indicates the values are dollars but the dollar signs aren't included with the actual numbers in the table or graph. For example, a graph titled "Monthly Sales ($USD Millions)" with *y*-axis labels of 10, 20, 30, 40, 50. I find this confusing. Including the "$" sign with each number eases the interpretation of the figures. Your audience doesn't have to remember they are looking at dollars because they are labeled explicitly. There are some elements that should always be retained with numbers, including dollar signs, percent signs, and commas in large numbers.

Decluttering: step-by-step

Now that we have discussed what clutter is, why it is important to eliminate it from our visual communications, and how to recognize it, let's look at a real-world example and examine how the process of identifying and removing clutter improves our visual and the clarity of the story that we're ultimately trying to tell.

Scenario: Imagine that you manage an information technology (IT) team. Your team receives tickets, or technical issues, from employees. In the past year, you've had a couple of people leave and decided

at the time not to replace them. You have heard a rumbling of complaints from the remaining employees about having to "pick up the slack." You've just been asked about your hiring needs for the coming year and are wondering if you should hire a couple more people. First, you want to understand what impact the departure of individuals over the past year has had on your team's overall productivity. You plot the monthly trend of incoming tickets and those processed over the past calendar year. You see that there is some evidence your team's productivity is suffering from being short-staffed and now want to turn the quick-and-dirty visual you created into the basis for your hiring request.

Figure 3.17 shows your original graph.

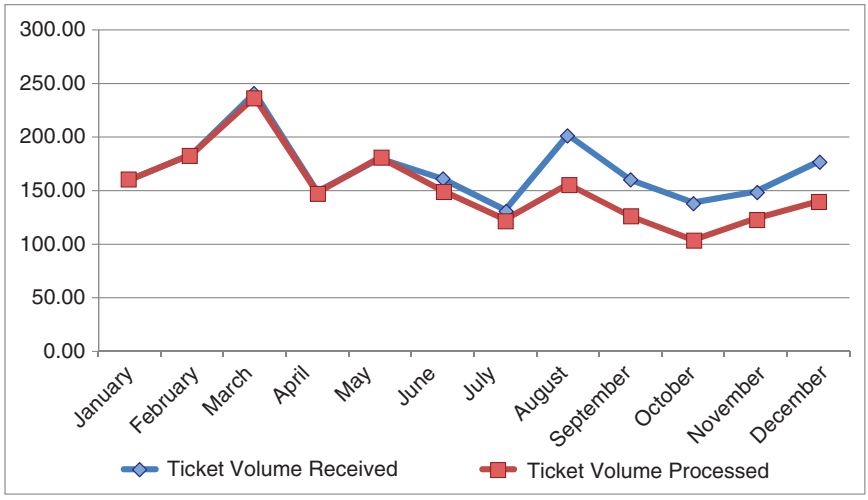


FIGURE 3.17 Original graph

Take another look at this visual with an eye toward clutter. Consider the lessons we've covered on Gestalt principles, alignment, white space, and contrast. What things can we get rid of or change? How many issues can you identify?

I identified six major changes to reduce clutter. Let's discuss each.

1. Remove chart border

Chart borders are usually unnecessary, as we covered in our discussion of the Gestalt principle of closure. Instead, think about using white space to differentiate the visual from other elements on the page as needed.

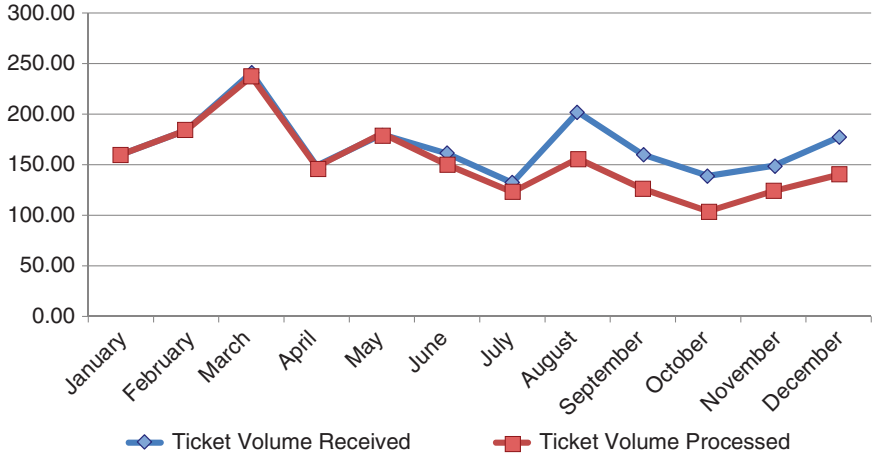


FIGURE 3.18 Remove chart border

2. Remove gridlines

If you think it will be helpful for your audience to trace their finger from the data to the axis, or you feel that your data will be more effectively processed, you can leave the gridlines. But make them thin and use a light color like grey. Do not let them compete visually with your data. When you can, get rid of them altogether: this allows for greater contrast, and your data will stand out more.

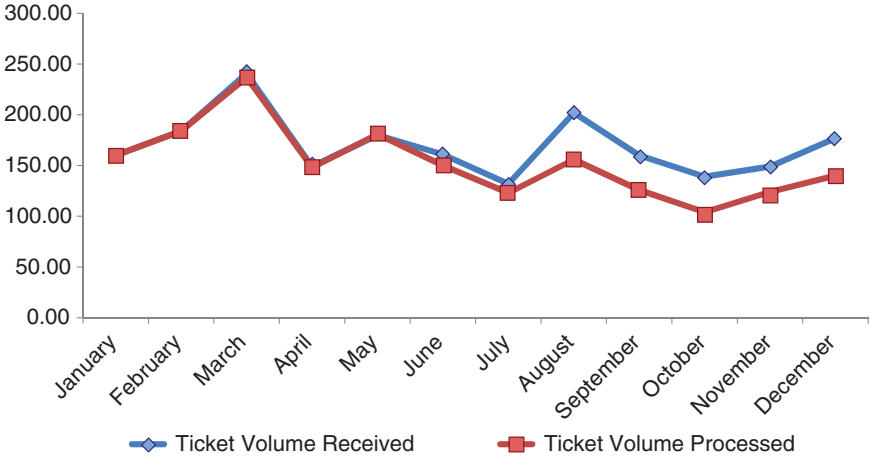


FIGURE 3.19 Remove gridlines

3. Remove data markers

Remember, every single element adds cognitive load on the part of your audience. Here, we're adding cognitive load to process data that is already depicted visually with the lines. This isn't to say that you should never use data markers, but rather use them on purpose and with a purpose, rather than because their inclusion is your graphing application's default.

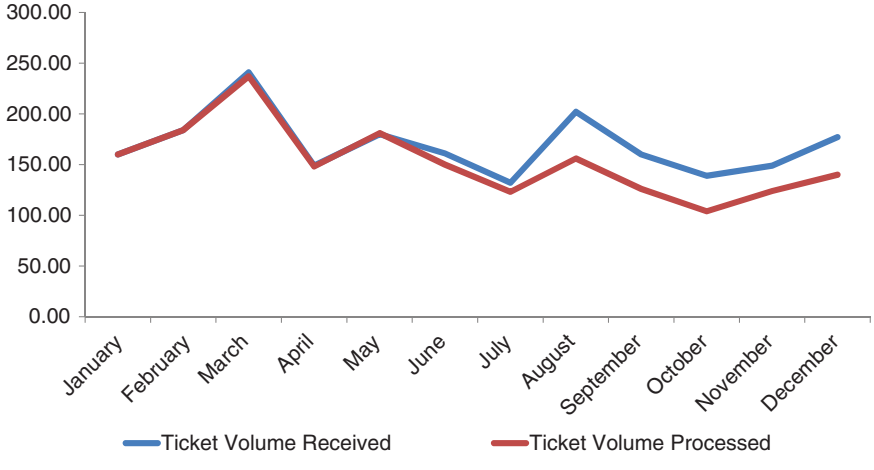


FIGURE 3.20 Remove data markers

4. Clean up axis labels

One of my biggest pet peeves is trailing zeros on *y*-axis labels: they carry no informative value, and yet make the numbers look more complicated than they are! Get rid of them, reducing their unnecessary burden on the audience's cognitive load. We can also abbreviate the months of the year so that they will fit horizontally on the *x*-axis, eliminating the diagonal text.

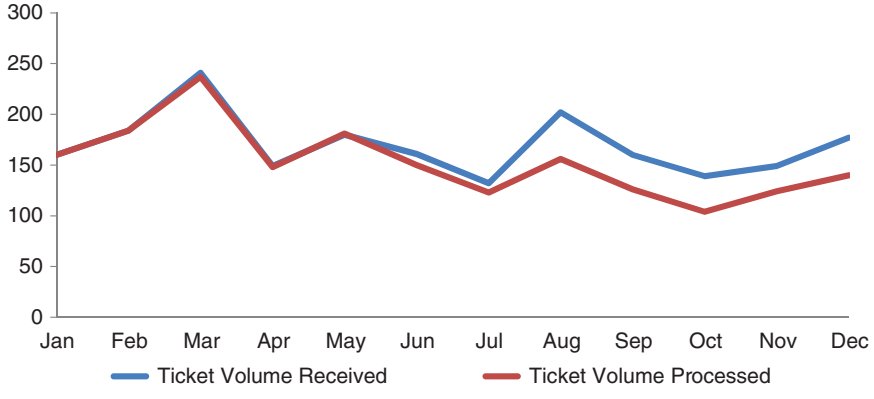


FIGURE 3.21 Clean up axis labels

5. Label data directly

Now that we have eliminated much of the extraneous cognitive load, the work of going back and forth between the legend and the data is even more evident. Remember, we want to try to identify anything that will feel like effort to our audience and take that work upon ourselves as the designers of the information. In this case, we can leverage the Gestalt principle of proximity and put the data labels right next to the data they describe.

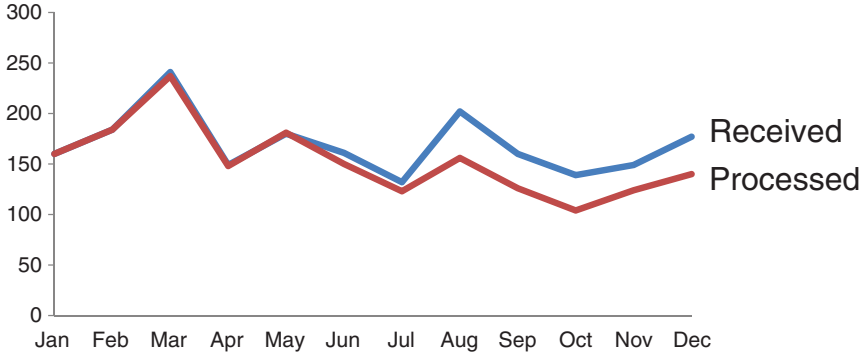


FIGURE 3.22 Label data directly

6. Leverage consistent color

While we leveraged the Gestalt principle of proximity in the prior step, let's also think about leveraging the Gestalt principle of similarity and make the data labels the same *color* as the data they describe. This is another visual cue to our audience that says, "these two pieces of information are related."

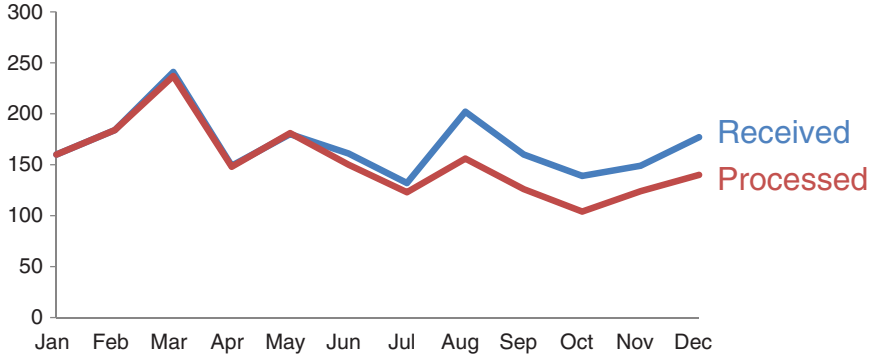


FIGURE 3.23 Leverage consistent color

This visual is not yet complete. But identifying and eliminating the clutter has brought us a long way in terms of reducing cognitive load and improving accessibility. Take a look at the before-and-after shown in Figure 3.24.

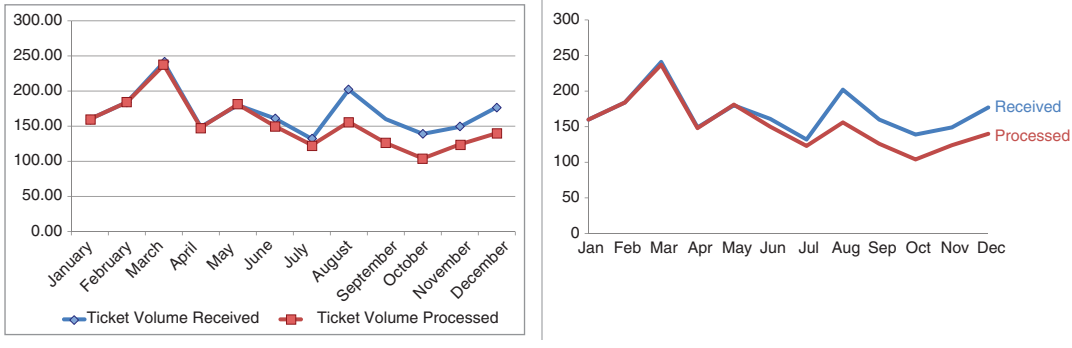


FIGURE 3.24 Before-and-after

In closing

Any time you put information in front of your audience, you are creating cognitive load and asking them to use their brain power to process that information. Visual clutter creates excessive cognitive load that can hinder the transmission of our message. The Gestalt Principles of Visual Perception can help you understand how your audience sees and allow you to identify and remove unnecessary visual elements. Leverage alignment of elements and maintain white space to help make the interpretation of your visuals a more comfortable experience for your audience. Use contrast strategically. Clutter is your enemy: ban it from your visuals!

You now know how to **identify and eliminate clutter**.