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.

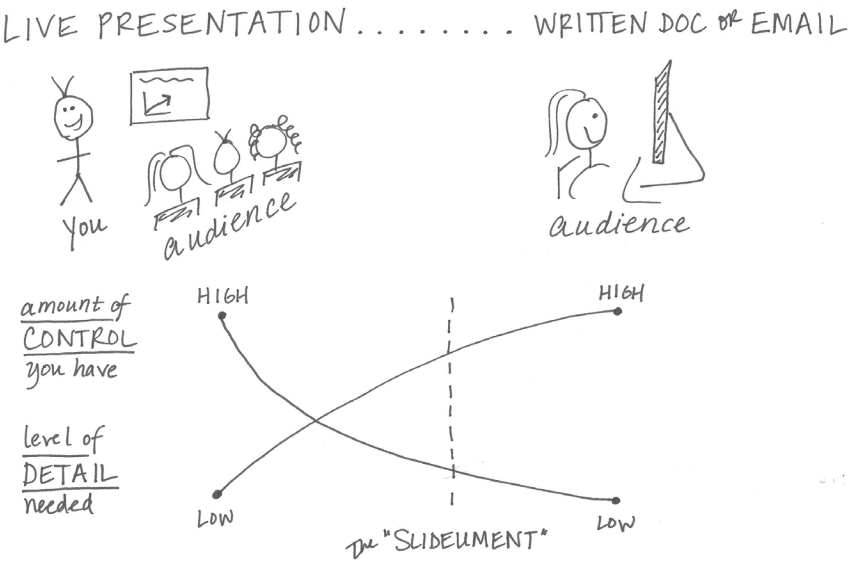


FIGURE 1.1 Communication mechanism continuum

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.

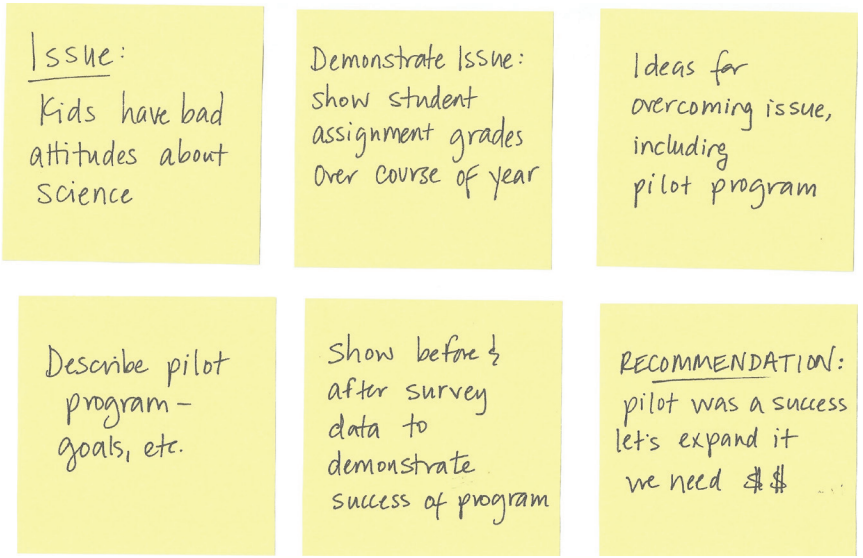


FIGURE 1.2 Example storyboard

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.

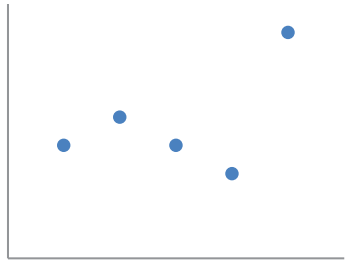
With that, consider your first lesson learned. You now **understand the importance of context**.

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.

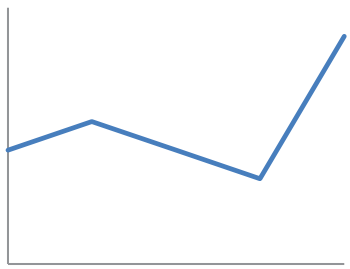
Simple text



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%

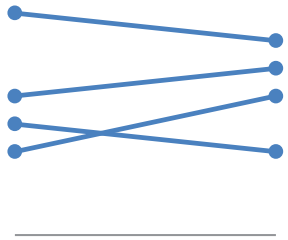
Table



Line

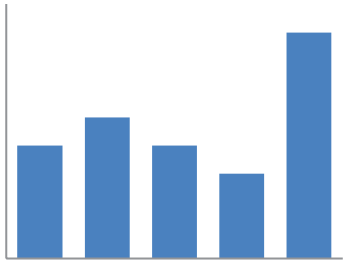
	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

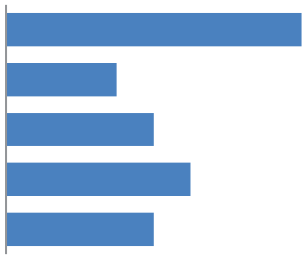


Slopegraph

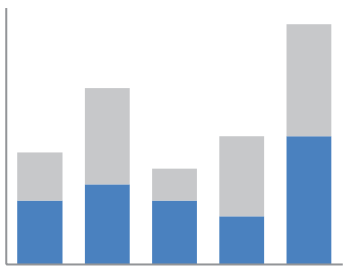
FIGURE 2.1 The visuals I use most



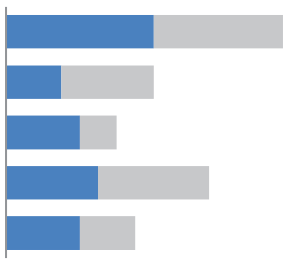
Vertical bar



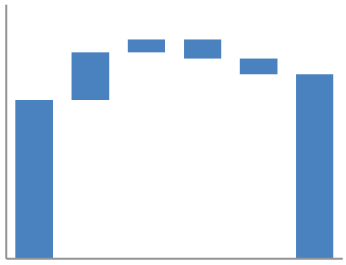
Horizontal bar



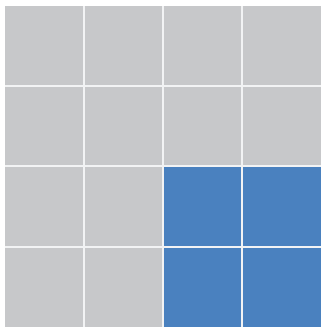
Stacked vertical bar



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.

Note: Based on children younger than 18. Their mothers are categorized based on employment status in 1970 and 2012.

Source: Pew Research Center analysis of March Current Population Surveys Integrated Public Use Microdata Series (IPUMS-CPS), 1971 and 2013

Adapted from PEW RESEARCH CENTER

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.

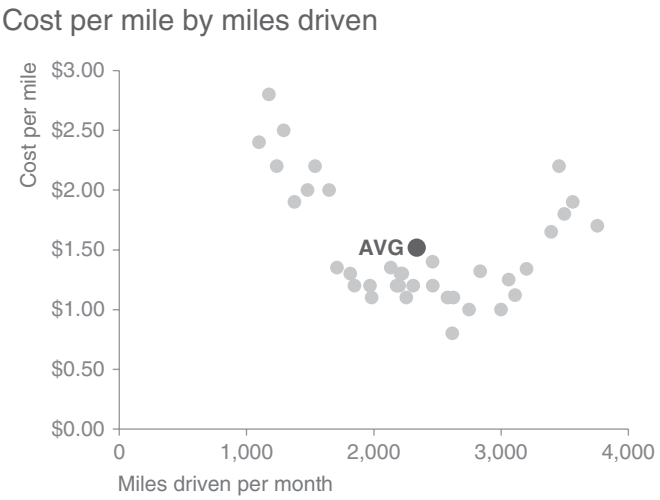


FIGURE 2.6 Scatterplot

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.

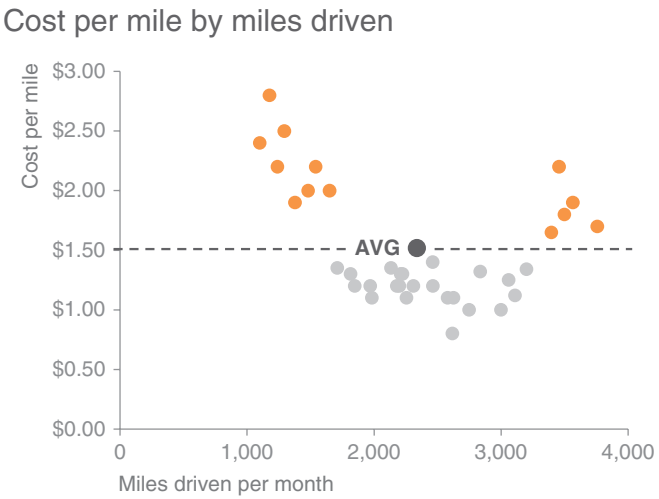


FIGURE 2.7 Modified scatterplot

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.

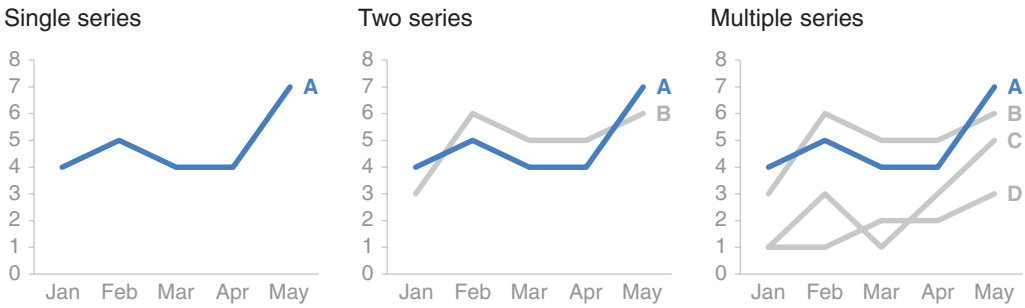


FIGURE 2.8 Line graphs

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.

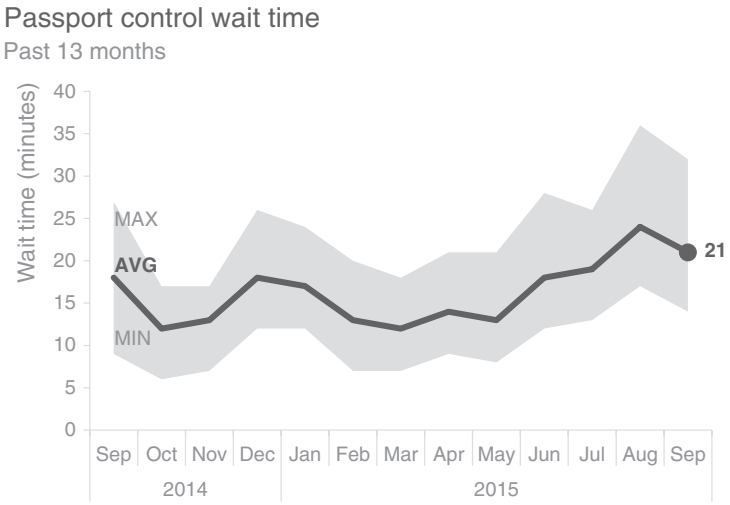


FIGURE 2.9 Showing average within a range in a line graph

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.

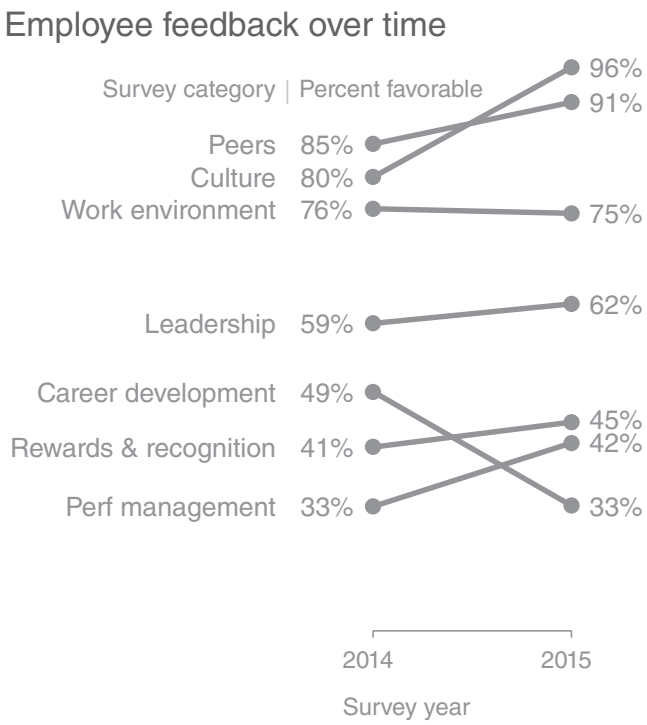


FIGURE 2.10 Slopegraph

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.

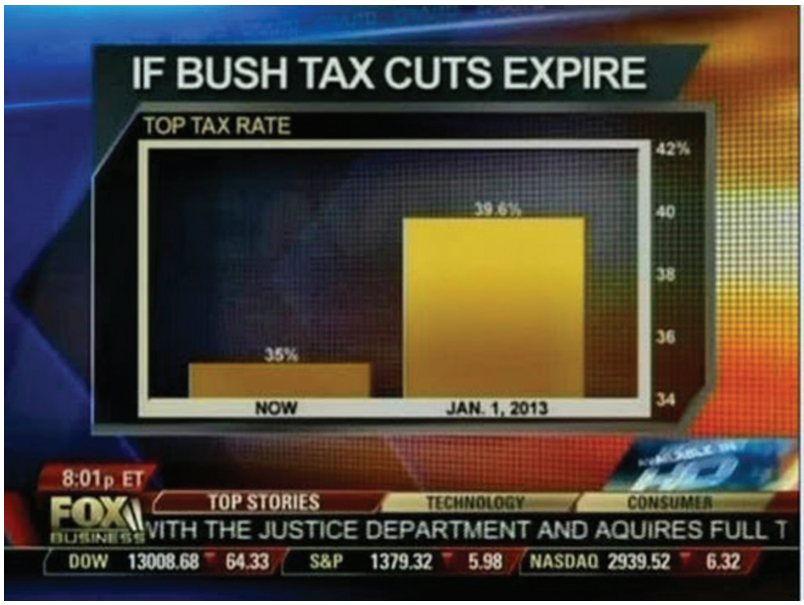


FIGURE 2.12 Fox News bar chart

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.

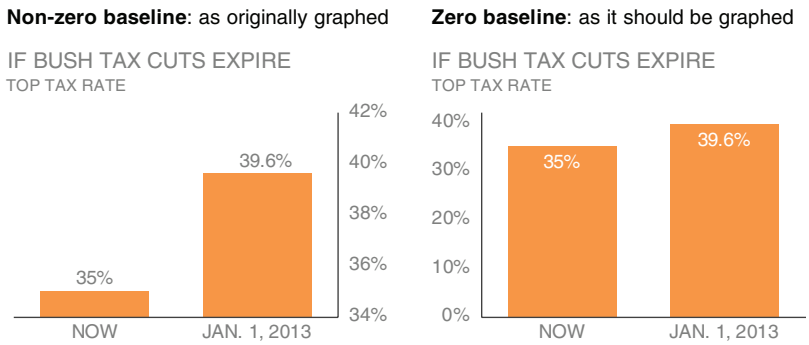


FIGURE 2.13 Bar charts must have a zero baseline

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.

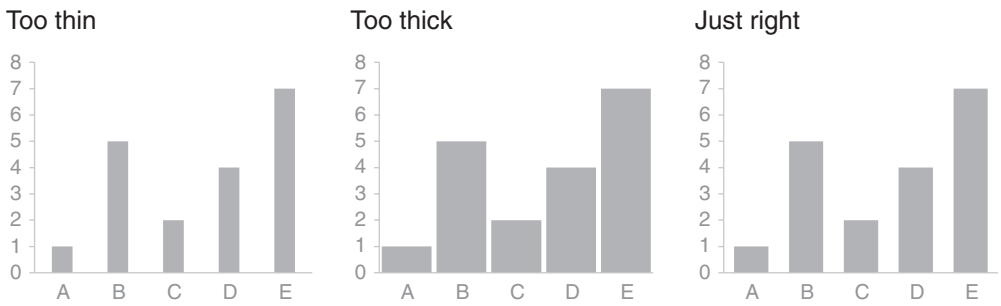


FIGURE 2.14 Bar width

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