CHAPTER 1

OUR BACKBONE

Why We Visualize

LEARNING OBJECTIVES

After reading this chapter, you will be able to:

- Articulate research-based reasons for pursuing better data visualization
- Justify your time and resource investment in data visualization
- Consider how data visualization has been used for deception
- Orient yourself to the skills you'll learn in this book

This chapter contains our justification for spending our time, energy, and resources on fiddling with our graphs. We address the very foundation of data visualization and the choices we need to make about the best chart type to use, and when to use it. This is the backbone of our work to visualize data, the reasoning we need to deliver to our boss when she or he asks why we are still dabbling in Excel when the report is due.

WHY WE VISUALIZE

I can remember back in my early days of grad school when I first began learning about organizational development and how companies operate. We learned that some successful companies achieved growth based on luck, chance, or the dream of a CEO. And other, more nimble and sustainable companies used data to inform their decisions. Yes, my younger readers, there used to be an era where we didn't collect data on our important efforts. Can you believe it? It's actually what made me pivot my career from teaching to data and research.

Since then, most organizations have invested significant time and money in becoming data-driven. Data are a great resource for making informed decisions more quickly, saving buckets of energy and lives down the road. Once organizations understood the power of data, they tried to get their arms around as much of it as possible.

Which brings us to today, where we have so much data we are suffocating in it. So now smart organizations are asking me how to cut through all the data they have and make it useful again. Here's the shortcut I use to navigate through all those data:

What's your point?

Seriously, that's the most important question to ask when pulling from collected data and creating a data visualization. It's the first thing I ask a client who sends me data for redesign. And it's the primary reason we visualize: because we have a point to communicate to the world. We have a compelling finding to share, a big idea revealed in our analysis that we need to say to people. A point.

Articulating the point generates an answer that drives nearly everything about visualizing those data. Here's how the conversation often goes:

Client:	"Thanks for working with us, Stephanie. We have these data from parents and students, and right now they are in a bar graph, and we are certain it could be displayed better—we just aren't sure how." (See Figure 1.1.)
Me:	"I can help with that, Client! What's your point?"
Client:	"Excuse me?"
Me:	"What's the point of showing these data about parent and student perspectives? Right now, it looks like you want people to compare parents and students. Is that your point?"
Client:	"Actually, no. And that's the most clarifying question you could have asked. Our point is that generally we expect students to report higher than parents on all of these questions, but our data showed that the students' expectations to go to college were way lower than their parents' expectations. That set off some alarm bells for us." (And this is when I silently pump my fist in the air, because the client answered the most important question and now I know how to better display these data.)
Me (after I catch my breath from	"The first thing we are going to do, then, is take what you just said and make it the headline of the graph. We are going to replace

all that fist pumping): this generic title with your main point. The next thing we will do is swap out a different graph type, maybe something like a slopegraph, since those are pretty good at highlighting when one thing is decreasing a lot and the rest are going up. Give me a day to play with some ideas, and let's talk tomorrow."

The Next Day

Me: "Good morning, Client! What did you think of that slopegraph I sent you?" (See Figure 1.2.)

Client: "It really does say exactly what we originally thought we needed to show. But I talked to my colleagues after our call yesterday and asked them, 'What's the point?' We decided that the real bottom-line point was that so few students have expectations to go to college. Forget the parents—that's a secondary issue right now. 'What's the point?' really helped us hone our thinking."

Me: "Ah well, in that case, you have other options for showing that point." (Telepathically sends new visual possibilities à la Figure 1.3.) "Maybe one of these?"

Client: "These are both right to the point. We will choose one today."

Figuring out your point sharpens the thinking and the messaging surrounding the data, and in doing so reveals the best way to visualize the data. When you get stuck with your graph, keep asking, "What's the point?" and craft an answer that speaks

Figure 1.1 /// Traditional clustered bar graphs can cloud the point.

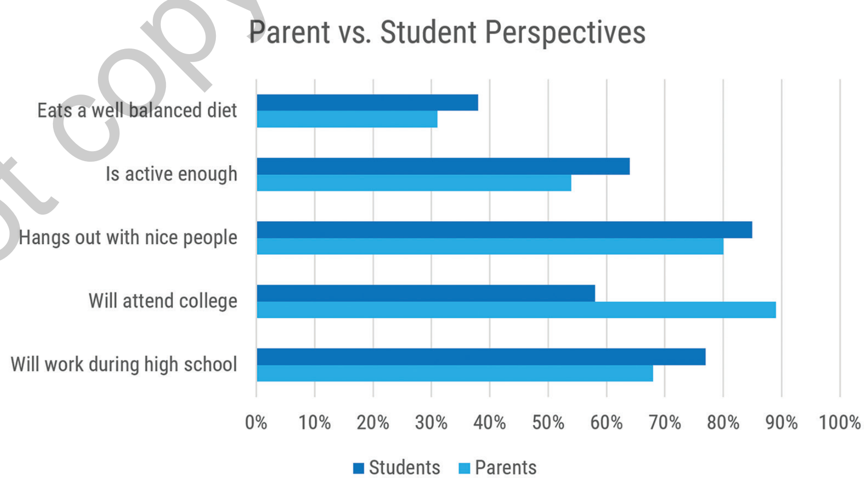


Figure 1.2 /// Slopegraphs are one way to compare two groups on multiple variables.

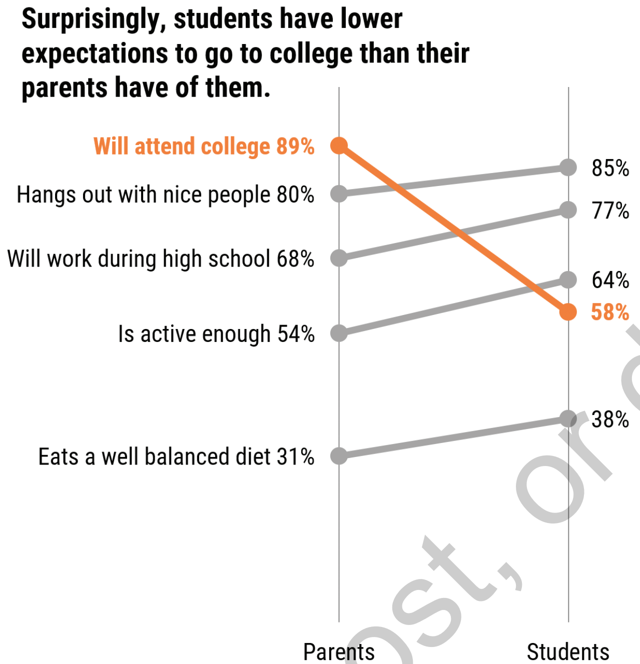
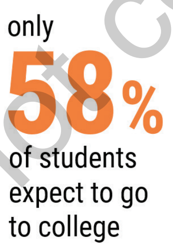
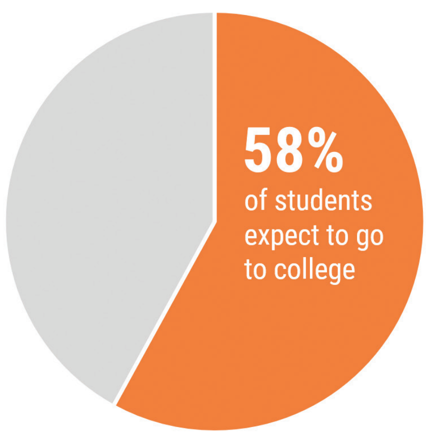


Figure 1.3 /// A single large number and a simple pie chart are two possible ways to help readers remember one important number.





to the audience you'll be presenting to. If you find you don't have a point, you probably shouldn't bother graphing the data. We visualize to communicate a point.

We also visualize to add legitimacy or credibility. People are persuaded by numbers and stories (de Graaf & Hustinx, 2011; Kosara & Mackinlay, 2013). When we can combine those things and tell stories with numbers, we have a communication powerhouse.

The research tells us that data are more persuasive when shown in graphs. Pandey, Manivannan, Nov, Satterthwaite, and Bertini (2014) presented mildly controversial topics to study participants. Some of the topic narrative contained simple column graphs, and some contained the same information in tables. The participants who saw it in graph format, particularly those who didn't have strong beliefs about the controversy beforehand, showed greater attitude change. In other words, people are more persuaded when they see data visually represented. In a supercool related study on political beliefs, Nyhan and Reifler (2013) found that misperception decreases when people are presented with (accurate) graphic representations of political information. One factor may be that we are primarily visual beings and that most of us, most of the time, are skimming the narrative for things that pop out at us and catch our attention (Evergreen, 2013). Data visualization does just that—it provides the pop.

Graphs and formulas seem to add credibility to data, even if they don't contain any new insights beyond what already exists in the narrative. Tal and Wansink (2014) experimented by including a graph (or a scientific formula) in materials about medication efficacy. They found that people who read the study materials believed the medications were more effective when the materials included a graph—even if the graph didn't contain substantial or additional information.

Of course, we use this power for good—to give more support and add credibility to our carefully researched points. But the same tools can be used to deceive.

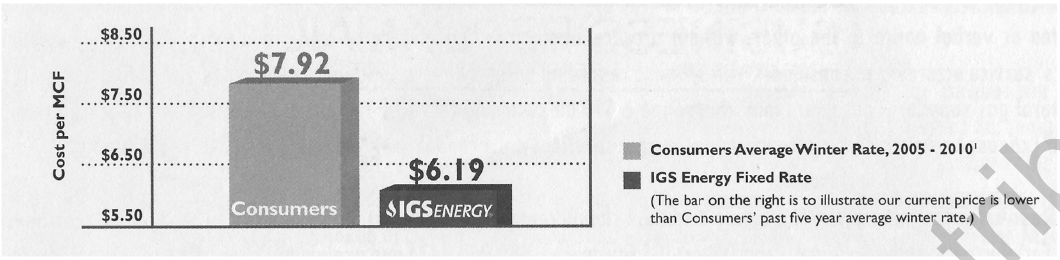
WHEN VISUALIZATION IS HARMFUL

At best, data visualization errors are unintentional mistakes that lead to misinformation. At worst, they are purposeful manipulations designed to influence the story a graph can tell. Elements like the scale of the axis or the size and shape of the graph can distort data and produce interpretation errors.

For example, take the advertisement in Figure 1.4, which arrived in my mail one day. Note how the y-axis begins at $5.50. This truncated axis cuts off most of the length of the columns so that it appears that the difference between Consumers and IGS Energy is greater than it really is. These errors create situations where data visualization is deceptive.

Changing aspects of the graph can lead to deception, whether intentional or benign. Pandey, Rall, Satterthwaite, Nov, and Bertini (2015) ran a study that included regular and distorted data visualizations. They manipulated the aspect

Figure 1.4 /// In a column graph, the axis should always start at 0. Otherwise the length of the bars sends a distorted message.



ratio of the graph and the y-axis in a couple of different ways and compared perceptions of these to perceptions of the same data in nondistorted graphs. The results were staggering: "the distorted charts [led] to responses between 58.5% and 129.5% bigger than the control condition" (p. 9). The effects were especially pronounced for line graphs. That said, there are justifiable reasons for truncating the y-axis on a line graph, and we will dig into this topic in Chapter 9. In such cases, the truncation is intentional, to better support honest decision making. It's a fine line to walk, because we must keep in mind that any alteration to the graph to change its shape can also alter the conclusions that can be drawn. Alteration to support decision making can be warranted. But distortion is real, common, and harmful.

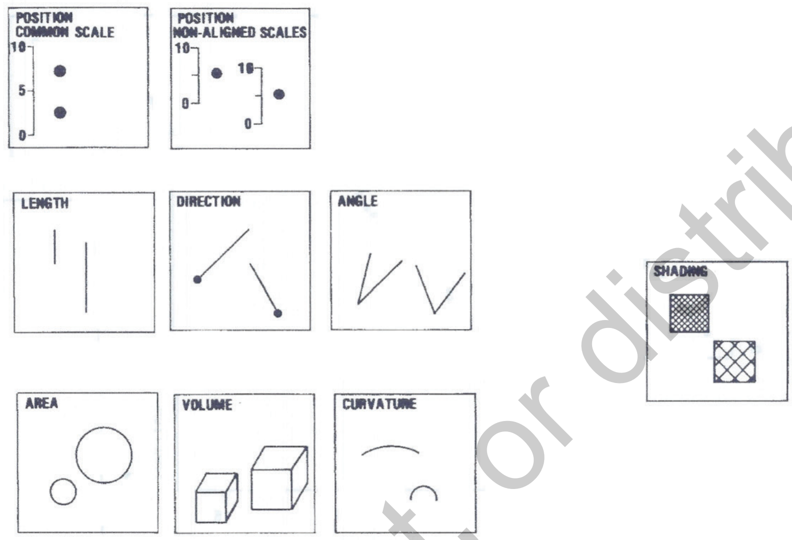
Data visualization is a powerful tool for communicating information. Thus, it is in everyone's best interest to learn how to display data in the best, most accurate way possible.

WHICH CHART TYPE IS BEST?

The surest way to defend against distortions and misrepresentations is to turn to the research about which type of graphs people interpret with accuracy. The foundational research in this area comes from Cleveland and McGill (1984). Parts of this research have been replicated or clarified in later research, which I mention in other chapters where it is relevant.

Cleveland is one of the grandfathers of data visualization, publishing pretty prolifically. Leading up to the 1984 publication, he ran many small studies testing how study subjects interpreted different graph types, essentially trying to figure out which graphs were the easiest and most accurate for people to understand. Easy and accurate. That's a nice goal, eh? Together with McGill, he published a hierarchy of graph types, placing the easiest and most accurate types of graphs at the top and the most confusing and error producing at the bottom (see Figure 1.5).

Figure 1.5 /// Cleveland and McGill offer a hierarchy of graph types, from most to least accurate.



Source: Adapted from Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association, 79*(387), 531–554.

At the top of Figure 1.5, Cleveland and McGill show that position on a common scale is the easiest visualization for people to interpret with accuracy. Position on a common scale? What does that look like? At the core, we are talking about dots on a line. You might be scratching your head at this one, or wondering if they meant a scatterplot. Excel's default chart options don't really include a graph type that reflects position on a common scale, but I introduce you to some in Chapter 3.

Next best, they said, is position on nonaligned scales. This means that if we have two graphs of dots side by side, we can compare and interpret them pretty well, as long as the scales are the same. The key here to me is that the scales have to be the same; otherwise that is part of what can be manipulated to misrepresent data. Still, I realize this isn't totally helpful information yet. Hang in there.

Below that level, we see length, direction, and angle. Length is how we encode column and bar charts. People are good at interpreting length accurately. Direction represents line charts and newer visualizations like slopegraphs. Angle is how pie charts display data. The researchers ran several small studies just using these three types of graphs to see how this order was going to shake out. They found that, time and again, angle produced the most errors. Pie charts produced the most errors.

You'd think the pain would stop there, but it gets worse. Turns out humans are super bad at interpreting area, volume, and curvature—the graph types shown at the bottom of the figure. Area is found in visuals like bubble graphs. Volume is anything 3-D. Curvature is how we would interpret a donut chart and other visualizations that can look more like art. The research does not support their functional use, but I will show the narrow acceptable uses of these graph types in other chapters.

Also listed at the bottom of this group is shading . . . but I've placed it off to the side of the hierarchy on purpose. Back in 1984, the only way researchers had to shade graphs was that terrible crosshatching pattern fill. That sort of shading caused optical illusions in readers; Tufte (2001) called this a moiré effect. Yikes! Thankfully, technology continues to evolve and increase our capability to visualize with color. Research now shows that people can distinguish between four shades of one color before things start to get difficult (Ware, 2013).

The point is that we should be striving to graph as high up in this hierarchy as possible so that it is easy for our audience to interpret our visualizations accurately. That said, we can't and shouldn't make everything into bar graphs. In addition to this hierarchy, our decision on which graph type to choose is based on the nature of our data and the audience we are speaking to, and we will spend the bulk of this book sorting out those issues.

TELL A STORY WITH DATA

Even though my favorite question in data visualization is "What's your point," I hear another one quite frequently: "How do I tell a story with data?" *Telling a story with data* must have been printed in some secret CEO newsletter somewhere, because it is a trendy question that doesn't always make sense for decision-making scenarios.

A story, traditionally, comprises a beginning, a middle, and an end. It has characters, a plot, a denouement, and a conclusion. Those elements make for great storytelling in a scenario like a TED stage. But do you really want all of that in your business meetings? Aren't those meetings long enough already?

Additionally, here's a sad truth: You know the typical way we were taught to discuss our research in academic settings? It goes something like this: introduction, background, literature review, methodology, discussion, conclusion. Outside academia, that sort of reporting makes our audiences wait until slide 100 before they get the bottom line they came to hear. But it is still a story arc, justifiable under the premise of *tell a story with data.*

No doubt, there will be times in your work life when a full story makes a lot of sense, where the customer is the hero and you are their sidekick, helping them save the world. But mostly, when we are in fast-paced, decision-making contexts, I don't think we actually want a story. We want interpretation. The audience wants the speaker to tell the point they think the audience needs to know, based on the available data evidence. Interpretation means we answer the question “So what?” with our graphs. Interpretation means we tell a one-sentence story as the title of our charts, capturing our most educated insights.

I just heard from a client—a senior vice president in a large company—who came from a data review meeting in which the other executives around the table spent the bulk of the hour trying to figure out the point the poor presenter was trying to make. The data, stuck in tables, weren’t helpful. The presenter didn’t have the tools to make clear graphs with succinct points. The presenter didn’t have this book. The hour came to a close, and everyone left the data review meeting without having made the decision they had been charged to make. This scenario repeats itself all over the world every single day. When you read and adopt the solutions I put forward throughout this book, you become the meeting rock star, ushering everyone through important decisions, and that is how you get a raise.

Beyond the awesome benefits you get as an individual data viz rock star, effective data visualization leads to clear conversations that support efficient decision making. And that leads to quicker, better-informed action. As a result, we end up building data-driven cultures, and I’ll tell you stories about how I’ve seen that happen many times over in the companies I work with.

HOW TO USE THIS BOOK

I wrote this book to help your data stories shine. A huge part of telling the right story is knowing how to pick the right type of graph. If you Google “Chart Chooser,” you will find a handful of other attempts to help you determine your graphing options. They all fall short for me, mainly because they are created from the point of view of a data visualizationist (visualizer? vizard?). By that, I mean they group chart types into broad categories like “Distribution,” which is not a user-friendly way to help you make your path down a decision tree. Few managers, team leaders, senior vice presidents, or even data junkies think in terms of “oh, I have a distribution here.” Rather, my audiences think in terms of “what we are trying to show” and the things they need the audience to do when viewing the data.

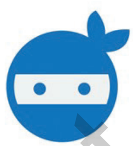
It’s “what we are trying to show” that makes up the table of contents of this book. Each chapter is a different data scenario. Inside each chapter, you’ll find the suite of graphing options that can best show that scenario. They don’t all have position on a common scale! That’s because some graph types are more appropriate for certain situations than others, and I’ve spelled out those considerations for you. Scan the chapter titles for the one that matches what you are trying to show, and then skip straight to that chapter to hone in on your data visualization possibilities. Flip to the inside front and back covers for Chart Chooser Cheat Sheets that spell out the right chart for the right data.

Some of these are probably new graph types that you haven't seen before. They are also the graph types that often fit your data story the best. Sure, they aren't a default option in Excel, but in each chapter I'll show you step-by-step instructions on how to make them. Excel (maybe surprisingly) is one of the most flexible data visualization tools at our disposal. For most of us, it requires no new expense and (also maybe surprisingly) not all that much new skill. We will chop up the defaults of Excel with some ninja skills to use what Excel does have and create powerhouse visualizations. If you are storing your data in another program, or even if you are analyzing your data in something like SPSS, you need to covert the data table to Excel to work with it. Don't worry—I'll show you how the data tables need to be set up, too. I'll hold your hand while you transform your work into something magical.

You'll find the icons from the Chart Chooser Cheat Sheets throughout the book, next to the instructions on how to craft the data visualization in Excel 2016. And I'm using a PC. The menus will look a little different if you are working in Excel on a Mac. That was a lie . . . they will look a lot different. But all of these same features are available in the more recent versions of Excel, regardless of the hardware you're using.

Starting off each set of instructions, you will also see a ninja score rating. This is my fairly arbitrary attempt to help you know how much work you are in for when you choose a graph type. The scale runs from 0, which means it's the easiest possible visualization to make or isn't made in Excel, to 10, which means we are going to engage in some hefty hacking of Excel's innards to pull out an amazing visual (see Figure 1.6).

Figure 1.6 /// Ninja levels tell you how much effort you'll need to put into the graph.



Levels 0–2

Minor adjustments to defaults required

3 minutes or less to a finished graph

Levels 6–8

Moderate adjustments to default graphs required

10 minutes or less to a finished graph

Levels 3–5

Some adjustments to default graphs required

6 minutes or less to a finished graph

Levels 9 & 10

Considerable adjustments to default graphs required

15 minutes or less to a finished graph, depending on the amount of data

You don't need tons of experience to start working on ninja level 10 visualizations—just some patience. Want an insider tip? Use the Chart Choosers to help you think through your graph type options and then sketch out your ideas using your own data before you crack open Excel. If you know what the end product should look like, you'll have an easier time constructing it with software. Sketching is where you work out the errors in your thinking. For more structured guidance on sketching data visualizations, check out *The Data Visualization Sketchbook* and search your favorite social media site for #evergreensketch to get inspired by others.

This skill of knowing how to push the right buttons to make the best graph is an increasingly valuable asset to organizations who want to make data-driven decisions. It's no wonder that IBM reported that job postings with "data visualization" as a required skill jumped 31% in 2016 (Markow, Braganza, Taska, Miller, & Hughes, 2017). That report went on to say that data-driven companies are more efficient and profitable than their competitors, and I see that in the companies I consult with. Here is how you leap ahead.

When you finish a new graph type, you're going to feel like a ninja rock star (you might think such a character doesn't exist, but just try it and you'll see). When your new graph types get accolades from your readers, you're going to feel like a superhero. Who knew being a data nerd could lead to such happiness? Well, it does. And more importantly, it leads to clear, effective communication.

/// EXERCISES

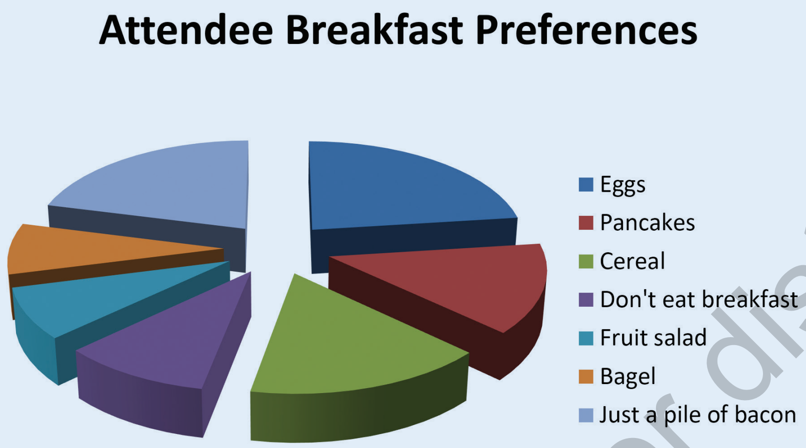
This 3-D thing is no joke. Let's say we're at your favorite conference and we're watching everyone eat breakfast from behind a fern. We tally up their choices into everyone's best friend: the 3-D exploding pie chart (see Figure 1.7).

While eggs were most popular, what was second? Just a pile of bacon or cereal? Please base your decision off the data in the chart, not your own personal preferences (I know how you feel about bacon). Hard to tell, isn't it? That's how much angle and area together can distort the data. For the answer in an improved visualization, check out **https://stephanieevergreen.com/books/**.

Download the practice file at **https://stephanieevergreen.com/books/**. The first time you make a new graph type, it may be easiest to work with the same data I'm using in my examples. Once you get the hang of it, you can apply the same steps to your own data. In the practice file, you'll see a tab for each chart type shown in this book with the data table already inserted.

While some would claim that stories must have a traditional arc, in which we leave the audience hanging from a cliff until the very end when we wow them with our conclusions, stories can also be very short and to the point. Here's one probably misattributed to Hemingway: "For sale: Baby shoes. Never worn." Ugh! Gets you right in the guts, doesn't it? Can you come up with a six-word story as the title of your graphs? Give it a try! For more six-word story inspiration, see **http://www.sixwordstories.net/**.

Figure 1.7 /// 3-D pie charts combine two things—area and angle—that people are bad at interpreting.



/// RESOURCES

Deception in data visualization is often unintentional. Check out the SlideShare compilation of graph violations that can lead to misinterpretation at **http://www.slideshare.net/powerfulpoint/presenting-data-webinar-presentation**.

As much as I'm on the Excel bandwagon, R software is also becoming increasingly popular (I feel as if just typing that sentence sounds the death knell) because it is open source—free—and so flexible that it can do nearly anything you can imagine . . . or write code for. Yes, R requires coding, but I put R tutorials for selected graphs from this book online at **https://stephanieevergreen.com/books/**.

/// REFERENCES

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531–554.

de Graaf, A., & Hustinx, L. (2011). The effect of story structure on emotion, transportation, and persuasion. *Information Design Journal*, *19*(2), 142–154.

Evergreen, S. D. H. (2018). *Presenting data effectively: Communicating findings for maximum impact* (2nd ed.). Thousand Oaks, CA: SAGE.

Kosara, R., & Mackinlay, J. (2013). Storytelling: The next step for visualization. *IEEE Computer*, *46*(5), 44–50.

Markow, W., Braganza, S., Taska, B., Miller, S. M., & Hughes, D. (2017). *The quant crunch: How the demand for data science skills is disrupting the job market.* Burning Glass Technologies: Boston, MA. Retrieved from https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IML14576USEN

Nyhan, B., & Reifler, J. (2013). *Blank slates or closed minds? The role of information deficits and identity threat in the prevalence of misperceptions.* Unpublished manuscript, Dartmouth College, Hanover, NH.

Pandey, A., Manivannan, A., Nov, O., Satterthwaite, M., & Bertini, E. (2014). The persuasive power of data visualization.

New York University Public Law and Legal Theory Working Papers, Paper 474.

Pandey, A. V., Rall, K., Satterthwaite, M., Nov, O., & Bertini, E. (2015, April 18–23). How deceptive are deceptive visualizations? An empirical analysis of common distortion techniques. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp. 1469–1478). Seoul, Republic of Korea.

Tal, A., & Wansink, B. (2014). Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. *Public Understanding of Science*, 1–9. doi: 10.1177/0963662514549688

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

Ware, C. (2013). *Information visualization: Perception for design* (3rd ed.). Waltham, MA: Morgan Kaufmann.