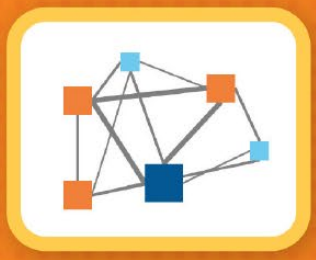
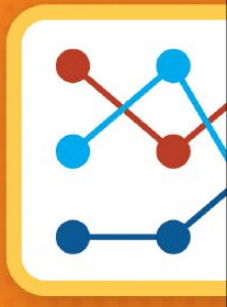
Stephanie D. H. EVERGREEN



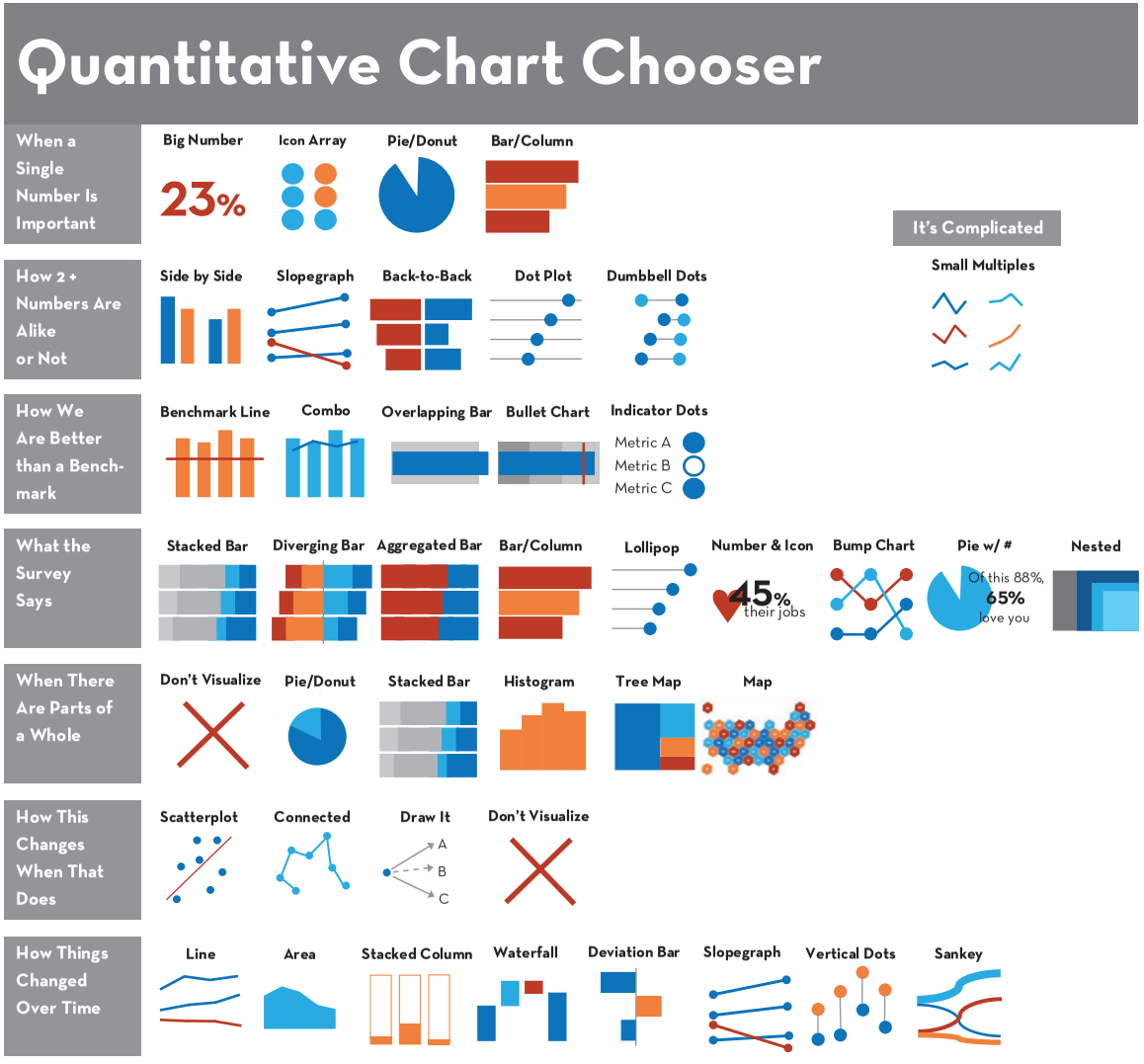
EFFECTIVE DATA VISUALIZATION

2 EDITION



The Right Chart for the Right Data

Quantitative Chart Chooser



Effective Data Visualization

Second Edition

Sara Miller McCune founded SAGE Publishing in 1965 to support the dissemination of usable knowledge and educate a global community. SAGE publishes more than 1000 journals and over 800 new books each year, spanning a wide range of subject areas. Our growing selection of library products includes archives, data, case studies and video. SAGE remains majority owned by our founder and after her lifetime will become owned by a charitable trust that secures the company's continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne

Effective Data Visualization

The Right Chart for the Right Data

Second Edition

Stephanie D. H. Evergreen
Evergreen Data & Evaluation, LLC

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London, EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
18 Cross Street #10–10/11/12
China Square Central
Singapore 048423

Printed in the United States of America

This book is printed on acid-free paper.

Library of Congress Cataloging-in-Publication Data

Names: Evergreen, Stephanie D. H., author.

Title: Effective data visualization : the right chart for the right data / Stephanie D. H. Evergreen.

Description: Second edition. | Thousand Oaks : Sage, [2020] | Includes bibliographical references and index.

Identifiers: LCCN 2018058818 | ISBN 9781544350882 (paperback : alk. paper)

Subjects: LCSH: Visual communication. | Charts, diagrams, etc. | Presentation graphics software. | Graphic design (Typography) | Information visualization.

Classification: LCC P93.5 .E937 2019 | DDC 302.23—dc23
LC record available at https://lccn.loc.gov/2018058818

Acquisitions Editor: Helen Salmon
Editorial Assistant: Megan O'Heffernan
Marketing Manager: Shari Countryman
Production Editor: Veronica Stapleton Hooper
Copy Editor: Rachel Keith
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Dennis Webb
Indexer: Beth Nauman-Montana
Cover Designer: Scott Van Atta

19 20 21 22 23 10 9 8 7 6 5 4 3 2 1

/// BRIEF CONTENTS

/// DETAILED CONTENTS

/// PREFACE

The moment you publish a book that is based on software, you accept the fact that some of the content is already obsolete and that a second edition is inevitable. This was actually a genius move on my part.

But this edition of *Effective Data Visualization* contains a lot more than updated screenshots of Excel. Hang on tight—I included nine entirely new quantitative graphs: overlapping bars (the number-one graph I recommend to the companies I consult), vertical dumbbell dot plots (my number-two most recommended graph), waffle charts, pictographs, bump charts, connected scatterplots, waterfalls, tile maps, and combination charts with target lines. Oh my gosh, you are going to learn so much. Of course, the quantitative chart chooser printed on the inside front cover has been updated to reflect these additions so you have a handy reference guide.

Did you look at the inside back cover too? This edition features a qualitative chart chooser as well. In fact, the entire qualitative chapter was scrapped and overhauled. We have packed over a dozen ways to visualize qualitative data into this chapter. You are going to be introduced to compelling ways to show qualitative data that you probably haven't seen before, including journey maps, histomaps, and spectrum displays. As with the rest of the lessons in this book, you won't need anything beyond Excel and PowerPoint to recreate these qualitative visuals. Qualitative data visualization may not be as well developed as the field of quantitative viz, but this chapter and the accompanying chart chooser are steps in that direction.

Even with all the new content to learn from, you'll want to review your favorites, because I've updated and simplified the directions, saving you valuable time you can put toward other ways you save the world. I've also packed each chapter with more support from peer-reviewed research around human perception, cognition, psychology, and data visualization so that you can have more confidence trying out new graph types. Be sure to download the updated sample data set from this book's associated website **(https://stephanieevergreen.com/books/)** so you can build amazing visuals right alongside me.

Oh, one more edition addition: Every image is now in full color. Color should make it approximately a million times easier to make sure your screen matches my screenshots. Easy, impactful dataviz. Does it get any better?

See you out at a workshop, a keynote, or at least in a third edition,

Catch up with Dr. Evergreen on Twitter or Instagram, both @evergreendata.

Stephanie Evergreen, PhD
Kalamazoo, Michigan, USA, and in an airplane somewhere

/// ACKNOWLEDGMENTS

At the end of a productive day of workshopping with a U.S. federal agency, I was trying desperately to get out of their highly secure complex and to my train before it left the station. I had called a car, but we had a hard time finding each other in the maze of buildings and security gates. When we finally landed in the same place at the same time, I politely asked him to step on it. The map said our trip would take 27 minutes, but my train was going to leave in 19. He hit the gas and I said a little prayer to the travel gods, silently promising that I would dedicate this book to this driver if he got me there on time. He took an alternate route and dropped me at the train station two full minutes ahead of departure. This book is for you, Karim.

This is my life now, sprinting from company to company, helping them use data to save the world. I love it, even if I have made that silent promise to add people to this book's acknowledgments hundreds of times as we've fought traffic or weather delays, all while trying to type these pages. Wherever I go, I have had massive support from the people in the service industry, who so often go underappreciated but did so much to make sure my needs were taken care of and I could just write. This book is for the airline stewards, bartenders, baristas, bed-and-breakfast managers, servers, and cabbies who made my life easier.

Similarly, this book is for my parents and friends and family who have cooked meals and taken wonderful care of my son and generally made it possible for me to squirrel away in a room between travel stops with my laptop and some data. I'm also so happy that Jennifer Lyons has joined me as coauthor of Chapter 8.

As always, I'm ever grateful for the support of my colleagues at SAGE, Megan O'Heffernan and Helen Salmon, as well as the input from generous peer reviewers:

David Boyns, California State University at Northridge

Thomas Cappaert, Rocky Mountain University of Health Professions

Michael G. Elasmar, Boston University

John O. Elliott, The Ohio State University and Ohio Health Research Institute

Brian Frederick, Bridgewater State University

David Han, University of Texas at San Antonio

Daniel Hawes, Kent State University

Mindy Hightower King, Indiana University

Kamir Kouzekanani, Texas A&M University–Corpus Christi

Jodi F. Paroff, NYU Robert F. Wagner Graduate School of Public Service

Shun-Yung Kevin Wang, University of South Florida–St. Petersburg

Finally, my stellar clients and colleagues contributed their own (often bad—but then improved!) work for public scrutiny, and that is hard and brave and I'm grateful.

/// ABOUT THE AUTHOR

Dr. Stephanie D. H. Evergreen is a sought-after speaker, designer, and researcher. She is best known for bringing a research-based approach to helping others better communicate their work through more effective graphs, slides, and reports. She holds a PhD from Western Michigan University in interdisciplinary research, which included a dissertation on the extent of graphic design use in written data reporting. Dr. Evergreen has trained audiences worldwide through keynote presentations and workshops for clients such as Verizon, Head Start, American Institutes for Research, Brookings Institute, the Ad Council, Boys and Girls Club of America, and the United Nations. She led the first known attempt to revamp the quality of presentations for an entire association: the Potent Presentations Initiative for the American Evaluation Association (AEA). She is the 2015 recipient of the AEA's Marcia Guttentag Promising New Evaluator Award, which recognizes early notable and substantial accomplishments in the field. Dr. Evergreen is coeditor and coauthor of two issues of *New Directions for Evaluation* on data visualization. She writes a popular blog on data presentation at StephanieEvergreen.com. Her book *Presenting Data Effectively: Communicating Your Findings for Maximum Impact,* was first published by SAGE in the fall of 2013 and was listed as number one in social science research on Amazon in the United States and United Kingdom for several weeks.

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.

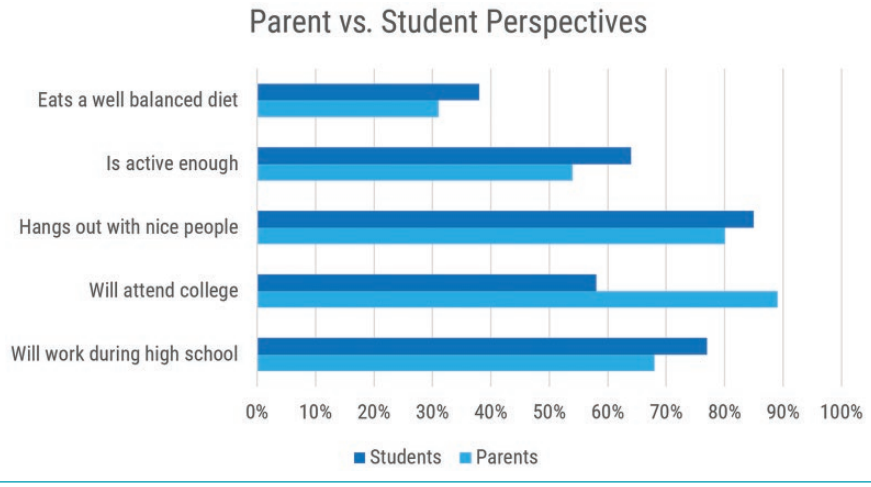


Figure 1.2 /// Slopegraphs are one way to compare two groups on multiple variables.

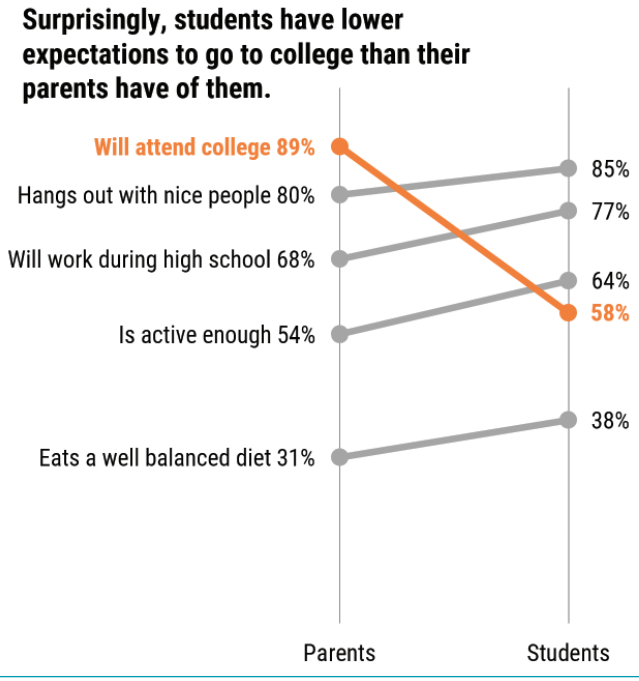
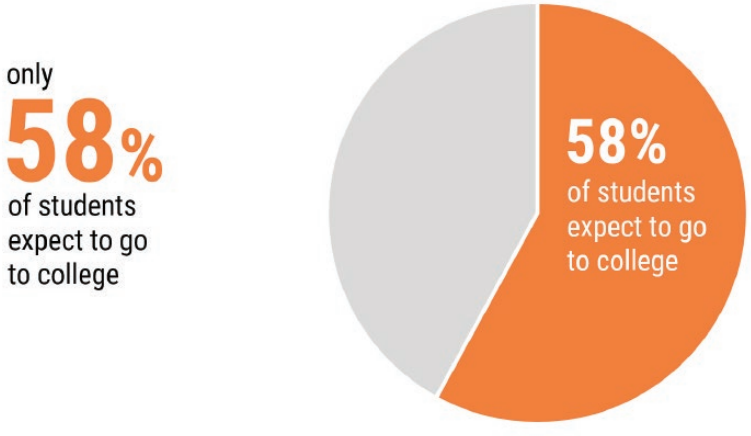


Figure 1.3 /// A single large number and a simple pie chart are two possible ways to help readers remember one important number.



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.

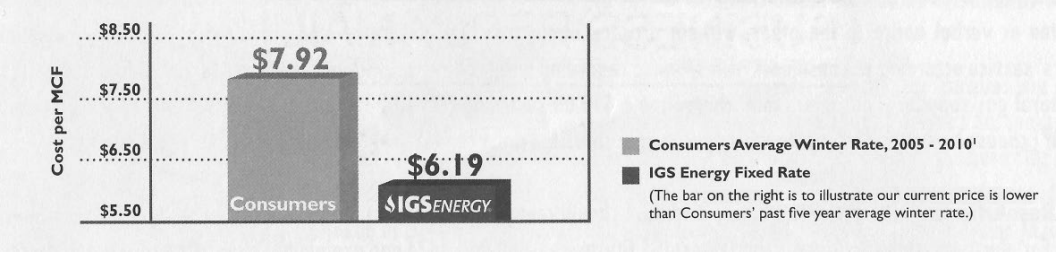
WHEN VISUALIZATION IS HARMFUL

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.

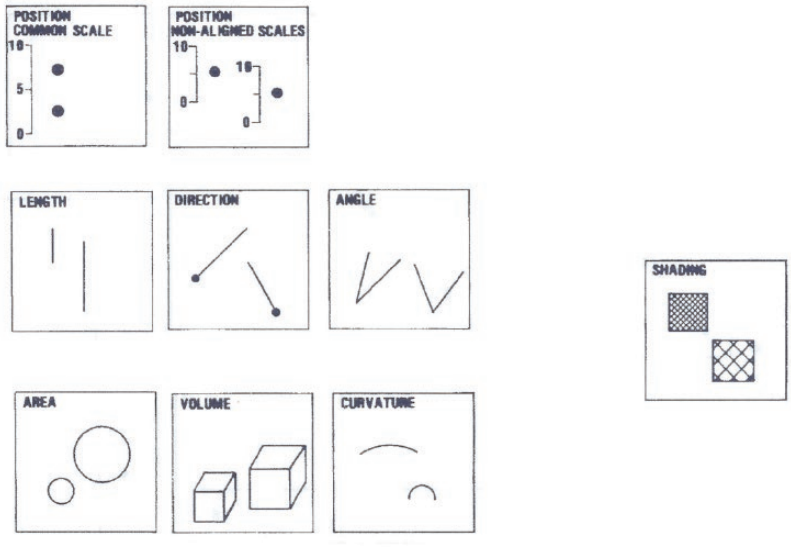
Data visualization is a powerful tool for communicating information. Thus, it is in everyone's best interest to learn how to display data in the best, most accurate way possible.

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.