

Cover image: Catherine Madden
Cover design: Flight Design Co.

Copyright © 2020 by Cole Nussbaumer Knaflic. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the Web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at www.wiley.com/go/permissions.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993, or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

ISBN 978-1-119-62149-2 (paper)
ISBN 978-1-119-62150-8 (ePub)
ISBN 978-1-119-62158-4 (ePDF)

Printed in the United States of America.

SKY10033004_020222

storytelling
with data *let's*
PRACTICE!

cole nussbaumer knaflic

illustrated
by
catherine
madden

WILEY

How to Build a Bridge!

close business partner

the right kind of print and electronic materials, and by forming alliances with other companies. Reprinted with permission of the author. May not be reprinted or reproduced without written permission from the book author or from such publications as *CD-ROMs*, *CD-ROMs* and *CD-ROMs*. You can download this material at <http://www.wiley.com> or www.wileyproducts.com.

Copyright © 1997 by John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in whole or in part, in any form or by any means,

electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the publisher.

Printed in the United States of America. 5 4 3 2 1 0 9 8 7 6 5 4 3 2 1

ISBN 0-471-32322-2 / 978-0-471-32322-0

WILEY

contents

acknowledgments	vii
about the author	ix
introduction	xi
chapter 1 understand the context	1
chapter 2 choose an effective visual	51
chapter 3 identify & eliminate clutter	107
chapter 4 focus attention	147
chapter 5 think like a designer	191
chapter 6 tell a story	235
chapter 7 practice more with Cole	285
chapter 8 practice more on your own	355
chapter 9 practice more at work	375
chapter 10 closing words	403
index	407

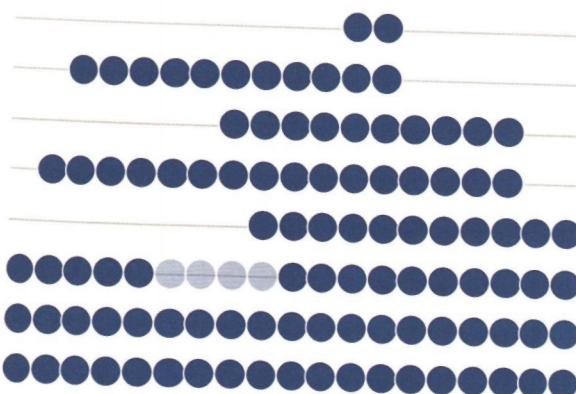
contents

iv	Introduction
x	Notes from the Author
ix	Notes from the Editor
1	Introduction to the Content
12	Score on Effective Interview
103	Inequity & Eliminate Culture
151	Focus on a Scenario
181	Shift into Observation
232	Identify a Story
282	Questions more with Cole
322	Practice more on Aont Omu
362	Practice more of Work
403	Chapter 10 - Guiding Words
403	Index

acknowledgments

Thanks to everyone who helped this book come to be...

2018		2019
A M J J A S O N D	J F M A M J J A S O	



Simon Beaumont, Lisa Carlson, Amy Cesal, Robert Crocker, Steven Franconeri, Megan Holstine, and Steve Wexler for your draft manuscript review and thoughtful feedback.

Kim Scheffler and Jasmine Kaufman for your sage guidance. Catherine Madden and Matt Meikle for your creativity and thought partnership. We've made a beautiful book!

Marika Rohn, my dear friend and editor extraordinaire: you've helped the words in my head make sense on paper!

Jody Riendeau for creating order, taking things off my plate, and bringing an awesome sense of levity to all that we do.

Elizabeth Ricks for your dedication and fantastic work sharing SWD lessons with the world. Congrats also on baby Henry!

Avery, Dorian, and Eloise Knaflic, my three beautiful children: you inspire me. You can be anything you want in life.

Randy Knaflic: you are my support, my voice of reason, my confidant, my everything. You hold my heart.

Thanks also to Bill Falloon, Mike Henton, Carly Hounsome, Steven Kyritz, Kimberly Monroe-Hill, Purvi Patel, Jean-Karl Martin, Amy Laundicano, Steve Cispke, RJ Andrews, Mike Cisneros, Alex Velez, Beatriz Tapia, Brenda Chi-Moran, and the team at Quad Graphics. Gratitude goes to all of our clients and everyone reading these words now (that's you!). Thanks for being on this amazing journey with me. Happy practicing!

about the author

Cole Nussbaumer Knaflic tells stories with data. She is the founder and CEO of *storytelling with data* (SWD) and author of the best-selling book, *storytelling with data: a data visualization guide for business professionals* (Wiley, 2015), which has been translated into a dozen languages, is used as a textbook by more than 100 universities, and serves as the course book for tens of thousands of SWD workshop participants. For nearly a decade, Cole and her team have delivered knock-out interactive learning sessions highly sought after by data-minded individuals, companies, and philanthropic organizations all over the world. They also help people create graphs that make sense and weave them into compelling stories through the popular SWD blog, podcast, and monthly challenge.

Prior to SWD, Cole's unique talent was honed through analytical roles in banking, private equity, and as a manager on the Google People Analytics team. At Google, she used a data-driven approach to inform innovative people programs and management practices and traveled to Google offices throughout the US and Europe to teach the course she developed on data visualization. Cole has acted as an adjunct faculty member at the Maryland Institute College of Art (MICA), where she taught Introduction to Information Visualization, and regularly guest lectures at prominent universities in the US and beyond.

Cole has a BS in Applied Math and an MBA from the University of Washington. When she isn't ridding the world of ineffective graphs, Cole is undertaking the adventures of parenting three young children with her husband at home in the Midwest and on travels abroad.

introduction

I often receive emails from people who have read my first book, *storytelling with data*, or attended one of our workshops by the same name. There are notes of encouragement, support for the work we're doing, and plenty of questions and requests. I especially love hearing the success stories: reports of having influenced a key business decision, spurred an overdue budget conversation, or prompted an action that positively impacted an organization's bottom line. The most inspiring accounts are those of personal growth and recognition. One grateful reader applied *storytelling with data* principles during an interview, helping him land a new job. All of this success is the result of people from different industries, functions, and roles committing time to improve their ability to communicate with data.

I also hear regularly from people who want *more*. They've read the book and understand the potential impact of telling stories with data, but struggle with the practical application to their own work. They have additional questions or feel they are facing nuanced situations that are keeping them from having the desired impact. It's clear that people crave more guidance and practice to help fully develop their data storytelling skills.

Others reach out who are—or would like to be—teaching the lessons outlined in *storytelling with data*. In many cases, they are university instructors (it's amazing to think that *storytelling with data* is used as a textbook at more than 100 universities around the world!) or they are a part of a learning and development function within an organization, interested in building an in-house course or training program. There are also leaders, managers, and individual contributors who want to upskill their teams or provide good coaching and feedback to others.

This book addresses all of these needs for individuals, teachers, and leaders. By sharing invaluable insight through many practical examples, guided practice, and open-ended exercises, I will help build your confidence and credibility when it comes to applying and teaching others to apply the *storytelling with data* lessons.

How this book is organized & what to expect

Each chapter starts with a brief recap of the key lessons that are covered in *storytelling with data*. This is followed by:

practice with Cole: exercises based on real-world examples posed for you to consider and solve, accompanied by detailed step-by-step illustration and explanation

practice on your own: more exercises and thought-provoking questions for you to work through individually without prescribed solutions

practice at work: thoughtful guidance and hands-on exercises for applying the lessons learned on the job, including practical instruction on when and how to solicit useful feedback and iterate to refine your work from good to great

Much of the content you'll encounter here is inspired by our *storytelling with data* workshops. Because these sessions span many industries, so do the examples upon which I'll draw. We'll navigate between different topics—from digital marketing to pet adoption to sales training—giving you a rich and varied set of situations to learn from as you hone your data storytelling skills.

Warning: this is not a traditional book that you sit and read. To get the most out of it, you'll want to make it a fully interactive experience. I encourage you to highlight, add bookmarks, and take notes in the margins. Expect to be flipping between pages and examples. Draw, discuss with others, and practice in your tools. This book should be beat up by the time you're done with it: that will be one indication that you've utilized it to the fullest extent!

How to use this book in conjunction with the original SWD

SWD: let's practice! works as a great companion guide to *storytelling with data: a data visualization guide for business professionals* (Wiley, 2015; henceforth referred to as *SWD*). It will not replace the in-depth lessons taught there, but rather augment them with additional dialogue, many more examples, and a focus on hands-on practice.

This book generally follows the same chapter structure as *SWD* with a couple of differences, as shown in Figure 0.1. Chapters 7, 8, and 9 are comprehensive exercises that offer additional guidance and practice applying the lessons covered throughout *SWD* and here.

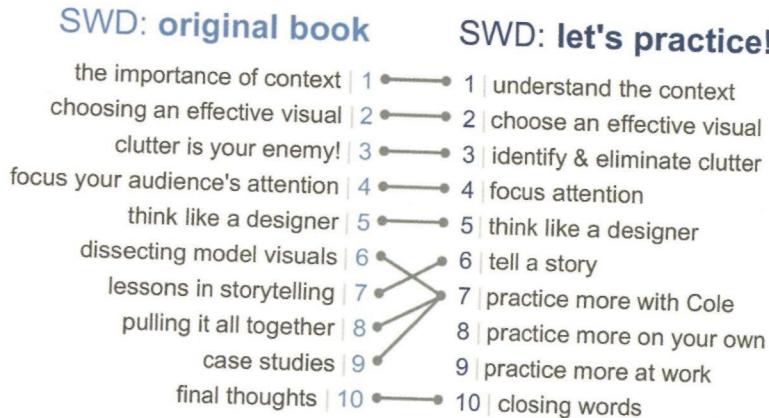


FIGURE 0.1 How *SWD* chapters correspond to this book

If you've picked up both *SWD* and *SWD: let's practice!*, you can use them in a couple of ways. You can read *SWD* once from start to finish to understand the big picture before digging into specifics. From there, you can determine which lessons you'd like to practice and can dive into the relevant sections within this book. Alternatively, you can peruse *SWD* one chapter at a time, then turn here to practice what you've read through hands-on exercises.

If you've already read *SWD*, feel free to jump right in as you will be familiar with these topics.

And if you've only bought this book, there is enough context within to give you the basics. You can always pick up a copy of *SWD* or check out the many resources at storytellingwithdata.com for supplemental guidance.

Do you want to learn or teach?

SWD: let's practice! was written with two different audiences in mind, united by a common goal—to communicate more effectively with data. Broadly, these two distinct groups are:

1. Those wanting to learn how to communicate more effectively with data, and
2. Those wanting to provide feedback, coach, or teach others how to communicate more effectively with data.

While the content is relevant for both groups, there will be subtle differences when it comes to getting the most out of it. Depending on your goal, the following strategies will maximize efficiency.

I want to learn to communicate more effectively with data

Because some later content builds upon or refers back to earlier content or exercises, begin with Chapter 1 and work through in numerical order. After that, you'll likely find yourself revisiting sections of interest and focusing your practice based on your specific needs and goals.

Start by reviewing the lesson recap for a given chapter. If you encounter anything that isn't familiar and you have access to *SWD*, turn back to the corresponding chapter for additional context.

After that, move straight into the *practice with Cole* exercises. First, work through each on your own—don't just jump to the solution (you're only cheating yourself!). If you're using this book with others, many of these activities lend themselves well to group discussion. The exercises in this section don't necessarily need to be worked through in order, though they do occasionally build upon prior exercises.

Once you've spent time on the given exercise (not just in your head: I strongly encourage you to write, draw, and use your tools), read through the provided solution. Observe where there are similarities and differences between that and your response. Be aware that there are very few situations where there is a single "right" answer. Some approaches are better than others, but there are usually numerous ways to solve a given problem. My solutions illustrate just one method that applies the lessons covered in *SWD*. Do read through all of the solutions, as many points of advice, tips, and nuances will arise that you will find helpful and insightful.

After completing the *practice with Cole* exercises, turn to the *practice on your own* section for more. These problems are similar to those in the first section, except that they don't include any predetermined solutions. If you are working in a group, have individuals first tackle a given exercise separately, then come together to present and discuss. Invariably, different people approach exercises in distinct ways, so you can learn a good deal through this sharing process. Confering with others is also great practice for talking through your design choices and decisions, which can further clarify thinking and help improve future application. Whether completing on your own or as part of a group, get feedback on your recommended approach. This will help you understand if what you propose is working, as well as where you can iterate to further improve effectiveness.

If, at any time, you find yourself with a current project that would benefit from applying the lessons outlined in a specific chapter, flip straight to the *practice at work* exercise section within that chapter. These contain guided practice that can be applied directly to real-world work situations. The more you practice implementing the various lessons in a work setting, the more they will become second nature.

Each chapter ends with discussion questions related to the lessons. Talk through these with a partner or perhaps even use as the basis of a larger book club conversation.

While the exercise sections in each chapter focus primarily on applying the given lesson, Chapters 7, 8, and 9 offer more comprehensive examples and exercises for applying the entire *storytelling with data* process. Chapter 7 (“practice more with Cole”) contains full-blown case studies presented for you first to solve, followed by my thought process for tackling and completing. Chapter 8 (“practice more on your own”) has additional case studies and robust exercises to practice the process without prescribed solutions. Chapter 9 (“practice more at work”) has tips on how to apply the *storytelling with data* process at work, guides to facilitate group learning, and assessment rubrics that you can use to evaluate your own work and seek feedback from others.

As part of your learning, it’s also imperative that you set specific goals. Communicate these to a friend, colleague, or manager. See Chapter 9 for more on this.

Next, let’s talk about how those interested in teaching others to effectively tell stories with data can use this book.

I want to provide feedback, coach, or teach others

You might be a manager or leader who wants to give good feedback on a graph or presentation from your team. Or perhaps you have a role in learning and development and are building training programs around how to communicate effectively with data. You may be a university instructor teaching students this important skill. In all of these scenarios, the chapter recap will provide an overview of the given lesson. After that, you will likely find the most value in the second and third exercise sections: *practice on your own* and *practice at work*. Each chapter ends with discussion questions that can be assigned, incorporated into tests, or used as the basis of group conversations.

The *practice on your own* section within each chapter contains targeted exercises helping those undertaking them practice the lessons outlined in the respective chapter and relevant section of SWD. These can be used as the basis of hands-on exercises in a classroom setting or assigned as homework. Some will also lend themselves well for use as group projects. These examples are provided without prescribed solutions. The problems in these sections can also work as models: consider where you could substitute data or visuals to create unique exercises.

Practice at work’s guided exercises can be used directly in a work setting as part of an ongoing program for professionals. They can be assigned, completed, and discussed in a group or classroom setting. Managers looking to develop their team’s skills may ask them to focus on specific exercises through their work or projects, or use with individuals as part of a goal-setting or career development process. For those teaching, Chapter 9 has additional *practice at work* exercises, including facilitator guides and assessment rubrics.

A quick note on tools

Many tools are available for visualizing data. You may use spreadsheet applications like Excel or Google Sheets. Perhaps you are familiar with chart creators such as Datawrapper, Flourish, or Infogram or data visualization software like Tableau or PowerBI. Maybe you write code in R or Python or leverage Javascript libraries like D3.js. Regardless of your tool of choice, pick one or a set of tools and get to know them as best you can so the instrument itself doesn't become a limiting factor for effectively communicating with data. No tool is inherently good or evil—pretty much any can be used well or not so well.

When it comes to undertaking the exercises in this book, you are encouraged to use whatever means for visualizing the data you have at your disposal. These may be tools you use currently, or possibly one or more that you'd like to learn. The visuals that illustrate the *practice with Cole* solutions were all created in Microsoft Excel. That said, this is certainly not your only choice and I welcome you to use other tools. We are also adding solutions built in other tools to our online library for you to explore.

On the topic of tools, there are a couple I highly recommend having on hand while reading this book: a pen or pencil and paper. You may consider dedicating a notebook to use as you work your way through the various exercises. Many direct you to write and sketch. There are important benefits to low-tech physical creation and iteration that we'll explore and practice, which can make the process of working in your technical tools more efficient.

Where to get the data

Downloads for the data throughout this book and for all of the visuals shown in the solutions for the *practice with Cole* exercises can be found at storytellingwith-data.com/letspractice/downloads.

Let's get started

There has never been a time in history where so many people have had access to so much data. Yet, our ability to tell stories with our graphs and visualizations has not kept pace. Organizations and individuals that want to move ahead must recognize that these skills aren't inherent and invest in their development. With a thoughtful approach, we can all tell inspiring and influential stories with our data.

I'm excited to help you take your data storytelling to the next level.

Let's practice!



4.0 *Writing*

10

11

12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

en een individueel niveau
of geslachtsmatig verschillend.
Kunnen we gezegd hebben dat
deelname evenementen de voor-

* Hoe kunnen we zeggen dat
deelnemers verschillend zijn
in belangstelling en interesses
in zowel het artiesten- als ander
vormen van cultuur? Welke vragen

staan op de lijst? Welk voor
uitgangspunt moet u volgen om
een goed inzicht te krijgen in
deelname aan evenementen?

en hoe kunnen we deze verschillen
van oorsprong en interesses
uiteenzetten? Welke vragen
moeten worden gesteld om deel-

* Hoe kunnen we zeggen dat
deelnemers verschillend zijn
in belangstelling en interesses
in zowel het artiesten- als ander

soorten cultuur? Welke vragen
moeten worden gesteld om deel-

name aan evenementen te begrijpen?
Welke verschillen bestaan er
tussen deelname aan evenementen
en andere vormen van cultuur? Welke
vragen moeten worden gesteld om
deelname aan evenementen te begrijpen?



chapter two

choose an effective visual

Once you've taken time to understand the context and planned your communication in a low-tech fashion, as we practiced in Chapter 1, comes the question: when I have some data I need to show, how do I do that in an effective way? This is the topic we'll tackle next.

There is no single "right" answer when it comes to how to visualize data. Any data can be graphed countless different ways. Often, it takes iterating—looking at the data one way, looking at it another way, and perhaps even another—to discover a view that will help us create that magical "ah ha" moment of understanding that graphs done well can do.

Speaking of iterating, we have some forthcoming exercises that will encourage you to do just that. Through the exercises in this chapter, we'll create and evaluate a number of different types of graphs, helping us understand both the advantages and limitations of different individual pictures of the data. Our go-tos will mainly be the usual suspects—lines and bars—but we'll look at some twists on graph types introduced in SWD as well.

Let's practice choosing an effective visual!

First, we'll review the main lessons from SWD Chapter 2.

SWD BOOK CHAPTER 2

FIRST, LET'S RECAP CHOOSING an EFFECTIVE VISUAL

SIMPLE TEXT

91%

Just because you have numbers doesn't mean you need a graph!

TABLE

	A	B	C
CATEGORY 1	152	211	403
CATEGORY 2	402	361	202
CATEGORY 3	352	171	391
CATEGORY 4	302	291	581

What is the main point I want to make?

OFTEN THERE ARE MORE EFFICIENT WAYS

Avoid using tables in live presentations because people stop listening & start reading

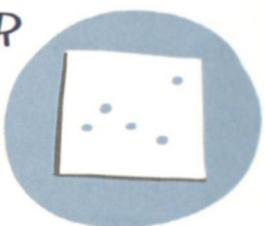
HEAT MAP

	A	B	C
CATEGORY 1	152	221	412
CATEGORY 2	402	361	202
CATEGORY 3	352	171	391
CATEGORY 4	302	291	581

Can work well when beginning to explore data and deciding where to dig further

EYES CAN EASILY PICK OUT BIG DIFFERENCES IN COLOR INTENSITY, but smaller ones don't stand out

SCATTER PLOT



Good for encoding data simultaneously on two axes to identify what relationships exist

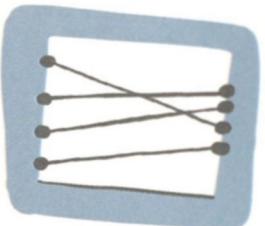
LINE



Rule: The lines that connect the dots have to make sense! Most effective with continuous data, often time

SLOPE GRAPH

A FANCY WORD FOR A LINE GRAPH WITH ONLY 2 POINTS ↑



Useful to focus on change between two points in time or difference between groups



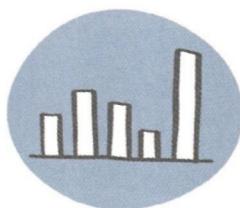
BAR CHARTS

Great for categorical data

Easy for our eyes - comparing heights to a consistent baseline

Rule:
Must have a zero baseline.
No exceptions!

VERTICAL

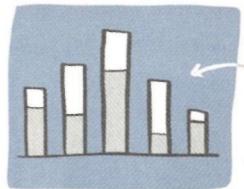


HORIZONTAL

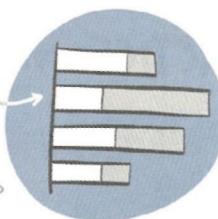


Good when category names are long

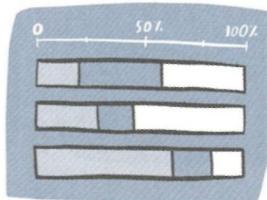
STACKED



OFTEN MISUSED... EASIER TO COMPARE TOTAL & FIRST SERIES, BUT SEGMENTS UP THE STACK DON'T LINE UP



100% STACKED



TWO BASELINES FOR COMPARISON

WATERFALL

ADDITIONS & DEDUCTIONS

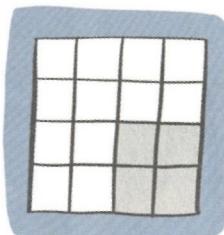


BEGINNING QUANTITY ENDING QUANTITY

Often used in finance to show variance to budget

SQUARE AREA

(AKA WAFFLE CHART)



THE GRID IS IMPORTANT BECAUSE WE TEND TO OVERESTIMATE AREAS

Good for showing numbers of very different magnitudes, or as an alternative to a pie chart

PRACTICE with COLE

- 2.1 improve this table
- 2.2 visualize!
- 2.3 let's draw
- 2.4 practice in your tool
- 2.5 how would you show this data?
- 2.6 visualize the weather
- 2.7 critique
- 2.8 what's wrong with this graph?

PRACTICE on your OWN

- 2.9 let's draw
- 2.10 practice in your tool
- 2.11 improve this visual
- 2.12 which graph would you choose?
- 2.13 what's wrong with this graph?
- 2.14 visualize and iterate
- 2.15 learn from examples
- 2.16 participate in #SWDchallenge

PRACTICE at WORK

- 2.17 draw it!
- 2.18 iterate in your tool
- 2.19 consider these questions
- 2.20 say it out loud
- 2.21 solicit feedback
- 2.22 build a data viz library
- 2.23 explore additional resources
- 2.24 let's discuss



PRACTICE with COLE

We'll start with basic tables and explore how visualizing data in graphs helps us more quickly see what's going on—as well as how different visuals cause us to identify new things and make varying design choices when graphing our data.

Exercise 2.1: improve this table

Frequently, when we first aggregate our data, we put it into a table. Tables allow us to scan rows and columns, reading the data and comparing the numbers. Let's look at an example table and explore both how we can improve it and take things a step further to visualize the data it contains.

Figure 2.1a shows the breakdown of new clients by tier for the recent year. Use this table to complete the following steps.

New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

FIGURE 2.1a Original table

STEP 1: Review the data in Figure 2.1a. What observations can you make? Do you have to make any assumptions when interpreting this data? What questions do you have about this data?

STEP 2: Consider the layout of the table in Figure 2.1a. Let's assume you've been told this information must be communicated in a table. Are there any changes you would make to the way the data is presented or the overall manner in which the table is designed? Download the data and create your improved table.

STEP 3: Let's assume the main comparison you want to make is between how accounts are distributed across the tiers compared to how revenue is distributed—and that you have the freedom to make bigger changes (it's not required to be a table). How would you visualize this data? Create a graph in the tool of your choice.

Solution 2.1: improve this table

STEP 1: When I encounter this table, I start reading and scanning down columns and across rows. In terms of specific observations, I might start by noticing that the majority of accounts are in Tiers B and C, while Tiers A and A+—though they don't make up a huge number (or percentage) of accounts—do make up a meaningful amount of revenue. In terms of questions, I wonder if the tiers are in order: I would think A+ belongs above A and am confused that they don't appear that way in the table (perhaps due to alphabetical sorting?).

I wish there was a "Total" row at the bottom, because in the absence of this I find myself wanting to add up numbers. In fact, it's when I start to do that when I notice some bigger issues. The third column (% Accounts)—which I assume means percent of total accounts—sums to 81.16%. The final column (% Revenue)—which I assume means percent of total revenue—sums to 95%. So now I'm unsure whether these really are percent of total or something else. If they are, then there must be some "Other" or "Non-tier" category that I'd want to include in order to have the full picture.

When I focus on the numbers themselves, two digits of significance (places past the decimal point) seem like a lot for the % Accounts column given the scale of the numbers. When showing data like this, you should be thoughtful about the appropriate level of detail. There isn't necessarily a single "right" answer, but you want to avoid too many digits of significance. This can make the numbers themselves harder to interpret and recall and may convey a false sense of accuracy. Is the difference between 7.08% and 7.09% meaningful? If not, we can drop a digit by rounding. Here, given the scale of the numbers and differences between them, I would round to whole numbers across all except the fourth column depicting revenue. There we are already summarizing in millions and it seems like we would lose important differences between the dollar volumes by rounding to a whole number, so there I'd round to one digit past the decimal point.

Figure 2.1b is an improved table that addresses the preceding points.

New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A+	19	2%	\$3.9	21%
A	77	7%	\$4.7	25%
B	338	31%	\$6.0	32%
C	425	39%	\$2.8	15%
D	24	2%	\$0.4	2%
All other	205	19%	\$0.9	5%
TOTAL	1,088	100%	\$18.7	100%

FIGURE 2.1b Slightly improved table



STEP 2: There are additional improvements I can make to this table. When tables are designed well, the actual design fades to the background so that we focus on the numbers in a way that makes sense. I recommend against shading every other row and instead am an advocate for white space (and limited light borders) to set apart columns and rows as needed. Speaking of white space, I typically avoid center-aligned text in graphs (because it creates hanging text and jagged edges that look messy) in favor of left- or right-aligning text. In the case of tables, however, I do sometimes opt for center alignment because of the separation this creates between columns (another common practice in tables is to right-align numbers or align by decimal point, which allows you to easily eyeball relative size). I can group the accounts-related columns and revenue-related columns with a single title (and under that, number and percent), which will reduce some redundancy of titles and also give me more space to be specific about what the columns represent. Doing so also allows me to make the columns narrower so the table overall takes up less space. These are some specific tips—I'll also put forth a couple of more general ones: consider the zigzagging “z” and where your eyes are drawn.

Consider the zigzagging “z”: Without other visual cues, your audience will typically start at the top left of your visual (for example, your table) and do zigzagging “z’s” across to take in the information. When we think about applying this to how we design our tables, it means you want to put the most important data at the top and at the left—when you can do so in the context of the overall data in a way that makes sense. In other words, if there are super-categories or data that needs to be taken into account together, keep them in the order that makes sense. In this particular example, I’d sort my tiers starting with the top (that is indeed A+) and decreasing as we move down the table. Going left to right, I’m happy enough with the way it is structured. I want to keep the distribution of accounts and percent of accounts next to each other since those relate to each other. If revenue were more important than accounts, I could move the two revenue columns leftwards, but I can also use other ways to focus attention there. Let’s discuss that next.

Where are your eyes drawn? Similarly to how we focus attention thoughtfully in graphs as part of explanatory analysis (something we’ll explore in detail in Chapter 4), we can also focus our audience’s attention in tabular data to establish hierarchy of information. This can be especially useful in instances where you can’t put the most important stuff leftward or at the top (because other constraints dictate the ordering). Despite this, you can still indicate relative importance to your viewers. Look back to Figure 2.1b: where are your eyes drawn? Mine go to the very first row where the column titles are Tier, # of Accounts, and so on. This isn’t even the data! Rather than use up ink and draw attention there, I can be conscious about where I want to direct attention in the data and take intentional steps to get my audience to look there. This can be done through sparing use of color or by outlining a specific cell or column or row. Adding visual aspects to some of the data in the table is another way to draw attention there: colors and pictures grab our attention when they are used judiciously.

If we assume the primary comparison we'd like our audience to make is between the distribution of % Accounts compared to % Revenue, I could apply heatmaping (using relative intensity of color to indicate relative value) to just those two columns. See Figure 2.1c.

New client tier share

TIER	ACCOUNTS		REVENUE	
	#	% OF TOT	\$M	% OF TOT
A+	19	2%	\$3.9	21%
A	77	7%	\$4.7	25%
B	338	31%	\$6.0	32%
C	425	39%	\$2.8	15%
D	24	2%	\$0.4	2%
All other	205	19%	\$0.9	5%
TOTAL	1,088	100%	\$18.7	100%

FIGURE 2.1c Table with heatmaping

As another approach, I could embed horizontal bar charts in place of the heatmap columns and allows us to see how the shape of the distribution varies across the given tier is harder, since these bars aren't aligned to a common baseline. Tip: If you are working in Excel, conditional formatting is available that will allow you to create heatmaping or embedded bars in a table with ease.

New client tier share

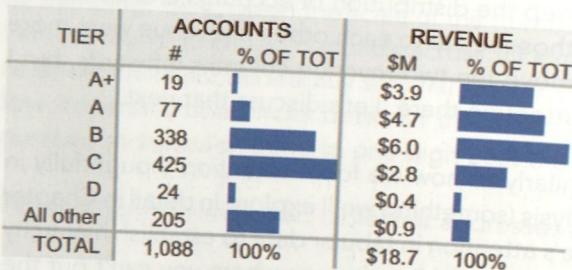


FIGURE 2.1d Table with embedded bars

STEP 3: Let's take it a step further and focus on the data that is in the bars in Figure 2.1d and review some different ways we could graph it. When I hear a term like "percent of total," it makes me think of parts of a whole—which might cause us to look to the pie chart. In this case, since we are interested in both % of Accounts and % of Revenue, we could depict this with a pair of pies. See Figure 2.1e.



New client tier share

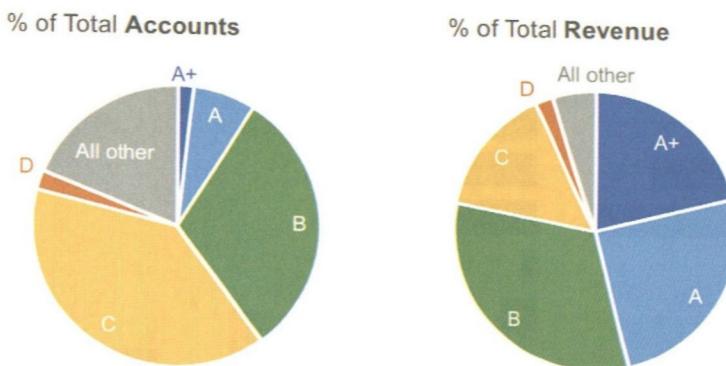


FIGURE 2.1e A pair of pies

I'm not a big fan of pies—I sometimes joke that there's one thing worse than a single pie: two pies!

Let me back up, though, and say that pies can work well if we want to make the point that one piece of the whole is very small, or another piece of the whole is very big. The challenge for me is that pies break down pretty quickly if we want to say anything more nuanced than that. This is because our eyes' ability to accurately measure and compare areas is limited, so when the segments are similar in size, it is difficult for us to assess which is bigger or by how much. If that's a comparison that is important, we'll want to represent it differently.

In this instance, the primary comparison we want our audience to make is between the various segments in the pie on the left and those in the pie on the right. This is difficult for two reasons: the area challenge mentioned above and the spatial separation between pies. This is further compounded by the fact that the segments are in different places on the right as a result of how the data differs between the breakdown on the left compared to the right. Basically, if any of the data is different between the pies (which it should be if we have something interesting to say about it!) then all the pieces are in different places across the two pies—making them hard to compare. In general, you want to identify the primary comparison you want your audience to make and put those things as physically close together and align to a common baseline to make that comparison easy.

Let's start by aligning each measure to its own baseline, with a view similar to the bars embedded in the table previously. See Figure 2.1f.

New client tier share

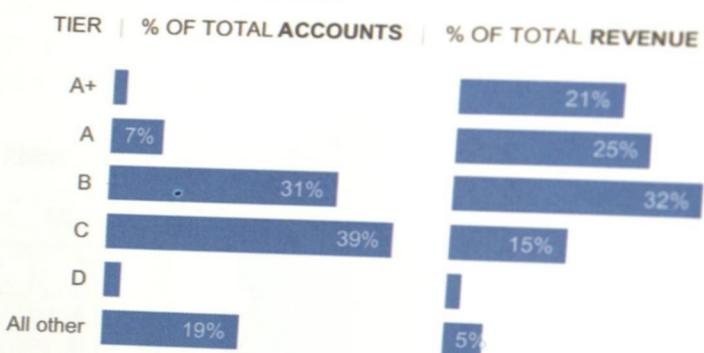


FIGURE 2.1f Two horizontal bar charts

In Figure 2.1f, it's very easy for us to compare the % of Total Accounts across tiers. It's also easy to compare the % of Total Revenue across tiers. I can attempt to compare accounts to revenue, but this is harder because they aren't aligned to a common baseline. If I want to allow for that as well, then I could pull both of these series into a single graph. See Figure 2.1g.

New client tier share

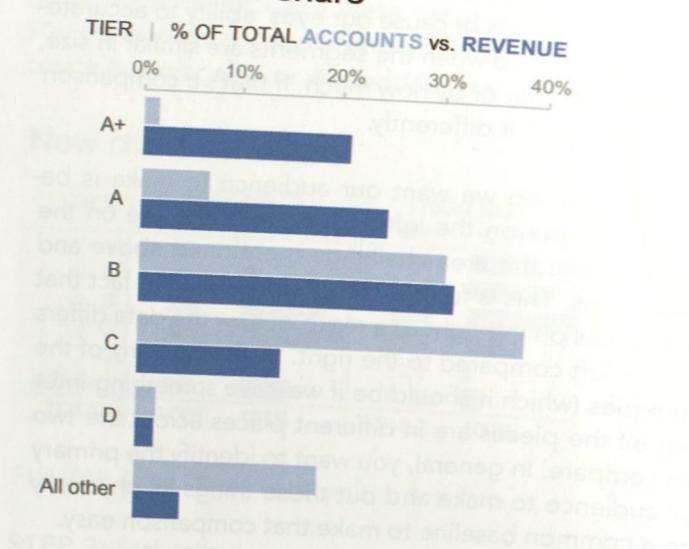


FIGURE 2.1g Horizontal dual series bar chart

With the arrangement in Figure 2.1g, the easiest comparison for me to make is, for a given tier, the % of Total Accounts compared to the % of Total Revenue. These elements are both the closest together and they are aligned to a common baseline. Bingo!



We could also flip this graph on its side into a vertical bar chart, or column chart. See Figure 2.1h.

New client tier share

% OF TOTAL ACCOUNTS vs. REVENUE

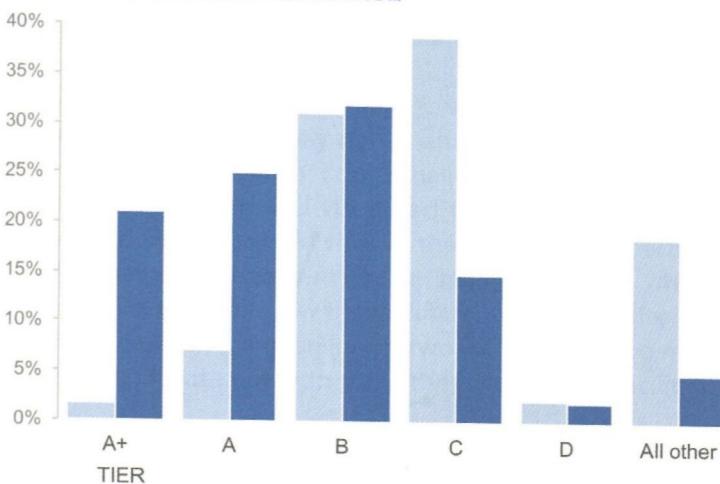


FIGURE 2.1h A vertical bar chart

When we depict data in this manner, the primary comparison our eyes are making is the endpoints of the paired bars relative to each other and to the baseline. Let's draw some lines to further highlight this comparison. See Figure 2.1i.

New client tier share

% OF TOTAL ACCOUNTS vs. REVENUE

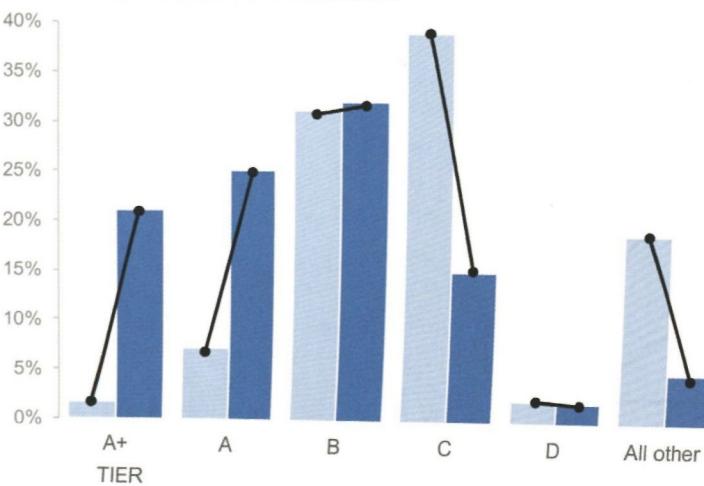


FIGURE 2.1i Let's draw some lines

Now that we've drawn the lines, we don't need the bars anymore. I've removed those in Figure 2.1j.

New client tier share

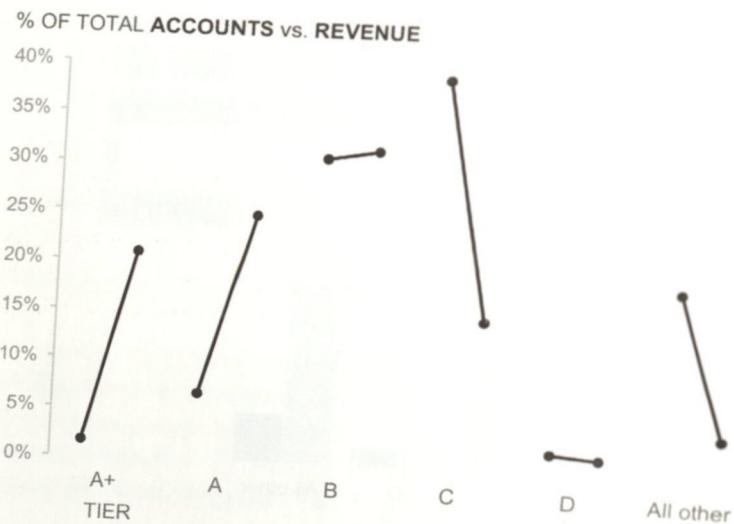


FIGURE 2.1j Take away the bars

Next, I'll collapse all of these lines and label everything directly. This yields the slopegraph shown in Figure 2.1k.

New client tier share

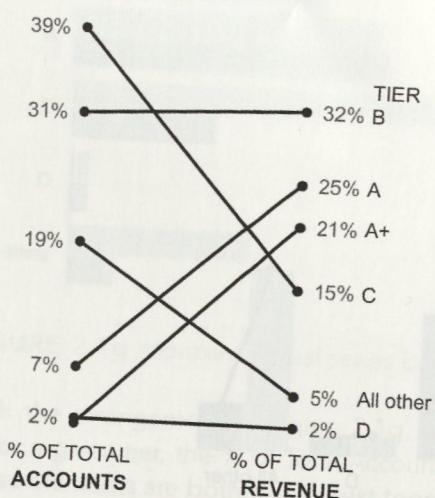


FIGURE 2.1k A slopegraph



Slopegraph is really just a fancy word for a line graph that only has two points in it. By drawing lines between the % of Total Accounts and % of Total Revenue for a given tier, we can quickly see where the two measures differ. Revenue as a proportion of total is quite a lot lower for Tier C and All other (indicated by lines sloping downwards), while revenue as a proportion of total is much higher for tiers A+ and A. In other words, though A+ and A make up a very small proportion of accounts (9% combined), together they account for nearly 50% of revenue!

We've looked at a number of ways to visualize this data. You likely made your own observations along the way about what worked well and what did not. What I've illustrated isn't exhaustive; I could have added a dot plot to the mix or calculated revenue per account and visualized that. That said, we don't typically have to go through every possible view of the data to find one that works. Perhaps both absolute values and percent of total are important, in which case the table might be the easiest way to show these different measures after all. If we can narrow our focus to a specific comparison or two, or a specific point we want to make, that will help us choose a way to show the data that will facilitate this.

Any data can be graphed countless different ways. This exercise illustrates how moving through different representations of our data allows us to more (or less) easily see different things. Allow yourself time to iterate and complete the additional exercises that will give you more practice at this important junction in the process!

Exercise 2.2: visualize!

Let's look at another table. The following shows the number of meals served each year as part of a corporate giving program. Spend a moment looking at the data. What is interesting about it?

Meals served over time

Campaign Year	Meals Served
2010	40,139
2011	127,020
2012	168,193
2013	153,115
2014	202,102
2015	232,897
2016	277,912
2017	205,350
2018	233,389
2019	232,797

FIGURE 2.2a Table showing meals served over time

Notice how much work it is to process a column of numbers like this. We read data that is presented to us in tabular form, which—though this may seem like a simple way to show the numbers—actually takes a ton of brainpower! When I scan these numbers, I see the jump from 2010 to 2011, and another between 2013 and 2014. You probably did, too. But if you're like me, it means you started at the top of the table and got there by scanning down the second column—comparing each new number to the one(s) before it.

Let's practice easing how hard our brains must work by making the data more visual. Download this data. Create the following visuals in the tool of your choice.

STEP 1: Apply heatmaping to the second column of values.

STEP 2: Create a bar graph.

STEP 3: Create a line graph.

STEP 4: Choose: which of the visuals you've created do you like best? Are there any other ways you would graph this data?

Solution 2.2: visualize!

Pretty much anything we do to visualize the data originally shown in table form in Figure 2.2a is going to make it more quickly understandable. Let's check out a few ways we can ease the processing.

STEP 1: First, let's apply some **heatmapping**. Most graphing applications have built-in functionality that will allow you to do this with ease. You can pick colors and choose how to apply them to the data. For example, I've created the following in Excel by applying conditional formatting to the second column of values. I indicated a 3-color scale, with lowest value white, 50th percentile light green and maximum value green. In some situations, you could add a legend to make it clear how to interpret the colors. In this case, I just want to give a general sense that more intense color represents bigger values and vice versa. Eyeballing it, this sense is intuitive given the numbers and relative intensity within the same hue.

Meals served over time

Campaign Year	Meals Served
2010	40,139
2011	127,020
2012	168,193
2013	153,115
2014	202,102
2015	232,897
2016	277,912
2017	205,350
2018	233,389
2019	232,797

FIGURE 2.2b Table with heatmapping

In Figure 2.2b, I'm perhaps more inclined to notice how much lower the number of meals served in 2010 was—it's totally white—less than a third of the next closest number! I can also quickly observe that 2016 had the greatest number of meals served without having to read the numbers. The relative intensity of color helps me more quickly interpret the relative quantitative values.

Related to this, I should point out that our eyes are pretty good at picking out big differences in intensity, but we have a harder time with more minor differences. This means that if there is something interesting about all of those medium shades of green, that's a little harder to quickly grasp and I might want to find a way to more fully visualize the values. Let's do that next.

STEP 2: Figure 2.2c shows a **bar graph** I could make based on this data. I chose to keep the y-axis for reference. Almost instantly, we can get a general sense of magnitude of the various bars. I thickened the bars from the default graph so there is less of a gap between them, which makes it easier for my eyes to follow along the tops of the bars and compare them to each other. I like the idea of bars.

We do have a continuous variable on the x-axis (time), but we can categorize it into years, which may make sense if we want to focus on a specific year at a time and have clear demarcation between the years.

Meals served over time

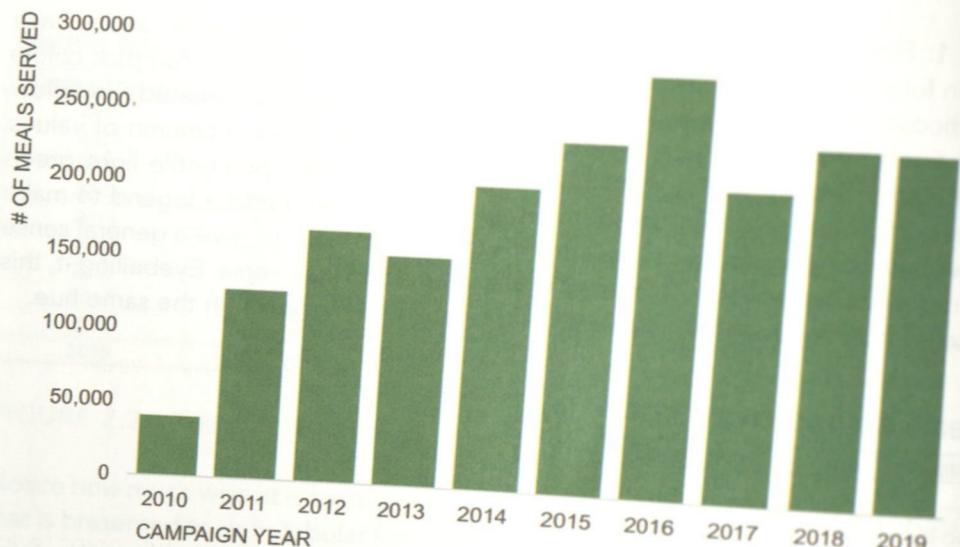


FIGURE 2.2c Bar chart

STEP 3: We can also show this data as a line graph; see Figure 2.2d. In this iteration, I decided to omit the y-axis and instead labeled only the beginning and end data points. This makes it easy (and obvious) for my audience to compare the number of meals served in 2010 to 2019. The rest of the values would have to be visually estimated. If there are other values you thought your audience would be particularly interested in (for example, the high point in 2016), you could also add data markers and labels to those points specifically.

When I remove the y-axis, I'll often use the subtitle space for the axis title. Here, given the graph title, you could probably argue that the subtitle is redundant and perhaps unnecessary. I'd rather be explicit so there is no question for my audience about what they are viewing. That said, another reasonable person might make a different decision.

I've used green in the visuals in this exercise, mainly to make it clear that—while I often default to blue—blue certainly isn't our only choice when it comes to using color in our visuals. We'll talk more about color as part of the exercises in Chapter 4.



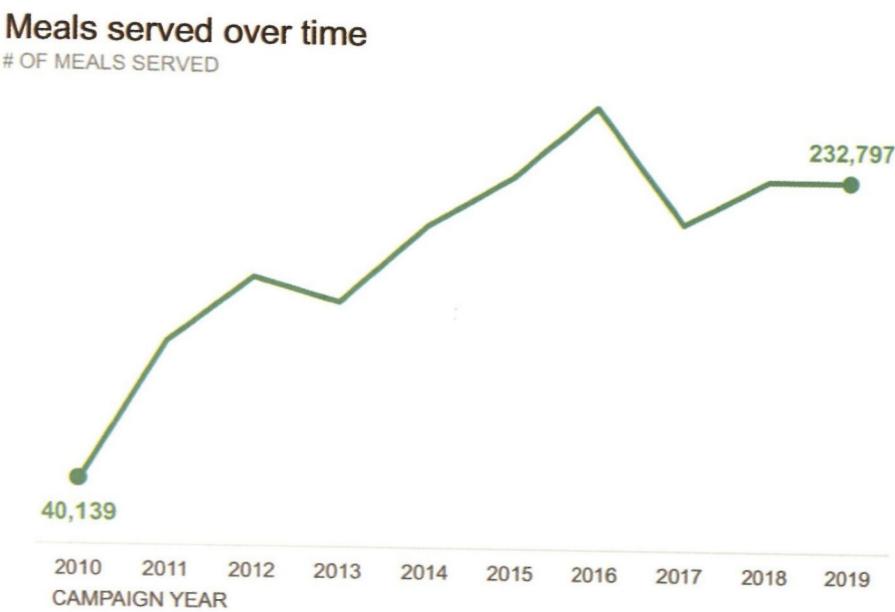


FIGURE 2.2d Line graph

You inevitably made different design choices with your heatmap, bar graph, and line chart, and that's totally fine. The examples here and throughout are meant to be illustrative, not prescriptive. We'll look more specifically at aspects of design in Chapter 5.

STEP 4: Which do I like best? When I look back over the visuals I created, I'm surprised at my own answer to this one. Going into it, I thought for sure I'd prefer the line. It's the cleanest and it takes up the least amount of ink. But seeing them together, and taking the limited context into account, I actually prefer the bar chart (Figure 2.2c). If there is a clear start and end to the program within each year, I'd provide this segmented picture. That said, I do think the overall trend is easier to see with the line graph. Additionally, if there were context that I wanted to annotate via text on the graph, I'd likely choose the line, which has more space to accommodate this.

As we saw in solution 2.1, this is another illustration that there is no single right approach for visualizing data. Two different people faced with the same data visualization challenge may opt for different approaches. Of utmost importance is that we are clear about what we want to enable our audience to see and choose a view that will help facilitate that.

Exercise 2.3: let's draw

One of the best tools we all have at our disposal when visualizing data is a blank piece of paper. If I'm ever feeling stuck or am looking for a creative solution, I get out a fresh sheet and start sketching. You don't need to be an artist to reap the important benefits of drawing. When working on paper, we remove the constraints of our tools (or what we know how to do with our tools). We are also less likely to form attachment to our work (the way we do after we've taken the time to create it with our computer). There's also simply something about empty space waiting to be filled that can help spark creativity.

Let's do a quick practice exercise using this important instrument: paper. The following graph (Figure 2.3a) shows capacity and demand measured in number of project hours over time. It is currently graphed as a horizontal bar chart. But is this the only way to show this data? Certainly not!

Get a blank piece of paper and set a timer for 10 minutes. How many different ways can you come up with to potentially visualize this data? Draw them! (Don't worry about plotting every specific data point exactly—make it quick and dirty to get an overall sense of what each visual could look like.) When the timer goes off, look over your sketches. Which do you like best and why?

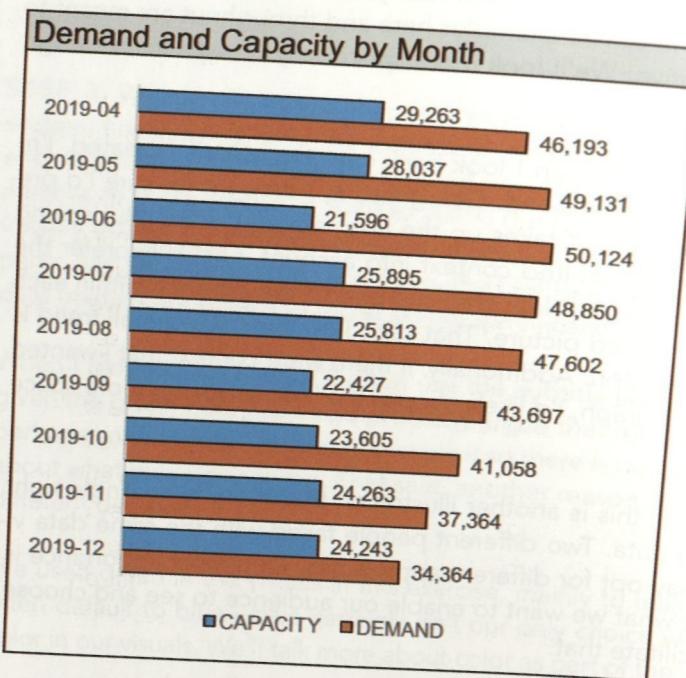


FIGURE 2.3a Let's draw this data!



Solution 2.3: let's draw!

After 10 minutes, my paper is filled with six different ways to depict the data. See Figure 2.3b.

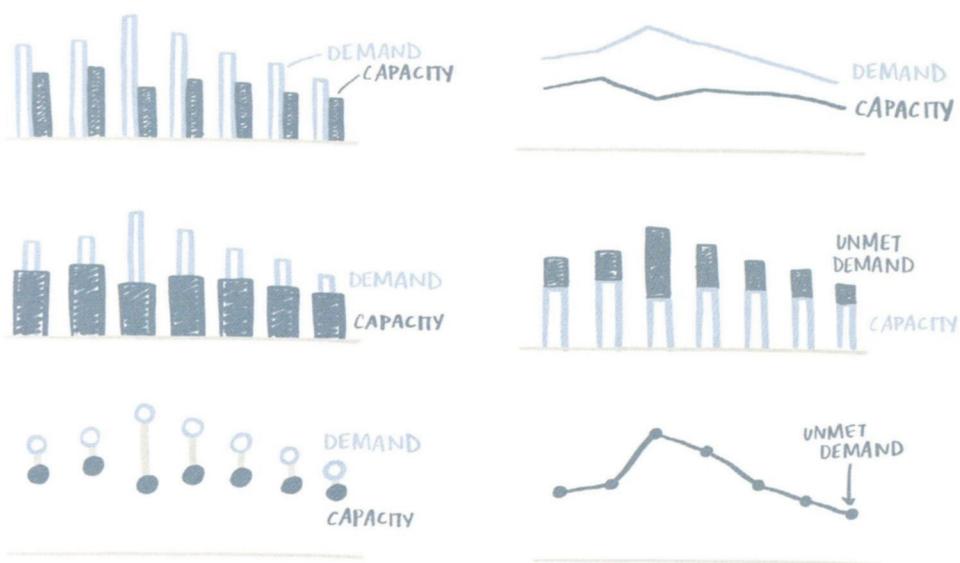


FIGURE 2.3b My data drawings

Starting at the top left, my initial sketch simply turns the horizontal bars upright so that time can move from left to right along the x-axis in a way that is intuitive. My second graph (top right) turns the bars into lines. I find it easier to focus on the gap with this view. But I wanted to do some more playing with bars of various forms, so my third iteration (middle left) goes back to those. I made Demand thinner and behind Capacity, in hopes that this would make it clear how much we're meeting out of the potential of what could be met. As a twist on that, we could stack the bars, which I've done in the middle right picture. The stacked series becomes the Unmet Demand (note that this stacked version only works if Demand is always greater than or equal to Capacity—it would get tricky if Demand were to fall below Capacity). My penultimate view (bottom left) re-envisioned the bars as dots and connects them to bring attention to the difference (this would still work if Demand falls below Capacity so long as we've made the Demand circles distinct from those representing Capacity, for example, by coloring them differently). My final illustration simply plots the trend of Unmet Demand. With this last one, we lose the context of the overall magnitude of Demand and Capacity, but depending on our goals, that may be okay.

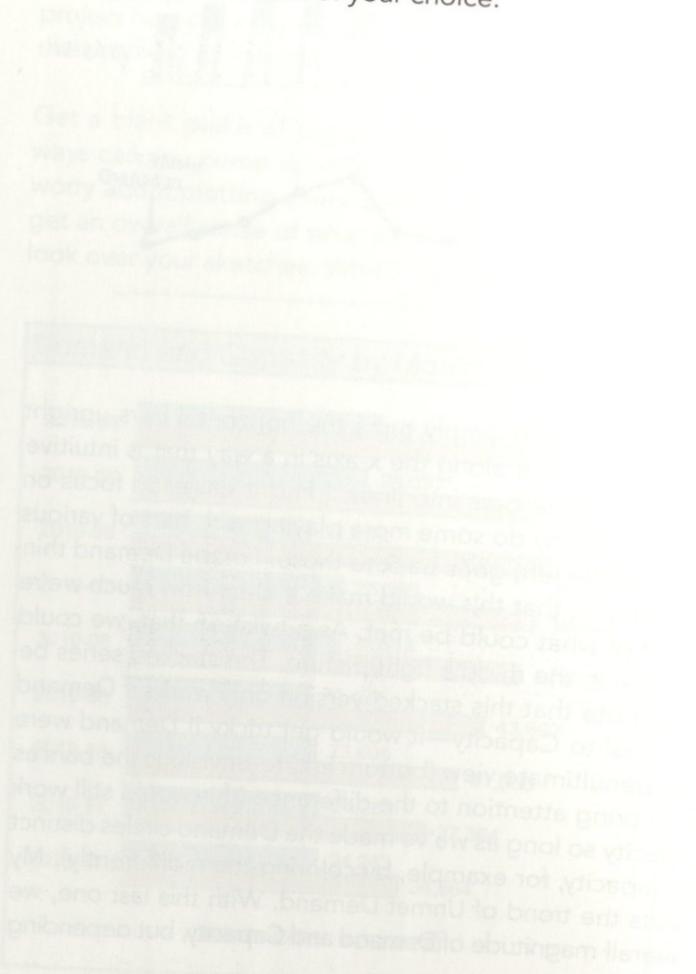
When it comes to which I like best, I prefer the stacked bars (middle right) if Demand is always higher than Capacity, as it is for the data we're graphing. That said, I think any of these views could potentially work. There are definitely other ways

to show this data as well. Compare your drawings to mine. Did you come up with any similar graphs? Where do our ideas differ? Which do you like best out of the full group (yours plus mine)?

Let's continue working with this data and determine how we can make one of the sketches come to life in our tools! Move on to Exercise 2.4.

Exercise 2.4: practice in your tool

Consider the sketches created as part of Exercise 2.3—both the ones you drew and the ones I sketched. Pick one (or more for extra credit!), download the data, and create in the tool of your choice.



Solution 2.4: practice in your tool

I'm an overachiever, so I created all of the views I drew by hand in Excel. See Figures 2.4a - 2.4f.

Basic bars. First is the basic bar graph, or column chart. See Figure 2.4a. I've intentionally filled in Capacity and left Demand as an outline to try to visually differentiate what we're able to meet compared to the unmet capacity. I don't love this graph—and I think I like it less than I did in my drawn version. I appreciate the idea of just having the outline for Capacity, yet I find the outline plus the white space between the bars visually jarring. I also feel this is the view out of all of them that directs the least amount of attention to the gap between Capacity and Demand, which seems like an important aspect of this data.

In this case I chose to use the subtitle space for my legend. I'll sometimes do this if there isn't an obvious place to label the data directly. As an alternative, I could try directly labeling the first or last set of bars and using those as my legend.

Demand vs capacity over time

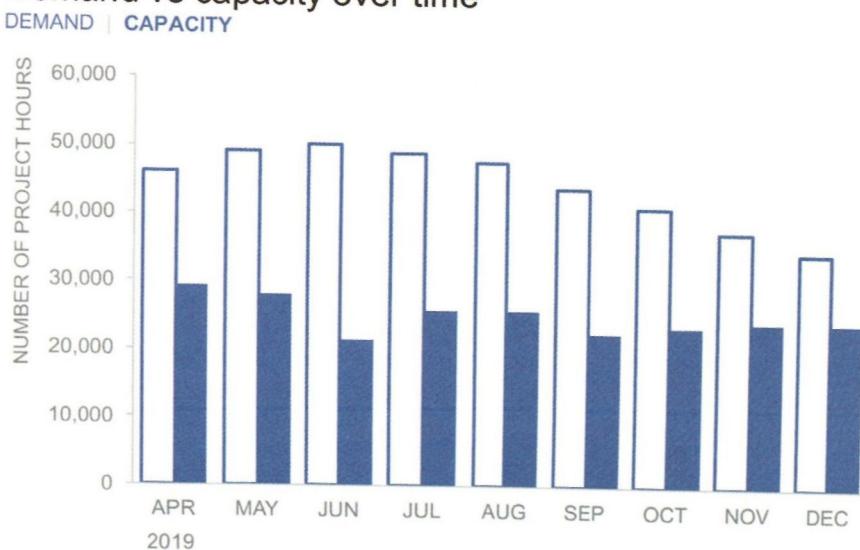


FIGURE 2.4a Basic bars

Line graph. The line graph is a cleaner design compared to the bars, because it simply takes up less ink. I chose to label the lines (and also added data labels) at the ends of the lines, eliminating any confusion over which series is which and reducing the work of going back and forth between a legend and the data. I like that the line allows us to focus on either Capacity or Demand, and it also makes the comparison of them really easy, so we can see the gap between the lines and quickly identify where it is growing and where it is shrinking. I bolded the Capac-

ity line so our attention would go there first, then see the context of the greater Demand. See Figure 2.4b.

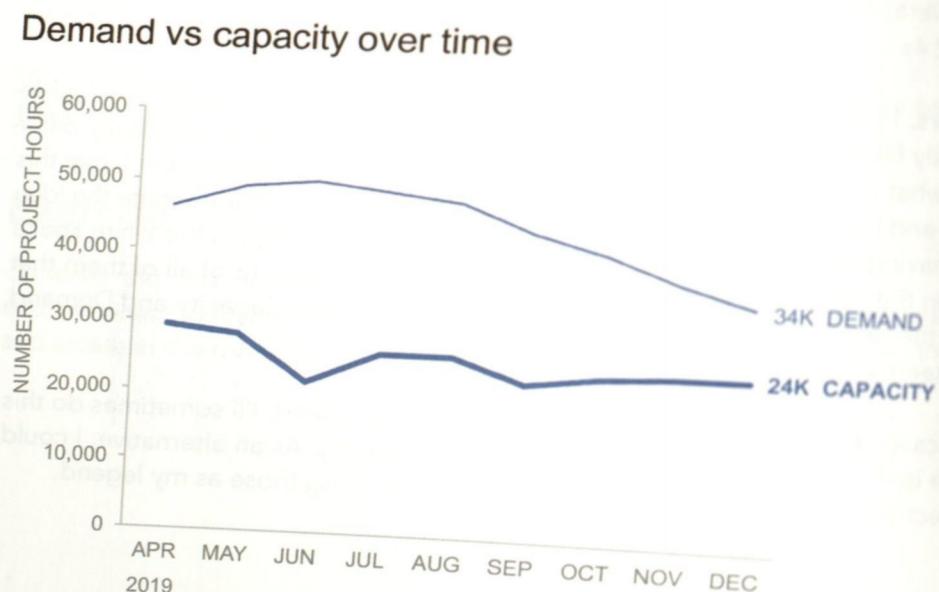


FIGURE 2.4b Line graph

Overlapping bars. We go back to bars in Figure 2.4c, which shows an atypical approach: making the bars overlap. I've made the Capacity series slightly transparent so that it is clear that the Demand series starts at zero and isn't meant to be interpreted as being stacked.

I like this iteration better than I anticipated when I sketched it on paper. That said, I could imagine an audience might find it confusing or off-putting since it doesn't look like a typical bar chart. If I wanted to use this graph, it would be a good one to show to a couple people and get feedback to see whether others find it confusing or if it could get the job done.

Demand vs capacity over time

DEMAND | CAPACITY

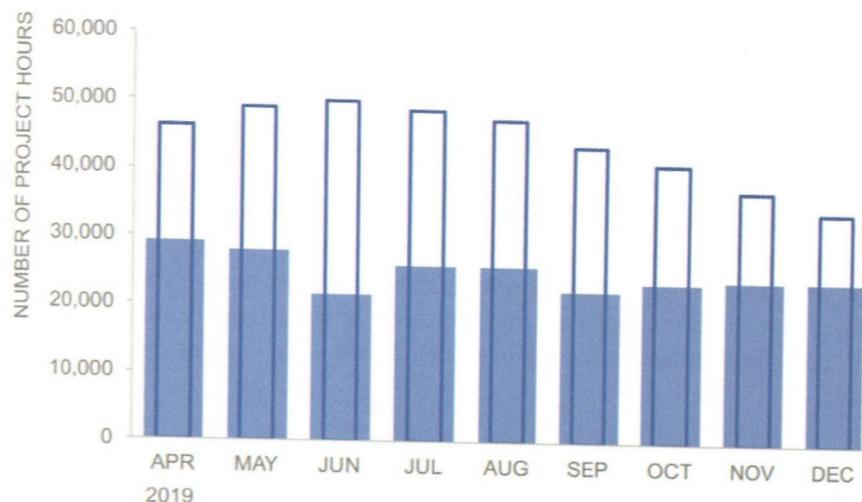


FIGURE 2.4c Overlapping bars

Stacked bars. With stacked bars, I kept Capacity plotted at the baseline, but then changed the second series to Unmet Demand so it could be stacked on top. I switched my emphasis to Unmet Demand, making it blue and rendering Capacity in a light grey. I like this view.

Demand vs capacity over time

CAPACITY | UNMET DEMAND

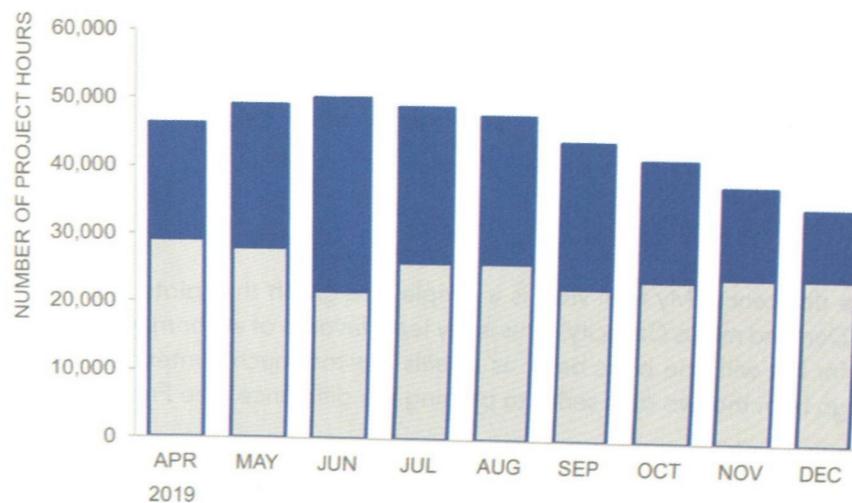


FIGURE 2.4d Stacked bars

Dot plot. This is another view that could catch my audience off guard. It feels intuitive to me, but I have to recognize that pretty much any way I graph this data is going to feel intuitive because I've spent time with the data: I know what it represents and what I want my audience to take away. It may not be as obvious to them, however. Again, soliciting feedback would be a good way to test and assess.

While I'm not sure I love this one, I am impressed at my own Excel wizardry used to create it. The circles are actually data markers on two line graphs (one for Demand, another for Capacity) where I've chosen not to show the actual line and made the data markers huge so I'd have room to center the data label within each point. The shaded region that connects the dots is Unmet Demand, which is a stacked bar that sits on top of a second inclusion of the Capacity series (unfilled so you don't see the bottom series in the stack). This is what I call brute force Excel at its finest!

Demand vs capacity over time

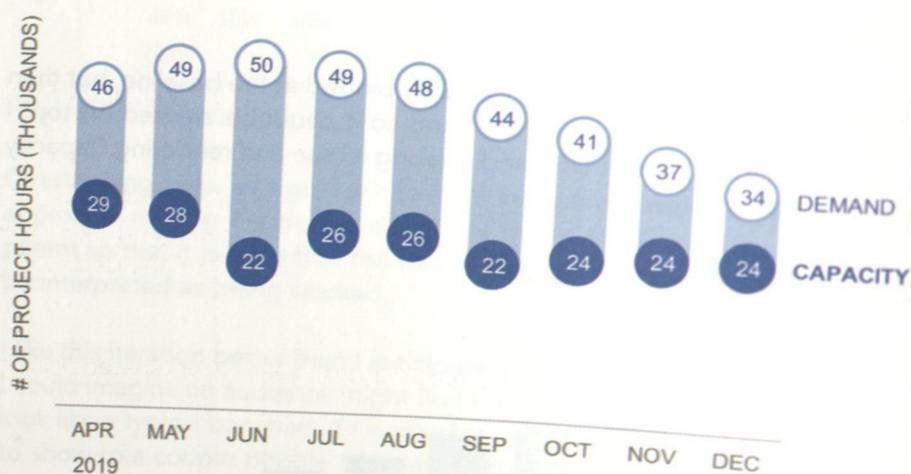


FIGURE 2.4e Dot plot

Graph the difference. My final view is a simple line graph that plots the Unmet Demand (Demand minus Capacity). This is my least favorite of all (or maybe I'd rate it as tied for last with the basic bars), as it feels like too much context is omitted when we go from the two data series to plotting the difference. See Figure 2.4f.

Unmet demand over time

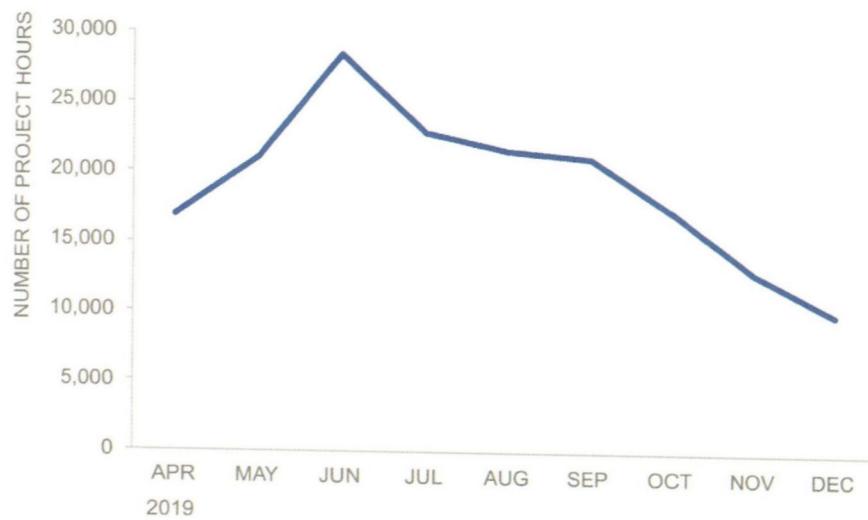


FIGURE 2.4f Graph the difference

How did the visual(s) you made in your tool turn out? Which out of all of them do you like best and why?

In absence of any context, I'd choose the Stacked Bar in Figure 2.4d. I like that it's easy to see both how Unmet Demand and Capacity are changing over time and I appreciate how this view makes it easy to focus attention on the decreasing Unmet Demand.

We will revisit this data in context of the broader dashboard it originated from in Chapter 6.

Exercise 2.5: how would you show this data?

The following table shows attrition rate for a 1-year associate training program for a given company. Spend a moment familiarizing yourself with it, then answer the following questions.

Year	Attrition Rate
2019	9.1%
2018	8.2%
2017	4.5%
2016	12.3%
2015	5.6%
2014	15.1%
2013	7.0%
2012	1.0%
2011	2.0%
2010	9.7%
AVG	7.5%

FIGURE 2.5a Attrition over time

QUESTION 1: How many different ways can you come up with to show this data? Draw or create in the tool of your choice.

QUESTION 2: How would you show the average in the various views you've created?

QUESTION 3: Which of the visuals you've created do you like best and why?

non-monotonicity in broadcast rebroadcast

After the first broadcast, the second broadcast is not rebroadcast

PROBLEMS WITH PREDICTION

Graph the difference between actual and predicted values

Demand (Demand minus Capacity). This is my error

It's hard for me with the basic tools, and it feels
when we go from the two data series to predicting



Solution 2.5: how would you show this data?

QUESTIONS 1 & 2: There are many potential ways we could show this data, depending on our audience and our goals. I came up with six different potential visuals and integrated the Average into each. Let's review and discuss each of these.

Simple text. Just because we have numbers doesn't mean we need a graph! In some situations we can simply communicate a number or two. For example, I could summarize all of this data by saying, "The attrition rate for this program has averaged 7.5% over the past ten years." This doesn't give us any sense of the range over time, or basis for comparison, which in some cases would be simplifying too much. If that's important, perhaps I could say something like, "The attrition rate has varied from 1% to 15% over the past decade, and was 9.1% in 2019." Or if I want to focus on more recent data, which may be more relevant, I could say, "The attrition rate for this program has increased in recent years, from 4.5% in 2017 to 9.1% in 2019."

Each time you create a visual, come up with a sentence that answers the question, "So what?" (Exercises 6.2, 6.7, 6.11, 7.5, and 7.6 will ask you to do this explicitly.) You may find that you can communicate with that sentence, eliminating the need for the graph altogether. When you do have more data you need to communicate, consider what context is helpful and how you can visualize it. Let's look at some ways to graph the data next.

Dot plot. I can use points to illustrate attrition rate (y-axis) by year (x-axis). I incorporated the average by adding a line to the graph, which allows us to easily see when we've been above and below average over time. See Figure 2.5b.

Attrition rate over time

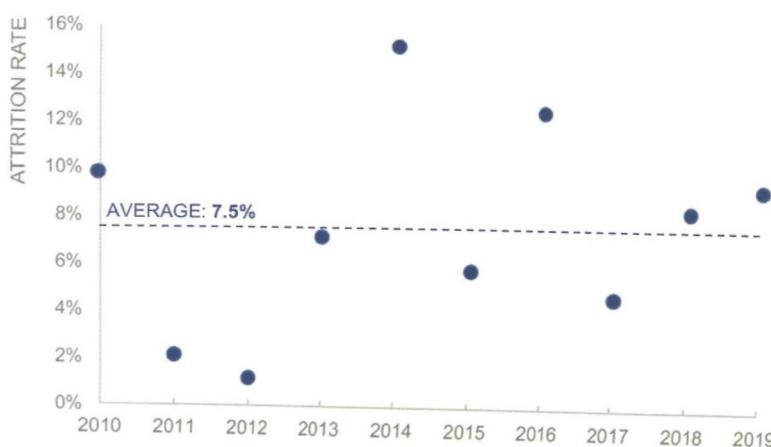


FIGURE 2.5b Dot plot

Line graph. Rather than plot the points, I could visualize the lines that connect them so we can more easily see the trend over time. Figure 2.5c illustrates this. I retained the thin dotted line for the Average, but moved my labeling of it (and also abbreviated) so that it would better fit given this new layout of the data. I also chose to put a data marker and label on the final data point. This makes the comparison between the most recent point of data and the average an obvious one for my audience.

Attrition rate over time

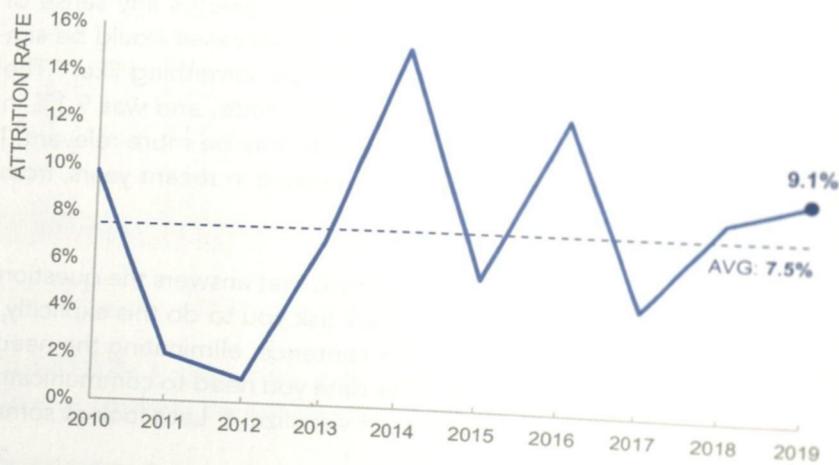


FIGURE 2.5c Line graph

I tried a second iteration on the line graph, using shaded area for the Average, rather than a line. See Figure 2.5d. I prefer the original line view in Figure 2.5c, but I could envision scenarios with different data that might cause me to choose another approach.

Attrition rate over time

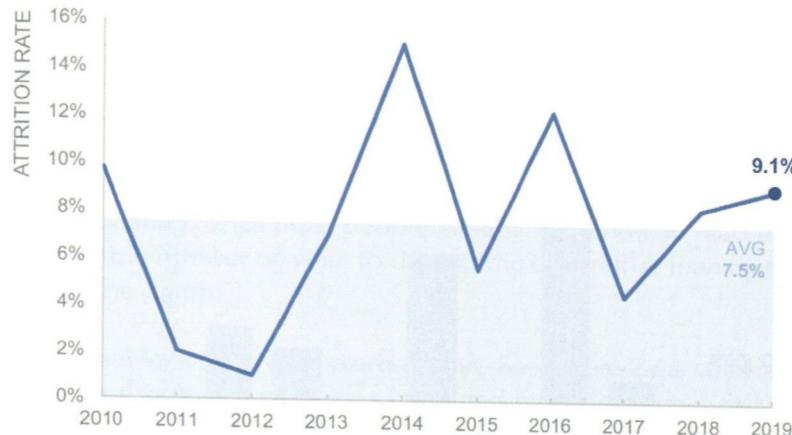


FIGURE 2.5d Line graph with shaded area depicting average

Area graph. After trying out area for the Average, I decided to switch it up and depict the Attrition Rate with area and revert back to a line for the Average. See Figure 2.5e. I chose a lighter blue for the Average line so it would show up both against the empty white background as well as when it overlaps the area encoding attrition rate. In each visual, I've labeled the Average differently—this is mainly due to the space available and the shape of it. Alternate views of the data may cause you to make other design modifications like this as well.

I don't love this one. It takes up a lot of ink for what we're trying to show and makes it seem like there's something important about the area under the curve, which isn't the case here. I don't use a lot of area graphs in general.

Attrition rate over time

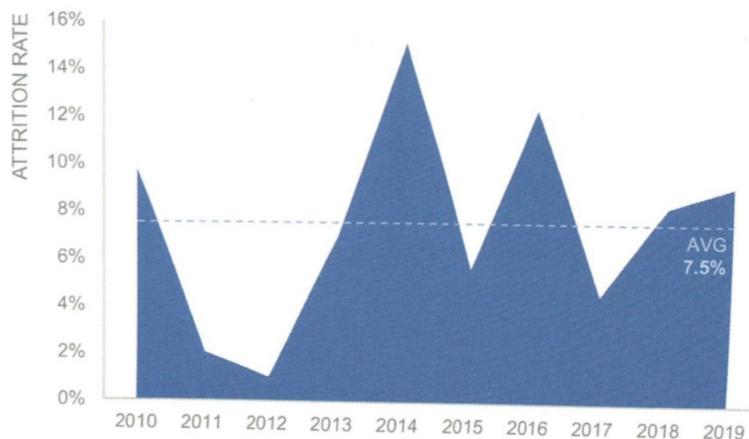


FIGURE 2.5e Area graph

Bar graph. Finally, I tried plotting this data as a bar chart. See Figure 2.5f. I preserved the Average as a line, labeling it again differently from prior views given the layout of the overall graph.

Attrition rate over time

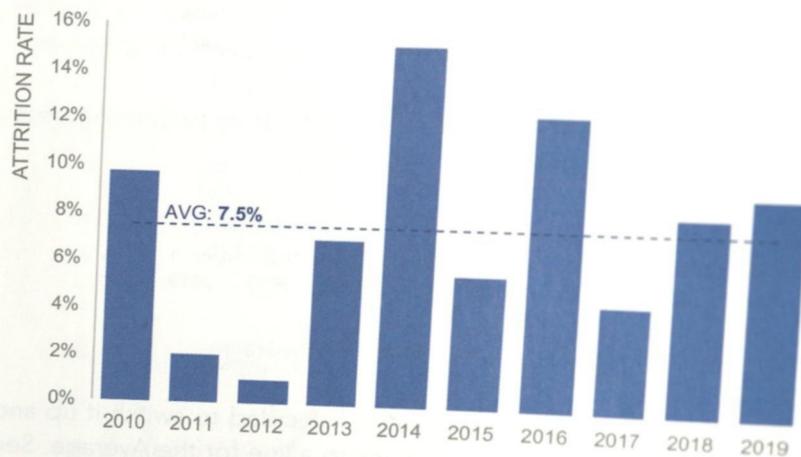


FIGURE 2.5f Bar graph

QUESTION 3: Which do I like best? I'm happy with the preceding bar chart, but it's easy to see the trend in Attrition Rate over time. I can easily compare it to the Average. It doesn't use a lot of ink, which leaves me space to add commentary if it makes sense to do so.

Exercise 2.6: let's visualize the weather

I'm a big fan of bar charts. They are easy to read—our eyes and brains are great at comparing lengths when aligned to a common baseline, which is what the bar chart does for us. We take in the information in bar charts by comparing the relative heights of the bars to each other and the baseline, so it's easy to see which is biggest and by how much. Also the familiarity of bar charts can be useful when communicating: since most people already know how to read them, they can focus their brainpower on what to do with the data rather than try to figure out how to read the graph.

Let's take a look at an example bar chart. See Figure 2.6a, which shows the weather forecast for the next six days measured by the expected daily high in degrees Fahrenheit.

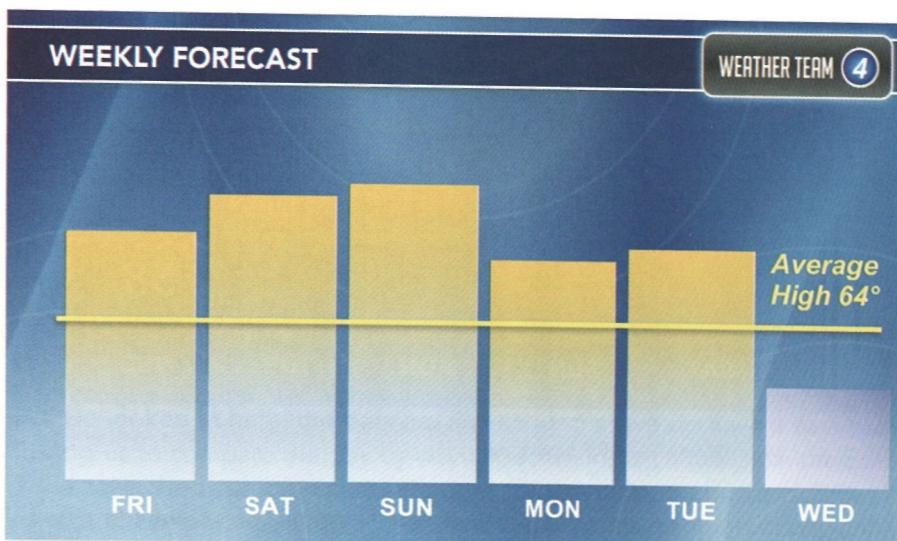


FIGURE 2.6a The weather forecast

QUESTION 1: Imagine you are preparing for a Sunday afternoon at the park. What temperature would you estimate for the high on Sunday?

QUESTION 2: You are planning your children's clothes for the coming week and trying to decide what type of jacket or coat they'll need midweek. What temperature might you estimate for the high on Wednesday?

QUESTION 3: What other observations can you make from this data?

Solution 2.6: visualize the weather

While the temperature may appear to be somewhere in the 90s on Sunday and in the 40s on Wednesday, that's not actually the case. Let's take a closer look.

It turns out that Sunday is 74 degrees while Wednesday is 58 degrees. How is that possible? The initial graph in Figure 2.6a does not have a y-axis that starts at zero. Rather, it begins at 50. This distorts the data, making it so we can't accurately compare the temperature day to day. See Figure 2.6b, which adds both the y-axis and data labels to the original graph.

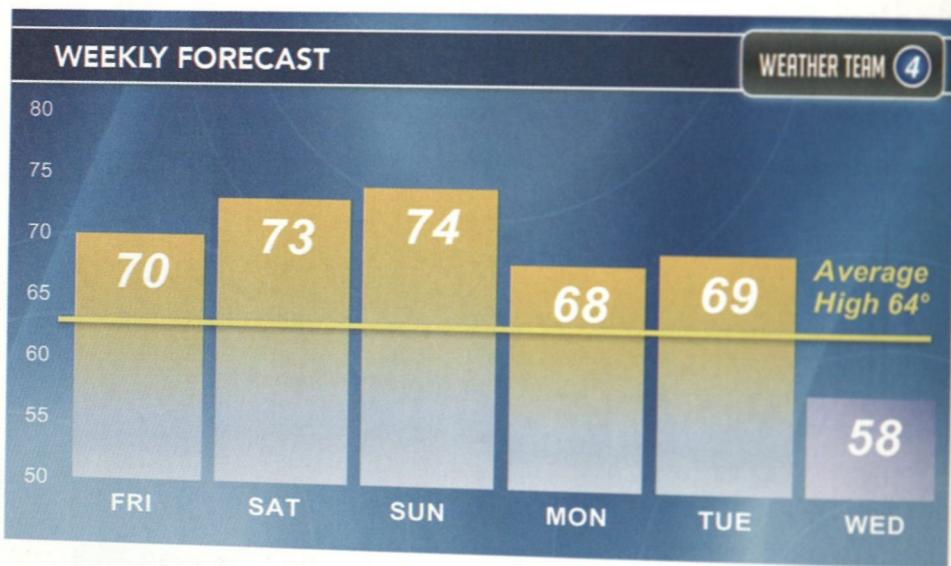


FIGURE 2.6b Bar charts must have a zero baseline!

Let's redesign the graph to start the y-axis at zero. Figure 2.6c shows the side-by-side. Notice the difference this makes in interpreting the data.



FIGURE 2.6c Let's compare the two graphs

What looked like a large deviation from the average on the left of Figure 2.6c looks comparatively a lot smaller on the right. With this view, you'd likely make a different decision when it comes to how thick the kids' coats should be on Wednesday!

There aren't a lot of hard-and-fast rules when it comes to visualizing data. But there are a few, and we've just witnessed one of them broken: bar charts must have a zero baseline. Because of the way our eyes compare the endpoints of the bars to each other and the baseline, we need the context of the full bar there in order to make that an accurate visual comparison.

There are no exceptions.

That said, this is not a rule that applies to all graphs. With bars, you can't chop or zoom because of the way we compare the ends of the bars relative to each other and the axis. But with points (scatterplots or dot plots) or lines (line graphs, slope-graphs), we focus primarily on the relative positions of the points in space, and in the case of line graphs, the relative slopes of the lines that connect the points. Mathematically, as we zoom, the relative positions and slopes remain constant. You still want to take context into account and avoid overzooming and making minor changes or differences look like a big deal. Though sometimes minor changes or differences are a big deal, so if you find yourself needing to change the axis to highlight this, reach for points or lines, not bars.

On a related note, I've heard the idea raised that a zero baseline for weather doesn't make sense, since temperatures can be negative, and zero (particularly on a Fahrenheit scale) isn't meaningful. In the case of a short-term weather forecast, like we looked at here, the bars are fine so long as we do have a zero baseline allowing us to compare the day by day expectations accurately. On the other hand, if we take the case of climate change, for example, a couple degrees change in global temperatures—which is nearly impossible to see in bars with a zero baseline—is meaningful. This isn't a good argument for changing to a non-zero baseline bar chart, but rather for not using bars to illustrate this data. We could shift to a line graph or graph the change in temperature instead of absolutes to bring focus to the small but meaningful differences. As always, we should step back and think critically about what we want to show, then choose an appropriate visual to facilitate this.

Exercise 2.7: critique!

Speaking of points (mentioned in the solution to the previous exercise), let's take a look at some next, in the context of critiquing a less than ideal graph.

See Figure 2.7a, which is a dot plot showing the bank index over time for a number of national banks. Assume you work at Financial Savings.

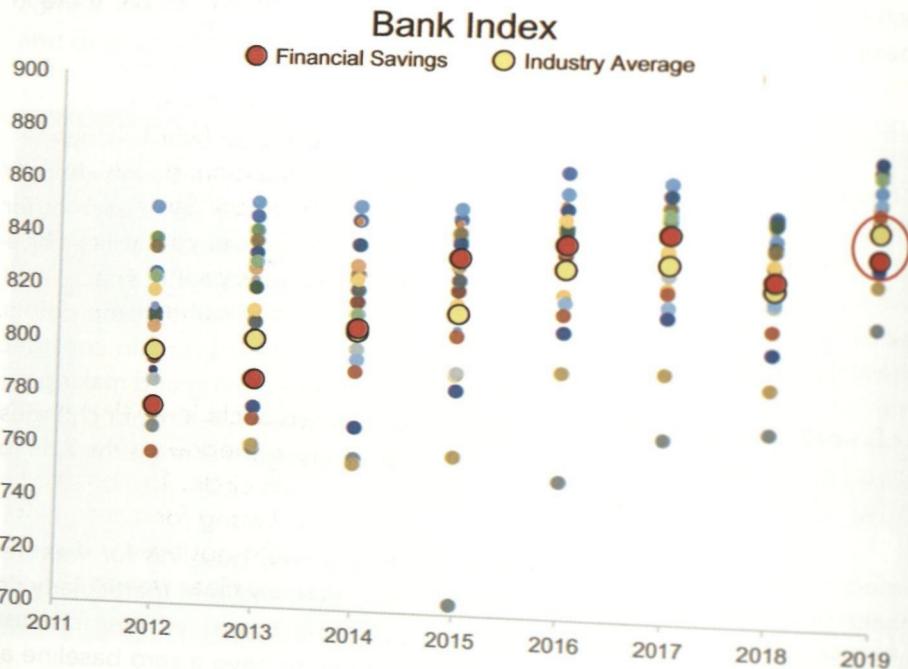


FIGURE 2.7a Bank index

QUESTION 1: What questions do you have about this data?

QUESTION 2: If you were designing the graph, what changes would you make? How would you visualize this data?

Solution 2.7: critique!

QUESTION 1: This graph seems to incite many more questions than it answers! My first question is: what exactly is the metric being plotted? I might assume "Bank Index" is some type of customer satisfaction score, and that the higher the number, the better. But what if this is really something like bank teller errors? I'd interpret that data very differently.

My next question is, do we need all of the data? We can see at the top that the red and yellow data points represent our company (Financial Savings) and the Industry Average, respectively (these also seem like odd color choices, though I guess they are bright in an attempt to stand out against all of the other color in this graph). I assume all of these dots roll up into the average (which is another question I have: is that the case?). This begs yet another question: do we need all of those individual data points or would showing only Financial Savings and the Industry Average work? When you consider getting rid of data, you always want to think through what context you lose when doing so. Here, by summarizing with the average, we'd lose line of sight to the spread across competitors. Depending on our goals, this may or may not be important.

In terms of other questions, I'm also curious what the red circle in 2019 is meant to highlight. I appreciate the thought process behind it: someone looked at this data and thought "I'd like you to look here" and drew a red circle. This presents a couple of challenges, however. First, there's so much competing for our attention in the graph with all the various colored dots that we might not even notice the red circle. Second, when we do notice it, it isn't immediately clear what it is trying to point out to us.

My final questions are: So what? What does this data show us? What's the story?

QUESTION 2: Let's shift from asking questions to redesigning how we show this data. It turns out the metric being graphed is branch satisfaction, where the higher the number, the better. I'll assume that we care most about how Financial Savings compares to the Industry Average. Simply making that decision means I can declutter this graph a ton and focus on the data points for Financial Savings and the Industry Average.

Speaking of data points, this data is over time. We can plot it as points, but I'd be apt to connect the points and display this data in a line graph. Lines will help us more easily see the change over time and can also help highlight interesting things when it comes to how these lines interact with each other: if one is always above the other, the lines will help us see the gap. If that's not the case, lines will help us see when one series crosses the other, which will be interesting as we try to answer the question, "So what?"

Figure 2.7b shows my makeover of this visual.

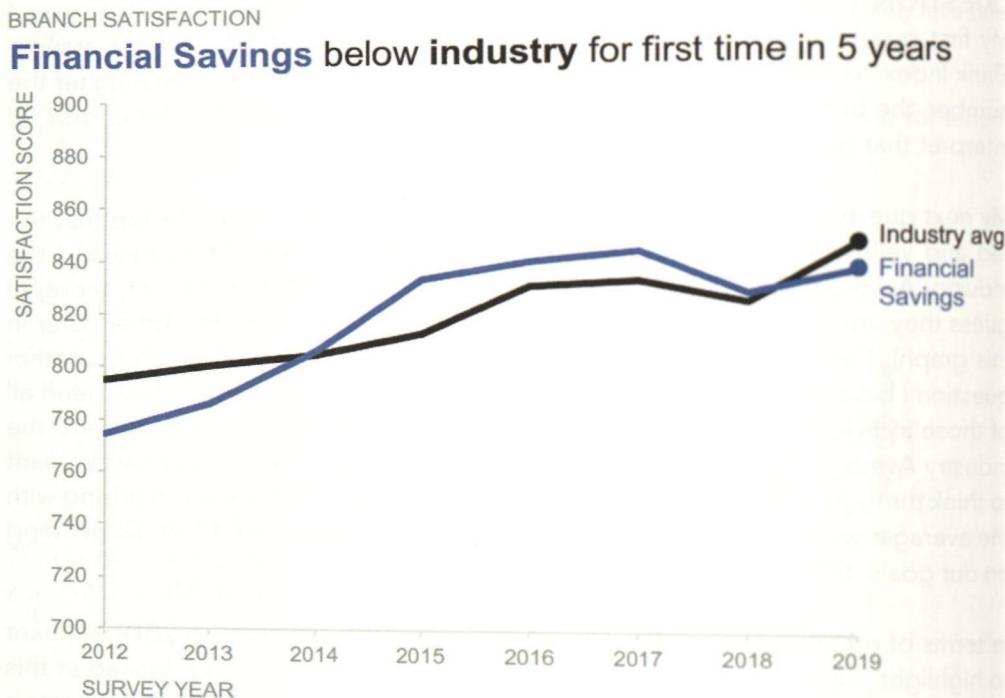


FIGURE 2.7b Revamped graph

Decluttering and changing the graph to lines helps us focus on the data. I've titled and labeled everything directly, so there's no need to make assumptions or hunt around for how to interpret the data. I used the title space to answer the question, "So what?"

With an additional understanding of what's driving the ups and downs in satisfaction for our business and the industry, I could take this even further. In a live meeting or presentation, I could build it line by line or time point by time point, which would allow me to focus my audience's attention as I talk through relevant context. If it needs to stand on its own, I could put text directly on the graph to annotate what's causing the changes we see. We'll look at a number of examples that employ these strategies as we get further along. We'll revisit this example and look at a scenario where we keep all of the original data in Chapter 4.

Next, let's redesign another graph.

Exercise 2.8: what's wrong with this graph?

Sometimes, we design a graph with the best of intentions, but inadvertently make things difficult for our audience. Let's take a look at an example where this is the case and discuss how we can improve it.

Continuing in the banking industry, next let's imagine you work as an analyst in consumer credit risk management. For those who may not be familiar—when people take out loans, some portion of those people don't repay them. These loans move through various levels of delinquency: 30-days past due, 60-days past due, and so on. Once they become 180 days past due, they are categorized as "Non-Performing Loans." After reaching this stage of delinquency, despite collection activities, many still don't repay and this results in a loss. Banks have to reserve money for these potential losses.

Now that we're past credit risk 101, let's talk about the data. You've been asked to create a graph showing how Non-Performing Loan (NPL) volume compares to the Loan Loss Reserves over time. Look at Figure 2.8a. Take note of how your eyes move as you process this information. What is confusing about this graph? How would you improve it?

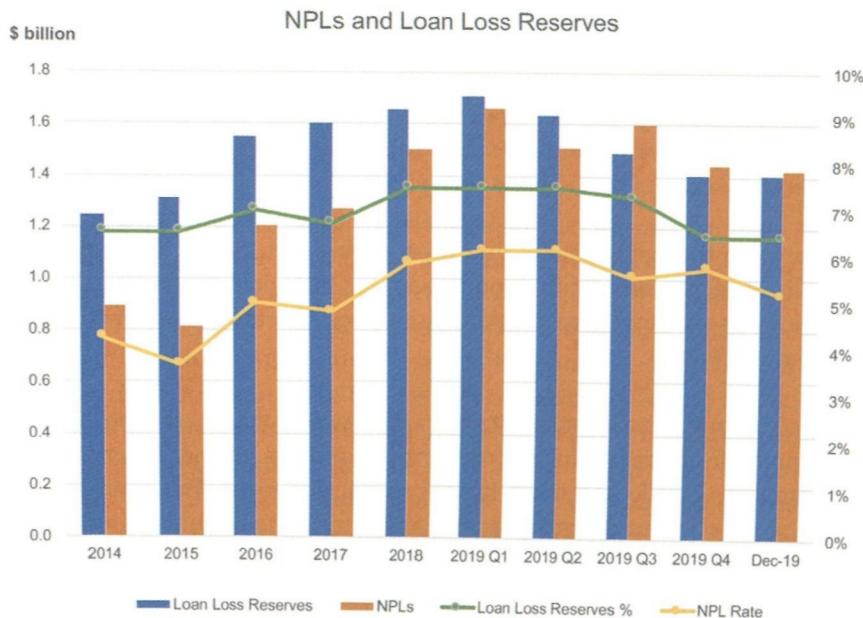


FIGURE 2.8a What is confusing in this graph?

Solution 2.8: what's wrong with this graph?

In considering how I process this data, I start out by doing a lot of back and forth between the bars, the lines, and the legend at the bottom of the graph to try to interpret the data. I scan the y-axes to see what those are. After some reading and thinking about it, I figured out that the lines—Loan Loss Reserves % and NPL Rate—are meant to be read against the secondary y-axis on the right-hand side of the graph. This means the bars—Loan Loss Reserves and NPLs—should be read against the primary y-axis on the left. This seems more difficult than it needs to be.

Going back to the Loan Loss Reserves % and NPL Rate: I'm not sure what the denominator is. I might assume that it's the total loan portfolio—but I'd rather this be made clear so I don't have to assume! I'm also unsure these lines add any value—they aren't adding any new information. There may be additional context that would cause me to make a different decision, but in the absence of that I'm going to focus on the volume—the dollars—and not confuse things by also showing the rate. As a bonus, this decision will eliminate the secondary y-axis (I recommend against the use of dual y-axes in general; for alternate approaches to the secondary y-axis, refer to Chapter 2 in SWD).

It's only after all this that I start paying attention to the x-axis and notice the biggest issue: we have inconsistent time intervals. Upon first glance (and several thereafter—perhaps you didn't even catch this problem), I started reading the x-axis to see it is in units of years and then assumed that continues to be the case as we move from left to right. When we read each label, however, we find that after 2018, the time interval changes to quarters, and after Q4 it seems we've broken out December on its own. This is not good!

I can appreciate the thought process that presumably led to this. December is probably the most recent month. Showing years for historical context is helpful, but then it's nice to also show greater granularity (e.g. quarterly, monthly) for the more recent time periods.

Sometimes inconsistent time intervals are a reality—we might be missing data or simply have something that occurs inconsistently over time. In that situation, we need to denote that visually and make it clear to our audience. The same bar or line shouldn't be used to represent a year and a quarter, as this can too easily lead to incorrect interpretation and false observations.

We have a couple of options for overcoming this challenge. If we have all of the quarterly data, I'd be apt to just plot that. Bars will get messy simply because there would be so many of them, but given that we're getting rid of the two data series that were originally depicted by lines, we could swap the original bars and plot the volume with lines. Or if for some reason we don't want to or can't show the quarterly data, and leave it all in one graph, one option would be to space the

x-axis such that each year takes up the same width that the four quarters together take up. If I did this, I would exclude year 2019 so there isn't redundancy between that and the four quarters that are broken out separately.

As another alternative, we could split this data into two graphs: one to show the annual data for 2014 through 2019, and then a second to break out the quarterly data just for 2019. This would allow me to title each explicitly and make the difference in time components clear. I would also compress the quarterly data more than the annual data to help visually reinforce the shorter time periods. The makeover incorporating these changes is shown in Figure 2.8b.

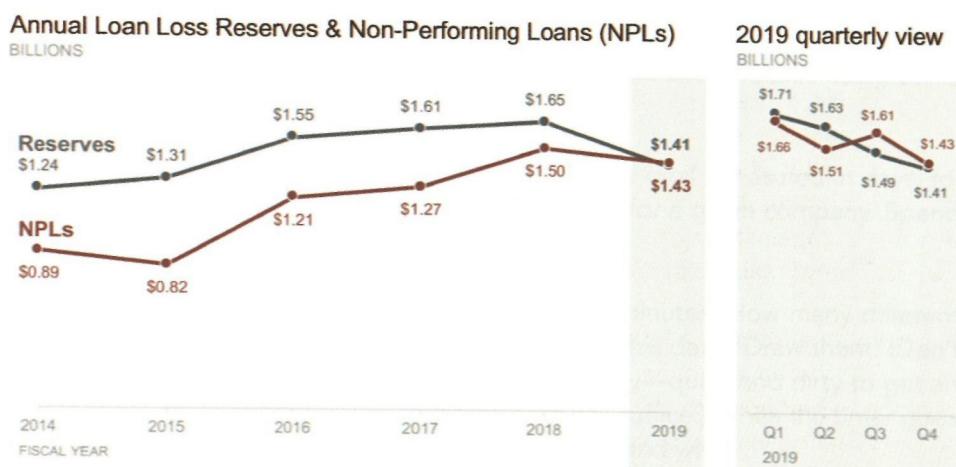


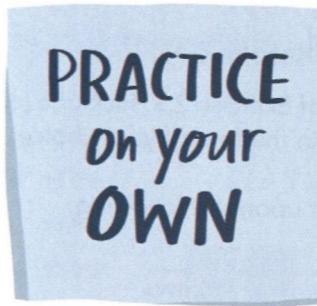
FIGURE 2.8b An alternative view

In this case, I chose to label the data directly so we can compare the volume of NPLs to that of Reserves without the work of having to estimate it from an axis. I maintained two decimal places of significance so there wouldn't be instances where two points of different heights shared the same value (for example, the third and fourth points on the Reserves line would both round to \$1.6M, which could cause confusion since the heights are visibly different) and so we can easily interpret the smaller but meaningful differences in the recent quarterly numbers. I used shading to tie the final data point in the first graph (2019) to the quarterly breakout for 2019 in the right graph. It is very important in showing the data like this to ensure that the y-axis minimum and maximum are set to be the same amount across the two graphs, so that the audience can compare the quarterly data points to the annual ones according to their relative height.

With this visual, I've taken away a lot of things that were making us do work in the original. Instead of trying to understand the graph, we can focus on the data. I can see that the gap between NPLs and our Loan Loss Reserves has narrowed markedly over time. Both were increasing through 2018, but decreased in 2019. 2019 marks the first time that NPL volume exceeds the Loan Loss Reserves. On a

quarterly basis, this happened in Q3 and Q4. This seems pretty important, so we should probably take some action!

Now that you've practiced with me, it's time to tackle some additional examples on your own.



Grab a pencil and paper! Let's draw some more, iterate in our tools, and improve less-than-ideal visuals through additional exercises.

Reminder:

The data and graphs can be downloaded at storytellingwithdata.com/letspractice/downloads

Exercise 2.9: let's draw

As illustrated in Exercise 2.3, some of our best tools for figuring out how to show our data are a blank piece of paper and pen or pencil. Let's practice using these important instruments!

The following data shows the average time to close a deal (measured in days) for direct and indirect sales teams across four products for a given company. Spend a moment to familiarize yourself with this data.

Get a blank piece of paper and set a timer for 10 minutes. How many different ways can you come up with to potentially visualize this data? Draw them! (Don't worry about plotting every specific data point exactly—quick and dirty to get an overall sense of what each visual could look like will suffice.) When the timer goes off, look over your sketches. Which do you like best and why?

Average time to close deal (days)

Product	Direct Sales	Indirect Sales	Total Sales
A	83	145	128
B	54	131	127
C	89	122	107
D	90	129	118

FIGURE 2.9a Average time to close deal

As part of this, what assumptions are you making about this data? What additional context do you wish you had?

Exercise 2.10: practice in your tool

STEP 1: Refer back to the sketches you created as part of Exercise 2.9. Pick one (or more for extra credit!), download the data, and create in the tool of your choice.

STEP 2: After creating your graph(s), pause and reflect upon the following.

QUESTION 1: What was helpful about sketching?

QUESTION 2: Did you find anything about the drawing process annoying or frustrating?

QUESTION 3: Was creating a graph in your tool different after first sketching it?

QUESTION 4: Can you envision using this approach (draw options first, then create in your tool) in the future? In what situations?

Write a few sentences summarizing your thoughts.

Exercise 2.11: improve this visual

Imagine you work for a regional health care center and want to assess the relative success of a recent flu vaccination education and administration program across your medical centers.

You have a dashboard where related metrics are reported and your colleague pulled the following visual from it. Take a moment to study Figure 2.11 and answer the following questions.



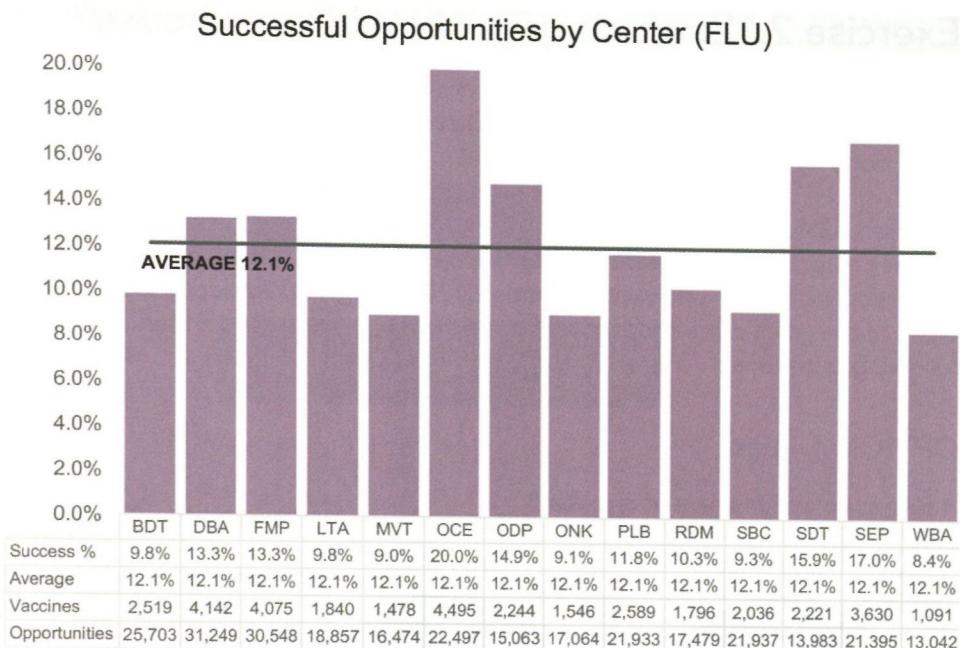


FIGURE 2.11 Original visual from dashboard

QUESTION 1: How is the data sorted? How else could we sort it? In what circumstances would you make a different decision about how to order the data?

QUESTION 2: There is currently a horizontal line to show the average. How do you feel about this? How else could you show the average?

QUESTION 3: What if there were a target—how might you incorporate it? Assume the target is 10%. How would you show this? Now assume the target is 25%. Does this change what you would show or how you would show it?

QUESTION 4: The graph contains a data table. Do you find this effective? What are the pros and cons of embedding a data table within a graph? Would you keep it or eliminate it in this case?

QUESTION 5: The graph currently shows the proportion who received the vaccination. What if you wanted to focus on the opportunity—the proportion who did not receive the vaccination—how could you visualize this?

QUESTION 6: How would you graph this data? Download it and create your ideal view in the tool of your choice.

Exercise 2.12: which graph would you choose?

Any set of data can be graphed many ways and varying views allow us to see different things. Let's look at a specific instance of numerous graphs plotting the same data.

You are visualizing data from your employee survey and want to show how employees responded this year compared to last year to the retention item "I plan to be working here in one year." Figures 2.12a through 2.12d depict four different views of the exact same data. Spend some time examining each, then answer the following questions.

OPTION A: pies

"I plan to be working here in one year"

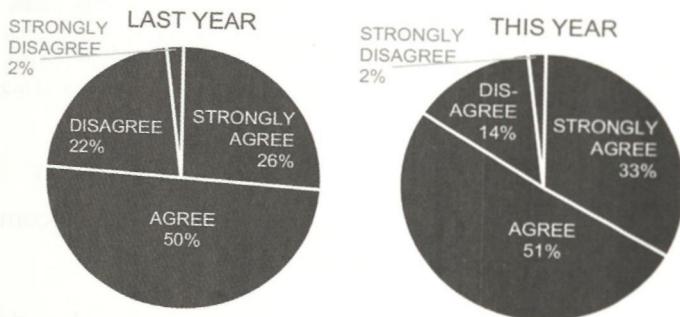


FIGURE 2.12a Pies

OPTION B: bars

"I plan to be working here in one year"

LAST YEAR | THIS YEAR

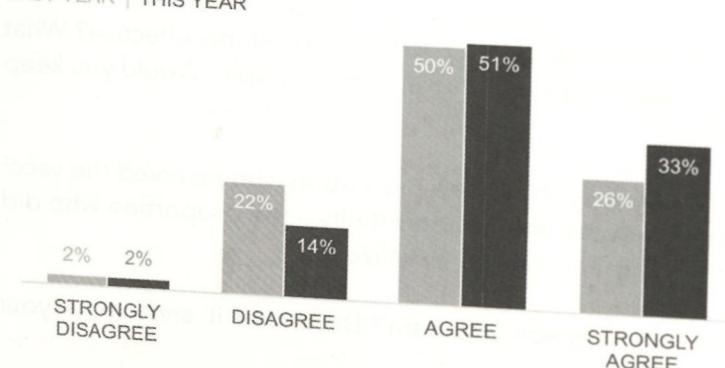


FIGURE 2.12b Bars

OPTION C: divergent stacked bars

"I plan to be working here in one year"



FIGURE 2.12c Divergent stacked bars

OPTION D: slopegraph

"I plan to be working here in one year"

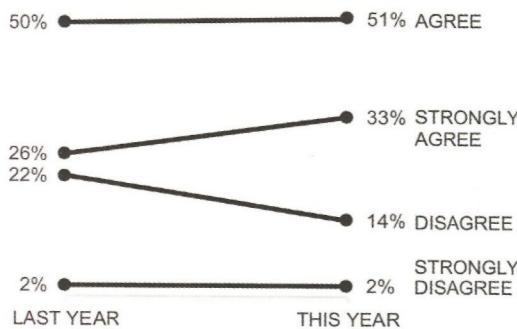


FIGURE 2.12d Slopegraph

QUESTION 1: What do you like about each graph? What can you easily see or compare?

QUESTION 2: What is difficult about the given view? Are there limitations or other considerations of which to be aware?

QUESTION 3: If you were tasked with communicating this data, which option would you choose and why?

QUESTION 4: Grab a friend or colleague and talk through the various options together. Do they agree with your preferred view, or is there a preference for another? Did your discussion highlight anything interesting that you hadn't previously considered?

Exercise 2.13: what's wrong with this graph?

Consider Figure 2.13, which shows response and completion rates for an email marketing campaign where email recipients were asked to complete a survey.

STEP 1: List three things that are not ideal about this graph. What makes it challenging?

STEP 2: For each of the three things you've listed, describe how you would overcome the given challenge.

STEP 3: Download the data. Create your visual that puts into practice the strategies you've outlined.

PRACTICE on your OWN

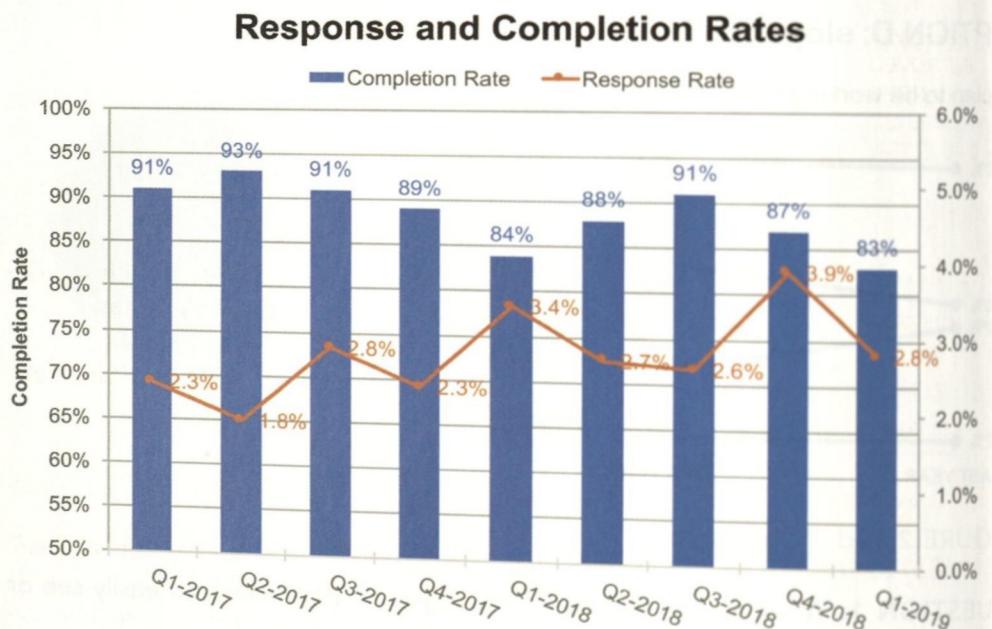


FIGURE 2.13 What's wrong with this graph?

Exercise 2.14: visualize & iterate

As we've seen through a number of examples so far, visualizing data—when done well—can help us spark a magical "ah ha" moment of understanding in our audience. But often it takes iterating—looking at the data numerous ways—both to better understand the nuances of the data and what we want to highlight, as well as to figure out a way that will work for our audience. Let's practice visualizing and iterating.

Imagine you work at a medical device company and are looking at data that shows patient-reported pain levels when a component of a certain device is turned on and when it is turned off. Figure 2.14 shows the data.

Patient-reported pain

PAIN LEVEL	DEVICE SETTING	
	ON	OFF
IMPROVED	58%	36%
UNCHANGED	32%	45%
WORSENERD	10%	19%
TOTAL	100%	100%

FIGURE 2.14 Let's visualize & iterate

STEP 1: Make a list: how many potential methods can you come up with to visualize this data? What different graphs might work? List as many as you can.

STEP 2: From the list you've made, create at least four different views of this data (draw or realize in the tool of your choice).

STEP 3: Answer the following related questions:

QUESTION 1: What do you like about each visual? What is easy to compare?

QUESTION 2: What considerations or limitations should you take into account with each?

QUESTION 3: Which view would you use if you were communicating this data?

Exercise 2.15: learn from examples

A great deal can be learned from the data visualizations that others create—both the good and the not so good. When you see a nice graph, pause and reflect: what makes it effective? What can you learn from it that you can apply in your own work? When you see a not-so-good example, do the same thing: stop and examine what was not done well and how you can avoid similar issues in your own work. Let's practice learning from examples.

Find a graph from the media that's done well and another that is less than ideal. Answer the following questions for each of these examples.

QUESTION 1: What do you like about it? What makes it effective? Make a list!

QUESTION 2: What do you not like about the example? What limits its effectiveness? How would you approach it differently?

QUESTION 3: What learnings from this process can you generalize to guide your future work?

PRACTICE on your OWN

Exercise 2.16: participate in #SWDchallenge

One of the best ways to learn is to do. The #SWDchallenge is a monthly challenge where readers of our blog practice and apply data visualization and storytelling skills. You can take part, too! Think of it as a safe space to try something new: test out a new tool, technique, or approach. Everyone is encouraged to participate and all backgrounds, experience levels, and tools are welcome.

We announce a new topic at storytellingwithdata.com at the beginning of each month. Participants have a set amount of time to find data and create and share their visual and related commentary. Historically, the focus has been on different graph types, but we sometimes change it up with a tip to try or a specific topic. This is meant to be a fun reason to flex your skills and share your work with others.

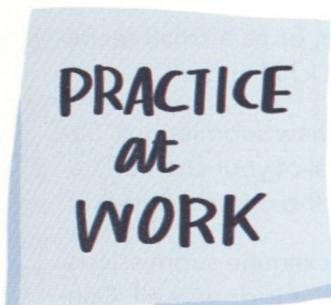
All submissions received by the deadline are shared in a recap post later in the given month. The monthly challenges and recap posts are archived at storytellingwithdata.com/SWDchallenge.

There are a number of exercises you can undertake related to this challenge. Visit storytellingwithdata.com/SWDchallenge and tackle one (or more!) of the following:

- **Participate!** Take part in the live challenge by creating and sharing your work. Or choose a past one as inspiration to flex your data visualization

skills. This can be done on your own, with a partner, or as a small team. Share your creations on social media, tagging #SWDchallenge.

- **Emulate!** Pick a recap post from the archives and review submissions. Select a visual you like and work to recreate it in the tool of your choice. Are there any aspects you would tackle differently than the original author?
- **Critique!** Select a recap post from the archives and examine submissions. Pick three you believe are effective and describe what is done well. Consider how you could generalize the learnings to apply to your own work. Pick three designs you believe are not ideal and reflect on the issues you see and how you could overcome them. What common challenges can you generalize from this that you can overcome in your own work?
- **Run your own challenge!** Get a group of colleagues or friends together, pick a past challenge (or create your own), and run your own version: everyone has a set amount of time to find data and create their visual. Share them with each other. Get together to discuss, giving everyone the opportunity to share their creation and receive feedback from others. Examine what you learn from this process that you can apply to future work. See Exercise 9.4 for more on how this fun process can help cultivate a feedback culture.



Next, let's explore how we can put the various lessons learned related to choosing an effective visual to use at work, including questions to ask yourself and how to get feedback.

Get a project in mind and tackle the following exercises!

Exercise 2.17: draw it!

Consider a current project where you need to visualize data. Grab a blank piece of paper and a pen or pencil. Set a timer for 10 minutes and see how many ideas you can sketch out for how to potentially show your data.

When the timer goes off, step back and take inventory of what you've created. Which view(s) do you like best? Why is that?

Show your sketches to someone else. Explain to them what you want to communicate. Which visual(s) do they like best? Why is that?

If you ever feel stuck or are looking for an innovative approach and having trouble coming up with one on your own, grab a conference room with a whiteboard and a creative colleague or two. Talk them through what you want to show. Start drawing and redrawing. Debate as you mock up the different views: what works well? What is lacking? Which visual(s) are worthy of creating in your tools? Can you do it on your own, or if not, what or who can help you realize your ideas?

Exercise 2.18: iterate in your tool

Allowing yourself time and flexibility to iterate through different views of your data allows you both to better understand the nuances and determine which way of showing the data might help you achieve that magical "ah ha" moment of understanding that you seek in your audience.

Take some data you'd like to visualize. Open your favorite graphing tool and start creating different visuals. How many ways can you come up with to look at the data? Set a timer for 30 minutes and iterate to create different views of the data in your graphing application.

When the timer goes off, assess for each: what are the pros and cons? What do you want to enable your audience to see? Which iteration(s) will facilitate this? If unsure, jump to Exercise 2.21, which provides some tips for soliciting feedback from others.

Exercise 2.19: consider these questions

When you create a graph, it's not surprising that it makes sense to you. You are familiar with the data, and you know what to look at and what is important. Don't assume this is necessarily true for your audience. After you've made your graph, ask the following questions to help determine whether further iteration is necessary.

- **What are you trying to show?** What do you want to enable your audience to do with your data? Does the visual you've created facilitate this? What takeaways are easiest to see? What comparisons are easiest to make? What things are harder to do given the way you are showing the data?
- **How important is it?** Is this a critical issue or merely something that people might find interesting? What are the stakes? Is it a scenario where quick and dirty is okay? What level of perfection is warranted? What level of accuracy is required?
- **Who is your audience?** Is your audience familiar with the data you are presenting or is it new? Does it fit in with their preconceived notions, or may it challenge a held belief? Does your audience expect the data to be presented in a certain way? What are the pros and cons of following the norm in this situation compared to doing something new or unexpected? What questions will your audience have and how can you anticipate and be prepared to seamlessly address them?
- **Is your audience familiar with the type of graph?** Anytime we use something less familiar to our audience, we are introducing a hurdle: we either have to get them to listen to us long enough to tell them how to read the graph, or get them to spend enough time with it to figure it out on their own. If you're using something less familiar, have a good reason for it. Does this view let your audience easily see something that would otherwise be difficult, or create a new insight that isn't possible with more familiar ways of graphing the data? Consider also: how much time do you want to spend talking about the graph—how much of your audience's brainpower do you want them to spend trying to understand the graph versus what the data in the graph shows?
- **How are you presenting the information?** Will you be there live in person to talk through the data, set context, and answer questions, or are you sending something around that has to be processed on its own? Especially in the case where you aren't there, you need to take intentional steps to make it clear to your audience what the graph represents, how to read it, and how you want them to process your data.

Exercise 2.20: say it out loud

After you've created your graph or your slide, practice talking through it out loud. If you'll be presenting the data in a live setting (a meeting or presentation), project it on the big screen and practice discussing it as you would in a meeting. Even in an instance when you will send it off for your audience to process the data on their own, there can be important benefits to talking through your graphs.

First, set up how to read the graph, what it shows, and what each axis represents. Then talk through the data and what important observations can be made. What you say may reveal pointers on how to iterate. If you find yourself saying things like, "This isn't important" or "Ignore that," these are cues for elements you might push to the background (or in some instances, eliminate entirely). Similarly, when you hear how you direct attention when talking through the data, consider how you can achieve this visually through the way you design the graph.

In the case where you will be presenting the data live, practicing out loud will also help make the ultimate delivery smoother. First, do this on your own. Once you feel good about that, practice talking through it with someone else and get their feedback. The next exercise (Exercise 2.21) provides more pointers for getting good graph feedback.

Want to learn more benefits to saying it out loud? Listen to Episode 6 of the *Storytelling with data* podcast (storytellingwithdata.com/podcast), which focuses on this topic.

Exercise 2.21: solicit feedback

You've created a graph and you think it's pretty awesome. The challenge is that you know your work better than probably anybody else and since you're the one who created the graph, of course it's going to make sense to you. But will it work well for your audience?

Or what about the scenario where you've iterated in your tool—you've created several different views of the data but aren't entirely sure which one will work best?

In each of these cases, I recommend soliciting feedback from others.

Create your visual or set of graphs and find a helpful friend or colleague. It can be someone without any context. Have them talk you through their thought process for taking in the information, including:

- What do they pay attention to?
- What questions do they have?
- What observations do they make?

This conversation can help you understand whether the visual you've created is serving its intended purpose, or if it isn't, give you pointers on where to concentrate your iterations. Ask questions. Discuss your design choices and talk about what is working effectively and what might not be as obvious to someone who is less close to the data. Seeking feedback from various sources can also be beneficial: think about when it would be helpful to get feedback from someone in a totally different role than your own.

Also, watch initial facial responses: there is a microsecond that passes before people censor their physical reactions. If you see any furrowing of brows or pursing of lips—any general face-scrunching—these are micro-cues that something may not be working quite right. Pay attention to these cues and work to refine your visuals. If people are having a hard time with your graph, don't assume it's them. Consider what you can do to make the information easier to take in: perhaps you can more clearly title or label, use sparing color to focus attention, or choose a different graph type to get your point across more easily.

You'll find additional guidance for giving and receiving effective feedback in Exercise 9.3.

Exercise 2.22: build a data viz library

Collect and build a library of the effective data visualization examples created and used at work. You can do this on your own, or this can be an excellent undertaking for a team or organization. Be thoughtful how you organize the content for easy searchability (for example, by graph type, topic, or tool). Make files available to download so others can see the specifics of how they were made and modify for use in their own work. You can also add effective examples that you encounter externally from the media, blogs, or #SWDchallenge.

Make effective data visualization a team goal. To ensure continued focus, host a regular friendly competition, where individuals can nominate their own or their colleagues' examples of effective data visualization. Each month or quarter, choose winners and archive their work in the shared library. This can be a great ongoing source of inspiration: if someone is feeling stuck, they have something to turn to and flip through for possible ideas. It is also an excellent resource for new hires, so they have examples of effective data visualization in your work environment, helping set the right expectations for their own work.

Exercise 2.23: explore additional resources

There are many additional resources out there when it comes to choosing effective graph or getting inspiration from other people's creations. Practicing, getting feedback, and iterating are keys to success. That said, here are a few chart choosers I'm aware of that you may find helpful when it comes to figuring out what graphs might work for your specific needs:

- **Chart Chooser** (Juice Analytics, labs.juiceanalytics.com/chartchooser). Use their filters to find the right chart type for your needs, download as Excel or PowerPoint templates and insert your own data.
- **The Chartmaker Directory** (Visualizing Data, chartmaker.visualisingdata.com). Explore the matrix of chart type by tool and click the circles to see solutions and examples.
- **Graphic Continuum** (PolicyViz, policyviz.com/?s=graphic+continuum). The poster includes more than 90 graphic types grouped into six categories. Also check out the related Match It Game and Cards.
- **Interactive Chart Chooser** (Depict Data Studio, depictdatastudio.com/charts). Explore the interactive chart chooser using filters.

Check out the following collections to browse other people's work for inspiration. For each graph you encounter, pause to reflect on what works well (or not so well) and consider how you can use (or avoid!) similar aspects in your own work:

- **Information Is Beautiful Awards** (informationisbeautifulawards.com). These annual awards celebrate excellence and beauty in data visualizations, infographics, interactives, and informative art. The archives contain hundreds of data visualizations.
- **Reddit: Data Is Beautiful** (reddit.com/r/dataisbeautiful). A place for visual representations of data: graphs, charts, and maps.
- **Tableau Public Gallery** (public.tableau.com/s/gallery). Stunning data visualization examples from across the web created with Tableau Public. In particular, check out the Greatest Hits Gallery using the drop-down menu.
- **The R Graph Gallery** (r-graph-gallery.com). Looking for inspiration or help? Here you will find hundreds of distinctive graphics made with the R programming language, including code.
- **Xenographics** (xeno.graphics). Xeno.graphics is a repository of novel, innovative, and experimental visualizations to help inspire, fight xenophobia and popularize new chart types.

Exercise 2.24: let's discuss

Consider the following questions related to Chapter 2 lessons and exercises. Discuss with a partner or group.

1. How is the way that we process tables different from how we process graphs? What are the pros and cons of presenting data in tabular form? In what circumstances does it make sense to use a table? In what scenarios should you avoid a table?
2. One common decision when graphing data is whether to have a y-axis that is titled and labeled or omit the axis and label the data directly. What considerations should you make when determining which is better for a given situation?
3. When is it okay to have a non-zero baseline when graphing data?
4. Why is paper a good tool for graphing data? Were the exercises in this chapter that asked you to draw helpful? Will you use this low-tech method in your work going forward? Why or why not?
5. What is the purpose of graphing a given set of data multiple ways? Why is it important to iterate and look at different views of your data? When will you take the time to do this going forward? When does it not make sense to spend time on this?
6. The examples in SWD and in this book are mostly basic charts: a lot of lines and bars. When does it make sense to use a graph that is more novel or less familiar? What are the pros and cons of using a graph that your audience may not have previously encountered? What steps can you take in this situation to help ensure success?
7. Are there any cases where data has historically been graphed a certain way by your team or your organization that you believe should be changed? How might you drive this change? What sort of resistance or pushback do you anticipate? How can you address this?
8. What is one specific goal you will set for yourself or your team related to the strategies outlined in this chapter? How can you hold yourself (or your team) accountable to this? Who will you turn to for feedback?