

cole nussbaumer knaflic

storytelling with data

a data
visualization
guide for
business
professionals

WILEY

storytelling with data

storytelling with data

a data visualization guide
for business professionals

cole nussbaumer knaflic

WILEY

Cover image: Cole Nussbaumer Knaflic
Cover design: Wiley

Copyright © 2015 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.

Library of Congress Cataloging-in-Publication Data:

ISBN 9781119002253 (Paperback)
ISBN 9781119002260 (ePDF)
ISBN 9781119002062 (ePub)

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

To Randolph

contents

foreword	ix
acknowledgments	xi
about the author	xiii
introduction	1
chapter 1 the importance of context	19
chapter 2 choosing an effective visual	35
chapter 3 clutter is your enemy!	71
chapter 4 focus your audience's attention	99
chapter 5 think like a designer	127
chapter 6 dissecting model visuals	151
chapter 7 lessons in storytelling	165
chapter 8 pulling it all together	187
chapter 9 case studies	207
chapter 10 final thoughts	241
bibliography	257
index	261

foreword

"Power Corrupts. PowerPoint Corrupts Absolutely."

—Edward Tufte, Yale Professor Emeritus¹

We've all been victims of bad slideware. Hit-and-run presentations that leave us staggering from a maelstrom of fonts, colors, bullets, and highlights. Infographics that fail to be informative and are only graphic in the same sense that violence can be graphic. Charts and tables in the press that mislead and confuse.

It's too easy today to generate tables, charts, graphs. I can imagine some old-timer (maybe it's me?) harrumphing over my shoulder that in *his* day they'd do illustrations by hand, which meant you had to think before committing pen to paper.

Having all the information in the world at our fingertips doesn't make it easier to communicate: it makes it harder. The more information you're dealing with, the more difficult it is to filter down to the most important bits.

Enter Cole Nussbaumer Knaflic.

I met Cole in late 2007. I'd been recruited by Google the year before to create the "People Operations" team, responsible for finding, keeping, and delighting the folks at Google. Shortly after joining I decided

¹ Tufte, Edward R. 'PowerPoint Is Evil.' *Wired Magazine*, www.wired.com/wired/archive/11.09/ppt2.html, September 2003.

we needed a People Analytics team, with a mandate to make sure we innovated as much on the people side as we did on the product side. Cole became an early and critical member of that team, acting as a conduit between the Analytics team and other parts of Google.

Cole always had a knack for clarity.

She was given some of our messiest messages—such as what exactly makes one manager great and another crummy—and distilled them into crisp, pleasing imagery that told an irrefutable story. Her messages of “don’t be a data fashion victim” (i.e., lose the fancy clipart, graphics and fonts—focus on the message) and “simple beats sexy” (i.e., the point is to clearly tell a story, not to make a pretty chart) were powerful guides.

We put Cole on the road, teaching her own data visualization course over 50 times in the ensuing six years, before she decided to strike out on her own on a self-proclaimed mission to “rid the world of bad PowerPoint slides.” And if you think that’s not a big issue, a Google search of “powerpoint kills” returns almost half a million hits!

In *Storytelling with Data*, Cole has created an of-the-moment complement to the work of data visualization pioneers like Edward Tufte. She’s worked at and with some of the most data-driven organizations on the planet as well as some of the most mission-driven, data-free institutions. In both cases, she’s helped sharpen their messages, and their thinking.

She’s written a fun, accessible, and eminently practical guide to extracting the signal from the noise, and for making all of us better at getting our voices heard.

And that’s kind of the whole point, isn’t it?

Laszlo Bock

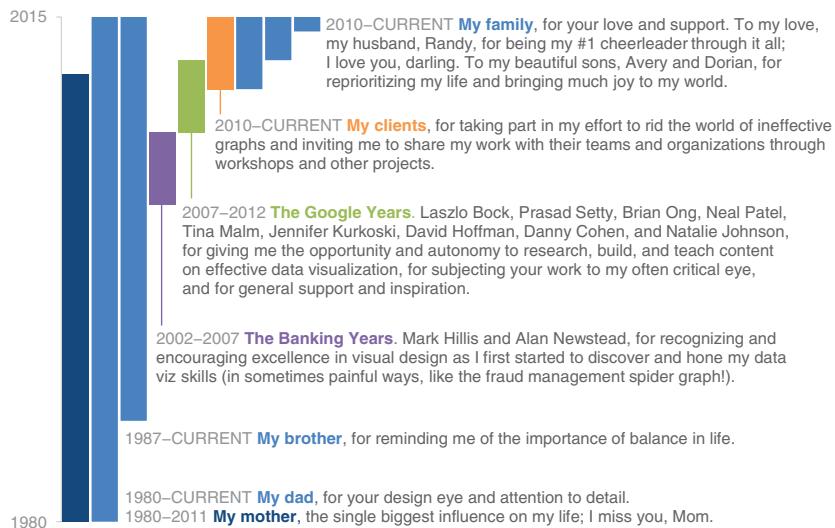
SVP of People Operations, Google, Inc.
and author of *Work Rules!*

May 2015

acknowledgments

My timeline of thanks

Thank you to...



Thank you also to everyone who helped make this book possible. I value every bit of input and help along the way. In addition to the people listed above, thanks to Bill Falloon, Meg Freeborn, Vincent Nordhaus, Robin Factor, Mark Bergeron, Mike Henton, Chris Wallace, Nick Wehrkamp, Mike Freeland, Melissa Connors, Heather Dunphy, Sharon Polese, Andrea Price, Laura Gachko, David Pugh, Marika Rohn, Robert Kosara, Andy Kriebel, John Kania, Eleanor Bell, Alberto Cairo, Nancy Duarte, Michael Eskin, Kathrin Stengel, and Zaira Basanez.

about the author

Cole Nussbaumer Knaflic tells stories with data. She specializes in the effective display of quantitative information and writes the popular blog storytellingwithdata.com. Her well-regarded workshops and presentations are highly sought after by data-minded individuals, companies, and philanthropic organizations all over the world.

Her unique talent was honed over the past decade through analytical roles in banking, private equity, and most recently 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, ensuring that Google attracted, developed, and retained great talent and that the organization was best aligned to meet business needs. Cole traveled to Google offices throughout the United States and Europe to teach the course she developed on data visualization. She has also acted as an adjunct faculty member at the Maryland Institute College of Art (MICA), where she taught Introduction to Information Visualization.

Cole has a BS in Applied Math and an MBA, both from the University of Washington. When she isn't ridding the world of ineffective graphs one pie at a time, she is baking them, traveling, and embarking on adventures with her husband and two young sons in San Francisco.

introduction

Bad graphs are everywhere

I encounter a lot of less-than-stellar visuals in my work (and in my life—once you get a discerning eye for this stuff, it's hard to turn it off). Nobody sets out to make a bad graph. But it happens. Again and again. At every company throughout all industries and by all types of people. It happens in the media. It happens in places where you would expect people to know better. Why is that?

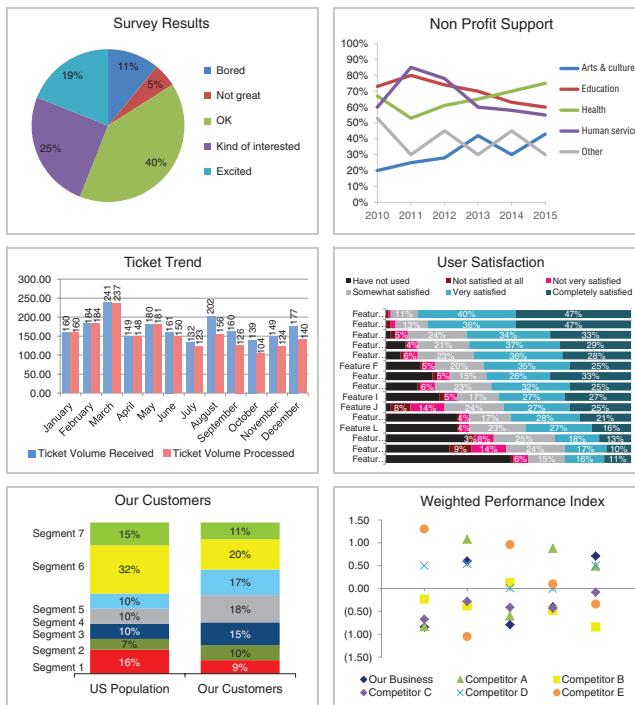


FIGURE 0.1 A sampling of ineffective graphs

We aren't naturally good at storytelling with data

In school, we learn a lot about language and math. On the language side, we learn how to put words together into sentences and into stories. With math, we learn to make sense of numbers. But it's rare that these two sides are paired: no one teaches us how to tell stories with numbers. Adding to the challenge, very few people feel naturally adept in this space.

This leaves us poorly prepared for an important task that is increasingly in demand. Technology has enabled us to amass greater and greater amounts of data and there is an accompanying growing desire to make sense out of all of this data. Being able to visualize data and tell stories with it is key to turning it into *information* that can be used to drive better decision making.

In the absence of natural skills or training in this space, we often end up relying on our tools to understand best practices. Advances in technology, in addition to increasing the amount of and access to data, have also made tools to work with data pervasive. Pretty much anyone can put some data into a graphing application (for example, Excel) and create a graph. This is important to consider, so I will repeat myself: anyone can put some data into a graphing application and create a graph. This is remarkable, considering that the process of creating a graph was historically reserved for scientists or those in other highly technical roles. And scary, because without a clear path to follow, our best intentions and efforts (combined with oft-questionable tool defaults) can lead us in some really bad directions: 3D, meaningless color, pie charts.

Skilled in Microsoft Office? So is everyone else!

Being adept with word processing applications, spreadsheets, and presentation software—things that used to set one apart on a resume and in the workplace—has become a minimum expectation for most employers. A recruiter told me that, today, having “proficiency in Microsoft Office” on a resume isn’t enough: a basic level of knowledge here is assumed and it’s what you can do above and beyond that will set you apart from others. Being able to effectively tell stories with data is one area that will give you that edge and position you for success in nearly any role.

While technology has increased access to and proficiency in tools to work with data, there remain gaps in capabilities. You can put some data in Excel and create a graph. For many, the process of data visualization ends there. This can render the most interesting story completely underwhelming, or worse—difficult or impossible to understand. Tool defaults and general practices tend to leave our data and the stories we want to tell with that data sorely lacking.

There is a story in your data. But your tools don’t know what that story is. That’s where it takes you—the analyst or communicator of the information—to bring that story visually and contextually to life. That process is the focus of this book. The following are a few example before-and-afters to give you a visual sense of what you’ll learn; we’ll cover each of these in detail at various points in the book.

The lessons we will cover will enable you to shift from simply showing data to **storytelling with data**.

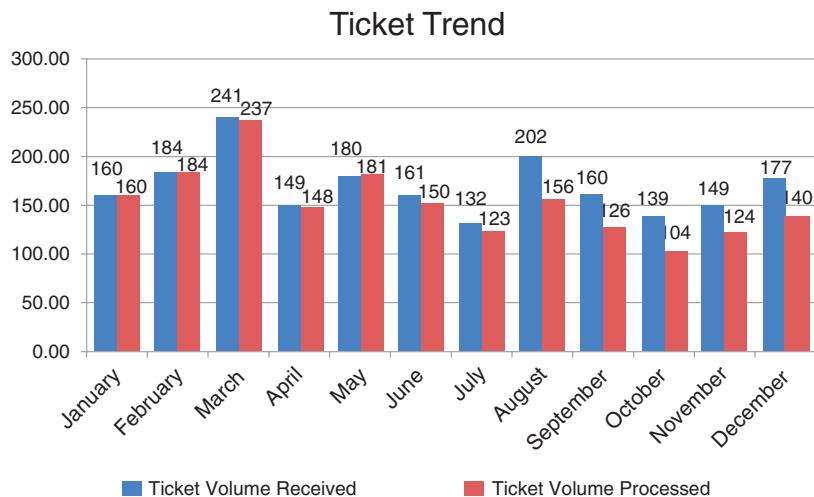
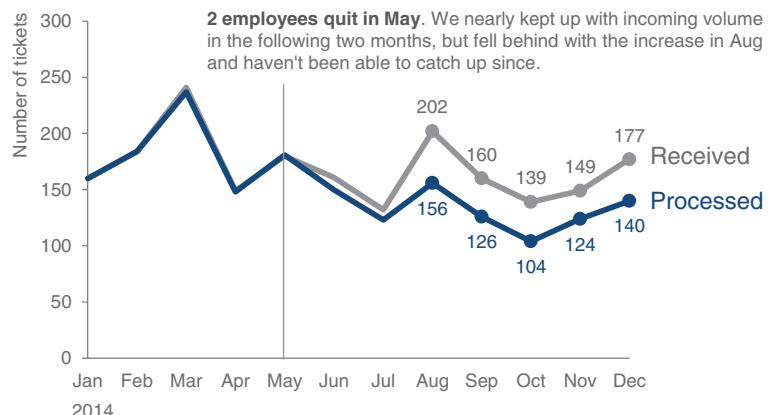


FIGURE 0.2 Example 1 (before): showing data

Please approve the hire of 2 FTEs

to backfill those who quit in the past year

Ticket volume over time



Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

FIGURE 0.3 Example 1 (after): storytelling with data

Survey Results

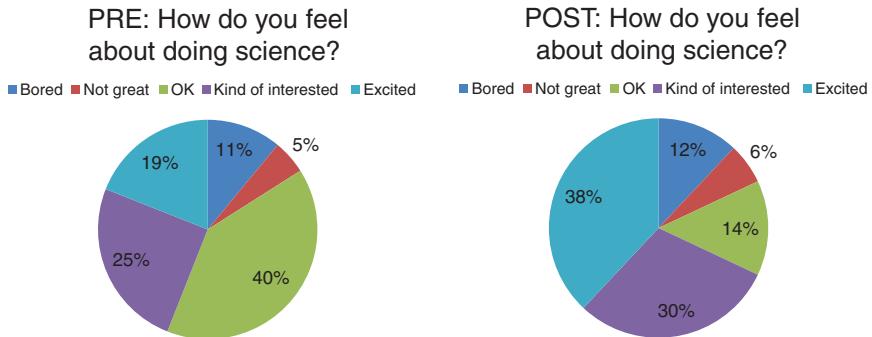
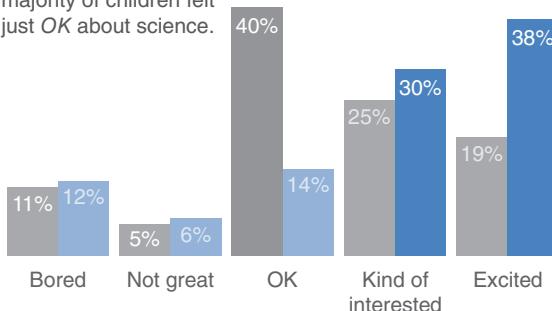


FIGURE 0.4 Example 2 (before): showing data

Pilot program was a success

How do you feel about science?

BEFORE program, the majority of children felt just *OK* about science.



AFTER program, more children were *Kind of interested* & *Excited* about science.

Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

FIGURE 0.5 Example 2 (after): storytelling with data

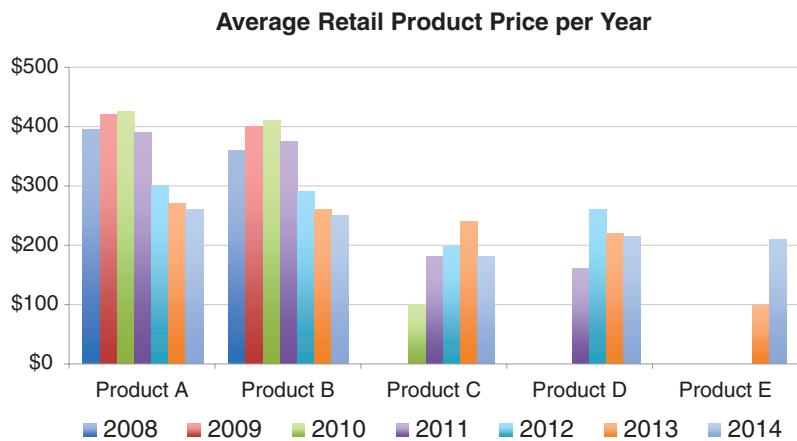


FIGURE 0.6 Example 3 (before): showing data

To be competitive, we recommend introducing our product *below the \$223 average price point in the \$150–\$200 range*

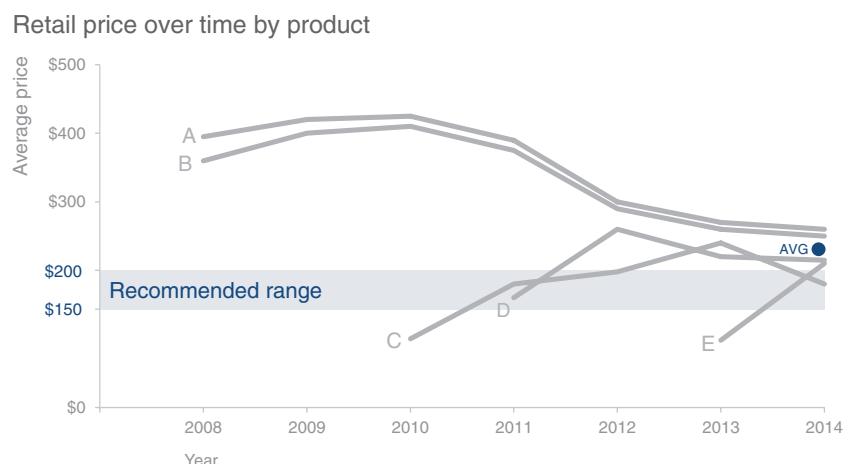


FIGURE 0.7 Example 3 (after): storytelling with data

Who this book is written for

This book is written for anyone who needs to communicate something to someone using data. This includes (but is certainly not limited to): analysts sharing the results of their work, students visualizing thesis data, managers needing to communicate in a data-driven way, philanthropists proving their impact, and leaders informing their board. I believe that anyone can improve their ability to communicate effectively with data. This is an intimidating space for many, but it does not need to be.

When you are asked to “show data,” what sort of feelings does that evoke?

Perhaps you feel uncomfortable because you are unsure where to start. Or maybe it feels like an overwhelming task because you assume that what you are creating needs to be complicated and show enough detail to answer every possible question. Or perhaps you already have a solid foundation here, but are looking for that something that will help take your graphs and the stories you want to tell with them to the next level. In all of these cases, this book is written with you in mind.

“When I’m asked to *show the data*, I feel...”

An informal Twitter poll I conducted revealed the following mix of emotions when people are asked to “show the data.”

Frustrated because I don’t think I’ll be able to tell the whole story.

Pressure to make it clear to whomever needs the data.

Inadequate. Boss: Can you drill down into that? Give me the split by x, y, and z.

Being able to tell stories with data is a skill that's becoming ever more important in our world of increasing data and desire for data-driven decision making. An effective data visualization can mean the difference between success and failure when it comes to communicating the findings of your study, raising money for your non-profit, presenting to your board, or simply getting your point across to your audience.

My experience has taught me that most people face a similar challenge: they may recognize the need to be able to communicate effectively with data but feel like they lack expertise in this space. People skilled in data visualization are hard to come by. Part of the challenge is that data visualization is a single step in the analytical process. Those hired into analytical roles typically have quantitative backgrounds that suit them well for the other steps (finding the data, pulling it together, analyzing it, building models), but not necessarily any formal training in design to help them when it comes to the communication of the analysis—which, by the way, is typically the only part of the analytical process that your audience ever sees. And increasingly, in our ever more data-driven world, those without technical backgrounds are being asked to put on analytical hats and communicate using data.

The feelings of discomfort you may experience in this space aren't surprising, given that being able to communicate effectively with data isn't something that has been traditionally taught. Those who excel have typically learned what works and what doesn't through trial and error. This can be a long and tedious process. Through this book, I hope to help expedite it for you.

How I learned to tell stories with data

I have always been drawn to the space where mathematics and business intersect. My educational background is mathematics and business, which enables me to communicate effectively with both sides—given that they don't always speak the same language—and help them better understand one another. I love being able to take

the science of data and use it to inform better business decisions. Over time, I've found that one key to success is being able to communicate effectively visually with data.

I initially recognized the importance of being skilled in this area during my first job out of college. I was working as an analyst in credit risk management (before the subprime crisis and hence before anyone really knew what credit risk management was). My job was to build and assess statistical models to forecast delinquency and loss. This meant taking complicated stuff and ultimately turning it into a simple communication of whether we had adequate money in the reserves for expected losses, in what scenarios we'd be at risk, and so forth. I quickly learned that spending time on the aesthetic piece—something my colleagues didn't typically do—meant my work garnered more attention from my boss and my boss's boss. For me, that was the beginning of seeing value in spending time on the visual communication of data.

After progressing through various roles in credit risk, fraud, and operations management, followed by some time in the private equity world, I decided I wanted to continue my career outside of banking and finance. I paused to reflect on the skills I possessed that I wanted to be utilizing on a daily basis: at the core, it was using data to influence business decisions.

I landed at Google, on the People Analytics team. Google is a data-driven company—so much so that they even use data and analytics in a space not frequently seen: human resources. People Analytics is an analytics team embedded in Google's HR organization (referred to at Google as "People Operations"). The mantra of this team is to help ensure that people decisions at Google—decisions about employees or future employees—are data driven. This was an amazing place to continue to hone my storytelling with data skills, using data and analytics to better understand and inform decision making in spaces like targeted hiring, engaging and motivating employees, building effective teams, and retaining talent. Google People Analytics is cutting edge, helping to forge a path that many other

companies have started to follow. Being involved in building and growing this team was an incredible experience.

Storytelling with data on what makes a great manager via Project Oxygen

One particular project that has been highlighted in the public sphere is the Project Oxygen research at Google on what makes a great manager. This work has been described in the *New York Times* and is the basis of a popular *Harvard Business Review* case study. One challenge faced was communicating the findings to various audiences, from engineers who were sometimes skeptical on methodology and wanted to dig into the details, to managers wanting to understand the big-picture findings and how to put them to use. My involvement in the project was on the communication piece, helping to determine how to best show sometimes very complicated stuff in a way that would appease the engineers and their desire for detail while still being understandable and straightforward for managers and various levels of leadership. To do this, I leveraged many of the concepts we will discuss in this book.

The big turning point for me happened when we were building an internal training program within People Operations at Google and I was asked to develop content on data visualization. This gave me the opportunity to research and start to learn the principles behind effective data visualization, helping me understand why some of the things I'd arrived at through trial and error over the years had been effective. With this research, I developed a course on data visualization that was eventually rolled out to all of Google.

The course created some buzz, both inside and outside of Google. Through a series of fortuitous events, I received invitations to speak at a couple of philanthropic organizations and events on the topic of data visualization. Word spread. More and more people were reaching out to me—initially in the philanthropic world, but increasingly in

the corporate sector as well—looking for guidance on how to communicate effectively with data. It was becoming increasingly clear that the need in this space was not unique to Google. Rather, pretty much anyone in an organization or business setting could increase their impact by being able to communicate effectively with data. After acting as a speaker at conferences and organizations in my spare time, eventually I left Google to pursue my emerging goal of teaching the world how to tell stories with data.

Over the past few years, I've taught workshops for more than a hundred organizations in the United States and Europe. It's been interesting to see that the need for skills in this space spans many industries and roles. I've had audiences in consulting, consumer products, education, financial services, government, health care, nonprofit, retail, startups, and technology. My audiences have been a mix of roles and levels: from analysts who work with data on a daily basis to those in non-analytical roles who occasionally have to incorporate data into their work, to managers needing to provide guidance and feedback, to the executive team delivering quarterly results to the board.

Through this work, I've been exposed to many diverse data visualization challenges. I have come to realize that the skills that are needed in this area are fundamental. They are not specific to any industry or role, and they can be effectively taught and learned—as demonstrated by the consistent positive feedback and follow-ups I receive from workshop attendees. Over time, I've codified the lessons that I teach in my workshops. These are the lessons I will share with you.

How you'll learn to tell stories with data: 6 lessons

In my workshops, I typically focus on five key lessons. The big opportunity with this book is that there isn't a time limit (in the way there is in a workshop setting). I've included a sixth bonus lesson that I've always wanted to share ("think like a designer") and also a lot more by way of before-and-after examples, step-by-step instruction, and insight into my thought process when it comes to the visual design of information.

I will give you practical guidance that you can begin using immediately to better communicate visually with data. We'll cover content to help you learn and be comfortable employing six key lessons:

1. Understand the context
2. Choose an appropriate visual display
3. Eliminate clutter
4. Focus attention where you want it
5. Think like a designer
6. Tell a story

Illustrative examples span many industries

Throughout the book, I use a number of case studies to illustrate the concepts discussed. The lessons we cover will not be industry—or role—specific, but rather will focus on fundamental concepts and best practices for effective communication with data. Because my work spans many industries, so do the examples upon which I draw. You will see case studies from technology, education, consumer products, the nonprofit sector, and more.

Each example used is based on a lesson I have taught in my workshops, but in many cases I've slightly changed the data or generalized the situation to protect confidential information.

For any example that doesn't initially seem relevant to you, I encourage you to pause and think about what data visualization or communication challenges you encounter where a similar approach could be effective. There is something to be learned from every example, even if the example itself isn't obviously related to the world in which you work.

Lessons are not tool specific

The lessons we will cover in this book focus on best practices that can be applied in any graphing application or presentation software. There are a vast number of tools that can be leveraged to tell effective stories with data. No matter how great the tool, however, it will never know your data and your story like you do. Take the time to learn your tool well so that it does not become a limiting factor when it comes to applying the lessons we'll cover throughout this book.

How do you do that in Excel?

While I will not focus the discussion on specific tools, the examples in this book were created using Microsoft Excel. For those interested in a closer look at how similar visuals can be built in Excel, please visit my blog at storytellingwithdata.com, where you can download the Excel files that accompany my posts.

How this book is organized

This book is organized into a series of big-picture lessons, with each chapter focusing on a single core lesson and related concepts. We will discuss a bit of theory when it will aid in understanding, but I will emphasize the practical application of the theory, often through specific, real-world examples. You will leave each chapter ready to apply the given lesson.

The lessons in the book are organized chronologically in the same way that I think about the storytelling with data process. Because of this and because later chapters do build on and in some cases refer back to earlier content, I recommend reading from beginning to end. After you've done this, you'll likely find yourself referring back to specific points of interest or examples that are relevant to the current data visualization challenges you face.

To give you a more specific idea of the path we'll take, chapter summaries can be found below.

Chapter 1: the importance of context

Before you start down the path of data visualization, there are a couple of questions that you should be able to concisely answer: Who is your audience? What do you need them to know or do? This chapter describes the importance of understanding the situational context, including the audience, communication mechanism, and desired tone. A number of concepts are introduced and illustrated via example to help ensure that context is fully understood. Creating a robust understanding of the situational context reduces iterations down the road and sets you on the path to success when it comes to creating visual content.

Chapter 2: choosing an effective visual

What is the best way to show the data you want to communicate? I've analyzed the visual displays I use most in my work. In this chapter, I introduce the most common types of visuals used to communicate data in a business setting, discuss appropriate use cases for each, and illustrate each through real-world examples. Specific types of visuals covered include simple text, table, heatmap, line graph, slopegraph, vertical bar chart, vertical stacked bar chart, waterfall chart, horizontal bar chart, horizontal stacked bar chart, and square area graph. We also cover visuals to be avoided, including pie and donut charts, and discuss reasons for avoiding 3D.

Chapter 3: clutter is your enemy!

Picture a blank page or a blank screen: every single element you add to that page or screen takes up cognitive load on the part of your audience. That means we should take a discerning eye to the elements we allow on our page or screen and work to identify those things that are taking up brain power unnecessarily and remove

them. Identifying and eliminating clutter is the focus of this chapter. As part of this conversation, I introduce and discuss the Gestalt Principles of Visual Perception and how we can apply them to visual displays of information such as tables and graphs. We also discuss alignment, strategic use of white space, and contrast as important components of thoughtful design. Several examples are used to illustrate the lessons.

Chapter 4: focus your audience's attention

In this chapter, we continue to examine how people see and how you can use that to your advantage when crafting visuals. This includes a brief discussion on sight and memory that will act to frame up the importance of preattentive attributes like size, color, and position on page. We explore how preattentive attributes can be used strategically to help direct your audience's attention to where you want them to focus and to create a visual hierarchy of components to help direct your audience through the information you want to communicate in the way you want them to process it. Color as a strategic tool is covered in depth. Concepts are illustrated through a number of examples.

Chapter 5: think like a designer

Form follows function. This adage of product design has clear application to communicating with data. When it comes to the form and function of our data visualizations, we first want to think about what it is we want our audience to be able to do with the data (function) and create a visualization (form) that will allow for this with ease. In this chapter, we discuss how traditional design concepts can be applied to communicating with data. We explore affordances, accessibility, and aesthetics, drawing upon a number of concepts introduced previously, but looking at them through a slightly different lens. We also discuss strategies for gaining audience acceptance of your visual designs.

Chapter 6: dissecting model visuals

Much can be learned from a thorough examination of effective visual displays. In this chapter, we look at five exemplary visuals and discuss the specific thought process and design choices that led to their creation, utilizing the lessons covered up to this point. We explore decisions regarding the type of graph and ordering of data within the visual. We consider choices around what and how to emphasize and de-emphasize through use of color, thickness of lines, and relative size. We discuss alignment and positioning of components within the visuals and also the effective use of words to title, label, and annotate.

Chapter 7: lessons in storytelling

Stories resonate and stick with us in ways that data alone cannot. In this chapter, I introduce concepts of storytelling that can be leveraged for communicating with data. We consider what can be learned from master storytellers. A story has a clear beginning, middle, and end; we discuss how this framework applies to and can be used when constructing business presentations. We cover strategies for effective storytelling, including the power of repetition, narrative flow, considerations with spoken and written narratives, and various tactics to ensure that our story comes across clearly in our communications.

Chapter 8: pulling it all together

Previous chapters included piecemeal applications to demonstrate individual lessons covered. In this comprehensive chapter, we follow the storytelling with data process from start to finish using a single real-world example. We understand the context, choose an appropriate visual display, identify and eliminate clutter, draw attention to where we want our audience to focus, think like a designer, and tell a story. Together, these lessons and resulting visuals and narrative illustrate how we can move from simply showing data to telling a story with data.

Chapter 9: case studies

The penultimate chapter explores specific strategies for tackling common challenges faced in communicating with data through a number of case studies. Topics covered include color considerations with a dark background, leveraging animation in the visuals you present versus those you circulate, establishing logic in order, strategies for avoiding the spaghetti graph, and alternatives to pie charts.

Chapter 10: final thoughts

Data visualization—and communicating with data in general—sits at the intersection of science and art. There is certainly some science to it: best practices and guidelines to follow. There is also an artistic component. Apply the lessons we've covered to forge your path, using your artistic license to make the information easier for your audience to understand. In this final chapter, we discuss tips on where to go from here and strategies for upskilling storytelling with data competency in your team and your organization. We end with a recap of the main lessons covered.

Collectively, the lessons we'll cover will enable you to tell stories with data. Let's get started!

chapter one

the importance of context

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

Exploratory vs. explanatory analysis

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

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

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

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

Recommended reading

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

Who, what, and how

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

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

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

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

Who

Your audience

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

You

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

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

Recommended reading

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

What

Action

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

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

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

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

Prompting action

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

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

Mechanism

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

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

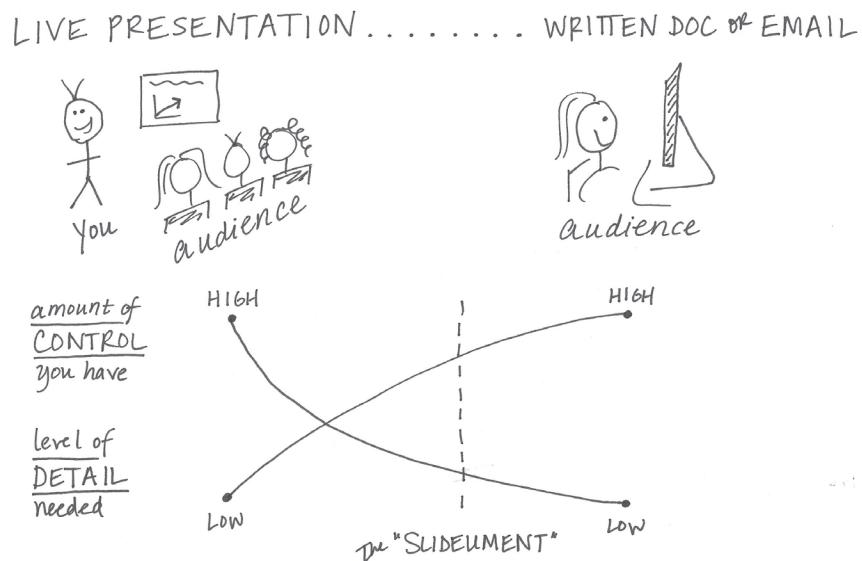


FIGURE 1.1 Communication mechanism continuum

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

For live presentations, practice makes perfect

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

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

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

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

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

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

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

Tone

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

How

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

Ignore the nonsupporting data?

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

Who, what, and how: illustrated by example

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

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

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

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

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

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

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

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

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

Consulting for context: questions to ask

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

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

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

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

The 3-minute story & Big Idea

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

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

3-minute story

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

Big Idea

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

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

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

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

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

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

Storyboarding

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

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

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

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

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

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

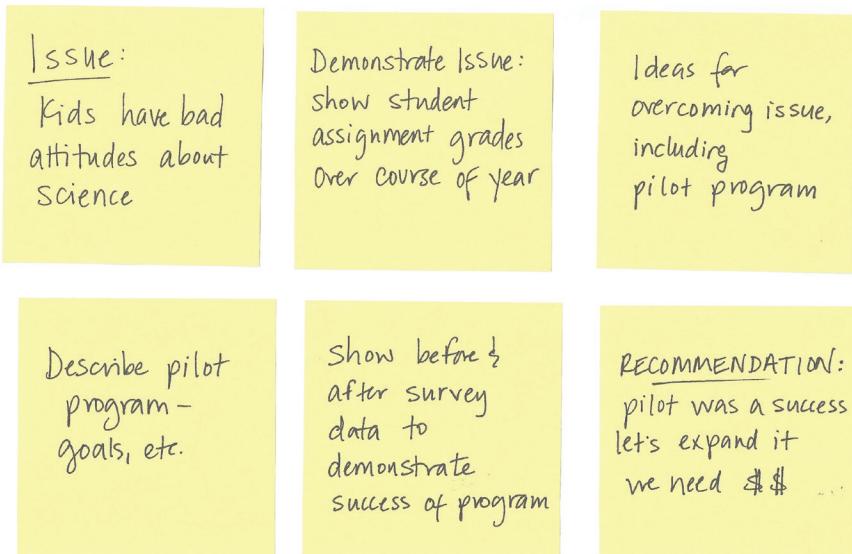


FIGURE 1.2 Example storyboard

In closing

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

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

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

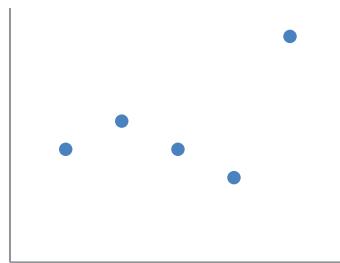
chapter two

choosing an effective visual

There are many different graphs and other types of visual displays of information, but a handful will work for the majority of your needs. When I look back over the 150+ visuals that I created for workshops and consulting projects in the past year, there were only a dozen different types of visuals that I used (Figure 2.1). These are the visuals we'll focus on in this chapter.

91%

Simple text



Scatterplot

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

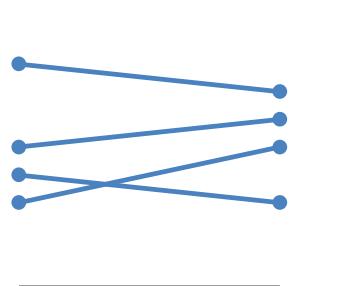
Table



Line

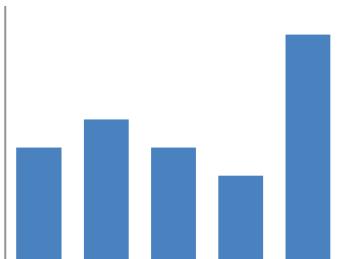
	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

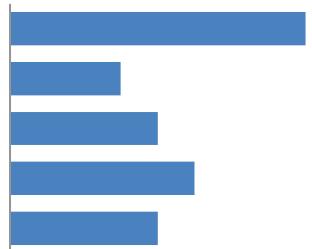


Slopegraph

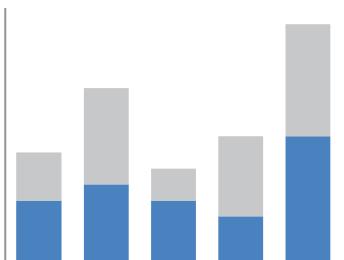
FIGURE 2.1 The visuals I use most



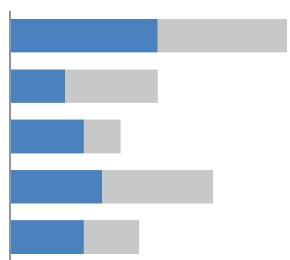
Vertical bar



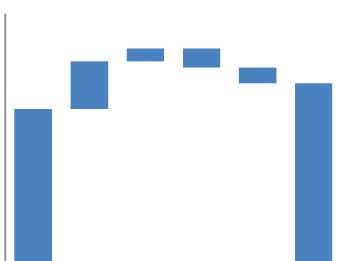
Horizontal bar



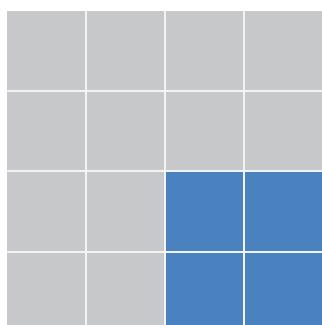
Stacked vertical bar



Stacked horizontal bar



Waterfall



Square area

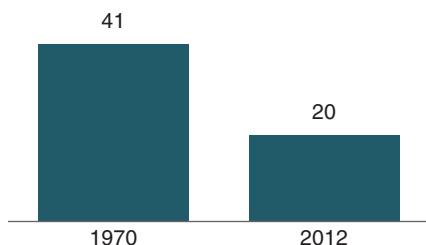
Simple text

When you have just a number or two to share, simple text can be a great way to communicate. Think about solely using the number—making it as prominent as possible—and a few supporting words to clearly make your point. Beyond potentially being misleading, putting one or only a couple of numbers in a table or graph simply causes the numbers to lose some of their oomph. When you have a number or two that you want to communicate, think about using the numbers themselves.

To illustrate this concept, let's consider the following example. A graph similar to Figure 2.2 accompanied an April 2014 Pew Research Center report on stay-at-home moms.

Children with a "Traditional" Stay-at- Home Mother

*% of children with a married
stay-at-home mother with a
working husband*



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

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

Adapted from PEW RESEARCH CENTER

FIGURE 2.2 Stay-at-home moms original graph

The fact that you have some numbers does not mean that you need a graph! In Figure 2.2, quite a lot of text and space are used for a grand total of two numbers. The graph doesn't do much to aid in the interpretation of the numbers (and with the positioning of the data labels outside of the bars, it can even skew your perception of relative height such that 20 is less than half of 41 doesn't really come across visually).

In this case, a simple sentence would suffice: *20% of children had a traditional stay-at-home mom in 2012, compared to 41% in 1970.*

Alternatively, in a presentation or report, your visual could look something like Figure 2.3.

20%

of children had a
traditional stay-at-home mom
in 2012, compared to 41% in 1970

FIGURE 2.3 Stay-at-home moms simple text makeover

As a side note, one consideration in this specific example might be whether you want to show an entirely different metric. For example, you could reframe in terms of the percent change: "The number of children having a traditional stay-at-home mom decreased more than 50% between 1970 and 2012." I advise caution, however, any time you reduce from multiple numbers down to a single one—think about what context may be lost in doing so. In this case, I find that the actual magnitude of the numbers (20% and 41%) is helpful in interpreting and understanding the change.

When you have just a number or two that you want to communicate: *use the numbers directly.*

When you have more data that you want to show, generally a table or graph is the way to go. One thing to understand is that people interact differently with these two types of visuals. Let's discuss each in detail and look at some specific varieties and use cases.

Tables

Tables interact with our verbal system, which means that we read them. When I have a table in front of me, I typically have my index finger out: I'm reading across rows and down columns or I'm comparing values. Tables are great for just that—communicating to a mixed audience whose members will each look for their particular row of interest. If you need to communicate multiple different units of measure, this is also typically easier with a table than a graph.

Tables in live presentations

Using a table in a live presentation is rarely a good idea. As your audience reads it, you lose their ears and attention to make your point verbally. When you find yourself using a table in a presentation or report, ask yourself: what is the point you are trying to make? Odds are that there will be a better way to pull out and visualize the piece or pieces of interest. In the event that you feel you're losing too much by doing this, consider whether including the full table in the appendix and a link or reference to it will meet your audience's needs.

One thing to keep in mind with a table is that you want the design to fade into the background, letting the data take center stage. Don't let heavy borders or shading compete for attention. Instead, think

of using light borders or simply white space to set apart elements of the table.

Take a look at the example tables in Figure 2.4. As you do, note how the data stands out more than the structural components of the table in the second and third iterations (light borders, minimal borders).

Heavy borders				Light borders				Minimal borders			
Group	Metric A	Metric B	Metric C	Group	Metric A	Metric B	Metric C	Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ	Group 1	\$X.X	Y%	Z,ZZZ	Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ	Group 2	\$X.X	Y%	Z,ZZZ	Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ	Group 3	\$X.X	Y%	Z,ZZZ	Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ	Group 4	\$X.X	Y%	Z,ZZZ	Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ	Group 5	\$X.X	Y%	Z,ZZZ	Group 5	\$X.X	Y%	Z,ZZZ

FIGURE 2.4 Table borders

Borders should be used to improve the legibility of your table. Think about pushing them to the background by making them grey, or getting rid of them altogether. The data should be what stands out, not the borders.

Recommended reading

For more on table design, check out Stephen Few's book, *Show Me the Numbers*. There is an entire chapter dedicated to the design of tables, with discussion on the structural components of tables and best practices in table design.

Next, let's shift our focus to a special case of tables: the heatmap.

Heatmap

One approach for mixing the detail you can include in a table while also making use of visual cues is via a heatmap. A heatmap is a way to visualize data in tabular format, where in place of (or in addition to) the numbers, you leverage colored cells that convey the relative magnitude of the numbers.

Consider Figure 2.5, which shows some generic data in a table and also a heatmap.

Table

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

Heatmap

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

FIGURE 2.5 Two views of the same data

In the table in Figure 2.5, you are left to read the data. I find myself scanning across rows and down columns to get a sense of what I'm looking at, where numbers are higher or lower, and mentally stack rank the categories presented in the table.

To reduce this mental processing, we can use **color saturation** to provide visual cues, helping our eyes and brains more quickly target the potential points of interest. In the second iteration of the table on the right entitled "Heatmap," the higher saturation of blue, the higher the number. This makes the process of picking out the tails of the spectrum—the lowest number (11%) and highest number (58)—an easier and faster process than it was in the original table where we didn't have any visual cues to help direct our attention.

Graphing applications (like Excel) typically have conditional formatting functionality built in that allows you to apply formatting like

that shown in Figure 2.5 with ease. Be sure when you leverage this to always include a legend to help the reader interpret the data (in this case, the LOW-HIGH subtitle on the heatmap with color corresponding to the conditional formatting color serves this purpose).

Next, let's shift our discussion to the visuals we tend to think of first when it comes to communicating with data: graphs.

Graphs

While tables interact with our verbal system, graphs interact with our visual system, which is faster at processing information. This means that a well-designed graph will typically get the information across more quickly than a well-designed table. As I mentioned at the onset of this chapter, there are a plethora of graph types out there. The good news is that a handful of them will meet most of your everyday needs.

The types of graphs I frequently use fall into four categories: points, lines, bars, and area. We will examine these more closely and discuss the subtypes that I find myself using on a regular basis, with specific use cases and examples for each.

Chart or graph?

Some draw a distinction between charts and graphs. Typically, "chart" is the broader category, with "graphs" being one of the subtypes (other chart types include maps and diagrams). I don't tend to draw this distinction, since nearly all of the charts I deal with on a regular basis are graphs. Throughout this book, I use the words *chart* and *graph* interchangeably.

Points

Scatterplot

Scatterplots can be useful for showing the relationship between two things, because they allow you to encode data simultaneously on a horizontal x-axis and vertical y-axis to see whether and what relationship exists. They tend to be more frequently used in scientific fields (and perhaps, because of this, are sometimes viewed as complicated to understand by those less familiar with them). Though infrequent, there are use cases for scatterplots in the business world as well.

For example, let's say that we manage a bus fleet and want to understand the relationship between miles driven and cost per mile. The scatterplot may look something like Figure 2.6.

Cost per mile by miles driven

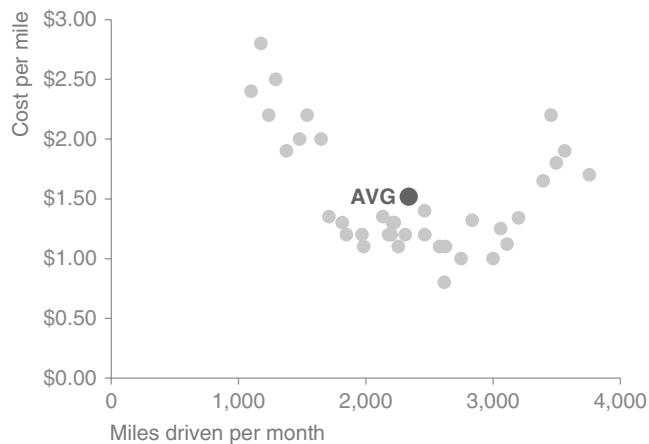


FIGURE 2.6 Scatterplot

If we want to focus primarily on those cases where cost per mile is above average, a slightly modified scatterplot designed to draw our eye there more quickly might look something like what is shown in Figure 2.7.

Cost per mile by miles driven

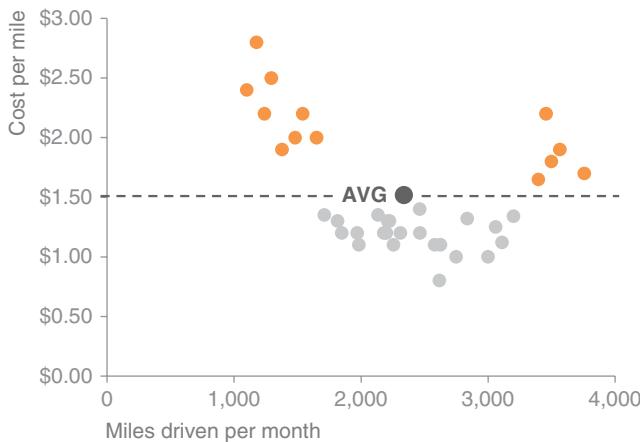


FIGURE 2.7 Modified scatterplot

We can use Figure 2.7 to make observations such as cost per mile is higher than average when less than about 1,700 miles or more than about 3,300 miles were driven for the sample observed. We'll talk more about the design choices made here and reasons for them in upcoming chapters.

Lines

Line graphs are most commonly used to plot continuous data. Because the points are physically connected via the line, it implies a connection between the points that may not make sense for categorical data (a set of data that is sorted or divided into different categories). Often, our continuous data is in some unit of time: days, months, quarters, or years.

Within the line graph category, there are two types of charts that I frequently find myself using: the standard line graph and the slopegraph.

Line graph

The line graph can show a single series of data, two series of data, or multiple series, as illustrated in Figure 2.8.

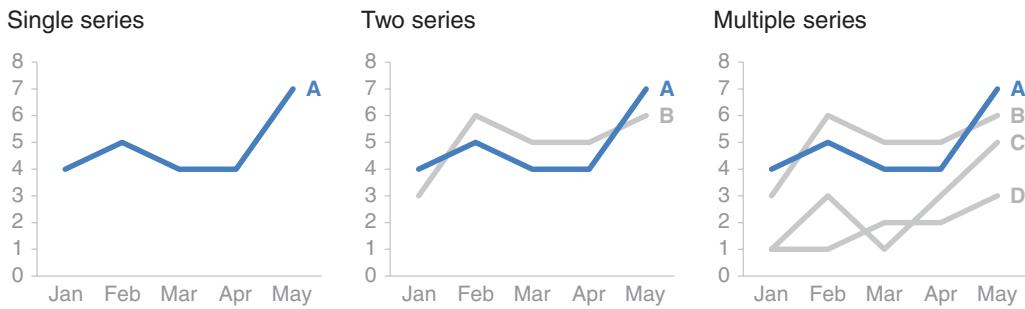


FIGURE 2.8 Line graphs

Note that when you're graphing time on the horizontal x-axis of a line graph, the data plotted must be in consistent intervals. I recently saw a graph where the units on the x-axis were decades from 1900 forward (1910, 1920, 1930, etc.) and then switched to yearly after 2010 (2011, 2012, 2013, 2014). This meant that the distance between the decade points and annual points looked the same. This is a misleading way to show the data. Be consistent in the time points you plot.

Showing average within a range in a line graph

In some cases, the line in your line graph may represent a summary statistic, like the average, or the point estimate of a forecast. If you also want to give a sense of the range (or confidence level, depending on the situation), you can do that directly on the graph by also visualizing this range. For example, the graph in Figure 2.9 shows the minimum, average, and maximum wait times at passport control for an airport over a 13-month period.

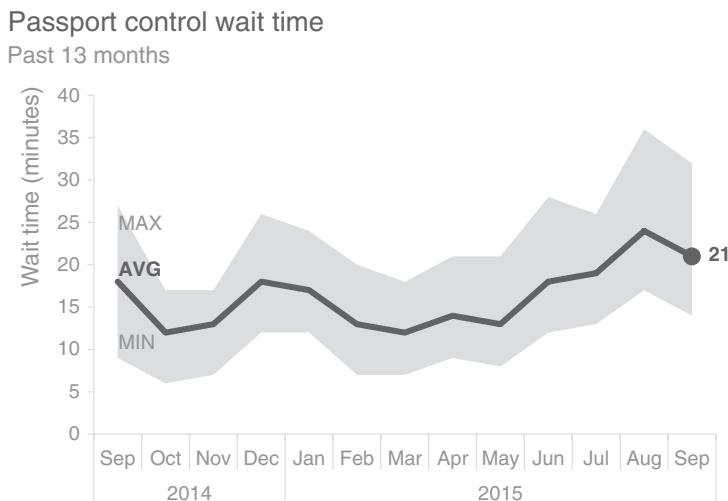


FIGURE 2.9 Showing average within a range in a line graph

Slopegraph

Slopegraphs can be useful when you have two time periods or points of comparison and want to quickly show relative increases and decreases or differences across various categories between the two data points.

The best way to explain the value of and use case for slopegraphs is through a specific example. Imagine that you are analyzing and communicating data from a recent employee feedback survey. To show the relative change in survey categories from 2014 to 2015, the slopegraph might look something like Figure 2.10.

Slopegraphs pack in a lot of information. In addition to the absolute values (the points), the lines that connect them give you the visual increase or decrease in rate of change (via the slope or direction) without ever having to explain that's what they are doing, or what exactly a "rate of change" is—rather, it's intuitive.



FIGURE 2.10 Slopegraph

Slopegraph template

Slopegraphs can take a bit of patience to set up because they often aren't one of the standard graphs included in graphing applications. An Excel template with an example slopegraph and instructions for customized use can be downloaded here: storytellingwithdata.com/slopegraph-template.

Whether a slopegraph will work in your specific situation depends on the data itself. If many of the lines are overlapping, a slopegraph may not work, though in some cases you can still emphasize a single series at a time with success. For example, we can draw attention

to the single category that decreased over time from the preceding example.

Employee feedback over time

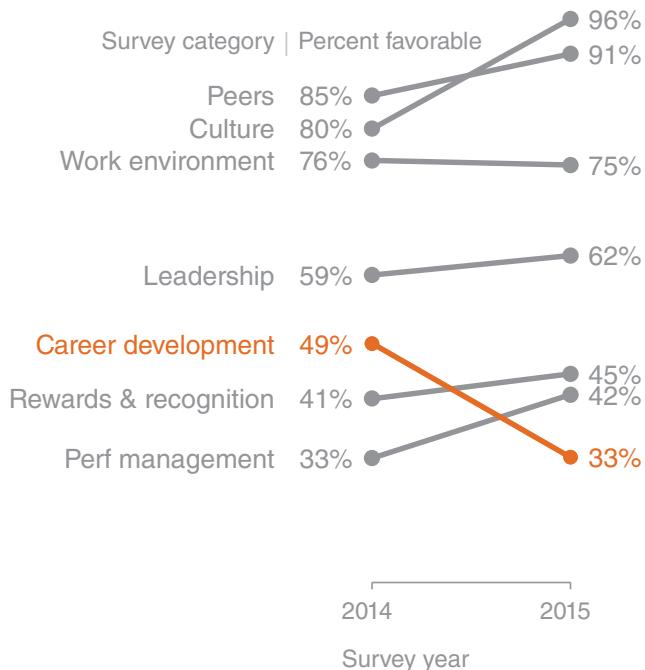


FIGURE 2.11 Modified slopegraph

In Figure 2.11, our attention is drawn immediately to the decrease in “Career development,” while the rest of the data is preserved for context without competing for attention. We will talk about the strategy behind this when we discuss preattentive attributes in Chapter 4.

While lines work well to show data over time, bars tend to be my go-to graph type for plotting categorical data, where information is organized into groups.

Bars

Sometimes bar charts are avoided because they are common. This is a mistake. Rather, bar charts should be leveraged because they are common, as this means less of a learning curve for your audience. Instead of using their brain power to try to understand how to read the graph, your audience spends it figuring out what information to take away from the visual.

Bar charts are easy for our eyes to read. Our eyes compare the end points of the bars, so it is easy to see quickly which category is the biggest, which is the smallest, and also the incremental difference between categories. Note that, because of how our eyes compare the relative end points of the bars, it is important that bar charts always have a zero baseline (where the x-axis crosses the y-axis at zero), otherwise you get a false visual comparison.

Consider Figure 2.12 from Fox News.

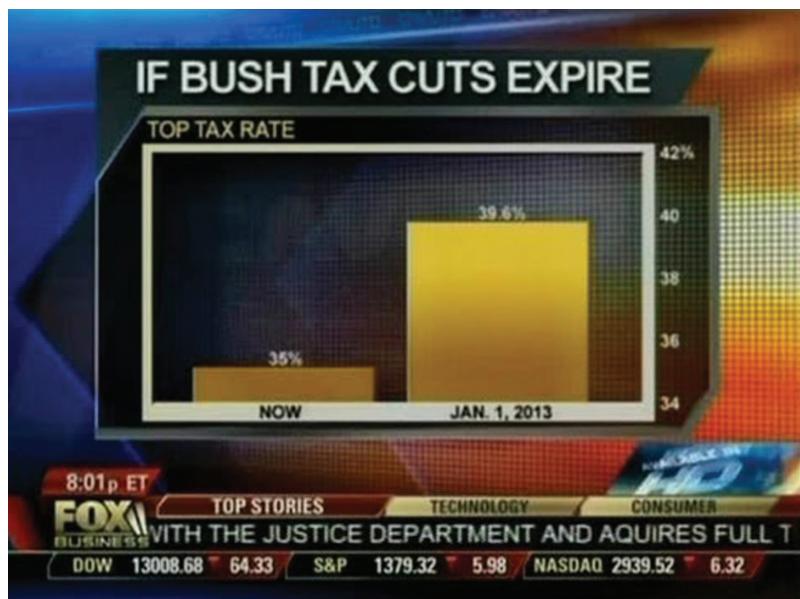


FIGURE 2.12 Fox News bar chart

For this example, let's imagine we are back in the fall of 2012. We are wondering what will happen if the Bush tax cuts expire. On the left-hand side, we have what the top tax rate is currently, 35%, and on the right-hand side what it will be as of January 1, at 39.6%.

When you look at this graph, how does it make you feel about the potential expiration of the tax cuts? Perhaps worried about the huge increase? Let's take a closer look.

Note that the bottom number on the vertical axis (shown at the far right) is not zero, but rather 34. This means that the bars, in theory, should continue down through the bottom of the page. In fact, the way this is graphed, the visual increase is 460% (the heights of the bars are $35 - 34 = 1$ and $39.6 - 34 = 5.6$, so $(5.6 - 1) / 1 = 460\%$). If we graph the bars with a zero baseline so that the heights are accurately represented (35 and 39.6), we get an actual visual increase of 13% ($(39.6 - 35) / 35$). Let's look at a side-by-side comparison in Figure 2.13.

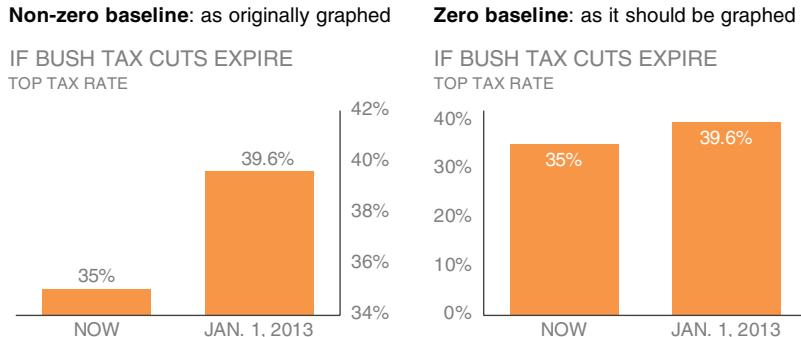


FIGURE 2.13 Bar charts must have a zero baseline

In Figure 2.13, what looked like a huge increase on the left is reduced considerably when plotted appropriately. Perhaps the tax increase isn't so worrisome, or at least not as severe as originally depicted. Because of the way our eyes compare the relative end points of the bars, it's important to have the context of the entire bar there in order to make an accurate comparison.

You'll note that a couple of other design changes were made in the remake of this visual as well. The y-axis labels that were placed on the right-hand side of the original visual were moved to the left (so we see how to interpret the data before we get to the actual data). The data labels that were originally outside of the bars were pulled inside to reduce clutter. If I were plotting this data outside of this specific lesson, I might omit the y-axis entirely and show only the data labels within the bars to reduce redundant information. However, in this case, I preserved the axis to make it clear that it begins at zero.

Graph axis vs. data labels

When graphing data, a common decision to make is whether to preserve the axis labels or eliminate the axis and instead label the data points directly. In making this decision, consider the level of specificity needed. If you want your audience to focus on big-picture trends, think about preserving the axis but deemphasizing it by making it grey. If the specific numerical values are important, it may be better to label the data points directly. In this latter case, it's usually best to omit the axis to avoid the inclusion of redundant information. Always consider how you want your audience to use the visual and construct it accordingly.

The rule we've illustrated here is that *bar charts must have a zero baseline*. Note that this rule does not apply to line graphs. With line graphs, since the focus is on the relative position in space (rather than the length from the baseline or axis), you can get away with a nonzero baseline. Still, you should approach with caution—make it clear to your audience that you are using a nonzero baseline and take context into account so you don't overzoom and make minor changes or differences appear significant.

Ethics and data visualization

But what if changing the scale on a bar chart or otherwise manipulating the data better reinforces the point you want to make? Misleading in this manner by inaccurately visualizing data is not OK. Beyond ethical concerns, it is risky territory. All it takes is one discerning audience member to notice the issue (for example, the y-axis of a bar chart beginning at something other than zero) and your entire argument will be thrown out the window, along with your credibility.

While we're considering lengths of bars, let's also spend a moment on the *width* of bars. There's no hard-and-fast rule here, but in general the bars should be wider than the white space between the bars. You don't want the bars to be so wide, however, that your audience wants to compare areas instead of lengths. Consider the following "Goldilocks" of bar charts: too thin, too thick, and just right.



FIGURE 2.14 Bar width

We've discussed some best practices when it comes to bar charts in general. Next let's take a look at some different varieties. Having a number of bar charts at your disposal gives you flexibility when

facing different data visualization challenges. We'll look at the ones I think you should be familiar with here.

Vertical bar chart

The plain vanilla bar chart is the vertical bar chart, or column chart. Like line graphs, vertical bar charts can be single series, two series, or multiple series. Note that as you add more series of data, it becomes more difficult to focus on one at a time and pull out insight, so use multiple series bar charts with caution. Be aware also that there is visual grouping that happens as a result of the spacing in bar charts having more than one data series. This makes the relative order of the categorization important. Consider what you want your audience to be able to compare, and structure your categorization hierarchy to make that as easy as possible.

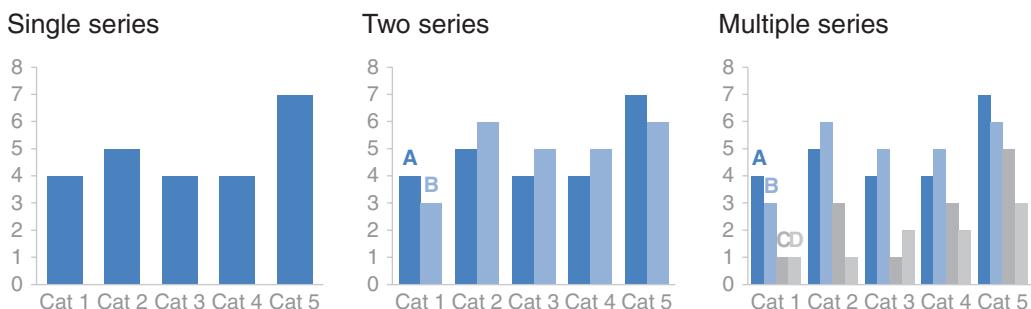


FIGURE 2.15 Bar charts

Stacked vertical bar chart

Use cases for stacked vertical bar charts are more limited. They are meant to allow you to compare totals across categories and also see the subcomponent pieces within a given category. This can quickly become visually overwhelming, however—especially given the varied default color schemes in most graphing applications (more to come on that). It is hard to compare the subcomponents across the various categories once you get beyond the bottom series (the one

directly next to the x-axis) because you no longer have a consistent baseline to use to compare. This makes it a harder comparison for our eyes to make, as illustrated in Figure 2.16.

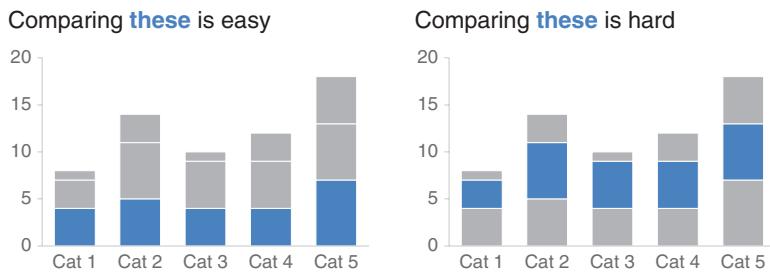


FIGURE 2.16 Comparing series with stacked bar charts

The stacked vertical bar chart can be structured as absolute numbers (where you plot the numbers directly, as shown in Figure 2.16), or with each column summing to 100% (where you plot the percent of total for each vertical segment; we'll look at a specific example of this in Chapter 9). Which you choose depends on what you are trying to communicate to your audience. When you use the 100% stacked bar, think about whether it makes sense to also include the absolute numbers for each category total (either in an unobtrusive way in the graph directly, or possibly in a footnote), which may aid in the interpretation of the data.

Waterfall chart

The waterfall chart can be used to pull apart the pieces of a stacked bar chart to focus on one at a time, or to show a starting point, increases and decreases, and the resulting ending point.

The best way to illustrate the use case for a waterfall chart is through a specific example. Imagine that you are an HR business partner and want to understand and communicate how employee headcount has changed over the past year for the client group you support.

A waterfall chart showing this breakdown might look something like Figure 2.17.

2014 Headcount math

Though more employees transferred out of the team than transferred in, aggressive hiring means overall headcount (HC) increased 16% over the course of the year.



FIGURE 2.17 Waterfall chart

On the left-hand side, we see what the employee headcount for the given team was at the beginning of the year. As we move to the right, first we encounter the incremental additions: new hires and employees transferring into the team from other parts of the organization. This is followed by the deductions: transfers out of the team to other parts of the organization and attrition. The final column represents employee headcount at the end of the year, after the additions and deductions have been applied to the beginning of year headcount.

Brute-force waterfall charts

If your graphing application doesn't have waterfall chart functionality built in, fret not. The secret is to leverage the stacked bar chart and make the first series (the one that appears closest to the x-axis) invisible. It takes a bit of math to set up correctly, but it works great. A blog post on this

topic, along with an example Excel version of the above chart and instructions on how to set one up for your own purposes can be downloaded at storytellingwithdata.com/waterfall-chart.

Horizontal bar chart

If I had to pick a single go-to graph for categorical data, it would be the horizontal bar chart, which flips the vertical version on its side. Why? Because it is *extremely easy to read*. The horizontal bar chart is especially useful if your category names are long, as the text is written from left to right, as most audiences read, making your graph legible for your audience. Also, because of the way we typically process information—starting at top left and making z's with our eyes across the screen or page—the structure of the horizontal bar chart is such that our eyes hit the category names before the actual data. This means by the time we get to the data, we already know what it represents (instead of the darting back and forth our eyes do between the data and category names with vertical bar charts).

Like the vertical bar chart, the horizontal bar chart can be single series, two series, or multiple series (Figure 2.18).

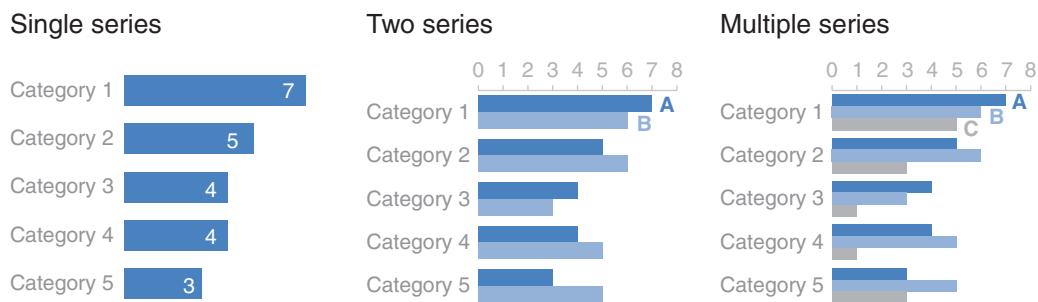


FIGURE 2.18 Horizontal bar charts

The logical ordering of categories

When designing any graph showing categorical data, be thoughtful about how your categories are ordered. If there is a natural ordering to your categories, it may make sense to leverage that. For example, if your categories are age groups—0–10 years old, 11–20 years old, and so on—keep the categories in numerical order. If, however, there isn't a natural ordering in your categories that makes sense to leverage, think about what ordering of your data will make the most sense. Being thoughtful here can mean providing a construct for your audience, easing the interpretation process.

Your audience (without other visual cues) will typically look at your visual starting at the top left and zigzagging in "z" shapes. This means they will encounter the top of your graph first. If the biggest category is the most important, think about putting that first and ordering the rest of the categories in decreasing numerical order. Or if the smallest is most important, put that at the top and order by ascending data values.

For a specific example about the logical ordering of data, check out case study 3 in Chapter 9.

Stacked horizontal bar chart

Similar to the stacked vertical bar chart, stacked horizontal bar charts can be used to show the totals across different categories but also give a sense of the subcomponent pieces. They can be structured to show either absolute values or sum to 100%.

I find this latter approach can work well for visualizing portions of a whole on a scale from negative to positive, because you get a consistent baseline on both the far left and the far right, allowing for easy

comparison of the left-most pieces as well as the right-most pieces. For example, this approach can work well for visualizing survey data collected along a Likert scale (a scale commonly used in surveys that typically ranges from Strongly Disagree to Strongly Agree), as shown in Figure 2.19.

Survey results

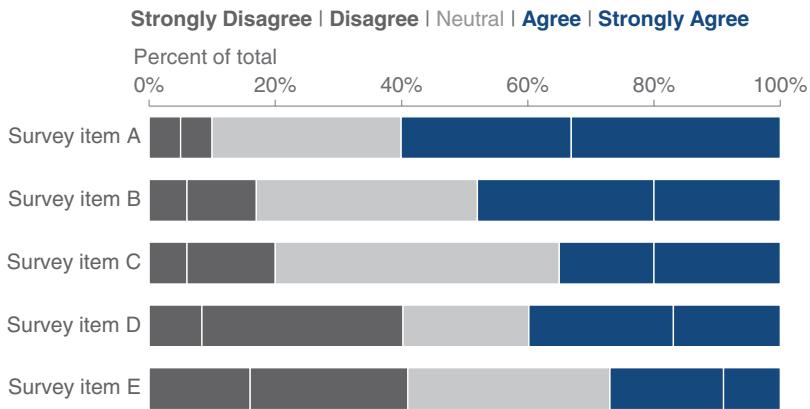


FIGURE 2.19 100% stacked horizontal bar chart

Area

I avoid most area graphs. Humans' eyes don't do a great job of attributing quantitative value to two-dimensional space, which can render area graphs harder to read than some of the other types of visual displays we've discussed. For this reason, I typically avoid them, with one exception—when I need to visualize numbers of vastly different magnitudes. The second dimension you get using a square for this (which has both height and width, compared to a bar that has only height or width) allows this to be done in a more compact way than possible with a single dimension, as shown in Figure 2.20.

Interview breakdown

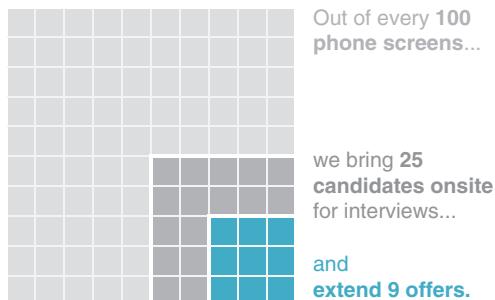


FIGURE 2.20 Square area graph

Other types of graphs

What I've covered up to this point are the types of graphs I find myself commonly using. This is certainly not an exhaustive list. However, they should meet the majority of your everyday needs. Mastering the basics is imperative before exploring novel types of data visualization.

There are many other types of graphs out there. When it comes to selecting a graph, first and foremost, choose a graph type that will enable you to clearly get your message across to your audience. With less familiar types of visuals, you will likely need to take extra care in making them accessible and understandable.

Infographics

*I*nfographic is a term that is frequently misused. An infographic is simply a graphical representation of information or data. Visuals coined *infographic* run the gamut from fluffy to informative. On the inadequate end of the spectrum,

they often include elements like garish, oversized numbers and cartoonish graphics. These designs have a certain visual appeal and can seduce the reader. On second glance, however, they appear shallow and leave a discerning audience dissatisfied. Here, the description of “information graphic”—though often used—is not appropriate. On the other end of the spectrum are infographics that live up to their name and actually inform. There are many good examples in the area of data journalism (for example, the *New York Times* and *National Geographic*).

There are critical questions information designers must be able to answer before they begin the design process. These are the same questions we’ve discussed when it comes to understanding the context for storytelling with data. Who is your audience? What do you need them to know or do? It is only after the answers to these questions can be succinctly articulated that an effective method of display that will best aid the message can be chosen. Good data visualization—infographic or otherwise—is not simply a collection of facts on a given topic; good data visualization tells a story.

To be avoided

We’ve discussed the visuals that I use most commonly to communicate data in a business setting. There are also some specific graph types and elements that you should avoid: pie charts, donut charts, 3D, and secondary y-axes. Let’s discuss each of these.

Pie charts are evil

I have a well-documented disdain for pie charts. In short, they are evil. To understand how I arrived at this conclusion, let’s look at an example.

The pie chart shown in Figure 2.21 (based on a real example) shows market share across four suppliers: A, B, C, and D. If I asked you to make a simple observation—which supplier is the largest based on this visual—what would you say?

Supplier Market Share

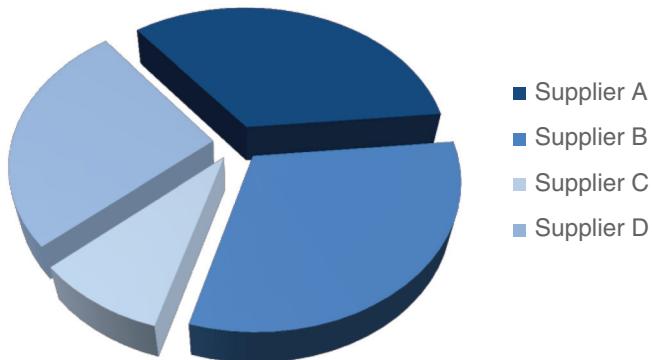


FIGURE 2.21 Pie chart

Most people will agree that “Supplier B,” rendered in medium blue at the bottom right, appears to be the largest. If you had to estimate what proportion supplier B makes up of the overall market, what percent might you estimate?

35%?

40%?

Perhaps you can tell by my leading questioning that something fishy is going on here. Take a look at what happens when we add the numbers to the pie segments, as shown in Figure 2.22.

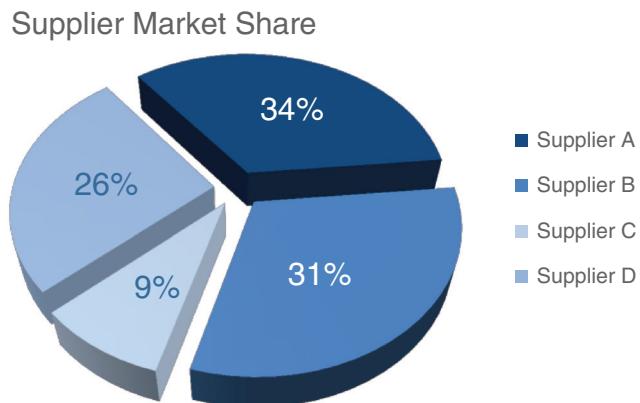


FIGURE 2.22 Pie chart with labeled segments

"Supplier B"—which *looks* largest, at 31%—is actually smaller than "Supplier A" above it, which looks smaller.

Let's discuss a couple of issues that pose a challenge for accurately interpreting this data. The first thing that catches your eye (and suspicion, if you're a discerning chart reader) is the 3D and strange perspective that's been applied to the graph, tilting the pie and making the pieces at the top appear farther away and thus smaller than they actually are, while the pieces at the bottom appear closer and thus bigger than they actually are. We'll talk more about 3D soon, but for now I'll articulate a relevant data visualization rule: *don't use 3D!* It does nothing good, and can actually do a whole lot of harm, as we see here with the way it skews the visual perception of the numbers.

Even when we strip away the 3D and flatten the pie, interpretation challenges remain. The human eye isn't good at ascribing quantitative value to two-dimensional space. Said more simply: *pie charts are hard for people to read*. When segments are close in size, it's difficult (if not impossible) to tell which is bigger. When they aren't close in size, the best you can do is determine that one is bigger than the other, but you can't judge by how much. To get over this, you can add data labels as has been done here. But I'd still argue the visual isn't worth the space it takes up.

What should you do instead? One approach is to replace the pie chart with a horizontal bar chart, as illustrated in Figure 2.23, organized from greatest to least or vice versa (unless there is some natural ordering to the categories that makes sense to leverage, as mentioned earlier). Remember, with bar charts, our eyes compare the end points. Because they are aligned at a common baseline, it is easy to assess relative size. This makes it straightforward to see not only which segment is the largest, for example, but also *how incrementally larger* it is than the other segments.

Supplier Market Share

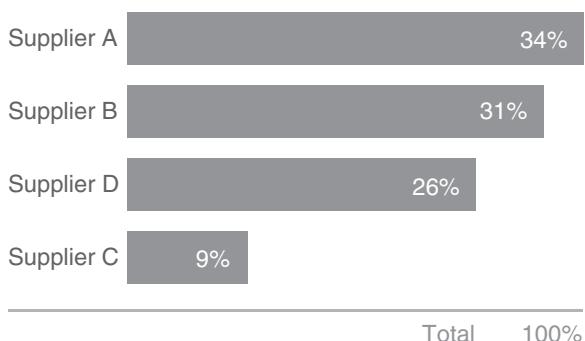


FIGURE 2.23 An alternative to the pie chart

One might argue that you lose something in the transition from pie to bar. The unique thing you get with a pie chart is the concept of there being a whole and, thus, parts of a whole. But if the visual is difficult to read, is it worth it? In Figure 2.23, I've tried to address this by showing that the pieces sum to 100%. It isn't a perfect solution, but something to consider. For more alternatives to pie charts, check out case study 5 in Chapter 9.

If you find yourself using a pie chart, pause and ask yourself: *why?* If you're able to answer this question, you've probably put enough thought into it to use the pie chart, but it certainly shouldn't be the first type of graph that you reach for, given some of the difficulties in visual interpretation we've discussed here.

While we're on the topic of pie charts, let's look quickly at another "dessert visual" to avoid: the donut chart.

The donut chart

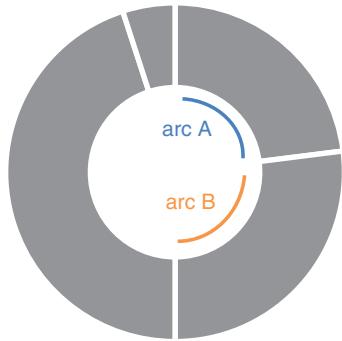


FIGURE 2.24 Donut chart

With pies, we are asking our audience to compare angles and areas. With a donut chart, we are asking our audience to compare one arc length to another arc length (for example, in Figure 2.24, the length of *arc A* compared to *arc B*). How confident do you feel in your eyes' ability to ascribe quantitative value to an arc length?

Not very? That's what I thought. Don't use donut charts.

Never use 3D

One of the golden rules of data visualization goes like this: never use 3D. Repeat after me: never use 3D. The only exception is if you are actually *plotting a third dimension* (and even then, things get really tricky really quickly, so take care when doing this)—and you should never use 3D to plot a single dimension. As we saw in the pie chart example previously, 3D skews our numbers, making them difficult or impossible to interpret or compare.

Adding 3D to graphs introduces unnecessary chart elements like side and floor panels. Even worse than these distractions, graphing

applications do some pretty strange things when it comes to plotting values in 3D. For example, in a 3D bar chart, you might think that your graphing application plots the front of the bar or perhaps the back of the bar. Unfortunately, it's often even less straightforward than that. In Excel, for example, the bar height is determined by an invisible tangent plane intersecting the corresponding height on the y-axis. This gives rise to graphs like the one shown in Figure 2.25.

Number of issues

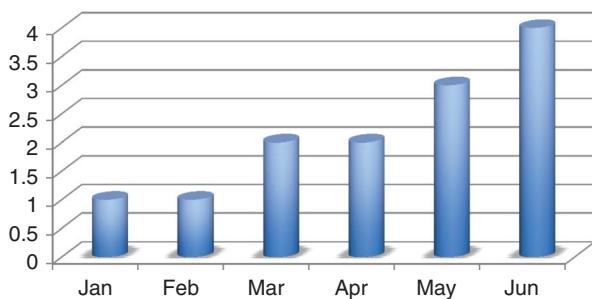


FIGURE 2.25 3D column chart

Judging by Figure 2.25, how many issues were there in January and February? I've plotted a single issue for each of these months. However, the way I read the chart, if I compare the bar height to the grid-lines and follow it leftward to the y-axis, I'd estimate visually a value of maybe 0.8. This is simply bad data visualization. Don't use 3D.

Secondary y-axis: generally not a good idea

Sometimes it's useful to be able to plot data that is in entirely different units against the same x-axis. This often gives rise to the secondary y-axis: another vertical axis on the right-hand side of the graph. Consider the example shown in Figure 2.26.

Secondary y-axis



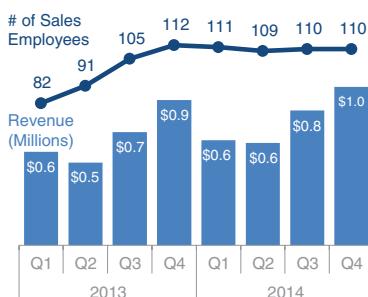
FIGURE 2.26 Secondary y-axis

When interpreting Figure 2.26, it takes some time and reading to understand which data should be read against which axis. Because of this, you should avoid the use of a secondary or right-hand y-axis. Instead, think about whether one of the following approaches will meet your needs:

1. Don't show the second y-axis. Instead, label the data points that belong on this axis directly.
2. Pull the graphs apart vertically and have a separate y-axis for each (both along the left) but leverage the same x-axis across both.

Figure 2.27 illustrates these options.

Alternative 1: label directly



Alternative 2: pull apart vertically

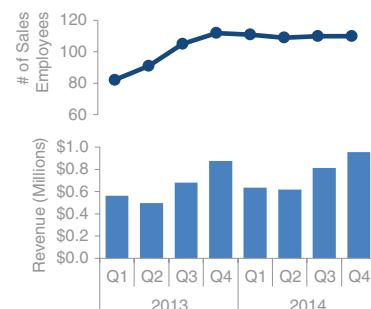


FIGURE 2.27 Strategies for avoiding a secondary y-axis

A third potential option not shown here is to link the axis to the data to be read against it through the use of color. For example, in the original graph depicted in Figure 2.26, I could write the left y-axis title “Revenue” in blue and keep the revenue bars blue while at the same time writing the right y-axis title “# of Sales Employees” in orange and making the line graph orange to tie these together visually. I don’t recommend this approach because color can typically be used more strategically. We’ll spend a lot more time discussing color in Chapter 4.

It is also worth noting that when you display two datasets against the same axis, it can imply a relationship that may or may not exist. This is something to be aware of when determining whether this is an appropriate approach in the first place.

When you’re facing a secondary y-axis challenge and considering which alternative shown in Figure 2.27 will better meet your needs, think about the level of specificity you need. Alternative 1, where each data point is labeled explicitly, puts more attention on the specific numbers. Alternative 2, where the axes are shown at the left, puts more focus on the overarching trends. In general, avoid a secondary y-axis and instead employ one of these alternate approaches.

In closing

In this chapter, we've explored the types of visual displays I find myself using most. There will be use cases for other types of visuals, but what we've covered here should meet the majority of everyday needs.

In many cases, there isn't a single correct visual display; rather, often there are different types of visuals that could meet a given need. Drawing from the previous chapter on context, most important is to have that need clearly articulated: *What do you need your audience to know?* Then choose a visual display that will enable you to make this clear.

If you're wondering *What is the right graph for my situation?*, the answer is always the same: whatever will be easiest for your audience to read. There is an easy way to test this, which is to create your visual and show it to a friend or colleague. Have them articulate the following as they process the information: where they focus, what they see, what observations they make, what questions they have. This will help you assess whether your visual is hitting the mark, or in the case where it isn't, help you know where to concentrate your changes.

You now know the second lesson of storytelling with data: how to **choose an appropriate visual display**.