STEVE WEXLER | JEFFREY SHAFFER | ANDY COTGREAVE

THE BIG BOOK OF DASHBOARDS

Visualizing Your Data Using Real-World Business Scenarios

WILEY

THE BIG BOOK OF DASHBOARDS

THE BIG BOOK OF DASHBOARDS

Visualizing Your Data Using Real-World Business Scenarios

STEVE WEXLER | JEFFREY SHAFFER | ANDY COTGREAVE

WILEY

Cover image: Course Metrics Dashboard by Jeffrey Shaffer
Cover design: Wiley

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

Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data

Names: Wexler, Steve, author. | Shaffer, Jeffrey, author. | Cotgreave, Andy, author.
Title: The big book of dashboards : visualizing your data using real-world business scenarios / Steve Wexler, Jeffrey Shaffer, Andy Cotgreave.
Description: Hoboken : Wiley, 2017. | Includes index.
Identifiers: LCCN 2016052146| ISBN 9781119282716 (paperback) | ISBN 9781119282785 (Adobe PDF) | ISBN 9781119282730 (epub)
Subjects: LCSH: Dashboards (Management information systems) | Organizational effectiveness—Evaluation. | Management—Evaluation. | BISAC: BUSINESS & ECONOMICS / Business Communication / Meetings & Presentations.
Classification: LCC HD30.213 .W43 2017 | DDC 658.4/038011—dc23 LC record available at https://lccn.loc.gov/2016052146

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Contents

Acknowledgments

From the three of us

Stephen Few, whose books have made a profound and lasting impression on us.

Alberto Cairo for his invaluable feedback and for his leadership in the data visualization community.

Our technical reviewers greatly improved our first drafts. Thanks to Troy Magennis, Andy Kirk, Jon Schwabish, Ariel Pohoryles, Trudy Weiss Craig, Michael Fry, Andy Kriebel, and a special thanks to Cole Nussbaumer Knaflic for introducing us to the Wiley team and who went far beyond our expectations with her detailed edits and comments.

All the contributors to this book gave significant time to tweak their dashboards according to our requests. We thank you for allowing us to include your work in the book.

Thanks, also, to Mark Boone, KK Molugu, Eric Duell, Chris DeMartini, and Bob Filbin for their efforts.

Our stellar team at Wiley: acquisitions editor Bill Falloon for fighting so hard on our behalf; editor Christina Verigan for her deft reworking and invaluable help optimizing flow; senior production editor Samantha Hartley for overseeing the daunting process of making this book a beautiful, tangible thing; copy editor Debra Manette for such detailed editing and insights; proofreader Hope Breeman for her meticulous proof check; the team at WordCo for a comprehensive index and marketing manager Heather Dunphy for her exceptional expertise in connecting author with audience.

From Steve

My wife, Laura, and my daughters, Janine and Diana, for the never-ending support and love.

Ira Handler and Brad Epstein, whose friendship, encouragement, and example have been a godsend for the past dozen years.

Joe Mako, who has always been willing to help me with "the difficult stuff" and provided much needed encouragement when I was starting out.

The Princeton University Triangle Club, where I learned how to bring talented people together to make wonderful things. Without my experiences there I don't know if I would have had the insight and ability to recruit my fellow authors.

Jeff and Andy, who not only made the book way better than it would have been had I tackled it on my own, but for providing me with one of the most rewarding and enriching experiences of my career. Your abilities, candor, humor, grit, patience, impatience, thoughtfulness, and leadership made for a remarkable ride.

Finally, to Liz, my wife, and my daughters, Beatrice and Lucy. Thank you for your support and the freedom to abandon you all on weekends, mornings, and evenings in order to compete this project. I could not have done it without you.

From Andy

I would like to thank Steve and Jeff for approaching me to join this project. I'd been procrastinating on writing a book for many years, and the opportunity to work with two passionate, skilled leaders was the trigger I needed to get going. I would like to thank them both for many hours of constructive debate (argument?) over the rights and wrongs of all aspects of dashboards and data visualization. It has been an enriching experience.

From Jeff

Thank you, Steve and Andy. It was a pleasure working with you guys. I will miss the collaboration, especially our many hours of discussion about data visualization and dashboard design.

A special thank you to Mary, my wife, and to Nina and Elle, my twin daughters, for sacrificing lots of family time over many long nights and weekends. I would not have been able to complete this project without your support.

About the Authors

Steve Wexler has worked with ADP, Gallup, Deloitte, Convergys, Consumer Reports, *The Economist*, ConEd, D&B, Marist, Tradeweb, Tiffany, McKinsey & Company, and many other organizations to help them understand and visualize their data. Steve is a Tableau Zen Master, Iron Viz Champion, and Tableau Training Partner.

His presentations and training classes combine an extraordinary level of product mastery with the real-world experience gained through developing thousands of visualizations for dozens of clients. In addition to his recognized expertise in data visualization and Tableau, Steve has decades of experience as a successful instructor in all areas of computer-based technology. Steve has taught thousands of people in both large and small organizations and is known for conducting his seminars with clarity, patience, and humor.

Website: DataRevelations.com

Jeffrey A. Shaffer is Vice President of Information Technology and Analytics at Recovery Decision Science and Unifund. He is also Adjunct Professor at the University of Cincinnati, where he teaches Data Visualization and was named the 2016 Outstanding Adjunct Professor of the Year.

He is a regular speaker on the topic of data visualization, data mining, and Tableau training at conferences, symposiums, workshops, universities, and corporate training programs. He is a Tableau Zen Master, and was the winner of the 2014 Tableau Quantified Self Visualization Contest, which led him to compete in the 2014 Tableau Iron Viz Contest. His data visualization blog was on the shortlist for the 2016 Kantar Information is Beautiful Awards for Data Visualization Websites.

Website: DataPlusScience.com

Andy Cotgreave is Technical Evangelist at Tableau Software. He has over 10 years' experience in data visualization and business intelligence, first honing his skills as an analyst at the University of Oxford. Since joining Tableau in 2011, he has helped and inspired thousands of people with technical advice and ideas on how to build a data-driven culture in a business.

In 2016 he ran the MakeoverMonday (http://www.makeovermonday.co.uk/) project with Andy Kriebel, a social data project which saw over 500 people make 3,000 visualizations in one year. The project received an honourable mention in the Dataviz

Project category of the 2016 Kantar Information is Beautiful Awards.

Andy has spoken at conferences around the world, including SXSW, Visualized, and Tableau's customer conferences. He writes a column for *Computerworld*, Living with Data (http://www.computerworld.com/blog/living-data/), as well as maintaining his own blog, GravyAnecdote.com.

Website: GravyAnecdote.com

Introduction

We wrote *The Big Book of Dashboards* for anyone tasked with building or overseeing the development of business dashboards. Over the past decade, countless people have approached us after training sessions, seminars, or consultations, shown us their data, and asked: *"What would be a really good way to show this?"*

These people faced a specific business predicament (what we call a "scenario") and wanted guidance on how to best address it with a dashboard. In reviewing dozens of books about data visualization, we were surprised that, while they contained wonderful examples showing why a line chart often works best for time-series data and why a bar chart is almost always better than a pie chart, none of them matched great dashboards with real-world business cases. After pooling our experience and enormous collection of dashboards, we decided to write our own book.

How This Book Is Different

This book is not about the fundamentals of data visualization. That has been done in depth by many amazing authors. We want to focus on proven, real-world examples and why they succeed.

However, if this is your first book about the topic of data visualization, we do provide a primer in Part I with everything you need to know to understand how the charts in the scenarios work. We also dearly hope it whets your appetite for more, which is why this section finishes with our recommended further reading.

How This Book Is Organized

The book is organized into three parts.

Part I: A Strong Foundation. This part covers the fundamentals of data visualization and provides our crash course on the foundational elements that give you the vocabulary you need to explore and understand the scenarios.

Part II: The Scenarios. This is the heart of the book, where we describe dozens of different business scenarios and then present a dashboard that "solves" the challenges presented in those scenarios.

Part III: Succeeding in the Real World. The chapters in this part address problems we've encountered and anticipate you may encounter as well. With these chapters—distilled from decades of real-world experience—we hope to make your journey quite a bit easier and a lot more enjoyable.

How to Use This Book

We encourage you to look through the book to find a scenario that most closely matches what you are tasked with visualizing. Although there might not be an exact match, our goal is to present enough scenarios that you can find something that will address your needs. The internal conversation in your head might go like this:

"Although my data isn't exactly the same as what's in this scenario, it's close enough, and this dashboard really does a great job of helping me and others see and understand that data. I think we should use this approach for our project as well."

For each scenario we present the entire dashboard at the beginning of the chapter, then explore how individual components work and contribute to the whole.

By organizing the book based on these scenarios and offering practical and effective visualization examples, we hope to make *The Big Book of Dashboards* a trusted resource that you open when you need to build an effective business dashboard. To ensure you get the most out of these examples, we have included a visual glossary at the back of this book. If you come across an unfamiliar term, such as "sparkline," you can look it up and see an illustration.

We also encourage you to spend time with *all* the scenarios and the proposed solutions as there may be some elements of a seemingly irrelevant scenario that may apply to your own needs.

For example, Chapter 11 shows a dashboard used by a team in the English Premier League to help players understand their performance. Your data might have nothing to do with sports, but the dashboard is a great example of showing current and historical performance. (See Figure I.1.) That might be something you have to do with your data. Plus, if you skip one scenario, you might miss a great example of the exact chart you need for your own solution.

We also encourage you to browse the book for motivation. Although a scenario may not be a perfect match, the thought process and chart choices may inspire you.

Succeeding in the Real World

In addition to the scenarios, an entire section of the book is devoted to addressing many practical and psychological factors you will encounter in your work. It's great to have theory- and evidenced-based research at your disposal, but what will you do when somebody asks you to make your dashboard "cooler" by adding packed bubbles and donut charts?

The three of us have a combined 30-plus years of hands-on experience helping people in hundreds of organizations build effective visualizations. We have fought (and sometimes lost) many "best practices" battles. But by having endured these struggles, we bring an uncommon empathy to the readers of this book.

We recognize that at times readers will be asked to create dashboards and charts that exemplify bad practice. For example, a client or a department head may stipulate using a particular combination of colors or demand a chart type that is against evidence-based data visualization best practices.

We hear you. We've been there.

FIGURE I.1 A player summary from an English Premier League Club

(Note: Fake data is used.)

Although the dashboard in Figure I.1 pertains to sports, the techniques are universal. Here the latest event is in yellow, the five most recent events are in red, and older events are in a muted gray. Brilliant.

We've faced many of the hurdles you will encounter and the concepts you will grapple with in your attempt to build dashboards that are informative, enlightening, and engaging. The essays in this section will help smooth the way for you by offering suggestions and alternatives for these issues.

What to Do and What *Not* to Do

Although the book is an attempt to celebrate good examples, we'll also show plenty of *bad* examples. We guarantee you will see this kind of work out in the wild, and you may even be asked to emulate it. We mark these "bad" examples with the cat icon shown in Figure I.2 so that you don't have to read the surrounding text to determine if the chart is something you should emulate or something you should avoid.

FIGURE I.2 If you see this icon, it means don't make a chart like this one.

Illustration by Eric Kim

What Is a Dashboard?

Ask 10 people who build business dashboards to define a dashboard and you will probably get 10 different definitions. For the purpose of this book, our definition is as follows:

> A dashboard is a visual display of data used to monitor conditions and/or facilitate understanding.

This is a broad definition, and it means that we would consider all of the examples listed below to be dashboards:

- An interactive display that allows people to explore worker compensation claims by region, industry, and body part
- A PDF showing key measures that gets e-mailed to an executive every Monday morning
- A large wall-mounted screen that shows support center statistics in real time
- A mobile application that allows sales managers to review performance across different regions and compare year-to-date sales for the current year with the previous year

Even if you don't consider every example in this book a true dashboard, we think you will find the discussion and analysis around each of the scenarios helpful in building your solutions. Indeed, we can debate the definition until we are blue in the face, but that would be a horrible waste of effort as it simply isn't that important. What is important—make that *essential*—is understanding how to combine different elements (e.g., charts, text, legends, filters, etc.) into a cohesive and coordinated whole that allows people to see and understand their data.

Final Thought: There Are No Perfect Dashboards

You will not find any perfect dashboards in this book.

In our opinion, there is no such thing as a perfect dashboard. You will never find one perfect collection of charts that ideally suits every person who may encounter it. But, although they may not be perfect, the dashboards we showcase in the book successfully help people see and understand data in the real world.

The dashboards we chose all have this in common: Each one demonstrates some great ideas in a way that is relevant to the people who need to understand them. In short, they all serve the end users. Would we change some of the dashboards? Of course we would, and we weigh in on what we would change in the author commentary at the end of each scenario. Sometimes we think a chart choice isn't ideal; other times, the layout isn't quite right; and in some cases, the interactivity is clunky or difficult. What we recognize is that every set of eyes on a dashboard will judge the work differently, which is something you also should keep in mind. Where you see perfection, others might see room for improvement. The challenge all the dashboard designers in this book have faced is balancing a dashboard's presentation and objectives with time and efficiency. It's not an easy spot to hit, but with this book we hope to make it easier for you.

Steve Wexler

Jeffrey Shaffer

Andy Cotgreave

PART I

A STRONG FOUNDATION

Chapter 1

Data Visualization: A Primer

This book is about real-world dashboards and why they succeed. In many of the scenarios, we explain how the designers use visualization techniques to contribute to that success. For those new to the field, this chapter is a primer on data visualization. It provides enough information for you to understand why we picked many of the dashboards. If you are more experienced, this chapter recaps data visualization fundamentals.

Why Do We Visualize Data?

Let's see why it's vital to visualize numbers by beginning with Table 1.1. There are four groups of numbers, each with 11 pairs. In a moment, we will create a chart from them, but before we do, take a look at the numbers. What can you see? Are there any discernible differences in the patterns or trends among them?

Let me guess: You don't really see anything clearly. It's too hard.

Before we put the numbers in a chart, we might consider their statistical properties. Were we to do that, we'd find that the statistical properties of each group of numbers are very similar. If the table doesn't show anything and statistics don't reveal much, what happens when we *plot* the numbers? Take a look at Figure 1.1.

Now do you see the differences? *Seeing* the numbers in a chart shows you something that tables and some statistical measures cannot. We visualize data to harness the incredible power of our visual system to spot relationships and trends.

This brilliant example is the creation of Frank Anscombe, a British statistician. He created this set

TABLE 1.1 Table with four groups of numbers: What do they tell you?

Group A		Group B		Group C		Group D	
x	y	x	y	x	y	x	y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

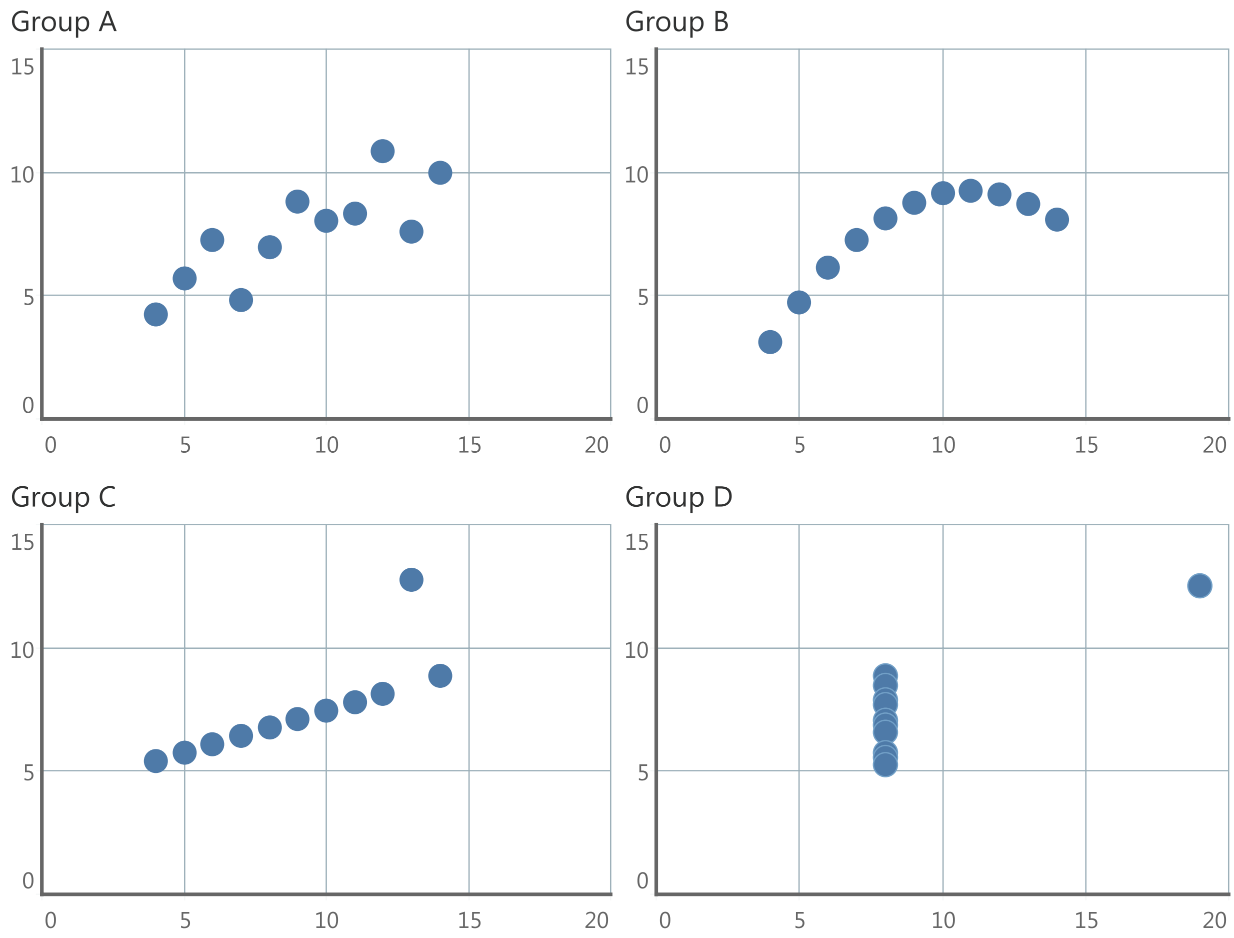

FIGURE 1.1 Now can you see a difference in the four groups?

of numbers—called "Anscombe's Quartet"—in his paper "Graphs in Statistical Analysis" in 1973. In the paper, he fought against the notion that "numerical calculations are exact, but graphs are rough."

Another reason to visualize numbers is to help our memory. Consider Table 1.2, which shows sales numbers for three categories, by quarter, over a four-year period. What trends can you see?

Identifying trends is as hard as it was with Anscombe's Quartet. To read the table, we need to look up every value, one at a time. Unfortunately, our short-term memories aren't designed to store many pieces of information. By the time we've reached the fourth or fifth number, we will have forgotten the first one we looked at.

Let's try a trend line, as shown in Figure 1.2.

TABLE 1.2 What are the trends in sales?

Category	2013 Q1	2013 Q2	2013 Q3	2013 Q4	2014 Q1	2014 Q2	2014 Q3	2014 Q4
Furniture	$463,988	$352,779	$338,169	$317,735	$320,875	$287,934	$319,537	$324,319
Office Supplies	$232,558	$290,055	$265,083	$246,946	$219,514	$202,412	$198,268	$279,679
Technology	$563,866	$244,045	$432,299	$461,616	$285,527	$353,237	$338,360	$420,018
Category	**2015 Q1**	**2015 Q2**	**2015 Q3**	**2015 Q4**	**2016 Q1**	**2016 Q2**	**2016 Q3**	**2016 Q4**
Furniture	$307,028	$273,836	$290,886	$397,912	$337,299	$245,445	$286,972	$313,878
Office Supplies	$207,363	$183,631	$191,405	$217,950	$241,281	$286,548	$217,198	$272,870
Technology	$333,002	$291,116	$356,243	$386,445	$386,387	$397,201	$359,656	$375,229

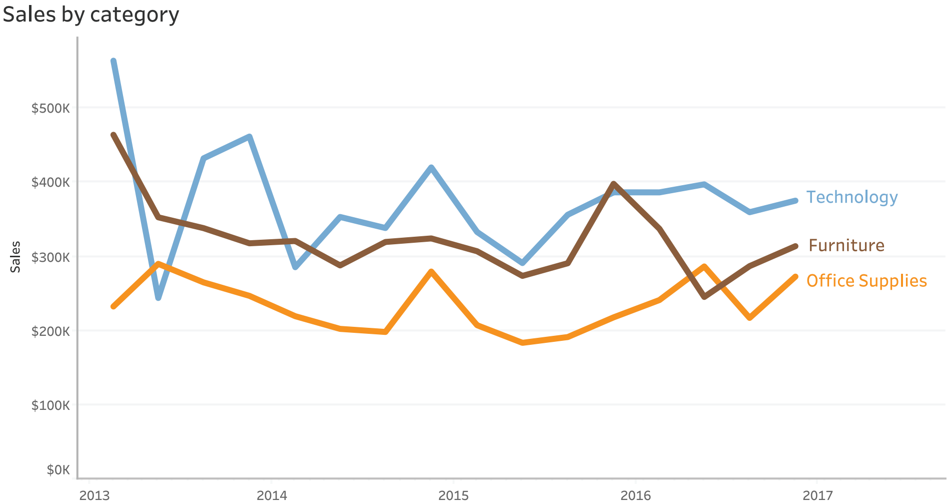

FIGURE 1.2 Now can you see the trends?

Now we have much better insight into the trends. Office supplies has been the lowest-selling product category in all but two quarters. Furniture trends have been dropping slowly over the time period, except for a bump in sales in 2015 Q4 and a rise in the last two quarters. Technology sales have mostly been the highest but were particularly volatile at the start of the time period.

The table and the line chart each visualized the same 48 data points, but only the line chart lets us see the trends. The line chart turned 48 data points into three *chunks* of data, each containing 16 data points. Visualizing the data hacks our short-term memory; it allows us to interpret large volumes of data instantly.

How Do We Visualize Data?

We've just looked at some examples of the power of visualizing data. Now we need to move on to *how* we build the visualizations. To do that, we first need to look at two things: preattentive attributes and types of data.

Preattentive Attributes

Visualizing data requires us to turn data into marks on a canvas. What kind of marks make the most sense? One answer lies in what are called "preattentive attributes." These are things that our brain processes in milliseconds, before we pay attention to everything else. There are many different types. Let's look at an example.

Look at the numbers in Figure 1.3. How many 9s are there?

How did you do? It's easy to answer the question—you just look at all the values and count the 9s—but it takes a long time. We can make one change to the grid and make it very easy for you. Have a look at Figure 1.4.

Now the task is easy. Why? Because we changed the color: 9s are red, and all the other numbers are light gray.

Color differences pop out. It's as easy to find one red 9 on a table of hundreds of digits as it is on a 10-by-10 grid. Think about that for a moment: Your brain registers the red 9s before you consciously addressed the grid to count them. Check out the grid of 2,500 numbers in Figure 1.5. Can you see the 9?

It's easy to spot the 9. Our eyes are amazing at spotting things like this.

2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

FIGURE 1.3 How many 9s are there?

2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

FIGURE 1.4 Now it's easy to count the 9s.

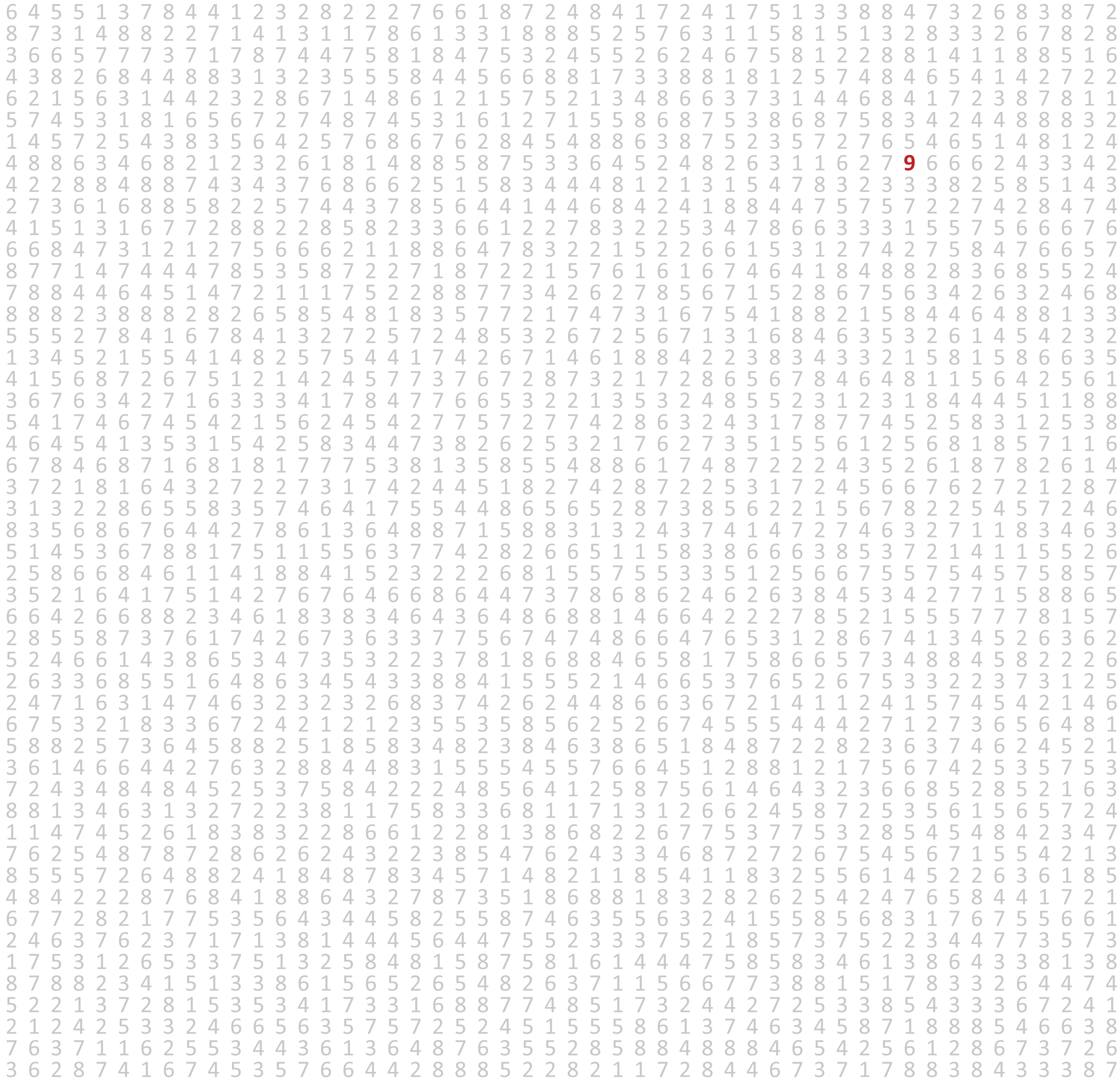

FIGURE 1.5 There is a single 9 in this grid of 2,500 numbers. We wager you saw it before you started reading any other numbers on this page.

2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

FIGURE 1.6 Differences in size are easy to see too.

2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

FIGURE 1.7 Coloring every digit is nearly as bad as having no color.

Color (in this case, hue) is one of several *preattentive attributes*. When we look at a scene in front of us, or a chart, we process these attributes in under 250 milliseconds. Let's try out a couple more preattentive features with our table of 9s. In Figure 1.6, we've made the 9s a different size from the rest of the figures.

Size and hue: Aren't they amazing? That's all very well when counting the 9s. What if our task is to count the frequency of *each* digit? That's a slightly more realistic task, but we can't just use a different color or size for each digit. That would defeat the preattentive nature of the single color. Look at the mess that is Figure 1.7.

It's not a complete disaster: If you're looking for the 6s, you just need to work out that they are red and then scan quickly for those. Using one color on a visualization is highly effective to make one category stand out. Using a few colors, as we did in Figure 1.2 to distinguish a small number of categories, is fine too. Once you're up to around eight to ten categories, however, there are too many colors to easily distinguish one from another.

To count each digit, we need to aggregate. Visualization is, at its core, about encoding aggregations, such as frequency, in order to gain insight. We need to move away from the table entirely and encode the frequency of each digit. The most effective way is to use length, which we can do in a bar chart. Figure 1.8 shows the frequency of each digit. We've also colored the bar showing the number 9.

Since the task is to count the 9s in the data source, the bar chart is one of the best ways to see the results. This is because length and position are best for quantitative comparisons. If we extend the example one final time and consider which numbers are most common, we could sort the bars, as shown in Figure 1.9.

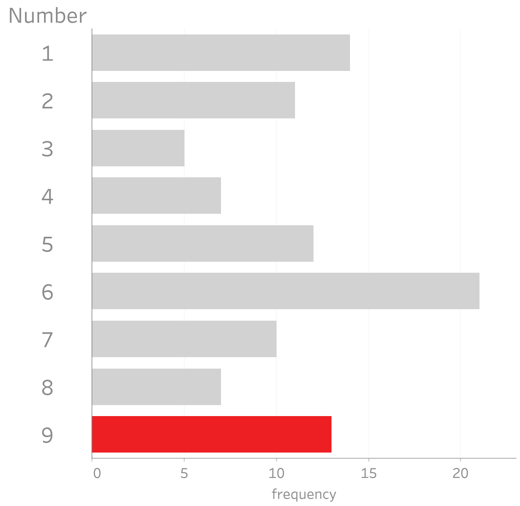

FIGURE 1.8 There are 13 9s.

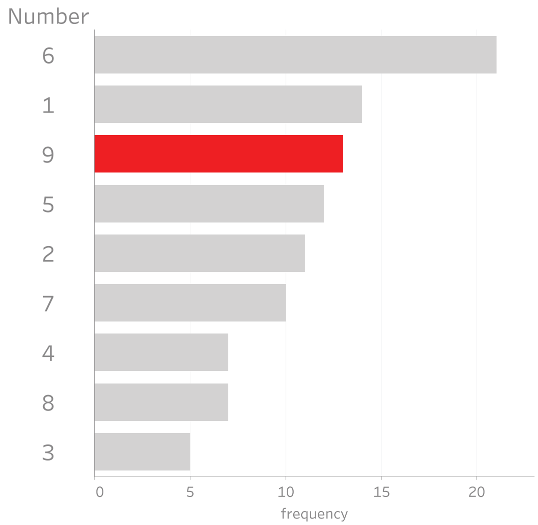

FIGURE 1.9 Sorted bar chart using color and length to show how many 9s are in our table.

This series of examples with the 9s reemphasizes the importance of visualizing data. As with Anscombe's Quartet, we went from a difficult-to-read table of numbers to an easy-to-read bar chart. In the sorted bar chart, not only can we count the 9s (the original task), but we also know that 9 was the third most common digit in the table. We can also see the frequency of every other digit.

The series of examples we just presented used color, size, and length to highlight the 9s. These are three of many preattentive attributes. Figure 1.10 shows 12 that are commonly used in data visualization.

Some of them will be familiar to you from charts you have already seen. Anscombe's Quartet (see Figure 1.1) used position and spatial grouping. The x- and y-coordinates are for position, while spatial grouping allows us to see the outliers and the patterns.

Preattentive attributes provide us with ways to encode our data in charts. We'll look into that in more detail in a moment, but not before we've talked about data.

To recap, we've seen how powerful the visual system is and looked at some visual features we can use to display data effectively. Now we need to look at the different types of data, in order to choose the best visual encoding for each type.

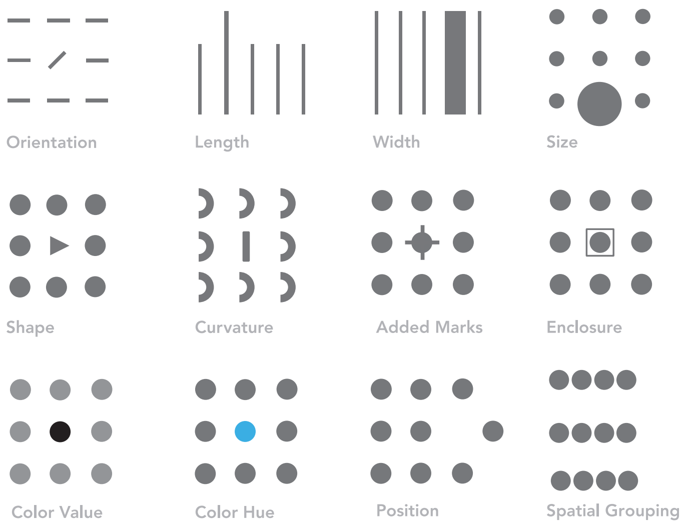

FIGURE 1.10 Preattentive features.

Types of Data

There are three types of data: categorical, ordinal, and quantitative. Let's use a photo to help us define each type.

Categorical Data

Categorical (or *nominal*) *data* represents *things*. These things are mutually exclusive labels without any numerical value. What nominal data can we use to describe the gentleman with me in the Figure 1.11?

- His *name* is Brent Spiner.
- By *profession* he is an actor.
- He played the *character* Data in the TV show *Star Trek: The Next Generation*.

FIGURE 1.11 One of your authors (Andy, on the right) with a celebrity.
Source: Author's photograph

Name, profession, character, and TV show are all categorical data types. Other examples include gender, product category, city, and customer segment.

Ordinal Data

Ordinal data is similar to categorical data, except it has a clear order. Referring to Brent Spiner:

- Brent Spiner's *date of birth* is Wednesday, February 2, 1949.
- He appeared in all seven *seasons* of *Star Trek: The Next Generation*.
- Data's *rank* was lieutenant commander.
- Data was the fifth of six *androids made* by Dr. Noonien Soong.

Other types of ordinal data include education experience, satisfaction level, and salary bands in an organization. Although ordinal values often have numbers associated with them, the interval between those values is arbitrary. For example, the difference in an organization between pay scales 1 and 2 might be very different from that between pay scales 4 and 6.

Quantitative Data

Quantitative data is the numbers. *Quantitative* (or *numerical*) data is data that can be measured and aggregated.

- Brent Spiner's *date of birth* is Wednesday, February 2, 1949.
- His *height* is 5 ft 9 in (180 cm) tall.
- He made 177 *appearances* in episodes of *Star Trek*.
- Data's positronic brain is capable of 60 trillion *operations per second*.

You'll have noticed that date of birth appears in both ordinal and quantitative data types. Time is unusual in that it can be both. In Chapter 31, we look in detail about how you treat time influences your choice of visualization types.

Other types of quantitative measures include sales, profit, exam scores, pageviews, and number of patients in a hospital.

Quantitative data can be expressed in two ways: as discrete or continuous data. Discrete data is presented at predefined, exact points—there's no "in between." For example, Brent Spiner appeared in 177 episodes of *Star Trek*; he couldn't have appeared in 177.5 episodes. Continuous data allows for the "in between," as there is an infinite number of possible intermediate values. For example, Brent Spiner grew to a height of 5 ft 9 in but at one point in his life he was 4 ft 7.5 in tall.

Encoding Data in Charts

We've now looked at preattentive attributes and the three types of data. It's time to see how to combine that knowledge into building charts. Let's look at some charts and see how they encode the different types of data. Sticking with *Star Trek*, Figure 1.12 shows the IMDB.com ratings of every episode of *Star Trek: The Next Generation*.

Table 1.3 shows the different types of data, what type it is, and how it's been encoded.

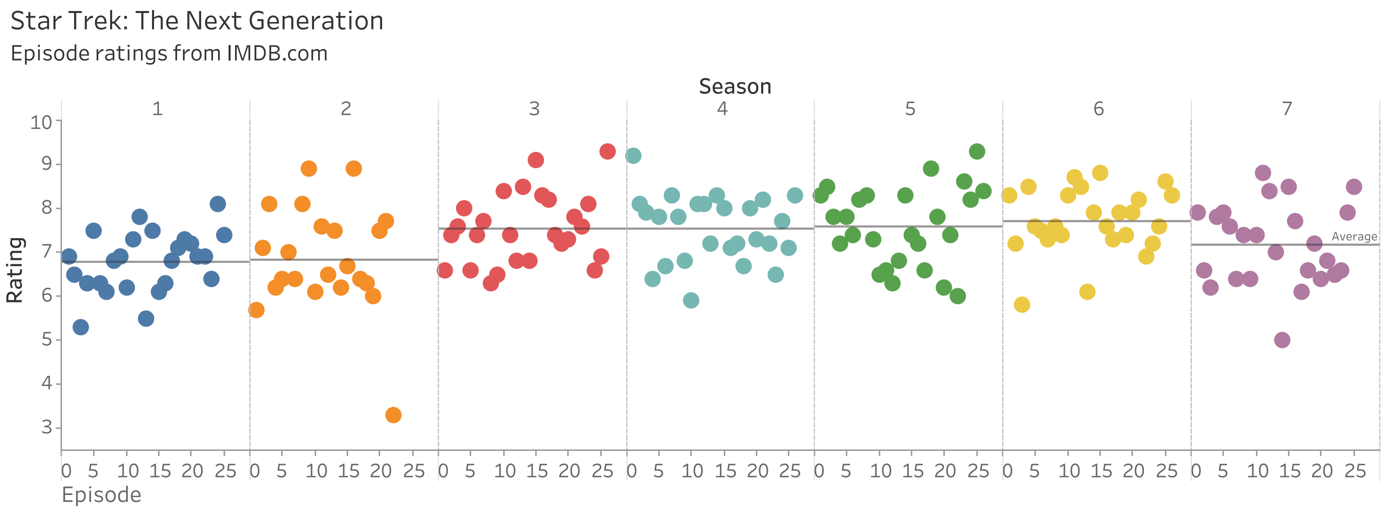

FIGURE 1.12 Every episode of *Star Trek: The Next Generation* rated.
Source: IMDB.com

TABLE 1.3 Data used in Figure 1.12.

Data	Data Type	Encoding	Note
Episode	Categorical	Position	Each episode is represented by a dot. Each dot has its own position on the canvas.
Episode Number	Ordinal	Position	The x-axis shows the number of each episode in each season.
Season	Ordinal	Color Position	Each season is represented by a different color (hue). Each season also has its own section on the chart.
IMDB rating	Ordinal	Position	The better the episode, the higher it is on the y-axis.
Average season rating	Quantitative	Position	The horizontal bar in each pane shows the average rating of the episodes in each season. There is some controversy over whether you should average ordinal ratings. We believe that the practice is so common with ratings it is acceptable.

Let's look at a few more charts to see how preattentive features have been used. Figure 1.13 is from *The Economist*. Look at each chart and see if you can work out which types of data are being graphed and how they are being encoded.

Table 1.4 shows how each data type is encoded.

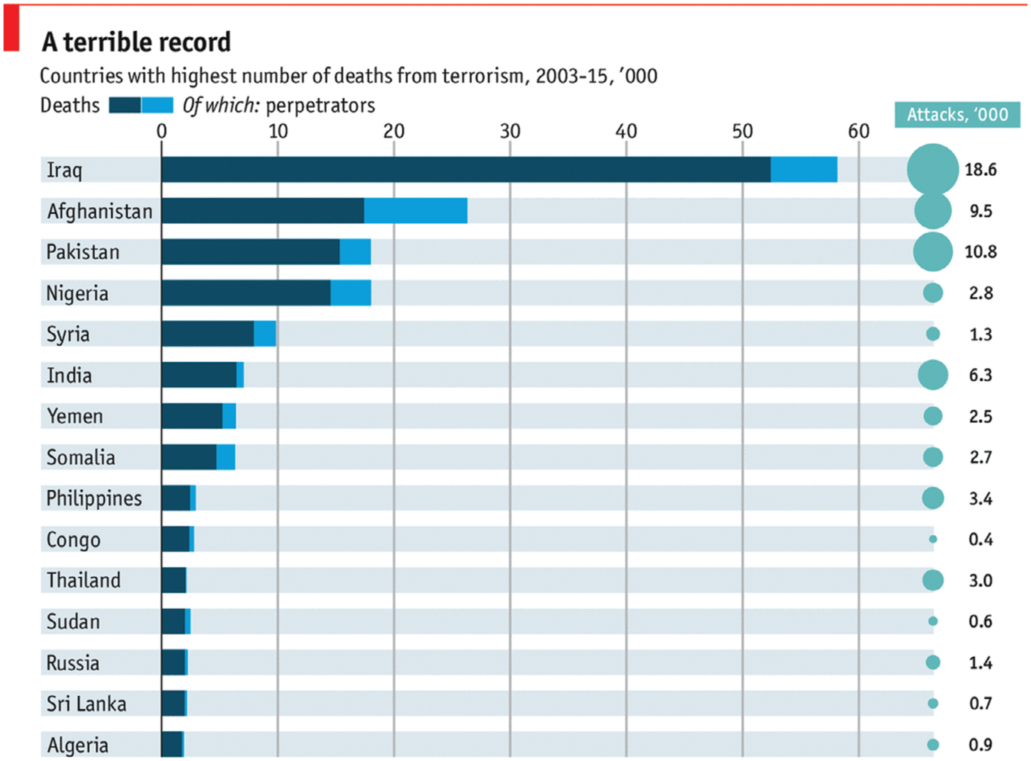

FIGURE 1.13 "A terrible record" from *The Economist*, July 2016.
Source: START, University of Maryland. *The Economist*, http://tabsoft.co/2agK3if

TABLE 1.4 Data used in the bar chart in Figure 1.13.

Data	Data Type	Encoding	Note
Country	Categorical	Position	Each country is on its own row (sorted by total deaths).
Deaths	Quantitative	Length	The length of the bar shows the number of deaths.
Death type	Categorical	Color	Dark blue shows deaths of victims, light blue shows deaths of the perpetrators.
Attacks	Quantitative	Size	Circles on the right are sized according to the number of attacks.

Let's look at another example. Figure 1.14 was part of the Makeover Monday project run by Andy Cotgreave and Andy Kriebel throughout 2016. This entry was by Dan Harrison. It takes data on malaria deaths from the World Health Organization. Table 1.5 describes the data used in the chart.

How did you do? As you progress through the book, stop and analyze some of the views in the scenarios: Think about which data types are being used and how they have been encoded.

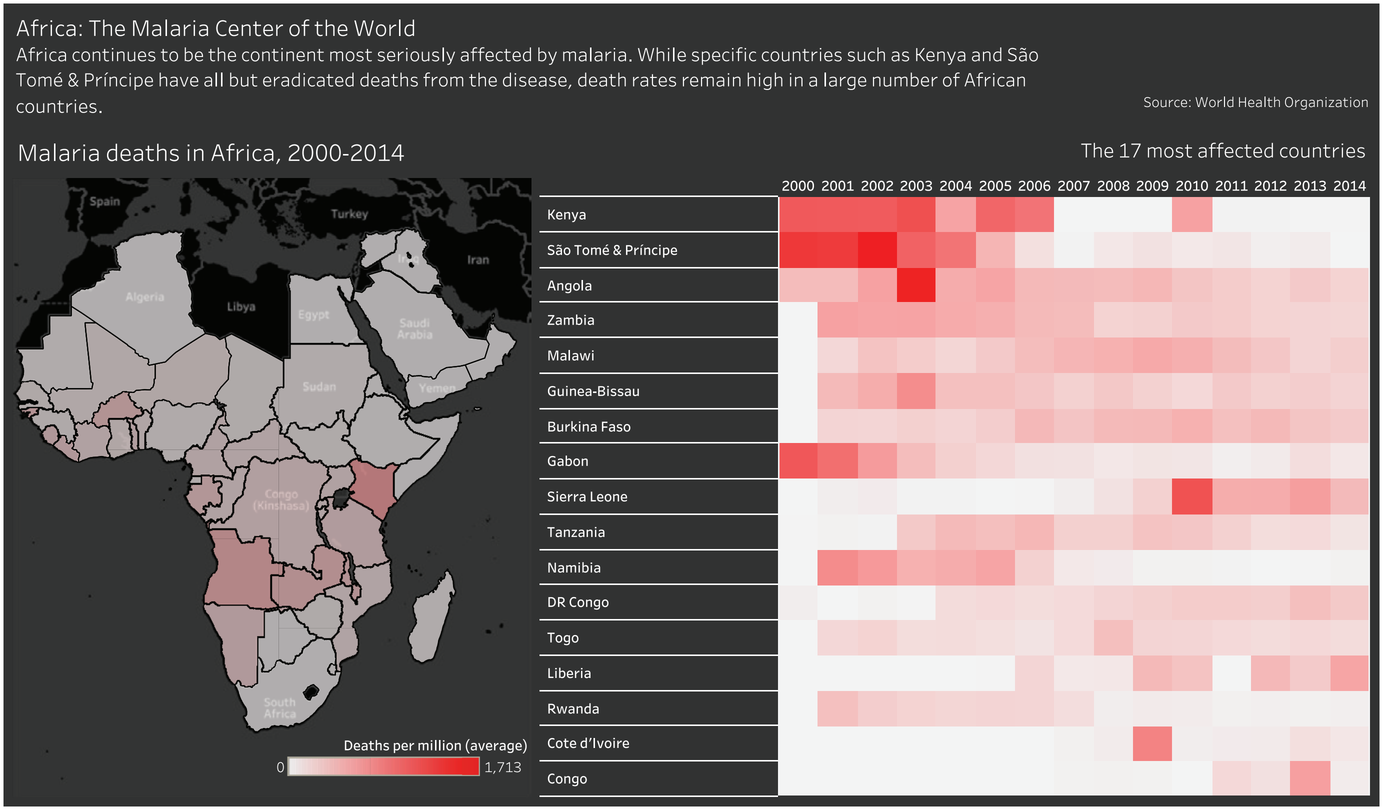

FIGURE 1.14 Deaths from malaria, 2000–2014.

Source: World Health Organization. Chart part of the Makeover Monday project

TABLE 1.5 Data used in the bar chart in Figure 1.14.

Data	Data Type	Encoding	Note
Country	Categorical	Position	The map shows the position of each country. In the highlight table, each country has its own row.
Deaths per million	Quantitative	Color	The map and table use the same color legend to show deaths per million people.
Year	Ordinal	Position	Each year is a discrete column in the table.

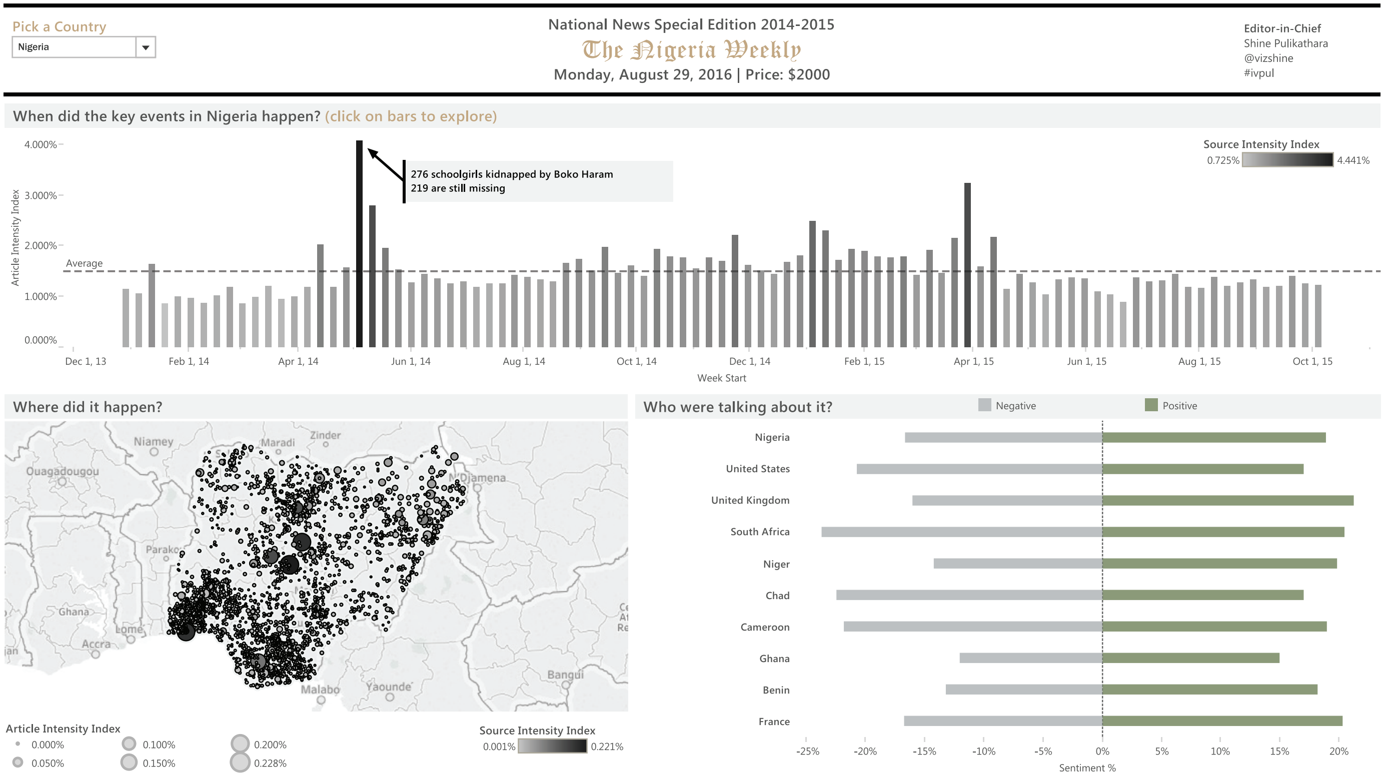

FIGURE 1.15 Winning visualization by Shine Pulikathara during the 2015 Tableau Iron Viz competition.

Source: Used with permission from Shine Pulikathara.

Color

Color is one of the most important things to understand in data visualization and frequently is misused. You should not use color just to spice up a boring visualization. In fact, many great data visualizations don't use color at all and are informative and beautiful.

In Figure 1.15, we see Shine Pulikathara's visualization that won the 2015 Tableau Iron Viz competition. Notice his simple use of color.

Color should be used purposefully. For example, color can be used to draw the attention of the reader, highlight a portion of data, or distinguish between different categories.

Use of Color

Color should be used in data visualization in three primary ways: *sequential*, *diverging*, and *categorical*.

In addition, there is often the need to *highlight* data or *alert* the reader of something important. Figure 1.16 offers an example of each of these color schemes.

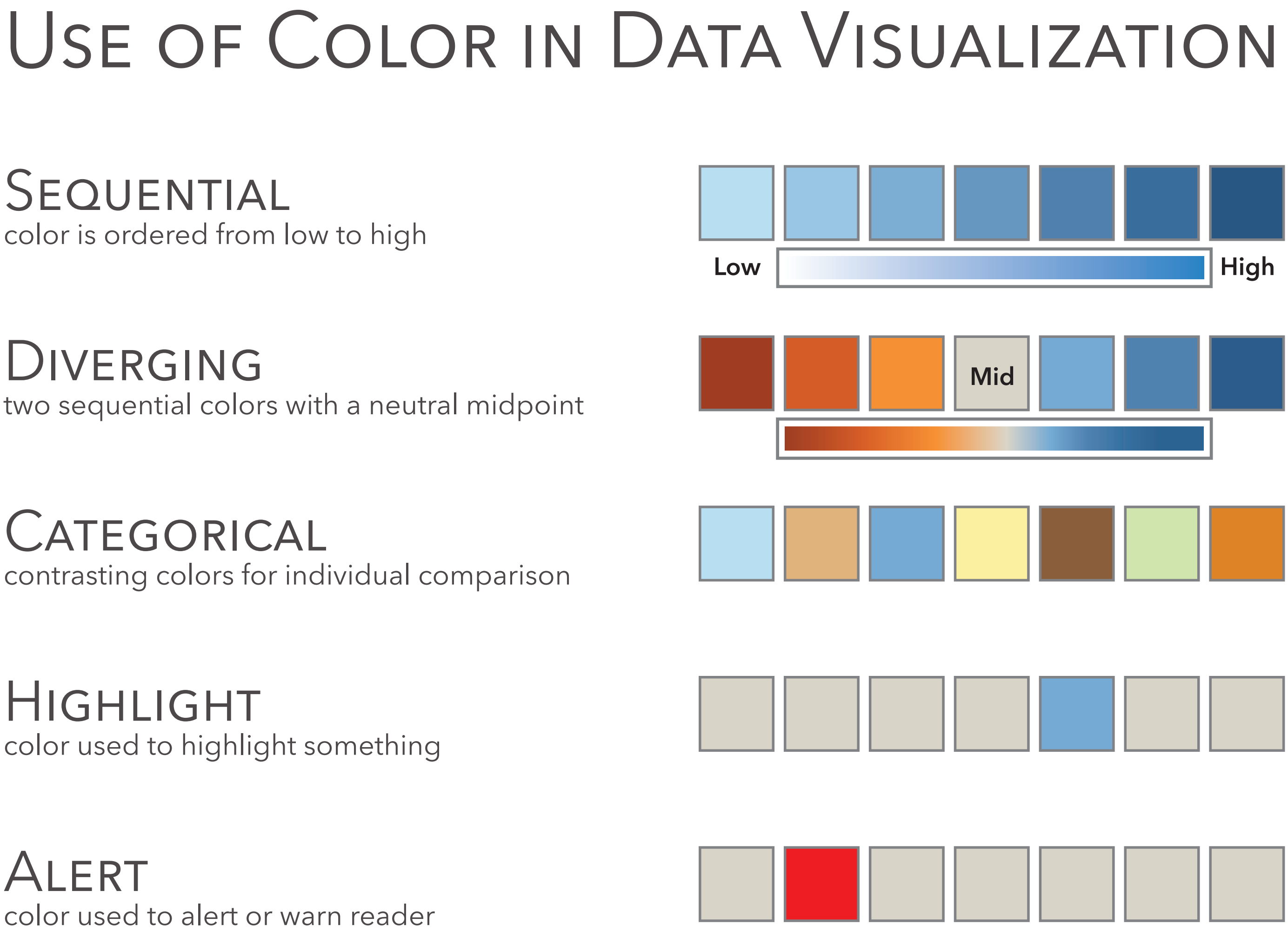

FIGURE 1.16 Use of color in data visualization.

Sequential color is the use of a single color from light to dark. An example is encoding the total amount of sales by state in blue, where the darker blue shows higher sales and a lighter blue shows lower sales. Figure 1.17 shows the unemployment rate by state using a sequential color scheme.

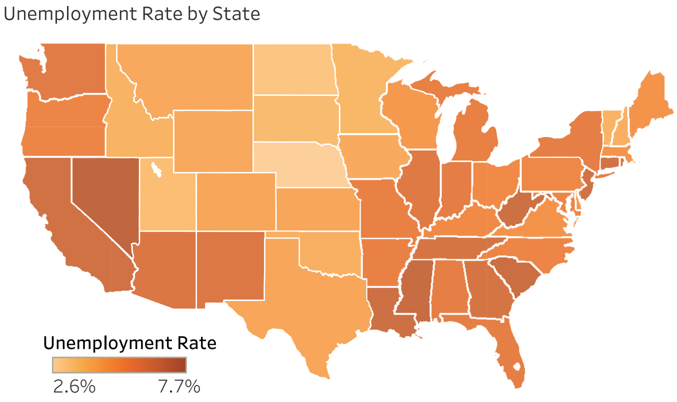

FIGURE 1.17 Unemployment rate by state using a sequential color scheme.

Diverging color is used to show a range diverging from a midpoint. This color can be used in the same manner as the sequential color scheme but can encode two different ranges of a measure (positive and negative) or a range of a measure between two categories. An example is the degree to which electorates may vote Democratic or Republican in each state, as shown in Figure 1.18.

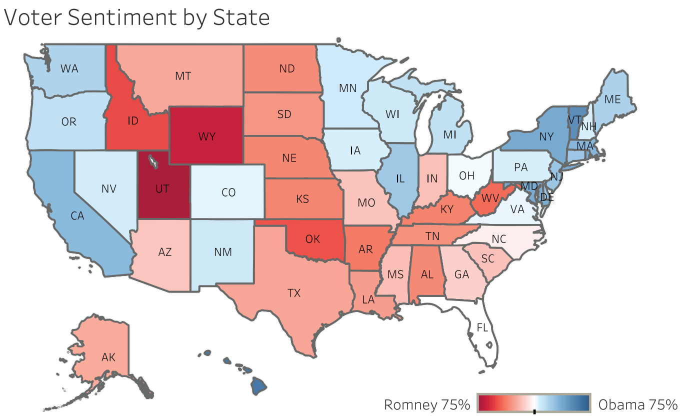

FIGURE 1.18 Degree of Democratic (blue) versus Republican (red) voter sentiment in each state.

Diverging color can also be used to show the weather, with blue showing the cooler temperatures and red showing the hotter temperatures. The midpoint can be the average, the target, or zero in cases where there are positive and negative numbers. Figure 1.19 shows an example with profit by state, where profit (positive number) is shown in blue and loss (negative number) is shown in orange.

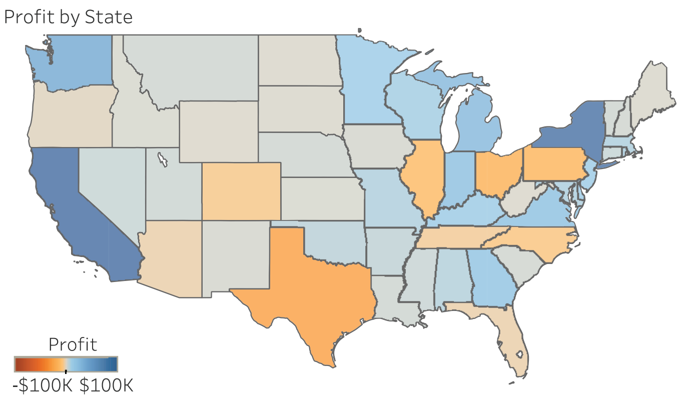

FIGURE 1.19 Profit by state using a diverging color scheme.

Categorical color uses different color hues to distinguish between different categories. For example, we can establish categories involving apparel (e.g., shoes, socks, shirts, hats, and coats) or vehicle types (e.g., cars, mini-vans, sport utility vehicles, and motorcycles). Figure 1.20 shows quantity of office supplies in three categories.

Highlight color is used when there is something that needs to stand out to the reader, but not alert or alarm them. Highlights can be used in a number of ways, as in highlighting a certain data point, text in a table, a certain line on a line chart, or a specific bar in a bar chart. Figure 1.21 shows a slopegraph with a single state highlighted in blue.

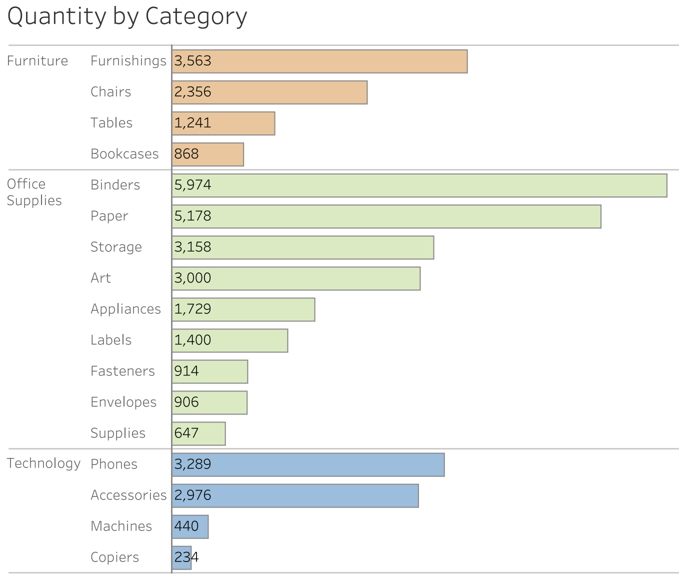

FIGURE 1.20 Quantity of office supplies in three categories using a categorical color scheme.

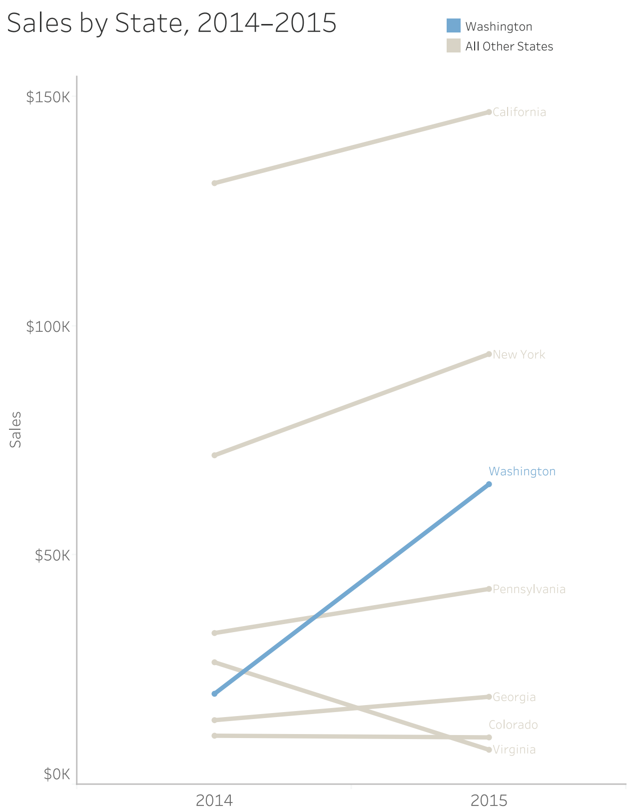

FIGURE 1.21 Slopegraph showing sales by state, 2014–2015, using a single color to highlight the state of Washington.

Alerting color is used when there is a need to draw attention to something for the reader. In this case, it's often best to use bright, alarming colors, which will quickly draw the reader's attention, as in Figure 1.22.

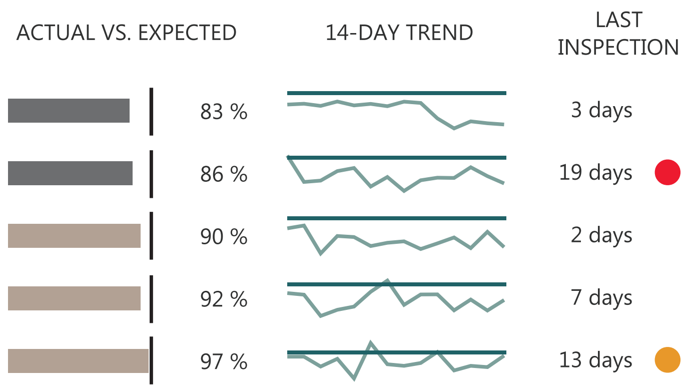

FIGURE 1.22 Red and orange indicators to alert the reader that something on the dashboard needs attention.

It is also possible to have a *categorical-sequential* color scheme. In this case, each category has a distinct hue that is darker or lighter depending on the measurement it is representing. Figure 1.23 shows an example of a four-region map using categorical colors (i.e., gray, blue, yellow, and brown) but at the same time encoding a measure in those regions using sequential color; let's assume that sales are higher in states with darker shading.

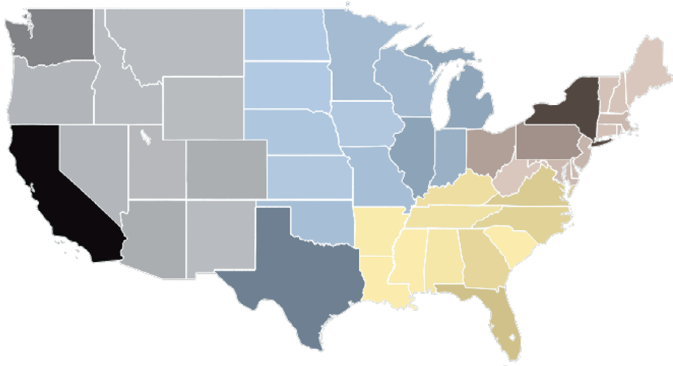

FIGURE 1.23 Sales by region using four categorical colors and the total sales shown with sequential color.

Color Vision Deficiency (Color Blindness)

Based on research (Birch 1993), approximately 8 percent of males have color vision deficiency (CVD) compared to only 0.4 percent of females. This deficiency is caused by a lack of one of three types of cones within the eye needed to see all color. The deficiency commonly is referred to as "color blindness", but that term isn't entirely accurate. People suffering from CVD can in fact see color, but they cannot distinguish colors in the same way as the rest of the population. The more accurate term is "color vision deficiency." Depending on which cone is lacking, it can be very difficult for people with CVD to distinguish between certain colors because of the way they see the color spectrum.

There are three types of CVD:

1. *Protanopia* is the lack of long-wave cones (red weak).
2. *Deuteranopia* is the lack of medium-wave cones (green weak).
3. *Tritanopia* is the lack of short-wave cones (blue). (This is very rare, affecting less than 0.5 percent of the population.)

CVD is mostly hereditary, and, as you can see from the numbers, it primarily afflicts men. Eight percent of men may seem like a small number, but consider that in a group of nine men, there is more than a 50 percent chance that one of them has CVD. In a group of 25 men, there is an 88 percent chance that one of them has CVD. The rates also increase among Caucasian men, reaching as high as 11 percent. In larger companies or when a data visualization is presented to the general public, designers must understand CVD and design with it in mind.

The primary problem among people with CVD is with the colors red and green. This is why it is best to avoid using red and green together and, in general, to avoid the commonly used traffic light colors. We discuss this issue further in Chapter 33 and offer some solutions for using red and green together.

Seeing the Problem for Yourself

Let's look at some examples of how poor choice of color can create confusion for people with CVD.

In Figure 1.24, the chart on the left uses the traditional traffic light colors red, yellow, and green. The example on the right is a protanopia simulation for CVD.

One common solution among data visualization practitioners is to use blue and orange. Using blue instead of green for *good* and orange instead of red for *bad* works well because almost everyone (with very rare exceptions) can distinguish blue and orange from each other. This blue-orange palette is often referred to as being "color-blind friendly."

Using Figure 1.25, compare the blue/orange color scheme and a protanopia simulation of CVD again.

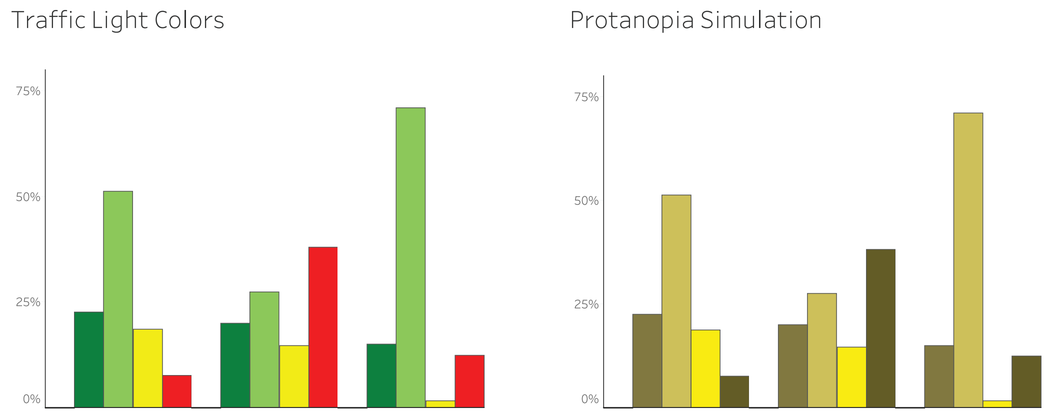

FIGURE 1.24 Bar chart using the traffic light colors and a protanopia simulation. Notice the red and green bars in the panel on the right are very difficult to differentiate from one another for a person with protanopia.

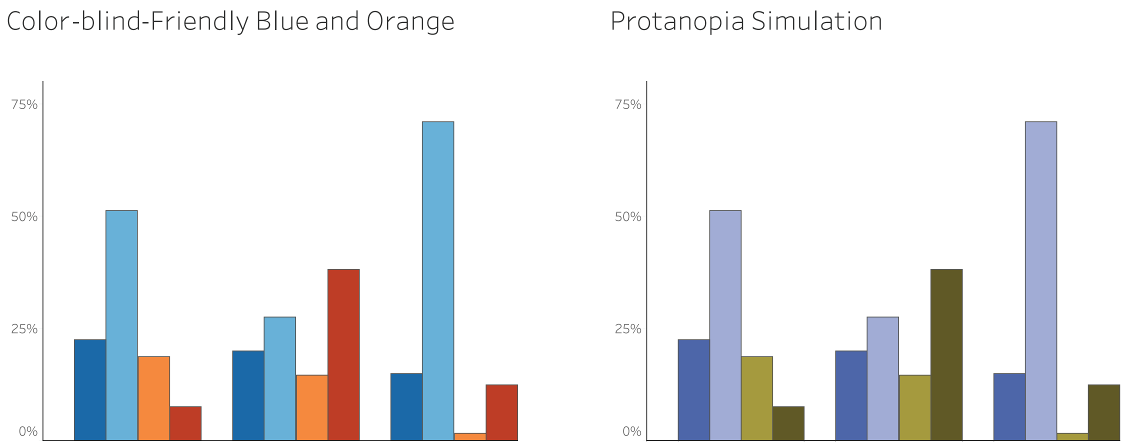

FIGURE 1.25 Bar chart using a color-blind-friendly blue and orange palette and a protanopia simulation.

The Problem Is Broader Than Just Red and Green

The use of red and green is discussed frequently in the field of data visualization, probably because the traffic light color palette is prevalent in many software programs and is commonly used in business today. It is common in Western culture to associate red with bad and green with good. However, it is important to understand that the problem in differentiating color for someone with CVD is much more complex than just red and green. Since red, green, and orange all appear to be brown for someone with strong CVD, it would be more accurate to say "Don't use red, green, brown, and orange together."

Figure 1.26 shows a scatterplot using brown, orange, and green together for three categories. When applying protanopia simulation, the dots in the scatterplot appear to be a very similar color.

One color combination that is frequently overlooked is blue and purple together. In a RGB (red-green-blue) color model, purple is achieved by using blue and red together. If someone with CVD has issues with red, then he or she may also have issues with purple, which would appear to look like blue. Other color combinations can be problematic as well. For example, people may have difficulty with pink or red used with gray or gray used together with brown.

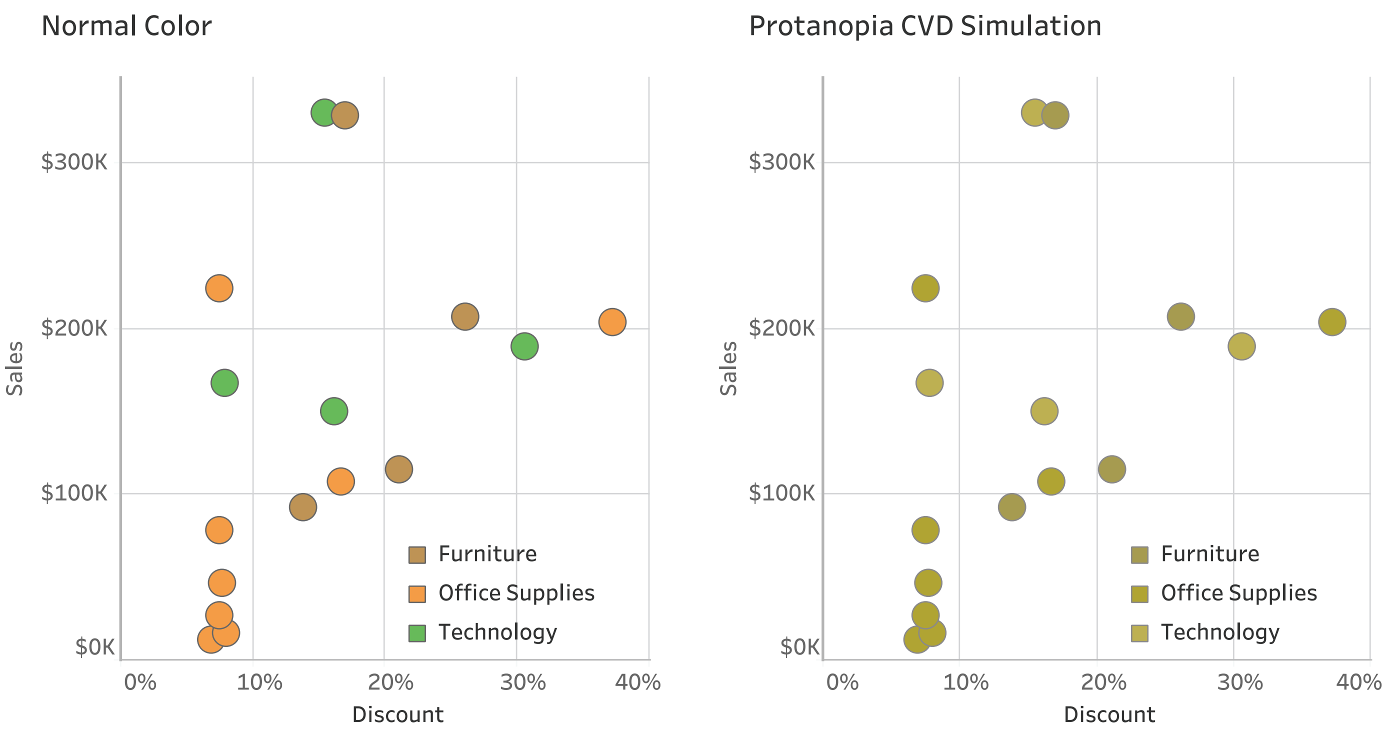

FIGURE 1.26 Scatterplot simulating color vision deficiency for someone with protanopia.

Figure 1.27 shows another scatterplot, this time using blue, purple, magenta, and gray. When applying deuteranopia simulation, the dots in the scatterplot appear to be a very similar color of gray.

It's important to understand these issues when designing visualizations. If color is used to encode data and it's necessary for readers to distinguish among colors to understand the visualization, then consider using color-blind-friendly palettes. Here are a few resources that you can use to simulate the various types of CVD for your own visualizations.

Adobe Illustrator CC. This program offers a built-in CVD simulation in the View menu under Proof Setup.

Chromatic Vision Simulator (free). Kazunori Asada's superb website allows users to upload images and simulate how they would appear to people with different form of CVD. See http://asada.tukusi.ne.jp/webCVS/

NoCoffee vision simulator (free). This free simulator for the Chrome browser allows users to simulate websites and images directly from the browser.

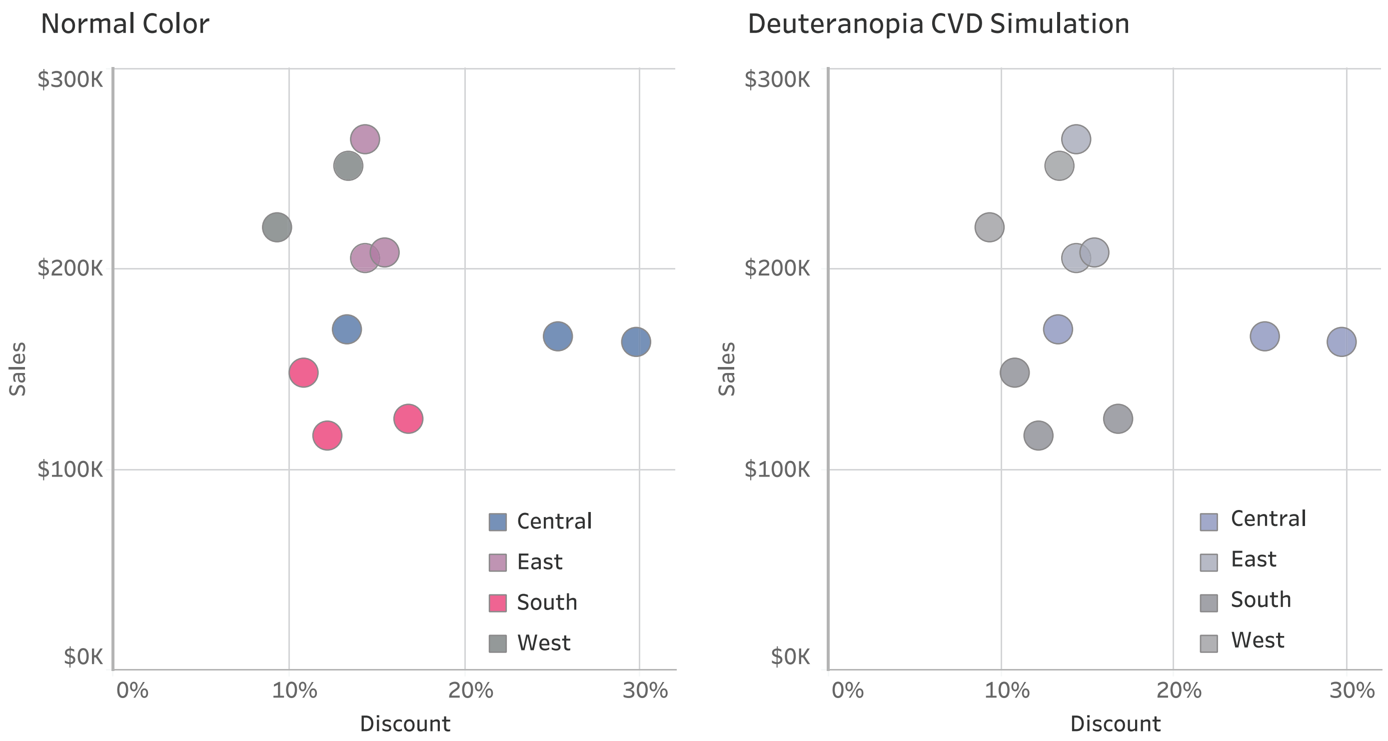

FIGURE 1.27 Scatterplot simulating color vision deficiency for someone with deuteranopia.

Common Chart Types

In this book, you will see many different types of charts. We explain in the scenarios why many of the charts were chosen to fulfill a particular task. In this section, we briefly outline the most common chart types. This list is intentionally short. Even if you use only the charts listed here, you would be able to cover the majority of needs when visualizing your data. More advanced chart types seen throughout the book are built from the same building blocks as these. For example, sparklines, which are shown in Chapters 6, 8, and 9, are a kind of line chart. Bullet charts, used in Chapter 17, are bar charts with reference lines and shading built in. Finally, waterfall charts, shown in Chapter 24, are bar charts where the bars don't have a common baseline.

Bar Chart

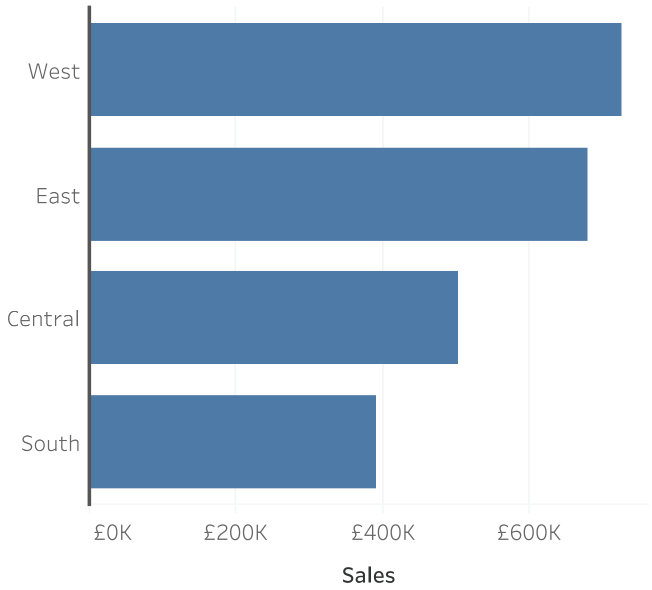

FIGURE 1.28 Bar chart.

A bar chart (see Figure 1.28) uses length to represent a measure. Human beings are extremely good at seeing even small differences in length from a common baseline. Bars are widely used in data visualization because they are often the most effective way to compare categories. Bars can be oriented horizontally or vertically. Sorting them can be very helpful because the most common task when bar charts are used is to spot the biggest/smallest items.

Time-Series Line Chart

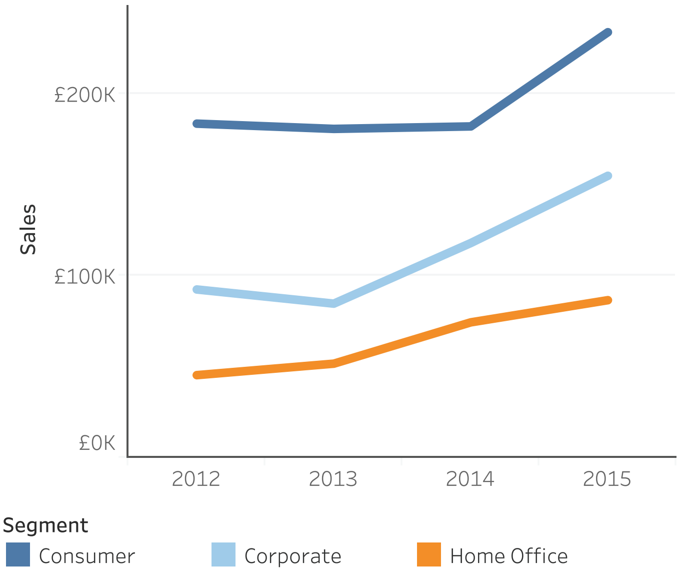

FIGURE 1.29 Time-series line chart.

Line charts (see Figure 1.29) usually show change over time. Time is represented by position on the horizontal x-axis. The measures are shown on the vertical y-axis. The height and slopes of the line let us see trends.