CENGAGE

Camm Cochran Fry Ohlmann

Data Visualization

Exploring and Explaining with Data

Data Visualization

Exploring and Explaining with Data

Jeffrey D. Camm
Wake Forest University

James J. Cochran
University of Alabama

Michael J. Fry
University of Cincinnati

Jeffrey W. Ohlmann
University of Iowa

CENGAGE

Australia • Brazil • Canada • Mexico • Singapore • United Kingdom • United States

Data Visualization: Exploring and Explaining with Data,
First Edition
Jeffrey D. Camm, James J. Cochran, Michael J. Fry, Jeffrey W. Ohlmann

SVP, Higher Education & Skills Product: Erin Joyner

VP, Higher Education & Skills Product: Michael Schenk

Product Director: Joe Sabatino

Senior Product Manager: Aaron Arnsparger

Senior Learning Designer: Brandon Foltz

Senior Content Manager: Conor Allen

Digital Delivery Lead: Mark Hopkinson

Marketing Director: Danae April

Executive Marketing Manager: Nate Anderson

IP Analyst: Ashley Maynard

IP Project Manager: Kelli Besse

Production Service: MPS Limited

Designer: Chris Doughman

Cover Image Source: iStockPhoto.com/mpilecky

Library of Congress Control Number: 2021930729

ISBN: 978-0-357-63134-8

Cengage
200 Pier 4 Boulevard
Boston, MA 02210
USA

Printed in the United States of America
Print Number: 01 Print Year: 2021

Brief Contents

Contents

About the Authors

Jeffrey D. Camm is Inmar Presidential Chair and Senior Associate Dean of Business Analytics in the School of Business at Wake Forest University. Born in Cincinnati, Ohio, he holds a B.S. from Xavier University (Ohio) and a Ph.D. from Clemson University. Prior to joining the faculty at Wake Forest, he was on the faculty of the University of Cincinnati. He has also been a visiting scholar at Stanford University and a visiting professor of business administration at the Tuck School of Business at Dartmouth College.

Dr. Camm has published more than 45 papers in the general area of optimization applied to problems in operations management and marketing. He has published his research in *Science*, *Management Science*, *Operations Research*, *INFORMS Journal on Applied Analytics*, and other professional journals. Dr. Camm was named the Dornoff Fellow of Teaching Excellence at the University of Cincinnati, and he was the 2006 recipient of the INFORMS Prize for the Teaching of Operations Research Practice. A firm believer in practicing what he preaches, he has served as an operations research consultant to numerous companies and government agencies. From 2005 to 2010 he served as editor-in-chief of the *INFORMS Journal on Applied Analytics* (formerly *Interfaces*). In 2016, Professor Camm received the George E. Kimball Medal for service to the operations research profession, and in 2017 he was named an INFORMS Fellow.

James J. Cochran is Associate Dean for Research, Professor of Applied Statistics, and the Rogers-Spivey Faculty Fellow at The University of Alabama. Born in Dayton, Ohio, he earned his B.S., M.S., and M.B.A. from Wright State University and his Ph.D. from the University of Cincinnati. He has been at The University of Alabama since 2014 and has been a visiting scholar at Stanford University, Universidad de Talca, the University of South Africa, and Pole Universitaire Leonard de Vinci.

Dr. Cochran has published more than 50 papers in the development and application of operations research and statistical methods. He has published in several journals, including *Management Science*, *The American Statistician*, *Communications in Statistics—Theory and Methods*, *Annals of Operations Research*, *European Journal of Operational Research*, *Journal of Combinatorial Optimization*, *INFORMS Journal on Applied Analytics*, and *Statistics and Probability Letters*. He received the 2008 INFORMS Prize for the Teaching of Operations Research Practice, 2010 Mu Sigma Rho Statistical Education Award, and 2016 Waller Distinguished Teaching Career Award from the American Statistical Association. Dr. Cochran was elected to the International Statistics Institute in 2005, named a Fellow of the American Statistical Association in 2011, and named a Fellow of INFORMS in 2017. He also received the Founders Award in 2014 and the Karl E. Peace Award in 2015 from the American Statistical Association, and he received the INFORMS President's Award in 2019.

A strong advocate for effective operations research and statistics education as a means of improving the quality of applications to real problems, Dr. Cochran has chaired teaching effectiveness workshops around the globe. He has served as an operations research consultant to numerous companies and not-for-profit organizations. He served as editor-in-chief of *INFORMS Transactions on Education* and is on the editorial board of *INFORMS Journal on Applied Analytics*, *International Transactions in Operational Research*, and *Significance*.

Michael J. Fry is Professor of Operations, Business Analytics, and Information Systems (OBAIS) and Academic Director of the Center for Business Analytics in the Carl H. Lindner College of Business at the University of Cincinnati. Born in Killeen, Texas, he earned a B.S. from Texas A&M University and M.S.E. and Ph.D. degrees from the University of Michigan. He has been at the University of Cincinnati since 2002, where he served as Department Head from 2014 to 2018 and has been named a Lindner Research Fellow. He has also been a visiting professor at Cornell University and at the University of British Columbia.

Professor Fry has published more than 25 research papers in journals such as *Operations Research*, *Manufacturing and Service Operations Management*, *Transportation Science*, *Naval Research Logistics*, *IIE Transactions*, *Critical Care Medicine*, and *Interfaces*. He serves on editorial boards for journals such as *Production and Operations Management*, *INFORMS Journal on Applied Analytics* (formerly *Interfaces*), and *Journal of Quantitative Analysis in Sports*. His research interests are in applying analytics to the areas of supply chain management, sports, and public-policy operations. He has worked with many different organizations for his research, including Dell, Inc., Starbucks Coffee Company, Great American Insurance Group, the Cincinnati Fire Department, the State of Ohio Election Commission, the Cincinnati Bengals, and the Cincinnati Zoo and Botanical Gardens. In 2008, he was named a finalist for the Daniel H. Wagner Prize for Excellence in Operations Research Practice, and he has been recognized for both his research and teaching excellence at the University of Cincinnati. In 2019, he led the team that was awarded the INFORMS UPS George D. Smith Prize on behalf of the OBAIS Department at the University of Cincinnati.

Jeffrey W. Ohlmann is Associate Professor of Business Analytics and Huneke Research Fellow in the Tippie College of Business at the University of Iowa. Born in Valentine, Nebraska, he earned a B.S. from the University of Nebraska and M.S. and Ph.D. degrees from the University of Michigan. He has been at the University of Iowa since 2003.

Professor Ohlmann's research on the modeling and solution of decision-making problems has produced more than two dozen research papers in journals such as *Operations Research*, *Mathematics of Operations Research*, *INFORMS Journal on Computing*, *Transportation Science*, and *European Journal of Operational Research*. He has collaborated with organizations such as Transfreight, LeanCor, Cargill, the Hamilton County Board of Elections, and three National Football League franchises. Because of the relevance of his work to industry, he was bestowed the George B. Dantzig Dissertation Award and was recognized as a finalist for the Daniel H. Wagner Prize for Excellence in Operations Research Practice.

Preface

Data Visualization: Exploring and Explaining with Data is designed to introduce best practices in data visualization to undergraduate and graduate students. This is one of the first books on data visualization designed for college courses. The book contains material on effective design, choice of chart type, effective use of color, how to explore data visually, how to build data dashboards, and how to explain concepts and results visually in a compelling way with data. In an increasingly data-driven economy, these concepts are becoming more important for analysts, natural scientists, social scientists, engineers, medical professionals, business professionals, and virtually everyone who needs to interact with data. Indeed, the skills developed in this book will be helpful to all who want to influence with data or be accurately informed by data.

The book is designed for a semester-long course at either the undergraduate or graduate level. The examples used in this book are drawn from a variety of functional areas in the business world including accounting, finance, operations, and human resources as well as from sports, politics, science, medicine, and economics. The intention is that this book will be relevant to students at either the undergraduate or graduate level in a business school as well as to students studying in other academic areas.

Data Visualization: Exploring and Explaining with Data is written in a style that does not require advanced knowledge of mathematics or statistics. The first five chapters cover foundational issues important to constructing good charts. Chapter 1 introduces data visualization and how it fits into the broader area of analytics. A brief history of data visualization is provided as well as a discussion of the different types of data and examples of a variety of charts. Chapter 2 provides guidance on selecting an appropriate type of chart based on the goals of the visualization and the type of data to be visualized. Best practices in chart design, including discussions of preattentive attributes, Gestalt principles, and the data-ink ratio, are covered in Chapter 3. Chapter 4 discusses the attributes of color, how to use color effectively, and some common mistakes in the use of color in data visualization. Chapter 5 covers the important topic of visualizing and describing variability that occurs in observed values. Chapter 5 introduces the visualization of frequency distributions for categorical and quantitative variables, measures of location and variability, and confidence intervals and prediction intervals.

Chapters 6 and 7 cover how to explore and explain with data visualization in detail with examples. Chapter 6 discusses the use of visualization in exploratory data analysis. The exploration of individual variables as well as the relationship between pairs of variables is considered. The organization of data to facilitate exploration is discussed as well as the effect of missing data. The special considerations of visualizing time series data and geospatial data are also presented. Chapter 7 provides important coverage of how to explain and influence with data visualization, including knowing your message, understanding the needs of your audience, and using preattentive attributes to better convey your message. Chapter 8 is a discussion of how to design and construct data dashboards, collections of data visualizations used for decision making. Finally, Chapter 9 covers the responsible use of data visualization to avoid confusing or misleading your audience. Chapter 9 addresses the importance of understanding your data in order to best convey insights accurately and also discusses how design choices in a data visualization affect the insights conveyed to the audience.

This textbook can be used by students who have previously taken a basic statistics course as well as by students who have not had a prior course in statistics. The two most technical chapters, Chapters 5 (Visualizing Variability) and 6 (Exploring Data Visually), do not assume a previous course in statistics. All technical concepts are gently introduced. For students who have had a previous statistics class, the statistical coverage in these chapters provides a good review within a treatment where the focus is on visualization. The book offers complete coverage for a full course in data visualization, but it can also support a basic statistics or analytics course. The following table gives our recommendations for chapters to use to support a variety of courses.

	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6	Chapter 7	Chapter 8	Chapter 9
	Intro	Chart Type	Design	Color	Variability	Exploring	Explaining	Dashboards	Truth
Full Data Visualization Course	●	●	●	●	●	●	●	●	●
Data Visualization Course Focused on Presentation	●	●	●	●			●		●
Part of a Basic Statistics Course		●	●	●	●	●			●
Part of an Analytics Course		●	●	●	●	●			●

Features and Pedagogy

The style and format of this textbook are similar to our other textbooks. Some of the specific features that we use in this textbook are listed here.

- Data Visualization Makeover: With the exception of Chapter 1, each chapter contains a Data Visualization Makeover. Each of these vignettes presents a real visualization that can be improved using the principles discussed in the chapter. We present the original data visualization and then discuss how it can be improved. The examples are drawn from many different organizations in a variety of areas including government, retail, sports, science, politics, and entertainment.
- Learning Objectives: Each chapter has a list of learning objectives of that chapter. The list provides details of what students should be able to do and understand once they have completed the chapter.
- Software: Because of its widespread use and ease of availability, we have chosen Microsoft Excel as the software to illustrate the best practices and principles contained herein. Excel has been thoroughly integrated throughout this textbook. Whenever we introduce a new type of chart or table, we provide detailed step-by-step instructions for how to create the chart or table in Excel. Step-by-step instructions for creating many of the charts and tables from the textbook using Tableau and Power BI are also available in MindTap.
- Notes and Comments: At the end of many sections, we provide Notes and Comments to give the student additional insights about the material presented in that section. Additionally, margin notes are used throughout the textbook to provide insights and tips related to the specific material being discussed.
- End-of-Chapter Problems: Each chapter contains at least 15 problems to help the student master the material presented in that chapter. The problems are separated into Conceptual and Applications problems. Conceptual problems test the student's understanding of concepts presented in the chapter. Applications problems are hands-on and require the student to construct or edit charts or tables.
- DATAfiles and CHARTfiles: All data sets used as examples and in end-of-chapter problems are Excel files designated as DATAfiles and are available for download by the student. The names of the DATAfiles are called out in margin notes throughout the textbook. Similarly, some Excel files with completed charts are available for download and are designated as CHARTfiles.

MindTap

MindTap is a customizable digital course solution that includes an interactive eBook, auto-graded exercises and problems from the textbook with solutions feedback, interactive visualization applets with quizzes, chapter overview and problem walk-through videos, and more! MindTap also includes step-by-step instructions for creating charts and tables from the textbook in Tableau and Power BI. Contact your Cengage account executive for more information about MindTap.

Instructor and Student Resources

Additional instructor and student resources for this product are available online. Instructor assets include an Instructor's Manual, Educator's Guide, PowerPoint® slides, a Solutions and Answers Guide, and a test bank powered by Cognero®. Student assets include data sets. Sign up or sign in at **www.cengage.com** to search for and access this product and its online resources.

ACKNOWLEDGMENTS

We would like to acknowledge the work of reviewers who have provided comments and suggestions for improvement of this first edition of this text. Thanks to:

Xiaohui Chang
Oregon State University

Wei Chen
York College of Pennsylvania

Anjee Gorkhali
Susquehanna University

Rita Kumar
Cal Poly Pomona

Barin Nag
Towson University

Andy Olstad
Oregon State University

Vivek Patil
Gonzaga University

Nolan Taylor
Indiana University

We are also indebted to the entire team at Cengage who worked on this title: Senior Product Manager, Aaron Arnsparger; Senior Content Manager, Conor Allen; Senior Learning Designer, Brandon Foltz; Digital Delivery Lead, Mark Hopkinson; Associate Subject-Matter Expert, Nancy Marchant; Content Program Manager, Jessica Galloway; Content Quality Assurance Engineer, Douglas Marks; and our Senior Project Manager at MPS Limited, Anubhav Kaushal, for their editorial counsel and support during the preparation of this text.

The following Technical Content Developers worked on the MindTap content for this text: Anthony Bacon, Philip Bozarth, Sam Gallagher, Anna Geyer, Matthew Holmes, and Christopher Kurt. Our thanks to them as well.

Jeffrey D. Camm
James J. Cochran
Michael J. Fry
Jeffrey W. Ohlmann

Chapter 1

Introduction

CONTENTS

LEARNING OBJECTIVES

After completing this chapter, you will be able to

LO 1 Define analytics and describe the different types of analytics

LO 2 Describe the different types of data and give an example of each

LO 3 Describe various examples of data visualization used in practice

LO 4 Identify the various charts defined in this chapter

You need a ride to a concert, so you select the Uber app on your phone. You enter the location of the concert. Your phone automatically knows your location and the app presents several options with prices. You select an option and confirm with your driver. You receive the driver's name, license plate number, make and model of vehicle, and a photograph of the driver and the car. A map showing the location of the driver and the time remaining until arrival is updated in real time.

Without even thinking about it, we continually use data to make decisions in our lives. How the data are displayed to us has a direct impact on how much effort we must expend to utilize the data. In the case of Uber, we enter data (our destination) and we are presented with data (prices) that allow us to make an informed decision. We see the result of our decision with an indication of the driver's name, make and model of vehicle, and license plate number that makes us feel more secure. Rather than simply displaying the time until arrival, seeing the progress of the car on a map gives us some indication of the driver's route. Watching the driver's progress on the app removes some uncertainty and to some extent can divert our attention from how long we have been waiting. What data are presented and how they are presented has an impact on our ability to understand the situation and make more-informed decisions.

A weather map, an airplane seating chart, the dashboard of your car, a chart of the performance of the Dow Jones Industrial Average, your fitness tracker—all of these involve the visual display of data. **Data visualization** is the graphical representation of data and information using displays such as charts, graphs, and maps. Our ability to process information visually is strong. For example, numerical data that have been displayed in a chart, graph, or map allow us to more easily see relationships between variables in our data set. Trends, patterns, and the distributions of data are more easily comprehended when data are displayed visually.

This book is about how to effectively display data to both discover and describe the information it contains data. We provide best practices in the design of visual displays of data, the effective use of color, and chart type selection. The goal of this book is to instruct you how to create effective data visualizations. Through the use of examples (using real data when possible), this book presents visualization principles and guidelines for gaining insight from data and conveying an impactful message to the audience.

With the increased use of analytics in business, industry, science, engineering, and government, data visualization has increased dramatically in importance. We begin with a discussion of analytics and data visualization's role in this rapidly growing field.

1-1 Analytics

Analytics is the scientific process of transforming data into insights for making better decisions.[1] Three developments have spurred the explosive growth in the use of analytics for improving decision making in all facets of our lives, including business, sports, science, medicine, and government:

- Incredible amounts of data are produced by technological advances such as point-of-sale scanner technology; e-commerce and social networks; sensors on all kinds of mechanical devices such as aircraft engines, automobiles, thermometers, and farm machinery enabled by the so-called Internet of Things; and personal electronic devices such as cell phones. Businesses naturally want to use these data to improve the efficiency and profitability of their operations, better understand their customers, and price their products more effectively and competitively. Scientists and engineers use these data to invent new products, improve existing products, and make new basic discoveries about nature and human behavior.

[1]We adopt the definition of analytics developed by the Institute for Operations Research and the Management Sciences (INFORMS).

- Ongoing research has resulted in numerous methodological developments, including advances in computational approaches to effectively handle and explore massive amounts of data as well as faster algorithms for data visualization, machine learning, optimization, and simulation.
- The explosion in computing power and storage capability through better computing hardware, parallel computing, and cloud computing (the remote use of hardware and software over the internet) enable us to solve larger decision problems more quickly and more accurately than ever before.

In summary, the availability of massive amounts of data, improvements in analytical methods, and substantial increases in computing power and storage have enabled the explosive growth in analytics, data science, and artificial intelligence.

Analytics can involve techniques as simple as reports or as complex as large-scale optimizations and simulations. Analytics is generally grouped into three broad categories of methods: descriptive, predictive, and prescriptive analytics.

Descriptive analytics is the set of analytical tools that describe what has happened. This includes techniques such as data queries (requests for information with certain characteristics from a database), reports, descriptive or summary statistics, and data visualization. Descriptive data mining techniques such as cluster analysis (grouping data points with similar characteristics) also fall into this category. In general, these techniques summarize existing data or the output from predictive or prescriptive analyses.

Predictive analytics consists of techniques that use mathematical models constructed from past data to predict future events or better understand the relationships between variables. Techniques in this category include regression analysis, time series forecasting, computer simulation, and predictive data mining. As an example of a predictive model, past weather data are used to build mathematical models that forecast future weather. Likewise, past sales data can be used to predict future sales for seasonal products such as snowblowers, winter coats, and bathing suits.

Prescriptive analytics are mathematical or logical models that suggest a decision or course of action. This category includes mathematical optimization models, decision analysis, and heuristic or rule-based systems. For example, solutions to supply network optimization models provide insights into the quantities of a company's various products that should be manufactured at each plant, how much should be shipped to each of the company's distribution centers, and which distribution center should serve each customer to minimize cost and meet service constraints.

Data visualization is mission-critical to the success of all three types of analytics. We discuss this in more detail with examples in the next section.

1-2 Why Visualize Data?

We create data visualizations for two reasons: exploring data and communicating/explaining a message. Let us discuss these uses of data visualization in more detail, examine the differences in the two uses, and consider how they relate to the types of analytics previously described.

Data Visualization for Exploration

Data visualization is a powerful tool for *exploring* data to more easily identify patterns, recognize anomalies or irregularities in the data, and better understand the relationships between variables. Our ability to spot these types of characteristics of data is much stronger and quicker when we look at a visual display of the data rather than a simple listing.

In chapter 2, we introduce a variety of different chart types and how to construct charts in Excel.

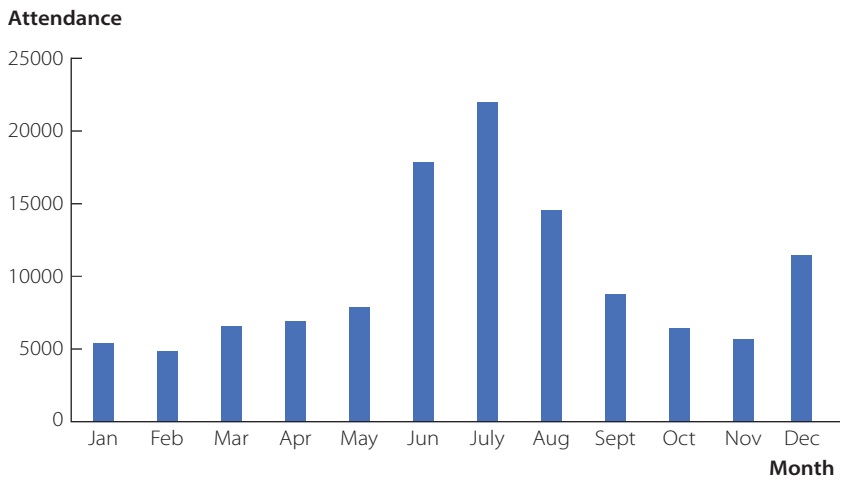
As an example of data visualization for exploration, let us consider the zoo attendance data shown in Table 1.1 and Figure 1.1. These data on monthly attendance to a zoo can be found in the file *Zoo*. Comparing Table 1.1 and Figure 1.1, observe that the pattern in the data is more detectable in the column chart of Figure 1.1 than in a table of numbers. A **column chart** shows numerical data by the height of the column for a variety of categories or time periods. In the case of Figure 1.1, the time periods are the different months of the year.

DATA file
Zoo

TABLE 1.1 Zoo Attendance Data

Month	Jan	Feb	Mar	Apr	May	Jun
Attendance	5422	4878	6586	6943	7876	17843
Month	**July**	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
Attendance	21967	14542	8751	6454	5677	11422

FIGURE 1.1 A Column Chart of Zoo Attendance by Month



Our intuition and experience tells us that we would expect zoo attendance to be highest in the summer months when many school-aged children are out of school for summer break. Figure 1.1 confirms this, as the attendance at the zoo is highest in the summer months of June, July, and August. Furthermore, we see that attendance increases gradually each month from February through May as the average temperature increases, and attendance gradually decreases each month from September through November as the average temperature decreases. But why does the zoo attendance in December and January not follow these patterns? It turns out that the zoo has an event known as the "Festival of Lights" that runs from the end of November through early January. Children are out of school during the last half of December and early January for the holiday season, and this leads to increased attendance in the evenings at the zoo despite the colder winter temperatures.

Visual data exploration is an important part of descriptive analytics. Data visualization can also be used directly to monitor key performance metrics, that is, measure how an organization is performing relative to its goals. A **data dashboard** is a data visualization tool that gives multiple outputs and may update in real time. Just as the dashboard in your car measures the speed, engine temperature, and other important performance data as you drive, corporate data dashboards measure performance metrics such as sales, inventory levels, and service levels relative to the goals set by the company. These data dashboards alert management when performances deviate from goals so that corrective actions can be taken.

Data dashboards are discussed in more detail in Chapter 8.

Visual data exploration is also critical for ensuring that model assumptions hold in predictive and prescriptive analytics. Understanding the data before using that data in modeling builds trust and can be important in determining and explaining which type of model is appropriate.

As an example of the importance of exploring data visually before modeling, we consider two data sets provided by statistician Francis Anscombe.[2] Table 1.2 contains these two data sets, each of which contains 11 *X*-*Y* pairs of data. Notice in Table 1.2 that both data sets have the same average values for *X* and *Y*, and both sets of *X* and *Y* also have the same standard deviations. Based on these commonly used summary statistics, these two data sets are indistinguishable.

Figure 1.2 shows the two data sets visually as scatter charts. A **scatter chart** is a graphical presentation of the relationship between two quantitative variables. One variable is shown on the horizontal axis and the other is shown on the vertical axis. Scatter charts are used to better understand the relationship between the two variables under consideration. Even though the two different data sets have the same average values and standard deviations of *X* and *Y*, the respective relationships between *X* and *Y* are different.

A scatter chart is often referred to as a scatter plot.

One of the most commonly used predictive models is linear regression, which involves finding the best-fitting line to the data. In the graphs in Figure 1.2, we show the best-fitting lines for each data set. Notice that the lines are the same for each data set. In fact, the measure of how well the line fits the data (expressed by a statistic labeled R^2) is the same (67% of the variation in the data is explained by the line). Yet, as we can see because we have graphed the data, in Figure 1.2a, fitting a straight line looks appropriate for the data set. However, as shown in Figure 1.2b, a line is not appropriate for data set 2. We will need to find a different, more appropriate mathematical equation for data set 2. The line shown in Figure 1.2 for data set 2 would likely dramatically overestimate values of *Y* for values of *X* less than 5 or greater than 14.

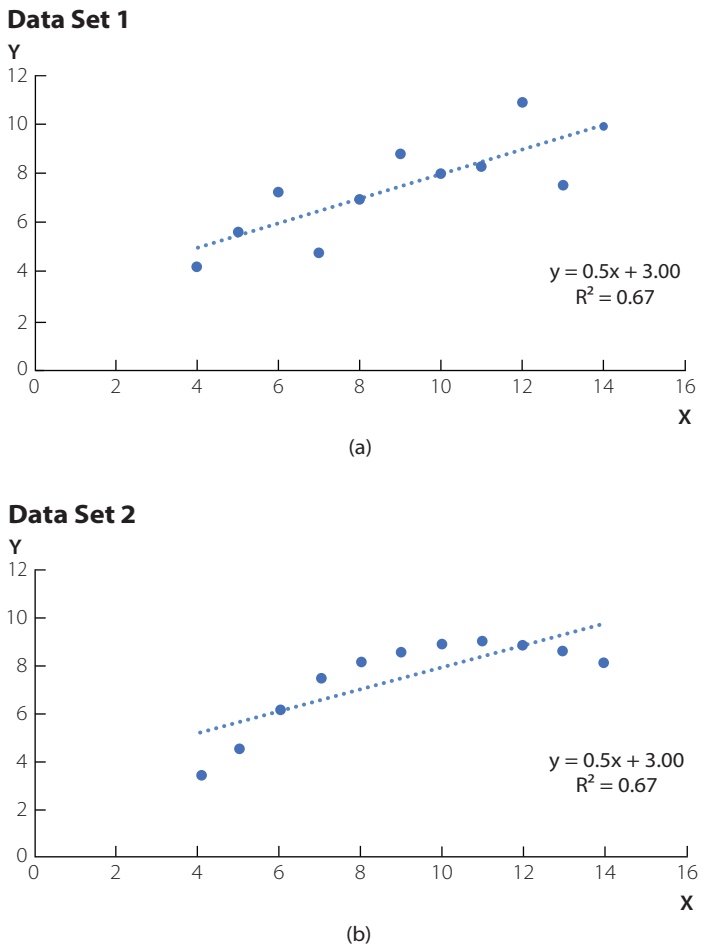
Hence, before applying predictive and prescriptive analytics, it is always best to visually explore the data to be used. This helps the analyst avoid misapplying more complex techniques and reduces the risk of poor results.

TABLE 1.2 Two Data Sets from Anscombe

	Data Set 1		Data Set 2	
	X	*Y*	*X*	*Y*
	10	8.04	10	9.14
	8	6.95	8	8.14
	13	7.58	13	8.74
	9	8.81	9	8.77
	11	8.33	11	9.26
	14	9.96	14	8.1
	6	7.24	6	6.13
	4	4.26	4	3.10
	12	10.84	12	9.13
	7	4.82	7	7.26
	5	5.68	5	4.74
Average	9	7.501	9	7.501
Standard Deviation	3.317	2.032	3.317	2.032

[2]Anscombe, F. J., "The Validity of Comparative Experiments," *Journal of the Royal Statistical Society*, Vol. 11, No. 3, 1948, pp. 181–211.

FIGURE 1.2 Anscombe's Data Displayed Graphically



DATA file
Anscombe

Data Visualization for Explanation

Data visualization is also important for *explaining* relationships found in data and for explaining the results of predictive and prescriptive models. More generally, data visualization is helpful in communicating with your audience and ensuring that your audience understands and focuses on your intended message.

Let us consider the article, "Check Out the Culture Before a New Job," which appeared in *The Wall Street Journal*.[3] The article discusses the importance of finding a good cultural fit when seeking a new job. Difficulty in understanding a corporate culture or misalignment with that culture can lead to job dissatisfaction. Figure 1.3 is a re-creation of a bar chart that appeared in this article. A **bar chart** shows a summary of categorical data using the length of horizontal bars to display the magnitude of a quantitative variable.

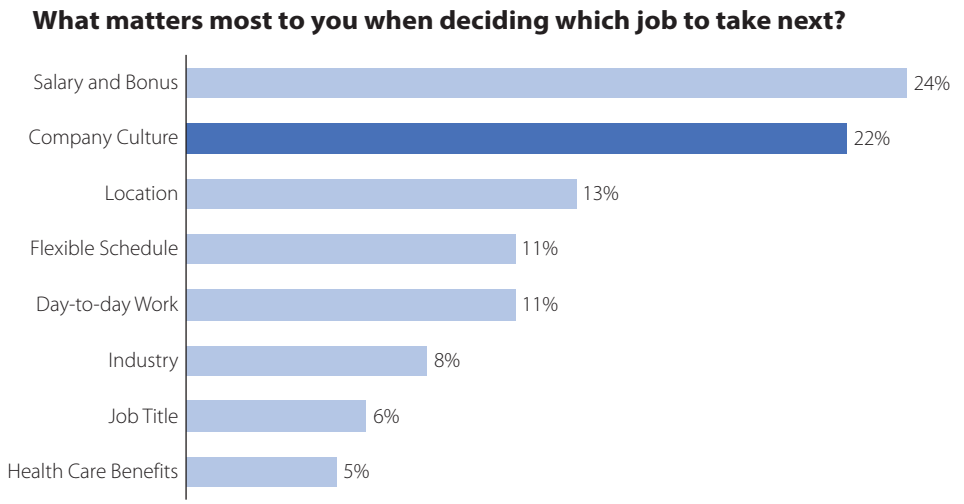
The chart shown in Figure 1.3 shows the percentage of the 10,002 survey respondents who listed a factor as the most important in seeking a job. Notice that our attention is drawn to the dark blue bar, which is "Company culture" (the focus of the

[3]Lublin, J. S. "Check Out the Culture Before a New Job," *The Wall Street Journal*, January 16, 2020.

article). We immediately see that only "Salary and bonus" is more frequently cited than "Company culture." When you first glance at the chart, the message that is communicated is that corporate culture is the second most important factor cited by job seekers. And as a reader, based on that message, you then decide whether the article is worth reading.

The effective use of color is discussed in more detail in Chapter 4.

FIGURE 1.3 A Bar Chart of Survey Results of Job Seekers



1-3 Types of Data

Different types of charts are more effective than others for certain types of data. For that reason, let us discuss the different types of data you might encounter.

The Dow Jones Industrial Average is a stock market index. It was created in 1896 by Charles Dow. The 30 companies that are included in The Dow change periodically to reflect changes in major corporations in the United States.

Table 1.3 contains information on the 30 companies that make up the Dow Jones Industrial Index (DJI). The table contains the company name, the stock symbol, the industry type, the share price, and the volume (number of shares traded). We will use the data contained in Table 1.3 to facilitate our discussion.

Quantitative and Categorical Data

Quantitative data are data for which numerical values are used to indicate magnitude, such as how many or how much. Arithmetic operations, such as addition, subtraction, multiplication, and division, can be performed on quantitative data. For instance, we can sum the values for Volume in Table 1.3 to calculate a total volume of all shares traded by companies included in the Dow, because Volume is a quantitative variable.

Categorical data are data for which categories of like items are identified by labels or names. Arithmetic operations cannot be performed on categorical data. We can summarize categorical data by counting the number of observations or computing the proportions of observations in each category. For instance, the data in the Industry column in Table 1.3 are categorical. We can count the number of companies in the Dow that are, for example, in the food industry. Table 1.3 shows two companies in the food industry: Coca-Cola and McDonald's. However, we cannot perform arithmetic operations directly on the data in the Industry column.

TABLE 1.3 Data for the Dow Jones Industrial Index Companies (April 3, 2020)

Company	Symbol	Industry	Share Price ($)	Volume
Apple Inc.	AAPL	Technology	241.41	32,470,017
American Express	AXP	Financial Services	73.6	9,902,194
Boeing	BA	Manufacturing	124.52	36,489,379
Caterpillar Inc.	CAT	Manufacturing	114.67	4,803,174
Cisco Systems	CSCO	Technology	39.06	21,235,157
Chevron	CVX	Petroleum	75.11	14,317,998
Disney	DIS	Entertainment	93.88	14,592,062
Goldman Sachs	GS	Financial Services	146.93	2,773,298
Home Depot, Inc.	HD	Retailing	178.7	6,762,357
IBM	IBM	Technology	106.34	3,909,196
Intel Corporation	INTC	Technology	54.13	23,906,062
Johnson & Johnson	JNJ	Pharmaceutical	134.17	9,409,033
JPMorgan Chase	JPM	Financial Services	84.05	20,363,095
Coca-Cola	KO	Food	43.83	13,294,556
McDonald's	MCD	Food	160.33	4,361,094
3M Company	MMM	Conglomerate	133.79	3,461,642
Merck & Co.	MRK	Pharmaceutical	76.25	9,181,539
Microsoft	MSFT	Technology	153.83	41,243,284
Nike	NKE	Apparel	78.86	8,297,443
Pfizer	PFE	Pharmaceutical	33.64	30,306,371
Procter & Gamble	PG	Consumer Goods	115.08	7,520,086
Travelers	TRV	Financial Services	93.89	1,595,000
UnitedHealth Group	UNH	Healthcare	229.49	4,356,992
Raytheon	UTX	Conglomerate	86.01	13,203,254
Visa	V	Financial Services	151.85	11,649,519
Verizon	VZ	Telecommunication	54.7	16,304,703
Walgreens	WBA	Retailing	40.72	6,489,129
Walmart	WMT	Retailing	119.48	9,390,287
Exxon Mobil	XOM	Petroleum	39.21	48,094,821

Cross-Sectional and Time Series Data

We distinguish between cross-sectional data and times series data. **Cross-sectional data** are collected from several entities at the same or approximately the same point in time. The data in Table 1.3 are cross-sectional because they describe the 30 companies that comprise the Dow at the same point in time (April 2020).

Time series data are data collected over several points in time (minutes, hours, days, months, years, etc.). Graphs of time series data are frequently found in business, economic, and science publications. Such graphs help analysts understand what happened in the past, identify trends over time, and project future levels for the time series.