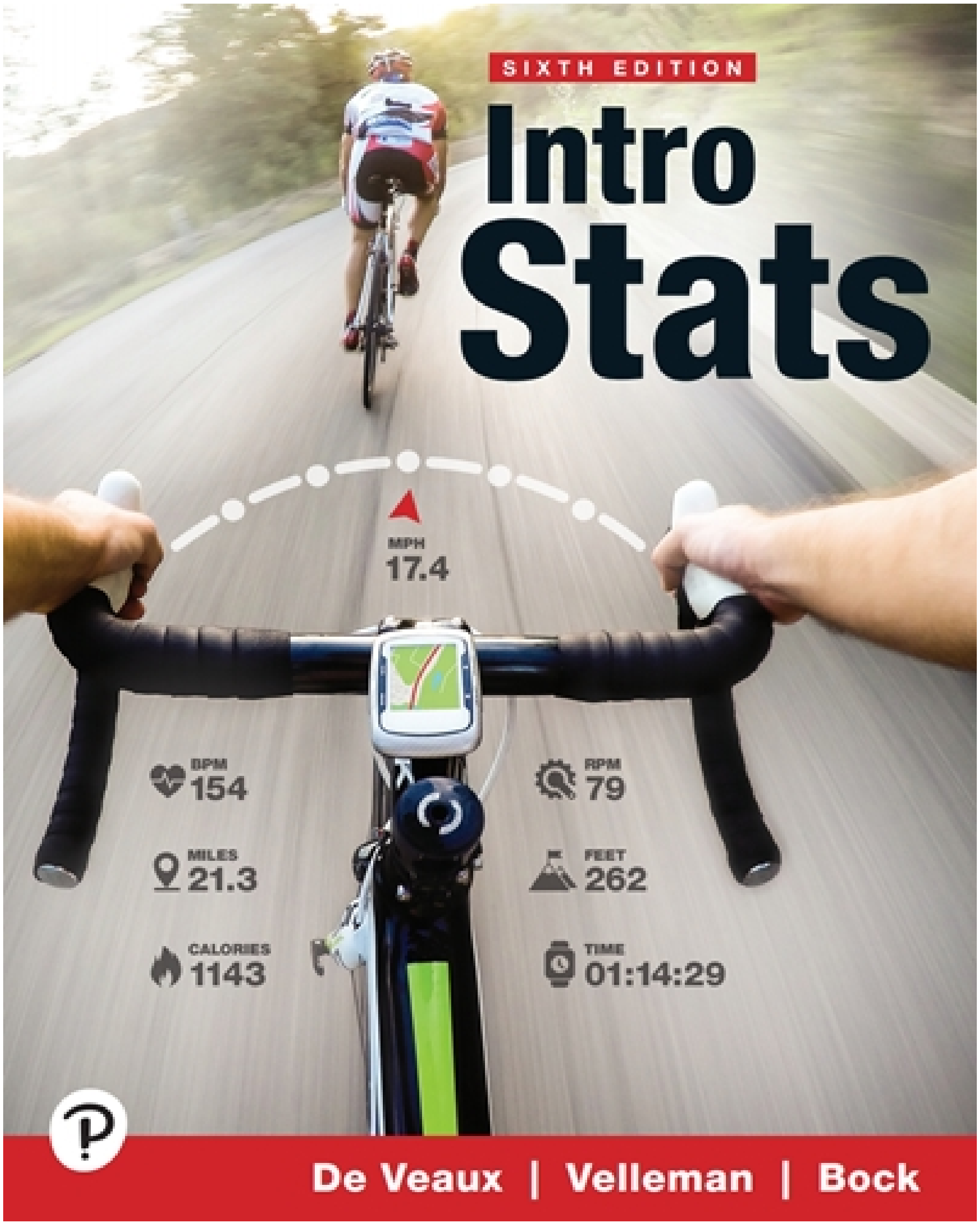
Solutions for Intro Stats 6th Edition by De Veaux

CLICK HERE TO ACCESS COMPLETE Solutions



Solutions

Instructor's Guide

Part I: Exploring and Understanding Data: Chapters 1–5

About This Guide

This Instructor's Guide is designed to help you prepare your classes and optimize your students' experience in this course. Every textbook is a compromise. Authors must decide how to order topics, what to say, and—even more important—what *not* to say. In this Instructor's Guide, we bring you behind the curtain to share with you how and why this book works the way it does. You'll find pages like this preceding each of the book's five parts in this Instructor's Edition. We'll explain our approach to the course and the reasoning behind the teaching decisions we've made. We'll summarize each chapter, highlighting the important concepts and pointing out where they'll show up later in the course. We'll offer some pedagogical suggestions based on our teaching experience, and include examples and activities you may choose to use in your class. The discussion here is at a higher level of statistical sophistication than in the book. We'll point forward to show how earlier topics are designed to prepare students for more advanced concepts and methods.[1]

Our discussions are structured to make it easy for you to find what you may be looking for. A typical chapter introduction contains some or all of the following features:

What's It About?

We summarize the topics covered in the chapter. What's more important, we tell the *story* of the chapter. Chapters in this book are not just sequences of definitions and equations. Each chapter starts with a real-world problem and then follows the data to illustrate new concepts and methods. There's a background story about how we can understand the world with data—the fundamental purpose of statistics and, we hope, the reason students are studying the subject. We know that many students are in class because the course is required. You may want to raise the question on the first day of class of *why* statistics is required by so many disciplines. (Most students never think about that.) We suggest that the reason is that we believe that our students know something about their respective fields and that statistics will help them combine that knowledge with data to achieve insights and make sound decisions. We offer a wide variety of real-world data sets and applications to illustrate this synergy.

We have structured the order of topics in the book so that each new statistics concept fits with what students have learned in previous chapters and what they will learn in subsequent ones. We give you the overview so you'll see how it all fits together.

Comments

This section explains the statistical and pedagogical reasons for the choices we've made about what to teach, how to present it, and in what order to discuss it. Some of these choices may differ from those made by other textbook authors. We try to point out these differences and explain our approach.

[1]If you are a student who has obtained an Instructor's Edition of the book, we urge you *not* to read these sections, as they may confuse you. We also caution you that, although you may believe that you now have the answers to all of the exercises in the back of the book, those answers are just the numbers and not the fully worked solutions. Don't try to turn those in without doing the work; you won't learn the material (which will hurt on the exams), and whoever grades your work will be able to tell that you copied it from the Instructor's Edition.

Random Matters

The most difficult ideas to teach in the introductory statistics course are the concept of a sampling distribution and the Central Limit Theorem. Both challenge students' intuitions ("The mean is actually *random*?!" "Yes, but it isn't haphazard; it follows a Normal model." "Yeah, sure. Tell me another." "OK, I will: The distribution of the data doesn't matter for this to be true." "?!?").

So, in this edition we have taken advantage of technology to spread out these concepts, introducing them at strategic points in the narrative that forms the book's story. We do this with new elements named **Random Matters**. As you read this Instructor's Guide for the first three parts of the text, you'll see how we lay a careful foundation for inference so that by the time we get to the Central Limit Theorem in Part IV, it is reassuring rather than challenging, telling students that what they've come to expect, really is true.

Many of the Random Matters elements describe experiments that students can perform using the online applications provided with the book or using many of the statistics packages commonly adopted with this book. We encourage that where possible. Alternatively (or additionally), we suggest that you demonstrate the experiments in class. (Don't worry: The Central Limit Theorem really is true, so your random results will show what you want them to.)

The Random Matters elements use special tools that we hope students will be able to access for themselves. The tools are available at the book's website and also at **astools.datadescription.com**. They run in any browser on a computer, tablet, or smart phone. Many RM elements lead students step-by-step through an experiment. Some tools, such as those for bootstrap calculations, can work with any dataset (tab-delimited text). Some exercises (marked with a "dice" symbol) ask students to use the random tools.

Ethics Matters

The GAISE Report challenges us to make discussions of ethics an integral part of the course. We have met that challenge with new "Ethics Matters" elements in each chapter. Some discuss ethical conduct, for example, the ASA *Ethical Guidelines* and the Belmont Report on Human Subjects Research. Some discuss sensitive topics such as the tension between an individual's right to identify their gender or race as they please and the expectation of individual privacy in data. Others recount current events, such as the statistically dubious claims of a widely advertised brain supplement, the secret collection of data by a dating site, and concerns about data ownership and privacy. And some challenge students to use their developing understanding of statistics to think about more technical questions such as ethical considerations in choosing predictors for a regression model, and the design of unbiased surveys. All are excellent fodder for classroom discussion.

Datasets

Throughout the text we refer to real data. Exercises that use real data have a "T" symbol in the margin. The data themselves (more than 300 datasets) are available on the book's website and also in the Data and Story Library, DASL, found at **dasl.datadescription.com**. DASL is a general-purpose archive of data for statistics instruction and can be searched by topic, dataset size, statistics topic, and of course, by dataset name. When an exercise is marked with a "T" symbol, the title of the exercise gives the name of the corresponding dataset.

Looking Ahead

The Looking Ahead sections point out ways that many of the ideas we introduce in early chapters foreshadow or pave the way for important features of later chapters. These are often good points to make in class to motivate students and to help them fit all these new concepts together into a coherent whole.

Class Do's

We offer pedagogical advice about approaches that have worked for us, ideas to stress, and other ways to highlight important concepts or take advantage of important features of this text. For example, we recommend that you emphasize that certain words are used by statisticians in ways that might be different from the ways students use them. Recent research suggests that it helps students if they keep their own glossary of terms—but only if they write it out for themselves. Copying and pasting definitions doesn't help at all.

The Importance of What You Don't Say

One of the challenges of teaching Intro Stats is that there's so much to say. It is easy to overwhelm a student who is seeing this material for the first time. Because deciding how much to say and when to say it is essential, we offer some suggestions about what *not* to say *quite yet* and what not to say at all.

Class Examples

It's always good to have another example for class. Students always want one more example. So we provide new examples. These sometimes include classroom materials in the form of worksheets or guided explorations.

At the End of Each Chapter

Each chapter ends with several elements designed to help students study the material:

Connections points out how the topics in the current chapter relate to those we've seen before and foreshadows some topics to come.

What Can Go Wrong? points out common errors in hopes of helping students to avoid them.

Chapter Review lists the main points (some might call them "learning objectives") of the chapter.

Review of Terms lists all the highlighted terms in the chapter and provides both a definition and a page reference back into the chapter.

When reviewing for exams, students might read the *Chapter Review* and check that they know all the *Terms*. That's a sound start.

Tech Support provides a starting place for several commonly used statistics packages. It is not meant to be thorough documentation, but just enough to get a student started. We provide support for *Data Desk*, *Excel*, *JMP*, *MINITAB*, *R*, *SPSS*, *StatCrunch*, and TI-83/84 calculators.

The book uses more than 250 data sets. These are available both at the book's website, www.pearsonhighered.com/dvb, and from the DASL archive at www.dasl.datadesk.com.

A Note on Computers

We are very much in favor of students using computers. We'd like students to try out the random experiments in the Random Matters sections for themselves rather than just trusting us or watching as the instructor demonstrates them. We hope they have a statistics package with which to make displays and compute analyses. Many exercises have data sets in the online archives. Some actually require that students use software to make plots or calculate results.

However, we are not in favor of students "taking notes" on computers in this class. Much of what they ought to be recording consists of plots, diagrams, equations, and other things better written by hand. And there is research showing that taking notes by hand just works better (e.g., www.psychologicalscience.org/news/releases/take-notes-by-hand-for-better-long-term-comprehension.html).

Just a Few More Words About the Text . . .

You'll see that this book isn't written like most other texts. We hope to entice students to read the book with a conversational style and to entertain them with occasional humor and stories. Don't be fooled into thinking that we take the subject itself lightly. We are professional statisticians and teachers. Our experience teaching, consulting, and solving real world problems informs many of our decisions. But our presentation is designed to entice students to read the book.

The 20 chapters of this book are grouped into five parts. Each chapter tells a story, but each part tells a larger tale and that fits into the narrative of the whole text. At the end of each part, you'll find an overview of the major concepts of that part, gathering the chapter stories into a larger whole. This overview is followed by a yet another set of exercises. These review exercises often integrate several concepts from different chapters and appear in no particular topic order, so they are particularly helpful in preparing students for tests.

Some Important Resources

No physical DVD accompanies this text. During the lifetime of this textbook series we have distributed CDs and then DVDs, but technology has moved on. Many computers lack a DVD player, which makes the distribution of data sets and software in this manner problematic for teachers and students alike. So now most of the materials previously available on DVD can be found online:

The *Data Desk* statistics program is available from datadesk.com. New tools that provide interactive versions of the distribution tables at the back of the book and tools for randomization inference methods and for sampling repeatedly from larger populations can be found online at astools.datadesk.com.

Test Bank and Resource Guide

This resource complements and expands on the commentary and teaching suggestions found in the Instructor's Guide sections of this Instructor's Edition. It includes chapter quizzes, unit tests, and suggestions for projects, as well as Web links and lists of other resources.

10-digit ISBN: 0134265408 13-digit ISBN: 9780134265407

Other Books and Magazines

- *Chance,* Springer-Verlag (1-800-SPRINGER). This magazine, published quarterly, provides articles about statistics as well as excellent examples and data sets to use in class.
- *Stats,* American Statistical Association. This is a magazine for students that provides articles about statistics and examples that you might find useful for class preparation.
- *Significance. Significance* is a joint effort of the Royal Statistical Society and the American Statistical Association. It aims to publish "a statistical view of what's going on in the world." Its online site, www.significancemagazine.org/view/index.html, is updated daily.

Internet URLs

The Internet is a valuable source of data sets, examples, tables, random numbers, and current events. The good news is that you can probably find almost everything you need or want to know there. The bad news is that the materials are not consistent or integrated. Be especially wary of introducing students to a variety of online applets, each with its own interface, notation, terminology, and assumptions.

Many of the data sets and examples in the book are sourced from Internet sites. The data used in the book are available online along with source URLs where available. Those URLs have gotten so long that you probably didn't want to type them anyway.

Other Sources of Data

www/dasl.datadesk.com
www.fedstats.gov
www.madd.org/stats/0,1056,1845,00.html (DWI statistics)
www.fbi.gov/ucr/01cius.htm (crime statistics)
www.amstat.org/publications/jse (*Journal of Statistics Education*)

CHAPTER 1

Stats Starts Here

What's It About?

Chapter 1 describes the important features of the text and then gets right down to business discussing data. We've given the chapter an unusual title and tried to grab students' attention with a humorous footnote. (Some students have e-mailed us to assure us that they *do* read the footnotes.) If we can get them to read three words and the footnote, maybe we can get them to read on. Then we talk about the importance of context (the W's), discuss variables, and make the distinction between categorical and quantitative data.

The first two chapters cover data, their display, and summaries. Many students will recognize some of the material from middle and high school, so our emphasis is on statistical thinking and we move at a rapid pace. Of course, we define terms and provide examples. But we also discuss *why* the methods presented are used and what we hope students will learn from them. These are concepts that appear throughout the course. What to look for in a histogram is more important than how to make one—the computer will do that. Even more important than what to look for is the underlying lesson that there are *reasons* for displaying and summarizing data. These reasons inform and motivate the entire course.

Comments

This is the students' first look at the style of the book, and we do lay it on more heavily in the first few pages than we will when discussing, say, confidence intervals. We want to shake things up. We want students to notice that this is not the same old math or science textbook they've seen before. And we'd like to get them on our side. That's the reason for the humor and self-deprecating remarks.

Every statistics text starts with a definition of *statistics*. We do too, but ours is different. And the difference matters. We say that statistics is *a way of reasoning* and that the goal is *to help us understand the world*. We've found it helpful to reinforce this idea throughout the semester, especially when we get into the methods sections of the course. This book is primarily about statistical thinking. Methods, definitions, and skills are all here, but each is presented with the purpose of understanding the world. That's why every example follows the *Think, Show, Tell* pattern, starting with careful reasoning and concluding with a sentence or two telling what we've learned about the world.

It is valuable to get students involved with data from the start. We don't take a "big picture" approach at this time. There will be plenty of time to build models and draw inferences later. For now, let's just get our hands dirty playing with the data. When students have a good sense of what kinds of things data can say to us, they learn to listen to the data. Throughout the course, we insist that no analysis of data is complete without telling what it means. This is where that understanding starts.

A Note on Features

We don't spend time in class on the features of the text. That material is meant to be read *by* students, not *to* them. Our goal here is to get the class moving and talking about data and what we can learn from data.

Looking Ahead

Technology plays an important role in this book. We expect students to use a calculator or statistics package for finding the numerical "answers." So we won't spend much time worrying about the calculation details, although we do expect students to understand what's happening. Instead, we focus on understanding and meaning. But the book is "technology neutral." The "computer output" in the book is designed to look a lot like many package results, but exactly match none of them. Students should feel comfortable using almost any statistics program or graphing calculator. This is a good time to introduce whatever technology your students will be using. Have them summarize the results of a survey of the class or some other small data set. Point them to the *Tech Support* sections at the end of most chapters. Those sections offer instructions to help students get started on several common statistics packages and calculators.

Class Do's

One of our favorite definitions is "Statistics is the art of distilling meaning from data." Data have a story to tell. Our objective is to uncover that story. Collect some data in class, and ask students to look for interesting facts hiding there. Get the class thinking about what the term "data" means. Students need to understand that data are not only numbers and that they must have a context (the W's). When data are quantitative, they should also have measurement units.

There are two ways we treat data: ***categorical*** and ***quantitative***. Don't get distracted by worrying about ratio, interval, and other distinctions. These are problematic and don't matter for the concepts and methods discussed in this book. Emphasize that the distinction between treating data as categorical or quantitative may be more about how *we* display and analyze data than it is about the variable itself. The variable "sex" is data, but just because we might label the males as 1 and the females as 0 doesn't mean that it's quantitative. On the other hand, taking the average of those 0's and 1's does give us the percentage of males. How about *age*? It is often quantitative, but it could be categorical if broken down only into *child*, *adult*, and *senior*. ZIP code is usually categorical, but if one business had an "average" ZIP code for their customers of 10000 while another had 90000, we'd know the latter had more customers in the western United States. Emphasize the importance of the context and the W's in summarizing these data.

Every discipline has its own vocabulary, and statistics is no exception. Students need to understand and use that vocabulary properly. Unfortunately, many statistics words have a common everyday usage that's not quite the same. We'll be pointing those out as we go along.

One of the first vocabulary words should be "variable." Point out that it does not mean the same thing as it did in Algebra. There, we call *x* a variable, but often that meant we didn't know its value. In statistics, a variable is an attribute or characteristic of an individual or object whose value varies from case to case.

Why do we talk about "a statistic" when we don't discuss "a mathematic" or "a physic"? Statistics is a whole that is made up of many parts, and each of those parts has its own meaning and its own story to tell. A *statistic* is a numerical summary of data. The book wisecracks that, contrary to an advertised saying, you can't be a statistic, only a datum.

Point out that summaries of data can be verbal, visual, and numerical. All are important. In fact, any complete analysis of data almost always includes all three of these.

The Importance of What You Don't Say

One of the reasons statistics can be difficult to teach is that we often deal with vague concepts. Students and teachers both like clear definitions; they're easier to teach, learn, and grade. But reasonable people can disagree about whether a histogram is symmetric or skewed, whether a straggling point is really an outlier or just the largest value, or whether two groups we want to compare vary about the same amount. It is important to allow students their own opinions and insights into data.[2]

This raises the issue of ethical practice in statistics. We are engaged in an honest search for truth and understanding, and that's what should guide our (and our students') judgments. Emphasize this point now to alert students that this isn't a course about calculating the right answer, but about understanding the world.

We are laying a foundation here. Stretching up to the attic at this point just makes everyone feel unsafe. Many fundamental statistics terms are left unmentioned in this chapter. You'll find it best to leave it that way. We'll get to them when the students have a safe place to file them along with their other knowledge. So we have an unusually long list of terms we recommend leaving for later in the course. In particular, avoid saying the following:

Population, Sample, Hypothesis, Inference. These are certainly important terms in this course, but we have no background for discussing them honestly now, so they would just be confusing and intimidating.

Nominal, Ordinal, Interval, Ratio. "Nominal" is used by some software packages as a synonym for "categorical" just as "continuous" is used for "quantitative." These distinctions arise from studies of measurement scales. But it isn't correct to claim that each variable falls into one of these categories. It is the use to which the data are put that determines what properties the variable must have. Ordinal categorical data may come up, but there are no special techniques for dealing with ordered categories in this course. And any differences between interval- and ratio-scaled data are commonly ignored in statistical analyses. If any of these terms were mentioned now, they'd never come up again anyway.

Probability, Correlation. Everyone has some intuitive sense of these terms, and we'll deal with them formally—but not for a while. Students may want to use these terms, but at this early stage in the course, we don't need them. Without background and careful definition, they are likely to be misused and can simply be frightening.

Ethics Matters

One ethics discussion deals with protecting human subjects (the Belmont Report). This can be excellent fodder for a classroom discussion. Is it important to collect data about human subjects? How can we do so while respecting individuals' rights? When facing a pandemic, is contact tracing appropriate? Should you be required to participate? A second discussion addresses data privacy and the large amount of data collected by online services. Chances are, students do not know of the National Strategy for Trusted Identities in Cyberspace—and they are likely to have opinions about it.

Few statistics texts address problems of data directly, but we think that's important—and of increasing importance. Don't skip over this part of the chapter.

Class Examples

You might take a quick class survey. We suggest asking for things like gender, political leaning (Liberal, Moderate, Conservative), number of siblings, number of countries visited, whether they play varsity sports, GPA, height, handedness (left or right), and shoe size. Be sure to include both categorical and quantitative values.[3] Try this question, after getting everyone's attention: Ask your students to pick a number at random between 1 and 10 and write it down

[2]There was a time, not long ago, when our students' first question, "Why am I taking this course?" was typically answered by "It's required. Sit down and be quiet." We propose that a better answer is "So you can learn how data can tell us about the world. Stand up and tell us what you see."

[3]Recall what you were really interested in at that age and try some carefully worded questions on those subjects.

quickly. (Later, you can look to see how "random" these numbers really are.) Have the data collected and duplicated, or put them up on a website before the next class. Ask students what story the data tell about their class. If you wish, you can make a stem-and-leaf display or scatterplot of some of the data immediately.

Hints: Data are rarely as simple as they seem. Suggest these variables, then pause for some discussion. Does touching down at an airport qualify as "visiting" a country? Does an only child count herself when counting siblings? Should shoe sizes be adjusted because men's and women's size 7 are different sizes? If you write with your left hand but throw with your right, are you left-handed? Give your students a chance, and they'll find other issues—and they'll be developing a healthy skepticism toward data. That's just what we want, so they'll value the tools that help them look at data more carefully.

If you don't specify units for height, you may get some values in centimeters. Alternatively, if you specify inches, you may get a 55 from someone who meant 5′5″. Those outliers make for good class discussion.

If you teach a large section, consider collecting data online. There are a number of services that will let you design an online survey and will host it for a modest price, letting students respond online at their convenience and providing you with anonymous and machine-readable responses. One we have used successfully is at www.surveymonkey.com.

Consider writing the numbers 17, 21, 44, and 76 on the board. Are those data? Context is critical—they could be test scores, ages in a golf foursome, or uniform numbers of the starting backfield on the football team. In each case, our reaction and what we might ask of the data change.

Run through some other examples of data, asking about the W's, the variables (what are they, what type is each used as, and what are the units), and so on.

- A *Consumer Reports* article on energy bars gave the brand name, flavor, price, number of calories, and grams of protein and fat.
- A report on the Boston Marathon listed each runner's gender, country, age, and time.

CHAPTER 2

Displaying and Describing Data

What's It About?

We introduce students to distributions, displays, and common summary statistics. The mathematics is easy, and the graphs are straightforward. We challenge students to uncover the story the data tell, and to write about it in complete sentences in context.

There's a lot of material in this chapter, but it should be review for most students. Almost all of it appears in the Common Core, and other parts are simply common in the news and online. We move through the material at a brisk pace for that reason. There's no need to linger over the calculation of the mean because most students know that, or how to make a histogram because we'll let technology do that. We devote most of our attention to discussing what to look for in data displays and how to think about what displays and summaries say about the data and about the world they describe.

We first consider summaries and displays of categorical data. We see frequency tables and relative frequency tables and display their contents with bar charts, pie charts, and ring charts. We then use pie charts in the first Random Matters exploration.

Random Matters

The **Random Matters** elements are designed strategically to introduce students to ideas about randomness, inference, and sampling distributions gradually and without overwhelming them. The first Random Matters example here gives the flavor.

If the chance of surviving the sinking of the *Titanic* had been "fair," then the distribution of survivors should look like a random sample of 712 of the people aboard. We present 24 pie charts of such random samples and hide among them a pie chart of the actual distribution of survivors. Can you pick it out?

There are several lessons here:

First, and most important, is that samples differ from one to the next. All these pie charts are different. Encourage students to look at them to get a sense of *how much* the samples vary. We said in Chapter 1 that statistics is about variation. That will be a theme throughout the book (as it is throughout statistics). This is a good opportunity to remind students of this theme.

Second, we look for exceptions that stand out against the background of underlying random variation, and identify them as important. That's the first step in thinking about hypothesis testing—but, of course, you shouldn't say so.

Third, we can use the computer to get an idea of how samples differ. We don't need to repeat the experiment by sinking the *Titanic* many times and allowing fair access to the boats to be able to see how much the results would vary.

Finally, there is the underlying concept that "fair" and "random" are concepts that go together. We all have a pretty common idea of what is "fair"; we have had one since we were children. We call upon students to consult their inner child to see that random allocations and random selections are fair.

The next section of the chapter takes up quantitative data. We examine histograms, stem-and-leaf displays, and dotplots. We want students to learn to describe the important features of a distribution: shape, center, spread, and any unusual features such as outliers, gaps, or clusters. Next, we introduce how to summarize the center of a distribution with the median or mean. Then we look at the spread with the IQR and standard deviation. We discuss the problems that arise when means and standard deviations are used for skewed data or in the presence of outliers; we recommend medians and IQRs for such data.

In Chapter 5, we'll make it clear that the central concept in all of statistics is that distances and differences are measured with a ruler based on the standard deviation. Encourage students to develop an intuitive sense of variation in data.

Why y? In this chapter it becomes clear that we use the letter y to denote the variable that we are summarizing, about which we would like to know something, or that we are modeling. This won't

APPENDIX A Answers

Here are the "answers" to the exercises for the chapters and the unit reviews. The answers are outlines of the complete solution. Your solution should follow the model of the Step-by-Step examples, where appropriate. You should explain the context, show your reasoning and calculations, and draw conclusions. For some problems, what you decide to include in an argument may differ somewhat from the answers here. But, of course, the numerical part of your answer should match the numbers in the answers shown.

Chapter 1

1. Retailers, and suppliers to whom they sell the information, will use the information about what products consumers buy to target their advertisements to customers more likely to buy their products.

2. Retailers will use the information about items you have purchased and viewed to target their recommendations to you. This will increase the likelihood that you will purchase the products they advertise to you.

3. Owners can advertise about the availability of parking. They can also communicate with businesses about hours when more spots are available and when they should encourage more business.

4. This rise and fall of temperature and water levels can help in planning for future problems and guide public policy to protect our safety.

5. The individual games.

6. The years.

7. The sample is about 5000 people; the *Who* is the selected subjects; the *What* includes medical, dental, and physiological measurements and laboratory test results.

8. The *Who* is the photos; the *What* could be facts about the photos.

9. a) Sample—A principal wants to know how many from each grade will be attending a performance of a school play.
b) Sample—A principal wants to see the trend in the amount of mathematics learned as students leave each grade.

10. a) ZIP codes simply identify a location. There are no units and there is little to be learned from numerically analyzing them.
b) ZIP codes begin numbering in the east and get higher as they move west.

11. Categorical.

12. Quantitative.

13. Quantitative.

14. Quantitative.

15. They might consider whether a person voted previously or whether they could name the candidates (indicating greater interest in the election).

16. They can use the models to predict the average temperature ten days in advance and compare their predictions to the actual temperatures.

17. Answers will vary.

18. Answers will vary.

19. *Who*—40 undergraduate women; *What*—Ability to differentiate gay men from straight men; *Population*—All women.

20. *Who*—Participants in the study; *What*—Recorded heart rate, oxygen consumption, and perceived exertion; *Population*—All adults.

21. *Who*—2500 cars; *What*—Distance from car to bicycle; *Population*—All cars passing bicyclists.

22. *Who*—30 similar companies; *What*—401(k) employee participation rates; *Population*—All similar companies.

23. *Who*—Coffee drinkers at a Newcastle University coffee station; *What*—Amount of money contributed; *Population*—All people in honor system payment situations.

24. *Who*—24 patients suffering from blindness; *What*—Effect of treatments on blindness; *Population*—All people suffering from these two types of blindness.

25. *Who*—474 participants in the San Antonio Longitudinal Study of Aging; *What*—Diet soda consumption and waist size change; *Population*—All diet soda drinkers.

26. *Who*—10 crankshafts at Cleveland Casting; *What*—The pouring temperature of molten iron; *Population*—All Cleveland Casting crankshafts.

27. *Who*—54 bears; *Cases*—Each bear is a case; *What*—Weight, neck size, length, and sex; *When*—Not specified; *Where*—Not specified; *Why*—To estimate weight from easier-to-measure variables; *How*—Researchers collected data on 54 bears they were able to catch.
Variable—Weight; *Type*—Quantitative; *Units*—Not specified;
Variable—Neck size; *Type*—Quantitative; *Units*—Not specified;
Variable—Length; *Type*—Quantitative; *Units*—Not specified;
Variable—Sex; *Type*—Categorical.

28. *Who*—Students; *Cases*—Each student is an individual case; *What*—Age, race or ethnicity, number of absences, grade level, reading score, math score, and disabilities/special needs; *When*—Current; *Where*—Not specified; *Why*—Keeping this information is a state requirement; *How*—The information is collected and stored as part of school records; *Variable*—Age; *Type*—Quantitative; *Units*—Not specified, probably years (perhaps years, months); *Variable*—Race or ethnicity; *Type*—Categorical; *Variable*—Days absent; *Type*—Quantitative; *Units*—Number of days; *Variable*—Current grade level; *Type*—Categorical (could be quantitative for some purposes); *Variable*—Standardized reading score; *Type*—Quantitative; *Units*—Not specified; *Variable*—Standardized math score; *Type*—Quantitative; *Units*—Not specified; *Variable*—Disability/special needs; *Type*—Categorical.

29. *Who*—Arby's sandwiches; *Cases*—Each sandwich is a case; *What*—Type of meat, number of calories, and serving size; *When*—Not specified; *Where*—Arby's restaurants; *Why*—To assess nutritional value of sandwiches; *How*—Report by Arby's restaurants; *Variable*—Type of meat; *Type*—Categorical; *Variable*—Number of calories; *Type*—Quantitative; *Units*—Calories; *Variable*—Serving size; *Type*—Quantitative; *Units*—Ounces.

30. *Who*—U.S. Adults; *What*—Age, Gender, Education, Marital Status, Belief in God, and Frequency of Prayer; *When*—May 2014; *Where*—U.S.; *Why*—Poll of Religious attitudes and practices; *Variables*—Age—quantitative, Gender—categorical, Education—quantitative, Marital status—categorical, Belief in God—categorical, Frequency of Prayer—categorical.

31. *Who*—Electric bicyces; *What*—motor size, max speed, wheel base, brand name, removable battery; *When*—May 2020; *Where*—no location specified; *Why*—Magazine review; *Variables*—Motor size, max speed, wheel base all—quantitative. Brand and whether battery can be removed are—categorical.

32. *Who*—385 species of flowers; *Cases*—Each of the 385 species at each of the 47 years is a case, for a total of 18,095 cases; *What*—Date of first flowering; *When*—Not specified; *Where*—Southern England; *Why*—The researchers believe that early flowering indicates a warming of the overall climate; *How*—Not specified; *Variables*—Date of first flowering; *Type*—Quantitative; *Units*—Days elapsed since January 1.

33. *Who*—Experiment subjects; *Cases*—Each subject is a case; *What*—Treatment (herbal cold remedy or sugar solution) and cold severity; *When*—Not specified; *Where*—Not specified; *Why*—To test efficacy of herbal remedy on common cold; *How*—The scientists set up an experiment; *Variable*—Treatment; *Type*—Categorical; *Variable*—Cold severity rating; *Type*—Quantitative (perhaps ordinal categorical); *Units*—Scale from 0 to 5; *Concerns*—The severity of a cold seems subjective and difficult to quantify. Scientists may feel pressure to report negative findings of herbal product.

34. *Who*—American vineyards; *Cases*—Each vineyard is a case; *What*—Size of vineyard, number of years in existence, state, varieties of grapes grown, average case price, gross sales, and percent profit; *When*—Not specified; *Where*—United States; *Why*—To provide information for American grape growers; *How*—Not specified; *Variable*—Vineyard size; *Type*—Quantitative; *Units*—Acres; *Variable*—Number of years in existence; *Type*—Quantitative; *Units*—Years; *Variable*—State; *Type*—Categorical; *Variable*—Varieties of grapes grown; *Type*—Categorical; *Variable*—Average case price; *Type*—Quantitative; *Units*—Not specified (dollars?); *Variable*—Gross sales; *Type*—Quantitative; *Units*—Not specified (dollars?); *Variables*—Percent profit; *Type*—Quantitative; *Units*—Percent.

35. *Who*—Streams; *Cases*—Each stream is a case; *What*—Name of stream, substrate of the stream, acidity of the water, temperature, BCI; *When*—Not specified; *Where*—Upstate New York; *Why*—To study ecology of streams; *How*—Not specified; *Variable*—Stream name; *Type*—Identifier; *Variable*—Substrate; *Type*—Categorical; *Variable*—Acidity of water; *Type*—Quantitative; *Units*—pH; *Variable*—Temperature; *Type*—Quantitative; *Units*—Degrees Celsius; *Variable*—BCI; *Type*—Quantitative; *Units*—Not specified.

36. *Who*—Each model of automobile; *Cases*—Each vehicle model is a case; *What*—Vehicle manufacturer, vehicle type, weight, horsepower, and gas mileage for city and highway driving; *When*—Current; *Where*—United States; *Why*—By the Environmental Protection Agency to track fuel economy of vehicles; *How*—The data are collected from the manufacturer of each model; *Variable*—Manufacturer; *Type*—Categorical; *Variable*—Vehicle type; *Type*—Categorical; *Variable*—Weight; *Type*—Quantitative; *Units*—Not specified (pounds); *Variable*—Horsepower; *Type*—Quantitative; *Units*—Not specified (horsepower); *Variable*—Gas mileage, city; *Type*—Quantitative; *Units*—Miles per gallon; *Variable*—Gas mileage, highway; *Type*—Quantitative; *Units*—Miles per gallon; *Concerns*—Do manufacturers' ratings of their own vehicles' gas mileage reflect customer experience?

37. *Who*—Dogs trained to identify odors; *What*—Whether dogs can be trained to identify COVID-19 by odor; *When*—2020; *Where*—Helsinki; *Why*—Possible way to identify COVID patients; *Variables*—Whether dogs could identify the scent—categorical.

38. *Who*—32 volunteers; *Cases*—Each volunteer is a case; *What*—Sex, height, handedness, distance walked, sideline crossed; *When*—Not specified; *Where*—Not specified; *Why*—To see if people naturally walk in circles; *How*—Collected during a test on a football field; *Variable*—Sex; *Type*—Categorical; *Variable*—Height; *Type*—Quantitative; *Units*—Not specified (inches?); *Variable*—Handedness; *Type*—Categorical; *Variable*—Distance; *Type*—Quantitative; *Units*—Yards; *Variable*—Sideline crossed; *Type*—Categorical.

39. *Who*—Kentucky Derby races; *What*—Date, winner, jockey, trainer, owner, and time; *When*—1875 to 2020; *Where*—Churchill Downs, Louisville, Kentucky; *Why*—Not specified (To see trends in horse racing?); *How*—Official statistics collected at race; *Variable*—Year; *Type*—Identifier and Quantitative; *Units*—Year; *Variable*—Winner; *Type*—Identifier; *Variable*—Jockey; *Type*—Categorical; *Variable*—Trainer; *Type*—Categorical; *Variable*—Owner; *Type*—Categorical; *Variable*—Time; *Type*—Quantitative; *Units*—Minutes and seconds.

40. *Who*—Indy 500 races; *What*—Year, winner, time, average speed; *When*—1911 to 2016; *Where*—Indianapolis; *Why*—Not specified (To see trends in auto racing?); *How*—Official statistics at race; *Variable*—Year; *Type*—Identifier and Quantitative; *Units*—Year; *Variable*—Winner; *Type*—Categorical; *Variable*—Time; *Type*—Quantitative; *Units*—Hours, minutes, and seconds; *Variable*—Average speed; *Type*—Quantitative; *Units*—Miles per hour.

41. a) Fonso
a) In 1895, from 1.5 miles to 1.25 miles
c) 124 seconds
d) Secretariat in 1973

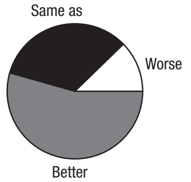
42. a) 88.618 mph
b) Twice
c) 6

Chapter 2

1.

Live comfortably	0.042
Meet their basic expenses with a little left over for extras	0.386
Just meet their basic expenses	0.479
Don't even have enough to meet basic expenses	0.086
Don't Know/No Answer	0.006

2.



3. a) Yes; each movie is categorized in a single genre.
b) Comedy.

4. a) Yes, each movie falls into only one rating category.
b) R

5. i. C,
ii. A,
iii. D,
iv. B

6. i. D,
ii. A,
iii. C,
iv. B

7. a) There are no data for those years.
b) Each year appears only once in the data set.
c) Most years, there were between 17,500 and 25,000 traffic deaths. There were also several years—possibly a second mode—with between 10,000 and 12,500 traffic deaths.

8. a) Two.
b) 1975 and 1980.
c) There have been fewer traffic deaths more recently than there were in 1975 and 1980.

9. Based on the height of the tallest points, about 85 of these 250 men have biceps close to 13 inches around. Most are between 12 and 15 inches around. But there are two as small as 10 inches and several that are 16 inches.

10. Yes, the dotplots look different. The plot based on inches has fewer values on the *x*-axis, so it shows less detail. This plot gives a better picture of the distribution.

11. It was most common for students to send just one e-mail message. Most sent five or fewer. But one student was an outlier, sending 21 e-mails.

12. a) The distributions are skewed to the right. Most states have smaller populations and fewer adoptions, but some big states have substantially more of each.
b) States with higher populations are likely to have more adoptions.
c) Report the number of adoptions per 100,000 people.

13. The distribution is mound-shaped and roughly symmetric.

14. The distribution is skewed to the right.

15. a) The distribution is slightly skewed to the right.
b) There is one mode around 1400. A second slight bump just below 3000 might be considered a mode.

16. a) The distribution is roughly symmetric and possibly bimodal.
b) There seem to be two modes, one between sizes 38 and 40 and another between sizes 44 and 46. This could be due to having data for both men and women. The lower mode might be for women and the upper for men.

17. a) The mean. The distribution is skewed to the left right, so the mean will be pulled by the tail in the higher direction.
b) The median is resistant to the skewed shape of the distribution, so it is a better choice.

18. a) The mean. The distribution is unimodal and skewed to the right, so the mean will be pulled by the tail toward the higher values.
b) The median is resistant to the skewed shape of the distribution, so it is a better choice for most summaries.

19. Because the distribution of bicep circumferences is unimodal and symmetric, the mean and the median should be very similar. The usual choice is to report the mean or to report both.

20. Because there are two modes, the mean and median are not helpful in reporting the story that the data tell. It is better to report the locations of the two modes.

21. a) IQR
b) For a skewed distribution, is it more appropriate to report the IQR. The skewness inflates the standard deviation.

22. a) IQR
b) For a skewed distribution, it is more appropriate to report the IQR. The skewness inflates the standard deviation.

23. Standard deviation. The distribution is reasonably symmetric and mound-shaped. The standard deviation is generally more useful whenever it is appropriate.

24. The data combine shoe sizes for men and for women. It isn't appropriate to summarize them as if they were a coherent collection of values.

25. Answers will vary.

26. Answers will vary.

27. Answers will vary.

28. Answers will vary.

29. Answers will vary.

30. Answers will vary.

31. Answers will vary.

32. Answers will vary.

33. a) Unimodal (near 0) and skewed to the right. Many seniors will have 0 or 1 speeding ticket. Some may have several, and a few may have more than that.
b) Probably unimodal and slightly skewed to the right. It is easier to score 15 strokes over the mean than 15 strokes under the mean.
c) Probably unimodal and symmetric. Weights may be equally likely to be over or under the average.
d) Probably bimodal. Men's and women's distributions may have different modes. The distribution may also be skewed to the right, since it is possible to have very long hair, but hair length can't be negative.

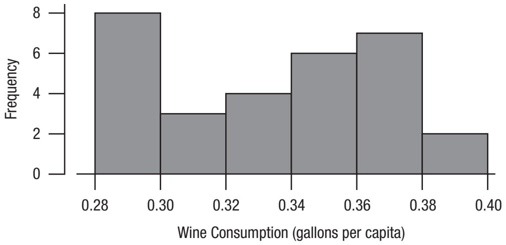
34. a) Bimodal because you have both players and parents. The distribution may also be skewed to the right, since parents' ages can be higher than the mean more easily than lower.

b) Unimodal and skewed to the right. There are probably many students with 0 or 1 sibling and some with 2 or more.
c) Unimodal and symmetric. It is unusual to have either very high or low pulse rates.
d) Uniform. Each face of the die has the same chance of coming up, so the number of times should be about the same for each face.

35. a) Thriller/Suspense
b) It is easy to tell from either chart; sometimes differences are easier to see on the bar chart because slices of the pie chart look too similar in size.

36. a) NC-17
b) It is easy to tell from either chart; sometimes differences are easier to see on the bar chart because slices of the pie chart look too similar in size.

37. a)



b) The median is 0.34 gallons per capita, and the quartiles are 0.3 gallons per capita and 0.37 gallons per capita, so the IQR is 0.07 gallons per capita.
c) This is a bimodal distribution with a peak between 0.28 and 30 gallons per capita and a second peak between 0.36 and 0.38 gallons per capita. The median is 0.34 gallons per capita and the IQR is 0.07 gallons per capita.

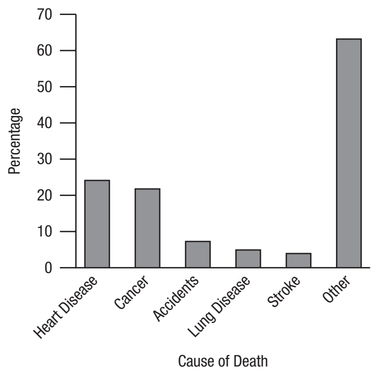
38. a)

```
11 | 3
11 | 79
12 | 0011122233333
12 | 5699
13 | 133444
13 | 5789
```

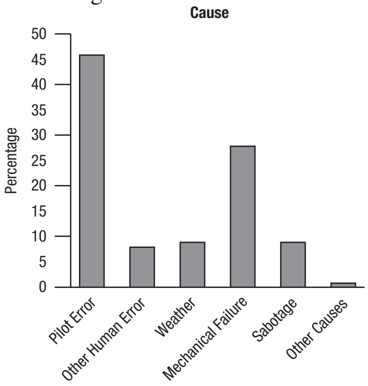
13|5 means 1.35.

b) The distribution appears bimodal with a median of 1.23 gallons per capita. The first and third quartiles are 1.21 and 1.33 gallons per capita, respectively.

39. a) Yes. We can add because these categories do not overlap. (Each person is assigned only one cause of death.)
b) 36.8%
c) Either a bar chart or pie chart with "other" added would be appropriate. A bar chart is shown.

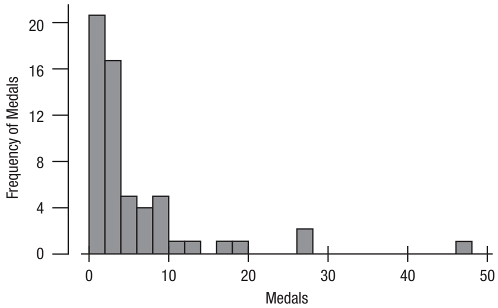


40. a) Yes, we can add because each crash was assigned to only one cause category.
b) The data are given to the nearest whole percent. If several of the categories were rounded up, an extra percentage point could be gained.
c)



41. a) There are too many categories to make a meaningful bar chart or pie chart by genre.
b) They combined several smaller categories into the category "Other."

42. a) There are too many categories to make a meaningful bar chart or pie chart by country.
b) It would be better to make a histogram of total medals.



43. The three-dimensional display distorts the sizes of the regions, violating the area principle.

44. a) The bars have false depth, which can be misleading. This is a bar chart, so the bars should have space between them. From a simple design standpoint, running the labels on the bars one way and the vertical axis labels the other way is awkward.
b) The percentages sum to 100%. This is unlikely if the respondents were asked to name three methods each. For example, it would be possible for 80% of respondents to use ice at some time and another 75% to use electric stimulation. This is a case where summing to 100% seems wrong rather than correct.

45. a) The distribution is bimodal. It looks like there may be two groups of cereals. The modes are near 13 g and 22 g. There are no outliers.
b) Corn Chex, Corn Flakes, Cream of Wheat (Quick), Crispix, Just Right Fruit & Nut, Kix, Nutri-Grain Almond-Raisin, Product 19, Rice Chex, Rice Krispies, Shredded Wheat 'n' Bran, Shredded Wheat Spoon Size, Total Corn Flakes, Triples

46. The distribution is skewed to the right and centered at around 30 to 31 minutes, with most observations between 29 and 32 minutes. The skewness is because it is easier to run much slower than usual but harder to run much faster than usual. Also, over a 10-year period, the runner may have slowed down due to injury or aging.

Instructor's Solutions Manual

Dirk Tempelaar

Maastricht University

Intro Stats

Sixth Edition

Richard D. De Veaux

Williams College

Paul F. Velleman

Cornell University

David E. Bock

Ithaca High School (Retired)

with contributions from

Brianna Heggeseth

Macalaster College

Susan Wang

Google, Inc

Pearson

Reproduced by Pearson from electronic files supplied by the author.

Pearson

ISBN-13: 978-0-13-735844-1
ISBN-10: 0-13-735844-X

Table of Contents

Chapter 1 – Stats Starts Here

Section 1.1

1. **Grocery shopping.** Discount cards at grocery stores allow the stores to collect information about the products that the customer purchases, what other products are purchased at the same time, whether or not the customer uses coupons, and the date and time that the products are purchased. This information can be linked to demographic information about the customer that was volunteered when applying for the card, such as the customer's name, address, sex, age, income level, and other variables. The grocery store chain will use that information to better market their products. This includes everything from printing coupons at the checkout that are targeted to specific customers to deciding what television, print, or Internet advertisements to use.

2. **Online shopping.** Amazon hopes to gain all sorts of information about customer behavior, such as how long they spend looking at a page, whether or not they read reviews by other customers, what items they ultimately buy, and what items are bought together. They can then use this information to determine which other products to suggest to customers who buy similar items, to determine which advertisements to run in the margins, and to determine which items are the most popular so these items come up first in a search.

3. **Parking lots.** The owners of the parking garage can advertise about the availability of parking. They can also communicate with businesses about hours when more spots are available and when they should encourage more business.

4. **Satellites and global climate change.** This rise and fall of temperature and water levels can help in planning for future problems and guide public policy to protect our safety.

Section 1.2

5. **Super Bowl.** When collecting data about the Super Bowl, the games themselves are the *Who.*

6. **Nobel laureates.** Each year is a case, holding all of the information about that specific year. Therefore, the year is the *Who.*

7. **Health records.** The sample is about 5,000 people, and the population is all residents of the United States of America. The *Who* is the selected subjects and the *What* includes medical, dental, and physiological measurements and laboratory test results.

8. **Facebook.** The *Who* is the 350 million photos. The *What* might be information about the photos, for example: file format, file size, time and date when uploaded, people and places tagged, and GPS information.

Section 1.3

9. Grade level.

a) If we are, for example, comparing the percentage of first-graders who can tie their own shoes to the percentage of second-graders who can tie their own shoes, grade-level is treated as categorical. It is just a way to group the students. We would use the same methods if we were comparing boys to girls or brown-eyed kids to blue-eyed kids.

b) If we were studying the relationship between grade-level and height, we would be treating grade level as quantitative.

10. ZIP codes.

a) ZIP codes are categorical in the sense that they correspond to a location. The ZIP code 14850 is a standardized way of referring to Ithaca, NY.

b) ZIP codes generally increase as the location gets further from the east coast of the United States. For example, one of the ZIP codes for the city of Boston, MA, is 02101. Kansas City, MO, has a ZIP code of 64101, and Seattle, WA, has a ZIP code of 98101. But Honolulu, HI, much further west than Seattle, has a ZIP code of 96701.

11. Gay marriage. The response is a categorical variable.

12. Gay marriage by party. The answer is a quantitative variable.

13. Medicine. The company is studying a quantitative variable.

14. Stress. The researcher is studying a quantitative variable.

Section 1.4

15. Voting and elections. Pollsters might consider whether a person voted previously or whether he or she could name the candidates. Voting previously and knowing the candidates may indicate a greater interest in the election.

16. Weather. Meteorologists can use the models to predict the average temperature ten days in advance and compare their predictions to the actual temperatures.

17. The News. Answers will vary.

18. The Internet. Answers will vary.

19. Gaydar. *Who* – 40 undergraduate women. *What* – Whether or not the women could identify the sexual orientation of men based on a picture. *Population of interest* – All women.

20. Hula-hoops. *Who* – An unknown number of participants. *What* – Heart rate, oxygen consumption, and rating of perceived exertion. *Population of interest* – All people.

21. Bicycle Safety. *Who* – 2,500 cars. *What* – Distance from the passing car to the bicycle (in inches). *Population of interest* – All cars passing bicyclists.

22. Investments. *Who* – 30 similar companies. *What* – 401(k) employee participation rates (in percent). *Population of interest* – All similar companies.

23. **Honesty.** *Who* – Workers who buy coffee in an office. *What* – amount of money contributed to the collection tray. *Population of interest* – All people in honor system payment situations.

24. Blindness. *Who* – 24 patients. *What* – Whether the patient had Stargardt's disease or dry age-related macular degeneration, and whether or not the stem cell therapy was effective in treating the condition. *Population of interest* – All people with these eye conditions.

25. Not-so-diet soda. *Who* – 474 participants. *What* – whether or not the participant drank two or more diet sodas per day, waist size at the beginning of the study, and waist size at the end of the study. *Population of interest* – All people.

26. Molten iron. *Who* – 10 crankshafts at Cleveland Casting. *What* – The pouring temperature (in degrees Fahrenheit) of molten iron. *Population of interest* – All crankshafts at Cleveland Casting.

27. Weighing bears. *Who* – 54 bears. *What* – Weight, neck size, length (no specified units), and sex. *When* – Not specified. *Where* – Not specified. *Why* - Since bears are difficult to weigh, the researchers hope to use the relationships between weight, neck size, length, and sex of bears to estimate the weight of bears, given the other, more observable features of the bear. *How* – Researchers collected data on 54 bears they were able to catch. *Variables* – There are 4 variables; weight, neck size, and length are quantitative variables, and sex is a categorical variable. No units are specified for the quantitative variables. *Concerns* – The researchers are (obviously!) only able to collect data from bears they were able to catch. This method is a good one, as long as the researchers believe the bears caught are representative of all bears with regard to the relationships between weight, neck size, length, and sex.

28. Schools. *Who* – Students. *What* – Age (probably in years, though perhaps in years and months), race or ethnicity, number of absences, grade level, reading score, math score, and disabilities/special needs. *When* – This information must be kept current. *Where* – Not specified. *Why* – Keeping this information is a state requirement. *How* – The information is collected and stored as part of school records. *Variables* – There are seven variables. Race or ethnicity, grade level, and disabilities/special needs are categorical variables. Number of absences (days), age (years?), reading test score, and math test score are quantitative variables. *Concerns* – What tests are used to measure reading and math ability, and what are the units of measure for the tests?

29. Arby's menu. *Who* – Arby's sandwiches. *What* – type of meat, number of calories (in calories), and serving size (in ounces). *When* – Not specified. *Where* – Arby's restaurants. *Why* – These data might be used to assess the nutritional value of the different sandwiches. *How* – Information was gathered from each of the sandwiches on the menu at Arby's. *Variables* – There are three variables. Number of calories (calories) and serving size (ounces) are quantitative variables, and type of meat is a categorical variable.

30. Religious Landscape. *Who* – U.S. Adults. *What* – Age, Gender, Education, Marital Status, Belief in God (certain, fairly certain, not certain, don't know, do not believe in God), and Frequency of Prayer (daily, weekly, monthly, seldom/never). *When* –2014. *Where* – United States. *Why* – The information was gathered for presentation in the Pew Research Center 2014 Religious Landscape study. *How* – Not specified. *Variables* – There are six variables. Gender, Marital status, Belief in God, Frequency of Prayer are categorical variables; Age (years) is a quantitative variable. Education could be quantitative (measured in years) or categorical if specified in levels rather than number of years.

31. E-bikes. *Who* – Electric bicycles. *What* – motor size (in watts), maximum speed (mph), wheel base (mm), brand name, whether the battery can be removed for security. *When* – May 2020. *Where* – no location specified. *Why* – *Bicycling* magazine review. *How* – Not specified. *Variables* – There are three quantitative variables: Motor size, maximum speed, wheel base. There are two categorical variables: Brand and whether the battery can be removed.

32. Flowers. *Who* – 385 species of flowers. *What* – Date of first flowering (in days). *When* – Not specified. *Where* – Southern England. *Why* – The researchers believe that this indicates a warming of the overall climate. *How* – Not specified. *Variables* – Date of first flowering is a quantitative variable. *Concerns* - Hopefully, date of first flowering was measured in days from January 1, or some other convention, to avoid problems with leap years.

33. Herbal medicine. *Who* – experiment volunteers. *What* – herbal cold remedy or sugar solution, and cold severity (0 to 5 scale). *When* – Not specified. *Where* – Major pharmaceutical firm. *Why* – Scientists were testing the efficacy of an herbal compound on the severity of the common cold. *How* – The scientists set up a controlled experiment. *Variables* – There are two variables. Type of treatment (herbal or sugar solution) is categorical, and severity rating is quantitative. *Concerns* – The severity of a cold seems subjective and difficult to quantify. Also, the scientists may feel pressure to report negative findings about the herbal product.

34. Vineyards. *Who* – American vineyards. *What* – Size of vineyard (in acres), number of years in existence, state, varieties of grapes grown, average case price (in dollars), gross sales (probably in dollars), and percent profit. *When* – Not specified. *Where* – United States. *Why* – Business analysts hoped to provide information that would be helpful to producers of American wines. *How* – Not specified. *Variables* – There are five quantitative variables and two categorical variables. Size of vineyard, number of years in existence, average case price, gross sales, and percent profit are quantitative variables. State and variety of grapes grown are categorical variables.

35. Streams. *Who* – Streams. *What* – Name of stream, substrate of the stream (limestone, shale, or mixed), acidity of the water (measured in pH), temperature (in degrees Celsius), and BCI (unknown units). *When* – Not specified. *Where* – Upstate New York. *Why* – Research is conducted for an Ecology class. *How* – Not specified. *Variables* – There are five variables. Name and substrate of the stream are categorical variables, and acidity, temperature, and BCI are quantitative variables.

36. Fuel economy. *Who* – Every model of automobile in the United States. *What* – Vehicle manufacturer, vehicle type, weight (probably in pounds), horsepower (in horsepower), and gas mileage (in miles per gallon) for city and highway driving. *When* – This information is collected currently. *Where* – United States. *Why* – The Environmental Protection Agency uses the information to track fuel economy of vehicles. *How* – The data are collected from the manufacturer of each model. *Variables* – There are six variables. City mileage, highway mileage, weight, and horsepower are quantitative variables. Manufacturer and type of car are categorical variables.

37. Dogs Detecting Coronavirus. *Who* – Dogs trained to identify odors. *What* – Whether dogs can be trained to identify COVID-19 by odor. *When* – 2020. *Where* – Helsinki. *Why* – Possible way to identify COVID patients. *How* – Not specified. *Variables* – There is 1 single variable. Whether dogs could identify the scent is a categorical variable.

38. Walking in circles. *Who* – 32 volunteers. *What* – Sex, height, handedness, the number of yards walked before going out of bounds, and the side of the field on which the person walked out of bounds. *When* – Not specified. *Where* – Not specified. *Why* – The researcher was interested in whether people walk in circles when lost. *How* – Data were collected by observing the people on the field, as well as by measuring and asking the participants. *Variables* – There are five variables. Sex, handedness, and side of the field are categorical variables. Height and number of yards walked are quantitative variables.

39. Kentucky Derby 2020. *Who* – Kentucky Derby races. *What* – Year, winner, jockey, trainer, owner, and time (in minutes, seconds, and hundredths of a second). *When* – 1875 – 2020. *Where* – Churchill Downs, Louisville, Kentucky. *Why* – Not specified. To examine the trends in the Kentucky Derby? *How* – Official statistics are kept for the race each year. *Variables* – There are six variables. Winner, jockey, trainer and owner are categorical variables. Year and duration are quantitative variables.

40. Indy 2020. *Who* – Indy 500 races. *What* – Year, driver, time (in minutes, seconds, and hundredths of a second), and speed (in miles per hour). *When* – 1911 – 2020. *Where* – Indianapolis, Indiana. *Why* – Not specified. To examine the trends in Indy 500 races? *How* – Official statistics are kept for the race every year. *Variables* – There are four variables. Driver is a categorical variable. Year, time, and speed are quantitative variables.

41. Kentucky Derby 2020 on the computer.

a) Fonso was the winning horse in 1880.

b) The length of the race changed in 1896, from 1.5 miles to 1.25 miles.

c) The winning time in 1974 was 124 seconds.

d) Secretariat ran the Derby in under 2 minutes in 1973, as did Monarchos in 2001. (Northern Dancer ran in exactly 2 minutes in 1964.)

42. Indy 500 2020 on the computer.

a) The average speed of the winner in 1920 was 88.619 miles per hour.

b) Bill Vukovich won the Indy 500 twice in the 1950s.

c) There were only six Indy 500 races in the 1940s.

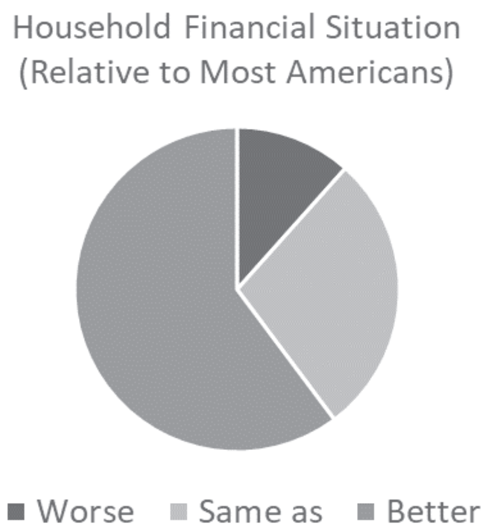
Chapter 2 – Displaying and Describing Data

Section 2.1

1. Financial situations of most Americans.

Live comfortably	0.0420
Meet their basic expenses with a little left over for extras	0.3863
Just meet their basic expenses	0.4794
Don't even have enough to meet basic expenses	0.0864
Don't Know/No Answer	0.0060

2. Comparing financial situations.



3. Movie genres.

a) A pie chart seems appropriate from the movie genre data. Each movie has only one genre, and the list of all movies constitute a "whole".

b) "Documentary" is the least common genre, with only 10 movies. It has the smallest region in the chart.

4. Movie ratings.

a) A pie chart seems appropriate for the movie rating data. Each movie has only one rating, and the list of all movies constitute a "whole". The percentages of each rating are different enough that the pie chart is easy to read, with the exception of the NC-17 category (there is less than 0.15% of the movies in this category).

b) The most common rating is R. It has the largest region on the chart.

5. Movie ratings.

i) C **ii)** A **iii)** D **iv)** B