PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

often leads to death from suicide or organ failure. A Canadian study surveyed 691 adolescent girls diagnosed with anorexia nervosa and their families.[13] Figure 1.9 is a histogram of the distribution of age at onset of anorexia for the 691 girls in the study (note that, if the onset occurred on the 12th birthday, for example, the onset is reported as 12.01 for the purpose of the histogram, to be coherent with customary age descriptions). Describe the shape of this distribution. Within which class does the midpoint of the distribution lie?

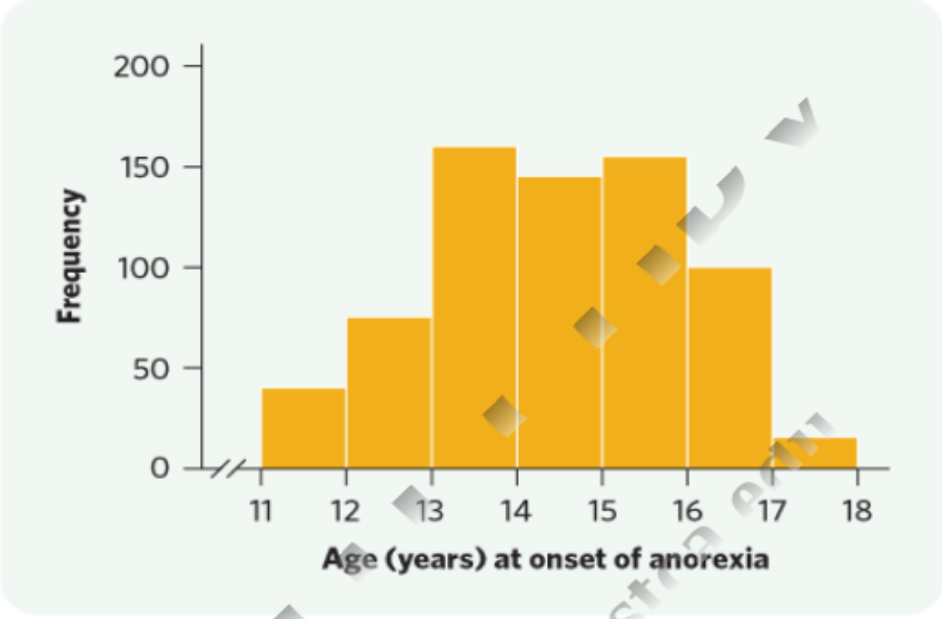

Figure 1.9
Baldi/Moore, *The Practice of Statistics in the Life Sciences*, 4e, © 2018 W. H. Freeman and Company

FIGURE 1.9 Histogram of age (in years) at onset of anorexia nervosa for 691 Canadian girls diagnosed with the disorder. The first class includes 11-year-old girls but excludes 12-year-olds.

PhotoStock-Israel/Getty Images

PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

often leads to death from suicide or organ failure. A Canadian study surveyed 691 adolescent girls diagnosed with anorexia nervosa and their families.[13] Figure 1.9 is a histogram of the distribution of age at onset of anorexia for the 691 girls in the study (note that, if the onset occurred on the 12th birthday, for example, the onset is reported as 12.01 for the purpose of the histogram, to be coherent with customary age descriptions). Describe the shape of this distribution. Within which class does the midpoint of the distribution lie?

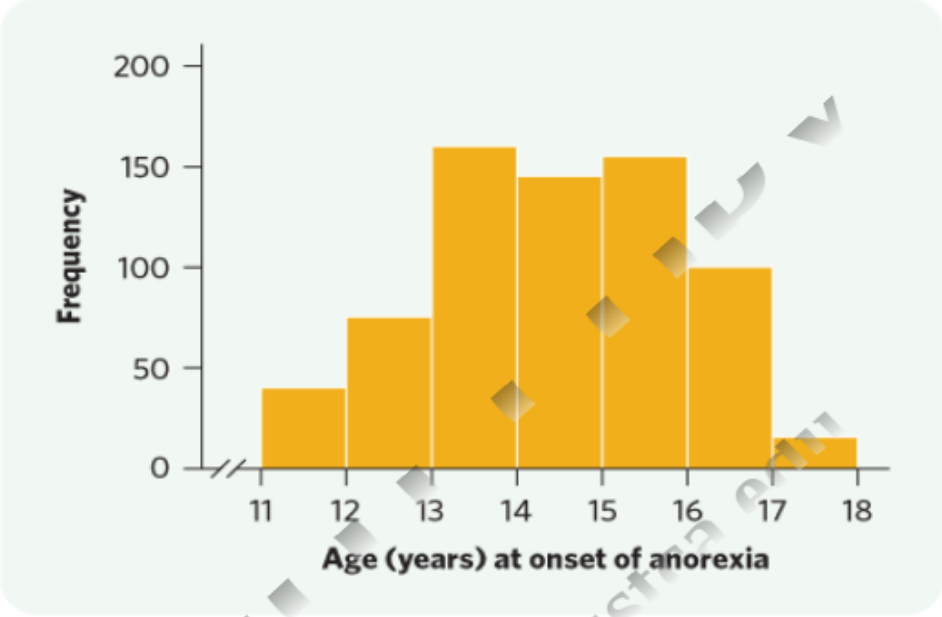

Figure 1.9
Baldi/Moore, *The Practice of Statistics in the Life Sciences*, 4e, © 2018 W. H. Freeman and Company

FIGURE 1.9 Histogram of age (in years) at onset of anorexia nervosa for 691 Canadian girls diagnosed with the disorder. The first class includes 11-year-old girls but excludes 12-year-olds.

PhotoStock-Israel/Getty Images

PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

QUANTITATIVE VARIABLES:
Dotplots

Histograms are not the only graphical display of quantitative distributions. *Dotplots* are also commonly used to diplay the distribution of quantitative data, especially for small data sets. They have the added advantage of displaying the **raw data;** that is, they show each one of the values in the data set.

DOTPLOT

To make a **dotplot:**

1. Sort the data set and plot each observation according to its numerical value along a labeled scaled axis.
2. Identical observations are typically stacked.

EXAMPLE 1.9 **Making a dotplot: sharks**

In Example 1.5 we saw how the body lengths of 44 great white sharks can be plotted in a histogram to more easily examine their distribution. Any raw data that can be displayed in a histogram can also be displayed in a dotplot, and vice versa. To make a dotplot of the shark data, create a one-dimensional graph with shark length on the axis (in feet). The axis may be horizontal or vertical, as desired, because there is only one axis. Figure 1.10(a) shows one dotplot of this data set. Here, data points with the same or very similar numerical values are shown as stacked dots. For example, two sharks had a body length of 18.7 feet and they are shown as two stacked dots. The sharks with body lengths of 12.1 and 12.2 feet are also shown as two stacked dots.

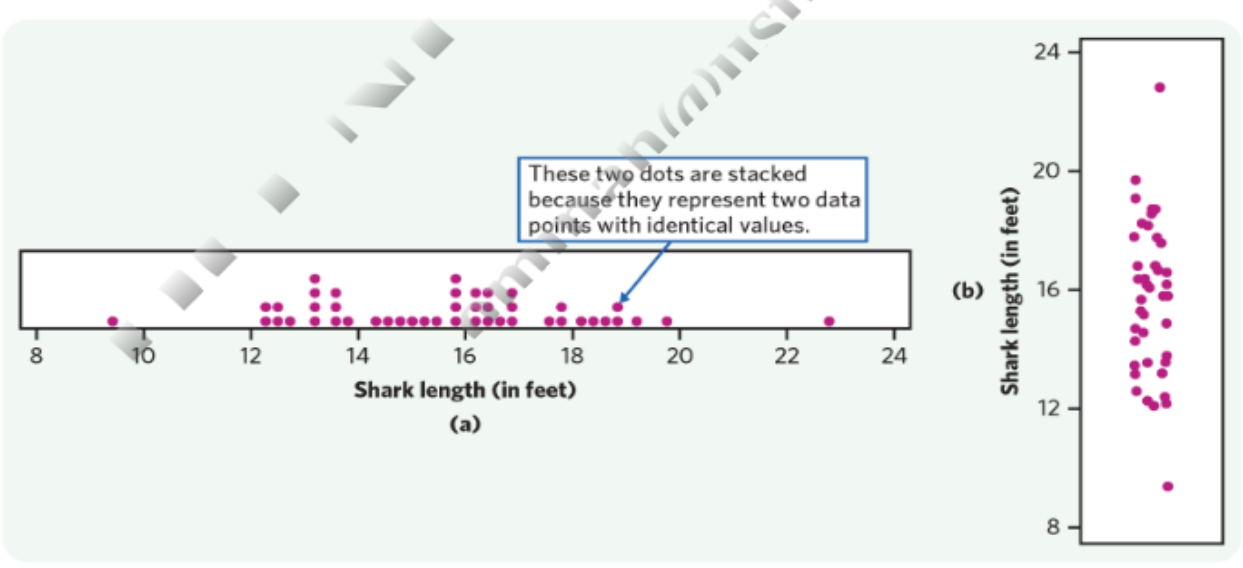

Figure 1.10
Baldi/Moore, *The Practice of Statistics in the Life Sciences*, 4e, © 2018 W. H. Freeman and Company

FIGURE 1.10 Dotplot of body length (in feet) for 44 great white sharks. (a) This dotplot is scaled horizontally, with similar values shown as stacked dots. (b) This dotplot is scaled vertically, with similar values distinguished by adding jitter between the dots.

Like a histogram, the dotplot shows the shape, center, and spread of the distribution as well as potential outliers. However, because the dotplot is one-dimensional, the distribution's shape and center are indicated by the density of dots rather than the height of the histogram bar. Notice how the dots in Figure 1.10(a) are

PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

fairly evenly spread in a roughly symmetric fashion, and that the two mild outliers stand apart from the rest of the data.

Dotplots with neatly stacked dots are easy to draw by hand for reasonably small data sets. When using statistical software, you will typically have the option of adding jitter between dots with similar numerical values so that they can be plotted at their exact location on the scaled axis while still being easily distinguishable, as shown in Figure 1.10(b).

APPLY YOUR KNOWLEDGE

1.9 California pertussis epidemic. In Example 1.7 we used a histogram to display the rates of pertussis cases per 100,000 persons in each of the 59 California counties for all of 2014. Create a dotplot of the data, by hand or using software. Compare your dotplot to the histogram in Figure 1.7 and explain the differences and similarities between the two plots.

1.10 Glucose levels. People with diabetes must monitor and control their blood glucose level. The goal is to maintain "fasting plasma glucose" between about 90 and 130 milligrams per deciliter (mg/dl). A study planned to compare the effectiveness of group instruction versus individual instruction on diabetes control. Here are the fasting plasma glucose levels for 18 diabetic patients who were given group instruction and 16 diabetic patients who were given individual instruction, five months after the end of instruction:[14]

Group instruction									
141	158	112	153	134	95	96	78	148	172
200	271	103	172	359	145	147	255		
Individual instruction									
128	195	188	158	227	198	163	164	159	128
283	226	223	221	220	160				

a. Make one dotplot containing both data sets. Software produces comparative dotplots easily. For handmade graphs, create a dotplot with one horizontal axis but two parallel lines, one over the other. Use one line to place the group instruction dots and the other line to place the individual instruction dots.

b. Describe the main features of each distribution. Are there any outliers?

c. How well is each group, as a whole, achieving the goal of controlling glucose level? How do the two groups differ with respect to this goal?

STATS IN YOUR WORLD

FLORENCE NIGHTINGALE

Florence Nightingale (1820–1910) dedicated her life to promoting and improving the field of nursing. She recorded data and displayed her findings in graphs that policymakers could understand. During the Crimean war, she showed that more soldiers died because of poor hospital conditions than from battle wounds. She also documented how casualties drastically dropped after sanitary conditions were improved. Nightingale was the first woman elected to the Royal Statistical Society, honoring her outstanding contributions.

PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

of the data.

Dotplots with neatly stacked dots are easy to draw by hand for reasonably small data sets. When using statistical software, you will typically have the option of adding jitter between dots with similar numerical values so that they can be plotted at their exact location on the scaled axis while still being easily distinguishable, as shown in Figure 1.10(b).

APPLY YOUR KNOWLEDGE

1.9 California pertussis epidemic. In Example 1.7 we used a histogram to display the rates of pertussis cases per 100,000 persons in each of the 59 California counties for all of 2014. Create a dotplot of the data, by hand or using software. Compare your dotplot to the histogram in Figure 1.7 and explain the differences and similarities between the two plots.

1.10 Glucose levels. People with diabetes must monitor and control their blood glucose level. The goal is to maintain "fasting plasma glucose" between about 90 and 130 milligrams per deciliter (mg/dl). A study planned to compare the effectiveness of group instruction versus individual instruction on diabetes control. Here are the fasting plasma glucose levels for 18 diabetic patients who were given group instruction and 16 diabetic patients who were given individual instruction, five months after the end of instruction:[14]

Group instruction									
141	158	112	153	134	95	96	78	148	172
200	271	103	172	359	145	147	255		
Individual instruction									
128	195	188	158	227	198	163	164	159	128
283	226	223	221	220	160				

a. Make one dotplot containing both data sets. Software produces comparative dotplots easily. For handmade graphs, create a dotplot with one horizontal axis but two parallel lines, one over the other. Use one line to place the group instruction dots and the other line to place the individual instruction dots.

b. Describe the main features of each distribution. Are there any outliers?

c. How well is each group, as a whole, achieving the goal of controlling glucose level? How do the two groups differ with respect to this goal?

STATS IN YOUR WORLD

FLORENCE NIGHTINGALE

Florence Nightingale (1820–1910) dedicated her life to promoting and improving the field of nursing. She recorded data and displayed her findings in graphs that policymakers could understand. During the Crimean war, she showed that more soldiers died because of poor hospital conditions than from battle wounds. She also documented how casualties drastically dropped after sanitary conditions were improved. Nightingale was the first woman elected to the Royal Statistical Society, honoring her outstanding contributions.

PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

130 milligrams per deciliter (mg/dl). A study planned to compare the effectiveness of group instruction versus individual instruction on diabetes control. Here are the fasting plasma glucose levels for 18 diabetic patients who were given group instruction and 16 diabetic patients who were given individual instruction, five months after the end of instruction:[14]

Group instruction									
141	158	112	153	134	95	96	78	148	172
200	271	103	172	359	145	147	255		
Individual instruction									
128	195	188	158	227	198	163	164	159	128
283	226	223	221	220	160				

a. Make one dotplot containing both data sets. Software produces comparative dotplots easily. For handmade graphs, create a dotplot with one horizontal axis but two parallel lines, one over the other. Use one line to place the group instruction dots and the other line to place the individual instruction dots.

b. Describe the main features of each distribution. Are there any outliers?

c. How well is each group, as a whole, achieving the goal of controlling glucose level? How do the two groups differ with respect to this goal?

FLORENCE NIGHTINGALE

Florence Nightingale (1820–1910) dedicated her life to promoting and improving the field of nursing. She recorded data and displayed her findings in graphs that policymakers could understand. During the Crimean war, she showed that more soldiers died because of poor hospital conditions than from battle wounds. She also documented how casualties drastically dropped after sanitary conditions were improved. Nightingale was the first woman elected to the Royal Statistical Society, honoring her outstanding contributions.

PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

TIME PLOTS

Many variables are measured at intervals over time. We might, for example, measure the height of a growing child or water precipitation at the end of each month. In these examples, our main interest is in change over time. To display change over time, make a *time plot*.

TIME PLOT

A **time plot** of a variable plots each observation against the time at which it was measured. Always put time on the horizontal scale of your plot and the variable you are measuring on the vertical scale. Connecting the data points by lines helps emphasize any change over time.

EXAMPLE 1.10 Water salinity in the Shark River

The Shark River runs through the southwestern portion of Everglades National Park in Florida and flows into the Gulf of Mexico. The U.S. Geological Survey closely monitors this important natural habitat. Figure 1.11 is a time plot of water salinity at the Gunboat Island station on the Shark River over a seven-day period in the fall of 2009. Salinity was recorded every 15 minutes from 12 A.M. (midnight) of September 30 to 12 A.M. (midnight) of October 7.[15]

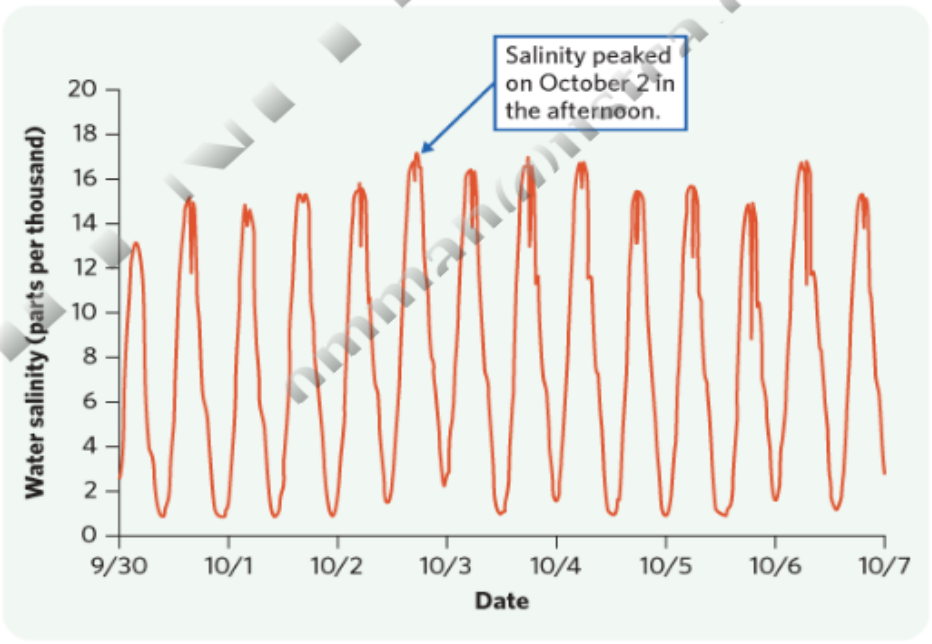

Figure 1.11
Baldi/Moore, *The Practice of Statistics in the Life Sciences*, 4e, © 2018 W. H. Freeman and Company

FIGURE 1.11 Time plot of Shark River water salinity in Everglades National Park over a seven-day period in 2009. The daily cycles reflect the influence of the tides from the Gulf of Mexico, into which the Shark River discharges.

When you examine a time plot, look once again for an overall pattern and for strong deviations from the pattern. Figure 1.11 shows strong **cycles,** representing regular up-and-down movements in water salinity. The cycles show the effects of the ocean tides on the salinity of the Shark River water. Water salinity in the river is highest twice a day, every day, coinciding with the high tides. This cyclical pattern is very clear and consistent, despite some variations in salinity from one day to the next.

PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

EXAMPLE 1.11 Global warming

Global warming is an important planet-wide ecological issue that has been hotly debated. Table 1.1 lists the annual global temperature anomaly (in degrees Celsius) from 1880 to 2012 based on data from recording stations around the world.[16] Individual annual temperature anomalies are computed locally by comparing the local annual sea surface temperature average with the local temperature reference (the 1951–1980 average). Both graphs in Figure 1.12 describe these data.

TABLE 1.1 Annual global temperature anomaly (degrees Celsius)

Year	Anomaly	Year	Anomaly	Year	Anomaly	Year	Anomaly	Year	Anomaly	Year	Anomaly
1880	−0.25	1903	−0.31	1926	−0.01	1949	−0.06	1972	0	1995	0.38
1881	−0.19	1904	−0.34	1927	−0.13	1950	−0.15	1973	0.14	1996	0.29
1882	−0.22	1905	−0.24	1928	−0.11	1951	−0.04	1974	−0.08	1997	0.39
1883	−0.23	1906	−0.2	1929	−0.24	1952	0.03	1975	−0.05	1998	0.56
1884	−0.3	1907	−0.38	1930	−0.06	1953	0.11	1976	−0.16	1999	0.32
1885	−0.3	1908	−0.34	1931	−0.01	1954	−0.1	1977	0.13	2000	0.33
1886	−0.25	1909	−0.35	1932	−0.06	1955	−0.1	1978	0.01	2001	0.48
1887	−0.35	1910	−0.33	1933	−0.17	1956	−0.17	1979	0.08	2002	0.56
1888	−0.26	1911	−0.33	1934	−0.05	1957	0.07	1980	0.18	2003	0.54
1889	−0.15	1912	−0.34	1935	−0.1	1958	0.08	1981	0.26	2004	0.48
1890	−0.36	1913	−0.32	1936	−0.04	1959	0.06	1982	0.05	2005	0.62
1891	−0.28	1914	−0.15	1937	0.08	1960	−0.01	1983	0.26	2006	0.54
1892	−0.32	1915	−0.09	1938	0.11	1961	0.08	1984	0.09	2007	0.56
1893	−0.31	1916	−0.3	1939	0.03	1962	0.04	1985	0.05	2008	0.44
1894	−0.33	1917	−0.4	1940	0.05	1963	0.08	1986	0.12	2009	0.59
1895	−0.27	1918	−0.32	1941	0.11	1964	−0.21	1987	0.26	2010	0.66
1896	−0.16	1919	−0.2	1942	0.04	1965	−0.11	1988	0.31	2011	0.54
1897	−0.12	1920	−0.18	1943	0.1	1966	−0.03	1989	0.19	2012	0.56
1898	−0.24	1921	−0.13	1944	0.2	1967	0	1990	0.38		
1899	−0.17	1922	−0.24	1945	0.07	1968	−0.04	1991	0.35		
1900	−0.1	1923	−0.2	1946	−0.04	1969	0.08	1992	0.12		
1901	−0.15	1924	−0.21	1947	0.01	1970	0.03	1993	0.14		
1902	−0.27	1925	−0.16	1948	−0.04	1971	−0.1	1994	0.23		

PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

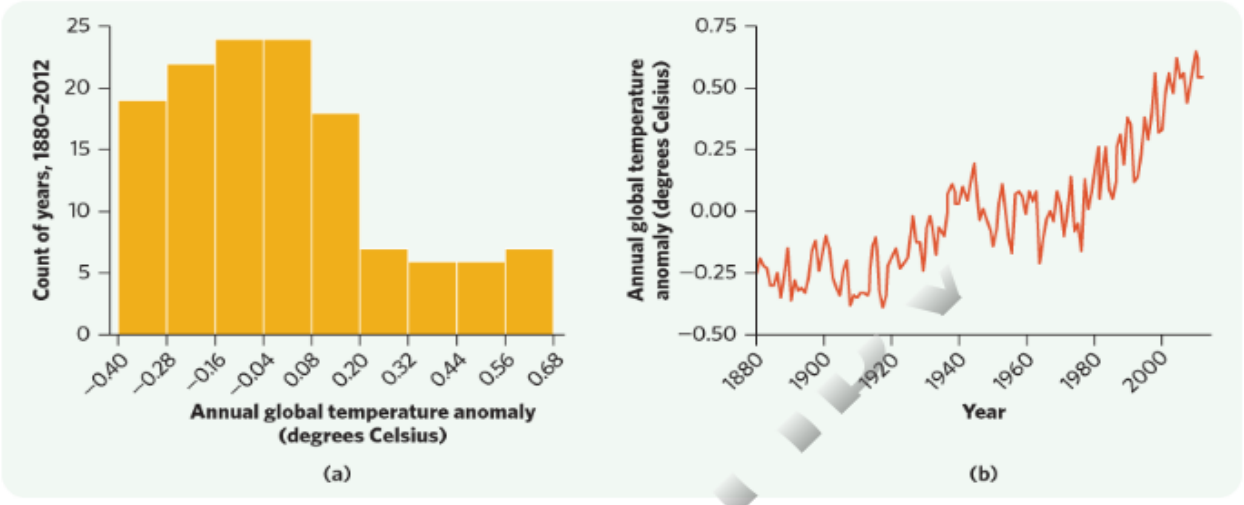

Figure 1.12
Baldi/Moore, *The Practice of Statistics in the Life Sciences*, 4e, © 2018 W. H. Freeman and Company

FIGURE 1.12 Histogram (a) and time plot (b) of the annual global temperature anomaly over the 133 years from 1880 to 2012. Data are from Table 1.1. Anomalies are relative to the 1951–1980 reference period.

© Eric Lefranc/Biosphoto

The histogram in Figure 1.12(a) shows the distribution of annual global temperature anomalies. The histogram is unimodal (single-peaked) and skewed to the right, with a center located at approximately −0.04 to 0.08 degrees Celsius. We might think that these data show just chance year-to-year fluctuations in global temperature anomalies, with a small proportion of larger anomalies.

Figure 1.12(b) is a time plot of the same data. For example, the first point lies above 1880 on the "Year" scale at height −0.25, the global temperature anomaly for 1880. This time plot tells a more interesting story than the histogram. There is a great deal of year-to-year variation, but there is also a clear increasing **trend** over time. That is, there is a long-term rise in annual global temperature anomalies. A trend in a time plot is a long-term upward or downward movement over time. The trend in the temperature data reflects a climate change: Global temperatures have been rising fairly steadily for more than a century, despite substantial year-to-year variations.

PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

DISCUSSION (Mis)adventures in data entry

Data are at the center of statistics, but how do you record data so that they can be analyzed easily? And what are some issues you should be concerned about? This discussion provides a few pointers. Failure to pay attention to these issues can lead to disastrous results, ranging from having to redo your whole analysis, to needing to collect new data or, worse yet, publishing incorrect findings and having to rescind them later.

Keeping detailed records

Scientific inquiries must be documented and you will be required to keep clear, detailed records of your findings. This is a necessary safeguard against dishonest conduct, but it is also a great help in data entry and proofing. To avoid accusations of data falsification or disputes over who discovered something first, using bound notebooks with page numbers and written dates to record data is good practice. You might also keep photographs of your findings or actual specimens. The discussion in Chapter 2 on dealing with outliers highlights how the first step after identifying an outlier is to check your data records and stored evidence to see if the outlier could have been a typographical or other simple data entry error.

Along with the actual data, you should record anything relevant to how the data were obtained, measured, or computed. For example, are values of weight recorded in grams, kilograms, ounces, or pounds, and are they self-reported (common in telephone surveys) or actually measured with a scale? If you computed body mass index from values of weight and height, exactly how did you perform that computation and did it involve rounding numerical values? This step will help you communicate your findings after your analysis is done, and it will be very useful if you ever find puzzling values during the course of your analysis.

Organizing data for use with statistical software

Whether you write your scientific findings on paper or record them electronically, at some point you will want to use statistical software for graphing and statistical analysis. Spreadsheet programs such as Microsoft Excel or LibreOffice Calc are commonly used for this purpose because they are widely available and because they allow data, notes, and even pictures to be saved into one convenient space. You can think of this approach as the electronic equivalent of your paper-based notebook. Proper statistical analysis, however, often requires that your data be organized in very specific ways.

Statistical software packages differ quite a bit in terms of which data format they require, sometimes even requiring different formats for different analyses. Let's just say that no one in the software business seems to have given serious consideration to the painful experiences of the users…Statisticians, however, generally agree that the best way to record your data is to use one row for each individual and one column for each variable. The table in Example 1.1 follows this format: Each tree shrew is given its own row, and every variable recorded about these tree shrews is given its own column (including a unique identifier, here the animal's given name).

Why is the "one individual, one row" approach optimal for data analysis? Let's consider the alternatives. (1) The researchers in Example 1.1 probably had information about each animal's sex and age before they obtained the data on paw usage, so it might have been tempting to use a different spreadsheet for the animals' personal information and for the experimental findings. However, this would make statistical analysis a lot more difficult if they wanted to, say, compare the male and female shrews. (2) Alternatively, the researchers could have recorded their findings on paw usage separately for the male and the female shrews, creating two different data tables. The challenge, then, would be performing an analysis using all shrews (regardless of sex).

When entering categorical values, be sure to be consistent with your choice of naming convention. Statistical software will consider "m" and "male" two different things, even if you understand that both

PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

mean the shrew is male. Some software packages won't accept text at all, and instead require that you use a numerical code, such as 0 for male and 1 for female. In such a case it would be better to use "Female" rather than "Sex" as the variable name and to record it as 0 for no and 1 for yes, which would be easier to interpret.

Lastly, there will be times when you have to deal with missing data. Maybe a tree shrew died before you collected all the data or it refused to perform one of the food tasks in your study—or perhaps you forgot to write that particular piece of data in your notebook…These things happen. But how do you handle it in your electronic data records? The answer, unfortunately, depends on which statistical software you use. Some applications mark missing data with an asterisk, others with a dot, the letters "NA," or a blank space. This can create serious complications if you need to transfer your data analysis from one software application to another. Statisticians sometimes recommend using a special code for missing data, such as 0 or −99 (using different codes for different causes for the missing data). This is a great strategy if the person recording the data is also the person performing the data analysis, but it can be disastrous if the person running the data analysis is not aware of this convention.

Checking for obvious errors, inconsistencies, and missing values

Data entry does not end with the creation of an electronic file of data values. You need to make sure your data are correct. Even data entry professionals are not 100% accurate. Every now and then, you can expect to face problems like spelling differences (as in the "m" and "male" mentioned earlier for categorical data), typos (notice how the values 9 and 0 are right next to each other on the keyboard's top row, and the values 9 and 6 on the numerical keypad), or a skipped row (this mistake is almost too easy to make). The real issue is catching them.

All of your data values should be realistic or at least plausible biologically. In an online class survey asking for height in inches, a student wrote "6.0" as the answer—this is definitely *not* realistic. If you do not catch this obvious error, all your analyses will be affected. In Chapters 2, 3, and 4, we discuss at length how outliers impact various statistical computations.

The best way to catch a mistake like an impossible height value is to plot your data. Stemplots, dotplots, and histograms are excellent ways to *see* unusual or implausible values. Numerical summaries and boxplots (described in Chapter 2) work, too. Time plots for data collected over a period of time are also useful for revealing events that influenced the data collection process (for example, data values becoming systematically larger after switching to a new measuring instrument or when taken by a different person). If your data are categorical, obtain a summary frequency table. You would see, for example, four values for Sex in such a table if you carelessly recorded sex as either m or male, and f or female.

Missing data can create substantial problems, especially if they go unnoticed. The publicly available Pima Indians diabetes data set, for example, contains biological data for 768 female members of the Pima Indian tribe. The data set is listed as having no missing data values. Yet, careful graphical exploration of the data reveals large numbers of zeros that are not biologically plausible (for example, for values of blood pressure or skinfold thickness). Perhaps even more troubling, it appears that many published analyses using this data set failed to identify the problem and instead treated such zeros as actual biological values.[19]

Don't make the same mistake. If you do not plot your data and carefully check them before running further statistical analyses, you could waste a lot of time and energy and, much worse, come to erroneous conclusions. Be smart. Plot your data!

APPLY YOUR KNOWLEDGE

1.13 Student data. Consider gathering some basic information about your classmates, perhaps during lecture or discussion.

PRINTED BY: Olivia Mah (ommah@usfca.edu). Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

a. You may be interested in the distribution of student heights. Which type of variable is student height? Explain how you would collect the data if the class had 30 or 40 students. Which kind of graph would you use to display the data?

b. You may also be interested in the class gender breakdown. Which type of variable is student gender? Explain how you would collect the data if the class had 30 or 40 students. Which kind of graph would you use to display the data?

c. We know that, overall, men tend to be taller than women. Therefore, it would be a good idea to graph and interpret the student height data separately for male and female students. Would the way you collected the data in parts a and b allow you to do this? How would you need to collect and organize the data to allow a comparison of heights for both genders? Which part of the Discussion on page 26 addresses this issue?

1.14 Temperatures records. NASA's Goddard Institute for Space Studies (GISS) allows you to download historical temperature records for locations worldwide. Go to `http://data.giss.nasa.gov/gistemp/station_data`, scroll down to "Download Station Data," and type "Los Angeles." (You may be interested in looking up your own location as well.) A new page appears with a link to the "Los Angeles California" station data. Click on this link to the next page, which should contain a time plot of the annual mean temperature (in degrees Celsius) for Los Angeles since 1880 and a link to "Download monthly data as text." (If this URL isn't working, you can find the data set on the companion website for this textbook, ex01-14.)

a. Open this data file in your preferred statistical software and create a dotplot of the meteorological annual mean temperatures (last column, "metANN"). Describe the distribution. Is there anything suspicious about the data? What could explain it?

b. Go back to the GISS website and look at the time plot of annual mean temperature. Notice that the line is interrupted in some parts of the graph. With this new insight, what do you think is the most likely explanation for the suspicious pattern in your dotplot? Which part of the Discussion on page 26 addresses this issue?

c. Now clean up the data so that only true temperature values will be plotted. You can do this by editing either the original text file or the data in your statistical software. Take the time to consider how your specific software handles missing data (if an empty cell won't work, see your software's help function).

d. Create both a dotplot and a time plot of the cleaned-up annual mean temperatures. Interpret your graphs and conclude in context.