(a) Read the graph: about what were the highest and lowest daily returns on the stock fund?

(b) Read the graph: the median return was about the same on both investments. About what was the median return?

(c) What is the most important difference between the two distributions?

96. **Income and education level** Each March, the Bureau of Labor Statistics compiles an Annual Demographic Supplement to its monthly Current Population Survey.[44] Data on about 71,067 individuals between the ages of 25 and 64 who were employed full-time were collected in one of these surveys. The boxplots below compare the distributions of income for people with five levels of education. This figure is a variation of the boxplot idea: because large data sets often contain very extreme observations, we omitted the individuals in each category with the top 5% and bottom 5% of incomes. Write a brief description of how the distribution of income changes with the highest level of education reached. Give specifics from the graphs to support your statements.
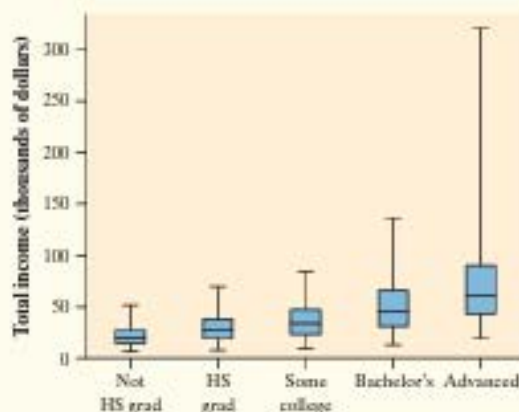
97. **Phosphate levels** The level of various substances in the blood influences our health. Here are measurements of the level of phosphate in the blood of a patient, in milligrams of phosphate per deciliter of blood, made on 6 consecutive visits to a clinic: 5.6, 5.2, 4.6, 4.9, 5.7, 6.4. A graph of only 6 observations gives little information, so we proceed to compute the mean and standard deviation.

(a) Find the standard deviation from its definition. That is, find the deviations of each observation from the mean, square the deviations, then obtain the variance and the standard deviation.

(b) Interpret the value of s_x you obtained in (a).

98. **Feeling sleepy?** The first four students to arrive for a first-period statistics class were asked how much sleep (to the nearest hour) they got last night. Their responses were 7, 7, 9, and 9.

(a) Find the standard deviation from its definition. That is, find the deviations of each observation from the mean, square the deviations, then obtain the variance and the standard deviation.

(b) Interpret the value of s_x you obtained in (a).

(c) Do you think it's safe to conclude that the mean amount of sleep for all 30 students in this class is close to 8 hours? Why or why not?

99. **Shopping spree** The figure displays computer output from Data Desk for data on the amount spent by 50 grocery shoppers.


(a) What would you guess is the shape of the distribution based only on the computer output? Explain.

(b) Interpret the value of the standard deviation.

(c) Are there any outliers? Justify your answer.

100. **C-sections** Do male doctors perform more cesarean sections (C-sections) than female doctors? A study in Switzerland examined the number of cesarean sections (surgical deliveries of babies) performed in a

year by samples of male and female doctors. Here are summary statistics for the two distributions:

	$\bar{x}$	s_x	Min.	Q_1	M	Q_3	Max.	IQR
Male doctors	41.333	20.607	20	27	34	50	86	23
Female doctors	19.1	10.126	5	10	18.5	29	33	19

(a) Based on the computer output, which distribution would you guess has a more symmetrical shape? Explain.

(b) Explain how the IQRs of these two distributions can be so similar even though the standard deviations are quite different.

(c) Does it appear that males perform more C-sections? Justify your answer.

101. **The IQR** Is the interquartile range a resistant measure of spread? Give an example of a small data set that supports your answer.

102. **Measuring spread** Which of the distributions shown has a larger standard deviation? Justify your answer.
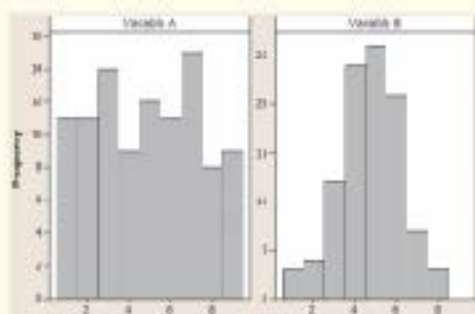
103. **SD contest** This is a standard deviation contest. You must choose four numbers from the whole numbers 0 to 10, with repeats allowed.

(a) Choose four numbers that have the smallest possible standard deviation.

(b) Choose four numbers that have the largest possible standard deviation.

(c) Is more than one choice possible in either (a) or (b)? Explain.

104. **What do they measure?** For each of the following summary statistics, decide (i) whether it could be used to measure center or spread and (ii) whether it is resistant.

(a) $\dfrac{Q_1 + Q_3}{2}$ (b) $\dfrac{\text{Max} - \text{Min}}{2}$

105. **SSHA scores** Here are the scores on the Survey of Study Habits and Attitudes (SSHA) for 18 first-year college women:

154	109	137	115	152	140	154	178	101
103	126	126	137	165	165	129	200	148

and for 20 first-year college men:

108	140	114	91	180	115	126
92	169	146	109	132	75	88
113	151	70	115	187	104	

Do these data support the belief that women have better study habits and attitudes toward learning than men? (Note that high scores indicate good study habits and attitudes toward learning.) Follow the four-step process.

106. **Hummingbirds and tropical flowers** Researchers from Amherst College studied the relationship between varieties of the tropical flower *Heliconia* on the island of Dominica and the different species of hummingbirds that fertilize the flowers.[45] Over time, the researchers believe, the lengths of the flowers and the forms of the hummingbirds' beaks have evolved to match each other. If that is true, flower varieties fertilized by different hummingbird species should have distinct distributions of length.

The table below gives length measurements (in millimeters) for samples of three varieties of *Heliconia*, each fertilized by a different species of hummingbird. Do these data support the researchers' belief? Follow the four-step process.

H. bihai

47.12	46.75	46.80	47.12	46.67	47.43	46.44	46.64
48.07	48.34	48.15	50.26	50.12	46.34	46.94	48.36

H. caribaea red

41.90	42.01	41.93	43.09	41.47	41.69	39.78	40.57
39.63	42.18	40.66	37.87	39.16	37.40	38.20	38.07
38.10	37.97	38.79	38.23	38.87	37.78	38.01	

H. caribaea yellow

36.78	37.02	36.52	36.11	36.03	35.45	38.13	37.10
35.17	36.82	36.66	35.68	36.03	34.57	34.63	

Multiple choice: Select the best answer for Exercises 107 to 110.

107. If a distribution is skewed to the right with no outliers,

(a) mean < median.   (d) mean > median.

(b) mean ≈ median.   (e) We can't tell without

(c) mean = median.   examining the data.

108. You have data on the weights in grams of 5 baby pythons. The mean weight is 31.8 and the standard deviation of the weights is 2.39. The correct units for the standard deviation are
    (a) no units—it's just a number.
    (b) grams.
    (c) grams squared.
    (d) pythons.
    (e) pythons squared.

109. Which of the following is least affected if an extreme high outlier is added to your data?
    (a) Median   (d) Range
    (b) Mean   (e) Maximum
    (c) Standard deviation

110. What are all the values that a standard deviation s_x can possibly take?
    (a) $s_x \geq 0$ (d) $-1 \leq s_x \leq 1$
    (b) $s_x > 0$ (e) Any number
    (c) $0 \leq s_x \leq 1$

Exercises 111 and 112 refer to the following setting. We used CensusAtSchool's "Random Data Selector" to choose a sample of 50 Canadian students who completed a survey in 2007–2008.

Method of communication
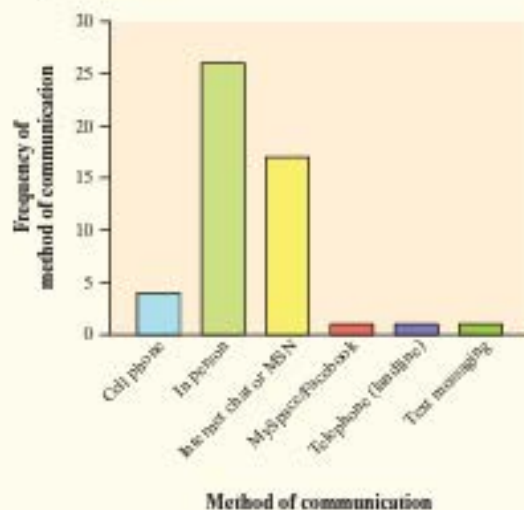
111. **Let's chat** (1.1) The bar graph displays data on students' responses to the question "Which of these methods do you most often use to communicate with your friends?"

(a) Would it be appropriate to make a pie chart for these data? Why or why not?

(b) Jerry says that he would describe this bar graph as skewed to the right. Explain why Jerry is wrong.

112. **How tall are you?** (1.2) Here are the heights (in centimeters) of 50 randomly selected Canadian students who participated in CensusAtSchool in 2007–2008:

166.5	170	178	163	150.5	169	173	169	171	166
190	183	178	161	171	170	191	168.5	178.5	173
175	160.5	166	164	163	174	160	174	182	167
166	170	170	181	171.5	160	178	157	165	187
168	157.5	145.5	156	182	168.5	177	162.5	160.5	185.5

Make an appropriate graph to display these data. Describe the shape, center, and spread of the distribution. Are there any outliers?

113. **Success in college** (1.1) The 2007 Freshman Survey asked first-year college students about their "habits of mind"—specific behaviors that college faculty have identified as being important for student success. One question asked students, "How often in the past year did you revise your papers to improve your writing?" Another asked, "How often in the past year did you seek feedback on your academic work?" The figure is a bar graph comparing male and female responses to these two questions.[46]
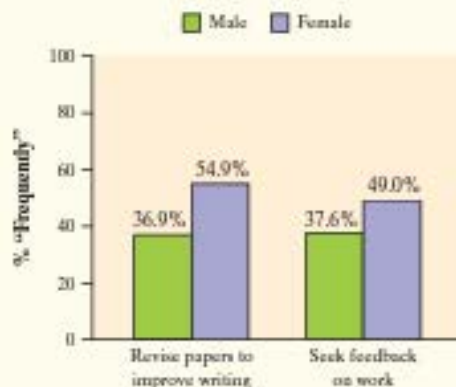

What does the graph tell us about the habits of mind of male and female college freshmen?

Chapter 1   Chapter Review

Statistics is the art and science of data. Data come from measuring one or more variables on some group of individuals. These variables can be categorical or quantitative. Data analysis is a process of describing data using graphs and numerical summaries. This chapter introduced the basic tools of data analysis for categorical and quantitative variables.

Bar graphs and sometimes pie charts are used to display distributions of categorical variables. Counts and percents help summarize the data. To describe the relationship between two categorical variables, start with a two-way table. Then compare appropriate percents.

For quantitative variables, the figure below outlines the data analysis process.
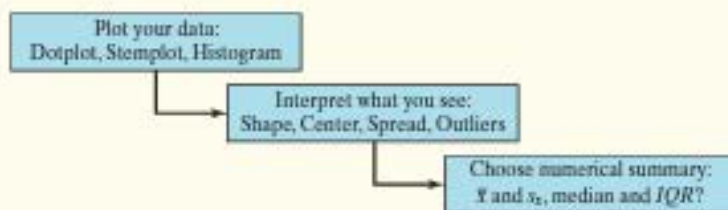
Chapter Review Exercises

These exercises are designed to help you review the important ideas and methods of the chapter. Relevant learning objectives are provided in bulleted form before each exercise.

- Identify the individuals and variables in a set of data.
- Classify variables as categorical or quantitative. Identify units of measurement for a quantitative variable.

R1.1. Hit movies According to the Internet Movie Database, *Avatar* is tops based on box office sales worldwide. The following table displays data on several popular movies.[47]

| Movie	Year	Rating	Time	Genre	Box office
Avatar	2009	PG-13	162	Action	2,714,767,458
Titanic	1997	PG-13	194	Drama	1,835,300,000
The Lord of the Rings: The Return of the King	2003	PG -13	201	Action	1,129,219,252
Pirates of the Caribbean: Dead Man's Chest	2006	PG-13	151	Action	1,060,332,628
The Dark Knight	2008	PG-13	152	Action	1,001,921,825

(a)   What individuals does this data set describe?

(b)   Clearly identify each of the variables. Which are quantitative? In what units do you think they are measured?

(c)   Describe the individual in the highlighted row.

- Make a bar graph of the distribution of a categorical variable or, in general, to compare related quantities.
- Recognize when a pie chart can and cannot be used.

R1.2. Movie ratings The movie rating system we use today was first established on November 1, 1968. Back then, the possible ratings were G, PG, R, and X. In 1984, the PG-13 rating was created. And in 1990, NC-17 replaced the X rating. Here is a summary of the ratings assigned to movies between 1968 and 2000: 8% rated G, 24% rated PG, 10% rated PG-13, 55% rated R, and 3% rated NC-17.[48] Make an appropriate graph for displaying these data.

- Identify what makes some graphs deceptive.

R1.3. I'd die without my phone! In a July 2008 survey of over 2000 U.S. teenagers by Harris Interactive, 47% said that "their social life would end or be worsened without their cell phone."[49] One survey question asked the teens how important it is for their phone to have certain features. The figure below displays data on the percent who indicated that a particular feature is vital.
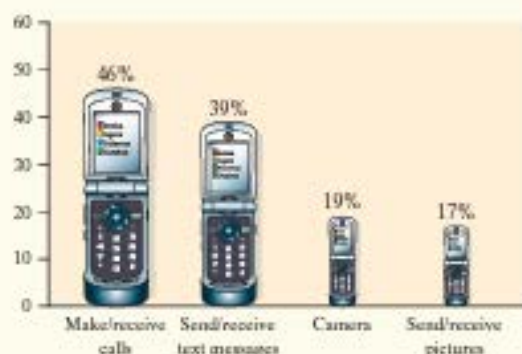
(a) Explain how the graph gives a misleading impression.
(b) Would it be appropriate to make a pie chart to display these data? Why or why not?
(c) Make a graph of the data that isn't misleading.

• From a two-way table of counts, answer questions involving marginal and conditional distributions.

R1.4. Facebook and age Is there a relationship between Facebook use and age among college students? The following two-way table displays data for the 219 students who responded to the survey.[50]

	Facebook User?	
Age	Yes	No
Younger (18–22)	78	4
Middle (23–27)	49	21
Older (28 and up)	21	46

(a) What percent of the students who responded were Facebook users? Is this percent part of a marginal distribution or a conditional distribution? Explain.
(b) What percent of the younger students in the sample were Facebook users? What percent of the Facebook users in the sample were younger students?

• Describe the relationship between two categorical variables by computing appropriate conditional distributions.

• Construct bar graphs to display the relationship between two categorical variables.

R1.5. Facebook and age Use the data in the previous exercise to determine whether there is an association between Facebook use and age. Give appropriate graphical and numerical evidence to support your answer.

• Make a dotplot or stemplot to display small sets of data.

• Describe the overall pattern (shape, center, spread) of a distribution, and identify any major departures from the pattern (like outliers).

R1.6. Density of the earth In 1798, the English scientist Henry Cavendish measured the density of the earth several times by careful work with a torsion balance. The variable recorded was the density of the earth as a multiple of the density of water. Here are Cavendish's 29 measurements:[51]

5.50	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65
5.57	5.53	5.62	5.29	5.44	5.34	5.79	5.10	5.27	5.39
5.42	5.47	5.63	5.34	5.46	5.30	5.75	5.68	5.85	

(a) Present these measurements graphically in a stemplot.
(b) Discuss the shape, center, and spread of the distribution. Are there any outliers?
(c) What is your estimate of the density of the earth based on these measurements? Explain.

• Make a histogram with a reasonable choice of classes.

• Identify the shape of a distribution from a dotplot, stemplot, or histogram as roughly symmetric or skewed.

• Make a boxplot.

R1.7. Guinea pig survival times Here are the survival times in days of 72 guinea pigs after they were injected with infectious bacteria in a medical experiment.[52] Survival times, whether of machines under stress or cancer patients after treatment, usually have distributions that are skewed to the right.

43	45	53	56	56	57	58	66	67	73	74	79
80	80	81	81	81	82	83	83	84	88	89	91
91	92	92	97	99	99	100	100	101	102	102	102
103	104	107	108	109	113	114	118	121	123	126	128
137	138	139	144	145	147	156	162	174	178	179	184
191	198	211	214	243	249	329	380	403	511	522	598

(a) Make a histogram of the data and describe its main features. Does it show the expected right-skew?

(b) Now make a boxplot of the data. Be sure to check for outliers.

- Interpret histograms.

R1.8. Household incomes Rich and poor households differ in ways that go beyond income. Following are histograms that compare the distributions of household size (number of people) for low-income and high-income households.[53] Low-income households had annual incomes less than \$15,000, and high-income households had annual incomes of at least \$100,000.

Household size, low income

Household size, high income
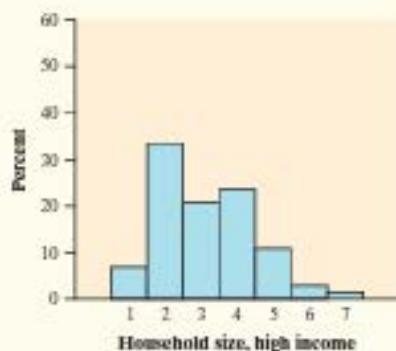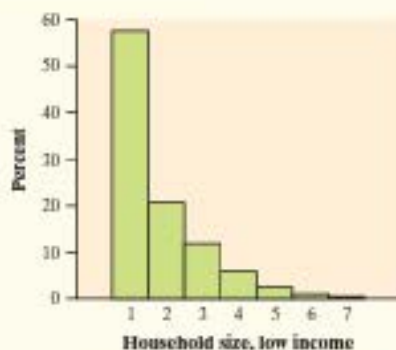
(a) About what percent of each group of households consisted of two people?

(b) What are the important differences between these two distributions? What do you think explains these differences?

Exercises R1.9 and R1.10 refer to the following setting. Do you like to eat tuna? Many people do. Unfortunately, some of the tuna that people eat may contain high levels of mercury. Exposure to mercury can be especially hazardous for pregnant women and small children. How much mercury is safe to consume? The Food and Drug Administration will take action (like removing the product from store shelves) if the mercury concentration in a six-ounce can of tuna is 1.00 ppm (parts per million) or higher.

What is the typical mercury concentration in cans of tuna sold in stores? A study conducted by Defenders of Wildlife set out to answer this question. Defenders collected a sample of 164 cans of tuna from stores across the United States. They sent the selected cans to a laboratory that is often used by the Environmental Protection Agency for mercury testing.[54]

- Interpret the value of a numerical summary in context: mean, median, *IQR*, standard deviation.

- Identify outliers using the 1.5 × *IQR* rule.

- Select the more appropriate measures of center and spread: $\bar{x}$ and s_y or the median and *IQR*.

R1.9. Mercury in tuna A histogram and some computer output provide information about the mercury concentration in the sampled cans (in parts per million, ppm).
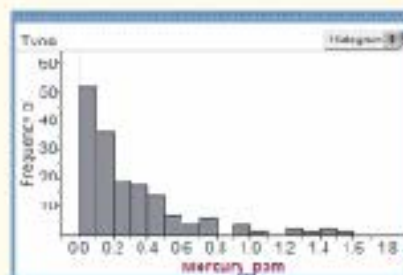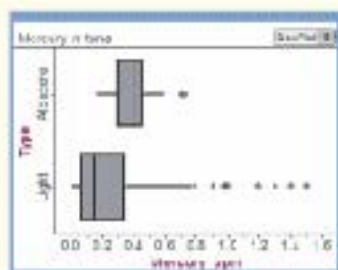

Descriptive Statistics: Mercury_ppm

Variable	N	Mean	StDev	Min
Mercury	164	0.285	0.300	0.012
Variable	Q_1	Med	Q_3	Max
Mercury	0.071	0.180	0.380	1.500

(a) Interpret the standard deviation in context.

(b) Determine whether there are any outliers.

(c) Describe the shape, center, and spread of the distribution.

- Use appropriate graphs and numerical summaries to compare distributions of quantitative variables.

R1.10. Mercury in tuna Is there a difference in the mercury concentration of light tuna and albacore tuna? Use the side-by-side boxplots and the computer output to write a few sentences comparing the two distributions.

Descriptive Statistics: Mercury_ppm

Type	N	Mean	StDev	Min
Albacore	20	0.401	0.152	0.170
Light	144	0.269	0.312	0.012

Type	Q_1	Med	Q_3	Max
Albacore	0.293	0.400	0.460	0.730
Light	0.059	0.160	0.347	1.500

Chapter 1 AP Statistics Practice Test

Section I: Multiple Choice *Select the best answer for each question.*

T1.1. You record the age, marital status, and earned income of a sample of 1463 women. The number and type of variables you have recorded is

(a) 3 quantitative, 0 categorical

(b) 4 quantitative, 0 categorical

(c) 3 quantitative, 1 categorical

(d) 2 quantitative, 1 categorical

(e) 2 quantitative, 2 categorical

T1.2. Consumers Union measured the gas mileage in miles per gallon of 38 vehicles from the same model year on a special test track. The pie chart provides information about the country of manufacture of the model cars tested by Consumers Union. Based on the pie chart, we conclude

(a) Japanese cars get significantly lower gas mileage than cars from other countries.

(b) U.S. cars get significantly higher gas mileage than cars from other countries.

(c) Swedish cars get gas mileages that are between those of Japanese and U.S. cars.

(d) Mercedes, Audi, Porsche, and BMW represent approximately a quarter of the cars tested.

(e) More than half of the cars in the study were from the United States.
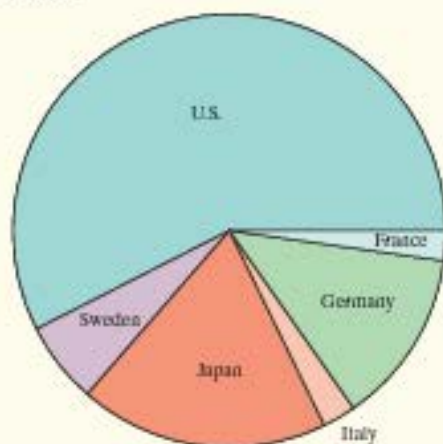
T1.3. Which of the following bar graphs is equivalent to the pie chart in Question T1.2?