INSTRUCTOR'S SOLUTIONS MANUAL

DIACRITECH

ELEMENTARY STATISTICS: PICTURING THE WORLD

SEVENTH EDITION

Ron Larson

Pennsylvania State University
The Behrend College

Betsy Farber

Bucks County Community College

Pearson

Reproduced by Pearson from electronic files supplied by the author.

Publishing as Pearson, 330 Hudson Street, NY NY 10013

ISBN-13: 978-0-13-468359-1
ISBN-10: 0-13-468359-5

CONTENTS

CHAPTER 1

Introduction to Statistics

1.1 AN OVERVIEW OF STATISTICS

1.1 TRY IT YOURSELF SOLUTIONS

1. The population consists of the responses of all ninth to twelfth graders in the United States. The sample consists of the responses of the 1501 ninth to twelfth graders in the survey. The sample data set consists of 1215 ninth to twelfth graders who said leaders today are more concerned with their own agenda than with achieving the overall goals of the organization they serve and 286 ninth to twelfth graders who did not say that.

2a. Population parameter, because the total spent on employees' salaries, $5,150,694, is based on the entire company.

b. Sample statistic, because 43% is based on a subset of the population.

3a. The population consists of the responses of all U.S. adults, and the sample consists of the responses of the 1000 U.S. adults in the study.

b. The part of this study that represents the descriptive branch of statistics involves the statement "three out of four adults will consult with their physician or pharmacist and only 8% visit a medication-specific website [when they have a question about their medication]."

c. A possible inference drawn from the study is that most adults consult with their physician or pharmacist when they have a question about their medication.

1.1 EXERCISE SOLUTIONS

1. A sample is a subset of a population.

2. It is usually impractical (too expensive and/or time consuming) to obtain all the population data.

3. A parameter is a numerical description of a population characteristic. A statistic is a numerical description of a sample characteristic.

4. The two main branches of statistics are descriptive statistics and inferential statistics.

5. False. A statistic is a numerical measure that describes a sample characteristic.

6. True

7. True

8. False. Inferential statistics involves using a sample to draw conclusions about a population.

9. False. A population is the collection of *all* outcomes, responses, measurements, or counts that are of interest.

10. False. A sample statistic can differ from sample to sample.

11. Population, because it is a collection of the salaries of each member of a Major League Baseball team.

12. Population, because it is a collection of the energy collected from all the solar panels on a photo voltaic power plant.

13. Sample, because the collection of the 300 people is a subset of the population of 13,000 people in the auditorium.

14. Population, because it is a collection of the revenue of all the stores at the shopping mall.

15. Sample, because the collection of the 10 patients is a subset of the population of 50 patients at the clinic.

16. Population, because it is a collection of the number of wireless devices in all U.S. households.

17. Population, because it is a collection of all the gamers' scores in the tournament.

18. Sample, because only the age of every fourth person entering the grocery store is recorded.

19. Population, because it is a collection of all the U.S. senators' political parties.

20. Sample, because the collection of the 20 air contamination levels is a subset of the population.

21. Population: Parties of registered voters
Sample: Parties of registered voters who respond to a survey

22. Population: Student donations at a food drive
Sample: Student donations of canned goods

23. Population: Ages of adults in the United States who own automobiles
Sample: Ages of adults in the United States who own Honda automobiles

24. Population: Incomes of home owners in Massachusetts
Sample: Incomes of home owners in Massachusetts with mortgages

25. Population: Collections of the responses of all U.S. adults
Sample: Collection of the responses of the 1020 U.S. adults surveyed
Sample data set: 42% of adults who said they trust their political leaders and 58% who said they did not

26. Population: Collection of fetal tobacco exposure of all infants
Sample: Collection of the fetal tobacco exposure of 203 infants
Sample data set: Infants with fetal tobacco exposure and their focused attention levels

27. Population: Collection of the influenza immunization status of all adults in the United States
Sample: Collection of the influenza immunization status of the 3301 U.S. adults surveyed
Sample data set: 39% of U.S. adults who received an influenza vaccine and 61% who did not

28. Population: Collection of the responses of travelers with pets in the world
Sample: Collection of the responses of the 1100 travelers surveyed with pets
Sample data set: 53% of respondents with pets who said they travel with their pets and 47% who said they did not

29. Population: Collection of the average hourly billing rates of all U.S. law firms
Sample: Collection of the average hourly billing rates for partners of the 159 U.S. law firms surveyed
Sample data set: The average hourly billing rate for partners of 159 U.S. law firms is $604.

30. Population: Collection of plans after high school of all students at a high school
Sample: Collection of plans after high school of 496 students surveyed at a high school
Sample data set: 95% of those surveyed who are planning to go to college and 5% who are not

31. Population: Collection of all U.S. adults
Sample: Collection of the responses of those suffering with chronic pain of the 1029 U.S. adults surveyed
Sample data set: 23% of respondents suffering with chronic pain who were diagnosed with a sleeping disorder and 77% who were not

32. Population: Collection of the responses of all preowned automobile shoppers
Sample: Collection of the responses of the 1254 preowned automobile shoppers surveyed
Sample data set: 5% of respondents shopping for preowned automobiles who bought extended warranties and 95% who did not

33. Population: Collection of all companies listed in the Standard & Poor's 500
Sample: Collection of the responses of the 54 Standard & Poor's 500 companies surveyed
Sample data set: Starting salaries of the 54 companies surveyed

34. Population: Collection of parents of 13- to 17-year-olds
Sample: Collection of responses of 1060 parents of 13- to 17-year-olds surveyed
Sample data set: 636 parents who said they check their teen's social media profile and 424 parents who did not

35. Sample statistic. The value $72,000 is a numerical description of a sample of average salaries

36. Sample statistic. The value 56.3% is a numerical description of a sample of college board members

37. Population Parameter. The 62 surviving passengers out of 97 total passengers is a numerical description of all of the passengers of the Hindenburg that survived.

38. Population parameter. The value 62% is a numerical description of the total number of governors.

39. Sample statistic. The value 7% is a numerical description of a sample of computer users.

40. Population parameter. The value 87% is a numerical description of the total number of voters.

41. Sample statistic. The value 80% is a numerical description of a sample of U.S. adults.

42. Population parameter. The score 20.6 is a numerical description of the ACT scores for all graduates.

43. The statement "23% of those suffering with chronic pain had been diagnosed with a sleep disorder" is an example of descriptive statistics. Using inferential statistics, you may conclude that an association exists between chronic pain and sleep disorders.

44. The statement "5% bought extended warranties" is an example of descriptive statistics. Using inferential statistics, you may conclude that most pre-owned automobile shoppers do not buy extended warranties.

45. Answers will vary.

46. Answers will vary.

47. The inference may incorrectly imply that exercise increases a person's cognitive ability. The study shows a slower decline in cognitive ability, not an increase.

48. The inference may incorrectly imply that obesity trends will continue in future years. Even though the obesity rates have been increasing, that does not mean the rates will continue to increase for eternity.

49. (a) The sample is the results on the standardized test by the participants in the study.

(b) The population is the collection of all the results of the standardized test.

(c) The statement "the closer that participants were to an optimal sleep duration target, the better they performed on a standardized test" is an example of descriptive statistics.

(d) Individuals who obtain optimal sleep will be more likely to perform better on a standardized test then they would without optimal sleep.

1.2 DATA CLASSIFICATION

1.2 TRY IT YOURSELF SOLUTIONS

1. The city names are nonnumerical entries, so these are qualitative data. The city populations are numerical entries, so these are quantitative data.

2. (1) Ordinal, because the data can be put in order.

(2) Nominal, because no mathematical computations can be made.

3. (1) Interval, because the data can be ordered and meaningful differences can be calculated, but it does not make sense to write a ratio using the temperatures.

(2) Ratio, because the data can be ordered, meaningful differences can be calculated, the data can be written as a ratio, and the data set contains an inherent zero.

1.2 EXERCISE SOLUTIONS

1. Nominal and ordinal

2. Ordinal, interval, and ratio

3. False. Data at the ordinal level can be qualitative or quantitative.

4. False. For data at the interval level, you can calculate meaningful differences between data entries. You cannot calculate meaningful differences at the nominal or ordinal levels.

5. False. More types of calculations can be performed with data at the interval level than with data at the nominal level.

6. False. Data at the ratio level can be placed in a meaningful order.

7. Quantitative, because dog weights are numerical measurements.

8. Quantitative, because carrying capacities are numerical measurements.

9. Qualitative, because hair colors are attributes.

10. Qualitative, because student ID numbers are labels.

11. Quantitative, because infant heights are numerical measurements.

12. Qualitative, because mammal species are labels.

13. Qualitative, because the poll responses are attributes.

14. Quantitative, because wait times are numerical measurements.

15. Interval. Data can be ordered and meaningful differences can be calculated, but it does not make sense to say one year is a multiple of another.

16. Ordinal. Data can be arranged in order, but differences between data entries are not meaningful.

17. Nominal. No mathematical computations can be made, and data are categorized using numbers.

18. Ratio. Data can be ordered and meaningful differences can be calculated. A length of 0 means it lasts for 0 seconds. A ratio of two data entries can be formed so that one data entry can be meaningfully expressed as a multiple of another.

19. Ordinal. Data can be arranged in order, but the differences between data entries are not meaningful.

20. Interval. Data can be ordered and meaningful differences can be calculated, but it does not make sense to say one time is a multiple of another.

21. Horizontal: Nominal; Vertical: Ratio

22. Horizontal: Ordinal; Vertical: Ratio

23. Horizontal: Nominal; Vertical: Ratio

24. Horizontal: Interval; Vertical: Ratio

25. (a) Interval (b) Nominal (c) Ratio (d) Ordinal

26. (a) Interval (b) Nominal (c) Interval (d) Ratio

27. Qualitative. Ordinal. Data can be arranged in order, but differences between data entries are not meaningful.

28. Qualitative. Nominal. No mathematical computations can be made, and data are categorized by political party.

29. Qualitative. Nominal. No mathematical computations can be made and data are categorized by region.

30. Quantitative. Interval. Data can be ordered and meaningful differences can be calculated, but it does not make sense to say that one score is a multiple of another.

31. Qualitative. Ordinal. Data can be arranged in order, but the differences between data entries are not meaningful.

32. Quantitative. Ratio. A ratio of two data entries can be formed, so one data entry can be expressed as a multiple of another.

33. An inherent zero is a zero that implies "none." Answers will vary.

34. Answers will vary.

1.3 DATA COLLECTION AND EXPERIMENTAL DESIGN

1.3 TRY IT YOURSELF SOLUTIONS

1. This is an observational study.

2. There is no way to tell why the people quit smoking. They could have quit smoking as a result of either chewing the gum or watching the DVD. The gum and the DVD could be confounding variables. To improve the study, two experiments could be done, one using the gum and the other using the DVD. Or just conduct one experiment using either the gum or the DVD.

3. Sample answer: Assign numbers 1 to 79 to the employees of the company. Use the table of random numbers and obtain 63, 7, 40, 19, and 26. The employees assigned these numbers will make up the sample.

4. (1) The sample was selected by using the students in a randomly chosen class. This is cluster sampling.

(2) The sample was selected by numbering each student in the school, randomly choosing a starting number, and selecting students at regular intervals from the starting number. This is systematic sampling.

1.3 EXERCISE SOLUTIONS

1. In an experiment, a treatment is applied to part of a population and responses are observed. In an observational study, a researcher measures characteristics of interest of a part of a population but does not change existing conditions.

2. A census includes the entire population; a sampling includes only a portion of the population.

3. In a random sample, every member of the population has an equal chance of being selected. In a simple random sample, every possible sample of the same size has an equal chance of being selected.

4. Replication is the repetition of an experiment under the same or similar conditions. Replication is important because it enhances the validity of the results.

5. False. A placebo is a fake treatment.

6. False. A double-blind experiment is used to decrease the placebo effect.

7. False. Using stratified sampling guarantees that members of each group within a population will be sampled.

8. False. A convenience sample is not representative of a population.

9. False. To select a systematic sample, a population is ordered in some way and then members of the population are selected at regular intervals.

10. True

11. Observational study. The study does not apply a treatment to the adults.

12. Experiment. The study applies a treatment (intensive program to lower systolic blood pressure) to the subjects.

13. Experiment. The study applies a treatment (different photographs) to the subjects.

14. Observational study. The study does not apply a treatment to the motorists.

15. Answers will vary. *Sample answer*: Starting at the left-most number in row 6:
28/70/35/17/09/94/45/64/83/96/73/78/
The numbers would be 28,70,35,17,9,94,45,64,83,96,73,78.

16. Answers will vary. *Sample answer*: Starting with the left-most number in row 10:
421/030/278/173/920/562/977/267/812/249/252/
The numbers would be 421,30,278,173,920,562,267,812,249,252.

17. Answers will vary.

18. Answers will vary.

19. (a) The experimental units are the 500 females ages 25 to 45 years old who suffer from migraine headaches. The treatment is the new drug used to treat migraine headaches.

(b) A problem with the design is that the sample is not representative of the entire population because only females ages 25 to 45 were used. To increase validity, use a stratified sample.

(c) For the experiment to be double-blind, neither the subjects nor the company would know whether the subjects are receiving the drug or the placebo.

20. (a) The experimental units are the 31 patients with type 2 diabetes. The treatment is the dietary supplement designed to control metabolism in patients with type 2 diabetes.

(b) A problem with the design is that the sample size is small. The experiment could be replicated to increase validity.

(c) In a placebo-controlled, double-blind experiment, neither the subject nor the experimenter knows whether the subject is receiving a treatment or a placebo. The experimenter is informed after all the data have been collected.

(d) Divide the subjects into age categories and then, within each age group, randomly assign subjects to either the treatment group or the control group.

21. Answers will vary. *Sample answer*: Number the volunteers from 1 to 18. Using the random number table in Appendix B, starting with the left-most number in row 16:
29/55/31/84/32/**13**/63/00/55/29/**02**/79/**18**/**10**/**17**/49/02/77/90/31/50/91/20/93/99
23/50/**12**/26/42/63/**08**/10/81/91/89/42/**06**/78/00/55/13/75/47/**07**/
Treatment group: Maria, Adam, Bridget, Carlos, Susan, Rick, Dan, Mary, and Connie.
Control group: Jake, Mike, Lucy, Ron, Steve, Vanessa, Kate, Pete, and Judy.

22. Answers will vary. *Sample answer*: Using a random number generator:
Treatment group: 1,2,3,4,5,6,7,9,12,15,18,20,22,23,26,27,28,30,31,32,33,34,35,36,37,38,41,42, 44,50,54,63,68,70,73,74,78,80,81,82,85,86,87,88,89
Control group: 8,10,11,13,14,16,17,19,21,24,25,29,39,40,43,45,46,47,48,49,51,52,53,55,56,57, 58,59,60,61,62,64,65,66,67,69,71,72,75,76,77,79,83,84,90.

23. Simple random sampling is used because each employee has an equal chance of being contacted, and all samples of 300 people have an equal chance of being selected. A possible source of bias is that the random sample may contain a much greater percentage of employees from one department than from others.

24. Convenience sampling is used because the students are chosen due to their convenience of location. Bias may enter into the sample because the students sampled may not be representative of the population of students.

25. Cluster sampling is used because the disaster area is divided into grids, and 30 grids are then entirely selected. A possible source of bias is that certain grids may have been much more severely damaged than others.

26. Systematic sampling is used because every tenth person entering the shopping mall is sampled. It is possible for bias to enter the sample if, for some reason, there is a regular pattern to people entering the shopping mall.

27. Stratified sampling is used because a sample is taken from each one-acre subplot (stratum).

28. Simple random sampling is used because each telephone number has an equal chance of being dialed, and all samples of 1012 phone numbers have an equal chance of being selected. The sample may be biased because telephone sampling only samples those individuals who have telephones, who are available, and who are willing to respond.

29. Census, because it is relatively easy to obtain the ages of the 115 residents.

30. Sampling, because the population of subscribers is too large to easily record their favorite movie type. Random sampling would be advised because it would be easy to randomly select subscribers and then record their favorite movie types.

31. The question is biased because it already suggests that eating whole-grain foods improves your health. The question might be rewritten as "How does eating whole-grain foods affect your health?"

32. The question is biased because it already suggests that text messaging while driving increases the risk of a crash. The question might be rewritten as "Does text messaging while driving affect the risk of a crash?"

33. The survey question is unbiased because it does not imply how much exercise is good or bad.

34. The question is biased because it already suggests that the media influences the opinions of voters. The question could be rewritten as "Does the media influence the opinions of voters?"

35. The households sampled represent various locations, ethnic groups, and income brackets. Each of these variables is considered a stratum. Stratified sampling ensures that each segment of the population is represented.

36. *Sample answer:* Observational studies may be referred to as natural experiments because they involve observing naturally occurring events that are not influenced by the study.

37. Answers will vary.

38. Answers will vary.

39. Open Question
Advantage: Allows respondent to express some depth and shades of meaning in the answer. Allows for new solutions to be introduced.
Disadvantage: Not easily quantified and difficult to compare surveys.

Closed Question
Advantage: Easy to analyze results.
Disadvantage: May not provide appropriate alternatives and may influence the opinion of the respondent.

CHAPTER 1 REVIEW EXERCISE SOLUTIONS

1. Population: Collection of the responses of all U.S. adults
Sample: Collection of the responses of the 4787 U.S. adults who were sampled
Sample data set: 15% of adults who use ride-hailing applications and 85% who do not

2. Population: Collection of the opinions on health care reform of all doctors in the St. Louis area
Sample: Collection of the opinions on health care reform of the 83 doctors in the St. Louis area who were sampled
Sample data set: Doctors in the St. Louis area and their opinions on health care reform

3. Population: Collection of the responses of all U.S. adults
Sample: Collection of the responses of the 2223 U.S. adults who were sampled
Sample data set: 62% of adults who would encourage a child to pursue a career as a video game developer or designer and 38% who would not

4. Population: Collection of the responses of all U.S. children and adults ages 16 years and older
Sample: Collection of the responses of the 1601 U.S. children and adults ages 16 and older who were sampled
Sample data set: 48% of children and adults who have visited a public library or a bookmobile over a recent span of 12 months and 52% who did not

5. Population parameter. The value $22.7 million is a numerical description of the total infrastructure-strengthening investments.

6. Sample statistic. The value 29% is a numerical description of a sample of U.S. voters.

7. Population Parameter. The 10 students minoring in physics is a numerical description of all math majors at a university.

8. Sample statistic. The value 30% is a numerical description of a sample of U.S. workers.

9. The statement "62% would encourage a child to pursue a career as a video game developer or designer" is an example of descriptive statistics. An inference drawn from the sample is that a majority of people encourage children to pursue a career as a video game developer or designer.

10. The statement "48% have visited a public library or a bookmobile over a recent span of 12 months" is an example of descriptive statistics. An inference drawn from the sample is that about half of U.S. children and adults ages 16 years and older have visited a public library or a bookmobile over a recent span of 12 months.

11. Quantitative, because ages are numerical measurements.

12. Qualitative, because zip codes are labels for customers.

13. Quantitative, because revenues are numerical measures.

14. Qualitative, because marital statuses are attributes.

15. Interval. The data can be ordered and meaningful differences can be calculated, but it does not make sense to say that 84 degrees is 1.05 times as hot as 80 degrees.

16. Ordinal. The data are qualitative and could be arranged in order of car size.

17. Nominal. The data are qualitative and cannot be arranged in a meaningful order.

18. Ratio. The data are quantitative, and it makes sense to say that $53.2 million is 1.12 times as much as $47.5 million.

19. Experiment. The study applies a treatment (drug to treat hypertension in patients with obstructive sleep apnea) to the subjects.

20. Observational study. The study does not attempt to influence the responses of the subjects and there is no treatment.

21. *Sample answer:* The subjects could be split into male and female and then be randomly assigned to each of the five treatment groups.

22. *Sample answer:* Number the volunteers and then use a random number generator to assign subjects randomly to one of the treatment groups or the control group.

23. Simple random sampling is used because random telephone numbers were generated and called. A potential source of bias is that telephone sampling only samples individuals who have telephones, who are available, and who are willing to respond.

24. Convenience sampling is used because the student sampled a convenient group of friends. The study may be biased toward the opinions of the student's friends.

25. Cluster sampling is used because each district is considered a cluster and every pregnant woman in a selected district is surveyed. A potential source of bias is that the selected districts may not be representative of the entire area.

26. Systematic sampling is used because every third car is stopped. A potential source of bias is that the street the law enforcement officials are using may be near a bar.

27. Stratified sampling is used because the population is divided by grade level and then 25 students are randomly selected from each grade level.

28. Convenience sampling is used because of the convenience of surveying people waiting for their baggage. A potential source of bias is that all of the people just got off an airplane.

29. Answers will vary. *Sample answer*: Sampling, because the population of students at the university is too large for their favorite spring break destinations to be easily recorded. Random sampling would be advised because it would be easy to select students randomly and then record their favorite spring break destination.

CHAPTER 1 QUIZ SOLUTIONS

1. Population: Collection of the school performance of all Korean adolescents
Sample: Collection of the school performance of the 359,264 Korean adolescents in the study

2. (a) Sample statistic. The value 52% is a numerical description of a sample of U.S. adults.

(b) Population Parameter. The 90% of members that approved the contract of the new president is a numerical description of all Board of Trustees members.

(c) Sample statistic. The value 25% is a numerical description of a sample of small business owners.

3. (a) Qualitative, because debit card personal identification numbers are labels and it does not make sense to find differences between numbers.

(b) Quantitative, because final scores are numerical measurements.

4. (a) Ordinal, because badge numbers can be ordered and often indicate seniority of service, but no meaningful mathematical computation can be performed.

(b) Ratio, because horsepower of one car can be expressed as a multiple of another.

(c) Ordinal, because data can be arranged in order, but the differences between data entries make no sense.

(d) Interval, because meaningful differences between years can be calculated, but a zero entry is not an inherent zero.

5. (a) Observational study. The study does not attempt to influence the responses of the subjects and there is no treatment.

(b) Experiment. The study applies a treatment (multivitamin) to the subjects.

6. Randomized block design

7. (a) Convenience sampling is used because all the people sampled are in one convenient location.

(b) Systematic sampling is used because every tenth machine part is sampled.

(c) Stratified sampling is used because the population is first stratified and then a sample is collected from each stratum.

8. Convenience sampling. People at campgrounds may be strongly against air pollution because they are at an outdoor location.

CHAPTER 1 TEST SOLUTIONS

1. (a) Sampling, because the population of New Jersey is too large for the most popular type of investment to be easily recorded. Random sampling would be advised because it would be easy to select people from New Jersey randomly and then record their most popular type of investment.

(b) Census, because the population is small and it is relatively easy to obtain the ages of the 30 employees.

2. (a) Sample statistic. The value of 72% is a numerical description of a sample of U.S. adults ages 18 years and older.

(b) Population parameter. The average evidence based reading and writing score of 543 is a numerical description of all test takers in a recent year.

3. (a) Stratified sampling is used because the high school students are divided into strata (male and female), and a sample is selected from each stratum.

(b) Simple random sampling is used because each customer has an equal chance of being contacted, and all samples of 625 customers have an equal chance of being selected.

(c) Convenience sampling is used because a sample is taken from members of a population that are readily available. The sample may be biased because the teachers at that school may not be representative of the population of teachers.

4. (a) Quantitative. Ratio. The number of employees are numerical measurements. A ratio of two data values can be formed, so it makes sense to say that 40 employees are twice as many as 20 employees.

(b) Quantitative. Interval. The grade point averages are numerical measurements. Data can be ordered and meaningful differences can be calculated, but it does not make sense to say that a person with a 3.8 GPA is twice as smart as a person with a 1.9 GPA.

5. (a) The survey question is unbiased.

(b) The question is biased because it already suggests that the town's ban on skateboarding in parks is unfair. The question could be written as "What are your thoughts on the town's ban on skateboarding in parks?"

6. (a) Population: Collection of the responses of all U.S. physicians
Sample: Collection of the 19,183 U.S. physicians who were sampled.

(b) Both. Location, employment status, benefits received, and speciality are qualitative because they are attributes. Income and time spent seeing patients per week are quantitative because they are numerical measurements.

(c) Nominal: location, employment status, benefits received, specialty
Ratio: income, time spent seeing patients per week

(d) Observational study. The study does not attempt to influence the responses of the physicians and there is no treatment.

Descriptive Statistics

2.1 FREQUENCY DISTRIBUTIONS AND THEIR GRAPHS

2.1 TRY IT YOURSELF SOLUTIONS

1. The number of classes is 6.

Min = 14, Max = 55, Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{55-14}{6} = 6.83 \Rightarrow 7$

The minimum data entry is a convenient lower limit for the first class. Then add the class width to get the lower limits of the other classes. The upper limits are one less than the lower limit of the next class.

Lower limit	Upper limit
14	20
21	27
28	34
35	41
42	48
49	55

Make a tally mark for each entry in the appropriate class. The number of tally marks for a class is the frequency of that class.

Class	Frequency, f
14-20	8
21-27	15
28-34	14
35-41	7
42-48	4
49-55	3

2. Find each midpoint, relative frequency, and cumulative frequency.

$$\text{Midpoint} = \frac{\text{(Lower class limit) + (Upper class limit)}}{2}$$

$$\text{Relative frequency} = \frac{\text{Class frequency}}{\text{Sample size}} = \frac{f}{n}$$

The cumulative frequency of a class is the sum of the frequencies of that class and all previous classes.

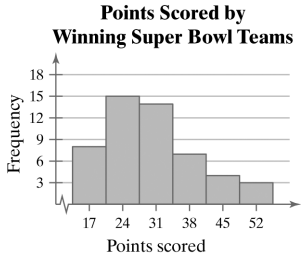
Class	f	Midpoint	Relative frequency	Cumulative frequency
14–20	8	$\frac{14+20}{2}=17$	$\frac{8}{51}\approx 0.1569$	8
21–27	15	$\frac{21+27}{2}=24$	$\frac{15}{51}\approx 0.2941$	$8+15=23$
28–34	14	$\frac{28+34}{2}=31$	$\frac{14}{51}\approx 0.2745$	$23+14=37$
35–41	7	$\frac{35+41}{2}=38$	$\frac{7}{51}\approx 0.1373$	$37+7=44$
42–48	4	$\frac{42+48}{2}=45$	$\frac{4}{51}\approx 0.0784$	$44+4=48$
49–55	3	$\frac{49+55}{2}=52$	$\frac{3}{51}\approx 0.0588$	$48+3=51$
	$\sum f = 51$		$\sum \frac{f}{n}=1$	

Sample answer: The most common range of points scored by winning teams is 21 to 27. About 14% of the winning teams scored more than 41 points.

3. Find the class boundaries. Because the data entries are integers, subtract 0.5 from each lower limit to find the lower class boundaries and add 0.5 to each upper limit to find the upper class boundaries.

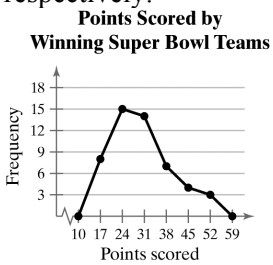
Class	Class Boundaries	Frequency, f
14–20	13.5–20.5	8
21–27	20.5–27.5	15
28–34	27.5–34.5	14
35–41	34.5–41.5	7
42–48	41.5–48.5	4
49–55	48.5–55.5	3

Use class midpoints for the horizontal scale and frequency for the vertical scale. (Class boundaries can also be used for the horizontal scale.)



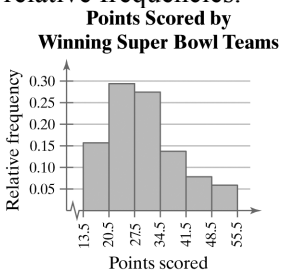
Sample answer: The most common range of points scored by winning teams is 21 to 27. About 14% of the winning teams scored more than 41 points.

4. To construct the frequency polygon, use the same horizontal and vertical scales that were used in the histogram labeled with the class midpoints in Try It Yourself 3. Then plot the points that represent the midpoint and frequency of each class and connect the points with line segments. Extend the left side and right side to one class width before the first class midpoint and after the last class midpoint, respectively.

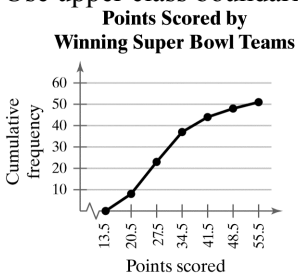


Sample answer: The frequency of points scored increases up to 24 points and then decreases.

5. Notice the shape of the relative frequency histogram is the same as the shape of the frequency histogram constructed in Try It Yourself 3. The only difference is that the vertical scale measures the relative frequencies.

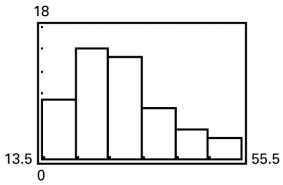


6. Use upper class boundaries for the horizontal scale and cumulative frequency for the vertical scale.



Sample answer: The greatest increase in cumulative frequency occurs between 20.5 and 27.5.

7.



2.1 EXERCISE SOLUTIONS

1. Organizing the data into a frequency distribution may make patterns within the data more evident. Sometimes it is easier to identify patterns of a data set by looking at a graph of the frequency distribution.

2. If there are too few or too many classes, it may be difficult to detect patterns because the data are too condensed or too spread out.

3. Class limits determine which numbers can belong to that class.
Class boundaries are the numbers that separate classes without forming gaps between them.

4. Relative frequency of a class is the portion, or percentage, of the data that falls in that class.
Cumulative frequency of a class is the sum of the frequencies of that class and all previous classes.

5. The sum of the relative frequencies must be 1 or 100% because it is the sum of all portions or percentages of the data.

6. A frequency polygon displays frequencies or relative frequencies whereas an ogive displays cumulative frequencies.

7. False. Class width is the difference between the lower (or upper limits) of consecutive classes.

8. True

9. False. An ogive is a graph that displays cumulative frequencies.

10. True

11. $\text{Class width} = \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{64-9}{7} \approx 7.9 \Rightarrow 8$
Lower class limits: 9, 17, 25, 33, 41, 49, 57
Upper class limits: 16, 24, 32, 40, 48, 56, 64

12. $\text{Class width} = \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{88-12}{6} \approx 12.7 \Rightarrow 13$
Lower class limits: 12, 25, 38, 51, 64, 77
Upper class limits: 24, 37, 50, 63, 76, 89

13. $\text{Class width} = \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{135-17}{8} = 14.75 \Rightarrow 15$
Lower class limits: 17, 32, 47, 62, 77, 92, 107, 122
Upper class limits: 31, 46, 61, 76, 91, 106, 121, 136

14. $\text{Class width} = \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{247-54}{10} = 19.3 \Rightarrow 20$
Lower class limits: 54, 74, 94, 114, 134, 154, 174, 194, 214, 234
Upper class limits: 73, 93, 113, 133, 153, 173, 193, 213, 233, 253

15. (a) Class width $= 11 - 0 = 11$

(b) and (c)

$$\text{Midpoint} = \frac{\text{(Lower class limit)} + \text{(Upper class limit)}}{2}$$

Find the class boundaries. Because the data entries are integers, subtract 0.5 from each lower limit to find the lower class boundaries and add 0.5 to each upper limit to find the upper class boundaries.

Class	**Midpoint**	**Class boundaries**
0 – 10	5	–0.5 – 10.5
11 – 21	16	10.5 – 21.5
22 – 32	27	21.5 – 32.5
33 – 43	38	32.5 – 43.5
44 – 54	49	43.5 – 54.5
55 – 65	60	54.5 – 65.5
66 – 76	71	65.5 – 76.5

16. (a) Class width $= 33 - 25 = 8$

(b) and (c)

$$\text{Midpoint} = \frac{\text{(Lower class limit)} + \text{(Upper class limit)}}{2}$$

Find the class boundaries. Because the data entries are integers, subtract 0.5 from each lower limit to find the lower class boundaries and add 0.5 to each upper limit to find the upper class boundaries.

Class	**Midpoint**	**Class boundaries**
25–32	28.5	24.5–32.5
33–40	36.5	32.5–40.5
41–48	44.5	40.5–48.5
49–56	52.5	48.5–56.5
57–64	60.5	56.5–64.5
65–72	68.5	64.5–72.5
73–80	76.5	72.5–80.5

17. Relative frequency $= \dfrac{\text{Class frequency}}{\text{Sample size}} = \dfrac{f}{n}$

The cumulative frequency of a class is the sum of the frequencies of that class and all previous classes.

Class	Frequency f	Midpoint	Relative frequency	Cumulative frequency
0 – 10	188	5	0.15	188
11 – 21	372	16	0.30	560
22 – 32	264	27	0.22	824
33 – 43	205	38	0.17	1029
44 – 54	83	49	0.07	1112
55 – 65	76	60	0.06	1188
66 – 76	32	71	0.03	1220
	$\sum f = 1220$		$\sum \frac{f}{n} = 1$	

18. Relative frequency $= \dfrac{\text{Class frequency}}{\text{Sample size}} = \dfrac{f}{n}$

The cumulative frequency of a class is the sum of the frequencies of that class and all previous classes.

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
25 – 32	86	28.5	0.24	86
33 – 40	39	36.5	0.11	125
41 – 48	41	44.5	0.11	166
49 – 56	48	52.5	0.13	214
57 – 64	43	60.5	0.12	257
65 – 72	68	68.5	0.19	325
73 – 80	40	76.5	0.11	365
	$\sum f = 365$		$\sum \frac{f}{n} \approx 1$	

19. (a) Number of classes: 7

(b) Greatest frequency: about 300
Least frequency: about 10

(c) Class width: 10

(d) *Sample answer:* About half of the employee salaries are between \$50,000 and \$69,000.

20. (a) Number of classes: 6

(b) Greatest frequency: 37
Least frequency: 1

(c) Class width: 53

(d) *Sample answer:* The heights of most roller coasters are less than 231 feet.

21. Identify the highest point and its respective class. Class with greatest frequency: 506 – 510
Identify the lowest point (not including the points on the horizontal axis) and its respective class. Class with least frequency: 474 – 478

22. Identify the highest point and its respective class. Class with greatest frequency: 3.5 – 4.5 miles
Identify the lowest point (not including the points on the horizontal axis) and its respective class. Class with least frequency: 0.5 – 1.5 miles

23. (a) Identify the tallest bar and its respective class. Class with greatest relative frequency: 35 – 36 centimeters
Identify the shortest bar and its respective class. Class with least relative frequency: 39 – 40 centimeters

(b) Greatest relative frequency ≈ 0.25
Least relative frequency ≈ 0.01

(c) *Sample answer*: From the graph, 0.25 or 25% of females have a fibula length between 35 and 36 centimeters.

24. (a) Identify the tallest bar and its respective class. Class with greatest relative frequency: 11 – 12 minutes
Identify the shortest bar and its respective class. Class with least relative frequency: 14 – 15 minutes

(b) Greatest relative frequency ≈ 38%
Least relative frequency ≈ 4%

(c) *Sample answer*: From the graph, about 0.75 or 75% of campus security response times are between 11 and 13 minutes.

25. (a) Locate the cumulative frequency of the highest (right-most) point. The number in the sample is 75.

(b) Locate the neighboring points where the pitch between them is the steepest. The greatest increase in frequency is from 158.5 – 201.5 pounds.

26. (a) Locate the cumulative frequency of the highest (right-most) point. The number in the sample is 77.

(b) Locate the neighboring points where the pitch between them is the steepest. The greatest increase in frequency is from 68 – 70 inches.

27. (a) Locate 201.5 on the horizontal axis and find the corresponding cumulative frequency at the point on the ogive: 47

(b) Locate 68 on the vertical axis and find the corresponding weight at the point on the ogive: 287.5 pounds

(c) Subtract the cumulative frequency for each weight: 62 – 22 = 40

(d) Subtract the cumulative frequency for bears weighing 330.5 pounds from the number in the sample: 75 – 69 = 6

28. (a) Locate 72 on the horizontal axis and find the corresponding cumulative frequency at the point on the ogive: 71

(b) Locate 15 on the vertical axis and find the corresponding height at the point on the ogive: 68 inches

(c) Subtract the cumulative frequency for each height: 71 – 15 = 56

(d) Subtract the cumulative frequency for adult males that are 70 inches tall from the number in the sample: 77 – 47 = 30

29. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{39-0}{5} = 7.8 \Rightarrow 8$

Class	**Frequency, *f***	**Midpoint**	**Relative frequency**	**Cumulative frequency**
0 –7	8	3.5	0.33	8
8–15	7	11.5	0.29	15
16–23	3	19.5	0.13	18
24–31	3	27.5	0.13	21
32–39	3	35.5	0.13	24
	$\sum f = 24$		$\sum \frac{f}{n} \approx 1$	

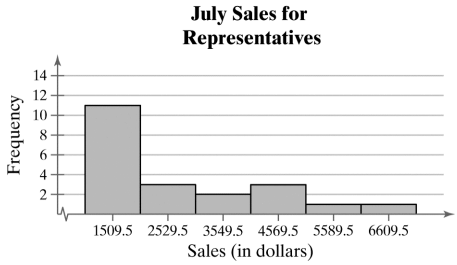
Class with greatest frequency: 0 – 7
Classes with least frequency: 16 – 23, 24 – 31, 32 – 39

30. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{530-30}{6} \approx 83.3 \Rightarrow 84$

Class	**Frequency, *f***	**Midpoint**	**Relative frequency**	**Cumulative frequency**
30–113	5	71.5	0.17	5
114–197	7	155.5	0.23	12
198–281	8	239.5	0.27	20
282–365	3	323.5	0.10	23
366–449	3	407.5	0.10	26
450–533	4	491.5	0.13	30
	$\sum f = 30$		$\sum \frac{f}{n} = 1$	

31. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{7119-1000}{6} \approx 1019.8 \Rightarrow 1020$

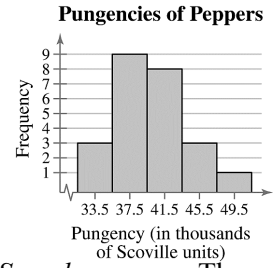
Class	Frequency, f	Mid-point	Relative frequency	Cumulative frequency
1000 –2019	11	1509.5	0.52	11
2020 –3039	3	2529.5	0.14	14
3040– 4059	2	3549.5	0.10	16
4060– 5079	3	4569.5	0.14	19
5080– 6099	1	5589.5	0.05	20
6100– 7119	1	6609.5	0.05	21
	$\sum f = 21$		$\sum \frac{f}{n} = 1$	



Sample answer: The graph shows that most of the sales representatives at the company sold from \$1000 to \$2019.

32. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{51-32}{5} = 3.8 \Rightarrow 4$

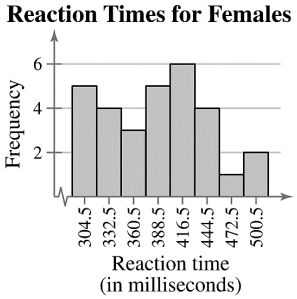
Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
32-35	3	33.5	0.1250	3
36-39	9	37.5	0.3750	12
40-43	8	41.5	0.3333	20
44-47	3	45.5	0.1250	23
48-51	1	49.5	0.0417	24
	$\sum f = 24$		$\sum \frac{f}{n} = 1$	



Sample answer: The graph shows that most of the pungencies of the peppers were between 36,000 and 43,000 Scoville units.

33. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{514-291}{8} = 27.875 \Rightarrow 28$

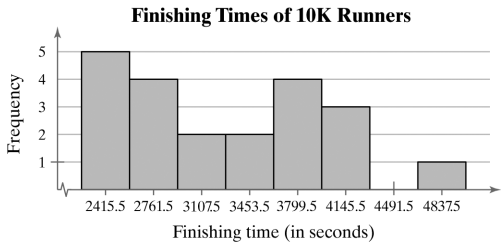
Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
291-318	5	304.5	0.1667	5
319-346	4	332.5	0.1333	9
347-374	3	360.5	0.1000	12
375-402	5	388.5	0.1667	17
403-430	6	416.5	0.2000	23
431-458	4	444.5	0.1333	27
459-486	1	472.5	0.0333	28
487-514	2	500.5	0.0667	30
	$\sum f = 30$		$\sum \frac{f}{n} = 1$	



Sample answer: The graph shows that the most frequent reaction times were between 403 and 430 milliseconds.

34. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{5008 - 2243}{8} = 345.625 \Rightarrow 346$

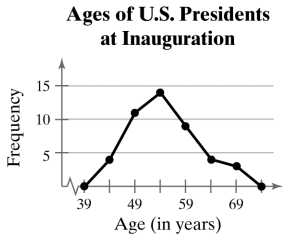
Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
2243–2588	5	2415.5	0.2381	5
2589–2934	4	2761.5	0.1905	9
2935–3280	2	3107.5	0.0952	11
3281–3626	2	3453.5	0.0952	13
3627–3972	4	3799.5	0.1905	17
3973–4318	3	4145.5	0.1429	20
4319–4664	0	4491.5	0.0000	20
4665–5010	1	4837.5	0.0476	21
	$\sum f = 21$		$\sum \frac{f}{n} = 1$	



Sample answer: The graph shows that the most frequent finishing times were from 2243 to 2588 seconds.

35. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{70-42}{6} \approx 4.7 \Rightarrow 5$

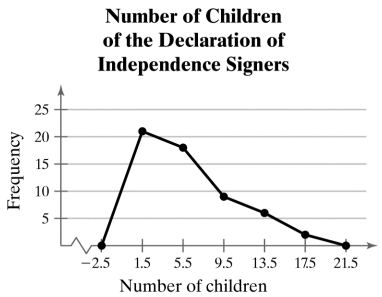
Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
42–46	4	44	0.0889	4
47–51	11	49	0.2444	15
52–56	14	54	0.3111	29
57–61	9	59	0.2000	38
62–66	4	64	0.0889	42
67–71	3	69	0.0667	45
	$\sum f = 45$		$\sum \frac{f}{n} = 1$	



Sample answer: The graph shows that the number of U.S. presidents who were 52 or older at inauguration was twice as many as those who were 51 and younger.

36. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{18-0}{5} = 3.6 \Rightarrow 4$

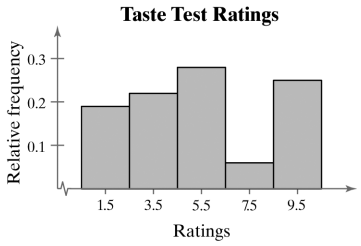
Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
0–3	21	1.5	0.3750	21
4–7	18	5.5	0.3214	39
8–11	9	9.5	0.1607	48
12–15	6	13.5	0.1071	54
16–19	2	17.5	0.0357	56
	$\sum f = 56$		$\sum \frac{f}{n} \approx 1$	



Sample answer: The graph shows that most of the signers of the Declaration of Independence had 7 or fewer children.

37. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{10-1}{5} = 1.8 \Rightarrow 2$

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
1–2	7	1.5	0.19	7
3–4	8	3.5	0.22	15
5–6	10	5.5	0.28	25
7–8	2	7.5	0.06	27
9–10	9	9.5	0.25	36
	$\sum f = 36$		$\sum \frac{f}{n} \approx 1$	



Class with greatest relative frequency: 5 – 6
Class with least relative frequency: 7 – 8

38. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{19-7}{5} = 2.4 \Rightarrow 3$

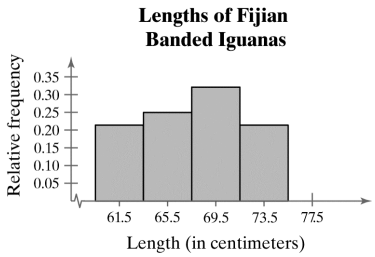
Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
7–9	8	8	0.2857	8
10–12	11	11	0.3929	19
13–15	8	14	0.2857	27
16–18	0	17	0.0000	27
19–21	1	20	0.0357	28
	$\sum f = 28$		$\sum \frac{f}{n} \approx 1$	

Class with greatest relative frequency: 10 – 12
Class with least relative frequency: 16 – 18

39. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{75-60}{5} = 3$

Notice that using a class width of 3 is not wide enough to include all the data with 5 classes. Therefore, use a class width of 4.

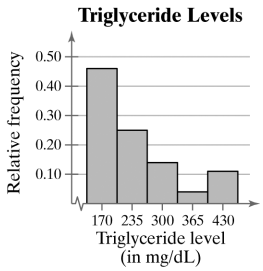
Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
60–63	6	61.5	0.2143	6
64–67	7	65.5	0.2500	13
68–71	9	69.5	0.3214	22
72–75	6	73.5	0.2143	28
76–79	0	77.5	0.0000	28
	$\sum f = 28$		$\sum \frac{f}{n} \approx 1$	



Class with greatest relative frequency: 68 – 71
Class with least relative frequency: 76 – 79

40. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{462-138}{5} = 64.8 \Rightarrow 65$

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
138–202	13	170	0.46	13
203–267	7	235	0.25	20
268–332	4	300	0.14	24
333–397	1	365	0.04	25
398–462	3	430	0.11	28
	$\sum f = 28$		$\sum \frac{f}{n} \approx 1$	



Class with greatest relative frequency: 138 – 202
Class with least relative frequency: 333 – 397

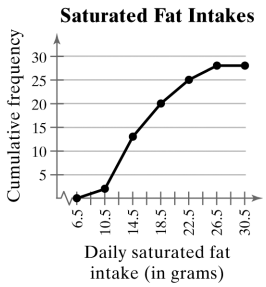
41. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{75-52}{6} \approx 3.8 \Rightarrow 4$

Class	Frequency, f	Relative frequency	Cumulative frequency
52–55	6	0.1714	6
56–59	4	0.1143	10
60–63	6	0.1714	16
64–67	10	0.2857	26
68–71	5	0.1429	31
72–75	4	0.1143	35
	$\sum f = 35$		$\sum \frac{f}{n} \approx 1$

Location of the greatest increase in frequency: 64 – 67

42. Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{26-7}{6} \approx 3.2 \Rightarrow 4$

Class	Frequency, f	Relative frequency	Cumulative frequency
7–10	2	0.0714	2
11–14	11	0.3929	13
15–18	7	0.2500	20
19–22	5	0.1786	25
23–26	3	0.1071	28
27–30	0	0.0000	28
	$\sum f = 28$	$\sum \frac{f}{n} \approx 1$	

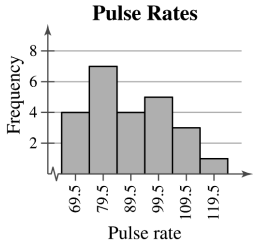


Location of the greatest increase in frequency: 11 – 14

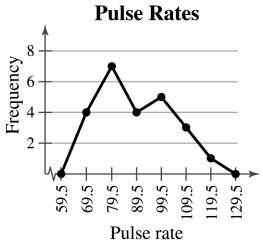
43. (a) Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{120-65}{6} \approx 9.2 \Rightarrow 10$

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
65-74	4	69.5	0.1667	4
75-84	7	79.5	0.2917	11
85-94	4	89.5	0.1667	15
95-104	5	99.5	0.2083	20
105-114	3	109.5	0.1250	23
115-124	1	119.5	0.0417	24
	$\sum f = 24$		$\sum \frac{f}{N} \approx 1$	

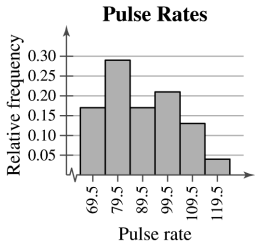
(b)



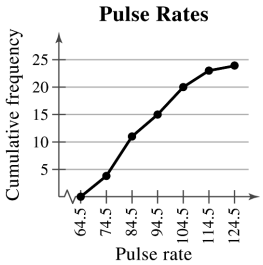
(c)



(d)



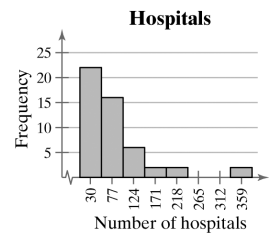
(e)



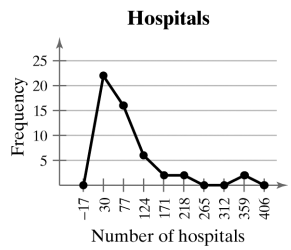
44. (a) Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{382-7}{8} = 46.875 \Rightarrow 47$

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
7–53	22	30	0.44	22
54–100	16	77	0.32	38
101–147	6	124	0.12	44
148–194	2	171	0.04	46
195–241	2	218	0.04	48
242–288	0	265	0.00	48
289–335	0	312	0.00	48
336–382	2	359	0.04	50
	$\sum f = 50$		$\sum \frac{f}{n} = 1$	

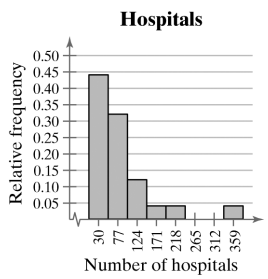
(b)



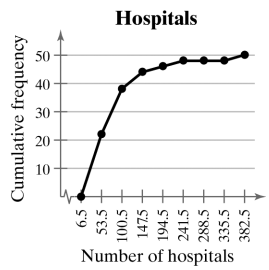
(c)



(d)

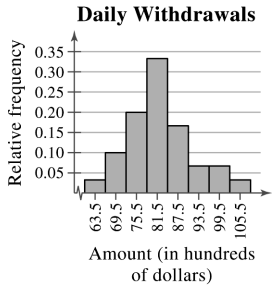


(e)



45. (a) Class width $= \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{104-61}{8} = 5.375 \Rightarrow 6$

Class	Frequency, f	Midpoint	Relative frequency
61-66	1	63.5	0.033
67-72	3	69.5	0.100
73-78	6	75.5	0.200
79-84	10	81.5	0.333
85-90	5	87.5	0.167
91-96	2	93.5	0.067
97-102	2	99.5	0.067
103-108	1	105.5	0.033
	$\sum f = 30$		$\sum \frac{f}{n} = 1$



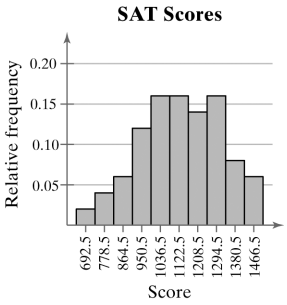
(b) 16.7%, because the sum of the relative frequencies for the last three classes is 0.167.

(c) $9700, because the sum of the relative frequencies for the last two classes is 0.10.

46. (a) $\text{Class width} = \dfrac{\text{Range}}{\text{Number of classes}} = \dfrac{1500-650}{10} = 85$

Notice that using a class width of 85 is not wide enough to include all the data with 10 classes. Therefore, use a class width of 86.

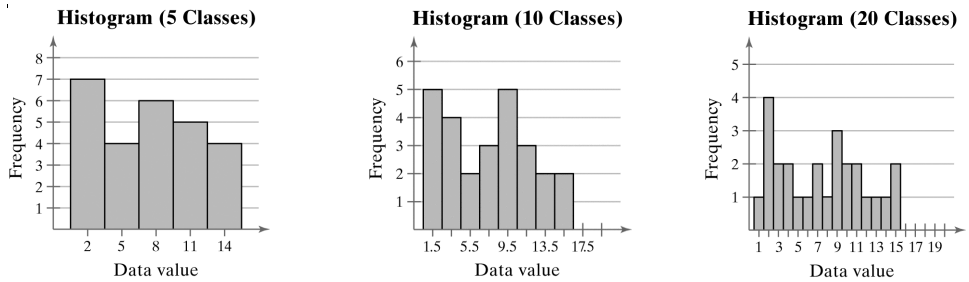
Class	Frequency, f	Midpoint	Relative frequency
650-735	1	692.5	0.02
736-821	2	778.5	0.04
822-907	3	864.5	0.06
908-993	6	950.5	0.12
994-1079	8	1036.5	0.16
1080-1165	8	1122.5	0.16
1166-1251	7	1208.5	0.14
1252-1337	8	1294.5	0.16
1338-1423	4	1380.5	0.08
1424-1509	3	1466.5	0.06
	$\sum f = 50$		$\sum \frac{f}{n} = 1$



(b) 64%; The portion of the scores greater than or equal to 1070 is 0.64.

(c) A score of 908 or above, because the sum of the relative frequencies of the class starting with 908 and all classes with higher scores is 0.88.

47.



In general, a greater number of classes better preserves the actual values of the data set but is not as helpful for observing general trends and making conclusions. In choosing the number of classes, an important consideration is the size of the data set. For instance, you would not want to use 20 classes if your data set contained 20 entries. In this particular example, as the number of classes increases, the histogram shows more fluctuation. The histograms with 10 and 20 classes have classes with zero frequencies. Not much is gained by using more than five classes. Therefore, it appears that five classes would be best.

2.2 MORE GRAPHS AND DISPLAYS

2.2 TRY IT YOURSELF SOLUTIONS

1. Because the data entries go from a low of 14 to a high of 55, use stem values from 1 to 5. List the stems to the left of a vertical line. For each data entry, list a leaf to the right of its stem.

1	4 6 6 6 7
2	0 0 0 1 1 1 3 3 4 4 4 4 6 7 7 7 7 7 8 9
3	0 1 1 1 1 2 2 3 4 4 4 4 5 5 5 7 8 8 9
4	2 3 6 8 9
5	2 5

Key: 1 | 4 = 14

Sample answer: Most of the winning teams scored between 20 and 39 points.

2. Use the leaves 0, 1, 2, 3, and 4 in the first stem row and the leaves 5, 6, 7, 8, and 9 in the second stem row.

1	4
1	6 6 6 7
2	0 0 0 1 1 1 3 3 4 4 4 4
2	6 7 7 7 7 7 8 9
3	0 1 1 1 1 2 2 3 4 4 4 4
3	5 5 5 7 8 8 9
4	2 3
4	6 8 9
5	2
5	5

Key: 1 | 4 = 14

Sample answer: Most of the winning teams scored from 20 to 35 points.