Complete Solutions Manual to Accompany

Statistics: Learning from Data

SECOND EDITION

Roxy Peck
California Polytechnic State University, San Luis Obispo

Tom Short
West Chester University of Pennsylvania

Prepared by

Stephen Miller
Winchester Thurston School, Pittsburgh, PA

CENGAGE Learning®

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

ISBN-13: 978-133755839-6
ISBN-10: 1-33755839-7

Cengage Learning
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at: **www.cengage.com/global**.

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions, visit **www.cengage.com**.

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com**.

Printed in the United States of America
1 2 3 4 5 6 7 17 16 15 14 13

Contents

Chapter 1 Collecting Data in Reasonable Ways

Section 1.2 Exercise Set 1

1.1: This is an observational study because the person conducting the study merely recorded (based on a survey) whether or not the boomers sleep with their phones within arm's length, and whether or not people ages 50 to 64 used their phones to take photos.

1.2: This is an observational study because the researchers recorded responses to questions on a survey. No men were assigned to different experimental groups.

1.3: This is an experiment because the researchers assigned different toddlers to experimental conditions (adult played with/talked to the robot or the adult ignored the robot).

1.4: This is an observational study because the researchers surveyed U.S. adults and drew a conclusion from the survey results; there were no experimental treatments assigned.

1.5: This is an experiment because the researchers assigned study participants to one of three treatment groups (meditation, distraction task, or relaxation technique).

Section 1.2 Exercise Set 2

1.6: This is an observational study based on results of a survey (no consumers were assigned to different experimental conditions).

1.7: This is an experiment because the participants (college students) were assigned to different experimental conditions (McDonald's Big Mac coupon or Subway 12-inch Italian BMT coupon).

1.8: This is an observational study because the researchers based their conclusions on the results of a survey. There was no assignment to different experimental conditions.

1.9: This is an experiment because the researchers assigned study participants to different experimental conditions (garlic supplement group or no garlic supplement group).

1.10: This is an experiment because the researchers assigned the children to different experimental groups (ball behaving as expected and ball behaving in an unexpected way).

Additional Exercises for Section 1.2

1.11: This is an observational study because the researchers observed the proportion of patients who got an infection in the two groups (overnight hospitalization or more than one night hospitalization); there was no assignment of subjects to experimental groups.

1.12: This is an experiment because the researcher assigned three of the statistics discussion sections to receive chocolate, and the remaining three did not receive chocolate (the chocolate or lack of chocolate was the experimental group).

1.13: This is an experiment because the study participants were assigned to one of the two experimental groups (how much would you pay for the mug or how much would you sell the mug for).

1.14: The study described was an experiment because the study participants were asked either the first or second question (the participants were assigned to one of the two experimental groups).

Section 1.3 Exercise Set 1

1.15: (a) The group of 716 bicycle fatalities represents a census of the 2008 bicycle fatalities. (b) Because the group of 716 represents a census, the average age of 41 years is a population characteristic.

1.16: The sample is the 100 San Fernando Valley residents, and the population of interest is all San Fernando Valley residents.

1.17: The headline referenced in the previous exercise is not appropriate because the Los Angeles County Metropolitan Transportation Authority only surveyed 100 residents from those who attended a community forum in Van Nuys on a Monday. Residents who could not attend the forum were not included in the survey. The residents who attended the forum are likely those who feel strongly about the issue. As such, it is not appropriate to generalize the survey results to all San Fernando Valley residents. The sample is biased, and we cannot be sure that the results of this survey are representative of the population as a whole. If results from this study are to be reported at all, a more appropriate headline might be "Over two-thirds of those attending a community forum OK with 1-cent transit tax."

1.18: There are several reasonable approaches. One is described here. Using the list of all students at the school, write their names on identical slips of paper. Thoroughly mix the slips of paper, and select 150 slips. The individuals whose names are on the slips of paper constitute the sample.

1.19: (a) The population of interest is all U.S. women. (b) Although the details of the sampling scheme are not presented, the sample size is large (which is generally desirable). However, not all states were represented in the sample; only women from Maryland, Minnesota, Oregon and Pennsylvania were included in the sample. As such, it might be difficult to generalize beyond the population of women in those four states. (c) Given that only women from four states were included in the sample, the sample is not likely to be

representative of the population of interest. (d) Selection bias is present because the selection method excluded women from all states other than Maryland, Minnesota, Oregon and Pennsylvania.

Section 1.3 Exercise Set 2

1.20: The given percentages are statistics, because they were computed from sample data obtained by a survey.

1.21: The group of soldiers surveyed represents a census. The reported percentage is a population characteristic, because it was computed using data from the entire population of soldiers.

1.22: (a) This is a convenience sample because there was no random selection of a top medical school or an expert within the school. (b) Because there was no random selection of the infectious disease expert, generalization to all doctors is not reasonable.

1.23: There are several reasonable approaches. One is described here. Write the names of all students enrolled at the college on identical slips of paper. Thoroughly mix the slips of paper and select 100 of the slips. The students whose names are on the 100 slips of paper constitute the simple random sample.

1.24: (a) The population of interest is all Arizona drivers. (b) Although the details of the sampling scheme are not presented, there was random selection of AAA Arizona members. Therefore, the sample was selected in a reasonable way. (c) This sample is not representative of the population of interest because only those people who are AAA Arizona members were included in the sample. (d) Selection bias is present because those drivers who are not AAA Arizona members have been excluded from the sample. Nonresponse bias could also be present because there was no indication of how many surveys were not returned.

Additional Exercises for Section 1.3

1.25: (a) One example of a leading question is “Knowing that there are health problems associated with consuming too much sugar, and that soft drinks contain large amounts of added sugar, should there be a tax on soft drinks to encourage people to consume less?” (b) A double-barreled question is “How satisfied are you with the food and service at this restaurant?” It would be better to split the bad question into two questions, such as “How satisfied are you with the food at this restaurant” and “How satisfied are you with the service at this restaurant.”

1.26: The population is all 7000 property owners in this particular rural county. The sample is the 500 property owners selected at random from the 7000 total owners in the county.

1.27: The population is the 5000 bricks in the lot available at the auction. The sample is the 100 bricks chosen for inspection.

1.28: The chairman does not understand the power of random selection. Random samples tend to reflect the distribution of voters in the population. Although it is possible to obtain a random sample that is not representative of the population, the risk of getting a sample that is not representative of the population does not depend on what fraction of the population is sampled. The random selection process allows us to be confident that the resulting sample will adequately reflect the population, even when the sample consists of only a relatively small fraction of the population.

1.29: Bias introduced through the two different sampling methods may have contributed to the different results. The online sample could suffer from voluntary response bias in that perhaps only those who feel very strongly would take the time to go to the website and register their vote. In addition, younger people might be more technologically savvy, and therefore the website might represent the views of younger people (particularly students) who support the parade. The telephone survey telephone responses might represent the view of permanent residents (as students might only use cell phones and not have a local phone number).

Section 1.4 Exercise Set 1

1.30: Random assignment allows the researcher to create groups that are equivalent, so that the subjects in each experimental group are as much alike as possible. This ensures that the experiment does not favor one experimental condition (playing Unreal Tournament 2004 or Tetris) over another.

1.31: (a) Allowing subject participants to choose which group they want to be in could introduce systematic differences between the two experimental conditions (compression socks group or regular athletic socks group), resulting in potential confounding. Those who would choose to wear compression socks might, in some way, be different from those who would choose the regular athletic socks. We would not know if differences in the measured variables from the blood samples between the two groups were due to the compression socks, or due to some inherent differences in the subjects who chose their experimental groups. (b) It would be good to have the runners be blind to the type of socks they were given to eliminate the possible psychological advantage the runners might have if they knew they were wearing compression socks. That way, we could better determine if the compression socks were the cause of any performance gains.

1.32: (a) The attending nurse was responsible for administering medication after judging the degree of pain and nausea, so the researchers did not want the nurse's personal beliefs about the different surgical procedures to influence measurements. (b) Because the

children who had the surgery could easily determine whether the surgical procedure was laparoscopic repair or open repair based on the type of incision.

1.33: There are several possible approaches. One is described here. Write each subject's name on identical slips of paper. Mix the slips of paper thoroughly and draw out slips one at a time. The names on the first 15 slips are assigned to the experimental condition of listening to a Mozart piano sonata for 24 minutes. The names on the next 15 slips are assigned to the experimental condition of listening to popular music for the same length of time. The remaining 15 names are assigned to the relaxation with no music experimental condition.

1.34: (1) Does a dietary supplement consisting of Omega 3, Omega 6, and antioxidants reduce hair loss in women with stage 1 hair loss? (2) The experimental conditions are the supplement and control. (3) The response variable is the change in hair density. (4) The experimental units are the 120 women who volunteered for the study. (5) Yes, the design incorporates random assignment of women to either the supplement group or the control group. (6) Yes, there was a control group. (7) Yes, there was blinding. The expert who determined the change in hair density did not know which of the women had taken the supplement. There is no indication that the women were blinded as to which treatment they received, although this could be incorporated.

Section 1.4 Exercise Set 2

1.35: Random assignment allows the researcher to create groups that are equivalent, so that the subjects in each experimental group are as much alike as possible. This ensures that the experiment does not favor one experimental condition (distraction or no distraction) over another.

1.36: (a) Allowing subject participants to choose which group they want to be in could introduce systematic differences between the two experimental conditions (knee replacement surgery with exercise and exercise therapy alone), resulting in potential confounding. Those who chose knee replacement surgery plus exercise might, in some way, be different from those who chose exercise therapy alone. We would not know if differences in pain relief between the two groups were due to the knee replacement surgery with exercise, or due to some inherent differences in the subjects who chose their experimental groups. (b) Because the purpose of this experiment is to determine whether knee replacement surgery with exercise provided more pain relief than exercise therapy alone, a control group could be used to provide a baseline against which the treatment groups can be compared to determine if the treatments had an effect.

1.37: Blinding of both the dog handlers and experimental observers is important so that the dogs are not intentionally or otherwise guided in determining which patients have cancer. The blinding guarantees that the dogs do not rely on any information other than the patient's breath.

1.38: There are several possible approaches. One is described here. Write each subject's name on identical slips of paper. Mix the slips of paper thoroughly and draw out slips one at a time. The names on the first 20 slips are assigned to one type of keyboard (experimental condition), and the remaining 20 are assigned to the other type of keyboard (the other experimental condition).

1.39: Was there a control group in which there were identical sheets of paper with no words written on them? Was there any random assignment of experimental units to treatment groups? How were the experimental units selected? How many water bottles were used in the study? Were the water bottles identical? How many bottles of water were used? What measurements were made on the water? Were measurements made both before and after the words were applied to the bottle? Who took the measurements? Was the person taking the measurements blinded to the presence or absence of words on the pieces of paper?

Additional Exercises for Section 1.4

1.40: The experimental conditions were the presence or absence of music with a vocal component. The response is the time required to complete the surgical procedure.

1.41: (a) Some surgical procedures are more complex and require a greater degree of concentration; music with a vocal component might be more distracting when the surgical procedure is more complex. (b) The temperature of the room might affect the comfort of the surgeon; if the surgeon is too hot or too cold, she or he might be uncomfortable, and therefore more easily distracted by the vocal component. (c) If the music is too loud, the surgeon might be distracted and unable to focus, regardless of the presence or absence of the vocal component. If the music is too soft, the surgeon might try to concentrate on listening to the vocal component, and therefore pay more attention to the music rather than the surgical procedure.

1.42: Random assignment of surgeons to music condition is important because there might be something inherently different about surgeons who want no vocals versus those who do want vocals. Random assignment ensures that the experiment does not favor one experimental condition over another.

1.43: This experiment could not have been double-blind because the surgeon would know whether or not there was a vocal component to the music.

1.44: Yes, the random assignment of subjects to experimental groups has been successful in creating groups that are similar in age. Both the LR and OR groups have similar maximum ages, and the LR group does have a few children with slightly lower ages than the OR group. Overall, however, the LR and OR groups are quite similar with respect to ages.

1.45: (a) Probably not, because the judges might not believe that Denny's food is as good as other restaurants. (b) Experiments are often blinded in this way to eliminate preconceptions about particular experimental treatments.

Section 1.5 Exercise Set 1

1.46: (a) This was most likely an observational study. (b) It is not reasonable to conclude that pushing a shopping cart causes people to be less likely to purchase junk food because the results of observational studies cannot be used to draw cause-and-effect conclusions.

1.47: (a) It is not reasonable to conclude that owning a dog causes higher heart rate variability. This was an observational study, so cause-and-effect conclusions cannot be drawn. (b) It is not reasonable to generalize the results of this survey to all adult Americans because the study participants were not randomly selected from the population of all adult Americans.

1.48: The researcher would have had to assign the nine cyclists at random to one of the three experimental conditions (chocolate milk, Gatorade, or Endurox).

1.49:

Study 1:

Question 1: This is an observational study.

Question 2: Yes, random selection was used.

Question 3: No, this was not an experiment so there were no experimental groups.

Question 5: It is reasonable to generalize to the population of students at this particular large college.

Study 2:

Question 1: This study was an experiment.

Question 2: Random selection was not used.

Question 3: There was no random assignment to experimental conditions (the grouping was based on gender).

Question 4: No, the conclusion is not appropriate because of confounding of gender and treatment (women ate pecans, and men did not eat pecans).

Question 5: It is not reasonable to generalize to a larger population.

Study 3:

Question 1: This is an observational study.

Question 2: There was no random selection.

Question 3: There was no random assignment to experimental groups.

Question 4: No, the conclusion is not appropriate because this was an observational study, and therefore cause-and-effect conclusions cannot be drawn.

Question 5: We cannot generalize to any larger population beyond the 200 volunteers.

Study 4:

Question 1: This is an experiment.

Question 2: There was no random selection from some population.

Question 3: Yes, there was random assignment to experimental groups.

Question 4: Yes, because this was a simple comparative experiment with random assignment of subjects to experimental groups. We can draw cause-and-effect conclusions.

Question 5: We cannot generalize to a larger population because there was no random selection from some population.

Study 5:

Question 1: This is an experiment.

Question 2: Yes, there was random selection from students enrolled at a large college.

Question 3: Yes, random assignment of subjects to experimental groups was used.

Question 4: Yes, because this was a simple comparative experiment with random assignment of subjects to experimental groups. We can draw cause-and-effect conclusions.

Question 5: Due to the random selection of students, we can generalize conclusions from this study to the population of all students enrolled at the large college.

Section 1.5 Exercise Set 2

1.50: (a) Random selection from the population of affluent Americans is required. (b) No, because the population sampled from was affluent Americans.

1.51: In order to determine if the conclusions implied by the headline are appropriate, I would need to know if dieters were randomly assigned to the experimental conditions (large fork or small fork). In order to generalize to the population of dieters, I would also want to know if the study participants were randomly selected from the population of dieters.

1.52: Random assignment ensures that the experiment does not favor one experimental condition (talking on the phone, not talking on the phone) over another. If the person crossing the virtual street was on the phone the first 10 crossings, and not on the phone the last 10 crossings, we wouldn't know if any difference between the treatments was due to the phone use or due to the person being either more or less aware of the surroundings for the last 10 crossings, for example.

1.53:

Study 1:

Question 1: This is an observational study.

Question 2: No, there was no random selection from a population.

Question 3: No, there was no random assignment to experimental groups.

Question 4: No, the conclusion that you can "strengthen your marriage with prayer" is not appropriate. There was no experiment conducted, so a cause-and-effect conclusion cannot be drawn.

Question 5: No, it is not reasonable to generalize conclusions from this study to some larger population because this was a voluntary response sample.

Study 2:

Question 1: This is an observational study.

Question 2: Yes, there was random selection from the population of AAUW members.

Question 3: There was no random assignment to experimental groups (this is not an experiment).

Question 4: No, the conclusion that you can "strengthen your marriage with prayer" is not appropriate. There was no experiment conducted, so a cause-and-effect conclusion cannot be drawn.

Question 5: Due to random selection, it is reasonable to generalize the conclusions from this study to the population of AAUW members.

Study 3:

Question 1: This was an observational study.

Question 2: No, there was no random selection from a population.

Question 3: No, there was no random assignment to experimental groups (this was an observational study, not an experiment).

Question 4: No, the conclusion that you can "strengthen your marriage with prayer" is not appropriate. Since this was an observational study, a cause-and-effect conclusion cannot be drawn.

Question 5: It is not reasonable to generalize conclusions from this study to a larger population because there was no random selection of study participants.

Additional Exercises for Section 1.5

1.54: It is not reasonable to conclude that being raised with two or more animals is the cause of the observed lower allergy rate. This was an observational study, so cause-and-effect conclusions cannot be drawn.

1.55: It might be that people who live in the South have a less healthy diet and exercise less than those in other parts of the country. As a result, the higher percentage of Southerners with high blood pressure might have nothing to do with living in the South.

1.56: This is an experiment.

1.57: There was no random selection from some population.

1.58: Yes, there was random assignment to experimental groups (portrait orientation or landscape orientation).

1.59: Yes, it is reasonable to draw the conclusion that reasoning using information displayed on a small screen is improved by turning the screen to landscape orientation because this was an experiment in which there was random assignment of subjects to experimental groups.

1.60: No, it is not reasonable to generalize the conclusions from this study to some larger population because there was no random selection of study participants from a population.

Chapter 1: Are You Ready to Move On? Chapter 1 Review Exercises

1.61: (a) This is an experiment due to the random assignment of subjects to experimental conditions (the five different rooms). (b) This is an observational study because there was no random assignment of subjects to experimental conditions; the researchers merely recorded what they observed on the MySpace pages. (c) This is an observational study

because there was no random assignment of subjects to experimental conditions; the researchers merely recorded the responses of the survey participants. (d) This is an experiment because of the random assignment of study participants (the adults with back pain) to experimental conditions (the four different treatments).

1.62: The population of interest is the 15,000 students at the college. The 200 students who were interviewed constitute the sample.

1.63: (a) 84% is a population characteristic. (b) 24.1 years is a statistic. (c) 22% is a population characteristic. (d) 6.4 days is a statistic. (e) 63 hours is a statistic.

1.64: (1) The study participants were volunteers and were not randomly selected. (2) The study participants were all students at Texas Women's University. (3) The study participants were all women (because they are students at a Women's university).

1.65: The council president should assign a unique identifying number to each of the names on the petition, numbered from 1 to 500. On identical slips of paper, write the numbers 1 to 500, with each number on a single slip of paper. Thoroughly mix the slips of paper and select 30 numbers. The 30 numbers correspond to the unique numbers assigned to names on the petition. These 30 names constitute the sample.

1.66: (a) (1) The patients are the population of interest. (2) The study description indicates no random selection of participants, so it does not appear as if the sample was selected in a reasonable way. (3) No, the sample is not likely to be representative of the population of interest. The sample consisted of only undergraduate students, so even if there was random selection of participants, the study results could not be generalized to the population of all patients. (4) It is likely that this study design is affected by selection bias because only undergraduate students were included in the study, thus systematically excluding all non-undergraduate students from the population of interest. (b) No, the stated conclusions are not reasonable because there was no random selection of study participants, and the study suffers from selection bias.

1.67: Without random assignment of the study participants to experimental condition, confounding could impact the conclusions of the study. For example, people who would choose an attractive avatar might be more outgoing and willing to engage than someone who would choose an unattractive avatar.

1.68: (a) Yes, by randomly selecting the 852 children to be in one experimental group (the book group), the remaining children, by default, are in the control group. (b) The control group allows the experimenter to assess how the response variable behaves when the treatment is not used. This provides a baseline against which the treatment groups can be compared to determine if the treatment has an effect. In this case, the researcher can determine whether

children given the reading books have better school performance, as measured by a reading test.

1.69: (a) It seems as if the alternate assignment to the experimental groups (large serving bowls, small serving bowls) would tend to produce groups that are similar. People who arrive to the party at approximately the same time might, in some way, be similar to each other, so dividing them into the different experimental groups as described would tend to make the two groups similar to each other. (b) Blinding ensures that individuals do not let personal beliefs influence their measurements. The research assistant that weighed the plates and estimated the calorie content of the food might (intentionally or not) have let her or his personal beliefs influence the estimate of the calorie content of the food on the plate.

1.70: There are several possible approaches. One is described here. Write each subject's name on identical slips of paper. Mix the slips of paper thoroughly and draw out slips one at a time. The names on the first 10 slips are assigned to the first hand drying method. The names on the next 10 slips are assigned to the second hand drying method. The remaining 10 names are assigned to the third hand drying method.

1.71: (a) (1) The experiment is designed to answer the question "Does using hand gestures help children learn math?" (2) The two experimental conditions are using hand gestures and not using hand gestures. (3) The response variable is the number correct on the six-problem test. (4) The experimental units are the 128 children in the study; they were selected because they were the children who answered all six questions on the pretest incorrectly. (5) Yes, the children were assigned randomly to one of the two experimental groups. (6) Yes, the control group is the experimental condition of not using any hand gestures. (7) There was no blinding and, indeed, it would not be possible to include blinding of subjects in this experiment (the children would know whether or not they were using hand gestures), and there is no need to blind the person recording the response because the test was graded with each answer correct or incorrect, so there is no subjectivity in recording the responses. (b) It seems as if the conclusions are reasonable because the subjects were assigned to the treatment groups at random.

1.72: (a) Yes, it is reasonable to generalize the stated conclusion to all 18-year-olds with a publically accessible MySpace web profile because the profiles were selected at random from all MySpace web profiles of 18-year-olds. (b) No, it is not reasonable to generalize the stated conclusion to all 18-year-old MySpace users because those users without publically accessible profiles were not included in the random selection process. (c) No, it is not reasonable to generalize the stated conclusion to all MySpace users because the study only included 18-year-old MySpace users.

1.73: (a) No, the 60 games selected were the 20 most popular (by sales) for each of three different gaming systems. The study excluded the games that were not in the top 20 most

popular (by sales). (b) It is not reasonable to generalize the researcher's conclusions to all video games due to selection bias (there was a systematic exclusion of those games not in the top 20 most popular (by sales)).

1.74: (a) The study described is not an experiment because there were no experimental conditions to which study participants were randomly assigned. (b) No, it is not reasonable to conclude that physical activity is the cause of the observed difference in body fat percentage. This was an observational study, and cause-and-effect conclusions cannot be drawn.

1.75:

Study 1:

Question 1: The study described is an observational study.

Question 2: No, there was no random selection from a population.

Question 3: No, there was no random assignment to experimental groups.

Question 4: No, it is not reasonable to conclude that taking calcium supplements is the cause of the increased heart attack risk.

Question 5: No, it is not reasonable to generalize conclusions from this study to a larger population because there was no random selection from a larger population.

Study 2:

Question 1: The study described is an observational study.

Question 2: Yes, there was random selection from the population of people living in Minneapolis who receive Social Security.

Question 3: No, there was no random assignment of subjects to experimental groups.

Question 4: No, it is not reasonable to conclude that taking calcium supplements is the cause of the increased heart attack risk.

Question 5: Yes, it is reasonable to generalize the results of this study to the population of people living in Minneapolis who receive Social Security.

Study 3:

Question 1: The study described is an experiment.

Question 2: Yes, there was random selection from the population of people living in Minneapolis who receive Social Security.

Question 3: No, there was no random assignment of subjects to experimental groups.

Question 4: No, it is not reasonable to conclude that taking calcium supplements is the cause of the increased risk of heart attack due to confounding and the lack of random assignment of subjects to experimental conditions. The participants in this study who did not have a previous history of heart problems were given the calcium supplement, and those with a history of heart problems were not given the supplement. It is not possible to determine the role of the calcium supplement because only those study participants who did not have a history of heart problems were given the supplement.

Question 5: It is possible to generalize the results from this study to the population of all people living in Minneapolis who receive Social Security. However, it is unclear (due to the confounding described in Question (4) what the conclusion would be.

Study 4:

Question 1: The study described is an experiment because there was random assignment of subjects to experimental conditions.

Question 2: No, there was no random selection from some larger population.

Question 3: Yes, there was random assignment of study participants to experimental groups.

Question 4: Yes, it is reasonable to conclude that taking calcium supplements is the cause of the increased risk of heart attack.

Question 5: No, it is not reasonable to generalize conclusions from this study to some larger population because of the lack of random selection of study participants from a population.

Chapter 2 Graphical Methods for Describing Data Distributions

Section 2.1 Exercise Set 1

2.1: (a) numerical, discrete (b) categorical (c) numerical, continuous (d) numerical, continuous (e) categorical

2.2: (a) discrete (b) continuous (c) discrete (d) discrete

2.3:

Data Set 1:

Question 1: There is one variable in the data set.

Question 2: The variable is categorical.

Question 3: The purpose of the graphical display is to summarize the data distribution.

Appropriate Graphical Display: Bar chart

Data Set 2:

Question 1: There is one variable in the data set.

Question 2: The variable is numerical.

Question 3: The purpose of the graphical display is to compare groups (full-time students or part-time students).

Appropriate Graphical Display: Comparative dotplot, comparative stem-and-leaf displays, and comparative histograms are all appropriate.

Data Set 3:

Question 1: There are two variables in the data set.

Question 2: The variable is numerical.

Question 3: The purpose of the graphical display is to investigate the relationship between two numerical variables.

Appropriate Graphical Display: Scatterplot

Data Set 4:

Question 1: There is one variable in the data set.

Question 2: The variable is categorical.

Question 3: The purpose of a graphical display is to compare groups (faculty, students)

Appropriate Graphical Display: Comparative bar chart

Data Set 5:

Question 1: There is one variable in the data set.

Question 2: The variable is numerical.

Question 3: The purpose of a graphical display is to summarize a data distribution.

Appropriate Graphical Display: Dotplot, stem-and-leaf, and histogram are all appropriate graphical displays.

Section 2.1 Exercise Set 2

2.4: (a) categorical

(b) categorical

(c) numerical – discrete

(d) numerical – continuous

(e) categorical

(f) numerical – continuous

2.5: (a) continuous (b) continuous (c) continuous (d) discrete

2.6: **Data Set 1:**

Question 1: There is one variable in the data set.

Question 2: The variable is numerical.

Question 3: The purpose of a graphical display is to summarize a data distribution.

Appropriate Graphical Display: Dotplot, stem-and-leaf, and histogram are all appropriate graphical displays.

Data Set 2:

Question 1: There is one variable in the data set.

Question 2: The variable is categorical.

Question 3: The purpose of a graphical display is to compare two groups (male and female color choice).

Appropriate Graphical Display: Comparative bar chart

Data Set 3:

Question 1: There are two variables in the data set.

Question 2: The variable is numerical.

Question 3: The purpose of a graphical display is to investigate the relationship between two numerical variables.

Appropriate Graphical Display: Scatterplot

Data Set 4:

Question 1: There is one variable in the data set.

Question 2: The variable is numerical.

Question 3: The purpose of a graphical display is to compare two groups.

Appropriate Graphical Display: Comparative dotplot, comparative stem-and-leaf displays, and comparative histograms are all appropriate.

Data Set 5:

Question 1: There is one variable in the data set.

Question 2: The variable is categorical.

Question 3: The purpose of a graphical display is to summarize a data distribution.

Appropriate Graphical Display: Bar chart.

Additional Exercises for Section 2.1

2.7: (a) numerical (b) numerical (c) categorical (d) numerical (e) categorical

2.8: Discrete: (b); Continuous: (a) and (d)

2.9: (a) categorical (b) numerical (c) numerical (d) categorical

2.10: Discrete: (b); Continuous: (c)

2.11: (a) numerical (b) numerical (c) numerical (d) categorical (e) categorical (f) numerical (g) categorical

2.12:

Question 1: There is one variable in the data set.

Question 2: The variable is categorical.

Question 3: The purpose is to summarize the data distribution.

Appropriate graphical display: Bar chart

2.13:

Question 1: There is one variable in the data set.

Question 2: The variable is numerical.

Question 3: The purpose is to compare groups (male students, female students).

Appropriate graphical display: Comparative dotplot, comparative stem-and-leaf, or histograms are all appropriate.

2.14:

Question 1: There are two variables in the data set.

Question 2: The variable is numerical.

Question 3: The purpose is to investigate the relationship between two numerical variables.

Appropriate graphical display: Scatterplot

2.15:

Question 1: There is one variable in the data set.

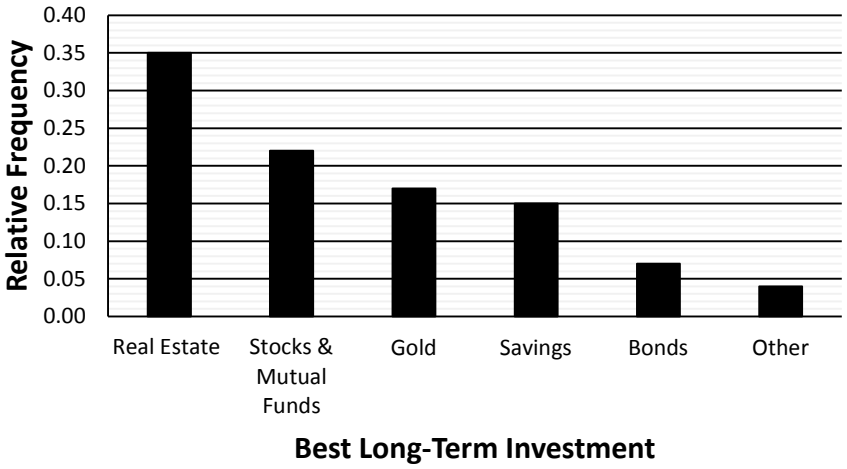
Question 2: The variable is numerical.

Question 3: The purpose of a graphical display is to summarize the data distribution.

Appropriate graphical display: Dotplots, stem-and-leaf plots, and histograms are all appropriate graphical displays.

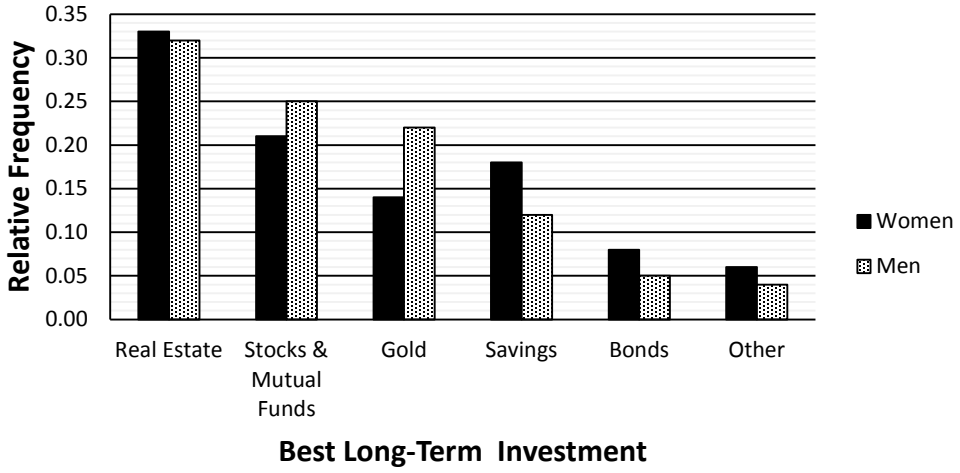
Section 2.2 Exercise Set 1

2.16: (a)



(b) Over half of the responses (57%) were from people who indicated that the best long-term investments were real estate (35%) and stocks & mutual funds (22%). The remaining 43% of respondents indicated that gold (17%), savings (15%), bonds (7%), and other (4%) were the best long-term investments.

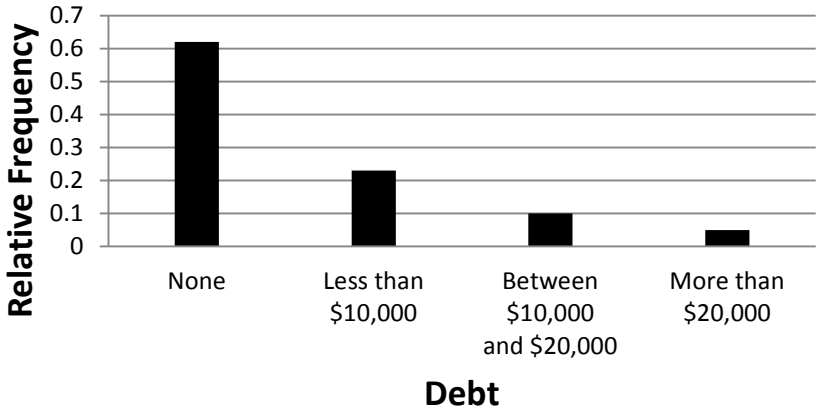
2.17: (a)



(b) In general, women and men have similar rankings for the best long-term investments. However, one notable difference is that women rank saving higher than gold, and men rank gold higher than savings.

Section 2.2 Exercise Set 2

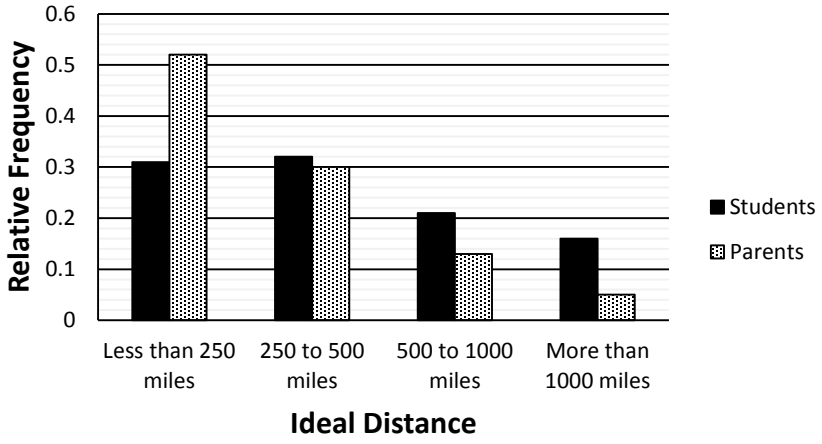
2.18: (a)



(b) Over 60% of students graduating with an AA degree from a public community college in 2008 graduate with no debt. As the amount of debt increases, fewer students reach that debt level. Twenty-three percent of students have less than $10,000 in debt, 10% have between $10,000 and $20,000 in debt, and only 5% have over $20,000 in debt.

2.19: (a) One would want to use relative frequencies when constructing a comparative bar chart to compare ideal distance for students and parents because the number of students (8,347) and number of parents (2,087) are not equal to each other.

(b)

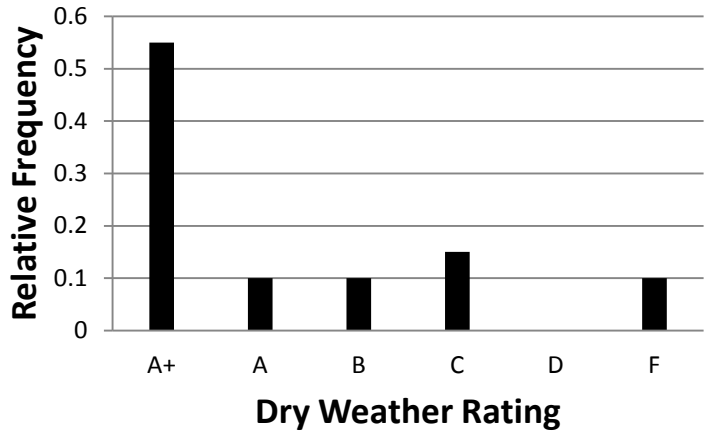


(c) Students tend to want to be farther away from home than their parents would like them to be. In fact, 37% of students want to be over 500 miles from home, whereas only 18% of parents want their children to be over 500 miles from home.

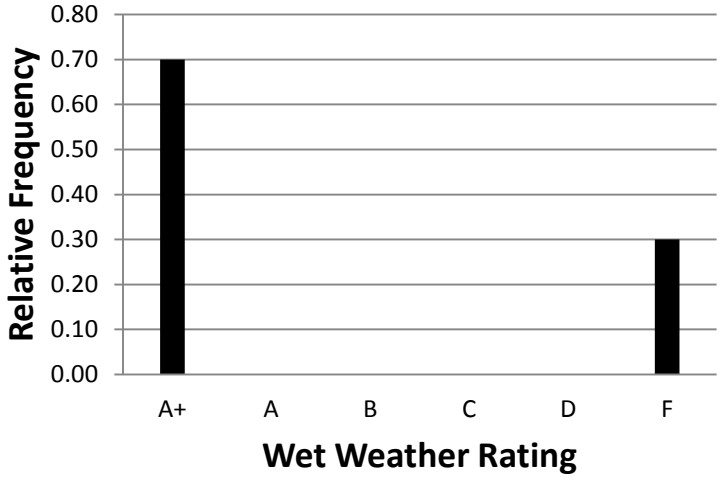
Additional Exercises for Section 2.2

2.20: (a)

Rating	Relative Frequency
A+	11/20 = 0.55
A	2/20 = 0.10
B	2/20 = 0.10
C	3/20 = 0.15
D	0/20 = 0.00
F	2/20 = 0.10



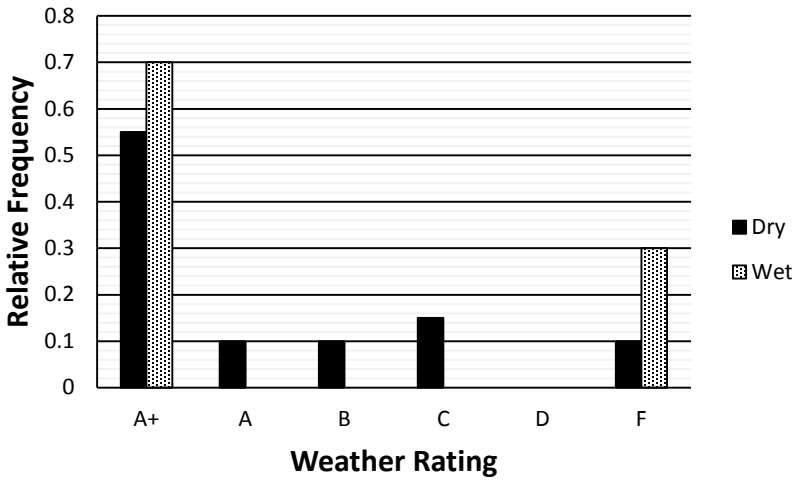
(b)



(c) Seventy-five percent (75%) of the dry weather ratings are B or higher, and 70% of wet weather ratings are B or higher, indicating that dry weather ratings are higher than wet weather ratings. Note that the wet weather ratings are only A+ or F, so the wet weather

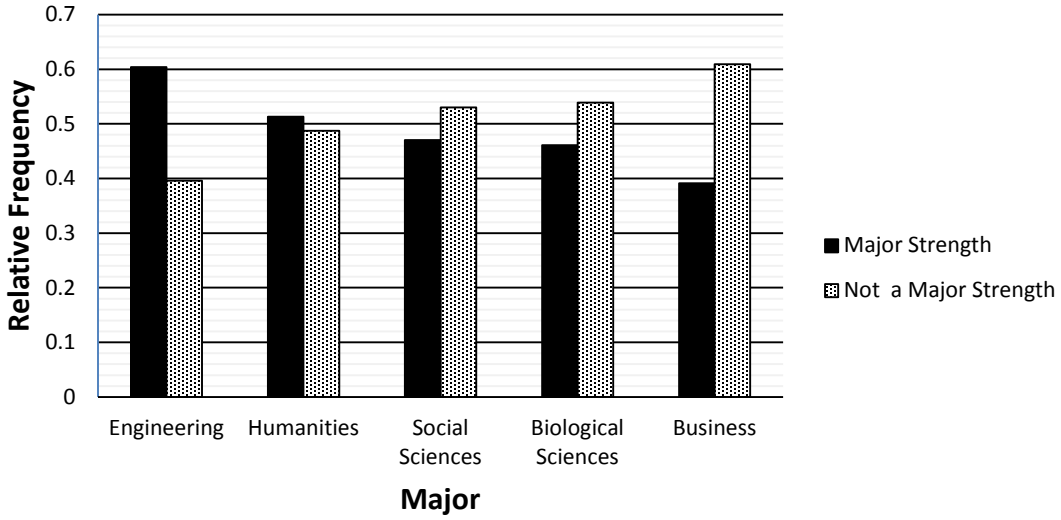
ratings are more extreme than dry weather ratings. If we only consider A+ ratings, then the wet weather ratings tend to be better than dry weather ratings because only 55% of dry weather ratings are A+, compared with 70% of wet weather ratings being A+.

(d)



The wet weather ratings are more extreme than the dry weather ratings (only ratings of A+ and F are found in the wet weather ratings). Dry weather ratings include all rating levels except for D.

2.21:

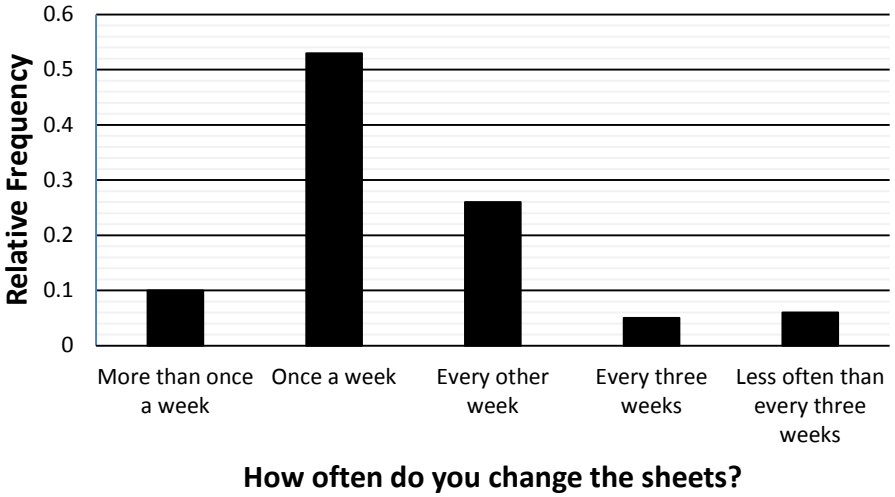


The majors in which a majority of students indicated that critical thinking is "a major strength" are engineering and humanities. A majority of students in other majors (social sciences, biological sciences, and business) indicated that critical thinking is "not a major strength." The greatest disparities between "a major strength" and "not a major strength" occurred with engineering and business majors, and humanities majors had the smallest difference between the two responses.

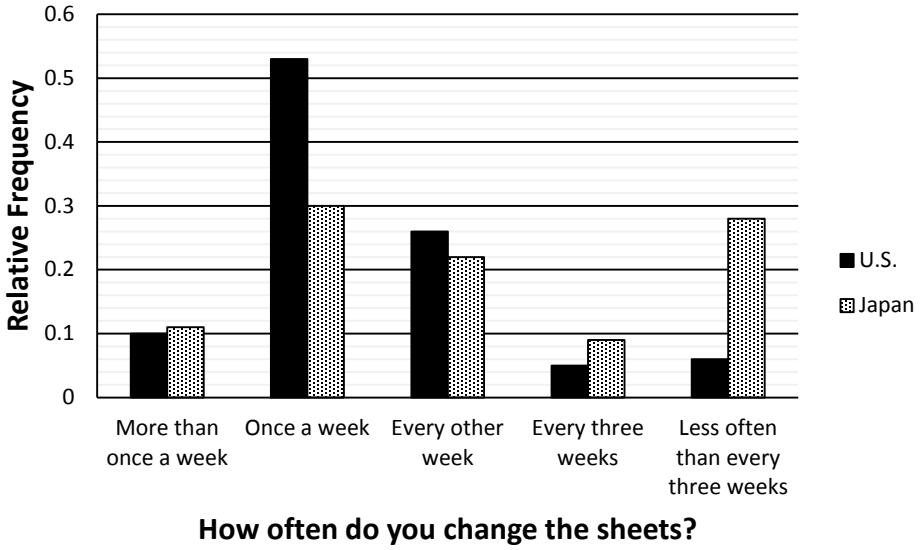
2.22: (a)

How Often	Relative Frequency
More than once a week	0.10
Once a week	0.53
Every other week	0.26
Every three weeks	0.05
Less often than every three weeks	0.06

(b)



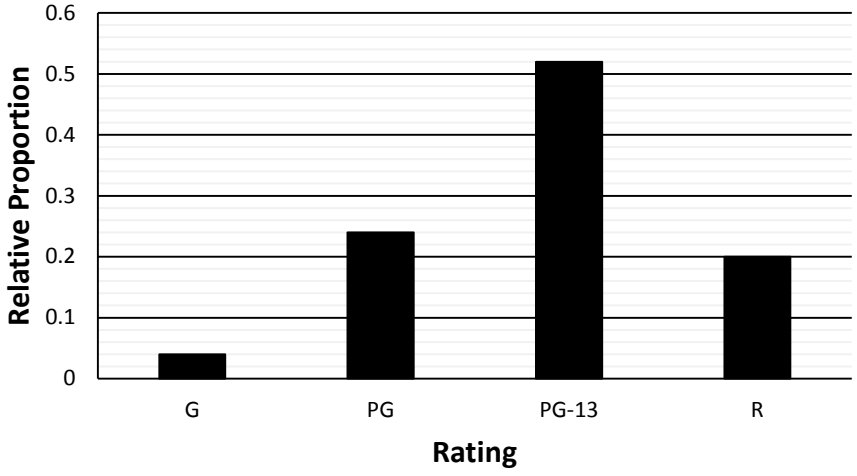
(c)



The biggest differences between the sheet-changing habits of Americans and Japanese are in the "once a week" and the "less often than every three weeks" categories. A much larger percentage of Americans change their sheets once a week than do the Japanese, and a much larger percentage of Japanese change their sheets less often than every three weeks. The remaining categories are fairly close to each other.

2.23: The relative frequency distribution is:

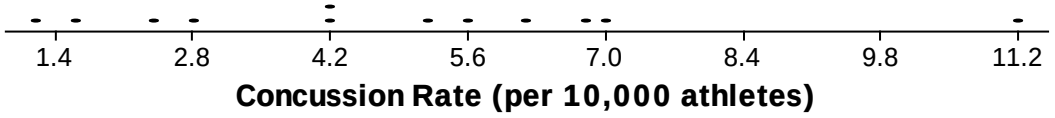
Rating	Relative Frequency
G	1/25 = 0.04
PG	6/25 = 0.24
PG-13	13/25 = 0.52
R	5/25 = 0.20



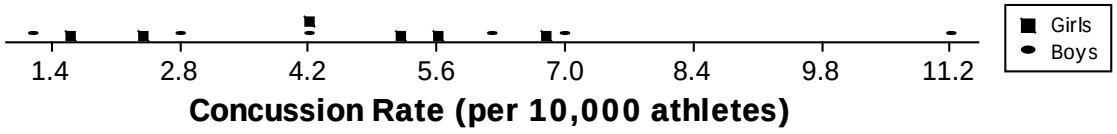
PG-13 is the rating with the highest relative proportion (0.52), followed by PG (0.24), R (0.20), and G (0.04). Seventy-two percent (72%) of the top 25 movies of 2015 are PG-13 or R, and the remaining 28% are rated G or PG.

Section 2.3 Exercise Set 1

2.24: (a) The dotplot below shows the concussion rate (concussions per 10,000 athletes).

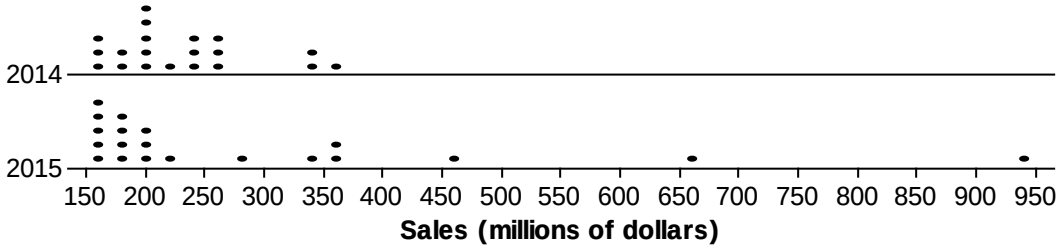


(b) The dotplot below shows the concussion rate (concussions per 10,000 athletes), with different symbols for boys and girls.



The sport with an unusually high (compared to all the other sports) concussion rate is football. Without considering football, the concussion rates for girls' sports is essentially the same as the concussion rate for boys' sports. However, if we consider football, the concussion rate for girls' sports tends to be lower than that for boys' sports.

2.25: (a)



(b) The shape of both distributions is positively skewed. The distribution of 2015 ticket sales is centered at about $280 million, which is higher than the center of the distribution of 2014 ticket sales, which is centered around $230 million. The lowest ticket sales for both 2014 and 2015 are approximately $150 million. Ticket sales for 2014 has a maximum value of approximately $350 million, which is much lower than the highest ticket sales for 2015 of $937 million. The spread between the lowest and highest values for 2014 is

approximately $200 million, which is less than the spread for 2015 of a little under $800 million.

2.26:

```
28 | 8
29 |
30 |
31 |
32 | 8
33 | 0
34 | 178
35 | 00145678899
36 | 238999
37 | 0034566777
38 | 01124558
39 | 00259
40 | 045
41 | 2
42 | 2

Legend: 34|1 = 34.1 years
```

The distribution of median ages is centered at approximately 37 years old, with values ranging from 28.8 to 42.2 years. The distribution is approximately symmetric (possibly slightly negatively skewed), with one possible outlier of 28.8 years.

2.27: (a)

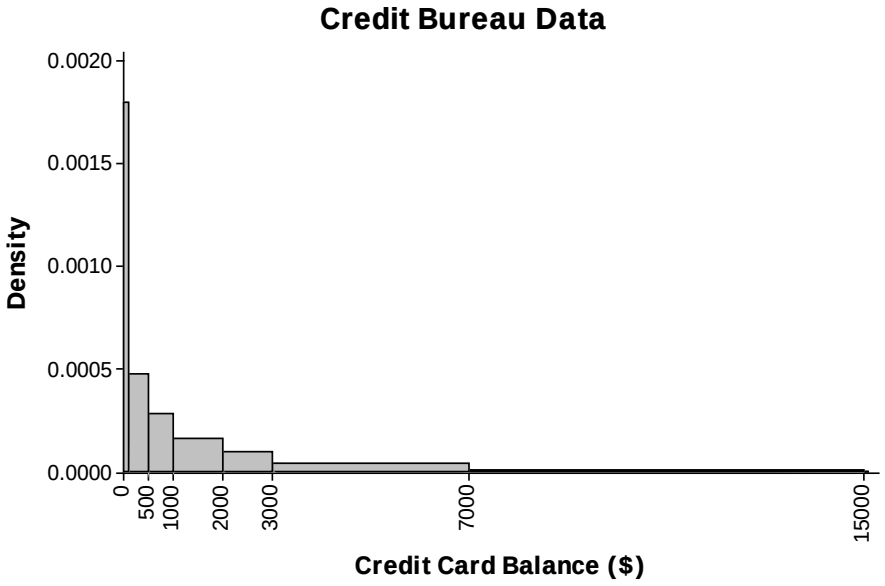
Very Large Urban Area		Large Urban Area
	1	023478
	2	369
8	3	0033589
99	4	0366
8711	5	012355
9730	6	
2	7	
	8	
3	9	

Legend: 4|6 = 46 extra hours per year

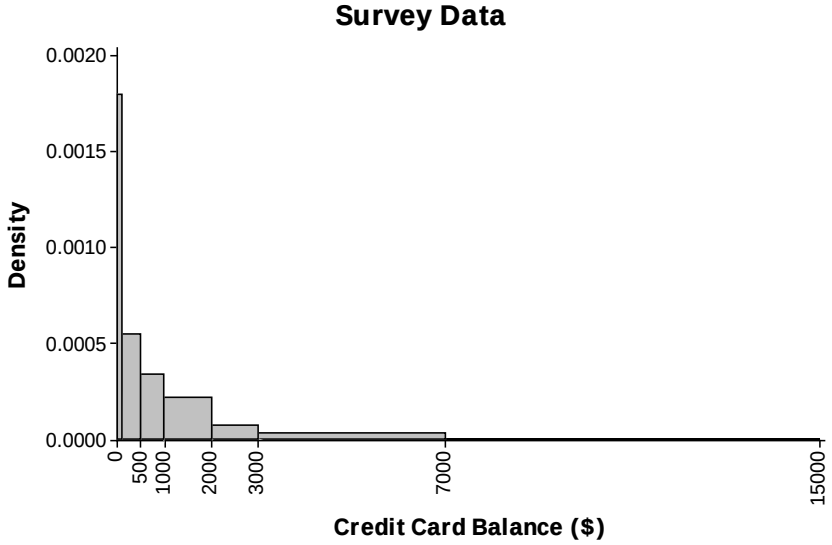
(b) The statement “The larger the urban areas, the greater the extra travel time during peak period travel” is generally consistent with the data. Although there is overlap between the times for the very large and large urban areas, the back-to-back stem-and-leaf plot shows that for the very large urban areas, the extra travel time during peak period is generally

longer than for the large urban areas. The extra travel time during peak period is centered at approximately 58 hours for the very large urban areas, which is higher than the center of approximately 34 hours for the large urban areas.

2.28: (a)



(b)



(c) The histograms are similar in shape. A notable difference is that the Credit Bureau data shows that 7% of students have credit card balances of at least $7000, but no survey respondent indicated a balance of at least $7000.

(d) Yes, it is likely that the high nonresponse rate for the survey may have contributed to the observed difference in the two histograms because students with credit card balances of $7000 or more might be too embarrassed to admit that they have such a high balance.

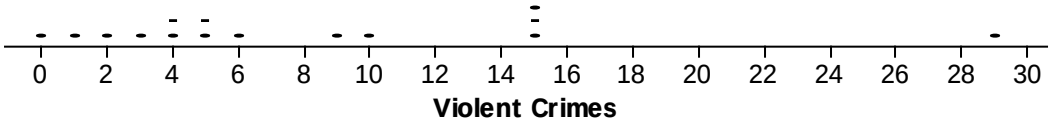
2.29: (a) If the exam is quite easy, the scores would be clustered at the high end of the scale, with a few low scores for the students who did not study. The histogram would be negatively skewed.

(b) If the exam is difficult, the scores would be clustered around a much lower median value. There might be a few high scores, therefore the histogram would be positively skewed.

(c) In this case, the histogram would be bimodal, with a cluster of high scores and a cluster of low scores.

Section 2.3 Exercise Set 2

2.30: (a)

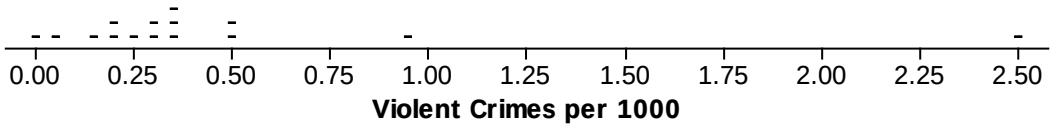


It appears as if there are two clusters of schools, those with 10 or fewer violent crimes, those with 15 or more. The schools that stand out from the rest are those with 15 or more violent crimes, namely, Florida State University, the University of Florida, the University of South Florida, and the University of Central Florida.

(b) The violent crime rates for all the schools are shown in the table below, followed by the corresponding dotplot.

University/College	Violent Crimes per 1000
Edison State College	0.195
Florida A&M University	0.931
Florida Atlantic University	0.163
Florida Gulf Coast University	0.355
Florida International University	0.189
Florida State University	0.367
New College of Florida	2.522
Pensacola State College	0.000

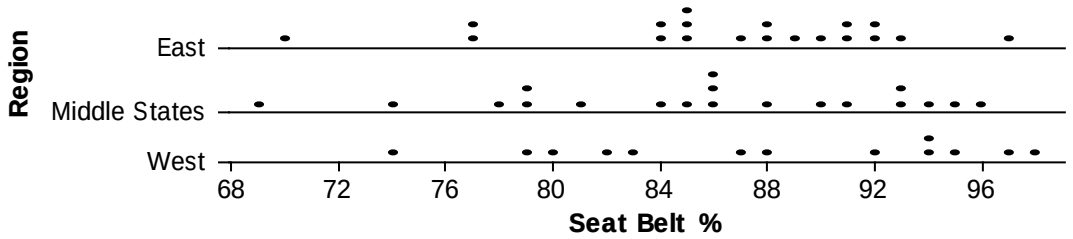
Santa Fe College	0.066
Tallahassee Community College	0.296
University of Central Florida	0.487
University of Florida	0.301
University of North Florida	0.249
University of South Florida	0.325
University of West Florida	0.477



No, the same schools do not stand out as unusual. Now only two schools seem to be unusual, namely, Florida A&M University (0.931 violent crimes per 1000) and New College of Florida (2.522 violent crimes per 1000).

(c) A typical (median) number of violent crimes among the Florida universities and colleges is 5. The dotplot shows that there are two groups of violent crime numbers, those less than or equal to 10, and those above 15. Overall, the number of violent crimes ranged between 0 and 29. Considering the crimes per 1,000 students, a typical (median) value is 0.3, with values that range between 0 and 2.522 violent crimes per 1,000 students. Both dotplots show that most schools on the list have relatively few violent crimes, shown by the higher density of violent crimes (in both raw counts and crimes per 1,000 students) at the low end of the scale. However, there are two schools (as noted in part (b) that have unusually high crimes per 1,000 students.

2.31: (a)



(b) There are no striking differences in seat belt percent for the three geographical regions. The distributions of seat belt percent for the three regions are similar, with the Middle States region having the lowest seat belt percent, and the Middle States and Eastern regions being similar to each other.

2.32: (a)

```
0 | 8
1 | 03455689
2 | 4555778
3 | 1269

Legend:  0|8 = 8% cardholders affected
```

(b) There are no unusually high or low percentages of cardholders affected by fraud. One value, 8%, is slightly lower than the others, and is the only single-digit percentage.

2.33: (a)

```
 0 | 0112445788
 1 | 135679
 2 | 011126
 3 | 0588
 4 | 0000348
 5 | 122344469
 6 | 68
 7 | 18
 8 | 45
 9 | 2
10 |
11 | 7
12 | 5

Legend: 7|1 = 7.1 percent change
```

(b) A typical percent change is approximately 3.8%. The distribution is positively skewed. There appear to be two outliers, one at 11.7% (District of Columbia), and the other at 12.5% (North Dakota).

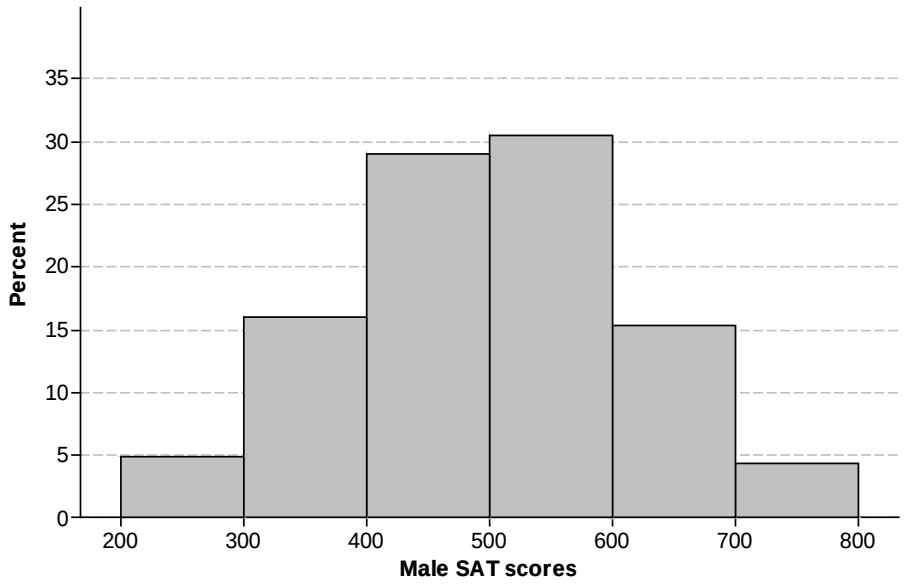
(c)

East		West
8875442110	0	
97651	1	3
62110	2	1
850	3	8
800	4	0034
9443	5	12246
	6	68
8	7	1
	8	45
	9	2
	10	
7	11	
	12	5

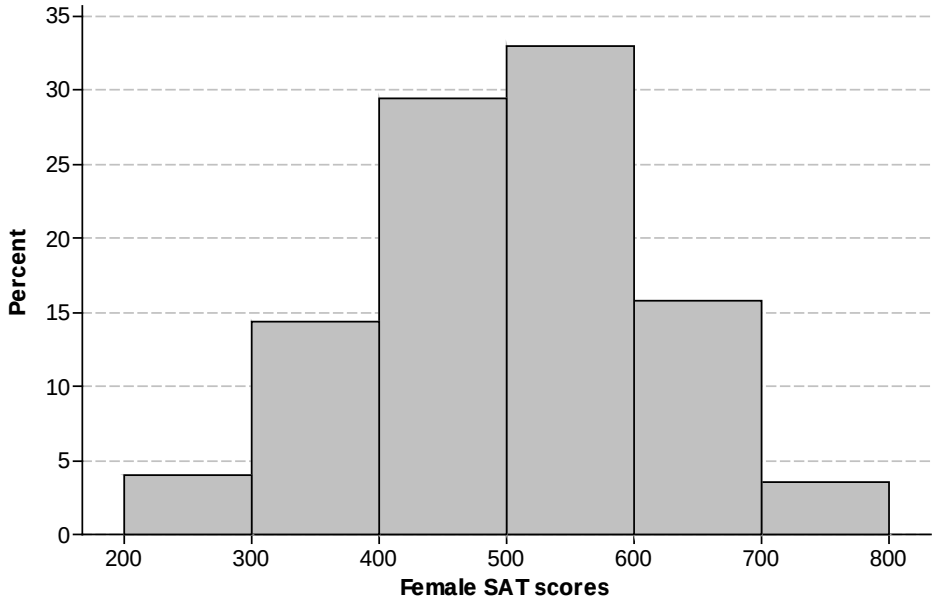
Legend: 1|3 = 13 percent change

In general, Western states have seen more growth than Eastern states. A typical percentage change for Western states is approximately 5.5%, which is greater than the typical percentage change for Eastern states of approximately 2.7%. In addition, each region has one large outlier (11.7% in the East (District of Columbia), and 12.5% in the West (North Dakota)). Both regions have roughly the same amount of variability. Specifically, the percentage change for Western states varies between 1.3% and 12.5%, which is about the same as for Eastern states, which varies between 0% and 11.7%. The Eastern distribution is positively skewed, and the Western distribution is roughly symmetric. Both distributions are unimodal, with a peak in the 0-1% interval for Eastern states and in the 5% interval for Western states.

2.34: (a)



(b)



(c) The distributions of male and female SAT scores are both approximately symmetric, and centered in the 500-600 interval. Both distributions range between the same values (200 to 800).

2.35: (a) Symmetric frequency distribution based on 70 observations.

Class Interval	Frequency
100 to <120	5
120 to <140	15
140 to <160	30
160 to <180	15
180 to <200	5

(b) Bimodal frequency distribution based on 70 observations.

Class Interval	Frequency
100 to <120	10
120 to <140	20
140 to <160	10
160 to <180	20
180 to <200	10

(c) Positively skewed frequency distribution based on 70 observations.

Class Interval	Frequency
100 to <120	30
120 to <140	20
140 to <160	10
160 to <180	5
180 to <200	5

(d) Negatively skewed frequency distribution based on 70 observations.

Class Interval	Frequency
100 to <120	5
120 to <140	5
140 to <160	10
160 to <180	20
180 to <200	30

Additional Exercises for Section 2.3

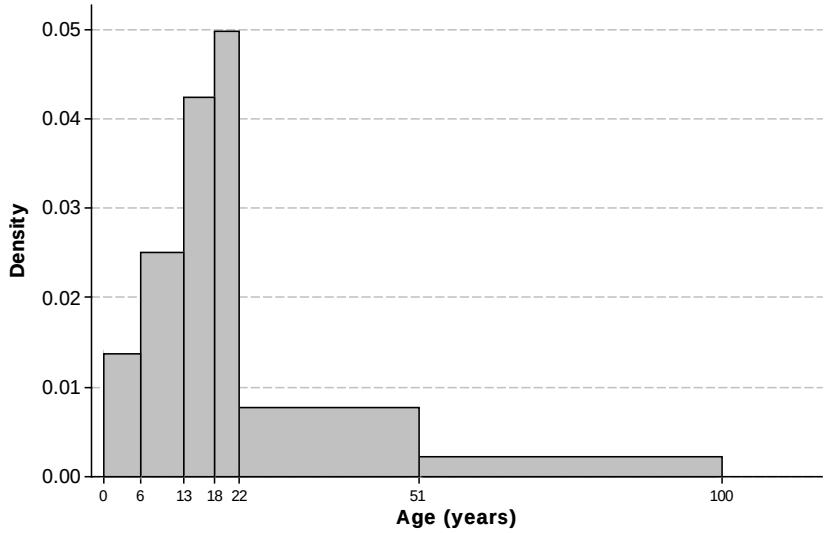
2.36: (a)

With Seat Belt Law Enforcement		Without Seat Belt Law Enforcement
43	0	
9988877766	0	599999
444333221000	1	0013334
87766555	1	57
30	2	2
	2	5

Legend: 1|3 = 13 traffic deaths per 100,000 population

(b) The distributions differ in shape, with the "with seat belt law enforcement" fatality rate being roughly symmetric, and the "without seat belt law enforcement" fatality rate being positively skewed. The center of the "with seat belt law enforcement" fatality rate distribution is approximately 12, and the center of the "without seat belt law enforcement" fatality rate is approximately 11. Both distributions have similar spreads of fatality rates.

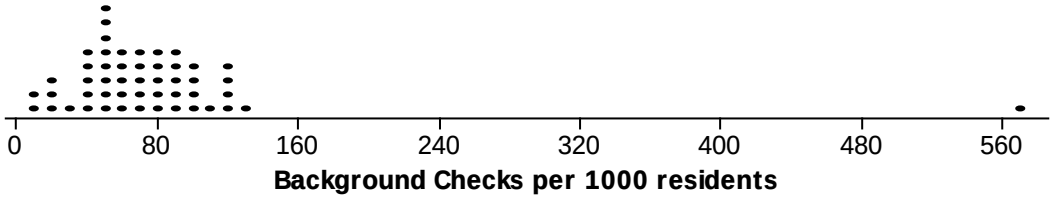
2.37:



Given that the eye is naturally drawn to large areas, the incorrect histogram exaggerates the proportion of clients in the wider age groups.

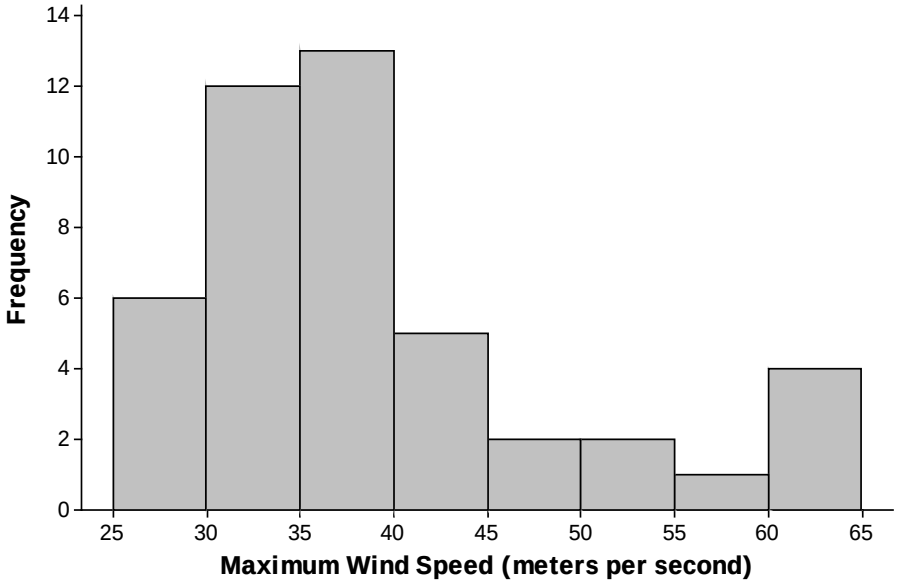
2.38: (a) The dotplot is skewed right (positively skewed), with seven states standing out as being unusual (Pennsylvania, Florida, North Carolina, Illinois, Texas, California, and Kentucky).

(b)

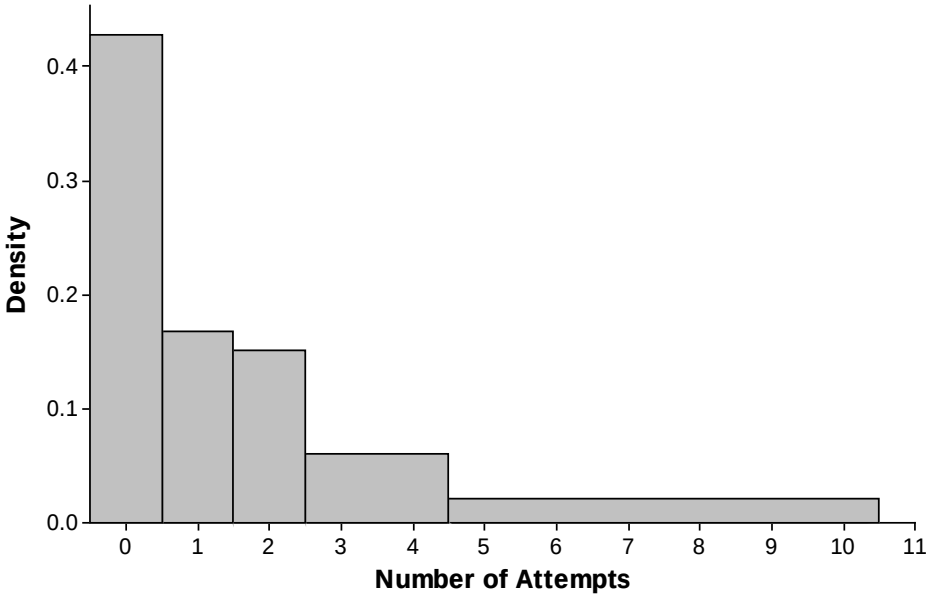


(c) The dotplots differ in that the dotplot in (b) shows only one unusual observation (Kentucky), compared with seven unusual values in the dotplot given in part (a). This indicates that, when one takes population size into consideration, most states are comparable in terms of background checks, with the exception of Kentucky.

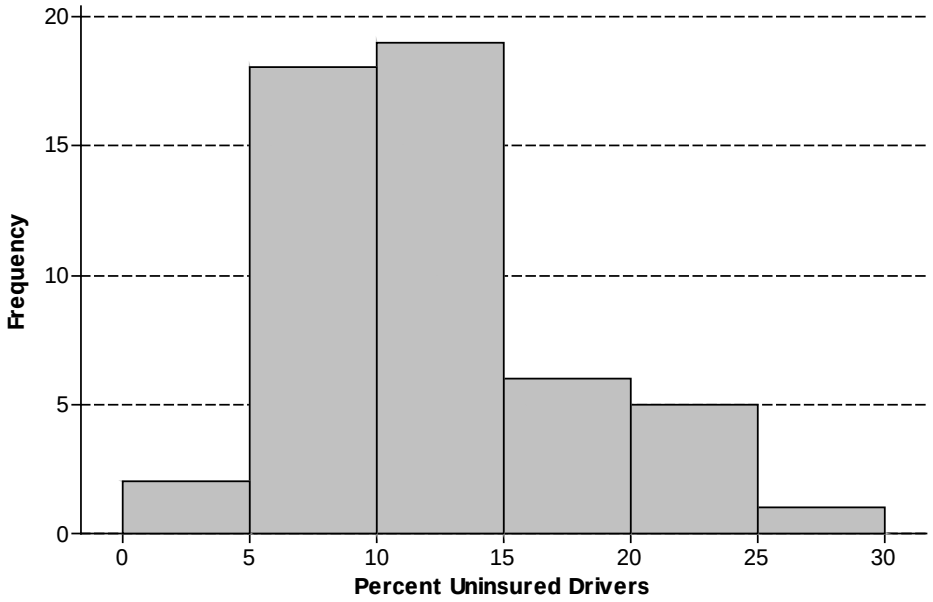
2.39: The distribution of wind speed is positively skewed and bimodal. There are peaks in the 35-40 m/s and 60-65 m/s intervals.



2.40:



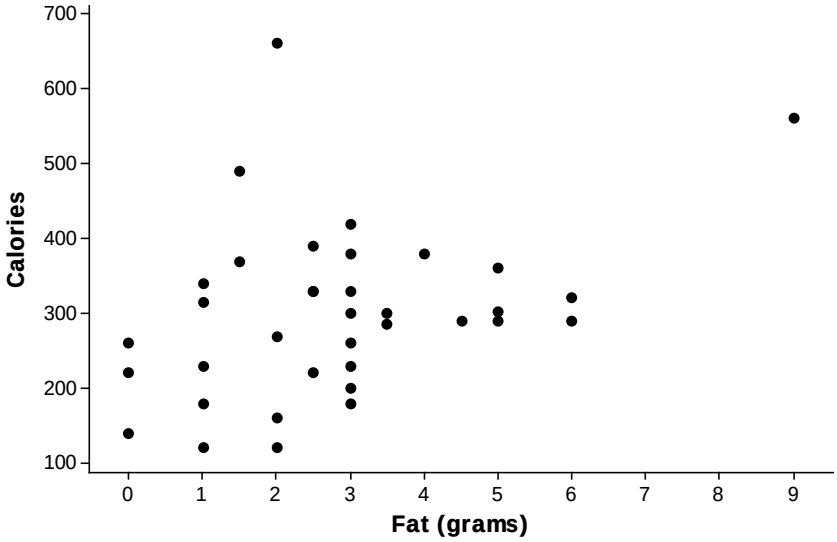
2.41:



The distribution of the percent of uninsured drivers is positively skewed, centered in the 10-15% interval. The percent uninsured drivers varies between a low of 3.9% and a high of 25.9%. Approximately 75% of states have 15% or less uninsured drivers.

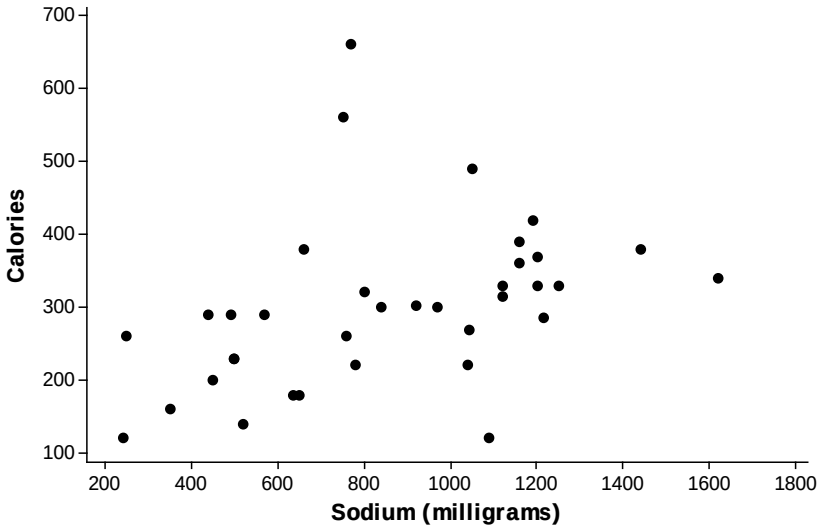
Section 2.4 Exercise Set 1

2.42: (a)



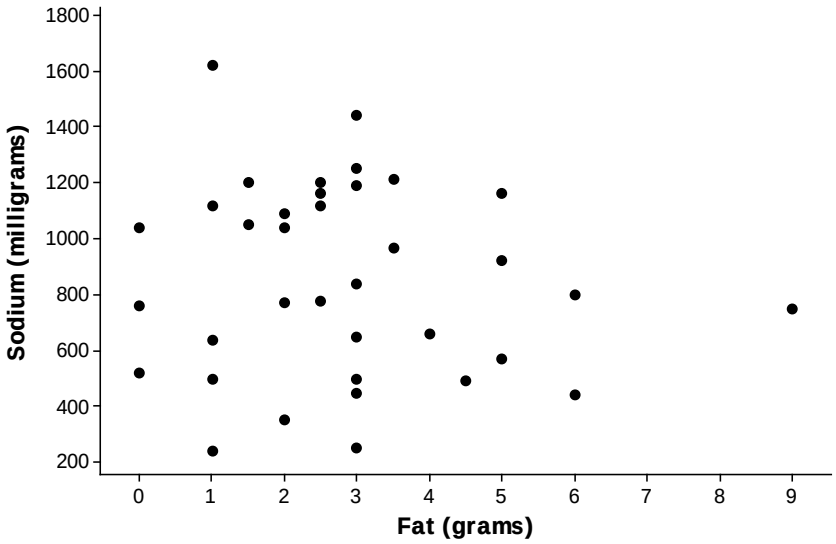
The scatterplot shows the expected positive relationship between grams of fat and calories. The relationship is weak.

(b)



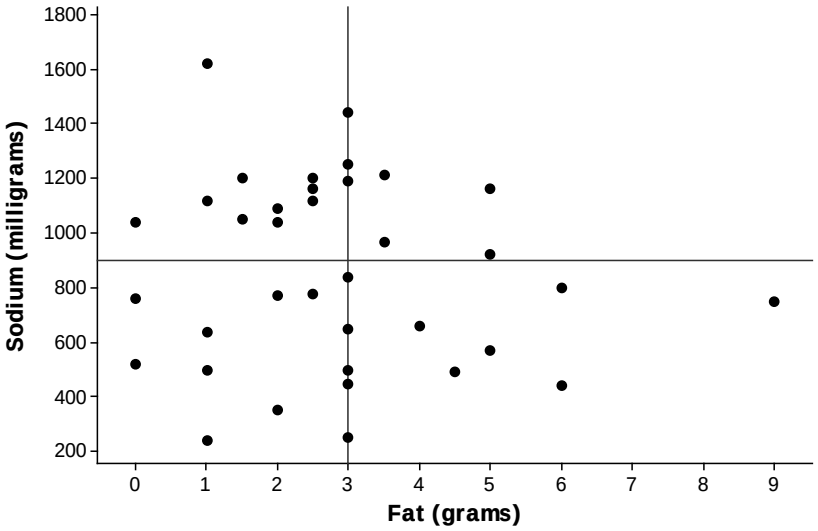
As was observed in the calories vs. fat scatterplot, there is also a weak, positive relationship between calories and sodium. The relationship between calories and sodium appears to be a little stronger than the calories vs. fat relationship.

(c)



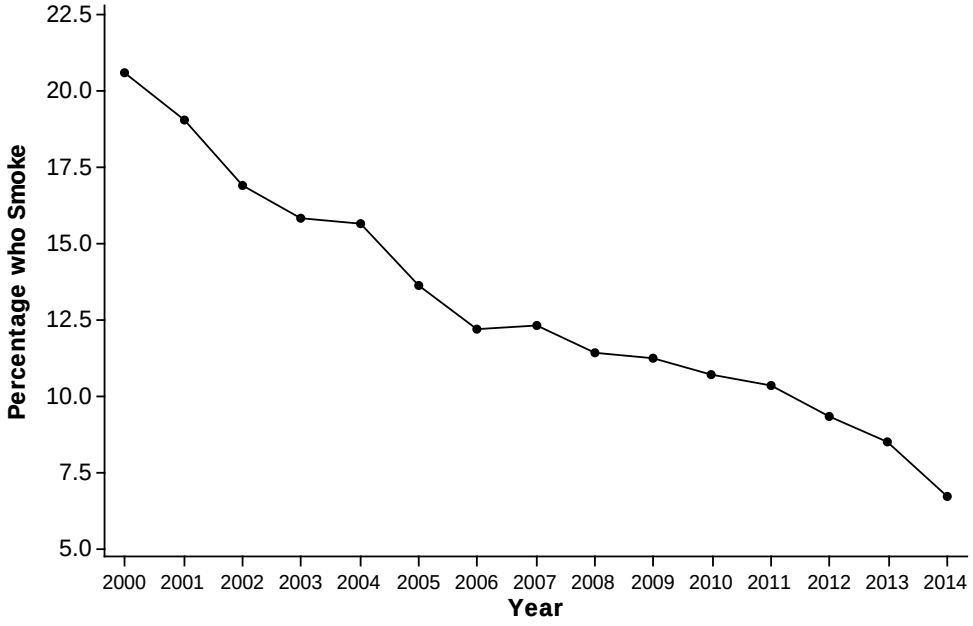
There is no apparent relationship between sodium and fat.

(d)



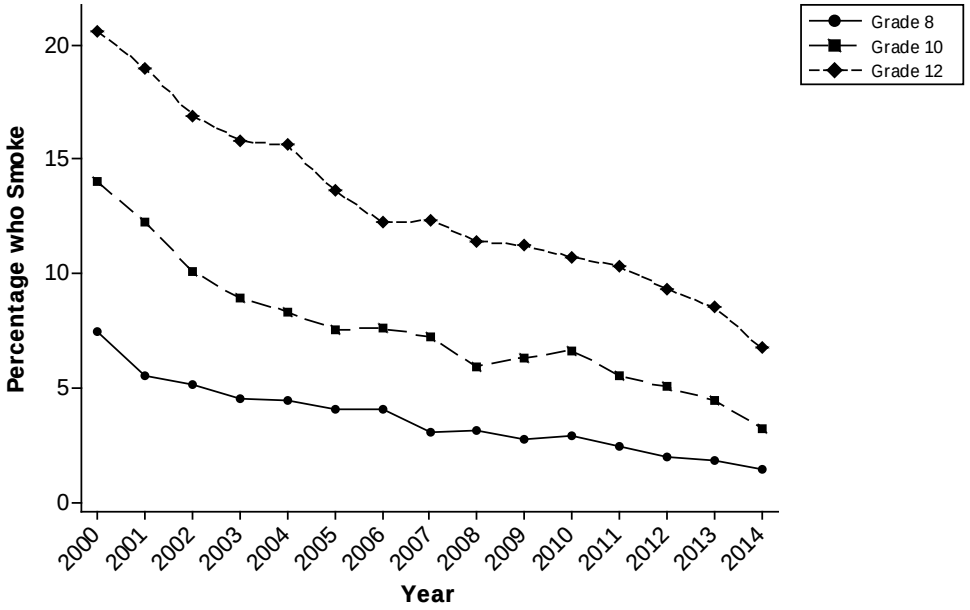
The lower left quadrant corresponds to healthier fast-food choices. This quadrant corresponds to food items with fewer than 3 grams of fat, and fewer than 900 milligrams of sodium, which are considered the healthier choices.

2.43:



There has been a downward trend in the percent of grade 12 students who smoke daily, from a high of 20.6% in 2000 down to 6.7% in 2014.

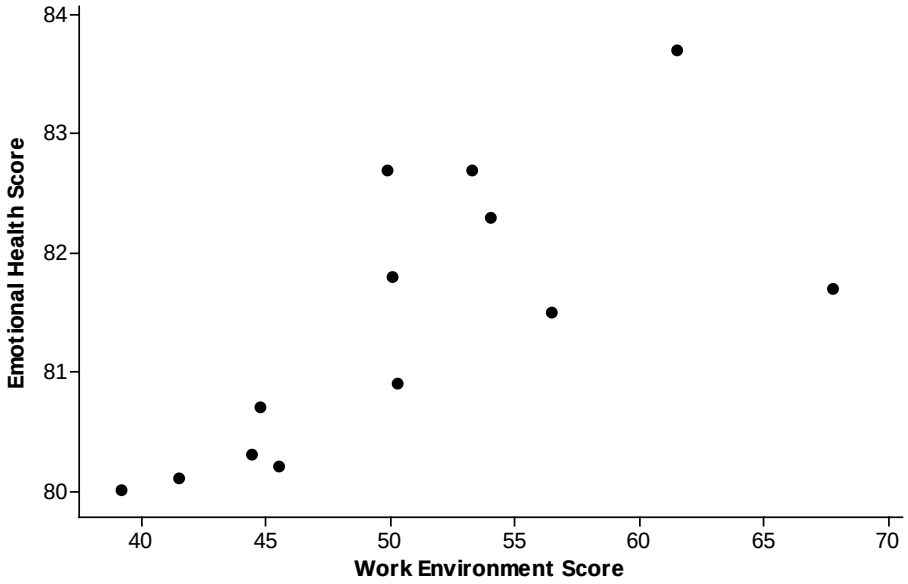
(b)



(c) There has been a downward trend in the percent of students who smoke regardless of grade level. For all the years in the data set, grade 8 students had a lower percentage of smokers than the other grades. The percentage of grade 10 students who smoke was between the percentages of grade 8 and grade 12, and the percentage of grade 12 students who smoke was the highest for all years in the data set. In addition to the maximum and minimum indicated in part (a), the percentages of grade 10 students who smoke fell from 14% in 2000 to 3.2% in 2014. Similarly, the percentage of grade 8 students who smoke fell from 7.4% in 2000 to 1.4% in 2014.

Section 2.4 Exercise Set 2

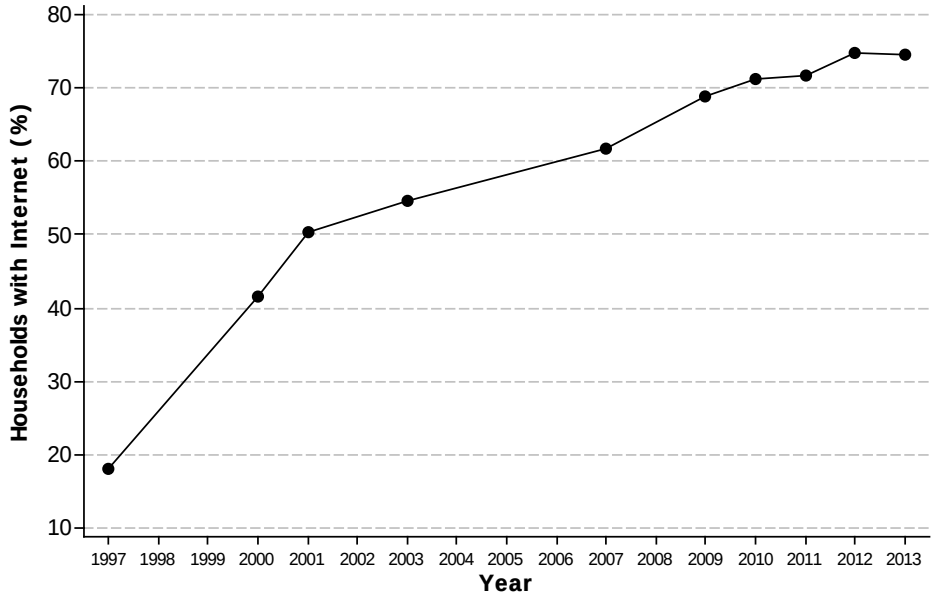
2.44: (a)



(b) There appears to be a positive association between emotional health score and work environment score. Yes, occupations with a higher work environment score tend to be associated with higher emotional health scores. There is one unusual value, namely, business owner, with a high work environment score and a middle environmental health score.

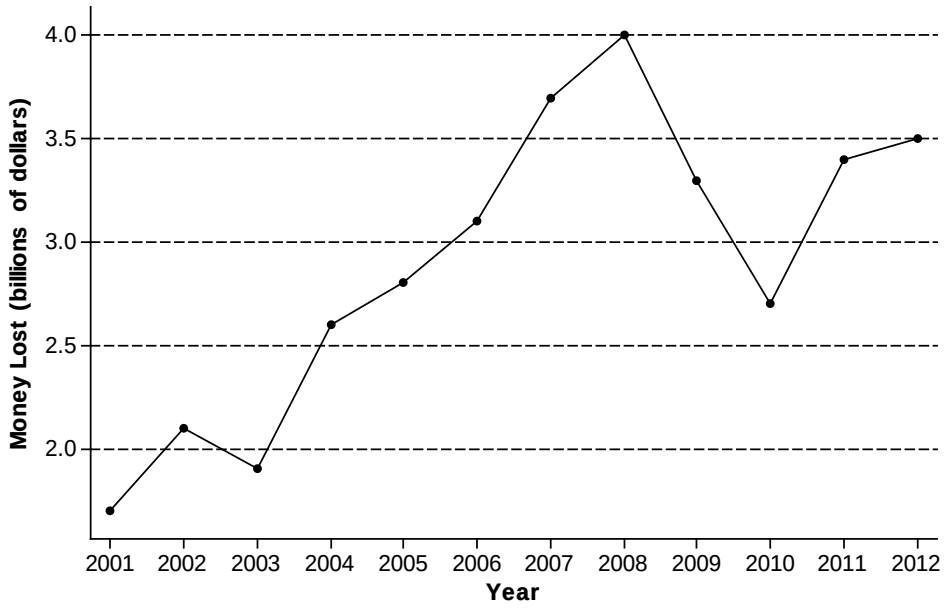
(c) The location of the point corresponding to teacher (work environment score = 49.9, emotional health score = 82.7) is consistent with the title of the article. Teachers have one of the highest emotional health score, but just a moderate (middle) work environment score.

2.45:



The percentage of households with internet access has increased over time, but not at a steady rate. The rate of growth from 1997 to 2001 was faster that the rate of growth from 2001 to 2013. This is shown on the time series plot by the steeper line from 1997 to 2001 compared with 2001 to 2013.

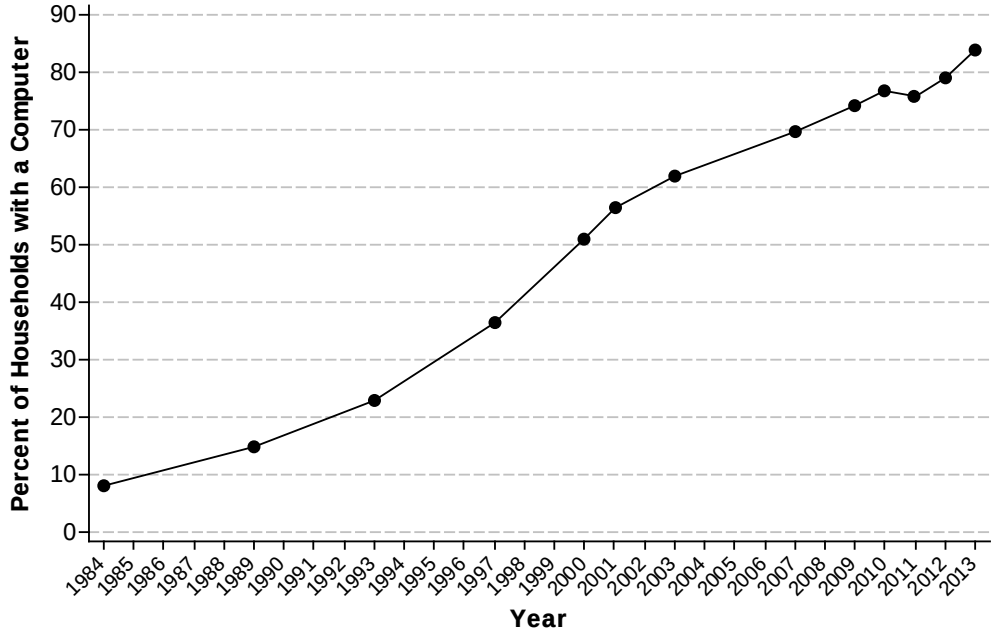
2.46:



The amount of money lost to online fraud has generally increased over time, but not at a steady rate. There are periods of time when the amount of money lost to online fraud has decreased (line segments with negative slopes), and other times when the amount of money lost has increased (line segments with positive slopes).

Additional Exercises for Section 2.4

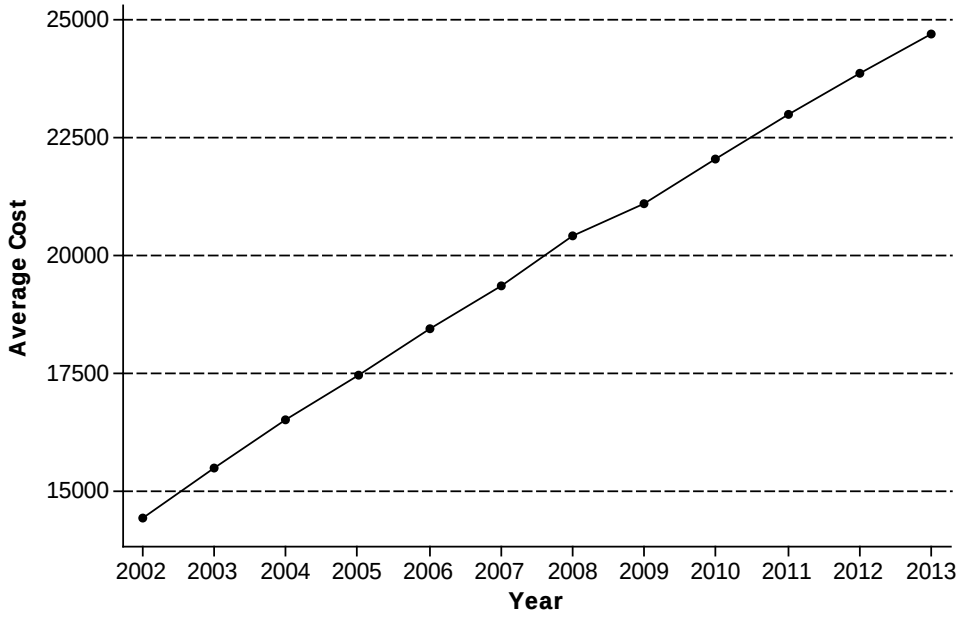
2.47: (a)



(b) The percentage of households with computers has increased over time, from a low of approximately 8% in 1984 to nearly 84% in 2013. The rate of increase of the percentage of households with a computer increased over time from 1984 to 2001, and the rate of increase remained roughly constant from 2001 to 2013.

2.48: There is a relatively weak positive relationship between price and overall score. More expensive laptops tend to earn higher overall scores, but this doesn't hold true for all laptops in this data set.

2.49:



There is a strong, positive trend in the average cost of per year for tuition, fees, and room and board for four-year public institutions in the U.S. The average cost has steadily increased from a low of $14,439 per year in 2002 to a high of $24,706 per year in 2013.

2.50: There is a strong negative linear association between racket resonance frequency and sum of peak-to-peak accelerations. There are two rackets whose data points follow the same trend as the remaining data points, but are separated from the other rackets. Those two points have much higher frequencies than the others, and lower sum of peak-to-peak accelerations.

Section 2.5 Exercise Set 1

2.51: (a) The response is a categorical variable.

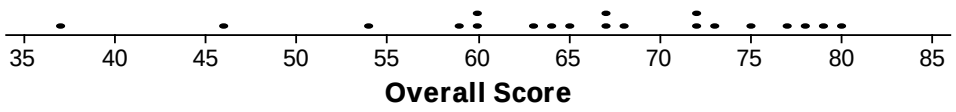
(b) A bar chart was used rather than a dotplot because the response is a categorical variable, and dotplots are used for numerical responses.

(c) This is not a correct representation of the response data because the percent values add up to over 100% (they add to 107%).

2.52: (a) Overall score is numerical. Grade is categorical. (b) The figure is equivalent to a segmented bar graph because the bar is divided into segments, with different shaded regions representing the different grades (“Top of the Class,” “Passing,” “Barely Passing,” and “Failing”), and the height of each segment is equal to the frequency for that category

(for example, there are 5 school districts in the Top of the Class category, 3 in the Passing category, and so on), thus making the area of each shaded region proportional to the relative frequencies for each grade. These conditions satisfy the definition of a segmented bar chart given in the textbook.

(c)



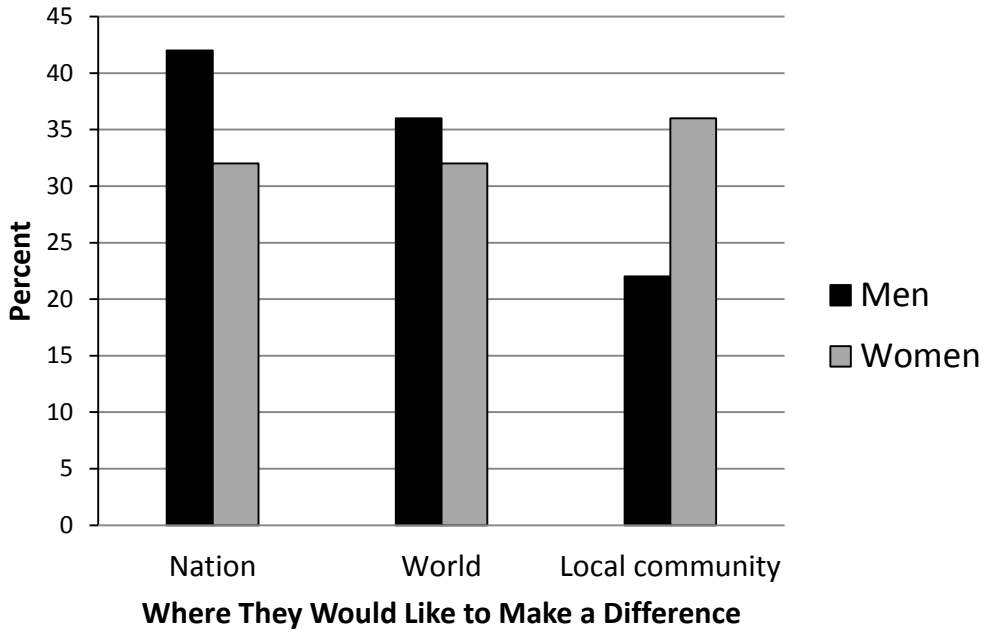
One alternate assignment of grades is to require that "Top of the Class" schools earn grades of 72 or higher, "Passing" schools earn between 66 and 71, "Barely Passing" schools earn between 61 and 65, and "Failing" schools earn 60 or below. This alternative is suggested because there appear to be clusters of dots on the dotplot that correspond to the alternate suggested ranges.

2.53: Answers may vary. Students and teachers differ in their opinions regarding use of Facebook and students being allowed to report on controversial issues in their student newspaper. For both questions, 61% of students agree with the statements given. However, the percentages of teachers who agree are much lower (39% and 29%, respectively, for the two statements). Similarly, a larger percentage of teachers disagree with the two statements (57% and 67%, respectively).

Section 2.5 Exercise Set 2

2.54: This graphical display is a segmented bar chart.

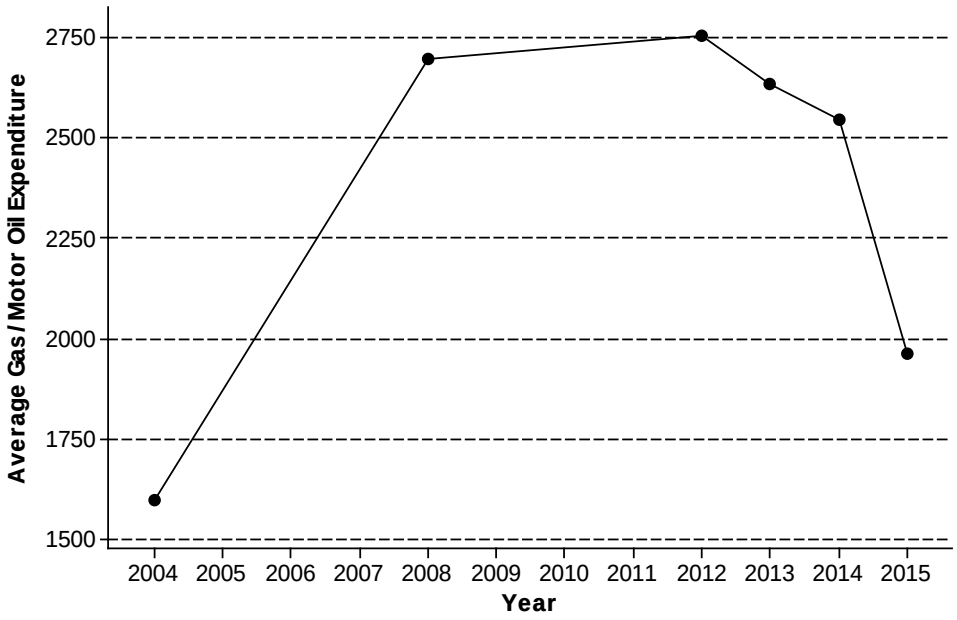
2.55: (a)



(b) The USA Today graph is constructed such that the height of the sleeves and the hands combined are in the correct proportions. Interpretation could be difficult due to the fact that the sleeves look like bars in a traditional bar chart, and readers might think those are drawn to the correct heights (which they are not). Note as well that the hands all have the same areas and that the heights of the sleeve-hand combinations are in proportion to their relative frequencies. However, the area of each sleeve-hand combination is not proportional to the corresponding relative frequency, which violates the area proportional to frequency principle.

2.56:

(a)

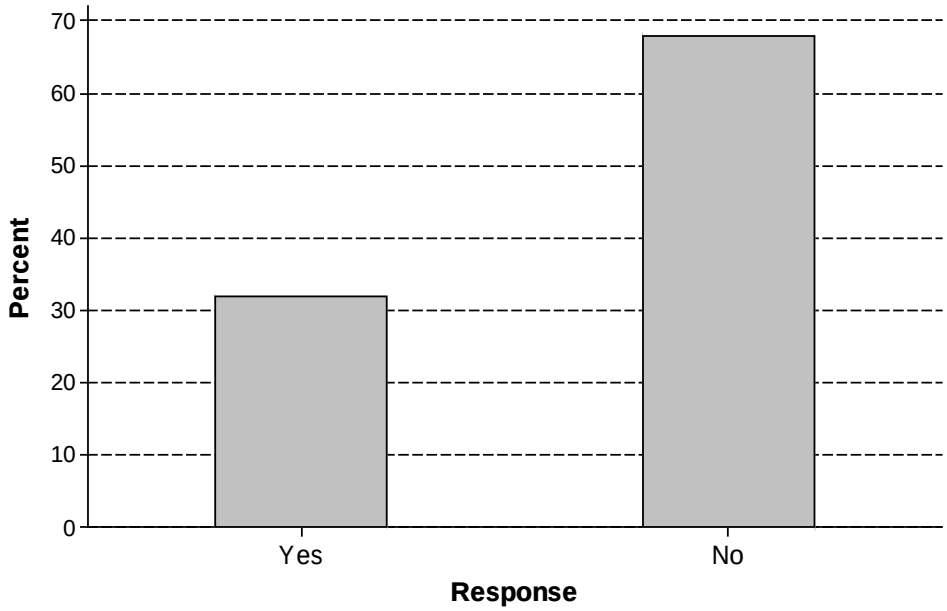


(b) The USA Today graph gives the impression of a fairly rapid increase from 2004 to 2008, when the increase actually happened over a period of four years. The time series plot more accurately represents the change over time.

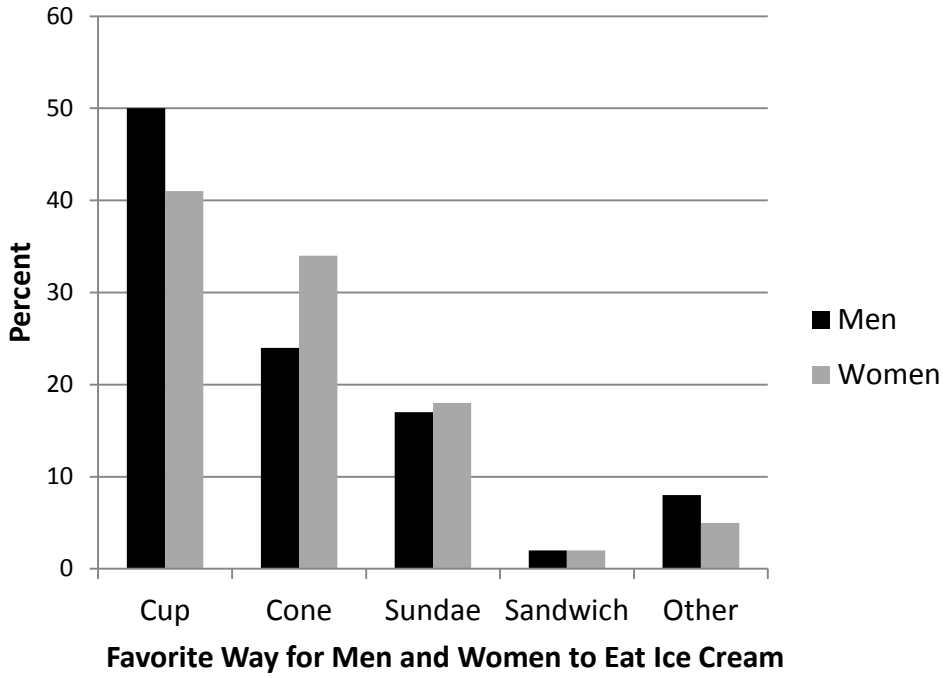
Additional Exercises for Section 2.5

2.57: (a) The areas in the display are not proportional to the values they represent. The "no" category seems to represent more than 68%.

(b)



2.58: This is not an effective summary of the data. This plot violates the area principle, and can give incorrect impressions of the percentages. The comparative bar chart below is more appropriate.



2.59: (a) This graph is similar to a pie chart because the portions of the house add up to 100%, as they should in a pie chart.

(b) Chart III is the segmented bar chart that is a graph of the data used to create the USA Today Snapshot graph. The height of the "Yes" portion of the bar in Chart III has a height that corresponds to 23%. The "No" portion of the bar in Chart III has a height that corresponds to 48%, which gives a total height of 71% for the combined "Yes" and "No" responses. Finally, the "Not sure" portion of the bar makes up the remaining 29%, making the total to 100%.

Chapter 2: Are You Ready to Move On?

2.60: (a) Gender, Brand of motorcycle, Telephone area code

(b) Number of previous motorcycles owned by purchaser

(c) Bar chart

(d) Dotplot

2.61:

Data Set 1

Question 1: There is one variable in the data set.

Question 2: The data set is numerical.

Question 3: The purpose of the graphical display is to summarize the data distribution.

Appropriate Graphical Display: Dotplot, stem-and-leaf display, or histogram

Data Set 2

Question 1: There are two variables in the data set.

Question 2: The data set is numerical.

Question 3: The purpose of the graphical display is to investigate the relationship between two numerical variables.

Appropriate Graphical Display: Scatterplot

Data Set 3

Question 1: There are two variables in the data set.

Question 2: The data set is categorical.

Question 3: The purpose of the graphical display is to compare groups.

Appropriate Graphical Display: Comparative Bar Chart

Data Set 4

Question 1: There is one variable in the data set.

Question 2: The data set is categorical.

Question 3: The purpose of the graphical display is to summarize the data distribution.

Appropriate Graphical Display: Bar chart

Data Set 5

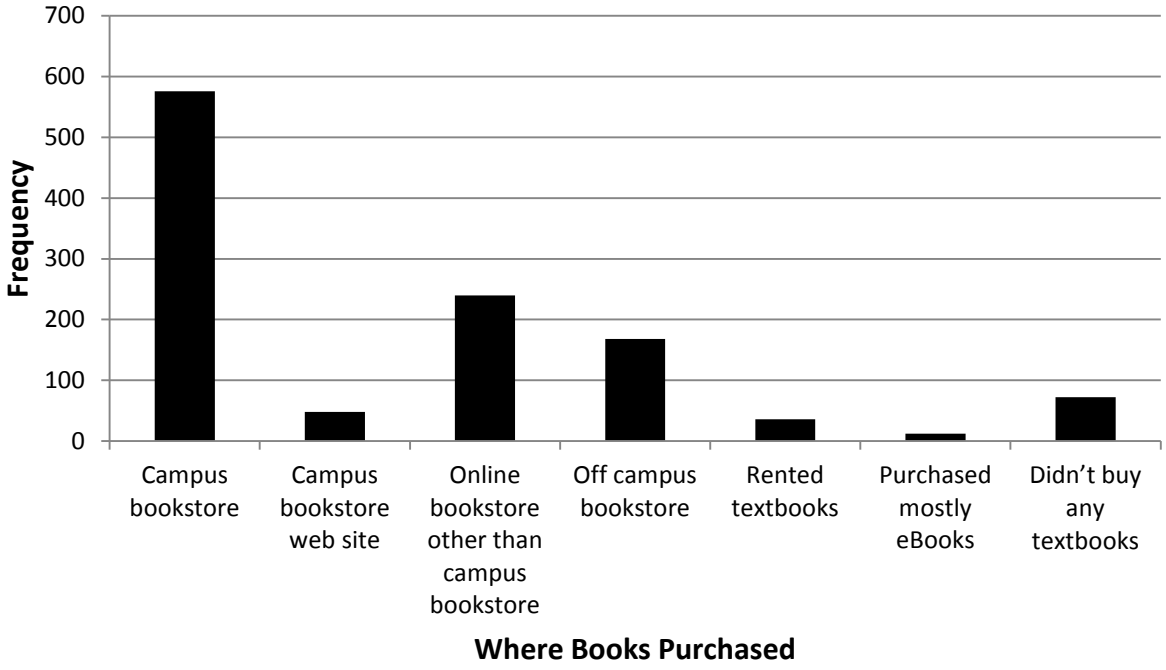
Question 1: There is one variable in the data set.

Question 2: The data set is numerical.

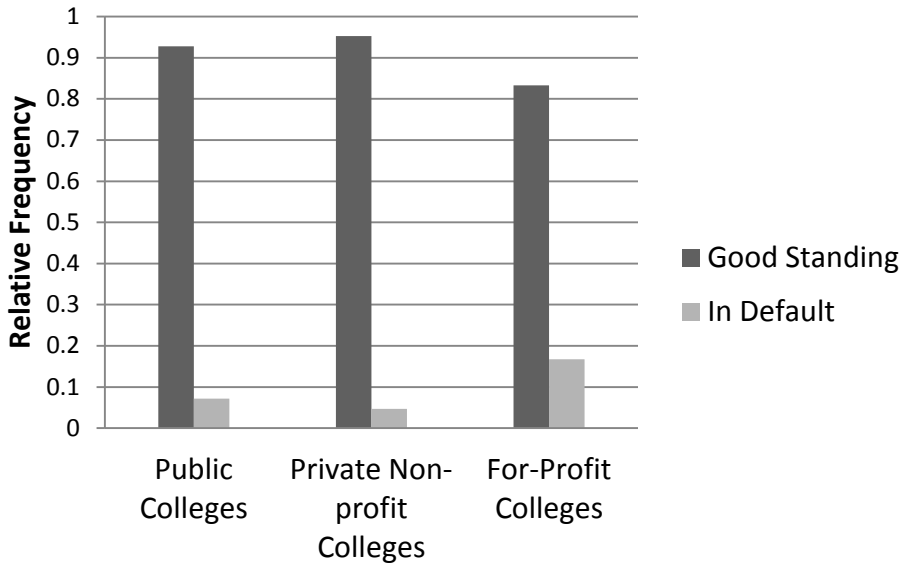
Question 3: The purpose of the graphical display is to compare groups.

Appropriate Graphical Display: Comparative dotplot, comparative stem-and-leaf display.

2.62: The bar chart shows that nearly half of students (48%) are purchasing their books primarily at campus bookstores. Additionally, 20% of students are purchasing books from online bookstores other than campus bookstores, and 14% of students are purchasing at off campus bookstores. The remaining categories each had fewer than 6% of students purchasing from those places.



2.63: (a)



(b) The aspect of the comparative bar chart that supports the statement "those who attended for-profit schools were more likely to default than those who attended public or private non-profit schools" is that the "In Default" bar in the "For-Profit Colleges" category is taller than either of the other "In Default" bars.

2.64: (a) I would use rate per 10,000 flights data to rank airlines based on flights delayed on the tarmac for at least 3 hours. The reason for this is because the airlines are different sizes, and have different total numbers of flights. In order to compare groups of different sizes, the airlines must be put on the same scale, namely, rate per 10,000 flights.

(b) Of the 17 airlines in the study, 13 had fewer than 2 delays per 10,000 flights and, there are 2 airlines that had between 2.5 and 3 delays per 10,000 flights. Two airlines, however, had much greater rates, namely, ExpressJet and Continental Airlines. These two airlines had the highest numbers of delays per 10,000 flights.

2.65: (a)

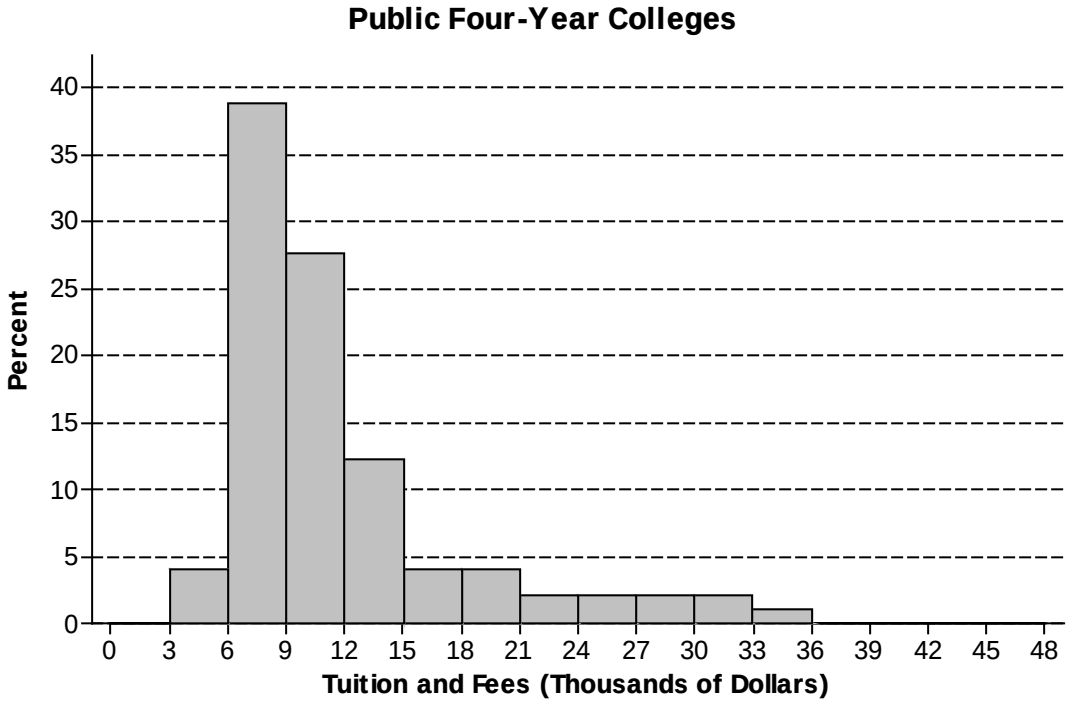
```
0 | 9
1 | 4667788899
2 | 00113333334556667789
3 | 000000011223333457
4 | 4
5 | 1

Legend: 1|4 = 14 cents per gallon
```

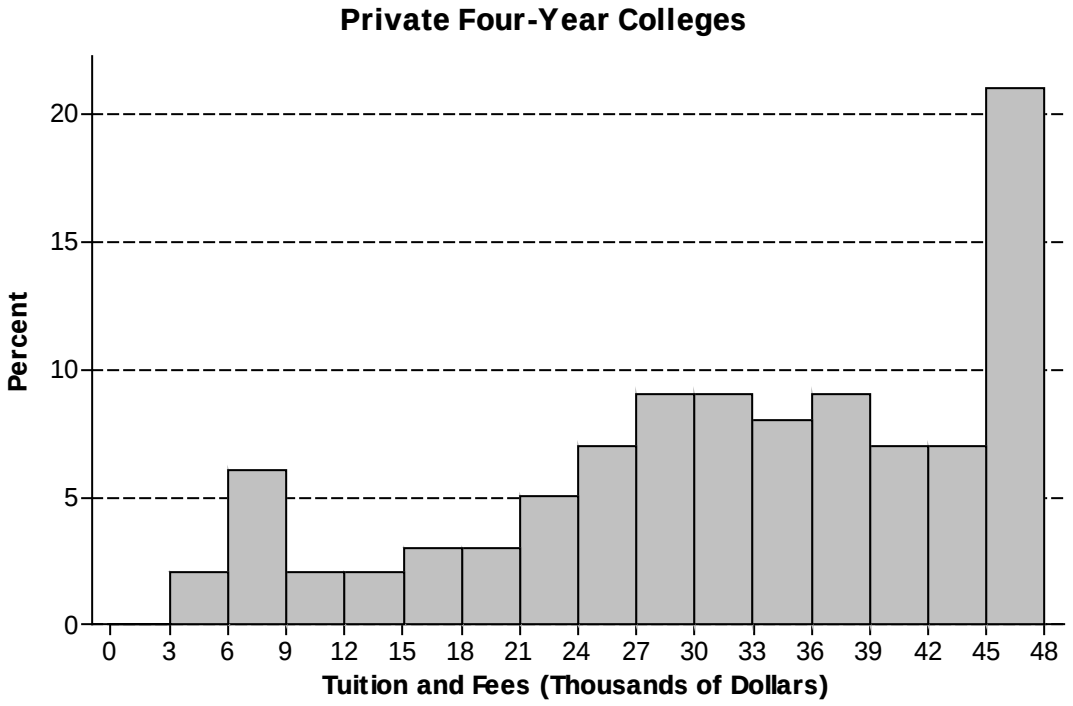
(b) The center is approximately 26 cents per gallon, and most states have a tax that is near the center value, with tax values ranging from 9 cents per gallon to 51 cents per gallon. The distribution is also approximately symmetric.

(c) The only value that might be considered unusual is the 51 cents per gallon tax in Pennsylvania. Although not an obvious outlier, it is approximately 7 cents per gallon higher than the next lower gasoline tax. There are no other states that make such a big jump to the next higher tax.

2.66: (a) The relative frequency histogram shows that the distribution of tuition and fees for public four-year colleges is positively skewed, with the most frequently occurring tuition and fee amounts between \$6,000 and \$9,000. The center of the distribution is in the range \$9,000 to \$12,000. The tuition and fees values range between a low of between \$3,000 and \$6000, to a high of somewhere between \$33,000 and \$36,000. There are no tuition and fee values above \$36,000.

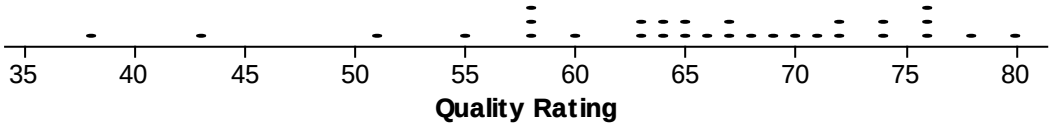


(b) The relative frequency bar chart shows that the distribution of tuition and fees for private non-profit four-year colleges is negatively skewed, with the most frequently occurring tuition and fee amounts between \$45,000 and \$48,000. The center of the distribution is between \$33,000 and \$36,000. The tuition and fees values range between a low of between \$3,000 and \$6,000 to a high of between \$45,000 and \$48,000. There are no tuition and fee values below \$3,000 or above \$48,000.



(c) The differences in the distributions of tuition and fees of public and private four-year colleges are in shape, center, and spread. The shapes are both skewed, with the public college tuition and fees distribution being positively skewed, and the private college tuition and fees distribution being negatively skewed. The centers are quite different, with the center of the private college tuition and fees higher than that for public colleges. Finally, spread of the tuition and fee amounts for private colleges is greater than that for public colleges, with the private colleges having higher tuitions than the public colleges.

2.67: (a)

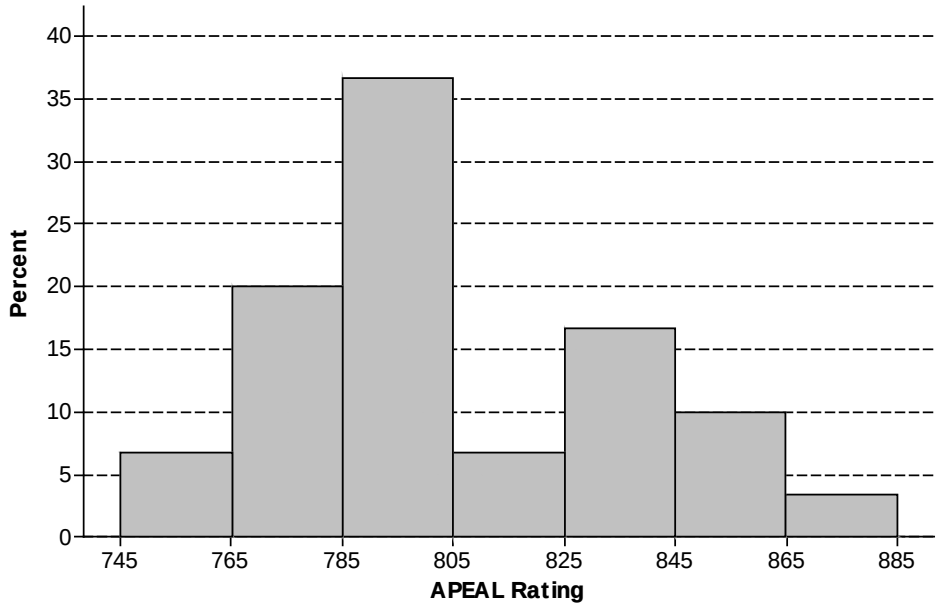


(b) A typical value for quality rating is approximately 66.

(c) Quality rating varies between a low score of 38 defects per 100 vehicles and a high score of 80 defects per 100 vehicles. Noting the minimum and maximum values on the dotplot supports this answer.

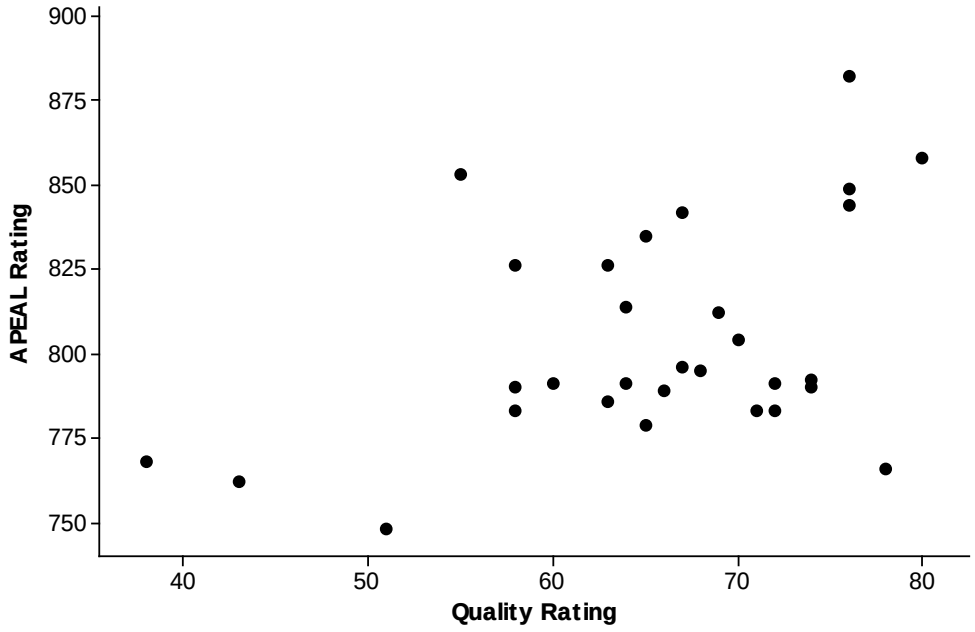
(d) Two brands (Fiat and Jeep) seem to stand out as having much lower quality ratings (defects per 100 vehicles) than the others, with values of 38 and 43, respectively.

(e)



The histogram is centered in the 785-805 range, with values that range between approximately 745 and 885. The distribution is bimodal and positively skewed.

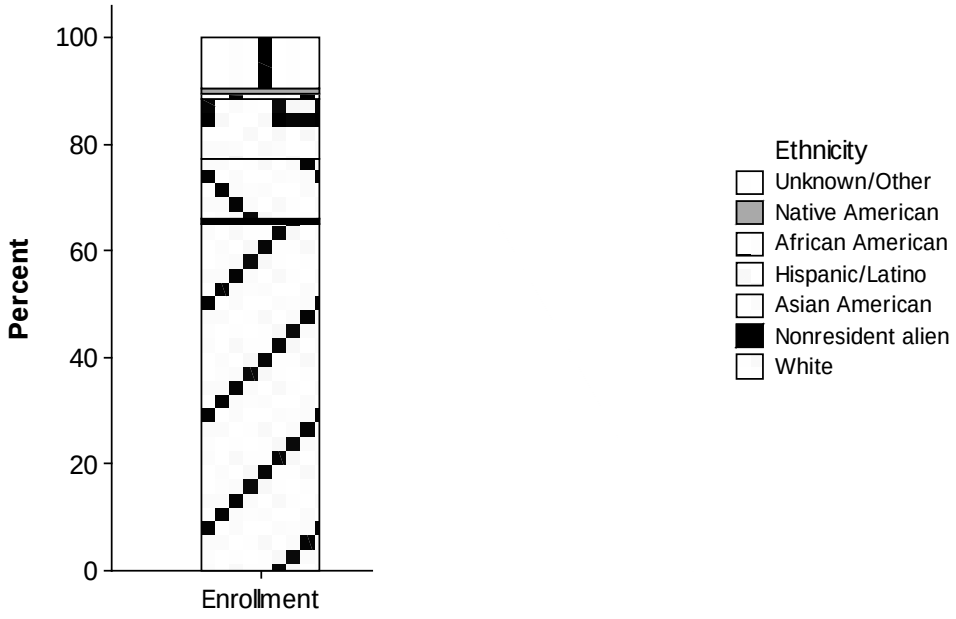
(f)



There is a weak positive association between customer satisfaction (as measured by the APEAL rating) and quality rating. Brands with a higher Quality Ratings (higher number of defects per 100 vehicles) tend to have higher APEAL ratings.

2.68: The time series plot is consistent with the statement of having seen steady growth in recycling and composting because those trends are both increasing. However, the statement about amounts landfilled have generally declined might be somewhat misleading, at least when the conclusion is based on the graph shown, with the given scale. Rather, it might be more appropriate to state that the amounts landfilled have remained roughly constant, with a slight decrease in the given time period.

2.69: (a)



(b) The segmented bar graph in part (a) is more informative because it is easier to get a sense of the percentages of each Ethnicity who are enrolled. Specifically, in the original graphical display with the "Nonwhite" category further subdivided, it could be difficult to compare the "Nonwhite" breakout categories with the categories remaining in the pie.

(c) The pie chart combined with the segmented bar graph could have been chosen because some of the pie slices might be very thin and hard to see, and too many pieces could be difficult to visually process. In addition, it could be that the enrollment census form had the category "Nonwhite," and then that could have been further subdivided.

(d) I would recommend the combination pie chart with segmented bar chart because of the number of categories. In the single pie chart, several categories are too small to distinguish

from each other. The same is true of the segmented pie chart. The combination pie chart with segmented bar chart gives a good sense of the overall percentages for the categories.

2.70: The first graph is not an appropriate representation. In that display, the heights of the "Z's" are drawn to correctly represent the percentages shown. However, the "Z's" are scaled both horizontally and vertically to keep the shape of the "Z" the same. This results in areas and volumes (since the "Z's" are three-dimensional) that are not in the correct proportions to the percentages being represented, which can be misleading. The second display is an appropriate representation. In this display, the heights are drawn appropriately to allow comparison of the different cities. Note also that the widths of each of the cars (which are used in place of bars in a bar chart) are all the same, so the area principle has not been violated.

2.71: The display is misleading because the area principle is violated. The areas of the cocaine mounds are not proportional to the relative frequencies being represented.