Complete Solutions Manual

Statistics
Learning from Data

Roxy Peck
California Polytechnic State University,
San Luis Obispo

Prepared by

Stephen Miller
Winchester Thurston School, Pittsburgh, PA

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

ISBN-13: 978-130508959-4
ISBN-10: 1-30508959-6

Cengage Learning
200 First Stamford Place, 4th Floor
Stamford, CT 06902
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 1.1 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 reviewed the history of the children who were participating in the long-term health study. No children 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 adult Americans 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 1.1 Exercise Set 2

1.6: This is an observational study based on results of a survey (no nurses 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 study participants to different experimental groups (vitamin supplement group or no vitamin supplement group).

Additional Exercises for Section 1.1

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 1.2 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 2,121 children between the ages of 1 and 4, and the population of interest is all children between the ages of 1 and 4.

1.17: No, it is not safe to generalize this result to the larger population of U.S. adults. The 6000 people who sent hair samples were not chosen using a random selection process. Rather, they voluntarily sent their hair samples.

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 but different 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 1.2 Exercise Set 2

1.20: The percentages are statistics, because they were computed from the results of a poll conducted by Travelocity.

1.21: The group of people surveyed represents a sample, and the percentages quoted are statistics (because they were computed from the sample).

1.22: (a) This was a convenience sample because the group of students was an easily available group to form a sample. (b) The estimate of the proportion of students who reported using illegal stimulants should not be generalized to all U.S. college students because this study used a convenience sample by only including students from one psychology class from a small, competitive college.

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 people who use public restrooms. (b) Although the details of the sampling scheme are not presented, the sample size is large (which is often desirable). One issue with how the sample was selected is that only people using public restrooms at airports in New York, Chicago, San Francisco, Dallas, Miami, and Toronto were included in the sample. (c) This sample is not representative of the population of interest because only those people at airports in these six cities were included in the sample. (d) Selection bias is present because those people using public restrooms at places other than airports in these six cities, and public restrooms in other cities in general, have been excluded from the sample.

Additional Exercises for Section 1.2

1.25: 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.

1.26: The population is all 2010 Toyota Camrys. The sample is the 6 2010 Toyota Camrys selected for the crash testing.

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 1.3 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 (tai chi group or control group), resulting in potential confounding. Those who would choose to do tai chi might, in some way, be different from those who would choose the control group. We would not know if differences in immunity between the two groups were due to the tai chi, 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 the tai chi treatment has an effect on immunity to a virus, a control group is needed to provide a baseline against which the treatment group can be compared to determine if the treatment has an effect.

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) Do ethnic group and gender influence the type of care that a heart patient receives? (2) The experimental conditions are the gender and race of the "patient" the doctor is shown. (3) The response variable is the type of care recommended for the heart patient. (4) The experimental units are the 720 primary care doctors at meetings of the American College of Physicians or the American Academy of Family Physicians. It is not clear how the physicians were chosen. (5) Yes, the design incorporates random assignment of doctors to view one of the four different videos through rolling a four-sided die. (6) No, there was no control group, as all the doctors were shown actions of some race or gender. The idea of a control group does not apply in this study. (7) There is no indication that the study includes blinding. There cannot be blinding in this study because the doctors know the gender and race of the "patients" they get.

Section 1.3 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 (viewing and discussing art group or hobbies and interests group), resulting in potential confounding. Those who chose to view and discuss art might, in some way, be different from those who chose to discuss their hobbies and interests. We would not know if differences in attitude, blood pressure, or constipation between the two groups were due to the art discussions, 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 viewing and discussing art has an effect on immunity to a virus, a control group is needed to provide a baseline against which the treatment group can be compared to determine if the treatment has 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 1.3

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 1.4 Exercise Set 1

1.46: 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.

1.47: (a) It is not reasonable to conclude that watching *Oprah* causes a decrease in cravings for fattening foods. 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 women in the United States because not all women watch daytime talk shows. It is not reasonable to generalize these results to all women who watch daytime talk shows because not all women who watch daytime shows access DietSmart.com. If there was no random selection of survey participants (which is often the case with surveys found on websites), then the results might be biased due to voluntary response of participants.

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 4: No, because this was not an experiment, cause-and-effect cannot be concluded.

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 1.4 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: 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.

1.52: 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 1.4

1.54: (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.

1.55: 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.

1.56: 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) 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 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 2:

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 (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 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 4:

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

Question 2: The data set 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 data set 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 data set 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 data set 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 data set 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 data set 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 data set 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 data set 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 data set 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 data set 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.

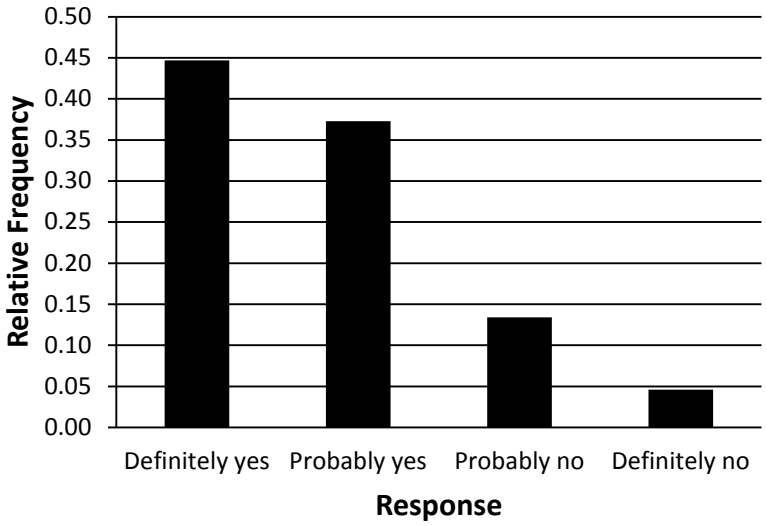
Question 2: The data set is numerical.

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.

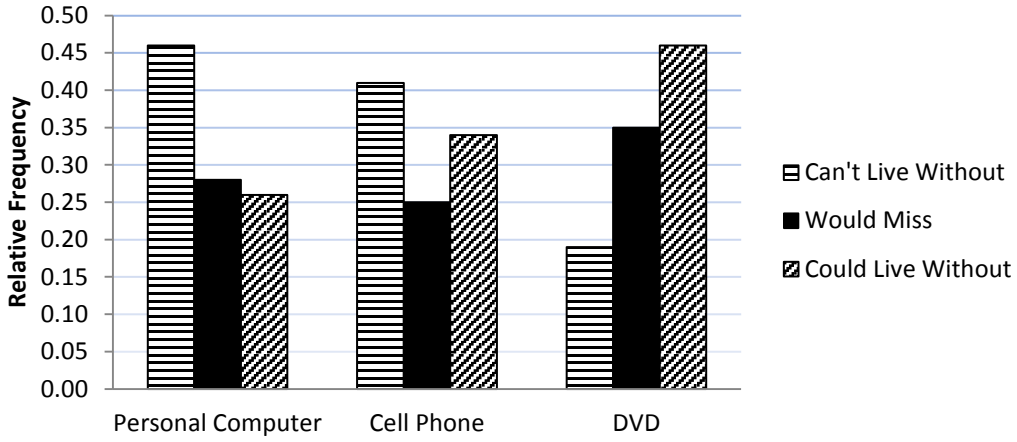
Section 2.2 Exercise Set 1

2.16: (a)



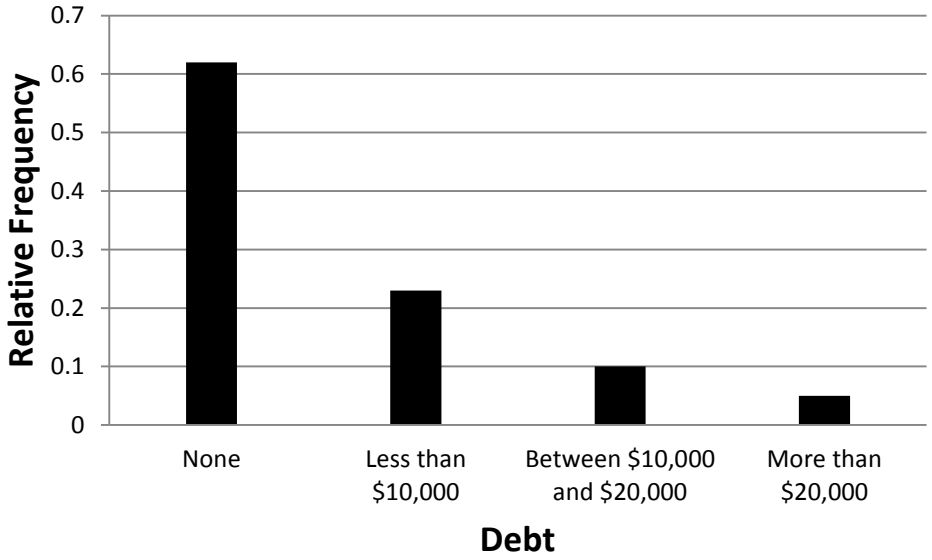
(b) "Senior Satisfaction! Over 80% say they would enroll again."

2.17:



Section 2.2 Exercise Set 2

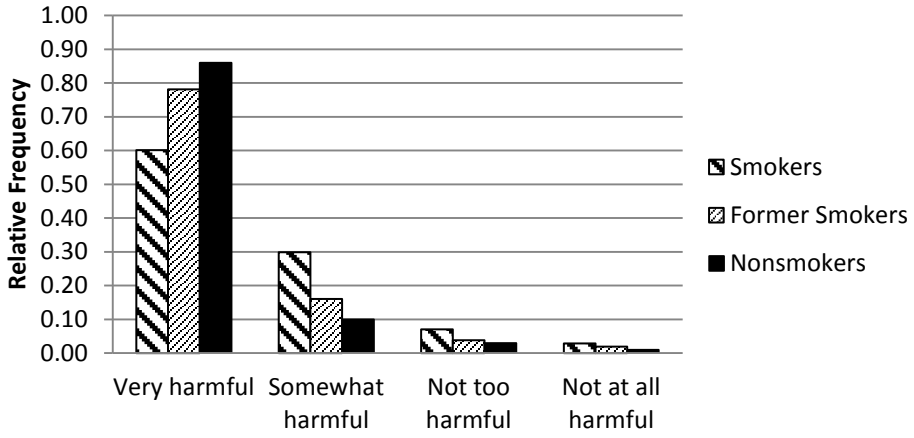
2.18: (a)



(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: The relative frequencies needed are shown in the table below, followed by the comparative bar chart.

Perceived Risk of Smoking	Smokers	Former Smokers	Nonsmokers
Very harmful	0.60	0.78	0.86
Somewhat harmful	0.30	0.16	0.10
Not too harmful	0.07	0.04	0.03
Not at all harmful	0.03	0.02	0.01

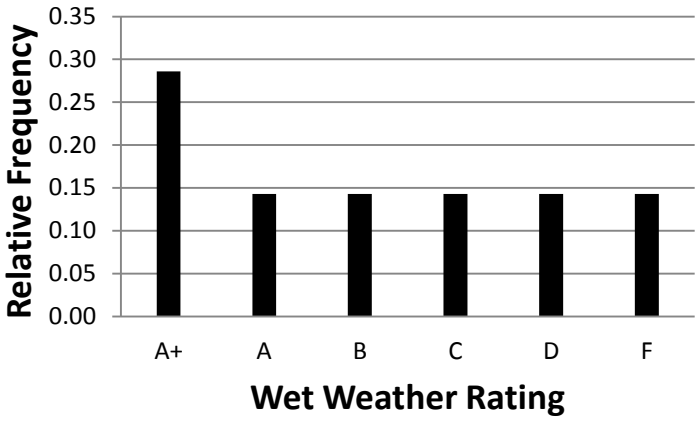


Within each category (smokers, former smokers, and nonsmokers), the ordering based on relative frequency of the perceived risks (very harmful, somewhat harmful, not too harmful, and not at all harmful) is the same. In the somewhat harmful, not too harmful, and not at all harmful categories, smokers had the highest relative frequencies, followed by former smokers, and then nonsmokers. The very harmful perceived risk category is different, in that the nonsmokers had the highest relative frequency, followed by the former smokers, and then the smokers.

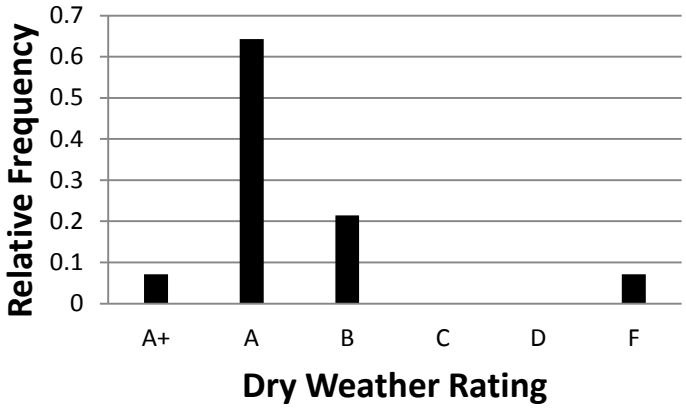
Additional Exercises for Section 2.2

2.20: (a)

Rating	Relative Frequency
A+	4/14 = 0.286
A	2/14 = 0.143
B	2/14 = 0.143
C	2/14 = 0.143
D	2/14 = 0.143
F	2/14 = 0.143

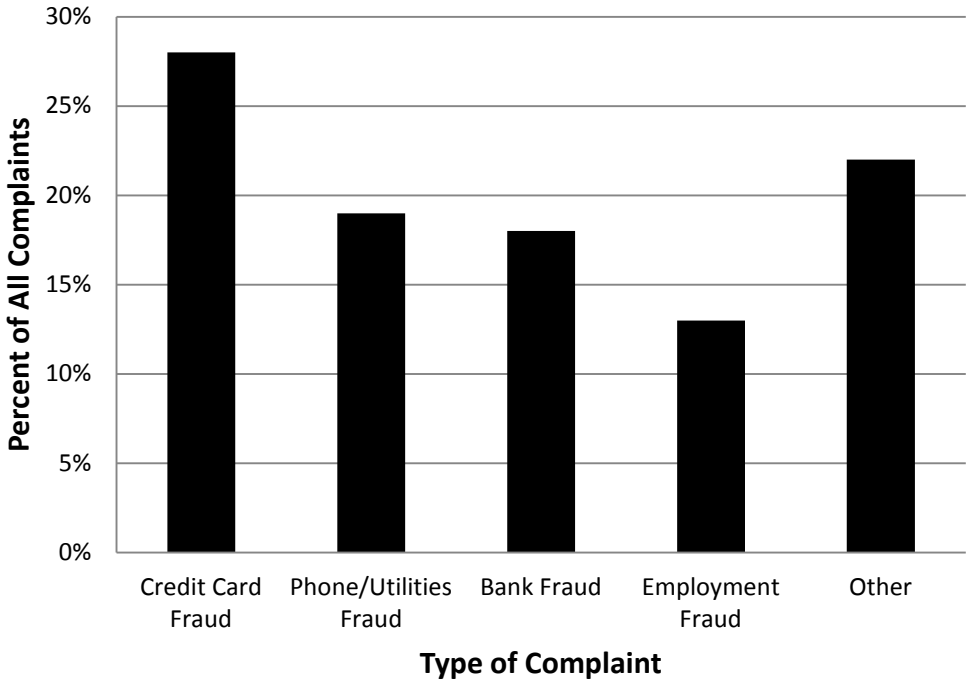


(b)



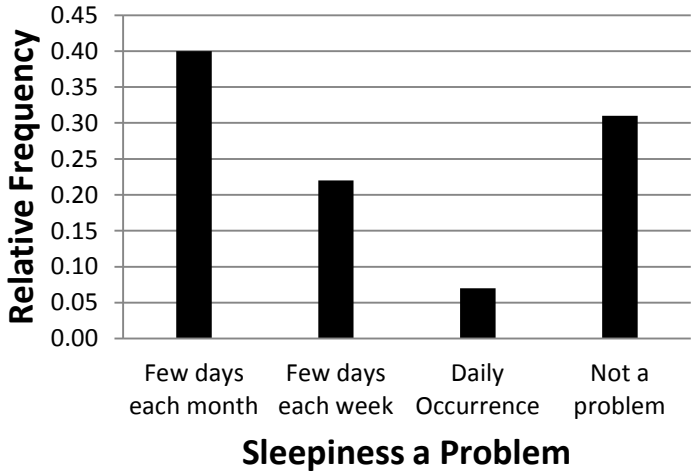
(c) Yes, the bar charts from (a) and (b) support the statement that beach water quality tends to be better in dry weather conditions. The bar charts show that in dry weather condtions, approximately 93% of the beaches have a rating of "B" or higher, whereas in wet weather, only approximately 57% of the beaches have a rating of "B" or higher.

2.21:



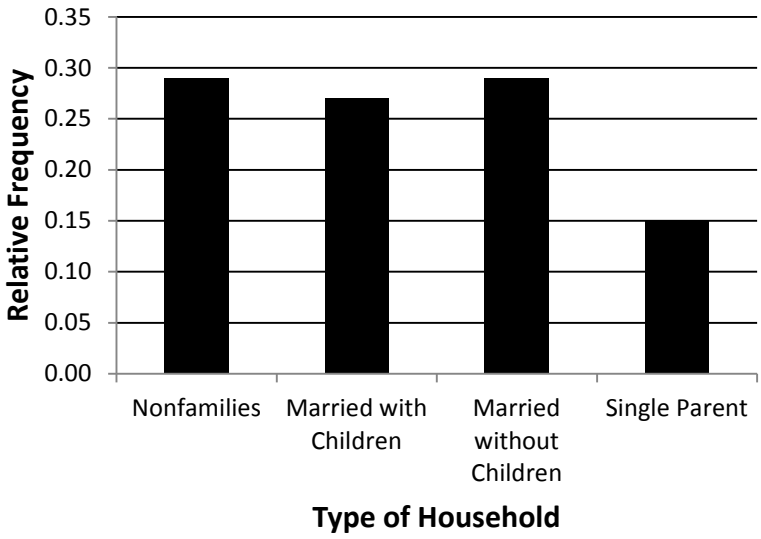
Credit card fraud is the most commonly occurring identity theft type. Although phone/utility, bank, and employment fraud each constitute a relatively large portion of overall type of identity theft, the collective "other fraud" category is greater than any one of these other three.

2.22: To construct a relative frequency bar chart, the sum of the relative frequencies for all the categories must add to 1 (or 100%). As such, the category "sleepiness on the job was not a problem" must account for 100 – (40 + 22 + 7) = 31% of those surveyed.



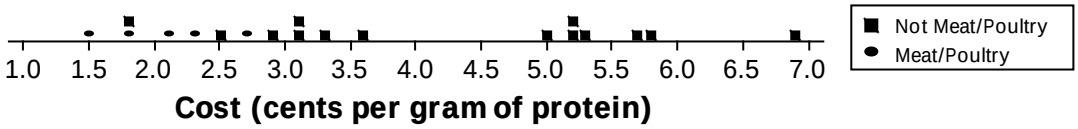
2.23: The relative frequency distribution is:

Type of Household	Relative Frequency
Nonfamilies	0.29
Married with Children	0.27
Married without Children	0.29
Single Parent	0.15



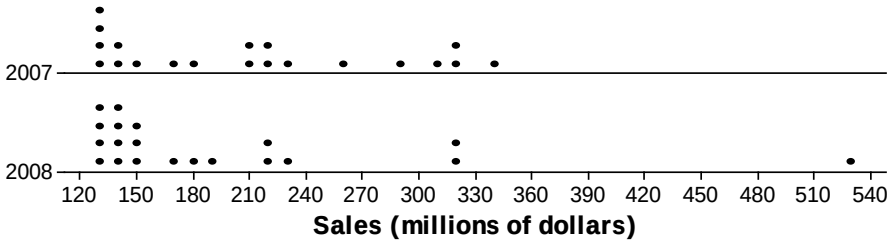
Section 2.3 Exercise Set 1

2.24: (a) The dotplot below shows the cost per gram of protein for 19 common food sources of protein.



(b) Because the meat and poultry products (represented by the circles on the dotplot) are located at generally smaller values of cost (in cents per gram of protein), they do appear to be a good value when compared to other sources of protein.

2.25: (a)



(b) The shape of both distributions is skewed toward larger values. The 2007 ticket sales is centered at about $210 million dollars, which is higher than the center of the 2008 ticket sales, which is centered around $150 million dollars. The lowest ticket sales for both 2007 and 2008 are approximately $127 million dollars. Ticket sales for 2008 has a maximum value of approximately $533 million dollars, which is much higher than the highest ticket sales for 2007. Without this extreme value, the spread between the lowest and highest values are approximately equal, with 2007 having a slightly higher spread than 2008.

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, with one possible outlier of 28.8 years.

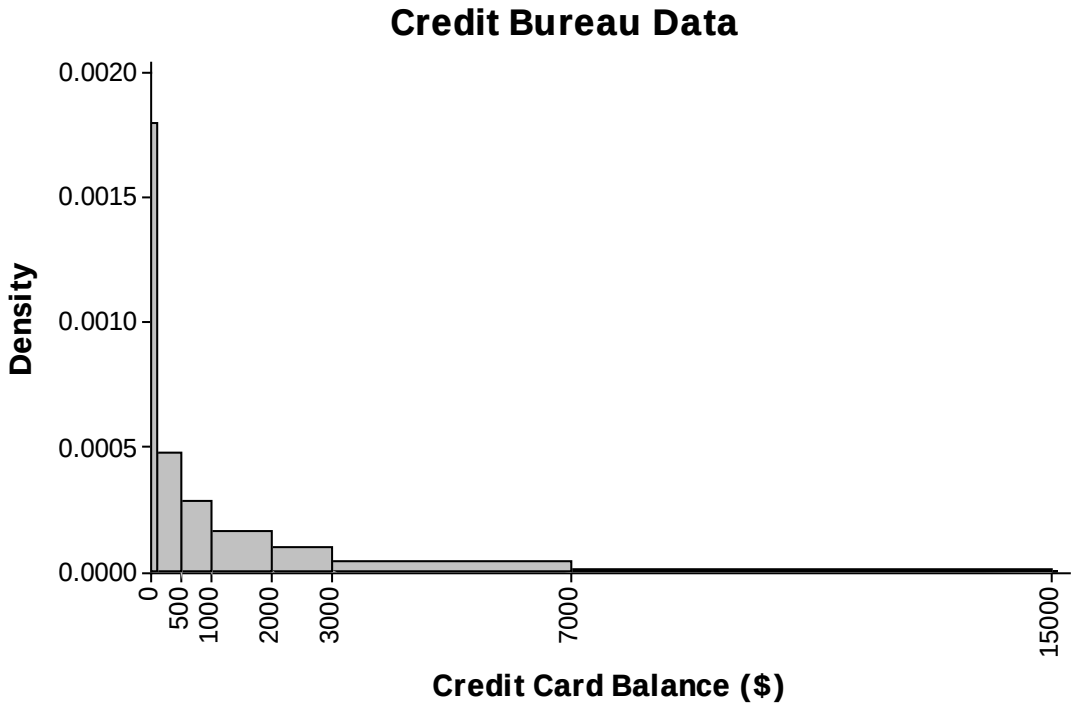
2.27: (a)

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	

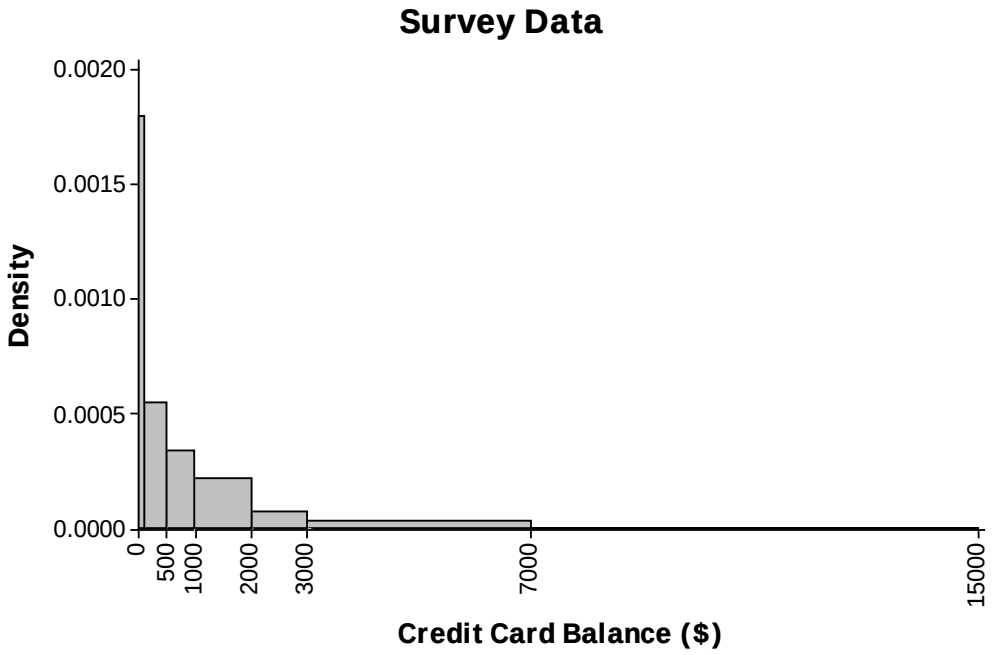
Legend: 4|6 = 46 extra hours per year

(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)


Credit Bureau Data
Density
0.0020
0.0015
0.0010
0.0005
0.0000
0
500
1000
2000
3000
7000
15000
Credit Card Balance ($)

(b)


Survey Data
Density
0.0020
0.0015
0.0010
0.0005
0.0000
0
500
1000
2000
3000
7000
15000
Credit Card Balance ($)

(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.

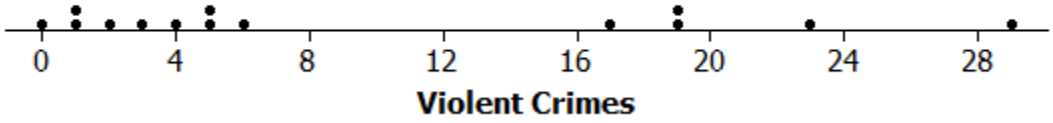
2.29: (a) If the exam is quite easy, the scores would be clustered at the high end of the scale, with a few low scores for the students who did not study. The histogram would be skewed toward low values.

(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 skewed toward higher values.

(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

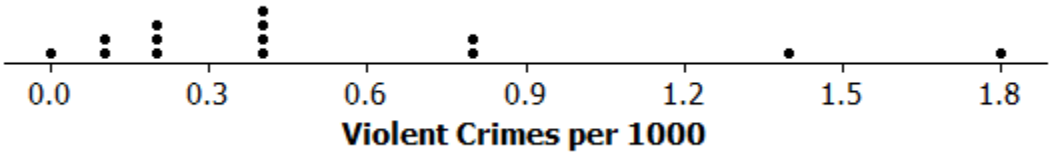
2.30: (a)



It appears as if there are two clusters of schools, those with 0 to 6 violent crimes, and those with 17 or more. The schools with 17 or more violent crimes are the University of Florida, University of Central Florida, University of South Florida, Florida A&M University, and Florida State University.

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

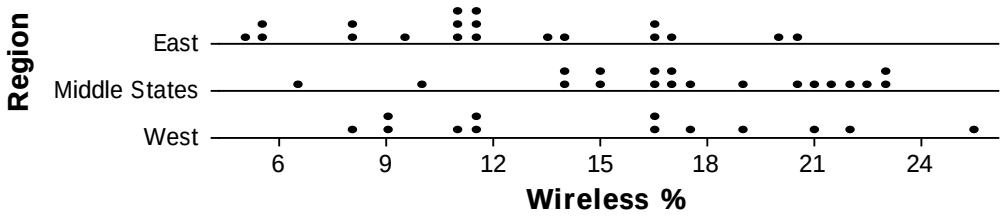
University/College	Violent Crimes per 1000
Florida A&M University	1.8
Florida Atlantic University	0.2
Florida Gulf Coast University	0.8
Florida International University	0.1
Florida State University	0.8
New College of Florida	1.4
Pensacola Junior College	0.2
Santa Fe Community College	0.2
Tallahassee Community College	0.0
University of Central Florida	0.4
University of Florida	0.4
University of North Florida	0.4
University of South Florida	0.4
University of West Florida	0.1



No, the same schools do not stand out as unusual. Now only two schools seem to be unusual, namely, Florida A&M University (1.8 violent crimes per 1000) and New College of Florida (1.4 violent crimes per 1000).

(c) A typical (median) number of violent crimes among the Florida universities and colleges is approximately 5. The dotplot shows that there are two groups of violent crime numbers, those below 6 and those above 17. Overall, the number of violent crimes ranged between 0 and 29. Considering the crimes per 1,000 students, a typical (median) value is 0.4, with values that range between 0 and 1.8 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 wireless percent for the three geographical regions. The distributions of wireless percent for the three regions are similar, with the Eastern region having the lowest wireless percent, and the Middle States and Western regions being similar to each other.

2.32: (a)

-1	100
-0	99998888776555555444433222211110
0	000011244577
1	179
2	2

Legend: 0 | 4 = 4% change

(b) The states that saw a marked increase in the number of 25- to 44-year-olds (Arizona, Idaho, Nevada, and Utah) are Western states that tend to be considered warmer states.

2.33: (a)

```
0 | 00333445555688888888999999
1 | 0001223344567
2 | 001123689
3 | 0
4 | 0
5 |
6 | 6

   Legend: 0|8 = 8% change
```

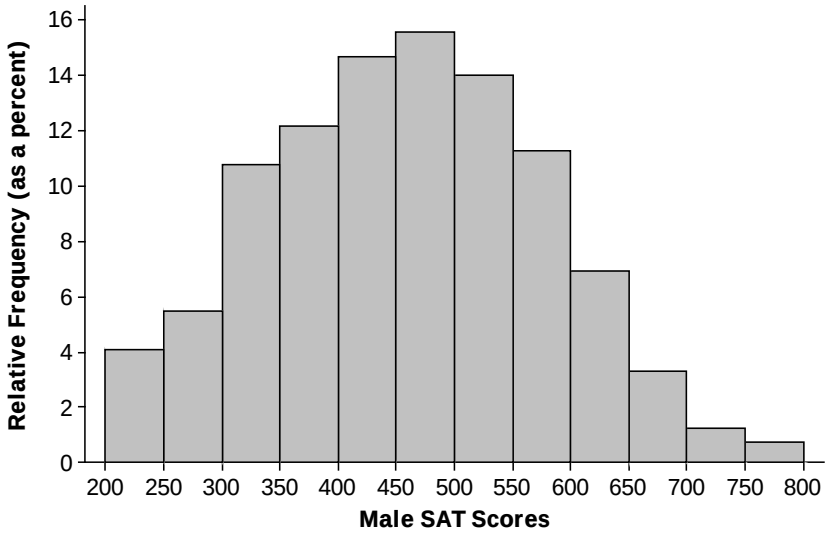
(b) A typical percent change is approximately 10%, with most of the values varying between about 0% and 30%. The distribution is unimodal with a peak in the 0-9 interval, and is positively skewed. There appear to be two outliers, one at 40% (Arizona), and an extreme outlier at 66% (Nevada).

(c)

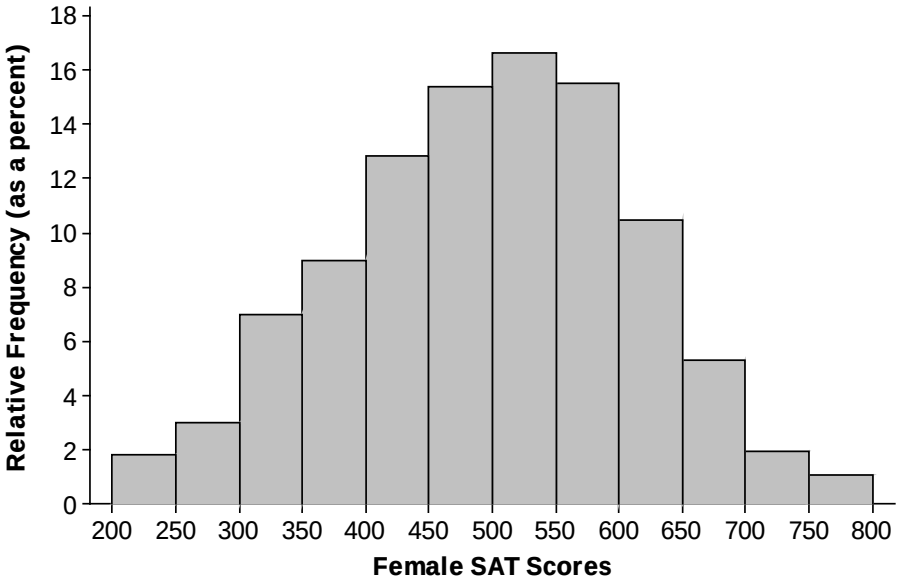
```
                East          West
999988865555443330 | 0 | 0888899
         7654321000 | 1 | 234
                631 | 2 | 001289
                    | 3 | 0
                    | 4 | 0
                    | 5 |
                    | 6 | 6
                          Legend:  1|2 = 12% change
```

In general, Western states have seen more growth than Eastern states. A typical percentage change for Western states is approximately 14%, which is greater than the typical percentage change for Eastern states of approximately 9%. In addition, a few Western states (Arizona and Nevada) have seen a much greater percentage increase than any of the Eastern states. Overall, there is greater variability in the percentage increase for Western states than that of Eastern states. Specifically, the percentage increase for Western states varies between 0 and 66%, which is larger than that for Eastern states, which range between 0 and 26%. Both distributions are skewed toward larger values (positively skewed). The distribution for Eastern states is unimodal, with a peak in the 0-9% interval. The distribution for Western states is bimodal, with peaks in the 0-9% and 20-29% intervals.

2.34: (a)



(b)



(c) The distribution of male SAT scores is approximately symmetric, centered at approximately 450-500. The distribution of female SAT scores differs in that it is negatively skewed and centered at a higher value, approximately 500-550. 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)

0	1111122222333444567777 88
1	01122233444567778
2	235
3	112
4	13
5	5
6	7

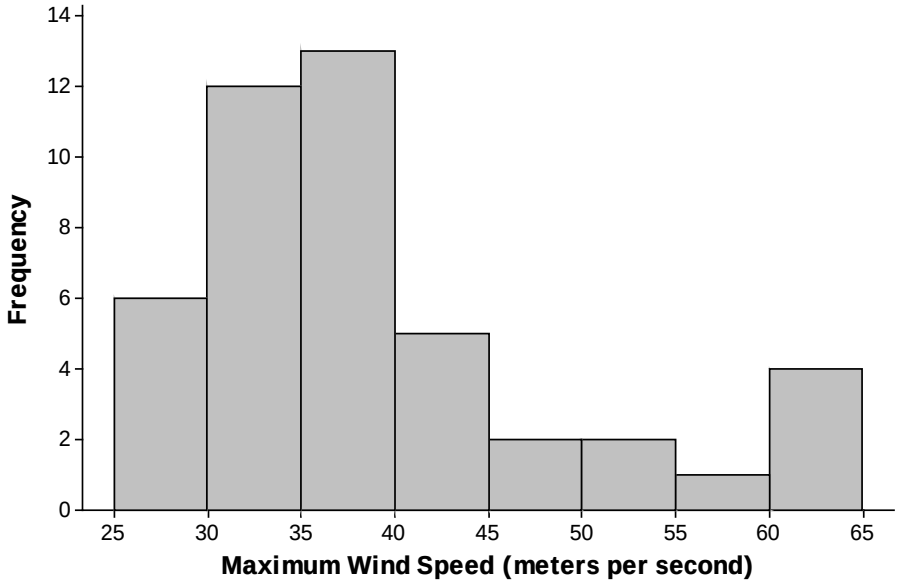
Legend: 2 | 2 = 2,200 (truncated to hundreds place)

(b) The distribution is strongly positively skewed. The center of the distribution is approximately 1,100 thousands. The four largest values could be outliers, given the large gap between the 3,200 thousand and 4,100 thousand values.

(c) This does not indicate that tobacco use is more of a problem in these states than elsewhere because these states also have extremely large populations, so it is likely that there would be more smokers in those states.

(d) I would not use the data in the table to compare states on the extent of tobacco use. Rather, I would want to consider tobacco users as a percent of total state population in order to better compare the states.

2.37: 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.

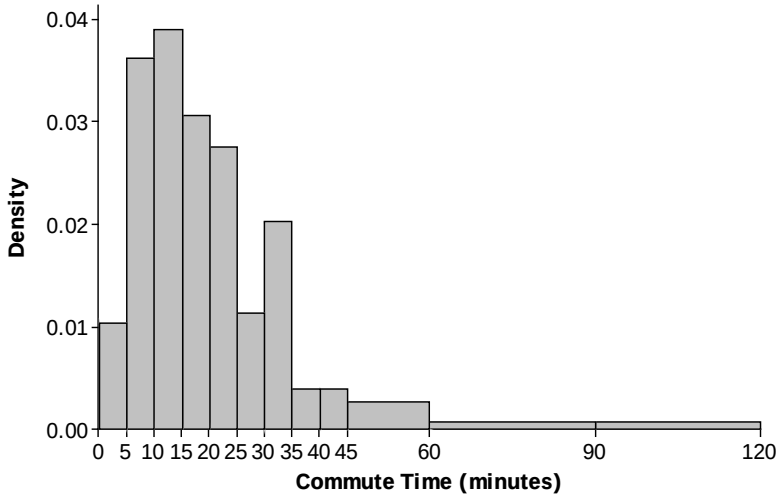


2.38: (a) Given that there are many commute times at the lower end of the distribution, more class intervals are required to show sufficient detail (hence the 5-minute interval width). Using that same interval width would likely give too many intervals. As such, wider intervals for the longer commute times are used.

(b)

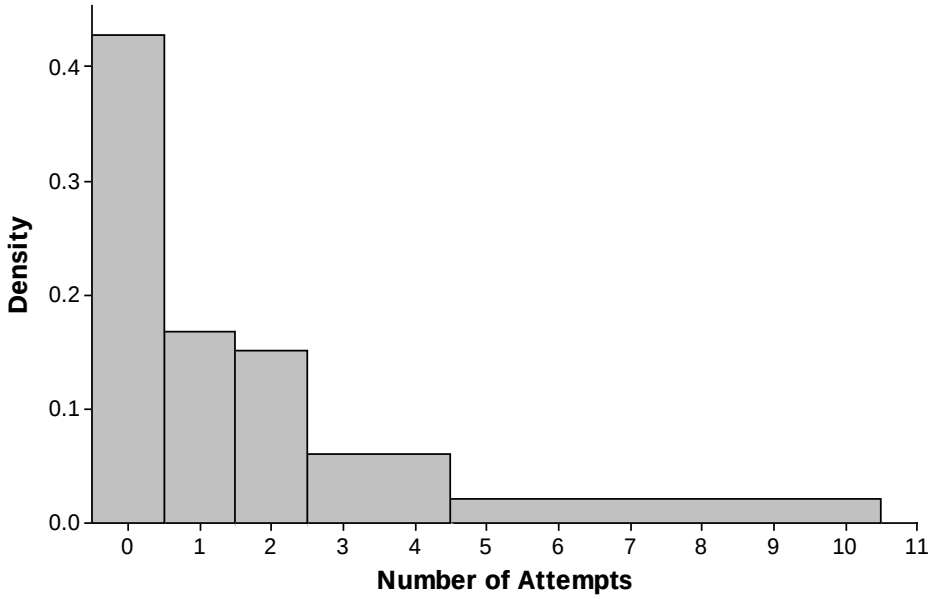
Commute Time	Frequency	Relative Frequency	Density
0 to < 5	5,200	0.051793	0.010359
5 to < 10	18,200	0.181275	0.036255
10 to < 15	19,600	0.195219	0.039044
15 to < 20	15,400	0.153386	0.030677
20 to < 25	13,800	0.137450	0.027490
25 to < 30	5,700	0.056773	0.011355
30 to < 35	10,200	0.101594	0.020319
35 to < 40	2,000	0.019920	0.003984
40 to < 45	2,000	0.019920	0.003984
45 to < 60	4,000	0.039841	0.002656
60 to < 90	2,100	0.020916	0.000697
90 to < 120	2,200	0.021912	0.000730

(c)



The density histogram is positively skewed and shows an unusual low number of commute times between 25 and 30 minutes when compared with adjacent commute times. The distribution is centered at approximately 15-20 minutes, which indicates that the typical commute time is between 15 and 20 minutes.

2.39:



2.40:

1	9
2	23788999
3	0011112233459
4	0123

Legend: 2 | 2 = 22 calories

The number of calories in these 26 brands of light beer varies between a low of 19 calories and a high of 43 calories. Of these beers, a typical calorie count is 31.

2.41:

Non-Disney		Disney
9521000	0	001233579
651	1	567
0	2	029
	3	
	4	
	5	4

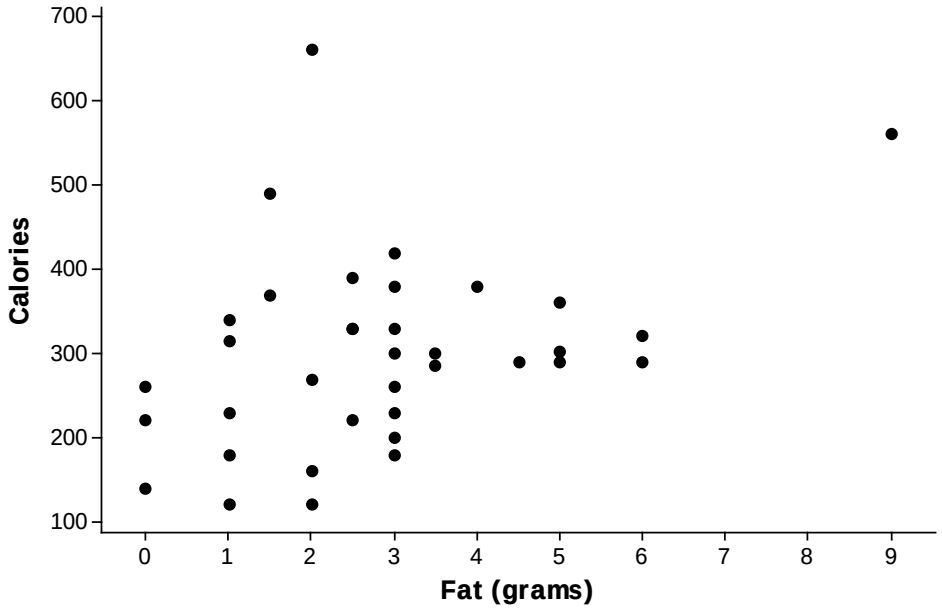
Legend: 1|5 = 150 seconds

In general, the Disney movies have longer tobacco exposure times than the non-Disney movies. Disney movies have a typical value of approximately 80 seconds, which is larger

than a typical value for the non-Disney movies of 50 seconds. In addition, there is more variability in the Disney tobacco exposure times than the others as well. Disney movies vary between 6 and 548 seconds, which is greater than the observed spread for the non-Disney movies, which vary between 1 and 205 seconds. Finally, there appears to be one outlier in the Disney movies (548 seconds), and no outliers in the non-Disney movies.

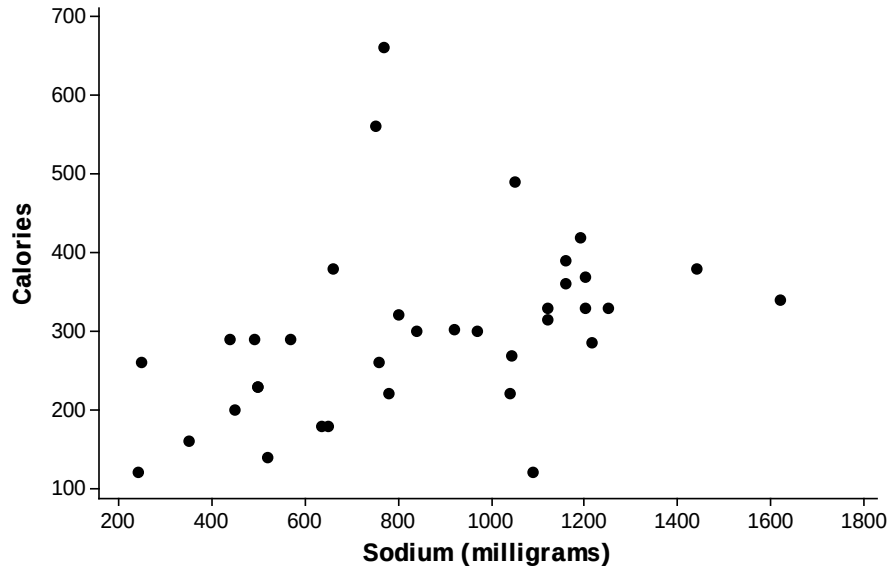
Section 2.4 Exercise Set 1

2.42: (a)



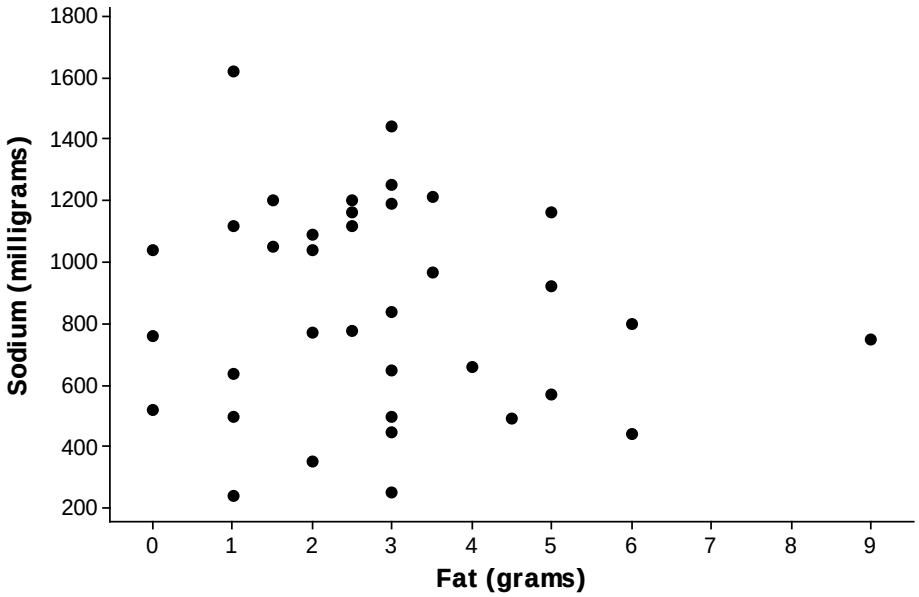
The scatterplot shows the expected positive relationship between grams of fat and calories. The relationship is weak.

(b)



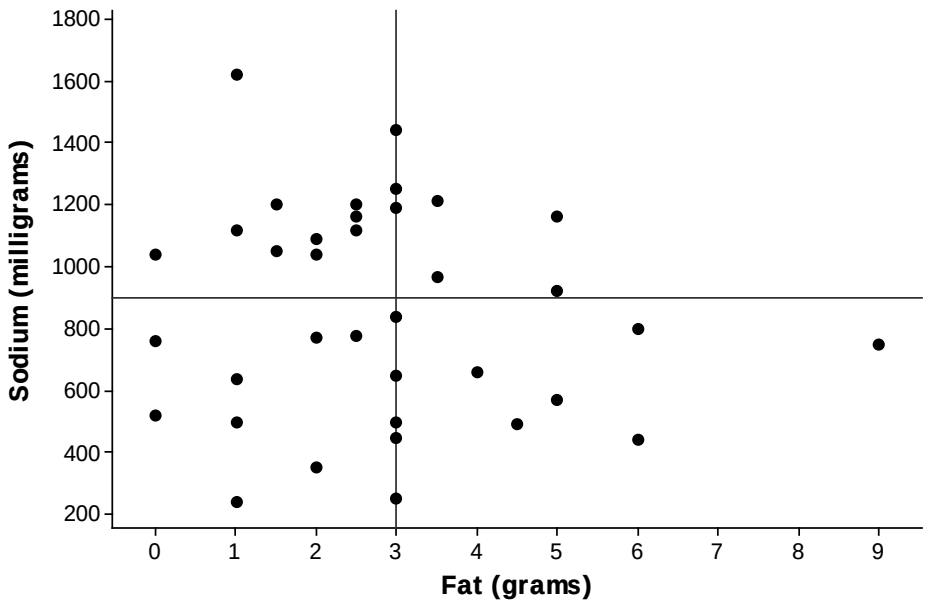
As was observed in the calories vs. fat scatterplot, there is also a weak, positive relationship between calories and sodium. The relationship between calories and sodium appears to be a little stronger than the calories vs. fat relationship.

(c)



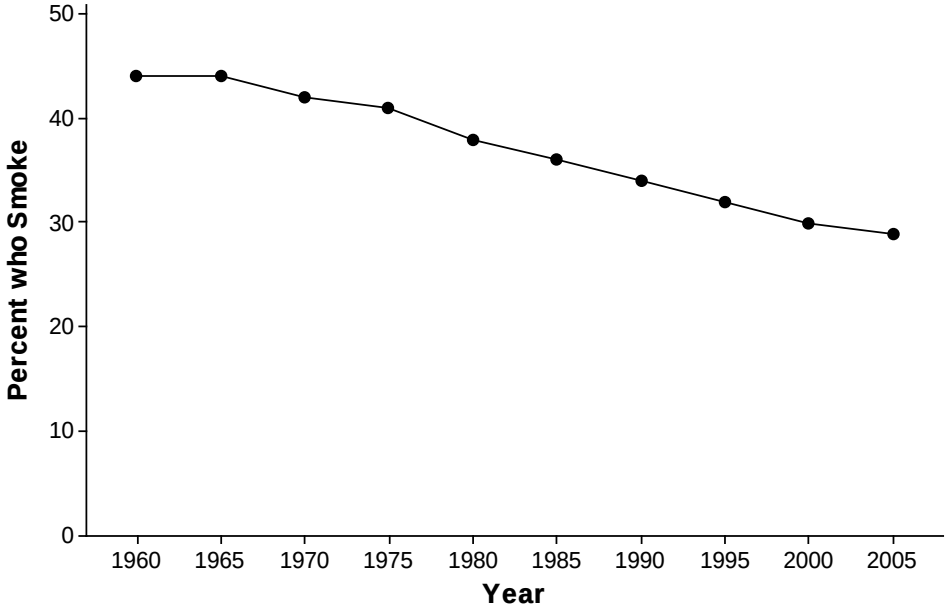
There is no apparent relationship between sodium and fat.

(d)



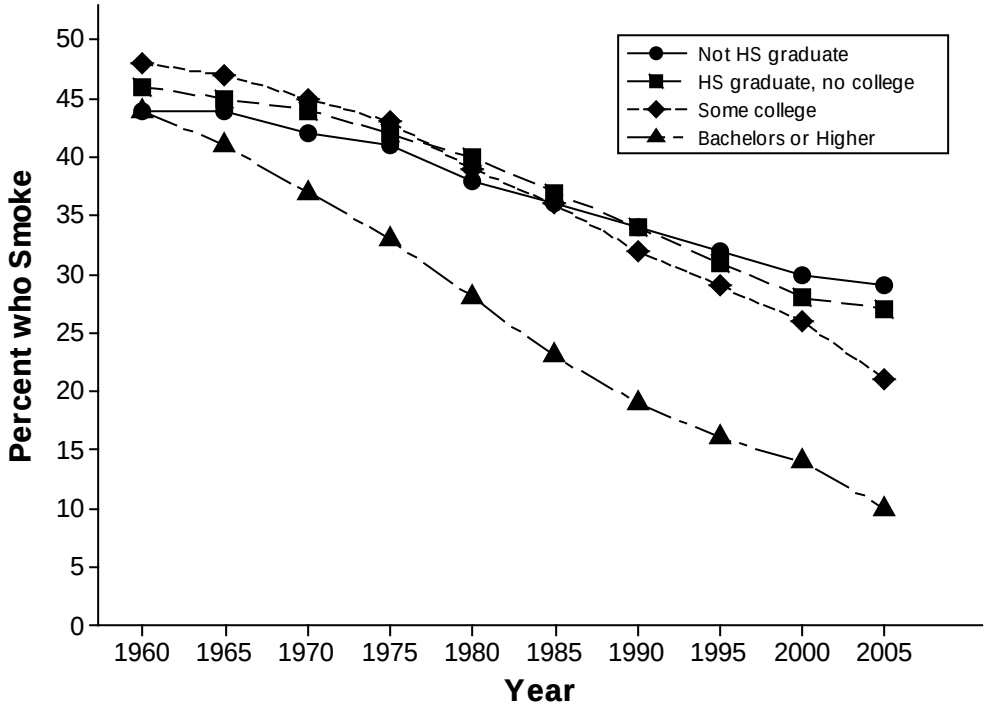
The lower left quadrant corresponds to healthier fast-food choices. This quadrant corresponds to food items with fewer than 3 grams of fat, and fewer than 900 milligrams of sodium, which are considered the healthier choices.

2.43:



There has been a steady downward trend in the percent of people who smoke among people who did not graduate from high school, from a high of 44% to 29% in 2005.

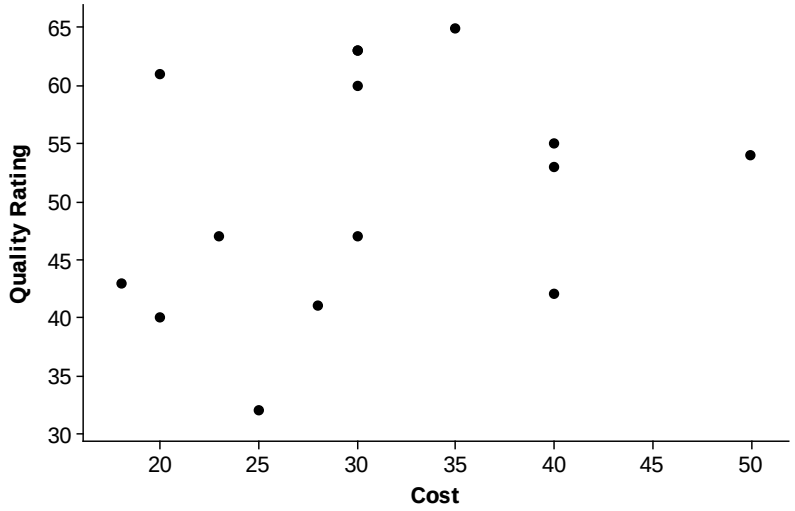
(b)



(c) There has been a steady downward trend in the percent of people age 25 or older who smoke regardless of education level. In 1960, regardless of education level, approximately the same percentage of people age 25 or older smoked (approximately 44-48%). Over time, however, the differences between the percent of people who smoke for different education levels has become more pronounced. In 2005, those people with bachelor's degrees or higher had the lowest smoking rate (10%), followed by those with some college (21%). The highest rates of smoking were found among those who either did not graduate from high school (29%), or graduated high school but did not attend college (27%).

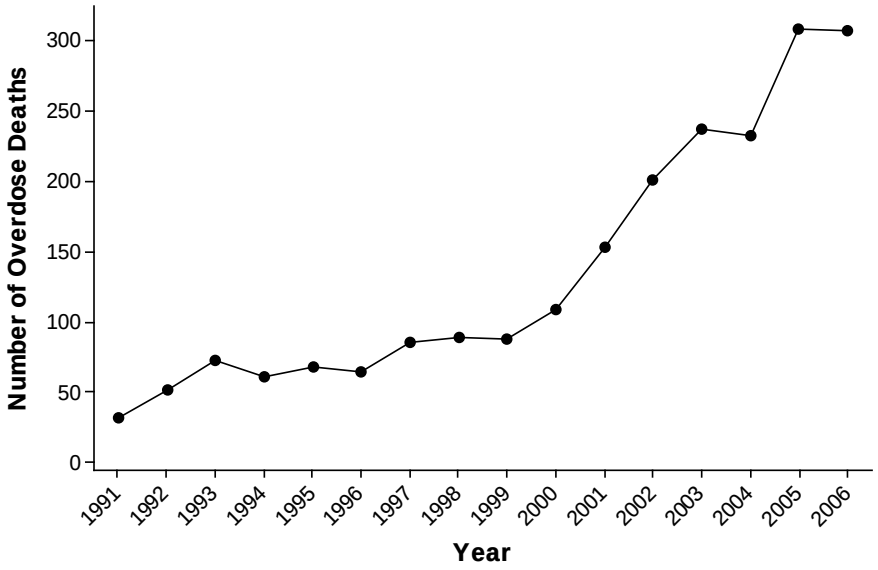
Section 2.4 Exercise Set 2

2.44: (a)



(b) There does not appear to be a relationship between cost and quality rating. The scatterplot does not indicate that the more expensive helmets tended to receive higher quality ratings.

2.45:



The number of overdose deaths has increased over time, but not at a steady rate. The greatest rate of increase occurred between 2000 and 2006.