

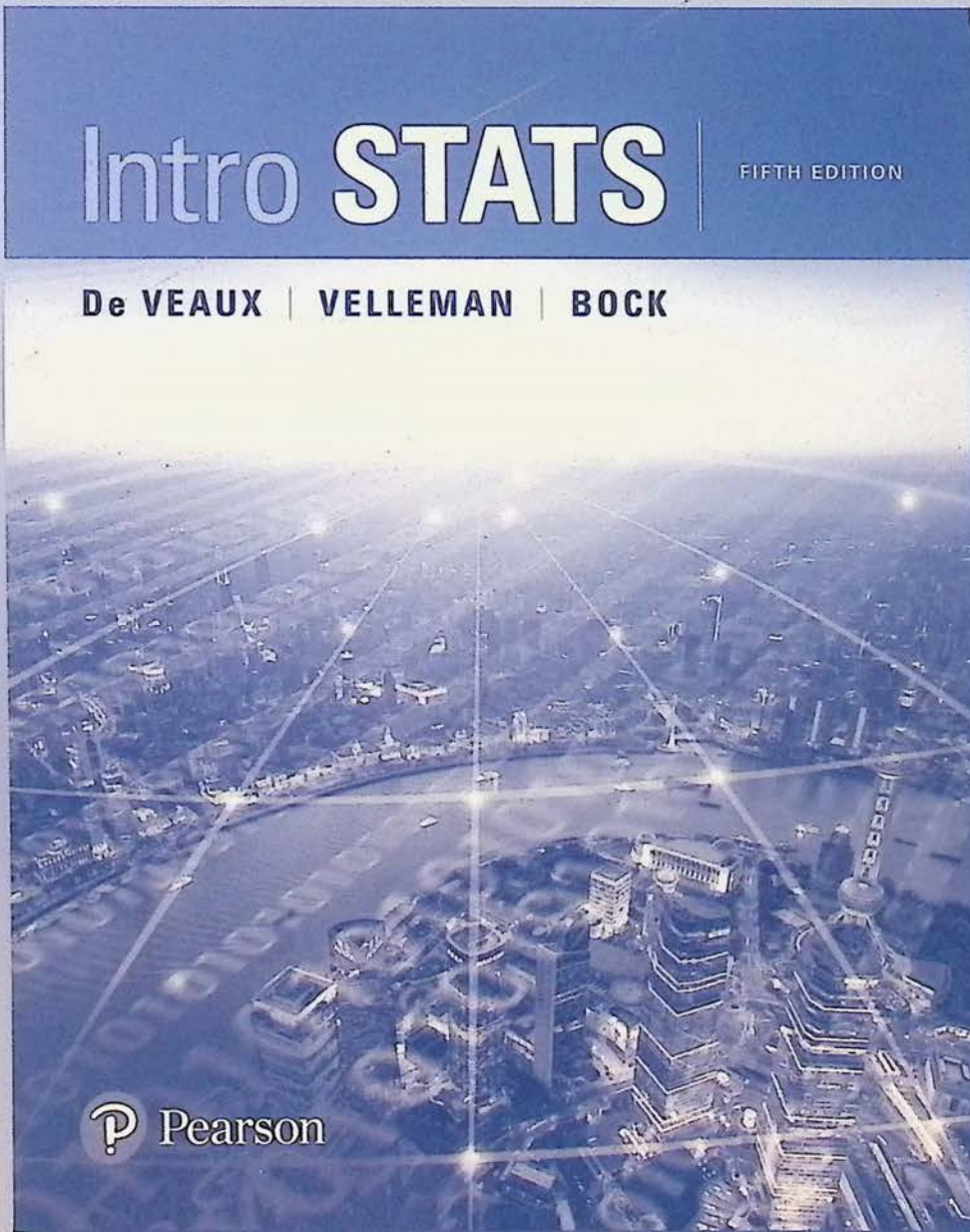
Student's Solutions Manual

William Craine III


Intro **STATS**

FIFTH EDITION

De VEAUX | VELLEMAN | BOCK



 Pearson

 Pearson

Contents

<i>Chapter 1</i>	Stats Starts Here	1
<i>Chapter 2</i>	Displaying and Describing Data	5
<i>Chapter 3</i>	Relationships Between Categorical Variables -Contingency Tables	24
<i>Chapter 4</i>	Understanding and Comparing Distributions	33
<i>Chapter 5</i>	The Standard Deviation as Ruler and the Normal Model	42
<i>Review of Part I</i>	Exploring and Understanding Data	55
<i>Chapter 6</i>	Scatterplots, Association, and Correlation	65
<i>Chapter 7</i>	Linear Regression	73
<i>Chapter 8</i>	Regression Wisdom	90
<i>Chapter 9</i>	Multiple Regression	110
<i>Review of Part II</i>	Exploring Relationships Between Variables	114
<i>Chapter 10</i>	Sample Surveys	126
<i>Chapter 11</i>	Experiments and Observational Studies	132
<i>Review of Part III</i>	Gathering Data	142
<i>Chapter 12</i>	From Randomness to Probability	148
<i>Chapter 13</i>	Sampling Distribution Models and Confidence Intervals for Proportions	162
<i>Chapter 14</i>	Inferences About Means	171
<i>Chapter 15</i>	Testing Hypotheses	184
<i>Chapter 16</i>	More About Tests and Intervals	200
<i>Review of Part IV</i>	From the Data at Hand to the World at Large	208
<i>Chapter 17</i>	Comparing Groups	223
<i>Chapter 18</i>	Paired Samples and Blocks	249
<i>Chapter 19</i>	Comparing Counts	260
<i>Chapter 20</i>	Inferences for Regression	272
<i>Review of Part V</i>	Inference for Relationships	287
<i>Parts I–V</i>	<i>Cumulative Review Exercises</i>	305

Chapter 1 - Stats Starts Here

Section 1.1

- 1. Grocery shopping.** Discount cards at grocery stores allow the stores to collect information about the products that the customer purchases, what other products are purchased at the same time, whether or not the customer uses coupons, and the date and time that the products are purchased. This information can be linked to demographic information about the customer that was volunteered when applying for the card, such as the customer's name, address, sex, age, income level, and other variables. The grocery store chain will use that information to better market their products. This includes everything from printing out coupons at the checkout that are targeted to specific customers to deciding what television, print, or Internet advertisements to use.
- 3. Parking lots.** The owners of the parking garage can advertise about the availability of parking. They can also communicate with businesses about hours when more spots are available and when they should encourage more business.

Section 1.2

- 5. Super Bowl.** When collecting data about the Super Bowl, the games themselves are the *Who*.
- 7. Health records.** The sample is about 5,000 people, and the population is all residents of the United States of America. The *Who* is the selected subjects and the *What* includes medical, dental, and physiological measurements and laboratory test results.

Section 1.3

- 9. Grade level.**
 - a)** If we are, for example, comparing the percentage of first-graders who can tie their own shoes to the percentage of second-graders who can tie their own shoes, grade-level is treated as categorical. It is just a way to group the students. We would use the same methods if we were comparing boys to girls or brown-eyed kids to blue-eyed kids.
 - b)** If we were studying the relationship between grade-level and height, we would be treating grade level as quantitative.
- 11. Voters.** The response is a categorical variable.
- 13. Medicine.** The company is studying a quantitative variable.

2 Part I Exploring and Understanding Data

Section 1.4

15. **Voting and elections.** Pollsters might consider whether a person voted previously or whether he or she could name the candidates. Voting previously and knowing the candidates may indicate a greater interest in the election.
17. **The News.** Answers will vary.
19. **Gaydar.** *Who* – 40 undergraduate women. *What* – Whether or not the women could identify the sexual orientation of men based on a picture. *Population of interest* – All women.
21. **Bicycle Safety.** *Who* – 2,500 cars. *What* – Distance from the bicycle to the passing car (in inches). *Population of interest* – All cars passing bicyclists.
23. **Honesty.** *Who* – Workers who buy coffee in an office. *What* – amount of money contributed to the collection tray. *Population of interest* – All people in honor system payment situations.
25. **Not-so-diet soda.** *Who* – 474 participants. *What* – whether or not the participant drank two or more diet sodas per day, waist size at the beginning of the study, and waist size at the end of the study. *Population of interest* – All people.
27. **Weighing bears.** *Who* – 54 bears. *What* – Weight, neck size, length (no specified units), and sex. *When* – Not specified. *Where* – Not specified. *Why* – Since bears are difficult to weigh, the researchers hope to use the relationships between weight, neck size, length, and sex of bears to estimate the weight of bears, given the other, more observable features of the bear. *How* – Researchers collected data on 54 bears they were able to catch. *Variables* – There are 4 variables; weight, neck size, and length are quantitative variables, and sex is a categorical variable. No units are specified for the quantitative variables. *Concerns* – The researchers are (obviously!) only able to collect data from bears they were able to catch. This method is a good one, as long as the researchers believe the bears caught are representative of all bears, in regard to the relationships between weight, neck size, length, and sex.
29. **Arby's menu.** *Who* – Arby's sandwiches. *What* – type of meat, number of calories (in calories), and serving size (in ounces). *When* – Not specified. *Where* – Arby's restaurants. *Why* – These data might be used to assess the nutritional value of the different sandwiches. *How* – Information was gathered from each of the sandwiches on the menu at Arby's, resulting in a census. *Variables* – There are three variables. Number of calories and serving size (ounces) are quantitative variables, and type of meat is a categorical variable.

- 31. Babies.** *Who* – 882 births. *What* – Mother’s age (in years), length of pregnancy (in weeks), type of birth (caesarean, induced, or natural), level of prenatal care (none, minimal, or adequate), birth weight of baby (unit of measurement not specified, but probably pounds and ounces), gender of baby (male or female), and baby’s health problems (none, minor, major). *When* – 1998-2000. *Where* – Large city hospital. *Why* – Researchers were investigating the impact of prenatal care on newborn health. *How* – It appears that they kept track of all births in the form of hospital records, although it is not specifically stated. *Variables* – There are three quantitative variables: mother’s age (years), length of pregnancy (, and birth weight of baby. There are four categorical variables: type of birth, level of prenatal care, gender of baby, and baby’s health problems.
- 33. Herbal medicine.** *Who* – experiment volunteers. *What* – herbal cold remedy or sugar solution, and cold severity (0 to 5 scale). *When* – Not specified. *Where* – Major pharmaceutical firm. *Why* – Scientists were testing the efficacy of an herbal compound on the severity of the common cold. *How* – The scientists set up a controlled experiment. *Variables* – There are two variables. Type of treatment (herbal or sugar solution) is categorical, and severity rating is quantitative. *Concerns* – The severity of a cold seems subjective and difficult to quantify. Also, the scientists may feel pressure to report negative findings about the herbal product.
- 35. Streams.** *Who* – Streams. *What* – Name of stream, substrate of the stream (limestone, shale, or mixed), acidity of the water (measured in pH), temperature (in degrees Celsius), and BCI (unknown units). *When* – Not specified. *Where* – Upstate New York. *Why* – Research is conducted for an Ecology class. *How* – Not specified. *Variables* – There are five variables. Name and substrate of the stream are categorical variables, and acidity, temperature, and BCI are quantitative variables.
- 37. Refrigerators.** *Who* – 353 refrigerator models. *What* – Brand, cost (probably in dollars), size (in cu. ft.), type, estimated annual energy cost (probably in dollars), overall rating, and repair history (in percent requiring repair over the past five years). *When* – 2013. *Where* – United States. *Why* – The information was compiled to provide information to the readers of *Consumer Reports*. *How* – Not specified. *Variables* – There are 7 variables. Brand, type, and overall rating are categorical variables. Cost, size, estimated energy cost, and repair history are quantitative variables.
- 39. Kentucky Derby 2016.** *Who* – Kentucky Derby races. *What* – Year, winner, jockey, trainer, owner, and time (in minutes, seconds, and hundredths of a second). *When* – 1875 – 2016. *Where* – Churchill Downs, Louisville, Kentucky. *Why* – Not specified. To examine the trends in the Kentucky Derby? *How* – Official statistics are kept for the race each year. *Variables* – There are 6 variables. Winner, jockey, trainer and owner are categorical variables. Date and duration are quantitative variables.

41. Kentucky Derby 2016 on the computer.

- a) Fonso was the winning horse in 1880.
- b) The length of the race changed in 1895, from 1.5 miles to 1.25 miles.
- c) The winning time in 1974 was 124 seconds.
- d) Secretariat ran the Derby in under 2 minutes in 1973, as did Monarchos in 2001.

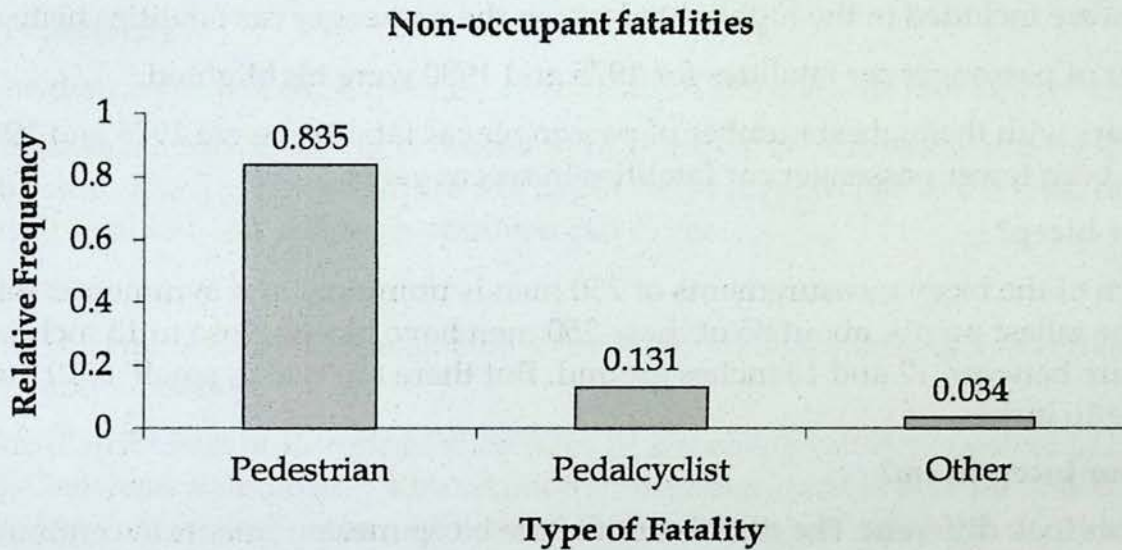
Chapter 2 – Displaying and Describing Data

Section 2.1

1. Automobile fatalities.

Subcompact and Mini	0.2658
Compact	0.2084
Intermediate	0.3006
Full	0.2069
Unknown	0.0183

2. Non-occupant fatalities.



3. Movie genres.

- A pie chart seems appropriate from the movie genre data. Each movie has only one genre, and the list of all movies constitute a “whole”.
- “Other” is the least common genre. It has the smallest region in the chart.

4. Movie ratings.

- A pie chart seems appropriate for the movie rating data. Each movie has only one rating, and the list of all movies constitute a “whole”. The percentages of each rating are different enough that the pie chart is easy to read.
- The most common rating is R. It has the largest region on the chart.

5. Movie ratings.

- i) C ii) A iii) D iv) B

6. Marriage in decline.

- i) D ii) A iii) C iv) B

6 Part I Exploring and Understanding Data

Section 2.2

7. Traffic Fatalities 2013.

- The gaps in the histogram for *Year* indicate that we do not have data for those years. This data set contains two variables for each case, and a histogram of the years doesn't give us much useful information.
- All of the bars in the *Year* histogram are the same height because each year only appears once in the data set.
- The distribution of passenger car fatalities has between 17,500 and 25,000 traffic fatalities per year in most years. There were also several years – possibly a second mode – with between 10,000 and 12,500 traffic fatalities.

8. Traffic Fatalities 2013 again.

- Two years were included in the highlighted bar in the passenger car fatalities histogram.
- The number of passenger car fatalities for 1975 and 1980 were highlighted.
- The two years with the highest number of passenger car fatalities were 1975 and 1980. There have been fewer passenger car fatalities in recent years.

9. How big is your bicep?

The distribution of the bicep measurements of 250 men is unimodal and symmetric. Based on the height of the tallest points, about 85 of these 250 men have biceps close to 13 inches around. Most are between 12 and 15 inches around. But there are two as small as 10 inches and several that are 16 inches.

10. How big is your bicep in cm?

Yes, the dotplots look different. The distribution of the bicep measurements in centimeters is still unimodal, but not as symmetric. The plot based on inches has fewer values on the *x*-axis, so it shows less detail. The measurements were rounded to the nearest half-inch after converting to inches, giving the plot a granular appearance. The plot in centimeters gives a better picture of the distribution.

11. E-mails.

The distribution of the number of emails received from each student by a professor in a large introductory statistics class during an entire term is skewed to the right, with the number of emails ranging from 1 to 21 emails. The distribution is centered at about 2 emails, with many students only sending 1 email. There is one outlier in the distribution, a student who sent 21 emails. The next highest number of emails sent was only 8.

12. Adoptions.

- The distributions of the number of adoptions and state populations are both skewed to the right. Most states have smaller populations and fewer adoptions, but some big states have substantially more of each.

- b) The distributions have similar shapes, since states with higher populations are likely to have more adoptions.
- c) The number of adoptions could be expressed as a rate. For example, report the number of adoptions per 100,000 people.

Section 2.3

13. Biceps revisited.

The distribution of the bicep measurements of 250 men is unimodal and roughly symmetric.

14. E-mails II.

The distribution of the number of emails received from each student by a professor in a large introductory statistics class during an entire term is skewed to the right.

15. Life expectancy.

- a) The distribution of life expectancies at birth in 190 countries is skewed to the left.
- b) The distribution of life expectancies at birth in 190 countries has one mode, at about 74 to 76 years. The fluctuations from bar to bar don't seem to rise to the level of defining additional modes, although opinions can differ.

16. Shoe sizes.

- a) The distribution of European shoe sizes of 269 college students is roughly symmetric and possibly bimodal.
- b) The distribution of European shoe sizes of 269 college students seems to have two modes, one between sizes 38 and 40, and another between sizes 44 and 46. This could be due to having data for both men and women. The lower mode may be the typical shoe size for women, and the upper mode may be the typical shoe size for men.

17. Life expectancy II.

- a) The distribution of life expectancies at birth in 190 countries is skewed to the left, so the median is expected to be larger than the mean. The mean life expectancy is pulled down toward the tail of the distribution.
- b) Since the distribution of life expectancies at birth in 190 countries is skewed to the left, the median is the better choice for reporting the center of the distribution. The median is more resistant to the skewed shape of the distribution.

18. Adoptions II.

- a) The distribution of the number of adoptions is skewed to the right, so the mean is expected to be larger. The mean number of adoptions is pulled up toward the tail of the distribution.
- b) Since the distribution of the number of adoptions is skewed to the right, the median is the better choice for reporting the center of the distribution. The median is more resistant to the skewed shape of the distribution.

19. How big is your bicep II?

Because the distribution of bicep circumferences is unimodal and symmetric, the mean and the median should be very similar. The usual choice is to report the mean or to report both.

20. Shoe sizes II.

Because the distribution of shoe sizes has two modes, the mean and median are not helpful in reporting the story that the data tell. It is better to report the locations of the two modes.

Section 2.5

21. Life expectancy III.

- a) We should report the IQR.
- b) Since the distribution of life expectancies at birth in 190 countries is skewed to the left, the better measure of spread is the IQR. The skewness of the distribution inflates the standard deviation.

22. Adoptions III.

- a) We should report the IQR.
- b) Since the distribution of the number of adoptions is skewed to the right, the IQR is the better choice for reporting the spread of the distribution. The skewness of the distribution inflates the standard deviation.

23. How big is your bicep III?

Because the distribution of bicep circumferences is unimodal and roughly symmetric, we should report the standard deviation. The standard deviation is generally more useful whenever it is appropriate. However, it would not be strictly wrong to use the IQR. We just prefer the standard deviation.

24. Shoe sizes III.

The data combine shoe sizes for men and for women. It isn't appropriate to summarize a bimodal distribution as if they were a single collection of values.

Chapter Exercises

- 25. **Graphs in the news.** Answers will vary.
- 26. **Graphs in the news II.** Answers will vary.
- 27. **Tables in the news.** Answers will vary.
- 28. **Tables in the news II.** Answers will vary.
- 29. **Histogram.** Answers will vary.
- 30. **Not a histogram.** Answers will vary.
- 31. **Centers in the news.** Answers will vary.
- 32. **Spreads in the news .** Answers will vary.

33. Thinking about shape.

- a) The distribution of the number of speeding tickets each student in the senior class of a college has ever had is likely to be unimodal and skewed to the right. Most students will have very few speeding tickets (maybe 0 or 1), but a small percentage of students will likely have comparatively many (3 or more?) tickets.
- b) The distribution of player's scores at the U.S. Open Golf Tournament would most likely be unimodal and slightly skewed to the right. The best golf players in the game will likely have around the same average score, but some golfers might be off their game and score 15 strokes above the mean. (Remember that high scores are undesirable in the game of golf!)
- c) The weights of female babies in a particular hospital over the course of a year will likely have a distribution that is unimodal and symmetric. Most newborns have about the same weight, with some babies weighing more and less than this average. There may be slight skew to the left, since there seems to be a greater likelihood of premature birth (and low birth weight) than post-term birth (and high birth weight).
- d) The distribution of the length of the average hair on the heads of students in a large class would likely be bimodal and skewed to the right. The average hair length of the males would be at one mode, and the average hair length of the females would be at the other mode, since women typically have longer hair than men. The distribution would be skewed to the right, since it is not possible to have hair length less than zero, but it is possible to have a variety of lengths of longer hair.

34. More shapes.

- a) The distribution of the ages of people at a Little League game would likely be bimodal and skewed to the right. The average age of the players would be at one mode and the average age of the spectators (probably mostly parents) would be at the other mode. The distribution would be skewed to the right, since it is possible to have a greater variety of ages among the older people, while there is a natural left endpoint to the distribution at zero years of age.
- b) The distribution of the number of siblings of people in your class is likely to be unimodal and skewed to the right. Most people would have 0, 1, or 2 siblings, with some people having more siblings.
- c) The distribution of pulse rate of college-age males would likely be unimodal and symmetric. Most males' pulse rates would be around the average pulse rate for college-age males, with some males having lower and higher pulse rates.
- d) The distribution of the number of times each face of a die shows in 100 tosses would likely be uniform, with around 16 or 17 occurrences of each face (assuming the die had six sides).

35. Movie genres again.

- a) Thriller/Suspense has a higher bar than Adventure, so it is the more common genre.
- b) It is easy to tell from either chart; sometimes differences are easier to see on the bar chart because slices of the pie chart look too similar in size.

10 Part I Exploring and Understanding Data

36. Movie ratings, again.

- The least common rating was NC-17. It has the shortest bar.
- It is easy to tell from either chart; sometimes differences are easier to see on the bar chart because slices of the pie chart look too similar in size.

37. Magnet Schools.

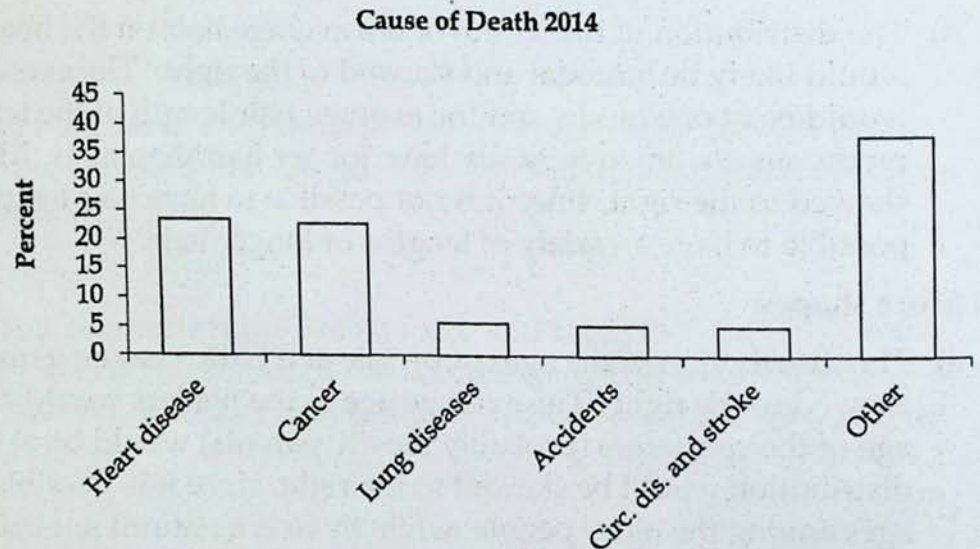
There were 1755 qualified applicants for the Houston Independent School District's magnet schools program. 53% were accepted, 17% were wait-listed, and the other 30% were turned away for lack of space.

38. Magnet schools again.

There were 1755 qualified applicants for the Houston Independent School District's magnet schools program. 29.5% were Black or Hispanic, 16.6% were Asian, and 53.9% were white.

39. Causes of death 2014.

- Yes, it is reasonable to assume that heart or lung diseases caused approximately 29% of U.S. deaths in 2014, since there is no possibility for overlap. Each person could only have one cause of death.

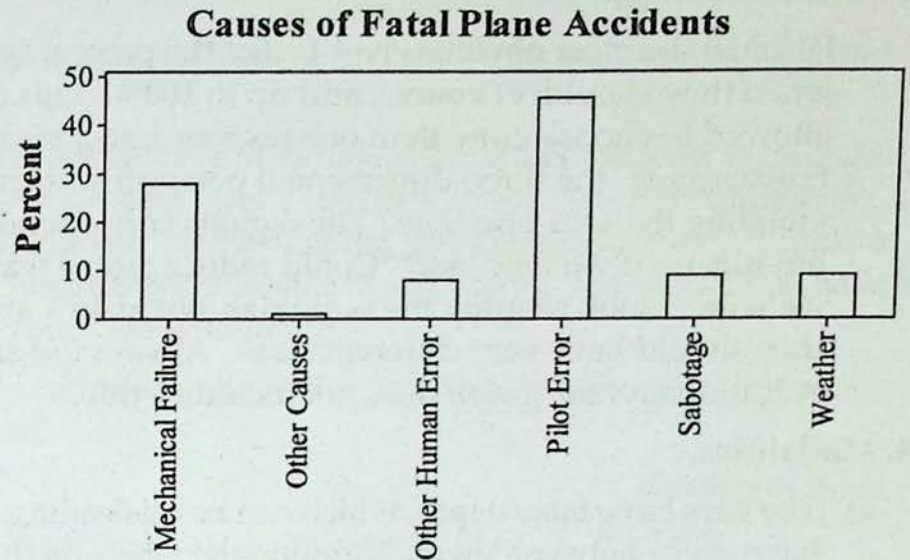


- Since the percentages listed add up to 61.9%, other causes must account for 38.1% of US deaths.
- A bar chart is a good choice (with the inclusion of the "Other" category). Since causes of US deaths represent parts of a whole, a pie chart would also be a good display.

40. Plane crashes.

- As long as each plane crash had only one cause, it would be reasonable to assume that weather or mechanical failures were the causes of about 37% of crashes.
- It is likely that the numbers in the table add up to 101% due to rounding.

- c) A relative frequency bar chart is a good choice. A pie chart would also be a good display, as long as each plane crash has only one cause.

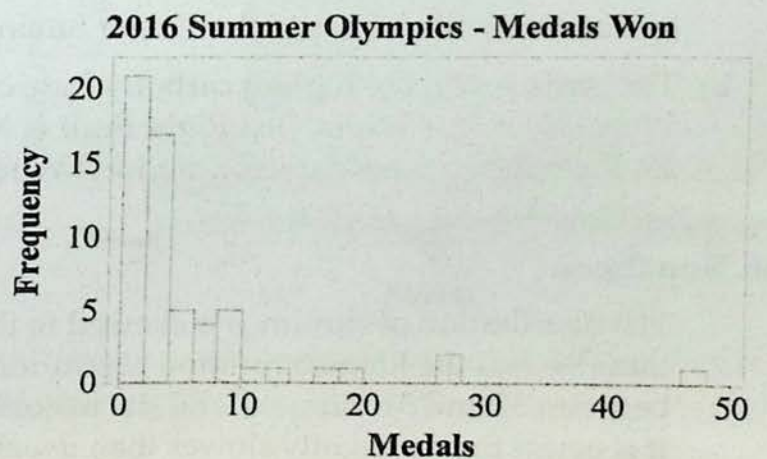


41. Movie genres once more.

- a) There are too many categories to construct an appropriate display. In a bar chart, there are too many bars. In a pie chart, there are too many slices. In each case, we run into difficulty trying to display genres that only represented a few movies.
- b) The creators of the bar chart included a category called "Other" for many of the genres that only occurred a few times.

42. Summer Olympics 2016.

- a) There are too many categories to construct an appropriate display. In a bar chart, there are too many bars. In a pie chart, there are too many slices. In each case, we run into difficulty trying to display those countries that didn't win many medals.
- b) Perhaps we are primarily interested in countries that won many medals. We might choose to combine all countries that won fewer than 6 medals into a single category. This will make our chart easier to read. But, we are probably more interested in the number of medals won overall, and don't need to know what countries won the medals. A histogram is probably the most appropriate display. Now we can see that most countries won very few medals, and a handful of countries won many medals per country.



43. Global warming.

Perhaps the most obvious error is that the percentages in the pie chart add up to 141%, when they should, of course, add up to 100%. This means that survey respondents were allowed to choose more than one response, so a pie chart is not an appropriate display. Furthermore, the three-dimensional perspective view distorts the regions in the graph, violating the area principle. The regions corresponding to "Could reduce global warming but unsure if we will" and "Could reduce global warming but people aren't willing to so we won't" look roughly the same size, but at 46% and 30% of respondents, respectively, they should have very different sizes. Always use simple, two-dimensional graphs. Additionally, the graph does not include a title.

44. Modalities.

- a) The bars have false depth, which can be misleading. This is a bar chart, so the bars should have space between them. Running the labels on the bars from top to bottom and the vertical axis labels from bottom to top is confusing.
- b) The percentages sum to 100%. Normally, we would take this as a sign that all of the observations had been correctly accounted for. But in this case, it is extremely unlikely. Each of the respondents was asked to list *three* modalities. For example, it would be possible for 80% of respondents to say they use ice to treat an injury, and 75% to use electric stimulation. The fact that the percentages total greater than 100% is not odd. In fact, in this case, it seems wrong that the percentages add up to 100%, rather than correct.

45. Cereals.

- a) The distribution of the carbohydrate content of breakfast cereals is bimodal, with a cluster of cereals with carbohydrate content around 13 grams of carbs and another cluster of cereals around 22 grams of carbs. The lower cluster shows a bit of skew to the left. Most cereals in the lower cluster have between 10 and 20 grams of carbs. The upper cluster is symmetric, with cereals in the cluster having between 20 and 24 grams of carbs.
- b) The cereals with the highest carbohydrate content are Corn Chex, Corn Flakes, Cream of Wheat (Quick), Crispix, Just Right Fruit & Nut, Kix, Nutri-Grain Almond-Raisin, Product 19, Rice Chex, Rice Krispies, Shredded Wheat 'n' Bran, Shredded Wheat Spoon Size, Total Corn Flakes, and Triples.

46. Run times.

The distribution of runtimes is skewed to the right. The shortest runtime was around 28.5 minutes and the longest runtime was around 35.5 minutes. A typical run time was between 30 and 31 minutes, and the majority of runtimes were between 29 and 32 minutes. It is easier to run slightly slower than usual and end up with a longer runtime than it is to run slightly faster than usual and end up with a shorter runtime. This could account for the skew to the right seen in the distribution.

47. Heart attack stays.

- a) The distribution of length of stays is skewed to the right, so the mean is larger than the median.
- b) The distribution of the length of hospital stays of female heart attack patients is bimodal and skewed to the right, with stays ranging from 1 day to 36 days. The distribution is centered around 8 days, with the majority of the hospital stays lasting between 1 and 15 days. There are a relatively few hospital stays longer than 27 days. Many patients have a stay of only one day, possibly because the patient died.
- c) The median and IQR would be used to summarize the distribution of hospital stays, since the distribution is strongly skewed.

48. Bird species 2013.

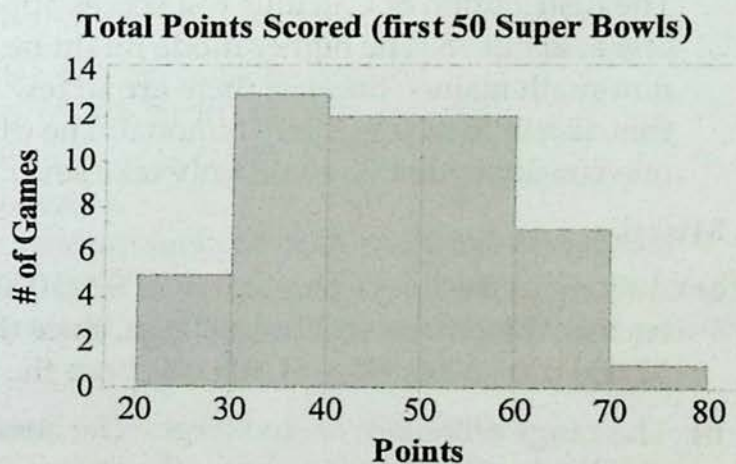
- a) The results of the 2013 Laboratory of Ornithology Christmas Bird Count are displayed in the stem and leaf display at the right.
- b) The distribution of the number of birds spotted by participants in the 2013 Laboratory of Ornithology Christmas Bird Count is skewed right, with a median of 117 birds. There are three high potential outliers, with participants spotting 150, 166, and 184 birds. With the exception of these outliers, most participants saw between 82 and 136 birds.

	Number of Birds
8	2368
9	78
10	1156
11	8
12	468
13	136
14	
15	0
16	6
17	
18	4

Key: 15 | 0 = 150 birds

49. Super Bowl points 2016.

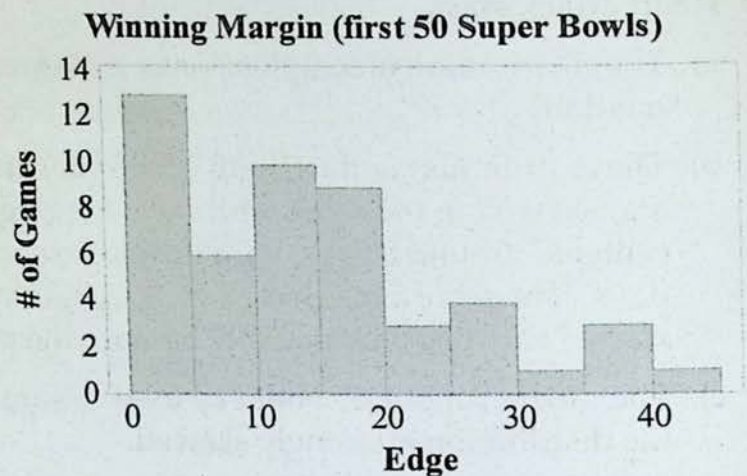
- a) The median number of points scored in the first 50 Super Bowl games is 46 points.
- b) The first quartile of the number of points scored in the first 50 Super Bowl games is 37 points. The third quartile is 55 points.
- c) In the first 50 Super Bowl games, the lowest number of points scored was 21, and the highest number of points scored was 75. The median number of points scored was 46, and the middle 50% of Super Bowls has between 37 and 55 points scored, making the IQR 18 points.



50. Super Bowl edge 2016.

- a) The median winning margin in the first 50 Super Bowl games is 12.5 points.
- b) The first quartile of the winning margin in the first 50 Super Bowl games is 4 points. The third quartile is 19 points.

- c) In the first 50 Super Bowl games the lowest winning margin was 1 point and the highest winning margin was 45 points. The median winning margin was 12.5 points, with the middle 50% of winning margins between 4 and 19 points, making the IQR 15 points.



51. Test scores, large class.

- a) The distribution of Calculus test scores is bimodal with one mode at about 62 and one at about 78. The higher mode might be math majors, and the lower mode might be non-math majors.
- b) Because the distribution of Calculus test scores is bimodal, neither the mean nor the median tells much about a typical score. We should attempt to learn if another variable (such as whether or not the student is a math major) can account for the bimodal character of the distribution.

52. Test scores, small class.

The distribution of Calculus test scores appears bimodal, with one mode at about 62 and one at about 78. The higher mode might be math majors, and the lower mode might be non-math majors. Because there are so few values in the middle bins, it is not as clear that this distribution is actually bimodal. The distribution might merely be skewed to the left, if one considers that it would only take several scores in the middle bars to fill in the "gap".

53. Mistake.

- a) As long as the boss's true salary of \$200,000 is still above the median, the median will be correct. The mean will be too large, since the total of all the salaries will decrease by $\$2,000,000 - \$200,000 = \$1,800,000$, once the mistake is corrected.
- b) The range will likely be too large. The boss's salary is probably the maximum, and a lower maximum would lead to a smaller range. The IQR will likely be unaffected, since the new maximum has no effect on the quartiles. The standard deviation will be too large, because the \$2,000,000 salary will have a large squared deviation from the mean.

54. Sick days.

- a) The company probably uses the mean, while the union uses the median number of sick days. The mean will likely be higher, since it is affected by probable right skew. Some employees may have many sick days, while most have relatively few.

- b) These choices for summaries make sense: The company cares about the number of days lost to sickness, so the mean (related to the total) is what matters. The union looks at individual workers and sees that the typical (median) worker doesn't miss many days for sickness. Additionally, each side is trying to make their side's position in the negotiation look better. The union benefits from a perception of lower numbers of sick days, while the company benefits from the perception of a higher number of sick days, which need to be curtailed.

55. Floods 2015.

- a) The mean annual number of deaths from floods is 81.95.
- b) In order to find the median and the quartiles, the list must be ordered.
 29 38 38 43 48 49 56 68 76 80 82 82 82 86 87 103 113 118 131 136 176
 The median annual number of deaths from floods is 82.
 Quartile 1 = 49 deaths, and Quartile 3 = 103 deaths.
 (Some statisticians consider the median to be separate from both the lower and upper halves of the ordered list when the list contains an odd number of elements. This changes the position of the quartiles slightly. If median is excluded, $Q1 = 48.5$, $Q3 = 108$. In practice, it rarely matters, since these measures of position are best for large data sets.)
- c) The range of the distribution of deaths is $\text{Max} - \text{Min} = 176 - 29 = 147$ deaths.
 The $\text{IQR} = Q3 - Q1 = 103 - 49 = 54$ deaths. (Or, the $\text{IQR} = 108 - 48.5 = 59.5$ deaths, if the median is excluded from both halves of the ordered list.)

56. Tornadoes 2015.

- a) The mean annual number of deaths from tornadoes is 81.43.
- b) In order to find the median and the quartiles, the list must be ordered.
 21 25 30 35 36 38 40 41 45 47 54 55 55 67 67 70 81 94 126 130 553
 The median annual number of deaths from tornadoes is 54.
 Quartile 1 = 38 deaths, and Quartile 3 = 70 deaths.
 (Some statisticians consider the median to be separate from both the lower and upper halves of the ordered list. This changes the position of the quartiles slightly. If median is included, $Q1 = 37$, $Q3 = 75.5$. In practice, it rarely matters, since these measures of position are best for large data sets.)
- c) The range of the distribution of deaths is $\text{Max} - \text{Min} = 553 - 21 = 532$ deaths.
 The $\text{IQR} = Q3 - Q1 = 70 - 38 = 32$ deaths. (Or, the $\text{IQR} = 75.5 - 37 = 38.5$ deaths, if the median is excluded from both halves of the ordered list.)

57. Floods 2105 II.

The distribution of deaths from floods is slightly skewed to the right and bimodal. There is one mode at about 40 deaths and one at about 80 deaths. There is one extreme value at 180 deaths.

58. Tornadoes 2015 II.

The distribution of deaths from tornadoes is slightly skewed to the right, with one extreme outlier at 553. The median is 54 deaths, and the IQR is 32 deaths.

59. Pizza prices.

The mean and standard deviation would be used to summarize the distribution of pizza prices, since the distribution is unimodal and symmetric.

60. Neck size.

The mean and standard deviation would be used to summarize the distribution of neck sizes, since the distribution is unimodal and symmetric.

61. Pizza prices again.

- a) The mean pizza price is closest to \$2.60. That's the balancing point of the histogram.
- b) The standard deviation in pizza prices is closest to \$0.15, since that is the typical distance to the mean. There are no pizza prices as far as \$0.50 or \$1.00.

62. Neck sizes again.

- a) The mean neck size is closest to 15 inches. That's the balancing point of the histogram.
- b) The standard deviation in neck sizes is closest to 1 inch, because a typical value lies about 1 inch from the mean. There are a few points as far away as 3 inches from the mean, and none as far away as 5 inches. Those are too large to be the standard deviation.

63. Movie lengths 2010.

- a) A typical movie would be around 105 minutes long. This is near the center of the unimodal and slightly skewed histogram, with the outlier set aside.
- b) You would be surprised to find that your movie ran for 150 minutes. Only 3 movies ran that long.
- c) The mean run time would probably be higher, since the distribution of run times is skewed to the right, and also has a high outlier. The mean is pulled towards this tail, while the median is more resistant. However, it is difficult to predict what the effect of the low outlier might be from just looking at the histogram.

64. Golf drives 2015.

- a) The distribution of golf drives is roughly unimodal and symmetric, with a typical drive of around 290 yards. Professional golfers on the men's PGA tour had drives that were as short as about 260 yards, and as long as about 320 yards.
- b) Approximately 25% of professional male golfers drive less than 280 yards.
- c) The actual mean drive is about 288.69 yards, so any estimate between 285 and 290 yards is reasonable.
- d) The distribution of golf drives is approximately symmetric, so the mean and the median should be relatively close. The actual median is 288.7.

65. Movie lengths 2010 II.

- a) i) The distribution of movie running times is fairly consistent, with the middle 50% of running times between 98 and 116 minutes. The interquartile range is 18 minutes.
- ii) The standard deviation of the distribution of movie running times is 16.6 minutes, which indicates that movies typically varied from the mean running time by 16.6 minutes.
- b) Since the distribution of movie running times is skewed to the right and contains an outlier, the standard deviation is a poor choice of numerical summary for the spread. The interquartile range is better, since it is resistant to outliers.

66. Golf drives II.

- a) i) The distribution of PGA golf drives is fairly consistent, with the middle 50% of the drives having distances between 282 and 294.5 yards. The interquartile range is 12.5 yards.
- ii) The standard deviation of the distribution of PGA golf drives is 9.8 yards, which indicates that golf drives are typically within 9.8 yards of the mean gold drive.
- b) Since the distribution of golf drives is reasonably symmetric, both the standard deviation and the interquartile range are reasonable measures of spread.

67. Movie budgets.

The industry publication is using the median, while the watchdog group is using the mean. It is likely that the mean is pulled higher by a few very expensive movies.

68. Cold weather.

- a) The mean temperature will be lower. The median temperature will not change, since the incorrect temperature is still the lowest temperature, and the median is based only on position.
- b) The range and standard deviation in temperature will both increase, since the incorrect temperature is more extreme than the correct temperature. The IQR will not change, since the both the correct and incorrect scores are below the first quartile, and the IQR measures the distance between the first and third quartiles.

69. Gasoline 2014.

a) Gasoline Prices

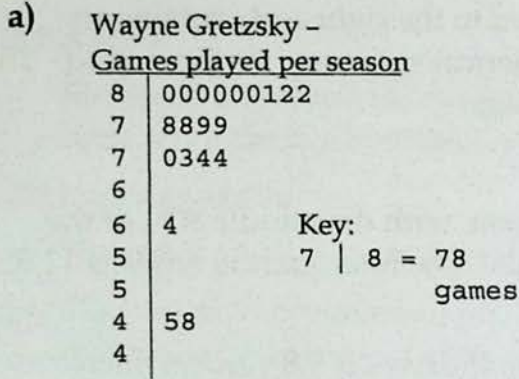
31		1
31		5
32		1233
32		6678
33		
33		9
34		23
34		556

Key : 32 | 1 = \$3.21/gal

18 **Part I Exploring and Understanding Data**

- b) The distribution of gas prices is bimodal, with two clusters, one centered around \$3.45 per gallon, and another centered around \$3.25 per gallon. The lowest and highest prices were \$3.11 and \$3.46 per gallon.
- c) There is a gap in the distribution of gasoline prices. There were no stations that charged between \$3.28 and \$3.39.

70. **The great one.**



- b) The distribution of the number of games played by Wayne Gretzky is unimodal and skewed to the left.
- c) Typically, Wayne Gretzky played about 80 games per season. The number of games played is tightly clustered in the upper 70s and low 80s.
- d) Two seasons are low outliers, when Gretzky played fewer than 50 games. He may have been injured during those seasons. Regardless of any possible reasons, these seasons were unusual compared to Gretzky's other seasons.

71. **States.**

- a) There are 50 entries in the stemplot, so the median must be between the 25th and 26th population values. Counting in the ordered stemplot gives median = 4.5 million people. The middle of the lower 50% of the list (25 state populations) is the 13th population, or 2 million people. The middle of the upper half of the list (25 state populations) is the 13th population from the top, or 7 million people. The IQR = $Q_3 - Q_1 = 7 - 2 = 5$ million people.
- b) The distribution of population for the 50 U.S. States is unimodal and skewed heavily to the right. The median population is 4.5 million people, with 50% of states having populations between 2 and 7 million people. There are two outliers, a state with 37 million people, and a state with 25 million people. The next highest population is only 19 million.

72. **Wayne Gretzky.**

- a) The distribution of the number of games played per season by Wayne Gretzky is skewed to the left, and has low outliers. The median is more resistant to the skewness and outliers than the mean.

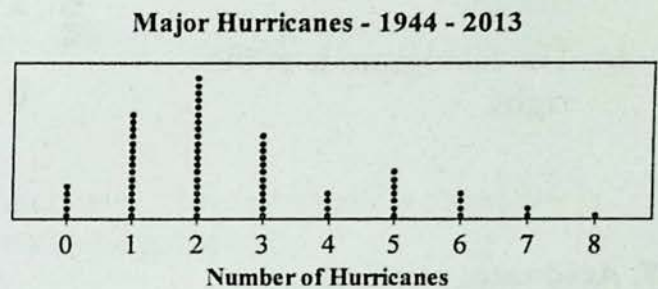
- b) The median, or middle of the ordered list, is 79 games. Both the 10th and 11th values are 79, so the median is the average of these two, also 79.
- c) The mean should be lower. There are two seasons when Gretzky played an unusually low number of games. Those seasons will pull the mean down.

73. A-Rod 2016.

The distribution of the number of homeruns hit by Alex Rodriguez during the 1994 - 2016 seasons is reasonably symmetric, with the exception of a second mode around 10 homeruns. A typical number of homeruns per season was in the high 30s to low 40s. With the exception of 5 seasons in which A-Rod hit 0, 0, 5, 7, and 9 homeruns, his total number of homeruns per season was between 16 and the maximum of 57.

74. Major hurricanes 2013.

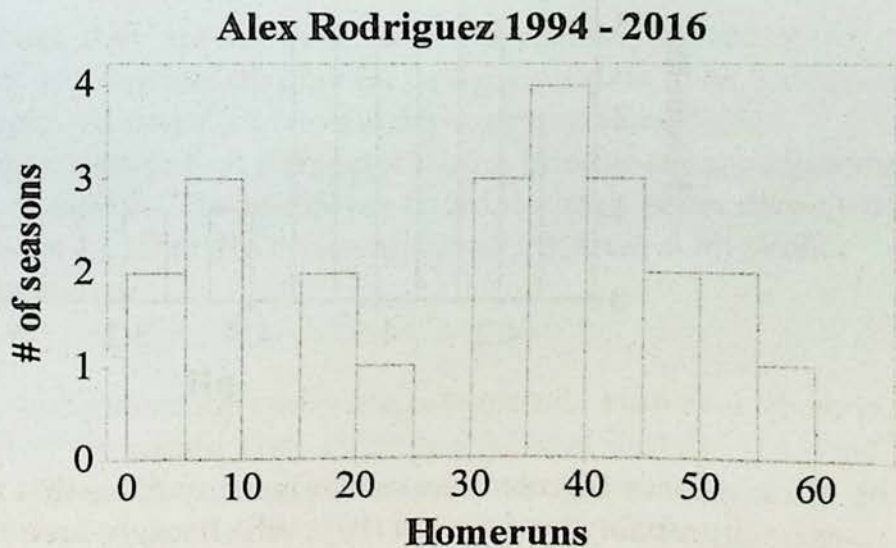
- a) A dotplot of the number of hurricanes each year from 1944 through 2013 is displayed. Each dot represents a year in which there were that many hurricanes.



- b) The distribution of the number of hurricanes per year is unimodal and skewed to the right, with center around 2 hurricanes per year. The number of hurricanes per year ranges from 0 to 8. There are no outliers. There may be a second mode at 5 hurricanes per year, but since there were only 6 years in which 5 hurricanes occurred, this may simply be natural variability.

75. A-Rod again 2016.

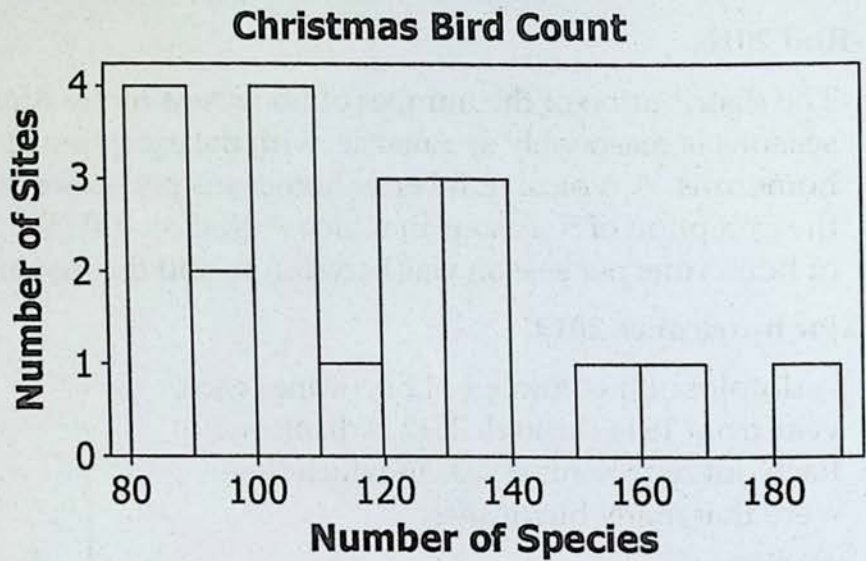
- a) This is not a histogram. The horizontal axis should contain the number of home runs per year, split into bins of a convenient width. The vertical axis should show the frequency; that is, the number of years in which A-Rod hit a number of home runs within the interval of each bin. The display shown is a bar chart/time plot hybrid that simply displays the data table visually. It is of no use in describing the shape, center, spread, or unusual features of the distribution of home runs hit per year by A-Rod.



- b) The histogram is at the right.

76. Return of the birds 2013.

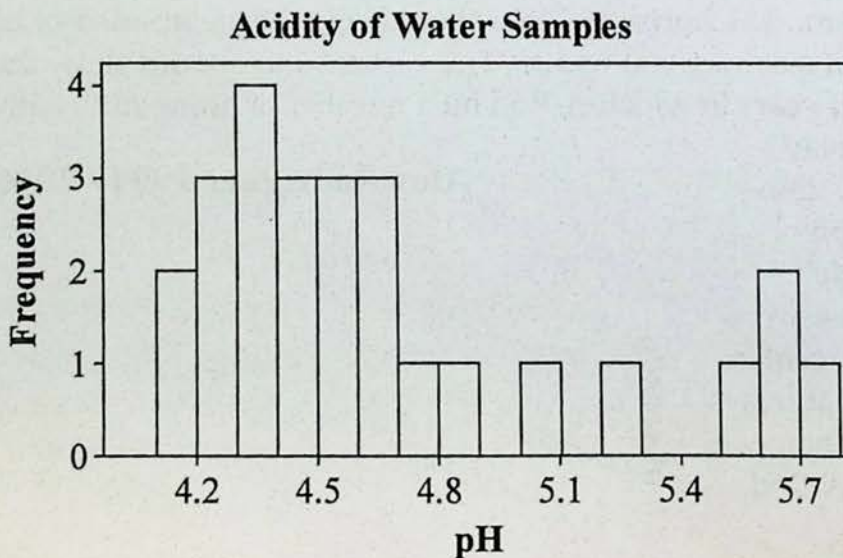
a) This is not a histogram. The horizontal axis should split the number of counts from each site into bins. The vertical axis should show the number of sites in each bin. The given graph is nothing more than a bar chart, showing the bird count from each site as its own bar. It is of absolutely no use for describing the shape, center, spread, or unusual features of the distribution of bird counts.



b) The histogram is at the right.

77. Acid rain.

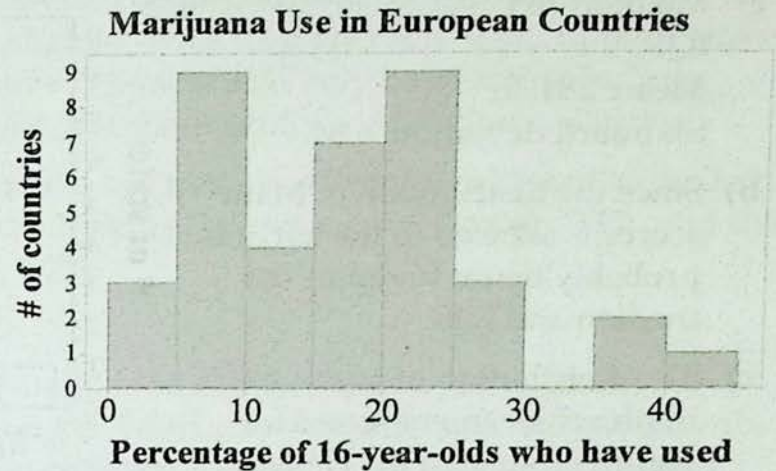
a) The distribution of the pH readings of water samples in Allegheny County, Penn. is bimodal. A roughly uniform cluster is centered around a pH of 4.4. This cluster ranges from pH of 4.1 to 4.9. Another smaller, tightly packed cluster is centered around a pH of 5.6. Two readings in the middle seem to belong to neither cluster.



b) The cluster of high outliers contains many dates that were holidays in 1973. Traffic patterns would probably be different then, which might account for the difference.

78. Marijuana 2015.

The distribution of the percentage of 16-year-olds in 38 countries who have used marijuana in 2015 is somewhat bimodal, with a group of countries having approximately 5 to 10% of 16-year-olds having used marijuana. Another group of countries has between 15% and 25% of teens who have used marijuana. Kosovo, at 2%, had the lowest percentage of 16-year-olds who have tried marijuana. Czech Republic had the highest percentage, at 42%. A typical country might have a percentage of approximately 18%, the median percentage of marijuana use among 16-year-olds.



79. Final grades.

The width of the bars is much too wide to be of much use. The distribution of grades is skewed to the left, but not much more information can be gathered.

80. Final grades revisited.

- a) This display has a bar width that is much too narrow. As it is, the histogram is only slightly more useful than a list of scores. It does little to summarize the distribution of final exam scores.
- b) The distribution of test scores is skewed to the left, with center at approximately 170 points. There are several low outliers below 100 points, but other than that, the distribution of scores is fairly tightly clustered.

81. Zip codes.

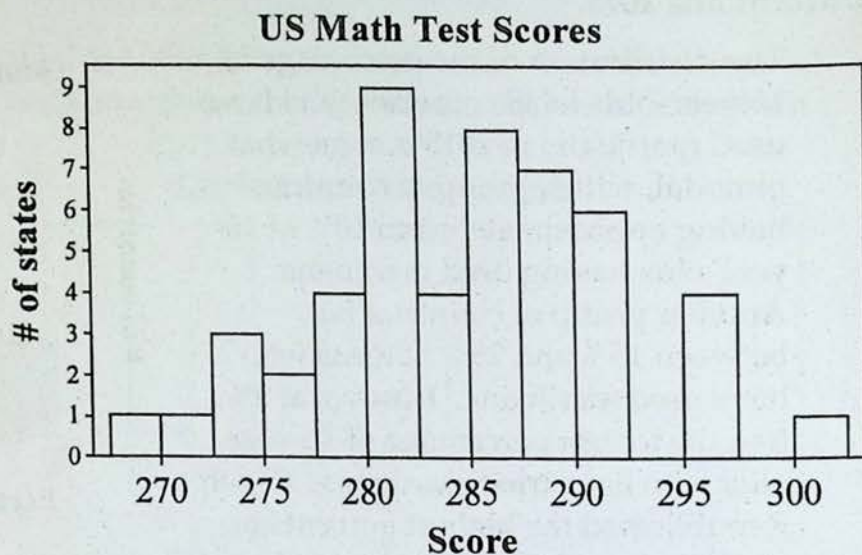
Even though zip codes are numbers, they are not quantitative in nature. Zip codes are categories. A histogram is not an appropriate display for categorical data. The histogram the Holes R Us staff member displayed doesn't take into account that some 5-digit numbers do not correspond to zip codes or that zip codes falling into the same classes may not even represent similar cities or towns. The employee could design a better display by constructing a bar chart that groups together zip codes representing areas with similar demographics and geographic locations.

82. Zip codes revisited

The statistics cannot tell us very much since zip codes are categorical. However, there is *some* information in the first digit of zip codes. They indicate a general East (0-1) to West (8-9) direction. So, the distribution shows that a large portion of their sales occurs in the West and another in the 32000 area. But a bar chart of the first digits would be the appropriate display to show this information.

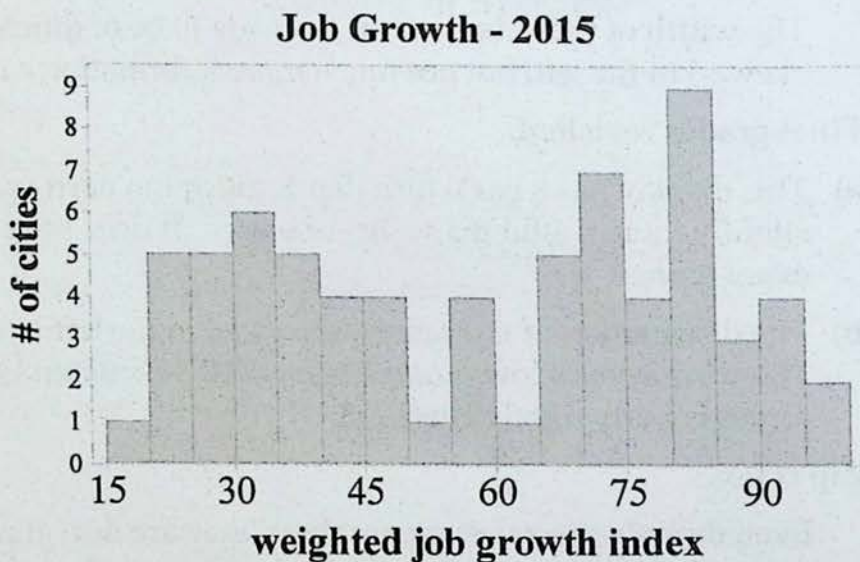
83. Math scores 2013.

- a) Median: 285
IQR: 9
Mean: 284.36
Standard deviation: 6.84
- b) Since the distribution of Math scores is skewed to the left, it is probably better to report the median and IQR.
- c) The distribution of average math achievement scores for eighth graders in the United States is skewed slightly to the left, and roughly unimodal. The distribution is centered at 285. Scores range from 269 to 301, with the middle 50% of the scores falling between 280 and 289.



84. Boomtowns 2015.

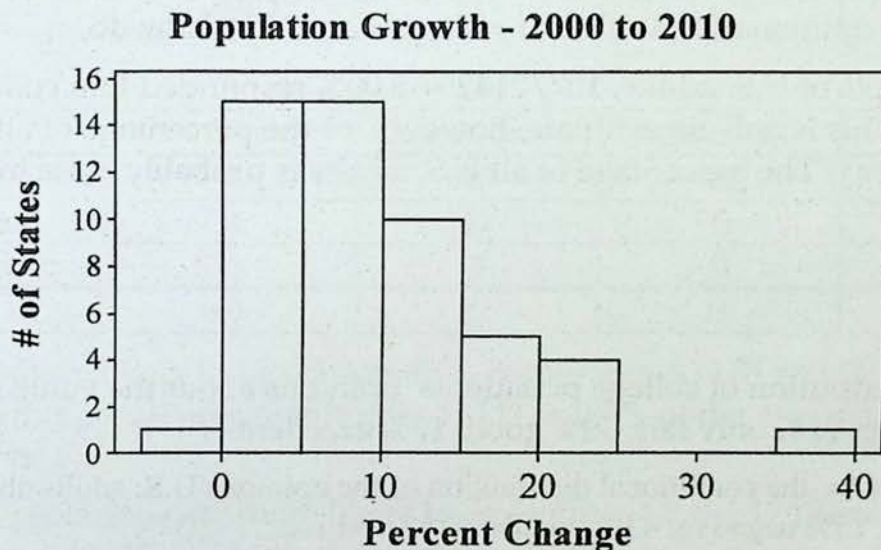
- a) A histogram of the job growth rates of NewGeography.com's best cities for job growth is at the right. A boxplot, stemplot, or dotplot would also have been an acceptable display.
- b) The mean weighted job rating index is 58.26% and the median weighted job rating index is 61.70%. The mean is lower than the median, indicating that the distribution is skewed to the left. At the very least, the distribution is bimodal, which makes the mean and median difficult to predict without calculations.
- c) It might be more informative to report one mode at about 33% and the other mode at about 80%.
- d) The standard deviation of the distribution of weighted job rating indices is 23.85% and the IQR is 45.1%.
- e) Neither is ideal. It might be more informative to discuss a measure of spread for each mode.
- f) If 49.23% were subtracted from each of the weighted job rating indices, the mean and median would each decrease by 49.23%. The standard deviation and the IQR would not change.



- g) If we were to set aside Austin-Round Rock, the 4th-highest weighted job rating index, the mean and standard deviation would decrease. The median and IQR would be relatively unaffected, although they would change slightly, since they are each based upon relative position. With the 4th-highest rating removed, there would only be 69 rating indices, instead of 70. This would cause the median and the quartiles to shift down slightly.
- h) The distribution of weighted job rating indices is bimodal and slightly skewed to the left. The lower mode is centered at a weighted job rating index of approximately 33%, and the upper mode is centered at approximately 80%.

85. Population growth 2010.

The distribution of population growth among the 50 United States and the District of Columbia is unimodal and skewed to the right. Most states experienced modest growth, as measured by percent change in population between 2000 and 2010. Nearly every state experienced positive growth, with the exception of Michigan. The median population growth was 7.8%, with the middle 50% of states experiencing between 4.30% and 14.10% growth, for an IQR of 9.80. The distribution contains one high outlier. Nevada experienced population growth of 35.1%.



86. Student survey.

Answers will vary. The best answers are about students and their responses, not about numbers and displays. Write what you learn about students, not about what your display looks like. Don't give the value of a summary statistic without a discussion of what it might mean.

Chapter 3 - Relationships Between Categorical Variables - Contingency Tables

Section 3.1

1. College value?

	Poor	Only Fair	Good	Excellent	DK/NA	Total
US Adults	321	900	750	107	64	2142
Presidents	32	222	622	179	0	1055
Total	353	1122	1372	286	64	3197

- The percent of college presidents who think that higher education provides a poor value is $32/1055 \approx 3\%$.
- $(750 + 107) / 2142 \approx 40\%$ of U.S. adults think that the value provided by the higher education system is either good or excellent.
- 15% of U.S. adults view college as a poor value, but only 3% of college presidents do. Similarly, U.S. adults are twice as likely to view college as an only fair value compared to the presidents (42% to 21%). Presidents are much more likely (76%) to rate colleges as a good or excellent value compared to U.S. adults (only 40%). So in short, college presidents have a much higher opinion of the value of college than U.S. adults do.
- In this random sample of U.S. adults, $107/2142 \approx 5.00\%$ responded that college provides an excellent value. This is only an estimate, however, of the percentage of all U.S. adults who feel the same way. The percentage of all U.S. adults is probably close to, but not exactly 5%.

Section 3.2

3. College value again.

- The conditional distribution of college presidents' opinions about the value of a college education is 3% poor; 21% only fair; 59% good; 17% excellent.
- Omitting the 64 DK/NA's, the conditional distribution of the opinions U.S. adults about the value of a college education is 15% negative; 43% middle; 41% positive.

Section 3.3

5. Diet and politics.

- The distribution of political alignment among carnivores is about 25% conservative, 40% moderate, and 35% liberal. Omnivores were generally less conservative and more liberal, with about 10% conservative, 35% moderate, and 55% liberal. Vegetarians were even more liberal than the other groups, with 0% conservative, 30% moderate, and 70% liberal.
- The differences are quite large. There appears to be a strong association between diet preference and political alignment.

7. Fish and prostate cancer revisited

- a) Looking at the horizontal axis only, approximately 7% of the men had prostate cancer.
- b) There are more men who didn't have cancer and never or seldom ate fish. The rectangles are approximately the same height, but the bar for "no cancer" is much wider.
- c) The percentage of men who never/seldom ate fish is lower in the group with no cancer than in the group with cancer. Disregard the width, and look only at the height to compare the conditional distribution of fish consumption within each cancer group. The bar for "never/seldom" is slightly shorter within the "no cancer" group.

Section 3.4

9. Diet and politics III.

Men

	Carnivore	Omnivore	Vegetarian	Total
Liberal	9	74	5	88
Moderate	12	54	1	67
Conservative	9	14	0	23
Total	30	142	6	178

Women

	Carnivore	Omnivore	Vegetarian	Total
Liberal	4	53	12	69
Moderate	4	27	6	37
Conservative	1	4	0	5
Total	9	84	18	111

- a) Men are more likely to be conservative carnivores. $9/178 \approx 5.1\%$ of the men are conservative carnivores, while only $1/111 \approx 0.9\%$ of the women are conservative carnivores.
- b) Liberal vegetarians are more likely to be women. Of the 17 liberal vegetarians, 12 of them are women. $12/17 \approx 70.6\%$ of liberal vegetarians are women.

Chapter Exercises

11. Movie genres and ratings.

- a) $452/1529 \approx 29.6\%$ of the films were rated R.
- b) $124/1529 \approx 8.1\%$ of the films were R-rated comedies.
- c) $124/452 \approx 27.4\%$ of the R-rated films were comedies.
- d) $124/312 \approx 39.7\%$ of the comedies were rated R.

13. Tables in the news. Answers will vary.

15. Poverty and region 2012.

The differences in poverty are not huge, but they may be real. The Northeast and Midwest have the lowest percentages of people living below the poverty level: 12.7% and 13.7%, respectively. In the West, 15.4% live below the poverty level, and the South has the highest rate at 16.8%.

17. Death from the sky.

- a) $100 - 60 - 30 - 1 - 0.2 - 0.17 = 8.63\%$ of estimated deaths are attributed to causes not listed here.
- b) Regardless of the type of display chosen, it is difficult to display causes of death with percentages as low as 0.2% and 0.17%.

19. Smoking.

- a) The smoking rate for 18-24-year-old men was 42.1% in 1974.
- b) From 1974 to 2014, the smoking rate for 18-24-year-old men dropped from 42.1% to 18.5%
- c) Men who were 18-24 years old in 1974 are in the 35-44 age group in 1994, the 45-54 age group in 2004, and the 2014. The smoking rate for this cohort has been decreasing through the years, from 42.1% to 33.2% to 26.7% to 18.8%. Although we don't have data on deaths in this table, it may very well be that the smokers have a higher death rate than the non-smokers, so this decrease doesn't necessarily mean that men in this cohort are quitting smoking.

21. Mothers and fathers 1965-2011.

- a) Fathers spend the vast majority of their time on paid work, while mothers spend more time on child care and house work.
- b) The time fathers spend on paid work has decreased, and the time they spend on child care and housework has increased. For mothers, the number of hours spent on paid work has significantly increased, and they have also increased their time spent on child care while reducing housework time.
- c) Parents are spending more time on child care and paid work (13 hours to 21 hours and 50 hours to 58 hours). The time spent on housework has decreased from 36 hours to 28 hours.
- d) Overall, parents in 2011 reported spending more time total on these tasks, a total of 107 hours in 2011 compared to 99 hours in 1965. Mothers increased their total working time by 3 hours, from 50 hours to 53 hours, while fathers increased their total working time by 5 hours, from 49 hours to 54 hours.

23. Teen smokers.

According to the Monitoring the Future study, teen smoking brand preferences differ somewhat by region. Although Marlboro is the most popular brand in each region, with about 58% of teen smokers preferring this brand in each region, teen smokers from the South prefer Newports at a higher percentage than teen smokers from the West, 22.5% to approximately 10%, respectively. Camels are more popular in the West, with 9.5% of teen smokers preferring this brand, compared to only 3.3% in the South. Teen smokers in the West are also more likely to have no particular brand than teen smokers in the South. 12.9% of teen smokers in the West have no particular brand, compared to only 6.7% in the South. Both regions have about 9% of teen smokers that prefer one of over 20 other brands.

25. Diet and politics IV.

- a) There are more men in the survey. The male columns are generally wider than the female columns.
- b) We can't compare the genders within each category of political ideology since the sample sizes differ. We can, however, note that the male bars are narrowest in the Liberal category, and widest in the Conservative category, indicating that a small share of liberals and a large share of conservatives among the men. The women show the opposite association. The widest female bar is Liberal, while the narrowest is Conservative. In other words, there is an association between gender and political ideology. Males tend to be more conservative and females tend to be more liberal.
- c) There is an association between politics and diet. Conservatives are more likely to be carnivores, while liberals are more likely to be vegetarians.
- d) The association between politics and diet seems to differ between men and women. Differences in vegetarianism across political ideology is more pronounced in females than in males. Differences in carnivorous eating habits across political ideology is more pronounced in males than females.

27. Job satisfaction.

- a) This is a table of column percents. The columns add up to 100%, while the rows do not.
- b)
 - i) This can't be found from the table. We don't know what the percent of respondents who are very satisfied.
 - ii) This can't be found from the table. We don't know what the percent of respondents who are dissatisfied.
 - iii) 39% of respondents who are dissatisfied with their current job are actually better off than their parents were at the same age.
 - iv) This can't be found from the table. We don't know what the percent of respondents who are very satisfied.

29. Seniors.

- a) A table with marginal totals is to the right. There are 268 White graduates and 325 total graduates. $268/325 \approx 82.5\%$ of the graduates are white.
- b) There are 42 graduates planning to attend 2-year colleges. $42/325 \approx 12.9\%$
- c) 36 white graduates are planning to attend 2-year colleges. $36/325 \approx 11.1\%$
- d) 36 white graduates are planning to attend 2-year colleges and there are 268 whites graduates. $36/268 \approx 13.4\%$
- e) There are 42 graduates planning to attend 2-year colleges, and 36 of them are white. $36/42 \approx 85.7\%$

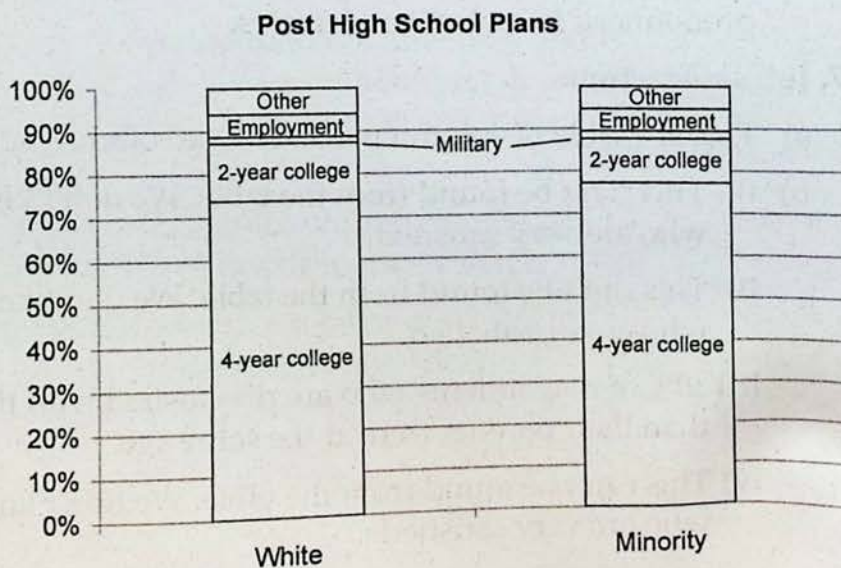
Plans	White	Minority	TOTAL
4-year college	198	44	242
2-year college	36	6	42
Military	4	1	5
Employment	14	3	17
Other	16	3	19
TOTAL	268	57	325

31. Movies 06-15.

- a) This is a table of column percents. The columns add up to 100%, while the rows do not.
- b) Movies rated G and PG have become slightly less common, while movies rated PG-13 and R have become slightly more common.
- c) For Dramas, the percentages of PG and R have decreased while the percentage of PG-13 dramas has increased significantly. For Comedies, there has been a large increase in the percentage of R-rated films.

33. More about seniors.

- a) For white students, 73.9% plan to attend a 4-year college, 13.4% plan to attend a 2-year college, 1.5% plan on the military, 5.2% plan to be employed, and 6.0% have other plans.
- b) For minority students, 77.2% plan to attend a 4-year college, 10.5% plan to attend a 2-year college, 1.8% plan on the military, 5.3% plan to be employed, and 5.3% have other plans.
- c) A segmented bar chart is a good display of these data.



- d) The conditional distributions of plans for Whites and Minorities are similar:
 White – 74% 4-year college, 13% 2-year college, 2% military, 5% employment, 6% other.
 Minority – 77% 4-year college, 11% 2-year college, 2% military, 5% employment, 5% other.
 Caution should be used with the percentages for Minority graduates, because the total is so small. Each graduate is almost 2%. Still, the conditional distributions of plans are essentially the same for the two groups. There is little evidence of an association between race and plans for after graduation.

35. Magnet schools revisited.

- a) There were 1755 qualified applicants to the Houston Independent School District’s magnet schools program. Of those, 292, or about 16.6% were Asian.
 b) There were 931 students accepted to the magnet schools program. Of those, 110, or about 11.8% were Asian.
 c) There were 292 Asian applicants. Of those, 110, or about 37.7%, were accepted.
 d) There were 1755 total applicants. Of those, 931, or about 53%, were accepted.

37. Back to school.

There were 1,755 qualified applicants for admission to the magnet schools program. 53% were accepted, 17% were wait-listed, and the other 30% were turned away. While the overall acceptance rate was 53%, 93.8% of Blacks and Hispanics were accepted, compared to only 37.7% of Asians, and 35.5% of whites. Overall, 29.5% of applicants were Black or Hispanics, but only 6% of those turned away were Black or Hispanic. Asians accounted for 16.6% of applicants, but 25.3% of those turned away. It appears that the admissions decisions were not independent of the applicant’s ethnicity.

39. Weather forecasts.

- a) The table shows the marginal totals.
 It rained on 34 of 365 days, or 9.3% of the days.
 b) Rain was predicted on 90 of 365 days.
 $90/365 \approx 24.7\%$ of the days.
 c) The forecast of Rain was correct on 27 of the days it actually rained and the forecast of No Rain was correct on 268 of the days it didn’t rain. So, the forecast was correct a total of 295 times. $295/365 \approx 80.8\%$ of the days.

		Actual Weather		Total
		Rain	No Rain	
Forecast	Rain	27	63	90
	No Rain	7	268	275
	Total	34	331	365

ISBN-13: 978-0-13-426535-3
ISBN-10: 0-13-426535-1

EAN



9 780134 265353 90000