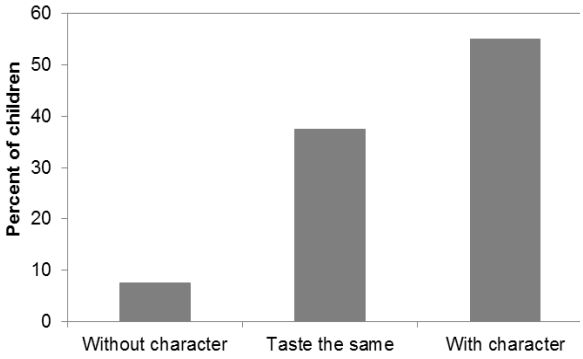
Chapter 1 Solutions

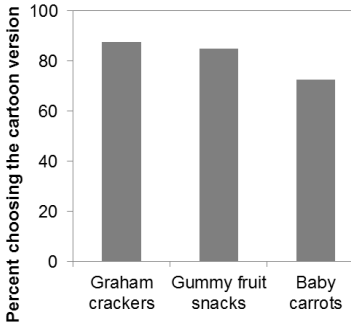
1.1. **(a)** The individuals in this exercise are the different brands of breakfast cereals. **(b)** The variables are manufacturer (categorical), preparation method (categorical), calories (quantitative), sugar content (quantitative), and fiber content (quantitative).

1.2. **(a)** The individuals in this exercise are the patients with first ever traumatic brain injury and similar healthy people with no prior brain trauma. **(b)** The variables are hours of sleep (quantitative) and signs of excessive daytime sleepiness (categorical).

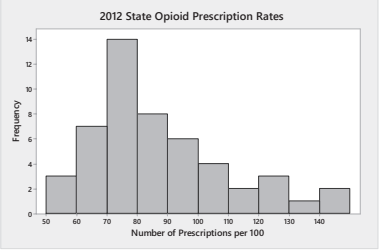
1.3. **(a)** The individuals in this exercise are the young children ages four to six. The variable is Taste preference. **(b)** Bar graph shown on right. **(c)** Yes, a pie chart would be appropriate because the three taste preferences make up one whole. **(d)** The data show that the majority of the children in this study preferred the crackers with a popular cartoon character, whereas less than 10% preferred the crackers without the character. This suggests that popular cartoon characters influence food choices in young children.



1.4. **(a)** The individuals in this exercise are the young children ages four to six. The variable is Food Item. **(b)** Bar graph shown on right.**(c)** A single pie chart would not be appropriate because the data for the three food types do not make up one whole. Instead, the percent who choose the cartoon version for any given food type is complementary to the percent who do not choose the cartoon version for that given food type. A separate pie chart would need to be made for each food type. **(d)** The data show the majority of children choose food types that display a cartoon character on the package for all food types.



1.5. The histogram to the right was created in MiniTab; some programs automatically include the lower end of each class while excluding the upper end, while others automatically include the upper end of each class while excluding the lower end. Either way is fine; compare your histogram to this one.



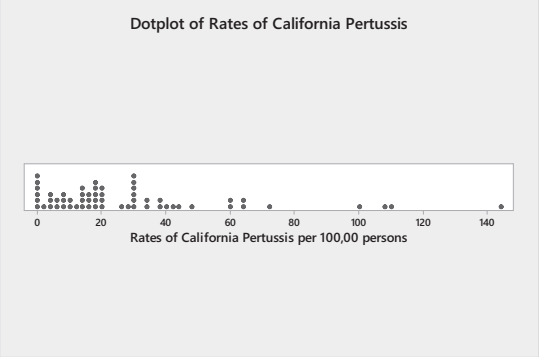
1.6. **(a)** The applet creates a histogram with 7 classes. **(b)** It is possible to get to one class ranging from 11 to 41 (not a very useful histogram). **(c)** The most classes the applet will allow is 15; the classes with the highest count have only 2 observations.

(d) Choices will vary; anything from about 6 to 10 classes is reasonable. A sample size of only 18 can be a little awkward to fit in a histogram because even small changes in class choice can have a substantial impact on the overall look of the histogram.

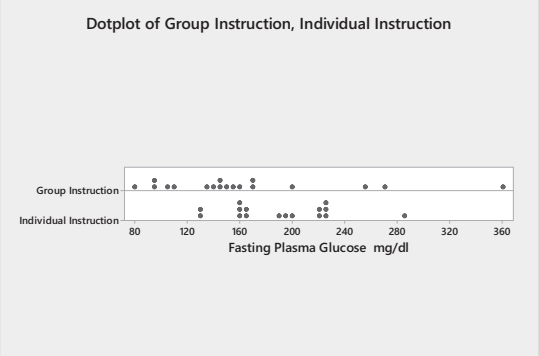
1.7. The distribution of the opioid prescription rates is single peaked and skewed to the right.

1.8. The histogram is clearly unimodal and symmetric, without outliers. The midpoint lies in the 14-to-15-years class, which represents 14-year-old girls.

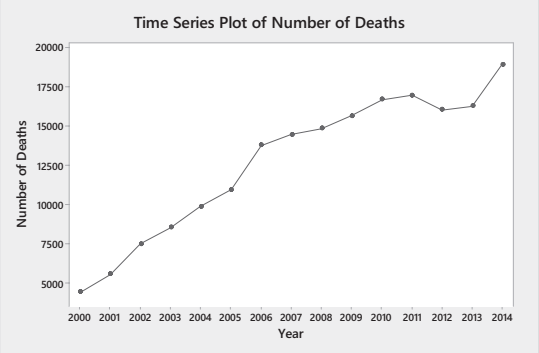
1.9. The dotplot shown to the right reveals the same right skewed distribution that the histogram shows. The dotplot better shows repeated values of the rates of pertussis per 100,000 persons than the histogram For example, there are six counties that had zero cases reported. The histogram shows that there is a count of 32 counties that have fewer than 20 cases of pertussis reported per 100,000 persons but does not show what those values are.



1.10. **(a)** Stacked dotplots shown to the right. **(b)** Glucose levels in the individual instruction group are less variable and appear symmetric with no outliers. Glucose levels in the class instruction group are slightly right skewed with an apparent outlier. **(c)** Overall, both groups have glucose levels that are not adequately controlled. Only four of the 18 individuals in the class instruction group had glucose levels in the desired range. In the individual instruction group, only two of the 16 diabetic patients maintained their fasting plasma glucose in the desired range.



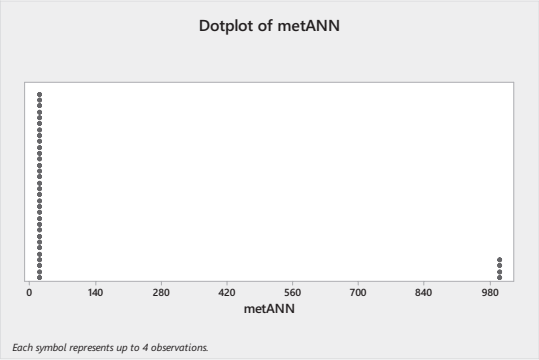
1.11. The time plot shows a clear increase in the number of deaths from prescription opioid pain relievers in the United States from the years 2000 to 2014.



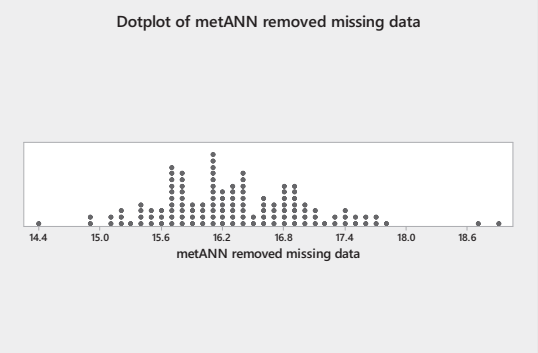
1.12. The data have both an obvious overall trend and clear cyclical variations. Monthly CO_2 levels vary seasonally, peaking in the spring and bottoming in the fall, creating annual cycles. But the big picture shows a clear upward trend reflecting increasing monthly CO_2 levels over the full 40-year period for which data exist. This suggests that we should think carefully about our global carbon emissions and possible action courses.

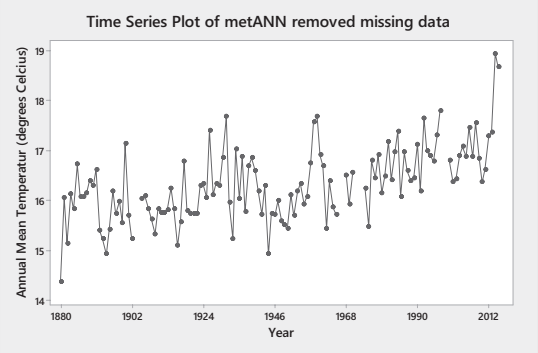
1.13. **(a)** Height is quantitative. You could collect the data in a spreadsheet with each row representing a student and a column representing height. The data could be displayed in a histogram or a dotplot. **(b)** Gender is categorical. In the spreadsheet there could be a column named "Female" and record a 0 for male and a 1 for female. The data could be displayed in a bar graph or a pie chart. **(c)** The data could be sorted by gender, and separate graphs could easily be constructed if the data were collected as described in parts (a) and (b). Organizing data for use with statistical software addressed this issue.

1.14. **(a)** The dotplot to the right shows the data is distinctly bimodal. It is suspicious and may be explained by a code for missing data. **(b)** The most likely explanation for the outliers in the dotplot is missing data. It seems that a 999.9 was used as a special code for years that did not have an annual mean temperature recorded. Organizing data for use with statistical software addressed this issue.



(c) The cells containing 999.9 were cleared in MiniTab to create the dotplot and time plot below.





(d) The time plot reveals regular up and down cycles with an overall increase in annual mean temperatures in Los Angeles from 1880 to 2017. The dotplot reveals that the distribution of annual mean temperatures is fairly symmetric with a two high values far removed from most of the data and one low value removed from most of the data.

1.15. (a) Pregnancy Status

1.16. (c) one categorical and one quantitative.

1.17. (c) The data can be displayed either on a pie chart or a bar graph because the categories represent the pieces of a whole.

1.18. (a) Kidney transplants represented nearly 61% (16,898/27,793) of all single-organ transplants in 2010.

1.19. (b) a bar graph that cannot be made into one pie chart.

1.20. (b) The majority of adults who live in high-income countries are overweight and obese.

1.21. (c) 31%

1.22. (a) The distribution of flea jumps is moderately right-skewed.

1.23. (b) single-peaked with two outliers.

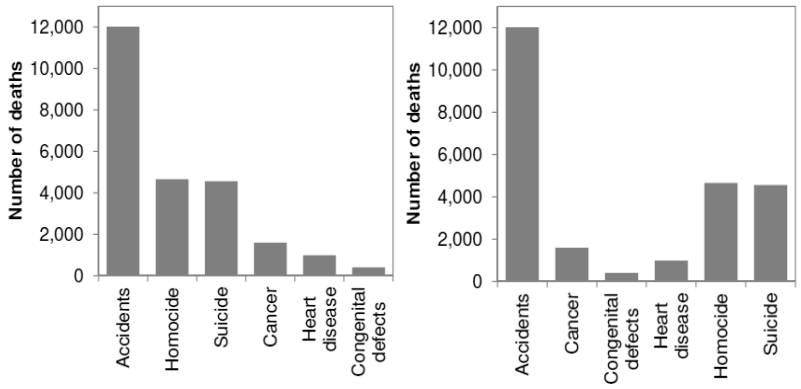
1.24. (b) roughly symmetric.

1.25. **(a)** Number of eggs is quantitative, a count. **(b)** Incubation period is quantitative, a length of time. **(c)** Parental care is categorical, 1 of only 3 possible choices. **(d)** Nest size is quantitative, a measure of size. **(e)** Presence of pesticides is categorical, 1 of only 2 possible choices.

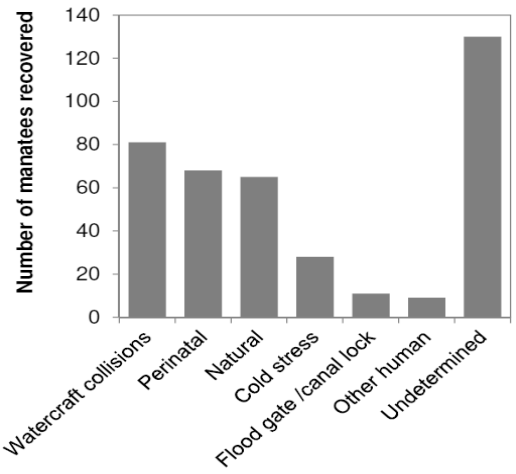
1.26. Answers will vary. Examples of categorical variables are gender (boy/girl), lunch type (school-provided/home-prepared), and milk consumption (yes/no). Examples of quantitative variables are age (in years), approximate number of calories consumed on the day before the interview (in Calories), and typical number of fruit and vegetable servings per day (a unitless count).

1.27. **(a)** The 53 lakes are the individuals. **(b)** There are 5 variables recorded, 4 quantitative and 1 categorical (age of data). Age of data is categorical because the only possible entries are recent data and year-old data (non-numerical).

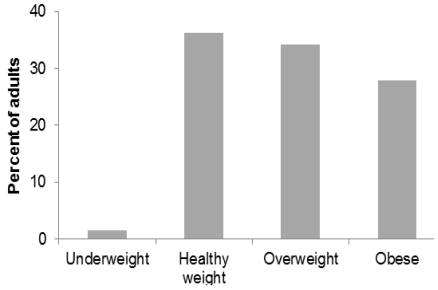
1.28. **(a)** Shown on the right. **(b)** In order to make a pie chart, we would need to know the total number of deaths in this age group (so that we could compute the number of deaths due to other causes).



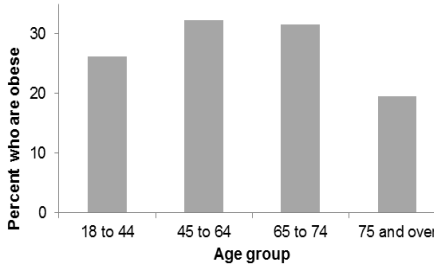
1.29. **(a)** There were 130 "undetermined" manatee deaths in Florida in 2012 (392 − 262) **(b)** "Watercraft collisions" made up 20.7% (81/392) of all 2012 manatee deaths in Florida. **(c)** Shown on the right. Of the deaths for which a cause could be identified, watercraft collisions was the most common, closely followed by perinatal and natural causes. **(d)** Yes, we could display these data in a pie chart because, together, they make up all the categories of manatee deaths.



1.30. **(a)** Shown on the right. There is a clear overweight crisis in the adult American population. **(b)** Yes, we could display these data in a pie chart because, together, the percents in all 4 weight categories make up the whole adult population. The pie chart would emphasize that more than half of the adult American population is either overweight or obese.



1.31. **(a)** Shown on the right. All age groups have a large rate of obesity. Adults 45 to 64 and 65 to 74 have the highest percentage of obese individuals. **(b)** No, we couldn't display these data in one pie chart because the percents in the 4 age groups do not make up the whole adult population. Each percent represents a different age group separately.



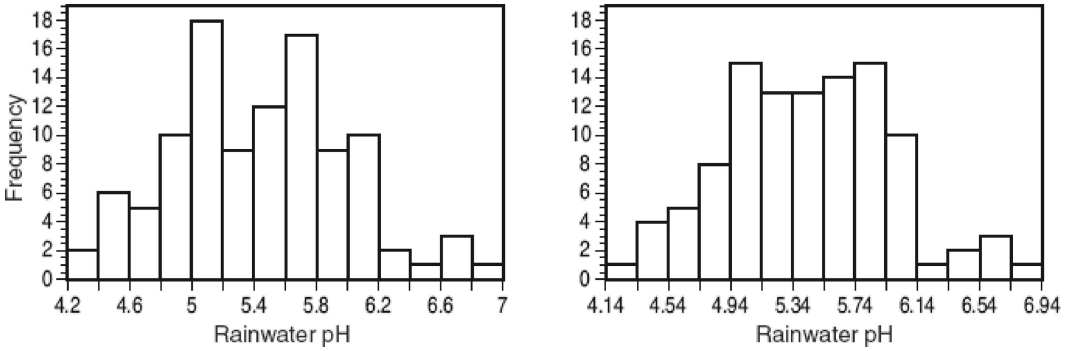
1.32. **(a)** The distribution is bimodal, with a spread of 0 to 75 percent of correct quiz answers. The first peak is located around 10 to 15 percent correct answers, and the second peak is located around 45 to 50 percent correct answers. **(b)** This suggests that there are two types of individuals claiming to have HSAM: individuals who perform poorly and are unlikely to truly have HSAM and individuals who perform much better and are more likely to truly have HSAM.

1.33. A bar graph would be appropriate for (a) and (d) because the data are categorical. A histogram would be appropriate for (b) and (c) because the data are quantitative.

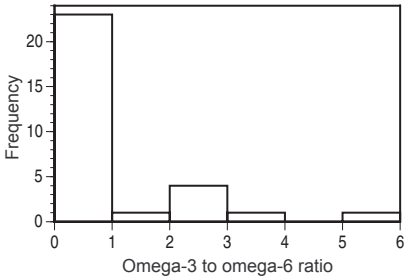
1.34. **(a)** The distribution is unimodal, moderately skewed to the right, without outliers. The spread of the distribution is 0 to 30 percent correct quiz answers. The center of the distribution is located somewhere between 5 and 15 percent correct answers. **(b)** The histogram for the control group is similar to the first peak of the bimodal

histogram for the group claiming to have HSAM. This confirms our initial suspicion that individuals who perform poorly (similar to the control group) are indeed unlikely to truly have HSAM, unlike the high-scoring individuals making up the second peak of the top histogram.

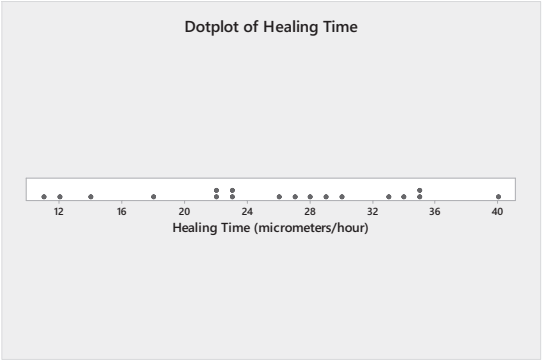
1.35. **(a)** The first histogram (on the left) shows possibly two modes: 5–5.2 and 5.6–5.8. **(b)** The second histogram (on the right) has peaks in locations close to those of the first, but these peaks are much less pronounced and more like a single but very wide peak.



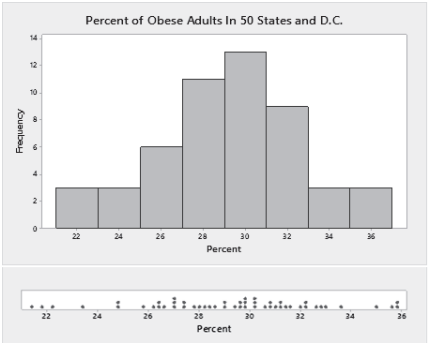
1.36. **(a)** The histogram is shown on the right. **(b)** The distribution is sharply skewed to the right. Only 7 of the 30 food oils have more omega-3 than omega-6 (ratio greater than 1), so most oils do not have the desired ratio. **(c)** The fish oils are 5 of the 7 oils that have a ratio over 1.



1.37. **(a)** Healing rates are fairly evenly spread out, with a slightly higher density near the center. **(b)** The stemplot reveals a symmetric distribution with no outliers. **(c)** Both the stemplot and the dotplot reveal the distribution is symmetric with no outliers and both display the raw data. The dotplot shows better how the data is spread out and the stemplot shows better the symmetric characteristic of the data.



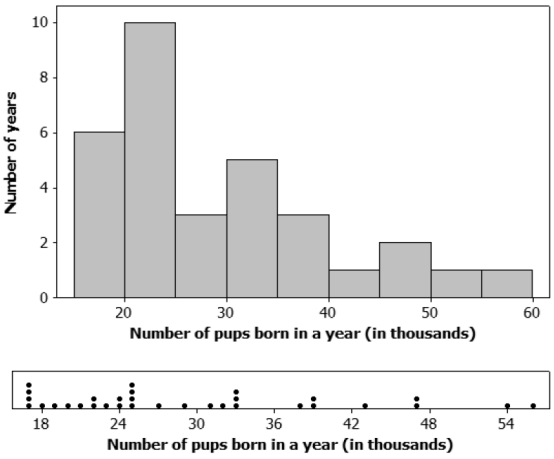
1.38. **(a)** The individuals are the set of people that self-reported their weight and height in each state and the District of Columbia. The variable is the percent of the population that is obese. The variable is quantitative. **(b)** The histogram and dotplot are shown to the right. The distribution of the percent of adults in each state and D.C. is approximately symmetric with no clear outliers. The center is around 29%, and the spread is from 21% to 37%.**(c)** Most of the Midwest, West, and Southeast have obesity rates between 20% and 35%. There are three states (Arkansas, Mississippi, and West Virginia) that have obesity rates greater than 35%. This geographic characteristic of the rates of obesity could not be identified from the data table, dotplot, or the histogram.



1.39. **(a)** Shown below. **(b)** The distribution of flower lengths is unimodal and skewed to the right.

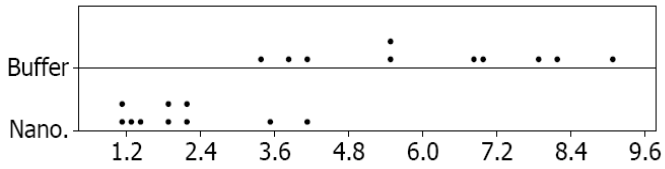
Stem	Leaf
46	3467889
47	114
48	1234
49	
50	13

1.40. The histogram and dotplot are shown on the right. The distribution is skewed to the right but does not have clear outliers. The center is around 25 pups per year, and the spread is from 17 to 56 pups per year.

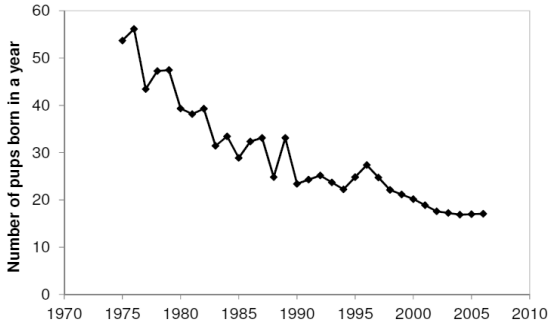


1.41. **(a)** Dotplots shown below. Tumor increase tends to be smaller with nanoparticle treatment. The distribution of tumor increases is skewed with nanoparticles, whereas it is somewhat symmetric and uniform with the buffer treatment. **(b)** The midpoints are about 6 with buffer and about 2 with nanoparticles. The most important difference revealed by dotplots is that nanoparticle treatment tumor increase tends to be much

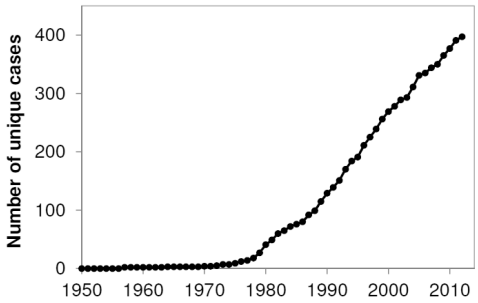
lower than buffer treatment with very little overlap. The study findings show there is a difference in treatment groups.



1.42. The time plot shows that the number of seal pups born each year has declined fairly steadily in that period, although more slowly so in recent years. This trend cannot be seen in a dotplot or a histogram.



1.43. **(a)** Time plot shown on the right. There is a clear increasing pattern since the mid-1970s, but no cyclical variations. The number of unique cases of herbicide resistance in weeds worldwide has increased drastically over time, from nearly nothing before the mid-1970s to nearly 400 in 2012. **(b)** The clear pattern over time would be missed in a histogram. It is always wise to plot data collected over time in a timeplot first.



1.44. **(a)** The percent of respondants that exercised regularly is the highest in summer and the lowest in winter.**(b)** The long term trend is a gradual increase in the percent of respondents who said they exercised regularly from 2008 to 2016.

LARGE DATA SET EXERCISES:
ANSWERS GROUPED IN "LARGE DATA SET" CHAPTER

Exercise 1.45 **Everglades.**