

Unit 1

Analyzing Data

Math Medic Unit 1: Analyzing Data

Lesson	Learning Targets
1.1 Statistical Studies	- Identify the population and sample in a statistical study. - Identify and classify the variable(s) in a study.
1.2 Analyzing Categorical Data	- Use bar charts and pie charts to analyze data for one categorical variable. - Use two-way tables and various bar graphs to analyze data for two categorical variables. - Be aware of misleading graphs.
1.3 Representing Categorical Data	- Compare distributions of categorical data using bar graphs and mosaic plots. - Determine if two categorical variables are associated.
1.4 Describing Quantitative Data	- Use dotplots, stem-and-leaf plots, and histograms to analyze quantitative data. - Describe the distribution of a quantitative variable (shape, outliers, center, variability).
1.5 Measuring Variability	- Calculate and interpret standard deviation. - Identify which summary statistics are resistant and nonresistant to outliers. - Compare the mean and median to determine the approximate shape of a distribution.
1.6 Comparing Quantitative Data	- Identify outliers in a quantitative data set. - Use a boxplot to analyze quantitative data. - Compare distributions of quantitative data.
1.7 Location in a Distribution	- Calculate and interpret percentiles for quantitative data. - Calculate and interpret z-scores for quantitative data. - Use z-scores to compare the relative location of values in different distributions.
1.8 Linear Transformations of Quantitative Data	- Determine what happens to the shape, center, and variability when adding/subtracting a constant or multiplying/dividing by a constant for quantitative data. - Determine the mean and standard deviation for a standardized (z-score) distribution.



Can Joy Detect Parkinson's Disease?



Joy Milne participated in a study where she was given 12 t-shirts, half of which were worn by Parkinson's patients, and half of which were worn by a control group. Joy correctly identified 11 out of the 12 shirts. Does this provide *convincing* evidence that Joy can smell Parkinson's?

1. Why would it be important to know that someone can smell Parkinson's disease?
2. How many correct decisions would you expect Joy to get out of 12 if she really couldn't smell Parkinson's (she was just guessing)? Explain.
3. Do we have some evidence that Joy can smell Parkinson's? Why?
4. How many correct decisions out of 12 would it take to *convince* you that Joy really could smell Parkinson's?



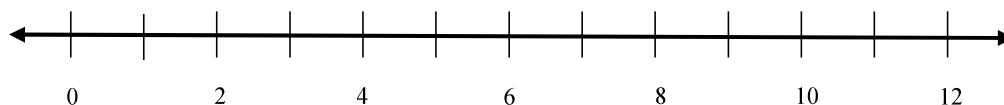
Let's investigate whether Joy's result could have happened purely by chance, just by guessing. Working in pairs, you will simulate this study. One person will be the experimenter and one person will be Joy and then you will switch.

Important: the experimenter should not reveal the truth for each shirt. They should simply record whether the guess was correct or incorrect.

4. As the experimenter, keep track of the results:

Guess	Tally	Frequency
Correct		
Incorrect		

5. Count up the number of correct decisions. Write the number on a sticker dot and bring it to the poster at the front of the room. Copy the dotplot here.



6. What does each dot represent?

7. Based on the class data, what percent of the simulations resulted in 11 or more correct identifications?

8. Based on these results, do we have convincing evidence that Joy can smell Parkinson's? Explain.

Babies With Pets May Be Less Likely to Develop Food Allergies

by Jamie Ducharme

[About 8% of children in the U.S. have a food allergy](#) — roughly double the percentage who [did in 2007](#). It's not entirely clear why more kids are being diagnosed with food allergies, both in the U.S. and [across the world](#) — but as these potentially life-threatening conditions grow more common, lots of researchers are studying how to [treat](#) or [prevent them](#).

[A study](#) published in *PLOS ONE* on March 29 offers one potential strategy: get a pet. Among thousands of babies in Japan, exposure to cats or dogs during pregnancy or early infancy was associated with a lower risk of developing food allergies, the researchers found.

That finding is based on data from the Japan Environment and Children's Study, an ongoing trial that includes about 100,000 women in Japan who became pregnant from January 2011 to March 2014 and their children. Kids in the study are tracked until they turn 13 so researchers can learn more about how their living conditions and environmental exposures affect their health.

The new study looked at about 66,000 children whose mothers provided information about their pet exposure and food allergies. Compared to children without pets, kids who were exposed to a cat or dog during pregnancy or early infancy were roughly 15% less likely to develop allergies to common triggers including eggs, milk, wheat, soybeans, and nuts by age three, according to their mothers' survey responses.

Other pets — including hamsters, turtles, and birds — didn't seem to provide the same benefits. Dogs kept indoors were also associated with a larger risk reduction than dogs kept outdoors.

Scientists don't fully understand the link between animal exposure and food allergies, the researchers write. One theory is that living with pets alters babies' [microbiomes](#) in a way that protects them from [food sensitivities](#), but more research is needed to say for sure.

The new study should be taken with a grain of salt, since it relies on parents' survey answers rather than objective allergy diagnostics. This sort of data also can't prove cause and effect — so, while the researchers considered factors like parental health, the family's socioeconomic status, and where they lived, it's possible there's another explanation for the link between pets and lower rates of food allergy.

[Not all previous studies](#) linking pets and food allergies have come to the same conclusion, the authors note. However, other researchers have found that kids who live with pets during infancy are less likely to develop environmental allergies, asthma, and other [respiratory illnesses](#). Pets are also associated with better mental and physical health among adults — so adopting a cat or dog may be beneficial for an entire household's health.

<https://time.com/6266337/pets-food-allergies-children/>



Study Says...



Statistical studies are everywhere. It is common to see headlines summarizing the findings from a study with an attention-grabbing one-liner. But what does the study actually say?

Read [this article](#) and answer the following questions:

1. What question were the researchers trying to answer?

2. How was the data collected for this study? Did they follow the participants over time or use past records? How do you know?

3. What groups are being compared in the study?

4. What information did the researchers record about the participants to answer their question?

5. What conclusion did the researchers make from the study?

6. Who was the conclusion about?

7. What potential problems are present in the study?

Lesson 1.1 – Statistical Studies

QuickNotes

Check Your Understanding

Researchers examined data from 112,916 Canadian public school students to see whether involvement in school music programs was associated with better scores on high school exams in English, math, and science. They found that students who participated in music activities (especially playing an instrument over multiple years) tended to achieve higher standardized exam scores than peers who did not participate in music.

1. Identify the population and sample.
2. Is this a prospective study or a retrospective study? How do you know?
3. Identify and classify the variable(s) in the study.
4. Describe at least one limitation of the study.



How Are Your Favorite Classes Related?



Is your favorite elective class associated with your favorite core class? We asked a group of students "What is your favorite elective class?" but forgot to record the number that said physical education.

Art	Music	Physical Education	World Language	Technology
6	9		4	4

1. Identify the individuals.

2. Clearly identify the variable being measured.

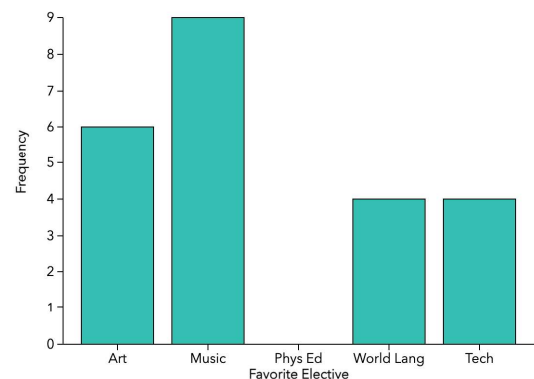
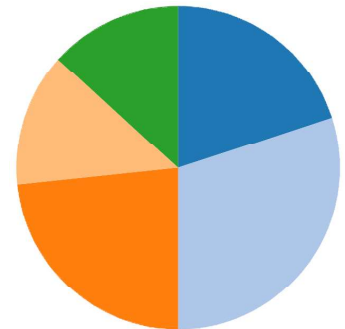
3. Is the variable categorical or quantitative?

4. If music was chosen by 30% of students, how many total students were sampled?

5. How many students chose physical education? Fill in the value in the table and complete the bar graph.

6. Label each section of the pie chart with the corresponding elective.

7. Here is the data broken down by favorite core class.



		Core Class		Total
		Math	English	
Elective class	Art	2	4	
	Music	5	4	
	P.E.	4	3	
	World Lang.	1	3	
	Technology	4	0	
Total				

Find the following:

% of all students who chose P.E.:

% of all students who chose Math and chose Art:

% of the students who chose Technology, given they chose math:



8. How many variables does the table have?
Are the variables categorical or quantitative?

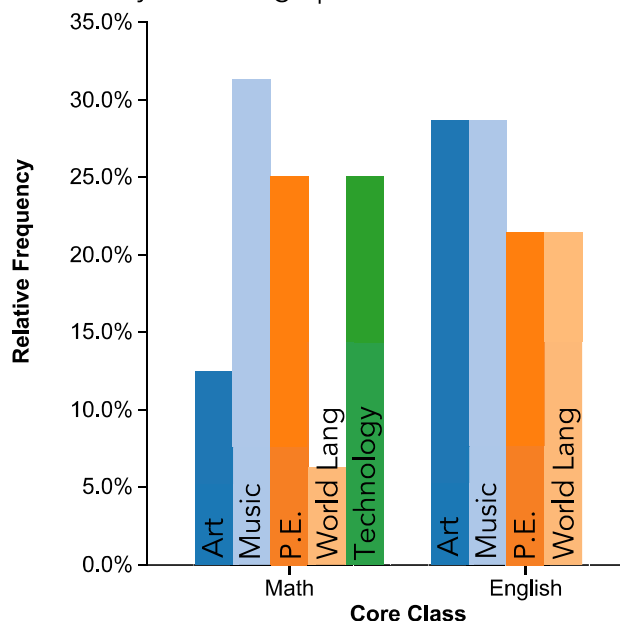
9. Which variable would best explain or predict the other variable?

**Elective
class**

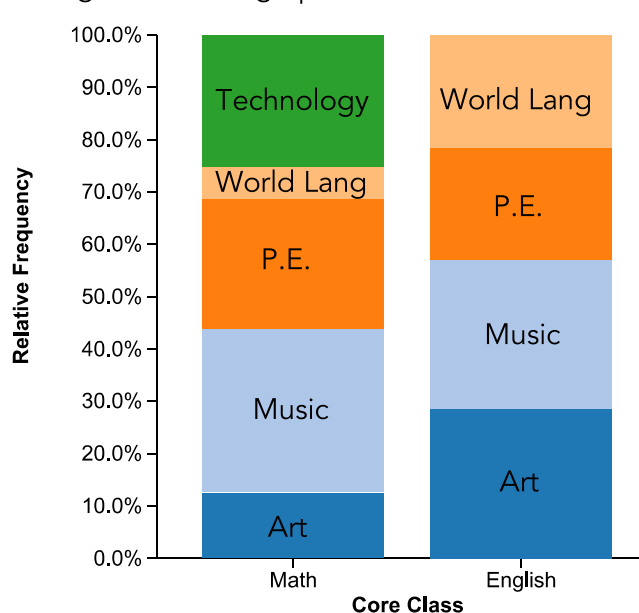
Core Class

	Math	English	Total
Art	2	4	6
Music	5	4	9
P.E.	4	3	7
World Lang.	1	3	4
Technology	4	0	4
Total	16	14	30

Side-by-side bar graph:



Segmented bar graph:



10. How do the bars in the side-by-side-bar graph relate to the bars in the segmented bar graph?

11. Is there an association between favorite core subject and favorite elective? If so, describe it.

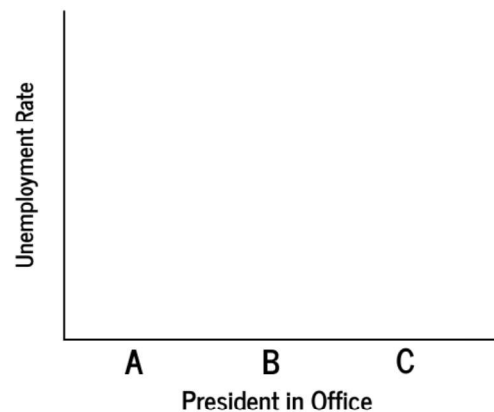
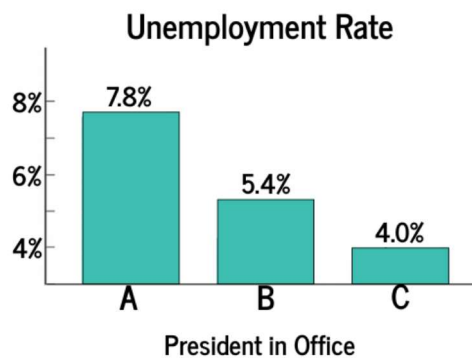
12. If there was not an association between favorite core subject and favorite elective, what would the graphs look like? Explain.

Lesson 1.2 – Analyzing Categorical Data

QuickNotes

Check Your Understanding:

1. Billy was working on a school project about unemployment rates at the end of three different presidencies for President A, President B, and President C. The following graph was displayed in his final report. Explain why the graph may be misleading and sketch a corrected version of the graph.



2. A real estate agent is collecting data on the number of houses built in his town's three neighborhoods during three different decades. The table below gives information.

	1960s	1970s	1980s
Shady Lane	40	30	10
Oakcrest	60	15	5
Pinewood Estates	0	45	15

- a. What proportion of the houses shown were built in Pinewood Estates?
- b. Find the distribution of Decade Built for the houses in this town using relative frequencies.
- c. What percent of homes were built in Oakcrest and in the 1960s?

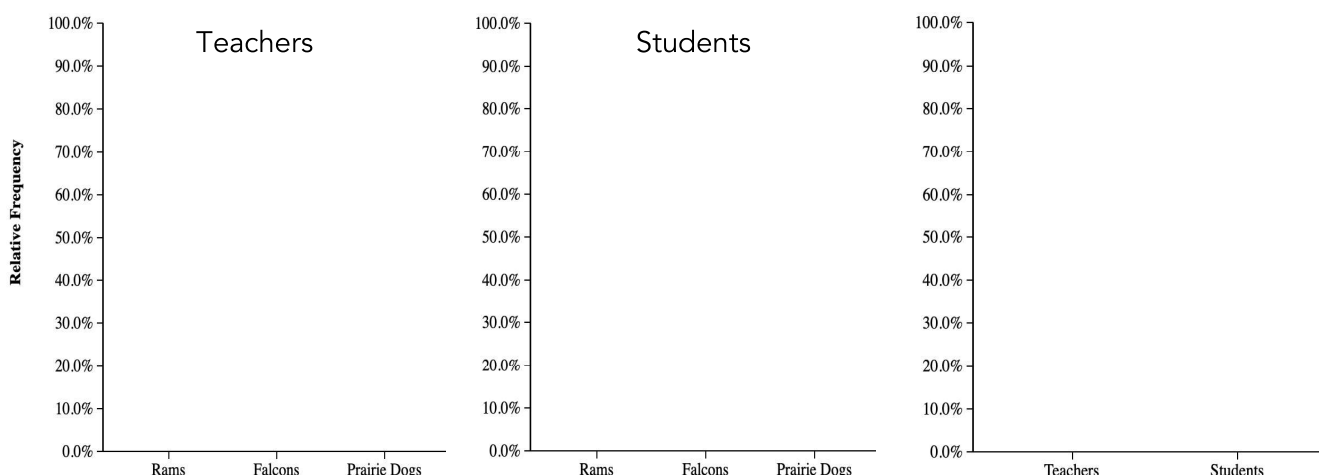


What Will Be the Mascot?

When the high school was built in 1969, the school needed to pick a mascot. The principal decided to have the students and teachers vote between three choices: rams, falcons, or prairie dogs. He took a random sample of students and a random sample of teachers. The results of the surveys are given in the table.

	Rams	Falcons	Prairie Dogs
Teachers	80%	10%	10%
Students	30%	60%	10%

1. Create two bar graphs below to display the results. Use three different colors for the bars.
2. Complete the third graph by taking each bar from the teacher sample and stacking them. Use the colors to mark each section. Do the same for the student sample.



3. According to your displays, which mascot appears to have the most support? Explain.
4. Upon hearing the results of the surveys, the students argued that the decision was incorrect because 100 teachers had been surveyed and 500 students had been surveyed. Use this information to fill in the table below with the number of responses.

	Rams	Falcons	Prairie Dogs
Teachers			
Students			

5. How many times more students were sampled than teachers? _____. How can you update the third graph above to take into account the sample size? Adjust your graph.
6. What should they make the EK mascot? Explain.

Lesson 1.3 – Representing Categorical Data

QuickNotes

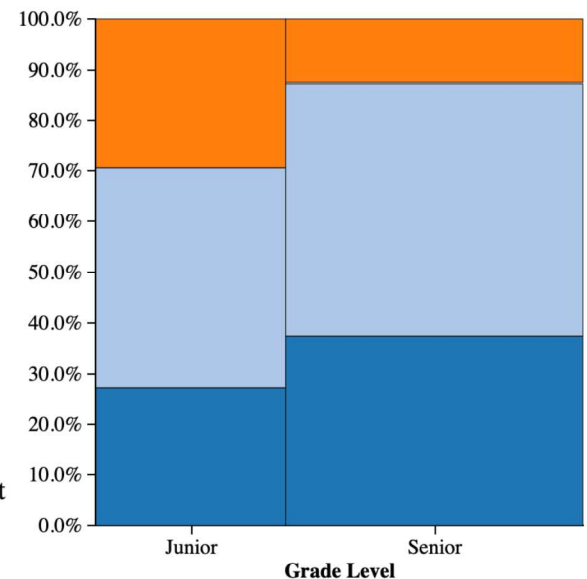
Check Your Understanding:

The following table gives the result of a random sample of upper-level students at Rocky Vista University (the Fighting Prairie Dogs!), along with a mosaic plot.

Employment Status	Grade Level	
	Junior	Senior
Currently working	14	30
Not working but have had a job	22	40
Never had a job	15	10

Employment status

- Never had a job
- Not working but did in the past
- Currently working



- a. Calculate the proportion of Juniors that are currently working, not working but have had a job, and never had a job.
- b. Calculate the proportion of Seniors that are currently working, not working but have had a job, and never had a job.
- c. Write a few sentences summarizing what the display in part (a) reveals about the association between grade level and job experience for the students in the sample.

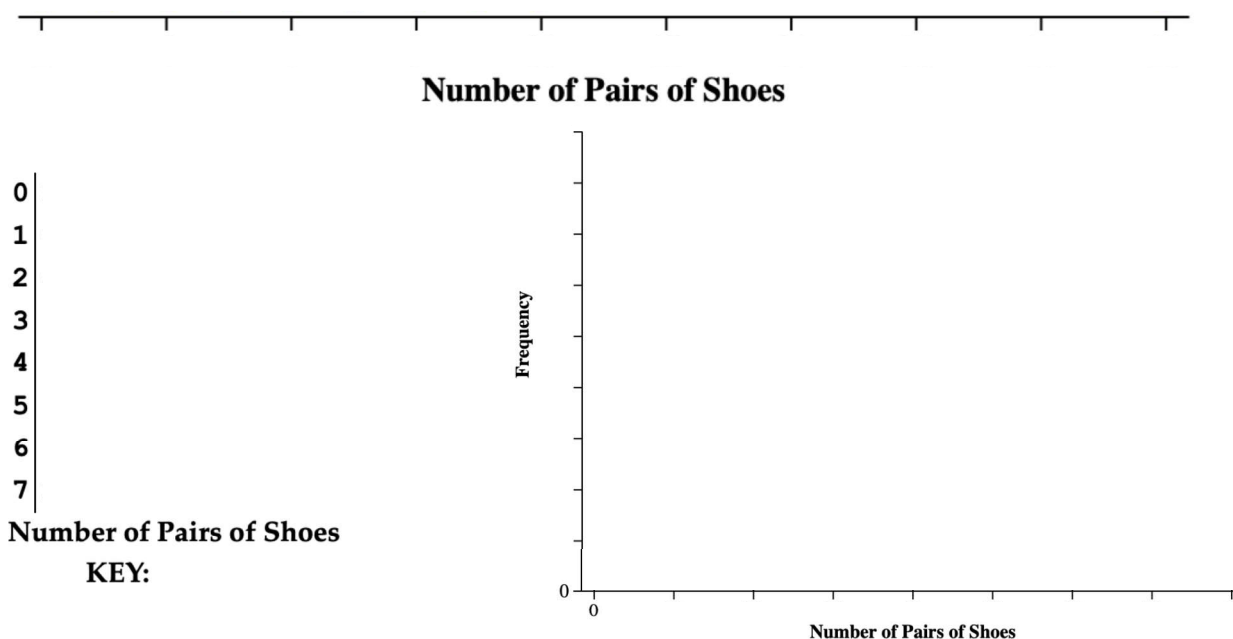


How Many Pairs of Shoes Do You Own?



How many pairs of shoes do you own? You can only count it as a pair if you have both shoes. If you are not sure how many pairs you own, make your best guess. Record your answer on the board.

1. Is "number of pairs of shoes" a categorical or quantitative variable?
2. Can this variable have decimal values?
3. Enter the data at stapplet.com. Sketch the dotplot, stem-and-leaf plot, and histogram.



4. How many pairs of shoes does the typical student in your class own?
5. Describe the distribution of the number of pairs of shoes for your class.

Shape:

Outliers:

Center:

Variability (spread):

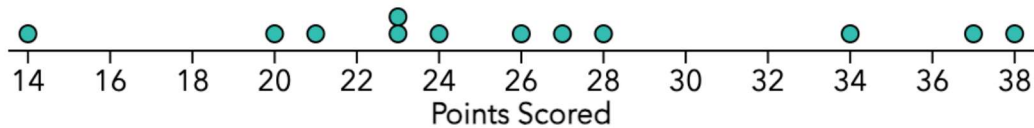
6. Which of the three types of display do you prefer? Why?

Lesson 1.4 – Describing Quantitative Data

QuickNotes

Check Your Understanding:

1. Mr. Wilcox is a huge fan of University of Michigan football. His favorite season was the 1997 season (a perfect season!). The dotplot shows the number of points scored by the U of M team in the 12 games that season.



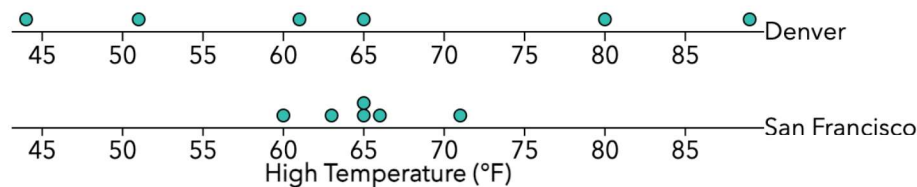
- (a) Use the dotplot to create a stem and leaf plot of the distribution.
- (b) Describe the shape of the distribution.
- (c) Without doing any calculations, do you think the mean would be greater than, less than, or roughly the same as the median? Explain.
- (d) Are there any potential outliers? Why?
- (e) What measure of center is most appropriate to describe the distribution? Explain.



Which City Has the Better Weather?



Both Denver and San Francisco are desirable locations to live, but which city would you say has the better weather? Below is a dotplot of the high temperatures from a random sample of 6 days from each city.



1. What is the same about the high temperatures in each city? What is different?
2. One way to measure variability for a set of data is with **range**. Calculate the range for each city.
3. Another way to measure variability is to find the **average distance of the values from the mean**.

- a. Complete the table for the Denver data.
The mean is 65 degrees Fahrenheit.

Value	Distance from mean	(Distance from mean) ²
44		
51		
61		
65		
80		
89		
Total:		
Average (Distance from mean) ² :		

- b. The average you calculated in the table is the average of the **squared distances** from the mean. How do we use this to find the **average distance from the mean**? Find it here.

4. Go to stapplet.com. Enter the data and find the summary statistics.

n	mean	SD	min	Q ₁	med	Q ₃	max

5. Let's add one more day to the data set (a very cold day with a high temperature of only 7° F). Add this value to the data set and record the new summary statistics. How did these change?

n	mean	SD	min	Q ₁	med	Q ₃	max

6. How would the average distance from the mean for San Francisco compare to Denver? Why?

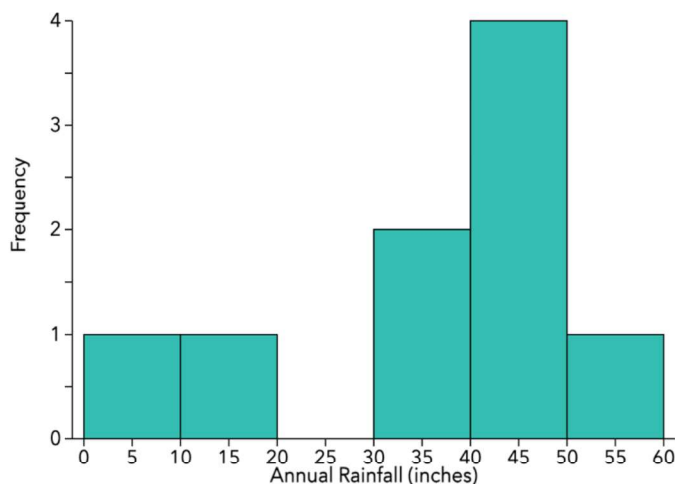
Lesson 1.5 – Measuring Variability

QuickNotes

Check Your Understanding

A researcher is interested in how much annual rainfall is typical in the United States. She selects a random sample of 9 cities in the U.S. and records the annual rainfall, in inches.

8.2
10.3
33.5
39.1
40.5
41.9
42.4
44.9
53.7



1. Calculate the mean annual rainfall for these cities.
2. Find the median annual rainfall for these cities.
3. Would you use the mean or the median to summarize the typical annual rainfall for a U.S. city? Explain.
4. Compare the mean and median to confirm the shape.
5. The standard deviation of the annual rainfall for these 9 cities is 15.52 inches. Interpret this value.

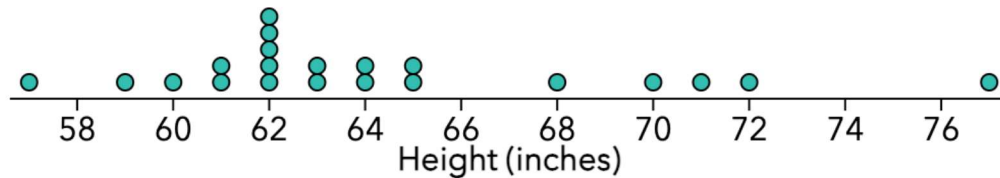


Where Do I Stand?



How does my height compare with the other AP Stats students in my class? To answer this question, Ashmita recorded the heights of everyone in her class. The heights (in inches) were:

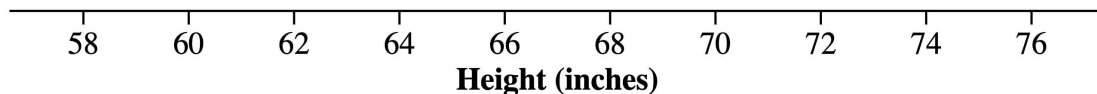
68 72 61 62 63 63 64 64 59 62 61 60 65 62 57 77 62 71 65 62 70



1. What is the median height? Describe how you found it.
2. Q_1 is the median of the lower half of the data. Q_3 is the median of the upper half of the data. Find Q_1 and Q_3 ?

3. Record the following values and then use them to make a boxplot.

Minimum: Q_1 : Median: Q_3 : Maximum:



4. The **interquartile range** (or IQR) is defined as $Q_3 - Q_1$. Find the IQR . Where do you see the IQR in the boxplot?

$IQR =$

5. An **outlier** is a data value that is way too small or way too big (using Rule 1 or Rule 2 below). Are there any outliers? Check using each rule separately and show your work.

Rule 1: way too small $< Q_1 - 1.5IQR$ way too big $> Q_3 + 1.5IQR$

OR

Rule 2: way too small $< \bar{x} - 2s_x$ way too big $> \bar{x} + 2s_x$

The mean is 64.3 inches and the standard deviation is 4.8 inches.

Lesson 1.6 – Comparing Quantitative Data

QuickNotes

Check Your Understanding:

Mr. Wilcox is a huge fan of University of Michigan football. His favorite season was the 1997 season (a perfect season!). Here parallel boxplots of the points scored by the 1997 University of Michigan football team and the archrival Michigan State University football team. Write a few sentences comparing the distributions.

