

## 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.



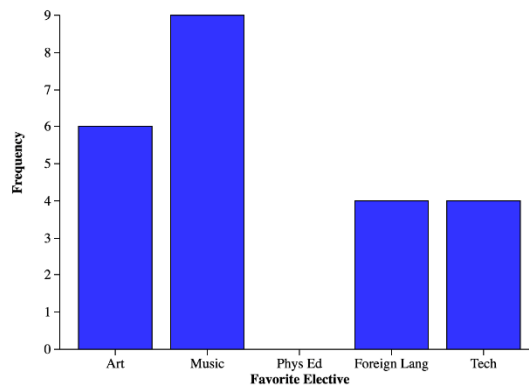
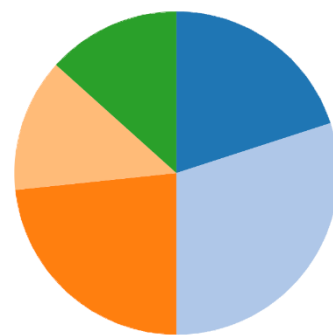
## How Are Your Favorite Classes Related?



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	Foreign 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.



Elective class	Core Class		Total
	Math	English	
Art	2	4	
Music	5	4	
P.E.	4	3	
Foreign Lang.	1	3	
Technology	4	0	
<b>Total</b>			

**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:

## Unit 1: Analyzing Data

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?

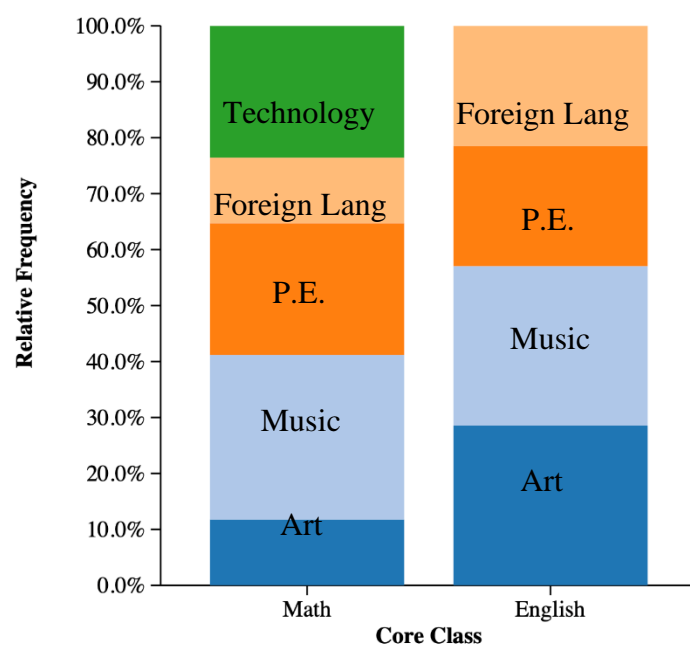
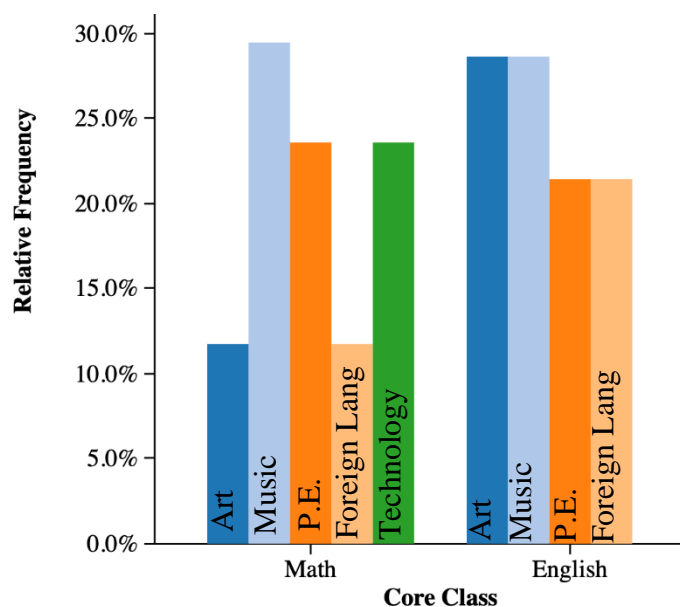
Name: \_\_\_\_\_ Core Class \_\_\_\_\_

**Elective  
class**

	Math	English	Total
Art	2	4	6
Music	5	4	9
P.E.	4	3	7
Foreign Lang.	1	3	4
Technology	4	0	4
<b>Total</b>	<b>16</b>	<b>14</b>	<b>30</b>

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.

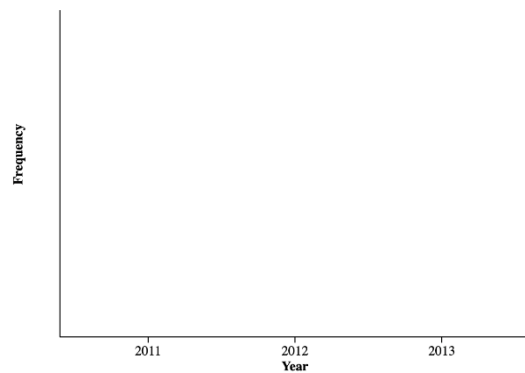
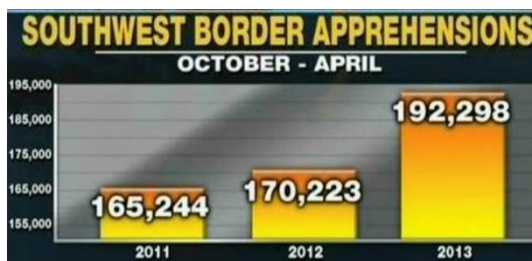
12. If there was not an association between favorite core subject and favorite elective, what would the graphs look like? Explain.

## Lesson 1.1 – Analyzing Categorical Data

QuickNotes

## Check Your Understanding:

1. The following graph was displayed by a national news organization. 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

- What proportion of the houses shown were built in Pinewood Estates?
- Find the distribution of Decade Built for the houses in this town using relative frequencies.
- 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.

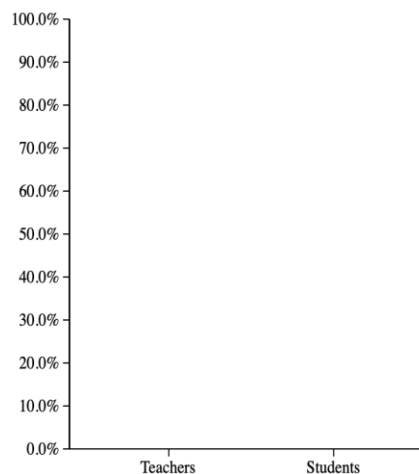
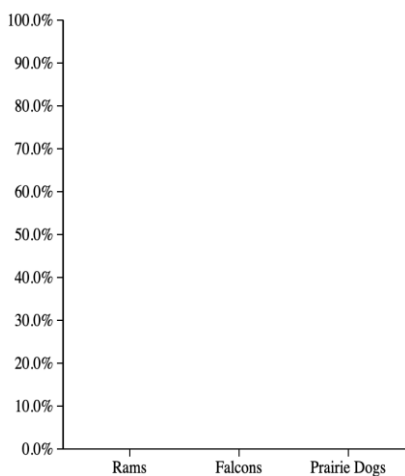
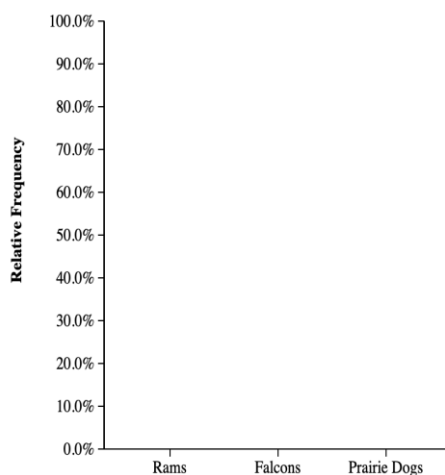


	Rams	Falcons	Prairie Dogs
Teachers	80%	10%	10%
Students	30%	60%	10%

- Create two bar graphs below to display the results. Use three different colors for the bars.
- 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.

Teachers

Students



- According to your displays, which mascot appears to have the most support? Explain.
- 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			

- 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.
- What should they make the EK mascot? Explain.

## Lesson 1.2 – Representing Categorical Data

QuickNotes

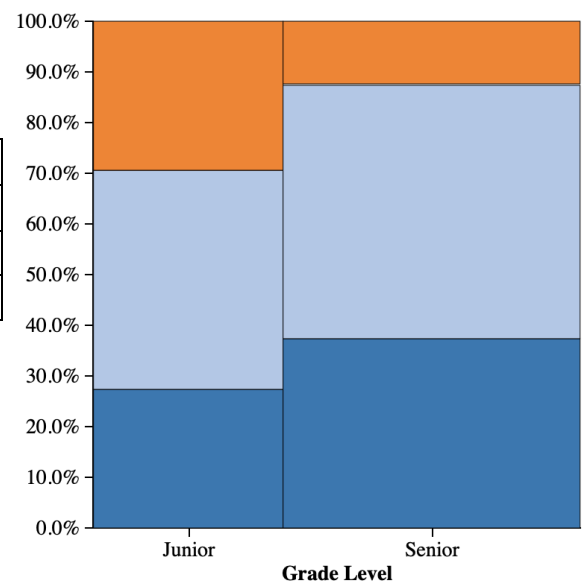
## Check Your Understanding:

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.

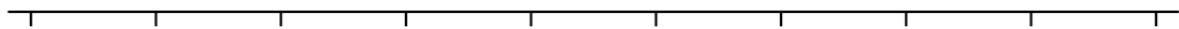


## How Many Pairs of Shoes Do You Own?



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 [statsmedic.com/applets](https://statsmedic.com/applets). Sketch the dotplot, stemplot, and histogram.

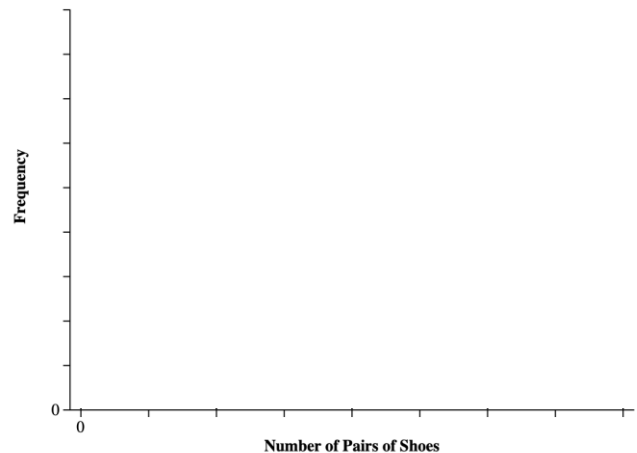


**Number of Pairs of Shoes**

0  
1  
2  
3  
4  
5  
6  
7

**Number of Pairs of Shoes**

**KEY:**



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?

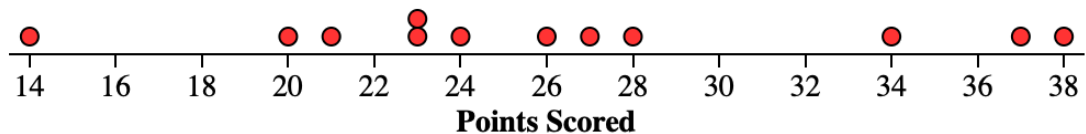


## Lesson 1.3 – Describing Quantitative Data

QuickNotes

## Check Your Understanding:

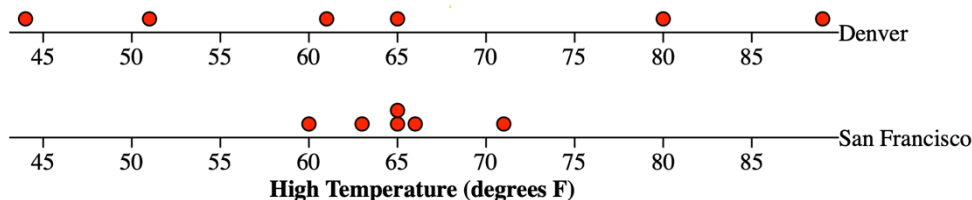
1. Mr. Wilcox is a huge fan of University of Michigan football. His favorite season was the 1997 season (a perfect season!). The dotplot shows the number of points scored by the U of M team in the 12 games that season.



- (a) Use the dotplot to create a stem and leaf plot of the distribution.
- (b) Describe the shape of the distribution.
- (c) Are there any potential outliers? Why?
- (d) 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.



- What is the same about the high temperatures in each city? What is different?
- One way to measure variability for a set of data is with **range**. Calculate the range for each city.
- 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) <sup>2</sup>
44		
51		
61		
65		
80		
89		
Total:		
Average (Distance from mean) <sup>2</sup> :		

- 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 [statsmedic.com/applets](https://statsmedic.com/applets). Enter

the data and find the summary statistics.

n	mean	SD	min	Q <sub>1</sub>	med	Q <sub>3</sub>	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 <sub>1</sub>	med	Q <sub>3</sub>	max

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

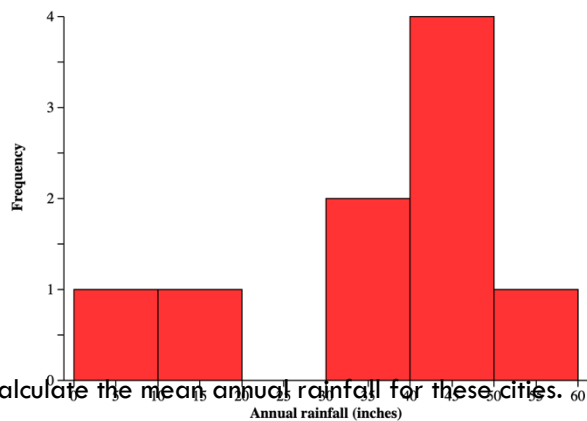
## Lesson 1.4 – Measuring Variability

QuickNotes

## Check Your Understanding

A researcher is interested in how much annual rainfall is typical in the United States. She takes 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



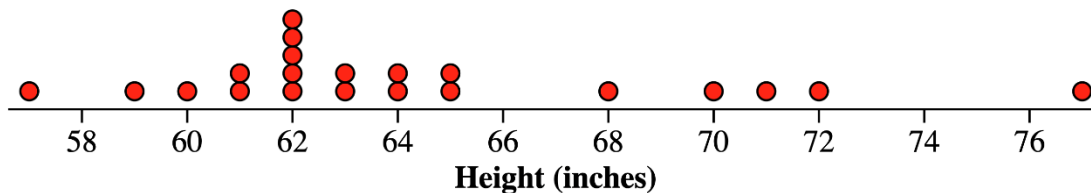
1. Calculate the mean annual rainfall for these cities.
2. Find the median annual rainfall for these cities.
3. Would you use the mean or the median to summarize the typical annual rainfall for a U.S. city? Explain.
4. The standard deviation of the annual rainfall for these 9 cities is 15.52 inches.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



## Where Do I Stand?

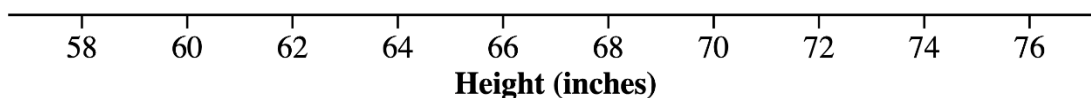
How does my height compare with the other AP Stats students in my class? To answer this question, Ashmita recorded the heights of everyone in her class. The heights (in inches) were:

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 the rules below). Are there any outliers? Show your work.

$$\text{way too small} < Q_1 - 1.5IQR \qquad \text{way too big} > Q_3 + 1.5IQR$$

6. Ashmita is 63 inches tall. How does her height compare with the other AP Stats students in her class?

## Lesson 1.5 – 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 archival Michigan State University football team. Write a few sentences comparing the distributions.

