Plan for Algebra 1 Unit 3: Two-Variable Statistics

Relevant Unit(s) to review: Grade 8 Unit 6: Associations in Data

Essential prior concepts to engage with this unit	• Create a scatter plot, and draw lines to fit data in a scatter plot. • Create and interpret a two-way table.
Brief narrative of approach	This unit formalizes a lot of learning that is introduced in grade 8. In grade 8, students informally constructed scatter plots and lines of fit, noticed linear patterns, and observed associations in categorical data using two-way tables. In this unit, students revisit two-way tables to find associations in categorical data using relative frequencies. The unit also builds on previous knowledge of scatter plots by assessing how well a linear model matches the data using residuals as well as the correlation coefficient for best-fit lines (found using technology). It closes with an exploration of the difference between correlation and causal relationships as well as an opportunity to apply this learning to anthropology and sports. Because of the substantial overlap between the two units, only three lessons are suggested for addition. Grade 8 Unit 6 Lesson 9 is included to give students an additional opportunity to explore and make sense of two-way tables and two-variable data. Additional grade 8 lessons are included to introduce scatter plots and develop students' informal sense of fit lines.

Lessons to Add	Lessons to Remove or Modify
1. 8.6.9: Activities 2 and 3. 2. 8.6.3: Scatter plots 3. 8.6.4: Fit lines	1. Remove Activity 1.3: This activity can be moved to practice outside of class. 2. Remove Activity 3.3: This activity invites students to invent their own data for imagined variables. It is an

	additional opportunity for students to explore association, but does not introduce new concepts. 3. Remove Lesson 6: This lesson explores residuals. No subsequent lessons address this topic, so if time is an issue, it can be removed. 4. Remove Lesson 10: This lesson is optional.
Lessons added: 3	Lessons removed: 3

Modified Plan for Algebra 1 Unit 3

Day	IM lesson	Notes
	A1.3 Check Your Readiness assessment	Check Your Readiness assessment: Pay careful attention to items 1–3. If students are successful with those items, then you can move more quickly through the grade 8 lessons.
1	8.6.9	Introduce students to two-way tables, including relative frequency. This provides students an opportunity to explore different representations of two-variable data, which should support their success in the first lessons of this unit, where they are expected to be able to interpret two-variable data.
2	A1.3.1 A1.3.2	Do Activities 1 and 2 from Lesson 1, followed by Activities 1 and 2 from Lesson 2. Do the cool-down from Lesson 1
3	A1.3.2 A1.3.3	Do Lesson 3 Activity 1, followed by Lesson 2 Activity 3, then Lesson 3 Activity 2. Finish with the cool-down for Lesson 3.
4	8.6.3	This lesson is added to introduce students to scatter plots and interpret the meaning of points on a scatter plot. Scatter plots are actually introduced in Lesson 2, so take extra care on the launch of Activity 2 of this lesson to make sure students understand how the table and the scatter plot are related.

5	8.6.4	This lesson is added to introduce students to fit lines.
6	A1.3.4	Students should be ready to look closer at the equations of the linear models and analyze the meaning of slope since they have thought about fit lines more informally in supplemental lessons.
7	A1.3.5	Fitting lines
8	A1.3.7	The correlation coefficient
9	A1.3.8	Using the correlation coefficient
10	A1.3.9	Causal relationships. Activity 2 is particularly important to emphasize.
11	assessment	End-of-Unit Assessment (remove item 1, which relies entirely on Lesson 6; remove or modify item 7 to assess understanding not related to residuals).

Priority and Category List for Lessons

High priority (+), Medium priority (0), Low priority (-)

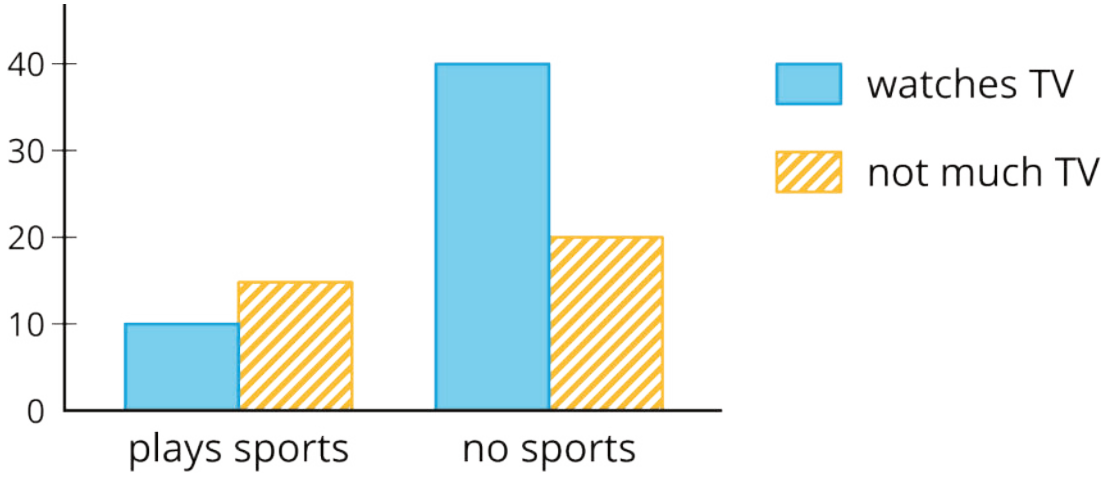
E: Explore, Play, and Discuss, D: Deep Dive, A: Synthesize and Apply

Lesson	Priority (+, 0, -)	Category (E, D, A)	Notes
A1.3.1	0	E	This lesson introduces two-way tables and connects to grade 8 work.
A1.3.2	0	D	This lesson builds on lesson one and introduces the idea of relative frequency.
A1.3.3	0	A	This lesson applies the learning from the first two lessons to introduce the concept of association between two variables.
A1.3.4	+	E	This lesson lays the foundation for the work with linear models of data for the rest of the unit.
A1.3.5	0	D	This lesson provides more practice and deeper understanding of fit lines.
A1.3.6	-	A	This lesson is about residuals. It is listed as lower priority, because it is the only lesson that addresses this concept, so if time is an issue, it can be cut without impacting learning in other lessons.
A1.3.7	+	E	This lesson introduces correlation coefficient.
A1.3.8	+	D	This lesson explores strength and sign of correlation coefficient.
A1.3.9	0	A	This lesson provides further opportunities to explore correlation coefficients. The second activity highlights the idea that association does not necessarily mean causation.
A1.3.10	-	A	This is an optional lesson that applies the learning of the unit.

Lesson 9: Looking for Associations

9.1: Notice and Wonder: Bar Association

What do you notice? What do you wonder?



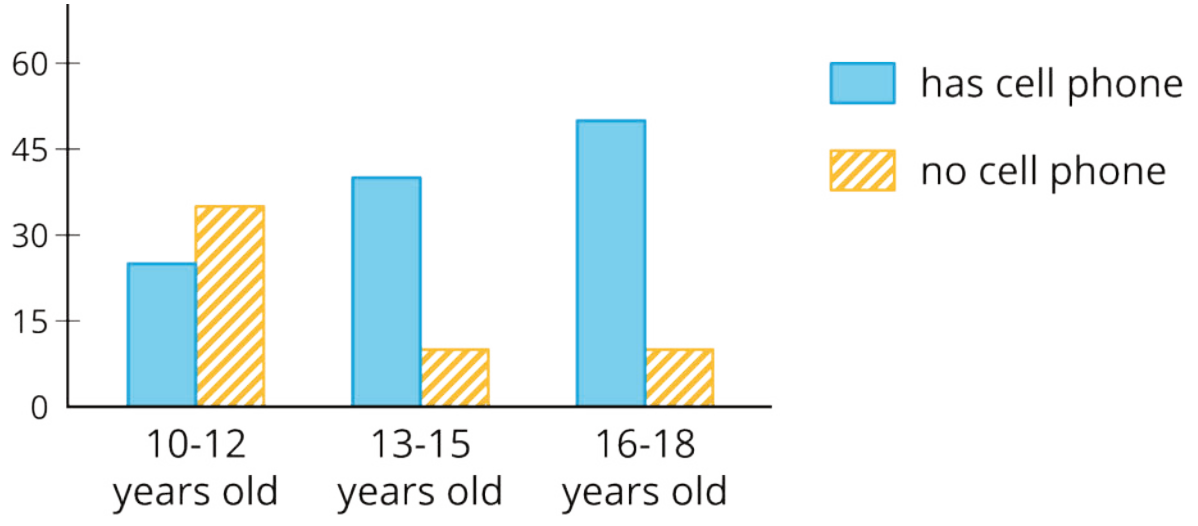
9.2: Card Sort: Matching Representations

Your teacher will hand out some cards.

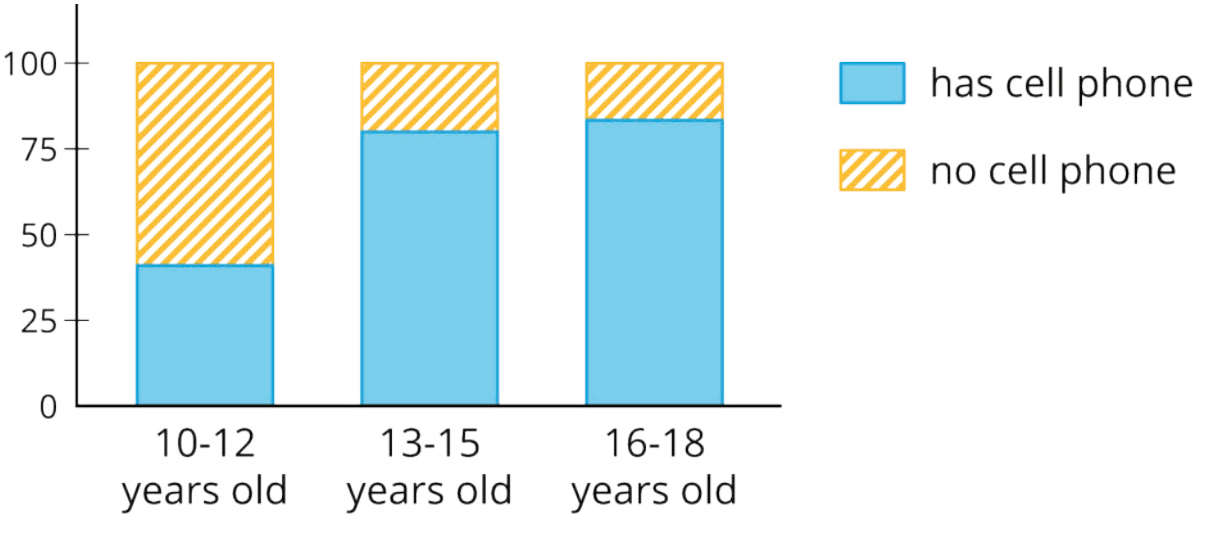
Some cards show **two-way tables** like this:

	has cell phone	does not have cell phone	total
10 to 12 years old	25	35	60
13 to 15 years old	40	10	50
16 to 18 years old	50	10	60
total	115	55	170

Some cards show bar graphs like this:



Some cards show **segmented bar graphs** like this:



The bar graphs and segmented bar graphs have their labels removed.

1. Put all the cards that describe the same situation in the same group.
2. One of the groups does not have a two-way table. Make a two-way table for the situation described by the graphs in the group.
3. Label the bar graphs and segmented bar graphs so that the categories represented by each bar are indicated.
4. Describe in your own words the kind of information shown by a segmented bar graph.

Are you ready for more?

One of the segmented bar graphs is missing. Construct a segmented bar graph that matches the other representations.

9.3: Building Another Type of Two-Way Table

Here is a two-way table that shows data about cell phone usage among children aged 10 to 18.

	has cell phone	does not have cell phone	total
10 to 12 years old	25	35	60
13 to 15 years old	40	10	50
16 to 18 years old	50	10	60
total	115	55	170

1. Complete the table. In each row, the entries for "has cell phone" and "does not have cell phone" should have the total 100%. Round entries to the nearest percentage point.

	has cell phone	does not have cell phone	total
10 to 12 years old	42%		
13 to 15 years old			100%
16 to 18 years old		17%	

   This is still a two-way table. Instead of showing *frequency*, this table shows **relative frequency**.

2. Two-way tables that show relative frequencies often don't include a "total" row at the bottom. Why?

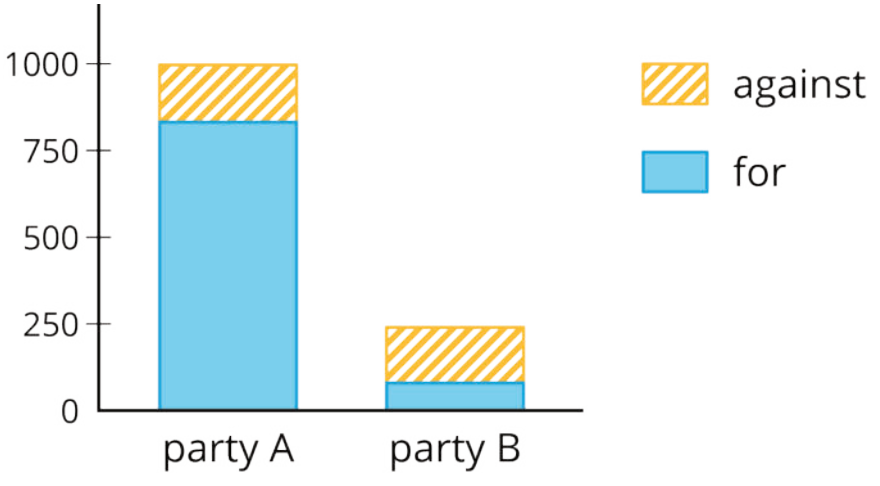
3. Is there an association between age and cell phone use? How does the two-way table of relative frequencies help to illustrate this?

Are you ready for more?

A pollster attends a rally and surveys many of the participants about whether they associate with political Party A or political Party B and whether they are for or against Proposition 3.14 going up for vote soon. The results are sorted into the table shown.

	for	against
party A	832	165
party B	80	160

- A news station reports these results by saying, "A poll shows that about the same number of people from both parties are voting against Proposition 3.14."
- A second news station shows this graphic.



1. Are any of the news reports misleading? Explain your reasoning.

2. Create a headline, graphic, and short description that more accurately represents the data in the table.

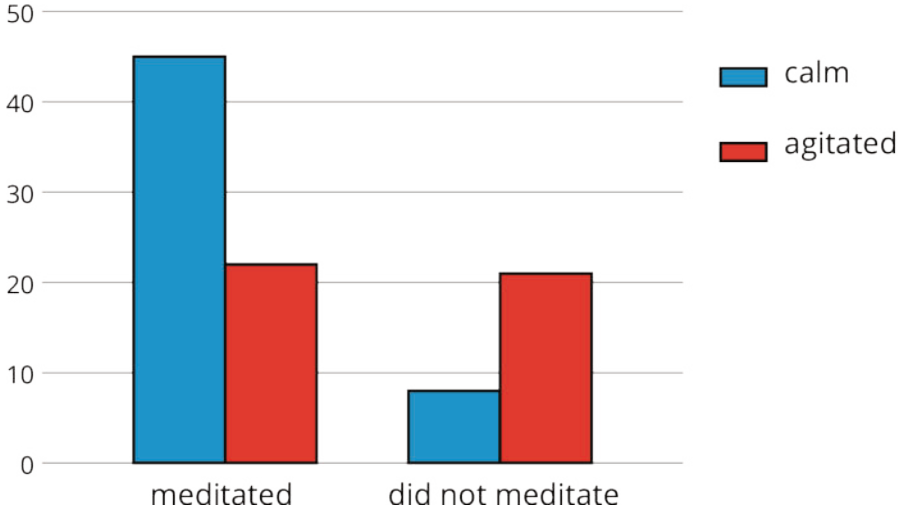
Lesson 9 Summary

When we collect data by counting things in various categories, like red, blue, or yellow, we call the data *categorical data*, and we say that color is a *categorical variable*.

We can use **two-way tables** to investigate possible connections between two categorical variables. For example, this two-way table of frequencies shows the results of a study of meditation and state of mind of athletes before a track meet.

	meditated	did not meditate	total
calm	45	8	53
agitated	23	21	44
total	68	29	97

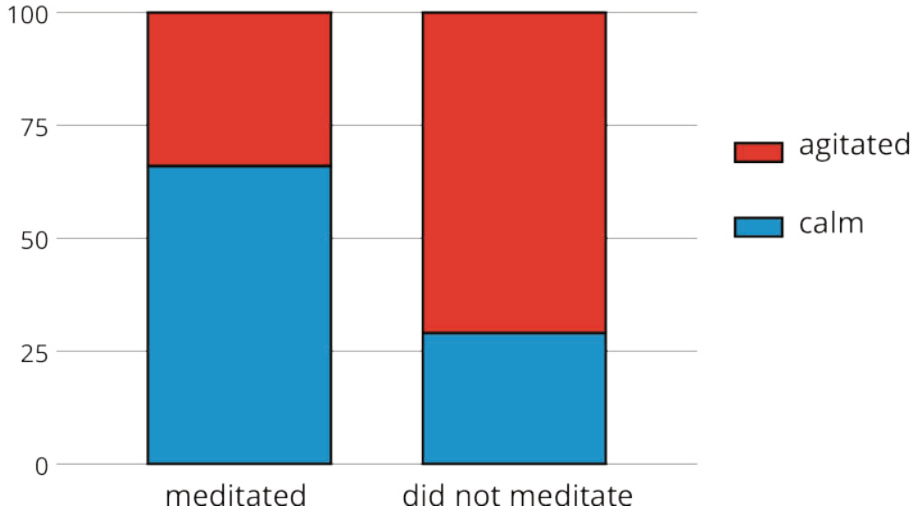
If we are interested in the question of whether there is an association between meditating and being calm, we might present the frequencies in a bar graph, grouping data about meditators and grouping data about non-meditators, so we can compare the numbers of calm and agitated athletes in each group.



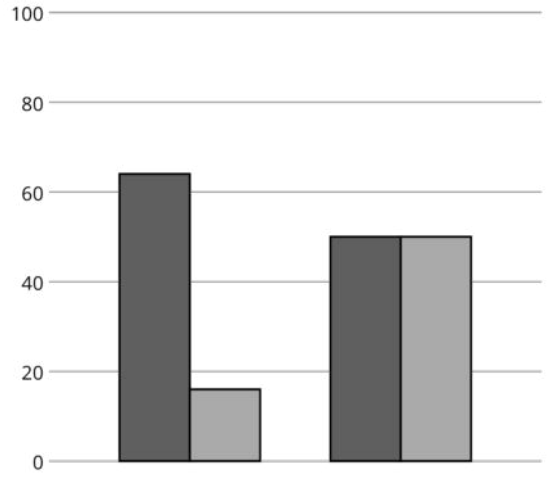
Notice that the number of athletes who did not meditate is small compared to the number who meditated (29 as compared to 68, as shown in the table).

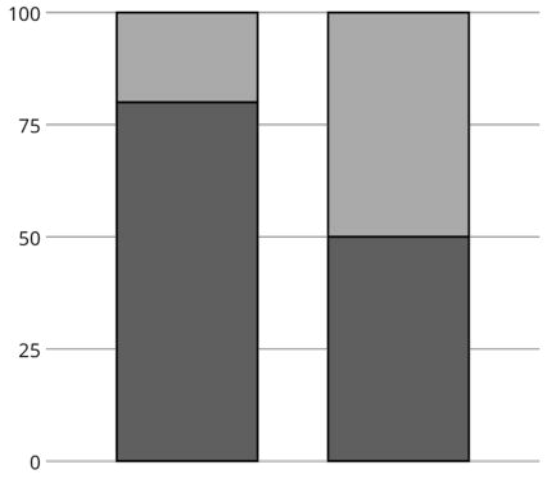
If we want to know the proportions of calm meditators and calm non-meditators, we can make a two-way table of **relative frequencies** and present the relative frequencies in a **segmented bar graph**.

	meditated	did not meditate
calm	66%	28%
agitated	34%	72%
total	100%	100%

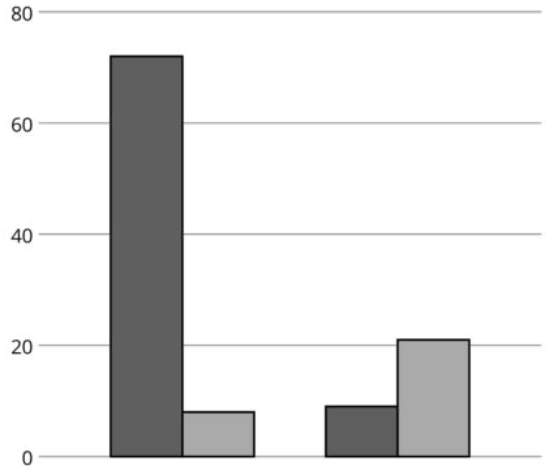


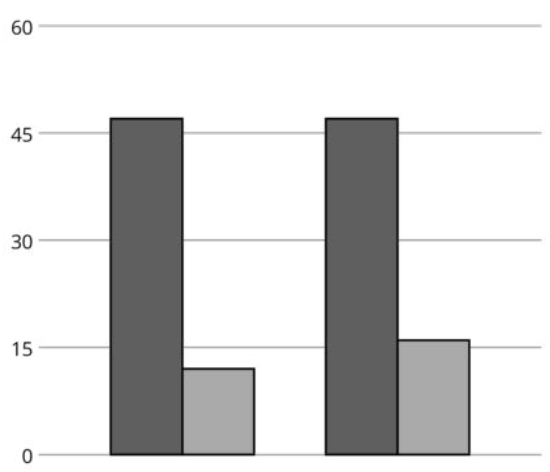
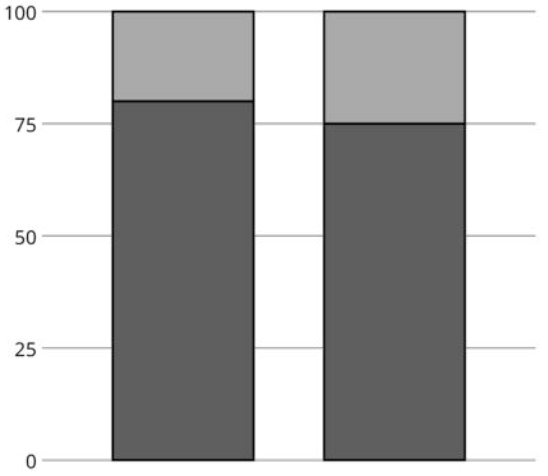
	liked the movie	disliked the movie	total
read the book	64	16	80
did not read the book	50	50	100
total	114	66	180





	satisfied with service	dissatisfied with service	total
satisfied with food	72	8	80
dissatisfied with food	9	21	30
total	81	29	110



	 60 45 30 15 0	 100 75 50 25 0

Lesson 9: Looking for Associations

Cool Down: Guitar and Golf

1. In a class of 25 students, some students play a sport, some play a musical instrument, some do both, some do neither. Complete the two-way table to show data that might come from this class.

	plays an instrument	does not play an instrument	total
plays a sport			
does not play a sport	5		
total			25

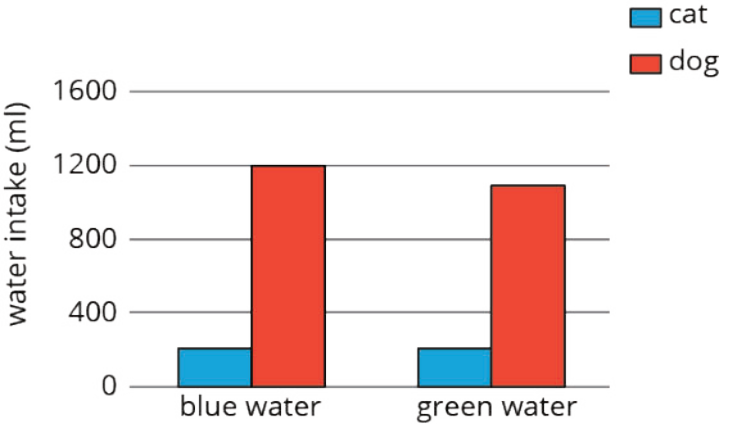
2. Using the entries from the previous table, complete this table so that it shows relative frequencies. Round entries to the nearest percentage point.

	plays an instrument	does not play an instrument	total
plays a sport			
does not play a sport			

Unit 6 Lesson 9 Cumulative Practice Problems

1. A scientist wants to know if the color of the water affects how much animals drink. The average amount of water each animal drinks was recorded in milliliters for a week and then graphed. Is there evidence to suggest an association between water color and animal?

	cat intake (ml)	dog intake (ml)	total (ml)
blue water	210	1200	1410
green water	200	1100	1300
total	410	2300	2710



2. A farmer brings his produce to the farmer's market and records whether people buy lettuce, apples, both, or something else.

	bought apples	did not buy apples
bought lettuce	14	58
did not buy lettuce	8	29

Make a table that shows the relative frequencies for each row. Use this table to decide if there is an association between buying lettuce and buying apples.

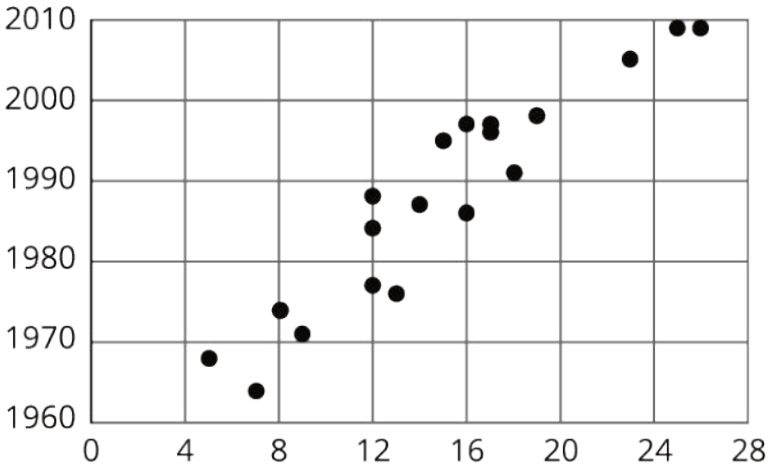
3. Researchers at a media company want to study news-reading habits among different age groups. They tracked print and online subscription data and made a 2-way table.

	internet media	print media
18-25 year olds	151	28
26-45 year olds	132	72
46-65 year olds	48	165

a. Create a segmented bar graph using one bar for each row of the table.

b. Is there an association between age groups and the method they use to read articles? Explain your reasoning.

4. Using the data in the scatter plot, what is a reasonable slope of a model that fits this data?



A. -2.5

B. -1

C. 1

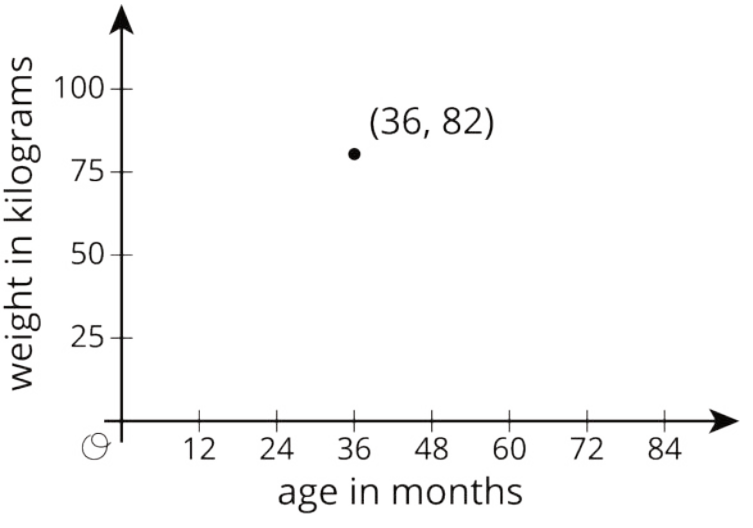
D. 2.5

(From Unit 6, Lesson 6.)

Lesson 3: What a Point in a Scatter Plot Means

3.1: The Giant Panda

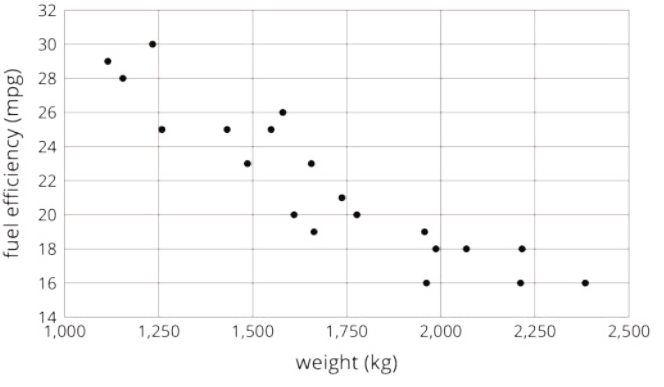
A giant panda lives in a zoo. What does the point on the graph tell you about the panda?



3.2: Weight and Fuel Efficiency

The table and scatter plot show weights and fuel efficiencies of 18 cars.

car	weight (kg)	fuel efficiency
A	1,549	25
B	1,610	20
C	1,737	21
D	1,777	20
E	1,486	23
F	1,962	16
G	2,384	16
H	1,957	19
I	2,212	16
J	1,115	29
K	2,068	18
L	1,663	19
M	2,216	18
N	1,432	25
O	1,987	18
P	1,580	26
Q	1,234	30
R	1,656	23



1. Which point in the scatter plot represents Car L's measurements?

2. What is the fuel efficiency of the car with the greatest weight?

3. What is the weight of the car with the greatest fuel efficiency?

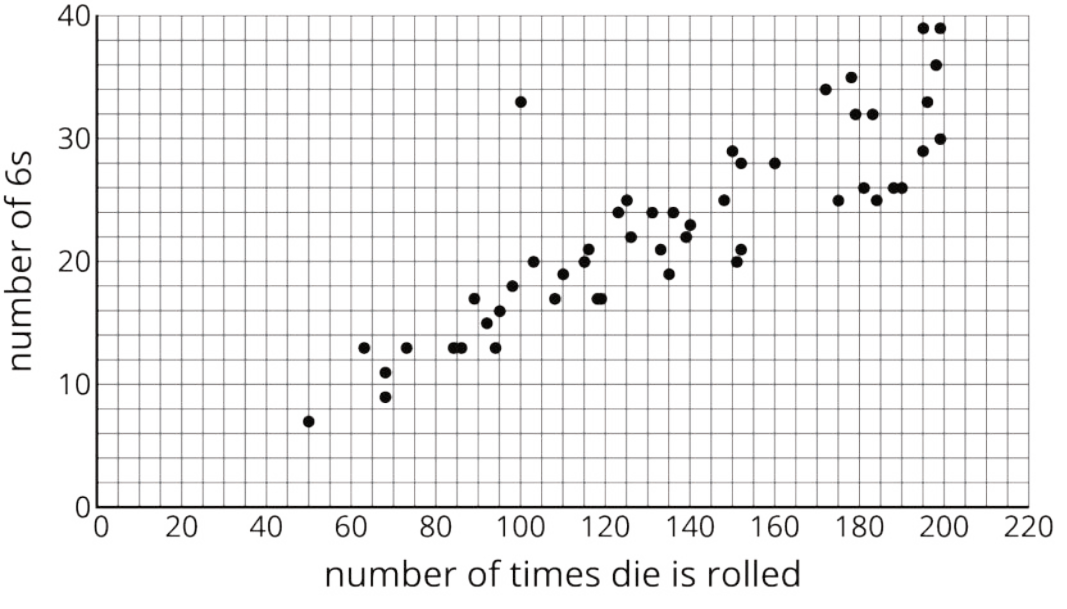
4. Car S weighs 1,912 kilograms and gets 16 miles per gallon. On the scatter plot, plot a point that represents Car S's measurements.

5. Cars N and O, shown in the scatter plot, are made by the same company. Compare their weights and fuel efficiencies. Does anything surprise you about these cars?

6. A different company makes Cars F and G. Compare their weights and fuel efficiencies. Does anything surprise you about these cars?

Are you ready for more?

After a board game competition, the tournament director collects 50 dice from the games played and rolls each one until he gets bored and tries a different one. The scatter plot shows the number of times he rolled each die and the number of 6s that resulted during those rolls.

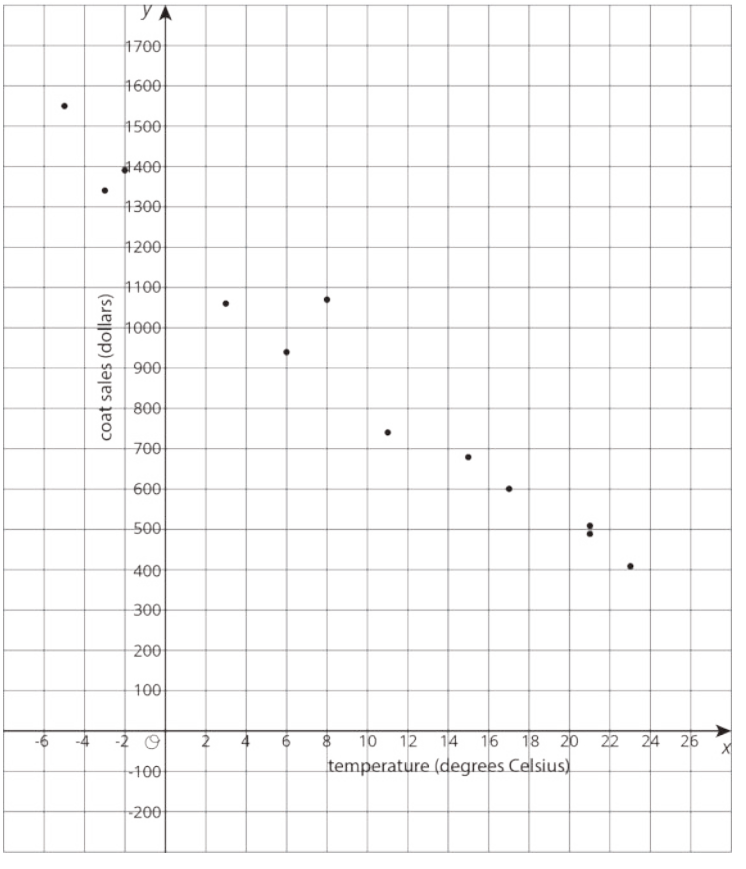


Select a point in the scatter plot and give its approximate coordinates, then tell the story of that point in the context of the problem.

3.3: Coat Sales

A clothing store keeps track of the average monthly temperature in degrees Celsius and coat sales in dollars.

temperature (degrees Celsius)	coat sales (dollars)
-5	1,550
-3	1,340
3	1,060
8	1,070
15	680
21	490
23	410
21	510
17	600
11	740
6	940
-2	1,390



1. What does the point $(15, 680)$ represent?

2. For the month with the lowest average temperature, estimate the total amount made from coat sales. Explain how you used the table to find this information.

3. For the month with the smallest coat sales, estimate the average monthly temperature. Explain how you used the scatter plot to find this information.

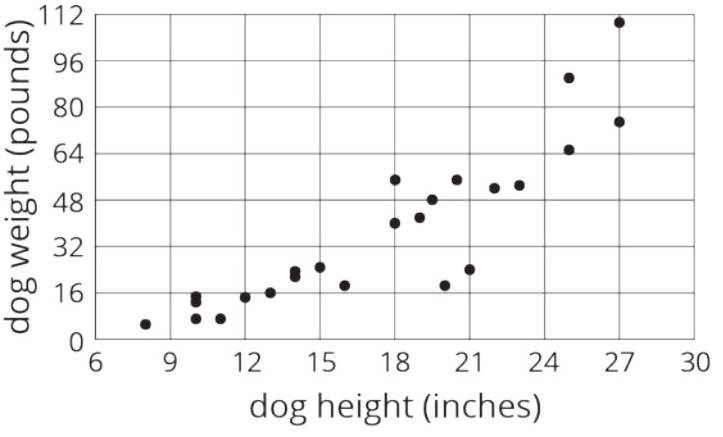
4. If there were a point at $(0, A)$ what would it represent? Use the scatter plot to estimate a value for A.

5. What would a point at $(B, 0)$ represent? Use the scatter plot to estimate a value for B.

6. Would it make sense to use this trend to estimate the value of sales when the average monthly temperature is 60 degrees Celsius? Explain your reasoning.

Lesson 3 Summary

Scatter plots show two measurements for each individual from a group. For example, this scatter plot shows the weight and height for each dog from a group of 25 dogs.



We can see that the tallest dogs are 27 inches, and that one of those tallest dogs weighs about 75 pounds while the other weighs about 110 pounds. This shows us that dog weight is not a function of dog height because there would be two different outputs for the same input. But we can see a general trend: Taller dogs tend to weigh more than shorter dogs. There are exceptions. For example, there is a dog that is 18 inches tall and weighs over 50 pounds, and there is another dog that is 21 inches tall but weighs less than 30 pounds.

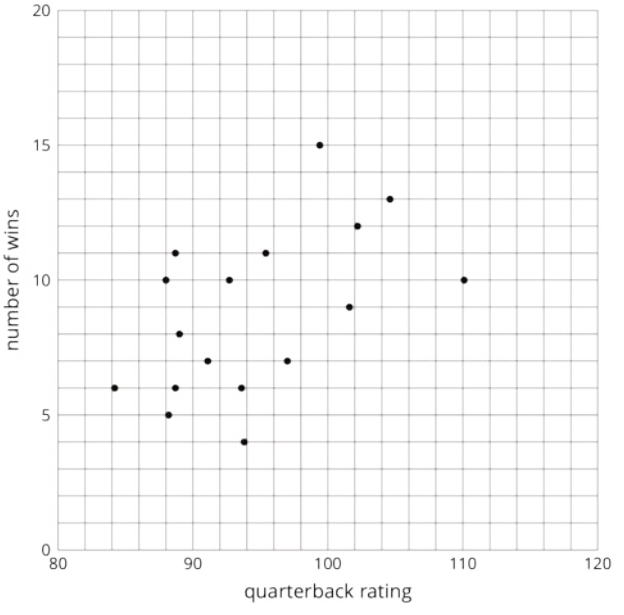
When we collect data by measuring attributes like height, weight, area, or volume, we call the data *numerical data* (or measurement data), and we say that height, weight, area, or volume is a *numerical variable*. Upcoming lessons will discuss how to identify and describe trends in data that has been collected.

Lesson 3: What a Point in a Scatter Plot Means

Cool Down: Quarterbacks

Here are a table and scatter plot that show ratings and wins for quarterbacks who started 16 games this season.

player	quarterback rating	number of wins
A	93.8	4
B	102.2	12
C	93.6	6
D	89	8
E	88.2	5
F	97	7
G	88.7	6
H	91.1	7
I	92.7	10
J	88	10
K	101.6	9
L	104.6	13
M	84.2	6
N	99.4	15
O	110.1	10
P	95.4	11
Q	88.7	11



1. Circle the point in the scatter plot that represents Player K's data.

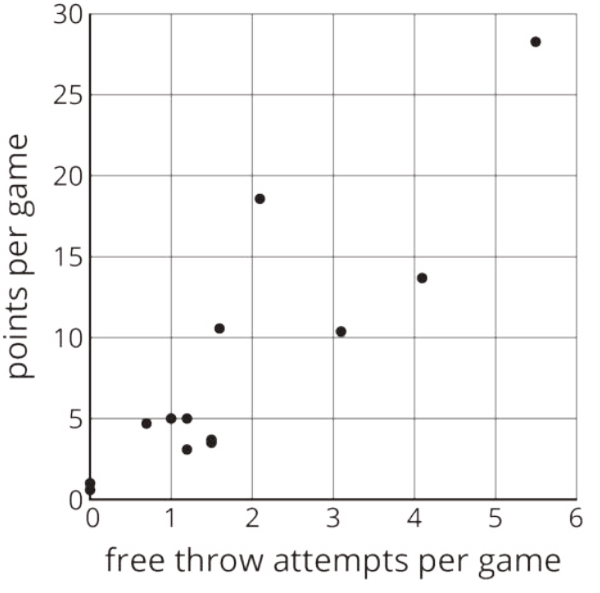
2. Which quarterback's data are represented by the point farthest to the left?

3. Player R is not included in the table because he did not start 16 games this year. He did have a quarterback rating of 99.4 and his team won 8 games. On the scatter plot, plot a point that represents Player R's data.

Unit 6 Lesson 3 Cumulative Practice Problems

1. Here is a table and a scatter plot that compares points per game to free throw attempts for a basketball team during a tournament.

player	free throw attempts	points
player A	5.5	28.3
player B	2.1	18.6
player C	4.1	13.7
player D	1.6	10.6
player E	3.1	10.4
player F	1	5
player G	1.2	5
player H	0.7	4.7
player I	1.5	3.7
player J	1.5	3.5
player K	1.2	3.1
player L	0	1
player M	0	0.8
player N	0	0.6



a. Circle the point that represents the data for Player E.

b. What does the point $(2.1, 18.6)$ represent?

c. In that same tournament, Player O on another team scored 14.3 points per game with 4.8 free throw attempts per game. Plot a point on the graph that shows this information.

2. Select **all** the representations that are appropriate for comparing exam score to number of hours of sleep the night before the exam.

   A. Histogram

   B. Scatter plot

   C. Dot plot

   D. Table

   E. Box plot

(From Unit 6, Lesson 2.)

3. A cylinder has a volume of 36π cm^3 and height h. Complete this table for the volume of other cylinders with the same radius but different heights.

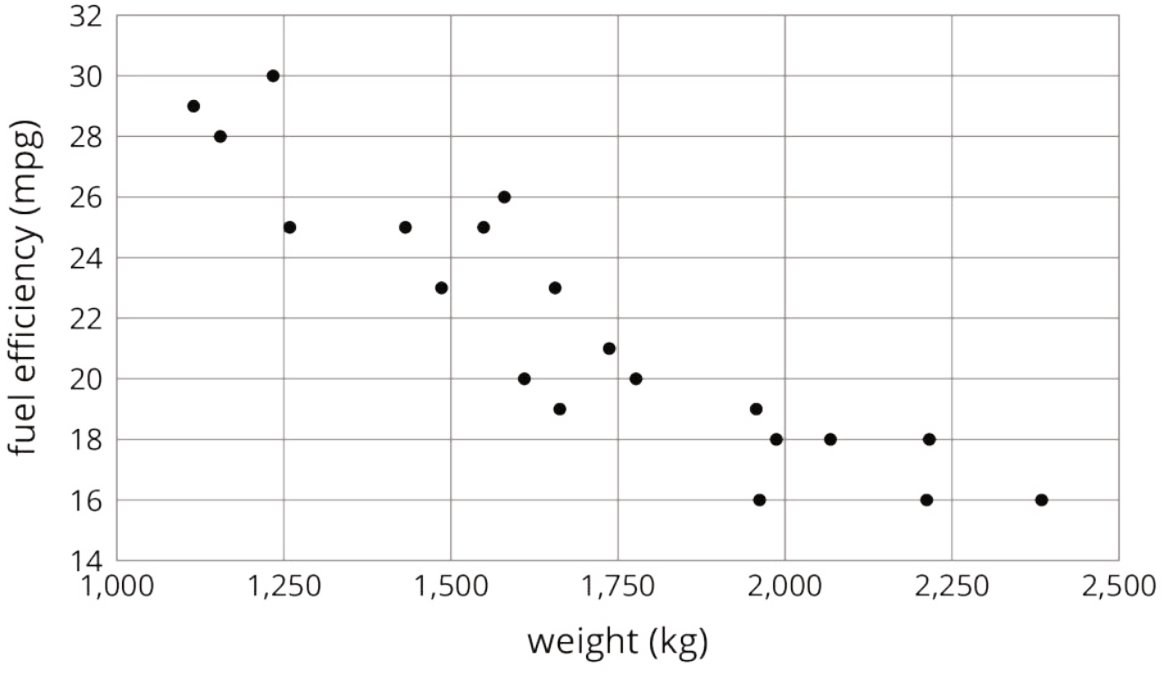
height (cm)	volume (cm^3)
h	36π
$2h$	
$5h$	
$\frac{h}{2}$	
$\frac{h}{5}$	

(From Unit 5, Lesson 17.)

Lesson 4: Fitting a Line to Data

4.1: Predict This

Here is a scatter plot that shows weights and fuel efficiencies of 20 different types of cars.



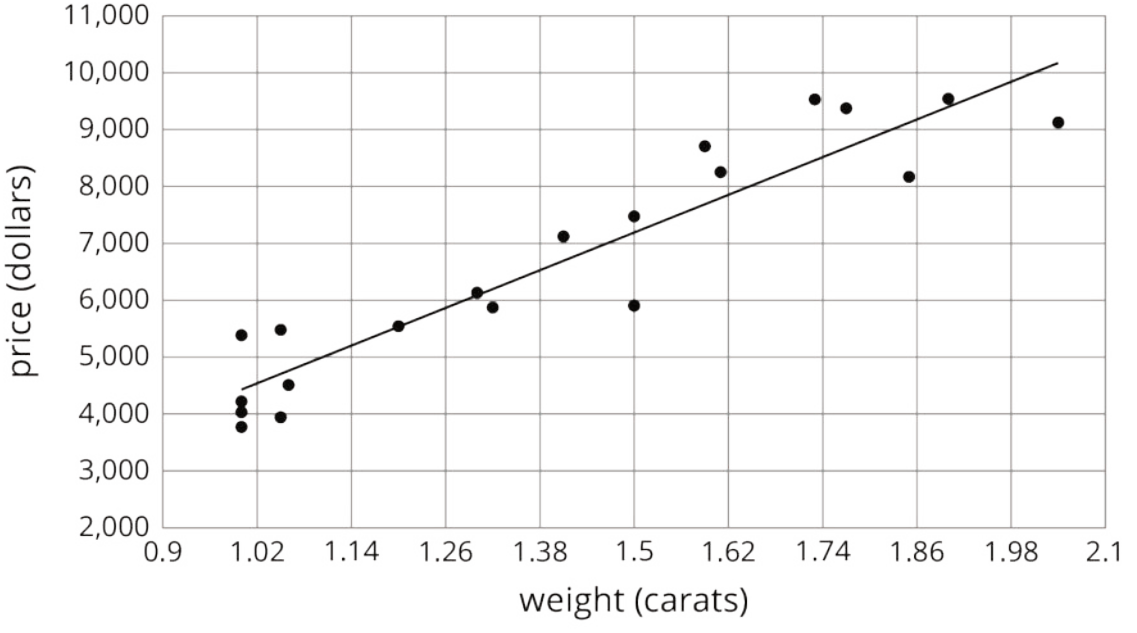
If a car weighs 1,750 kg, would you expect its fuel efficiency to be closer to 22 mpg or to 28 mpg? Explain your reasoning.

4.2: Shine Bright

Here is a table that shows weights and prices of 20 different diamonds.

weight (carats)	actual price (dollars)	predicted price (dollars)
1	3,772	4,429
1	4,221	4,429
1	4,032	4,429
1	5,385	4,429
1.05	3,942	4,705
1.05	4,480	4,705
1.06	4,511	4,760
1.2	5,544	5,533
1.3	6,131	6,085
1.32	5,872	6,195
1.41	7,122	6,692
1.5	7,474	7,189
1.5	5,904	7,189
1.59	8,706	7,686
1.61	8,252	7,796
1.73	9,530	8,459
1.77	9,374	8,679
1.85	8,169	9,121
1.9	9,541	9,397
2.04	9,125	10,170

The scatter plot shows the prices and weights of the 20 diamonds together with the graph of $y = 5{,}520x - 1{,}091$.



The function described by the equation $y = 5{,}520x - 1{,}091$ is a *model* of the relationship between a diamond's weight and its price.

This model *predicts* the price of a diamond from its weight. These predicted prices are shown in the third column of the table.

1. Two diamonds that both weigh 1.5 carats have different prices. What are their prices? How can you see this in the table? How can you see this in the graph?

2. The model predicts that when the weight is 1.5 carats, the price will be $7,189. How can you see this in the graph? How can you see this using the equation?

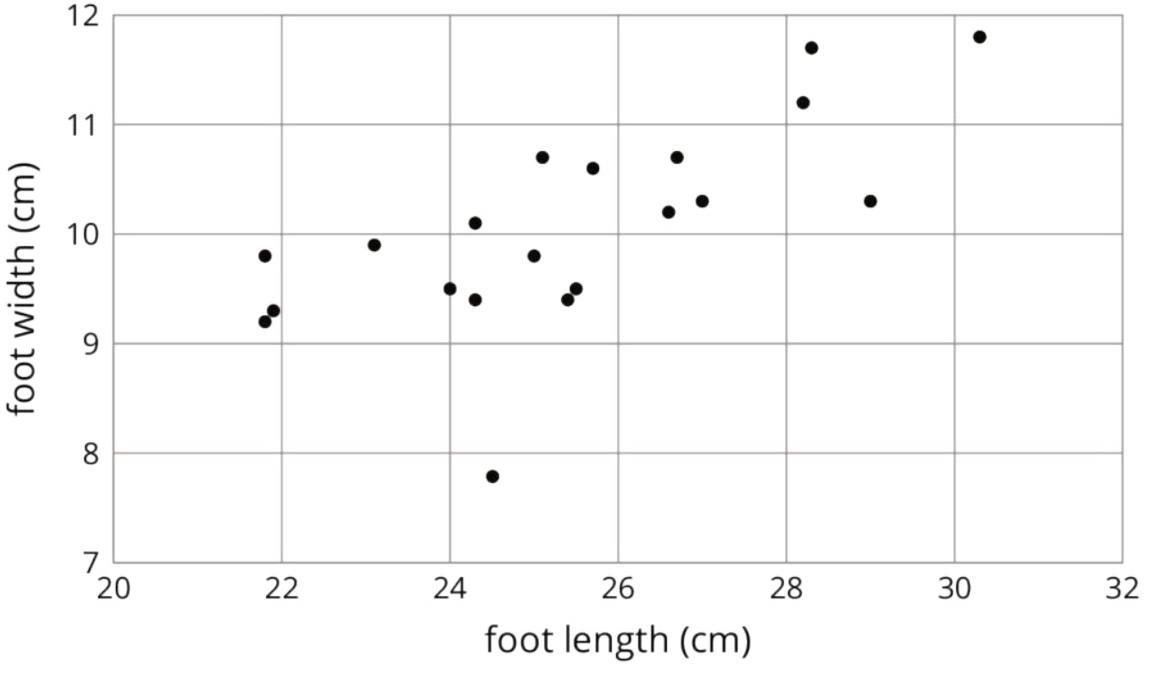
3. One of the diamonds weighs 1.9 carats. What does the model predict for its price? How does that compare to the actual price?

4. Find a diamond for which the model makes a very good prediction of the actual price. How can you see this in the table? In the graph?

5. Find a diamond for which the model's prediction is not very close to the actual price. How can you see this in the table? In the graph?

4.3: The Agony of the Feet

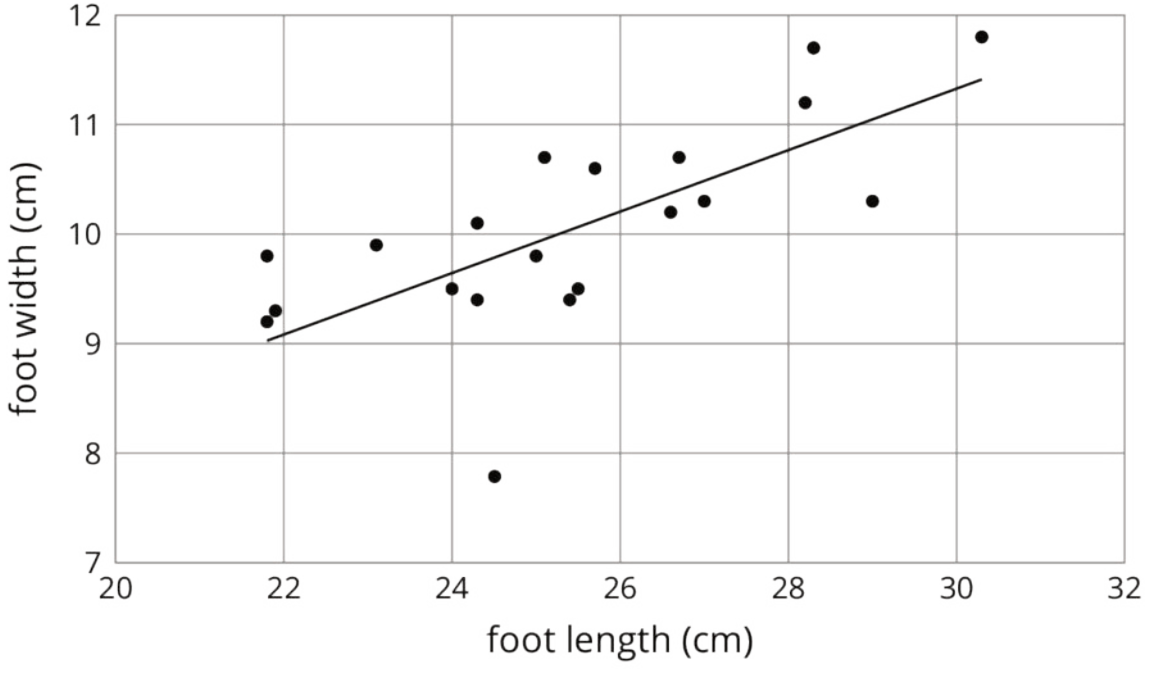
Here is a scatter plot that shows lengths and widths of 20 different left feet.



1. Estimate the widths of the longest foot and the shortest foot.

2. Estimate the lengths of the widest foot and the narrowest foot.

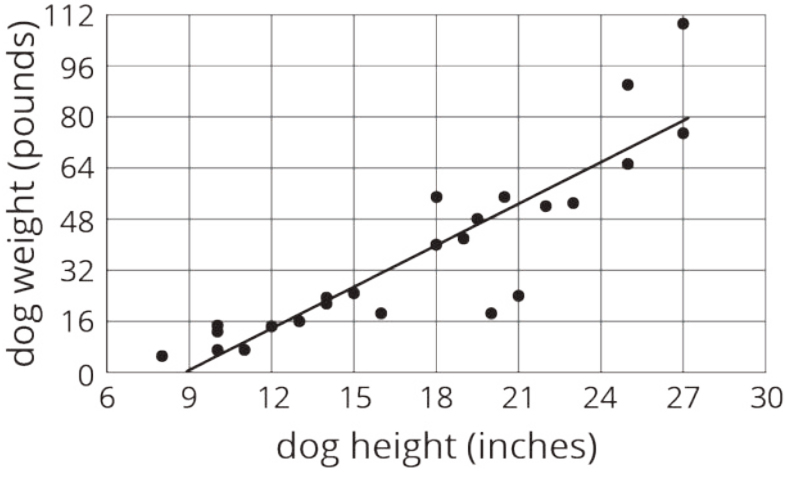
3. Here is the same scatter plot together with the graph of a model for the relationship between foot length and width.



Circle the data point that seems weird when compared to the model. What length and width does that point represent?

Lesson 4 Summary

Sometimes, we can use a linear function as a model of the relationship between two variables. For example, here is a scatter plot that shows heights and weights of 25 dogs together with the graph of a linear function which is a model for the relationship between a dog's height and its weight.



We can see that the model does a good job of predicting the weight given the height for some dogs. These correspond to points on or near the line. The model doesn't do a very good job of predicting the weight given the height for the dogs whose points are far from the line.

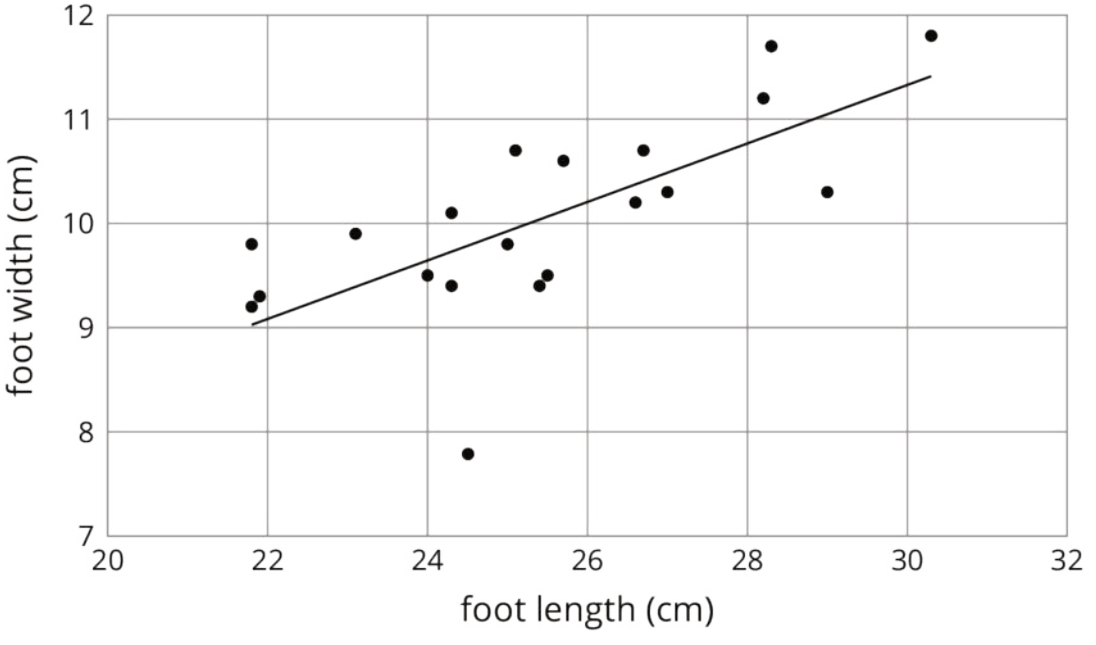
For example, there is a dog that is about 20 inches tall and weighs a little more than 16 pounds. The model predicts that the weight would be about 48 pounds. We say that the model *overpredicts* the weight of this dog. There is also a dog that is 27 inches tall and weighs about 110 pounds. The model predicts that its weight will be a little less than 80 pounds. We say the model *underpredicts* the weight of this dog.

Sometimes a data point is far away from the other points or doesn't fit a trend that all the other points fit. We call these **outliers**.

Lesson 4: Fitting a Line to Data

Cool Down: A 1 Foot Foot

Here is a scatter plot that shows lengths and widths of 20 left feet, together with the graph of a model of the relationship between foot length and width.



1. Draw a box around the point that represents the foot with length closest to 29 cm.

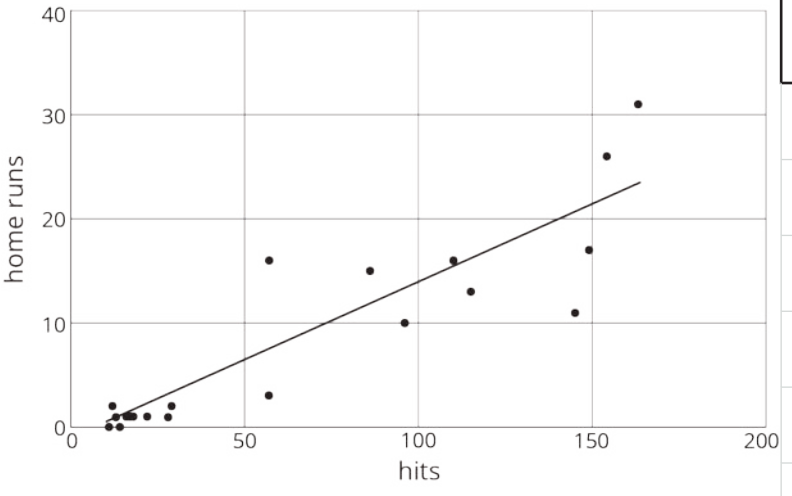
2. What is the approximate width of this foot?

3. What width does the model predict for a foot with length 29 cm?

Unit 6 Lesson 4 Cumulative Practice Problems

1. The scatter plot shows the number of hits and home runs for 20 baseball players who had at least 10 hits last season. The table shows the values for 15 of those players.

   The model, represented by $y = 0.15x - 1.5$, is graphed with a scatter plot.



hits	home runs	predicted home runs
12	2	0.3
22	1	1.8
154	26	21.6
145	11	20.3
110	16	15
57	3	7.1
149	17	20.9
29	2	2.9
13	1	0.5
18	1	1.2
86	15	11.4
163	31	23
115	13	15.8
57	16	7.1
96	10	12.9

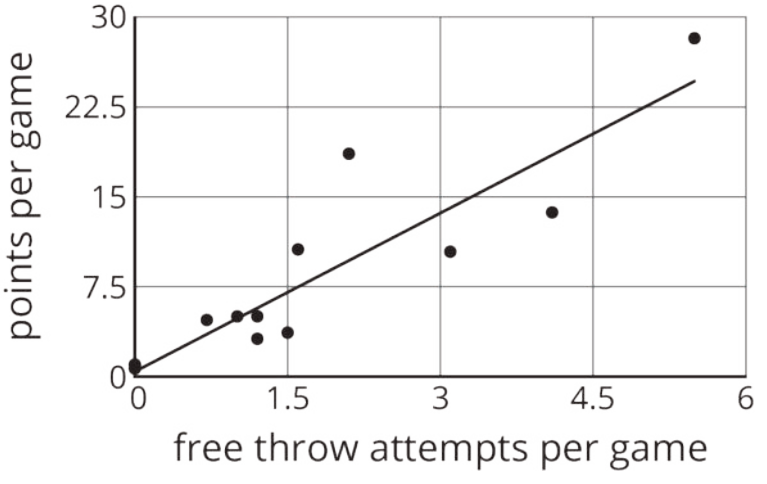
   Use the graph and the table to answer the questions.

   a. Player A had 154 hits in 2015. How many home runs did he have? How many was he predicted to have?

   b. Player B was the player who most outperformed the prediction. How many hits did Player B have last season?

   c. What would you expect to see in the graph for a player who hit many fewer home runs than the model predicted?

2. Here is a scatter plot that compares points per game to free throw attempts per game for basketball players in a tournament. The model, represented by $y = 4.413x + 0.377$, is graphed with the scatter plot. Here, x represents free throw attempts per game, and y represents points per game.



a. Circle any data points that appear to be outliers.

b. What does it mean for a point to be far above the line in this situation?

c. Based on the model, how many points per game would you expect a player who attempts 4.5 free throws per game to have? Round your answer to the nearest tenth of a point per game.

d. One of the players scored 13.7 points per game with 4.1 free throw attempts per game. How does this compare to what the model predicts for this player?