NAME DATE PERIOD

Unit 6, Lesson 1: Organizing Data

Let's find ways to show patterns in data

1.1: Notice and Wonder: Messy Data

Here is a table of data. Each row shows two measurements of a triangle.

length of short side (cm)	length of perimeter (cm)
0.25	1
2	7.5
6.5	22
3	9.5
0.5	2
1.25	3.5
3.5	12.5
1.5	5
4	14
1	2.5

What do you notice? What do you wonder?

NAME DATE PERIOD

1.2: Seeing the Data

Here is the table of isosceles right triangle measurements from the warm-up and an empty table.

length of short sides (cm)	length of perimeter (cm)
0.25	1
2	7.5
6.5	22
3	9.5
0.5	2
1.25	3.5
3.5	12.5
1.5	5
4	14
1	2.5

length of short sides (cm)	length of perimeter (cm)

1. How can you organize the measurements from the first table so that any patterns are easier to see? Write the organized measurements in the empty table.

2. For each of the following lengths, estimate the perimeter of an isosceles right triangle whose short sides have that length. Explain your reasoning for each triangle.
   a. length of short sides is 0.75 cm

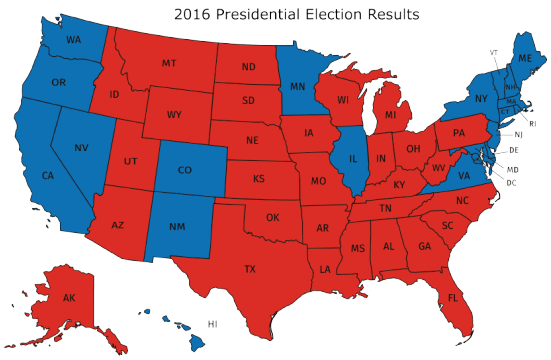
NAME DATE PERIOD

b. length of short sides is 5 cm

c. length of short sides is 10 cm

Are you ready for more?

In addition to the graphic representations of data you have learned, there are others that make sense in other situations. Examine the maps showing the results of the elections for United States president for 2012 and 2016. In red are the states where a majority of electorate votes were cast for the Republican nominee. In blue are the states where a majority of the electorate votes were cast for the Democrat nominee.



1. What information can you see in these maps that would be more difficult to see in a bar graph showing the number of electorate votes for the 2 main candidates?

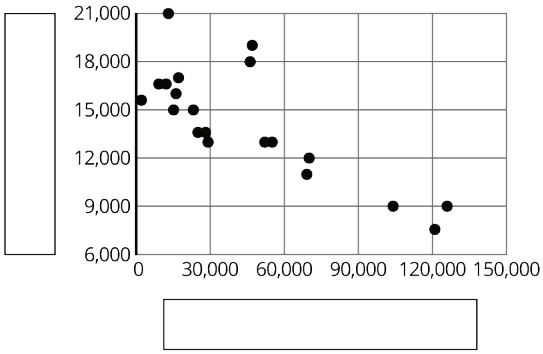
2. Why are these representations appropriate for the data that is shown?

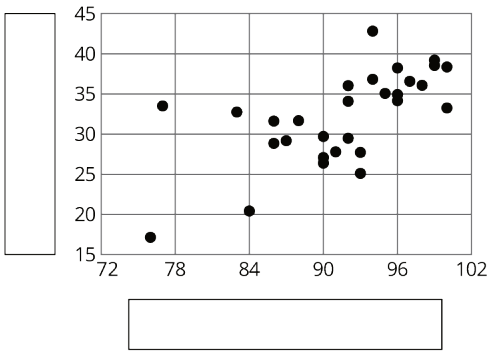
NAME DATE PERIOD

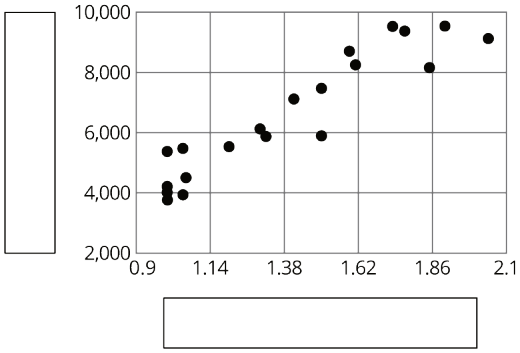
1.3: Tables and Their Scatter Plots

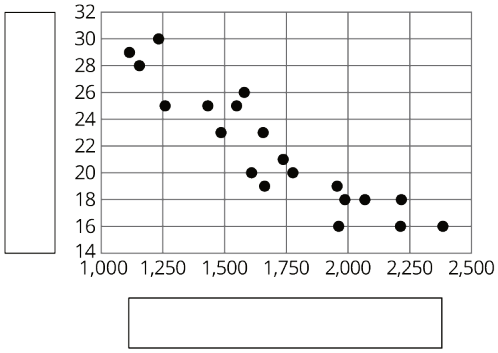
Here are four **scatter plots**. Your teacher will give you four tables of data.

- Match each table with one of the scatter plots.
- Use information from the tables to label the axes for each scatter plot.









NAME DATE PERIOD

Lesson 1 Summary

Consider the data collected from pulling back a toy car and then letting it go forward. In the first table, the data may not seem to have an obvious pattern. The second table has the same data and shows that both values are increasing together.

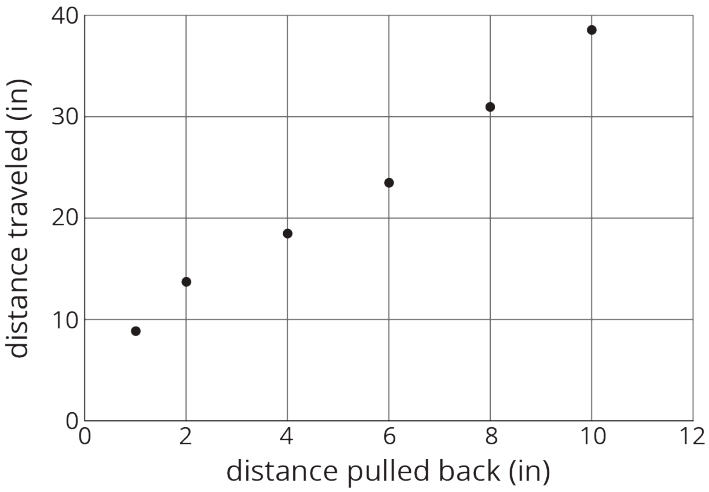
Unorganized table

distance pulled back (in)	distance traveled (in)
6	23.57
4	18.48
10	38.66
8	31.12
2	13.86
1	8.95

Organized table

distance pulled back (in)	distance traveled (in)
1	8.95
2	13.86
4	18.48
6	23.57
8	31.12
10	38.66

A scatter plot of the data makes the pattern clear enough that we can estimate how far the car will travel when it is pulled back 5 inches.



Patterns in data can sometimes become more obvious when reorganized in a table or when represented in **scatter plots** or other diagrams. If a pattern is observed, it can sometimes be used to make predictions.

Lesson 1 Glossary Terms

- scatter plot

NAME DATE PERIOD

Unit 6, Lesson 1: Organizing Data

1. Here is data on the number of cases of whooping cough from 1939 to 1955.

year	number of cases
1941	222,202
1950	120,718
1945	133,792
1942	191,383
1953	37,129
1939	103,188
1951	68,687
1948	74,715
1955	62,786
1952	45,030
1940	183,866
1954	60,866
1944	109,873
1946	109,860
1943	191,890
1949	69,479
1947	156,517

a. Make a new table that orders the data by year.

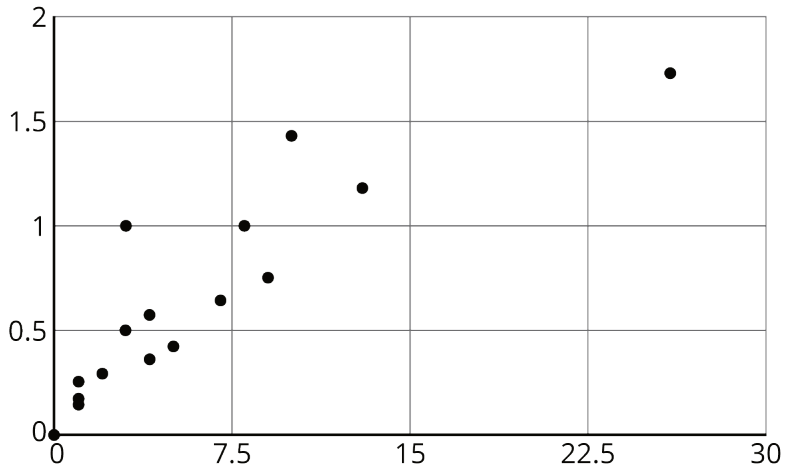
b. Which years in this period of time had fewer than 100,000 cases of whooping cough?

NAME DATE PERIOD

c. Based on this data, would you expect 1956 to have closer to 50,000 cases or closer to 100,000 cases?

2. In volleyball statistics, a block is recorded when a player deflects the ball hit from the opposing team. Additionally, scorekeepers often keep track of the average number of blocks a player records in a game. Here is part of a table that records the number of blocks and blocks per game for each player in a women's volleyball tournament. A scatter plot that goes with the table follows.

blocks	blocks per game
13	1.18
1	0.17
5	0.42
0	0
0	0
7	0.64



Label the axes of the scatter plot with the necessary information.

3. A cylinder has a radius of 4 cm and a height of 5 cm.

a. What is the volume of the cylinder?

b. What is the volume of the cylinder when its radius is tripled?

c. What is the volume of the cylinder when its radius is halved?

(from Unit 5, Lesson 18)

NAME DATE PERIOD

Unit 6, Lesson 2: Plotting Data

Let's collect and display some data about the class.

2.1: Representing Data

Lin surveyed 30 students about the longest time they had ever run. Andre asked them about their favorite color. How could Lin and Andre represent their data sets? Would they represent them in the same way? Why or why not?

2.2: Gathering Data

Are older students always taller? Do taller students tend to have bigger hands? To investigate these questions, the class will gather data.

- A person's *arm span* is the distance between the tips of their index fingers, when their arms are fully spread out.
- A person's *hand span* is the distance from the tip of their thumb to the tip of their little finger, when their fingers are fully spread out.

1. Each partner should:
   - Measure the other partner's height, arm span, and hand span for their right hand to the nearest centimeter.
   - Record the other partner's measurements and age (in months) in the table.

	height (cm)	**arm span (cm)**	**hand span (cm)**	**age (months)**
partner A				
partner B				

NAME DATE PERIOD

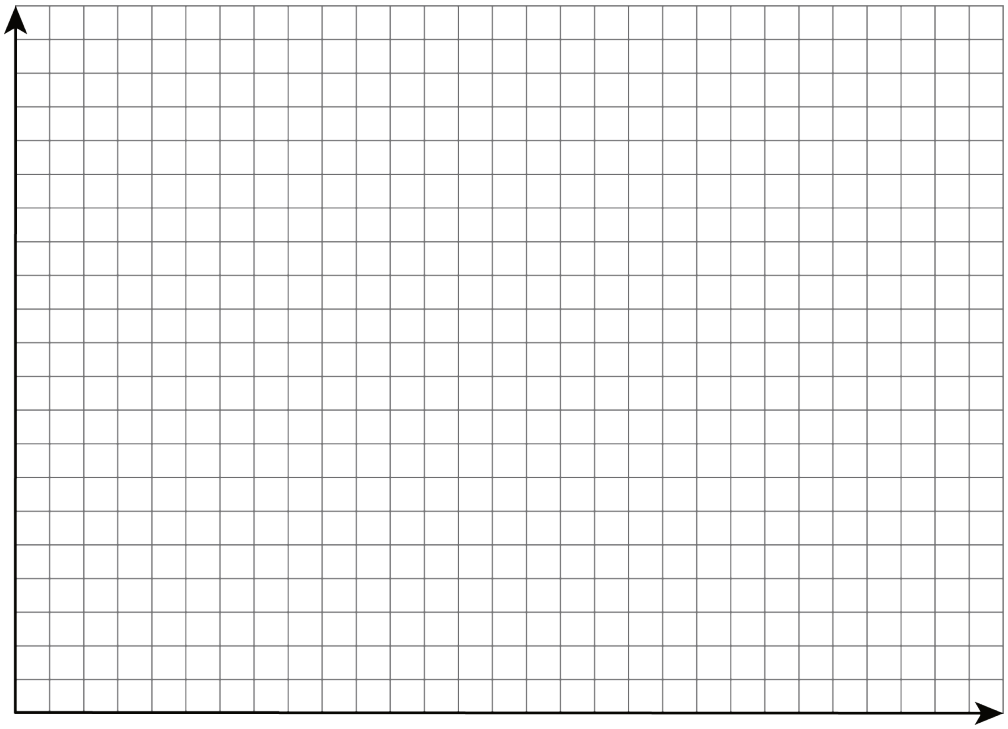
2. One partner records the data from your table in a table of data for the entire class.

2.3: Scatter Plots

m.openup.org/1/8-6-2-3

1. What types of graphical representations could be used to show the class's height measurements? Make a graphical representation of the class's height measurements.

2. Choose a color and use it to plot a point on the coordinate plane that represents your own height and hand span. Then, in the same color, plot a second point that represents your partner's height and hand span.



3. In a different color, plot the height and hand span of each student in your class, making a scatter plot of the heights and hand spans for the entire class.

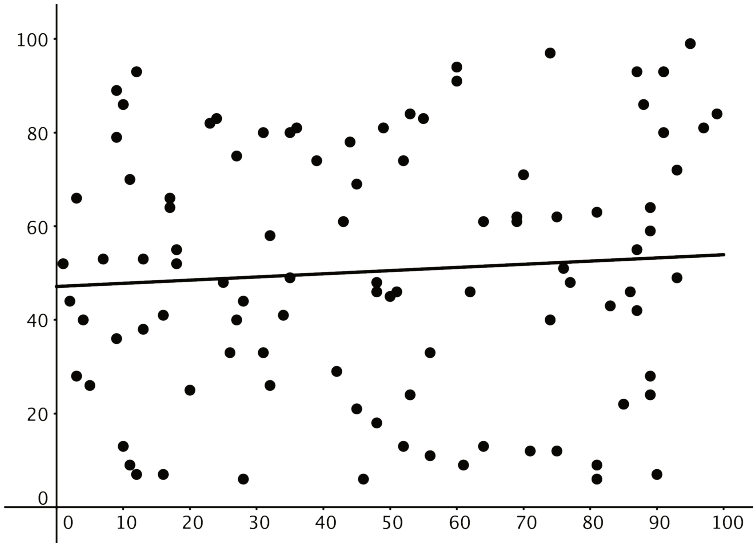
NAME DATE PERIOD

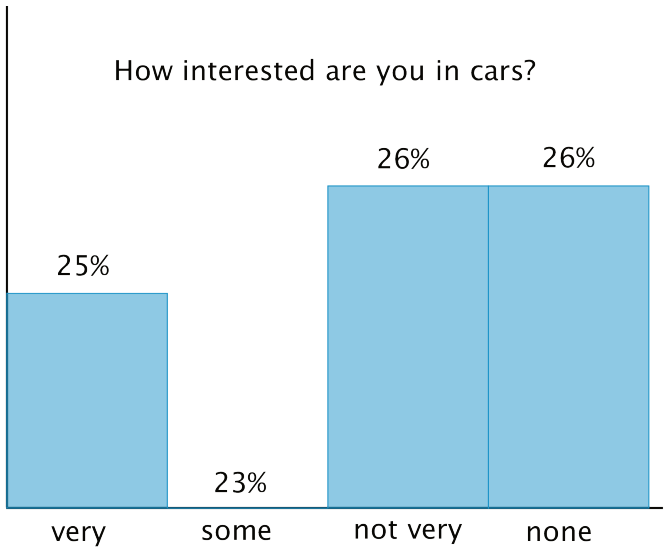
4. Based on your scatter plot, answer these questions:
   a. Do taller students in your class tend to have bigger hands? Explain how you know.

   b. Is hand span a linear function of height? Explain how you know.

Are you ready for more?

Although the data may be accurate, displaying the data incorrectly can tell the wrong story. What is wrong with each of these graphic representations of the data?



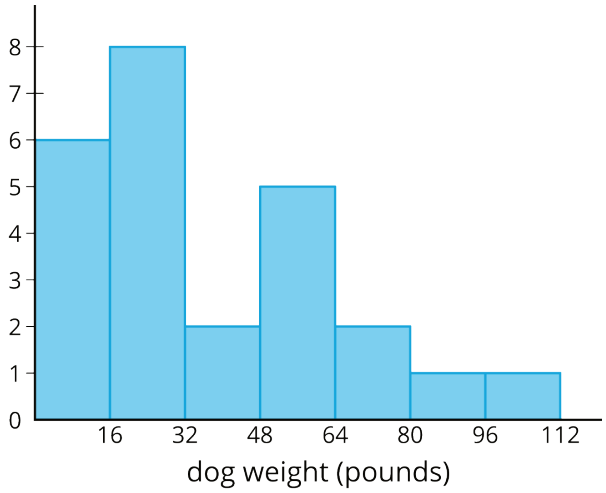


NAME DATE PERIOD

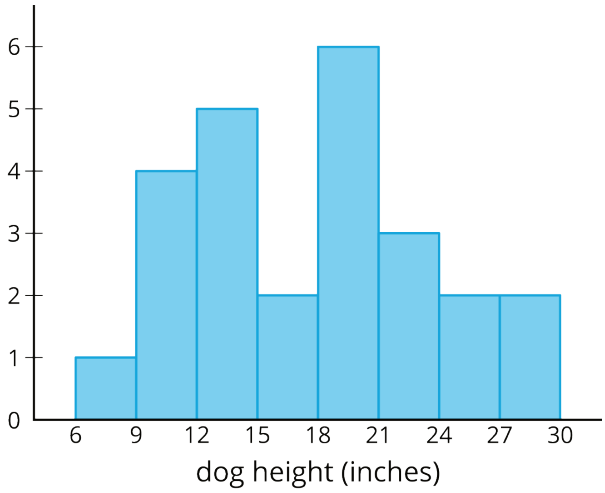
Lesson 2 Summary

Histograms show us how measurements of a single attribute are distributed. For example, a veterinarian saw 25 dogs in her clinic one week. She measured the height and weight of each dog.

This histogram shows how the weights of the dogs are distributed.

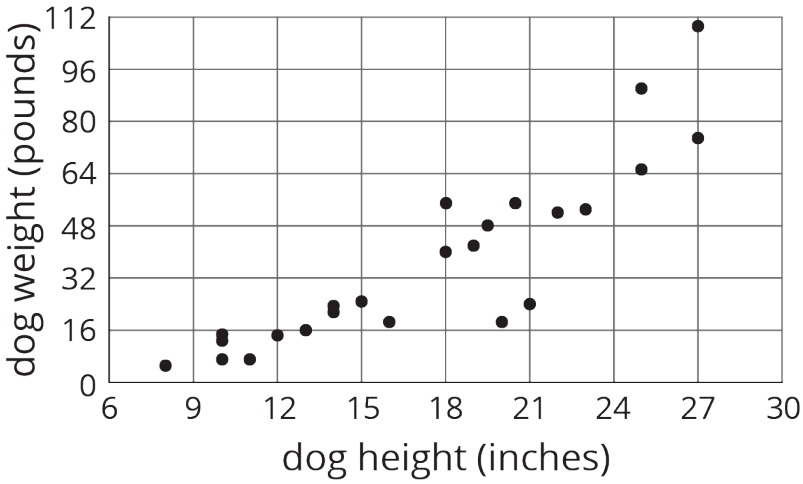


This histogram shows how the heights of the dogs are distributed.



These histograms tell us how the weights of the dogs and how the heights of dogs were distributed. But, they do not give any evidence of a connection between a dog's height and its weight.

Scatter plots allow us to investigate possible connections between two attributes. In this example, each plotted point corresponds to one of the 25 dogs, and its coordinates tell us the height and weight of that dog. Examination of the scatter plot allows us to see a connection between height and weight for the dogs.



NAME DATE PERIOD

Unit 6, Lesson 2: Plotting Data

1. In hockey, a player gets credited with a “point” in their statistics when they get an assist or goal. The table shows the number of assists and number of points for 15 hockey players after a season.

assists	points
22	28
16	18
46	72
19	29
13	26
9	13
16	22
8	18
12	13
12	17
37	50
7	12
17	34
27	58
18	34

Make a scatter plot of this data. Make sure to scale and label the axes.

2. Select **all** the representations that are appropriate for comparing bite strength to weight for different

NAME DATE PERIOD

carnivores.

A. Histogram
B. Scatter plot
C. Dot plot
D. Table
E. Box plot

3. When is it better to use a table? When is it better to use a scatter plot?

4. There are many cylinders with radius 6 meters. Let h represent the height in meters and V represent the volume in cubic meters.

   a. Write an equation that represents the volume V as a function of the height h.

   b. Sketch the graph of the function, using 3.14 as an approximation for π.

   c. If you double the height of a cylinder, what happens to the volume? Explain this using the equation.

   d. If you multiply the height of a cylinder by $\frac{1}{3}$, what happens to the volume? Explain this using the graph.

(from Unit 5, Lesson 17)

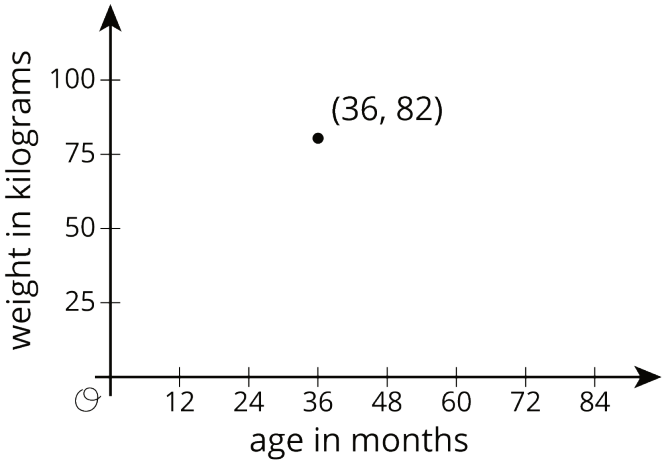
NAME DATE PERIOD

Unit 6, Lesson 3: What a Point in a Scatter Plot Means

Let's investigate points in scatter plots.

3.1: The Giant Panda

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



NAME DATE PERIOD

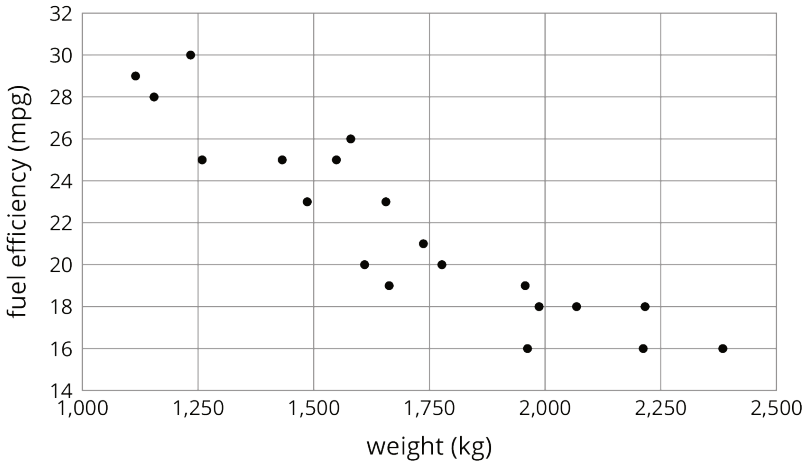
3.2: Weight and Fuel Efficiency

m.openup.org/1/8-6-3-2

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

NAME DATE PERIOD

car	weight (kg)	fuel efficiency (mpg)
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?

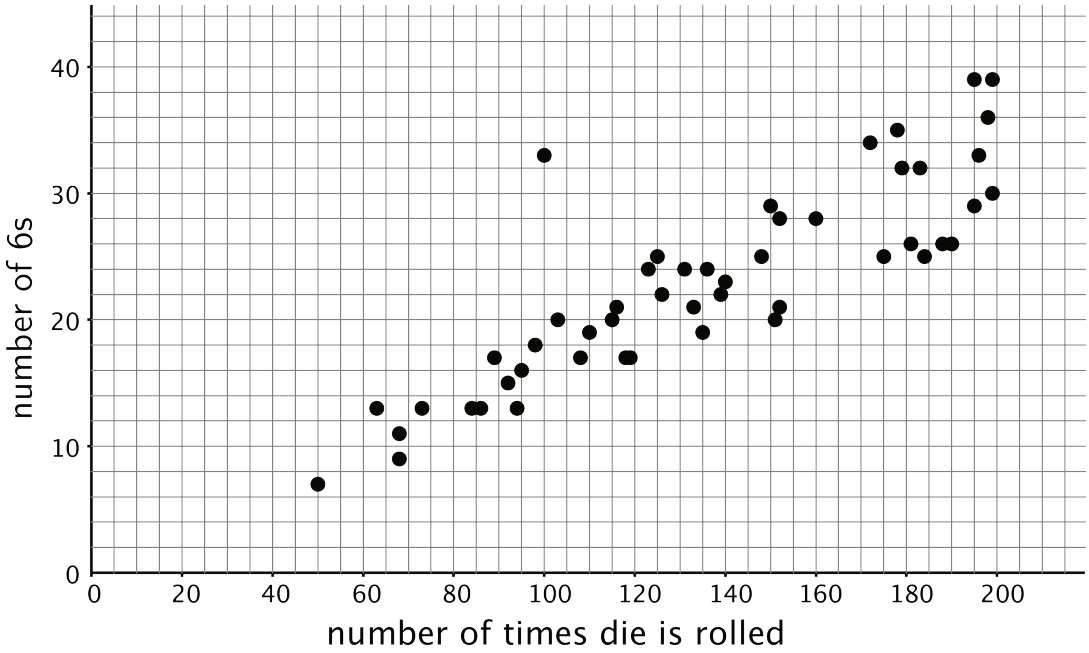
6. A different company makes Cars F and G. Compare their weights and fuel efficiencies. Does anything surprise you about these cars?

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

NAME DATE PERIOD

during those rolls.



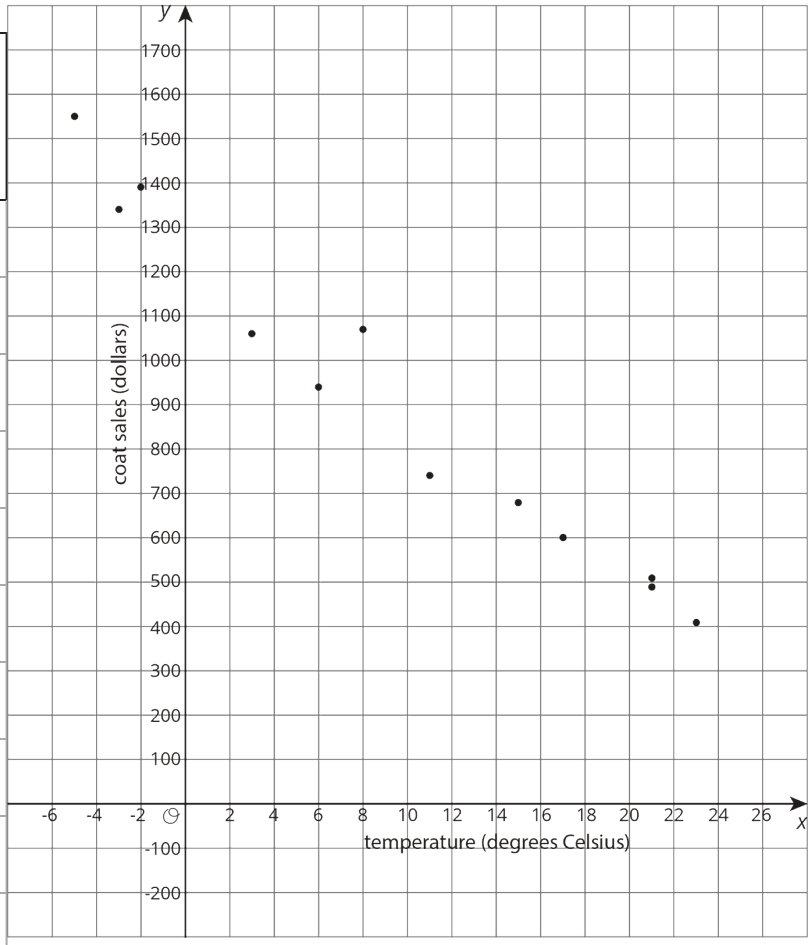
Select a point in the scatter plot and give its approximate coordinates, then tell the story of that point in the context of the problem.

NAME DATE PERIOD

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.

NAME DATE PERIOD

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.

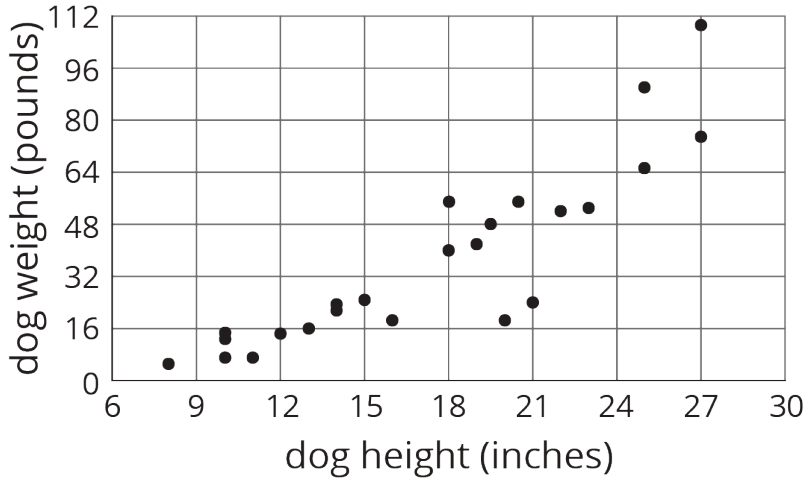
5. What would a point at $(B, 0)$ represent? Use the scatter plot to estimate a value for B.

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.

NAME DATE PERIOD

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.

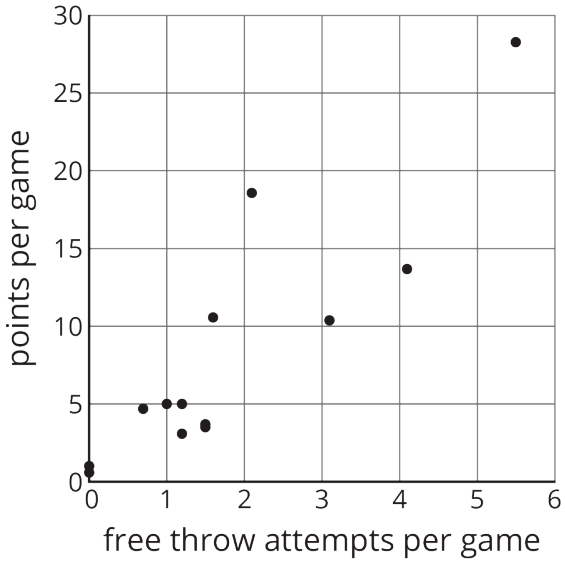
NAME DATE PERIOD

Unit 6, Lesson 3: What a Point in a Scatter Plot Means

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

NAME DATE PERIOD



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

NAME DATE PERIOD

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

(from Unit 5, Lesson 17)

NAME DATE PERIOD

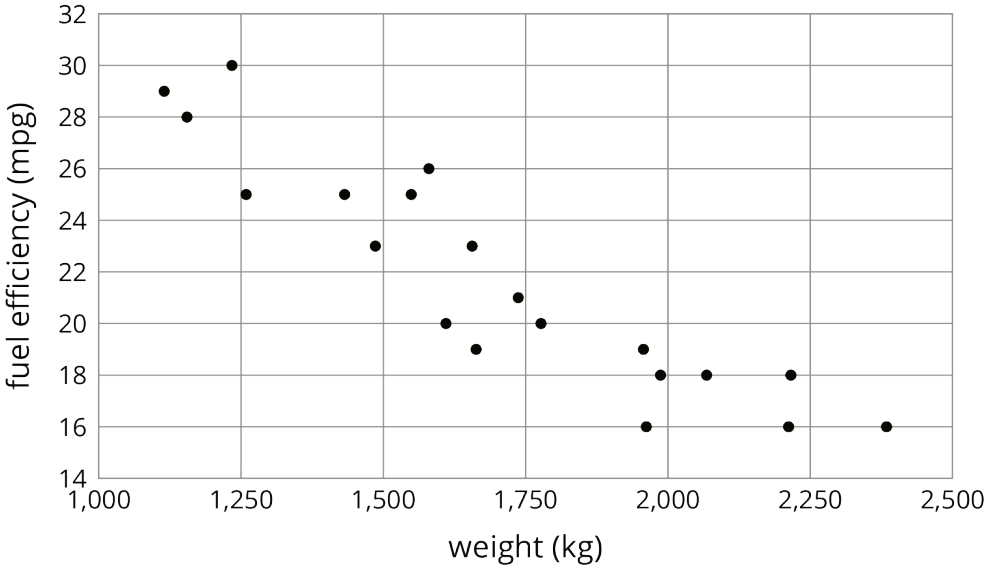
Unit 6, Lesson 4: Fitting a Line to Data

Let's look at the scatter plots as a whole.

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.



NAME DATE PERIOD

4.2: Shine Bright

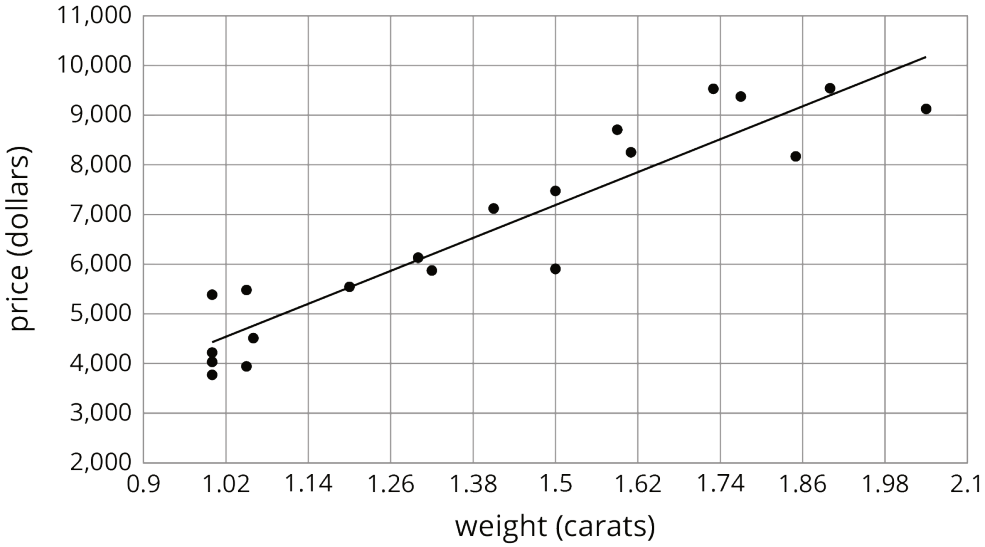
m.openup.org/1/8-6-4-2

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

NAME DATE PERIOD

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?

NAME DATE PERIOD

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?

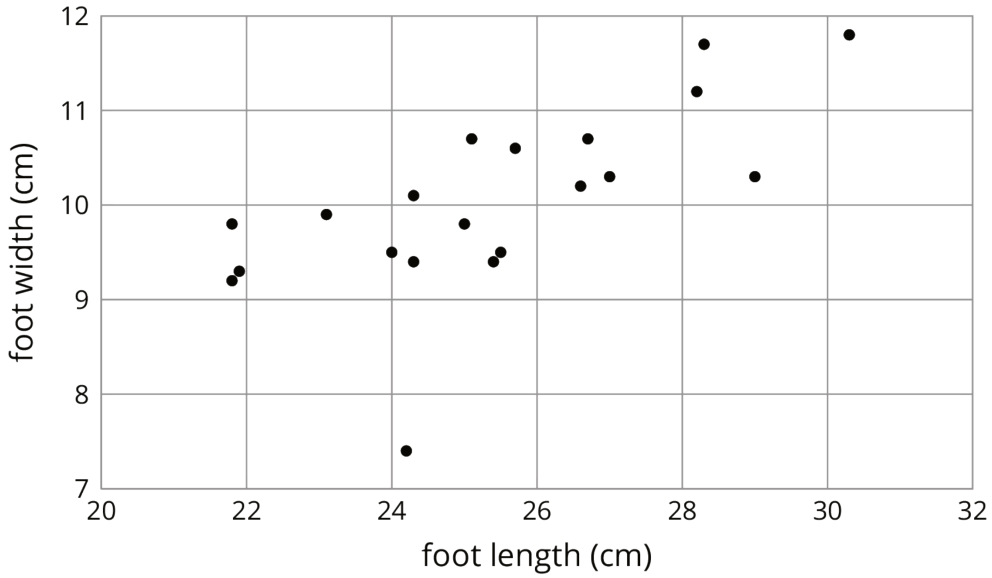
4. Find a diamond for which the model makes a very good prediction of the actual price. How can you see this in the table? In the graph?

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

m.openup.org/1/8-6-4-3

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

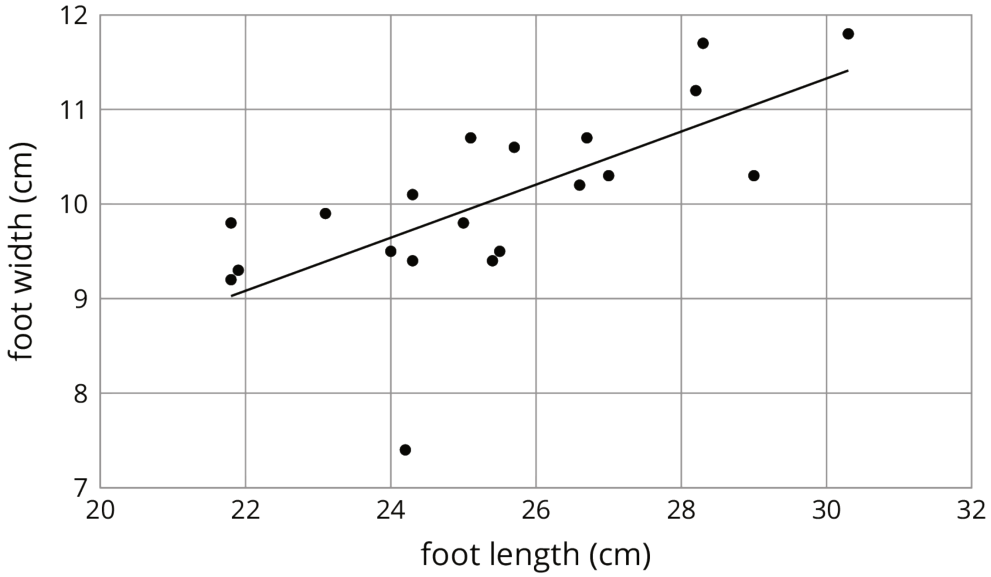


NAME DATE PERIOD

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

2. Estimate the widths 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.

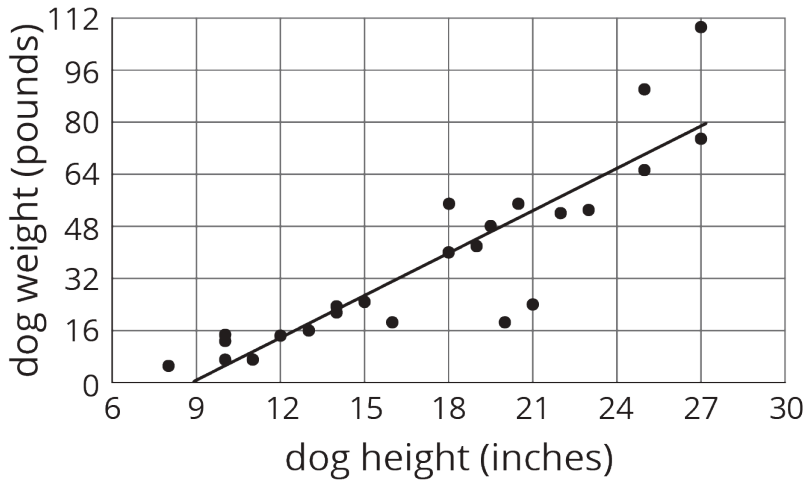


Circle the data point that seems weird when compared to the model. What length and width does that point represent?

NAME DATE PERIOD

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 Glossary Terms

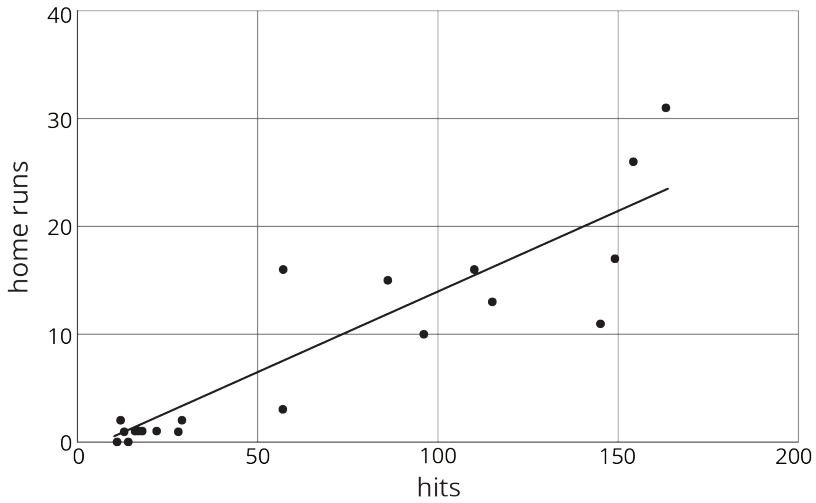
- outlier

NAME DATE PERIOD

Unit 6, Lesson 4: Fitting a Line to Data

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?

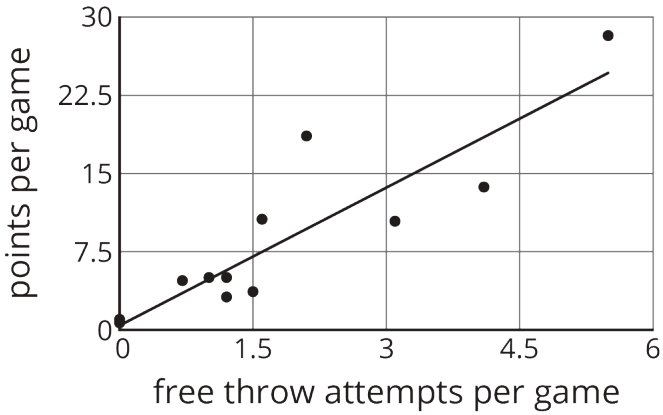
   b. Player B was the player who most outperformed the prediction. How many hits did Player B have last season?

   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

NAME DATE PERIOD

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?

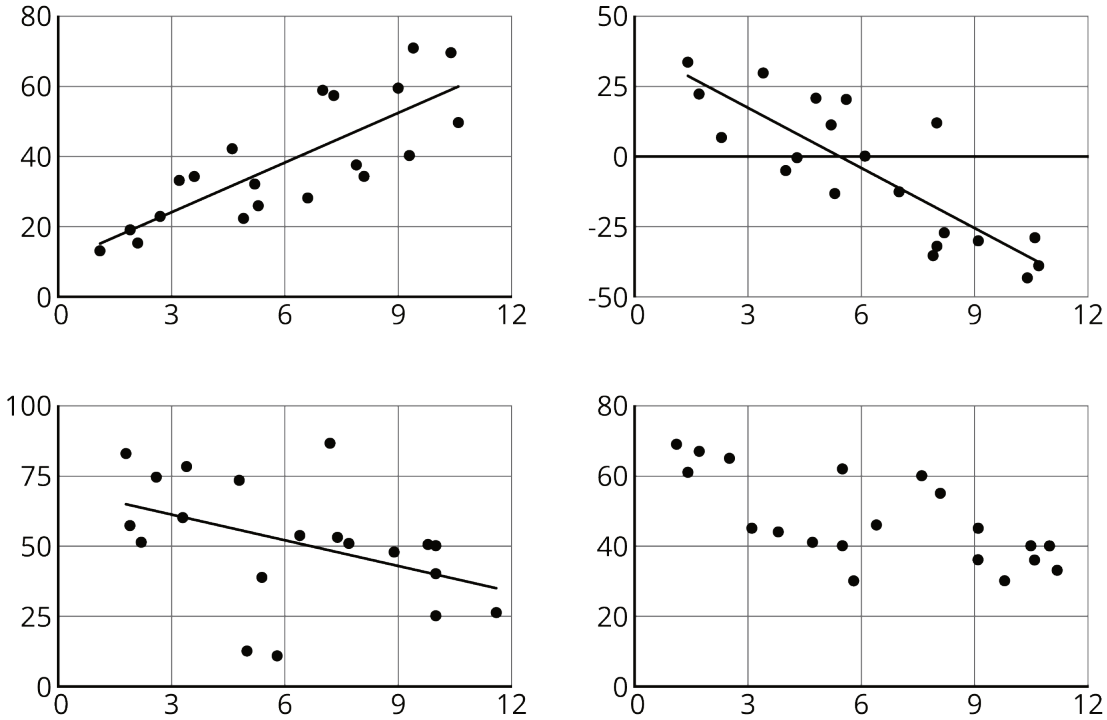
NAME DATE PERIOD

Unit 6, Lesson 5: Describing Trends in Scatter Plots

Let's look for associations between variables.

5.1: Which One Doesn't Belong: Scatter Plots

Which one doesn't belong?



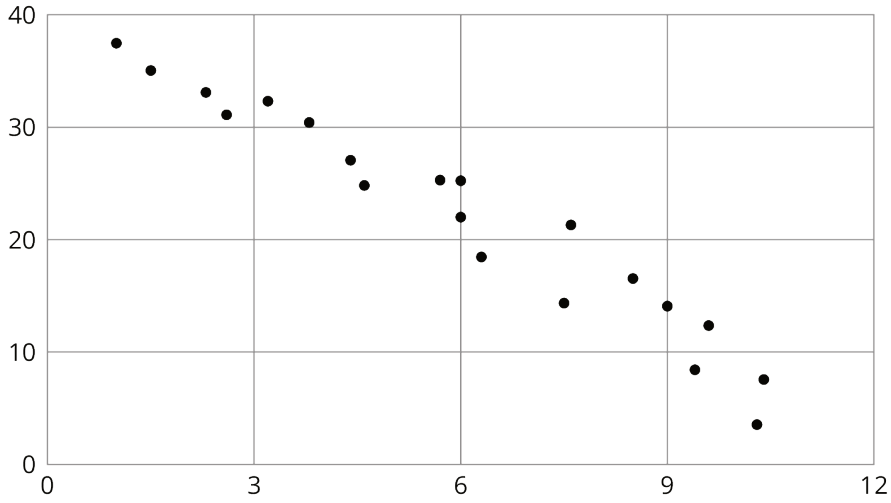
NAME DATE PERIOD

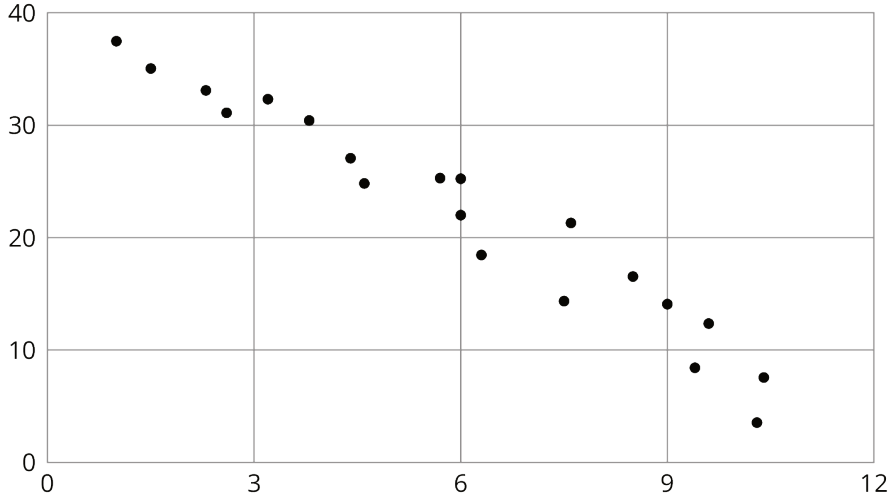
5.2: Fitting Lines

m.openup.org/1/8-6-5-2

Your teacher will give you a piece of pasta and a straightedge.

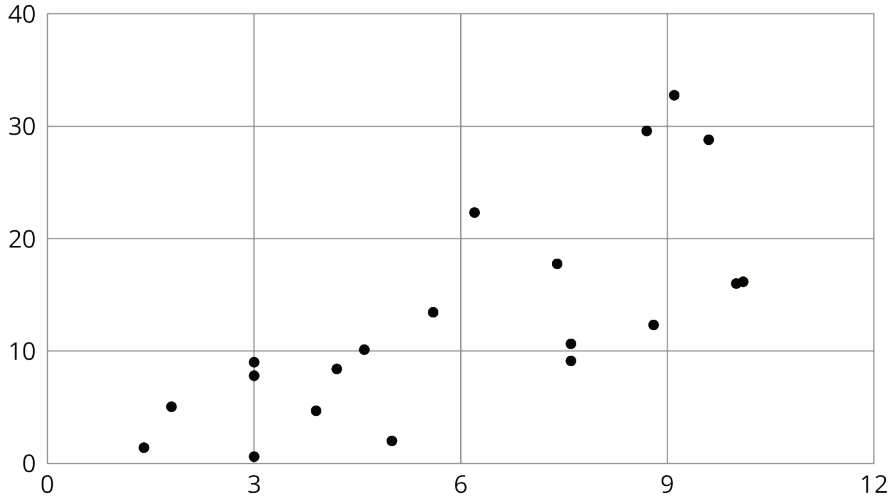
1. Here are two copies of the same scatter plot. Experiment with drawing lines to fit the data. Pick the line that you think best fits the data. Compare it with a partner's.

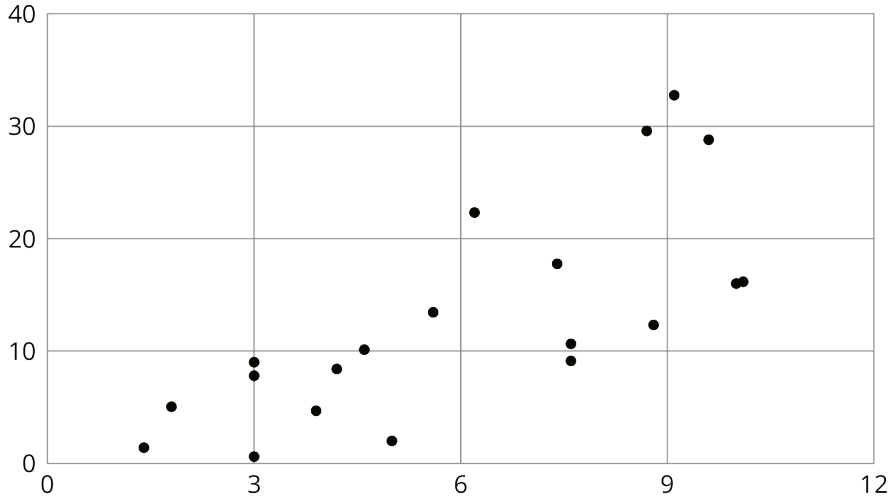




NAME DATE PERIOD

2. Here are two copies of another scatter plot. Experiment with drawing lines to fit the data. Pick the line that you think best fits the data. Compare it with a partner's.



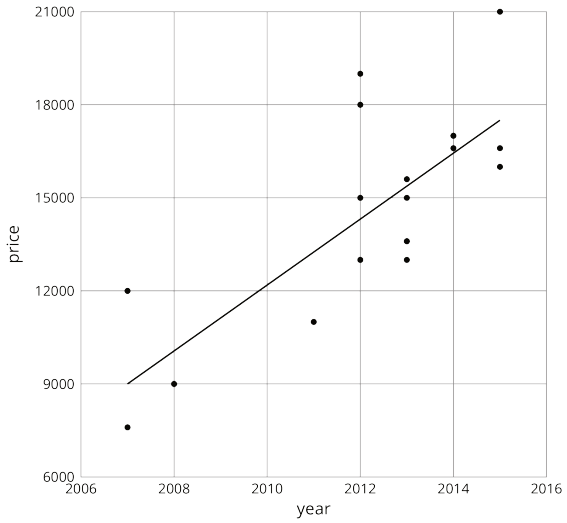


3. In your own words, describe what makes a line fit a data set well.

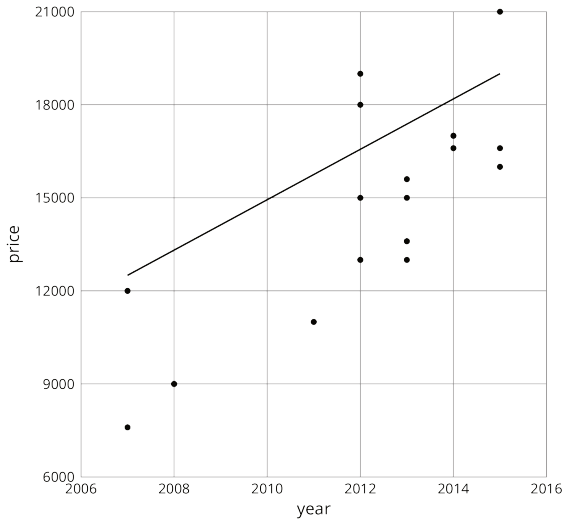
NAME DATE PERIOD

5.3: Good Fit Bad Fit

The scatter plots both show the year and price for the same 17 used cars. However, each scatter plot shows a different model for the relationship between year and price.



A



B

1. Look at Diagram A.
   a. For how many cars does the model in Diagram A make a good prediction of its price?

   b. For how many cars does the model underestimate the price?

   c. For how many cars does it overestimate the price?

2. Look at Diagram B.
   a. For how many cars does the model in Diagram B make a good prediction of its price?

   b. For how many cars does the model underestimate the price?

   c. For how many cars does it overestimate the price?

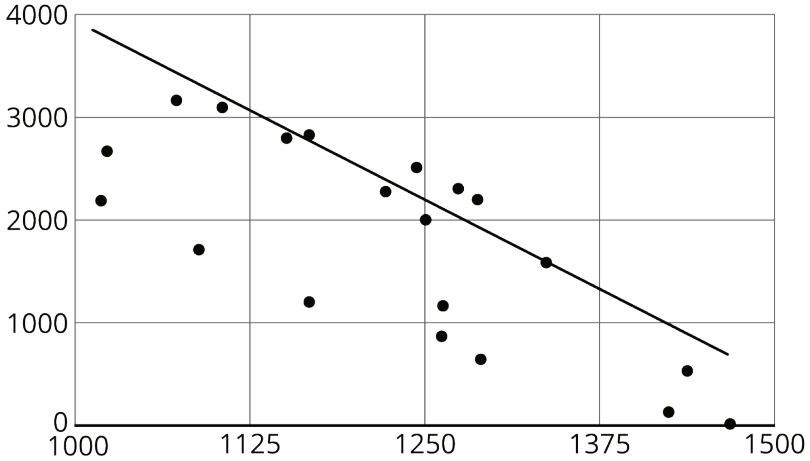
3. For how many cars does the prediction made by the model in Diagram A differ by more than $3,000? What about the model in Diagram B?

NAME DATE PERIOD

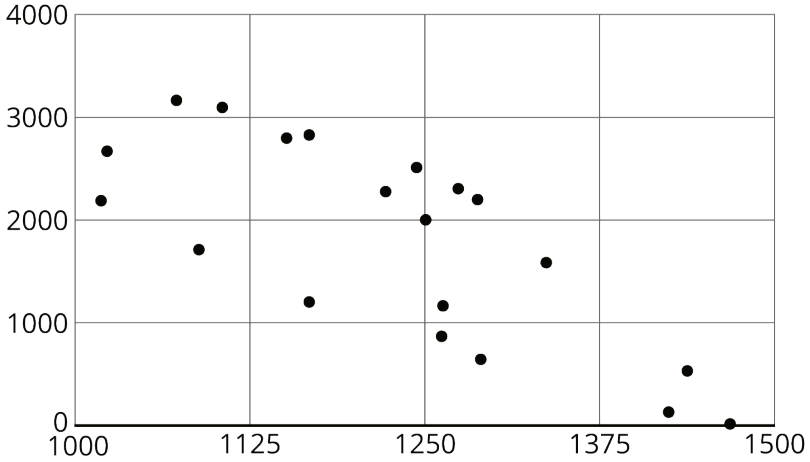
4. Which model does a better job of predicting the price of a used car from its year?

5.4: Practice Fitting Lines

1. Is this line a good fit for the data? Explain your reasoning.

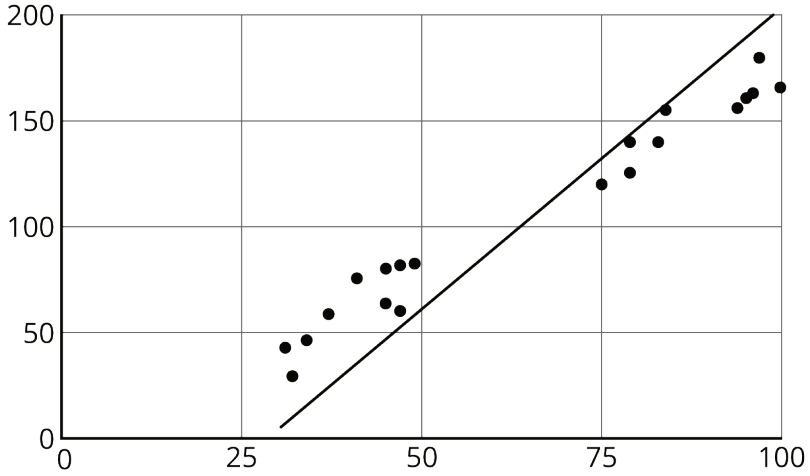


2. Draw a line that fits the data better.

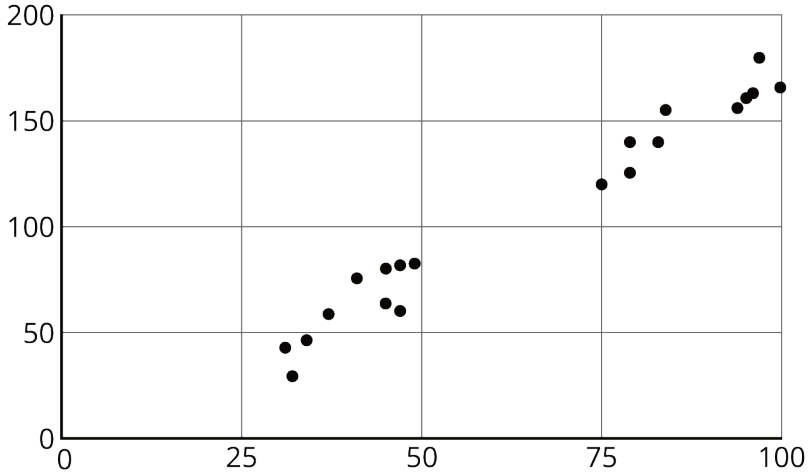


NAME DATE PERIOD

3. Is this line a good fit for the data? Explain your reasoning.

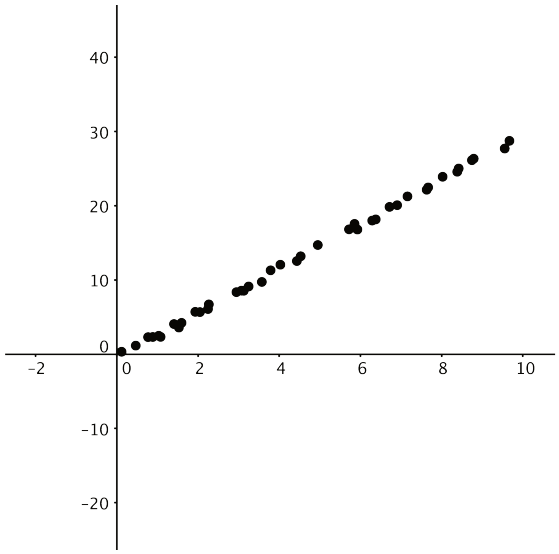


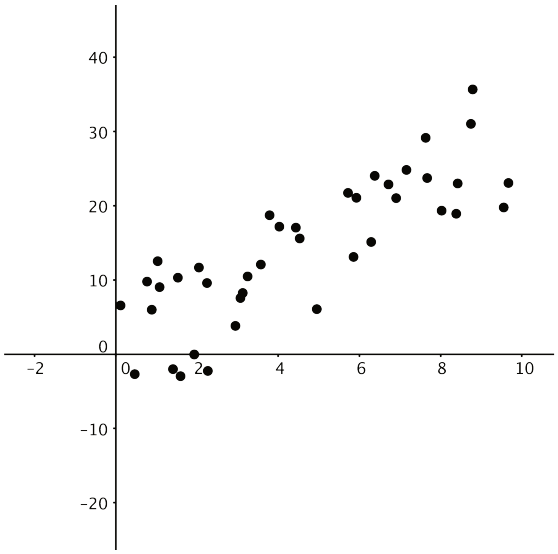
4. Draw a line that fits the data better.

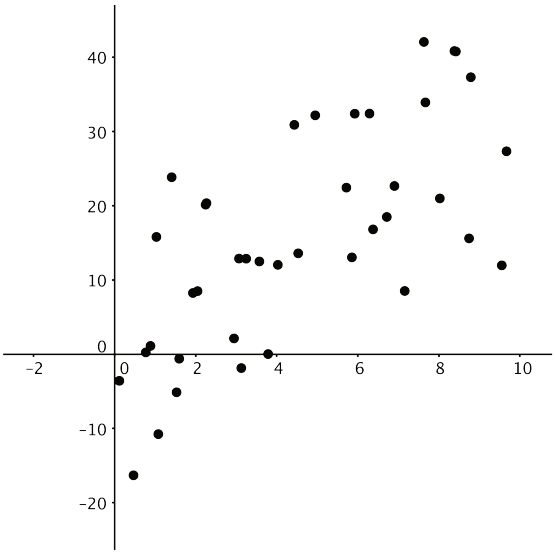


NAME DATE PERIOD

Are you ready for more?







These scatter plots were created by multiplying the x-coordinate by 3 then adding a random number between two values to get the y-coordinate. The first scatter plot added a random number between -0.5 and 0.5 to the y-coordinate. The second scatter plot added a random number between -2 and 2 to the y-coordinate. The third scatter plot added a random number between -10 and 10 to the y-coordinate.

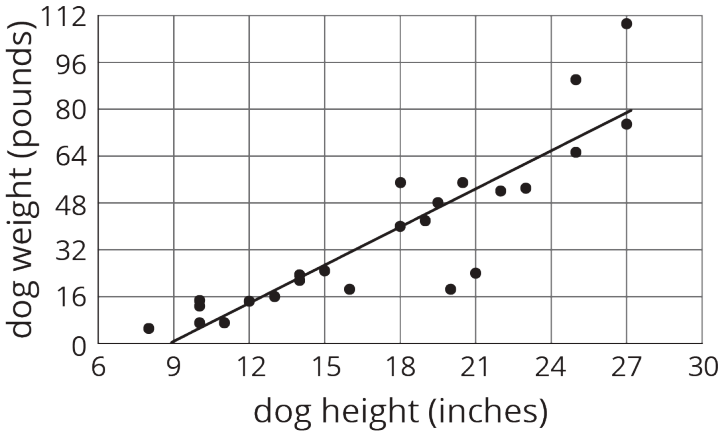
1. For each scatter plot, draw a line that fits the data.

2. Explain why some were easier to do than others.

NAME DATE PERIOD

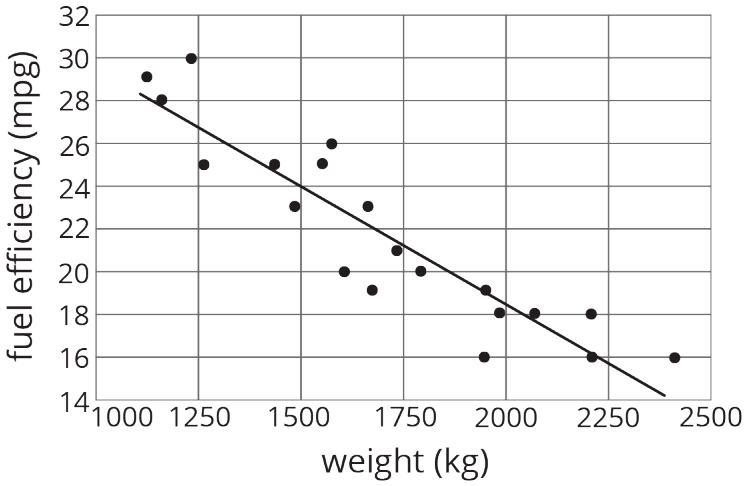
Lesson 5 Summary

When a linear function fits data well, we say there is a *linear association* between the variables. For example, the relationship between height and weight for 25 dogs with the linear function whose graph is shown in the scatter plot.



Because the model fits the data well and because the slope of the line is positive, we say that there is a **positive association** between dog height and dog weight.

What do you think the association between the weight of a car and its fuel efficiency is?



Because the slope of a line that fits the data well is negative, we say that there is a **negative association** between the fuel efficiency and weight of a car.

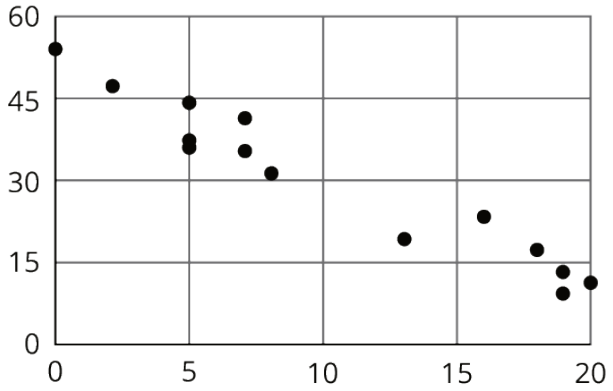
Lesson 5 Glossary Terms

- negative association
- positive association

NAME DATE PERIOD

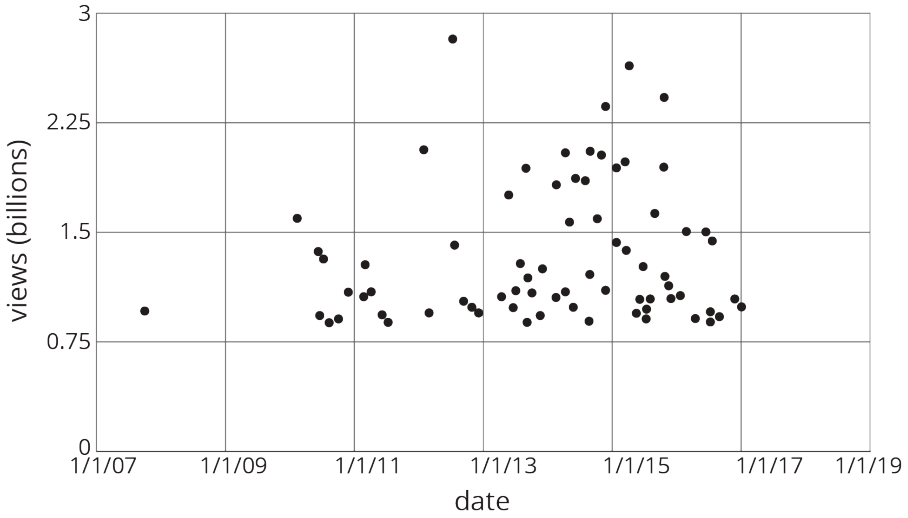
Unit 6, Lesson 5: Describing Trends in Scatter Plots

1. a. Draw a line that you think is a good fit for this data. For this data, the inputs are the horizontal values, and the outputs are the vertical values.



b. Use your line of fit to estimate what you would expect the output value to be when the input is 10.

2. Here is a scatter plot that shows the most popular videos in a 10-year span.



a. Use the scatter plot to estimate the number of views for the most popular video in this 10-year span.

b. Estimate when the 4th most popular video was released.

(from Unit 6, Lesson 3)

NAME DATE PERIOD

3. A recipe for bread calls for 1 teaspoon of yeast for every 2 cups of flour.

   a. Name two quantities in this situation that are in a functional relationship.

   b. Write an equation that represents the function.

   c. Draw the graph of the function. Label at least two points with input-output pairs.

(from Unit 5, Lesson 8)

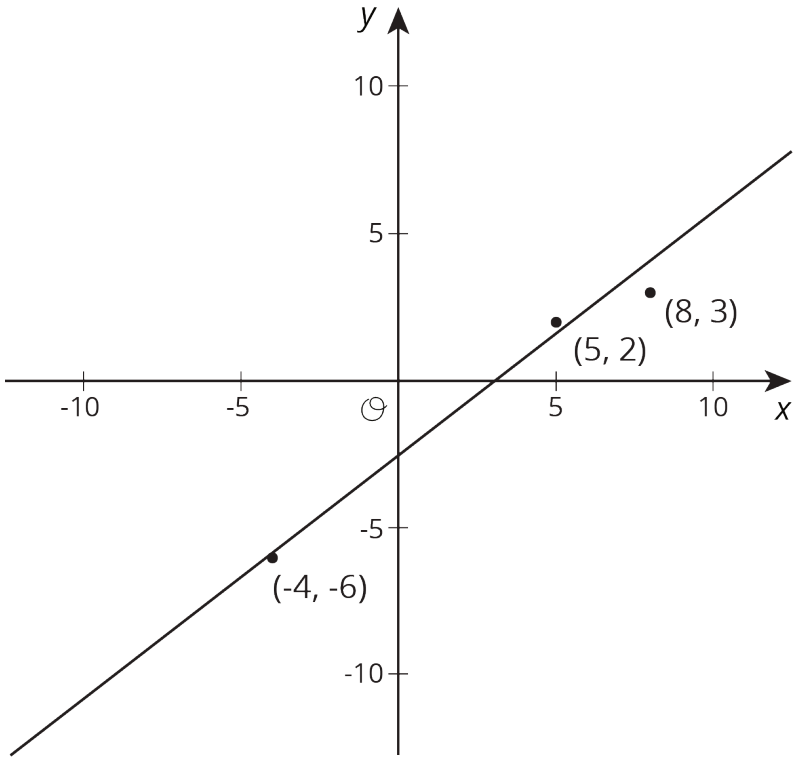
NAME DATE PERIOD

Unit 6, Lesson 6: The Slope of a Fitted Line

Let's look at how changing one variable changes another.

6.1: Estimating Slope

Estimate the slope of the line.

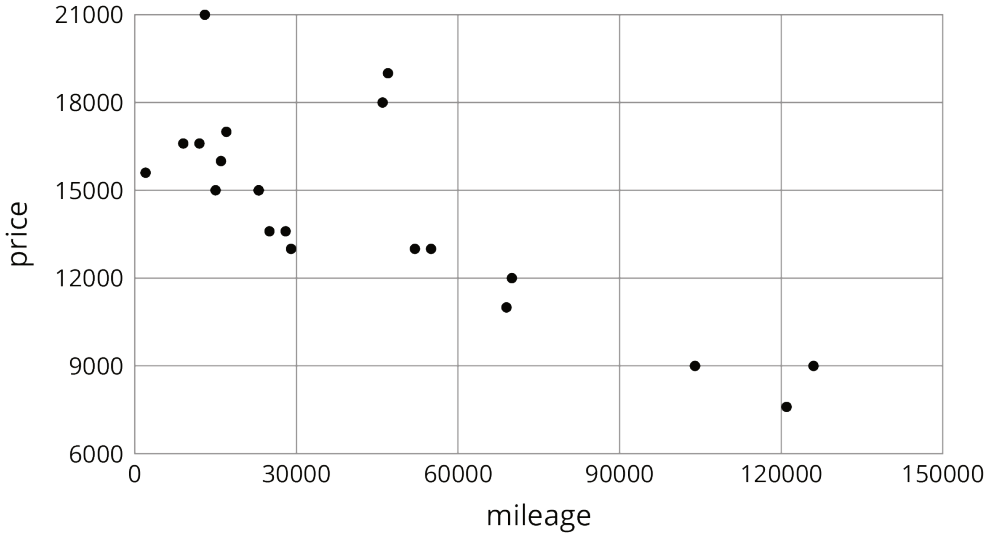


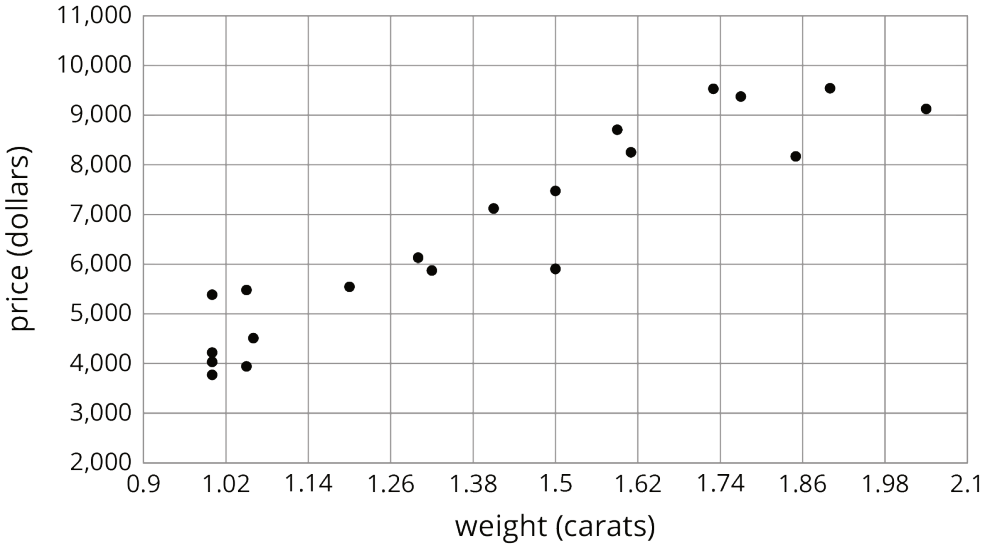
6.2: Describing Linear Associations

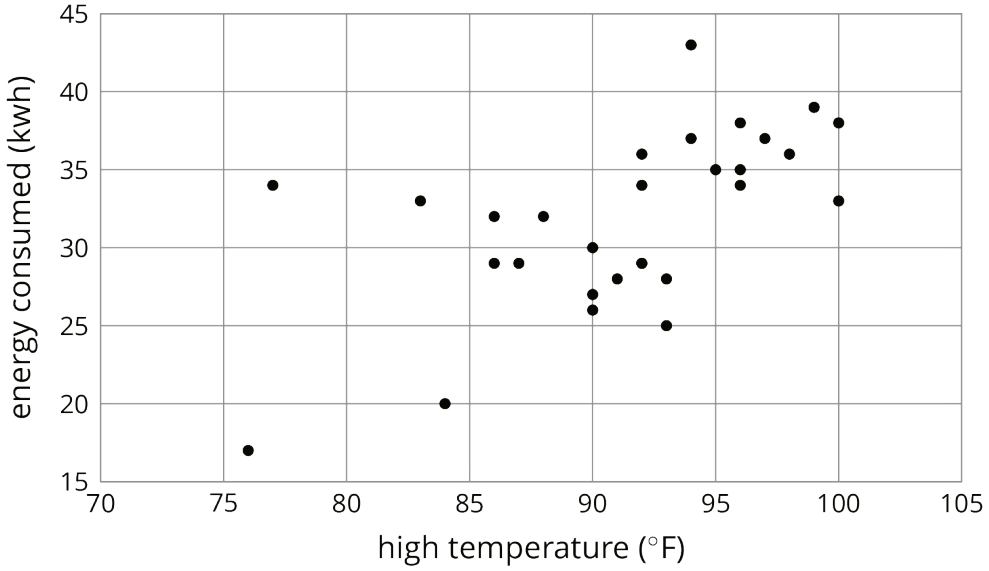
For each scatter plot, decide if there is an association between the two variables, and describe the situation using one of these sentences:

- For these data, as _____ increases, _____ tends to increase.
- For these data, as _____ increases, _____ tends to decrease.
- For these data, _____ and _____ do not appear to be related.

NAME DATE PERIOD







NAME DATE PERIOD

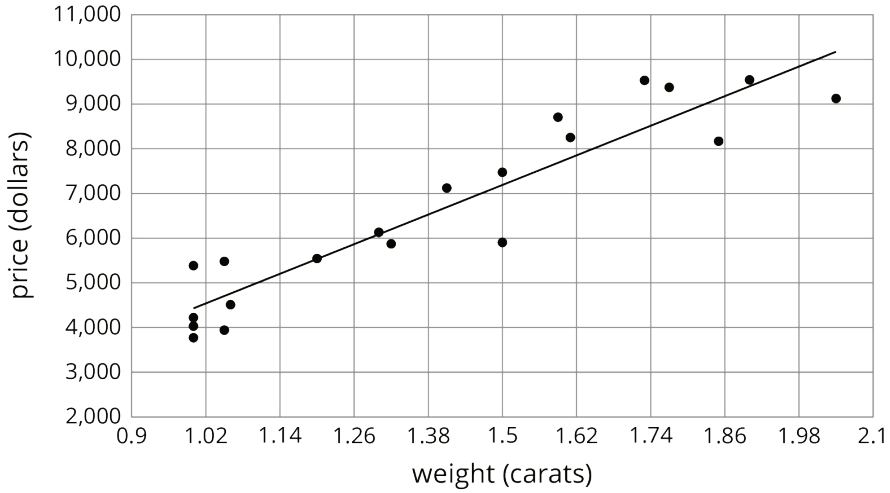
6.3: Interpreting Slopes

For each of the situations, a linear model for some data is shown.

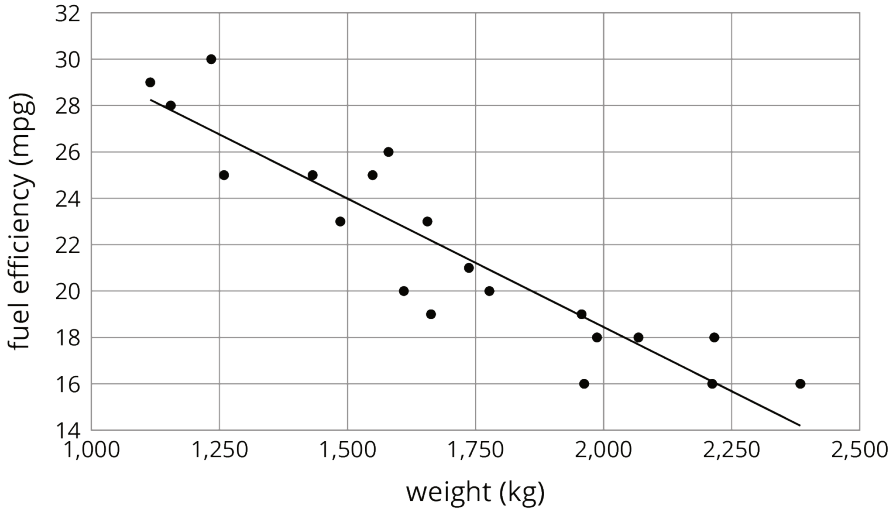
1. What is the slope of the line in the scatter plot for each situation?

2. What is the meaning of the slope in that situation?

$y = 5{,}520.619x - 1{,}091.393$

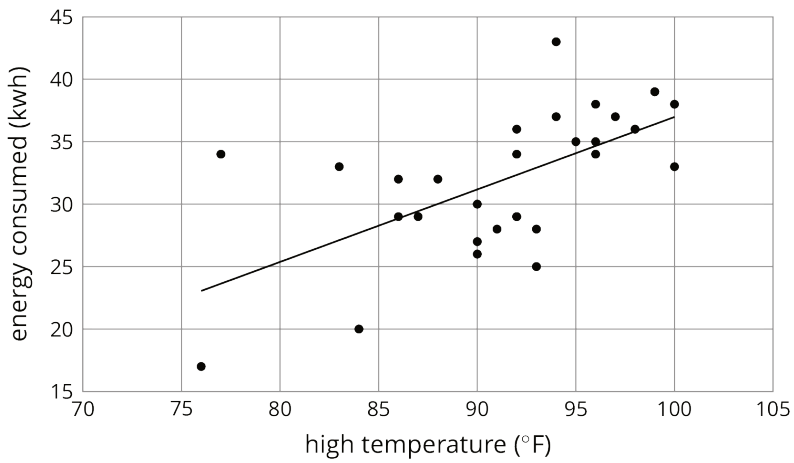


$y = -0.011x + 40.604$



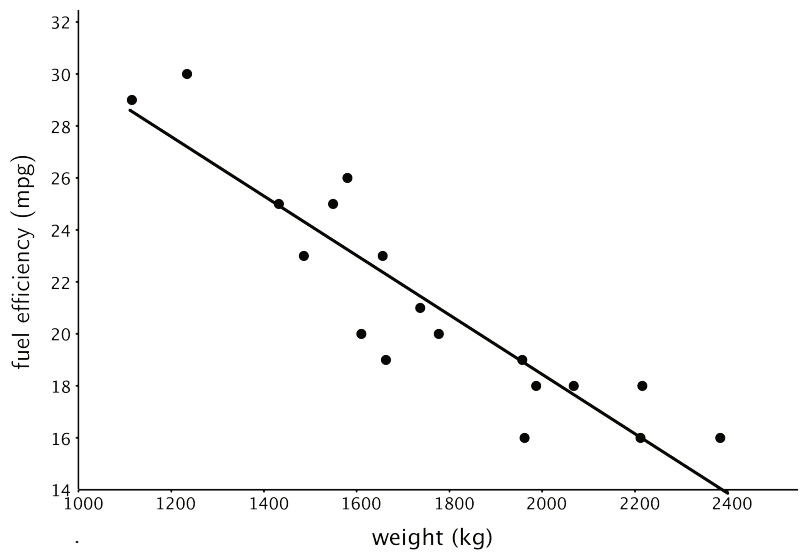
NAME DATE PERIOD

$y = 0.59x - 21.912$



Are you ready for more?

The scatter plot shows the weight and fuel efficiency data used in an earlier lesson along with a linear model represented by the equation $y = -0.0114x + 41.3021$.



1. What is the value of the slope and what does it mean in this context?

2. What does the other number in the equation represent on the graph? What does it mean in context?

NAME DATE PERIOD

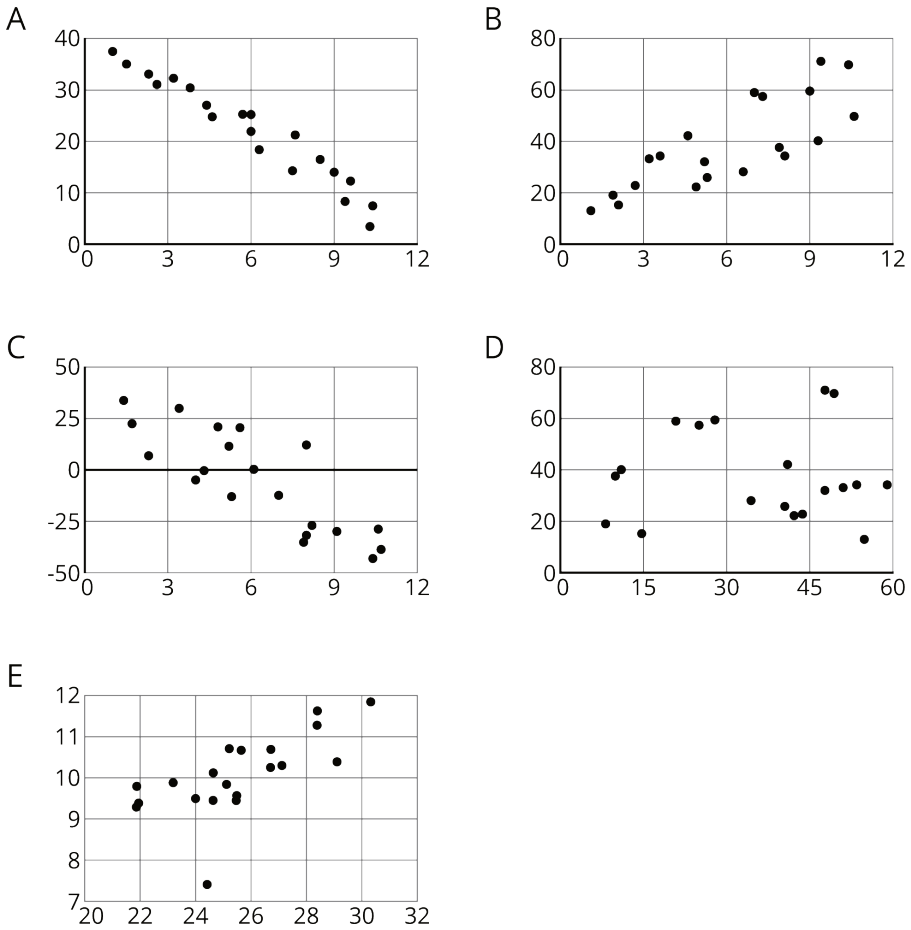
3. Use the equation to predict the fuel efficiency of a car that weighs 100 kilograms.

4. Use the equation to predict the weight of a car that has a fuel efficiency of 22 mpg.

5. Which of these two predictions probably fits reality better? Explain.

6.4: Positive or Negative?

1. For each of the scatter plots, decide whether it makes sense to fit a linear model to the data. If it does, would the graph of the model have a positive slope, a negative slope, or a slope of zero?

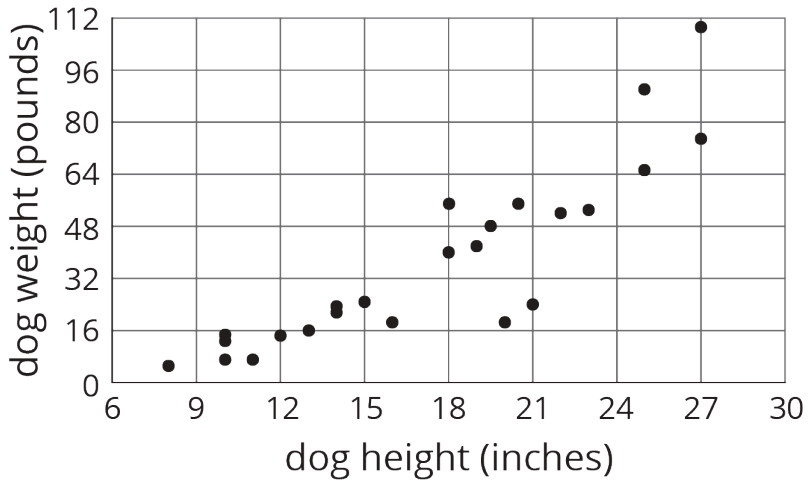


2. Which of these scatter plots show evidence of a positive association between the variables? Of a negative association? Which do not appear to show an association?

NAME DATE PERIOD

Lesson 6 Summary

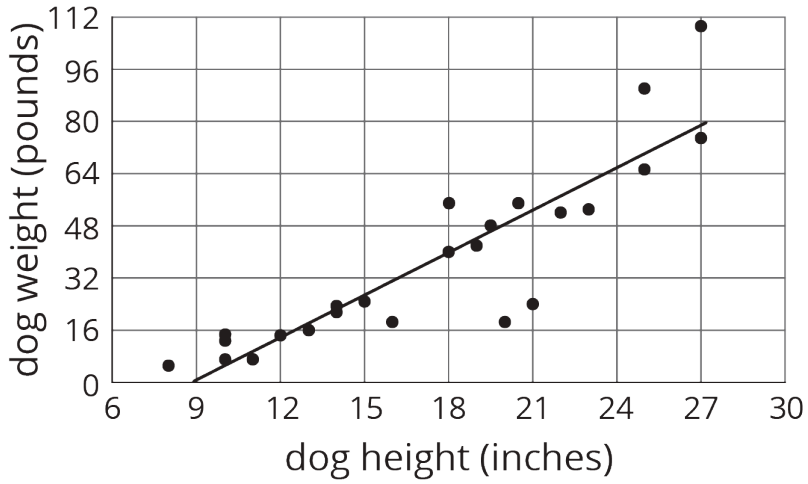
Here is a scatter plot that we have seen before. As noted earlier, we can see from the scatter plot that taller dogs tend to weigh more than shorter dogs. Another way to say it is that weight tends to increase as height increases. When we have a positive association between two variables, an increase in one means there tends to be an increase in the other.



We can quantify this tendency by fitting a line to the data and finding its slope. For example, the equation of the fitted line is

$$w = 4.27h - 37$$

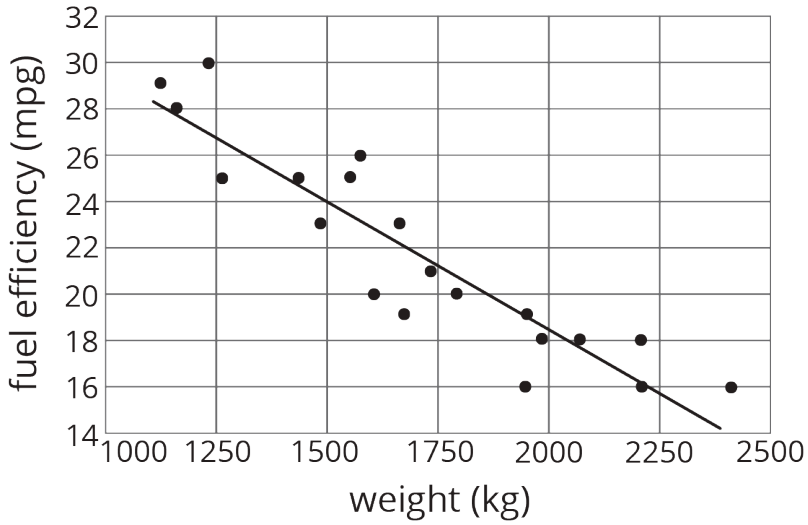
where h is the height of the dog and w is the predicted weight of the dog.



The slope is 4.27, which tells us that for every 1-inch increase in dog height, the weight is predicted to increase by 4.27 pounds.

NAME DATE PERIOD

In our example of the fuel efficiency and weight of a car, the slope of the fitted line shown is -0.01.

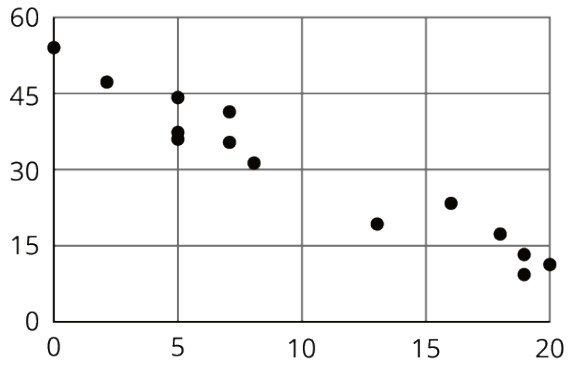


This tells us that for every 1-kilogram increase in the weight of the car, the fuel efficiency is predicted to decrease by 0.01 miles per gallon. When we have a negative association between two variables, an increase in one means there tends to be a decrease in the other.

NAME DATE PERIOD

Unit 6, Lesson 6: The Slope of a Fitted Line

1. Which of these statements is true about the data in the scatter plot?



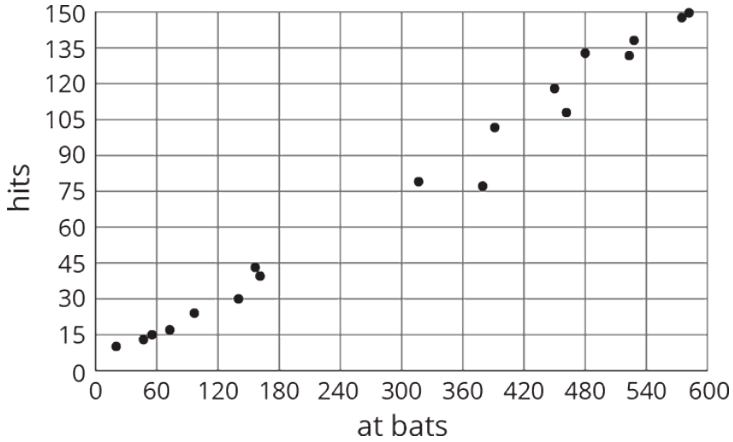
A. As x increases, y tends to increase.

B. As x increases, y tends to decrease.

C. As x increases, y tends to stay unchanged.

D. x and y are unrelated.

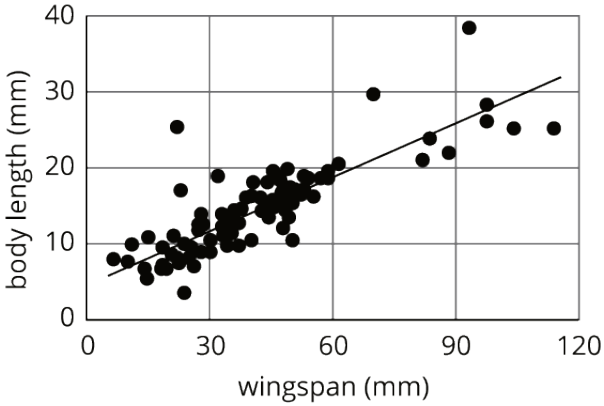
2. Here is a scatter plot that compares hits to at bats for players on a baseball team.



Describe the relationship between the number of at bats and the number of hits using the data in the scatter plot.

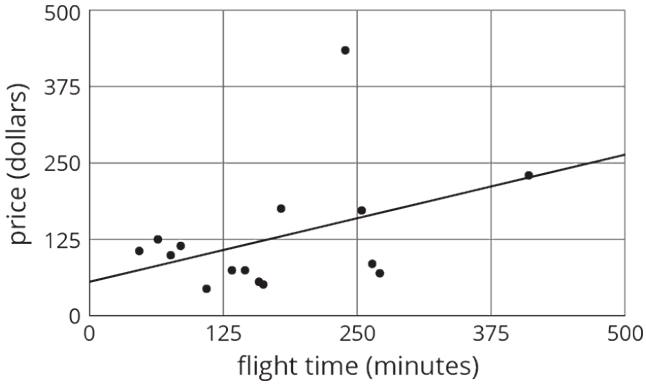
3. The linear model for some butterfly data is given by the equation $y = 0.238x + 4.642$. Which of the following best describes the slope of the model?

NAME DATE PERIOD



A. For every 1 mm the wingspan increases, the length of the butterfly increases 0.238 mm.
B. For every 1 mm the wingspan increases, the length of the butterfly increases 4.642 mm.
C. For every 1 mm the length of the butterfly increases, the wingspan increases 0.238 mm.
D. For every 1 mm the length of the butterfly increases, the wingspan increases 4.642 mm.

4. Nonstop, one-way flight times from O'Hare Airport in Chicago and prices of a one-way ticket are shown in the scatter plot.



a. Circle any data that appear to be outliers.

b. Use the graph to estimate the difference between any outliers and their predicted values.

(from Unit 6, Lesson 4)

5. Solve: $\begin{cases} y = -3x + 13 \\ y = -2x + 1 \end{cases}$

(from Unit 4, Lesson 14)

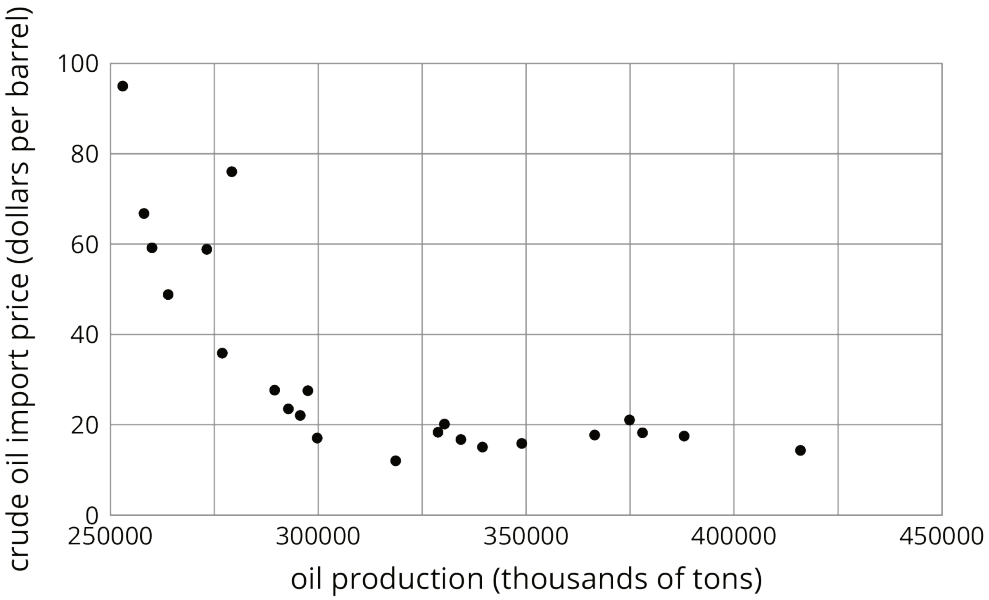
NAME DATE PERIOD

Unit 6, Lesson 7: Observing More Patterns in Scatter Plots

Let's look for other patterns in data.

7.1: Notice and Wonder: Nonlinear Scatter Plot

What do you notice? What do you wonder?



7.2: Scatter Plot City

Your teacher will give you a set of cards. Each card shows a scatter plot.

1. Sort the cards into categories and describe each category.

2. Explain the reasoning behind your categories to your partner. Listen to your partner's reasoning for their categories.

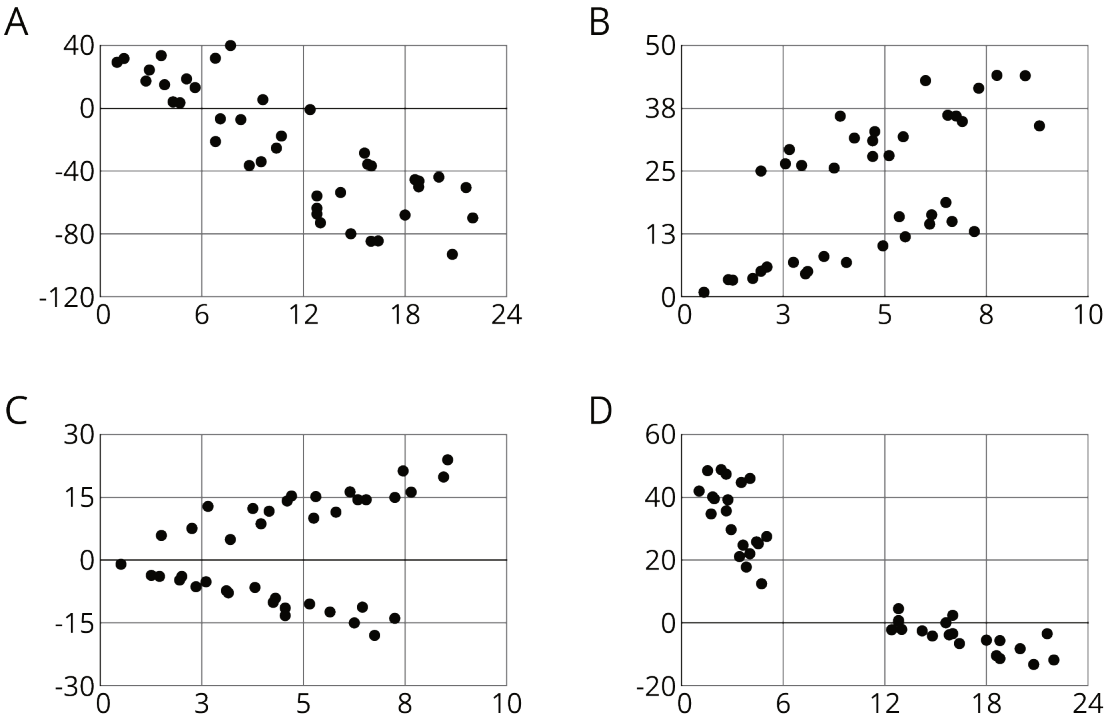
3. Sort the cards into two categories: positive associations and negative associations. Compare your sorting with your partner's and discuss any disagreements.

4. Sort the cards into two categories: linear associations and non-linear associations. Compare your sorting with your partner's and discuss any disagreements.

NAME DATE PERIOD

7.3: Clustering

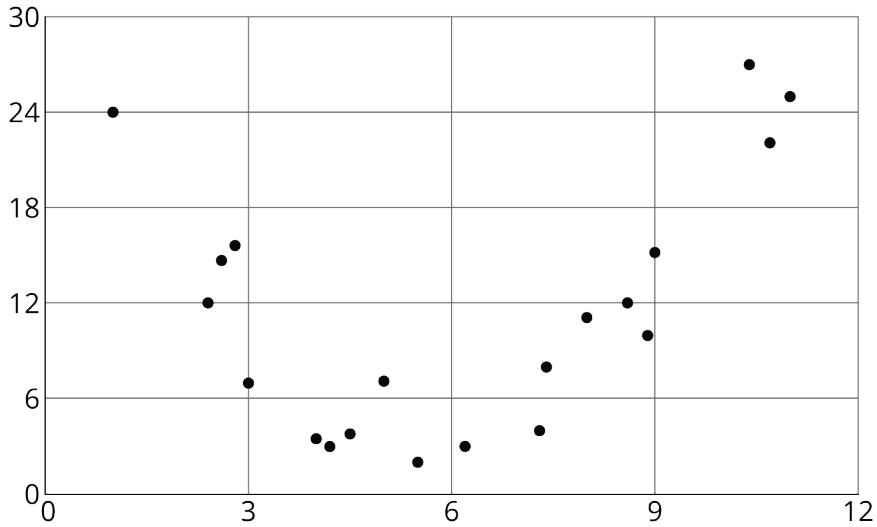
How are these scatter plots alike? How are they different?



NAME DATE PERIOD

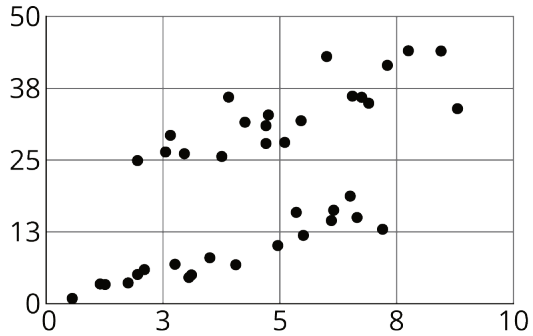
Lesson 7 Summary

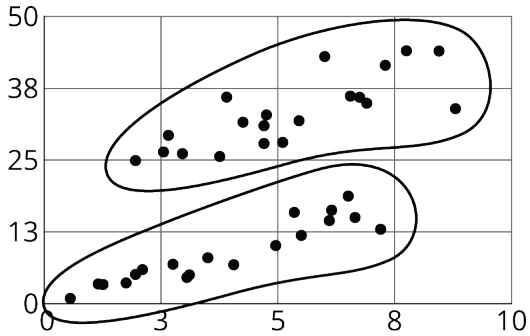
Sometimes a scatter plot shows an association that is *not* linear:



We call such an association a *non-linear association*. In later grades, you will study functions that can be models for non-linear associations.

Sometimes in a scatter plot we can see separate groups of points.



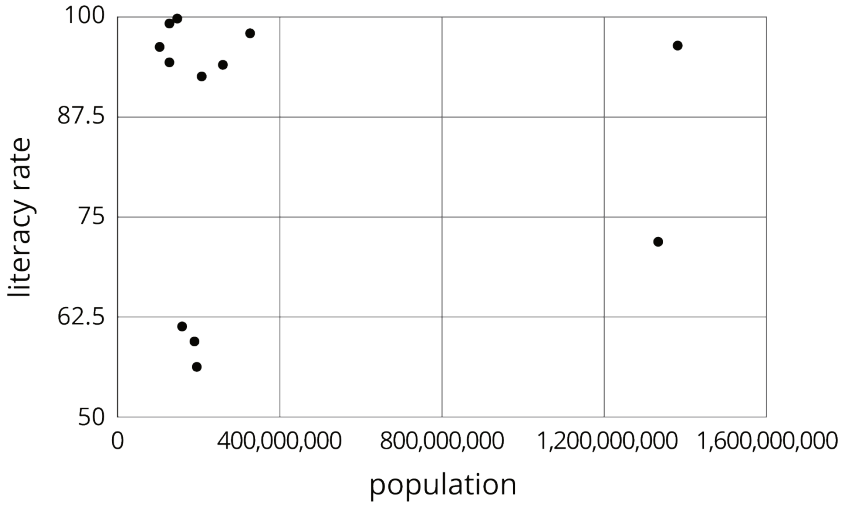


We call these groups *clusters*.

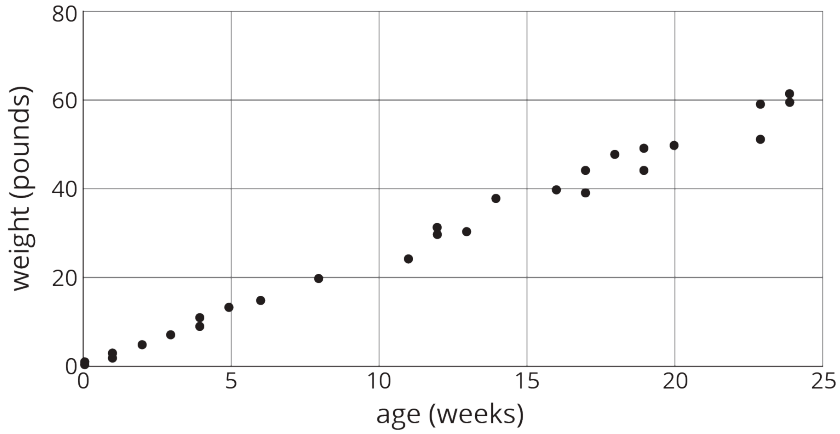
NAME DATE PERIOD

Unit 6, Lesson 7: Observing More Patterns in Scatter Plots

1. Literacy rate and population for the 12 countries with more than 100 million people are shown in the scatter plot. Circle any clusters in the data.



2. Here is a scatter plot:



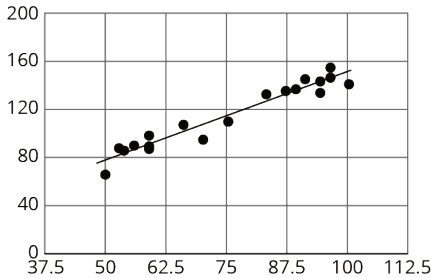
Select **all** the following that describe the association in the scatter plot:

A. Linear association
B. Non-linear association
C. Positive association
D. Negative association
E. No association

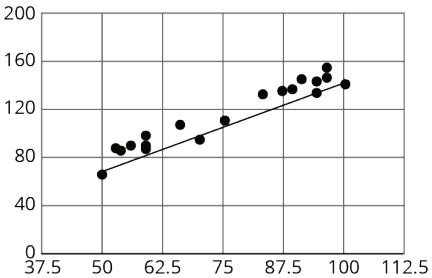
NAME DATE PERIOD

3. For the same data, two different models are graphed. Which model more closely matches the data? Explain your reasoning.

A



B



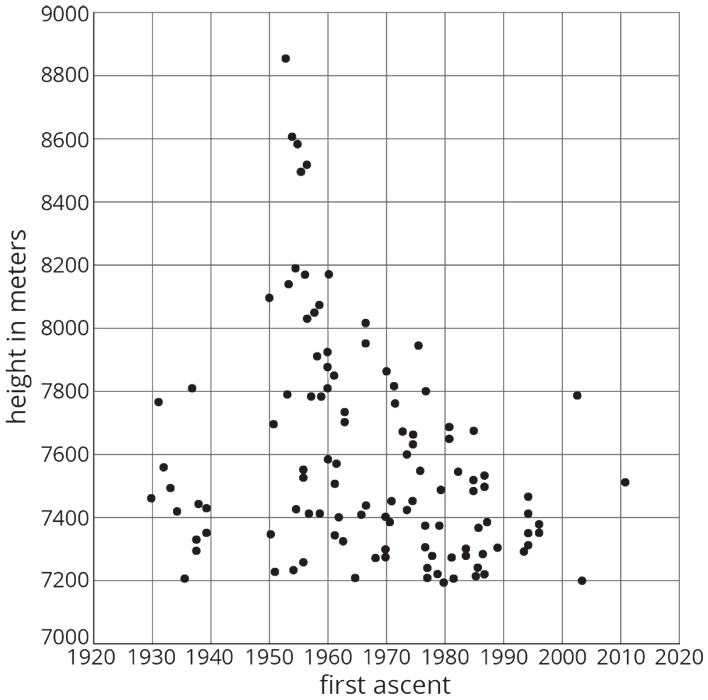
(from Unit 6, Lesson 5)

4. Here is a scatter plot of data for some of the tallest mountains on Earth.

The heights in meters and year of first recorded ascent is shown. Mount Everest is the tallest mountain in this set of data.

a. Estimate the height of Mount Everest.

b. Estimate the year of the first recorded ascent of Mount Everest.



(from Unit 6, Lesson 3)

5. A cone has a volume V, radius r, and a height of 12 cm.

a. A cone has the same height and $\frac{1}{3}$ of the radius of the original cone. Write an expression for its volume.

NAME DATE PERIOD

b. A cone has the same height and 3 times the radius of the original cone. Write an expression for its volume.

(from Unit 5, Lesson 18)

NAME DATE PERIOD

Unit 6, Lesson 8: Analyzing Bivariate Data

Let's analyze data like a pro.

8.1: Speed vs. Step Length

A researcher found an association between a dog's stride length and its speed: the longer a dog's steps, the faster it goes. The predicted speed in meters per second, s, as a function of step length in meters, l, is

$$s = 4l - 1.6$$

What does the rate of change of the function tell you about the association between stride length and speed?

NAME DATE PERIOD

8.2: Animal Brains

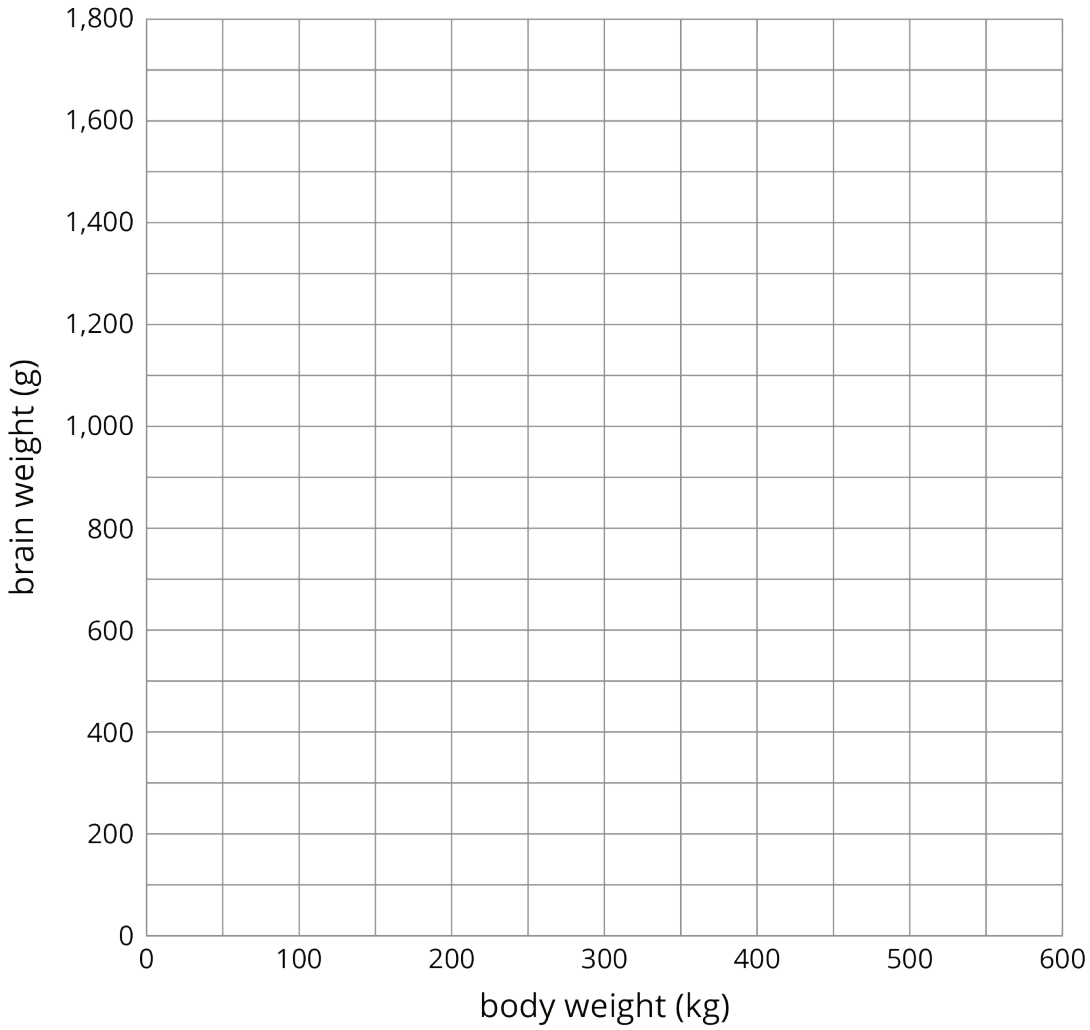
m.openup.org/1/8-6-8-2

Is there an association between the weight of an animal's body and the weight of the animal's brain?

Use the data in the table to make a scatter plot. Are there any outliers?

animal	body weight (kg)	brain weight (g)
cow	465	423
grey wolf	36	120
goat	28	115
donkey	187	419
horse	521	655
potar monkey	10	115
cat	3	26
giraffe	529	680
gorilla	207	406
human	62	1,320
rhesus monkey	7	179
kangaroo	35	56
sheep	56	175
jaguar	100	157
chimpanzee	52	440
pig	192	180

NAME DATE PERIOD



1. After removing the outliers, does there appear to be an association between body weight and brain weight? Describe the association in a sentence.

2. Using a piece of pasta and a straightedge, fit a line to your scatter plot, and estimate its slope. What does this slope mean in the context of brain and body weight?

NAME DATE PERIOD

3. Does the fitted line help you identify more outliers?

Are you ready for more?

Use one of the suggestions or find another set of data that interested you to look for associations between the variables.

- Number of wins vs number of points per game for your favorite sports team in different seasons
- Amount of money grossed vs critic rating for your favorite movies
- Price of a ticket vs stadium capacity for popular bands on tour

After you have collected the data,

1. Create a scatter plot for the data.

2. Are any of the points very far away from the rest of the data?

3. Would a linear model fit the data in your scatter plot? If so, draw it. If not, explain why a line would be a bad fit.

4. Is there an association between the two variables? Explain your reasoning.

NAME DATE PERIOD

8.3: Equal Body Dimensions

m.openup.org/1/8-6-8-3

Earlier, your class gathered data on height and arm span.

1. Sometimes a person's arm span is the same as their height. Is this true for anyone in the class?

2. Make a scatter plot for the arm span and height data, and describe any association.

3. Is the line $y = x$ a good fit for the data? If so, explain why. If not, find the equation of a line that fits the data better.

4. Examine the scatter plot. Which person in your class has the *largest* ratio between their arm span and their height? Explain or show your reasoning.

NAME DATE PERIOD

Lesson 8 Summary

People often collect data in two variables to investigate possible associations between two numerical variables and use the connections that they find to predict more values of the variables. Data analysis usually follows these steps:

1. Collect data.

2. Organize and represent the data, and look for an association.

3. Identify any outliers and try to explain why these data points are exceptions to the trend that describes the association.

4. Find a function that fits the data well.

Although computational systems can help with data analysis by graphing the data, finding a function that might fit the data, and using that function to make predictions, it is important to understand the process and think about what is happening. A computational system may find a function that does not make sense or use a line when the situation suggests that a different model would be more appropriate.

NAME DATE PERIOD

Unit 6, Lesson 8: Analyzing Bivariate Data

1. Different stores across the country sell a book for different prices. The table shows the price of the book in dollars and the number of books sold at that price.

price in dollars	number sold
11.25	53
10.50	60
12.10	30
8.45	81
9.25	70
9.75	80
7.25	120
12	37
9.99	130
7.99	100
8.75	90

a. Draw a scatter plot of this data. Label the axes.

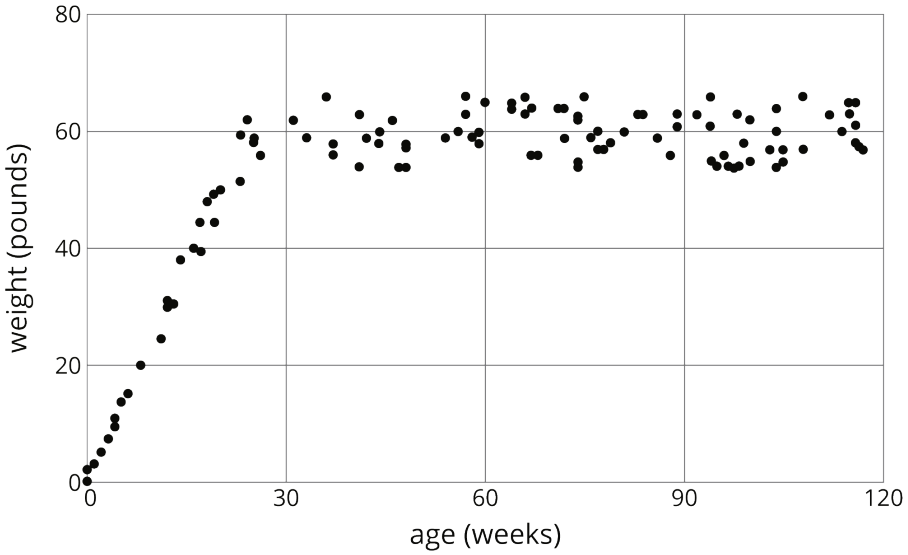
b. Are there any outliers? Explain your reasoning.

c. If there is a relationship between the variables, explain what it is.

d. Remove any outliers, and draw a line that you think is a good fit for the data.

2. Here is a scatter plot:

NAME DATE PERIOD

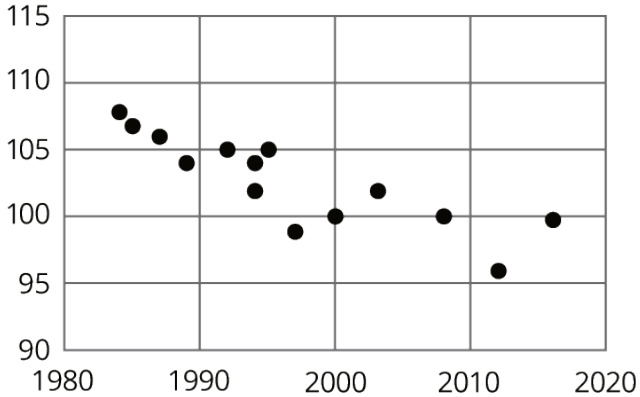


Select **all** the following that describe the association in the scatter plot:

A. Linear association
B. Non-linear association
C. Positive association
D. Negative association
E. No association

(from Unit 6, Lesson 7)

3. Using the data in the scatter plot, what can you tell about the slope of a good model?



A. The slope is positive.
B. The slope is zero.
C. The slope is negative.

NAME DATE PERIOD

D. There is no association.

(from Unit 6, Lesson 6)

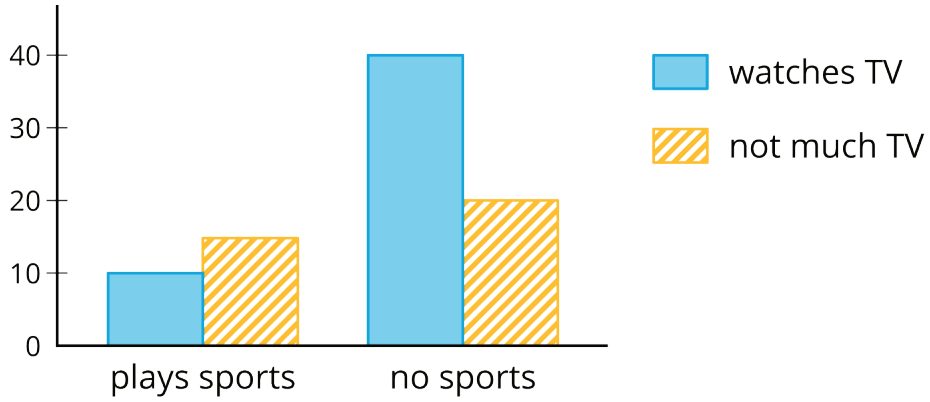
NAME DATE PERIOD

Unit 6, Lesson 9: Looking for Associations

Let's look for associations in data.

9.1: Notice and Wonder: Bar Association

What do you notice? What do you wonder?



9.2: Matching Representations Card Sort

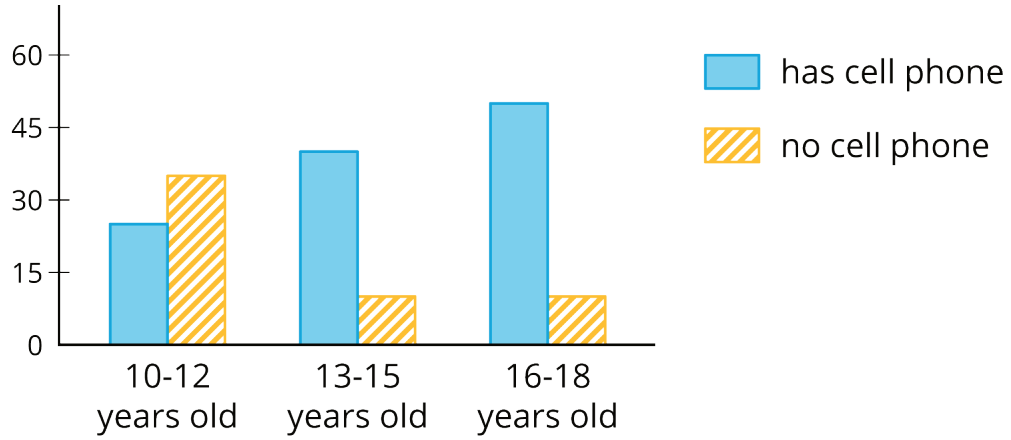
Your teacher will hand out some cards.

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

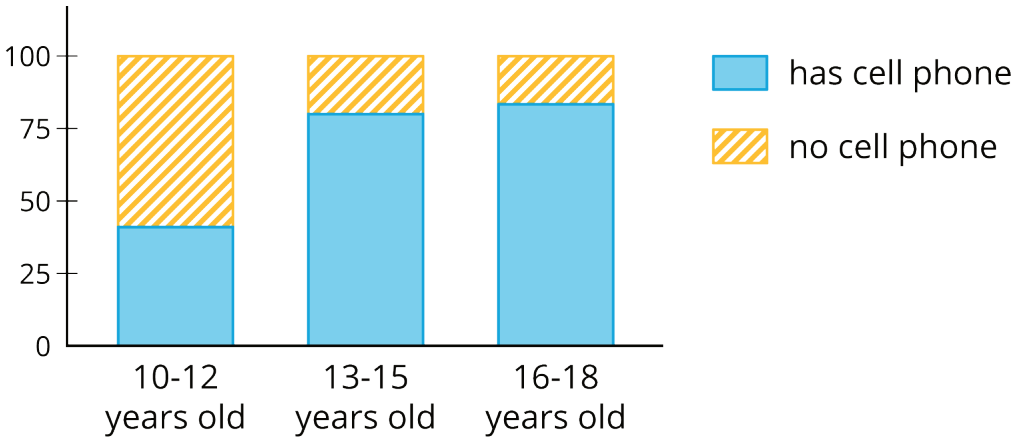
	has cell phone	does not have cell phone	total
10 to 12 years old	25	35	60
13 to 15 years old	40	10	50
16 to 18 years old	50	10	60
total	115	55	170

NAME DATE PERIOD

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

NAME DATE PERIOD

matches the other representations.

NAME DATE PERIOD

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?

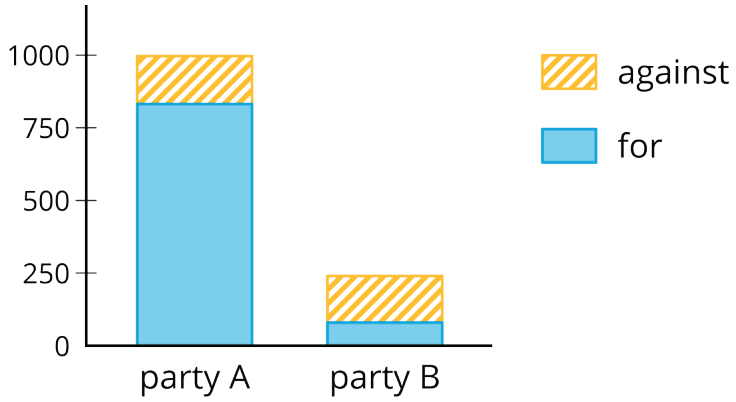
NAME DATE PERIOD

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.

NAME DATE PERIOD

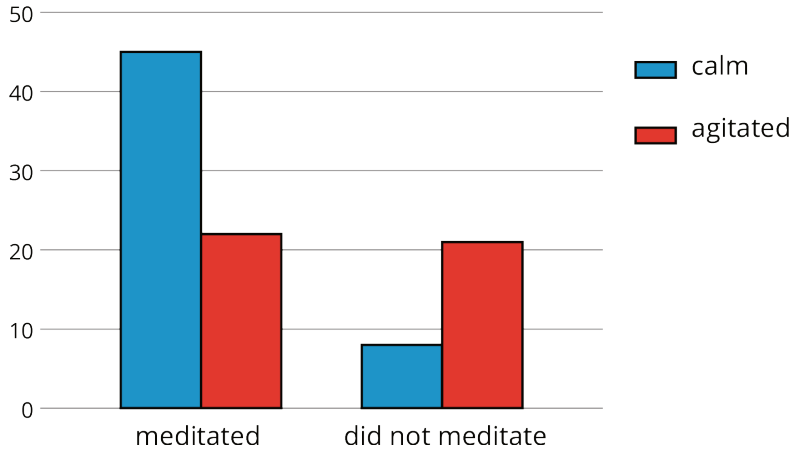
Lesson 9 Summary

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.

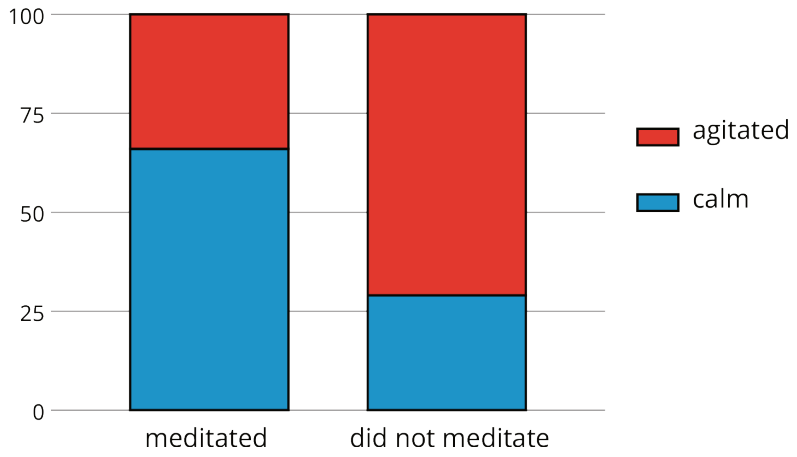


Notice that the number of athletes who did not meditate is small compared to the number who meditated (29 as compared to 68, as shown in the table).

NAME DATE PERIOD

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%



Lesson 9 Glossary Terms

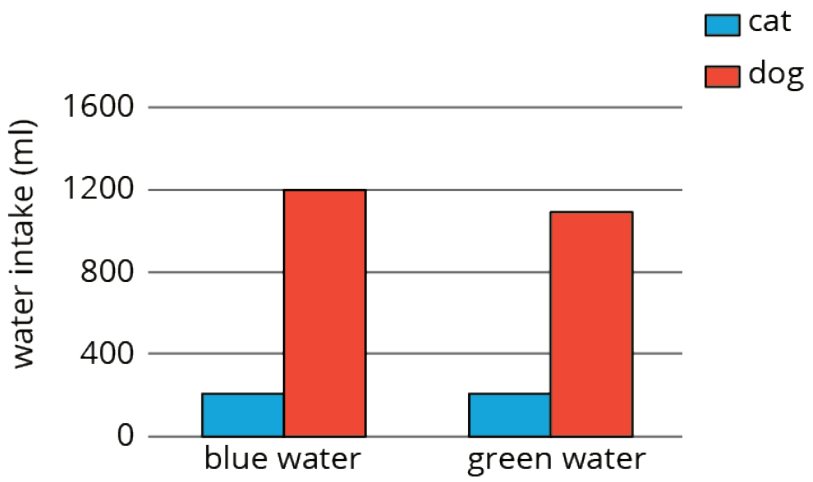
- segmented bar graph
- relative frequency
- two-way table

NAME DATE PERIOD

Unit 6, Lesson 9: Looking for Associations

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.

NAME DATE PERIOD

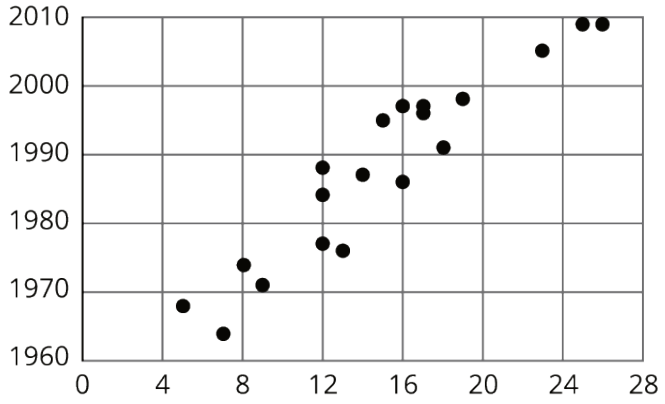
3. Researchers at a media company want to study news-reading habits among different age groups. They tracked print and online subscription data and made a 2-way table.

	internet articles	**print articles**
18–25 year olds	151	28
26–45 year olds	132	72
45–65 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)

NAME DATE PERIOD

Unit 6, Lesson 10: Using Data Displays to Find Associations

Let's use data displays to find associations.

10.1: Sports and Musical Instruments

For a survey, students in a class answered these questions:

- Do you play a sport?
- Do you play a musical instrument?

1. Here is a two-way table that gives some results from the survey. Complete the table, assuming that all students answered both questions.

	plays instrument	**does not play instrument**	**total**
plays sport	5		16
does not play sport			
total		15	25

2. To the nearest percentage point, what percentage of students who play a sport *don't* play a musical instrument?

3. To the nearest percentage point, what percentage of students who *don't* play a sport also *don't* play a musical instrument?

NAME DATE PERIOD

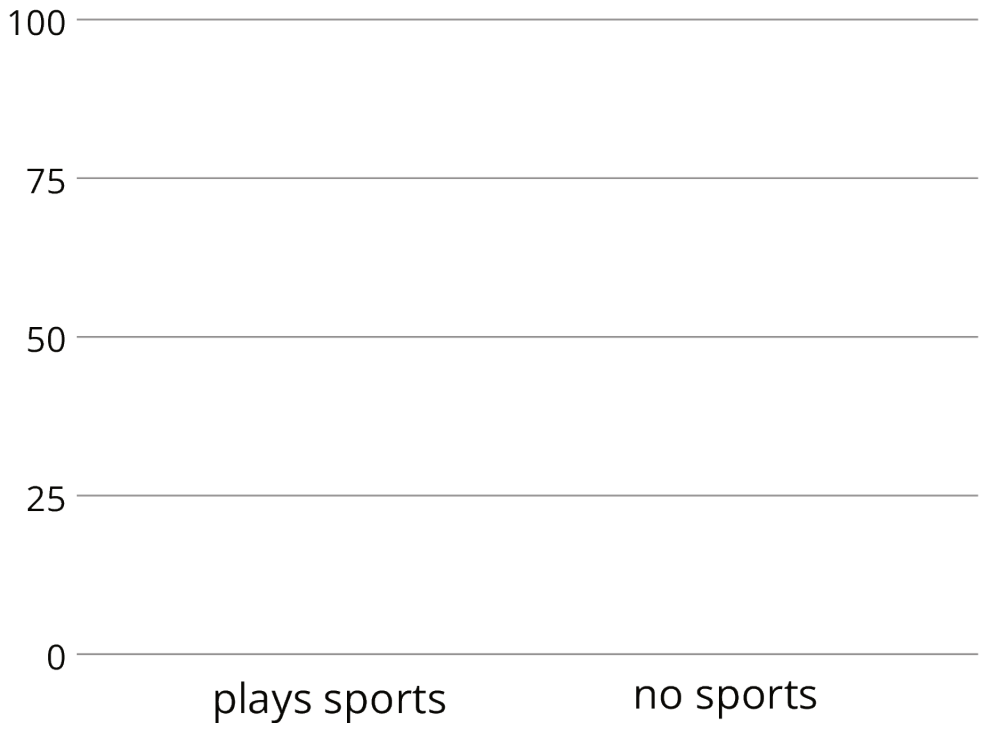
10.2: Sports and Music Association

Your teacher will give you a two-way table with information about the number of people in your class who play sports or musical instruments.

1. Complete this table to make a two-way table for the data from earlier. The table will show relative frequencies *by row*.

	plays instrument	**does not play instrument**	**row total**
plays sport			100%
does not play sport			100%

2. Make a segmented bar graph for the table. Use one bar of the graph for each row of the table.



NAME DATE PERIOD

3. Complete the table to make a two-way table for the data from earlier. The table will show relative frequencies *by column*.

	plays instrument	**does not play instrument**
plays sport		
does not play sport		
column total	100%	100%

4. Using the values in the table, make a segmented bar graph. Use one bar of the graph for each column of the table.

5. Based on the two-way tables and segmented bar graphs, do you think there is an association between playing a sport and playing a musical instrument? Explain how you know.

NAME DATE PERIOD

10.3: Colored Erasers

An eraser factory has five machines. One machine makes the eraser shapes. Then each shape goes through the red machine, blue machine, yellow machine, or green machine to have a side colored.

The manager notices that an uncolored side of some erasers is flawed at the end of the process and wants to know which machine needs to be fixed: the shape machine or some of the color machines. The manager collected data on the number of flawed and unflawed erasers of each color.

	unflawed	**flawed**	**total**
red	285	15	300
blue	223	17	240
yellow	120	80	200
green	195	65	260
total	823	177	1000

1. Work with a partner. Each of you should make one segmented bar graph for the data in the table. One segmented bar graph should have a bar for each *row* of the table. The other segmented bar graph should have one bar for each *column* of the table.

2. Are the flawed erasers associated with certain colors? If so, which colors? Explain your reasoning.

NAME DATE PERIOD

Are you ready for more?

Based on the federal budgets for 2009, the table shows where some of the federal money was expected to go. The values are in billions of U.S. Dollars.

	United States	Japan	United Kingdom
defense	718.4	42.8	49.2
education	44.9	47.5	113.9

1. Why would a segmented bar graph be more useful than the table of data to see any associations between the country and where the money is spent?

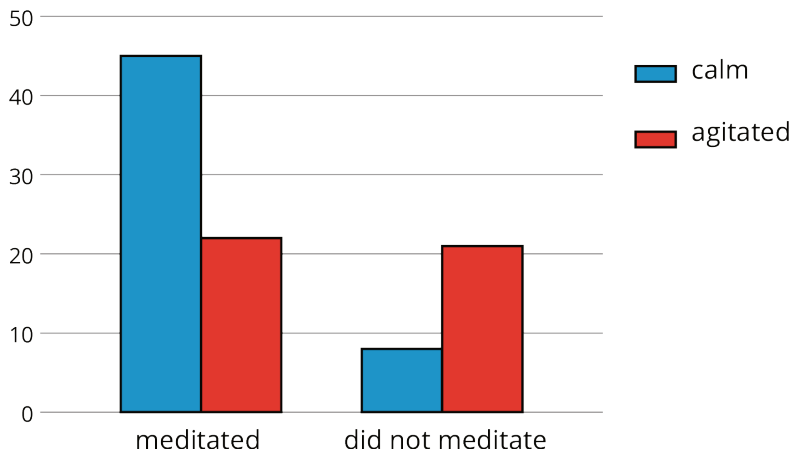
2. Create a segmented bar graph that represents the data from the table.

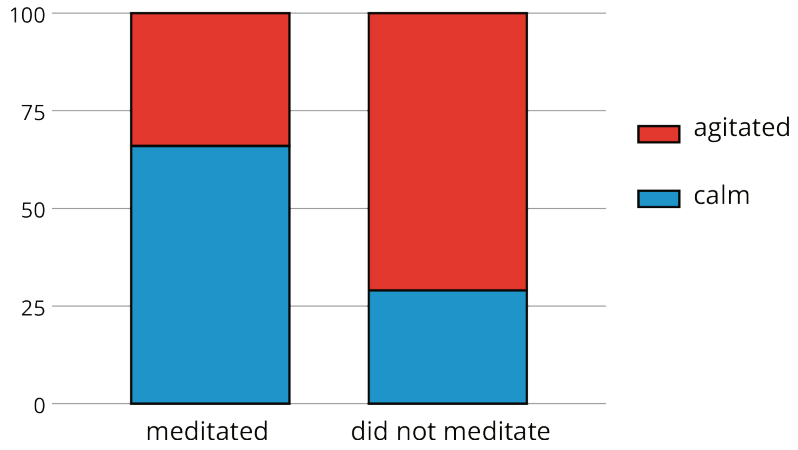
3. Is there an association between the country's budget and their spending in these areas? Explain your reasoning.

NAME DATE PERIOD

Lesson 10 Summary

In an earlier lesson, we looked at data on meditation and state of mind in athletes.





Is there an association between meditation and state of mind?

The bar graph shows that more athletes were calm than agitated among the group that meditated, and more athletes were agitated than calm among the group that did not. We can see the proportions of calm meditators and calm non-meditators from the segmented bar graph, which shows that about 66% of athletes who meditated were calm, whereas only about 27% of those who did not meditate were calm.

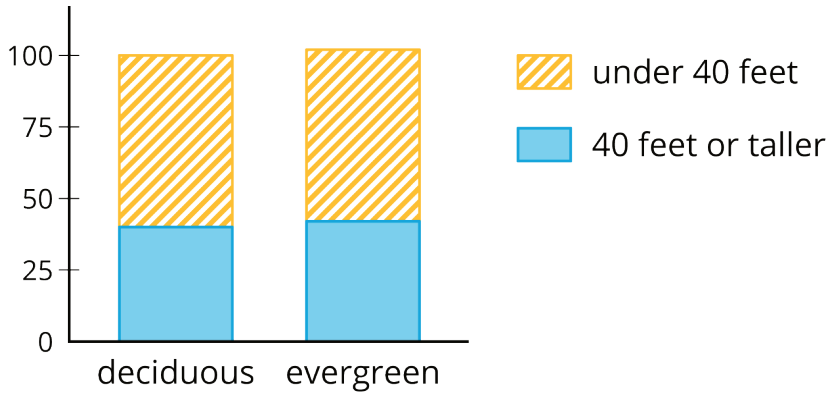
This does not necessarily mean that meditation causes calm; it could be the other way around, that calm athletes are more inclined to meditate. But it does suggest that there is an association between meditating and calmness.

NAME DATE PERIOD

Unit 6, Lesson 10: Using Data Displays to Find Associations

1. An ecologist is studying a forest with a mixture of tree types. Since the average tree height in the area is 40 feet, he measures the height of the tree against that. He also records the type of tree. The results are shown in the table and segmented bar graph. Is there evidence of an association between tree height and tree type? Explain your reasoning.

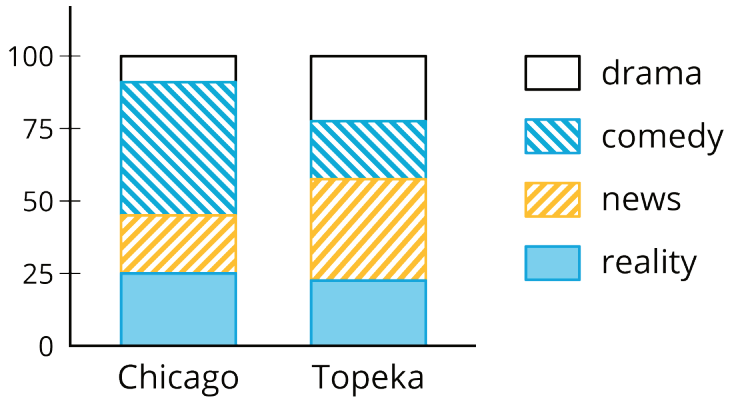
	under 40 feet	40 feet or taller	total
deciduous	45	30	75
evergreen	14	10	24
total	59	40	99



2. Workers at an advertising agency are interested in people's TV viewing habits. They take a survey of people in two cities to try to find patterns in the types of shows they watch. The results are recorded in a table and shown in a segmented bar graph. Is there evidence of different viewing habits? If so, explain.

	reality	news	comedy	drama
Chicago	50	40	90	20
Topeka	45	70	40	45

NAME DATE PERIOD



3. A scientist is interested in whether certain species of butterflies like certain types of local flowers. The scientist captures butterflies in two zones with different flower types and records the number caught. Do these data show an association between butterfly type and zone? Explain your reasoning.

	zone 1	**zone 2**
eastern tiger swallowtail	16	34
monarch	24	46

NAME DATE PERIOD

Unit 6, Lesson 11: Gone In 30 Seconds

Let's gather and analyze some timing data.

11.1: Measuring 30 Seconds

In this activity, you'll get two chances to guess at how long 30 seconds is, then look for an association between the two guesses of all students.

1. Work with a partner. Follow the instructions listed here to gather your data.
   - One of you will hold a stopwatch where the other person cannot see it.
   - The person holding the stopwatch says "go" and starts the timer.
   - The other person says "stop" when they think 30 seconds have passed.
   - The person holding the stopwatch will stop the timer, then report and record the time to the nearest second.
   - The person holding the stopwatch will give a second chance, repeating the experiment.
   - After *both* times are recorded, switch roles.
2. Record the group data in this table. When you finish, a group member should give the data to the teacher.

name	time 1	time 2

3. Look at your data. Comparing Time 1 to Time 2, do you think there is a positive association, a negative association, or no association? Discuss your thinking with your group.

NAME DATE PERIOD

4. What are some ways you could organize and represent the entire class's data?

NAME DATE PERIOD

5. Make a scatter plot of the entire class's data and look for patterns. Identify any outliers and the type of any association you observe.

6. Draw two lines on your scatter plot: a vertical line and a horizontal line, each representing 30 seconds for one trial. Use the table for the class's data to complete this two-way table.

	Time 2 < 30 sec	**Time 2 = 30 sec**	**Time 2 > 30 sec**	**total**
Time 1 < 30 sec				
Time 1 = 30 sec				
Time 1 > 30 sec				
total				

7. Use the two-way table to decide whether there is an association between Time 1 and Time 2. Explain how you know.