Topic B:

Bivariate Numerical Data

8.SP.A.1, 8.SP.A.2

Focus Standard:	8.SP.A.1	Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.
	8.SP.A.2	Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.
Instructional Days:	4	
Lesson 6:	Scatter Plots (P)[1]	
Lesson 7:	Patterns in Scatter Plots (P)	
Lesson 8:	Informally Fitting a Line (P)	
Lesson 9:	Determining the Equation of a Line Fit to Data (P)	

In Topic B, students connect their study of linear functions to applications involving bivariate data. A key tool in developing this connection is a scatter plot. In Lesson 6, students construct scatter plots and focus on identifying linear versus nonlinear patterns (**8.SP.A.1**). They distinguish positive linear association and negative linear association based on the scatter plot. Students describe trends in the scatter plot, along with clusters and outliers (points that do not fit the pattern). In Lesson 8, students informally fit a straight line to data displayed in a scatter plot (**8.SP.A.2**) by judging the closeness of the data points to the line. In Lesson 9, students interpret and determine the equation of the line they fit to the data and use the equation to make predictions and to evaluate possible association of the variables. Based on these predictions, students address the need for a "best-fit" line, which is formally introduced in Grade 9.

[1] Lesson Structure Key: **P**-Problem Set Lesson, **M**-Modeling Cycle Lesson, **E**-Exploration Lesson, **S**-Socratic Lesson

Lesson 6: Scatter Plots

Student Outcomes

- Students construct scatter plots.
- Students use scatter plots to investigate relationships.
- Students understand the distinction between a statistical relationship and a cause-and-effect relationship.

Lesson Notes

This lesson is the first of a set of lessons dealing with relationships between numerical variables. In this lesson, students learn how to construct a scatter plot and look for patterns in the scatter plot which suggest that there is a statistical relationship between two numerical variables.

Classwork

Example 1 (5 minutes)

Spend a few minutes introducing the context of this example. Make sure that students understand that in this context, an observation can be thought of as an ordered pair consisting of the value for each of two variables.

Example 1

A bivariate data set consists of observations on two variables. For example, you might collect data on 13 different car models. Each observation in the data set would consist of an (x, y) pair.

x = weight (in pounds, rounded to the nearest 50 pounds)

and

y = fuel efficiency (in miles per gallon, mpg.)

The table below shows the weight and fuel efficiency for 13 car models with automatic transmissions manufactured in 2009 by Chevrolet.

Model	Weight (pounds)	Fuel Efficiency (mpg)
1	3,200	23
2	2,550	28
3	4,050	19
4	4,050	20
5	3,750	20
6	3,550	22
7	3,550	19
8	3,500	25
9	4,600	16
10	5,250	12
11	5,600	16
12	4,500	16
13	4,800	15

Scaffolding:

- Point out to students that the word "bivariate" is composed of the prefix "bi" and the stem "variate."
- "Bi" means "two."
- "Variate" indicates a "variable."
- The focus in this lesson is on two numerical variables.

Scaffolding:

- ELL students new to the curriculum may be familiar with the metric system (kilometers, kilograms, and liters) but unfamiliar with the English system (miles, pounds, and gallons).
- It may be helpful to provide conversions:
  1 kg ≈ 2.2 lb.
  1 lb. ≈ 0.45 kg
  1 km≈ 0.62 mi.
  1 mi. ≈ 1.61 km

Exercises 1–3 (10–12 minutes)

After students have had a chance to think about Exercise 1, make sure that everyone understands what an observation (an ordered pair) represents in the context of this example. Relate plotting the point that corresponds to the first observation to students' previous work with plotting points in a rectangular coordinate system. As a way of encouraging the need to look at a graph of the data, consider asking students to try to determine if there is a relationship between weight and fuel efficiency by just looking at the table. Allow students time to complete the scatter plot and complete Exercise 3. Have students share their answers to Exercise 3.

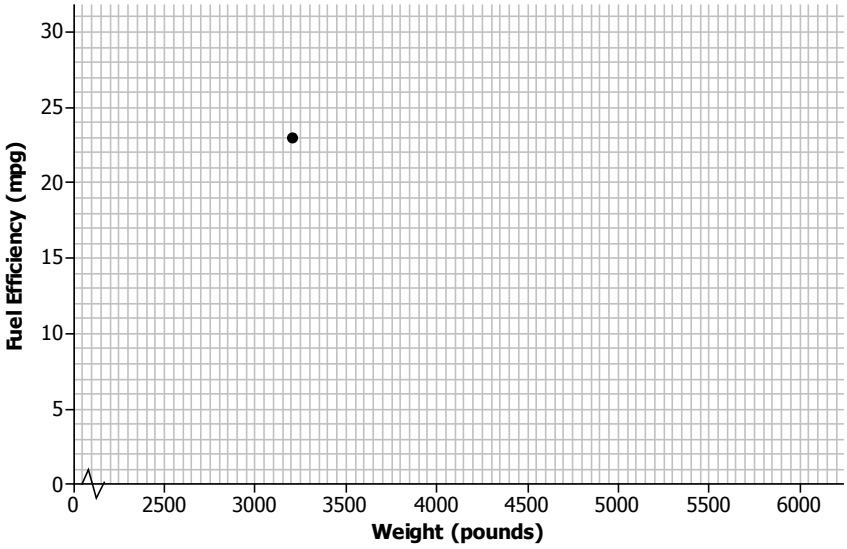
Exercises 1–3

1. **In the table above, the observation corresponding to model 1 is $(3200, 23)$. What is the fuel efficiency of this car? What is the weight of this car?**

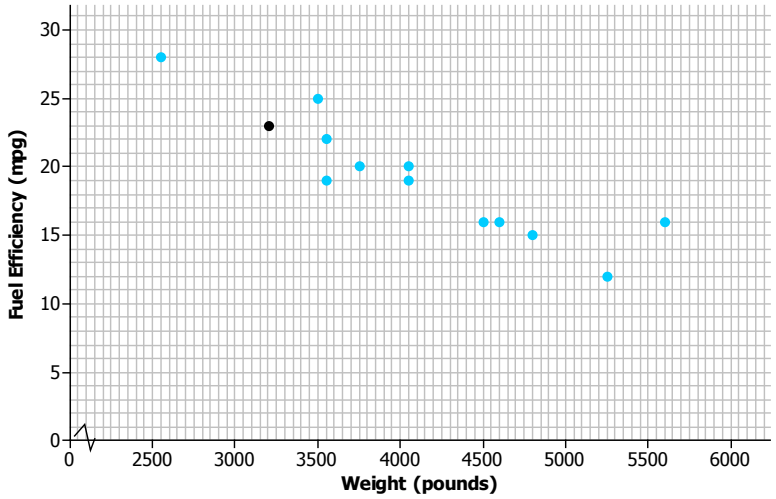
   The fuel efficiency is 23 miles per gallon, and the weight is $3,200$ pounds.

One question of interest is whether there is a relationship between the car weight and fuel efficiency. The best way to begin to investigate is to construct a graph of the data. A *scatter plot* is a graph of the (x, y) pairs in the data set. Each (x, y) pair is plotted as a point in a rectangular coordinate system.

For example, the observation $(3200, 23)$ would be plotted as a point located above $3{,}200$ on the x-axis and across from 23 on the y-axis, as shown below.



2. Add the points corresponding to the other 12 observations to the scatter plot.



MP.7

3. Do you notice a pattern in the scatter plot? What does this imply about the relationship between weight (x) and fuel efficiency (y)?

There does seem to be a pattern in the plot. Larger weights tend to be paired with smaller fuel efficiencies, so it looks like heavier cars generally have lower fuel efficiency.

Exercises 4–8 (6 - 8 minutes)

These exercises give students additional practice creating a scatter plot and identifying a pattern in the plot. Students should work individually on these exercises and then discuss their answers to Exercises 7 and 8 with a partner. However, some ELL students may benefit from paired or small group work, particularly if their English literacy is not strong.

Exercises 4–8

Is there a relationship between price and the quality of athletic shoes? The data in the table below are from the Consumer Reports website.

x = price (in dollars)

and

y = Consumer Reports quality rating

The quality rating is on a scale of 0 to 100, with 100 being the highest quality.

Shoe	Price (dollars)	Quality Rating
1	65	71
2	45	70
3	45	62
4	80	59
5	110	58
6	110	57
7	30	56
8	80	52
9	110	51
10	70	51

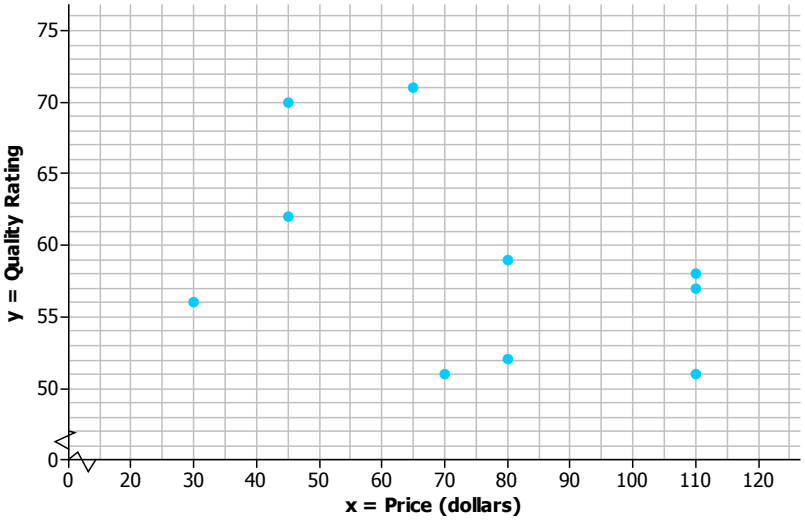
4. One observation in the data set is $(110, 57)$. What does this ordered pair represent in terms of cost and quality?

The pair represents a shoe that costs $\$110$ with a quality rating of 57.

5. To construct a scatter plot of these data, you need to start by thinking about appropriate scales for the axes of the scatter plot. The prices in the data set range from $\$30$ to $\$110$, so one reasonable choice for the scale of the x-axis would range from $\$20$ to $\$120$, as shown below. What would be a reasonable choice for a scale for the y-axis?

Sample response: The smallest y-value is 51, and the largest value is 71. So, the y-axis could be scaled from 50 to 75.

6. Add a scale to the y-axis. Then, use these axes to construct a scatter plot of the data.



7. **Do you see any pattern in the scatter plot indicating that there is a relationship between price and quality rating for athletic shoes?**

   Answers will vary. Students may say that they do not see a pattern or they may say that they see a slight downward trend.

8. **Some people think that if shoes have a high price, they must be of high quality. How would you respond?**

   Answers will vary. The data do not support this. Students will either respond that there does not appear to be a relationship between price and quality, or if they saw a downward trend in the scatter plot, they might even indicate that the higher priced shoes tend to have lower quality. Look for consistency between the answer to this question and how students answered the previous question.

Scaffolding:

- For more complicated and reflective answers, consider allowing ELL students to use one or more of the following options: collaborate with a same-language peer, frame their response in an illustration, or provide a first-language narration or response.

Example 2 (5–10 minutes): Statistical Relationships

This example makes a very important point. As you discuss this example with the class, make sure students understand the distinction between a statistical relationship and a cause-and-effect relationship. After discussing the example, ask students if they can think of other examples of numerical variables that might have a statistical relationship, but which probably do not have a cause-and-effect relationship.

Example 2

A pattern in a scatter plot indicates that the values of one variable tend to vary in a predictable way as the values of the other variable change. This is called a *statistical relationship*. In the fuel efficiency and car weight example, fuel efficiency tended to decrease as car weight increased.

This is useful information, but be careful not to jump to the conclusion that increasing the weight of a car *causes* the fuel efficiency to go down. There may be some other explanation for this. For example, heavier cars may also have bigger engines, and bigger engines may be less efficient. You cannot conclude that changes to one variable *cause* changes in the other variable just because there is a statistical relationship in a scatter plot.

Exercises 9–10 (5 minutes)

Students can work individually or with a partner on these exercises. Then, confirm answers as a class.

Exercises 9–10

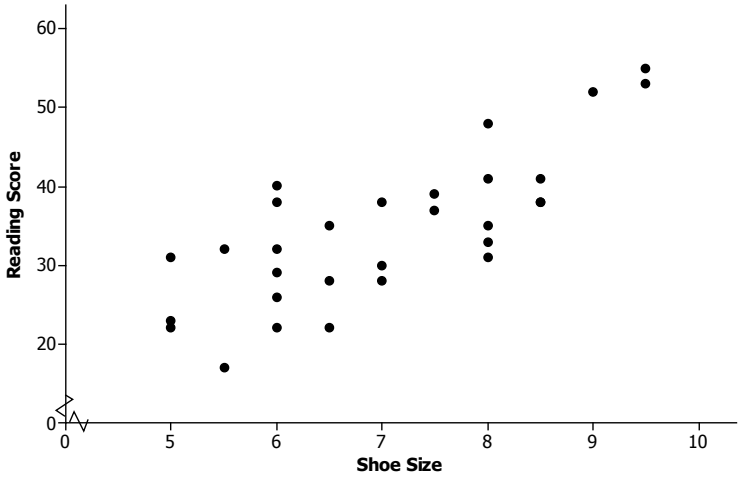
9. **Data were collected on**

x = **shoe size**

and

y = **score on a reading ability test**

for 30 elementary school students. The scatter plot of these data is shown below. Does there appear to be a statistical relationship between shoe size and score on the reading test?



Possible answer: The pattern in the scatter plot appears to follow a line. As shoe sizes increase, the reading scores also seem to increase. There does appear to be a statistical relationship because there is a pattern in the scatter plot.

10. **Explain why it is not reasonable to conclude that having big feet causes a high reading score. Can you think of a different explanation for why you might see a pattern like this?**

Possible answer: You cannot conclude that just because there is a statistical relationship between shoe size and reading score that one causes the other. These data were for students completing a reading test for younger elementary school children. Older children, who would have bigger feet than younger children, would probably tend to score higher on a reading test for younger students.

Closing (3 minutes)

Consider posing the following questions; allow a few student responses for each:

- Why is it helpful to make a scatter plot when you have data on two numerical variables?
  - *A scatter plot makes it easier to see patterns in the data and to see if there is a statistical relationship between the two variables.*

- Can you think of an example of two variables that would have a statistical relationship but not a cause-and-effect relationship?
  - *Possible response: One famous example is the number of people who must be rescued by lifeguards at the beach and the number of ice cream sales. Both of these variables have higher values when the temperature is high and lower values when the temperature is low. So, there is a statistical relationship between them—they tend to vary in a predictable way. However, it would be silly to say that an increase in ice cream sales causes more beach rescues!*

Lesson Summary

- **A scatter plot is a graph of numerical data on two variables.**
- **A pattern in a scatter plot suggests that there may be a relationship between the two variables used to construct the scatter plot.**
- **If two variables tend to vary together in a predictable way, we can say that there is a statistical relationship between the two variables.**
- **A statistical relationship between two variables does not imply that a change in one variable causes a change in the other variable (a cause-and-effect relationship).**

Exit Ticket (5 minutes)

Name _____ Date_____

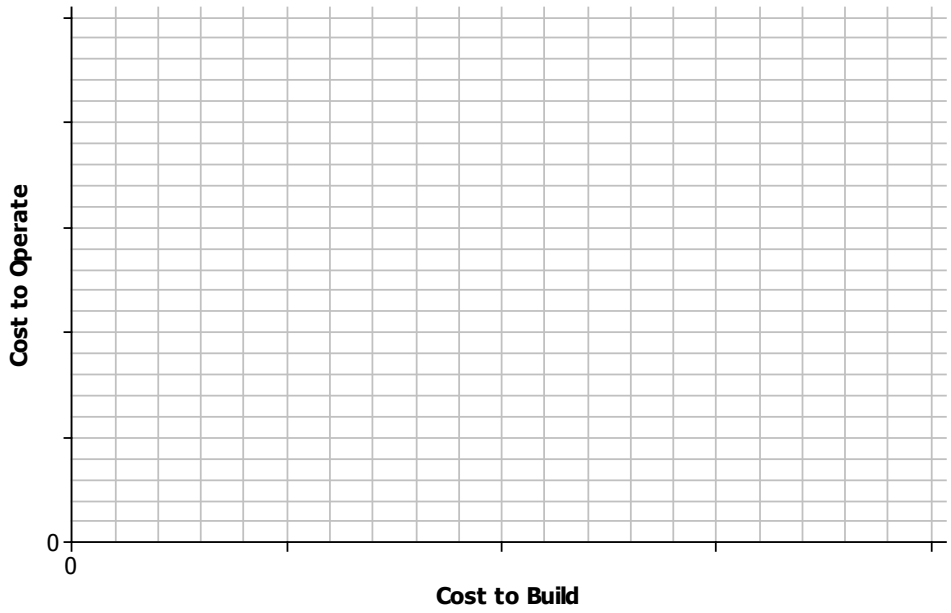
Lesson 6: Scatter Plots

Exit Ticket

Energy is measured in kilowatt hours. The table below shows the cost of building a facility to produce energy and the ongoing cost of operating the facility for five different types of energy.

Type of Energy	Cost to Operate (cents per kilowatt hour)	Cost to Build (dollars per kilowatt hour)
Hydroelectric	0.4	2,200
Wind	1.0	1,900
Nuclear	2.0	3,500
Coal	2.2	2,500
Natural Gas	4.8	1,000

1. Construct a scatter plot of the cost to build the facility (x) and the cost to operate the facility (y). Use the grid below, and be sure to add an appropriate scale to the axes.



2. Do you think that there is a statistical relationship between building cost and operating cost? If so, describe the nature of the relationship.

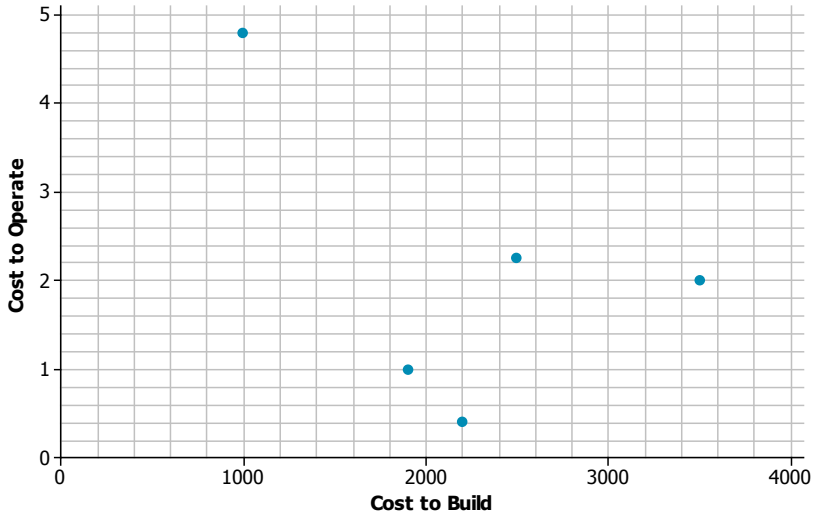
3. Based on the scatter plot, can you conclude that decreased building cost is the cause of increased operating cost? Explain.

Exit Ticket Sample Solutions

Energy is measured in kilowatt hours. The table below shows the cost of building a facility to produce energy and the ongoing cost of operating the facility for five different types of energy.

Type of Energy	Cost to Operate (cents per kilowatt hour)	Cost to Build (dollars per kilowatt hour)
Hydroelectric	0.4	$2,200$
Wind	1.0	$1,900$
Nuclear	2.0	$3,500$
Coal	2.2	$2,500$
Natural Gas	4.8	$1,000$

1. **Construct a scatter plot of the cost to build the facility (x) and the cost to operate the facility (y). Use the grid below, and be sure to add an appropriate scale to the axes.**



2. **Do you think that there is a statistical relationship between building cost and operating cost? If so, describe the nature of the relationship.**

   Answers may vary. Sample answer: Yes, because it looks like there is a downward pattern in the scatter plot. It appears that the types of energy that have facilities that are more expensive to build are less expensive to operate.

3. **Based on the scatter plot, can you conclude that decreased building cost is the cause of increased operating cost? Explain.**

   Sample answer: No. Just because there may be a statistical relationship between cost to build and cost to operate does not mean that there is a cause-and-effect relationship.

Problem Set Sample Solutions

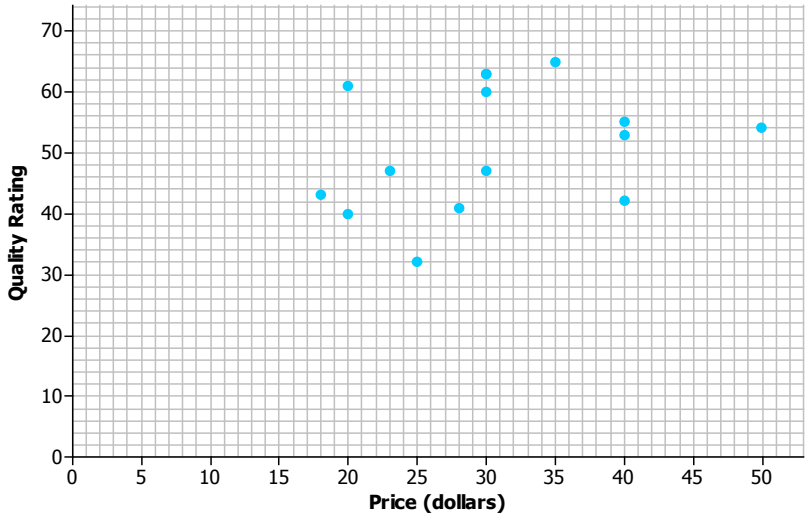
The problem set is intended to reinforce material from the lesson and have students think about the meaning of points in a scatter plot, clusters, positive and negative linear trends, and trends that are not linear.

1. The table below shows the price and overall quality rating for 15 different brands of bike helmets.

Data Source: www.consumerreports.org

Helmet	Price (dollars)	Quality Rating
A	35	65
B	20	61
C	30	60
D	40	55
E	50	54
F	23	47
G	30	47
H	18	43
I	40	42
J	28	41
K	20	40
L	25	32
M	30	63
N	30	63
O	40	53

Construct a scatter plot of price (x) and quality rating (y). Use the grid below.



2. Do you think that there is a statistical relationship between price and quality rating? If so, describe the nature of the relationship.

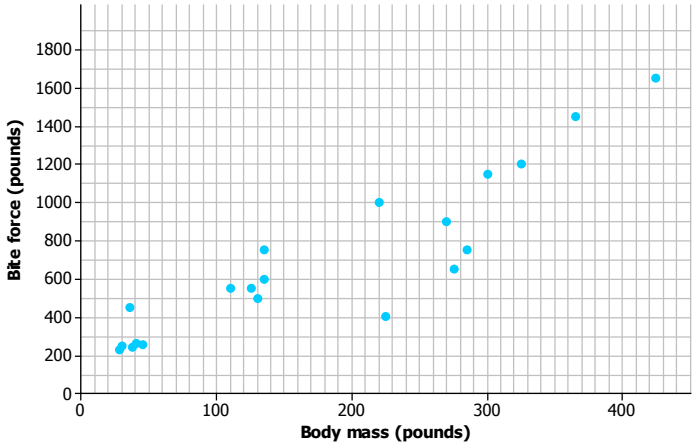
Sample answer: No. There is no pattern visible in the scatter plot. There does not appear to be a relationship between price and the quality rating for bike helmets.

3. Scientists are interested in finding out how different species adapt to finding food sources. One group studied crocodiles to find out how their bite force was related to body mass and diet. The table below displays the information they collected on body mass (in pounds) and bite force (in pounds).

Species	Body mass (pounds)	Bite force (pounds)
Dwarf crocodile	35	450
Crocodile F	40	260
Alligator A	30	250
Caiman A	28	230
Caiman B	37	240
Caiman C	45	255
Croc A	110	550
Nile crocodile	275	650
Croc B	130	500
Croc C	135	600
Croc D	135	750
Caiman D	125	550
Indian Gharial croc	225	400
Crocodile G	220	1,000
American Croc	270	900
Croc D	285	750
Croc E	425	1,650
American Alligator	300	1,150
Alligator B	325	1,200
Alligator C	365	1,450

Data Source: PLoS One Greg Erickson biomechanics, Florida State University

Construct a scatter plot of body mass (x) and bite force (y). Use the grid below, and be sure to add an appropriate scale to the axes.



4. Do you think that there is a statistical relationship between body mass and bite force? If so, describe the nature of the relationship.

Sample answer: Yes, because it looks like there is an upward pattern in the scatter plot. It appears that alligators with larger body mass also tend to have greater bite force.

5. Based on the scatter plot, can you conclude that increased body mass causes increased bite force? Explain.

Sample answer: No. Just because there is a statistical relationship between body mass and bite force does not mean that there is a cause-and-effect relationship.

Lesson 7: Patterns in Scatter Plots

Student Outcomes

- Students distinguish linear patterns from nonlinear patterns based on scatter plots.
- Students describe positive and negative trends in a scatter plot.
- Students identify and describe unusual features in scatter plots, such as clusters and outliers.

Lesson Notes

This lesson asks students to look for and describe patterns in scatter plots. It provides a foundation for later lessons in which students will use a line to describe the relationship between two numerical variables when the pattern in the scatter plot is linear. Students will distinguish between linear and nonlinear relationships as well as positive and negative linear relationships. The terms *clusters* and *outliers* are also introduced, and students look for these features in scatter plots and investigate what clusters and outliers reveal about the data.

Classwork

Example 1 (3–5 minutes)

Spend a few minutes going over the three questions posed as a way to help students structure their thinking about data displayed in a scatter plot. Students should see that looking for patterns in a scatter plot is a logical extension of their work in the previous lesson where they learned to make a scatter plot. Make sure that students understand the distinction between a positive linear relationship and a negative linear relationship before moving on to Exercises 1–5. Students will have a chance to practice answering these questions in the exercises that follow. To highlight MP.7, consider asking students to examine the five scatter plots and describe their similarities and differences before telling students what to look for.

Example 1

In the previous lesson, you learned that when data is collected on two numerical variables, a good place to start is to look at a scatter plot of the data.

When you look at a scatter plot, you should ask yourself the following questions:

1. **Does it look like there is a relationship between the two variables used to make the scatter plot?**
2. **If there is a relationship, does it appear to be linear?**
3. **If the relationship appears to be linear, is the relationship a positive linear relationship or a negative linear relationship?**

Scaffolding:

- Point out to students that the word "relationship" is not the same as the use of the word describing a familial connection, such as a sister or cousin.
- In this lesson, a "relationship" indicates that two numerical variables have a connection that can be described either verbally or with mathematical symbols.

Scaffolding:

- For ELL students, the teacher may need to read aloud the information in Example 1, highlighting each key point with a visual example as students record it in a graphic organizer for reference.

To answer the first question, look for patterns in the scatter plot. Does there appear to be a general pattern to the points in the scatter plot, or do the points look as if they are scattered at random? If you see a pattern, you can answer the second question by thinking about whether the pattern would be well-described by a line. Answering the third question requires you to distinguish between a positive linear relationship and a negative linear relationship. A positive linear relationship is one that is described by a line with a positive slope. A negative linear relationship is one that is described by a line with a negative slope.

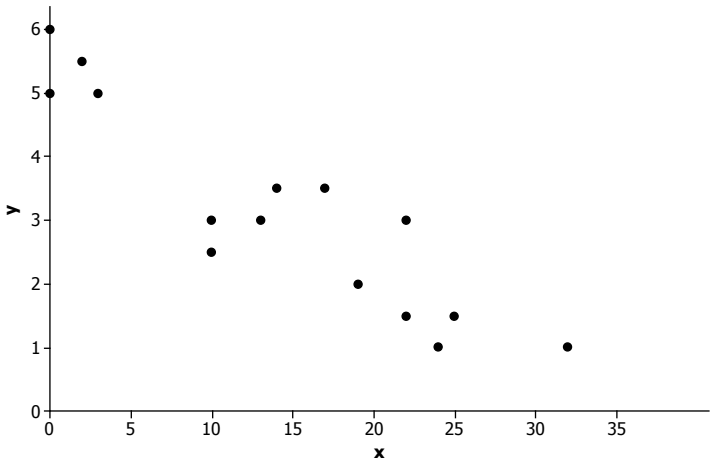
Exercises 1–5 (8–10 minutes)

You may want to answer Exercise 1 as part of a whole class discussion, and then allow students to work individually or in pairs on Exercises 2–5. Have students share answers to these exercises and discuss any of the exercises where there is disagreement on the answers. Additionally, point out to students that scatter plots that more closely resemble a linear pattern are sometimes called "strong." Scatter plots that are linear but not as close to a line are sometimes known as "weak." A linear relationship may sometimes be referred to as "strong positive," "weak positive," "strong negative," or "weak negative." Consider using these terms with students as you discuss their scatter plots.

Exercises 1–5

Take a look at the following five scatter plots. Answer the three questions above for each scatter plot.

1. Scatter plot 1



Is there a relationship?

Yes.

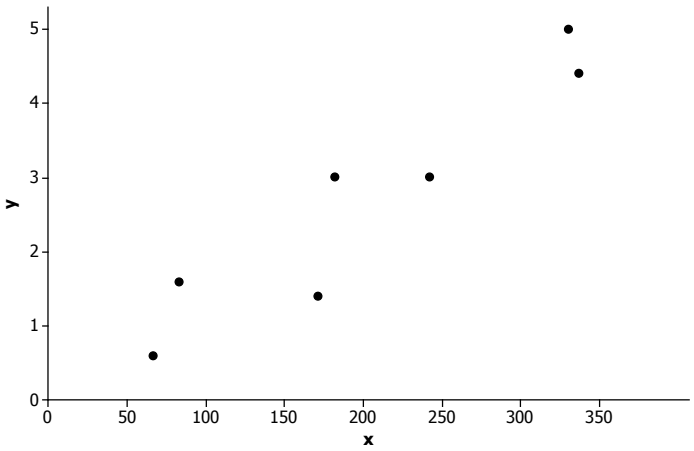
If there is a relationship, does it appear to be linear?

Yes.

If the relationship appears to be linear, is it a positive or negative linear relationship?

Negative.

2. Scatter plot 2



Is there a relationship?

Yes.

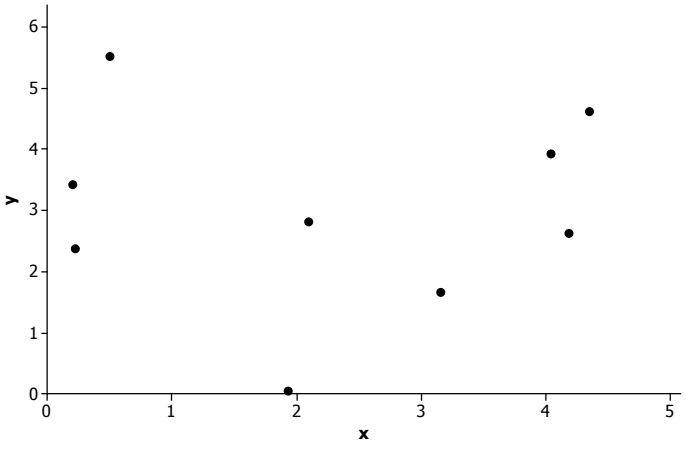
If there is a relationship, does it appear to be linear?

Yes.

If the relationship appears to be linear, is it a positive or negative linear relationship?

Positive.

3. Scatter plot 3



Is there a relationship?

No.

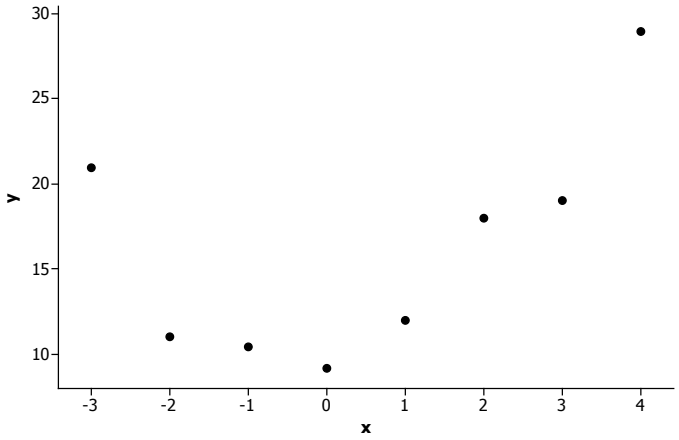
If there is a relationship, does it appear to be linear?

Not applicable.

If the relationship appears to be linear, is it a positive or negative linear relationship?

Not applicable.

4. Scatter plot 4



Is there a relationship?

Yes.

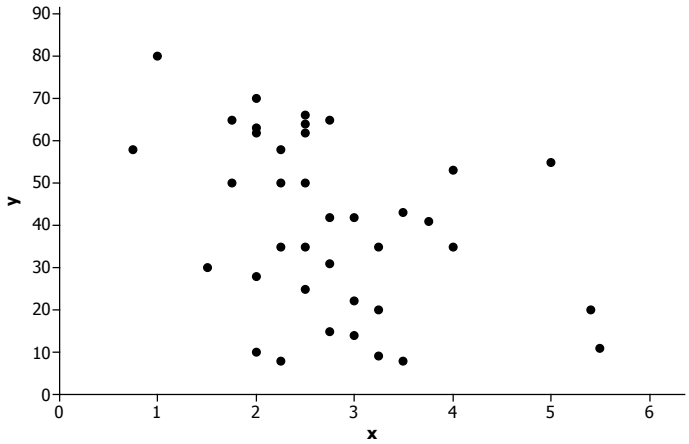
If there is a relationship, does it appear to be linear?

No.

If the relationship appears to be linear, is it a positive or negative linear relationship?

Not applicable.

5. Scatter plot 5



Is there a relationship?

Yes.

If there is a relationship, does it appear to be linear?

Yes.

If the relationship appears to be linear, is it a positive or negative linear relationship?

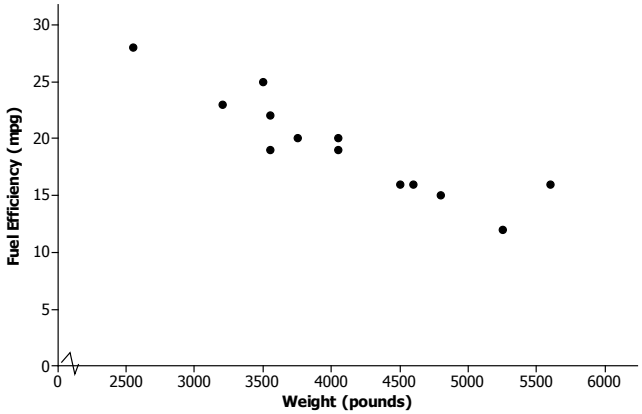
Negative.

Exercises 6–9 (10 minutes)

Let students work in pairs on Exercises 6–9. Encourage students to use terms such as linear and nonlinear and positive and negative in their descriptions. Also, remind students that their descriptions should be written making use of the context of the problem. Point out that a good description would provide answers to the three questions they answered in the previous exercises.

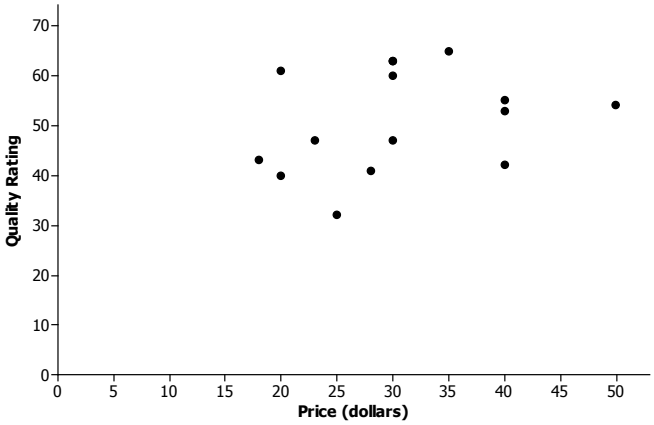
Exercises 6–9

6. Below is a scatter plot of data on weight (x) and fuel efficiency (y) for 13 cars. Using the questions at the beginning of this lesson as a guide, write a few sentences describing any possible relationship between x and y.



Possible answer: There appears to be a negative linear relationship between fuel efficiency and weight. Students may note that this is a fairly strong negative relationship. The cars with greater weight tend to have lower fuel efficiency.

7. Below is a scatter plot of data on price (x) and quality rating (y) for 14 bike helmets. Using the questions at the beginning of this lesson as a guide, write a few sentences describing any possible relationship between x and y.

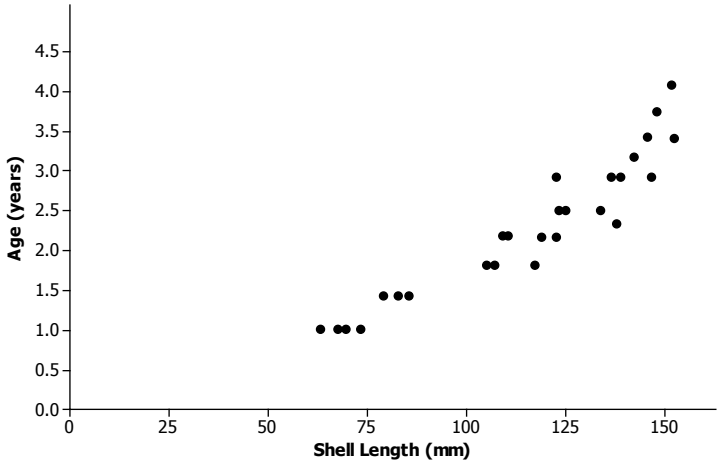


Possible answer: There does not appear to be a relationship between quality rating and price. The points in the scatter plot appear to be scattered at random, and there is no apparent pattern in the scatter plot.

Scaffolding:

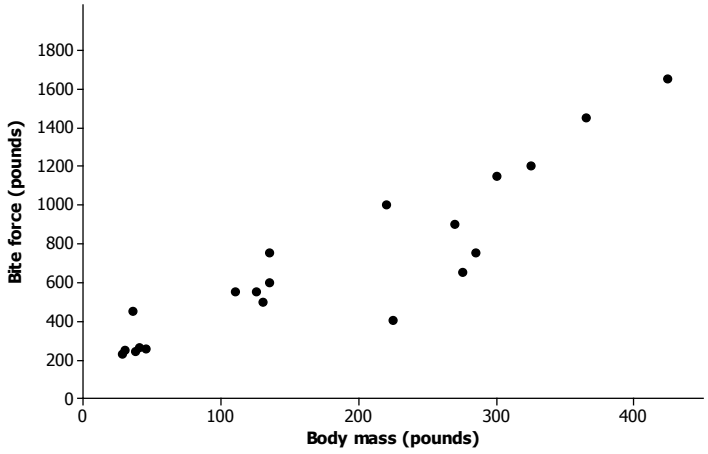
- It may be helpful to provide sentence frames on the classroom board to help students articulate their observations.
- For example, "I see a negative or positive linear relationship between ____ and ____. The higher or lower the ____, the higher or lower the ____."

8. Below is a scatter plot of data on shell length (x) and age (y) for 27 lobsters of known age. Using the questions at the beginning of this lesson as a guide, write a few sentences describing any possible relationship between x and y.



Possible answer: There appears to be a relationship between shell length and age, but the pattern in the scatter plot is curved rather than linear. Age appears to increase as shell length increases, but the increase is not at a constant rate.

9. Below is a scatter plot of data from crocodiles on body mass (x) and bite force (y). Using the questions at the beginning of this lesson as a guide, write a few sentences describing any possible relationship between x and y.



Possible answer: There appears to be a positive linear relationship between bite force and body mass. For crocodiles, the greater the body mass, the greater the bite force tends to be. Students may notice that this is a positive relationship, but not quite as "strong" as the relationship noted in Exercise 6.

Example 2: Clusters and Outliers (5 minutes)

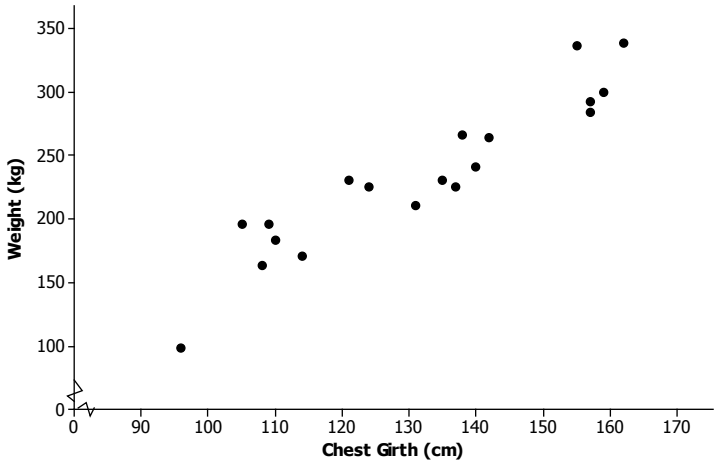
Spend a few minutes introducing the meaning of the terms *clusters* and *outliers* in the context of scatter plots. You might ask students to sketch a scatter plot that has an outlier and a scatter plot that has two clusters as a way of checking their understanding of these terms before moving on to the exercises that follow.

Example 2

In addition to looking for a general pattern in a scatter plot, you should also look for other interesting features that might help you understand the relationship between two variables. Two things to watch for are as follows:

- **Clusters: Usually the points in a scatter plot form a single cloud of points, but sometimes the points may form two or more distinct clouds of points. These clouds are called clusters. Investigating these clusters may tell you something useful about the data.**
- **Outliers: An outlier is an unusual point in a scatter plot that does not seem to fit the general pattern or that is far away from the other points in the scatter plot.**

The scatter plot below was constructed using data from a study of Rocky Mountain elk ("Estimating Elk Weight from Chest Girth," *Wildlife Society Bulletin*, 1996). The variables studied were chest girth in cm (x) and weight in kg (y).



Scaffolding:

- ELL students will need the chance to practice using the terms *clusters* and *outliers* in both oral and written contexts. Sentence frames may be useful for students to communicate initial ideas.

Scaffolding:

- The terms "elk" and "girth" may not be familiar to ELL students.
- An "elk" is a large mammal, similar to a deer.
- "Girth" refers to the measurement around something. For this problem, girth refers to the measurement around the elk from behind the front legs and under the belly. A visual aid of an elk (found on several websites) would help explain an elk's chest girth.
- Consider providing students with sentence frames or word banks, and allow students to respond in their first language to these exercises.

Exercises 10–12 (8 minutes)

Have students work individually or in pairs on Exercises 10–12. Then, have students share answers to these exercises and discuss any of the exercises where there is disagreement on the answers.

Exercises 10–12

10. Do you notice any point in the scatter plot of elk weight versus chest girth that might be described as an outlier? If so, which one?

Possible answer: The point in the lower left hand corner of the plot corresponding to an elk with a chest girth of about 96 cm and a weight of about 100 kg could be described as an outlier. There are no other points in the scatter plot that are near this one.

11. **If you identified an outlier in Exercise 10, write a sentence describing how this data observation differs from the others in the data set.**

    Possible answer: This point corresponds to an observation for an elk that is much smaller, both in terms of chest girth and weight, than the other elk in the data set.

12. **Do you notice any clusters in the scatter plot? If so, how would you distinguish between the clusters in terms of chest girth? Can you think of a reason these clusters might have occurred?**

    Possible answer: Other than the outlier, there appear to be three clusters of points. One cluster corresponds to elk with chest girths between about 105 cm and 115 cm. A second cluster includes elk with chest girths between about 120 cm and 145 cm. The third cluster includes elk with chest girths above 150 cm. It may be that age and sex play a role. Maybe the cluster with the smaller chest girths includes young elk. The two other clusters might correspond to females and males if there is a difference in size for the two sexes for Rocky Mountain elk. If we had data on age and sex, we could investigate this further.

Closing (3–5 minutes)

Consider posing the following questions; allow a few student responses for each.

- Why do you think it is a good idea to look at a scatter plot when you have data on two numerical variables?
  - *Possible response: Looking at a scatter plot makes it easier to see if there is a relationship between the two variables. It is hard to determine if there is a relationship when you just have the data in a table or a list.*
- What should you look for when you are looking at a scatter plot?
  - *Possible response: First, you should look for any general patterns. If there are patterns, you then want to consider whether the pattern is linear or nonlinear, and if it is linear, whether the relationship is positive or negative. Finally, it is also a good idea to look for any other interesting features, such as outliers or clusters. The closer the points are to a line, the "stronger" the linear relationship.*

Scaffolding:

- Allowing ELL students to brainstorm with a partner first may elicit a greater response in the whole-group setting.

Lesson Summary

- **A scatter plot might show a linear relationship, a nonlinear relationship, or no relationship.**
- **A positive linear relationship is one that would be modeled using a line with a positive slope. A negative linear relationship is one that would be modeled by a line with a negative slope.**
- **Outliers in a scatter plot are unusual points that do not seem to fit the general pattern in the plot or that are far away from the other points in the scatter plot.**
- **Clusters occur when the points in the scatter plot appear to form two or more distinct clouds of points.**

Exit Ticket (5 minutes)

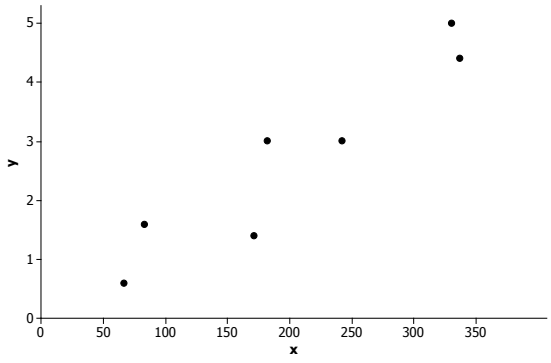
Name _____ Date_____

Lesson 7: Patterns in Scatter Plots

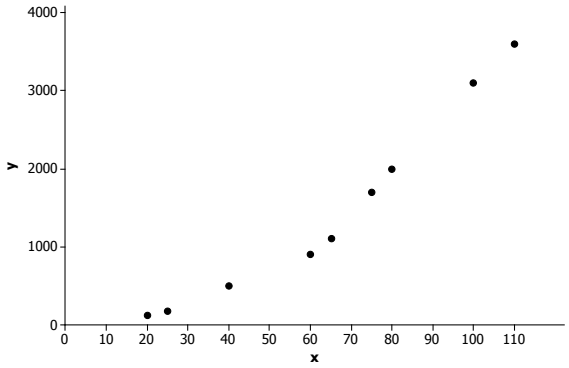
Exit Ticket

1. Which of the following scatter plots shows a negative linear relationship? Explain how you know.

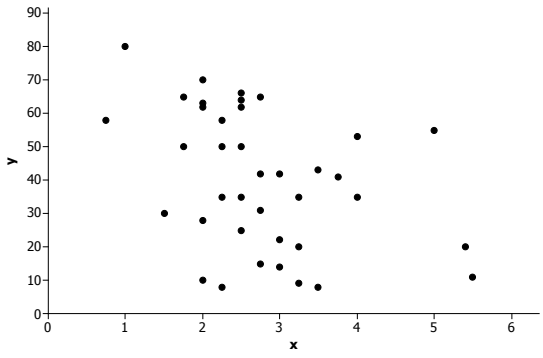
Scatter plot 1



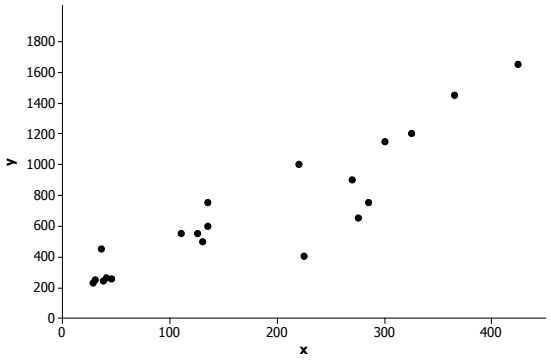
Scatter plot 2



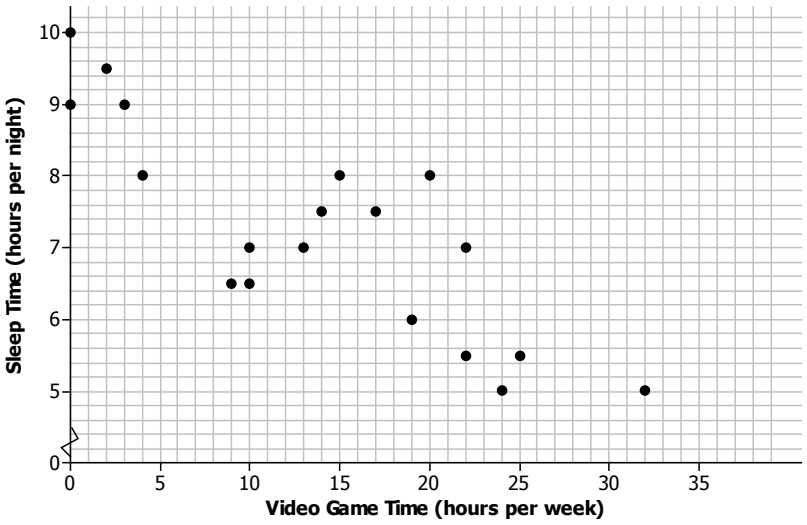
Scatter plot 3



Scatter plot 4



2. The scatter plot below was constructed using data from eighth-grade students on time spent playing video games per week (x) and number of hours of sleep per night (y). Write a few sentences describing the relationship between sleep time and time spent playing video games for these students. Are there any noticeable clusters or outliers?

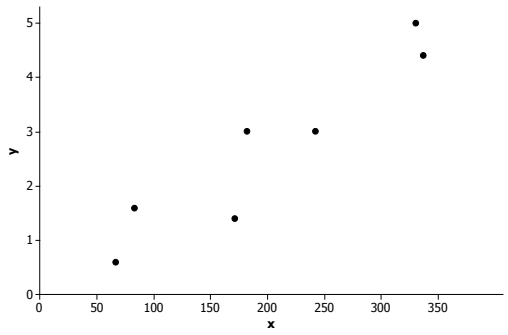


3. In a scatter plot, if the values of y tend to increase as the value of x increases, would you say that there is a positive relationship or a negative relationship between x and y? Explain your answer.

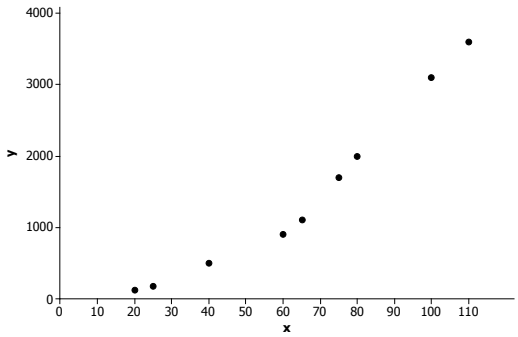
Exit Ticket Sample Solutions

1. Which of the following scatter plots shows a negative linear relationship? Explain how you know.

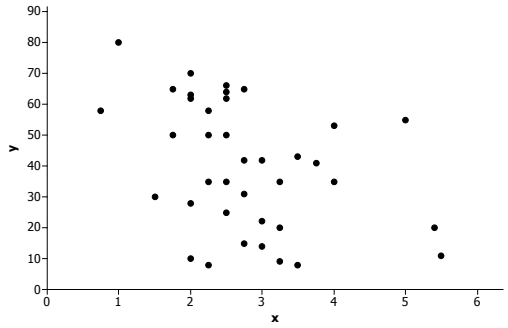
Scatter plot 1



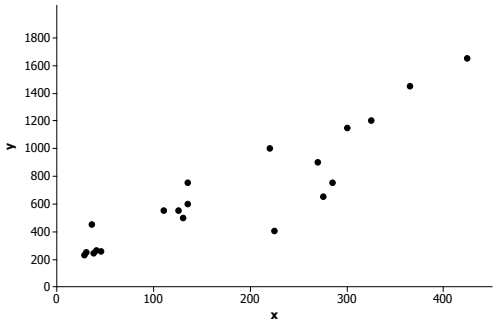
Scatter plot 2



Scatter plot 3

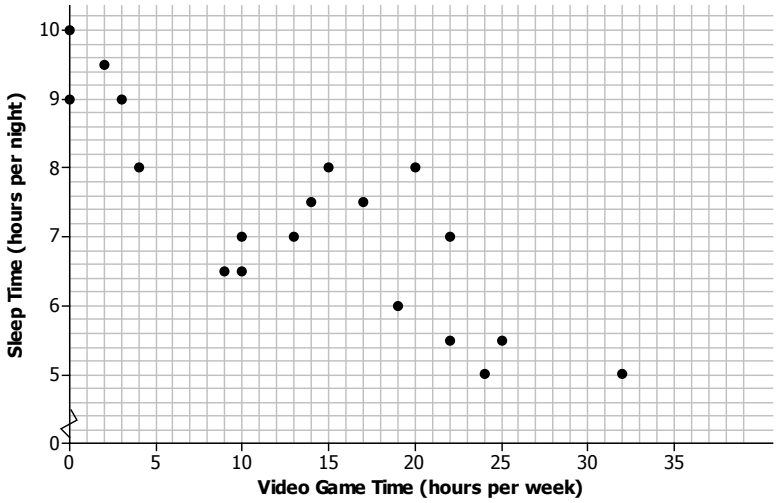


Scatter plot 4



Scatter plot 3, because it is the only one where the y-values tend to decrease as the value of x increases.

2. The scatter plot below was constructed using data from eighth-grade students on time spent playing video games per week (x) and number of hours of sleep per night (y). Write a few sentences describing the relationship between sleep time and time spent playing video games for these students. Are there any noticeable clusters or outliers?



Answers will vary. Sample response: There appears to be a negative linear relationship between the number of hours per week a student plays video games and the number of hours per night the student sleeps. As video game time increases, number of hours of sleep tends to decrease. There is one observation that might be considered an outlier—the point corresponding to a student who plays video games 32 hours per week. Other than the outlier, there are two clusters—one corresponding to students who spend very little time playing video games and a second corresponding to students who play video games between about 10 and 25 hours per week.

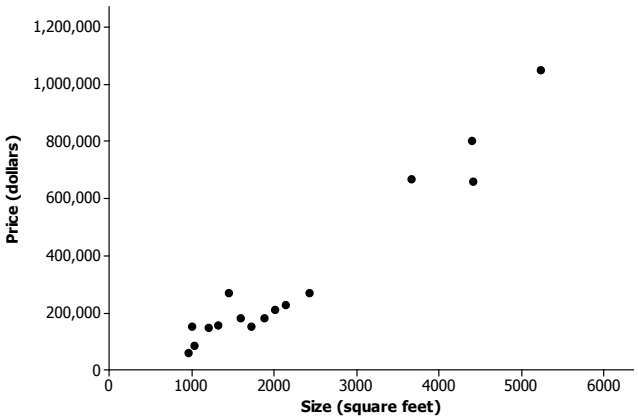
3. In a scatter plot, if the value of y tends to increase as the value of x increases, would you say that there is a positive relationship or a negative relationship between x and y?

A positive relationship. If the value of y increases as the value of x increases, the points go up on the scatter plot as you go left to right.

Problem Set Sample Solutions

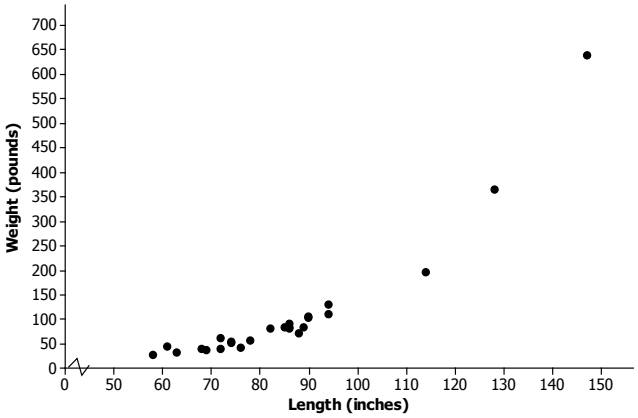
The Problem Set is intended to reinforce material from the lesson and have students think about the meaning of points in a scatter plot, clusters, positive and negative linear trends, and trends that are not linear.

1. **The scatter plot below was constructed using data size in square feet (x) and price in dollars (y) for a sample of houses. Write a few sentences describing the relationship between price and size for these houses. Are there any noticeable clusters or outliers?**



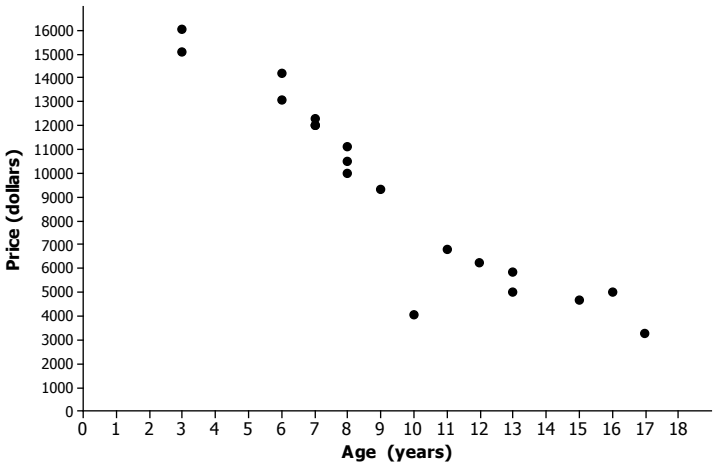
Answers will vary. Possible response: There appears to be a positive linear relationship between size and price. Price tends to increase as size increases. There appear to be two clusters of houses—one that includes houses that are less than $3,000$ square feet in size and another that includes houses that are more than $3,000$ square feet in size.

2. **The scatter plot below was constructed using data on length in inches (x) and weight in pounds (y) for a sample of alligators. Write a few sentences describing the relationship between weight and length for these alligators. Are there any noticeable clusters or outliers?**



Answers will vary. Possible response: There appears to be a positive relationship between length and weight, but the relationship is not linear. Weight tends to increase as length increases. There are three observations that stand out as outliers. These correspond to alligators that are much bigger in terms of both length and weight than the other alligators in the sample. Without these three alligators, the relationship between length and weight would look linear. It might be possible to use a line to model the relationship between weight and length for alligators that have lengths of less than 100 inches.

3. The scatter plot below was constructed using data on age in years (x) and price in dollars (y) for a sample of Honda Civics. Write a few sentences describing the relationship between price and age for these cars. Are there any noticeable clusters or outliers?

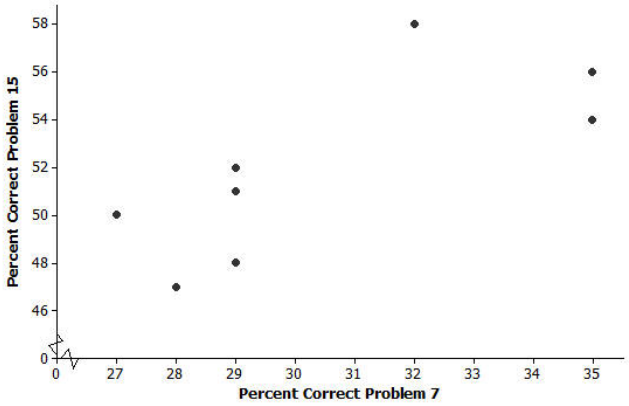


Answers will vary. Possible response: There appears to be a relatively strong negative linear relationship between price and age. Price tends to decrease as age increases. There is one car that looks like an outlier—the car that is 10 years old. This car has a price that is lower than expected based on the pattern of the other points in the scatter plot.

4. Samples of students in each of the U.S. states periodically take part in a large-scale assessment called the National Assessment of Educational Progress (NAEP). The table below shows the percent of students in the northeastern states (as defined by the U.S. Census Bureau) who answered problems 7 and 15 correctly on the 2011 eighth-grade test. The scatter plot shows the percent of eighth-grade students who got problems 7 and 15 correct on the 2011 NAEP.

State	% Correct Problem 7	% Correct Problem 15
Connecticut	29	51
New York	28	47
Rhode Island	29	52
Maine	27	50
Pennsylvania	29	48
Vermont	32	58
New Jersey	35	54
New Hampshire	29	52
Massachusetts	35	56

Percent Correct for Problems 7 and 15 on 2011 Eighth-Grade NAEP



a. **Why does it appear that there are only eight points in the scatter plot for nine states?**

Two of the states, New Hampshire and Rhode Island, had exactly the same percent correct on each of the questions, (29, 52).

b. **What is true of the states represented by the cluster of five points in the lower left corner of the graph?**

Answers will vary; those states had lower percentages correct than the other three states in the upper right.

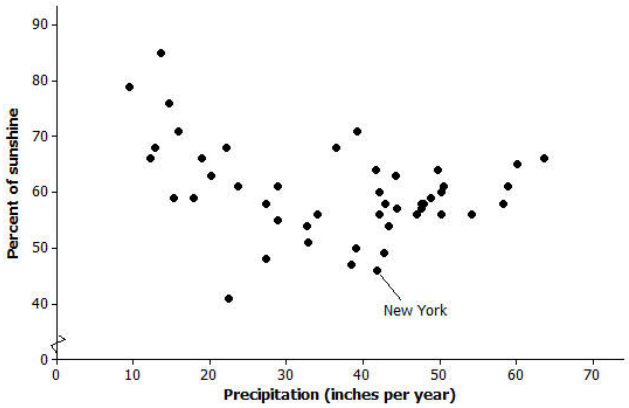
c. **Which state did the best on these two problems? Explain your reasoning.**

Answers will vary; some students might argue that Massachusetts at (35, 56) did the best. Even though Vermont actually did a bit better on problem 15, it was lower on problem 7.

d. **Is there a trend in the data? Explain your thinking.**

Answers will vary; there seems to be a weak positive linear trend as a large percent correct on one question suggests a large percent correct on the other, and a low percent on one suggests a low percent on the other.

5. The plot below shows the mean percent of sunshine during the year and the mean amount of precipitation in inches per year for the states in the United States.



Data Source: www.currentresults.com/Weather/US/average-annual-state-sunshine.php
www.currentresults.com/Weather/US/average-annual-state-precipitation.php

a. Where on the graph are the states that have a large amount of precipitation and a small percent of sunshine?

Those states will be in the lower right hand corner of the graph.

b. New York State is the point $(46, 41.8)$. Describe how the mean amount of precipitation and percent of sunshine in New York compare to the rest of the United States.

New York has a little over 40 inches of precipitation per year and is sunny about 45% of the time. It has a smaller percent of sunshine over the year than most states and is about in the middle of the states in terms of the amount of precipitation, which goes from about 10 to 65 inches per year.

c. Write a few sentences describing the relationship between mean amount of precipitation and percent of sunshine.

There is a weak negative relationship, or the more precipitation, the less percent of sun. If you took away the three states at the top left with a large percent of sun and very little precipitation, the trend would not be as pronounced. The relationship is not linear.

6. Think about shaking hands with people.

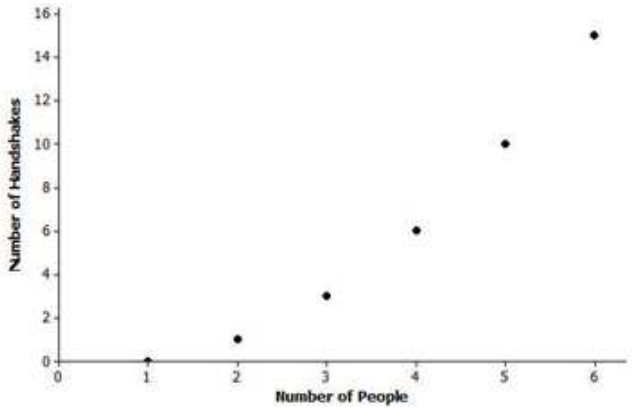
a. If three people are in a room and everyone shakes hands with everyone else, how many handshakes will there be?

Three handshakes.

b. Make a table for the number of handshakes in the room for one to six people. You may want to make a diagram or list to help you count the number of handshakes.

Number people	Handshakes		Number people	Handshakes
1	0		6	15
2	1			
3	3			
4	6			
5	10			

c. Make a scatter plot of number of people (x) and number of handshakes (y). Explain your thinking.



d. Does the trend seem to be linear? Why or why not?

The trend is increasing, but it is not linear. As the number of people increases, the number of handshakes also increases. It does not increase at a constant rate.

Lesson 8: Informally Fitting a Line

Student Outcomes

- Students informally fit a straight line to data displayed in a scatter plot.
- Students make predictions based on the graph of a line that has been fit to data.

Lesson Notes

In this lesson, students investigate scatter plots of data and informally fit a line to the pattern observed in the plot. Students then make predictions based on their line. Students informally evaluate their predictions based on the fit of the line to the data.

Classwork

Example 1 (2–3 minutes): Housing Costs

Introduce the data presented in the table and the scatter plot of the data. Ask students:

- Examine the scatter plot. What trend do you see? How would you describe this trend?
  - *It appears to be a positive linear trend. The scatter indicates that the larger the size, the higher the price.*

MP.2 & MP.7

Scaffolding:

- The terms "house" and "home" are used interchangeably throughout the example.
- This may be confusing for ELL students and should be clarified.

(Note: Ensure students are provided with an opportunity to explain why they think there is a positive linear trend between price and size.)

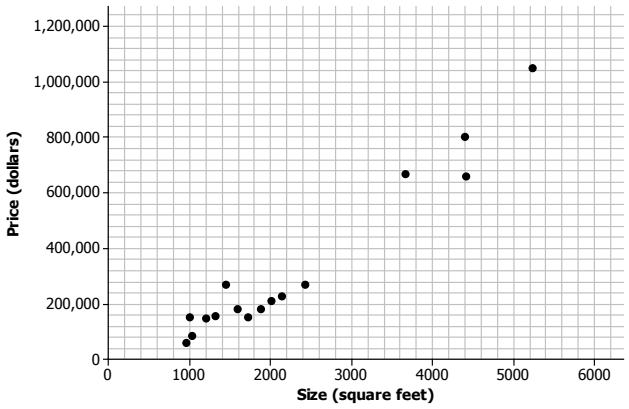
Example 1: Housing Costs

Let's look at some data from one midwestern city that indicates the sizes and sale prices of various houses sold in this city.

Size (square feet)	Price (dollars)		Size (square feet)	Price (dollars)
5,232	1,050,000		1,196	144,900
1,875	179,900		1,719	149,900
1,031	84,900		956	59,900
1,437	269,900		991	149,900
4,400	799,900		1,312	154,900
2,000	209,900		4,417	659,999
2,132	224,900		3,664	669,000
1,591	179,900		2,421	269,900

Data Source: http://www.trulia.com/for_sale/Milwaukee,WI/5_p accessed 7/13/2013

A scatter plot of the data is given below.



Exercises 1–6 (15 minutes)

MP.6

In these exercises, be sure that students retain the units as they write and discuss the solutions, being mindful of the mathematical practice standard of attending to precision. Students might use a transparent ruler or a piece of uncooked spaghetti to help draw and decide where to place their lines. To avoid problems with the size of the numbers and to have students focus on drawing their lines, the teacher should provide a worksheet for students with the points already plotted on a grid. Students should concentrate on the general form of the scatter plot rather than worrying too much about the exact placement of points in the scatter plot. The primary focus of the work in these exercises is to have students think about the trend, use a line to describe the trend, and make predictions based on the line.

Work through the exercises as a class, allowing time to discuss multiple responses.

Exercises 1–6

1. **What can you tell about the price of large homes compared to the price of small homes from the table?**

   Answers will vary. Students should make the observation that, overall, the larger homes cost more and the smaller homes cost less. However, it is hard to generalize because one of the smaller homes costs nearly $150,000.

2. **Use the scatter plot to answer the following questions.**

   a. **Does the plot seem to support the statement that larger houses tend to cost more? Explain your thinking.**

      Yes, because the trend is positive; the larger the size of the house, the more the house tends to cost.

   b. **What is the cost of the most expensive house, and where is that point on the scatter plot?**

      The house with a size of 5,232 square feet costs $1,050,000, which is the most expensive. It is in the upper right corner of the scatter plot.

   c. **Some people might consider a given amount of money and then predict what size house they could buy. Others might consider what size house they want and then predict how much it would cost. How would you use the above scatter plot?**

      Answers will vary. Since the size of the house is on the horizontal axis and the price is on the vertical axis, the scatter plot is set up with price as the dependent variable and size as the independent variable. This is the way you would set it up if you wanted to predict price based on size. Although various answers are appropriate, move the discussion along using size to predict price.

d. **Estimate the cost of a $3,000$ square foot house.**

Answers will vary. Reasonable answers range between $\$300,000$ and $\$600,000$.

e. **Do you think a line would provide a reasonable way to describe how price and size are related? How could you use a line to predict the price of house if you are given its size?**

Answer will vary; however, use this question to develop the idea that a line would provide a way to estimate the cost given the size of a house. The challenge is how to make that line. Note: Students are encouraged in the next exercise to first make a line, and then evaluate whether or not it fits the data. This will provide a reasonable estimate of the cost of a house in relation to its size.

3. **Draw a line in the plot that you think would fit the trend in the data.**

Answers will vary. Discuss several of the lines students have drawn by encouraging students to share their lines with the class. At this point, do not evaluate the lines as good or bad. Students may want to know a precise procedure or process to draw their lines. If that question comes up, indicate to students that a procedure will be developed in their future work (Grade 9) with statistics. For now, the goal is to simply draw a line that can be used to describe the relationship between the size of a home and its cost. Indicate that strategies for drawing a line will be explored in Exercise 5. Use the lines provided by students to evaluate the predictions in the following exercise. These predictions will be used to develop a strategy for drawing a line. Use the line drawn by students to highlight their understanding of the data.

MP.1

4. **Use your line to answer the following questions:**

a. **What is your prediction of the price of a $3,000$ square foot house?**

Answers will vary. A reasonable prediction is around $\$500,000$.

b. **What is the prediction of the price of a $1,500$ square foot house?**

Answers will vary. A reasonable prediction is around $\$200,000$.

Scaffolding:

- Point out to students that the word "trend" is not connected to the use of this word in describing fashion or music. (For example, "the trend in music is for more use of drums.")
- In this lesson, "trend" describes the pattern or lack of a pattern in the scatter plot.
- Ask students to highlight words that they think would describe a trend in the scatter plots that are examined in this lesson.
- For ELL students, it should be explained that the scatter plot may be referred to as just "plot."

Display various predictions students found for these two examples. You might use a chart similar to the following to discuss the different predictions.

Student	*Estimate of the price for a* 3,000 *square foot house*	*Estimate of the price for a* 1,500 *square foot house*
Student 1	$300,000	$100,000
Student 2	$600,000	$400,000

MP.1

Discuss that predictions vary as a result of the different lines that students used to describe the pattern in the scatter plot. What line makes the most sense for this data?

Before you discuss answers to that question, encourage students to explain how they drew their line and why their predictions might have been higher (or lower) than other students. For example, students with lines that are visibly above most of the points may have predictions that are higher than the predictions of students with lines below several of the points. Ask students to summarize their theories of how to draw a line as "strategy" for drawing a line. After they provide their own descriptions, provide students an opportunity to think about the following strategies that might have been used to draw a line.

5. Consider the following general strategies used by students for drawing a line. Do you think they represent a good strategy in drawing a line that will fit the data? Explain why or why not, or draw a line for the scatter plot using the strategy that would indicate why it is or why it is not a good strategy.

   a. Laure thought she might draw her line using the very first point (farthest to the left) and the very last point (farthest to the right) in the scatter plot.

      Answers will vary. This may work in some cases, but those points might not capture the trend in the data. For example, the first point in the lower left might not be in line with the other points.

   b. Phil wants to be sure that he has the same number of points above and below the line.

      Answers will vary. You could draw a nearly horizontal line that has half of the points above and half below, but that might not represent the trend in the data at all. Note: For many students just starting out, this seems like a reasonable strategy, but it often can result in lines that clearly do not fit the data. As indicated, drawing a nearly horizontal line is a good way to indicate that this is not a good strategy.

   c. Sandie thought she might try to get a line that had the most points right on it.

      Answers will vary. That might result in, perhaps, three points on the line (knowing it only takes two to make a line), but the others could be anywhere. The line might even go in the wrong direction. Note: For students just beginning to think of how to draw a line, this seems like a reasonable goal; however, point out that this strategy, may result in lines that are not good for predicting price.

   d. Maree decided to get her line as close to as many of the points as possible.

      Answers will vary. If you can figure out how to do this, Maree's approach seems like a reasonable way to find a line that takes all of the points into account.

6. Based on the strategies discussed in Exercise 5, would you change how you draw a line through the points? Explain your answer.

   Answers will vary based on how a student drew his or her original line. Summarize that the goal is to draw a line that is as close as possible to the points in the scatter plot. More precise methods are developed in Algebra I.

Example 2 (2–3 minutes): Deep Water

Introduce students to the data in the table. Pose the questions in the text and allow for multiple responses.

Example 2: Deep Water

Does the current in the water go faster or slower when the water is shallow? The data on the depth and speed of the Columbia River at various locations in Washington state listed below can help you think about the answer.

Depth and Velocity in the Columbia River, Washington State

Depth (feet)	Velocity (feet/second)
0.7	1.55
2.0	1.11
2.6	1.42
3.3	1.39
4.6	1.39
5.9	1.14
7.3	0.91
8.6	0.59
9.9	0.59
10.6	0.41
11.2	0.22

Data Source: www.seattlecentral.edu/qelp/sets/011/011.html

Scaffolding:

- The word "current" has multiple meanings that ELL students may be familiar with from a social studies class (current events) or from a science class (electrical current).
- In this example, the "current" refers to the flow of the river.

a. What can you tell about the relationship between the depth and velocity by looking at the numbers in the table?

Answers will vary. The depths are not in order, so it is very hard to see any pattern.

b. If you were to make a scatter plot of the data, which variable would you put on the horizontal axis and why?

Answers will vary. It might be easier to measure the depth and use that information to predict the velocity of the water, so the depth should go on the horizontal axis.

Exercises 7–9 (12–15 minutes)

MP.6

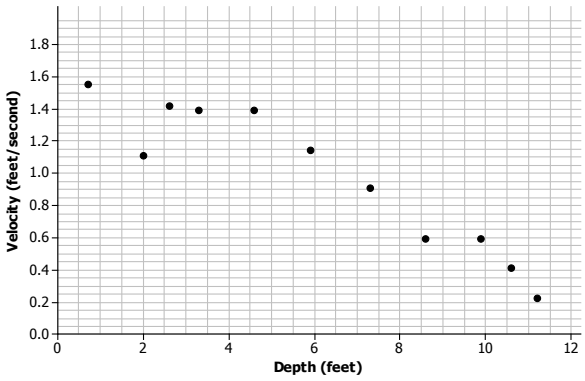
These exercises engage students in a context where the trend has a negative slope. Again, students should pay careful attention to units and interpretation of rate of change. They evaluate the line by assessing its closeness to the data points. Let students work with a partner. If time allows, discuss the answers as a class.

Scaffolding:

ELL students may need support in recognizing the relationship between the words "depth" and "deep."

Exercises 7–9

7. A scatter plot of the Columbia River data is shown below.



a. Choose one point in the scatter plot and describe what it means in terms of the context.

Answers will vary. For example, $(4.6, 1.39)$ would represent a place in the river that was 4.6 feet deep and had a velocity of 1.39 ft./sec.

b. Based on the scatter plot, describe the relationship between velocity and depth.

The deeper the water, the slower the current tends to be.

c. How would you explain the relationship between the velocity and depth of the water?

Answers will vary. Sample response: Velocity may be a result of the volume of water. Shallow water has less volume, and as a result, the water runs faster. Note: Students may have several explanations. For example, they may say that depth is a result of less water runoff; therefore, water depth increases.

d. If the river is two feet deep at a certain spot, how fast do you think the current would be? Explain your reasoning.

Answers will vary. It could be around 1.11 ft./sec. or it could be closer to 1.42 ft./sec., which is more in line with the pattern for the other points.

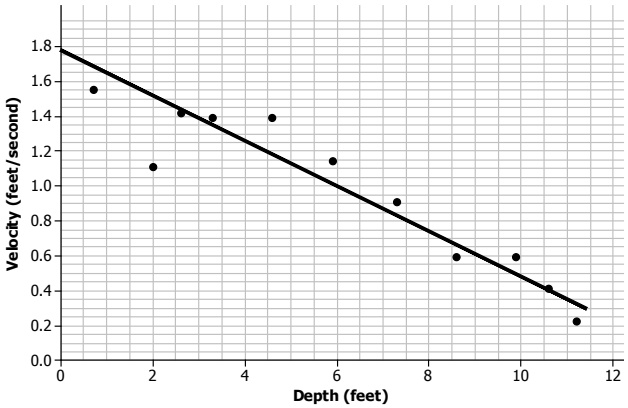
8. **Consider the following questions:**

   a. **If you draw a line to represent the trend in the plot, will it make it easier to predict the velocity of the water if you know the depth? Why or why not?**

   Answers will vary. A line will help you determine a better prediction for 1.5 ft. or 5 ft., where the points are a bit scattered.

   b. **Draw a line that you think does a reasonable job of modeling the trend on the scatter plot above. Use the line to predict the velocity when the water is 8 feet deep.**

   Answers will vary. A line is drawn in the following graph. Using this line, when the water is 8 feet deep, the velocity is predicted to be 0.76 ft./sec.



9. **Use the line to predict the velocity for a depth of 8.6 feet. How far off was your prediction from the actual observed velocity for the location that had a depth of 8.6 feet?**

   Answers will vary. Sample response: The current would be moving at 0.68 ft./sec. The observed velocity was 0.59 ft./sec., so the line predicted a velocity that was 0.07 ft./sec. slower than the observed value.

Closing (5 minutes)

Consider posing the following questions; allow a few student responses for each.

- How do scatter plots and tables of data differ in helping you understand the "story" when looking at bivariate numerical data?
  - *The numbers in a table can give you a sense of how big or small the values are, but it is easier to see a relationship between the variables in a scatter plot.*
- What is the difference between predicting an outcome by looking at a scatter plot and predicting the outcome using a line that models the trend?
  - *When you look at the plot, the points are sometimes very scattered, and for a given value of an independent variable, some values you might be interested in may not be included in the data set. Using a line takes all of the points into consideration, and your prediction is based on an overall pattern rather than just one or two points.*

- In a scatter plot, which variable goes on the horizontal axis and which goes on the vertical axis?
  - *The independent variable (or the variable not changed by other variables) goes on the horizontal axis and the dependent variable (or the variable to be predicted by the independent variable) goes on the vertical axis.*

Lesson Summary

- **When constructing a scatter plot, the variable that you want to predict (i.e., the dependent or response variable) goes on the vertical axis. The independent variable (i.e., the variable not changed by other variables) goes on the horizontal axis.**
- **When the pattern in a scatter plot is approximately linear, a line can be used to describe the linear relationship.**
- **A line that describes the relationship between a dependent variable and an independent variable can be used to make predictions of the value of the dependent variable given a value of the independent variable.**
- **When informally fitting a line, you want to find a line for which the points in the scatter plot tend to be close to the line.**

Exit Ticket (5 minutes)

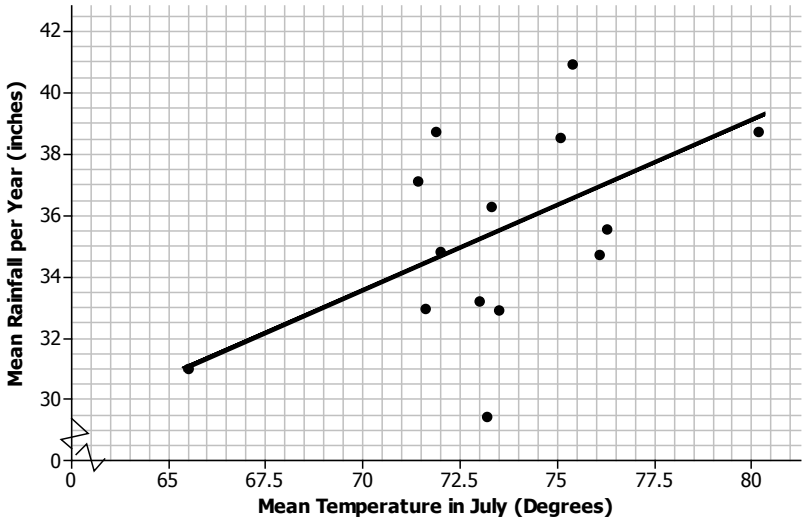
Name _____ Date_____

Lesson 8: Informally Fitting a Line

Exit Ticket

The plot below is a scatter plot of mean temperature in July and mean inches of rain per year for a sample of midwestern cities. A line is drawn to fit the data.

July Temperatures and Rainfall in Selected Midwestern Cities



Data Source: http://countrystudies.us/united-states/weather/

1. Choose a point in the scatter plot and explain what it represents.

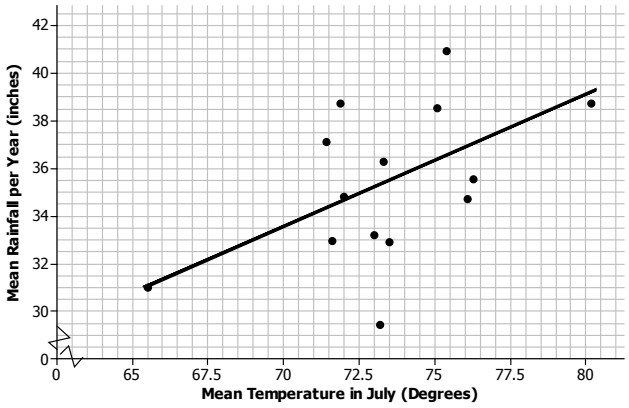
2. Use the line provided to predict the mean number of inches of rain per year for a city that has a mean temperature of 70° in July.

3. Do you think the line provided is a good one for this scatter plot? Explain your answer.

Exit Ticket Sample Solutions

The plot below is a scatter plot of mean temperature in July and mean inches of rain per year for a sample of midwestern cities.

July Temperatures and Rainfall in Selected Midwestern Cities



Data Source: http://countrystudies.us/united-states/weather/

1. **Choose a point in the scatter plot and explain what it represents.**

   Answers will vary. Sample response: The point at about $(72, 35)$ represents a midwestern city where the mean temperature in July is about $72°$ and where the rainfall per year is about 35 inches.

2. **Use the line provided to predict the mean number of inches of rain per year for a city that has a mean temperature of $70°$ in July.**

   Response: Predicted rainfall $= 33$ inches of rain per year. (Some students will state approximately 33.5 inches of rain.)

3. **Do you think the line provided is a good one for this scatter plot? Explain your answer.**

   Response: Yes, the line follows the general pattern in the scatter plot, and it does not look like there is another area in the scatterplot where the points would be any closer to the line.

Problem Set Sample Solutions

1. The table below shows the mean temperature in July and the mean amount of rainfall per year for 14 cities in the Midwest.

City	Mean Temperature in July (Degrees Fahrenheit)	Mean Rainfall per Year (inches)
Chicago, IL	73.3	36.27
Cleveland, OH	71.9	38.71
Columbus, OH	75.1	38.52
Des Moines, IA	76.1	34.72
Detroit, MI	73.5	32.89
Duluth, MN	65.5	31.00
Grand Rapids, MI	71.4	37.13
Indianapolis, IN	75.4	40.95
Marquette, MI	71.6	32.95
Milwaukee, WI	72.0	34.81
Minneapolis–St. Paul, MN	73.2	29.41
Springfield, MO	76.3	35.56
St. Louis, MO	80.2	38.75
Rapid City, SD	73.0	33.21

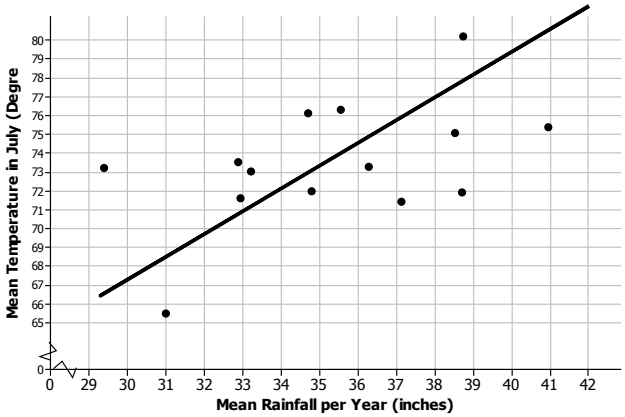
Data Source: http://countrystudies.us/united-states/weather/

a. What do you observe from looking at the data in the table?

Answers will vary. Many of the temperatures were in the 70s, and many of the mean inches of rain were in the 30s. It also appears that, in general, as the rainfall increased, the mean temperature also increased.

b. Look at the scatter plot below. A line is drawn to fit the data. The plot in the Exit Ticket had the mean July temperatures for the cities on the horizontal axis. How is this plot different, and what does it mean for the way you think about the relationship between the two variables, temperature and rain?

July Rainfall and Temperatures in Selected Midwestern Cities



This scatter plot has the labels on the axes reversed: (mean inches of rain, mean temperature). This is the scatter plot I would use if I wanted to predict the mean temperature in July knowing the mean amount of rain per year.

c. The line has been drawn to model the relationship between the amount of rain and the temperature in those midwestern cities. Use the line to predict the mean July temperature for a midwestern city that has a mean of 32 inches of rain per year.

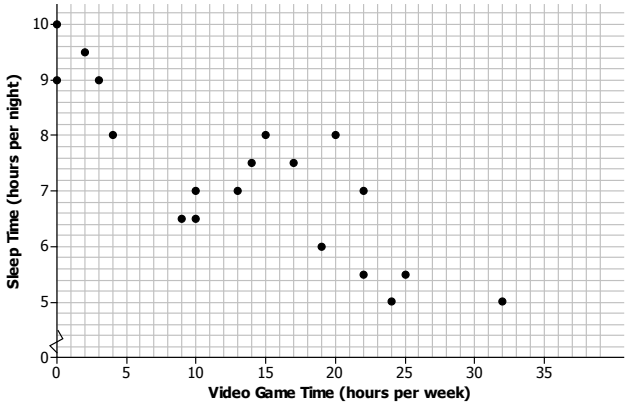
Answers will vary. For 32 inches of rain per year, the line indicates a mean July temperature of approximately 70 degrees.

d. For which of the cities in the sample will the line do the worst job of predicting the mean temperature? The best? Explain your reasoning with as much detail as possible.

Answers will vary. I looked for points that were really close to the line and ones that were far away. The line prediction for temperature would be farthest off for Minneapolis—for 29.41 in. of rain in Minneapolis, the line predicted approximately 67 degrees, whereas the actual mean temperature in July was 73.2°. The line predicted very well for Milwaukee—for 32.95 in. of rain in Milwaukee, the line predicted approximately 73 degrees, whereas the actual mean temperature in July was 72° and was only off by about 1 degree. The line was also close for Marquette—for 34.81 inches of rain in Marquette, the like predicted approximately 71 degrees, whereas the actual mean temperature in July was 71.6° and was only off by about 1 degree.

2. The scatter plot below shows the results of a survey of eighth-grade students who were asked to report the number of hours per week they spend playing video games and the typical number of hours they sleep each night.

Mean Hours Sleep per Night vs. Mean Hours Playing Video Games per Week



a. What trend do you observe in the data?

The more hours that students play video games, the fewer hours they tend to sleep.

b. What was the fewest number of hours per week that students who were surveyed spent playing video games? The most?

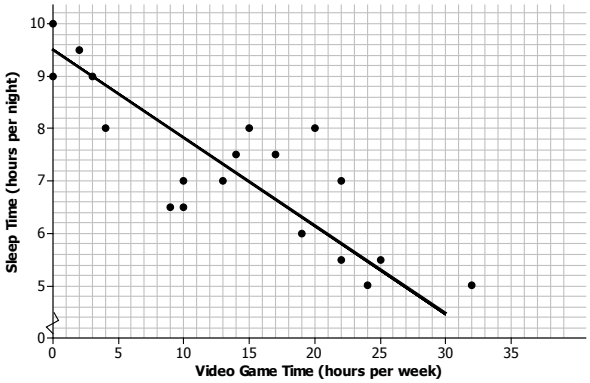
Two students spent 0 hours and one student spent 32 hours per week per week playing games.

c. What was the fewest number of hours per night that students who were surveyed typically slept? The most?

The fewest hours of sleep per night was around 5 hours and the most was around 10 hours.

d. Draw a line that seems to fit the trend in the data and find its equation. Use the line to predict the number of hours of sleep for a student who spends about 15 hours per week playing video games.

Answers will vary. A student who played 15 hours per week playing games would get about 7 hours of sleep per night.



3. Scientists can take very good pictures of alligators from airplanes or helicopters. Scientists in Florida are interested in studying the relationship between the length and the weight of alligators in the waters around Florida.

a. Would it be easier to collect data on length or weight? Explain your thinking.

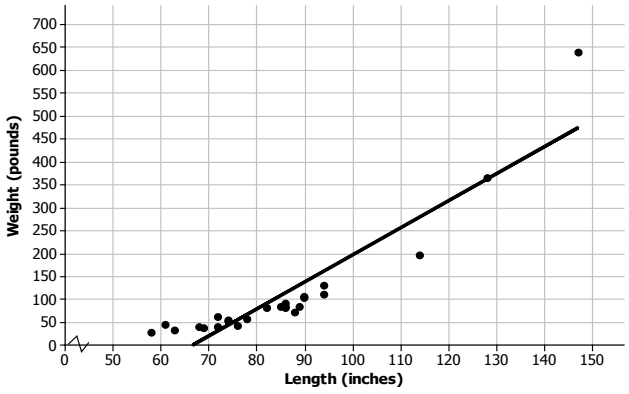
Answers will vary. You could measure the length from the pictures, but you would have to actually have the alligators to weigh them.

b. Use your answer to decide which variable you would want to put on the horizontal axis and which variable you might want to predict.

You would probably want to predict the weight of the alligator knowing the length; therefore, the length would go on the horizontal axis and the weight on the vertical axis.

4. Scientists captured a small sample of alligators and measured both their length (in inches) and weight (in pounds). Torre used their data to create the following scatter plot and drew a line to capture the trend in the data. She and Steve then had a discussion about the way the line fit the data. What do you think they were discussing and why?

Alligator length (in.) and weight (lb.)



Data Source: http://exploringdata.net/stories.htm#alligatr

Answers will vary. Sample response: The pattern in the scatter plot is curved instead of linear. All of the data points in the middle of the scatter plot fall below the line, and the line does not really capture the pattern in the scatter plot. A line does not pass through the cluster of points between 60 to 80 inches in length that fit the other points. A model other than a line might be a better fit.

Lesson 9: Determining the Equation of a Line Fit to Data

Student Outcomes

- Students informally fit a straight line to data displayed in a scatter plot.
- Students determine the equation of a line fit to data.
- Students make predictions based on the equation of a line fit to data.

Lesson Notes

In this lesson, students informally fit a line to data by drawing a line that describes a linear pattern in a scatter plot and then using their lines to make predictions. They determine the equation of the line and informally analyze different lines fit to the same data. This lesson begins developing ideas that are the foundation for finding an objective way to judge how well a line fits the trend in a scatter plot and the notion of a "best-fit" line in Algebra I.

Classwork

Example 1 (5 minutes): Crocodiles and Alligators

Discuss the data presented in the table and scatter plot. You might start by asking if students are familiar with crocodiles and alligators and how they differ. Ask students if they can imagine what a bite force of 100 pounds would feel like. Ask them if they know what body mass indicates. If students understand that body mass is an indication of the weight of a crocodile and bite force is a measure of the strength of a crocodile's bite, the data can be investigated even if they do not understand the technical definitions and how these variables are measured. Also, ask students if any other aspects of the data surprised them. For example, did they realize that there are so many different species of crocodiles? Did the wide range of body mass and bite force surprise them? If time permits, you may want to suggest that students do further research on crocodiles.

Example 1: Crocodiles and Alligators

Scientists are interested in finding out how different species adapt to finding food sources. One group studied crocodiles to find out how their bite force was related to body mass and diet. The table below displays the information they collected on body mass (in pounds) and bite force (in pounds).

Crocodilian Biting

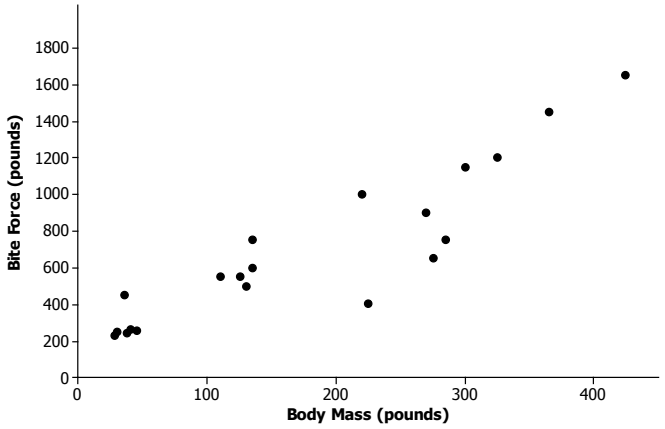
Species	Body mass (pounds)	Bite force (pounds)
Dwarf Crocodile	35	450
Crocodile F	40	260
Alligator A	30	250
Caiman A	28	230
Caiman B	37	240
Caiman C	45	255
Croc A	110	550
Nile Crocodile	275	650
Croc B	130	500
Croc C	135	600
Croc D	135	750
Caiman D	125	550
Indian Gharial Croc	225	400
Crocodile G	220	1,000
American Croc	270	900
Croc D	285	750
Croc E	425	1,650
American Alligator	300	1,150
Alligator B	325	1,200
Alligator C	365	1,450

Data Source: PLoS One Greg Erickson biomechanics, Florida State University

Scaffolding:

- The word "crocodilian" refers to any reptile of the order Crocodylia.
- This includes crocodiles, alligators, caimans, and gavials. Showing students a visual aid with pictures of these animals may aid in their comprehension.

As you learned in the previous lesson, it is a good idea to begin by looking at what a scatter plot tells you about the data. The scatter plot below displays the data on body mass and bite force for the crocodiles in the study.



Exercises 1–5 (15 minutes)

Exercises 1 through 5 ask students to consider the fit of a line. Each student (or small group of students) draws a line that would be a good representation of the trend in the data. Students evaluate their lines and the lines of the four students introduced in Exercise 4.

In Exercise 2, students draw a line they think will be a good representation of the trend in the data. Ask them to compare their line with other students. As a group, decide who might have the best line, and ask why students think a particular line is a better line. Have groups share their ideas. Point out that it would be helpful to agree on a standard method for judging the fit of a line. One method is to look at how well the line predicts for the given data or how often it is over or under the actual or observed value.

Scaffolding:

Point out to ELL students that the terms "body mass" and "weight" are used interchangeably in this lesson.

Exercises 1–5

1. **Describe the relationship between body mass and bite force for the crocodiles shown in the scatter plot.**

   As the body mass increases, the bite force tends to also increase.

2. **Draw a line to represent the trend in the data. Comment on what you considered in drawing your line.**

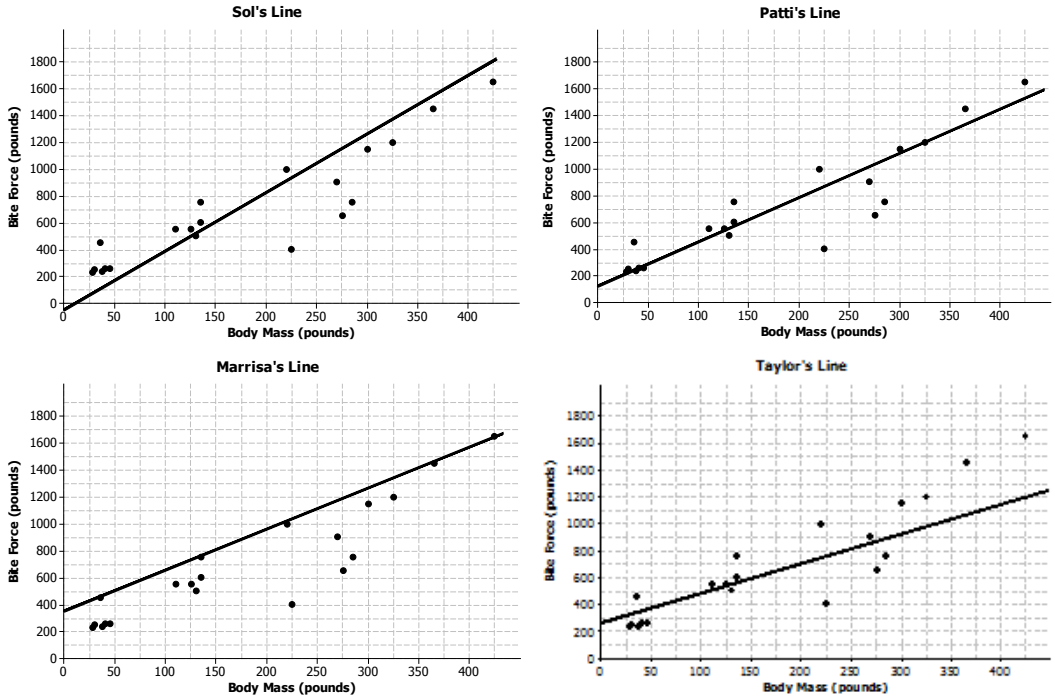
   The line should be as close as possible to the points in the scatter plot. Students explored this idea in Lesson 8.

3. **Based on your line, predict the bite force for a crocodile that weighs 220 pounds. How does this prediction compare to the actual bite force of the crocodile in the data set that weighed 220 pounds?**

   Answers will vary. A reasonable prediction is around 650 to 700 pounds. The actual bite force was 1,000 pounds, so the prediction based on the line was not very close for this crocodile.

MP.2

4. **Several students decided to draw lines to represent the trend in the data. Consider the lines drawn by Sol, Patti, Marrisa, and Taylor, which are shown below.**



For each student, indicate whether or not you think the line would be a good line to use to make predictions. Explain your thinking.

a. Sol's line

In general, it looks like Sol's line overestimates the bite force for heavier crocodiles and underestimates the bite force for crocodiles that do not weigh as much.

b. Patti's line

Patti's line looks like it fits the data well, so it would probably produce good predictions. The line goes through the middle of the points in the scatter plot, and the points are fairly close to the line.

c. Marrisa's line

It looks like Marrisa's line overestimates the bite force because almost all of the points are below the line.

d. Taylor's line

It looks like Taylor's line tends to underestimate the bite force. There are many points above the line.

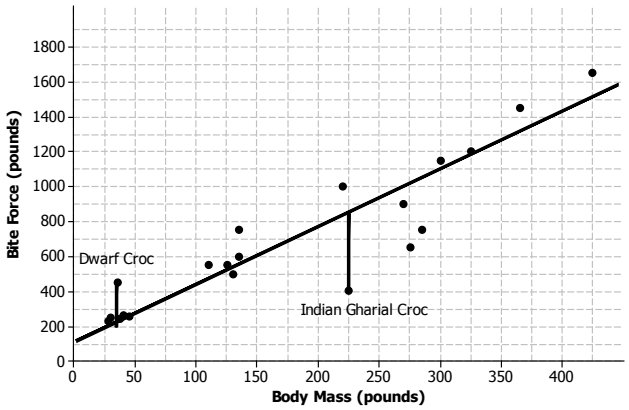
5. What is the equation of your line? Show the steps you used to determine your line. Based on your equation, what is your prediction for the bite force of a crocodile with a bite force of 200 pounds?

Answers will vary. Students have learned from previous modules how to find the equation of a line. Anticipate that students will first determine the slope based on two points on their lines. Students then use a point on the line to obtain an equation in the form $y = mx + b$ (or $y = a + bx$). Students will use their line to predict a bite force for a crocodile weighing 200 pounds. A reasonable answer would be around 800 pounds.

Exercise 6 (5 minutes)

Exercise 6

6. Patti drew vertical line segments from two points to the line in her scatter plot. The first point she selected was for a Dwarf Croc. The second point she selected was for an Indian Gharial Crocodile.



a. Would Patti's line have resulted in a predicted bite force that was closer to the actual bite force for the Dwarf Crocodile or for the Indian Gharial Crocodile? What aspect of the scatter plot supports your answer?

The prediction would be closer to the actual bite force for the Dwarf Crocodile. That point is closer to the line (the vertical line segment connecting it to the line is shorter) than the point for the Indian Gharial Crocodile.

b. Would it be preferable to describe the trend in a scatter plot using a line that makes the differences in the actual and predicted values large or small? Explain your answer.

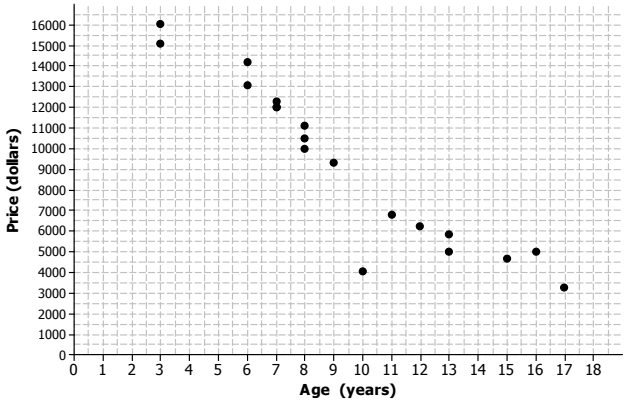
It would be better for the differences to be as small as possible. Small differences are closer to the line.

Exercise 7 (15 minutes): Used Cars

This exercise provides additional practice for students. Students use the equation of a line to make predictions and informally assess the fit of the line.

Exercise 7: Used Cars

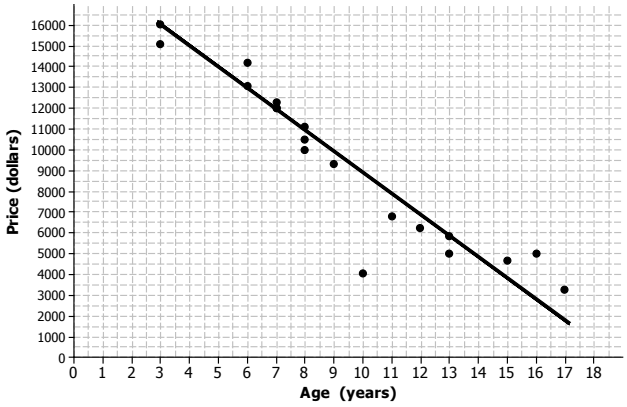
7. The plot below shows the age (in years) and price (in dollars) of used Honda Civic cars that were advertised in a local newspaper.



a. Based on the scatter plot above, describe the relationship between the age and price of the used cars.

The older the car, the lower the price tends to be.

b. Nora drew a line she thought was close to many of the points and found the equation of the line. She used the points $(13, 6000)$ and $(7, 12000)$ on her line to find the equation. Explain why those points made finding the equation easy.



The points are at the intersection of the grid lines in the graph, so it is easy to determine the coordinates of these points.

c. Find the equation of Nora's line for predicting the price of a used car given its age. Summarize the trend described by this equation.

Using the points, the equation is $y = -1,000x + 19,000$, or $Price = -1,000(age) + 19,000$. The slope of the line is negative, so the line indicates that the price of used cars decreases as cars get older.

d. For which car in the data set would the predicted value based on the line be farthest from the actual value? How can you tell?

It would be farthest for the car that is 10 years old. It is the point in the scatter plot that is farthest from the line.

e. What does the equation predict for the cost of a 10-year-old car? How close was the prediction using the line to the actual cost of the 10-year-old car in the data set? Given the context of the data set, do you think the difference between the predicted price and the actual price is large or small?

The line predicts a 10-year-old car would cost about $\$9,000$. $-1,000(10) + 19,000 = 9,000$. Compared to $\$4,040$ for the 10-year-old car in the data set, the difference would be $\$4,960$. The prediction is off by about $\$5,000$, which seems like a lot of money, given the prices of the cars in the data set.

f. Is $\$5,000$ typical of the differences between predicted prices and actual prices for the cars in this data set? Justify your answer.

No, most of the differences would be much smaller than $\$5,000$. Most of the points are much closer to the line, and most predictions would be within about $\$1,000$ of the actual value.

Closing (3–5 minutes)

- When you use a line to describe a linear relationship in a data set, what are characteristics of a good fit?
  - *The line should be as close as possible to the points in the scatter plot. The line should go through the "middle" of the points.*

Lesson Summary

- **A line can be used to represent the trend in a scatter plot.**
- **Evaluating the equation of the line for a value of the independent variable will determine a value predicted by the line.**
- **A good line for prediction is one that goes through the middle of the points in a scatter plot and for which the points tend to fall close to the line.**

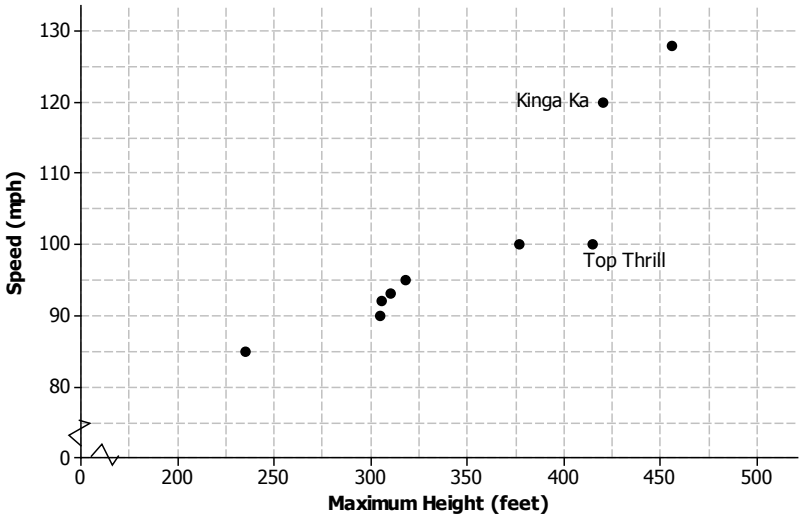
Exit Ticket (5 minutes)

Name _____ Date_____

Lesson 9: Determining the Equation of a Line Fit to Data

Exit Ticket

1. A scatter plot of the height and speed of some of the world's fastest roller coaster rides is indicated below. Draw a line that you think is a good fit for the data.



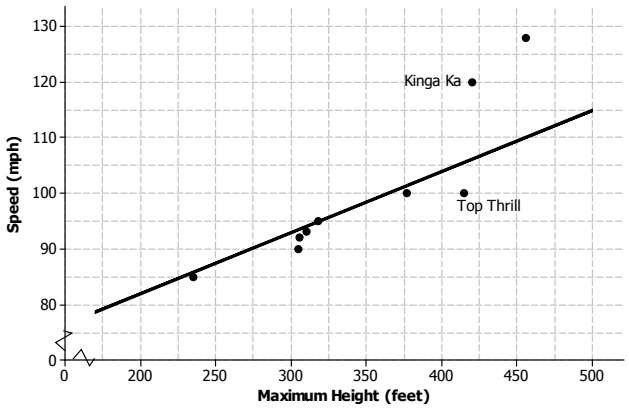
2. Find the equation of your line. Show your steps.

3. For the two roller coasters identified in the scatter plot, find the approximate difference between the observed speeds and the predicted speeds using the line.

Exit Ticket Sample Solutions

1. **A scatter plot of the height and speed of some of the world's fastest roller coaster rides is indicated below. Draw a line that you think is a good fit for the data.**

   Student would draw a line based on the goal of a best fit for the given scatter plot. A possible line is drawn below.



2. **Find the equation of your line. Show your steps.**

   Answers will vary based on the line drawn. Let S equal the speed of the roller coaster and H equal the maximum height of the roller coaster.

   $$m = \frac{115 - 85}{500 - 225} \approx 0.11$$

   $$S = 0.11H + b$$
   $$85 = 0.11(225) + b$$
   $$b \approx 60$$

   Therefore, the equation of the line drawn in Problem 1 is $S = 0.11H + 60$.

3. **For the two roller coasters identified in the scatter plot, find the approximate difference between the observed speeds and the predicted speeds using the line.**

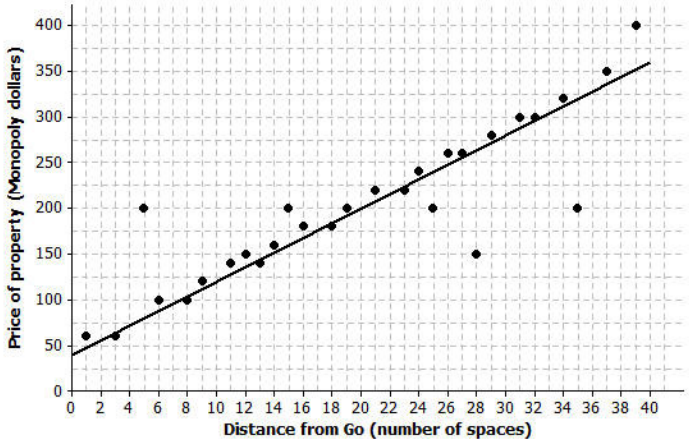
   Answers will vary depending on the line drawn by a student or the equation of the line. For the Top Thrill, the maximum height is about 415 feet and the speed about 100 miles per hour. The line indicated in Problem 2 predicts a speed of 106 miles per hour, so the difference is about 6 miles per hour over the actual speed. For the Kinga Ka, the maximum height is about 424 feet with a speed of 120 miles per hour. The line predicts a speed of about 107 miles per hour, for a difference of 13 miles per hour under the actual speed. (Students can use the graph or the equation to find the predicted speed.)

Problem Set

1. Monopoly is a board game that is popular in many countries. The scatter plot below shows the distance from "Go" to a property (in number of spaces moving from "Go" in a clockwise direction) and the price of the properties on the Monopoly board. The equation of the line is $P = 8x + 40$, where P represents the price (in Monopoly dollars) and x represents the distance (in number of spaces).

Distance from "Go" (number of spaces)	Price of Property (Monopoly dollars)	Distance from "Go" (number of spaces)	Price of Property (Monopoly dollars)
1	60	21	220
3	60	23	220
5	200	24	240
6	100	25	200
8	100	26	260
9	120	27	260
11	140	28	150
12	150	29	280
13	140	31	300
14	160	32	300
15	200	34	320
16	180	35	200
18	180	37	350
19	200	39	400

Price of Property vs. Distance from "Go" in Monopoly



a. Use the equation to find the difference (observed value − predicted value) for the most expensive property and for the property that is 35 spaces from "Go."

The most expensive property is 39 spaces from "Go" and costs $400. The price predicted by the line would be $8(39) + 40$ or $352. Observed price – predicted price would be $400 − $352 = $48. The price predicted for 35 spaces from "Go" would be $8(35) + 40 = $320. Observed price – predicted price would be $200 − $320 = −$120.

b. Four of the points seem to lie in a horizontal line. What do these points have in common? What is the equation of the line containing those four points?

These points all have the same price. The equation of the line through those points would be Price = $200.

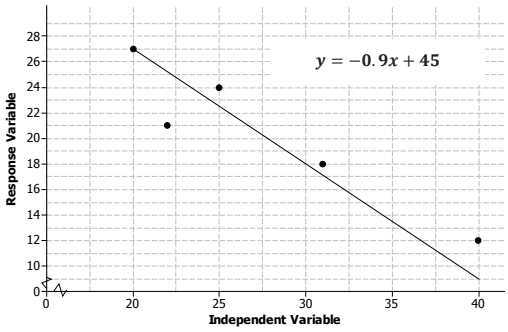
c. The four points described in part (b) are the railroads. If you were fitting a line to predict price with distance from "Go," would you use those four points? Why or why not?

Answers will vary. Because the four points are not part of the overall trend in the price of the properties, I would not use them to determine a line that describes the relationship. I can show this by finding the total error to measure the fit of the line.

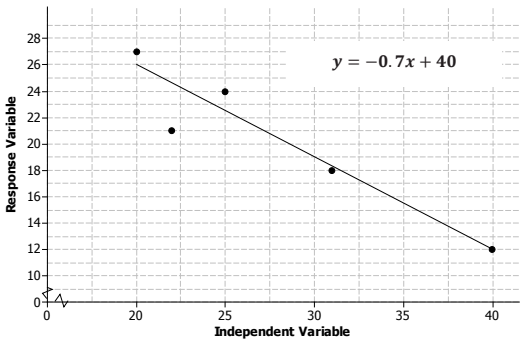
2. The table below gives the coordinates of the five points shown in the scatter plots that follow. The scatter plots show two different lines.

Data Point	Independent Variable	Response Variable
A	20	27
B	22	21
C	25	24
D	31	18
E	40	12

Line 1



Line 2



a. Find the predicted response values for each of the two lines.

Independent	*Observed Response*	*Response Predicted by Line 1*	*Response Predicted by Line 2*
20	27	27	26
22	21	25.2	24.6
25	24	22.5	22.5
31	18	17.1	18.3
40	12	9	12

b. For which data points is the prediction based on Line 1 closer to the actual value than the prediction based on Line 2?

Only for data point A. For data point C, both lines are off by the same amount.

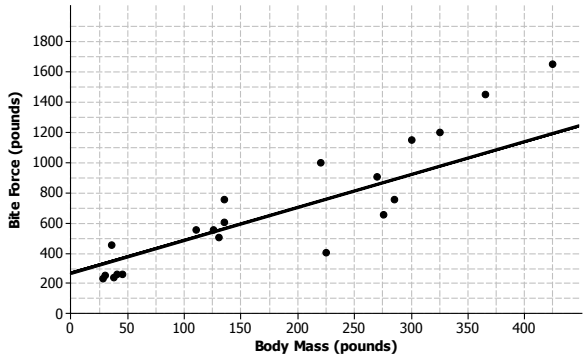
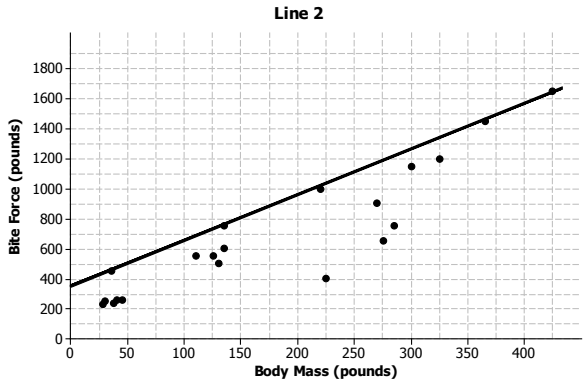
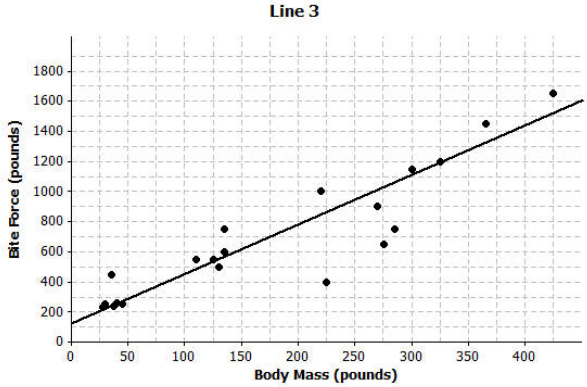
c. Which line (Line 1 or Line 2) would you select as a better fit?

Line 2 because it is closer to more of the data points.

3. The scatter plots below show different lines that students used to model the relationship between body mass (in pounds) and bite force (in pounds) for crocodiles.

   a. Match each graph to one of the equations below and explain your reasoning. Let B represent bite force (in pounds) and W represent body mass (in pounds).

Equation 1	Equation 2	Equation 3
$B = 3.28W + 126$	$B = 3.04W + 351$	$B = 2.16W + 267$

Equation: *3* *The intercept of 267 appears to match the graph, which has the second largest intercept.*	**Line 1** 
Equation: *2* *The intercept of Equation 2 is larger, so it matches Line 2, which has a y-intercept closer to 400.*	**Line 2** 
Equation: *1* *The intercept of Equation 1 is the smallest, which seems to match the graph.*	**Line 3** 

b. **Which of the lines do you think would be a better fit for the trend in the data? Explain your thinking.**

Answers will vary. Line 3 would be better than the other two lines. Line 1 is not a good fit for larger weights, and Line 2 is above nearly all of the points and pretty far away from most of them. It looks like Line 3 would be closer to most of the points.

4. **Comment on the following statements:**

a. **A line modeling a trend in a scatter plot always goes through the origin.**

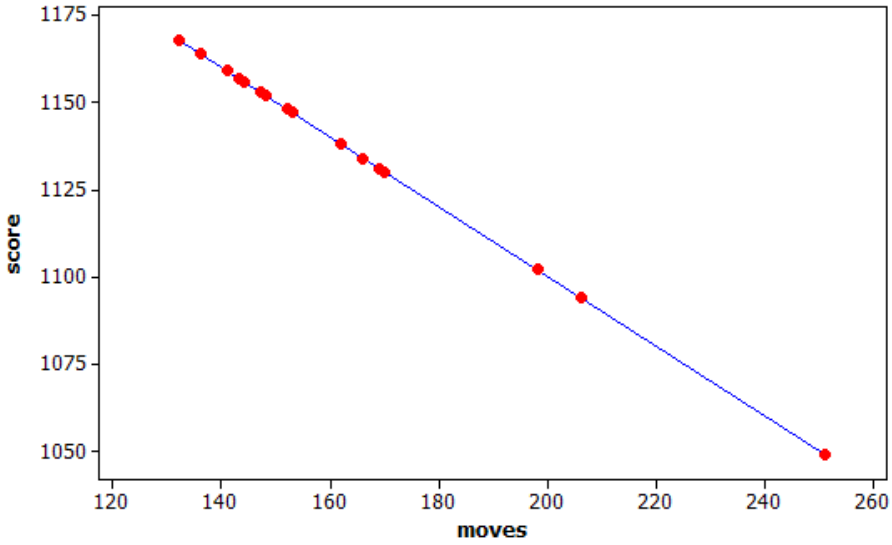
Some trend lines will go through the origin, but others may not. Often, the value $(0, 0)$ does not make sense for the data.

b. **If the response variable increases as the independent variable decreases, the slope of a line modeling the trend will be negative.**

If the trend is from the upper left to the lower right, the slope for the line will be negative because for each unit increase in the independent variable, the response will decrease.

Name _____ Date _____

1. Many computers come with a "solitaire" card game. The player moves cards in certain ways to complete specific patterns. The goal is to finish the game in the shortest number of moves possible, and a player's score is determined by the number of moves. A statistics teacher played the game 16 times and after each game recorded the number of moves and the final score. The line represents the linear function that is used to determine the score from the number of moves.



a. Was this person's average score closer to 1130 or 1110? Explain how you decided.

b. The first two games she played took 169 moves (1131 points) and 153 moves (1147 points). Based on this information, determine the equation of the linear function used by the computer to calculate the score from the number of moves. Explain your work.

c. Based on the linear function, each time the player makes a move, how many points does he or she lose?

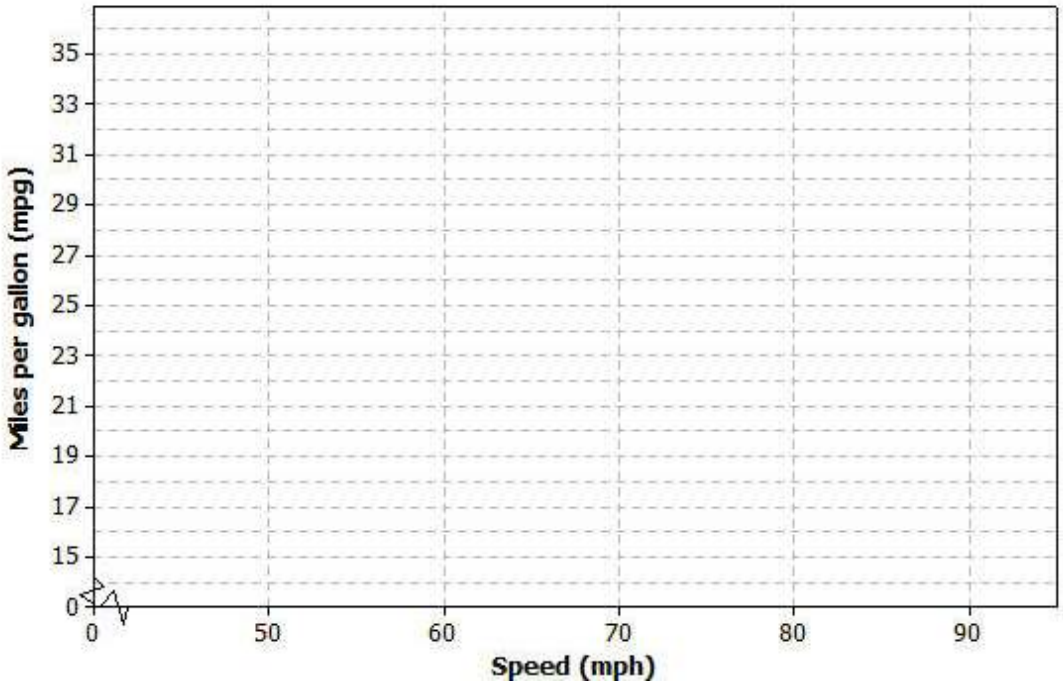
d. Based on the linear function, how many points does the player start with in this game? Explain your reasoning.

2. To save money, drivers often try to increase their mileage, which is measured in miles per gallon (mpg). One theory is that speed traveled impacts miles per gallon. Suppose the following data are recorded for five different 300-mile tests, with the car traveling at different speeds in miles per hour (mph) for each test.

Speed (mph)	Miles per gallon (mpg)
50	32
60	29
70	24
80	20
90	17

a. For the data in this table, is the association positive or negative? Explain how you decided.

b. Construct a scatter plot of these data using the following coordinate grid. The vertical axis represents the miles per gallon (mpg) and the horizontal axis represents the speed in miles per hour (mph).



c. Draw a line on your scatter plot that you think is a reasonable model for predicting the miles per gallon from the car speed.

d. Estimate and interpret the slope of the line you found in part (c).

Suppose additional data were measured for three more tests. These results have been added to the previous tests and the combined data are shown in the table below.

Speed (mph)	Miles per gallon (mpg)
20	25
30	27
40	30
50	32
60	29
70	24
80	20
90	17

e. Does the association for these data appear to be linear? Why or why not?

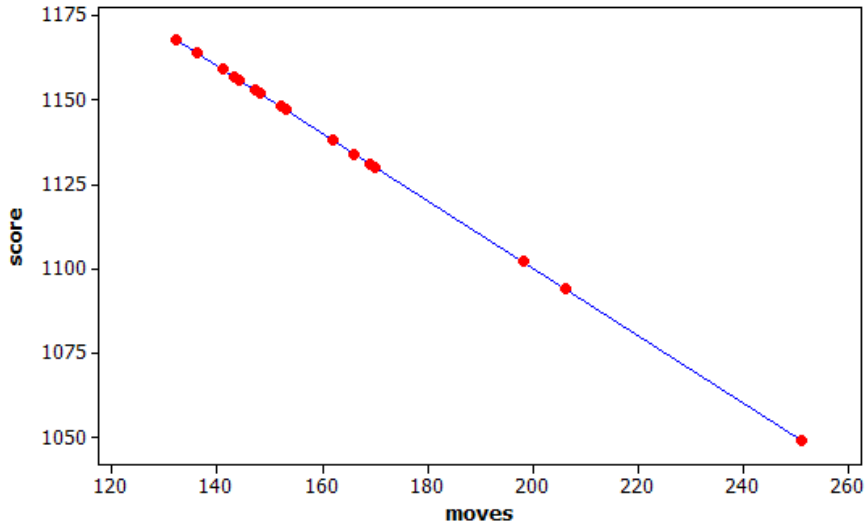
f. If your only concern was miles per gallon and you had no traffic constraints, what speed would you recommend traveling based on these data? Explain your choice.

A Progression Toward Mastery					
Assessment Task Item		**STEP 1** Missing or incorrect answer and little evidence of reasoning or application of mathematics to solve the problem.	**STEP 2** Missing or incorrect answer but evidence of some reasoning or application of mathematics to solve the problem.	**STEP 3** A correct answer with some evidence of reasoning or application of mathematics to solve the problem. **OR** An incorrect answer with substantial evidence of solid reasoning or application of mathematics to solve the problem.	**STEP 4** A correct answer supported by substantial evidence of solid reasoning or application of mathematics to solve the problem.
1	a **8.SP.A.1**	Student solution does not appear to utilize the information from the graph.	Student chooses 1110 based solely on it being the midpoint of the y-axis values.	Student chooses 1130 but reasoning is incomplete or missing.	Student chooses 1130 based on the higher concentration of red dots around those y-values.
	b **8.F.B.4**	Student cannot obtain a line.	Student only attempts to eyeball line from graph.	Student approach is reasonable but does not obtain the correct line, e.g., interchanges slope and intercept in equation or inverse of slope equation is set up or insufficient work is shown.	Student finds correct equation (or with minor errors) from slope $= \frac{(1131-1147)}{169-153} = -1$, and intercept from $1131 = a - 169$, so $a = 1308$. Equation: $y - hat = 1300 - x$, where $y =$ points and $x =$ number of moves.
	c **8.F.B.4**	Student makes no use of given data.	Student does not recognize this as question about slope.	Student only estimates from graph.	Student reports slope (-1) found in (b).
	d **8.F.B.4**	Student makes no use of given data.	Student does not recognize this as a question about intercept.	Student only estimates from graph or solves equation with moves = 0 without recognizing a connection to the equation.	Student reports intercept (1300) found in (b).

2	a 8.F.B.4	Student does not make use of data in table or context.	Student answers based solely on the content, e.g., faster cars will be less fuel-efficient.	Student refers to scatter plot in (b) or makes a minor error (e.g., misspeaks and describes a negative association, but appears to unintentionally call it a positive association).	Student notes that mpg values are decreasing while speeds (mph) are increasing, so negative association. Student could also solve for slope and note sign of slope.
	b 8.SP.A.1	Student does not make use of given data.	Student does not have the correct number of dots.	Student reverses roles of sped and miles per gallon	Scatter plot has five dots in correct locations.
	c 8.SP.A.2	Student does not answer question.	Student does not draw a line but rather connects the dots.	Line is used but does not reasonably describe the behavior of the plotted data.	Line reasonably summarizes the behavior of the data.
	d 8.F.B.4	Student does not utilize data given in problem.	Student uses correct approach, but makes major calculation error, uses only values from table, or fails to interpret slope.	Student uses correct approach, but makes minor errors in calculation or in interpretation.	Student estimates coordinates for two locations and determines change in y-values divided by change in x-values, e.g., $(70,25)$ and $(80,20)$ which yields $\left(-\frac{5}{10}\right) = -0.5$, and interprets this as the decrease in mpg per additional mph in speed.
	e 8.F.B.5	Student does not examine the increasing or decreasing pattern in the values.	Student attempts to sketch a graph of the data and focuses on overall pattern but does not see the change in the direction of the association.	Student focuses only on how the change in the miles per gallon is not constant without noticing the change in sign of the differences.	Student comments on the increasing then decreasing behavior of the mpg column as the mph column steadily increases.
	f 8.F.B.4	Student does not address the question.	Student answers around 55 mph based only on anecdote and does not provide any reasoning.	Student gives a reasonable estimate but does not fully justify the choice.	Student gives justification for a speed between 40 and 50 mph or at 50 mph based on the association "peaking" at 50 mph.

Name _____ Date _____

1. Many computers come with a "solitaire" card game. The player moves cards in certain ways to complete specific patterns. The goal is to finish the game in the shortest number of moves possible, and a player's score is determined by the number of moves. A statistics teacher played the game 16 times and after each game recorded the number of moves and the final score. The line represents the linear function that is used to determine the score from the number of moves.



a. Was this person's average score closer to 1130 or 1110? Explain how you decided.

Most of the games had scores between 1125 and 1175. The mean score will be closer to 1130.

b. The first two games she played took 169 moves (1131 points) and 153 moves (1147 points). Based on this information, determine the equation of the linear function used by the computer to calculate the score from the number of moves. Explain your work.

The difference in the scores is 1131 - 1147 or -16.
The difference in the number of moves is 169 - 153 = 16
The slope is -16/16 or -1. This means that
1131 = intercept - 169, so intercept equals 1300
Score = 1300 - moves

c. Based on the linear function, each time the player makes a move, how many points does he or she lose?

One point lost per move.

d. Based on the linear function, how many points does the player start with in this game? Explain your reasoning.

1300, or the score when the number of moves equals 0.

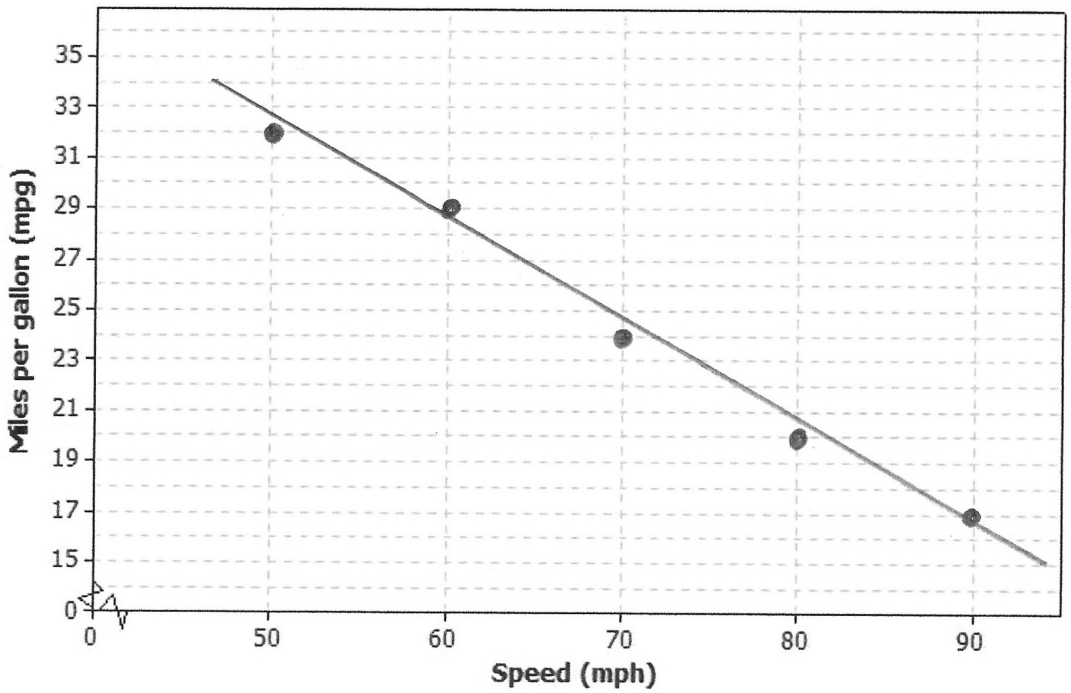
2. To save money, drivers often try to increase their mileage, which is measured in miles per gallon (mpg). One theory is that speed traveled impacts miles per gallon. Suppose the following data are recorded for five different 300-mile tests, with the car traveling at different speeds in miles per hour (mph) for each test.

Speed (mph)	Miles per gallon (mpg)
50	32
60	29
70	24
80	20
90	17

a. For the data in this table, is the association positive or negative? Explain how you decided.

As the speed increases in miles per hour, the miles per gallon decrease. This describes a negative association.

b. Construct a scatter plot of these data using the following coordinate grid. The vertical axis represents the miles per gallon (mpg) and the horizontal axis represents the speed in miles per hour (mph).



c. Draw a line on your scatter plot that you think is a reasonable model for predicting the miles per gallon from the car speed.

d. Estimate and interpret the slope of the line you found in part (c).

Two points are about (50, 32) and (90, 17).
So, slope $\approx \frac{32-17}{50-90} = -.375$
Each increase of 1 mph in speed predicts a decrease of .375 mpg.

Suppose additional data were measured for three more tests. These results have been added to the previous tests and the combined data are shown in the table below.

Speed (mph)	Miles per gallon (mpg)
20	25
30	27
40	30
50	32
60	29
70	24
80	20
90	17

e. Does the association for these data appear to be linear? Why or why not?

No, the values mostly increase and then mostly decrease. There is no fixed rate of increase or decrease.

f. If your only concern was miles per gallon and you had no traffic constraints, what speed would you recommend traveling based on these data? Explain your choice.

About 50 mph. It is around 50mph that the mpg stops increasing and starts to decrease.