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.

Scaffolding:

- Point out to students that in this lesson, the meaning of 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.

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 scatterplots show trends in bivariate data.

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

a. **Does it look like there is a relationship between the two variables used to make the scatter plot?**

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

c. **If the relationship appears to be linear, is the relationship a positive linear relationship or a negative linear relationship?**

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.

Scaffolding:

For English language learners, 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.

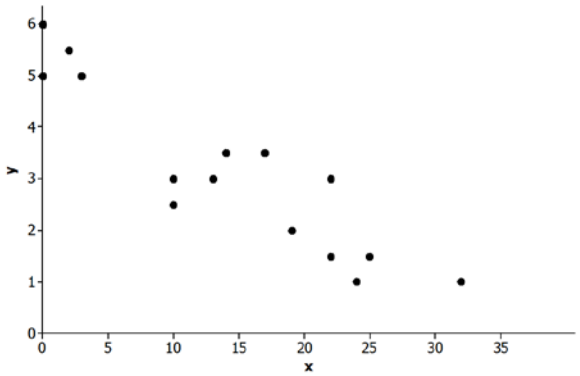
Exercises 1–5 (8–10 minutes)

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 1–9

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.

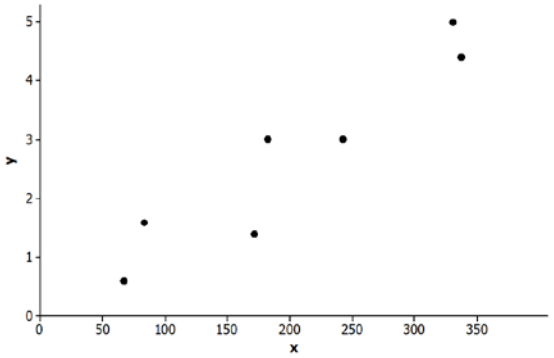
If there is a relationship, does it appear to be linear?

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.

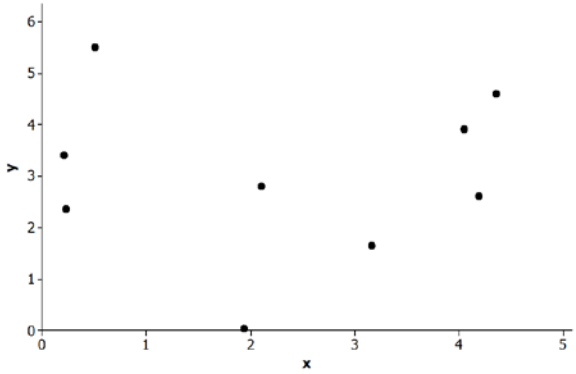
If there is a relationship, does it appear to be linear?

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.

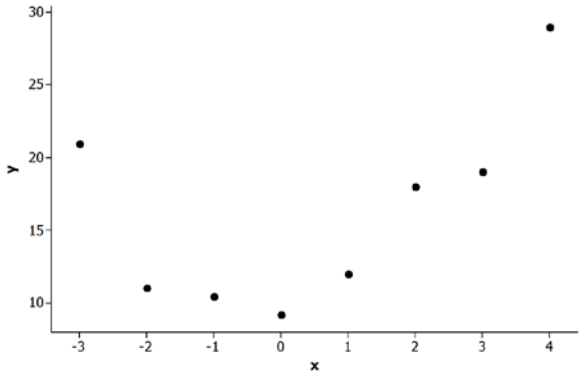
If there is a relationship, does it appear to be linear?

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.

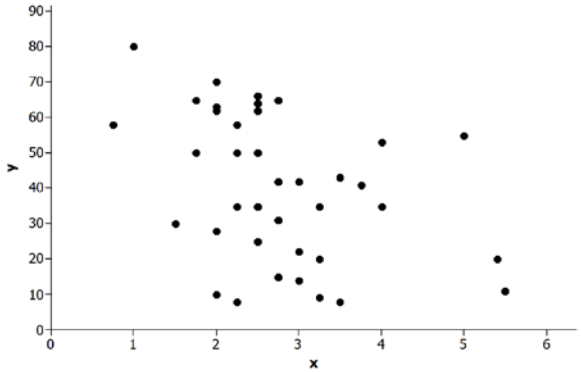
If there is a relationship, does it appear to be linear?

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.

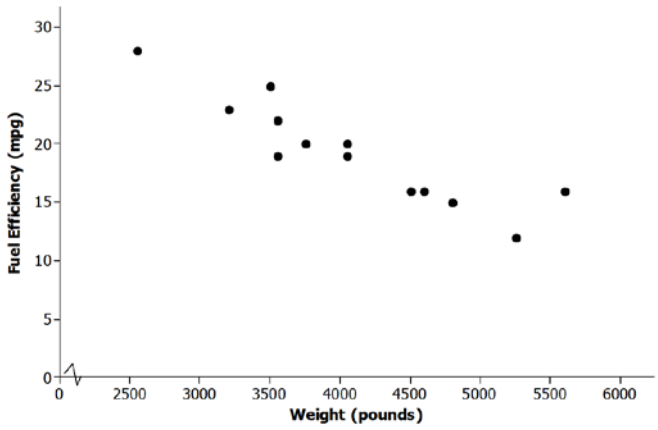
Exercises 6–9 (10 minutes)

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.

Scaffolding:

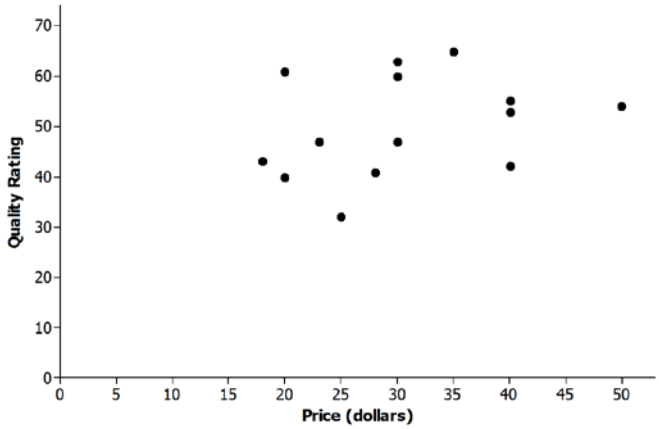
- It may be helpful to provide sentence frames on the classroom board to help students articulate their observations.
- For example, "I see a negative or positive linear relationship between ____ and ____. The higher or lower the ____, the higher or lower the ____."

6. **Below is a scatter plot of data on weight in pounds (x) and fuel efficiency in miles per gallon (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.**



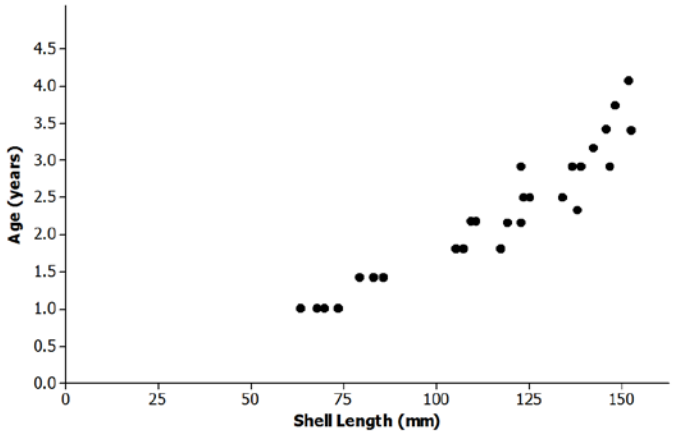
Possible response: There appears to be a negative linear relationship between fuel efficiency and weight. Students may note that this is a fairly strong negative relationship. The cars with greater weight tend to have lesser fuel efficiency.

7. **Below is a scatter plot of data on price in dollars (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.**



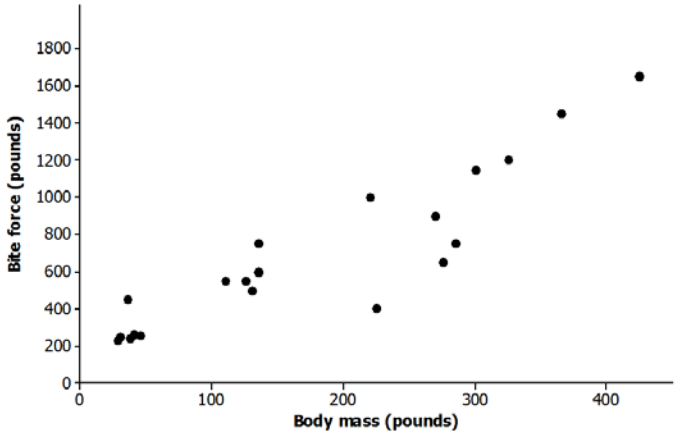
Possible response: There does not appear to be a relationship between quality rating and price. The points in the scatter plot appear to be scattered at random, and there is no apparent pattern in the scatter plot.

8. Below is a scatter plot of data on shell length in millimeters (x) and age in years (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 response: 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 in pounds (x) and bite force in pounds (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 response: 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 (5 minutes): Clusters and Outliers

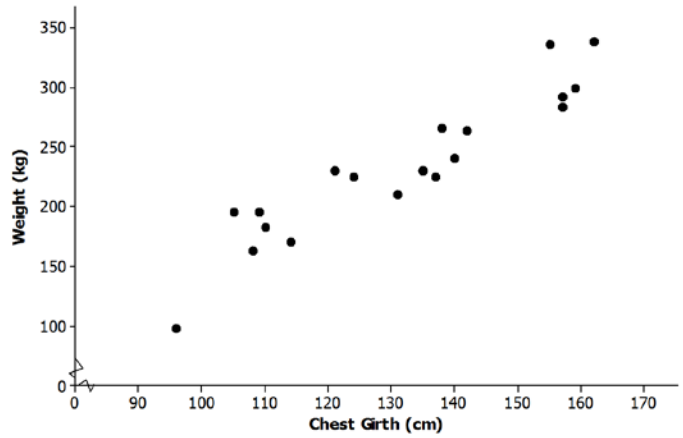
Spend a few minutes introducing the meaning of the terms *clusters* and *outliers* in the context of scatter plots. You might ask students to sketch a scatter plot that has an outlier and a scatter plot that has two clusters as a way of checking their understanding of these terms before moving on to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

Example 2: Clusters and Outliers

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 centimeter (x) and weight in kilogram (y).



Scaffolding:

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

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 response: 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 English language learners 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)

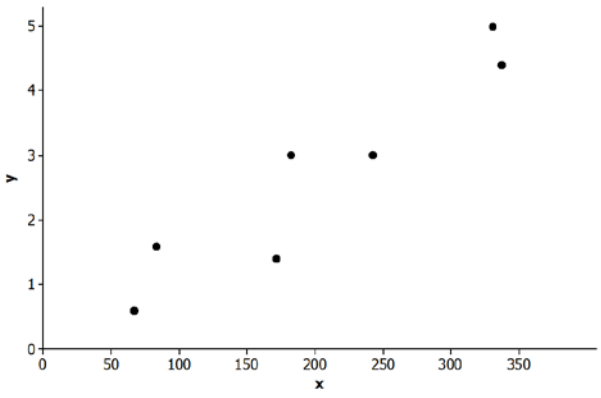
Name _____ Date_____

Lesson 7: Patterns in Scatter Plots

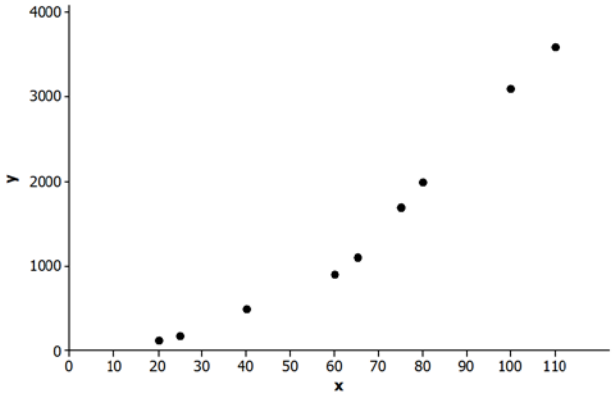
Exit Ticket

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

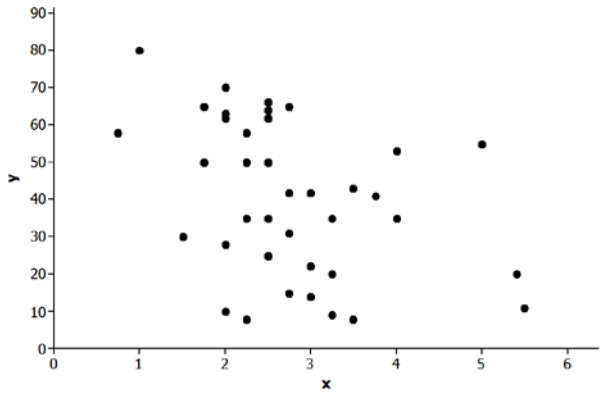
Scatter Plot 1



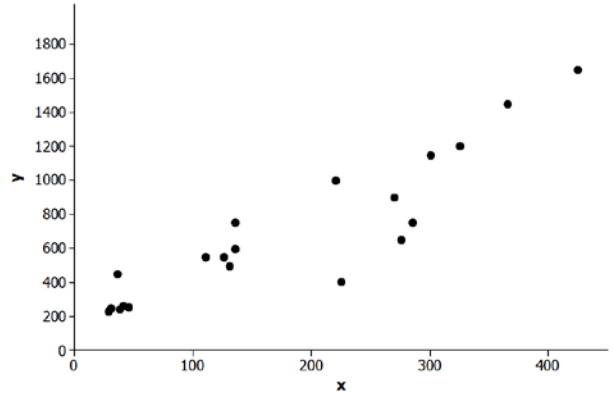
Scatter Plot 2



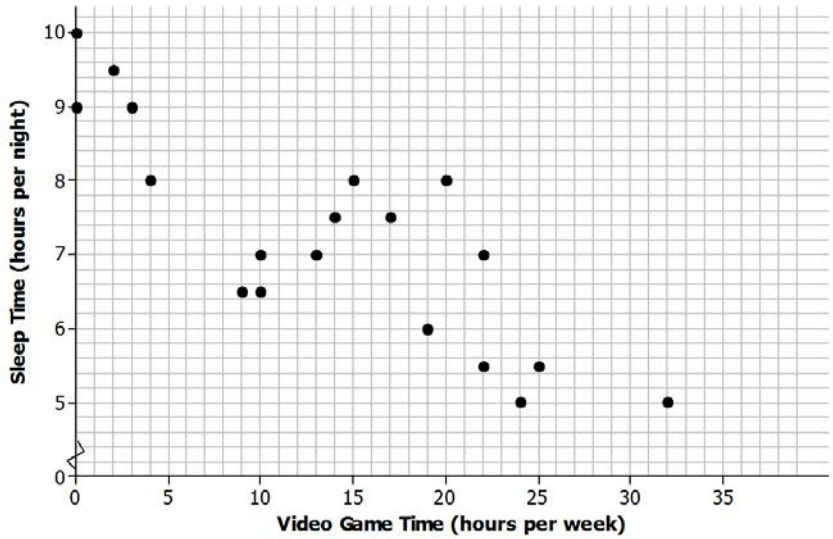
Scatter Plot 3



Scatter Plot 4



2. The scatter plot below was constructed using data from eighth-grade students on time spent in hours 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?

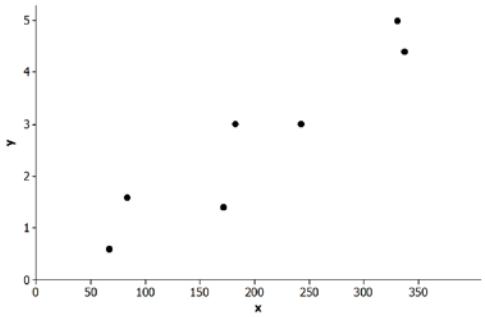


3. In a scatter plot, if the values of y tend to increase as the value of x increases, would you say that there is a positive relationship or a negative relationship between x and y? Explain your answer.

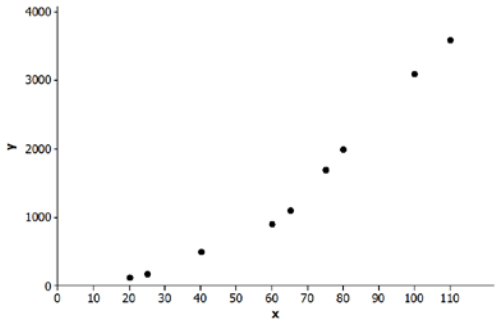
Exit Ticket Sample Solutions

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

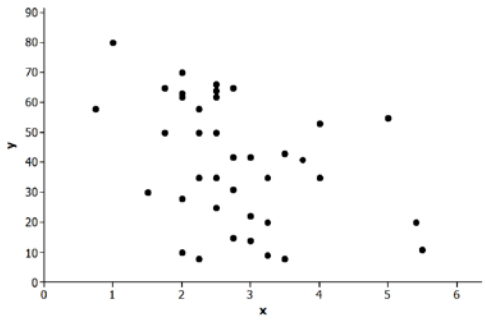
Scatter Plot 1



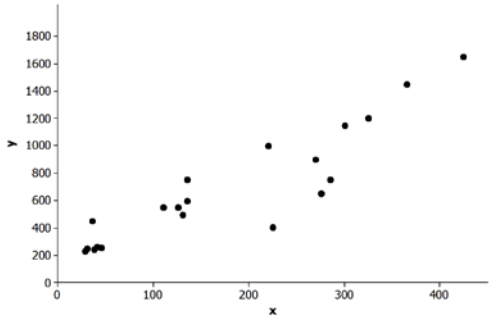
Scatter Plot 2



Scatter Plot 3

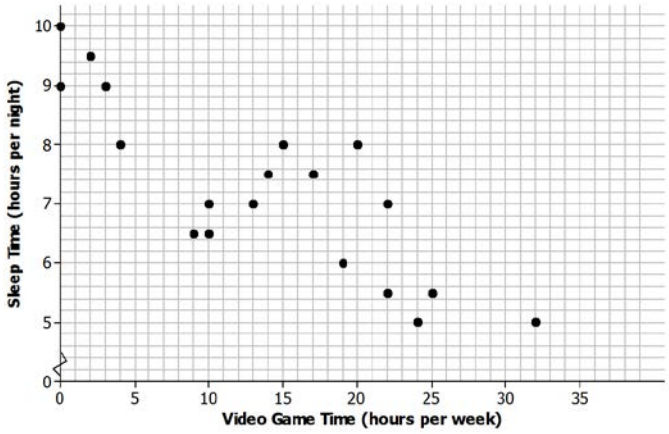


Scatter Plot 4



Only Scatter plot 3 shows a negative linear relationship because 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 in hours 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, the 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.

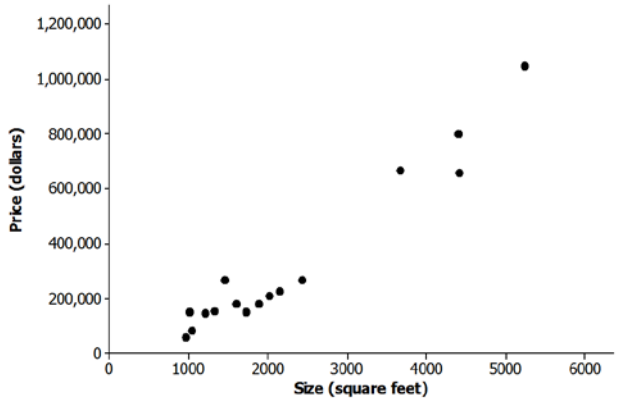
3. In a scatter plot, if the value of y tends to increase as the value of x increases, would you say that there is a positive relationship or a negative relationship between x and y?

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

Problem Set Sample Solutions

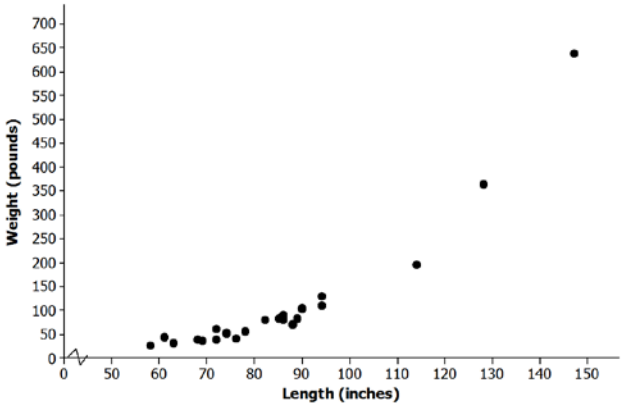
The Problem Set is intended to reinforce material from the lesson and have students think about the meaning of points in a scatter plot, clusters, positive and negative linear trends, and trends that are not linear.

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



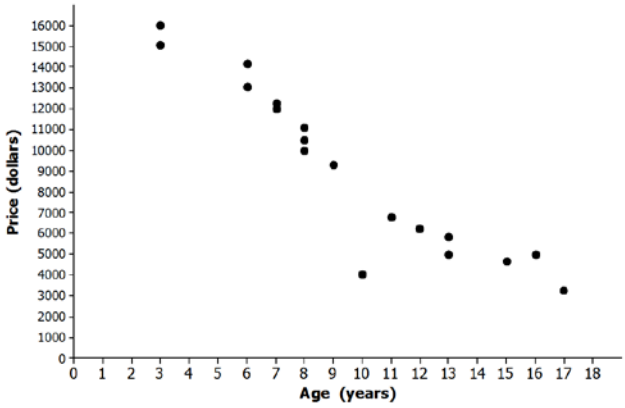
Answers will vary. Possible response: There appears to be a positive linear relationship between size and price. Price tends to increase as size increases. There appear to be two clusters of houses—one that includes houses that are less than $3,000$ square feet in size and another that includes houses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

2. **The scatter plot below was constructed using data on length in inches (x) of several alligators and weight in pounds (y). 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 fewer than 100 inches.

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

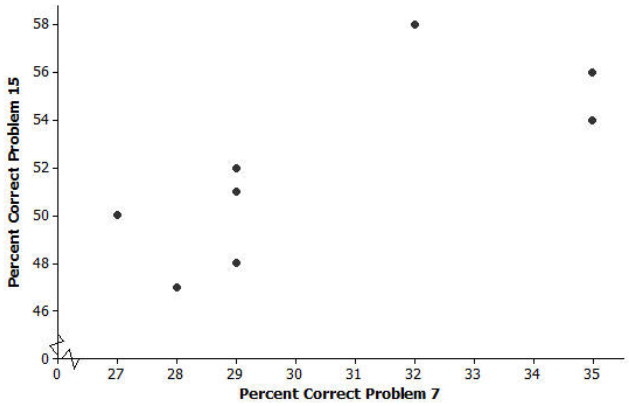


Answers will vary. Possible response: There appears to be a relatively strong negative linear relationship between price and age. Price tends to decrease as age increases. There is one car that looks like an outlier—the car that is 10 years old. This car has a price that is lower than expected based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

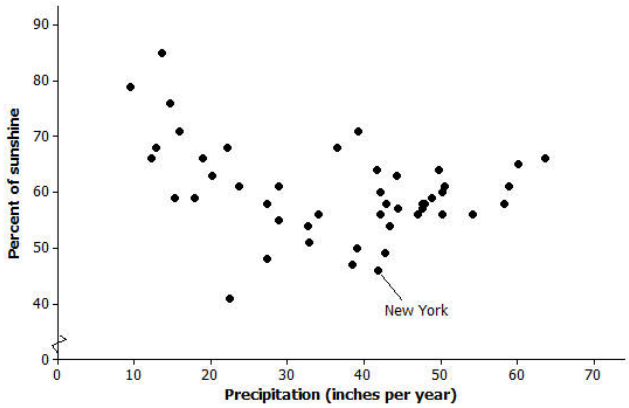
c. Which state did the best on these two problems? Explain your reasoning.

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. The state of New York 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. At a dinner party, every person shakes hands with every other person present.

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

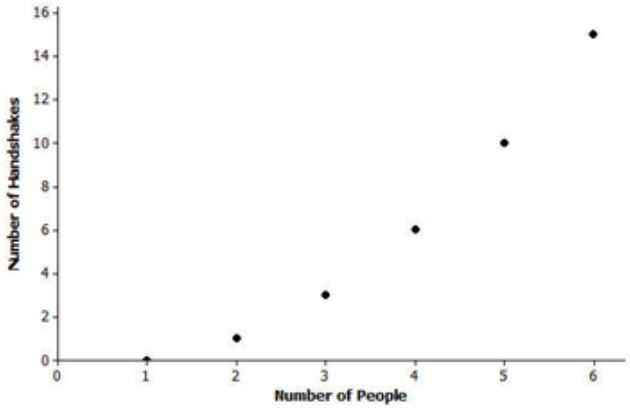
Three handshakes

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
1	0
2	1
3	3

Number People	Handshakes
4	6
5	10
6	15

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.