Lesson 7: Patterns in Scatter Plots

Classwork

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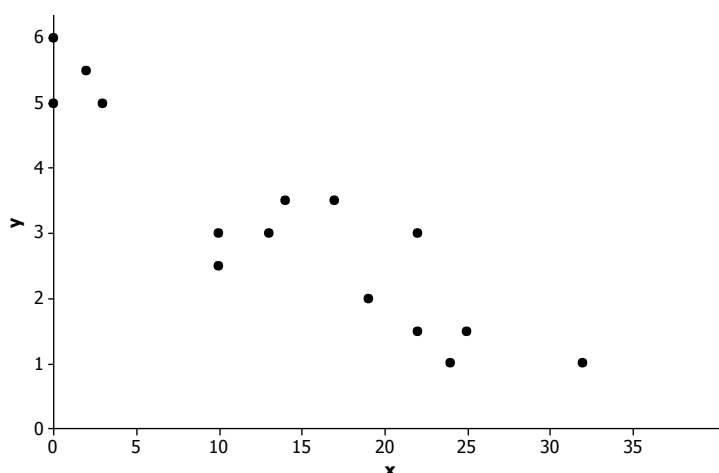
1. Does it look like there is a relationship between the two variables used to make the scatter plot?
2. If there is a relationship, does it appear to be linear?
3. If the relationship appears to be linear, is the relationship a positive linear relationship or a negative linear relationship?

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

Take a look at the following five scatter plots. Answer the three questions above for each scatter plot.
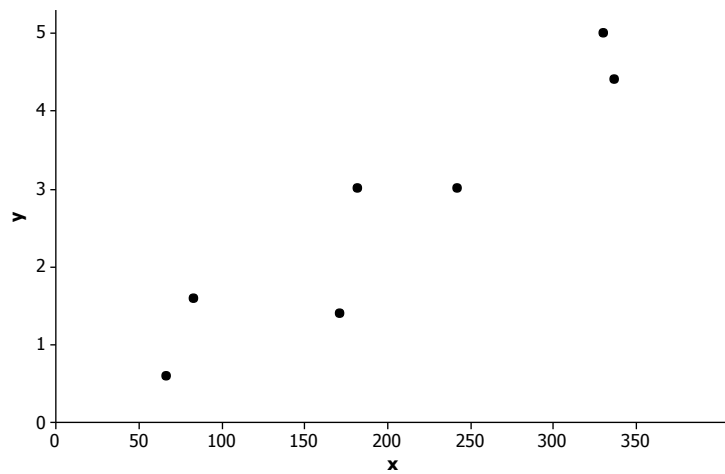
1. Scatter plot 1

Is there a relationship?

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

If the relationship appears to be linear, is it a positive or negative linear relationship?
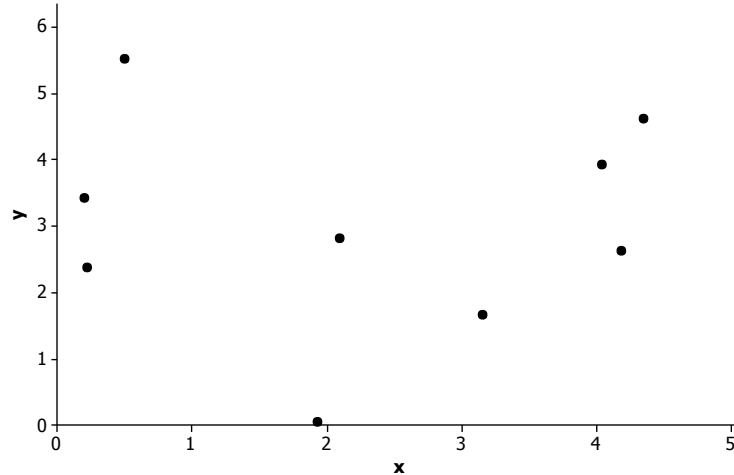
2. Scatter plot 2

Is there a relationship?

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

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

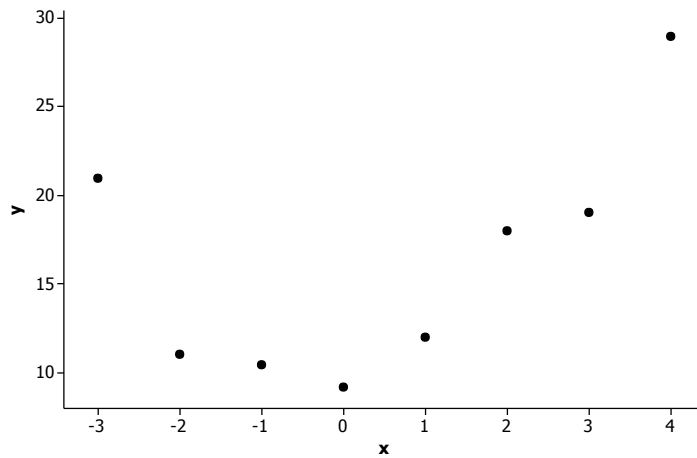3. Scatter plot 3

Is there a relationship?

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

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

4. Scatter plot 4

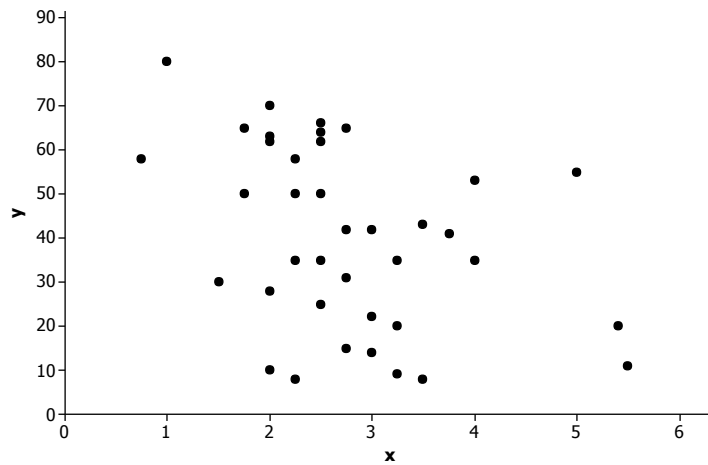
Is there a relationship?

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

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

5. Scatter plot 5


Is there a relationship?

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

If the relationship appears to be linear, is it a positive or negative linear relationship?
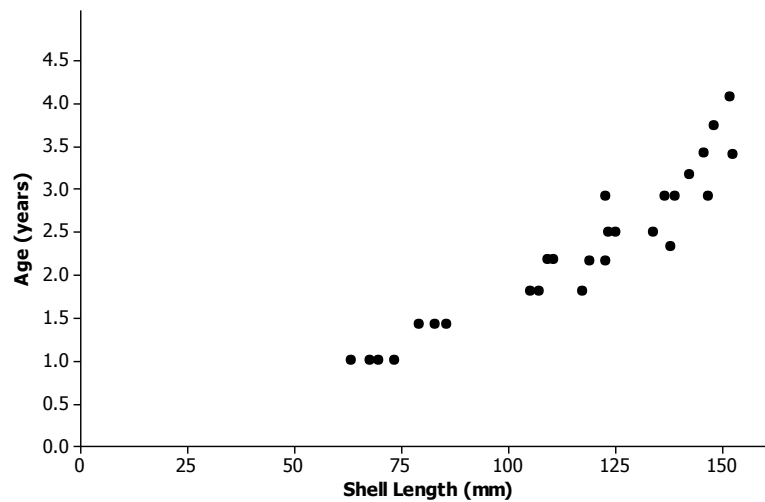
Exercises 6–9

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.
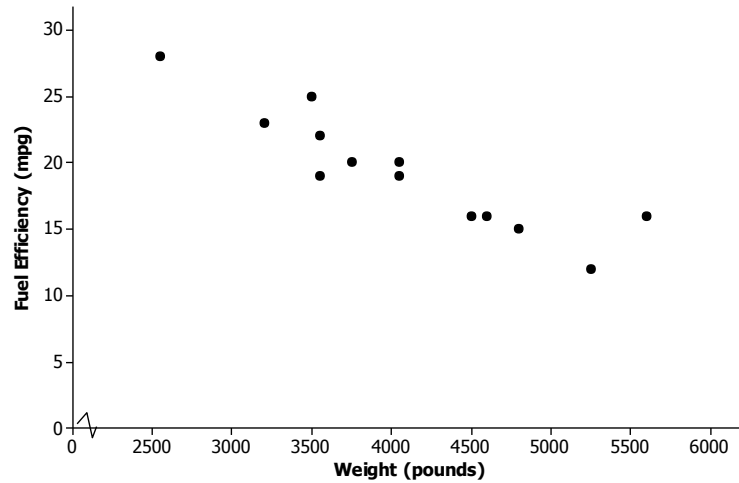

7. Below is a scatter plot of data on price (x) and quality rating (y) for 14 bike helmets. Using the questions at the beginning of this lesson as a guide, write a few sentences describing any possible relationship between x and y.
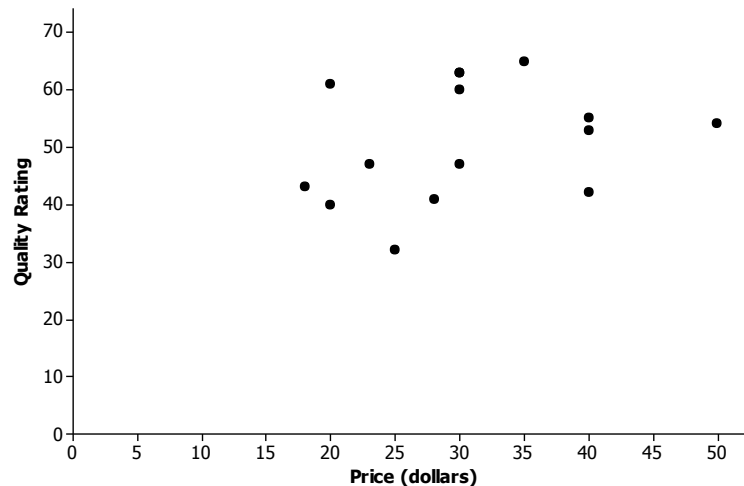
8.   Below is a scatter plot of data on shell length (x) and age (y) for 27 lobsters of known age. Using the questions at the beginning of this lesson as a guide, write a few sentences describing any possible relationship between x and y.

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.
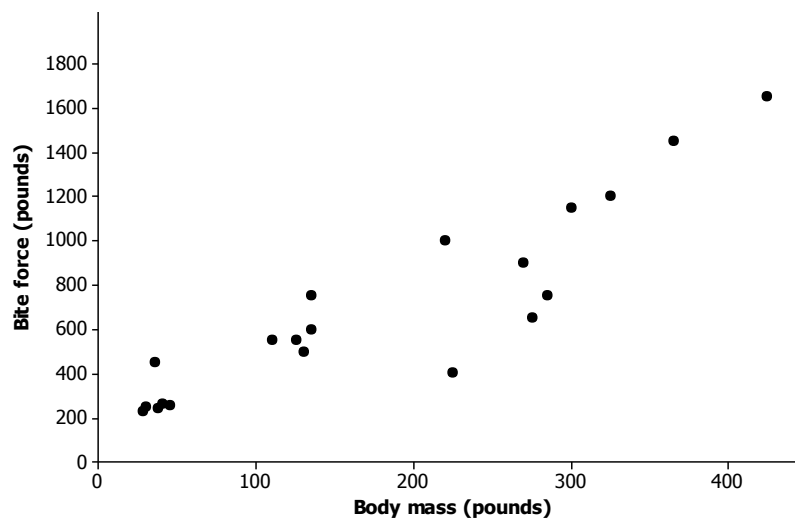
Example 2

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).
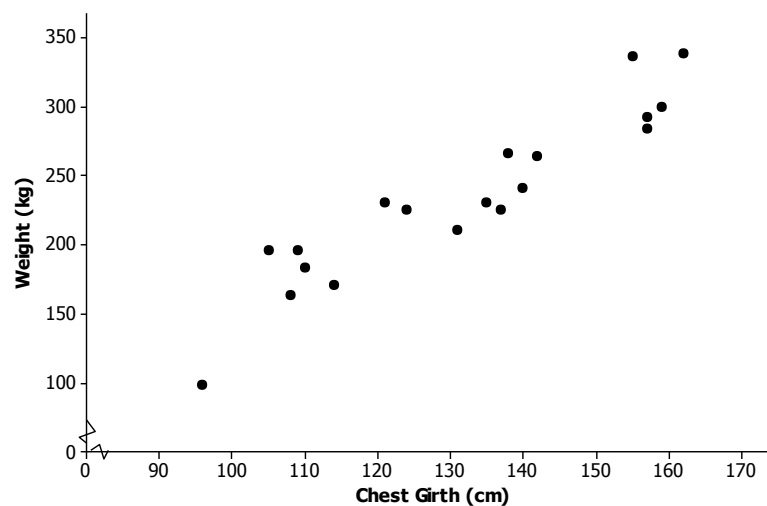

Exercises 10–12

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?

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.

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?


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.


Problem Set

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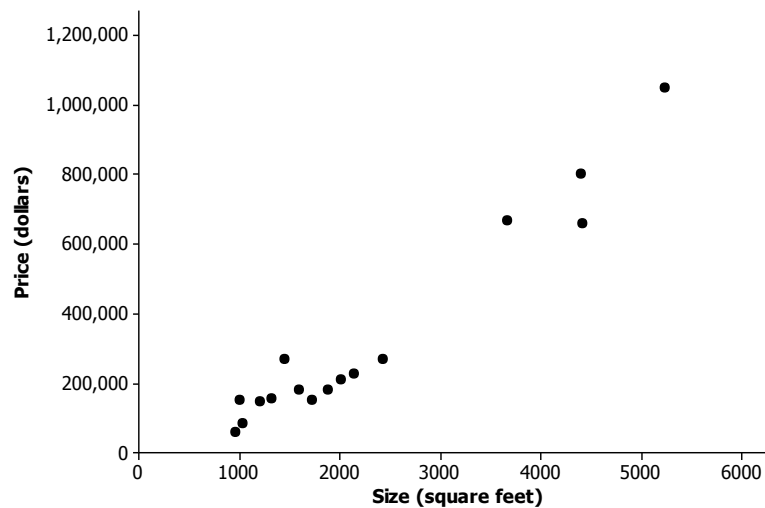
2. The scatter plot below was constructed using data on length in inches (x) and weight in pounds (y) for a sample of alligators. Write a few sentences describing the relationship between weight and length for these alligators. Are there any noticeable clusters or outliers?
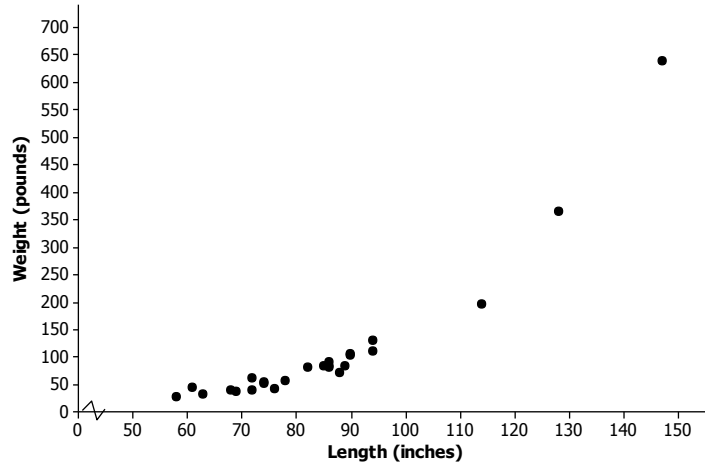

3. The scatter plot below was constructed using data on age in years (x) and price in dollars (y) for a sample of Honda Civics. Write a few sentences describing the relationship between price and age for these cars. Are there any noticeable clusters or outliers?
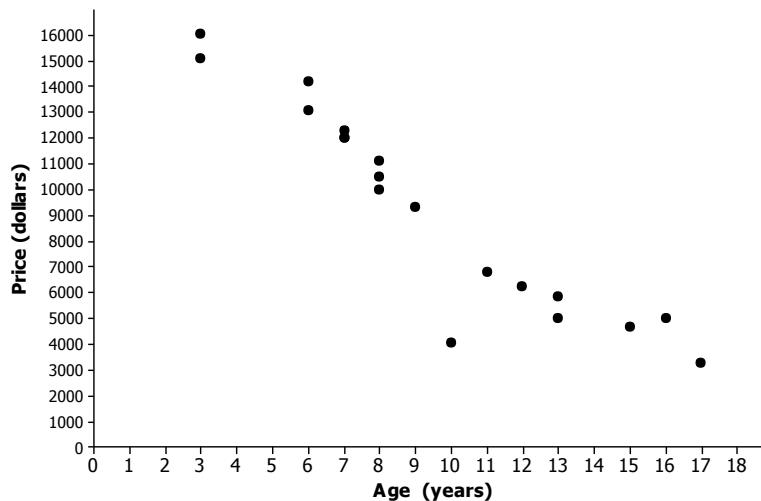
4. Samples of students in each of the U.S. states periodically take part in a large-scale assessment called the National Assessment of Educational Progress (NAEP). The table below shows the percent of students in the northeastern states (as defined by the U.S. Census Bureau) who answered problems 7 and 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
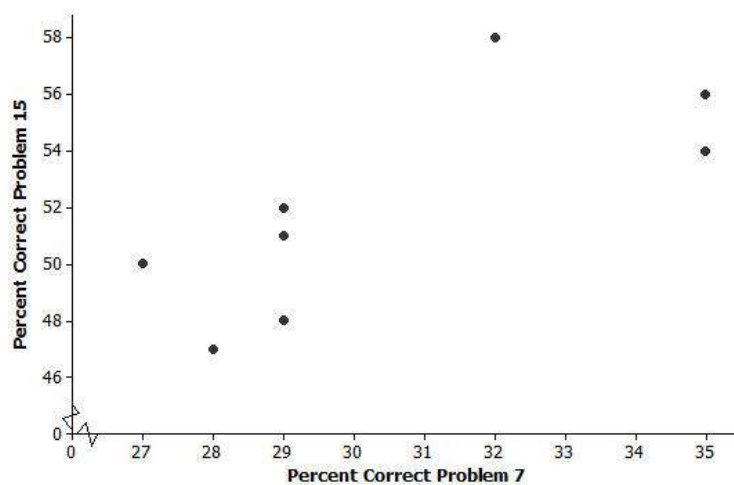
a. Why does it appear that there are only eight points in the scatter plot for nine states?

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

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

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

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.
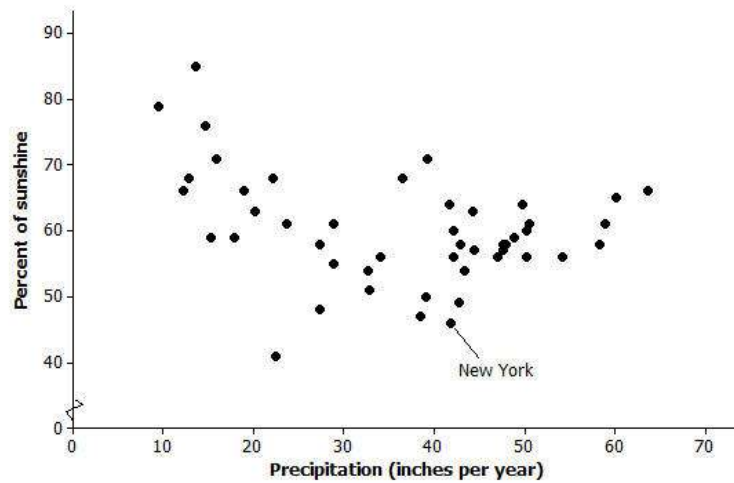

Data Source: www.currentresults.com/Weather/US/average-annual-state-sunshine.php
www.currentresults.com/Weather/US/average-annual-state-precipitation.php

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

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.

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

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?

    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

    c. Make a scatter plot of number of people (x) and number of handshakes (y). Explain your thinking.
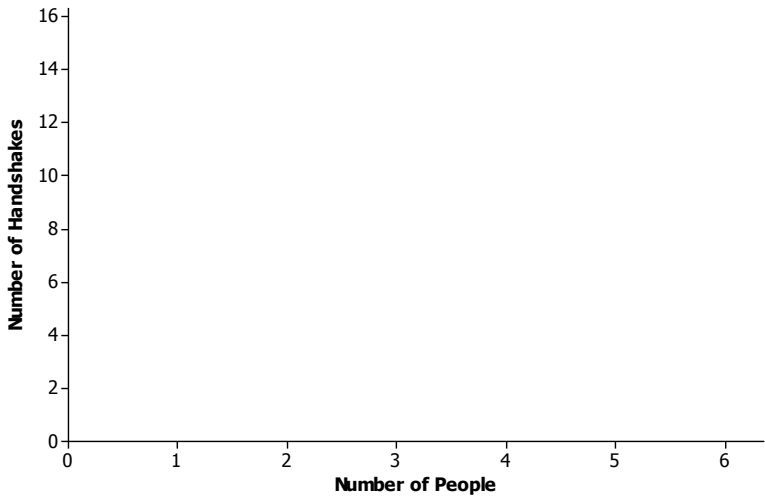
    d. Does the trend seem to be linear? Why or why not?