# 4.4    Scatter Plots and Lines of Fit

**Essential Question**   How can you use a scatter plot and a line of fit to make conclusions about data?

A **scatter plot** is a graph that shows the relationship between two data sets. The two data sets are graphed as ordered pairs in a coordinate plane.

**EXPLORATION 1**    Finding a Line of Fit

**Work with a partner.** A survey was taken of 179 married couples. Each person was asked his or her age. The scatter plot shows the results.

a. Draw a line that approximates the data. Write an equation of the line. Explain the method you used.

b. What conclusions can you make from the equation you wrote? Explain your reasoning.
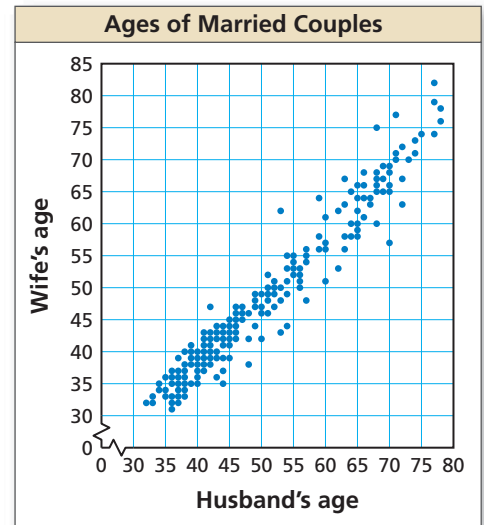
**REASONING QUANTITATIVELY**

To be proficient in math, you need to make sense of quantities and their relationships in problem situations.

**Ages of Married Couples**



**EXPLORATION 2**    Finding a Line of Fit

**Work with a partner.** The scatter plot shows the median ages of American women at their first marriage for selected years from 1960 through 2010.

a. Draw a line that approximates the data. Write an equation of the line. Let $x$ represent the number of years since 1960. Explain the method you used.

b. What conclusions can you make from the equation you wrote?

c. Use your equation to predict the median age of American women at their first marriage in the year 2020.

**Ages of American Women at First Marriage**
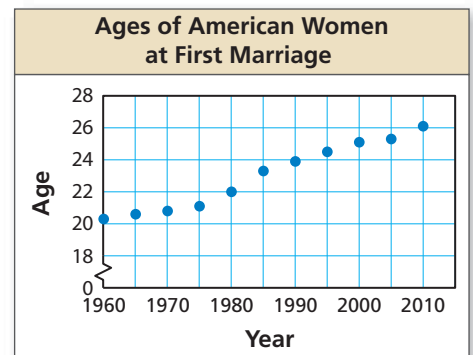


## Communicate Your Answer

3. How can you use a scatter plot and a line of fit to make conclusions about data?

4. Use the Internet or some other reference to find a scatter plot of real-life data that is different from those given above. Then draw a line that approximates the data and write an equation of the line. Explain the method you used.

## What You Will Learn

▶ Interpret scatter plots.

▶ Identify correlations between data sets.
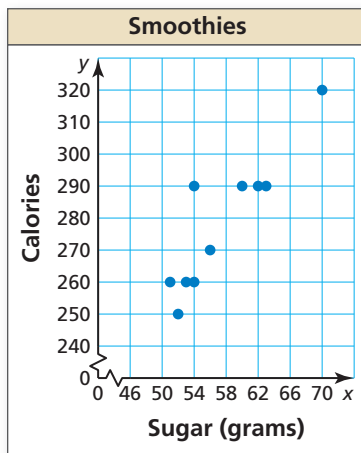
▶ Use lines of fit to model data.

## Interpreting Scatter Plots

### Core Concept

**Scatter Plot**

A **scatter plot** is a graph that shows the relationship between two data sets. The two data sets are graphed as ordered pairs in a coordinate plane. Scatter plots can show trends in the data.

### EXAMPLE 1   Interpreting a Scatter Plot

The scatter plot shows the amounts $x$ (in grams) of sugar and the numbers $y$ of calories in 10 smoothies.



**a.** How many calories are in the smoothie that contains 56 grams of sugar?

**b.** How many grams of sugar are in the smoothie that contains 320 calories?

**c.** What tends to happen to the number of calories as the number of grams of sugar increases?

### SOLUTION

**a.** Draw a horizontal line from the point that has an $x$-value of 56. It crosses the $y$-axis at 270.

▶  So, the smoothie has 270 calories.

**b.** Draw a vertical line from the point that has a $y$-value of 320. It crosses the $x$-axis at 70.

▶  So, the smoothie has 70 grams of sugar.

**c.** Looking at the graph, the plotted points go up from left to right.

▶  So, as the number of grams of sugar increases, the number of calories increases.
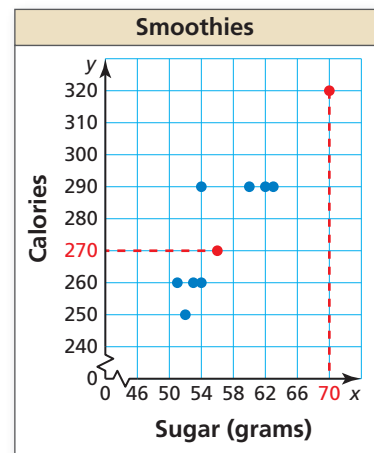


## Monitoring Progress   🔊 Help in English and Spanish at *BigIdeasMath.com*
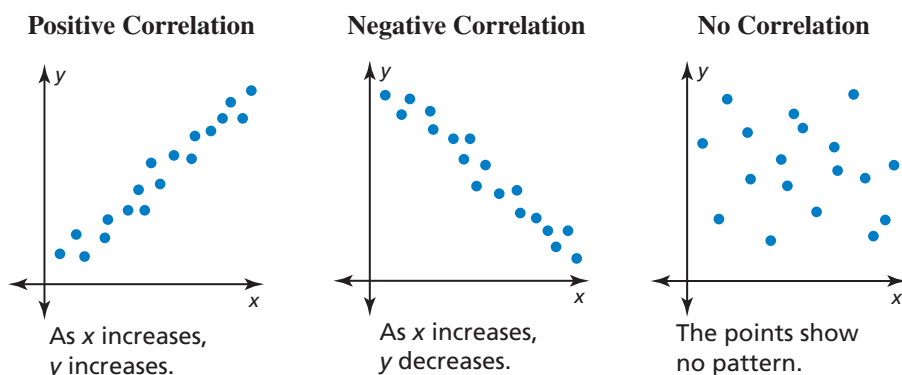
**1.** How many calories are in the smoothie that contains 51 grams of sugar?

**2.** How many grams of sugar are in the smoothie that contains 250 calories?

# Identifying Correlations between Data Sets

A **correlation** is a relationship between data sets. You can use a scatter plot to describe the correlation between data.
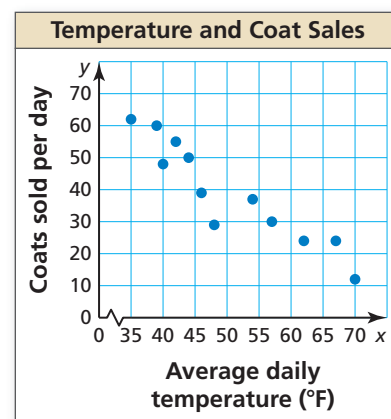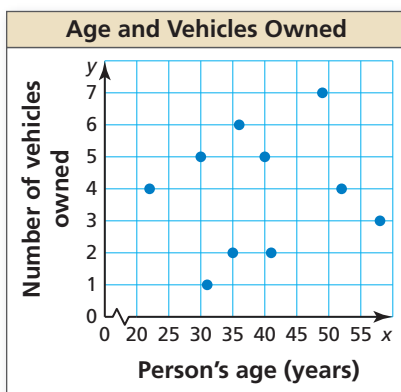
**Positive Correlation**



As *x* increases, *y* increases.

**Negative Correlation**



As *x* increases, *y* decreases.

**No Correlation**



The points show no pattern.

**EXAMPLE 2**   **Identifying Correlations**

Tell whether the data show a *positive*, a *negative*, or *no* correlation.

**a.** age and vehicles owned

**b.** temperature and coat sales at a store



**Age and Vehicles Owned**

**Temperature and Coat Sales**

**SOLUTION**

**a.** The points show no pattern. The number of vehicles owned does not depend on a person's age.

▶  So, the scatter plot shows no correlation.

**b.** As the average temperature increases, the number of coats sold decreases.

▶  So, the scatter plot shows a negative correlation.

## Monitoring Progress   🔊 Help in English and Spanish at *BigIdeasMath.com*

**Make a scatter plot of the data. Tell whether the data show a *positive*, a *negative*, or *no* correlation.**

**3.**

| Temperature (°F), *x* | 82 | 78 | 68 | 87 | 75 | 71 | 92 | 84 |
|---|---|---|---|---|---|---|---|---|
| Attendees (thousands), *y* | 4.5 | 4.0 | 1.7 | 5.5 | 3.8 | 2.9 | 4.7 | 5.3 |

**4.**

| Age of a car (years), *x* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Value (thousands), *y* | $24 | $21 | $19 | $18 | $15 | $12 | $8 | $7 |

## Using Lines of Fit to Model Data

When data show a positive or negative correlation, you can model the *trend* in the data using a line of fit. A **line of fit** is a line drawn on a scatter plot that is close to most of the data points.

> **STUDY TIP**
>
> A line of fit is also called a *trend line*.

## Core Concept

### Using a Line of Fit to Model Data

**Step 1** Make a scatter plot of the data.

**Step 2** Decide whether the data can be modeled by a line.

**Step 3** Draw a line that appears to fit the data closely. There should be approximately as many points above the line as below it.

**Step 4** Write an equation using two points on the line. The points do not have to represent actual data pairs, but they must lie on the line of fit.

---

**EXAMPLE 3** **Finding a Line of Fit**

The table shows the weekly sales of a DVD and the number of weeks since its release. Write an equation that models the DVD sales as a function of the number of weeks since its release. Interpret the slope and $y$-intercept of the line of fit.

| Week, $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Sales (millions), $y$ | $19 | $15 | $13 | $11 | $10 | $8 | $7 | $5 |

### SOLUTION

**Step 1** Make a scatter plot of the data.

**Step 2** Decide whether the data can be modeled by a line. Because the scatter plot shows a negative correlation, you can fit a line to the data.

**Step 3** Draw a line that appears to fit the data closely.

**Step 4** Write an equation using two points on the line. Use (5, 10) and (6, 8).



**DVD Sales** (scatter plot: Sales (millions of dollars) vs. Week, with points labeled (5, 10) and (6, 8))

The slope of the line is $m = \dfrac{8 - 10}{6 - 5} = -2$.

Use the slope $m = -2$ and the point (6, 8) to write an equation of the line.

$y - y_1 = m(x - x_1)$      <span style="color:red">Write the point-slope form.</span>

$y - 8 = -2(x - 6)$      <span style="color:red">Substitute −2 for $m$, 6 for $x_1$, and 8 for $y_1$.</span>

$y = -2x + 20$      <span style="color:red">Solve for $y$.</span>

▶ An equation of the line of fit is $y = -2x + 20$. The slope of the line is $-2$. This means the sales are decreasing by about $2 million each week. The $y$-intercept is 20. The $y$-intercept has no meaning in this context because there are no sales in week 0.

---
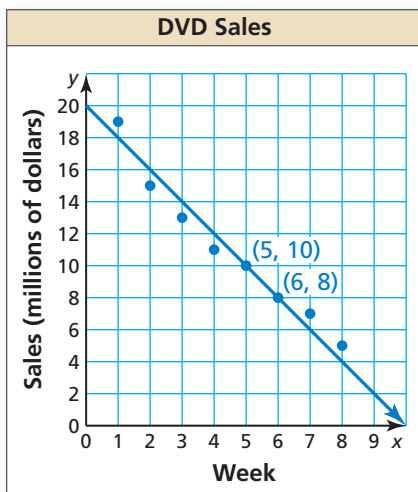
## Monitoring Progress 🔊 Help in English and Spanish at *BigIdeasMath.com*

5. The following data pairs show the monthly income $x$ (in dollars) and the monthly car payment $y$ (in dollars) of six people: (2100, 410), (1650, 315), (1950, 405), (1500, 295), (2250, 440), and (1800, 375). Write an equation that models the monthly car payment as a function of the monthly income. Interpret the slope and $y$-intercept of the line of fit.
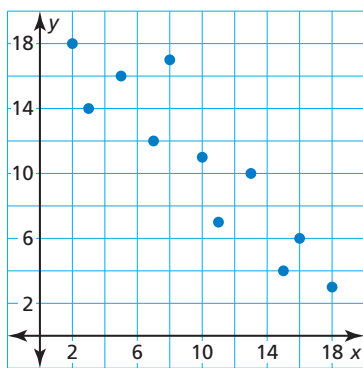
### Vocabulary and Core Concept Check

1. **COMPLETE THE SENTENCE** When data show a positive correlation, the dependent variable tends to _____ as the independent variable increases.

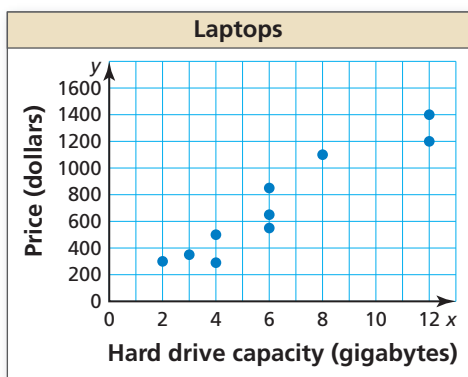2. **VOCABULARY** What is a line of fit?

### Monitoring Progress and Modeling with Mathematics

In Exercises 3–6, use the scatter plot to fill in the missing coordinate of the ordered pair.

3. (16,    )

4. (3,    )

5. (    , 12)

6. (    , 17)



7. **INTERPRETING A SCATTER PLOT** The scatter plot shows the hard drive capacities (in gigabytes) and the prices (in dollars) of 10 laptops. *(See Example 1.)*



a. What is the price of the laptop with a hard drive capacity of 8 gigabytes?

b. What is the hard drive capacity of the $1200 laptop?

c. What tends to happen to the price as the hard drive capacity increases?

8. **INTERPRETING A SCATTER PLOT** The scatter plot shows the earned run averages and the winning percentages of eight pitchers on a baseball team.
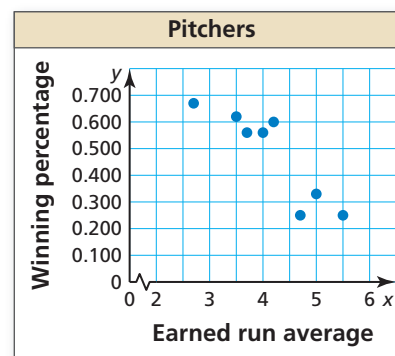


a. What is the winning percentage of the pitcher with an earned run average of 4.2?

b. What is the earned run average of the pitcher with a winning percentage of 0.33?

c. What tends to happen to the winning percentage as the earned run average increases?

In Exercises 9–12, tell whether *x* and *y* show a *positive*, a *negative*, or *no* correlation. *(See Example 2.)*

9.


10.


11.


12.
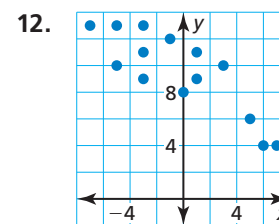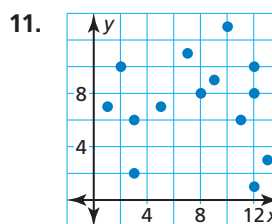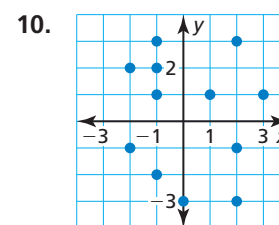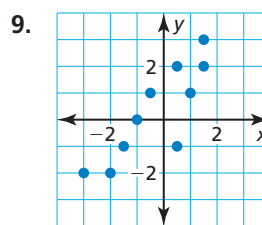
**In Exercises 13 and 14, make a scatter plot of the data. Tell whether *x* and *y* show a *positive*, a *negative*, or *no* correlation.**

**13.**

| x | 3.1 | 2.2 | 2.5 | 3.7 | 3.9 | 1.5 | 2.7 | 2.0 |
|---|-----|-----|-----|-----|-----|-----|-----|-----|
| y | 1   | 0   | 1   | 2   | 0   | 2   | 3   | 2   |

**14.**

| x | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|---|----|----|----|----|----|----|----|----|
| y | 67 | 67 | 50 | 33 | 25 | 21 | 19 | 4  |

**15. MODELING WITH MATHEMATICS** The table shows the world birth rates *y* (number of births per 1000 people) *x* years since 1960. *(See Example 3.)*

| x | 0    | 10   | 20   | 30   | 40   | 50   |
|---|------|------|------|------|------|------|
| y | 35.4 | 33.6 | 28.3 | 27.0 | 22.4 | 20.0 |

a. Write an equation that models the birthrate as a function of the number of years since 1960.

b. Interpret the slope and *y*-intercept of the line of fit.

**16. MODELING WITH MATHEMATICS** The table shows the total earnings *y* (in dollars) of a food server who works *x* hours.

| x | 0 | 1  | 2  | 3  | 4  | 5  | 6   |
|---|---|----|----|----|----|----|-----|
| y | 0 | 18 | 40 | 62 | 77 | 85 | 113 |

a. Write an equation that models the server's earnings as a function of the number of hours the server works.

b. Interpret the slope and *y*-intercept of the line of fit.

**17. OPEN-ENDED** Give an example of a real-life data set that shows a negative correlation.

**18. MAKING AN ARGUMENT** Your friend says that the data in the table show a negative correlation because the dependent variable *y* is decreasing. Is your friend correct? Explain.

| x | 14 | 12 | 10 | 8  | 6  | 4  | 2  |
|---|----|----|----|----|----|----|----|
| y | 4  | 1  | 0  | −1 | −2 | −4 | −5 |

**19. USING TOOLS** Use a ruler or a yardstick to find the heights and arm spans of five people.

a. Make a scatter plot using the data you collected. Then draw a line of fit for the data.

b. Interpret the slope and *y*-intercept of the line of fit.

**20. THOUGHT PROVOKING** A line of fit for a scatter plot is given by the equation $y = 5x + 20$. Describe a real-life data set that could be represented by the scatter plot.

**21. WRITING** When is data best displayed in a scatter plot, rather than another type of display, such as a bar graph or circle graph?

**22. HOW DO YOU SEE IT?** The scatter plot shows part of a data set and a line of fit for the data set. Four data points are missing. Choose possible coordinates for these data points.
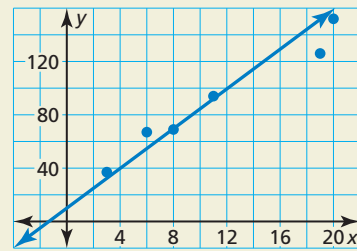


**23. REASONING** A data set has no correlation. Is it possible to find a line of fit for the data? Explain.

**24. ANALYZING RELATIONSHIPS** Make a scatter plot of the data in the tables. Describe the relationship between the variables. Is it possible to fit a line to the data? If so, write an equation of the line. If not, explain why.

| x | −12 | −9 | −7 | −4 | −3 | −1 |
|---|-----|----|----|----|----|----|
| y | 150 | 76 | 50 | 15 | 10 | 1  |

| x | 2 | 5  | 6  | 7  | 9  | 15  |
|---|---|----|----|----|----|-----|
| y | 5 | 22 | 37 | 52 | 90 | 226 |

---

**Maintaining Mathematical Proficiency** Reviewing what you learned in previous grades and lessons

**Evaluate the function when $x = -3, 0,$ and $4$.** *(Section 3.3)*

**25.** $g(x) = 6x$

**26.** $h(x) = -10x$

**27.** $f(x) = 5x - 8$

**28.** $v(x) = 14 - 3x$