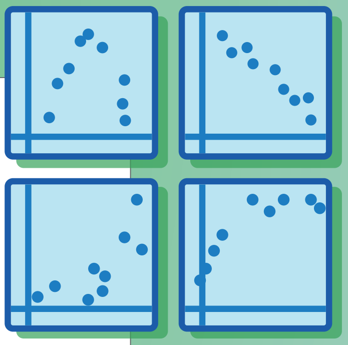
Caregiver Support | Grade 8 Unit 6



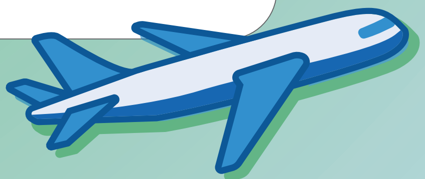
Unit 6

Associations in Data

Data is one way we make sense of the world around us. Organizing and displaying data can allow us to describe trends and make predictions. In this unit, you will investigate data points that represent two pieces of information.

Essential Questions

- What is a scatter plot and what can it tell you?
- How can lines help you model data on a scatter plot?
- How can you analyze data with two variables that are categories instead of numbers?



Summary | Lesson 1

You can organize and display data that includes numbers in different ways, including in a table and in a scatter plot. A **scatter plot** is a set of disconnected data points plotted on a coordinate plane.

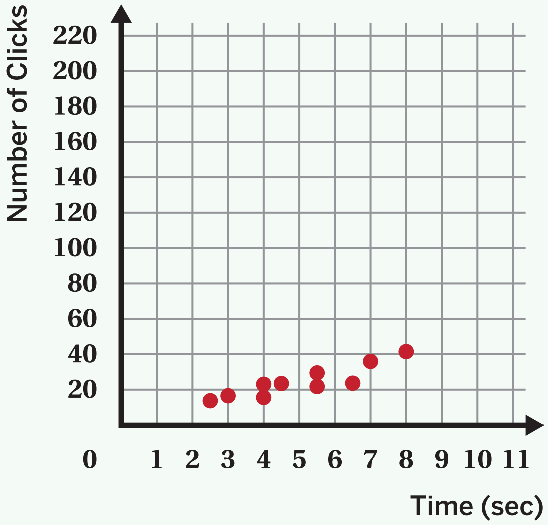
A table and a scatter plot both display the same data, but can be helpful in different ways. For example, you can use a scatter plot to investigate connections between two variables, while a table is helpful for looking for the exact values of specific data points.

Here is data showing the amount of time in seconds and the number of clicks of the button.

Table

Time (sec)	Number of Clicks
2.5	14
3	17
4	16
4	23
4.5	24
5.5	22
5.5	30
6.5	24
7	36
8	42

Scatter Plot

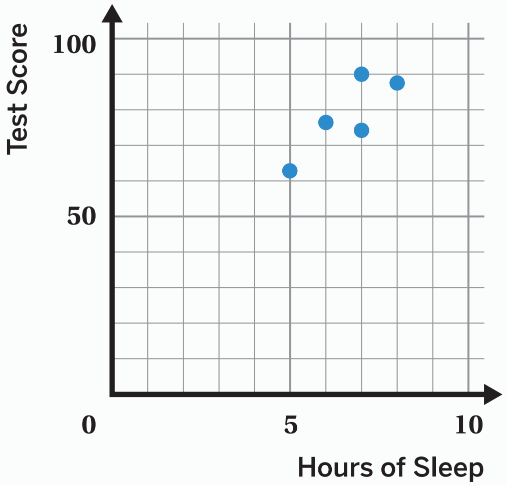


Try This

For extra credit, a group of students participated in a study where they recorded the number of hours they slept the night before a test and their test scores.

Here is a scatter plot of the data.

- Ayaan: 7 hours, score of 74
- Emika: 6 hours, score of 76
- Inola: 8 hours, score of 88
- Kwasi: 5 hours, score of 63
- Zoe: 7 hours, score of 90



a What is another way you could organize or display this data?

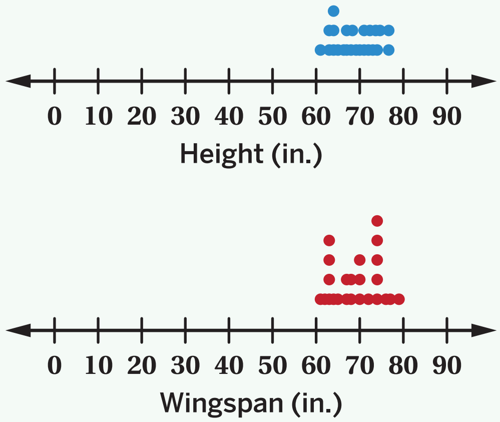
b What is one advantage of organizing the data into a scatter plot?

Summary | Lesson 2

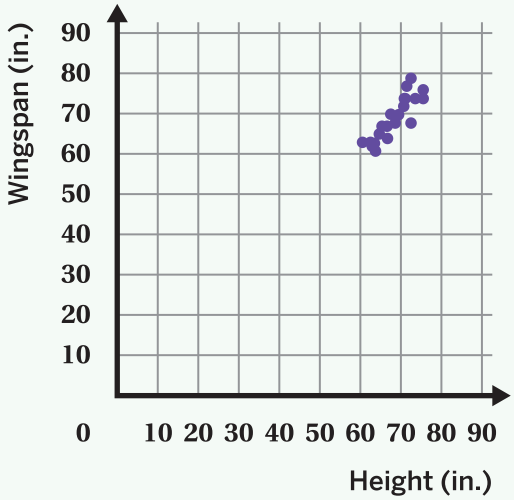
Data presented as numbers, quantities, or measurements that can be compared in a meaningful way is called *numerical data*, or *quantitative data*. You can investigate *univariate data*, which involves one variable, and *bivariate data*, which involves two variables.

There are different ways to represent numerical data. A *dot plot* shows data for one variable and a scatter plot shows data for two variables at the same time. Seeing two numerical variables at the same time allows us to notice trends and connections.

Dot Plot



Scatter Plot

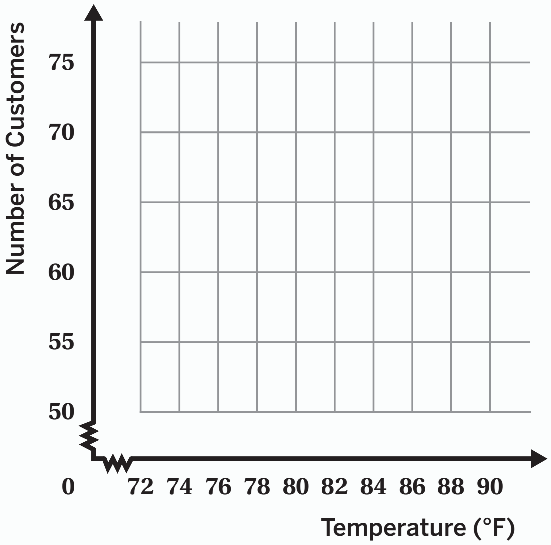


Try This

During one week, an ice cream stand collected data on the temperature outside and the number of customers.

a Create a scatter plot of this data.

Day	Temperature (°F)	Number of Customers
Monday	85	58
Tuesday	83	55
Wednesday	90	68
Thursday	75	50
Friday	85	72

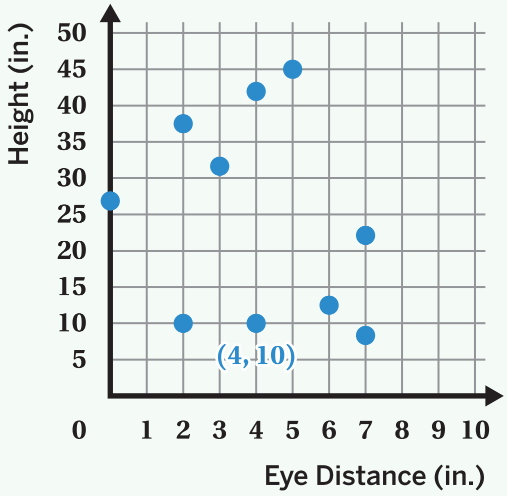


b Write a question that you can answer based on the scatter plot.

Summary | Lesson 3

A point on a scatter plot represents two pieces of information. The axis labels tell you how to interpret the coordinates of each point.

In this example, the point $(4, 10)$ represents a robot with an eye distance of 4 inches and a height of 10 inches.



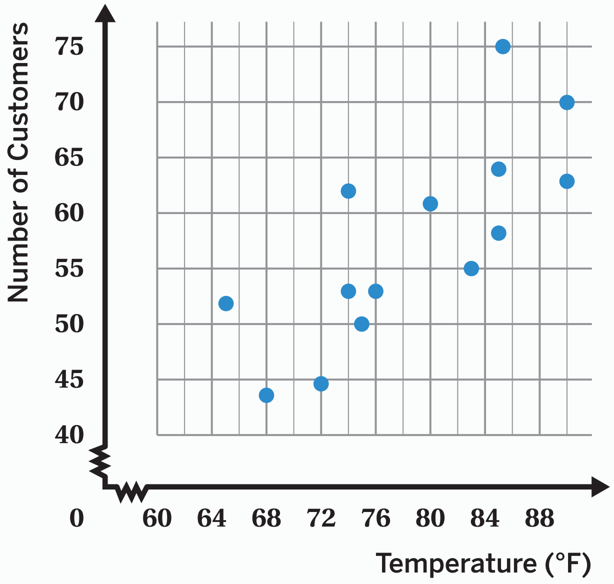
Try This

An ice cream stand collected data on the temperature outside and the number of customers over time.

a. Circle the point that represents the day the temperature reached 72°F outside.

b. Estimate the number of customers on that day.

c. What does the point $(85, 75)$ represent?

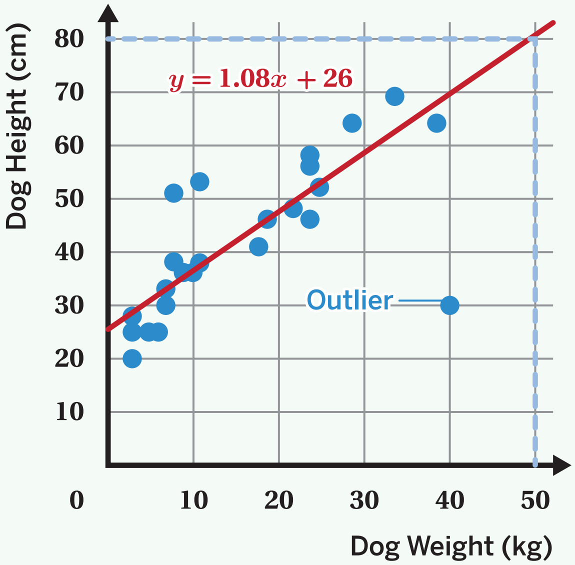


A **linear model** is a line on a scatter plot that helps identify trends in data more clearly. You can also use a linear model to make a prediction.

For example, there are two ways you can use this linear model to predict a dog's height when it weighs 50 kilograms.

- Use the graph to locate 50 on the x-axis and follow it up to meet the linear model, which shows a y-value of 80. This means when the dog's weight is 50 kilograms, its height is 80 centimeters.
- Use the equation for the linear model, $y = 1.08x + 26$, by replacing x with 50 and evaluating for y, which is approximately 80 centimeters.

You can identify an **outlier** by looking for points that are far away from the other points and from the predicted values. The point $(40, 30)$ is an outlier on the graph of dog weights and heights.

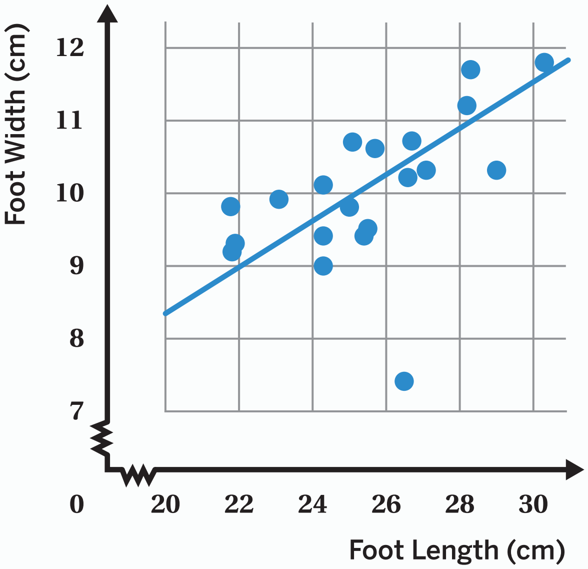


Try This

This scatter plot shows data collected about the length and width of people's feet.

a Use the linear model to predict the foot width of a person whose foot length is 28 centimeters.

b Circle the outlier on the graph.

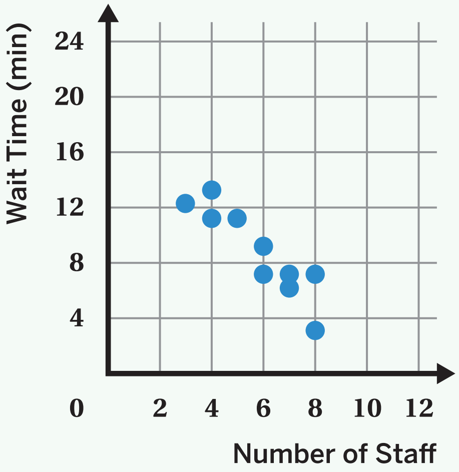


You can use a scatter plot to help identify patterns in data points and relationships between two variables.

For example, this scatter plot shows data about how long customers waited at a drive-thru restaurant and the number of staff working at that time.

The scatter plot shows both specific information and general trends, including:

- When 3 staff were working, the wait time was about 12 minutes.
- The more staff there are, the shorter the wait time seems to be.

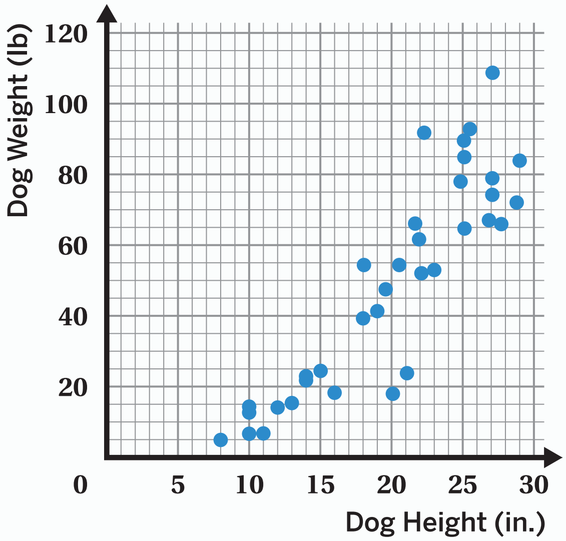


Try This

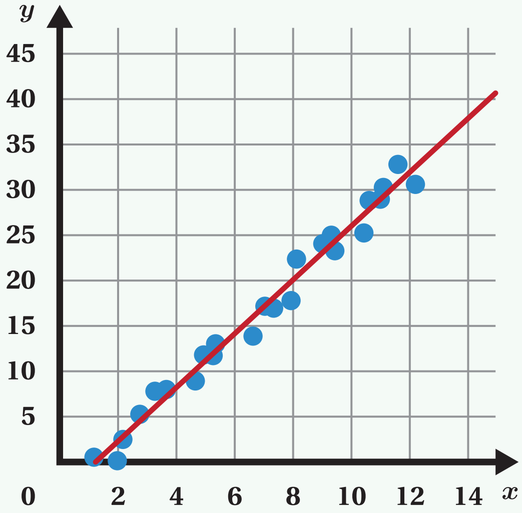
This scatter plot shows the heights and weights of 35 dogs.

a What does the point $(8, 5)$ represent in this situation?

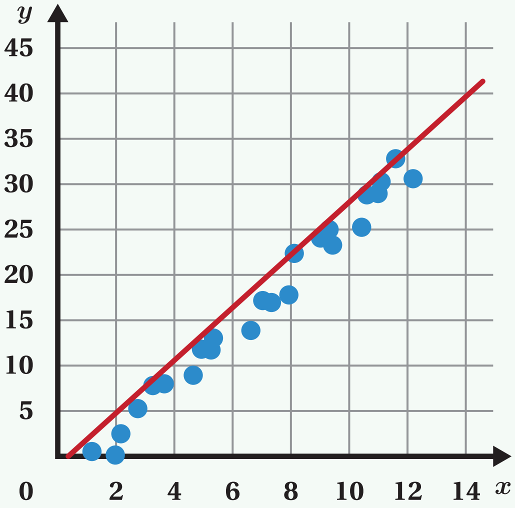
b Identify a general trend based on the scatter plot.



When creating a line of fit for a scatter plot, it's important to determine how well the line fits the data. A good line of fit follows the trend of the data, is as close to the plotted points as possible, and has about the same number of points above and below the line. The line may pass through some, all, or none of the points.



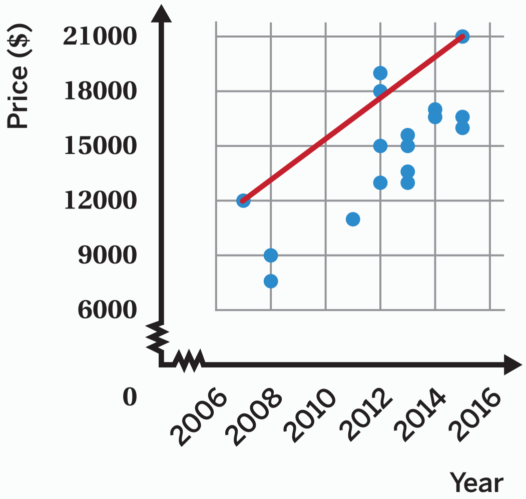
This line is a good fit for the data.



This line is not a good fit for the data.

Try This

Here is data about the price of a used car and the year that the car was manufactured.

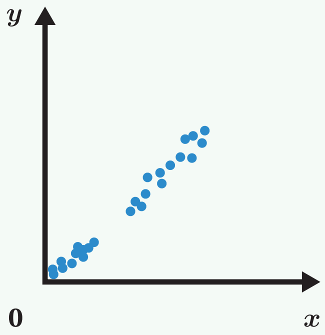


a) Explain why this model isn't a good fit for the data.

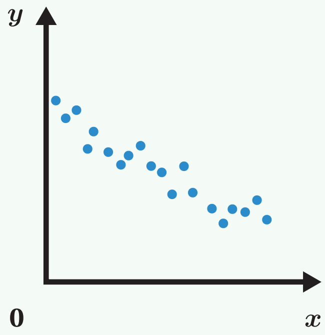
b) Draw a line of fit that better represents the data for this scatter plot.

Summary | Lesson 7

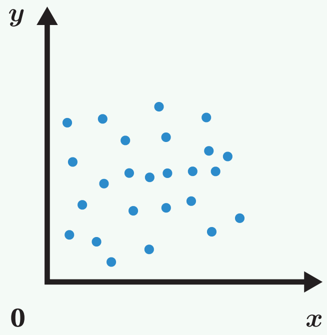
If two variables on a scatter plot are related, there is an **association**. The slope of a linear model can help determine the type of association. A positive association means that when one variable increases, the other also increases. A negative association means that when one variable increases, the other decreases. If the scatter plot shows no clear trend between the two variables, then the variables have no association.



Positive association



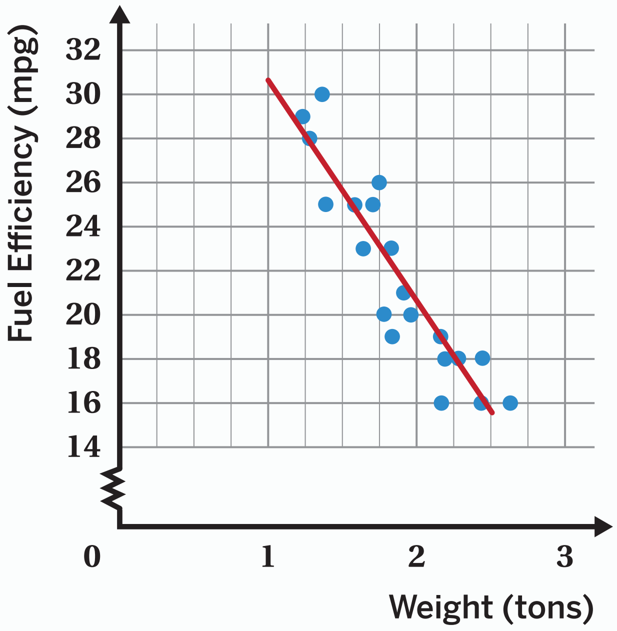
Negative association



No association

Try This

Here is a scatter plot that shows data on the weight and fuel efficiency (miles driven per gallon of fuel) of 21 cars.



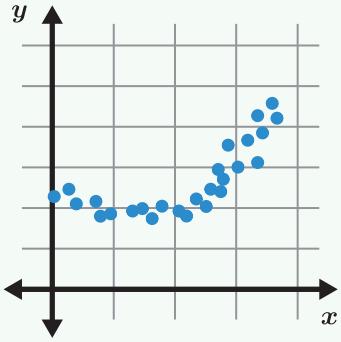
a What is the type of association between weight and fuel efficiency?

b The line of fit has a slope of about -10. What does this number mean for the weight of a car and its predicted fuel efficiency?

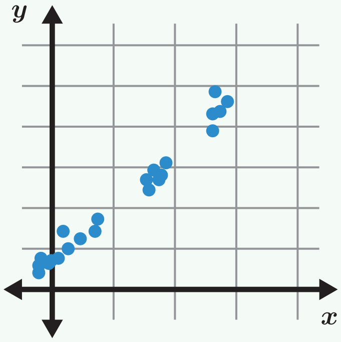
Summary | Lesson 8

When you can model data on a scatter plot with a straight line, we say it has a linear association. Data that can't be modeled by a straight line has a non-linear association. Sometimes groups of data points appear close together, which are called **clusters**.

This scatter plot is an example of a non-linear association, without clusters.



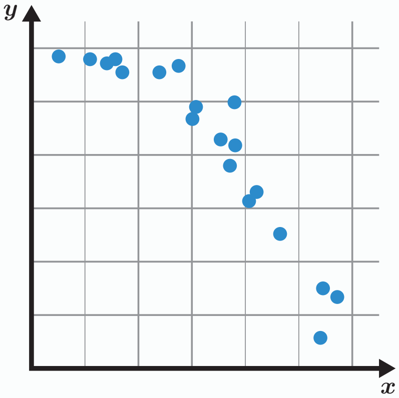
This scatter plot is an example of a linear association, with clusters.



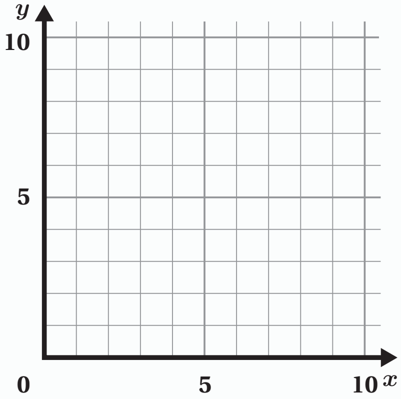
Try This

a Describe this scatter plot using at least two terms from the word bank.

positive association	negative association	clusters
linear association	non-linear association	outlier



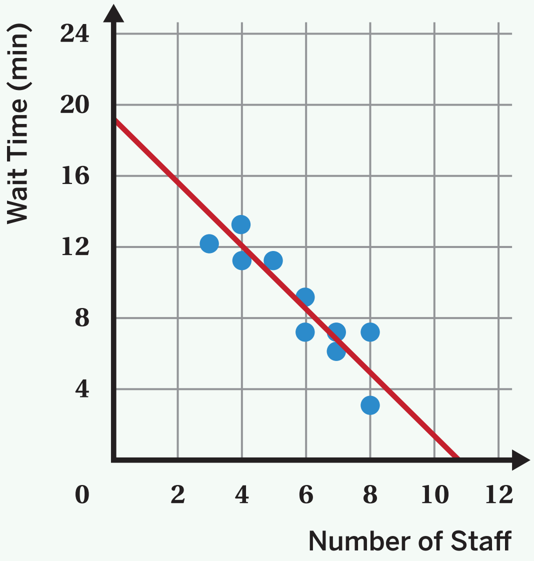
b Plot points to create a scatter plot that shows a negative linear association with clustering.



By understanding the association between two variables, you can make predictions about unknown values. When there's a linear association, using a linear model can often make predictions more accurate.

For example, this scatter plot shows data about how many minutes customers waited at a drive-through restaurant and the number of staff working at that time. This data can be modeled by the equation $y = -1.75x + 19$.

- The *slope* of the linear model is -1.75, which means that if the number of staff increases by 1 person, the wait time decreases by 1.75 minutes.
- The linear model predicts that if there are 2 staff working, the wait time will be approximately 15.5 minutes.
- But the linear model also predicts that when there are 0 staff working, the wait time will be 19 minutes, which is impossible!



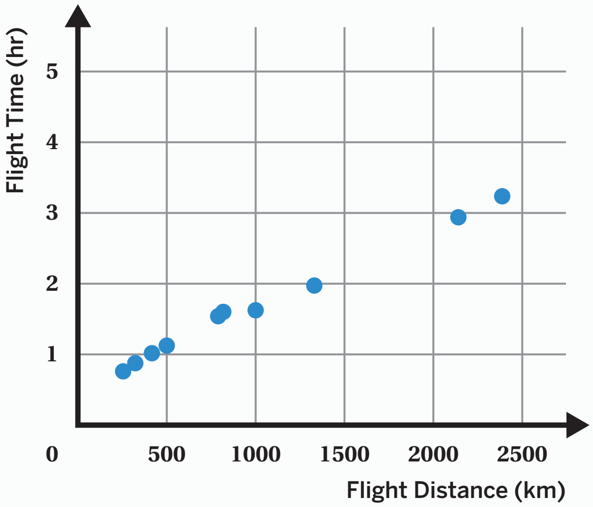
Try This

This scatter plot shows the distances and times for a set of flights.

a Draw a line of fit that best represents the data.

b Describe the association between flight distance and flight time.

c Use your model to predict the flight time when the flight distance is 2,000 kilometers.



Summary | Lesson 10

You can use a **two-way table** to compare two variables of *categorical data*, which is data that can be sorted into categories.

This two-way table shows data about whether students meditated on a certain day and whether they felt calm or agitated that day. Each entry in the table represents the **frequency**, or the number of times, that a category appears in the data set.

	Meditated	Did Not Meditate	Total
Calm	45	8	53
Agitated	23	21	44
Total	68	29	97

You can use these two-way tables to investigate possible connections between variables. In the example, we can see there's a connection between meditating and feeling calm since a majority of the people who felt calm also meditated.

Try This

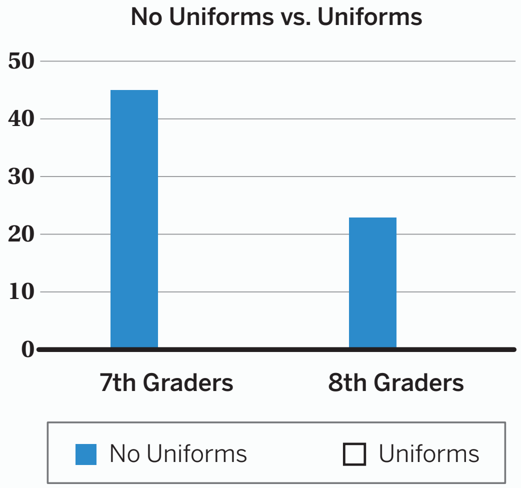
Students took a survey on whether they want to wear uniforms or not. The table shows their responses.

	No Uniforms	Uniforms	Total
7th Graders	45		53
8th Graders	23	21	
Total		29	97

a Complete the two-way table.

b What does the number 53 mean in this situation?

c Complete the bar graph for this situation with the data about students who want uniforms.



Summary | Lesson 11

You can use specific types of two-way tables and bar graphs to show frequencies and percentages within data sets.

The **relative frequency** of a category is the fraction or percentage of the data set that's in that category. A two-way table of relative frequencies shows the fraction or percentage of each category instead of the number of data points.

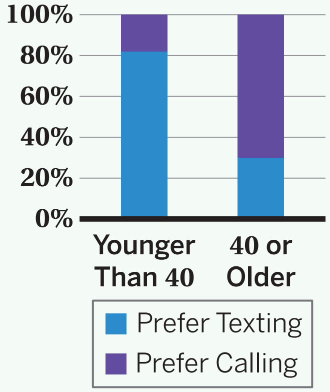
Relative Frequencies

	Prefer Texting	Prefer Calling	Total
Younger Than 40	82%	18%	100%
40 or Older	33%	67%	100%

A **segmented bar graph** compares different categories within a data set. Each bar represents all the data within one category, or 100%. The bars are each separated into parts, or segments, that show what percentage each part makes up of the whole category.

We can use representations like these to identify associations between two categorical variables. For example, the table and graph show an association between categorical variables, age, and communication preference.

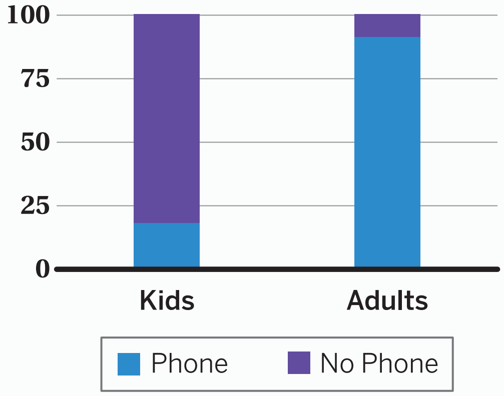
Segmented Bar Graph



Try This

Here are two representations of relative frequency.

Cell Phone Ownership



Lucky Socks and Winning Bingo

	Winners	Losers	Total
Lucky Socks	80%	20%	100%
Regular Socks	79%	21%	100%

a. Is there an association between age and cell phone ownership? Explain your thinking.

b. Is there an association between wearing lucky socks and winning bingo? Explain your thinking.

Try This | Answer Key

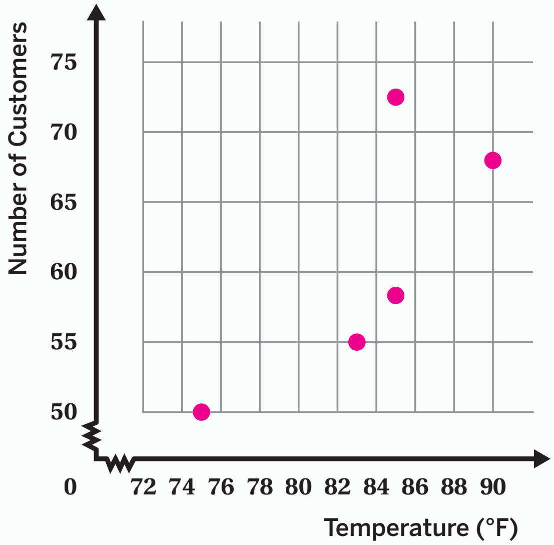
Lesson 1

a *Responses vary*. Create a table that sorts the data by hours of sleep or score.

b *Responses vary*. In a scatter plot, it is easier to see the relationship between both variables.

Lesson 2

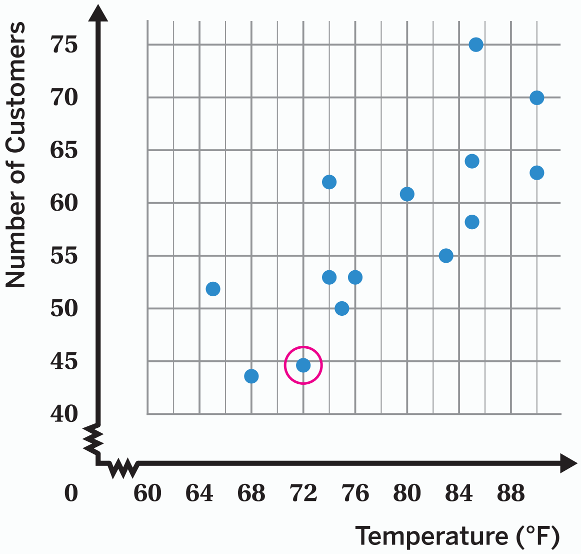
a



b *Responses vary*. How are temperature and number of customers related? What trends, if any, can be seen from the scatter plot?

Lesson 3

a



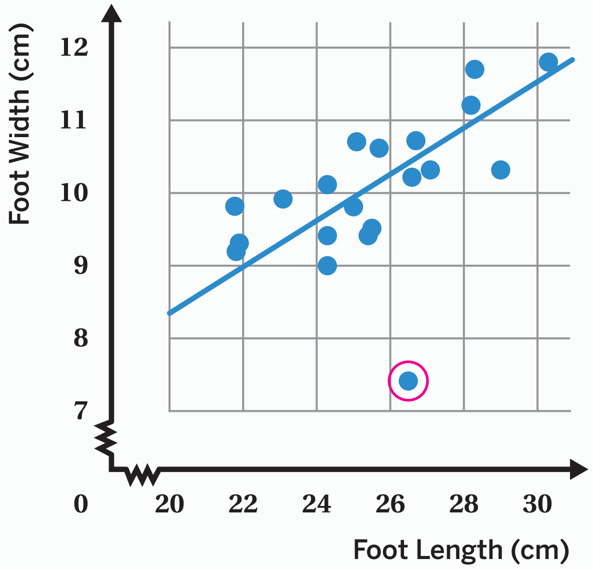
b There were approximately 44 customers when the temperature outside was 72°F.

c The point (85, 75) represents a temperature of 85°F on a day with 75 customers.

Lesson 4

a Responses between 10.75 and 10.95 centimeters are considered correct.

b



Lesson 5

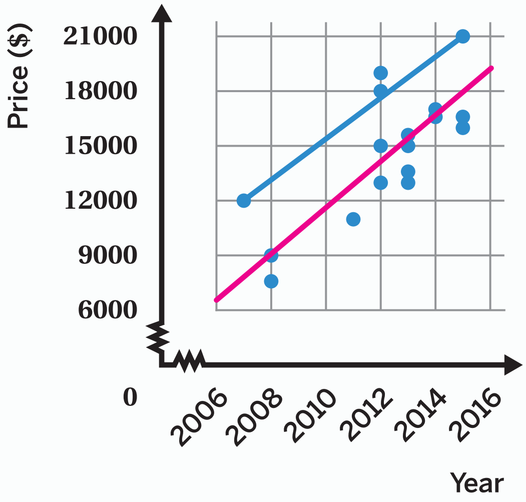
a The point (8, 5) represents a dog that is 8 inches tall and weighs 5 pounds.

b According to the scatter plot, a general trend is that taller dogs weigh more.

Lesson 6

a *Responses vary*. This model isn't a good fit because more of the data is below the line than above the line. This means that the line would overpredict most of the data.

b Sample shown on graph.



Lesson 7

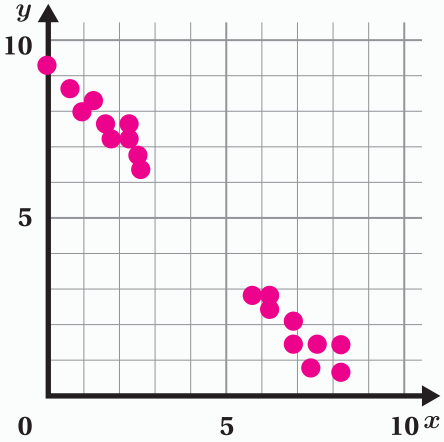
a *Responses vary*. There is a negative association between weight and fuel efficiency. As the weight of the car increases, the fuel efficiency decreases.

b If the weight of a car increases by 1 ton, the model predicts that the car's fuel efficiency will decrease by 10 miles per gallon.

Lesson 8

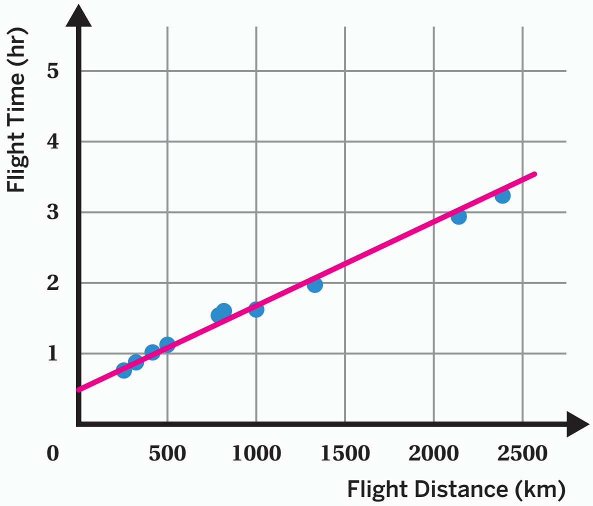
a *Responses vary*. This scatter plot shows a negative non-linear association without clusters.

b *Responses vary*. Sample shown on graph.



Lesson 9

a



b *Responses vary*. Flight distance and flight time have a positive linear association meaning that as flight distance increases, flight time increases.

c *Responses vary*. A flight with a distance of 2,000 kilometers will have a flight time of just under 3 hours.

Try This | Answer Key

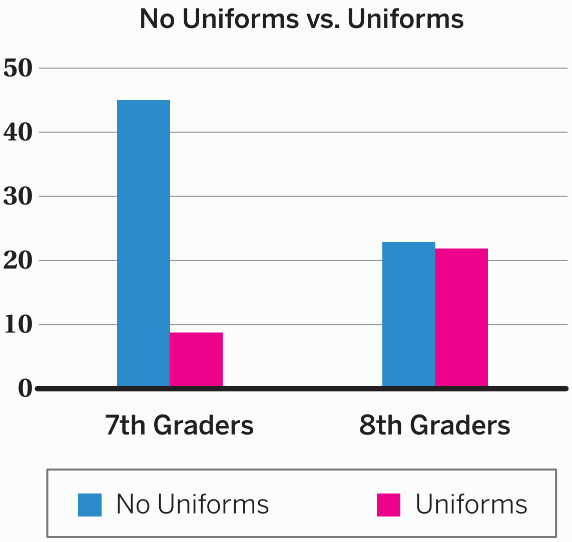
Lesson 10

a

	No Uniforms	Uniforms	Total
7th Graders	45	**8**	53
8th Graders	23	21	**44**
Total	**68**	29	97

b In this situation, 53 represents the total number of 7th graders that responded to the survey.

c



Lesson 11

a Yes. *Explanations vary.* There is an association between age and cell phone ownership. The percentage of adults who own a cell phone is much larger than the percentage of kids who own a cell phone.

b No. *Explanations vary.* There isn't an association between wearing lucky socks and winning bingo. The percentage of bingo winners wearing lucky socks is only 1% greater than the percentage of bingo winners wearing regular socks.