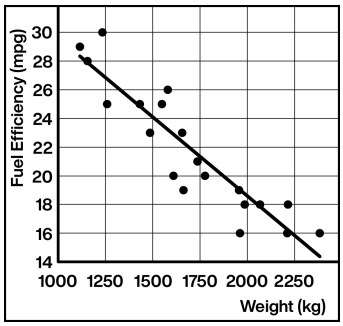
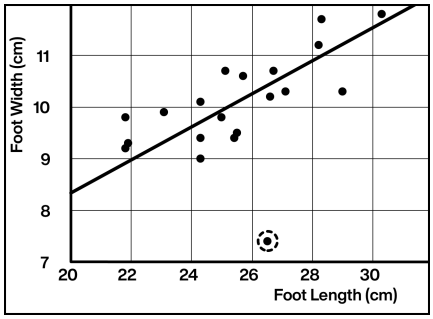
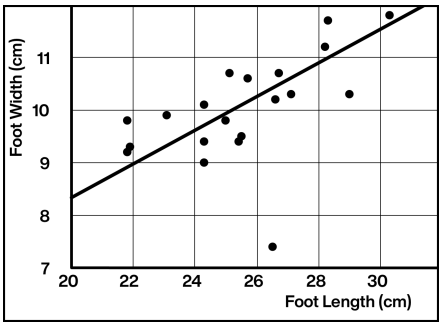
Associations in Data
Student Guide

Math 7 Unit 6 Accelerated

desmos

Unit 8.6, Student Goals and Glossary

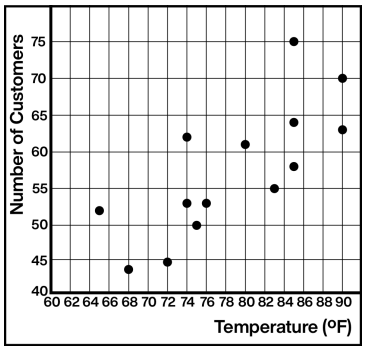
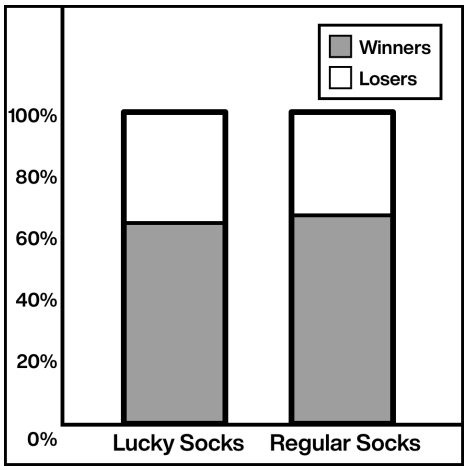
Glossary

Term	Definition
negative association	A negative association is a relationship between two quantities where one tends to decrease as the other increases. 
outlier	An outlier is a data value that is far from the other values in the data set. 
positive association	A positive association is a relationship between two quantities where one tends to increase as the other increases. 
relative frequency	The relative frequency of a category tells us the fraction or percent of the data set that is in this category.

	Meditated	**Did Not Meditate**
Calm	66%	28%
Agitated	34%	72%
Total	100%	100%

desmos

Unit 8.6, Student Goals and Glossary

scatter plot	A scatter plot is a set of disconnected data points plotted on a coordinate plane. It allows us to investigate connections between two variables. 
segmented bar graph	A segmented bar graph compares two categories within a data set. The whole bar represents all the data within one category. Then, each bar is separated into parts (segments) that show the percentage of each part in the second category. 
two-way table	A two-way table provides a way to compare two categorical variables. It shows one of the variables across the top and the other down one side. Each entry in the table is the frequency or relative frequency of the category shown by the column and row headings.

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

desmos

Unit 8.6, Family Resource

Unit 6 Summary

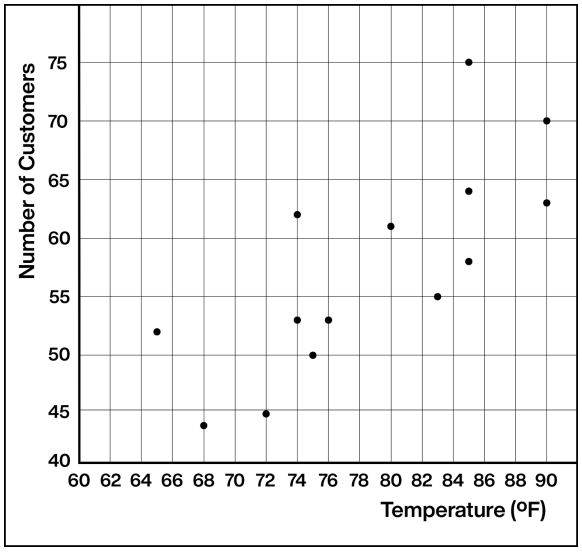
Prior Learning	Math 8, Unit 6	Future Learning
Grade 3 ● Drawing and interpreting bar graphs Math 6 ● Plotting points in the coordinate plane ● Shape, center, spread, and outliers of data in one variable Math 8, Unit 3 ● Linear relationships	● Organizing numerical data ● Analyzing numerical data ● Categorical data	High School ● Fitting data with nonlinear functions ● Correlation coefficients

Organizing Numerical Data

Lists, sorted tables, scatter plots, and dot plots are all ways we can organize numerical data.

Scatter plots show us how two different variables are related.

This is data for an ice cream stand collected on temperature and number of customers over time.



With a scatter plot, we can investigate visual patterns and make predictions.

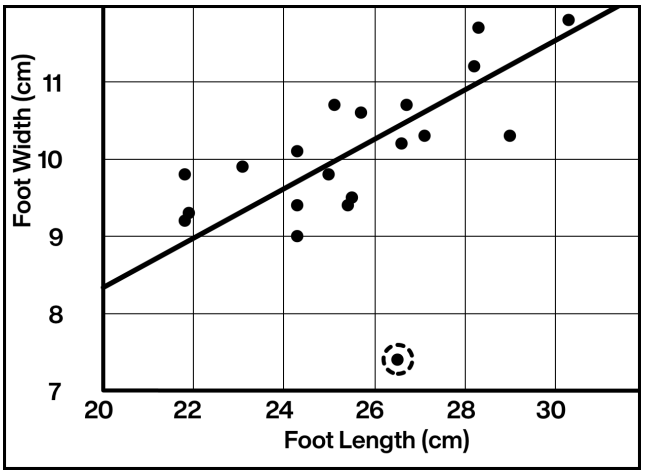
Some questions scatter plots can help answer:

- Is there an association between the outside temperature and the number of customers at an ice cream store?
- What is the predicted number of customers if the temperature is 78ºF?

desmos

Unit 8.6, Family Resource

Analyzing Numerical Data



We say that there is a **positive association** between foot length and foot width because in general, longer feet are wider than shorter feet.

The line drawn on the graph shows the overall trend and can help us find the predicted foot width for a given foot length.

Points that are not close to the line and to most of the data are called **outliers**.

Categorical Data

When we collect data by counting things in various categories, such as tall or short, we call that categorical data. To help organize categorical data, we can use two-way tables and bar graphs.

This table shows states of mind of athletes during a meet and whether or not they meditated beforehand.

23 of the people who meditated were anxious. Only 21 of the people who did not meditate were anxious.

Does this mean that meditation has no impact or even a slight negative association with mood?

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

It can be more helpful to examine the percentages (called *relative frequencies*) in each category.

Of the people who meditated, 66% were calm. 28% of the people who did not meditate were calm.

The group that meditated has a lower percentage of athletes who are anxious.

	Meditated	Did Not Meditate
Calm	66%	28%
Anxious	34%	72%
Total	100%	100%

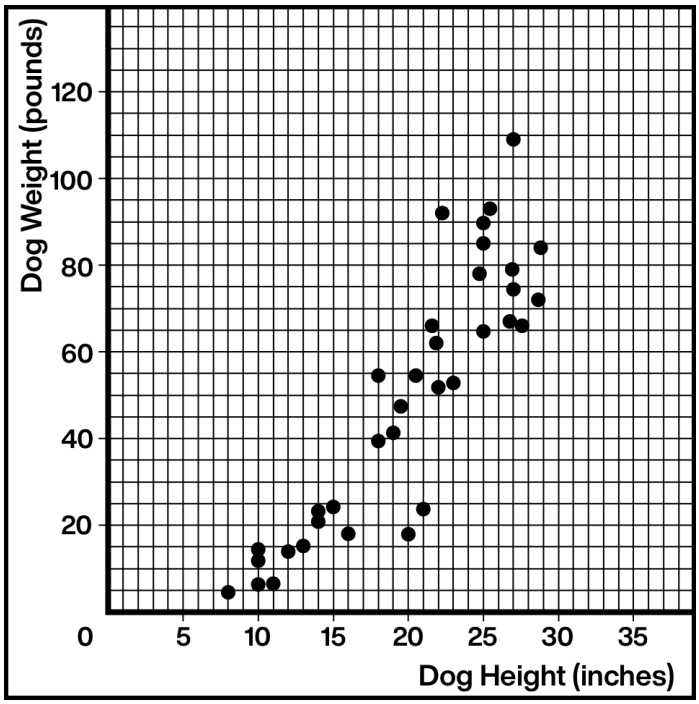
desmos

Unit 8.6, Family Resource

Try This at Home

Organizing Numerical Data

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



1.1 Add a point to represent a dog that is 15 inches tall and weighs 35 pounds.

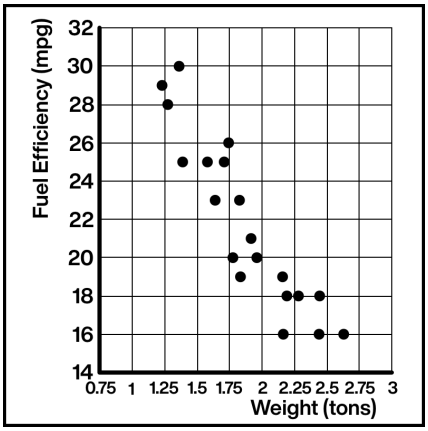
1.2 Add a point to represent a dog that weighs 100 pounds.

1.3 How many dogs in the set are about 25 inches tall? Explain how you know.

1.4 What is the heaviest weight for a dog in the set? Explain how you know.

Analyzing Numerical Data

Here is data on the weight of 21 cars and their fuel efficiency (miles driven for each gallon of gas).



2.1 How many cars have a fuel efficiency that is greater than 22 miles per gallon? Explain your thinking.

2.2 Do the variables in the scatter plot show a positive association or a negative association? Explain your thinking.

2.3 What is the predicted fuel efficiency of a car that weighs 1.5 tons?

2.4 Add an outlier to the scatter plot. Explain why this point is an outlier.

desmos

Unit 8.6, Family Resource

Categorical Data

This data is about people in various age groups and whether they use their cell phone as their alarm clock.

	Uses Cell Phone as Alarm	Does Not Use Cell Phone as Alarm	Total
18 **to** 29 **Years Old**	47	16	63
30 **to** 49 **Years Old**	66	23	87
50 + **Years Old**	31	39	70
Total	144	78	220

3.1 Fill in the blanks with the relative frequencies for each row in the table below. In other words, calculate the percent of people in each age group who use their phone as an alarm.

	Uses Cell Phone as Alarm	Does Not Use Cell Phone as Alarm	Total
18 **to** 29 **Years Old**	75%		100%
30 **to** 49 **Years Old**			
50 + **Years Old**			

3.2 Is there an association between cell phone alarm use and age for 18- to 29-year-olds and 30- to 49-year-olds? Explain your thinking.

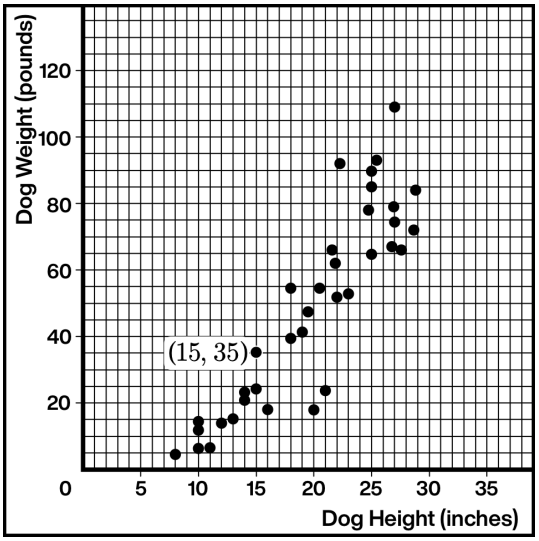
3.3 Is there an association between cell phone alarm use and age for the youngest age bracket and the 50 + age bracket? Explain your thinking.

desmos

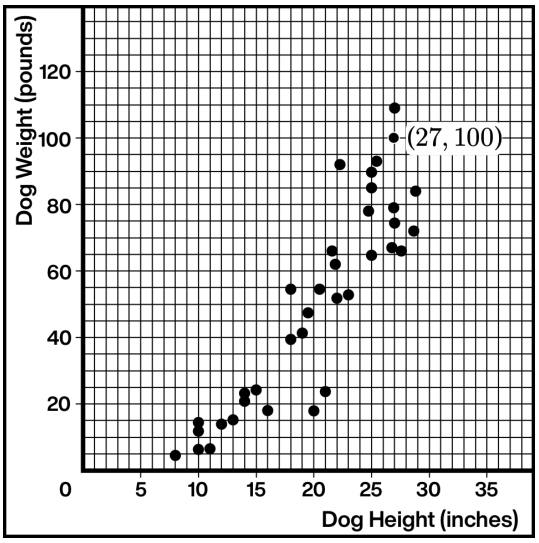
Unit 8.6, Family Resource

Solutions:

1.1



1.2 A point that represents a dog with a weight of 100 pounds will have a y-value of 100, such as the point in the graph below.



1.3 Each point on the scatter plot represents one dog. Since the x-axis represents the height of a dog in inches, we can find the number of points that have an x-value of 25 to help us determine how many dogs are about 25 inches tall. There are 5 points that have an x- value of roughly 25, so there are 5 dogs that are about 25 inches tall.

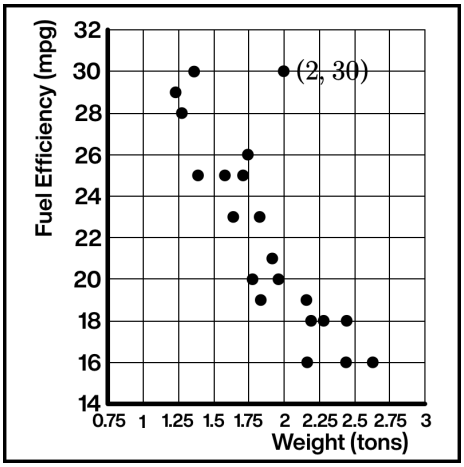
1.4 Since the y-axis represents the weight of a dog in pounds, we can find the heaviest dog by finding the y-value of the highest point on the scatter plot, which is about 110. This means that the heaviest dog in the set is about 110 pounds.

2.1 Since the y-axis represents fuel efficiency in miles per gallons, we can find the points on the graph that describe cars with a fuel efficiency greater than 22 miles per gallon by finding the number of points that have a y-value greater than 22. There are nine points with a y-value greater than 22, so there are nine cars with a fuel efficiency that is greater than 22 miles per gallon.

2.2 There is a negative association between weight and fuel efficiency because as the weight of the car increases, the fuel efficiency decreases.

2.3 Using the line of fit, we can predict that a 1.5-ton car will have a fuel efficiency of about 25.5 miles per gallon.

2.4 $(2, 30)$ is one example of an outlier. It is far from the rest of the data and has far greater fuel efficiency than the line of fit predicts for a car that weighs 2 tons.



3.1 $\frac{47}{63}$ is approximately 0.75, so the percentage of 18- to 29-years-olds who use a cell phone as an alarm clock is about 75%.

	Uses Cell Phone as Alarm	**Does Not Use Cell Phone as Alarm**	**Total**
18 **to** 29 **Years Old**	75%	25%	100%
30 **to** 49 **Years Old**	76%	24%	100%
$50+$ **Years Old**	44%	56%	100%

3.2 There is not an association between cell phone alarm use and age for 18- to 29-year-olds and 30- to 49-year-olds because the relative frequencies are very similar (75% vs. 76% and 25% vs. 24%).

3.3 Using a cell phone as an alarm is associated with being in the younger age brackets. About 75% of 18**-** to 29-year olds use their cell phone as an alarm, but only 44% of people 50 years or older do.

desmos

Unit 8.6, Lesson 1: Notes Name _____

Organizing Data

Learning Goal(s):

- **I can organize data to notice patterns more clearly.**
- **I can describe the advantages and disadvantages of organizing data in different ways.**

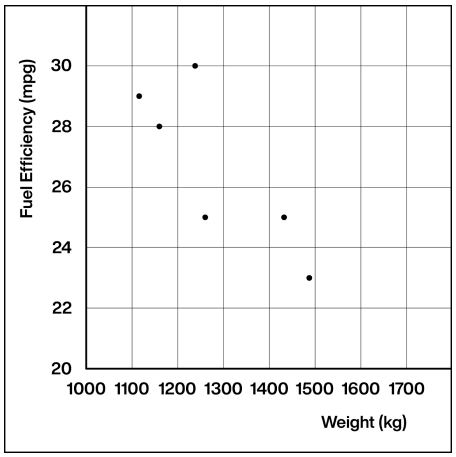
We can organize and display data with two variables in different ways.

Latifa is curious about different cars and their fuel efficiency (miles driven for each gallon of gas).

Table

Car Weight (kg)	Fuel Efficiency (mpg)
1116	28.93
1160	27.88
1238	29.94
1260	24.95
1432	25
1487	22.96

Scatter Plot



Predict the fuel efficiency of a typical car that weighs 1600 kilograms.

Approximately 21 miles per gallon.

A teacher asked her students how many hours of sleep they had the night before a test.

How might you organize or display this data?

Responses vary.

- **Sort the table by hours of sleep or score.**
- **Create a scatter plot.**

Why might someone want to organize it this way?

Responses vary.

- **In a sorted table, it is easier to see patterns.**
- **In a scatter plot, it is easier to see the relationship between both variables.**

Student	Hours of Sleep	Score
Ayaan	7	74
Emika	6	76
Inola	8	88
Kwasi	5	63
Zoe	7	90

Summary Question

What is one advantage of representing data in . . .

. . . a scatter plot? ***Responses vary.*** . . . a table? ***Responses vary.***

It is easier to see trends in the data. **It is easier to see values of individual data points.**

desmos

Unit 8.6, Lesson 2: Notes Name _____

Plotting Data

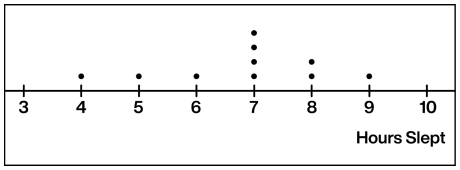
Learning Goal(s):

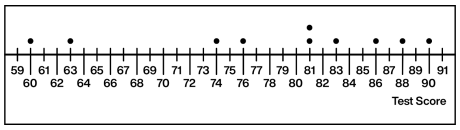
- **I can compare and contrast two different ways to display data (a dot plot and a scatter plot).**
- **I can draw a scatter plot to represent data.**

Representing data with a scatter plot is different from ways we have represented data before.

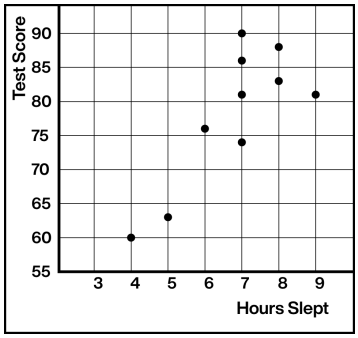
A teacher asked her students how many hours of sleep they had the night before a test.

Dot Plots





Scatter Plot



What is different about the two ways of representing the data?

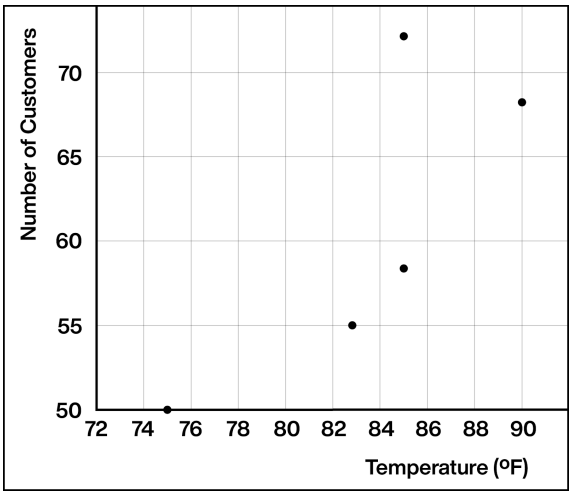
Responses vary.

Dot plots separate each variable and scatter plots show both variables at once.

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

Create a scatter plot of this data.

Day	Temperature (°F)	Number of Customers
Monday	85	58
Tuesday	83	55
Wednesday	90	63
Thursday	75	50
Friday	85	72



Summary Question

Scatter plots allow us to investigate possible connections between two numerical variables. Explain what this sentence means in your own words.

Responses vary. **This sentence means that scatter plots help us see possible relationships between two variables that involve counting or numbers.**

desmos

Unit 8.6, Lesson 3: Notes Name _____

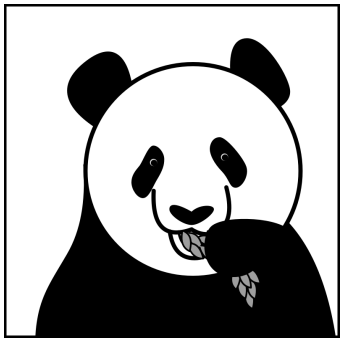
What a Point in a Scatter Plot Means

Learning Goal(s):

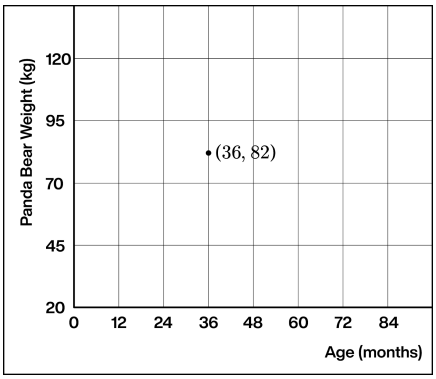
- **I can describe the meaning of a point on a scatter plot in context.**

Scatter plots are made up of many individual data points. What do each of those points represent?

A giant panda lives in a zoo. What does the point on the graph tell you about the panda?



The panda is 36 months old and weighs 82 kilograms.



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

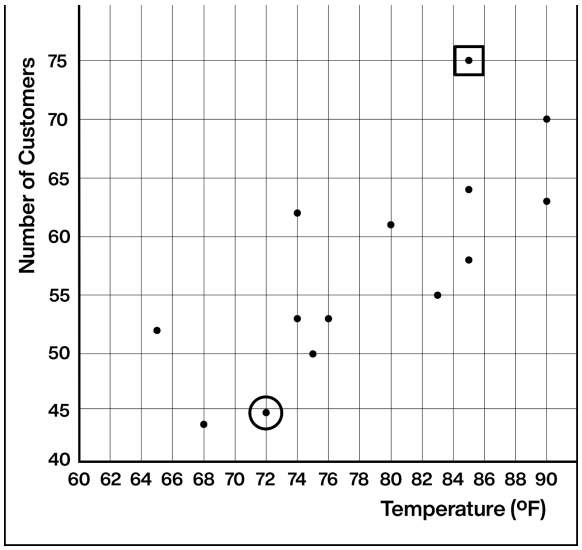
Put a circle around the data point that represents the day it was 72° F outside.

Put a square around the day when the stand had the most number of customers.

Why might the ice cream stand want to collect and visualize this data?

Responses vary.

The ice cream stand might be curious if there is an association between the weather and how many customers they have so they can plan how much ice cream to make that day.



Summary Question

Describe a strategy to determine what a single point on a scatter plot means.

One strategy to determine what a point means is to look at the x-value and the label on the x-axis to figure out what one variable represents. Do the same with the y-value and y-axis to determine the other variable.

desmos

Unit 8.6, Lesson 4: Notes Name _____

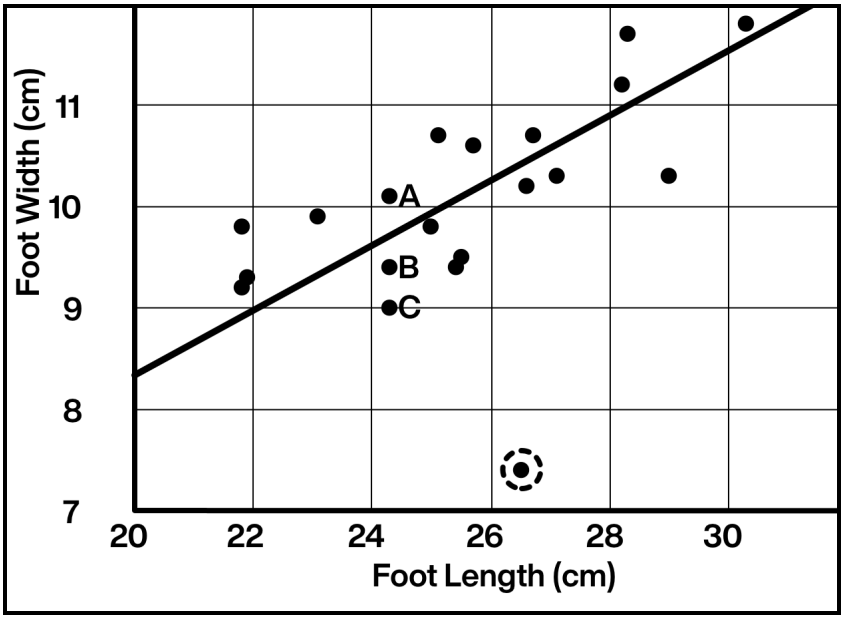
Lines of Fit and Outliers

Learning Goal(s):

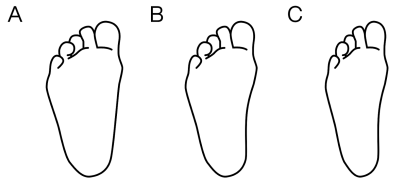
- **I can use a line of fit to predict values not in the data.**
- **I can identify outliers on a scatter plot.**

What if we want to make predictions about data not in the original data set? We can use linear functions to model data on a scatter plot. Models typically fit some data points well and not others.

This is data collected about different feet's length and width.



These three feet are all size 8 (about 24 centimeters long).



Label each foot on the scatter plot.

Response on the graph.

Which foot does the linear model fit best? **B** Explain your thinking.

Responses vary. **Foot B is closest to the fitted line.**

Circle the outlier on this graph. On the right, draw what the outlier foot might look like. ***Response on the graph.***

Is the outlier wider or less wide than predicted for its length?
Responses vary. **The outlier foot will be much less wide than predicted.**

Responses vary.

Summary Question

What does it mean for a data point to be an outlier?

Responses vary**. A data point is an outlier if it is far away from the other points.**

desmos

Unit 8.6, Lesson 5: Notes Name _____

Fitting a Line to Data

Learning Goal(s):

- **I can draw a line to fit data in a scatter plot.**
- **I can describe features of a line that fits data well.**

For any given data set with a linear association, there are infinite linear models we can draw. How do we decide what linear models are good fits for the data?

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

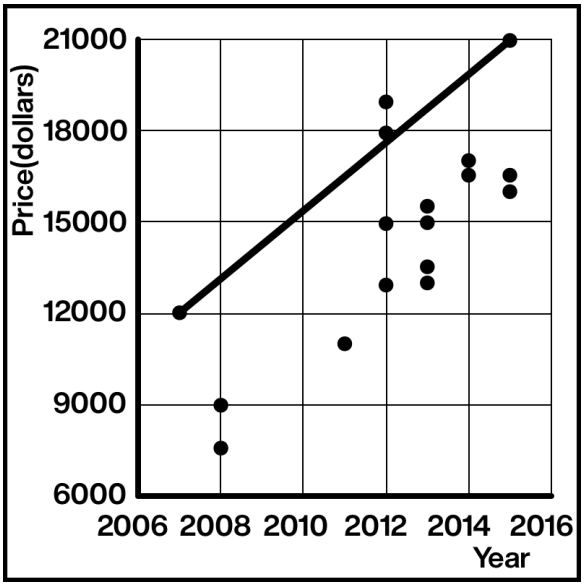
Saanvi drew this line of fit for the data.

Why might she have chosen this line?

Responses vary. **This line goes through the leftmost point on the scatter plot and one of the rightmost points on the scatter plot.**

Explain why this model is not a good fit for the data.

Responses vary. **This model is not 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.**



Responses vary.

Draw a linear model that fits the data better. Explain how you chose your model.

Responses vary. **This model is a better fit because the line has a positive trend, and about half of the data points are below the line of fit and half of the points are above the line of fit.**

Summary Question

Describe some characteristics of a line that is a good fit for the data in a scatter plot.

Responses vary.

- **The direction (or slope) of the line matches the trend of the data.**
- **The line passes through the "middle" of the data.**
- **The points are as close as possible to the line.**

desmos

Unit 8.6, Lesson 6: Notes Name _____

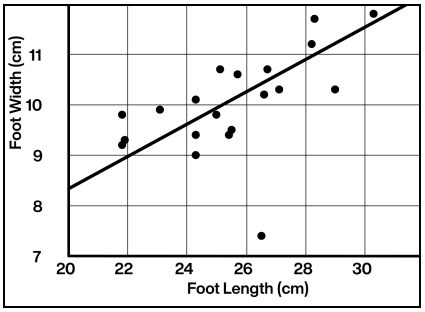
The Slope of a Fitted Line

Learning Goal(s):

- **I can explain whether data in a scatter plot has a positive association, a negative association, or neither.**
- **I can interpret the slope of a line fit to data in a real-world situation.**

Sometimes we want to know how two variables are related. In this case, we can use the slope of a linear model to explain how increasing one variable typically changes the other.

Here is a scatter plot of foot length and width for various feet.

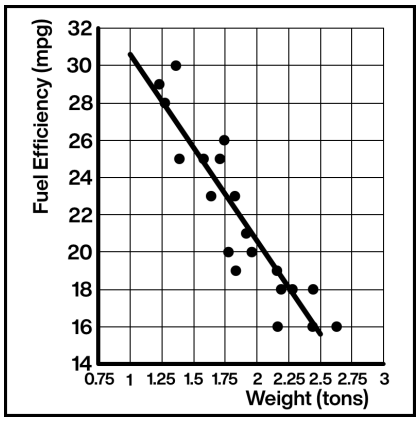


As foot length increases, foot width tends to **increase**.

This means there is a **positive association**/negative association/ no association between foot length and foot width.

The slope of the fitted line is about 0.32. If the length of a foot increases by 1 **cm**, the model predicts that its width will **increase**/decrease by 0.32 **cm**.

Here is data on the weight of 21 cars and their fuel efficiency (miles driven for each gallon of gas).



Describe the relationship between weight and fuel efficiency.

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

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

If the weight of a car increases by 1 ton, the model predicts that its fuel efficiency will decrease by 10 mi./ gal.

Summary Question

When looking at a scatter plot of data, how can we tell if there is . . .

. . . a positive association?	. . . a negative association?	. . . no association?
As one variable increases, the other variable tends to increase.	**As one variable increases, the other variable tends to decrease.**	**There is no clear pattern or trend.**

desmos

Unit 8.6, Lesson 7: Notes Name _____

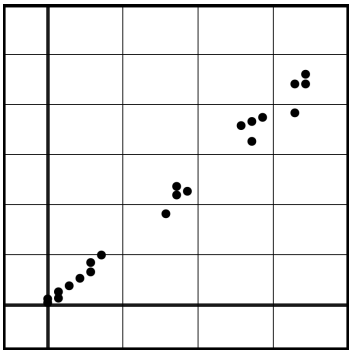
Observing More Patterns in Scatter Plots

Learning Goal(s):

- **I can use a scatter plot to decide if two variables have a linear association and make connections to what the data represents.**
- **I can pick out clusters in data and make connections to what the data represents.**

Sometimes the points in a scatter plot show an association, and sometimes there is no association.

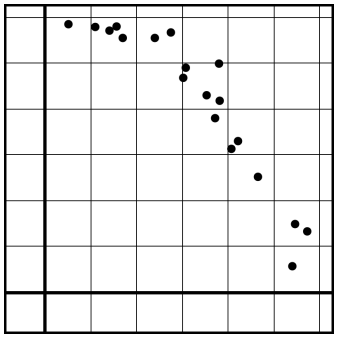
Circle the terms that describe the association in each scatter plot.



Positive / negative / no association

Linear / non-linear association

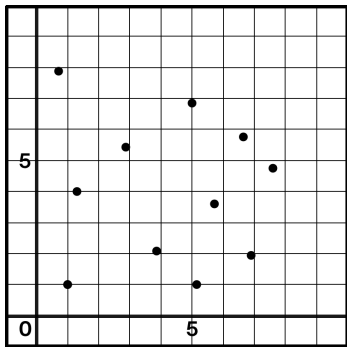
With / without clustering



Positive / **negative** / no association

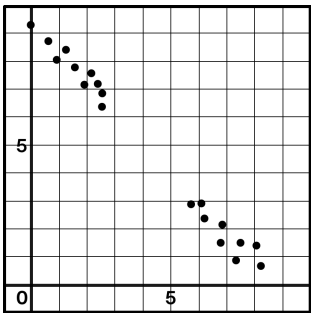
Linear / **non-linear** association

With / **without** clustering



Draw a scatter plot that shows no association.

Responses vary.



Draw a scatter plot that shows a negative linear association with clustering.

Responses vary.

Summary Question

What is a strategy you can use to decide if two variables have a linear association?

Responses vary. **Draw a line through the middle of the points with a slope that has the same sign as the association. If the points follow the pattern of the line, it is a linear association.**

desmos

Unit 8.6, Lesson 8: Notes Name _____

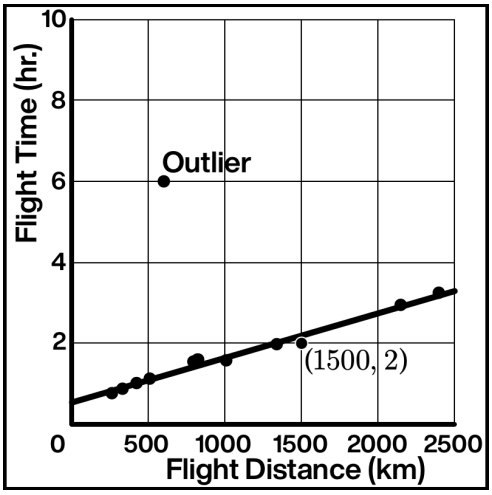
Analyzing Bivariate Data

Learning Goal(s):

- **I can create a scatter plot and draw a line to fit the data, and identify outliers that appear in the data.**
- **I can use associations between two variables to make predictions.**

People often collect data to investigate possible associations between two numerical variables and use the connections that they find to predict more values of the variables.

The scatter plot shows flight distances and times for a set of flights.



Sketch a line on the scatter plot that fits the data well.

Add a point to the scatter plot that shows a $1,500$-kilometer flight with a flight time of 2 hours.

Add an outlier to the scatter plot.

Explain why this point is an outlier.

Responses vary. **This point is an outlier because it is very far from the rest of the data, and it is not close to the line. This flight took much longer than the typical flight of its distance.**

Describe the association between flight distance and flight time.

Responses vary. **As flight distance increases, the flight time tends to increase.**

Use your model to predict the y-value of a point on the scatter plot with $x = 2000$. $y = 2.75$

What does this point tell you about the flight distance and flight time for the airplane?

Responses vary. **A flight that is $2,000$ kilometers will take around 2.75 hours.**

Summary Question

What are some things that are important to remember when analyzing a scatter plot?

Responses vary.

- **When analyzing a scatter plot, look for trends in the data to decide if there is an association between the two variables.**
- **If an association is linear, then there is a line that can be drawn to fit the data.**
- **Points that are far from the line and from the rest of the data are called outliers.**

desmos

Unit 8.6, Lesson 9: Notes Name _____

Two-Way Tables and Bar Graphs

Learning Goal(s):

- **I can identify and represent the same data in bar graphs and in two-way frequency tables.**

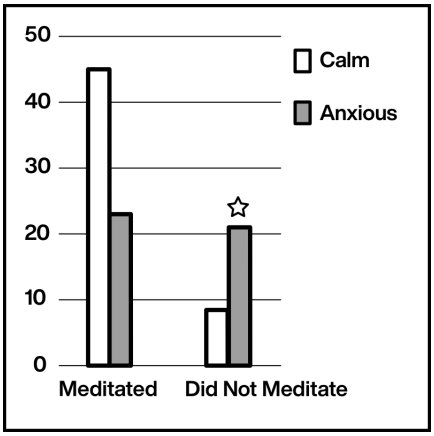
When we collect data by measuring attributes, such as height, we call that numerical data. When we collect data by counting things in various categories, such as tall or short, we call that categorical data. To help organize categorical data, we can use two-way tables and bar graphs.

These are the results of a study on meditation and athletes' state of mind before a track meet.

Two-Way Table

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

Bar Graph



Fill in the missing values in the table.

Add a bar above to represent the number of people who did not meditate and were calm.

Add a star where 21 appears in the bar graph. What does 21 mean in this scenario?

Responses vary. **There were 21 athletes in the study who were anxious and did not meditate.**

Circle 44 in the two-way table. What does 44 mean in this scenario?

Responses vary. **In total, there were 44 athletes in the study who were anxious.**

Summary Question

What are some advantages to displaying information in . . .

. . . a two-way table?

Responses vary. **You can see the precise numbers in each category.**

. . . a bar graph?

Responses vary. **You can compare data in different categories by looking at the heights of the bars.**

desmos

Unit 8.6, Lesson 10: Notes Name _____

Using Data Displays to Find Associations

Learning Goal(s):

- **I can use relative frequencies in tables and in segmented bar graphs to decide if there is an association between two variables.**

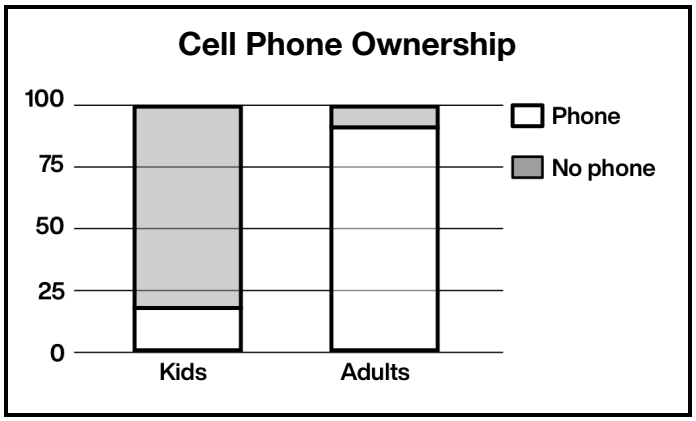
Flu Treatment Results

	Treatment A	Treatment B
Improved Health	57.5%	41.7%
No Improvement	42.5%	58.3%
Total	100%	100%

57.5% of people who took Treatment A had improved health, whereas 41.7% of those who took Treatment B had improved health.

This means there **is** / is not an association between treatment and improved health.

For each situation, decide if there is an association. Explain your thinking.



(Circle one) There **is** / is not an association between age and cell phone ownership.

Explain your thinking: ***Responses vary.***
The percentage of adults who use a cell phone (about 90%) is much larger than the percentage of kids who use one (about 20%).

Lucky Socks and Winning

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

(Circle one) There is / **is not** an association between wearing lucky socks and winning.

Explain your thinking: ***Responses vary.***
The percentage of winners wearing lucky socks is only 1% greater than those wearing regular socks, so there is not an association.

Summary Question

How can you tell when there is a possible association between variables?

Responses vary. **There is a possible association when there is big difference between how the total (100%) is split between different groups.**

desmos

Unit 8.6, Lesson 11: Notes Name _____

Creating Data Representations

Learning Goal(s):

- **I can make relative frequency tables and segmented bar graphs from frequency tables.**
- **I can use a representation of data to decide if there is an association between two variables.**

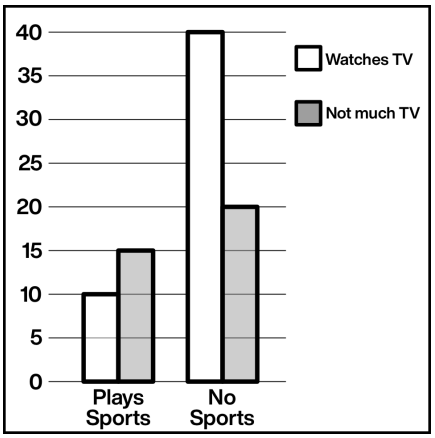
These data displays show the results of a survey of sports playing and TV watching of a group of students.

Fill in the missing information so that all of the data displays represent the same information.

Two-Way Table

	Watches TV	Not Much TV
Plays Sports	10	15
No Sports	40	20

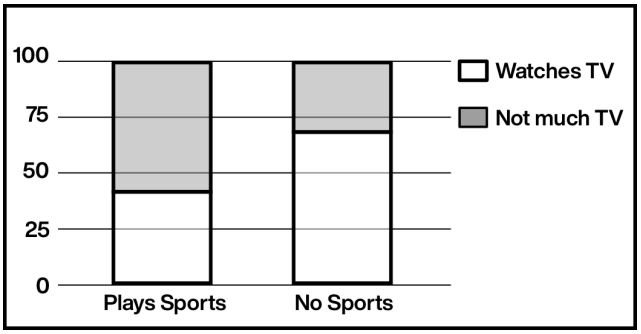
Bar Graph



Relative Frequency Table

	Watches TV	Not Much TV	Total
Plays Sports	40%	60%	100%
No Sports	66.7%	33.3%	100%

Segmented Bar Graph



Is there an association between playing sports and watching TV? Explain your thinking.

***Responses vary. Yes.* The segmented bar graph helps me see that students who play a sport are less likely to watch TV (only 40%), and students who don't play a sport are more likely to watch TV (66.7%).**

Summary Question

What are some things to remember when making relative frequency tables or segmented bar charts?

***Responses vary.* Take your percentages based on the totals that you are looking at and make sure they add up to 100%. The axis labels in the segmented bar graph are the same as the totals in the relative frequency table.**