First published in 2013 by the University of Utah in association with the Utah State Office of Education.

Table of Contents

Chapter 6: Statistics-Investigate Patterns of Association in Bivariate Data (2 weeks)

Utah Core Standard(s):

- Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association. (8.SP.1)
- Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line. (8.SP.2)
- Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. *For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height*. (8.SP.3)
- Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects. Use relative frequencies calculated for rows or columns to describe possible association between the two variables. *For example, collect data from students in your class on whether or not they have a curfew on school nights and whether or not they have assigned chores at home. Is there evidence that those who have a curfew also tend to have chores? (8.SP.4)*

Academic Vocabulary: Experiment, outcomes, sample space, random variables, realizations, quantitative (numerical) variables, categorical variables, univariate data, bivariate data, scatter plot, association, positive association, negative association, no apparent association, linear association, non-linear association, weak association, strong association, perfect association, cluster, outlier, line of best fit, linear model, prediction function, two-way frequency table, marginal frequencies, relative frequencies.

Chapter Overview:

Up to this point, students have been studying data that falls on a straight line. Most of the time data given in the real world is not perfect; however, often the data is associated with patterns that can be described mathematically. In this chapter, students will investigate patterns of association in bivariate data by constructing and interpreting scatters plots, fitting a linear function to scatter plots that suggest a linear association, and using the prediction function to solve real world problems and make predictions. In addition they explore categorical bivariate data by constructing and interpreting two-way frequency tables.

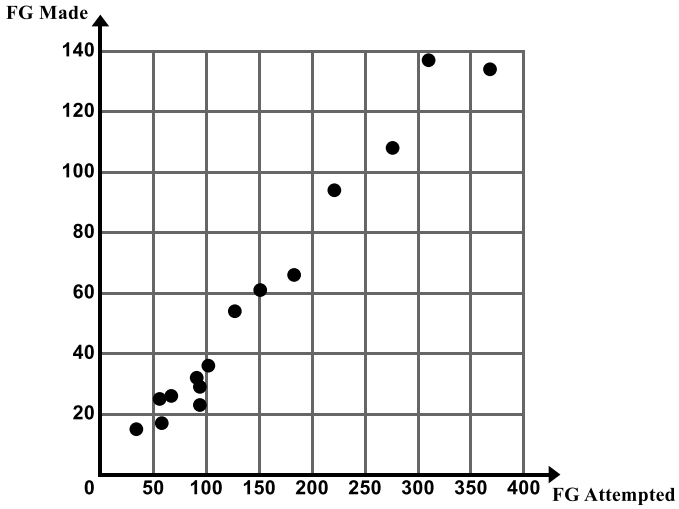
Connections to Content:

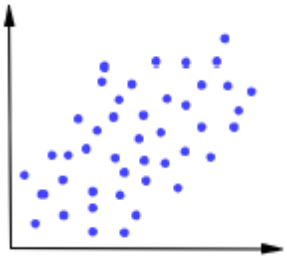
Prior Knowledge: Until 8^{th} grade, the study of statistics has centered on univariate data. Students have created and analyzed univariate data displays, describing features of the data and calculating numerical measures of center and spread. In 8^{th} grade, students have the opportunity to apply what they have learned about the coordinate plane and linear functions in order to analyze and interpret bivariate data and construct linear models for data sets that suggest a linear association.

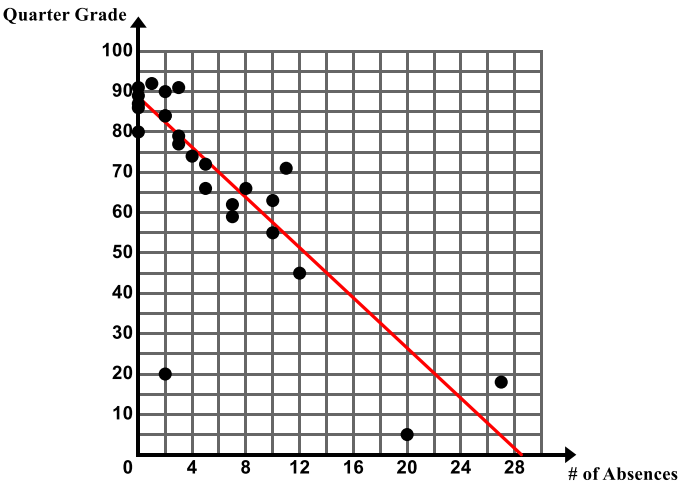
Future Knowledge: Students will more formally fit a linear, as well as additional types of functions, to bivariate data using technology. They will also calculate correlation coefficients, a numerical measure for determining the strength of a linear association. Students will also use residual plots as a tool for assessing the fit of a linear model. Students will also continue with the study of two-way frequency tables.

MATHEMATICAL PRACTICE STANDARDS:

	**Make sense of problems and persevere in solving them.**	Emina loves to eat tomatoes from her garden in Salt Lake City. She asked her friend Renzo, "Don't you just love tomatoes?" Renzo crinkled his nose and replied, "Ew, tomatoes gross me out! When I see them in the grocery store, I just keep on walking." Renzo's response prompted Emina to think, "I don't buy tomatoes at the grocery store either, because I grow them in my garden. The tomatoes from my garden are delicious, whereas grocery store tomatoes look less appealing to me. I wonder if there is an association between enjoying tomatoes and having a garden at home?" *In the problem above the student must help Emina determine if there is an association between liking tomatoes and having a garden at home. They organize collected data into a two-way frequency table and then analyze it. Students must problem solve as they decide how to organize their data and as they determine what the data is telling them.*
	**Reason abstractly and quantitatively.**	The table gives data relating the number of oil changes every two years to the cost of car repairs. *Table not shown due to space.* Plot the data on the graph provided, with the number of oil changes on the horizontal axis. You will need to define your own scale. Write a prediction function in slope-intercept form that you could use to predict the cost of repairs, y, for any number of oil changes, x. Compare your prediction with that of a partner. Use your prediction function to predict how much a person would spend on car repairs if they were to get 8 oil changes. Compare your prediction with that of a partner *Throughout the chapter, students analyze displays of numeric data sets (in tables and in graphs). If the data sets suggest a linear association, students construct a linear function to model the situation. These functions are an abstract way to represent the associations suggested by the data sets.*

	Construct viable arguments and critique the reasoning of others.	Using the scatter plot, determine if there is a relationship between field goals attempted and field goals made. Describe any trends or patterns you observe in the data.  *Throughout the chapter, students are asked to create a scatter plot of a given data set and analyze the scatter plot to determine if there is an association between two variables. They look for trends and patterns, including clusters and outliers. They provide explanations related to the context for the associations, trends, and patterns. Students are making **arguments** about the data and are asked to **support** their arguments with data and critical thinking about the context and limitations of the data.*
	Model with mathematics.	Students will say a selected tongue twister one at a time. In the first trial, only the first student will say the tongue twister; in the second trial, only the first and second students will say the tongue twister, etc. In each trial, one person will be added to the chain of tongue twisters and the total elapsed time will be recorded. Tongue twisters: A. Work will win when wishy-washy wishing won't. B. Three witches wished three wishes, but which witch wished which wish. C. Peter Piper picked a peck of pickled peppers. D. Picky people pick Peter Pan peanut butter it is the only peanut butter picky people pick. *Throughout the chapter students will fit a linear model to several real-life situations that suggest a linear association. Students will construct prediction functions for lines of best fit and use the functions to make predictions and solve real-world problems.*
	Use appropriate tools strategically.	*Online software and graphing calculators are important tools that can be used to display and analyze large data sets and construct functions to model data sets. Additionally, many of the skills that students have learned up to this point will become a tool they will rely on in order to construct linear functions for data sets that suggest a linear association.*

	Attend to precision.	The following table shows the weight of an English Mastiff from birth to age 60 weeks. *Table not shown due to space.* Create a scatter plot of the data on the grid below. Describe any patterns of association you see in this scatter plot. Use the context to give possible explanations as to why these trends, patterns, and associations exist. *When students create scatter plots in this chapter, they must determine how to scale each axis appropriately and ensure that they are graphing the data points accurately in order to determine whether an association exists between the two variables and in order to write a function that models the data.*
	Look for and make use of structure.	Describe the association between x and y. Circle any clusters in the data. Put a star by any points that appear to be outliers.  *In order to describe the association between x and y, students must examine the structure of the data points on the graph. If there is an association, students must determine the following: Is it linear or non-linear? Is the association positive or negative? Is the association weak or strong? Do there appear to be any outliers or clusters?*

	**Look for and express regularity in repeated reasoning.**	The following scatter plot shows the final grade in Ms. Ganchero's math class for students and the number of times they are absent. Explain the meaning of the slope and *y*-intercept in the context.  *Throughout the chapter, students must determine whether the relationship between two quantities suggests a linear association. In the case of a linear association, slope is a calculation that is repeated – linear functions grow at a constant rate. For data that resembles a line, students will write a prediction function for a line of best fit drawn through the data and explain the meaning of the slope in the context.*

6.0 Anchor Problem: Tongue Twisters

Students will say a selected tongue twister one at a time. In the first trial, only the first student will say the tongue twister; in the second trial, only the first and second students will say the tongue twister, etc. In each trial, one person will be added to the chain of tongue twisters and the total elapsed time will be recorded.

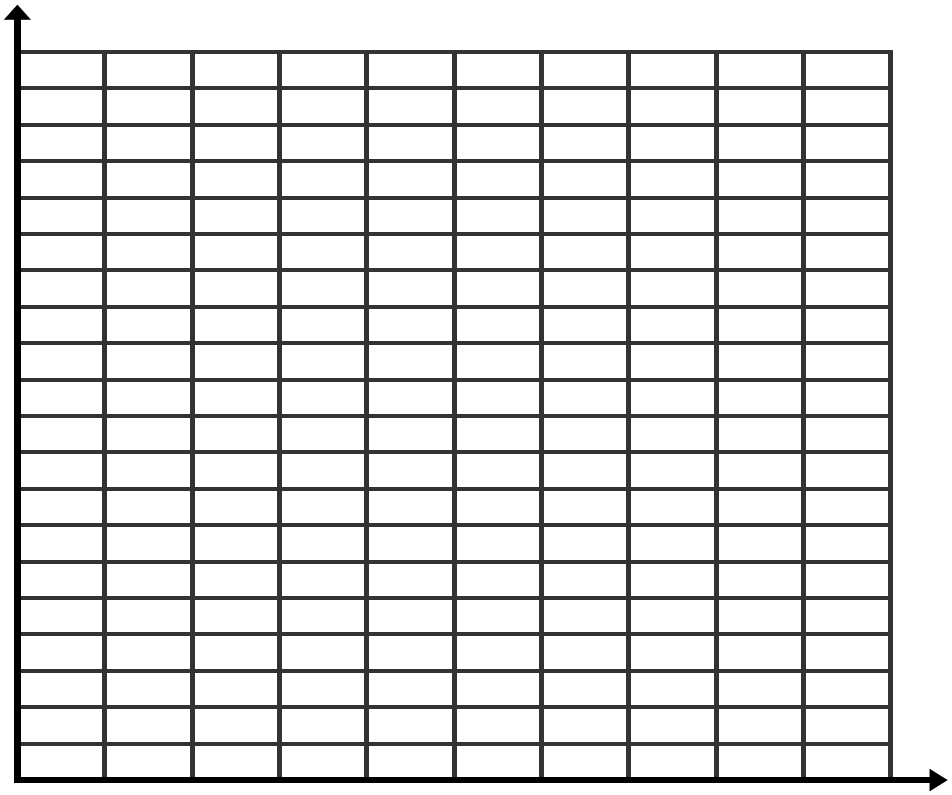
Tongue twisters:

A. Work will win when wishy-washy wishing won't.
B. Three witches wished three wishes, but which witch wished which wish.
C. Peter Piper picked a peck of pickled peppers.
D. Picky people pick Peter Pan peanut butter it is the only peanut butter picky people pick.

1. In the table below, record the class data for each Tongue Twister.

Number of people	Tongue Twister A (time)	Tongue Twister B (time)	Tongue Twister C (time)	Tongue Twister D (time)
1				
2				
3				
4				
5				
6				
7				
8				
9				

2. Make a scatter plot using different colors for each tongue twister's data. Make sure you label and title the graph.



3. Observe the different data sets. What observations can you make about the data sets?

4. Choose a tongue twister. How long would it take 25 people to say each tongue twister? Explain how you determined your answer. Using the same tongue twister, determine how many people can say the tongue twister in 2 minutes.

Section 6.1: Construct and Interpret Scatter Plots for Bivariate Data

Section Overview: In this section we continue our study of bivariate data, specifically quantitative or numerical data. In 7th grade students engaged in the study of univariate data. We begin this section with a problem that deals with univariate data and then use the same context to explore a bivariate data set. As in the case of univariate data, analysis of bivariate measurement data graphed on a scatterplot proceeds by describing shape, center, and spread. Later, we are introduced to Izumi and her basketball statistics and use her data throughout the chapter to build upon the concepts of analyzing bivariate data. In this section students learn how to construct, read, and interpret a scatter plot. Throughout the section students investigate and describe trends and patterns of association between two variables and interpret these associations in a variety of real-world situations.

Concepts and Skills to be Mastered:
By the end of this section students should be able to:

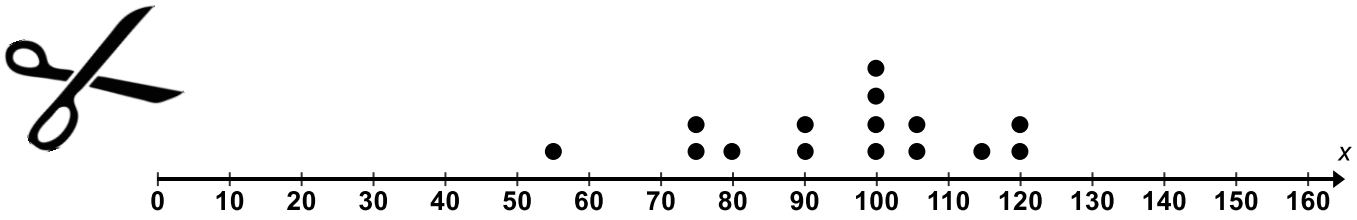
1. Read and interpret a scatter plot.
2. Construct a scatter plot for bivariate data.
3. Describe patterns of association in a scatter plot.

6.1a Class Activity: Read and Interpret a Scatter Plot

1. Jenny is a hair stylist. She decides to record the amount of money she makes in tips over a 15-day period. She records the following data:

Day	Amount of Money Made in Tips (dollars)
1	120
2	75
3	80
4	100
5	115
6	100
7	55
8	90
9	100
10	120
11	90
12	105
13	105
14	75
15	100

To better visualize the data, Jenny makes a dot plot of the data.

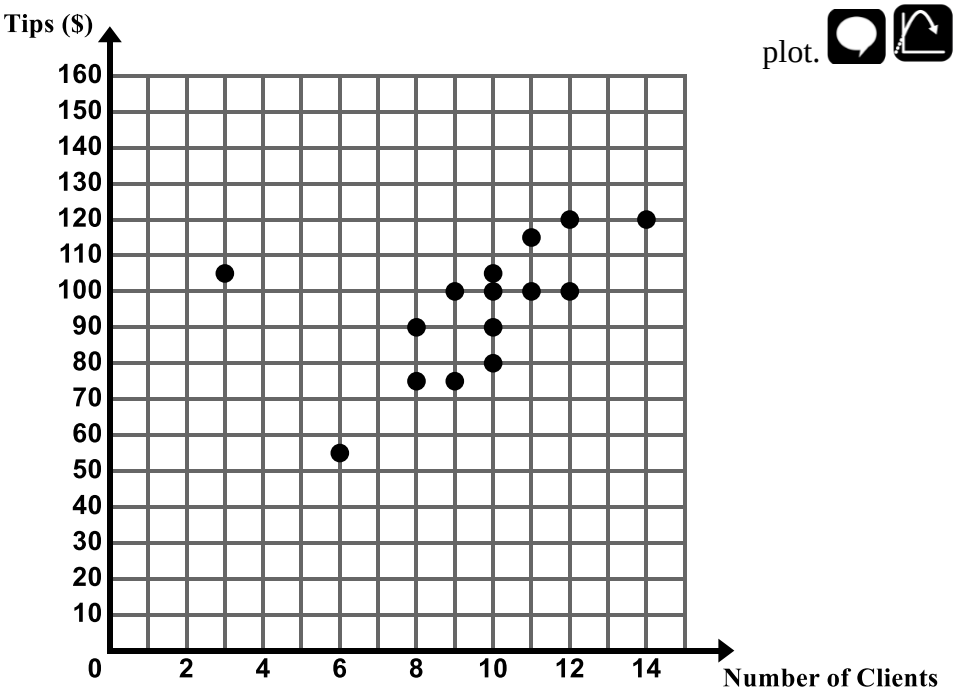


a. Make some observations about the data shown in the dot plot.

2. Jenny then asks herself the following question: "I wonder if the amount I make in tips is associated to the number of clients I have each day?" She looks back through her appointment book and records the number of clients she had on each of the 15 days. She records the following data.

Day	Number of Clients	Amount of Money Made in Tips (dollars)
1	12	120
2	8	75
3	10	80
4	12	100
5	11	115
6	9	100
7	6	55
8	8	90
9	10	100
10	14	120
11	10	90
12	10	105
13	3	105
14	9	75
15	11	100

To better visualize the data, Jenny makes a **scatter plot** of the data. A **scatter plot** is a graph in the coordinate plane of the set of all (x, y) ordered pairs of bivariate data.



a. Make some observations about the scatter plot.

Directions: Determine if the following scenarios represent univariate or bivariate data.

3. Lucas conducts an experiment where he records the number of speeding tickets issued in Iron County in a given year along with the average price of gasoline for that same given year. He collects this data from the year 1972 through 2012.

4. Lea conducts an experiment where she records the heights of all the NBA basketball players on the Miami Heat's roster for the 2014 season.

5. Adel conducts an experiment where she records the selling price of several homes in a neighborhood.

6. Adel conducts an experiment where she records the selling price and square footage of homes in a neighborhood.

7. Lisa conducts an experiment on the number of times a person works out a week and the person's weight.

In this chapter, we will focus our study on bivariate data sets and we will explore the relationship between two variables of interest.

Izumi is the score keeper for her school's basketball team. Izumi's responsibilities as score keeper are to keep a record for several plays during the 2012-2013 season. The basketball plays are listed below.

- Total number of field goals made.

  In basketball a field goal is the result of the player successfully shooting the basketball through the hoop, regardless of whether it is a two point shot or a three point shot. This does not include foul shots.
- The total number of field goals attempted.

  A field goals attempt results when a player tries to make a field goal, an attempt is made whether or not the ball goes through the hoop.
- The total number of assists.

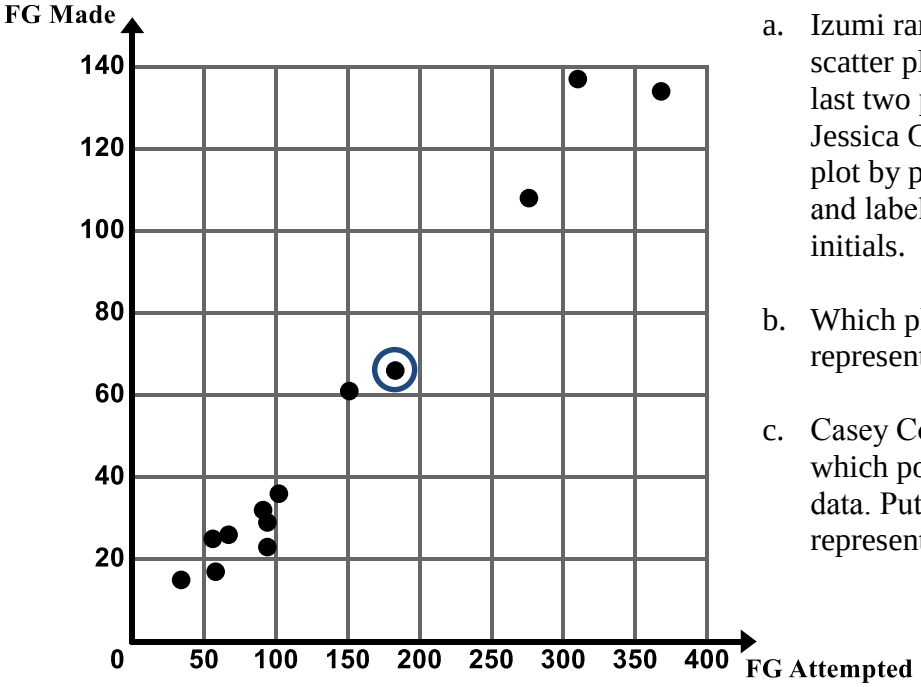
  An assist results when the player passes the ball to a teammate who then scores.
- The total number of rebounds

  A rebound results when the player retrieves the ball from an unsuccessful field goal attempt.

The table given below shows the record that Izumi made regarding the number of field goals attempted and the number of field goals made.

Player	Field Goals Attempted	Field Goals Made
Amber Carlson	34	15
Casey Corbin	368	134
Joan O'Connell	94	23
Monique Ortiz	102	36
Maria Ferney	91	32
Amelia Krebs	310	137
Tonya Smith	56	25
Juanita Martinez	58	17
Sara Garcia	151	61
Alicia Mortenson	67	26
Parker Christiansen	94	29
Rachel Reagan	183	66
Paula Lyons	276	108
Thao Ho	221	94
Jessica Geffen	127	54

8. As Izumi examines the data she wonders, "Is there is an association between the number of field goals made and the number of field goals attempted?" To further investigate the relationship between these two random variables, "Field Goals Made" and "Field Goals Attempted" Izumi makes a **scatter plot** of the data as shown below.



a. Izumi ran out of time while creating her scatter plot and did not plot the data for the last two players in the table, Thao Ho and Jessica Geffen. Help Izumi finish the scatter plot by plotting the data for these players and labeling the points with these players' initials.

b. Which player does the circled data point represent?

c. Casey Corbin sees Izumi's graph and asks which point on the scatter plot represents her data. Put Casey's initials by the point that represents his data.

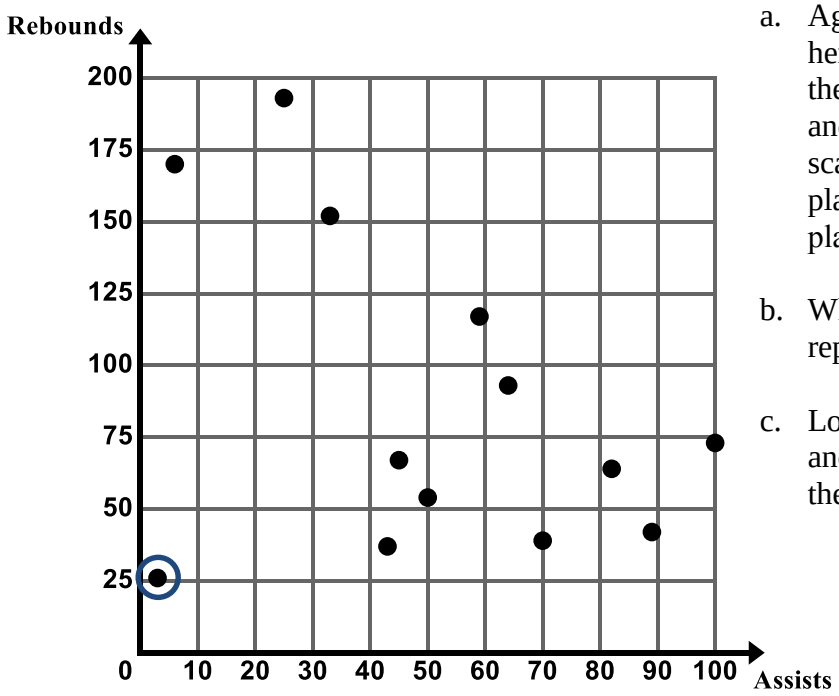
d. Using the scatter plot, determine if there is a relationship between field goals attempted and field goals made. Describe any trends or patterns you observe in the data. 

e. Can you think of another variable that when graphed with field goals made would have a positive association?

9. In addition to data about field goals, Izumi is curious about the relationship between the number of assists and the number of rebounds a player makes in a season. In order to study this relationship, Izumi gathers data on the number of assists and rebounds each player makes during the season. Izumi's Assist and Rebound data are given in the following table.

Player	Assists	Rebounds
Amber Carlson	82	64
Casey Corbin	6	170
Joan O'Connell	43	37
Monique Ortiz	50	54
Maria Ferney	89	42
Amelia Krebs	25	193
Tonya Smith	70	39
Juanita Martinez	3	26
Sara Garcia	100	73
Alicia Mortenson	33	152
Parker Christiansen	64	93
Rachel Reagan	45	67
Paula Lyons	59	117
Thao Ho	15	179
Jessica Geffen	30	113

Izumi made the scatter plot of assists and rebounds shown below to help her better visualize the data.



a. Again, Izumi ran out of time while creating her scatter plot and did not plot the data for the last two players in the table, Thao Ho and Jessica Geffen. Help Izumi finish the scatter plot by plotting the data for these players and labeling the points with these players' initials.

b. Which player does the circled data point represent?

c. Locate the data points for 3 different players and put the initials of the players next to their data point.

d. Izumi notices the circled data point stands out noticeably from the general behavior of the data set. We call this point an **outlier**. Provide an explanation as to why this player's data does not fit with the rest of the data.

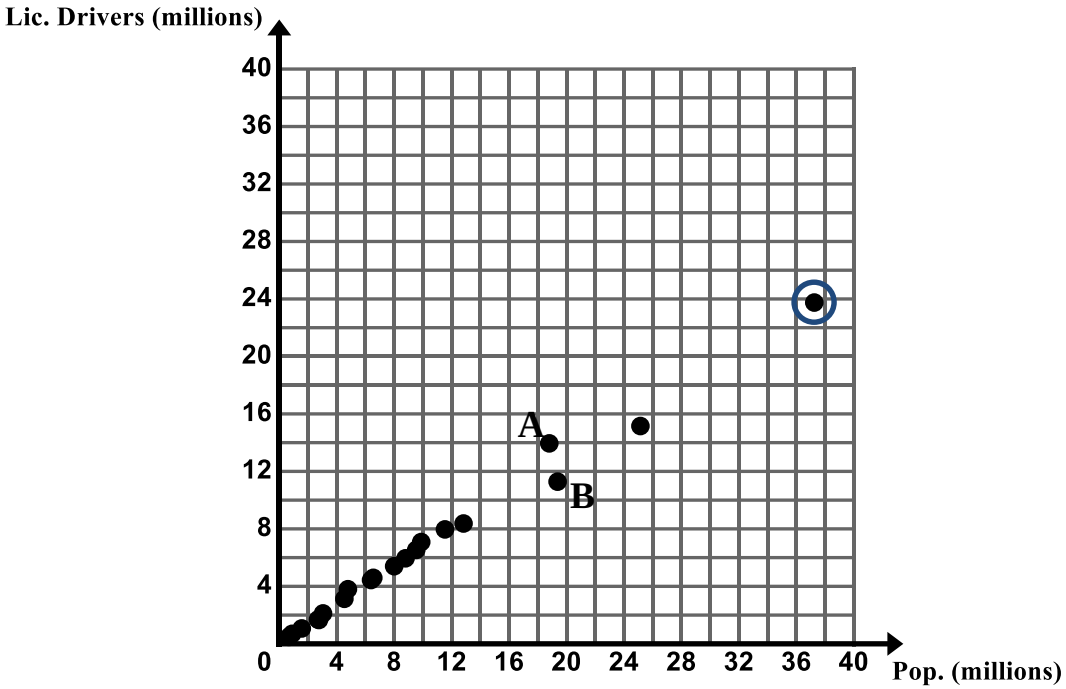
e. Using the scatter plot, determine if there is a relationship between number of assists and number of rebounds. Describe any trends or patterns you observe in the data.

f. Can you think of another variable that when graphed with field goals made would have a negative association?

10. Which data set appears to have a stronger association: the relationship between number of field goal made and number of field goal attempts or the relationship between number of rebounds and number of assists?

6.1a Homework: Read and Interpret a Scatter Plot

1. The U.S. Census Bureau collects data about the people and economy in the United States. The graph below shows the population (in millions) and the number of licensed drivers (in millions) for 20 different states for the year 2010.



a. What does the circled data point (37.25, 23.75) represent in the context?

b. In 2010, Texas had a population of approximately 25.15 million people and had approximately 15.2 million licensed drivers. Put a star by the data point that represents Texas.

c. What does the graph show about the relationship between a state's population and the number of licensed drivers in the state?

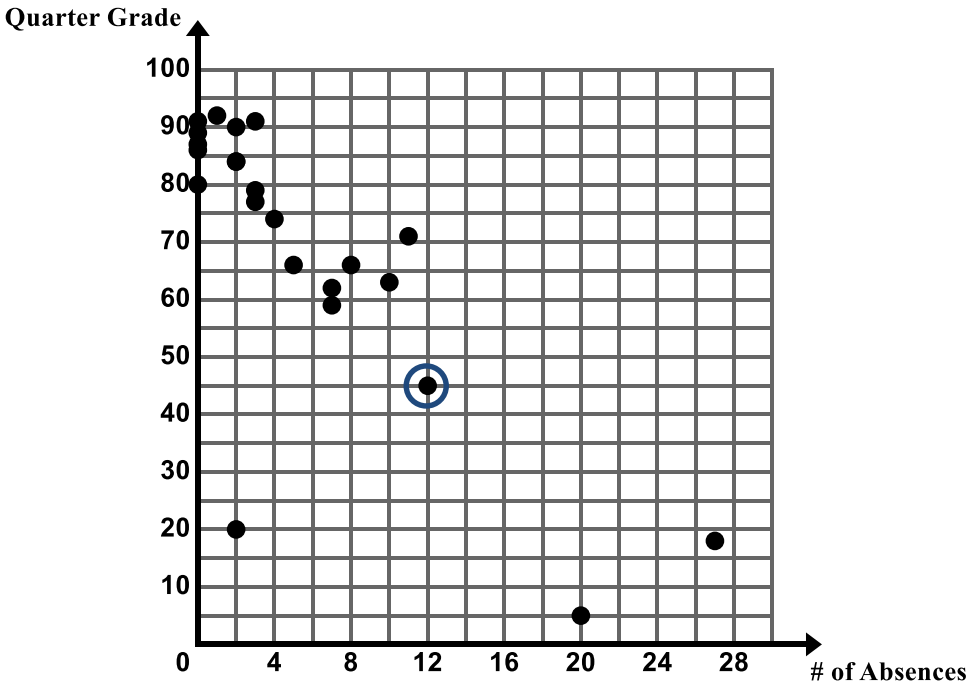
d. If a state has a population of approximately 32 million people, approximately how many licensed drivers would you expect to find in the state based on the trend in the scatter plot?

e. If a state has approximately 12 million licensed drivers in a state, what would you expect the population to be in that state based on the trend in the scatter plot?

f. Compare data points A and B.

g. Data point A represents the state of Florida and data point B represents the state of New York. Provide an explanation as to why New York has more total people than Florida but fewer licensed drivers.

2. Ms. Ganchero is a math teacher. She wonders if there is an association between the number of absences a student has in her class and the grade they earn at the end of the quarter. In order to analyze this relationship, Ms. Ganchero created the scatter plot below which shows the number of absences a student has in a quarter and their final grade at the end of the quarter.



a. While reviewing the scatter plot, Ms. Ganchero realized that she did not plot the data for two students. Rachel was absent 5 times and received a final grade of 72 and Lydia was absent 10 times and received a final grade of 55. Plot and label these two data points on the scatter plot above.

b. What does the circled data point represent in the context?

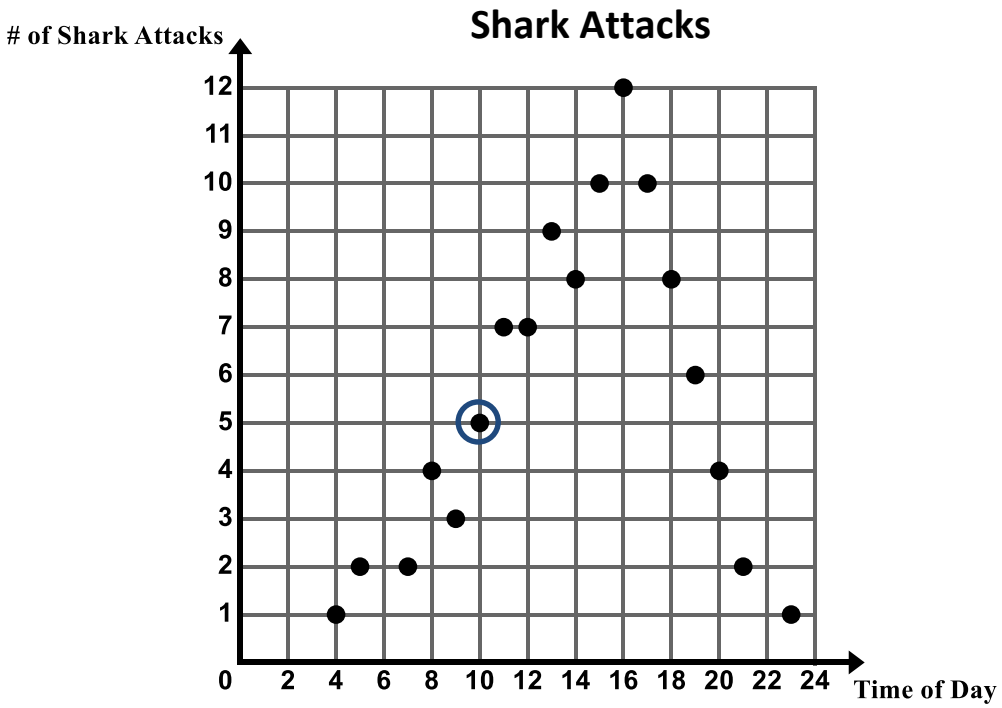
c. Provide an explanation for the cluster of points in the upper left corner of the graph.

d. Do there appear to be any outliers in the data? If yes, what are they? Provide an explanation for the outlier(s).

e. Does the scatter plot suggest a relationship between absences and grade? Describe any trends or patterns you observe in the data.

3. A long stretch of a popular beach is overseen by the local coast guard. Over a period of 60 years the coast guard has kept track of the number of shark attacks occurring along the coast as well as the hour during the day in which the attack occurred. The table and corresponding scatter plot show this data. *__Note:__ The time of day is given by a 24 hour clock, also known as military time.

Hour during the day	Number of Shark Attacks
04:00	1
05:00	2
07:00	2
08:00	4
09:00	3
10:00	5
11:00	7
12:00	7
13:00	9
14:00	8
15:00	10
16:00	12
17:00	10
18:00	8
19:00	6
20:00	4
21:00	2
23:00	1



a. What does the circled data point represent in the context?

b. Describe the association that exists between the time of day and the number of shark attacks. Give a possible explanation as to why this graph is shaped the way it is.

For tomorrow's class, you will need data on the height and shoe size of 5 people. Be sure to gather this data from different aged people – younger siblings, older siblings, parents, grandparents. Record your data here for tomorrow's class.

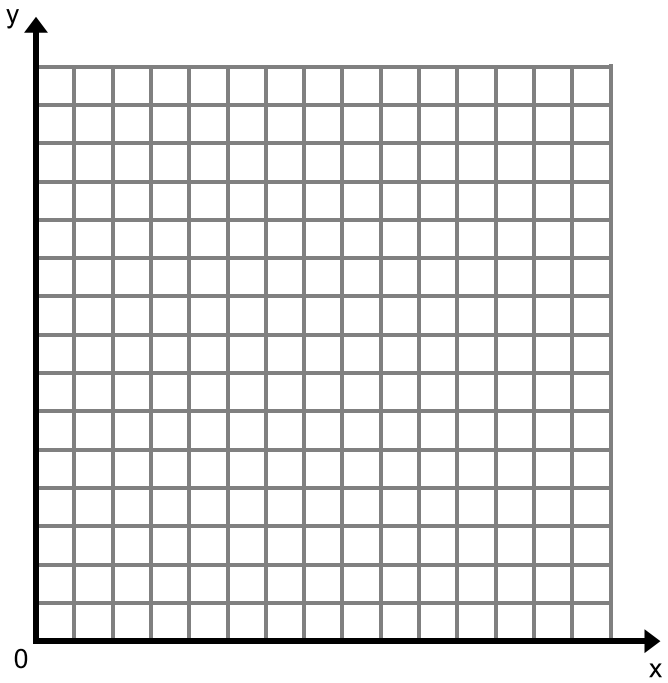
6.1b Class Activity: Create and Analyze a Scatter Plot

1. Do you anticipate an association between a person's height and their shoe length?
   a. Make a prediction.

   b. Collect your class data in the table below.

	Height	Shoe Length
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		
11.		
12.		
13.		
14.		
15.		

   c. Make a scatter plot of the data.



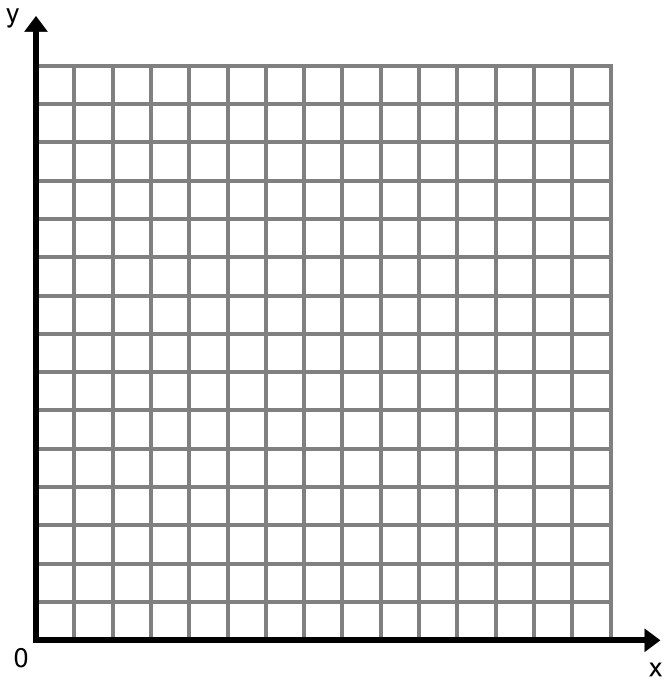
   d. Using the scatter plot, determine if there is an association between a person's shoe length and height. Describe any trends or patterns you observe in the data including clusters and outliers.

2. Is there an association between the number of letters in a person's first name and the number of letters in a person's last name?
    a. Make a prediction.

    b. Collect your class data in the table below.

Person's first and last name	Number of letters in their first name	Number of letters in their last name

    c. Make a scatter plot of the data.



    d. Using the scatter plot, determine if there is an association between the number of letters in a person's first name and the number of letters in their last name. Describe any trends or patterns you observe in the data including clusters and outliers.

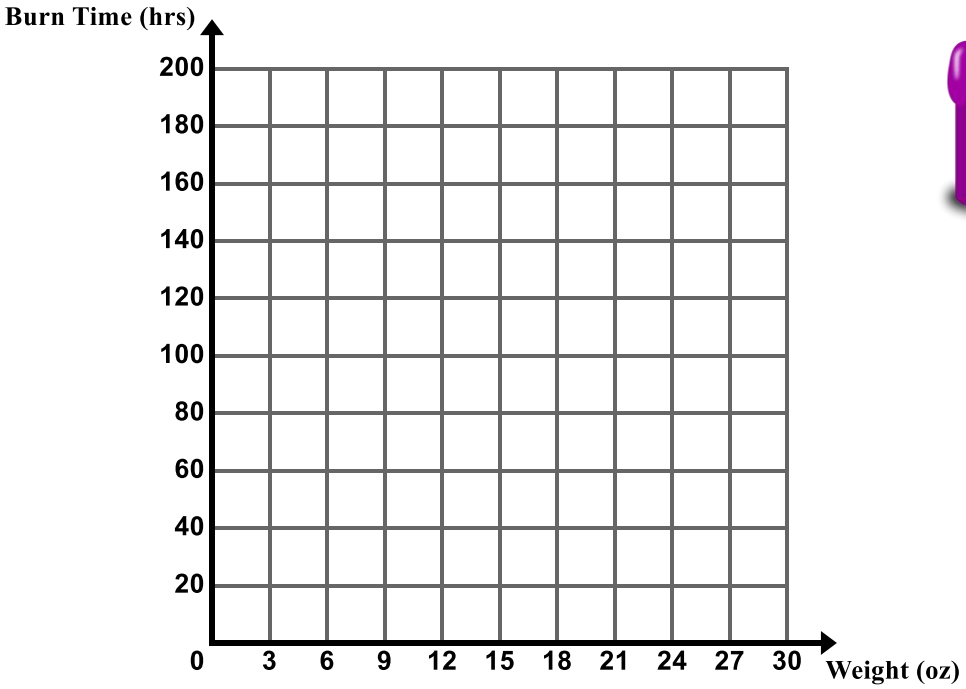
6.1b Homework: Create and Analyze a Scatter Plot

1. Is there an association between the weight of a candle and the amount of time it burns?
   a. Make a prediction.

A company that manufactures candles tests the amount of time it takes for several candles of several different weights to burn. The results are shown in the table below.

Candle Weight (ounces)	2	2	2	3	3	4	4	5	5	5	10	10	10	16	16	16	22	22	22	26	26
Burn Time (hours)	15	16	18	20	33	34	35	38	40	36	40	80	80	95	100	98	120	125	175	174	180

   b. Make a scatter plot of the data on the graph provided.



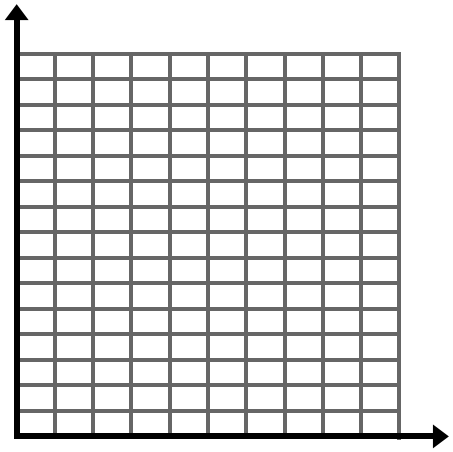
   c. Using the scatter plot, determine if there is an association between the weight of a candle and how long it burns. Describe any trends or patterns you observe in the data including clusters and outliers.

   d. **Bonus:** How much would a candle have to weigh to burn for one year?

2. Create scatter plots of the following sets of data. Think about how to scale each axis based on the data set.

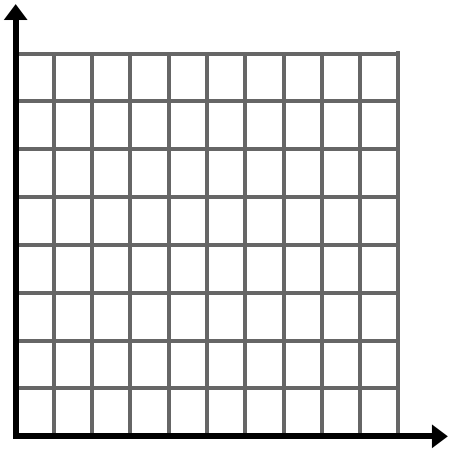
a.

x	1	2	3	4	5
y	2	5	9	12	14



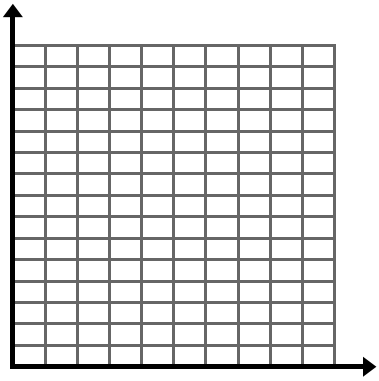
b.

x	0	2	2	4	4	5	6	6	7	8
y	5	6	5	5	7	6	4	6	5	6



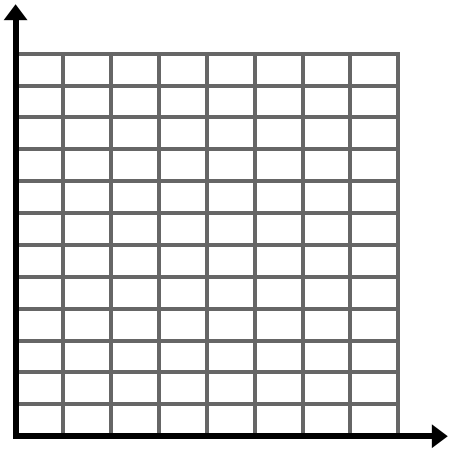
c.

x	0	1	2	3	4	5
y	1.4	1.7	2	2.2	2.4	2.8



d.

x	10	10	20	30	30	40	40	50	60	80
y	9	10	9	8	9	7.5	8	7	6	5



6.1c Classwork: Patterns of Association

So far in our study of bivariate data, we have seen data sets that show different **types of association** between two variables. There are many ways that we can describe the association (if there is one) between two variables. Common ways to talk about the association of two variables are shown in the table below. **Sketch scatter plots that correspond to each of the four associations described.**

1. **Positive Linear Association**	2. **Negative Linear Association**
3. **No Apparent Association**	4. **Nonlinear Association**

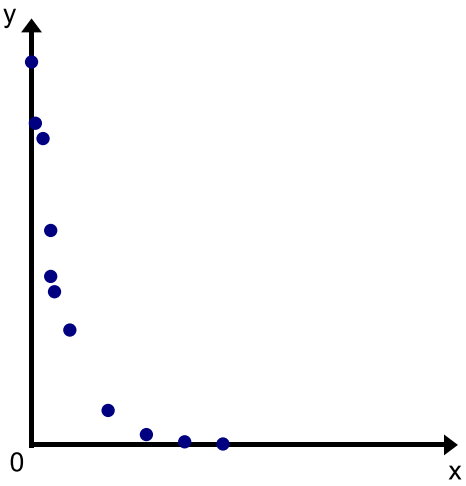
If the variables show a linear association, we can determine whether that relationship is **strong**, **weak**, or **perfect**. Imagine drawing a line through the center of the points—EYEBALLING the line. If the data points are closely packed around your line, the linear relationship is a **strong** one. If the data points are more spread out from the line, the linear relationship is a **weak** one. If your data points fall on a straight line, the linear association is **perfect**.

We may also observe the following patterns in our data:

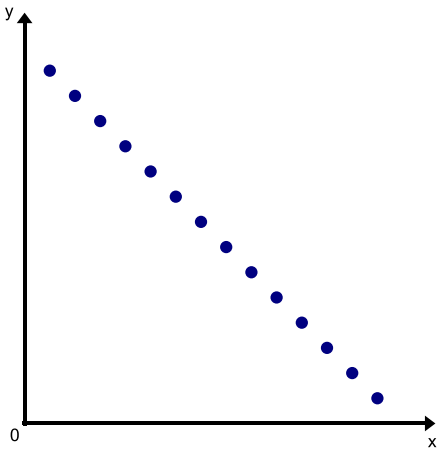
- **Clusters -** A cluster is a set of points that are in close proximity to each other.
- **Outliers -** An outlier is a data point that noticeably stands out from the general behavior of the data set.

Directions: Describe the association between *x* and *y* using the terms from the previous page. Circle any clusters in the data. Put a star by any points that appear to be outliers.

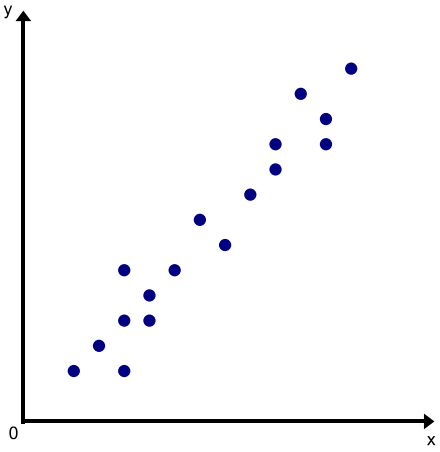
5.



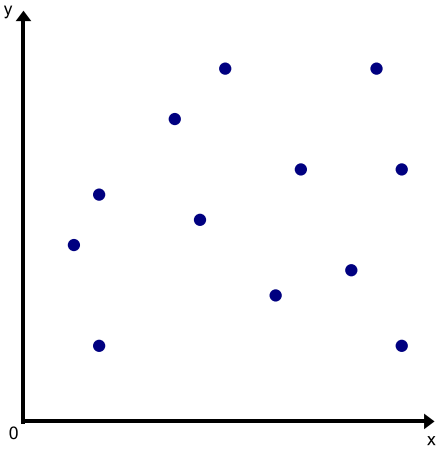
6.



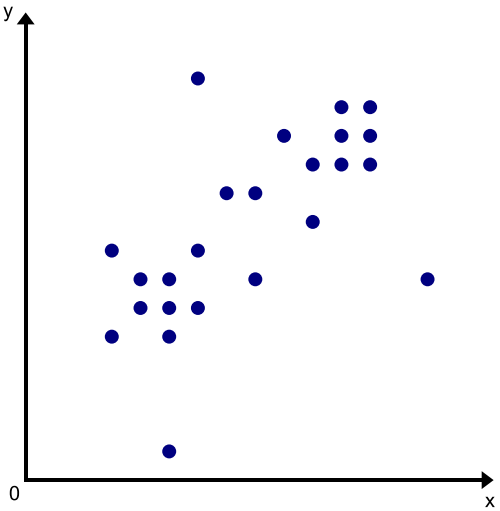
7.



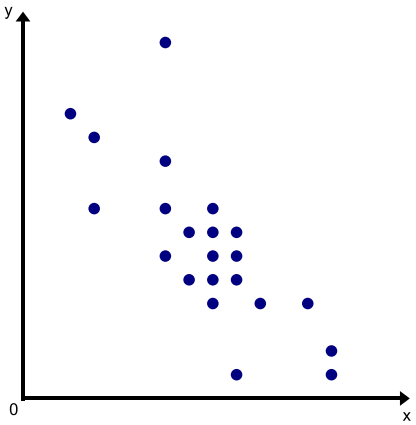
8.



9.

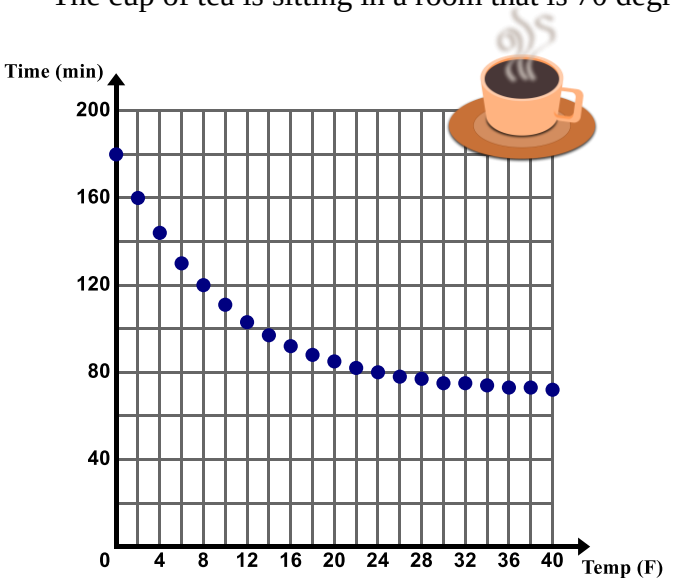


10.

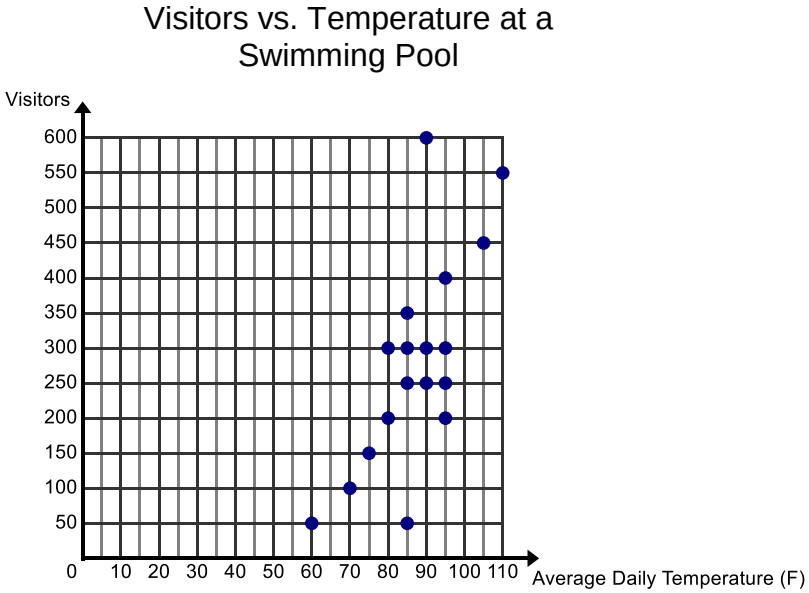


Directions: Examine the following scatter plots. Describe the association between the two variables. Circle any clusters in the data. Put a star by any points that appear to be outliers. Use the context to give possible explanations as to why these trends, patterns, and associations exist.

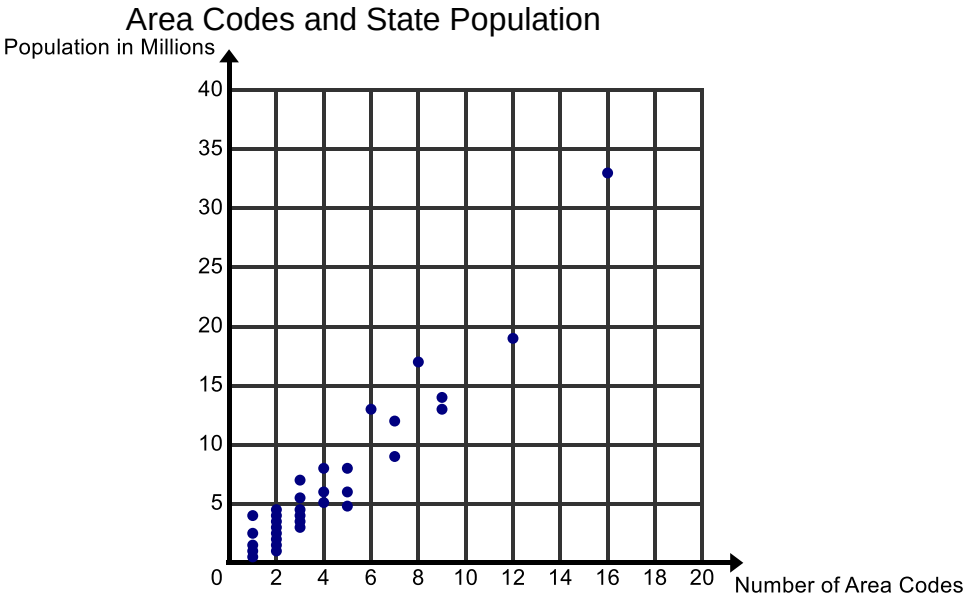
11. The scatter plot given below shows the temperature of a cup of tea sitting on the counter for 30 minutes. The cup of tea is sitting in a room that is 70 degrees.



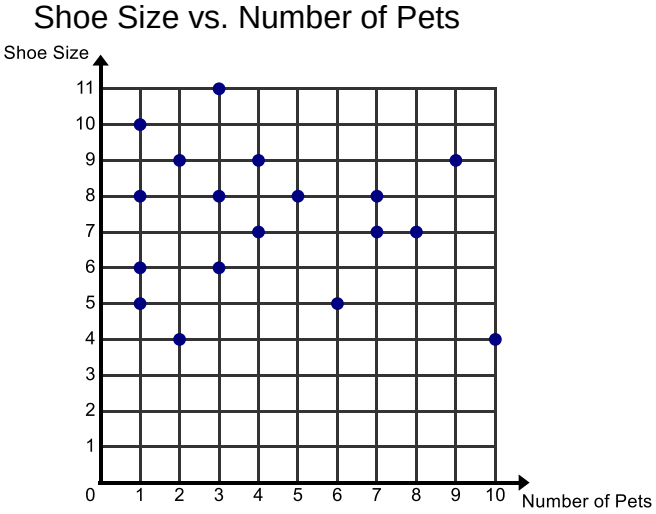
12. The Paradise Pool records the average daily temperature and the number of visitors to their pool for 18 days throughout the month of July. On July 24th, to celebrate Pioneer Day, admission is half off. The average daily temperature on that day is 90 degrees.



13. The scatter plot below show the population (in millions) and number of area codes for some states in the United States.



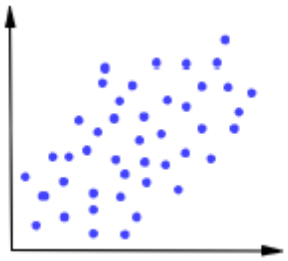
14. Holly's math teacher asks her to conduct her own survey to study different types of association. She chooses to investigate the number of pets a person has and their shoe size.



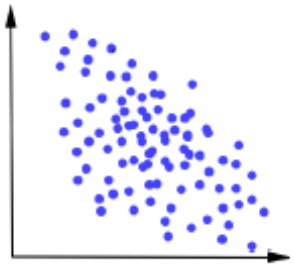
6.1c Homework: Patterns of Association

Directions: Describe the association between *x* and *y*. Circle any clusters in the data. Put a star by any points that appear to be outliers.

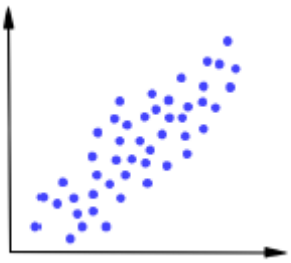
1.



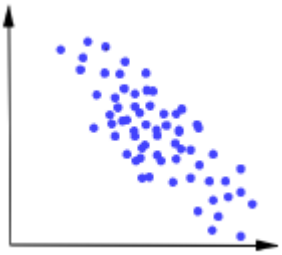
2.



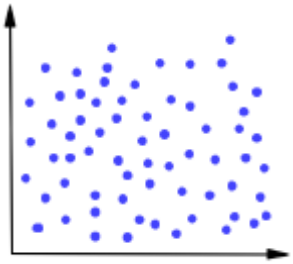
3.



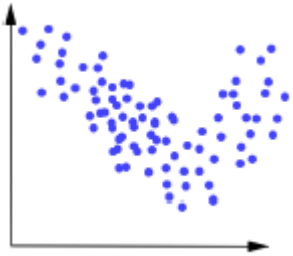
4.



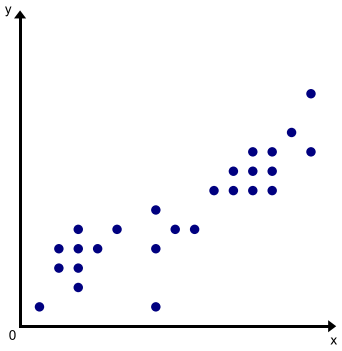
5.



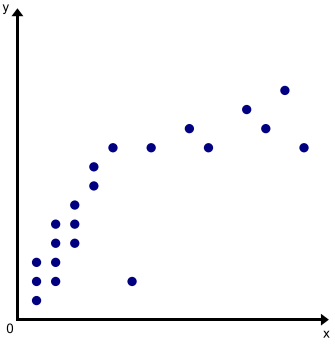
6.



7.



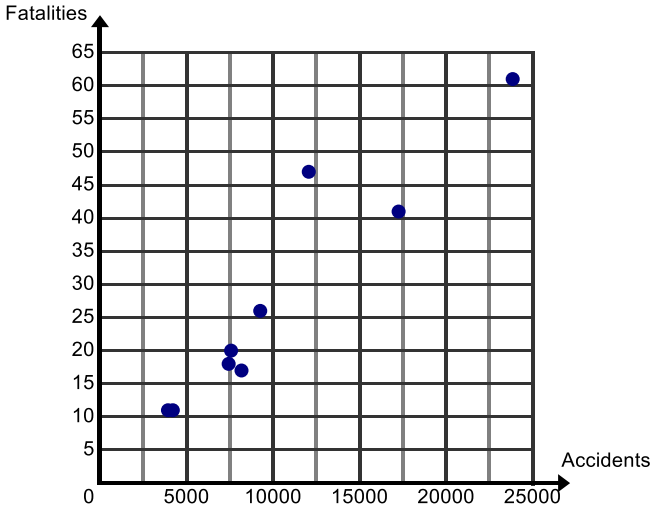
8.



Directions: Examine the following scatter plots. Describe the association between the two variables. Circle any clusters in the data. Put a star by any points that appear to be outliers. Use the context to give possible explanations as to why these trends, patterns, and associations exist.

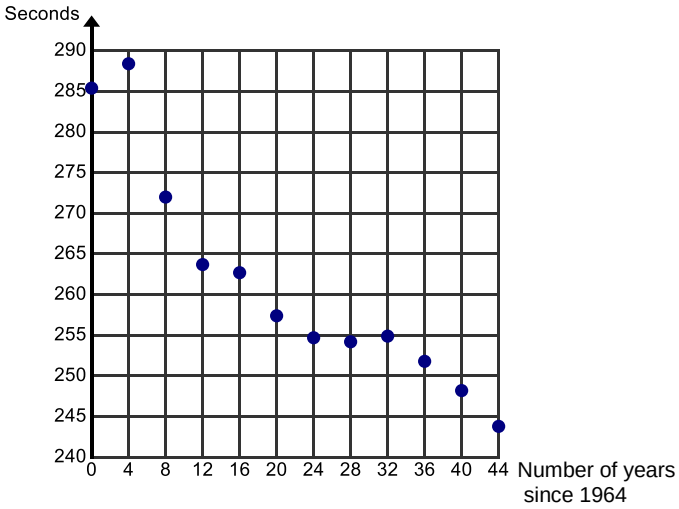
9. For Heidi's Driver's Education class, she finds data about the number of car accidents and fatalities (deaths) from car accidents for teens in the Western United States.

Fatalities vs. Accidents for Teen Drivers in 2006 in the Western United States



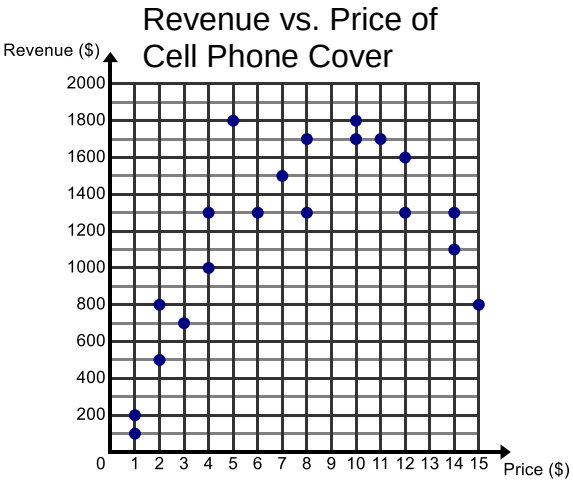
10. Winning times for the Men's Individual Swimming Medley in the Olympics from 1964-2008 are in the plot below. Michael Phelps' times are the last two entries.

400-Meter Individual Swimming Medley in Olympics (1964 – 2008)



Bonus: Research, collect, and analyze Olympic data for other events that interest you.

11. Hannah has a kiosk in the mall where she is selling Cell Phone Covers. She records how much money she makes (revenue) based on the price she charges for the covers.



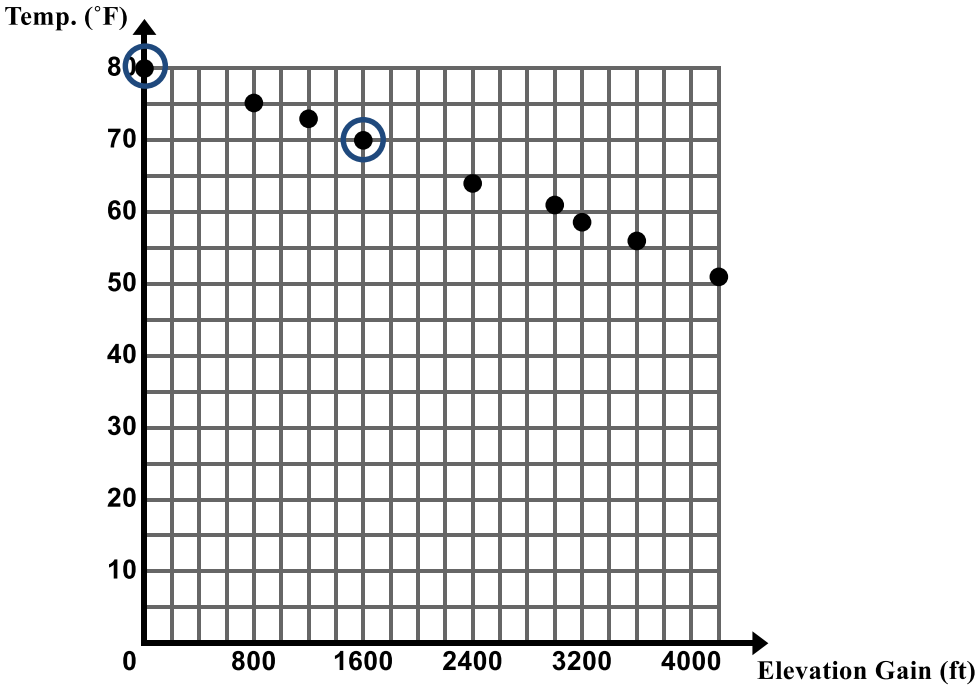
6.1d Self-Assessment: Section 6.1

Consider the following skills/concepts. Rate your comfort level with each skill/concept by checking the box that best describes your progress in mastering each skill/concept. Sample problems for each standard can be found on the following page(s).

Skill/Concept	Minimal Understanding 1	Partial Understanding 2	Sufficient Mastery 3	Substantial Mastery 4
1. Read and interpret a scatter plot.				
2. Construct a scatter plot for bivariate data.				
3. Describe patterns of association in a scatter plot.				

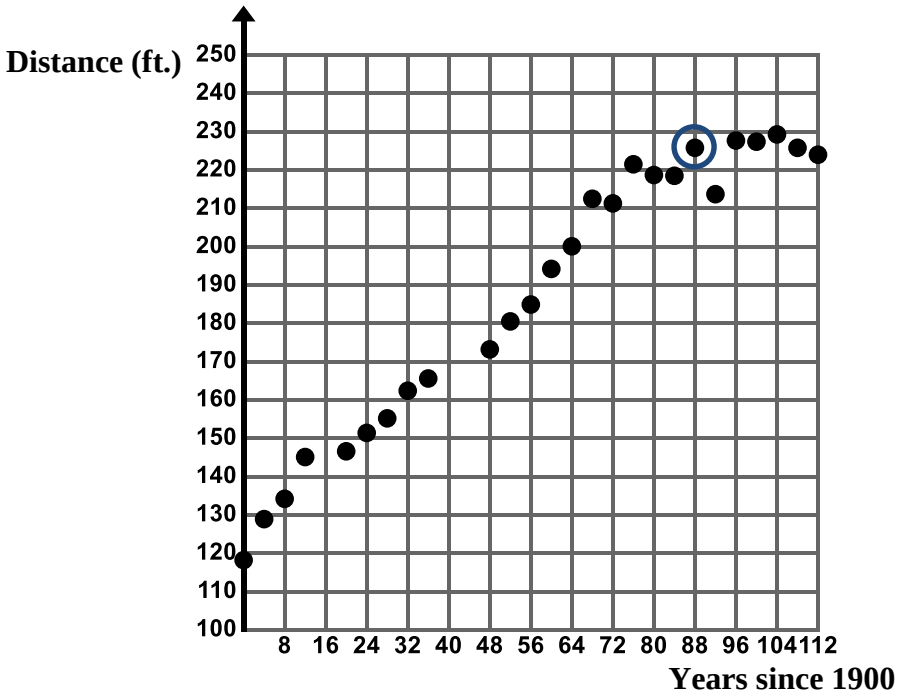
1. The following graph shows the temperature at the start of a popular hiking trail and at various points along the hike (for use with Skill/Concepts #1 and #3).
   a. What do the circled data points represent in the context?

   b. Describe any patterns of association you see in this scatter plot. Use the context to give possible explanations as to why these trends, patterns, and associations exist.



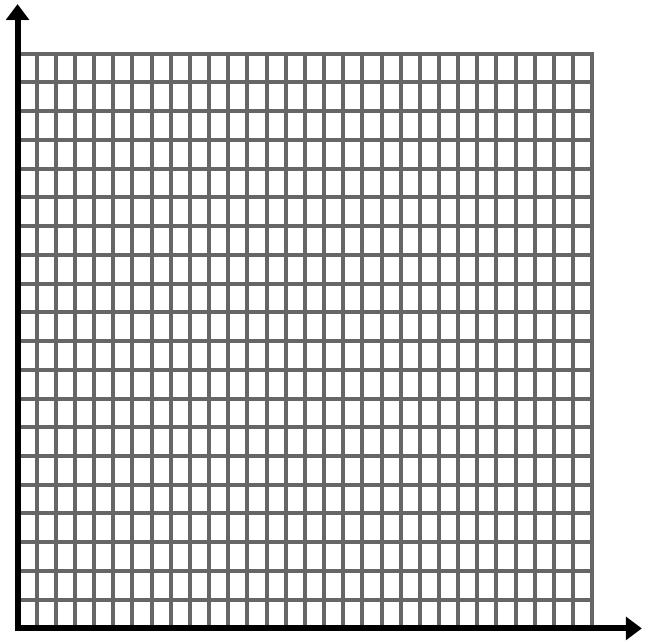
2. The following graph shows the distance, in feet, of the winning Olympic discus throws for men from 1900 to 2012 (for use with Skill/Concepts #1 and #3).
   a. What does the circled data point (88, 225.8) represent in the context?

   b. Virgilijus Alekna of Lithuania holds the Olympic record for discus in the 2004 Summer Olympics in Athens. Circle this data point on the scatter plot.
   c. Describe any patterns of association you see in this scatter plot. Use the context to give possible explanations as to why these trends, patterns, and associations exist.



3. The following table shows the weight of an English Mastiff from birth to age 60 weeks (for use with Skill/Concepts #1, 2 and #3).
    a. Create a scatter plot of the data on the grid below.
    b. Describe any patterns of association you see in this scatter plot. Use the context to give possible explanations as to why these trends, patterns, and associations exist.

Age (weeks)	0	4	8	9	10	11	12	16	20	24	28	32	36	40	44	48	52	56	60
Weight (lbs.)	1.4	15	29	33	36	40	45	60	80	100	125	140	155	165	170	175	180	185	188

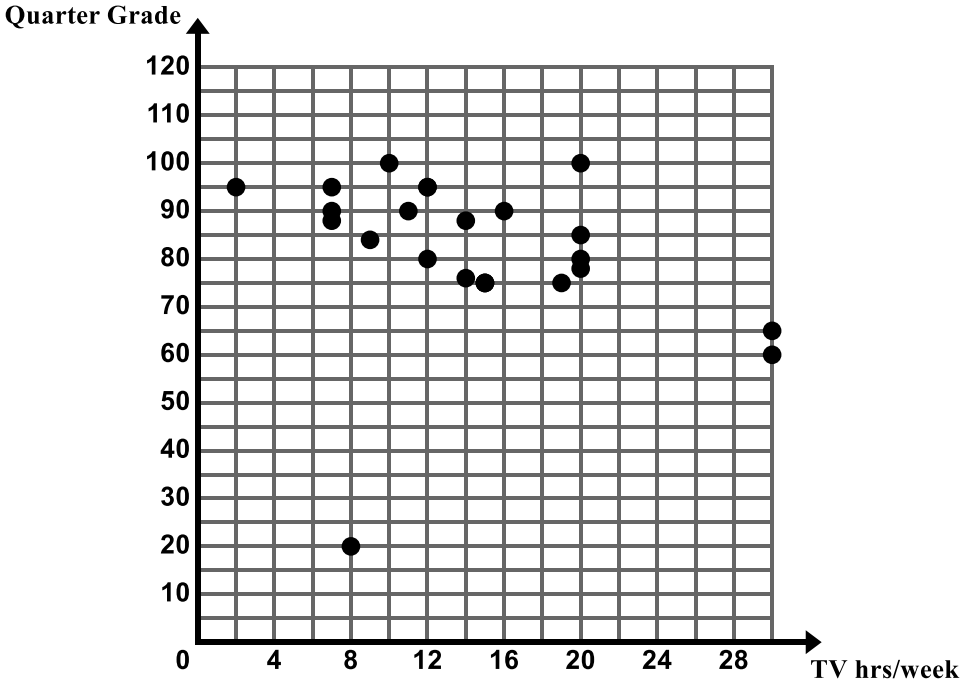


4. Mr. Clark's math classes gathered data on the average number of hours of television a student watches each week and the student's final grade at the end of the quarter. The scatter plot below shows the data. (for use with Skill/Concepts #1 and #3).
    a. Describe any patterns of association you see in this scatter plot. Use the context to give possible explanations as to why these trends, patterns, and associations exist.

    b. Can you think of a variable that when graphed with quarter grade would have a positive association?

    c. Can you think of a different variable that when graphed with quarter grade would have a negative association?

    d. Can you think of a variable that when graphed with quarter grade would have no apparent association?



Section 6.2 Construct a Linear Model to Solve Problems

Section Overview:
In this section, students continue to construct and interpret scatter plots. For scatter plots that suggest a linear association, students informally fit a straight line to the data and assess the model fit by judging the closeness of the data points to the line. They also analyze how outliers affect a line of best fit and reason about whether to drop outliers from a data set. Students then construct functions to model the data sets that suggest a linear association and use the functions to make predictions and solve real-world problems, noting that limitations exist for extreme values of x. Students will interpret the slope and y-intercept of the prediction function in context. Throughout the section students must use a critical eye, keeping in mind that most statistical data is subjective and has limitations. Students will also rely on their knowledge of the subject matter as they analyze the data.

Concepts and Skills to be Mastered:
By the end of this section students should be able to:
1. Draw a line of best fit for linear models.
2. Informally assess the model fit by judging the closeness of the data points to the line.
3. Write a prediction function for the line of best fit.
4. Explain the meaning of the slope and y-intercept of the prediction function in context.
5. Use the prediction function of a linear model to solve problems.

These practice standards are central to this entire section and chapter.

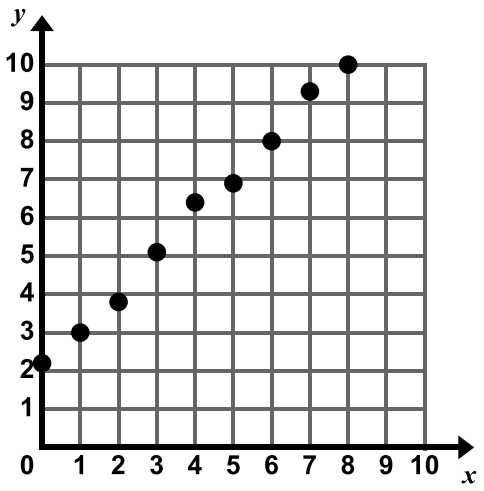
6.2a Classwork: Lines of Best Fit

Most real-world data does not fall perfectly on a line. However, if the data on a scatter plot resembles a line, we can fit a line to the data, write a function for the line, and use this function to solve problems and make predictions.

The line that you use to represent the data is called **the line of best fit**. We will refer to the function you write for the line of best fit as the **prediction function**. The most common way to find the line of best fit is to use the "eye-balling" technique. Simply try to draw a straight line that best fits the data.

Directions: In #1 and 2, observe the data sets and take note of any associations you see, draw a line of best fit, write a prediction function, and use your function to predict the value of y when $x = 12$ and when $x = 100$.

1.



a. Observations:

b. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

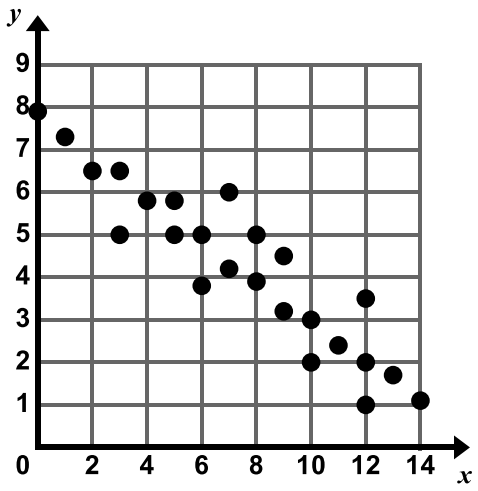
c. Estimate the slope and y-intercept of your line.

$m \approx$ _____ $b \approx$ _____

d. Write a prediction function for the data set.

e. Use your prediction function to find the value of y when $x = 12$ and when $x = 100$.

2.



a. Observations:

b. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

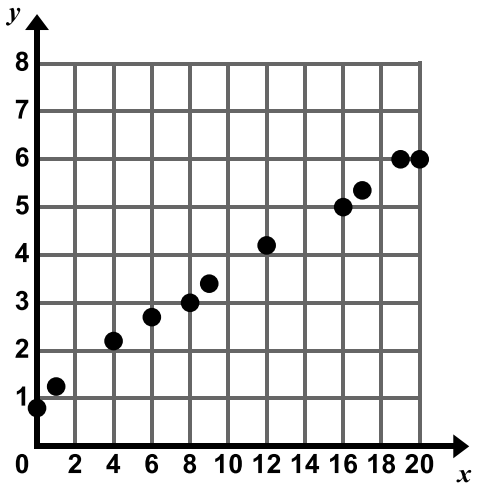
c. Estimate the slope and y-intercept of your line.

$m \approx$ _____ $b \approx$ _____

d. Write a prediction function for the data set.

e. Use your prediction function to find the value of y when $x = 12$ and when $x = 100$.

3.



a. Observations:

b. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

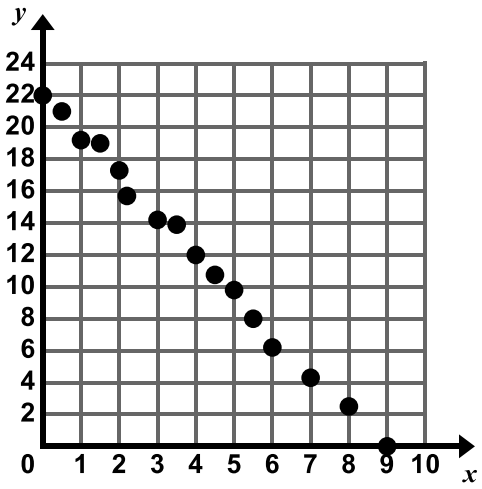
c. Estimate the slope and *y*-intercept of your line.

$m \approx$ _____ $b \approx$_____

d. Write a prediction function for the data set.

e. Use your prediction function to find the value of *y* when $x = 12$ and when $x = 100$.

4.



a. Observations:

b. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

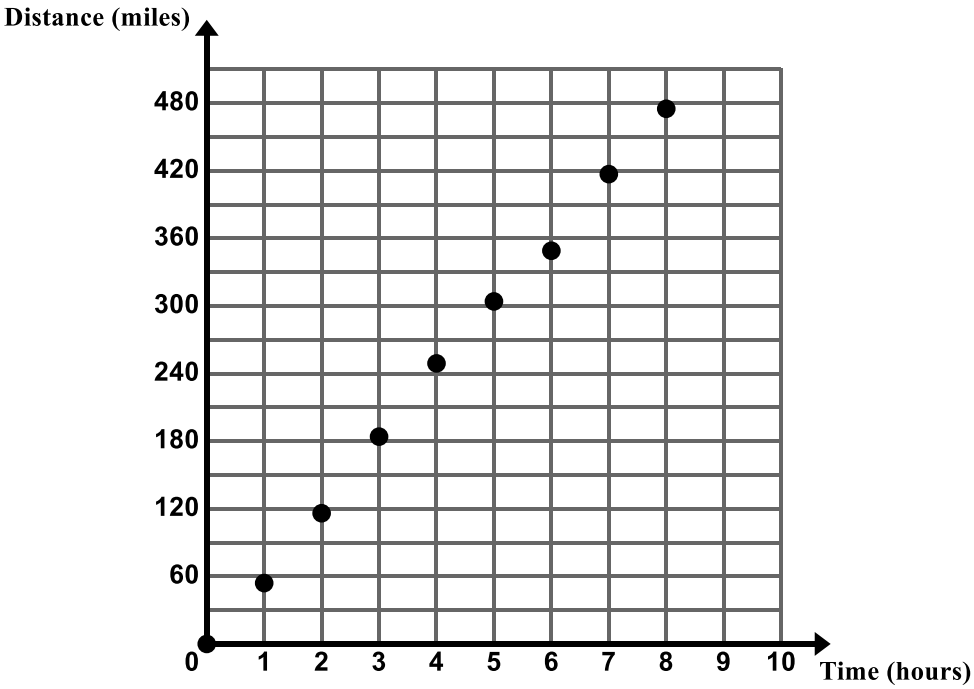
c. Estimate the slope and *y*-intercept of your line.

$m \approx$ _____ $b \approx$ _____

d. Write a prediction function for the data set.

e. Use your prediction function to find the value of *y* when $x = 12$ and when $x = 100$.

5. Camilo and his family are taking a road trip. The graph below shows the total distance the family traveled over an eight hour period.



a. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

b. Estimate the slope and *y*-intercept of your line.

$m \approx$ _____ $b \approx$ _____

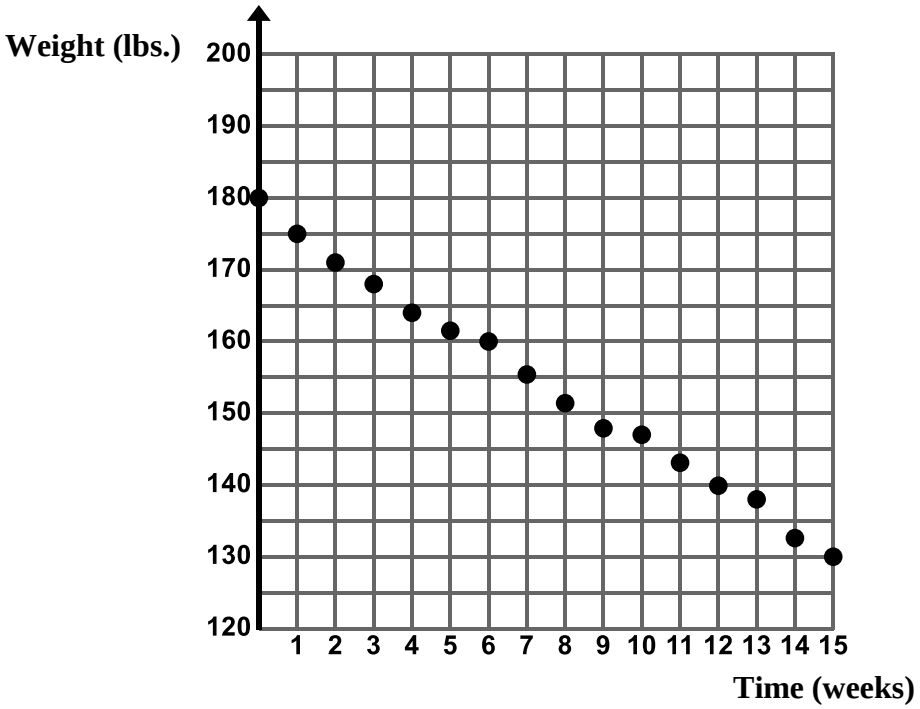
c. Write a prediction function for the data set.

d. What does the slope represent in the context?

e. What does the *y*-intercept represent in the context?

f. Predict how far Camilo and his family will have driven after 10 hours if this trend continues.

6. The scatter plot below shows the weight, in pounds, of a person who is on a strict diet.



a. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

b. Estimate the slope and *y*-intercept of your line.
$m \approx$ _____ $b \approx$ _____

c. Write a prediction function for the data set.

d. What does the slope represent in the context?

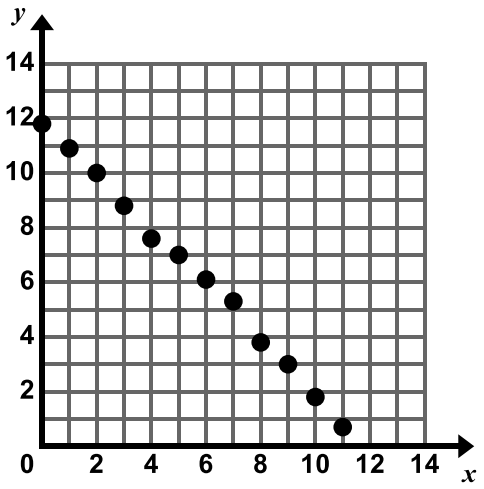
e. What does the *y*-intercept represent in the context?

f. Predict this person's weight after 18 weeks if this trend continues.

6.2a Homework: Lines of Best Fit

Directions: In #1 and 2, observe the data sets and take note of any associations you see, draw a line of best fit, write a prediction function, and use your function to predict the value of y when $x = 12$ and when $x = 100$.

1.



a. Observations:

b. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

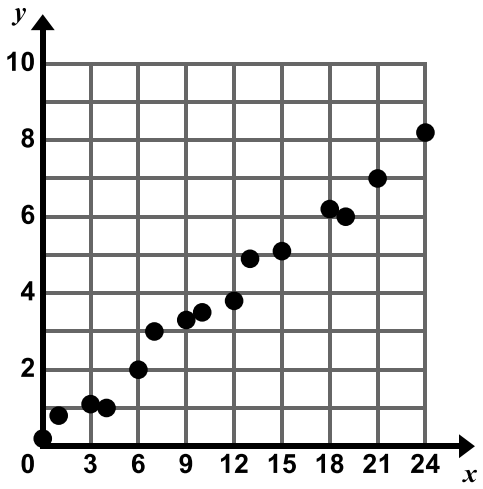
c. Estimate the slope and y-intercept of your line.

$m \approx$ _____ $b \approx$ _____

d. Write a prediction function for the data set.

e. Use your prediction function to find the value of y when $x = 12$ and when $x = 100$.

2.



a. Observations:

b. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

c. Estimate the slope and y-intercept of your line.

$m \approx$ _____ $b \approx$ _____

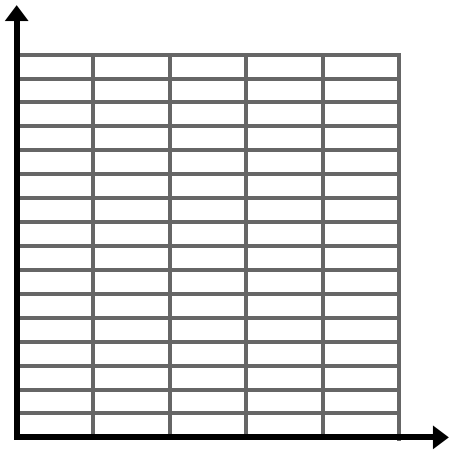
d. Write a prediction function for the data set.

e. Use your prediction function to find the value of y when $x = 12$ and when $x = 100$.

3. Use the table of data shown below to answer the questions that follow.

x	1	1	1	2	2	3	3	3	4	4	4	5
y	5	6	8	8	10	9	10	12	11	12	15	14

a. Create a scatter plot of the data on the grid below.



b. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

c. Estimate the slope and y-intercept of your line.

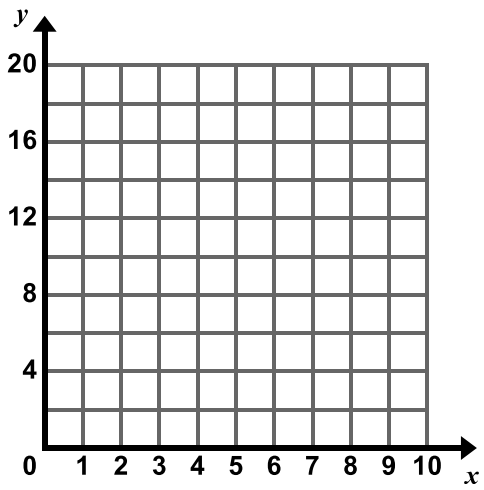
$m \approx$ _____ $b \approx$ _____

d. Write a prediction function for the data set.

4. Use the table of data shown below to answer the questions that follow.

x	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7
y	16	15	12	13	12	11	10	8	7	6	5	4	3	2	0

a. Create a scatter plot of the data on the grid below.



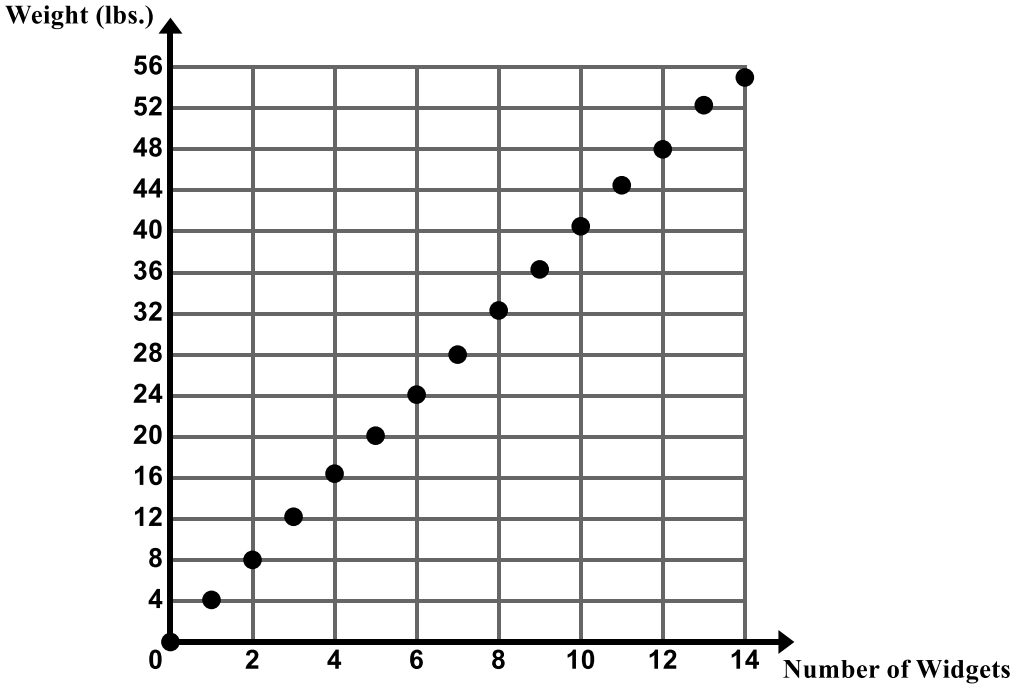
b. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

c. Estimate the slope and y-intercept of your line.

$m \approx$ _____ $b \approx$ _____

d. Write a prediction function for the data set.

5. Company XYZ makes and sells widgets. The following graph shows the weight of widgets and the number of widgets put on a scale.



a. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

b. Estimate the slope and *y*-intercept of your line.

$m \approx$ _____ $b \approx$ _____

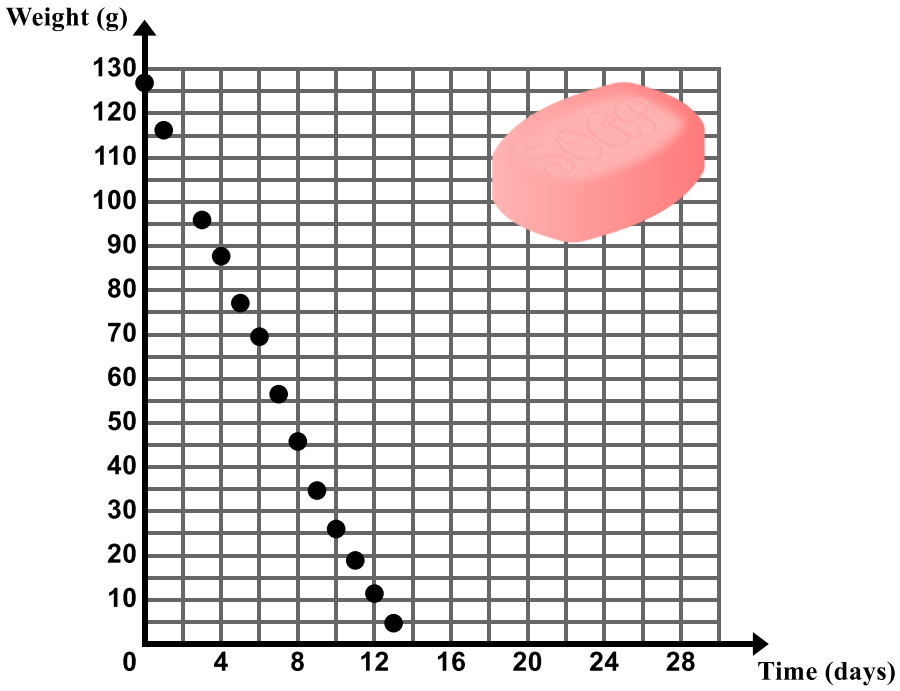
c. Write a prediction function for the data set.

d. What does the slope represent in the context?

e. What does the *y*-intercept represent in the context?

f. Predict the weight of 50 widgets.

6. Chad was trying to determine how quickly his family goes through a bar of soap in the shower. He took the weight of the soap in the shower over a period of several days.



a. Using a ruler, draw a line of best fit through the data points that captures the general trend of the data.

b. Estimate the slope and *y*-intercept of your line.

$m \approx$ _____ $b \approx$ _____

c. Write a prediction function for the data set.

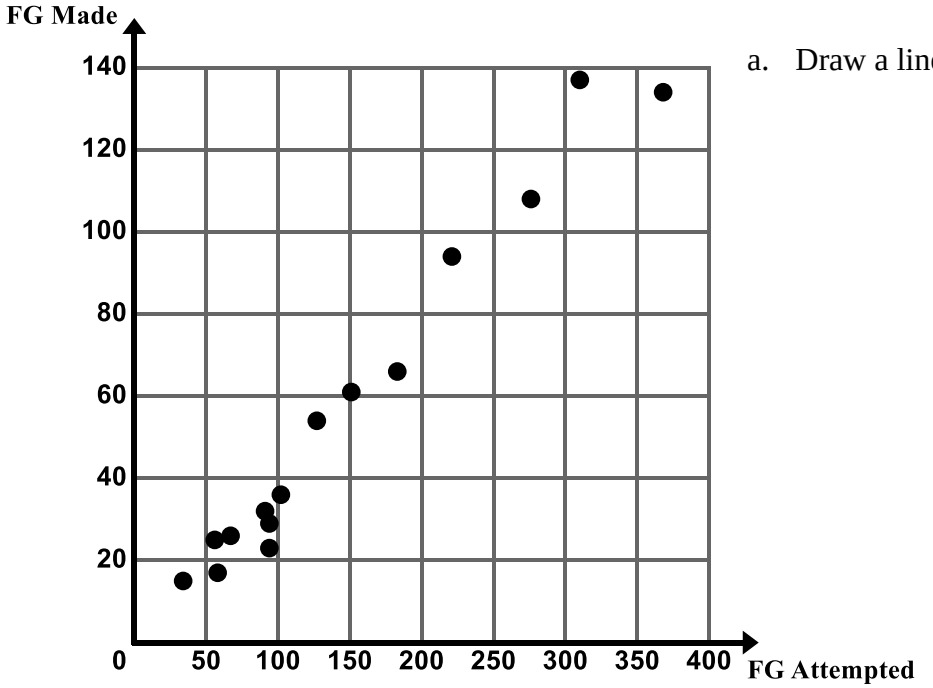
d. What does the slope represent in the context?

e. What does the *y*-intercept represent in the context?

6.2b Class Activity: Fit a Linear Model to Bivariate Data

Let's revisit some examples from section 1 where the two variables of interest had a linear association and determine a line of best fit for the data.

1. Once again refer back to Izumi's basketball statistics. Look at the scatter plot for Field Goals Made and Field Goals Attempted.



a. Draw a line of best fit on the scatter plot.

b. Write a prediction function for the line of best fit you drew.

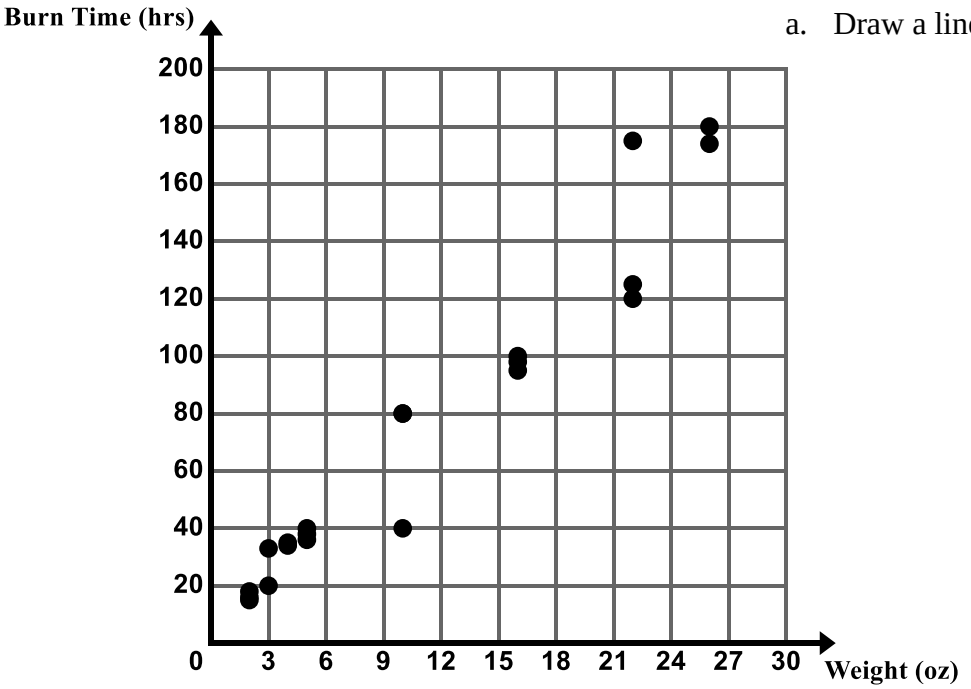
c. Explain the meaning of the slope and y-intercept in the context. 

d. Use your prediction function to predict the number of field goals a person would make if they attempted 500 field goals.

e. Use your prediction function to predict the number of field goals a person would make if they attempted 102 field goals.

f. Is the association between number of field goals attempted and number of field goals made strong or weak? Justify your answer.

2. The following scatter plot shows the burn time for candles of various weights.



a. Draw a line of best fit on the scatter plot.

b. Write a prediction function for the line of best fit you drew.

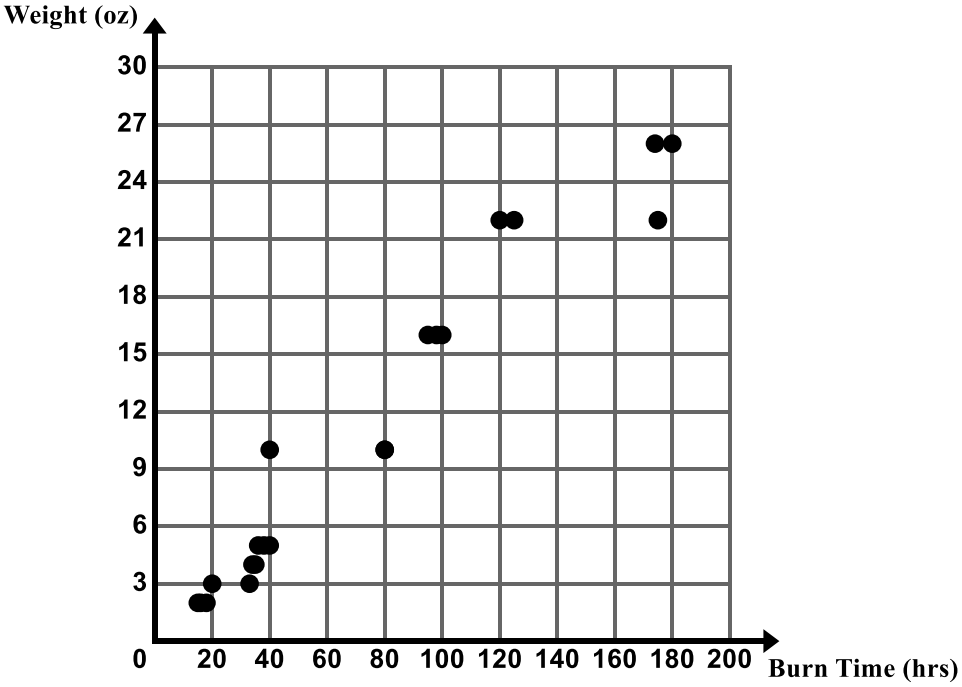
c. Explain the meaning of the slope and *y*-intercept in the context.

d. Use your prediction function to predict the burn time for a candle that weighs 40 ounces.

e. If candle burns out at 500 hours, predict how much the candle weighs.

f. What do you think would happen if we changed the graph above so that burn time was on the *x*-axis and weight was on the *y*-axis? Would our data still resemble a line? What would happen to the slope and *y*-intercept of the line of best fit?

3. The following scatter plot shows the burn time for candles of various weights. This time, burn time has been graphed on the *x*-axis and weight has been graphed on the *y*-axis.



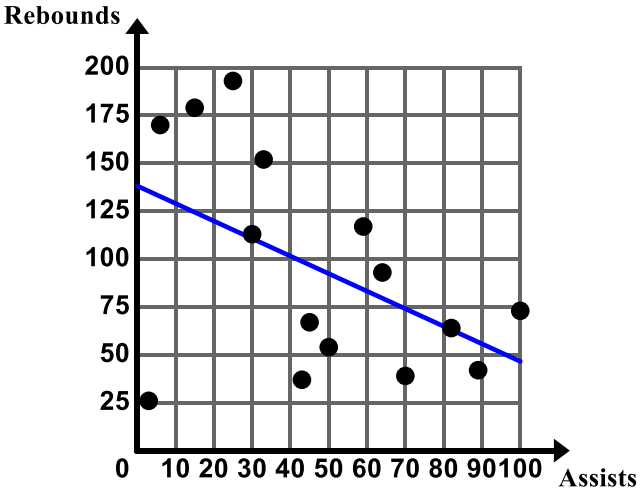
a. Was your prediction on the previous page correct?

b. Draw a line of best fit on the scatter plot.

c. Write a prediction function for the line of best fit you drew.

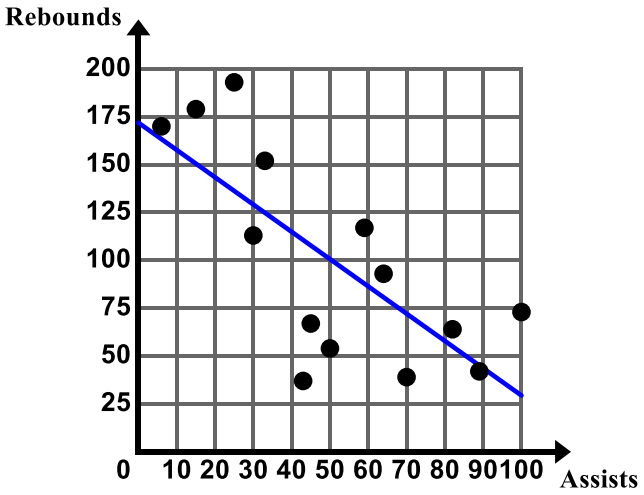
d. How does this new function compare to your equation in #2? What accounts for this change?

4. Software programs and graphing calculators can be used to draw lines of best fit. Izumi used a graphing calculator to generate a line of best fit for her data on assists and rebounds. The graph below shows the line of best fit generated by the calculator.



a. After creating this line of best fit, Izumi decided that it might be best to drop the outlier (3, 26) from her data set. Is it reasonable for Izumi to drop the outlier from her data set? Why or why not? Assume this player joined the team midway through the season.

After dropping the outlier, Izumi used the calculator to generate a new line of best fit.



b. Analyze the differences in the two lines. What did the outlier do to the line of best fit generated by the calculator?

c. Write a prediction function for the line of best fit generated by the calculator with the data set that does **not** include the outlier.

d. Explain the meaning of the slope and *y*-intercept in the context. 

e. Use your function to predict the number of rebounds a random player would have if they made 110 assists throughout the season? 150 assists? Explain the limitations that the data exhibits.

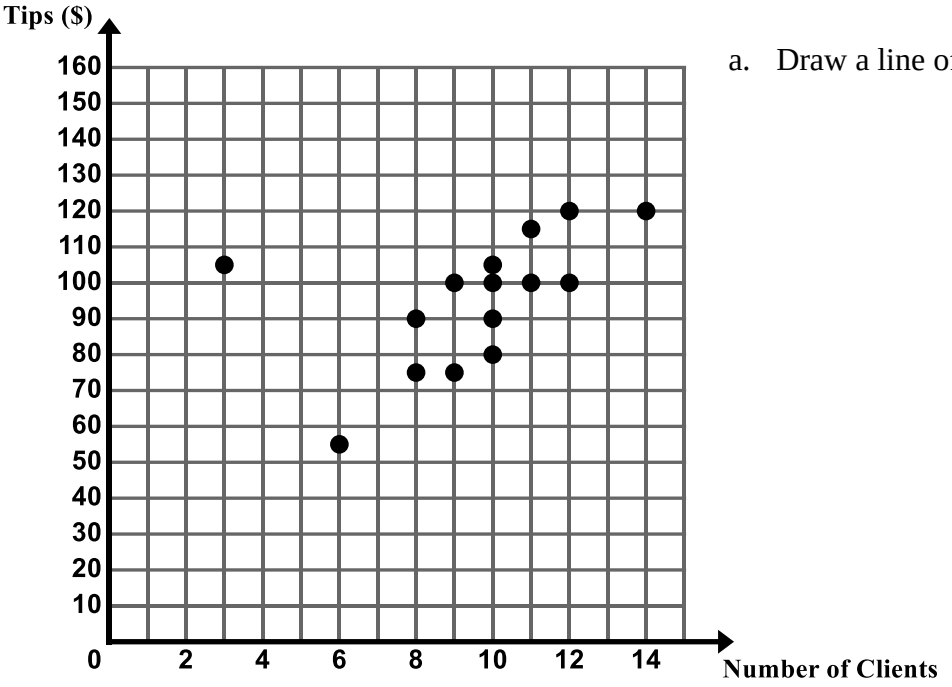

f. Similarly use your function to predict the number of assists a random player would have if they made 150 rebounds throughout the season.

5. Which scatter plot, the Field Goals Made vs. Field Goals Attempts or Rebounds vs. Assists, is more closely aligned with its line of best fit? Justify your answer. What does this tell us about the strength of each of the associations? What does this tell us about the accuracy of using each of the prediction functions to make predictions?

6.2b Homework: Fit a Linear Model to Bivariate Data

Directions: For the following problems, draw a line of best fit, write a prediction function, and use your function to make predictions. **Prior to drawing your line of best fit, determine whether you should remove any outliers from your data set.**

1. The following scatter plot shows the amount of money Jenny makes in tips based on how many clients she has in a day.



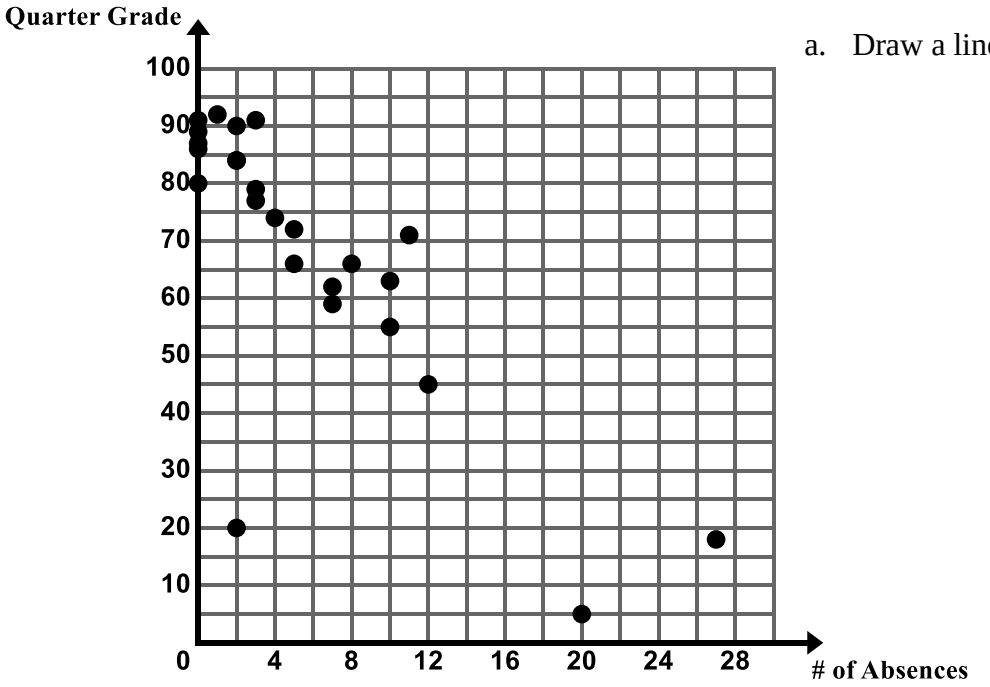
a. Draw a line of best fit on the scatter plot.

b. Write a prediction function for the line of best fit you drew.

c. Explain the meaning of the slope and *y*-intercept in the context.

d. Use your prediction function to predict the amount Jenny would make in tips if she had 18 clients in one day.

2. The following scatter plot shows the final quarter grade in Ms. Ganchero's math class for students vs. the number of times they are absent.



a. Draw a line of best fit on the scatter plot.

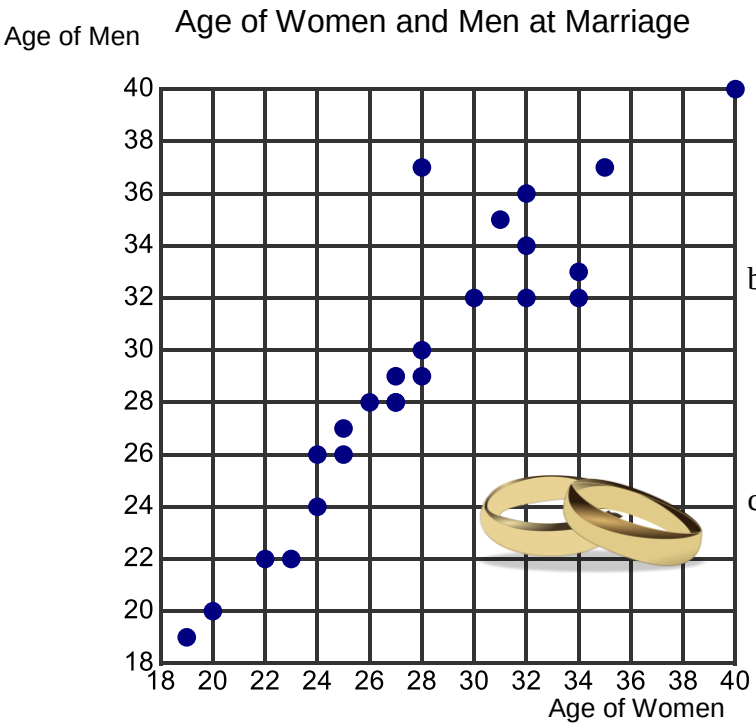
b. Write a prediction function for the line of best fit you drew.

c. Explain the meaning of the slope and *y*-intercept in the context.

d. Use your prediction function to predict the final grade of a student who is absent 16 times.

e. Use your prediction function to predict how many times a student is absent who receives a final grade of 5 in the class.

3. Bethany is interested in the relationship between the age of when men and women get married. She surveys 24 couples and asks them the age in which they got married for the first time. A scatter plot of her data is below.



a. Describe the association between the two variables. Circle any clusters in the data. Put a star by any points that appear to be outliers.

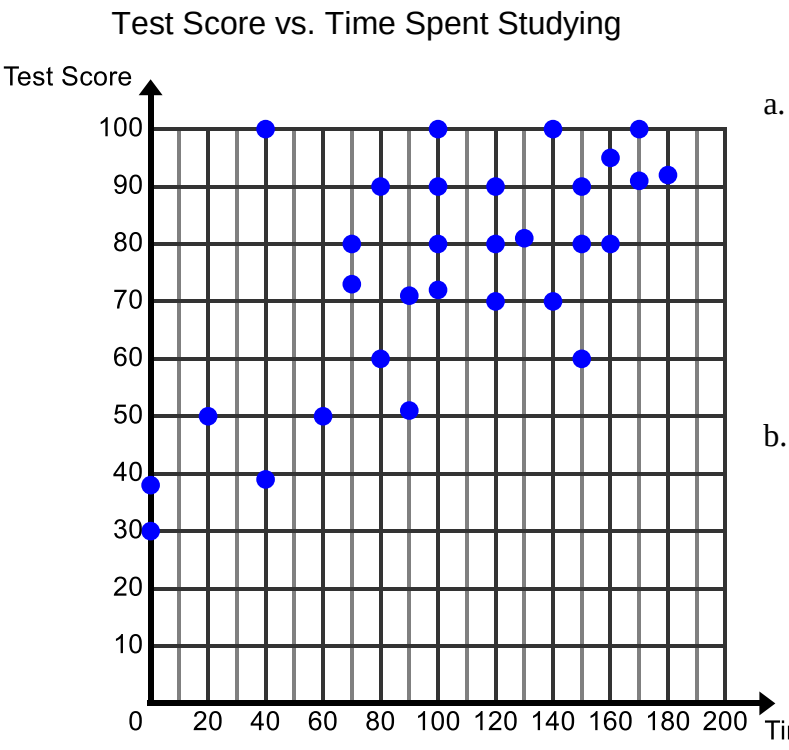
b. Provide an explanation for any clusters of data or outliers.

c. Draw a line of best fit on the scatter plot.

d. Write a prediction function for the line of best fit you drew.

e. Use your prediction function to predict the age of a man when he gets married if the woman that he marries is 38.

4. Jenna is interested in the association between the time spent studying for a test and the score that is earned. She surveys 30 people about the time they spent studying for a test and the score that they earned on the test. Her data is in the scatter plot below.



a. Describe the association between the two variables. Circle any clusters in the data. Put a star by any points that appear to be outliers.

b. Provide an explanation for any clusters of data or outliers.

c. Draw a line of best fit on the scatter plot.

d. Write a prediction function for the line of best fit you drew.

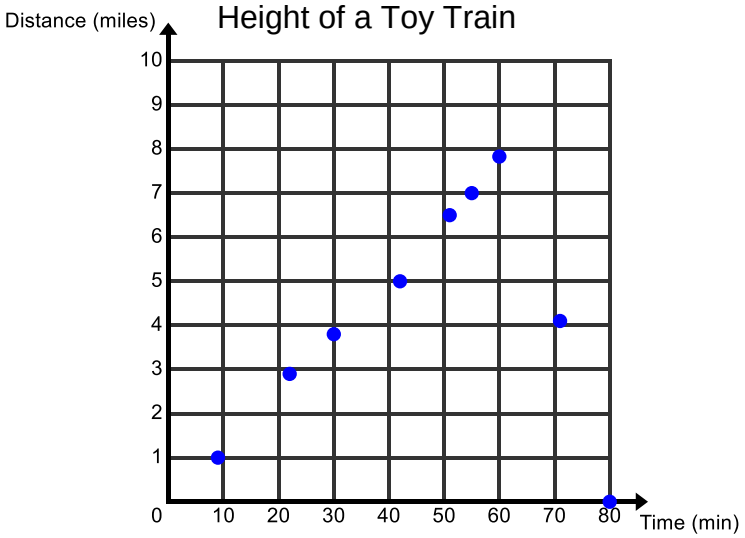
e. Explain the meaning of the slope and *y*-intercept of your line of best fit in the context.

f. Use your prediction function to predict the score for a person who studies for 160 minutes.

g. Compare and contrast the prediction calculated using the equation with the actual data points of the people who studied for 160 minutes.

h. Does the association between these two variables appear to be weak or strong? Provide an explanation regarding why the strength is this way.

5. A scatter plot given below is about the height of a toy train attached to a weather balloon. A GPS (global positioning system) records the height of the toy train about every ten minutes that it is in the air. When the train reaches the stratosphere the weather balloon pops.

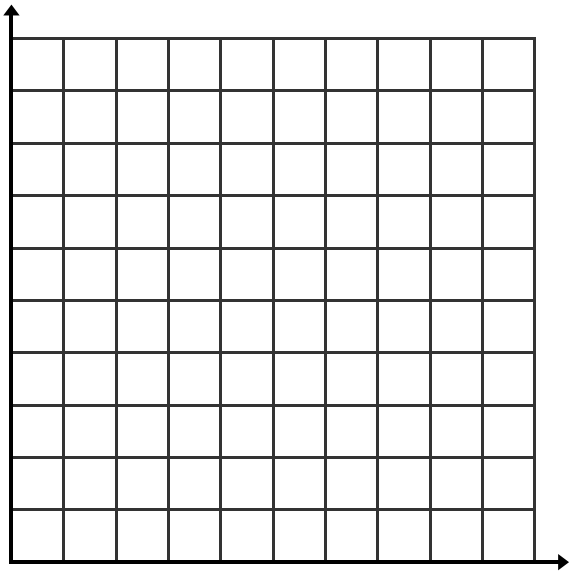


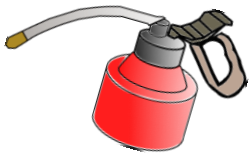
a. What kind of association exists for this data?

b. Would it be feasible to draw a line of best fit for this data? Why or why not.

6. The table gives data relating the number of oil changes every two years to the cost of car repairs.
   a. Plot the data on the graph provided, with the number of oil changes on the horizontal axis. You will need to define your own scale.

Oil Changes	3	5	2	3	1	4	6	4	3	2	0	10	7
Repair Costs	$300	$300	$500	$400	$700	$400	$100	$250	$450	$650	$600	$0	$150





   b. Write a sentence describing the association between the number of oil changes and the cost of car repairs. Is the association weak or strong?

   c. Are there any outliers or clusters that affect the data?

   d. Draw a line of best fit for the data. Assess how well the line fits the data.

   e. What is the slope of the line of best fit and what does it represent?

   f. What is the *y*-intercept of the line and what does it represent?

g. Write a prediction function in slope-intercept form that you could use to predict the cost of repairs, *y*, for any number of oil changes, *x*. Compare your prediction with that of a partner.

h. Use your prediction function to predict how much a person would spend on car repairs if they were to get 8 oil changes. Compare your prediction with that of a partner.

i. If a person spent $1,000 dollars on car repairs how many oil changes would you expect them to have?

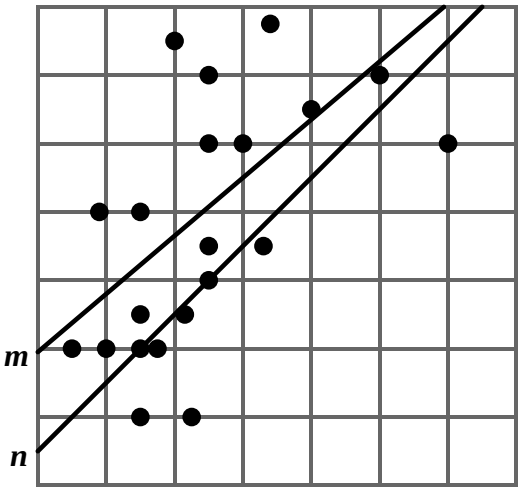
j. Based off of this data what would you recommend as the ideal number of oil changes to get every two years.

6.2c Self-Assessment: Section 6.2

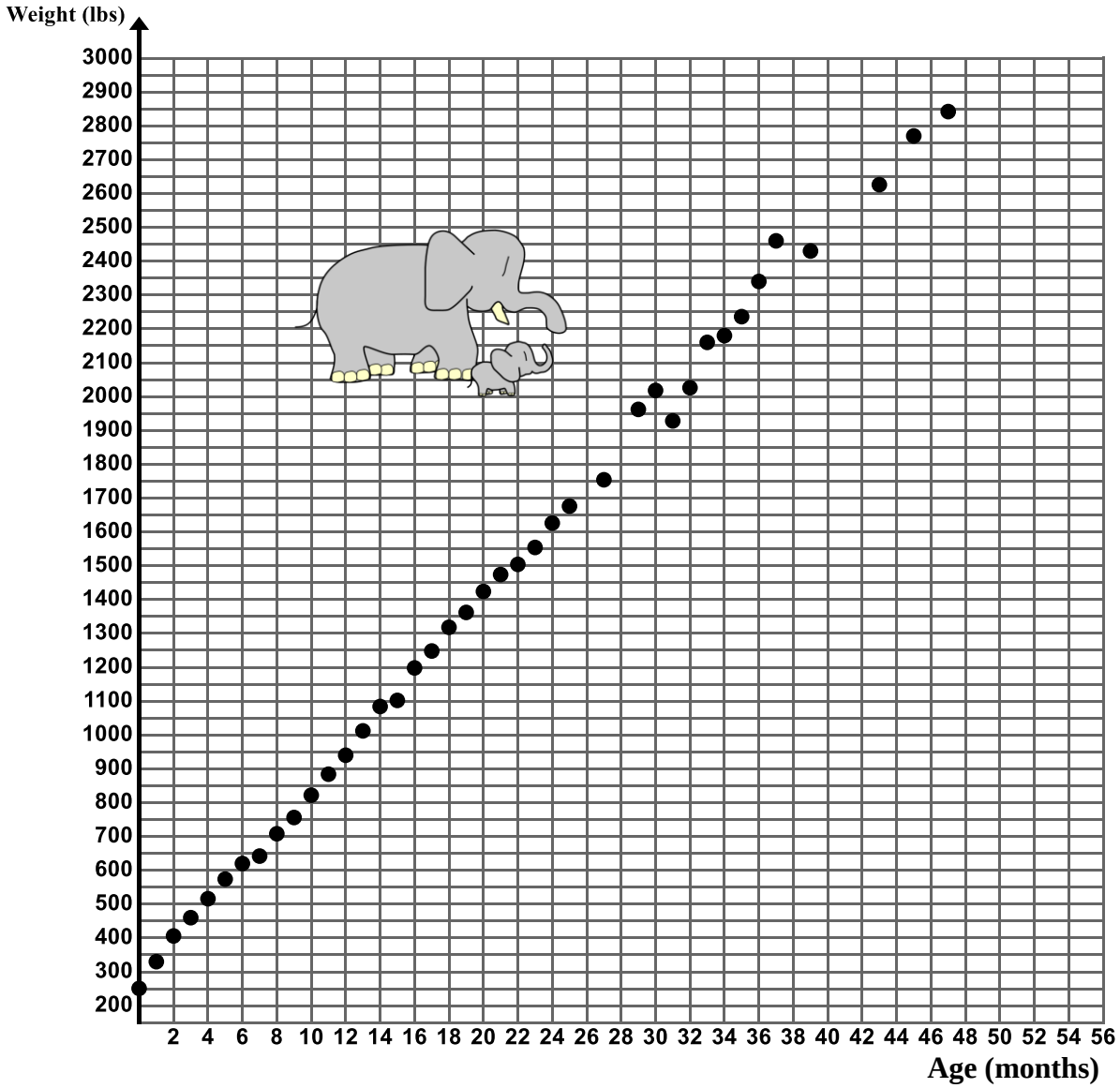
Consider the following skills/concepts. Rate your comfort level with each skill/concept by checking the box that best describes your progress in mastering each skill/concept. Sample problems for each standard can be found on the following page(s).

Skill/Concept	**Minimal Understanding 1**	**Partial Understanding 2**	**Sufficient Mastery 3**	**Substantial Mastery 4**
1. Draw a line of best fit for linear models.				
2. Informally assess the model fit by judging the closeness of the data points to the line.				
3. Write a prediction function for the line of best fit.				
4. Explain the meaning of the slope and *y*-intercept of the prediction function in context.				
5. Use the prediction function of a linear model to solve problems.				

1. Which line, *m* or *n*, is the best fit for the data? Justify your answer (for use with Skill/Concepts #2).



2. The following scatter plot show the weight of Zuri, a female African elephant born at Utah's Hogle Zoo on August 10, 2009 (for use with Skill/Concepts #1 – 5).



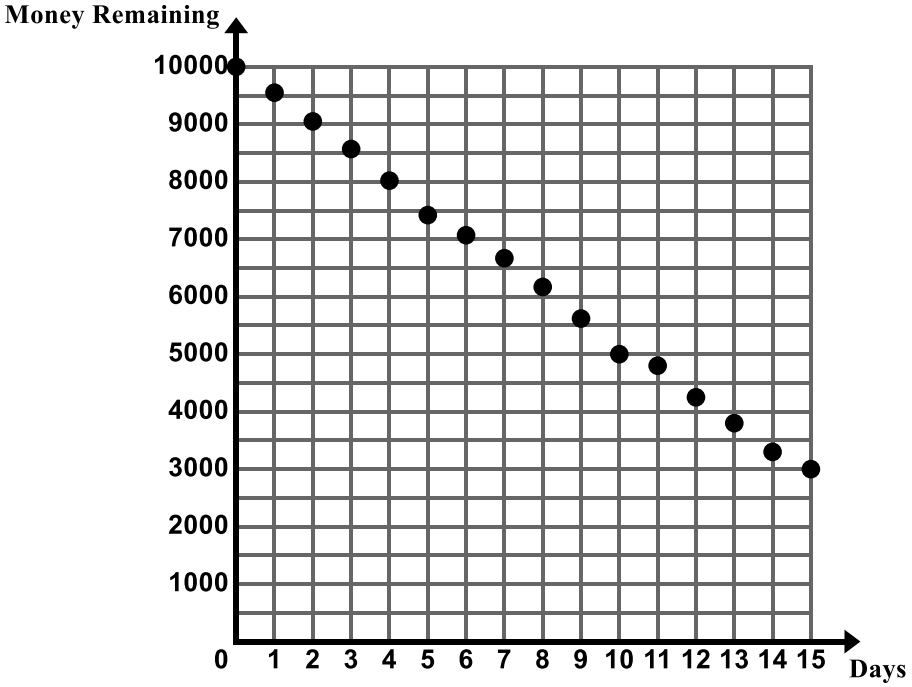
a. Describe the association between the two variables.

b. Draw a line of best fit on the scatter plot.
c. Write a prediction function for the line of best fit you drew.
d. Explain the meaning of the slope and *y*-intercept of your line of best fit in the context.

e. Use your prediction function to predict the weight of Zuri at 56 months.
f. Adult female African elephants typically weigh between 8,000 and 11,000 pounds. If Zuri's growth rate continues to follow the pattern shown in the graph above, how long will it take for her to be full grown?

Source: *Data provided by Utah's Hogle Zoo*

3. The Burgess family took a 15-day vacation to southern California and visited several popular theme parks during their trip. The graph below shows the amount of money the Burgess family had remaining at the end of each day of their trip (for use with Skill/Concepts #1 – 5).



a. Describe the association between the two variables.

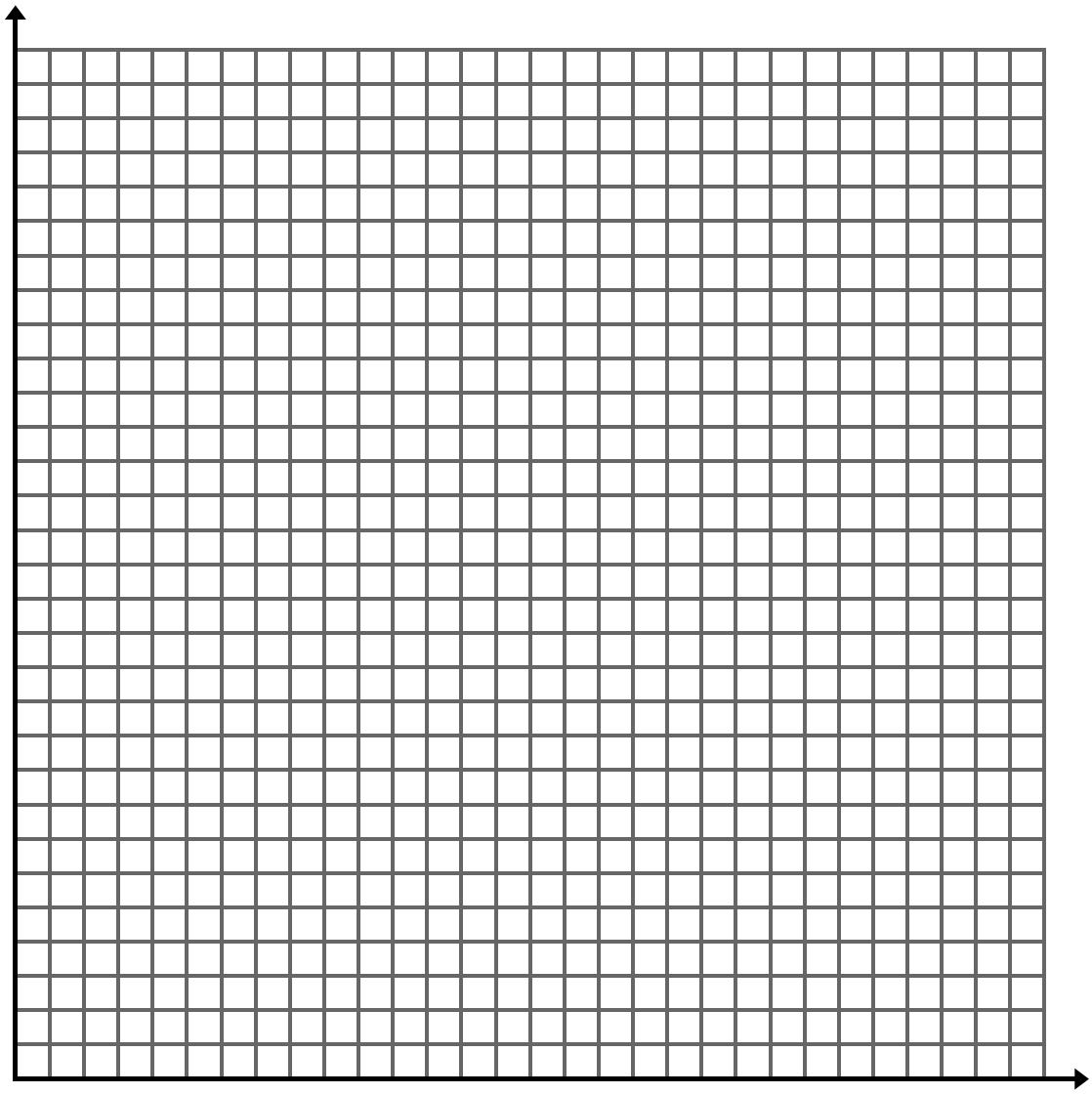
b. Draw a line of best fit on the scatter plot.

c. Write a prediction function for the line of best fit you drew.

d. Explain the meaning of the slope and *y*-intercept of your line of best fit in the context.

e. Use your prediction function to predict how much money the Burgess family will have at the end of Day 18 if they extend the length of their trip.

4. Gather data to determine whether there is an association between the height of a person and the length of their arm span. The arm span of a person is the length from one end of an individual's arms (measured at the fingertips) to the other end when the arms are raised parallel to the ground at shoulder height (for use with Skill/Concepts #1 – 5).
    a. Create a scatter plot of the data on the grid below.
    b. Describe any patterns of association you see in this scatter plot. Use the context to give possible explanations as to why these trends, patterns, and associations exist.
    c. If the plot suggests a linear association, draw a line of best fit and write a prediction function.
    d. If the plot suggests a linear association, explain the meaning of the slope and *y*-intercept in the context.



Section 6.3 Construct and Interpret Two-Way Frequency Tables to Analyze Categorcial Data

Section Overview:
At the beginning of this section students are introduced to a new type of random variable – a categorical random variable. Up to this point in the chapter, students have been studying quantitative random variables. Quantitative random variables have a cardinal numerical value. Categorical random variables are those that represent some quality or name. Categorical data is often represented and summarized in a two-way frequency table. In this section, students learn what a two-way frequency table is and how to read it. They complete two-way frequency tables by filling in missing data. As the section progresses, students begin to formally interpret the frequency tables. They calculate and analyze relative frequencies (for rows, columns, and the entire table) to describe possible associations between the two variables and use these associations to make decisions. Finally, students conduct a survey of their own involving categorical random variables, summarize their data in a two-way frequency table, and analyze the data to determine if an association exists between the two variables of interest.

Concepts and Skills to be Mastered:
By the end of this section students should be able to:
1. Read and understand a two-way frequency table.
2. Construct a two-way frequency table for categorical data.
3. Calculate and analyze relative frequencies (for rows, columns, and the entire table) to describe possible associations between the two variables and to make decisions.

6.3a Class Activity: Construct Two-Way Frequency Tables using Categorical Data

There are two different types of random variables when looking at bivariate data; **quantitative random variables** and **categorical random variables**. So far in this chapter, we have been studying **quantitative random variables**. Quantitative random variables can be counted or measured. For example, we can count the number of assists and rebounds that a player on Izumi's team had during the team. We can count the amount that Jenny made in tips each day. We can measure a person's shoe size and their height. We can measure the amount of time it takes to say a tongue twister. A **categorical random variable** represents a quality or a name. Suppose we were interested in determining if there is an association between a person's gender and whether or not that person has pierced ears. We would interview people and classify them as male or female and as yes (ears pierced) or no (ears not pierced). Suppose we were interested in whether a person's favorite color is associated with their favorite holiday. We would categorize a person according to their favorite color (red, orange, yellow, etc.) and their favorite holiday (Christmas, Thanksgiving, Halloween, Hanukah, etc.) 

Directions: Determine if the following random variables represent data that is Quantitative or Categorical.

1. Gender of babies born in the Riverton Hospital for the month of June

2. Thickness of the plastic for various types of water bottles

3. Favorite ice cream flavor chosen from the following options; chocolate, vanilla, or strawberry

4. The number of pages you can read of your favorite book before you fall asleep

In the previous sections we summarized and displayed quantitative data using a **scatter plot**. In this section, we will summarize and display categorical bivariate data using a **two-way frequency table**. A two-way frequency table is "two-way" because each bivariate data entry is composed of an ordered pair from two categorical random variables.

Suppose we were interested in whether there is an association between a person's gender (male/female) and whether or not they smoke (smoker/non-smoker). The following ordered pairs are possible outcomes for our experiment:

(female, non-smoker) (female, smoker) (male, non-smoker) (male, smoker)

The table is a "frequency" table because the cell entries count the number of data points that fall into each combination of categories.

In this section, we will construct two-way frequency tables and analyze the tables to determine if there is an association between the two variables of interest.

5. Carlos enjoys spending time with his friends. He feels sad when one of his friends cannot hang out with him. Often when one of his friends cannot hang out with him it is because they are either doing their chores or they cannot stay out late at night. Carlos notices that it tends to be the same group of friends that have curfews on school nights who also have chores to do at home. He wonders, "In general, do students at my school who have chores to do at home tend to also have curfews at night?"

   Carlos decides to conduct an experiment to help answer his question. He randomly surveys 52 students at his school, asking each student if they have a curfew and if they have to do household chores. He organizes his findings into the frequency table below.

	Has A Curfew	**No Curfew**	**Total**
Has Chores	26	9	
No Chores	5	12	
Total			

Directions: Use the table to answer each question below.

a. How many students have a curfew and have chores?

b. How many students have no curfew and have chores?

c. How many students have no curfew and no chores?

It is also possible to calculate the frequencies for "Total" column and "Total" row. These frequencies represent the total count of one variable at a time.

d. Find the frequencies for the Total column and Total row by adding up the numbers in each column and row. Write these numbers in the table above.

e. How many of the students surveyed have chores?

f. How many of the students surveyed have a curfew?

You can also calculate how many total students that were surveyed by adding up the frequencies in the "Total" row and "Total" column.

g. Add the entries in the Total row and the Total column and put this number in the cell in the bottom left corner. Does this number match how many students that Carlos said he was going to survey?

6. Emina loves to eat tomatoes from her garden in Salt Lake City. She asked her friend Renzo, "Don't you just love tomatoes?" Renzo crinkled his nose and replied, "Ew, tomatoes gross me out! When I see them in the grocery store, I just keep on walking." Renzo's response prompted Emina to think, "I don't buy tomatoes at the grocery store either, because I grow them in my garden. The tomatoes from my garden are delicious, whereas grocery store tomatoes look less appealing to me. I wonder if there is an association between enjoying tomatoes and having a garden at home."

   She decides to survey 100 randomly selected Salt Lake City vegetable eating residents and asks each of them two questions: 1. Do you primarily obtain your vegetables at the grocery store (including food pantry), the farmer's market, or your home garden (assume they grow tomatoes in their home garden)? Do you like tomatoes? Her results are summarized in the table below.

	Grocery Store	**Farmer's Market**	**Home Garden**	**Total**
Likes Tomatoes	50	4	12	
Dislikes Tomatoes	30	1	3	
Total				

   a. Fill in the frequencies for the Total column and Total row in the table.

   b. Check to make sure that you found the above frequencies correctly by finding the total number of people surveyed.

   c. How many people get their tomatoes at the farmer's market and dislike tomatoes?

   d. How many people get their tomatoes from a home garden and like tomatoes?

   e. How many people get their tomatoes from the grocery store?

   f. How many people like tomatoes?

Emina is not quite sure if her data suggests an association between enjoying tomatoes and having a garden. We will further investigate this relationship in the next section.

7. Use the given information to complete the two-way frequency table about the eating habits of 595 students at Copper Ridge Middle School.
   - 190 male students eat breakfast regularly out of 320 total males surveyed.
   - 295 students do not eat breakfast regularly
   - 165 females do not eat breakfast regularly

   a. Fill in the missing information.

	Male	Female	Total
Eat breakfast regularly			
Do not eat breakfast regularly			
Total			

   b. How many females total were surveyed?

   c. How many people surveyed eat breakfast regularly?

   d. How many people total were surveyed?

   e. How many males surveyed do not eat breakfast regularly?

   f. How many females surveyed eat breakfast regularly?

   g. What percentage **of the total number of people surveyed** eat breakfast regularly?

   h. What percentage **of the females surveyed** eat breakfast regularly?

   i. What percentage **of the people who eat breakfast regularly** are male?

   j. What percentage **of the total number of people surveyed** are females who do not eat breakfast regularly?

   k. Make up your own problem similar to the problems in parts g. – j. Have a partner answer your question.

   l. Make up a different problem similar to the problems in parts g. – j. Have a partner answer your question.

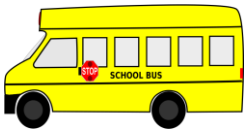
8. The data given in the table below is about modes of transportation to and from school at Brookside High School.

a. Fill in the missing information.

	Walk	Car	Bus	Cycle	Total
Male		28			129
Female	46		12	17	92
Total		45	27	69	

b. How many males ride their bikes to school?

c. How many females take the bus to school?



d. How many females were surveyed?

e. How many students were surveyed?

f. What percentage **of the total number of people surveyed** walk to school?

g. What percentage **of the total number of people surveyed** are females that bike to school?

h. What percentage **of the males surveyed** cycle to school?

i. Make up your own problem similar to the problems in parts f. – h. Have a partner answer your question.

j. Make up a different problem similar to the problems in parts f. – h. Have a partner answer your question.

9. Keane collects data about the number of people who own a smart phone and if they also own an MP3 player. He gives you the following information.
    - 25 people surveyed owned smart phones
    - 20 people that own a smart phone do not own an MP3 player
    - 9 people do not own smart phones but they do own an MP3 player
    - 24 people do not own an MP3 player

    a. Design and complete a two-way frequency table to show the display the data.

    b. How many people did Keane survey?

    c. How many people own a smart phone and an MP3 player?

    d. How many people own an MP3 player?

10. Tamra wondered if there is an association between age and favorite flavor of ice cream (choices: chocolate, strawberry, and vanilla). She surveyed 200 children in different age ranges. The table below shows the results of her survey.

    Tamra gives you the following information.

    - $\frac{1}{2}$ of the children surveyed chose chocolate as their favorite flavor
    - 25% of the children surveyed were in the age range of 8 – 12 years old
    - $\frac{2}{5}$ of the children surveyed were in the age range of 13 – 17 years old
    - 50% of the children in the age range of 3 – 7 years old chose chocolate as their favorite flavor
    - 50 children chose strawberry as their favorite flavor

    a. Complete the two-way frequency table to display the data.

	Chocolate	**Vanilla**	**Strawberry**	**Total**
Ages 3 – 7				
Ages 8 – 12	25		12	
Ages 13 – 17			12	
Total				200

6.3a Homework: Construct a Two-Way Frequency Table

1. In Miss Marble's music collection there are…
   - 208 songs in total
   - She has 150 songs in her "Workout Music" playlist
   - 162 of the songs in the total music collection are Pop songs
   - 38 Classical songs are in her "Music for Studying" playlist

   a. Complete the table for about the Miss Marble's music collection.

	Workout Music	**Music for Studying**	**Totals**
Classical			
Pop			
Totals			

   b. How many total songs are in her "Music for Studying" playlist?

   c. How many classical songs are in her "Workout Music" playlist?

   d. What percentage **of songs in the collection** are pop?

   e. What percentage **of songs in the collection** are for studying?

   f. What percentage **of the classical music** is music for studying?

   g. What percentage **of songs in the collection** are classical music for studying?

2. Laura was driving home from school and texting her mom at the same time. She did not notice that she was speeding and a police officer pulled her over and gave her a traffic citation. She wonders if there is an association between people who regularly text while driving and if they have received a traffic citation in the last 2 years. She conducts a survey among 50 drivers and records some data in the table below.

   a. Fill in the missing information in the frequency table below.

	Regularly Texts While Driving	**Never Texts While Driving**	**Totals**
No traffic citations			
Has received a traffic citation in the last two years.	18	5	
Totals	25		50

   b. How people regularly text while driving?

   c. How many people have no traffic citations and regularly text while driving?

3. Paul tosses a dice and spins a coin 150 times as part of an experiment. He records 71 heads and a six 21 times. On 68 occasions, he gets neither a head nor a six. Complete the table.

	Six	**Not a Six**	**Totals**
Head			
Tail			
Totals			

   a. How many times did he toss a tails and a six?

   b. How many times did he toss a heads?

4. The 300 members of a tennis club are classified by gender and whether or not they are over 18. You are given the following information about the members of the club.
    - 36 are under 18 and female
    - 159 are over 18 and male
    - 180 are male

    a. Design and complete a two-way table to show this information.

    b. How many members of the club are female?

    c. How many member of the club are over 18 and female?

    d. What percentage **of the members** are female?

    e. What percentage **of the members** are age 18 and over?

    f. What percentage **of the members** are males under age 18?

    g. What percentage **of the members age 18 and over** are male?

5. Susan loves social media and is interested in at what age people prefer different social media outlets. She groups people into the following age groups, middle school age, high school age, and college age. She then asks 75 people what their favorite form of social media is, Twitter, Instagram, or Facebook.
   a. Fill in the missing information in the frequency table below.

	Facebook	Instagram	Twitter	Totals
Middle School		5	3	
High School	10	10		27
College		7		24
Total	31		22	

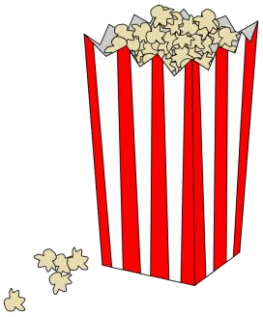
   b. How many Middle School aged people were surveyed?

   c. How many people prefer Instagram?

   d. How many college age people prefer Facebook?

   e. How many high school aged people prefer Twitter?

6. Julie wants to know if there is an association between gender and the type of movie a person prefers. She surveys 500 people and discovers the following.
   - 35% of the people surveyed prefer comedy movies
   - $\frac{3}{10}$ of the people surveyed prefer action movies
   - 95 people surveyed prefer romance movies
   - Of the females surveyed, $\frac{2}{7}$ prefer romance movies
   - 35% of the males surveyed prefer comedy movies



   a. Complete the two-way frequency table to display the data.

	Romance	Comedy	Action	Drama	Totals
Male					
Female				52	280
Total					

6.3b Class Activity: Interpret Two-Way Frequency Tables

Now that we are comfortable making a two-way frequency table we are going to see what conclusions we can draw from them. 

1. The table below displays the data Julie gathered on gender and the type of movie a person prefers. Use **numerical evidence** from the table to answer the questions below.

	Romance	**Comedy**	**Action**	**Drama**	**Totals**
Male	15	77	100	28	220
Female	80	98	50	52	280
Total	95	175	150	80	500

   a. Julie is showing a movie at a party at which males and females will be present. Which type or types of movies should Julie show?

   b. Julie is showing a movie at a party at which only males will be present. Which type or types of movies should Julie show?

   c. Julie is showing a movie at a party at which only females will be present. Which type or types of movies should Julie show?

   d. Determine whether the following statement is true or false based on the data in the table. Put a "T" on the line if it is true and an "F" on the line if it is false. **Use numerical evidence to support your answer.**

      _____ Males and females have an equal likelihood of choosing comedy movies.

2. The table below show the results of the data Tamra collected on age and favorite flavor of ice cream (choices: chocolate, strawberry, and vanilla). Use **numerical evidence** from the table to answer the questions below.

	Chocolate	**Vanilla**	**Strawberry**	**Total**
Ages 3 – 7	35	9	26	70
Ages 8 – 12	25	13	12	50
Ages 13 – 17	40	28	12	80
Total	100	50	50	200

a. Tamra is in charge of buying ice cream for a pre-school carnival. Which type or types of ice cream should she purchase?

b. Tamra is in charge of buying ice cream for a neighborhood picnic at which all ages of children will attend. What type or types of ice cream should she buy?

c. Determine whether the following statements are true or false based on the data in the table. Put a "T" on the line if the statements are true and an "F" on the line if the statements are false. **Use numerical evidence to support your answer.**

_____ Children in all of the age ranges have an equal likelihood of choosing chocolate.

_____ Children in the age ranges 8 – 12 and 13 – 17 have an equal likelihood of choosing strawberry.

_____ As students get older they tend to like vanilla more.

3. Refer back to Carlos' data regarding chores and curfew.

	Has A Curfew	**No Curfew**	**Totals**
Has Chores	26	9	35
No Chores	5	12	17
Totals	31	21	52

a. Analyze the two-way table. What arguments can you make about the data? Use numerical evidence to support your answer.

b. Is there an association between kids having chores and having a curfew? Use numerical evidence to support your answer.

4. Let's revisit Emina and her tomatoes.

	Grocery Store	Farmer's Market	Home Garden	Totals
Likes Tomatoes	50	4	12	66
Dislikes Tomatoes	30	1	3	34
Totals	80	5	15	100

a. Analyze the two-way table. What arguments can you make about the data? Use numerical evidence to support your answer.

b. Is there an association between growing your own tomatoes (having a home garden) and whether or not you like tomatoes?

5. In the previous section you made a frequency table about gender and eating breakfast.

	Male	Female	Totals
Eat breakfast regularly	190	110	300
Do not eat breakfast regularly	130	165	295
Totals	320	275	595

   a. Is there an association between gender and whether or not a person eats breakfast regularly.

6. Eddy wanted to determine whether there is an association between gender and whether or not a person has their ears pierced. He collected data from a random sample of young adults ages 13 – 18.

	Has Pierced Ears	Does not have Pierced Ears	Totals
Male	19	71	**90**
Female	84	4	**88**
Totals	**103**	**75**	**178**

   a. Is there an association between gender and whether or not a person has their ears pierced?

6.3b Homework: Interpret Two-Way Frequency Tables

1. **Modes of Transportation:** Recall the data gathered from Brookside High School about modes of transportation and gender. Use **numerical evidence** from the table to answer the questions below.

	Walk	Car	Bus	Cycle	Total
Male	34	28	15	52	129
Female	46	17	12	17	92
Total	80	45	27	69	221

Directions: Answer the following questions about the data collected:

a. What percentage **of students surveyed** take the bus to school?

b. What percentage **of students surveyed** are males who walk to school?

c. Based off of the table above what is the most popular mode of transportation for the sample population.

d. What is the preferred method of transportation for females? Use numerical evidence to support your answer.

e. What is the preferred method of transportation for males? Use numerical evidence to support your answer.

f. Is taking the bus more common with males or females?

2. **Cell Phones and MP3 Players:** Recall the two-way table you made in the previous section about Keane's data on Cell Phones and MP3 Players below. Use **numerical evidence** from the table to answer the questions below.

	Owns a smart phone	**Does not own a smart phone**	**Total**
Owns an MP3 player	5	9	14
Does not own an MP3 Player	20	4	24
Total	25	13	38

a. What percentage **of the people surveyed** own a smart phone?

b. What percentage **of the people surveyed** do not own a smart phone but own an MP3 player?

c. What percentage **of the people surveyed** own a smart phone and an MP3 player?

d. Is there an association between owning a smart phone and owning an MP3 player? Use numerical evidence to support your answer.

3. **Music:** Use the two-way frequency table given below about Miss Marbles' music playlists to answer the following questions.

	Workout Music	**Music for Studying**	**Totals**
Classical	8	38	46
Pop	142	20	162
Totals	150	58	208

a. Is there an association between what Miss Marble is doing (exercising or studying) and what she is listening to? Use numerical evidence to support your answer.

4. **Texting While Driving:** Use the two-way given below about texting while driving to answer the questions that follow.

	Regularly Texts While Driving	**Never Texts While Driving**	**Totals**
No traffic citations	7	20	27
Has received a traffic citation in the last two years.	18	5	23
Totals	25	25	50

a. What percentage of people regularly text while driving?

b. What percentage of people have not received a traffic citation in the last two years?

c. What percentage of people regularly text and have received a traffic citation in that last two years?

d. What percentage of people who never text have no traffic citations?

e. What percentage of people who regularly text while driving have received a traffic citation in the last two years?

f. Out of all the people who have received a traffic citation in the last two years, what percentage of them text regularly?

g. What type of association exists between texting while driving and receiving traffic citations? Use numerical evidence to support your answer.

5. **Social Media:** Use the two-way frequency table given below to answer the questions that follow.

	Facebook	Instagram	Twitter	Totals
Middle School	16	5	3	24
High School	10	10	7	27
College	5	7	12	24
Total	31	22	22	75

a. Analyze the two-way table. What arguments can you make about the data? Use numerical evidence to support your answer.

6.3c Class Activity: Conduct a Survey

Is there an association between whether a student plays a sport and whether he or she plays a musical instrument? **This problem was adapted from an Illustrative Mathematics task.*
To investigate these questions, ask 20 students in your class to answer the following two questions:

1. Do you play a sport? (yes or no)

2. Do you play a musical instrument? (yes or no)

3. Record the answers in the table below.

Student Name	Sport?	Musical Instrument?
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		

4. Summarize the data into a clearly labeled frequency table.

Use the tables that you made above to answer the following questions.

5. What percentage of students play a sport and a musical instrument?

6. What percentage of students that play a sport also play a musical instrument?

7. What percentage of students that do not play a sport play a musical instrument?

8. What percentage of musical instrument players do not play a sport?

9. Based on the class data, do you think there is an association between playing a sport and playing an instrument? Use numerical evidence to support your answer.

6.3d Self-Assessment: Section 6.3

Consider the following skills/concepts. Rate your comfort level with each skill/concept by checking the box that best describes your progress in mastering each skill/concept. Sample problems for each standard can be found on the following page(s).

Skill/Concept	Minimal Understanding 1	Partial Understanding 2	Sufficient Mastery 3	Substantial Mastery 4
1. Read and understand a two-way frequency table.				
2. Construct a two-way frequency table for categorical data.				
3. Calculate and analyze relative frequencies (for rows, columns, and the entire table) to describe possible associations between the two variables and to make decisions.				

1. Lisa is the owner of a local gym and is trying to determine if there is an association between gender and a person's favorite workout class. She gathers data and organizes it into the two-way frequency table shown below.

	Zumba	Spinning	Weight Lifting	Step	Total
Male	2		25	5	
Female		16			
Total	45	30	35		150

a. Complete the table.
b. How many females chose Zumba as their favorite workout class?
c. How many males chose spinning as their favorite workout class?
d. How many females were surveyed?
e. How many people were surveyed?
f. What percentage of the people surveyed chose step as their favorite class?
g. What percentage of the people who chose spinning as their favorite class are male?
h. What percentage of the males surveyed chose weight lifting as their favorite class?
i. Based on the data, do you think there is an association between gender and a person's favorite workout class? Use numerical evidence to support your claim.
j. Are there any other conclusions you can draw from the table? Use numerical evidence to support your claims.