## Table of Contents

**WF**

**CHAPTER 6: STATISTICS-INVESTIGATE PATTERNS OF ASSOCIATION IN BIVARIATE DATA (2 WEEKS)........ 2**

*6.0 Anchor Problem: Tongue Twisters........ 7*

SECTION 6.1: CONSTRUCT AND INTERPRET SCATTER PLOTS FOR BIVARIATE DATA........ 9

*6.1a Class Activity: Read and Interpret a Scatter Plot........ 10*

*6.1a Homework: Read and Interpret a Scatter Plot........ 16*

*6.1b Class Activity: Create and Analyze a Scatter Plot........ 19*

*6.1b Homework: Create and Analyze a Scatter Plot........ 21*

*6.1c Classwork: Patterns of Association........ 23*

*6.1c Homework: Patterns of Association........ 27*

*6.1d Self-Assessment: Section 6.1........ 30*

SECTION 6.2 CONSTRUCT A LINEAR MODEL TO SOLVE PROBLEMS........ 34

*6.2a Classwork: Lines of Best Fit........ 35*

*6.2a Homework: Lines of Best Fit........ 39*

*6.2b Class Activity: Fit a Linear Model to Bivariate Data........ 44*

*6.2b Homework: Fit a Linear Model to Bivariate Data........ 50*

*6.2c Self-Assessment: Section 6.2........ 57*

SECTION 6.3 CONSTRUCT AND INTERPRET TWO-WAY FREQUENCY TABLES TO ANALYZE CATEGORCIAL DATA........ 61

*6.3a Class Activity: Construct Two-Way Frequency Tables using Categorical Data........ 62*

*6.3a Homework: Construct a Two-Way Frequency Table........ 69*

*6.3b Class Activity: Interpret Two-Way Frequency Tables........ 73*

*6.3b Homework: Interpret Two-Way Frequency Tables........ 78*

*6.3c Class Activity: Conduct a Survey........ 82*

*6.3d Self-Assessment: Section 6.3........ 84*

***Important Note for Chapter 6***
The study of statistics can be somewhat subjective. Many of the observations and described associations are open to interpretation and rich discussion. The emphasis should be on a student's ability to make **arguments** about the data and to **support** their arguments with numerical evidence.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

# 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.

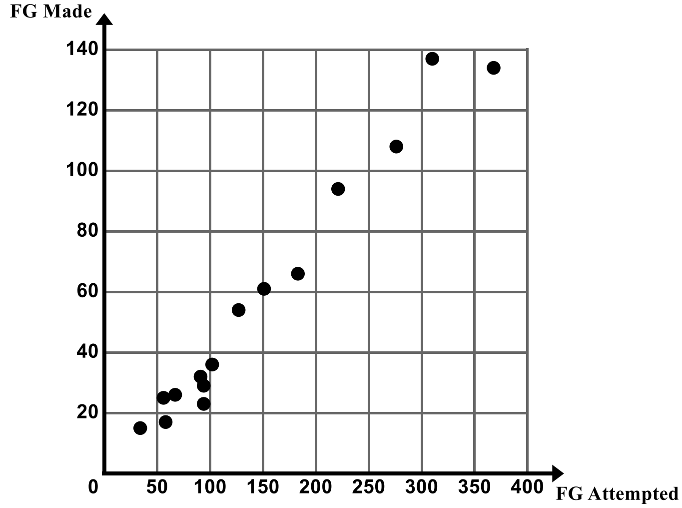

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

## 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?"<br><br>*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.<br>*Table not shown due to space.*<br><br>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.<br><br>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.<br><br>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<br><br>*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.* |

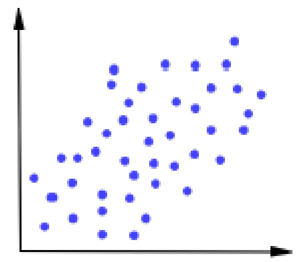


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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

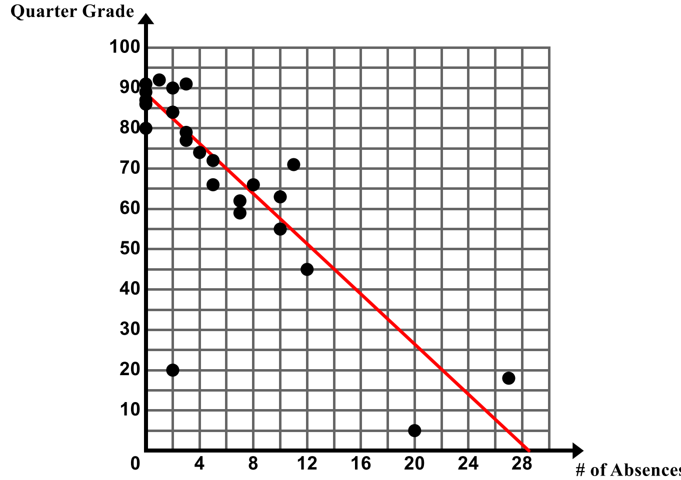


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

| | | |
|---|---|---|
| | **Attend to precision.** | The following table shows the weight of an English Mastiff from birth to age 60 weeks.<br>*Table not shown due to space.*<br><br>Create a scatter plot of the data on the grid below.<br>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.<br><br>*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.<br><br>*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?* |



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

<table>
<tr>
<td></td>
<td>Look for and express regularity in repeated reasoning.</td>
<td>
The following scatter plot shows the final grade in Ms. Ganchero’s math class for students and the number of times they are absent.<br><br>
Explain the meaning of the slope and y-intercept in the context.<br>

<br>
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.
</td>
</tr>
</table>



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

**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.

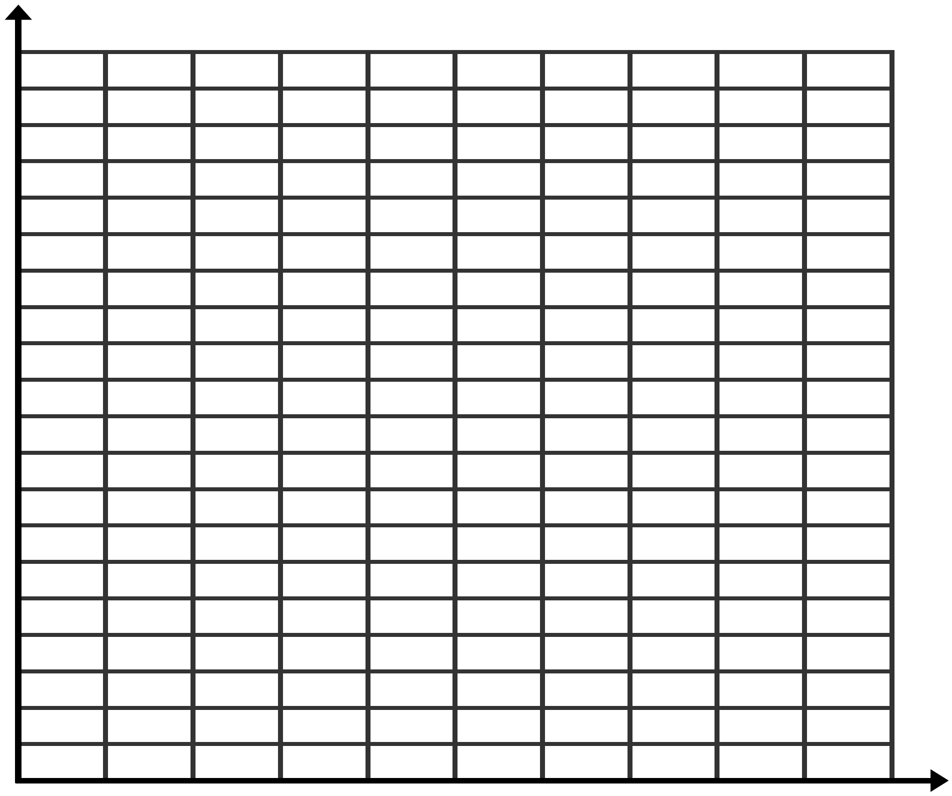
1. In the table below, record the class data for each Tongue Twister.

| 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 | | | | |



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

# 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 $7^{th}$ 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.


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

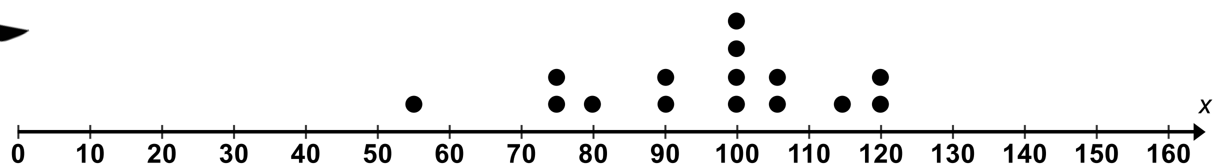
**6.1a Class Activity: Read and Interpret a Scatter Plot**

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 |

Problems 1 and 2 provide students with an opportunity to connect what they have learned in 6th/7th grade with what they will learn in 8th grade. Problem 1 is a review of 6th and 7th grade content where students learned to display and analyze **univariate** data (collections or counts of measurements of one variable). Students have learned how to describe the **shape** (normal, skewed right, skewed left), **center** (mean or median) and **spread** (mean absolute deviation and interquartile range) of univariate data. They have learned to describe areas where the data shows **clustering** and identify points that appear to be **outliers.** In 8th grade, students connect this learning to work with **bivariate** data – data that corresponds to two variables. Just as in the study of univariate data, students can describe the "shape" (this cloud of points resembles a line) of bivariate data. They can think of the "center" as a line drawn through the center of the points that captures the essence of the data and as "spread" as referring to how far the data points stray from this line (weak/strong/perfect association). Students will also observe clusters of data and outliers.

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.
Some possible observations might be: The average amount she makes is around $100. The data does not appear to be very spread out. The point 55 appears to be an outlier and may pull the average down. What could have caused this outlier? She can usually expect to make between $75 and $120 a day.

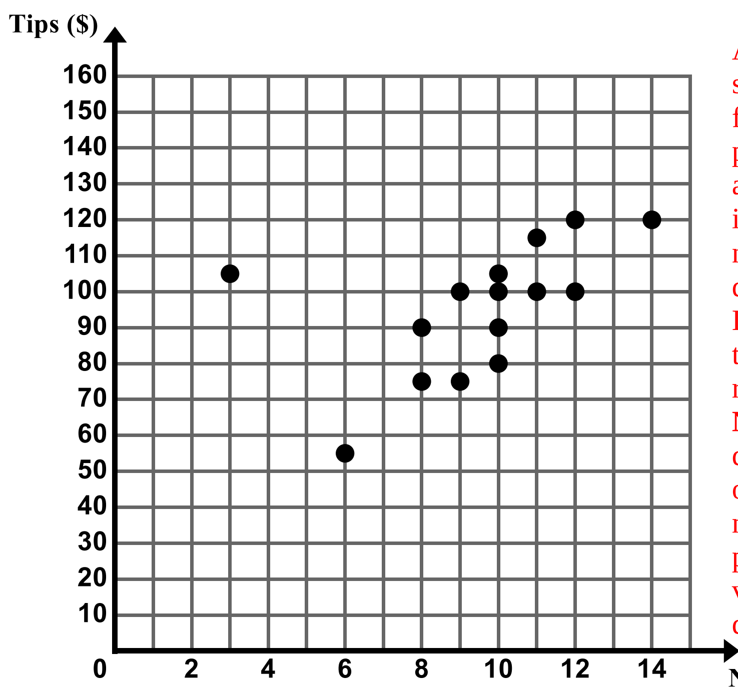


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.
Consider what different points in the scatter plot mean in the context. For example, find the point on the scatter plot that represents the day where Jenny had 10 clients and made $80 (this is the point (10, 80) on the scatter plot).




a. Make some observations about the scatter plot.

At this point, the observations students are making should be very informal. Students may observe the following: The data resembles a line. There is a positive association between number of clients and amount made in tips (i.e. as the number of clients increases so does the amount of tips). The data does not appear to be very spread out. The point $(3, 105)$ does not appear to fit with the rest of the data set. Data points that do not appear to fit with the rest of the data set are called outliers. Brainstorm possible reasons for this outlier.
**Note:** The dot plot on the previous page is displaying **univariate** data (each of the data points or dots corresponds to one variable, the amount of money made in tips on a given day). On the scatter plot to the left, each data point corresponds to two variables (**number of clients <u>and</u> tips** on a given day). This is an example of **bivariate data**.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

**Directions:** Determine if the following scenarios represent univariate or bivariate data.
Univariate data deals with a single variable. Bivariate data involves two variables. For example, in the dot plot on page 10, each data point involves a single variable – the amount of money Jenny makes in tips. This is an example of univariate data. In the example on pg. 11, each data point corresponds to two variables – the number of clients Jenny has in a day **and** the amount she makes in tips. This is an example of 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. Bivariate Data

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. Univariate Data

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.*

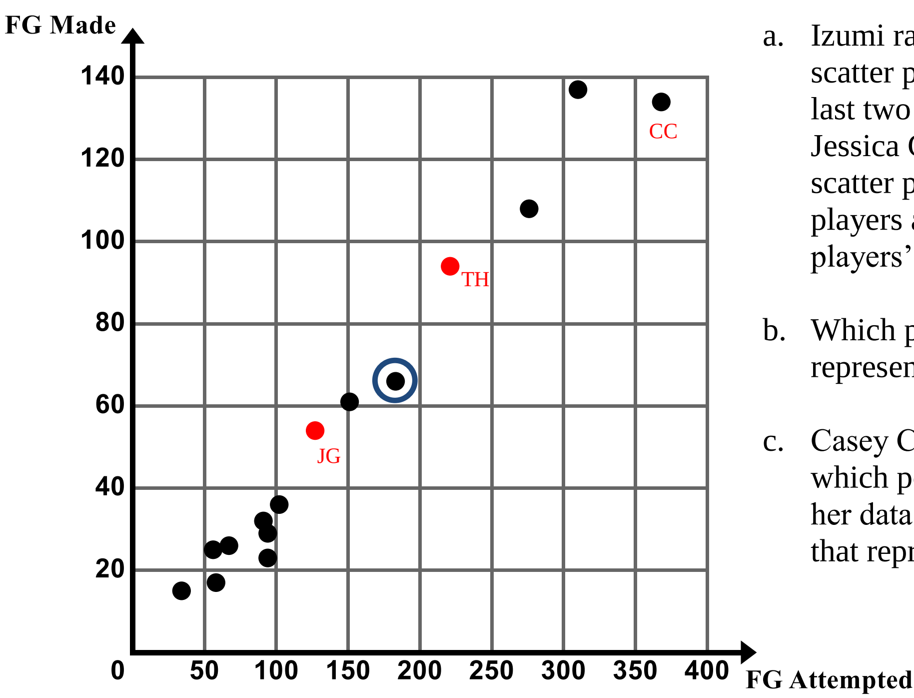

©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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 |

As you examine the data in the table, ask yourself the following questions: Who scored the most points? Is this person the best player? Who are the best players on the team and why? What are plausible reasons that some people are attempting so few shots?

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. See graph.

b. Which player does the circled data point represent? Rachel Reagan

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. See graph.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

There appears to be a positive association between the two variables: as the number of field goal attempts increases, so does the number of field goals made. Although the data does not fit on a straight line, it resembles a line with a positive slope. There appears to be a cluster of points in the domain of 50 to 100, meaning several players attempted between 50 and 100 shots.

e. Can you think of another variable that when graphed with field goals made would have a positive association? Answers will vary. Possible answers playing time, time spent practicing, height, etc.

Important Vocabulary:

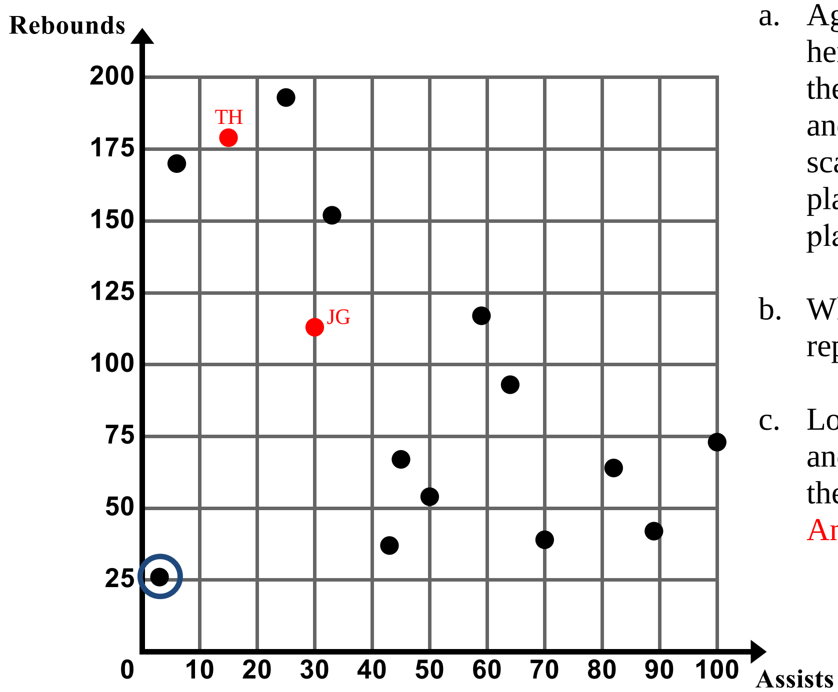
- An **experiment** is *any process or study that results in the collection of data.* Izumi is conducting an experiment to determine if there is a relationship between number of field goals attempted and number of field goals made.
- The **sample space** is the *set of all possible outcomes of a particular experiment (in Izumi's case the sample space is the data gathered from her team). Izumi is gathering data on two* ***random variables*** *(number of field goals attempted (x) and number of field goals made (y)). A random variable is a variable that takes on different values as a result of the outcomes of an experiment.*
- A **realization or observation** is the specific value that a random variable may assume. The data point (102, 36) represents a specific value for the random variables, which happens to correspond to the player Monique Ortiz.

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. Again, you can review statistics terminology: experiment, sample space, random variable, realization. You can also discuss why this is bivariate data.

| 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 |


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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. See graph.

b. Which player does the circled data point represent? Juanita Martinez

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

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.
Player moved to the school after the season had started and joined the team late. Player did not have a lot of playing time due to skill or injury. Player shot from the outside perimeter a lot so did not assist or rebound very much.

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.
Allow students to articulate what they see in the graph. Surface the following ideas: Although the data does not fit on a straight line, it resembles a line with a negative slope. There appears to be a negative association between the two variables: as the number of assists increases, the number of rebounds decreases. A plausible reason for this is the position being played – a person making assists is less likely to be close to the basket for a rebound.

f. Can you think of another variable that when graphed with field goals made would have a negative association? Answers will vary. Possible answers # of games missed, amount of time spent on bench

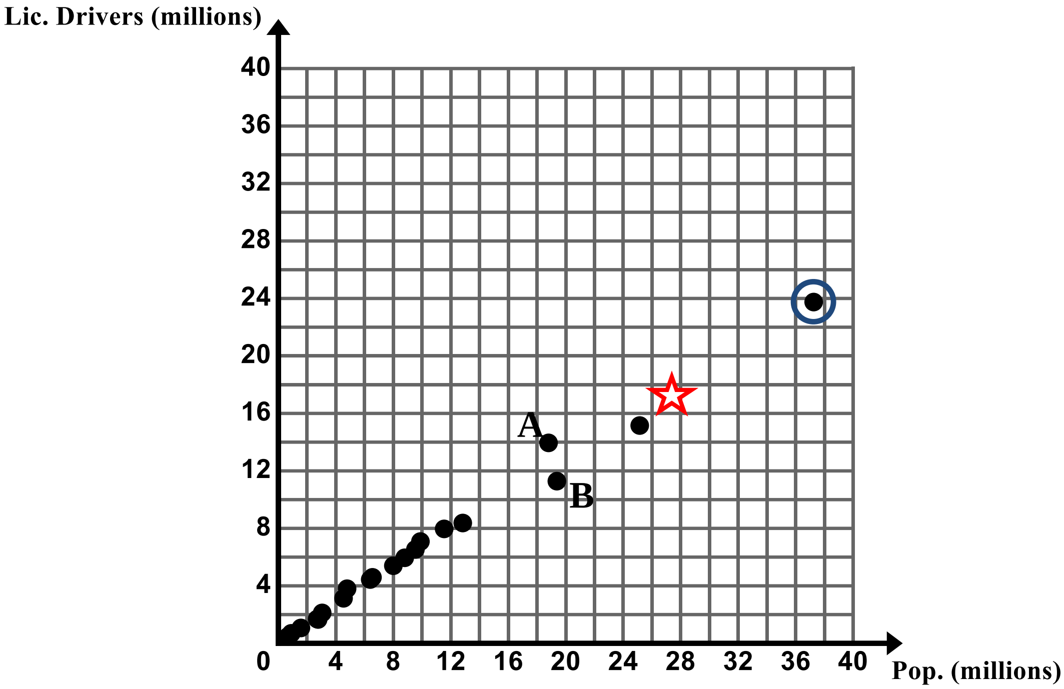
10. Which data set appears to have a stronger association: the relationship between number of field goal made and number of field goal attempts or the relationship between number of rebounds and number of assists?



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

**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. See graph.

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

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? Approximately 18 – 18.5 million people

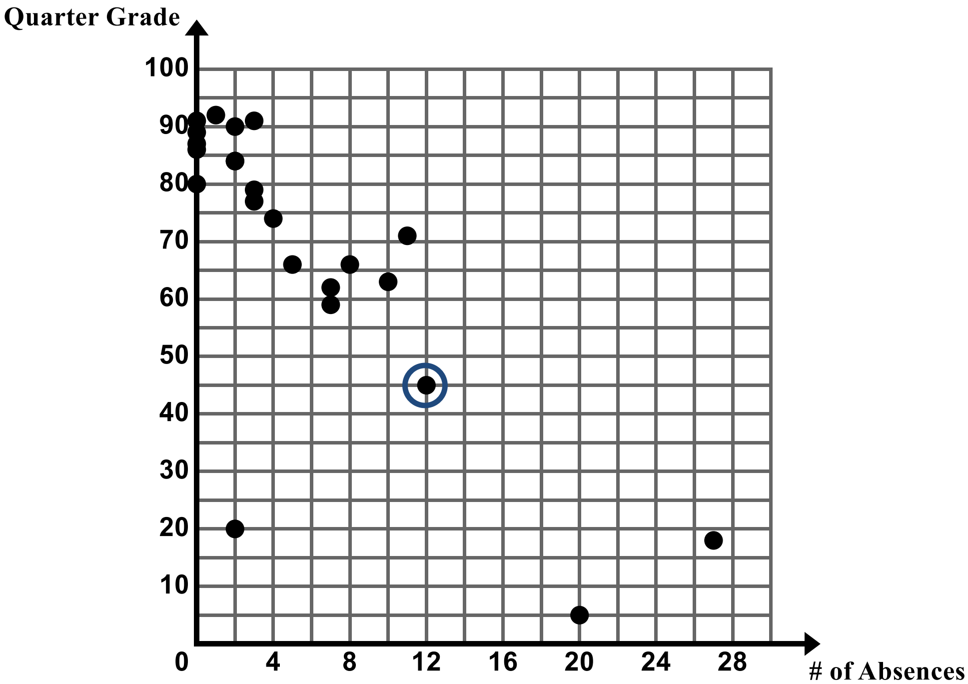
f. Compare data points A and B.

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. Plausible explanations may include: The public transit in NY is very good so people don't need cars as much. New York roads are more congested so driving is not a great way to get around. New York is less spread out than Florida. Parking is more expensive in New York.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.
Most students do not miss that much school so it is reasonable that we would see a cluster in the domain of 0 – 3 absences.

d. Do there appear to be any outliers in the data? If yes, what are they? Provide an explanation for the outlier(s). The point (2, 20) appears to be an outlier. A student who was absent only 2 times received a final grade of 20. Some plausible reasons for this – the student did not do his/her homework, the student did not pay attention in class.

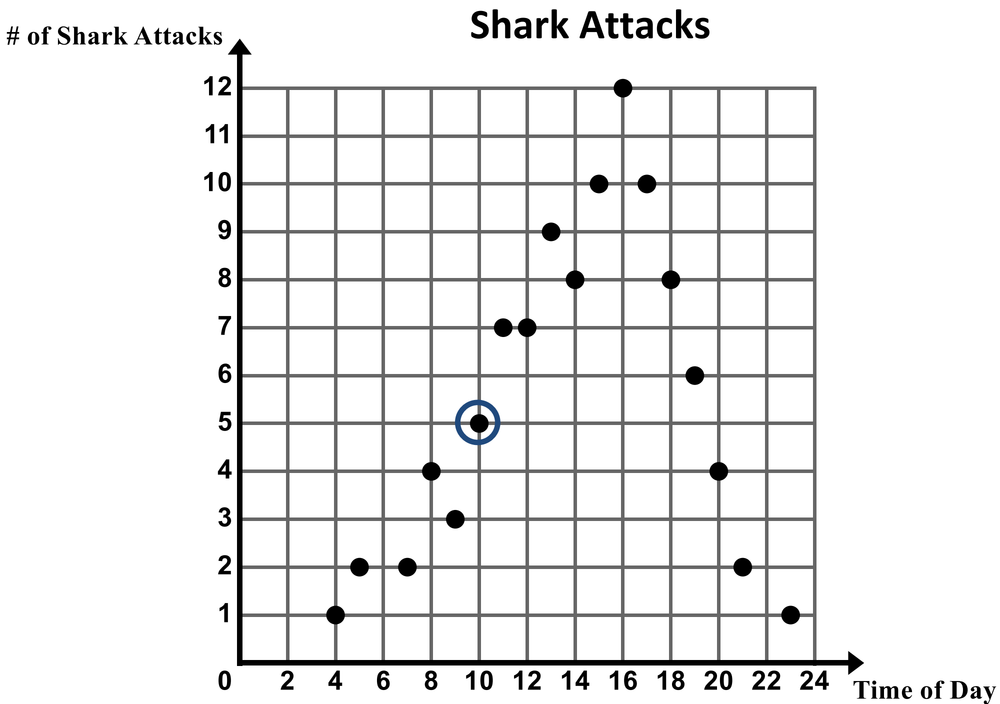
e. Does the scatter plot suggest a relationship between absences and grade? Describe any trends or patterns you observe in the data.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

While the data does show a pattern, the pattern is **non-linear** or curved. As the time of day increases over the interval from 0 to 16:00 hours the number of shark attacks also increases. As the hours increase over the interval from 16:00 to 23:00 hours the number of shark attacks decreases. A possible explanation is that as the day progresses the temperature gets warmer and more people go to the beach and get in the water. Then as the temperature begins to cool down less people will be in the water. The more people in the water the greater the likelihood of someone being attacked by a shark.

**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.**



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

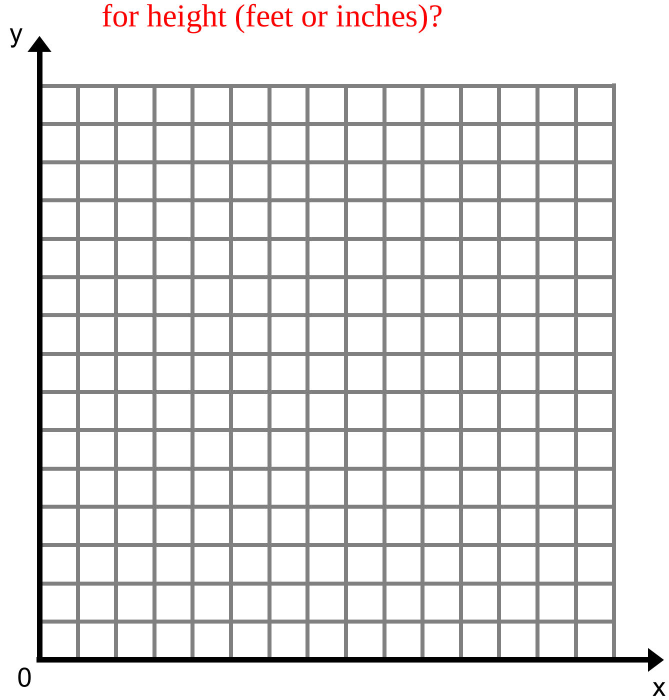
## 6.1b Class Activity: Create and Analyze a Scatter Plot

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. Hint: Come up with a strategy for how to create this graph depending on your data set. Review key graphing concepts: Which variable will be our independent/dependent? How should we scale the graph? What unit of measure should we use for height (feet or inches)?




   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.

   Likely, your data will show a positive linear association between these two variables. Plausible reasons for outliers may be someone with a larger shoe size that has not gone through their growth spurt yet.



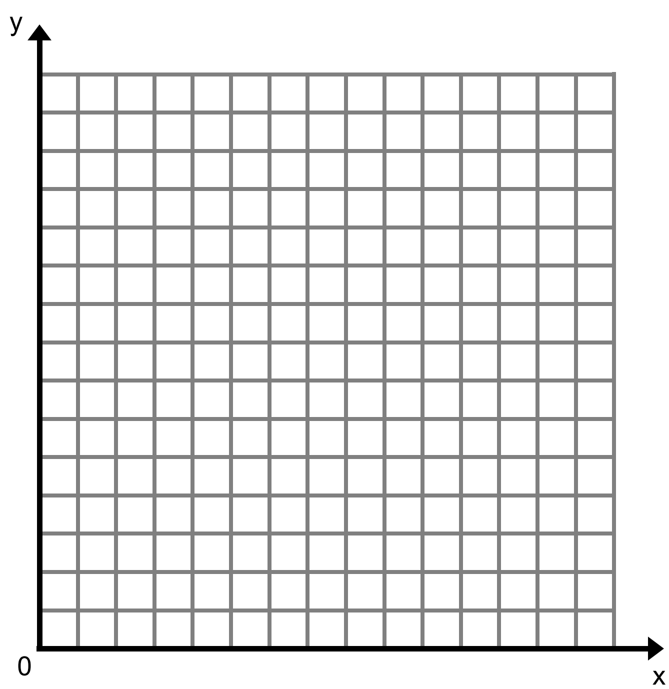
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

   Answers will vary depending on your data. Likely your data will show that there is no apparent association between the number of letters in a person's first name and the number of letters in their last name.


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

## 6.1b Homework: Create and Analyze a Scatter Plot

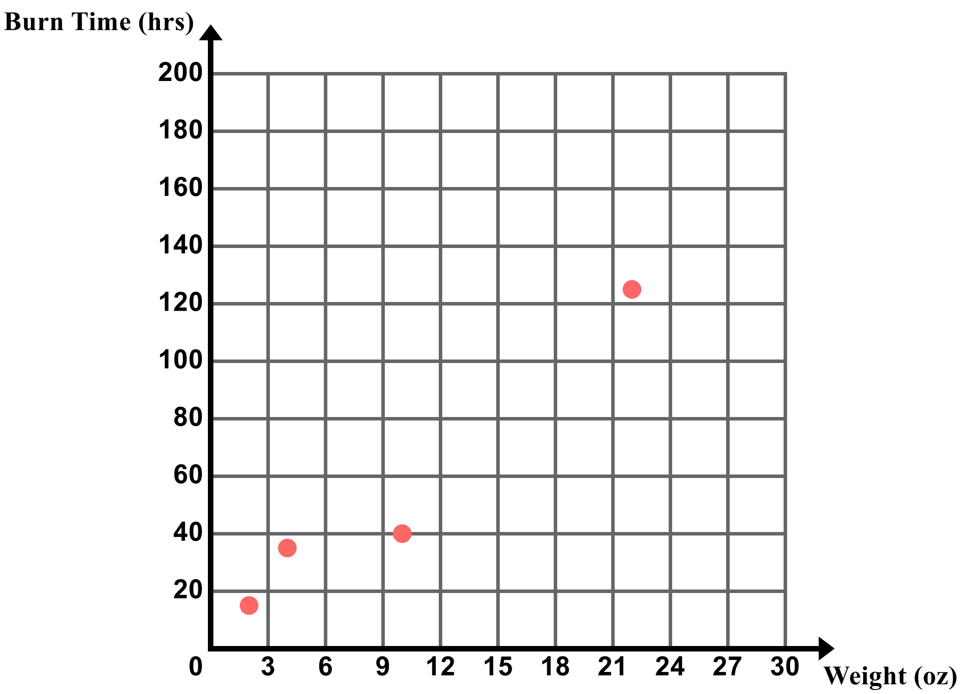
1. Is there an association between the weight of a candle and the amount of time it burns?
   a. Make a prediction.

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.

The points (2, 15), (4, 35), (10, 40), and (22, 125) have been graphed for students as an example. Students should graph these points and the rest of the points.




   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.

This indicates a strong positive linear association. As the weight of the candle increases the amount of time it burns also increases. There is a cluster of data in the lower corner, perhaps many candles made are between 2 and 5 ounces in weight.

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

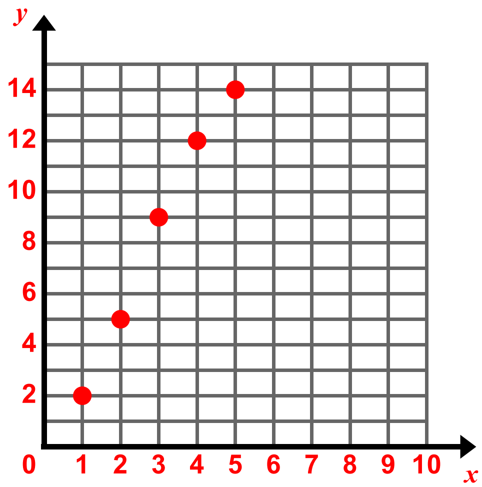


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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

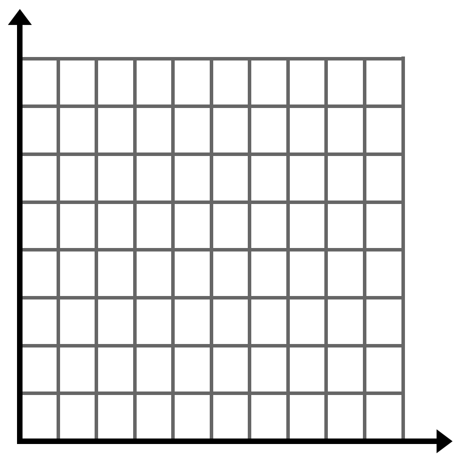
a.

| $x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y$ | 2 | 5 | 9 | 12 | 14 |



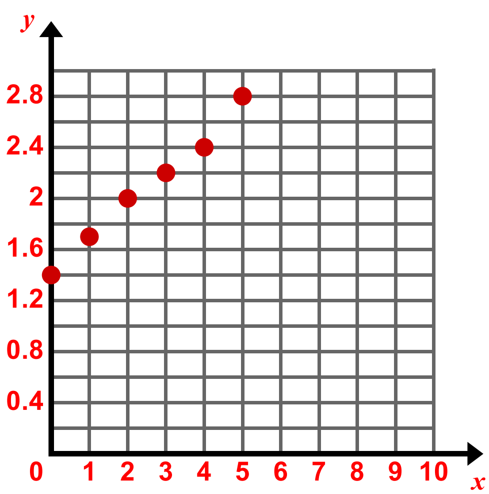

b.

| $x$ | 0 | 2 | 2 | 4 | 4 | 5 | 6 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 5 | 6 | 5 | 5 | 7 | 6 | 4 | 6 | 5 | 6 |



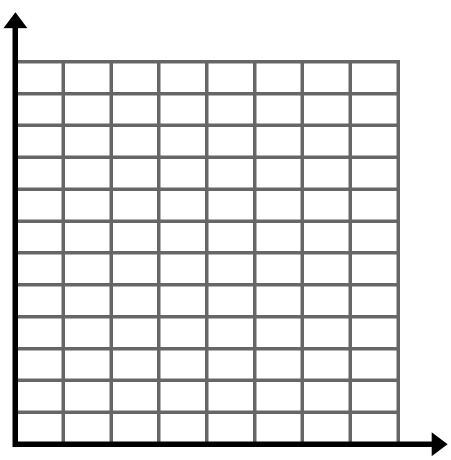
c.

| $x$ | 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 |





©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

**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**<br>Graphs will vary. See #7 and #9 on pg. 24 for sample plots. | 2. **Negative Linear Association**<br>Graphs will vary. See #6 and #10 on pg. 24 for sample plots. |
| 3. **No Apparent Association**<br>Graphs will vary. See #8 on pg. 24 for a sample plot. | 4. **Nonlinear Association**<br>Graphs will vary. See #5 on pg. 24 for a sample plot. |

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:

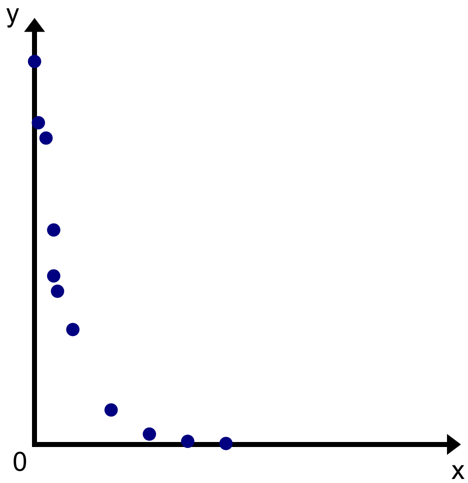
- **Clusters -** A cluster is a set of points that are in close proximity to each other.
- **Outliers -** An outlier is a data point that noticeably stands out from the general behavior of the data set.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

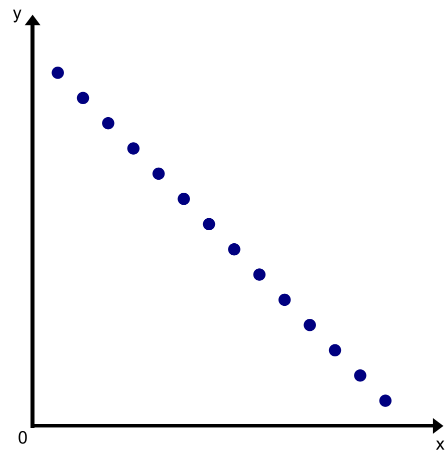
**Directions:** Describe the association between *x* and *y* using the terms from the previous page. Circle any clusters in the data. Put a star by any points that appear to be outliers.

5.



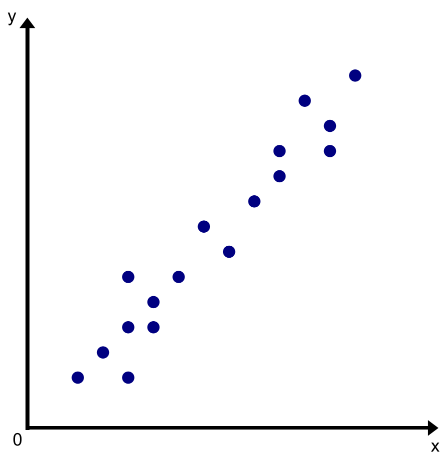

non-linear association

6.



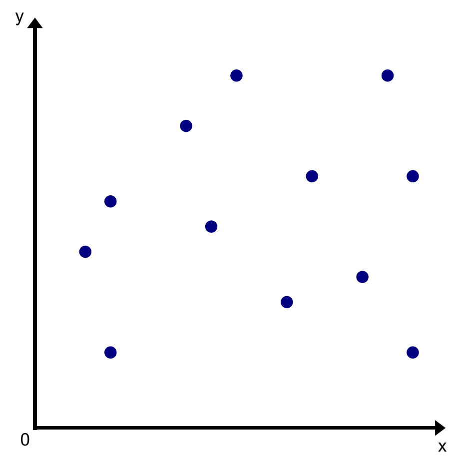

perfect negative linear association

7.



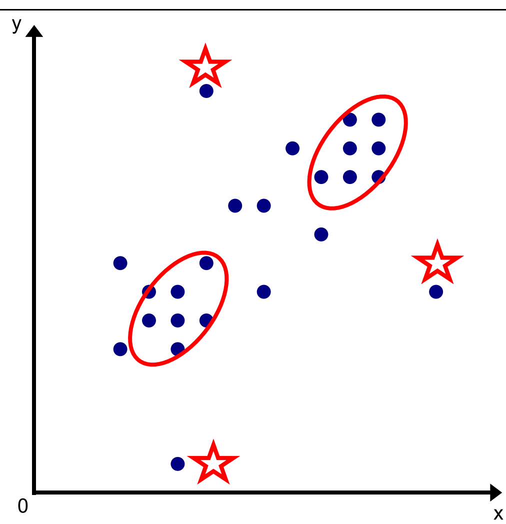

strong, positive linear association

8.



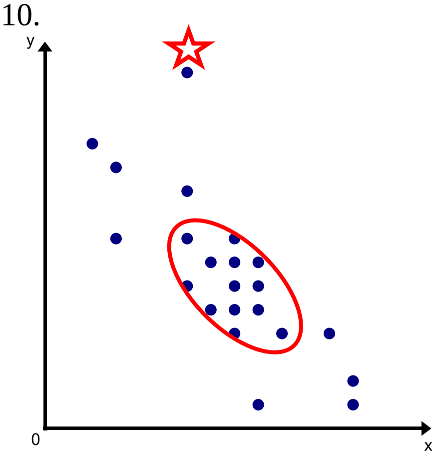

no apparent association

9.




positive linear association that is moderate

10.



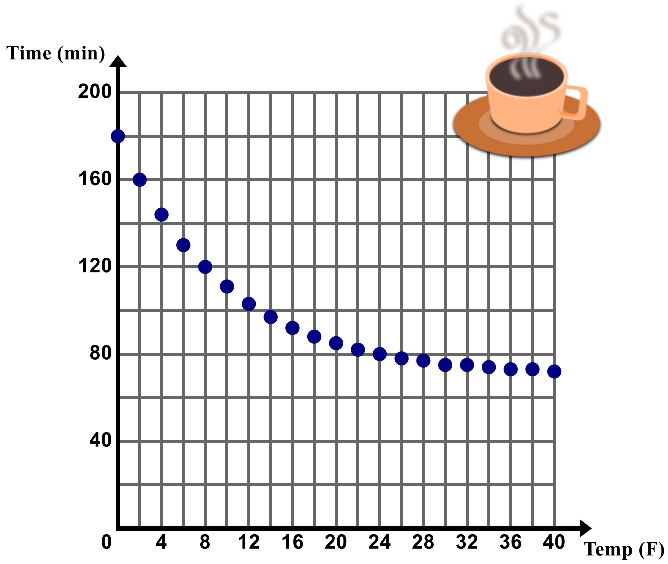

negative linear association that is moderate



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

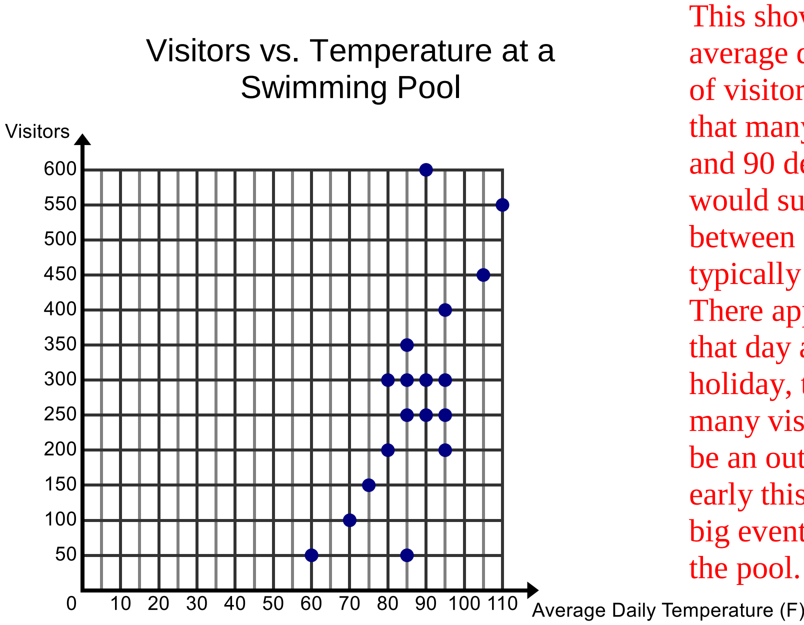
**Directions:** Examine the following scatter plots. Describe the association between the two variables. Circle any clusters in the data. Put a star by any points that appear to be outliers. Use the context to give possible explanations as to why these trends, patterns, and associations exist.

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.




Non-linear association. When the cup of tea is initially poured its temperature decreases rapidly at first and then the temperature decreases at a slower rate. The temperature of the tea drops until it reaches the temperature of the room which 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.

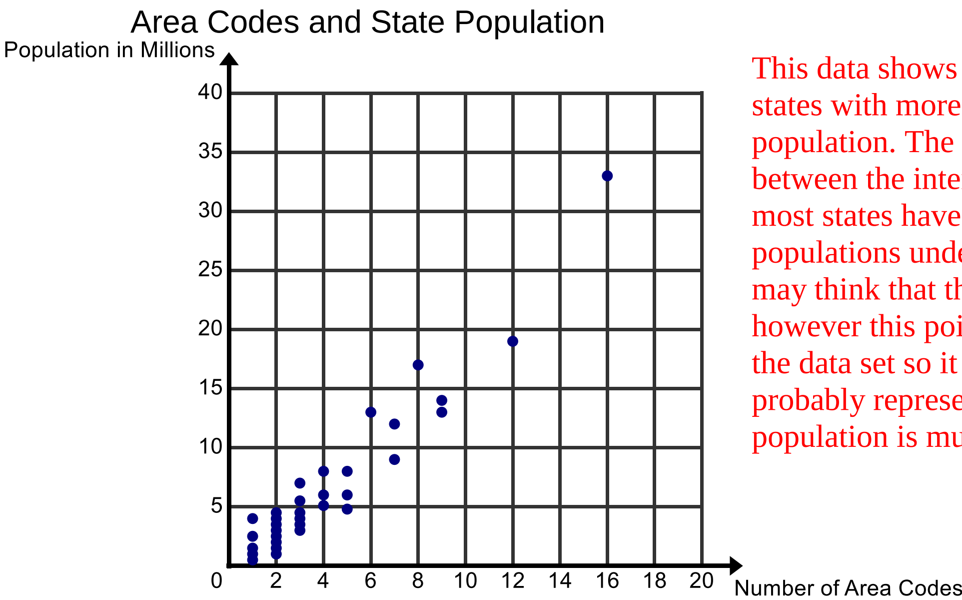



This shows a positive linear association - as the average daily temperature increases the number of visitors to the pool also increases. It appears that many of the data points cluster between 70 and 90 degrees and 200 to 300 visitors. This would suggest that the pool regularly has between 200 to 300 people and that people typically visit the pool in this temperature range. There appears to be an outlier at (90, 600). On that day admission was half off and it was also a holiday, that would explain why there where so many visitors. Also the point (85, 50) appears to be an outlier as well – maybe the pool closed early this day for cleaning or maybe there was a big event in town that drew people away from the pool.



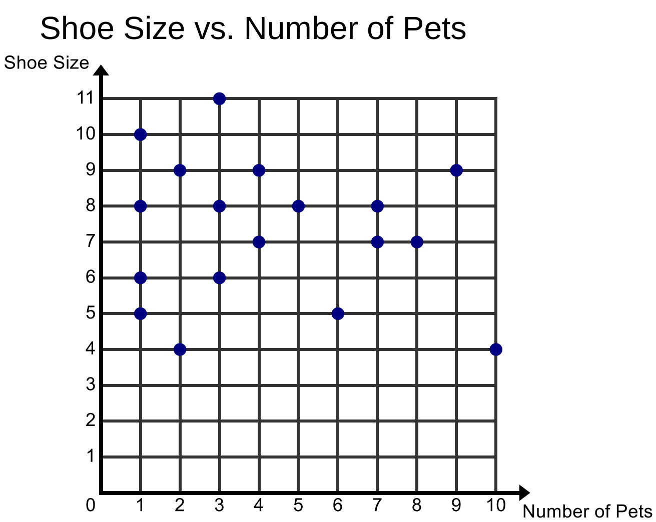
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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




This data shows a positive linear association, states with more area codes have a higher population. The majority of the data is clustered between the intervals of 1 and 5 suggesting that most states have between 0 and 5 area codes and populations under 15 million people. Students may think that the point (16, 34) is an outlier; however this point still fits the general trend of the data set so it is not an outlier. This point probably represents California since its population is much greater than other 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.




This scatter plot strongly that there is no association between the number of pets a person has and their shoe size. It also shows that most people surveyed had one pet.

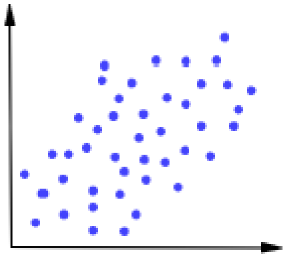


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

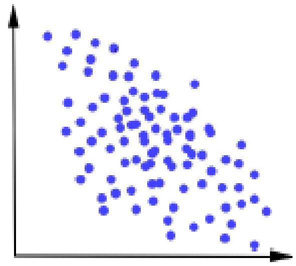
**6.1c Homework: Patterns of Association**

**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.

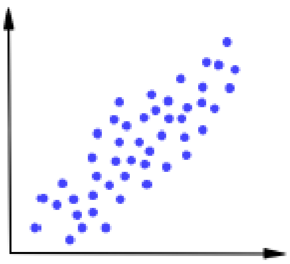
1.



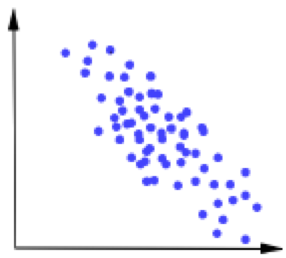
2.



3.

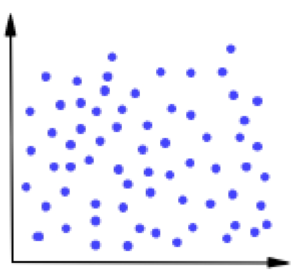


4.



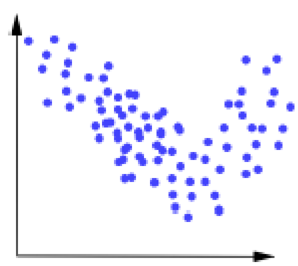
strong negative linear association

5.

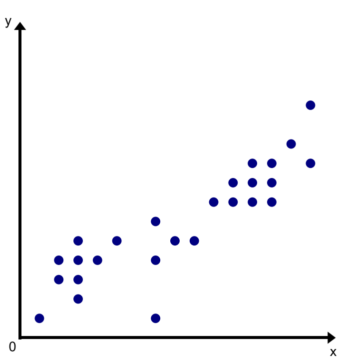


no apparent association

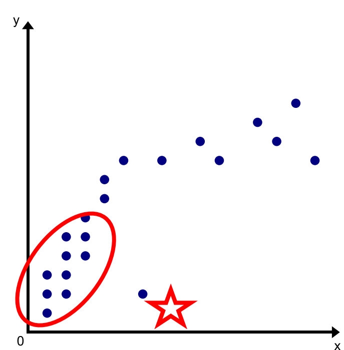
6.



7.




8.




non-linear association

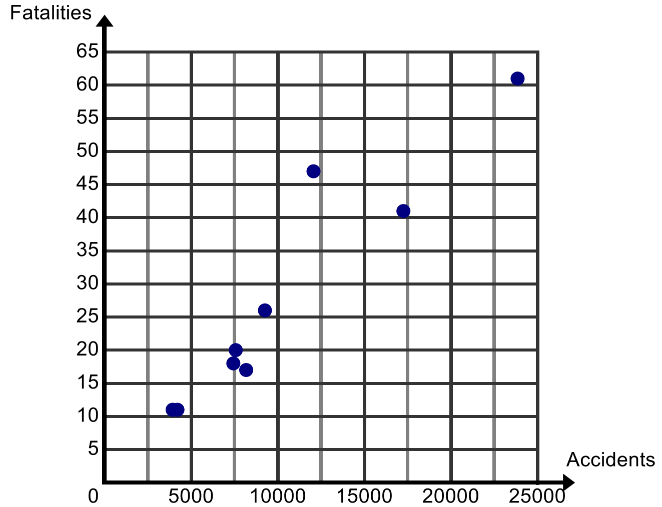


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

**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.

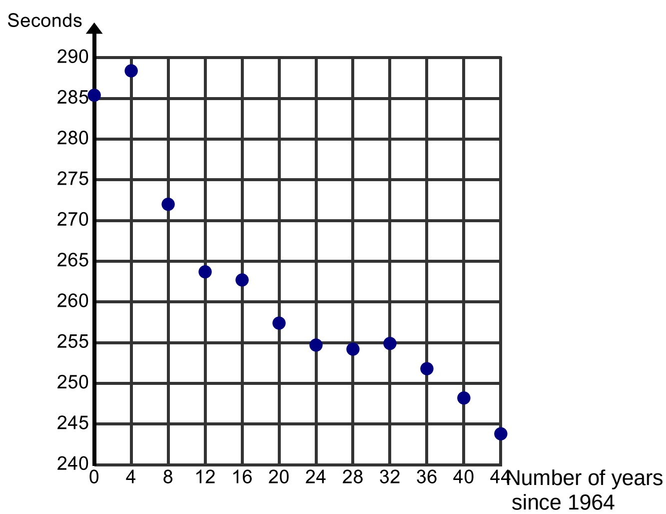
9. For Heidi's Driver's Education class, she finds data about the number of car accidents and fatalities (deaths) from car accidents for teens in the Western United States.

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)

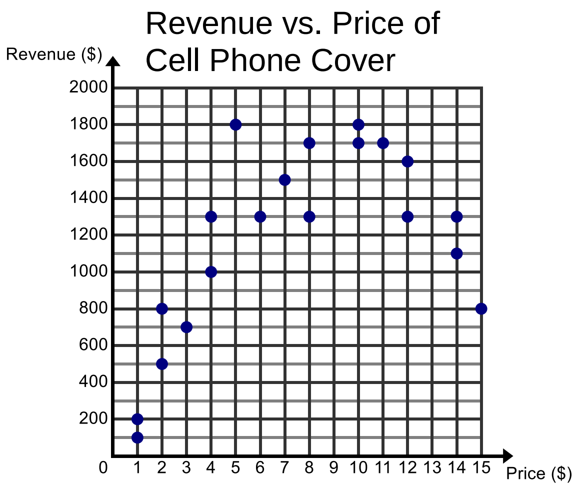



**Bonus:** Research, collect, and analyze Olympic data for other events that interest you.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.




This scatter plot shows a non-linear association. Students may think that the point (5, 1800) appears to be an outlier but this is questionable – there is not really enough data to tell. This graph shows that the optimal price to charge for a cell phone cover is around $10. You may choose to further discuss this plot with students – if you don't charge very much for a cover, you may sell a lot of covers but not make as much in revenue because you are not charging very much. If you charge too much for a cover you will not sell as many so will not make as much. There is a selling price that optimizes the amount you make.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

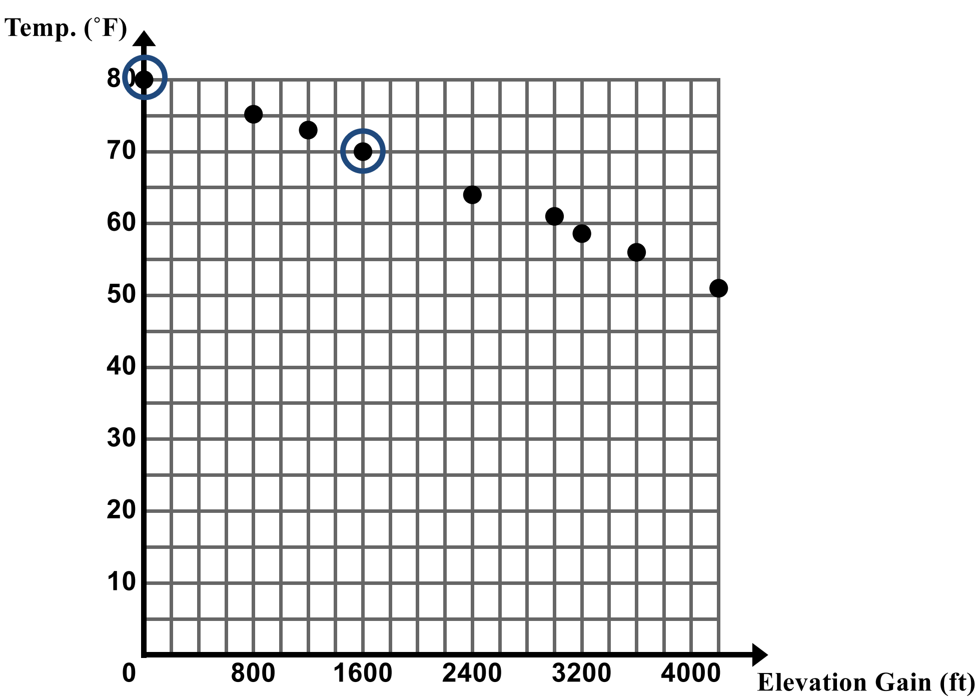
**6.1d Self-Assessment: Section 6.1**

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.

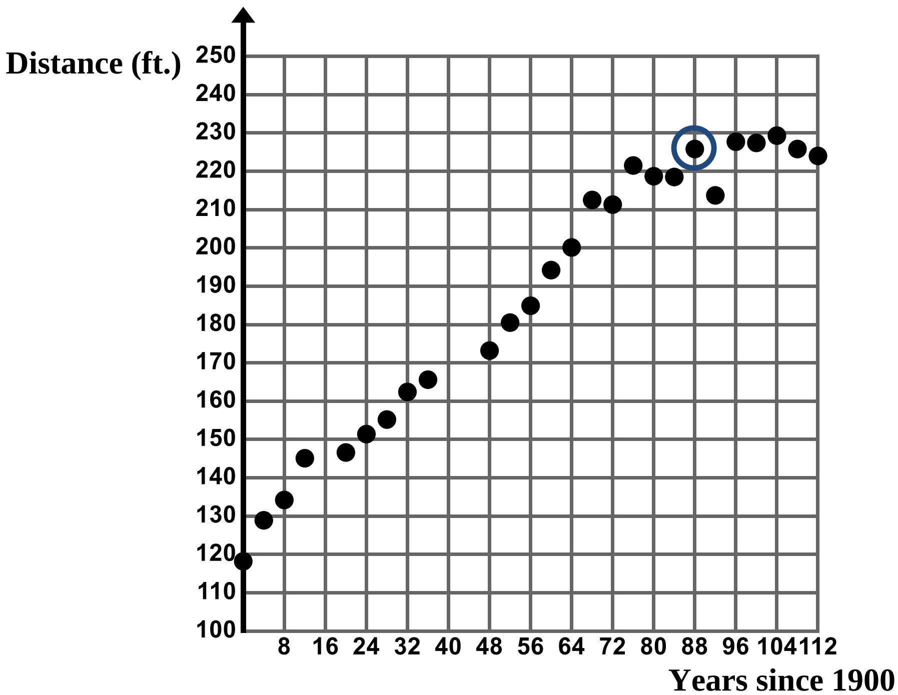





©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

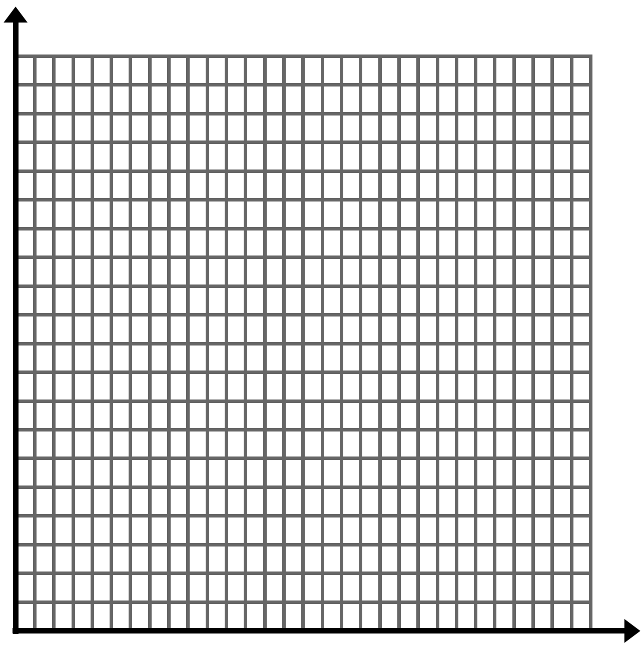





©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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 |





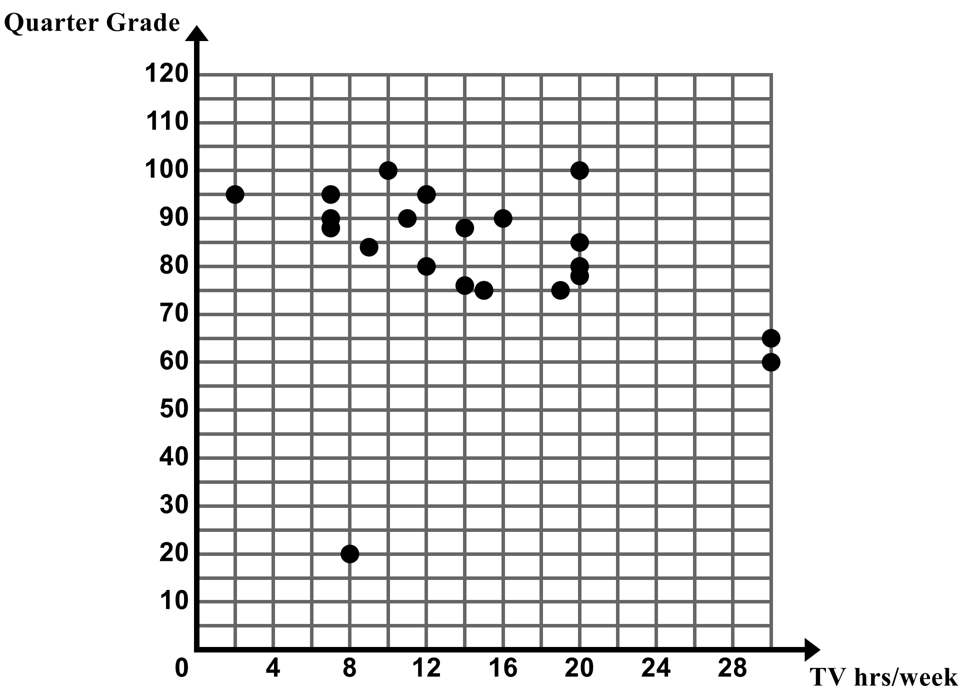
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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?






©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

# 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.

In this section, talk to students about strategies for drawing a line of best fit. The goal is to draw a line that best approximates the data. Sometimes it helps to think of the points as a cloud of points - the goal is to draw a line that captures the essence of the shape of this cloud. It may pass through some of the points, all of the points, or none of the points. Students can use a strand of uncooked spaghetti to help them to determine where to place the line of best fit. Talk to students about how we can assess the fit of the line we drew – check to see how closely the points are packed around the line. For the purposes of writing an equation for this line of best fit, it sometimes helps to have the line pass through two integer points; however this is not necessary. Encourage students to use multiple points to determine the prediction function – use two points that are close together and then choose two points that are farther apart and compare. Keep in mind throughout this chapter that the line of best fit will depend upon the method used to find it, and will vary from student to student, so prediction functions and predictions will vary from the key.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

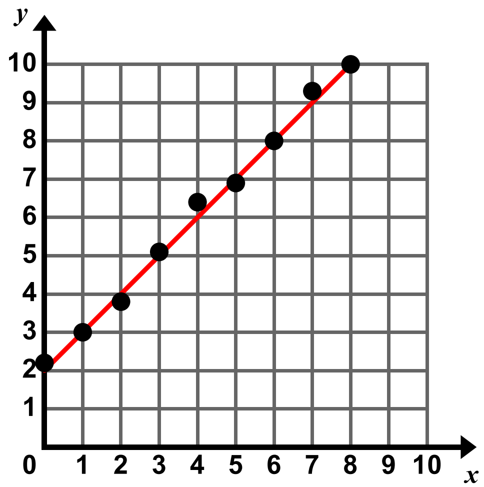
**6.2a Classwork: Lines of Best Fit**

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:

Strong positive linear association

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

See notes in red on pg. 34.

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

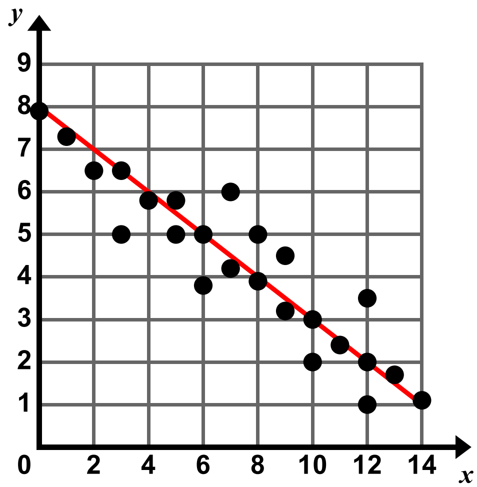
$m \approx$ __1___ $b \approx$ __2___

One way to find the slope and $y$-intercept of the line of best fit is to eyeball it from the graph. You may also choose 2 points on or close to the line and use these points to find the slope and $y$-intercept. For example, you may use the points (0, 2) and (5, 7). For additional help on finding the slope from 2 points, refer to chapters 2 and 3.

d. Write a prediction function for the data set. $y \approx x + 2$

e. Use your prediction function to find the value of $y$ when $x = 12$ and when $x = 100$. $y \approx 14$ and $y \approx 102$ repectively

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.

$m \approx$ __$-\frac{1}{2}$___ $b \approx$ __8___

Note: In this problem, it seems reasonable to use the points (0, 8) and (6, 5) to find the slope and $y$-intercept of the line of best fit.

d. Write a prediction function for the data set. $y = -\frac{1}{2}x + 8$

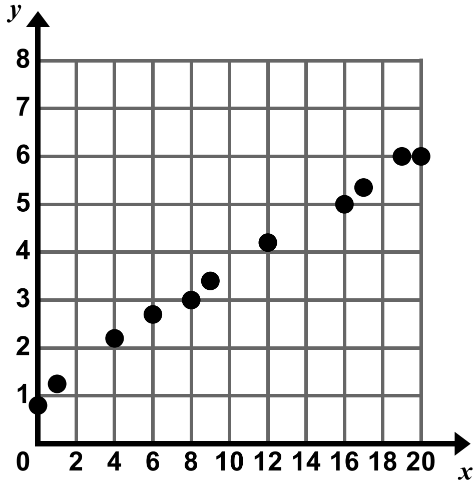
For help on writing the equation of a line, refer to Chapter 3.

e. Use your prediction function to find the value of $y$ when $x = 12$ and when $x = 100$. $y \approx 2$ and $y \approx -42$ repectively



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

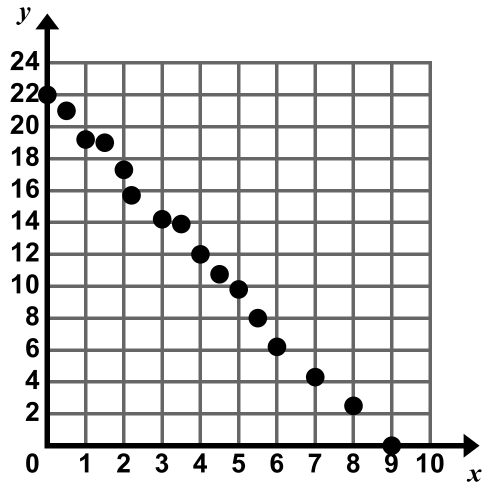
c. Estimate the slope and *y*-intercept of your line.

*m* ≈ _____ *b* ≈_____

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* ≈ _____ *b* ≈ _____

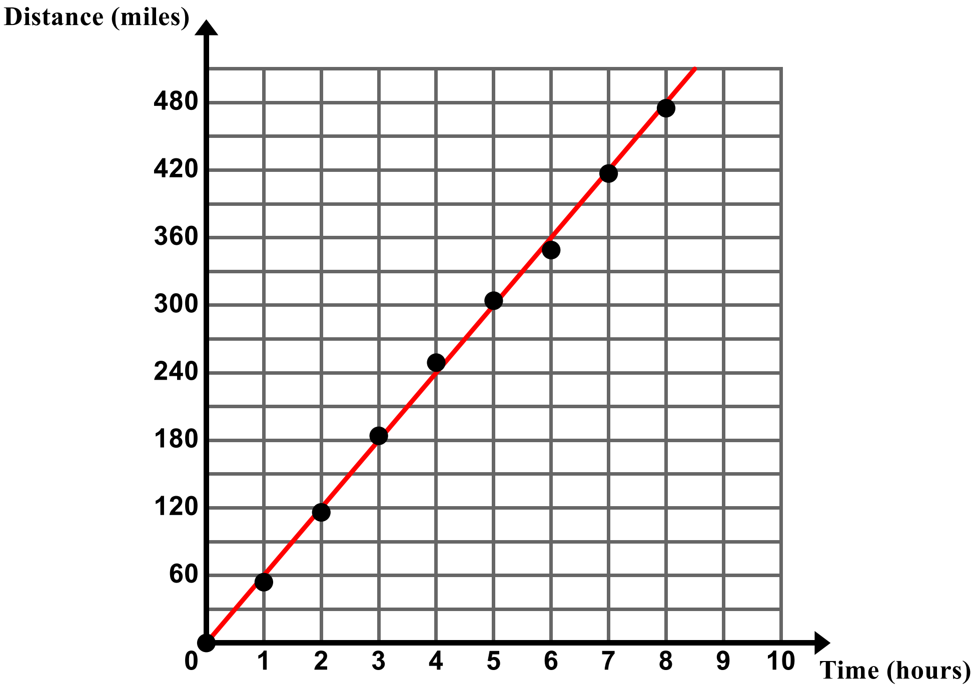
d. Write a prediction function for the data set.

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



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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. Be sure to pay attention to the scale of the graph. Remember, to find the slope of the line, determine the $\frac{\text{rise}}{\text{run}}$.

$m \approx$ __60___ $b \approx$ __0___

c. Write a prediction function for the data set.
$d \approx 60t$ (Remember, the equation of a line is $y = mx + b$ where $m$ represents the slope of the line and $b$ represents the $y$-intercept.)

d. What does the slope represent in the context?
The average speed of the trip is 60 mph. Consider the units associated with the rise and the run. The units associated with the rise are on the $y$-axis (distance in miles) and the units associated with the run (time in hours) are on the $x$-axis. $\frac{\text{rise}}{\text{run}} = \frac{\text{miles}}{\text{hour}}$. The units will help to interpret the slope in context.

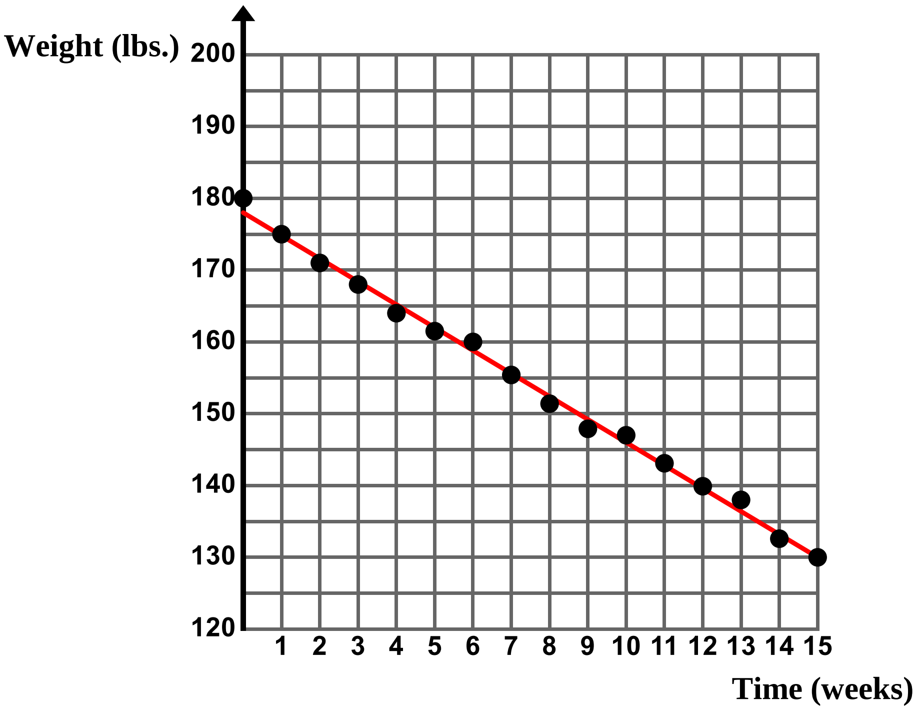
e. What does the $y$-intercept represent in the context? At time 0, Camilo and his family had not traveled any distance – they had not started their trip.

f. Predict how far Camilo and his family will have driven after 10 hours if this trend continues.
Approximately 600 miles



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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?
Hint: Write the units for the $y$ and $x$ values as rise over run to interpret the slope. The unit on the $y$-axis is pounds and the unit on the $x$-axis is weeks. See #5 part d for an additional example.

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

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

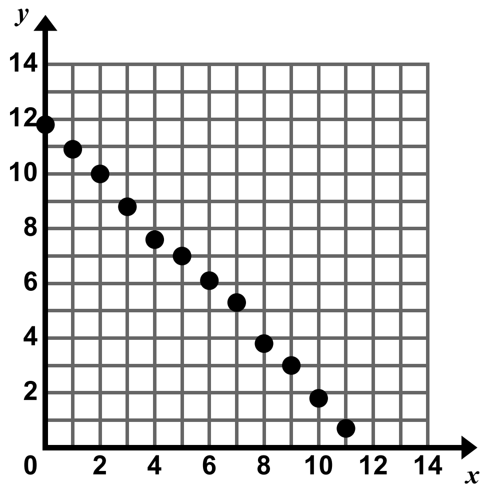


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

## 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.

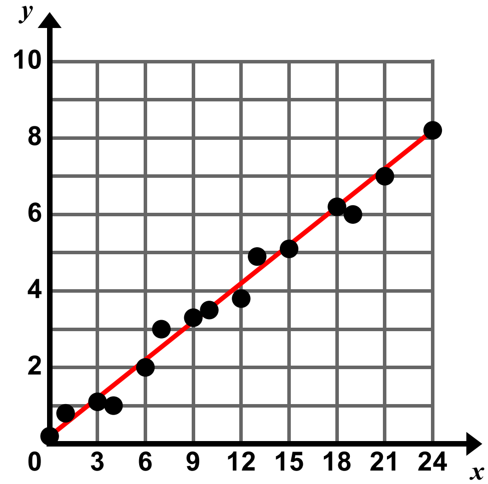
c. Estimate the slope and *y*-intercept of your line.

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

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:
Strong positive linear relationship

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

See notes in red on pg. 34 for help with drawing a line of best fit.

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

Students can eyeball the slope and *y*-intercept from the graph. Alternatively, students can use 2 points that fall on or near the line. For example, in this problem, students may use the points (0, 0) and (15, 5) to find the slope. For help on how to find the slope of a line, refer to chapters 2 and 3.

$m \approx$ __$\frac{1}{3}$__ $b \approx$ _0.2____

d. Write a prediction function for the data set.

$$y \approx \frac{1}{3}x + 0.2$$

Remember the equation of a line is $y = mx + b$ where $m$ is the slope and $b$ is the $y$-intercept.

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

To determine the value of *y* when *x* = 12, substitute in 12 for *x* into the prediction function from part d and solve.

$y \approx 4.2$ and $y \approx -33.5$ repectively

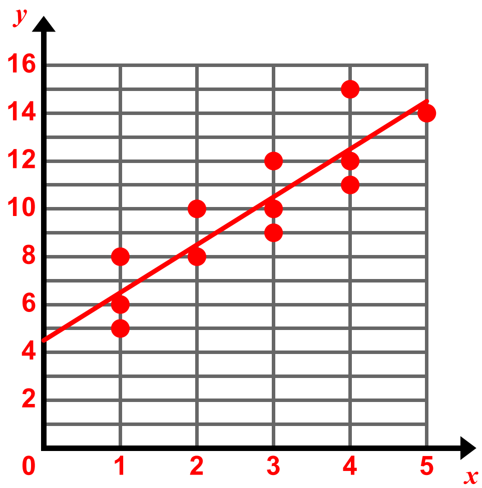


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

$m \approx$ __2___ $b \approx$ __4.5___

d. Write a prediction function for the data set. $y \approx 2x + 4.5$

Remember to write the equation for a line, we must find the **slope** and ***y*-intercept**.

**Slope:** To find the slope of a line, determine $\frac{\text{rise}}{\text{run}}$ from the graph or choose two points and use the slope formula $\frac{y_2-y_1}{x_2-x_1}$. We can use any two points that are on our line of best fit or near it. For ease of calculation, it often makes sense to use two integer points (points on the corner of the grid). For this problem, we might use the points $(4, 12)$ and $(1, 6)$. Using these points and the slope formula above, $\frac{6-12}{1-4} = \frac{-6}{-3} = 2$. What if we used the points $(1, 6)$ and $(5, 14)$? We would still end up with a slope of 2. It is important to remember that your answer might not match the key exactly – remember, a line of best fit is an estimate – it captures the essence of the data.

***y*-intercept:** One way to find the $y$-intercept is to just estimate it from the graph. Our line of best fit crosses the $y$-axis at approximately 4.5. If we use the two points $(4, 12)$ and $(1, 6)$ from above, we can solve for $b$. Using the point $(4, 12)$ and the slope we calculated above (2):

$y = mx + b$
$12 = 2(4) + b$
$12 = 8 + b$
$4 = b$

Notice that when we estimate the $y$-intercept from the graph, we get 4.5 whereas when we solve for the $y$-intercept using two points, we get 4. Either answer is acceptable. Remember the line of best fit is an estimate of the data. Your answers should be close to the answer given but will not always be exactly the same.

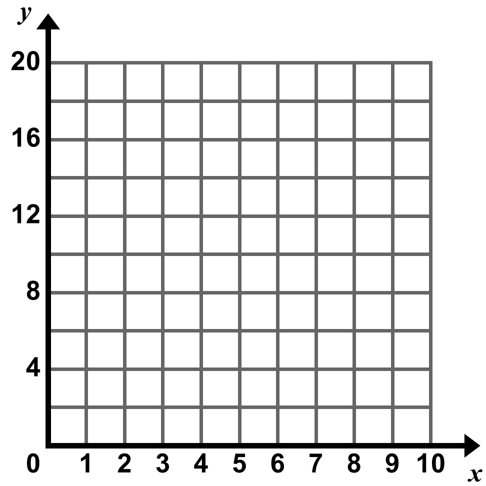


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

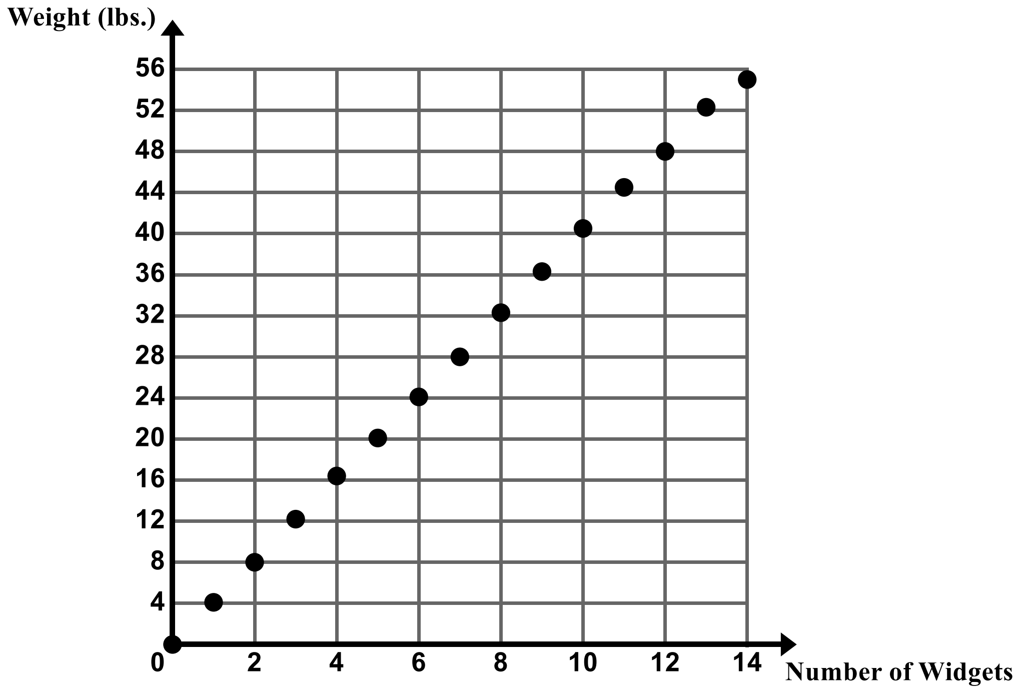
$m \approx$ _______ $b \approx$ _______

d. Write a prediction function for the data set.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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? See #5d in the class activity for help

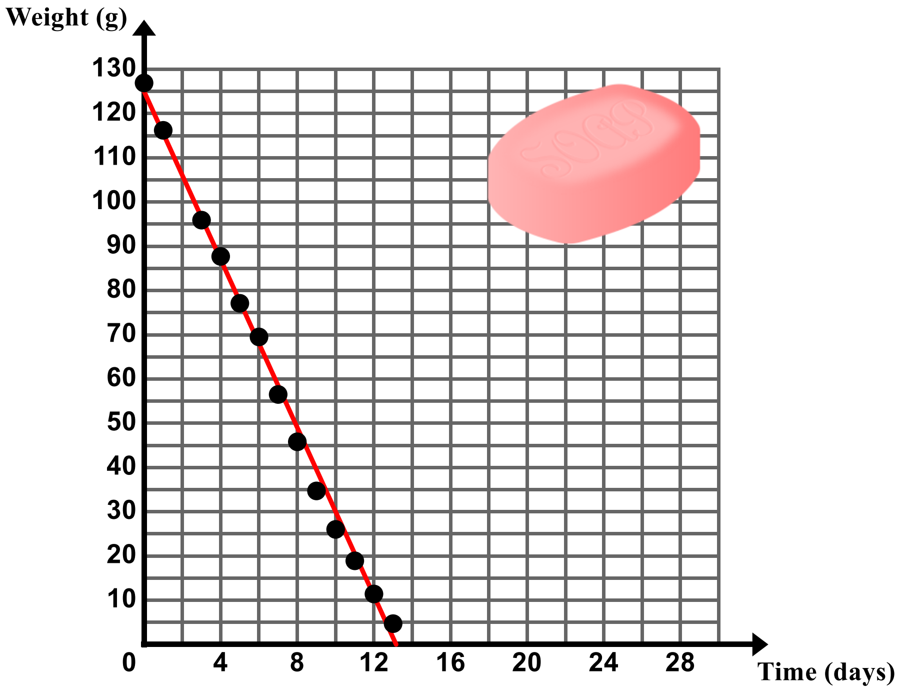
e. What does the $y$-intercept represent in the context? See #5e in the class activity for help

f. Predict the weight of 50 widgets.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

$m \approx$ __-9.5___ $b \approx$ __125___

Remember, your slope and *y*-intercept may vary slightly from the answers given.

c. Write a prediction function for the data set. $y \approx -9.5x + 125$

d. What does the slope represent in the context? Each day, Chad's family uses approximately 9.5 g. of soap

e. What does the *y*-intercept represent in the context? A new bar of soap weighs approximately 125 g.

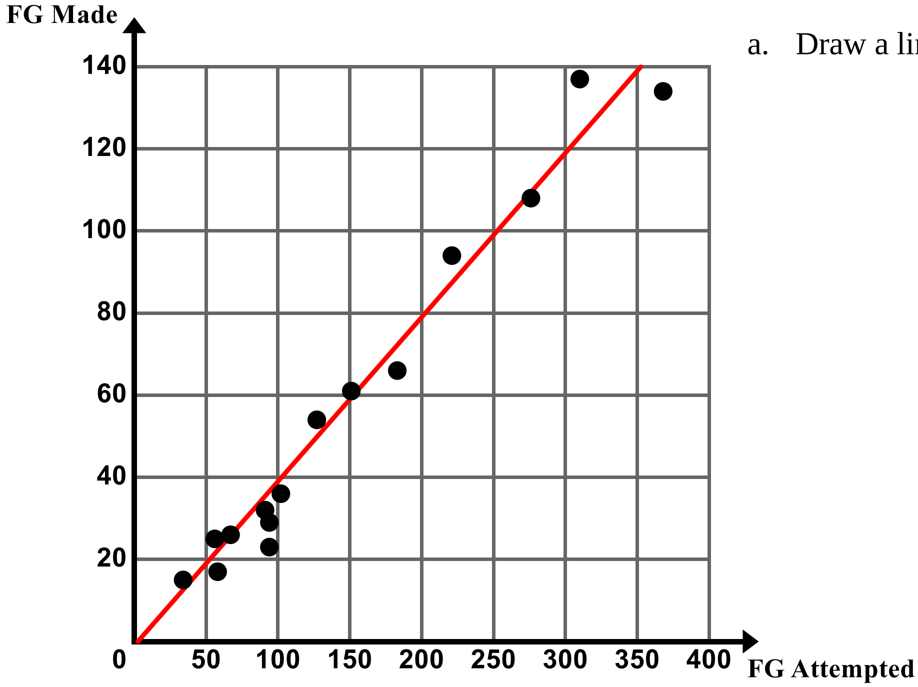


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

**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.


Discuss with students which two parameters we need to find in order to write an equation that shows the relationship between two variables. For ease, we often choose two points that fall on the corners of the grid (integer points). These may or may not be actual data points and they may or may not fall on the line (we can use corner points that are very close to the line).

In this problem, it seems reasonable to use the points (150, 60) and (200, 80) – you may choose to use different points and your equation will vary slightly. You may look at the graph and decide it is reasonable to say the *y*-intercept is 0 and use this as one of your points. For help calculating the slope from two points and determining the *y*-intercept, refer to #3d in 6.2a Homework.

(150, 60) and (200, 80)

$y \approx \frac{2}{5}x$

If your data shows a positive association (as it does above), then your slope should also be positive.


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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

**Slope:** For each additional 5 shots attempted, 2 would be made. If you are thinking of the slope as 0.4, you may say that for each additional shot attempted, 0.4 shots are made. The first explanation seems to make more sense in the context than saying that part of a shot is made. It will help students if you have them label their slope with the appropriate quantities: $\frac{\text{2 field goals made}}{\text{5 field goals attempted}}$

***y*-intercept:** The $y$-intercept would tell us that if a person attempts 0 shots, they will make 0 shots – makes a lot of sense. Further, even with 1 shot, we're not likely to score. However at 2 we likely will. The interpretation of the $y$-intercept will not always be this straightforward as we will see in upcoming examples.

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

A person that has 500 field goal attempts would likely make 200 field goals. To solve this problem, use your equation from part b, substitute in what you know (in this case we know the person attempted 500 field goals), and solve for the unknown. $y = \frac{2}{5}x \rightarrow y = \frac{2}{5}(500) \rightarrow y = 200$.

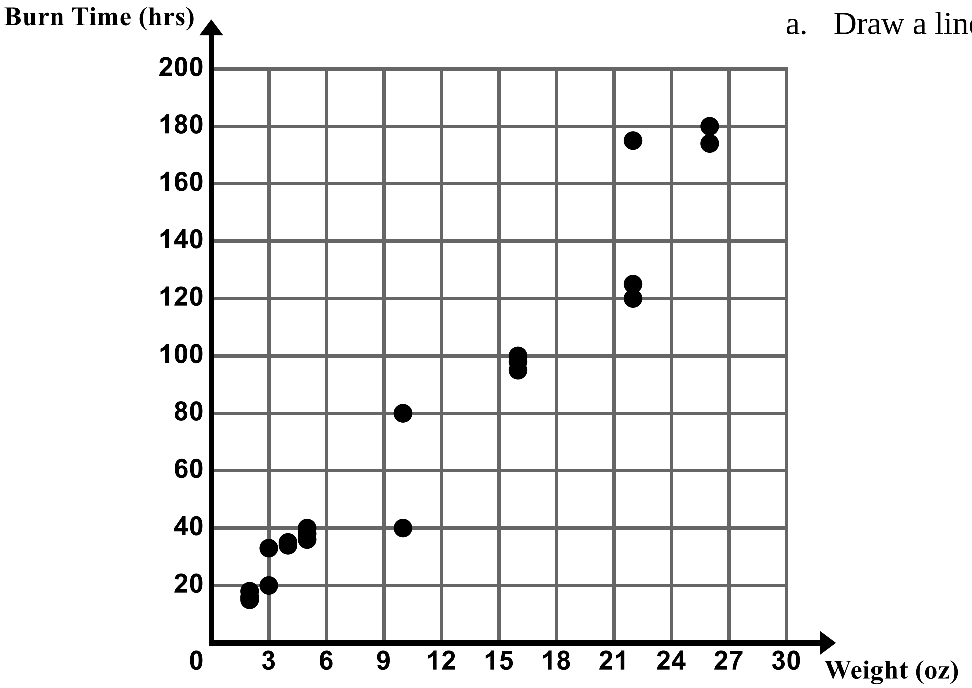
e. Use your prediction function to predict the number of field goals a person would make if they attempted 102 field goals. Again, use the equation from above, substitute in what you know, and solve for the unknown. Using the equation above, the answer is 40.8. Making part of a shot does not make sense – it seems reasonable to round this to 41. After you find the prediction using your line of best fit, observe the actual data point of a player who attempted 102 field goals. This person made 36 of them. This is pretty close to the prediction made by the equation. Distinguish between the realization (actual data point) and the prediction from the equation.

f. Is the association between number of field goals attempted and number of field goals made strong or weak? Justify your answer. Now that students have drawn a line through the data, they can more easily see that this is a strong linear association. If you observe the vertical distance from each of the data points to the line, you see that the vertical distance is small for most data points. Students will study this idea more formally in Secondary I when they calculate correlation coefficients and residuals.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

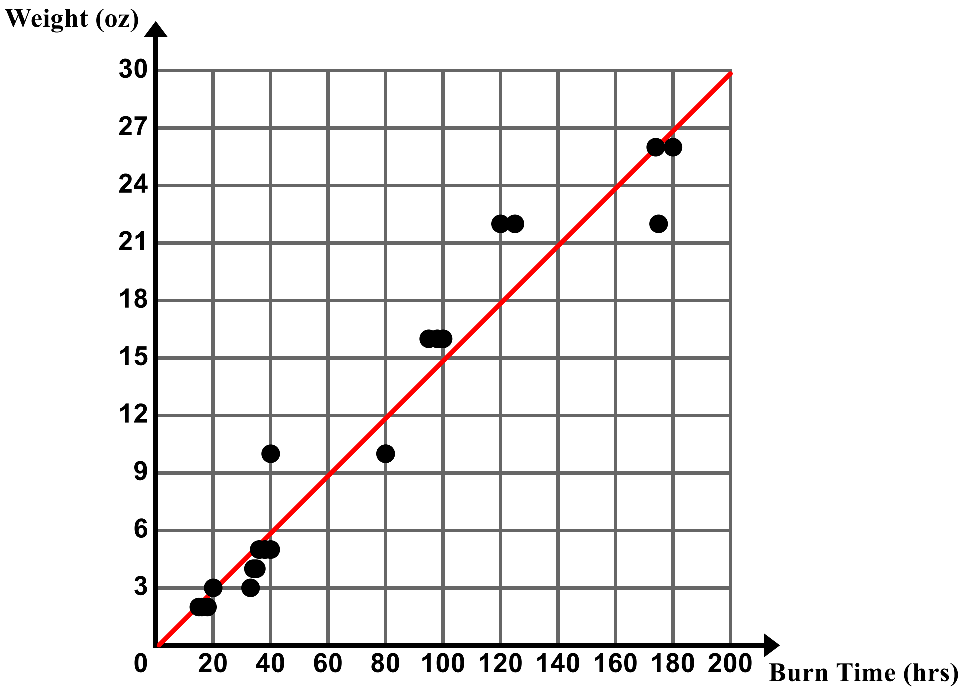
e. If candle burns out at 500 hours, predict how much the candle weighs.

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?



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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? Answers will vary.

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

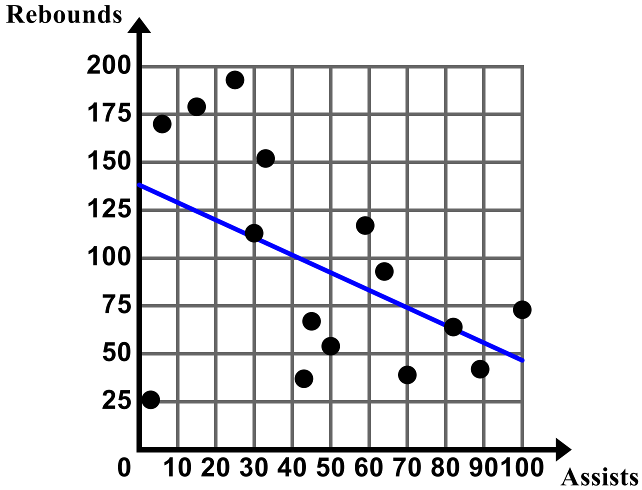
c. Write a prediction function for the line of best fit you drew.
It seems reasonable to use the points (20, 2) and (176, 26) giving the following equation:
$y \approx .15x$
However, it is perfectly acceptable for students to "eyeball" the line and then estimate the slope.

d. How does this new function compare to your equation in #2? What accounts for this change?
Since we have changed our *x* and *y* variables on the graph, the slope will be inverted. In #2, our slope was $\frac{40}{6}$ while here our slope is $\frac{6}{40}$



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

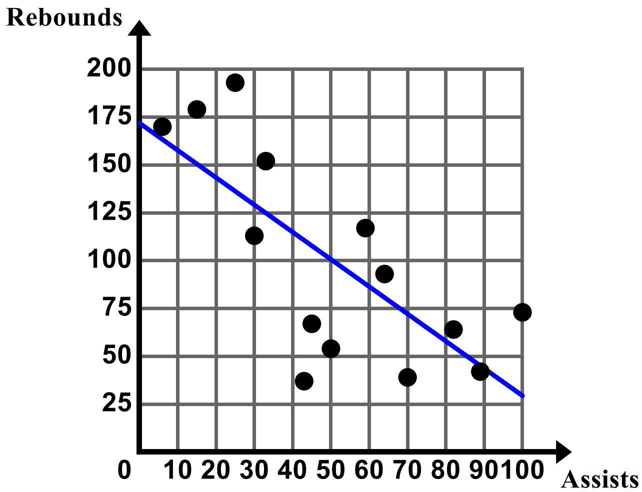
4. Software programs and graphing calculators can be used to draw lines of best fit. Izumi used a graphing calculator to generate a line of best fit for her data on assists and rebounds. The graph below shows the line of best fit generated by the calculator.




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.

There is not a complete set of data for her so it does make sense to remove this outlier from the data set.

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?

The outlier was pulling the line of best fit down and it made it less steep. With the outlier removed, the line of best fit better captures the association between assists and rebounds on this team.

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.

It seems reasonable to use the points (50, 100) and (70, 75) to write our equation:

$$y \approx -\frac{5}{4}x + 162.5$$



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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

The **slope** seems to indicate a negative association. It appears that for every additional assist that a player makes the number of rebounds they make likely decreases by 1.25. Or, for each additional 4 assists a player makes, the number of rebounds they make likely decreases by 5 or vice-versa (for each additional 5 rebounds a person makes, the number of assists they make decreases by 4). Again, it is recommended to have students label the quantities associated with the rise and run in the slope in order to better interpret the slope in context: $-\frac{5 \text{ rebounds}}{4 \text{ assists}}$

The ***y*-intercept** indicates that, for a random player on the team, if they were to have 0 assists you could expect them to also have made 162.5 rebounds. This is a situation where you can talk with students about thinking critically about the data. Is it feasible for a player to have 0 assists and make 162.5 rebounds? If a player has 0 assists, they likely did not play much so would not have this many rebounds. This shows some of the limitations of the data.

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.

Use your prediction equation from above, substitute in what you know, and solve for the unknown. $y = -\frac{5}{4}x + 162.5 \rightarrow y = -\frac{5}{4}(110) + 162.5 \rightarrow y = 25$. For a random player on the team you could expect them to make 25 rebounds throughout the season if they have 110 assists. However if a player had 150 assists the equation yields -25 rebounds. This is impossible; there are limitations on this data for extreme values. You will notice that none of the players even had 150 assists so this may not even be a realistic question to ask.

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

If a random player has 150 rebounds you would expect them to have 10 assists. Again, use your prediction equation from above and substitute in what you know. This time we know the number of rebounds, which is represented by *y* in our equation, is equal to 150.

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?

The data in the field goals made vs. field goals attempted plot is more closely aligned with its line of best fit. This indicates a strong relationship and the function can likely be used to make more accurate predictions about the data. We can see in the case of the rebounds vs. assists, the vertical distance from each of the data points to the line is larger than in the case of the shots made vs. shots attempted. Still, the data points are not that far from the line in the rebounds vs assists, so the strength is likely moderate as opposed to weak.

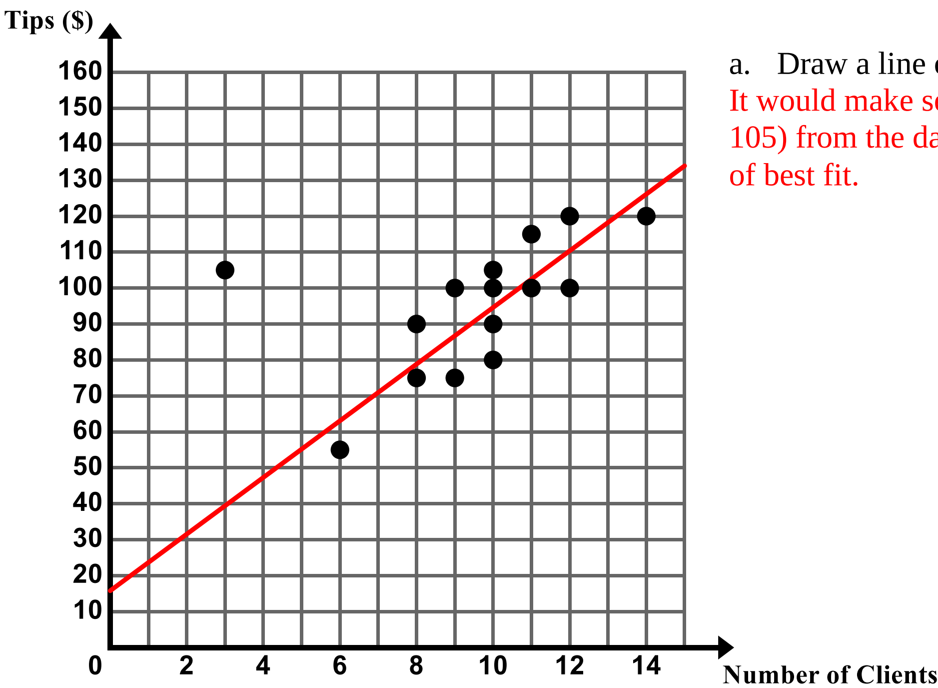


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

**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.
It would make sense to remove the outlier (3, 105) from the data set prior to drawing the line of best fit.

b. Write a prediction function for the line of best fit you drew.
Here is a particular eyeball result:
$y \approx 8x + 16$
Note: Equations may vary.

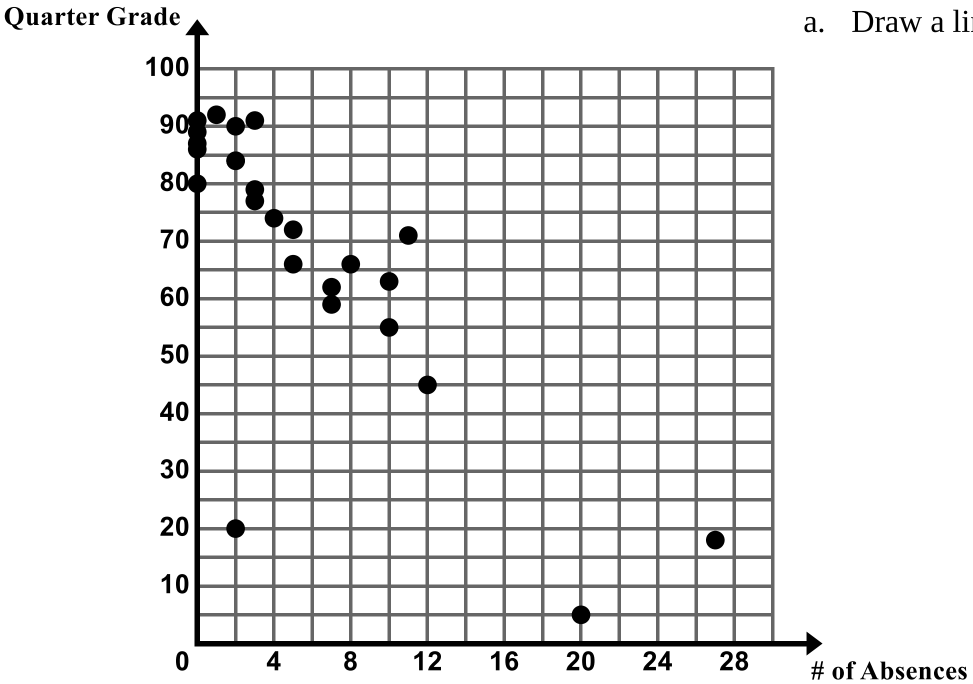
c. Explain the meaning of the slope and *y*-intercept in the context.
Slope: For each additional client that Jenny sees, she will make an additional $8 in tips. The *y*-intercept indicates that she would make $16 in tips if she sees 0 clients.

d. Use your prediction function to predict the amount Jenny would make in tips if she had 18 clients in one day.
$160



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

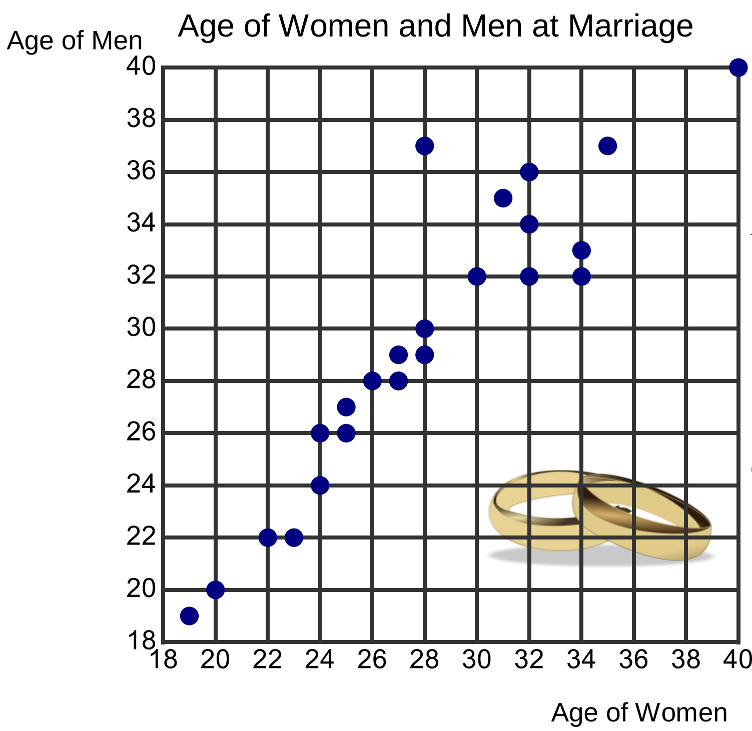
d. Use your prediction function to predict the final grade of a student who is absent 16 times.

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



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

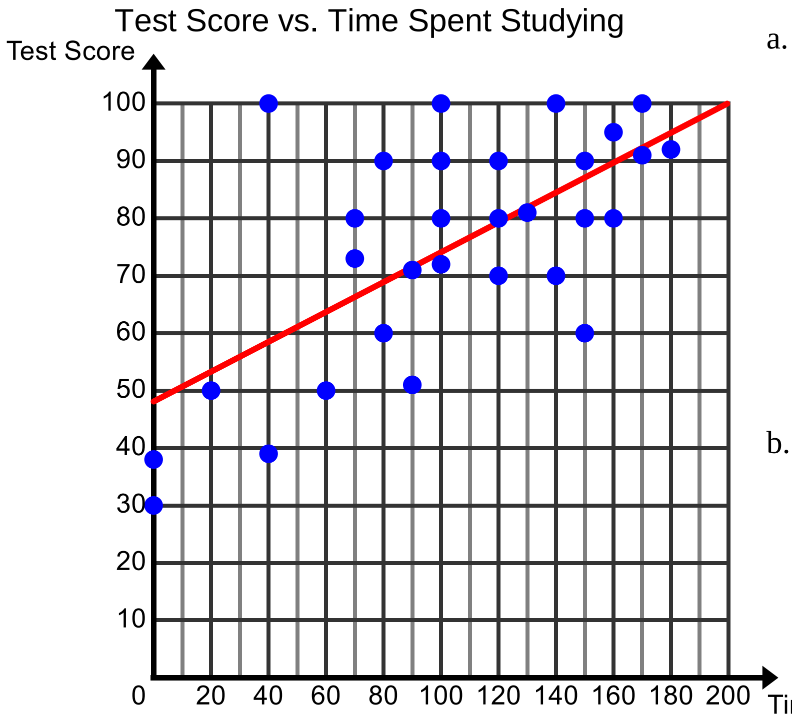
d. Write a prediction function for the line of best fit you drew.

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



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.
As the time spent studying increases the test score also increases. This shows a positive linear association that is fairly weak. Most of the data points appear to be clustered between the time intervals of 80 to 160 minutes. There is an outlier at (40, 100).

b. Provide an explanation for any clusters of data or outliers.
This person did not study as long as other people but still earned a 100 on the test. One possible explanation is that this person paid very close attention in class or they may have taken the class before.

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

d. Write a prediction function for the line of best fit you drew.
$y \approx \frac{1}{4}x + 50$
Equations may vary; see #3d in 6.2a Homework for a detailed explanation on how to write the equation of a line.

e. Explain the meaning of the slope and *y*-intercept of your line of best fit in the context.
Slope: A person receives an additional point on the test for each additional 4 minutes they study
*y*-intercept: A person who does not study at all can expect to earn a 50 on the test.

f. Use your prediction function to predict the score for a person who studies for 160 minutes.
90 (To answer this question, use the equation you found in part d, substitute in what you know ($x = 160$) and solve for the unknown ($y$ = score on test).

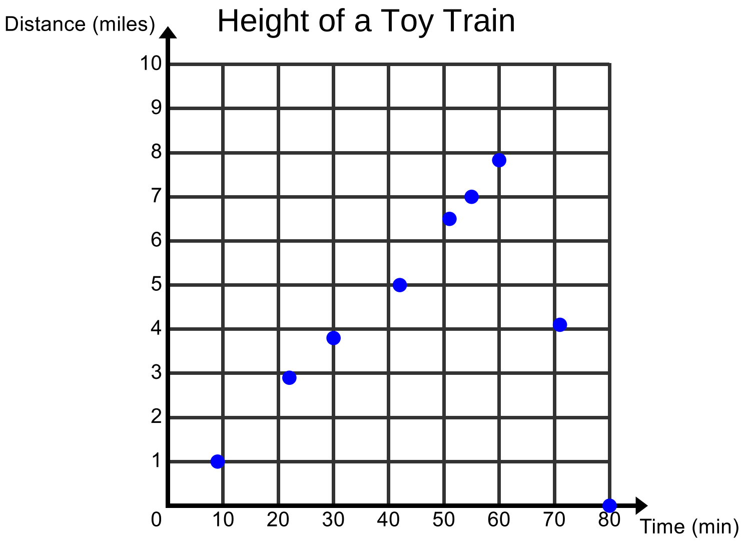
g. Compare and contrast the prediction calculated using the equation with the actual data points of the people who studied for 160 minutes.
The realizations or actual data points of people who studied for 160 minutes are (160, 95) and (160, 80) so the prediction is a fairly good average of these two data points and a good prediction of what a student might do.

h. Does the association between these two variables appear to be weak or strong? Provide an explanation regarding why the strength is this way.
These data points are not extremely close to the line of best fit, indicating that the association is not really strong. There are many other factors that contribute to how well a student does on a test.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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?
The data shows a nonlinear association.

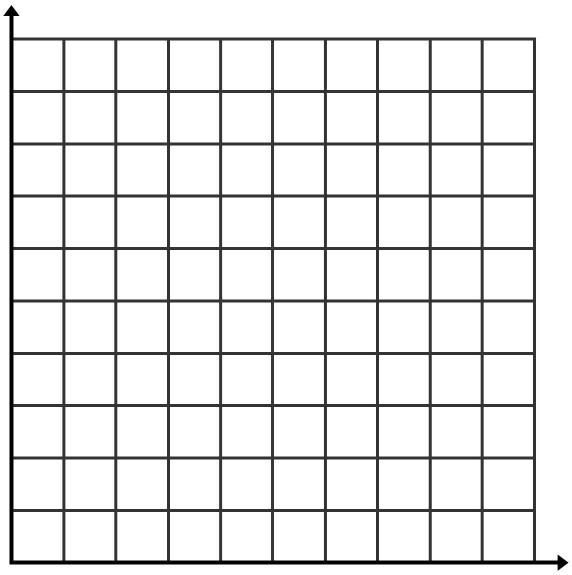
b. Would it be feasible to draw a line of best fit for this data? Why or why not.
No, the data is not linear so a line of best fit would not work for this data.

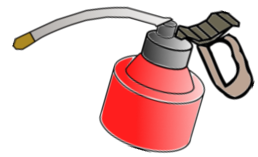
The issue may arise that this scatter plot is linear up to a point. This is true, however, over the entire domain or the time interval from 0 to 80 minutes it is not linear.


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.



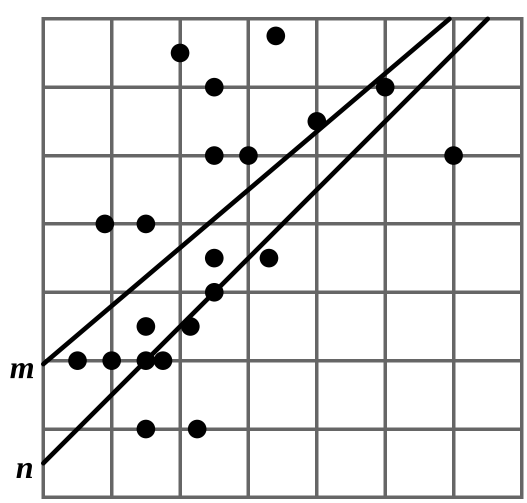
©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

**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. | | | | |

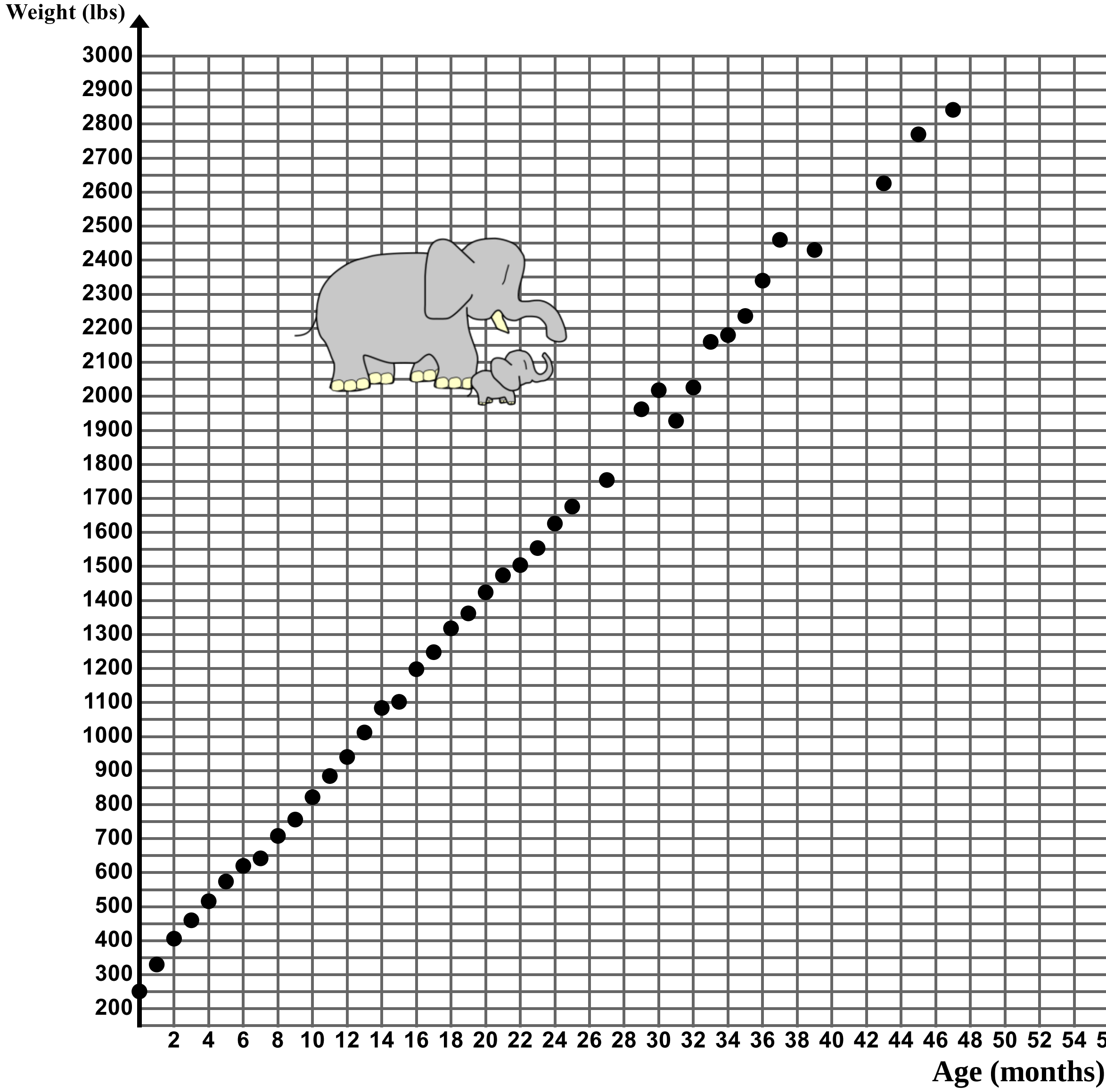
1. Which line, *m* or *n*, is the best fit for the data? Justify your answer (for use with Skill/Concepts #2).






©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

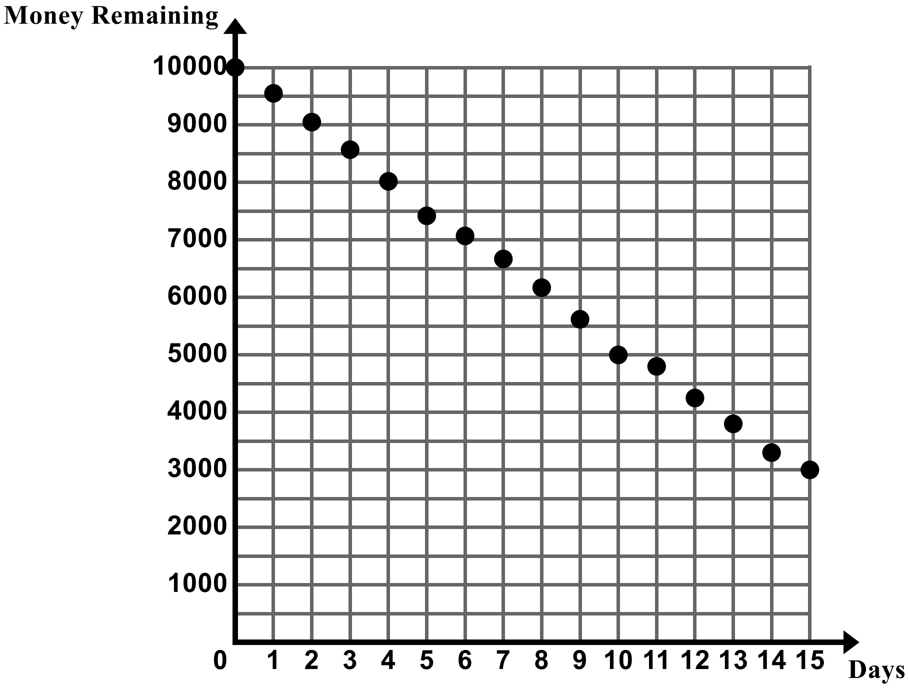
d. Explain the meaning of the slope and *y*-intercept of your line of best fit in the context.

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*


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.

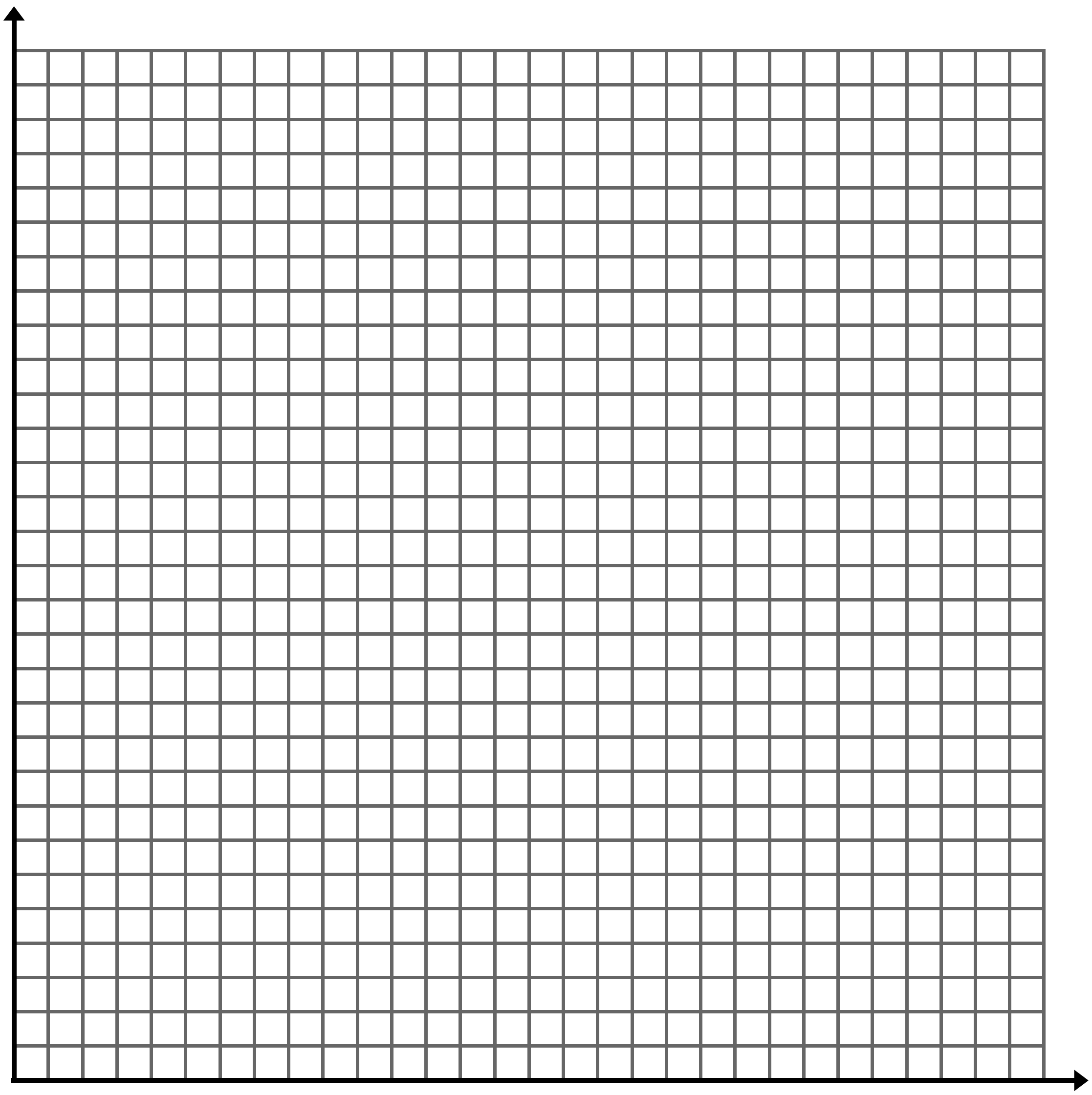
d. Explain the meaning of the slope and *y*-intercept of your line of best fit in the context.

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.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.





©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

# 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.


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

## 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 Izuhmi'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
   Categorical

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

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.


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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 | 35 |
| **No Chores** | 5 | 12 | 17 |
| **Total** | 31 | 21 | 52 |

Notice that the numbers in the row total sum to 52 (31 + 21) and the numbers in the column total sum to 52 (35 + 17). These should always sum to the **total number of people surveyed**.

**Directions:** Use the table to answer each question below.

a. How many students have a curfew and have chores?
   26
b. How many students have no curfew and have chores?
   9
c. How many students have no curfew and no chores?
   12

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.
   See table.
e. How many of the students surveyed have chores?
   35
f. How many of the students surveyed have a curfew?
   31

The frequencies calculated in parts d, e, and f are called *marginal* frequencies. They are located in the margins of the table. The frequencies found within the body of the table are called *joint* frequencies. These will be more formally discussed in Secondary 1.

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?
   Yes


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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** | 190 | 110 | 300 |
| **Do not eat breakfast regularly** | 130 | 165 | 295 |
| **Total** | 320 | 275 | 595 |

   b. How many females total were surveyed?
   275
   c. How many people surveyed eat breakfast regularly?
   300
   d. How many people total were surveyed?
   595
   e. How many males surveyed do not eat breakfast regularly?
   130
   f. How many females surveyed eat breakfast regularly?
   110
   g. What percentage **of the total number of people surveyed** eat breakfast regularly?
   $\frac{300}{595} = 50.4\%$
   h. What percentage **of the females surveyed** eat breakfast regularly?
   $\frac{110}{275} = 40\%$
   i. What percentage **of the people who eat breakfast regularly** are male?
   $\frac{190}{300} = 63.3\%$

   j. What percentage **of the total number of people surveyed** are females who do not eat breakfast regularly?
   $\frac{165}{595} = 27.7\%$

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

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


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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. Examine the table to see which pieces you are able to fill in first. For example, you can start by filling in the total number of people surveyed by adding the number of males and females surveyed in the column total. You can also determine the number of females that take a car to school. If there is a cell you don't have enough information to fill in, try a different one first, and then go back to that cell.

| | 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?

$\frac{80}{221} = 36.2\%$

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

$\frac{17}{221} = 7.7\%$

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.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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. Below is one way to set up the table. You may also switch the row and column headings (MP3 players as your column headings and smart phones as your row headings).

| | **Owns a smart phone** | **Does not own a smart phone** | **Total** |
|---|---|---|---|
| **Owns an MP3 player** | | | |
| **Does not own an MP3 Player** | | | |
| **Total** | | | 38 |

   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?



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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

In this problem, we are working backwards. Instead of being given the counts as in previous problems, we are given the percentages and we have to find the counts. In this piece of the problem, $\frac{1}{2}$ of the total number of children chose chocolate. $\frac{1}{2}$ of 200 is equal to 100. Now we can fill in the column total for chocolate with 100.

- 25% of the children surveyed were in the age range of 8 – 12 years old

In this piece of the problem, we know that 25% or $\frac{1}{4}$ of the children surveyed are in the age range of 8 – 12 years old. $\frac{1}{4}$ of 200 is equal to 50. You can also multiply 0.25 and 200 which also equals 50. Now we can fill this in the row total for ages 8 – 12.

- $\frac{2}{5}$ of the children surveyed were in the age range of 13 – 17 years old

To solve this piece of the problem, we need to find $\frac{2}{5}$ of 200 or $\frac{2}{5} \times 200 = 80$. If students struggle with the math, determine $\frac{1}{5}$ of 200 which is 40 and multiply it by 2.

- 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** | 35 | 9 | 26 | 70 |
| **Ages 8 – 12** | 25 | 13 | 12 | 50 |
| **Ages 13 – 17** | 40 | 28 | 12 | 80 |
| **Total** | 100 | 50 | 50 | 200 |


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

## 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** | 8 | 38 | 46 |
| **Pop** | 142 | 20 | 162 |
| **Totals** | 150 | 58 | 208 |

Notice that the numbers in the row total sum to 208 (150 + 58) and the numbers in the column total also sum to 208 (46 + 162). These should always sum to the **total number of people surveyed**.

b. How many total songs are in her "Music for Studying" playlist?
58

c. How many classical songs are in her "Workout Music" playlist?
8

d. What percentage **of songs in the collection** are pop?
$\frac{162}{208} = 77.9\%$

e. What percentage **of songs in the collection** are for studying?
$\frac{58}{208} = 27.9\%$

f. What percentage **of the classical music** is music for studying?
$\frac{38}{46} = 82.6\%$ Notice how the denominator changes here. We are looking for the percentage **of classical music** as opposed to the percentage **of the total** as in parts d. and e.

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



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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** | | 61 | |
| **Tail** | | | 79 |
| **Totals** | 21 | | |

   a. How many times did he toss a tails and a six?
   11

   b. How many times did he toss a heads?


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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. See class activity #9 for help setting up a two-way frequency table.

| | | | |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |

   b. How many members of the club are female?
   120

   c. How many member of the club are over 18 and female?
   84

   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?


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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. See class activity #10 for help with this problem.
   - 35% of the people surveyed prefer comedy movies To solve this piece of the problem, you must determine what 35% of 500 is: $0.35 \times 500 = 175$. Now you can fill in the table with this piece of data.
   - $\frac{3}{10}$ of the people surveyed prefer action movies
   - 95 people surveyed prefer romance movies
   - Of the females surveyed, $\frac{2}{7}$ prefer romance movies Be careful with this piece, it is $\frac{2}{7}$ **of the females surveyed** that prefer romance: $\frac{2}{7} \times 280 = 80$.
   - 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** | 80 | | | 52 | 280 |
| **Total** | | 175 | | | 500 |


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

**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?
The most popular types of movies among males and females tend to be comedies and action movies. On could make an argument that Julie should choose comedy because males and females have an equal likelihood of preferring comedies (35% of the males and 35% of the females chose comedy). For action movies, 45% of the males prefer action movies while only about 18% of the females prefer action movies.

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?
Again, either comedy or action would be a good choice. Of the males surveyed, 45% prefer action movies while 35% prefer comedy movies.

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?
Julie should probably either choose romance or comedy. Of the females surveyed, about 29% prefer romance and about 35% prefer comedy. Only about 18% prefer action and about 19% prefer drama.

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.**

___T___ Males and females have an equal likelihood of choosing comedy movies. Make sure that students see that we need to consider the counts in the table in relationship to the totals. Upon first glance, students may think that females have a greater likelihood of preferring comedies because the count is higher in the table (98 vs. 77); however more females were surveyed (280 vs. 220). If we look at the percentage **of males** and **of females** who prefer comedies, we see that both equal 35% so according to this data, males and females have an equal likelihood of choosing comedies.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.**

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

50% of the children in all age ranges prefer chocolate

___F___ Children in the age ranges 8 – 12 and 13 – 17 have an equal likelihood of choosing strawberry. Of the children in the age range 8 – 12, 24% chose strawberry. Of the children in the age range 13 – 17, 15% chose strawberry.

___T___ As students get older they tend to like vanilla more. About 13% of the children in the age range of 3 – 7 prefer vanilla, 26% of the children in the age range 8 – 12 prefer vanilla, and 35% of the children in the age range 13 – 17 prefer vanilla



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.
Again, students will draw a variety of conclusions – 80% of the people surveyed buy their tomatoes at the grocery store, 5% at a farmer's market and 15% grow their own tomatoes. 66% of the people surveyed like tomatoes, 34% do not. 3% of the people surveyed grow their own tomatoes and dislike tomatoes.

b. Is there an association between growing your own tomatoes (having a home garden) and whether or not you like tomatoes?
There are many arguments that students can make and they are all valid **as long as the students support their arguments with mathematical evidence**. Students may make the following arguments. **Of the people who like tomatoes,** what percentage buys their tomatoes at the grocery store ($\frac{50}{66}$ or roughly 76%)? Buys their tomatoes at a farmer's market ($\frac{4}{66}$ or roughly 6%)? Grow them in their garden ($\frac{12}{66}$ or roughly 18%)? But does this really answer our question since so many people get their tomatoes at the grocery store in the first place. Let's examine some column frequencies and see what we can find out? **Of the people who buy their tomatoes at the grocery store,** what percentage like tomatoes ($\frac{50}{80}$ or roughly 62.5%)? **Of the people who buy their tomatoes at the grocery store,** what percentage **do not like tomatoes** ($\frac{30}{80}$ or roughly 37.5%). How about people who have a home garden? **Of the people who have a home garden,** what percentage like tomatoes ($\frac{12}{15}$ or roughly 80%)? **Of the people who have a home garden,** what percentage **do not like tomatoes** ($\frac{3}{15}$ or roughly 20%). These relative frequencies seem to tell us that there is an association between people who grow their own tomatoes and people who like tomatoes. If we add in the data from the farmer's market, we support the argument that people prefer the taste of tomatoes that are fresh and locally grown. This conclusion makes sense - after all, wouldn't we expect people who plant tomatoes or buy them at the market to like them in the first place?


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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? This is a good problem for discussing how you need to look at several different angles of a two-way frequency table in order to draw valid conclusions. What if you only calculated the percentage of students who eat breakfast regularly (50.4%) and the percentage of students who do not eat breakfast regularly (49.6%)? One might conclude that this demonstrates that there is no association between gender and whether or not a person eats breakfast regularly because an equal percentage eat breakfast and do not eat breakfast. But what if your sample space included more males than females or vice-versa? Let's look at it from another angle. **Of the people who are male,** what percentage eat breakfast regularly (59.3%). **Of the people who are female,** what percentage eat breakfast regularly? (40%). It seems as though males tend to eat breakfast more regularly. This would indicate that there is a weak association between gender and whether or not a person eats breakfast.

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? Yes, numerical evidence would suggest a strong association. 96% of females have their ears pierced while only 21% of males have their ears pierced. Of the people that have pierced ears, roughly 18% are men and roughly 82% are women.


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

**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? 12%

b. What percentage **of students surveyed** are males who walk to school? 15%

c. Based off of the table above what is the most popular mode of transportation for the sample population. Walking is the most popular mode of transportation with 36% of the student population that walk to school.

d. What is the preferred method of transportation for females? Use numerical evidence to support your answer. Walking, 50% of females walk to school.

e. What is the preferred method of transportation for males? Use numerical evidence to support your answer. Riding their bike (cycling), 40% of males bike to school.

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



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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. There are many valid arguments that can be made here **as long as the arguments are supported with numerical evidence from the table**.

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.
Yes, there does appear to be an association between what Miss Marble is doing and what type of music she prefers. There are many arguments that can be made to support this conclusion. One possible argument is, while working out 95% of Miss Marbles' workout music is Pop.


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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. The calculations you made in parts a. – f. should help you to determine if there is an association. Also, use these calculations to back up your arguments with **numerical evidence**.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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. If you are stumped on this problem, start by determining some relative frequencies. For example, determine what percentage of college age students prefer twitter. Then determine what percentage of middle school students prefer twitter. Do similar calculations for Facebook and Instagram. This should help you to identity associations and draw conclusions about the data.

.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

## 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 | | |


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

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.



©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.

**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.


©2014 University of Utah Middle School Math Project in partnership with the Utah State Office of Education. Licensed under Creative Commons, cc-by.