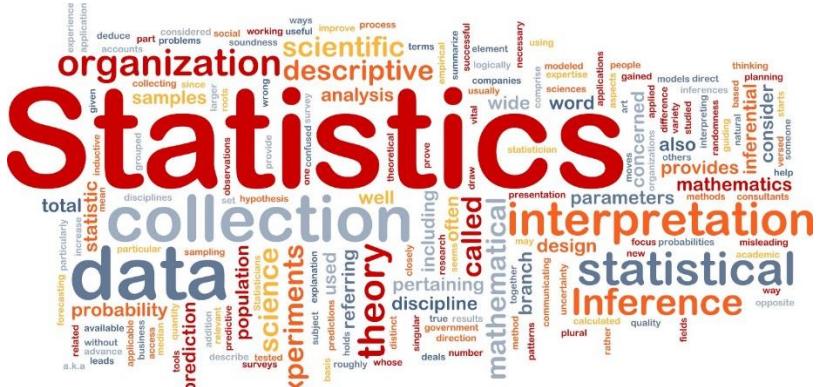


# Statistics



## **Unit 4**

## **Standards:**

**8.SP.1** – 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.2** – 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.3** – Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept.

**8.SP.4** – 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.

Name: \_\_\_\_\_ Period: \_\_\_\_\_

## Lesson #46 Categorical Data

\_\_\_\_\_ is data consisting of numbers.

Below are some measurement data about Ms. Robert's 8<sup>th</sup> grade math class.

• Number of boys.....	12
• Number of girls.....	18
• Number of students who ate breakfast this morning.....	20
• Number of students who passed the math test today.....	25

Use the data to answer the questions below.

1. How many students are in the class? \_\_\_\_\_

2. What percent of the students in the class are boys?

What percent of the students in the class are girls?

Why must the sum of these percents equal 100%?

3. What percent of the students ate breakfast this morning?

What percent of the students passed the math test today?

Why don't the sum of these percents equal 100%?

Below are some measurement data about a survey collected at School Orientation.

• Number of parents for uniforms.....	224
• Number of parents against uniforms.....	76
• Number of students for uniforms.....	98
• Number of students against uniforms....	202

1. How many people took the survey?

2. What percent of students surveyed are for uniforms?

What percent of students surveyed are against uniforms?

What percent of parents surveyed are for uniforms?

What percent of parents surveyed are against uniforms?

Why must the sum of the parents percents and students percents equal 100%?

Why must the sum of the percents for uniforms not equal 100%?

\_\_\_\_\_ is data sorted into categories, such as colors, ranges of measurements, or other attributes of the data.

**For each topic below, list categories for organizing collected data. Then create a possible survey question that you could use to collect data with each category.**

Topic	Possible Categories	Possible Survey Question(s)
Example: Hair color	Black, brown, blonde, red, other	What color is your hair? OR Do you have dark hair?
Music		
Art		
Fruit		
Ecosystems		

1. If you were reading a table that had categorical data labeled “swimming,” “football,” and “basketball,” what might be the larger topic that is being explored?

## HW #46 Categorical Data

State whether each question in the table below is a categorical or numerical survey question.

Question	Categorical (C) or Numerical (N)?
1. Are you male or female?	
2. What is your favorite color?	
3. How many dress shirts do you own?	
4. Are you an only child?	
5. How many siblings do you have?	
6. What were the class scores on the test?	
7. What types of birds did we observe today?	
8. What are the income levels of different cities in my state?	

9. State whether the question is categorical or numerical. Then rewrite the question in the other form.

“How many video games do you own?” Type: \_\_\_\_\_

Rewritten in other form:

10. Create a categorical question and then rewrite it as a numerical question.

## Lesson #47 Bivariate Categorical Data

\_\_\_\_\_ is data that has two variables based on the same population.

The results of an 8<sup>th</sup> grade survey about favorite color are below.

	Pink	Red	Blue	Purple	TOTAL
Boys	2	10	10	3	
Girls	8	2	3	7	
TOTAL					

1. Complete the table by finding totals for the rows and the columns.
2. How many students total were surveyed? n = \_\_\_\_\_
3. How many girls were surveyed? \_\_\_\_\_
4. What percent of girls preferred pink? \_\_\_\_\_
5. How many boys were surveyed? \_\_\_\_\_
6. What percent of boys preferred pink? \_\_\_\_\_
7. How many students preferred pink? \_\_\_\_\_
8. What percent of the students who preferred pink are girls? \_\_\_\_\_
9. What percent of the students who preferred pink are boys? \_\_\_\_\_
10. Compare the **questions** from problems 4 and 8. How are they different?
  
11. Compare the **answers** from problems 4 and 8. How are they different?
  
12. The table above is called a two – way table. Explain what you think a two – way table is in your own words?

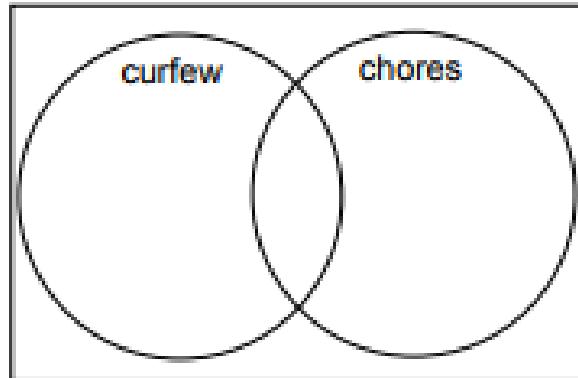
Ten different 8<sup>th</sup> grades were asked the following questions:

- Do you have a curfew?
- Are you assigned chores at home

Data was collected on their responses and recorded in the table below.

Students A through J										
	A	B	C	D	E	F	G	H	I	J
Curfew	Yes	No	No	Yes	Yes	No	Yes	No	No	Yes
Chores	Yes	No	Yes	Yes	No	No	Yes	Yes	No	Yes

1. Use the sections in the Venn diagram to record the number of students in each category. Be sure to include the number of students with neither chores nor curfew outside the circles.
2. How many students were surveyed? \_\_\_\_\_  
This is called the number of observations (n)  
Or sample size for the survey.
3. How many students surveyed had chores?
4. How many students surveyed had curfew?
5. What percent of students had both chores and a curfew?
6. What percent of students had neither chores nor a curfew?
7. What percent of students who had chores also had a curfew?
8. What percent of student who had a curfew also had chores?
9. Jarred thinks that the most students have both a curfew and chores. Does the data support his claim? Explain.



## HW #47 Two – way Frequency Tables

A frequency table is a table that lists items and the number of times they occur in a data set.

1. Use the data from today's lesson to complete the table.

	Students with Curfew	Students with No Curfew	TOTAL
Students with Chores			
Students with No Chores			
TOTAL			

Based on this table:

2. Circle the number that indicates the sample size for the whole survey.

What percent of all students had chores?

3. Draw a square around the number that indicates the total number of students who had chores?

What percent of students who had chores also had a curfew?

4. Draw a triangle around the number that indicates the total number of student who did not have a curfew.

What percent of the students who did not have a curfew also did not have chores?

5. Draw a parallelogram around the number that indicates the total number of students who had a curfew.

What percent of students who had a curfew also had chores?

6. Mikayla thinks that most students who were surveyed who had chores were more likely to have a curfew. Does the data support Mikayla's claim? Explain.

## Lesson #48 Relative Frequency Tables

A frequency table with raw data can be used to create a \_\_\_\_\_ that contains percents. The strength of any association or relationship between two variables can be easier to describe using percents.

1. Here is one way to create a relative frequency table about curfews and chores. Use data from the previous page to complete the table.

**Table 1: Curfew and Chores**

	Curfew	No Curfew	TOTAL
Chores (n = _____)			100%
No Chores (n = 4)	$\frac{1}{4} = 25\%$		100%

2. Here is another way to create a relative frequency table about curfews and chores. Use the data from the previous page to complete this table.

**Table 2: Curfew and Chores**

	Curfew (n = 5)	No Curfew (n = _____)
Chores		
No Chores	$\frac{1}{5} = 20\%$	
TOTAL	100%	100%

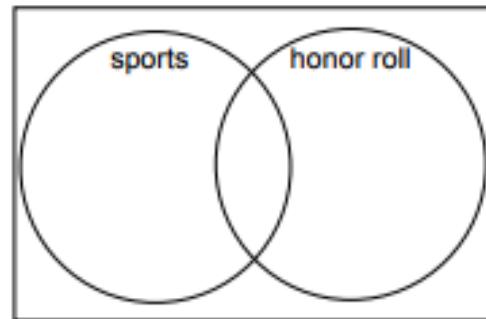
3. Circle the percent that compares curfews and chores in each table. Why is this percent different in the two tables?
4. Mikayla thought that students who were surveyed who had chores were more likely to have a curfew as well. Do these relative frequency tables support her claim? Explain.
5. Connor states, “Of students who were surveyed, students without curfews are less likely to have chores compared to those with a curfew.” Draw squares around the percents that show the association between not having a curfew and not having chores. Then determine if Connor’s reasoning is correct.

Jose collected data on the grades of athletes in his high school. He asked 500 students the following questions:

- Do you play a varsity sport?
  - Did you make honor roll last semester?
1. Jose lost some of his work. Help him figure out the missing values in his frequency table, and then record data in the venn diagram.

**Academics and Athletics**

	sport	no sport	TOTAL
honor roll	110		240
no honor roll		220	
TOTAL	150		500



2. Complete these relative frequency tables for Jose's data.

**Table 1:  
Do Athletes Earn Better Grades?**

	sport	no sport	TOTAL
honor roll (n = _____)			100%
no honor roll (n = _____)			100%

**Table 2:  
Do Athletes Earn Better Grades?**

	sport (n = _____)	no sport (n = _____)
honor roll		
no honor roll		
TOTAL	100%	100%

3. Jose's conjecture was that students who played sports had a greater likelihood of making the honor roll compared to students who do not play sports.

Explain why the relative frequency tables make this association easier to see compared to the frequency table and venn diagram above.

## **HW #48 Relative Frequency Tables**

**Use the Academics and Sports tables from today's lesson to answer the questions that follow.**

1. Which relative frequency table (1 or 2) can be used to answer the question, "What percent of students who made honor roll also played a sport? \_\_\_\_\_

Answer the question: \_\_\_\_\_ %

Write another question that can be answered from this table and answer it.

2. Which relative frequency table (1 or 2) can be used to answer the question, "What percent of students who played sports did not make the honor roll? \_\_\_\_\_

Answer the question: \_\_\_\_\_ %

Write another question that can be answered from this table and answer it.

3. Jose's conjecture was that students who played sports had a greater likelihood of making the honor roll compared to students who do not play sports. Which of the percentages in the relative frequency tables best support his argument? Explain.

4. Jose is going to present to the school board and argue against proposed budget cuts to the athletic program. What are three statistics that he could use in his presentation to show that participation in athletics helps support academic success?

## Lesson #49 Scatterplots and Linear Association

### Income and Education

1. Examine the table to the right.  
What does the data say?

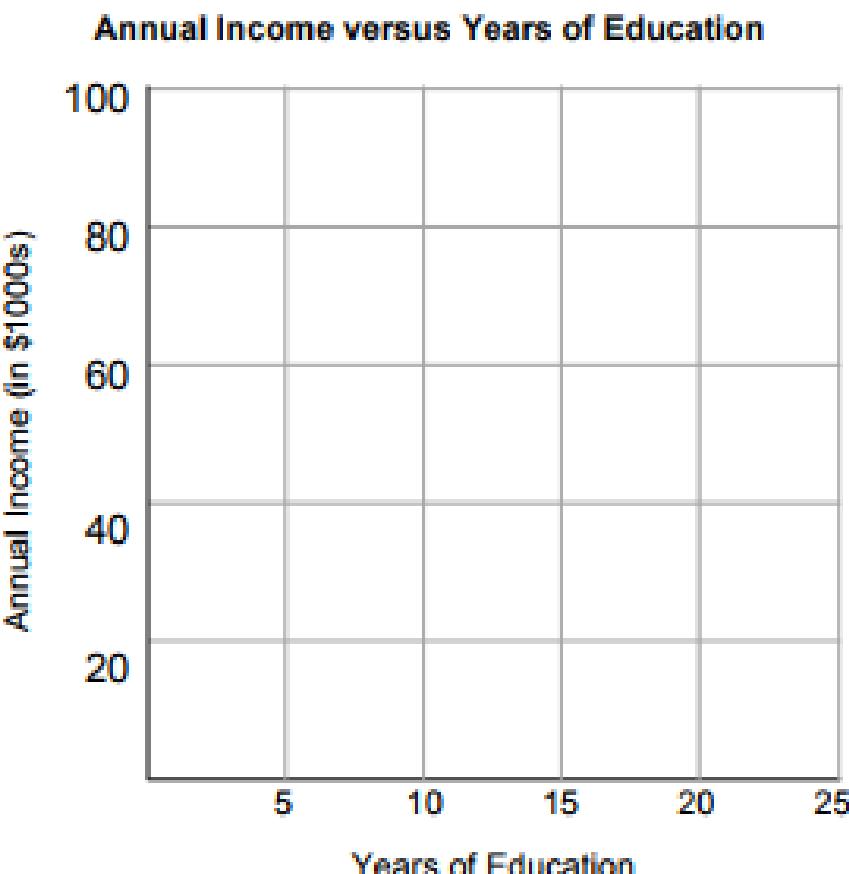
Level of Education	Average Years of Education	Average Annual Income (in \$1000s)
Not a High School Graduate	10	23
High School Graduate	12	32
Some College, No Degree	13	36
Associate Degree	14	38
Bachelor Degree	16	53
Master Degree	17	63
Doctorate Degree	20	81

2. Write sentences to describe what these order pairs from the table represent.

- a. (12, 32)
- b. (20, 81)

3. Graph the data from the table.

4. Describe the shape of the graphed data.



## Unemployment and Education

1. Examine the table to the right. What does the data say?

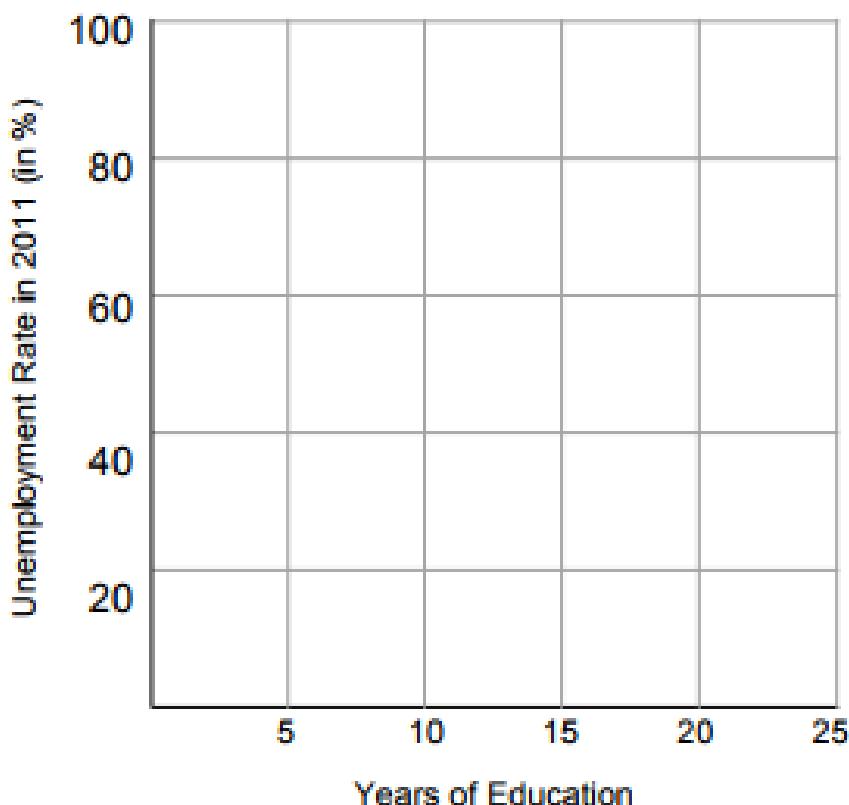
Level of Education	Average Years of Education	Unemployment Rate (in %)
Not a High School Graduate	10	14.1
High School Graduate	12	9.4
Some College, No Degree	13	8.7
Associate Degree	14	6.8
Bachelor Degree	16	4.9
Master Degree	17	3.6
Doctorate Degree	20	2.4

2. Write sentences to describe what these order pairs from the table represent.

- a. (10, 14.1)
- b. (20, 2.4)

**Unemployment Rate versus Years of Education**

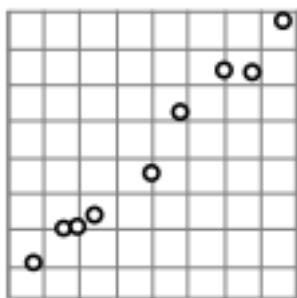
3. Graph the data from the table.



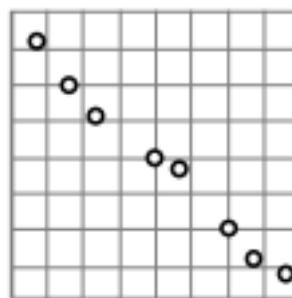
4. Describe the shape of the graphed data.

## HW #49 Scatterplots and Linear Association

- Fill in each blank with positive or negative.



We say the linear association is **strong** if it "clusters around a line."



---

association

---

association

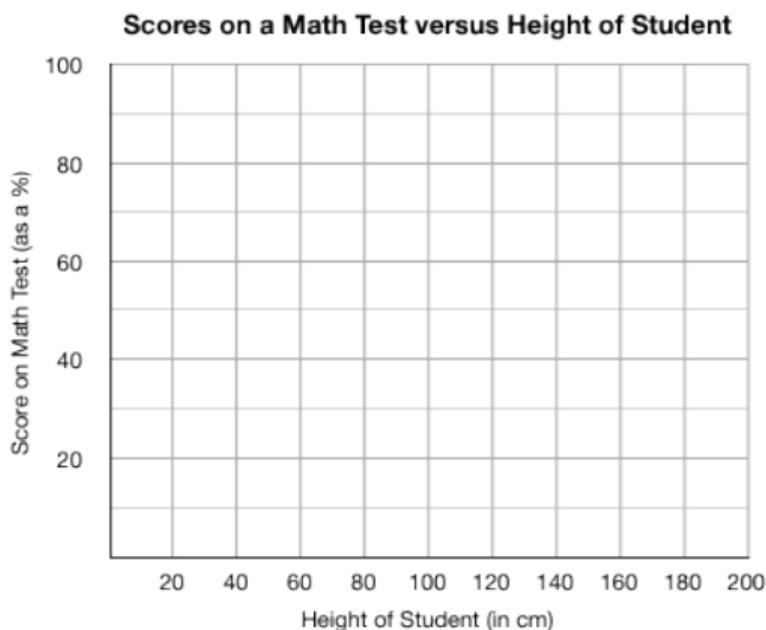
Use the "Income and Education" data:	Use the "Unemployment and Education" data:
2. Does the association between income and education appear to be strong? Explain.	6. Does the association between Unemployment and Education appear to be strong? Explain.
3. What kind of association does there appear to be between income and education? _____.	7. What kind of association does there appear to be between Unemployment and Education? _____.
4. In general, we can say that the _____ education people have, the <u>more</u> income they will likely earn.	8. In general, we can say that the _____ education people have, the <u>more</u> likely they could be unemployed.
5. In general, we can also say that the <u>less</u> education people have, the _____ income they will likely earn.	9. In general, we can also say that the <u>more</u> education people have, the _____ likely they could be unemployed.

- What are some possible reasons why these associations might exist?

## Lesson #50 Describing Linear Association

1. Examine the table below. Then make a graph of the data.

Height of Student (in cm)	Score on Math Test (as a %)
150	50
130	30
140	70
120	90
180	40
100	80
150	95
120	55



We say the association is weak or does not exist if the data “clusters in a cloud.”

2. Does there appear to be an association between the scores on the math test and the height of a student? Explain.

3. Examine the pairs of variables below. What kind of association (if any) do you think might exist for each pair? Be prepared to share your reasoning with others.

Pairs of variables	Kind of association (if any)
Number of sodas consumed and number of cavities.	
Age and quality of eyesight.	
Number of traffic tickets and cost of car insurance.	
Shoe size and number of pets at home.	
Life expectancy and income.	

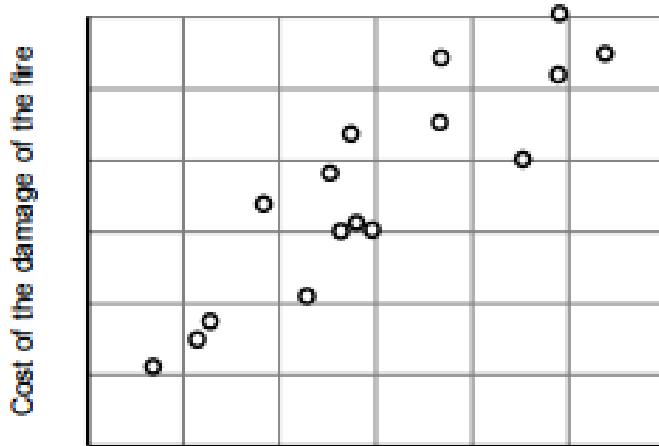
4. What are some other examples of variables that might have a positive or negative association?

\_\_\_\_\_ does not imply \_\_\_\_\_.

\_\_\_\_\_ is defined as the action of causing something. There is causation if one variable causes the other to happen.

1. What variables are compared on this graph?

**Cost of Fires vs Number of Firefighters Responding**



2. What does the data tell you?

3. Chase said, the graph shows “Firefighters cause damage because the more firemen who respond to a fire alarm, the more costly the damage of the fire.”

a. What is wrong with Chase’s statement?

b. What other variables might be a cause of this association between cost of the damage and the number of firefighters?

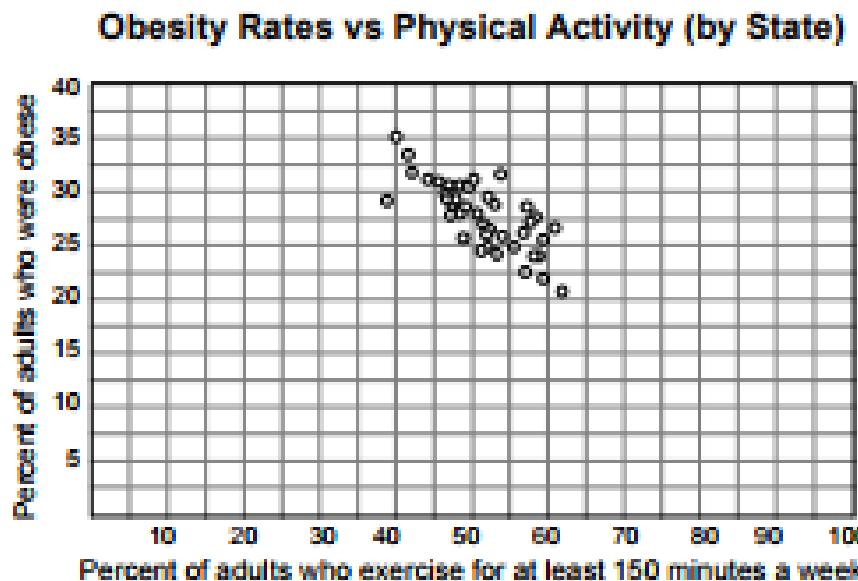
These statements confuse association with causation. For each statement, list other factors that might be a cause for the association.

Example	Other possible causes
4. People who walk faster tend to live longer. Therefore, if you want to live longer, walk faster.	
5. As sales of ice cream increase, the rate of drowning deaths increases. Therefore, ice cream causes people to drown.	
6. The more cell phones a country has, the longer the life expectancy of people in that country. Therefore, cell phones cause you to live longer.	

## HW #50 Describing Linear Association

The graph below compares data on exercise and obesity for 46 states. Data for the remaining 4 states is shown in the table.

- Graph the remaining states.



State	Exercise Rate (%)	Obesity Rate (%)
Washington	54.3	26.5
West Virginia	43.1	32.4
Wisconsin	57.5	27.7
Wyoming	53.3	25.0

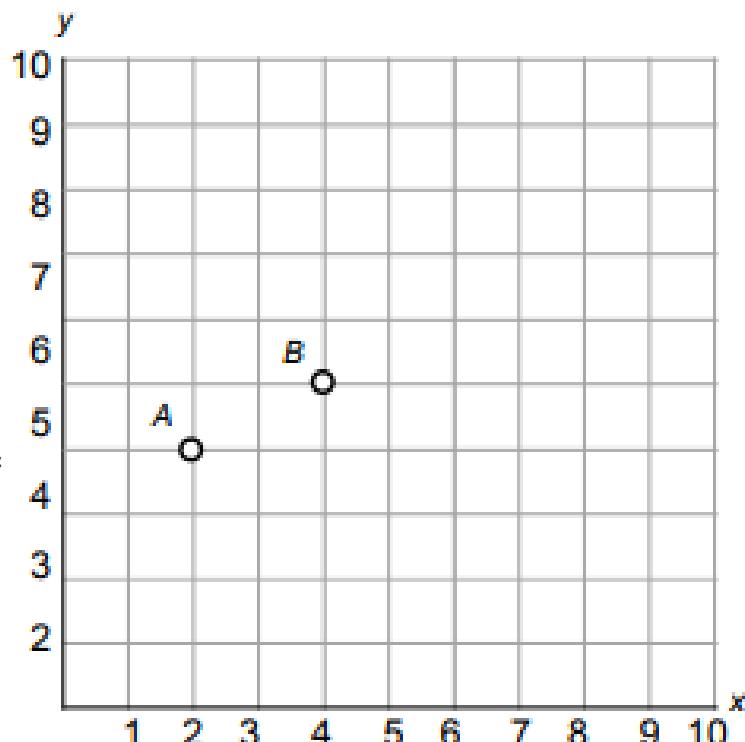
- Explain what Washington's data point (54.3, 26.5) means in the context of the problem.
- Describe any association that can be seen in the graph.
- Fill in the blanks with “higher” or “lower” so that the statements accurately reflect the association.

States that have a higher exercise rate tend to have a \_\_\_\_\_ obesity rate.  
States that have a \_\_\_\_\_ exercise rate tend to have a higher obesity rate.
- On July 10, 2012, the *New York Times* published an article titled “Despite Obesity Concerns, Gym Classes Are Cut.”  
Is this headline related to the graph above? Explain.

## Lesson #51 Line of Best Fit

- What are the coordinates for points A and B?
- Use a straight edge to draw a line that Goes through both points and extends to the left and right of the points.
- Use the graph to find the missing value of each point on the line listed below.

(6, \_\_\_\_\_) (3, \_\_\_\_\_) (\_\_\_\_\_, 7)



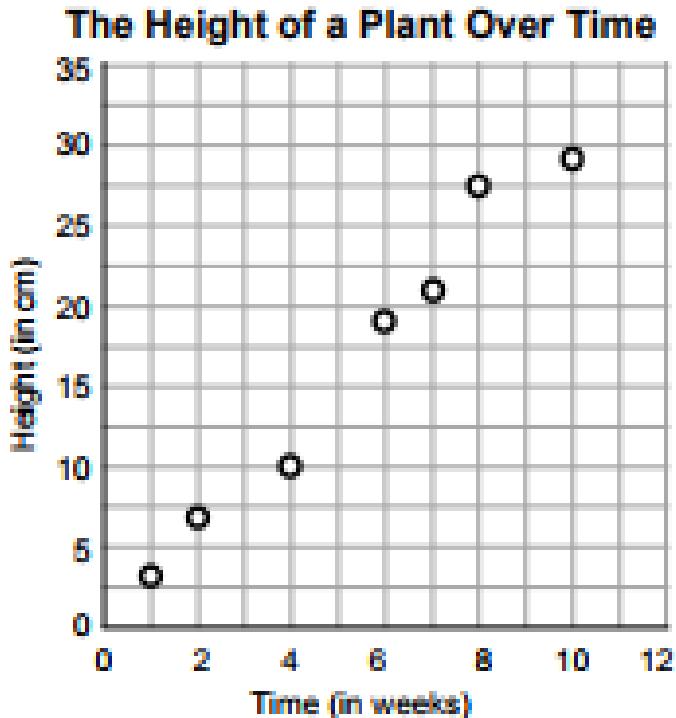
- The line through A and B has slope = \_\_\_\_\_ and y-intercept = \_\_\_\_\_.
- Write an equation for this line. \_\_\_\_\_
- Substitute each ordered pair from problem 3 into this equation to prove that they are on the line.

--	--	--

A \_\_\_\_\_ for a scatter plot is a straight line that best represents the data points in the scatter plot.

The table below shows bivariate measurement data of the height of a plant (in cm) over a period of weeks.

Time (in weeks)	Height (in cm)
1	3
2	7
4	10
6	19
7	21
8	25
10	29



1. Explain what the data shows.
2. Draw a straight line through (1, 3) and (7, 21). Does this line capture the overall trend in the data? Explain.
3. Use a colored pencil to draw a line through the points (2, 7) and (4, 10). Is this a good line of best fit? Explain.
4. Use a different colored pencil to draw a line through the points (3, 8) and (9, 27). Does this represent a good line of best fit? Explain.

5. Use the line through  $(1, 3)$  and  $(7, 21)$  to estimate how tall the plant will be after 12 weeks.

Use the same line to estimate how tall the plant will be after 50 weeks.

6. Use the coordinates  $(1, 3)$  and  $(7, 21)$  to find a linear function that could model the growth of the plant. Write the function in slope intercept form.

7. Explain what the slope of this line represents in the context of this problem.

8. Explain what the y-intercept of this line represents in the context of this problem.

Use your model (linear function) to answer the following questions.

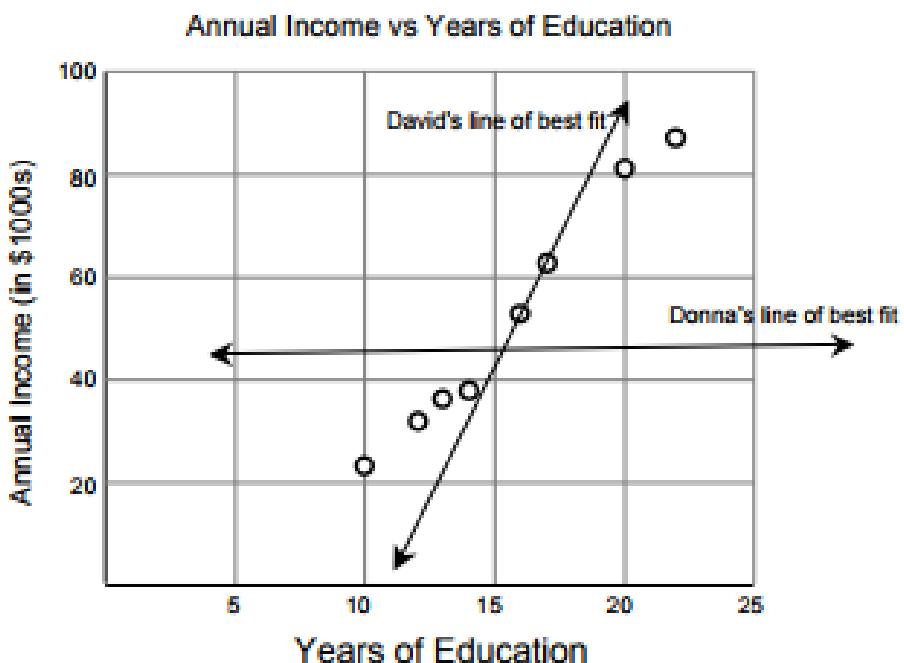
9. How tall will the plant be in 50 weeks.

10. How long will it take your plant to grow to height of 2 meters (200 cm)?

## HW #51 Line of Best Fit

Two students drew lines of best fit as show below.

Average Years of Education	Average Annual Income (in \$1000)
10	23
12	32
13	36
14	38
16	53
17	63
20	81
22	87

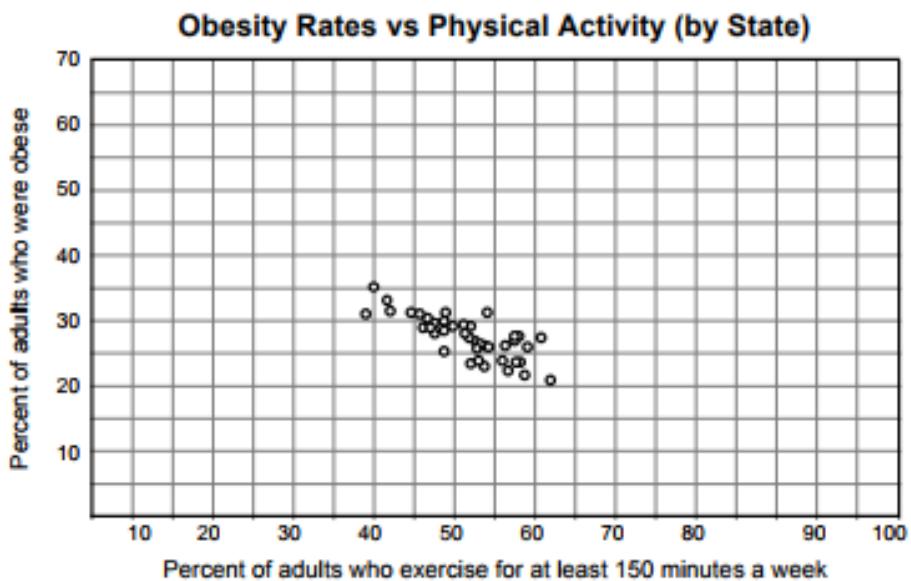


1. Why aren't either of the lines drawn considered the best fit? Explain.
2. Draw a line of best fit through the points (12, 32) and (22, 87). Then write the equation of the line in slope intercept form.
3. Explain what the slope represents in the context of this problem.
4. Use your model to predict the average income of someone with 28 years of education.

## Lesson #52 Line of Best Fit Continued

1. Draw a line of best fit through the points  $(40, 35)$  and  $(60, 25)$ .

2. Find the equation of the line.



3. Explain what the slope means in the context of the problem.

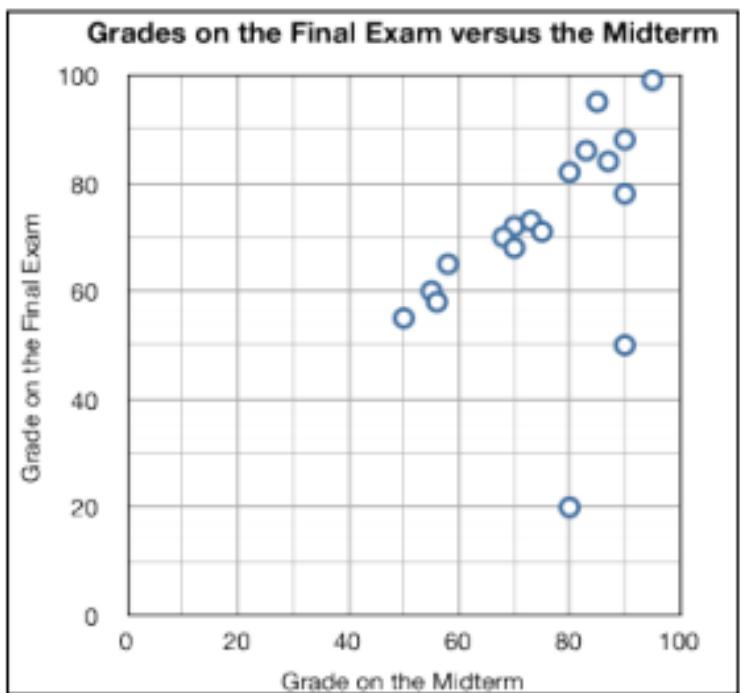
4. Use your equation to predict the obesity rate in a state where 50% of the population exercises.

5. The governor of Oregon wants to lower the obesity rate in her state to less than 20%. What does your model predict that the exercise rate should be for her to meet this goal?

An \_\_\_\_\_ of a data set is a data value that is unusually small or large relative to the overall pattern of values in the data set.

The graph on the right shows data for 18 math students (grades on the final exam versus their grades of the midterm exam).

1. What does the data point (95, 88) mean in the context of this problem?
2. Circle any data points that appear to be potential outliers. What effect do these potential outliers have on the pattern of association in the data?
3. If you ignore these potential outliers, what happens to the apparent strength of the association?
4. Draw a line of best fit.
5. What does the slope of the line mean in the context of this problem.



## HW #52 Line of Best Fit Continued

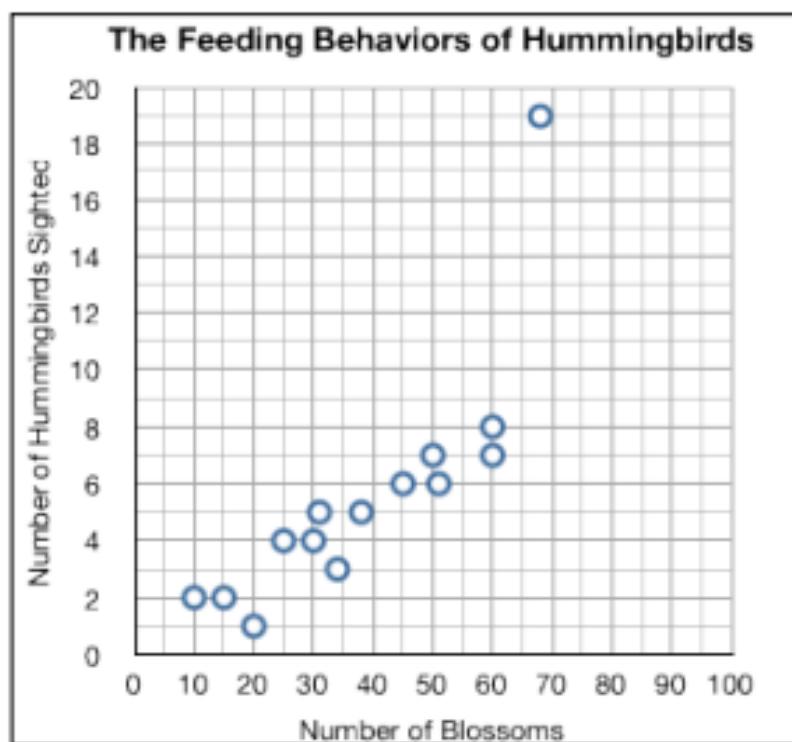
Billy is conducting a study on the feeding habits of hummingbirds. For an hour each day, he counted the number of flower blossoms in his garden and the number of different hummingbirds he sees.

His data set is graphed on the scatter plot below.

1. Does there appear to be a potential outlier in Billy's data?

Explain.

2. Ignore any potential outliers, describe any association in the data.



3. Draw a line of best fit on your graph through the points (10, 2) and (60, 8). Determine the equation of the line.

4. Interpret the slope in the context of the problem.

5. Use your line of best fit to answer the following questions:

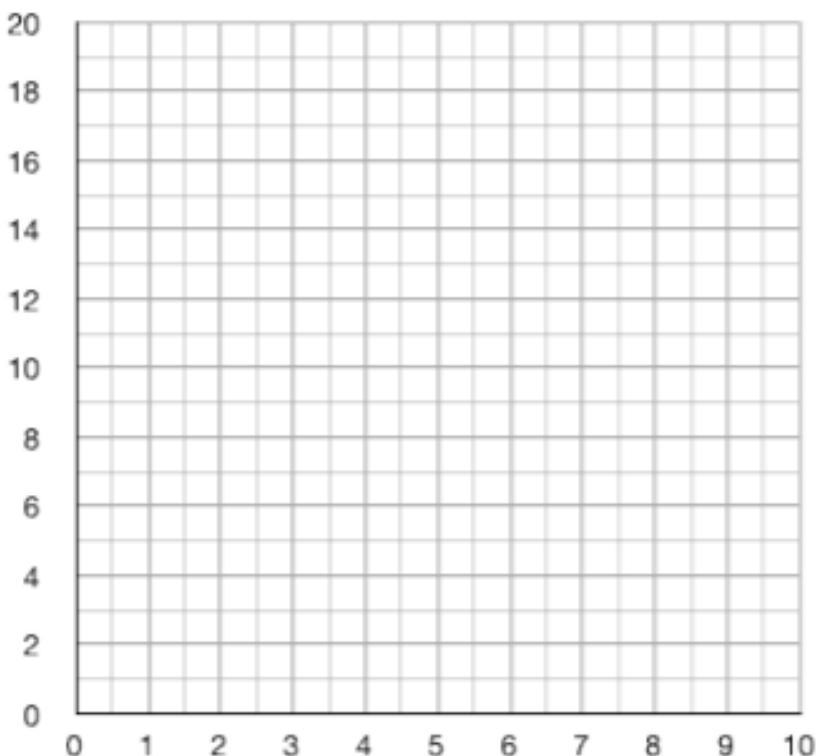
- a. How many hummingbirds can Billy expect to see if he counts 80 flower blossoms?

- b. How many flower blossoms can Billy expect to count if he sees 3 hummingbirds?

## Lesson #53 Scatter Plots Continued

Ricardo is doing research about water usage at a summer math camp for high school students. He measured the depth of the water in the well each week. His data is recorded in the table below.

Weeks	Depth of Water in the Well (in feet)
1	18
2	15
3	14
4	10
5	9

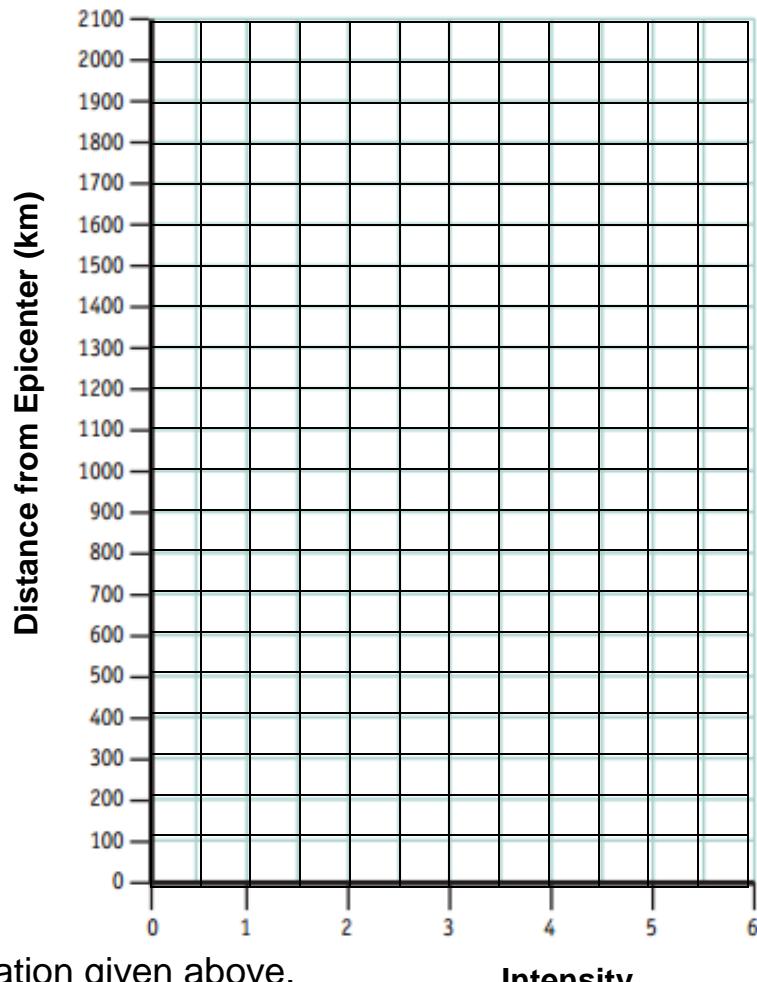


1. Plot the measurement data on the graph. Make sure you label each axis appropriately and give your graph and title.
2. Describe any patterns you see and make a conjecture about what is happening to the depth of the water.
3. Draw a line of best fit through points (1, 18) and (4, 10). Find the equation of the line. What is the equation of the line?
4. Interpret the slope in the context of the problem.
5. If the summer camp goes on for a total of 10 weeks, will there be enough water for the campers? Use your function to make a prediction.

## HW #53 Scatter Plots Continued

Actuaries use information about the magnitude and severity of earthquakes in a particular geographic area to help insurance companies determine how to serve their customers. The magnitude describes the size of the earthquake at the source. The distance is from Parkfield, CA, a town famous for its seismic activity.

Intensity	Distance from Epicenter (km)
IV	89
V	35
I	934
V	47
I	1431
IV	99
II	497
I	2019
III	482
III	778
III	401
IV	294
II	930
II	781
V	129



1. Create a scatterplot of the information given above.
2. Find a possible line of best fit. What is the function rule for the line?
3. What association do you notice based on the information plotted here? Is the relationship weak or strong?
4. Do you think it would cost more or less to insure customers in the Parkfield region against earthquake damage than it would customers in places where there are no recorded earthquakes? Why?

## Lesson #54 Scatter Plots Continued

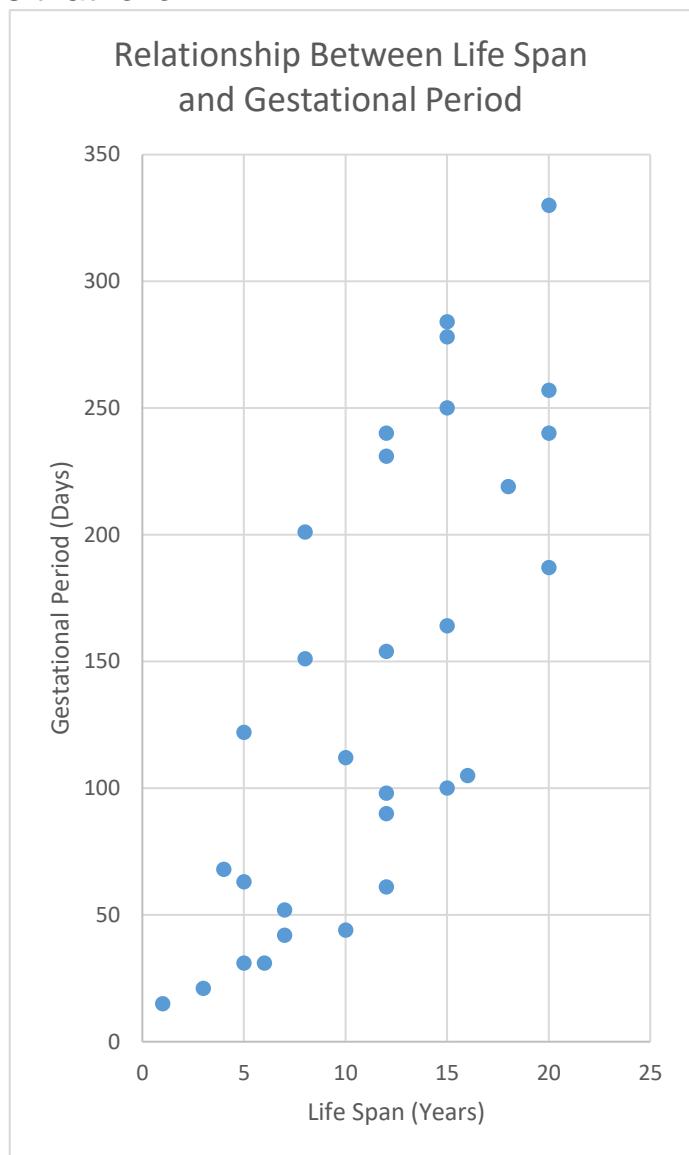
Below is a scatter plot that shows the relationship between life span and gestational period for 30 mammals. Answer the questions that follow.

1. Describe the association between life span and length of gestation period.

2. Draw a line of best fit through points  $(5, 30)$  and  $(20, 330)$ . What is the equation of the line?

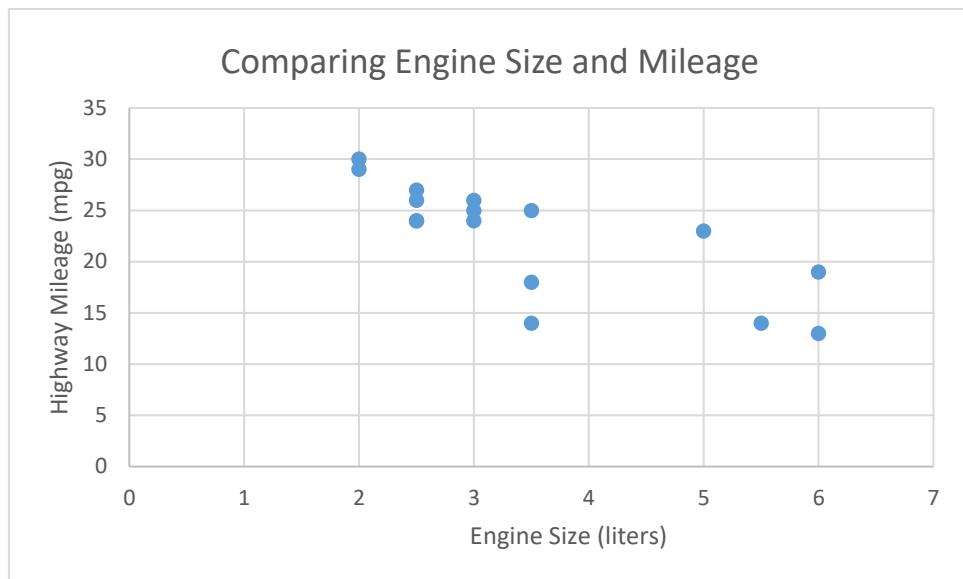
3. Describe the slope in the context of the problem.

4. Use your model to predict the gestational period of a mammal with a life span of 28 years.
5. Use your model to predict the life span of a mammal with a gestational period of 400 days.



## HW #54 Scatter Plots Continued

At right is a scatterplot that shows the relationship between engine size and highway mileage.



1. Describe the association between engine size and highway mileage.
2. Draw a line of best fit through the points (2, 29) and (6, 13). What is the function rule for that line?
3. Interpret the slope in the context of the problem.
4. Use your model to predict the highway mileage of a vehicle with an 8 liter engine.
5. Use your model to determine the engine size of a vehicle that gets 32 mpg.