Name _____ Period _____ Date _____

UNIT 6
STUDENT PACKET

Math Links
GRADE 8

BIVARIATE DATA

Parent (or Guardian) signature _____

MY WORD BANK

Explain the mathematical meaning of each word or phrase, using pictures and examples when possible. See Student Resources for definitions and examples.

association	bivariate data
categorical data	frequency table relative frequency table
line of best fit	numerical data measurement data
outlier	two-way table

STACKING CUPS

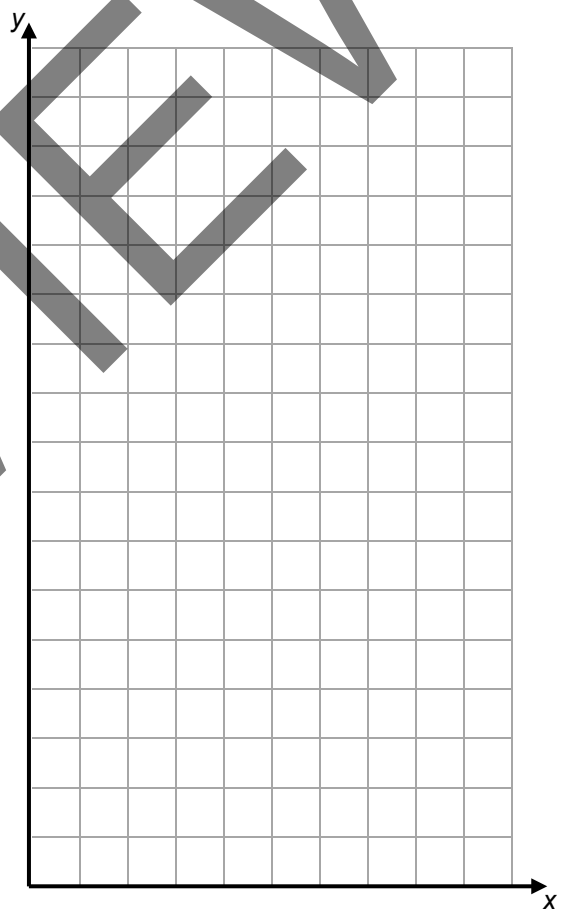
Follow your teacher's directions for (1) – (3).

(1)

Number of cups (*x*)		1	2	3	4	5	6
Height in cm (*y*)							

(2) Graph the data.

(3)



STACKING CUPS
Continued

4. The independent variable (input) is…

   The dependent variable (output) is…

5. Describe the height increases in words.

6. Approximate the slope of the line drawn.

7. Write an input-output rule for the line drawn (an equation in *x* and *y*) that relates height to number of cups stacked.

8. Use your rule to estimate the height of a stack of 100 cups.

9. Even though you sketched a line to approximate the data, why do ordered pairs that correspond to non-whole-number *x*-values make no sense in this context?

NUMERICAL DATA

We will construct scatter plots for bivariate numerical data, investigate patterns of association, and interpret the data.

[8.SP.1; SMP2, 3, 4, 8]

GETTING STARTED

Use the data below for this page and the next.

Level of Education	**Average Years of Education** (“input”)	**Average Annual Income** in $1,000s (“output 1”)	**Unemployment Rate** in % (“output 2”)
< high school diploma	10	32.2	11.7
high school diploma	12	40.6	9
some college, no degree	13	45.6	8.3
associate’s degree	14	48.8	7.1
bachelor’s degree	16	67.9	5.5
master’s degree	18	80.3	4.1
doctoral degree	20	98	3.1

(US Bureau of Labor Statistics, 2020)

1. Explain what each ordered pair below means in the contexts given above.

a. (input, output 1) ⟶ (12, 40.6)	b. (input, output 2) ⟶ (12, 9)
c. (input, output 1) ⟶ (16, 67.9)	d. (input, output 2) ⟶ (16, 5.5)

2. Record the meanings of numerical data, measurement data, and bivariate data in **My Word Bank**.

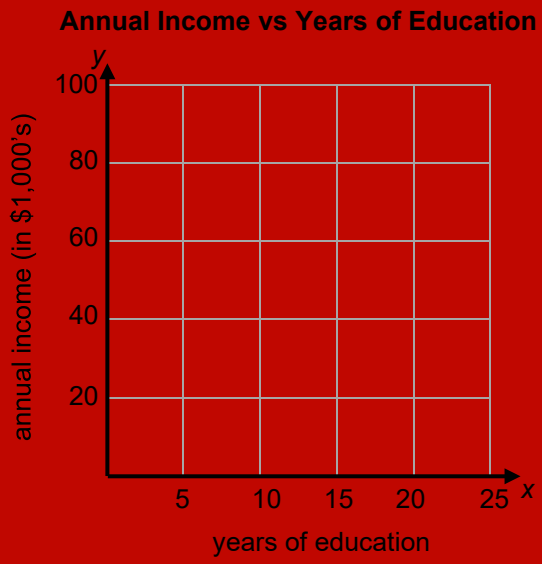
3. Was the data you collected in **Stacking Cups** numerical data? _____

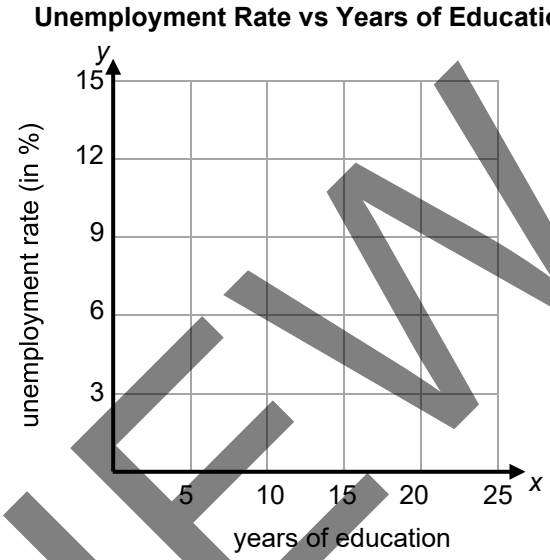
   Measurement data? _____ Bivariate data? _____

LINEAR ASSOCIATION

Follow your teacher's directions for (1) – (3)

(1)



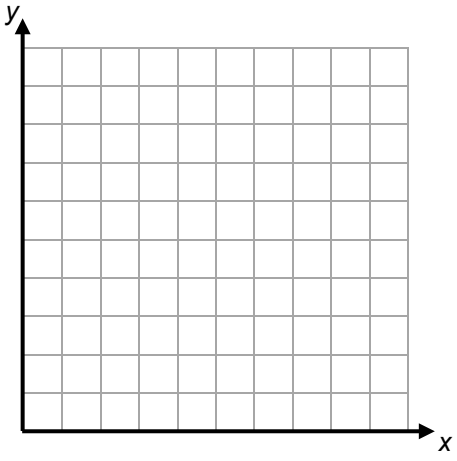


(2)

(3)

4. Make a scatter plot of the data below.

Height of Student (in cm)	Score on Math Test (as a %)
150	45
130	30
140	70
120	90
180	40
100	80
150	95
120	55



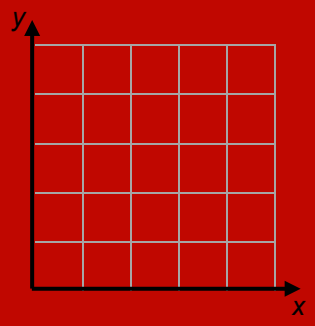
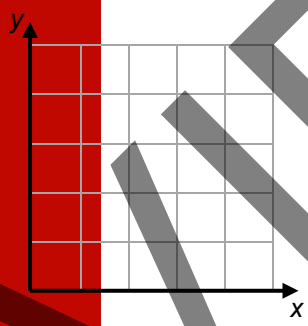
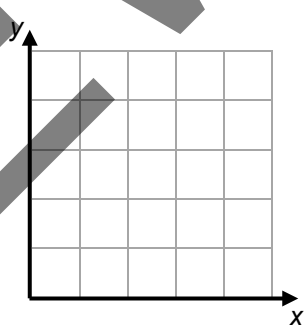
We say the association is weak or does not exist if the data "clusters in a cloud."

5. Explain what you think the association is between the math test score data and the student height data.

6. Record the meaning of association in **My Word Bank**.

PRACTICE 1

1. Look at the sets of (x, y) ordered pairs below, all without contexts. Predict the kind of association each has, if any, by observing patterns in the data. Graph points to verify predictions.

Set 1	Set 2	Set 3
(0, 5) (0.5, 4) (1, 4.2) (1.5, 4) (2, 3.5) (3, 3.8) (3.5, 2.6) (4, 1.9) (5, 1.8)	(0, 3.5) (0.5, 1) (1, 4.8) (1.5, 4.2) (2, 1.5) (3, 0) (3.5, 3.5) (4, 1.1) (5, 3.2)	(0, 0.5) (0.5, 0.9) (1, 1.1) (1.5, 2.5) (2, 2) (3, 2.8) (3.5, 2.6) (4, 3.9) (5, 4.1)
Prediction: (before graphing)	Prediction: (before graphing)	Prediction: (before graphing)
 Verification: (after graphing)	 Verification: (after graphing)	 Verification: (after graphing)

2. Examine the pairs of variables below, all without numerical data. Describe the kind of association you might expect.

Variables	Kind of association
a. Number of sodas consumed and number of cavities	
b. Age and quality of eyesight	
c. Number of traffic tickets and cost of car insurance	
d. Shoe size and number of pets at home	
e. Life expectancy and annual income	

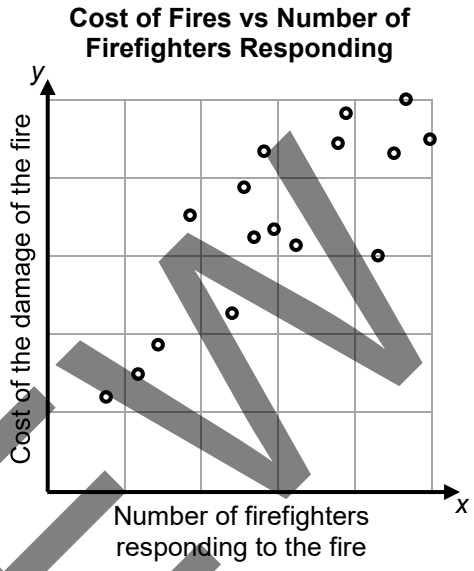
3. Write pairs of variables with each kind of association. (Create examples you've not already seen in this unit.)

Positive	Negative	Weak or None

ASSOCIATION AND CAUSATION

Follow your teacher's directions for (1) – (2).

(1)



(2)

The statements in problems 3 – 5 confuse association with causation. For each statement, list other factors that *might* be a cause for the association. Then discuss with others.

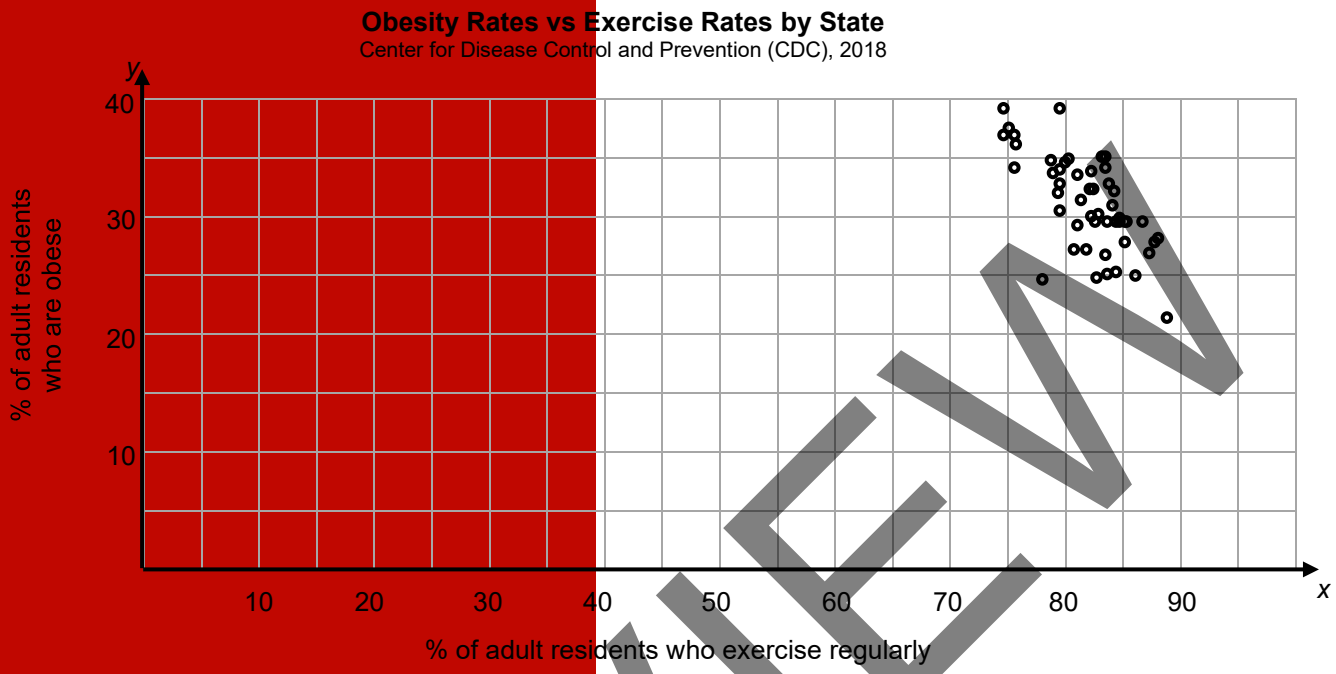
3. People who walk faster tend to live longer. Therefore, if you want to live longer, walk faster.

4. As sales of ice cream increase, the rate of drowning deaths increases. Therefore, ice cream causes people to drown.

5. The more cell phones a country has as a percentage of population, the longer the life expectancy of people in that country. Therefore, cell phones cause you to live longer.

6. Explain what the phrase "association does not imply causation" means in your own words.

PRACTICE 2



1. List the variables compared on the graph above.

2. How many data points do you think are on the scatter plot? Does the data appear to cluster anywhere?

3. Explain what each data point below means in context.

Colorado (83.8, 22.7)	Mississippi (69.1, 39.5)

PRACTICE 2
Continued

4. Explain what stands out regarding the data points for the two states above.

5. What appears to be the relationship between exercise and obesity? Include the type of association you observe, if any.

6. What might be some reasons for high rates of obesity in the United States?

LINES OF BEST FIT

We will construct linear models for measurement and other numerical data clustered around a straight line. We will draw lines of best fit, estimate their equations, interpret slope and *y*-intercept in context, and use the equations as models to make predictions. We will explore the effects of potential outliers on data. We will observe associations that are not linear.

[8.SP.1, 8.SP.2, 8.SP.3, 8.F.4, 8.F.5, SMP1, 2, 3, 4, 5, 6, 7]

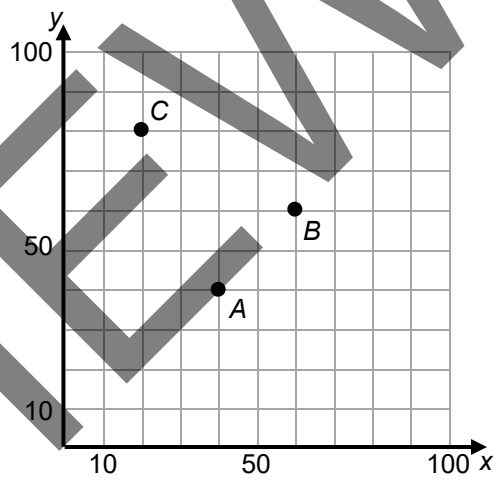
GETTING STARTED

1. Write the (*x*, *y*) coordinates for each point.

   A (_____, _____)

   B (_____, _____)

   C (_____, _____)



2. Use a straightedge to draw lines $\overleftrightarrow{AB}$, $\overleftrightarrow{AC}$, and $\overleftrightarrow{CB}$. Extend lines to cross the *x*-axis and *y*-axis when possible.

3. Fill in the missing information for each line.

	$\overleftrightarrow{AB}$	$\overleftrightarrow{CB}$	$\overleftrightarrow{AC}$
three points on the line (use the graph)	(50, _____) (100, _____) (0, _____)	(40, _____) (80, _____) (0, _____)	(30, _____) (60, _____) (0, _____)
***y*-intercept**			
slope			
equation (in slope-intercept form)			
another point on the line (use the equation)	(-53, _____)	(-12, _____)	$\left(-\frac{1}{2}, \_\_\_\_\_\right)$

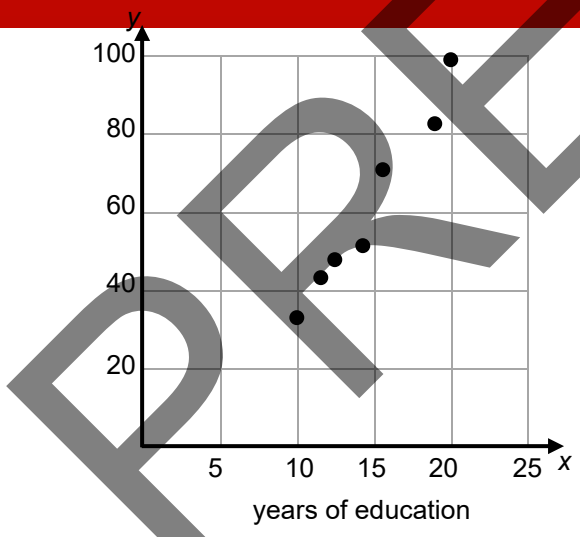
EDUCATION DATA REVISITED

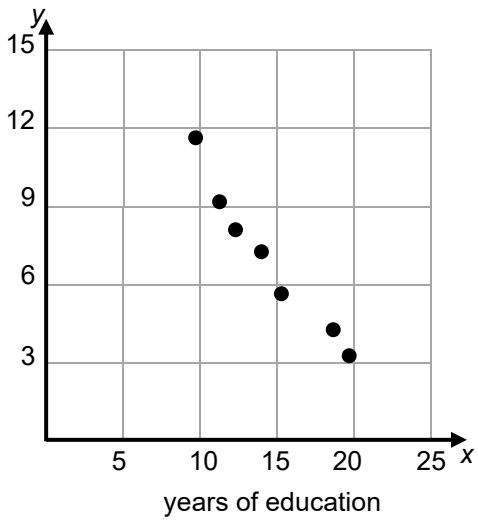
Follow your teacher's directions using data and graphs from the previous lesson.

Level of Education	Average Years of Education	Average Annual Income in $1,000s	Unemployment Rate in %
< high school diploma	10	32.2	11.7
high school diploma	12	40.6	9
some college, no degree	13	45.6	8.3
associate's degree	14	48.8	7.1
bachelor's degree	16	67.9	5.5
master's degree	18	80.3	4.1
doctoral degree	20	98	3.1

(1)

(2)





ANALYZING EDUCATION DATA

1. Compare your linear model with up to three classmates. Then answer the following.

   a. Is your line of best fit in the exact same place as theirs on the graph?

   b. Are your values for the slope and *y*-intercept exactly the same?

   c. Is it possible that there could be more than one line of best fit that is accurate enough? Explain.

2. Explain what the slope represents in the context of this problem for the graph on the left.

   According to your model, how much is a year of school worth in terms of a year's income?

3. Explain what the slope represents in the context of this problem for the graph on the right.

   According to your model, how much does a year of school affect employment?

ANALYZING EDUCATION DATA
Continued

4. Is it reasonable to use the *y*-intercept to make predictions about the income of a person with zero years of education? Explain.

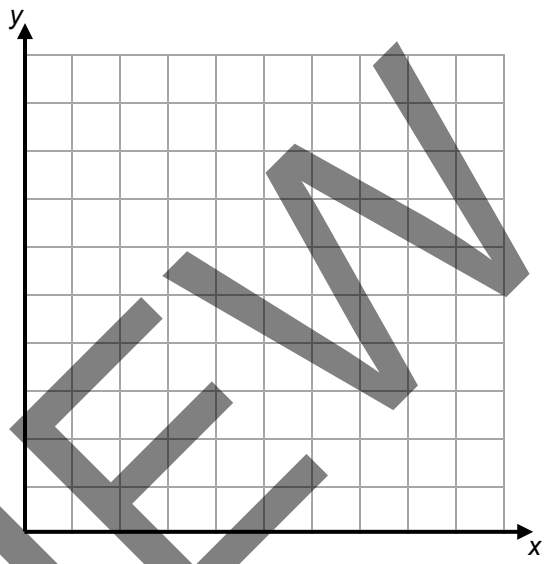
5. Madison graduates from high school at age 18, and plans to work for 50 years, earning an average of $40,600 per year. Piper graduates from college at age 22, and plans to work for 46 years, earning an average of $67,900 per year. Compare their lifetime expected earnings. What are some other factors you might want to consider in this analysis?

6. Record the meaning of line of best fit in **My Word Bank**.

PRACTICE 3

The table shows bivariate measurement data of a plant's growth.

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



1. Label and scale the grid.
   Then graph the data from the table.

2. Explain what you think the association is between the variables and what it means.

3. Draw an estimate for a line of best fit. Then write the following.

   a. An estimate for the slope of the line.

   b. An estimate for the *y*-intercept.

   c. The equation of this line in slope-intercept form.

4. Explain what each represents in the context of the problem.

a. The slope	b. The *y*-intercept

5. Use your equation to predict…

a. plant height at 40 weeks.	b. time for the plant to grow to 100 cm.

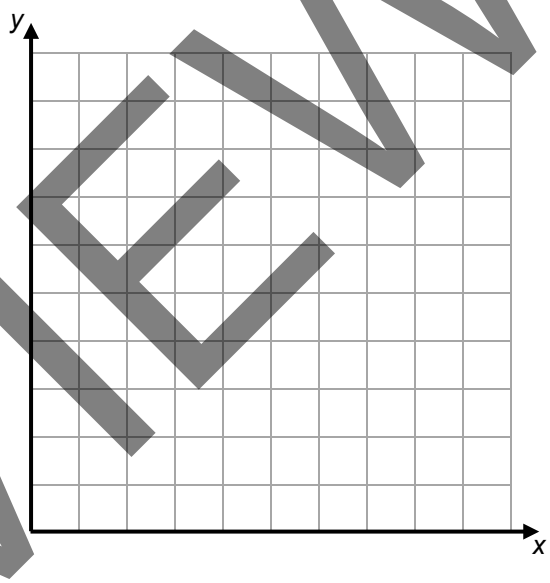
6. What are some issues that could arise by using a line of best fit to predict the height of the plant after large amounts of time passing, like 100 weeks?

PRACTICE 4

1. Go back to **Stacking Cups** and revisit the equation you wrote in problem 7. Explain why you think this is a line of best fit, or improve your estimate.

Maxine measured and recorded the height of a bowl. Then she placed a second bowl inside the first, measured and recorded the new height, and continued this process a few more times.

Number of Bowls (x)	1	2	3	4	5	6
Height in cm (y)	4.2	6.3	8.2	10.3		



2. Write reasonable heights for 5 and 6 bowls in the table. Label and scale the graph. Graph the six points.

3. Explain what you think the association is between the variables and what it means.

4. Estimate a line of best fit (draw it) and write its equation.

5. Explain what each represents in the context of the problem.

a. The slope	b. The y-intercept

6. Use your equation to predict…

a. the height of 20 bowls.	b. number of bowls it takes to reach 100 cm.

OBESITY RATES BY STATE

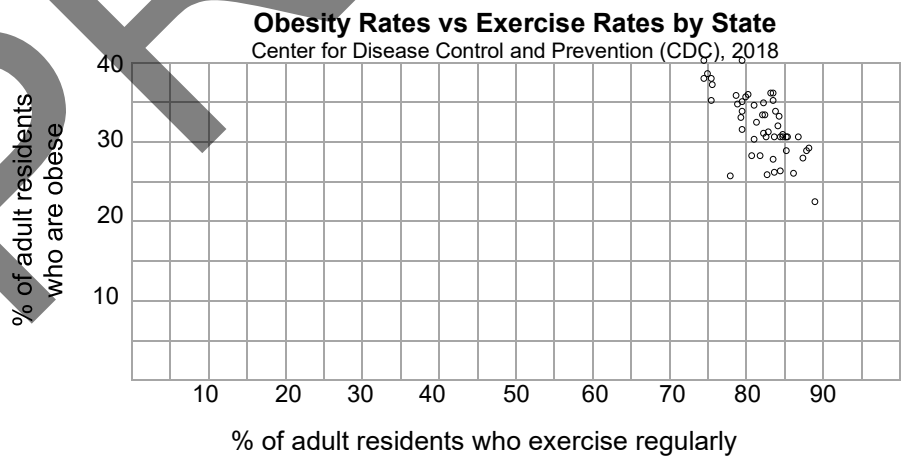
The graph at the bottom of the page is from **Practice 2**. Use it to answer the following questions.

1. Estimate (draw) a line of best fit on the graph below and write its equation. Since the *y*-intercept is above the graph, you may want to use a ruler.

2. Explain what the slope and *y*-intercept represent in the context of the problem.

3. Does the scatterplot represent a function? _____ The line of best fit? ____

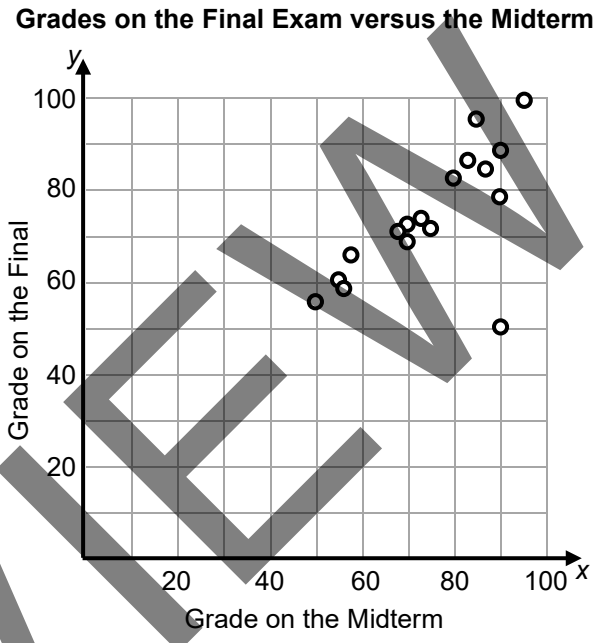
4. Use your equation to predict the obesity rate in a state where 50% of the population exercises.	5. The CDC recommends to governors of all states to set a goal of 20% for their obesity rates. What does your model predict that the exercise rate should be to meet this goal?



OUTLIERS

1. You may have learned about outliers in a previous grade. Record the meaning of outlier in **My Word Bank**.

The graph to the right shows data for 18 high school math students in one class.



2. What does the data point (80, 82) mean in the context of this problem?

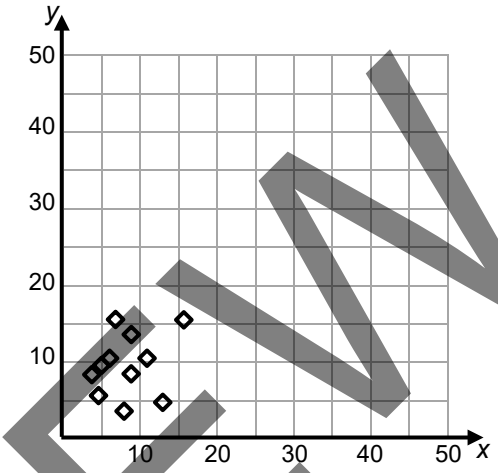
3. Circle the data point that appears to be a potential outlier. If you removed this data, what happens to the apparent strength of the association?

4. Draw a line of best fit on the graph that excludes the potential outlier and write its equation.

OUTLIERS
Continued

Potential outliers can also trick us into seeing patterns that are not really there.

5. Do you think the data graphed on the right shows a strong association?



6. Graph the points (45, 45) and (35, 40).

   Explain whether these new points appear to be potential outliers, and what effect, if any, they might have on the appearance of an association.

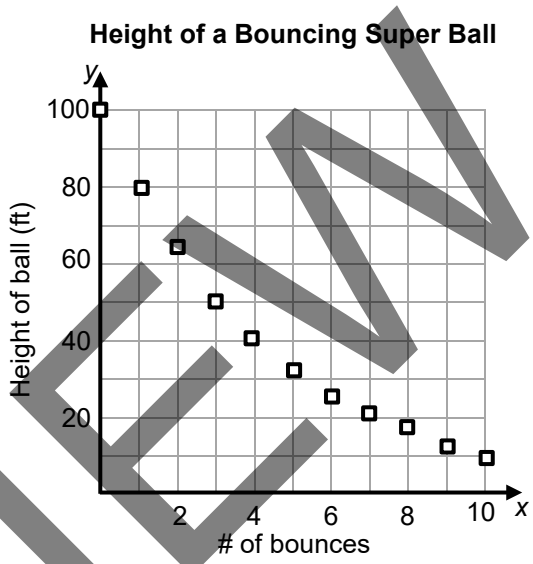
7. Make up a title for the graph above and label the axes in a way that the data might fit your context.

NONLINEAR ASSOCIATIONS

On this page and the next are two examples of bivariate measurement data that are not linear.

The graph below shows the height of a bouncing ball measured at the top of each bounce.

1. Explain what the data point (2, 64) means in the context of the problem.



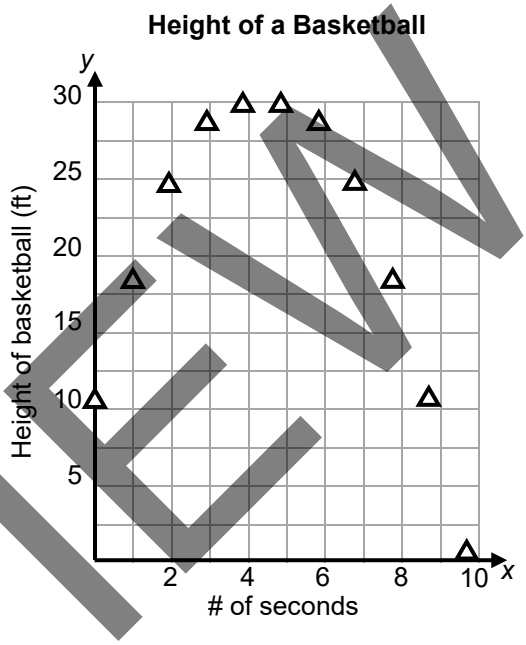
2. Explain what the data says. Include why there is an association, but it is not linear.

3. Do there appear to be any outliers in the data?

NONLINEAR ASSOCIATIONS
Continued

The graph below shows the height of a basketball measured over a period of time after it is thrown in the air.

4. Explain what the data points (0, 10) and (9, 10) mean in the context of the problem.



5. Explain what the data says. Include why there is an association, but it is not linear.

6. Do there appear to be any outliers in the data?

7. Choose a graph. Explain why you think it is realistic (or not).

CATEGORICAL DATA

We will use two-way tables to display the frequencies and relative frequencies of categorical data. We will examine patterns of association in bivariate categorical data, and draw conclusions about possible associations.

[8.SP.4; SMP2, 3, 4, 7]

GETTING STARTED

1. Record the meaning of frequency table and relative frequency table in **My Word Bank**.

Below is the start of two tables created in Ms. Costello's 8th grade math class. Recall that *n* is used to represent the total number in a population.

2. Complete Table I below.
   a. How did you determine the number of students in the class?

   b. Explain why the percent total in Table I is 100% (or very close to it).

3. Complete Table II below.
   a. Can a student have more than one type of pet? _____
   b. The sum of the 2nd column is _____. How does this value compare to *n*?

   c. Explain why the percent total (the sum of the 3rd column) in Table II is not 100%?

Table I
How many dogs do you own?

Number of Dogs	Number of students *n* = _____	Percent of students
0	20	
1	8	
2	2	
3 or more	0	
Total		

Table II
What pets do you own?

Animal	Number of students *n* = _____	Percent of students
Cats	6	
Birds		10%
Dogs	10	
Other	1	
None		50%

WHAT IS CATEGORICAL DATA?

Follow your teacher's directions for (1).

(1)		
Data type	**A survey question**	**Possible responses**
Categorical data		
Numerical data		

Complete each table by writing appropriate survey question(s) and possible responses.

2.	**MUSIC PREFERENCES**	
Data type	**A survey question**	**Possible responses**
Categorical data		

3.	**INVOLVEMENT IN THE PERFORMING ARTS**	
Data type	**A survey question**	**Possible responses**
Numerical data		

4.	**STREAMING VIDEOS**	
Data type	**A survey question**	**Possible responses**
Categorical data		

Make up your own topic.

5.		
Data type	**A survey question**	**Possible responses**
Categorical Data		
Numerical data		

6. Record the meaning of categorical data in **My Word Bank**.

TWO-WAY TABLES

Follow your teacher's directions for (1) – (3). Round to the nearest percent.

(1)

Table I

(2)

(3)

Table II

(*n* = _____)			

State which table has the best information for the question. If a calculator computation is required, write the expression along with the answer.

4. How many are male?	5. What percent are male?	6. How many lived?
7. What percent lived?	8. Out of all the males, what percent lived?	9. Estimate the ratio of males to females who died?

10. What conclusions can you draw from the data. Use the data to explain whether it was better to be male or female.

11. Record the meaning of <u>two-way table</u> in **My Word Bank**.

PRACTICE 5

Ten 8th graders were asked the following questions:

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

Data was collected on their responses and recorded.

	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 data above. Complete the tables below.

Table I: Curfews and Chores Frequency Table

	Students with Curfew	**Students with No Curfew**	**Total**
Students with Chores			
Students with No Chores			
TOTAL			

Table II: Curfews and Chores Relative Frequency Table

(*n* = ____)	**Students with Curfew**	**Students with No Curfew**	**Total**
Students with Chores			
Students with No Chores			
TOTAL			

In your tables:

2. Draw a circle around the total number of students.
3. Draw a triangle around the total number of students with chores.
4. Draw a square around the percent of students with no curfew.
5. Draw a trapezoid around the percent of students who had chores and a curfew.

PRACTICE 5
Continued

Complete the following using the bivariate categorical data on the previous page. Write which table has the best information for the question. If a computation is required, write the expression along with the answer.

6. How many students had neither chores nor curfew?	7. What percent of students had no chores and no curfew?
8. What percent of students had no chores?	9. What percent of students who had no chores also had no curfew?
10. What percent of students who had no curfew also had no chores?	11. What percent of students who had a curfew also had chores?

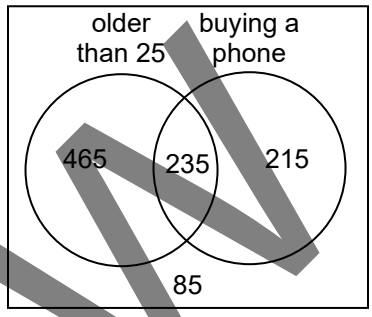
12. Vinnie says that according to Table II, only 10% of the students have no chores. What is mistaken with Vinnie's reasoning?

13. Raji thinks that students who don't have chores are more likely not to have a curfew either. Does the data support Raji's claim? Explain.

14. Do you think there might be some associations in this data? Explain.

A MARKETING DECISION

You are part of a marketing team that is planning a cell phone advertising campaign in movie theaters. You collect data to determine whether to pitch your campaign to a younger audience or an older one. You asked the following questions of 1,000 movie-goers, and recorded the data in a Venn Diagram (to the right).



- Will you buy a new phone within the next year?
- Are you older than 25?

1. Use data from the Venn Diagram to complete the two-way frequency table.

	Age > 25	**Age ≤ 25**	**Total**
buy phone			
won't buy phone			
Total			1,000

2. Write at least four statements that might help you determine whether to pitch to a younger audience or an older one. Clearly show how you used the data. Use frequencies and relative frequencies to formulate your statements.

3. Use your statements and data to explain whether you think the cell phone company should advertise in movie theaters targeting a younger audience or an older audience.

REVIEW

POSTER PROBLEMS: BIVARIATE DATA

Part 1: Your teacher will divide you into groups.

- Identify members of your group as A, B, C, or D.
- Each group will start at a numbered poster. Our group start poster is _____.
- Each group will have a different colored marker. Our group marker is _____.

Part 2: Do the problems on the posters by following your teacher's directions.

Poster 1 (or 5)	Poster 2 (or 6)	Poster 3 (or 7)	Poster 4 (or 8)
(1, 6) (2, 6) (3, 6)	(1, 55) (2, 60) (3, 40)	(1, 70) (2, 60) (3, 80)	(2, 0) (3, 30) (4, 27)
(4, 12) (5, 9) (6, 6)	(4, 40) (5, 40) (6, 25)	(4, 70) (5, 100) (6, 80)	(4, 30) (5, 30) (5, 33)
(7, 18) (8, 12)	(7, 25) (8, 15)	(7, 100) (8, 100)	(8, 12) (8, 9)
(9, 21) (10, 24)	(9, 5) (10, 20)	(9, 90) (10, 120)	(9, 9) (9, 12)
(11, 18) (12, 18)	(11, 10) (12, 0)	(11, 90) (12, 120)	(10, 9) (12, 36)

A. Number the axes and graph the ordered pairs.

B. Draw a line of best fit and estimate an equation in slope-intercept form (if possible).

C. Explain the association in words.

D. Come up with a reasonable context, and write an appropriate title and axes labels.

Part 3: Return to your seats. Work with your group, and show all work.

Go back to your start problem.

1. Write a few comments critiquing the answers on the poster.

2. Offer an alternative context for the data.

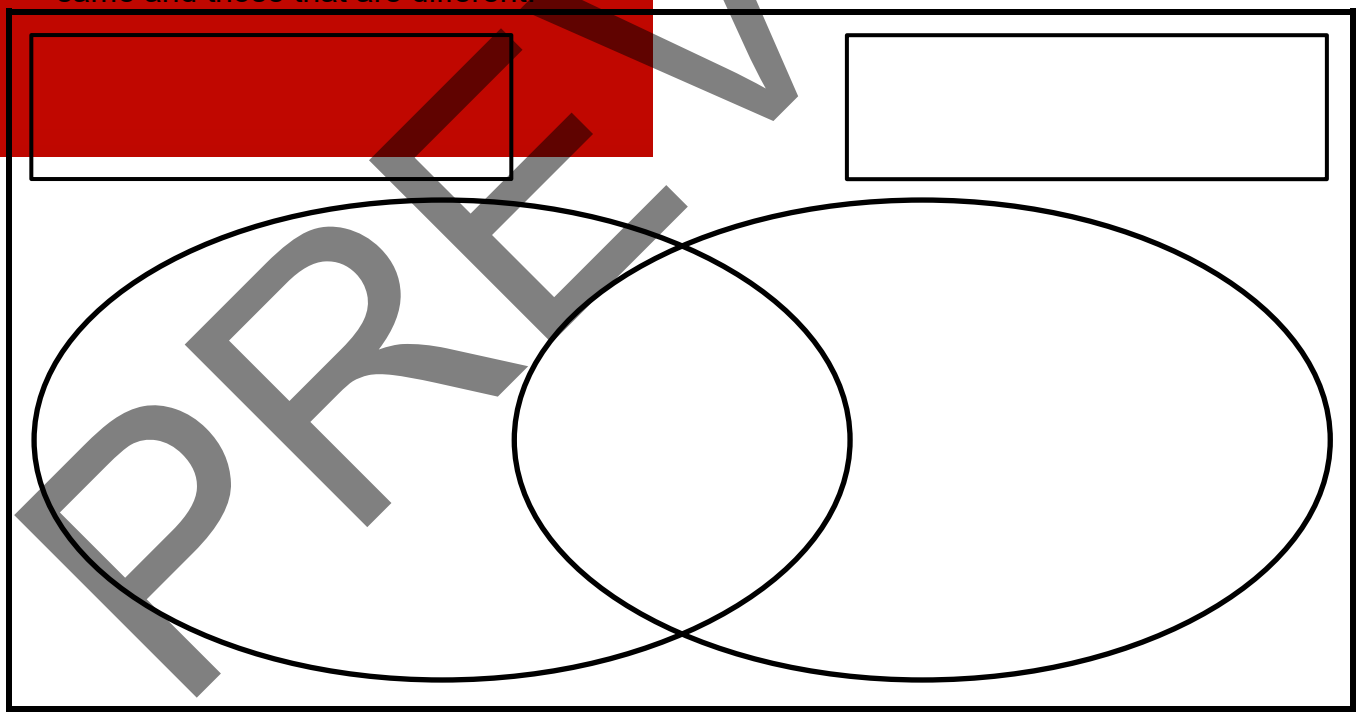
MATCH AND COMPARE SORT: BIVARIATE DATA

Your teacher will give you some cards. Cut them out.

1. Individually, match words with descriptions. Record results.

Card set △			Card set ◯		
Card number	**word**	**Card letter**	**Card number**	**word**	**Card letter**
I			**I**		
II			**II**		
III			**III**		
IV			**IV**		

2. Partners, choose a pair of numbered matched cards and record the attributes that are the same and those that are different.



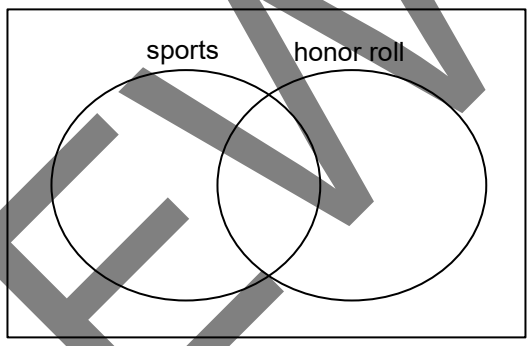
3. Partners, choose another pair of numbered matched cards and discuss the attributes that are the same and those that are different.

FIGHTING STEREOTYPES

Jose hears frequently that athletes don't perform well academically, but he believes that all students can be successful in school. He is disturbed by the hurtful stereotypes of certain populations.

Jose collected some data at the local high school to try to find an association between participation in school sports and being on the honor roll. Students were asked at random:

- Did you play a sport last semester?
- Did you make honor roll last semester?



1. Each member of your group will get one or more cards with information on it. You may NOT show cards to anyone, but you may read yours aloud as many times as needed. Together, fill in all of the missing information in the Venn diagram at the right. Then complete the two-way tables below.

Table I: **Academics and Athletics** (frequencies)

	sport	**no sport**	**Total**
honor roll			
no honor roll			
Total			

Table II: **Academics and Athletics** (relative frequencies)

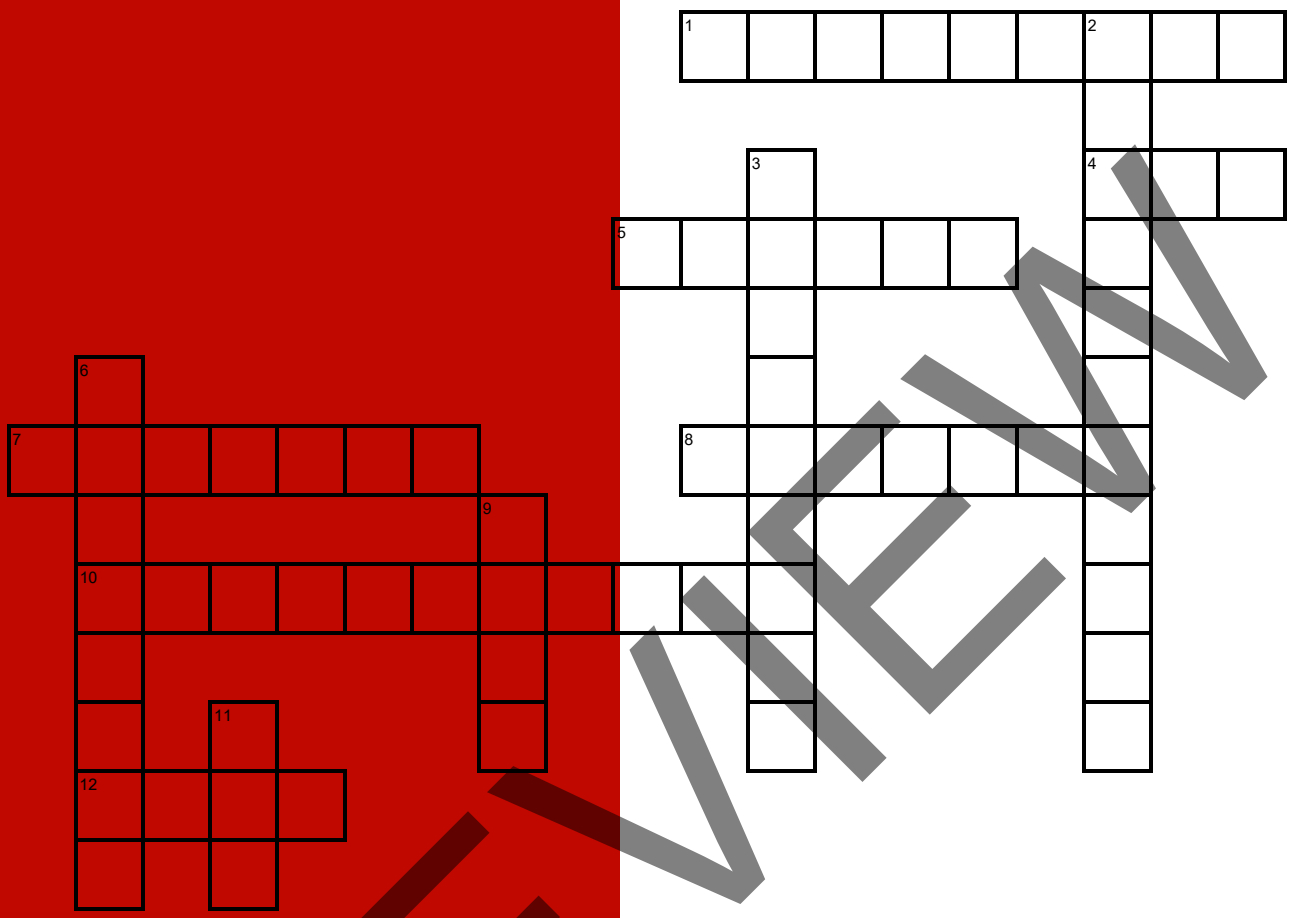
(*n* = _____)

	sport	**no sport**	**Total**
honor roll			
no honor roll			
Total			

2. What associations, if any, do you see in the data? What conjecture might you make about how well athletes do academically at this school? Clearly show how you used the data. Use frequencies and relative frequencies to formulate your statements.

3. Are there other stereotypes you hear about that you think ought to be questioned?

FOCUS ON VOCABULARY



Across

1 data whose responses are numbers

4 number of variables in bivariate data

5 questionnaire used to collect data

7 line that approximates a linear relationship (2 words)

8 unusually large or small value in a data set

10 relationship among variables (it does not imply causation)

12 a sorting circle diagram

Down

2 data sorted by attributes

3 table that lists counts from a population

6 type of frequency table that displays percents

9 analyzed with statistics or displays

11 number of variables in univariate data

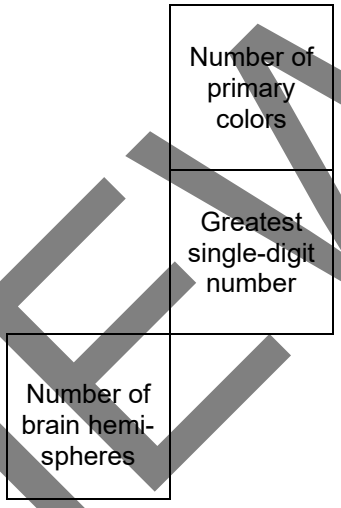
SPIRAL REVIEW

1. **Alge-Grid: What's the *a*?** Each clue gives the value of a corresponding cell. Use clues to find *a*, which has the same value in all cells. Once evaluated, the cells will contain the whole numbers 1 – 9, exactly once each.

The Alge-Grid

$a + 3$	$[9(a-1)]^{\frac{1}{3}}$	$\sqrt{a} - 1$
$a \times a^0$	$(a-1)^2$	$a^2 - 2a - 2$
$\frac{1}{2}a$	$(a-2)^3$	$a + 1$

The Clues



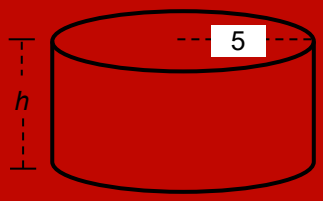
2. Solve each equation. Check by substitution in to the original equation.

a. $-4(g - 8) = 28$	b. $54 = -\left(\frac{x}{4} - 6\right)$
c. $\frac{m-8}{4} = -4$	d. $3.2(p - 5.5) + 2p = -33.2$

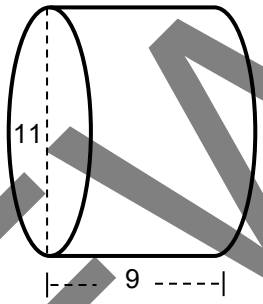
SPIRAL REVIEW
Continued

Find the indicated measurements.

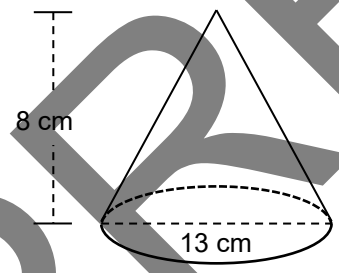
3. A cylinder has a volume of 706.5 ft^3 and radius of 5 ft. Find the height. Use $\pi = 3.14$.



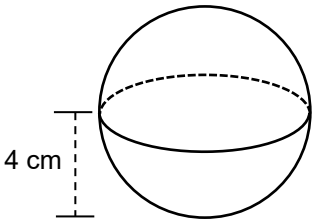
4. A cylinder has a diameter of 11 cm and height 9 cm. Find its volume. Leave in terms of π.



5. A cone has dimensions shown. Find its volume. Use $\pi = 3.14$.



6. A sphere has a radius of 4 cm. Find its volume. Use $\pi = 3.14$.



SPIRAL REVIEW
Continued

7. Compute.

a. $2^3 \bullet 2^{-2}$	b. $15^{-9} \bullet 15^9$	c. $3^3 \bullet 3^{-2}$
d. $\frac{(5^2)^{-3}}{5^{-8}}$	e. $(2^3)^4 \bullet (4^2)^{-3}$	f. $\frac{10^7 \bullet 10^5}{10^{10}}$

8. Fill in the table below.

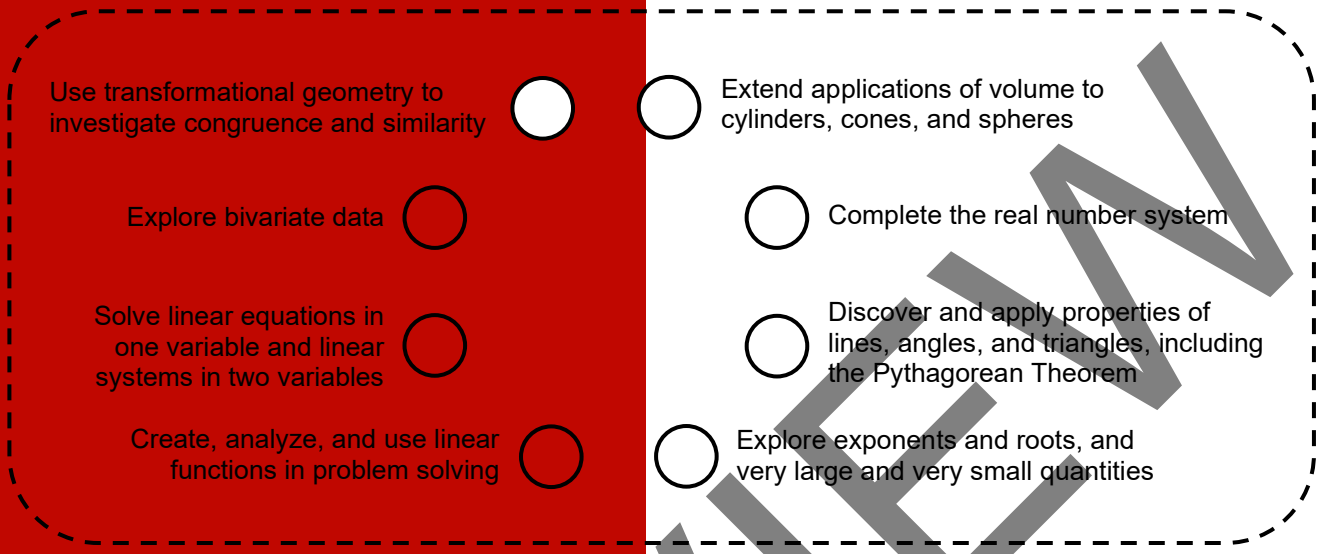
	Standard notation	Product of a number between 1 and 10, and a multiple of 10	Scientific notation
a.	56,000,000	$5.6 \times 10{,}000{,}000$	$5.6 \times 10^{\square}$
b.		$8.85 \times 100{,}000$	
c.			2.7×10^6

9. The sun is expected to die out in about 8 billion years. The average human life expectancy is 80 years. How many lifetimes will the sun exist? Use scientific notation.

10. A cough expels about 20,000 germs per droplet. Each cough carries about 3,000 droplets. How many germs are expelled in a single cough? Use scientific notation.

REFLECTION

1. **Big Ideas**. Shade all circles that describe big ideas in this unit. Draw lines to show connections that you noticed.



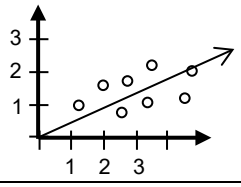
Give an example from this unit of one of the connections above.

2. **Unit Progress.** Go back to **Monitor Your Progress** on the cover and complete or update your responses. Explain something you understand better now than before.

3. **Mathematical Practice.** Describe a situation where paying attention to reasoning of others in class played an important part of understanding the data presented to you [SMP3]. Then circle one more SMP on the back of this packet that you think was addressed in this unit and be prepared to share an example.

4. **Making Connections.** Describe a situation where a story was communicated with data that is of interest to you now or in the future.

STUDENT RESOURCES

Word or Phrase	Definition
association	In statistics, an association between two variables is a relationship between the variables, so that the variables are statistically dependent. In the case of numerical variables, if the relationship is linear, we refer to a linear association between the variables.
bivariate data	Bivariate data is data that has two variables. Bivariate data can be represented by ordered pairs. A list of country of origin and batting average for each baseball player is a bivariate data set with one categorical variable and one numerical variable.
bivariate numerical data	Bivariate numerical data is data that has two numerical variables. Bivariate numerical data can be represented by a scatter plot, so that the relationship (if any) between the variables is more easily seen. A list of heights and weights for each player on a football team is a bivariate numerical data set.
categorical data	Categorical data is data sorted into categories, such as colors, ranges of measurements, or other attributes of the data. Generally, there are only finitely many categories.
data set	A data set is a collection of pieces of information about a population, often numbers, obtained from observation, questioning, or measuring.
frequency table	A frequency table is a table that lists items and the number of times they occur in a data set.
line of best fit	A line of best fit for a scatter plot is a straight line that best represents (in some sense) the data points in the scatter plot. 
measurement data	Measurement data is numerical data that comes from making measurements. Measurement data can be obtained by measuring such things as heights, weights, temperatures, lengths, areas, and volumes.
numerical data	Numerical data is data consisting of numbers. The numbers allow for statistical calculations, such as finding the mean or median.
outlier	An outlier of a data set is a data value that is unusually small or unusually large relative to the overall pattern of values in the data set. For the data set {1, 1, 1, 3, 5, 6, 6, 7, 23}, the data value 23 is a potential outlier.

Word or Phrase	Definition
population	In statistics, the population refers to the source of a data set. If we wish to make statistical inferences about the students at a school, we may take a random sample of the students, or we may gather data from all the students. In either case, the population refers to the students in the school.
relative frequency table	A relative frequency table is a frequency table that lists items and the proportion (or percent) of times they occur.
statistical question	A statistical question is a question where numerical data that has potential for variability can be collected and analyzed for the purpose of answering the question. A statistical question: "How much TV do middle school students watch on average?" NOT a statistical question: "How many hours of TV did you watch last week?"
two-way table	A two-way table is a table that displays bivariate categorical data, in which the rows correspond to the categories of one variable, and the columns correspond to the categories of the other. A two-way table that includes the number of data observations is called a "two-way frequency table". A two-way table that includes the percentage of the number of data observations relative to the total number of observations is called a "two-way relative frequency table".

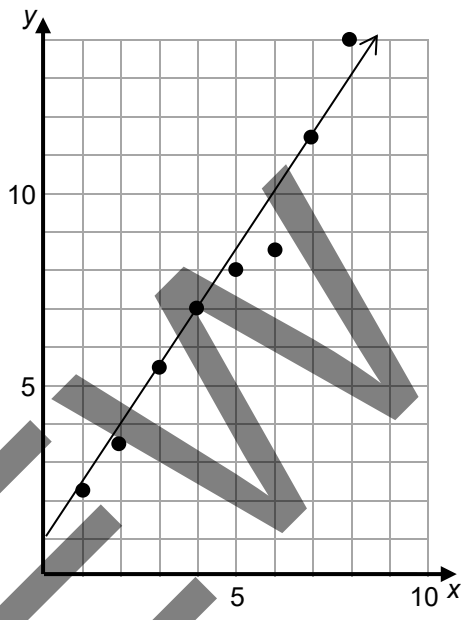
Numerical Data
Numerical data is data consisting of numbers. Measurement data is numerical data that comes from making measurements. Numerical survey questions are used to collect numerical data. Numerical data typically come from counting or measurements. Examples of numerical survey questions include: • How many dogs do you own? (a counting question) • How many minutes did you exercise last week? (a measurement question) Some ways to report one-variable (or univariate) numerical data include: • Measures of center such as mean, median, mode • Measures of spread such as range, mean absolute deviation (MAD), and 5-number summary • Data displays such as tables, line plots, histograms, and box plots Some ways to report two-variable (or bivariate data) numerical data include: • Tables • Graphs • Equations

Lines of Best Fit

A line of best fit for a scatter plot is a straight line that best represents (in some sense) the data points in the scatter plot.

Example: When the data in the table below is graphed in a scatter plot, the data points cluster along a straight line. We conclude that there is likely a linear association between x and y. One possible such line may be estimated by the equation graphed below, $y = \frac{3}{2}x + 1$. Using a graphing calculator, another estimated equation is given as $y = 1.6x + 0.3$ (not graphed).

x	1	2	3	4	5	6	7	8
y	2.2	3.5	5.5	7	8	8.5	11.5	14

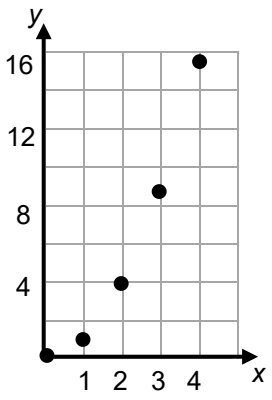


Nonlinear Associations

Not all associations are linear. Here is an example of a scatter plot of bivariate data that appears to have a nonlinear association.

Example: For this data set, the graphed points do not fall in a linear pattern. They increase at an increasing rate.

x	0	1	2	3	4
y	0	1	4	9	16

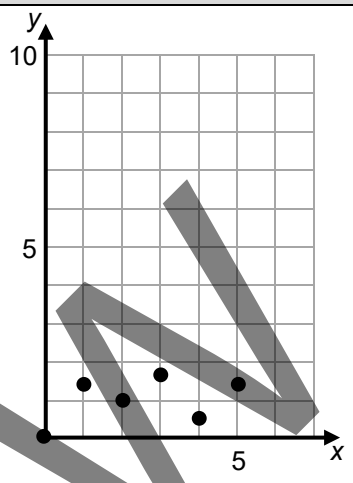


Outliers

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

Outliers can create the illusion that an association exists when one does not. They can also distract us from seeing an association when there clearly is one.

Example 1: In the scatter plot to the right, the data point (6, 10) is a potential outlier. Its *y*-coordinate 10 appears to be unusually large compared to the other *y*-coordinates.



Example 2: In a 6th grade classroom, students were asked how many pets they had. All students but one replied with numbers of pets that ranged from 0 to 8. That one pet owner said she had 40 fish. This number of fish appears to be an outlier, because it is unusually large compared to the other numbers of pets.

Categorical Data

Categorical data is data sorted into categories, such as colors, ranges of measurements, or other attributes of the data. Generally, there are only finitely many categories.

Categorical survey questions are used to collect categorical data. Responses to these questions are usually in words. Examples of categorical survey questions include:

- What types of pets do you own? (Answers include dog, cat, bird, no pets, etc.)
- Do you have a curfew? (A yes-no answer)

Some ways to report one-variable categorical data include

- Frequency tables
- Relative frequency tables
- Pie charts (circle graphs)
- Bar graphs

Some ways to report two-variable categorical data include:

- Two-way frequency tables
- Two-way relative frequency tables

COMMON CORE STATE STANDARDS

STANDARDS FOR MATHEMATICAL CONTENT	
8.F.B	**Use functions to model relationships between quantities.**
8.F.4	Construct a function to model a linear relationship between two quantities. Determine the rate of change and initial value of the function from a description of a relationship or from two (x, y) values, including reading these from a table or from a graph. Interpret the rate of change and initial value of a linear function in terms of the situation it models, and in terms of its graph or a table of values.
8.F.5	Describe qualitatively the functional relationship between two quantities by analyzing a graph (e.g., where the function is increasing or decreasing, linear or nonlinear). Sketch a graph that exhibits the qualitative features of a function that has been described verbally.
8.SP.A	**Investigate patterns of association in bivariate data.**
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. *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.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. *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?*

STANDARDS FOR MATHEMATICAL PRACTICE	
SMP1	Make sense of problems and persevere in solving them.
SMP2	Reason abstractly and quantitatively.
SMP3	Construct viable arguments and critique the reasoning of others.
SMP4	Model with mathematics.
SMP5	Use appropriate tools strategically.
SMP6	Attend to precision.
SMP7	Look for and make use of structure.
SMP8	Look for and express regularity in repeated reasoning.