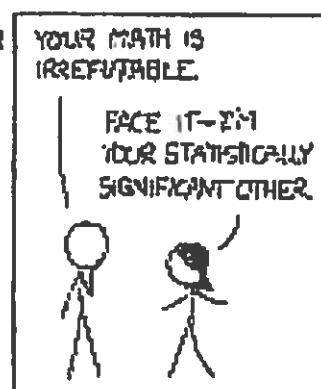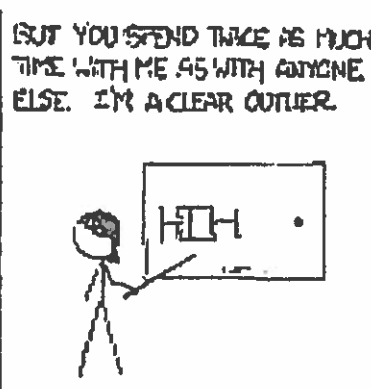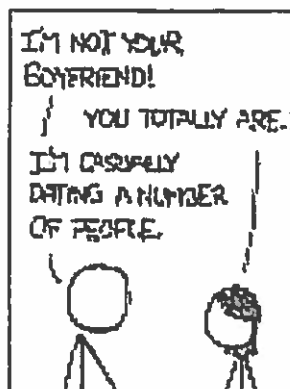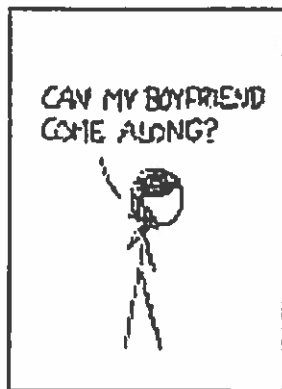# Unit 6: Statistics

I can:

- create a box plot, dot plot, or histogram, and explain its interpretation
- draw conclusions from a dot plot, histogram, and/or box plot
- determine in which cases the mean or median is a better measure of center
- determine in which cases the interquartile range is a better measure of spread
- use appropriate statistics to compare two or more data sets
- interpret how an extreme data point effects a data set
- identify which measure is the most appropriate representation of the center of the data set
- draw a symmetrical distribution of data points
- interpret how an extreme data point effects the spread
- identify which measure is the most appropriate representation of the spread of a data set
- construct two-way frequency tables for categorical data.
- interpret joint, marginal and conditional relative frequencies in context.
- explain possible associations between categorical data in two-way tables.
- identify and describe trends in the data.
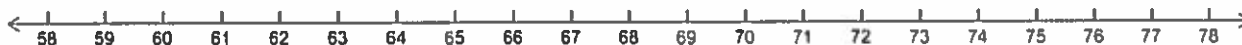- distinguish between correlation and causation.

Where Do I Stand?

How does my height compare with other Algebra 1 students?

In pairs, measure each other's height, rounded to the nearest inch.
Record your height on the dotplot at the front of the room (females use red, males use green).
Make a line at the front of the room, shortest to tallest.

1. Record the dotplot

2. What is the median height? Describe how you found it.

3. What is Q_1 and Q_3? Describe how you found them.

4. Record the following values and then use them to make a boxplot.

Minimum: Q_1:   Median: Q_3:   Maximum:
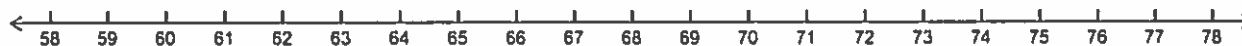
4. The **interquartile range** (or *IQR*) is defined as $Q_3 - Q_1$. Find the *IQR*. Where do you see the *IQR* in the boxplot?

5. An **outlier** is a data value that is way too small or way too big (using the rules below). Are there any outliers? Show your work.

Way too small $< Q_1 - 1.5IQR$

Way too big $> Q_3 + 1.5IQR$

6. Now we will separate our data into two groups, females and males.

Heights for females – find the following values and then make a boxplot.

Minimum: Q_1:   Median: Q_3:   Maximum:

Heights for males – find the following values and then make a boxplot.

Minimum: Q_1:   Median: Q_3:   Maximum:

Write a few sentences <u>comparing</u> the distribution of heights for girls with the distribution of heights for boys.

TheStatsMedic

Learning Task: Math Class

Name_____   Date_____

Mr. Turner has two Math 2 classes. With one class, he lectured and the students took notes. In the other class, the students worked in small groups to solve math problems. After the first test, Mr. Turner recorded the student grades to determine if his different styles of teaching might have impacted student learning.

     Class 1: 80, 81, 81, 75, 70, 72, 74, 76, 77, 77, 77, 79, 84, 88, 90, 86, 80, 80, 78, 82

     Class 2: 70, 90, 88, 89, 86, 86, 86, 86, 84, 82, 77, 79, 84, 84, 84, 86, 87, 88, 88, 88

1. Analyze his student grades by calculating the mean, median, mean absolute deviation, and interquartile range. Which class do you think was the lecture and which was the small group? Why?

2. Draw histograms to easily compare the shapes of the distributions.

3. Which measure of center and spread is more appropriate to use? Explain.

Georgia Department of Education
Georgia Standards of Excellence Framework
GSE Algebra I • Unit 6

Graphic Organizer: Measures of Center and Spread

Mean	Median

Measures of Center

Measures of Spread

| Mean Absolute Deviation	Interquartile Range
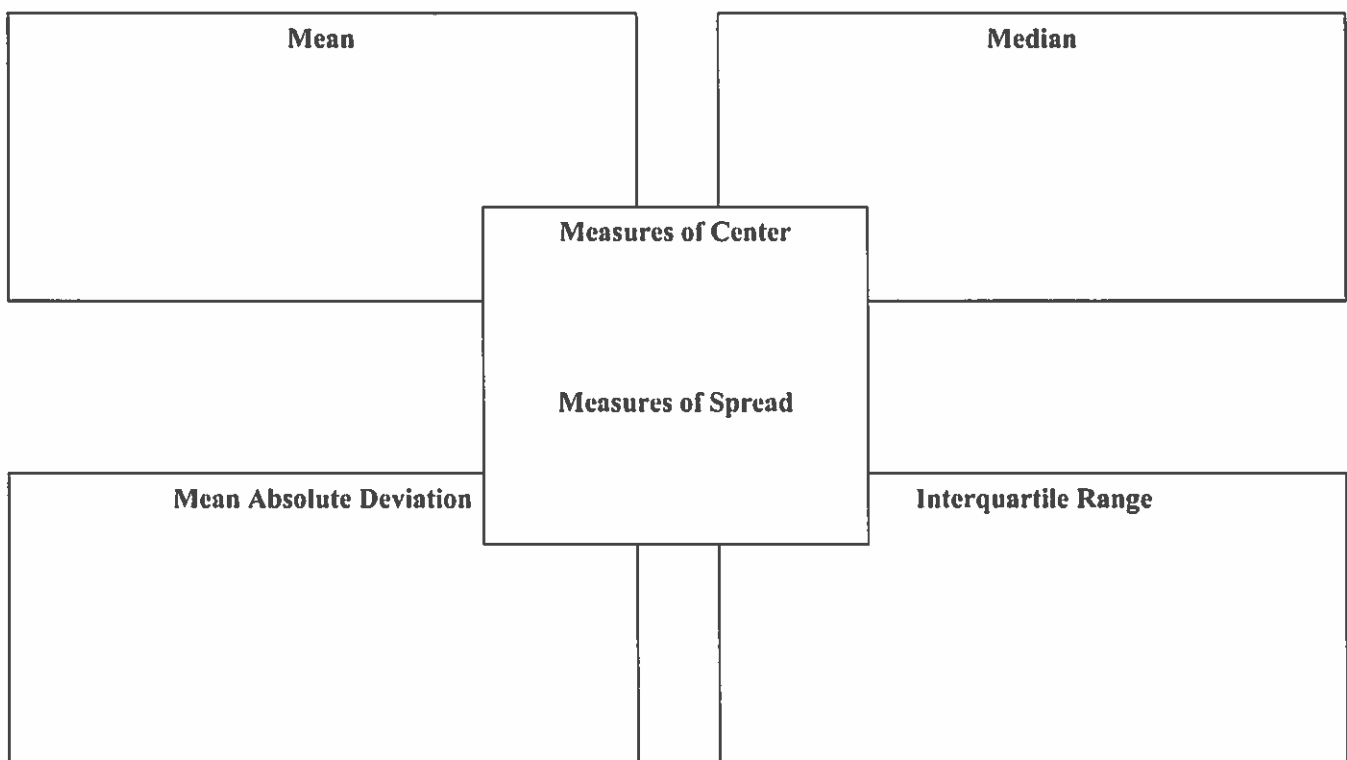
6

Name: _____   Date: _____

GRAPHICALLY REPRESENTING DATA

Quantitative data on a single variable is often collected in order to understand how a characteristic of a group differs amongst the group members or between groups. When we ask a question like "How old is a typical fast food worker?" it is helpful to take a survey and then see graphically how the ages differ amongst the group.
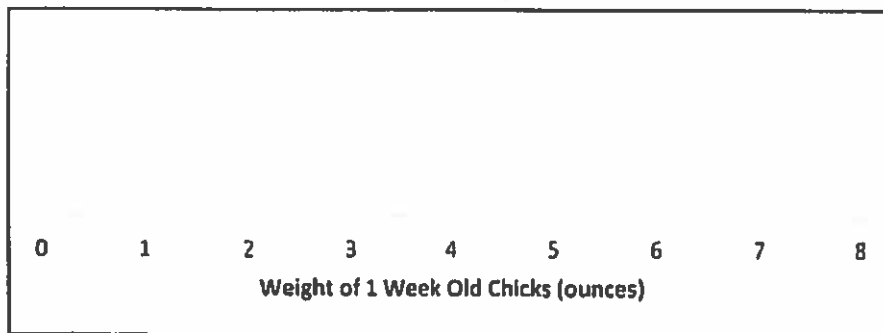
Exercise #1: Charlie's Food Factory currently employs 28 workers whose ages are shown below on a dot plot. Answer the following questions based on this plot.

Ages of Workers at Charlie's Food Factory
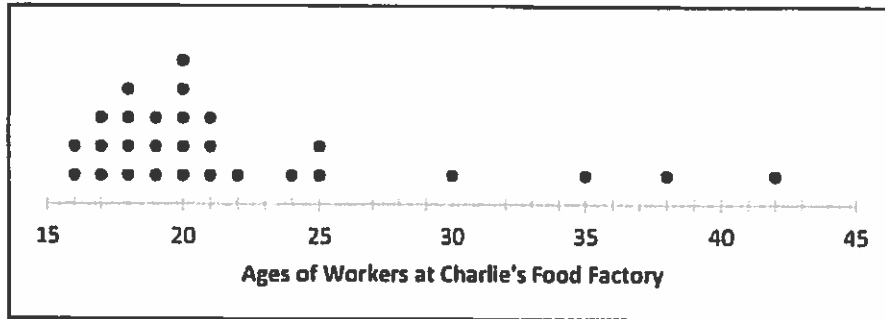
(a) How many of the workers are 18 years old?

(b) What is the range of the ages of the workers?

(c) Would you consider this distribution symmetric?

(d) The mean (average) age for a worker is 22 years old. Why is this average not representative of a typical worker?

Exercise #2: A farm is studying the weight of baby chickens (chicks) after 1 week of growth. They find the weight, in ounces, of 20 chicks. The weights are shown below. Construct a dot plot on the axes given.

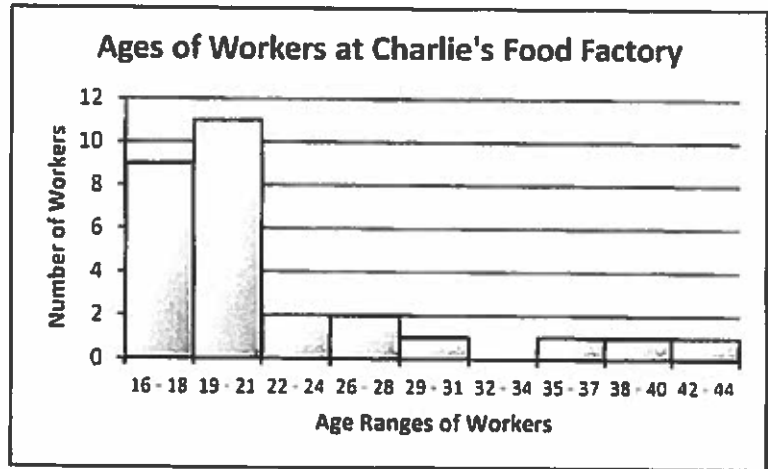2, 1, 3, 4, 2, 2, 3, 1, 5, 3, 4, 4, 5, 6, 3, 8, 5, 4, 6, 3

Weight of 1 Week Old Chicks (ounces)

Exercise #3: The following **histogram** shows the ages of the workers at Charlie's Food Factory (from Exercise #1) but in a different format.

(a) How many workers have ages between 19 and 21 years?

(b) What is the disadvantage of a histogram compared to a dot plot?

(c) Does the histogram have any advantages over the dot plot?

Ages of Workers at Charlie's Food Factory

Number of Workers

Age Ranges of Workers

16 - 18 19 - 21 22 - 24 26 - 28 29 - 31 32 - 34 35 - 37 38 - 40 42 - 44

Exercise #4 The 2006 – 2007 Arlington High School Varsity Boy's basketball team had an excellent season, compiling a record of 15 – 5 (15 wins and 5 losses). The total points scored by the team for each of the 20 games are listed below in the order in which the games were played:

76, 55. 76, 64, 46, 91, 65, 46, 45, 53, 56, 53, 57, 67, 58, 64. 67, 52, 58, 62

(a) Complete the frequency table below.

POINTS SCORED	TALLY	FREQUENCY
40 - 49		
50 - 59		
60 - 69		
70 - 79		
80 - 89		
90 - 99		

(b) Construct the histogram below.
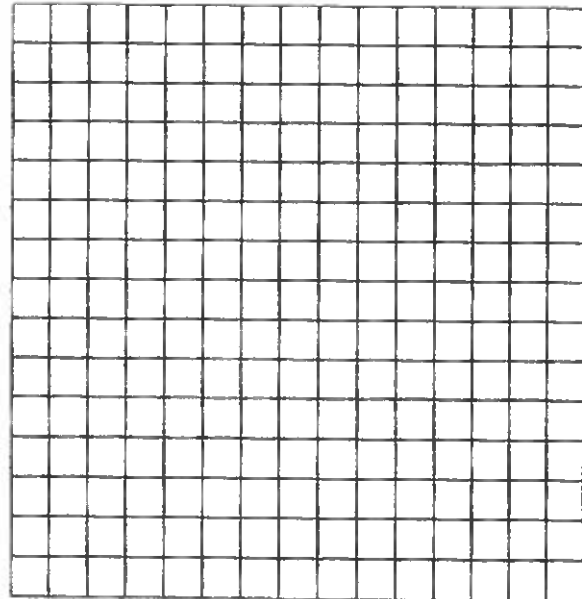
Name: _____   Date: _____

GRAPHICALLY REPRESENTING DATA
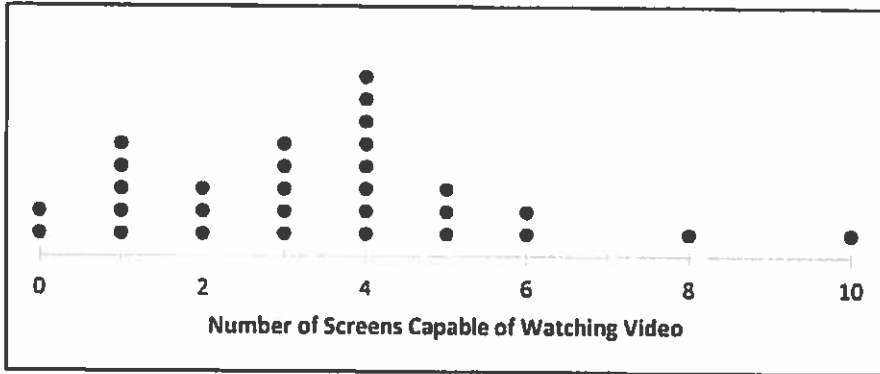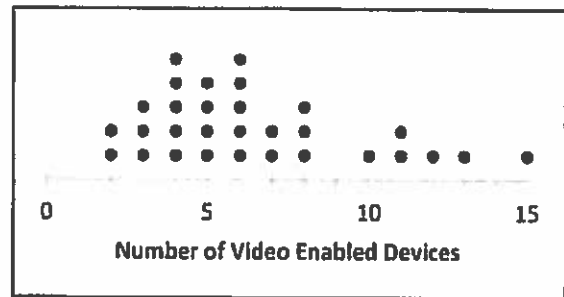
A local marketing company did a survey of 30 households to determine how many devices the household contained that family members watched video on (i.e. TV's, tablets, smart phones, etcetera). The dot plot of the responses is shown below.
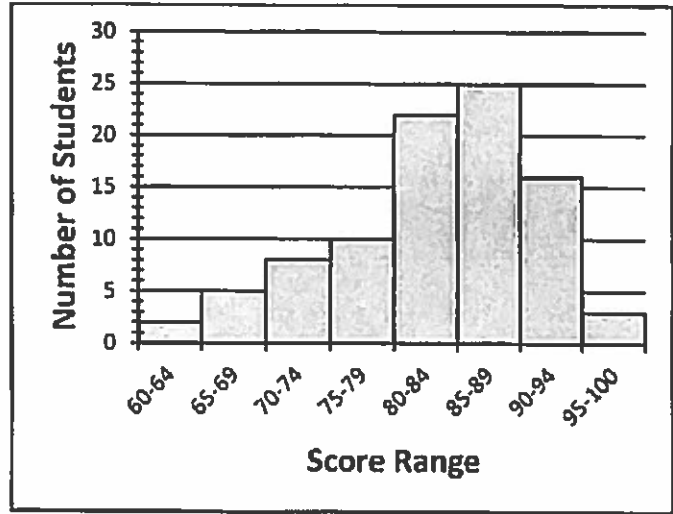

Number of Screens Capable of Watching Video

1. How many households have three devices capable of showing video on them?

    (1) 1   (3) 7

    (2) 2   (4) 5

2. More households had 4 devices to watch video on than any other number. Which of the following is closest to the percent of households that have 4 devices?

    (1) 22%   (3) 27%

    (2) 34%   (4) 45%

3. The marketing company would like to claim that the majority of households have either 3 or 4 screens capable of watching video on. Does the information displayed on the dot plot support this claim? Explain your reasoning.

4. The same marketing company then surveyed 30 households that contained at least one teenager. The dot plot for the video enabled devices is shown below. The mean number of screens for the first survey was 3.4. Based on the second dot plot, do you think its mean will be higher or lower? Explain.

Number of Video Enabled Devices

COMMON CORE ALGEBRA 1, UNIT #10 – STATISTICS – LESSON #1
eMATHINSTRUCTION, RED HOOK, NY 12571, © 2013

On a recent Precalculus quiz, Mr. Weiler found the following distribution of scores, which are arranged in 5 point intervals (with the exception of the last interval).

5. How many students scored in the 75 to 79 point range?

   (1) 8   (3) 25

   (2) 10   (4) 5

6. Students do not pass the quiz if they receive lower than a 70. How many students did not pass?

   (1) 8   (3) 7

   (2) 5   (4) 15

7. How many total student took the quiz?

   (1) 25   (3) 56

   (2) 104   (4) 91

8. Twenty-two students scored in the 80 to 84 range on this test. Does the histogram provide us with enough information to conclude that a student must have scored on 82 on this test? Explain your thinking.

9. A random survey of 100 cars found the following frequency distribution for the fuel efficiency of the car, as measured in miles per gallon. Construct a histogram below that effectively shows the distribution of this data set.

Fuel Efficiency (miles per gallon)	Number of Cars
10 to 14	4
15 to 19	17
20 to 24	36
25 to 29	24
30 to 34	10
35 to 39	6
40 to 44	3
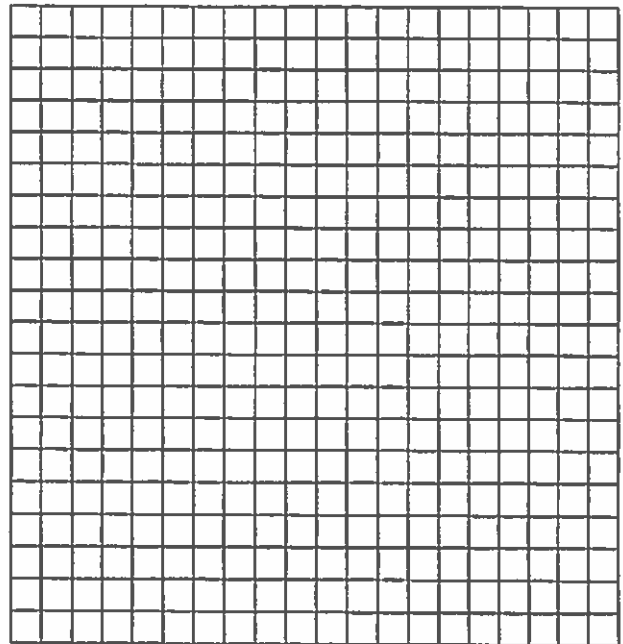
Name: _____   Date: _____

QUARTILES AND BOX PLOTS

Another visual representation of how a data set is **distributed** comes in the form of a box plot. We create box plots by dividing the data up roughly into quarters by finding the **quartiles** of the data set.

Exercise #1: Shown below are the scores 16 students received on a math quiz.

52, 60, 66, 66, 68, 72, 72, 73, 74, 75, 80, 82, 84, 91, 92, 98

(a) What is the median of this data set?

(b) Find the **range** of the data set (defined as the difference between the largest data value and the smallest data value).

(c) What is the median of the lower half of this data set (known as the **first quartile**, Q_1)?

(d) What is the median of the upper half of this data set (known as the **third quartile**, Q_3)?

The first and third quartiles are sometimes known as the lower and upper quartiles, respectively. The quartiles, the median, and the lowest and highest values in a data set comprise what is known as the **five number summary** and can be graphically represented on a **box plot**. This type of plot is also sometimes known as a **box and whiskers plot**.

Exercise #3: Using the same data set construct a box plot on the number line given below.

Exercise #4: The ages of the 15 employees of the Red Hook Curry House are given below.

$$16, 17, 17, 18, 19, 22, 25, 26, 29, 33, 33, 37, 40, 42, 44$$

(a) Determine the median and quartile values for this data set.

(b) Create a box-and-whiskers diagram below.

Exercise #5: Twenty of Mr. Ouimet's physics students recently took a quiz. The results of this quiz are shown in the following box-and-whiskers diagram. Assume that all scores are whole numbers.
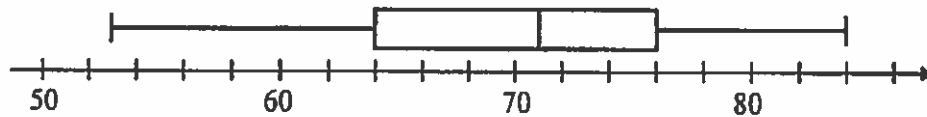
(a) What was the median score on Mr. Ouimet math quiz?

(b) What was the range of the scores on Mr. Ouimet's physics quiz?

(c) What score was greater than or equal to 75% of all other scores on this quiz?

(d) Mr. Ouimet regularly sets the passing grade on his quizzes to be the score of the lower quartile. What is the passing grade on this quiz?

Exercise #6: Which of the following box plots shows a data set with the greatest median?

(1)

(3)

(2)

(4)
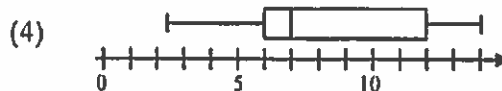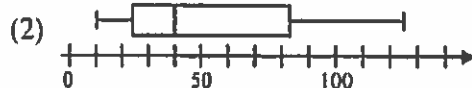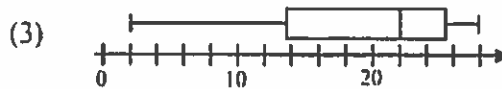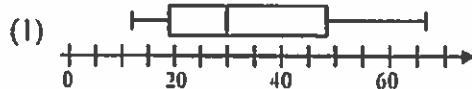
Name: _____   Date: _____

QUARTILES AND BOX PLOTS

FLUENCY

1. Which of the following data sets, given in ascending order, has the greatest range?

    (1) $\{3, 4, 7, 10, 18\}$ (3) $\{-2, 5, 8, 11, 26\}$

    (2) $\{65, 66, 70, 72\}$ (4) $\{-5, -2, 4, 7, 10\}$

2. Given the box plot shown below, which of the following represents the third quartile value for this data set?

    (1) 12   (3) 6

    (2) 18   (4) 19
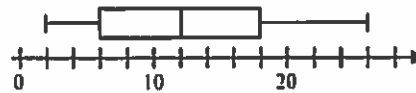
3. Given the box plot shown below, which of the following represents the range of this data set?

    (1) 110   (3) 60

    (2) 40   (4) 75
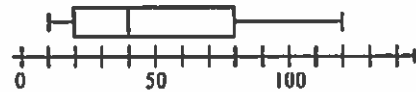
4. According to the following box-and-whiskers diagram, which of the following values represents the lower quartile of this data set?

    (1) 20   (3) 28

    (2) 13   (4) 16
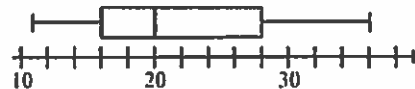
5. Which of the following box-and-whiskers diagram represents a data set whose median value is equal to 65?

    (1)

    (2)

    (3)

    (4)

14
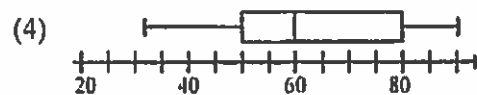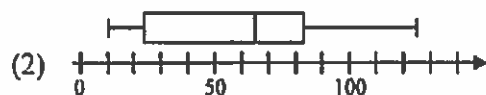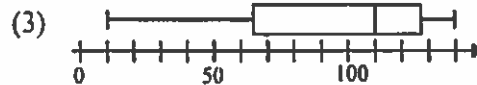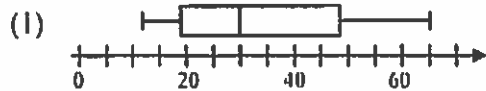
APPLICATIONS

6. The ages of 12 retail workers are given in the data set below.

$$17, 18, 18, 19, 20, 21, 22, 23, 25, 25, 34, 47$$

(a) Calculate the five number summary. Label each of the five numbers with what they represents (i.e. min, max, lower quartile, etc.).

(b) Create a box-and-whiskers diagram of this data set below.

```
  +----+----+----+----+----+----+----+----+---->
  10   20   30   40   50
```

7. Mr. Ramirez gives a math test and records the grades of his 17 students as follows:

$$67, 72, 74, 74, 78, 80, 80, 82, 85, 85, 86, 87, 90, 92, 92, 95, 98$$

Create a box-and-whisker diagram of this data set below.

```
  +----+----+----+----+----+----+----+----+---->
  60   70   80   90   100
```
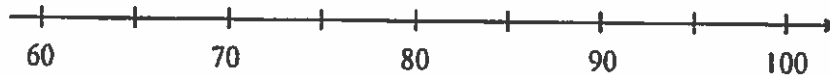
8. The speeds, in miles per hour, of 24 cars on a particular road are recorded and represented on the box-and-whiskers diagram shown below. Answer each of the following questions based on this diagram.

(a) What is the range of this data set?

```
      +-----[====|====]----+
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+->
  20     30     40     50
```
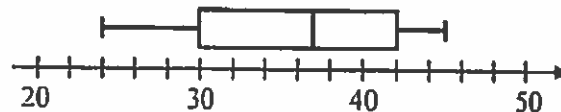
(b) What is the maximum speed of the 24 drivers?

(c) How many drivers drove between 30 and 42 miles per hour?

(d) If the speed limit on this part of the road is 35 miles per hour, are more people speeding or are more people going below the speed limit? Justify.

Notes: Comparing Distributions

When you compare two or more data sets, focus on four features:
- **Center.** Graphically, the center of a distribution is the point where about half of the observations are on either side.
- **Spread**. The spread of a distribution refers to the variability of the data. If the observations cover a wide range, the spread is larger. If the observations are clustered around a single value, the spread is smaller.
- **Shape**. The shape of a distribution is described by symmetry, skewness, number of peaks, etc.
- **Unusual features**. Unusual features refer to gaps (areas of the distribution where there are no observations) and outliers.

SPREAD
The spread of a distribution refers to the variability of the data. If the data cluster around a single central value, the spread is smaller. The further the observations fall from the center, the greater the spread or variability of the set.

Less Spread   **More Spread**
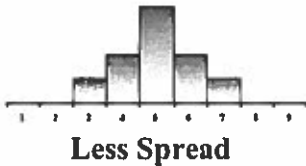
SHAPE
The shape of a distribution is described by symmetry, number of peaks, direction of skew, or uniformity

symmetric, unimodal,   **skewed right**   **skewed left**
bell-shaped

uniform   **symmetric, bimodal**   **non-symmetric, bimodal**
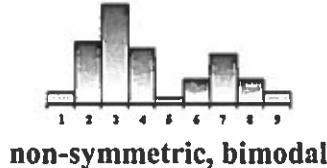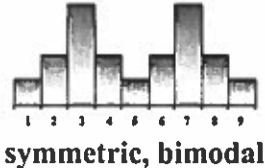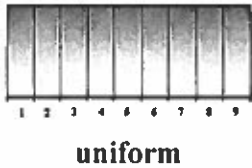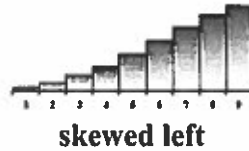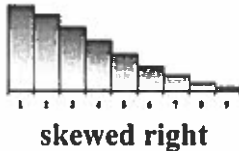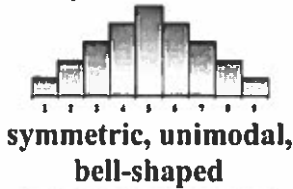
UNUSUAL FEATURES
Sometimes, statisticians refer to unusual features in a set of data. The two most common unusual features are gaps and outliers.
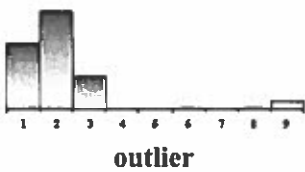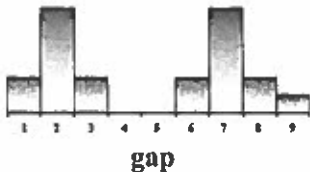
gap   **outlier**

16

Performance Task: The Basketball Star

Name_____   Date_____

Bob believes he is a basketball star and so does his friend Alan.

Bob's Points per Game	Alan's Points per Game
8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8	1, 3, 0, 2, 4, 5, 7, 7, 8, 10, 4, 4, 3, 2, 5, 6, 6, 6, 8, 8, 10, 11, 11, 10, 12, 12, 5, 6, 8, 9, 10, 15, 10, 12, 11, 11, 6, 7, 7, 8

1. Create dot plots for Bob and Alan's last forty games.

2. Create box plots for Bob and Alan's last forty games.

3. Create histograms of both Bob's and Alan's data.

4. Which graphical representation best displayed Bob's and Alan's data?

5. Use summary statistics to compare Bob and Alan's points per game.

6. Describe each person's data in terms of center, spread, and shape.

18

Name: _____   Date: _____

TWO WAY FREQUENCY TABLES

So far we have worked with **quantitative data** for a single variable, for example weight of baby chicks or number of video enabled devices. We can also work with **categorical data** or data that shows how many things surveyed fall into a **given category**.

Exercise #1: Let's do a quick categorical survey in this class. By a show of hands, determine how many students fall into each of the following categories for eye color.

Brown Eyes   Blue Eyes   Green Eyes   Other

Although surveys of data that contain only one category are interesting, statisticians are often interested in how responses to two categories relate to one another. For example, we may want to know how a person's gender (one category) affects what profession (a second category) they would prefer when they grow up. We may want to know if a person's hair color (one category) has any relationship to their eye color (a second category). This type of data is summarized in a **two-way frequency table**.

Exercise #2: A class of 20 students recorded their hair color and eye color which are shown in the two-way frequency table below.

		Hair Color			
		Black	Blond	Red	Total
Eye Color	Blue	3	4	1	8
	Brown	5	2	0	7
	Green	1	1	3	5
	Total	9	7	4	20

(a) How many students had blond hair and blue eyes?   (b) How many students had red hair?

(c) Construct a table that shows the **joint relative frequencies** and the **marginal relative frequencies** for the data above.

		Hair Color			
		Black	Blond	Red	Total
Eye Color	Blue				
	Brown				
	Green				
	Total				

We would like to understand **associations** or **trends** within the data set, i.e. would a response to one category tell us something about the response to the other category?

Exercise #3: Let's see if there is a connection between eye color and hair color by using **conditional relative frequencies.**

		Hair Color			
		Black	Blond	Red	Total
Eye Color	Blue	3	4	1	8
	Brown	5	2	0	7
	Green	1	1	3	5
	Total	9	7	4	20

(a) What is the conditional relative frequency of having green eyes if you have red hair? (This is equivalent to asking what the percent of people with red hair have green eyes.)

(b) What is the conditional relative frequency of having green eyes if you have black hair?

(c) Does it appear that having green eyes has a dependency or at least an association with having red hair? Explain.

(d) Is it more likely that a person with black hair has blue eyes or that a person with blond hair has brown eyes? Use conditional marginal frequencies to support your answer.

Exercise #4: A survey of 52 graduating seniors was conducted to determine if there was a connection between the gender of the student and whether they were going on to college. Based on this data, what is more likely: that someone going to college is female or that someone who is female is going to college? These may seem like the same thing, but are quite different.

	Gender		
	Male	Female	Total
Going to College	16	13	29
Not Going to College	14	9	23
Total	30	22	52

Name: _____   Date: _____

TWO WAY FREQUENCY TABLES

A survey was done to determine the relationship between gender and subject preference. A total of 56 students were surveyed to determine if they liked math, English, social studies, or science as their favorite subject. The results were then broken down based on whether the respondent was male or female.

	Math	English	Social Studies	Science	Total
Female	8	6	11	5	30
Male	10	4	8	4	26
Total	18	10	19	9	56

1. Which of the following is closest to the joint relative frequency of being a male who likes social studies?

   (1) 0.42   (3) 0.31

   (2) 0.14   (4) 0.56

2. Which of the following is the marginal relative frequency of liking math?

   (1) $\dfrac{18}{36}$ (3) $\dfrac{10}{18}$

   (2) $\dfrac{8}{10}$ (4) $\dfrac{18}{56}$

3. What percent of female students liked English as their favorite subject?

   (1) 20%   (3) 11%

   (2) 16%   (4) 60%

4. A person looking at this table concludes that it is more likely that a female student will like social studies than a male student will like math. Is this correct? Justify your answer.

5. Is it more likely that a person who likes social studies will be female or that a person who is female will like social studies? Justify.

Demographers are trying to understand the association between where a person lives and how they commute to work. They survey 100 people in three cities with the results shown below.

	Car	Train	Walk	Total
New York	5	25	10	40
Los Angeles	18	12	5	35
Chicago	8	14	3	25
Total	31	51	18	100

6. Fill in the table below with the relative frequencies.

	Car	Train	Walk	Total
New York				
Los Angeles				
Chicago				
Total				

7. Given that a person rides a train to work, what is the conditional relative frequency that they live in New York?

   (1) 0.25   (3) 0.49

   (2) 0.63   (4) 0.82

8. If a person lives in Los Angeles, what is the conditional relative frequency that they drive a car?

   (1) 0.42   (3) 0.68

   (2) 0.16   (4) 0.51

9. Which of the following is the marginal frequency of walking to work?

   (1) 18%   (3) 25%

   (2) 60%   (4) 44%

10. Is a person more likely to ride a train if they live in New York or if they live in Chicago? Justify your answer.

Scatter Plots

a graph with points plotted to show the _____
between _____ _____ of data

Distance to School (miles)	Length of Commute (minutes)
1	5
3	5
5	10
6	12
2	4
3	6
7	15

Distance Students Commute to School

Commute Time (minutes) — Distance (miles)

Types of Association

Positive Association

Negative Association

No Association

Warm-up

24

26

Name: _____ Date: _____

27
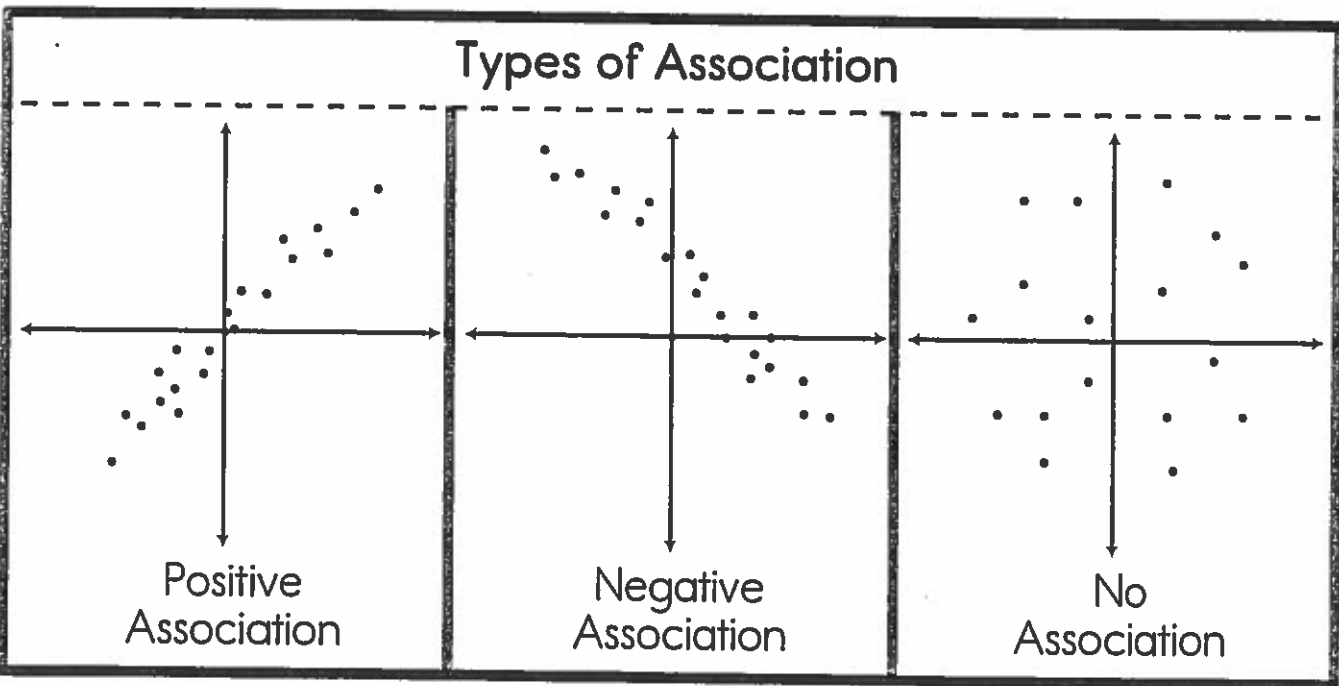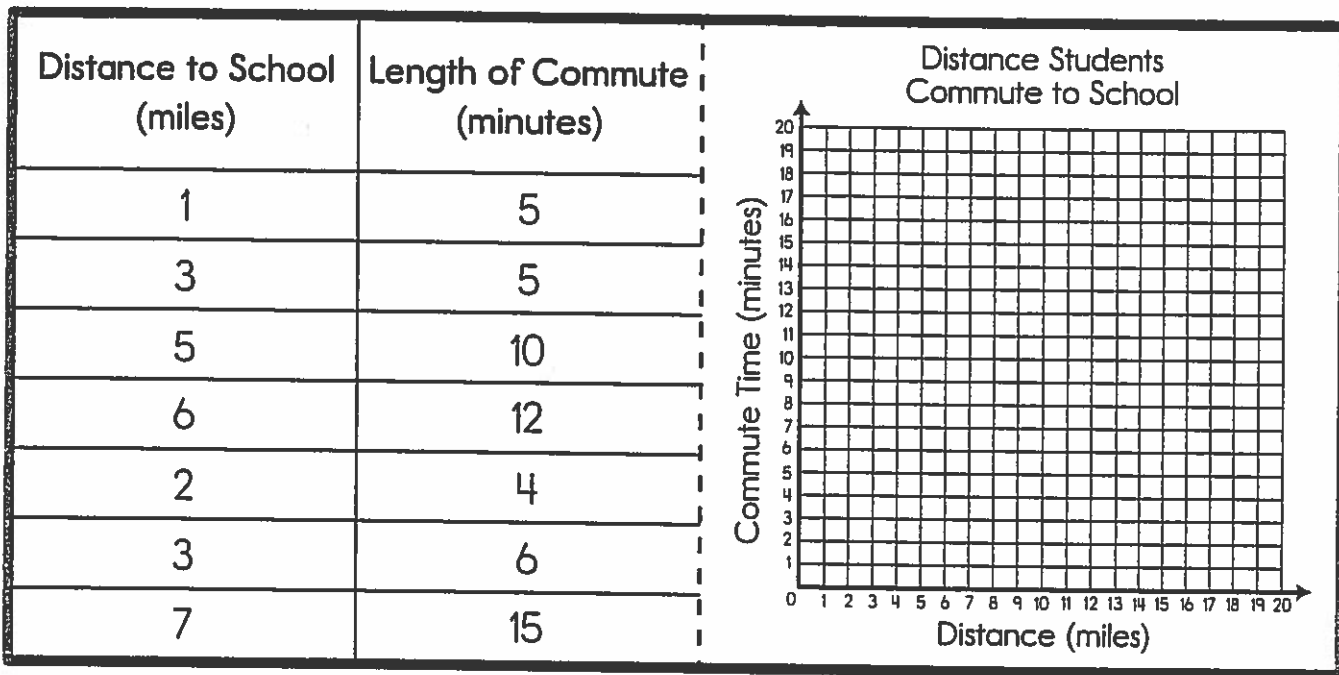
BIVARIATE DATA ANALYSIS
Scatter Plots

Oftentimes, statistical studies are done where data is collected on **two variables** instead of one in order to establish whether there is a **relationship** between the **two variables**. This is called a **bivariate data analysis**.

Exercise #1: A survey was taken of 10 low and high temperatures, in Fahrenheit, in the month of April to try to establish a relationship between a day's low temperature and high temperatures.

Low Temperature, x	26	28	30	32	34	35	37	38	41	45
High Temperature, y	49	50	57	54	60	58	64	66	63	72

(a) Construct a scatter plot of this bivariate data set on the grid below.

(b) Draw a line of best fit through this data set.

(c) Calculate the slope of this line by picking off two points (not necessarily data points).
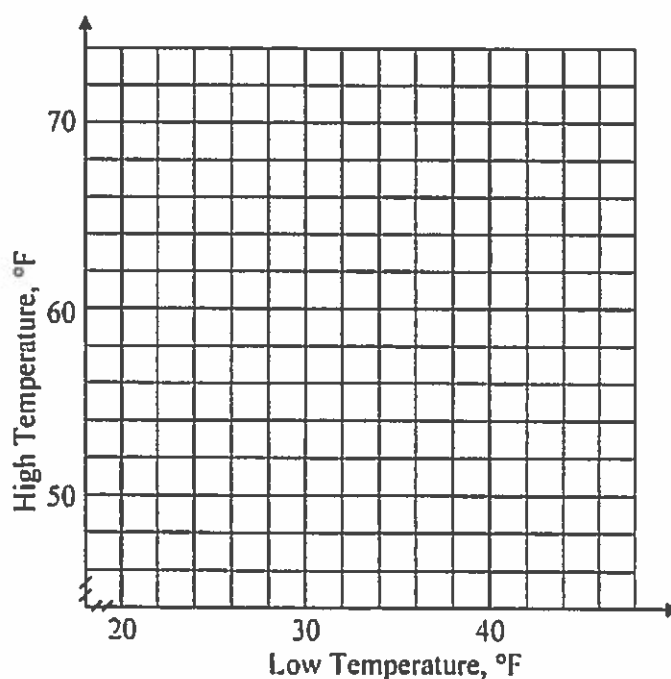
(d) Use your line of best fit to estimate the high temperature for a day in April given that the low temperature was 42 degrees. Illustrate your answer on your graph.

(e) Would you characterize the relationship between the low and high temperature as a **positive correlation** or a **negative correlation**? Explain.

Two variables can have a **strong relationship** with one another, as seen on a scatterplot, but might not have a **causal relationship**. A causal relationship exists when the change in one variable actually **causes the change** in the other (or is one of the primary causes).

Exercise #2: In each of the following scenarios, two variables are given that if plotted would have a strong correlation (a scatterplot where the data falls nearly in a line). Determine if there exists a causal relationship between the two variables. If so, which variable causes the other?

(a) The high temperature in New York City and the number of bottles of water sold.

(b) A person's height and a person's shoe size.

(c) A person's weight loss and the number of hours a person spends in the gym per week.

(d) The years of education a person achieves and the salary that person starts at upon entering the work force.

Variables can have extremely strong **correlations** but no causal relationship. This is often the case if there is a third variable that causes both (known interestingly enough as a **lurking variable**).

Exercise #3: The table below shows the number of firefighters required to fight a given fire versus the dollar damage done to the house by the fire.

Number of firefighters	2	3	5	8	9	12	16
Damage done by fire (in dollars)	2,932	9,750	15,575	23,190	22,900	35,400	52,900

(a) Are the data positively or negatively correlated? How can you tell?

(b) Does the number of firefighters cause the damage done to the house? If not, what hidden variable is causing both variables to change?

Name: _____   Date: _____

BIVARIATE DATA ANALYSIS
Scatter Plots

1. A survey was done at Ketcham High School to determine the effect of time spent on studying and grade point average. The table below shows the results for 10 students randomly selected.

Study time (Hours per week)	2	4	5	7	10	12	14	17	19	20
GPA (out of 100)	64	71	69	74	81	86	84	94	91	96

(a) Create a scatter plot for this data set on the grid provided. Draw a best fit line through the scatterplot drawn.

Grade Point Average

Hours Studying

(b) Determine the equation for the best fit line for this data set. Use two points from the line you drew in (a) to determine the slope and estimate the y-intercept graphically. Round your slope to the nearest *tenth* and the y-intercept to the nearest whole number.
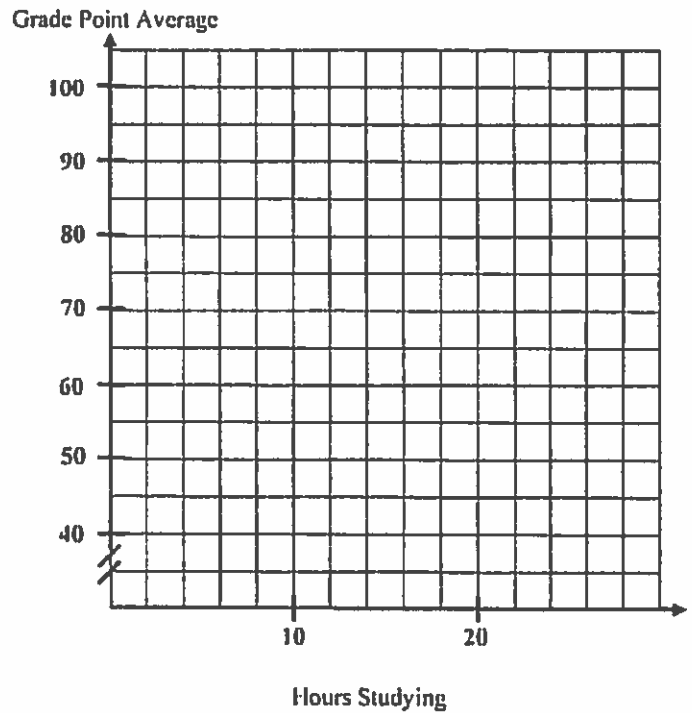
(c) Use your answer from part (b) to determine the expected GPA from studying for 8 hours per week. Round your answer to the nearest whole number.

(d) Is there a causal relationship between these two variables? If so, which variable causes the other?

30

2. A survey was done to determine if there was any connection between the price that people pay for their most expensive car and the current value of their house. The results, for eight participants, are given below.

Car Cost, x (in dollars)	11,500	14,750	19,500	26,750	32,900	43,000	45,750	54,500
House Value, y (in dollars)	160,000	195,000	255,000	400,000	440,000	525,000	475,000	725,000

A computer was used to determine the line of best fit. Its equation was:

$$y = 12x + 33,766$$

(a) Use the line of best fit to predict the house value of a person whose most expensive car costs $19,500.

(b) Was the prediction in (a) an overestimate or underestimate of the actual house value? Explain.

(c) Is there a positive or negative correlation between these two variables? Explain.

(d) Is there a causal relationship between these two variables? If you answer yes, then determine which variable causes the other. If you answer no, then explain a third variable that could be causing both.

3. It has been noted that on any given day, there is a strong correlation between the number of ice-cream cones sold and the number of people who go swimming. Is there a causal relationship between the eating ice-cream and going swimming? If not, what could be causing this strong correlation?

Scatter Plots and Trend Lines
Extension: Correlation, Lines of Fit, and Predictions

Essential question: *How can you decide whether a correlation exists between paired numerical data and, if so, what is the line of fit for that data?*

Video Tutor

**MCC9–12.S.ID.8**

1 ENGAGE \ **Understanding Correlation**

When two real-world variables (such as height and weight or latitude and average temperature) are measured from the same things (the same people, places, etc.), you obtain a set of paired numerical data that you can plot as points in the coordinate plane to create a data display called a *scatter plot*. Sometimes the scatter plot will show a linear pattern. When it does, the linear pattern may be tight (that is, the points lie very close to a line), or it may be loose (that is, the points are more dispersed about a line). The degree to which a scatter plot shows a linear pattern is an indicator of the strength of a **correlation** between the two variables.

Mathematicians have defined a measure of the direction and magnitude of a correlation. This measure is called the **correlation coefficient** and is denoted by r. When the points in a scatter plot all lie on a line that is not horizontal, r has a value of 1 if the line rises from left to right and a value of -1 if the line falls from left to right. The correlation coefficient takes on values between -1 and 1 in cases where the points are not perfectly linear.
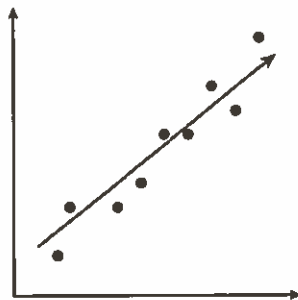
Strong positive correlation
r is close to 1.
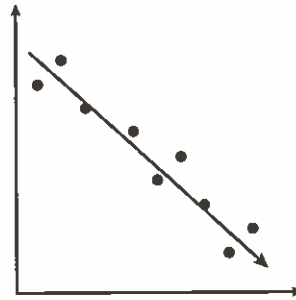
Strong negative correlation
r is close to -1.
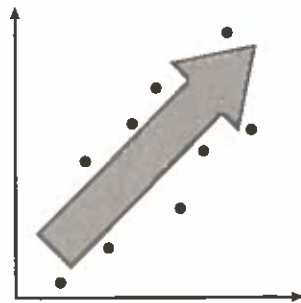
Weak positive correlation
r is closer to 0.5 than to 0 or 1.

Weak negative correlation
r is closer to -0.5 than to 0 or -1.

32

Mean Absolute Deviation

Find the mean absolute deviation

10, 7, 13, 10, 8

Data	Mean	Difference	Positive Value
		Sum:	
		Count:	

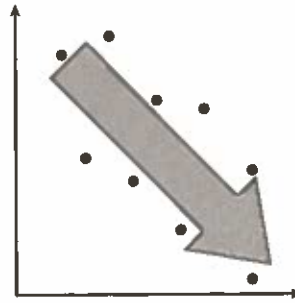The average of the "Positive Value" column — Mean Absolute Deviation:

Find the mean absolute deviation

110, 114, 104, 108, 106

Data	Mean	Difference	Positive Value
		Sum:	
		Count:	

The average of the "Positive Value" column — Mean Absolute Deviation:

Find the mean absolute deviation

87, 75, 85, 77, 74, 82, 90, 88, 79, 81

Data	Mean	Difference	Positive Value
		Sum:	
		Count:	

The average of the "Positive Value" column — Mean Absolute Deviation:

Find the mean absolute deviation

15, 17, 15, 17, 21, 17, 15, 23, 20, 18

Data	Mean	Difference	Positive Value
		Sum:	
		Count:	

The average of the "Positive Value" column — Mean Absolute Deviation:
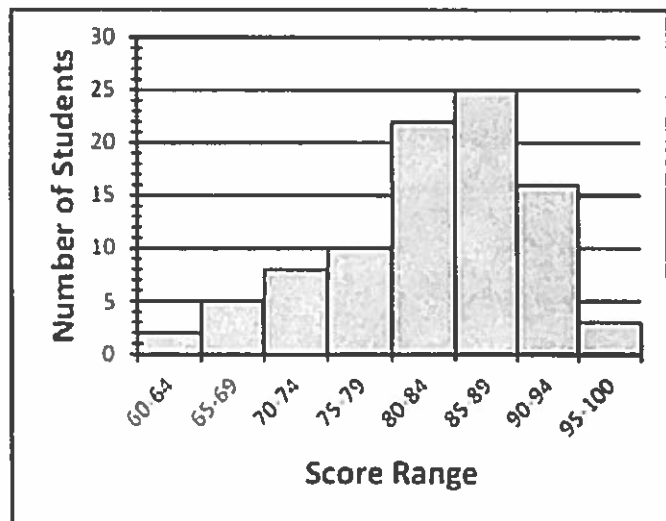
34

Data Representations Overview Homework

A local marketing company did a survey of 30 households to determine how many devices the household contained that family members watched video on (i.e. TV's, tablets, smart phones, etcetera). The dot plot of the responses is shown below.

Number of Screens Capable of Watching Video

1. How many households have three devices capable of showing video on them?

   (1) 1   (3) 7

   (2) 2   (4) 5

2. More households had 4 devices to watch video on than any other number. Which of the following is closest to the percent of households that have 4 devices?

   (1) 22%   (3) 27%

   (2) 34%   (4) 45%

On a recent Precalculus quiz, Mr. Weiler found the following distribution of scores, which are arranged in 5 point intervals (with the exception of the last interval)

5. How many students scored in the 75 to 79 point range?

   (1) 8   (3) 25

   (2) 10   (4) 5

6. Students do not pass the quiz if they receive lower than a 70. How many students did not pass?

   (1) 8   (3) 7

   (2) 5   (4) 15

7. How many total student took the quiz?

   (1) 25   (3) 56

   (2) 104   (4) 91

Score Range

8. Twenty-two students scored in the 80 to 84 range on this test. Does the histogram provide us with enough information to conclude that a student must have scored on 82 on this test? Explain your thinking.

36

6. The ages of 12 retail workers are given in the data set below

$$17, 18, 18, 19, 20, 21, 22, 23, 25, 25, 34, 47$$

(a) Calculate the five number summary. Label each of the five numbers with what they represents (i.e. min. max, lower quartile, etc.)

(b) Create a box-and-whiskers diagram of this data set below

```
  +----+----+----+----+----+----+----+----+----+----+--->
  10        20        30        40        50
```

A survey was done to determine the relationship between gender and subject preference. A total of 56 students were surveyed to determine if they liked math, English, social studies, or science as their favorite subject. The results were then broken down based on whether the respondent was male or female.

	Math	English	Social Studies	Science	Total
Female	8	6	11	5	30
Male	10	4	8	4	26
Total	18	10	19	9	56

1. Which of the following is closest to the joint relative frequency of being a male who likes social studies?

   (1) 0.42   (3) 0.31

   (2) 0.14   (4) 0.56

2. Which of the following is the marginal relative frequency of liking math?

   (1) $\dfrac{18}{36}$ (3) $\dfrac{10}{18}$

   (2) $\dfrac{8}{10}$ (4) $\dfrac{18}{56}$

3. What percent of female students liked English as their favorite subject?

   (1) 20%   (3) 11%

   (2) 16%   (4) 60%

4. A person looking at this table concludes that it is more likely that a female student will like social studies than a male student will like math. Is this correct? Justify your answer

ת