2019 AP Statistics – Summer Work

Please complete this packet in its entirety, this is Unit 1 of AP Statistics. By completing this unit over the summer we will have close to a month to review in April and May before the AP exam. When we return to school in the fall we will spend the first class period reviewing the packet and answering any questions you have. You will have a test over this packet the second class period of the new year and I will collect your work for this packet at that time.

Unit 1 – Exploring Data

Introduction

Statistics is the science of data. The volume of data available to us can be overwhelming. In many cases, the data is trying to tell us a story. To hear what the data re saying, we need to help them speak by organizing, displaying, summarizing, and asking questions. This is **data analysis**.

Any set of data contains information about some group of **individuals**. The characteristics we measure on each individual are called **variables**.

Definitions

Individuals are the objects describes by a set of data. Individuals may be people, animals, or things.

A **variable** is any characteristic of an individual. A variable can take different values for different individuals.

A **categorical variable** places an individual into one of several groups or categories. This is typically a word. For example, hair color (blond), gender (female), state you were born in (Nebraska).

A **quantitative variable** takes numerical values for which it makes sense to find an average. For example, age (56), height (67 inches), number of pets (2).

Example – Census at School

10 Canadian students completed a survey and the table below displays their data

Province	Gender	Languages Spoken	Handed	Height (cm)	Wrist circum. (mm)	Preferred communication
Saskatchewan	Male	1	Right	175	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178	174	Facebook
Ontario	Male	2	Right	169	160	Cell phone
Ontario	Female	2	Right	166	65	In person
Nunavut	Male	1	Right	168.5	160	Text
Ontario	Female	1	Right	166	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text
Ontario	Female	2	Right	150.5	187	Text
Ontario	Female	1	Right	171	180	Text

1a) Who are the individuals in this data set?
1b) What variables were measured?
1c) Identify each variable as categorical or quantitative.
1d) Describe the individual in the highlighted row.

Quantitative variables may take values that are very close together or values that are quite spread out. Categorical variables sometimes have similar counts in each category and sometimes don't. We call the pattern of variation of a variable its **distribution**.

Definition: Distribution

The **distribution** of a variable tells us what values the variable takes and how often it takes these values.

Section 1

Analyzing Categorical Data

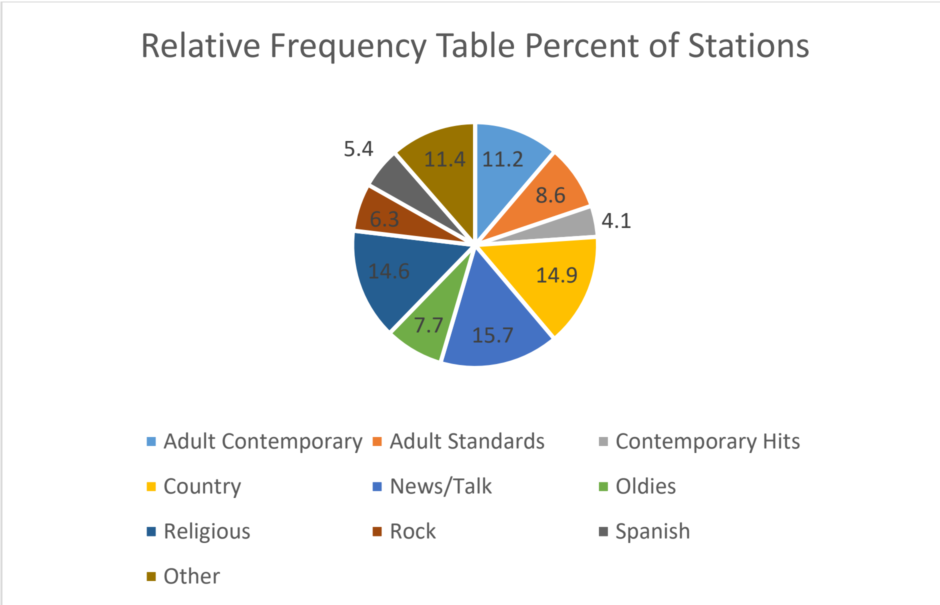
Columns of numbers take time to read. You can use a **pie chart** or a **bar graph** to display the distribution of a categorical variable more vividly.

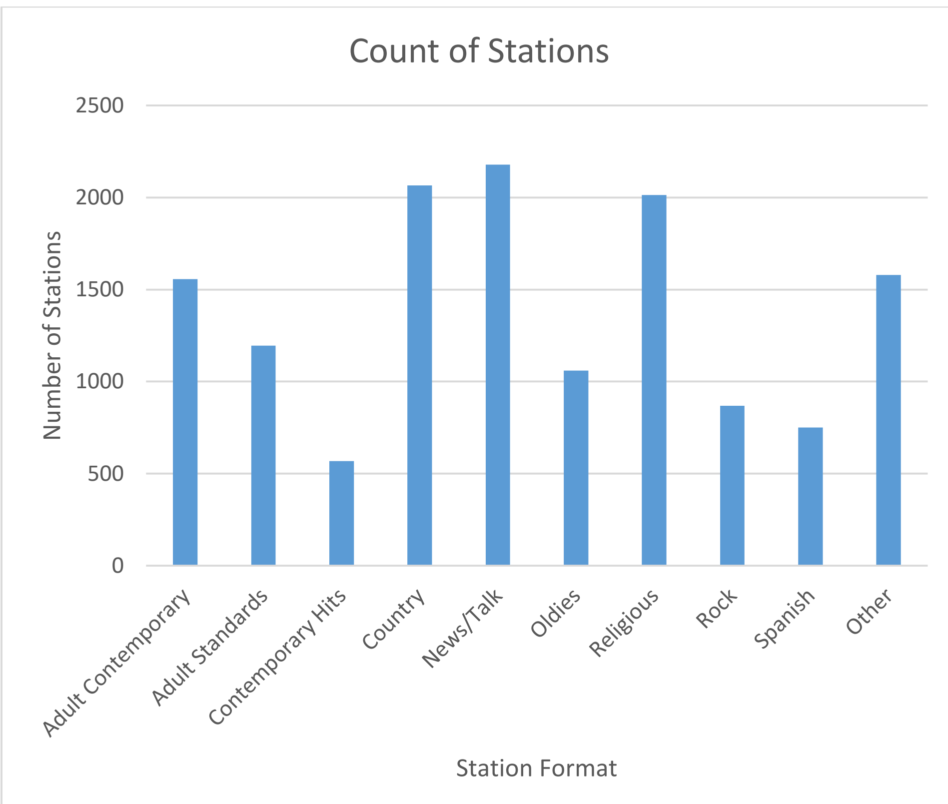
The table on the left is a **frequency table**. It displays counts (frequencies) of radio stations in each category. The table on the right is a **relative frequency table**. This table shows the percents of stations in each category.

Example – Radio Station Formats

Due to rounding error, the table below does not total 100%.

Frequency table			**Relative Frequency Table**	
Format	**Count of Stations**		**Format**	**Percent of Stations**
Adult Contemporary	1556		Adult Contemporary	11.2
Adult Standards	1196		Adult Standards	8.6
Contemporary Hits	569		Contemporary Hits	4.1
Country	2066		Country	14.9
News/Talk	2179		News/Talk	15.7
Oldies	1060		Oldies	7.7
Religious	2014		Religious	14.6
Rock	869		Rock	6.3
Spanish	750		Spanish	5.4
Other	1579		Other	11.4
Total	**13,838**		**Total**	**99.9**





Bar graphs and pie charts can both be easily constructed using Excel or Google Sheets. Both pie charts and bar graphs are useful in displaying categorical data.

In a bar graph it is VERY important to scale your axis accurately. Also, in a true bar graph, the bars do not touch each other.

Example – I'm Gonna Be Rich!

A survey of 4826 randomly selected young adults (aged 19 to 25) asked, "What do you think the chances are you will have much more than a middle-class income at age 30?" The table below shows the responses.

Young adults by gender and chance of getting rich			
	Gender		
Opinion	**Female**	**Male**	**Total**
No Chance	96	98	**194**
Some Chance	426	286	**712**
A 50-50 Chance	696	720	**1416**
A Good Chance	663	758	**1421**
Almost Certain	486	597	**1083**
Total	**2367**	**2459**	**4826**

This is a **two-way table** because it describes two categorical variables, gender and opinion about becoming rich. Opinion is the *row variable*. Gender is the *column variable*.

The total column at the right of the table contains the row totals. The total row at the bottom contains the totals for the column variables. These are distributions for each variable separately and are called **marginal distributions** because they appear in the margins of the table.

Definition – Marginal Distribution

The **marginal distribution** of one of the categorical variables in a two-way table of counts is the distribution of values of that variable among all individuals described by the table.

Percents are often more informative than counts, especially when we are comparing groups of different sizes. We can display the marginal distribution of opinion in percents by dividing each row total by the table total and converting to a percent. For example

$$\frac{Almost\ Certain\ Total}{Table\ Total} = \frac{1083}{4826} = 0.224 = 22.4\%$$

2a) Use the data in the two-way table to calculate the marginal distribution in percents of opinions.
2b) Make a graph to display the marginal distribution in percents.
2c) Describe what you see.

Example - Superpowers

A random sample of 415 children aged 9 to 17 from the United Kingdom and the United States completed a survey. Each student's country was recorded along with which superpower they would most like to have. The data is summarized in the table below.

	Country	
Superpower	**U. K.**	**U. S.**
Fly	54	45
Freeze Time	52	44
Invisibility	30	37
Superstrength	20	23
Telepathy	44	66

3a) Use the two way table to calculate the marginal distribution in percents of superpower preference.
3b) Make a graph to display the marginal distribution
3c) Describe what you see

A two-way table contains much more information than the two marginal distributions of opinion alone and gender alone. Marginal distributions tell us nothing about the relationship between two variables. To describe a relationship between two categorical variables, we must calculate some well-chosen percents from the counts given in the body of the table.

Young adults by gender and chance of getting rich			
	Gender		
Opinion	**Female**	**Male**	**Total**
No Chance	96	98	**194**
Some Chance	426	286	**712**
A 50-50 Chance	696	720	**1416**
A Good Chance	663	758	**1421**
Almost Certain	486	597	**1083**
Total	**2367**	**2459**	**4826**

We can study the opinions of women alone by looking only at the "Female" column in the two-way table. To find the percent of young women who think they are almost certain to be rich by age 30, divide the count of such women by the total number of women, the column total:

$$\frac{women\ who\ are\ almost\ certain}{column\ total} = \frac{486}{2367} = 0.205 = 20.5\%$$

Doing this for all five entries in the "Female" column gives the **conditional distribution** of opinion among women.

Definition: Conditional distribution

A **conditional distribution** of a variable describes the values of that variable among individuals who have a specific value of another variable. There is a separate conditional distribution for each value of the other variable.

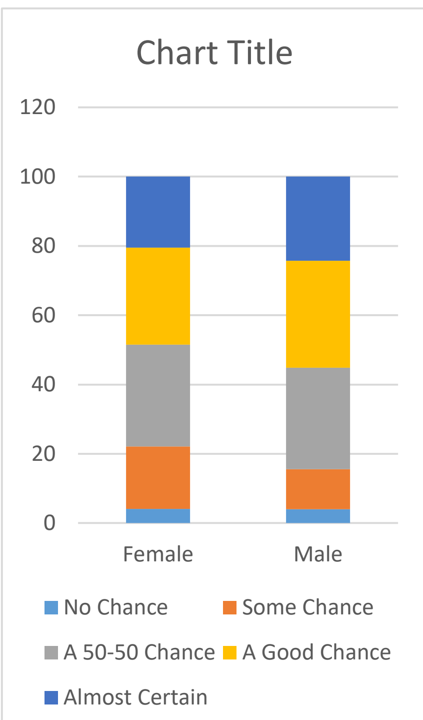
Video Link: https://www.youtube.com/watch?v=FWrEaSaW2mc

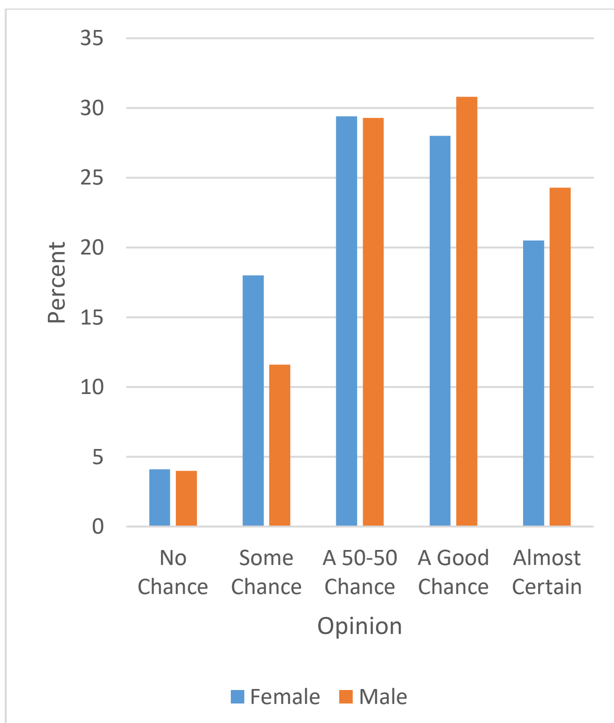
4) Calculate the conditional distribution of opinion among the young men.

5) Based on the survey data, can we conclude that young men and women differ in their opinions about the likelihood of future wealth? Give appropriate evidence to support your answer.

We could use a segmented bar graph to compare the distributions of male and female responses in the previous example. The figure below shows the completed graph. Each bar has five segments – one for each of the opinion categories.

A side-by-side bar graph, also shown below, can also make comparison easier.





Both graphs provide evidence of an **association** between gender and opinion.

Definition: Association

We say that there is an **association** between two variables if knowing the value of one variable helps predict the value of the other. If knowing the value of one variable does not help you predict the value of the other, then there is no association between the variables.

6a) Find the conditional distributions of superpower preference among the students from the United Kingdom and the United States.
6b) Make an appropriate graph to compare the conditional distributions
6c) Is there an association between country of origin and superpower preference? Give appropriate evidence to support your answer.

Section 2

Displaying Quantitative Data with Graphs

One of the simplest graphs to construct and interpret when working with quantitative data is a **dotplo**t. Each data value is shown as a dot above its location on a number line.

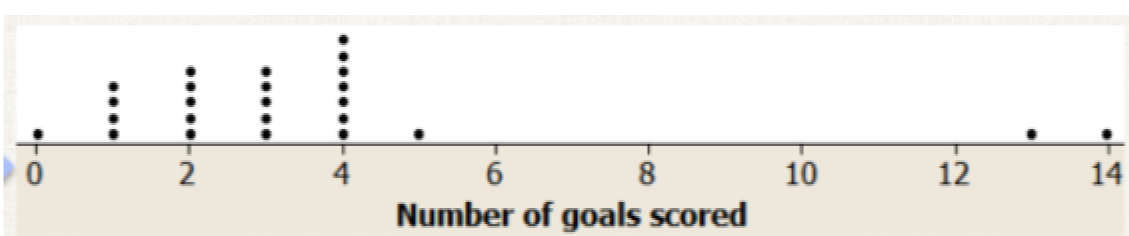
Example – GOOOOAAAAALLLLLL!

How good was the 2012 U.S. Women's soccer team? Here are the data on the number of goals scored by the team in the 12 months prior to the 2012 Olympics.

1	3	1	14	13	4	3	4	2	5	2	0	4
1	3	4	3	4	2	4	3	1	2	4	2	

Here are the steps in making a dotplot

- Draw a horizontal axis (number line) and label it with the variable name.
- Scale the axis so it covers the maximum and minimum values
- Mark a dot above the location on the horizontal axis corresponding to each data value.



How to examine and describe the distribution of a quantitative variable

In any graph, look for the **overall pattern** and for striking **departures** from that pattern.

- You can describe the overall pattern of a distribution by its **shape, center, and spread**.
- An important kind of departure is an **outlier**, and individual value that falls outside the overall pattern.

Shape: The dotplot has a peak at 4, a single main cluster of dots between 0 and 5, a large gap between 5 and 13.

Center: The midpoint of the 25 values shown in the graph is the 13th value, which is about 3 goals.

Spread: The data vary from 0 to 14 goals scored.

Outliers: The games in which the team scored 13 and 14 goals stand out from the overall pattern of the distribution and could be labeled as possible outliers.

Describing Shape

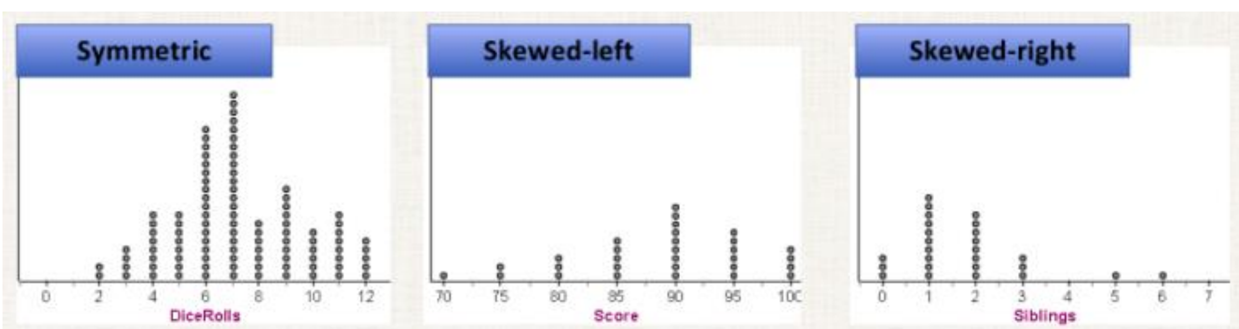
When describing shape, look for peaks, clusters, gaps, symmetry, or clear skewness.

Definition: Symmetric and skewed distributions

A distribution is roughly **symmetric** if the right and left sides of the graph are approximately mirror images of each other.

A distribution is **skewed to the right** if the right side of the graph (larger values) is much longer than the left side. It is **skewed to the left** if the left side of the graph is much longer than the right side.

***** ***The direction of skewness is the direction of the long tail*** *****



Using the graph of siblings above right answer the following questions

7a) Describe the shape of the distribution.
7b) Describe the center of the distribution.
7c) Describe the spread of the distribution.
7d) Identify any potential outliers.

Another simple graphical display is a **stemplot** (also called a stem-and-leaf plot). Here is an example that shows how to make a stemplot.

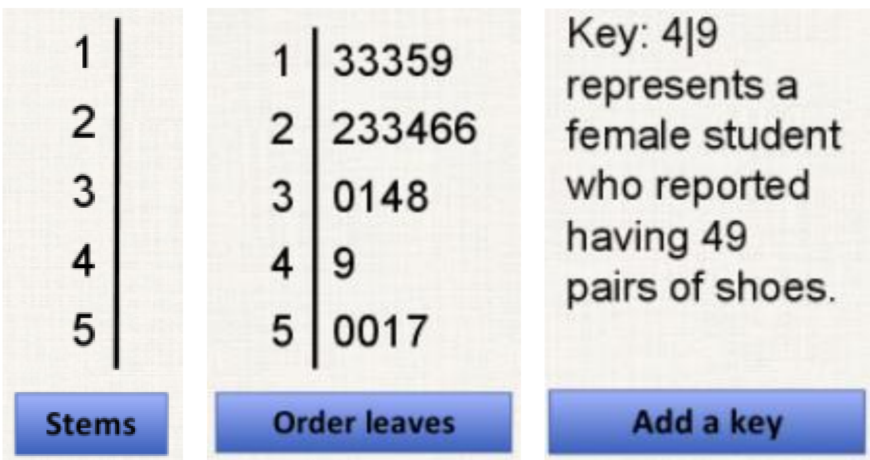
Example – How Many Shoes

A group of AP Statistics students surveyed 20 female students from their school and recorded the number of pairs of shoes each student reported having. Here are the data:

50	26	26	31	57	19	24	22	23	38	13	50
13	34	23	30	49	13	15	51				

Here are the steps in making a stemplot

- Organize the data from smallest to largest
- This data consists of two digit numbers. The tens digit will be the stem and the ones digit will be the leaf
- Write the stems in a vertical column with the smallest at the top and draw a vertical line at the right of the column
- Write each leaf to the right of the stem
- Add a key that explains in context what the stems and leaves represent.



Sometimes when data values are bunched up you can get a better picture of the data by **splitting stems** (make two rows for each stem, 0-4 and 5-9).

If you want to compare two sets of data it can be helpful to put both sets of data on a **back-to-back stemplot**.

The same AP Statistics asked 20 male students how many pairs of shoes they have. This is the data

14	7	6	5	12	38	8	7	10	10	10	11
4	5	22	7	5	10	35	7				

This is what the data looks like in a back-to-back stemplot with split stems.

Females		Males
	0	4
	0	555677778
333	1	0000124
95	1	
4332	2	2
66	2	
410	3	
8	3	58
	4	
9	4	
100	5	
7	5	

Key: 4|9 represents a student who reported having 49 pairs of shoes.

Example – Who is taller?

Who is taller, males or females? A sample if 14-year-olds from the United Kingdom was randomly selected and heights were measured (in cm).

Male:	154	157	187	163	167	159	169	162	176	177	151	175
	174	165	165	183	180							
Female:	160	169	1521	167	164	163	160	163	169	157	158	153
	161	165	165	159	168	153	166	158	158	166		

8) Create a back-to-back split stem stemplot comparing the heights of males and females

Histograms

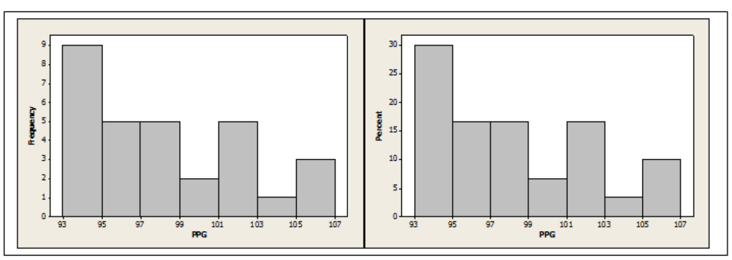
One very common type of graph used to display quantitative data is a **histogram**.

To illustrate this we are going to use the following data

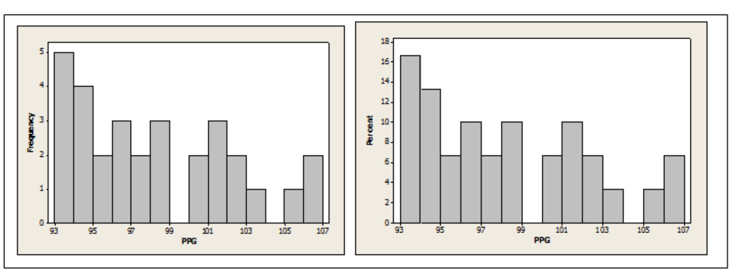
The table presents the average points scored per game (PPG) for the 30 NBA teams in the 2012–2013 regular season.

Team	PPG	Team	PPG	Team	PPG
Atlanta Hawks	98.0	Houston Rockets	106.0	Oklahoma City Thunder	105.7
Boston Celtics	96.5	Indiana Pacers	94.7	Orlando Magic	94.1
Brooklyn Nets	96.9	Los Angeles Clippers	101.1	Philadelphia 76ers	93.2
Charlotte Bobcats	93.4	Los Angeles Lakers	102.2	Phoenix Suns	95.2
Chicago Bulls	93.2	Memphis Grizzlies	93.4	Portland Trail Blazers	97.5
Cleveland Cavaliers	96.5	Miami Heat	102.9	Sacramento Kings	100.2
Dallas Mavericks	101.1	Milwaukee Bucks	98.9	San Antonio Spurs	103.0
Denver Nuggets	106.1	Minnesota Timberwolves	95.7	Toronto Raptors	97.2
Detroit Pistons	94.9	New Orleans Hornets	94.1	Utah Jazz	98.0
Golden State Warriors	101.2	New York Knicks	100.0	Washington Wizards	93.2

Because the smallest value is 93.2 and the largest value is 106.1, we are going to use bars of width 2 starting at 93. Here are a frequency histogram and a relative frequency (percents) for this data.



Now here are histograms using class widths of 1 starting at 93. The choice of class width can influence the appearance of a histogram. In the histograms below, we can see gaps from 99 to 100 and from 104 to 105.



What percent of your state's residents were born outside the United States?

State	Percent	State	Percent	State	Percent
Alabama	2.8	Louisiana	2.9	Ohio	3.6
Alaska	7.0	Maine	3.2	Oklahoma	4.9
Arizona	15.1	Maryland	12.2	Oregon	9.7
Arkansas	3.8	Massachusetts	14.1	Pennsylvania	5.1
California	27.2	Michigan	5.9	Rhode Island	12.6
Colorado	10.3	Minnesota	6.6	South Carolina	4.1
Connecticut	12.9	Mississippi	1.8	South Dakota	2.2
Delaware	8.1	Missouri	3.3	Tennessee	3.9
Florida	18.9	Montana	1.9	Texas	15.9
Georgia	9.2	Nebraska	5.6	Utah	8.3
Hawaii	16.3	Nevada	19.1	Vermont	3.9
Idaho	5.6	New Hampshire	5.4	Virginia	10.1
Illinois	13.8	New Jersey	20.1	Washington	12.4
Indiana	4.2	New Mexico	10.1	West Virginia	1.2
Iowa	3.8	New York	21.6	Wisconsin	4.4
Kansas	6.3	North Carolina	6.9	Wyoming	2.7
Kentucky	2.7	North Dakota	2.1		

Histograms on your calculator

1. Enter the data for the percent of state residents born outside the United States in your Statistics/List Editor.

    Press $\boxed{\text{Stat}}$ and right arrow over to edit
    Type the values into L1

2. Set up a histogram in the Statistics Plots Menu

    Press $\boxed{2^{nd}}$ $\boxed{\text{Y=}}$ (STAT PLOT)
    Press $\boxed{\text{ENTER}}$ or $\boxed{1}$ to go into plot 1
    Turn Plot 1 on
    Choose the third type (looks like a histogram)
    Xlist should say L1
    Freq should say 1
    If you have a choice of colors (depends upon what calculator you have) go crazy and choose whatever color you are fond of today.

3. Use ZoomStat to let the calculator choose classes (bar widths) and set an appropriate window.

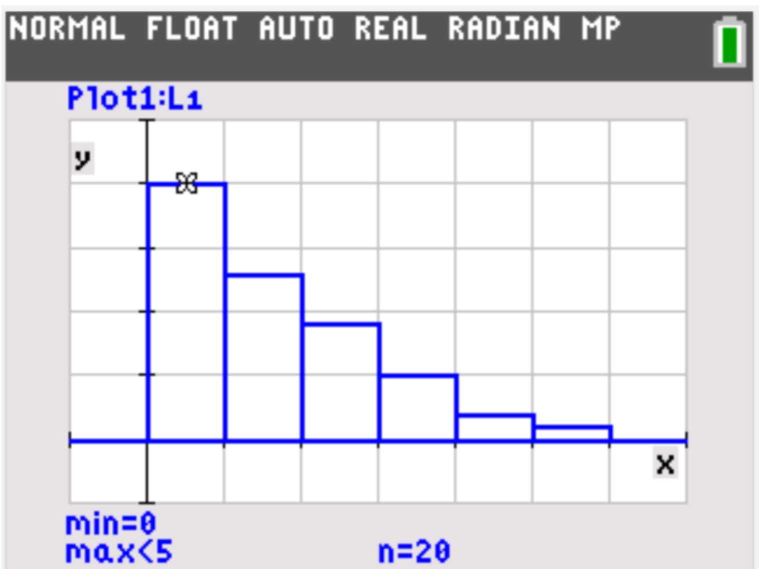
    Press $\boxed{\text{ZOOM}}$ and choose ZoomStat
    Choose TRACE and right and left arrow to see how the calculator set the classes (bar widths). Usually the calculator does a terrible job of choosing how wide to make the bars, so we need to fix this.

4. Adjust the classes

Press WINDOW and enter the following values

Xmin = -5
Xmax = 35
Xscl = 5
Ymin = -5
Ymax = 25
Yscl = 5

5. Now doesn't that look better? Your graph should look like this.



Video Link: https://www.youtube.com/watch?v=J3rTwaCQuUU&t=14s

Notice a major difference between a bar graph and a histogram is that on a histogram the bars touch (because they are numbers, not categories).

Section 3

Describing Quantitative Data with Numbers

How long do people spend traveling to work? Here are the travel times for 15 randomly chosen workers in North Carolina.

30	20	10	40	25	20	10	60	15	40	5	30
12	10	10									

We want to describe the center and spread of this set of data.

The most common measure of center is the ordinary average, or **mean**.

Definition: The mean $\bar{x}$

To find the mean $\bar{x}$ (pronounced "x-bar") of a set of observations, add their values and divide buy the number of observations.

$$\bar{x} = \frac{\sum x_i}{n}$$

9a) Find the mean travel time for the 15 workers.
9b) Calculate the mean again excluding the person who reported that their travel time was 60 minutes.
9c) What do you notice?

This exercise illustrates an important weakness of using the mean as a measure of center. The mean is very sensitive to the influence of extreme observations (outliers).

Another measure of center is the **median**, or midpoint, of a distribution.

Definition: median

The **median** is the midpoint of a distribution, the number such that about half of the observations are smaller and about half are larger. To find the median of a distribution:

- Arrange the observations in order of size, from smallest to largest
- If the number of observations is odd, the median is the number in the center of the list
- If the number of observations is even, the median is the average of the two numbers in the center of the list.

10) Find the median travel time for the 15 workers

Should you choose the mean or the median as a measure of center?

If the distribution of the data is skewed, or contains extreme values, you should use the median as a measure of center. The mean is affected by the skew or outliers and can present an inaccurate representation of center.

Measuring Spread: Range and Interquartile Range

A measure of center alone can be misleading. The mean annual temperature in San Francisco, California is 57 degrees, the same as Springfield, Missouri. But the wardrobe needed to live in these two cities is very different.

The simplest measure of variability is the **range.** To compute the range of a data set, subtract the smallest value from the largest value. For the travel time data the range is 60 – 5 = 55 minutes.

You can improve your description of the data by also looking at the spread of the middle half of the data. To do this:

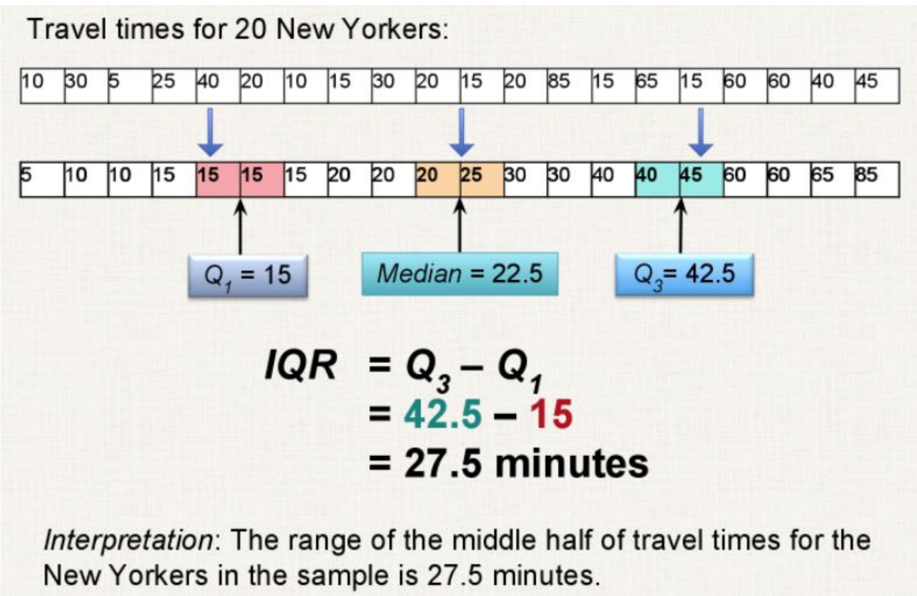
Count how many data points you have

The **first quartile** (Q1) lies one quarter up the list, starting with the smallest number

The second quartile is the median

The **third quartile** (Q3) lies three-quarters of the way up the list

From Q1 to Q3 is the middle half of the data. The **interquartile range** (IQR) measures the range of the middle half of the data, or Q3 – Q1.



In addition to serving as a measure of spread, the interquartile range (IQR) is used as part of a rule for identifying outliers. Video Link: https://www.youtube.com/watch?v=Cm_852R8JPw

Definition: The 1.5 x IQR rule for outliers

Call an observation an outlier if it falls more than 1.5 x IQR above Q3 or below Q1.

Does the 1.5 x IQR rule identify any outliers for the New York travel times above?

Q1 = 15 Q3 = 42.5 IQR = 27.5

1.5 (27.5) = 41.25

Q1 – 1.5 x IQR = 15 – 41.25 = -26.25

Q3 + 1.5 x IQR = 42.5 + 41.25 = 83.75

This means that any data values that are below -26.25 or above 83.75 would be classified as outliers. So for the New York travel time data, 85 is an outlier.

11) Determine if there are any outliers for the North Carolina travel times.

The Five-Number Summary and Boxplots

To get a quick summary of center and spread for a data set, the **five number summary** is often used.

> **Definition: The five number summary**
>
> The **five number summary** of a distribution consists of the smallest observation, the first quartile, the median, the third quartile, and the largest observation, written in order from smallest to largest. That is, the five number summery is
>
> Minimum, Q1, Median, Q3, Maximum

Example – Home Run King

Here are the number of home runs that Hank Aaron hit in each of his 23 seasons:

13	27	26	44	30	39	40	34	45	44	24	32
44	39	29	44	38	47	34	40	20	12	10	

To find the five-number summary:

1) Put the data in order from smallest to largest
2) Identify the median (middle)
3) Find Q1 (middle of the bottom half)
4) Find Q3 (middle of the top half)

10 12 13 20 24 26 27 29 30 32 34 34 38 39 39 40 40 44 44 44 44 45 47

So the five number summary is: 10 26 34 40 47

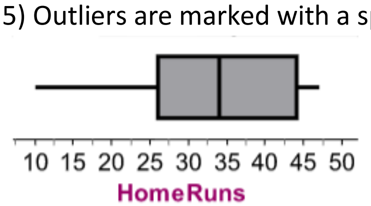
Are there any outliers?

26 – 1.5(44-26) = -1 44 + 1.5(44-26) = 71 so there are no outliers

The five number summary of a distribution leads to a new graph, the boxplot (sometimes called a box and whisker plot).

How to make a boxplot

1) Make a number line that stretches from the minimum to the maximum (at least)
2) A central box is drawn from Q1 to Q3
3) A line in the box marks the median
4) Lines (called whiskers) extend from the box out to the smallest and largest observations that are not outliers.
5) Outliers are marked with a special symbol such as a dot or an asterisk (*)



> 12) Construct a boxplot for the New York travel times.

How to sort data, do five number summaries, and boxplots on your calculator.

(Please don't hate me for not showing you this earlier)

First let's put some data in your calculator to play with.

1) Enter the travel time for North Carolina into L1 and the travel time for New York into L2. (see page 13 if you need instructions for this again).

2) Set up your stat plot. Plot 1 should be turned on, but this time instead of selecting the histogram icon for type you will choose the 4th one on that row (it's to the right of the histogram). Xlist should be L1, Freq should be 1, pick whatever mark and color you want.

3) Press zoom and select zoomstat.

4) If you press trace and use your left and right arrows you can see the five number summaries displayed on your screen.

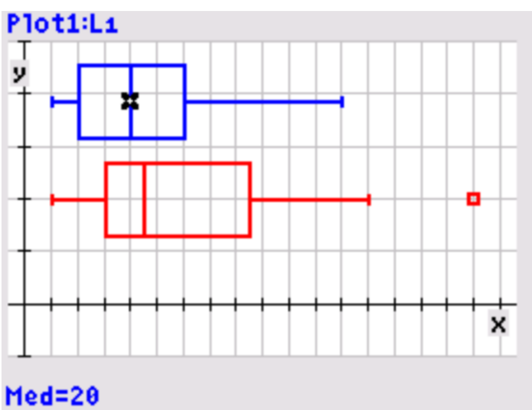
5) Now go back to your stat plot and turn on plot 2. For plot 2 choose the same type, xlist should be L2 (press 2nd, 2), freq 1, pick whatever mark you want. If you have a color calculator choose a different color than the one you chose for the first boxplot.

6) Press zoom and select zoomstat.

7) Press trace and you can use your up and down arrows to switch between boxplots and your left and right arrows to see the five number summary for each. Notice that the outlier is shown.

Video Link: https://www.youtube.com/watch?v=VvCw5MRo1P4

Your screen should look like this



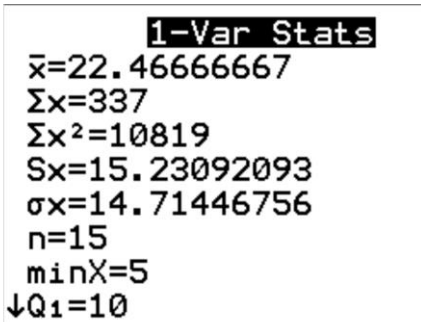
You can also get the mean and the five number summary alone without doing a boxplot.

1)Press 2^{nd}, Quit to go back to your home screen

2) Press STAT, right arrow over to CALC, and choose #1 (1-Var Stats)

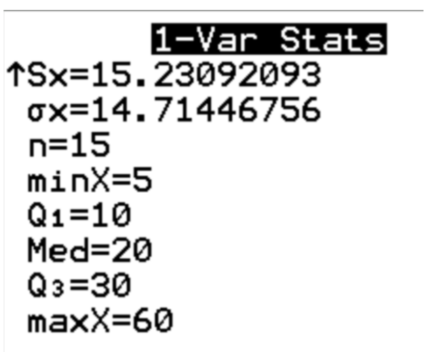
3) List should say L1, you can leave frequency blank, and press calculate.

You should have this on your screen



The number at the top, $\bar{x} = 22.47$ is the mean

The small arrow at the bottom left tells you that there is more stuff if you arrow down. Arrow down and your screen should look like this.



There is your five number summary. You can do the same thing for the data in L2.

The calculator will also sort the data for you.

1) Press 2^{nd}, Quit again to go back to your home screen

2) Press STAT, #2 – SortA

The A stands for ascending

3) Now you need to tell the calculator which list to sort. Press 2^{nd}, 1, close the parenthesis, and press enter and it will sort L1. Press 2^{nd}, 2, close the parenthesis, and press enter and it will sort L2. If you go back and look at your lists now they will be in order from smallest to largest.

Let's practice this calculator stuff

Start by clearing the travel time data out of your lists. The easy way to do this is press STAT, and then Enter. This should get you back to your lists. Up arrow so that L1 is highlighted. Press CLEAR. Then down arrow. Up arrow so that L2 is highlighted, Press CLEAR. Then down arrow. This should dump all the old data from your lists.

DO NOT press delete (instead of clear). If you do this the calculator will literally delete the L1 instead of clearing the data from L1. If you already made this mistake, DO NOT panic. Press STAT, #5: Set Up Editor, and then press ENTER and your lists will reappear.

Okay, now that you have empty lists to work with, enter the following data into L1.

The 2011 roster of the Dallas Cowboys football team included 8 offensive linemen. Their weights in pounds were

310 307 345 324 305 301 290 307

Use your calculator to

13a) Sort the data from smallest to largest
13b) Find the mean and the five number summary
13c) Create a box plot
13d) Determine if there are any outliers

Measuring Spread: The standard Deviation

The five number summary is not the most common numerical description of a distribution. That distinction belongs to a combination of the mean to measure center and the **standard deviation** to measure spread. The standard deviation and its close relative, the **variance,** measure spread by looking at how far the observations are from the mean.

Example – Foot lengths

Here are the foot lengths (in centimeters) for a random sample of seven 14 year olds

25 22 20 25 24 24 28

The mean foot length is 24 cm.

To calculate the standard deviation for this data set by hand (you will not have to do this, but I think going through this once so you can see where standard deviation comes from is helpful in understanding the concept of standard deviation, I will show you how to do this on your calculator in a minute or two) you need to find how far away each of these data points is from the mean, square them, and then add them up.

x	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
25	25 – 24 = 1	$(1)^2 = 1$
22	22 – 24 = −2	$(-2)^2 = 4$
20	20 – 24 = −4	$(-4)^2 = 16$
25	25 – 24 = 1	$(1)^2 = 1$
24	24 – 24 = 0	$(0)^2 = 0$
24	24 – 24 = 0	$(0)^2 = 0$
28	28 – 24 = 4	$(4)^2 = 16$
	Sum = 0	Sum = 38

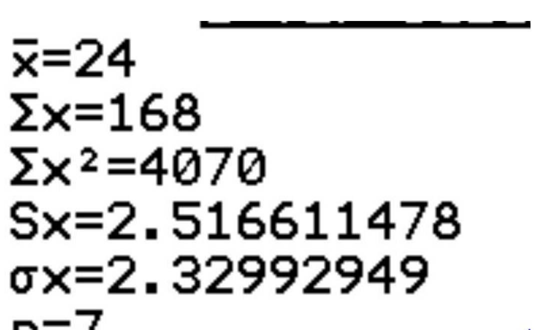
The variance is $\frac{38}{7-1} = 6.33$. The standard deviation is $\sqrt{6.33} = 2.52$ cm. This 2.52 cm is the typical distance each foot length is from the mean.

To find the variance and standard deviation on your calculator:

Enter the foot lengths into L1

Press STAT, arrow over to CALC, Do 1-VAR Stat on L1, CALCULATE

You should get this screen



See the mean? $\bar{x} = 24$. The standard deviation is labeled $\sigma x = 2.32992949$. To get the variance, take the standard deviation number and square it.

The heights (in inches) of five starters on a basketball team are 67, 72, 76, 76, and 84.

14a) Find the mean 14b) Find the standard deviation and the variance 14c) Interpret the standard deviation in this setting.

At this point you should begin reviewing for the test over this material. In your textbook additional problems can be found on pages 6-7, 20-24, 41-48, 69-73, and 76-81.

You are **<u>not</u>** required to do all of these problems. I will only be collecting your work for the questions contained in this packet. Do whatever you need to do to get enough practice to feel comfortable with this material. If you choose odd numbered problems the answers are in the back of the textbook.

Your test will consist of both multiple choice and free response questions.

Answers

1a) The 10 students who participated in the survey

1b and c)

Province – categorical

Gender – categorical

Languages Spoken – quantitative

Handed – categorical

Height – quantitative

Wrist – quantitative

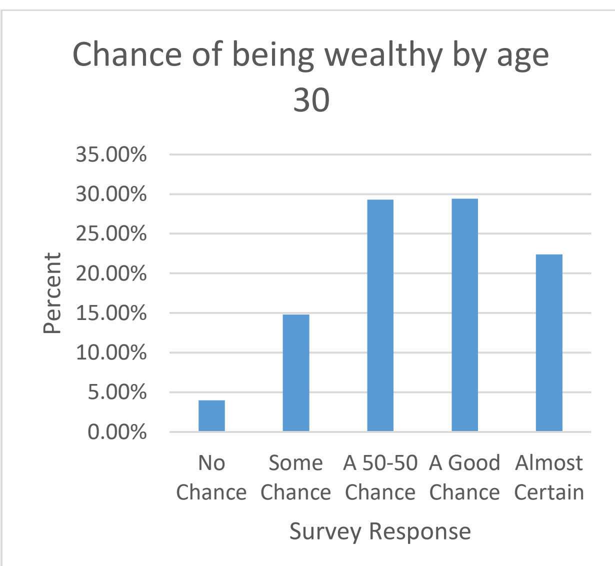
Preferred communication – categorical

1d) The student lives in Ontario, is male, speaks four languages, is left handed, 157.5 cm tall, has a wrist circumference of 147 mm and prefers to communicate with texting.

2a)

Response	**Percent**
No Chance	4.0%
Some Chance	14.8%
A 50-50 Chance	29.3%
A Good Chance	29.4%
Almost Certain	22.4%

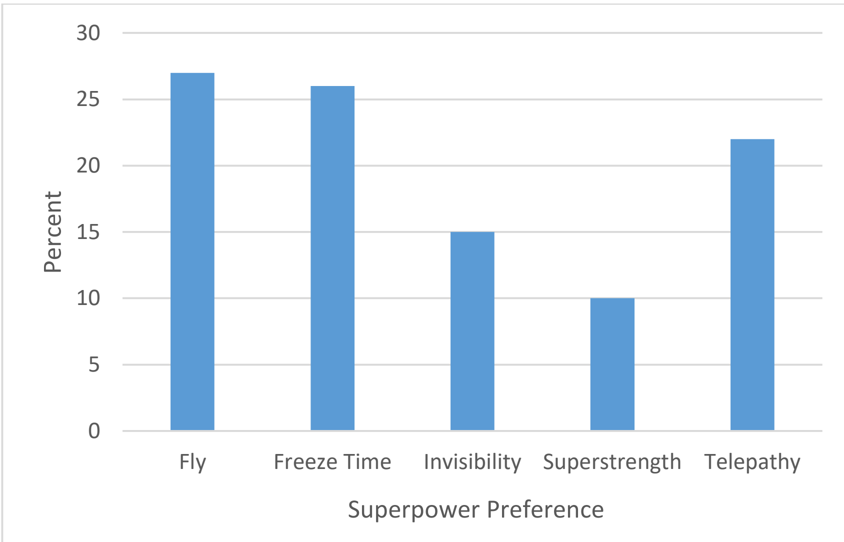
2b)



2c) It seems that many young adults are optimistic about their future income. Over 50% of those who responded to the survey felt that they had “a good chance” or were “almost certain” to be rich by age 30.

3a) Fly 23.9%, Freeze time 23.1%, Invisibility 16.1%, Superstrength 10.4%, Telepathy 26.5%

3b)



3c) It appears that telepathy, ability to fly, and ability to freeze time were the most popular choices, with about 25% of students choosing each one. Invisibility was the 4th most popular and superstrength was the least popular.

4)

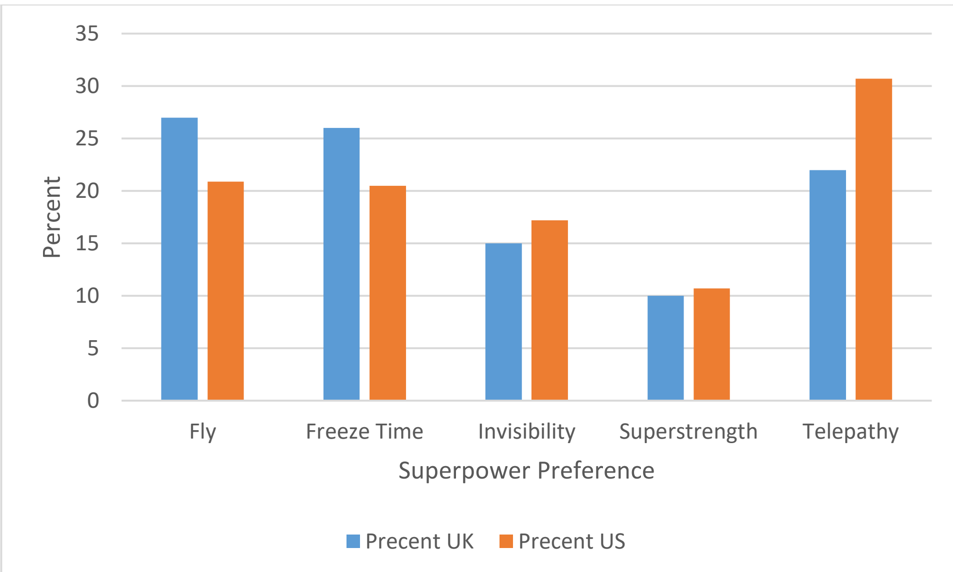
Conditional distribution of opinion among men	
Response	**Percent**
No Chance	4.0%
Some Chance	11.6%
A 50-50 Chance	29.3%
A Good Chance	30.8%
Almost Certain	24.3%

5) Based on the sample data, men seem somewhat more optimistic about their future income than women. Men were less likely to say that they have “some chance but probably not” than women (11.6% vs. 18.0%). Men were more likely to say that they have “a good chance” (30.8% vs. 28.0%) or are “almost certain” (24.3% vs. 20.5%) to have much more than a middle-class income by age 30 than women were.

6a)

Superpower	Precent	
	UK	US
Fly	27	20.9
Freeze Time	26	20.5
Invisibility	15	17.2
Superstrength	10	10.7
Telepathy	22	30.7

6b)



6c) There is an association between country of origin and superpower preference. Students in the U.K. are more likely to choose flying and freezing time, while students in the U.S. are more likely to choose invisibility or telepathy. Superstrength is about equally likely in both countries.

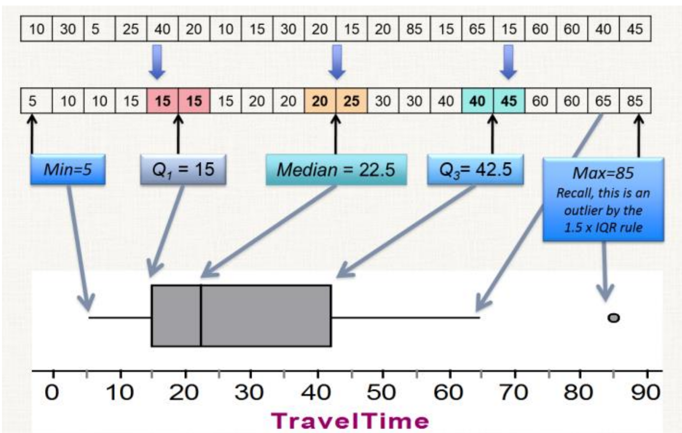
7a) This distribution is skewed to the right.
7b) The midpoint of the 28 values is between 1 and 2.
7c) The number of siblings varies from 0 to 6.
7d) There are two potential outliers at 5 and 6 siblings.

8)

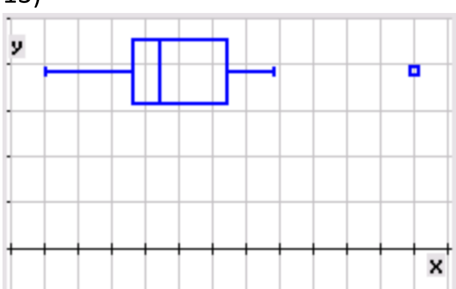
Female		Male
332	15	14
98887	15	79
433100	16	23
99876655	16	5579
	17	4
	17	567
	18	03
	18	7

Key: 15|1 represents a student who is 151 cm tall

9a) 22.5 minutes
9b) 19.8 minutes
9c) This one observation increased the mean by 2.7 minutes
10) 20 minutes
11) Values above 60 or below -20 would be outliers, so there are no outliers in this data set, although yes the 60 is really close.
12)



13)



1-Var Stats
$\bar{x}$=311.125

minX=290
Q_1=303
Med=307
Q_3=317
maxX=345

14a) 75
14b) Standard deviation = 6.24 inches   variance = 39 inches
14c) The players' height typically vary by about 6.24 inches from the mean height of 75 inches.