Name: ____ Block:______________

AP Stats Summer Assignment 2020

What you need to know about AP Statistics…

The AP Stats course is designed to make you think differently about data, in a time where data is often being misconstrued. We explore data through different mediums leading up to the most important topic in all of statistics (where you see the infamous "a new study found…". The four main topics covered in this class include: Descriptive Statistics, Experimental Design, Probability, and Inference.

Workload. Traditionally, we do most learning in class and homework is optional. With the unknown of how the fall will look, I cannot entirely speak to how the workload will look, but I anticipate the bulk of the learning will come on your end through videos and other resources (but hopefully not!).

Grades. Each unit will consist of ***one investigation (lab), two - three quizzes,*** and a ***unit test***, all of which count as summative grades. The AP test consists of 40 multiple choice questions and 6 free response. Tests and quizzes are designed like the AP test with a mix of multiple choice and free response questions.

Technology. You will need a graphing calculator for this course. Either a Ti-84 or NSpire will suffice. If you do not have a graphing calculator, Androids have an emulator called Wabbitemu which has the Ti-84 as well. For iOS devices, the GrafNCalc83 is not the Ti-84, but it has all the features needed. There is also Desmos, however, the learning curve is slightly different. As of now College Board does not allow Desmos, but this may change.

The summer assignment… Please review the resources provided for chapter 1 and complete the questions that accompany it. The powerpoints and questions are split up by section so that you can work on one at a time.

It should be noted that this summer assignment does not count as a summative grade. Don't stress, you have seen most of this material before in Algebra 1 and 2, it just adds a few new topics. Understanding this material will only benefit you as we dive into AP Statistics in the fall.

Please feel free to reach out to me with any questions, comments, or concerns. I look forward to a great year and am excited to meet all of you!

Dermot McGee
dermot.mcgee@lcps.org

Introduction 1.0 and 1.1 – Statistics: The Science and Art of Data/Analyzing Categorical Data

1. What variables are measured? Identify each as categorical or quantitative. In what units were the quantitative variables measured?

State	Number of Family Members	Age	Gender	Marital Status	Total Income	Travel time to work
Kentucky	2	61	Female	Married	21000	20
Florida	6	27	Female	Married	21300	20
Wisconsin	2	27	Male	Married	30000	5
California	4	33	Female	Married	26000	10
Michigan	3	49	Female	Married	15100	25
Virginia	3	26	Female	Married	25000	15
Pennsylvania	4	44	Male	Married	43000	10
Virginia	4	22	Male	Never married/ single	3000	0
California	1	30	Male	Never married/ single	40000	15
New York	4	34	Female	Separated	30000	40

2. A sample of 200 children from the United Kingdom ages 9-17 was selected from the CensusAtSchool website (www.censusatschool.com). The gender of each student was recorded along with which super power they would most like to have: invisibility, super strength, telepathy (ability to read minds), ability to fly, or ability to freeze time. Here are the results:

	Female	Male	Total
Invisibility	17	13	30
Super Strength	3	17	20
Telepathy	39	5	44
Fly	36	18	54
Freeze Time	20	32	52
Total	115	85	200

a. What proportion of males want the power of invisibility?

b. What proportion of females want the power of freeze time?

c. What proportion of children that want the power of telepathy are male?

d. What proportion of children that want the power of fly are female?

3. Create a well labeled segmented bar graph of the marginal distributions of power preference and gender. Be sure to include a key.

Key:

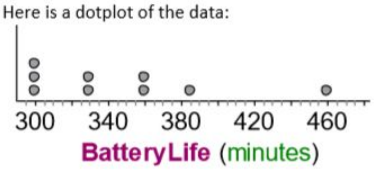
4. Based on the graphs above, can we conclude that boys and girls differ in their preference of superpower? Give appropriate evidence to support your answer.

Section 1.2 – Displaying Quantitative Data with Graphs

Smart Phone Battery Life

Smart Phone	Battery Life (minutes)
Apple iPhone	300
Motorola Droid	385
Palm Pre	300
Blackberry Bold	360
Blackberry Storm	330
Motorola Cliq	360
Samsung Moment	330
Blackberry Tour	300
HTC Droid	460

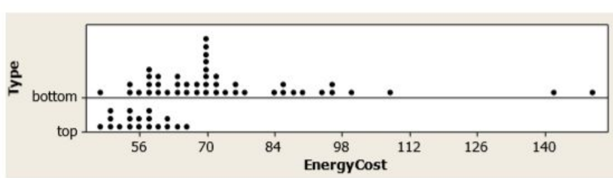
Here is a dotplot of the data:



1. Describe the shape, center, and spread of the distribution. Are there any (potential) outliers?

Top vs. Bottom Freezers

How do the annual energy costs (in dollars) compare for refrigerators with top freezers and refrigerators with bottom freezers? The data below is from the May 2015 issue of Consumer Reports.



2. Compare the distributions of annual energy costs for these two types of refrigerators.

Who's Taller?

Which gender is taller? A sample of 14-year-olds from the United Kingdom was randomly selected using the CensusAtSchool website.

Here are the heights of the students (in cm):

Male: 154, 157, 187, 163, 167, 159, 169, 162, 176, 177, 151, 175, 174, 165, 165, 183, 180 Female: 160, 169, 152, 167, 164, 163, 160, 163, 169, 157, 158, 153, 161, 165, 165, 159, 168, 153, 166, 158, 158, 166

Here is a back-to-back stemplot comparing male and female heights:

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.

3. Compare the distributions of height for females and males.

Section 1.3 - Describing Quantitative Data with Numbers

McDonald's Beef Sandwiches

Here are data for the amount of fat (in grams) for McDonald's been sandwiches:

Sandwich	Fat (g)
Hamburger	9 g
Cheeseburger	12 g
Double Cheeseburger	23 g
McDouble	19 g
Quarter Pounder®	19 g
Quarter Pounder® with Cheese	26 g
Double Quarter Pounder® with Cheese	42 g
Big Mac®	29 g
Big N' Tasty®	24 g
Big N' Tasty® with Cheese	28 g
Angus Bacon & Cheese	39 g
Angus Deluxe	39 g
Angus Mushroom & Swiss	40 g
McRib ®	26 g
Mac Snack Wrap	19 g

Use a graphing calculator to find the following:

Mean	
Median	
5 Number Summary	
IQR	
Are there any outlier/s using the IQR*1.5 Rule?	

The Previous Home Run King

Using a graphing calculator, create a box plot using the data below. Be sure to identify each number in a five number summary and any outliers using the IQR*1.5 Rule.

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

Who Has More Contacts – Males or Females?

The following data show the number of contacts that a sample of high school students had in their phones. Do the data give convincing statistical evidence that one gender has more contacts than the other? You need both graphical and numerical evidence.

Male: 124 41 29 27 44 87 85 260 290 31 168 169 167 214 135 114 105 103 96 144
Female: 30 83 116 22 173 155 134 180 124 33 213 218 183 110

AP Stats Chapter 1

Chapter 1.0/1.1 What is Statistics and how do I interpret it?

Statistics is the science of Data and Variability

This course is broken up into two categories

Descriptive Statistics | Statistical Inference

Population vs. Samples

Population all data of a particular set

Sample a subset of a population

This class is a sample of the entire Rock Ridge population.

When you make a calculation from a population it is called a parameter

When you make a calculation from a sample it is called a statistic

Individuals are objects described by a set of data. who/what

Variable is any characteristic of an individual. This can be broken up into two groups.

- **Categorical Variables –** places an individual into one of several groups or categories
  Can't take an average of
  Examples – color , ice cream flavor, gender

- **Quantitative Variables** – takes numerical values for which it makes sense to find an average
  Can take average and measure in units
  Examples – Height, Weight, Age, Shoe Size

Examples:

1. Jake is a car buff who wants to find out more about the vehicles that students at his school drive. He gets permission to go to the student parking lot and record data. Later, he does some research about each model of car online. Finally, Jake makes a spreadsheet that includes each car's model, year, color, number of miles on the car, gas mileage, weight, and whether it has a navigation system.

   a. Who/what are the individuals in Jake's study?

   Cars in the student parking lot

   b. What variables did Jake measure? Identify each as categorical or quantitative.

   Categorical: Model, color, navigation system
   Quantitative: Number of miles on car, gas mileage, weight

   Year?

2. The Gallup Poll conducted a representative telephone survey of 1180 American voters during the first quarter of 1999. Among the reported results were the voter's region (Northeast, South, etc.), age, party affiliation, and whether or not the person had voted in the previous election.

   a. What is the population of interest for this problem?

   American Voters

   b. What is the sample for this problem?

   1180 American Voters

   c. What variables were measured? Identify each as categorical or quantitative.

   Categorical: Region, party affiliation, voted in previous election

   Quantitative: Age

Analyzing Categorical Data

A recent survey of people at the Alamo movie theater were asked their favorite type of movie genre.

Frequency Table

Genre	Frequency Count
Comedy	67
Romance	26
Drama	44
Horror	43
Sci-Fi	36
Family/Animated	12
Total	**228**

What is the relative frequency of … %

a. Comedy 67/228 = 29.4%

b. Drama 44/228 = 19.3%

c. Sci-fi 36/228 = 15.8%

d. Horror 43/228 = 18.9%

****The sum of all of the relative frequencies should always equal 100 or round to it. Typically round all decimals to the tenth or hundredth place.**

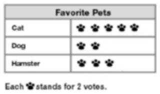
When your data is categorical it can be displayed 2 different ways.

Bar graph (overlap can happen; Show % in groups) or Pie chart/ circle graph (no overlap; part of a whole (100%))

Things to keep in mind when making a bar graph.

1. It is called a BARgraph for a reason. USE BARS NOT PICTURES!

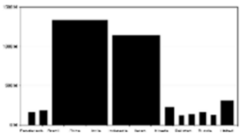
NO!



BAD!

2. Watch your scales! Make sure all the bars are equally wide.

EEEEK!

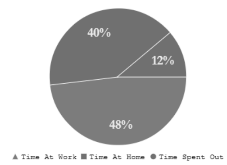


BAD!

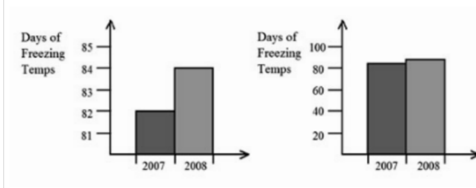
3. Make bars disjointed. They can go in any order because they are categorical.

4. It is rare that you would ever have to draw a pie chart by hand, there are many websites out there that will do it for you.

How my time is spent in a week?



A good pie chart



They display the same info, but notice how simple changes can change the interpretation. Data is in the eye of the beholder

Two way tables are created when a situation contains 2 categorical sets of data (the row variable and the column variable). We can focus on the count or the percent.

This is known as a **contingency table**. The relationship here is gender and most used phone app.

Young adults by gender and most used phone app — phone app by gender

	Female	Male	Total
Twitter	63	58	121
Instagram	79	33	112
Facebook	39	28	67
Total	181	119	300

Female
Twitter: $\frac{63}{181} = 34.8\%$
Insta: $\frac{79}{181} = 43.6\%$
Fb: $39/181 = 21.5\%$

male
$58/119 = 48.7\%$
$33/119 = 27.7\%$
$28/119 = 23.5\%$

The **marginal distribution** of one of the categorical variables in a two-way table of counts. It's the distribution of values of that variable among all individuals by the table. (think row or column)

APPS: Twitter: $\frac{121}{300} = 40.3\%$ Insta: $\frac{112}{300} = 37.3\%$ Fb: $67/300 = 22.3\%$

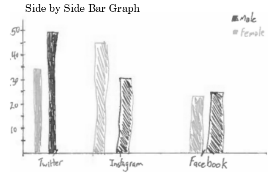
A **conditional distribution** of a variable describes the values of that variable among individuals who have a specific value of another variable.

listed above: if we said conditional distribution for phone app by gender

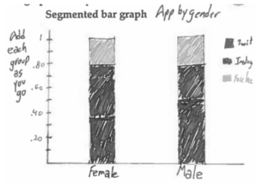
the other alternative would be gender by phone app which would list all apps followed by male/female for both

Here is how to display either marginal or conditional distributions

Side by Side Bar Graph



Segmented bar graph App by gender



So how can we tell if there is an association between two categorical variables?

- Find the conditional distribution for one variable by another.

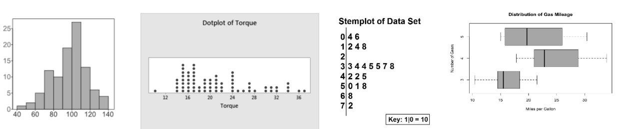
- If the percentages of each group are relatively the same across the categories we can say that there is no association (or independent).

- If there is a notable difference we can say that there is an association (or dependent)

Using the example from before, it appears that because there is a notable difference between each of the groups for the most used app and gender, there is an association between them.

Chapter 1.2 Displaying and Describing Quantitative Data

We will be using Histograms, Dotplots, Stem/Leaf plots, and Boxplots to describe quantitative data



Review of Shape

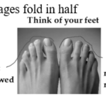
Symmetric – if the right and left sides of the graph are approximately mirror images fold in half

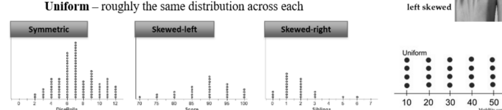
Skewed right – tall on the left side with a long low tail on the right

Skewed left – tall on the right side with a long low tail on the left

Uniform – roughly the same distribution across each

Think of your feet





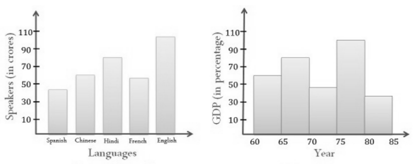
Shape is basically asking "Where is the tail?"

Quantitative Data continued... Histograms

Quantitative variables often take many values. A graph of the distribution may be clearer if nearby values are grouped together. The most common graph of the distribution of one quantitative variable is a **histogram.**

How to make a histogram:

1) Divide the range of data into classes of equal width.
2) Find the count (frequency) or percent (relative frequency) of individuals in each class.
3) Label and scale your axes and draw the histogram. The height of the bar equals its frequency. Adjacent bars should touch, unless a class contains no individuals.



Bar Graph **Histogram**

Making a stemplot

Splitting Stems

When data values are "bunched up", we can get a better picture of the distribution by **splitting stems.**

96	4
96	78
97	224
97	677888999
98	00000122222233344444
98	56666777777888888889
99	00112234
99	9
100	0
100	8

Back-to-Back Stemplots

Two distributions of the same quantitative variable can be compared using a **back-to-back stemplot** with a common stem

pulse rate

before		after
9 8 8 8	6	
8 6 6 4 1 1 0	7	
8 8 6 2	8	6 7 8 8
6 0	9	0 2 2 4 5 8 9 9
4	10	0 4 4
0	11	8
	12	4 4

The purpose of making a graph is to understand the data. After you make a graph you should always think "What do I see?" remember your SOCS

Shape ______ Outlier ______

Center ______ Spread ______

Example: A cashier at Target tracks how much 30 randomly selected shoppers spend by gender.

Female

$133	$132	$95	$97	$118
$87	$97	$120	$139	$120
$119	$120	$127	$133	$137

Male

$82	$103	$138	$127	$96
$83	$116	$100	$109	$96
$94	$105	$108	$112	$114

Create a back-to-back stemplot.

F		M
7	8	2 3
7 7 5	9	4 6 6
	10	0 3 5 8 9
8	11	2 4 6
7 0 0 0	12	7
9 7 3 3 2	13	8

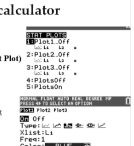
Compare the two distributions (think about SOCS)

There does not appear to be any outliers in either group.
The data of female shoppers appears skewed left while men are roughly symmetric. Females have a larger average of dollars spent at Target compared to men (about $120 and $105 respectively). The female distribution has a gap and an uneven spread of data compared to the male's.

How to construct in the calculator

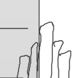
Let's look at how we can create a histogram on our calculators.

1. Once data is entered into L1 press **2nd Y= (to take you to Stat Plot)**
2. Select Plot 1 and turn it on. Under Type change the graph to a Histogram.
3. The X-List should be whatever list you placed the data (in most cases this will just be L1) You can get L1 by pressing **2nd** and the number **1**. Leave Freq as 1.
4. Press the **Zoom** button and press **9** to get to **Zoom Stat** this is a zoom feature we will use a lot in this class.

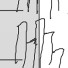


The many shapes of Histograms

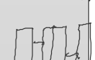
- Peaks or bumps in a histogram are called **modes**



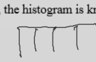
- Histograms with one peak is called **unimodal**
- Histograms with two peaks is called **bimodal**



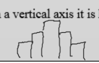
- Histograms with three or more peaks is called **multi-modal**



- If bars are roughly the same height, the histogram is known as **uniform**



- If the histogram can be folded on a vertical axis it is known as **symmetric**



Describing a distribution

It is very rare that you would need to draw a histogram on the AP Exam. Often times you have to interpret one, so make sure you have a solid understanding of **S.O.C.S.**

The histograms below show the distribution of ages for female and male heart attack patients. Make a comparison of the two models.