Welcome to Probability and Statistics!

This summer packet is for all students enrolled in **Probability and Statistics** at Herndon High School for Fall 2021.

This summer assignment is not *required*, but it is *strongly recommended*. The exercises will give you the opportunity to self-assess how prepared you are for **Probability and Statistics** this year.

Provided at the end is a key for you to check your work. Be sure to keep track of sticky spots and ask questions when we return. You are also welcome to reach out to us over the summer; our contact information is below.
Dan Logan: drlogan1@fcps.edu
Carol Treakle: cltreakle@fcps.edu

FCPS recommend activities for each level of mathematics are also posted on the Herndon High School website. Both resources will help you prepare for next year.

Have a great summer – we are looking forward to meeting you in August!

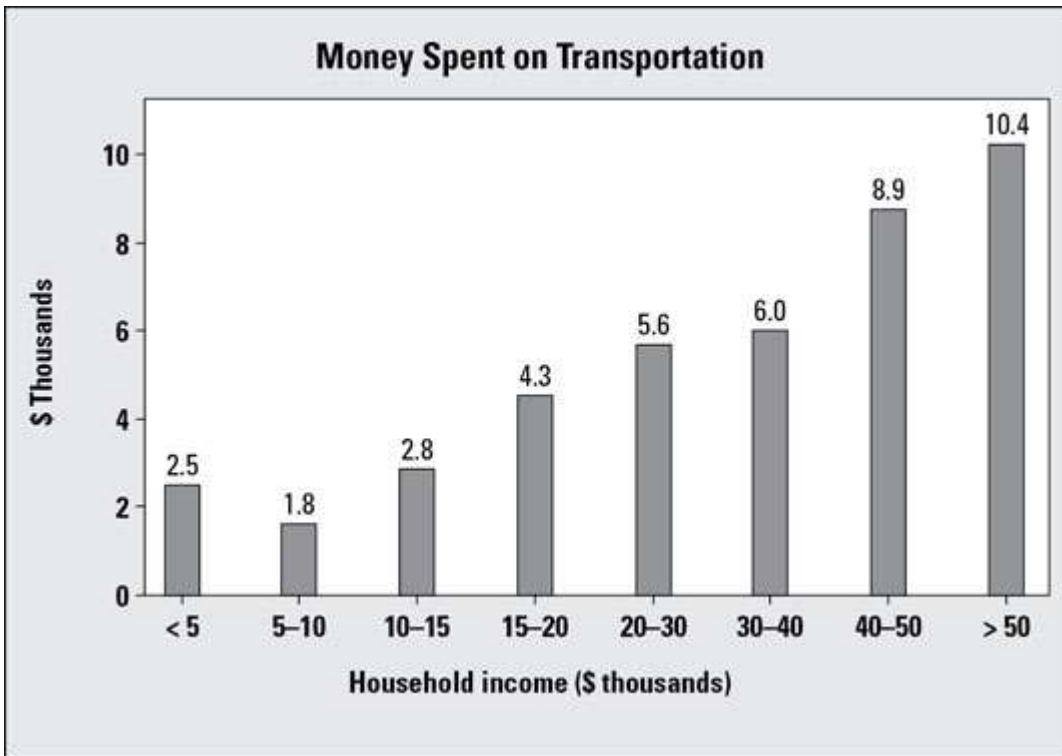**As you work through the packet, keep track of the following:**

"Things I learned, but forget how to do:"	"Things I never learned:"

Probability and Statistics Summer 2021 Packet

We will be creating and interpreting different types of graphs in the course. One of the goals of the class is to become informed interpreters of graphs.

I. Watch the video: The Beauty of Data Visualization
    https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization

II. Answer the following questions about the bar graph below:

Money Spent on Transportation

Household income ($ thousands)	$ Thousands
< 5	2.5
5–10	1.8
10–15	2.8
15–20	4.3
20–30	5.6
30–40	6.0
40–50	8.9
> 50	10.4

1) Do we know who collected this information? _____

2) Do we know how the information was collected? _____

3) Do we know when the information was collected? _____

4) Is it important to know how the information was obtained for this graph? Why or why not?

5) How much money does someone who makes $25,000 spend on transportation? _____

6) What is the minimum amount spent on transportation? _____

7) What is the maximum amount spent on transportation? _____

8) What is the minimum amount of household income? _____

9)Does money spent on transportation increase with every increase in income? If not, at what income levels does it not increase?

10)Does it make sense that money spent on transportation generally increases with income level? Why or why not?

11)Now, calculate the amount spent on transportation as a percent of total income:

Percent is: part/whole   2,500 spent on transportation/household income of 5,000 = 2500/5000 = .5 = 50%

< 5   _____50%___

5-10   _____

10-15   _____

20-30   _____

30-40   _____

40-50   _____

>50   _____

12) Does your conclusion about the amount of money spent on transportation change based on these figures? If so, how? Which category spends the most percent of their income on transportation?

13) What is a problem with the way the groupings are set up? For example, which group would you put someone with an income of exactly 40,000
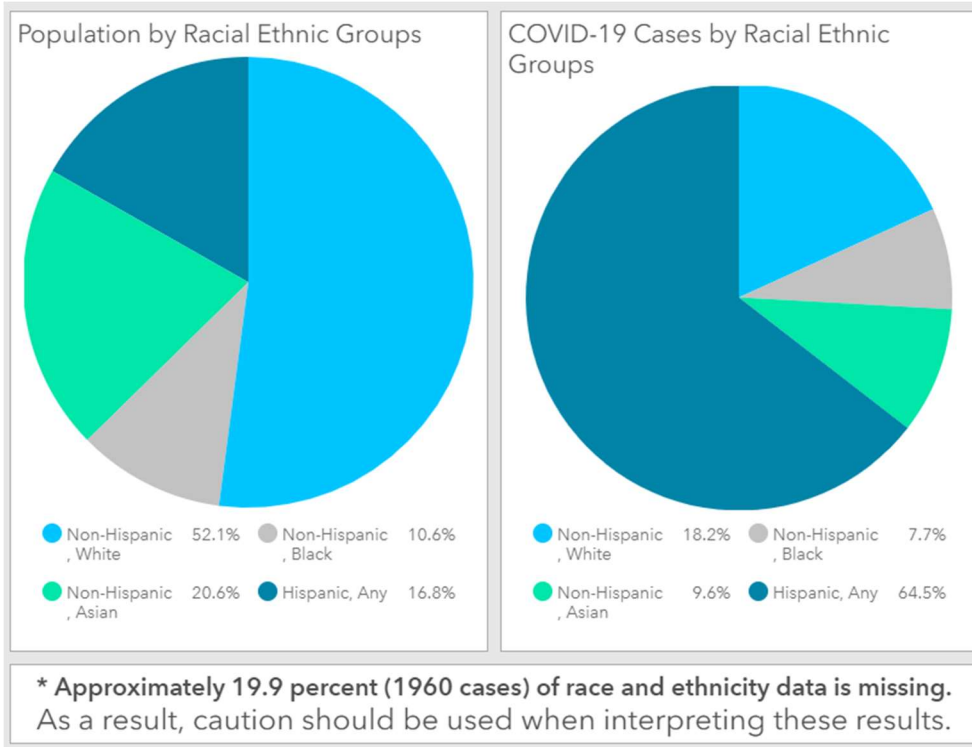
14)How could you change the groupings so that an income could not be placed in two categories?

15) What else do you notice about the graph groupings that is inconsistent and could cause a misrepresentation of the data?

III.   Watch the following video on how statistics can be misleading:
         https://www.ted.com/talks/mark_liddell_how_statistics_can_be_misleading

         Then use the following graph to answer the questions:

Fairfax County Health Department Cases by Racial Ethnic Group

Population by Racial Ethnic Groups

COVID-19 Cases by Racial Ethnic Groups

● Non-Hispanic, White	52.1%	● Non-Hispanic, Black	10.6%
● Non-Hispanic, Asian	20.6%	● Hispanic, Any	16.8%

● Non-Hispanic, White	18.2%	● Non-Hispanic, Black	7.7%
● Non-Hispanic, Asian	9.6%	● Hispanic, Any	64.5%

* Approximately 19.9 percent (1960 cases) of race and ethnicity data is missing. As a result, caution should be used when interpreting these results.

Note: Data is from May 15th

1) Which ethnic group has the highest population? _____
2) Which ethnic group has the highest number of Covid 19 cases? _____
3) What are some possible reasons that the group with the highest population does not have the highest number of Covid-19 cases?

4) How could county health experts use this information to provide Covid-19 support services?

5)   Why should "caution be used when interpreting these results" ?

6) Do the graphs give a true representation of population data and Covid-19 cases?

IV.   Watch the following video on relative vs absolute risks:
       https://www.ted.com/talks/gerd_gigerenzer_why_do_people_fear_the_wrong_things

       Think about the following: Imagine you could lower your chance of being struck by lightning from 0.5% to 0.4%. Which makes you feel safer, a 0.1% decrease in chance that you will be struck by lightning, or being 20% less likely to be struck by lightning? Think about why being clear about the use of relative and absolute risks when giving people information.
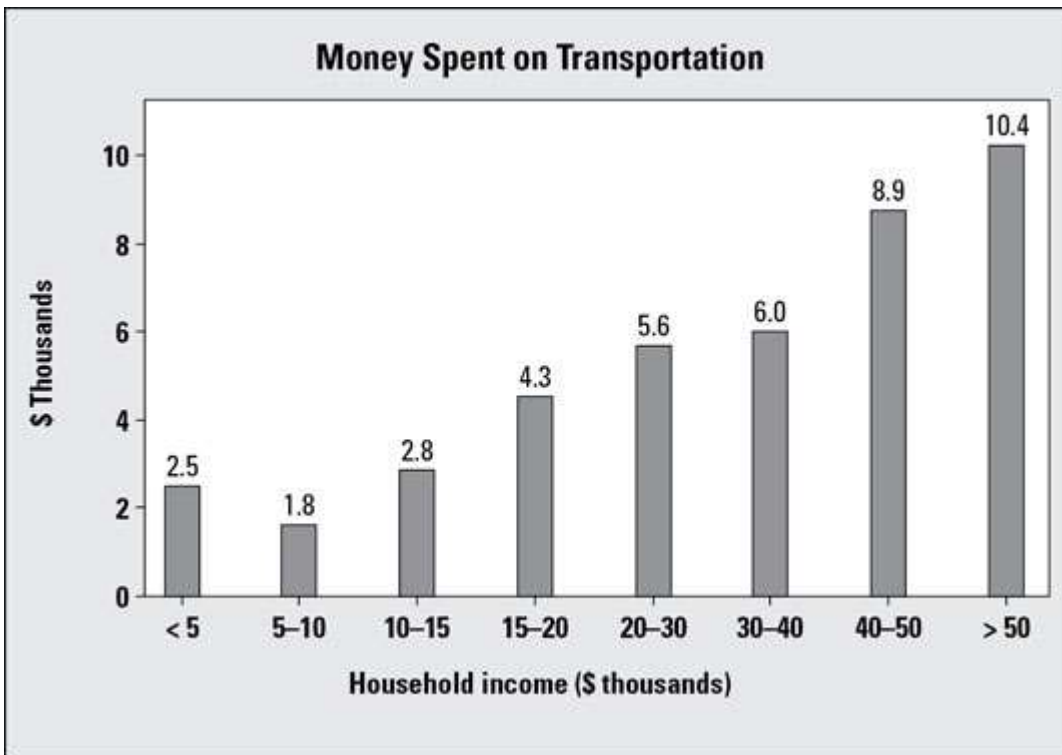
V.   And last, watch the following video on how why taking Statistics can be more important than Calculus:
       https://www.ted.com/talks/arthur_benjamin_teach_statistics_before_calculus

       Struggling in mathematics is not the enemy, any more than sweating is the enemy in basketball; it is part of the process, and a clear sign of being in the game. Math asks our students to think in ways they are not used to thinking; they will be asked to look at the obvious in ways they're not accustomed to, and then we'll ask them to explore the not-so-obvious in similar ways. A rigor of thinking and a clarity of expression is demanded that will stretch them beyond familiar styles. It will also require an honest pursuit; there really are no shortcuts.

Answer the following questions about the bar graph below:

Money Spent on Transportation

A bar graph titled "Money Spent on Transportation" with the y-axis labeled "$ Thousands" ranging from 0 to 10, and the x-axis labeled "Household income ($ thousands)". The bars from left to right:
- < 5: 2.5
- 5–10: 1.8
- 10–15: 2.8
- 15–20: 4.3
- 20–30: 5.6
- 30–40: 6.0
- 40–50: 8.9
- > 50: 10.4

1) Do we know who collected this information? _____NO_____

2) Do we know how the information was collected? _____NO_____

3) Do we know when the information was collected? _____NO_____

4) Is it important to know how the information was obtained for this graph? Why or why not?

YES.

MUST CONSIDER BIAS AND RANDOMNESS TO GAGE ACCURACY OF RESULTS TO APPLY TO A WHOLE POPULATION

5) How much money does someone who makes $25,000 spend on transportation? _____$5,600.00____

6) What is the minimum amount spent on transportation? _____$1,800.00_____

7) What is the maximum amount spent on transportation? _____$10,400.00___

8) What is the minimum amount of household income? _____Less than $5,000.00_____

9)Does money spent on transportation increase with every increase in income? If not, at what income levels does it not increase?

NO Between less than $5,000.00 and $5,000 - $10,000 it goes down from $2,500.00 to $1,800.00

10)Does it make sense that money spent on transportation generally increases with income level? Why or why not?

YES more income means more money available for "extras" such as cars, gas, planes, trains.

Lower income means that more household income must be used for basic necessities such as housing, food, clothing.

11)Now, calculate the amount spent on transportation as a percent of total income:

Percent is: part/whole   2,500 spent on transportation/household income of 5,000 = 2500/5000 = .5 = 50%

< 5 _____50%____

5 – 10 _____18%_____ 1800/10000

10-15 _____18.7%_____   2800/15000

20-30 _____18.7%_____   5600/30000

30-40 _____15%_____ 6000/40000

40-50 ___17.8%_____   8900/50000

>50 _____17.3%_____ example: 10400/60000

12) Does your conclusion about the amount of money spent on transportation change based on these figures? If so, how? Which category spends the most percent of their income on transportation?

> Yes. Those earning less than $5000.00 per year spend half (50%) of their income on transportation which leaves much less money available for other purchases and luxuries.

> Those in the $30000.00-$40000.00 range spend the least percent of their income on transportation at 15%

> Those with the most income spend the second smallest percent of income on transportation at 17.3%

> Median household income in Fairfax County, VA is currently: $122,227 (per DATAUSA 6/13/2020).

> 10400/122227 = 8.5%

My conclusion changes from: Those who make the most income spend the most on transportation, to: The higher the income, smaller percent of income spent on transportation leaving more money available for other types of spending.

13) What is a problem with the way the groupings are set up? For example, which group would you put someone with an income of exactly 40,000

The number appears in two separate categories which results in confusing results
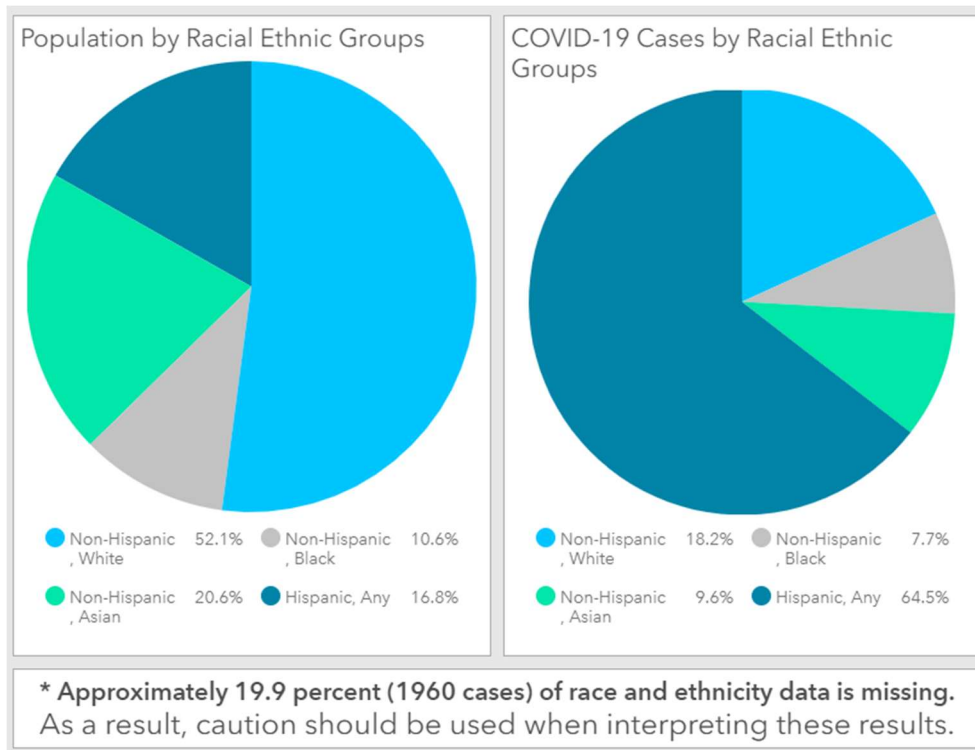
14)How could you change the groupings so that an income could not be placed in two categories?

Categories should end at .99   for example: 0 – 4.99.   5 – 9.99 10-14.99 etc

15) What else do you notice about the graph groupings that is inconsistent and could cause a misrepresentation of the data?

Category groupings are not consistent. The first four categories increase by $5000.00 each, but after the 20000-30000 category the increase is by $10000.00. This can cause a misrepresentation of the true data.

Fairfax County Health Department Cases by Racial Ethnic Group

Population by Racial Ethnic Groups

COVID-19 Cases by Racial Ethnic Groups

Non-Hispanic, White	52.1%	Non-Hispanic, Black	10.6%
Non-Hispanic, Asian	20.6%	Hispanic, Any	16.8%

Non-Hispanic, White	18.2%	Non-Hispanic, Black	7.7%
Non-Hispanic, Asian	9.6%	Hispanic, Any	64.5%

* Approximately 19.9 percent (1960 cases) of race and ethnicity data is missing. As a result, caution should be used when interpreting these results.

Note: Data is from May 15th

7) Which ethnic group has the highest population? _____Non Hispanic, white_____
8) Which ethnic group has the highest number of Covid 19 cases? _____Hispanic, any_____
9) What are some possible reasons that the group with the highest population does not have the highest number of Covid-19 cases?
   Lower income levels may lead to less opportunity for medical care and participation in social distancing and mask wearing. If more essential workers are from this category, they may have higher exposure rates.
10) How could county health experts use this information to provide Covid-19 support services?
   Health experts and government resources (money) should be allocated to the Hispanic population to fight Covid-19
11) Why should "caution be used when interpreting these results" ?
   The statement that "approximately 19.9% of data is missing. Which means there is almost 20% of data missing so the results may not be as they appear.
12) Do the graphs give a true representation of population data and Covid-19 cases?
   Maybe. We don't know as they admit 19.9% of the data is missing.