

AP Stats Summer assignment!
Braintree High School
Mr. Cooper: raymond.cooper@braintreeschools.org

The goal of this summer assignment is to get you in a statistical frame of mind and familiar with necessary vocabulary and fundamental skills entering this course. Part of the assignment this summer is within the design of this summer task; it's your job to research and develop your *own* thoughts and opinions about the following vocabulary and develop your own skills through *collaboration with peers, watching videos, and/or and by no means limited to reading references from the vast resource: the INTERNET!* We have to start THINKING and QUESTIONING and DISCUSSING and INTERPRETTING! Notice it's a math class and I didn't say solving... hmm.

Concern 1) "Mr. Cooper, doesn't this mean that everyone in a *math* class could have different answers and skills coming back from summer?"

Response 1) YES! Welcome to statistics! Without exactly THESE DIFFERENCES we wouldn't have a course!

Concern 2) Do you suggest we use multiple sources and not just the first answer we find?

Response 2) YES, and if you're ready for the rigor I hope to bring to AP Statistics you'll put in the time necessary to MASTER the concepts being developed this summer!

Calculator! We are going to make these machines do more this year than you ever knew was possible! I know the most about the TI-83 and 84 models. If you have or purchase a different model make sure you keep the manual, it will totally work, but we might have to do some discovery. Here's a list of approved machines: <https://apstudent.collegeboard.org/apcourse/ap-calculus-ab/calculator-policy>

Over the course of the summer you need to start building a solid foundation of statistics and expanding curiosity! This upcoming year it is mandatory for you to be engaged in the classroom, with your peers, questioning, discussing, collaborating, synthesizing information; without an active mind in our active classroom our educational experience will fall short. No pressure, but this starts now!

Google Classroom: If you're ready and sticking with it, then the assignment can be found on my website. Head to your google classroom account using your **@braintreeschools.org** account

Access code: **srgivk**

You should gain access to Mr. Cooper's class and we're in group: AP Statistics 2019-2020.

The documents can be found under folders tab within the group in the Summer Assignment file or just right in the stream! *If you do not have an account please contact me and we'll make sure you have an updated password* raymond.cooper@braintreeschools.org

- All parts of the assignment must be completed and handed in OR submitted by **September 9th**. You are able to submit paragraphs and vocab online and bring the analysis questions with you to school.
- *Your term grades will be made up of 90% assessments and 10% combined: work, homework, AP practice.*
- *This summer assignment will make up 20% of your work grade for 1st term but more importantly lay the foundation for the rest of the year.*
- *No portion of the assignment will be accepted late, and failure to complete will be an extreme set back moving forward with the curriculum.*
- *Most of this will be reviewed, but this is essentially chapter 1 and will be assessed within the first weeks of school.*
- *It's your job to research and develop your own thoughts and opinions about the following vocabulary and develop your own skills through *collaboration with peers, watching videos, and/or and by no means limited to reading references from the vast resource: the INTERNET!**

The assignment is made up of 5 parts for a total of 45 points

- **Vocab:** You should use the references available in the reference folder to discover understanding and mastery of the words from this list. **I'll collect the conceptual design project tied to the vocab assignment. 3points**
- **Chart:** The chart table is a tool to help you understand many of the data displays we'll utilize to kick off the year.
- **Paragraphs:** directions are in attachment. **3 points each**
- **Analysis:** These culminating problems have their own directions included in the problems and really get into the ideas we all should have a good grasp of coming back to school in September! A-C: **4 points each**, D is worth 5 points.
- **Investigative task:** This is a claim that you'll have to investigate, research, and provide evidence that supports or refutes said claim. This task has its own rubric worth: **10 points.**

This is most of the 1st chapter! There will be an AP style assessment within the first weeks from our first day together 1 and this material will be necessary for your success.

Vocabulary! These are some of the important words that we'll be making reference to all year long and it's *very* important to see how these words are similar, different, and relate to each other. Define and describe the following terms. Once you have an understanding create some sort of concept mapping/chart/web/diagram to help highlight the relationships between many of these ideas!

- | | | | | |
|--------------|------------------|----------------|--------------|--------------------|
| ● Data | ● Random | ● Continuous | ● Relative | ● Shape |
| ● Population | Variables | ● Distribution | frequency | ● Central Tendency |
| ● Sample | ● Quantitative | ● Frequency | ● Cumulative | ● Spread |
| ● Parameter | ● Qualitative or | ● Cumulative | Relative | ● Outlier |
| ● Statistics | Categorical | frequency | Frequency | ● Lurking variable |
| ● Census | ● Discrete | | ● Patterns | ● Skewed |

Displays	Sketch	Data Type	Pros	Cons	When/Why would you use it
Pie Chart					
Bar Graph					
Segmented Bar graph					
Stem and Leaf plot					
Dot Plot					
Box-and Whisker plot					
Histogram					
Ogive					
Scatterplot					

Paragraphs! These answers need to have evidence and support to justify correctness. You don't have to cite sources, for each answer, BUT you should for the summer as a whole if you are going beyond the sources I've provided.

I've set these up on the google classroom site as an "assignment" but you have all summer to submit or type up responses.

If this doesn't work out you can share the responses to school in one document:

Raymond.cooper@braintreeschools.org

Thank you!

- 1) What is statistics and how is it used? Be thoughtful in your reflection, use examples, and consider how statistics is relevant to YOU.
- 2) How are displaying distributions with graphs and describing distributions with numbers connected?
- 3) Why do you believe Statistics should be a class taken by high school students? What do you hope to gain from taking a class in Statistics?

Initial references:

- https://ted.com/talks/alan_smith_why_we_re_so_bad_at_statistics?utm_source=tedco_mshare&utm_medium=email&utm_campaign=tedsread
 - http://www.ted.com/talks/arthur_benjamin_s_formula_for_changing_math_education
 - <http://college.usatoday.com/2015/04/08/voices-statistics-might-be-the-most-important-class-you-take-in-college/>
- 4) Describe in detail your understanding of standard deviation. What is it? Why does it exist? Is there a difference between deviation and standard deviation? How is it used?
 - 5) Discuss the importance of having reliable data. Questions that might help: What kind of information can be counted on or believable? Can all statistics be trusted? What might influence data even when data can be trusted?

Analysis! For each of the following, provide a complete solution to the problem described (“Complete” solutions include explanations/work/labels/interpretations; not just answers). If you are unable to “solve/finish” a problem, you should write down thoughts and ideas describing how far you can work with the data set. 4 problems A – D! These can be submitted as a google doc or on your own paper written work. YOU WILL HAVE TO FIGURE OUT HOW TO DO THESE ON YOUR OWN, research!

A) In a study designed to determine the average death age reported for the population of a major US city, a statistician randomly selected 31 obituaries from the city’s largest newspaper. The sample consisted of 14 males and 17 females – their age of death is listed in the table below:

Female	75	77	82	75	74	89	87	44	91	94	60	68	57	84	79	75	74
Male	55	60	62	71	74	79	84	90	98	61	70	58	70	93			

- 1) Calculate the following statistics for the male and female data separately
 - a. Mean
 - b. Median
 - c. First quartile
 - d. 3rd Quartile
 - e. Range
 - 2) Create a back to back stem-and-leaf plot.
 - 3) Use the 1.5IQR outlier rule (might have to look it up) to justify if any data points are outliers.
 - 4) If there’s outliers, remove them, and recalculate the statistics for the data.
 - 5) Describe how the “outlier” affected each statistic.
 - 6) Construct both parallel box-and-whisker plots *and* back to back stem and leaf plots.
 - 7) Using your displays, describe the general shape and distribution of the data, the center, the spread.
 - 8) A statistician needs to interpret data in order to convey thoughts. Write a paragraph comparing and contrasting the data sets and provide insight into the similarities or differences for age of death for men and women in this particular city.
- B) Suppose a set of data consists of 33 whole number observations. Its five number summary is (min, Q1, median, Q3, max) = (16, 20, 22, 30, 46). Think box and whisker plot
- 1) What is the range of the data?
 - 2) How many observations are strictly less than 22?
 - 3) Is it possible that there is no observation equal to 22, explain?
 - 4) How many observations are strictly less than 20?
 - 5) Is it possible that there is no observation equal to 20, explain?
 - 6) Construct a modified box plot (investigate how it’s different, and show any calculations needed).

- C) According to a press release and data on car thefts in 2012, the 2009 Toyota Camry (a 3 year old car in that year) was the most stolen car in 2002. Further, according to the data the 2000 Camry, 2001 Camry, and 2008 Camry were also all among the top ten most stolen cars in the respective years. The press release claims that the most compelling reason for these cars being stolen is for parts but does not discuss any other variables.
- 1) The press release claims that these cars are stolen mostly for their parts. Can you think of any OTHER reasons that such a car would be a prime target for thieves (think about statistics – provide at least two statistical/mathematical reasons in your answer)?
 - 2) Based on the article, someone claims that because there is a high correlation between the age of the car and its theft (i.e., as the car gets older, it becomes more likely that it becomes a target for theft), it's obvious that car thieves prefer older cars. Does the high correlation mean that the age of the car causes the car to become a target? Do car thieves really prefer older cars or is there something else occurring – what are the possible confounding variables?
- D) Reference the excel sheet which presents data about individual states that relate to education. Study of a data set with many variables begins by examining each variable by itself:
- 1) Make a graphical display of the population of the states. Briefly describe the shape, center, and spread of the distribution of the population. Explain why the shape of the distribution is not surprising. Are there any states that you consider outliers?
 - 2) Make a stemplot of the distribution of the percent of high school seniors who take the SAT in the various states. Briefly describe the overall shape of the distribution. Find the midpoint of the data and make this value on the stemplot. Explain why describing the center is not very for a distribution with this shape.
 - 3) Make a graph to display the distribution of average teachers' salaries for the states. Is there a clear overall pattern? Are there any outliers or other notable deviations from the pattern?
 - 4) The "Percent no HS" column gives the percent of adult population in each state who did not graduate high school. We want to compare the percent of people without a high school education in the northeastern and the southern states. Take the northeastern states to be those in the MA – mid Atlantic and NE – New England regions. The southern states are those in the SA – south Atlantic and ESC – East South Central regions. Leave out the District of Columbia. ***tasks on the next page***
 - a. Make numerical summaries and graphs to compare the two distributions.
 - b. Write a paragraph about the comparisons between these two regions.

Investigative Task!

Statistics is all about using evidence to make predictions about some population of interest. This is our first attempt at this, by the end of the year you'll know exactly how to provide a well researched and statistically sound response to the following claim. This is our first attempt, so using the rubric and your inquisitive mind your job is to provide "statistical" evidence to support or refute the claim:

There is no significant difference in the average tuition between private and state undergraduate tuition in the United States.

Report (submitted electronically: slides, docs, something you like)	10 points
1. Introduction	2 pt
a. Statement of the question you are answering, why it's even a question?	
b. Population of interest who might be interested in your findings.	
c. Parameter of interest	
d. Background information, why this is of concern?	
2. Data Collection	2 pts
a. How was data collected?	
b. Why you trust the information?	
c. Is it representative? Why?	
3. Data Analysis	3 pts
a. Provide data in tabular form (or other)	
b. Graphical Displays	
c. Numerical summaries.	
d. Comparing and contrasting information.	
4. Conclusion	3 points
a. Does this evidence support or refute the initial claim?	
i. why? how?	
b. Are your findings significant?	
c. Ways to improve your project or findings. No matter how well you did there's always something that could be revised and improved for next time	

Here are a few resources that I've started collecting... do not limit yourself to these!

- <http://www.stattrek.com/>
 - I'd start with the tutorial.
- <http://www.apstatsguy.com/>
 - I'd say use this guy all year, but to kick things off the summer videos are helpful.
- <http://nces.ed.gov/programs/digest/>
- <http://onlinestatbook.com/2/index.html>
- <https://www.khanacademy.org/math/probability>

- We'll be starting with descriptive statistics, but most of the course is available to continue referencing.

Data on education in the United States								
State	Region	Population (1000)	Percent taking 2007	SAT Math 2007	SAT Critical Reading 2007	SAT Writing 2007	Percent no HS diploma	Teachers' pay (\$1000)
AL	ESC	4849	9	556	563	554	31.2	47.9
AK	PAC	736	48	517	519	491	13.4	65.46
AZ	MTN	6731	32	525	519	491	19.3	49.9
AR	WSC	2966	5	566	578	565	31.7	46.6
CA	PAC	38802	49	516	499	498	21.8	69.3
CO	MTN	5355	24	565	560	549	14.6	49.8
CT	NE	3596	84	512	510	511	18.8	69.3
DE	SA	935	72	496	497	486	21.3	59.7
DC	SA	659	78	462	478	471	24.9	53.5
FL	SA	19893	65	496	497	479	23.1	46.6
GA	SA	10097	69	495	494	483	27.2	52.9
HI	PAC	1419	61	506	484	473	17.1	54.3
ID	MTN	1634	19	539	541	519	17.3	49.7
IL	ENC	12880	8	611	594	588	21.8	59.1
IN	ENC	6596	62	507	497	479	22.4	50.1
IA	WNC	3107	4	613	608	586	17.9	50.9
KS	WNC	2904	8	590	583	569	16.7	47.4
KY	ESC	4413	10	565	567	553	33.4	50.2
LA	WSC	4649	7	567	569	563	28.4	51.3
ME	NE	1330	100	465	466	457	19.2	48.4
MD	SA	5976	70	502	500	496	19.6	64.2
MA	NE	6745	89	522	513	511	18	72.3
MI	ENC	9909	9	579	568	553	21.3	61.5
MN	WNC	5457	9	603	596	577	15.6	56.2
MS	ESC	2994	4	549	568	560	33.7	41.8
MO	WNC	6063	6	594	594	587	24.1	47.5
MT	MTN	1023	28	543	538	522	17.5	48.8
NE	WNC	1881	6	585	579	562	16.8	49
NV	MTN	2839	41	506	500	480	19.3	56
NH	NE	1326	83	521	521	512	15.8	55.6
NJ	MA	8938	82	510	495	494	21.3	68.8
NM	MTN	2085	12	546	555	540	22.9	45.4
NY	MA	19746	89	505	491	482	23.6	75.2
NC	SA	9943	71	509	495	482	28.7	45.7

State	Region	Population (1000)	Percent taking 2007	SAT Math 2007	SAT Critical Reading 2007	SAT Writing 2007	Percent no HS diploma	Teachers' pay (\$1000)
ND	WNC	739	4	596	584	562	22.1	47.3
OH	ENC	11594	27	542	536	522	21.9	56.3
OK	WSC	3878	6	571	578	559	23.4	44.3
OR	PAC	3970	54	526	522	502	16.4	57.6
PA	MA	12787	75	499	493	482	23.2	63
RI	NE	1055	68	498	496	492	26.1	63.4
SC	SA	4832	62	496	488	475	29.7	48.3
SD	WNC	853	3	602	589	567	22	39.1
TN	ESC	6549	13	569	574	568	31.3	47.5
TX	WSC	26956	52	507	492	482	25.9	48.8
UT	MTN	2942	6	556	558	544	13.3	49.4
VT	NE	626	67	518	516	508	17.8	52.5
VA	SA	8326	73	511	511	498	23.5	48.7
WA	PAC	7061	53	531	526	510	14.9	52.2
WV	SA	1850	22	528	520	512	33.1	45.4
WI	ENC	5757	6	598	587	575	19.4	53.8
WY	MTN	584	8	571	565	544	15.6	56.7