AP Statistics Summer Assignment

First, welcome to AP Statistics. This course will likely be unique among the math courses you've taken in its emphasis on your ability to **think, reason, explain,** and **support** rather than simply performing rudimentary computations. It is unlikely that you'll find yourself questioning "when would someone ever use this?" very often, if ever.

Second, to succeed in this course, you should be competent in basic algebra. However, as mentioned, advanced computational skill is **not** necessary in this course (though it's obviously good to have anyway). Throughout this course, you will familiarize yourself with a number of new concepts and explore some older, familiar ones in more depth than before.

Third, you'll need to have your own calculator and laptop and bring them (charged) to class **every day**. We will use the TI-Nspire calculator for almost all of the topics that we cover in this class, and we'll also use various applets and computer demonstrations throughout the year to help supplement what we learn in class. Although not necessarily an emphasis in the AP Statistics curriculum, Statistics largely goes hand-in-hand with various types of technology.

Finally, you will need to be an active participant in this course. This means you **must** be willing to work with your classmates and with me during the year (*and to have a good time while doing it*). If you are not interested in demonstrating a strong work ethic, never enjoy the mental challenge of a good question, and would rather be inactive during class time, then this is probably not the course for you.

Please obtain your textbook (*Stats: Modeling the World* 4th edition, by Bock, Velleman, and De Veaux) from the main office when you are able. If you have trouble procuring it, let me know - this will help you work through some parts of the packet. Collaboration/teamwork is okay, but plagiarism/copying is not. Don't be intimidated by the apparent length of the assignment. Part 2 could be a slog, but you'll get through it just fine. Part 1 is short and opinion-based, and Part 3 should be fairly quick (and possibly entertaining) after having done the work in Part 2. This will count for about 10% of your first quarter grade, plus the three chapters covered will be assessed with a test/quiz in the first few weeks of school.

I am SO tremendously excited to work with you next year, and I hope that you have a great summer! If at any point you are stuck on something in this assignment or just want to bounce an idea/thought off of me, feel free to email me at jfitzpatrick@ionaprep.org or post in Google Classroom. I check both of these constantly.

Due Dates

Please work through the parts in order.
I sequenced the assignment this way for a reason.

Part 0 - Join Google Classroom with the code **lgik6z2** or via invitation in email (I will email some time in June if you haven't joined with the code yet)

Part 1 is opinion-based. It is due July 31 at 11:59pm (before September 1) via Google Classroom.

Part 2 is reading/summarizing the first three chapters of the textbook and completing a few exercises. It is due the first day of school (scheduled as Friday, September 6) on Google Classroom.

Part 3 is critiquing poorly made graphs from popular media sources by applying principles you've learned. I think it should be somewhat entertaining (it's funny to me, anyway). It is due the second day of school (scheduled as Monday, September 9).

All parts will be collected via Google Classroom, so if you are drawing your dotplot or histogram by hand, make sure you put a photo of it into your work with Google Docs or something like that.

(No excuses unless cleared WELL in advance)

Part 1

"Statistics" is a word that gets thrown around a lot in many different contexts. However, it is not always clear what the intended meaning is *or* what the reader understands it to mean. So, I want to know what it means to each of you - right now, before having studied Statistics (or even having completed the rest of the summer assignment).

When you hear the word "Statistics", what do you think it means? What does it call to mind? Given what you think it means, how do you expect to see it applied in questions that you'd encounter on an exam, and how do you think it is utilized in "real life"? Please try to address all of these questions in your response. You can do a little research if you really want to, but I do want to get an idea of what you think of Stats *before* actually learning about it for real.

This is a very low-pressure, low-stakes part of the assignment. There is no wrong answer (aside from a lack of thought and effort). I just want you to start thinking about Statistics and how you might see it applied. **No need for more than a page**.

This is due on Google Classroom on **July 31**. There's no bonus for handing it in early, but I'm very interested to read these, so it would be ideal for me to not have all of them submitted at 11:59pm that night. Also, as it's admittedly a low-pressure part of the assignment that I *want* you to do before actually learning anything new, I'd advise you to knock this part out sooner than later, both for my sake (getting to read them) and yours (getting an early start on your work).

Part 2

This summer, you'll be responsible for reading and developing some comfort with the material in the first three chapters of the textbook (there are 26 in total). Before starting to read each chapter, I recommend looking at the outlines (starting on the next page) for each of the three to see what information from the chapters is important/should be your focus. Fill in the outlines (or write in your notebook) for each chapter as you go.

The chapters you'll read are:

Ch. 1 – Stats Starts Here, which provides an overview of some basic terms and descriptions of how we'll approach problems and situations in Statistics

Ch. 2 – Displaying and Describing Categorical Data, which provides an overview for how we deal with categorical (non-numerical) data

Ch. 3 – Displaying and Summarizing Quantitative Data, which provides an overview for how we deal with quantitative (numerical) data

Our textbook is generally regarded as quite readable among AP textbooks. That said, some of the information (especially Ch. 3 – it's quite a long chapter – it used to be split into two chapters in older editions of the textbook) may be dense, so I recommend spreading out the time that you work on this so that you're not experiencing a huge time crunch at the end of the summer (especially if you're taking other AP's that also have assignments). Once you've completed the outlines, there are a few questions for you to complete based on that material.

I'll re-emphasize here that **if you have questions, I am always available by email** or Google Classroom and would be happy to assist you however I can.

Chapter 1: Stats Starts Here

Key Vocabulary:

- Statistics
- data, datum
- variation
- individual
- respondent
- subject
- participant
- experimental unit
- observation
- variable
- categorical
- quantitative

1. Name three things you learned about *Statistics* in Chapter 1.

   -
   -
   -

2. The authors claim that this book is very different from a typical mathematics textbook. Would you agree or disagree, based on what you read in Chapter 1? Explain.

3. According to the authors, what are the "three simple steps to doing *Statistics* right?"

4. What do the authors refer to as the “W’s of data?”

5. Why must data be in context (the W’s)?

6. Explain the difference between a *categorical variable* and a *quantitative variable*. Give an example of each.

Chapter 2: Displaying and Describing Categorical Data

Key Vocabulary:

- frequency table
- relative frequency table
- distribution
- bar chart
- pie chart
- contingency table
- marginal distribution
- conditional distribution
- independent
- segmented bar chart
- Simpson's Paradox

1. According to the authors, what are the three rules of data analysis?

2. Explain the difference between a frequency table and a relative frequency table.

3. When is it appropriate to use a bar chart?

4. When is it appropriate to use a pie chart?

5. When is it appropriate to use a contingency table?

6. What does a marginal distribution show?

7. When is it appropriate to look at a conditional distribution?

8. What does it mean for two variables to be independent?

9. How does a segmented bar chart compare to a pie chart?

10. Explain what is meant by Simpson's Paradox.

Chapter 3: Displaying and Summarizing Quantitative Data

Key Vocabulary:

- distribution
- histogram
- relative frequency histogram
- stem-and-leaf display
- dotplot
- shape
- center
- spread
- mode
- unimodal
- bimodal
- multimodal
- uniform
- symmetric
- tail center
- spread
- midrange
- median
- range
- quartile
- interquartile range
- percentile
- five-number summary
- boxplot
- mean
- standard deviation
- variance
- skewed
- outliers
- gaps
- time plot
- re-expressing data

Calculator Skills:

- display a histogram
- Boxplot
- modified boxplot
- 1-Var Stats

1. What is meant by a *distribution*?

2. Explain the difference between a *histogram* and a *relative frequency histogram*.

3. In what ways are *histograms* similar to *stem-and-leaf displays*?

4. Name some advantages and disadvantages of *stem-and-leaf displays.*

5. When is it more appropriate to use a *histogram* rather than a *stem-and-leaf display*?

6. Name some advantages and disadvantages of *dotplots.*

7. When describing a *distribution*, what three things should you always mention?

8. What should you look for when describing the *shape* of a *distribution*?

9. In general, what is meant by the *center* of a *distribution*?

10. In general, what is meant by the *spread* of a *distribution*?

11. When is it appropriate to use a *time plot* to display quantitative data?

12. Explain the difference between *range* and *interquartile range*. Why is the *interquartile range* often a better measure of the spread of a distribution?

13. What are some advantages of *boxplots*?

14. What are some disadvantages of *boxplots*?

15. When is it more appropriate to use the *mean* as a measure of center rather than the *median*? Why?

16. When is it more appropriate to use the *median* as a measure of center rather than the *mean*? Why?

17. When do the *mean* and *median* have the same value?

18. Describe the relationship between *variance* and *standard deviation*.

Ch. 1 example- The 2.5-mile Indianapolis Motor Speedway has been the home to a race on Memorial Day weekend nearly every year since 1911. Even during the first race, there were controversies. Ralph Mulford was given the checkered flag first but took three extra laps just to make sure he'd completed 500 miles. When he finished, another driver, Ray Harroun, was being presented with the winner's trophy, and Mulford's protests were ignored. Harroun averaged 74.6 mph for the 500 miles. In 2013, the winner, Tony Kanaan, averaged 187.433 mph.

Here are the data for the first five races and five recent Indianapolis 500 races.

Year	Winner	Time	Average Speed (mph)
1911	Ray Harroun	6:42:08.039	74.602
1912	Joe Dawson	6:21:06.144	78.719
1913	Jules Goux	6:35:05.108	75.933
1914	René Thomas	6:03:45.060	82.474
1915	Ralph DePalma	5:33:55.619	89.840
. . .			
2009	Hélio Castroneves	3:19:34.6427	150.318
2010	Dario Franchitti	3:05:37.0131	161.623
2011	Dan Wheldon	2:56:11.7267	170.265
2012	Dario Franchitti	2:58:51	167.734
2013	Tony Kanaan	2:40:03.4181	187.433

a. Identify the W's
b. Name the variables
c. Specify for each variable whether its use indicates that it should be treated as categorical or quantitative
d. For any quantitative variable, identify the units in which it was measured (or note that they were not provided).

Ch. 2 examples:

1. In 1997, there were 92,353 deaths from accidents in the United States. Among these were 42,340 deaths from motor vehicle accidents, 11,858 from falls, 10,163 from poisoning, 4,051 from drowning, and 3,601 from fires.

a. Find the percent of accidental deaths from each of these causes, rounded to the nearest percent. What percent of accidental deaths were due to other causes?

b. The most common way of displaying this type of data is a bar chart. Would it also be correct to use a pie chart to display these data? Explain.

2. Students in an Intro Stats course were asked to describe their politics as liberal, moderate, or conservative. Here are the results:

		Politics			
		L	M	C	Total
Sex	Female	35	36	6	77
	Male	50	44	21	115
	Total	85	80	27	192

a. What percent of the class is male?

b. What percent of the class considers themselves to be conservative?

c. What percent of the males in the class consider themselves to be conservative?

d. What percent of all students in the class are males who consider themselves to be conservative?

e. Find the conditional distribution (percentages) of political views amongst both the males and the females.

f. Do the variables *politics* and *sex* appear to be independent? Explain.

Ch. 3 example: In October 2011, 16 gas stations in eastern Wisconsin, posted these prices for a gallon of regular gasoline:

3.43	3.46	3.43	3.59	3.65	3.63	3.62	3.65
3.66	3.31	3.35	3.42	3.41	3.46	3.47	3.48

a. Make a dotplot of these data. Make a tick mark on the axis for every 5 cents.

b. Make a histogram of these data. Use "bins" (widths of the bars of 5 cents).
Remember that histograms should NOT have space between the bars.
Hint: the histogram's bars' heights should be similar in appearance to the dotplot

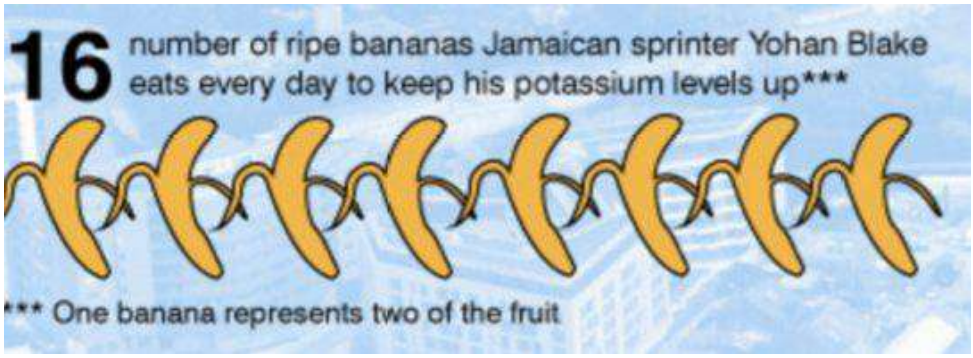
c. Describe the shape, center, and spread of this distribution.

d. What unusual feature do you see in this distribution?

Part 3

Because of how useful statistics are and how they are found all throughout the world – media, science, etc. – we encounter many statistical displays out "in the wild" (in those other contexts – outside of a math/Stats class). However, not everyone is an expert in statistics and, in particular, not an expert in *displaying* statistics. Whether they are convoluted, mislabeled, misleading, or simply the wrong way to display a certain type of data, many displays are bad and can improved with even a small amount of statistical knowledge.

You learned about a number of different statistical displays in Chapters 2 and 3. More importantly, you learned about the principles that make a statistical display useful. We're now going to put this practical knowledge to use. On the next couple of pages, you'll find a number of statistical displays that miss the mark – something about them is misleading, obscures the meaning of the data, or is just wrong. For each one, write a few sentences describing what is wrong with that particular graph/statistical display and how it might be improved (this can include changing the type of display/graph completely, if another type would be better). Some errors may be obvious and some may be a bit more subtle.

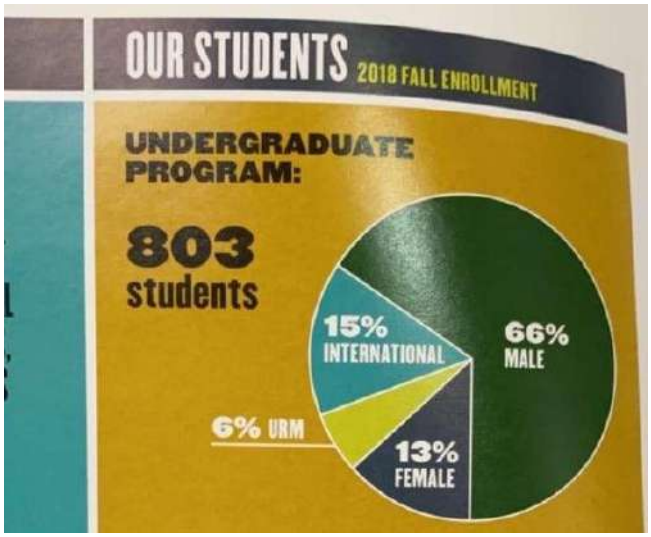
1. 

2. 

3. Line graph showing the changes of copper prices during the Covid-19 epidemic in Chile

4. Pie chart showing the results of a “favorite color” survey

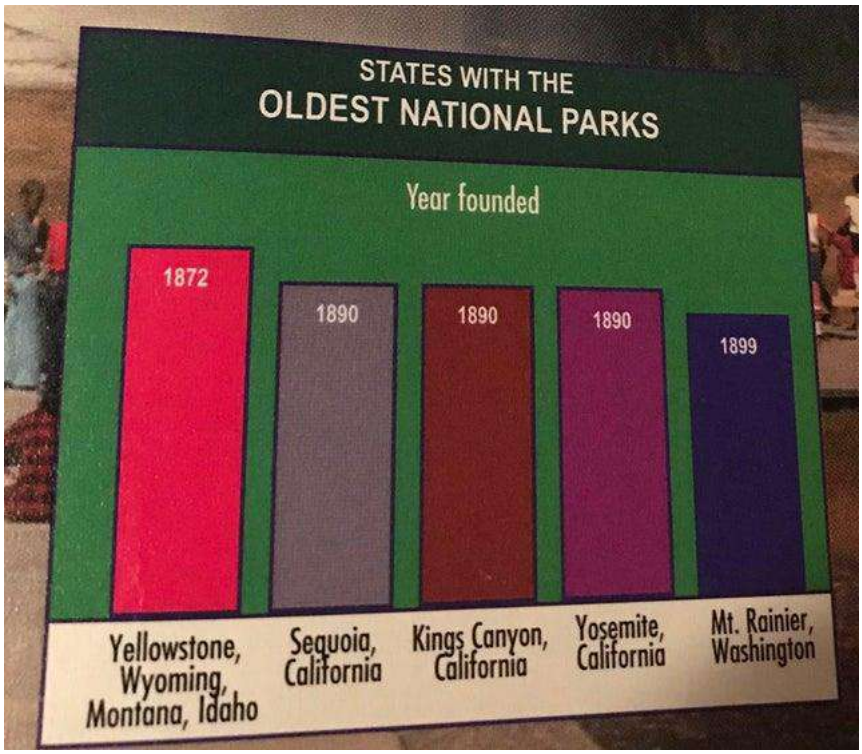
5. Pie chart of demographics in an undergraduate program – “international” means a student from outside the U.S. and “URM” means “underrepresented minority”.



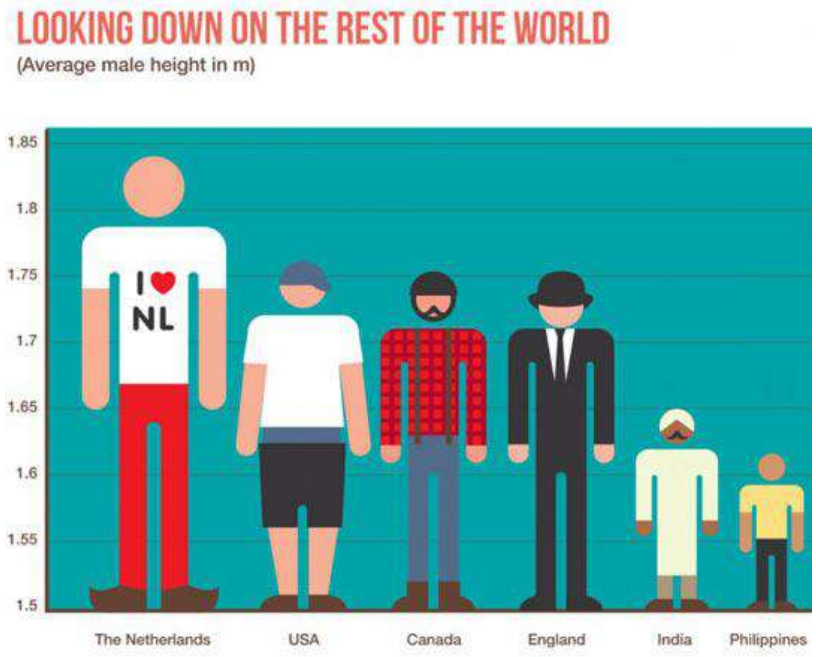
6. Display of polling results for political parties in a British Parliamentary election

7. Display of election results from Venezuela

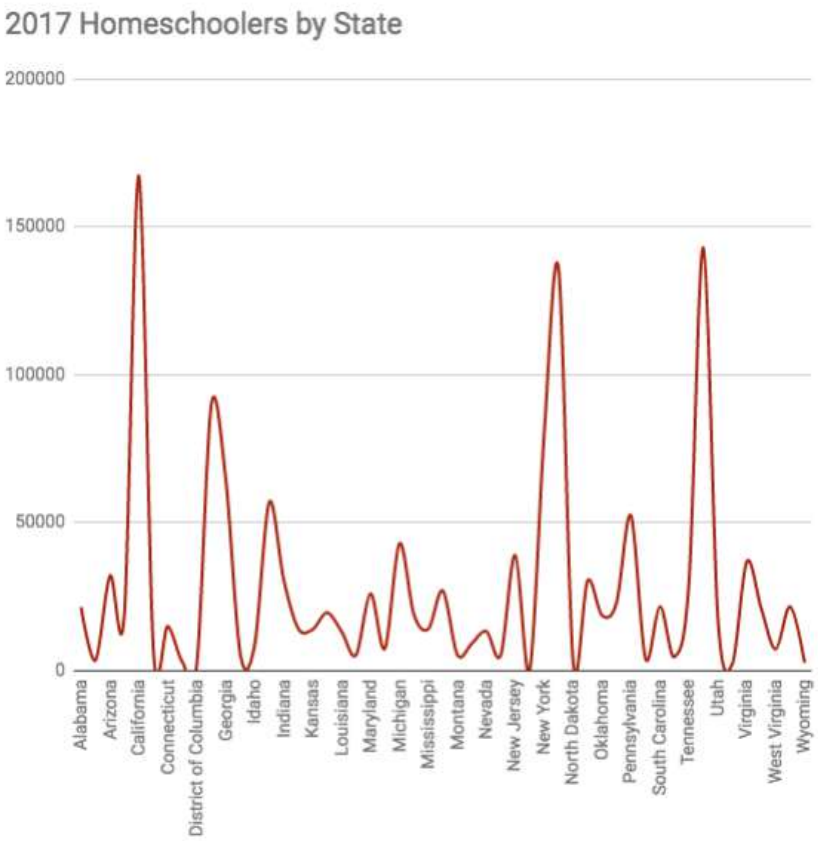
8.



9.


LOOKING DOWN ON THE REST OF THE WORLD
(Average male height in m)
1.85
1.8
1.75
1.7
1.65
1.6
1.55
1.5
I ♥ NL
The Netherlands
USA
Canada
England
India
Philippines

10.


2017 Homeschoolers by State
200000
150000
100000
50000
0
Alabama
Arizona
California
Connecticut
District of Columbia
Georgia
Idaho
Indiana
Kansas
Louisiana
Maryland
Michigan
Mississippi
Montana
Nevada
New Jersey
New York
North Dakota
Oklahoma
Pennsylvania
South Carolina
Tennessee
Utah
Virginia
West Virginia
Wyoming