Guidelines for HW Submissions

You should follow these guidelines for your homework submission:

- As a general rule, explanation/discussion is expected for **ALL** questions, unless you are explicitly told no explanation is needed. Many questions assigned (some are from the textbook, and some are presented to you inline on the assignment page) will be worded in such a way that it may seem that no explanation is required, but the graders WILL look for one. Of course, graders are looking for the correct answer, but the rationale is also very important. If you create any graphics, describe the trends you see, or, if you have multiple graphs for a given question, compare/contrast them! If you arrived at an answer and found it was easy, then by all means make the explanation simple. If you find yourself doing similar steps for similar problems, it is perfectly valid to state, 'as shown previously'.

- When showing work, you must make **clear** what operation you've tried to perform. For instance, the difference between
$$\binom{5}{3} \text{ and } \binom{5}{3}$$
is critical. The left refers to the fraction 'five-thirds', the right refers to the combination calculation 'five-choose-three'. It can be hard for the graders to decide what is the intention of the student.

- **DO NOT SUBMIT PHOTOS OF HANDWRITTEN WORK.** The goal of the HW is not to make you spend hours formatting. If you are having a hard time making some notation work, please post on Piazza "how do I make an integral symbol in Latex," and someone should be able to help. Or, if you cannot achieve the desired effect, just describe the computation you tried to do: "The integral of x^2 with limits of integration -2 and 2 is..."

- The data we provide is often not formatted conveniently. This is a very real issue in the world and we leave it up to you how best to deal with it. You can try to find a formatted version of the dataset (make sure that the data is what you think it is), try to run it through some sort of online formatter, or manually format it yourself by carefully replacing spaces with commas. **You should not share your version of the formatted data.**

- When submitting assignments, you should place each question starting on a new page. (So if the assignment outline has questions 1,2,3,4,5, make sure to start Question 1 on a new page, Question 2 on a new page, etc.) The tool used for submission in this course, Gradescope, will prompt you to select which pages correspond to which questions upon submission. You are allowed to select multiple pages for a single question so it is perfectly ok for explanations/answers to spill onto subsequent pages. Everything is digital so no need to worry about wasting any paper!

Naturally this brings up the question of "Where do I do this homework?" Our recommendation is Overleaf, the online Latex Editor. Latex is incredibly useful as it ensures proper formatting and consistency, and is widely used in academic/technical publications. At the end of the day, Latex is just a tool for fancy formatting and page consistency. And much like learning anything new, the more you practice, work, or play with it, the easier working with Latex will feel in the future.

Of course, similar effects can be achieved with both Google Docs and Microsoft Word. Do feel free to use those tools. These word editors usually have some method to insert 'Math equations', and we recommend you use these features for both easing the task of the graders but also helping you get used to expressing your thoughts in mathematical notation.

The rest of this document is an example submission for what homework may look like. Consider the following as the questions for the homework: (Note: These questions are derived from questions and analysis present in the textbook.)

1. Using the following **dataset** (see description of the dataset's columns **here**), answer the following questions:

   A Were more boys or girls surveyed?

   B What is the average age of 4th grades, 5th graders, and 6th graders?

   C What trait do 4th graders think makes someone popular? What trait do 5th graders think makes someone popular? 6th graders? Construct bar charts for each grade illustrating what they think is the most important for being popular. (For these charts, you should count based only off of how many times each trait was rated most important.)

   D For boys and girls of all grades, create charts to show what they think makes someone popular.

2. Suppose there are two car factories, JrGrafo and SrGrafo. These factories occasionally will create broken cars that appear externally to be fully functional. JrGrafo creates 20 cars, 4 of which are secretly broken. SrGrafo creates 120 cars, 10 of which are secretly broken. All 140 cars are thoroughly mixed up and put in the GrafoFamily car lot. You decide to randomly purchase a car from this lot.

   A What is the probability that the car you buy is broken?

   B What is the probability that the car you buy is from the SrGrafo factory?

   C Suppose you have found out that the car you purchased is broken. Now what is the probability that the car is from the SrGrafo factory?

   D You decided to return the broken car to the GrafoFamily, and the GrafoFamily decides that there is nothing wrong with the car and returns the broken car to somewhere random in the lot. (That is, all 140 cars are in the lot again). You tell your friend John Doe to buy 6 cars randomly from the lot. What is the probability that all 6 cars John purchases are broken?

3. You have a sample of 50 mice. The mean weight of the mice in this sample is 34.7 grams. The standard error in the weight of mice is 0.044 grams. Using a significance level of 5%, assess the following claim "the mean weight of all mice is 35 grams."

Question 1

A **More girls were surveyed.**

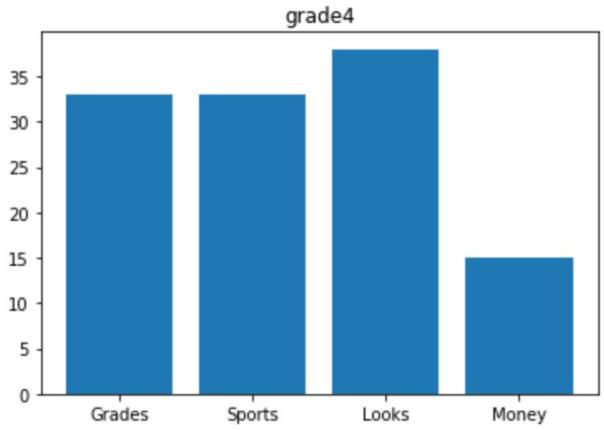
There were 227 boys surveyed and 251 girls surveyed.

B **The average age of 4th graders is 9.168, of 5th graders is 10.347, of 6th graders is 11.311.**

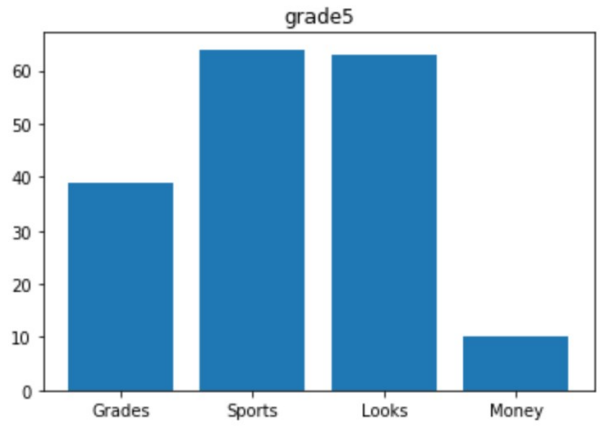
[Side note: you can consider taking an average of numbers the same as adding numbers. You do not need to explain how an average is done, but rather over what set of numbers it's done.]

C Plots for what each grade level found most important to being popular. The numbers above each graph represent the heights of each bar.

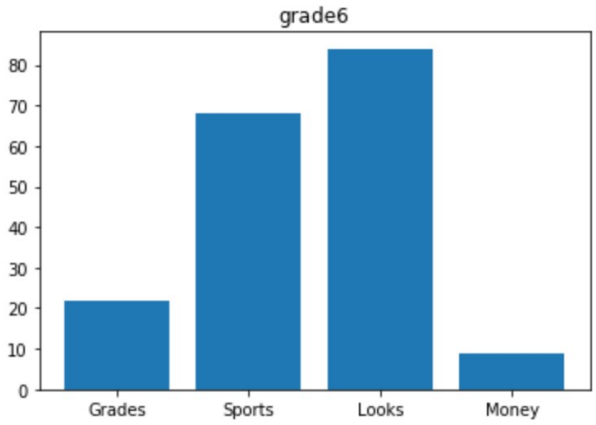
```
[33, 33, 38, 15]
```




```
[39, 64, 63, 10]
```



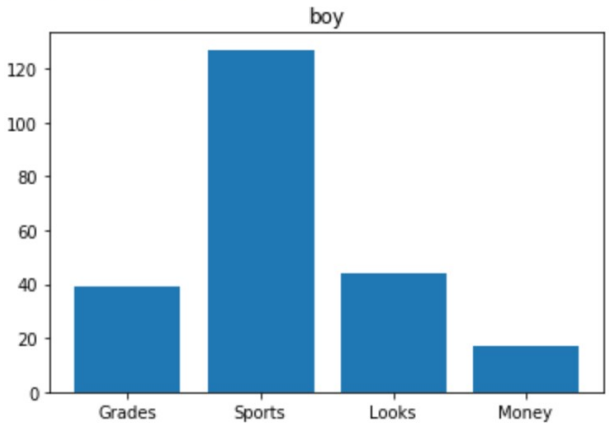

```
[22, 68, 84, 9]
```



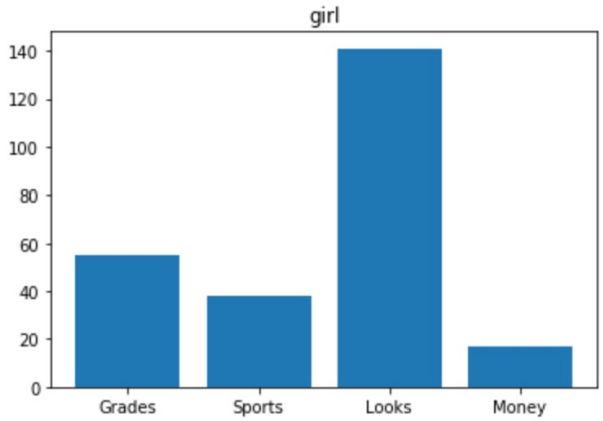
It seems that money is always the least important across all 3 grade levels, and as kids grow older, the relative importance of grades decreases. Sports and Looks are consistently important.

D Plots of the values for what boys and girls found most important to being popular.

```
[39, 127, 44, 17]
```




```
[55, 38, 141, 17]
```



It seems that Boys typically find that sports makes someone popular, whereas girls typically find looks makes someone popular. Both groups find money the least important.

Question 2

A There are 14 broken cars and 140 total cars. The probability that any random car is broken is $\boxed{\frac{14}{140}}$.

B There are 120 cars from the SrGrafo factory so the probability that any random car is a SrGrafo car is $\boxed{\frac{120}{140}}$.

C If the car is broken, then the size of the denominator must be 14 since there are a total of 14 cars. The SrGrafo factory produced 10 of these cars so the probability is $\boxed{\frac{10}{14}}$

D There are initially 14 broken cars and 140 total cars. The probability here is $\frac{14}{140}$. Next there are now 13 broken cars and 139 total cars. The probability here is $\frac{13}{139}$. Each successive car will reduce the number of broken cars and total cars by 1. Our answer is the product of the first 6 of these terms:

$$\boxed{\frac{14}{140} \cdot \frac{13}{139} \cdot \frac{12}{138} \cdot \frac{11}{137} \cdot \frac{10}{136} \cdot \frac{9}{135}}$$

Question 3

Let σ represent the stderr. We know that:

$$\sigma = 0.044$$

Next we must compute the test statistic,

$$\text{Let } g = \frac{\text{sample mean} - \text{hypothesized mean}}{\sigma} = \frac{34.7 - 35}{.044} \approx -6.8181$$

We evaluate the two-sided p-value. Since $N > 30$, we can use the normal distribution approximation:

$$p = 1 - \frac{1}{\sqrt{2\pi}} \int_{-|g|}^{|g|} exp\left(\frac{-x^2}{2}\right) dx \approx 0$$

We can confidently say that the hypothesis is unlikely to be true given our sample and our desired significance level. We reject the claim.