Introduction:

For this practical assignment, I am interested in the question "Between college male students and college female students, who spends more time on social media every day?" I came up with this idea because I often noticed many people like to look at their phones in many places. For example, while waiting at the light rail stations, sitting in the classroom before class starts and having coffee at coffee shops, I can see people looking at their phones from time to time. I let $\mu(m)$= the mean amount of time that college male students spend on social media every day. $\mu(f)$= the mean amount of time that college female students spend on social media every day. In this case, I hypothesize:

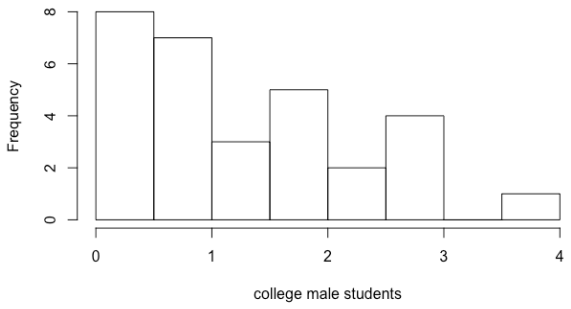
Null Hypothesis: $\mu(f)= \mu(m)$

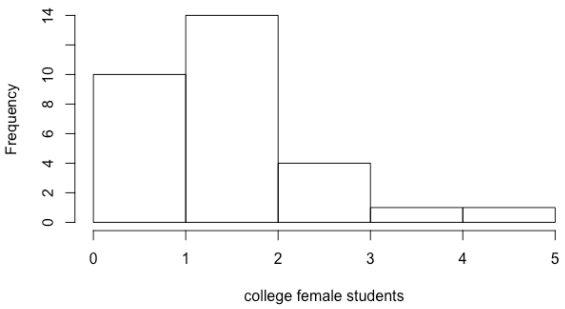
Alterative Hypothesis: $\mu(f)>\mu(m)$

In my observation about most posts on all the social media apps I have currently, I noted that I see more posts that are from female. Therefore, I assume that female college students generally spend more time on social media. When I tried to gather my data, I asked many people I know in different ways. I sometimes asked in person, and mostly I asked my friends directly online. Since I only gathered data from people around me, it is not a random sample. I asked them "How much time do you think you spend on social media every day generally?" When I asked them, I asked them to measure as accurate as possible before they answered the question. At the end, I have data from just 30 college male students and 30 college female students.

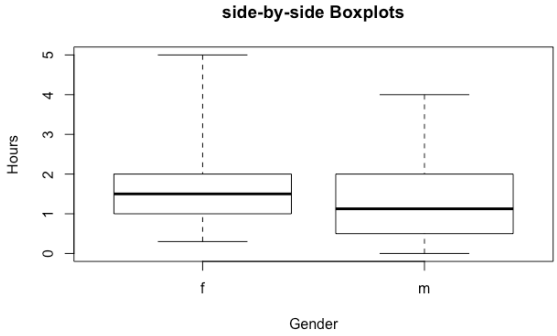
Graphs and Descriptive Statistics:

I have two independent samples; therefore, I will be conducting a two-sample t test. And the following is the histograms for college male students and college female students. We note that both of them are not bell shape, and the histogram for college female students is approximately right-skewed.





Also, this boxplot I made has a comparison between male and female students. From the boxplot, we know that the median of time spent on social media is higher among college female students. The range is larger among college female students. However, the IQR is larger among college male students.



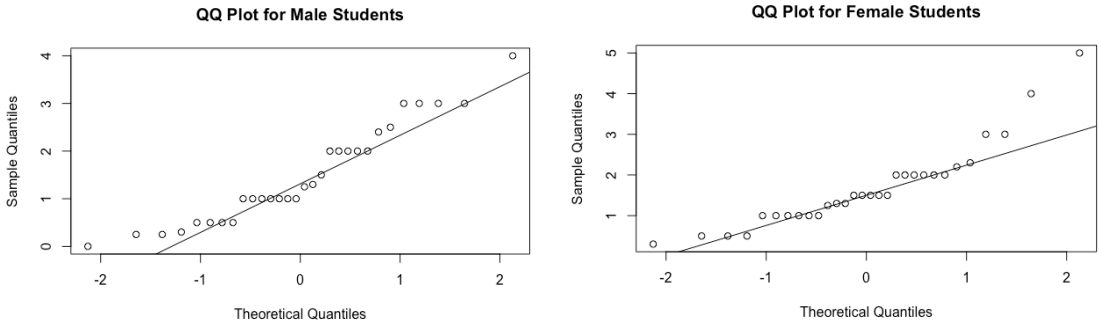
By using R and create the following table, we know that the mean amount of time spent on social media is higher among college female students in this sample. The standard deviation is slightly higher among college male students, which means variations within these two groups are almost the same.

	Min	IQR	Median	Mean	SD	Max
Hours(Male)	0.000	1.375	1.125	1.492	1.033	4.000
Hours(Female)	0.300	1.000	1.500	1.688	1.028	5.000

Verifying Necessary Data Conditions:

The assumption for a two-sample t test is: (1) There are two independent random samples. (2) Approximately normal distribution for each group.
The two samples I collect are independent. However, these two samples are not random since I only collect data from the people around me. Therefore, in my further analysis, I need to keep in mind that the result I will get cannot represent the whole population.

In order to verify whether the data is normally distributed, I can use QQ plot:



By looking at these two QQ plots for male students and female students. The distribution of these two samples are approximately normal but not perfectly. From what we learned from class, the two-sample t test is robust to non-normality for large sample size. Since each of my samples has 30 individuals. The analysis should not be affected by non-normality.

Conducting a Hypothesis Test:

In order to conduct a two-sample t test, I used t.test function R to conduct the test with the confident level 95%. The test statistic is -0.73895, p-value is 0.4629. This p-value indicates that if the null hypothesis is true, then the probability of obtaining a test statistic at least as extreme as the observed under the null hypothesis is 0.4629.

Therefore, in the sample, since p-value is greater than significant level 0.05, I failed to reject the null hypothesis. I don't have enough evidence to prove that the mean amount of time that female college students and male college students spend on social media is different. Therefore, the results are not statistically significant. Since there is not a big difference between the mean of college male students and the mean of college female students, the results are not practically significant.

Since I failed to reject the null hypothesis, it is possible for me to make type II error. It is difficult to calculate the probability of making type II error under the alternative hypothesis. However, I know that the probability of type II error decreases as the probability of type I error increases.

By using R, the confidence interval I get is (-0.7294097, 0.3360764), which means that the difference of the mean amount of time spent on social media between college female students and college male students is between -0.7294097 and 0.3360764. Since 0 is contained in this interval, I can agree with the conducted hypothesis test. The possible value between these two means lies within this interval.

Conclusion and Summary:

I conducted this project because I wondered whether gender can determine the amount of time spent on social media. I collected those data by simply asking people around me, and I concluded that there is not such as big difference between college female students and college male students.

Since the sample data I collected is not random, my results cannot represent the general population because I only gather data from people around me. In order to improve this project, I will have to conduct a survey for a greater population on campus. Also, it will be better to collect data from other schools as well.