Name: Melissa Selva

UmichID: 52046884

What dataset are you working with: comic\_characters

List 3 questions that you can ask with your dataset.

Q1: Is a character’s gender related to the number of appearances in the comics?

Q2: Is the mean number of appearances of a comic character greater than 2,000?

Q3:

List the associated null hypothesis for each question:

Q1: The number of appearances of a male comic character is not different than the number of appearances of a female comic character.

Q2:  
Q3:

What statistical test(s) will you use to answer each of the questions:

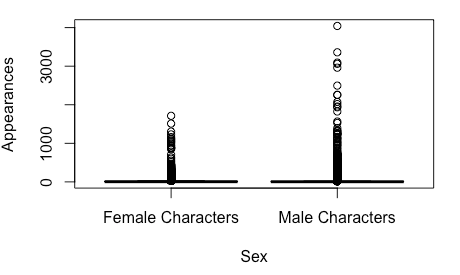
Q1: Two sample t-test.

Q2:

Q3:

Make a visual plot showing the relationship that you will analyze statistically (e.g. boxplot for t-test or ANOVA; scatterplot for regression; table for chi-square).

Q1:



Q2:

Q3:

Do your data meet the assumptions required for the statistical test you want to run? Please state the assumptions you examined and whether or not your data meet those assumptions:

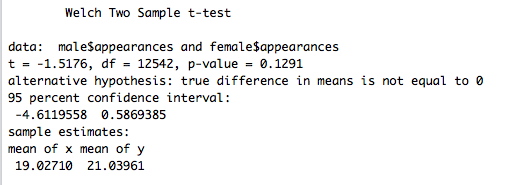
Q1: We do not know this for sure, but are assuming that our sample was randomly selected and observations are independent. Another key assumption is that values are nearly normal, but since our samples sizes are so large, we can assume normality and proceed with running the statistical test.

Q2:

Q3:

Run the statistical test! Put your results here:

Q1:



Q2:

Q3:

Interpret your results!

Q1: With a p-value that is not less than the typical alpha value of 0.05, we cannot reject the null hypothesis that “the number of appearances of a male comic character is not different than the number of appearances of a female comic character”. From these results, we might be able to conclude that the number of appearances does differ significantly between male and female comic characters. However, the samples are of very different sizes. The male group is over 16,000 values, while the female group is only 5,800 values. Therefore, even if there might be a difference in number of appearances between male and female comic characters, it could be masked in the vast difference in sample sizes.

Q2:

Q3: