Name: Lauren Edson

UmichID: 71631166

What dataset are you working with: comic\_characters

List 3 questions that you can ask with your dataset.

Q1: Is sex related to a character’s alignment?

Q2: Is sex related to the character’s number of appearances?

Q3: Is number of appearances related to if the character is alive or dead?

List the associated null hypothesis for each question:

Q1: The sex of the character and their alignment is independent.

Q2: The mean number of appearances is the same for every gender  
Q3: The mean number of appearances is the same for living and dead.

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

Q1: chi-square

Q2: one-way anova and Tukey

Q3: two-sample t-test

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:

Contingency table

align

sex Bad Characters Good Characters Neutral Characters

Agender Characters 20 10 13

Female Characters 1573 2490 836

Genderfluid Characters 0 1 1

Genderless Characters 11 6 3

Male Characters 7561 4809 1799

Transgender Characters 1 0 0

align

sex Reformed Criminals

Agender Characters 0

Female Characters 1

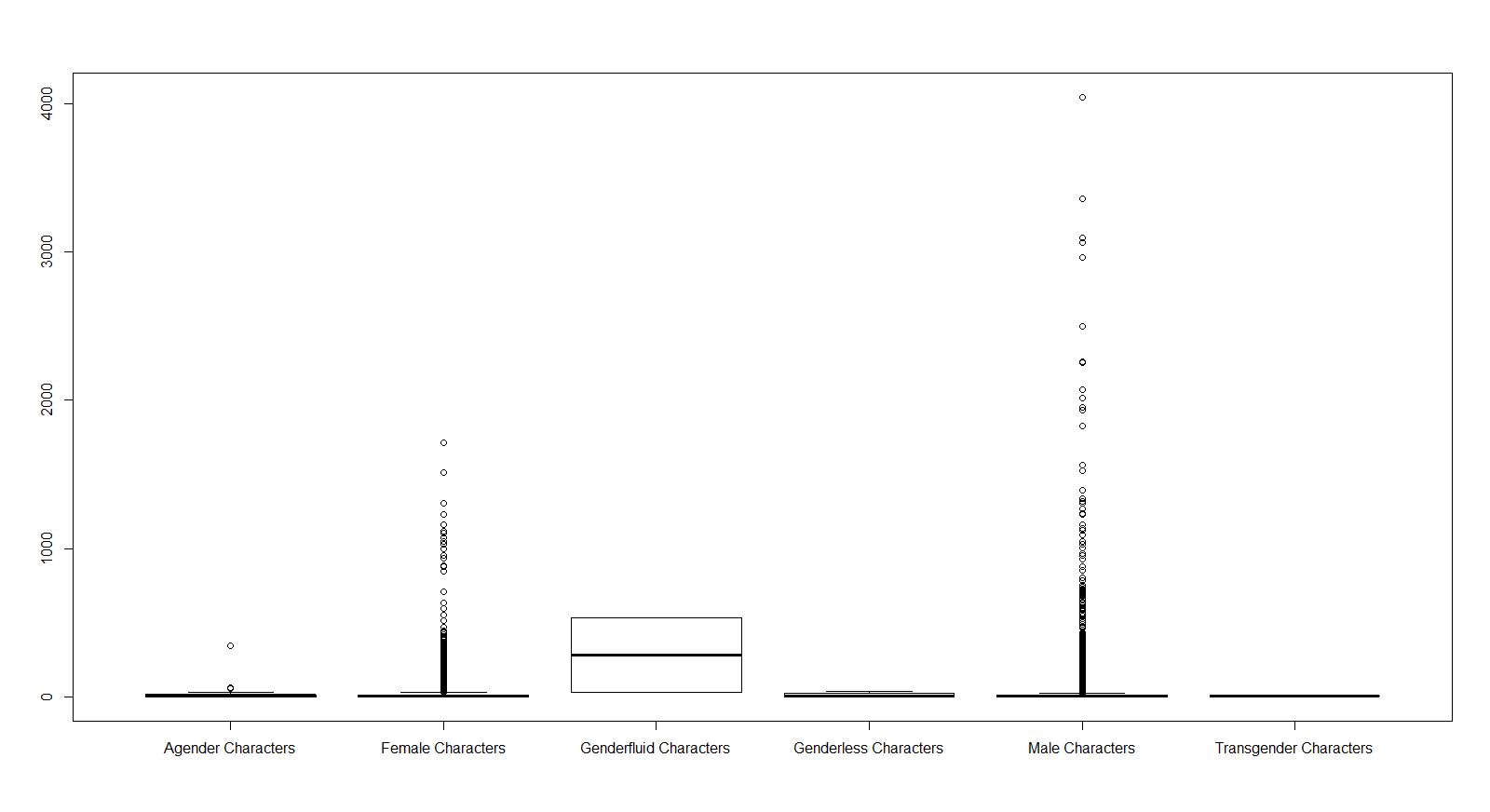
Genderfluid Characters 0

Genderless Characters 0

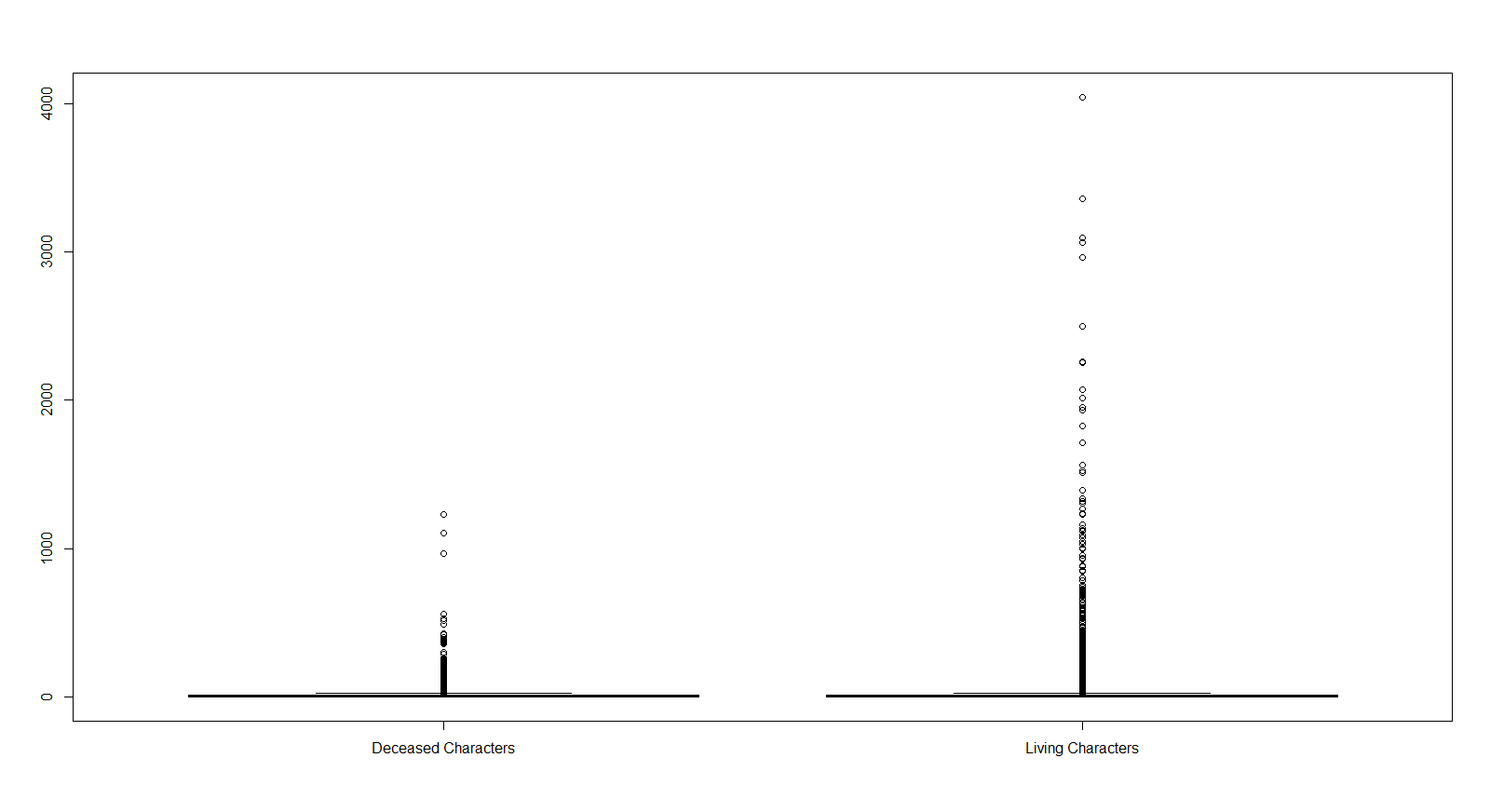
Male Characters 2

Transgender Characters 0

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:

Both variables are categorical. Each observation is independent. This meets the sample size assumption because more than 20% of the cells are populated with <5.

Q2:

We assume a normal distribution because the data set is large. Also, samples are independent because each character is only assigned one gender type. However, the data do not meet the assumption of equal variances; clearly, the ‘male’ and ‘female’ groups have very large variances and the other groups do not.

Q3:

The observations are independent. Since the sample size is large, we can assume the distribution is normal. The dependent variable is continuous. Equal variance can be tested in R; our data do not have equal variances.

Run the statistical test! Put your results here:

Q1:

# Pearson's Chi-squared test

# data: tbl

# X-squared = 680.45, df = 15, p-value < 2.2e-16

Q2:

Anova:

# Df Sum Sq Mean Sq F value Pr(>F)

#sex 5 155747 31149 3.409 0.00443 \*\*

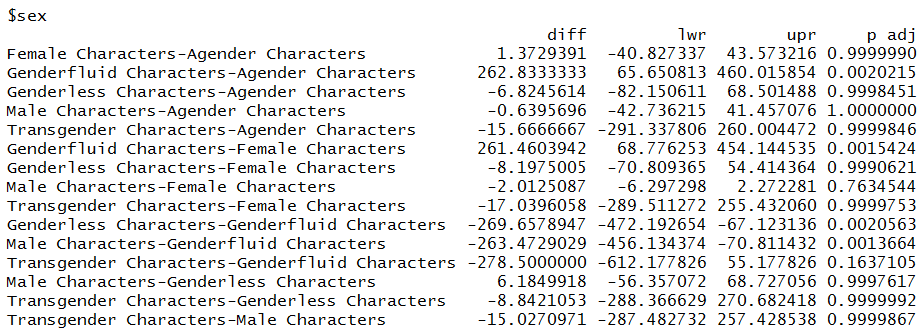
#Residuals 20963 191574001 9139

#---

#Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

#2303 observations deleted due to missingness

TukeyHSD:



Q3:

\_ran out of time to finish this question\_

Interpret your results!

Q1:

Since the p-value is less than 0.05, we reject the null hypothesis. We can take this to mean that the character’s sex and alignment are not independent.

Q2:

The anova test tells us that at least one mean is significantly different than the others; one gender has a mean that is significantly different. The Tukey test tells us if there is a difference between each combination of gender groups; we see that the genderfluid characters are significantly different than all other groups.

Q3:

\_ran out of time to finish this question\_