Name:

UmichID:

What dataset are you working with: biopic, nfl\_fav\_team

List 3 questions that you can ask with your dataset.

Q1: Does whether or not the subject of the movie is a person of color have an affect on the box office results?

Q2: Is there a correlation between race of subject and box office results?

Q3: Does ownership affect player likeability, coaching behavior, and on field behavior.

List the associated null hypothesis for each question:

Q1: There is no difference in the means.

Q2: There is no correlation.  
Q3: There is no correlation.

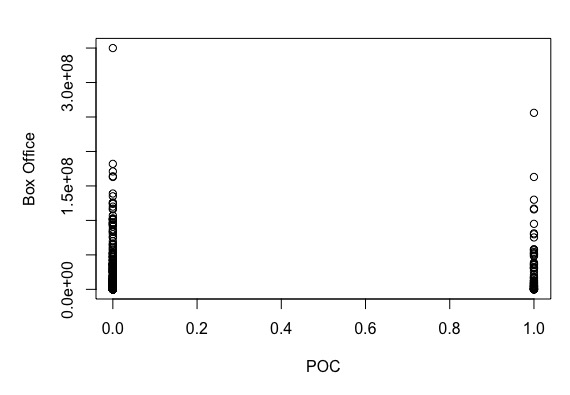
What statistical test(s) will you use to answer each of the questions:

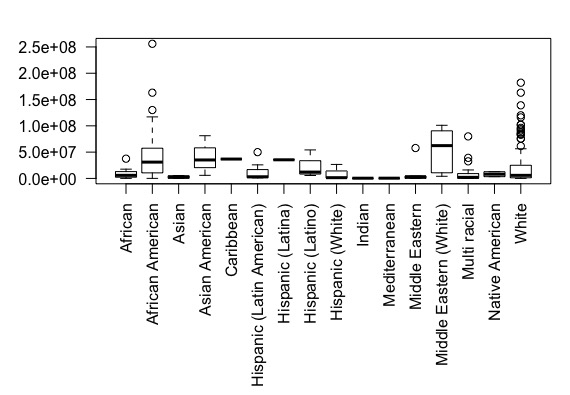
Q1: T-test

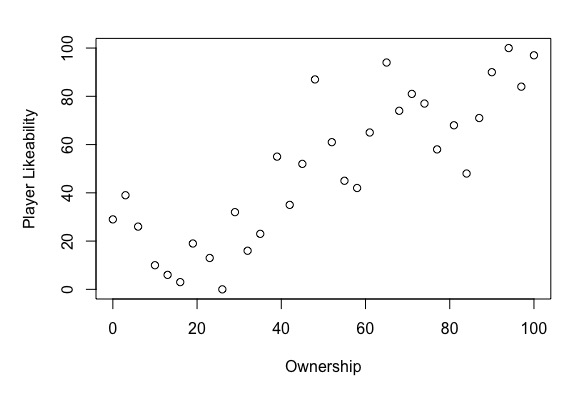
Q2: ANOVA

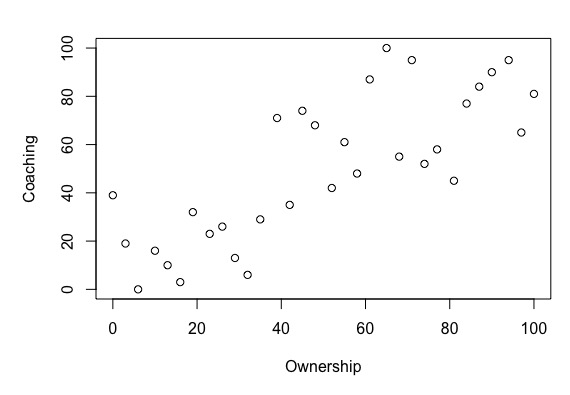
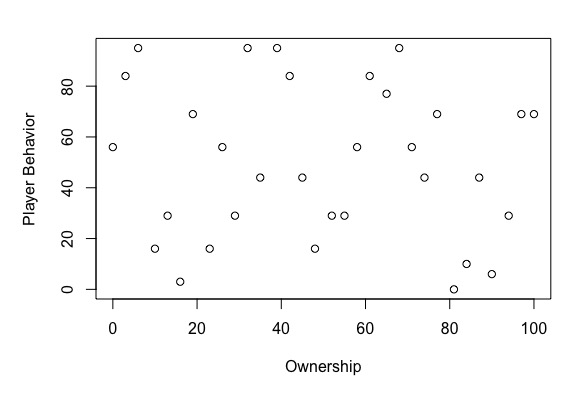
Q3: Linear regression

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: I examined the variance, which was not equal, so a Welch’s was used. Distribution was not normal, however sample size was more than 30.

Q2:

Q3:

Run the statistical test! Put your results here:

Q1: Welch Two Sample t-test

data: POCN[, "box\_office"] and POCY[, "box\_office"]

t = -1.4513, df = 99.779, p-value = 0.1498

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-18266663 2832524

sample estimates:

mean of x mean of y

21639075 29356145

Q2:

Q3:

Interpret your results!

Q1: There is a difference in mean box office earnings ($29,356,145 for a subject that is a person of color; $21,639,075 for not), however the p-val is above 0.05 so I will not reject the null hypothesis that there is no difference.

Q2:

Q3: