Name: Katie Pritchard

UmichID: kpritc

What dataset are you working with: tennis\_events\_time

List 3 questions that you can ask with your dataset.

Q1: Is there a significant different in match length between matches played on clay, grass, carpet, or hard surfaces?

Q2: Are matches played on hard courts longer than those played on clay?

Q3: Are matches played on grass courts longer than those played on carpet courts?

List the associated null hypothesis for each question:

Q1: There is no significant difference in length of matches played on clay, grass, carpet, and hard courts.

Q2: There is no significant difference in seconds added to match length on those played on clay v. hard courts.

Q3: There is no significant difference in seconds added to match length on those played on grass v. carpet courts.

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

Q1: One-way ANOVA

Q2: Two-sample t-test

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:



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 assume that the data is normally distributed, samples are independent, and that each population has the same variance. Since the p value in running a shapiro test <0.05, we conclude that our data is not normally distributed. Since the p value in running a var test is <0.05, we cannot conclude that the samples have equal variances.

Q2: We assume that the data is normally distributed, samples are independent, and that each population has the same variance. I ran Shapiro tests for both hard and clay surfaces and the p values for both were > 0.05, which means the data is normally distributed. I ran a var test and since p>0.05, we conclude that the samples have equal variances.

Q3: We assume that the data is normally distributed, samples are independent, and that each population has the same variance. I ran Shapiro tests for both grass and carpet surfaces and while the grass data is normally distributed, the p value for carpet surfaces is <0.05, which means the data is not normally distributed. I ran a var test and since p>0.05, we conclude that the samples have equal variances.

Run the statistical test! Put your results here:

Q1: P < 2e-16

Q2: p < 2.2e-16

Q3: p = 7.027\*e-05

Interpret your results!

Q1: Since p<0.05, we reject the null hypothesis and conclude that there is a significant difference in match length between matches played on clay, carpet, grass, or hard surface. We cannot run a Tukey test because we don’t have an equal sample size for each group.

Q2: Since p<0.05, we reject the null hypothesis that the mean time added to match length is the same on clay and hard courts and conclude that they are significantly different.

Q3: Since p<0.05, we reject the null hypothesis and conclude that the mean time added to match length is significantly different on grass v. carpet courts.