Assignment - 2

Session 6 – Visualization

and Plotting

1. Import the Titanic Dataset from the link Titanic Data Set.

Perform the following:

a. Is there any difference in fares by different class of tickets?

Note - Show a boxplot displaying the distribution of fares by class

Ans:

> cor(titanic$fare, titanic$pclass, use="complete.obs")

[1] -0.5586287

> aggregate(fare~pclass,titanic,median)

pclass fare

1 1 60.0000

2 2 15.0458

3 3 8.0500

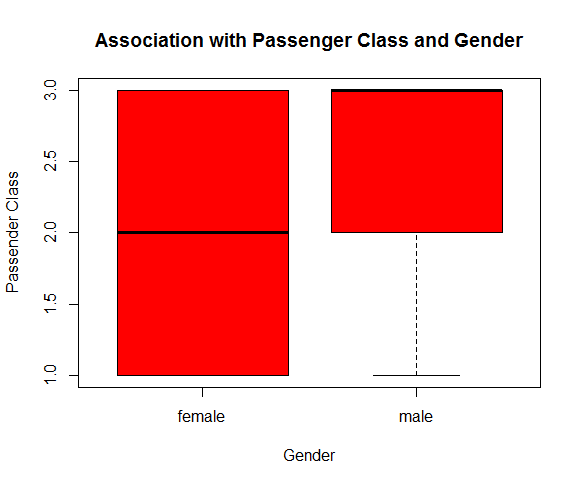
boxplot(fare~pclass, data=titanic, xlab="fare", ylab="pclass", main="Boxplot Displaying The Distribution Of Fares By Class", col="Blue")

Boxplot saved as .png file in GitHub link: https://github.com/Manishchugh79/DATA-ANALYTICS-WITH-R-EXCEL-TABLEAU\_Session6Assignment6.2

> summary(titanic$fare)

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

0.000 7.896 14.454 33.295 31.275 512.329 1



b. Is there any association with Passenger class and gender?

Note – Show a stacked bar chart

Ans:

Yes, there can be an association established between pclass and sex

> tapply(titanic$pclass, titanic$sex, shapiro.test)

$`female`

Shapiro-Wilk normality test

data: X[[i]]

W = 0.75037, p-value < 2.2e-16

$male

Shapiro-Wilk normality test data: X[[i]]

W = 0.70214, p-value < 2.2e-16

> bartlett.test(titanic$pclass~titanic$sex)

Bartlett test of homogeneity of variances

data: titanic$pclass by titanic$sex

Bartlett's K-squared = 2.5363, df = 1, p-value = 0.1113

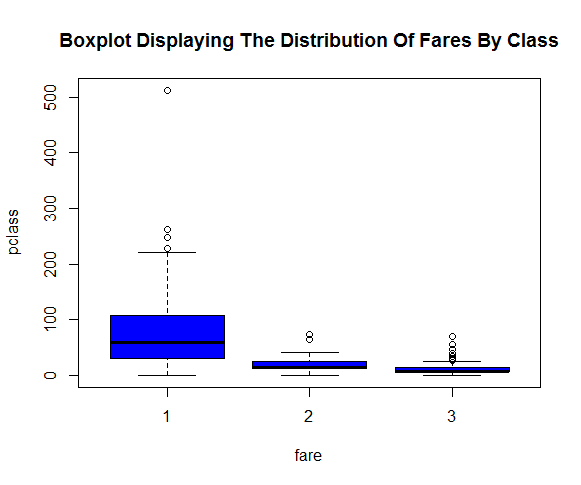
> oneway.test(titanic$pclass ~ titanic$sex)

One-way analysis of means (not assuming equal variances)

data: titanic$pclass and titanic$sex

F = 19.863, num df = 1.00, denom df = 908.08, p-value = 9.358e-06

plot(titanic$sex, titanic$pclass, xlab="Gender", ylab="Passender Class", col="red", main="Association with Passenger Class and Gender")

****