Assignment - 3\_5 Session - 3

#Q1: 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

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

# Note - Show a stacked bar chart

#Solution1

#a)

#import the titanic dataset

library(readxl)

titanic <- read\_xls("titanic3.xls")

#boxplot

boxplot(fare~pclass,data= titanic,

main="Fares Versus Pclass",xlab="Class",ylab="Fares",col=rainbow(3))

#Class 1 has the highest range of fares and high avg fares, then class 2 and class 3

#b)

#stacked bar chart

counts<-table(titanic$sex,titanic$pclass)

barplot(counts, main = "Distribution of Class by gender", xlab="Pclass", col=c("blue", "green"), legend = c("Female","Male"), names.arg = c("Pclass1st", "Pclass2nd","Pclass3rd"))

#we can do chisq test also for checking association

chisq.test(titanic$pclass ,titanic$sex)

#ho:there is no association

#since p value is 0.0002064<0.05

#we reject the null hypothesis and thus there is association