SESSION 6 – ASSIGNMENT 6.2

Date: 21st Jan 2019

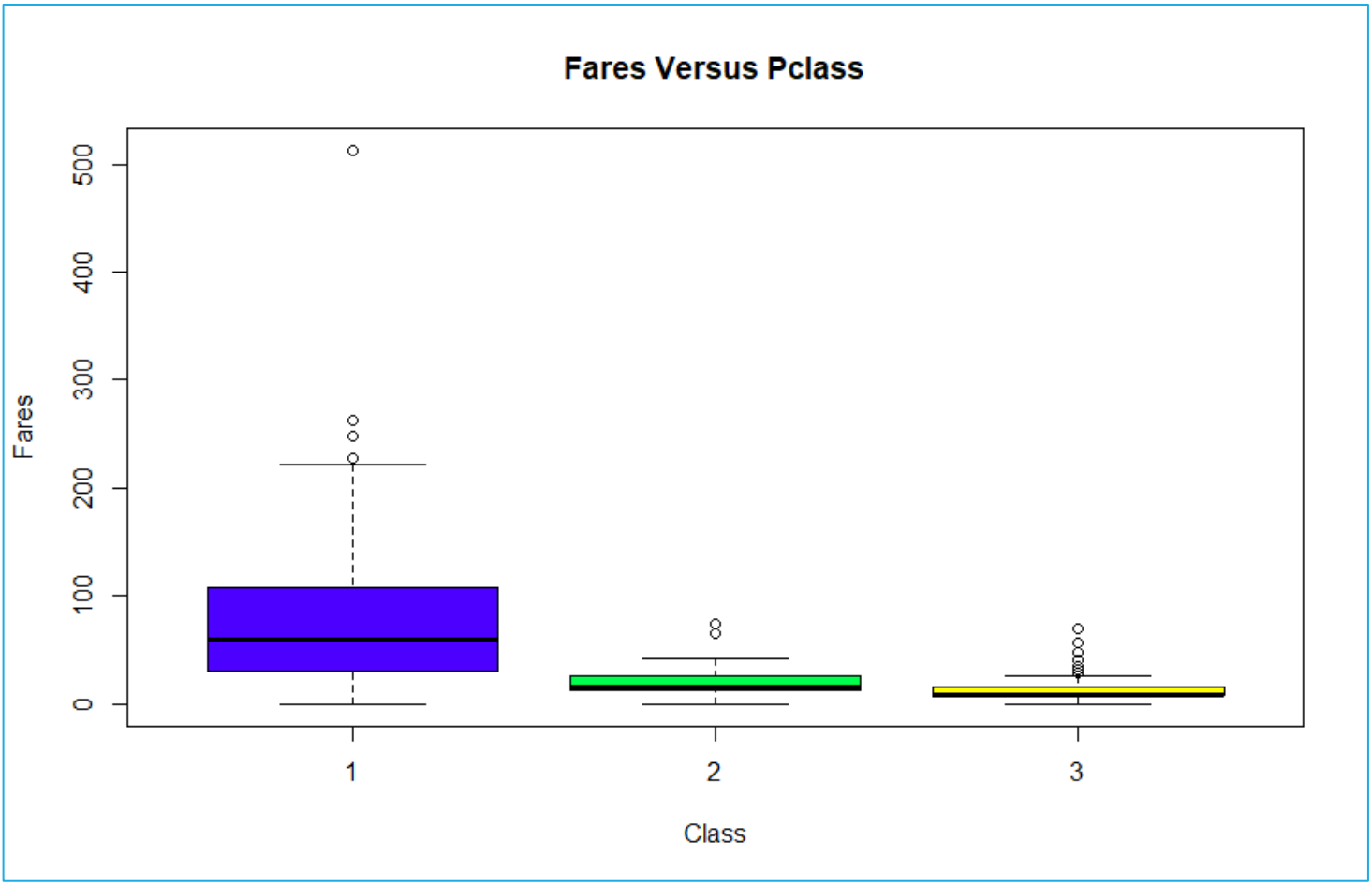
1. **Import the Titanic Dataset from the following link:**
   https://drive.google.com/file/d/1JTJCjdGuUxzKXYlwOavwovB01k6FWg 3r/view?ts=5b42ea10
   Perform the below operations:
   a. **Is there any difference in fares by different class of tickets? (Note- show a boxplot displaying the distribution of fares by class)**
   b. **Is there any association with Passenger class and gender? (Note- show a stacked bar chart)**

```
#a. Is there any difference in fares by different class of tickets?
#Note - Show a boxplot displaying the distribution of fares by class use titanic dataset

library(readr)
titanic <- read_csv("G:/DATA ANALYTICS/DATA/titanic3.csv")
str(titanic)

View(titanic)

boxplot(fare~pclass,data= titanic,
      main="Fares Versus Pclass",xlab="Class",ylab="Fares",col=topo.colors(3))
```

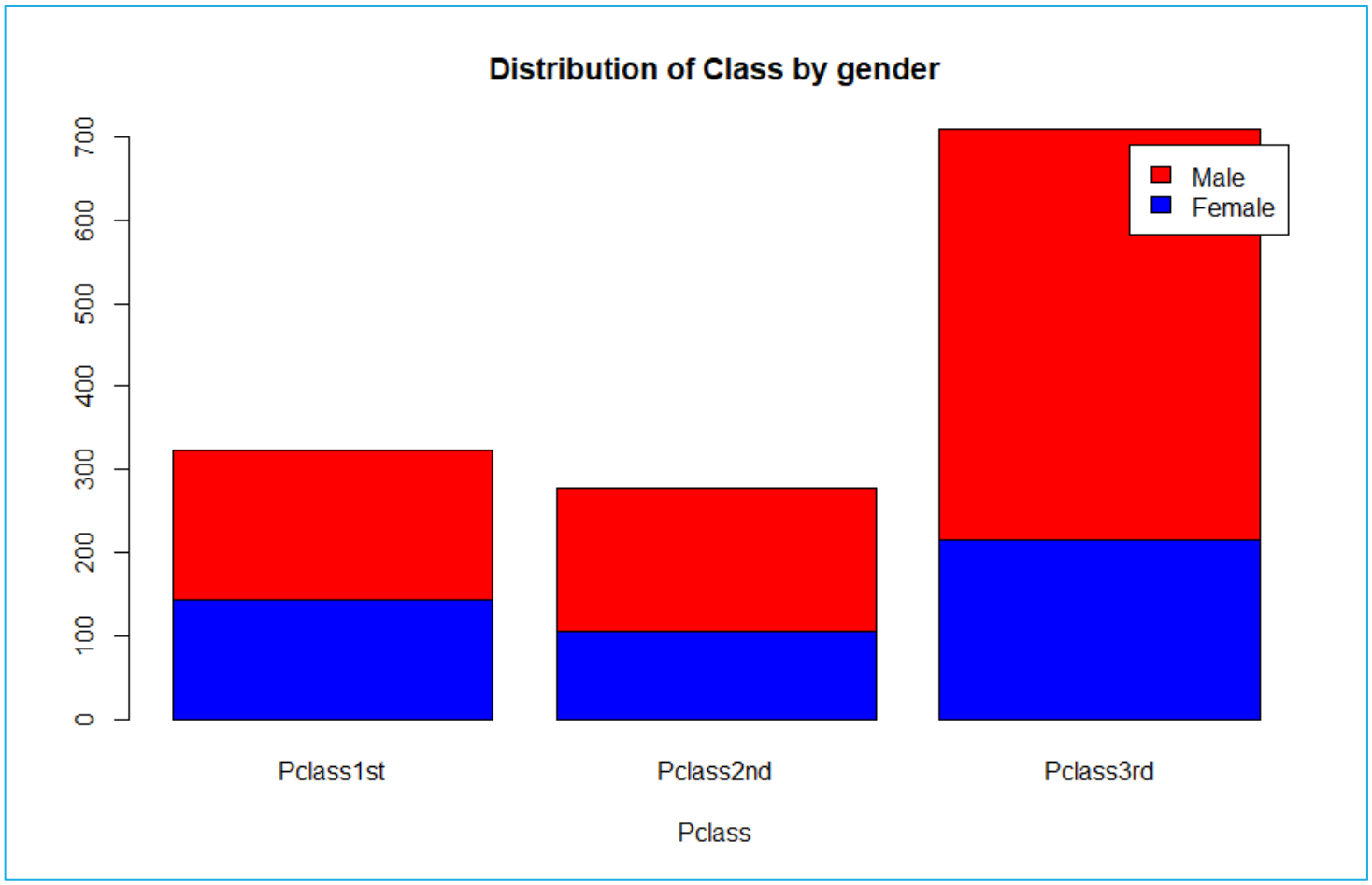


 # Answer1.b. Is there any association with Passenger class and gender?
 # Note - Show a stacked bar chart

```
counts<-table(titanic$sex,titanic$pclass)
barplot(counts, main = "Distribution of Class by gender", xlab="Pclass", col=c("blue", "red"),
legend = c("Female","Male"), names.arg = c("Pclass1st", "Pclass2nd","Pclass3rd"))


#another way --> chisq test 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 say there is association
```


Distribution of Class by gender
0
100
200
300
400
500
600
700
Male
Female
Pclass1st
Pclass2nd
Pclass3rd
Pclass