ACADGILD

SESSION 6: Visualization & Plotting

Assignment 2

PROBLEM STATEMENT

1. Import the Titanic Dataset from the following link:
   https://drive.google.com/file/d/1JTJCjdGuUxzKXYlwOavwovB01k6FWg3r/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

SOLUTION:

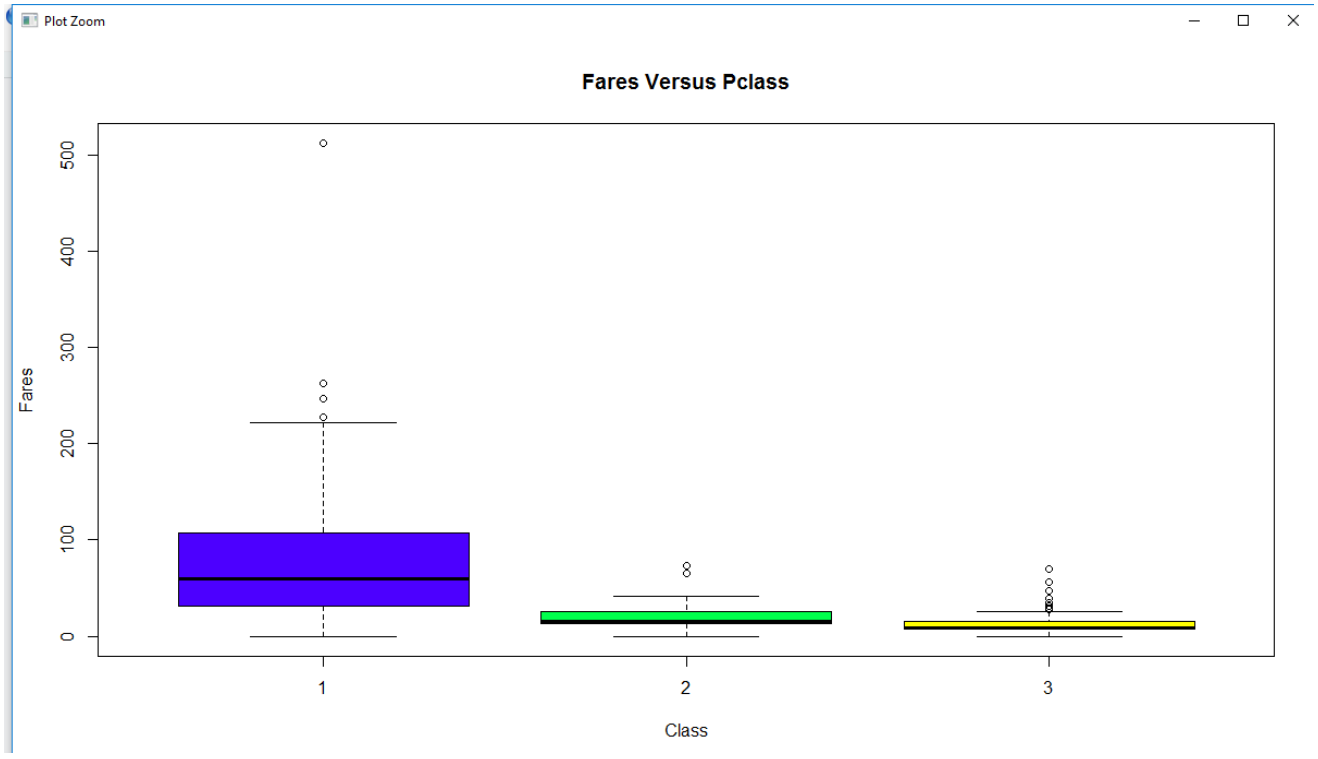
A. **The R-script for the given problem is as follows:**

```
#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("F:/ACADGILD - Online Course/1. DATA SETS/titanic3.csv")
str(titanic)
View(titanic)

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

The output of the R-Script (from Console window) is given as follows:

```
> library(readr)
> titanic <- read_csv("F:/ACADGILD - Online Course/1. DATA
SETS/titanic3.csv")
Parsed with column specification:
cols(
  pclass = col_double(),
  survived = col_double(),
  name = col_character(),
  sex = col_character(),
  age = col_double(),
  sibsp = col_double(),
  parch = col_double(),
  ticket = col_character(),
  fare = col_double(),
  cabin = col_character(),
  embarked = col_character(),
  boat = col_character(),
  body = col_double(),
  home.dest = col_character()
)
> str(titanic)
Classes ‘spec_tbl_df’, ‘tbl_df’, ‘tbl’ and 'data.frame': 1310 obs.
of  14 variables:
 $ pclass   : num  1 1 1 1 1 1 1 1 1 1 ...
 $ survived : num  1 1 0 0 0 1 1 0 1 0 ...
 $ name     : chr  "Allen, Miss. Elisabeth Walton" "Allison,
Master. Hudson Trevor" "Allison, Miss. Helen Loraine" "Allison,
Mr. Hudson Joshua Creighton" ...
 $ sex      : chr  "female" "male" "female" "male" ...
 $ age      : num  29 0.917 2 30 25 ...
 $ sibsp    : num  0 1 1 1 1 0 1 0 2 0 ...
 $ parch    : num  0 2 2 2 2 0 0 0 0 0 ...
 $ ticket   : chr  "24160" "113781" "113781" "113781" ...
 $ fare     : num  211 152 152 152 152 ...
 $ cabin    : chr  "B5" "C22 C26" "C22 C26" "C22 C26" ...
 $ embarked : chr  "S" "S" "S" "S" ...
 $ boat     : chr  "2" "11" NA NA ...
 $ body     : num  NA NA NA 135 NA NA NA NA NA 22 ...
```

```
 $ home.dest: chr  "St Louis, MO" "Montreal, PQ / Chesterville,
ON" "Montreal, PQ / Chesterville, ON" "Montreal, PQ /
Chesterville, ON" ...
 - attr(*, "spec")=
  .. cols(
  ..   pclass = col_double(),
  ..   survived = col_double(),
  ..   name = col_character(),
  ..   sex = col_character(),
  ..   age = col_double(),
  ..   sibsp = col_double(),
  ..   parch = col_double(),
  ..   ticket = col_character(),
  ..   fare = col_double(),
  ..   cabin = col_character(),
  ..   embarked = col_character(),
  ..   boat = col_character(),
  ..   body = col_double(),
  ..   home.dest = col_character()
  .. )
> View(titanic)
> boxplot(fare~pclass,data= titanic,
+         main="Fares Versus
Pclass",xlab="Class",ylab="Fares",col=topo.colors(3))
```

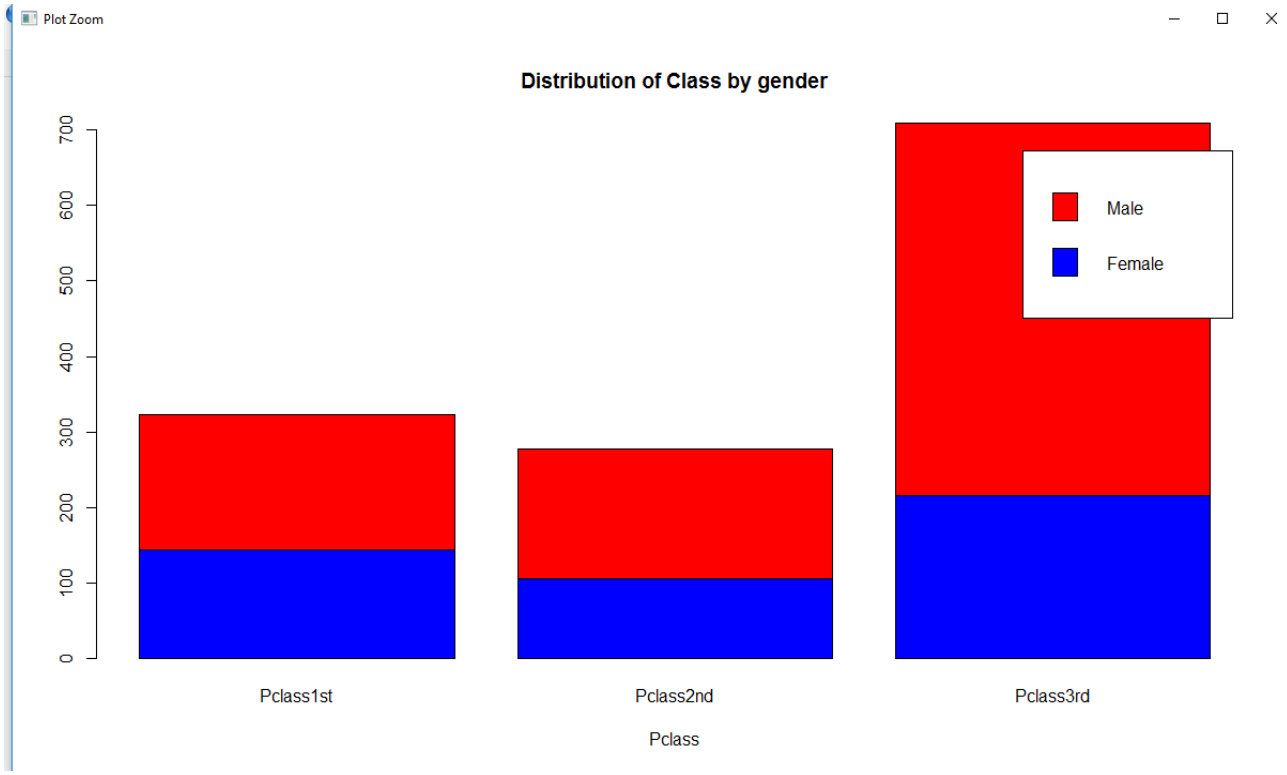


B. **The R-script for the given problem is as follows:**

```
#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
```

The output of the R-Script (from Console window) is given as follows:

```
> 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)

        Pearson's Chi-squared test

data:  titanic$pclass and titanic$sex
X-squared = 20.379, df = 2, p-value = 3.757e-05

>
> #ho:there is no association
> #ho:there is no association
> #since p value is 0.0002064<0.05
> #ho:there is no association
> #since p value is 0.0002064<0.05
> #we reject the null hypothesis and thus say there is
association
```