ACADGILD

SESSION 6:
Visualization & Plotting
Assignment 2

Submitted by: Munmun Ghosal
Login Id: munmun55@gmail.com
(M):+91-8007178659

Table of Contents

1. 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

2. Solution

Import the Titanic Dataset

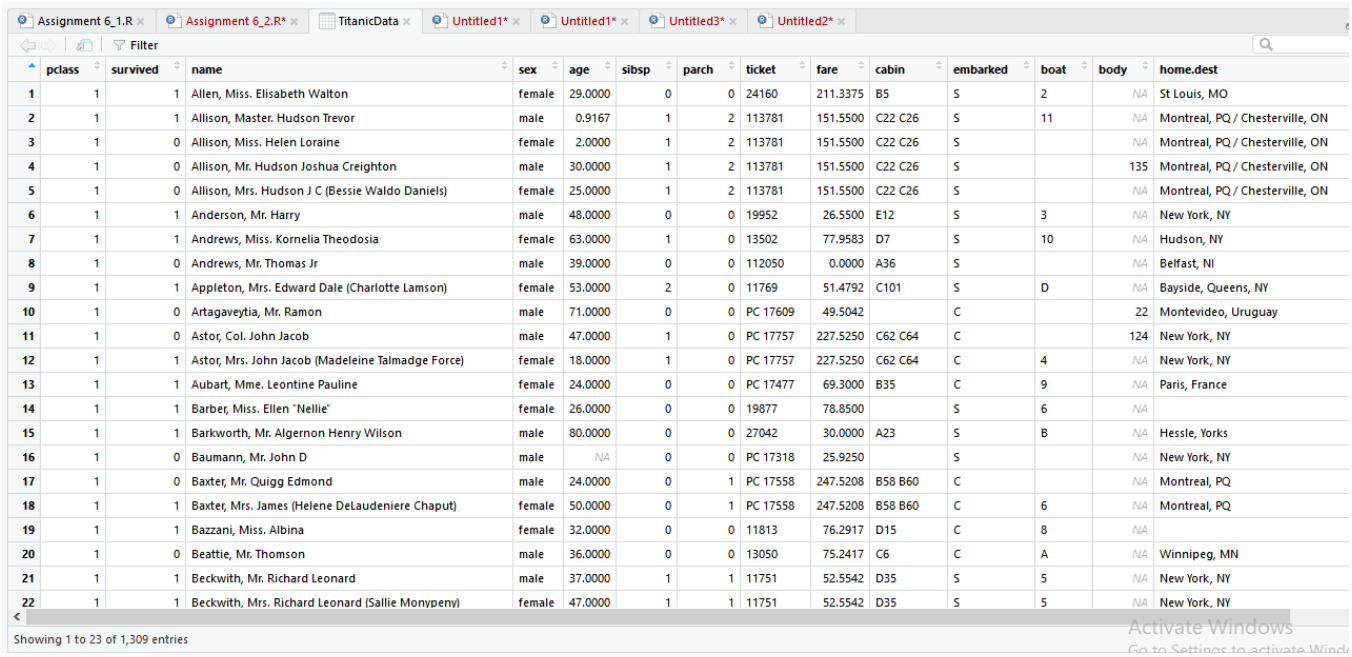
The R-script for the given problem is as follows:

```
library("readr")
TitanicData <- read.csv("E:/munmun_acadgild/acadgild data analytics/supporting files/titanic3.csv")
View(TitanicData)
str(TitanicData)
```

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

```
> library("readr")
> TitanicData <- read.csv("E:/munmun_acadgild/acadgild data analytics/supporting files/titanic3.csv")
> View(TitanicData)
> str(TitanicData)
'data.frame':   1309 obs. of  14 variables:
 $ pclass   : int  1 1 1 1 1 1 1 1 1 1 ...
 $ survived : int  1 1 0 0 0 1 1 0 1 0 ...
 $ name     : Factor w/ 1307 levels "Abbing, Mr. Anthony",..: 22 24 25 26 27 31 46 47 51 55 ...
 $ sex      : Factor w/ 2 levels "female","male": 1 2 1 2 1 2 1 2 1 2 ...
 $ age      : num  29 0.917 2 30 25 ...
 $ sibsp    : int  0 1 1 1 1 0 1 0 2 0 ...
 $ parch    : int  0 2 2 2 2 0 0 0 0 0 ...
 $ ticket   : Factor w/ 929 levels "110152","110413",..: 188 50 50 50 50 125 93 16 77 826 ...
 $ fare     : num  211 152 152 152 152 ...
 $ cabin    : Factor w/ 187 levels "","A10","A11",..: 45 81 81 81 81 151 147 17 63 1 ...
 $ embarked : Factor w/ 4 levels "","C","Q","S": 4 4 4 4 4 4 4 4 4 2 ...
 $ boat     : Factor w/ 28 levels "","1","10","11",..: 13 4 1 1 1 14 3 1 28 1 ...
 $ body     : int  NA NA NA 135 NA NA NA NA NA 22 ...
 $ home.dest: Factor w/ 370 levels "","?Havana, Cuba",..: 310 232 232 232 232 238 163 25 23 230 ...
>
```

The titanic dataset is shown as follows:



Assignment 6_1.R | Assignment 6_2.R* | TitanicData | Untitled1* | Untitled1* | Untitled3* | Untitled2*

Filter

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
1	1	1	Allen, Miss. Elisabeth Walton	female	29.0000	0	0	24160	211.3375	B5	S	2	NA	St Louis, MO
2	1	1	Allison, Master. Hudson Trevor	male	0.9167	1	2	113781	151.5500	C22 C26	S	11	NA	Montreal, PQ / Chesterville, ON
3	1	0	Allison, Miss. Helen Loraine	female	2.0000	1	2	113781	151.5500	C22 C26	S		NA	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ / Chesterville, ON
5	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1	2	113781	151.5500	C22 C26	S		NA	Montreal, PQ / Chesterville, ON
6	1	1	Anderson, Mr. Harry	male	48.0000	0	0	19952	26.5500	E12	S	3	NA	New York, NY
7	1	1	Andrews, Miss. Kornelia Theodosia	female	63.0000	1	0	13502	77.9583	D7	S	10	NA	Hudson, NY
8	1	0	Andrews, Mr. Thomas Jr	male	39.0000	0	0	112050	0.0000	A36	S		NA	Belfast, NI
9	1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53.0000	2	0	11769	51.4792	C101	S	D	NA	Bayside, Queens, NY
10	1	0	Artagaveytia, Mr. Ramon	male	71.0000	0	0	PC 17609	49.5042		C		22	Montevideo, Uruguay
11	1	0	Astor, Col. John Jacob	male	47.0000	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY
12	1	1	Astor, Mrs. John Jacob (Madeleine Talmadge Force)	female	18.0000	1	0	PC 17757	227.5250	C62 C64	C	4	NA	New York, NY
13	1	1	Aubart, Mme. Leontine Pauline	female	24.0000	0	0	PC 17477	69.3000	B35	C	9	NA	Paris, France
14	1	1	Barber, Miss. Ellen "Nellie"	female	26.0000	0	0	19877	78.8500		S	6	NA	
15	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0000	0	0	27042	30.0000	A23	S	B	NA	Hessle, Yorks
16	1	0	Baumann, Mr. John D	male	NA	0	0	PC 17318	25.9250		S		NA	New York, NY
17	1	0	Baxter, Mr. Quigg Edmond	male	24.0000	0	1	PC 17558	247.5208	B58 B60	C		NA	Montreal, PQ
18	1	1	Baxter, Mrs. James (Helene DeLaudeniere Chaput)	female	50.0000	0	1	PC 17558	247.5208	B58 B60	C	6	NA	Montreal, PQ
19	1	1	Bazzani, Miss. Albina	female	32.0000	0	0	11813	76.2917	D15	C	8	NA	
20	1	0	Beattie, Mr. Thomson	male	36.0000	0	0	13050	75.2417	C6	C	A	NA	Winnipeg, MN
21	1	1	Beckwith, Mr. Richard Leonard	male	37.0000	1	1	11751	52.5542	D35	S	5	NA	New York, NY
22	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0000	1	1	11751	52.5542	D35	S	5	NA	New York, NY

Showing 1 to 23 of 1,309 entries

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

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

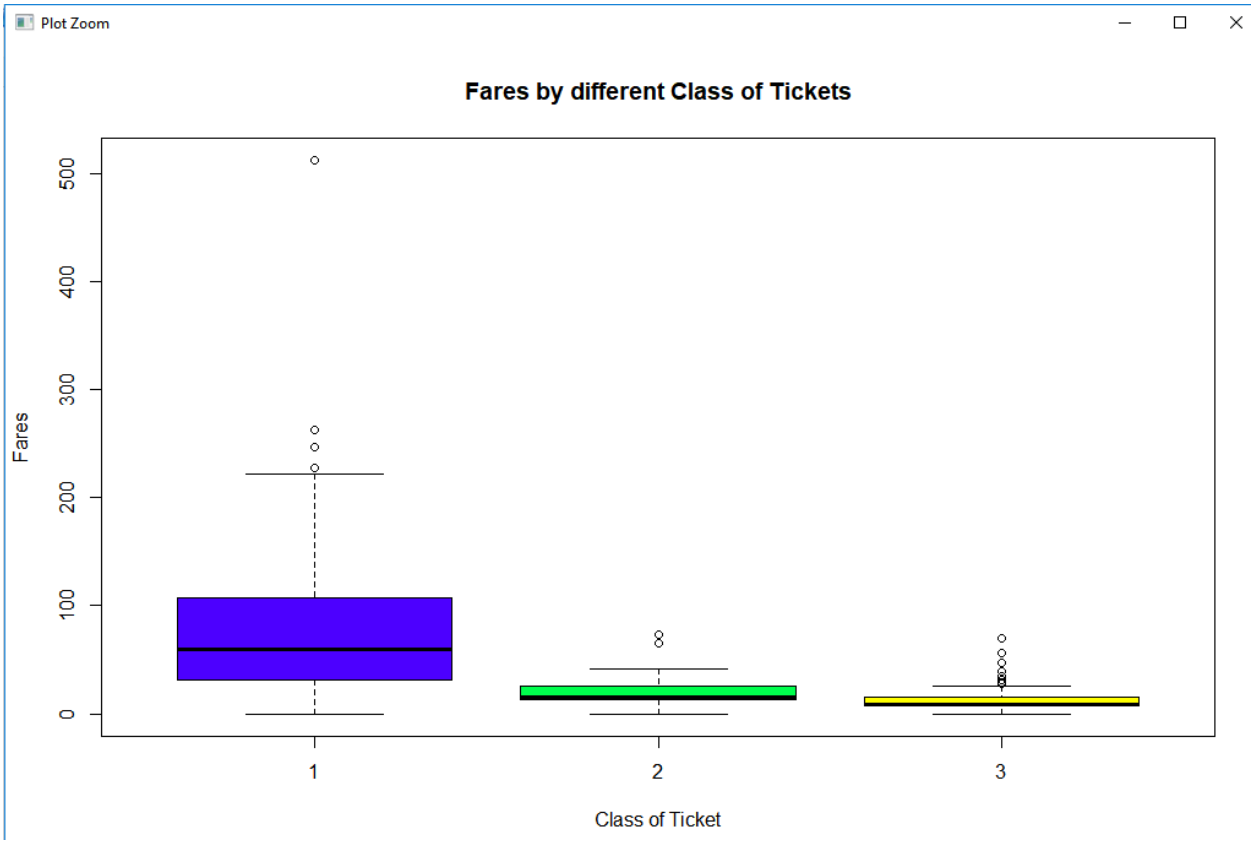
The R-script for the given problem is as follows:

```
library("readr")
TitanicData <- read.csv("E:/munmun_acadgild/acadgild data analytics/supporting
files/titanic3.csv")
View(TitanicData)
str(TitanicData)

colnames(TitanicData) <-
c("Pclass","Survived","Name","Sex","Age","SibSp","Parch","Ticket","Fare",
                "Cabin","Embarked","Boat","Body","destination")

Titanic <- TitanicData %>% mutate(Pclass = as.factor(Pclass))  # Passenger class as factor
str(Titanic)
View(Titanic)

boxplot(Fare~Pclass, data = Titanic, col = topo.colors(3),
      xlab = "Class of Ticket", ylab = "Fares", main = "Fares by different Class of Tickets")
```

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

```
> library("readr")
> TitanicData <- read.csv("E:/munmun_acadgild/acadgild data
analytics/supporting files/titanic3.csv")
> View(TitanicData)
> str(TitanicData)
'data.frame':  1309 obs. of  14 variables:
 $ pclass   : int  1 1 1 1 1 1 1 1 1 1 ...
 $ survived : int  1 1 0 0 0 1 1 0 1 0 ...
 $ name     : Factor w/ 1307 levels "Abbing, Mr. Anthony",..: 22 24 25 26 27
31 46 47 51 55 ...
 $ sex      : Factor w/ 2 levels "female","male": 1 2 1 2 1 2 1 2 1 2 ...
 $ age      : num  29 0.917 2 30 25 ...
 $ sibsp    : int  0 1 1 1 1 0 1 0 2 0 ...
 $ parch    : int  0 2 2 2 2 0 0 0 0 0 ...
 $ ticket   : Factor w/ 929 levels "110152","110413",..: 188 50 50 50 50 125
93 16 77 826 ...
 $ fare     : num  211 152 152 152 152 ...
 $ cabin    : Factor w/ 187 levels "","A10","A11",..: 45 81 81 81 81 151 147
17 63 1 ...
 $ embarked : Factor w/ 4 levels "","C","Q","S": 4 4 4 4 4 4 4 4 4 2 ...
 $ boat     : Factor w/ 28 levels "","1","10","11",..: 13 4 1 1 1 14 3 1 28 1
...
 $ body     : int  NA NA NA 135 NA NA NA NA NA 22 ...
 $ home.dest: Factor w/ 370 levels "","?Havana, Cuba",..: 310 232 232 232 232
238 163 25 23 230 ...
>
> colnames(TitanicData) <-
c("Pclass","Survived","Name","Sex","Age","SibSp","Parch","Ticket","Fare",
+                           "Cabin","Embarked","Boat","Body","destination")
>
> Titanic <- TitanicData %>% mutate(Pclass = as.factor(Pclass))  # Passennger
class as factor
> str(Titanic)
'data.frame':  1309 obs. of  14 variables:
 $ Pclass     : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
 $ Survived   : int  1 1 0 0 0 1 1 0 1 0 ...
 $ Name       : Factor w/ 1307 levels "Abbing, Mr. Anthony",..: 22 24 25 26
27 31 46 47 51 55 ...
 $ Sex        : Factor w/ 2 levels "female","male": 1 2 1 2 1 2 1 2 1 2 ...
 $ Age        : num  29 0.917 2 30 25 ...
 $ SibSp      : int  0 1 1 1 1 0 1 0 2 0 ...
 $ Parch      : int  0 2 2 2 2 0 0 0 0 0 ...
 $ Ticket     : Factor w/ 929 levels "110152","110413",..: 188 50 50 50 50
125 93 16 77 826 ...
 $ Fare       : num  211 152 152 152 152 ...
 $ Cabin      : Factor w/ 187 levels "","A10","A11",..: 45 81 81 81 81 151
147 17 63 1 ...
 $ Embarked   : Factor w/ 4 levels "","C","Q","S": 4 4 4 4 4 4 4 4 4 2 ...
 $ Boat       : Factor w/ 28 levels "","1","10","11",..: 13 4 1 1 1 14 3 1 28
1 ...
 $ Body       : int  NA NA NA 135 NA NA NA NA NA 22 ...
 $ destination: Factor w/ 370 levels "","?Havana, Cuba",..: 310 232 232 232
232 238 163 25 23 230 ...
```

```
> View(Titanic)
>
> boxplot(Fare~Pclass, data = Titanic, col = topo.colors(3),
+         xlab = "Class of Ticket", ylab = "Fares", main = "Fares by
different Class of Tickets")
```



Conclusion/Interpretation:

- Yes. Fares are different as per Class of Ticket.

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

Note- show a stacked bar chart

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

```
A <- table(Titanic$Sex, Titanic$Pclass)
A
bp <- barplot(A, col= rainbow(length(A)), legend = rownames(A),
        main = "Passenger class and gender",
        xlab = "Class of Ticket", ylab = "No. of Passangers by Gender")
```

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

```
> A <- table(Titanic$Sex, Titanic$Pclass)
> A
```

```
          1   2   3
  female 144 106 216
  male   179 171 493
> bp <- barplot(A, col= rainbow(length(A)), legend = rownames(A),
+               main = "Passenger class and gender",
+               xlab = "Class of Ticket", ylab = "No. of Passangers by Gender")
```

Conclusion/Interpretation:

- Male passengers are more than female in each class .
- The percentage of male passengers over Female Passengers is more in class 3 as compared to class 1 & 2