ACADGILD

SESSION 6: Visualization & Plotting

Assignment 2

Table of Contents

1. Introduction

This assignment will help you understand the concepts learnt in the session.

2. Objective

This assignment will test your skills on Visualization and Plotting operations in R.

3. Prerequisites

Not applicable.

4. Associated Data Files

Not applicable.

5. 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?

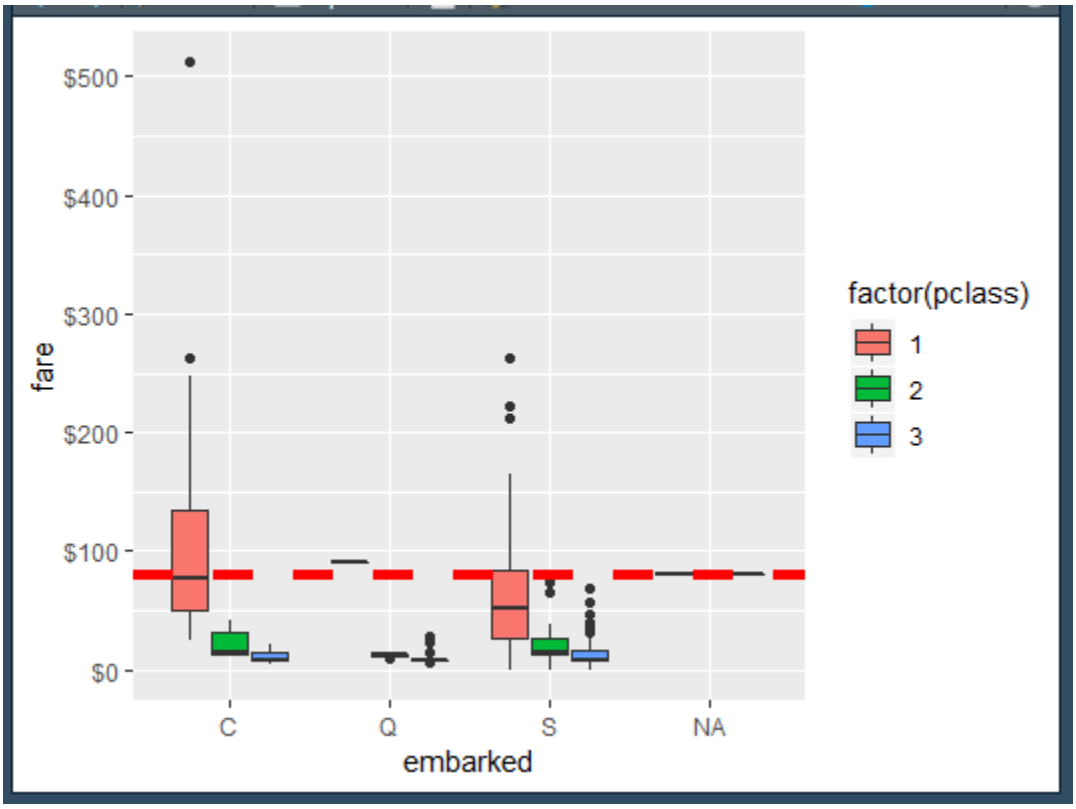
   Note- show a stacked bar chart

SOLUTIONS
Library:ggplot2
Library:dplyr

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

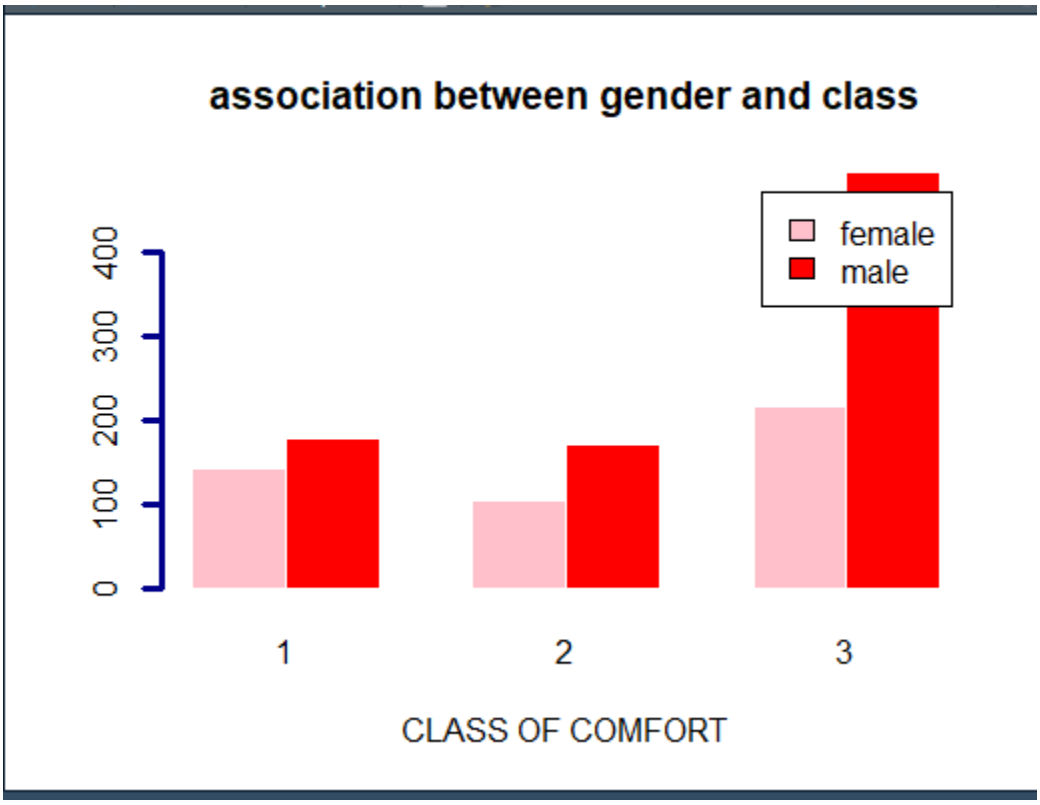
   Note- show a boxplot displaying the distribution of fares by class

SOLUTION:
```
# This variable appears to have a lot of missing values
 full$cabin[1:28]
# The first character is the deck. For example:
 strsplit(full$cabin[2], NULL)[[1]]
# Create a Deck variable. Get passenger deck A - F:
full$Deck<-factor(sapply(full$cabin, function(x) strsplit(x, NULL)[[1]][1]))
# Passengers 62 and 830 are missing Embarkment
full[c(62, 830), 'embarked']
# Get rid of our missing passenger IDs
embark_fare <- full %>%
 filter(ticket != 62 & ticket != 830)
# Use ggplot2 to visualize embarkment, passenger class, & median fare
library(ggplot2)
 ggplot(embark_fare, aes(x = embarked, y = fare, fill = factor(pclass))) +
geom_boxplot() +
geom_hline(aes(yintercept=80),
colour='red', linetype='dashed', lwd=2) +
scale_y_continuous(labels=dollar_format())
```



b.Is there any association with Passenger class and gender?
Note- show a stacked bar chart
SOLUTION: as per the data , there is association between gender and passenger class

```
asc=table(full$sex, full$pclass)
barplot(asc, legend=rownames(asc),beside = T, main = "association between gender and class",
xlab = "CLASS OF COMFORT", col = c('pink','red'), fg="darkblue", lwd=3,border = F)
```


association between gender and class
female
male
0
100
200
300
400
1
2
3
CLASS OF COMFORT