



ACADGILD

SESSION 6: Visualization & Plotting

Assignment 2

Table of Contents

1.Introduction
3

2.Objective
3

3.Prerequisites
3

4.Associated Data Files
3

5.Problem Statement
3

Data Analytics

6.Expected Output
3

7.Approximate Time to Complete Task
3

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/1JTJCidGuUxzKXYlwOavwovB01k6FWg3r/view?ts=5b42ea10>

Perform the below operations:

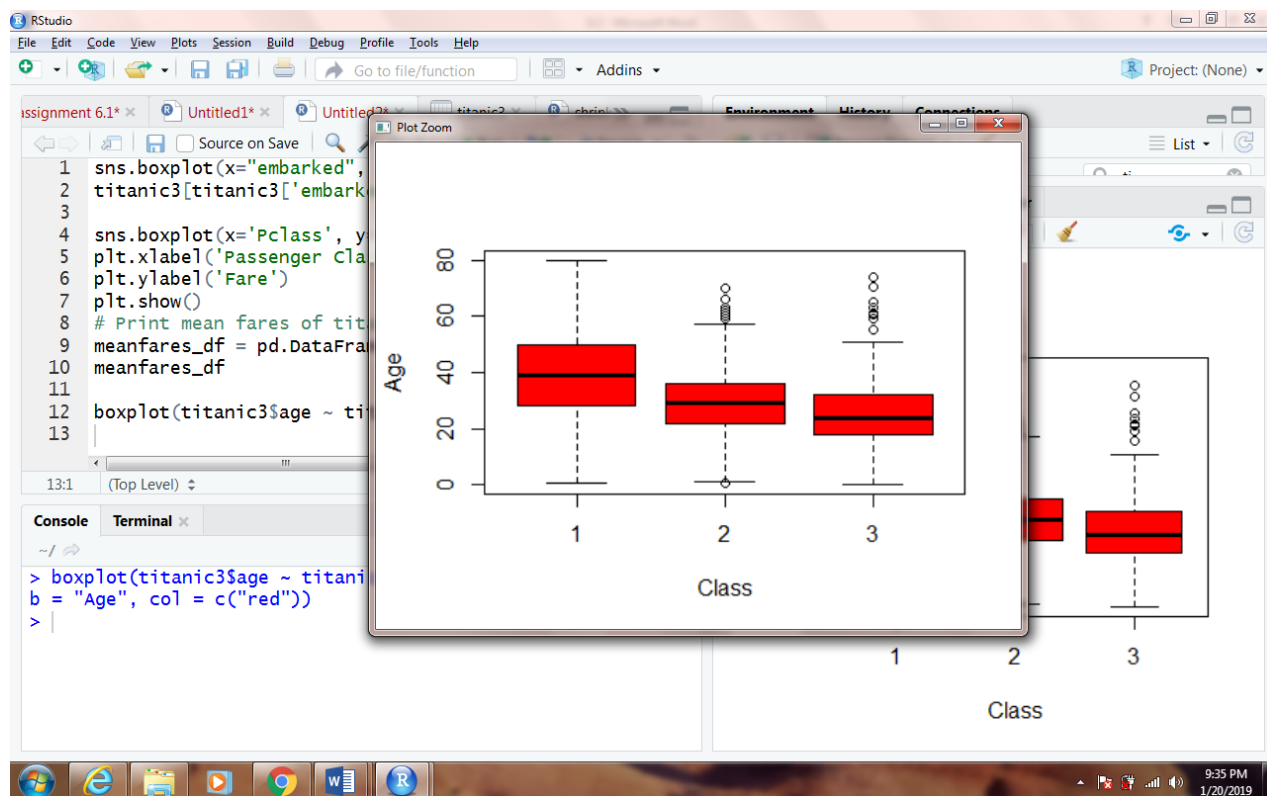
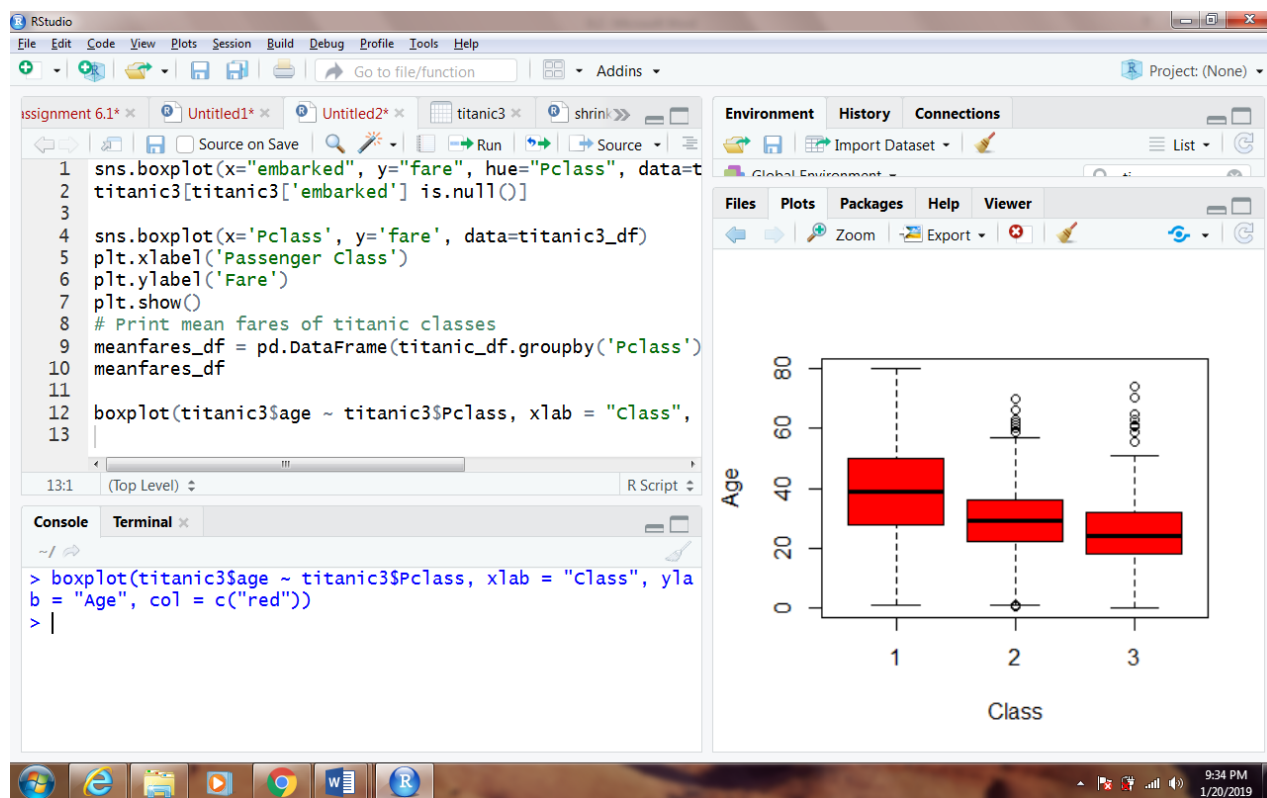
- a. Is there any difference in fares by different class of tickets?

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

Ans :

```
boxplot(train$Age ~ train$Pclass, xlab = "Class", ylab = "Age", col = c("red"))
```

Data Analytics



Data Analytics

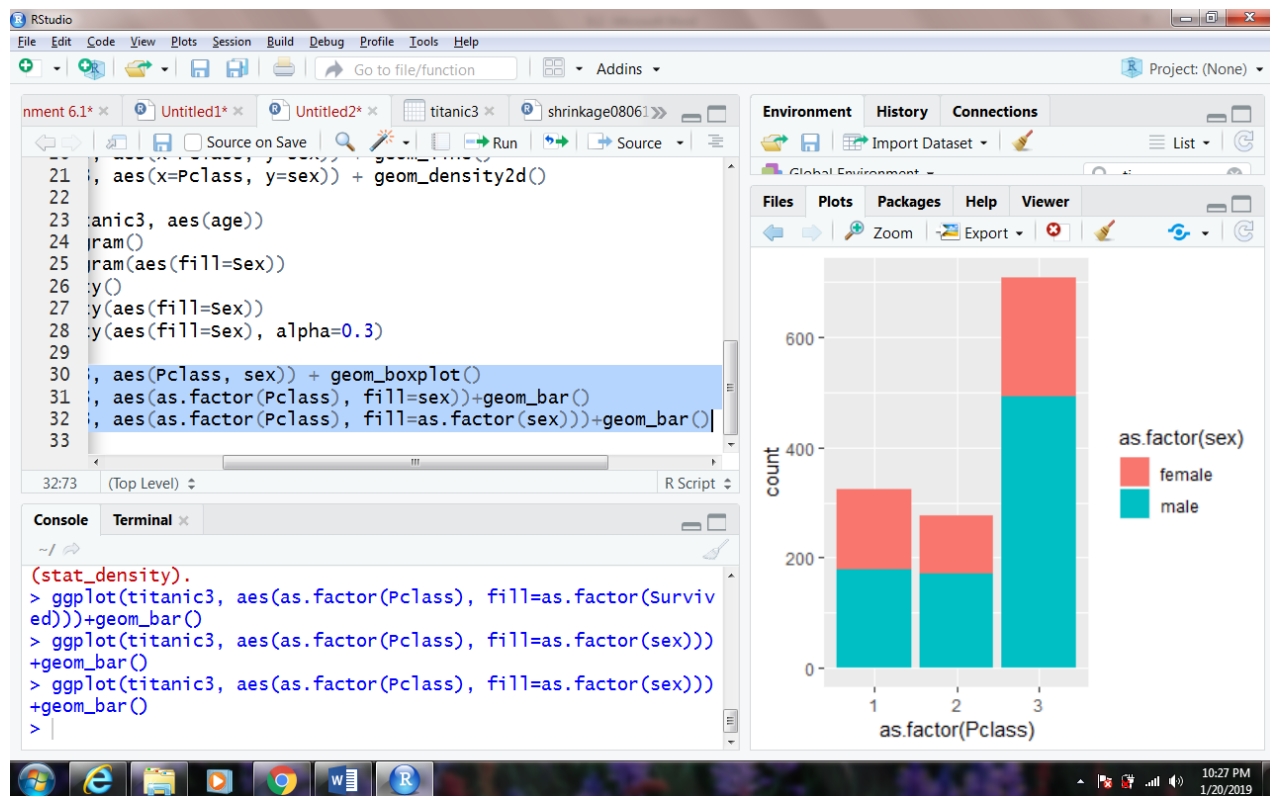
Interpretation: After analyzing this box plot of Class and Age, we can infer that the median value of Class 1 passengers is higher than that of Class 2 and Class 3. It means Class 2 and Class 3 had younger passengers than Class 1. Infact, class 3 had the youngest age group of passengers. The tiny circles outside the range of Class 2 and Class 3 represents the outlier values.

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

Note- show a stacked bar chart

```
ggplot(titanic3, aes(Pclass, sex)) + geom_boxplot()  
ggplot(titanic3, aes(as.factor(Pclass), fill=sex))+geom_bar()  
ggplot(titanic3, aes(as.factor(Pclass), fill=as.factor(sex)))+geom_bar()
```

Data Analytics



Male gender have higher contribution in P class as compared to Female.

6. Expected Format

Data Analytics

1. R file should be submitted where applicable.
2. R file should be in PDF or in .r format
3. Proper screenshots of the outputs should be submitted as well
4. The r codes, if submitted in any other format, will be subjected to deduction in marks

Note: Your solution will not be entertained if it is any other format, e.g., .zip, .doc, .rtf etc.

7. Approximate Time to Complete Task

30 mins.