**Internship Program on Data Science**

**Using R-Programming**

**R – Hackathon**

**Done By**

**M.Sham Ganesh**

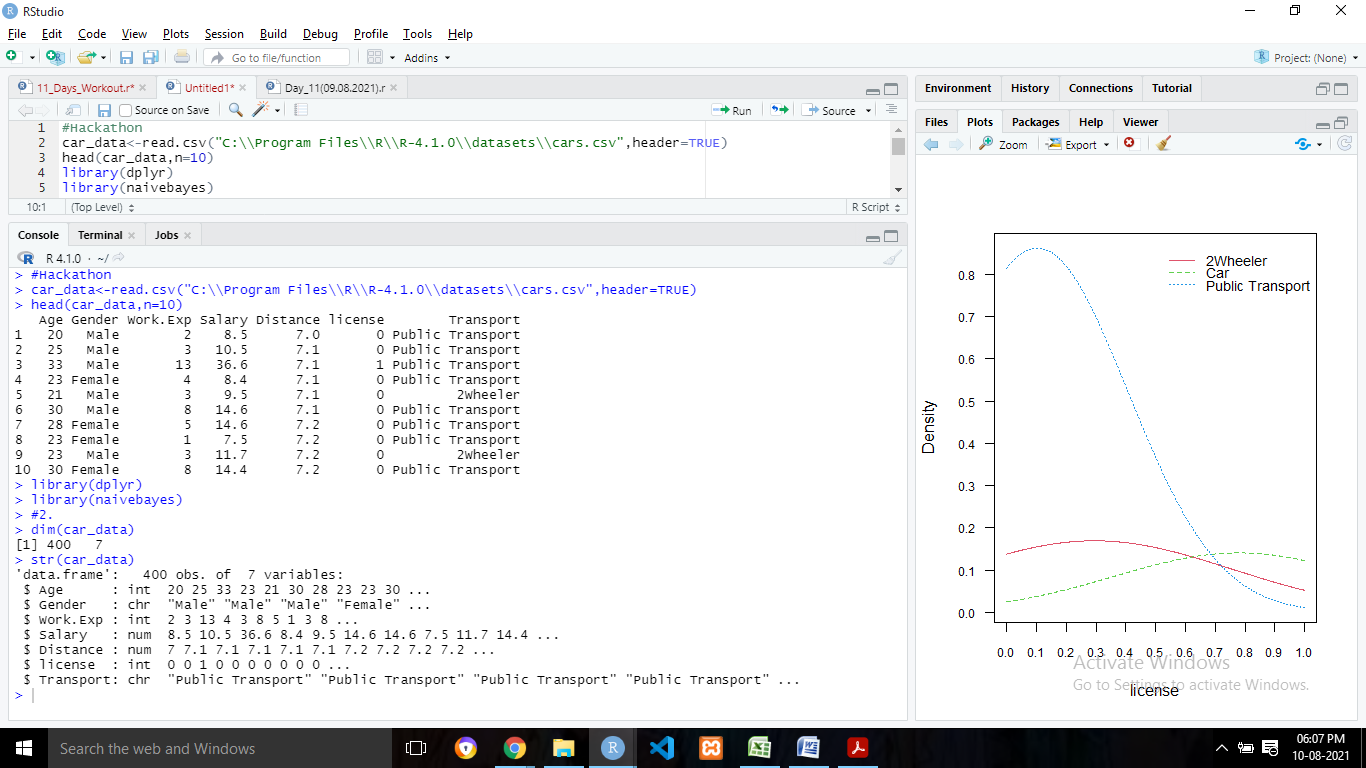
**2. With the car data, find the following - Attach the R code with Output**

**a. Number of records and columns**

dim(car\_data)

**b. Structure of dataset**

str(car\_data)



**c. Summary of the dataset**

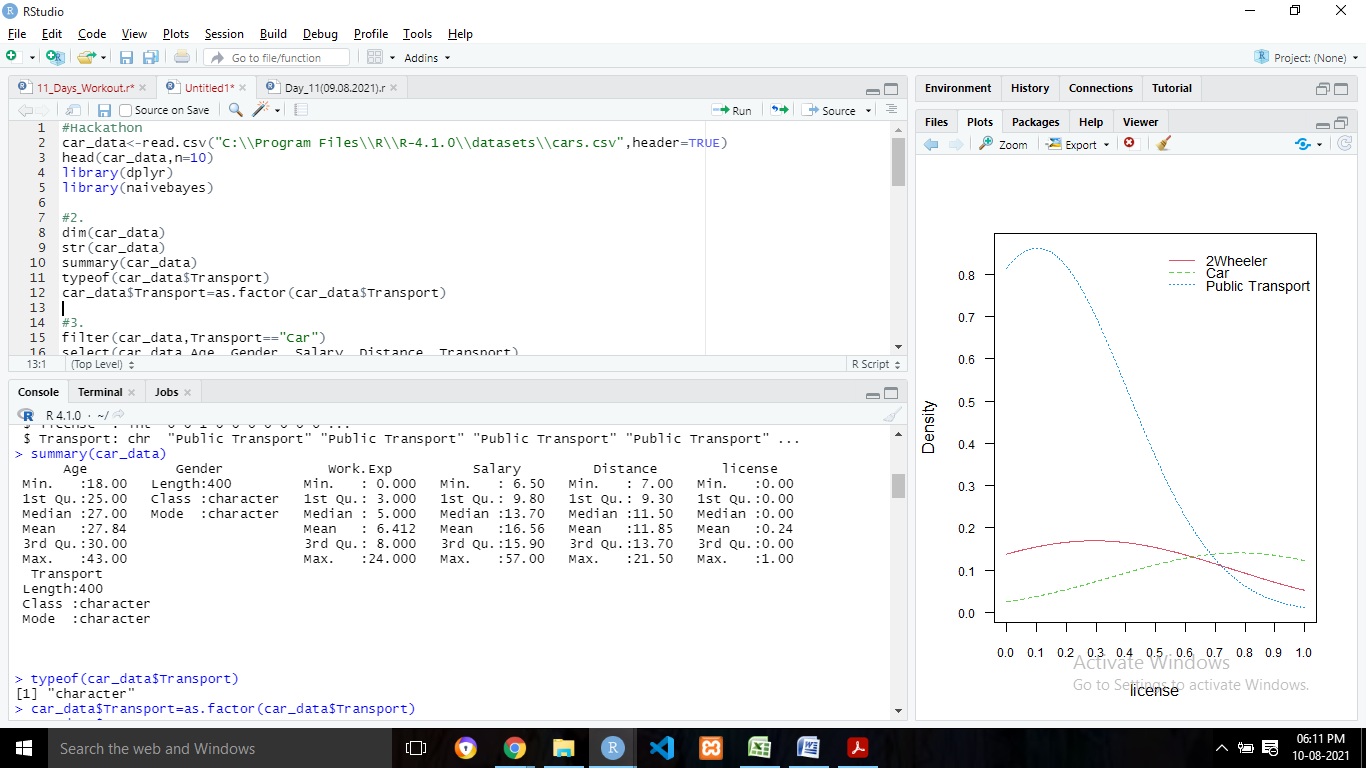
summary(car\_data)

**d. Data type of attribute “Transport”**

typeof(car\_data$Transport)

**e. Convert attribute “Transport” to categorical data type**

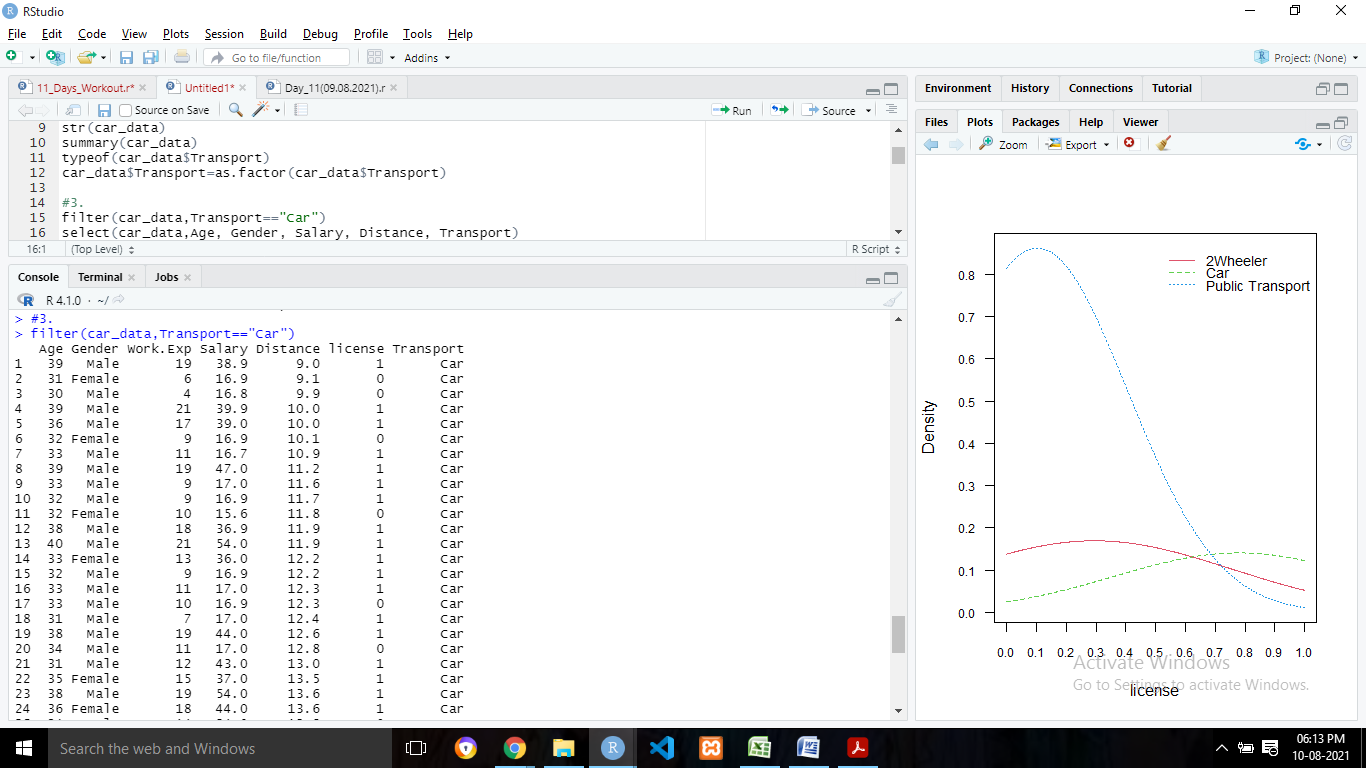
car\_data$Transport=as.factor(car\_data$Transport)



**3. Using car Data, generate the following output - Attach the R code and output**

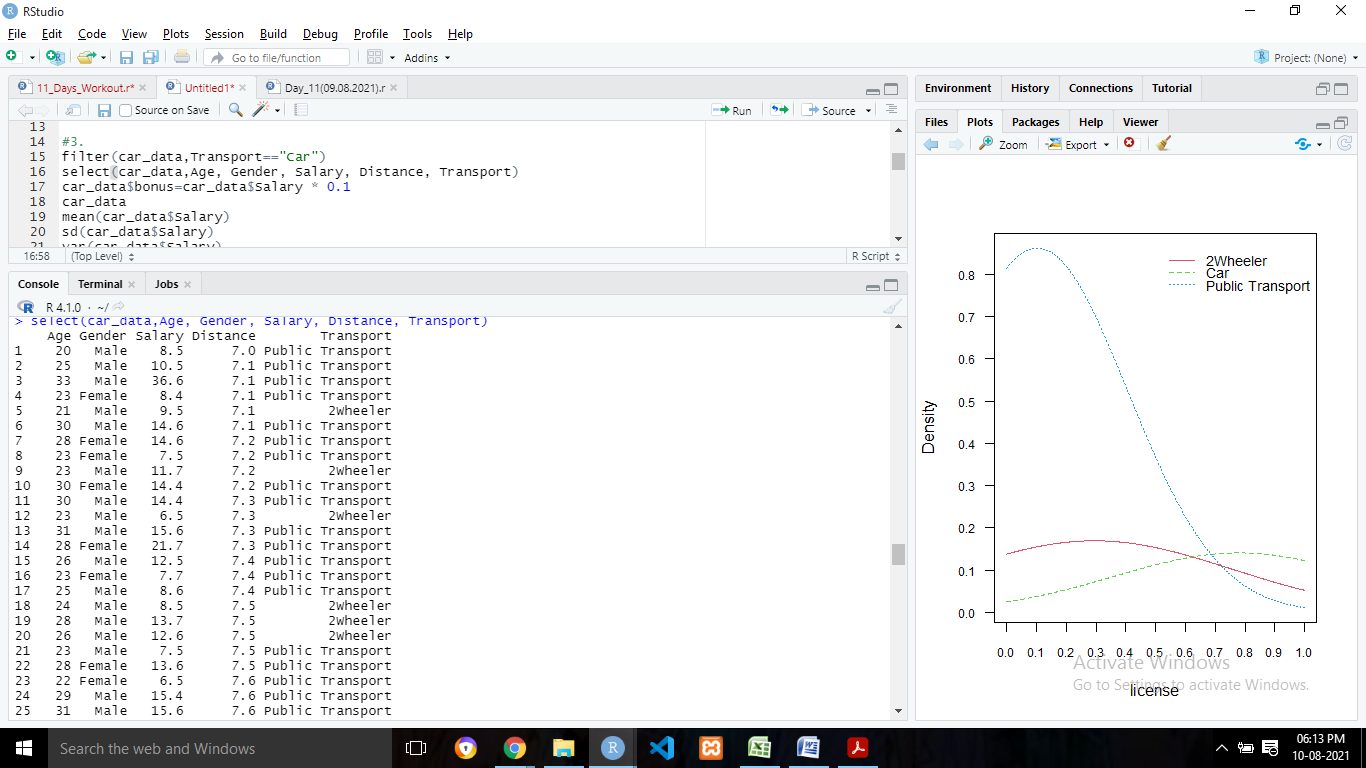
**a. List the records with Transport = “car”**

filter(car\_data,Transport=="Car")



**b. List the datasets with Age, Gender, Salary, Distance, Transport**

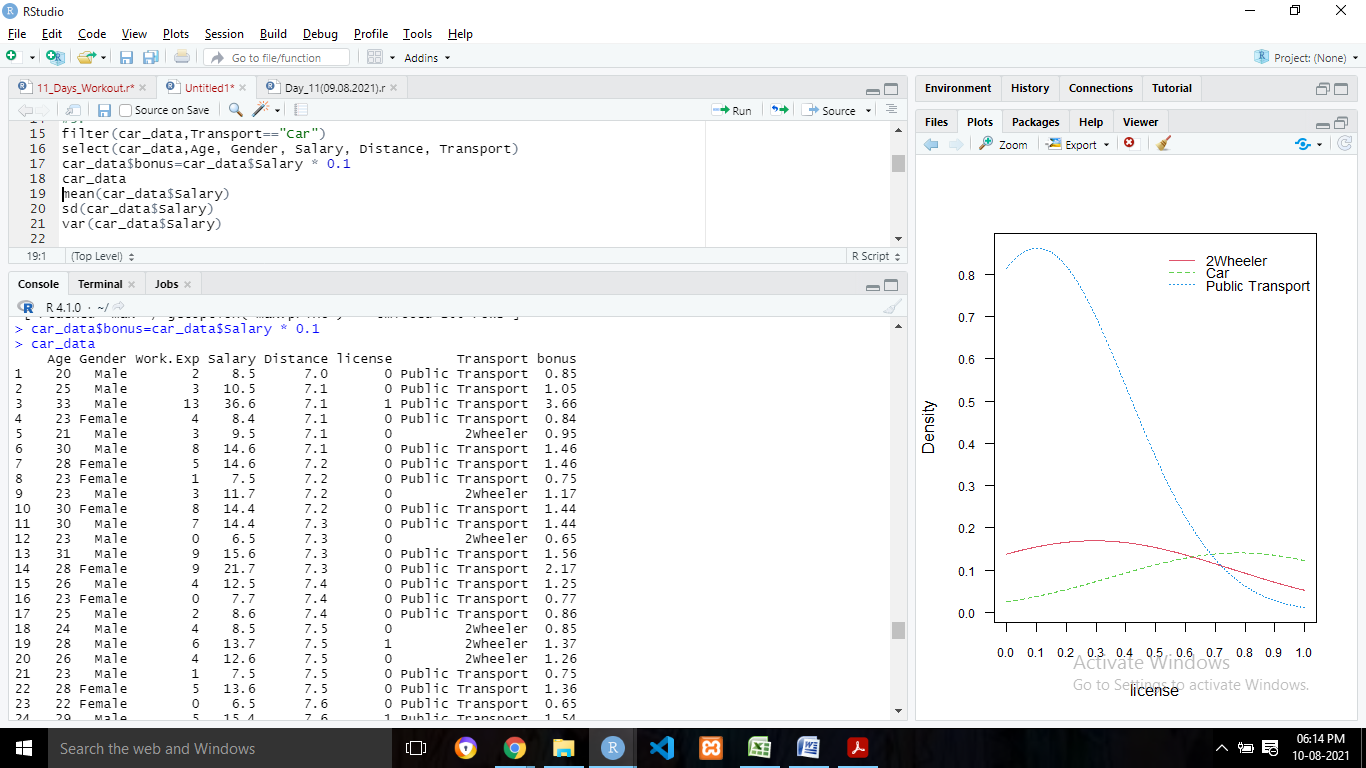
select(car\_data,Age, Gender, Salary, Distance, Transport)

****

**c. Add a new attribute and display the output**

car\_data$bonus=car\_data$Salary \* 0.1

car\_data

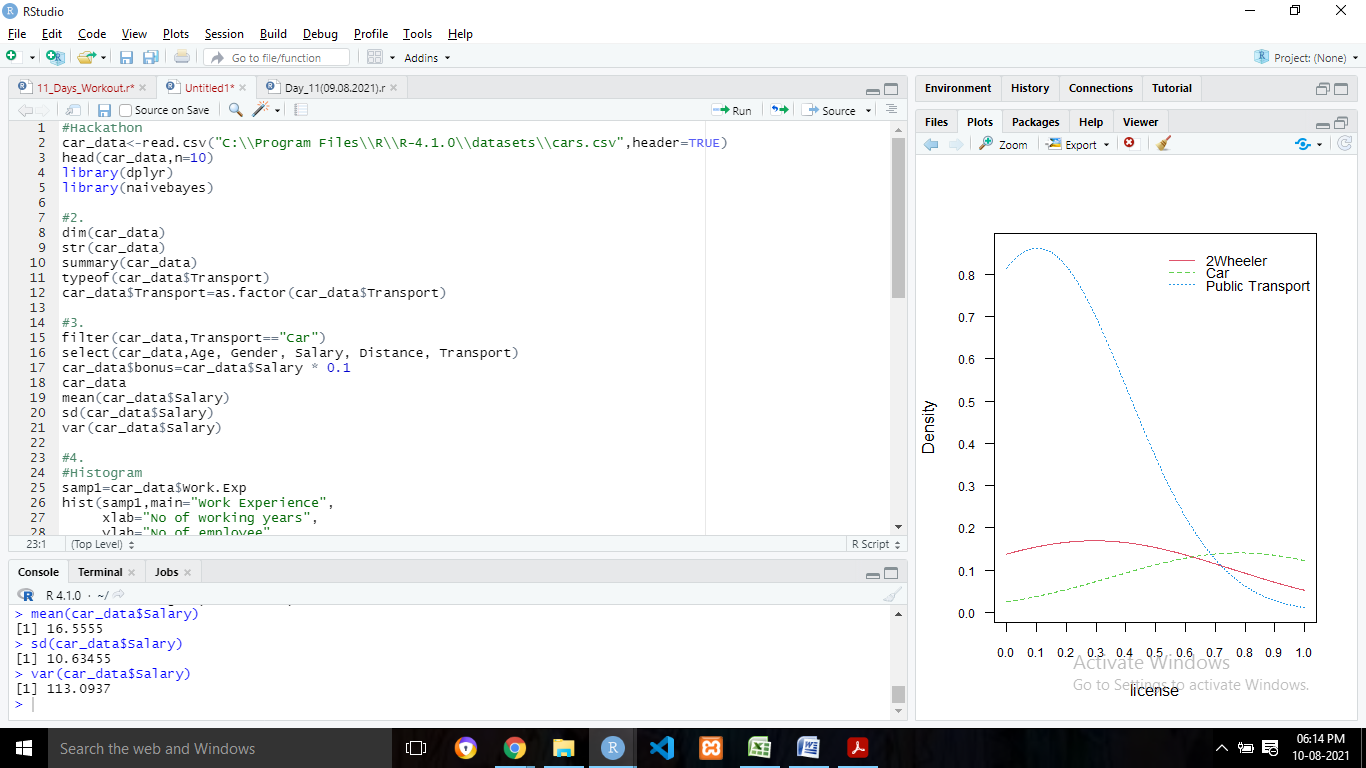
****

**d. Display the mean, standard deviation and variance of the Salary attribute**

mean(car\_data$Salary)

sd(car\_data$Salary)

var(car\_data$Salary)



**4. Using Car Dataset, apply the following Visualization Techniques to the appropriate attributes – Attach the R code and output**

**a. Histogram**

#Histogram

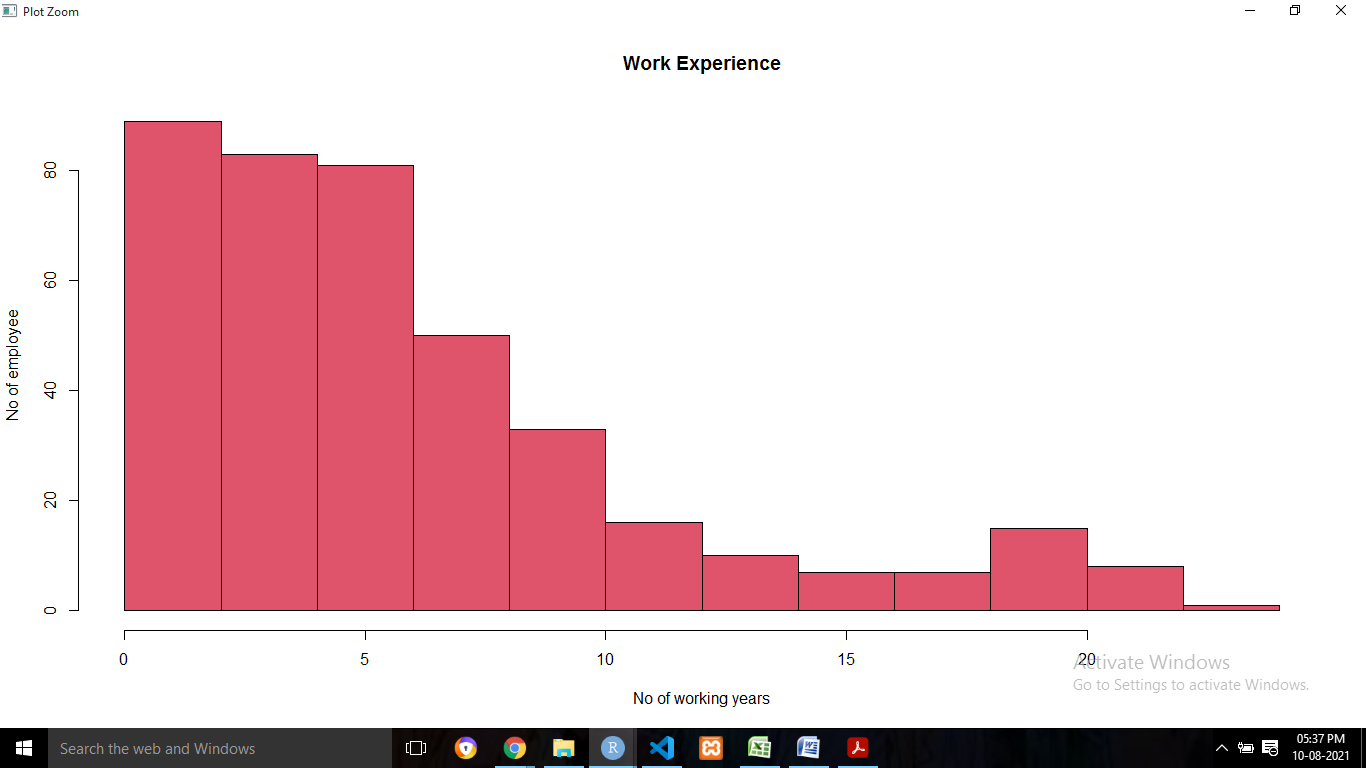
samp1=car\_data$Work.Exp

hist(samp1,main="Work Experience",

xlab="No of working years",

ylab="No of employee",

col=2)



**b. Bar plot**

#Bar plot

samp2=table(car\_data$Gender)

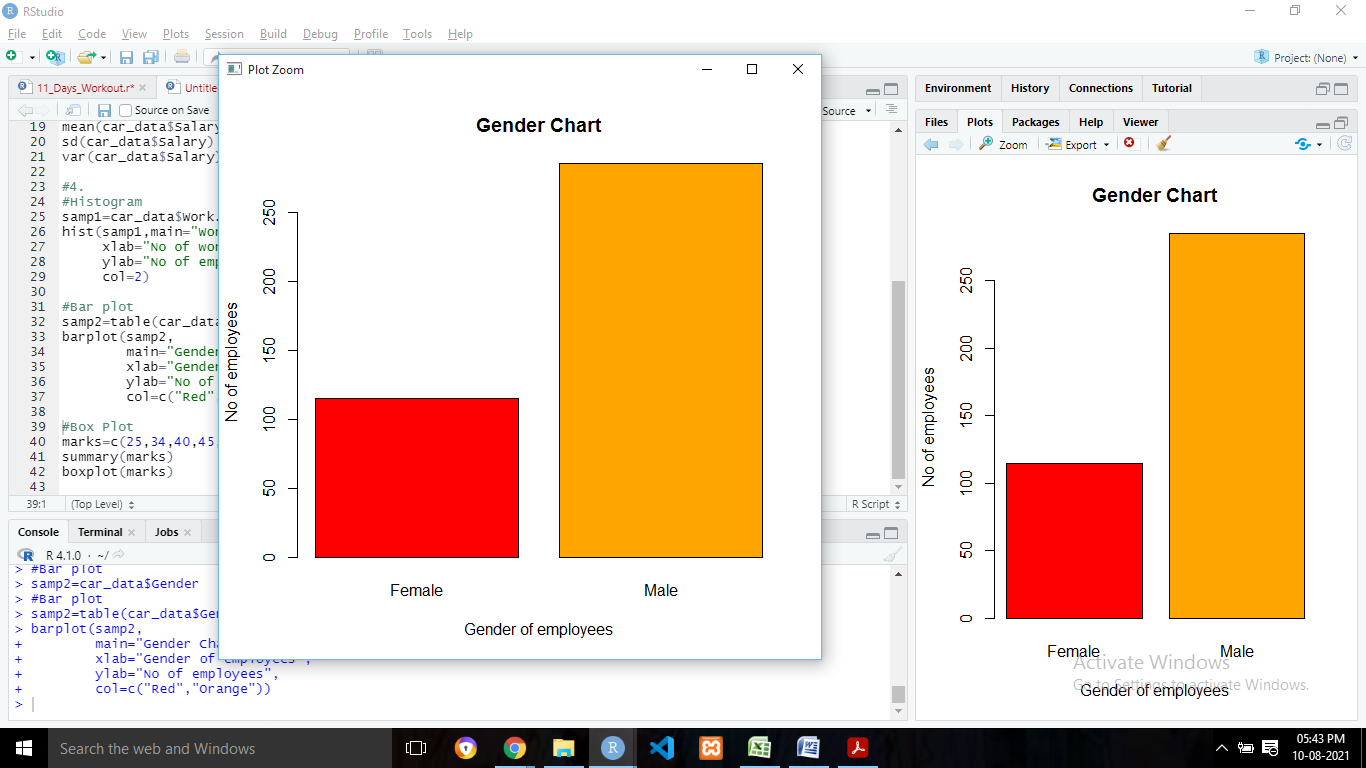
barplot(samp2,

main="Gender Chart",

xlab="Gender of employees",

ylab="No of employees",

col=c("Red","Orange"))



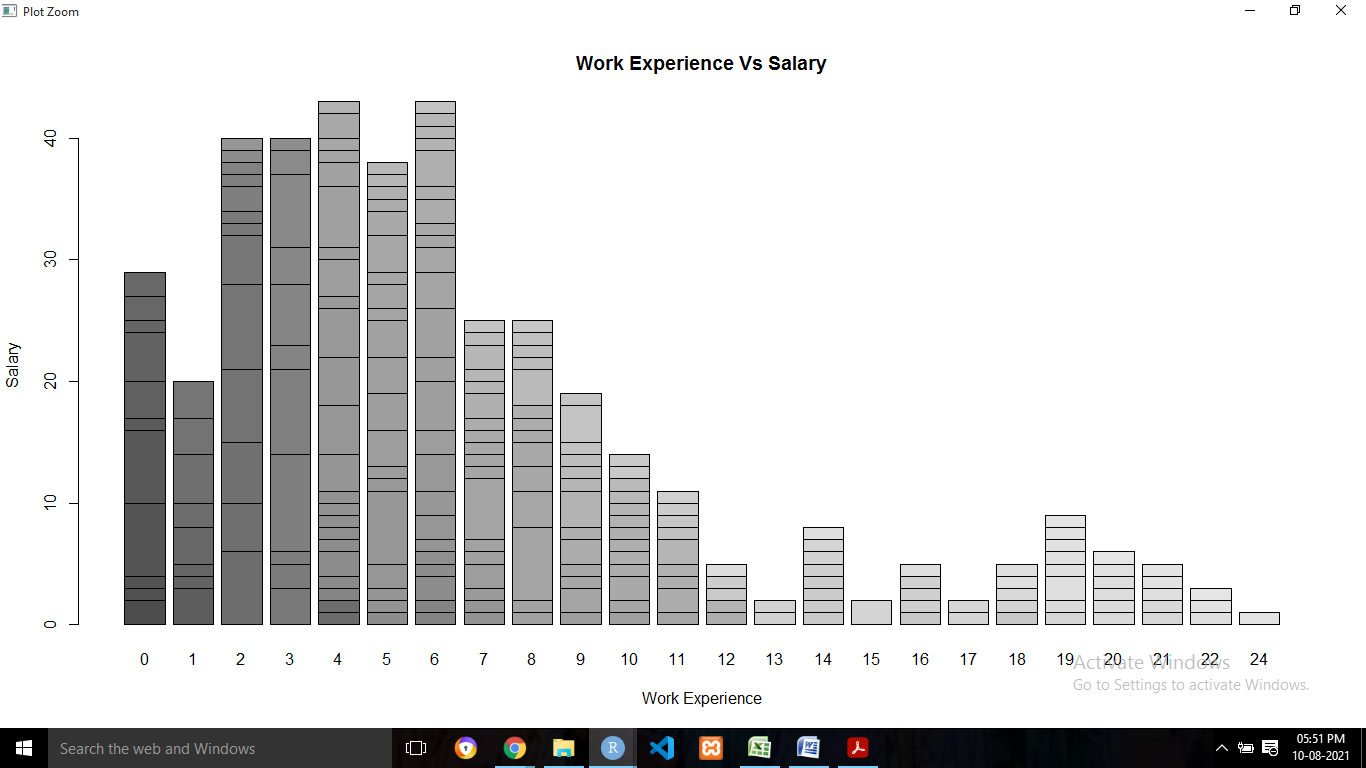
samp3=table(car\_data$Salary,car\_data$Work.Exp)

barplot(samp3,

main="Work Experience Vs Salary",

xlab="Work Experience",

ylab="Salary")

****

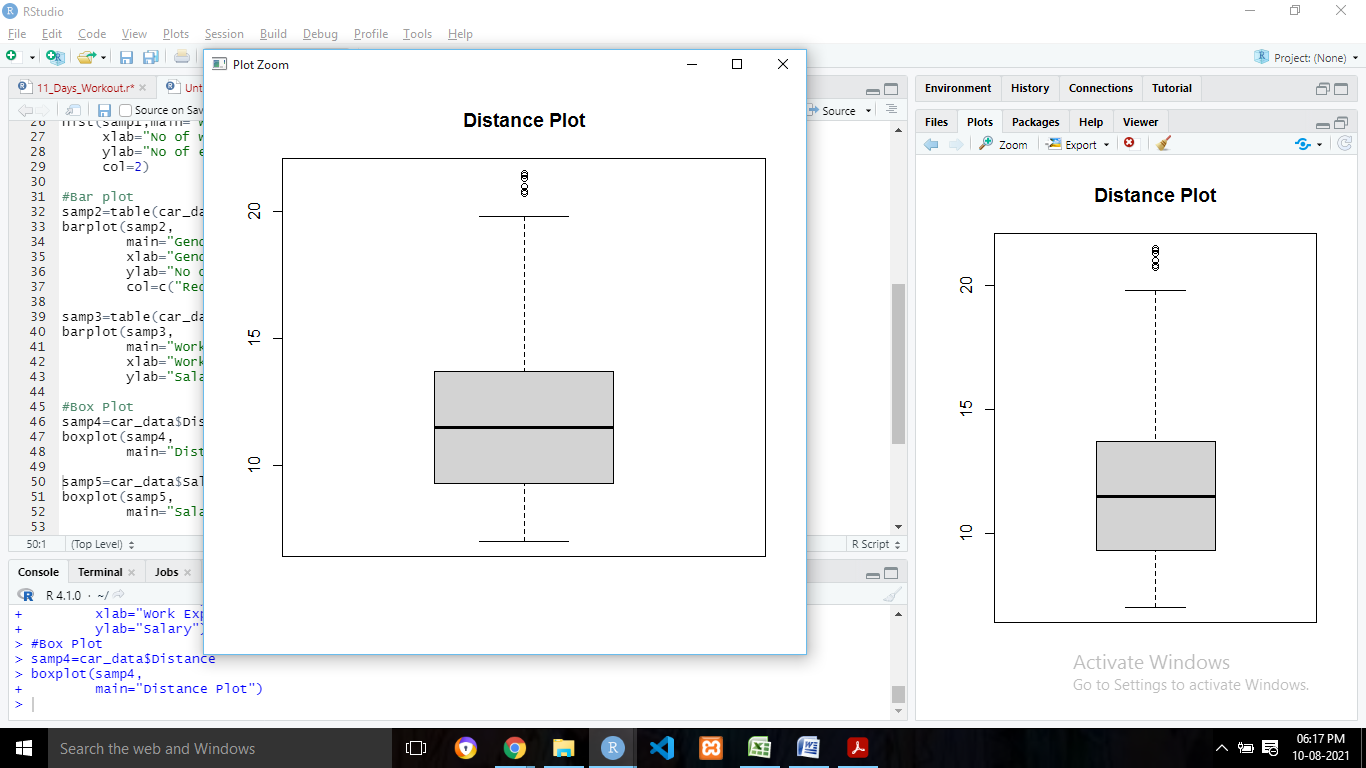
**c. Box plot**

#Box Plot

samp4=car\_data$Distance

boxplot(samp4,

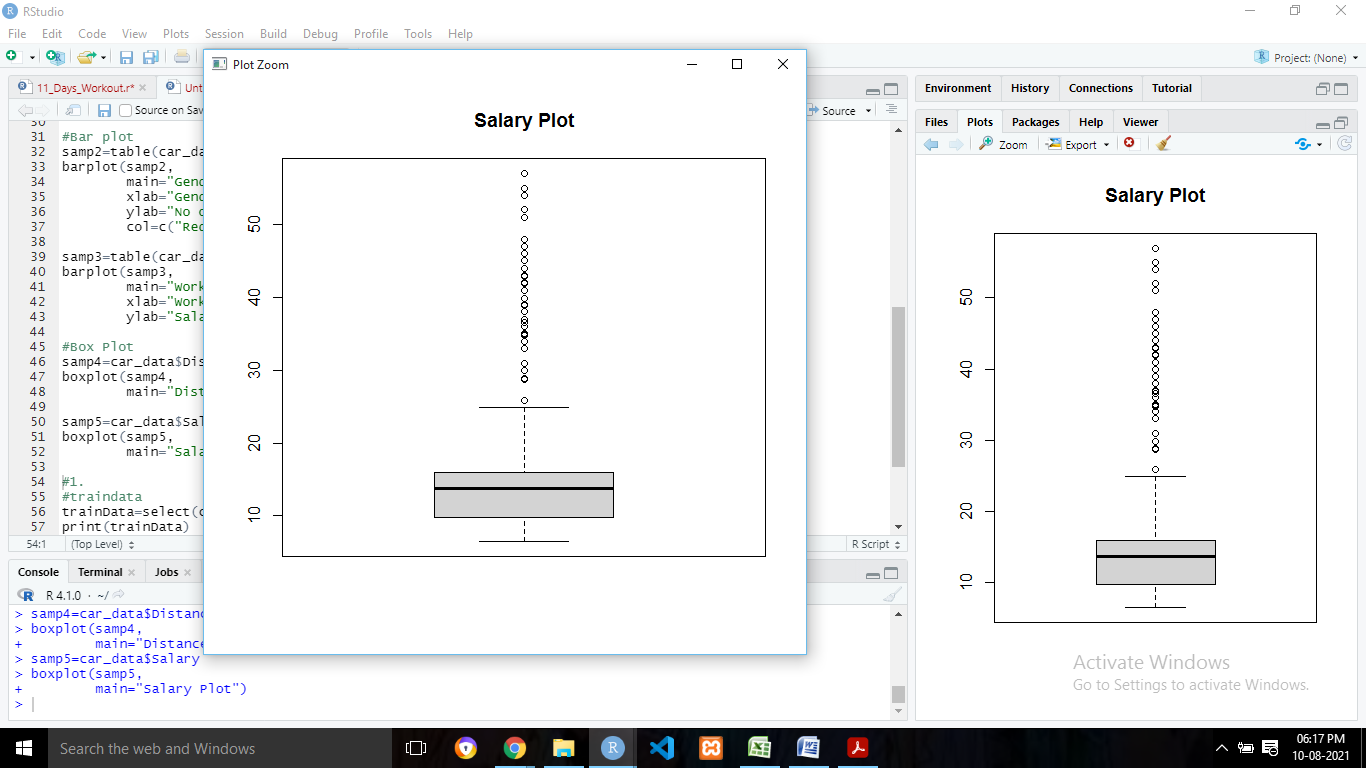
main="Distance Plot")



samp5=car\_data$Salary

boxplot(samp5,

main="Salary Plot")



**1. The attached car data have 400 records, with attributes - Age, Gender, Work Exp, Salary, Distance, License, Transport.**

**Build a model and specify, what would be your prediction regarding their choice of transport for the following Employee. (Attach the R code, model output and predicted output of the following record)**

#traindata

trainData=select(car\_data,Age,Gender,Work.Exp,Salary,Distance,license,Transport)

print(trainData)

naive\_model=naive\_bayes(Transport~ Age+Gender+Work.Exp+Salary+Distance+license,

data=trainData,type="C-classification")

plot(naive\_model)

print(naive\_model)

#predict on test Data

test=data.frame(Age=30,Gender="Male",Work.Exp=5,Salary=20,Distance=7,license=1)

result=predict(naive\_model,test)

print(result)

