SESSION 8: Exploratory Data Analytics

Assignment 1

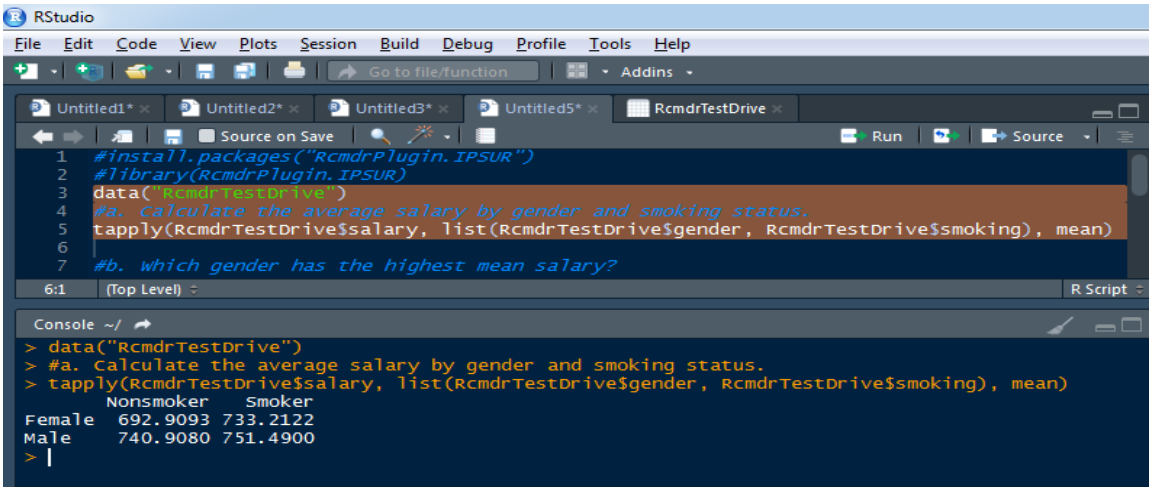
1. Use the package -RcmdrPlugin.IPSUR.
   data(RcmdrTestDrive)
   And perform the below operations:

   a. Calculate the average salary by gender and smoking status.

Answer:

tapply(RcmdrTestDrive$salary, list(RcmdrTestDrive$gender, RcmdrTestDrive$smoking), mean)

Output:



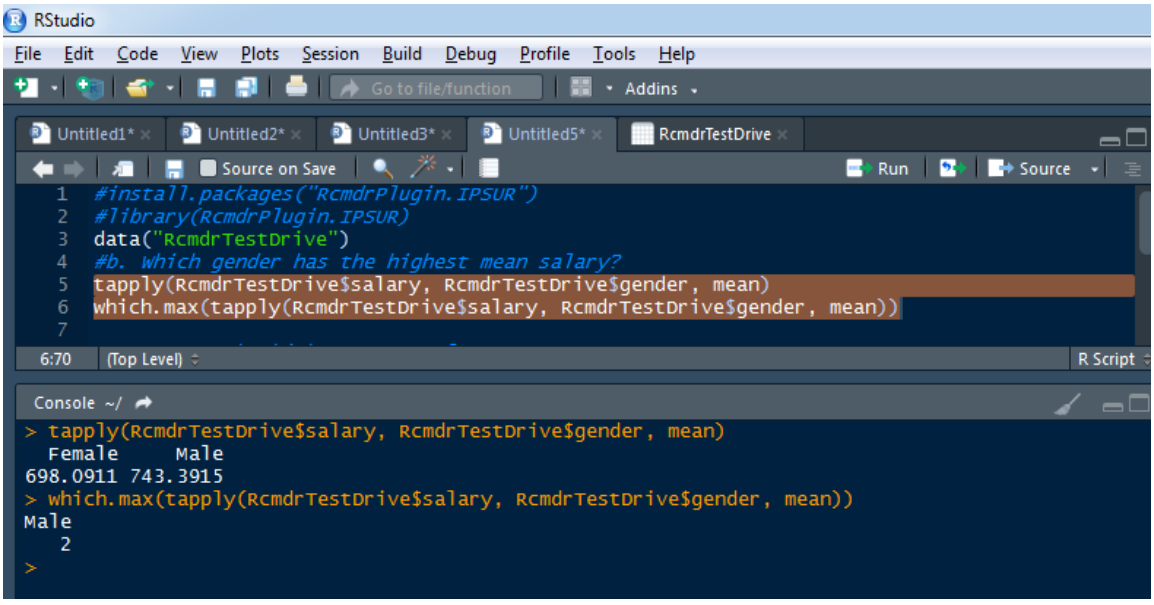
   b. Which gender has the highest mean salary?

Answer:

tapply(RcmdrTestDrive$salary, RcmdrTestDrive$gender, mean)

which.max(tapply(RcmdrTestDrive$salary, RcmdrTestDrive$gender, mean))

Output:

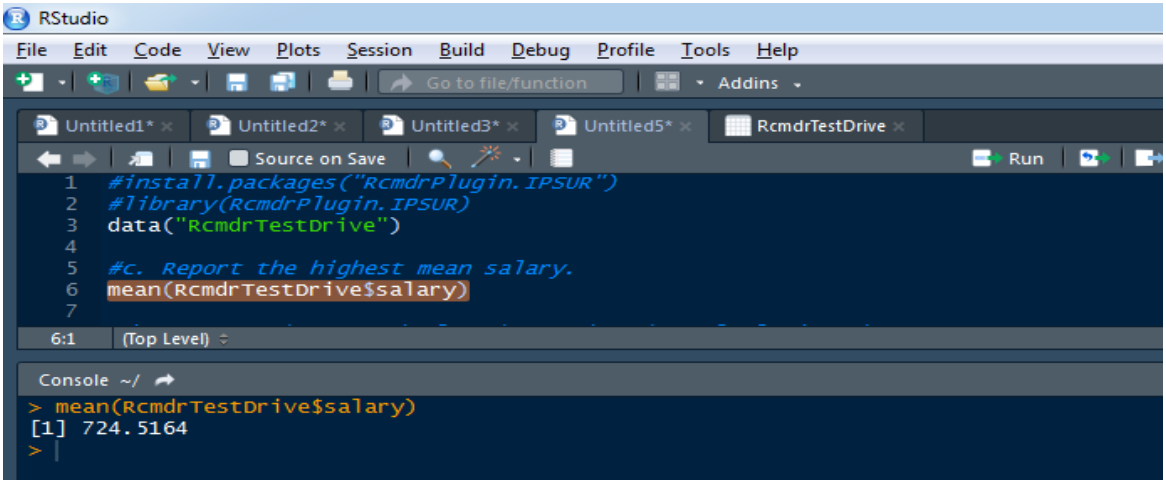


c. Report the highest mean salary.

Answer:

mean(RcmdrTestDrive$salary)

Output:



d. Compare the spreads for the genders by calculating the standard deviation of salary by gender.

Answer:

tapply(RcmdrTestDrive$salary, RcmdrTestDrive$gender, sd)

boxplot(salary~gender, data=RcmdrTestDrive, main="Salary v/s Gender", xlab="Gender", ylab="Salary")

Output: