Session-8-Assignment-1

Aditya Mulay

July 4, 2018

#1. Use the package RcmdrPlugin.IPSUR. data(RcmdrTestDrive) and perform the below operations:  
library(RcmdrPlugin.IPSUR)  
head(RcmdrTestDrive)

## order smoking gender race before after salary reduction parking  
## 1 1 Nonsmoker Female Caucasian 72.6 75.2 618.65 9 2  
## 2 2 Nonsmoker Male AfricanAmer 75.3 73.2 544.56 62 1  
## 3 3 Nonsmoker Female Caucasian 75.5 74.5 550.24 19 4  
## 4 4 Nonsmoker Female Caucasian 71.3 74.6 616.16 30 1  
## 5 5 Nonsmoker Female Hispanic 74.3 73.8 543.39 105 1  
## 6 6 Nonsmoker Male Caucasian 73.0 73.6 692.09 43 1

# a. Calculate the average salary by gender and smoking status.  
gender <- tapply(RcmdrTestDrive$salary, RcmdrTestDrive$gender, mean)  
print(gender)

## Female Male   
## 698.0911 743.3915

smoking <- tapply(RcmdrTestDrive$salary, RcmdrTestDrive$smoking, mean)  
print(smoking)

## Nonsmoker Smoker   
## 719.3792 746.3494

# b. Which gender has the highest mean salary?  
print(gender)

## Female Male   
## 698.0911 743.3915

print("Male has the highest mean salary")

## [1] "Male has the highest mean salary"

# c. Report the highest mean salary.  
max(RcmdrTestDrive$salary)

## [1] 1156.16

# d. Compare the spreads for the genders by calculating the standard deviation of salary by gender.  
tapply(RcmdrTestDrive$salary, RcmdrTestDrive$gender, sd)

## Female Male   
## 130.7053 158.5423

boxplot(salary~gender,data= RcmdrTestDrive,main="salary versus gender",xlab="gender",ylab="salary",col=topo.colors(2))

