

Problem Statement

i. Use the package -RcmdrPlugin.IPSUR.

data(RcmdrTestDrive)

and perform the below operations:

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

Answer :

The R-script for the given problem is as follows:

```
library(Rcmdr)

library(RcmdrPlugin.IPSUR)

data(RcmdrTestDrive)

RcmdrTestDrive

library(dplyr)

str(RcmdrTestDrive)

#Data <- RcmdrTestDrive

AvgSalary <- RcmdrTestDrive%>%group_by(gender, smoking)%>%

select(smoking, gender, salary)%>%summarise(mean(salary))

AvgSalary <- as.data.frame(AvgSalary)

AvgSalary$meansalary <- AvgSalary$`mean(salary)`

AvgSalary

stripchart(meansalary ~ gender, vertical=TRUE, method="jitter",

          ylab="meansalary", data=AvgSalary)
```

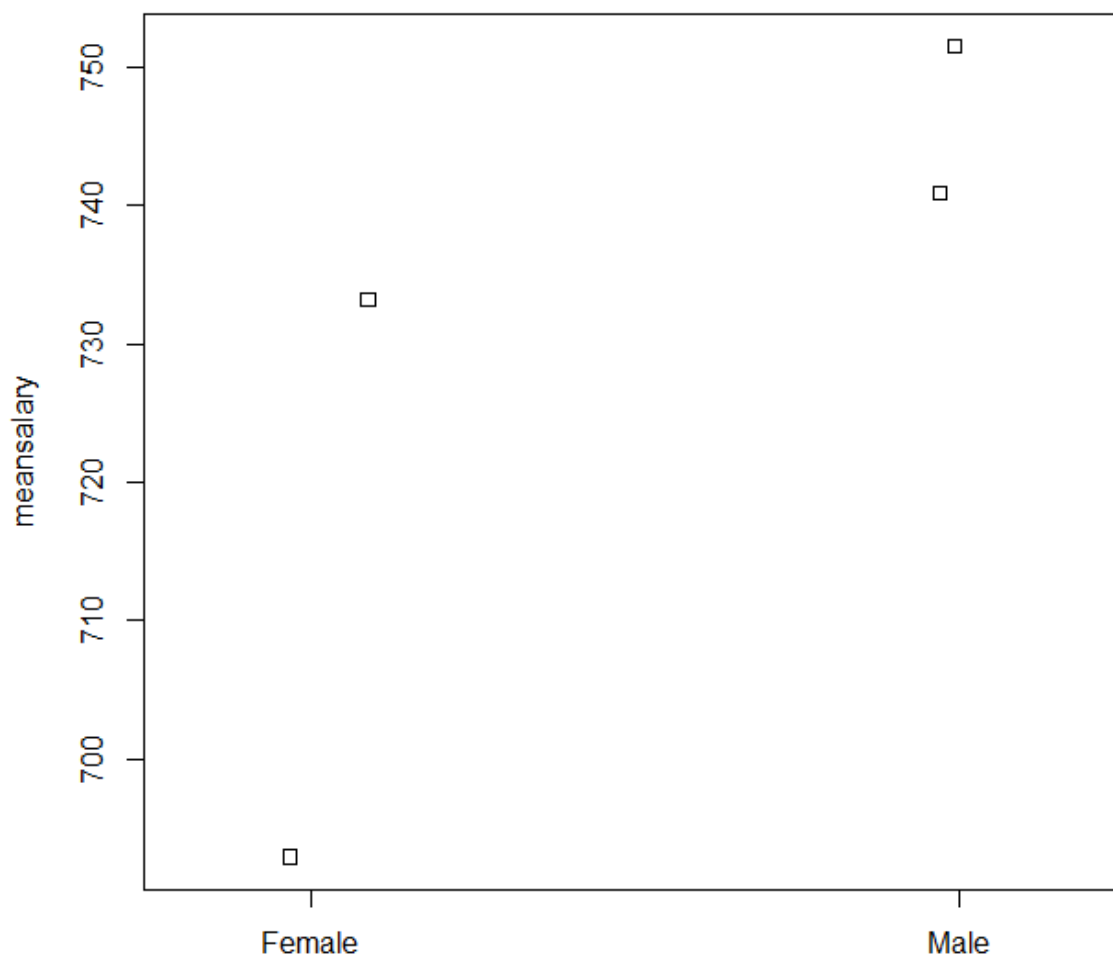
The output of the R-Script (from Console window) is given as follows:

```
> library(Rcmdr)
> library(RcmdrPlugin.IPSUR)
> data(RcmdrTestDrive)
> RcmdrTestDrive
```

```
> library(dplyr)
> str(RcmdrTestDrive)

> AvgSalary <- RcmdrTestDrive%>%group_by(gender, smoking)%>%
+ select(smoking, gender, salary)%>%summarise(mean(salary))
> AvgSalary <- as.data.frame(AvgSalary)
> AvgSalary$meansalary <- AvgSalary$`mean(salary)`
> AvgSalary

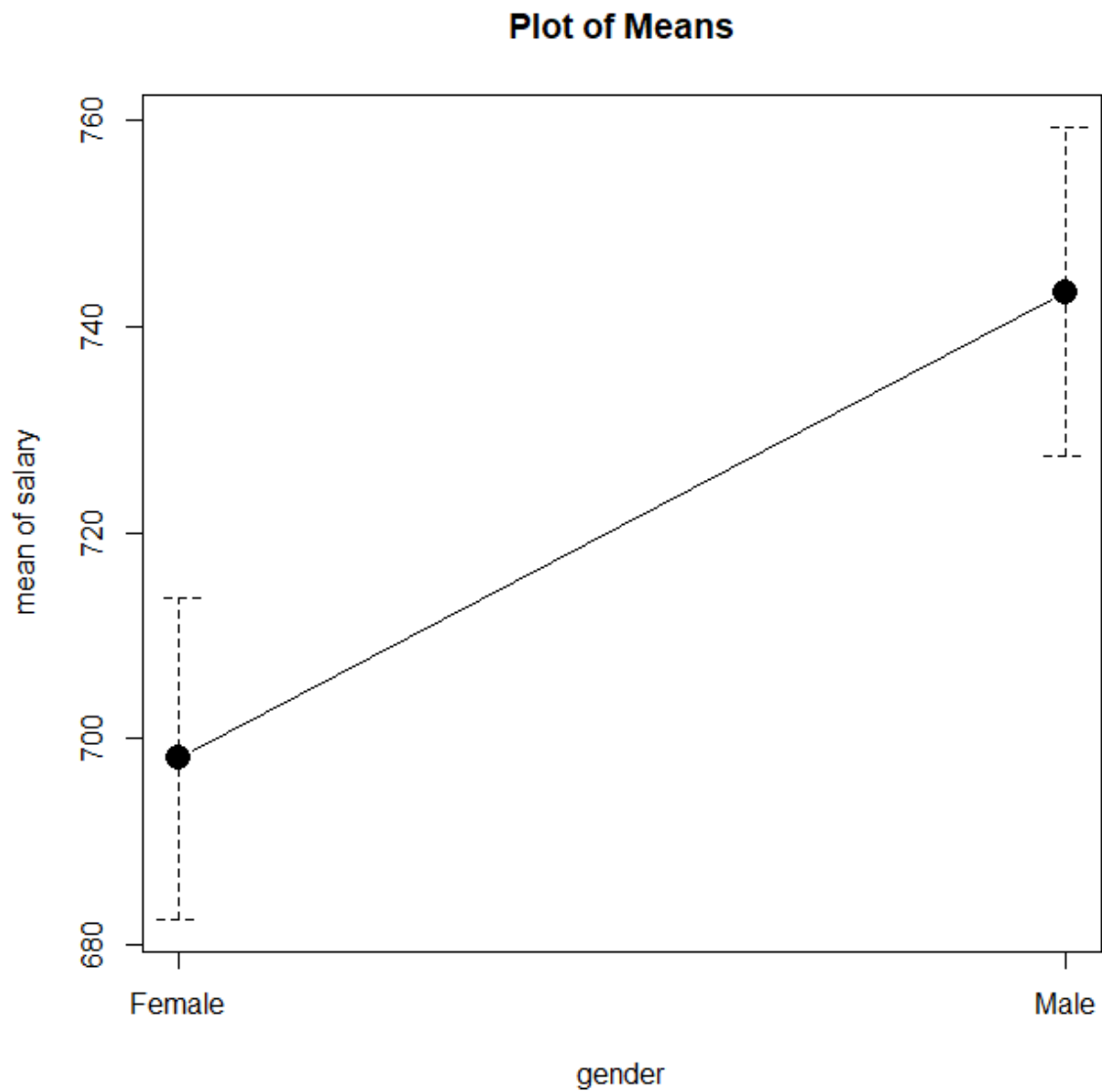
> stripchart(meansalary ~ gender, vertical=TRUE, method="jitter",
+ ylab="meansalary", data=AvgSalary)
```



b. Which gender has the highest mean salary?

Answer :

```
> with(RcmdrTestDrive, plotMeans(salary, gender, error.bars="se"))
```



c. Report the highest mean salary.

Answer :

```
meansalary <- as.data.frame(RcmdrTestDrive%>%group_by(gender)%>%
```

```

select(gender,salary)%>%summarise(mean(salary))

meansalary$meansalary <- meansalary$`mean(salary)`

meansalary

meansalary[which.max(meansalary$meansalary),]

bp <- barplot(meansalary$meansalary, xlab = names(meansalary),

  ylab = "Mean Salary",

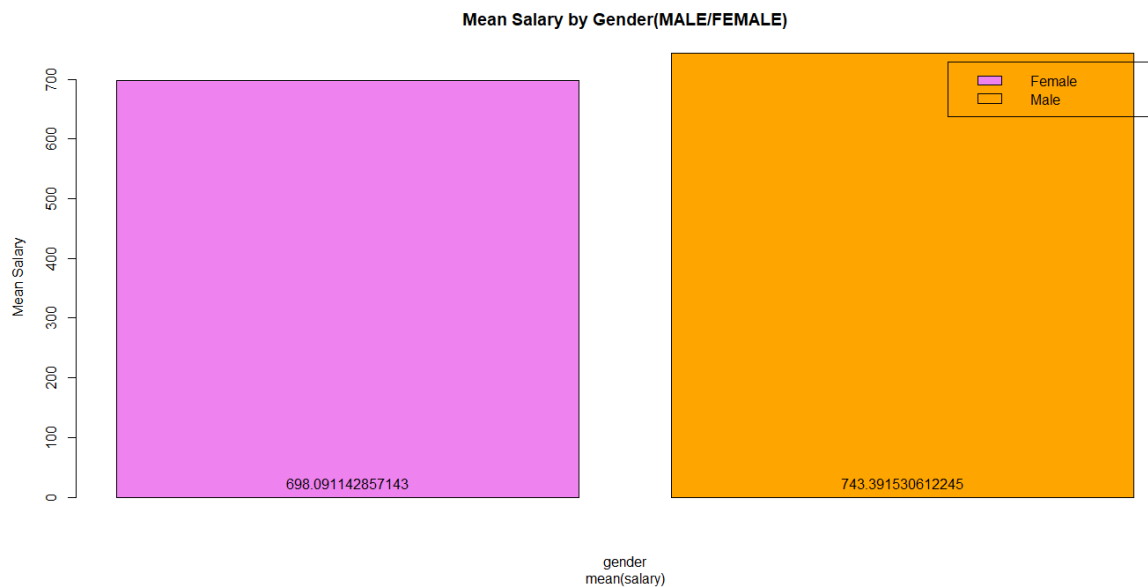
  main = "Mean Salary by Gender(MALE/FEMALE)",

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

  legend = meansalary$gender)

text(bp, 0, meansalary$meansalary, cex = 1, pos = 3)

```



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

Answer :

```

str(RcmdrTestDrive)

MaleSalary <- RcmdrTestDrive%>%select(gender, salary)%>%filter(gender == "Male")

FemaleSalary <- RcmdrTestDrive%>%select(gender, salary)%>%filter(gender == "Female")

```

```
par(mfrow = c(1,2))
```

```
M <- density(MaleSalary$salary)
```

```
plot(M, type="l", main="Male Salary Distribution", col = "Red")
```

```
N <- density(FemaleSalary$salary)
```

```
plot(N, type = "l", main = "Female Salary Distribution", col = "Blue")
```

