Meher Shrishti Nigam - 20BRS1193

Foundations Of Data Analytics

LAB EXERCISE 8

CODE:

```
# L7+L8
# Meher Shrishti Nigam
# CSE AI + Robotics
# 20BRS1193
# LAB 8

# Use the newsurvey data obtained by cleaning 'na' values in survey data of MASS package to do the following:

rm(list=ls())

install.packages('ggplot2')

library(ggplot2)

library(dplyr)

library(MASS)

df <- na.omit(survey)

# 1. Plot a bar graph for the number of male and female participants in the survey. Provide

# the title as "Male and Female participants", y-axis label as "frequency" and specify

# the colours for the bars.

sex <- table(df$Sex)

barplot(sex, main = 'Male and Female Participants', names.arg = c('Male','Female'), ylab = 'Frequency')

# 2. Plot a bar graph for the number of left handers and right handers in the survey.

# Provide the title as "Left Handers and Right Handers", y-axis label as "count" and

# specify the colours for the bars.

handedness <- table(df$W.Hnd)

barplot(handedness, main='Left Handers and Right Handers', names.arg = c('Left', 'Right'), ylab = 'Count')
```

 # 3. Plot the distribution between male left handers and female left handers using bar

 # chart. Provide the title as “Female Left Handers and Male Left Handers , y-axis label

 # as “count” and specify the colours for the bars.

```
lefthanded <- table((filter(df, W.Hnd == 'Left'))$Sex)

barplot(lefthanded, main='Female Left Handers and Male Left Handers', names.arg = c('Male', 'Female'), col = c("red", "green"), ylab = 'Count')
```

 # 4. Draw the distribution of smoking habits of male left handers using pie chart.

```
male.left.handers <- table((filter(df, W.Hnd == 'Left', Sex == 'Male'))$Smoke)

labels <- c('Never', 'Regul', 'Heavy', 'Occas')

pie(male.left.handers, labels, radius = 1)
```

 # 5. Draw the histogram of age distribution with the title as ‘Age distribution’ and xlabel

 # as ‘Age range’ and ylabel as ‘frequency’.

```
ggplot(df,aes(x=Age))+geom_histogram()+labs(title='Age Distribution', x='Age Range', y='frequency')
```

 # 6. Plot the density distribution of age distribution with title as ‘Age distribution’ and

 # xlabel as ‘Age range’ and ylabel as ‘density’.

```
ggplot(df,aes(x=Age))+geom_density()+labs(title='Age Distribution', x='Age Range', y='density')
```

 # 7. Create a suitable grid for projecting the multiple charts obtained earlier.

```
par(mfrow=c(2,2),mar=c(2,5,2,1),las=1, bty='n')

barplot(sex, main = 'Male and Female Participants', names.arg = c('Male','Female'), ylab = 'Frequency')

barplot(handedness, main='Left Handers and Right Handers', names.arg = c('Left', 'Right'), ylab = 'Count')

barplot(lefthanded, main='Female Left Handers and Male Left Handers', names.arg = c('Male', 'Female'), col = c("red", "green"), ylab = 'Count')

pie(male.left.handers, labels, radius = 1)
```

 # 8. Reveal the relationship between the age and writing hand span using scatter plot.

```
plot(df$Age, df$Wr.Hnd, xlab='Age', ylab='Writing Hand Span', main='Relationship between Age and Writing Hand Span')
```

 # 9. Plot the relationship between age, height and writing hand span in a single chart.

```
library(lattice)

dev.off()

splom(df[c(2,10,12)])
```

 # 10. Plot the relationship between height and writing hand span

```
par(mar=c(3,3,3,3))

plot(df$Height, df$Wr.Hnd, xlab='Height', ylab='Writing Hand Span', main='Rel btw Height and Writing Hand Span')
```

 # 11. Plot the relationship between height and writing hand span based on gender and left and right handers.

```
femright <- data.frame(filter(df, Sex == 'Female', W.Hnd == 'Right'))

femright$hand.gender <- "Female Right Handed"

femleft <- data.frame(filter(df, Sex == 'Female', W.Hnd == 'Left'))

femleft$hand.gender <- "Female Left Handed"

maleright <- data.frame(filter(df, Sex == 'Male', W.Hnd == 'Right'))

maleright$hand.gender <- "Male Right Handed"

maleleft <- data.frame(filter(df, Sex == 'Male', W.Hnd == 'Left'))

maleleft$hand.gender <- "Male Left Handed"

df1 <- rbind(maleleft, maleright, femleft, femright)

ggplot(df1,aes(x=Height,y=Wr.Hnd,color=hand.gender))+geom_point()
```

 # 12. Draw the boxplot for pulse rate to analyse the five summary statistics. Provide

 # appropriate title and label.

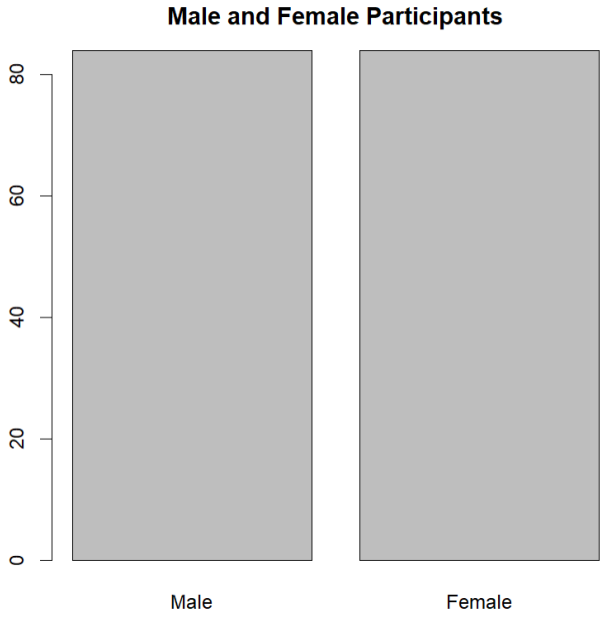
```
boxplot(df$Pulse, xlab='Boxplot', ylab='Pulse', label="Pulse Rate")

summarize(df, mean(Pulse,na.rm=TRUE))

summarize(df, median(Pulse,na.rm=TRUE))

summarize(df, IQR(Pulse,na.rm=TRUE))

summarize(df, sd(Pulse,na.rm=TRUE))

summarize(df, var(Pulse,na.rm=TRUE))
```

OUTPUT:

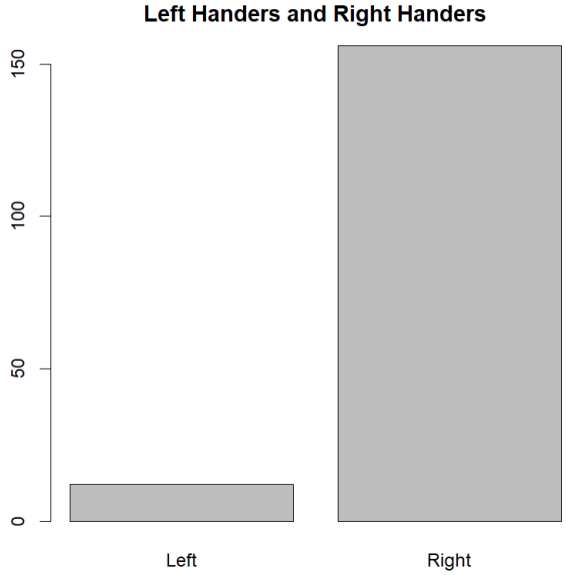
> # Use the newsurvey data obtained by cleaning 'na' values in survey data of MASS package to do the following:

> rm(list=ls())
> library(ggplot2)
> library(dplyr)
> library(MASS)
> df <- na.omit(survey)

Q1) > sex <- table(df$Sex)
> barplot(sex, main = 'Male and Female Participants', names.arg = c('Male','Female'), ylab = 'Frequency')



Q2) > handedness <- table(df$W.Hnd)
> barplot(handedness, main='Left Handers and Right Handers', names.arg = c('Left', 'Right'), ylab = 'Count')



Q3)

> lefthanded <- table((filter(df, W.Hnd == 'Left'))$Sex)

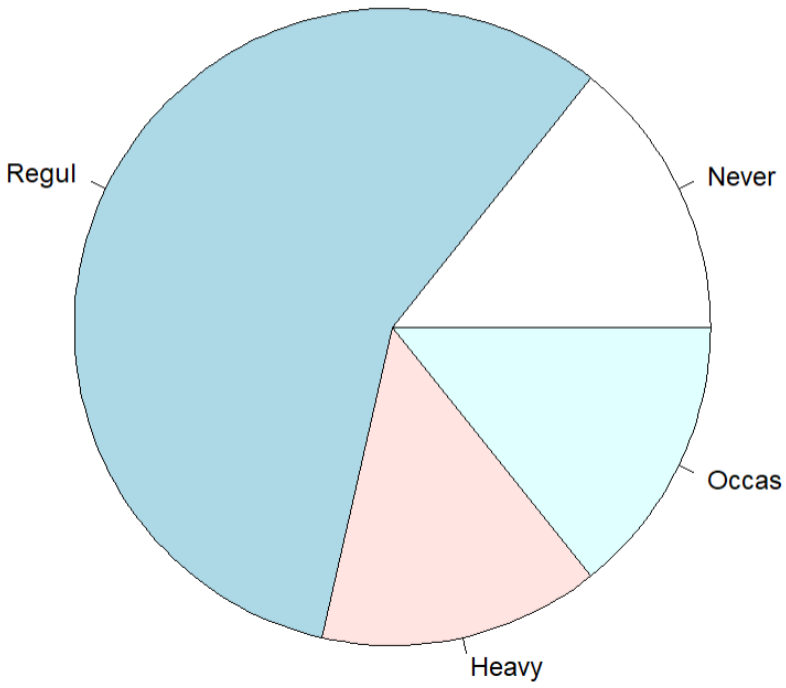
> barplot(lefthanded, main='Female Left Handers and Male Left Handers', names.arg = c('Male', 'Female'), col = c("red", "green"), ylab = 'Count')

Q4)

> male.left.handers <- table((filter(df, W.Hnd == 'Left', Sex == 'Male'))$Smoke)

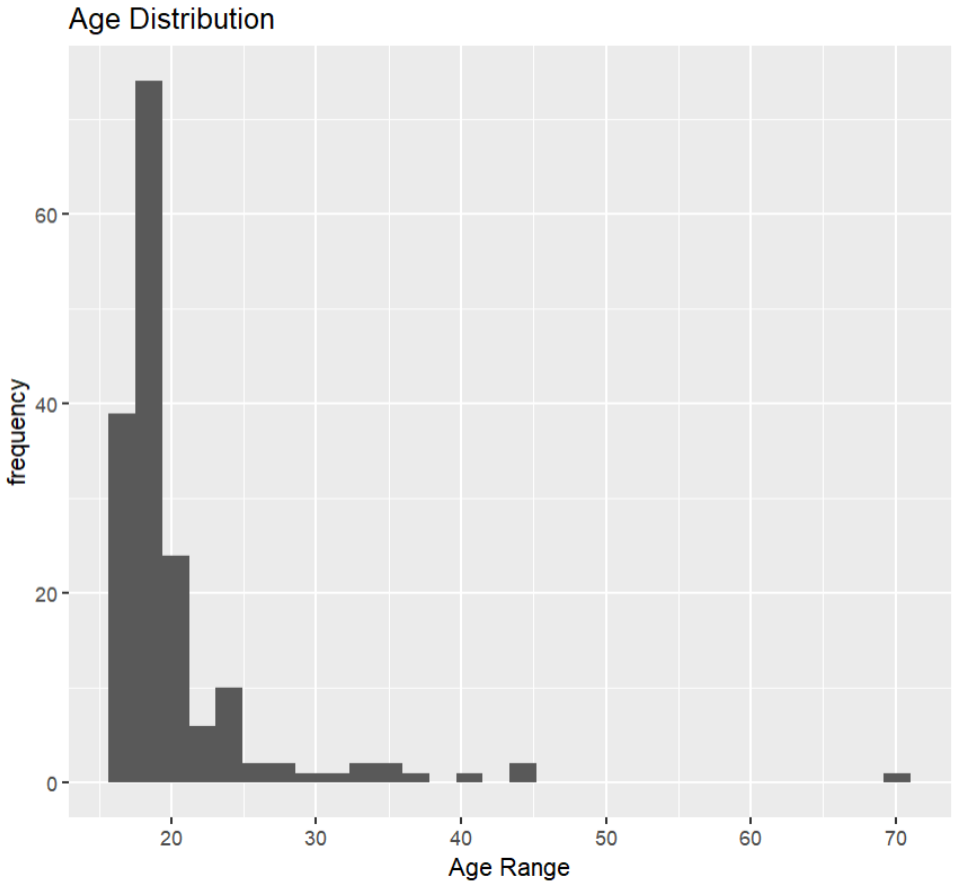
> labels <- c('Never', 'Regul', 'Heavy', 'Occas')

> pie(male.left.handers, labels, radius = 1)



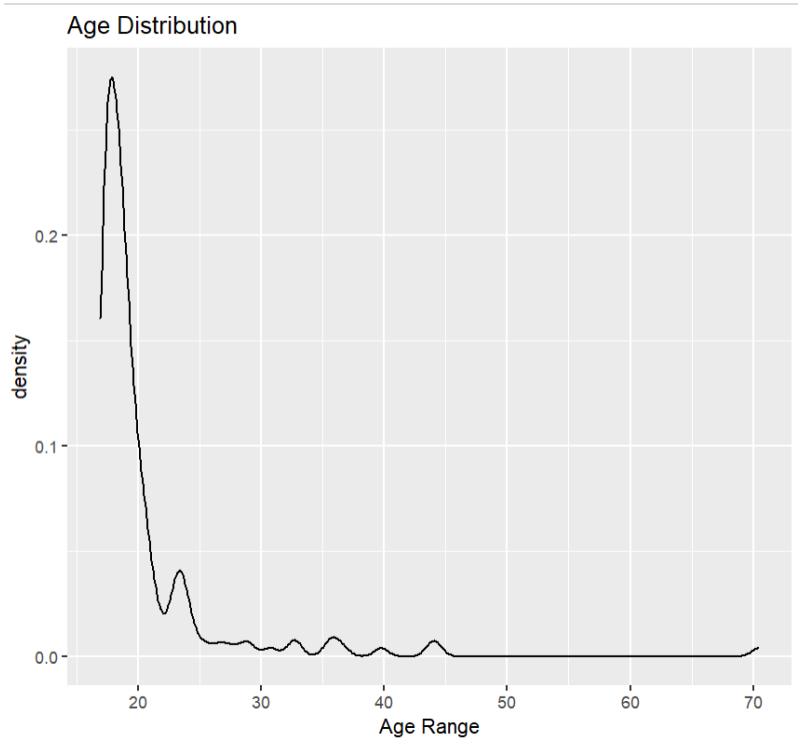
Q5)

> ggplot(df,aes(x=Age))+geom_histogram()+labs(title='Age Distribution', x='Age Range', y='frequency')



Q6)

> ggplot(df,aes(x=Age))+geom_density()+labs(title='Age Distribution', x='Age Range', y='density')



Q7)

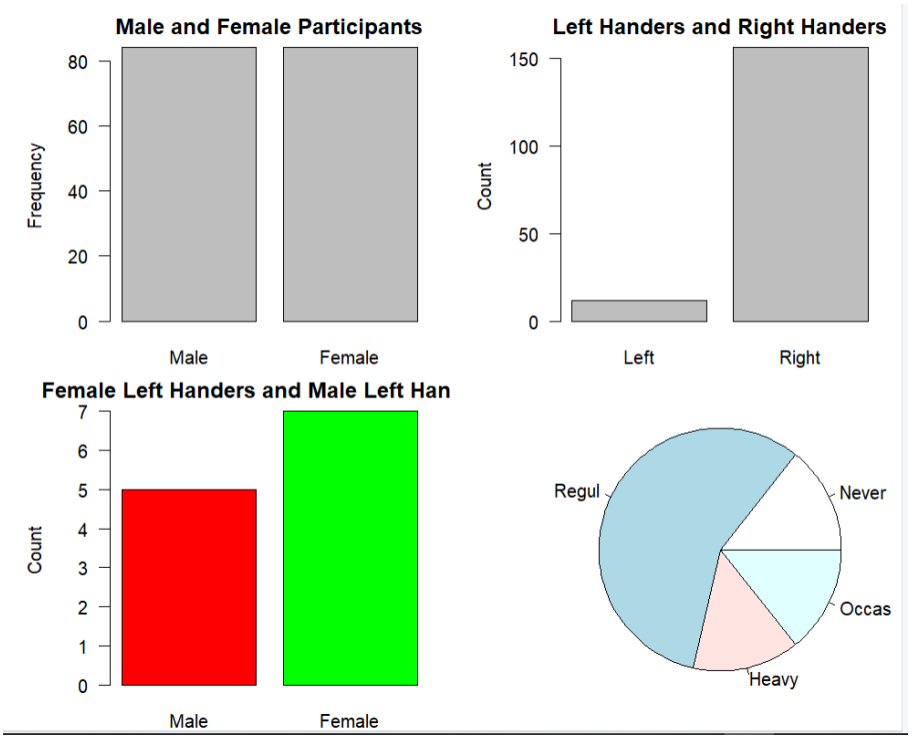
> par(mfrow=c(2,2),mar=c(2,5,2,1),las=1, bty='n')

> barplot(sex, main = 'Male and Female Participants', names.arg = c('Male','Female'), ylab = 'Frequency')

> barplot(handedness, main='Left Handers and Right Handers', names.arg = c('Left', 'Right'), ylab = 'Count')

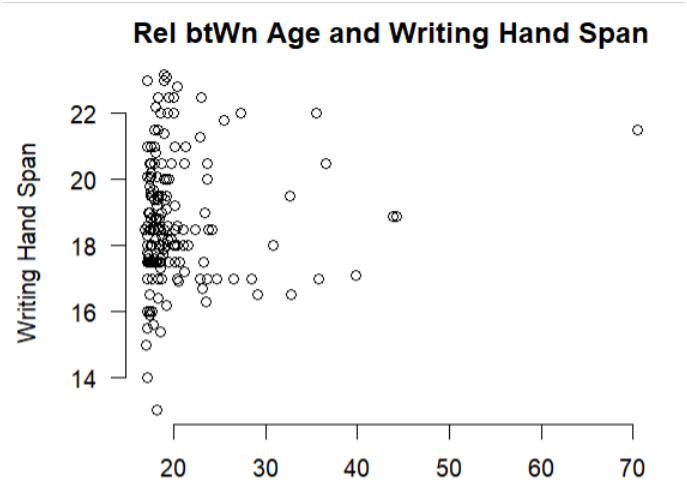
> barplot(lefthanded, main='Female Left Handers and Male Left Handers', names.arg = c('Male', 'Female'), col = c("red", "green"), ylab = 'Count')

> pie(male.left.handers, labels, radius = 1)

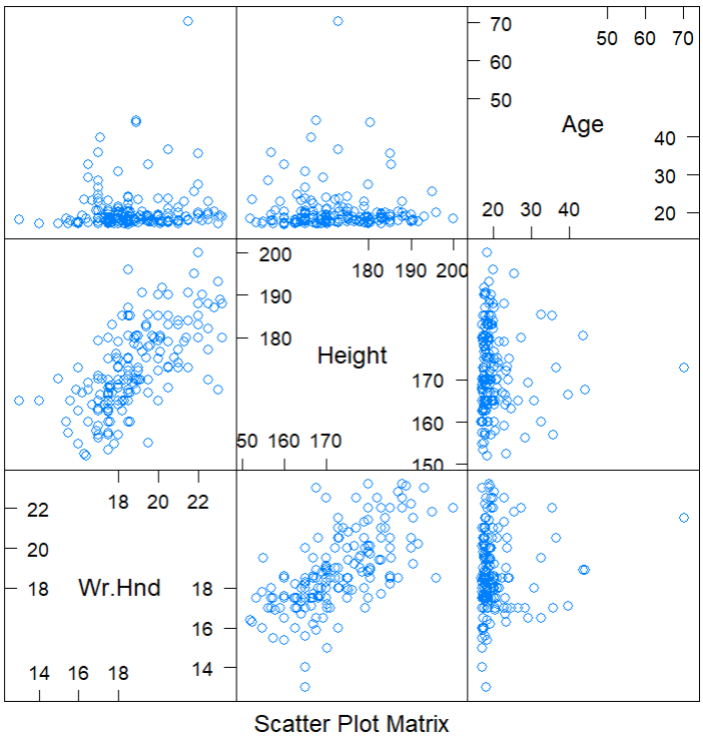


Q8)

> plot(dfAge, dfWr.Hnd, xlab='Age', ylab='Writing Hand Span', main='Relationship between Age and Writing Hand Span')



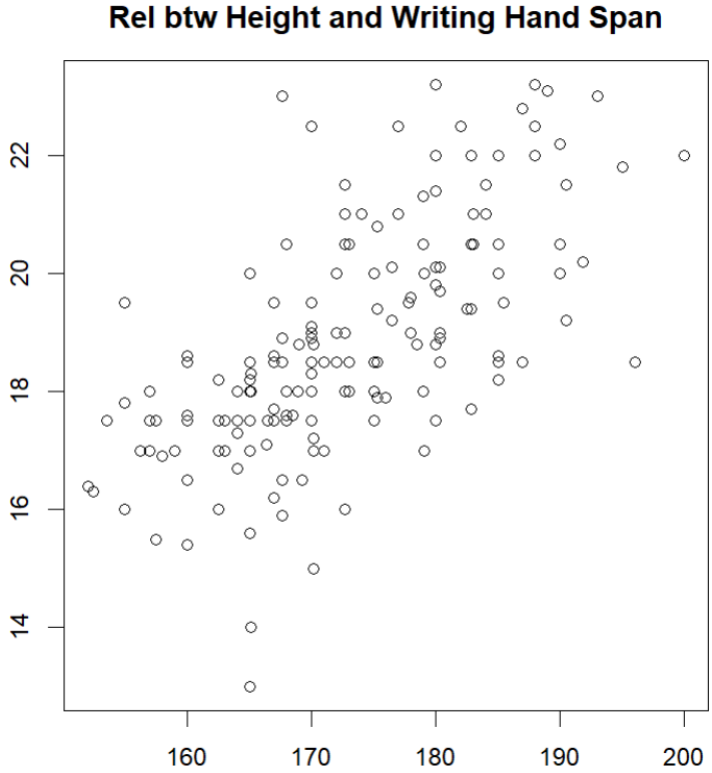
Q9) > library(lattice)
> dev.off()
> splom(df[c(2,10,12)])



Q10)

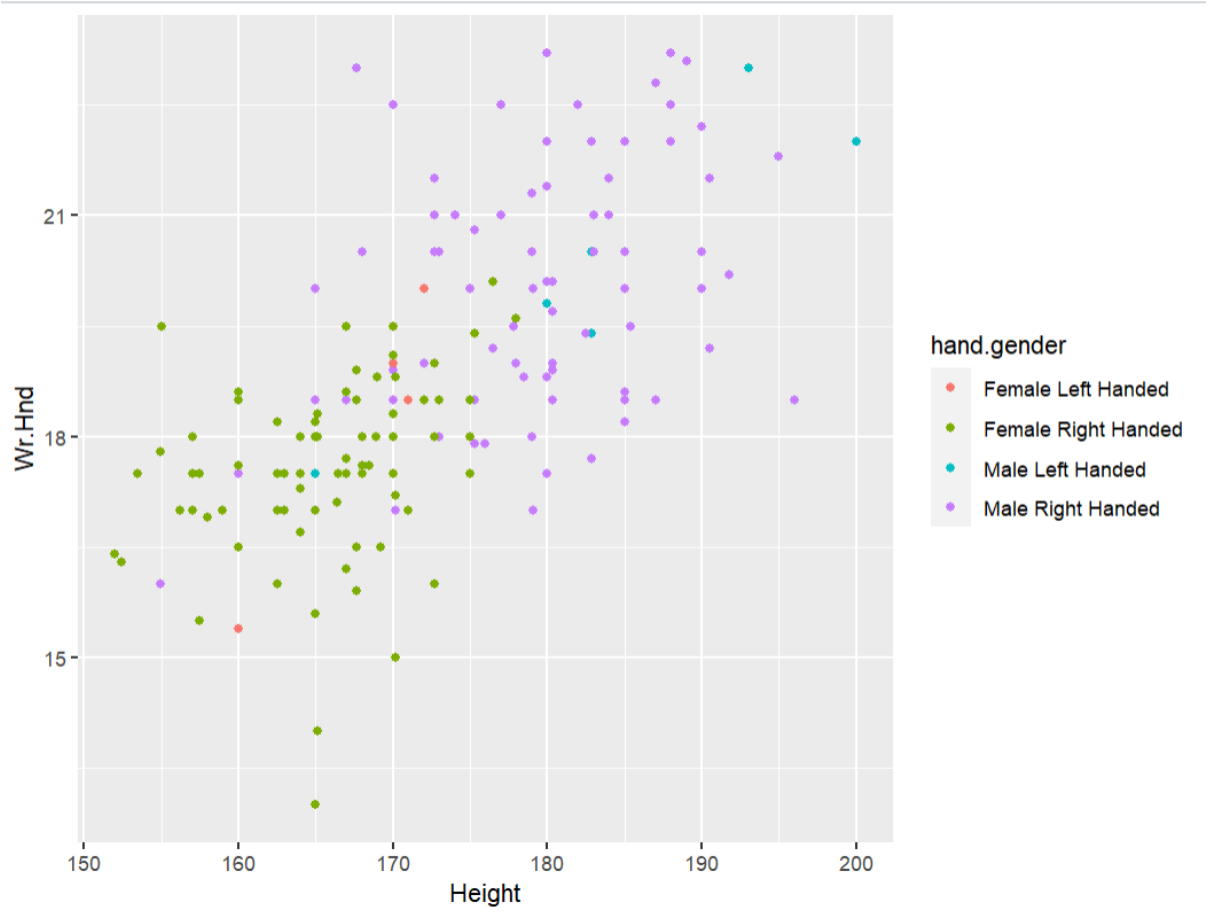
> par(mar=c(3,3,3,3))

> plot(df$Height, df$Wr.Hnd, xlab='Height', ylab='Writing Hand Span', main='Rel btw Height and Writing Hand Span')



Q11)

```
> femright <- data.frame(filter(df, Sex == 'Female', W.Hnd == 'Right'))
> femright$hand.gender <- "Female Right Handed"
> femleft <- data.frame(filter(df, Sex == 'Female', W.Hnd == 'Left'))
> femleft$hand.gender <- "Female Left Handed"
> maleright <- data.frame(filter(df, Sex == 'Male', W.Hnd == 'Right'))
> maleright$hand.gender <- "Male Right Handed"
> maleleft <- data.frame(filter(df, Sex == 'Male', W.Hnd == 'Left'))
> maleleft$hand.gender <- "Male Left Handed"
> df1 <- rbind(maleleft, maleright, femleft, femright)
> ggplot(df1,aes(x=Height,y=Wr.Hnd,color=hand.gender))+geom_point()
```



Q12)

```
> boxplot(df$Pulse, xlab='Boxplot', ylab='Pulse', label="Pulse Rate")
> summarize(df, mean(Pulse,na.rm=TRUE))
  mean(Pulse, na.rm = TRUE)
1                  74.02381
```

```
> summarize(df, median(Pulse,na.rm=TRUE))
  median(Pulse, na.rm = TRUE)
1                          72
> summarize(df, IQR(Pulse,na.rm=TRUE))
  IQR(Pulse, na.rm = TRUE)
1                    13.25
> summarize(df, sd(Pulse,na.rm=TRUE))
  sd(Pulse, na.rm = TRUE)
1                11.53747
> summarize(df, var(Pulse,na.rm=TRUE))
  var(Pulse, na.rm = TRUE)
1                 133.1132
```