Assignment-1

1. How many patients have heart disease?

   Code:

```
x<-length(which(mydata$heart_problem == 1))
cat("No. of patients have heart disease: ",x)
```

   Output:

```
> x<-length(which(mydata$heart_problem == 1))
> cat("No. of patients have heart disease: ",x)
No. of patients have heart disease:  101
```

2. What is the average Cholesterol level of people with heart disease and without heart disease? What is the standard deviation?

   Code:

```
m<-tapply(RawData$chol,RawData$heart_problem,mean)
s<-tapply(RawData$chol,RawData$heart_problem,sd)
cat("average Cholesterol level of people with heart disease is: ",m[2])
cat("standard deviation of Cholesterol level of people with heart disease is: ",s[2])
cat("average Cholesterol level of people without heart disease is: ",m[1])
cat("standard deviation of cholesterol level of people without heart disease is: ",s[1])
```

   Output:

```
> m<-tapply(RawData$chol,RawData$heart_problem,mean)
> cat("average Cholesterol level of people with heart disease is: ",m[2])
average Cholesterol level of people with heart disease is:  269.1881
> cat("average Cholesterol level of people without heart disease is: ",m[1])
average Cholesterol level of people without heart disease is:  239.9529
> s<-tapply(RawData$chol,RawData$heart_problem,sd)
> cat("standard deviation of Cholesterol level of people with heart disease is: ",s[2])
standard deviation of Cholesterol level of people with heart disease is:  79.91116
> cat("standard deviation of cholesterol level of people without heart disease is: ",s
[1])
standard deviation of cholesterol level of people without heart disease is:  56.69453
> plot
```

	WITH HEART DISEASE	WITHOUT HEART DISEASE
AVERAGE	269.1881	239.9529
STANDARD DEVIATION	79.91116	56.69453

3. What is the median and average age of people with cholesterol higher than 240.0, higher than 240.0 with heart disease, higher than 240.0 without heart disease?
   Code:

```
x<-RawData[RawData$chol>240,]
a<-mean(x$age)
median_a<-median(x$age)
avg<-tapply(x$age,x$heart_problem,mean)
median<-tapply(x$age,x$heart_problem,median)
cat("average age of people with cholestrol higher than 240 is: ",a)
cat("median of age with cholestrol higher than 240 is: ",median_a)
cat("average age of people with heart disease and cholesterol higher than 240 is: ",avg[2])
cat("median of age for people with heart disease and cholesterol higher than 240 is: ",median[2])
cat("average age of people without heart disease and cholesterol higher than 240 is: ",avg[1])
cat("median of age for people without heart disease and cholesterol higher than 240 is: ",median[1])
```

   Output:

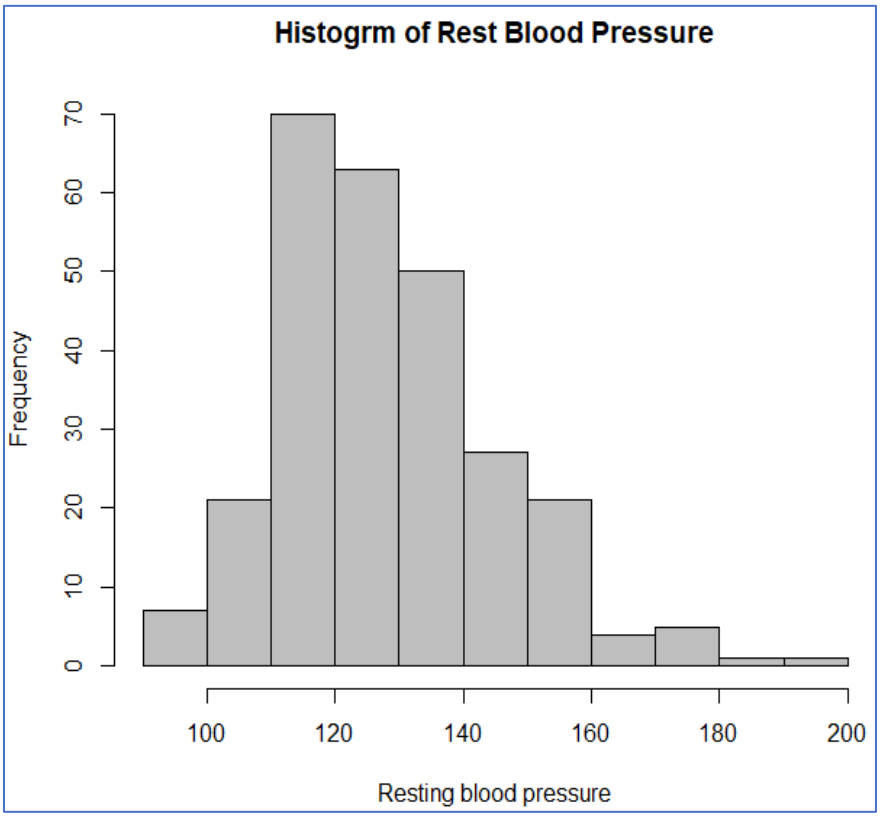
```
> x<-RawData[RawData$chol>240,]
> a<-mean(x$age)
> median_a<-median(x$age)
> avg<-tapply(x$age,x$heart_problem,mean)
> median<-tapply(x$age,x$heart_problem,median)
> cat("average age of people with cholestrol higher than 240 is: ",a)
average age of people with cholestrol higher than 240 is:  48.35252
> cat("median of age with cholestrol higher than 240 is: ",median_a)
median of age with cholestrol higher than 240 is:  49
> cat("average age of people with heart disease and cholesterol higher than 240 is: ",avg[2])
average age of people with heart disease and cholesterol higher than 240 is:  49.41935
> cat("median of age for people with heart disease and cholesterol higher than 240 is: ",median[2])
median of age for people with heart disease and cholesterol higher than 240 is:  50
> cat("average age of people without heart disease and cholesterol higher than 240 is: ",avg[1])
average age of people without heart disease and cholesterol higher than 240 is:  47.49351
> cat("median of age for people without heart disease and cholesterol higher than 240 is: ",median[1])
median of age for people without heart disease and cholesterol higher than 240 is:  48
```

	Age	With Heart Disease	Without Heart Disease
AVERAGE	48.35252	49.41935	47.49351
MEDIAN	49	50	48

4. Create a histogram of resting blood pressure.
   Code:

```
hist(RawData$trestbps,
     main="Histogrm of Rest Blood Pressure",
     xlab="Resting blood pressure",
     col="Grey")
```

Output:

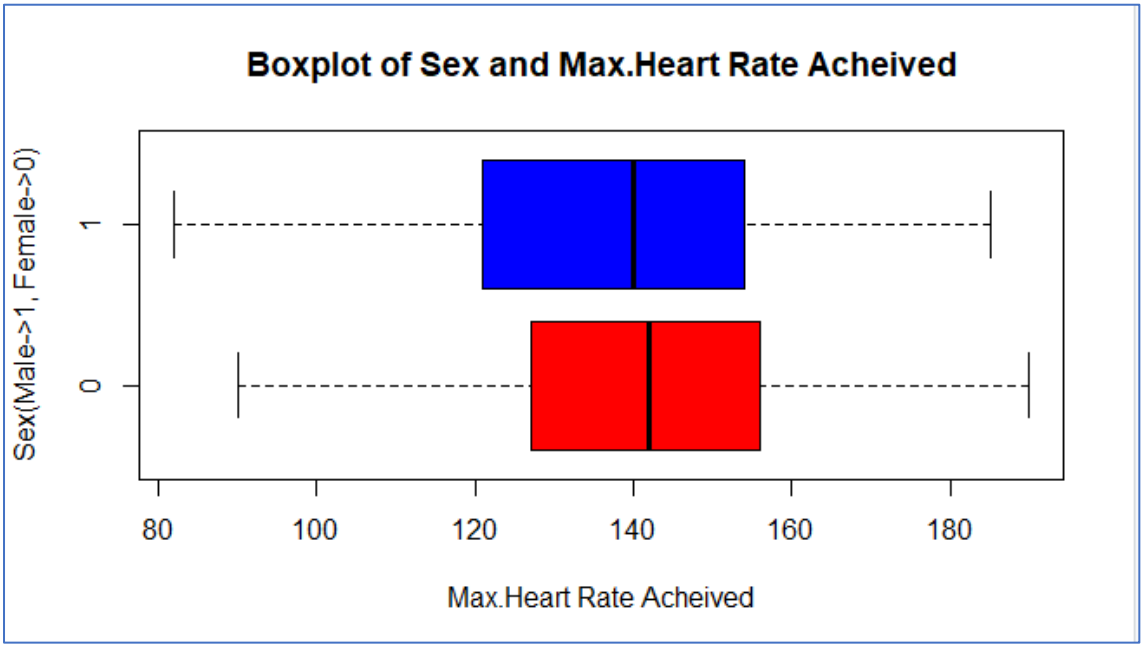


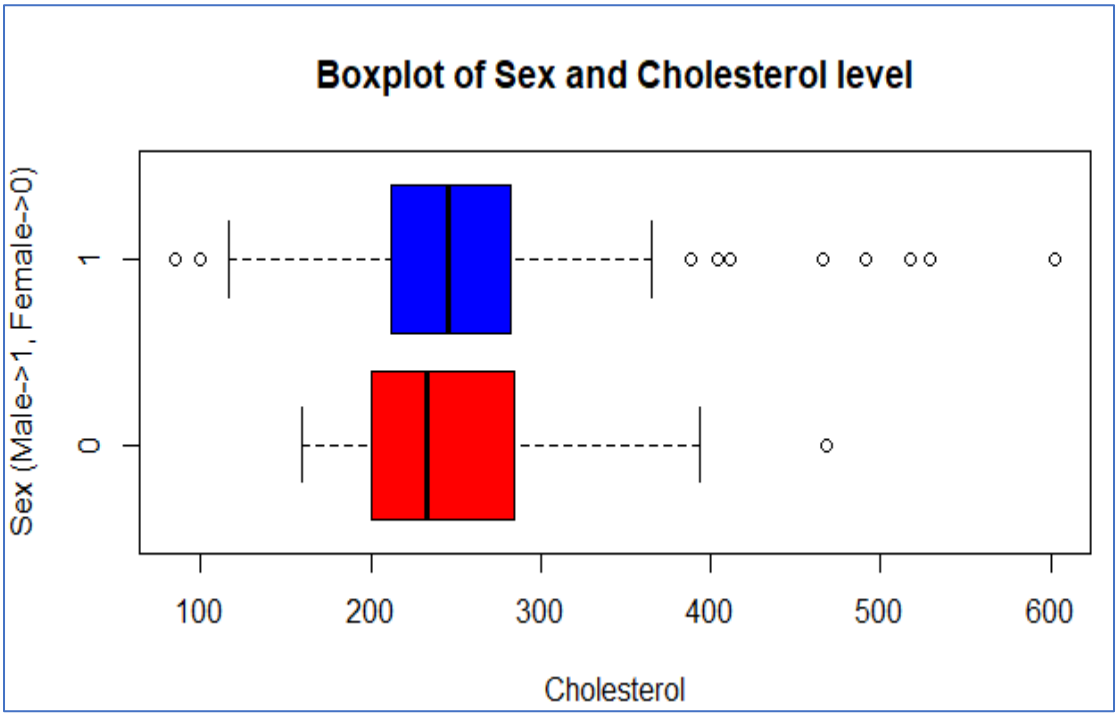
5. Create boxplots based on the sex of the patients for the following attributes:
   - cholesterol level
   - maximum heart rate achieved

Code:

```
boxplot(RawData$thalach~RawData$sex,
        main = "Boxplot of Sex and Max.Heart Rate Acheived",
        xlab="Max.Heart Rate Acheived",
        ylab="Sex(Male->1, Female->0)",
        col=c("Red","Blue"),
        horizontal = TRUE)
boxplot(RawData$chol~RawData$sex,
        main = "Boxplot of Sex and Cholesterol level",
        xlab="Cholesterol",
        ylab="Sex (Male->1, Female->0)",
        col=c("Red","Blue"),
        horizontal = TRUE)
```

Output:





6. For each Box plot, answer the following questions:
    - What is the H-Spread (Q3-Q1) of cholesterol level for male and females?

Code:

```
q1<-tapply(RawData$chol,RawData$sex,quantile,c(0.25,0.75),na.rm=TRUE)
q1
```

Output:

```
> q1<-tapply(RawData$chol,RawData$sex,quantile,c(0.25,0.75),na.rm=TRUE)
> q1
$`0`
   25%    75%
201.25 282.75

$`1`
25% 75%
212 282
```

H-Spread of Female = Q3-Q1 = 282.75-201.25 =81.50
H-Spread of Male = Q3-Q1 =282-212 = 70

- What are the Lower Hinge and Upper Hinge values for maximum heart rate for male and female?

Code:

```
q2<-tapply(RawData$thalach,RawData$sex,quantile,c(0.25,0.75),na.rm=TRUE)
q2
```

Output:

```
> q2<-tapply(RawData$thalach,RawData$sex,quantile,c(0.25,0.75),na.rm=TRUE)
> q2
$`0`
25% 75%
127 156

$`1`
25% 75%
121 154
```

	Lower Hinge	**Upper Hinge**
Male	121	154
Female	127	156

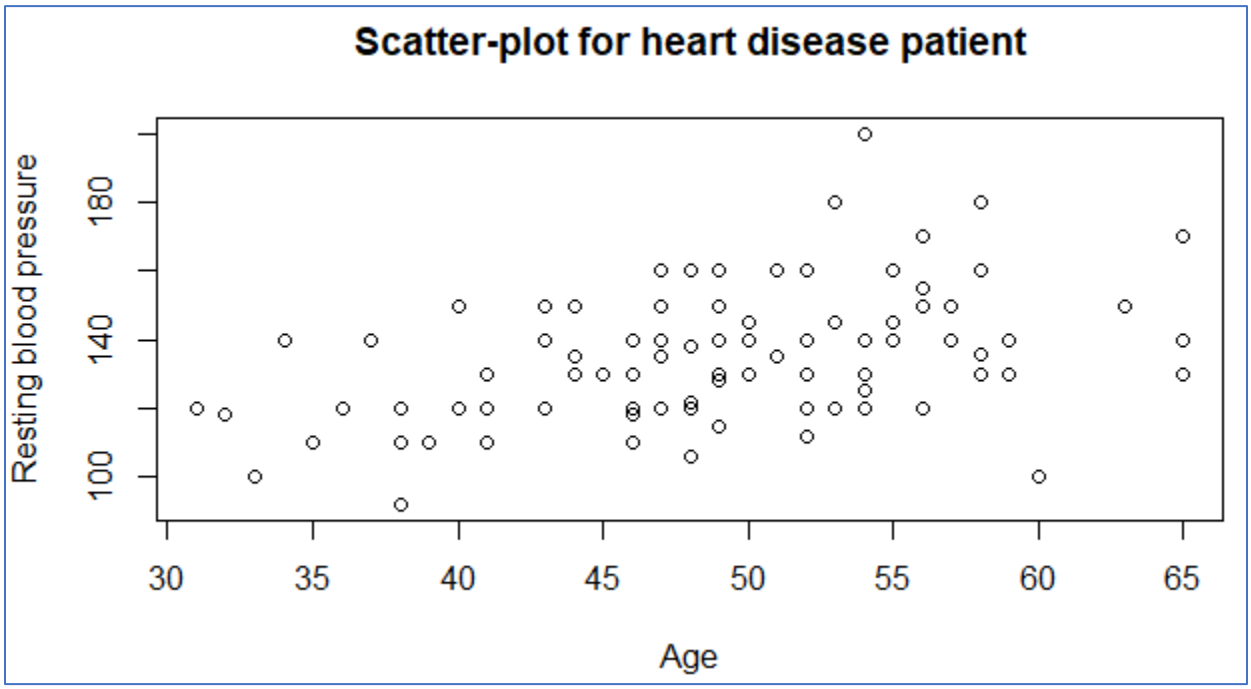
7. In order to find if two attributes are related and their values change together, we can use Scatter plot. Follow the instruction below and answer the questions:

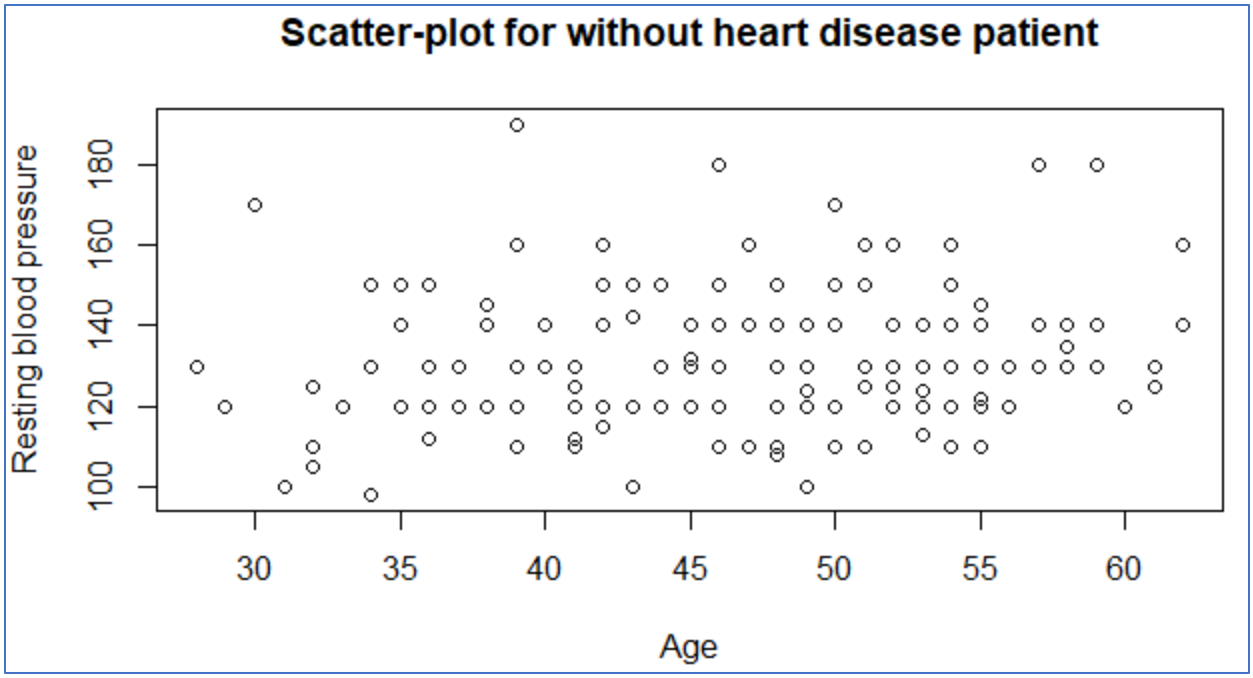
- Create two scatter plots of age and resting blood pressure for people with heart disease and without heart disease. Is there any visual correlation?

Code:

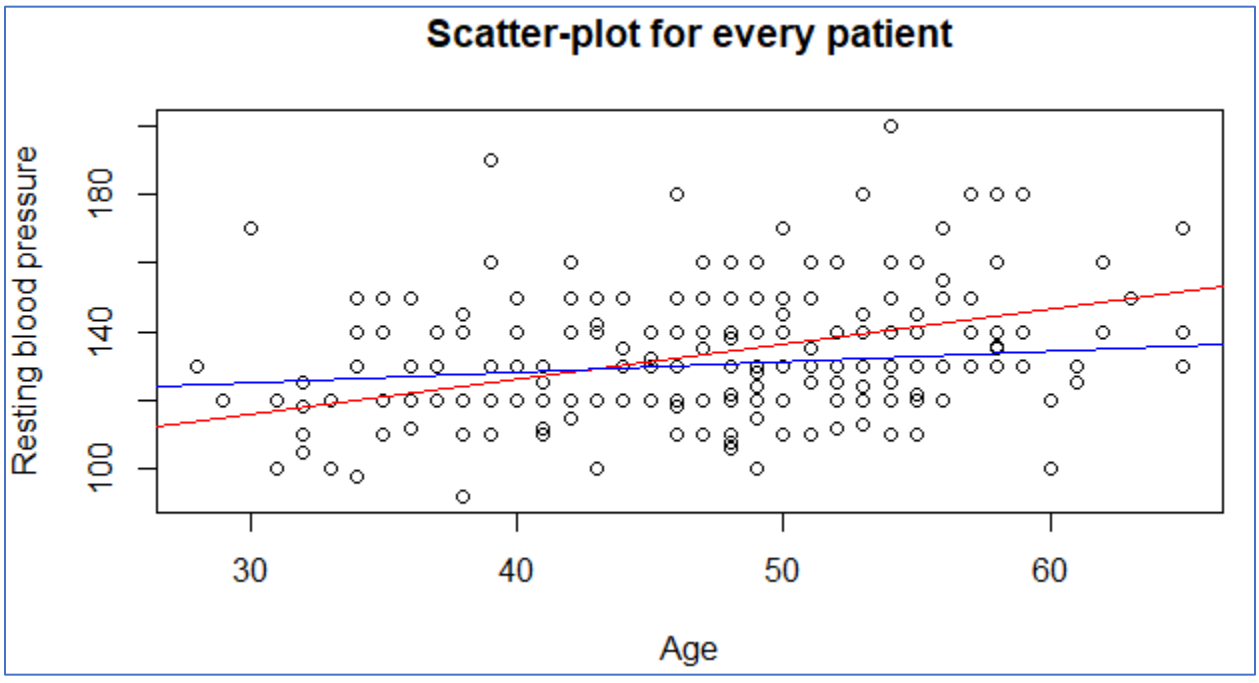
```
hd<-RawData[RawData$heart_problem==1,]
plot(hd$age,hd$trestbps,
     xlab="Age",
     ylab="Resting blood pressure",
     main="Scatter-plot for heart disease patient")

whd<-RawData[RawData$heart_problem==0,]
plot(whd$age,whd$trestbps,
     xlab="Age",
     ylab="Resting blood pressure",
     main="Scatter-plot for without heart disease patient")

plot(RawData$age,RawData$trestbps,xlab="Age",
     ylab="Resting blood pressure",
     main="Scatter-plot for every patient")

abline(lm(hd$trestbps~hd$age),col="Red")
abline(lm(whd$trestbps~whd$age),col="Blue")
```

Output:





As we can see from the below scatter plot and regression line of the above two scatter plots, the patients with heart disease and without it is correlated as both has positive p-value.



- Calculate the average resting blood pressure of each age (HINT : Use Group by for age) for people with heart disease. Calculate the average resting blood pressure of each age (HINT : Use Group by for age) for people without heart disease.

Code:

```
x<-RawData[RawData$heart_problem==1,]
avg<-tapply(x$trestbps,x$age,mean)
avg

x1<-RawData[RawData$heart_problem==0,]
avg1<-tapply(x1$trestbps,x1$age,mean)
avg1
```

Output:

```
> x<-RawData[RawData$heart_problem==1,]
> avg<-tapply(x$trestbps,x$age,mean)
> avg
      31       32       33       34       35       36       37       38       39       40       41       43
120.0000 118.0000 100.0000 140.0000 110.0000 120.0000 140.0000 107.3333 110.0000 135.0000 120.0000 132.5000
      44       45       46       47       48       49       50       51       52       53       54       55
138.3333 130.0000 123.0000 141.0000 140.7500 136.1429 139.0000 147.5000 136.5000 148.3333 137.5000 145.0000
      56       57       58       59       60       63       65
149.0000 145.0000 147.2000 135.0000 100.0000 150.0000 146.6667
```

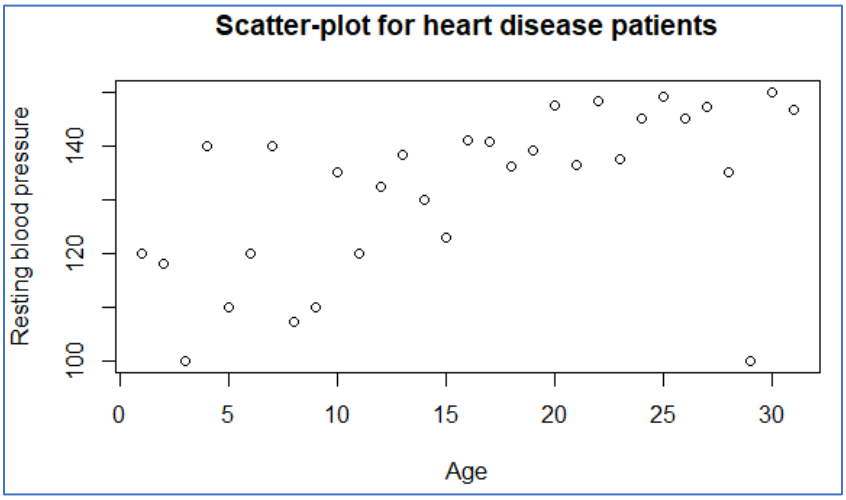
```
> x1<-RawData[RawData$heart_problem==0,]
> avg1<-tapply(x1$trestbps,x1$age,mean)
> avg1
      28       29       30       31       32       33       34       35       36       37       38       39
130.0000 120.0000 170.0000 100.0000 113.3333 120.0000 126.0000 132.5000 128.0000 127.1429 135.0000 132.5000
      40       41       42       43       44       45       46       47       48       49       50       51
132.5000 120.2857 132.1429 124.5714 130.0000 132.4000 134.2857 137.5000       NA 122.8000 135.7143 133.5714
      52       53       54       55       56       57       58       59       60       61       62
133.5714 128.3750 133.3333 125.7000 127.5000 150.0000 133.7500 145.0000 120.0000 127.5000 150.0000
```

- Now create two scatter plots using the previous results. Do you see a pattern now? Do people without heart disease have higher blood pressure as they age than people with heart disease?

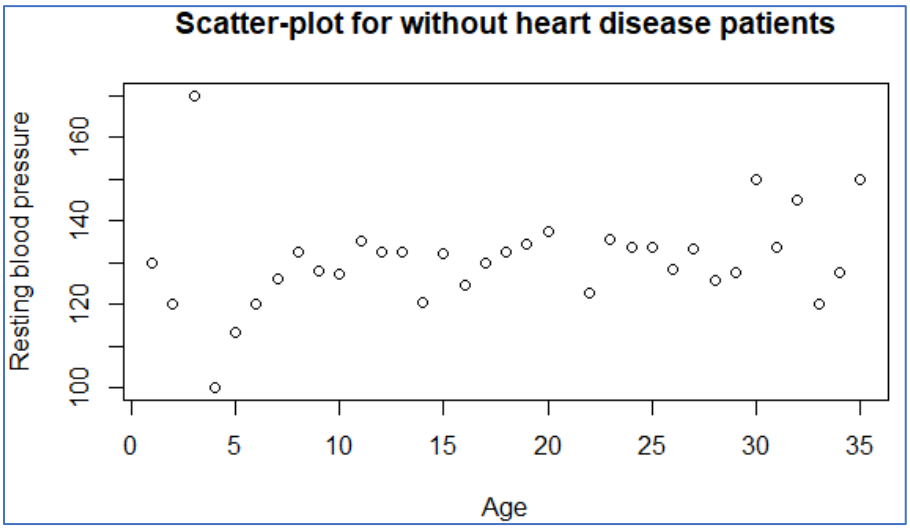
Code:

```
plot(avg,xlab="Age",
     ylab="Resting blood pressure",
     main="Scatter-plot for heart disease patients")

plot(avg1,xlab="Age",
     ylab="Resting blood pressure",
     main="Scatter-plot for without heart disease patients")
```

Output:



As per above scatter plot, its look like as age increases resting blood pressure of patients also increases.



For those who don't have heart disease, above plot indicates, as age increases resting blood pressure of patients also increases same as heart disease patients but from above two plots we can see slope of without heart disease is less that the patients who have.

8. Compare the resting blood pressure of people with heart disease and without.

Code:

```
s<-tapply(RawData$trestbps,RawData$heart_problem,summary)
s
```

Output:

```
> s<-tapply(RawData$trestbps,RawData$heart_problem,summary)
> s
$`0`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   98.0   120.0   130.0   130.5   140.0   190.0       1

$`1`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   92.0   120.0   135.0   135.7   150.0   200.0
```