

# Heart Disease

Ismail Shaikh

11/14/2021

## *Loading Packages*

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(corrplot)

## corrplot 0.91 loaded
```

## *Importing the data*

```
data <- read.csv("/Users/ismailshaikh/Documents/heart.csv")
```

## *Exploring the data*

```
head(data)

##   Age Sex ChestPainType RestingBP Cholesterol FastingBS RestingECG MaxHR
## 1  40  M             ATA        140         289          0    Normal   172
## 2  49  F             NAP        160         180          0    Normal   156
## 3  37  M             ATA        130         283          0         ST    98
## 4  48  F             ASY        138         214          0    Normal  108
## 5  54  M             NAP        150         195          0    Normal  122
## 6  39  M             NAP        120         339          0    Normal  170
##   ExerciseAngina Oldpeak ST_Slope HeartDisease
## 1              N     0.0       Up            0
## 2              N     1.0      Flat            1
## 3              N     0.0       Up            0
## 4              Y     1.5      Flat            1
## 5              N     0.0       Up            0
## 6              N     0.0       Up            0

names(data)
```

```
## [1] "Age"           "Sex"           "ChestPainType" "RestingBP"
## [5] "Cholesterol"   "FastingBS"     "RestingECG"     "MaxHR"
## [9] "ExerciseAngina" "Oldpeak"       "ST_Slope"       "HeartDisease"
```

```
glimpse(data)
```

```
## Rows: 918
## Columns: 12
## $ Age          <int> 40, 49, 37, 48, 54, 39, 45, 54, 37, 48, 37, 58, 39,
49,...
## $ Sex          <chr> "M", "F", "M", "F", "M", "M", "F", "M", "M", "F", "
F", ...
## $ ChestPainType <chr> "ATA", "NAP", "ATA", "ASY", "NAP", "NAP", "ATA", "A
TA",...
## $ RestingBP    <int> 140, 160, 130, 138, 150, 120, 130, 110, 140, 120, 1
30, ...
## $ Cholesterol  <int> 289, 180, 283, 214, 195, 339, 237, 208, 207, 284, 2
11, ...
## $ FastingBS    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0...
## $ RestingECG   <chr> "Normal", "Normal", "ST", "Normal", "Normal", "Norm
al",...
## $ MaxHR        <int> 172, 156, 98, 108, 122, 170, 170, 142, 130, 120, 14
2, 9...
## $ ExerciseAngina <chr> "N", "N", "N", "Y", "N", "N", "N", "N", "Y", "N", "
N", ...
## $ Oldpeak      <dbl> 0.0, 1.0, 0.0, 1.5, 0.0, 0.0, 0.0, 0.0, 1.5, 0.0, 0
.0, ...
## $ ST_Slope     <chr> "Up", "Flat", "Up", "Flat", "Up", "Up", "Up", "Up",
"Fl...
## $ HeartDisease <int> 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1,
0, 1...
```

```
summary(data)
```

```
##      Age          Sex      ChestPainType      RestingBP
## Min.   :28.00    Length:918    Length:918    Min.    :  0.0
## 1st Qu.:47.00    Class :character  Class :character  1st Qu.:120.0
## Median :54.00    Mode  :character  Mode  :character  Median :130.0
## Mean   :53.51                                     Mean   :132.4
## 3rd Qu.:60.00                                     3rd Qu.:140.0
## Max.   :77.00                                     Max.   :200.0
## Cholesterol      FastingBS      RestingECG      MaxHR
## Min.    :  0.0    Min.    :0.0000    Length:918    Min.    : 60.0
## 1st Qu.:173.2    1st Qu.:0.0000    Class :character  1st Qu.:120.0
## Median :223.0    Median :0.0000    Mode  :character  Median :138.0
## Mean    :198.8    Mean    :0.2331                                     Mean   :136.8
## 3rd Qu.:267.0    3rd Qu.:0.0000                                     3rd Qu.:156.0
## Max.    :603.0    Max.    :1.0000                                     Max.   :202.0
## ExerciseAngina    Oldpeak      ST_Slope      HeartDisease
## Length:918        Min.    :-2.6000    Length:918    Min.    :0.0000
```

```
## Class :character 1st Qu.: 0.0000 Class :character 1st Qu.:0.0000
## Mode :character Median : 0.6000 Mode :character Median :1.0000
## Mean : 0.8874 Mean :0.5534
## 3rd Qu.: 1.5000 3rd Qu.:1.0000
## Max. : 6.2000 Max. :1.0000
```

***Checking if we have missing values and if yes, what is the count of NULL values in the dataset***

```
sum(is.na(data))
```

```
## [1] 0
```

***Data types of the attributes***

```
str(data)
```

```
## 'data.frame': 918 obs. of 12 variables:
## $ Age : int 40 49 37 48 54 39 45 54 37 48 ...
## $ Sex : chr "M" "F" "M" "F" ...
## $ ChestPainType : chr "ATA" "NAP" "ATA" "ASY" ...
## $ RestingBP : int 140 160 130 138 150 120 130 110 140 120 ...
## $ Cholesterol : int 289 180 283 214 195 339 237 208 207 284 ...
## $ FastingBS : int 0 0 0 0 0 0 0 0 0 ...
## $ RestingECG : chr "Normal" "Normal" "ST" "Normal" ...
## $ MaxHR : int 172 156 98 108 122 170 170 142 130 120 ...
## $ ExerciseAngina: chr "N" "N" "N" "Y" ...
## $ Oldpeak : num 0 1 0 1.5 0 0 0 0 1.5 0 ...
## $ ST_Slope : chr "Up" "Flat" "Up" "Flat" ...
## $ HeartDisease : int 0 1 0 1 0 0 0 0 1 0 ...
```

***Performing data transformation for better representation of the data***

```
newdata <- data%>%
  mutate(Sex = if_else(Sex == "M", 'Male', 'Female'),
         ChestPainType = if_else(ChestPainType == "TA", "Typical Angina",
                                if_else(ChestPainType == "ATA", "Atypical Angina",
                                         if_else(ChestPainType == "NAP", "Non-Anginal Pain", "Asymptomatic"))),
         FastingBS = if_else(FastingBS == 1, ">120", "<=120"),
         ExerciseAngina = if_else(ExerciseAngina == "N", "No", "Yes"),
         HeartDisease = if_else(HeartDisease == 1, "Yes", "No")
  ) %>%
  mutate_if(is.character, as.factor)%>%
  dplyr::select(Age, Sex, ChestPainType, RestingBP, Cholesterol, FastingBS, RestingECG, MaxHR, ExerciseAngina, Oldpeak, ST_Slope, HeartDisease)

head(newdata)
```

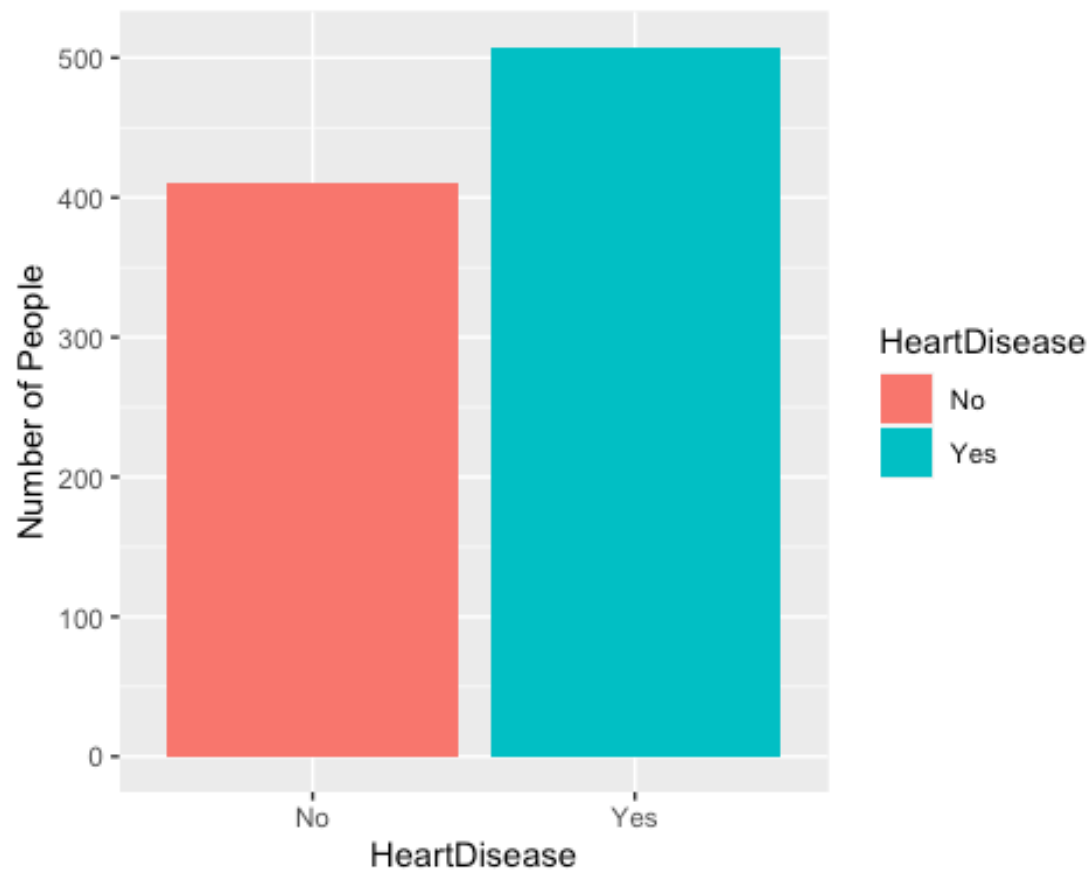
##	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR
## 1	40	Male	Atypical Angina	140	289	<=120	Normal	172
## 2	49	Female	Non-Anginal Pain	160	180	<=120	Normal	156
## 3	37	Male	Atypical Angina	130	283	<=120	ST	98
## 4	48	Female	Asymptomatic	138	214	<=120	Normal	108
## 5	54	Male	Non-Anginal Pain	150	195	<=120	Normal	122
## 6	39	Male	Non-Anginal Pain	120	339	<=120	Normal	170

##	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
## 1	No	0.0	Up	No
## 2	No	1.0	Flat	Yes
## 3	No	0.0	Up	No
## 4	Yes	1.5	Flat	Yes
## 5	No	0.0	Up	No
## 6	No	0.0	Up	No

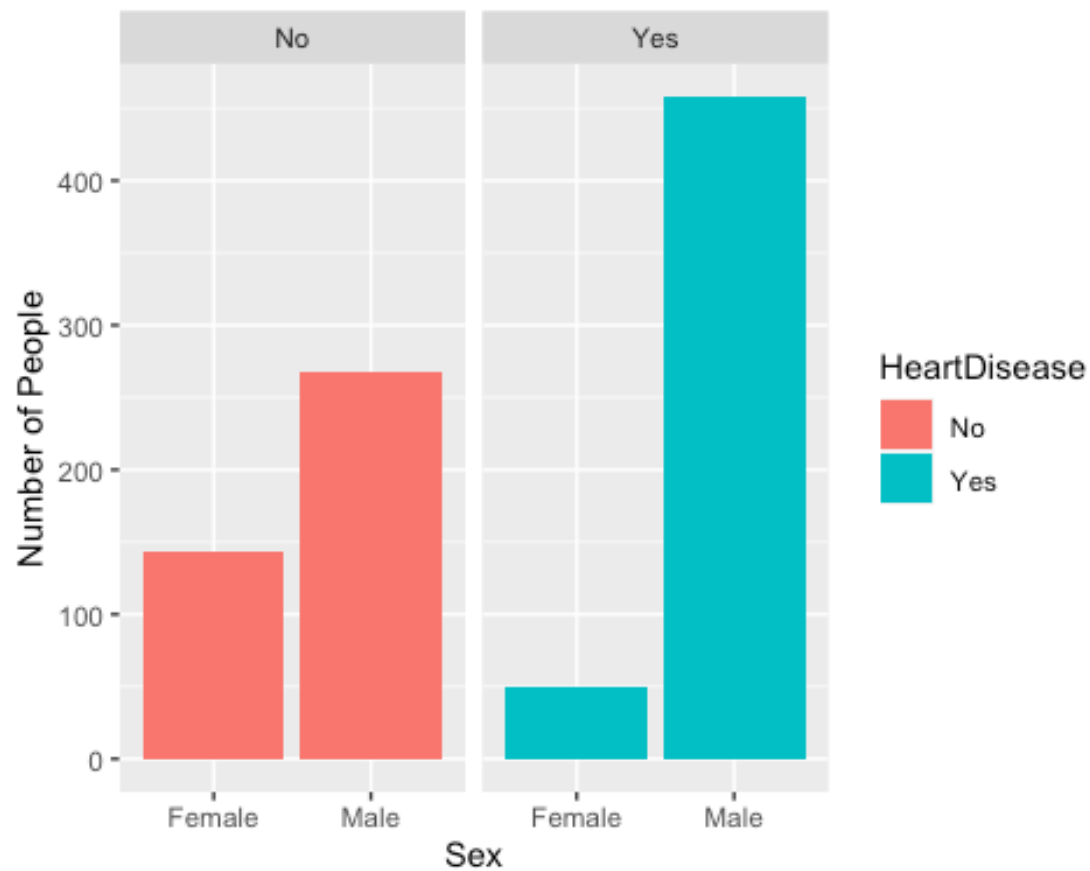
### ***Number of people having Heart Disease***

```
ggplot(newdata, aes(x = HeartDisease, fill = HeartDisease))+
  geom_bar()+
  ylab("Number of People")
```



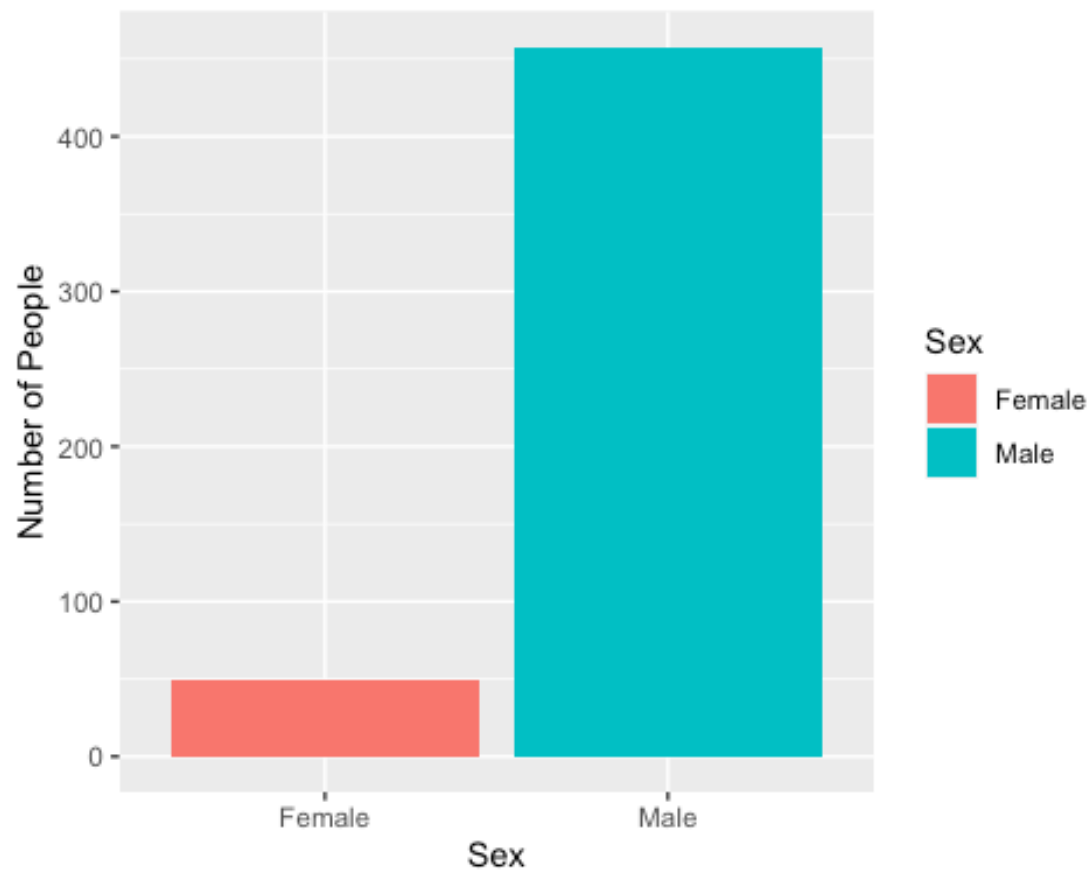
***Number of Male and Female having Heart Disease and not having Heart Disease***

```
ggplot(newdata,aes(x = Sex, y = ..count..))+  
  geom_bar(aes(fill = HeartDisease))+  
  facet_grid(.~HeartDisease)+  
  ylab("Number of People")
```



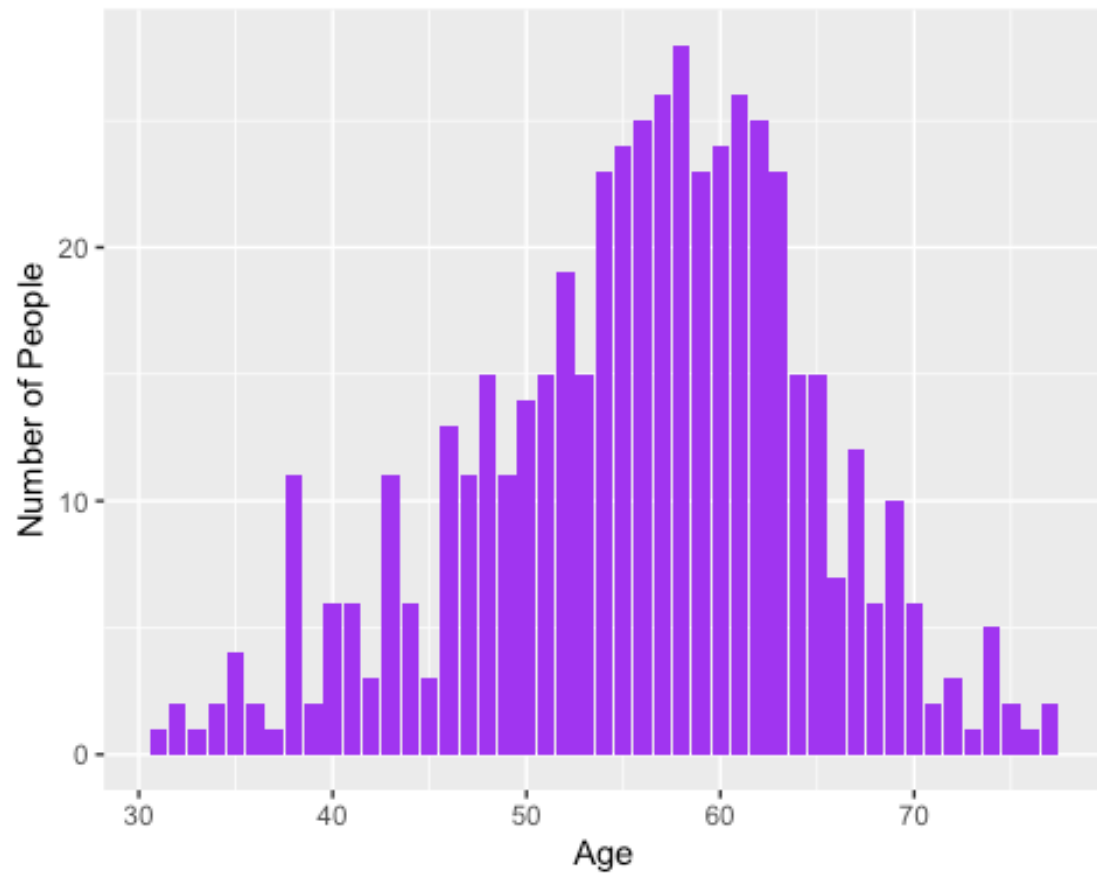
### ***Number of Male and Female having Heart Disease***

```
data2<- newdata%>%  
  filter(HeartDisease == "Yes")  
  
ggplot(data2,aes(x = Sex))+  
  geom_bar(aes(fill = Sex))+  
  ylab("Number of People")
```



### ***Age range of People with Heart Disease***

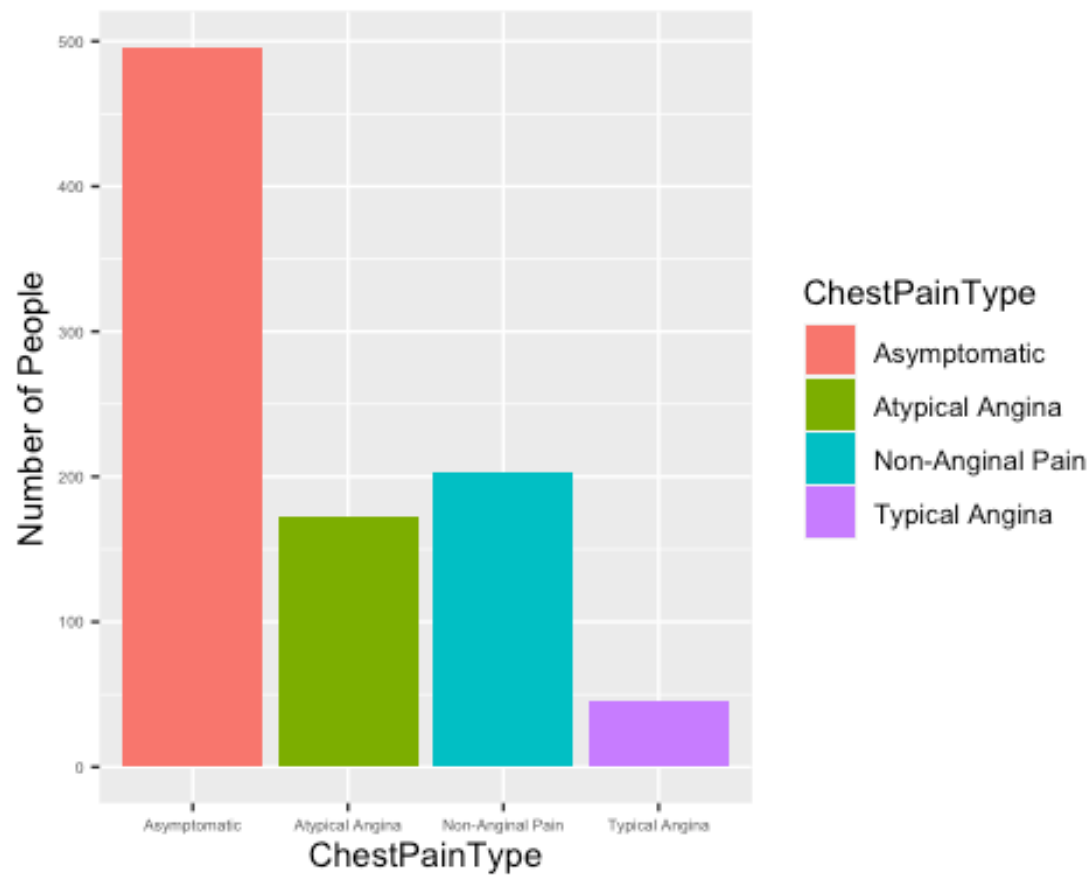
```
data2<- newdata%>%  
  filter(HeartDisease == "Yes")  
  
ggplot(data2,aes(x = Age))+  
  geom_bar(fill = "Purple")+  
  ylab("Number of People")
```



### ***ChestPain V/S Number of People***

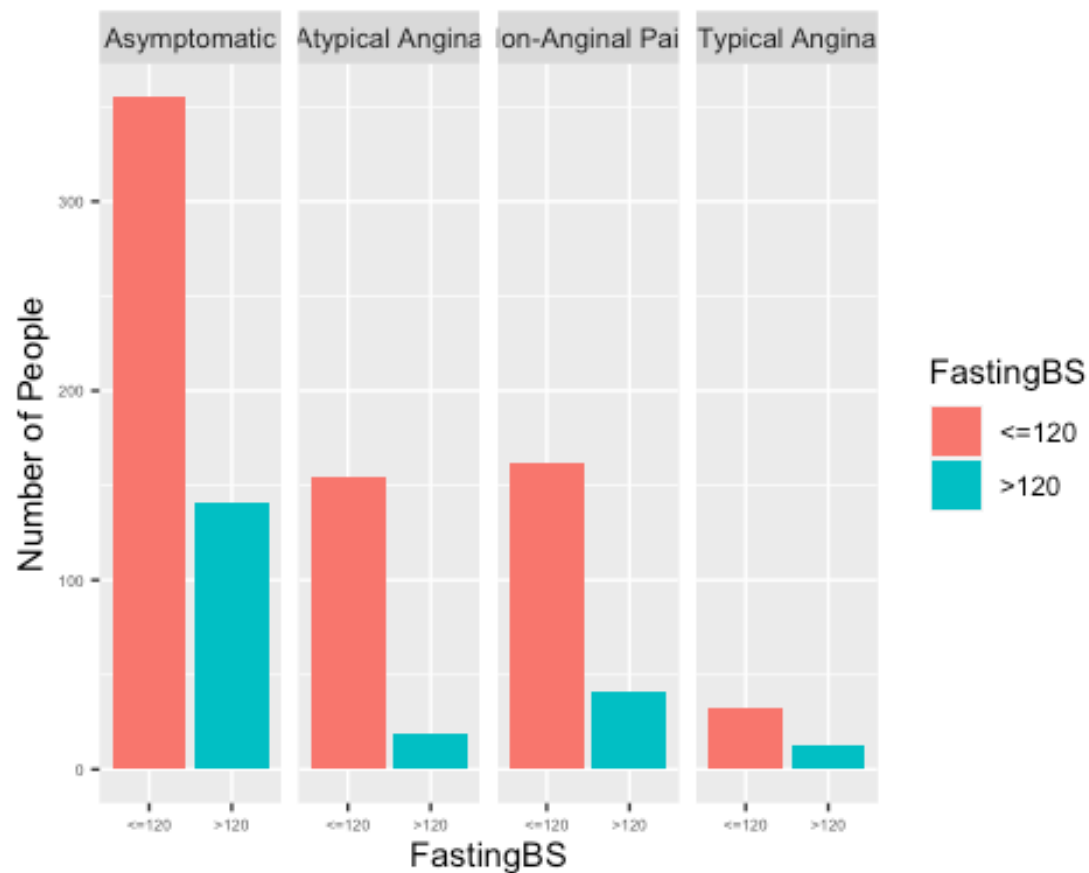
```
ggplot(newdata,aes(x = ChestPainType,fill = ChestPainType))+  
  geom_bar()+  
  ylab("Number of People")+  
  theme(axis.text.x = element_text(size = 5),  
        axis.text.y = element_text(size = 5))
```





***Blood pressure for different categories of chest pain***

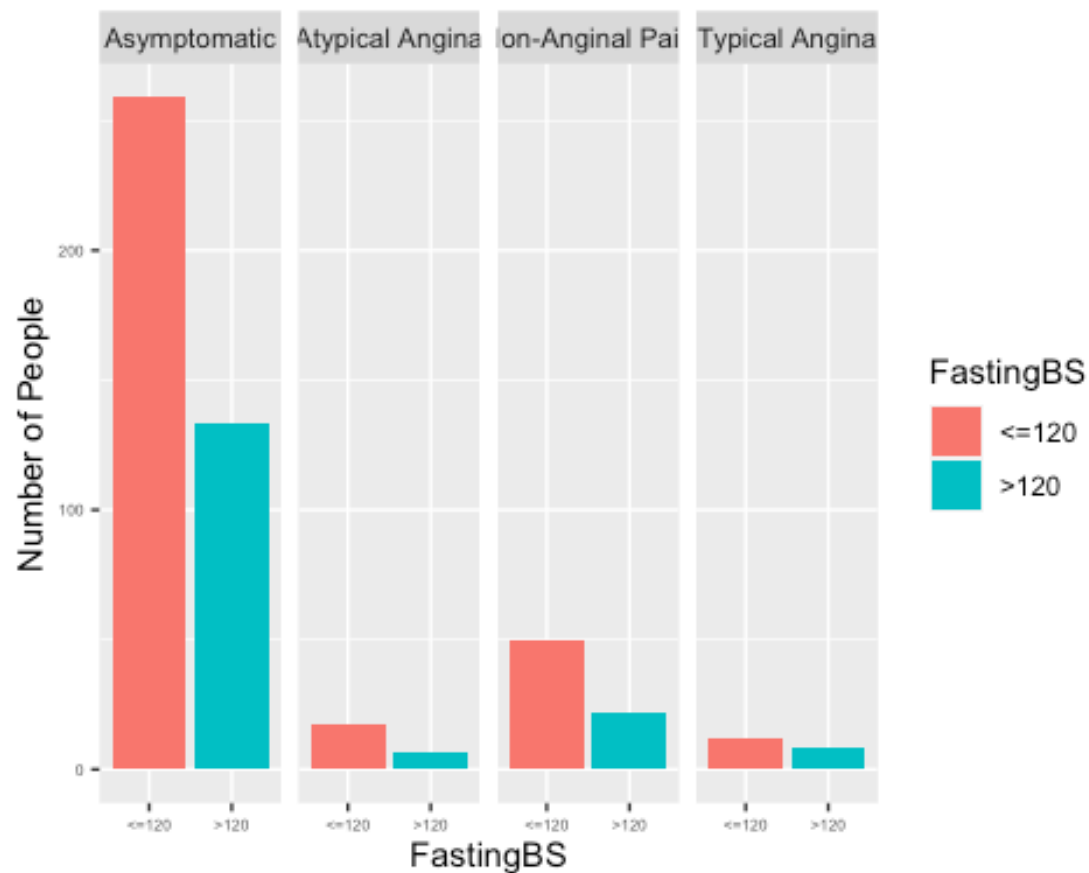
```
ggplot(newdata,aes(x = FastingBS,fill = FastingBS))+  
  geom_bar()+  
  ylab("Number of People")+  
  facet_grid(~ChestPainType)+  
  theme(axis.text.x = element_text(size = 5),  
        axis.text.y = element_text(size = 5))
```



***Blood Pressure for different categories of Chest Pain of people who have heart Disease***

```
data2 <- newdata%>%
  filter(HeartDisease == "Yes")

ggplot(data2, aes(x = FastingBS, fill = FastingBS)) +
  geom_bar() +
  ylab("Number of People") +
  facet_grid(~ChestPainType) +
  theme(axis.text.x = element_text(size = 5),
        axis.text.y = element_text(size = 5))
```



## Correlation

```
cr <- newdata[,c(1,4,5,8,10)]
cor(cr, method = 'pearson')
```

```
##           Age  RestingBP Cholesterol  MaxHR  Oldpeak
## Age      1.0000000  0.2543994 -0.09528177 -0.3820447  0.25861154
## RestingBP 0.25439936  1.0000000  0.10089294 -0.1121350  0.16480304
## Cholesterol -0.09528177  0.1008929  1.00000000  0.2357924  0.05014811
## MaxHR      -0.38204468 -0.1121350  0.23579240  1.0000000 -0.16069055
## Oldpeak     0.25861154  0.1648030  0.05014811 -0.1606906  1.00000000
```

```
cr <- newdata[,c(1,4,5,8,10)]
cor(cr, method = 'kendall')
```

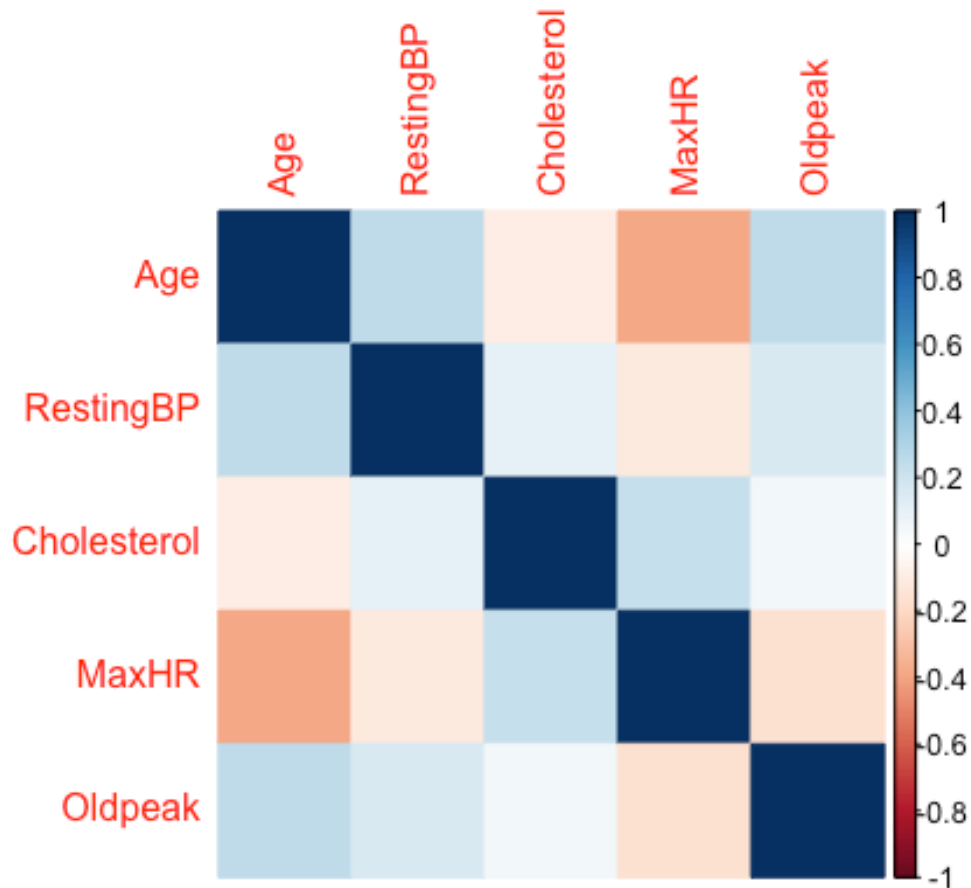
```
##           Age  RestingBP Cholesterol  MaxHR  Oldpeak
## Age      1.0000000  0.19827847 -0.03225405 -0.25312318  0.21777358
## RestingBP 0.19827847  1.00000000  0.07731297 -0.07559161  0.13023678
## Cholesterol -0.03225405  0.07731297  1.00000000  0.12598700  0.03805694
## MaxHR      -0.25312318 -0.07559161  0.12598700  1.00000000 -0.15031874
## Oldpeak     0.21777358  0.13023678  0.03805694 -0.15031874  1.00000000
```

```
cr <- newdata[,c(1,4,5,8,10)]
cor(cr, method = 'spearman')
```

```
##           Age  RestingBP Cholesterol  MaxHR  Oldpeak
## Age      1.00000000  0.2793901 -0.04715589 -0.3650251  0.29825005
## RestingBP 0.27939008  1.0000000  0.10948098 -0.1075661  0.17531294
## Cholesterol -0.04715589 0.1094810  1.00000000  0.1838998  0.05172277
## MaxHR     -0.36502506 -0.1075661  0.18389978  1.0000000 -0.20511293
## Oldpeak    0.29825005  0.1753129  0.05172277 -0.2051129  1.00000000
```

### Correlation Plot

```
corrplot(cor(cr),method = "color")
```



### Correlation Heat Map

```
palette = colorRampPalette(c("green", "white", "Blue")) (20)
heatmap(x = cor(cr), col = palette, symm = TRUE)
```

