

# Data Visualization of Heart Disease Dataset UCI

## Data Visualization of Heart Disease dataset of UCI

In this *R* notebook we are going to explore the data analytics and data visualization power of *R*.

In this example we are going to analyze the heart disease database from UCI machine library.

The dataset contains 76 predictors(features) and 303 observations. Patients with heart disease is binary coded as **Presence** given as 1 and **No Presence** as 0. The prerequisite to run in R Markdown is download the CSV data file in your working directory. This can be done by setting the current working directory as follows in R chunk: `setwd("C:\\Users\\RajuPC\\Documents\\MyR")`

First load the supporting *R* libraries

```
setwd("C:\\Users\\RajuPC\\Documents\\MyR") # This is how we set Working Directory
library(tidyverse) # A high efficient data viz and manipulation R Library
library(caret) # A collection of Machine Learning Libraries
library(plotly) # A interactive Graphing System
```

Loading of UCI heart disease data.

```
#Load the CSV data file
hci<-read_csv("heart.csv")
```

```
## Parsed with column specification:
## cols(
##   age = col_integer(),
##   sex = col_integer(),
##   cp = col_integer(),
##   trestbps = col_integer(),
##   chol = col_integer(),
##   fbs = col_integer(),
##   restecg = col_integer(),
##   thalach = col_integer(),
##   exang = col_integer(),
##   oldpeak = col_double(),
##   slope = col_integer(),
##   ca = col_integer(),
##   thal = col_integer(),
##   target = col_integer()
## )
```

```
# Convert sex attribute to character for plotting purpose
hci$sex <- as.character(hci$sex)
hci$sex[hci$sex== 1] <- "Male"
hci$sex[hci$sex== 0] <- "Female"

summary(hci) #Descriptive statistics
```

##	age	sex	cp	trestbps
##	Min. :29.00	Length:303	Min. :0.000	Min. : 94.0
##	1st Qu.:47.50	Class :character	1st Qu.:0.000	1st Qu.:120.0
##	Median :55.00	Mode :character	Median :1.000	Median :130.0
##	Mean :54.37		Mean :0.967	Mean :131.6
##	3rd Qu.:61.00		3rd Qu.:2.000	3rd Qu.:140.0

```
## Max. :77.00 Max. :3.000 Max. :200.0
## chol fbs restecg thalach
## Min. :126.0 Min. :0.0000 Min. :0.0000 Min. : 71.0
## 1st Qu.:211.0 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:133.5
## Median :240.0 Median :0.0000 Median :1.0000 Median :153.0
## Mean :246.3 Mean :0.1485 Mean :0.5281 Mean :149.6
## 3rd Qu.:274.5 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:166.0
## Max. :564.0 Max. :1.0000 Max. :2.0000 Max. :202.0
## exang oldpeak slope ca
## Min. :0.0000 Min. :0.00 Min. :0.000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.00 1st Qu.:1.000 1st Qu.:0.0000
## Median :0.0000 Median :0.80 Median :1.000 Median :0.0000
## Mean :0.3267 Mean :1.04 Mean :1.399 Mean :0.7294
## 3rd Qu.:1.0000 3rd Qu.:1.60 3rd Qu.:2.000 3rd Qu.:1.0000
## Max. :1.0000 Max. :6.20 Max. :2.000 Max. :4.0000
## thal target
## Min. :0.000 Min. :0.0000
## 1st Qu.:2.000 1st Qu.:0.0000
## Median :2.000 Median :1.0000
## Mean :2.314 Mean :0.5446
## 3rd Qu.:3.000 3rd Qu.:1.0000
## Max. :3.000 Max. :1.0000
```

```
tbl_df(hci) # A nicer view of the data as a table
```

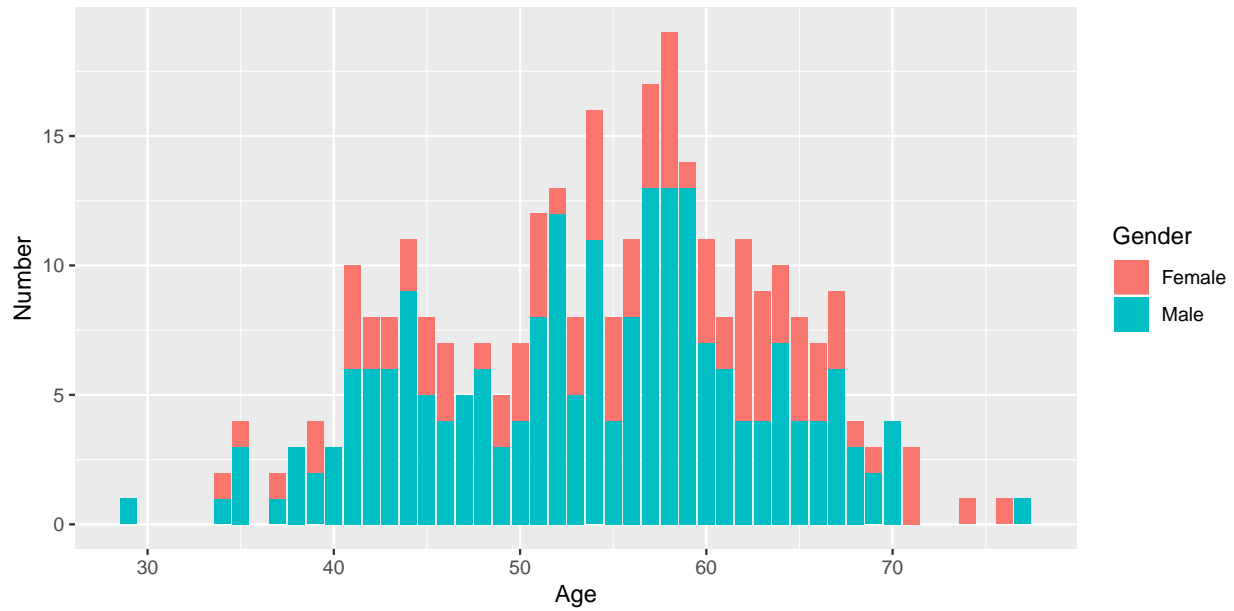
```
## # A tibble: 303 x 14
## age sex cp trestbps chol fbs restecg thalach exang oldpeak
## <int> <chr> <int> <int> <int> <int> <int> <int> <int> <dbl>
## 1 63 Male 3 145 233 1 0 150 0 2.3
## 2 37 Male 2 130 250 0 1 187 0 3.5
## 3 41 Fema~ 1 130 204 0 0 172 0 1.4
## 4 56 Male 1 120 236 0 1 178 0 0.8
## 5 57 Fema~ 0 120 354 0 1 163 1 0.6
## 6 57 Male 0 140 192 0 1 148 0 0.4
## 7 56 Fema~ 1 140 294 0 0 153 0 1.3
## 8 44 Male 1 120 263 0 1 173 0 0
## 9 52 Male 2 172 199 1 1 162 0 0.5
## 10 57 Male 2 150 168 0 1 174 0 1.6
## # ... with 293 more rows, and 4 more variables: slope <int>, ca <int>,
## # thal <int>, target <int>
```

Convert following predictors as factor for plotting

```
#Convert following predictors as factor for plotting
hci$sex<-as.factor(hci$sex)
hci$cp<-as.factor(hci$cp)
hci$thal<-as.factor(hci$thal)
hci$ca<-as.factor(hci$ca)
```

Distribution of Male and Female population across Age parameter

```
p1<-hci %>% ggplot(aes(x=age,fill=sex))+geom_bar()+xlab("Age") +
  ylab("Number")+ guides(fill = guide_legend(title = "Gender"))
plot(p1)
```



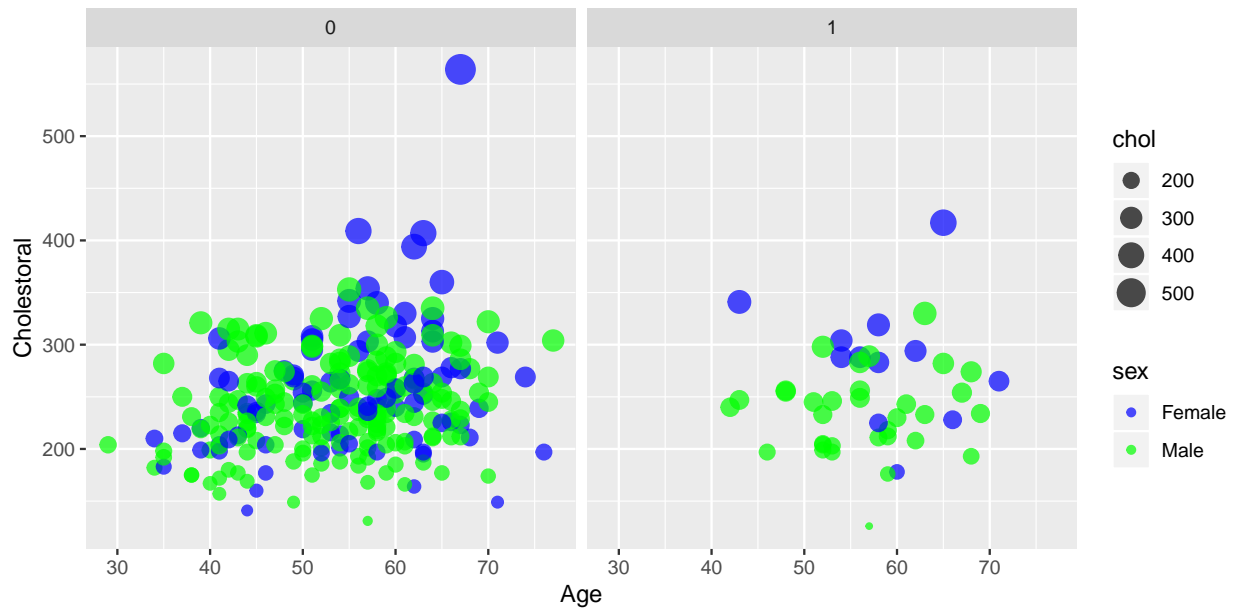
Representation of Cholestoral level

```
p2<-hci %>% ggplot(aes(x=age,y=chol,col=sex, size=chol))+geom_point(alpha=0.4)+xlab("Age") +
  ylab("Cholestoral")+guides(fill = guide_legend(title = "Gender"))+scale_colour_manual(values
plot(p2)
```



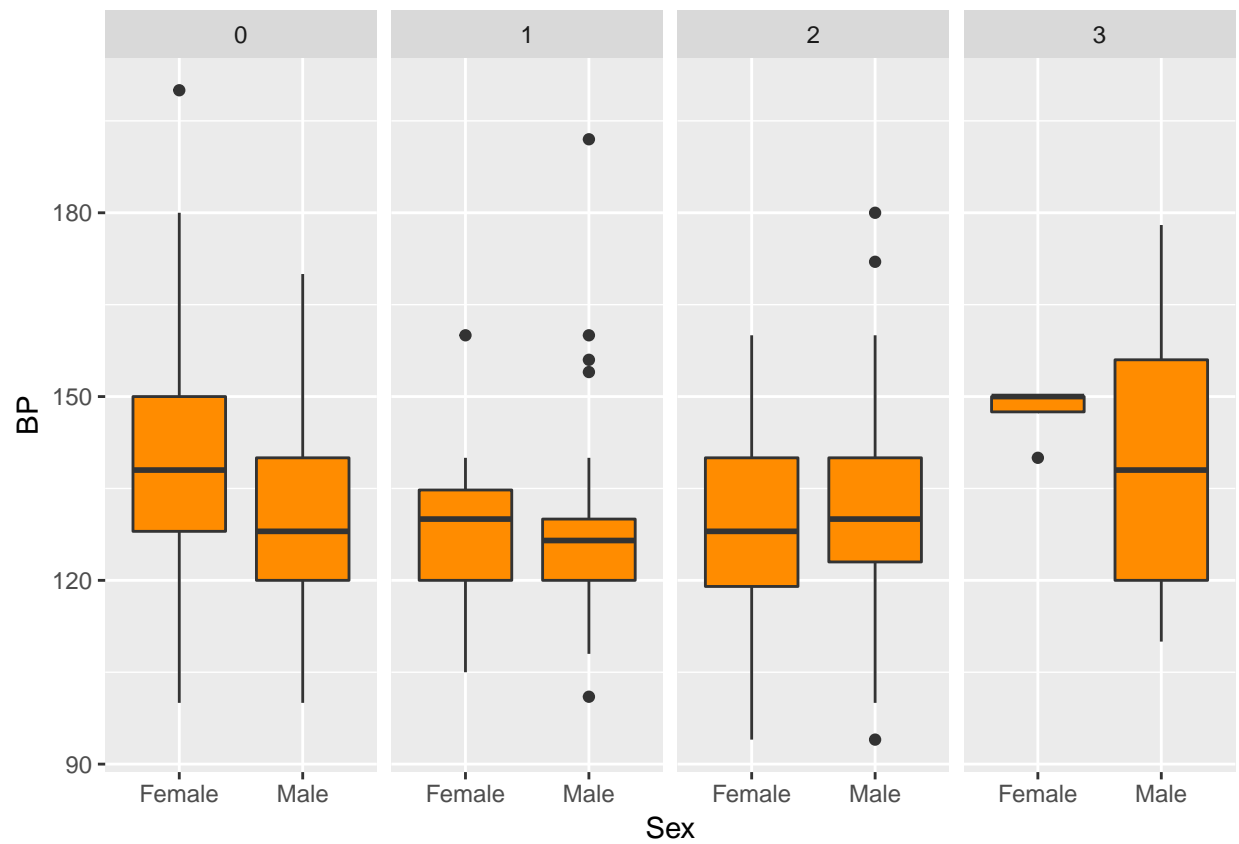
Representation of Cholestoral level across different defect conditions

```
p3<-hci %>% ggplot(aes(x=age,y=chol,col=sex, size=chol))+geom_point(alpha=0.7)+xlab("Age") +
  ylab("Cholestoral")+facet_grid(.~fbs)+scale_colour_manual(values = c( "blue", "green"))
plot(p3)
```



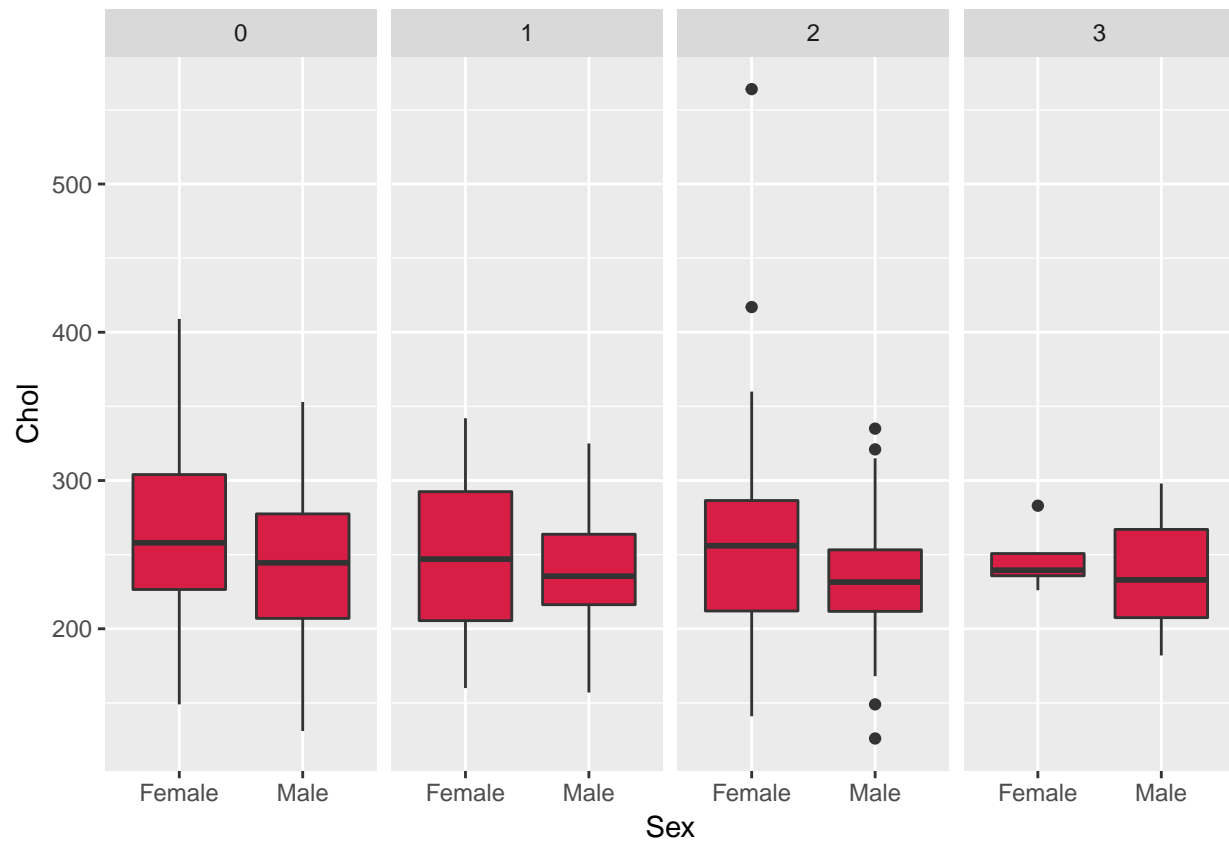
Comparison of Blood pressure across pain type (0~3)

```
p4<-hcai%>%ggplot(aes(x=sex,y=trestbps))+geom_boxplot(fill="darkorange")+xlab("Sex")+ylab("BP")+facet_grid(pain_type~.)
```



Comparison of Cholestoral across pain type (0~3)

```
p5<-hci%>%ggplot(aes(x=sex,y=chol))+geom_boxplot(fill="#D81E44")+xlab("Sex")+ylab("Chol")+facet_grid(~c
plot(p5)
```



Relation between Gender, Age, Cholestoral, BP

```
# Scatterplot
gg <- ggplot(hci, aes(x=age, y=chol, col=sex)) +
  geom_point(aes( size=trestbps),shape=1,alpha=0.6) +
  geom_smooth(method="loess", se=F) +scale_colour_manual(values = c( "red", "blue"))
plot(gg)
```

