Portfolio

Jehwan Bang

18/07/2021

AGE - Age of the patient
SEX - Sex of the patient, where (0:female , 1: male)
CP - Chest pain type, where (0: typical angina, 1: atypical angina, 2: non-anginal pain, 3: asymptomatic)
TRTBPS - Resting blood pressure in (MM|HG)
CHOL - Cholestrol in (MG|HG) fetched via bmi sensor
FBS - Prediabetes (fasting blood sugar > 120 MG/DL), where (0:false, 1:true)
RESTECG - Resting electrocardiographic results, where (0: normal, 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
THALACH -Maximum heart rate achieved
EXANG - Exercise induced angina, where (0: no, 1: yes)
OLDPEAK - Previous peak
SLP - Slope
CAA - Number of major vessels (0-4)
THALL - Thalium stress test result (0,3)
OUTPUT - Target variable, where (0: less chance of heart attack, 1: more chance of heart attack)

install.packages("dplyr")
install.packages("ggplot2")
install.packages("highcharter")

```
nm <- suppressPackageStartupMessages  # no message
setwd("C:/Users/je111/Downloads")
nm(library(dplyr))
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
nm(library(ggplot2))
nm(library(highcharter))
```

```
## Warning: package 'highcharter' was built under R version 3.6.3
```

```
heart = read.csv("heart.csv",header=TRUE)
head(heart)
```

	age <int>	sex <int>	cp <int>	trtbps <int>	chol <int>	fbs <int>	restecg <int>	thalachh <int>	exng <int>
1	63	1	3	145	233	1	0	150	0
2	37	1	2	130	250	0	1	187	0
3	41	0	1	130	204	0	0	172	0
4	56	1	1	120	236	0	1	178	0
5	57	0	0	120	354	0	1	163	1
6	57	1	0	140	192	0	1	148	0

6 rows | 1-10 of 15 columns

```
summary(heart)
```

```
##       age             sex               cp            trtbps     
##  Min.   :29.00   Min.   :0.0000   Min.   :0.000   Min.   : 94.0  
##  1st Qu.:47.50   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:120.0  
##  Median :55.00   Median :1.0000   Median :1.000   Median :130.0  
##  Mean   :54.37   Mean   :0.6832   Mean   :0.967   Mean   :131.6  
##  3rd Qu.:61.00   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:140.0  
##  Max.   :77.00   Max.   :1.0000   Max.   :3.000   Max.   :200.0  
##       chol            fbs             restecg          thalachh    
##  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  
##       exng           oldpeak          slp              caa        
##  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  
##      thall           output     
##  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
```

```
str(heart)
```

```
## 'data.frame':    303 obs. of  14 variables:
##  $ age     : int  63 37 41 56 57 57 56 44 52 57 ...
##  $ sex     : int  1 1 0 1 0 1 0 1 1 1 ...
##  $ cp      : int  3 2 1 1 0 0 1 1 2 2 ...
##  $ trtbps  : int  145 130 130 120 120 140 140 120 172 150 ...
##  $ chol    : int  233 250 204 236 354 192 294 263 199 168 ...
##  $ fbs     : int  1 0 0 0 0 0 0 0 1 0 ...
##  $ restecg : int  0 1 0 1 1 1 0 1 1 1 ...
##  $ thalachh: int  150 187 172 178 163 148 153 173 162 174 ...
##  $ exng    : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ oldpeak : num  2.3 3.5 1.4 0.8 0.6 0.4 1.3 0 0.5 1.6 ...
##  $ slp     : int  0 0 2 2 2 1 1 2 2 2 ...
##  $ caa     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ thall   : int  1 2 2 2 2 1 2 3 3 2 ...
##  $ output  : int  1 1 1 1 1 1 1 1 1 1 ...
```

Data Type is int except for the oldpeak which is num.

```
glimpse(heart)
```

```
## Rows: 303
## Columns: 14
## $ age      <int> 63, 37, 41, 56, 57, 57, 56, 44, 52, 57, 54, 48, 49, 64, ~
## $ sex      <int> 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1,~
## $ cp       <int> 3, 2, 1, 1, 0, 0, 1, 1, 2, 2, 0, 2, 1, 3, 3, 2, 2, 3, 0,~
## $ trtbps   <int> 145, 130, 130, 120, 120, 140, 140, 120, 172, 150, 140, 1~
## $ chol     <int> 233, 250, 204, 236, 354, 192, 294, 263, 199, 168, 239, 2~
## $ fbs      <int> 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,~
## $ restecg  <int> 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1,~
## $ thalachh <int> 150, 187, 172, 178, 163, 148, 153, 173, 162, 174, 160, 1~
## $ exng     <int> 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,~
## $ oldpeak  <dbl> 2.3, 3.5, 1.4, 0.8, 0.6, 0.4, 1.3, 0.0, 0.5, 1.6, 1.2, 0~
## $ slp      <int> 0, 0, 2, 2, 2, 1, 1, 2, 2, 2, 2, 2, 2, 1, 2, 1, 2, 0, 2,~
## $ caa      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ thall    <int> 1, 2, 2, 2, 2, 1, 2, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ output   <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
```

```
heart <- unique(heart)
glimpse(heart)
```

```
## Rows: 302
## Columns: 14
## $ age      <int> 63, 37, 41, 56, 57, 57, 56, 44, 52, 57, 54, 48, 49, 64, ~
## $ sex      <int> 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1,~
## $ cp       <int> 3, 2, 1, 1, 0, 0, 1, 1, 2, 2, 0, 2, 1, 3, 3, 2, 2, 3, 0,~
## $ trtbps   <int> 145, 130, 130, 120, 120, 140, 140, 120, 172, 150, 140, 1~
## $ chol     <int> 233, 250, 204, 236, 354, 192, 294, 263, 199, 168, 239, 2~
## $ fbs      <int> 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,~
## $ restecg  <int> 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1,~
## $ thalachh <int> 150, 187, 172, 178, 163, 148, 153, 173, 162, 174, 160, 1~
## $ exng     <int> 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,~
## $ oldpeak  <dbl> 2.3, 3.5, 1.4, 0.8, 0.6, 0.4, 1.3, 0.0, 0.5, 1.6, 1.2, 0~
## $ slp      <int> 0, 0, 2, 2, 2, 1, 1, 2, 2, 2, 2, 2, 2, 1, 2, 1, 2, 0, 2,~
## $ caa      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ thall    <int> 1, 2, 2, 2, 2, 1, 2, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ output   <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
```

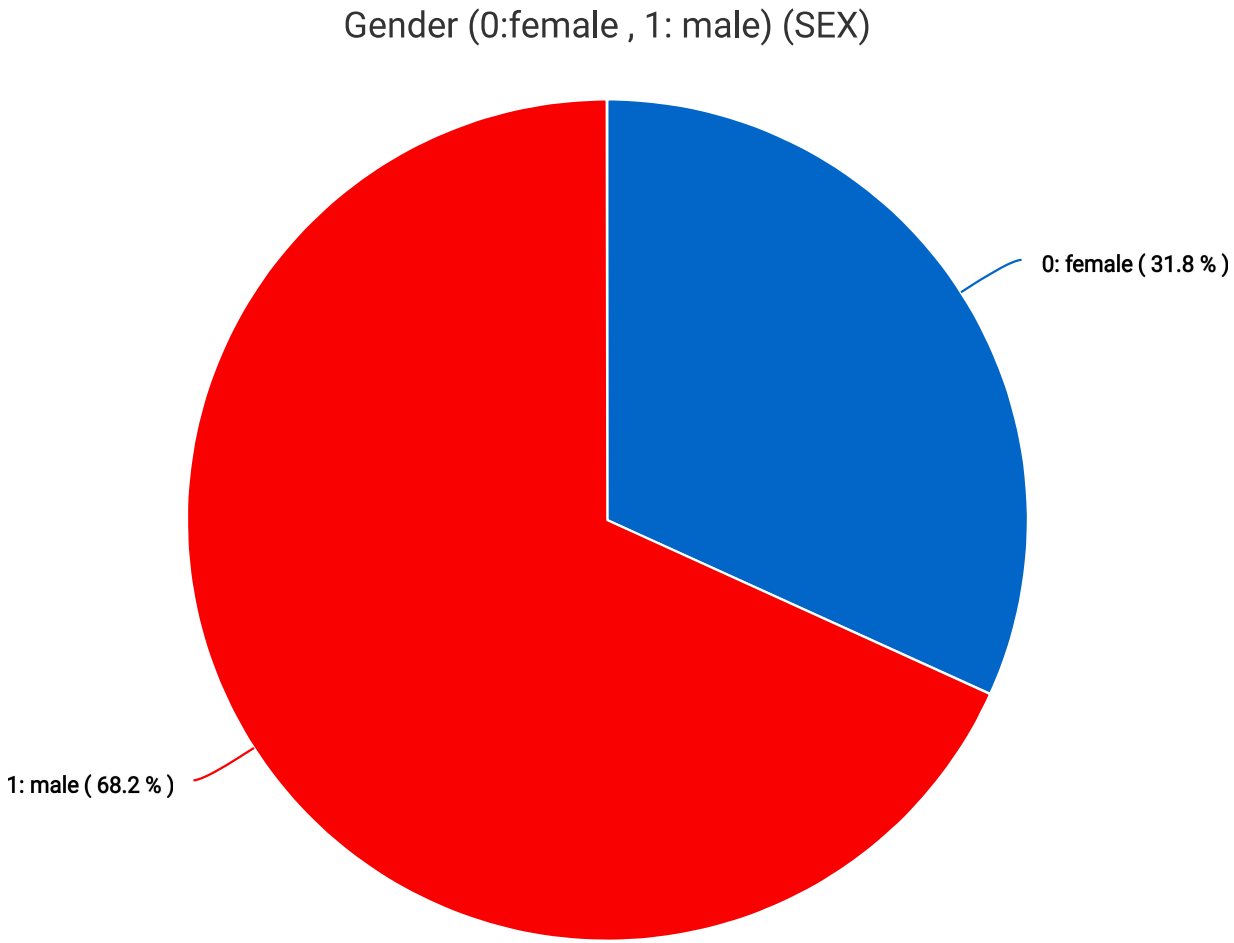
We do see that one duplicate is removed from 303 Rows to 302 Rows.

```
nu_var <- heart %>% dplyr::select("age","trtbps","chol","thalachh","oldpeak")
cat_var <- heart %>% dplyr::select("sex","cp","fbs","restecg","exng","slp","caa","thall","output")

names <- c(1:9)
cat_var[,names] <-lapply(cat_var[,names],factor)
str(cat_var)
```

```
## 'data.frame':    302 obs. of  9 variables:
##  $ sex    : Factor w/ 2 levels "0","1": 2 2 1 2 1 2 1 2 2 2 ...
##  $ cp     : Factor w/ 4 levels "0","1","2","3": 4 3 2 2 1 1 2 2 3 3 ...
##  $ fbs    : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 1 2 1 ...
##  $ restecg: Factor w/ 3 levels "0","1","2": 1 2 1 2 2 2 1 2 2 2 ...
##  $ exng   : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...
##  $ slp    : Factor w/ 3 levels "0","1","2": 1 1 3 3 3 2 2 3 3 3 ...
##  $ caa    : Factor w/ 5 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ thall  : Factor w/ 4 levels "0","1","2","3": 2 3 3 3 3 2 3 4 4 3 ...
##  $ output : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
```

SEX - Sex of the patient, where (0:female , 1: male)

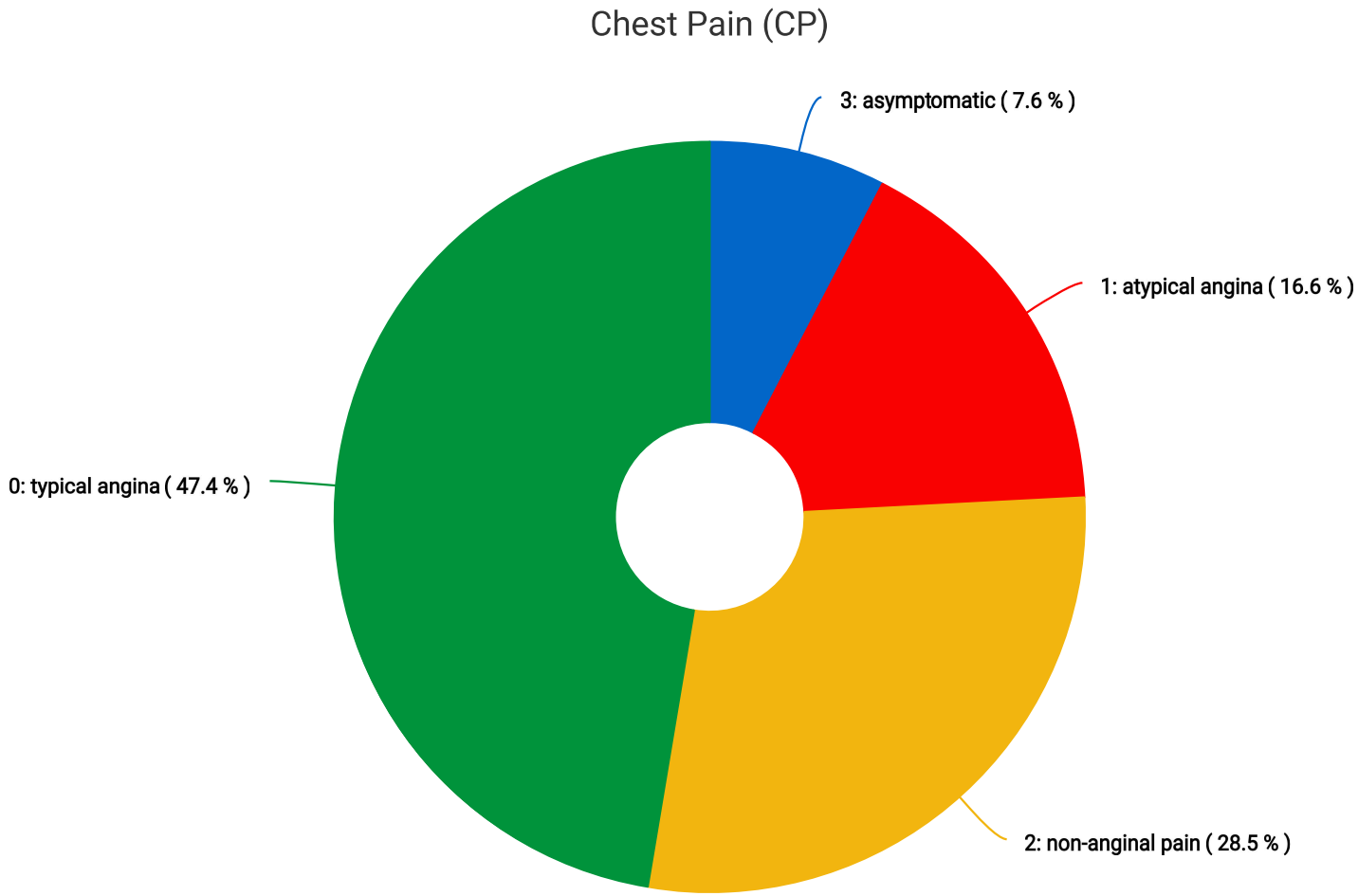
```
new_sex_str = function(bool_var){
  char_list = c(1, length(bool_var))
  for (i in 1:length(bool_var)) {
    var = bool_var[i]
    nchar = 'unknown'

    if(var == 1){
        nchar = '1: male'}
    else{
        nchar = '0: female'}

     char_list[i] = nchar}

    return(char_list)
}

heart_sex <- heart %>%
    select(sex) %>%
    group_by(sex) %>%
    summarise(Count = n()) %>%
    arrange(Count) %>%
    mutate(
        gender = new_sex_str(sex),
        percentage = paste(round(Count / sum(Count) * 100, 1), "%")
    )


heart_sex %>%
  hchart(type = "pie", hcaes(x = paste(gender, ' \t(', percentage, ')' ), y = Count)) %>%
    hc_xAxis(categories = paste(heart_sex$gender, ' (', heart_sex$percentage, ')' )) %>%
    hc_title(text = "Gender (0:female , 1: male) (SEX)") %>%
    hc_add_theme(hc_theme_google())
```



Chest Pain (CP)

CP - Chest pain type, where (0: typical angina, 1: atypical angina, 2: non-anginal pain, 3: asymptomatic)

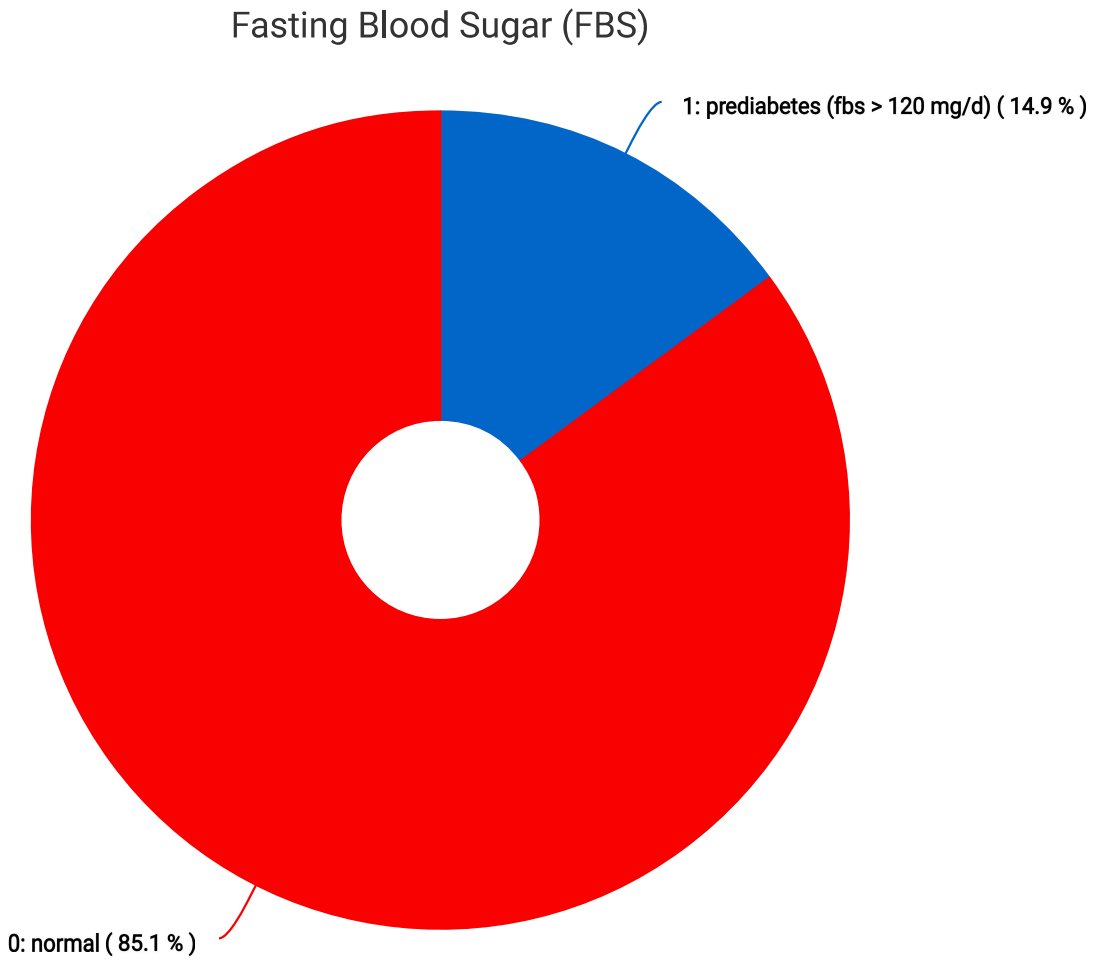
```
new_cp_str = function(bool_var){
  char_list = c(1, length(bool_var))
  for (i in 1:length(bool_var)) {
    var = bool_var[i]
    nchar = 'unkown'

    if(var == 0){
        nchar = '0: typical angina'}

    if(var == 1){
        nchar = '1: atypical angina'}

    if(var == 2){
        nchar = '2: non-anginal pain'}

    if(var == 3){
        nchar = '3: asymptomatic'}

     char_list[i] =  nchar}

    return(char_list)
 }


heart_cp <- heart %>%
    select(cp) %>%
    group_by(cp) %>%
    summarise(Count = n()) %>%
    arrange(Count) %>%
    mutate(cp_str = new_cp_str(cp),
        percentage = paste(round(Count / sum(Count) * 100, 1), "%"))

heart_cp %>%
  hchart(type = "pie", hcaes(x = paste(cp_str, ' \t(', percentage, ')' ), y = Count)) %>%
    hc_xAxis(categories = paste(heart_cp$cp_str, ' (', heart_cp$percentage, ')' )) %>%
    hc_title(text = 'Chest Pain (CP)') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE, alpha=0, beta=0)) %>%
    hc_plotOptions(pie=list(innerSize= 100, depth= 50))
```



Fasting Blood Suger (FBS)

FBS - Prediabetes (fasting blood sugar > 120 MG/DL), where (0:false, 1:true)

```
new_fbs_str = function(bool_var){
  char_list = c(1, length(bool_var))
  for (i in 1:length(bool_var)) {
    var = bool_var[i]
    nchar = 'unknown'

    if(var == 1){
        nchar = '1: prediabetes (fbs > 120 mg/d)'}
    else{
        nchar = '0: normal'}

     char_list[i] =  nchar}

    return(char_list)
}



heart_fbs <- heart %>%
    select(fbs) %>%
    group_by(fbs) %>%
    summarise(Count = n()) %>%
    arrange(Count) %>%
    mutate(fbs_str = new_fbs_str(fbs),
        percentage = paste(round(Count / sum(Count) * 100, 1), "%")
    )

heart_fbs %>%
  hchart(type = "pie", hcaes(x = paste(fbs_str, ' \t(', percentage, ')' ), y = Count)) %>%
    hc_xAxis(categories = paste(new_fbs_str(heart_fbs$fbs_str), ' (', heart_fbs$percentage, ')' )) %>%
    hc_title(text = 'Fasting Blood Sugar (FBS)') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE, alpha=0)) %>%
    hc_plotOptions(pie=list(innerSize= 100,
                            depth= 50))
```

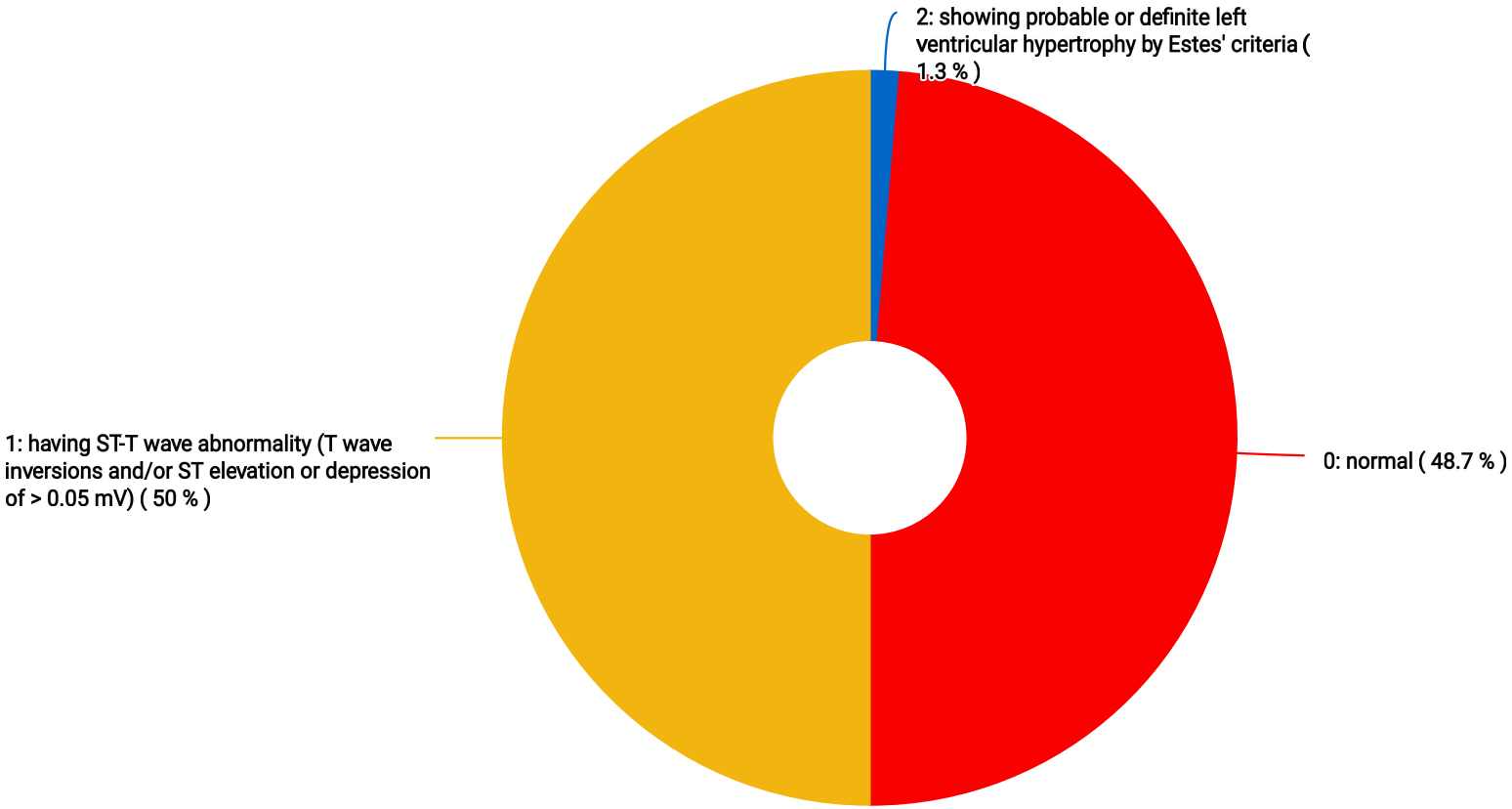


Resting Electrocardiographic Results (RESTECG)

RESTECG - Resting electrocardiographic results, where (0: normal, 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), 2: showing probable or definite left ventricular hypertrophy by Estes' criteria

```
new_restecg_str = function(bool_var){
  char_list = c(1, length(bool_var))
  for (i in 1:length(bool_var)) {
    var = bool_var[i]
    nchar = 'unknown'

    if(var == 0){
        nchar = '0: normal'}
    if(var == 1){
        nchar = '1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)'}
    if(var == 2){
        nchar = "2: showing probable or definite left ventricular hypertrophy by Estes' criteria"}

     char_list[i] =  nchar
   }

    return(char_list)
 }

heart_restecg <- heart %>%
    select(restecg) %>%
    group_by(restecg) %>%
    summarise(Count = n()) %>%
    arrange(Count) %>%
    mutate(restecg_str = new_restecg_str(restecg),
        percentage = paste(round(Count / sum(Count) * 100, 1), "%")
    )

heart_restecg %>%
  hchart(type = "pie", hcaes(x = paste(restecg_str, ' \t(', percentage, ')' ), y = Count)) %>%
    hc_xAxis(categories = paste(heart_restecg$restecg_str, ' (', heart_restecg$percentage, ')' )) %>%
    hc_title(text = 'Resting Electrocardiographic Results (RESTECG)') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE, alpha=0)) %>%
    hc_plotOptions(pie=list(innerSize= 100,
                            depth= 50))
```

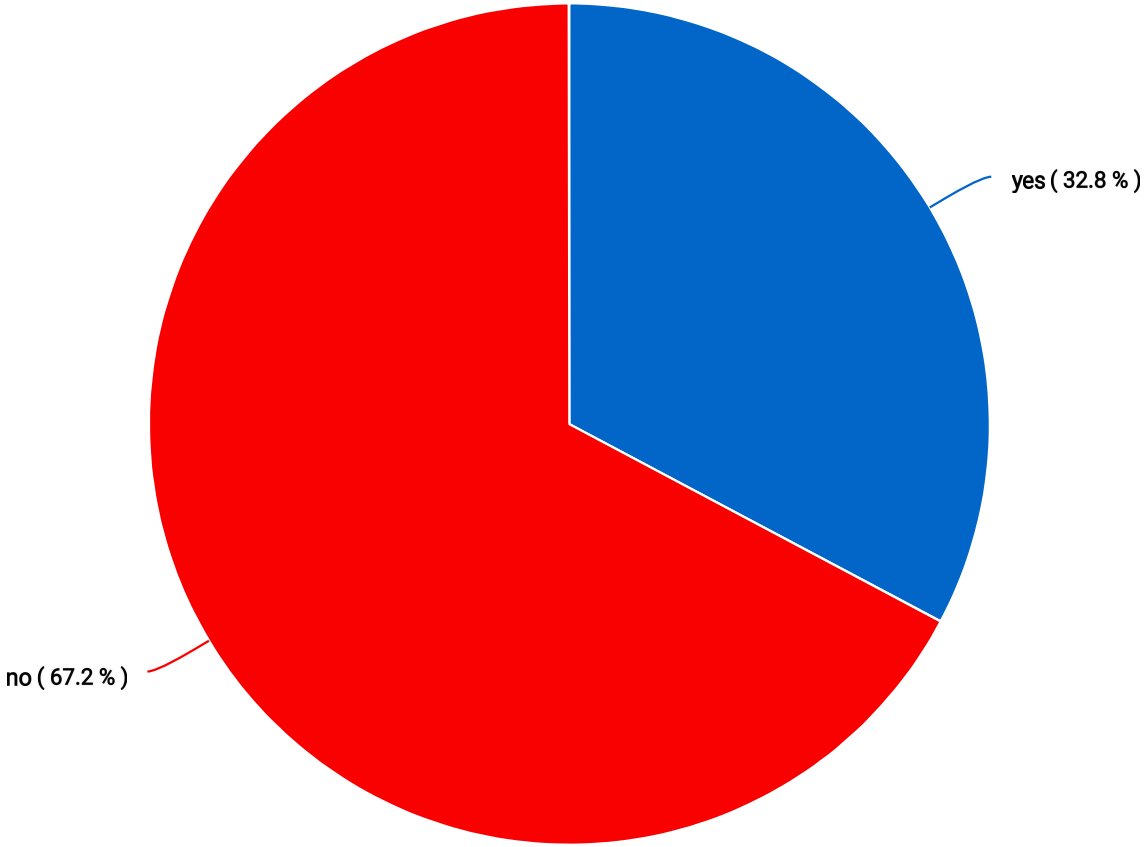
Resting Electrocardiographic Results (RESTECG)



EXANG - Exercise induced angina, where (0: no, 1: yes)

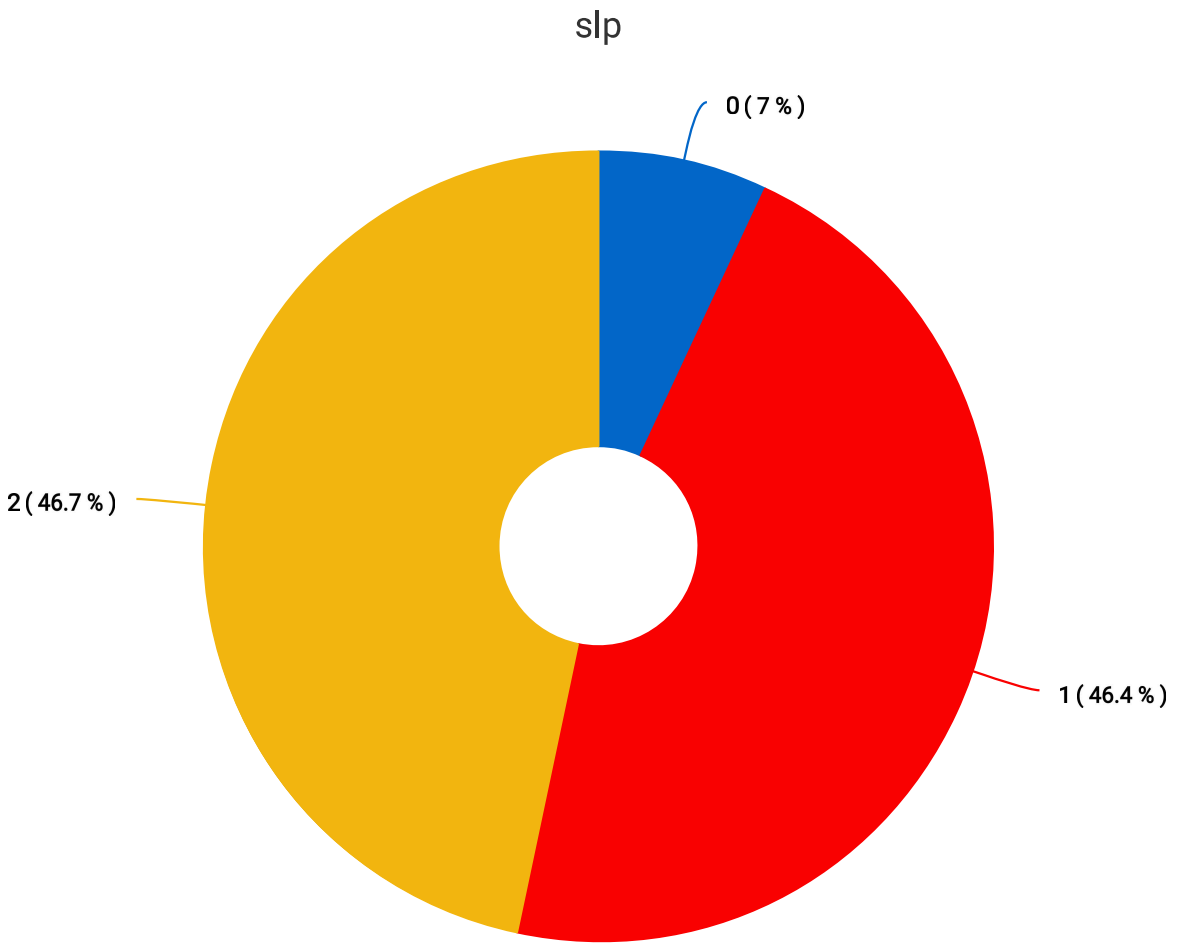
```
new_exng_str = function(bool_var){
  char_list = c(1, length(bool_var))
  for (i in 1:length(bool_var)) {
    var = bool_var[i]
    nchar = 'unknown'

    if(var == 1){
        nchar = 'yes'}
    else{
        nchar = 'no'}

     char_list[i] =  nchar}

    return(char_list)
 }

heart_exng <- heart %>%
    select(exng) %>%
    group_by(exng) %>%
    summarise(Count = n()) %>%
    arrange(Count) %>%
    mutate(
        exng_txt = new_exng_str(exng),
        percentage = paste(round(Count / sum(Count) * 100, 1), "%")
    )
heart_exng %>%
  hchart(type = "pie", hcaes(x = paste(exng_txt, ' \t(', percentage, ')' ), y = Count)) %>%
    hc_xAxis(categories = paste(heart_exng$exng_txt, ' (', heart_exng$percentage, ')' )) %>%
    hc_title(text = 'Exercise induced angina, where (0: no, 1: yes) (EXNG)') %>%
    hc_add_theme(hc_theme_google())
```

Exercise induced angina, where (0: no, 1: yes) (EXNG)



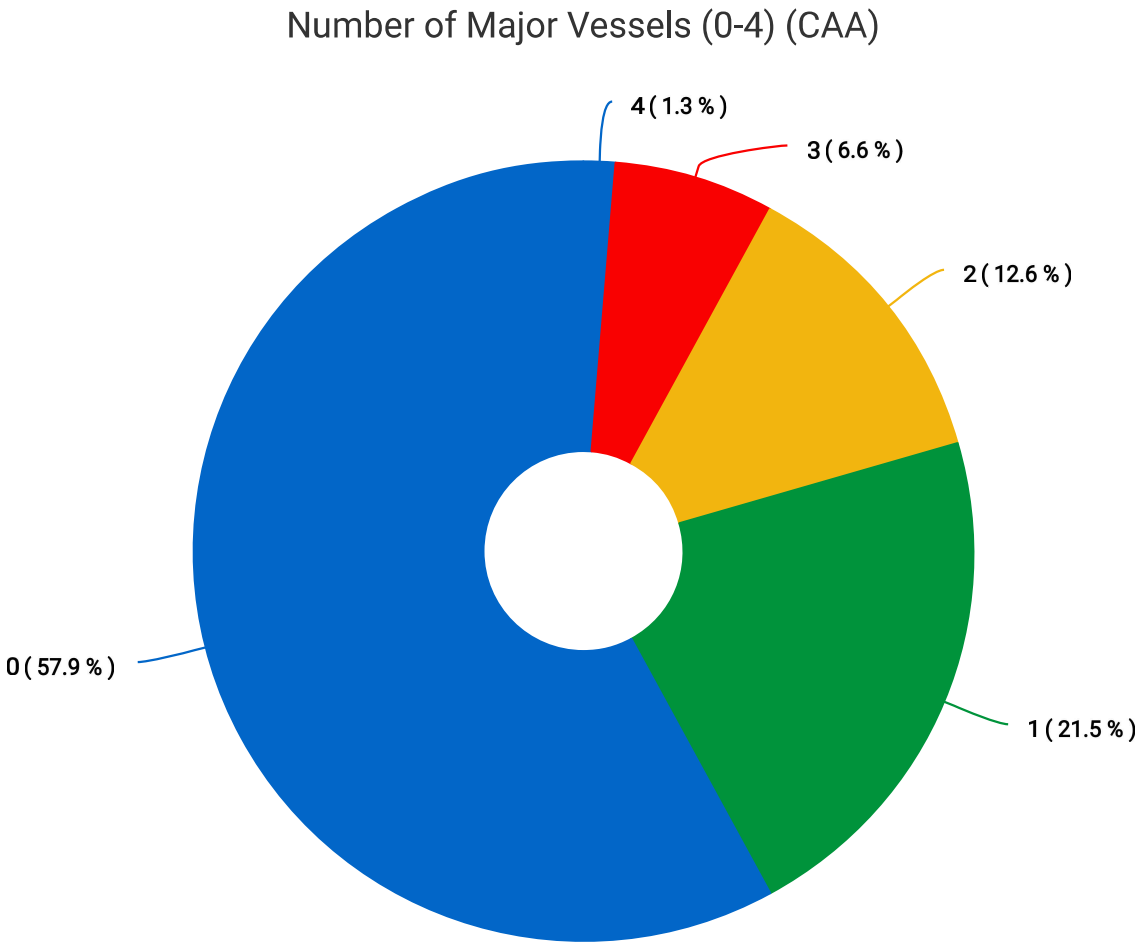
SLP - Slope

```
heart_slp <- heart %>%
    select(slp) %>%
    group_by(slp) %>%
    summarise(Count = n()) %>%
    arrange(Count) %>%
        mutate(
        percentage = paste(round(Count / sum(Count) * 100, 1), "%")
    )


heart_slp %>%
  hchart(type = "pie", hcaes(x = paste(slp, ' \t(', percentage, ')' ), y = Count)) %>%
    hc_xAxis(categories = paste(heart_slp$slp, ' (', heart_slp$percentage, ')' )) %>%
    hc_title(text = 'slp') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE, alpha=0)) %>%
    hc_plotOptions(pie=list(innerSize= 100,
                            depth= 50))
```



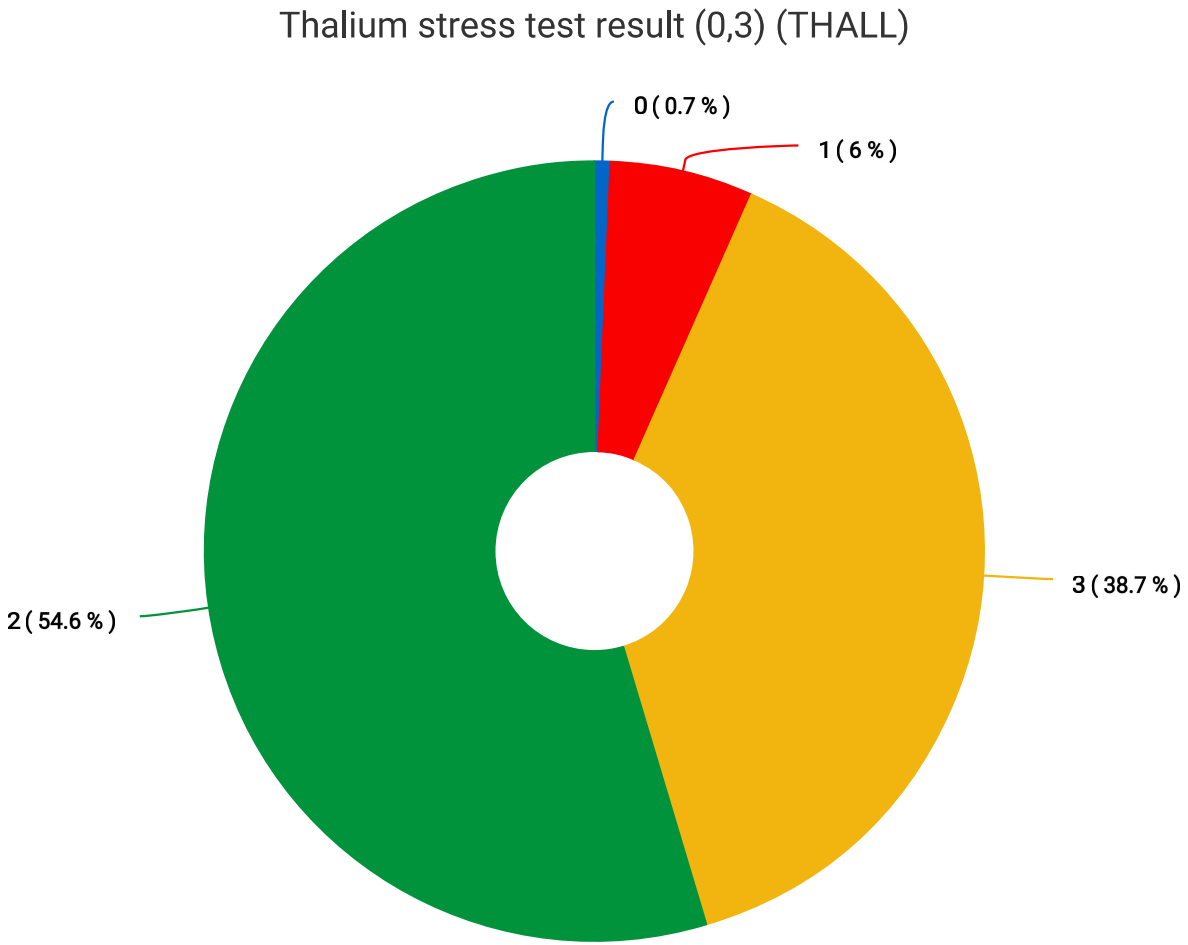
CAA - Number of major vessels (0-4)

```
heart_caa <- heart %>%
    select(caa) %>%
    group_by(caa) %>%
    summarise(Count = n()) %>%
    arrange(Count) %>%
        mutate(
        percentage = paste(round(Count / sum(Count) * 100, 1), "%")
    )


heart_caa %>%
  hchart(type = "pie", hcaes(x = paste(caa, ' \t(', percentage, ')' ), y = Count)) %>%
    hc_xAxis(categories = paste(heart_caa$caa, ' (', heart_caa$percentage, ')' )) %>%
    hc_title(text = 'Number of Major Vessels (0-4) (CAA)') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE, alpha=0)) %>%
    hc_plotOptions(pie=list(innerSize= 100,
                            depth= 50))
```



THALL - Thalium stress test result (0,3)

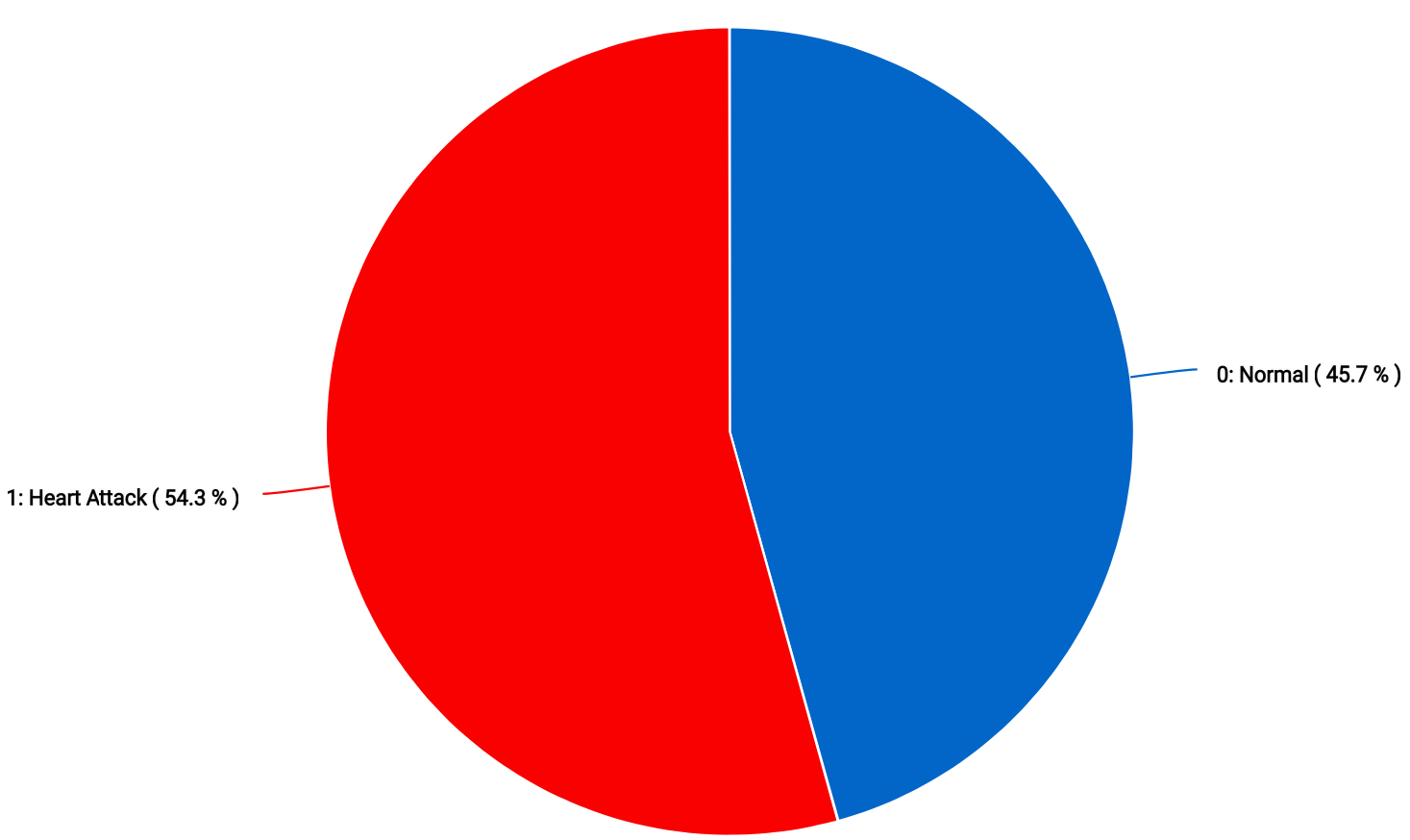
```
heart_thall <- heart %>%
    select(thall) %>%
    group_by(thall) %>%
    summarise(Count = n()) %>%
    arrange(Count) %>%
        mutate(
        percentage = paste(round(Count / sum(Count) * 100, 1), "%")
    )
heart_thall %>%
  hchart(type = "pie", hcaes(x = paste(thall, ' \t(', percentage, ')' ), y = Count)) %>%
    hc_xAxis(categories = paste(heart_thall$thall, ' (', heart_thall$percentage, ')' )) %>%
    hc_title(text = 'Thalium stress test result (0,3) (THALL)') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE, alpha=0)) %>%
    hc_plotOptions(pie=list(innerSize= 100,
                            depth= 50))
```



Heart Attack (OUTPUT)

OUTPUT - Target variable, where (0: less chance of heart attack, 1: more chance of heart attack)

```
new_output_str = function(bool_var){
  char_list = c(1, length(bool_var))
  for (i in 1:length(bool_var)) {
    var = bool_var[i]
    nchar = 'unknown'

    if(var == 0){
        nchar = '0: Normal'}

    if(var == 1){
        nchar = '1: Heart Attack'}

     char_list[i] =  nchar}

    return(char_list)
 }

heart_output <- heart %>%
    select(output) %>%
    group_by(output) %>%
    summarise(Count = n()) %>%
    arrange(Count) %>%
    mutate(
        heart_attack = new_output_str(output),
        percentage = paste(round(Count / sum(Count) * 100, 1), "%")
    )


heart_output %>%
  hchart(type = "pie", hcaes(x = paste(heart_attack, ' \t(', percentage, ')' ), y = Count)) %>%
    hc_xAxis(categories = paste(heart_output$heart_attack, ' (', heart_output$percentage, ')' )) %>%
    hc_title(text = 'Heart Attack (0: less chance, 1: more chance (OUTPUT)') %>%
    hc_add_theme(hc_theme_google())
```

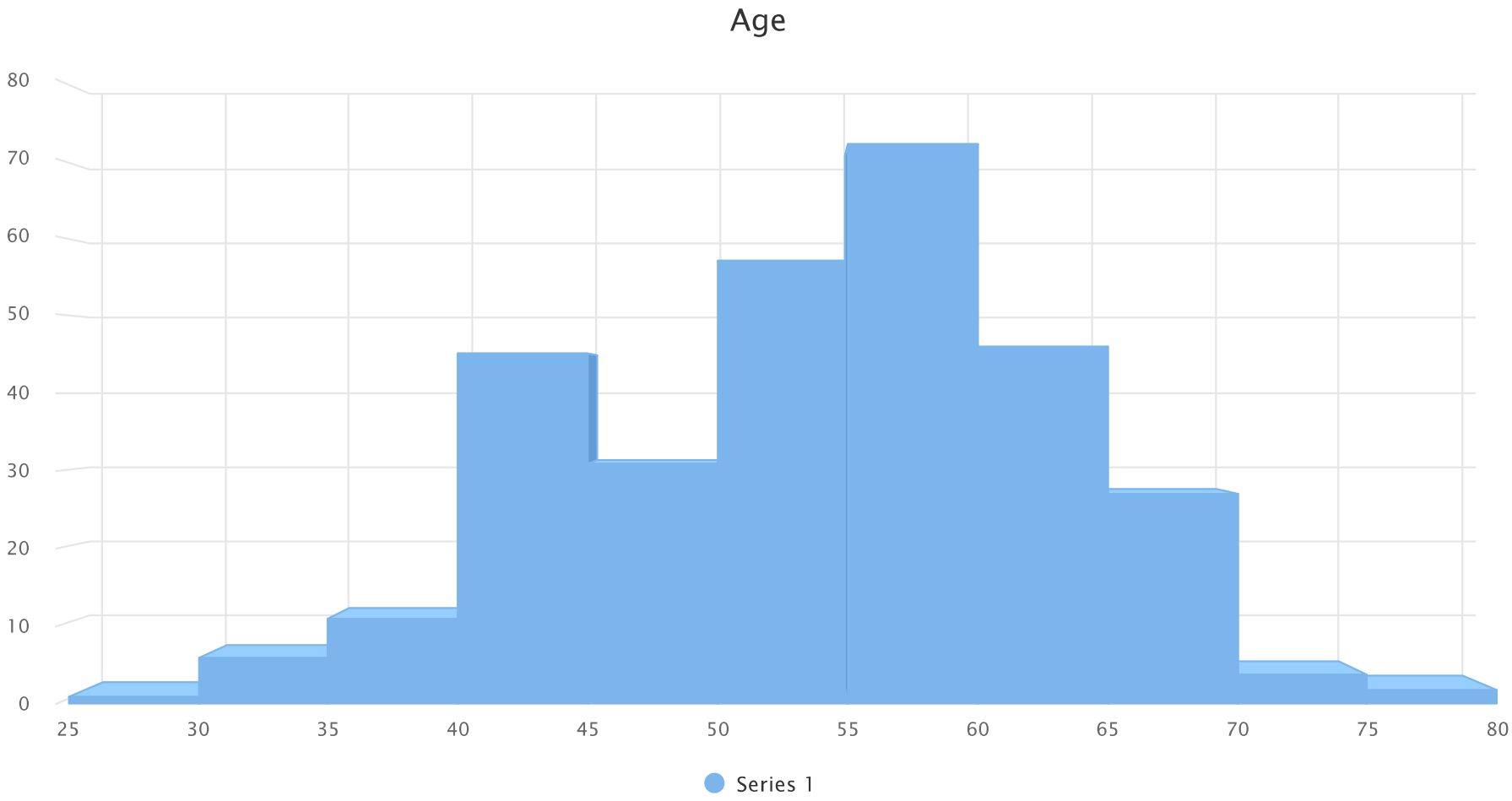

Heart Attack (0: less chance, 1: more chance (OUTPUT)
0: Normal (45.7 %)
1: Heart Attack (54.3 %)

```
print_histogram = function(name, title){
   keeps <- c(name, 'output')
   frame = heart[ , (names(heart) %in% keeps)]

  normal_chart <- hchart(frame[, (names(frame) %in% name)], type = "column") %>%
    hc_title(text= title) %>%
    hc_xAxis(title = name) %>%
    hc_chart(options3d=list(enabled=TRUE, alpha=0, beta=0,
                            depth=100, viewDistance=20)) %>%
    hc_plotOptions(column=list(depth= 100))

   htmltools::tagList(normal_chart)

}
```

```
summary(heart$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   29.00   48.00   55.50   54.42   61.00   77.00
```

```
print_histogram('age', 'Age')
```

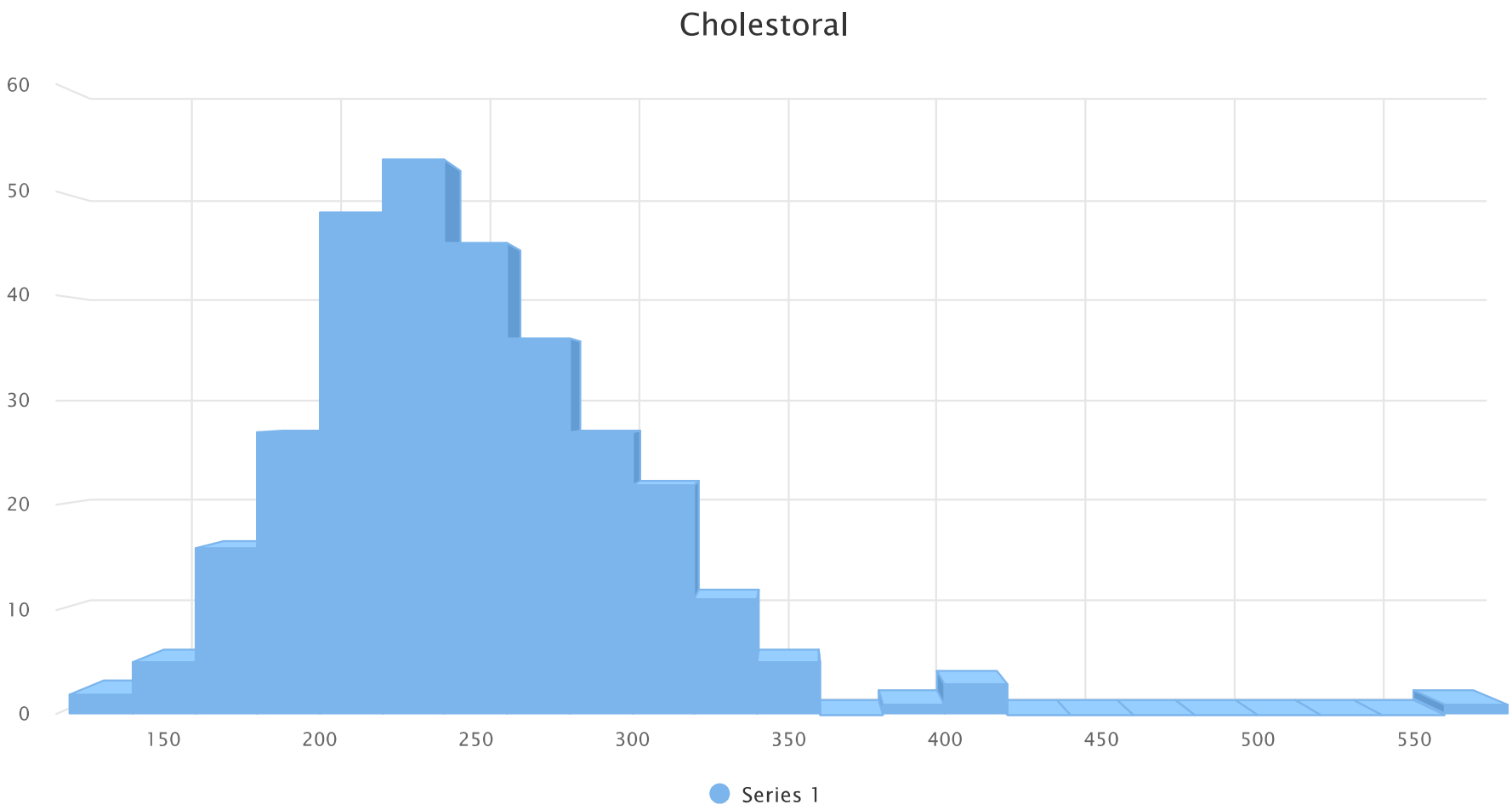


It is close to normally distributed

```
summary(heart$chol)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   126.0   211.0   240.5   246.5   274.8   564.0
```

```
print_histogram('chol', 'Cholestoral')
```

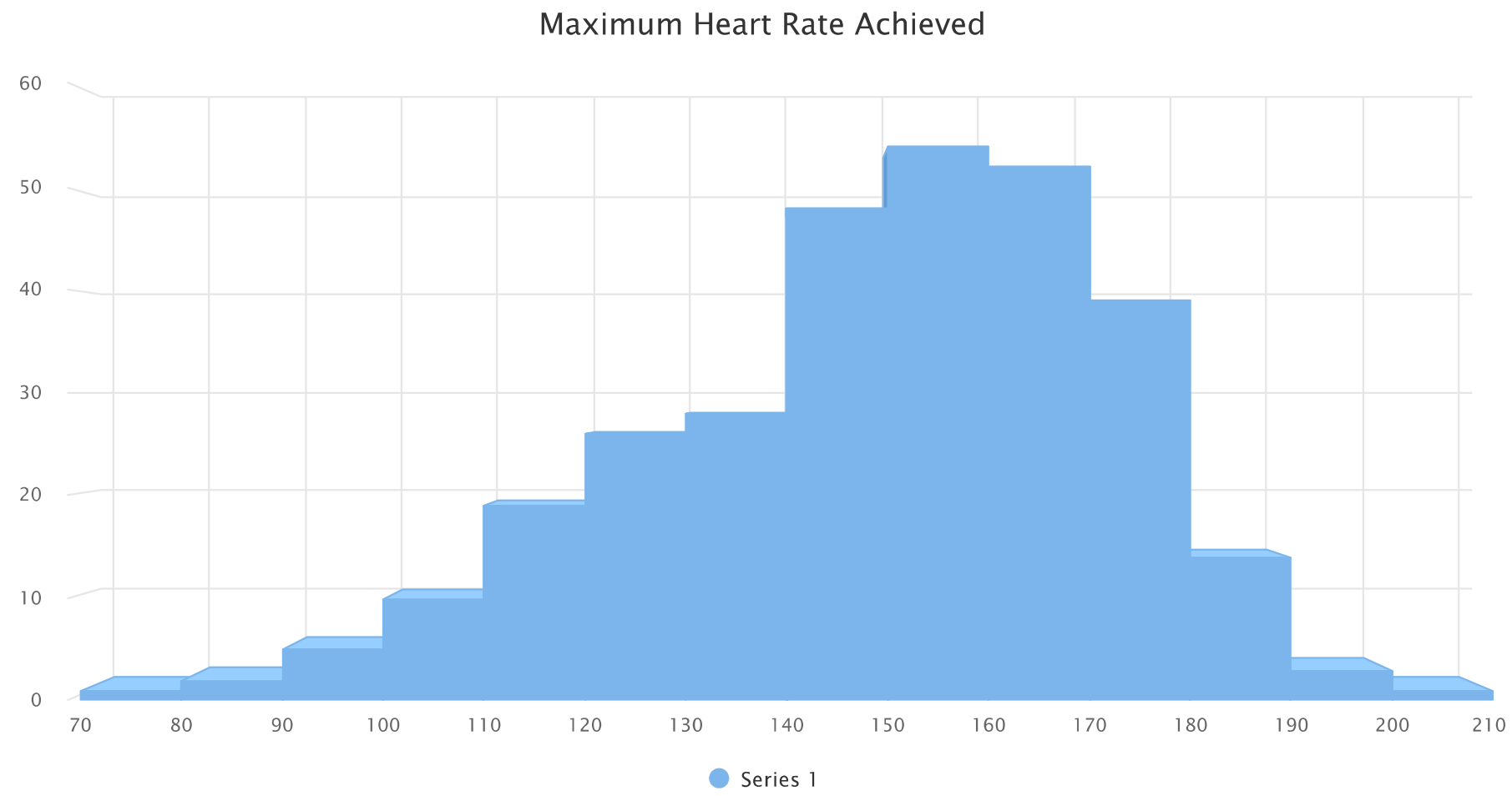


It is skewed to the right

```
summary(heart$thalachh)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    71.0   133.2   152.5   149.6   166.0   202.0
```

```
print_histogram('thalachh', 'Maximum Heart Rate Achieved')
```

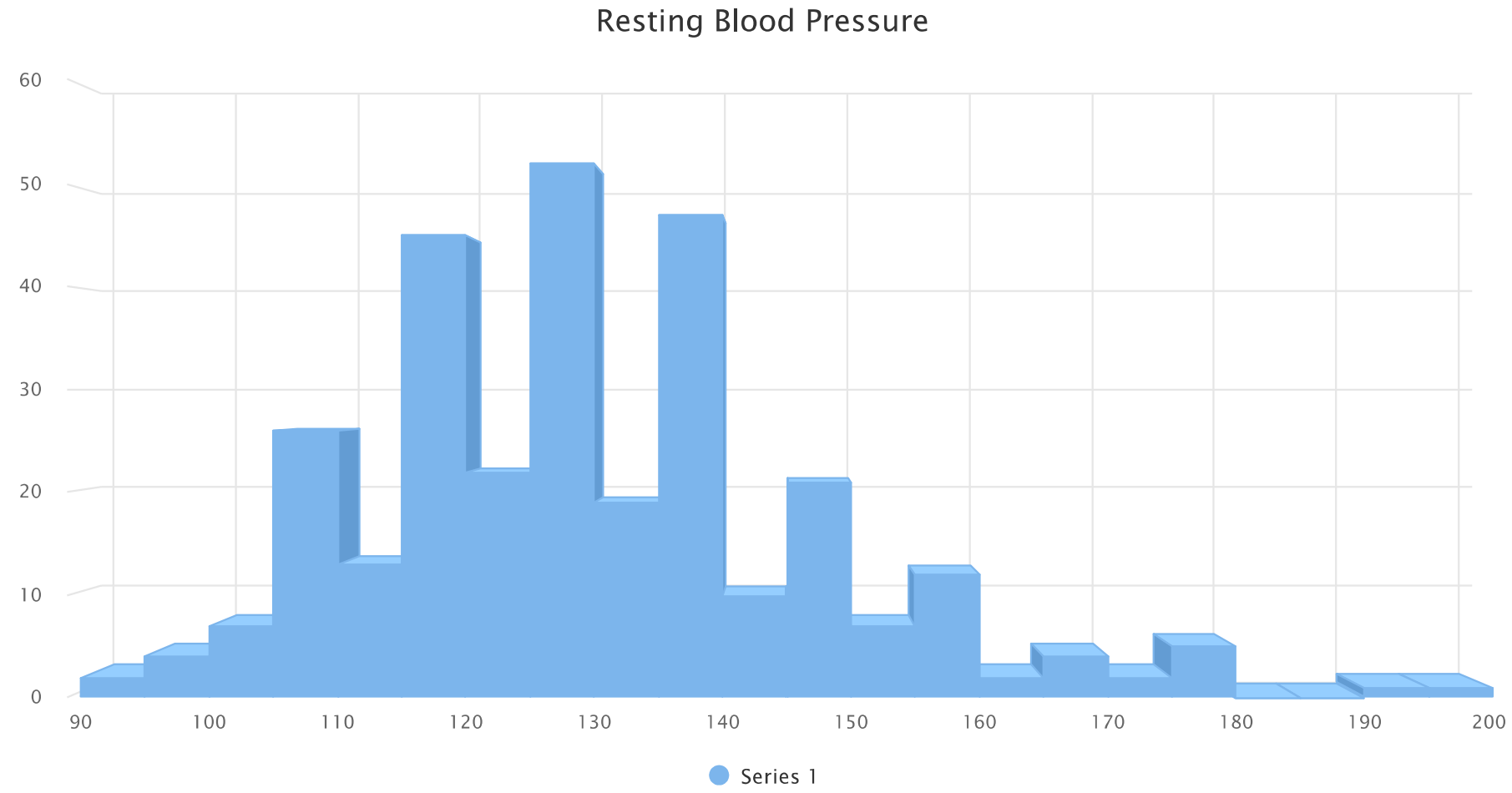


It is skewed to the left

```
summary(heart$trtbps)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    94.0   120.0   130.0   131.6   140.0   200.0
```

```
print_histogram('trtbps', 'Resting Blood Pressure')
```

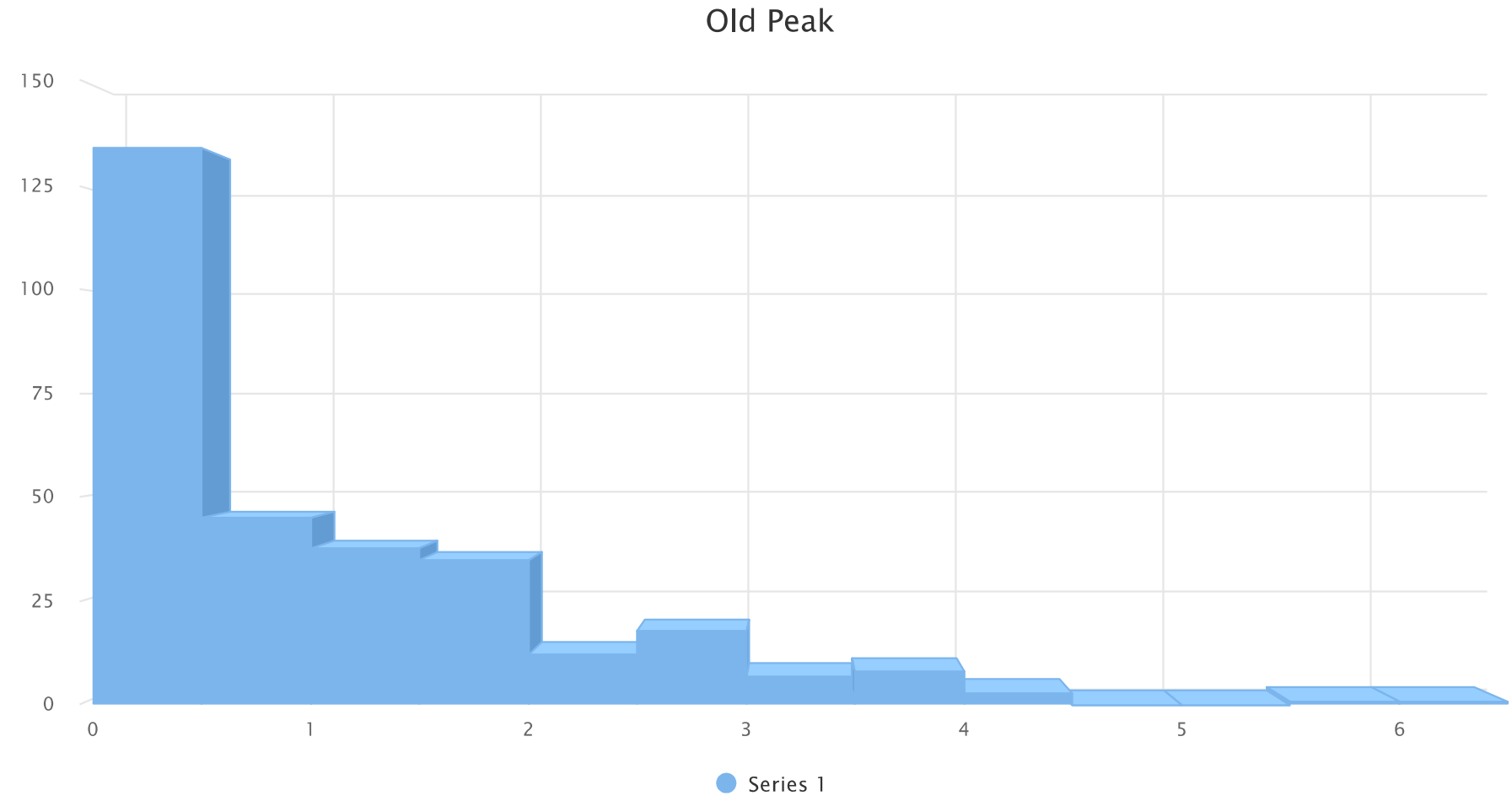


It is skewed to the right

```
summary(heart$oldpeak)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.800   1.043   1.600   6.200
```

```
print_histogram('oldpeak', 'Old Peak')
```



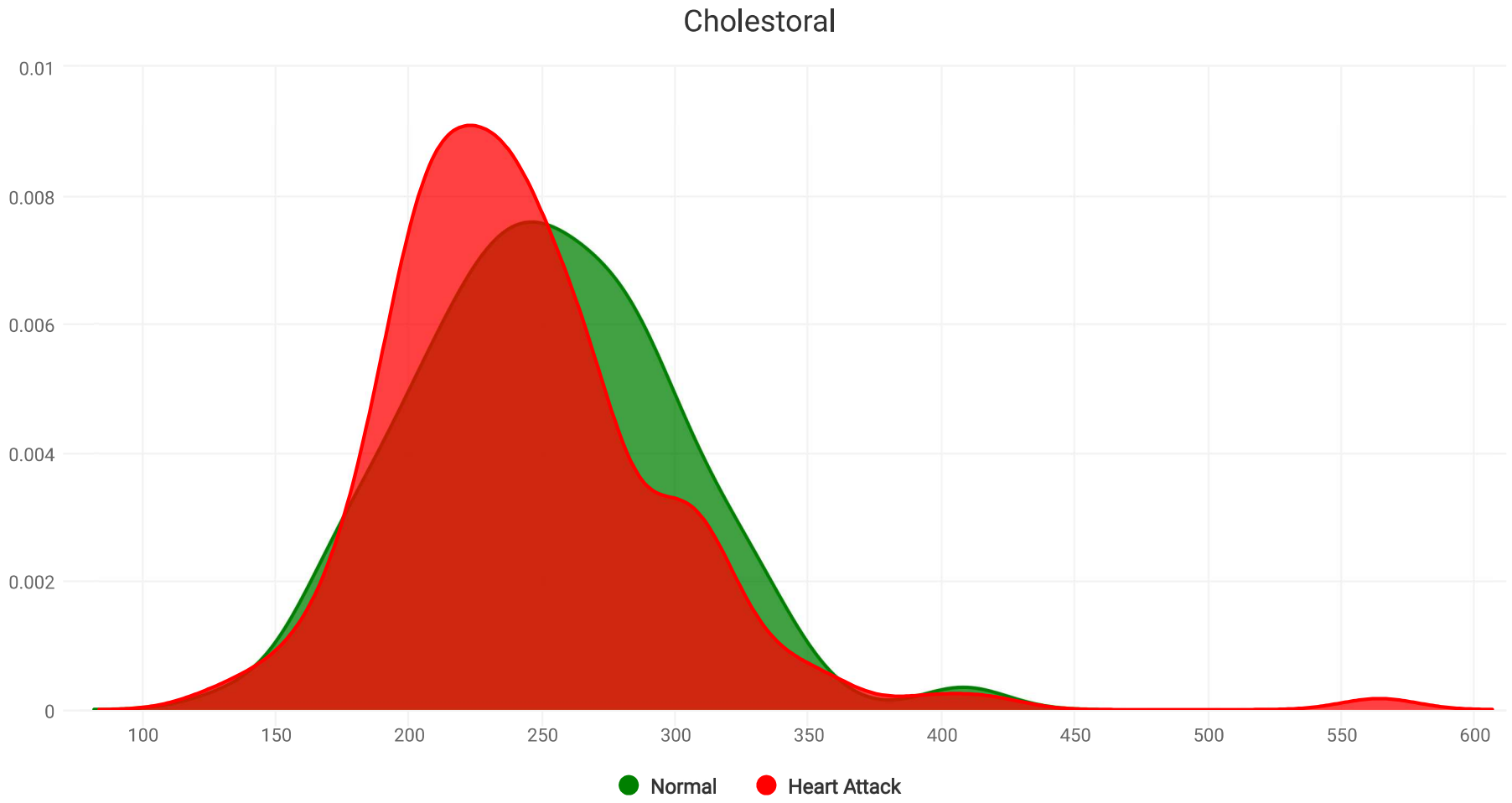
It is skewed to the right

```
print_dense_histogram = function(name, title){

    keeps <- c(name, 'output')
    frame = heart[ , (names(heart) %in% keeps)]

    normal <- frame %>% filter(output == 0)
    heart_attack <- frame %>% filter(output == 1)

   dense_plot <- hchart(
      density(normal[, (names(normal) %in% name)]), type = "area",
      color = "green",
       name = "Normal"
    ) %>%
      hc_add_series(
      density(heart_attack[, (names(heart_attack) %in% name)]), type = "area",
      color = "red",
      name = "Heart Attack"
   ) %>%
    hc_title(text= title) %>%
    hc_xAxis(title = name) %>%
    hc_add_theme(hc_theme_google())

  htmltools::tagList(dense_plot)

}
```

```
print_dense_histogram('age', 'Age')
```

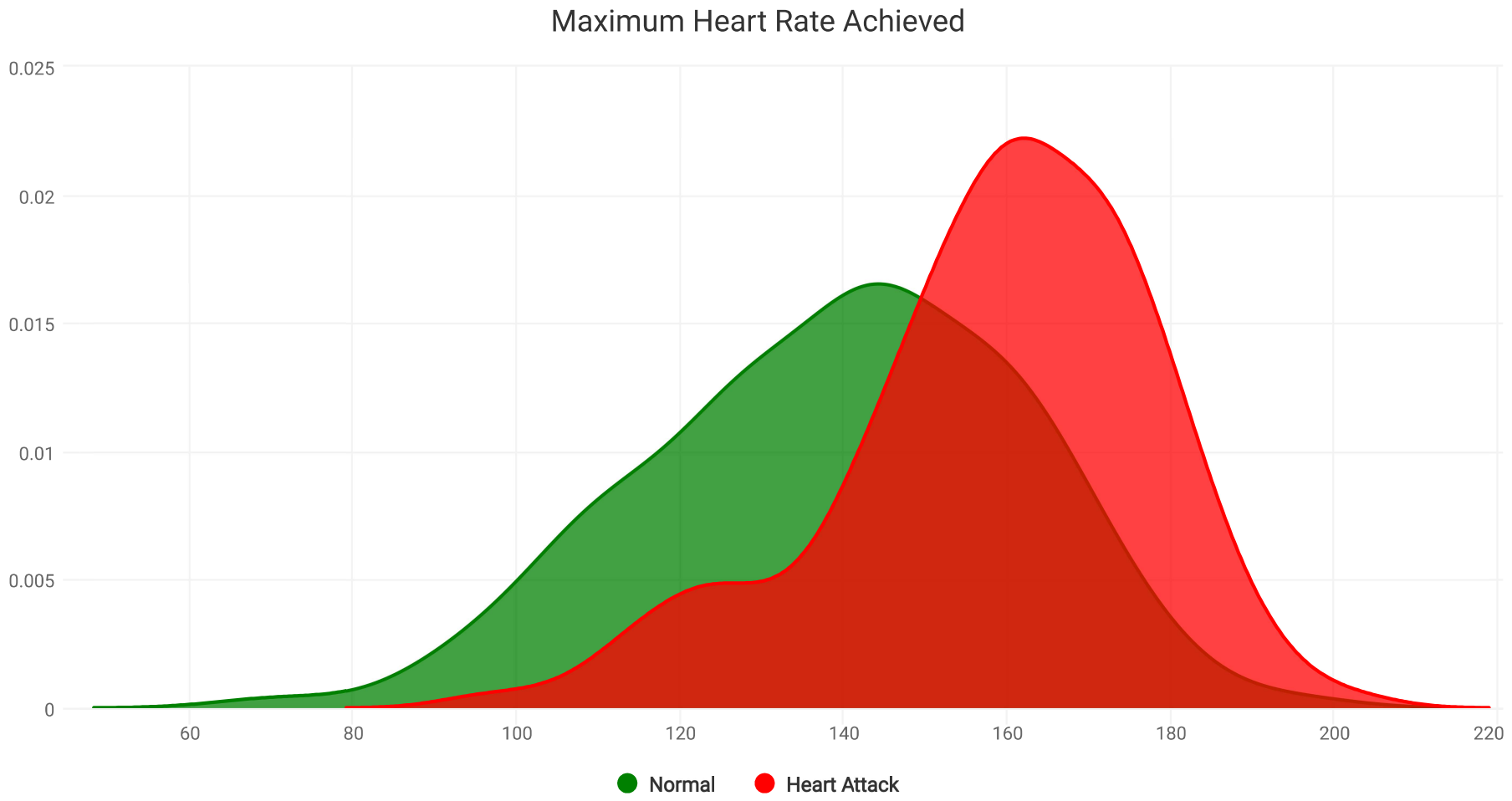
We do see that there is more heart attack for age (20-53 and 70-80+)

```
print_dense_histogram('chol', 'Cholestoral')
```



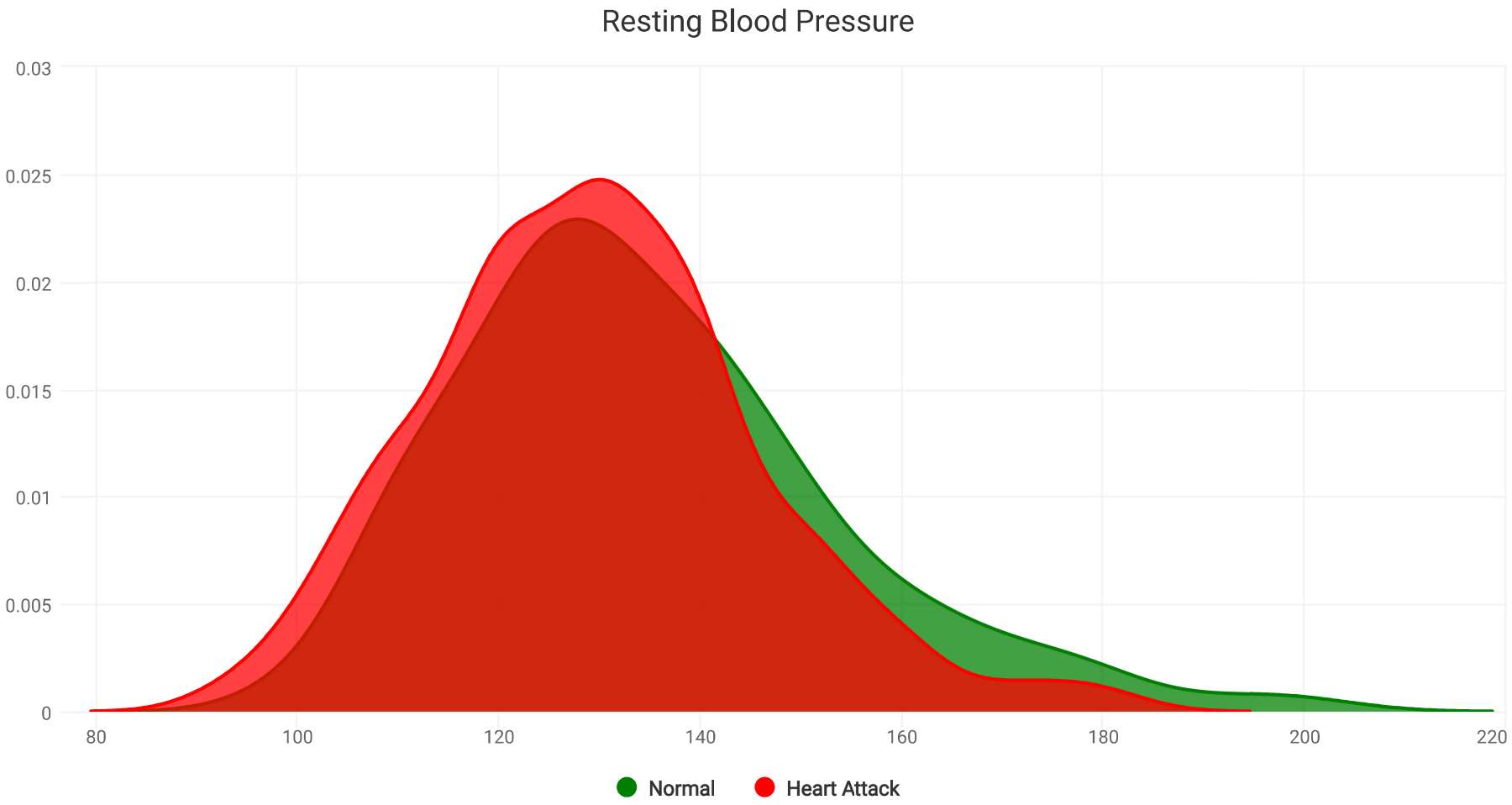
Low cholestrol (180-250)and high cholestroal (over 500) is more prone to heart attack

```
print_dense_histogram('thalachh', 'Maximum Heart Rate Achieved')
```



Maximum heart rate after 150 is more prone to heart attack

```
print_dense_histogram('trtbps', 'Resting Blood Pressure')
```



Resting blood pressure less than 140 is more prone to heart attack

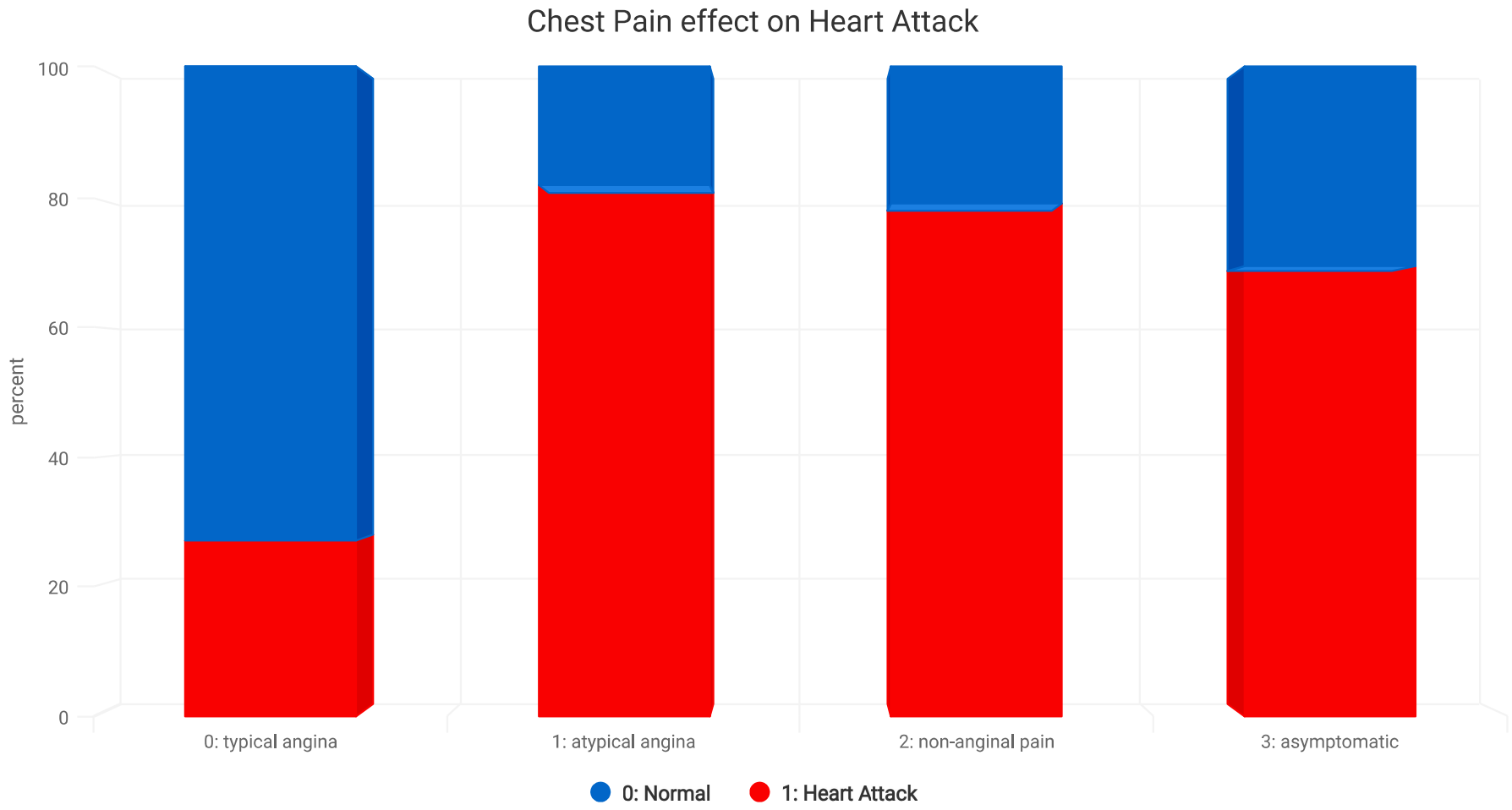
```
print_dense_histogram('oldpeak', 'Old Peak')
```

Less old peak is more prone to heart attack

```
cp_heart <- heart %>%
        select(output, cp) %>%
        group_by(output, cp) %>%
        summarize(count = n()) %>%
        mutate(
            cp_str = new_cp_str(cp),
            is_trouble = new_output_str(output),
            percentage = as.integer(round(count / sum(count) * 100, 1)),
            percentage_str = paste0(round(count / sum(count) * 100, 1), "%")
        )
```

```
## `summarise()` has grouped output by 'output'. You can override using the `.groups` argument.
```

```
cp_heart %>%
  hchart(type = "column", hcaes(x = paste(cp_str, ' \t(', percentage_str, ')' ), y=count, group = is_trouble)) %>%
    hc_plotOptions(column = list(stacking = "percent")) %>%
    hc_xAxis(title = "Chest Pain", categories = cp_heart$cp_str) %>%
    hc_yAxis(title = list(text = 'percent')) %>%
    hc_title(text = 'Chest Pain effect on Heart Attack') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE,
                            depth=100, viewDistance=25)) %>%
    hc_plotOptions(column=list(depth= 100))
```

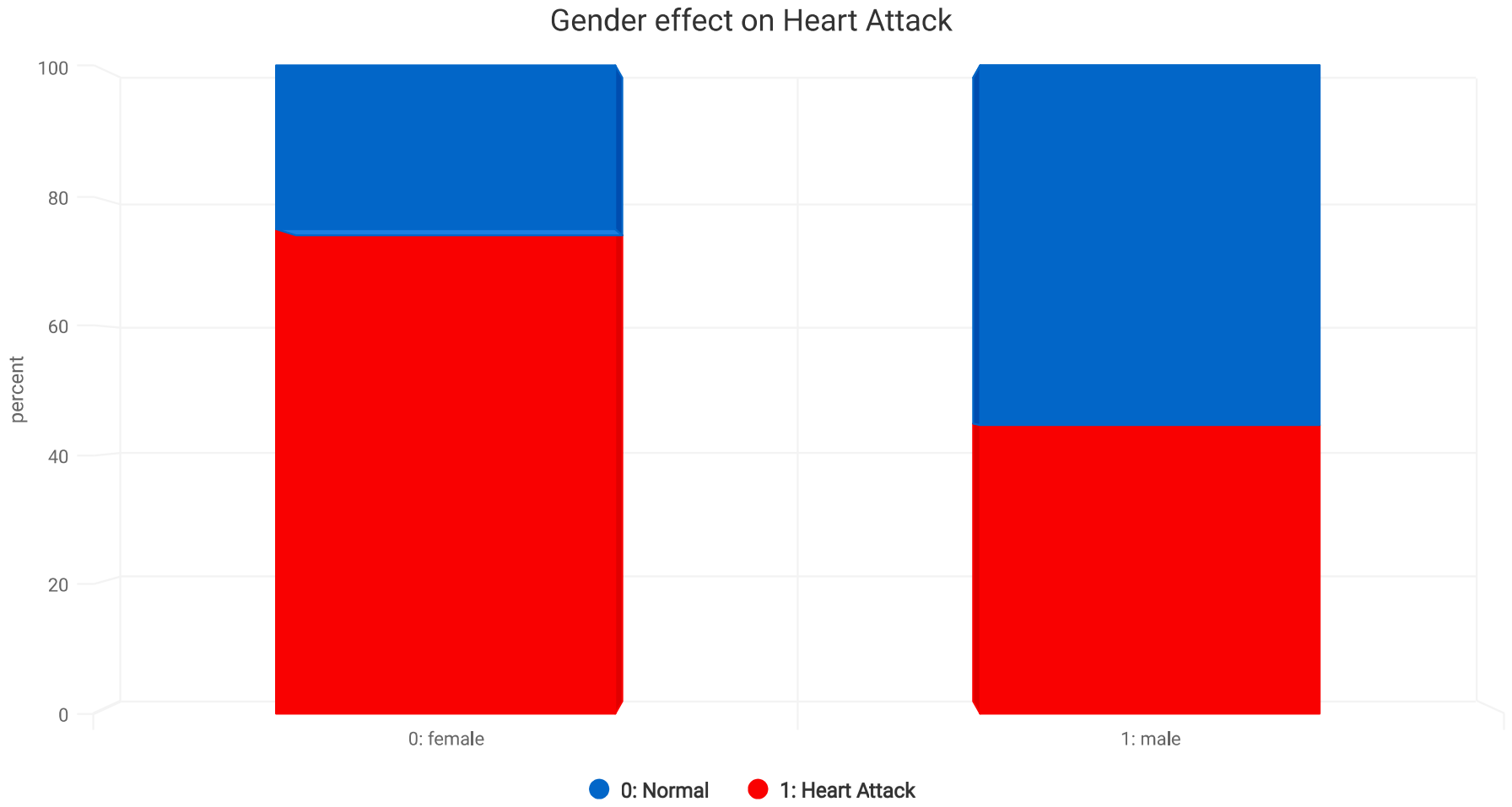


Chest pain type 1-3 is more prone to heart attack

```
gender_heart <- heart %>%
        select(output, sex) %>%
        group_by(output, sex) %>%
        summarize(count = n()) %>%
        mutate(
            gender = new_sex_str(sex),
            is_trouble = new_output_str(output),
            percentage = as.integer(round(count / sum(count) * 100, 1)),
            percentage_str = paste0(round(count / sum(count) * 100, 1), "%")
    )
```

```
## `summarise()` has grouped output by 'output'. You can override using the `.groups` argument.
```

```
gender_heart %>%
  hchart(type = "column", hcaes(x = paste(gender, ' \t(', percentage_str, ')' ), y=count, group = is_trouble)) %>%
    hc_plotOptions(column = list(stacking = "percent")) %>%
    hc_xAxis(title = "Gender", categories = gender_heart$gender) %>%
    hc_yAxis(title = list(text = 'percent')) %>%
    hc_title(text = 'Gender effect on Heart Attack') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE,
                            depth=100, viewDistance=25)) %>%
    hc_plotOptions(column=list(depth= 100))
```

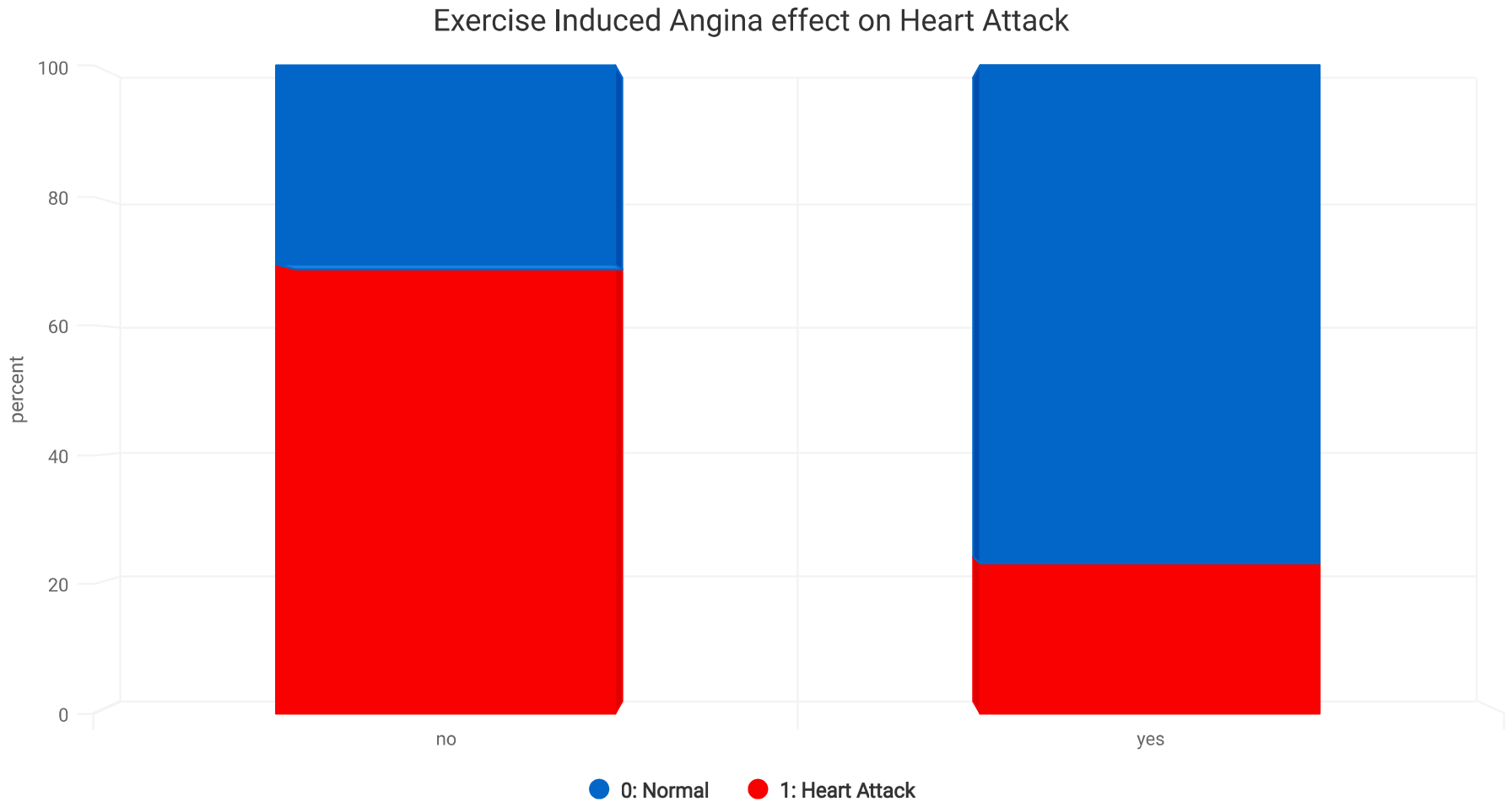


Female is more prone to heart attack

```
exng_heart <- heart %>%
        select(output, exng) %>%
        group_by(output, exng) %>%
        summarize(count = n()) %>%
        mutate(
            exng_str = new_exng_str(exng),
            is_trouble = new_output_str(output),
            percentage = as.integer(round(count / sum(count) * 100, 1)),
            percentage_str = paste0(round(count / sum(count) * 100, 1), "%")
        )
```

```
## `summarise()` has grouped output by 'output'. You can override using the `.groups` argument.
```

```
exng_heart %>%
  hchart(type = "column", hcaes(x = paste(exng_str, ' \t(', percentage_str, ')' ), y=count, group = is_trouble)) %>%
    hc_plotOptions(column = list(stacking = "percent")) %>%
    hc_xAxis(title = "Exercise Induced Angina", categories = exng_heart$exng_str) %>%
    hc_yAxis(title = list(text = 'percent')) %>%
    hc_title(text = 'Exercise Induced Angina effect on Heart Attack') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE,
                            depth=100, viewDistance=25)) %>%
    hc_plotOptions(column=list(depth= 100))
```

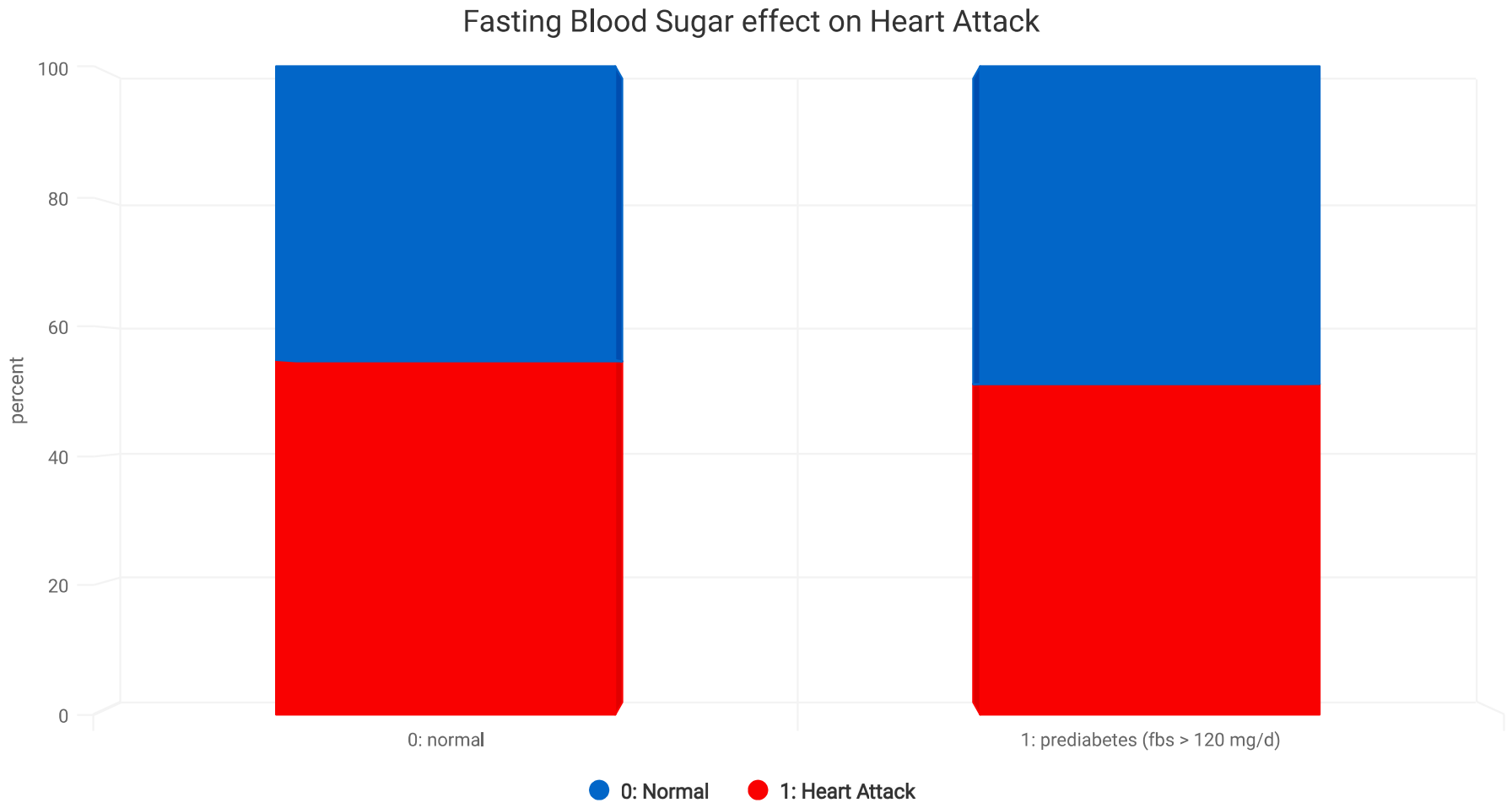


No exercise induced angina is more prone to heart attack

```
fbs_heart <- heart %>%
        select(output, fbs) %>%
        group_by(output, fbs) %>%
        summarize(count = n()) %>%
        mutate(
            fbs_str = new_fbs_str(fbs),
            is_trouble = new_output_str(output),
            percentage = as.integer(round(count / sum(count) * 100, 1)),
            percentage_str = paste(round(count / sum(count) * 100, 1), "%")
        )
```

```
## `summarise()` has grouped output by 'output'. You can override using the `.groups` argument.
```

```
fbs_heart %>%
  hchart(type = "column", hcaes(x = paste(fbs_str, ' \t(', percentage_str, ')' ), y=count, group = is_trouble)) %>%
    hc_plotOptions(column = list(stacking = "percent")) %>%
    hc_xAxis(title = "Fasting Blood Sugar", categories = fbs_heart$fbs_str) %>%
    hc_yAxis(title = list(text = 'percent')) %>%
    hc_title(text = 'Fasting Blood Sugar effect on Heart Attack') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE,
                            depth=100, viewDistance=25)) %>%
    hc_plotOptions(column=list(depth= 100))
```

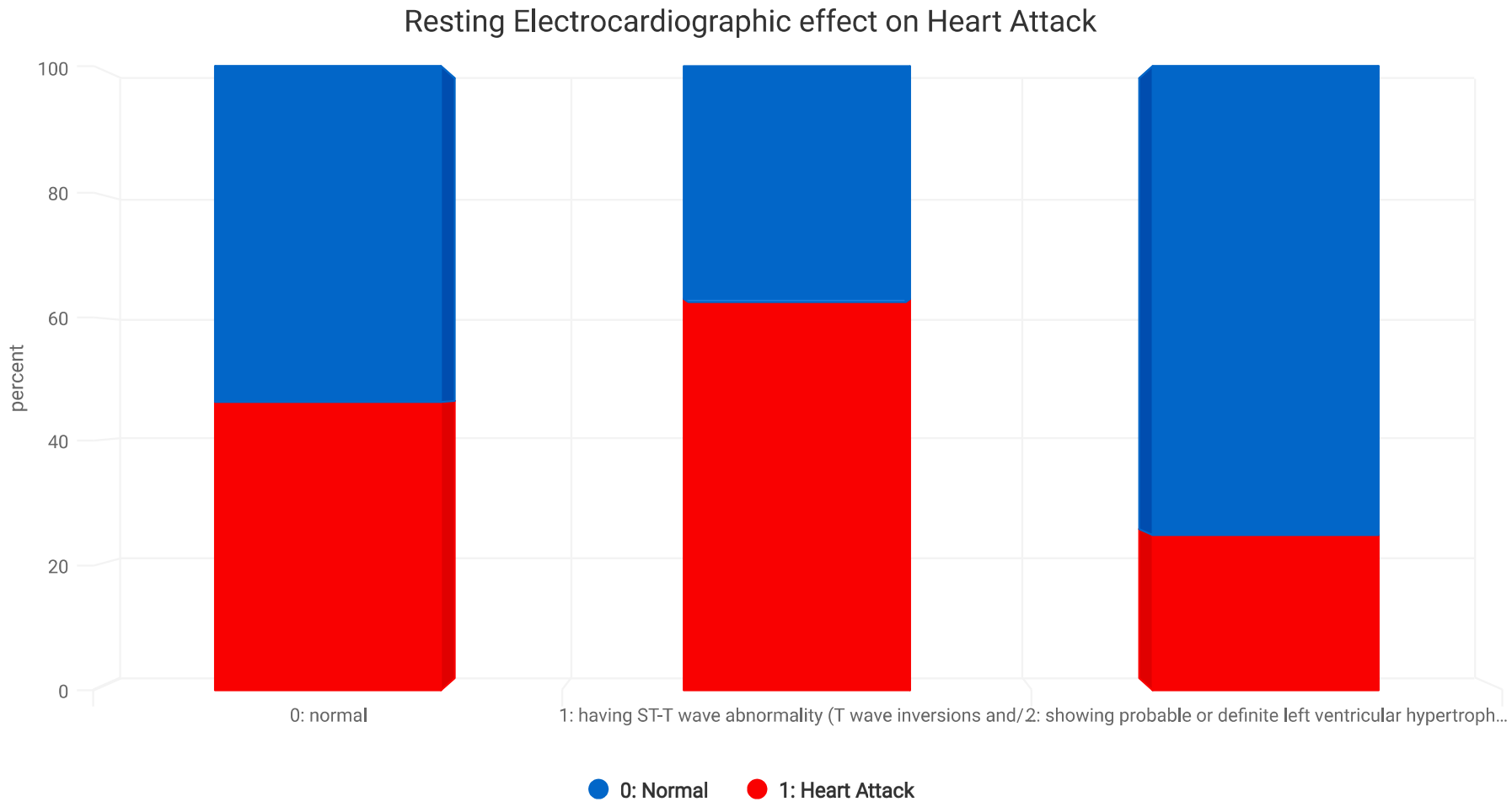


fbs is not really impacting heart attack much

```
restecg_heart <- heart %>%
        select(output, restecg) %>%
        group_by(output, restecg) %>%
        summarize(count = n()) %>%
        mutate(
            res_str = new_restecg_str(restecg),
            is_trouble = new_output_str(output),
            percentage = as.integer(round(count / sum(count) * 100, 1)),
            percentage_str = paste0(round(count / sum(count) * 100, 1), "%")
        )
```

```
## `summarise()` has grouped output by 'output'. You can override using the `.groups` argument.
```

```
restecg_heart %>%
  hchart(type = "column", hcaes(x = paste(res_str, ' \t(', percentage_str, ')' ), y=count, group = is_trouble)) %>%
    hc_plotOptions(column = list(stacking = "percent")) %>%
    hc_xAxis(title = "Resting Electrocardiographic Results", categories = restecg_heart$res_str) %>%
    hc_yAxis(title = list(text = 'percent')) %>%
    hc_title(text = 'Resting Electrocardiographic effect on Heart Attack') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE,
                            depth=100, viewDistance=25)) %>%
    hc_plotOptions(column=list(depth= 100))
```

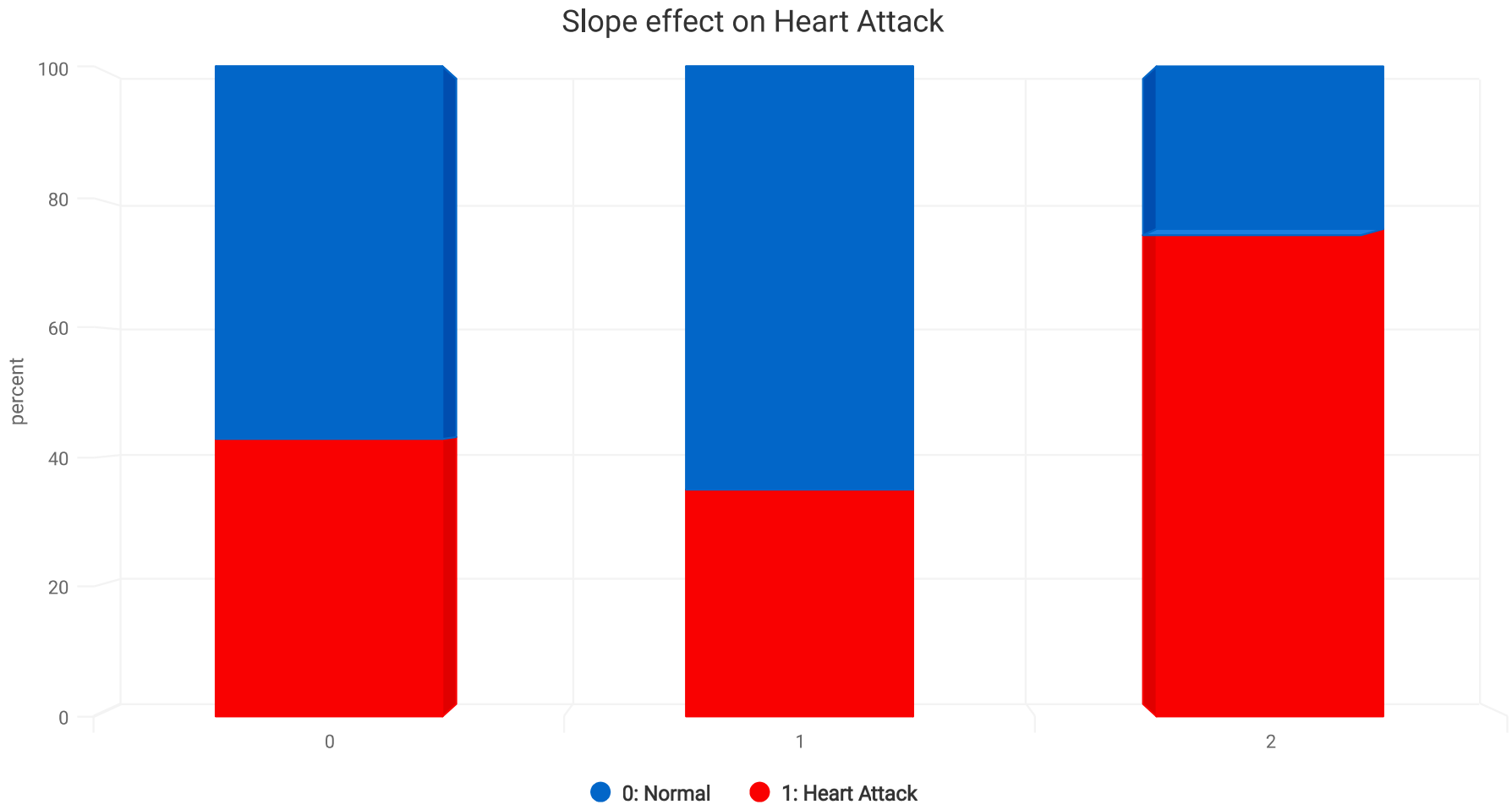


Resting electrocardiographic 2 is less risky to heart attack

```
slp_heart <- heart %>%
        select(output, slp) %>%
        group_by(output, slp) %>%
        summarize(count = n()) %>%
        mutate(
            slp_str =  slp,
            is_trouble = new_output_str(output),
            percentage = as.integer(round(count / sum(count) * 100, 1)),
            percentage_str = paste0(round(count / sum(count) * 100, 1), "%")
        )
```

```
## `summarise()` has grouped output by 'output'. You can override using the `.groups` argument.
```

```
slp_heart %>%
  hchart(type = "column", hcaes(x = paste(slp_str, ' \t(', percentage_str, ')' ), y=count, group = is_trouble)) %>%
    hc_plotOptions(column = list(stacking = "percent")) %>%
    hc_xAxis(title = "Slope", categories = slp_heart$slp_str) %>%
    hc_yAxis(title = list(text = 'percent')) %>%
    hc_title(text = 'Slope effect on Heart Attack') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE,
                            depth=100, viewDistance=25)) %>%
    hc_plotOptions(column=list(depth= 100))
```

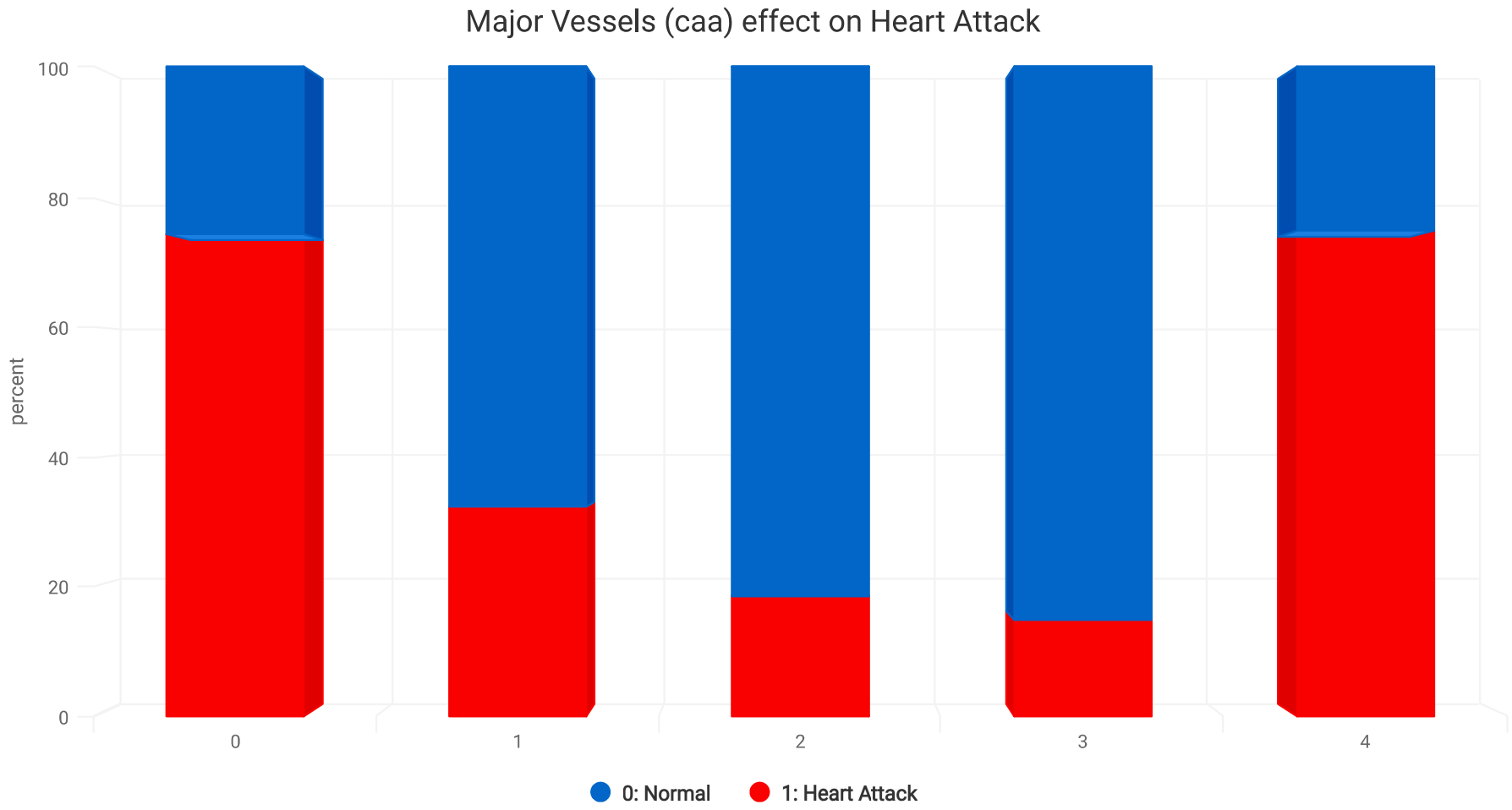


slope 2 is more prone to heart attack

```
caa_heart <- heart %>%
        select(output, caa) %>%
        group_by(output, caa) %>%
        summarize(count = n()) %>%
        mutate(
            caa_str = caa,
            is_trouble = new_output_str(output),
            percentage = as.integer(round(count / sum(count) * 100, 1)),
            percentage_str = paste0(round(count / sum(count) * 100, 1), "%")
        )
```

```
## `summarise()` has grouped output by 'output'. You can override using the `.groups` argument.
```

```
caa_heart %>%
    hchart(type = "column", hcaes(x = paste(caa_str, ' \t(', percentage_str, ')' ), y=count, group = is_trouble)) %>%
    hc_plotOptions(column = list(stacking = "percent")) %>%
    hc_xAxis(title = "Number of Major Vessels (0-4)", categories = caa_heart$caa_str) %>%
    hc_yAxis(title = list(text = 'percent')) %>%
    hc_title(text = 'Major Vessels (caa) effect on Heart Attack') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE,
                            depth=100, viewDistance=25)) %>%
    hc_plotOptions(column=list(depth= 100))
```

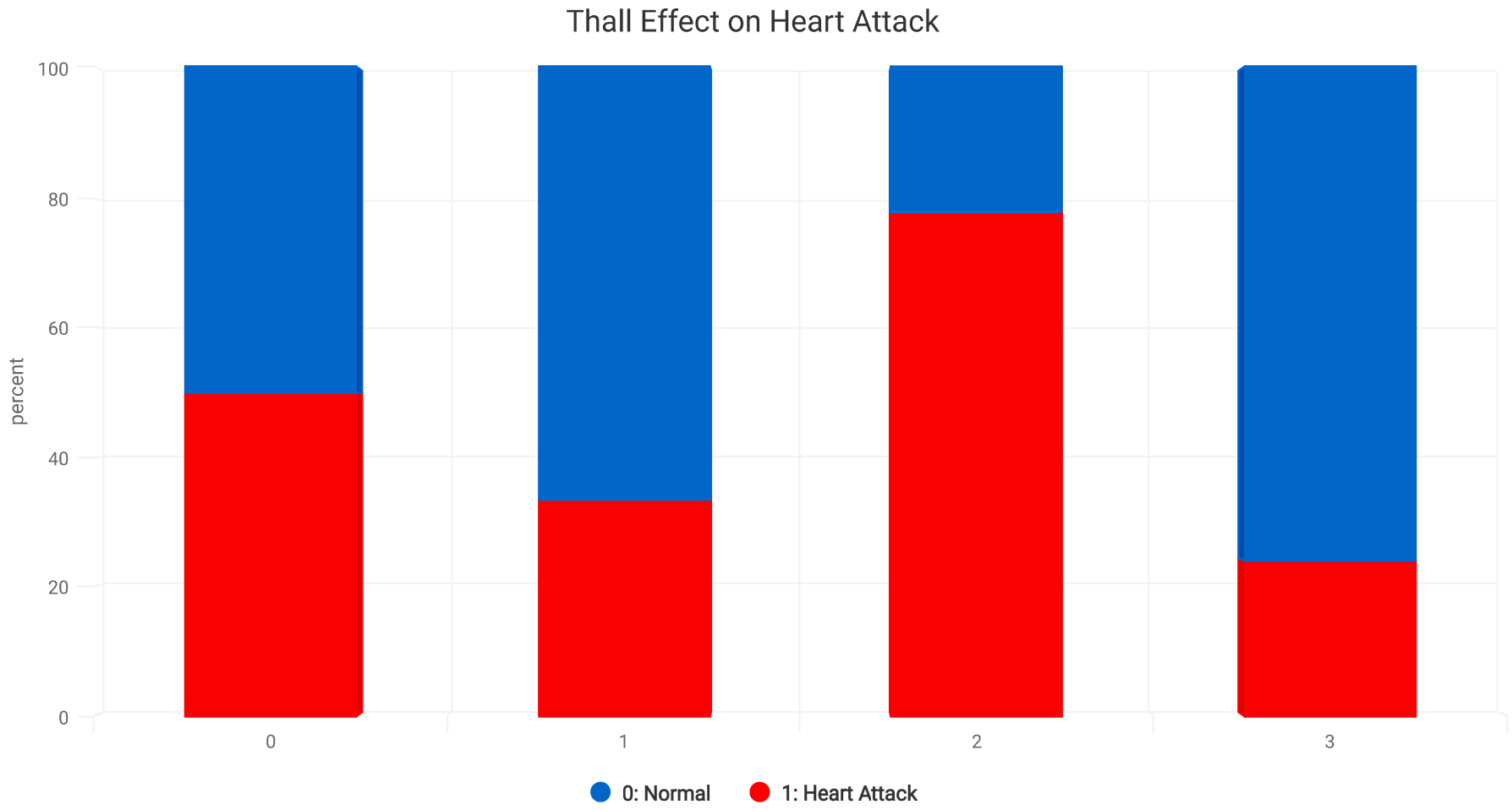


We do see that major vessels 0 and 4 is more prone to heart attack and 1-3 is less risky to heart attack

```
thall_heart <- heart %>%
        select(output, thall) %>%
        group_by(output, thall) %>%
        summarize(count = n()) %>%
        mutate(
            thall_str = thall,
            is_trouble = new_output_str(output),
            percentage = as.integer(round(count / sum(count) * 100, 1)),
            percentage_str = paste0(round(count / sum(count) * 100, 1), "%")
        )
```

```
## `summarise()` has grouped output by 'output'. You can override using the `.groups` argument.
```

```
thall_heart %>%
    hchart(type = "column", hcaes(x = paste(thall_str, ' \t(', percentage_str, ')' ), y=count, group = is_trouble)) %>%
    hc_plotOptions(column = list(stacking = "percent")) %>%
    hc_xAxis(title = "Thall(0-3)", categories = thall_heart$thall_str) %>%
    hc_yAxis(title = list(text = 'percent')) %>%
    hc_title(text = 'Thall Effect on Heart Attack') %>%
    hc_add_theme(hc_theme_google()) %>%
    hc_chart(options3d=list(enabled=TRUE,
                            depth=100, viewDistance=70)) %>%
    hc_plotOptions(column=list(depth= 100))
```



thall type 2 is more prone to heart attack and 1 and 3 is less risky to heart attack

```
cor(heart)[,14]
```

```
##         age         sex          cp      trtbps        chol         fbs
## -0.22147583 -0.28360936  0.43207959 -0.14626866 -0.08143720 -0.02682597
##     restecg    thalachh        exng     oldpeak         slp         caa
##  0.13487445  0.41995504 -0.43560076 -0.42914583  0.34393953 -0.40899198
##       thall      output
## -0.34310071  1.00000000
```

We have positive correaltion with cp,restecg,thalachh,slp. The chance of heart attack will increase when cp increases and it is same for restecg, thalachh, and slp. For negative correlation, the chance of heart attack will increase when age is decreasing and that is same for the rest of negative correlations.

#Using logliklihood of poisson distribution

#```{r} #model = glm(output~1, family=poisson(link=log), data=heart) #summary(model)

#model1 = glm(output~trtbps+chol+fbs+restecg, family=poisson(link=log), #data=heart) #summary(model1)

#model2 = glm(output~age+sex+cp+trtbps+chol+fbs+restecg+thalachh+exng+oldpeak+sl #p+caa+thall, family=poisson(link=log), data=heart) #summary(model2)

#anova(model, model1,model2, test="Chisq") #```