HEART

ilke

2022-10-14

```
getwd()
```

```
## [1] "/home/ilke/Documents/github/r"
```

```
setwd("/home/ilke/Downloads")

df<- read.csv("heart.csv",sep=",", header=TRUE,stringsAsFactors = FALSE)
```

```
head(df)
```

```
##   age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp caa thall output
## 1  63   1  3    145  233   1       0     150    0     2.3   0   0     1      1
## 2  37   1  2    130  250   0       1     187    0     3.5   0   0     2      1
## 3  41   0  1    130  204   0       0     172    0     1.4   2   0     2      1
## 4  56   1  1    120  236   0       1     178    0     0.8   2   0     2      1
## 5  57   0  0    120  354   0       1     163    1     0.6   2   0     2      1
## 6  57   1  0    140  192   0       1     148    0     0.4   1   0     1      1
```

exang =egzerzise bağlı engina: (1 = evet; 0 = hayır)

cp : Göğüs Ağrısı tipi Göğüs ağrısı tipi Değer 1: tipik anjina Değer 2: atipik anjina Değer 3: anjinal olmayan ağrı Değer 4: asemptomatik

trtbps : dinlenme kan basıncı (mm Hg olarak)

chol : BMI sensörü aracılığıyla alınan mg/dl cinsinden kolestoral

fbs : (açlık kan şekeri > 120 mg/dl) (1 = doğru; 0 = yanlış

rest_ecg : dinlenme elektrokardiyografik sonuçları 0 değeri: normal Değer 1: ST-T dalga anormalliğine sahip olmak (T dalgası inversiyonları ve/veya ST yükselmesi veya > 0.05 mV depresyonu) Değer 2: Estes kriterlerine göre olası veya kesin sol ventrikül hipertrofisini gösteriyor

thalach: ulaşılan maksimum kalp atış hızı

target : 0= daha az kalp krizi geçirme olasılığı 1= daha fazla kalp krizi geçirme olasılığı

```
colnames(df)
```

```
##  [1] "age"      "sex"      "cp"       "trtbps"   "chol"     "fbs"
##  [7] "restecg"  "thalachh" "exng"     "oldpeak"  "slp"      "caa"
## [13] "thall"    "output"
```

```
summary(df)
```

```
##       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
```

```
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
```

```
glimpse(df)
```

```
## Rows: 303
## Columns: 14
## $ age      <int> 63, 37, 41, 56, 57, 57, 56, 44, 52, 57, 54, 48, 49, 64, 58, 5…
## $ sex      <int> 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1…
## $ cp       <int> 3, 2, 1, 1, 0, 0, 1, 1, 2, 2, 0, 2, 1, 3, 3, 2, 2, 3, 0, 3, 0…
## $ trtbps   <int> 145, 130, 130, 120, 120, 140, 140, 120, 172, 150, 140, 130, 1…
## $ chol     <int> 233, 250, 204, 236, 354, 192, 294, 263, 199, 168, 239, 275, 2…
## $ fbs      <int> 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 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, 1, 1…
## $ thalachh <int> 150, 187, 172, 178, 163, 148, 153, 173, 162, 174, 160, 139, 1…
## $ exng     <int> 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 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, 0.2, 0…
## $ slp      <int> 0, 0, 2, 2, 2, 1, 1, 2, 2, 2, 2, 2, 1, 2, 1, 2, 0, 2, 2, 1…
## $ caa      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0…
## $ thall    <int> 1, 2, 2, 2, 2, 1, 2, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3…
## $ output   <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1…
```

```
#kategorik verileri faktor türüne dönüştürme


df$sex <- as.factor(df$sex)
df$output <- as.factor(df$output)
df$caa <- as.factor(df$caa)
df$exng <- as.factor(df$exng)
df$thall <- as.factor(df$thall)
df$fbs <- as.factor(df$fbs)
df$restecg <- as.factor(df$restecg)
df$slp <- as.factor(df$slp)
df$cp <- as.factor(df$output)
```

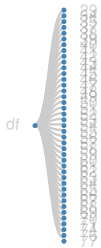
```
glimpse(df)
```

```
## Rows: 303
## Columns: 14
## $ age      <int> 63, 37, 41, 56, 57, 57, 56, 44, 52, 57, 54, 48, 49, 64, 58, 5…
## $ sex      <fct> 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1…
## $ cp       <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1…
## $ trtbps   <int> 145, 130, 130, 120, 120, 140, 140, 120, 172, 150, 140, 130, 1…
## $ chol     <int> 233, 250, 204, 236, 354, 192, 294, 263, 199, 168, 239, 275, 2…
## $ fbs      <fct> 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0…
## $ restecg  <fct> 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1…
## $ thalachh <int> 150, 187, 172, 178, 163, 148, 153, 173, 162, 174, 160, 139, 1…
## $ exng     <fct> 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 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, 0.2, 0…
## $ slp      <fct> 0, 0, 2, 2, 2, 1, 1, 2, 2, 2, 2, 2, 2, 1, 2, 1, 2, 0, 2, 2, 1…
## $ caa      <fct> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0…
## $ thall    <fct> 1, 2, 2, 2, 2, 1, 2, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3…
## $ output   <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1…
```

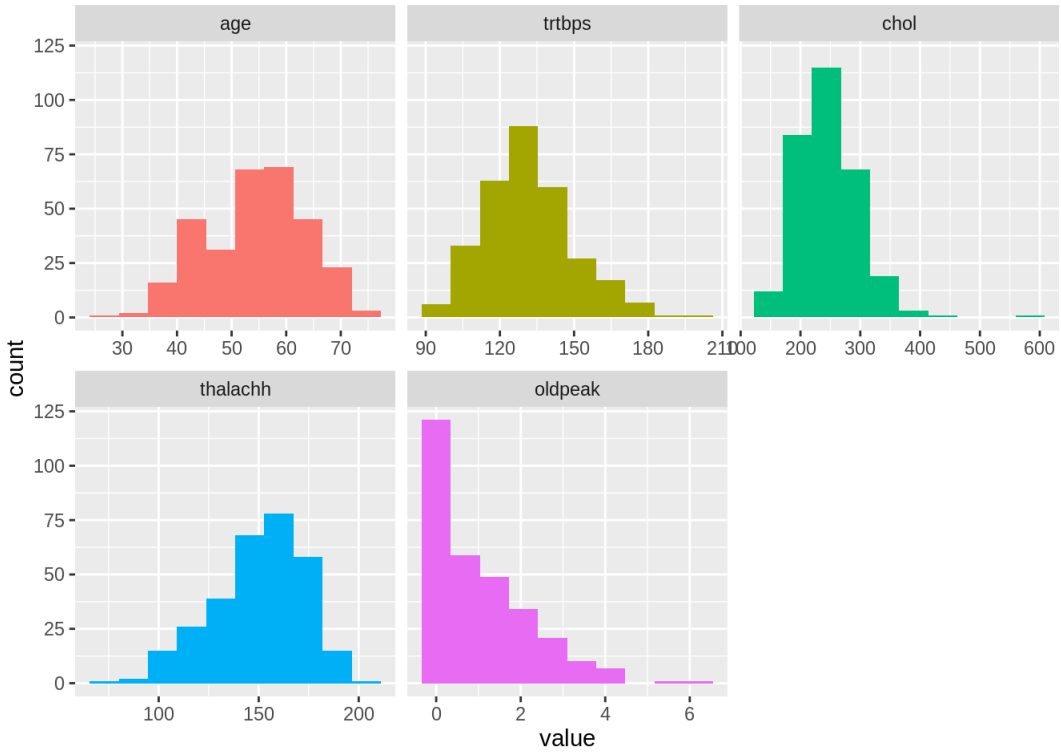
```
df[df==""] <- NA
sum(is.na(df))        #eksik gözlem yok. temiz veri seti
```

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

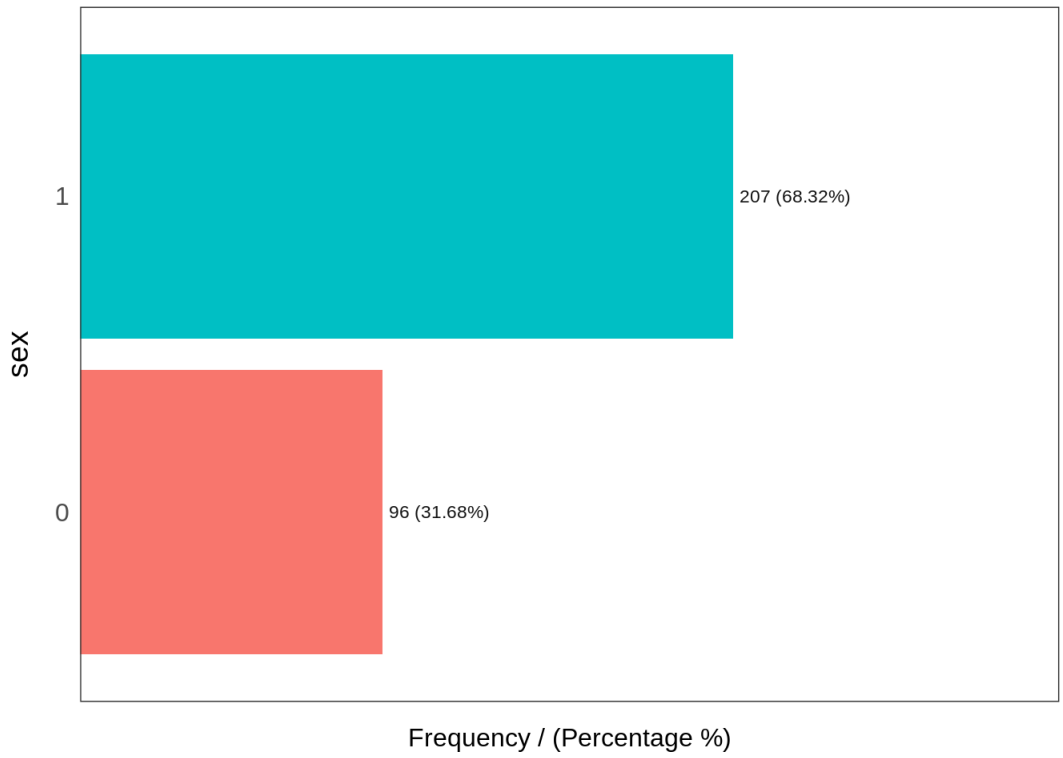
```
library(d3Tree)
d3tree(list(root = df2tree(rootname ='df',
                           struct = as.data.frame(df)),
                           layout = 'collapse'))
```

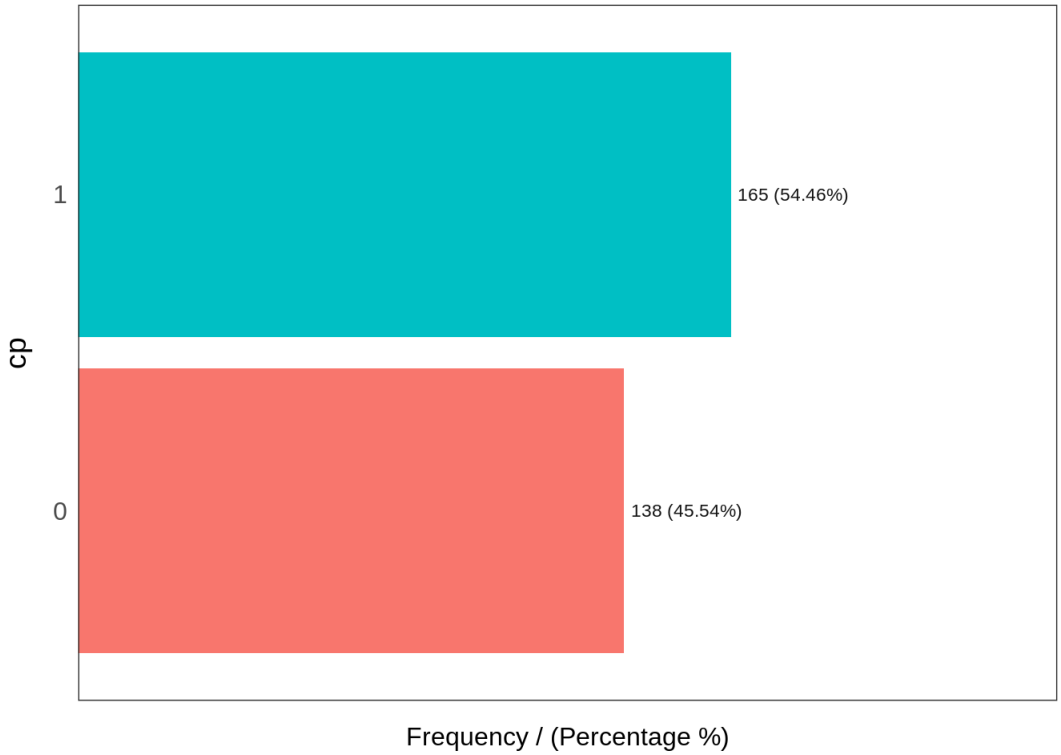


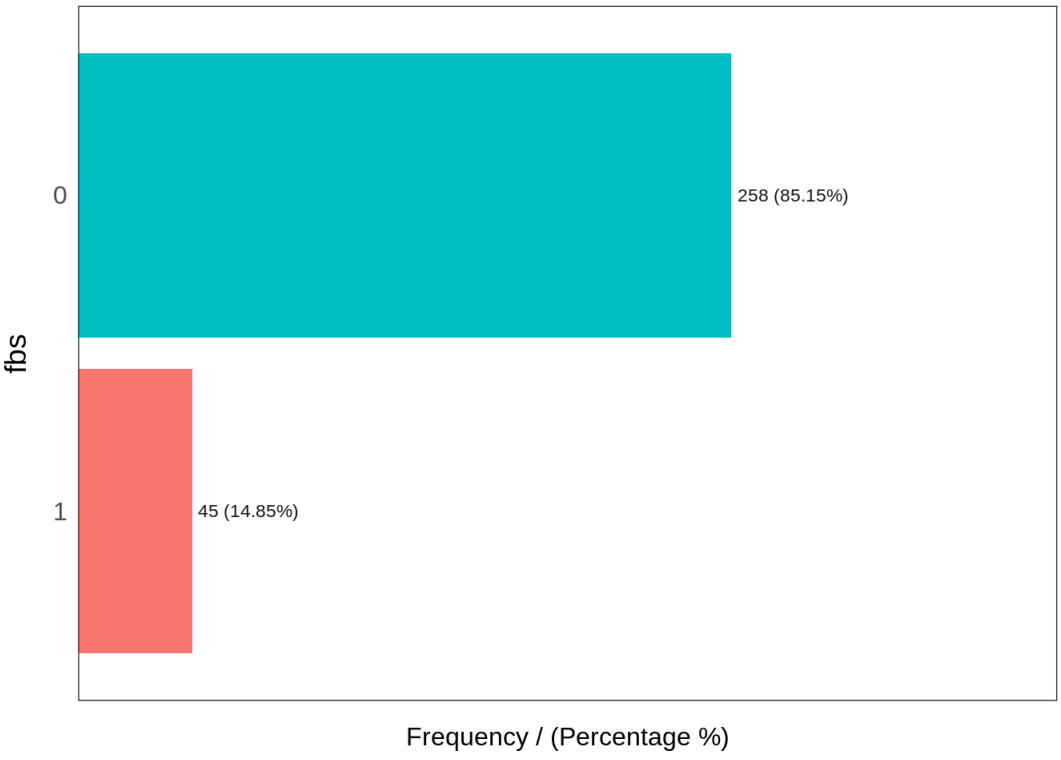

```
library(funModeling)
profiling_num(df)   #sürekli değişkenlerin özet bazı istatistiklerini verir
plot_num(df)  #genel histogram
```

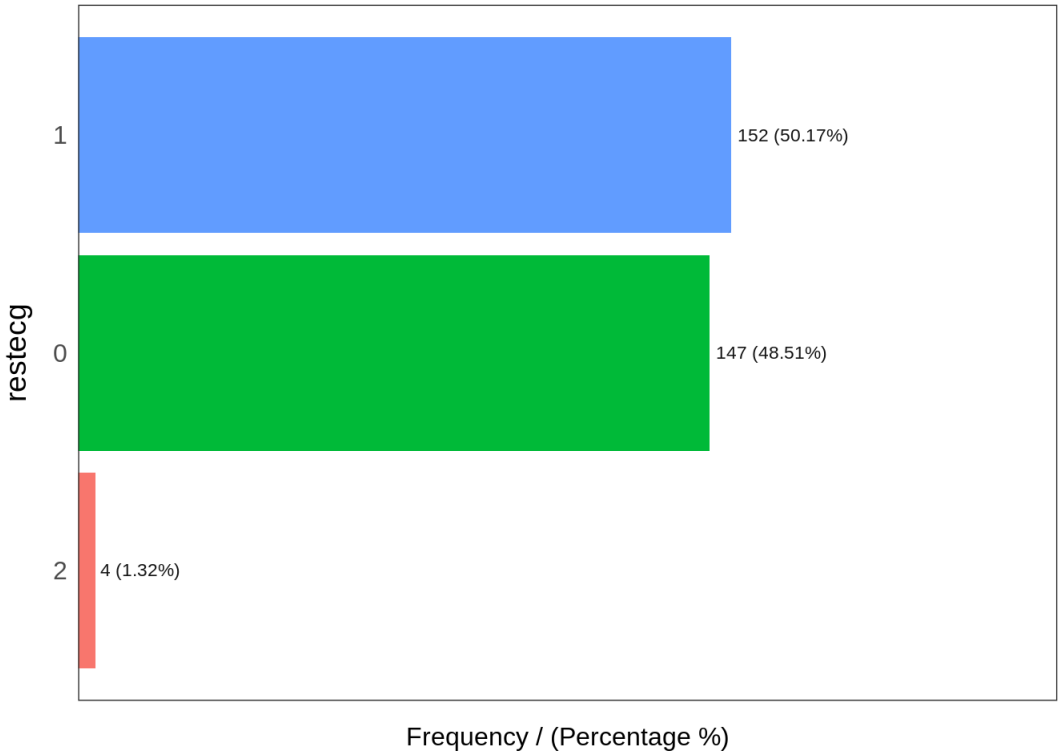


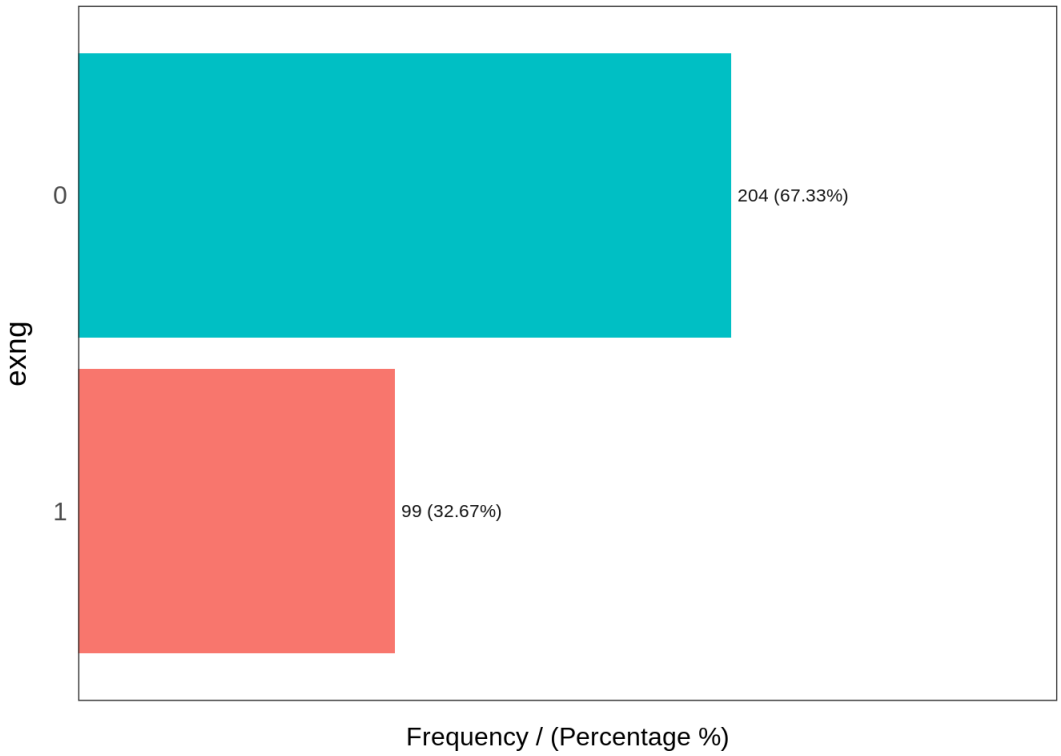

```
freq(df)      #detaylı data frame,grafikler verir
```

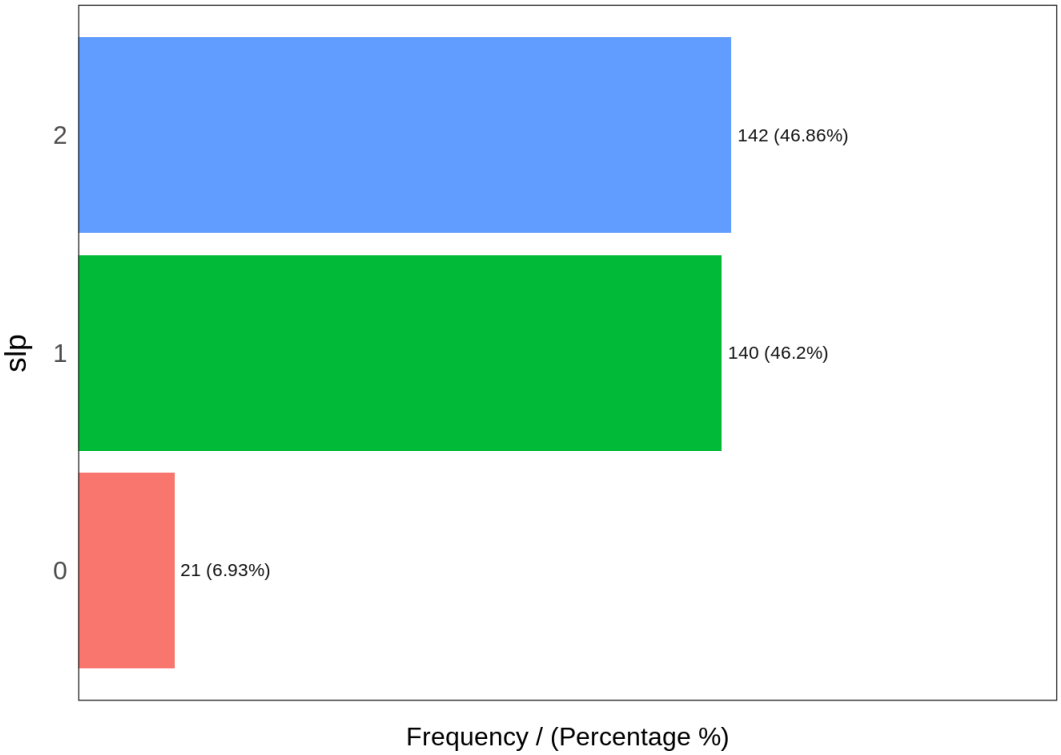


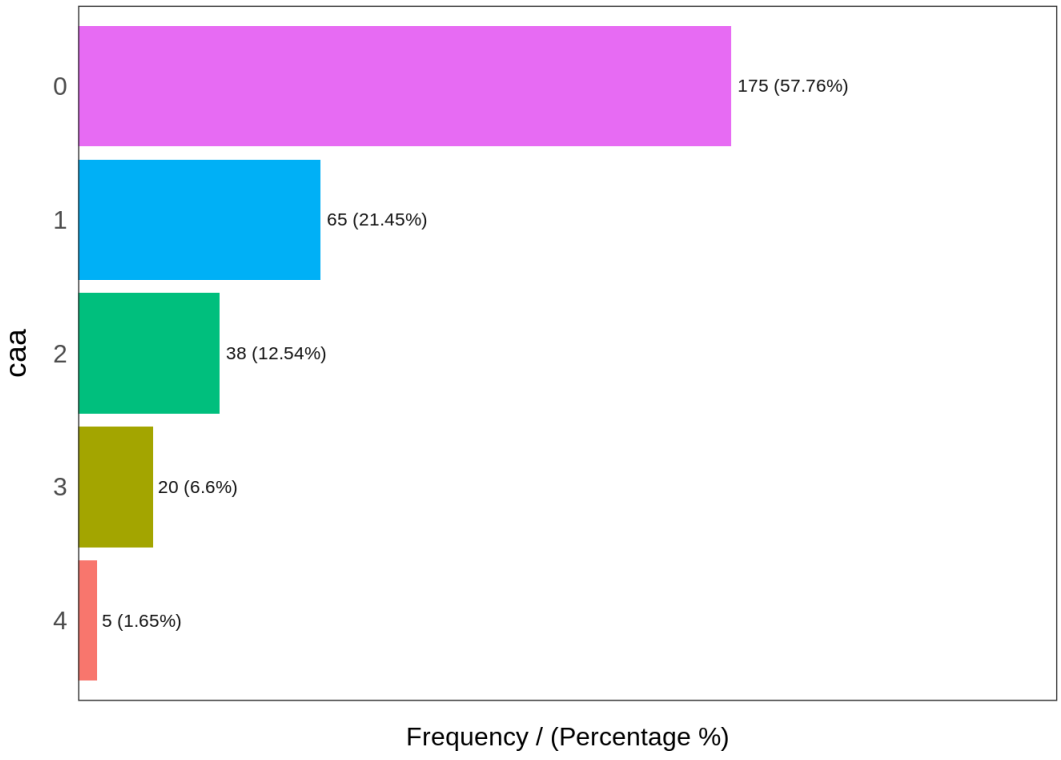

cp
1
0
165 (54.46%)
138 (45.54%)
Frequency / (Percentage %)

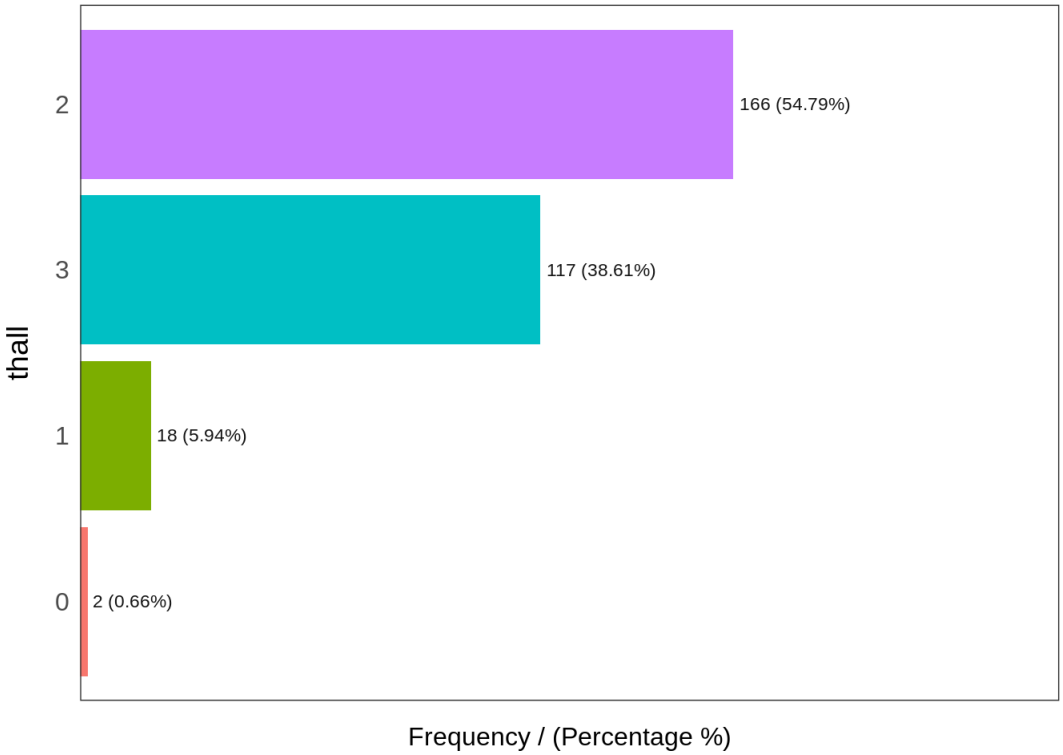

fbs
0
1
258 (85.15%)
45 (14.85%)
Frequency / (Percentage %)

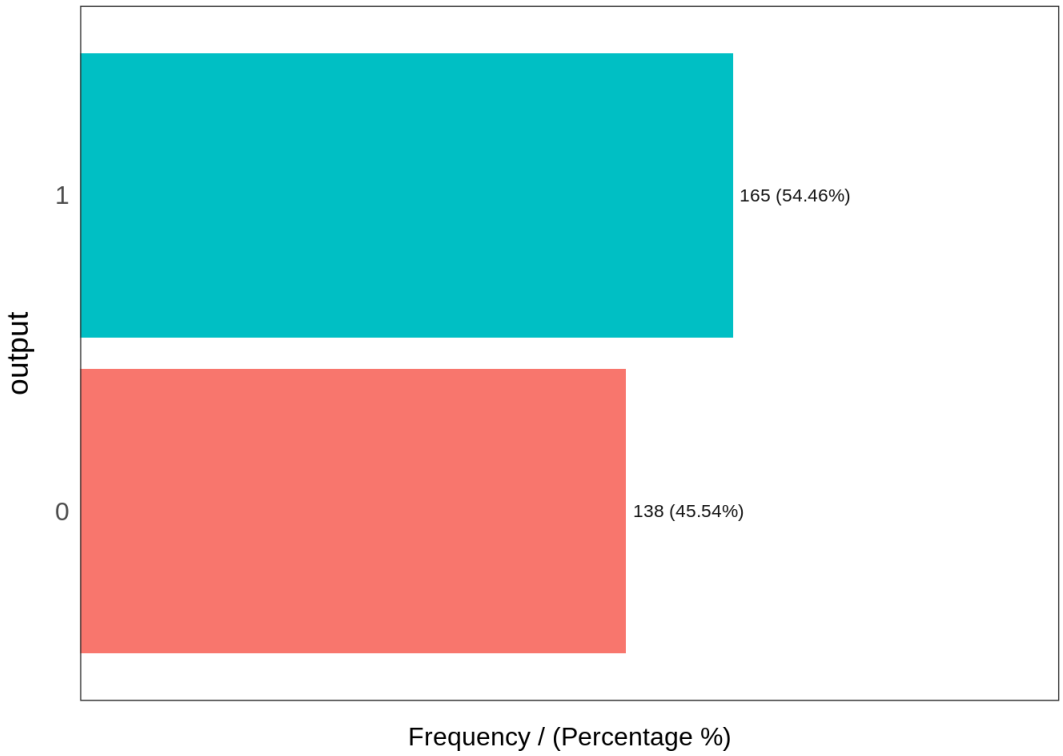

restecg
1
0
2
152 (50.17%)
147 (48.51%)
4 (1.32%)
Frequency / (Percentage %)


exng
0
1
204 (67.33%)
99 (32.67%)
Frequency / (Percentage %)


slp
2
1
0
142 (46.86%)
140 (46.2%)
21 (6.93%)
Frequency / (Percentage %)


caa
0
1
2
3
4
175 (57.76%)
65 (21.45%)
38 (12.54%)
20 (6.6%)
5 (1.65%)
Frequency / (Percentage %)






```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✔ tibble  3.1.8     ✔ purrr   0.3.4
## ✔ tidyr   1.2.0     ✔ stringr 1.4.0
## ✔ readr   2.1.2     ✔ forcats 0.5.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter()    masks stats::filter()
## ✖ dplyr::lag()       masks stats::lag()
## ✖ Hmisc::src()       masks dplyr::src()
## ✖ Hmisc::summarize() masks dplyr::summarize()
```

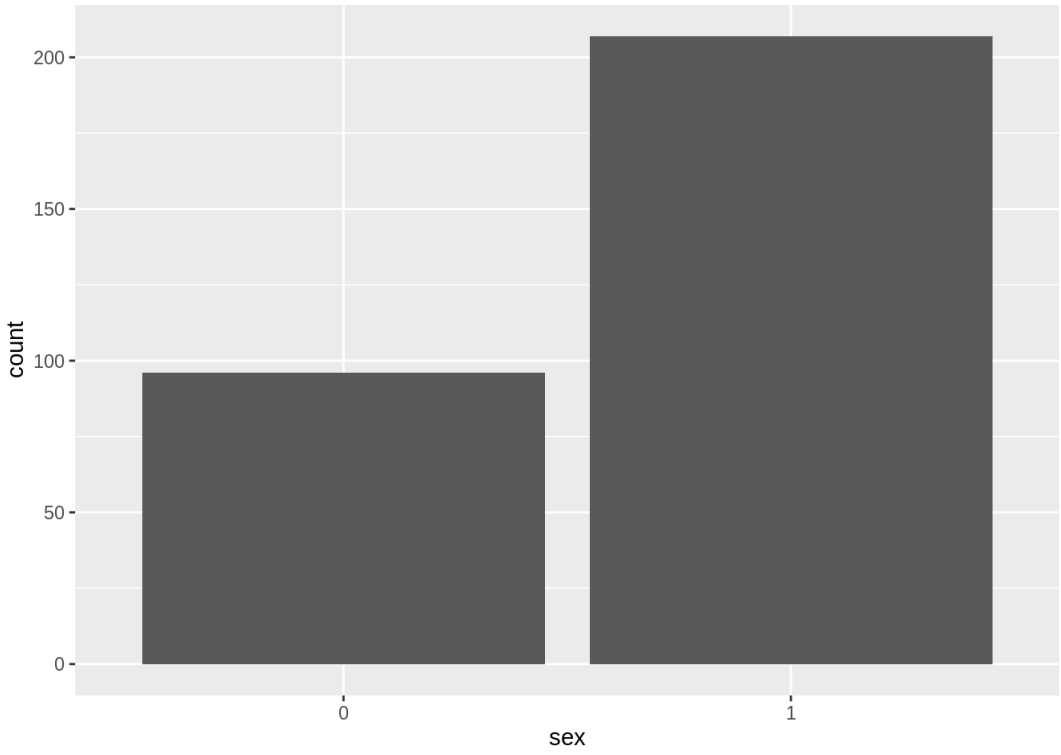
```
library(ggplot2)
library(plotly)
```

```
##
## Attaching package: 'plotly'
##
## The following object is masked from 'package:Hmisc':
##
##     subplot
##
## The following object is masked from 'package:ggplot2':
##
##     last_plot
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following object is masked from 'package:graphics':
##
##     layout
```

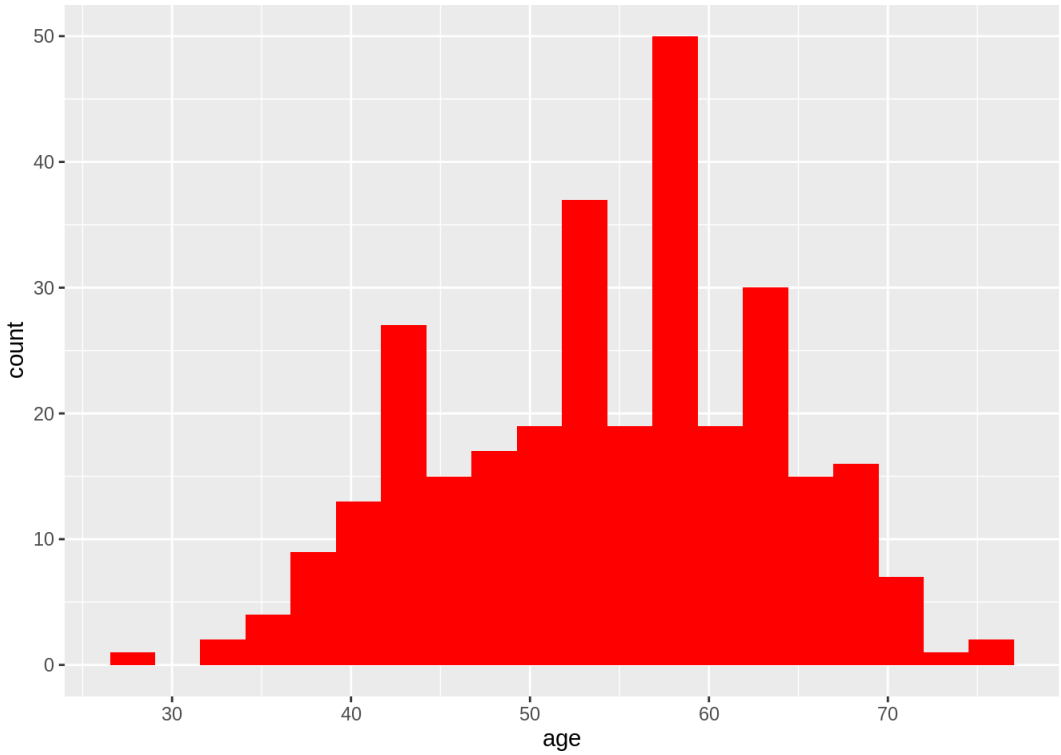
```
cinsiyet<- df %>% group_by(sex) %>% tally() %>% arrange(n, decreasing=T)
cinsiyet
```

```
## # A tibble: 2 × 2
##   sex       n
##   <fct> <int>
## 1 0        96
## 2 1       207
```

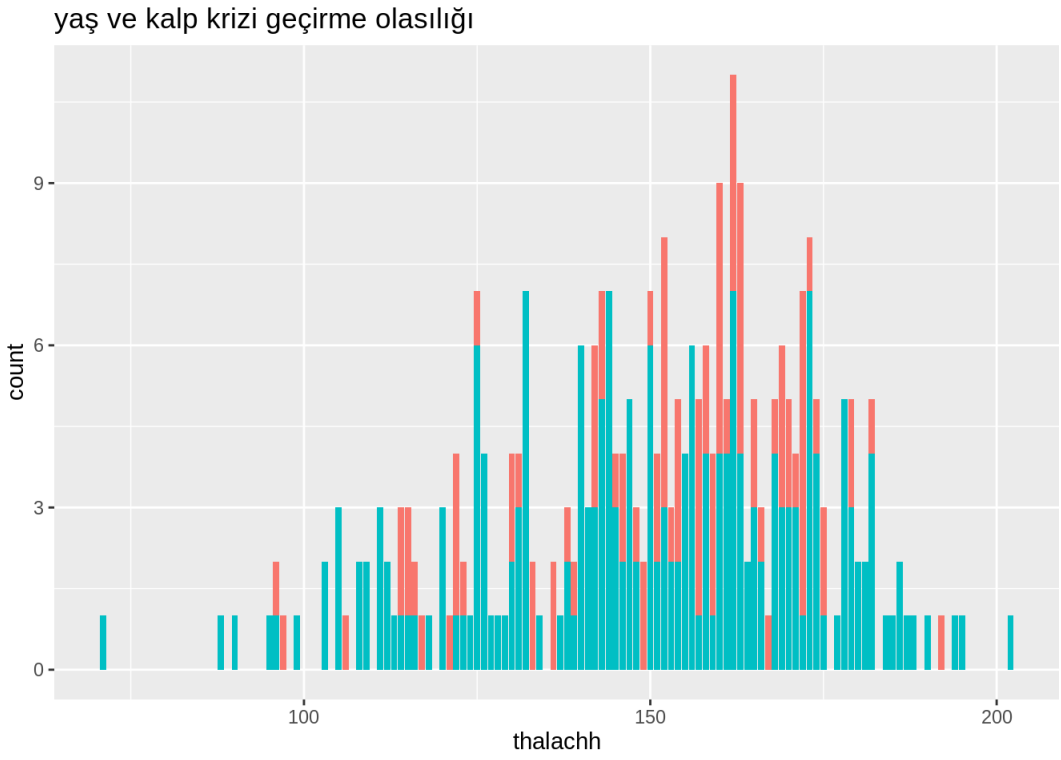
```
ggplot(data = df) +
 geom_bar(mapping = aes(x = sex))
```




```
ggplot(df ,aes(x=age)) +geom_histogram(bins=20,fill='red')
```




```
#maksimum kalp ve cinsiyet


ggplot(df, aes(x = thalachh, fill=factor(sex)))+
    geom_bar() +
    theme(legend.position = "yaş") +
    ggtitle("yaş ve kalp krizi geçirme olasılığı")
```

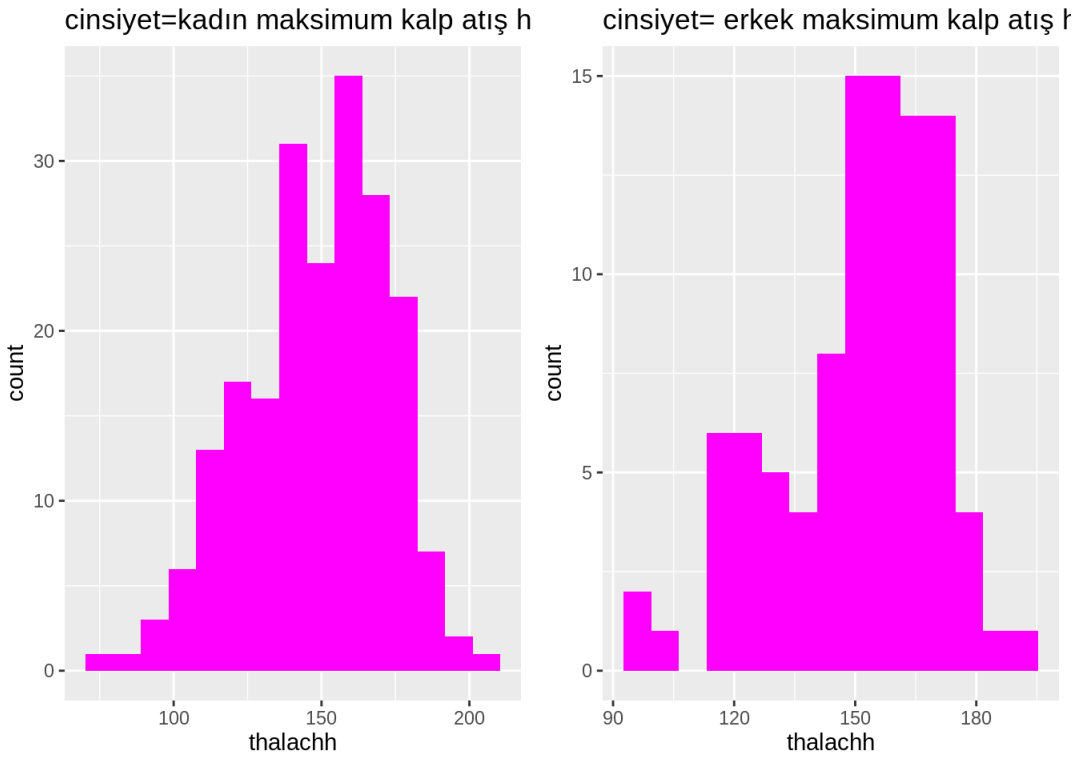



```
library(gridExtra)
```

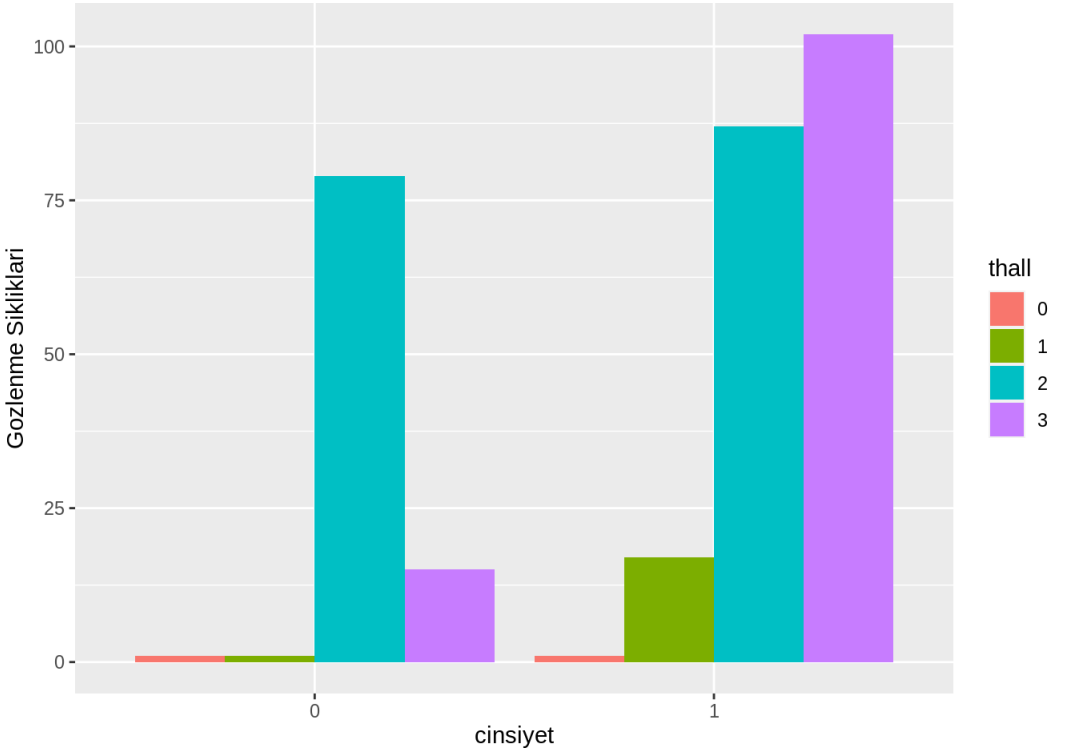
```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

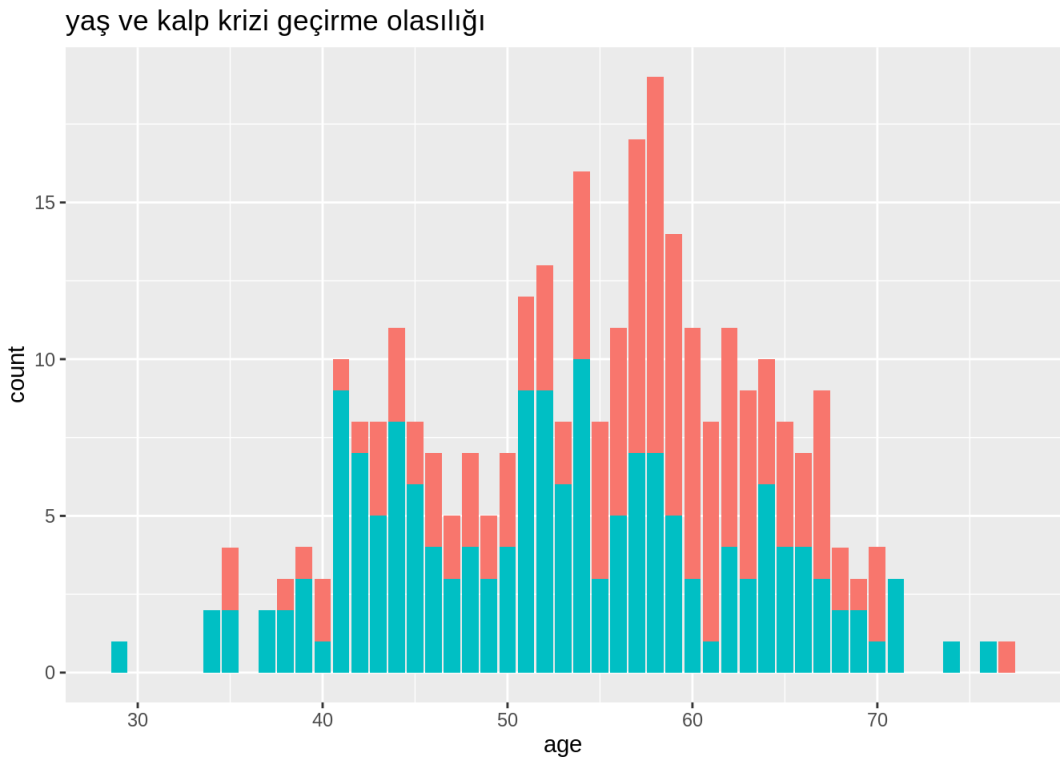
```
df1<- df %>% filter(sex == '1')

df2<- df %>% filter(sex == '0')
df11<- ggplot(df1,aes(x=thalachh))+geom_histogram(bins=15,fill='magenta') + ggtitle("cinsiyet=kadın maksimum kalp atış hızı")

df22<- ggplot(df2,aes(x=thalachh))+geom_histogram(bins=15,fill='magenta') + ggtitle("cinsiyet= erkek maksimum kalp atış hızı")

grid.arrange(df11,df22,nrow=1)
```



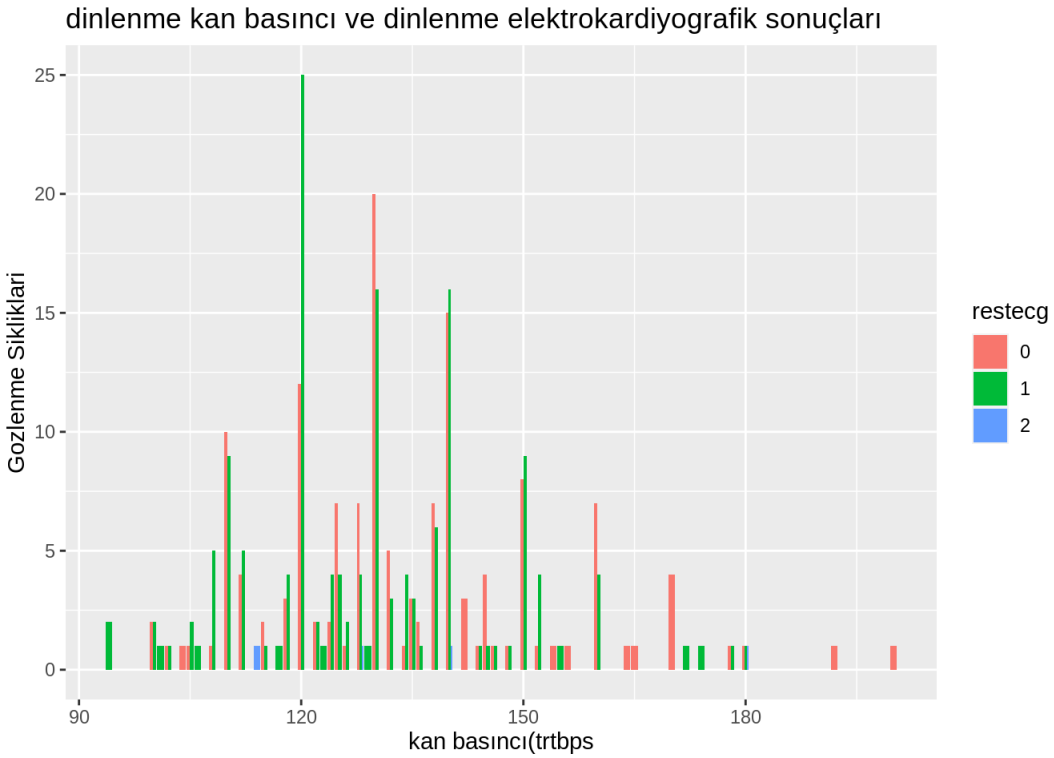

```
ggplot(df, aes(factor(sex), fill = thall)) +
  geom_bar(position = position_dodge()) +
  xlab("cinsiyet") +
  ylab("Gozlenme Sikliklari")
```



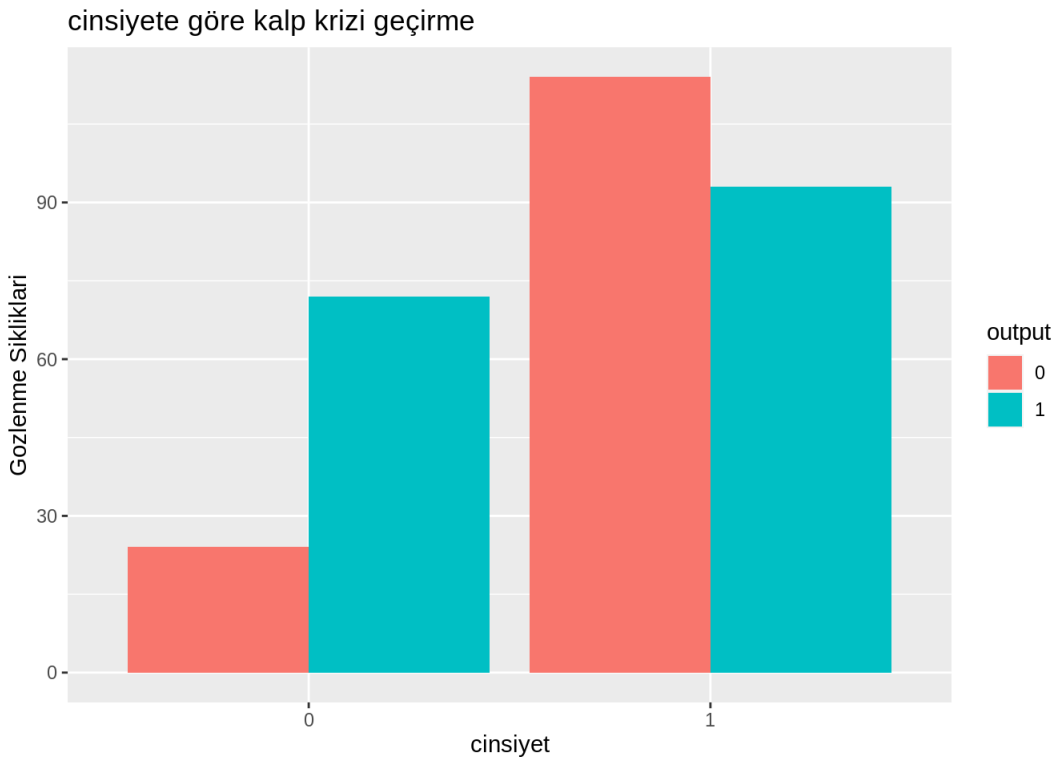

```
ggplot(df, aes(x = age, fill=factor(output)))+
    geom_bar() +
    theme(legend.position = "output") +
    ggtitle("yaş ve kalp krizi geçirme olasılığı")
```




```
ggplot(df, aes(trtbps, fill = restecg)) +
  geom_bar(position = position_dodge()) +
  ggtitle("dinlenme kan basıncı ve dinlenme elektrokardiyografik sonuçları") +
  xlab("kan basıncı(trtbps") +
  ylab("Gozlenme Sikliklari")
```




```
ggplot(df, aes(factor(sex), fill = output)) +
  geom_bar(position = position_dodge()) +
  ggtitle("cinsiyete göre kalp krizi geçirme ") +
  xlab("cinsiyet") +
  ylab("Gozlenme Sikliklari")
```

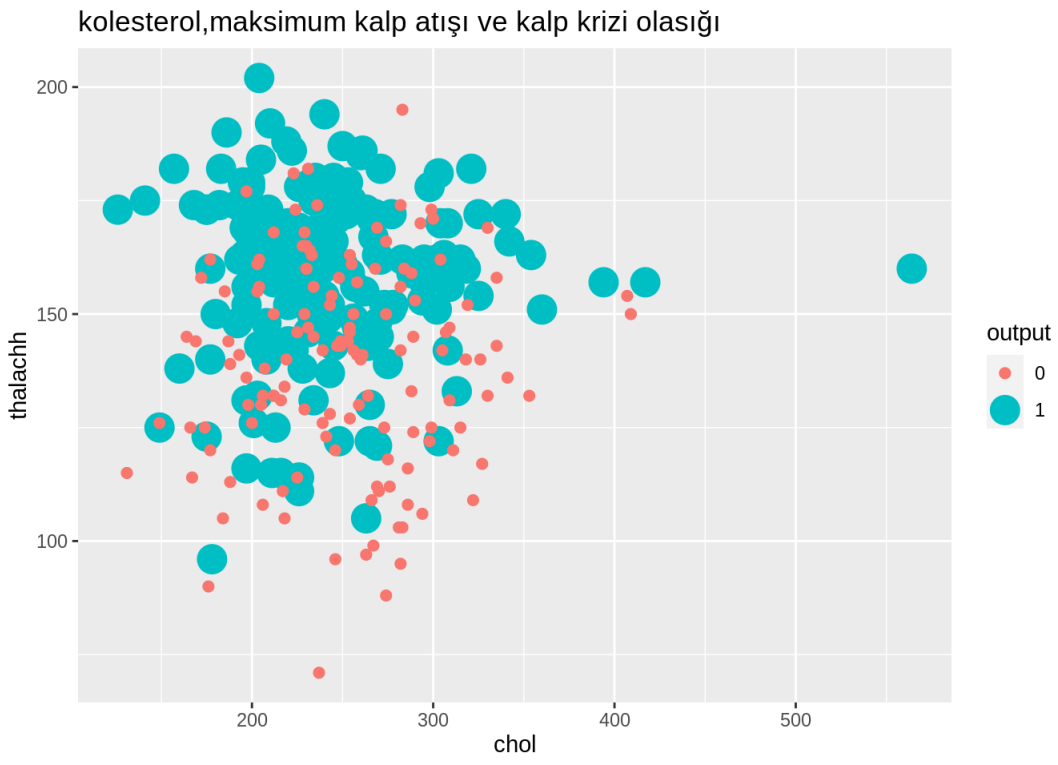



```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

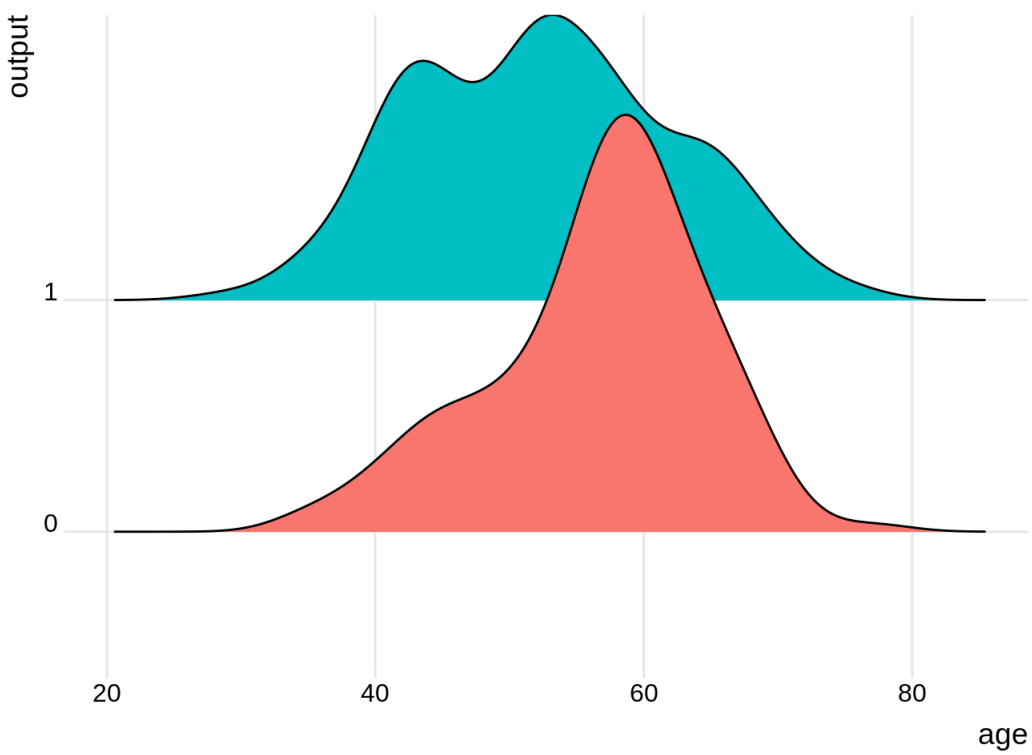
```
ggplot(df, aes(chol, thalachh, color = output, size = output )) +
  geom_point() +
   ggtitle("kolesterol,maksimum kalp atışı ve kalp krizi olasığı")
```

```
## Warning: Using size for a discrete variable is not advised.
```



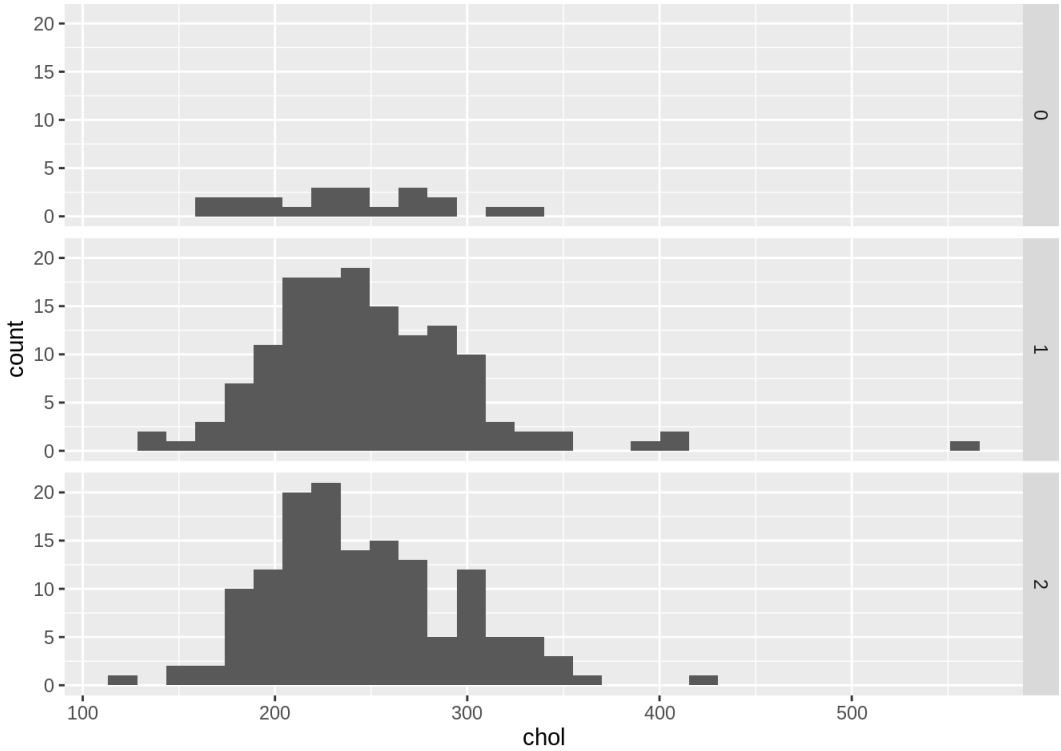

```
library(ggridges)

ggplot(df, aes(x = age, y = output, fill = cp)) +
  geom_density_ridges() +
  theme_ridges() +
  theme(legend.position = "none")
```

```
## Picking joint bandwidth of 2.8
```



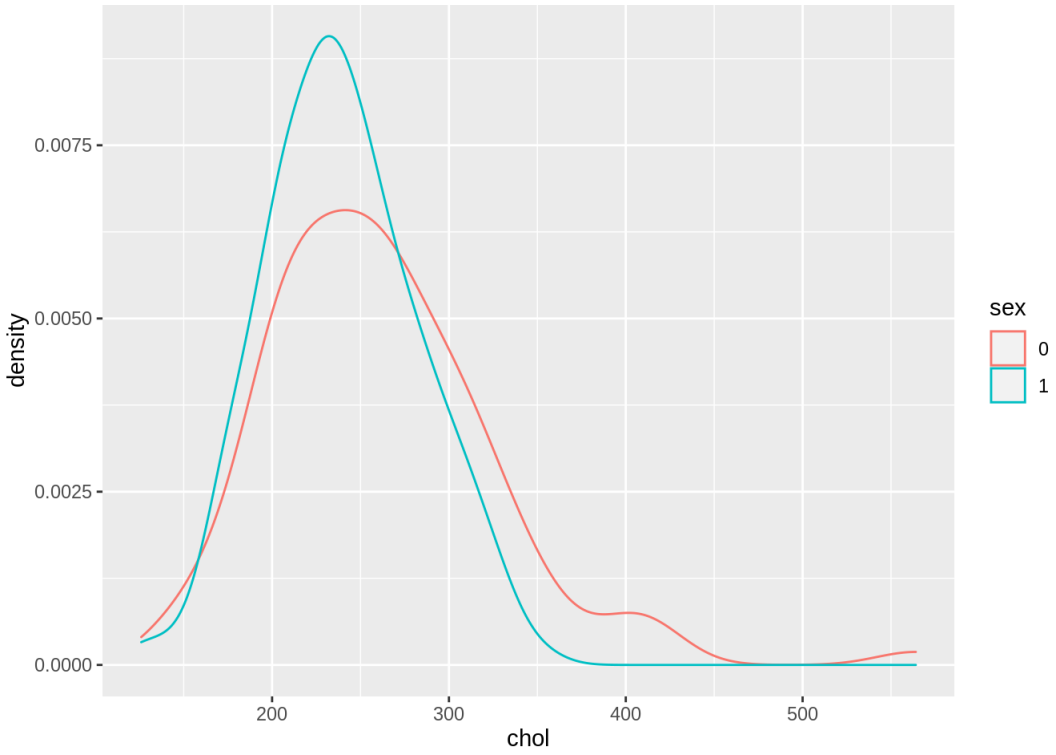

```
ggplot(df, aes(chol)) +
  geom_histogram() +                #ayrı ayrı verir
  facet_grid(slp ~. )
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

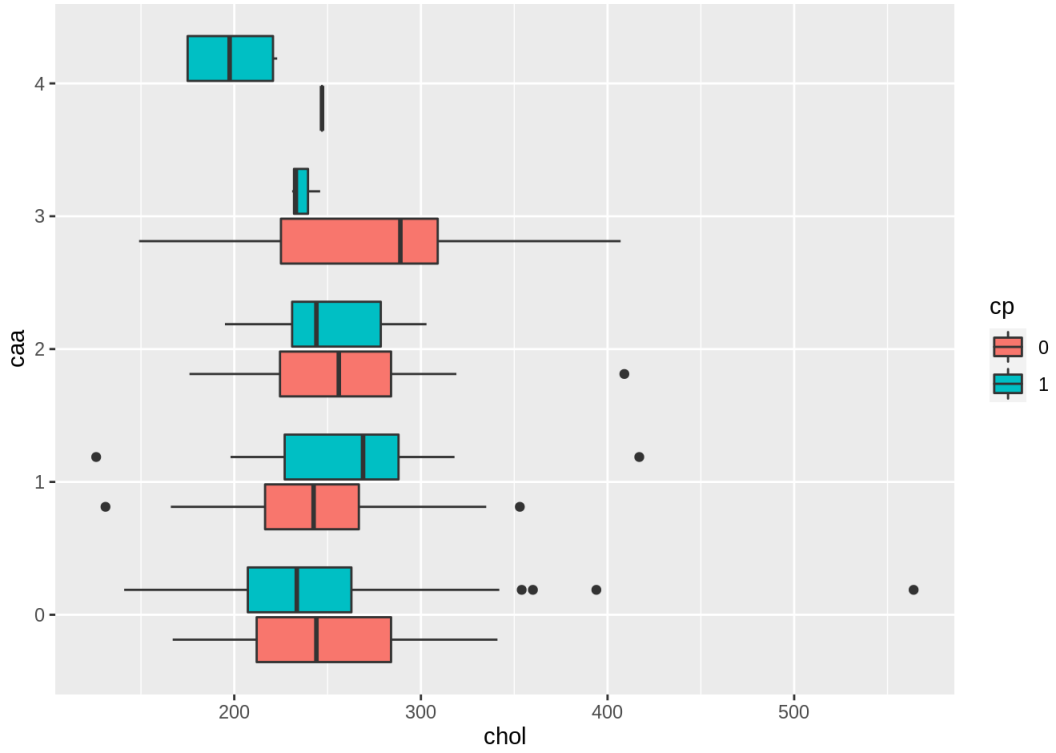



```
#çoklu frekans
ggplot(df, aes(chol, y = ..density..)) +
  geom_density(aes(colour = sex), binwidth = 500)
```

```
## Warning: Ignoring unknown parameters: binwidth
```

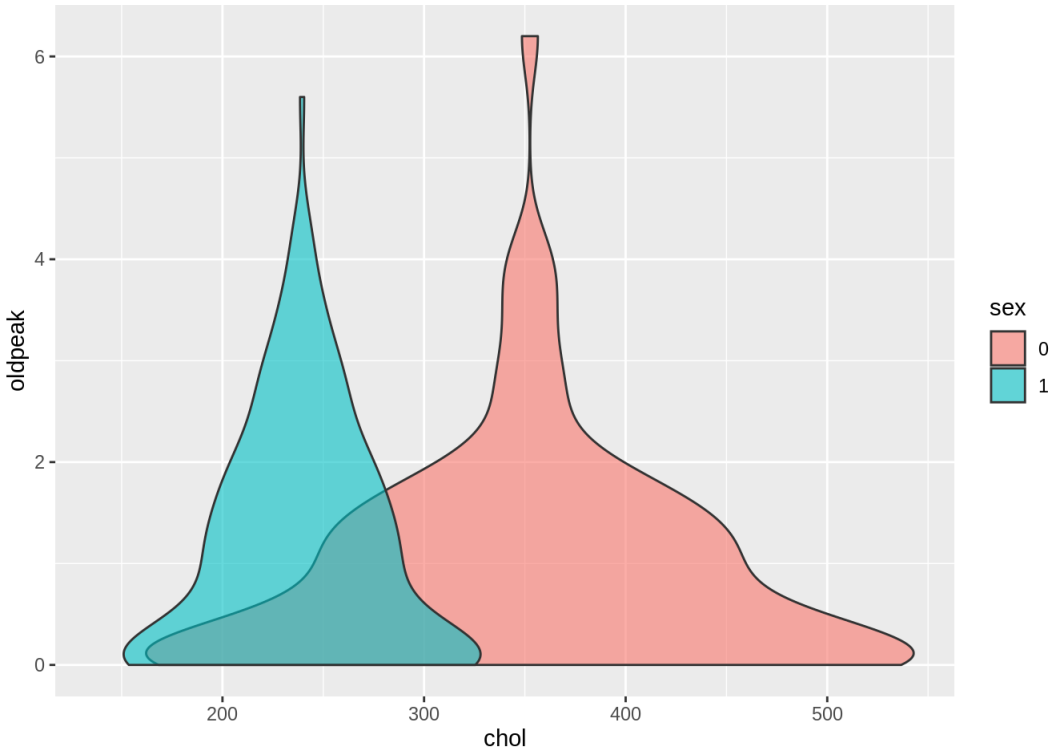



```
ggplot(df, aes(x = chol , y = caa, fill = cp)) +
      geom_boxplot()
```




```
ggplot(df, aes(x = chol, y = oldpeak, fill = sex)) +
      geom_violin(alpha = 0.6)
```

```
## Warning: position_dodge requires non-overlapping x intervals
```


6
4
2
0
oldpeak
200
300
400
500
chol
sex
0
1