

title: "logistic regression for the heart disease UCI"

author: "Rania Ahmed"

date: "October 29, 2019"

output:

pdf_document: default

html_document:

df_print: paged

Source: open source data <https://www.kaggle.com/ronitf/heart-disease-uci>

Description:

age: age in years

sex: (1 = male; 0 = female)

cp: chest pain type – Value 0: asymptomatic – Value 1: atypical angina – Value 2: non-anginal pain – Value 3: typical angina

trestbps: resting blood pressure (in mm Hg on admission to the hospital)

chol: serum cholestoral in mg/dl

fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

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

thalach: maximum heart rate achieved

exang: exercise induced angina (1 = yes; 0 = no)

oldpeak: ST depression induced by exercise relative to rest

slope: the slope of the peak exercise ST segment – 0: downsloping; 1: flat; 2: upsloping
ca: number of major vessels (0-3) colored by flourosopy

thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

target: 1 (disease) or 0 (no disease)

Primary outcome is getting the reasons of Heart Disease UDI

```
heart <- read.csv(file.choose(), header = T)
attach(heart)
View(heart)
library(knitr)

sex_dummy <- ifelse(heart$sex=="1", "male", "female")
cpasym <- ifelse(heart$cp=="0", "1", "0")
cpatyp <- ifelse(heart$cp=="1", "1", "0")
cpnon_ang <- ifelse(heart$cp=="2", "1", "0")
cptyp <- ifelse(heart$cp=="3", "1", "0")
fbs_dummy <- ifelse(fbs=="1", "true", "false")
restecg_notnorm <- ifelse(heart$restecg=="0", "1", "0")
restecg_norm <- ifelse(heart$restecg=="1", "1", "0")
restecg_abnor <- ifelse(heart$restecg=="2", "1", "0")
exang_dummy <- ifelse(heart$exang=="1", "yes", "no")
slope_down <- ifelse(heart$slope=="0", "1", "0")
slope_flat <- ifelse(heart$slope=="1", "1", "0")
slope_up <- ifelse(heart$slope=="2", "1", "0")
thal_nor <- ifelse(heart$thal=="1", "1", "0")
thal_zero <- ifelse(heart$thal=="0", "1", "0")
thal_fixed <- ifelse(heart$thal=="2", "1", "0")
thal_reves <- ifelse(heart$thal=="3", "1", "0")
target_dummy <- ifelse(heart$target=="1", "disease", "no disease")
ex <- cbind(sex_dummy,
cpasym, cpatyp, cpnon_ang, cptyp, fbs_dummy, restecg_notnorm, restecg_norm, re
stecg_abnor, exang_dummy, slope_down, slope_flat, slope_up, thal_nor, thal_ze
ro, thal_fixed, thal_reves, target_dummy)
newheart <- cbind(heart, ex)
write.csv(newheart, file = "newheart.csv")

attach(newheart)

## The following objects are masked _by_ .GlobalEnv:
##
##      cpasym, cpatyp, cpnon_ang, cptyp, exang_dummy, fbs_dummy,
##      restecg_abnor, restecg_norm, restecg_notnorm, sex_dummy,
##      slope_down, slope_flat, slope_up, target_dummy, thal_fixed,
##      thal_nor, thal_reves, thal_zero

## The following objects are masked from heart:
##
##      ca, chol, cp, exang, fbs, Ã..age, oldpeak, restecg, sex,
##      slope, target, thal, thalach, trestbps

summary(newheart)
```

```

##      age      sex      cp      trestbps
## 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      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      sex_dummy      cpasym      cpatyp
cpnon_ang
## Min.   :0.000   Min.   :0.0000   female: 96   0:160   0:253   0:216
## 1st Qu.:2.000   1st Qu.:0.0000   male  :207   1:143   1: 50   1: 87
## 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
##      ctpyp      fbs_dummy      restecg_notnorm      restecg_norm      restecg_abnor
## 0:280   false:258   0:156           0:151           0:299
## 1: 23    true : 45   1:147           1:152           1: 4
##
##
##
##      exang_dummy      slope_down      slope_flat      slope_up      thal_nor      thal_zero
thal_fixed
## no :204           0:282           0:163           0:161           0:285           0:301           0:137
## yes: 99          1: 21           1:140           1:142           1: 18           1: 2           1:166
##
##
##
##      thal_reves      target_dummy
## 0:186      disease :165
## 1:117      no disease:138
##
##

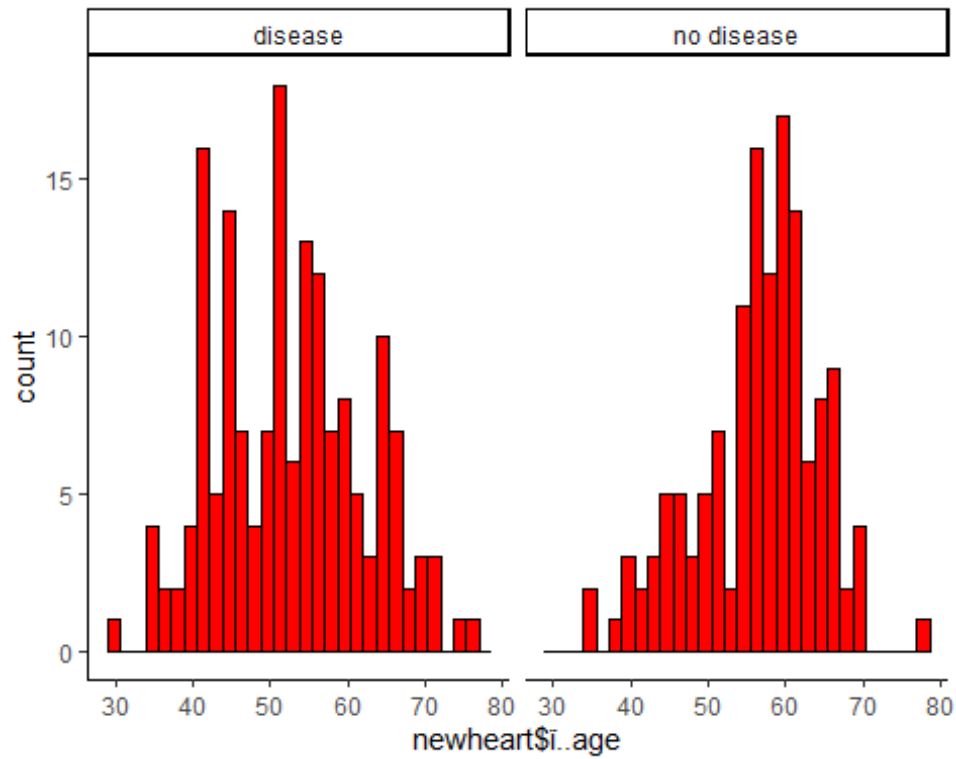
```

```
##  
##
```

Analysis of variances

1- Age:

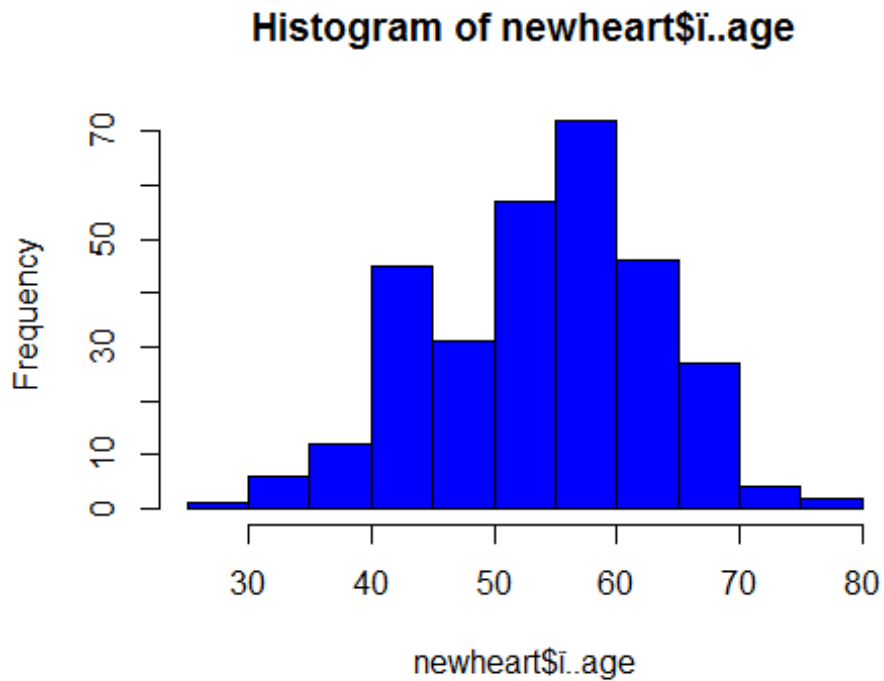
```
library(dplyr)  
  
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union  
  
library(finalfit)  
library(ggplot2)  
age <- group_by(newheart, newheart$target_dummy)%>%  
  summarise(count=n(), mean=mean(newheart$Ã~..age),  
    sd=sd(newheart$Ã~..age), var= var(newheart$Ã~..age))  
age  
  
## # A tibble: 2 x 5  
##   `newheart$target_dummy` count  mean    sd   var  
##   <fct>                <int> <dbl> <dbl> <dbl>  
## 1 disease                165  54.4  9.08  82.5  
## 2 no disease             138  54.4  9.08  82.5  
  
ggplot(newheart, aes(newheart$Ã~..age, fill=newheart$Ã~..age))+geom_histo  
gram  
(stat="bin", color="black", fill="red")+facet_wrap(~target_dummy, ncol=2, s  
cale="fixed")+theme_classic()  
  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
summary(newheart$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  29.00   47.50   55.00   54.37   61.00   77.00
```

```
hist(newheart$age, col="blue")
```



```
library(moments)
skewness(newheart$ĩ~..age)

## [1] -0.2014597

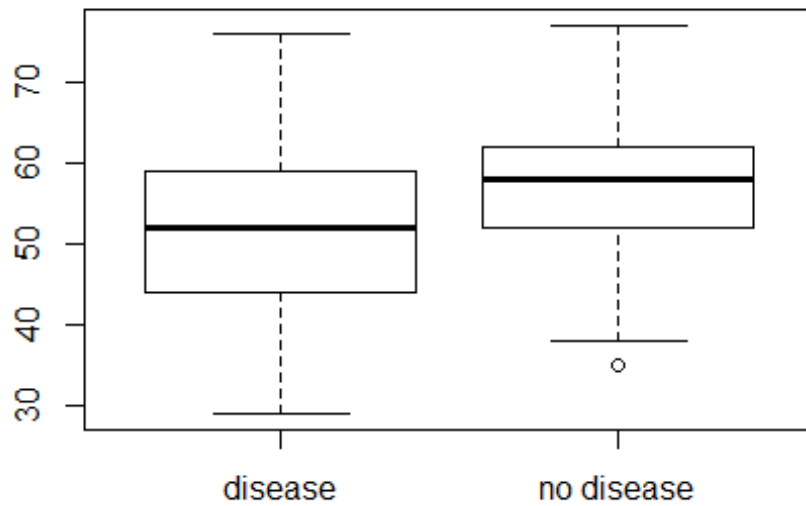
kurtosis(newheart$ĩ~..age)

## [1] 2.447001

shapiro.test(newheart$ĩ~..age)

##
## Shapiro-Wilk normality test
##
## data:  newheart$ĩ~..age
## W = 0.98637, p-value = 0.005798

boxplot(newheart$ĩ~..age~newheart$target_dummy)
```



```
wilcox.test(newheart$age~newheart$target_dummy, mu=0,  
alternative="two.sided", var.equal=T, conf.level=0.95)  
  
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: newheart$age by newheart$target_dummy  
## W = 8240.5, p-value = 3.439e-05  
## alternative hypothesis: true location shift is not equal to 0
```

Mean of the age of the data is 54.4 for patients who suffering disease and 54.4 for those don't suffer from disease while standard deviation of the sample is 9.08 for both patients. The range of age is between 29 and 77 years old.

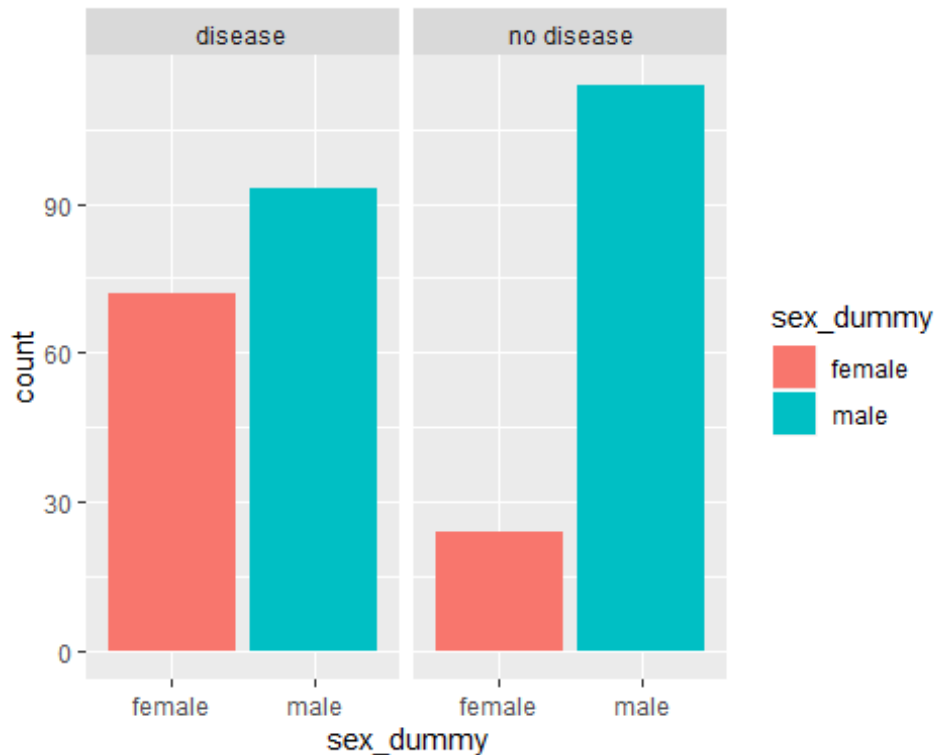
checking the normality:

the data is highly skewed and kurtosis test is out of its acceptable range. Statistical test(shapiro test) is 0.0058 which is less than 0.05(significant level). so the sample is not normally distributed. the data is violated.

The p value of wilcoxon rank test is 3.439e-05 which is less than 0.05(significance level alpha) so there is significant difference between thier medians and reject null hypothesis.

2- SEX:

```
sex <- as.data.frame(newheart %>%  
  
summary_factorlist("sex_dummy", "target_dummy", p=F, add_dependent_label  
= T, cont_cut = 1))  
sex  
  
##   Dependent: sex_dummy          female      male  
## 1      target_dummy    disease 72 (43.6)  93 (56.4)  
## 2              no disease 24 (17.4) 114 (82.6)  
  
library(ggplot2)  
ggplot(newheart, aes(sex_dummy, fill=sex_dummy))+geom_bar(stat="count")+f  
acet_wrap(~target_dummy, ncol=2, scale="fixed")+theme_get()
```

```
chisq.test(table(target_dummy, sex_dummy))

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(target_dummy, sex_dummy)
## X-squared = 22.717, df = 1, p-value = 1.877e-06
```

the count of diseased male are 93 patients (56.4%) and female are 72 patients (43.6%) while healthy male are 114 (82.6%) and female 24(17.4%)

p value of chi square test is 1.877e-06 is less than 0.05 (significance level alpha) so there is significant difference between female and male who are suffering from heart disease UDI and will reject null hypothesis.

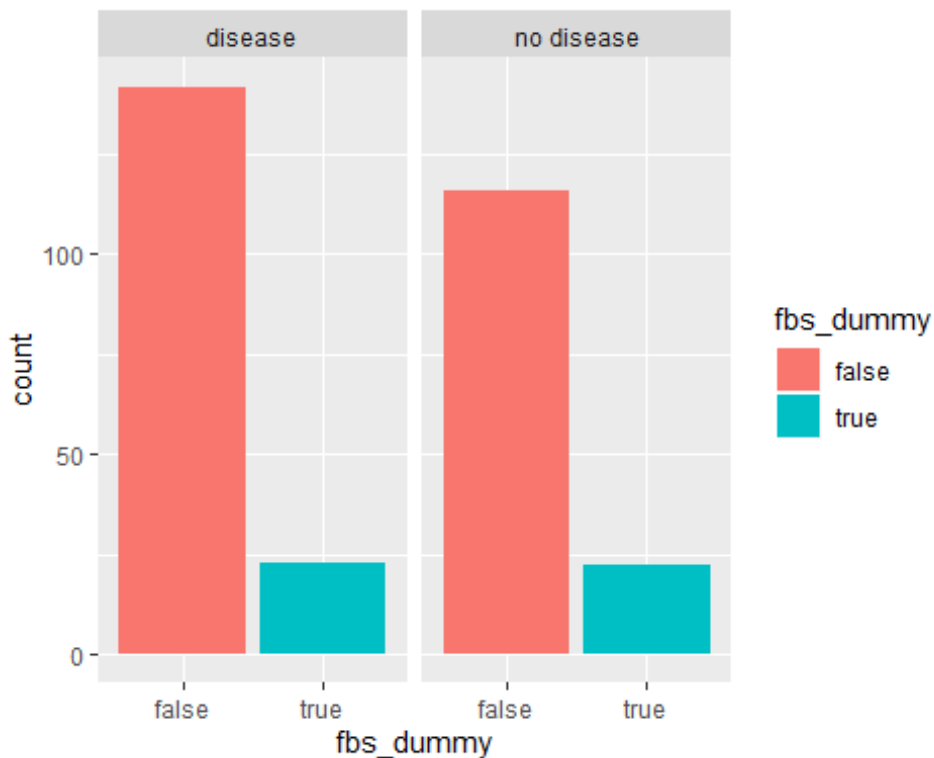
3- Fbs:

```
fbs <- as.data.frame(newheart %>%
  summary_factorlist("fbs_dummy", "target_dummy",
```

```
p=F, add_dependent_label = T, cont_cut = 1))
fbs
```

```
##   Dependent: fbs_dummy           false      true
## 1      target_dummy    disease 142 (86.1) 23 (13.9)
## 2              no disease 116 (84.1) 22 (15.9)
```

```
ggplot(newheart, aes(fbs_dummy, fill=fbs_dummy))+geom_bar(stat =
"count")+facet_wrap(~target_dummy)+theme_get()
```



```
chisq.test(table(newheart$target_dummy, newheart$fbs_dummy))
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: table(newheart$target_dummy, newheart$fbs_dummy)
## X-squared = 0.10627, df = 1, p-value = 0.7444
```

p value of chisquare test is 0.744 which is greater than 0.05(significance level alpha) so there is no significant difference between diabetic patients and heart disease UDI so we fail to reject null hypothesis.

4- trestbps:

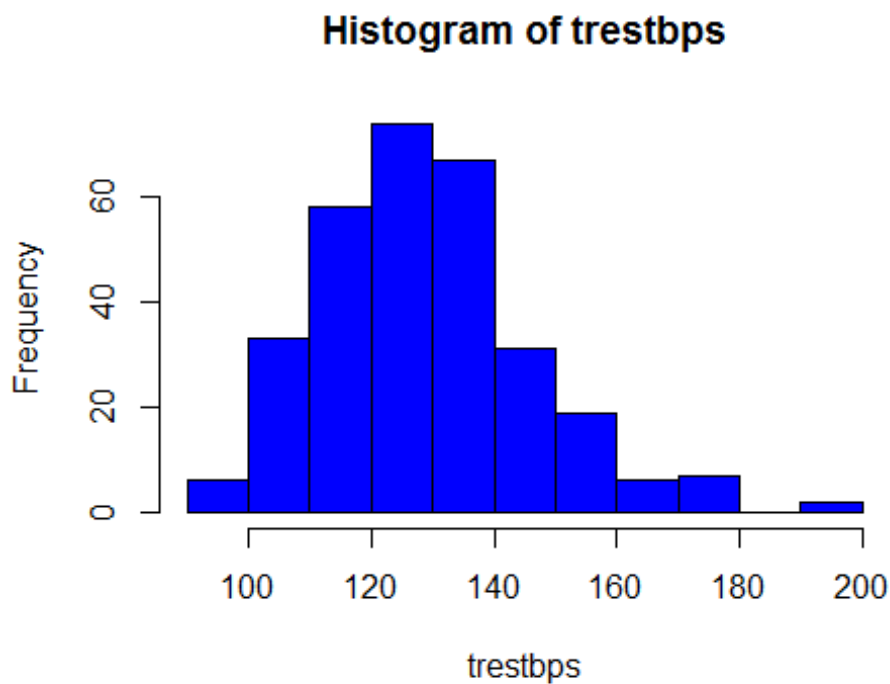
```
tres <- group_by(newheart, target_dummy) %>% summarise(count=n(), mean=
mean(trestbps), sd=sd(trestbps), var=var(trestbps))
tres
```

```
## # A tibble: 2 x 5
##   target_dummy count  mean    sd   var
##   <fct>         <int> <dbl> <dbl> <dbl>
## 1 disease         165  129.  16.2  261.
## 2 no disease      138  134.  18.7  351.
```

```
summary(trestbps)
```

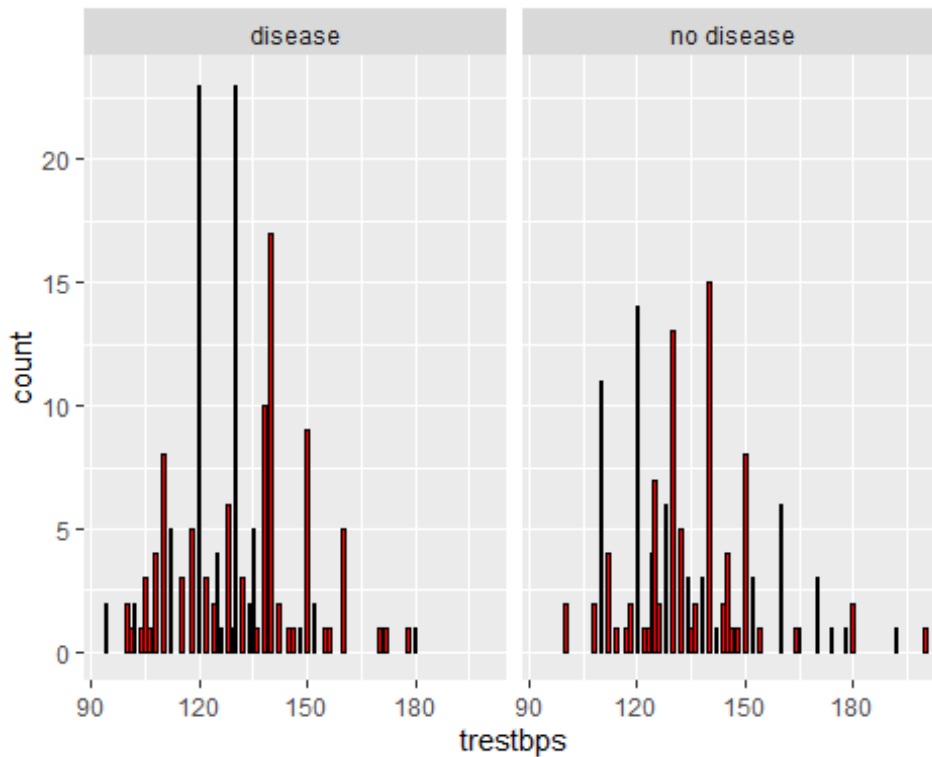
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      94.0   120.0   130.0   131.6   140.0   200.0
```

```
hist(trestbps, col="blue")
```



```
ggplot(newheart,aes(trestbps,fill=trestbps))+geom_histogram(stat="count",color="black",fill="red")+facet_wrap(~target_dummy)+theme_get()
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



```
skewness(trestbps)
```

```
## [1] 0.7102301
```

```
kurtosis(trestbps)
```

```
## [1] 3.894057
```

```
shapiro.test(trestbps)
```

```
##
```

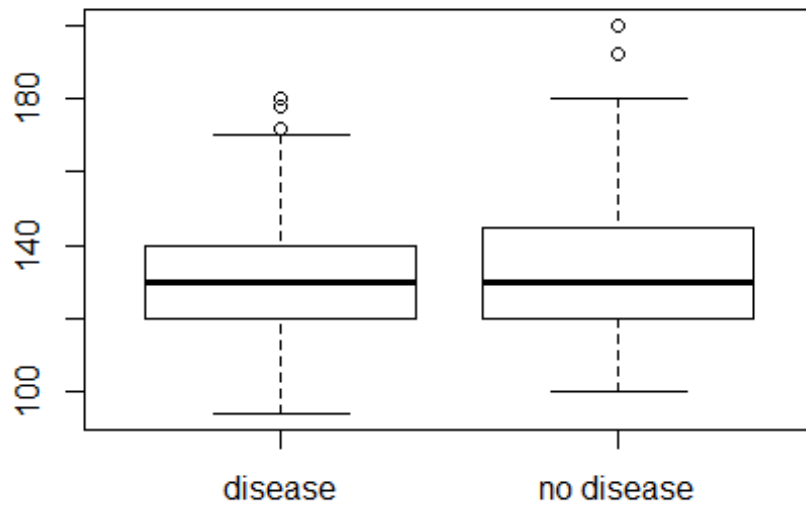
```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: trestbps
```

```
## W = 0.96592, p-value = 1.458e-06
```

```
boxplot(trestbps~target_dummy)
```



```
wilcox.test(trestbps~target_dummy, mu=0, alternative = "two.sided",  
var.equal= F, data = newheart)
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: trestbps by target_dummy  
## W = 9784.5, p-value = 0.03465  
## alternative hypothesis: true location shift is not equal to 0
```

The mean of trestbps (resting blood pressure) is 129.3 in patients suffering from heart disease UDI while in healthy patients is 134.4. the standard deviation in patients suffering HDI is 16.16 while in healthy patient is 18.73.the range of trestbps from 94 to 200 mm Hg.

check normality:

in numerical tests of normality , data is highly skewed and out of acceptable range of kurtosis test.

in statistical test , data is violated . p value is less than 0.05 which is less than the significance level (0.05). so the data is not normally distributed.

in wilcoxon rank test , p-value is less than 0.05 so there is significant difference between presence of HDI and trestbps so we reject null hypothesis.

5-Chol:

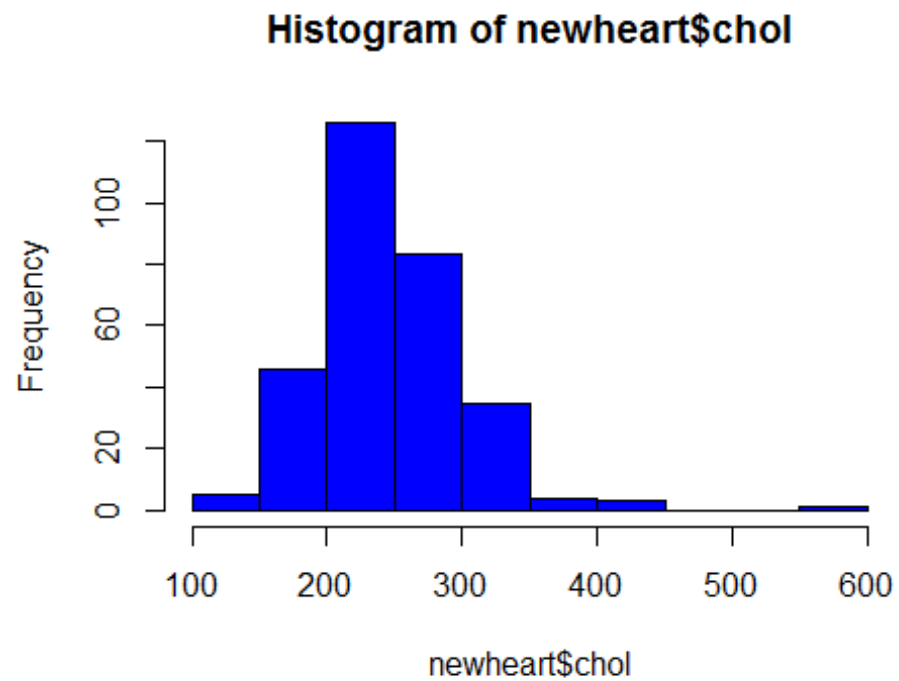
```
cholcoun <- group_by(newheart, target_dummy)%>% summarise(count=n(),
mean=mean(chol), sd=sd(chol), var=var(chol))
cholcoun

## # A tibble: 2 x 5
##   target_dummy count  mean    sd   var
##   <fct>         <int> <dbl> <dbl> <dbl>
## 1 disease         165  242.   53.6 2868.
## 2 no disease      138  251.   49.5 2446.

summary(newheart$chol)

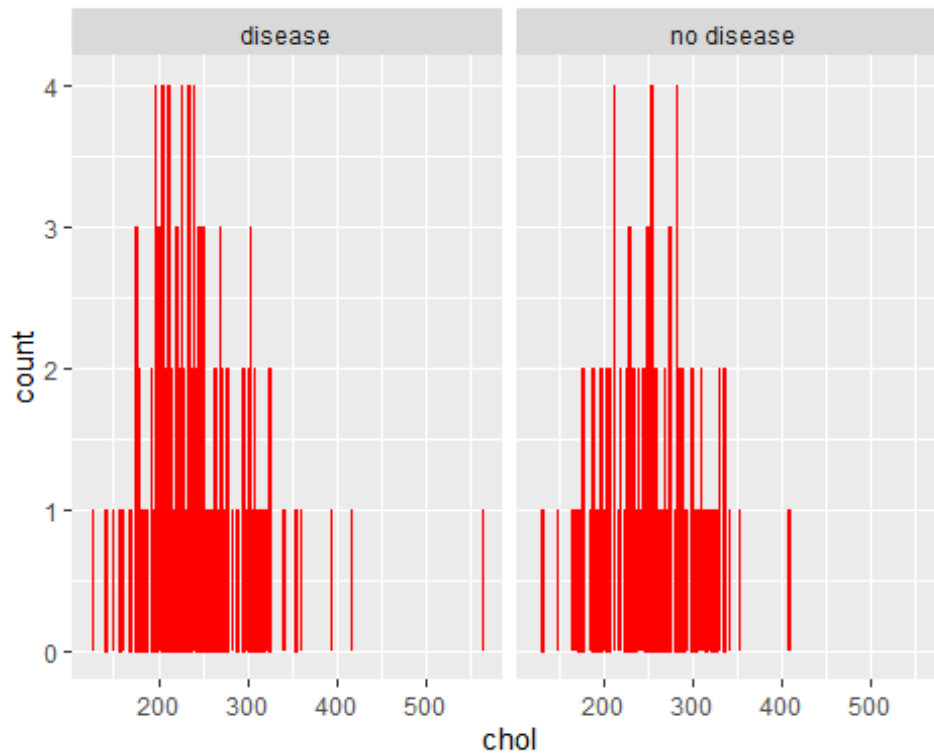
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  126.0   211.0   240.0   246.3   274.5   564.0

hist(newheart$chol, col = "blue")
```

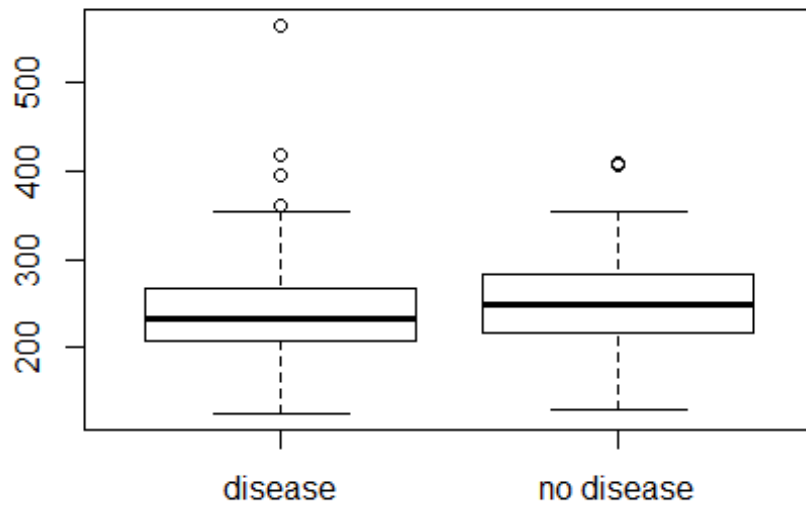


```
ggplot(newheart, aes(chol, fill=chol))+geom_histogram(stat="count", color="red", fill="black")+facet_wrap(~target_dummy)+theme_get()
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



```
skewness(newheart$chol)
## [1] 1.137733
kurtosis(newheart$chol)
## [1] 7.411682
shapiro.test(newheart$chol)
##
## Shapiro-Wilk normality test
##
## data:  newheart$chol
## W = 0.94688, p-value = 5.365e-09
boxplot(newheart$chol~target_dummy)
```

```
wilcox.test(newheart$chol~target_dummy, mu=0, alternative="two.sided",  
var.equal = F, data = newheart)
```

```
##
```

```
## Wilcoxon rank sum test with continuity correction
```

```
##
```

```
## data: newheart$chol by target_dummy
```

```
## W = 9789.5, p-value = 0.03572
```

```
## alternative hypothesis: true location shift is not equal to 0
```

the mean of Cholestrol level in patients suffering HDI is 242.23 while in healthy patients is 251.1. The standard deviation of chol. level in patients suffering HDI is 53.55 while in healthy patients is 49.45. the cholestrol level is from 126 to 564 mg/dl.

check normality:

in numerical test of normality, the data is highly skewed and out of acceptable range of kurtosis test.

in statistical test of normality, the data is violated , p- value is less than 0.05 so the data is not normally distributed.

in wilcoxon rank test p- value is less than 0.05 so there is significant difference between presence of HDI and cholesterol level. so we reject null hypothesis.

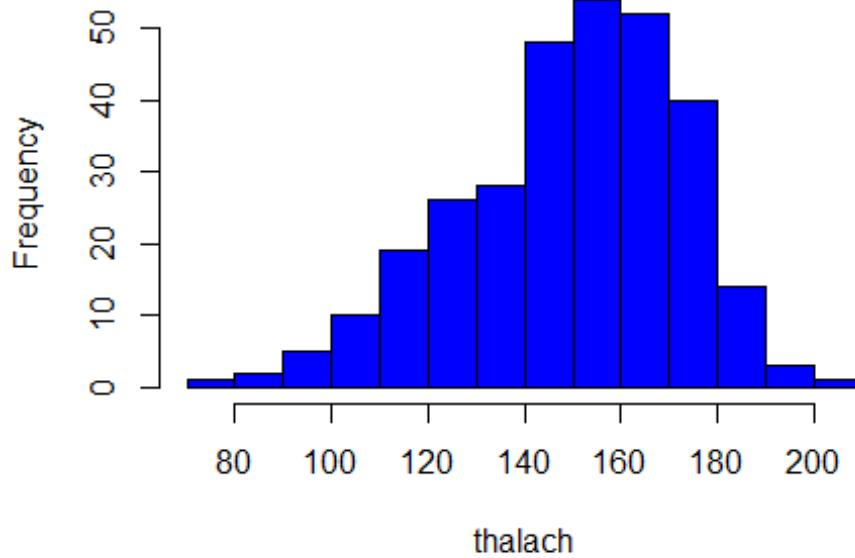
6- thalach:

```
thalachmean <- group_by(newheart, target_dummy) %>%
  summarise(count=n(), mean= mean(thalach), sd=sd(thalach), var=
  var(thalach))
thalachmean

## # A tibble: 2 x 5
##   target_dummy count  mean    sd   var
##   <fct>         <int> <dbl> <dbl> <dbl>
## 1 disease         165  158.  19.2  368.
## 2 no disease      138  139.  22.6  511.

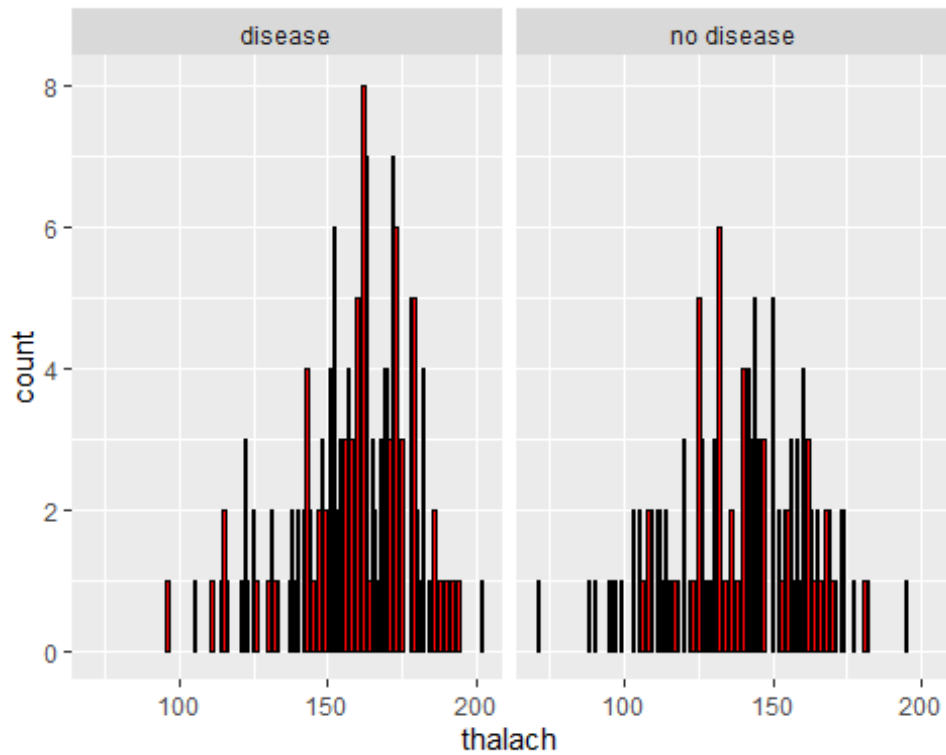
hist(thalach, col="blue")
```

Histogram of thalach



```
ggplot(newheart, aes(thalach, fill=thalach))+geom_histogram(stat="count",  
color="black", fill="red")+facet_wrap(~target_dummy)+theme_get()
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



```
summary(thalach)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      71.0  133.5   153.0   149.6  166.0   202.0
```

```
skewness(thalach)
```

```
## [1] -0.5347455
```

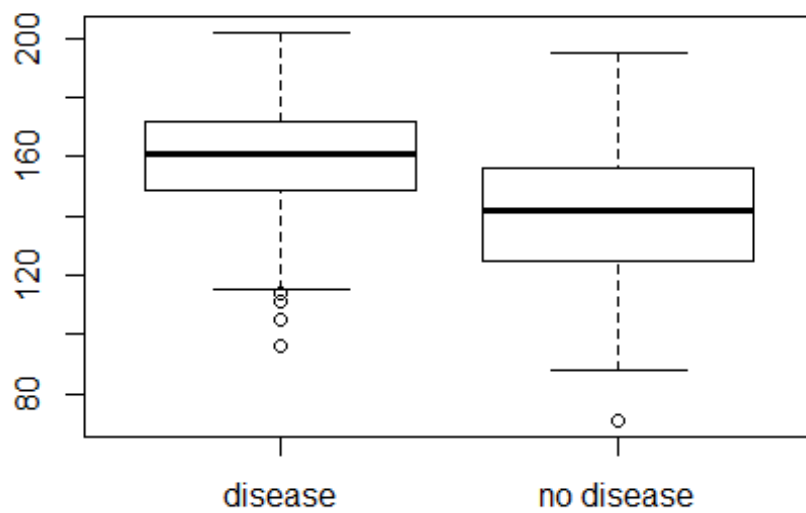
```
kurtosis(thalach)
```

```
## [1] 2.919311
```

```
shapiro.test(thalach)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  thalach
## W = 0.97632, p-value = 6.621e-05
```

```
boxplot(thalach~target_dummy)
```



```
wilcox.test(thalach~target_dummy, mu=0, alternative="two.sided",  
var.equal= F, data = newheart)
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: thalach by target_dummy  
## W = 17038, p-value = 9.797e-14  
## alternative hypothesis: true location shift is not equal to 0
```

the mean of thalach(maximum heart rate achieved) in patients with HDI is 158.47 while in healthy patients is 139.1. the standard deviation of thalach in patients with HDI is 19.17 while in healthy patients is 22.59. thalach is from 71 to 202 beat/sec (max.).

check normality:

in numerical test, the data is highly skewed and out of acceptable range of kurtosis test

in statistical test, the data is violated. p-value is less than 0.05 so the data is not normally distributed.

in wilcoxon test, p-value is less than 0.05 so there is significant difference between presence of HDI and thalach so we reject null hypothesis.

7- oldpeak:

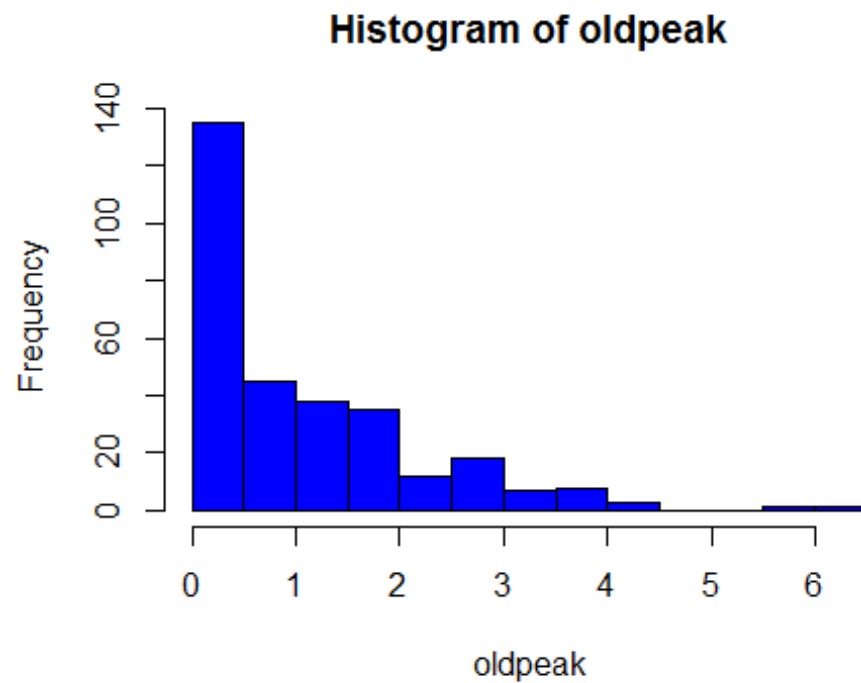
```
oldpeakmean <- group_by(newheart, target_dummy)%>% summarise(count=n(),
mean=mean(oldpeak), sd=sd(oldpeak), var= var(oldpeak))
oldpeakmean

## # A tibble: 2 x 5
##   target_dummy count  mean    sd   var
##   <fct>         <int> <dbl> <dbl> <dbl>
## 1 disease         165 0.583 0.781 0.609
## 2 no disease      138 1.59  1.30  1.69

summary(oldpeak)

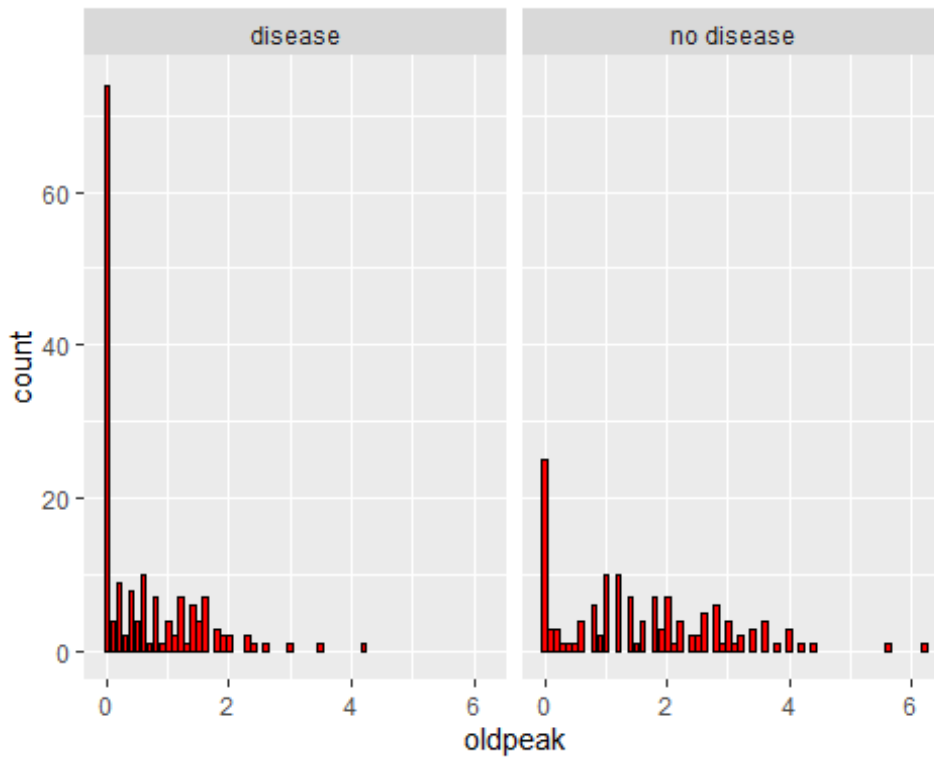
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   0.00   0.80   1.04   1.60   6.20

hist(oldpeak, col="blue")
```

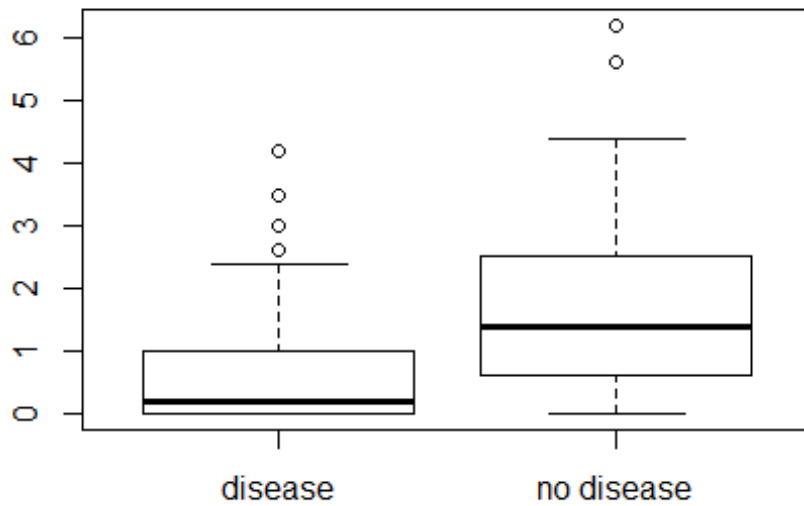


```
ggplot(newheart, aes(oldpeak, fill=oldpeak))+geom_histogram(stat="count",  
color="black", fill="red")+facet_wrap(~target_dummy)+theme_get()
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



```
skewness(oldpeak)
## [1] 1.263426
kurtosis(oldpeak)
## [1] 4.530193
shapiro.test(oldpeak)
##
## Shapiro-Wilk normality test
##
## data:  oldpeak
## W = 0.84418, p-value < 2.2e-16
boxplot(oldpeak~target_dummy)
```

```
wilcox.test(oldpeak~target_dummy, mu=0, alternative= "two.sided",  
var.equal= F, data = newheart)
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: oldpeak by target_dummy  
## W = 5922, p-value = 2.407e-13  
## alternative hypothesis: true location shift is not equal to 0
```

The mean of oldpeak(ST depression induced by exercise relative to rest) in patients with HDI is 0.58 while in healthy patients is 1.59. The standard deviation in patients with HDI is 0.78 while in healthy patients is 1.3. The range of oldpeak is from 0 to 6.20

Check normality:

in numerical test , data is highly skewed and out of the acceptable range of kurtosis test.

statistical test, the data is violated. p-value is less than 0.05 so the data is not normally distributed.

in wilcoxon rank test, p-value is less than 0.05 so there is significant difference between presence of HDI and oldpeak so we reject null hypothesis.

8- ca:

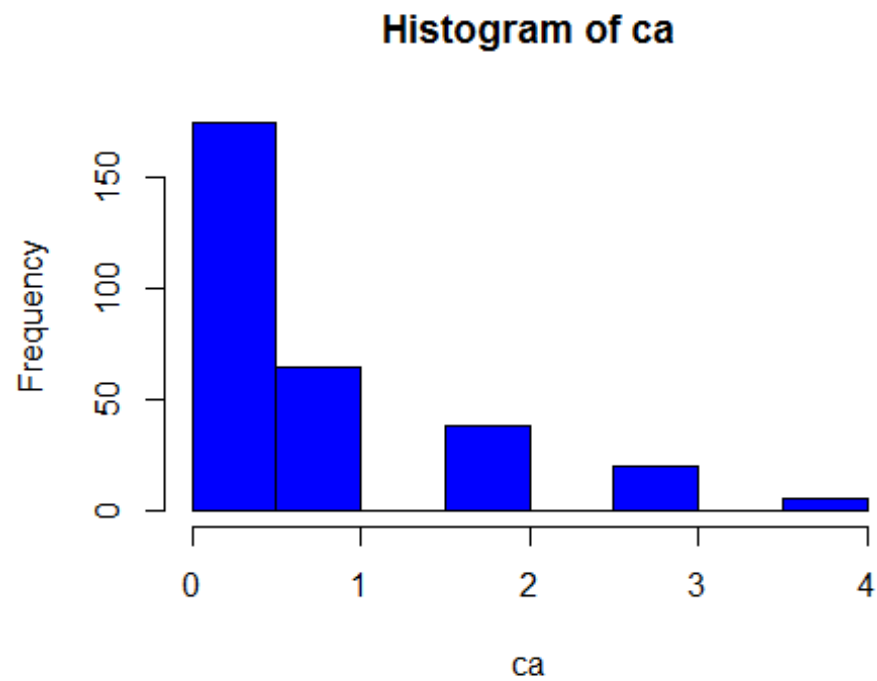
```
camean <- group_by(newheart, target_dummy)%>% summarise(count=n(),
mean=mean(ca), sd=sd(ca), var= var(ca))
camean

## # A tibble: 2 x 5
##   target_dummy count  mean    sd   var
##   <fct>         <int> <dbl> <dbl> <dbl>
## 1 disease         165 0.364 0.849 0.721
## 2 no disease      138 1.17  1.04  1.09

summary(ca)

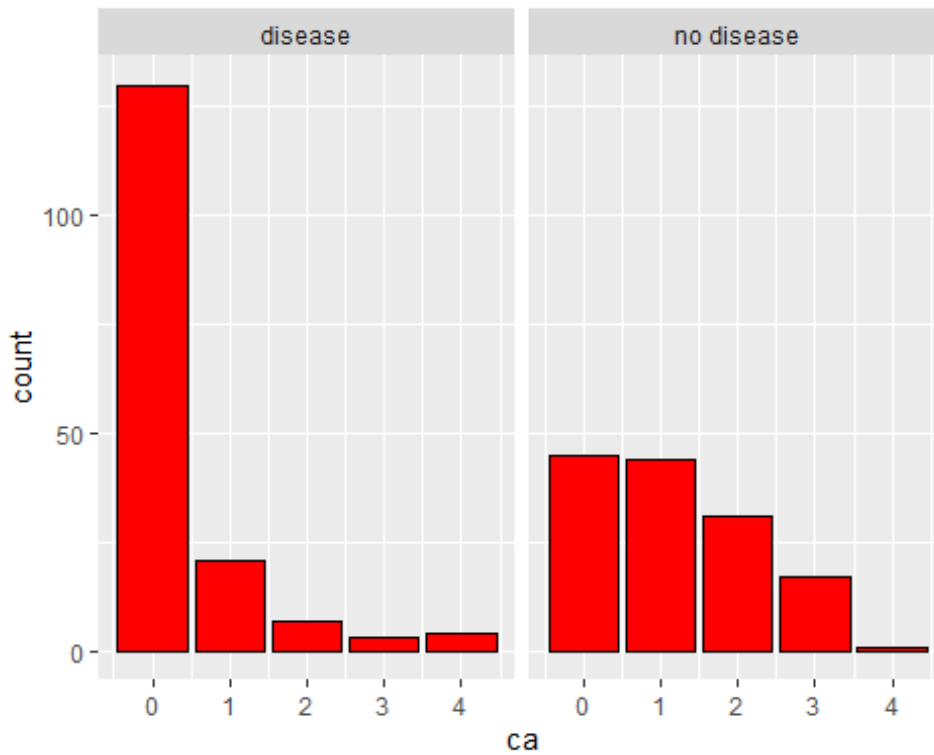
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000 0.0000  0.0000  0.7294  1.0000  4.0000

hist(ca, col="blue")
```



```
ggplot(newheart, aes(ca, fill=ca))+geom_histogram(stat="count", color="black", fill="red")+facet_wrap(~target_dummy)+theme_get()
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



```
skewness(ca)
```

```
## [1] 1.303926
```

```
kurtosis(ca)
```

```
## [1] 3.805731
```

```
shapiro.test(ca)
```

```
##
```

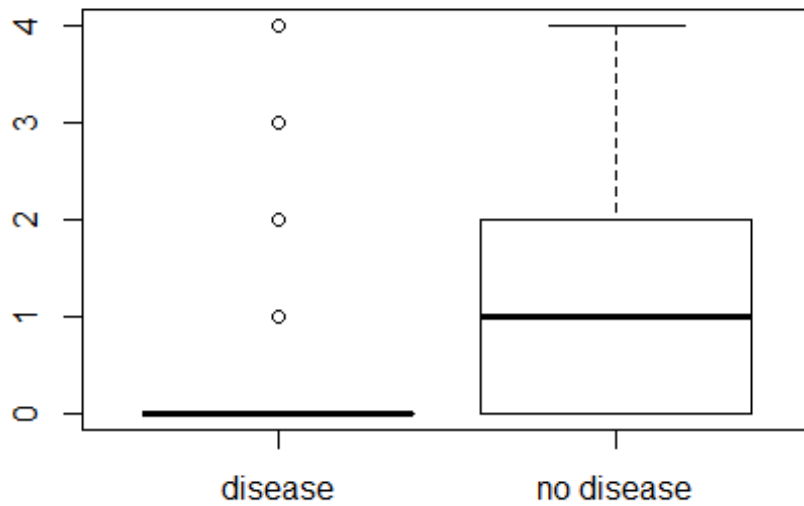
```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: ca
```

```
## W = 0.72812, p-value < 2.2e-16
```

```
boxplot(ca~target_dummy)
```



```
wilcox.test(ca~target_dummy, mu=0, alternative="two.sided", var.equal=F, data= newheart)
```

```
##
```

```
## Wilcoxon rank sum test with continuity correction
```

```
##
```

```
## data: ca by target_dummy
```

```
## W = 5999, p-value = 1.841e-15
```

```
## alternative hypothesis: true location shift is not equal to 0
```

The mean of ca(number of major vessels(0-3)colored by flourosopy) in patients with HDI is 0.36 while inn healthy patients is 1.17. The standard deviation in patients with HDI is 0.85 while in healthy patients is 1.04.

Check normality:

in numerical test: the data is highly skewed and out of acceptable range of kurtosis test.

in statistical test: the data is violated. p-value is less than 0.05. so the data is not normally distributed.

wilcoxon rank test: p-value is less than 0.05 so there is significant difference between presence of HDI and number of coloured major vessels(ca). we reject null hypothesis.

9-restecg:

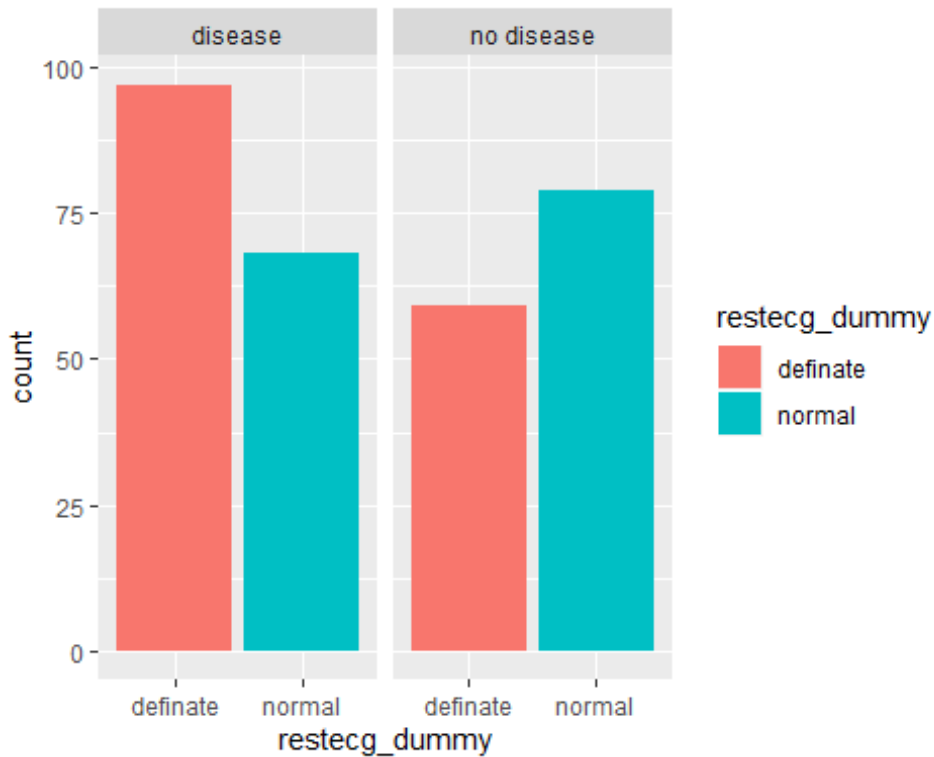
```
newheart$restecg <- as.character(newheart$restecg)
restecg_dummy <- ifelse(restecg=="0","normal","definate")
restecgcoun<- as.data.frame(newheart %>%
summary_factorlist("restecg","target_dummy",p=F, cont_cut = 1))

## Warning in chisq.test(tab, correct = FALSE): Chi-squared
approximation may
## be incorrect

restecgcoun

##           label      levels      0      1      2
## 1 target_dummy    disease 68 (41.2) 96 (58.2) 1 (0.6)
## 2                no disease 79 (57.2) 56 (40.6) 3 (2.2)

library(ggplot2)
ggplot(newheart, aes(restecg_dummy, fill= restecg_dummy))+geom_bar(stat
= "count")+facet_wrap(~target_dummy)+ theme_get()
```



```
chisq.test(table(target_dummy, restecg_dummy))

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(target_dummy, restecg_dummy)
## X-squared = 7.1064, df = 1, p-value = 0.007681

fisher.test(table(target_dummy,restecg_dummy))

##
## Fisher's Exact Test for Count Data
##
## data:  table(target_dummy, restecg_dummy)
## p-value = 0.00578
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.177711 3.099989
## sample estimates:
## odds ratio
## 1.90589
```

chisquare test is incorrect so we will apply fisher's exact test.

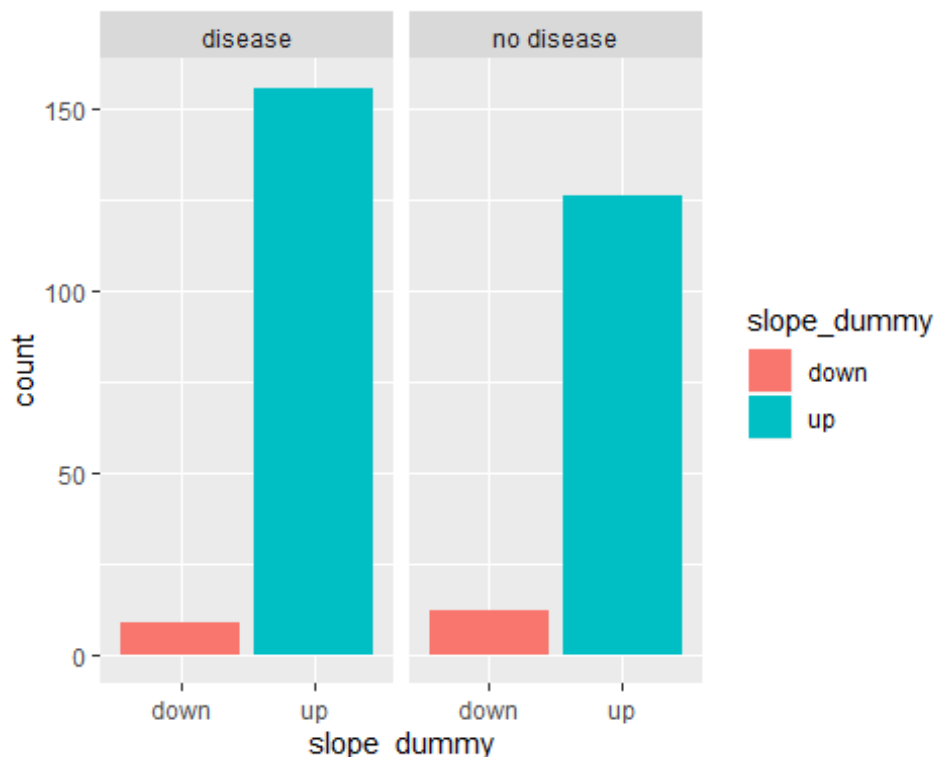
in fisher's exact test P-value is less than 0.05 so there is significant difference between presence of HDI and restecg(resting electrocardiograph results). we reject null hypothesis.

10- slope:

```
newheart$slope <- as.character(newheart$slope)
slope_dummy <- ifelse(slope=="0", "down", "up")
slopecoun <- table(target_dummy, slope_dummy)
slopecounper <- as.data.frame(newheart %>%
  summary_factorlist("slope",
    "target_dummy", p=F, add_dependent_label = T, cont_cut = 1))
slopecounper
```

Dependent: slope		0	1	2
1	target_dummy	disease 9 (5.5)	49 (29.7)	107 (64.8)
2		no disease 12 (8.7)	91 (65.9)	35 (25.4)

```
ggplot(newheart, aes(slope_dummy, fill= slope_dummy))+geom_bar(stat =
"count")+facet_wrap(~target_dummy)+ theme_get()
```




```

chisq.test(slopecoun)

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  slopecoun
## X-squared = 0.77294, df = 1, p-value = 0.3793

fisher.test(slopecoun)

##
## Fisher's Exact Test for Count Data
##
## data:  slopecoun
## p-value = 0.3641
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.2182915 1.6260889
## sample estimates:
## odds ratio
##  0.6067889

```

chisquare test is incorrect. so we will apply fisher's exact test. p-value in fisher's exact test is less than 0.05 so there is significant difference between presence of HDI and type of slope. we will reject null hypothesis.

11- thal:

```

newheart$thal <- as.character(newheart$thal)
thalcoun <- as.data.frame(newheart%>%
  summary_factorlist("thal", "target_dummy",
    p=F, add_dependent_label = T, cont_cut = 1))

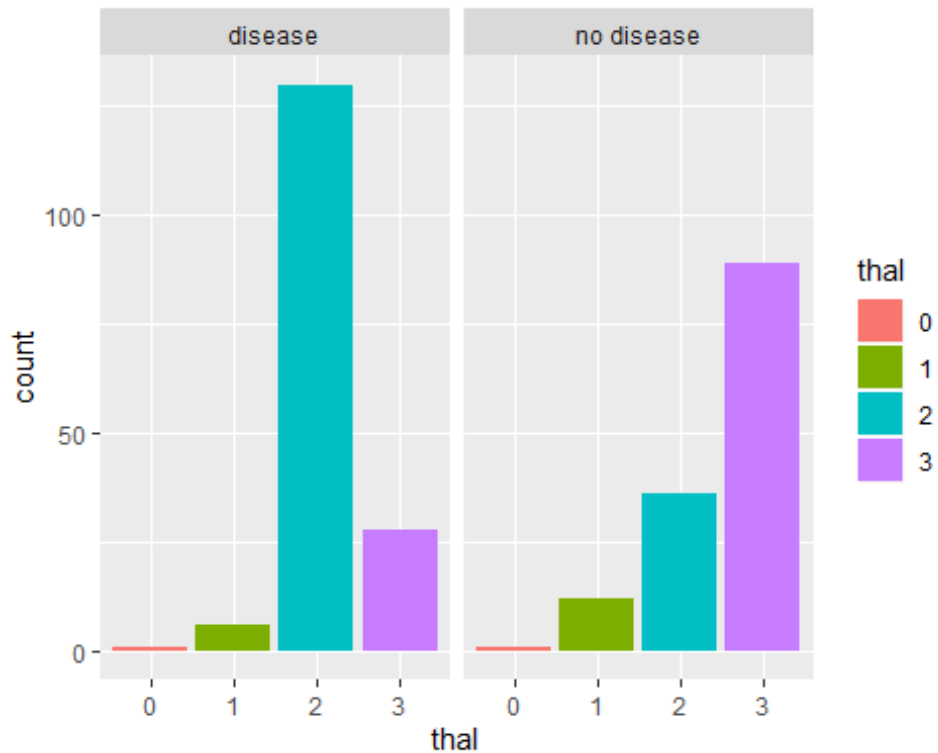
## Warning in chisq.test(tab, correct = FALSE): Chi-squared
## approximation may
## be incorrect

thalcoun

##   Dependent: thal           0           1           2           3
## 1 target_dummy disease 1 (0.6)  6 (3.6) 130 (78.8) 28 (17.0)
## 2                no disease 1 (0.7) 12 (8.7)  36 (26.1) 89 (64.5)

ggplot(newheart, aes(thal, fill=thal))+geom_bar(stat =
"count")+facet_wrap(~target_dummy)+theme_get()

```



```
chisq.test(table(target_dummy, thal))

## Warning in chisq.test(table(target_dummy, thal)): Chi-squared
## approximation
## may be incorrect

##
## Pearson's Chi-squared test
##
## data:  table(target_dummy, thal)
## X-squared = 85.304, df = 3, p-value < 2.2e-16

fisher.test(table(target_dummy, thal))

##
## Fisher's Exact Test for Count Data
##
## data:  table(target_dummy, thal)
## p-value < 2.2e-16
## alternative hypothesis: two.sided
```

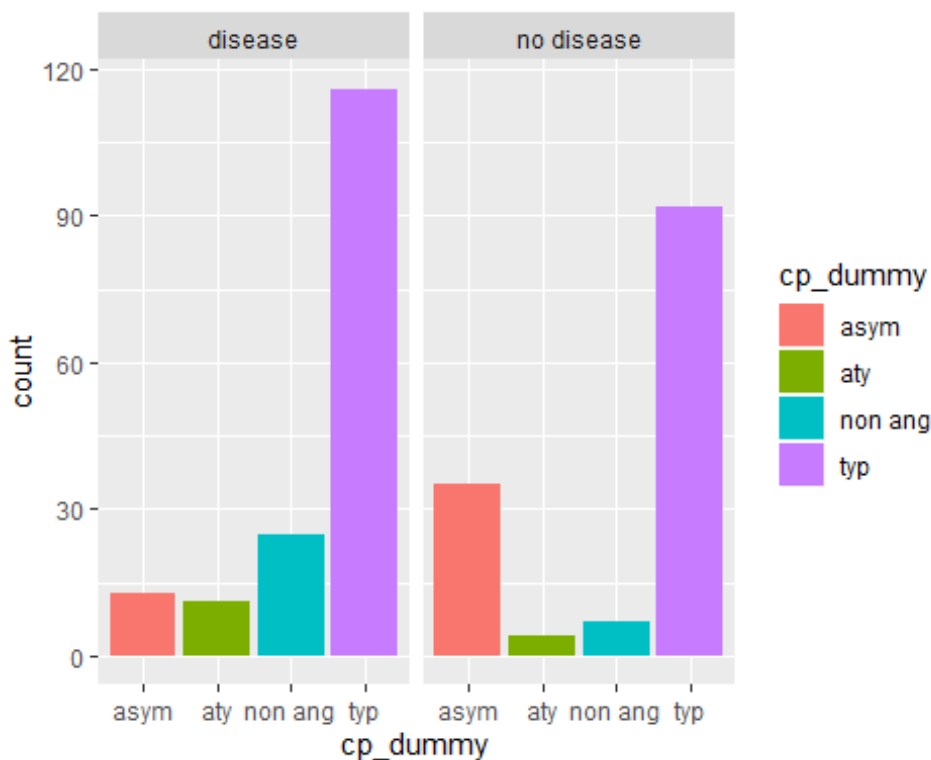
chisquare test is incorrect so we will apply fisher's exact test. p-value in fisher test is less than 0.05 so there is significant difference between presence of HDI and that types. we will reject null hypothesis.

12- cp:

```
newheart$cp <- as.character(newheart$cp)
cp_dummy <- ifelse(cp==c("0","1","2"),c("asym","aty","non ang"),"typ")
cpcoun <- as.data.frame(newheart %>%
  summary_factorlist("cp", "target_dummy", p=F,
    add_dependent_label = T, cont_cut = 1))
cpcoun
```

Dependent: cp		0		1		2		3	
## 1	target_dummy	disease	39 (23.6)	41 (24.8)	69 (41.8)	16 (9.7)			
## 2		no disease	104 (75.4)	9 (6.5)	18 (13.0)	7 (5.1)			

```
library(ggplot2)
ggplot(newheart, aes(cp_dummy, fill= cp_dummy))+geom_bar(stat =
  "count")+facet_wrap(~target_dummy)+theme_get()
```



```
chisq.test(table(target_dummy, cp))
```

```
##
## Pearson's Chi-squared test
##
## data: table(target_dummy, cp)
## X-squared = 81.686, df = 3, p-value < 2.2e-16
```

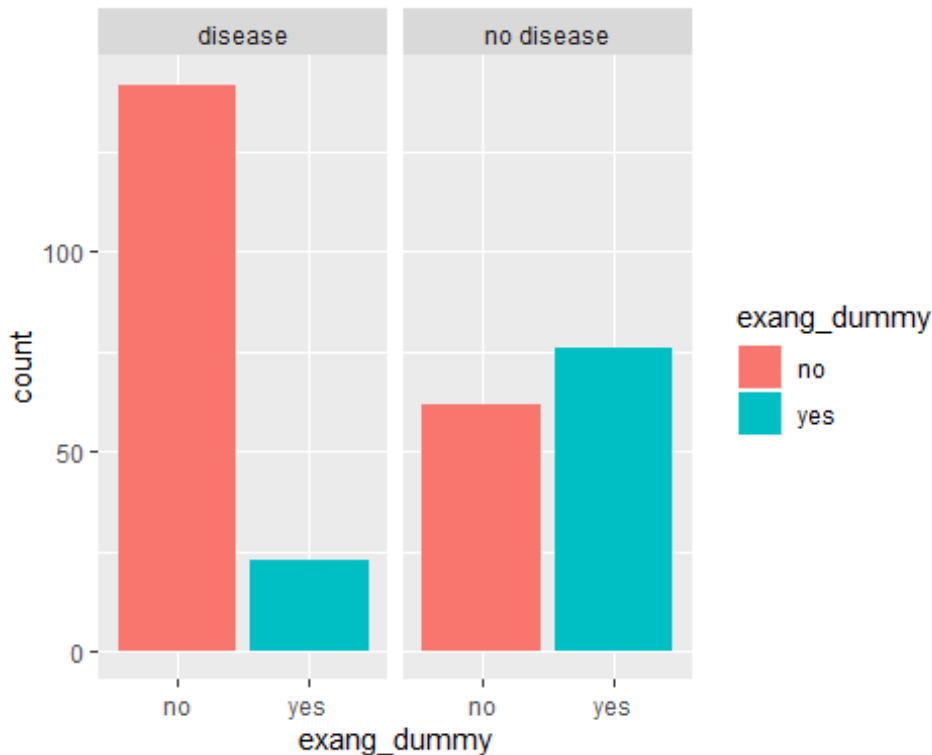
p-value in chisquare test is less than 0.05 so there is significant difference between presence of HDI and Cp types. we will reject null hypothesis.

12-exang:

```
newheart$exang <- as.character(newheart$exang)
exangcoun <- as.data.frame(newheart%>%
  summary_factorlist("exang_dummy",
    "target_dummy", p=F, add_dependent_label = T, cont_cut = 1))
colnames(exangcoun) <- c("dependent:exang", "", "no excercise",
  "excerise")
exangcoun
```

	dependent:exang		no excercise	excerise
## 1	target_dummy	disease	142 (86.1)	23 (13.9)
## 2		no disease	62 (44.9)	76 (55.1)

```
ggplot(newheart, aes(exang_dummy,
  fill=exang_dummy))+geom_bar(stat="count")+facet_wrap(~target_dummy)+
  theme_get()
```



```
chisq.test(table(target_dummy, exang_dummy))

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(target_dummy, exang_dummy)
## X-squared = 55.945, df = 1, p-value = 7.454e-14
```

p-value of chisquare test is less than 0.05 so there is significant difference between presence of HDI and exercise induced angina. we will reject null hypothesis.

```
newheart1 <- newheart[ , -c(3,7,11,13,15,20,24,31,32,33,34,35)]
newheart1$cpatyp <- as.numeric(newheart1$cpatyp)
newheart1$cpasym <- as.numeric(newheart1$cpasym)
newheart1$cpnon_ang <- as.numeric(newheart1$cpnon_ang)
newheart1$cptyp <- as.numeric(newheart1$cptyp)
newheart1$restecg_notnorm <- as.numeric(newheart1$restecg_notnorm)
newheart1$restecg_norm <- as.numeric(newheart1$restecg_norm)
newheart1$restecg_abnor <- as.numeric(newheart1$restecg_abnor)
newheart1$slope_down <- as.numeric(newheart1$slope_down)
newheart1$slope_flat <- as.numeric(newheart1$slope_flat)
newheart1$slope_up <- as.numeric(newheart1$slope_up)
```

```
newheart1$thal_nor <- as.numeric(newheart1$thal_nor)
newheart1$thal_zero <- as.numeric(newheart1$thal_zero)
newheart1$thal_fixed <- as.numeric(newheart1$thal_fixed)
newheart1$exang <- as.numeric(newheart1$exang)
```

```
View(newheart1)
```

logistic regression

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
corelations <- cor(newheart1)
corelations
```

```
##           ãˆ~..age           sex      trestbps           chol
## ãˆ~..age      1.00000000 -0.098446602  0.27935091  0.213677957
## sex          -0.09844660  1.000000000 -0.05676882 -0.197912174
## trestbps      0.27935091 -0.056768824  1.00000000  0.123174207
## chol          0.21367796 -0.197912174  0.12317421  1.000000000
## fbs           0.12130765  0.045031789  0.17753054  0.013293602
## thalach       -0.39852194 -0.044019908 -0.04669773 -0.009939839
## exang         0.09680083  0.141663811  0.06761612  0.067022783
## oldpeak       0.21001257  0.096092877  0.19321647  0.053951920
## ca            0.27632624  0.118261413  0.10138899  0.070510925
## target       -0.22543872 -0.280936576 -0.14493113 -0.085239105
## cpasym        0.13825269  0.089621434  0.02144716  0.070683691
## cpatyp        -0.14739877 -0.041248836 -0.08185063 -0.012749604
## cponon_ang    -0.05943205 -0.116588013 -0.04510713 -0.037918239
## cptyp         0.04751617  0.088043159  0.15134350 -0.050589086
## restecg_notnorm 0.14139844  0.036538679  0.13324945  0.166564516
## restecg_norm   -0.16065360 -0.011940132 -0.14657621 -0.174393593
## restecg_abnor  0.08462123 -0.107689392  0.05863509  0.034614914
## slope_down     0.03053643  0.046183063  0.12238362 -0.045339897
## slope_flat     0.17720114 -0.009157059  0.02520741  0.051176659
## slope_up      -0.19257988 -0.014356333 -0.08747167 -0.028053783
## thal_nor       0.06375812  0.141135216  0.07636870 -0.093560219
## thal_zero     -0.01677850 -0.032092529 -0.01687041 -0.057126565
## thal_fixed    -0.13589099 -0.376364934 -0.14071190 -0.004591380
##           fbs           thalach           exang           oldpeak
## ãˆ~..age      0.121307648 -0.398521938  0.09680083  0.210012567
## sex          0.045031789 -0.044019908  0.14166381  0.096092877
## trestbps      0.177530542 -0.046697728  0.06761612  0.193216472
## chol          0.013293602 -0.009939839  0.06702278  0.053951920
## fbs           1.000000000 -0.008567107  0.02566515  0.005747223
## thalach       -0.008567107  1.000000000 -0.37881209 -0.344186948
## exang         0.025665147 -0.378812094  1.00000000  0.288222808
## oldpeak       0.005747223 -0.344186948  0.28822281  1.000000000
## ca            0.137979327 -0.213176928  0.11573938  0.222682322
```

## target	-0.028045760	0.421740934	-0.43675708	-0.430696002
## cpasym	-0.060190075	-0.376560502	0.46906073	0.280243114
## cpatyp	-0.060649324	0.248317168	-0.23386291	-0.277512939
## cponon_ang	0.083678880	0.165475348	-0.27102836	-0.132443676
## cptyp	0.055511402	0.079081546	-0.09338433	0.086959155
## restecg_notnorm	0.077404511	-0.073952802	0.08405914	0.100140887
## restecg_norm	-0.066344065	0.101512064	-0.09377636	-0.138486464
## restecg_abnor	-0.048304858	-0.120870361	0.04272883	0.168172094
## slope_down	0.105284377	-0.056590111	0.05925270	0.394252942
## slope_flat	-0.033359989	-0.419626768	0.25768650	0.310986156
## slope_up	-0.020255395	0.448042809	-0.28760649	-0.511356313
## thal_nor	0.091351214	-0.159748161	0.06307303	0.104634556
## thal_zero	0.080568494	-0.050428545	0.03011330	-0.037945863
## thal_fixed	-0.086773812	0.294801040	-0.32853931	-0.339085606
##	ca	target	cpasym	cpatyp
## ãˆ·.age	0.27632624	-0.225438716	0.138252689	-0.14739877
## sex	0.11826141	-0.280936576	0.089621434	-0.04124884
## trestbps	0.10138899	-0.144931128	0.021447156	-0.08185063
## chol	0.07051093	-0.085239105	0.070683691	-0.01274960
## fbs	0.13797933	-0.028045760	-0.060190075	-0.06064932
## thalach	-0.21317693	0.421740934	-0.376560502	0.24831717
## exang	0.11573938	-0.436757083	0.469060731	-0.23386291
## oldpeak	0.22268232	-0.430696002	0.280243114	-0.27751294
## ca	1.00000000	-0.391723992	0.218226250	-0.13471514
## target	-0.39172399	1.000000000	-0.516014642	0.24587910
## cpasym	0.21822625	-0.516014642	1.000000000	-0.42027424
## cpatyp	-0.13471514	0.245879095	-0.420274238	1.000000000
## cponon_ang	-0.08899838	0.316742158	-0.599985532	-0.28213534
## cptyp	-0.07049557	0.086956875	-0.270952355	-0.12741180
## restecg_notnorm	0.08267863	-0.159775193	0.114074988	-0.09352970
## restecg_norm	-0.08964150	0.175321804	-0.128731176	0.10522641
## restecg_abnor	0.03066020	-0.068410240	0.064422311	-0.05141847
## slope_down	-0.04221809	-0.063553672	0.028349913	-0.05129836
## slope_flat	0.14192321	-0.362053300	0.237724494	-0.19799440
## slope_up	-0.12030579	0.394066369	-0.251934784	0.22392087
## thal_nor	0.06661843	-0.106588974	0.098025220	-0.03649584
## thal_zero	-0.05823589	-0.007292627	0.004580603	-0.03623736
## thal_fixed	-0.22133111	0.527333550	-0.349920500	0.20735417
##	cponon_ang	cptyp	restecg_notnorm	
restecg_norm				
## ãˆ·.age	-0.059432046	0.047516172	0.141398441	-
0.1606535989				
## sex	-0.116588013	0.088043159	0.036538679	-
0.0119401319				
## trestbps	-0.045107126	0.151343499	0.133249452	-
0.1465762142				
## chol	-0.037918239	-0.050589086	0.166564516	-
0.1743935935				
## fbs	0.083678880	0.055511402	0.077404511	-
0.0663440652				

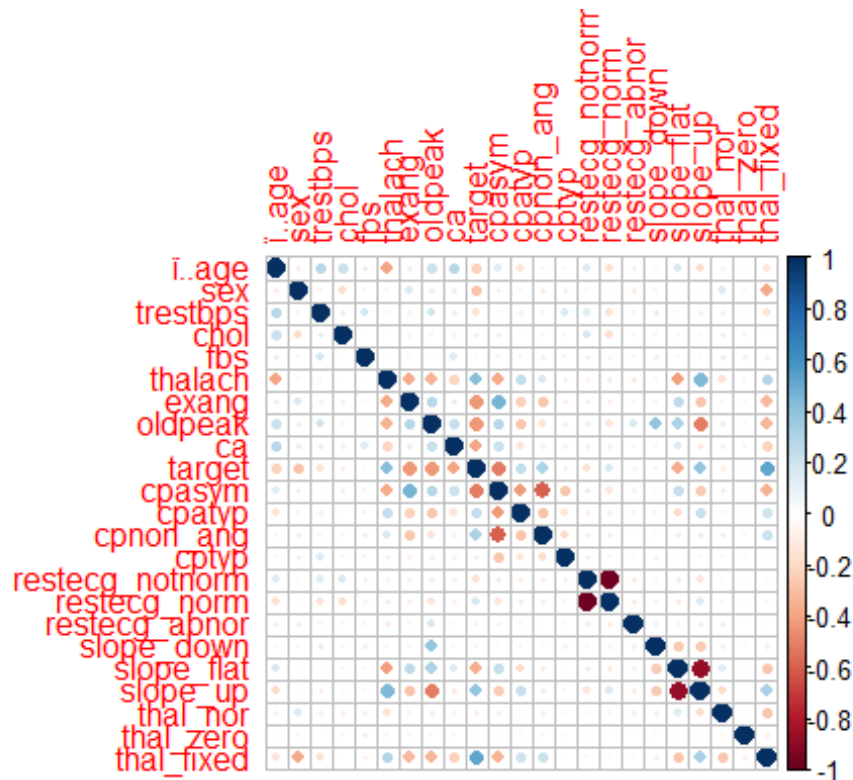
## thalach	0.165475348	0.079081546	-0.073952802	
0.1015120636				
## exang	-0.271028362	-0.093384327	0.084059136	-
0.0937763598				
## oldpeak	-0.132443676	0.086959155	0.100140887	-
0.1384864643				
## ca	-0.088998383	-0.070495572	0.082678626	-
0.0896415044				
## target	0.316742158	0.086956875	-0.159775193	
0.1753218037				
## cpasym	-0.599985532	-0.270952355	0.114074988	-
0.1287311765				
## cpatyp	-0.282135342	-0.127411798	-0.093529701	
0.1052264113				
## cpono_ang	1.000000000	-0.181893698	-0.090611084	
0.0927383817				
## cptyp	-0.181893698	1.000000000	0.070849902	-
0.0632518363				
## restecg_notnorm	-0.090611084	0.070849902	1.000000000	-
0.9739343615				
## restecg_norm	0.092738382	-0.063251836	-0.973934362	
1.0000000000				
## restecg_abnor	-0.009492059	-0.033149677	-0.112276992	-
0.1160453444				
## slope_down	-0.029575622	0.068981473	0.047110442	-
0.0658744119				
## slope_flat	-0.105320120	0.009321326	0.107012036	-
0.1222149625				
## slope_up	0.120275914	-0.044421040	-0.130890663	
0.1556295839				
## thal_nor	-0.097774600	0.033404087	0.035403938	-
0.0566769101				
## thal_zero	0.038353497	-0.023362357	0.002422274	-
0.0002690243				
## thal_fixed	0.210130448	0.009998345	-0.033629927	
0.0361539249				
##	restecg_abnor	slope_down	slope_flat	slope_up
## A`..age	0.084621232	0.03053643	0.177201137	-0.192579878
## sex	-0.107689392	0.04618306	-0.009157059	-0.014356333
## trestbps	0.058635085	0.12238362	0.025207406	-0.087471668
## chol	0.034614914	-0.04533990	0.051176659	-0.028053783
## fbs	-0.048304858	0.10528438	-0.033359989	-0.020255395
## thalach	-0.120870361	-0.05659011	-0.419626768	0.448042809
## exang	0.042728832	0.05925270	0.257686502	-0.287606491
## oldpeak	0.168172094	0.39425294	0.310986156	-0.511356313
## ca	0.030660200	-0.04221809	0.141923208	-0.120305790
## target	-0.068410240	-0.06355367	-0.362053300	0.394066369
## cpasym	0.064422311	0.02834991	0.237724494	-0.251934784
## cpatyp	-0.051418469	-0.05129836	-0.197994399	0.223920867
## cpono_ang	-0.009492059	-0.02957562	-0.105320120	0.120275914


```

## ctyp          -0.033149677  0.06898147  0.009321326 -0.044421040
## restecg_notnorm -0.112276992  0.04711044  0.107012036 -0.130890663
## restecg_norm   -0.116045344 -0.06587441 -0.122214962  0.155629584
## restecg_abnor   1.000000000  0.08228948  0.066803931 -0.108623961
## slope_down     0.082289481  1.000000000 -0.252903773 -0.256280951
## slope_flat     0.066803931 -0.25290377  1.000000000 -0.870365063
## slope_up       -0.108623961 -0.25628095 -0.870365063  1.000000000
## thal_nor       0.093258346  0.09633889  0.131149275 -0.180060652
## thal_zero      -0.009428143 -0.02224420  0.006205428  0.005121513
## thal_fixed     -0.011121108 -0.09151157 -0.275317554  0.321639566
##               thal_nor   thal_zero   thal_fixed
## Ñ..age         0.06375812 -0.0167785044 -0.135890992
## sex            0.14113522 -0.0320925291 -0.376364934
## trestbps       0.07636870 -0.0168704125 -0.140711896
## chol          -0.09356022 -0.0571265654 -0.004591380
## fbs           0.09135121  0.0805684944 -0.086773812
## thalach       -0.15974816 -0.0504285454  0.294801040
## exang         0.06307303  0.0301133015 -0.328539306
## oldpeak       0.10463456 -0.0379458634 -0.339085606
## ca            0.06661843 -0.0582358939 -0.221331107
## target       -0.10658897 -0.0072926270  0.527333550
## cpasym        0.09802522  0.0045806031 -0.349920500
## cpatyp       -0.03649584 -0.0362373551  0.207354172
## cpnon_ang     -0.09777460  0.0383534965  0.210130448
## ctyp          0.03340409 -0.0233623568  0.009998345
## restecg_notnorm 0.03540394  0.0024222742 -0.033629927
## restecg_norm   -0.05667691 -0.0002690243  0.036153925
## restecg_abnor   0.09325835 -0.0094281428 -0.011121108
## slope_down     0.09633889 -0.0222442027 -0.091511566
## slope_flat     0.13114927  0.0062054284 -0.275317554
## slope_up       -0.18006065  0.0051215126  0.321639566
## thal_nor       1.000000000 -0.0204854530 -0.276635291
## thal_zero      -0.02048545  1.0000000000 -0.089727488
## thal_fixed     -0.27663529 -0.0897274882  1.000000000

```

```
corrplot(corelations, method = "circle")
```



```
newheart1$target <- as.factor(newheart1$target)
logisticmodel <- glm(target~., family = binomial, data = newheart1)
summary(logisticmodel)
```

```
##
## Call:
## glm(formula = target ~ ., family = binomial, data = newheart1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7777  -0.3544   0.1525   0.5302   2.6007
##
## Coefficients: (3 not defined because of singularities)
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    4.3847107   8.4336529   0.520 0.603129
## i.age         -0.0005734   0.0235952  -0.024 0.980614
## sex           -1.5149396   0.5212317  -2.906 0.003655 **
## trestbps      -0.0170729   0.0107004  -1.596 0.110589
## chol          -0.0043317   0.0038894  -1.114 0.265394
## fbs           0.1764007   0.5661856   0.312 0.755375
## thalach       0.0171314   0.0107357   1.596 0.110545
## exang        -0.7630837   0.4260285  -1.791 0.073268 .
## oldpeak      -0.4892926   0.2258040  -2.167 0.030243 *
## ca           -0.8331781   0.2043120  -4.078 4.54e-05 ***
## cpasym       -2.0159122   0.6506319  -3.098 0.001946 **
## cpatyp       -1.0326821   0.7621467  -1.355 0.175429
## cpnon_ang    -0.0706805   0.6475534  -0.109 0.913083
```

```
## ctyp                NA                NA                NA                NA
## restecg_notnorm    0.2767289    2.2672126    0.122 0.902854
## restecg_norm       0.8469354    2.2656601    0.374 0.708543
## restecg_abnor      NA                NA                NA                NA
## slope_down         -0.2015612    0.9382445   -0.215 0.829901
## slope_flat         -0.9212254    0.4568067   -2.017 0.043731 *
## slope_up           NA                NA                NA                NA
## thal_nor           1.3414378    0.7576463    1.771 0.076638 .
## thal_zero          -0.4732491    2.3013525   -0.206 0.837072
## thal_fixed         1.3800697    0.4101153    3.365 0.000765 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 417.64  on 302  degrees of freedom
## Residual deviance: 201.69  on 283  degrees of freedom
## AIC: 241.69
##
## Number of Fisher Scoring iterations: 6
```

In this data we found that there are significant difference between the presence of heart disease or not and some variances which are sex, cpasym, oldpeak , ca , slope-flate and thalfixed.

so we will repeat our model with these variances to detect the interpretation.

```
logisticmode2 <- glm(target~slope_up+restecg_abnor+cptyp, family =
binomial, data = newheart1)
summary(logisticmode2)

##
## Call:
## glm(formula = target ~ slope_up + restecg_abnor + cptyp, family =
binomial,
##   data = newheart1)
##
## Deviance Residuals:
##    Min       1Q   Median       3Q      Max
## -2.0734  -0.9155   0.7678   0.7678   1.6651
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.8860     1.4205  -2.032   0.0422 *
```

```

## slope_up      1.7237      0.2591      6.652 2.9e-11 ***
## restecg_abnor -0.4456      1.1678     -0.382 0.7028
## ctyp          0.9549      0.5008      1.907 0.0566 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 417.64  on 302  degrees of freedom
## Residual deviance: 364.94  on 299  degrees of freedom
## AIC: 372.94
##
## Number of Fisher Scoring iterations: 4

logisticmode3 <-
glm(target~sex+oldpeak+ca+cpasym+slope_flat+thal_fixed+cptyp, family =
binomial, data = newheart1)
summary(logisticmode3)

##
## Call:
## glm(formula = target ~ sex + oldpeak + ca + cpasym + slope_flat +
##      thal_fixed + ctyp, family = binomial, data = newheart1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5536  -0.4854   0.2158   0.4905   2.3367
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   4.0757     1.4229   2.864 0.004178 **
## sex           -1.0397     0.4407  -2.359 0.018314 *
## oldpeak       -0.5527     0.1747  -3.164 0.001559 **
## ca            -0.7504     0.1805  -4.158 3.21e-05 ***
## cpasym        -1.9073     0.3679  -5.184 2.18e-07 ***
## slope_flat    -1.0611     0.3670  -2.891 0.003839 **
## thal_fixed     1.3953     0.3653   3.820 0.000133 ***
## ctyp           0.1815     0.5957   0.305 0.760540
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 417.64  on 302  degrees of freedom
## Residual deviance: 221.99  on 295  degrees of freedom
## AIC: 237.99
##
## Number of Fisher Scoring iterations: 5

```

```

logisticmode4 <- glm(target~sex+oldpeak+ca+thal_fixed+cpasym, family =
binomial, data = newheart1)
summary(logisticmode4)

##
## Call:
## glm(formula = target ~ sex + oldpeak + ca + thal_fixed + cpasym,
##      family = binomial, data = newheart1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4293  -0.4715   0.2291   0.5799   2.4464
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    2.4555     0.9413   2.609  0.00909 **
## sex            -0.7802     0.4116  -1.896  0.05799 .
## oldpeak        -0.6991     0.1721  -4.061 4.88e-05 ***
## ca             -0.7304     0.1719  -4.249 2.15e-05 ***
## thal_fixed      1.5664     0.3538   4.427 9.57e-06 ***
## cpasym         -1.9607     0.3388  -5.787 7.17e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 417.64  on 302  degrees of freedom
## Residual deviance: 230.51  on 297  degrees of freedom
## AIC: 242.51
##
## Number of Fisher Scoring iterations: 5

```

the interpretation

```

library(Greg)

## Loading required package: forestplot
## Loading required package: grid
## Loading required package: magrittr
## Loading required package: checkmate
## Loading required package: Gmisc
## Loading required package: Rcpp
## Loading required package: htmlTable

printCrudeAndAdjustedModel(logisticmode4)[-1,]

```

Crude

Adjusted

Variable

OR

2.5 % to 97.5 %

OR

2.5 % to 97.5 %

sex

0.27

0.16 to 0.46

0.46

0.20 to 1.02

oldpeak

0.39

0.29 to 0.51

0.50

0.35 to 0.69

ca

0.40

0.30 to 0.53

0.48

0.34 to 0.67

thal_fixed

10.52

6.25 to 18.17

4.79

2.41 to 9.72

cpasym

0.10

0.06 to 0.17

0.14

0.07 to 0.27

Crude odd ratio(OR):

First interpretation for crude value which means the value of OR(odd ratio) in the presence of other variables and about 95% of true odds is lying in between the mentioned upper and lower limits:

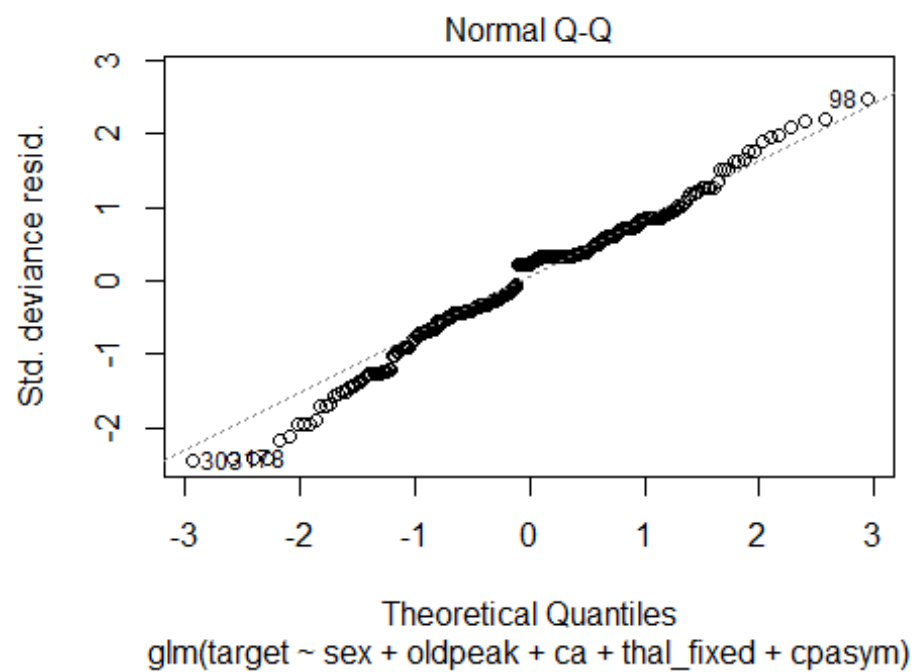
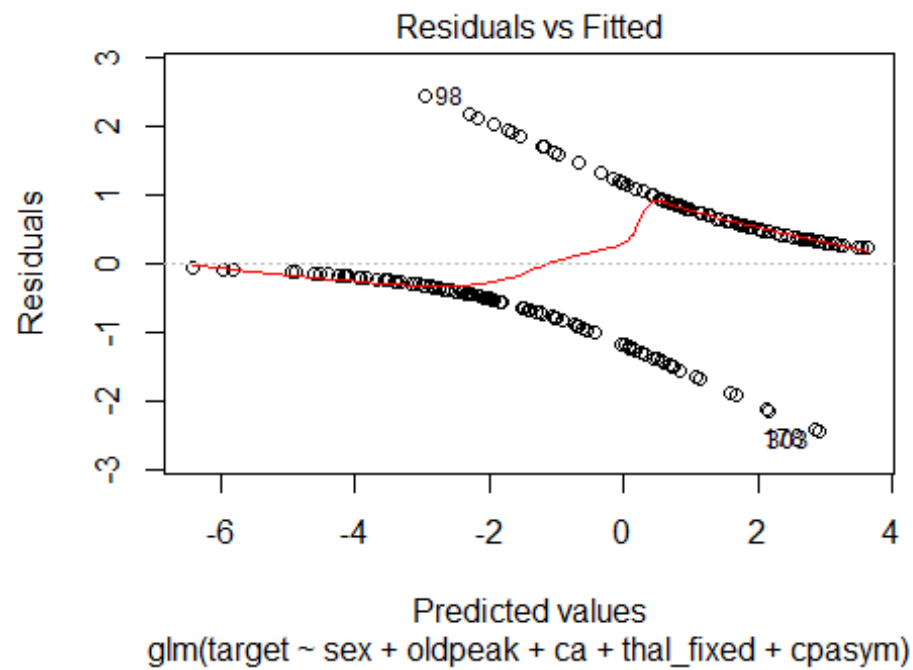
- 1- the odds of presence heart disease in male patients taking the consideration oldpeak, ca, cpasym and thal-fixed is 0.27 which means the probability of presence of UDI in male patient is 21.27% then in female is 78.73% so female is more susceptible for HDI.
- 2- the odds of presence UDI oldpeak taking the consideration the rest of significant variances is 0.39 which means the probability of presence of UDI in oldpeak is 28.05%
- 3- the odds of presence UDI in ca taking the consideration the rest of other significant variances is 0.4 which means the probability of presence of UDI in thalab is 28.58%
- 4- the odds of presence UDI in thal_fixed taking the consideration the rest of other significant variances is 10.52 which means that its probability is 91.31%
- 5- the odds of presence of UDI in cp_asym taking the consideration the rest of significant variances is 0.10 which means that its probability is 9.09%

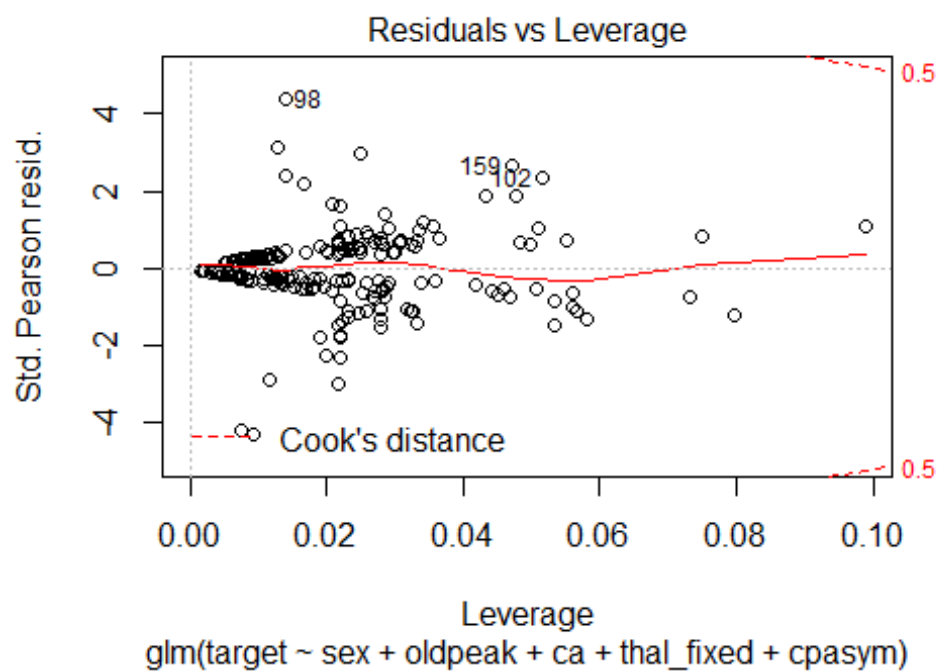
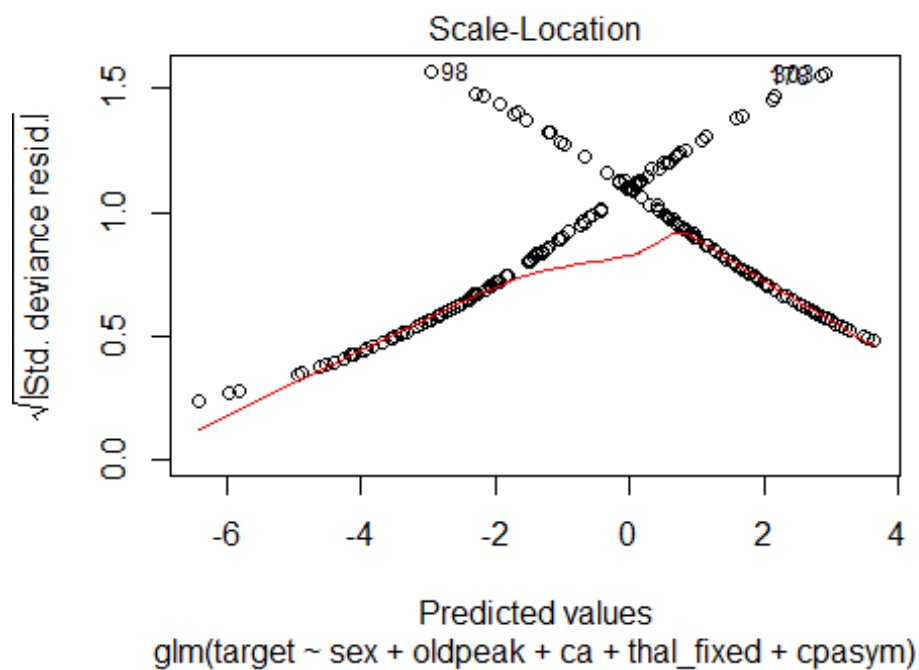
second interpretation is adjusted value which means the value of OR(odd ratio) only without any effect of other significant variables and 95% of true odds is lying between the mentioned upper and lower limits:

- 1- the odds of presence of HDI in male patients is 0.46 and its probability is 31.5% in male and thus in female is 68.5% so female is more susceptible for HDI
- 2-the odds of presence of HDI in oldpeak is 0.50 and its probability is 33.33%
- 3-the odds of presence of HDI in ca is 0.48 and its probability is 32.43%
- 4- the odds of presence of HDI in thal_fixed is 4.79 and its probability is 82.72%
- 5- the odds of presence of HDI in cp_asym is 0.14 and its probability is 12.28%

check residuals

```
plot(logisticmode4)
```





In normal Q-Q curve the residuals is in street line which means that it is acceptable residuals.