

```
library(tidyverse)

Registered S3 methods overwritten by 'ggplot2':
  method      from
read.xml.response xml2

--- Attaching packages --- tidyverse 1.2.1 ---
✔ ggplot2 3.1.1    ✔ purrr 0.3.2
✔ tidbale 2.1.1    ✔ dplyr 0.8.0.1
✔ tidyr 0.8.3      ✔ stringr 1.4.0
✔ readr 1.3.1      ✔ forcats 0.4.0

--- Conflicts --- tidyverse_conflicts() ---
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag() masks stats::lag()
```

```
In [3]: data <- read.csv("heart.csv")
head(data)
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1	
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1	
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1	
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1	
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1	
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1	

```
In [4]: tail(data)
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
298	59	1	0	164	176	1	0	90	0	1.0	1	2	1	0
299	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
300	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
301	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
302	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
303	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

```
In [5]: glimpse(data)
```

Observations: 303
Variables: 14
\$ age <int> 63, 37, 41, 56, 57, 57, 56, 44, 52, 57, 54, 48, 49, 64, 58, ...
\$ sex <int> 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, ...
\$ cp <int> 3, 2, 1, 1, 1, 0, 0, 1, 1, 2, 2, 0, 2, 1, 3, 3, 2, 2, 3, 0, 3, ...
\$ trestbps <int> 145, 130, 130, 120, 120, 140, 140, 120, 172, 150, 140, 130, ...
\$ chol <int> 233, 250, 204, 236, 354, 192, 294, 263, 199, 168, 239, 275, ...
\$ fbs <int> 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, ...
\$ restecg <int> 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, ...
\$ thalach <int> 150, 187, 172, 178, 163, 148, 153, 173, 162, 174, 160, 139, ...
\$ exang <dbl> 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, 0.2, ...
\$ slope <int> 0, 0, 2, 2, 1, 1, 2, 2, 2, 2, 2, 1, 2, 1, 2, 0, 2, 2, ...
\$ ca <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, ...
\$ thal <int> 1, 2, 2, 2, 2, 1, 2, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ...
\$ target <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...

```
In [6]: ncol(data)
```

14

```
In [8]: nrow(data)
```

303

```
In [9]: colnames(data)
```

- 1. 'age'
- 2. 'sex'
- 3. 'cp'
- 4. 'trestbps'
- 5. 'chol'
- 6. 'fbs'
- 7. 'restecg'
- 8. 'thalach'
- 9. 'exang'
- 10. 'oldpeak'
- 11. 'slope'
- 12. 'ca'
- 13. 'thal'
- 14. 'target'

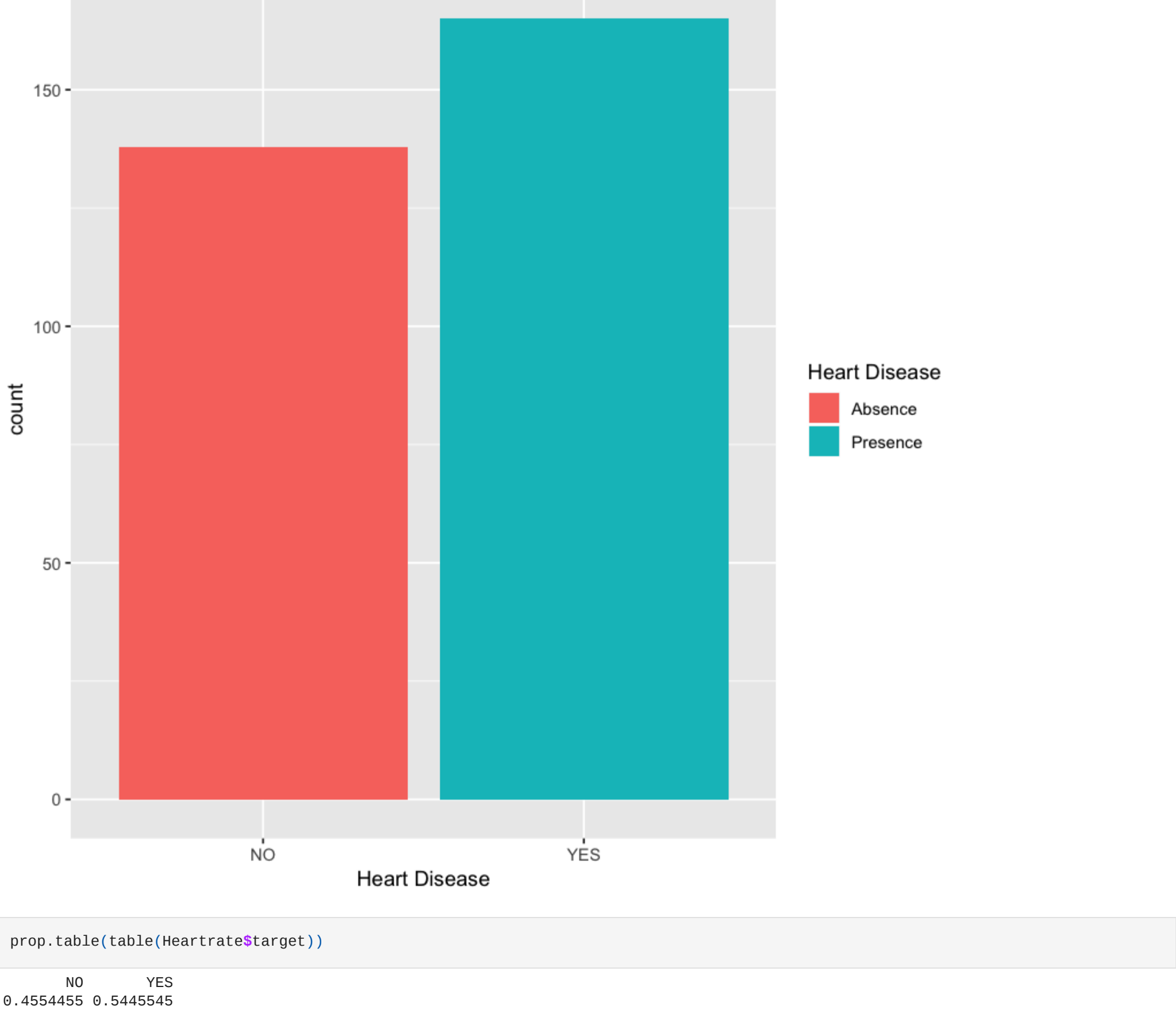
```
In [12]: summary(data)
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
Min.	:29.00	Min.	:0.0000	Min.	:0.000	Min.	:0.000	Min.	:94.0					
1st Qu.	:47.50	1st Qu.	:0.0000	1st Qu.	:0.000	1st Qu.	:0.000	1st Qu.	:120.0					
Median	:55.00	Median	:1.0000	Median	:1.000	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	exang	oldpeak	slope	ca	thal	target				
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	thal	target								
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							

```
In [32]: #Data Transformation
Heartrate <- data %>%
  mutate(sex = if_else(sex == 1, "MALE", "FEMALE"),
         fbs = if_else (fbs ==1, ">120", "<=120"),
         exang = if_else (exang ==1, "YES", "NO"),
         cp = if_else (cp == 1, "ATYPICAL ANGINA",
                      if_else(cp == 2, "NON-ANGINAL PAIN", "ASYMPTOMATIC")),
         restecg = if_else(restecg == 0, "NORMAL",
                           if_else(restecg == 1, "ABNORMALITY", "PROBABLE OR DEFINITE")),
         slope = as.factor(slope),
         ca = as.factor(ca),
         thal = as.factor(thal),
         target = if_else(target ==1, "YES", "NO")) %>%
  mutate_if(is.character, as.factor)%>%
  dplyr::select(target, sex, fbs, exang, cp, restecg, slope, ca, thal, everything())
```

```
In [23]: #Data Visualization
# ---Bar Plot for target (heart disease)

ggplot(Heartrate, aes(x=Heartrate$target, fill=Heartrate$target))+
  geom_bar()+
  xlab("Heart Disease")+
  ylab("count")+
  ggtitle("Prsence & Absence of Heart Disease")+
  scale_fill_discrete(name= 'Heart Disease', labels =c("Absence", "Presence"))
```

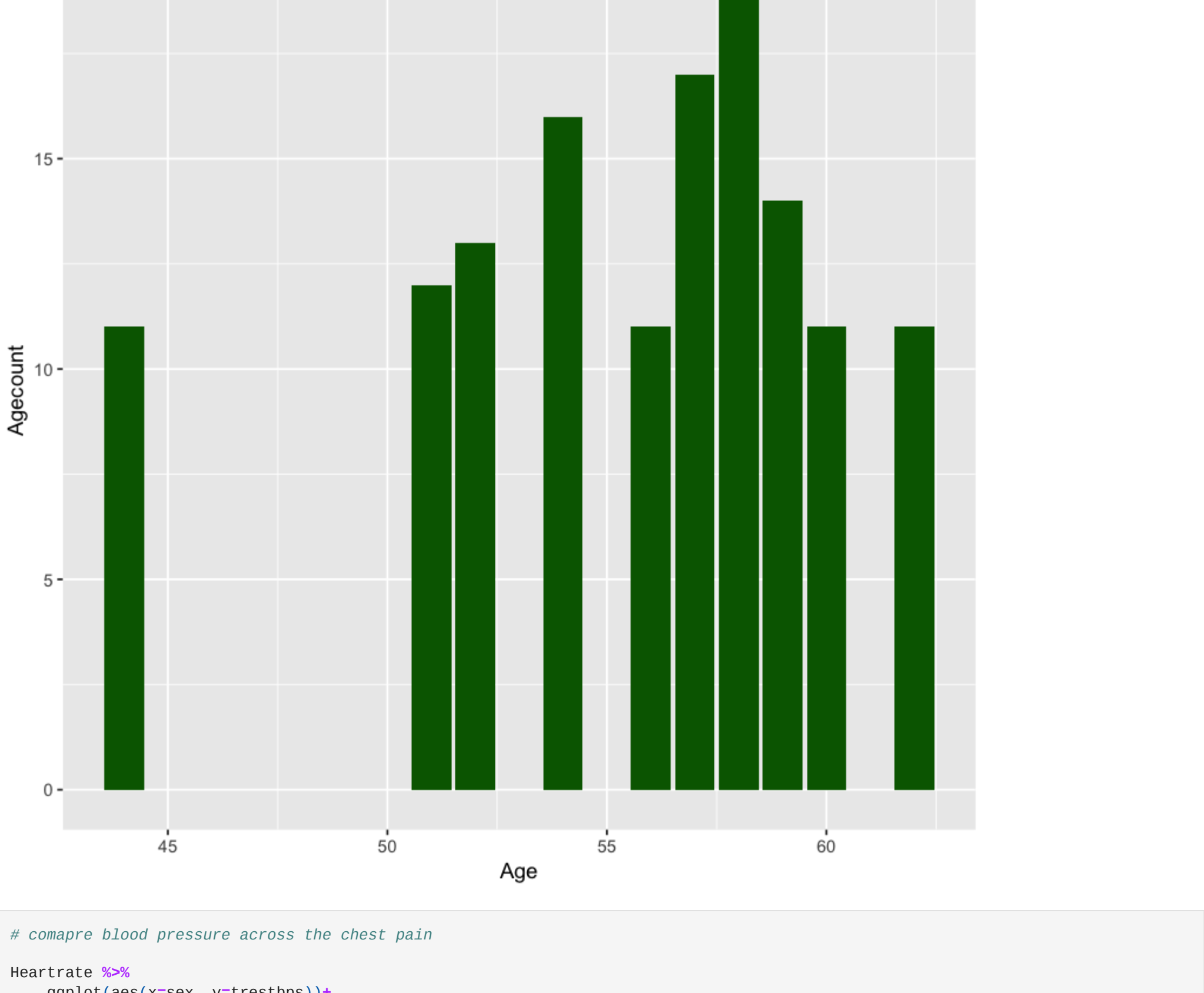


```
In [24]: prop.table(table(Heartrate$target))
```

	NO	YES
	0.4554455	0.5445545

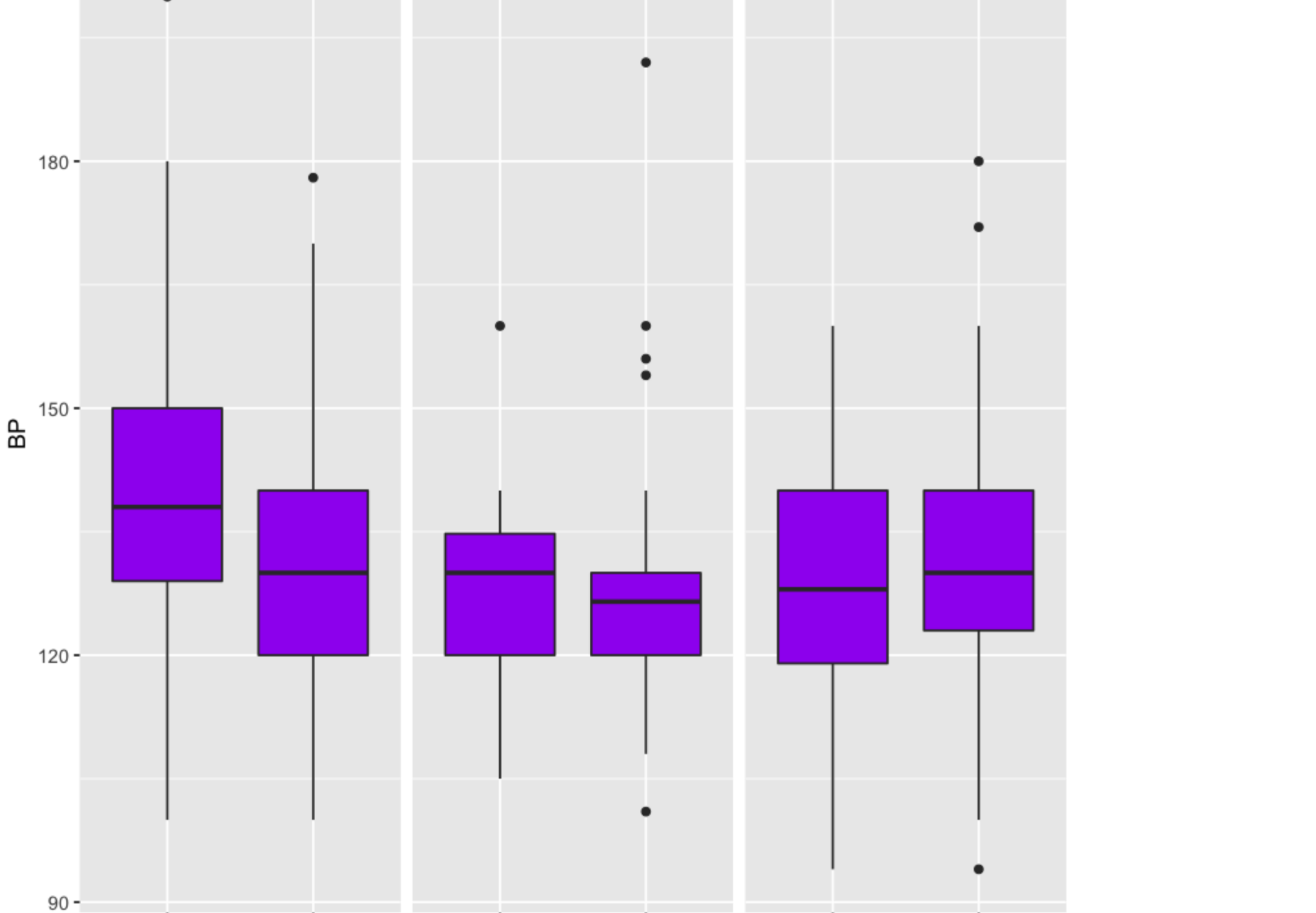
```
In [28]: # count the frequency of the values of age

Heartrate %>%
  group_by(age) %>%
  count() %>%
  filter(n>10) %>%
  ggplot() +
  geom_col(aes(age,n),fill = 'darkgreen') +
  ggtitle("Age Analysis") +
  xlab("Age")+
  ylab("Agecount")
```

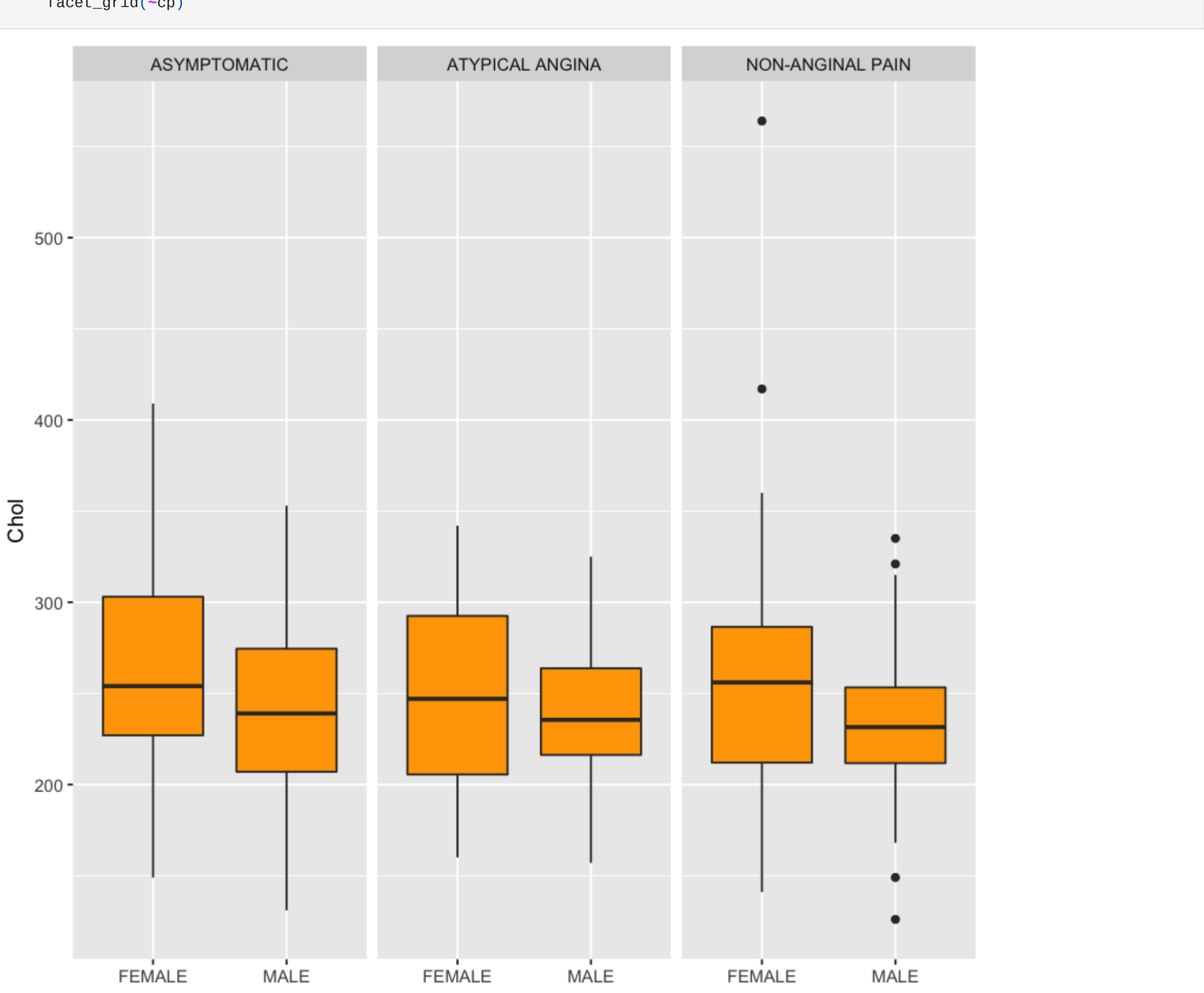


```
In [31]: # comapre blood pressure across the chest pain

Heartrate %>%
  ggplot(aes(x=sex, y=trestbps))+
  geom_boxplot(fill='purple')+
  xlab('sex')+
  ylab('BP')+
  facet_grid(~cp)
```



```
In [33]: Heartrate %>%
  ggplot(aes(x=sex, y=chol))+
  geom_boxplot(fill='orange')+
  xlab('sex')+
  ylab('chol')+
  facet_grid(~cp)
```



```
In [34]: #Correlation

install.packages("corrplot")
install.packages("ggplot2")
```

Updating HTML index of packages in '.Library'
Making 'packages.html' ... done
also installing the dependencies 'isoband', 'rlang'

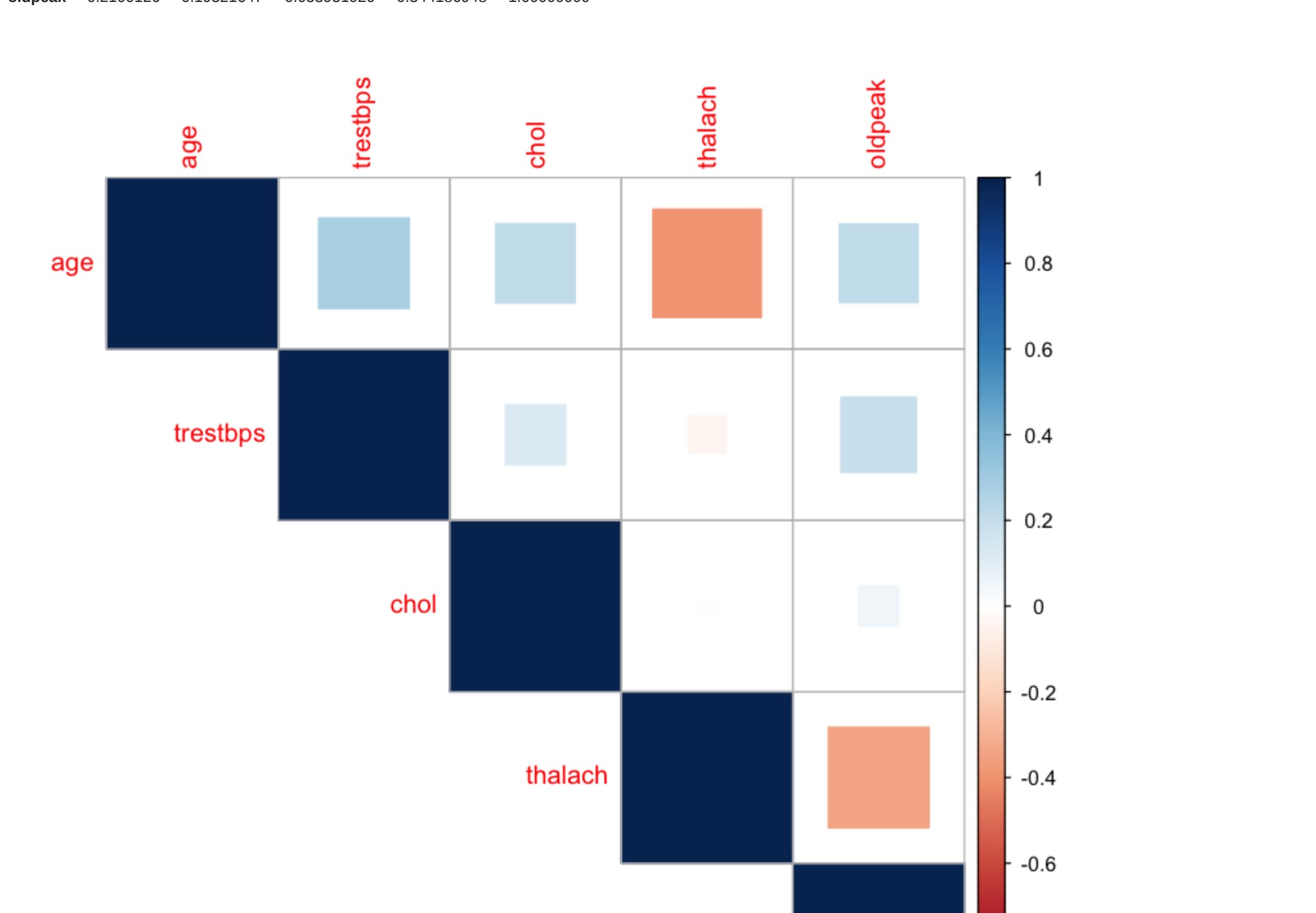
Warning message in install.packages("ggplot2"):
"installation of package 'isoband' had non-zero exit status"Warning message in install.packages("ggplot2"):
"installation of package 'rlang' had non-zero exit status"Warning message in install.packages("ggplot2"):
"installation of package 'ggplot2' had non-zero exit status"Updating HTML index of packages in '.Library'
Making 'packages.html' ... done

```
In [35]: library(corrplot)
library(ggplot2)
```

corrplot 0.92 loaded

```
In [36]: cor_heart <- cor(Heartrate[,10:14])
cor_heart

corrplot(cor_heart,method='square', type='upper')
```



```
In [39]: ??method
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```