About Dataset

Context: The leading cause of death in the developed world is heart disease. Therefore there needs to be work done to help prevent the risks of of having a heart attack or stroke.

Content: Use this dataset to predict which patients are most likely to suffer from a heart disease in the near future using the features given.

ge : Age of the patient

Sex : Sex of the patient

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

ca: number of major vessels (0-3)

cp : Chest Pain type chest pain type

```
Value 1: typical angina
Value 2: atypical angina
Value 3: non-anginal pain
Value 4: asymptomatic
```

trtbps : resting blood pressure (in mm Hg)

chol : cholestoral in mg/dl fetched via BMI sensor

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

rest_ecg : resting electrocardiographic results

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

thalach : maximum heart rate achieved

target : 0= less chance of heart attack 1= more chance of heart attack

In [80]:
```
library(dplyr)
library(gmodels)
library(descr)
library(ggplot2)
```

In [2]:
```
heart=read.csv(file.choose(),header = T,stringsAsFactors = T)
```

In [3]:
```
head(heart)
```

ï..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 [81]:
```
glimpse(heart)
```

```
Rows: 303
Columns: 14
$ ï..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~
$ trestbps <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~
$ thalach  <int> 150, 187, 172, 178, 163, 148, 153, 173, 162, 174, 160, 139, 1~
$ exang    <int> 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~
$ slope    <int> 0, 0, 2, 2, 2, 1, 1, 2, 2, 2, 2, 2, 2, 1, 2, 1, 2, 0, 2, 2, 1~
$ ca       <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0~
$ thal     <int> 1, 2, 2, 2, 2, 1, 2, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3~
$ target   <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
```

In [82]:
```
dim(heart)
```

1. 303
2. 14

In [83]:
```
heart_data=heart %>%
  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, "Present", "Absent")
         )
```

In [84]:
```
# Covrting the variables into factor
heart_data$sex = as.factor(heart_data$sex)
heart_data$cp = as.factor(heart_data$cp)
heart_data$fbs = as.factor(heart_data$fbs)
heart_data$restecg = as.factor(heart_data$restecg)
heart_data$exang = as.factor(heart_data$exang)
heart_data$slope = as.factor(heart_data$slope)
heart_data$cp=as.factor(heart_data$cp)
heart_data$thal = as.factor(heart_data$thal)
heart_data$target = as.factor(heart_data$target)
```

In [85]:
```
dim(heart_data)
```

1. 303
2. 14

In [86]:
```
str(heart_data)
```

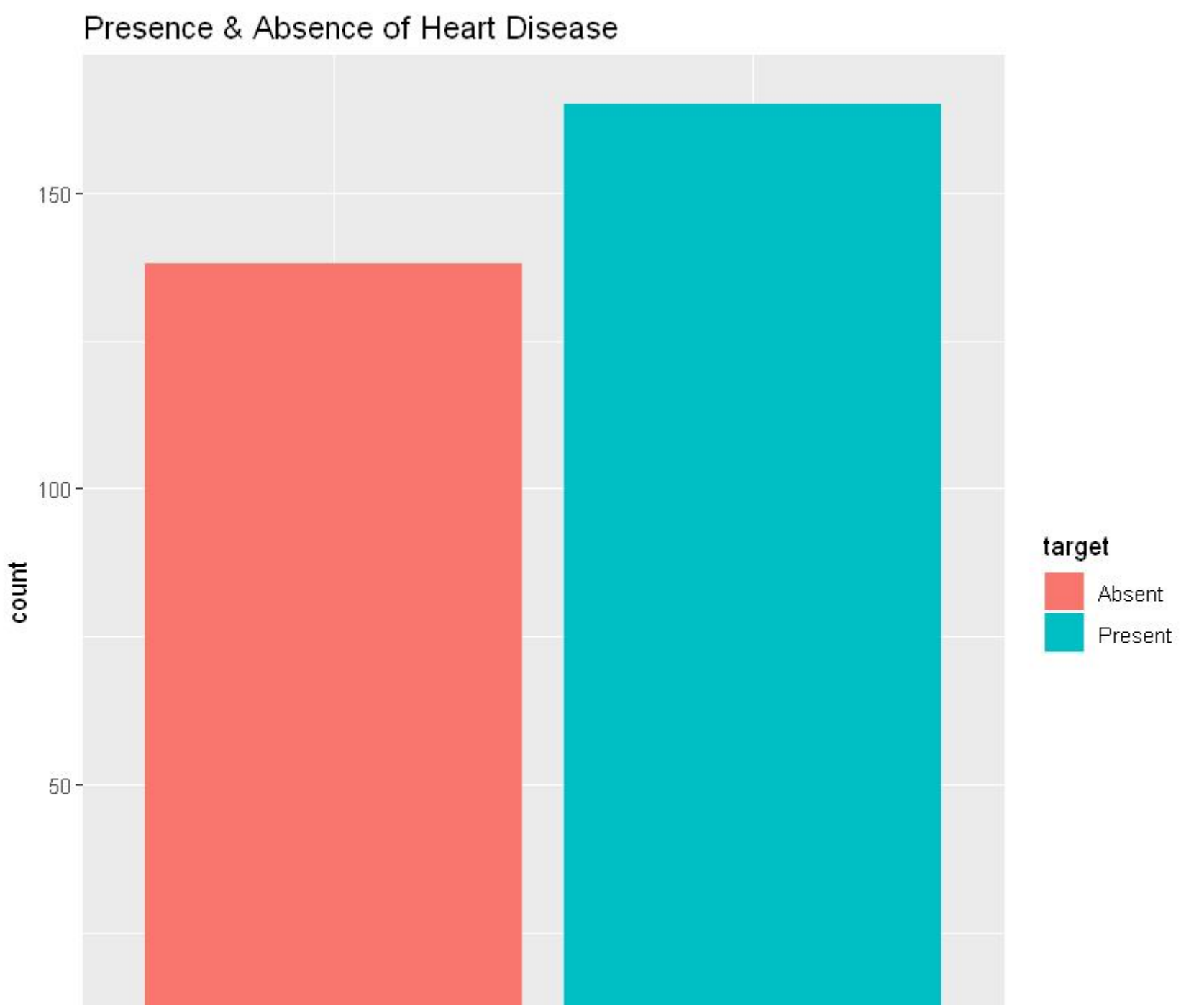
```
'data.frame':   303 obs. of  14 variables:
 $ ï..age  : int  63 37 41 56 57 57 56 44 52 57 ...
 $ sex     : Factor w/ 2 levels "FEMALE","MALE": 2 2 1 2 1 2 1 2 2 2 ...
 $ cp      : Factor w/ 3 levels "ASYMPTOMATIC",..: 1 3 2 2 1 1 2 2 3 3 ...
 $ trestbps: int  145 130 130 120 120 140 140 120 172 150 ...
 $ chol    : int  233 250 204 236 354 192 294 263 199 168 ...
 $ fbs     : Factor w/ 2 levels "<=120",">120": 2 1 1 1 1 1 1 1 2 1 ...
 $ restecg : Factor w/ 3 levels "ABNORMALITY",..: 2 1 2 1 1 1 2 1 1 1 ...
 $ thalach : int  150 187 172 178 163 148 153 173 162 174 ...
 $ exang   : Factor w/ 2 levels "NO","YES": 1 1 1 1 2 1 1 1 1 1 ...
 $ oldpeak : num  2.3 3.5 1.4 0.8 0.6 0.4 1.3 0 0.5 1.6 ...
 $ slope   : Factor w/ 3 levels "0","1","2": 1 1 3 3 3 2 2 3 3 3 ...
 $ ca      : Factor w/ 5 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ thal    : Factor w/ 4 levels "0","1","2","3": 2 3 3 3 3 2 3 4 4 3 ...
 $ target  : Factor w/ 2 levels "Absent","Present": 2 2 2 2 2 2 2 2 2 2 ...
```

In [87]:
```
CrossTable(heart$target)
```

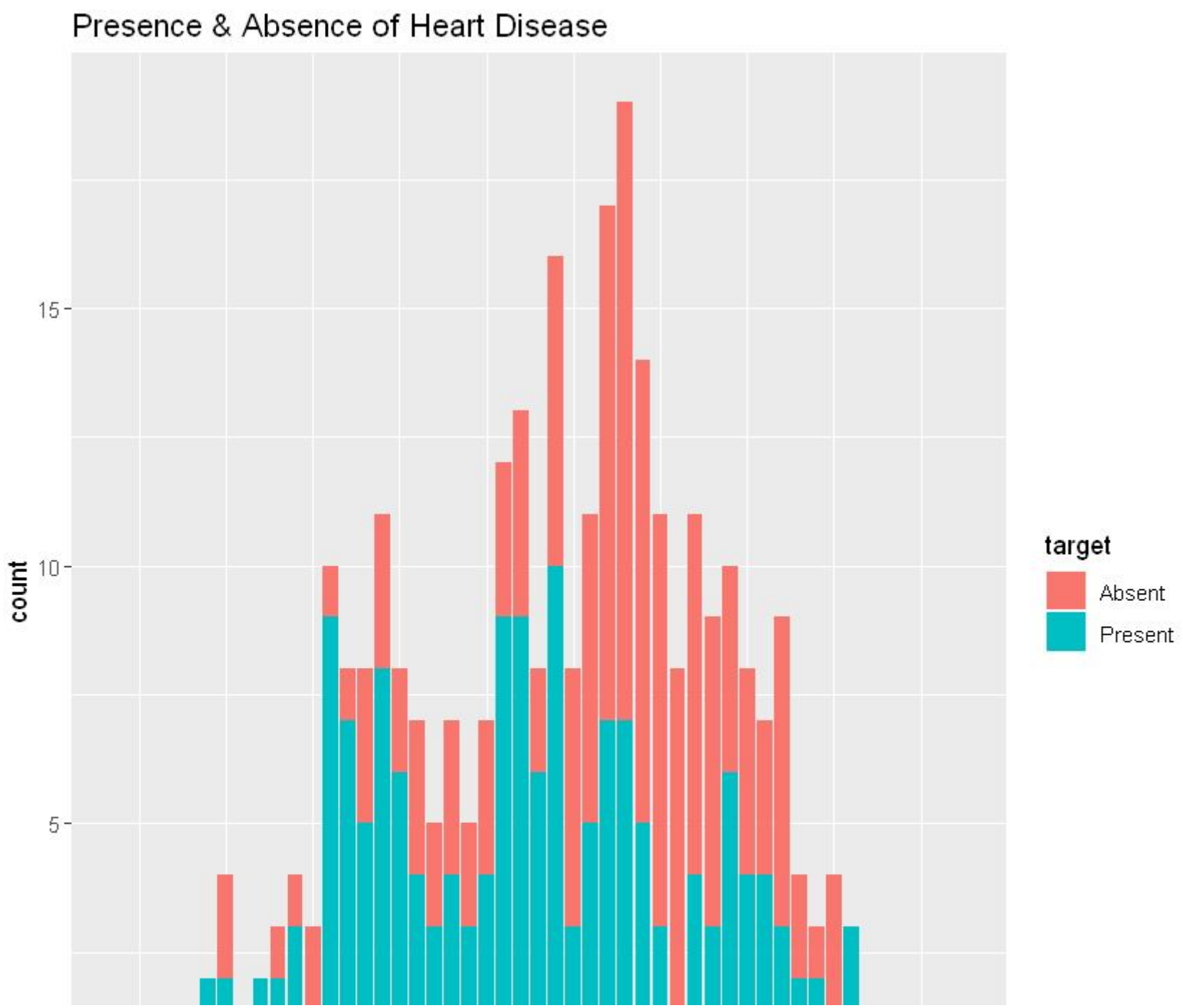
```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|-------------------------|

|       0 |       1 |
|---------|---------|
|     138 |     165 |
|   0.455 |   0.545 |
|---------|---------|
```

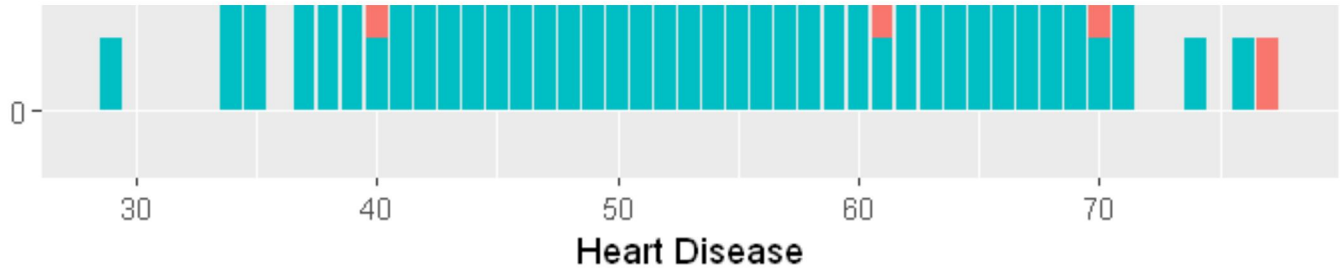
Data Visulization

In [88]:
```
# Bar plot for heart disease

ggplot(heart_data, aes(x=heart_data$target, fill=target))+
   geom_bar()+
   xlab("Heart Disease")+
   ylab("count")+
   ggtitle("Presence & Absence of Heart Disease")
```


Presence & Absence of Heart Disease
150
100
50
count
target
Absent
Present

In [89]:
```
# age wise heart disease data
ggplot(heart_data, aes(x=heart_data$ï..age, fill=target))+
   geom_bar()+
   xlab("Heart Disease")+
   ylab("count")+
   ggtitle("Presence & Absence of Heart Disease")
```


Presence & Absence of Heart Disease
count
15
10
5
target
Absent
Present



Dividing the age into different age group for better understanding

In [90]:
```
heart_data$age_bar=if_else(heart_data$ï..age>70,">70 years",if_else(heart_data$ï..age>60,"60-70",if_else(heart_data$ï..age>50,
                                                                                                          if_else(heart_data$ï..age>40,"4
                                                                                                                  if_else(heart_data$ï..age>30,"30-4
```

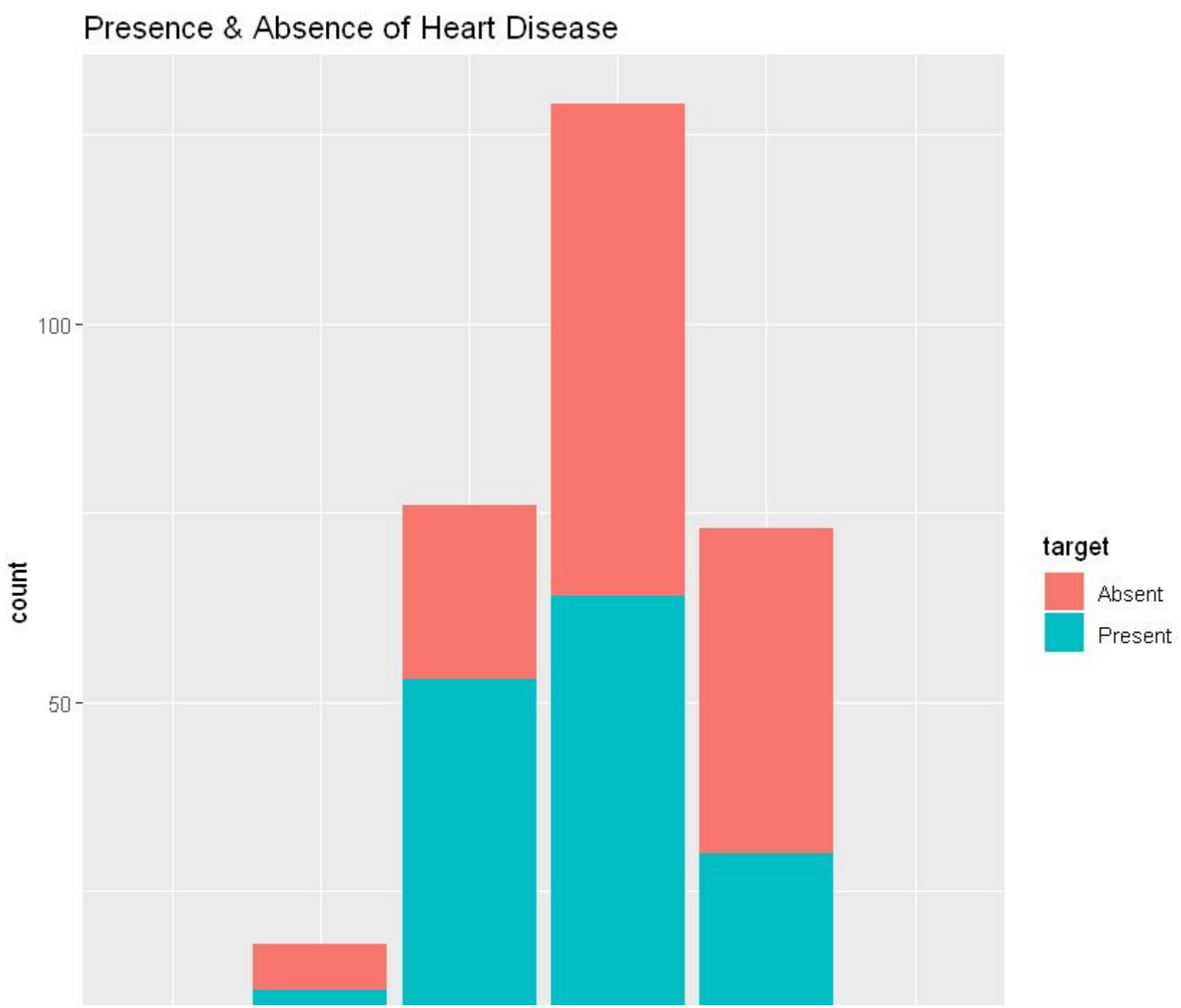
In [91]:
```
tail(heart_data)
```

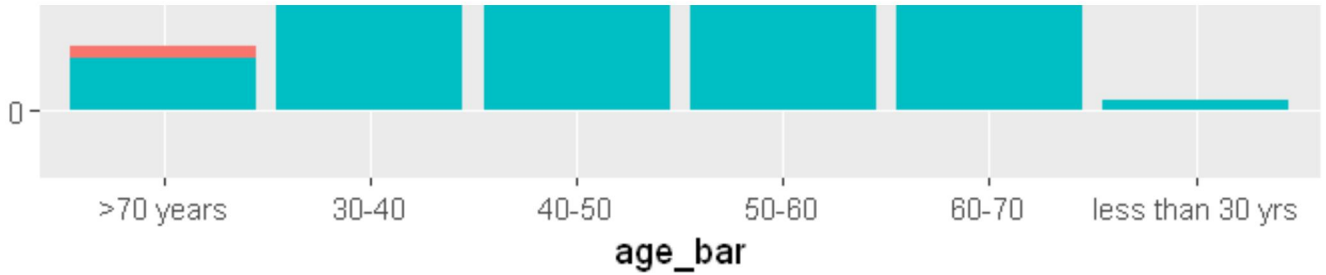
	ï..age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	age_bar
298	59	MALE	ASYMPTOMATIC	164	176	>120	NORMAL	90	NO	1.0	1	2	1	Absent	50-60
299	57	FEMALE	ASYMPTOMATIC	140	241	<=120	ABNORMALITY	123	YES	0.2	1	0	3	Absent	50-60
300	45	MALE	ASYMPTOMATIC	110	264	<=120	ABNORMALITY	132	NO	1.2	1	0	3	Absent	40-50
301	68	MALE	ASYMPTOMATIC	144	193	>120	ABNORMALITY	141	NO	3.4	1	2	3	Absent	60-70
302	57	MALE	ASYMPTOMATIC	130	131	<=120	ABNORMALITY	115	YES	1.2	1	1	3	Absent	50-60
303	57	FEMALE	ATYPICAL ANGINA	130	236	<=120	NORMAL	174	NO	0.0	1	1	2	Absent	50-60

In [92]:
```
heart_data$age_bar=as.factor(heart_data$age_bar)
```

In [93]:
```
ggplot(heart_data,aes(x=age_bar,fill=target))+geom_bar()+ ggtitle("Presence & Absence of Heart Disease")
crosstab(heart_data$age_bar,heart_data$disease,prop.r = T,plot = FALSE)
```

```
Error in table(dep, indep): all arguments must have the same length
Traceback:

1. crosstab(heart_data$age_bar, heart_data$disease, prop.r = T,
 .     plot = FALSE)
2. table(dep, indep)
3. stop("all arguments must have the same length")
```

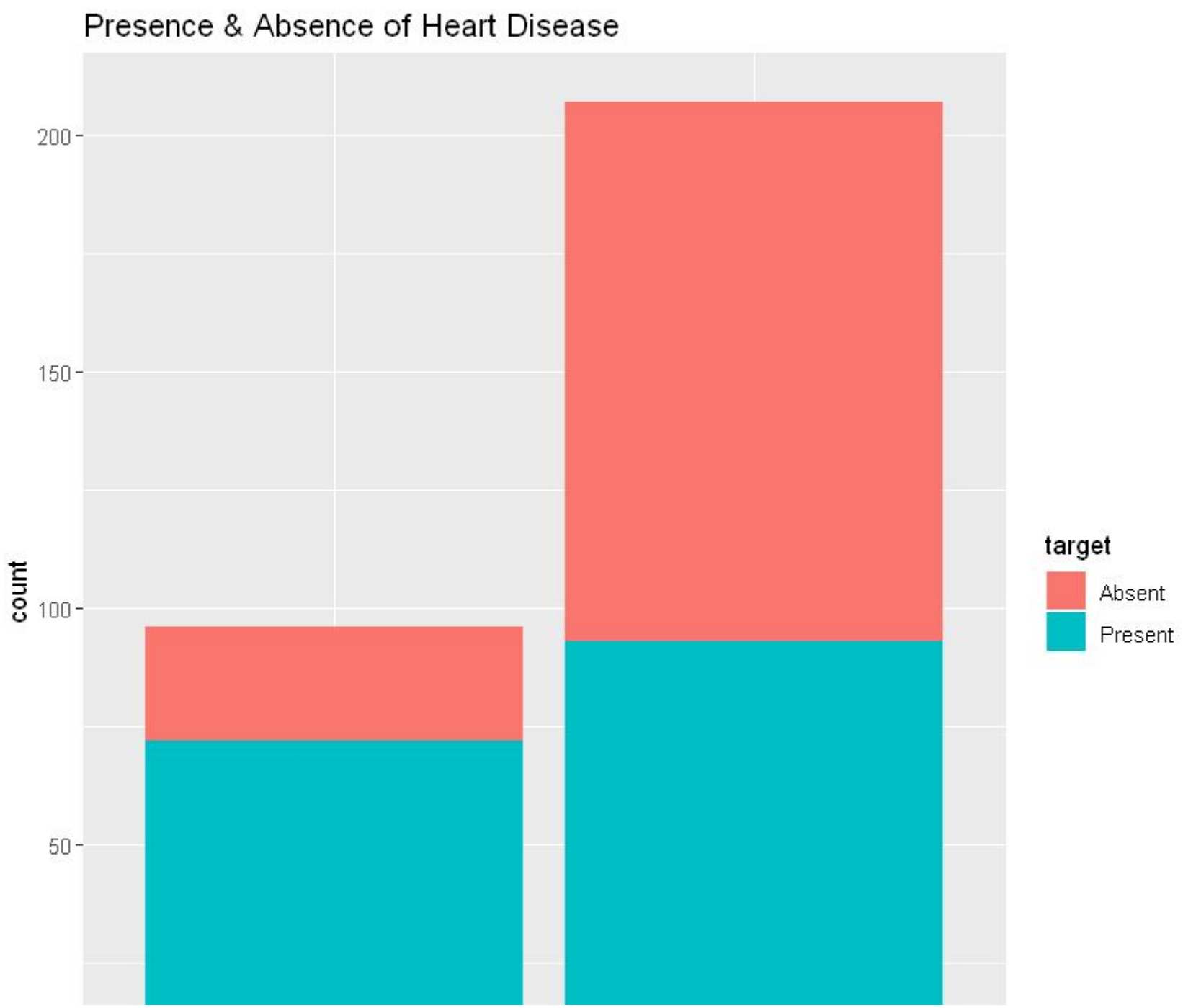

Presence & Absence of Heart Disease
count
100
50
target
Absent
Present



In [ ]:
```
table(heart_data$sex)
# There is a imbalance between male and female data
```

In [94]:
```
# comapre Heart disease Genderwise

heart_data %>%
  ggplot(aes(x=sex,fill=target))+
  geom_bar()+ ggtitle("Presence & Absence of Heart Disease")
crosstab(heart_data$sex,heart_data$disease,prop.r = T,plot = FALSE)
```

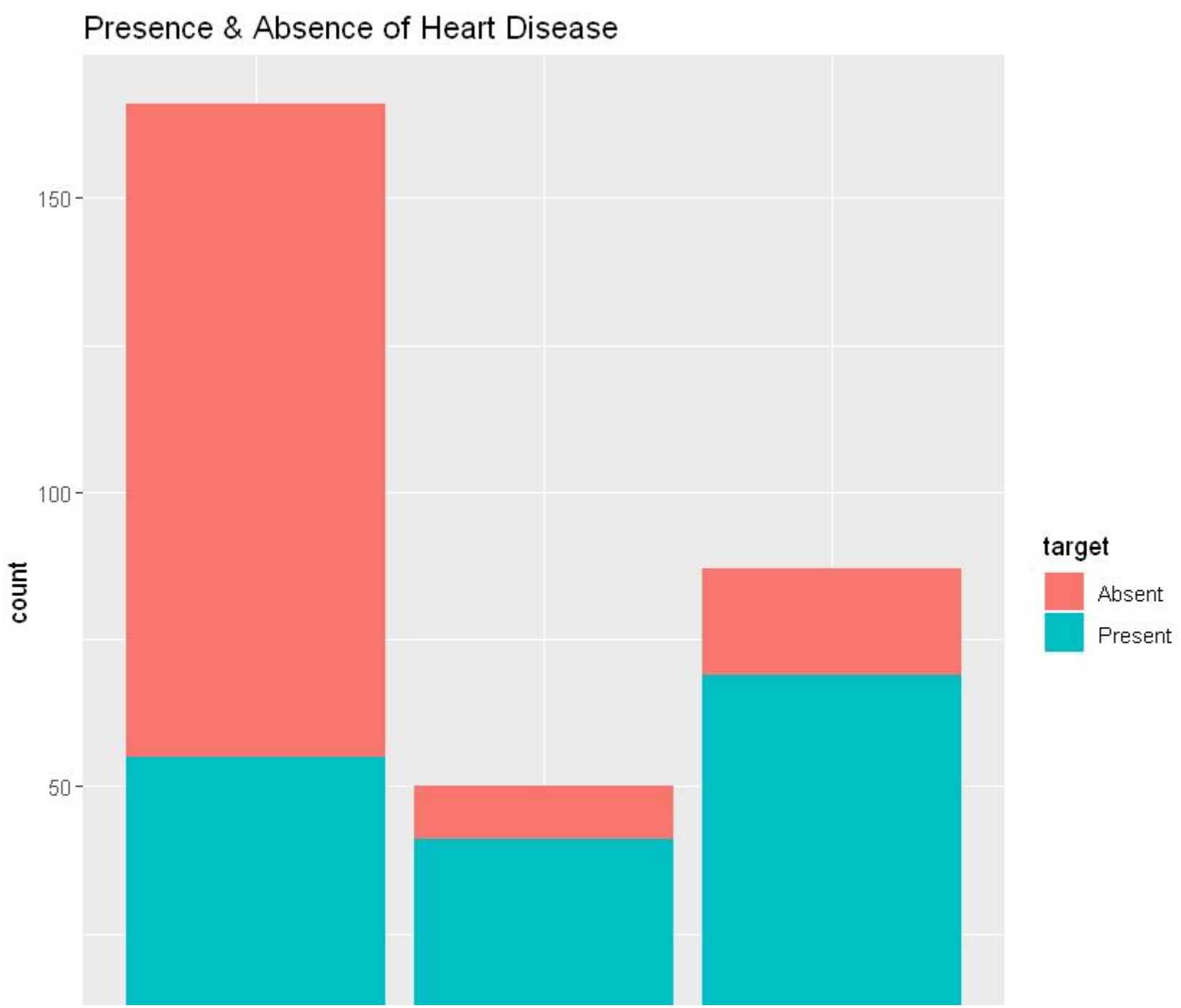
```
Error in table(dep, indep): all arguments must have the same length
Traceback:

1. crosstab(heart_data$sex, heart_data$disease, prop.r = T, plot = FALSE)
2. table(dep, indep)
3. stop("all arguments must have the same length")
```

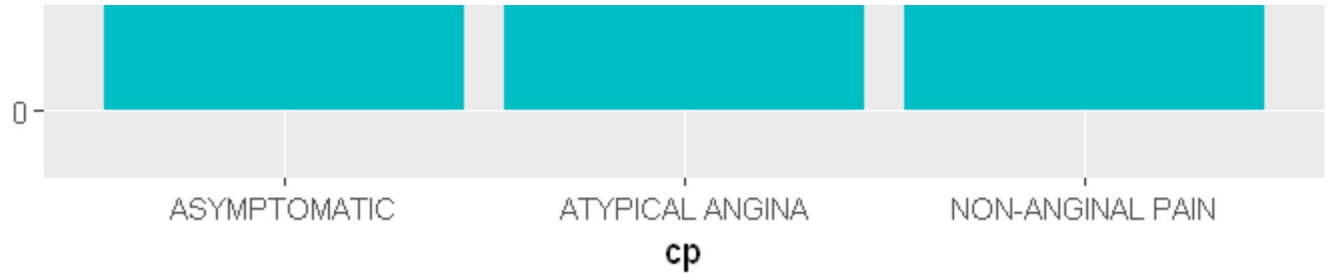

Presence & Absence of Heart Disease
count
200
150
100
50
target
Absent
Present

Female are more prone to heart disease as compared to male.

In [95]:
```
# Chest pain analysis
#cp

#Chest pain type

heart_data %>%
  ggplot(aes(x=cp,fill=target))+
  geom_bar()+ ggtitle("Presence & Absence of Heart Disease")
crosstab(heart_data$cp,heart_data$disease,prop.r = T,plot = FALSE)
```

```
Error in table(dep, indep): all arguments must have the same length
Traceback:

1. crosstab(heart_data$cp, heart_data$disease, prop.r = T, plot = FALSE)
2. table(dep, indep)
3. stop("all arguments must have the same length")
```


Presence & Absence of Heart Disease
150
100
50
count
target
Absent
Present



Here we see that it is very difficult to tell whether a patient has a heart disease attending just to the symptoms of the patients.

In [96]:
```
#trestbps

# Resting blood pressure in millimeters of mercury (mm Hg) when the patient was admitted to the hospital.

resting_bp= heart_data %>%
  mutate(resting_bps=case_when(100<=trestbps & trestbps<=130~'Normal',
                   130<trestbps & trestbps<=140~'Moderately high',
                   140<trestbps & trestbps<=160~'Very high',

                   TRUE~'>160 extreamly high'))
```

In [97]:
```
resting_bp
```

ï..age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	age_bar	rest
63	MALE	ASYMPTOMATIC	145	233	>120	NORMAL	150	NO	2.3	0	0	1	Present	60-70	Very
37	MALE	NON-ANGINAL PAIN	130	250	<=120	ABNORMALITY	187	NO	3.5	0	0	2	Present	30-40	Norr
41	FEMALE	ATYPICAL ANGINA	130	204	<=120	NORMAL	172	NO	1.4	2	0	2	Present	40-50	Norr
56	MALE	ATYPICAL ANGINA	120	236	<=120	ABNORMALITY	178	NO	0.8	2	0	2	Present	50-60	Norr
57	FEMALE	ASYMPTOMATIC	120	354	<=120	ABNORMALITY	163	YES	0.6	2	0	2	Present	50-60	Norr
57	MALE	ASYMPTOMATIC	140	192	<=120	ABNORMALITY	148	NO	0.4	1	0	1	Present	50-60	Mo
56	FEMALE	ATYPICAL ANGINA	140	294	<=120	NORMAL	153	NO	1.3	1	0	2	Present	50-60	Mo
44	MALE	ATYPICAL ANGINA	120	263	<=120	ABNORMALITY	173	NO	0.0	2	0	3	Present	40-50	Norr
52	MALE	NON-ANGINAL PAIN	172	199	>120	ABNORMALITY	162	NO	0.5	2	0	3	Present	50-60	e
57	MALE	NON-ANGINAL PAIN	150	168	<=120	ABNORMALITY	174	NO	1.6	2	0	2	Present	50-60	Very
54	MALE	ASYMPTOMATIC	140	239	<=120	ABNORMALITY	160	NO	1.2	2	0	2	Present	50-60	Mo
48	FEMALE	NON-ANGINAL PAIN	130	275	<=120	ABNORMALITY	139	NO	0.2	2	0	2	Present	40-50	Norr
49	MALE	ATYPICAL ANGINA	130	266	<=120	ABNORMALITY	171	NO	0.6	2	0	2	Present	40-50	Norr
64	MALE	ASYMPTOMATIC	110	211	<=120	NORMAL	144	YES	1.8	1	0	2	Present	60-70	Norr
58	FEMALE	ASYMPTOMATIC	150	283	>120	NORMAL	162	NO	1.0	2	0	2	Present	50-60	Very
50	FEMALE	NON-ANGINAL PAIN	120	219	<=120	ABNORMALITY	158	NO	1.6	1	0	2	Present	40-50	Norr

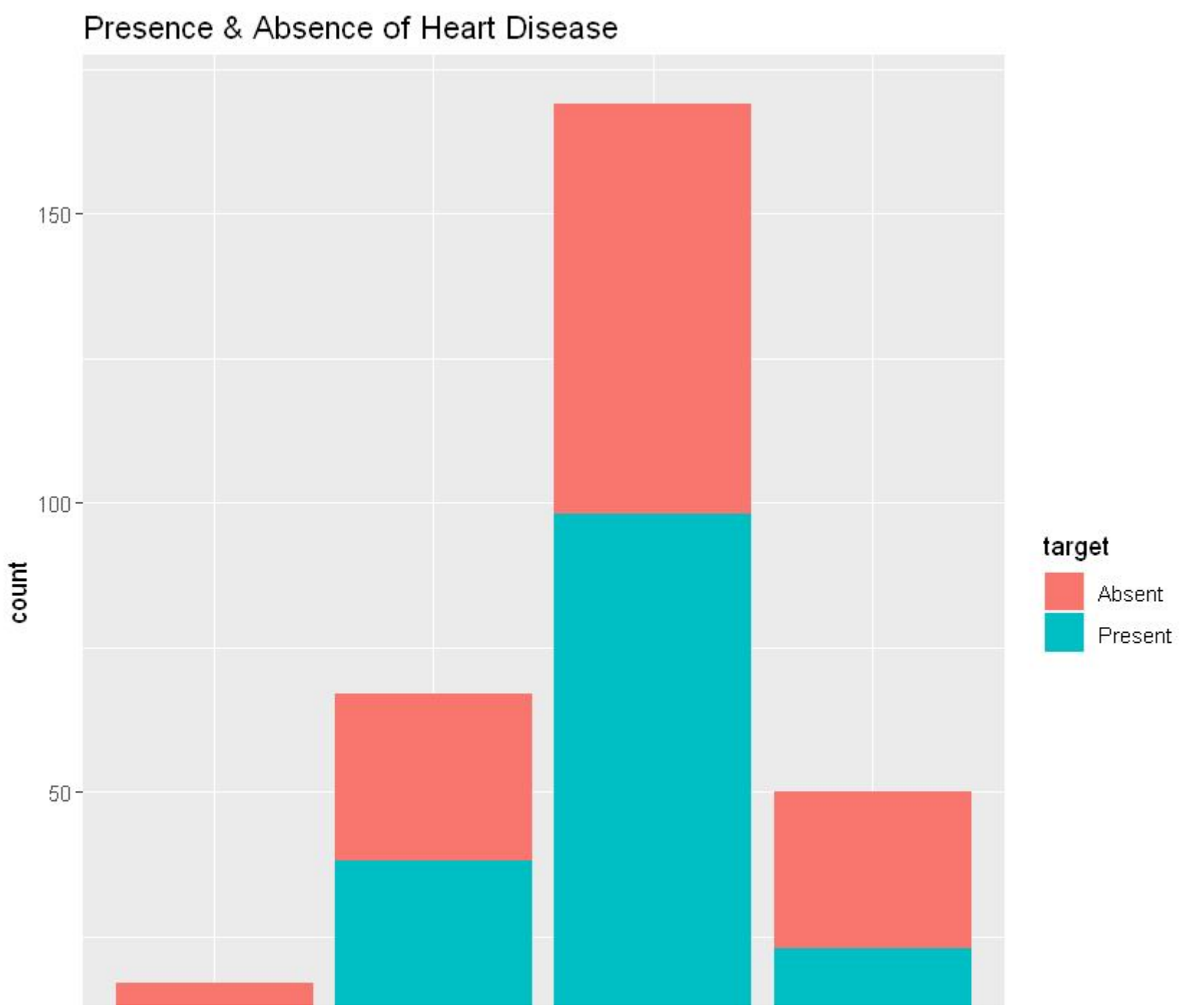
ï..age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	age_bar	rest
58	FEMALE	NON-ANGINAL PAIN	120	340	<=120	ABNORMALITY	172	NO	0.0	2	0	2	Present	50-60	Norr
66	FEMALE	ASYMPTOMATIC	150	226	<=120	ABNORMALITY	114	NO	2.6	0	0	2	Present	60-70	Very
43	MALE	ASYMPTOMATIC	150	247	<=120	ABNORMALITY	171	NO	1.5	2	0	2	Present	40-50	Very
69	FEMALE	ASYMPTOMATIC	140	239	<=120	ABNORMALITY	151	NO	1.8	2	2	2	Present	60-70	Mo
59	MALE	ASYMPTOMATIC	135	234	<=120	ABNORMALITY	161	NO	0.5	1	0	3	Present	50-60	Mo
44	MALE	NON-ANGINAL PAIN	130	233	<=120	ABNORMALITY	179	YES	0.4	2	0	2	Present	40-50	Norr
42	MALE	ASYMPTOMATIC	140	226	<=120	ABNORMALITY	178	NO	0.0	2	0	2	Present	40-50	Mo
61	MALE	NON-ANGINAL PAIN	150	243	>120	ABNORMALITY	137	YES	1.0	1	0	2	Present	60-70	Very
40	MALE	ASYMPTOMATIC	140	199	<=120	ABNORMALITY	178	YES	1.4	2	0	3	Present	30-40	Mo
71	FEMALE	ATYPICAL ANGINA	160	302	<=120	ABNORMALITY	162	NO	0.4	2	2	2	Present	>70 years	Very
59	MALE	NON-ANGINAL PAIN	150	212	>120	ABNORMALITY	157	NO	1.6	2	0	2	Present	50-60	Very
51	MALE	NON-ANGINAL PAIN	110	175	<=120	ABNORMALITY	123	NO	0.6	2	0	2	Present	50-60	Norr
65	FEMALE	NON-ANGINAL PAIN	140	417	>120	NORMAL	157	NO	0.8	2	1	2	Present	60-70	Mo
53	MALE	NON-ANGINAL PAIN	130	197	>120	NORMAL	152	NO	1.2	0	0	2	Present	50-60	Norr
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
58	MALE	ASYMPTOMATIC	100	234	<=120	ABNORMALITY	156	NO	0.1	2	1	3	Absent	50-60	Norr
47	MALE	ASYMPTOMATIC	110	275	<=120	NORMAL	118	YES	1.0	1	1	2	Absent	40-50	Norr

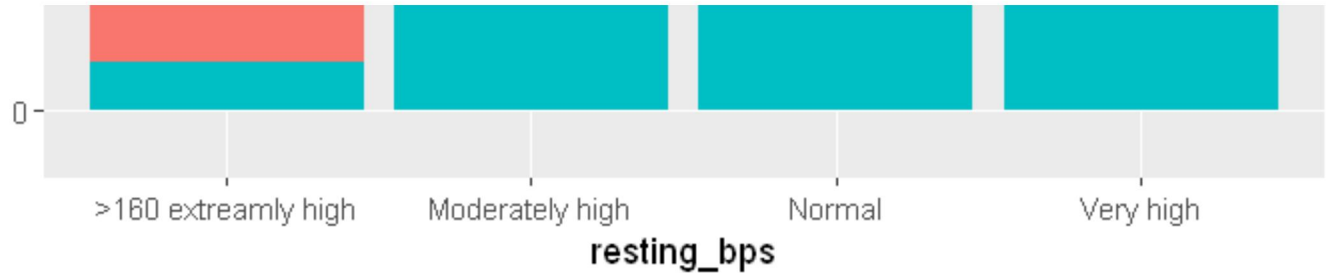
ï..age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	age_bar	rest
52	MALE	ASYMPTOMATIC	125	212	<=120	ABNORMALITY	168	NO	1.0	2	2	3	Absent	50-60	Norr
58	MALE	ASYMPTOMATIC	146	218	<=120	ABNORMALITY	105	NO	2.0	1	1	3	Absent	50-60	Very
57	MALE	ATYPICAL ANGINA	124	261	<=120	ABNORMALITY	141	NO	0.3	2	0	3	Absent	50-60	Norr
58	FEMALE	ATYPICAL ANGINA	136	319	>120	NORMAL	152	NO	0.0	2	2	2	Absent	50-60	Mo
61	MALE	ASYMPTOMATIC	138	166	<=120	NORMAL	125	YES	3.6	1	1	2	Absent	60-70	Mo
42	MALE	ASYMPTOMATIC	136	315	<=120	ABNORMALITY	125	YES	1.8	1	0	1	Absent	40-50	Mo
52	MALE	ASYMPTOMATIC	128	204	>120	ABNORMALITY	156	YES	1.0	1	0	0	Absent	50-60	Norr
59	MALE	NON-ANGINAL PAIN	126	218	>120	ABNORMALITY	134	NO	2.2	1	1	1	Absent	50-60	Norr
40	MALE	ASYMPTOMATIC	152	223	<=120	ABNORMALITY	181	NO	0.0	2	0	3	Absent	30-40	Very
61	MALE	ASYMPTOMATIC	140	207	<=120	NORMAL	138	YES	1.9	2	1	3	Absent	60-70	Mo
46	MALE	ASYMPTOMATIC	140	311	<=120	ABNORMALITY	120	YES	1.8	1	2	3	Absent	40-50	Mo
59	MALE	ASYMPTOMATIC	134	204	<=120	ABNORMALITY	162	NO	0.8	2	2	2	Absent	50-60	Mo
57	MALE	ATYPICAL ANGINA	154	232	<=120	NORMAL	164	NO	0.0	2	1	2	Absent	50-60	Very
57	MALE	ASYMPTOMATIC	110	335	<=120	ABNORMALITY	143	YES	3.0	1	1	3	Absent	50-60	Norr
55	FEMALE	ASYMPTOMATIC	128	205	<=120	PROBABLE OR DEFINITE	130	YES	2.0	1	1	3	Absent	50-60	Norr
61	MALE	ASYMPTOMATIC	148	203	<=120	ABNORMALITY	161	NO	0.0	2	1	3	Absent	60-70	Very
58	MALE	ASYMPTOMATIC	114	318	<=120	PROBABLE OR DEFINITE	140	NO	4.4	0	3	1	Absent	50-60	Norr

ï..age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	age_bar	rest
58	FEMALE	ASYMPTOMATIC	170	225	>120	NORMAL	146	YES	2.8	1	2	1	Absent	50-60	e
67	MALE	NON-ANGINAL PAIN	152	212	<=120	NORMAL	150	NO	0.8	1	0	3	Absent	60-70	Very
44	MALE	ASYMPTOMATIC	120	169	<=120	ABNORMALITY	144	YES	2.8	0	0	1	Absent	40-50	Norr
63	MALE	ASYMPTOMATIC	140	187	<=120	NORMAL	144	YES	4.0	2	2	3	Absent	60-70	Mo
63	FEMALE	ASYMPTOMATIC	124	197	<=120	ABNORMALITY	136	YES	0.0	1	0	2	Absent	60-70	Norr
59	MALE	ASYMPTOMATIC	164	176	>120	NORMAL	90	NO	1.0	1	2	1	Absent	50-60	e
57	FEMALE	ASYMPTOMATIC	140	241	<=120	ABNORMALITY	123	YES	0.2	1	0	3	Absent	50-60	Mo
45	MALE	ASYMPTOMATIC	110	264	<=120	ABNORMALITY	132	NO	1.2	1	0	3	Absent	40-50	Norr
68	MALE	ASYMPTOMATIC	144	193	>120	ABNORMALITY	141	NO	3.4	1	2	3	Absent	60-70	Very
57	MALE	ASYMPTOMATIC	130	131	<=120	ABNORMALITY	115	YES	1.2	1	1	3	Absent	50-60	Norr
57	FEMALE	ATYPICAL ANGINA	130	236	<=120	NORMAL	174	NO	0.0	1	1	2	Absent	50-60	Norr

In [98]:
```
resting_bp %>%
  ggplot(aes(x=resting_bps,fill=target))+
  geom_bar()+ ggtitle("Presence & Absence of Heart Disease")
crosstab(resting_bp$resting_bps,heart_data$disease,prop.r = T,plot = FALSE)
```

```
Error in table(dep, indep): all arguments must have the same length
Traceback:

1. crosstab(resting_bp$resting_bps, heart_data$disease, prop.r = T,
 .     plot = FALSE)
2. table(dep, indep)
3. stop("all arguments must have the same length")
```

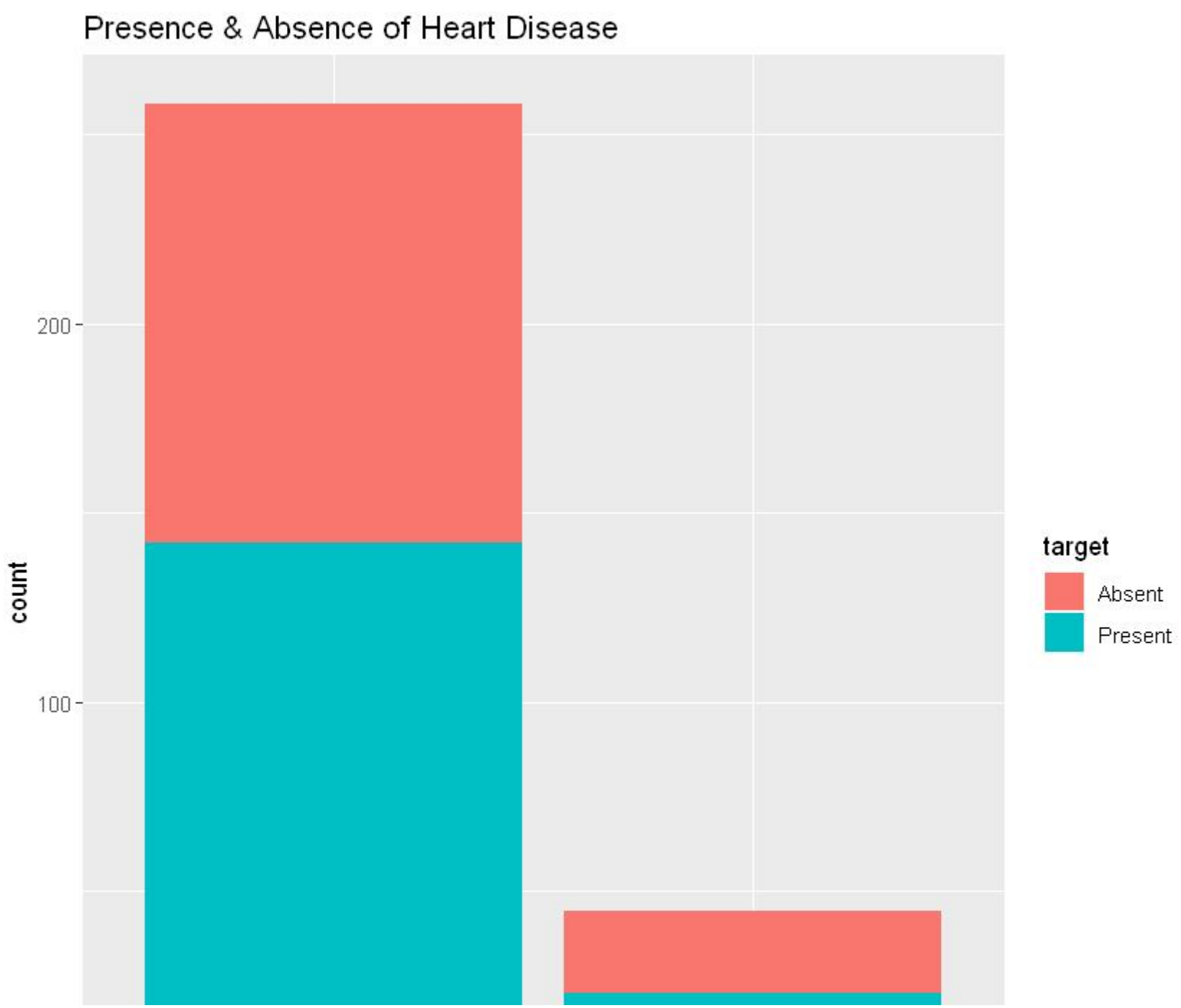

Presence & Absence of Heart Disease
150
100
50
count
target
Absent
Present



In [ ]:
```
# Cholestrol
ggplot(heart_data, aes(chol, fill=target)) +
  geom_histogram(binwidth=10) +
 ggtitle("Presence & Absence of Heart Disease")
```

In [99]:
```
heart_data$fbs=as.factor(heart_data$fbs)
ggplot(heart_data, aes(fbs, fill=target)) +
  geom_bar() + ggtitle("Presence & Absence of Heart Disease")


crosstab(heart_data$fbs,heart_data$disease,prop.r = T,plot = FALSE)
```

```
Error in table(dep, indep): all arguments must have the same length
Traceback:

1. crosstab(heart_data$fbs, heart_data$disease, prop.r = T, plot = FALSE)
2. table(dep, indep)
3. stop("all arguments must have the same length")
```

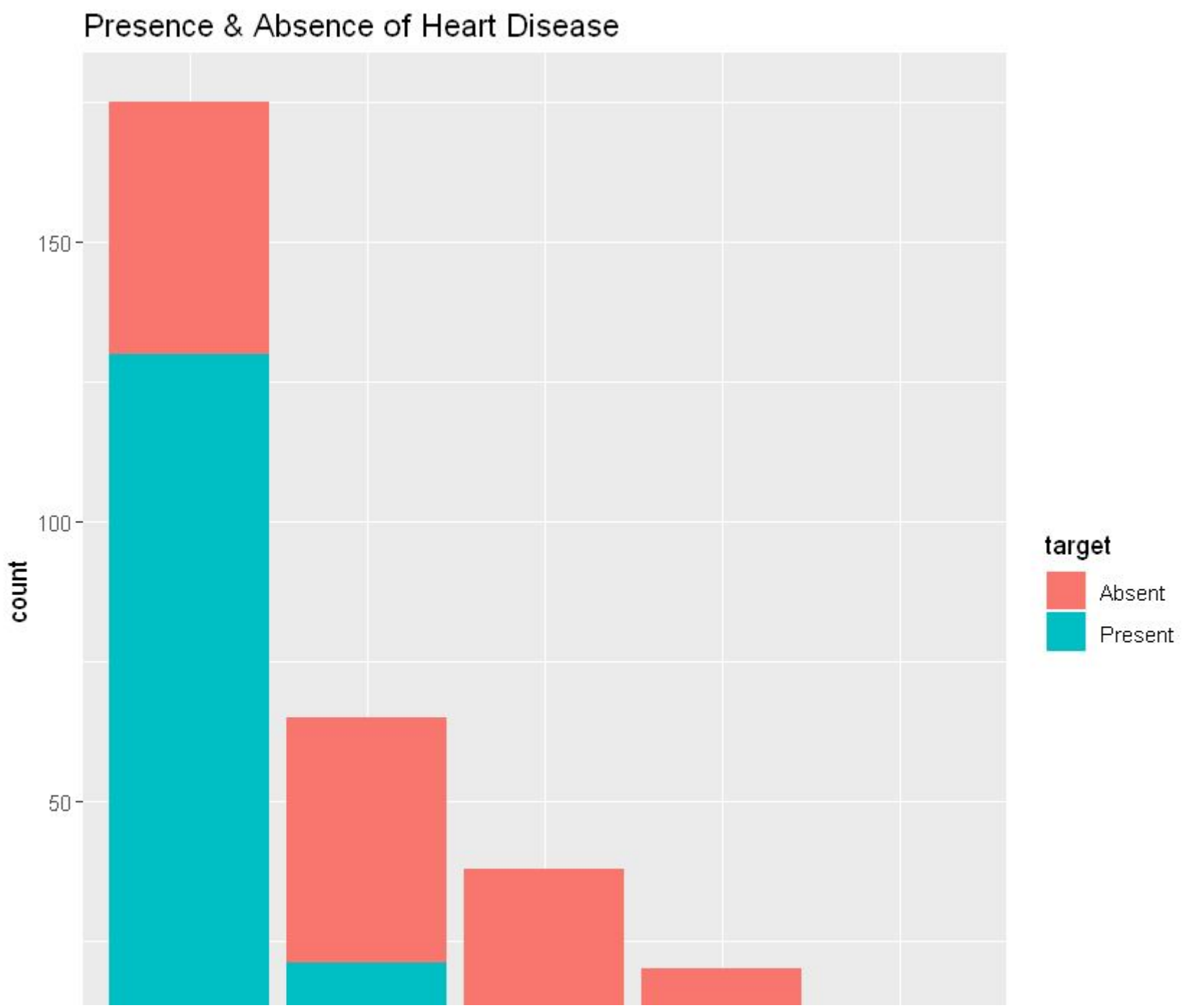

Presence & Absence of Heart Disease
count
200
100
target
Absent
Present

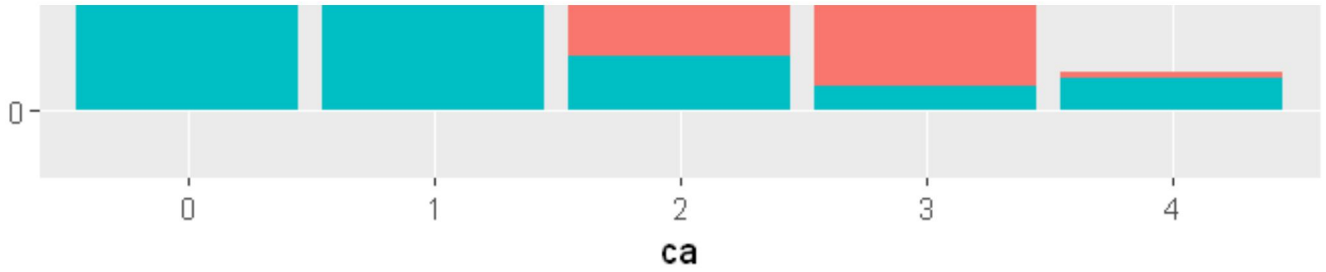
There is no concluding evidence about the sugar level effect on heart disease.

In [100…

```
#ca
#This feature refers to the number of narrow blood vessels seen

ggplot(heart_data, aes(ca, fill=target)) +
  geom_bar() + ggtitle("Presence & Absence of Heart Disease")

crosstab(heart_data$ca,heart_data$disease,prop.r = T,plot = FALSE)
```

```
Error in table(dep, indep): all arguments must have the same length
Traceback:

1. crosstab(heart_data$ca, heart_data$disease, prop.r = T, plot = FALSE)
2. table(dep, indep)
3. stop("all arguments must have the same length")
```


Presence & Absence of Heart Disease
150
100
50
count
target
Absent
Present



In [ ]:
```
# Corelation between independent variable
library(corrplot)
corr_heart= heart_data[,sapply(heart_data,is.numeric)]
str(corr_heart)
cor(corr_heart)
```

In [ ]:

Model Building

In [101…
```
library(caret)
```

TraindataRows = createDataPartition(heart_data$target, p=0.7, list= FALSE) trainData = heart_data[TraindataRows,] testData = heart_data[-TraindataRows,]

In [102…
```
AUC = list()
Accuracy = list()
```

Logistics Regression

In [ ]:
```
install.packages('e1071', dependencies = TRUE)
```

In [ ]:
```
install.packages("pROC")
```

In [ ]:
```
set.seed(10)
logRegModel = train(target ~ ., data=trainData, method = 'glm', family = 'binomial')
logRegPrediction = predict(logRegModel, testData)
logRegPredictionprob = predict(logRegModel, testData, type='prob')[2]
logRegConfMat = confusionMatrix(logRegPrediction, testData[,"target"])
#ROC Curve
library(pROC)
AUC$logReg = roc(as.numeric(testData$target),as.numeric(as.matrix((logRegPredictionprob))))$auc
Accuracy$logReg = logRegConfMat$overall['Accuracy']
```

In [ ]:
```
# Support Vector Machine


install.packages('kernlab')
fitControl = trainControl(method = "repeatedcv",
                          number = 5,
                          repeats = 5,
                          ## Estimate class probabilities
                          classProbs = TRUE,
                          ## Evaluate performance using
                          ## the following function
                          summaryFunction = twoClassSummary)


set.seed(10)
svmModel = train(target ~ ., data = trainData,
                 method = "svmRadial",
                 trControl = fitControl ,
                 preProcess = c("center", "scale"),
                 tunelength = 8,
                 metric = "ROC")
svmPrediction = predict(svmModel, testData)
svmPredictionprob = predict(svmModel, testData, type='prob')[2]
svmConfMat = confusionMatrix(svmPrediction, testData[,"target"])
#ROC Curve
library(pROC)
AUC$svm = roc(as.numeric(testData$target),as.numeric(as.matrix((svmPredictionprob))))$auc
Accuracy$svm = svmConfMat$overall['Accuracy']
```

In [ ]:
```
# Random Forest
library(randomForest)
set.seed(10)
RFModel = randomForest(target ~ .,
                       data=trainData,
                       importance=TRUE,
                       ntree=200)
#varImpPlot(RFModel)
RFPrediction = predict(RFModel, testData)
RFPredictionprob = predict(RFModel,testData,type="prob")[, 2]

RFConfMat = confusionMatrix(RFPrediction, testData[,"target"])

AUC$RF = roc(as.numeric(testData$target),as.numeric(as.matrix((RFPredictionprob))))$auc
Accuracy$RF = RFConfMat$overall['Accuracy']
```

In [ ]:
```
row.names = names(Accuracy)
col.names = c("AUC", "Accuracy")
cbind(as.data.frame(matrix(c(AUC,Accuracy),nrow = 3, ncol = 2,
                           dimnames = list(row.names, col.names))))
```

In [ ]:
```
logRegConfMat
```

So here accuracy rate is 86 % which is quite good.

In [ ]:
```
RFConfMat
```

In [ ]:
```
svmConfMat
```

So here Logistic Regression model gives higest accuracy.

In [ ]:

In [ ]: