Heart Rate Predictions

Pooja Rathee

26 April 2022

Read the following packages

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

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

```
library(ggplot2)
```

Read the file

```
heart_disease <- read_csv("Heart_Disease_Prediction.csv")
```

```
## Rows: 270 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): Heart Disease
## dbl (13): Age, Sex, Chest pain type, BP, Cholesterol, FBS over 120, EKG resu...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Viewing the table and column names in the file

```
colnames(heart_disease)
```

```
##  [1] "Age"                    "Sex"
##  [3] "Chest pain type"        "BP"
##  [5] "Cholesterol"            "FBS over 120"
##  [7] "EKG results"            "Max HR"
##  [9] "Exercise angina"        "ST depression"
## [11] "Slope of ST"            "Number of vessels fluro"
## [13] "Thallium"               "Heart Disease"
```

```
View(heart_disease)
```

Rename the columnes to make it more consistent

```
heart_disease1 <-
  rename(heart_disease, "age"= "Age" , "chest_pain_type" = "Chest pain type" , "chl"  = "Cholesterol" ,
```

Doing some analytical calculations alongwith formatting

```
# Format the heart rate
is.factor(heart_disease1$max_hr)
```

```
## [1] FALSE
```

```
heart_disease1$max_hr <- as.numeric(as.character(heart_disease1$max_hr))
is.numeric(heart_disease1$max_hr)
```

```
## [1] TRUE
```

```
# Find the mean, median, max and min values
mean(heart_disease1$max_hr) #straight average
```

```
## [1] 149.6778
```

```
median(heart_disease1$max_hr) #midpoint number in the ascending array of heart rate
```

```
## [1] 153.5
```

```
max(heart_disease1$max_hr) #max rate
```

```
## [1] 202
```

```r
min(heart_disease1$max_hr) #min rate
```

```
## [1] 71
```

```r
# Formatting the Sex column
is.numeric(heart_disease1$Sex)
```

```
## [1] TRUE
```

```r
heart_disease1$Sex <- as.factor(as.numeric(heart_disease1$Sex))
is.factor(heart_disease1$Sex)
```

```
## [1] TRUE
```

Summarize the column of Heart Rate and aggregate further it with presence and absence of heart diseases

```r
summary(heart_disease1$max_hr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    71.0   133.0   153.5   149.7   166.0   202.0
```

```r
# Compare presence and absence of disease with heart rate
aggregate(heart_disease1$max_hr ~ heart_disease1$heart_dis, FUN = mean)
```

```
##   heart_disease1$heart_dis heart_disease1$max_hr
## 1                  Absence              158.3333
## 2                 Presence              138.8583
```

```r
aggregate(heart_disease1$max_hr ~ heart_disease1$heart_dis, FUN = median)
```

```
##   heart_disease1$heart_dis heart_disease1$max_hr
## 1                  Absence                161.0
## 2                 Presence                141.5
```

```r
aggregate(heart_disease1$max_hr ~ heart_disease1$heart_dis, FUN = max)
```

```
##   heart_disease1$heart_dis heart_disease1$max_hr
## 1                  Absence                  202
## 2                 Presence                  195
```

```r
aggregate(heart_disease1$max_hr ~ heart_disease1$heart_dis, FUN = min)
```

```
##   heart_disease1$heart_dis heart_disease1$max_hr
## 1                  Absence                   96
## 2                 Presence                   71
```

```r
# Aggregate further it with age category
aggregate(heart_disease1$max_hr ~ heart_disease1$heart_dis + heart_disease1$age, FUN = mean)
```

```
##    heart_disease1$heart_dis heart_disease1$age heart_disease1$max_hr
## 1                    Absence                 29              202.0000
## 2                    Absence                 34              183.0000
## 3                    Absence                 35              182.0000
## 4                   Presence                 35              143.0000
## 5                    Absence                 37              178.5000
## 6                   Presence                 38              182.0000
## 7                    Absence                 39              180.5000
## 8                   Presence                 39              140.0000
## 9                    Absence                 40              178.0000
## 10                  Presence                 40              147.5000
## 11                   Absence                 41              163.3750
## 12                  Presence                 41              158.0000
## 13                   Absence                 42              165.2857
## 14                  Presence                 42              125.0000
## 15                   Absence                 43              168.0000
## 16                  Presence                 43              128.0000
## 17                   Absence                 44              172.8750
## 18                  Presence                 44              165.0000
## 19                   Absence                 45              161.3333
## 20                  Presence                 45              147.0000
## 21                   Absence                 46              160.0000
## 22                  Presence                 46              137.0000
## 23                   Absence                 47              149.5000
## 24                  Presence                 47              135.0000
## 25                   Absence                 48              170.0000
## 26                  Presence                 48              161.3333
## 27                   Absence                 49              165.3333
## 28                  Presence                 49              132.5000
## 29                   Absence                 50              160.5000
## 30                  Presence                 50              139.0000
## 31                   Absence                 51              149.4444
## 32                  Presence                 51              145.6667
## 33                   Absence                 52              170.0000
## 34                  Presence                 52              163.0000
## 35                   Absence                 53              147.8000
## 36                  Presence                 53              125.0000
## 37                   Absence                 54              159.7000
## 38                  Presence                 54              127.8333
## 39                   Absence                 55              158.0000
## 40                  Presence                 55              126.2500
## 41                   Absence                 56              164.3333
## 42                  Presence                 56              129.5000
## 43                   Absence                 57              158.7143
## 44                  Presence                 57              123.0000
## 45                   Absence                 58              155.0000
## 46                  Presence                 58              145.4000
## 47                   Absence                 59              161.8000
## 48                  Presence                 59              143.5714
## 49                   Absence                 60              142.3333
```

```
## 50                Presence                60             151.3333
## 51                 Absence                61             137.0000
## 52                Presence                61             143.8333
## 53                 Absence                62             151.5000
## 54                Presence                62             123.4286
## 55                 Absence                63             167.0000
## 56                Presence                63             150.5000
## 57                 Absence                64             131.6000
## 58                Presence                64             129.2500
## 59                 Absence                65             149.0000
## 60                Presence                65             143.2500
## 61                 Absence                66             134.3333
## 62                Presence                66             139.0000
## 63                 Absence                67             158.0000
## 64                Presence                67             119.2000
## 65                 Absence                68             133.0000
## 66                Presence                68             150.0000
## 67                 Absence                69             141.0000
## 68                Presence                69             146.0000
## 69                 Absence                70             143.0000
## 70                Presence                70             115.3333
## 71                 Absence                71             139.0000
## 72                 Absence                74             121.0000
## 73                 Absence                76             116.0000
## 74                Presence                77             162.0000
```

Piping the data

```
heart_disease1 %>%
  group_by(age,heart_dis) %>%
  summarise(avg_max_hr = mean(max_hr)) %>%
  arrange(age,heart_dis)
```

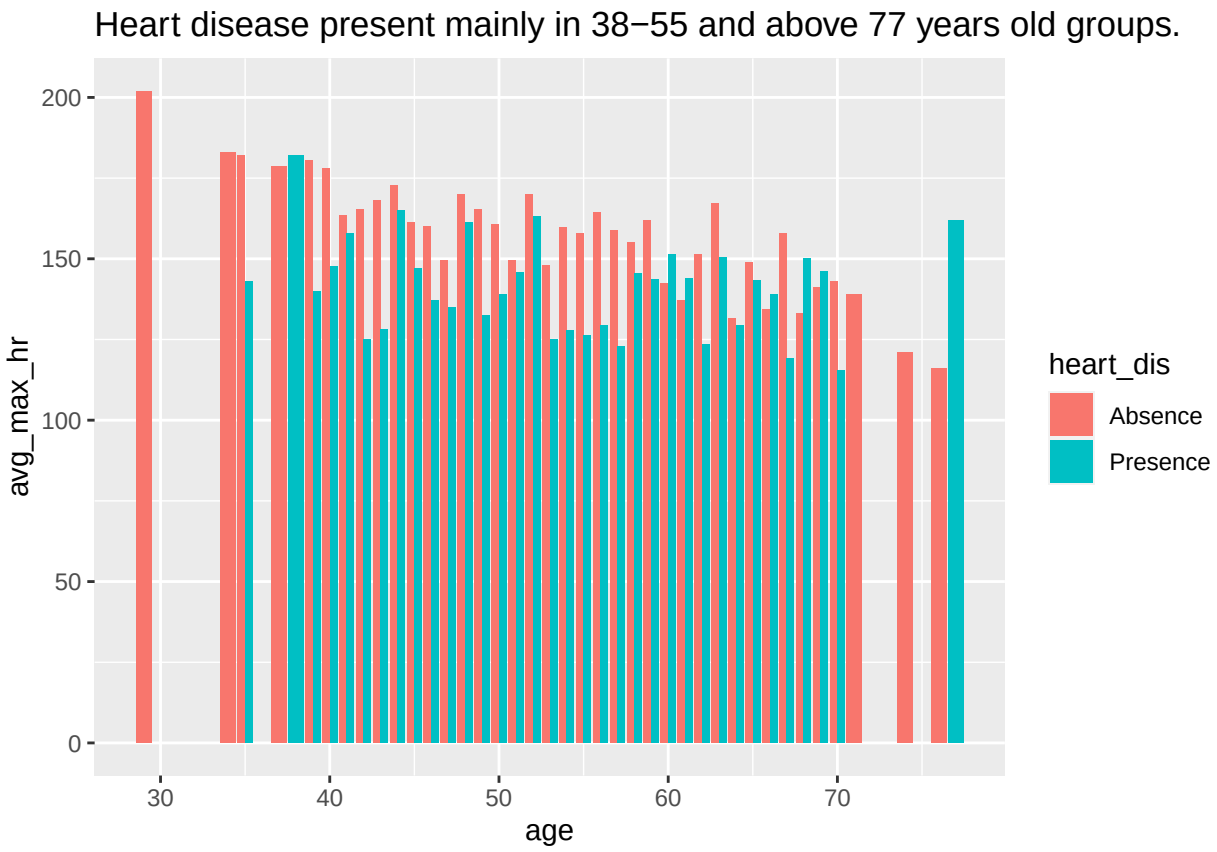
```
## `summarise()` has grouped output by 'age'. You can override using the `.groups`
## argument.
```

```
## # A tibble: 74 x 3
## # Groups:   age [41]
##      age heart_dis avg_max_hr
##    <dbl> <chr>          <dbl>
##  1    29 Absence         202
##  2    34 Absence         183
##  3    35 Absence         182
##  4    35 Presence        143
##  5    37 Absence         178.
##  6    38 Presence        182
##  7    39 Absence         180.
##  8    39 Presence        140
##  9    40 Absence         178
## 10    40 Presence        148.
## # ... with 64 more rows
```

Let's do some visualization

```
heart_disease1 %>%
  group_by(age,heart_dis) %>%
  summarise(avg_max_hr = mean(max_hr)) %>%
  arrange(age,heart_dis) %>%
 ggplot(aes(x = age, y = avg_max_hr, fill = heart_dis)) +
  geom_col(position = "dodge") +
  labs(title = "Heart disease present mainly in 38-55 and above 77 years old groups.")
```

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

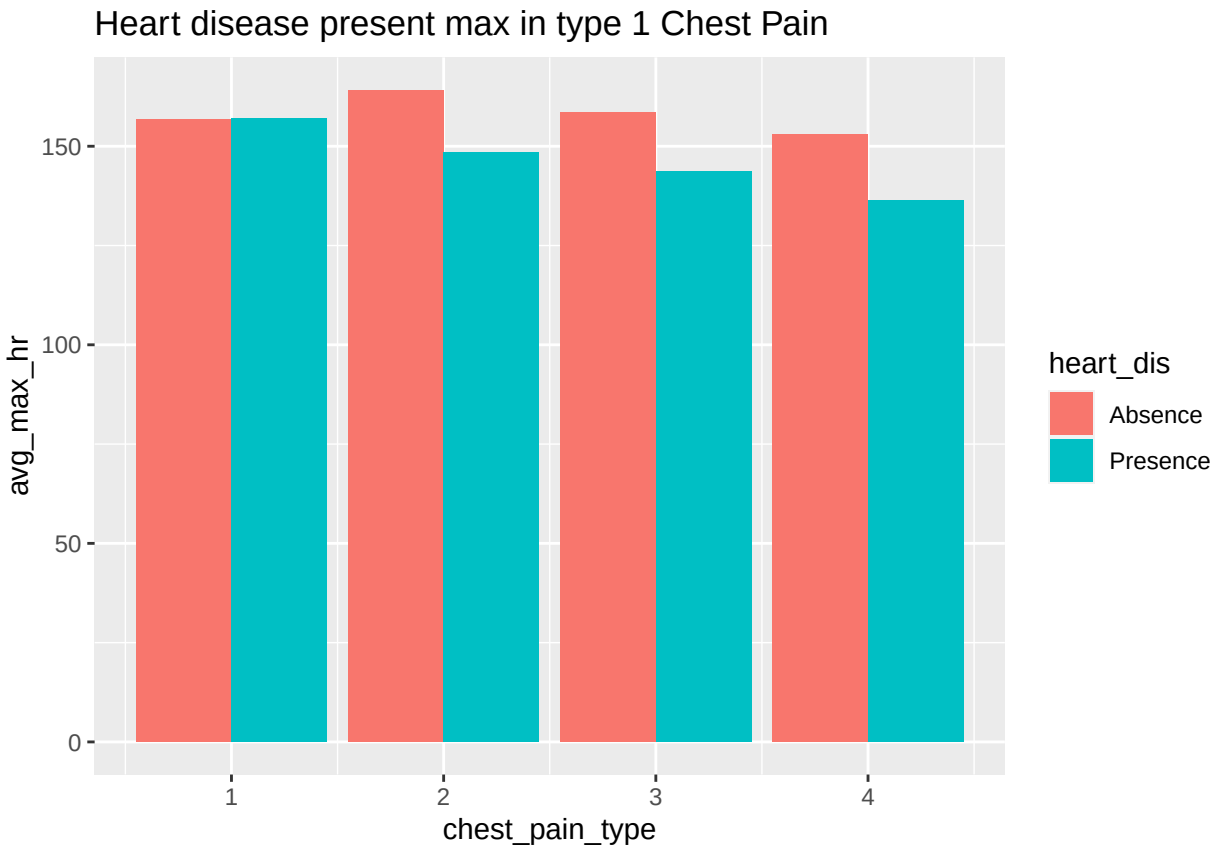


Heart diseases mainly present in 38-70 age groups people and as heart rate without any disease is generally increases in 27-28 age group youngsters. Doing some more

```
heart_disease1 %>%
  group_by(chest_pain_type,heart_dis) %>%
  summarise(avg_max_hr = mean(max_hr)) %>%
  arrange(chest_pain_type,heart_dis) %>%
 ggplot(aes(x = chest_pain_type, y = avg_max_hr, fill = heart_dis)) +
  geom_col(position = "dodge") +
  labs(title = "Heart disease present max in type 1 Chest Pain")
```

```
## `summarise()` has grouped output by 'chest_pain_type'. You can override using
```

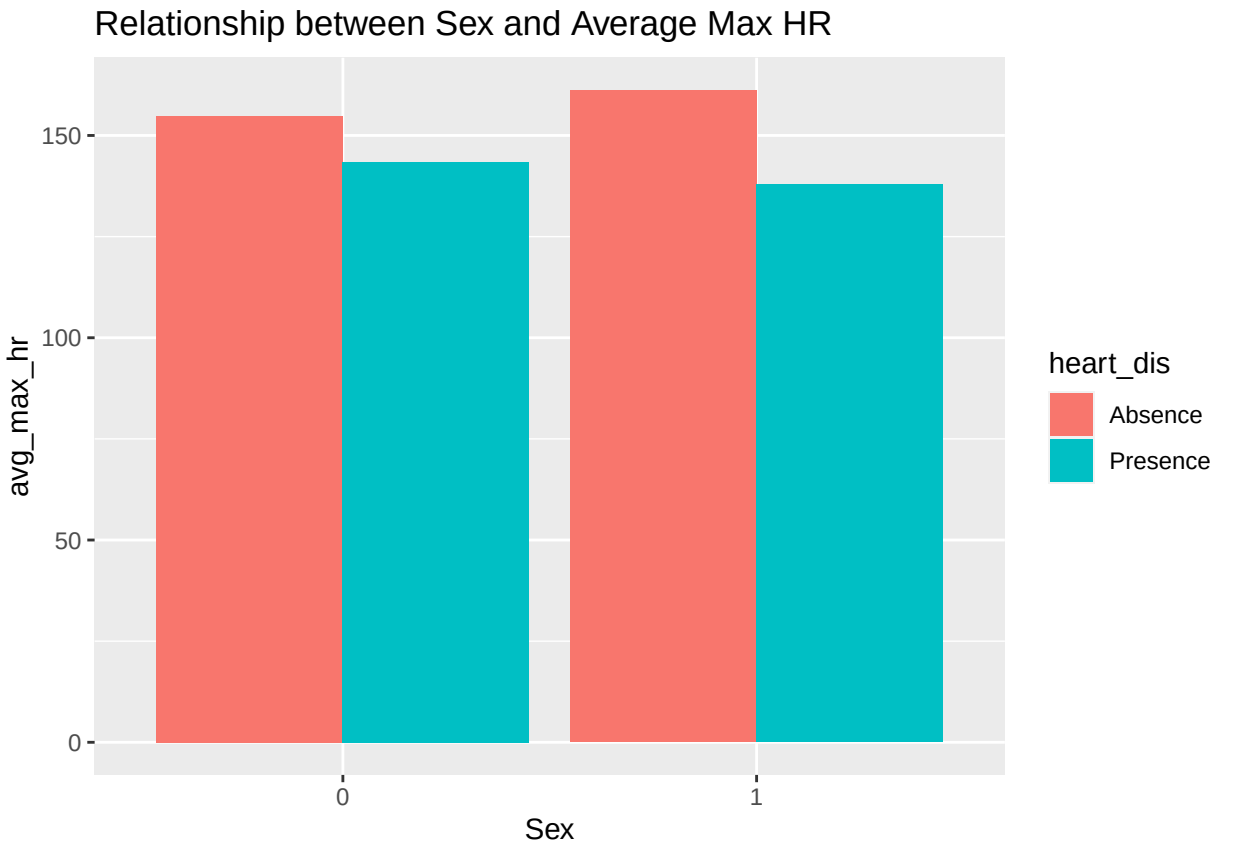
```
## the '.groups' argument.
```



Heart rate generally high in individuals that are having some diseases affecting from type 1 chest pain as compare to other types while type 2 chest pain generally having max high rate without any diseases.

```
heart_disease1 %>%
  group_by(Sex,heart_dis) %>%
  summarise(avg_max_hr = mean(max_hr)) %>%
  arrange(Sex,heart_dis) %>%
 ggplot(aes(x = Sex, y = avg_max_hr, fill = heart_dis)) +
  geom_col(position = "dodge") +
  labs(title = "Relationship between Sex and Average Max HR")
```

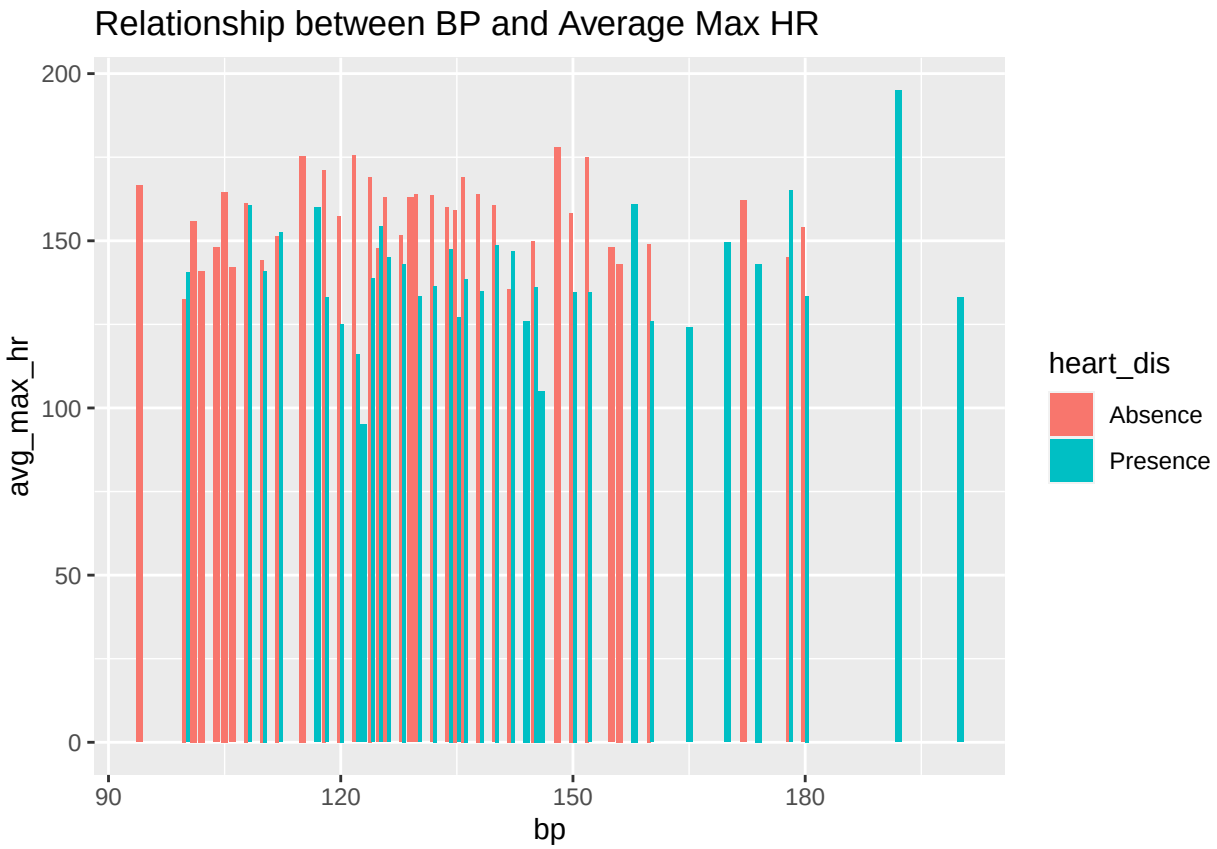
```
## 'summarise()' has grouped output by 'Sex'. You can override using the '.groups'
## argument.
```



Level 1 Sex category having 140 heart rate in individuals that are affected by diseases while in level 2 Sex category having 165 beats in individuals that are free from any diseases.

```
heart_disease1 %>%
  group_by(bp,heart_dis) %>%
  summarise(avg_max_hr = mean(max_hr)) %>%
  arrange(bp,heart_dis) %>%
 ggplot(aes(x = bp, y = avg_max_hr, fill = heart_dis)) +
  geom_col(position = "dodge") +
  labs(title = "Relationship between BP and Average Max HR")
```

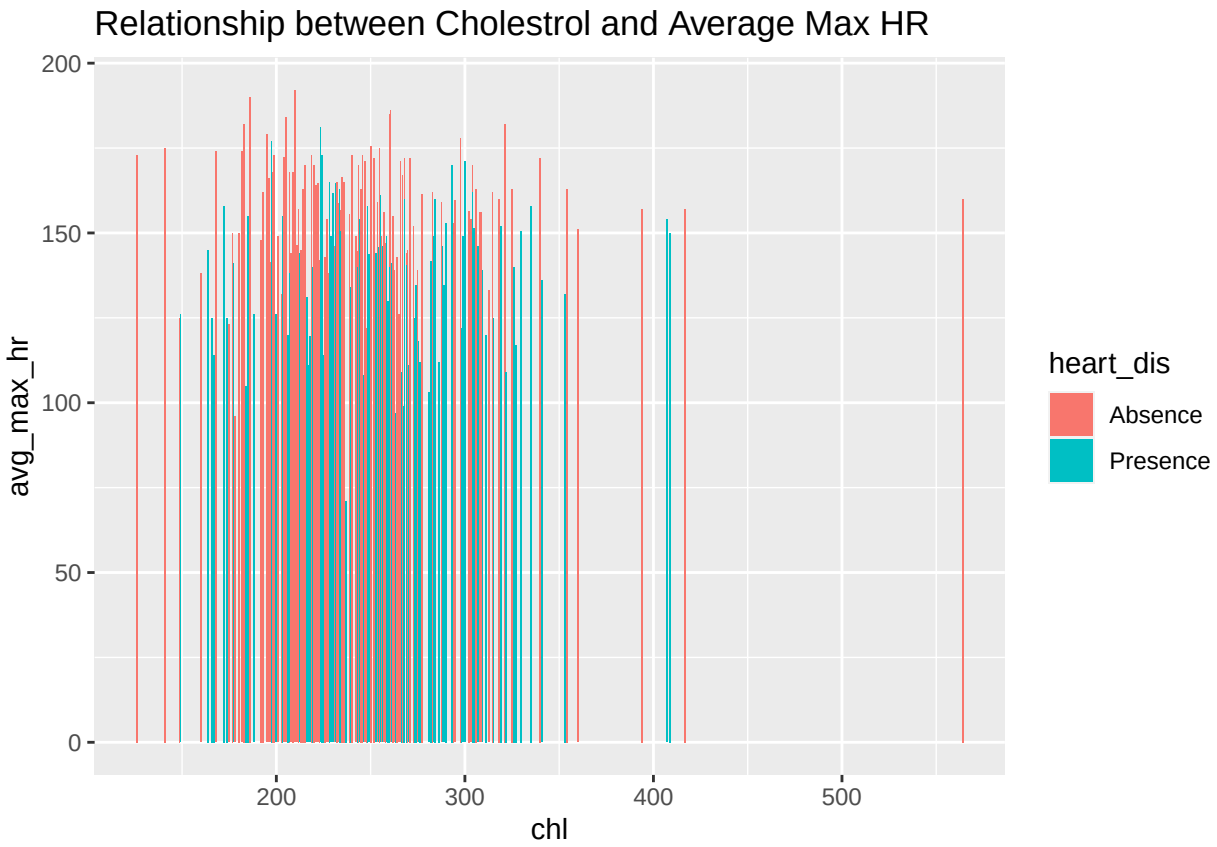
```
## `summarise()` has grouped output by 'bp'. You can override using the `.groups`
## argument.
```



Heart rate generally reaching at high level of the individuals having BP > 190 upto 190 beats.

```
heart_disease1 %>%
  group_by(chl,heart_dis) %>%
  summarise(avg_max_hr = mean(max_hr)) %>%
  arrange(chl,heart_dis) %>%
 ggplot(aes(x = chl, y = avg_max_hr, fill = heart_dis)) +
  geom_col(position = "dodge") +
  labs(title = "Relationship between Cholestrol and Average Max HR")
```

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



Max Heart Disease present in individuals having chl in 200 - 300 range and heart rate reaching maximum upto 180 beats approximately at 220 level of cholestrol.

```
heart_disease1 %>%
  group_by(ST_depression,heart_dis) %>%
  summarise(avg_max_hr = mean(max_hr)) %>%
  arrange(ST_depression,heart_dis) %>%
 ggplot(aes(x = ST_depression, y = avg_max_hr, fill = heart_dis)) +
  geom_col(position = "dodge") +
  labs(title = "Relationship between ST Depression and Average Max HR")
```

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

ST Depression with diseases having max heart rate is of value 3.8 and having heart rate of 170 beats and without disease is of 0.7 and having heart rate of 180 beats.