

Assignment_1

wliu16

2022-09-15

```
setwd("~/Desktop/R/64060/Assignment_1")
```

```
health_data <- read.table("500_Person_Gender_Height_Weight_Index.csv", TRUE, ",")  
str(health_data)
```

```
## 'data.frame':    500 obs. of  4 variables:  
## $ Gender: chr  "Male" "Male" "Female" "Female" ...  
## $ Height: int  174 189 185 195 149 189 147 154 174 169 ...  
## $ Weight: int  96 87 110 104 61 104 92 111 90 103 ...  
## $ Index : int  4 2 4 3 3 3 5 5 3 4 ...
```

```
summary(health_data)
```

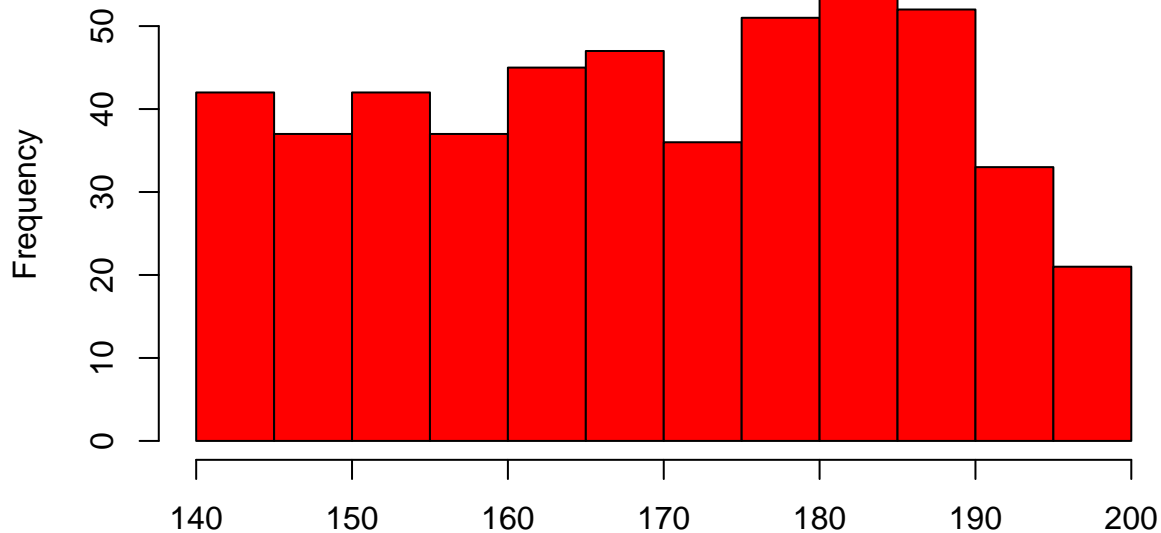
##	Gender	Height	Weight	Index
##	Length:500	Min. :140.0	Min. : 50	Min. :0.000
##	Class :character	1st Qu.:156.0	1st Qu.: 80	1st Qu.:3.000
##	Mode :character	Median :170.5	Median :106	Median :4.000
##		Mean :169.9	Mean :106	Mean :3.748
##		3rd Qu.:184.0	3rd Qu.:136	3rd Qu.:5.000
##		Max. :199.0	Max. :160	Max. :5.000

```
head(health_data)
```

##	Gender	Height	Weight	Index
## 1	Male	174	96	4
## 2	Male	189	87	2
## 3	Female	185	110	4
## 4	Female	195	104	3
## 5	Male	149	61	3
## 6	Male	189	104	3

```
hist(health_data$Height, col = 'red')
```

Histogram of health_data\$Height



health_data\$Height

Comments:

There are over 50 individuals between 180-190cm in height which is the most popular height group in this dataset

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
table(health_data$Index)
```

```
##
##    0    1    2    3    4    5
## 13  22  69  68 130 198
```

```
health_data_arrange1 <- health_data %>% arrange(Height)
View(health_data_arrange1)
```

```
health_data_filter1 <- health_data %>% filter(Index == 3)
View(health_data_filter1)
```

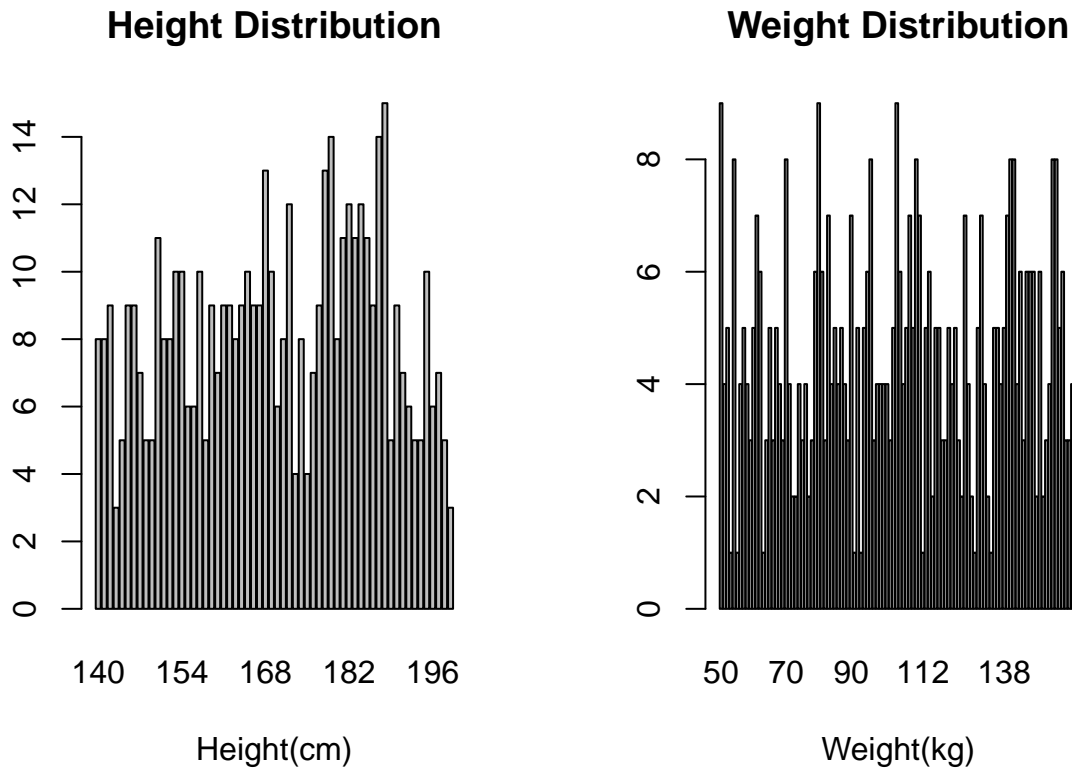
Comments: count the number of individuals by fitness levels 1-5 while 1 as extra weak, 3 as normal and 5 as extra obese arrange a new dataframe by Height column on ascending order filter a new dataframe by fitness level is normal (Index = 3)

```
Height_M <- health_data$Height/100
BMI <- health_data$Weight/(Height_M^2)
head(BMI)
```

```
## [1] 31.70828 24.35542 32.14025 27.35043 27.47624 29.11453
```

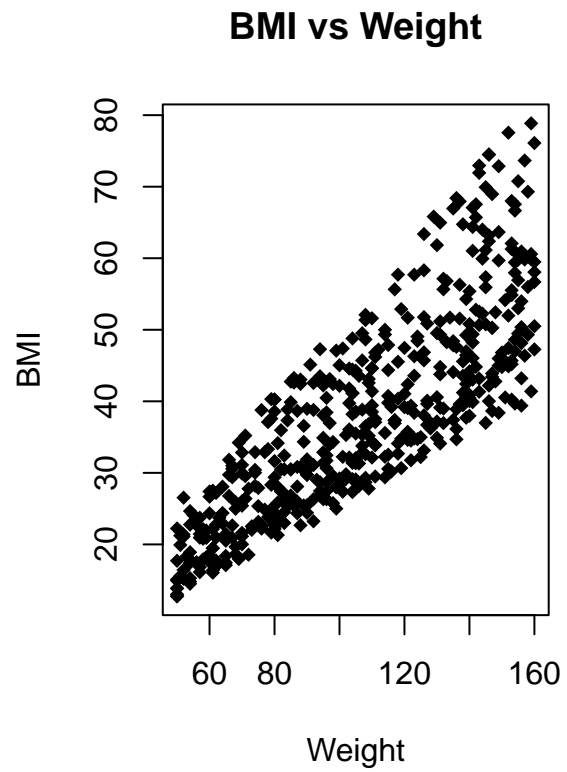
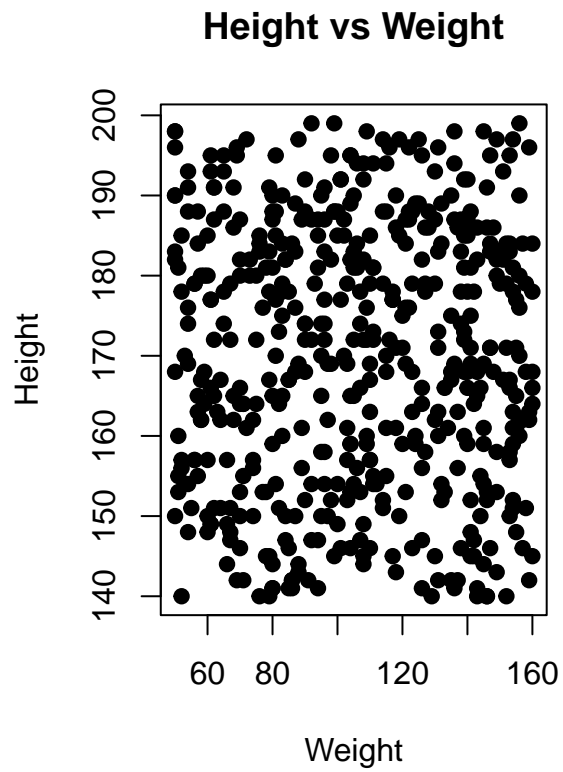
Comments: convert the height to meters, calculate and print first 6 rows of BMI

```
par(mfrow = c(1,2))
counts_h <- table(health_data$Height)
counts_w <- table(health_data$Weight)
barplot(counts_h, main = "Height Distribution", xlab= "Height(cm)")
barplot(counts_w, main = "Weight Distribution", xlab= "Weight(kg)")
```



Comments: plots two bargraphs on the distribution of height and weight. We can see most height is more concentrated towards 180-190cm while weight is more spread out evenly between 50-150 kg.

```
par(mfrow = c(1,2))
plot(health_data$Weight, health_data$Height, main = "Height vs Weight",
     xlab = "Weight", ylab= "Height", pch=19)
plot(health_data$Weight, BMI, main = "BMI vs Weight",
     xlab = "Weight", ylab= "BMI", pch =18)
```



Comments: There is no obvious correlation between height and weight but BMI is positively correlated with weight. The greater the weight is, the higher BMI is.