

R CODES FOR DATA SCIENCE PROJECT

```
DataSc <- read.csv("C:/Users/Iqra/Downloads/data science project.csv")
View(DataSc)
```

```
#installing packages and loading libraries.
```

```
install.packages("tidyverse")
```

```
library(tidyverse)
```

```
install.packages("readxl")
```

```
library(readxl)
```

```
install.packages("dplyr")
```

```
library(dplyr)
```

```
install.packages("purrr")
```

```
library(purrr)
```

```
#datacleaning/transforming
```

```
dim(DataSc)
```

```
DataSc %>% summarise(count = sum(is.na(education_ID)))
```

```
  DataSc %>%
```

```
    summarise(count = sum(is.na(Financial_status)))
```

```
  DataSc%>%
```

```
    summarise(count=sum(is.na(Education_ID)))
```

```
DataSc %>%
```

```
  group_by(Body_Weight) %>%
```

```
  summarise(mean_value = mean(numeric_variable))
```

```
DataS <- na.omit(DataSc)
```

```
dim(DataS)
```

```
#Explatory Data Analysis
```

```
# Summary statistics
```

```
summary(DataS)
```

```
head(DataS)
```

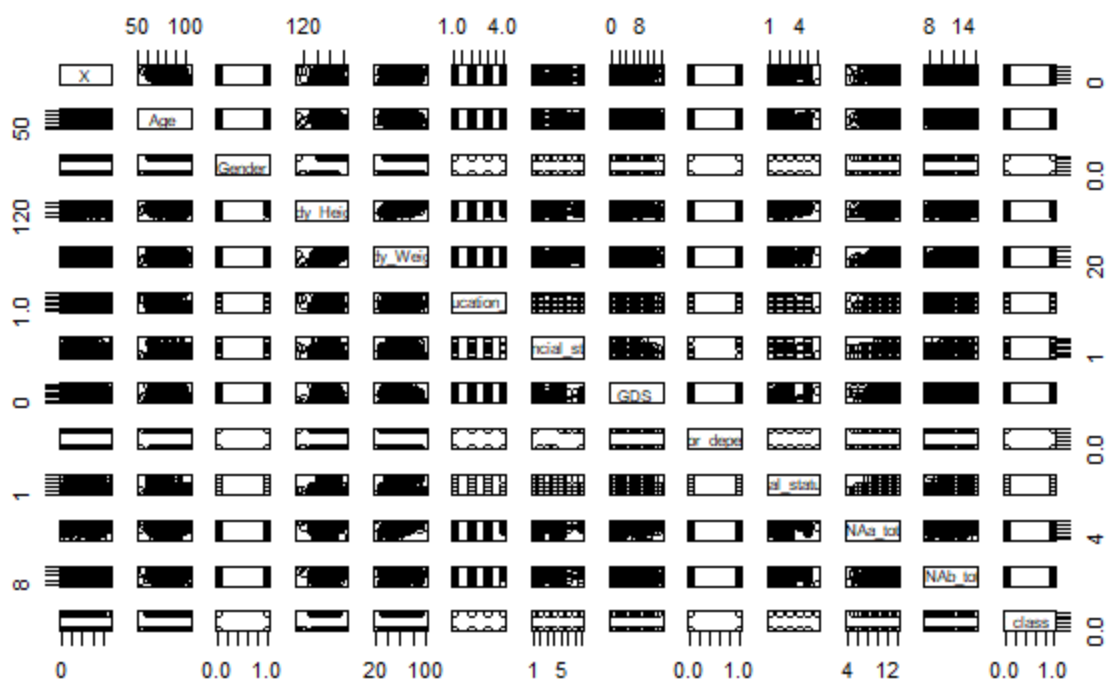
```
tail(DataS)
```

```
# Correlation matrix
```

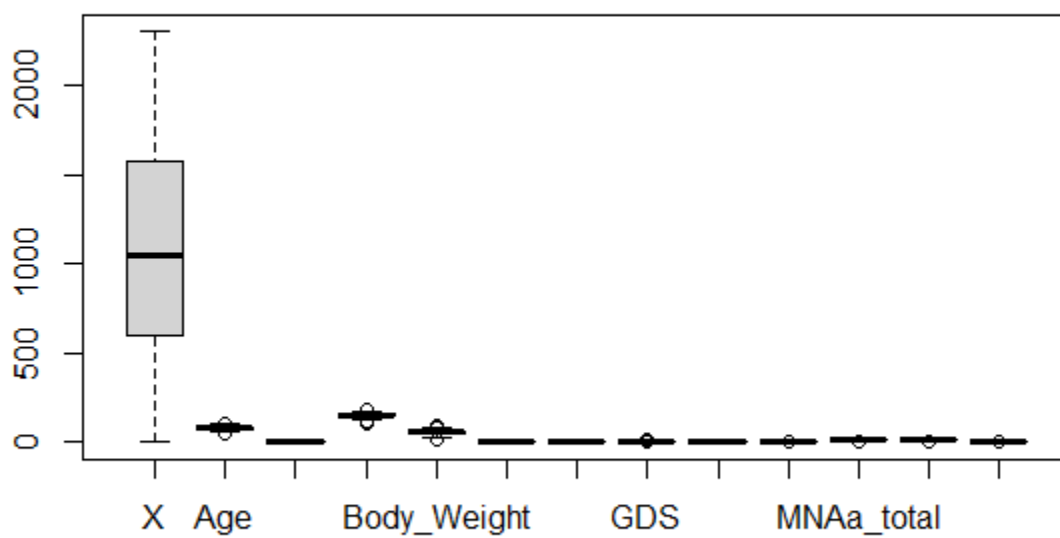
```
cor(DataS)
```

```
# Scatter plot matrix
```

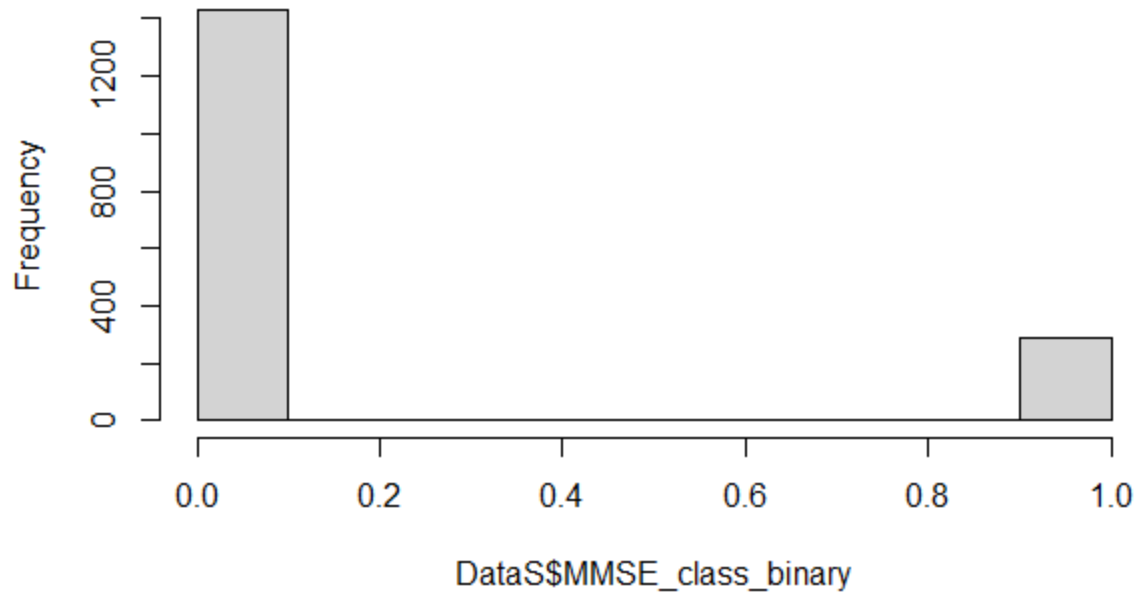
```
pairs(DataS)
```



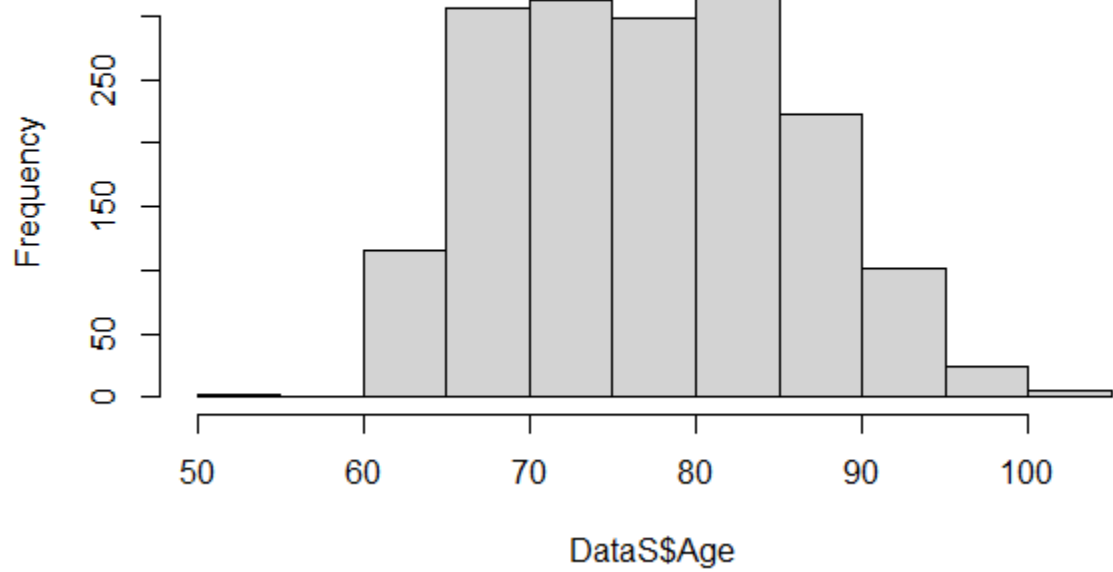
Boxplots
boxplot(DataS)



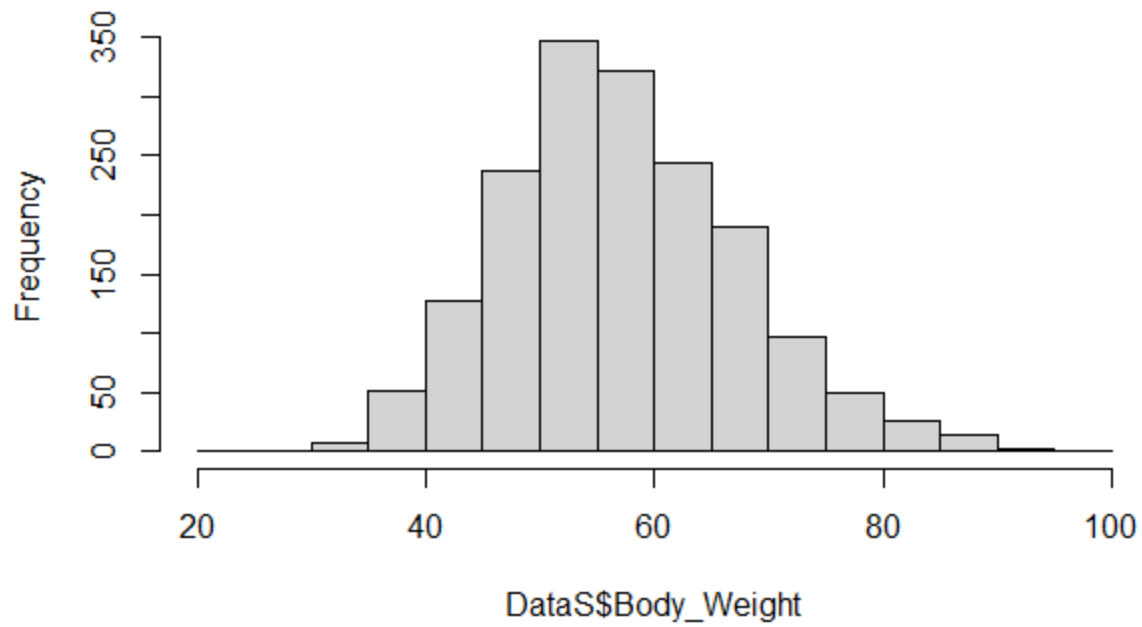
Histogram of Data\$MMSE_class_binary



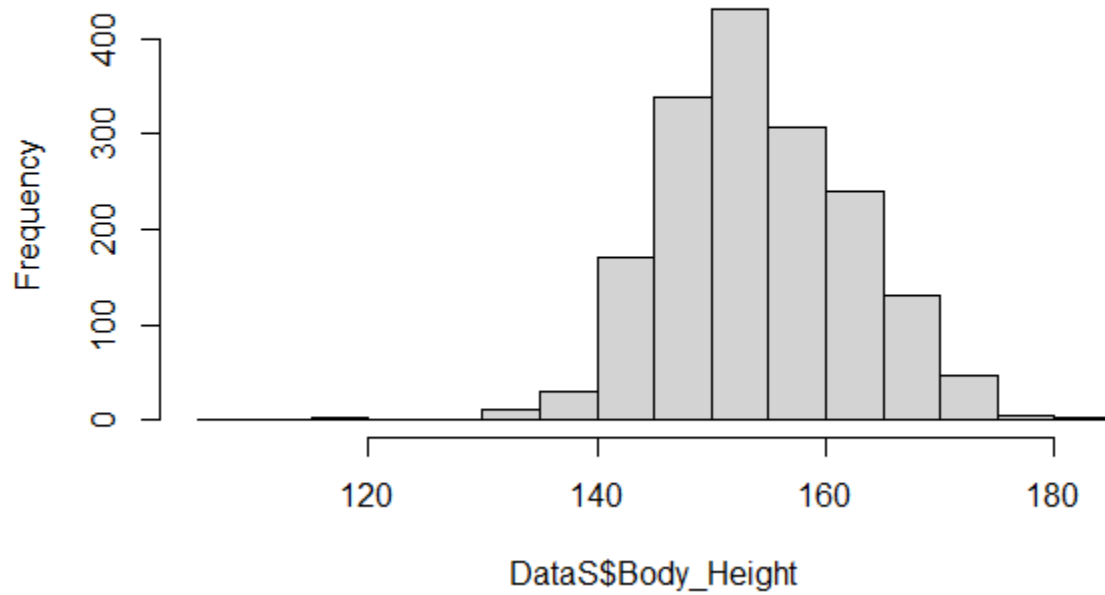
Histogram of Data\$Age



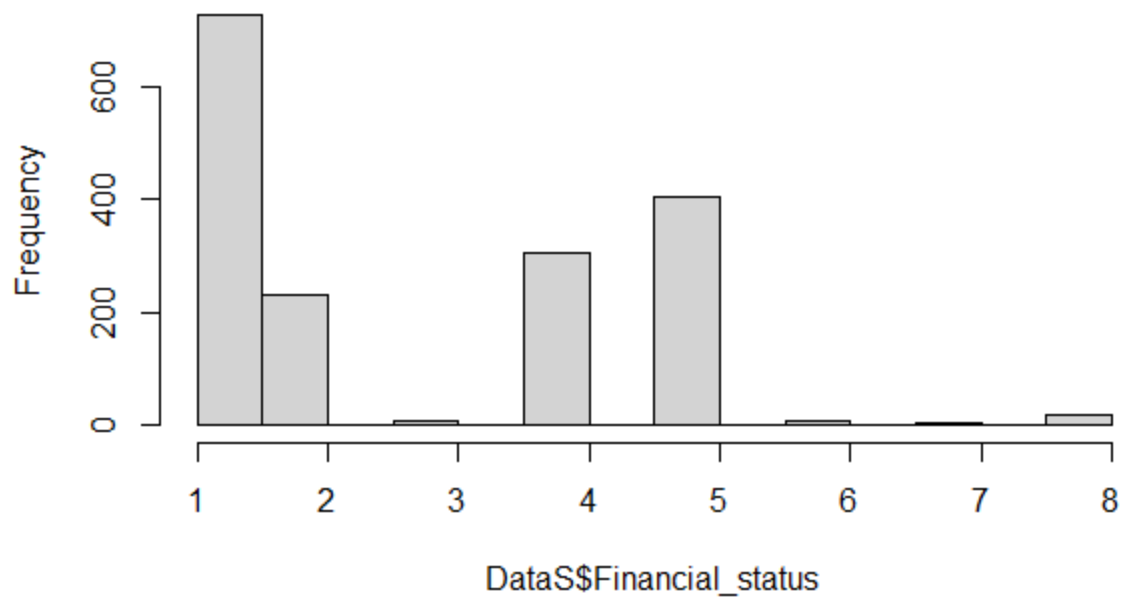
Histogram of Data\$Body_Weight



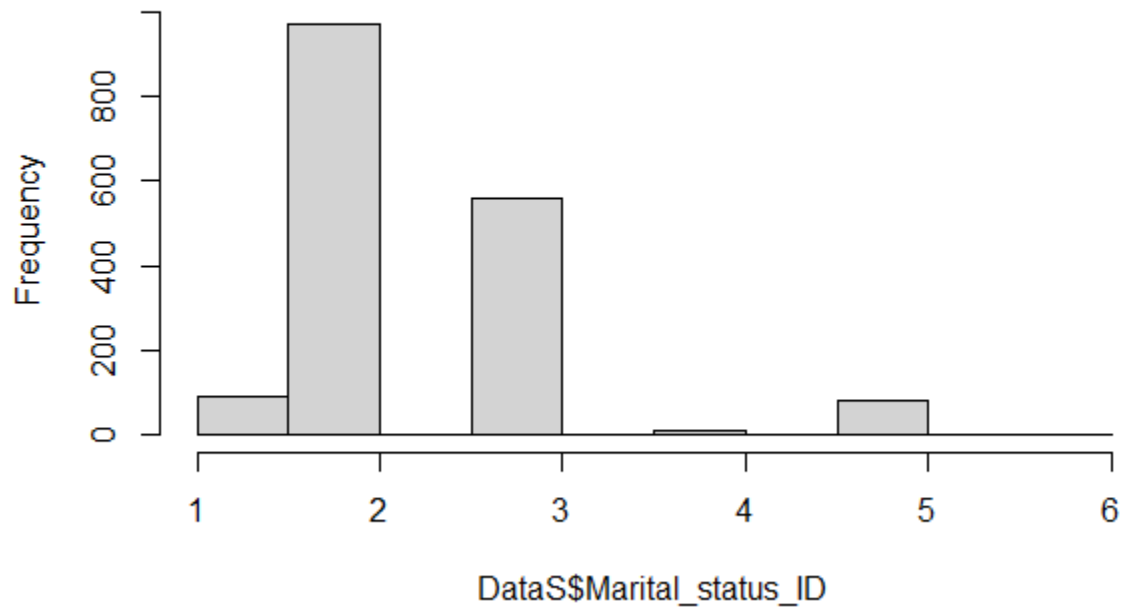
Histogram of Data\$Body_Height



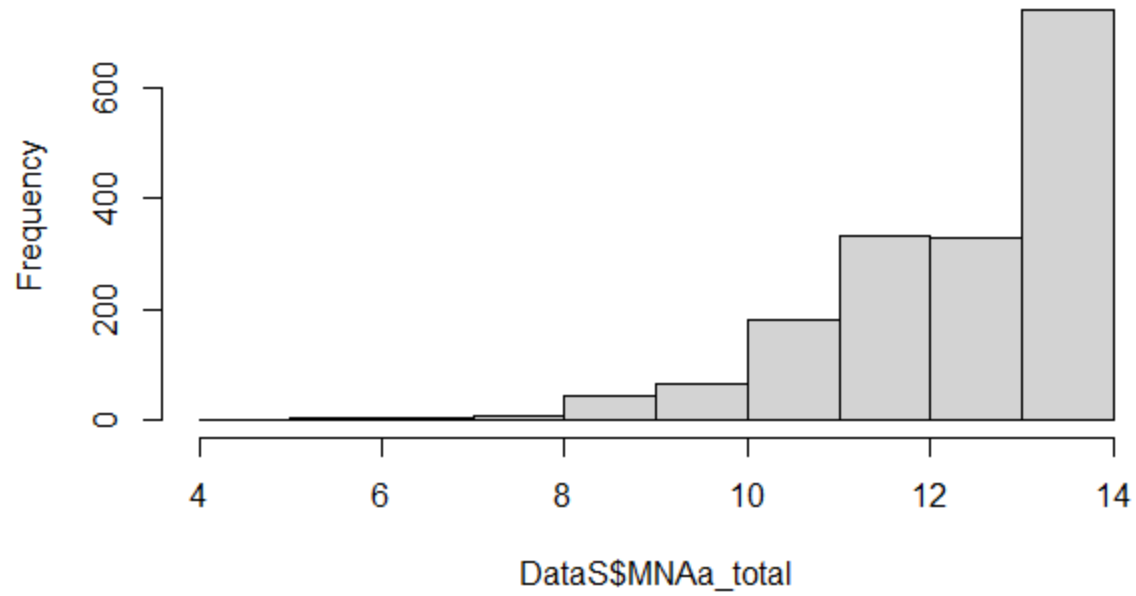
Histogram of Data\$Financial_status



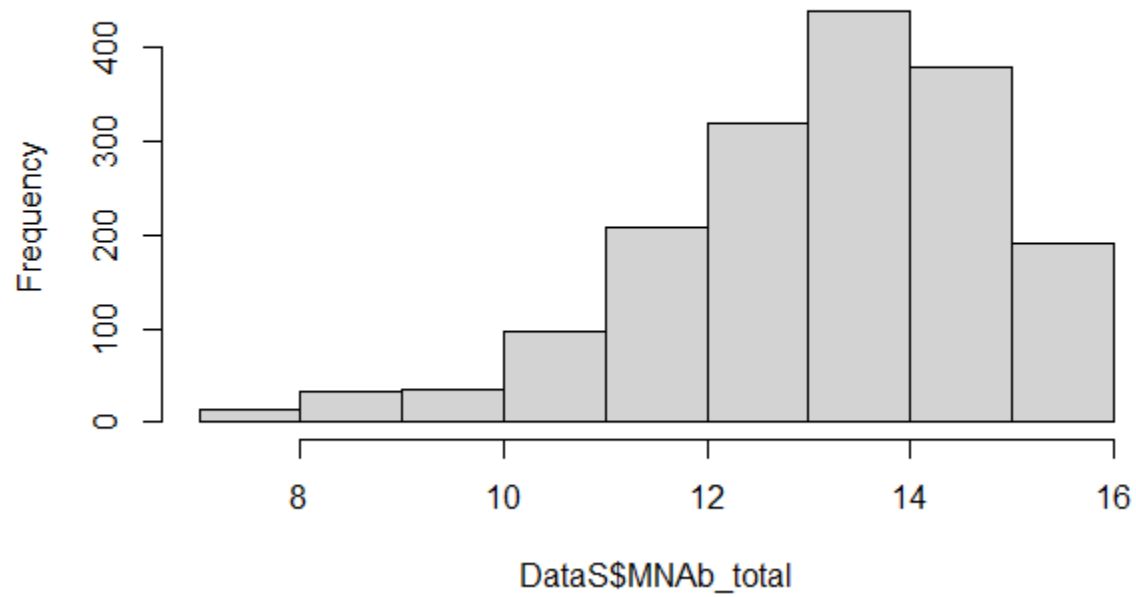
Histogram of Data\$Marital_status_ID



Histogram of Data\$MNAa_total



Histogram of Data\$MNAb_total



```
# Histograms  
hist(Data$MMSE_class_binary)
```

```

hist(DataS$Age)
hist(DataS$Body_Weight)
hist(DataS$Body_Height)
hist(DataS$Financial_status)
hist(DataS$Marital_status_ID)
hist(DataS$MNAa_total)
hist(DataS$MNAb_total)

```

```

# Bar plot
DAge <- DataS$Age
DAge
table(DAge)
barplot(table(DAge))

```

```
table(DAge)
```

```

DAge
 51  55  60  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77
78
  1   1   1   6  11  25  40  33  36  52  70  89  60  76  70  47  58  61  53  59
59
 79  80  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98
99
 58  70  77  67  60  55  64  74  44  42  40  23  29  24  24  18   7  10   4   5
5
100 101 102 104

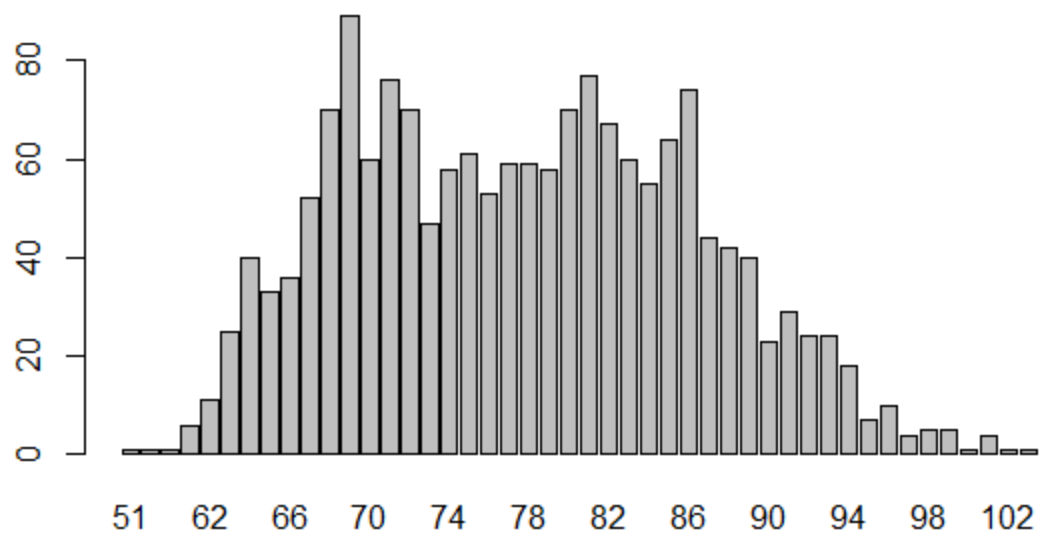
```

```
table(DFinancial)
```

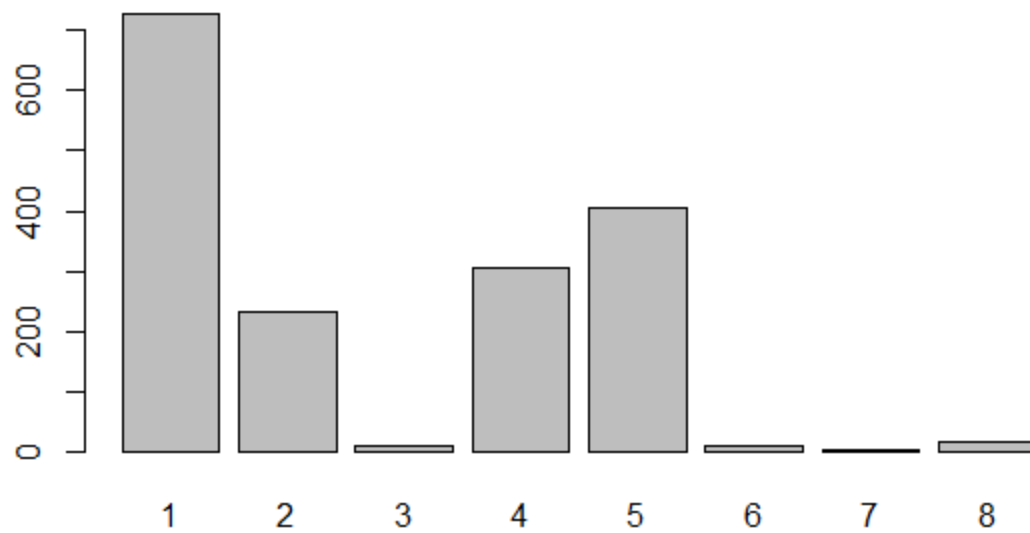
```

DFinancial
  1   2   3   4   5   6   7   8
726 233  10 307 405  10   6  18

```



```
DFinancial <- DataS$Financial_status  
DFinancial  
table(DFinancial)  
barplot(table(DFinancial))
```

```
DMMSE <- DataS$MMSE_class_binary
```

```
DMMSE
```

```
table(DMMSE)
```

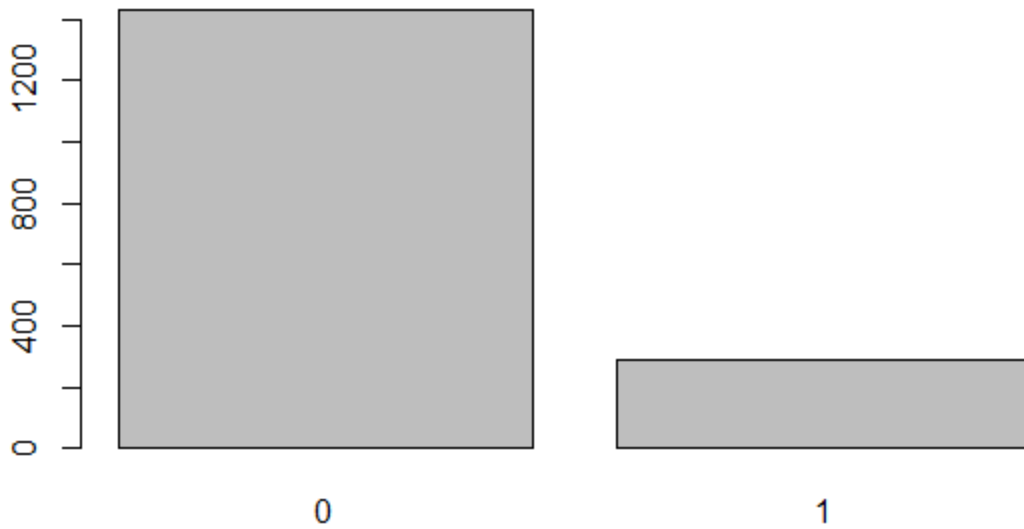
```
barplot(table(DMMSE))
```

```
table(DMMSE)
```

```
DMMSE
```

```
0    1
```

```
1428 287
```



```
DMartial <- DataS$Marital_status_ID
```

```
DMartial
```

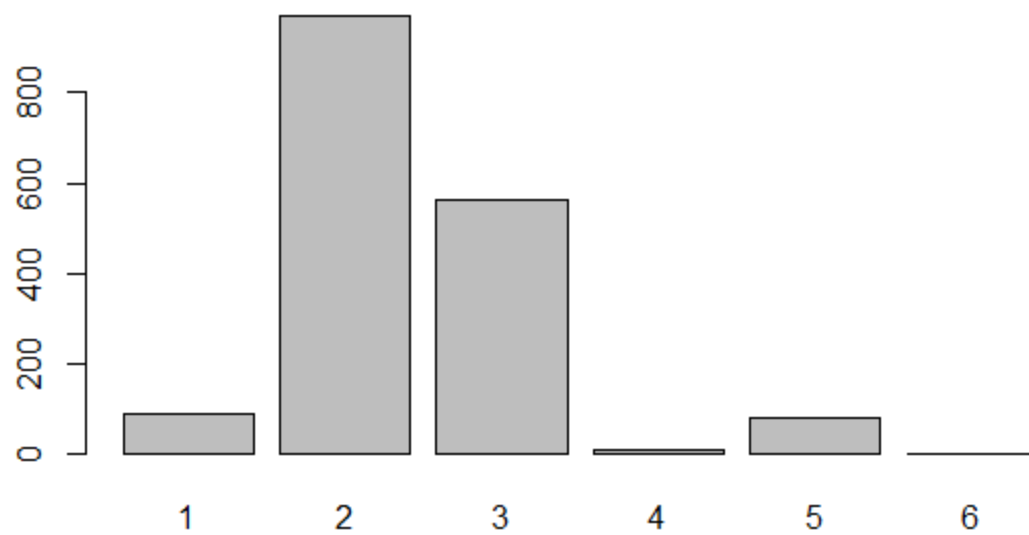
```
table(DMartial)
```

```
barplot(table(DMartial))
```

```
table(DMartial)
```

```
DMartial
```

1	2	3	4	5	6
90	969	562	12	81	1



```
mean(Data$Age)
```

```
[1] 77.73294
```