

ASSIGNMENT1..

MUNERAH

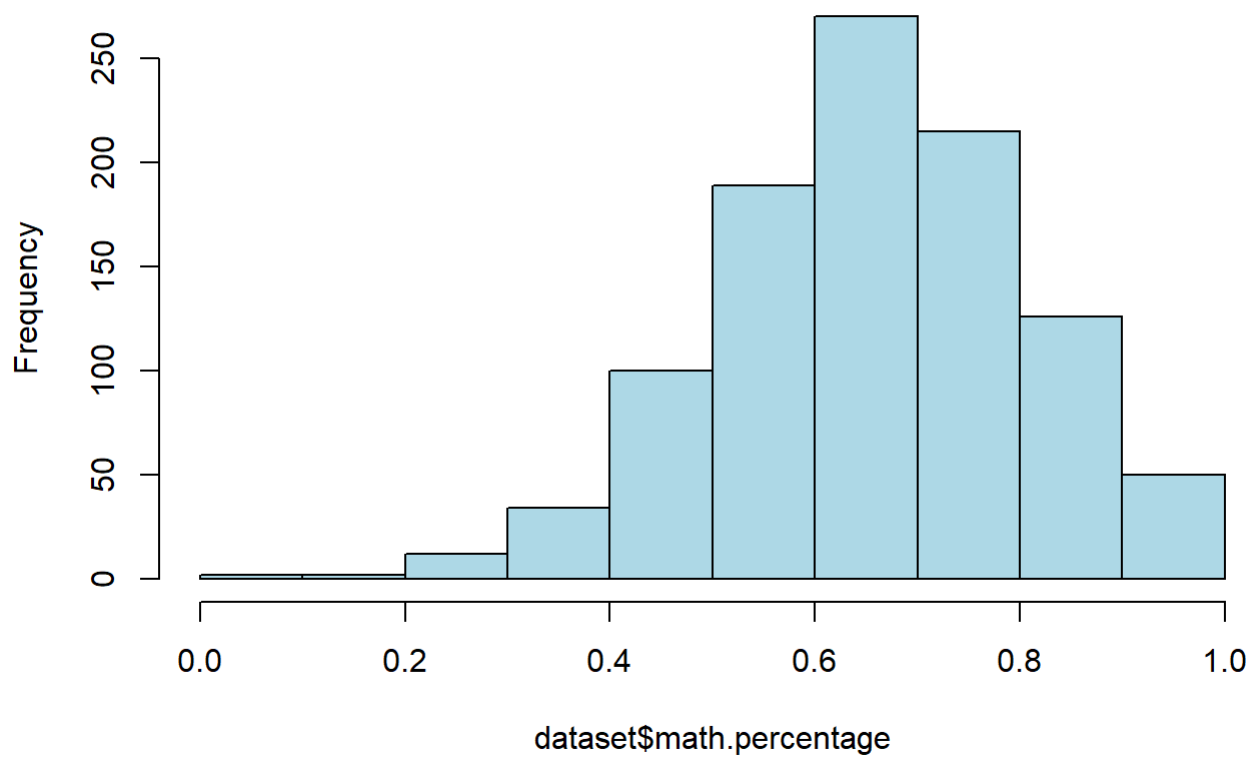
9/24/2021

```
### import the dataset into R
dataset <- read.csv("C:/Users/mnooo/Desktop/Datasets/Student.csv")
## summary
summary(dataset)
```

```
##           X           race.ethnicity      parental.level.of.education
## Min.      : 0.0      Length:1000      Length:1000
## 1st Qu.:249.8      Class :character    Class :character
## Median :499.5      Mode  :character    Mode  :character
## Mean      :499.5
## 3rd Qu.:749.2
## Max.      :999.0
##      lunch      test.preparation.course math.percentage
## Length:1000      Length:1000      Min.      :0.0000
## Class :character  Class :character    1st Qu.:0.5700
## Mode  :character  Mode  :character    Median :0.6600
##                                     Mean      :0.6609
##                                     3rd Qu.:0.7700
##                                     Max.      :1.0000
## reading.score.percentage writing.score.percentage      sex
## Min.      :0.1700      Min.      :0.1000      Length:1000
## 1st Qu.:0.5900      1st Qu.:0.5775      Class :character
## Median :0.7000      Median :0.6900      Mode  :character
## Mean      :0.6917      Mean      :0.6805
## 3rd Qu.:0.7900      3rd Qu.:0.7900
## Max.      :1.0000      Max.      :1.0000
```

```
### Descriptive Statistics for quantitive data
hist(dataset$math.percentage,col = 'light blue')
```

Histogram of dataset\$math.percentage



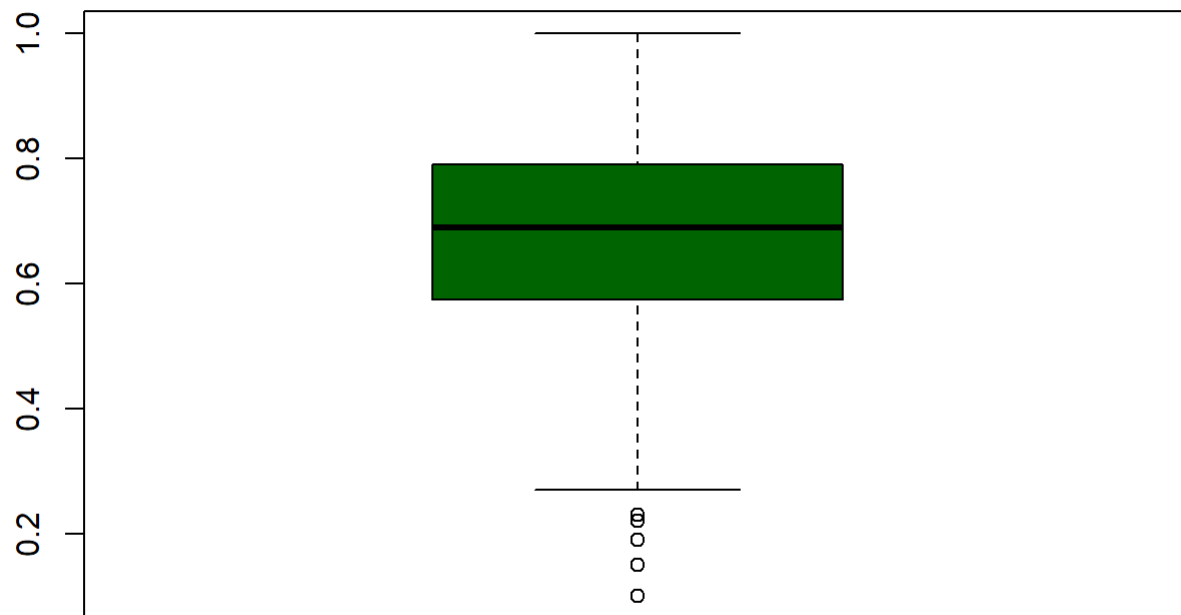
```
mean(dataset$math.percentage)
```

```
## [1] 0.66089
```

```
median(dataset$math.percentage)
```

```
## [1] 0.66
```

```
boxplot(dataset$writing.score.percentage,col = 'dark green')
```



```
###Categorical Variables
class(dataset$sex)
```

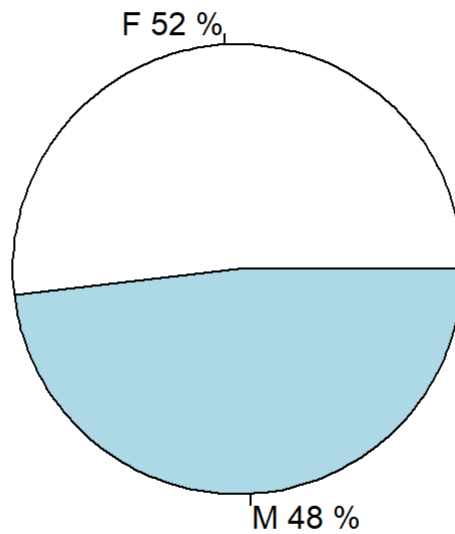
```
## [1] "character"
```

```
table(dataset$sex)
```

```
##
##   F   M
## 518 482
```

```
mypct=round((table(dataset$sex))/(sum(table(dataset$sex)))*100)
lbls=paste(names(table(dataset$sex)),mypct,"%")
pie(table(dataset$sex),lbls,main = "Gender Percentage")
```

Gender Percentage



```
###variable transformation
```

```
## adding new colmn for the GPA
library(tidyverse)
```

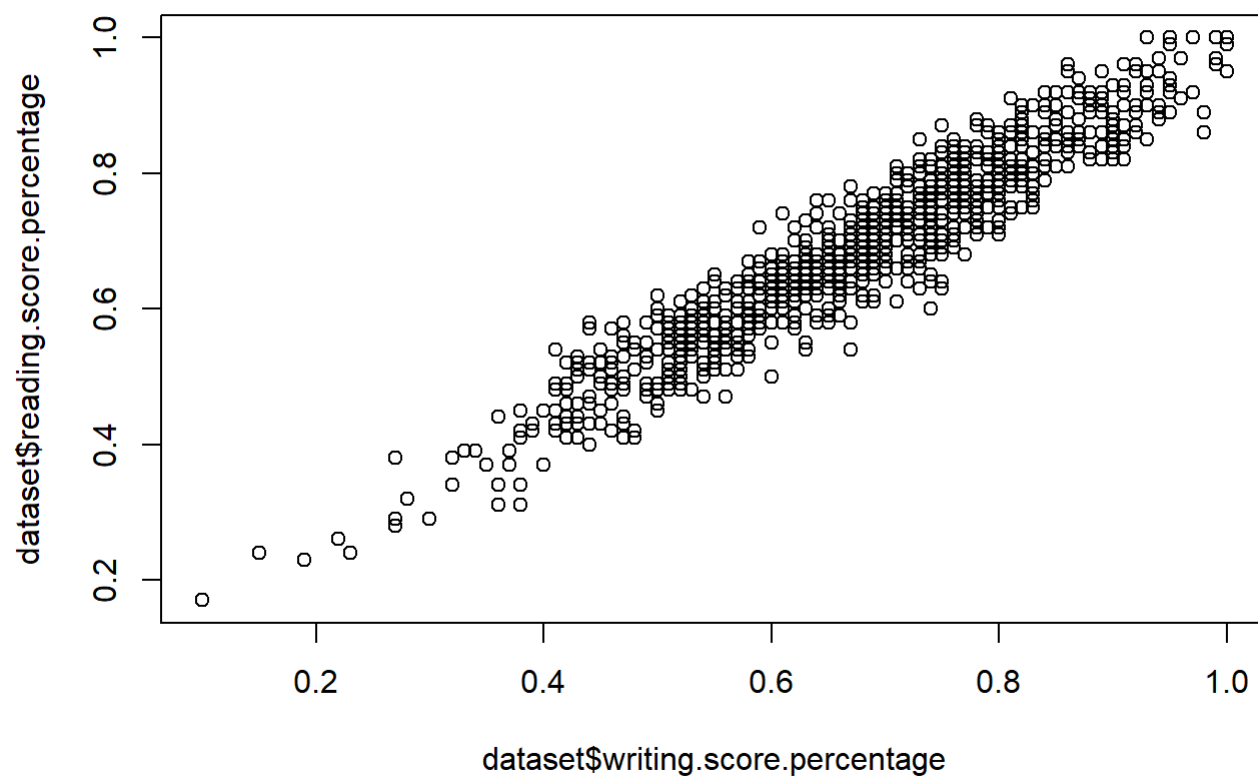
```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.4      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1
```

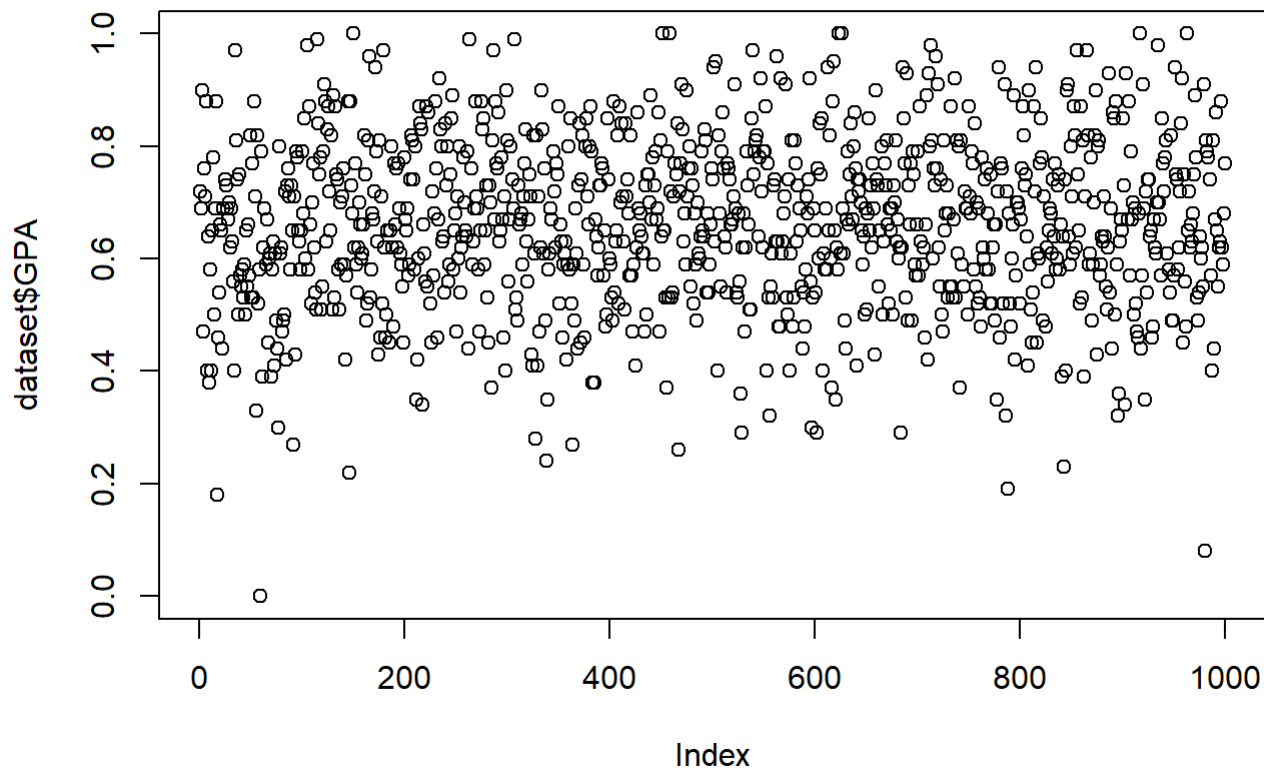
```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
dataset_mutate <- dataset %>% mutate(GPA = (dataset$math.percentage + dataset$reading.score.perc
centage + dataset$writing.score.percentage)/3)
View(dataset_mutate)
```

```
### Plot
plot(dataset$writing.score.percentage,dataset$reading.score.percentage,col='black')
```



```
plot(dataset$math.percentage,dataset$GPA,col='black')
```



SOURCE OF THE DATA SET

<https://www.kaggle.com/spscientist/students-performance-in-exams> (<https://www.kaggle.com/spscientist/students-performance-in-exams>)