

FML ASSIGNMENT_1 RAJESH

```
#Started by importing the dataset from an Excel file.
```

```
#I have obtained my dataset from kaggle(Global YouTube Statistics [https://www.kaggle.com/datasets/nelgiriwithana/global-youtube-statistics-2023]).
```

```
library(readxl)
```

```
data1 <- read_excel("C:\\Users\\yadla sreebhavya\\Downloads\\GYS\\GlobalYouTubeStatistics.xlsx")
```

```
# removed rows with missing values (NAs) from the dataset using the na.omit()
```

```
data1 <- na.omit(data1)
```

```
#generated summary statistics for the quantitative variables and categorical variables in the dataset.
```

```
summary(data1)
```

```
##      Youtuber      subscribers      video views      Title
## Length:848      Min.   : 12300000      Min.   :0.000e+00      Length:848
## Class :character 1st Qu.: 14500000      1st Qu.:4.295e+09      Class :character
## Mode  :character Median : 17600000      Median :7.762e+09      Mode  :character
##                      Mean  : 22962618      Mean   :1.137e+10
##                      3rd Qu.: 24100000      3rd Qu.:1.383e+10
##                      Max.   :245000000      Max.   :2.280e+11
##      Country      channel_type      video_views_rank
## Length:848      Length:848      Min.   :      1
## Class :character Class :character 1st Qu.:     281
## Mode  :character Mode  :character Median :     783
##                      Mean   : 174422
##                      3rd Qu.:  2113
##                      Max.   :4057901
## video_views_for_the_last_30_days lowest_yearly_earnings
## Min.   :1.000e+00      Min.   :      0
## 1st Qu.:2.587e+07      1st Qu.:   69050
## Median :7.149e+07      Median :  197450
## Mean   :1.827e+08      Mean   :  483875
## 3rd Qu.:1.754e+08      3rd Qu.:  504000
## Max.   :6.589e+09      Max.   :10200000
## highest_yearly_earnings
## Min.   :      0
## 1st Qu.: 1100000
## Median : 3150000
## Mean   : 7747703
## 3rd Qu.: 8100000
## Max.   :163400000
```

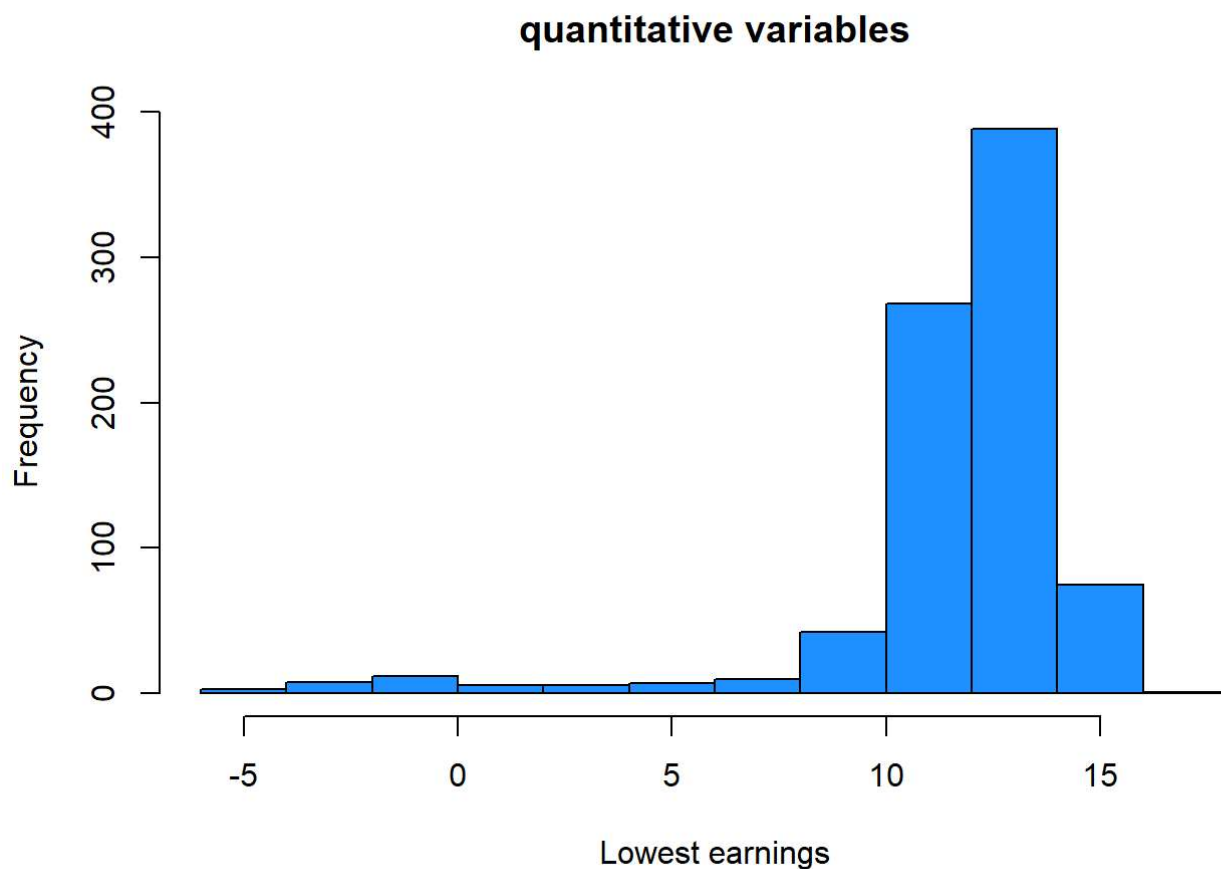
```
#Transforming a Numeric Variable (Log Transformation)
```

```
data1$lowest_yearly_earnings <- log(data1$lowest_yearly_earnings)  
head(data1$lowest_yearly_earnings)
```

```
## [1] 15.732433 -3.218876 15.201805 15.590463 15.520259 14.603968
```

```
#Histogram for "lowest_yearly_earnings"
```

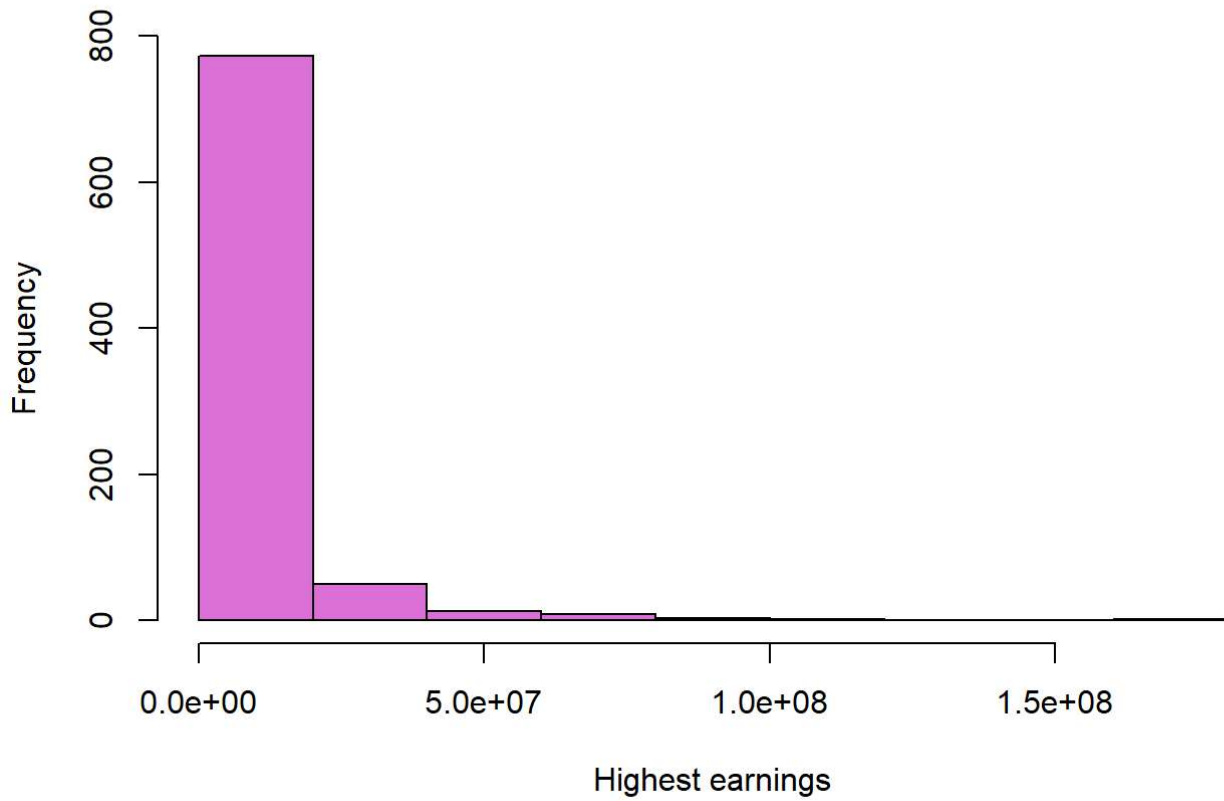
```
hist(data1$lowest_yearly_earnings, main = "quantitative variables", xlab = "Lowest earnings",  
col = 'dodgerblue')
```



```
# Histogram for "highest_yearly_earnings" (Similar to Chunk 6)
```

```
hist(data1$highest_yearly_earnings, main = "quantitative variables", xlab = "Highest earnings",  
col = 'orchid')
```

quantitative variables



#Scatterplot for Subscribers vs. Views

```
plot(data1$subscribers, data1$`video views`, main = "Scatterplot of Subscribers vs No. of Views", xlab = "subscribers", ylab = "Views", col = "firebrick", pch = 18)
```

Scatterplot of Subscribers vs No. of Views

