FML ASSIGNMENT MANOJ

```
#I have taken my dataset from kaggle(Most streamed spotify songs 2023 {https://www.kaggle.co
m/datasets/nelgiriyewithana/top-spotify-songs-2023})
#Imported my dataset from excel file
library(readxl)
newdata <- read_excel("C:\\Users\\yadla sreebhavya\\Downloads\\spotify data 20232.xlsx")
```

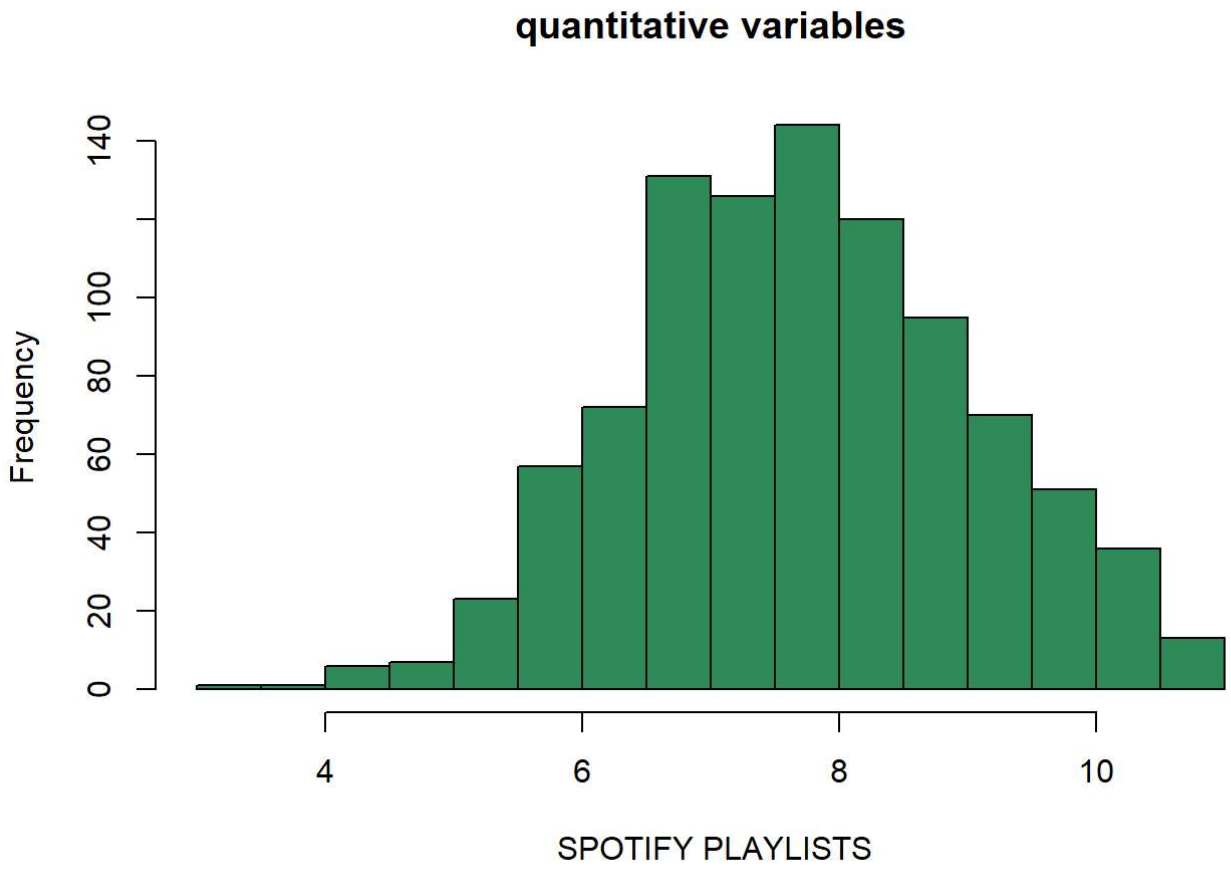
```
#Produced summary statistics for both the numerical and categorical variables within the data
set.
summary(newdata)
```

```
##   track_name        artist(s)_name      artist_count   released_year
##  Length:953         Length:953         Min.   :1.000   Min.   :1930
##  Class :character   Class :character   1st Qu.:1.000   1st Qu.:2020
##  Mode  :character   Mode  :character   Median :1.000   Median :2022
##                                        Mean   :1.556   Mean   :2018
##                                        3rd Qu.:2.000   3rd Qu.:2022
##                                        Max.   :8.000   Max.   :2023
##  in_spotify_playlists   streams          in_apple_playlists in_deezer_playlists
##  Min.   :   31        Length:953         Min.   :  0.00     Min.   :    0.0
##  1st Qu.:  875        Class :character   1st Qu.: 13.00     1st Qu.:   13.0
##  Median : 2224        Mode  :character   Median : 34.00     Median :   44.0
##  Mean   : 5200                           Mean   : 67.81     Mean   :  385.2
##  3rd Qu.: 5542                           3rd Qu.: 88.00     3rd Qu.:  164.0
##  Max.   :52898                           Max.   :672.00     Max.   :12367.0
##      mode
##  Length:953
##  Class :character
##  Mode  :character
##
##
##
```

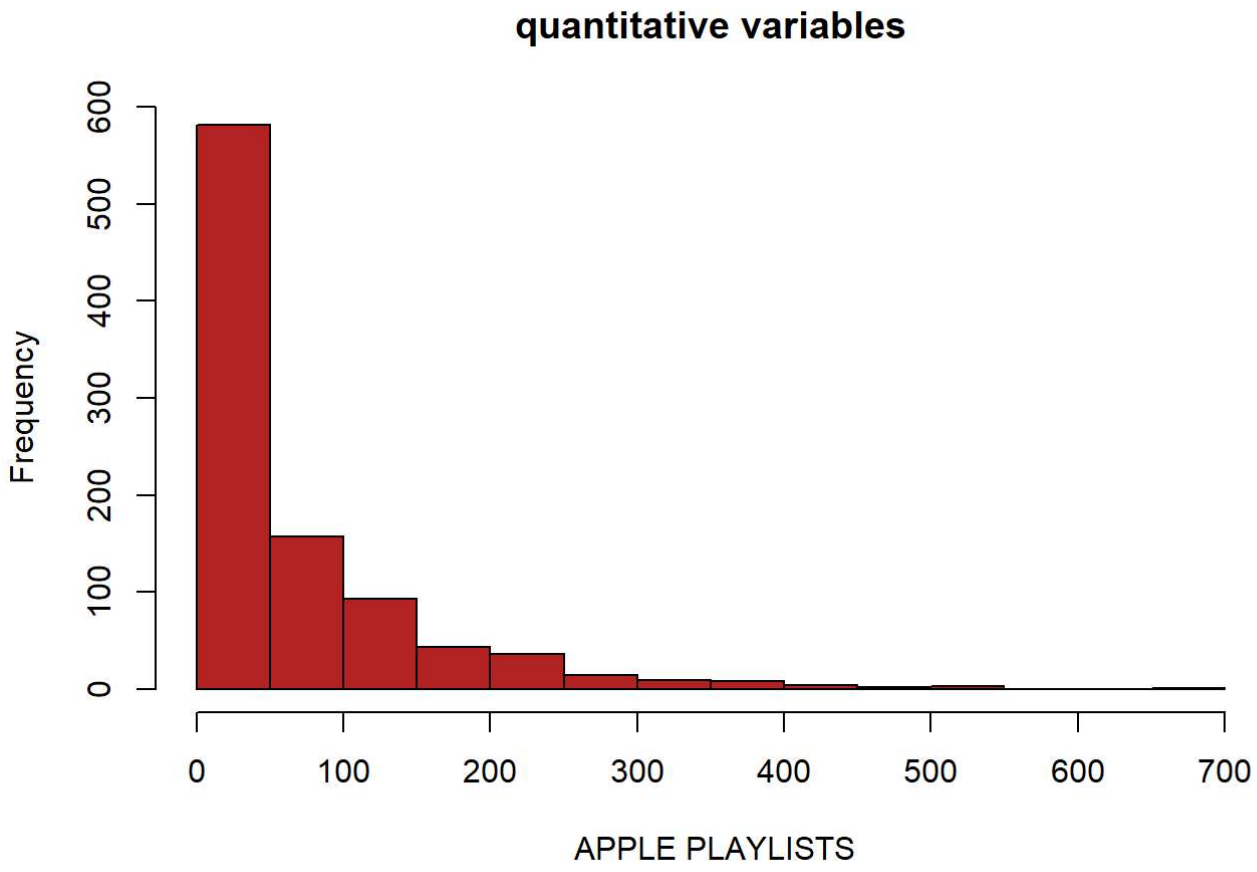
```
#Transforming a Numeric Variable using log Transformation

newdata$in_spotify_playlists <- log(newdata$in_spotify_playlists)
head(newdata$in_spotify_playlists)
```

```
## [1] 6.315358 7.295735 7.242082 8.969287 8.049746 7.689829
```

```
#Histogram for "SPOTIFY PLAYLISTS"

hist(newdata$in_spotify_playlists, main = "quantitative variables", xlab = " SPOTIFY PLAYLIST
S", col ='seagreen')
```

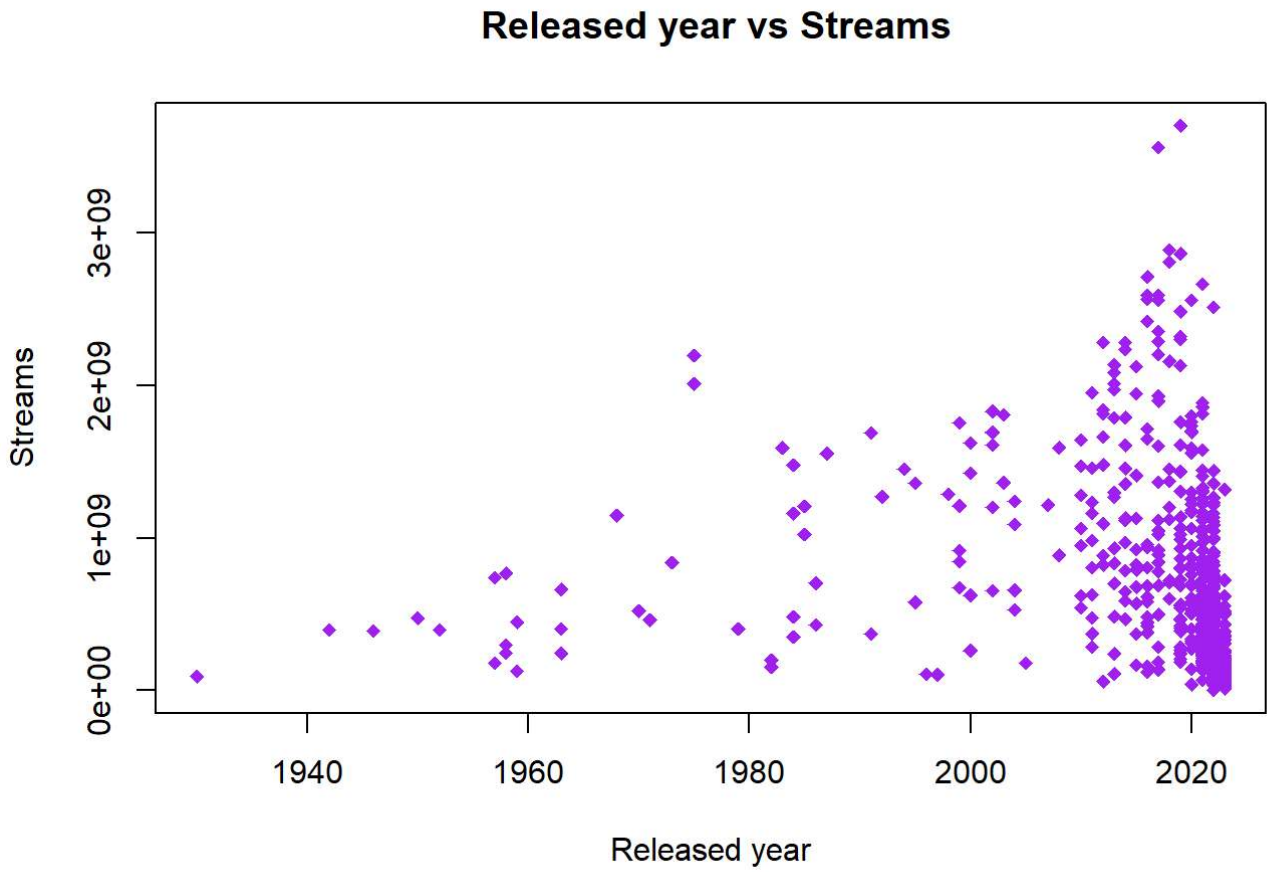



```
 #Histogram for "APPLE PLAYLISTS"

hist(newdata$in_apple_playlists, main = "quantitative variables", xlab = " APPLE PLAYLISTS",
col ='firebrick')
```




```
#Scatterplot for Released Year vs. Streams

plot(newdata$released_year,newdata$streams ,main = " Released year vs Streams", xlab = "Relea
sed year", ylab = "Streams", col = "purple", pch = 18)
```

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): NAs introduced by coercion
```


Released year vs Streams
Streams
0e+00
1e+09
2e+09
3e+09
1940
1960
1980
2000
2020
Released year