BA 64060-FML Assignment 1

2023-09-09

I have downloaded the data set from Kaggle and below is the URL for source data

# URL for the dataset

[**https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023**](https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023)

With reference of source data I have Modified the data which is required for now

spotify <- read.csv("spotify-2023\_1.csv", nrows = 50, header=TRUE, sep = ",")  
  
spotify <- na.omit(spotify)  
  
head(spotify)

## artists\_name released\_year released\_month released\_day spty\_plists  
## 1 Myke Towers 2023 3 23 1474  
## 2 Olivia Rodrigo 2022 6 30 1397  
## 3 Taylor Swift 2019 8 23 7858  
## 4 Bad Bunny 2018 5 18 3133  
## 5 Dave, Central Cee 2015 6 1 2186  
## 6 Eslabon Armado 2000 3 16 3090  
## apple\_plist deezer\_plist bpm mode  
## 1 48 58 92 Major  
## 2 94 91 138 Major  
## 3 116 125 170 Major  
## 4 84 87 144 Minor  
## 5 67 88 141 Major  
## 6 34 43 148 Minor

# Descriptive statistics for quantitative variables

quantative\_vars <- spotify[, c("released\_day", "bpm")]  
  
summary(quantative\_vars)

## released\_day bpm   
## Min. : 1.00 Min. : 67.0   
## 1st Qu.: 8.25 1st Qu.:100.0   
## Median :18.50 Median :125.5   
## Mean :16.84 Mean :125.9   
## 3rd Qu.:24.00 3rd Qu.:143.2   
## Max. :31.00 Max. :204.0

# Descriptive statistics for categorical variables

Categorical\_var <- spotify[, c("released\_year", "released\_month")]  
  
table(Categorical\_var)

## released\_month  
## released\_year 1 2 3 4 5 6 7 8 9 10 11 12  
## 1999 0 0 0 0 0 0 1 0 0 0 0 0  
## 2000 0 0 1 0 0 0 0 0 0 0 0 0  
## 2005 0 0 0 1 1 0 0 0 0 0 0 0  
## 2007 0 0 1 0 0 0 0 0 0 0 0 0  
## 2013 1 0 0 0 0 0 0 0 0 0 0 0  
## 2014 2 0 0 0 0 0 0 0 0 0 0 0  
## 2015 0 0 0 0 0 1 0 0 0 0 0 0  
## 2016 0 0 0 0 0 0 0 0 1 0 1 0  
## 2017 0 0 0 0 0 0 1 0 0 0 0 0  
## 2018 0 0 0 0 1 0 0 0 0 1 0 0  
## 2019 0 0 0 0 0 0 0 1 0 0 0 0  
## 2020 0 0 0 0 0 0 0 0 0 0 1 0  
## 2022 0 0 2 0 1 1 0 1 0 1 0 2  
## 2023 2 4 3 2 3 8 4 0 0 0 0 0

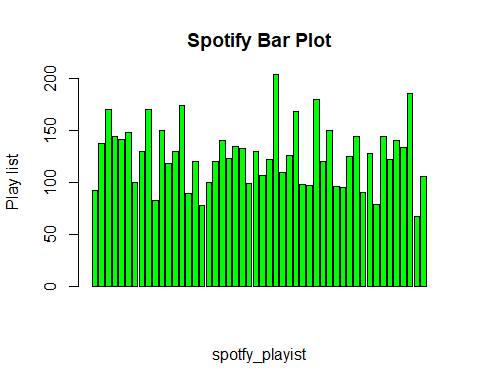
# Transform Spotify\_Playlist to Rank wise

log\_rank <- log(spotify$spty\_plists)  
  
print(log\_rank)

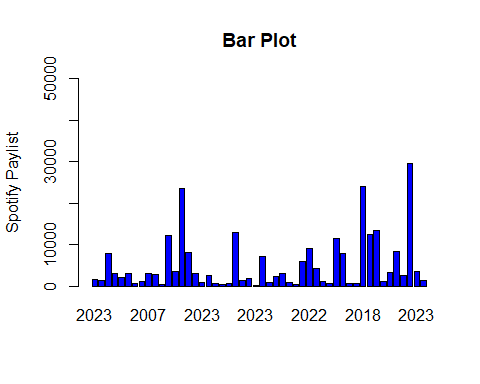
## [1] 7.295735 7.242082 8.969287 8.049746 7.689829 8.035926 6.570883  
## [8] 6.999422 7.990577 7.964156 6.045005 9.410092 8.168486 10.067942  
## [15] 9.000730 7.986845 6.771936 7.867106 6.390241 5.805135 6.246107  
## [22] 9.461799 7.180070 7.573017 5.521461 8.869539 6.755769 7.791523  
## [29] 8.002360 6.761573 5.583496 8.705497 9.114050 8.362642 7.047517  
## [36] 6.510258 9.344347 8.965718 6.369901 6.539586 10.089718 9.432043  
## [43] 9.502039 7.018402 8.123261 9.039671 7.817223 10.293365 8.133881  
## [50] 7.168580

# Bar Plot

# Create a bar plot  
  
barplot(spotify$bpm,  
 xlab = 'spotfy\_playist', ylim = c(0,200), ylab = 'Play list',  
 main = 'Spotify Bar Plot', col = 'green')



barplot(spotify$spty\_plists, names.arg = spotify$released\_year, col = "blue", main = "Bar Plot", ylab = "Spotify Paylist", ylim = c(0,50000),  
 num\_ticks <- 5)



# Scatter Polt

plot(spotify$released\_year, spotify$deezer\_plist,   
 type = "p",   
 col = "blue",   
 main = "Scatter Plot",   
 xlab = "Released Year",   
 ylab = "Deezer Playlist")

