

# Assignment\_1

2023-09-09

*##Dataset used is Spotify Music Data, retrieved from Kaggle. The dataset details the top songs from 2010-2019, including artist + title, genre, year, and quantitative song details such as BPM.*

```
install.packages("ggplot2", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/13308/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)

## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\13308\AppData\Local\Temp\RtmpIh1dQL\downloaded_packages

library("ggplot2")

## Warning: package 'ggplot2' was built under R version 4.2.3

library(readr)

## Warning: package 'readr' was built under R version 4.2.3

Spotify_Music_Data <- read_csv("Spotify_Music_Data.csv")

## New names:
## • `` -> `...1`

## Rows: 603 Columns: 15
## — Column specification
##
## Delimiter: ","
## chr (3): title, artist, top genre
## dbl (12): ...1, year, bpm, nrgy, dnce, dB, live, val, dur, acous, spch,
pop
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this
message.

View(Spotify_Music_Data)

##Summary and descriptive statistics of variables
```

```
summary(Spotify_Music_Data)
```

```
##      ...1      title      artist      top genre
## Min.   : 1.0    Length:603    Length:603    Length:603
## 1st Qu.:151.5   Class :character  Class :character  Class :character
## Median :302.0   Mode  :character  Mode  :character  Mode  :character
## Mean   :302.0
## 3rd Qu.:452.5
## Max.   :603.0
##      year      bpm      nrgy      dnce
## Min.   :2010   Min.   : 0.0   Min.   : 0.0   Min.   : 0.00
## 1st Qu.:2013   1st Qu.:100.0 1st Qu.:61.0 1st Qu.:57.00
## Median :2015   Median :120.0 Median :74.0 Median :66.00
## Mean   :2015   Mean  :118.5 Mean  :70.5 Mean  :64.38
## 3rd Qu.:2017   3rd Qu.:129.0 3rd Qu.:82.0 3rd Qu.:73.00
## Max.   :2019   Max.   :206.0 Max.   :98.0 Max.   :97.00
##      dB      live      val      dur
## Min.   :-60.000   Min.   : 0.00   Min.   : 0.00   Min.   :134.0
## 1st Qu.: -6.000   1st Qu.: 9.00   1st Qu.:35.00   1st Qu.:202.0
## Median : -5.000   Median :12.00   Median :52.00   Median :221.0
## Mean   : -5.579   Mean  :17.77   Mean  :52.23   Mean  :224.7
## 3rd Qu.: -4.000   3rd Qu.:24.00   3rd Qu.:69.00   3rd Qu.:239.5
## Max.   : -2.000   Max.   :74.00   Max.   :98.00   Max.   :424.0
##      acous      spch      pop
## Min.   : 0.00   Min.   : 0.000   Min.   : 0.00
## 1st Qu.: 2.00   1st Qu.: 4.000   1st Qu.:60.00
## Median : 6.00   Median : 5.000   Median :69.00
## Mean   :14.33   Mean  : 8.358   Mean  :66.52
## 3rd Qu.:17.00   3rd Qu.: 9.000   3rd Qu.:76.00
## Max.   :99.00   Max.   :48.000   Max.   :99.00
```

```
summary(Spotify_Music_Data$artist)
```

```
##      Length      Class      Mode
##      603 character character
```

```
summary(Spotify_Music_Data$bpm)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0  100.0   120.0   118.5  129.0   206.0
```

```
summary(Spotify_Music_Data$dur)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     134.0  202.0   221.0   224.7  239.5   424.0
```

*##Converting variable "bpm" from integer to numeric*

```
Spotify_Music_Data$bpm <- as.numeric(Spotify_Music_Data$bpm)
```

*##Plotting of dB variable- loudness of song. No significant findings in*

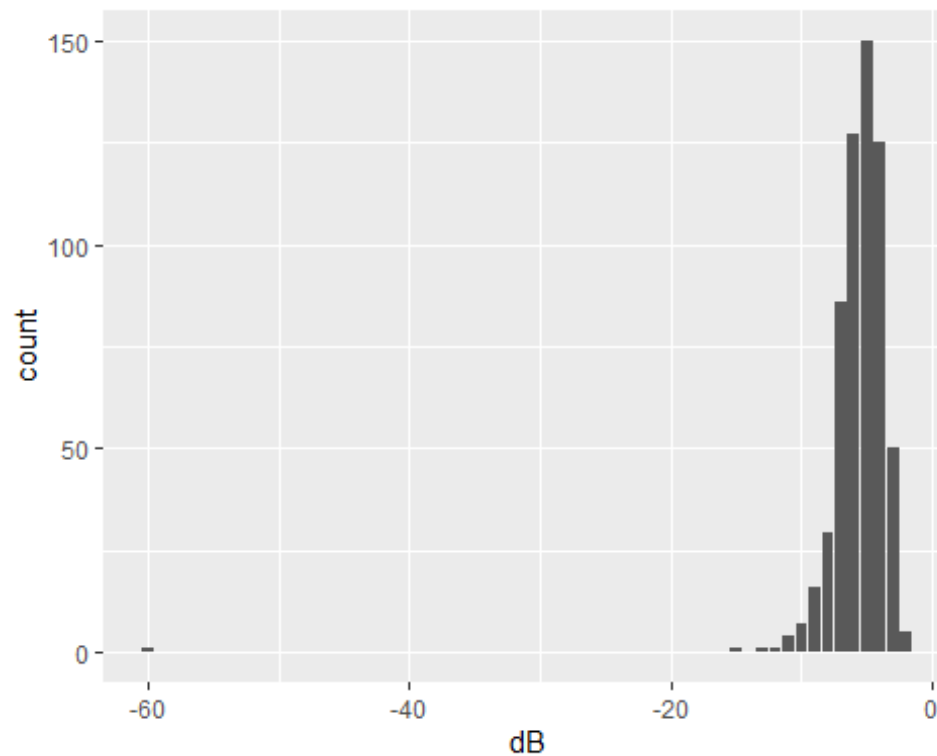
*results.*

```
install.packages("ggplot2")
```

```
## Warning: package 'ggplot2' is in use and will not be installed
```

```
library(ggplot2)
```

```
ggplot(data = Spotify_Music_Data, aes(x = dB)) + geom_bar()
```



*##Scatterplot- bpm of songs, per year. No significant findings in results.*

```
ggplot(data = Spotify_Music_Data, aes(x = year, y = bpm)) + geom_point()
```

