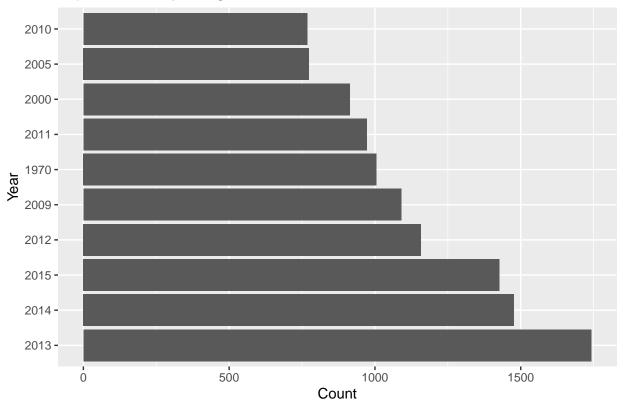
## R. Notebook

```
library(dplyr)
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##
       filter, lag
## The following objects are masked from 'package:base':
##
##
       intersect, setdiff, setequal, union
library(ggplot2)
library(tm)
## Loading required package: NLP
##
## Attaching package: 'NLP'
## The following object is masked from 'package:ggplot2':
##
##
       annotate
library(wordcloud)
## Loading required package: RColorBrewer
library(viridis)
## Loading required package: viridisLite
music <- read.csv("~/Software/musics/output.csv")</pre>
count(music)
##
## 1 21428
head(music)
##
     id
                                             title
                                                                 album artist name
## 1 1
                                 Good Ol' 39 Days #3 (Deluxe Version) The Script
## 2 2
           A Great Big Sled (feat. Toni Halliday) (RED) Christmas EP The Killers
## 3 3
                                      Brutal Love
                                                                 ¡TRÉ!
                                                                         Green Day
                                       Love Me Do
                                                                     1 The Beatles
## 5 5 Low Low (ft Tee Cee & Nipsey Hussle & RJ)
                                                            10 Summers DJ Mustard
## 6 6
                                       Chandelier 1000 Forms Of Fear
                genre year length bitrate lyrics album_id artist_id
## 1 Alternative Rock 2012 264045 320000
                                                        54
                                                                  26
                                              NA
## 2
              Holiday 2011 259030 320000
                                              NA
                                                       375
                                                                 286
## 3
                 Rock 2012 296356 320000
                                                      1000
                                                                 591
                                              NA
```

```
## 4
                  Pop 2000 141349 192000
                                               NA
                                                       617
                                                                 458
## 5
              Hip-Hop 2014 156290 320000
                                               NA
                                                       204
                                                                 183
## 6
                  Pop 2014 216120 320000
                                               NA
                                                       610
                                                                  74
cbind(lapply(lapply(music, is.na), sum))
               [,1]
## id
               0
## title
               0
## album
               0
## artist name 0
## genre
               0
## year
               0
## length
               0
## bitrate
               0
## lyrics
               21428
## album_id
               0
## artist_id
music <- subset(music, select=-c(lyrics, id, album_id, artist_id))</pre>
head(music)
##
                                          title
                                                              album artist_name
## 1
                              Good Ol' 39 Days #3 (Deluxe Version) The Script
## 2
        A Great Big Sled (feat. Toni Halliday)
                                                (RED) Christmas EP The Killers
## 3
                                   Brutal Love
                                                              ¡TRÉ!
                                                                      Green Day
## 4
                                    Love Me Do
                                                                  1 The Beatles
## 5 Low Low (ft Tee Cee & Nipsey Hussle & RJ)
                                                         10 Summers DJ Mustard
                                    Chandelier 1000 Forms Of Fear
## 6
                                                                            Sia
                genre year length bitrate
##
## 1 Alternative Rock 2012 264045 320000
              Holiday 2011 259030 320000
## 3
                 Rock 2012 296356 320000
## 4
                  Pop 2000 141349 192000
## 5
              Hip-Hop 2014 156290 320000
## 6
                  Pop 2014 216120 320000
top_10 <- music %>% count(year) %>% arrange(desc(n)) %>% top_n(10)
## Selecting by n
filtered <- music %>% filter(year %in% top_10$year) %>% mutate(Years = factor(year, levels=top_10$year)
ggplot(filtered, aes(x=Years)) +
 geom bar() + coord flip() +
 labs(x = "Year", y = "Count", title = "Top 10 Years by Song Count")
```



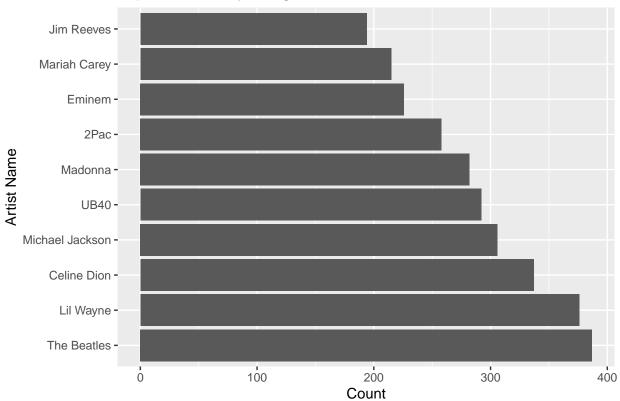


```
top_10_art <- music %>% count(artist_name) %>% arrange(desc(n)) %>% top_n(10)
```

```
## Selecting by n
```

```
## Selecting by n
filtered_art <- music %>% filter(artist_name %in% top_10_art$artist_name) %>% mutate(Artists = factor(a
ggplot(filtered_art, aes(x=Artists), fill=Artists) +
    geom_bar() + coord_flip() +
    labs(x = "Artist Name", y = "Count", title = "Top 10 Artists by Song Count")
```

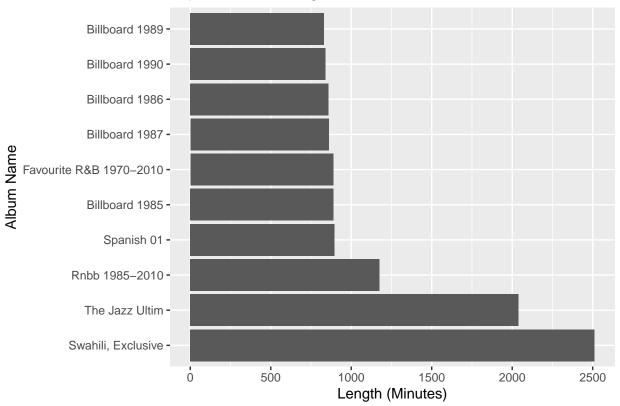
Top 10 Artists by Song Count



```
album_lengths <- music %>% group_by(album) %>% summarize(length_minutes = sum(length)/1000/60) %>% arraw
## Selecting by length_minutes
album_lengths$album <- factor(album_lengths$album, levels = album_lengths$album)

ggplot(album_lengths, aes(x = album, y = length_minutes)) +
    geom_bar(stat = "identity") +
    coord_flip() +
    labs(x = "Album Name", y = "Length (Minutes)", title = "Top 10 Album Lengths")</pre>
```





```
corpus <- Corpus(VectorSource(music$album))</pre>
corpus <- tm_map(corpus, content_transformer(tolower))</pre>
## Warning in tm_map.SimpleCorpus(corpus, content_transformer(tolower)):
## transformation drops documents
corpus <- tm_map(corpus, removePunctuation)</pre>
## Warning in tm_map.SimpleCorpus(corpus, removePunctuation): transformation drops
## documents
corpus <- tm_map(corpus, removeNumbers)</pre>
## Warning in tm_map.SimpleCorpus(corpus, removeNumbers): transformation drops
## documents
corpus <- tm_map(corpus, removeWords, stopwords("english"))</pre>
## Warning in tm_map.SimpleCorpus(corpus, removeWords, stopwords("english")):
## transformation drops documents
corpus <- tm_map(corpus, stripWhitespace)</pre>
## Warning in tm_map.SimpleCorpus(corpus, stripWhitespace): transformation drops
## documents
term_doc_matrix <- TermDocumentMatrix(corpus)</pre>
word_frequency <- rowSums(as.matrix(term_doc_matrix))</pre>
wordcloud(words = names(word_frequency), freq = word_frequency, min.freq = 40)
```

```
robot collection slow archiveorg raymond dead piece makiadi presents sinner season now night season reloaded soundtrack reloaded soundtrack reloaded soundtrack reloaded soundtrack reloaded soundtrack reloaded soundtrack something hill something hill number as something hill high beatles with sparty story of the plesongs dedication stevie money back mixtape blood americantown astillicome mariah back mixtape blood americantown stillicome mariah back mixtape blood americantown stillicome mariah back mixtape blood americantown stillicome mariah platinum jah empire girl bobby shalloyn of lets new past hank shallow platinum jah empire girl bobby shalloyn of last platinum jah empire girl bobby shalloyn of last wandross luther jackson paramore blue world hall mathers hardwell colour future bad days world hall mathers hardwell colour future bad days street loaded classic wonder shallow wonder street loaded classic wonder shallow black girl instrumentalmasters eyes black greatest deluxe ymcmb jazzz
```

```
corpus <- Corpus(VectorSource(music$title))</pre>
corpus <- tm_map(corpus, content_transformer(tolower))</pre>
## Warning in tm_map.SimpleCorpus(corpus, content_transformer(tolower)):
## transformation drops documents
corpus <- tm_map(corpus, removePunctuation)</pre>
## Warning in tm_map.SimpleCorpus(corpus, removePunctuation): transformation drops
## documents
corpus <- tm_map(corpus, removeNumbers)</pre>
## Warning in tm map.SimpleCorpus(corpus, removeNumbers): transformation drops
## documents
corpus <- tm map(corpus, removeWords, stopwords("english"))</pre>
## Warning in tm map.SimpleCorpus(corpus, removeWords, stopwords("english")):
## transformation drops documents
corpus <- tm_map(corpus, stripWhitespace)</pre>
## Warning in tm_map.SimpleCorpus(corpus, stripWhitespace): transformation drops
## documents
term_doc_matrix <- TermDocumentMatrix(corpus)</pre>
word_frequency <-rowSums(as.matrix(term_doc_matrix))</pre>
wordcloud(words = names(word frequency), freq = word frequency, min.freq = 50)
```

day years light acoustic mine way ways each something woman stop your years light acoustic mine was woman a girl angel turn a part life a get b never right years light and interludes ong fire of everything loststay want want another chris need better version rungonna lights walk bonus dream lil thing intro real walk bonus dream lil thing intro real light youreyoung heaven cant boy long let city a without blue side little little live nothing money still rock black high gone blues two radiobad prince white hold way crazy ways ways crazy like a way crazy wayne come thats time track