

Lab Assignment 3: Billboard Hot 100 Songs

Instructions

- Please answer these questions using code.
- The **only** R printouts should be the answers to the questions. Make sure your code does not display any extra information.
- You do not need to write any text for this assignment; only code.
- Submit your .html file to the assignment on PolyLearn.

Introduction

Today, we will study song popularity. In the US, the Billboard Hot 100 is a list that comes out every week, showing the 100 most played songs that week. The following code will load a dataset of Billboard Hot 100 songs, as well as the tidyverse and lubridate packages. More information about the creation of this dataset, as well as some analyses by the author, can be found here: <https://mikekling.com/analyzing-the-billboard-hot-100/>. The dataset you are provided is a limited version of the full data, containing: - The title - The artist - The highest rank the song ever reached (1 is the best) - The number of weeks the song was on the chart - The latest date the song appeared on the Billboard Hot 100

```
library(tidyverse)
library(lubridate)
songs <- read.table(
  "https://www.dropbox.com/s/jrwjthqo9b5o07g/billboard_songs.txt?dl=1",
  header = TRUE,
  stringsAsFactors = FALSE)
```

Advice

This is a very large dataset! Consider testing out your steps on a smaller subset of songs - say, the first 100 on the list - before you run your code and try to figure out errors on the full 34,605 songs. `slice()` will help you make a smaller dataset.

Questions

- 0) [Not graded, but useful!] Make your variables easier to work and/or adjust variable types.

Hint: use `str_sub` to find the year, month, and day of the chart leaving date, then create a datetime object from these.

- 1) What 10 songs spent the longest on the charts? Give only the title, artist, and weeks.
- 2) What date did the oldest song(s) in this dataset leave the charts? Give only the date.
- 3) What songs could have been played at your 16th birthday party? That is, which songs overall peaked at #1 and **entered** the charts within a couple months (before or after) your 16th birthday? Give only the song title, artist, and date of chart entry.
- 4) What is the most common word, at least 4 letters long, used in the title of any song? Give only the word itself, and its count.

```
## Some help for you:

titles <- songs %>%
  pull(title)

four_char_words <-
  # get a vector of all 4+ character words that appear in titles
  # unlist() might be useful

# You can simply copy this code
# but try to understand it!
four_char_words %>%
  table() %>%
  sort() %>%
  rev() %>%
  head(1)
```

5) Which five artists were the most successful in terms of:

- Number of songs that were #1 hits
- Total weeks any of their songs spent on the chart.

For this question, you may ignore collaborations; i.e., “Ariana Grande” is considered a different artist from “Jessie J, Ariana Grande & Nicki Minaj”

Hint: the function `n()` might help you summarize.

6) Which artist has **featured** on the most Billboard charting songs? Which artist has **collaborated** on the most Billboard charting songs?

Definitions:

RAE SREMMURD featuring NICKI MINAJ & YOUNG THUG

In this string, Nicki Minaj and Young Thug are considered to be **featured**.

JESSIE J, ARIANA GRANDE & NICKI MINAJ

In this string, Jessie J and Ariana Grande and Nicki Minaj are all considered to have **collaborated** on the song.

Hint: Take a similar approach from Question 4: Find a way to create a list of all collaborating/featured artists, then use `table()`, etc to see how often each name appears.