Week 10 Programming Assignment

Instructor: Chien-Hao Fu

4/18/2023

Due: 4/24/2023

Data Source: `billboard_lyrics.csv`

50 Years of Pop Music Lyrics (1965-2015). Author: Kaylin Pavlik (Github: walkerkq). Accessed 5/06/2019. Downloaded from https://github.com/walkerkq/musiclyrics

Preparation

1. Load the package `tidyverse`

2. Download the following files: "`billboard_lyrics.csv`", and "`W10 Student Submission.docx`"

3. Import the csv file "`billboard_lyrics.csv`" and save it as dataframe: `songs`

4. Take a moment to examine the data frame

Instruction

Task: Find the frequency of a particular word by year in Billboard 100 lists

Key Steps

1. Pick one or two words that interest you

   - Preferably words that often occur in the lyrics of pop music
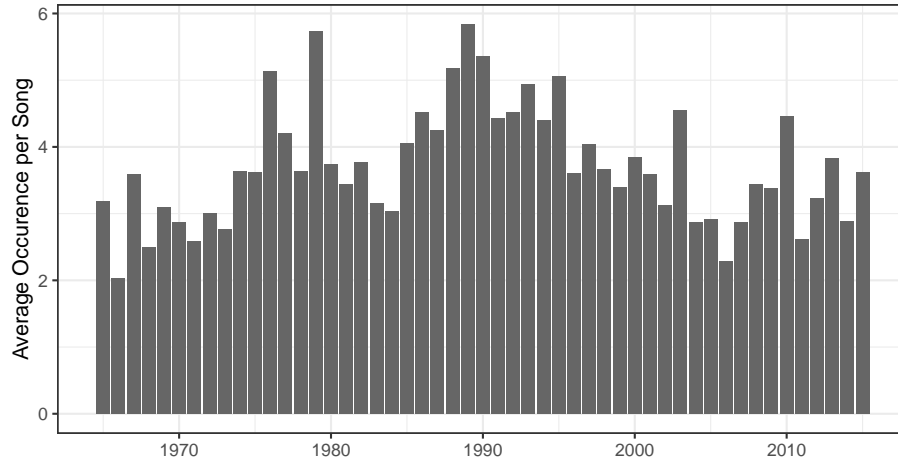
2. Generate a new variable to indicate the number of occurrence of the word

   - Tips: Use `mutate()` and `str_count()`

3. Calculate the average frequency per song by year for your chosen words

   - Tips: Use `group_by()` and `summarise()`
   - Use the `mean()` function in `summarise()` to calculate the average word frequency
   - Set the option: `na.rm=TRUE` to remove missing values in calulation

```
...
summarise( mean = mean(word_count, na.rm = TRUE)) ...
```

4. Use the data frame generated in Step 3 to draw a bar chart, in which `Year` is shown on the x-axis and the average frequency is on the y-axis
5. Save your graph(s) and source codes in the doc file "W10 Student Submission.docx"
6. Briefly comment on your charts

Your bar charts may look similar to the following (However, you should choose different words to investigate.):

Average occurence of the word "love" in the lyrics
of Billboard Top 100 Songs, 1965–2015

Average occurence of the word "hate" in the lyrics
of Billboard Top 100 Songs, 1965–2015