Week 12 Programming Assignment

Instructor: Chien-Hao Fu

5/09/2019

Due: 5/15/2019

Data Source

- `transcript.txt`
  摘錄自批踢踢實業坊 | 看板: Gossiping | 作者: Crossmy88. https://www.ptt.cc/bbs/Gossiping/M.1556942183.A.25A.html
- `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: "transcript.txt", "billboard_lyrics.csv", and "W12 Student Submission.docx"
3. Import the text file "transcript.txt" with `read_file()` and save it as a single string: `transcript`
4. Import the csv file "billboard_lyrics.csv" and save it as dataframe: `songs`
5. Take a moment to examine these data frames

Part 1: Use `str_count` to count Chinese words

- Task: use the function `str_count` to count the number of occurrences of the following words in the string `transcript`
  - 黃捷
  - 韓國瑜
  - 發大財
  - 發財
- Fill in your answers in Part I of the document "W12 Student Submission.docx"

Part 2: Changes in Lyrics over Time

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

Key Steps

1. Pick one or two words that interests you
   - Preferably words that often occurs 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 "W12 Student Submission.docx"
6. Briefly comment on your charts

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