Love, the heart of Songs' Spirituality

Xudong Guo

September 17, 2019

Overview

Always motivated and inspired by different types of songs, I am fanatical about the subconscious part of the author's mind when a song lyric was generated. This project renders me great opportunity in exploring intuitive logics between words and sentences by using text mining as well as natural language processing methods.

Introduction

The processed_lyrics dataset was provided by Professor and modified from the original one so that important and meaningful words were stemmed from the song lyrics. We have 125704 rows and 7 columns which including variables "song", "year", "artist", "genre", "lyrics", "id" and "stemmedwords". The years of songs ranging from 112 to 2016 and genres classified into 12 categories: "Country", "Electronic", "Folk", "Hip-Hop", "Indie", "Jazz", "Metal", "Pop", "R&B", "Rock", "Other" and "Not Available" which are the two major variables I will mainly focus on besides variable "stemmedwords".

The Analysis of Song Lyrics
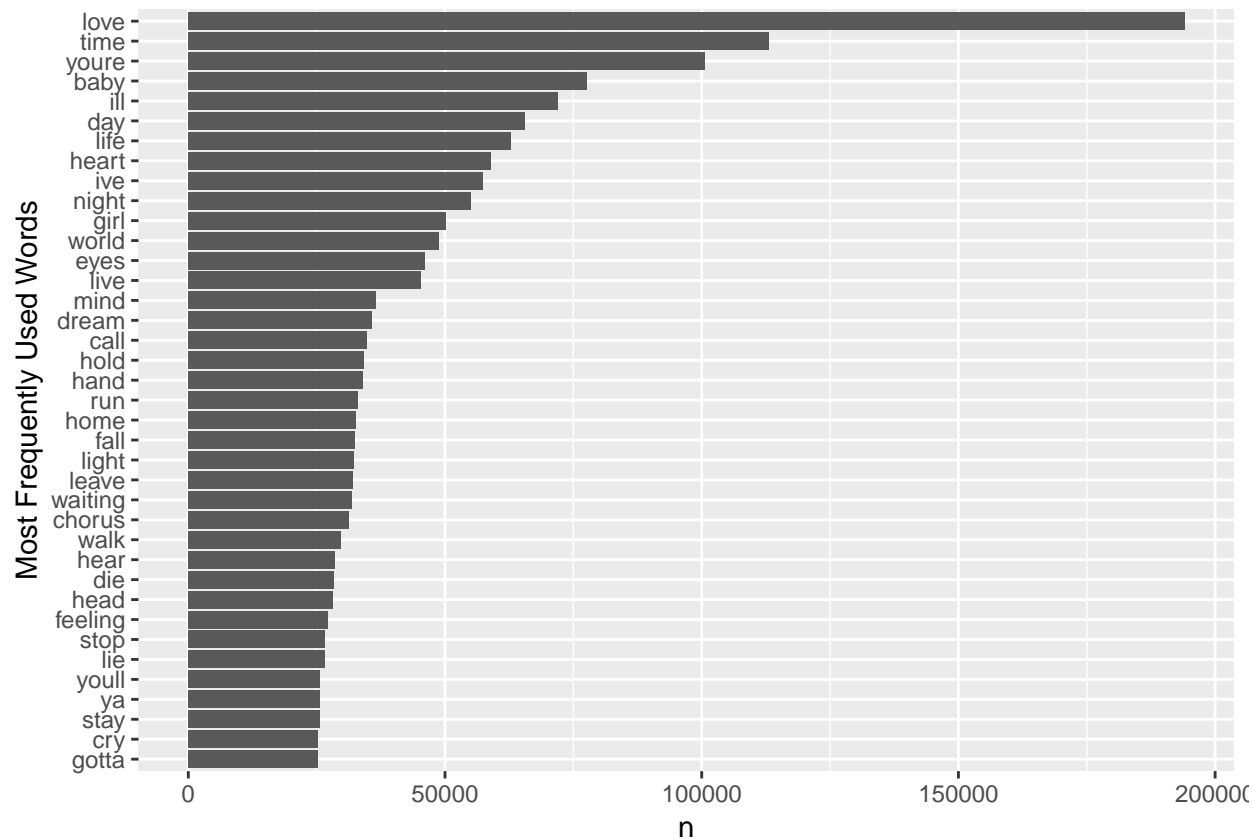
When taking all the song lyrics as a whole and looking at each individual words, not surprisingly, the word "love" appears most frequently then follows with "time", "youre" and "baby". The word "ill" appears 71915 times which far exceeds my expectation.

The Sentiment Analysis

```
## Joining, by = "word"
## Joining, by = "word"

##              genre  word     n      freq
## 1             Pop   love 53150 2.8427020
## 2             R&B   love  6055 2.7851886
## 3            Jazz   love  9409 2.2826298
## 4   Not Available  love  6843 2.0580451
## 5             R&B   baby  4128 1.8988040
## 6           Other   love   268 1.8231293
## 7         Country  love 13308 1.7663924
## 8      Electronic  love  4865 1.7300853
## 9         Hip-Hop  love 15234 1.7107243
## 10          Other  baby   212 1.4421769
## 11        Hip-Hop  shit 14150 1.5889949
## 12        Hip-Hop bitch 10152 1.1400337
## 13          Other   ill   140 0.9523810
## 14        Hip-Hop   ill  7556 0.8485121
## 15          Other bitch   115 0.7823129
## 16        Hip-Hop money  6908 0.7757440
## 17            Pop   ill 14061 0.7520458
## 18          Other  shit   104 0.7074830
## 19            R&B   ill  1526 0.7019319
## 20        Hip-Hop   hit  6028 0.6769231
```

By referring to the nrc lexicon, I am mostly interested in the negative emotion: anger as well as the positive emotion: joy. When grouping by genre, words representing joy such like "love" and "baby" have significant appearances. For the word "love", Pop, R&B and Jazz music dominate the top three ranking in terms of frequency, that is, the number of "love" appearances over all word counts in each genre. On the other hand, Hip-Hop music has the highest frequency of negative words like "shit", "ill" which are representing anger.

```
## Joining, by = "word"
```
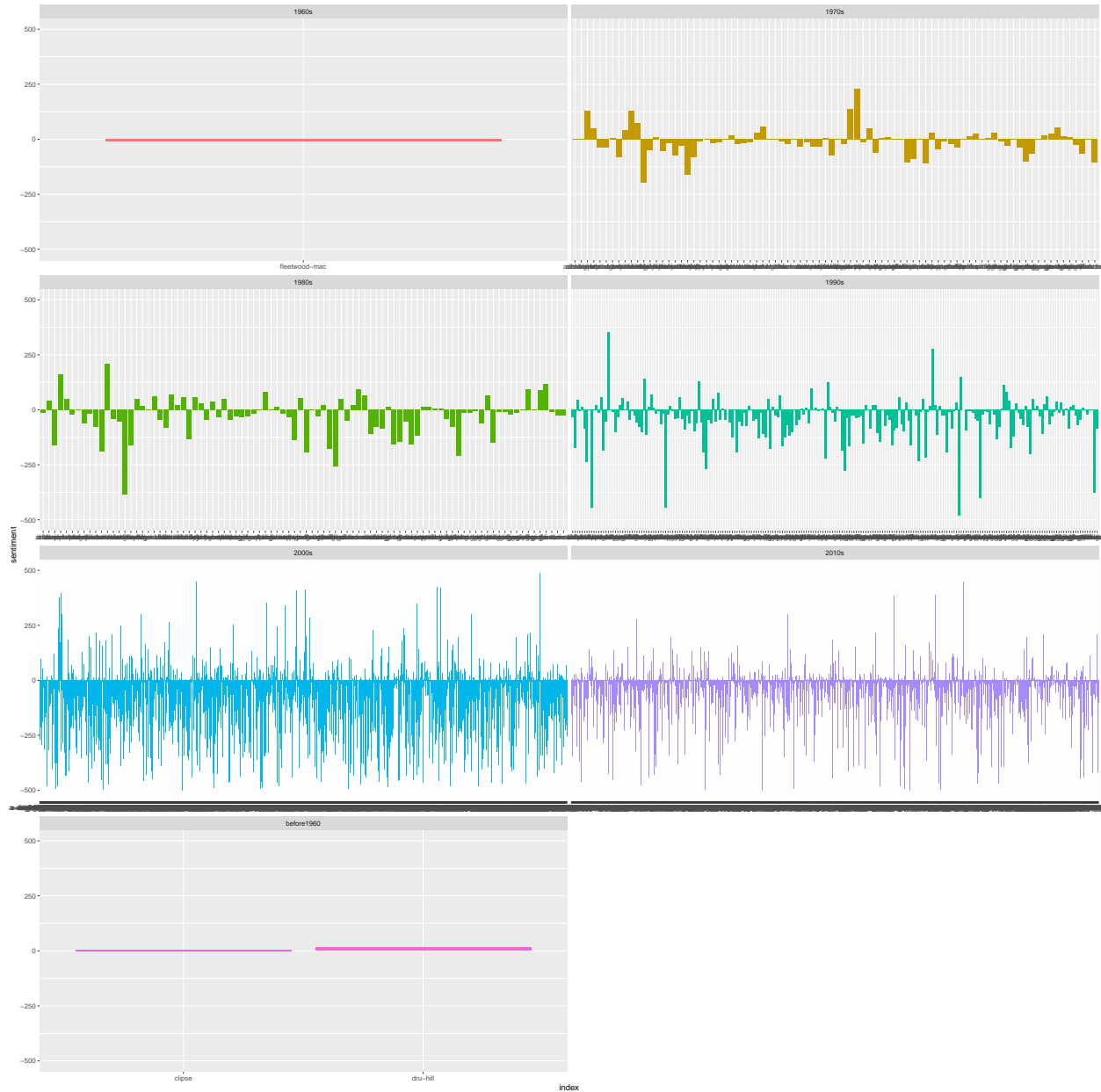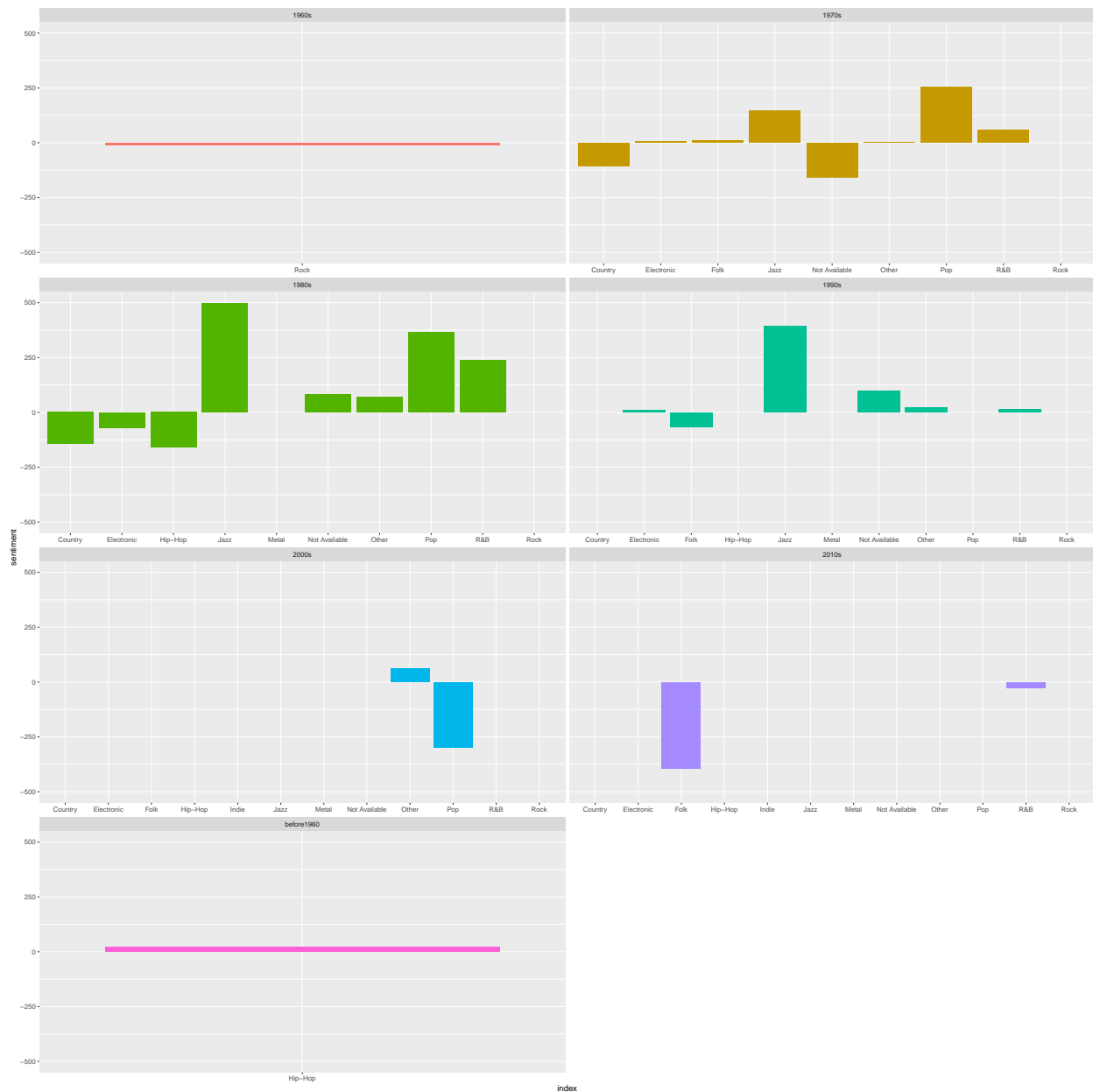
From the sentiment analysis plots

which categorized by lyrics genre and due to the frequency calculation in previous parts, it's not counterintuitive that Hip-Hop always delivers negative emotions as well as Indie, Metal and Folk. On the contrary, most Jazz lyrics are full of positive emotions such as love and happiness.

```
## Joining, by = "word"
## Joining, by = "word"
```

Based on the graphs above, we notice an emerging pattern of the number of artists. Ironically, despite of the massive production of song lyrics, the expression of anger and sadness (negative emotions) is on the rise too. In the 1980s and 1990s, most genres were in positive attitudes especially Jazz, Pop and R&B. However, after stepping into the new millenium, people managed to use Pop music to deliver their negative emotions. In addition, the type of genres significantly changed: people don't like Jazz anymore, instead, more and more folk music comes out.

```
## Joining, by = "word"
```

```
## Selecting by n
```
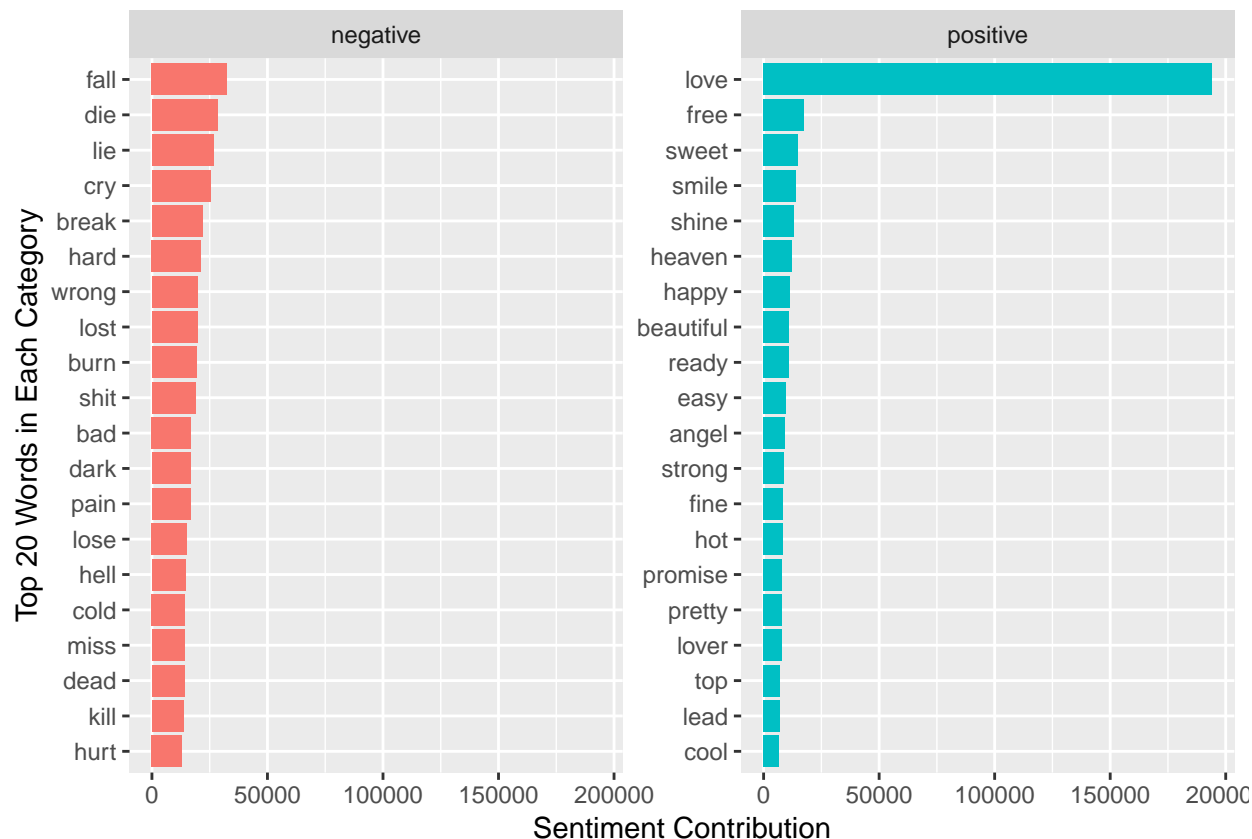
So besides the word "love", what other vocabs those artists use to deliver their either bad or good emotions? For the positive emotion, "free", "sweet", "smile", "shine" are frequently used.For the negative emotion, "fall", "die", "lie" and "cry" are commonly applied.

Bigrams Analysis

```
## # A tibble: 10 x 3
##    wordone wordtwo     n
##    <chr>   <chr>   <int>
##  1 love    baby     4463
##  2 baby    love     4123
##  3 close   eyes     4077
##  4 fall    love     3828
##  5 live    life     3378
##  6 rock    roll     2934
##  7 set     free     2875
##  8 break   heart    2737
##  9 chorus  x        2729
## 10 love    youre    2650
```

When using two-grams approach, the most likely combination is "love baby" then follows with "baby love", "close eyes" and "fall love" which mostly in positive emotions. Unlike positive attitudes are often expressed in two words, negative emotions are delivered in a single word.

```
## # A tibble: 10 x 3
##    wordone wordtwo     n
##    <chr>   <chr>   <int>
```

```
##  1 love    baby     4463
##  2 baby    love     4123
##  3 fall    love     3828
##  4 love    youre    2650
##  5 true    love     2346
##  6 love    time     2335
##  7 ill     love     2308
##  8 youre   love     2169
##  9 time    love     2123
## 10 love    ill      2085
```

Although there are a bunch of negative emotions in the song lyrics right now, love is still deeply rooted in people's mind. Subconsciously, no one would deny the strength of love and we use love to reveal their true needs. The table above shows the most frequently combinations with at least a word "love". Words such like "time", "baby", "true" are with "love" quite often.

Based on the bigrams visualization, "love" is always the heart of lyrics which connecting with "sweet", "forever", "life", "day", etc. In additions, some combinations are also popular such like "light shine", "heart broken", "doesn't matter", "wind blow", etc.

Conclusions

La Vie en Rose, my most favorite jazz song, is always inspirational. Unsurprisingly, based on the sentimental analysis, jazz music is the genre which lyrics are most likely delivering positive emotions. There is a strong pattern that more and more negative emotions are involved into song lyrics. Meanwhile, people's taste has changed from Jazz, R&B to Pop and Folk music. "Love" is the most frequently used word in song lyrics which always combined with "sweet", "forever", "baby", "heart", etc. On the other hand, "fall" and "die" are

the most frequently used words with negative emotions.