

Music Lyrics Analysis

02.137 Digital Humanities
Group Members:
 Cordeiro Dominic Gabriel Tan
 Liu Menglu
 Siow Lee Sei
 Zhang Manyue

Introduction

Music is an art form that delivers messages to the world or promoting an idea they believe in. Musicians normally express their feelings or ideas through the lyrics. Therefore, it is very interesting to do analysis on the lyrics to find out the hidden meaning.

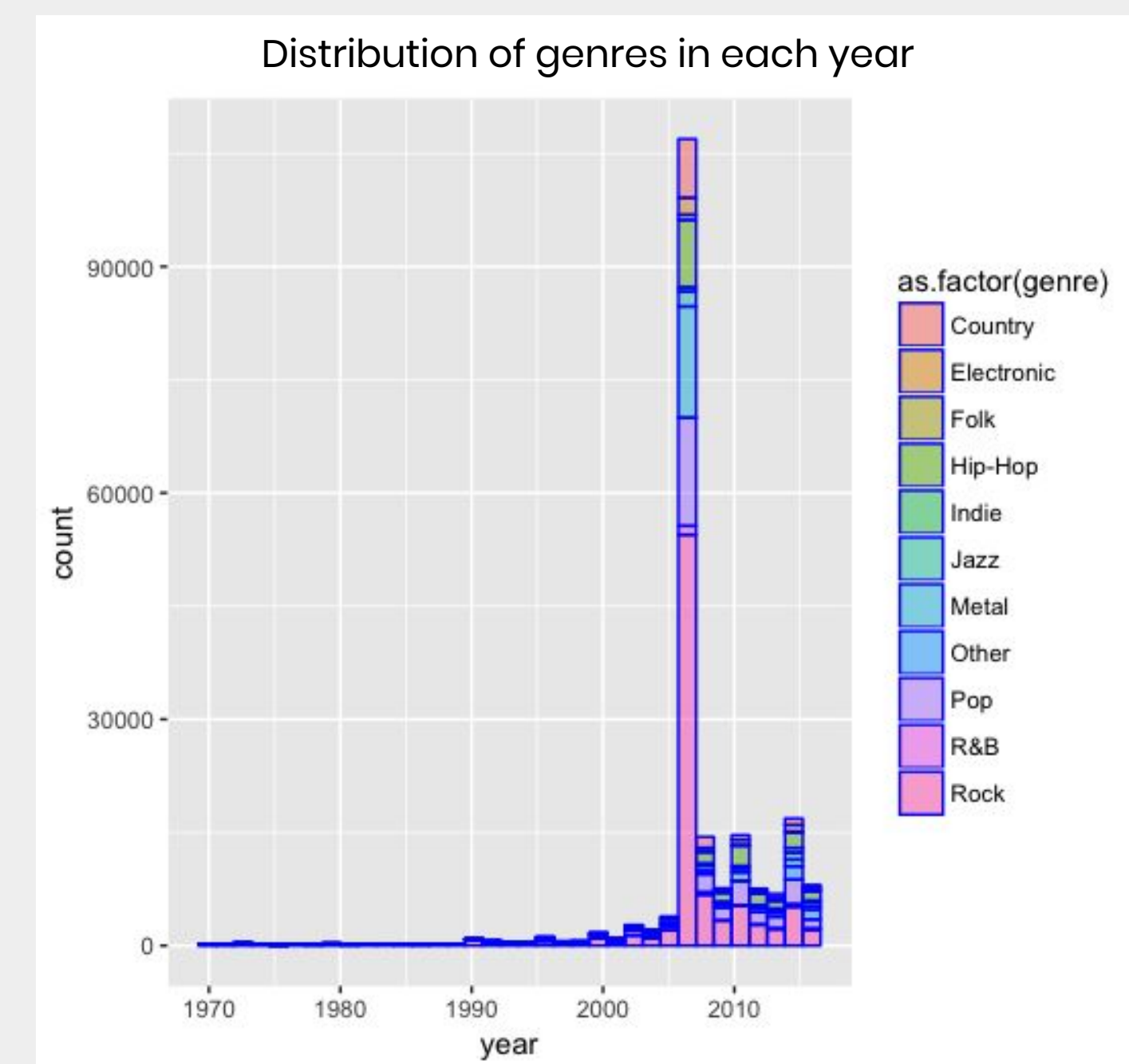
In this poster, we will explore trends in music lyrics from the 1970 to 2016. There are in total of eleven genres of lyrics that we explored, which are Pop, Indie, R&B, Folk, Electronics, Jazz, Country, Metal, Rock, Other and Hip-Hop.

For the analysis on lyrics, we have done several computations on the lyrics of each song in our dataset of more than 220000 songs. Our computations include sentiment analysis, frequency analysis, the use of word clouds and polynomial fitting.



Our dataset

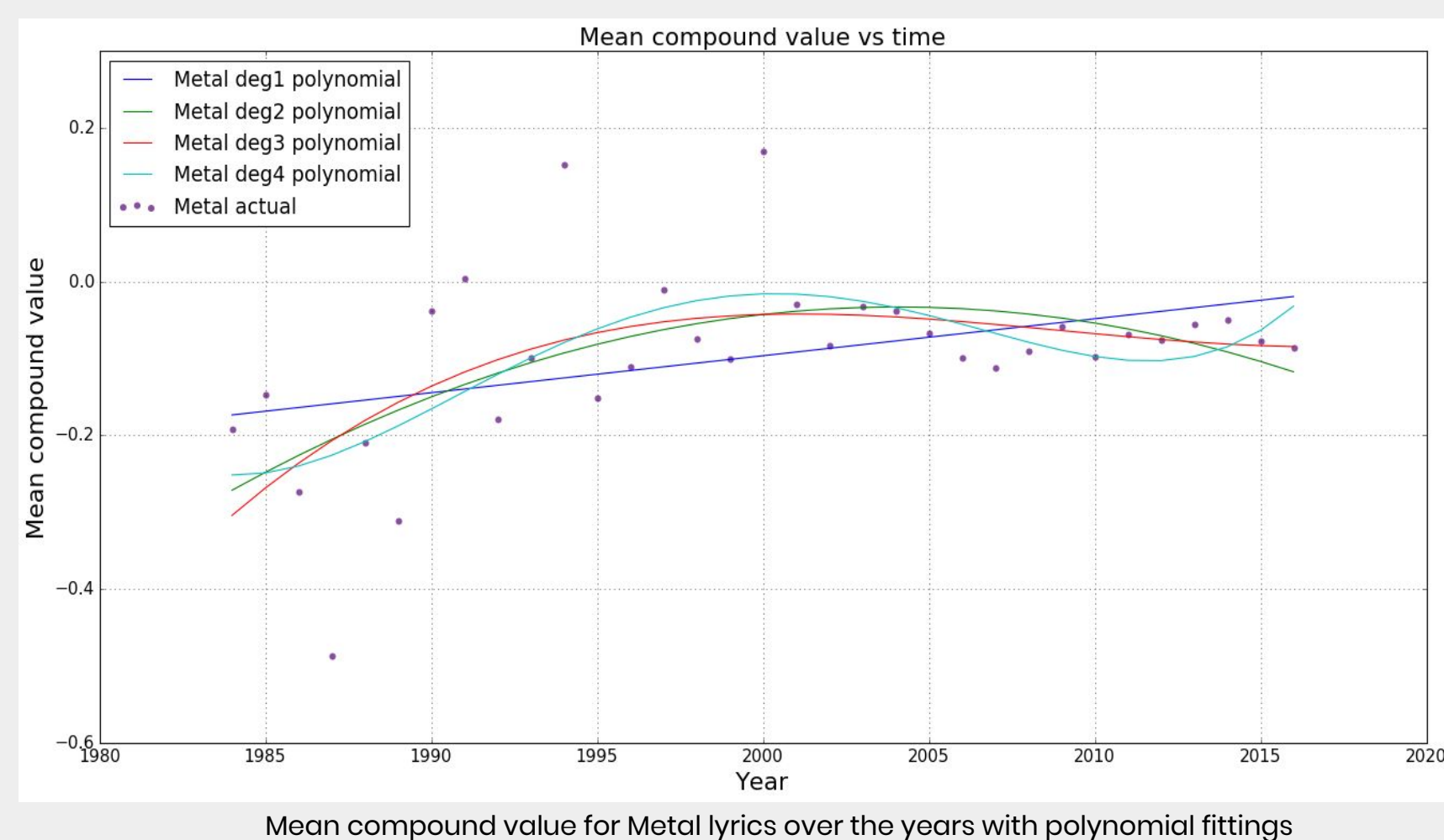
We have gathered as many lyrics as possible of the songs from the past few decades. The final set of data we obtained did not have even distribution of songs from each genre in each year. In year 2006, there was a total number of over 100,000 songs we have found which formed half of our dataset.



Trends in genres

After sentiment analysis on the lyrics are done, we get a score called compound value. A positive value indicates that the song has a positive meaning, negative values indicate that the song has negative meanings and zero indicating that it is neutral. A mean of this value for all songs in a genre of a particular year gives us a general idea on the sentiments of the song of that year.

The mean compound value of certain genres such as metal, is quite interesting. Looking at the trend overtime by doing a polynomial fitting, we see that the mean compound value tapers off over the years. This shows that metal songs are becoming less "negative" overtime. Do note that modern day metal songs (year 2010 onwards) still have a mean compound value that is negative, though not as negative as they were in the 1980s.



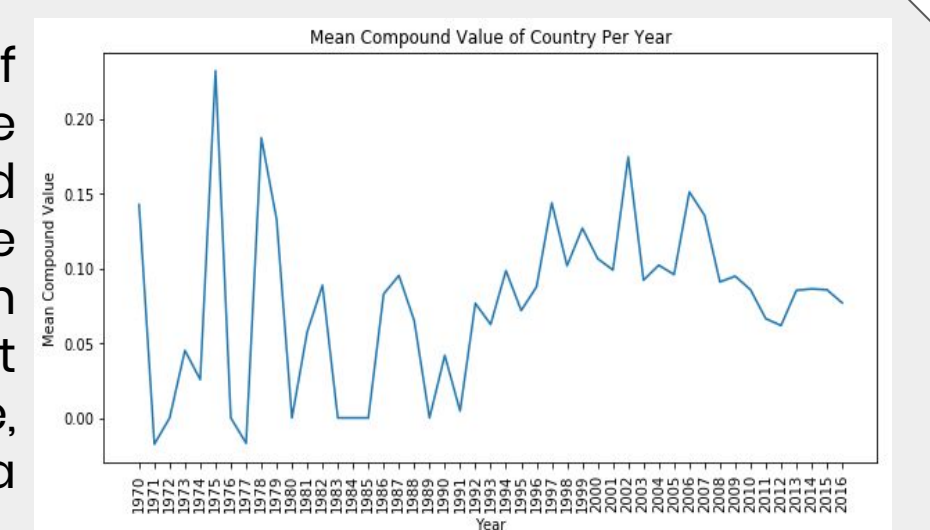
Style in genres

From the results of sentiment analysis on the lyrics of each genre over the years, we realised that there are some genres with fluctuating mean compound values at the start of 1970s and it becomes stabilize as the year progress. Therefore, we came out with the hypothesis that the musicians probably do not have a unified style in writing the lyrics for that genre, but as the year goes on, they have finally reached a unified style of the lyrics.

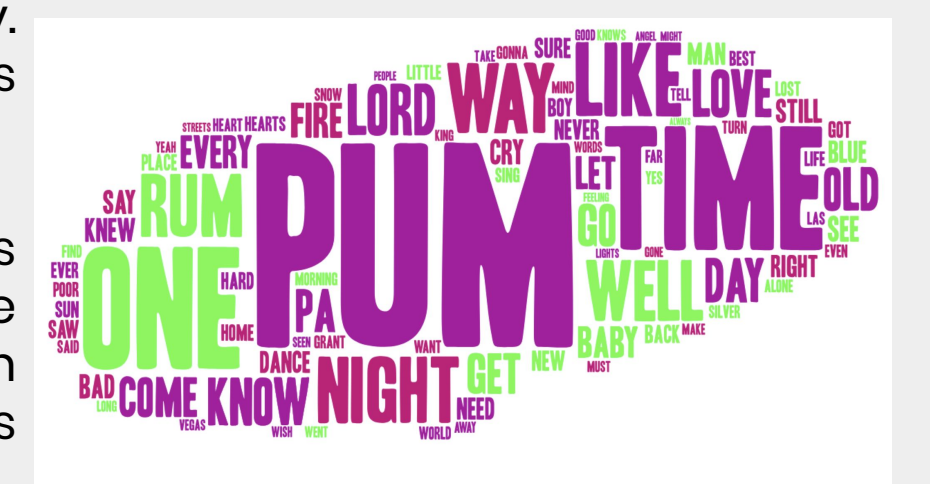
One of the genres with such trend is Country. Counting of words in lyrics around 1970s and 2010s have been done and the results are compared.

We found out that most of the words used in 2010s are mostly related to love, while there are more special words like 'angel', 'king' and 'lord' which indicates there is a variety of themes in Country lyrics in 1970s.

Moreover, 'pum' is the most frequently used word in 1970s lyrics. It is part of the phrase 'pa rum pum pum pum' which imitates the sound of a drum and is commonly used as chorus. However, in 2010s, there is no any special words can be used to imitate the sound of an instrument, and thus this might hints that the musicians have changed the style of writing the chorus for Country lyrics.



Mean compound value for Country lyrics over the years



Word cloud for 1970s Country lyrics



Word cloud for 2010s Country lyrics

The love songs

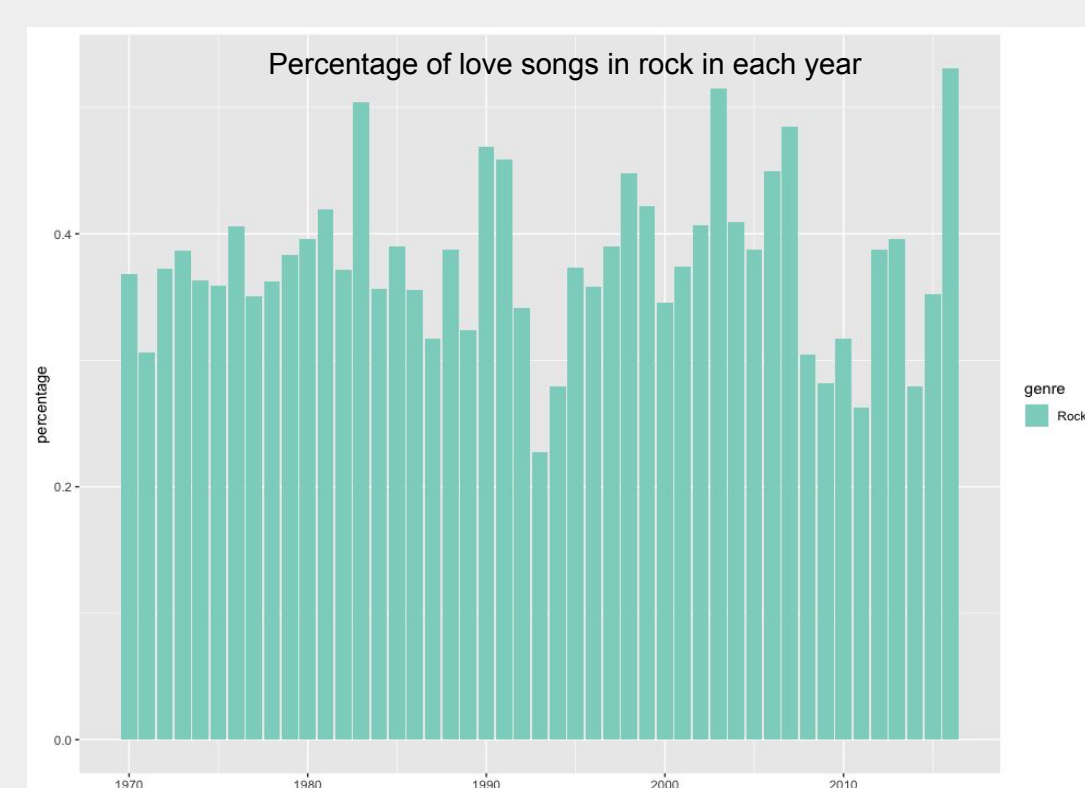
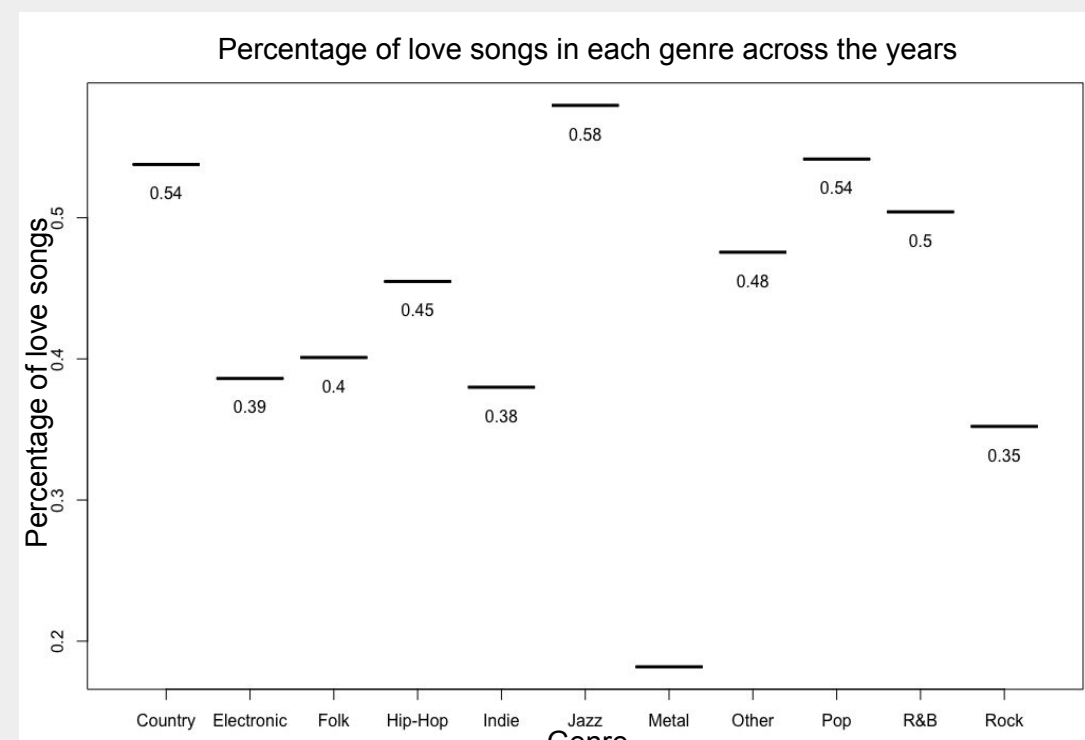
Based on the results from word cloud, we observed that the frequency of appearance of the word "love" is usually amongst the top five ranking in almost all genres. We then came out with a hypothesis of which there are more love songs than other types of songs in each music genre.

We started with identifying love songs in our dataset. If the word "love" was used as a lyric, the song would be recognized and labelled as a "love song".

We found out that most genres consist of more than 40% love songs. An interesting finding was that even for rock music, 35% of them are love songs. This percentage has not fallen in the range of 30% to 40% over the years.



Word cloud for rock music



Limitations

We have made some valuable observations as stated in the other sections. However, we have found some shortcomings for our model during the analysis process:

- Most frequently used words "like", "know", and "love" is generally used for all kind of genres, they could be used in many context. Therefore, to determine the style of the lyrics by word frequency is not sufficient.
- As the number of songs in each genre for each year is not the same, determining the change in style of songs over years by comparing the mean compound value and word clouds might not be accurate..

