

SI330W18 Project Proposal
Libby Murray

Both music and language are complex communication systems that have deep ties across cultures. One way to look at the effect music has is the artist's relationship with their users. Social media is a big driver for success for celebrities and their business, but does this carry over for musicians? This is why I'm interested in understanding what effect does the activity of an artist, via the content they post, have on the popularity, the ranking, of their music. The main question I hope to answer is: Do more involved artist on social media have more success? Additionally, how much is the language used in an artist's music reflected in the language used on social media?

Data sources:

- Name: Spotify artists (81323 artists on Spotify)
 - a. Size: 4.15 MB
 - b. Location: <https://www.kaggle.com/ehcall/spotify-artists/data>
 - c. Format: CSV
 - d. Access method: download CSV file from Kaggle
- Name: Twitter
 - a. Location: Twitter.com
 - b. Format: JSON
 - c. Access method: Web API
- Name: Top Spotify tracks of 2017 (Spotify's playlist of the songs streamed most often over the course of 2017)
 - a. Size: 12.84 KB
 - b. Location: <https://www.kaggle.com/nadintamer/top-tracks-of-2017/data>
 - c. Format: CSV
 - d. Access method: download CSV file from Kaggle
- Name: Spotify's worldwide daily song ranking (200 daily most streamed songs in 53 countries) -- How a songs popularity varies over time
 - a. Size : 43.07MB
 - b. Location: <https://www.kaggle.com/edumucelli/spotify-worldwide-daily-song-ranking/data>
 - c. Format: CSV
 - d. Access method: download CSV file from Kaggle

Data manipulation:

- For initial processing I don't see any data manipulation I will need to do at this time
- For combining the data, I will use regex functions to find common words, and relationships between groups with NLTK.
- The new information will be: how many genres are represented, how many of their song reach the top charts, and what are the most common words used for artist in a specific genre

Data visualization: Timeline- Number of tweet from artists & number of streams on spotify over the span of 2017 (<http://timeline.knightlab.com>)

