DSCI 235 Final Assignment Proposal

Ethan Creagar

We've often been told that a lot of Pop music sounds the same and is "formulaic". To examine this question, I found a dataset containing metrics of 41,106 songs from 1960 to 2015. The dataset contains basic musical metrics such as beats per minute, tempo, key, time signature, and song length, as well as other categories compiled by Spotify's web API that contains metrics like instrumentalness, danceability, and energy. With this data, I am intending to answer questions about how pop music can be categorized, what factors are most instrumental in making a song a pop-song, and what songs follow this mold best. Specifically, I intend to answer the following questions:

1) Is pop music formulaic?

- What percent of (hit) songs are within a certain interval for each category?
- How much do hit songs deviate from the means vs. non-hit songs?

2) What are the best predictors for what will become a hit song?

- This question can be answered using some of the same metrics as before. Which categories do hit songs deviate from non-hit songs the most?

3) What is the 'ideal' pop song?

- Using the mean and standard deviation metrics above, how should artists write songs if they want to get on the top 100 lists?
- What songs follow this mold best? Which popular songs stray the furthest from this mold?

4) How has popular music changed through the last 50 years?

- Plot the means of categories for hit songs over decades and see which categories have seen the largest increase/decrease from the 1960s to now.

5) Who is the top artist of the last 50 years?

- Who appeared the most times on the top 100 list?
- What percentage of their songs made top 100?

These questions can be answered through means, standard deviations, and graphs. I plan to run similar operations on each of the data columns to determine in which columns the largest differences between hit song and non-hit songs lie in.