



This is an analysis of American culture and society from 1950 to 2015 through the lens of pop song lyrics.



Studying music and its lyrics can bring us into the mind of the musician, but studying popular music can tell us what the masses want to hear. By examining the lyrics of the billboard top 100 songs from 1950 to 2015, I aim to discover the various themes and messages behind pop music, and at which points in time those themes become more or less prevalent than others.

Agenda

1. Data
2. Word Count & Word Use
3. Sentiment Analysis
4. Topic Modeling Process

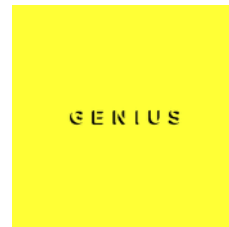
I'll first explain how I gathered my data, I'll talk about an analysis I performed using the word counts of the various songs. I'll briefly touch on a sentiment analysis of the lyrics throughout the decades, and lastly explain my topic modeling process.

Data

Song Title, Artist, Release Year



Lyrics

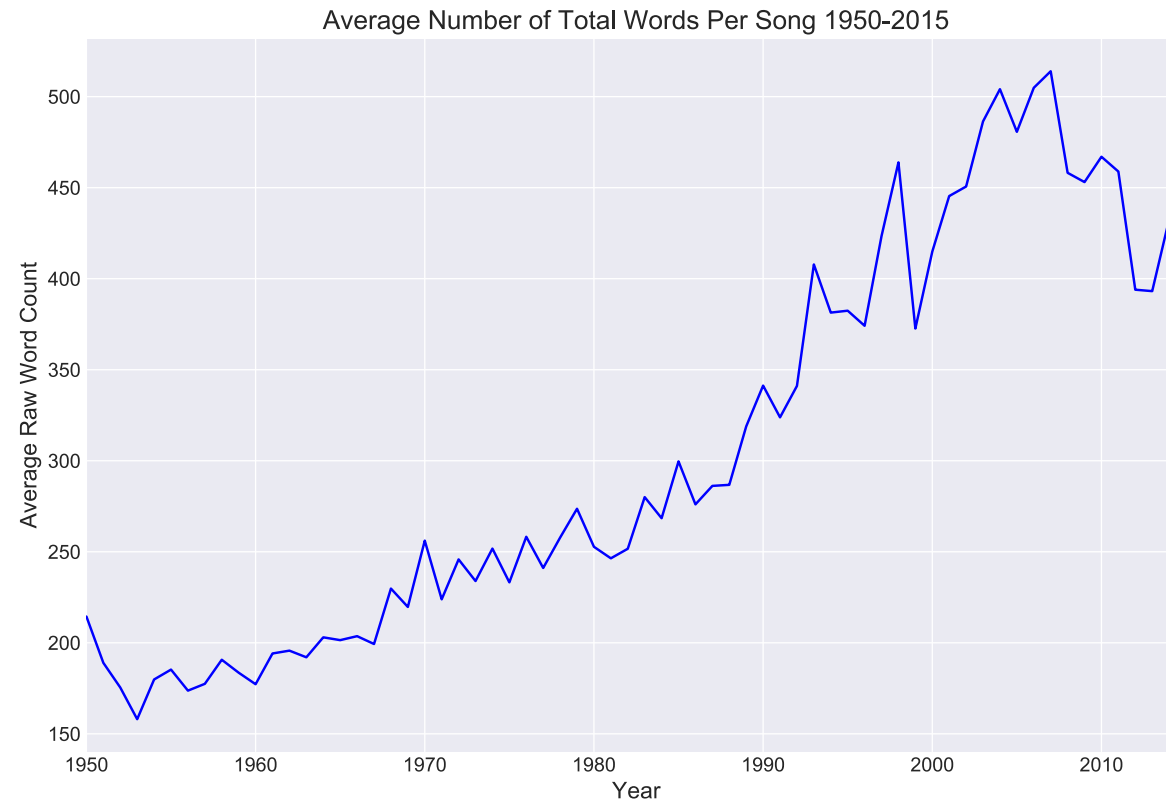


I obtained the song metadata from a CSV file on r/datasets.

I then got the URLs for the lyrics through the Genius API and scraped the content using Beautiful Soup. Any remaining songs I wasn't able to get from Genius I scraped directly from MetroLyrics and LyricsWiki.

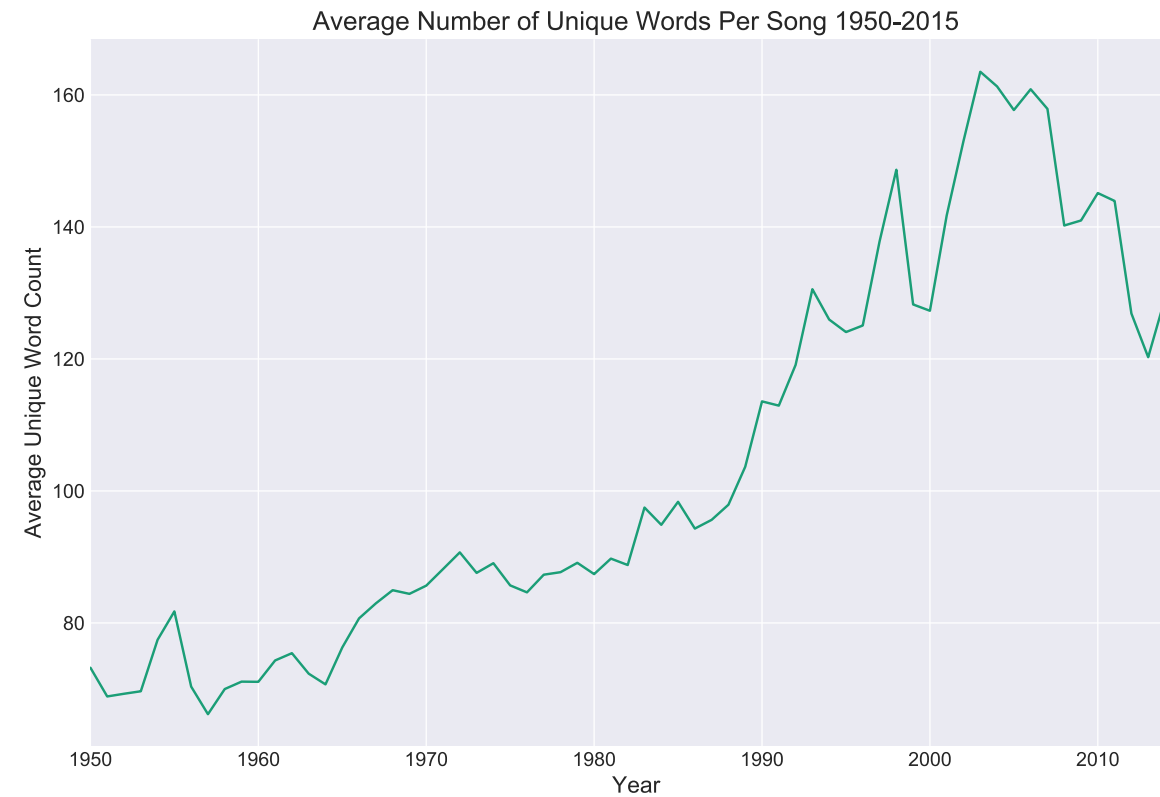
In total I was able to gather 5,800 songs - there were a fair amount of lyrics from the 1950s that weren't available on the internet, and I'm missing some lyrics throughout the years due to naming inconsistencies between my dataset and the websites.

Raw Word Count



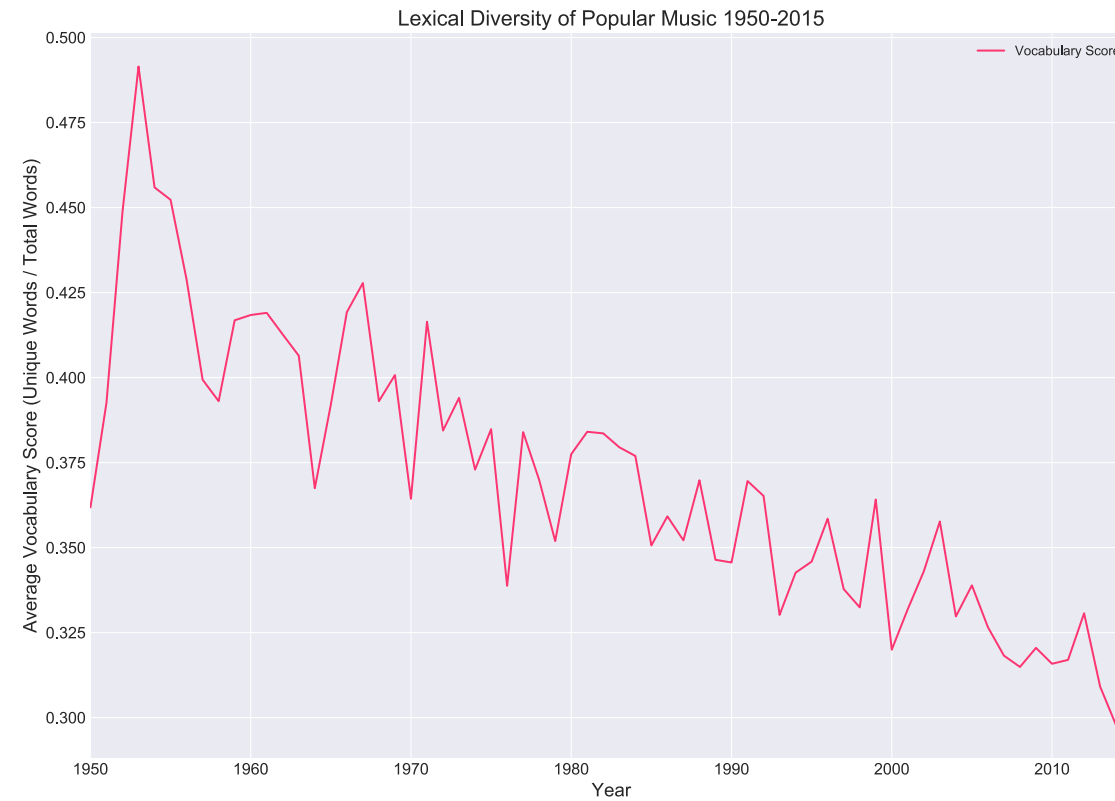
Ok, let's talk about raw word count. Throughout the years it is clear to see that pop musicians are using a greater number of total words in their music, with a drop off near the mid 2000s. However, this could likely be due to increased lyric repetition and is not a very telling metric on its own.

Unique Word Count



Similar to raw word count, the number of unique words used in pop music has also seen a steady increase over the years with a dip near the turn of the century. Unique word count is a stronger indication of superior articulation than raw word count, and one could argue that songs with more unique words are intended to contain more complex meaning. However, we need to turn to vocabulary to confirm whether this is true.

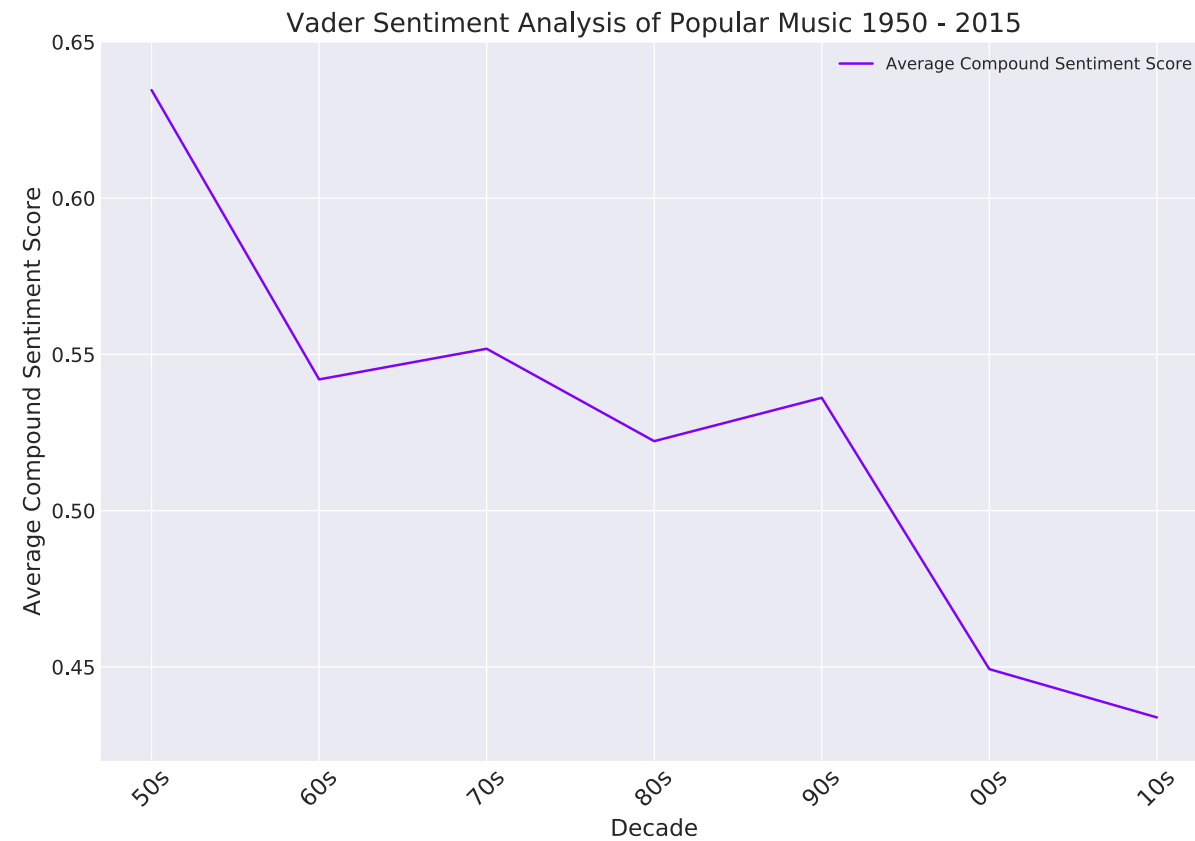
Lexical Diversity



Lexical diversity, also known as ‘vocabulary’ is the number of unique words in a song divided by the number of total words. A song with a higher vocabulary score is said to contain more complex meaning. This figure indicates increased word repetition and suggests that over the years we are talking more, aka using more words, but saying less, aka conveying a weaker message to our audience.

However I have a caveat, word repetition in music has rhythmic function and a lot of pop music uses repetition as a sonic device. The words ‘ooh’, ‘ah’ and ‘yeah’ are examples of words that are more likely to carry sound value as opposed to semantic meaning. So a song with a lower vocabulary score does not necessarily mean it is simple, because the decision to repeat certain words is often based on musicality.

Sentiment Analysis - Vader



Using Vader Sentiment Analysis, I calculated the average compound sentiment score for each decade of music. A song with a compound sentiment score of greater than .05 is said to be a positive song, so although the average pop song in American society technically has a positive sentiment (as every song on this graph is greater than .05), we can conclude that sentiment is declining and popular music today tends to touch on more negative subjects than it did in the past.

Topic Modeling

- **Regex, NLTK for Lemming & Stopwords**
- **TF-IDF**
- **NMF**
- **Fine Tuning, Interpretation**

So onto topic modeling. I first used regex to do an initial cleaning of my corpus and used the NLTK libraries to lemmatize the lyrics and add in certain stopwords. TF-IDF was used to vectorize the raw text so it could be fed into a model. For NMF, I extracted 30 topics after much tuning and focused on seeing which songs favored certain topics by reverse engineering the process. In the end, out of those 30 topics I honed in on 10 that explicitly displayed themes commonly found in pop music.

Interactively Exploring Topics

Exploring Themes in Popular Music

Using plotly, I created an interactive visualization which allows you to explore the trends in those 10 different topics over time.

Thank You

Any Questions?