Chris Trost & Reilly Koren

DHUM 1100

4 October 2021

Text Analysis in Voyant & R: Sentiment and Word Choice of Music

Question:

Critics claim that mainstream music genres like Rap, Pop, and Rock are lyrically simple or promote negative messages. From a text analysis standpoint, do mainstream music genres appear any more negative or simple than a more respected medium like poetry? Further, have these genre's lyrics gotten simpler over time, and how do they compare to one another?

Predictions

Our hypothesis is that across the genres, rap will have more variety in vocabulary than pop, more similar to genres such as rap and poetry. However, the sentiment and use of vulgar language will be the highest in the genres of rap and rock, which are more rebellious in nature than poetry and popular music. This language will give rap/hip-hop and rock more negative sentiment scores, although it will be interesting to see if rap has become more or less negative over time.

Due to the progression of genres such as rock and rap becoming mainstream, we predict that over time, each of these genres will decrease in variety of vocabulary to have a more general appeal and decrease in rebellion. However, we feel that pop music will remain constant, as it is the most widely listened to genre.

Genre	**Artists**
Rap	Biggie, Kanye West, Drake
Pop	Michael Jackson, Britney Spears, Justin Bieber
Rock	Jimi Hendrix, Nirvana, Nickelback/Blink-182
Poets	Dickinson, Nursery Rhymes/Dr Seuss (to compare for lyric simplicity)

Voyant Analysis

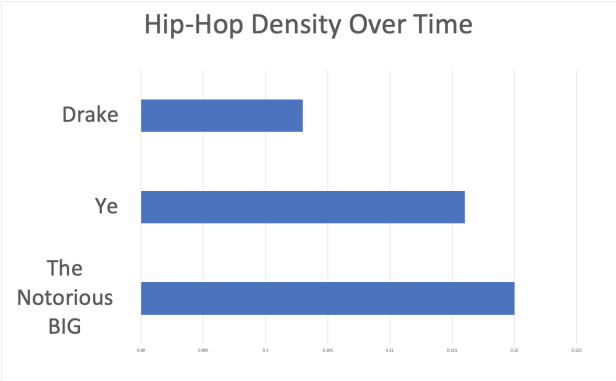
Artist	Unique Words	Total Words	Ratio (Density)
Biggie	5,053	41,944	0.120
Ye	4,221	36,340	0.116
Drake	3,281	31,846	0.103
Michael Jackson	3,249	49,895	0.065
Britney	49,895	1,752	0.062
Beiber	29,531	1,965	0.067
Hendrix	14,274	1,840	0.129
Nirvana	12,345	1,477	0.120
Blink-182	19,746	1,845	0.093
Dickinson	29,692	5,636	0.190
Dr. Seuss	6,343	845	0.133
Nursery Rhymes	17,062	2,403	0.141

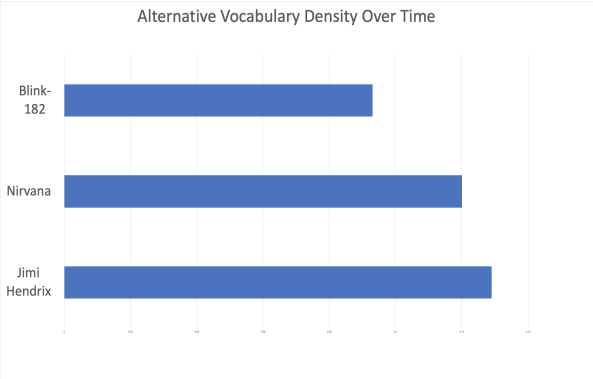
With Voyant, we were able to get interesting data on vocabulary density and sentiment that helped us answer our questions.

To answer these questions surrounding vocabulary density and unique words, we extracted vocabulary density measurements and compared them with singular artists across genres and through time. We chose three artists from each genre, with each artist being progressively newer than the previous. Further, as a baseline to compare to, we also extracted similar data from poetry over time.

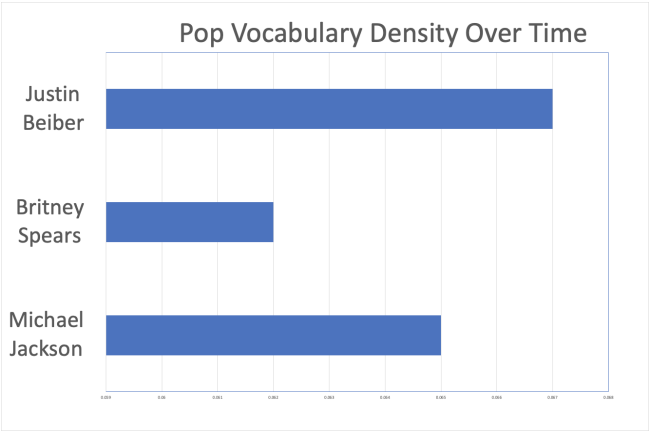
Using the measure of vocab density, which is a statistic that takes the number of unique words divided by the total number of words, we quantified the vocabulary of certain artists.
What we discovered in our study was that genres that have progressively become more popular and mainstream (Rap, Alternative/ Rock) throughout their existence display a decreasing variety of

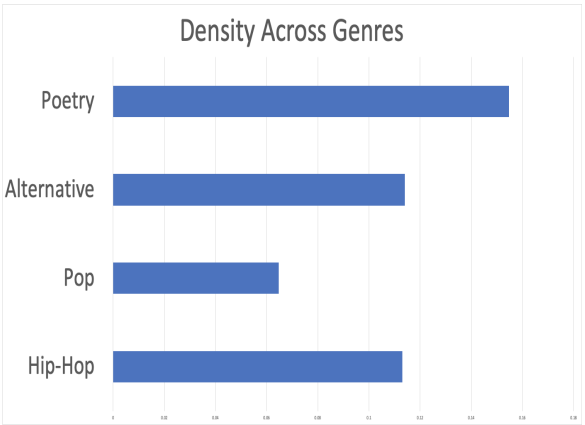
vocabulary as artists become closer to modern. Examples of this would be someone like Drake, a modern hip-hop artist with extremely high radio play and wide appeal, who uses less unique words relative to total words. This could perhaps allude to more simple song structures or a conscious effort to include vocabulary that has a wider appeal.





In contrast, the genre of pop did not have this trend. Rather, as pop artists became more modern, there was no vocabulary trend. This could possibly suggest that since pop by nature appeals to the widest audience, these simpler song structures and more consistent themes lead to lesser variety in vocabulary. Further, this is backed by the idea that in our study, pop music had the lowest average vocabulary variety compared to other genres, while poetry understandably had the highest.





Overall, our conclusions from the study using Voyant are that for our selected artists, there seems to be a negative correlation between vocabulary density and time within non-pop genres. Furthermore, of the most popular genres, rap has the most vocabularic variety.

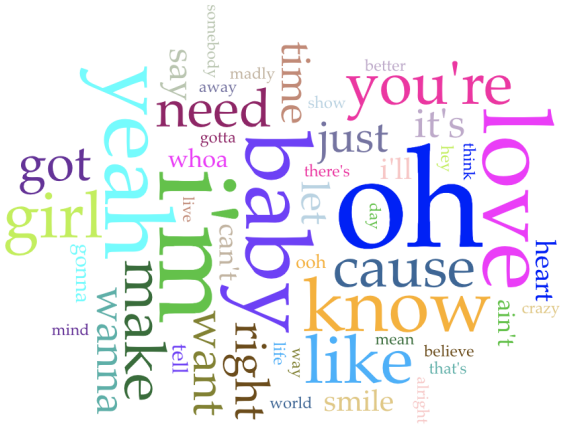
Comparison

Our experience with Voyant was two-fold: while we really enjoyed the visualization opportunities within the platform, we found it hard to draw real conclusions from. Word clouds can reveal a lot about an artist, but we found Voyant more useful for tracking density than sentiment analysis. Due to the ability to have more personalization and curation towards solving our exact question, we preferred using R in this project due to its functionality in this regard.

However, the simplicity of Voyant made it a seamless process to extract figures on unique words and create a graph comparing them. While each had pros and cons, we found it best to focus on Voyant for unique words and on R for sentiment analysis, given that R's sentiment packages are so robust.

Voyant Sentiment

To do some preliminary work on sentiment with Voyant, we found it most salient to compare word clouds rather than draw conclusions from a single one. For example, when comparing Beiber (left) to The Notorious BIG, one can draw the conclusion that the themes of the two artists greatly differ. Beiber's music discusses need, longing, and love, with far more romantic themes than Biggie Smalls, who is more in line with the stereotypical hip-hop themes of vulgar language, self-confidence (I'm), overall offensiveness and obsessions with money and women. Less romantic and widely appealing, far more raw, real, and unfiltered emotion with violent themes. However, these are just broader interpretations. For the data, we turn to R:



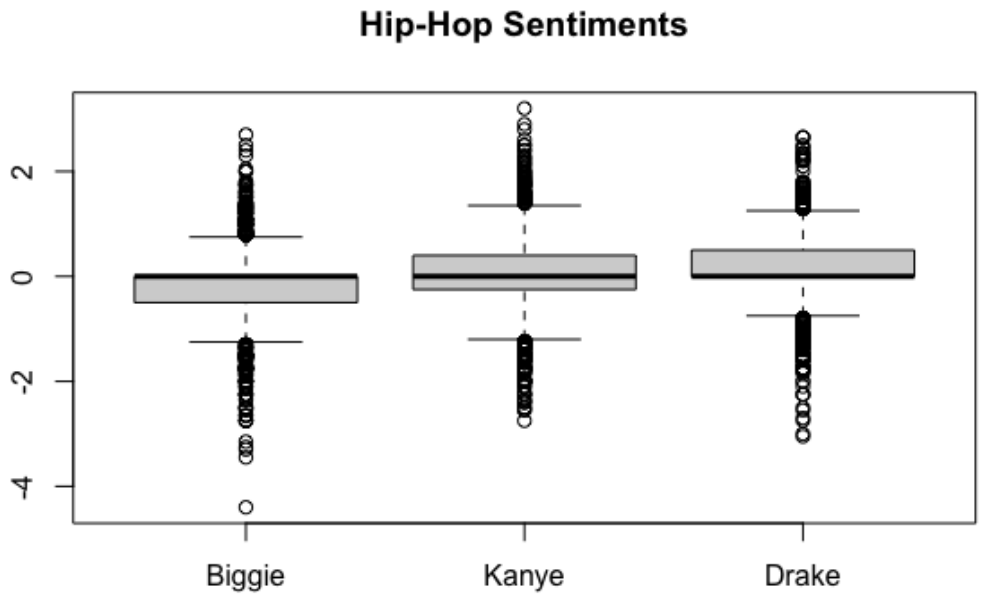
Analysis in R

To analyze sentiment in R, we were able to analyze individual word sentiment and lyric-by-lyric sentiment. For each word, the NRC package groups each word into one of eight categories: Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, and Trust. Additionally, the package also finds whether word choices appear positive or negative, and keeps a running count of the words that fit all of these sentiments. In our project's Github repository, we put two bar graphs for each artist, one representing the number of positive and negative words and another representing the respective word counts for each emotion. On top of this, when given an array of sentences, or in this case lyrics, the 'get_sentiment' function assigns numerical sentiment values, where positive values represent positive sentiments, and vice versa. To make use of this functionality, we split each text file by line and passed each resulting array as a parameter to the get_sentiment function. From this result, we then calculated the mean (average) sentiment of each artist's lyrics, as well as the minimum, maximum, and the range from the 25th to 75th percentiles. Below is the most frequently-used word sentiment, positive/negative polarity, and the mean lyric sentiment for each of the twelve artists we looked at:

Artist	Most Frequent Word Sentiments	More Positive or Negative Words?	Mean Lyric Sentiment
Biggie	Fear, Anger, Trust	Negative	-0.16
Kanye	Trust, Fear, Sadness	Negative	0.006
Drake	Trust, Anticipation, Fear	Positive	0.073
Michael Jackson	Fear, Sadness, Trust	Negative	0.052
Britney Spears	Fear, Trust, Anger	Positive	0.073
Justin Bieber	Trust, Joy, Anticipation	Positive	0.19
Jimi Hendrix	Fear, Anticipation, Trust	Negative	0.15
Nirvana	Trust, Sadness, Fear	Negative	-0.005
Blink-182	Fear, Sadness, Anger	Negative	-0.14
Emily Dickinson	Trust, Fear, Sadness	Positive	0.018
Nursery Rhymes	Anticipation, Joy, Anger	Negative	0.091
Dr. Seuss	Trust, Joy, Fear	Positive	0.020

From performing this analysis, there are a few important observations that we made. One key point we noticed is that despite stigma around various genres, none of the four genres we looked at displayed entirely positive word choice and lyric sentiment. In fact, several musical artists displayed more positive lyrics than the children's literature we looked at. There was also some interesting overlap in frequent word sentiments between artists from very different genres. For example, Kanye West's lyrics had the exact same most frequent emotions, Trust, Fear, and Sadness, in the exact same order no less, as 19th-century poet Emily Dickinson. That said, some of our findings in lyric sentiment were more unsurprising, including the angst of Nirvana and Blink-182's lyrics leading to relatively negative sentiment scores or the upbeat lyrics of Justin Bieber leading to a positive sentiment score with mostly positive word choices as well.

To better explore some trends over time within genres, we plotted each of the artists' lyric sentiment scores in a box plot with the other artists from their respective genre. A few interesting trends include Hip-Hop getting relatively more positive from Biggie to Drake, while rock/alternative becoming more negative in sentiment from Jimi Hendrix to grunge and pop punk artists. An example of one of these boxplots is given below, and all four of these boxplots are also available in the Github repository as well.



Conclusion

From performing this text analysis, the main takeaway we had is that music artists and genres cannot be boiled down to stereotypes and assumptions. Most of our predictions were based on the general perception of each genre, like hip-hop being more influenced by violence and negative themes or rock and pop having simpler, more repetitive lyrics. Instead, we found that while some artists fit these general trends, many did not, leading to surprising similarities between rappers and poets or rock artists and pop stars. While perhaps overgeneralizing in some cases, R and Voyant both proved to be valuable tools in analyzing changing trends in music lyrics and sentiment.