Lyrical Transformation of

Billboard HOT 100

DH 100 Theory and Methods with Professor Adam Anderson | Alisha Lewis | 30 June 2021

"Music expresses feeling and thought, without language; it was below and before speech, and it is above and beyond all words." - Robert G. Ingersoll

Outline

- ★ Research Questions
- ★ Project Importance
- ★ Jupyter Notebook
- ★ Github Repository
- ★ Poster
- ★ Conclusions
- ★ Works Cited

Research Questions

★ How have Billboard Hot 100 lyrics transformed to reflect society since 1950?

★ How has technology become prevalent in popular music lyrics?

★ Have swear words become more popular in Billboard Hot 100 lyrics?

★ How have the most common lyrics changed?

★ Are lyric phrases more or less unique in current times?

★ What are other general shifts in music lyric trends?

Why it Matters

Music tends to reflect the current state of society. As our society has gotten more progressive and open, so have our popular music lyrics. By looking at how lyrics have changed since 1950, an arugment can be made regarding what is considered acceptable in the United States. For example, if technology is mentioned more often, it could be reflecting the rise of technology in our society. By analyzing different themes in in popular music lyrics, commentary on the values of our society can argued.

billboard THE HOT 100

Jupyter Notebook Experience

★ Adapted from Eunice Chan's work on Billboard Hot 100 and Rishabh Parekh's web scraping work

★ Difficult to web scrape from Genius

★ Trouble scraping songs with featured artists and older songs

★ Self generated dictionaries for technology and swear words

★ Useful for creating visuals and tables that show trends

```python
for word in song:

    #Count number of words in first person singular
    if word in first_person_single:
        words_first_single += 1

    #Count number of words in first person plural
    if word in first_person_plural:
        words_first_plural += 1

    #Count number of words in second person
    if word in second_person:
        words_second += 1

    #Count number of male pronouns
    if word in male:
        words_male += 1

    #Count number of female pronouns
    if word in female:
        words_female += 1

    #Count number of technology references
    if word in technology:
        words_tech += 1

    #Count number of swear words
    if word in swear_words:
        words_swear += 1
```

GitHub Repository

alishalewis Update README.md — ce307f1 4 days ago — ⏱ 25 commits

📄 Alisha Lewis DigHum100 Poster - P...	Add files via upload	4 days ago
📄 Data_Analysis (1).ipynb	Add files via upload	4 days ago
📄 Extracting_Billboard_100_Genius (1)...	Add files via upload	4 days ago
📄 LICENSE	Initial commit	18 days ago
📄 README.md	Update README.md	4 days ago
📄 alisha data (2).csv	Add files via upload	4 days ago
📄 results with years.csv	Add files via upload	4 days ago

README.md ✏

Welcome! I am currently working on a DigHum100 project analyzing the lyrics of US Billboard Hot 100 Songs from 1950-2018. I am a recent UC Berkeley grad with a Bachelor of Arts in Cogntive Science and a minor in Data Science. Check out my google colab notebook for this project here.

★ Serves as a great place to store documents
★ Useful for sharing with peers
★ Collaborative workspace
★ Good backup in case files are lost
★ Peers can elaborate and create further work

Lyrical Transformation of Billboard HOT 100

DH 100 Theory and Methods | Alisha Lewis | 27 June 2021

Project Work Flow

- Upload 'Billboard 100(1950-2018).csv' to Jupyter Notebook
- Webscrape genius.com for Lyrics Using Beautiful Soup
- Perform Basic Data Analysis
- Preprocessing Lyrics Without Punctuation, Lyrics by Phrases, Lyrics by Words with NLTK
- Analyze Unique Artists, Artist Frequency, Unique Phrases, Average Character and Word Count
- Analyze Frequency of Parts of Speech, Speculative Words, Personal Pronouns, and Lexical Diversity
- Explore Freuqency of Technology and Swear Words
- Network Graph and Visualizations
- Sentiment Analysis (By Phrases)
- Explore Common Words, Direct Count (Without Stopwords), TF-IDF, Topic Modeling
- Collaborate and Share Work

Introduction:

The content of popular music has changed drastically over the years and often highlights the beliefs of our society. Through looking at American music lyric trends, values and cultural shifts can be analyzed. While studies have concluded that music has gotten more negative and less about love, this project aims to analyze how much more music mentions technology and swear words, along with more general analysis. America has progressed to be more accepting of these topics in popular culture, and this study hopes consider if this is reflected within popular music lyrics.

Research Question: How have Billboard Hot 100 Lyrics Changed Since 1950 to Reflect Societal Values?

Description of Data Set:

This project will use a dataset on Billboard Hot 100 songs from 1950-2018 created by Eunice Chan. This dataset contains the Billboard Hot 100 song title and artist for every year from 1950 till 2018. Eunice manually inputted lyricis for the top two songs of every year creating 136 data points, but this project relies on having more data for more detailed analysis. To learn how to web scrape lyrics from genius.com, work from Rishabh Parekh's project will be adapted to this project. Using Parekh's technique, the lyrics for the songs in Chan's csv file were scraped. Some lyrics were unable to be found so there are 2,540 song lyrics in the expanded data set. This dataset includes the lyrics along with the artist, song, and year. While Chan's and Parkekh's work has insipred much of the organization and basic data analysis, this project hopes to expand by looking at specific themes in a wider dataset of lyrics such as technology and swear words. With a much larger dataset, this project is able to produce more reliable results that show better trends about lyrics have changed generally. To analyze specific shifts in references to technology and swear words, a dictionary was manually created to track how many times those words appear in the lyrics. Using data analysis, NLTK, and visualizations, this project will highlight how Billboard Hot 100 lyrics have changed since 1950, reflecting a shift in societal values.

Unique Phrases vs Decade
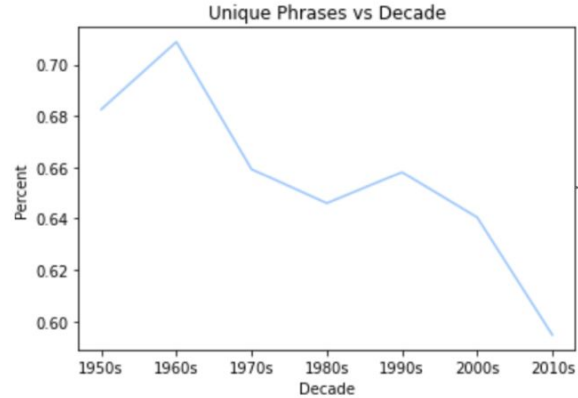
Interpreting Results:

In the graph "Unique Phrases vs Decade," there appears to be a steep decline in the amount of unqiue phrases since 1950. While there was a slight bump in 1990, there is an overall downward trend. In the words clouds for the 1950s and 2010s, there are both similarities and differences. While words such as "love" and "baby" are not the most popular words by 2010, they are still common enough to be featured in the word cloud. There are words that disappeared completeley from the most common words since 1950, such as "rock" and "kiss." Additionally, words have appeared in the 2010s that were not present before like "ain't." In the "Technology References" graph there are two spikes in the 1970s and 2010s. The 1990s appear to have the lowest amount of technology references. Looking at the "Use of Swear Words" graph, there is a steep incline of swear word usage starting in the 1990s. It seems to have leveled off slightly in the 2010s, but there is still an increase. Prior to that, there was a slight increase in the 1970s.

Discussion of Results:

It could be hypothesized that the trends these graphs are correlated to shifts in our societal values. As our society has become more progressive and open, those shifts have been reflected in our popular music. A possible explanation for a decrease in unique phrases is the rise in rap music. This genre uses shorter, more repetitive phrases and began to be popular around the 1970s. When considering shifts in the most common words, it can be assumed that this is due to changes in our language. For example, it would make sense that "ain't"did not appear in the 1950s most common word cloud because it was not apart of the slang at that time. The lyrics seem to be less about love, suggesting that the US has become less romantic and could be focused on other topics such as sex. Furthermore, words like "rock" may be less common because of a shift in genre popularity. Rock was the most popular genre in the 1950s, but hip hop popularity has risen in 2010. When considering references to technology, there is a spike in the 1970s, which is when the first personal computers were created. There is another spike in the 2010s, which could be explained by the more common use of cell phones. It is unusual that there were more references to technology in the 1950s and 1960s than there were in the 1990s and 2000s. It is possible this could be due to the selection of words in the technology dictionary. The use of swear words starkly increased in the 1980s, which could be due to the popularity of rap starting int the 1970s. Swearing also suggests a more casual societal standards.

Most Common Words of the 1950s

Most Common Words of the 2010s

Technology References

Use of Swear Words
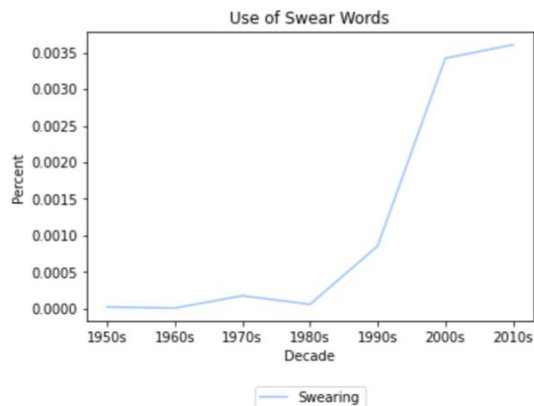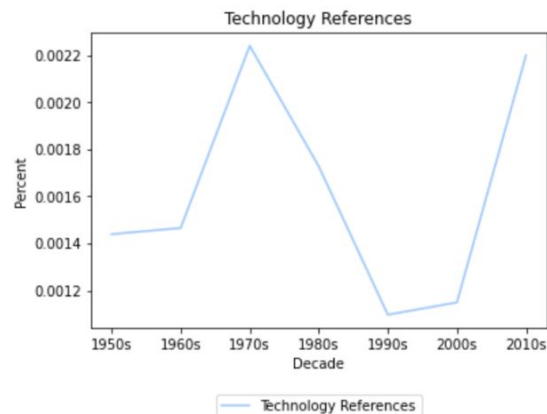
Conclusions:

As discussed, the song lyrics of popular music can indicate shifts in genre popularity, societal values, and historic events. Through analyzing trends in Uc popular music lyrics, the rise of technology, rap, slang, and a more progressive society has been highlighted. Future work could include expanding the dictionaries to include more words of association or exploring topic modeling to look at technology and swearing. If topic modeling were to be adapted to this, more topics could be explored. Further studies could also look at genre, gender identification of artsists, and use of slang. This project could be improved with a larger dataset. While this dataset is much larger than the original, the webscraper failed to get the lyrics for many songs that had featured artists. Featuring has become much more popular in more recent times and is more common in certain genres, so this could have skewed the results. Additionally, some of the older songs lyrics were not on genius, so there is missing data there. The dataset could be expanded to twice it's size if these bugs in web scraping can be fixed.

Works Cited:

Github Repo
More Negative Lyrics
Chan, Billboard Hot 100
Parekh, DH Project
History of Computers
History of Rap
Coding help and feedback from Luc Le Pottier, Ash Tan, and Professor Anderson

Billboard

Conclusions

★ Popular music lyrics show shifts in societal values and reflect historical trends
★ Lyrics were about technology in times of invention such as the computer and cell phone
★ Swear words have become much more common in our lyrics, reflecting a more casual atmosphere
★ There are less unique phrases in lyrics, demonstrating that popular music is more repetitive

Video

Jupyter Notebook Experience

★ Adapted from Eunice Chan's work on Billboard Hot 100 and Rishabh Parekh's web scraping work
★ Difficult to web scrape from Genius
★ Trouble scraping songs with featured artists and older songs
★ Self generated dictionaries for technology and swear words
★ Useful for creating visuals and tables that show trends

```python
for word in song:

    #Count number of words in f
    if word in first_person_single:
        words_first_single += 1

    #Count number of words in first person plural
    if word in first_person_plural:
        words_first_plural += 1

    #Count number of words in second person
    if word in second_person:
        words_second += 1

    #Count number of male pronouns
    if word in male:
        words_male += 1

    #Count number of female pronouns
    if word in female:
        words_female += 1

    #Count number of technology references
    if word in technology:
        words_tech += 1

    #Count number of swear words
    if word in swear_words:
        words_swear += 1
```
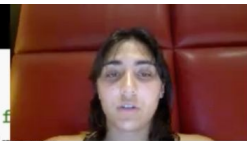
Works Cited

Thank you to:

<u>Eunice Chan</u>

<u>Rishabh Parekh</u>

Luc Le Pottier

Prof. Adam Anderson

Ash Tan

Images:

<u>Slide 1</u>

<u>Slide 2</u>

<u>Slide 3</u>

<u>Slide 4</u>

<u>Slide 9</u>

<u>Slide 10</u>