

## OVERVIEW

# Project iMuse: Popular Lyrics in Popular Songs

IMUSE: JACK ANSTEY, ANNA LU  
APRIL 13TH, 2023

Project iMuse provides the unique ability to view the most used words in popular music over the past 50+ years in conjunction with how they were used in songs through sentiment analysis.

Project iMuse further provides the ability to look at subsets of popular words by allowing a user to change the range of years considered or to look at only a subset of word types which includes: names, foods, colors, and swear words.

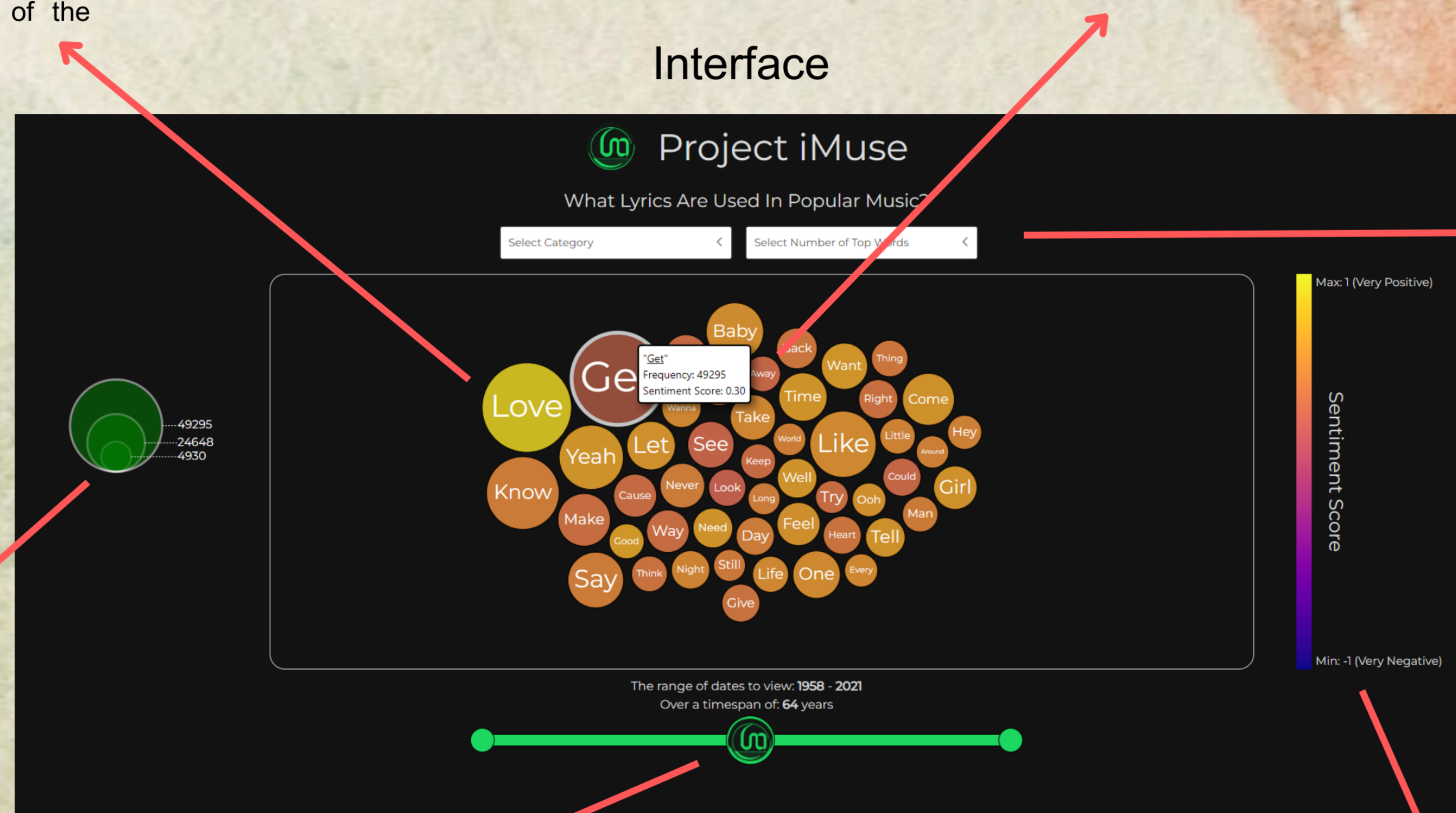
## Research Questions

- What types of words are commonly used in popular songs with positive versus negative sentiment and how do they change over time?
- How does the language used in popular songs differ between different time periods or musical eras?
- Is there a correlation between the use of certain words in popular songs and their commercial success?

The bubble size reflects the frequency of the word used during the input year range, while the color encodes for the sentiment score of the words' context

The pop-up tooltip represents the total frequency of the on-hovered word during the given year range. The sentiment score is calculated through a proposed formula that takes the contextual sentiment score of the word and times the frequency of the word being used for that given year range divided by the overall frequency.

The legend associates the bubble sizes with dynamically adjusted frequencies for the input year range such that the users can perceive the frequency of words accurately and understand the scale of the bubbles readily.



We implemented a preset word category as well as the limits for the number of word bubbles displayed on canvas. The preset categories include Vanilla, Name, Food, Profanity, and Color. The max limit of displays is 100 while the min is 10.

The left and right of the slider allow the user to choose a start and end point for a given time range and the middle handle is for the user to drag through the timeline for that fixed time interval to see different displays of word bubbles and monitor the size and sentiment score oscillate.

We used a color-blind friendly gradient for color legend to reflect sentiment scores

## Dataset

- We used Python, Genius, and Spotify's API to gather lyrics and publication dates for popular songs from the Billboard Hot 100 between 1958 and 2021.
- We collected lyrics for 23,195 out of the 29,680 unique songs in the dataset.
- The resulting dataset had two columns: publication year and lyrics, and was further processed for visualization.

## Methods

- Lyrics Cleaning:
  - Remove Punctuations
  - Expand Contractions
  - Remove Stopwords (NLTK)
  - Lemmatize Words (NLTK)
- Sentiment Analysis were performed using VADER tool, assigning a compound sentiment score ranging from -1 to 1.
- A weighted average sentiment score was calculated for every word each year.

## Example Finding

- Various assumptions can be concluded depending on one's usage of the webpage. Potential findings can be:
  - There is a discernible negative trend in the lyrical content of songs over the years.
  - A comparative analysis reveals that the sentiment polarity of language in older music exhibits greater contrast in comparison to that of more recent music.
  - The color "Blue" is the most popular color in lyrics throughout years.

## Reflection & Conclusion

- Potential improvements includes:
  - Remove more general words and only leave ones with contextual meaning
  - Add ability to search individual words
  - Facilitate a smoother transition in the visualization of the variation in bubble size when scrolling through timeline.
- Overall, iMuse allows users to customize their experience by exploring subsets of popular words based on various criteria, and has the potential to offer insights into the evolution of language and cultural trends in popular music.