Hip Hop Talks: An Analysis of Past and Present Hip Hop Lyrics

Yodahe Alemu

Massachusetts Institute of Technology

ykinsew@mit.edu

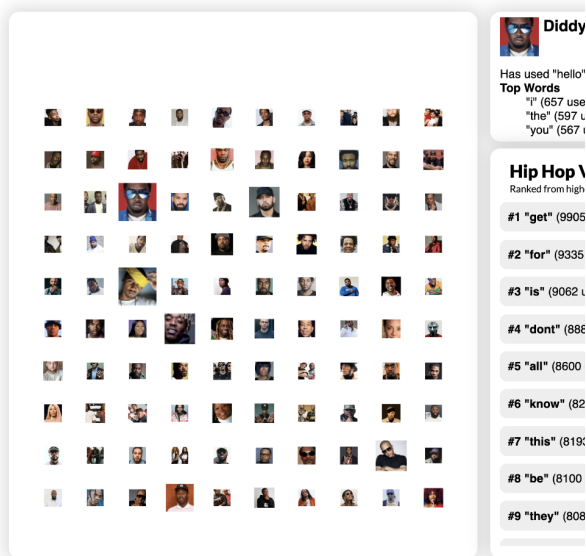Joshua Verdejo

Massachusetts Institute of Technology

jover@mit.edu

1 Introduction

Hip hop has defined our culture for generations, starting as a form of self expression and protest [1], and becoming one of the most popular styles of music in the world, while still retaining its roots [2]. While hip-hop is a wildly popular form of music, it's not often that we take a more critical look into the words the phrases in the context that many of our favorite artists are using.

Now some "hip-hop scholars" may do this more often, however for the majority of us, if the melody is catchy and we know the chorus, it's good enough to sing at concerts or parties. Because of this, our visualization hopes to shed some more light into the words that are being used by our favorite artists, in our favorite songs. However, more importantly, this visualization tool allows users to look at trends of usage overtime, allowing users to see when words became popular, and when they died off. Hip-hop talks also hopes to inspire users to want to create their own lyrics, and in doing so the goal of the tool is also to work with data visualization and allow users to create their own music by guiding them through the lens of an artist of their choosing.

After careful consideration, our group decided to focus on creating this visualization because we both enjoy hip hop. Also, we wanted to create a way for people in our generation, where short attention spans reign supreme, to also be able to enjoy slowing down and understanding what words popular artists use, how they have changed, and gather a deeper appreciation for hip hop in general.

2 Related Work

In this section, we discuss a brief overview of the history of Hip Hop and how it has previously impacted culture. We then discuss how breaking down hip hop lyrics allows for deeper appreciation of concepts we hoped to illuminate in our visualization.

2.1 History of Hip Hop

Hip hop started in the Bronx [3], and is often broken down into five primary categories: MCing, DJing,

1

Breakdancing, Street Art, and Knowledge [4]. Over time, there has been a significant shift from the traditional hip hop pillars, to a much more significant focus on MCing, DJing, and knowledge, as the largest hip hop stars are often lauded for their lyrical prowess and music genius.

2.2 Deconstructing Hip Hop

Because of this shift in focus, understanding the words being spoken is that much more important. It is because of this shift in focus that many youtube series have taken off based on the analysis of modern hip hop lyrics [5] and production [6]. With the advent of better technology, visualizations have also improved, and discussions have become more collaborative. Websites such as Genius [7] have allowed for multiple inputs as to how hip hop lyrics should be interpreted, and visualizations such as many of those debuted in The Pudding [8] have created graphics for us to look at hip hop in a new light. In this spirit, our visualization works to bring modern visualizations to both old and new school hip hop by analyzing trends in word usage over history.

3 Methods

Our development process can be broken down into three stages - data wrangling, data manipulation, and data visualization.

3.1 Data Wrangling

Data was created by scraping the Genius database. Multiple passes of the database were required, as none of the information needed previously existed in the database. There were other APIs pursued as well, such as AZlyrics, and Lyrics.com, however, both of these unfortunately had limit requests that made it difficult to request the information needed, as requests were made on a song by song basis. Genius allowed for requests to be made on an artist by artist basis, which was much more suited for our implementation.

3.2 Data Manipulation

There were multiple methods we used to separate our data. First, we generated all lyrics from the top 50 songs from 100 hip hop artists. From there, we stripped the data to get rid of information that was not necessary, such as apostrophes, dashes, and song notations such as "chorus" and "intro". After this was done, we summed the number of occurrences of each song, for each artist, and then added all of these together in order to generate a comprehensive word usage list. For the individual datasets, the same procedure was followed, except instead of adding all songs together we only added all songs by a given artist.

Next, the temporal information was handled in a similar fashion, however, the website had to be scraped more for more information. Most songs belonged to albums, and each album had its own release year. Using this information, we matched songs with the albums they belonged in in order to know what years should be used for each song. Finally, for the word recommender, the lyrics were redistributed into pairs, and arranged in a trie-esque structure, such that each word pointed to a dictionary of words that followed the key, along with the probability of that word following the keyed word.

3.3 Data Visualizations

For the data visualizations, we used d3.js in order to both build the website as well as the visualizations on the page. This includes the grid of artists which illustrates how much each artist used a given word, the bar chart indicating the most likely next word, and the bar chart detailing word usage over time.

User interactions took one of three forms - typing into an input bar, clicking on vocabulary words, or hovering over charts to bring up a tooltip. The actual website is implemented such that the majority of changes occur as a result of the user entering a word. In this way, the typing and vocabulary words serve as methods to mutate the data presented to the user, whereas hovering (which is naturally a more passive interaction) serves to explain the data that is already presented.

Typing a word, or clicking on a word on the vocabulary chart, will shift the artist matrix, the word usage chart, and the word suggestion chart. Our intended main interaction was just the ability to type in a word, as this created a more fun and playful experience with the visualization. However, after user testing we realized that it would be helpful to include a vocabulary list which would help give insight to what words were actually included in our dataset.

The hovering interaction, when used in the respective bar graphs, will create a tooltip that displays the number of times a word was used by all artists in a year, the number of times a word was used by a specific artist in a year, or the frequency by which a word occurs after the selected word. The hovering can also be used in the artist matrix, such that hovering over an artist's square will show more information about them in the artist description box. This more passive interaction style is simply used to bring up further explanations about the data that is currently displayed on screen.

We simplified our color palette to only include three colors - black, white, and green. Designing around such a simple color palette allows us to make sure that the visualizations use very clear methods for displaying information. We're also able to use the green color, which stands out very well against both the white and black, as a way to draw special attention and link different aspects of the visualization. The color green in our visualization typically corresponds to data that is artist-specific. For example, when a user selects one of the artists in the grid, that artist gains a green outline; and the bar graphs shift to the green color in order to indicate that they are now displaying information specific to the selected artist.

4   Results

In this section, two case studies are examined: one in which a user wants to look at historical trends, and another in which a user is looking at creating a song.

4.1   Historical View on Slang

In this example, a user wants to understand the historical significance of the word "yo". They are interested in seeing if the word was used more frequently in the past, as they realized they had not heard the word as often in recent music. By simply typing "yo" in the text input section, the Word Usage Over Time section dynamically updates to the figure shown below.

Word Usage Over Time
How has the word "yo" been used over time?
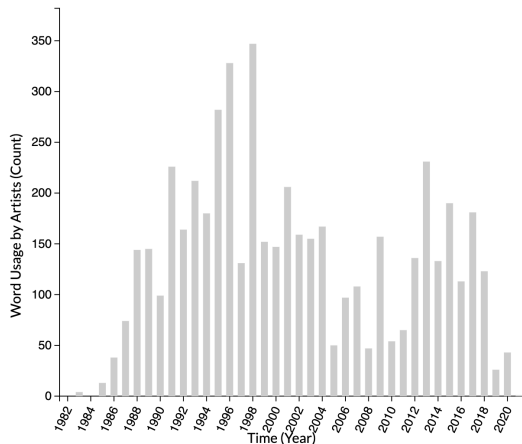
Here, the user can tell that their suspicions were confirmed, as the word "yo" was used heavily up until approximately the year 2000, where there was a significant dropoff, followed by another drop shortly after. At this point, the user is curious as to the usage patterns of one of their favorite older rappers (Busta Rhymes), and how it compares to a newer rapper they enjoy (Cardi B). By simply clicking on each artists, the graphs update for Busta Rhymes (left) and Cardi B. (right) respectively.

4.2   Word Suggestion Tool
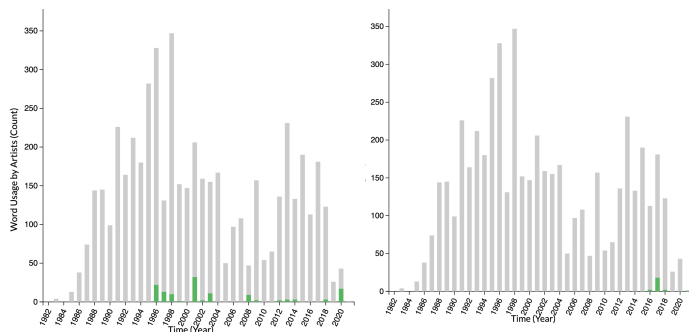
In this example, a user is writing lyrics, and wants to channel the style of one of their favorite rappers, Eminem. The user knows they want to say something personal, so they start with "I feel", and simply by typing into the text input and following some of the suggestions, end with "I feel like you don't want to be the same you", which inspires them to finish the verse.
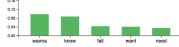
5   Discussion

Hip Hop Talks allows users to develop further insight into the Hip Hop industry and the language that is used in its music than would be possible by simply listening to the original music. Based on the feedback we received from our peers and other smaller groups of peo[ple, our work benefits primarily from its ability to compare artists within the industry and its playful design.

5.1   Artist Comparison

In our original implementation of Hip Hop Talks, we were simply focused on the music industry as a whole and exploring how different words were used; however, after receiving lots of positive feedback about how our visualization allowed users to compare artists, we decided to hone in on this feature, hopefully allowing further fruitful insights into Hip Hop.

Our visualization enables effortless comparison of the language used by multiple artists across the industry from different genres and time periods. Our core method for doing so was the artist grid, which allows users to see how an artist uses a word in direct comparison to other artists. Many users who tested our visualization found this to be a very intuitive and insightful method for expressing this data. Removing the numbers and using a purely visual comparison system - the scale of the artist's image - seems to have been a lot more approachable while still conveying a good amount of information.

The temporal bar chart also received a lot of positive feedback as it was a good way to see how an artist compared numerically to the music industry. Additionally, the bar chart gave users added insight as to how the artist compared to other artists from different time periods.

5.2   Playful Interaction

Our visualization encourages a playful interaction from the user. This starts from the typing input (as we described above) but the smooth and reactive animations and visualizations also play into this feeling. Many people commented on how this visualization came together to create a fun experience which encouraged them to explore further. For example, one of our peers explained that they enjoyed exploring the temporal data and trying to remember the songs of their favorite artists from each year that included the words.

While a lot of visualizations tend to be more

serious and structured, there's a lot of benefit to designing an interaction to be playful and fun. Creating an interactive visualization doesn't provide much insight to a user who doesn't feel compelled to actually interact with the work. Although we had intentionally included these playful design choices from the beginning, we were still happily surprised by the positive response we received.

6 Future Work

In this section, we discuss how each section could be improved and changed in the future, from word suggestions, to visual encodings.

6.1 Word Suggestion Tool

Future implementations of this project could include a bigger focus on the word generation, using natural language processing and machine learning in order to understand what the user has already written and provide more relevant information.

6.2 Customizable Data

We partially implemented a method for allowing users to choose another artist to add to the visualization, but weren't able to fully realize the solution graphically. Most of the code is in the python scripts, so adding that feature could be an interesting extra feature for users.

6.3 Visual Encoding of Information

In the future, there are many changes that could be made to the visual representation of the data. We could introduce color customizers for users, make the text bigger, even make components in the website modular, so users would be able to drag and place components on different locations in the website.

7 References

[1] SpearIt, . "Sonic Jihad—Muslim Hip Hop in the Age of Mass Incarceration." *FIU Law Review*, vol. 11, no. 1, 2015, doi:10.25148/lawrev.11.1.15.

[2] "Hip-Hop Is the Most Listened to Genre in the World, According Spotify Analysis of 20 Billion Tracks." The Independent, Independent Digital News and Media, 14 July 2015, www.independent.co.uk/arts-entertainment/music/news/hip-hop-most-listened-genre-world-according-spotify-analysis-20-billion-tracks-10388091.html.

[3] "DJ Kool Herc DJs His First Block Party (His Sister's Birthday) at 1520 Sedgwick Avenue, Bronx, New York." The Guardian, Guardian News and Media, 12 June 2011, www.theguardian.com/music/2011/jun/13/dj-kool-herc-block-party.

[4] "UHHM Presents the [R]Evolution of Hip Hop Exhibit." Universal Hip Hop Museum, 26 Mar. 2021, www.uhhm.org/revolution-of-hip-hop/.

[5] "The Formula." *NPR*, NPR, www.npr.org/series/822013650/the-formula.

[6] "Deconstructed." *Genius*, 2021, genius.com/shows/deconstructed

[7] "Genius" *Genius, 2021* www.genius.com

[8] Daniels, Matt. "Rappers, Sorted by the Size of Their Vocabulary." *The Pudding*, 21 Jan. 2019, pudding.cool/projects/vocabulary/index.html.