Song Repetition Analysis

Elliot Maringer, Lucas Strauss, Wyatt Denu & Jake Clauson

2025-05-15

## Intro

In today’s music industry, repetitiveness is often cited as a catchy or commercially successful songs. However, defining what makes a song “repetitive” is far from straightforward. Repetitiveness can refer to repeated words, phrases, melodies, or even structural elements such as choruses. In this project, we aim to explore lyrical repetitiveness across Billboard’s year-end top songs list over time, using a combination of methods and lyrical analysis. Our goal is to identify trends in repetitiveness, investigate whether songs have become more or less repetitive over the decades, and consider how different metrics of repetitiveness might provide further insight.

To do this, we built a dataset of Billboard’s top songs for each year since 1946. For each song, we scraped the lyrics from songlyrics.com and extracted metadata from the Spotify Web API. While some data—such as genre—were incomplete, our dataset still allows for a robust analysis of lyrical repetitiveness over time using multiple approaches.

## Methods

To measure lyrical repetitiveness in popular songs, we developed four different analytical techniques. Each method captures a unique dimension of repetition, from word variety to structural reuse, and together they provide a more complete picture of how repetition functions in music over time.

### **Unique-to-Total Word Ratio**

This method helps highlight the variety of words in a song. Values from this method will be in the range [0,1] where a value closer to 0 indicates the song repeated a word many times and a value close to 1 indicates the song used very few repeated words. This method offers a high-level view of lyrical variety and is useful in comparing how word usage and patters may have evolved over time

### **Title Word Repetition**

This method helps highlight a very common trend in songs of repeating the title. The title of a song very often in the focal point or the “message” of a songs so it is often repeated for a greater effect. This approach shows a stylistic trend that is especially popular in pop and radio songs where memorability and recognizably are critical for their commercial success.

### **Line or Verse Repetition**

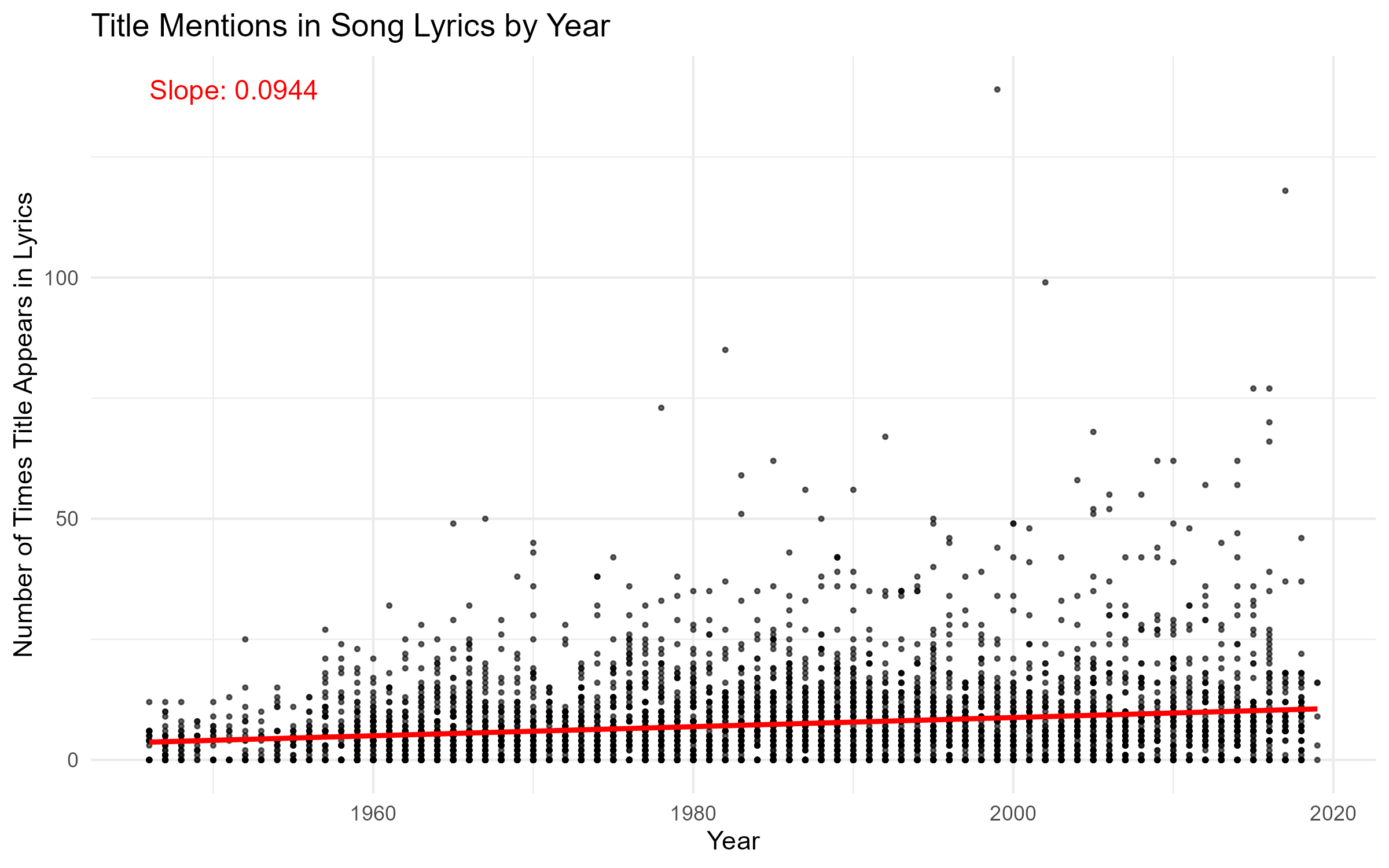
As with the method above this also highlights a very common trend in pop or radio songs. Many songs have a hook or chorus that is repeated throughout the songs to again increase the lyrical affect. The chorus is usually separated by verses which help reduce the songs repetitiveness. By measuring how often linear were repeated we can assess how much or a song’s structure is built around repitition. Songs with highly repetitive choruses or hooks will score high with this metric

### **Most Repeated Non-Stop Words**

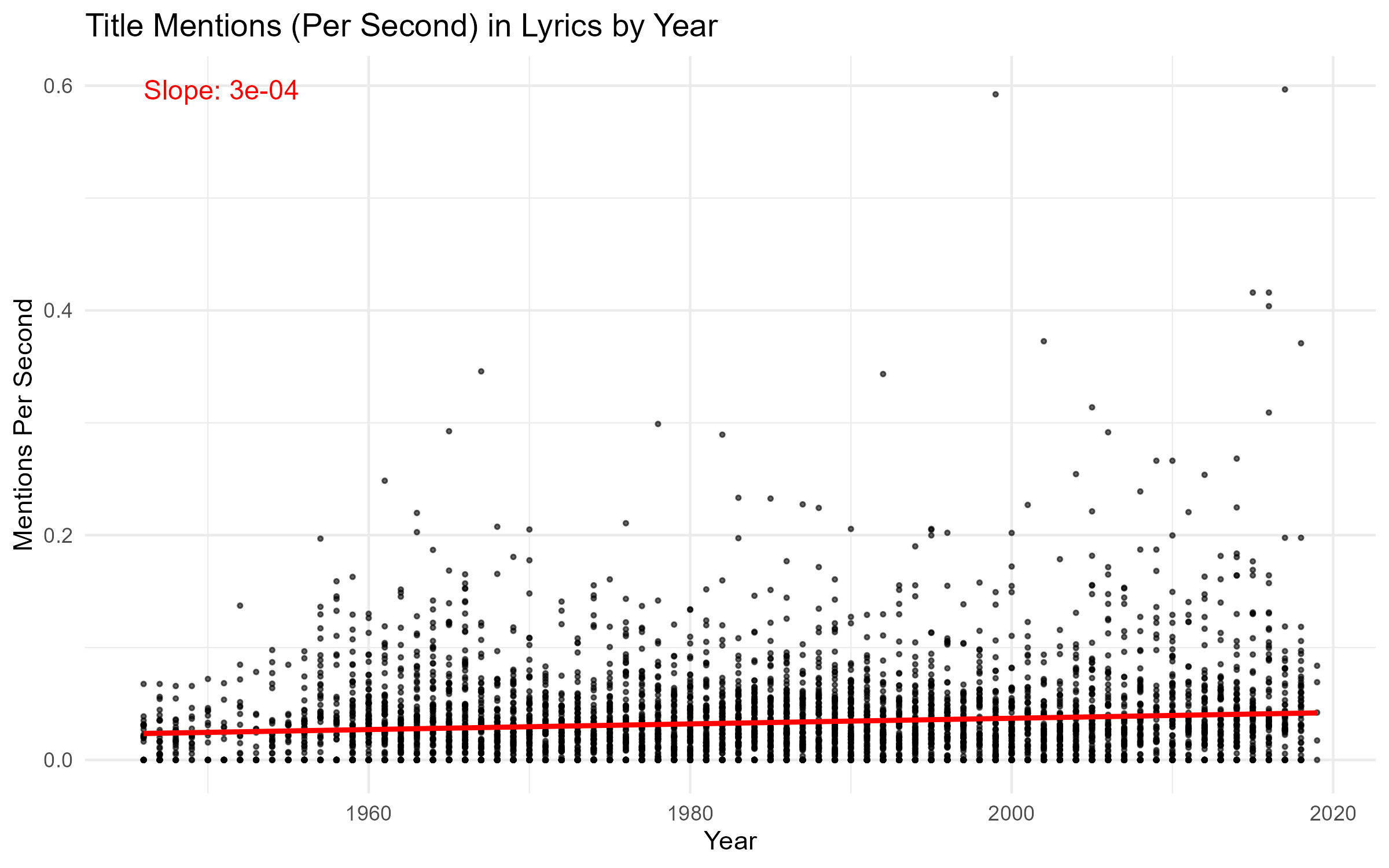
This technique identifies the most frequently used meaningful word in a song’s lyrics. To focus on relevant content, we first remove English stop words like “the,” “and,” or “but” that contribute little to no meaning. This metric reflects how strongly a particular concept, theme, or emotion is emphasized through repetition. For example, a love song might frequently repeat a word like “baby” or “heart,” revealing its emotional focus. By excluding filler words, we better isolate the song’s lyrical intent.

## Analysis

### Title Word Repitition

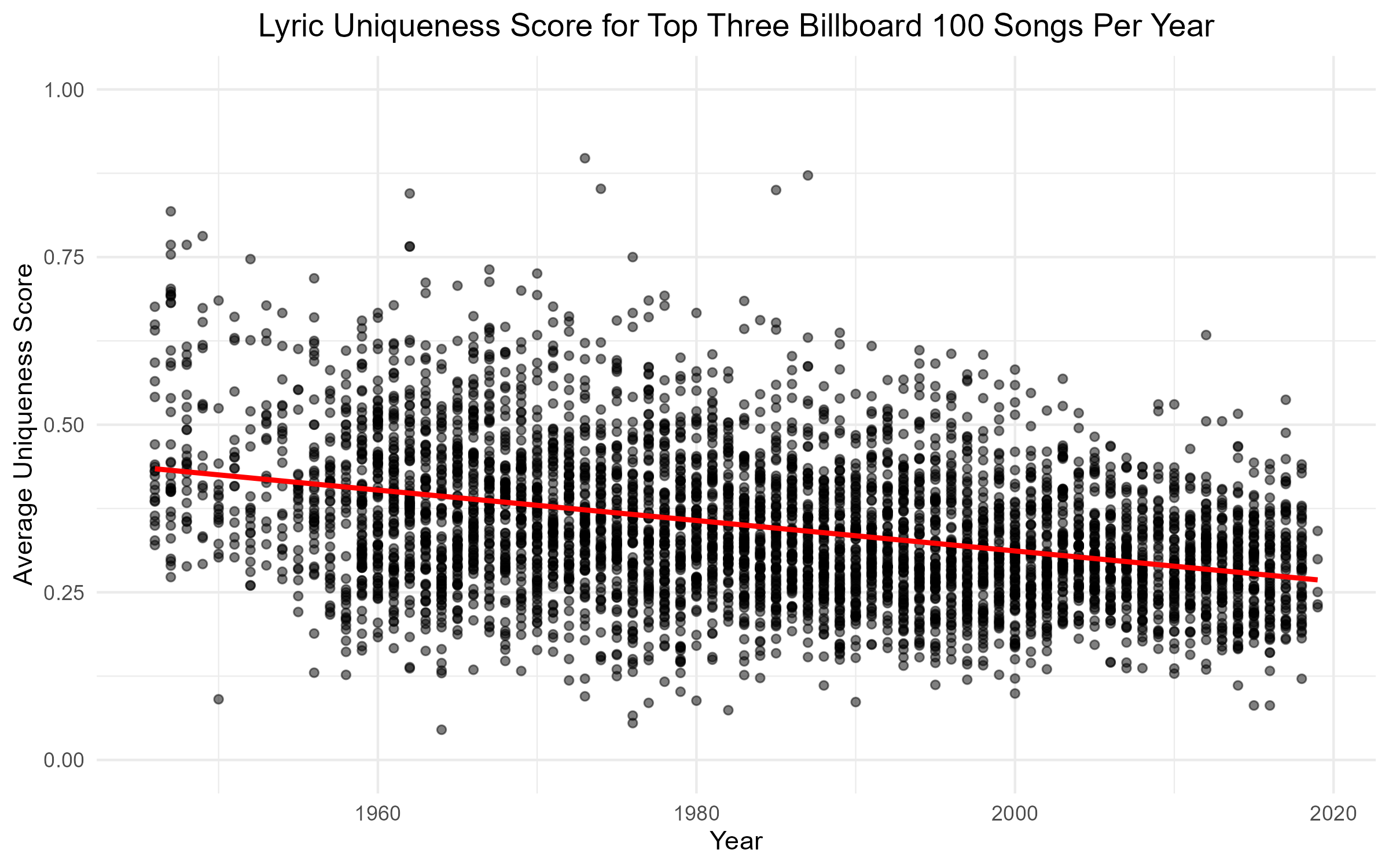


The above plot visualizes the relationship between the number of times a song title appears in its lyrics and the year the song was released. Each point represents an individual song, and a linear regression line was fitted to observe the trend over time. The slope of the line is 0.0944, indicating a slight upward trend on average, newer songs mention their titles slightly more often than older ones. While the slope is small, it still suggests a gradual increase in title repetition over the decades, potentially reflecting changes in songwriting style or audience preferences for more memorable lyrics.



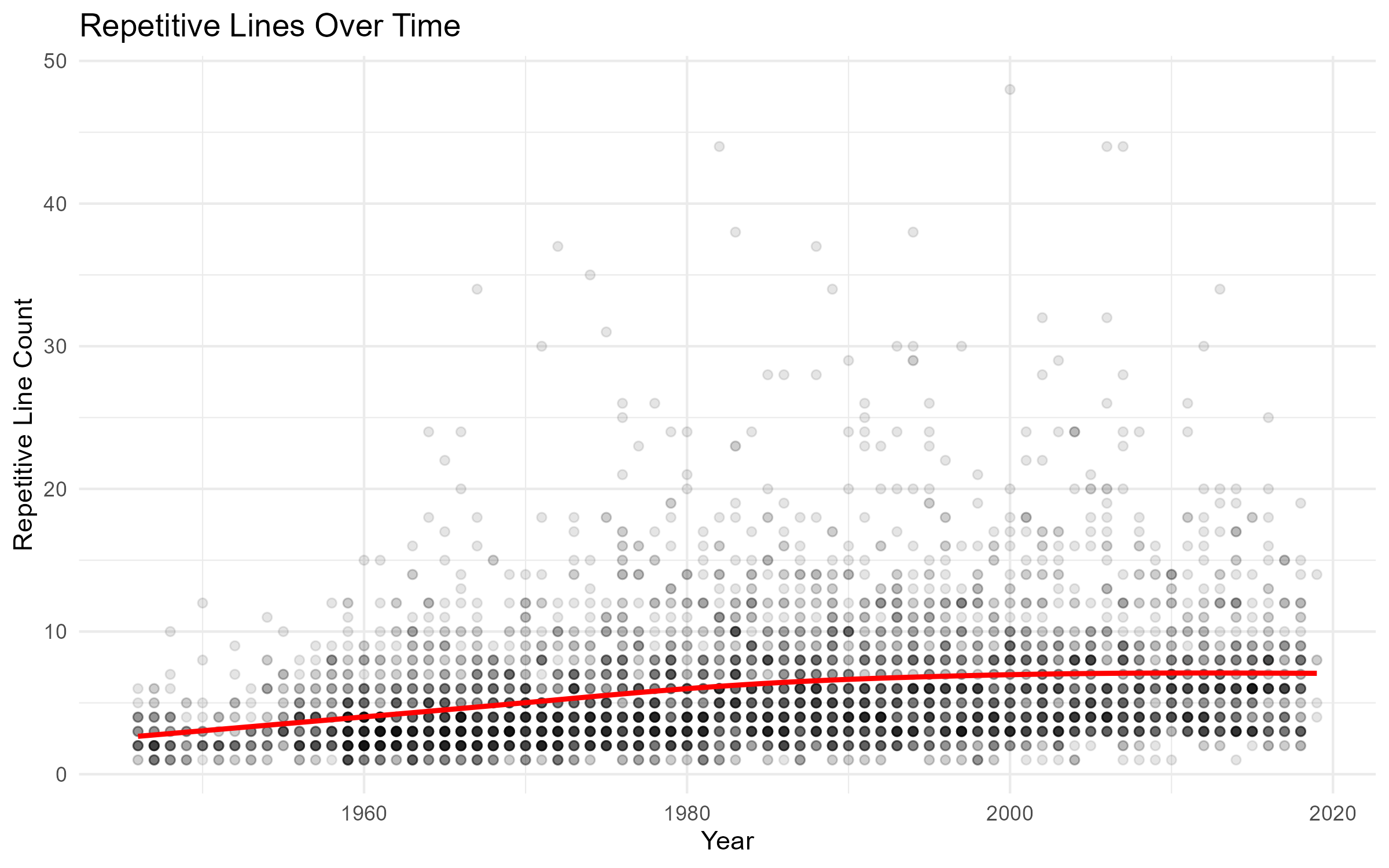
The second plot analyzes title mentions per second, accounting for the length of each song, to determine if repetition has become more concentrated over time. This normalization helps address differences in song duration across eras. The fitted linear model produces a much smaller slope of 0.0003 (3e-04), suggesting an almost flat trend. This implies that although newer songs may say their titles more times overall, the rate at which they do so within the duration of a song hasn’t changed significantly.

### Unique-to-Total

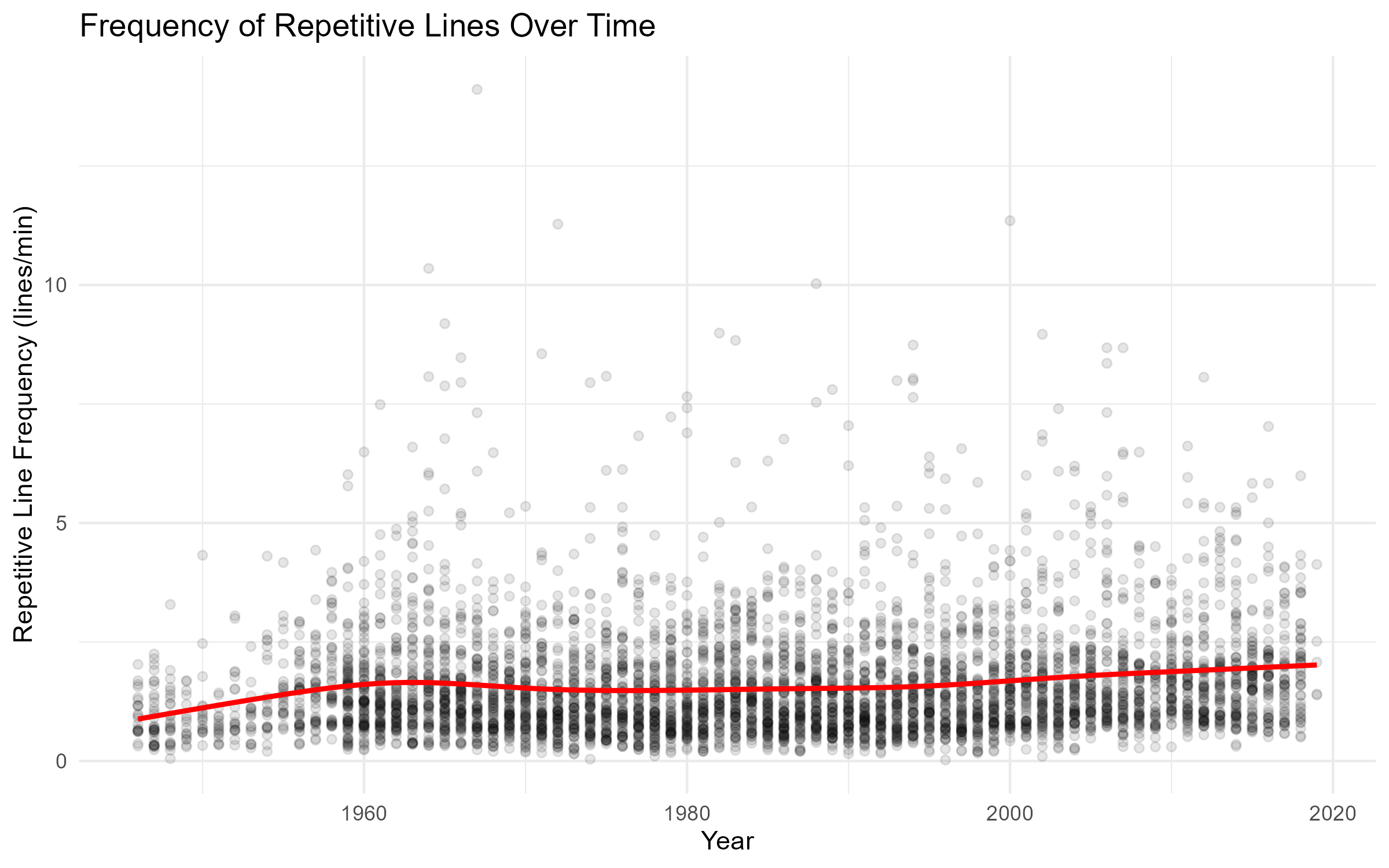


For this graphic produced, the factors that were taken into consideration were the top three billboard 100 songs in each year and how unique the lyrics were for those songs in their respective years. To determine how unique the songs were for each of the years, I computed a uniqueness score that consisted of the total amount of unique words in each song divided by the total number of words in each song. As seen in the graph, there is a moderate negative correlation between song lyric uniqueness and increasing years. Over time, it is safe to say that the amount of unique words that are used in songs that reach the top charts of the billboard 100 have decreased. ### Title Word Repetition

### Line or Verse Repetition



This graph displays the number of times the most common line in a song is repeated over the year it was on the top 100 songs list. Over time, there appears to be a positive correlation between the repetitive lines and the year it was released. Therefore, there is evidence to suggest that lines are becoming repeated more and more as time goes on. However, between 2000 and 2020, there is very little increase in repetitive lines, suggesting that the repetitiveness of songs is stagnating.

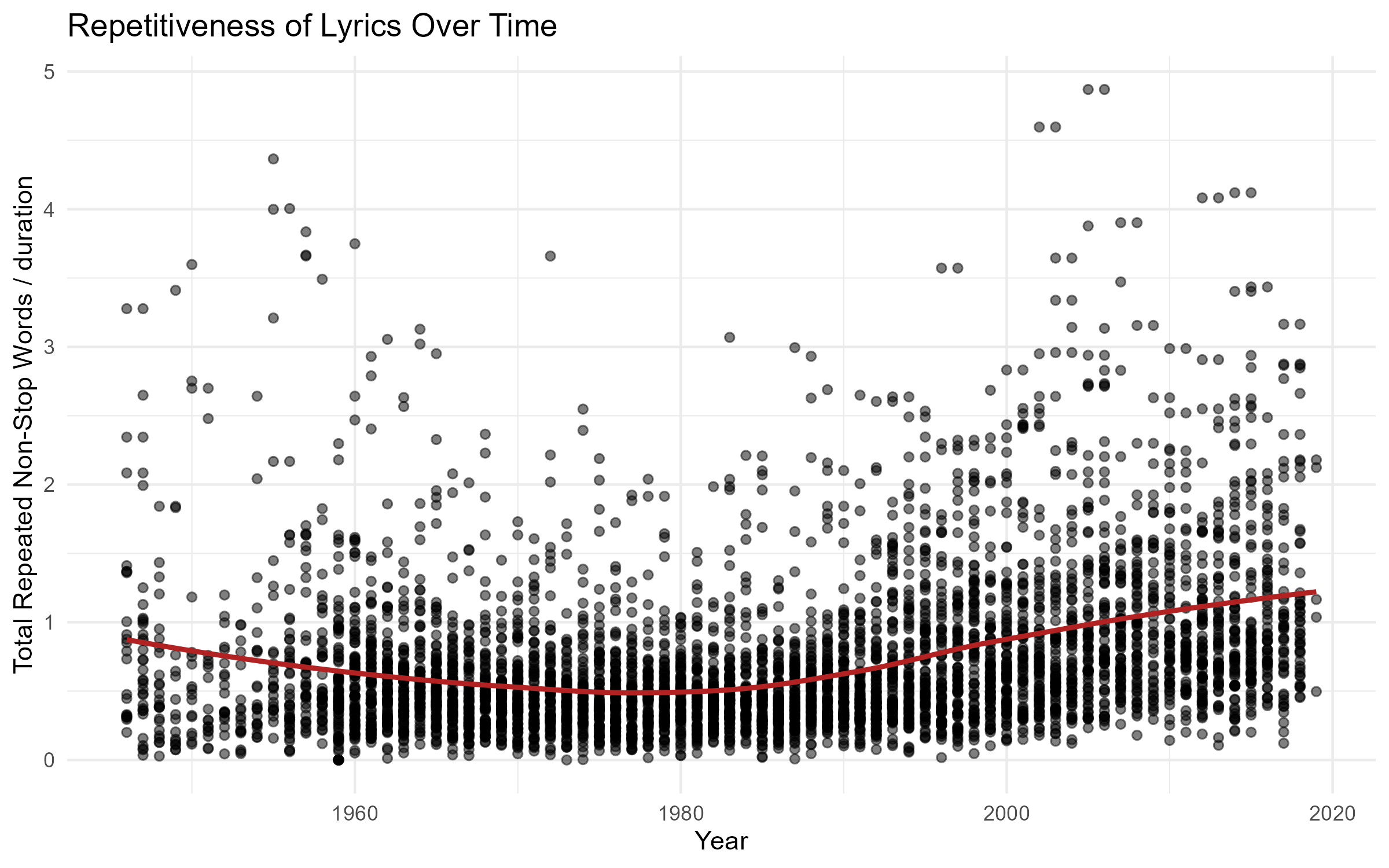


This graph displays the frequency of repetitive lines in repetitive lines per minute. This graph takes into account the duration of the song when analyzing repetitiveness. This graph suggests that repetitiveness took a sharp increase between 1945 and 1960, a slight decrease between 1960 and 1980, and a slow increase between 1980 and 2020. Overall, there is a positive correlation between repetitive line frequency and time. This would suggest that songs are becoming slowly more repetitive. This graph provides a slightly better estimate of repetitiveness than the last one because it takes into account the duration of each song. Therefore, longer songs are not considered more repetitive than shorter songs.

### Most Repeated Non-Stop Words



The above graphics shows our analysis of repetitiveness of lyrics over time in relation to non-stop words. The first plot is just our standard method where we are counting the number of repeated non-stop words. In this plot we can see a consistent upward trend from 1960 to around 2010 when it peaks. This indicates that more modern songs tend to repeat meaningful words more than older songs with a slight tapering in the most recent decade.



The second plot takes into consideration the duration of the song where we divide the total number of repeats by the duration. This helps account for longer songs naturally having more words and naturally more repeated words. We can see a U-shaped pattern that indicates a decrease in repetitiveness per unit time from 1960-1980. After this we see a drastic increase in our repetitiveness score. This indicates that modern artists are cramming more repeated words in their songs with respect to duration.

## Conclusion

Music has been paramount in human culture and society for thousands of years. It can evoke the strongest emotions and elicit the most meaningful memories. Music brings people together in a way that few things do. However, as with everything in life, change is inevitable. Music has changed countless times, sometimes more dramatically than at other times. This project aims to study the change in the repetitiveness of popular music over the past century.

Our analyses indicate there is a definitive increase in repetitiveness over time. Each method of analysis considered both total repetitions and repetitions relative to the song length. All four methods of determining repetitiveness show an increase in repetitiveness over time. Repetitiveness can be classified in many ways, but our analyses show an increase in repeated words, lines, or phrases and an overall decrease in lyrical variety. In essence, modern music repeats the same phrases more commonly and contains fewer unique phrases.

This analysis provides insight into the direction music is heading. With the evolution of digital media, music is as large as ever. While the causes of the increase in repetition were not studied in this analysis, there are plenty of theories to help explain. Are humans simply more attracted to songs with simpler lyrics? Are repetitive songs more popular because the lyrics are easier to remember? Perhaps the digital world has prompted artists to spend less time and effort creating longer, more unique songs because of the ease of widespread media attention. The reasons remain unexplored, yet one thing remains certain: songs are undoubtedly becoming more redundant.