**LOVE IS ALL YOU COUNT: REPORT**

Lorenza Culotta and Amanda Seagle

**Scope**

Data analysis of lyrical trends in love songs through time, from 1960s to 2010s

**Questions**

How do people express their love through music today vs 50 years ago?

* Has the popularity of love songs increased or decreased by decade?
* Has the number of words in each love song changed over time?
* Do the words used in love songs throughout the decades change?
* Is there an emergence or disappearance of certain words used in love songs over time?

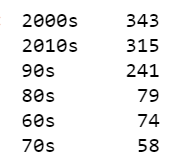
**Process**

**Part1: Retrieve top love songs for different decades**

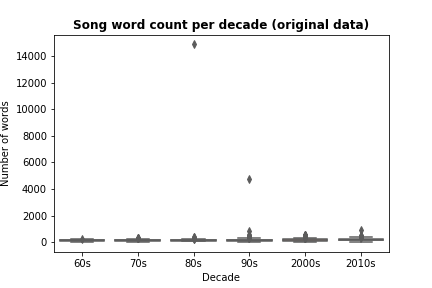
1. Our group decided to pursue this question after developing an informal hypothesis that our generation defines and talks about “love” in a different way than people have in decades past. An informal, non-statistical, highly biased analysis among our group ended with us thinking that words related to physical “love” appear more often in music today, and led us to hypothesize that we would end up with highly variant results that would include words that we see as being more prevalent in the 2010s. We thought we might see a higher occurrence of swear words and sexual words and were interested to see if our data supported this.

1. Our process began with narrowing down the scope of our question and agreeing on the process of defining a “love” song. Our group went through multiple APIs attempting to find one that could define love through data, most of which used some form of AI to determine song valence and mood. None of these were accessible to us, so we decided to rely on the Last.FM API, a database of over 30,000 songs appearing on Last.Fm’s top songs charts, and tagged as love songs. We found minimal documentation of how these songs were tagged, but we had to assume it was through a manual process because there was no mention of AI.

**Question: What's the number of love songs per decade?**

1. We called 2000 love songs from the Last.Fm API, and then used the lyricsgenius module to pull each song’s lyrics and release date through Genius.com. Once we had our dataset, we realized that we had a massive overrepresentation of songs released between 2000-2009 (1034 vs the next highest number, 315 from 2010-2019), and decided to pull songs that Last.FM ranked in the top 350 between 2000-2009 rather than use all 2000 songs, or even out our song list. We originally had wanted to focus primarily on the 1960s, but even after calling 2000 songs had only retrieved 74 songs from that decade. We decided that our dataset was small enough that we could analyze multiple decades.

Our results answered our question and told us that there is no shortage of popular love songs from the past twenty years, and negated our original thinking that millennial artists are more creative with their song topic choices, and that listeners are less interested in love songs than they used to be.

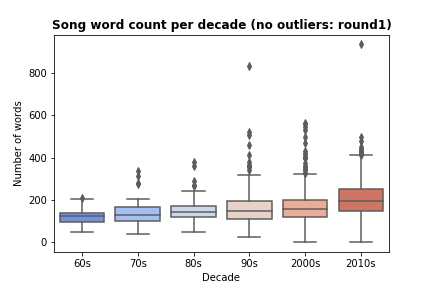
**Part2: Lyrics cleanup**

1. To clean the lyrics, our group removed punctuation by importing the string library, articles and “stop words” (meaning commonly used words that support sentence structure) by using the Natural Language Toolkit Corpus’s pre-defined set of stop words, and words related to song structure that we defined on our own after viewing the results of our lyrics call. We also had to remove some artist names that Genius had accidentally include in the lyrics body, in particular, “Chris Martin” appeared so many times in Genius’s top charts that he appeared in our list of “most used words” for 2000-2010 before we removed him. After cleaning up the lyrics, we plotted the total number of words in each song by decade.

Original Plot: Number of words for each love song

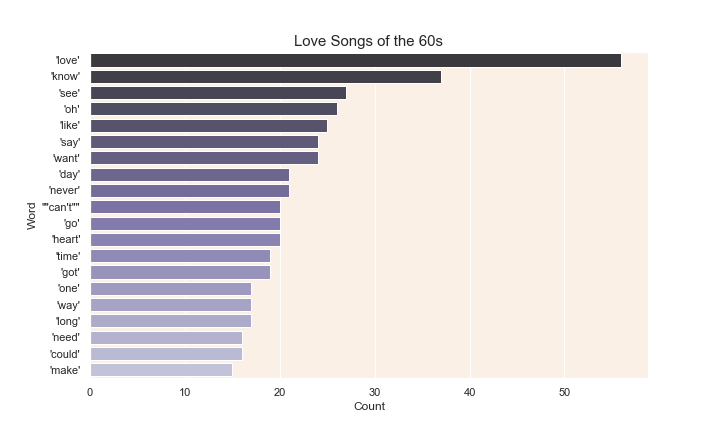
1. After plotting the songs, we were able to identify major outliers, and then ran a OneWay ANOVA for the resulting dataset. Many of our outliers were bad data, one of which was an error from Genius that attributed the lyrics to “The Devil Glitch” by Chris Butler to the Beck song “Everybody’s Gotta Learn Sometimes.” A quick listen to the Beck song told us that the lyrics were misattributed, and a quick Google of the Chris Butler song told us that this is the longest song ever written – definite outlier! Other outliers were songs that had multiple lyrical parts that overlapped each other, like in “Brown Skin Lady” by Black Star, which was removed.

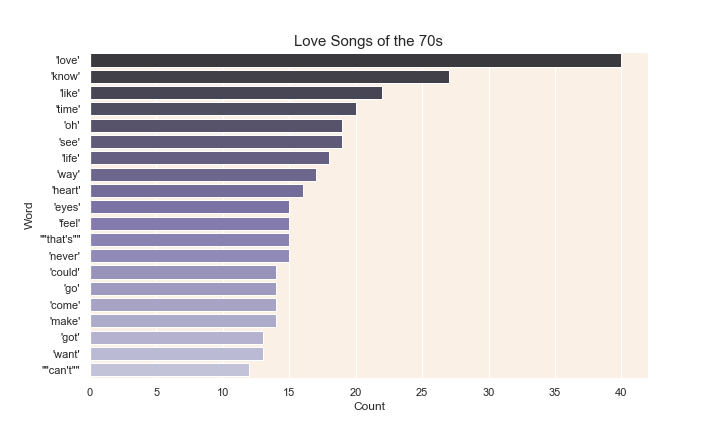
**Part 3: Analysis**

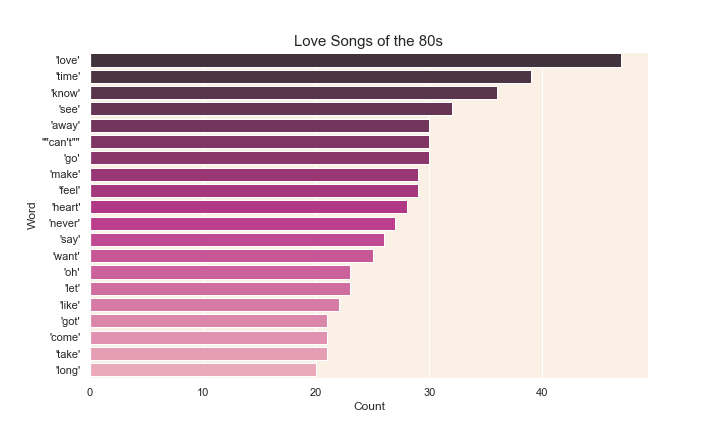
**Question: Has the number of words in love songs increased over time?**

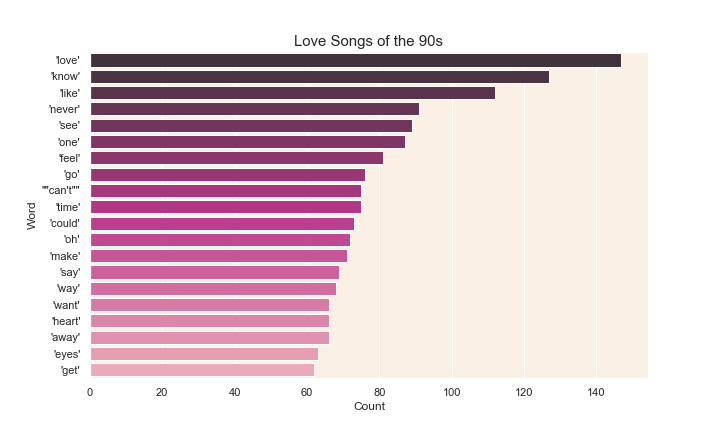
1. We hypothesized that the mean number of words per song would increase decade by decade, and our null hypothesis was that there would be no statistically significant variance between the means of song word counts. Our results supported our hypothesis, as the resulting p-value was significantly less than .05 (8.3958361954e-22), meaning that there are 22 zeroes before our 8 and the mean word count has increased over time. Love songs from the current decade have significantly more words than they did 50 years ago.

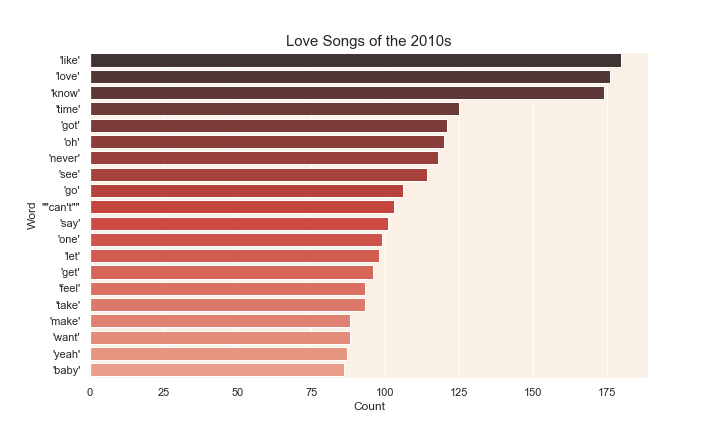
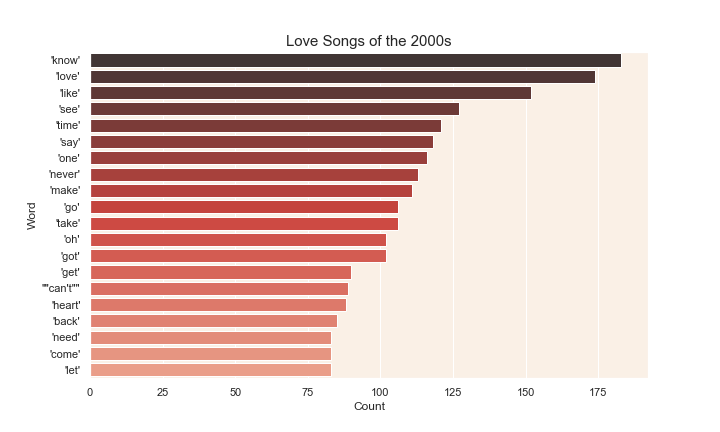
**Question: Have the words we used to express love through lyrics changed since 1960?**

1. ****The genesis of our analysis was based on the hypothesis that there would be a noticeable variance between the most popular words used in love songs by decade. To get our results, we analyzed lists of unique words used in lyrics for each of our songs, again removing stop words, song structure words, and artist names. We decided to use unique words to avoid songs that repeat phrases like “oh baby” ad nauseum, thus increasing the word appearance by decade, from skewing our results. We then created a loop to count the appearance of each word in every song, and plotted our results:



****

****

****

As you can see, the words “love”, “know” and “like” comprise the top three most used words for every decade but the sixties and eighties, where “like” ranks number five and 16 respectively. We plotted 20 words for 6 separate decades, giving us a total of 120 words. However, only 33 out of 120 (or less than 30%) were unique words. Of these unique words, 24% appeared on every list, and 66% appeared on at least 3 lists.

Our resulting list of unique words and their appearance over in the top 20 of each decade is as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| like | 6 | want | 5 | need | 2 |
| love | 6 | go | 5 | away | 2 |
| know | 6 | heart | 5 | eyes | 2 |
| time | 6 | one | 4 | long | 2 |
| got | 6 | feel | 4 | yeah | 1 |
| oh | 6 | could | 3 | baby | 1 |
| never | 6 | way | 3 | cant | 1 |
| see | 6 | let | 3 | back | 1 |
| make | 6 | get | 3 | life | 1 |
| can't | 5 | take | 3 | that's | 1 |
| say | 5 | come | 3 | day | 1 |

From these results we can conclude that we’ve talked about love the same way for many decades, and our original hypothesis does not apply.

* Our final question, “Is there an emergence or disappearance of certain words used in love songs over time?” was assigned to our teammate who left the project early on, and we were not able to answer it in time.