Dear Reader,

This essay was the most interesting to write by far, as it was a topic I chose myself and have a lot of interest in. However, it was pretty difficult and time consuming due to the code and data analysis aspect of the paper. All in all, I enjoyed it a lot even though I spent so many hours on it. My paper is very different from my previous paper, as it is not so much centered around a tension and motivating question. The big focus of my paper is data analysis, and I have 3 main sections in which I complete data analyses. I feel that the weakest point in my paper is my literature review, just because I spent the least amount of time on it. However, I feel that it does touch on all the relevant literature in the field and is a good summary of the gaps in knowledge.

I will send the Python code in a separate email, and both .pdf and .docx files will be posted to the Blackboard. This will be done to ensure that all figures are properly formatted, as I have been experiencing glitches with them on my end.

Thank you for reading,

Adam

The Data of Lyrics:

An Analysis of the Dynamicity of Music Lyrics Since 1965 and their effects on Social Trends

1. Introduction

Song lyrics and music have been coupled since their creation, generating a strong emotional response when the two are effectively paired. However, music and lyrics both serve their specific purposes. For example, music has been shown to increase “social cohesion” of groups and also “increases health and life expectancy” (MacDonald, Kreutz, and Mitchell 441). Lyrics, on the other hand, are less studied. As quoted by Victor Hugo, “Music expresses that which cannot be put into words and that which cannot remain silent.” As Hugo articulates, one major purpose of song lyrics in music is about making a statement about the world. Lyrics are used to send a message in a powerful manner which could not be done through simple prose.

Due to the close relationship between music, lyrics, and emotion (Dediego 2), it is likely that music lyrics will be closely correlated with real-life measures of human emotion and action. It is a fact that different styles of music effectively incite different emotions. These emotions vary based on genre, tone, and lyrics. Lyrics however, can oftentimes be understood in multiple ways. It has been shown that “emotion is strongly related to most people’s primary motives for listening to music” (Juslin and Laukka 1). Some explicit trends in music lyrics over time have been examined. Christenson, Haan-Rietdijk, Robert, and Bogt found that sex and romance is among one the most common themes in popular music, but the proportion of this theme in lyrics remains relatively constant (3). Additionally, Napier and Shamir have noted extreme differences between the music of the 50s and the popular music today, stating that “anger, disgust, fear, sadness, and conscientiousness have increased significantly, while joy, confidence, and openness expressed in pop song lyrics have declined" (2). It is generally known that the content of popular music lyrics is changing over time in various semantic subgroups. However, it remains an open question whether certain trends in lyrical frequencies over the years can be related to social and empirical data from the real world, as everyone listens and understands music differently.

Early research has noted that popular music culture strongly affects teens. For example, Cole states that The Beatles had a very significant impact on young men's clothing and hairstyles during their popular years (389). Johnstone and Katz found that teenage female music preference was closely related to her residential neighborhood within a town and her popularity levels (23). Both of these examples reflect music and lyrics' direct impacts on social outcomes. These social outcomes, fashion choices and music preference, are just small examples of what could be major correlations between lyrical content and real-world trends.

Not all research is in agreement that lyrics may reflect real-world trends, however. Robinson and Hirsch argue that, although teens may have listened to popular protest songs, most of them were unaware of the actual meaning behind the lyrics and the social change that the songs were supposed to inspire (12). Additionally, Horton found that music listeners oftentimes don't even know the lyrics to songs, other than the chorus, furthering the claim that lyrics would have no effect on real-world outcomes.

In order to solve this tension, several statistical analyses will be completed on the lyrics of top 100 Billboard ranked songs from 1965 to 2015. Real-world trends that will be examined will include economics (GDP, unemployment) and violence (crime rates, murders). Additionally, sentiment analysis will be completed among lyrics in order to examine trends in lyrics over time.

2. Methodology

2.1 Data collection

Lyric data was collected from Kaggle.com from the “Billboard 1964-2015 Songs + Lyrics” data set. This data set includes the top 100 Billboard ranked songs from each year from 1965 to 2015. Data columns include rank (of 100), song title, artist name, year, and lyrics. The lyrics column contains all of the lyrics of the given songs, as they would be sang. In total, there were 5100 songs from 51 years. 247/5100, or about 4.8% of songs, were marked as “NA” or “instrumental”, so they were not included in lyrical data analysis.

2.2 Data Analysis

Section A: Word Clouds

Data was analyzed using a novel Python script to clean, sort, graph, and regress data points. Several analyses were completed, which will be outlined here. To begin, in order to get an idea of the general mood of each decade (1960s - 2010s), the most frequent 30 single-word lyrics of each decade were arranged into word clouds. Stop words, or meaningless words normally filtered out prior to natural language processing, cluttered the word clouds and had to be removed. The code was altered to ignore stop words, and the list of ignored words can be found in the appendix. Additionally, to avoid significant yet prevalent words from cluttering the word clouds, a ‘z-score’ was calculated for each word by determining the ratio of each word’s actual occurrences to expected occurrences within a decade. From there, six word clouds were created with the highest z-scoring words of each decade. These words vary greatly among the decades and help to define the significant themes of the popular lyrics from each decade.

Section B: Frequency Distribution and Zipf's Law

In continuation, the overall frequency distribution of all words (excluding stop words) will be graphed and compared to a best-fitting curve. The curve, in this case, is hypothesized to be in the form described by Zipf's law, where k is an arbitrary constant, n is the integer frequency ranking of the given word, and f is the number of occurrences:

$$f = \frac{k}{n}$$

Zipf's law is a linguistic formula which states the frequency of word occurrences is inversely proportional to a given word's frequency ranking among a large text (Piantadosi 1). For example, the 2^{nd} most common word in a data set has a 50% probability of appearing as the most common word. The 3^{rd} most common word has a 33.3% probability of appearing, and so on. This law is being used in this context because previous scholars have successfully modeled past literature using the relationship (Piantadosi 1). However, there do not seem to be studies centered around the law's effectiveness concerning music lyrics, which will be examined here. If Zipf's law is shown to hold true for this data set, this provides evidence that song lyrics are normally distributed in frequency, just as other large text databases have proven to be.

Section C: Real-World Correlations

The final analysis completed among the data concerns trends in specific word subgroups and their correlations to related real-world measures. There will be a total of four lyrical subgroups and corresponding real-world measures. The subgroups are as follows: Positive, Negative, Financial, and Violent. Each subgroup is comprised of 6-17 of the most common overall words which could be classified into each group. Highest frequency words were chosen because they have the largest effect on the subgroup total, as word frequencies decrease with

Zipf's function. The words in each subgroup were hand-coded and may be a minor source of human error and bias. Exact lists of the words included in each subgroup are shown below.

- <u>Positive</u>: love, heart, good, dance, sweet, light, together, best, dream, friends, party, soul, dreams, sing, smile, kind
- <u>Negative</u>: gone, leave, bad, hard, alone, cry, break, rain, lonely, miss, lose, lost, end, hurt, cold, die, tears
- <u>Financial</u>: money, dollar, cash, bill, bills, ballin
- <u>Violent</u>: shit, high, fuck, ass, club, bitch, fight, drink, bang, gun, fuckin, drunk, kill, shoot, hell

These subgroups were assembled using the most common words which could be classified into each category and had a total frequency over 100 occurrences.

Using the four subgroups, the data was then correlated to real-world data sets in order to search for correlations between common lyric sentiments and empirical data. To begin, the Financial subgroup was compared to US closing GDP per year and average unemployment rate. This data was gathered from the Global Financial Data database. Specific data sets used were "United States Real GDP in 2012 Dollars" and "United States Unemployment Rate - 25 years and over."

The Violent subgroup was then compared with average violent and property crime rates in the US. Violent crimes include murder rape, robbery, and aggravated assault (FBI - UCR). Property crimes include burglary, larceny, and motor vehicle theft (FBI - UCR). Data was gathered from the FBI's Uniform Crime Reporting (UCR) Statistics.

Finally, Positive and Negative subgroups were compared to each other. In their comparison with each other, averages and sample standard deviations were calculated for the percent frequencies of words in both the Positive and Negative subgroups. From there, the number of standard deviations in which each year differed from its respective subgroup average was calculated (ΔSTD). The ΔSTD values of the Positive and Negative subgroups were then graphed against each other, and linear best-fitting models were created for each subgroup in order to compare changes over time.

It is hypothesized that, assuming lyrics accurately reflect author sentiment, the Financial and Violent subgroups will be significantly correlated with their respective real-world data sets. The specificities of the correlations (positive or negative) will provide evidence into how certain words are being utilized by authors in their songs. Additionally, it is hypothesized that the Positive subgroup will not be significantly correlated with the Negative subgroup. All correlation tests were completed on Microsoft Excel using the "Data Analysis" function and selecting "Regression." An alpha of 0.05 was used to determine significance.

3. Results

Section A: Word Clouds

Six word clouds were created using the 30 highest z-scoring lyrics of each decade (1960s-2010s). They are shown here below in Figure 1. Data is located in the appendix.

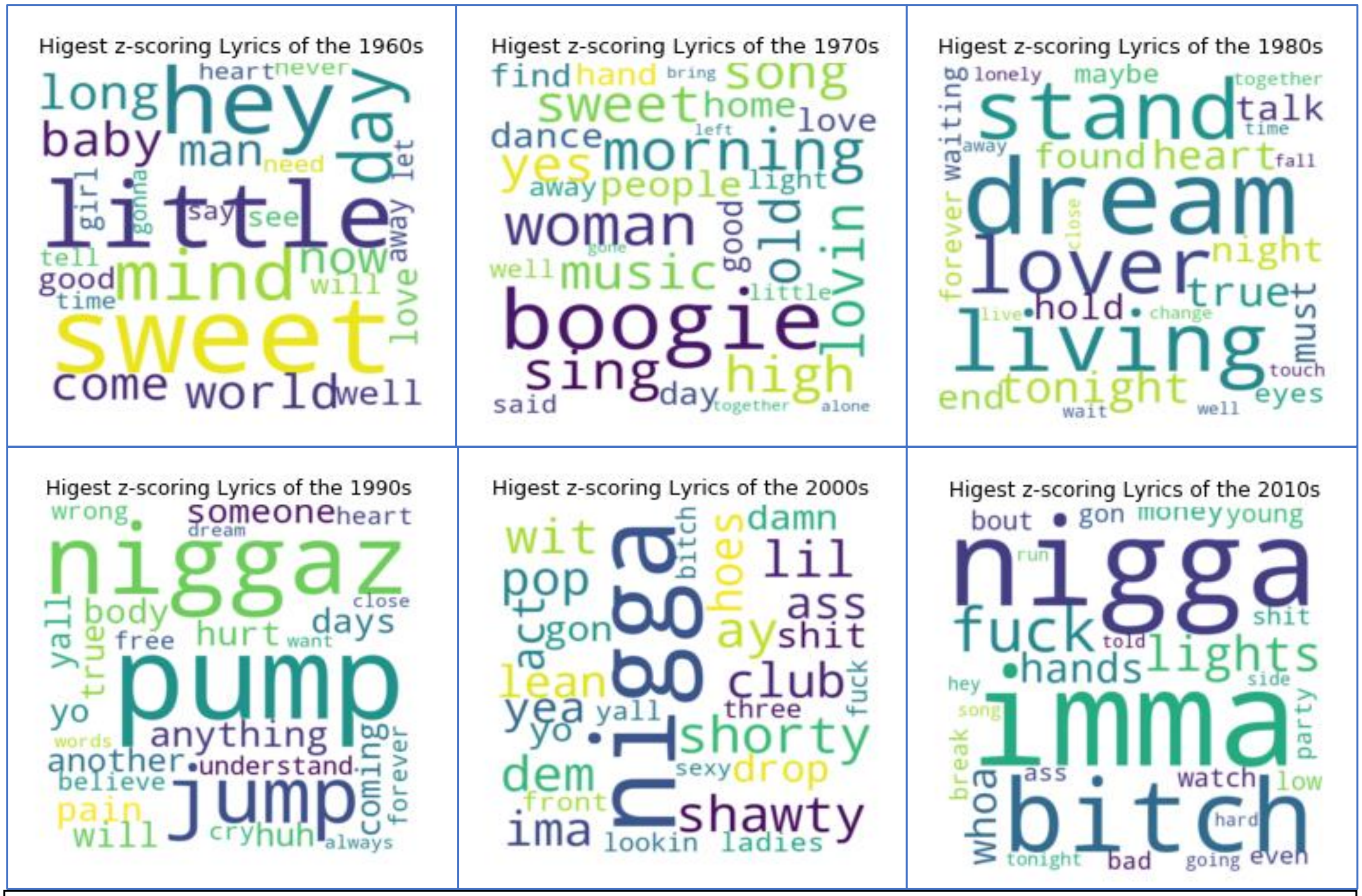

Figure 1: Highest 30 z-scoring lyrics from each decade since the 1960s. The z-scores were calculated by finding the ratio of actual occurrences to expected occurrences per decade.

Section B: Frequency Distribution and Zipf's Law

The overall frequency distribution of all words besides stop words was then graphed and fit to a best-fitting curve. The graph is shown below in Figure 2.

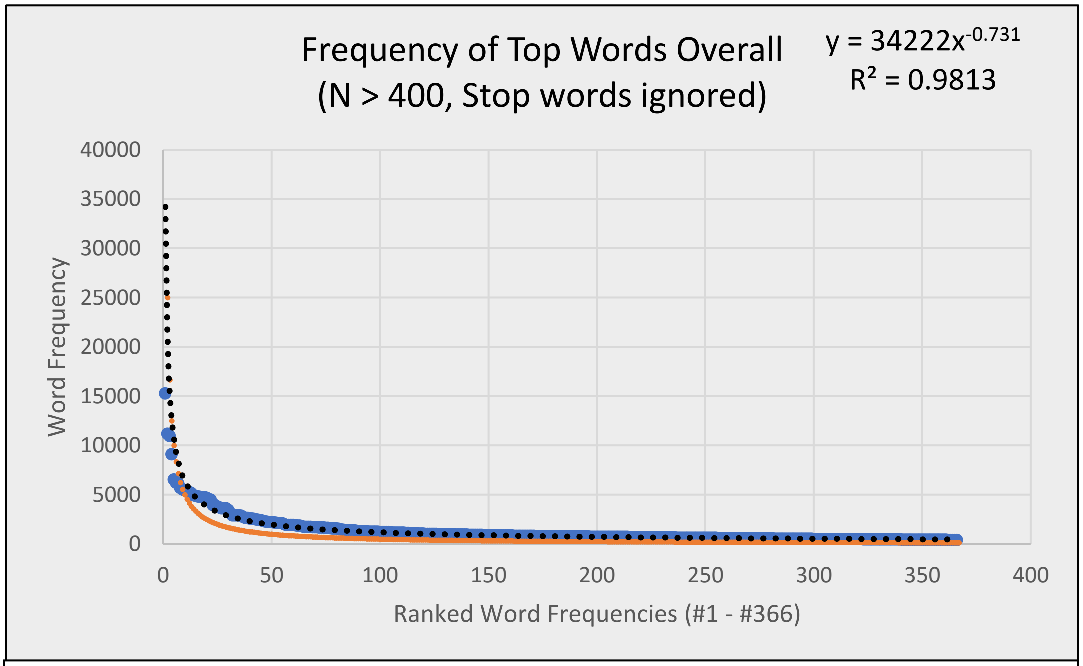

Figure 2: Frequency distribution of all words appearing at a frequency N > 400 occurrences. Stop words are ignored in this analysis. The data is graphed in blue, the best-fitting curve is in black, and the Zipf function is plotted in orange.

The data (graphed in blue) is closely fitted by the best-fitting curve (graphed in dotted black). The best-fitting curve, in this case was a power curve with the equation:

$$y = \frac{34{,}222}{x^{0.731}}$$

The R^2 value of this regression was 0.9813, meaning the model fits the data extremely closely. This model also closely resembles the shape of the Zipf's law function (graphed in orange). The Zipf function graphed is in the form:

$$y = \frac{k}{n} = \frac{50{,}000}{x}$$

The value for k was arbitrarily chosen as 50,000 to fit the given data.

Section C: Real-World Correlations

Relevant words were grouped into four categories according to their connotations (Positive, Negative, Financial, or Violent). Percent frequencies were calculated for every year for each of the four categories. The Positive subgroup had an average percent frequency of 2.14%, sample standard deviation of 0.61%, maximum of 3.61%, and minimum of 1.05%. The Negative subgroup had an average percent frequency of 0.91%, sample standard deviation of 0.16%, maximum of 1.35%, and minimum of 0.59%. The Financial subgroup had an average percent frequency of 0.09%, sample standard deviation of 0.07%, maximum of 0.35%, and minimum of 0.01%. The Violent subgroup had an average percent frequency of 0.33%, sample standard

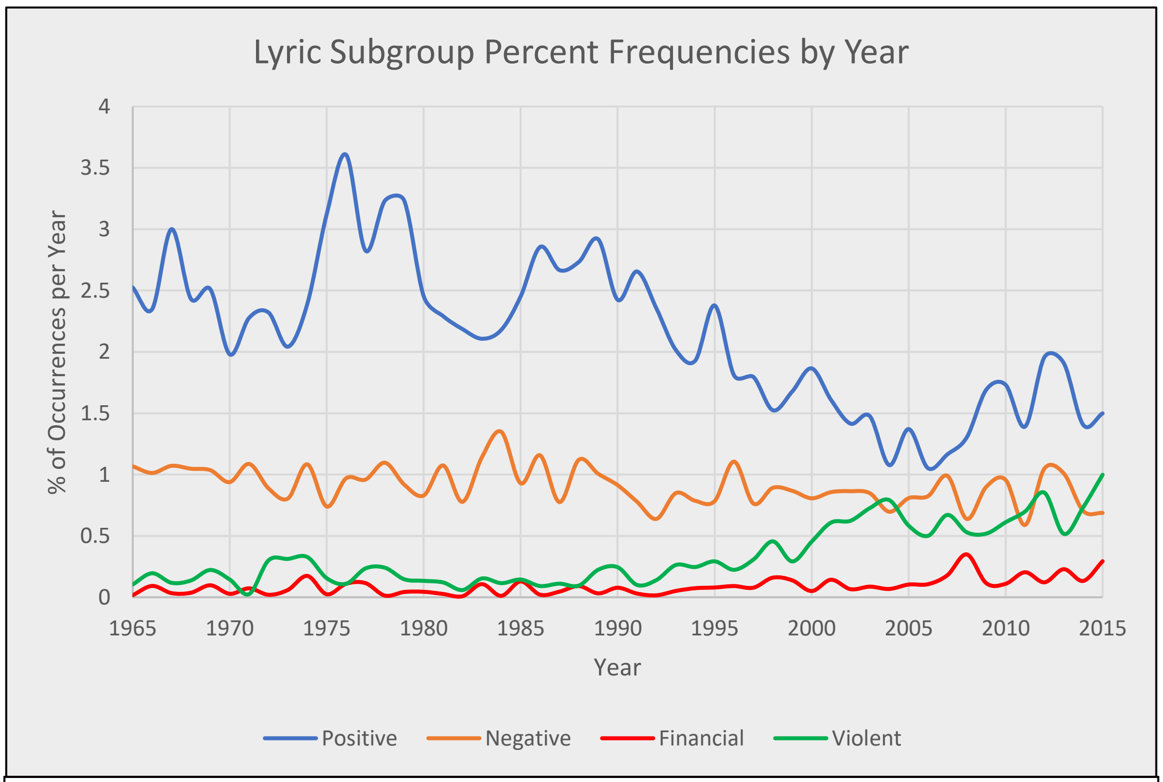

Figure 3: Lyric subgroup % frequencies by year. Blue denotes Positive subgroup, orange denotes Negative subgroup, green denotes Financial subgroup, and red denotes Violent subgroup.

deviation of 0.24%, maximum of 1.00%, and minimum of 0.03%. The percent frequency values of each subgroup for every year is graphed below in Figure 3.

The Financial subgroup was correlated to two real-world financial metric: US closing GDP and unemployment rate over 25. The Financial subgroup percent frequency was significantly correlated to US real GDP in millions of dollars, with a P-value of $P<0.0001$, below the alpha of 0.05. The regression between the Financial subgroup and the average US unemployment rate over 25 years was not significant, with a P-value of 0.6102. The two relationships are graphed below in Figures 4 and 5, including equations for the best-fit linear model and their corresponding R^2 values.

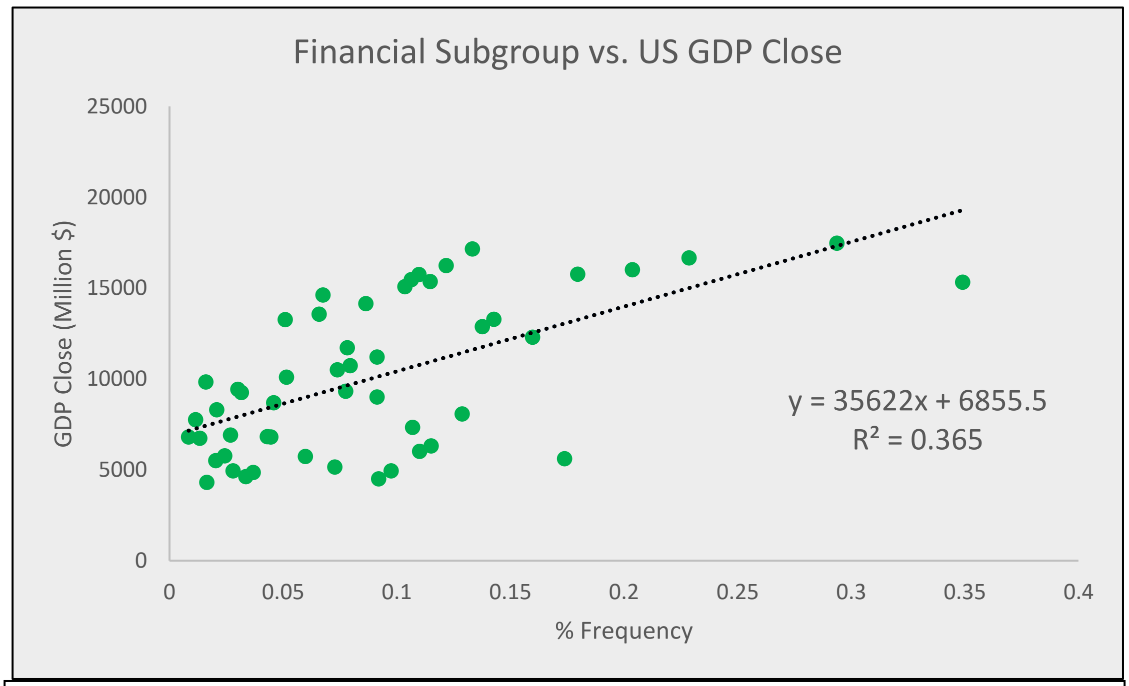

Figure 4: Financial subgroup % frequency vs. US GDP close in millions of dollars. The equation of the line of best fit is $y = 35622x + 6855.5$ and $R^2 = 0.365$, meaning there is a slight positive correlation between the two variables.

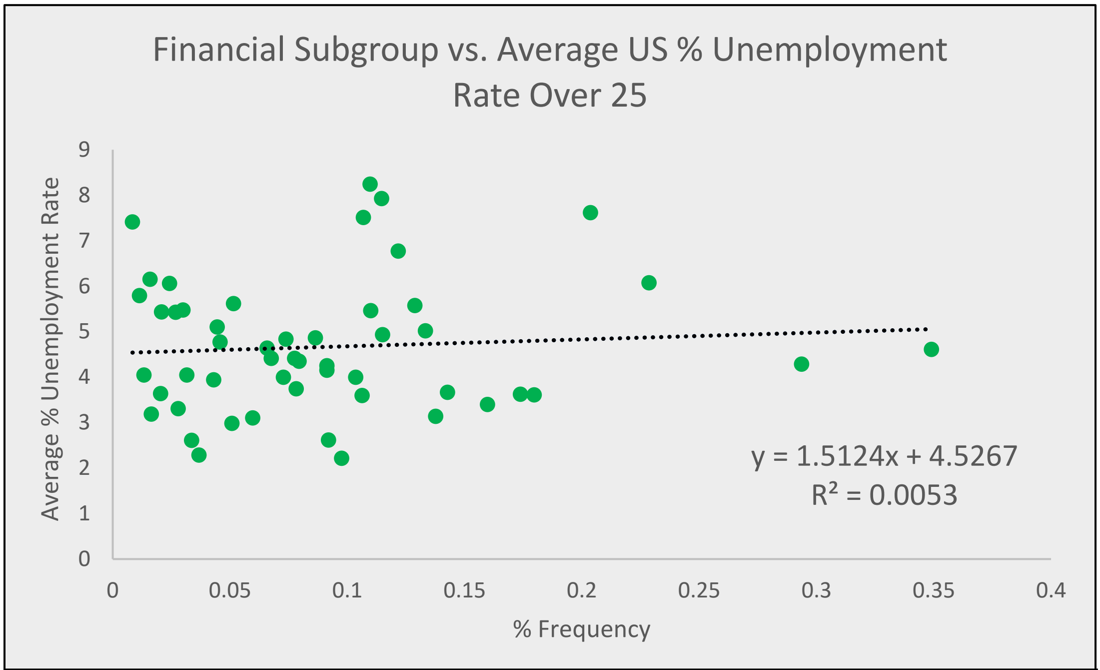

Figure 5: Financial subgroup % frequency vs. US average unemployment rate over 25 years. The equation of the line of best fit is $y = 1.5124x + 4.5267$ and $R^2 = 0.0053$, meaning there is no correlation between the two variables.

The Violent subgroup was correlated to two real-world crime/violence metrics: property crime rate and violent crime rate. The Violent subgroup percent frequency was significantly correlated to the national property crime rate, with a P-value of $P<0.0001$. The Violent subgroup was not significantly correlated with the national violent crime rate, with a P-value of 0.0987. However, one facet of violent crime rate, “murder and nonnegligent manslaughter rate,” is significantly correlated with the Violent subgroup frequency, with a P-value of $P<0.0001$. The three relationships are graphed below in Figures 6-8, including equations for the best-fit linear model and their corresponding R^2 values.

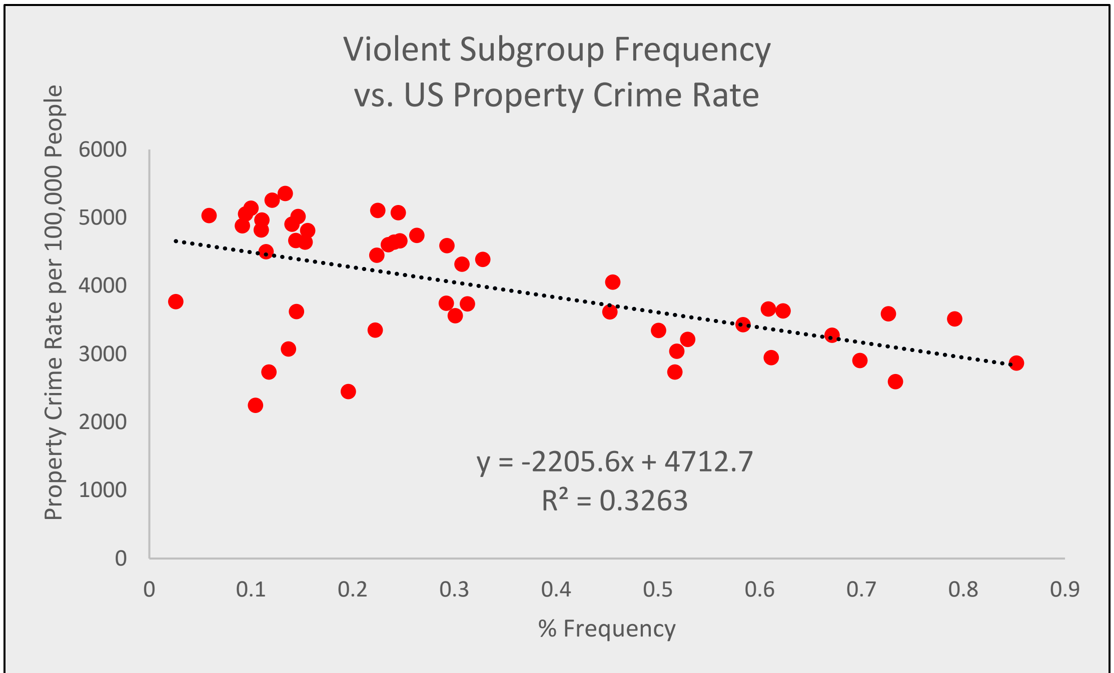

Figure 6: Violent subgroup % frequency vs. US property crime rate. The equation of the line of best fit is y = -2205.6x + 4712.7 and R^2 = 0.3263, meaning there is a slight negative correlation between the two variables.

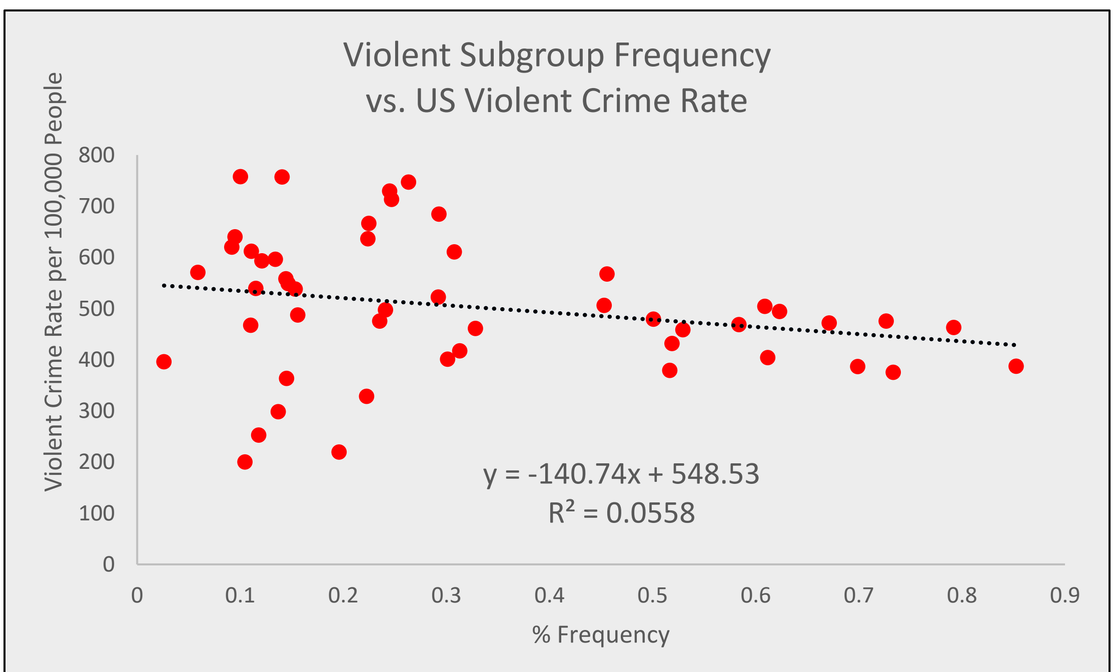

Figure 7: Violent subgroup % frequency vs. US violent crime rate. The equation of the line of best fit is y = -140.74x + 548.53 and R^2 = 0.0558, meaning there is no correlation between the two variables.

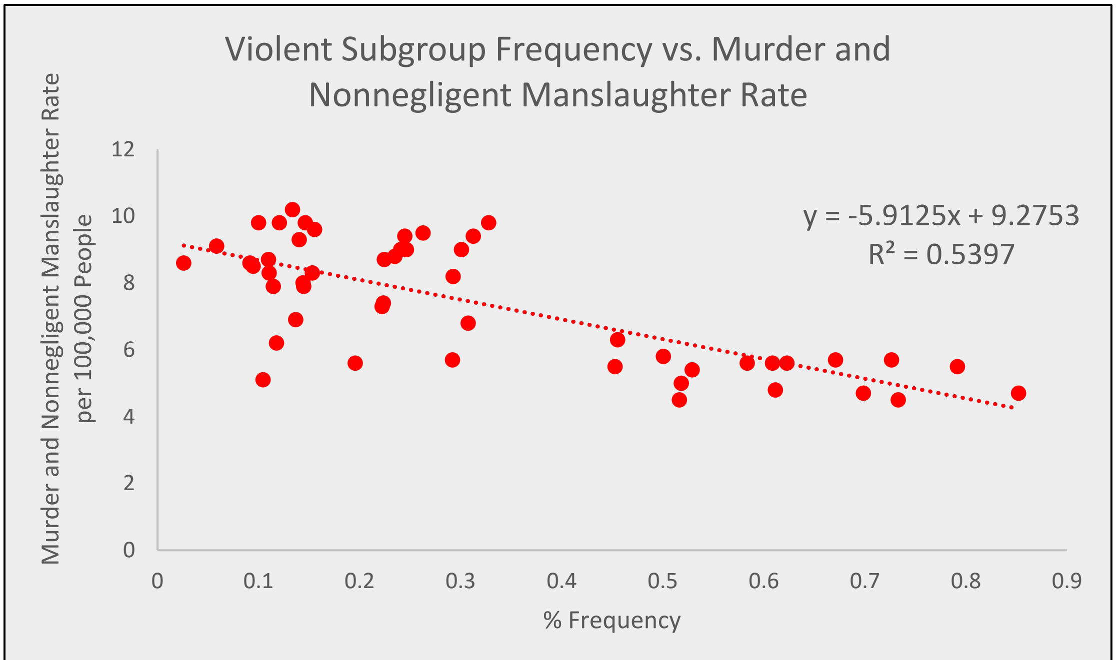

Figure 8: Violent subgroup % frequency vs. US murder and nonnegligent manslaughter rate. The equation of the line of best fit is y = -5.9125x + 9.2753 and R^2 = 0.5397, meaning there is a strong negative correlation between the variables.

Lastly, the Positive and Negative subgroups were directly compared with each other, by analyzing how the number of standard deviations from the mean (ΔSTD) changed over time. The average ΔSTD for both subgroups was 0, as the values were centered around 0. The maximum Positive ΔSTD was 2.41, and the minimum Positive ΔSTD was -1.79. The maximum Negative ΔSTD was 2.77, and the minimum Negative ΔSTD was -2.04. The two subgroups were significantly statistically correlated, with a P-value of 0.0039 and an R^2 value of 0.1578. A graph of the two subgroups for ΔSTD over time is shown below in Figure 9.

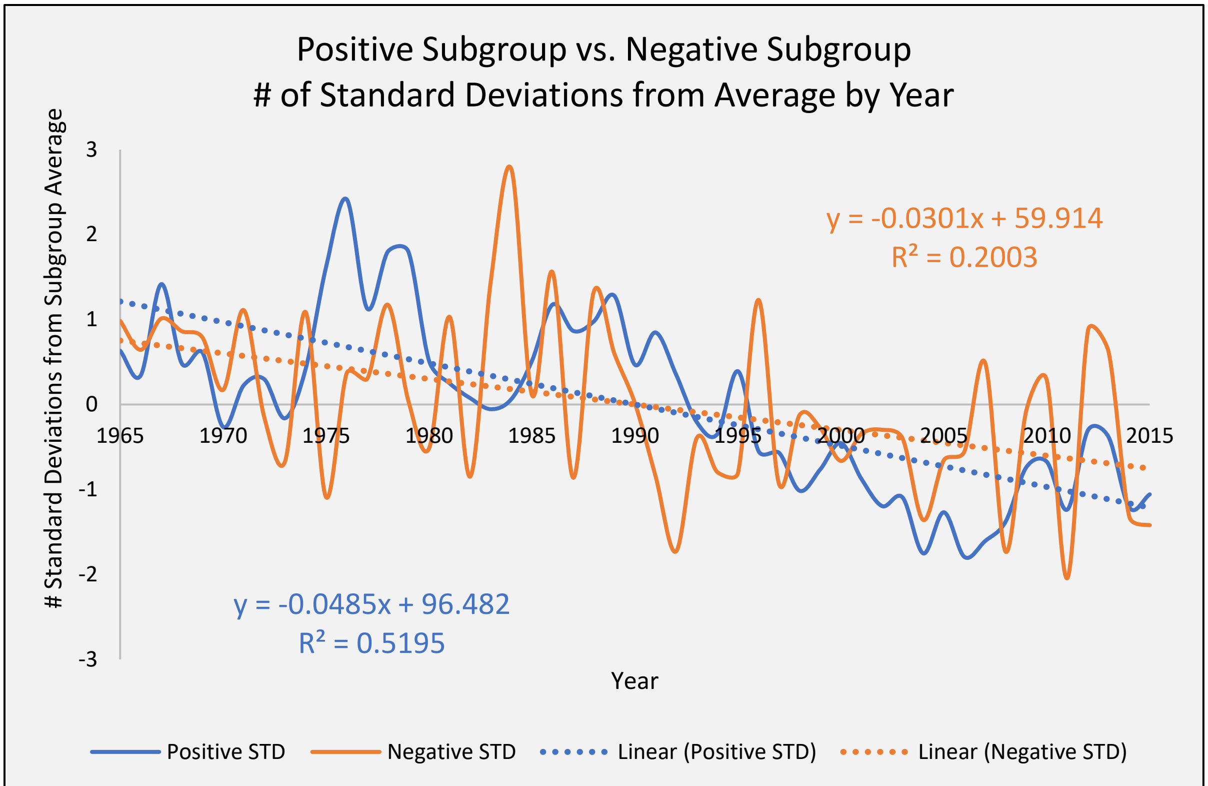

Figure 9: Positive subgroup (Blue) and Negative subgroup (Orange) ΔSTD per year. The line of best fit for the Positive subgroup (Dotted blue) has an equation of $y = -0.0485x + 96.482$ and $R^2 = 0.5195$. The line of best fit for the Negative subgroup (Dotted orange) has an equation of $y = -0.0301x + 59.914$ and $R^2 = 0.2003$. Therefore, both subgroups have a slight downwards trend over time.

4. Discussion

Section A: Word Clouds

From the word clouds portrayed in paragraph 3, section A, it can be seen that each decade is defined by unique, characteristic lyrics. Word clouds, and therefore this section of the analysis, are very qualitative, so statistical tests were not completed in this section. Over the decades, 1960s-2010s, it can be seen clearly how Positive words (sweet, good, baby) decreased over time and Violent words (bitch, fuck) increased over time. Negative and Financial words, however, were too evenly distributed over the decades to be represented significantly in word clouds. This is because z-scores were used to create word clouds, meaning that words that differed the most from their expected values had the highest z-score values.

Section B: Frequency Distribution and Zipf's Law

The frequency distribution of words (with frequencies > 400) did, in fact, follow a Zipf function. Therefore, this finding supports the hypothesis stated in chapter 2.2, section B. Since lyrical data follows this Zipf distribution, and so do large examples of literature, literature-processing techniques can be effectively used in the same manner to analyze this data set, and we can be sure that this is not a source of error in data analysis.

Section C: Real-World Correlations

The percent frequency of the Financial subgroup was found to be significantly correlated to annual closing GDP but not with the unemployment rate. The trendline for the regression between % frequency and closing GDP had a positive slope and an R^2 of 0.365, displaying a slight positive correlation between the variables. Roughly, a higher frequency of Financial word occurrences signified a higher closing GDP in any one year. However, as GDP is a widescale, national measure, it is unlikely that it is altered by song lyrics. Therefore, it can be reasonably assumed that as GDP increases, songwriters mention money-related terms in their lyrics more often. The opposite follows, in that, as GDP decreases, songwriters mention Financial words less often. This result leads to one possible conclusion about the purpose of Financial terms in songs. Since the terms are mentioned in times of greater economic prosperity, Financial terms are likely used to brag about a state of wealth, rather than sing about a state of poverty or financial insecurity.

The percent frequency of the Violent subgroup was found to be significantly correlated to property crime rate and murder/nonnegligent manslaughter rate, but not violent crime rate. The linear trendline for the regression between the Violent subgroup and property crime rate had a negative slope and an R^2 of 0.3263, signifying a slight negative correlation between the two variables. The regressions including violent crime and murder/nonnegligent manslaughter rate both had negative slopes with R^2 values of 0.0558 and 0.5397, respectively. Therefore, analyzing the two significant cases, higher frequencies of Violent words in any given year led to lower rates of property crime and murder. As these crimes are done on an individual level (unlike GDP, which is national), it can be argued that the relationship between the variables acts in either direction. For example, greater mentions of violence in music may effectively satisfy criminals' urge to commit crimes. Therefore, less crimes are committed as Violent words are mentioned in songs more often. Conversely, greater crime rates may lead to decreased frequencies of violence in songs. One possible explanation for this pattern may be that songwriters may become numb or accustomed to hearing of common crimes, but when crimes are rarer within a year, they individually affect the author to a greater extent.

Positive and Negative percent frequencies were shown to be significantly positively correlated with each other, with a P-value of 0.0039 and an R^2 value of 0.1578. This means that as one subgroup increased or decreased, the other followed. Finally, as can be determined from Figure 9, both Positive and Negative terms are decreasing within popular music. From the linear model, the frequency of Positive terms decrease at roughly 0.05 standard deviations per year, whereas the frequency of Negative terms decrease at roughly 0.03 standard deviations per year. The R^2 values of the Positive and Negative regression models are 0.5195 and 0.2003, respectively, describing moderate negative correlations in the data. Therefore, from the past to the present, song lyrics are becoming less strongly Positive or Negative, in favor of other emotions. For example, Violent and Financial themes increased over time, as shown in Figure 3. Additionally, other emotions, such as love/romance, which were not studied, may have increased as well. Neutral words may also be more common due to evolutions in how songwriters write and use the English language in their songs over time.

5. Conclusion

From the data analyses and relationship that have been completed and examined throughout this paper, one can confidently conclude that lyrics have a significant social impact in the United States. In both the contexts of economics and crime statistics, relevant lyrics have effectively been correlated to major empirical data sets. It has been shown that lyrical semantics fluctuate over time and correlate to real-world trends. Additionally, lyrical distributions follow the same laws that govern literature and other large bodies of text. These analyses shed light on the shear significance of music and lyrics in popular culture since 1965 until today. Without music, the world would be a different place. Music and lyrics are intertwined in almost every aspect of life, which helps to explain their significant relationship with social trends. Most things in life, whether it be sports, religion, school, work, or something else can integrate music in some way or another. Music is, by definition, the combination of vocal or instrumental sounds, in such a form as to produce beauty and emotion. With lyrics, music transcends pleasing the ear and fulfills the greater purpose of stimulating emotion and triggering social change.

6. Appendix

Stop words: you, i, the, a, and, to, me, it, my, in, on, oh, im, we, yeah, la, NA, is, that, your, be, of, all, dont, so, for, just, do, with, its, but, no, got, get, can, what, when, this, youre, if, up, she, some, much, our, his, her, about, theres, then, da, every, shes, ooh, that, could, had, wont, where, at, ill, isnt, cant, whos, na, put, take, have, they, how, are, youve, was, not, were, there, an, em, uh, id, am, him, from, he, as, ive, ya, by, or

The z-scores of the top 30 words from each decade:

1960s word cloud:

```
sweet: 2.481307751555933
little: 1.7527576964966587
hey: 1.7123491332505116
mind: 1.7069597418346178
day: 1.655849655444104
too: 1.5840945951225263
world: 1.541407725455383
long: 1.524749382681868
baby: 1.5161941287343261
come: 1.4804745386774756
now: 1.430982539326433
man: 1.39309431550251
well: 1.3831752714251875
love: 1.355619132525775
good: 1.319006435974412
away: 1.1881424791752713
girl: 1.1847559997175143
see: 1.1211608068153092
will: 1.1123625001990431
say: 1.1110333215922088
let: 1.1013037936071677
heart: 1.0983956884195176
never: 1.0971420935556382
tell: 1.094686120126041
need: 1.0754326608350053
time: 1.0415164646503903
down: 1.0348528318747343
here: 0.9977823005127454
gonna: 0.9912142789512228
been: 0.966168356714837
```

1970s word cloud:

```
boogie: 5.942541000044395
woman: 2.6678143954078006
```

```
morning: 2.5002791158288327
lovin: 2.1326167704839483
sing: 2.125523594602532
sweet: 2.0615868184671036
old: 1.9496583547203494
high: 1.9350336319470136
song: 1.8641496847900636
yes: 1.8289236312820998
music: 1.8288178520090355
people: 1.7973857369055153
home: 1.7511432353633873
dance: 1.7385420337249242
day: 1.4836267202710125
hand: 1.4727725768421733
love: 1.4550382323642483
find: 1.45159465959129
good: 1.451271093897271
away: 1.4301972714069748
light: 1.4170556175810376
said: 1.3960707379699662
well: 1.387838924154645
little: 1.3686321474909213
together: 1.3655804777302638
alone: 1.3593385180543782
again: 1.355302445473744
left: 1.3298953347170845
bring: 1.3195055274146072
gone: 1.3132531777184737
```

1980s word cloud:

```
living: 1.9832258298198777
stand: 1.9820667785796482
lover: 1.9641820829961856
dream: 1.7927226447223479
tonight: 1.7669878011997404
true: 1.6801518144006067
heart: 1.62430800376583
dreams: 1.6117256675528842
night: 1.6033943259169467
found: 1.5975627822589609
talk: 1.5837623633254196
hold: 1.5745864516027925
must: 1.5619173652105864
end: 1.5515331769175256
eyes: 1.5279518343027503
waiting: 1.521658562197936
forever: 1.519433342847099
maybe: 1.5107808814478565
together: 1.508558552991414
lonely: 1.502880381668961
wait: 1.4423102366801281
away: 1.4244653349211769
touch: 1.4228642124015953
```

```
close: 1.4148138913214647
time: 1.4081287472997646
should: 1.4017173536664278
live: 1.3986846788839202
fall: 1.3955942728070807
change: 1.388438086438992
well: 1.3859229573612568
```

1990s word cloud:

```
pump: 3.642628005311301
niggaz: 3.0865926190644317
jump: 1.8685319196055594
anything: 1.6663578975564601
someone: 1.499091297020203
days: 1.4926235620418147
body: 1.4896632367262022
pain: 1.4636701299850587
hurt: 1.4579418000566817
ever: 1.457764640600171
will: 1.453305798307597
because: 1.3860779785159714
another: 1.376648366843677
true: 1.3553639218254574
coming: 1.3551120670425287
yall: 1.3548072981828228
huh: 1.3162552801768157
yo: 1.3128270148251333
understand: 1.3110481164310612
believe: 1.295539551636477
wrong: 1.2922756611794537
heart: 1.290374231437468
cry: 1.2822790355062417
forever: 1.276410481567998
free: 1.2700883545348796
always: 1.255369895162589
words: 1.2532077583532202
dream: 1.2489625333162822
close: 1.2350747189177431
want: 1.2275180443495846
```

2000s word cloud:

```
shorty: 3.055513979550214
ay: 2.854073700015856
lil: 2.825208226437946
shawty: 2.823989150612083
dem: 2.767114963690536
wit: 2.670837371712317
u: 2.6432846937102172
ima: 2.5719632961675423
club: 2.5328904762115507
pop: 2.4758416705177435
```

```
lean: 2.1677369292977575
hoes: 2.1626087919185464
ass: 2.131707192794037
act: 2.0886508331929945
yea: 2.0865916966502476
drop: 2.0755832992098866
gon: 2.017630581002091
damn: 1.9998650458841263
shit: 1.955713958534462
yo: 1.9426144739537152
ladies: 1.884818968228293
front: 1.8730084167866383
nigga: 1.8634056227680378
yall: 1.7764589855378197
niggas: 1.674064001521691
lookin: 1.6568689905460565
bitch: 1.6293098214352018
fuck: 1.6246315381367142
three: 1.6237890696547976
sexy: 1.6099537928767746
```

2010s word cloud

```
imma: 5.507934162332987
bitch: 2.9647563291061942
fuck: 2.7825784665844937
niggas: 2.677356480876512
lights: 2.285430746210569
hands: 2.15270960353383
whoa: 2.15221881549353
money: 2.040345638776549
bout: 1.8966037030643481
nigga: 1.854902240202194
watch: 1.8277621244434223
young: 1.759460414294178
gon: 1.7414819192084234
low: 1.7002004563141866
ass: 1.6895473719998642
shit: 1.6501945605402402
even: 1.63530355138032
bad: 1.5951006244967663
party: 1.569349293885531
break: 1.5471629656651587
tonight: 1.5264559640031963
hard: 1.4839783477541564
hey: 1.4711499427601054
song: 1.4641432130283503
side: 1.445727399722363
like: 1.435423599834254
these: 1.426134446770158
going: 1.4170148943769454
told: 1.4132435467190774
run: 1.4014519981058655
```

7. Works Cited

Christenson, Peter G., et al. “What Has America Been Singing about? Trends in Themes in the U.S. Top-40 Songs: 1960–2010.” *Psychology of Music*, vol. 47, no. 2, 2018, pp. 194–212., doi:10.1177/0305735617748205.

Cole, Richard R. “Top Songs in the Sixties.” *American Behavioral Scientist*, vol. 14, no. 3, 1971, pp. 389–400., doi:10.1177/000276427101400311.

DeDiego, Amanda C. “The Use of Song Lyrics as an Expressive Arts Tool in Counseling: A Literature Review.” *American Counseling Association*, doi:10.1891/9780826177025.0001.

“Estimated Crime in United States-Total.” *Uniform Crime Reporting Statistics*, www.ucrdatatool.gov/Search/Crime/State/RunCrimeStatebyState.cfm.

Johnstone, John, and Elihu Katz. “Youth and Popular Music: A Study in the Sociology of Taste.” *American Journal of Sociology*, vol. 62, no. 6, 1957, pp. 563–568., doi:10.1086/222104.

Juslin, Patrik N., and Petri Laukka. “Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening.” *Journal of New Music Research*, vol. 33, no. 3, 2004, pp. 217–238., doi:10.1080/0929821042000317813.

MacDonald, Raymond A. R., et al. *Music, Health, and Wellbeing*. Oxford University Press, 2013.

Napier, Kathleen, and Lior Shamir. “Quantitative Sentiment Analysis of Lyrics in Popular Music.” *Journal of Popular Music Studies*, vol. 30, no. 4, 2018, pp. 161–176., doi:10.1525/jpms.2018.300411.

Piantadosi, Steven T. “Zipf’s Word Frequency Law in Natural Language: A Critical Review and Future Directions.” *Psychonomic Bulletin & Review*, vol. 21, no. 5, 2014, pp. 1112–1130., doi:10.3758/s13423-014-0585-6.

Robinson, J P, and P J Hirsch. "Teenage Response to Rock and Roll Protest Songs." *University of Michigan Institute of Social Research*, 1969.