Rock and Roll Sub-Genres

Jack Stapholz[1]

University of California - Los Angeles[a]

(Dated: 24 March 2024)

Any sort of music genre can be further divided into countless niche sub-genres. However, the more we partition these genres, the aspects that characterize each sub-genre becomes unclear. Using text mining and analysis, we can precisely define what defines a song of a given sub-genre and even classify which sub-genre a given song belongs to. In this study, we will copmile the lyrics to songs in multiple sub-genres of rock and roll music and tokenize the terms for analysis.

[a] jstapholz@ucla.edu

I. INTRODUCTION

A. Motivation

The motivation of this research project is to explore various rock and roll sub-genres, and how they are similar to each other. It is also of interest to be able to tell if a certain artist has unique lyrics, allowing us to predict who the artist is through the lyrics of their songs.

B. Research Problem

In order to narrow down our sample size, we're going to take a look at rock bands with a popularity score of 4 or higher. Artists are going to be one of a few from various sub-genres, such as '90s grunge, classic rock, and heavy metal. After building a corpus from our lyric data, we can clean up the text and tokenize. Using our tokenized words, we can then analyze sentiments and identify frequent and infrequent terms. One of the more important aspects in this project is modeling the topics through the lyrics of each band's songs. This will allow us to differentiate the aspects of each sub genre and uniqueness of individual artists. Based on the predictive ability of the lyrics of rock and roll sub-genres, we can determine how unique each one is from each other. Music data was derived from a dataset posted on Kaggle (Neisse, 2022).

C. Expected Outcome

The expected outcome of this research project is that specific sub-genres will have similar characteristics (sentiments, word choices, complexity, etc.). Even further, artists within these sub-genres will differentiate themselves from the other artists.

II. INTRODUCTION TO THE DATA

A. Data Selection

In order to reduce complexity of the study and allow for limitations in computing power, we will take a look at 3-4 artists from each sub-genre. From the '90s grunge sub-genre, we are analyzing Alice in Chains, Nirvana, and Pearl Jam. Classic Rock bands in this study include Bon Jovi, Johnny Cash, Queen, and the Rolling Stones. Metallica, Slipknot, and Iron Maiden are the metal bands that are being studied.

B. Previewing the data

```
## [1] 2524    9
```

```
##                    SName         Artist                    Genres Popularity
## 1 Sympathy For The Devil Rolling Stones Classic Rock; Rock; Blues       12.8
##   language Songs                                              SLink
## 1       en   438 /the-rolling-stones/sympathy-for-the-devil.html
##                   ALink
```

```
## 1 /the-rolling-stones/
```

Our lyrics data has 2,524 rows (each corresponding to a song) and 9 columns. Along with lyric data, each song comes with a song name, artist name, list of genre tags, popularity score, language, and total number of songs belonging to the artist.

III. PROCESSING THE DATA

A. Building the Corpus

A corpus is constructed with each song as a document. Other information contained in the data set is stored in the corpus as metadata.

B. Cleaning the Data

Extensive steps are taken to clean the raw lyric data from the dataset. First, symbols such as quotation marks, currency symbols, and accented letters are removed. Then, all letters are transformed to lower case. Both numbers and remaining punctuation are removed, along with stop words. Whitespace is stripped, and other miscellaneous words such as "ahhhhhhhhhhhhhhhhhhhh" are removed.

C. Tokenization

Lyrics are tokenized and categorized primarily based on genre.

D. Term Document Matrix

The following output is a summary of term document matrices of each genre.

```
## <<TermDocumentMatrix (terms: 5475, documents: 448)>>
## Non-/sparse entries: 21759/2431041
## Sparsity           : 99%
## Maximal term length: 17
## Weighting          : term frequency (tf)
```

```
## [1] "foxymophandlemama"
```

```
## <<TermDocumentMatrix (terms: 11301, documents: 1584)>>
## Non-/sparse entries: 99718/17801066
## Sparsity           : 99%
## Maximal term length: 21
## Weighting          : term frequency (tf)
```

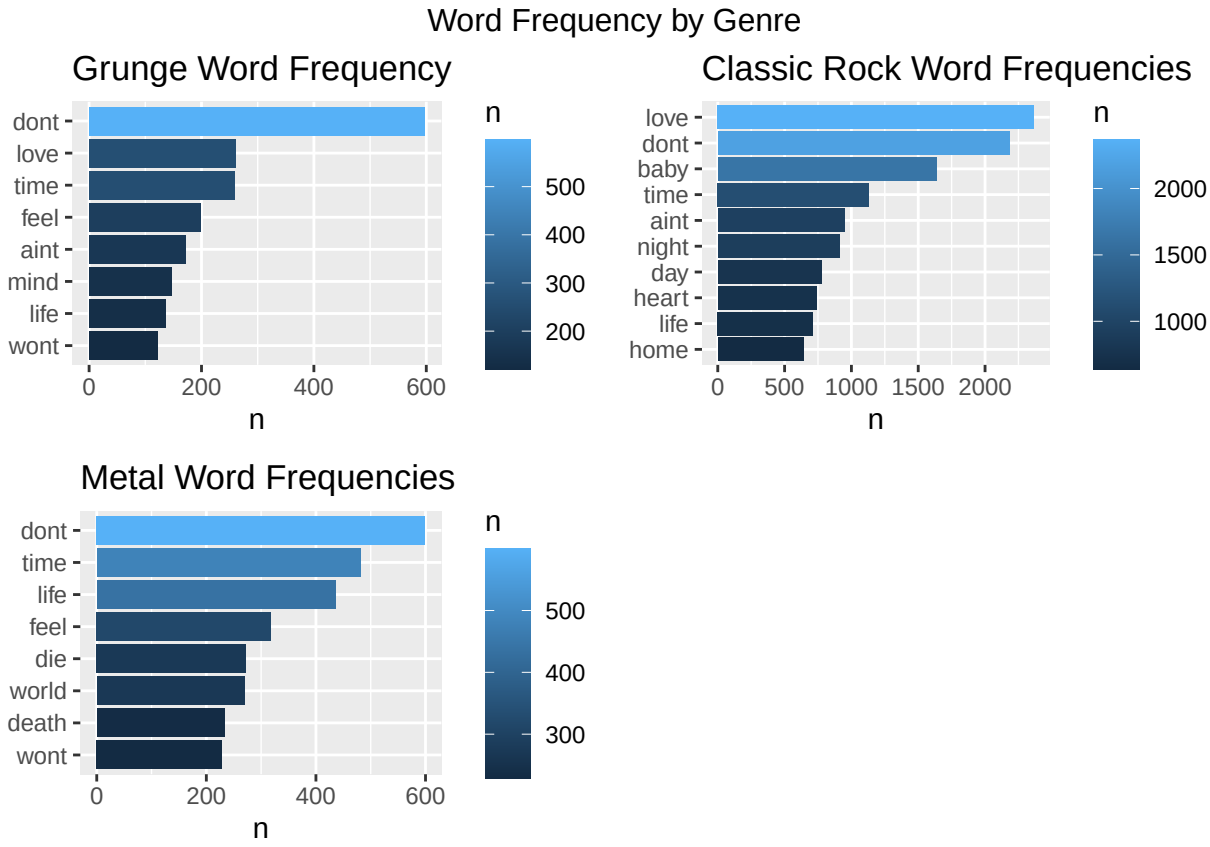
```
## [1] "parahumanoidarianised"
```

```
## <<TermDocumentMatrix (terms: 7215, documents: 492)>>
## Non-/sparse entries: 35588/3514192
## Sparsity           : 99%
## Maximal term length: 16
## Weighting          : term frequency (tf)
```

```
## [1] "sleepinspecially"
```

E. Word Frequency

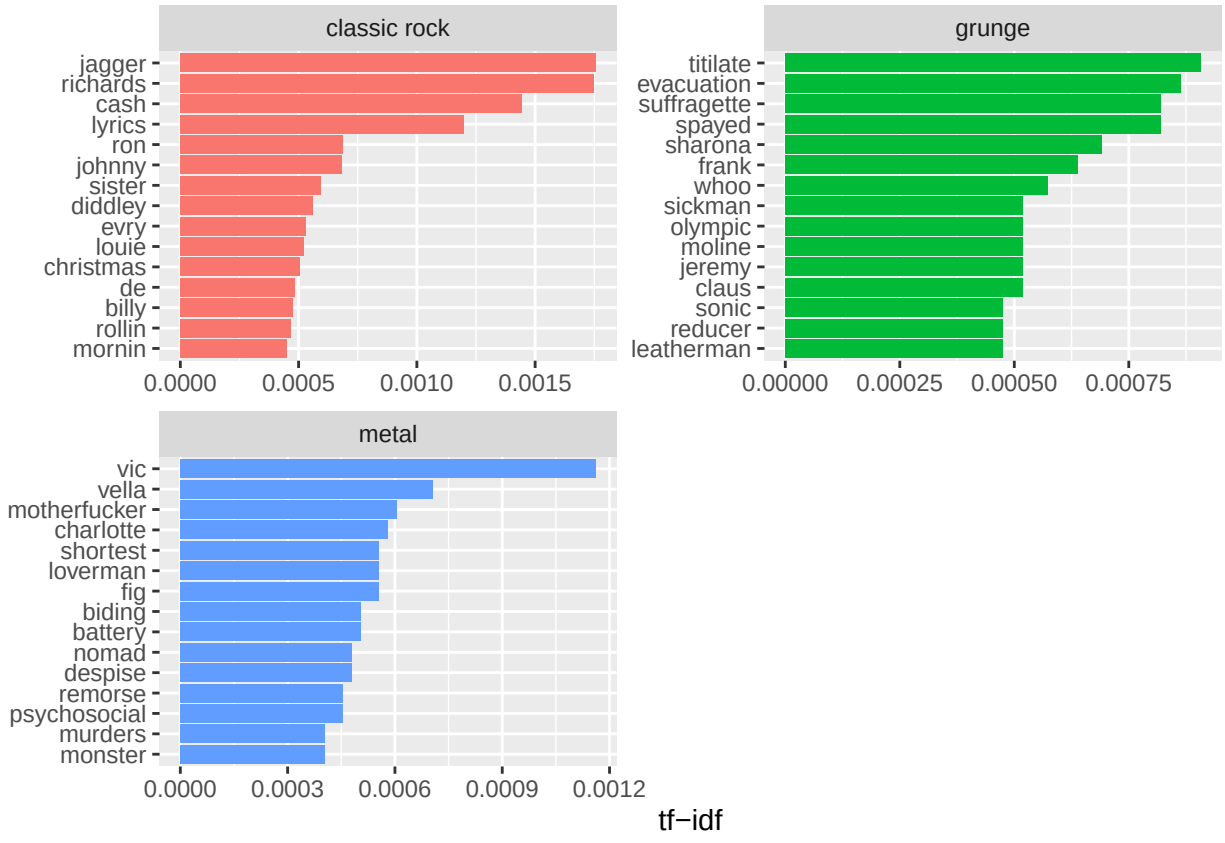
The following output of plots is a summary of the most frequent terms from each genre. The cutoffs for grunge, classic rock, and metal are n > 120, n > 600, and n > 225 respectively.



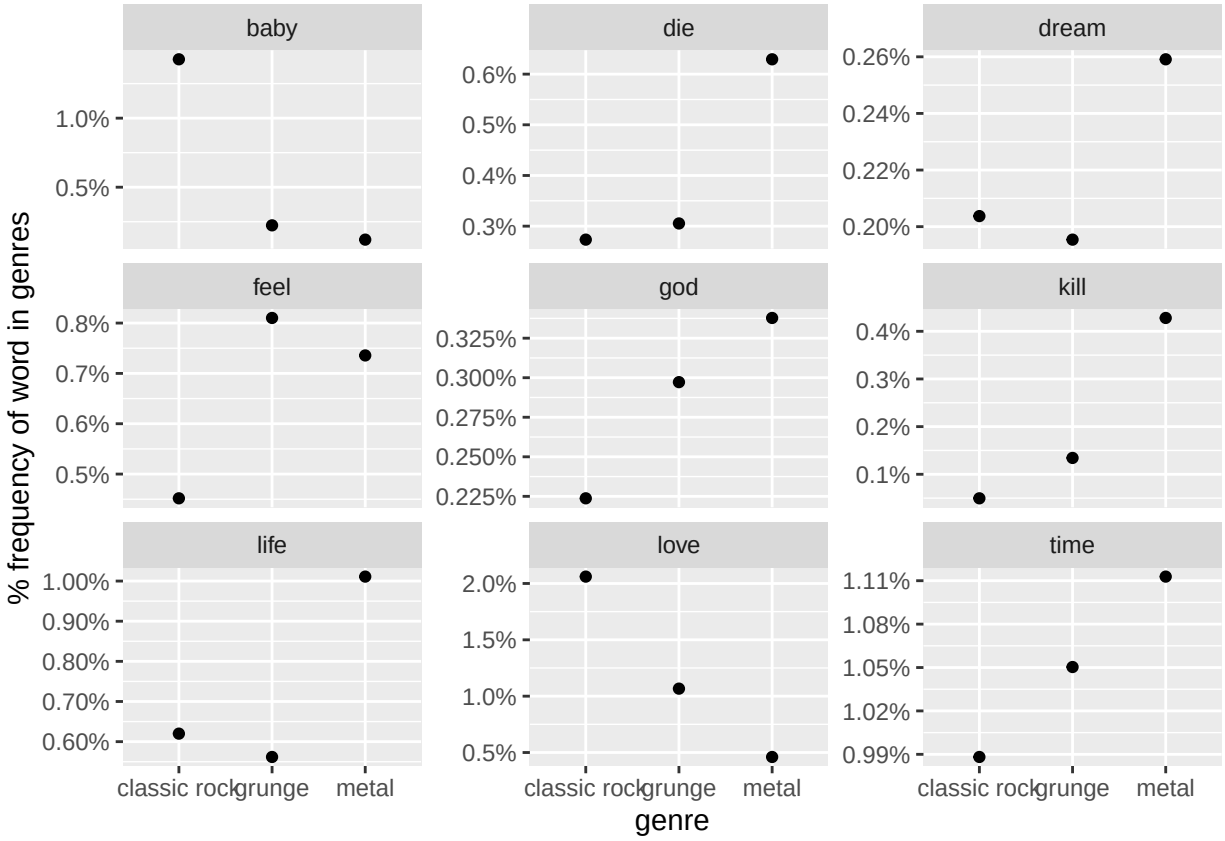
The tf-idf plots show the terms that are the most unique to each genre. Terms from grunge and classic rock genres are a combination of names and other unusual terms. Terms from the metal genre are characterized as exceptionally dark and vulgar.

```
## # A tibble: 3 x 7
##   word  genre            n  total     tf   idf tf_idf
##   <chr> <chr>        <int>  <int>  <dbl> <dbl>  <dbl>
## 1 love  classic rock  2366 120069 0.0197     0      0
```

```
## 2 dont  classic rock  2182 120069 0.0182     0       0
## 3 baby  classic rock  1637 120069 0.0136     0       0
```

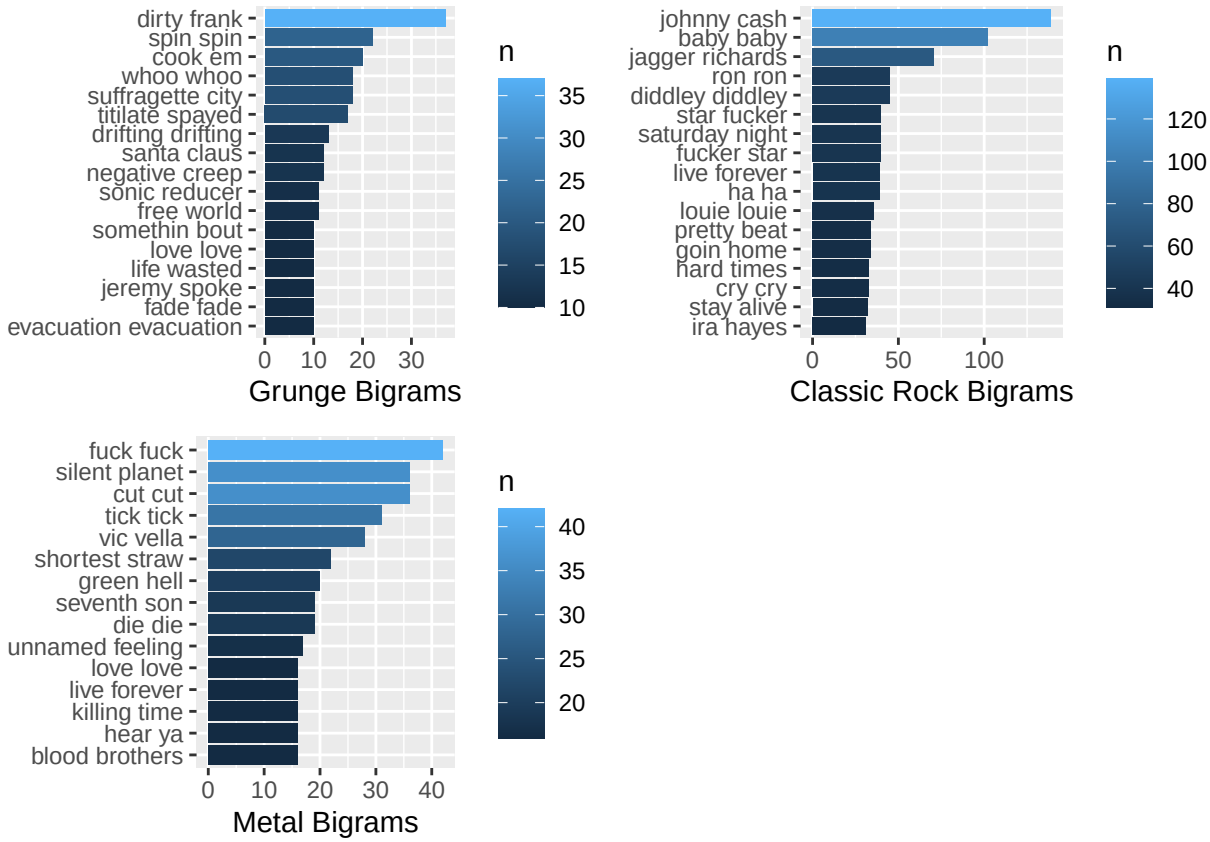


Taking a look at individual terms, there is a large gap between the frequency in each genre. Negative words like die and kill are very common in metal, and positive words such as love and baby are most common in classic rock.



F. N-Grams

N-Grams, specifically bi-grams as visualized by the plots below, show strong connections between two words that provide a deeper insight into things discussed within song lyrics. Similar to the tf-idf plots, names are common bi-grams. Other bi-grams shown below typically come from a specific song and are repeated commonly in a specific song.



G. Text Visualization

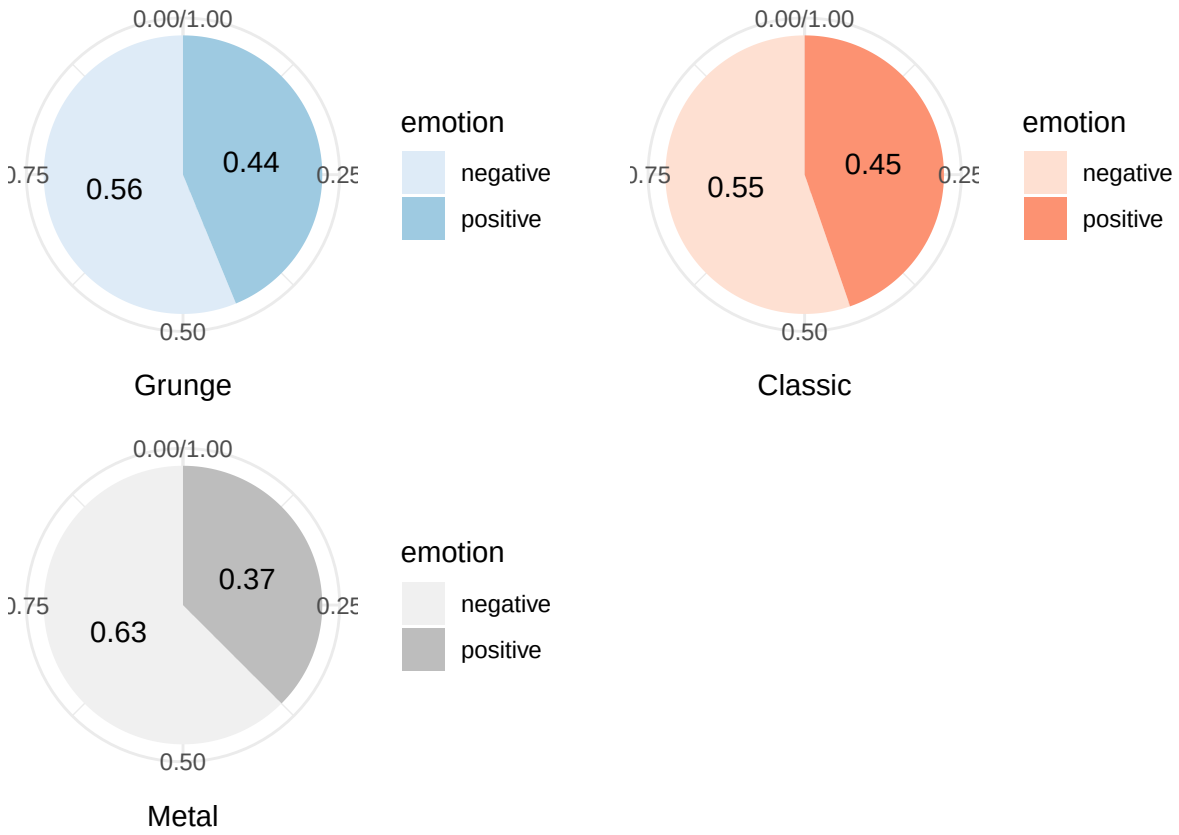
IV. SENTIMENT ANALYSIS

Pie charts of positive vs. negative sentiments are insightful to determine the general mood of songs from the genres covered in this study. Grunge and classic rock songs are practically equal when it comes to positivity with a ratio of ~55/45 positive to negative. These graphs provide a great insight into how metal differs from other rock sub-genres, since it has a ratio of 63/37. This is strong evidence to suggest metal music is much more negative than other rock and roll songs.

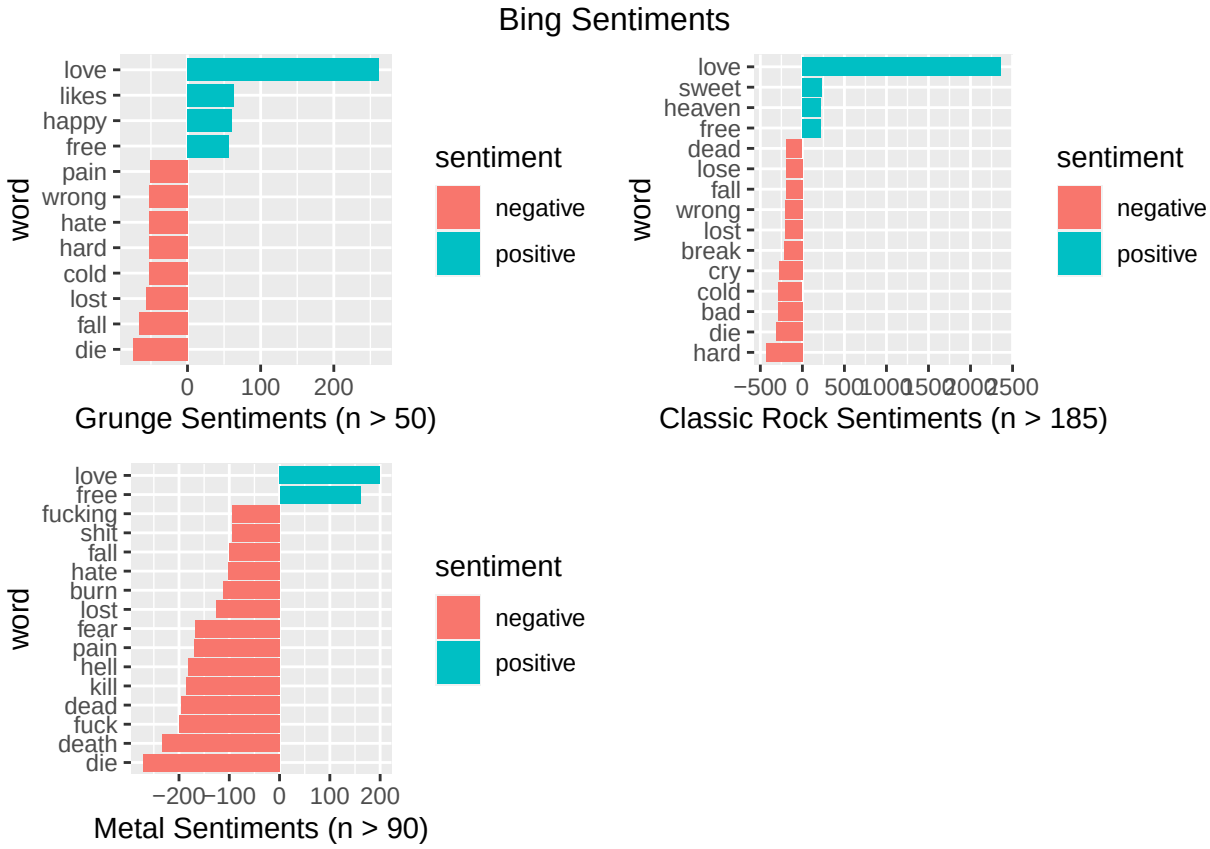
FIG. 1. Wordcloud2 of Grunge Words.

FIG. 2. Wordcloud2 of Classic Rock Words.

FIG. 3. Wordcloud2 of Metal Words.



These graphs show the frequency of sentiments within the rock and roll sub-genres. Given a specific cut-off relative to the number of terms from each genre, each plot shows a larger variety of terms used to describe negative sentiments. However, only heavy metal music has a more frequent use of negative words compared to positive ones. This is another strong piece of evidence to show how exactly metal differentiates from grunge and classic rock.

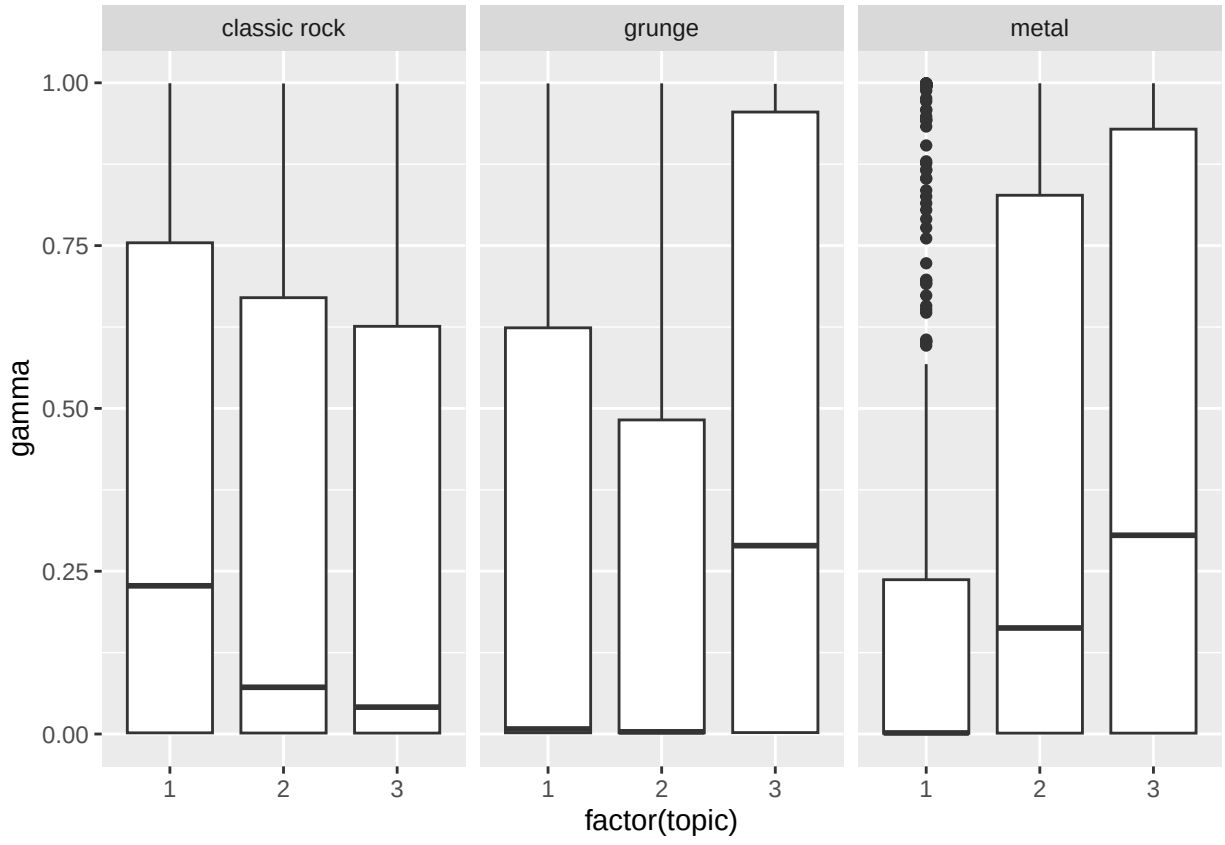


V. TOPIC MODELING

Topic modeling from the music tokens provides evidence contrary to the claim that these sub-genres are different from each other. There is strong cross over in pretty much every topic in every genre.

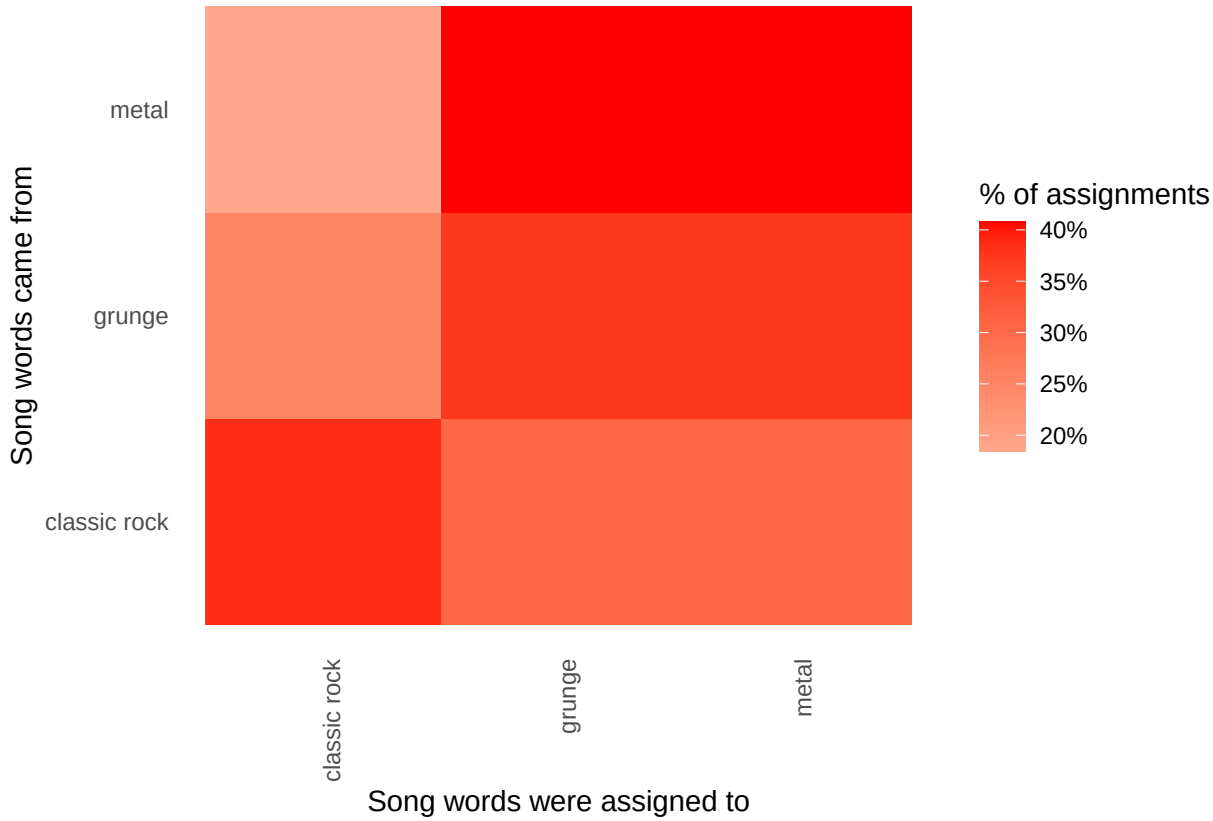
```
## # A tibble: 12 x 3
##    topic term     beta
##    <int> <chr>   <dbl>
##  1     1 love  0.0378
##  2     1 baby  0.0277
```

```
##  3     1 time  0.0145
##  4     1 heart 0.00857
##  5     2 time  0.00804
##  6     2 life  0.00793
##  7     2 home  0.00782
##  8     2 die   0.00690
##  9     3 life  0.0111
## 10     3 time  0.00902
## 11     3 night 0.00884
## 12     3 feel  0.00821
```



The following misclassification chart shows that there is immense crossover between topics discussed in metal and grunge music. However, classic rock differentiates itself from the other two genres in terms of topics. Classic rock is the genre most often misclassified.

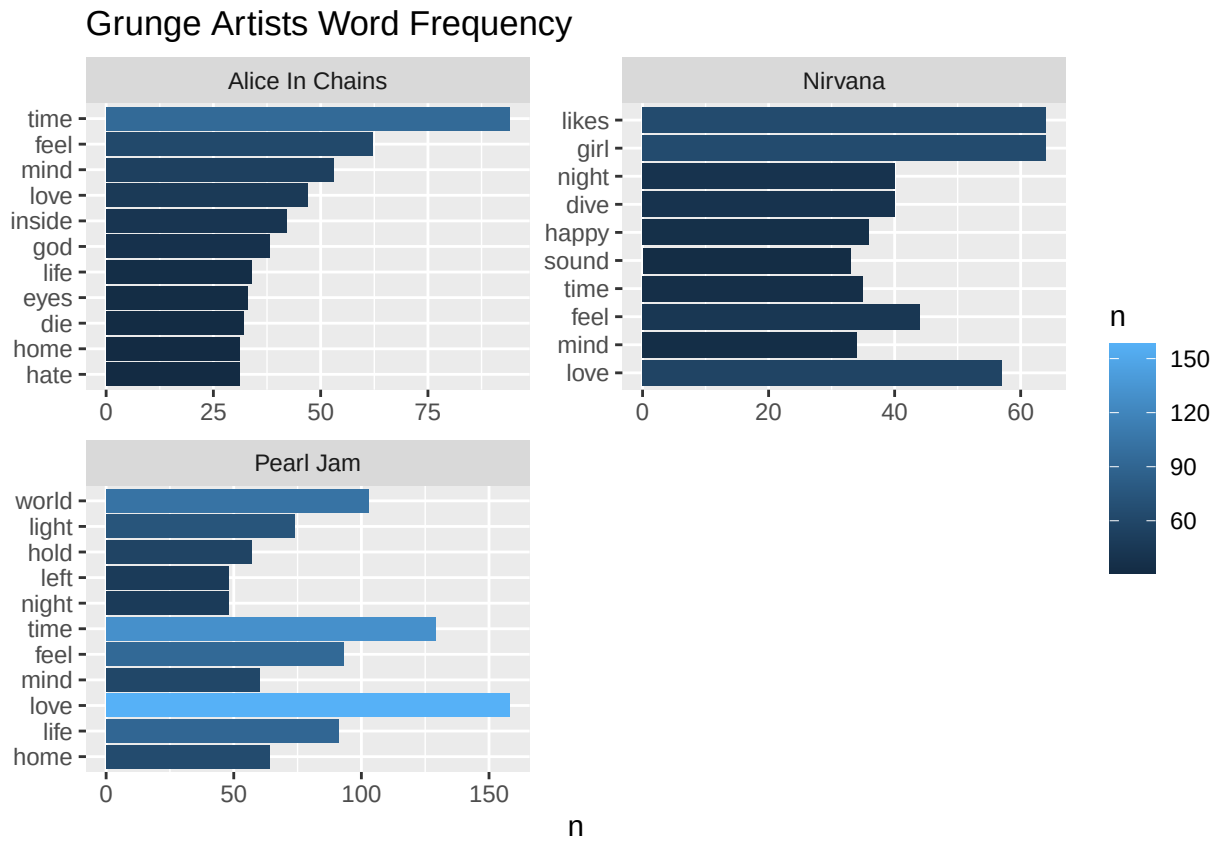
```
## classic rock
##          984
```



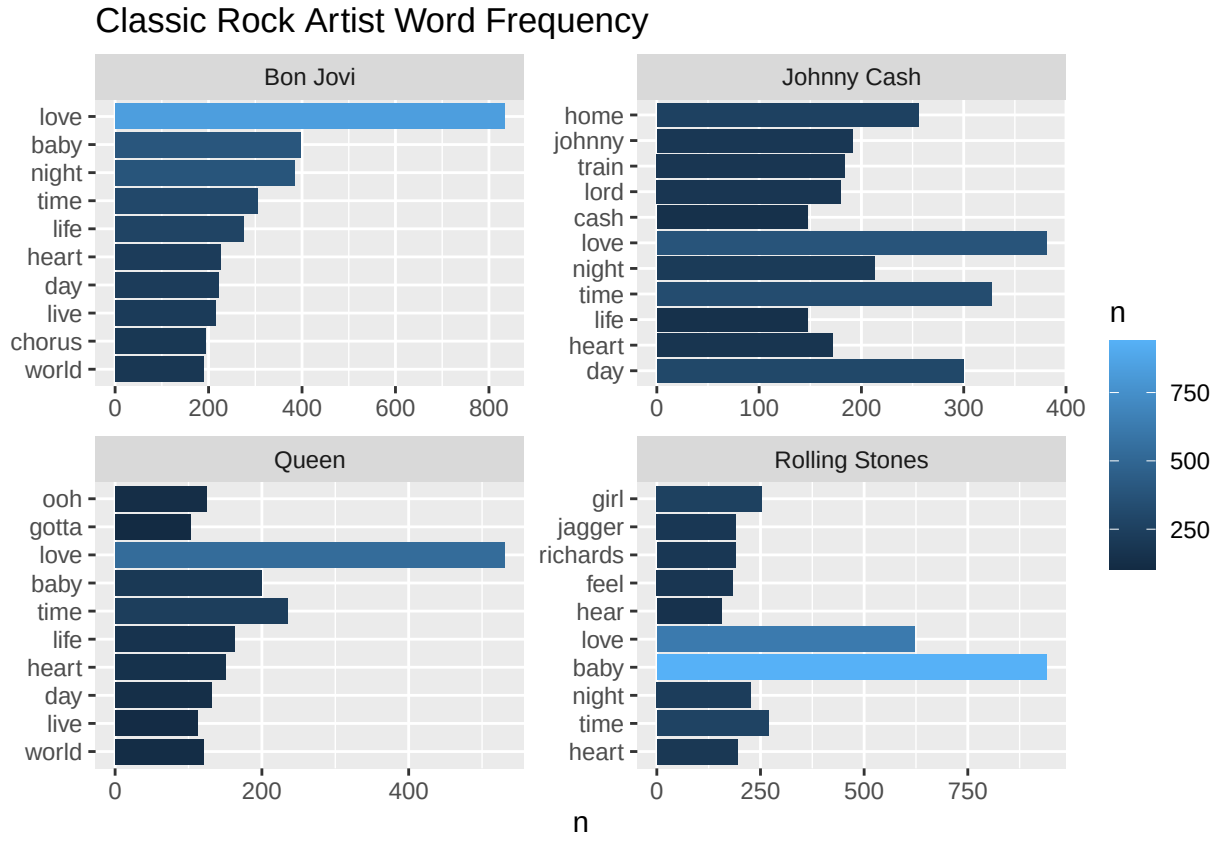
VI. COMPARING ARTISTS

A. Term Frequency

All of the grunge artists use a similar set of words at a similar frequency.

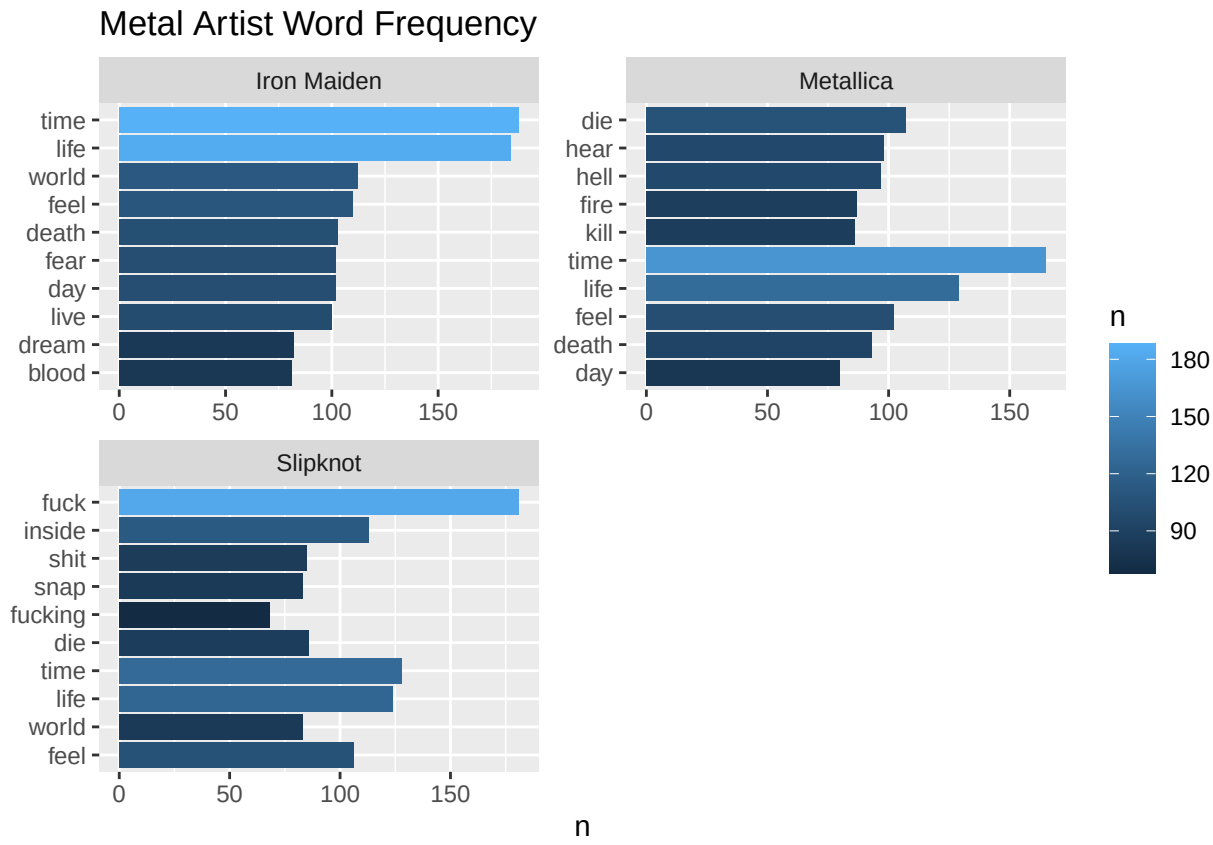

Grunge Artists Word Frequency
Alice In Chains
time
feel
mind
love
inside
god
life
eyes
die
home
hate
0
25
50
75
Nirvana
likes
girl
night
dive
happy
sound
time
feel
mind
love
0
20
40
60
Pearl Jam
world
light
hold
left
night
time
feel
mind
love
life
home
0
50
100
150
n
150
120
90
60
n
n

145


Classic Rock Artist Word Frequency
Bon Jovi
love
baby
night
time
life
heart
day
live
chorus
world
0
200
400
600
800
Johnny Cash
home
johnny
train
lord
cash
love
night
time
life
heart
day
0
100
200
300
400
Queen
ooh
gotta
love
baby
time
life
heart
day
live
world
0
200
400
Rolling Stones
girl
jagger
richards
feel
hear
love
baby
night
time
heart
0
250
500
750
n
750
500
250
n
n

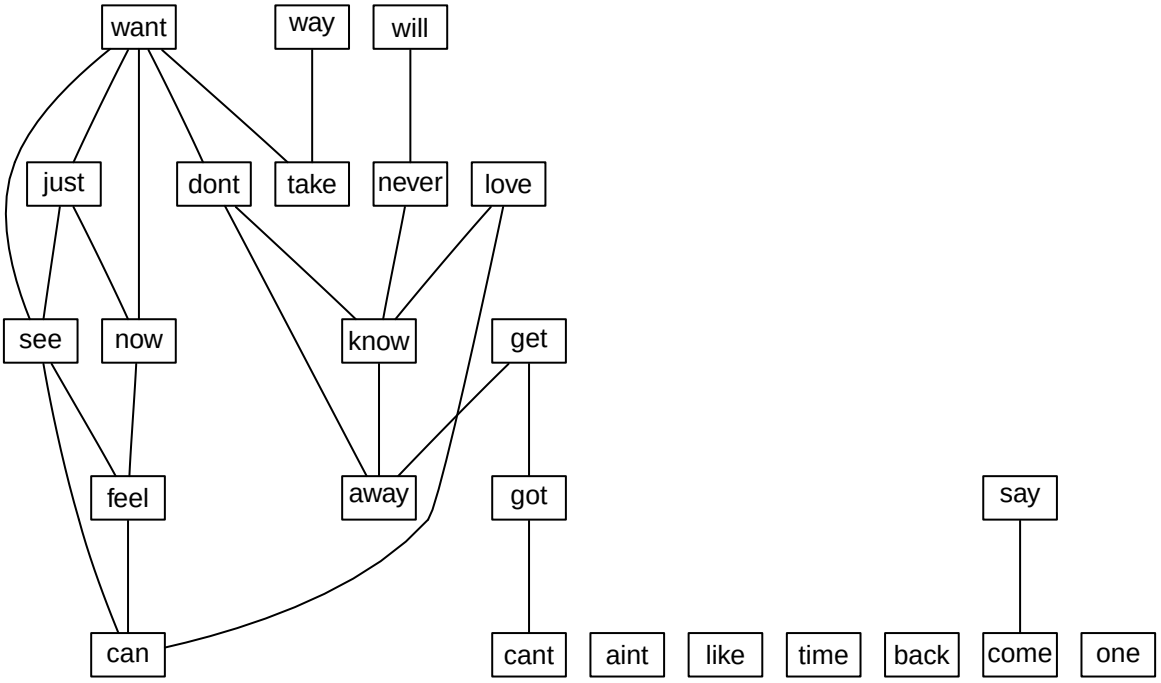
146

The same pattern holds with metal artists.



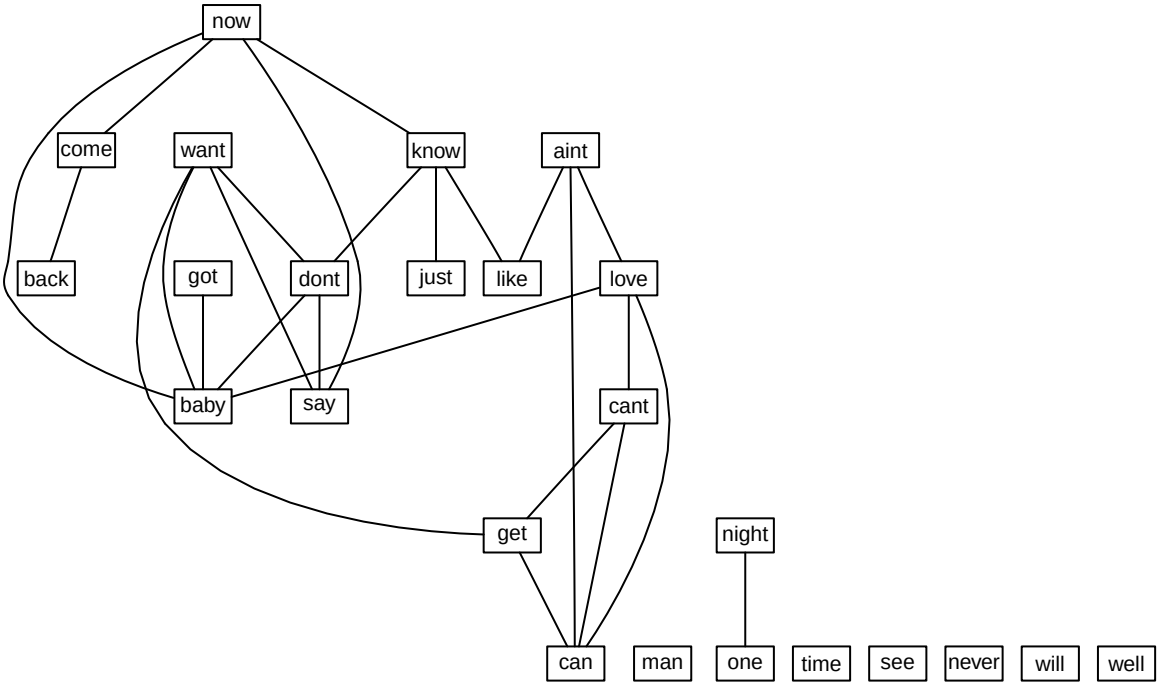
Even within the same sub-genre, terms have a low correlation with each other. The following plot is composed of grunge terms with a minimum frequency of 150 and a correlation threshold of 0.10.

Grunge Term Correlation Plot (Correlation Threshold of 0.10)

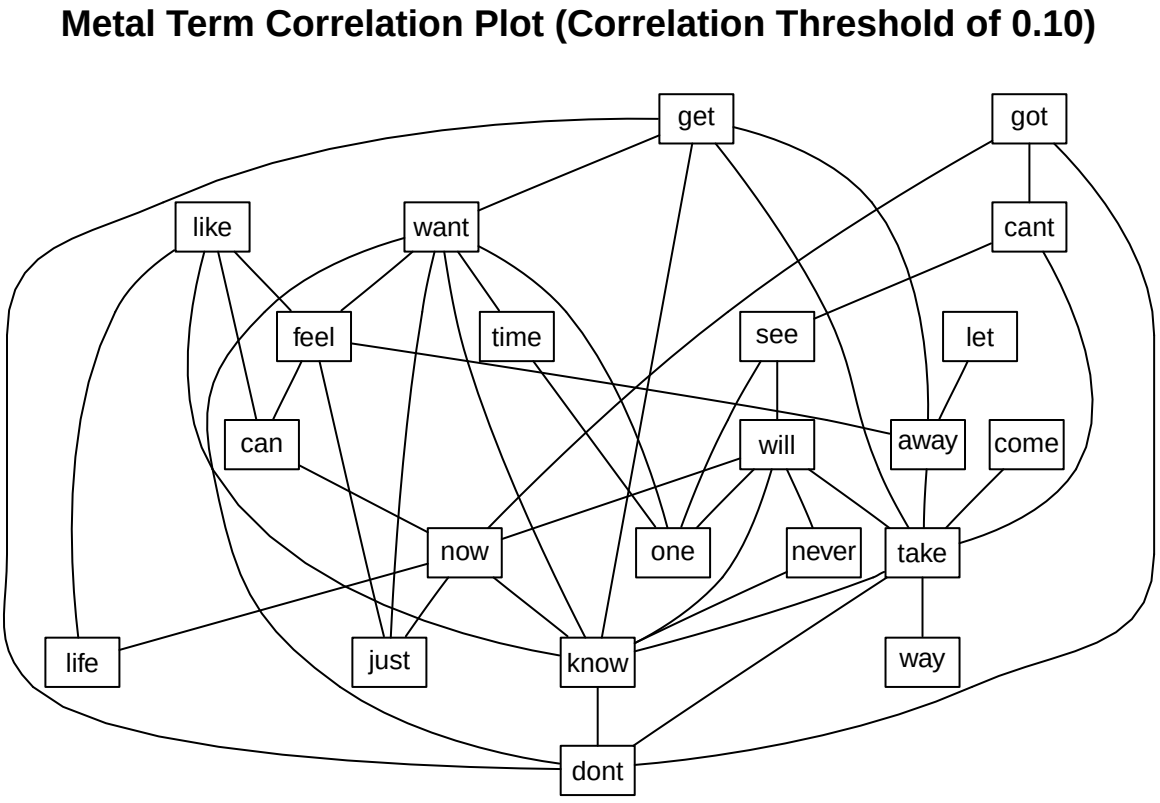


152

Classic Rock Term Correlation Plot (Correlation Threshold of 0.10)



153

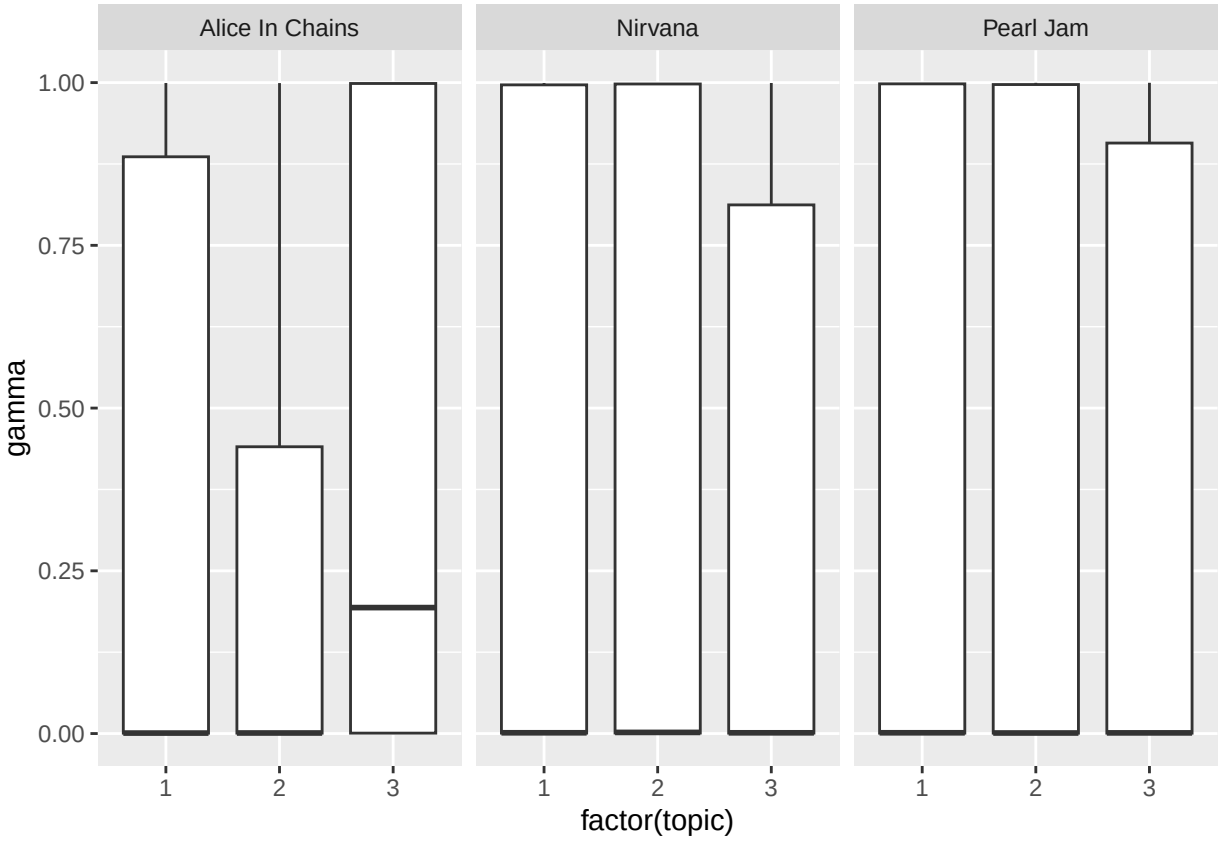


154

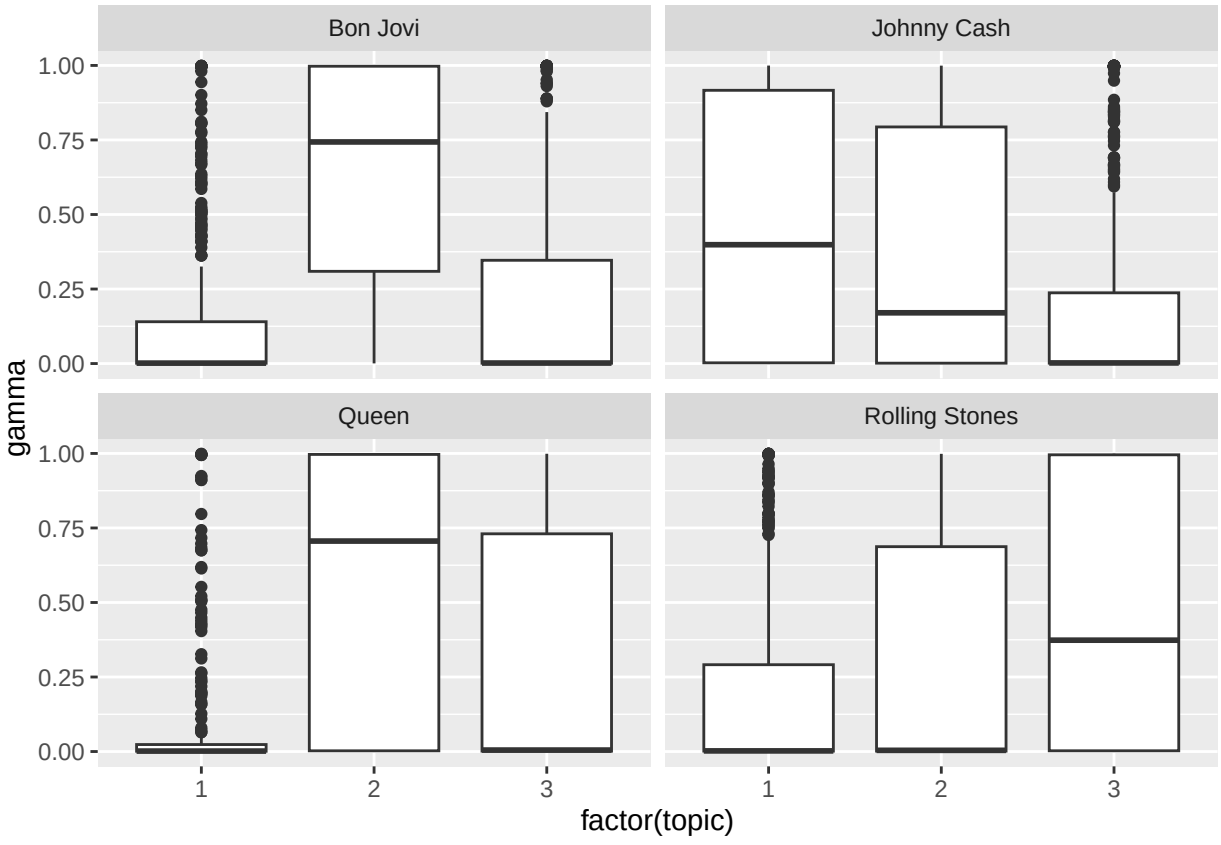
155 **B. Topic Modeling**

156 As with the genre comparison, all of the grunge artists have large overlaps with each

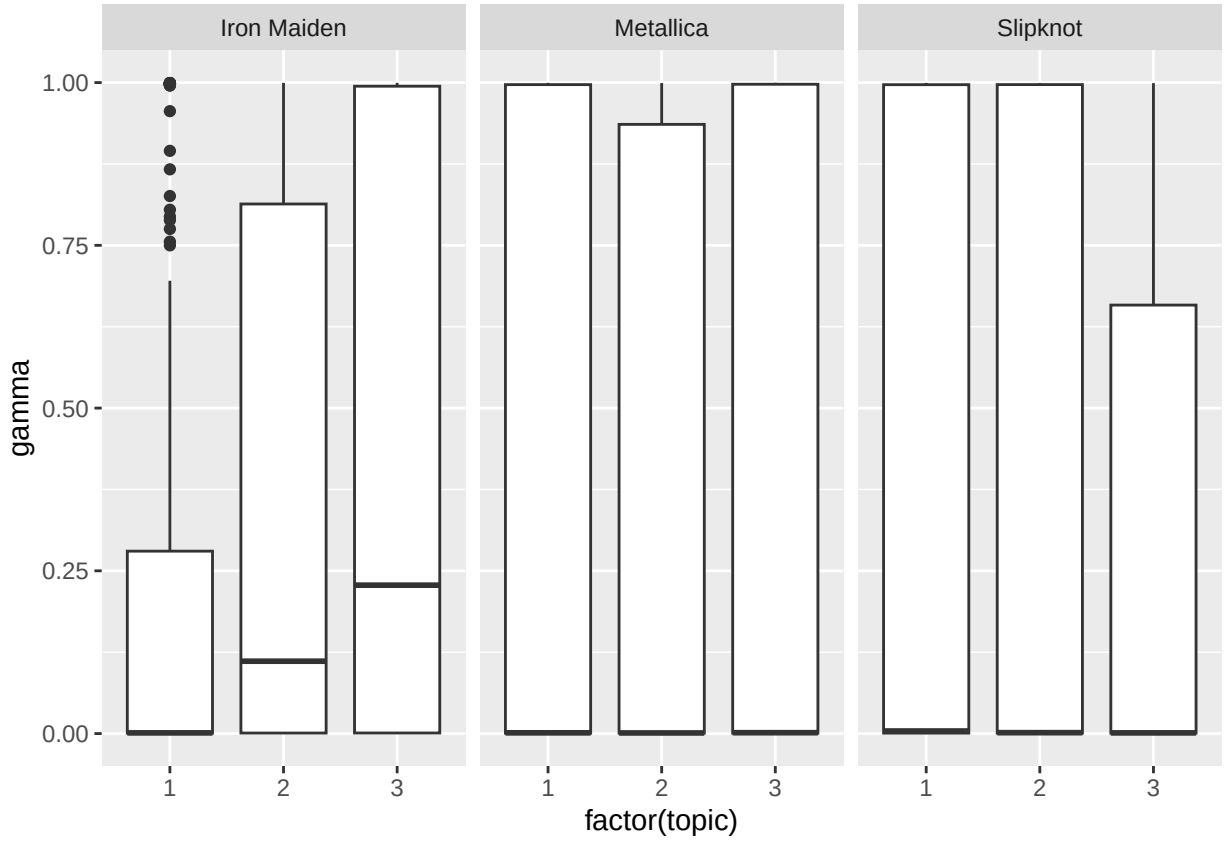
157 other. However, only one of the artist distinguishes themselves from the others.


Alice In Chains
Nirvana
Pearl Jam
gamma
1.00
0.75
0.50
0.25
0.00
1
2
3
factor(topic)

158

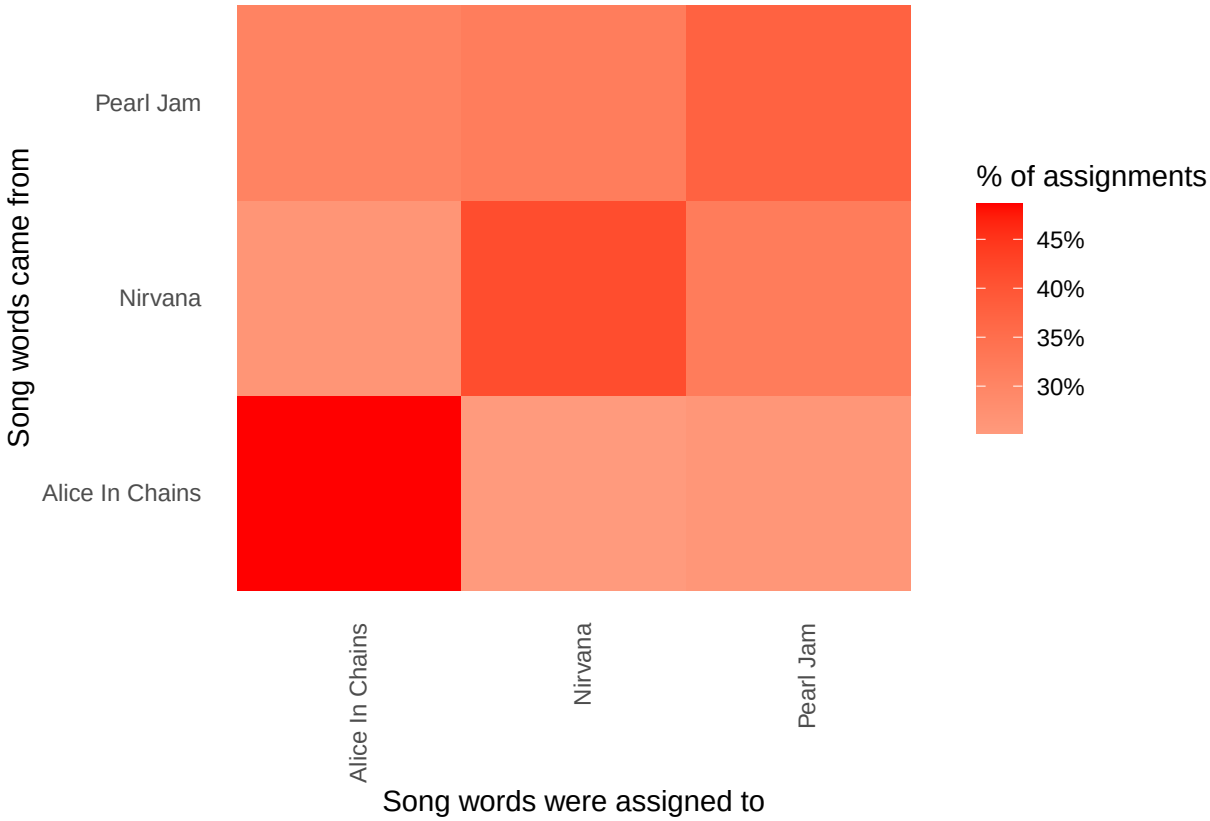

Bon Jovi
Johnny Cash
Queen
Rolling Stones
gamma
1.00
0.75
0.50
0.25
0.00
1
2
3
factor(topic)

159

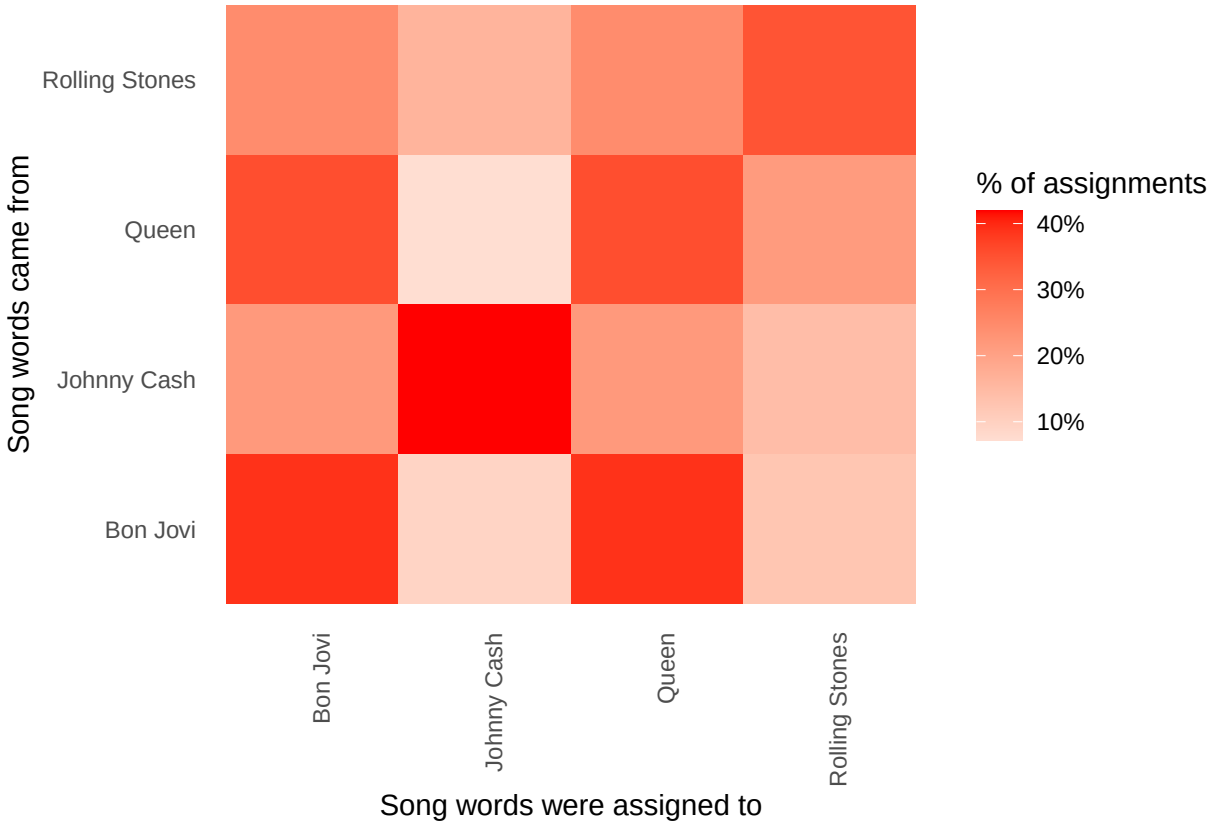


In this comparison chart, a majority of grunge songs are correctly identified with their corresponding artist, however they all have considerable overlap.

```
## Pearl Jam
##       139
```

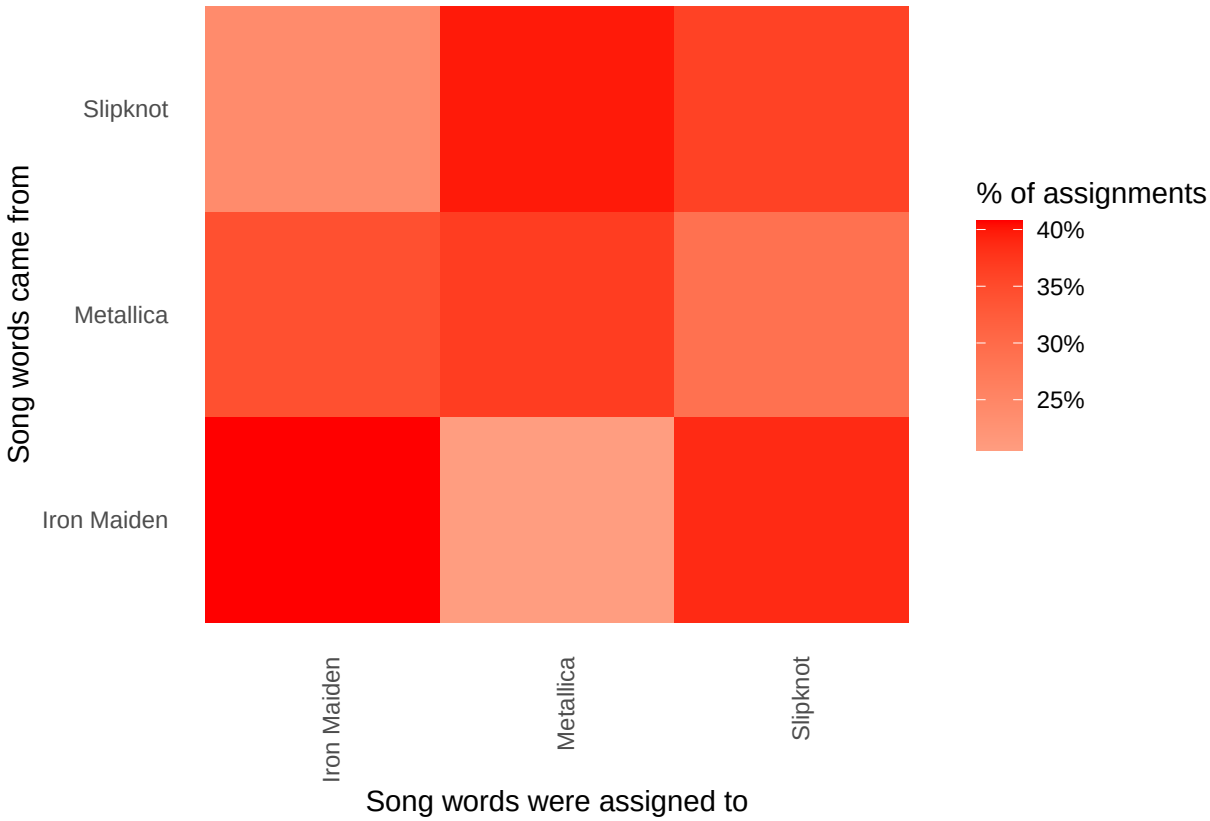



```
## Johnny Cash
##         513
```



Metal artists have considerably less distinction from each other.

```
## Iron Maiden
##         114
```



VII. CONCLUSION

Based on the above analysis, we can conclude that there are specific differences between sub-genres of rock and roll. Metal is classified by a variety of strongly negative and vulgar terms. However, grunge and classic rock are somewhat similar in this category. The ratio of positive to negative terms for these two genres are roughly equal. Topic modeling paints a different story. It suggests large overlap between metal and grunge in the things discussed in their lyrics. Classic rock stands by itself in terms of the ideas discussed in its music. Artists within a specific genre use similar terminology, however grunge artists write lyrics on unique topics. The same cannot be said for metal artists, as it is almost impossible to

determine correctly which artist a particular song belongs to. However, classic rock artists fall somewhere in the middle, as out of the four artists a song typically is assigned mostly to one or two.

Overall, although grunge and classic rock uses similar wording, grunge's lyrics more closely align with heavy metal. Each sub-genre typically has its own set of commonly used terms, and they have a varying degree of originality and uniqueness.

Neisse, A. (**2022**). "Song lyrics from 79 musical genres" (Kaggle).