What is Iron Maiden, Taylor Swift and Pharrell Williams singing? A primer on text mining.*

Walter Guillioli

April 03, 2025

Introduction

I love music and data science. Purpose of this paper is to apply main text analytics techniques to explore lyrics from different authors.

We will see things like sentiment analysis, word clouds, topic model and network bigrams.

Purpose is to process lyrics and get idea of what they sign without reading lyrics.

A key companino for this was the book tidy text oby x (add link)

Data and Methods

Lyrics from 4 albums were dowloaded from the azlyrics.com website. They were manually copy pasted to text documents and read into R. Two albums for Iron Maiden - Number of the Beast (1982) and X-Factor (1995). For Taylor Swift, 1989 from 2014. And Pharrell Williams album x from x.

Data was read from txt into data frame. The techniques applied where: list all of them.

48 songs were analyzed from these albums: Iron Maiden: Number of the Beast 9,
Iron Maiden: X Factor 11, Pharrell Williams: GIRL 12, Taylor Swift: 1989 16.

Table 1 shows the names of the songs of each album.

```
## # A tibble: 48 x 3
##    artist      album               song
##    <chr>       <chr>               <chr>
##  1 Iron Maiden Number of the Beast Invaders
##  2 Iron Maiden Number of the Beast Children of the Damend
##  3 Iron Maiden Number of the Beast The Prisoner
##  4 Iron Maiden Number of the Beast 22 Acacia Avenue
##  5 Iron Maiden Number of the Beast The Number Of The Beast
##  6 Iron Maiden Number of the Beast Run to the Hills
##  7 Iron Maiden Number of the Beast Gangland
##  8 Iron Maiden Number of the Beast Hallowed be thy Name
##  9 Iron Maiden Number of the Beast Total Eclipse
## 10 Iron Maiden X Factor            Sign of the Cross
## # i 38 more rows
```

*The data and scripsts posted here: https://github.com/wguillioli

The lyrics were split into invidual words for analysis and common words (stop words) like a, on, the were removed. Table 2 shows number of words for analysis:

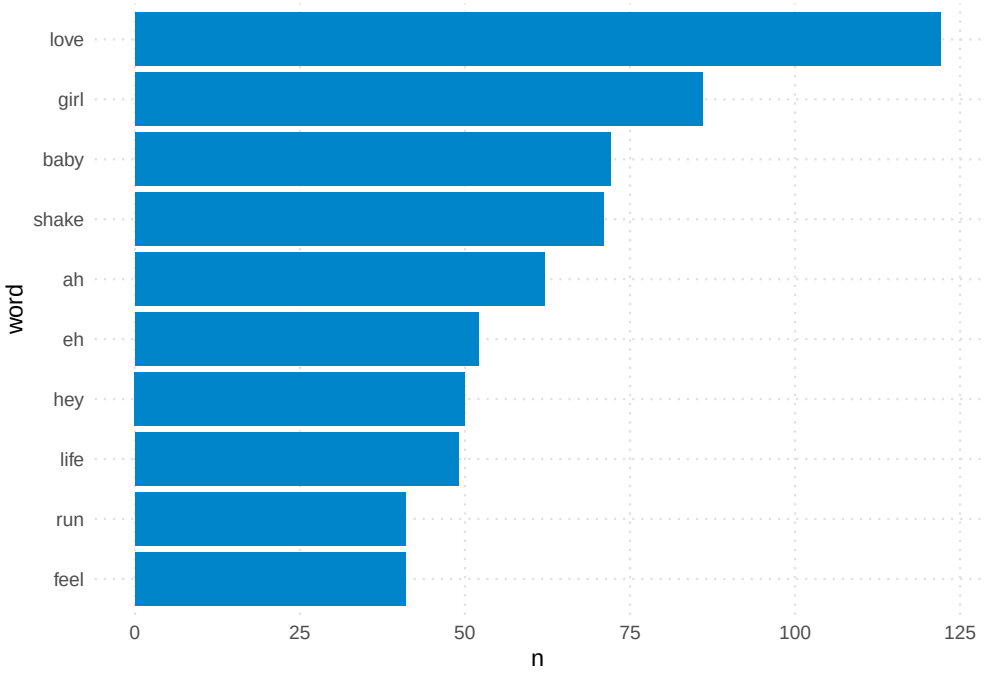
```
## # A tibble: 4 x 2
##   artist_album                      number_of_words
##   <chr>                                       <int>
## 1 Iron Maiden: Number of the Beast              800
## 2 Iron Maiden: X Factor                         871
## 3 Pharrell Williams: GIRL                      1437
## 4 Taylor Swift: 1989                           2000
```

Results

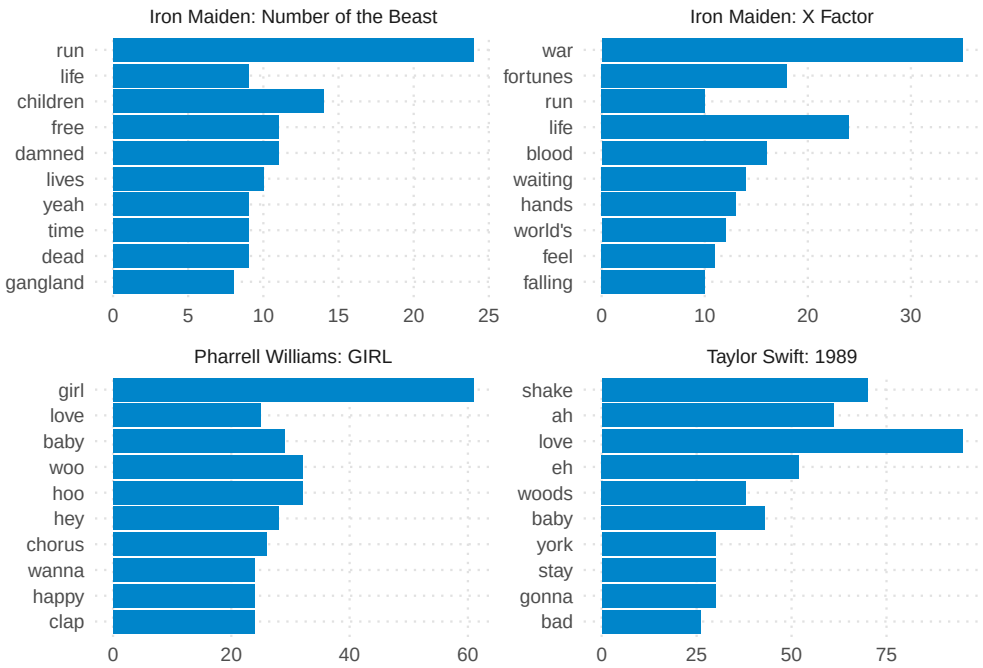
To analyze text we start with x, then do y and then z.

Top words used on the lyrics

We can start by plotting the top 10 most common words across all albums.



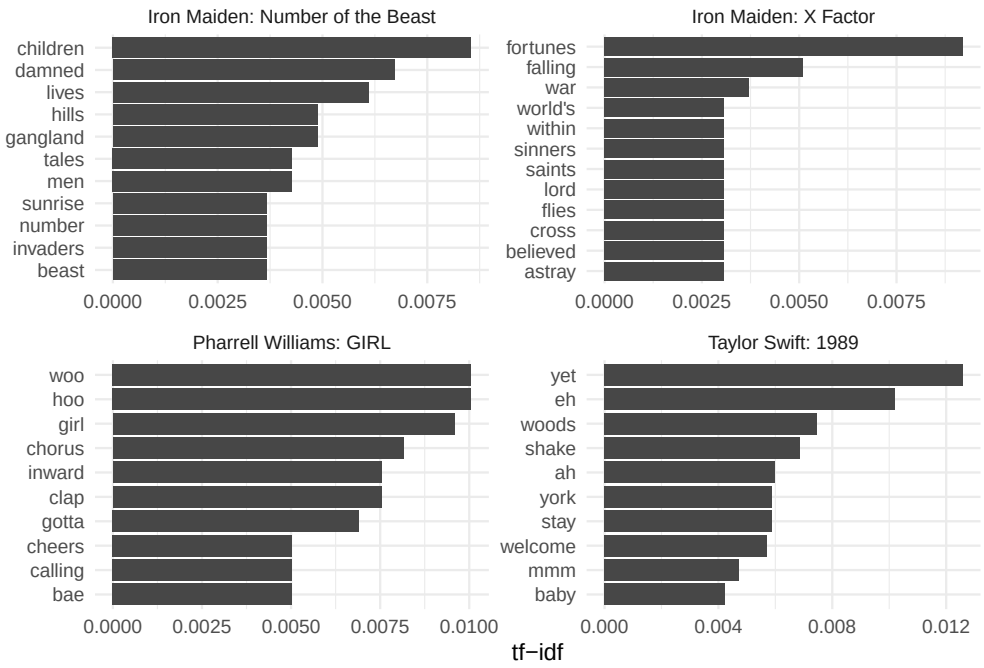
This is relatively interesting but not useful since we know some artists use more words than others. So let's do the same but by album.



Alternative we could use a more advanced way via tfidf and these are the results. so discuss and pick one and / or send one to appendix.

```
## Joining with `by = join_by(artist_album)`
```

```
## # A tibble: 2,846 x 6
##    artist_album                     word         n      tf   idf  tf_idf
##    <chr>                            <chr>    <int>   <dbl> <dbl>   <dbl>
##  1 Taylor Swift: 1989               yet         64 0.00907 1.39  0.0126
##  2 Taylor Swift: 1989               eh          52 0.00737 1.39  0.0102
##  3 Pharrell Williams: GIRL          hoo         32 0.00725 1.39  0.0100
##  4 Pharrell Williams: GIRL          woo         32 0.00725 1.39  0.0100
##  5 Pharrell Williams: GIRL          girl        61 0.0138  0.693 0.00958
##  6 Iron Maiden: X Factor            fortunes    18 0.00663 1.39  0.00920
##  7 Iron Maiden: Number of the Beast children    14 0.00616 1.39  0.00854
##  8 Pharrell Williams: GIRL          chorus      26 0.00589 1.39  0.00816
##  9 Pharrell Williams: GIRL          clap        24 0.00544 1.39  0.00754
## 10 Pharrell Williams: GIRL          inward      24 0.00544 1.39  0.00754
## # i 2,836 more rows
```



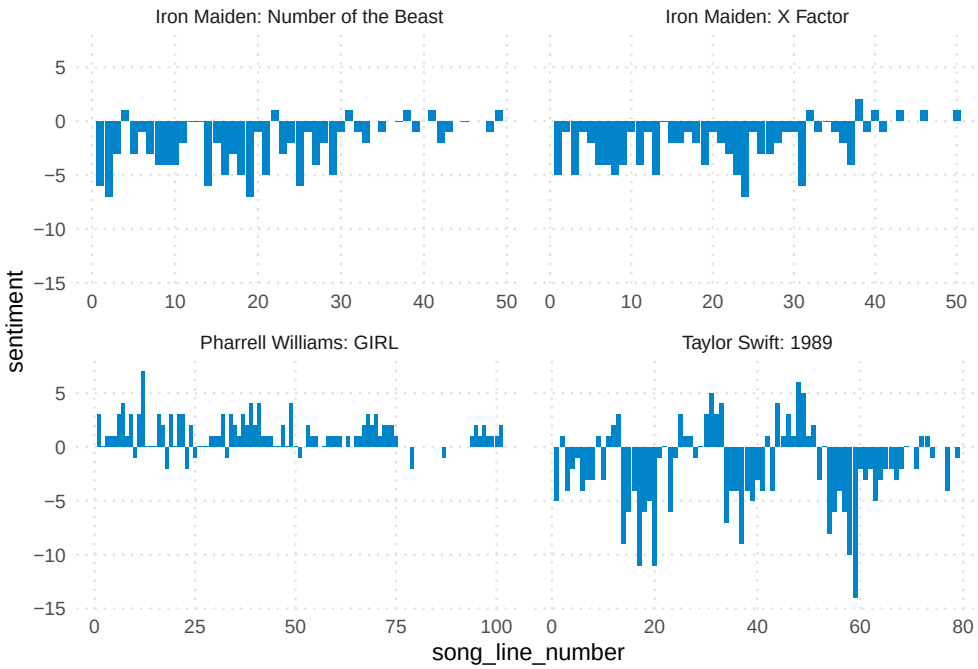
More informally we can see some wordclouds to get an idea of what they talk. This just confirms what we saw above in more detail.

Sentiment

The top words above give an idea of what they talk about but it's just the start. The next thing is to see if they sing about happy, sad or angry stuff. Probably a bit of everything.

We start with a simple analysis to see if the sentiment is positive or negative. In the chart below we see the sentiment evolution of the albums as the album progress where each bar represents a line in the song. Also note that number of lines varies and Maiden sings with less words.

```
## Joining with ‘by = join_by(word)‘
```



Several points to highlight from this chart. Iron Maiden tends to be very negative while Taylor Swift is mostly negative with some exceptions. Pharrell Williams on the contrary seems to be a very happy and uplifting singer. We will explore why and more in the next section.

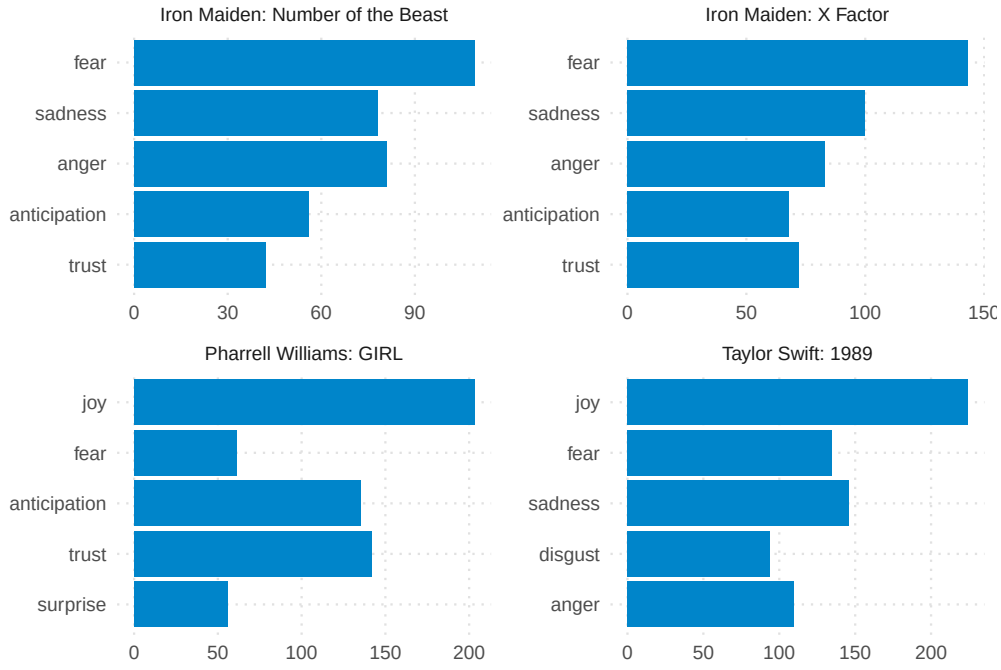
Another thing we can do is to see the sentiment for each album. Maiden is motly about fear, sadness and anger. While Pharrell sings about Joy and Trust. While Taylor seems to be happier but sometimes is inundated with sadness and fear. As we will see later this is a diffrent fear than the one Maiden shows.

```
## Joining with `by = join_by(word)`
```

```
## # A tibble: 20 x 3
## # Groups:   artist_album [4]
##    artist_album                     sentiment        n
##    <chr>                            <chr>        <int>
##  1 Iron Maiden: Number of the Beast fear           109
##  2 Iron Maiden: Number of the Beast anger           81
##  3 Iron Maiden: Number of the Beast sadness         78
##  4 Iron Maiden: Number of the Beast anticipation    56
##  5 Iron Maiden: Number of the Beast trust           42
##  6 Iron Maiden: X Factor            fear           143
##  7 Iron Maiden: X Factor            sadness        100
##  8 Iron Maiden: X Factor            anger           83
##  9 Iron Maiden: X Factor            trust           72
## 10 Iron Maiden: X Factor            anticipation    68
## 11 Pharrell Williams: GIRL          joy            203
## 12 Pharrell Williams: GIRL          trust          142
## 13 Pharrell Williams: GIRL          anticipation   135
## 14 Pharrell Williams: GIRL          fear            61
## 15 Pharrell Williams: GIRL          surprise        56
## 16 Taylor Swift: 1989               joy            224
## 17 Taylor Swift: 1989               sadness        146
```

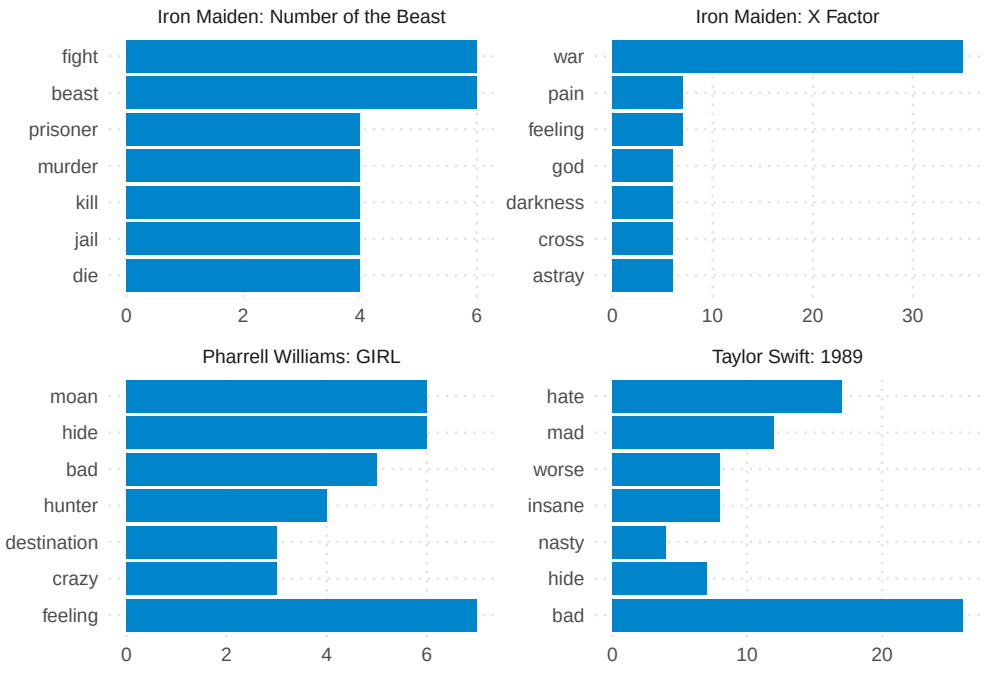
```
## 18 Taylor Swift: 1989              fear           135
## 19 Taylor Swift: 1989              anger          110
## 20 Taylor Swift: 1989              disgust         94
```

```
## Joining with `by = join_by(word)`
```



A common feeling across all albums is the sentiment of fear. So next we explore the top fear words of each album.

```
## Joining with `by = join_by(word)`
```

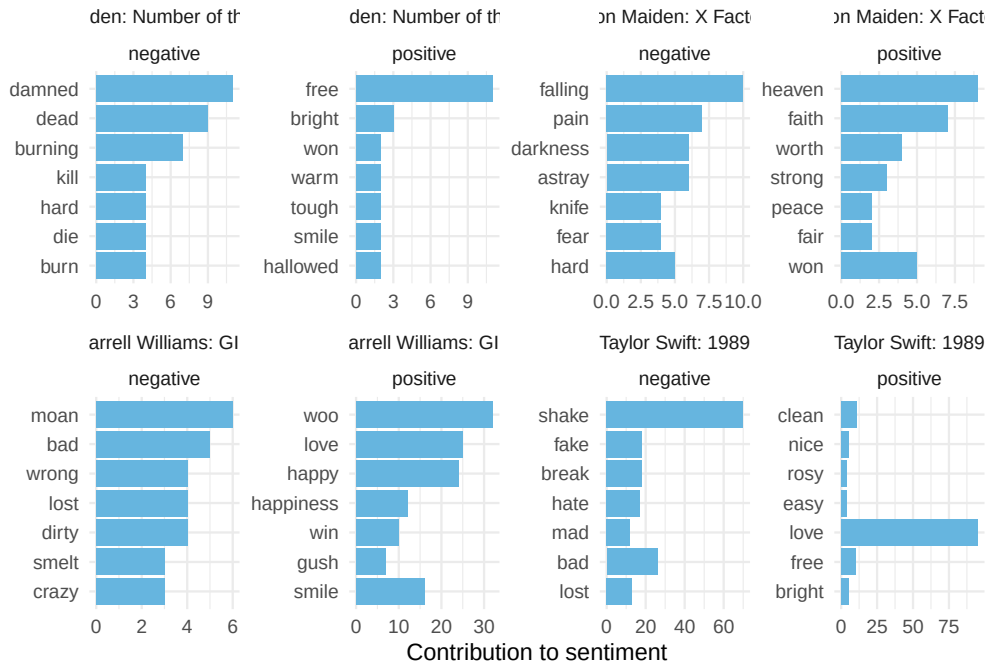


So even though they are all fear words, the meaning is different. TAke Iron Maiden is talking about fight, kill, murder, beast, war. Whille Pharrell talks about his feelings and moan. Even though hunter appears it's a love song about hunting for his girl. Taylor swift with bad, hate and mad clearly is talking about a bad romance.

Positive and negative words

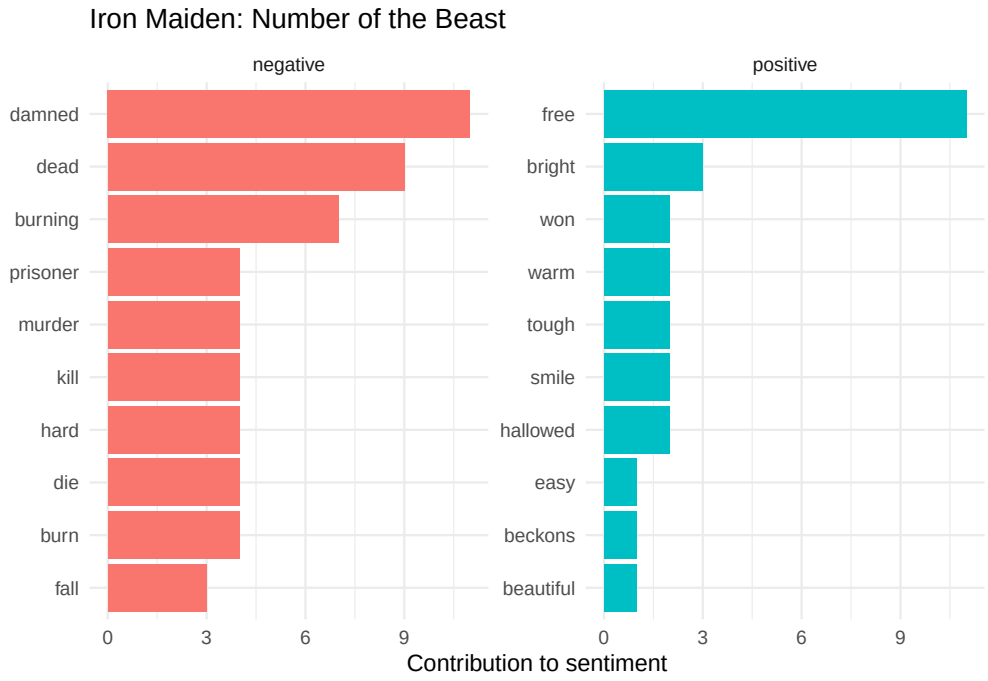
Exploring the words further we can take a look at the most positive and negative words per album.

```
## Joining with `by = join_by(word)`
```



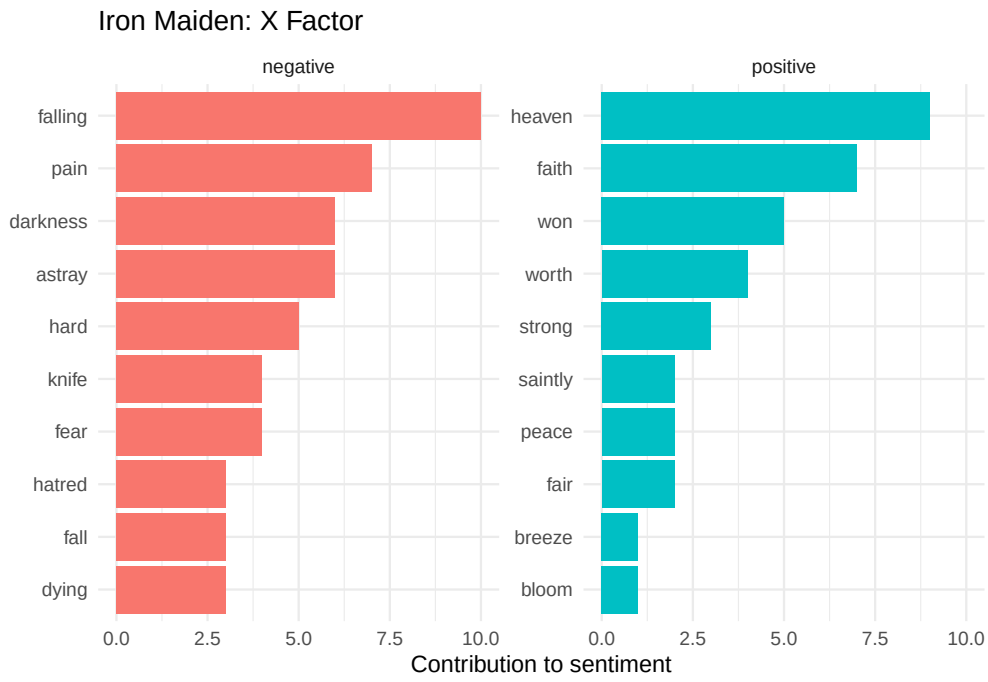

```
## [1] 1
```

```
## Joining with `by = join_by(word)`
```



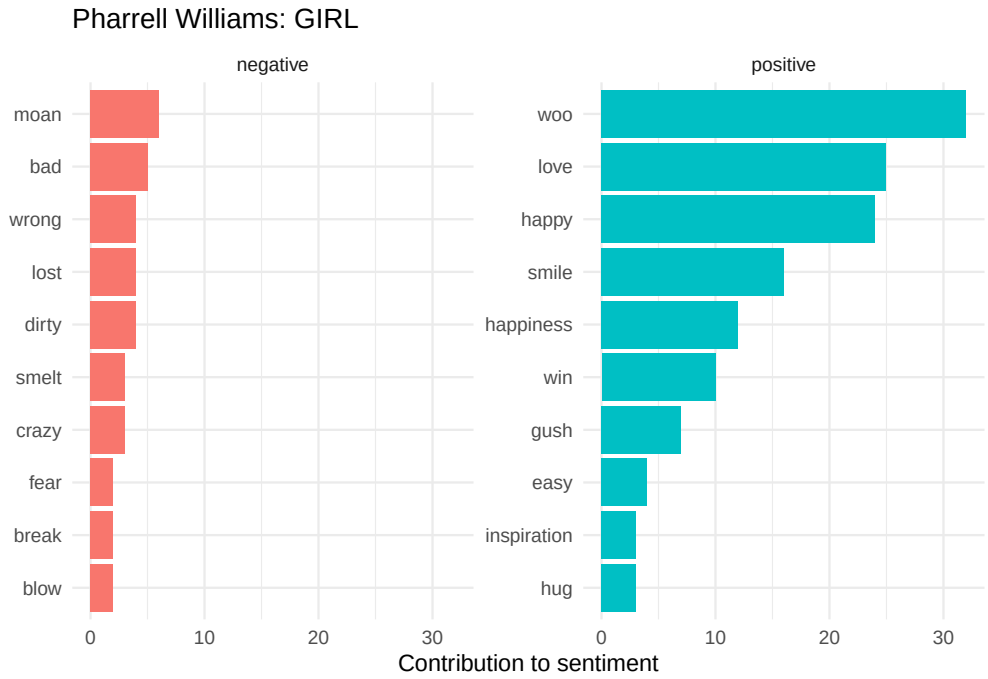

```
## [1] 2
```

```
## Joining with `by = join_by(word)`
```



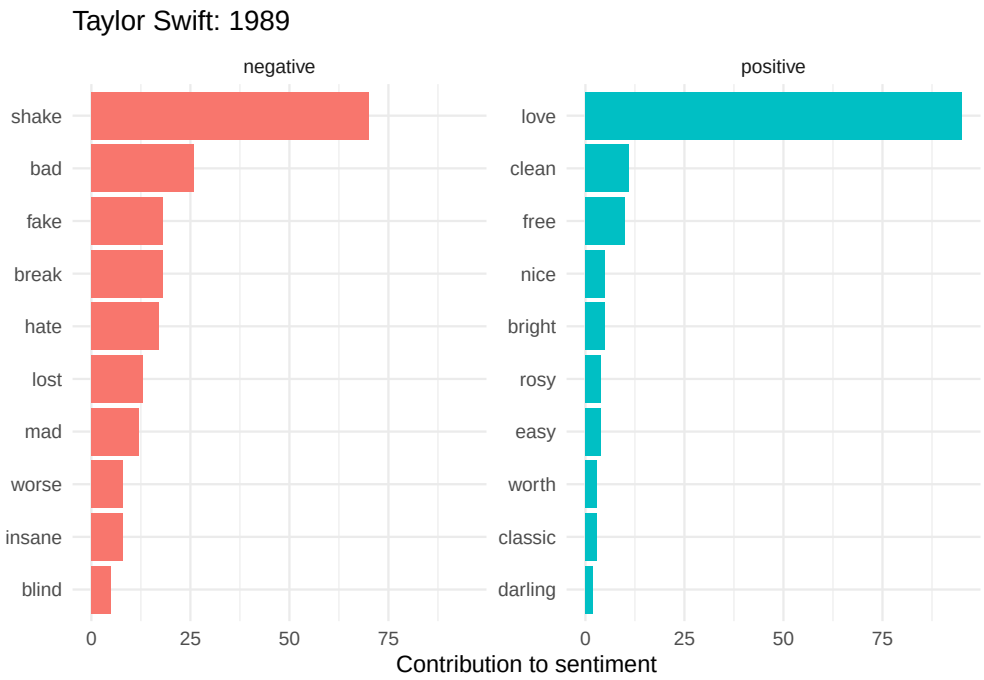

```
## [1] 3
```

```
## Joining with `by = join_by(word)`
```




```
## [1] 4
```

```
## Joining with `by = join_by(word)`
```



Maiden best is interesting. Free is positive but it's actually from running free and it's someone escaping from the police. Hallowed is also from a song called hallowed by the name from someone on death row. so this can be misleading.

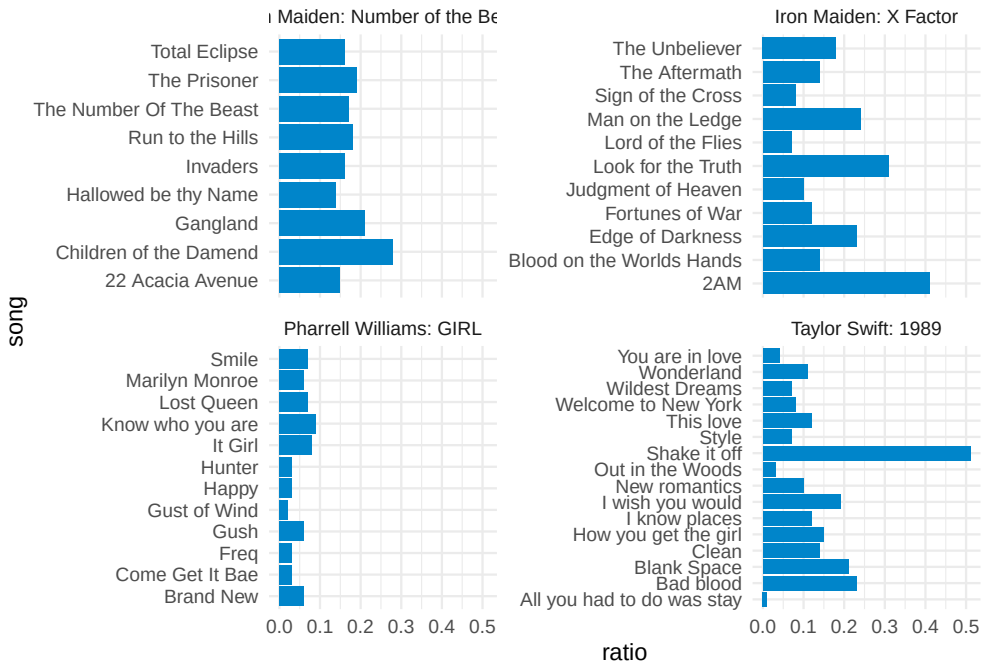
Pharrell, appreciate positivism with woo (part of woo-hoo) and in general talks about happiness. Moan comes from sex so huhh.

Taylor we can see that likes to talk about love - good or bad? and shake is that from shake what? can be from shake the body soo...

Sentiment of each song

A next obvious quesitno is beyond words. How does each song contribute to the feelings shows above. For that we calculate a negative ratio of each song which is basically # of negative words over total words so a higher number means a negative song. results are this.

```
## `summarise()` has grouped output by 'artist_album'. You can override using the
## `.groups` argument.
## Joining with `by = join_by(word)`
## `summarise()` has grouped output by 'artist_album'. You can override using the
## `.groups` argument.
```



Makes sense for Maiden cause Chilre of damned talks about homeless kids and 2am talks about contemplating suicidine.

Not surpriseing Williams is an optimistic fellow.

And Taylow, Shake it off, is misleading cause it's not a negative song. it's more about somebody that likes to go out and party and "shakes off" criticisim from others so could be positive. example fo limitation. next is bad blood, which is about a failed relatioship that went from love to hate so it makes sense.

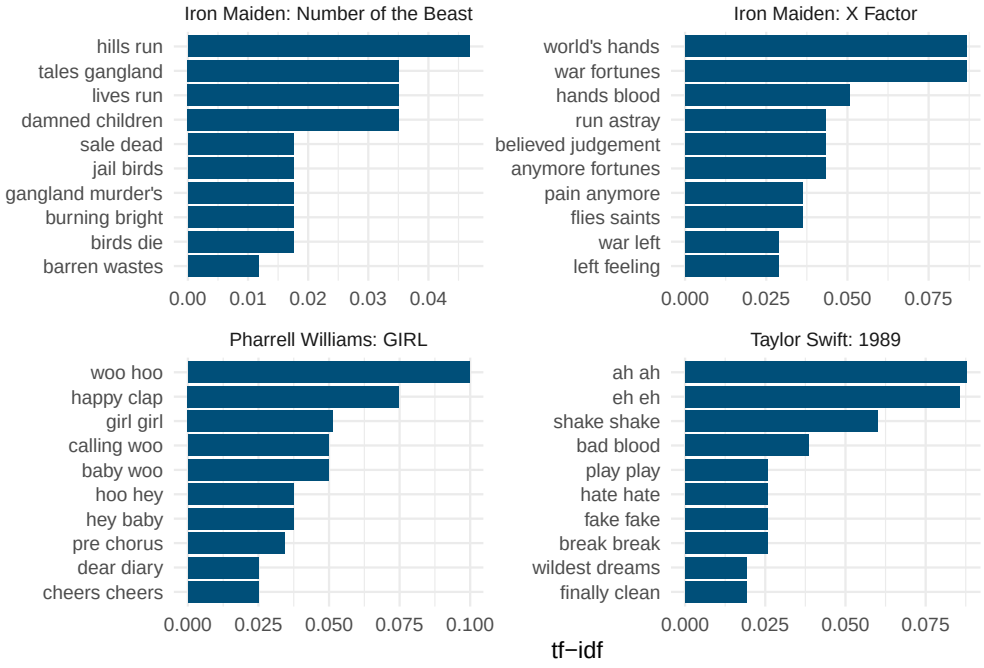
Looking beyond one word

As seen, one word is limited so here we look at combinations of two words that give more context. So what are the top 10 combinations of pairs of words per album?

```
## # A tibble: 7,141 x 2
##    bigram        n
##    <chr>     <int>
##  1 i i          99
##  2 of the       90
##  3 in the       89
##  4 are we       68
##  5 out of       61
##  6 i wish       59
##  7 and i        55
##  8 this love    52
##  9 oh oh        50
## 10 yet are      43
## # i 7,131 more rows

## # A tibble: 751 x 3
##    word1   word2     n
```

```
##    <chr>   <chr> <int>
##  1 ah      ah       41
##  2 girl    girl     41
##  3 eh      eh       40
##  4 woo     hoo      32
##  5 shake   shake    28
##  6 happy   clap     24
##  7 bad     blood    18
##  8 baby    woo      16
##  9 calling woo      16
## 10 break   break    12
## # i 741 more rows
```



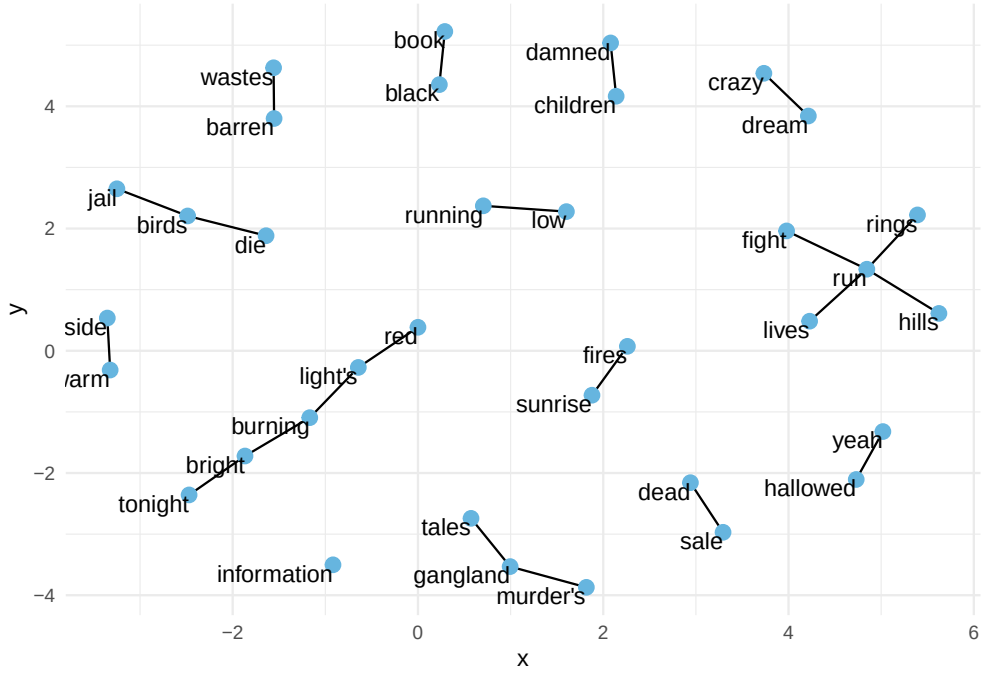
This gives more insight. for example for xfactor above fortune was positive but it's really talking about war fortunes which is an irony for all the bad stuff that happends after the wars.

Interesting to see how Williams and Pharrell don't use real words and more "ah ah", "eh eh", "woo hoo".

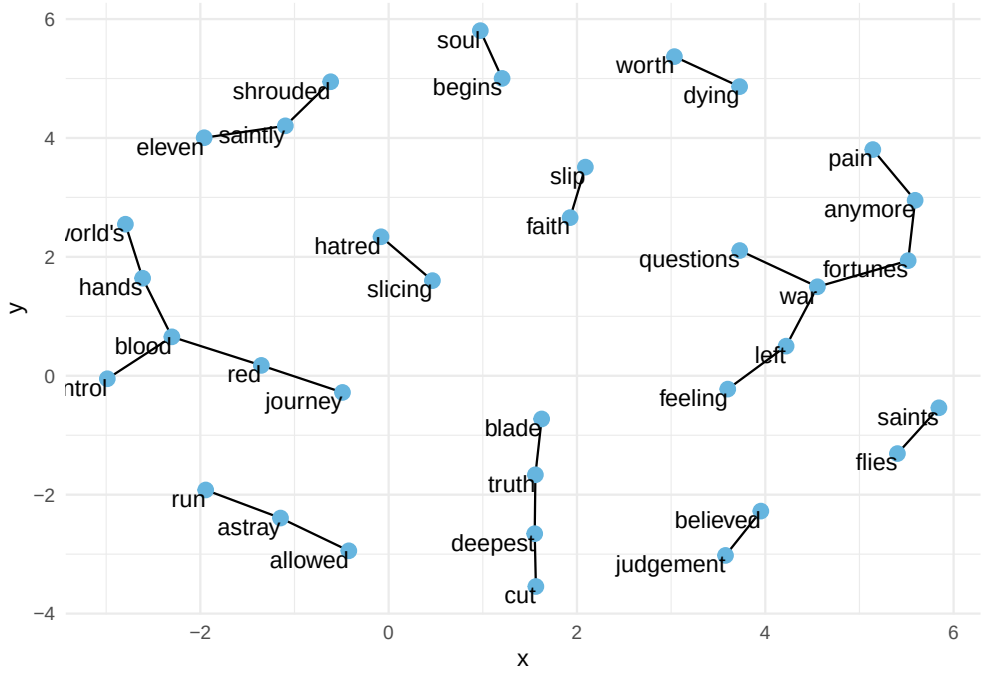
Network graphs of these bigrams

naturally we want to see how these words are interconnected to see about topics they talk about.

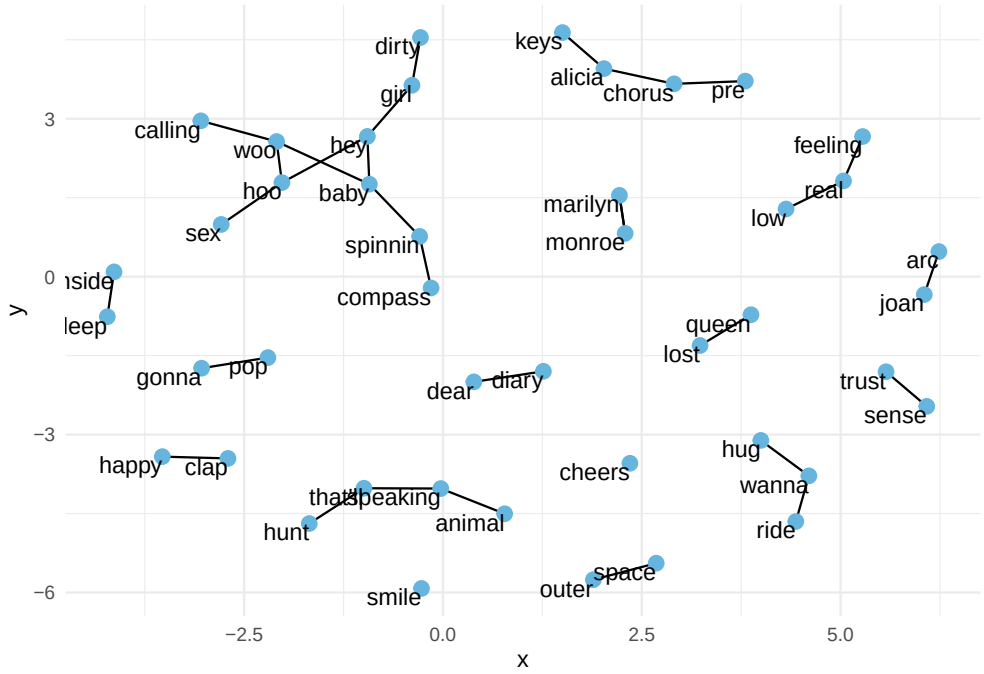
```
plot_graph(albums_to_parse[1], 1)
```



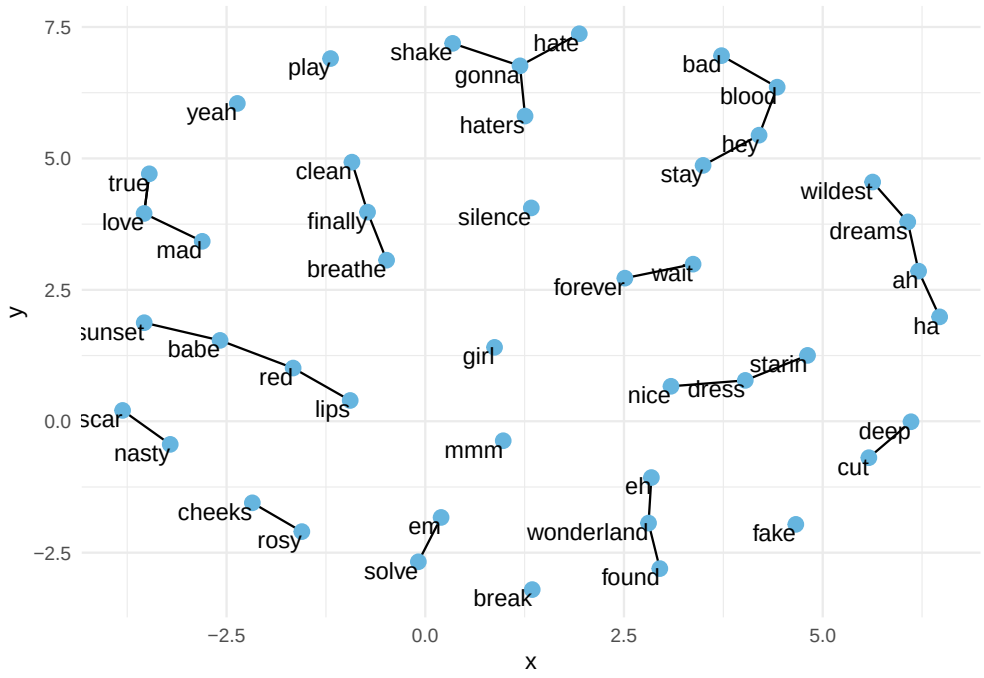

```
plot_graph(albums_to_parse[2], 1)
```



```
plot_graph(albums_to_parse[3], 2)
```




```
plot_graph(albums_to_parse[4], 3)
```



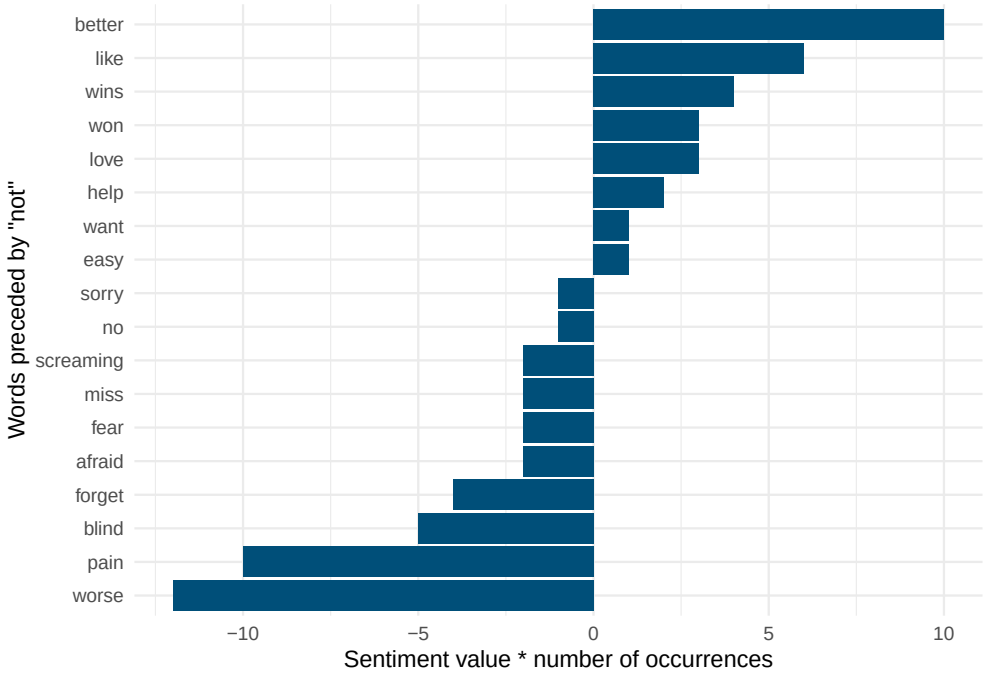
taylor, she talks about all she;s gonna do as is shake. but more context is needed cause paople are gonna hate her. another interesting, is she's asking about stay with someone with bad blood.

for williams we can see girl connected to dirty gilr and hey girl and hey baby so sounds about happy love. he also wants to hug.

Maiden beast, see how more context is given. run is connected to varios words and when reading the lyrics one can see that its about running to the hills (escaping), and run for your lives (escaping).

Negated words

Another importantthing to consider is that happy when doing one word is treated as positive so let's say "i am happy" and "i am not happy" will read happy and that is that. so that is why combining two words give context. so it's intersting to see what words were kind of misclassified with negated words before. so i see what words were c("not", "no", "never", "without") after thse ones and how much they misled the scores.

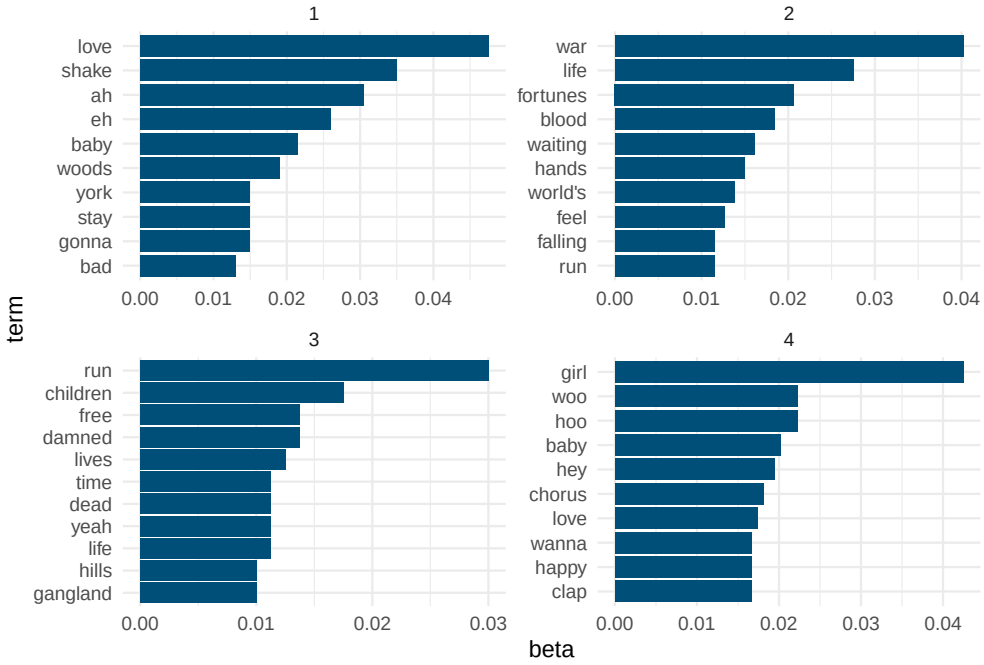


so for example, worse its the most negative here. but how about it say i am not worse, that would kind of be positive.

let's take like, if it's never like then we have a problem.

Topic modeling

Finally imagine we didn't have the albums. we just knew songs. well there is a way to group them by "topic". so I ovbiously tune the algo to 4 topics.



if I compre this charts with above it would be easy to identify the 4 albums with these 4 topics.

Conclusions

1. one word analysis can be misleading

References

Appendix