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

Walter Guillioli

April 04, 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.

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:

Table X
Number of songs and words per album

Album	# Songs	# Words
Iron Maiden: Number of the Beast	9	800
Iron Maiden: X Factor	11	871
Pharrell Williams: GIRL	12	1437
Taylor Swift: 1989	16	2000

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

---

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

Table X
Songs per album

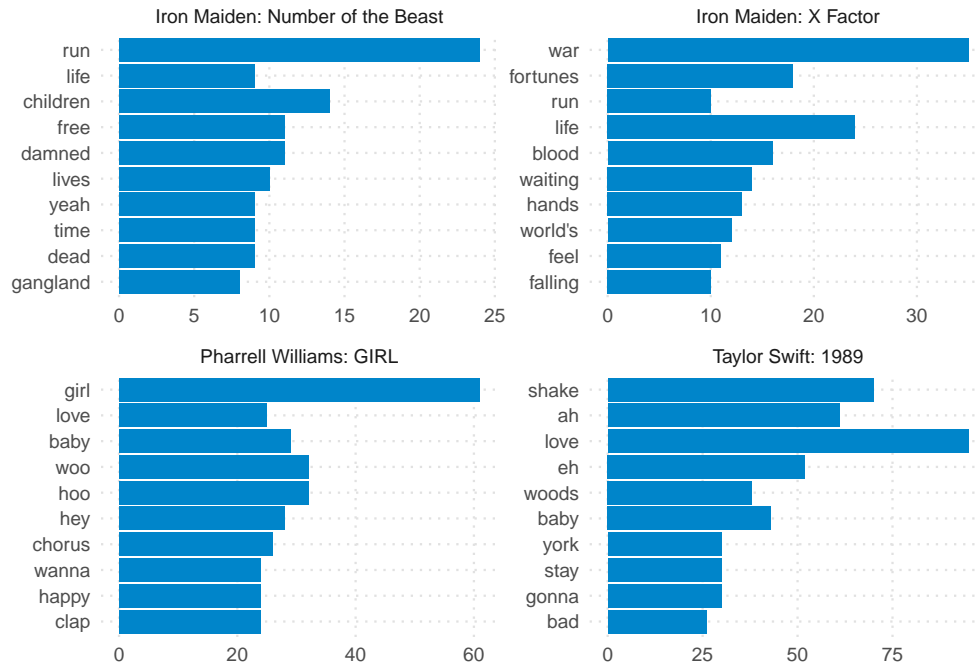| Taylor Swift: 1989	Pharrell Williams: GIRL	Iron Maiden: Sign of the Cross	Iron Maiden: Number of the Beast
Welcome to NY	Marilyn Monroe	Sign of the Cross	Invaders
Blank Space	Brand New	Lord of the Flies	Children Of the Damned
Style	unter	Man on the Ledge	The Prisoner
Out in the Woods	Gush	Fortunes of War	22 Acacia Avenue
All you had to...	Happy	Look for the Truth	Number Of The Beast
Shake it off	Come Get It Bae	The Aftermath	Run to the Hills
I wish you would	Gust of Wind	Judgment of Heaven	Gangland
Bad blood	Lost Queen	Blood on the Worlds Hands	Hallowed be thy Name
Wildest Dreams	Freq	Edge of Darkness	Total Eclipse
How you get the girl	Know who you are	2AM	
This love	It Girl	The Unbeliever	
I know places	Smile		
Clean			
Wonderland			
You arein love			
New romantics			

Results

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

Top words used on the lyrics

We simply remove stop words and count and get this. Alternative we could use a more advanced way via tfidf to appendix
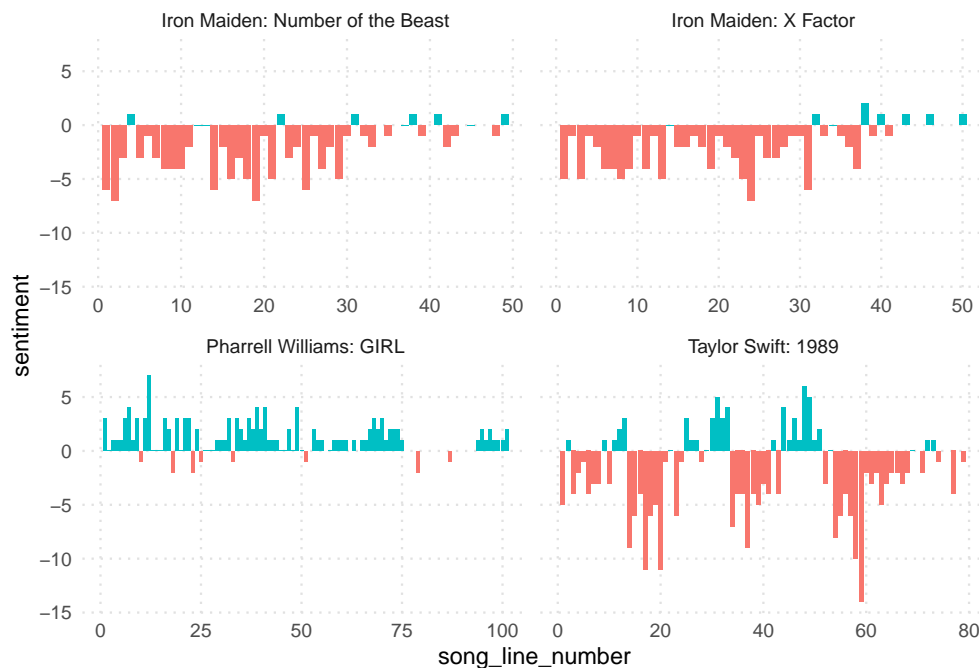
This is relatively interesting but not useful since we know some artists use more words than others. So let's do the same but by album.

Sentiment evolution in the album

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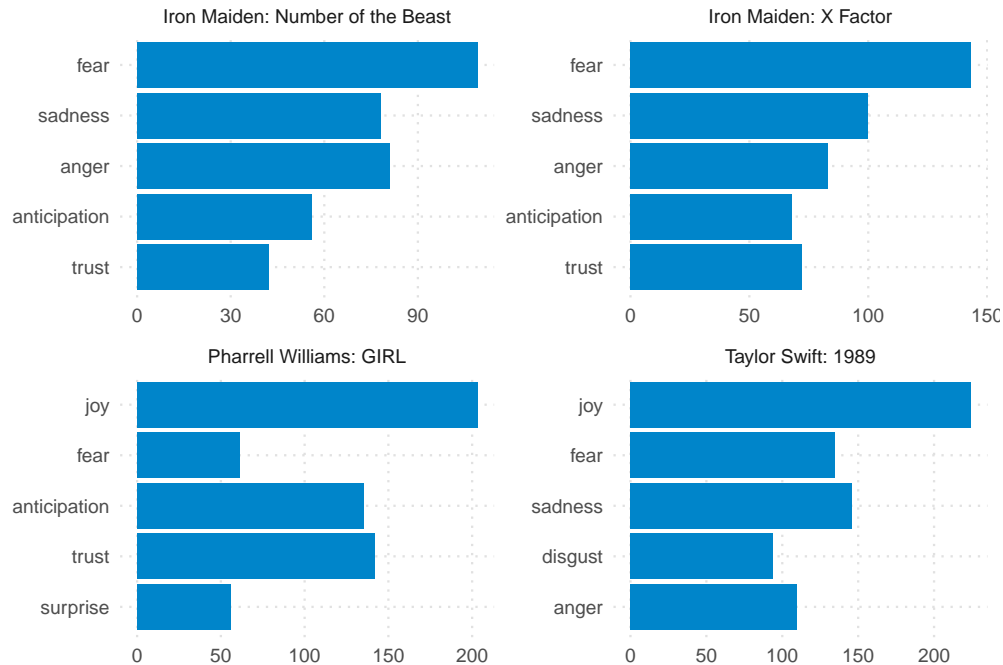
We start with a simple analysis to see if the sentiment is positive or negative. In the chart below we see the sentiment evolution of the albums as the album progress where each bar represents a line in the song. Also note that number of lines varies and Maiden sings with less words.

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.
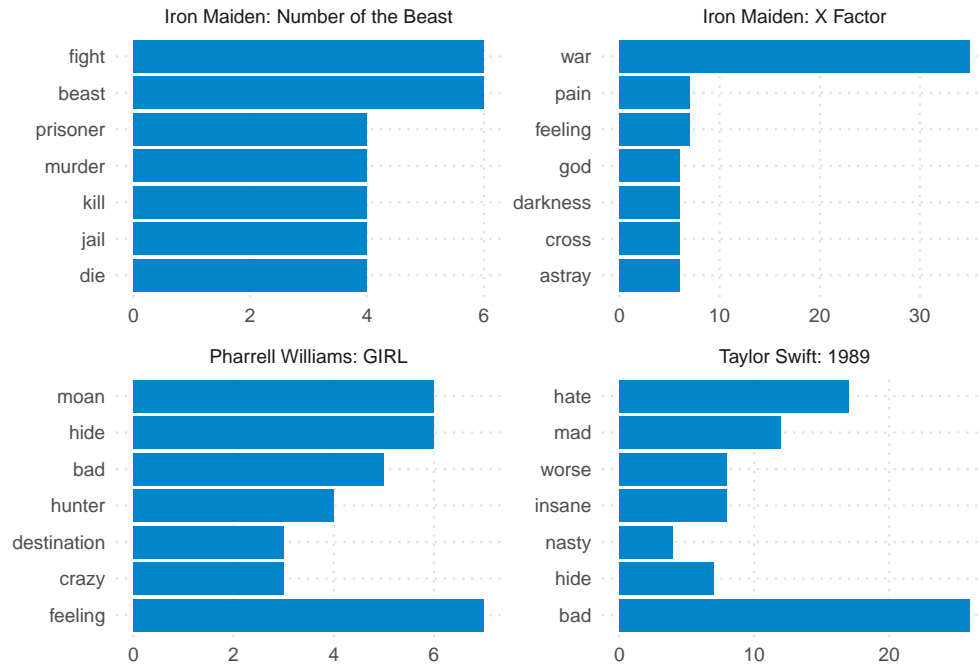
Main feelings from each album

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.

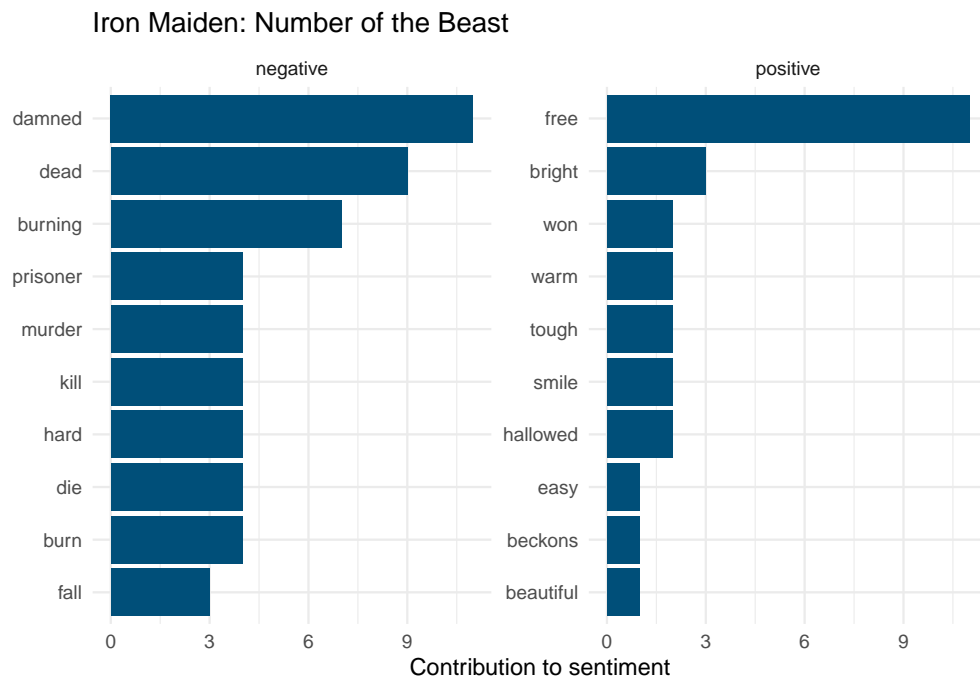**What are they afraid of?**

A common feeling across all albums is the sentiment of fear. So next we explore the top fear words of each album.
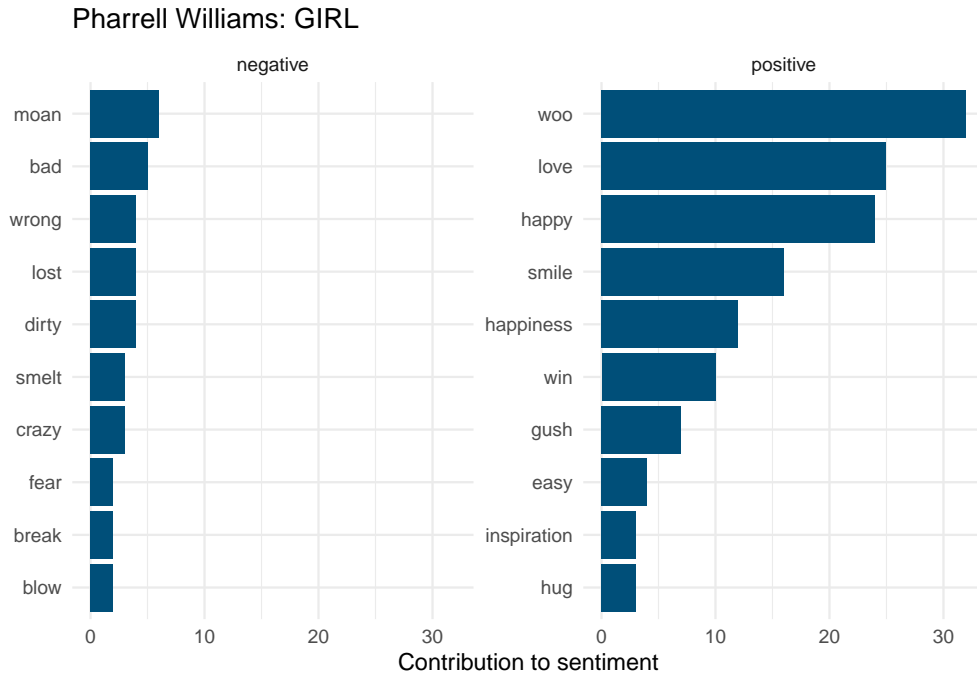
So even though they are all fear words, the meaning is different. TAke Iron Maiden is talking about fight, kill, murder, beast, war. Whille Pharrell talks about his feelings and moan. Even though hunter appears it's a love song about hunting for his girl. Taylor swift with bad, hate and mad clearly is 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.

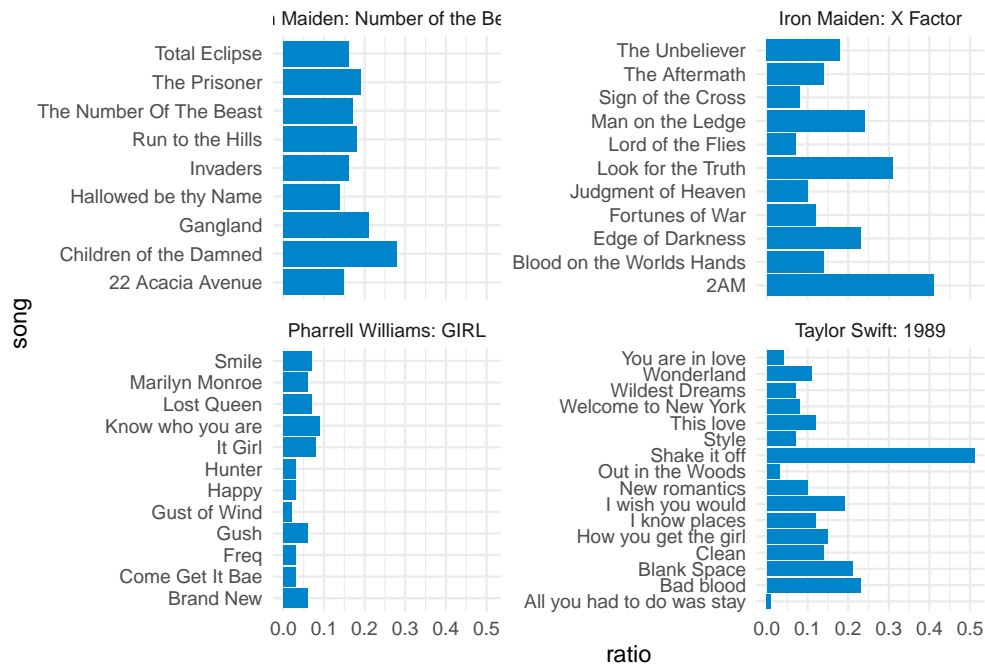Iron Maiden: Number of the Beast

Pharrell Williams: GIRL

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. . .

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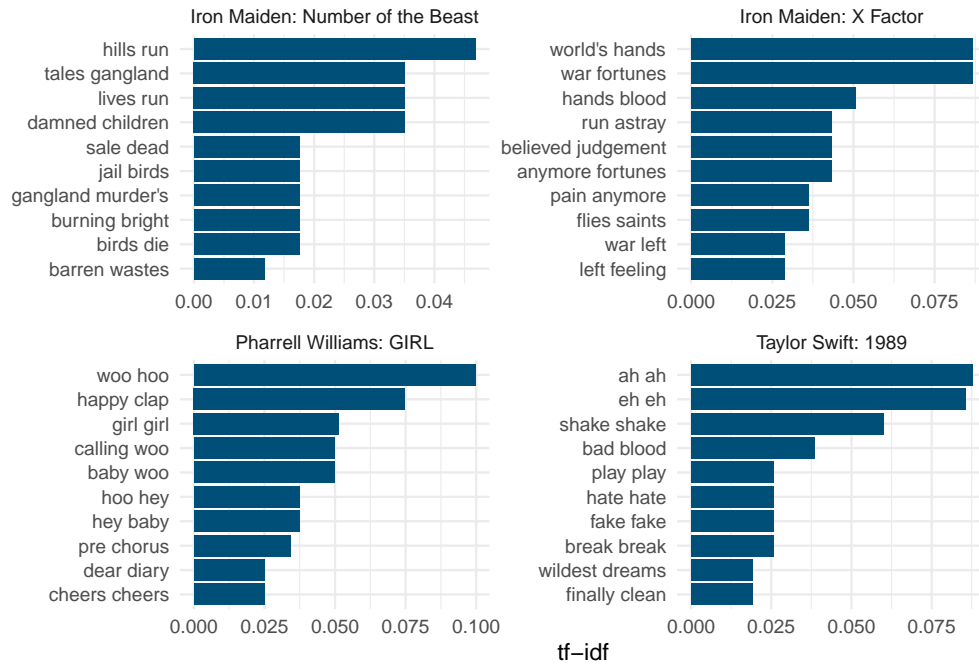
Makes sense for Maiden cause Chilre of damned talks about homeless kids and 2am talks about contemplating suicidine.

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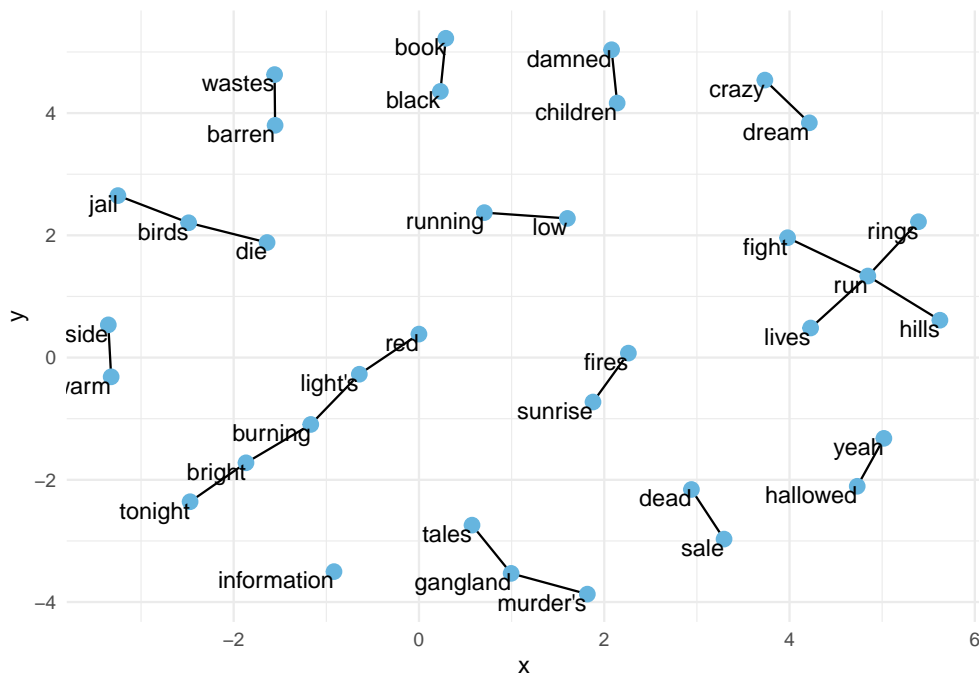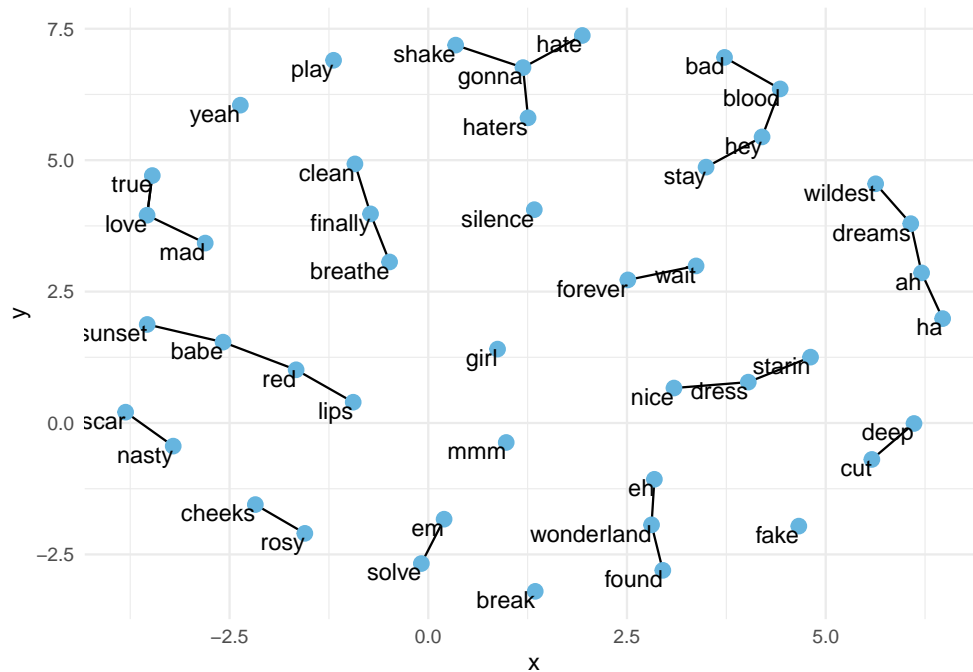
This gives more insight. for example for xfactor above fortune was positive but it's really talking about war fortunes which is an irony for all the bad stuff that happends after the wars.

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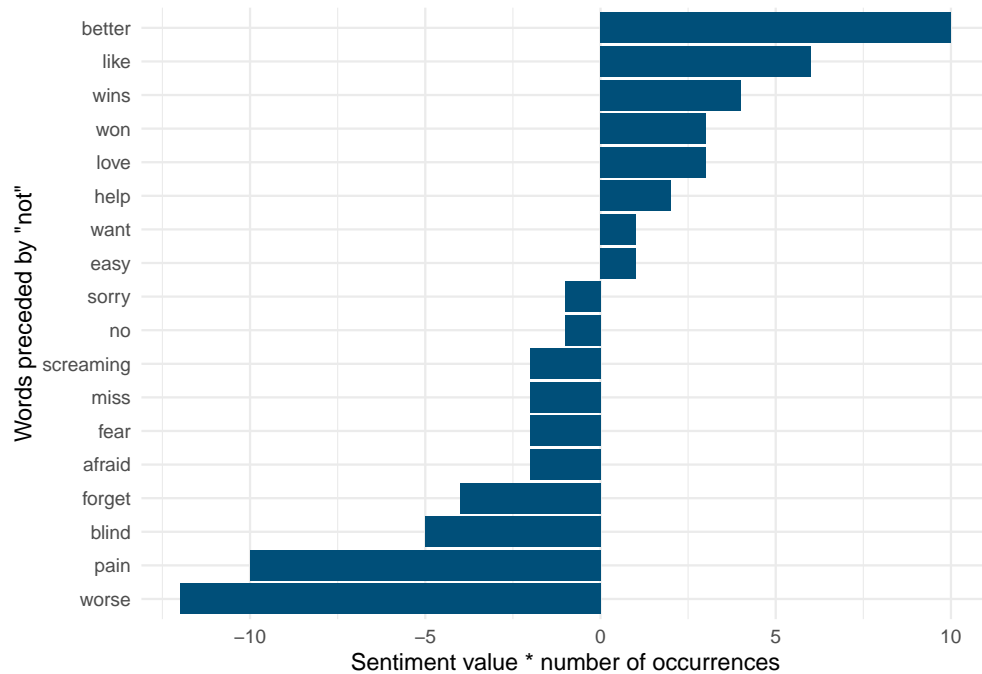
taylor, she talks about all she;s gonna do as is shake. but more context is needed cause paople are gonna hate her. another interesting, is she's asking about stay with someone with bad blood.

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).

Impact of hegated 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.

A note on 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. see appendix
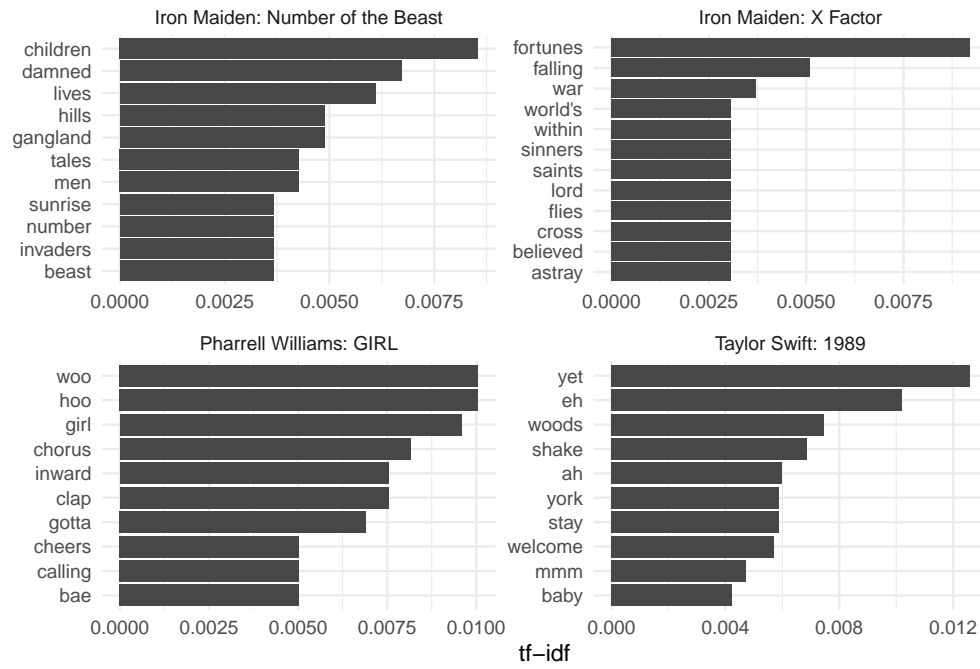
Conclusions
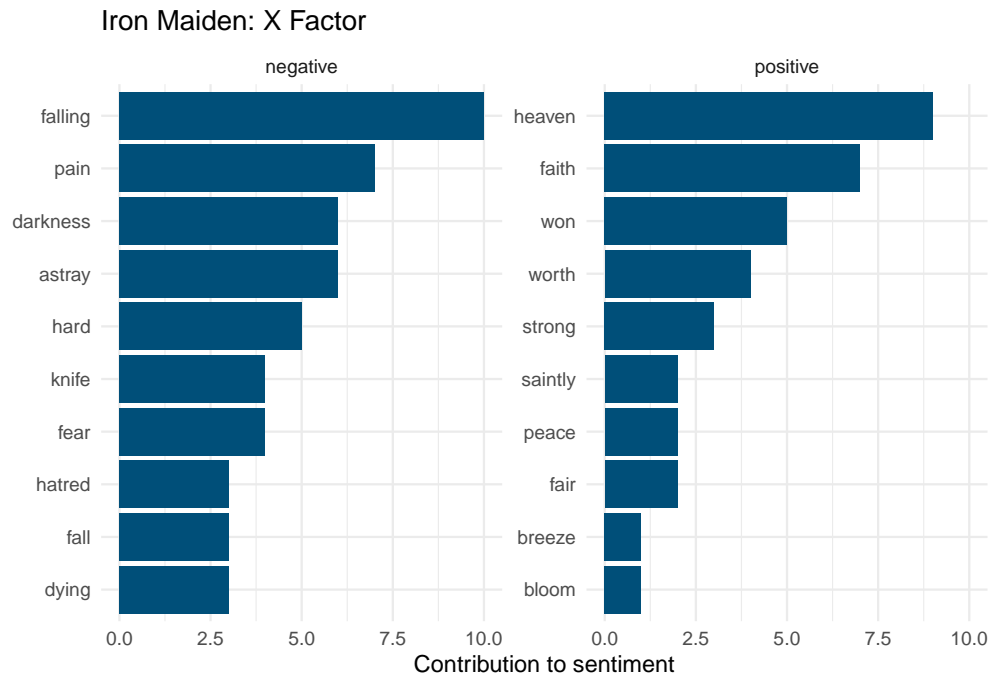
1. one word analysis can be misleading
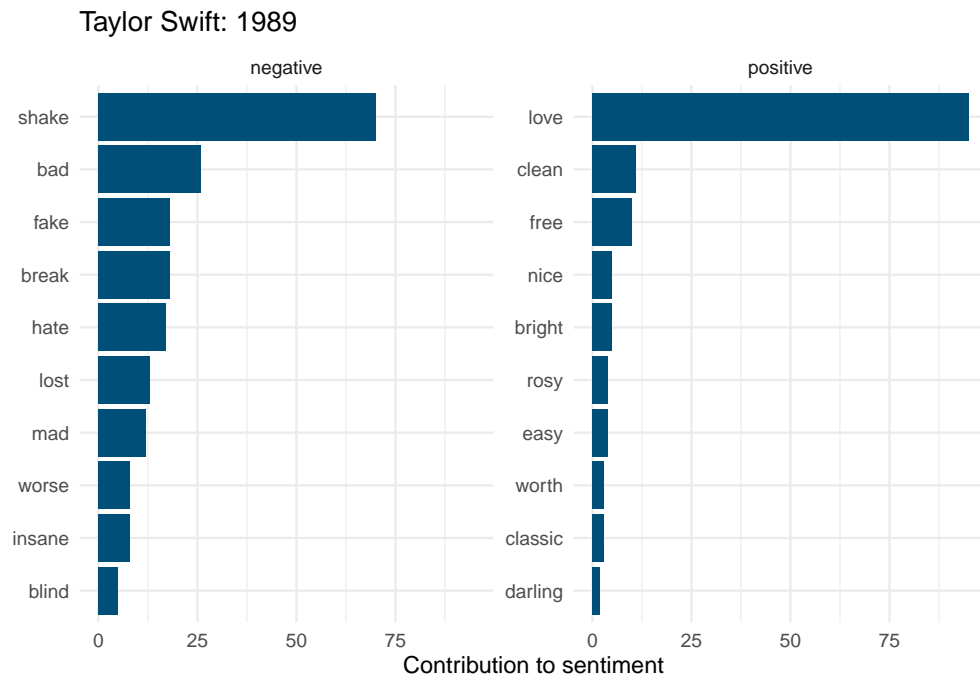
References

Appendix

Appendix 1: TFIDF

Iron Maiden: Number of the Beast

children
damned
lives
hills
gangland
tales
men
sunrise
number
invaders
beast

0.0000   0.0025   0.0050   0.0075

Iron Maiden: X Factor

fortunes
falling
war
world's
within
sinners
saints
lord
flies
cross
believed
astray

0.0000   0.0025   0.0050   0.0075

Pharrell Williams: GIRL

woo
hoo
girl
chorus
inward
clap
gotta
cheers
calling
bae

0.0000 0.0025 0.0050 0.0075 0.0100

Taylor Swift: 1989

yet
eh
woods
shake
ah
york
stay
welcome
mmm
baby

0.000   0.004   0.008   0.012

tf−idf

Appendix 2

```
print_pos_neg_words(2)
```

Iron Maiden: X Factor

negative	positive
falling	heaven
pain	faith
darkness	won
astray	worth
hard	strong
knife	saintly
fear	peace
hatred	fair
fall	breeze
dying	bloom

Contribution to sentiment

```
print_pos_neg_words(4)
```

Taylor Swift: 1989

negative	positive
shake	love
bad	clean
fake	free
break	nice
hate	bright
lost	rosy
mad	easy
worse	worth
insane	classic
blind	darling

Contribution to sentiment
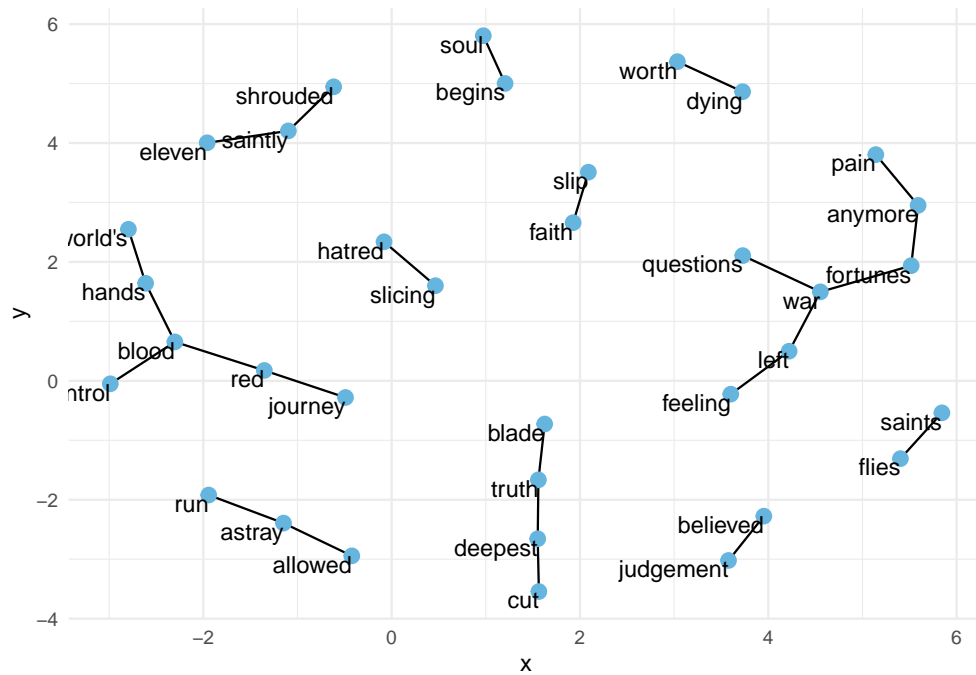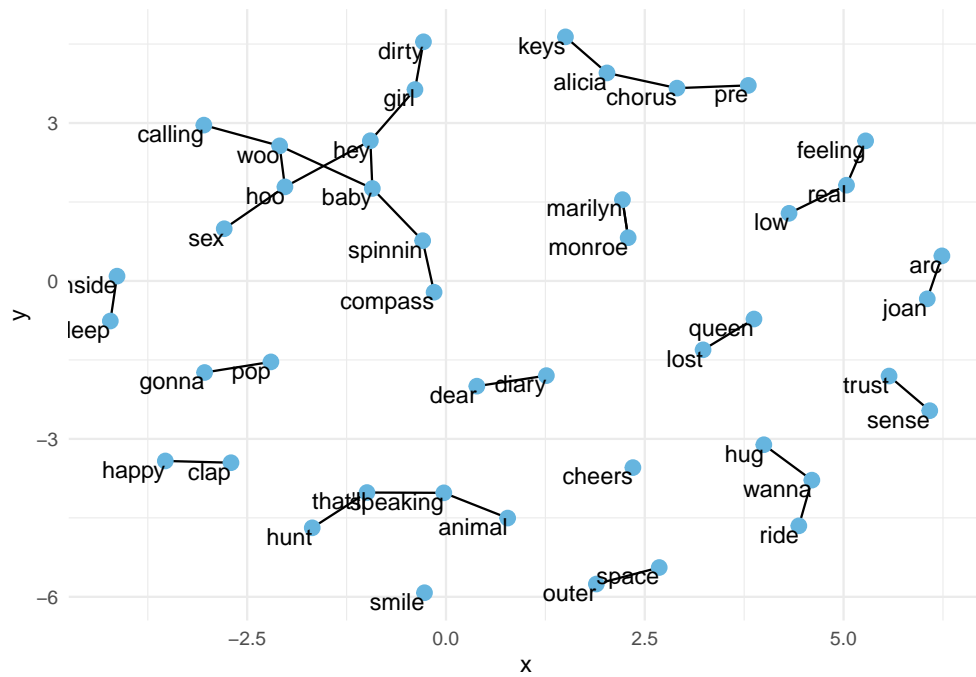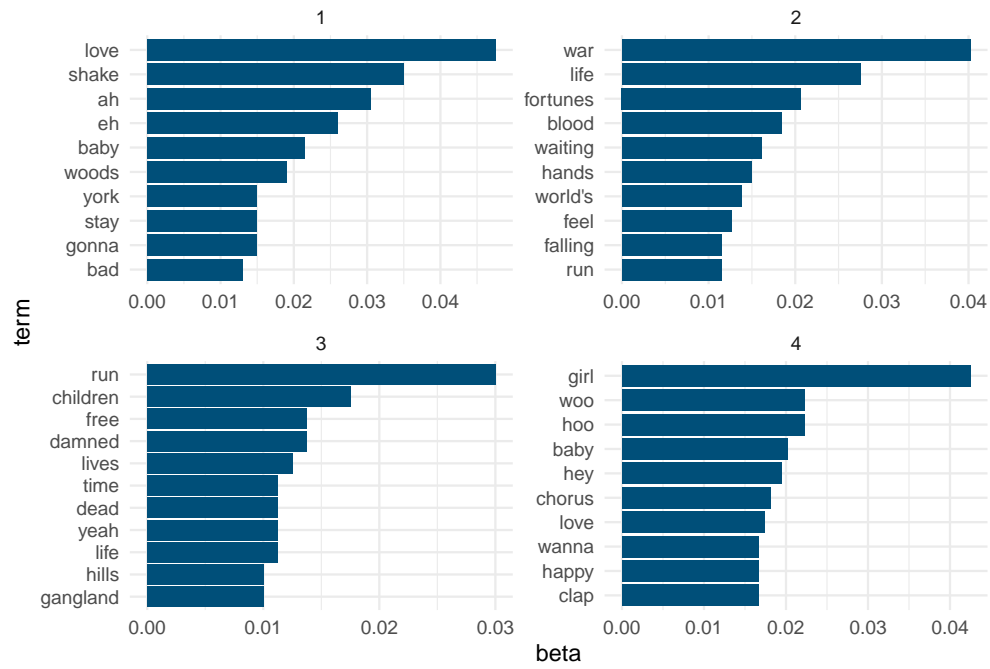
Appendix 3

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



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

Appendix 4

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