

NLP Analysis for Game of Thrones

Mingjie Zhao

Mar. 20, 2020

1. Introduction

Many people around me are fans of Game of Thrones, and last summer I heard a lot of complaints about the show's ending. For whatever reason, my brain has had difficulty memorizing character names, especially when the relationships of characters are complicated. So I gave up watching the show after a few episodes. However, I think it would be very interesting to analyze the words of the characters. The show creates a fantastic world and I am curious to learn about the world through the characters's words (much faster than watching the whole show which includes 73 episodes), and that's the motivation of this project.

2. Methodology

In this project, I picked four characters with the most lines. You may be wondering by now about who are the most talkative ones in this show. Well, the answer will be provided below in the analysis. Since characters may speak around certain topics based on their character settings, I created LDA (Latent Dirichlet Allocation) models of these 4 characters to see if there are main topics from what they said. I also performed sentiment analysis on their words to see how they feel about by living in the fantastic world. Data visualizations are created by multiple plots and some findings are provided at the end of the project.

3. Analysis

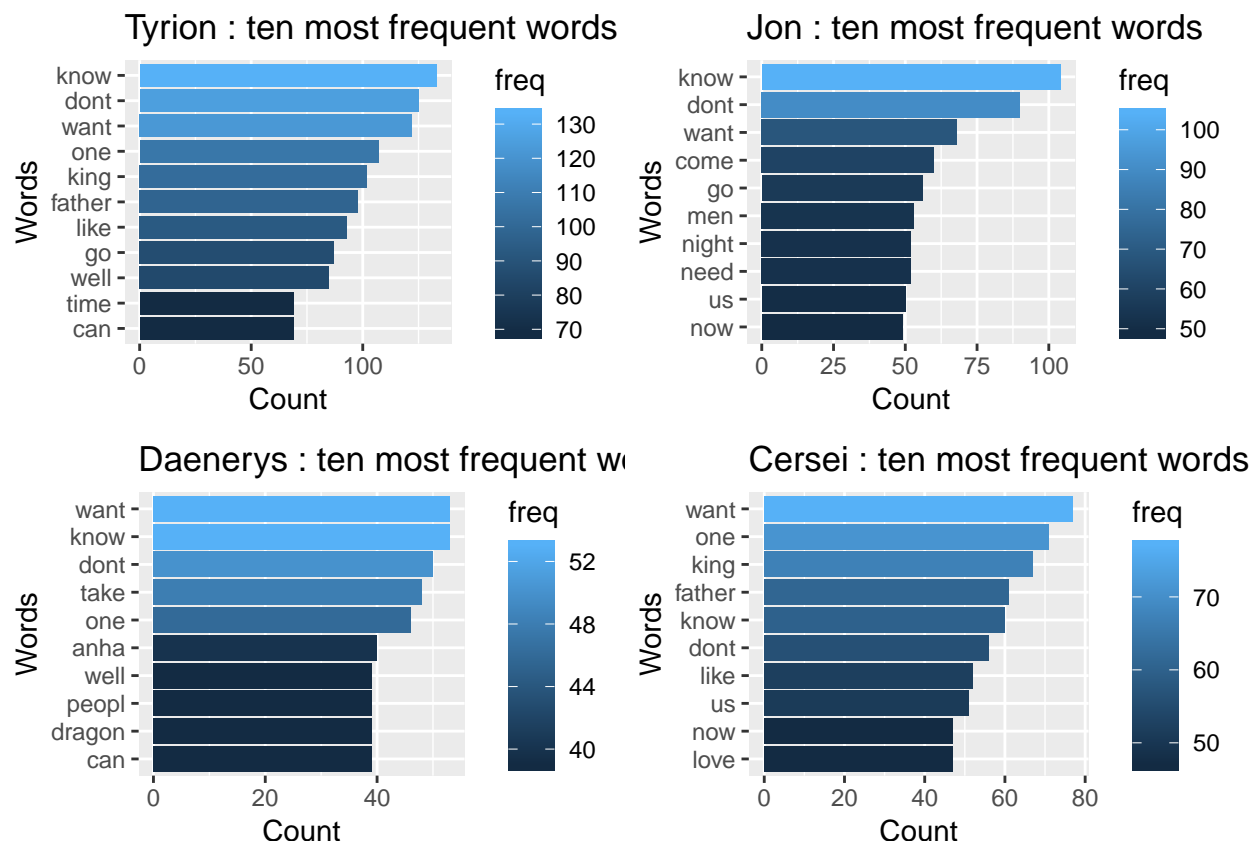
1) Data preprocessing

Firstly, read in (a) dataset which was found on Github (json file provided by <https://github.com/jeffreylancaster/game-of-thrones>). The original dataset is a huge list of list and I manipulated the dataset to get a list of the characters and a new dataframe to count for numbers of sentence for each character. We can see that the average number of sentences in the show across all the characters is 181.5, while the most talkative character said 1704 sentences. However, there are a lot of characters who only said one sentence in the show. And the 4 characters who are on my list of top 4 talkative ones are Tyrion Lannister, Jon Snow, Daenerys Targaryen and Cersei Lannister. I hope this is not a surprising result to you if you are a fan, because what they've said would be very important here in my analysis below.

2) Word frequency analysis

To deal with the text data, I created a function to do data manipulation including removing punctuation, numbers and stopwords. I also did word stemming to remove affixes. Then I created a chart to show which words are used the most by these characters. It's interesting to see that these top words are highly correlated with each character's background. For example, Tyrion used "king" and "father" a lot, because he spent a lot of court business dealing

with nobles. Also Tyrion lived in the shadow of his father Tywin Lannister for a long time, until he killed his own father. Daenerys liked to say “dragon” and “people”, since she is the mother of three dragons, and is considered to be the “Mother of dragons”. And Cersei was the ruling queen of the seven kingdoms, until losing it to Dany. She married King Robert, and is mother of two kings. That may explain her favorite words like “king”, “father” and “love”.



3) LDA and topic model

Latent Dirichlet Allocation (LDA) is a widely used model in natural language processing (NLP) as an unsupervised learning method. In this project, I am trying to find 3 topics for each character. A function was written to perform the LDA modeling and a chart was plotted for the top 4 words in the 3 topics, to visualize the results and understand the topics that were extracted from these words.

It’s interesting to see that some of the topics correlate closely with the show. For example, the third topic of Jon includes key words like “wall”, “night” and “watch”. This probably correlates with the story that Jon has served in the Night Watch, first as a personal steward of the then Lord Commander, and then became a Lord Commander of the Night Watch himself. The main duty of the Night Watch is to guard the Wall. Also, keywords in Topic 3 include other languages spoken by Dany, as she is ruling many people of different cultures and races. Another good example is the Topic 2 of Cersei include keywords like “love”, “brother”, and “daughter”, while Cersei has a relationship and affair with her twin brother Jaime Lannister.

```
## [1] "Tyrion term"
##      Topic 1 Topic 2 Topic 3
```

```

## [1,] "want"  "yes"   "dont"
## [2,] "can"   "father" "king"
## [3,] "man"   "like"   "know"
## [4,] "good"  "right"  "go"

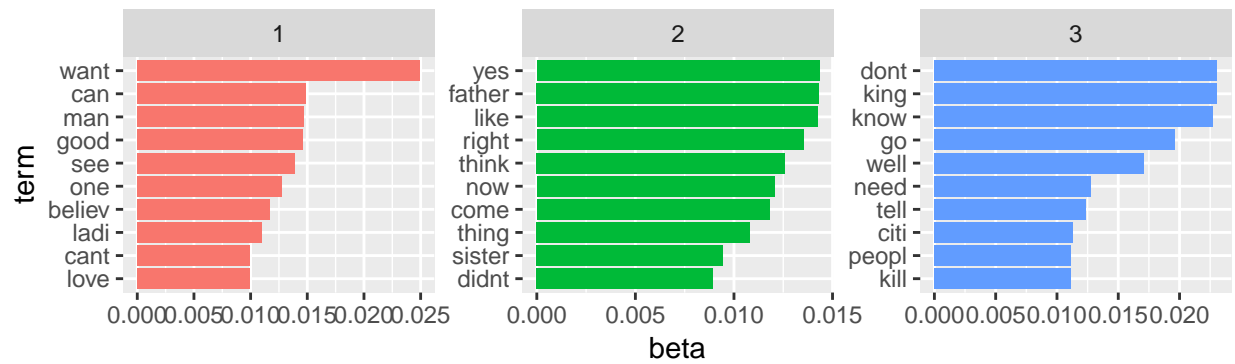
## [1] "Jon term"
##      Topic 1 Topic 2 Topic 3
## [1,] "dont"  "know"   "go"
## [2,] "want"  "one"    "men"
## [3,] "need"  "father" "come"
## [4,] "king"  "right"  "wall"

## [1] "Daenerys term"
##      Topic 1 Topic 2 Topic 3
## [1,] "know"  "can"    "anha"
## [2,] "dont"  "one"    "ma"
## [3,] "want"  "kill"   "vos"
## [4,] "take"  "fight"  "yeri"

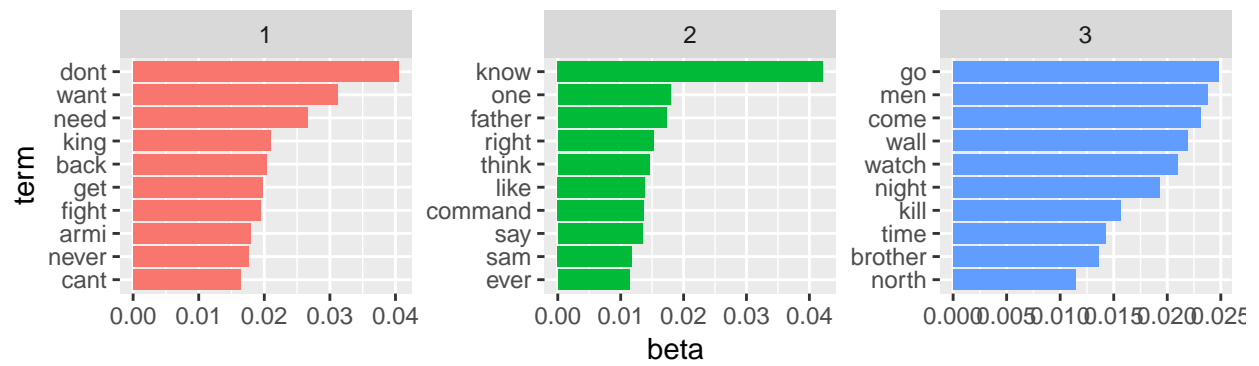
## [1] "Cersei term"
##      Topic 1 Topic 2 Topic 3
## [1,] "father" "one"     "king"
## [2,] "us"      "love"    "want"
## [3,] "hes"     "brother" "good"
## [4,] "like"    "know"    "dont"

```

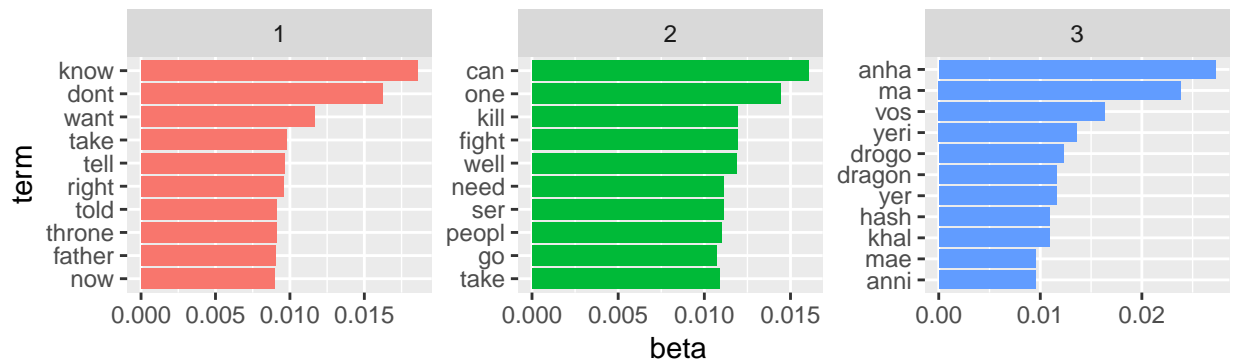
Tyrion : Top terms in each LDA topic



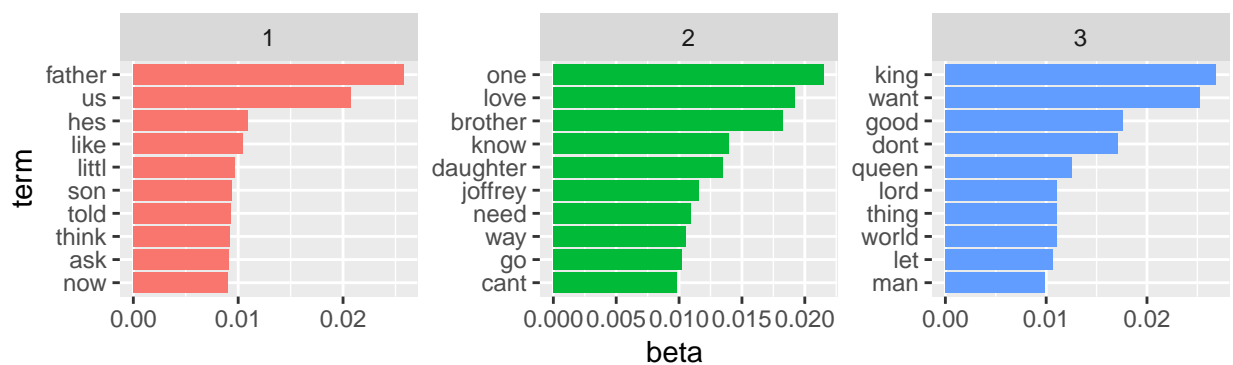
Jon : Top terms in each LDA topic



Daenerys : Top terms in each LDA topic



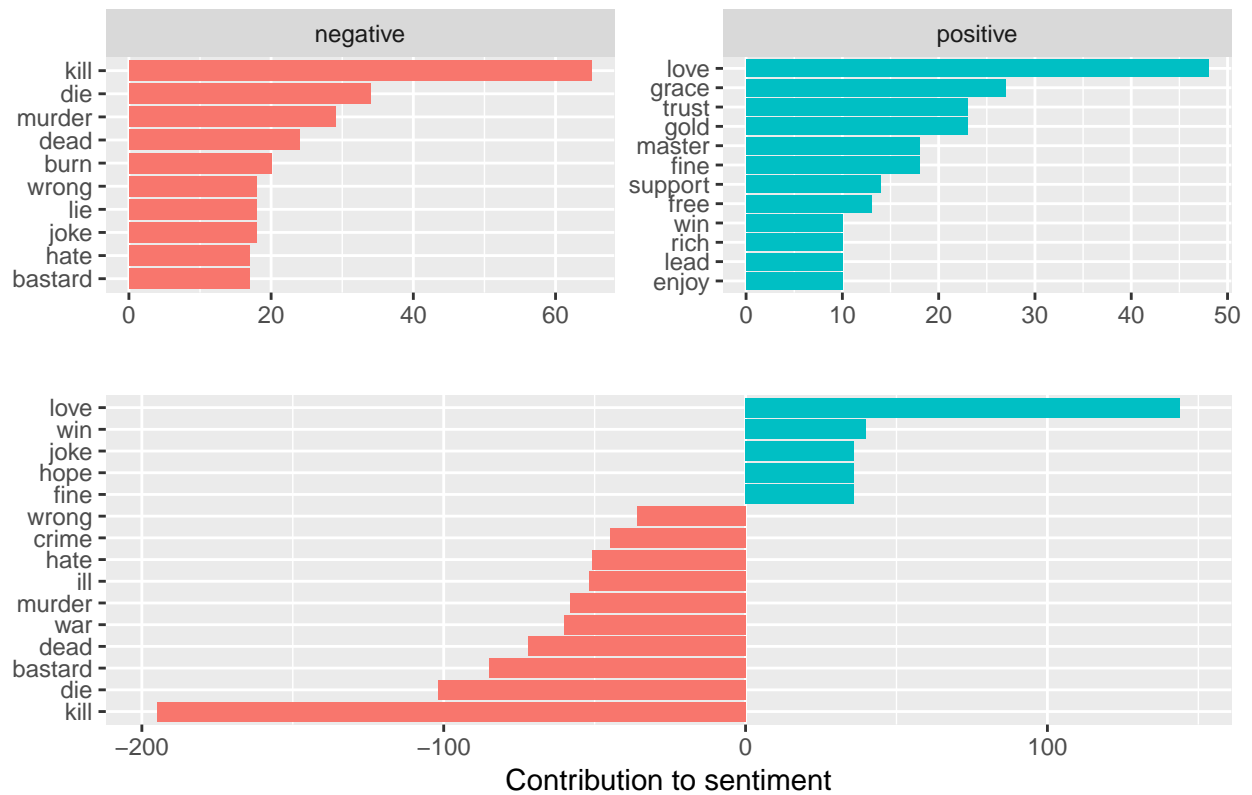
Cersei : Top terms in each LDA topic



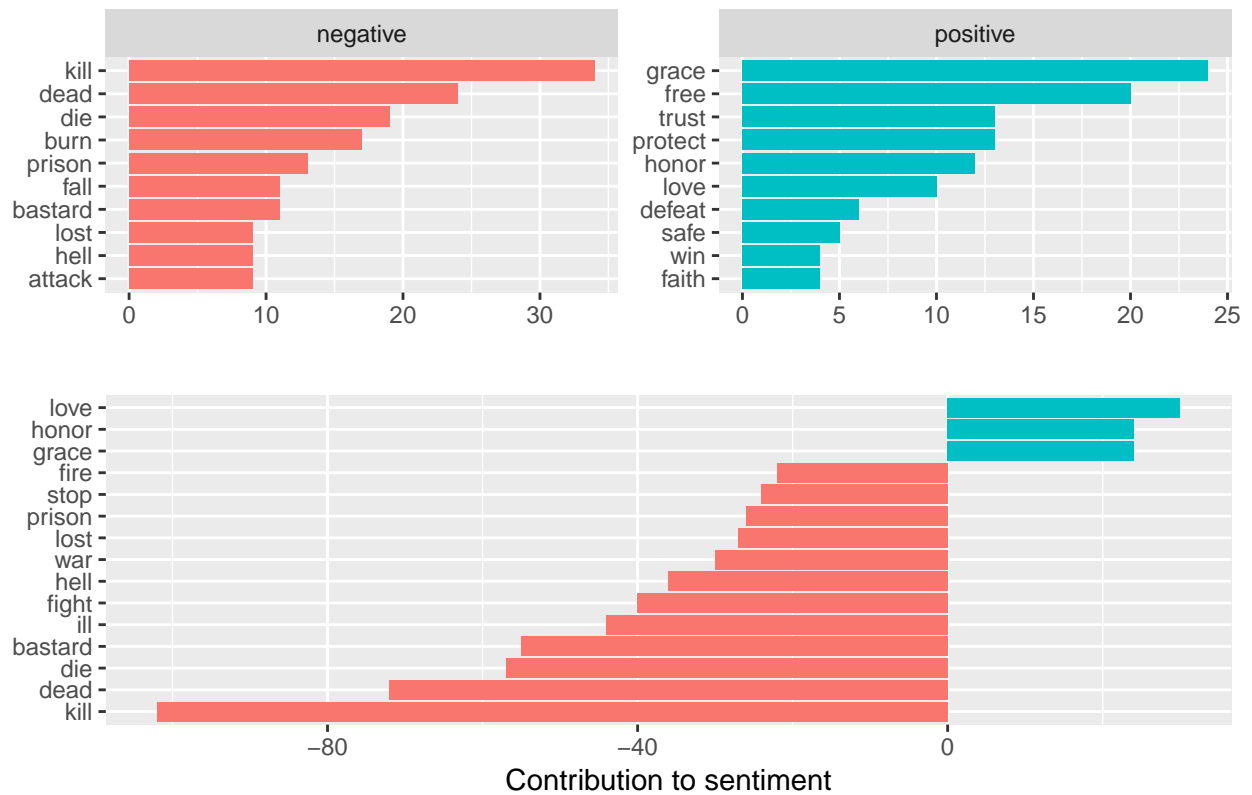
4) Sentiment analysis

It is also insightful to perform sentiment analysis to check emotions within the lines. Two plots were made for each character. The first plot shows the top words contributing the most to positive and negative emotions, and the second plot shows the top 15 emotional words that contribute to sentiment. It is interesting to see that the word “love” is the top positive word appeared in the analysis results for all 4 characters. Additionally, it is worth noticing that in the second plot (Contribution to sentiment) for all four main characters, there are more negative words contributing to sentiment than positive words, indicating a negative attitude in lines for the four characters.

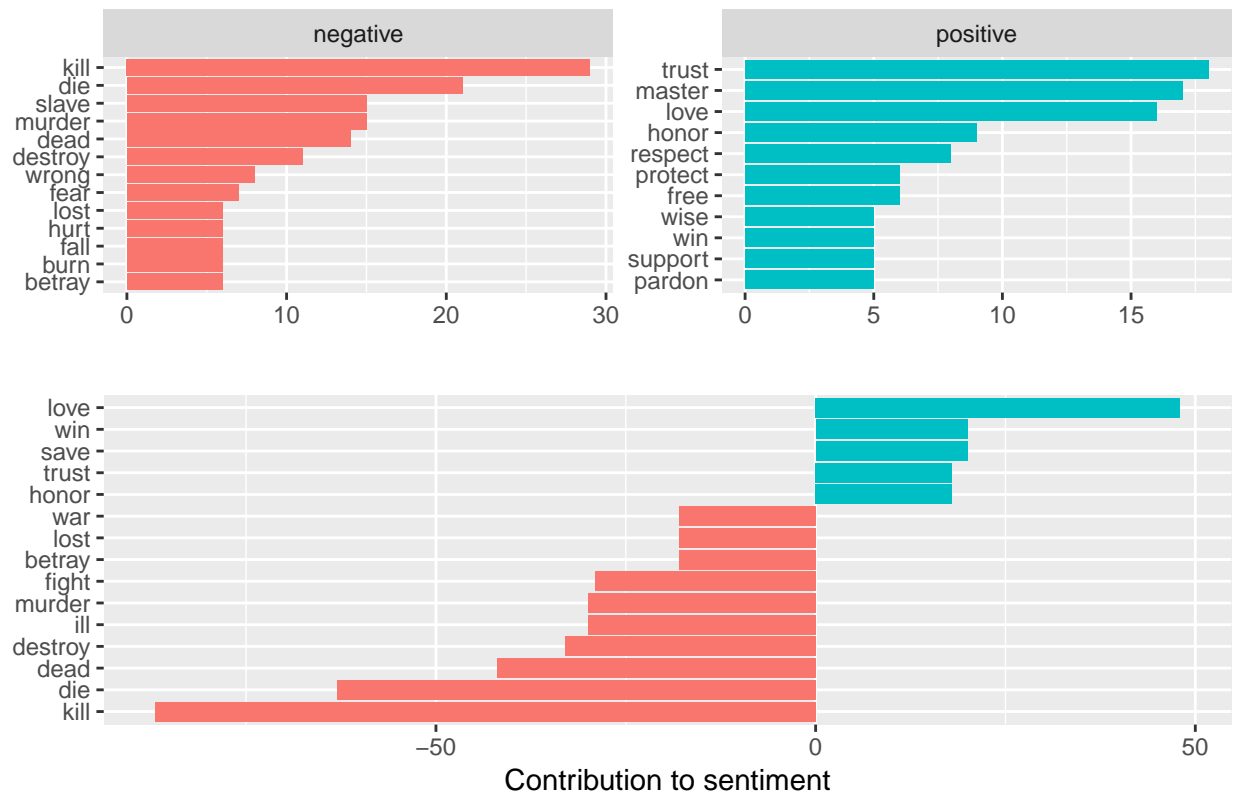
Tyrion : words contributions



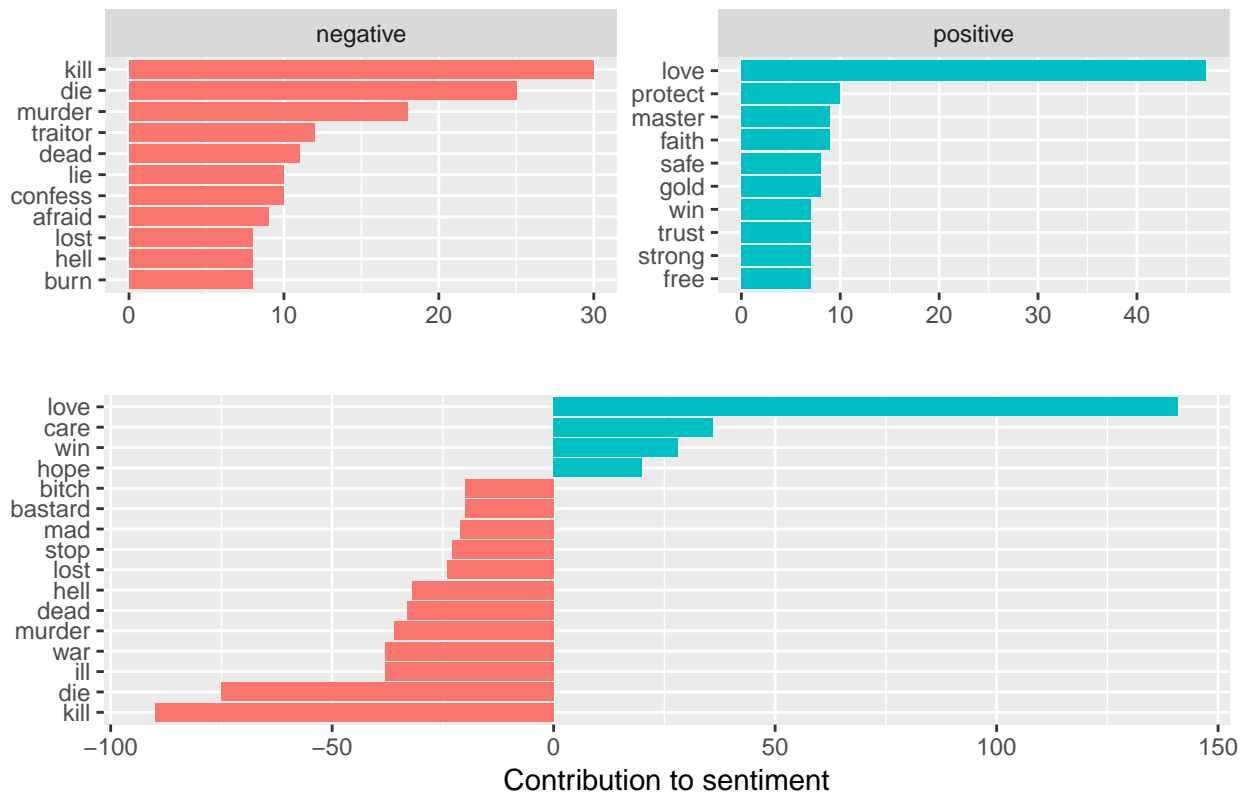
Jon : words contributions



Daenerys : words contributions



Cersei : words contributions



5) Word cloud

Last but not least, I created a word cloud for each character to visualize the emotional words in lines. We can see that on the negative side, “kill” and “die” are really eyes-catching, while “love” and “grace” pop up in the positive side.



Jon Snow

negative



positive

Daenerys Targaryen

negative



positive

Cersei Lannister

negative



positive

3. Conclusions

In this project, I used 5 ways to analyze the lines from Game of Thrones with NLP techniques. As a summary the LDA and topic model do a good job in terms of representing the stories of each character. The sentiment analyses provide insights about the emotions of these characters. It is interesting to see that most of the results correspond to the background of the chatacters. This show mostly provides the audience with negative emotions, but I am happy to see that love is the most used positive word.

Reference

Julia Silge and David Robinson, “Text Mining with R”