Heart of Darkness - Sentiment Analysis (Task 2)

Zara Waheed

Introduction

This report is a text and sentiment analysis of the novel Heart of Darkness by Joseph Conrad. Different lexicons are used to see the overall sentiment of the book.
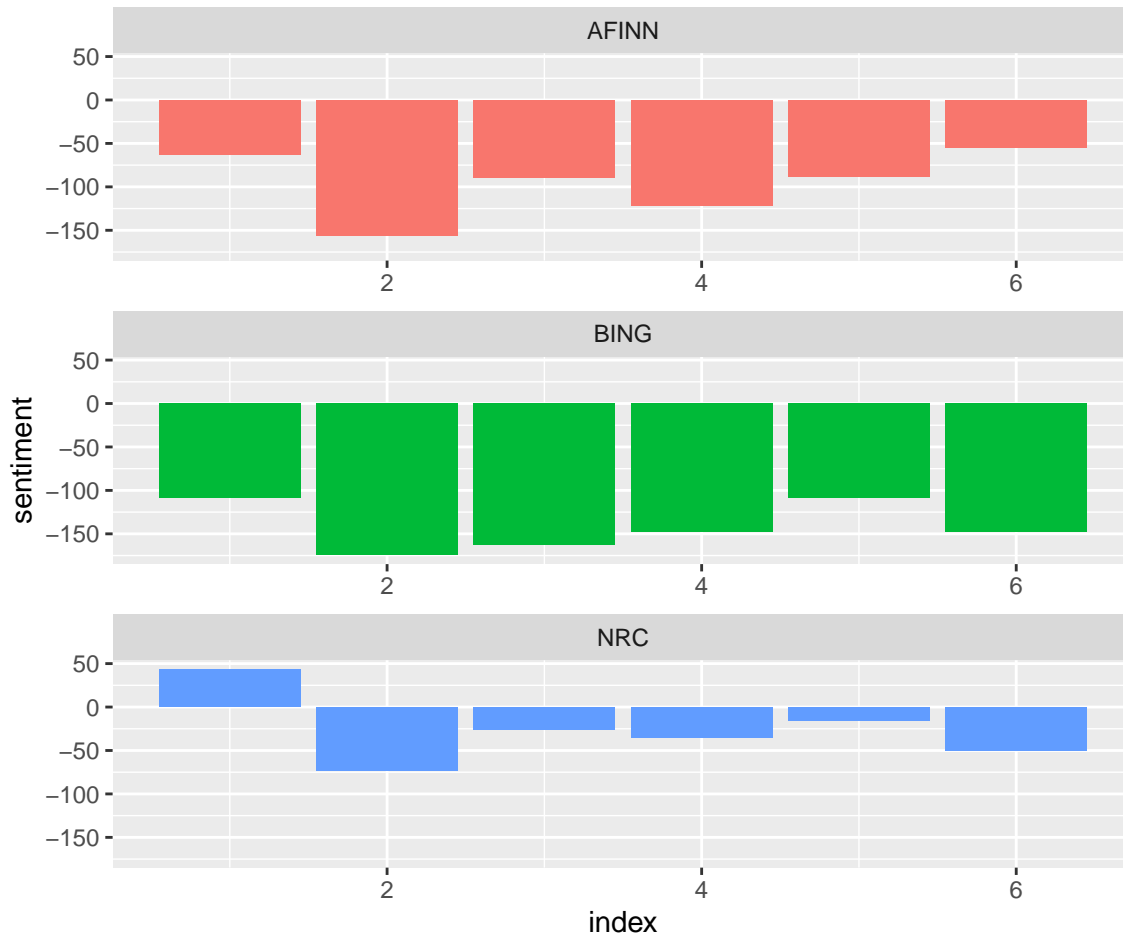
Initial steps

First, I did some cleaning of the text file of the book in order to make easier to work with. After reading in the book from Gutenberger, I manually removed the table of contents, along with any stop words throughout. I read it into RStudio and used
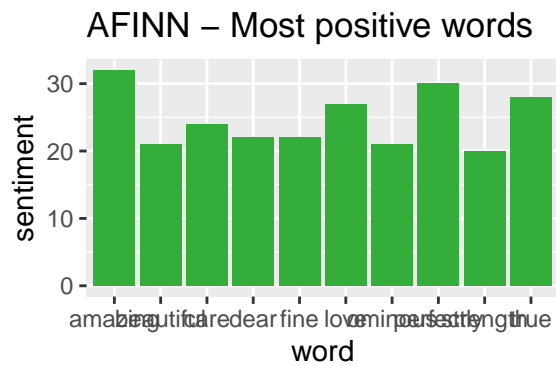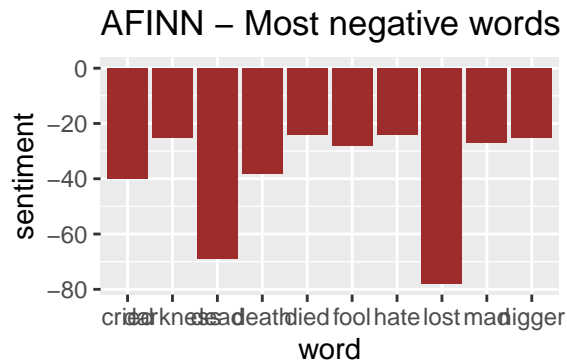
Sentiment Analysis

I explored three lexicons when I did a sentiment analysis for Heart of Darkness. Three different lexicons were used namely, "Afinn", "Bing", and "Nrc". Afinn contains sentiment scores for words, Bing has positive negative categories and NRC has positive, negative and many other categories as well. I created 3 separate data frames for the three lexicons and bound them together. They each divided the sentiments by chapter into scores and positive or negative. From the plots below, we can get an idea of what words were used more

and what the behaviors are in different chapters. There were only 6 chapters in the book so looking at their difference was difficult.
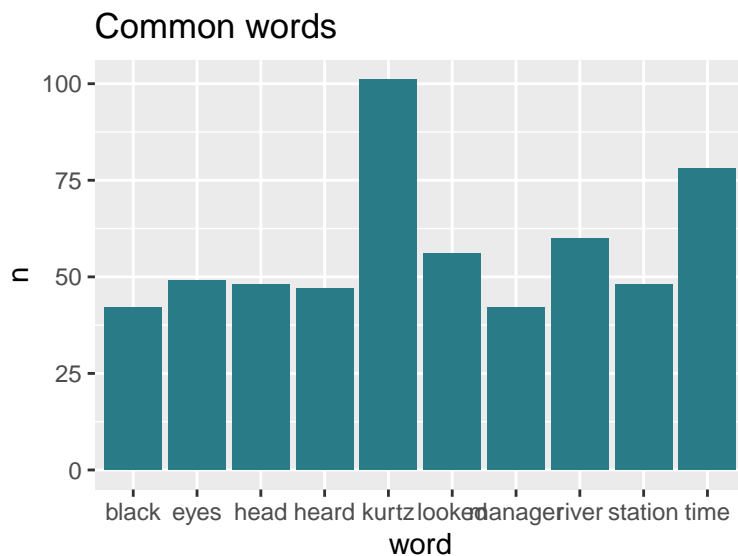

The sentiments of the book were overall very negative since it's about the hardships faced by our protagonist who is travelling to meet a man, reputed to be an idealistic man of great abilities.

The words that contributed the most as well as the least to these negative scores can be seen below by lexicon.

AFINN – Most negative words

AFINN – Most positive words

Words used most often

Just as an interesting thing to look at, the plot below shows the most commonly used words in the book after the cleaning and removal of stop words. This indicates that these two characters are two of the main characters in the book. This is not all too surprising considering the name of the book.

Common words

I also experimented with creating a simple worcloud to show the top 100 words used.

```
## NULL
```

\end