Text Analysis of The Picture of Dorian Gray

Rose Determan

Synopsis and Characters

"The saying 'be careful what you wish for' has arguably never been more apt in literature than it is in this classic novel. When the young Victorian heartthrob Dorian Gray is influenced by Lord Henry Wotton's warning that he only has 'a few years in which to live really, perfectly, and fully' due to the transiency of his youthful beauty, he wishes for his portrait to change with time instead. Little does he know that he will soon stumble down the rocky road of moral corruption, committing one bad deed after another, destroying relationships with the people he meets at the same time as any good reputation he used to possess." *Source: https://www.theguardian.com/childrens-books-site/2014/feb/13/review-picture-of-dorian-gray-oscar-wilde*

Main Characters:
Dorian Gray: Young man highly aware that his youth and beauty are fleeting and is obsessed with his image. He dies at the conclusion of the novel.
Lord Henry Wotton: Exercises influence over Dorian. A bit of a philosopher and revolutionary.
Basil Hallward: Talented painter who creates a portrait of Dorian and is enamored with him. He is killed by Dorian.
Sibyl Vane: Young actress. Dorian's love interest in the novel. She dies by suicide after Dorian breaks her heart.
Source: https://www.sparknotes.com/lit/doriangray/characters/

Bag of Words Analysis

Methods

1. Import text from Project Gutenberg
2. Label rows with chapter ID and row ID
3. **Tokenize** the dataframe. This step takes a dataframe of lines of text and separates it into individual words.
4. Remove **stop words** which are common words such as "every", "with", or "doing" (selected from `stop_words` dataset).
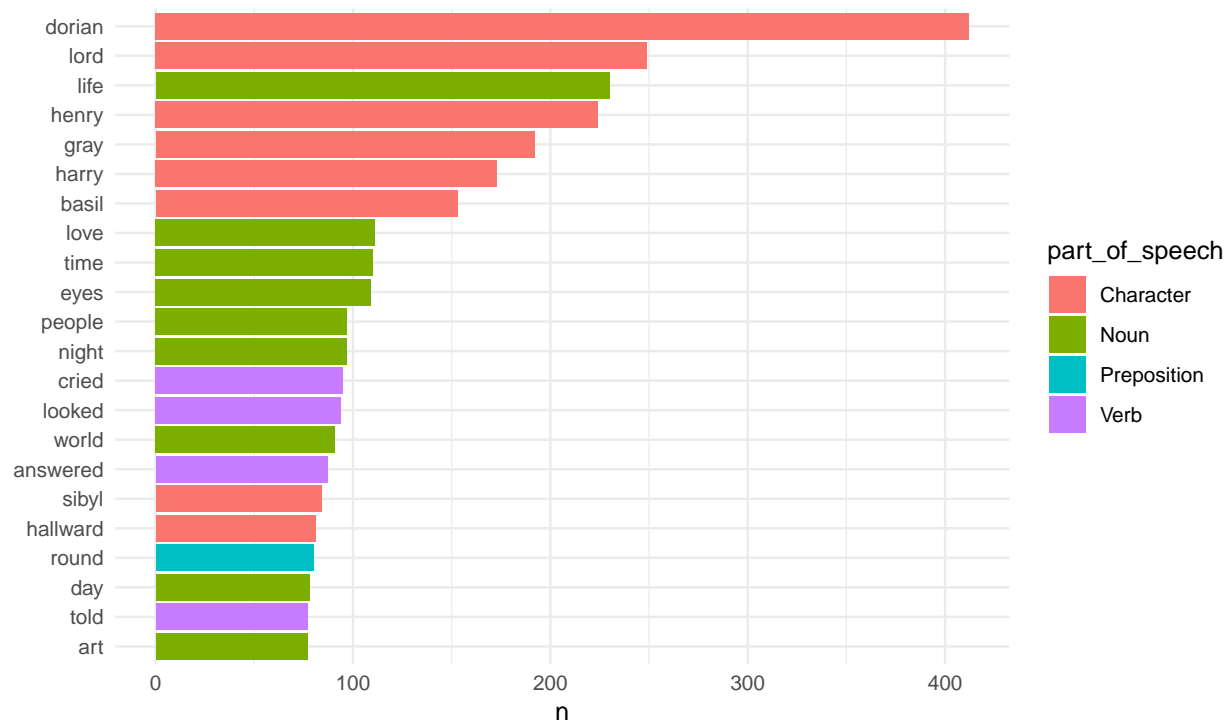5. Do the analysis

Figure 1 shows the count of the most common non-stop words in the Picture of Dorian Gray. As we might expect words related to characters are most common. Interestingly the word "round" is one of the most common words. In the novel, Wilde does not use round as an adjective, but rather as a preposition, for example, "I turned half-way round, and saw Dorian Gray for the first time."

Sentiment Flow with lexicons

First using the Bing lexicon, I calculated the overall sentiment of the text. The Bing lexicon assigns a positive or negative sentiment to words. For example "beautiful" is coded as positive, and "corrupt" is coded as negative. If there are more negative words, as is true in *The Picture of Dorian Gray*, the result is a negative integer.

Next I used the Afinn Lexicon which codes words with integers from -5 to 5. The only +5 word in the text is "superb," and there are no -5 words identified. Level -4 rated words include "torture" and "hell", for example. The average word sentiment is -0.2, and there were more negative words identified than positive ones.

For the 3rd example following the textbook, I used the NRC Lexicon to identify sentiments. The NRC Lexicon codes words with a sentiment word, for example, the word "beautiful" is coded as both "joy" and "positive," while "accident" is coded as "fear", "negative", "sadness", and "surprise." I was surprised to find that more words are coded as positive than negative using the NRC lexicon.

Extra Lexicon

I've selected a lexicon that has common words from each decade.Words have changed meaning over time and in a sentiment analysis these differences should be considered. Since *The Picture of Dorian Gray* was published in 1890, I've selected the lexicon for the decade 1890. The lexicon can be accessed from the following link: *https://nlp.stanford.edu/projects/socialsent/*.

Based on Figure 2, the lexicon with words and sentiments from 1890 matches well with the other lexicons. There are slight variations, as we would expect, but the trends are largely the same. The shared pattern is more distinct in the second half of the text. The average sentiment based on the year specific lexicon is 0.02, which suggests a slightly positive sentiment. This lexicon identifies "admirable", "beauty", and "fine" as the
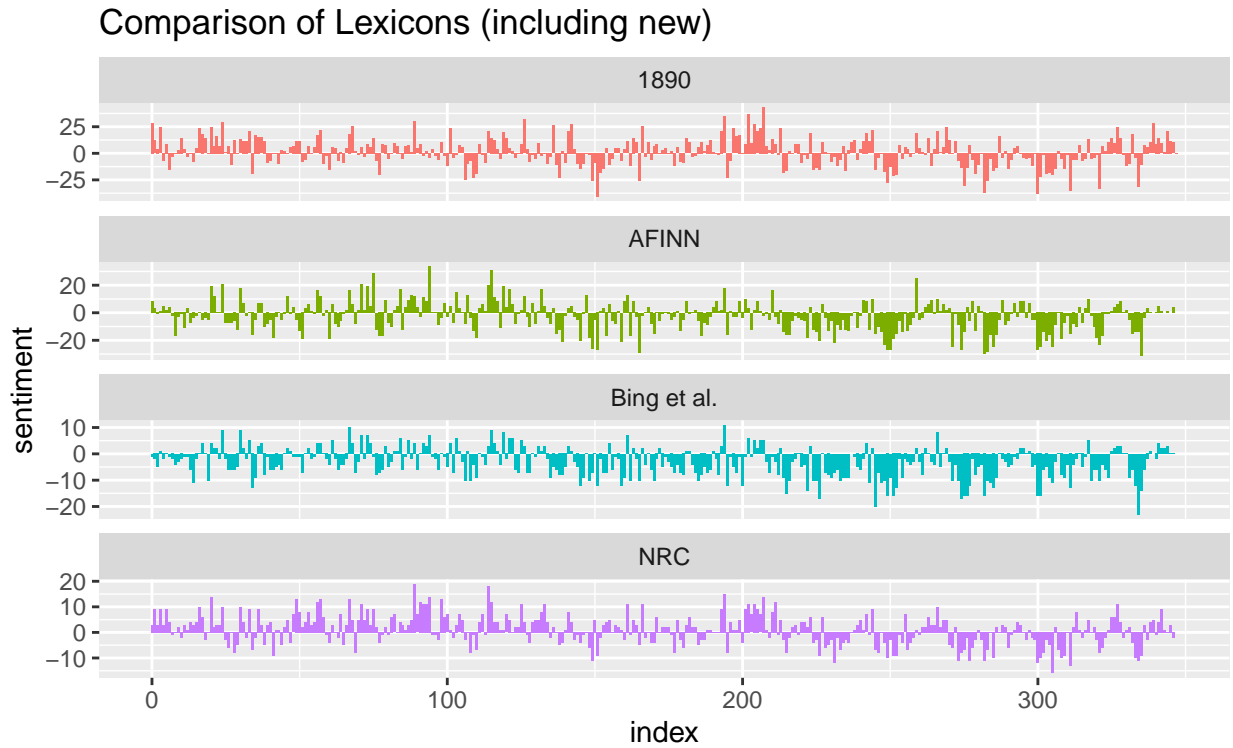
Figure 1: Comparison of sentiment ratings based on lexicon. Each bar represents the sum of word sentiments in groups of 25 lines. Note that around index 250 (or around line 6250) there is a distinct group of negiive sentiments identified in all four lexicons.

strongest positive words, and "cruelty", "terrible", and "wicked" as the strongest negitive words.

William L. Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky. Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora. ArXiv preprint (arxiv:1606.02820). 2016. https://nlp.stanford.edu/proje cts/socialsent/

Sentence Level Analysis

Methods

1. Set up true number space

2. Read book with true numbers

3. Use a query to access database with sentences

4. Complete sentence level analysis

Sentimentr

Using a similar plotting method, we can see the trends in sentence sentiment over the course of the text. This plot is similar to the bag of words plot, but a bit more scattered. For example, in the bag of words analysis, around index 250, there is a distinct negative sentiment detected in all of the lexicons, but in this analysis there is not the same pattern detected.
This section of negative text corresponds with an area of conflict in the text where Dorian murders Basil. Here are two example sentences.

1 "The mad passions of a hunted animal stirred within him, and he loathed the man who was seated at the table, more than in his whole life he had ever loathed anything" (Sentimentr sentence rating: -0.451, Paragraph rating: -0.132) 2. "Gradually the events of the preceding night crept with silent blood-stained feet into his brain, and reconstructed themselves there with terrible distinctness. (Sentimentr rating: -0.073, Paragraph rating: -0.327)

Both of these sentences are clearly negative and are coded as such. The paragraphs are also coded as having a negative sentiment.

Character Analysis

Histograms of character sentiments

I selected four of the main characters, and using the sentence sentiment values created histograms for the sentiment of sentences that mention the character's name. This method is not without fault though, since I do not identify sentences such as, "As a rule, he is charming to me, and we sit in the studio and talk of a thousand things" where Basil is describing his friendship with Dorian, because Basil only refers to Dorian as "he" rather than by his name.
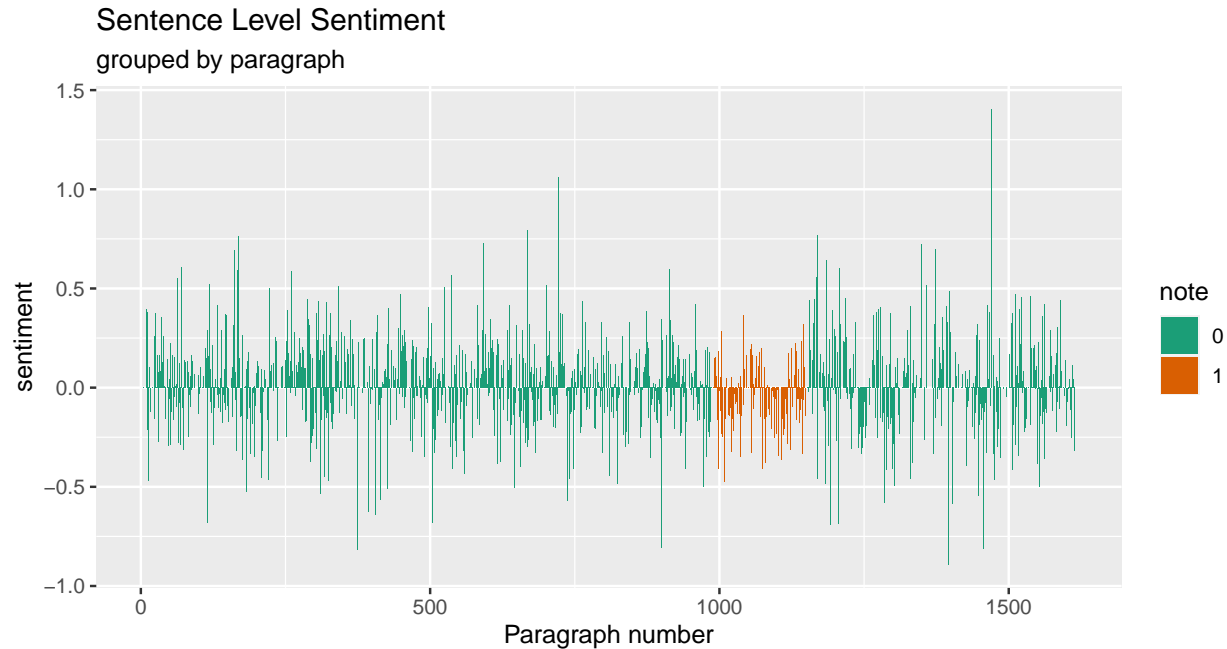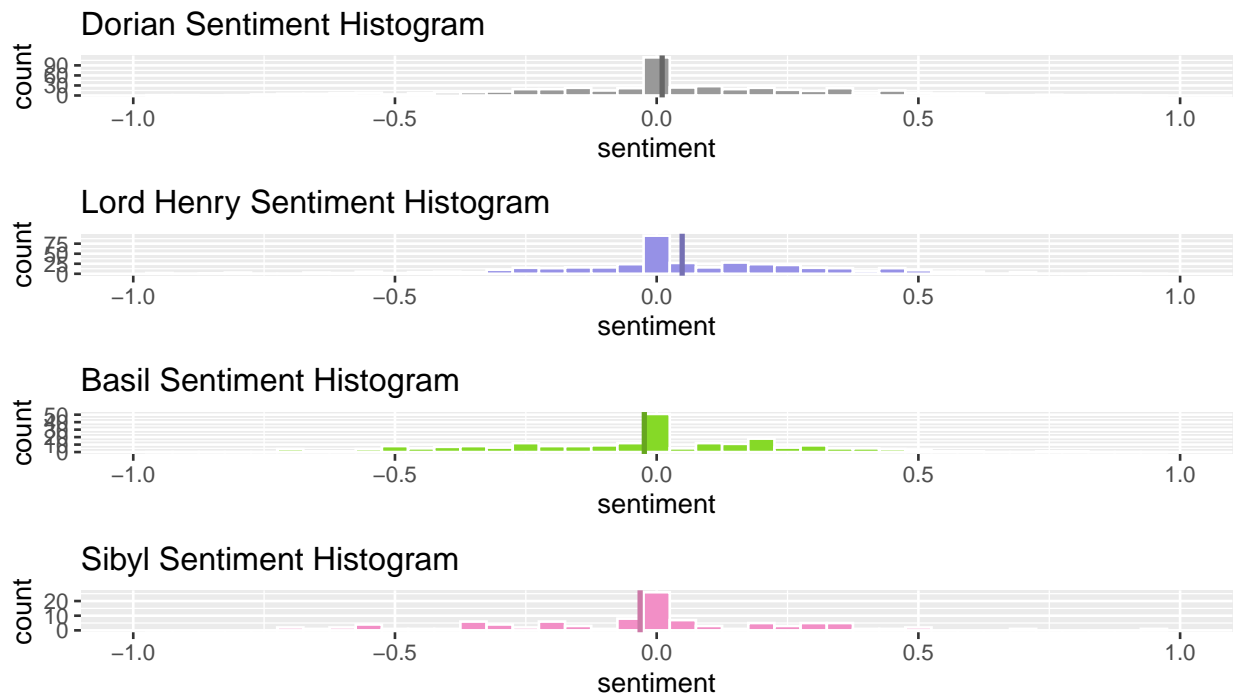
Sentence Level Sentiment
grouped by paragraph

Figure 2: Sentence level sentment grouped by paragraph. The pattern is similar to others that use a bag of words analysis, but the pattern is not the same. The orange bars represent the two chapters where Dorian kills Basil. Here some negative sentiment is detected, but these are not the most negative paragraphs identified.

Dorian Sentiment Histogram

Lord Henry Sentiment Histogram

Basil Sentiment Histogram

Sibyl Sentiment Histogram

Next, I plotted the character sentiment over time. The light gray bars represent all of the text and the black bars represent the text where the character is called by name. Some patterns arise that inform the plot of the book. For example, Dorian is the main character since there are sentences that refer to him throughout the text. We can also see where Sibyl dies approximately half way through the text in her plot since there is
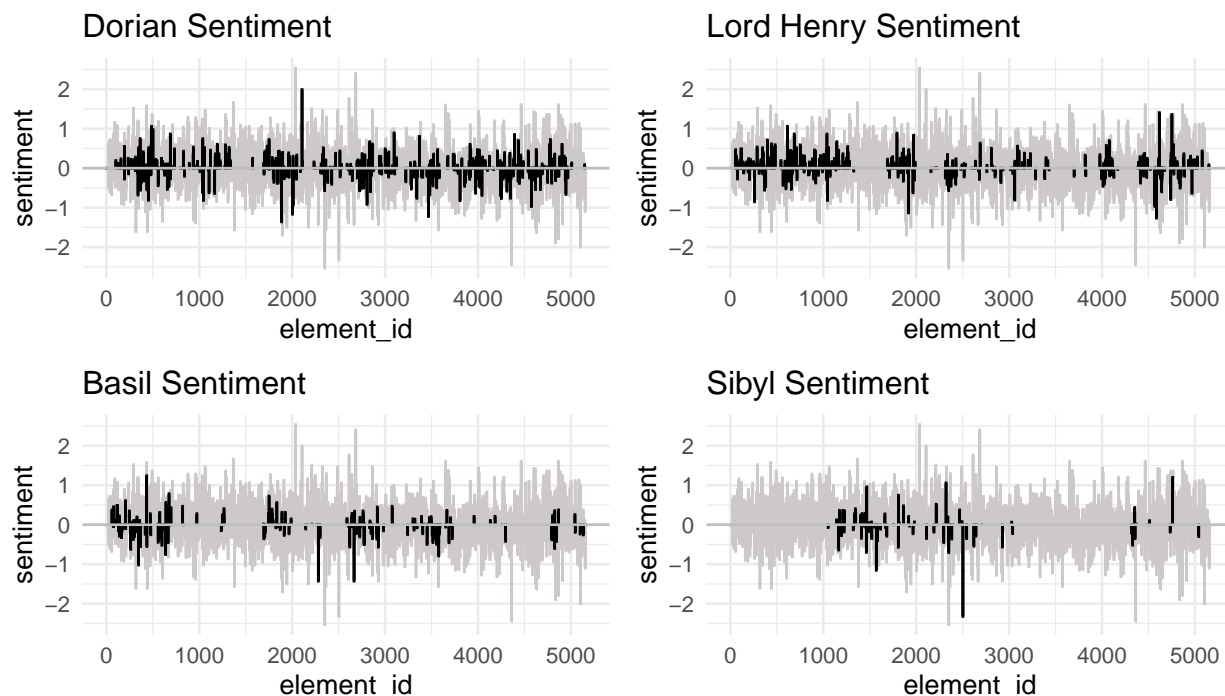
a strong negative sentiment identified.

Figure 3: Timeline of sentence sentiment colored by character occurance

Character co-occurance and sentiment histograms

A limitation of this analysis is that it does not differentiate between the character speaking or narrative. The following sentence is a simple narrative sentence. "Dorian Gray frowned and turned his head away." It is clear that Dorian is having some type of negative experience. In contrast, it is difficult to contextualize each sentence as a relationship between characters. For example, in this passage Basil is speaking about the inspiration that Dorian provides him. Wilde writes, "He is all my art to me now," said the painter, gravely... "What the invention of oil-painting was to the Venetians, the face of Antinoüs was to late Greek sculpture, and the face of Dorian Gray will some day be to me." Here, in several sentences, Basil conveys a positive sentiment towards Dorian, but in the analysis, only the name "Dorian Gray" is tagged. There is not a convenient way to indicate who is speaking about whom and how that relates to the sentiment. None the less, this analysis provides some insights into character relationships.

I would be curious to extract information about how characters feel about eachother. For example, Basil is enamored with Dorian, but their friendship breaks down. For example, this passage narrates Dorian's feelings about Basil, "Basil had said things to him that were unbearable, and that he had yet borne with patience. The murder had been simply the madness of a moment."
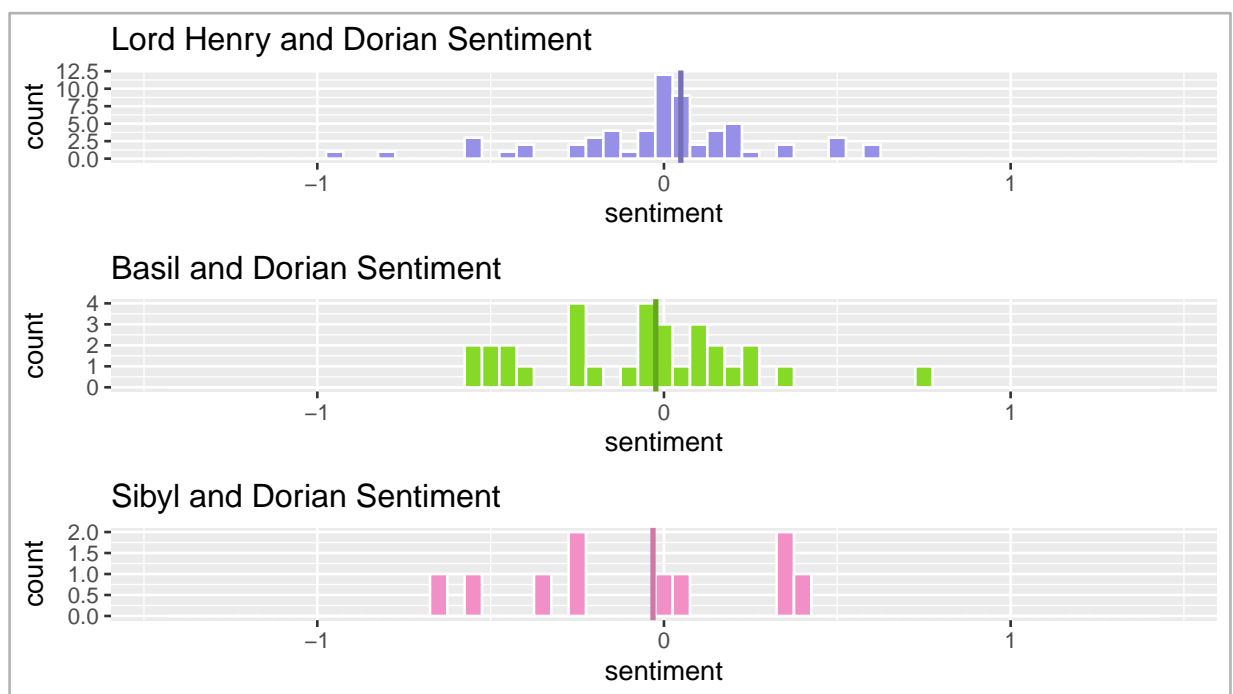
Figure 4: Co-occurance of character's sentiment histogram