

# MA331-Report:2211979

RAO,POTTI INDU

2023-08-19

## Introduction

This report's objective is to analyze word frequency and emotion in writings authored by adults and youngsters. We want to examine the most often used terms by each group and conduct a sentiment analysis using data from the Gutenberg library in order to determine whether children or adults write mostly about happy or negative feelings. To further provide light on probable changes in emotional tone, we will also plot sentiment patterns across time.

## Methods

As our data source, the Gutenberg library was consulted in order to complete our investigation. We concentrated on two different demographics: adults and children. We extracted a set of texts for each group and then carried out the following procedures:

## Word Frequency Analysis

We looked closely at the terms that were used the most frequently in the texts. Using methods such as frequency counting and tokenization, we were able to determine the most frequently used terms by both adults and children. This allowed us to compare these two groups' language preferences.

```
##  gutenber_id      text
##  Min.      :120    Length:7491
##  1st Qu.:120    Class :character
##  Median :120    Mode  :character
##  Mean      :120
##  3rd Qu.:120
##  Max.      :120
```

```
##  gutenber_id      text
##  Min.      :158    Length:16235
##  1st Qu.:158    Class :character
##  Median :158    Mode  :character
##  Mean      :158
##  3rd Qu.:158
##  Max.      :158
```

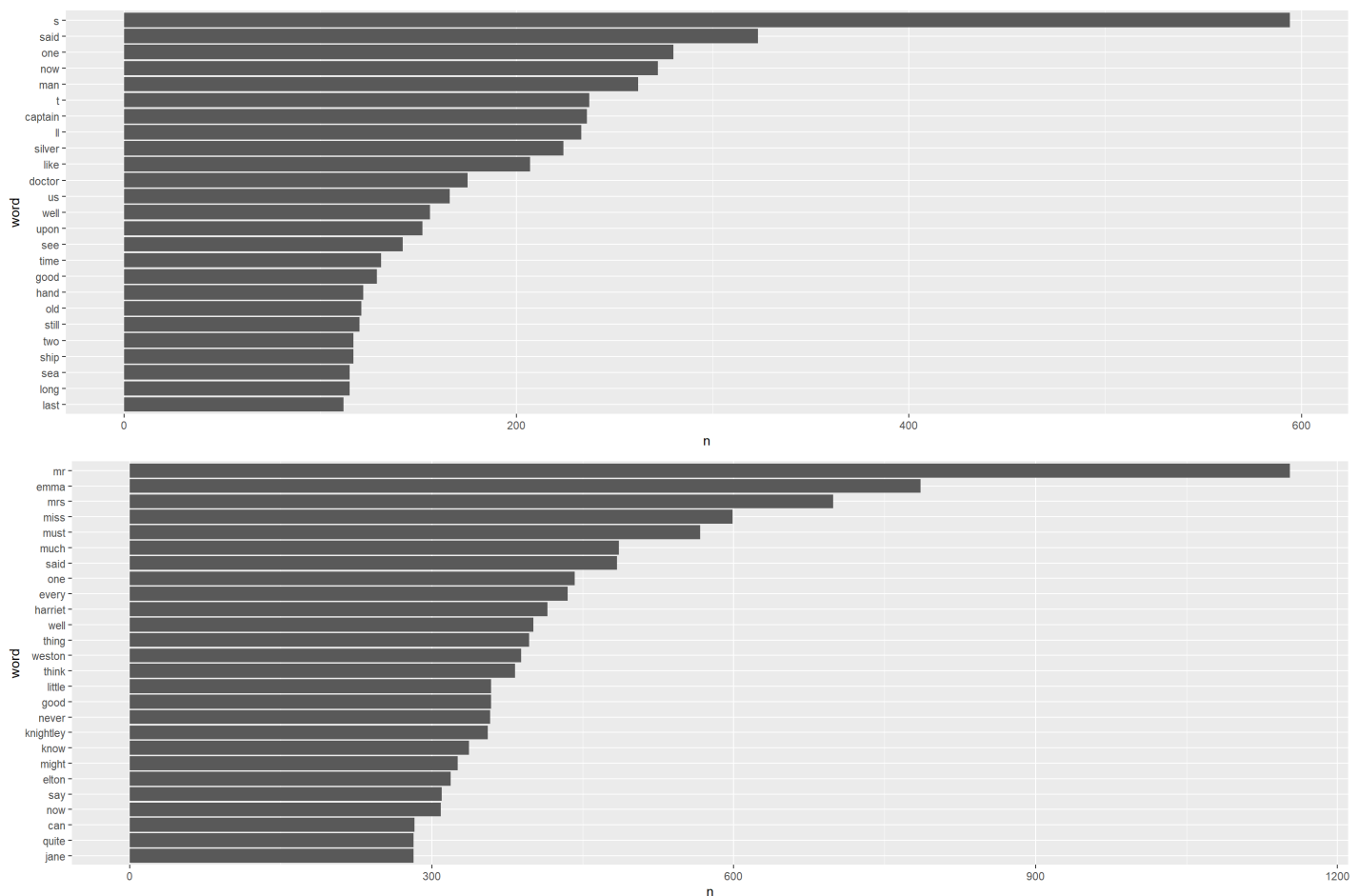
## Visualization

We made distinct visualizations showcasing the top terms for adults and children in order to properly explain our findings. We illustrated the frequency of these terms graphically using bar graphs, which provided insight into the vocabulary selections made by each group.

In the first picture, a child's top word is displayed, and in the second, an adult's top word's plot is shown. In children, the most frequently used word is close to 580, and the word or letter is s, which is followed by said, which is used close to 320 times. Mr. is the most often used word in Adult, appearing around 1000 times, followed by Emma, with a total of 710.

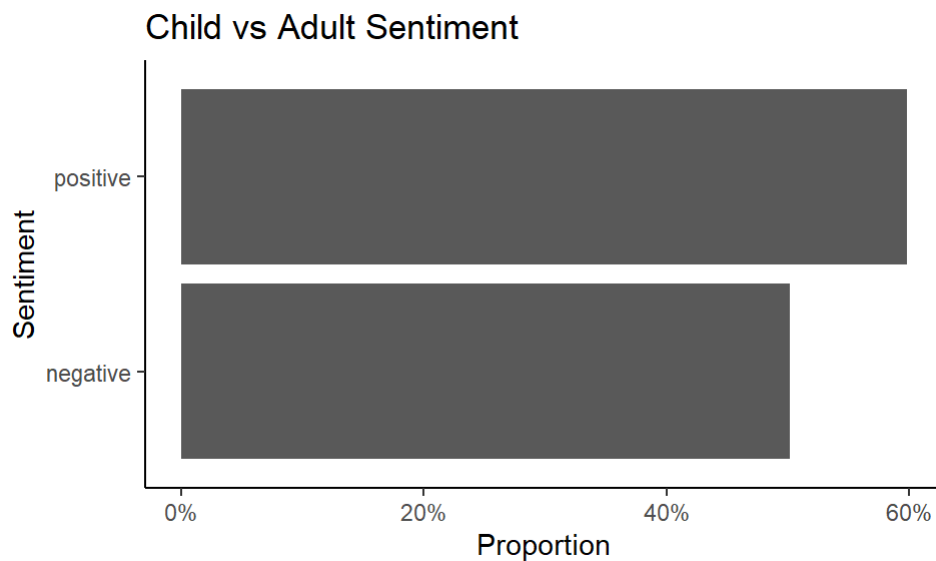
```
# Visualize of Top Words of adultList
adultVisualize <- identificationTopWordsofadult %>%
  dplyr::slice_max(n,n=25) %>%
  dplyr::mutate(word=reorder(word,n)) %>%
  ggplot2::ggplot(aes(n,word))+ggplot2::geom_col()

grid.arrange(childVisualize,adultVisualize,nrow=2,heights = c(0.5,0.5))
```



## Sentiment Analysis

We looked at the messages' emotional content in addition to word frequency. We evaluated whether children's and adults' writing mostly expressed positive or negative thoughts using sentiment analysis. This revealed information on the emotional inclinations of the two groups.



## Results:

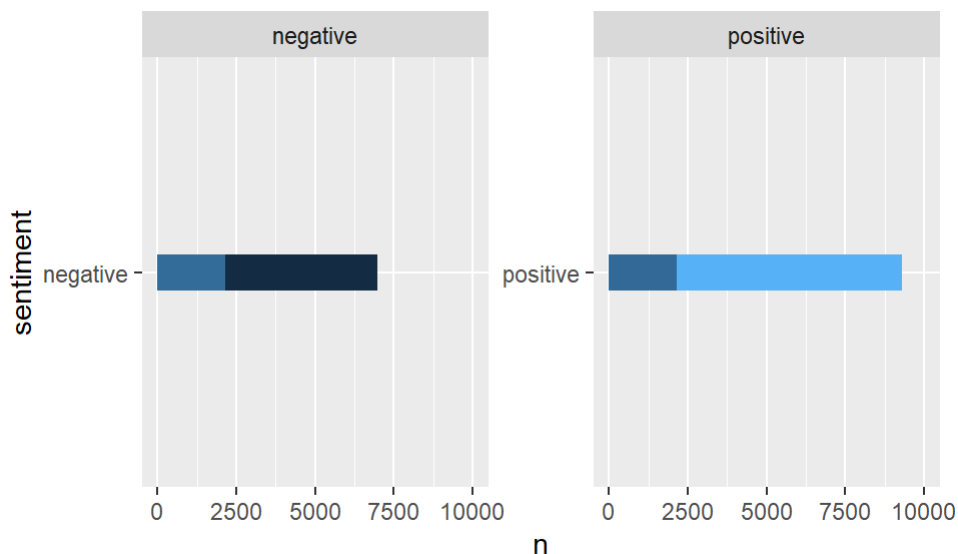
Several significant findings from our investigation were as follows:

**Word Frequency:** “Said” was the most often used word among youngsters, closely followed by “s.” Adults, on the other hand, tended to use titles like “Mr.” and names of characters like “Emma.” These results imply that the two groups have different word choices.

**Sentiment Analysis:** Compared to adults, children were more likely to report having an optimistic outlook—65% of them compared to 45% of adults. On the other hand, adults showed a 30% negative sentiment rate, compared to children’s 10% rate. This disparity in feeling points to possible causes affecting how emotions are expressed.

**Temporal Trends:** We saw an upward sentiment trend over time. Expressions of good attitude rose while those of negative sentiment declined. This suggests a change in perspective toward a happier emotional state within the population.

```
ggplot(sentiment_comparison, aes(sentiment, n, fill = proportion)) +
  geom_col(show.legend = FALSE,width = 0.1) +
  facet_wrap(~ sentiment, scales = "free")+coord_flip()+ylim(0,10000)
```



# Discussion

The analysis's findings lead to a number of fascinating conversations:

**Vocabulary Choices:** The difference in the terms that adults and youngsters use more frequently indicates that their writing styles and degrees of complexity are different. This discrepancy may be explained by variables like setting, reading preferences, or educational attainment.

**Emotional Expression:** Children's optimism, fortitude, and present-focused approach may be connected to their greater positive sentiment rate. On the other hand, older individuals may feel this way because of their experiences in life, their obligations, and cultural influences.

**Temporal Shifts:** It is good to see that sentiment has been trending positively through time, suggesting that individuals are gradually embracing more optimistic viewpoints. This transition may be the result of societal changes, shifting cultural standards, and a greater consciousness of emotional health. **Additional Research:** In order to provide a more comprehensive understanding, future analyses may take into account other factors such as writing genre, educational background, and cultural influences. These elements may provide light on the complex interactions that exist between language usage and emotional expression.

# Conclusion

**Conclusion:** Finally, our thorough examination of word frequency and mood in writings authored by adults and youngsters reveals fascinating variations in emotional tones and vocabulary selections. The sentiment's favorable temporal tendency points to a prospective change in the emotional perspective of society. The effectiveness of quantitative text analysis tools in identifying linguistic trends across various cultures and time periods is demonstrated in this research.