

Shakespeare Play Analysis

Import and Tidy Data

The purpose of the sample project is describing the overall atmosphere of the prominent plays of Shakespeare. We will be analyzing words from the plays to conclude whether his masterpieces are either sentimentally positive or negative depending on the type such as Comedy or Tragedy. Hypothetically, Comedy should have more positive atmosphere and vice versa. First, we will count how many titles and types are in Shakespeare.

```
# Import Six Shakespeare Acts' Data
shakespeare = read_csv("shakespeare.csv")

# Pipe the shakespeare data frame to the next line
shakespeare %>%
  # Use count to find out how many titles/types there are
  count(title, type) %>%
  arrange(desc(n))
```

```
## # A tibble: 6 x 3
##   title                                type      n
##   <chr>                                <chr>   <int>
## 1 Hamlet, Prince of Denmark          Tragedy 6776
## 2 The Tragedy of Romeo and Juliet   Tragedy 4441
## 3 The Merchant of Venice             Comedy  4225
## 4 Much Ado about Nothing             Comedy  3799
## 5 A Midsummer Night's Dream         Comedy  3459
## 6 The Tragedy of Macbeth             Tragedy 3188
```

Hamlet, Prince of Denmark, seems to have the most lines compared to other five Shakespeare Plays, having **6776** lines. Next, we will redistribute texts into words and respective line numbers.

```
tidy_shakespeare <- shakespeare %>%
  # Group by the titles of the plays
  group_by(title) %>%
  # Define a new column linenumbr
  mutate(linenumbr = row_number()) %>%
  # Transform the non-tidy text data to tidy text data
  unnest_tokens(word, text) %>%
  ungroup() %>%
  select(-X1)
head(tidy_shakespeare)
```

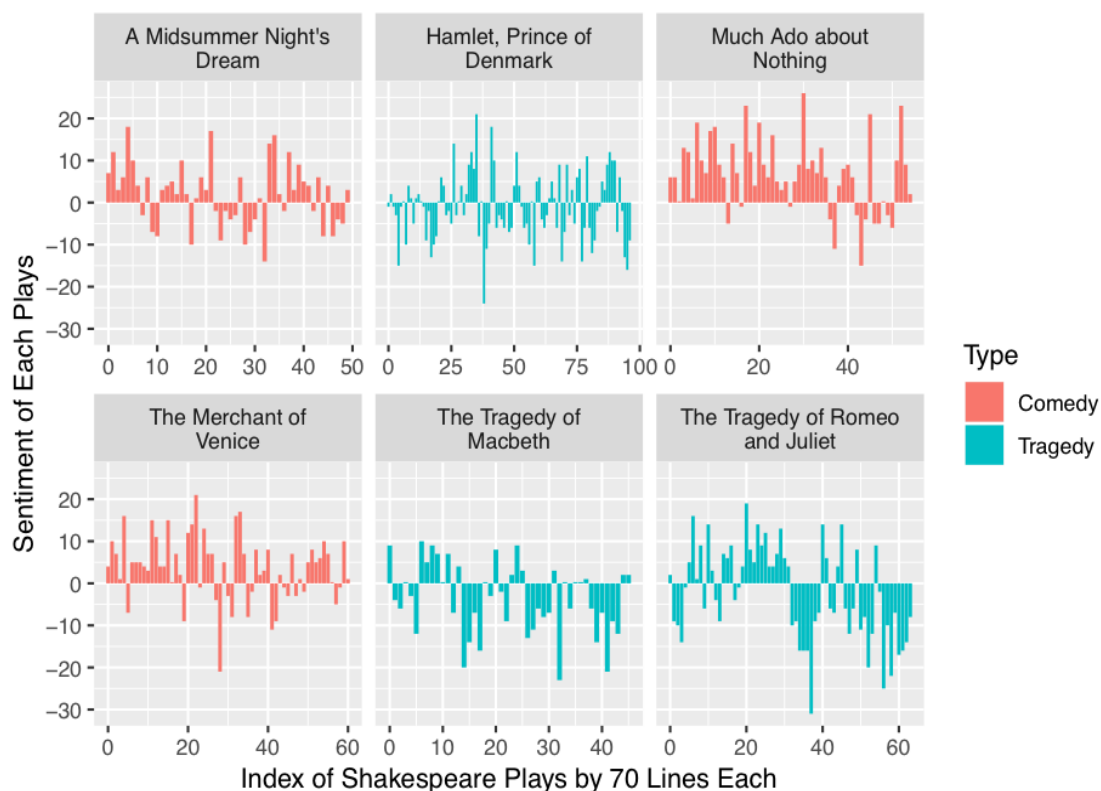
```
## # A tibble: 6 x 4
##   title                                type  linenumbr word
##   <chr>                                <chr>    <int> <chr>
## 1 The Tragedy of Romeo and Juliet Tragedy      1 the
## 2 The Tragedy of Romeo and Juliet Tragedy      1 complete
## 3 The Tragedy of Romeo and Juliet Tragedy      1 works
## 4 The Tragedy of Romeo and Juliet Tragedy      1 of
## 5 The Tragedy of Romeo and Juliet Tragedy      1 william
## 6 The Tragedy of Romeo and Juliet Tragedy      1 shakespeare
```

Plot the Cleaned Shakespeare Data

Next, we are going to use `tidy_shakespeare` to analyze Six Shakespeare plays. Which how does each sentiment fluctuate from the beginning to the end?

Plot Sentiment Changes with Respect to PLAYS

```
tidy_shakespeare %>%  
  inner_join(get_sentiments("bing")) %>%  
  count(title, type, index = linenumber %/% 70, sentiment) %>%  
  spread(sentiment, n, fill = 0) %>%  
  mutate(sentiment = positive - negative,  
         fixed_title = str_wrap(title, width = 20)) %>%  
  # Put index on x-axis, sentiment on y-axis, and map comedy/tragedy to fill  
  ggplot(aes(x = index, y = sentiment, fill = type)) +  
  # Make a bar chart with geom_col()  
  geom_col() +  
  # Separate panels for each title with facet_wrap()  
  facet_wrap(~ fixed_title, scales = "free_x") +  
  labs(x = "Index of Shakespeare Plays by 70 Lines Each",  
       y = "Sentiment of Each Plays") +  
  scale_fill_discrete(name = "Type")
```



Conclusion

As expected, comedic Shakespeare plays tend to have more positive sentiments while tragic acts tend to be more negative. Notice that *The Tragedy of Romeo and Juliet* follows the somewhat positive introduction and then the tragic end.