

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. 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)
```

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

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 linenumbers
  mutate(linenumbers = 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  linenumbers 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

```
# 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")
```

