Project1

Orianne Luo

9/28/2021

```
data<-read.csv(file = 'C:/Users/orian/Downloads/philosophy_data.csv/philosophy_data.csv')
```

Load all the required packages:

```
library(dplyr)
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(ggplot2)
```

What are all the schools, and how many authors do we have for each school?

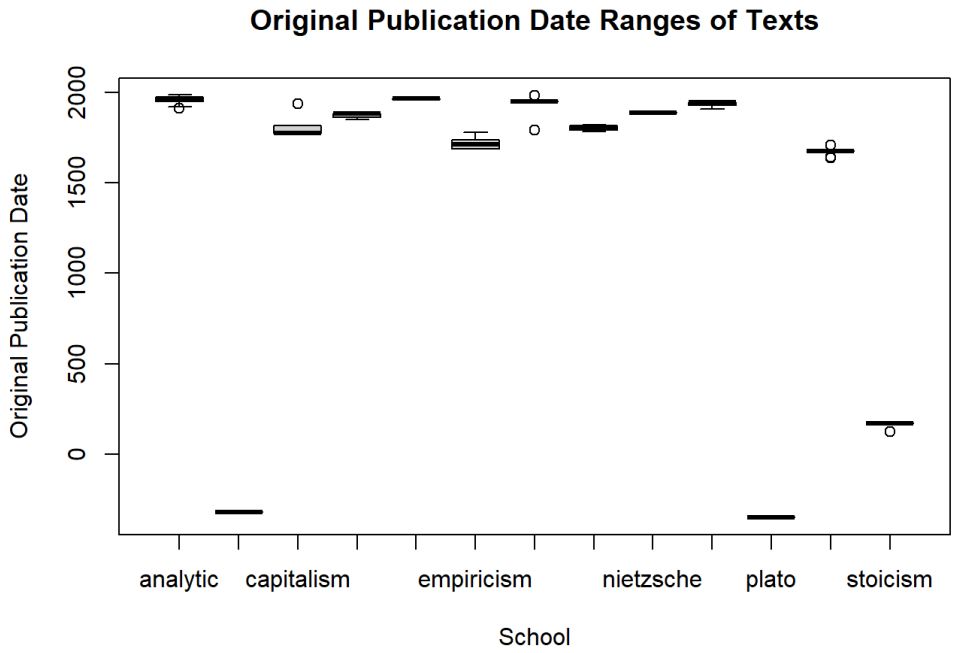
```
data %>% group_by(school) %>% summarise(author_count=n_distinct(author))
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 13 x 2
##    school          author_count
##    <chr>                  <int>
##  1 analytic                   7
##  2 aristotle                  1
##  3 capitalism                 3
##  4 communism                  2
##  5 continental                3
##  6 empiricism                 3
##  7 feminism                   3
##  8 german_idealism            3
##  9 nietzsche                  1
## 10 phenomenology              3
## 11 plato                      1
## 12 rationalism                4
## 13 stoicism                   2
```

What are the date ranges for these different schools? We can see that the original publication dates for these works are all pretty modern, except for Aristotle, Plato, and Stoicism.

```
boxplot(original_publication_date~school, data=data, main="Original Publication Date Ranges of Texts", xlab="School", ylab=
"Original Publication Date")
```



Question 1: Which philosophers have the longest sentences?

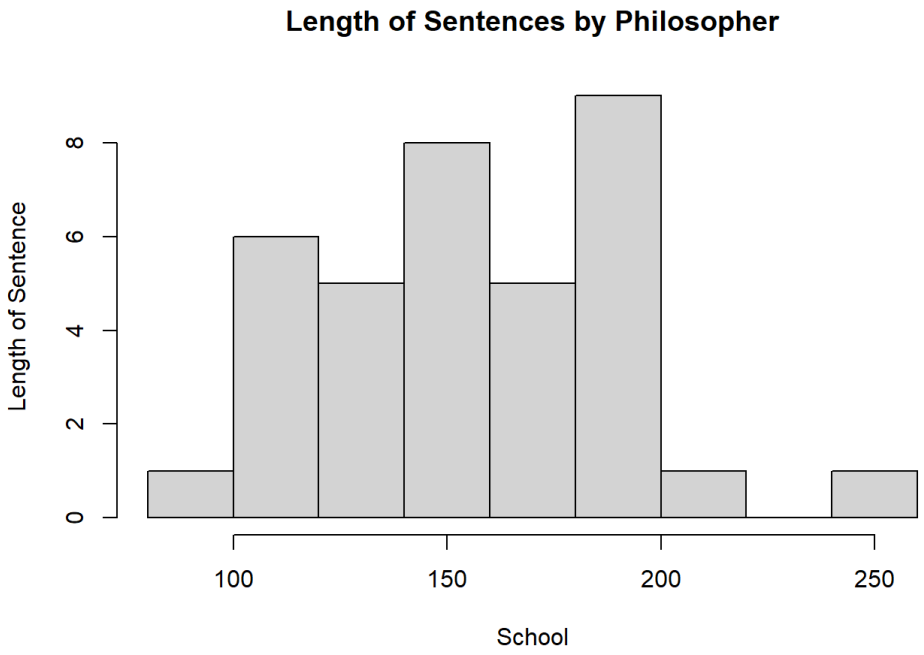
```
slength<-data %>% group_by(author) %>% summarize(mean_sentence_length=mean(sentence_length)) %>% arrange(desc(mean_sentence_
length))

slength
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 36 x 2
##    author         mean_sentence_length
##    <chr>                         <dbl>
##  1 Descartes                      247.
##  2 Locke                          200.
##  3 Kant                           198.
##  4 Keynes                         197.
##  5 Wollstonecraft                 191.
##  6 Foucault                       190.
##  7 Ricardo                        186.
##  8 Husserl                        185.
##  9 Smith                          185.
## 10 Lenin                          181.
## # ... with 26 more rows
```

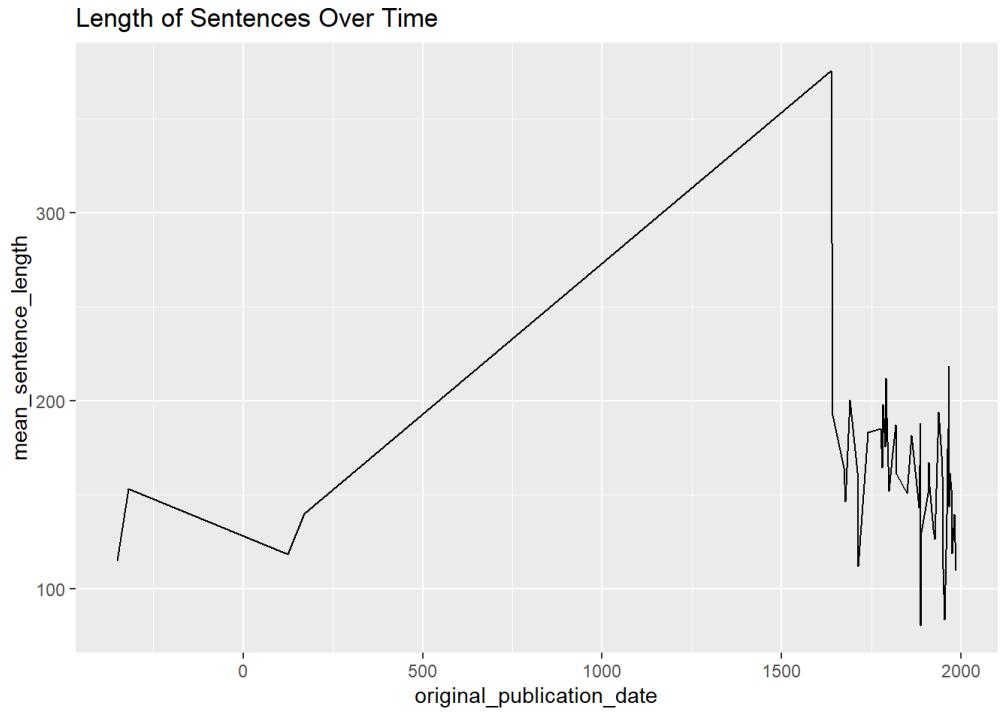
We can create a visualization of the mean sentence lengths of these different philosophers. As we can see from the visualization, Descartes had much longer sentences than the other philosophers, over 200 words on average, while Wittgenstein had shorter sentences on average than other philosophers, less than 90 words on average. Overall, the distribution looks slightly skewed to the left.

```
hist(slength$mean_sentence_length, main="Length of Sentences by Philosopher", xlab="School", ylab="Length of Sentence")
```



Does the year of publication affect sentence length? We can see a sharp increase in the average sentence length in the 1500s, and then a lot of variation in the sentence length into the present day.

```
slengthy<-data %>% group_by(original_publication_date) %>% summarize(mean_sentence_length=mean(sentence_length)) %>% arrange
(desc(mean_sentence_length))

ggplot(slengthy, aes(x=original_publication_date, y=mean_sentence_length)) +
  geom_line() +
  ggtitle("Length of Sentences Over Time")
```



Question 2: Which topics were philosophers of each school most interested in? Which topics were philosophers most interested in overall? We can explore this by looking at the most common words overall, and then by school.

```
library(tidytext)
```

```
data1 <- data %>% separate_rows(tokenized_txt) %>% filter(tokenized_txt!="")
```

After cleaning the data, count the number of times each word appears for each of the schools. I removed the stop words to get the most common words. It seems like overall, the philosophers were most interested in “time,” “nature,” “world,” “reason,” and “true.”

```
data1 %>% group_by(tokenized_txt) %>% count() %>% arrange(desc(n)) %>% anti_join(stop_words, by=c("tokenized_txt"="word"))
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 89,497 x 2
## # Groups:   tokenized_txt [89,497]
##    tokenized_txt     n
##    <chr>         <int>
##  1 time          15307
##  2 nature        12796
##  3 world         12088
##  4 reason        11464
##  5 true           9435
##  6 form           9262
##  7 sense          9255
##  8 body           9047
##  9 object         8848
## 10 knowledge      8117
## # ... with 89,487 more rows
```

As expected, we can see that there is a good deal of diversity in the most commonly mentioned words/topics among different philosopher schools. See the bottom of the notebook for a wordcloud for all of these schools.

```
data1 %>% anti_join(stop_words, by=c("tokenized_txt"="word")) %>% group_by(school, tokenized_txt) %>% count() %>% arrange(de
sc(n)) %>% group_by(school) %>% slice(1:5)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 65 x 3
## # Groups:   school [13]
##    school    tokenized_txt     n
##    <chr>     <chr>         <int>
##  1 analytic  true           3055
##  2 analytic  sense          2518
##  3 analytic  theory         2138
##  4 analytic  language       1853
##  5 analytic  world          1835
##  6 aristotle animals        2508
##  7 aristotle time           2379
##  8 aristotle body           2058
##  9 aristotle nature         2047
## 10 aristotle motion         1912
## # ... with 55 more rows
```

Question 3: How do different schools of thought think differently about certain topics?

Good: How do the different schools view Good differently? The Analytic school mainly relates good with “reason,” “sense,” and “true,” indicating an emphasis on rationality. On the other hand, the Aristotle school most commonly draws comparisons with the other extreme, including words like “Bad” and “Evil,” while also associating Good with “Excellence” and “Pleasure.” The schools Capitalism and Communism mainly use “Good” in relation to products, as indicated by “price,” “money,” “time,” and “production.” The Continental school associates Good with “sense,” “nature,” and

“common,” implying an intrinsic existence. The Empiricism school juxtaposes Good with “Power” and “Evil,” but also mentions “reason” and “nature,” indicating an objective view of what is good. The school of Feminism most commonly mentions Good in relation to “woman,” “women,” and “mother.” On the other hand, German idealism most commonly mentions Good with relation to “reason,” “moral,” “law,” and “evil,” indicating an importance placed on upholding what is good. The Nietzsche school most commonly mentioned good in relation with “evil,” “god,” and “people,” indicating a higher level of good in relation with god. On the other hand, the Phenomenology mentions good in relation with “conscience,” “life,” and “reason,” indicating an importance placed on adhering to Good. The school of Plato mentions Good in relation with “bad,” “people,” “city,” and “soul,” indicating an importance placed on bringing the people closer towards what is good. The school of Rationalism mentions good with relation more abstract concepts such as “god,” “evil,” and “love.” Finally, the school of Stoicism mostly mentions good with relation to “thou,” “doth,” and “nature,” indicating a more subjective emphasis placed on Good with respect to the reader.

```
data_good<-data1[grepl("good", data1$lemmatized_str, fixed = TRUE),]

data_good %>% anti_join(stop_words, by=c("tokenized_txt"="word")) %>% filter(tokenized_txt!="good") %>% group_by(school, tok
enized_txt) %>% count() %>% arrange(desc(n)) %>% group_by(school) %>% slice(1:5)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 65 x 3
## # Groups:   school [13]
##    school    tokenized_txt     n
##    <chr>     <chr>         <int>
##  1 analytic  reason           87
##  2 analytic  sense            51
##  3 analytic  true             43
##  4 analytic  theory           37
##  5 analytic  belief           35
##  6 aristotle bad             328
##  7 aristotle excellence      224
##  8 aristotle pleasure        195
##  9 aristotle evil            162
## 10 aristotle nature          162
## # ... with 55 more rows
```

Nature: We can see that some schools, such as Analytic and Phenomenology, are much more scientific and consider “laws,” “world,” and “sense,” while other groups, like Aristotle, are more concerned with more literal things like “animals” and “body.” Similarly, Communism and Capitalism schools primarily view nature in relation to the economy. Plato, on the other hand, discusses nature mainly with relation to people, using the words “soul” and “human.”

```
data_nature<-data1[grepl("nature", data1$lemmatized_str, fixed = TRUE),]

data_nature %>% anti_join(stop_words, by=c("tokenized_txt"="word")) %>% filter(tokenized_txt!="nature") %>% group_by(school,
tokenized_txt) %>% count() %>% arrange(desc(n)) %>% group_by(school) %>% slice(1:5)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 65 x 3
## # Groups:   school [13]
##    school    tokenized_txt     n
##    <chr>     <chr>         <int>
##  1 analytic  intrinsic        74
##  2 analytic  laws             74
##  3 analytic  sense             35
##  4 analytic  knowledge        32
##  5 analytic  question         29
##  6 aristotle animals         166
##  7 aristotle body            127
##  8 aristotle reason          110
##  9 aristotle matter          106
## 10 aristotle movement        105
## # ... with 55 more rows
```

True/Truth: We can see that many schools, especially Analytic and Aristotle, consider “false” in relation with truth. On the other hand, Capitalism and Communism still consider truth in relation with the economy, while Feminism still considers truth in relation with “woman.” Empiricism considers truth with relation to “ideas” and “knowledge,” while Phenomenology considers truth in relation to “world” and “knowledge.” Similarly, the school of Plato also often mentions “knowledge” in relation to truth.

```
data_true1<-data1[grepl("true", data1$lemmatized_str, fixed = TRUE),]
data_true2<-data1[grepl("truth", data1$lemmatized_str, fixed = TRUE),]

data_true<-rbind(data_true1, data_true2)
data_true %>% anti_join(stop_words, by=c("tokenized_txt"="word")) %>% filter(tokenized_txt!="true" & tokenized_txt!="truth")
%>% group_by(school, tokenized_txt) %>% count() %>% arrange(desc(n)) %>% group_by(school) %>% slice(1:5)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 65 x 3
## # Groups:   school [13]
##    school    tokenized_txt     n
##    <chr>     <chr>         <int>
##  1 analytic  false           521
##  2 analytic  truths          388
##  3 analytic  sentence        313
##  4 analytic  proposition     283
##  5 analytic  sense           265
##  6 aristotle false           299
##  7 aristotle conclusion      128
##  8 aristotle belongs          97
##  9 aristotle white            92
## 10 aristotle time             87
## # ... with 55 more rows
```

Question 4: How has the commonness of certain topics changed over time? Look at the number of times that the word is mentioned among the excerpts for the year that the text was originally published.

```
top5<-data1 %>% group_by(tokenized_txt) %>% count() %>% arrange(desc(n)) %>% anti_join(stop_words, by=c("tokenized_txt"="wor
d")) %>% head(5) %>% select(tokenized_txt)
```

```
data_count<-data1 %>% inner_join(top5, by="tokenized_txt") %>% group_by(tokenized_txt, original_publication_date) %>% count
() %>% arrange(desc(n))
```

We can see that Nature and Time were very common words among earlier philosophers. There was also a spike in the number of occurrences of the words Reason and Nature in the 1700s and 1800s, likely as a response to the schools of philosophy that arose from the Enlightenment. The word World went from being a not very common word to becoming one of the most common words among modern philosophers. This indicates a shift in the focus of philosophers, from focusing mainly on more natural phenomena to the popularity of scientific thinking and reasoning, and finally a greater awareness of the world around us.

```
library(ggplot2)
ggplot(data_count, aes(x=original_publication_date, y=n, group=tokenized_txt, color=tokenized_txt)) +
  geom_line()
```