

# R Notebook

[Code ▾](#)[Hide](#)

```
data <- read.csv('~/.gr5243/philosophy_data.csv')
```

[Hide](#)

```
unique(data$author)
```

```
[1] "Plato"      "Aristotle"  "Locke"      "Hume"       "Berkeley"
"Spinoza"    "Leibniz"    "Descartes"
[9] "Malebranche" "Russell"    "Moore"      "Wittgenstein" "Lewis"
"Quine"      "Popper"     "Kripke"
[17] "Foucault"    "Derrida"    "Deleuze"    "Merleau-Ponty" "Husserl"
"Heidegger"   "Kant"       "Fichte"
[25] "Hegel"      "Marx"       "Lenin"      "Smith"       "Ricardo"
"Keynes"     "Epictetus"  "Marcus Aurelius"
[33] "Nietzsche"   "Wollstonecraft" "Beauvoir"   "Davis"
```

[Hide](#)

```
unique(data$original_publication_date)
```

```
[1] -350 -320 1689 1739 1779 1713 1710 1677 1637 1641 1674 1921 1912 1910 1953 1985 1950 1959 1
972 1975 1963 1961 1966 1967 1968 1945 1936 1907 1927
[30] 1788 1790 1781 1798 1817 1807 1820 1883 1848 1862 1776 125 170 1888 1886 1887 1792 1949 1
981
```

[Hide](#)

```
library(tidytext)
data_words <- data %>%
  unnest_tokens(word, tokenized_txt)
data_count <- data_words %>%
  group_by(author, word) %>%
  count()
data(stop_words)
data_count_use <- data_count %>%
  anti_join(stop_words)
```

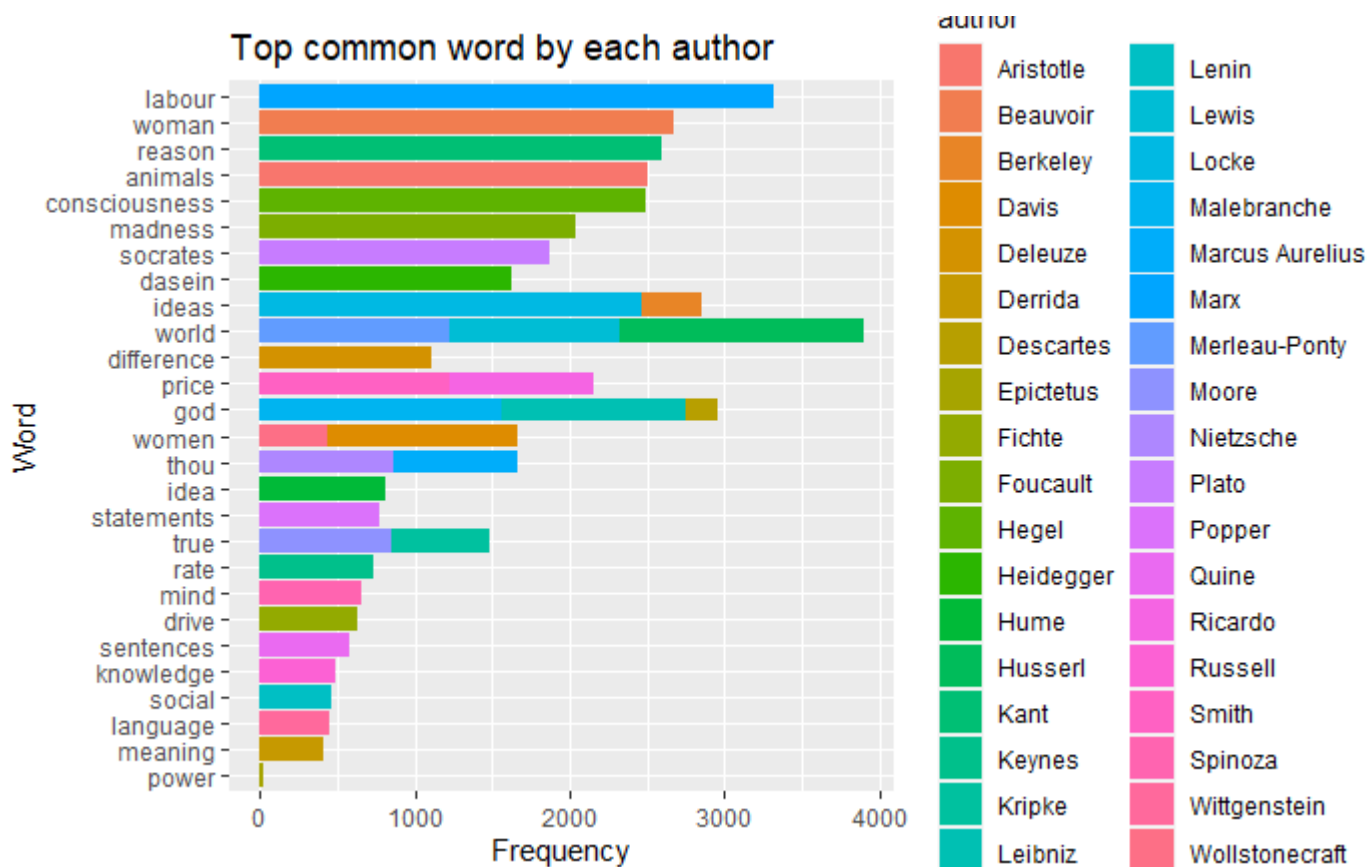
```
Joining, by = "word"
```

[Hide](#)

```

topwords <- data_count_use %>%
  group_by(author) %>%
  top_n(1, n)
ggplot(topwords, aes(x = reorder(word, n), y = n, fill = author)) +
  geom_col() +
  xlab(NULL) +
  coord_flip() +
  ggtitle("Top common word by each author") +
  xlab("Word") +
  ylab("Frequency")

```



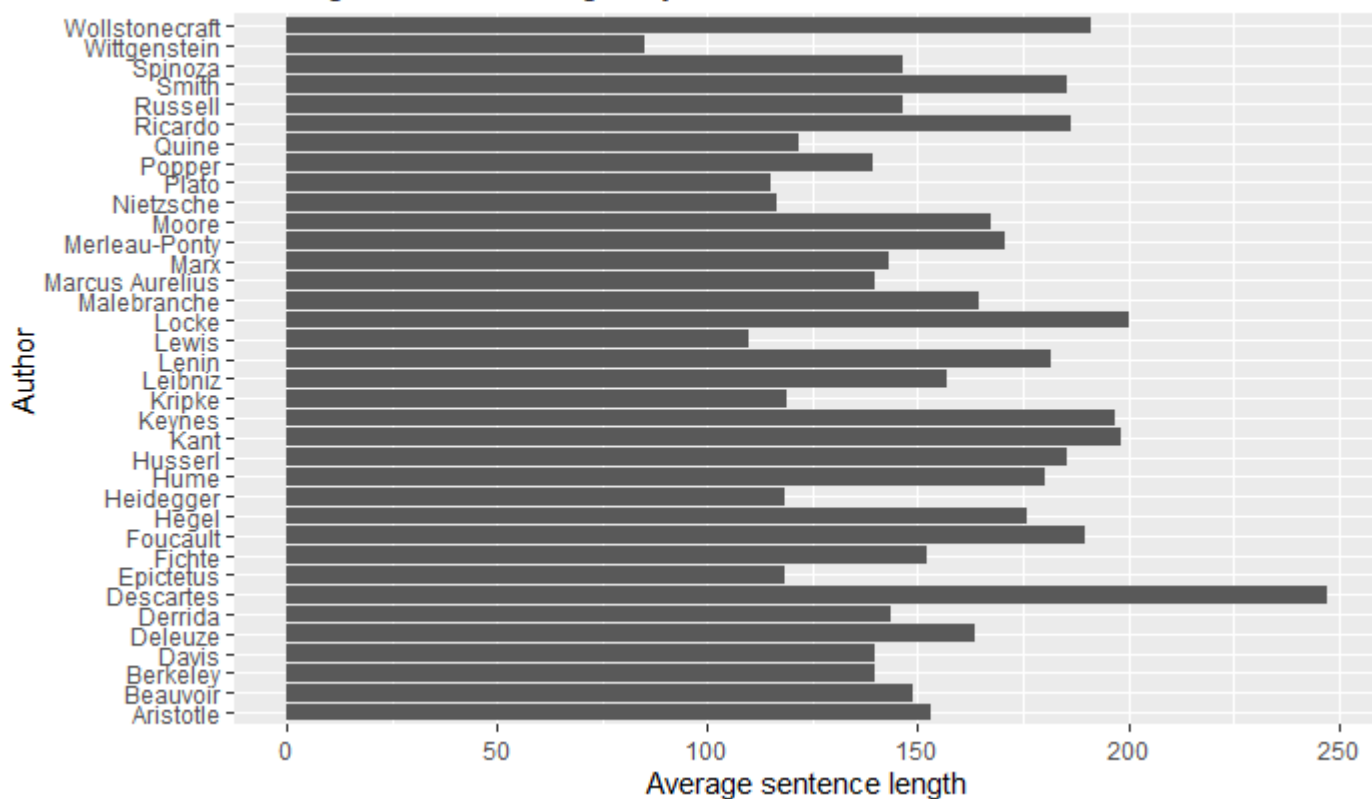
Hide

```

library(dplyr)
library(ggplot2)
data_avg <- data %>%
  group_by(author) %>%
  summarize(avg_sentence_length = mean(sentence_length))
ggplot(data_avg, aes(x = avg_sentence_length, y = author)) +
  geom_bar(stat = "identity") +
  ggtitle("Average sentence length by author") +
  xlab("Average sentence length") +
  ylab("Author")

```

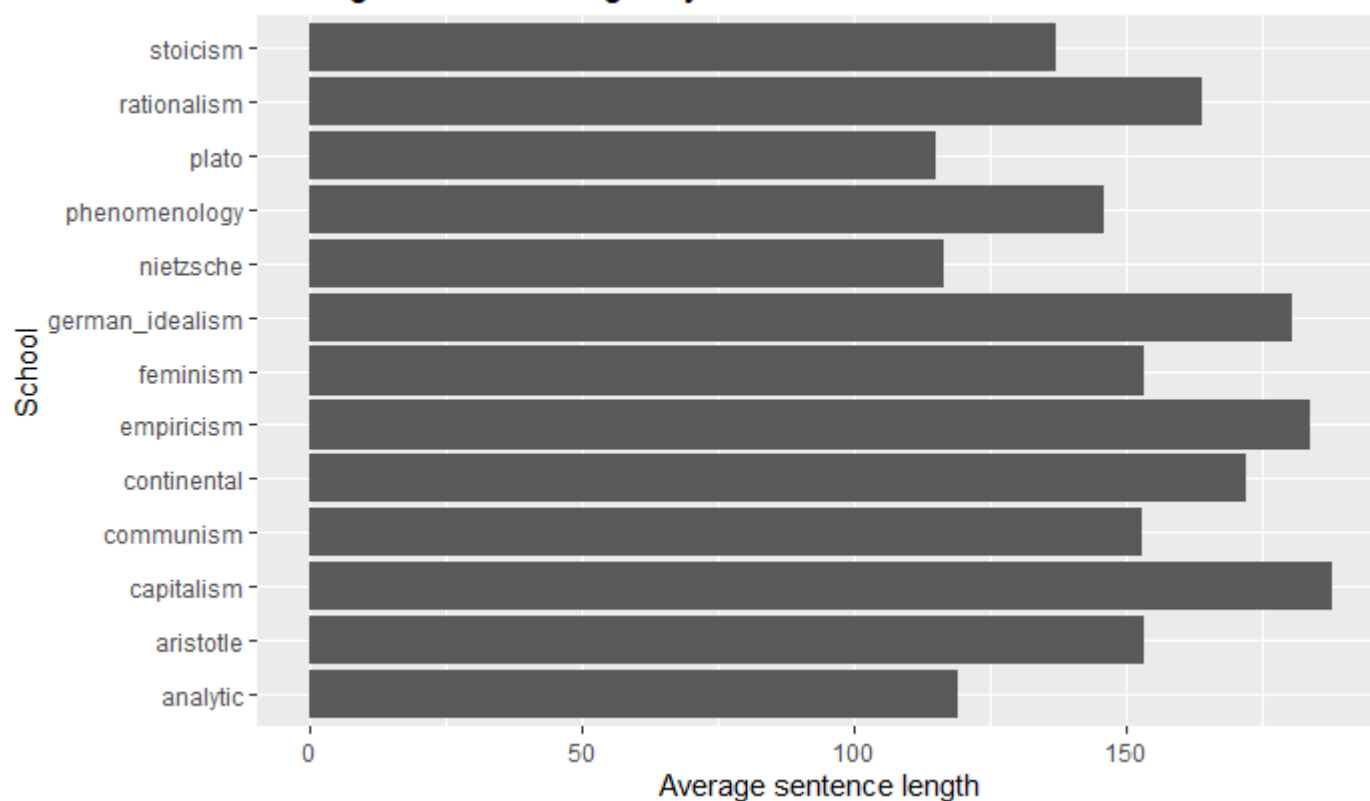
Average sentence length by author



Hide

```
library(ggplot2)
data_avgschool <- data %>%
  group_by(school) %>%
  summarize(avg_sentence_length1 = mean(sentence_length))
ggplot(data_avgschool, aes(x = avg_sentence_length1, y = school)) +
  geom_bar(stat = "identity") +
  ggtitle("Average sentence length by school") +
  xlab("Average sentence length") +
  ylab("School")
```

Average sentence length by school



Hide

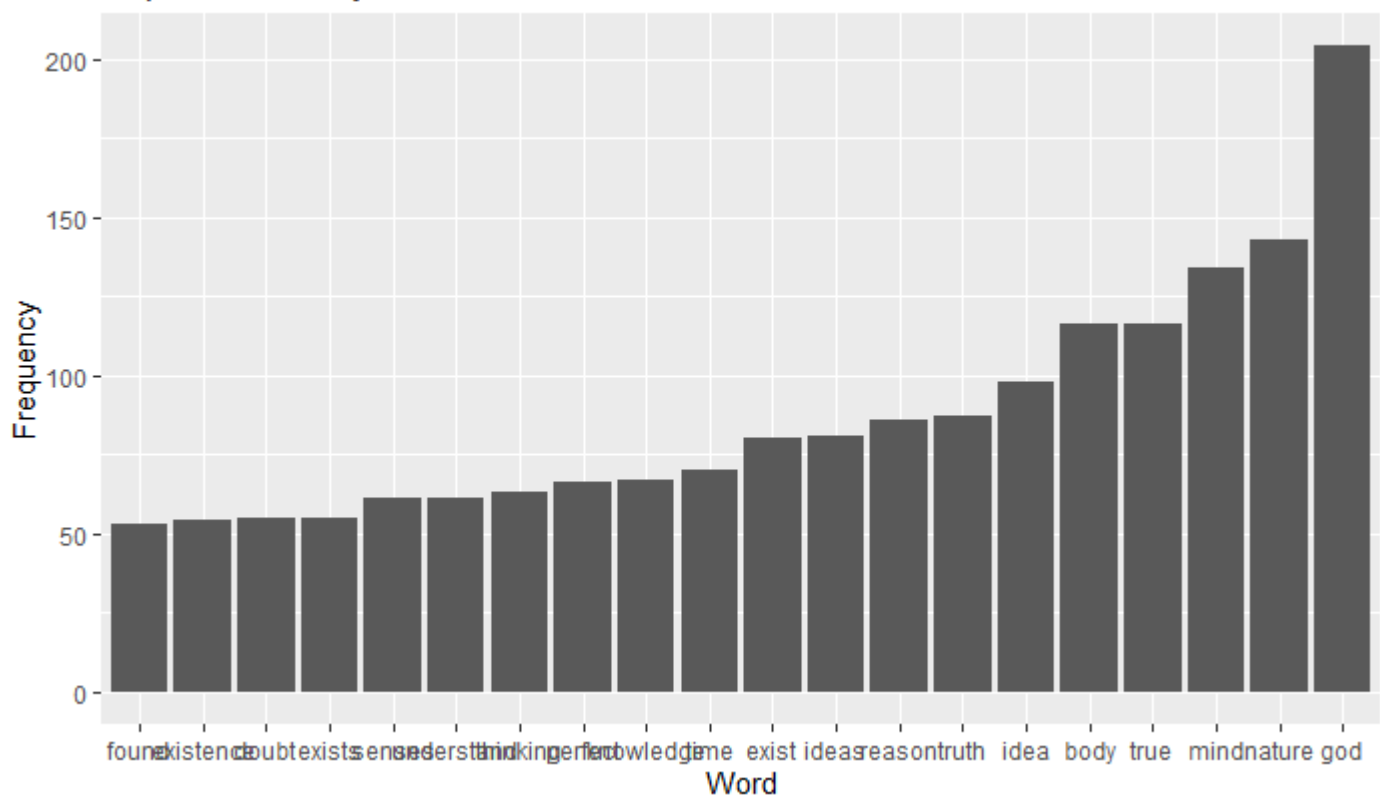
```
descartes <- data %>% filter(author == "Descartes")
descartes_word <- descartes %>% unnest_tokens(word, tokenized_txt) %>% count(word, sort = TRUE)
descartes_word_use <- descartes_word %>%
  anti_join(stop_words)
```

Joining, by = "word"

Hide

```
descartes_top <- head(descartes_word_use, 20)
descartes_bot <- tail(descartes_word_use, 20)
ggplot(descartes_top, aes(x = reorder(word, n), y = n)) +
  geom_col() +
  xlab(NULL) +
  ggtitle("Top 20 words by Descartes") +
  xlab("Word") +
  ylab("Frequency")
```

## Top 20 words by Descartes



Hide

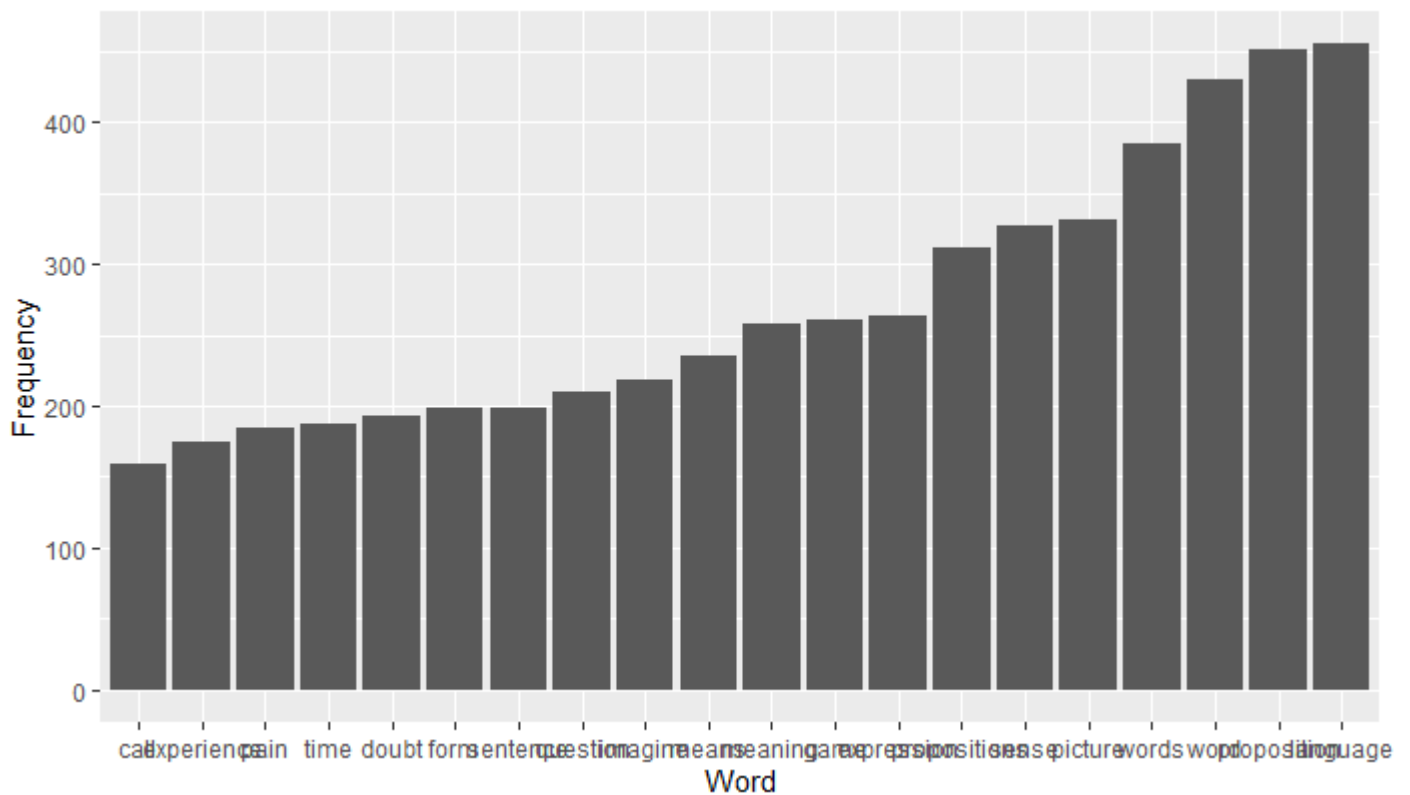
```
wittgenstein <- data %>% filter(author == "Wittgenstein")
wittgenstein_word <- wittgenstein %>% unnest_tokens(word, tokenized_txt) %>% count(word, sort = TRUE)
wittgenstein_word_use <- wittgenstein_word %>%
anti_join(stop_words)
```

Joining, by = "word"

Hide

```
wittgenstein_top <- head(wittgenstein_word_use, 20)
wittgenstein_bot <- tail(wittgenstein_word_use, 20)
ggplot(wittgenstein_top, aes(x = reorder(word, n), y = n)) +
  geom_col() +
  xlab(NULL) +
  ggtitle("Top 20 words by Wittgenstein") +
  xlab("Word") +
  ylab("Frequency")
```

## Top 20 words by Wittgenstein



Hide

```
library(tidytext)
library(textdata)
get_sentiments("nrc")
```

word	sentiment
<chr>	<chr>
abacus	trust
abandon	fear
abandon	negative
abandon	sadness
abandoned	anger
abandoned	fear
abandoned	negative
abandoned	sadness
abandonment	anger
abandonment	fear

1-10 of 13,872 rows

Previous 1 2 3 4 5 6 ... 100 Next

[Hide](#)

```
d_nrc_positive <- get_sentiments("nrc") %>%  
  filter(sentiment == "positive")  
  
data_words %>%  
  filter(author == "Descartes") %>%  
  inner_join(d_nrc_positive) %>%  
  count(word, sort = TRUE)
```

Joining, by = "word"

word	n
<chr>	<int>
god	204
true	116
truth	87
reason	86
knowledge	67
perfect	66
existence	54
found	53
greater	49
perceive	45

1-10 of 519 rows

Previous 1 2 3 4 5 6 ... 52 Next

[Hide](#)

```
d_nrc_negative <- get_sentiments("nrc") %>%  
  filter(sentiment == "negative")  
  
data_words %>%  
  filter(author == "Descartes") %>%  
  inner_join(d_nrc_negative) %>%  
  count(word, sort = TRUE)
```

Joining, by = "word"

word	n
<chr>	<int>
doubt	55

word	n
<chr>	<int>
contrary	34
case	33
error	33
deceived	30
pain	25
difficulty	20
difficulties	18
ignorant	18
possessed	17
1-10 of 366 rows	Previous 1 2 3 4 5 6 ... 37 Next

Hide

```
w_nrc_positive <- get_sentiments("nrc") %>%
  filter(sentiment == "positive")

data_words %>%
  filter(author == "Wittgenstein") %>%
  inner_join(w_nrc_positive) %>%
  count(word,sort = TRUE)
```

Joining, by = "word"

word	n
<chr>	<int>
proposition	451
word	430
sense	327
question	210
kind	167
don	145
logical	142
true	134
truth	134
feeling	108



1-10 of 561 rows

Previous 1 2 3 4 5 6 ... 57 Next

Hide

```
w_nrc_negative <- get_sentiments("nrc") %>%
  filter(sentiment == "negative")

data_words %>%
  filter(author == "Wittgenstein") %>%
  inner_join(w_nrc_negative) %>%
  count(word, sort = TRUE)
```

Joining, by = "word"

word	n
<chr>	<int>
words	385
case	262
sentence	199
doubt	193
pain	185
feeling	108
wrong	99
mistake	92
negation	40
subject	39

1-10 of 511 rows

Previous 1 2 3 4 5 6 ... 52 Next

Hide

```
library(sentimentr)
descartes_data <- data[data$author == "Descartes",]
descartes_data1 <- descartes_data$sentence_spacy
sentiment_scores <- sentiment(descartes_data1)
```

Warning: Each time `sentiment` is run it has to do sentence boundary disambiguation when a raw `character` vector is passed to `text.var`. This may be costly of time and memory. It is highly recommended that the user first runs the raw `character` vector through the `get\_sentences` function.

Hide

```
head(sentiment_scores,20)
```

element_id <int>	sentence_id <int>	word_count <int>	sentiment <dbl>
1	1	166	0.25147277
2	1	49	0.16142857
3	1	94	0.09282791
4	1	19	0.21794495
5	1	46	0.50130265
6	1	52	0.13174130
7	1	17	-0.31529631
8	1	76	0.05505978
9	1	86	0.55315517
10	1	111	0.80797074

1-10 of 20 rows

Previous **1** 2 Next

Hide

```
library(sentimentr)
wittgenstein_data <- data[data$author == "Wittgenstein",]
wittgenstein_data1 <- wittgenstein_data$sentence_spacy
sentiment_scores1 <- sentiment(wittgenstein_data1)
```

Warning: Each time `sentiment` is run it has to do sentence boundary disambiguation when a raw `character` vector is passed to `text.var`. This may be costly of time and memory. It is highly recommended that the user first runs the raw `character` vector through the `get\_sentences` function.

Hide

```
head(sentiment_scores1,20)
```

element_id <int>	sentence_id <int>	word_count <int>	sentiment <dbl>
1	1	17	0.00000000
2	1	16	0.00000000
3	1	18	0.00000000
4	1	7	0.00000000
5	1	9	-0.33333333

element_id <int>	sentence_id <int>	word_count <int>	sentiment <dbl>			
6	1	13	0.00000000			
7	1	52	0.31895261			
8	1	12	0.11547005			
9	1	9	-0.08333333			
10	1	5	-0.11180340			
1-10 of 20 rows			Previous	1	2	Next