philosophy data story

Yunfan LIu yl5111

2023-01-19

Step 0: Load Libraries and Data Set

```
library(dplyr)
library(vistime)
library(lubridate)
library(plotly)
library(tm)
library(SnowballC)
library(wordcloud)
library(RColorBrewer)
library(wordcloud2)
```

If R cannot open the file, please go to README.md under data folder, and download philosophy_data.csv from the link. This happens because the csv file is over 300 MB.

Load Data

```
pdata= read.csv(file="../data/philosophy_data.csv",header=TRUE)
head(pdata)
```

```
##                      title author school
## 1 Plato - Complete Works  Plato  plato
## 2 Plato - Complete Works  Plato  plato
## 3 Plato - Complete Works  Plato  plato
## 4 Plato - Complete Works  Plato  plato
## 5 Plato - Complete Works  Plato  plato
## 6 Plato - Complete Works  Plato  plato
##
## 1  What's new, Socrates, to make you leave your usual haunts in the Lyceum and spend your time here l
## 2                                            Surely you are not prosecuting anyone befo
## 3                                  The Athenians do not call this a prosecution bu
## 4
## 5                Someone must have indicted you, for you are not going to tell me that you l
## 6                                                                    But som
##
## 1  What's new, Socrates, to make you leave your usual haunts in the Lyceum and spend your time here l
## 2                                            Surely you are not prosecuting anyone befo
## 3                                  The Athenians do not call this a prosecution bu
## 4
## 5                Someone must have indicted you, for you are not going to tell me that you l
## 6                                                                    But som
##   original_publication_date corpus_edition_date sentence_length
## 1                      -350                1997             125
## 2                      -350                1997              69
```

```
## 3                         -350          1997        74
## 4                         -350          1997        21
## 5                         -350          1997       101
## 6                         -350          1997        34
##
## 1  what's new, socrates, to make you leave your usual haunts in the lyceum and spend your time here b
## 2                                           surely you are not prosecuting anyone befor
## 3                                   the athenians do not call this a prosecution but
## 4
## 5                  someone must have indicted you, for you are not going to tell me that you
## 6                                                              but some
##
## 1 ['what', 'new', 'socrates', 'to', 'make', 'you', 'leave', 'your', 'usual', 'haunts', 'in', 'the',
## 2                                                         ['surely', 'you
## 3                                                ['the', 'athenia
## 4
## 5                  ['someone', 'must', 'have', 'indicted', 'you', 'for', 'you', 'are',
## 6
##
## 1    what be new , Socrates , to make -PRON- leave -PRON- usual haunt in the Lyceum and spend -PRON-
## 2                                           surely -PRON- be not prosecute anyo
## 3                                   the Athenians do not call this a pro
## 4
## 5                  someone must have indict -PRON- , for -PRON- be not go to tell -PRON
## 6
```

Q1: How many philosophers? Who are they?

Since the data set is big and new to me, and it seems many rows have the same author. First, let's see how many philosophers are there, and who are they?

```
unique(pdata$author)
```

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

```
length(unique(pdata$author))
```

```
## [1] 36
```

So, in our data set, we have 36 different philosophers.

Q2: How many sentences each philosopher said? Who said the most sentences?

Since the data set is about the maxims those philosophers said, I am wondering how many sentences each philosopher said in their life time, and who is the most long-winded philosopher?

First, I build a new data frame which counts the number of sentences each philosopher said, it's called author_count.

Then, I choose to visualize the result in a bar graph. Due to limited space, each philosopher is represented by the first 3 letters of his name.

```
author_count = pdata %>% group_by(author) %>% summarise(total_count=n(),
            .groups = 'drop')
head(author_count)
```

```
## # A tibble: 6 x 2
##    author     total_count
##    <chr>            <int>
## 1 Aristotle        48779
## 2 Beauvoir         13017
## 3 Berkeley          2734
## 4 Davis             3059
## 5 Deleuze          12540
## 6 Derrida           5999
```

```
author_3=list()
for (i in author_count$author){
  author_3=append(author_3,substr(toString(i),1,3))}
```

```
barplot(author_count$total_count,col=rainbow(36),width=4,
        names.arg=author_3,
        cex.names=0.25,space=4, cex.axis=0.7)
```
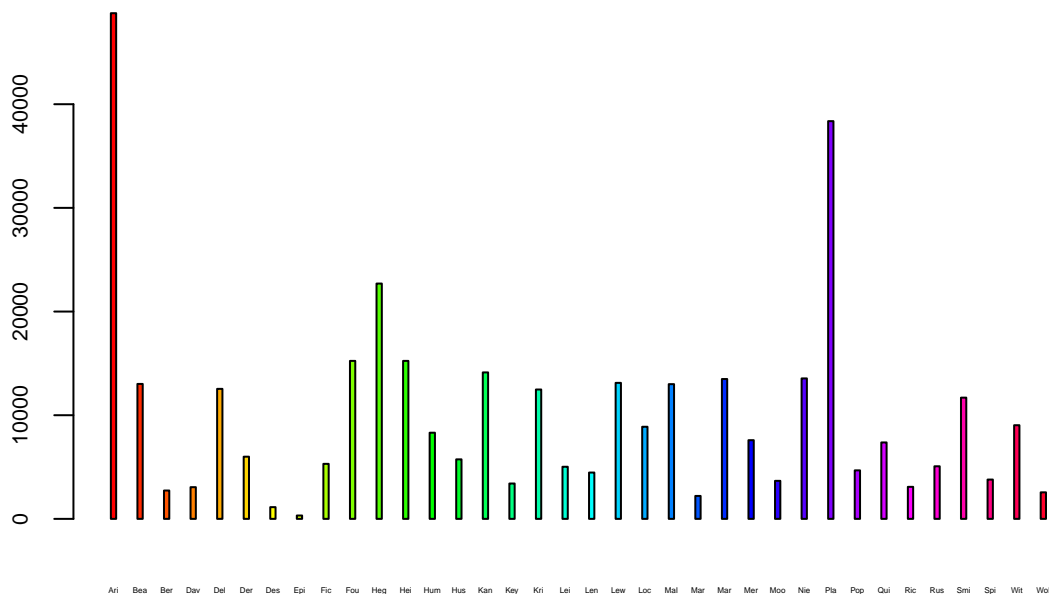

Aristotle said the most maxims, with over 40000 sentences.

Q3: How many shcools do they have? Which school has the most philosophers supported?

I noticed that the data set also provides school information. I am wondering how many schools are there in this data set, and which school is the most powerful?

I start to build a data frame called school_count, which counts the number of philosophers in each school.

I choose to visualize the result in a pie chart.

```
unique(pdata$school)
```

```
##  [1] "plato"          "aristotle"      "empiricism"     "rationalism"
##  [5] "analytic"       "continental"    "phenomenology"  "german_idealism"
##  [9] "communism"      "capitalism"     "stoicism"       "nietzsche"
## [13] "feminism"
```

```
school_count = pdata %>% group_by(school) %>% summarise(
  num_of_author=n_distinct(author),.groups = "drop")
head(school_count)
```

```
## # A tibble: 6 x 2
##   school      num_of_author
##   <chr>               <int>
## 1 analytic                7
## 2 aristotle               1
## 3 capitalism              3
## 4 communism               2
## 5 continental             3
## 6 empiricism              3
```

```
pct <- round(school_count$num_of_author/sum(school_count$num_of_author)*100)
lables <- school_count$school
lables = paste(lables, pct)
lables <- paste(lables,"%",sep="")
pie(school_count$num_of_author, labels =lables, main="PIE FOR SCHOOL",cex=0.9)
```

PIE FOR SCHOOL
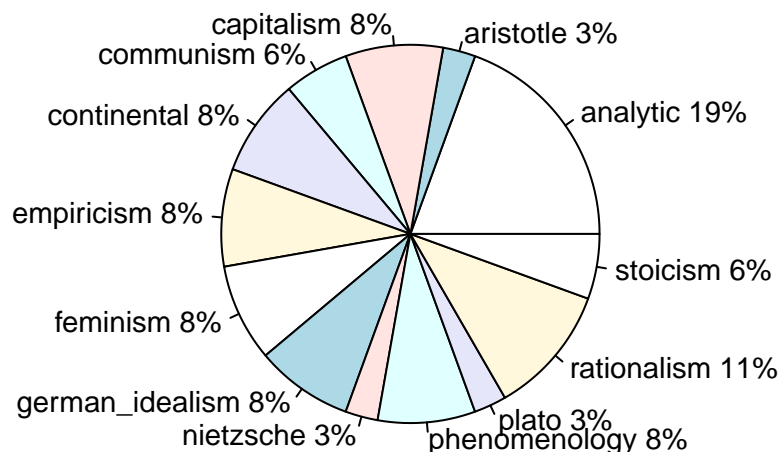
It is clear that most philosophers are in analytic school.

Q4: How did schools develop?

Since I explored the schools, and found out their appeal towards philosophers. I think one reason for the difference in number of supporters is time, for example, some schools developed earlier. So, i want to show a time line about philosophy school development.

```
school_time = pdata %>% group_by(school) %>% summarise(original_publication_date,
          .groups = 'drop')
```

```
school_time= school_time %>% distinct()
school_time= school_time %>% arrange(original_publication_date)
head(school_time)

## # A tibble: 6 x 2
##   school      original_publication_date
##   <chr>                           <int>
## 1 plato                            -350
## 2 aristotle                        -320
## 3 stoicism                          125
## 4 stoicism                          170
## 5 rationalism                      1637
## 6 rationalism                      1641
## transfer integer year into date type
start_i=c(125,1637,1689,1776,1781,1792,1848,1886,1907,1910,1961)

start_1=as.Date(ISOdate(start_i, 1, 1))
a= c(make_date(year = -350L),make_date(year = -320L))

start=c(a,start_1)

end_i= c(170,1710,1779,1936,1820,1981,1883,1888,1950,1985,1972)
end_1= as.Date(ISOdate(end_i, 1, 1))
b= c(make_date(year = -350L),make_date(year = -320L))
end= c(b,end_1)

timedata= data.frame(event= unique(school_time$school),
                     start= start,
                     end= end)
gg_vistime(timedata,
           optimize_y = FALSE,
           linewidth = 8,
           show_labels=TRUE,
           background_lines=3,
           title="Philosophy Development"
           )
```

Philosophy Development
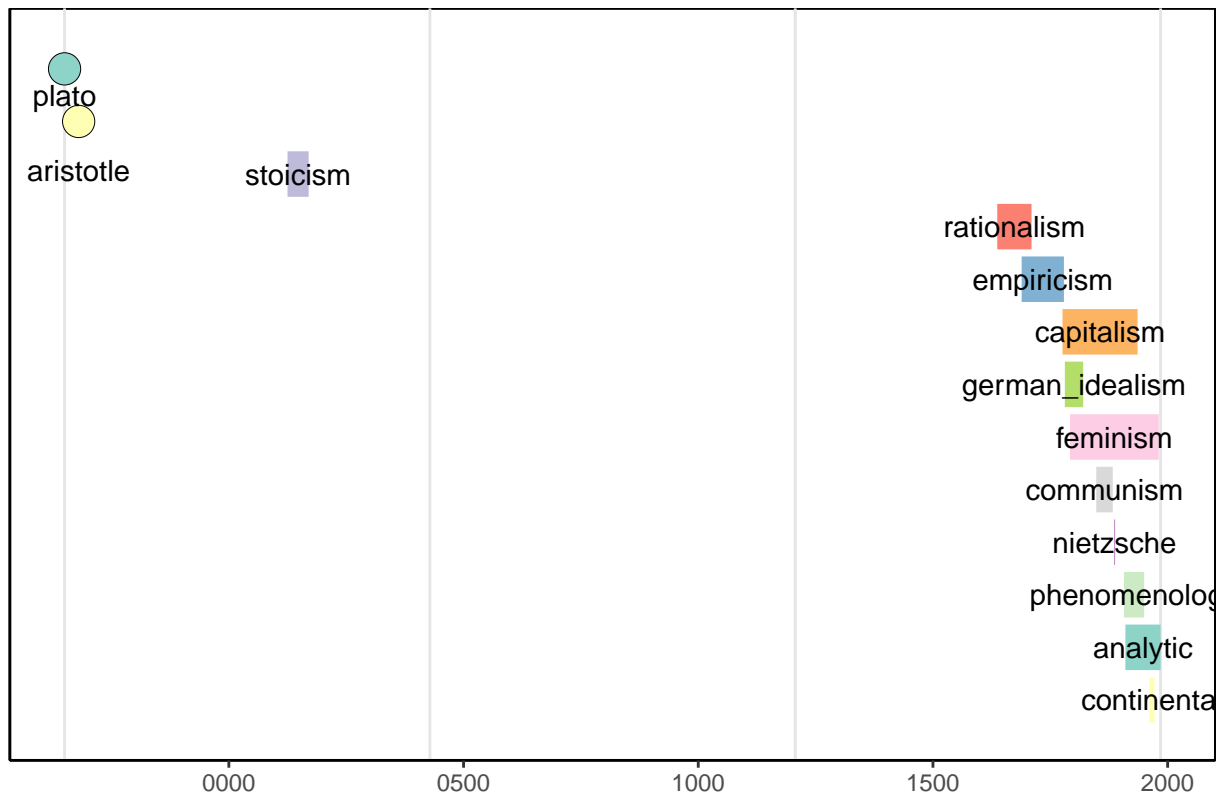
Although the data set says all Plato and Aristotle sentences were in 350 and 320 B.C., it does not mean that the 2 schools ended and there were no philosophy in medieval Age.I did some research and found out that Plato and Aristotle were the two leading influences on medieval thought.

Q5: What is the general topic for all philosophers? For each school?

I would like to know what is the biggest topic of philosophy from ancient times to modern times. I remove the pointless words like "one, can, will, things. . . "

Actually, before i remove the pointless words, guess what is the most common word in philosophy? I thought it would be "the", or "is, are", but it is "one". Philosophers love to use "one" as subject, for example, they would say "One is . . . ". I am curious about how many sentences have "one".

How many maxims have "one"?

```
count=0
for (i in pdata$sentence_lowered){
  if (grepl("one",i,fixed=TRUE)){
    count=count+1
  }
}
(count/nrow(pdata))*100
```

```
## [1] 16.95722
```

16.95722% of maxims have "one".

Common Word ALL Philosophers

```r
text_all = removePunctuation(pdata$sentence_lowered)

text_all= gsub("one","",text_all)
text_all= gsub("can",'',text_all)
text_all= gsub("must",'',text_all)
text_all= gsub("will",'',text_all)
text_all= gsub("may",'',text_all)
text_all= gsub("things",'',text_all)
text_all= gsub("thing",'',text_all)
text_all= gsub("also",'',text_all)
text_all= gsub("even",'',text_all)
text_all= gsub("therefore",'',text_all)
text_all= gsub("say",'',text_all)
text_all= gsub("said",'',text_all)
text_all= gsub("just",'',text_all)
text_all= gsub("man",'',text_all)
text_all= gsub("every",'',text_all)

wordcloud(text_all, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```

By School

Analytic

```
df_anal =pdata[pdata$school =='analytic',]

text_anal = removePunctuation(df_anal$sentence_lowered)
text_anal= gsub("one","",text_anal)
text_anal= gsub("can",'',text_anal)
text_anal= gsub("must",'',text_anal)
text_anal= gsub("will",'',text_anal)
text_anal= gsub("may",'',text_anal)
text_anal= gsub("things",'',text_anal)
text_anal= gsub("thing",'',text_anal)
text_anal= gsub("also",'',text_anal)
text_anal= gsub("even",'',text_anal)
text_anal= gsub("therefore",'',text_anal)
text_anal= gsub("say",'',text_anal)
text_anal= gsub("said",'',text_anal)
text_anal= gsub("just",'',text_anal)
text_anal= gsub("man",'',text_anal)
text_anal= gsub("every",'',text_anal)


wordcloud(text_anal, scale=c(3,0.4), max.words=250, random.order=FALSE,
          rot.per=0.5
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```

Plato

```r
df_plato =pdata[pdata$school =='plato',]

text_plato = removePunctuation(df_plato$sentence_lowered)

text_plato= gsub("one","",text_plato)
text_plato= gsub("can",'',text_plato)
text_plato= gsub("must",'',text_plato)
text_plato= gsub("will",'',text_plato)
text_plato= gsub("may",'',text_plato)
text_plato= gsub("things",'',text_plato)
text_plato= gsub("thing",'',text_plato)
text_plato= gsub("also",'',text_plato)
text_plato= gsub("even",'',text_plato)
text_plato= gsub("therefore",'',text_plato)
text_plato= gsub("say",'',text_plato)
text_plato= gsub("said",'',text_plato)
text_plato= gsub("just",'',text_plato)
text_plato= gsub("man",'',text_plato)
text_plato= gsub("every",'',text_plato)

wordcloud(text_plato, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```


Aristotle

```r
df_aristotle =pdata[pdata$school == 'aristotle',]

text_aristotle = removePunctuation(df_aristotle$sentence_lowered)

text_aristotle= gsub("one","",text_aristotle)
text_aristotle= gsub("can",'',text_aristotle)
text_aristotle= gsub("must",'',text_aristotle)
text_aristotle= gsub("will",'',text_aristotle)
text_aristotle= gsub("may",'',text_aristotle)
text_aristotle= gsub("things",'',text_aristotle)
text_aristotle= gsub("thing",'',text_aristotle)
text_aristotle= gsub("also",'',text_aristotle)
text_aristotle= gsub("even",'',text_aristotle)
text_aristotle= gsub("therefore",'',text_aristotle)
text_aristotle= gsub("say",'',text_aristotle)
text_aristotle= gsub("said",'',text_aristotle)
text_aristotle= gsub("just",'',text_aristotle)
text_aristotle= gsub("man",'',text_aristotle)
text_aristotle= gsub("every",'',text_aristotle)


wordcloud(text_aristotle, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```


Capitalism

```
df_capitalism =pdata[pdata$school == 'capitalism',]

text_capitalism = removePunctuation(df_aristotle$sentence_lowered)

text_capitalism= gsub("one","",text_capitalism)
text_capitalism= gsub("can",'',text_capitalism)
text_capitalism= gsub("must",'',text_capitalism)
text_capitalism= gsub("will",'',text_capitalism)
text_capitalism= gsub("may",'',text_capitalism)
text_capitalism= gsub("things",'',text_capitalism)
text_capitalism= gsub("thing",'',text_capitalism)
text_capitalism= gsub("also",'',text_capitalism)
text_capitalism= gsub("even",'',text_capitalism)
text_capitalism= gsub("therefore",'',text_capitalism)
text_capitalism= gsub("say",'',text_capitalism)
text_capitalism= gsub("said",'',text_capitalism)
text_capitalism= gsub("just",'',text_capitalism)
text_capitalism= gsub("man",'',text_capitalism)
text_capitalism= gsub("every",'',text_capitalism)


wordcloud(text_capitalism, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```


German_idealism

```r
df_ger =pdata[pdata$school == 'german_idealism',]

text_ger = removePunctuation(df_ger$sentence_lowered)

text_ger= gsub("one","",text_ger)
text_ger= gsub("can",'',text_ger)
text_ger= gsub("must",'',text_ger)
text_ger= gsub("will",'',text_ger)
text_ger= gsub("may",'',text_ger)
text_ger= gsub("things",'',text_ger)
text_ger= gsub("thing",'',text_ger)
text_ger= gsub("also",'',text_ger)
text_ger= gsub("even",'',text_ger)
text_ger= gsub("therefore",'',text_ger)
text_ger= gsub("say",'',text_ger)
text_ger= gsub("said",'',text_ger)
text_ger= gsub("just",'',text_ger)
text_ger= gsub("man",'',text_ger)
text_ger= gsub("every",'',text_ger)


wordcloud(text_ger, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```


Continental

```
df_con =pdata[pdata$school == 'continental',]

text_con = removePunctuation(df_con$sentence_lowered)

text_con= gsub("one","",text_con)
text_con= gsub("can",'',text_con)
text_con= gsub("must",'',text_con)
text_con= gsub("will",'',text_con)
text_con= gsub("may",'',text_con)
text_con= gsub("things",'',text_con)
text_con= gsub("thing",'',text_con)
text_con= gsub("also",'',text_con)
text_con= gsub("even",'',text_con)
text_con= gsub("therefore",'',text_con)
text_con= gsub("say",'',text_con)
text_con= gsub("said",'',text_con)
text_con= gsub("just",'',text_con)
text_con= gsub("man",'',text_con)
text_con= gsub("every",'',text_con)


wordcloud(text_con, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```


Empiricism

```
df_emp =pdata[pdata$school == 'empiricism',]

text_emp = removePunctuation(df_emp$sentence_lowered)

text_emp= gsub("one","",text_emp)
text_emp= gsub("can",'',text_emp)
text_emp= gsub("must",'',text_emp)
text_emp= gsub("will",'',text_emp)
text_emp= gsub("may",'',text_emp)
text_emp= gsub("things",'',text_emp)
text_emp= gsub("thing",'',text_emp)
text_emp= gsub("also",'',text_emp)
text_emp= gsub("even",'',text_emp)
text_emp= gsub("therefore",'',text_emp)
text_emp= gsub("say",'',text_emp)
text_emp= gsub("said",'',text_emp)
text_emp= gsub("just",'',text_emp)
text_emp= gsub("man",'',text_emp)
text_emp= gsub("every",'',text_emp)


wordcloud(text_emp, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```
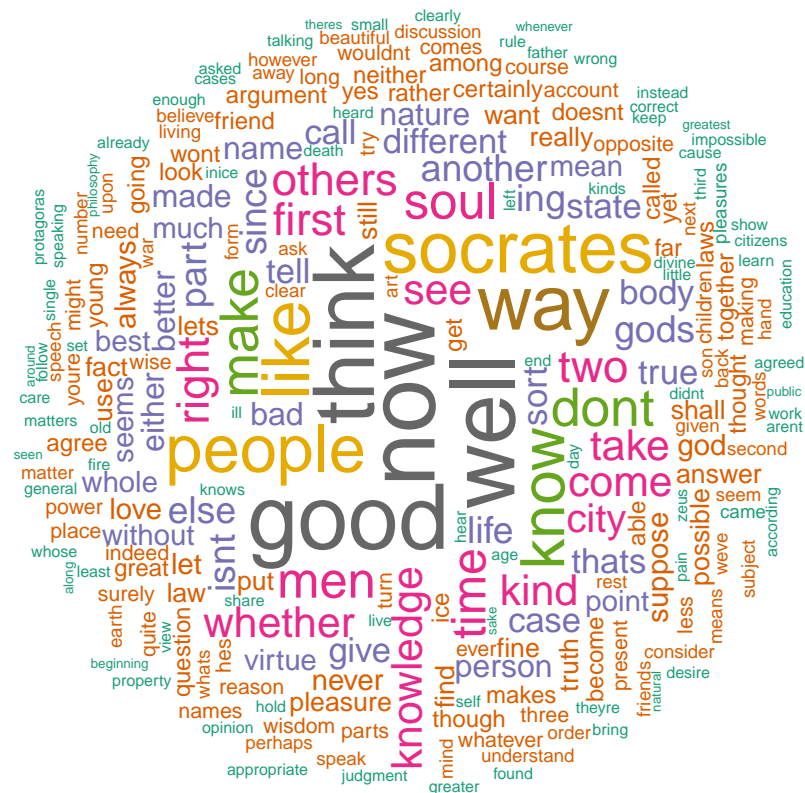

Rationalism

```
df_rat =pdata[pdata$school == 'rationalism',]

text_rat = removePunctuation(df_rat$sentence_lowered)

text_rat= gsub("one","",text_rat)
```

```
text_rat= gsub("can",'',text_rat)
text_rat= gsub("must",'',text_rat)
text_rat= gsub("will",'',text_rat)
text_rat= gsub("may",'',text_rat)
text_rat= gsub("things",'',text_rat)
text_rat= gsub("thing",'',text_rat)
text_rat= gsub("also",'',text_rat)
text_rat= gsub("even",'',text_rat)
text_rat= gsub("therefore",'',text_rat)
text_rat= gsub("say",'',text_rat)
text_rat= gsub("said",'',text_rat)
text_rat= gsub("just",'',text_rat)
text_rat= gsub("man",'',text_rat)
text_rat= gsub("every",'',text_rat)


wordcloud(text_rat, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
      , use.r.layout=TRUE
      , colors=brewer.pal(8, "Dark2"))
```


Phenomenology

```
df_phe =pdata[pdata$school == 'phenomenology',]

text_phe = removePunctuation(df_phe$sentence_lowered)

text_phe= gsub("one","",text_phe)
text_phe= gsub("can",'',text_phe)
text_phe= gsub("must",'',text_phe)
text_phe= gsub("will",'',text_phe)
text_phe= gsub("may",'',text_phe)
text_phe= gsub("things",'',text_phe)
text_phe= gsub("thing",'',text_phe)
```

```r
text_phe= gsub("also",'',text_phe)
text_phe= gsub("even",'',text_phe)
text_phe= gsub("therefore",'',text_phe)
text_phe= gsub("say",'',text_phe)
text_phe= gsub("said",'',text_phe)
text_phe= gsub("just",'',text_phe)
text_phe= gsub("man",'',text_phe)
text_phe= gsub("every",'',text_phe)


wordcloud(text_phe, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```
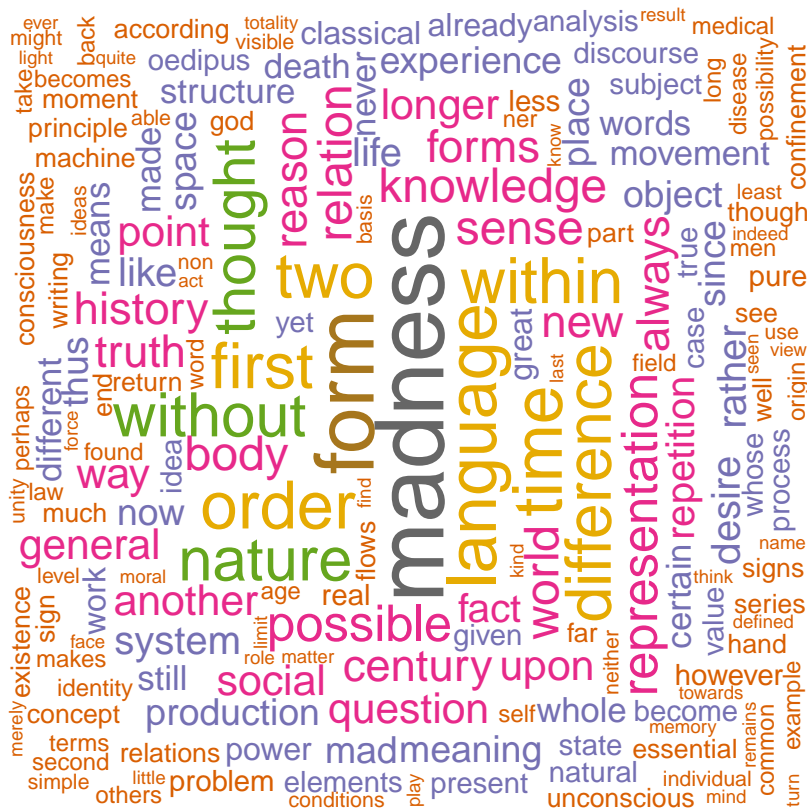

Communism

```r
df_comm =pdata[pdata$school == 'communism',]

text_comm = removePunctuation(df_comm$sentence_lowered)

text_comm= gsub("one","",text_comm)
text_comm= gsub("can",'',text_comm)
text_comm= gsub("must",'',text_comm)
text_comm= gsub("will",'',text_comm)
text_comm= gsub("may",'',text_comm)
text_comm= gsub("things",'',text_comm)
text_comm= gsub("thing",'',text_comm)
text_comm= gsub("also",'',text_comm)
text_comm= gsub("even",'',text_comm)
text_comm= gsub("therefore",'',text_comm)
text_comm= gsub("say",'',text_comm)
text_comm= gsub("said",'',text_comm)
text_comm= gsub("just",'',text_comm)
```

```r
text_comm= gsub("man",'',text_comm)
text_comm= gsub("every",'',text_comm)


wordcloud(text_comm, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```


Stoicism

```r
df_sto =pdata[pdata$school == 'stoicism',]

text_sto = removePunctuation(df_sto$sentence_lowered)

text_sto= gsub("one","",text_sto)
text_sto= gsub("can",'',text_sto)
text_sto= gsub("must",'',text_sto)
text_sto= gsub("will",'',text_sto)
text_sto= gsub("may",'',text_sto)
text_sto= gsub("things",'',text_sto)
text_sto= gsub("thing",'',text_sto)
text_sto= gsub("also",'',text_sto)
text_sto= gsub("even",'',text_sto)
text_sto= gsub("therefore",'',text_sto)
text_sto= gsub("say",'',text_sto)
text_sto= gsub("said",'',text_sto)
text_sto= gsub("just",'',text_sto)
text_sto= gsub("man",'',text_sto)
text_sto= gsub("every",'',text_sto)


wordcloud(text_sto, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```
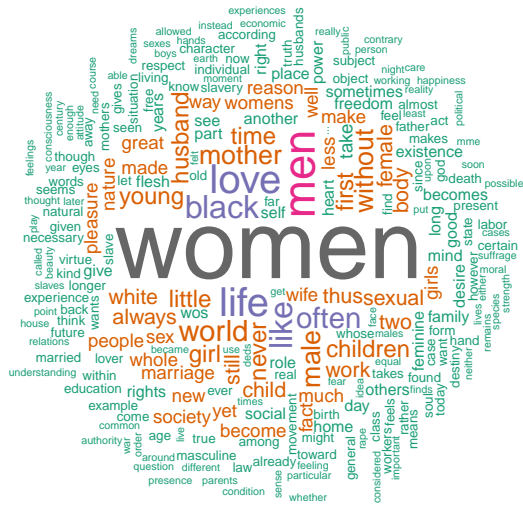
- thou means you in old days.

Nietzsche

```
df_nie =pdata[pdata$school == 'nietzsche',]

text_nie = removePunctuation(df_nie$sentence_lowered)

text_nie= gsub("one","",text_nie)
text_nie= gsub("can",'',text_nie)
text_nie= gsub("must",'',text_nie)
text_nie= gsub("will",'',text_nie)
text_nie= gsub("may",'',text_nie)
text_nie= gsub("things",'',text_nie)
text_nie= gsub("thing",'',text_nie)
text_nie= gsub("also",'',text_nie)
text_nie= gsub("even",'',text_nie)
text_nie= gsub("therefore",'',text_nie)
text_nie= gsub("say",'',text_nie)
text_nie= gsub("said",'',text_nie)
text_nie= gsub("just",'',text_nie)
text_nie= gsub("man",'',text_nie)
text_nie= gsub("every",'',text_nie)


wordcloud(text_nie, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```

- Zarathustra is also known as Zoroaster. He is said to have been an Iranian prophet.

Feminism

```
df_fem =pdata[pdata$school == 'feminism',]

text_fem = removePunctuation(df_fem$sentence_lowered)

text_fem= gsub("one","",text_fem)
text_fem= gsub("can",'',text_fem)
text_fem= gsub("must",'',text_fem)
text_fem= gsub("will",'',text_fem)
text_fem= gsub("may",'',text_fem)
text_fem= gsub("things",'',text_fem)
text_fem= gsub("thing",'',text_fem)
text_fem= gsub("also",'',text_fem)
text_fem= gsub("even",'',text_fem)
text_fem= gsub("therefore",'',text_fem)
text_fem= gsub("say",'',text_fem)
text_fem= gsub("said",'',text_fem)
text_fem= gsub("just",'',text_fem)
text_fem= gsub("man",'',text_fem)
text_fem= gsub("every",'',text_fem)


wordcloud(text_fem, scale=c(3,0.2), max.words=250, random.order=FALSE,
          rot.per=0.35
        , use.r.layout=TRUE
        , colors=brewer.pal(8, "Dark2"))
```

Conclusion

1. Aristotle said the largest amount of maxims, with over 40000 sentences.
2. Analytic has 19% of philosophers, other schools have almost the same amount of supporters.