

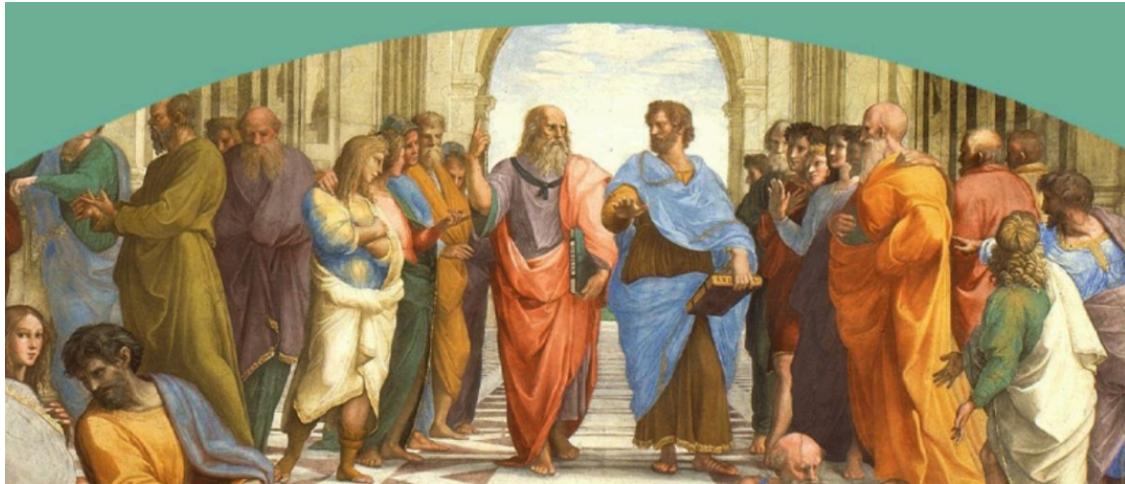
# project 1 story without code

September 21, 2022

```
[1]: from IPython.display import Image, display
```

```
[2]: Image(filename='/Users/lsx/Desktop/title.png')
```

```
[2]:
```



## 1 How did this philosophy evolved through the years?

## 2 How Hard is it to read Philosophy?

Shuangxian Li

### 2.1 Introduction

Literally, the term “philosophy” means, “love of wisdom.” In a broad sense, the study of the fundamental nature of knowledge, reality, and existence, especially when considered as an academic discipline. Since as philosophy school develops, human society also have significant change. It’s worthwhile to study the change of philosophy during the past years. Then, we will answer these following questions which help us analyze our process.

- What is the theme of these philosophy schools?
- Are there some similarities among different philosophy schools?

Additionally, since many people are not familiar with philosophy schools and texts, it's meaningful for us to do some difficulty sorting in order to select the appropriate books for readers. Therefore we can also answer other questions below: - What are the readabilities of philosophy texts? - if the text I want to read is not on among those texts, how do I know whether they are difficult to read?

## 2.2 Data

The data was taken from Kaggle database <https://www.kaggle.com/kouroshhalizadeh/history-of-philosophy>. It contains over 300,000 sentences from over 50 texts spanning 13 major schools of philosophy. The represented schools are: Plato, Aristotle, Empiricism, Rationalism, Analytic Philosophy, Continental Philosophy, Phenomenology, German Idealism, Communism, Capitalism, Stoicism, Nietzsche, Feminism.

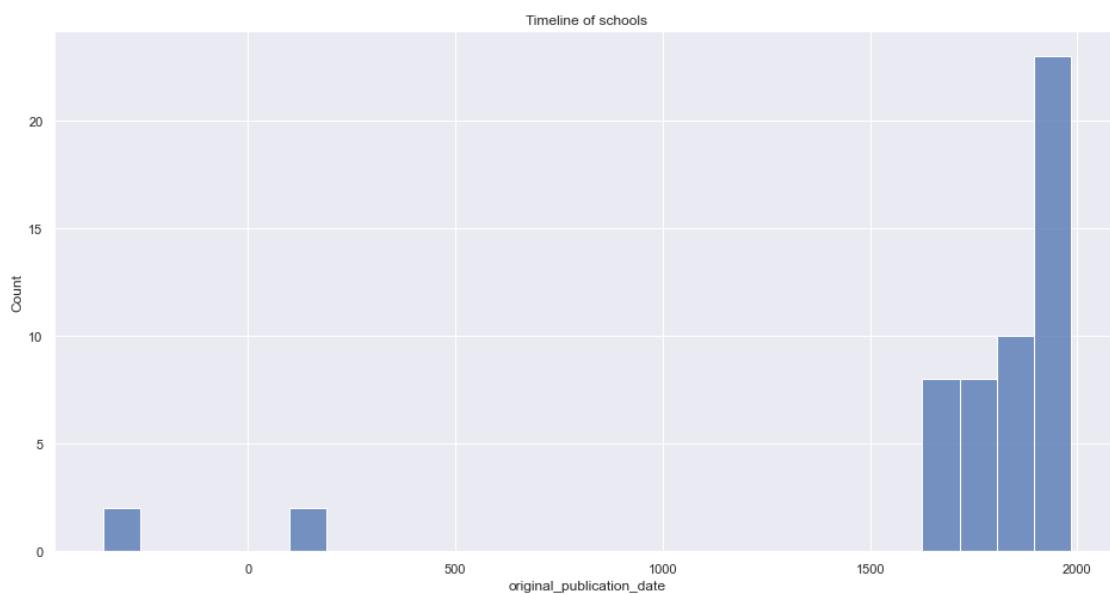
## 2.3 Data exploration

The data and plots show that the first philosophical works are from 350BC, when the Greek philosopher Plato founded Plato school. The era represents the rapid development of Greek civilization. However, there is an obvious decrease in the number of philosophical publications during the period from 500 - 1500(known as the Middle Ages). That was an era of Population decline, counterurbanisation, the collapse of centralized authority, invasions, and mass migrations of tribes, which had begun in Late Antiquity, continued into the Early Middle Ages. It is logical for the decline of philosophy. During the 1700s, the emergence of the Renaissance prompted philosophy works to appear again. Additionally, philosophy schools developed in succession, which means our philosophical thoughts indeed evolve as the change of world society.

Moreover, when I deep-dive into each schools, it can be seen that Empiricism and Rationalism are longest lasting school of thoughts. It is so interesting that both two thoughts are opposite.

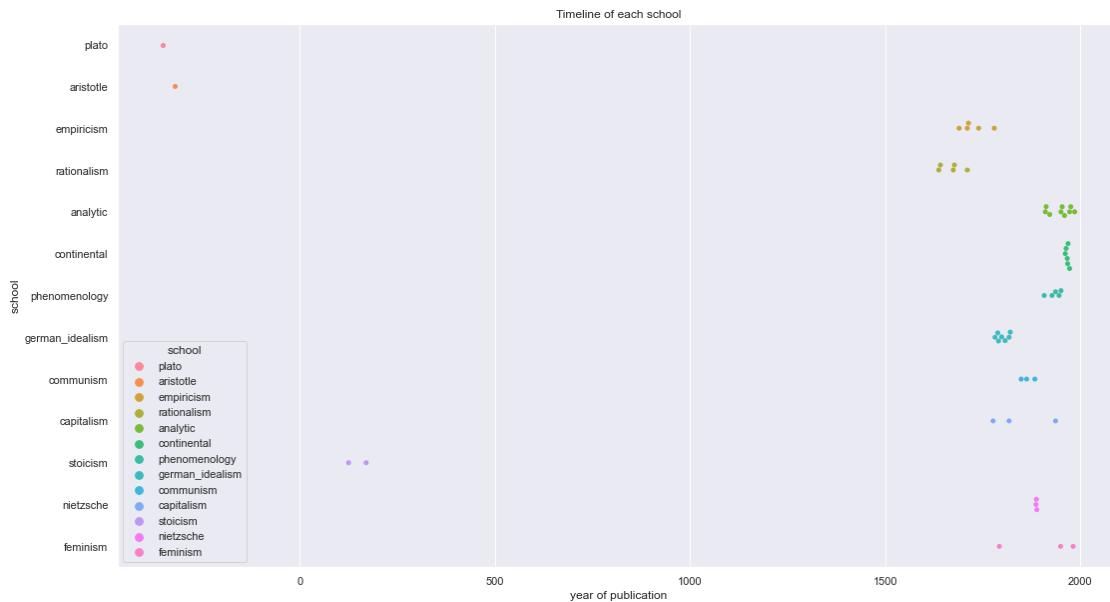
```
[3] : Image(filename='/Users/lsx/Desktop/Timeline of schools.png')
```

```
[3] :
```



```
[4]: Image(filename='/Users/lsx/Desktop/Timeline of schools2.png')
```

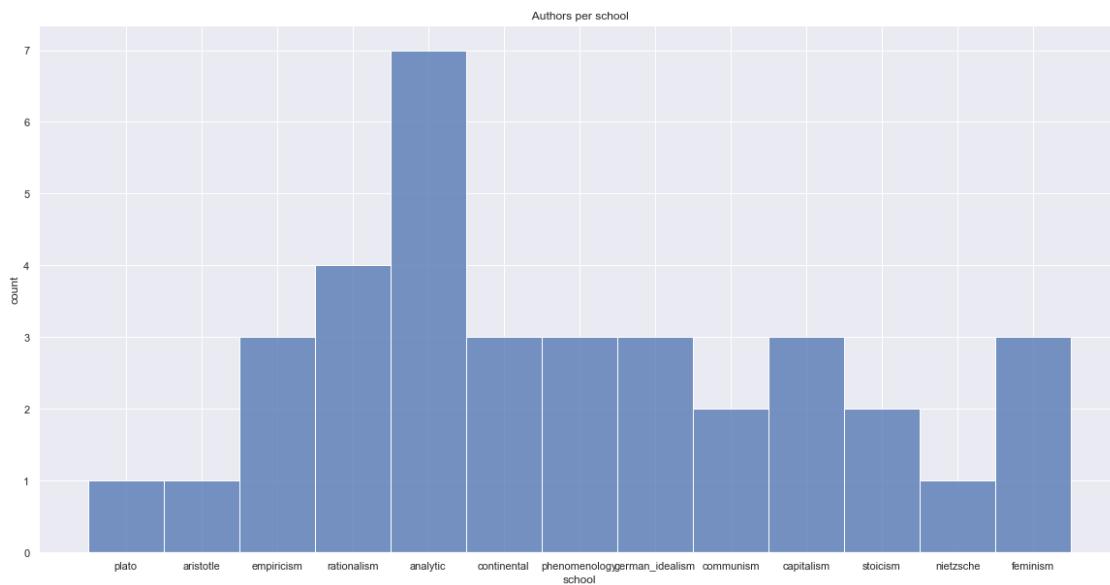
[4]:



Let's have look at some details about philosophical schools and authors. The plots below indicates that the number of works and authors for each school separately. Obviously, no school we have available more than 11 books and some schools indeed are represented from one book which also written by only one author except from nietzsche. It seems make sense that Plato and Aristotle schools represent theories of the two authors themselves separately.

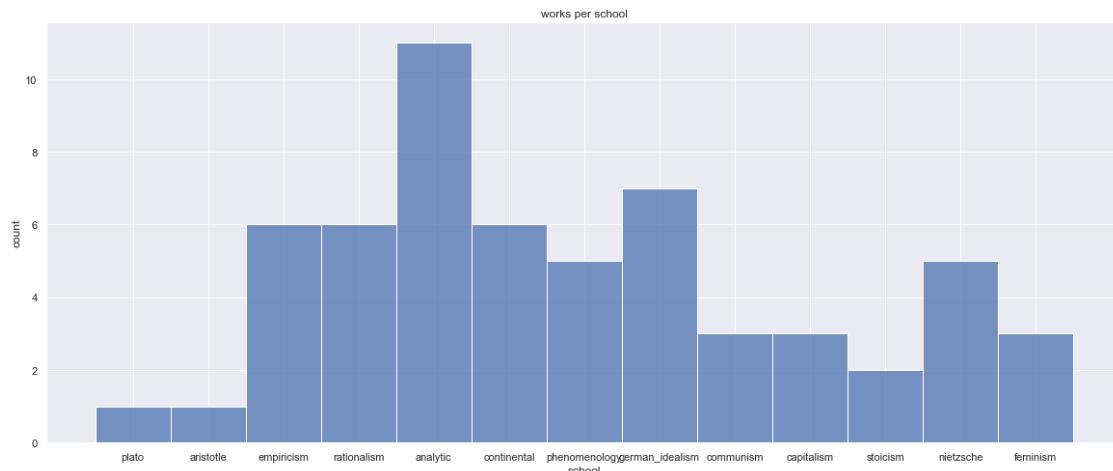
```
[5]: Image(filename='/Users/lsx/Desktop/Authors per school.png')
```

[5]:



```
[6] : Image(filename='/Users/lsx/Desktop/works per school.png')
```

```
[6] :
```



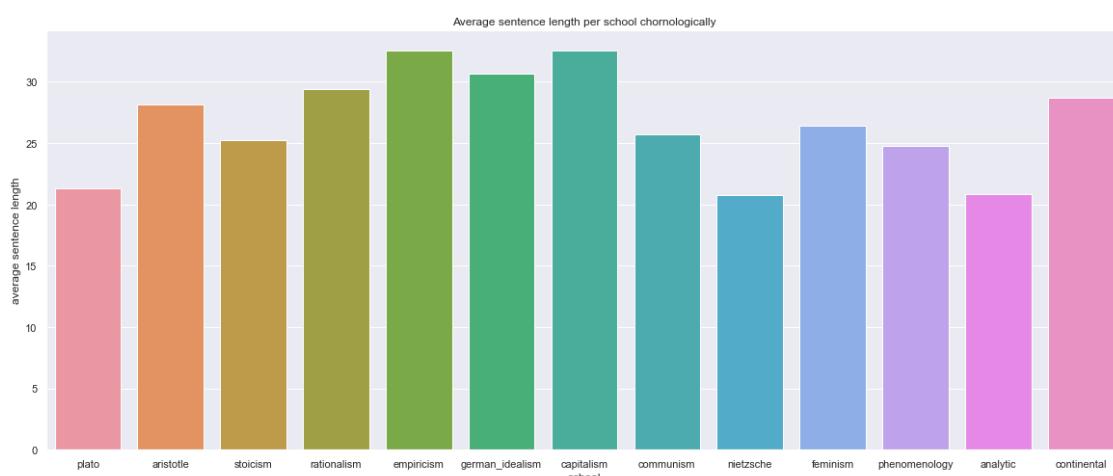
How does the trend of philosophy evolution shown? Let's analyze average sentence lengths of books for each school.

As shown in the following graphs, the size of the bullets in the second picture shows the size of sentences in the dataset for each school. Specifically, the most ancient philosophy school is Plato which highly focus on small average sentence length in books. In addition, the opposite philosophy schools, Empiricism and Rationalism, have the largest average sentence length based on very difficult texts and more knowledge. According to the general trend, ancient philosophers initially seemed to use short and simple sentences in their philosophical texts and sentences became longer and more complex as time passed. Then, the recent philosophers begin to focus on laconism, so short and simple sentences in books made their appearance again.

However, this hypothesis is challenged by the existance of two outliers, namely schools of aristotle and continental. It seems hard to get a specific relationship between time and laconism since some influence of other factors like punctuation in texts.

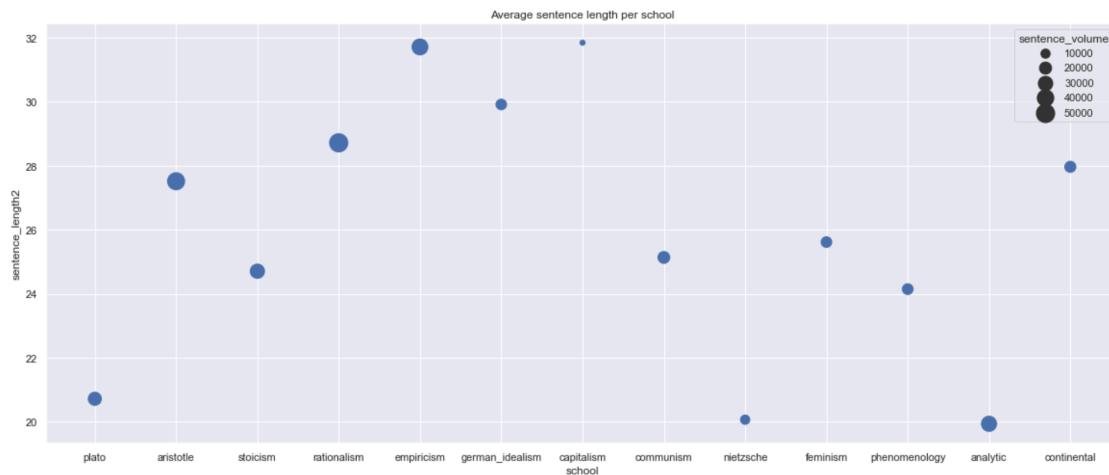
```
[7] : Image(filename='/Users/lsx/Desktop/Average sentence length per school.png')
```

```
[7] :
```



```
[33]: Image(filename='/Users/lsx/Desktop/avg_sentence.png')
```

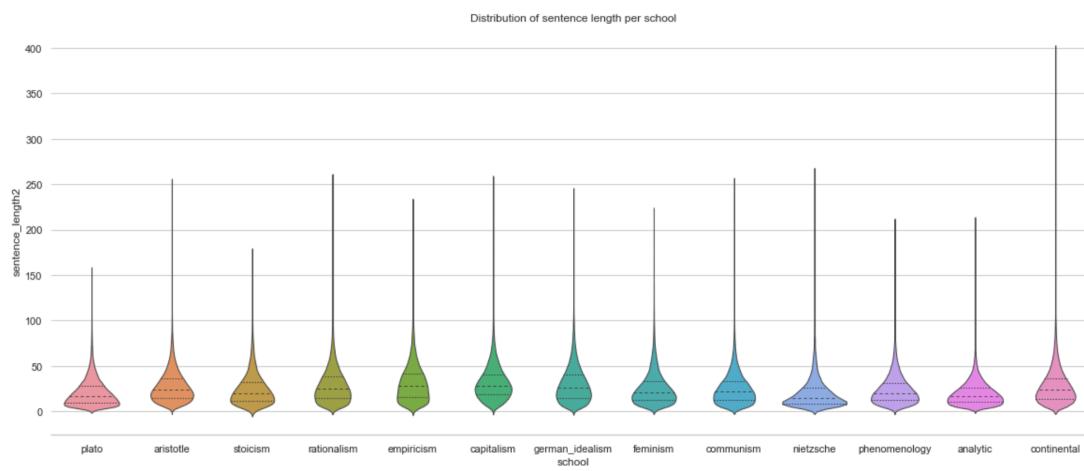
[33]:



Let's try distribution plot of sentence length per school and this plot is more easy to be understand. It is obvious that there exist some sentences in the continental school is much higher than others. We guess the effect of punctuation may be important.

```
[34]: Image(filename='/Users/lsx/Desktop/Sentence_length_distribution.png')
```

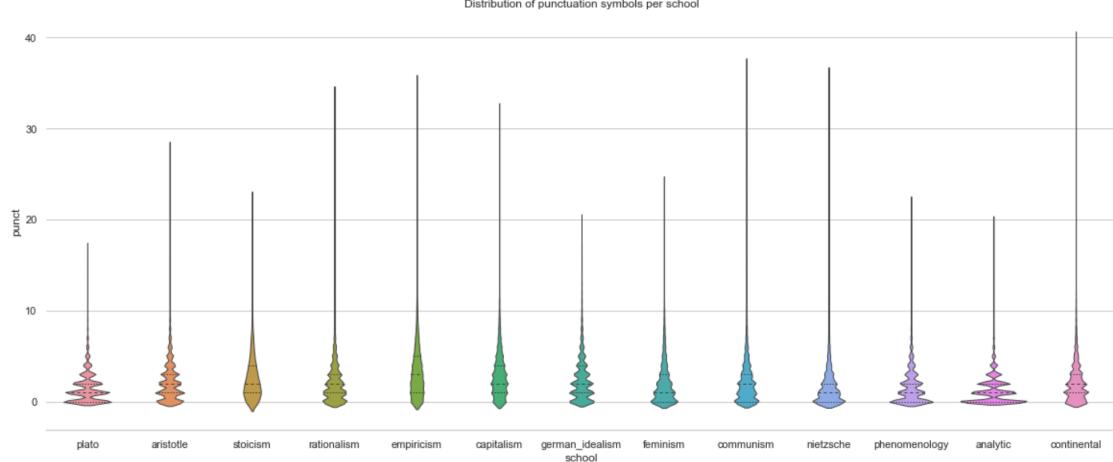
[34]:



Next, the plot below shows the distribution of ‘,’ and ‘;’ punctuations symbols per school. It indicates that the schools have the more sentences and also have more ‘,’ and ‘;’ punctuations symbols.

```
[35]: Image(filename='/Users/lsx/Desktop/punctuation_distribution.png')
```

[35] :



## 2.4 Data processing

Data cleaning and transformation are essential for our following analysis and answer the questions we asked before. I have listed below some commons text preprocessing steps which will be used:

- Removing Extra Whitespaces
- Removing Punctuations
- Removing Stopwords
- Lowercasing all letters
- Lemmatization

## 2.5 Wordclouds

What is the theme of these philosophy schools?

Now we can answer this question by wordclouds, presenting the most used words for every school. In order to extract Let us know have a look at the concepts introduced by each school. Below we can see the worldclouds, presenting the most used words for every school. In order to extract the most frequent words accurately, we removed those meaningless stopwords (such as and, the, of).

[28] : `display(Image(filename='/Users/lsx/Desktop/a.png', embed=True))`



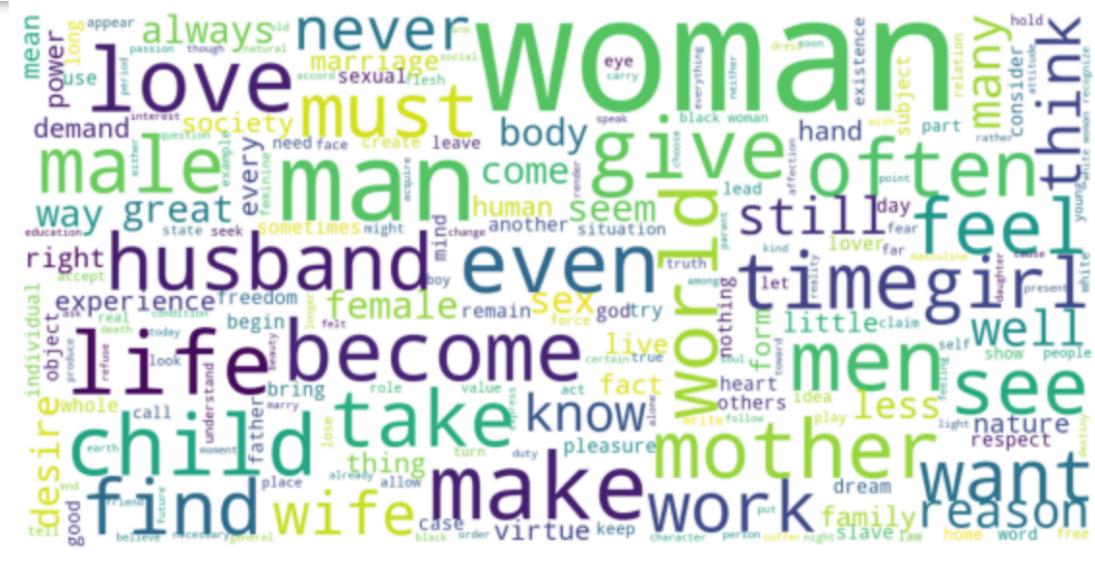
```
[29]: display(Image(filename='/Users/lsx/Desktop/b.png', embed=True))
```



```
[30]: display(Image(filename='/Users/lsx/Desktop/c.png', embed=True))
```



```
[20]: display(Image(filename='/Users/lsx/Desktop/13.png', embed=True))
```



## FEMINISM

We can get some key words for each school by the wordclouds chronological; and compare it with our knowledge about the respective philosophical below. - Plato: think, man, good, well, people, and socrate - Aristotle: thing, animal, man, body, time ,reason, - Stoicism: thing ,man , unto, nature,either, whatsoever

- Rationalism: thing , god, mind, body, reason,soul, nature
- Empiricism: thing ,ideal, men, reason, mind, object, man, body
- Capitalism: either,time,country,capital, though ,produce,tax
- German idealism:: concept, nature, object, form, determination, universal
- Communism: work, capital, hand, capitalist, labourer
- Nietzsche: even,thing,great,man , life, zarathustra
- Analytic: think, mean, case, fact, know, see ,give,way,even
- Phenomenology: world, object, mean, experience,time, consciousness
- Feminism: woman, man, husband, child, life, love
- Continental: form, madness, order, language, relation

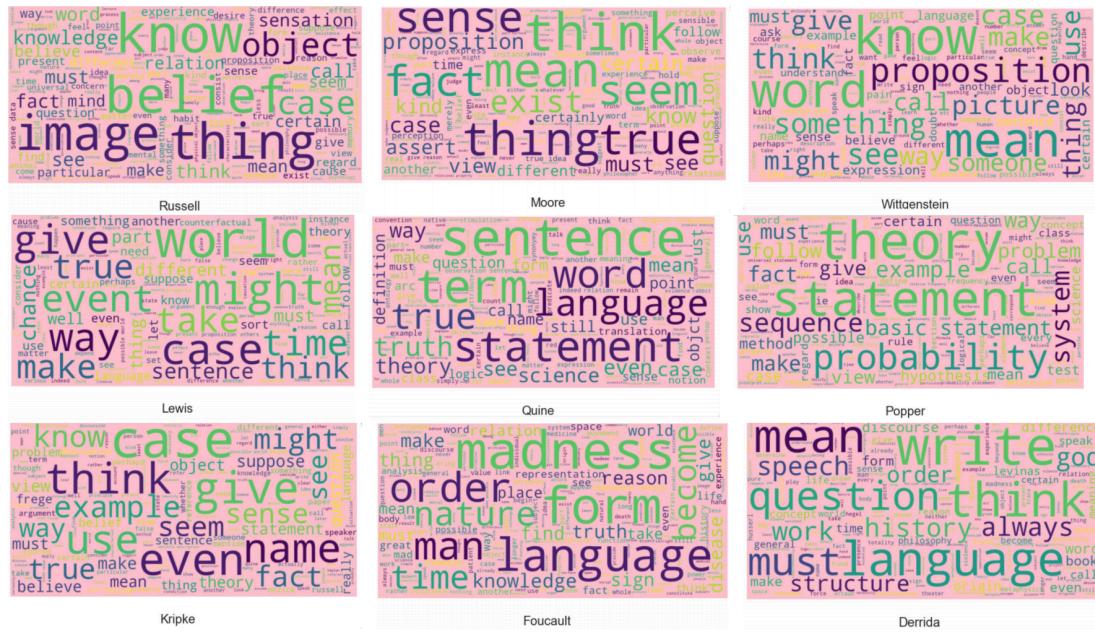
The results of the wordclouds also clearly display the differences in thought and writing between each philosophical school. For example, the Rationalism school focuses on god and soul but the capitalism and Communism tend to focus on ideas such as labour and nation.

Moreover, we also observe that older schools such as the Plato school frequently use antiquated terms such as thyself and hath which do not appear in the writing of the newer schools.

```
[21]: display(Image(filename='/Users/lsx/Desktop/001.png', embed=True))
```



```
[25]: display(Image(filename='/Users/lsx/Desktop/002.png', embed=True))
```



```
[26]: display(Image(filename='/Users/lsx/Desktop/003.png', embed=True))
```



```
[27]: display(Image(filename='/Users/lsx/Desktop/004.png', embed=True))
```



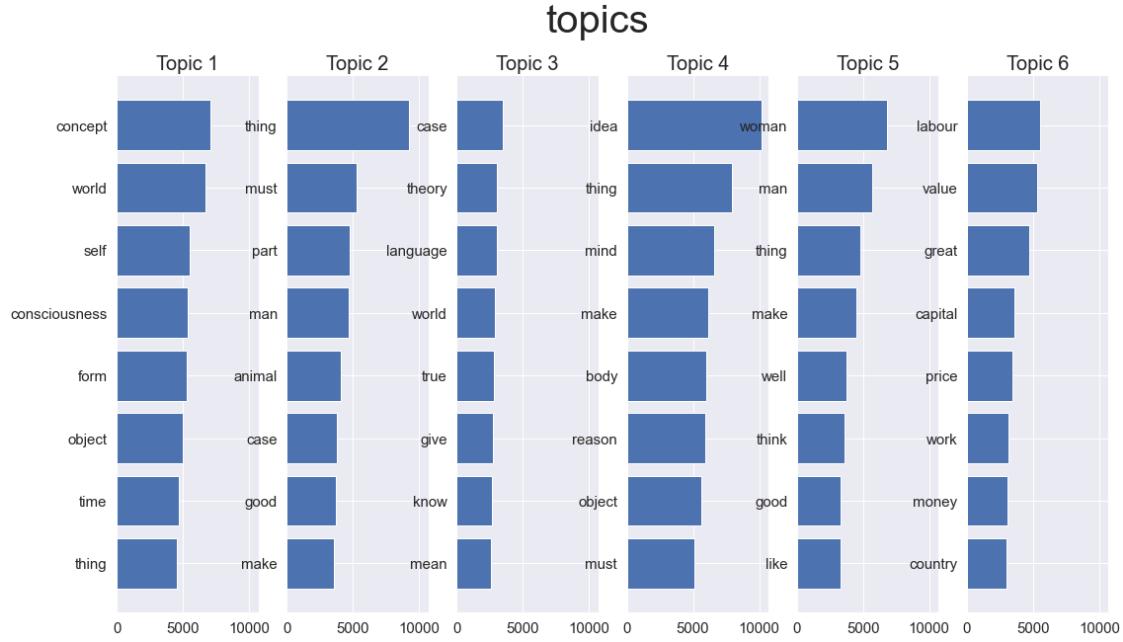
Notably, we see authors such as Spinoza, Leibniz, Descartes, and Malebranche referencing “God” a lot. Some philosophiers like Locke, Hume, Berkeley seem to be have a lot of text regarding “Ideas”. Other obvious observations include Marx and Lenin who focus on labour, production, the state, etc. which makes sense in their contexts as political figures rather than philosophers. Similarly, Smith, Ricardo, and Keynes also make reference to government but seem to be interested in the idea of capital and economy. Besides, Wollstoncraft, Beauvoir and Davis can be grouped as talking a lot about women which makes sense as women’s rights activists.

## 2.6 Topic Modeling

Now it would be interesting to investigate if there exist overlap in the different philosophical schools. Topic modeling could help us to find the difference and similarities of philosophical thoughts and observe their evolution.

Specifically, we will try to generate 6 different classifications from the available books and investigate how they correlate to each other. The most occurred words for the created topics are plotted below.

```
[23]: display(Image(filename='/Users/lsx/Desktop/LDA_topics.png', embed=True))
```



We immediatley see that the created classifications include words that are very representative of specific topics. For example, texts included in Topic 5 are mostly ‘genetated’ by worlds such as capital, labour, woman. We immediately see that texts refering to communism or capitalism must have a high probability to be mostly classified in this topic (we use the word mostly as LDA performs a sort of soft assignment of documents to topics).

Another observation is that the plot is consistent with the hypothesis that the different schools share common themes.(similarity) We can see how different topics share common words in the figure above.

We now perform the following vizualization that allowes us to: Manually select each topic to view its top most frequent and/or “relevant” terms, using different values of the `n` parameter. This can help whith assigning a human interpretable name or “meaning” to each topic. Explore the Intertopic Distance Plot.This can help to learn about how topics relate to each other, including potential higher-level structure between groups of topics. This shown a transformation for dimentionality reduction and uses only the first 2 components.

#mind, reason, soul, god, labor, and woman.

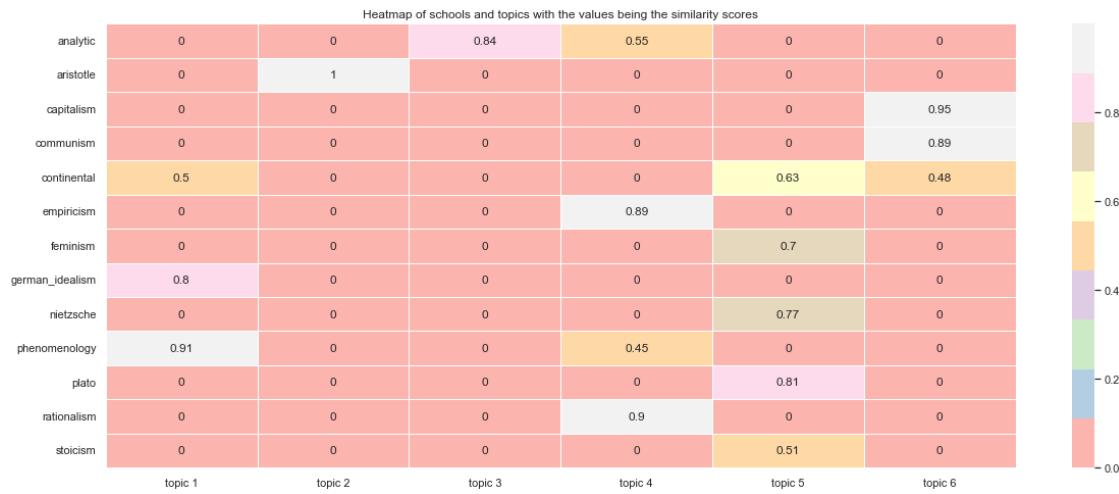
we can try to assign names to the topics using figure’s labeling : - Topic1: Consciousness and concept - Topic2: human and nature - Topic3: theory and language - Topic4: thinking and human mind - Topic5: man and woman - Topic6: Labour and capital

It’s easy to find the level of similarity between different philosophy schools in the following heatmap below.

The results depict evolution of philosophical thinking. We can find that how plato and stoicism, two of the most ancient philosophical schools share common themes with nietzsche and feminism. That means some ideas are processed and developed into new texts rather than disappeared.

Most schools have satisfactory classification but some schools like continental philosophy is misclassified in **man and woman** group, Additionally, certain topics are mentioned by more than 3 schools, for example, such as “**woman and man**” and “**thinking and mind**”, since the model always classified some texts into various classes. One unique observation is that the idiomorphism of the language used by aristotle and the fact that he considers nature, animals, human bodies are among the reasons why he is clustered in a separate topic.

```
[24]: display(Image(filename='/Users/lsx/Desktop/heatmap.png', embed=True))
```



## Readability Readability is a measure of how easy a piece of text is to read. The level of complexity of the text, its familiarity, legibility and typography all feed into how readable your text is.

## 2.7 Readability scores

In order to calculate the scores, we need to calculate 3 components: Total number of sentences in a text, total number of words in a text ,total number of syllables in a text.

We can use the above statistics to calculate Flesch Reading Ease interpret the score.

```
[36]: display(Image(filename='/Users/lsx/Desktop/readability scores equation.png', embed=True))
```

- **Flesch Reading Ease:**

$$206.835 - 1.015 \left( \frac{\text{total words}}{\text{total sentences}} \right) - 84.6 \left( \frac{\text{total syllables}}{\text{total words}} \right)$$

We simply implement the formula with the functions defined before and get calculation result below.

```
[46]: display(Image(filename='/Users/lsx/Desktop/score.png', embed=True))
```

```
: print(df_title["Flesch_score"].describe())
print("-----")
print(df_title["grade_level"].describe())

count    59.000000
mean     46.035593
std      11.866437
min      13.000000
25%     38.950000
50%     45.400000
75%     52.700000
max      75.400000
Name: Flesch_score, dtype: float64
-----
count    59.000000
mean     13.728814
std      3.402993
min      6.000000
25%     12.000000
50%     14.000000
75%     15.000000
max      28.000000
Name: grade_level, dtype: float64

: print("The book that is most difficult to read is '{t}' by {a} originally published in {y} with a Flesch reading-ease score of {s}"
      .format(t=df_title.index[0], a=df_title.iloc[0,0], y=df_title.iloc[0,1], s=df_title.iloc[0,-2]))

print("The book that is easiest to read is '{t}' by {a} originally published in {y} with a Flesch reading-ease score of {s}"
      .format(t=df_title.index[-1], a=df_title.iloc[-1,0], y=df_title.iloc[-1,1], s=df_title.iloc[-1,-2]))
```

The book that is most difficult to read is "Discourse On Method" by Descartes originally published in rationalism with a Flesch reading-ease score of 13.0

The book that is easiest to read is "Thus Spake Zarathustra" by Nietzsche originally published in nietzsche with a Flesch reading-ease score of 75.4

Generally, the range of readability is actually very wide for the 59 texts. The readability score ranges from 75.4 to 13.0, which means some texts are extremely difficult to read but also some are relatively easy.

It is beneficial for new readers to begin with those books with higher readability scores.

Let's analyze the readability scores deeper. The table below is classifications of difficulty scores which we will use later.

```
[37]: display(Image(filename='/Users/lsx/Desktop/table.png', embed=True))
```

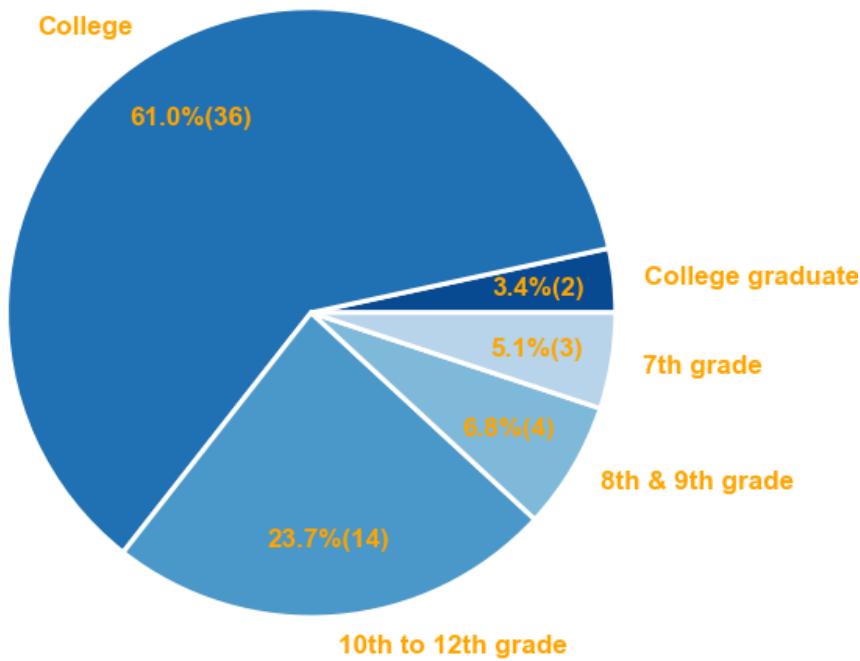
Score	School level	Notes
100.00–90.00	5th grade	Very easy to read. Easily understood by an average 11-year-old student.
90.0–80.0	6th grade	Easy to read. Conversational English for consumers.
80.0–70.0	7th grade	Fairly easy to read.
70.0–60.0	8th & 9th grade	Plain English. Easily understood by 13- to 15-year-old students.
60.0–50.0	10th to 12th grade	Fairly difficult to read.
50.0–30.0	College	Difficult to read.
30.0–10.0	College graduate	Very difficult to read. Best understood by university graduates.
10.0–0.0	Professional	Extremely difficult to read. Best understood by university graduates.

We can convert the score into the corresponding school levels and use a pie chart to visualize the findings.

From the graph, generally, the philosophy texts are hard to read since over 60% of texts have college level. This level of difficulty to read partially explains why American high schools rarely offer philosophy classes and why French high school students find understanding class contents difficult. However, only 2 texts are college graduate level readings and no book has professional level. It indicates that people feel hard to fully understand the philosophy works despite most texts of philosophy we can read with difficulty. This seems be why so many people regarding philosophy to be difficult. .

```
[38]: display(Image(filename='/Users/lsx/Desktop/pie_chart.png', embed=True))
```

## Flesch reading-ease test of philosophy books



## 2.8 Categorical Readability

Now, in order to know the difficulty of some texts which are not among those 59 texts. It would be interesting to show the relationship between readability and time/ schools/ authors.

```
[40]: display(Image(filename='/Users/lsx/Desktop/Flesch reading-ease score over time.  
→png', embed=True))
```



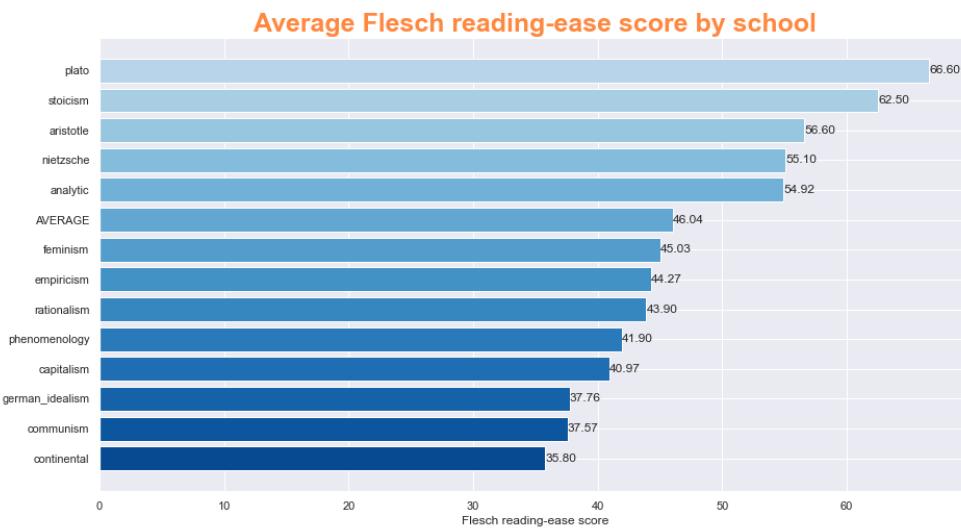
Unfortunately, the original publication date can not be as the indicator of texts readability. However, we can try to investigate those texts which published after 1600 since most publications were after 1600.

```
[43]: display(Image(filename='/Users/lsx/Desktop/score_over_time.png', embed=True))
```



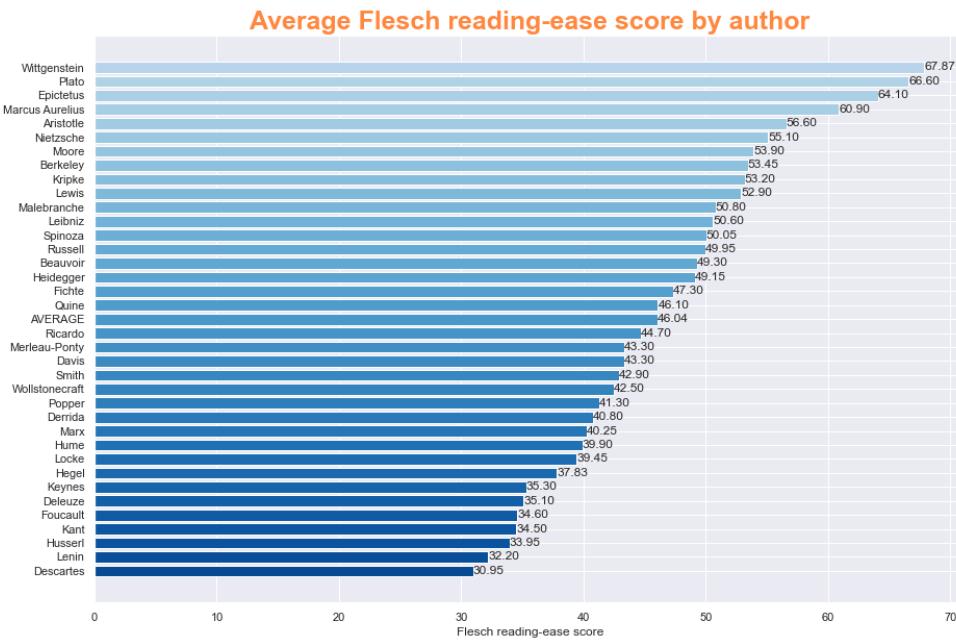
Although we limited the publication time after 1600, there is no clear trend between the readability score and time.

```
[44]: display(Image(filename='/Users/lsx/Desktop/score_by_school.png', embed=True))
```



when we analyze readability of texts from different schools. those texts from Plato, Stoicism, Aristotle, Nietzsche, and Analytic schools of thought have relative higher readability but texts from Continental, Communism, German Idealism are among the vary difficult to read.

```
[45]: display(Image(filename='/Users/lsx/Desktop/score_by_author.png', embed=True))
```



From the graph above, the works of Wittgenstein, Plato, Epictetus and Marcus Aurelius are relatively easy to read but the works of Descartes, Lenin, Husserl, Kant and many others are much harder to read.

```
[ ]:
```

## 2.9 Conclusions

In this project we studied 13 schools of philosophy about their theme and similarities. Besides, some topics according to emotion of philosophers changed as the time passed.

As philosophy express minds of people, it is expected that philosophical ideas grow and progress as society do. The similarities in different schools indicates that to be familiar with many philosophy schools are important for those people who desire to understand philosophy deeper.

As the analysis demonstrates, the majority (over 60%) of the texts require people have at least college level education if they want to read based on the Flesch reading ease score.

Not all the philosophy texts are difficult to read . For example, 59 texts has a wide spread of reading ease score from 75.4 (7th grade) to 13.0 (College graduate). Therefore people with less experience reading philosophy texts ,such as high school students, can begin with easier texts and they will be able to understand some of those texts.

## 2.10 Limitation

- some schools like plato only has one author or book, we can't get the accurate analysis from the little data.
- when we remove the stopwords, some words that deleted may affect little or much that we are not sure.
- we can also analyze how people's emotions towards specific topics (e.g., woman, mind) changed over time by sentiment analysis, but I don't enough time and have problem in building the function of emotions.
- For text difficulty sorting, we should know the difference between understanding and reading. Although we have the formula as an indication to get text reading scores, the score can not translate into the difficulty level after completely understanding readings because of some abstract content of texts

[ ]: