Feel the Philosophy: Insight into Schools and Authors

1. Introduction

“We live in the best of all possible worlds” - Gottfried Wilhelm Leibniz

The quote above by the famous philospher Leibniz can't be more true when it comes to knowledge and philosophy is what guides it. It allows us to think critically, ask important questions and fuels progress. But understanding philosophy in itself is important and that is what you should be ready for in the article.

In the below analyses, I used the philosphy data with more than 350 thousand rows to understand the underlying 'School of thoughts' of philosophy and also the authors who have contributed towards it.

I utilized Python libraries for plots, text mining, word cloud and sentiment analysis for the same.

The analyses are divided into four sections:

1. Timelines for 'School of Thoughts'
2. Most frequently used words among the schools and how they fit with the thought principle
3. Do the School of thoughts overlap?
4. Looking into the Philosophers:
    - How the sentence complexity varies for each author?
    - Emotion and Sentiment analysis for each author

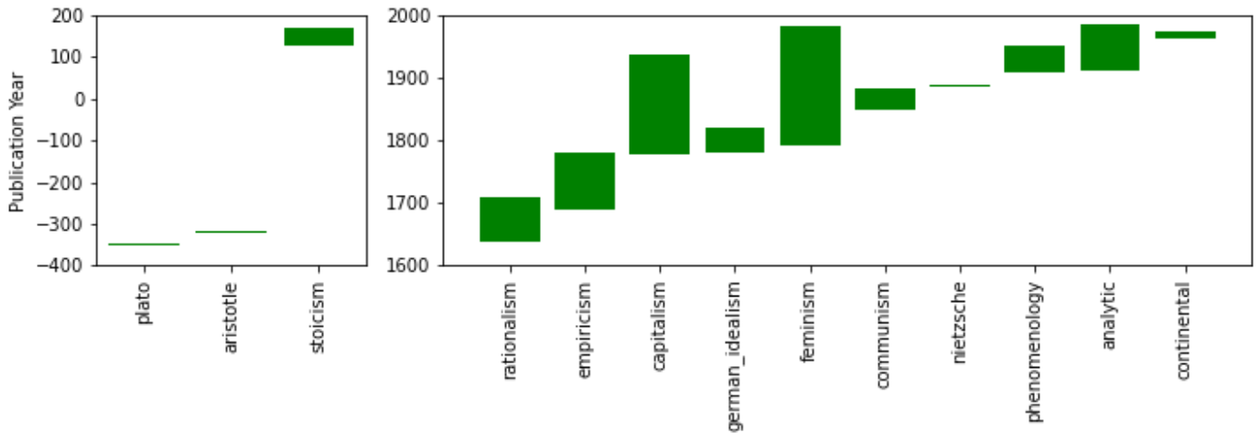
2. Data Source

The data was taken from Kaggle database https://www.kaggle.com/kouroshalizadeh/history-of-philosophy (https://www.kaggle.com/kouroshalizadeh/history-of-philosophy).

3. Analysis

3a. Timelines for 'School of thoughts'

It is known that during a given period in history a particular way of thinking is more common. I used the original publication dates of different books to identify the *eras* for each school of thought. The plot below shows the duration for each school of thought and is ordered based on their earliest publication date.



From the figure above, it can be seen that Capitalism and Feminism are longest lasting school of thoughts. And if we were to compare schools that are considered opposites, for example, Capitalism and Communism - Capitalism shows to have lasted way longer than communism.

Additionally, when I deep-dive into each schools, the timeline seems to overlap major events contributing to that thought process. Specifically, if we look at 'Feminism' school of thought, it overlaps with the Women's Rights Movement (the first wave of feminism) which lasted from 1848–1917. This was largely discussed topic at that time due to the fight for voting rights for women. Following this, the second wave for feminism went from 1963 to 1980's. We can clearly conclude that these movements have played towards the popularity of this schools of thought.

Furthermore, we can clearly see that the most modern schools of thoughts are 'Analytic', and 'Continental' based on their publication dates.

3b. Most frequently used words among schools and how they fit with the thought principle

In the next analysis, I used the given sentences from different publications to create word clouds based on frequency of use for each schools.

The word clouds are made up of top fifty most frequent words in each school and their size depends on the frequency. They follow the expected pattern and show the words at the center of the ideology to be frequently used.

Schools of philosophy - Word Cloud

Analytic

Aristotle

Capitalism

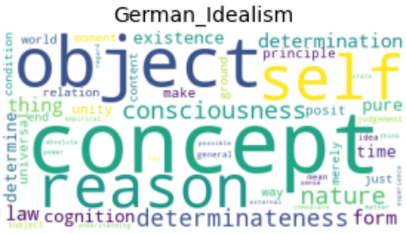
Communism

Continental

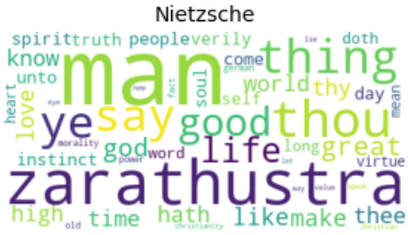
Empiricism

Feminism

German_Idealism



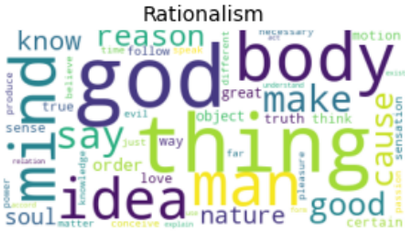
Nietzsche



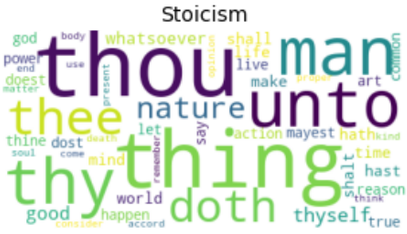
Phenomenology

Plato

Rationalism



Stoicism



Let's look at a couple of schools above:

- 'Analytic' school has an emphasis on language, so the words like *'say', 'sentence', 'know'* lie in the most commonly used words.
- 'Capitalism' school has an emphasis on economic system within nations and hence the words like *'price', 'country', 'money'* are very frequently used, whereas for communism, 'labour' and 'value' are more frequent.
- 'Phenomenology' school studies the consciousness and experience. Words like *'world', 'thing', 'time'* and others are very relevant to it.
- We can explore other schools similarly to develop an in-depth understanding.

3c. Do the School of thoughts overlap?

Next question that came to my mind was: seeing as there are so many schools that exist, is there a possibility that any of the schools are very closely related? This could be either because of similar thought process, or maybe because they are opposite.

In the table below, we can see the schools which have a significant number of overlapping top words among them. We chose to look at only top five school pairs as they represent almost 50% overlap in the top words.

	School_A	School_B	Common_word_count
0	empiricism	rationalism	29
1	rationalism	aristotle	27
2	empiricism	aristotle	24
3	aristotle	analytic	23
4	plato	aristotle	23

Focusing on the top overlapping schools. Empiricism tells us that the knowledge about the world comes from experience or sense and its perception. On the other hand, rationalism believes that the knowledge can exist even without experience and that experience is not fully reliable to rest our knowledge. However, they both want to seek knowledge that is more or less scientific and have complemented each other, the way mathematic and physics complement one other even though their base is entirely different.

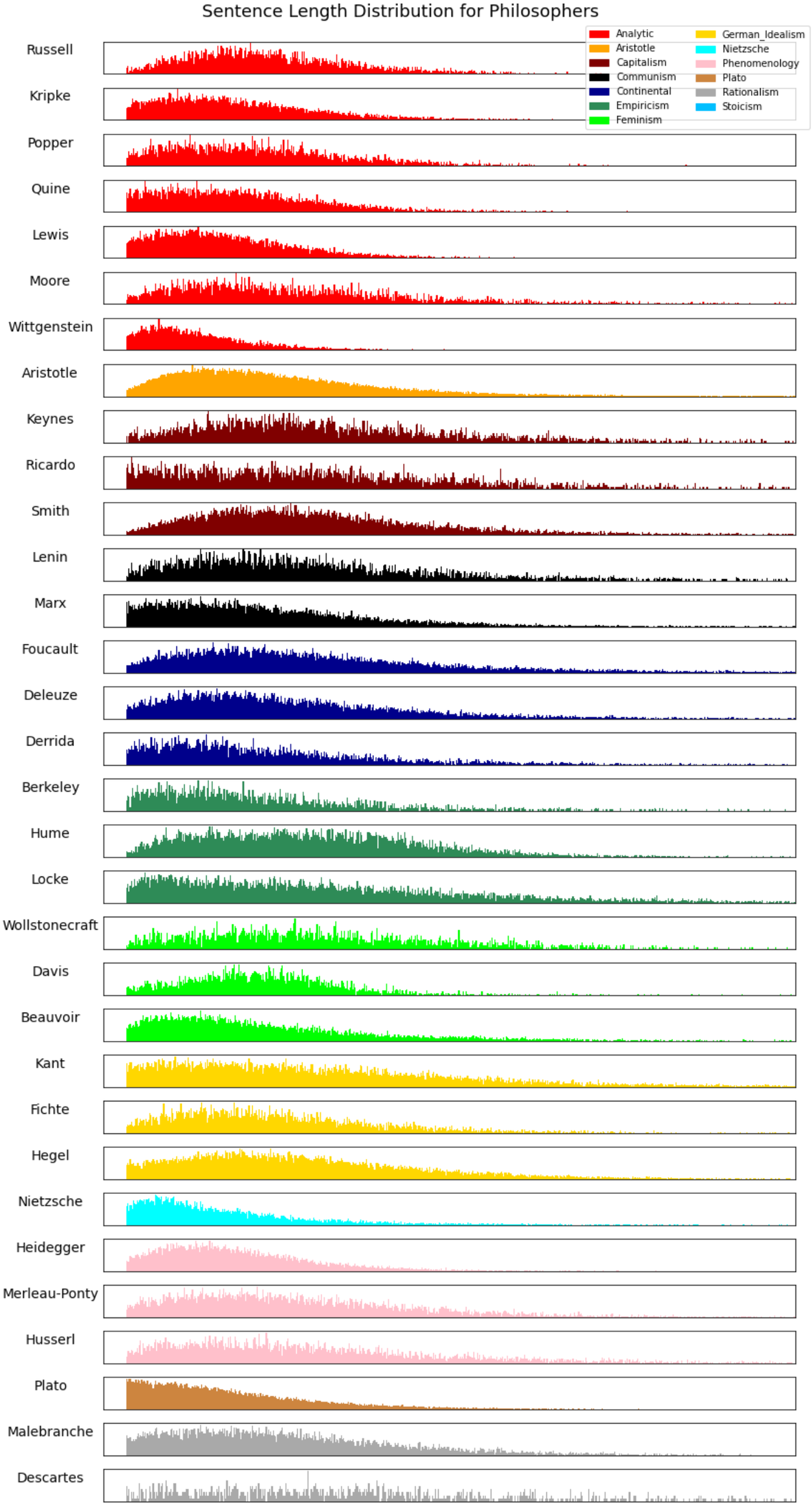
3d. Looking into the Philosophers:

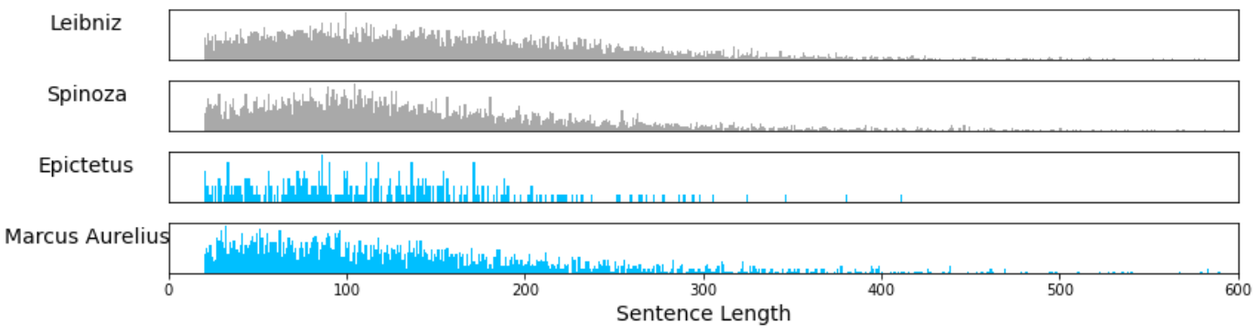
3d.1. How the sentence complexity varies for each author?

Now, as I was looking into individual philosophers, I decided to start by looking at the complexity of their sentences. I furhter wanted to see if there is a pattern with respect to these philosophers' schools of thought.

From the below chart, following observations were made:

- On an average, most of the philosophers have a sentence length between 50-100 characters.
- All the authors of 'Capitalism' school have a higher average sentence length (~150 characters).
- Authors of 'Rationalism' school have more variation in sentence length than other authors.
- Plato shows to have compact sentences on majority than all other authors.
- Authors of 'Feminism' school do not have a common average for sentence length.


Sentence Length Distribution for Philosophers
Analytic
Aristotle
Capitalism
Communism
Continental
Empiricism
Feminism
German_Idealism
Nietzsche
Phenomenology
Plato
Rationalism
Stoicism
Russell
Kripke
Popper
Quine
Lewis
Moore
Wittgenstein
Aristotle
Keynes
Ricardo
Smith
Lenin
Marx
Foucault
Deleuze
Derrida
Berkeley
Hume
Locke
Wollstonecraft
Davis
Beauvoir
Kant
Fichte
Hegel
Nietzsche
Heidegger
Merleau-Ponty
Husserl
Plato
Malebranche
Descartes



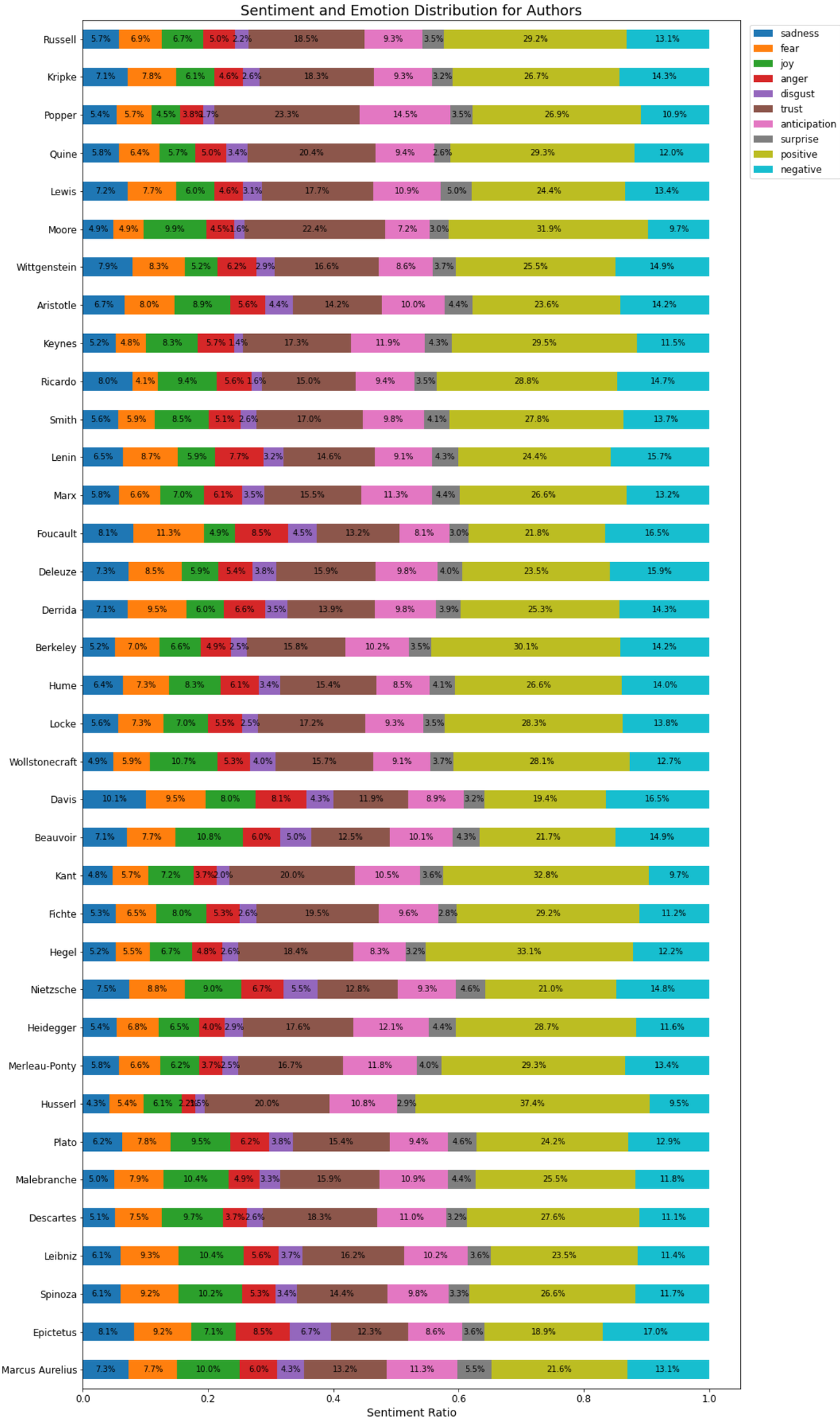
3d.2. Emotion and Sentiment analysis for each author

Sentiment and emotion of writing is very important to analyze in philosophy. In the chart below, we can see the ratio of each sentiments and emotions for all publications of an author.

Firstly, looking at a high-level, we notice that all authors show a higher ratio of 'trust' emotion in their sentences. Emotions like 'sadness', 'fear', and 'joy' have almost similar ratios for all authors. Ratio of 'anticipation' to 'surprise' is also similar. Next, when we try to identify pattern, we observe that the authors of 'German-idealism' and 'Empiricism' schools have very high positive sentiment.

Then we look at individual philosophers and notice some exceptions.

- Epictetus, Nietzsche, and Beauvoir have more than 5% of 'disgust' emotion in their sentences.
- Husserl has a very low 'anger' emotion ratio.
- Davis and Epictetus have nearly similary positive and negative sentiment ratios.


Sentiment and Emotion Distribution for Authors
sadness
fear
joy
anger
disgust
trust
anticipation
surprise
positive
negative
Russell 5.7% 6.9% 6.7% 5.0% 2.2% 18.5% 9.3% 3.5% 29.2% 13.1%
Kripke 7.1% 7.8% 6.1% 4.6% 2.6% 18.3% 9.3% 3.2% 26.7% 14.3%
Popper 5.4% 5.7% 4.5% 3.8% 1.7% 23.3% 14.5% 3.5% 26.9% 10.9%
Quine 5.8% 6.4% 5.7% 5.0% 3.4% 20.4% 9.4% 2.6% 29.3% 12.0%
Lewis 7.2% 7.7% 6.0% 4.6% 3.1% 17.7% 10.9% 5.0% 24.4% 13.4%
Moore 4.9% 4.9% 9.9% 4.5% 1.6% 22.4% 7.2% 3.0% 31.9% 9.7%
Wittgenstein 7.9% 8.3% 5.2% 6.2% 2.9% 16.6% 8.6% 3.7% 25.5% 14.9%
Aristotle 6.7% 8.0% 8.9% 5.6% 4.4% 14.2% 10.0% 4.4% 23.6% 14.2%
Keynes 5.2% 4.8% 8.3% 5.7% 1.4% 17.3% 11.9% 4.3% 29.5% 11.5%
Ricardo 8.0% 4.1% 9.4% 5.6% 1.6% 15.0% 9.4% 3.5% 28.8% 14.7%
Smith 5.6% 5.9% 8.5% 5.1% 2.6% 17.0% 9.8% 4.1% 27.8% 13.7%
Lenin 6.5% 8.7% 5.9% 7.7% 3.2% 14.6% 9.1% 4.3% 24.4% 15.7%
Marx 5.8% 6.6% 7.0% 6.1% 3.5% 15.5% 11.3% 4.4% 26.6% 13.2%
Foucault 8.1% 11.3% 4.9% 8.5% 4.5% 13.2% 8.1% 3.0% 21.8% 16.5%
Deleuze 7.3% 8.5% 5.9% 5.4% 3.8% 15.9% 9.8% 4.0% 23.5% 15.9%
Derrida 7.1% 9.5% 6.0% 6.6% 3.5% 13.9% 9.8% 3.9% 25.3% 14.3%
Berkeley 5.2% 7.0% 6.6% 4.9% 2.5% 15.8% 10.2% 3.5% 30.1% 14.2%
Hume 6.4% 7.3% 8.3% 6.1% 3.4% 15.4% 8.5% 4.1% 26.6% 14.0%
Locke 5.6% 7.3% 7.0% 5.5% 2.5% 17.2% 9.3% 3.5% 28.3% 13.8%
Wollstonecraft 4.9% 5.9% 10.7% 5.3% 4.0% 15.7% 9.1% 3.7% 28.1% 12.7%
Davis 10.1% 9.5% 8.0% 8.1% 4.3% 11.9% 8.9% 3.2% 19.4% 16.5%
Beauvoir 7.1% 7.7% 10.8% 6.0% 5.0% 12.5% 10.1% 4.3% 21.7% 14.9%
Kant 4.8% 5.7% 7.2% 3.7% 2.0% 20.0% 10.5% 3.6% 32.8% 9.7%
Fichte 5.3% 6.5% 8.0% 5.3% 2.6% 19.5% 9.6% 2.8% 29.2% 11.2%
Hegel 5.2% 5.5% 6.7% 4.8% 2.6% 18.4% 8.3% 3.2% 33.1% 12.2%
Nietzsche 7.5% 8.8% 9.0% 6.7% 5.5% 12.8% 9.3% 4.6% 21.0% 14.8%
Heidegger 5.4% 6.8% 6.5% 4.0% 2.9% 17.6% 12.1% 4.4% 28.7% 11.6%
Merleau-Ponty 5.8% 6.6% 6.2% 3.7% 2.5% 16.7% 11.8% 4.0% 29.3% 13.4%
Husserl 4.3% 5.4% 6.1% 2.2% 1.5% 20.0% 10.8% 2.9% 37.4% 9.5%
Plato 6.2% 7.8% 9.5% 6.2% 3.8% 15.4% 9.4% 4.6% 24.2% 12.9%
Malebranche 5.0% 7.9% 10.4% 4.9% 3.3% 15.9% 10.9% 4.4% 25.5% 11.8%
Descartes 5.1% 7.5% 9.7% 3.7% 2.6% 18.3% 11.0% 3.2% 27.6% 11.1%
Leibniz 6.1% 9.3% 10.4% 5.6% 3.7% 16.2% 10.2% 3.6% 23.5% 11.4%
Spinoza 6.1% 9.2% 10.2% 5.3% 3.4% 14.4% 9.8% 3.3% 26.6% 11.7%
Epictetus 8.1% 9.2% 7.1% 8.5% 6.7% 12.3% 8.6% 3.6% 18.9% 17.0%
Marcus Aurelius 7.3% 7.7% 10.0% 6.0% 4.3% 13.2% 11.3% 5.5% 21.6% 13.1%
0.0
0.2
0.4
0.6
0.8
1.0
Sentiment Ratio

4. Conclusion

From the analysis above, we observe that different schools of thoughts have vastly different word use and the way in which the philosophers used to write. This is also impacted by the era of philosophy. The analysis also shows us how the schools of thought impact writing complexity and sometimes emotions.

5. Future Work

In the future, this research could be extended to analyzing philosophers on a publication level. It could also include additional datasets having more publication texts, and philosopher demographics to note if location affects the style of writing or ideology.

APPENDIX

6. Python Code

In [1]:
```
# Importing Libraries

import pandas as pd
import matplotlib.pyplot as plt
from wordcloud import WordCloud
from sklearn.feature_extraction import text
from sklearn.feature_extraction.text import TfidfVectorizer
import matplotlib.patches as mpatches
from nrclex import NRCLex
import numpy as np
from tabulate import tabulate
```

In [2]:
```
# Loading the data

phil_data_raw = pd.read_csv(r'C:\Users\aakan\Downloads\philosophy_data1.csv')
```

In [3]:
```
# Timeline of school of philosophy Dataset

Q1_data = phil_data_raw[['school']].groupby(['school']).max()
Q1_data['min_pub'] = phil_data_raw[['school','original_publication_date']].gro
upby(['school']).min()
Q1_data['max_pub'] = phil_data_raw[['school','original_publication_date']].gro
upby(['school']).max()
Q1_data['lim_pub'] = Q1_data['max_pub'] - Q1_data['min_pub']

Q1_data.loc[Q1_data['lim_pub'] == 0, 'lim_pub'] = 2
Q1_data.reset_index(inplace=True)
Q1_data.sort_values(by=['min_pub'], inplace = True)

Q1_data_pos = Q1_data.loc[Q1_data['max_pub'] > 200]
Q1_data_neg = Q1_data.loc[Q1_data['max_pub'] < 200]
```

In [14]:
```
fig, (ax2, ax1) = plt.subplots(1, 2, sharex=False, sharey=False , gridspec_kw=
{'width_ratios': [1, 3]})
fig.subplots_adjust(hspace=0.05)

ax1.bar(Q1_data_pos['school'], Q1_data_pos['lim_pub'], bottom = Q1_data_pos['m
in_pub'], color = 'green')
ax2.bar(Q1_data_neg['school'], Q1_data_neg['lim_pub'], bottom = Q1_data_neg['m
in_pub'], color = 'green')

ax1.set_ylim(1600, 2000)
ax2.set_ylim(-400, 200)

fig.tight_layout(pad=0.0)
ax1.set_xticklabels(Q1_data_pos['school'], rotation=90)
ax2.set_xticklabels(Q1_data_neg['school'], rotation=90)

fig.suptitle('Active Years for schools of philosophy', fontsize=18.0, y = 1.1)
ax2.set_ylabel('Publication Year')
fig.set_figwidth(10)

plt.tight_layout()
plt.savefig(r'C:\Users\aakan\OneDrive\Documents\Spring 2022\Applied Data Scien
ce\GitHub\spring-2022-prj1-aa4863\figs\section1.png', dpi=fig.dpi)
plt.show()
```

In [5]:
```
# Q2 Most frequently Used words

schools = sorted(list(set(phil_data_raw['school'])))
schools_words = {}

for school in schools:
    schools_words[school] = ' '.join(phil_data_raw[phil_data_raw['school'] ==
school]['lemmatized_str'].tolist())
```

In [6]:
```
# with column 'lemmatized_str'

corpus = []
schools = []
for school in schools_words:
    schools.append(school)
    corpus.append(schools_words[school])

schools = sorted(schools)

my_stop_words = text.ENGLISH_STOP_WORDS.union(["pron"])

vectorizer = TfidfVectorizer(stop_words = my_stop_words)
vecs = vectorizer.fit_transform(corpus)
feature_names = vectorizer.get_feature_names()
dense = vecs.todense()
lst1 = dense.tolist()
df = pd.DataFrame(lst1, columns=feature_names)

clouds_dict1 = {}

for i in range(len(df)):
    Cloud = WordCloud(background_color = "white", max_words = 50).generate_fro
m_frequencies(df.iloc[i])
    clouds_dict1[schools[i]] = Cloud
```

In [13]:
```
fig, axes = plt.subplots(5, 3, sharex=False, sharey=False)

fig.delaxes(axes[4,1])
fig.delaxes(axes[4,2])

for i in range(len(schools)):
    x = i // 3
    y = i % 3
    axes[x,y].imshow(clouds_dict1[schools[i]])
    axes[x,y].set_axis_off()
    axes[x,y].set_title(schools[i].title(), fontsize=14.0)

fig.tight_layout(pad=0.5)
fig.suptitle('Schools of philosophy - Word Cloud', fontsize=18.0, y = 0.97)
fig.set_figwidth(12)
fig.set_figheight(15)
plt.tight_layout()
plt.savefig(r'C:\Users\aakan\OneDrive\Documents\Spring 2022\Applied Data Scien
ce\GitHub\spring-2022-prj1-aa4863\figs\section2.png', dpi=fig.dpi)
plt.show()
```

In [15]:

```
# Q3 Overlap between schools

top_words = {}

for school in schools:
    top_words[school] = list(clouds_dict1[school].words_.keys())

common_words = pd.DataFrame({'School_A':[],'School_B':[],'Common_words':[],'Co
mmon_word_count':[]})

for i in range(len(schools)):
    for j in range(i+1, len(schools)):
        common_words = common_words.append({'School_A': schools[i],
                                            'School_B': schools[j],
                                            'Common_words': sorted(set(top_wor
ds[schools[i]])&set(top_words[schools[j]]), key = lambda k : top_words[schools
[i]].index(k)),
                                            'Common_word_count': len(sorted(se
t(top_words[schools[i]])&set(top_words[schools[j]]), key = lambda k : top_word
s[schools[i]].index(k)))},
                                           ignore_index = True)

common_words.sort_values(by = 'Common_word_count', ascending = False, inplace
= True)

overlap = common_words[0:5][['School_A','School_B','Common_word_count']].reset
_index()
overlap = overlap.drop('index',1)

print(tabulate(overlap, headers = 'keys', tablefmt = 'fancy_grid'))
```

In [16]:
```
# Q4 : Writing style of different pholisophers
# Part A : Sentence Length

colors = ['red', 'orange', 'maroon', 'black', 'darkblue', 'seagreen', 'lime',
'gold', 'cyan', 'pink', 'peru', 'darkgrey', 'deepskyblue']

school_color = {}

for i in range(len(colors)):
    school_color[schools[i]] = colors[i]

legend_col = []

for i in range(len(colors)):
    x = mpatches.Patch(color = colors[i] , label=schools[i].title())
    legend_col.append(x)

school_author = phil_data_raw[['school','author']].drop_duplicates().sort_valu
es(by = 'school').reset_index()

word_len_dict = {}

for author in school_author['author']:
    word_len_dict[author] = phil_data_raw[phil_data_raw['author'] == author][
'sentence_length'].tolist()

fig, axes = plt.subplots(36, 1, sharex=False, sharey=False)

for i in range(len(school_author)):
    axes[i].hist(word_len_dict[school_author['author'][i]], bins=range(min(wor
d_len_dict[school_author['author'][i]]), max(word_len_dict[school_author['auth
or'][i]]) + 1, 1),
                 alpha=1.0, color=school_color[school_author['school'][i]])
    axes[i].axes.yaxis.set_ticks([])
    if i<len(school_author)-1:
        axes[i].axes.xaxis.set_ticks([])

    axes[i].set_xlim(0,600)
    axes[i].set_ylabel(school_author['author'][i].title(), fontsize=14.0, labe
lpad = 55.0, rotation = 0)

axes[35].set_xlabel('Sentence Length', fontsize=14.0)
fig.suptitle('Sentence Length Distribution for Philosophers', fontsize=18.0, y
= 1.0)
fig.set_figwidth(12)
fig.set_figheight(25)
fig.legend(handles = legend_col, loc = 'upper right', ncol = 2, bbox_to_anchor
=(1.0, 0.99))
plt.tight_layout()
plt.savefig(r'C:\Users\aakan\OneDrive\Documents\Spring 2022\Applied Data Scien
ce\GitHub\spring-2022-prj1-aa4863\figs\section4a.png', dpi=fig.dpi)
plt.show()
```

In [23]:
```
# Part B: Sentiment and Emotion Analysis for Philosophers

authors = list(school_author['author'])[::-1]
authors_words = {}

for author in authors:
    authors_words[author] = ' '.join(phil_data_raw[phil_data_raw['author'] ==
author]['sentence_lowered'].tolist())

senti_ratio = []

for author in authors:
    text_object = NRCLex(text = authors_words[author])
    freq = text_object.affect_frequencies
    freq['author'] = author
    senti_ratio.append(freq)

senti_ratio = pd.DataFrame(senti_ratio)
senti_ratio = senti_ratio[['sadness', 'fear', 'joy', 'anger', 'disgust', 'trus
t', 'anticipation', 'surprise', 'positive', 'negative', 'author']]
```

In [25]:
```
senti_ratio.plot(
    x = 'author',
    kind = 'barh',
    stacked = True,
    title = 'Sentiment and Emotion Distribution for Authors',
    fontsize = 12.0,
    mark_right = True,
    figsize = (15,25),
    sort_columns = True)

df_total = senti_ratio.sum(numeric_only = True, axis = 1)
df_rel = senti_ratio[senti_ratio.columns[:-1]].div(df_total, 0)*100

for n in df_rel:
    for i, (cs, ab, pc) in enumerate(zip(senti_ratio.iloc[:, :-1].cumsum(1)[n
], senti_ratio[n], df_rel[n])):
        plt.text(cs - ab / 2, i, str(np.round(pc, 1)) + '%', va = 'center', ha
= 'center')

plt.xlabel('Sentiment Ratio', fontsize=14.0)
plt.ylabel('', fontsize=0.0)
plt.title('Sentiment and Emotion Distribution for Authors', fontsize = 18.0)
plt.legend(loc = 'upper right', bbox_to_anchor=(1.16, 1.0), fontsize = 12.0)
plt.tight_layout()
plt.savefig(r'C:\Users\aakan\OneDrive\Documents\Spring 2022\Applied Data Scien
ce\GitHub\spring-2022-prj1-aa4863\figs\section4b.png', dpi=fig.dpi)
plt.show()
```