ETA 5001 Final Project

Introduction

In this project, I attempt to create a digital corpus of the Benjamin Jowwett translations of Plato. These works are central to the Western philosophical canon. This means that these works have been studied frequently by all types of academics. In this paper, I utilize some exploratory text analytics (ETA) methods. In analyzing Plato's corpus in this manner, I want to answer the following questions:

1. What, apart from time, distinguishes Plato's early, middle and late works?
2. Does Plato's works converge on a set of themes, and if so, what are they?
3. Can we account on why the *Theaetetus, Statesmen* and *Sophists* were published in the same dramatic universe, even though each work was published at different stages of Plato's life?

Corpus/Pre-Processing

The corpus consists of 25 translations of Plato by Benjamin Jowwett. These translations exclude works that are of doubtful authorship. For example, the *Epigrams,* as well as the *Clitophon* and *Demodocus* are not included in the corpus. These omissions ensure that the corpus analysed here is slightly different from modern standard translations. [1]

In terms of preprocessing, I created three different tables. The first table is the LIB table, which consists of the name of the work, and the period in which the work was published. This is demarcated by 'early, middle and late.' I then created a TOKEN table, annotated with NLTK features (F2). One problem in analysing Plato's works in this manner is that most of Plato's works do not have chapters. Rather, the dialogues are intended to be read in one, continuous manner. Plato only has two works (*The Republic* and *Laws*) which are divided into books and chapters. For this reason, I have treated most of Plato's dialogues as one chapter. As for the processing of the *Republic* and *Laws,* I have divided them into distinct books. This approach is not ideal. Many of Plato's dialogues could be chunked into smaller arguments. Nevertheless, this requires more annotation than time has permitted. From the TOKEN table, I also created a VOCAB table, with tf_idf scores.

Can We Distinguish Plato's Early, Middle and Later Works?

There has been much scholarly debate on what stylistically separates Plato's various periods. The later works are said to be more free flowing, not following the strict dialectical structure set out by the earlier dialogues, such as the *Euthyphro.* However, barring these qualitative comments, I set out to try to answer whether there exists any thematic differences. In doing this, I conducted a PCA of the available corpus.

---

[1] I refer here to the Cooper edition of Plato's works

One issue I ran into almost immediately was the decision on how many components I was to include. The scree plot for this was inconclusive.
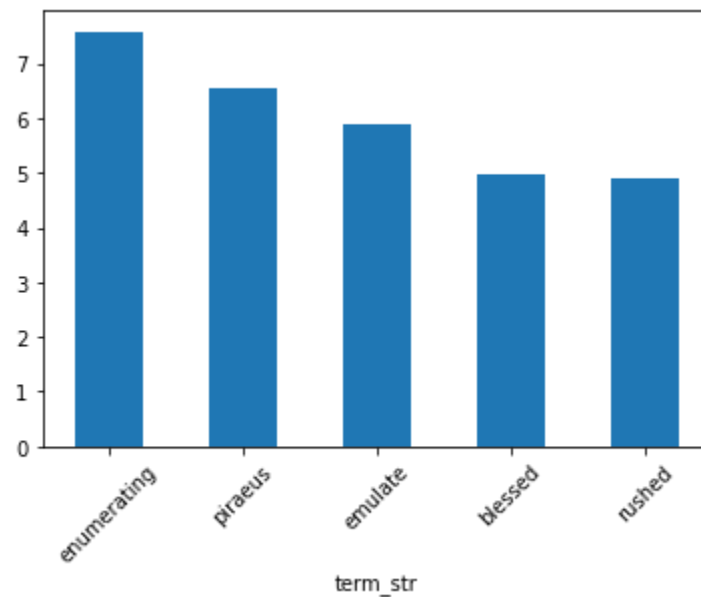

Fig I: PCA Scree Plot

I settled on two components, since the differences between the first and second were the largest. The loadings for PC1 and PC2 mainly consist of words about certainty and lacking respectively.[2] I projected the component loadings onto a two-dimensional space, and received the following results:

---

[2] The words that make up PC1 are the following: 'undoubtedly assented latter 1b pray asked probable milton term inevitable.' The words that make up PC2 are the following: 'rep welcome please bad rare without stasinus honour same come'
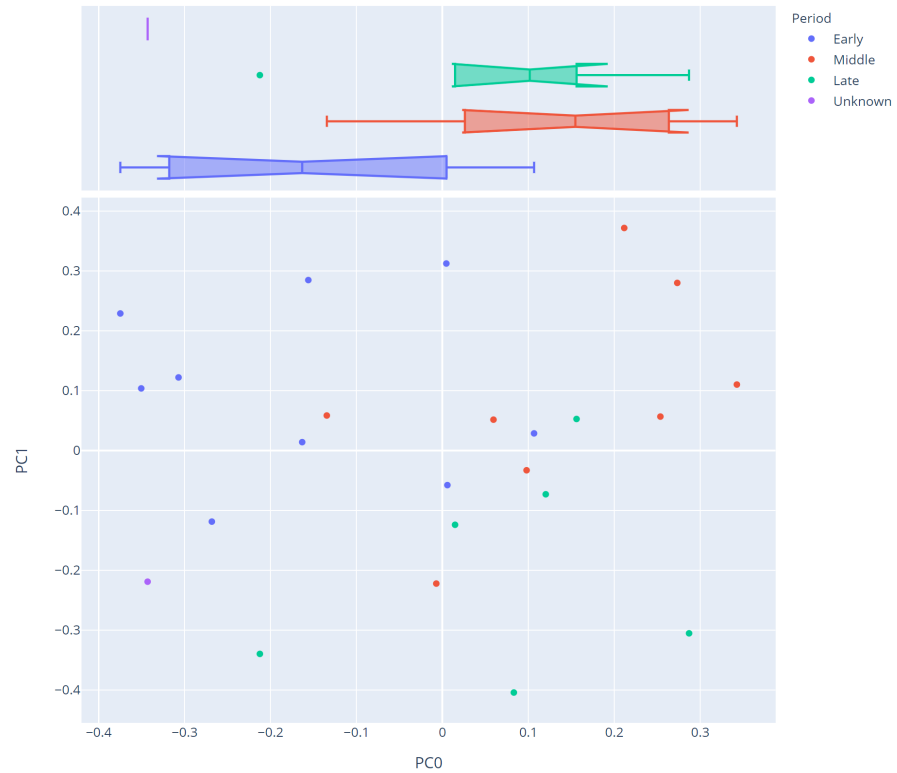
Fig II: PCA1 and 2

Early works score much lower on the certainty component. Nevertheless, it is interesting to note that there is a non-significant overlap between Plato's Middle and Late works. Nevertheless, the Later works score somewhat lower on the *certainty* component. This might be due to the freer structure adapted in these later dialogues.

Does Plato's Works Converge on a Set of Themes?

One important hermeneutic challenge in Plato interpretation is the investigation into whether Plato was united. We know that Plato was not systematic, and that his reflections often depended on the flow of each individual dialogue. But this does not mean that Plato's works were not united. In this sense, a corpus of work is united if each dialogue, on the whole, converges on a set of themes. In investigating this, I analyzed the available corpus through topic modelling. The results for topic weights are shown below.
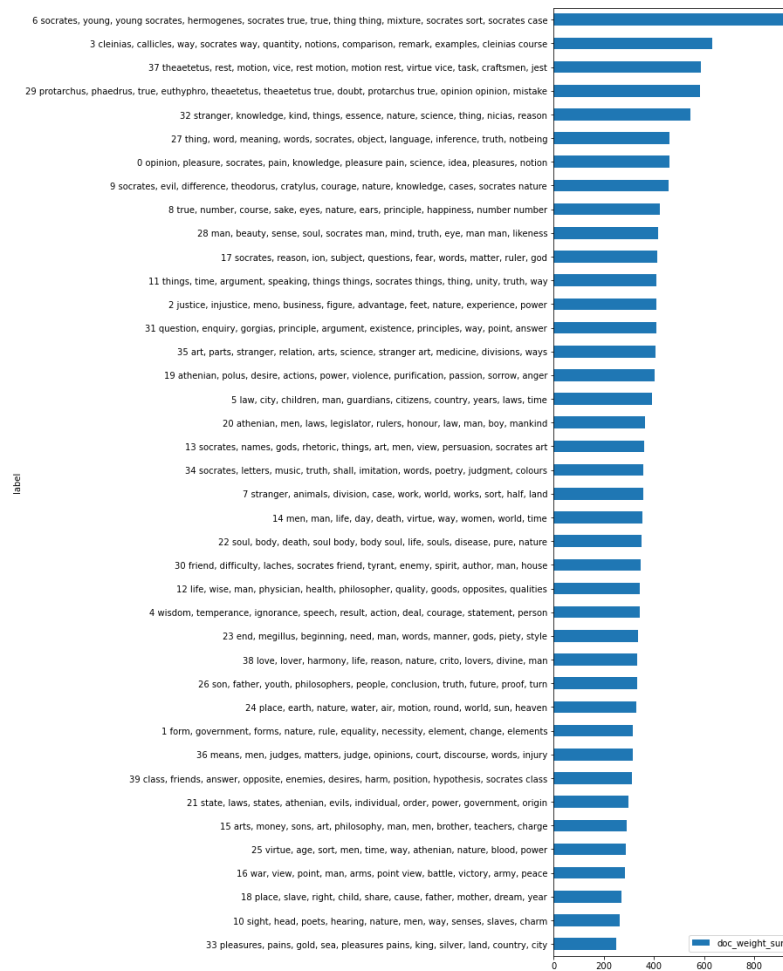
Fig III: Top Topics

Most medieval philosophers focused on Plato's works on virtue. But it is clear that, through examining the topic models, virtue was not Plato's main focus. Rather, it was the nature of 'being,' or the ontological question, that occupied most of Plato's time. Some examples of this could be given in Topic 3, which discusses the classic Western philosophical debate on 'quantity.' Another topic is topic 32, which talks about the relationship between the 'essence' of things. This does not mean that the medieval philosophers were not true to Plato's works. Topic 0 is also one of the top five Plato topic models, and discusses the moral worth of pleasure. Nevertheless, it is the discussion of the nature of things, or 'ontology,' that occupies most of Plato's corpus. Both are not mutually exclusive. But one can definitely be persuaded by Plato's metaphysics, and still be on the fence about Plato's ethics, and vice-versa.

We can cluster these topics in the following manner:

('30 friend, difficulty, laches, socrates friend, tyrant, enemy, spirit, author, man, house',)
('3 cleinias, callicles, way, socrates way, quantity, notions, comparison, remark, examples, cleinias course',)
('4 wisdom, temperance, ignorance, speech, result, action, deal, courage, statement, person',)
('2 justice, injustice, meno, business, figure, advantage, feet, nature, experience, power',)
('39 class, friends, answer, opposite, enemies, desires, harm, position, hypothesis, socrates class',)
('31 question, enquiry, gorgias, principle, argument, existence, principles, way, point, answer',)
('16 war, view, point, man, arms, point view, battle, victory, army, peace',)
('11 things, time, argument, speaking, things things, socrates things, thing, unity, truth, way',)
('27 thing, word, meaning, words, socrates, object, language, inference, truth, notbeing',)
('24 place, earth, nature, water, air, motion, round, world, sun, heaven',)
('18 place, slave, right, child, share, cause, father, mother, dream, year',)
('26 son, father, youth, philosophers, people, conclusion, truth, future, proof, turn',)
('28 man, beauty, sense, soul, socrates man, mind, truth, eye, man man, likeness',)
('12 life, wise, man, physician, health, philosopher, quality, goods, opposites, qualities',)
('22 soul, body, death, soul body, body soul, life, souls, disease, pure, nature',)
('29 protarchus, phaedrus, true, euthyphro, theaetetus, theaetetus true, doubt, protarchus true, opinion opinion, mistake',)
('8 true, number, course, sake, eyes, nature, ears, principle, happiness, number number',)
('37 theaetetus, rest, motion, vice, rest motion, motion rest, virtue vice, task, craftsmen, jest',)
('35 art, parts, stranger, relation, arts, science, stranger art, medicine, divisions, ways',)
('15 arts, money, sons, art, philosophy, man, men, brother, teachers, charge',)
('32 stranger, knowledge, kind, things, essence, nature, science, thing, nicias, reason',)
('7 stranger, animals, division, case, work, world, works, sort, half, land',)
('20 athenian, men, laws, legislator, rulers, honour, law, man, boy, mankind',)
('14 men, man, life, day, death, virtue, way, women, world, time',)
('5 law, city, children, man, guardians, citizens, country, years, laws, time',)
('36 means, men, judges, matters, judge, opinions, court, discourse, words, injury',)
('33 pleasures, pains, gold, sea, pleasures pains, king, silver, land, country, city',)
('25 virtue, age, sort, men, time, way, athenian, nature, blood, power',)
('10 sight, head, poets, hearing, nature, men, way, senses, slaves, charm',)
('1 form, government, forms, nature, rule, equality, necessity, element, change, elements',)
('38 love, lover, harmony, life, reason, nature, crito, lovers, divine, man',)
('23 end, megillus, beginning, need, man, words, manner, gods, piety, style',)
('21 state, laws, states, athenian, evils, individual, order, power, government, origin',)
('19 athenian, polus, desire, actions, power, violence, purification, passion, sorrow, anger',)
('9 socrates, evil, difference, theodorus, cratylus, courage, nature, knowledge, cases, socrates nature',)
('6 socrates, young, young socrates, hermogenes, socrates true, true, thing thing, mixture, socrates sort, socrates case',)
('13 socrates, names, gods, rhetoric, things, art, men, view, persuasion, socrates art',)
('34 socrates, letters, music, truth, shall, imitation, words, poetry, judgment, colours',)
('17 socrates, reason, ion, subject, questions, fear, words, matter, ruler, god',)
('0 opinion, pleasure, socrates, pain, knowledge, pleasure pain, science, idea, pleasures, notion',)

Fig IV: Topic Clusters

Some interesting clusters include the following:

- Cluster with Topic IDs 25,10,1, 38, 21 and 19: This cluster discusses what it means to be 'virtuous.' They also connect the virtue of the 'polis' with themes of 'personal virtue,' a uniquely Athenian theme
- Cluster with Topic IDs 35, 15, 32 and 7: This cluster discusses 'strangers' and knowledge. This cluster will become very important in analyzing the *Theaetetus* trilogy
- Cluster with Topic IDs 2 and 4: This cluster discusses the connection between 'wisdom' and 'temperance.' This is a very interesting cluster, as it also reflects the classic Athenian belief that the springs of justice is from wisdom only

Similarities and Differences between *Theaetetus, Statesmen* and *Sophists*

An interesting issue in Plato interpretation is the understanding of the dramatic logic behind the dialogues. Plato's dialogues, apart from being philosophically dense, is also a primarily dramatic universe. There are certain recurring characters. However, the *Theaetetus, Statesmen* and *Sophist* trilogy represent something very interesting. The *Theaetetus* and the *Statesmen* are two of Plato's middle dialogues. Both concern knowledge, and the nature of applied philosophy. The *Sophist,* on the other hand, is a *Late*-period dialogue. Nevertheless, all three dialogues supposedly happen on the same day. Moreover, the same speakers (Theaetetus, Theodorus) reappear. Why is this so?

One approach I tried was sentiment analysis. This yielded some interesting results.
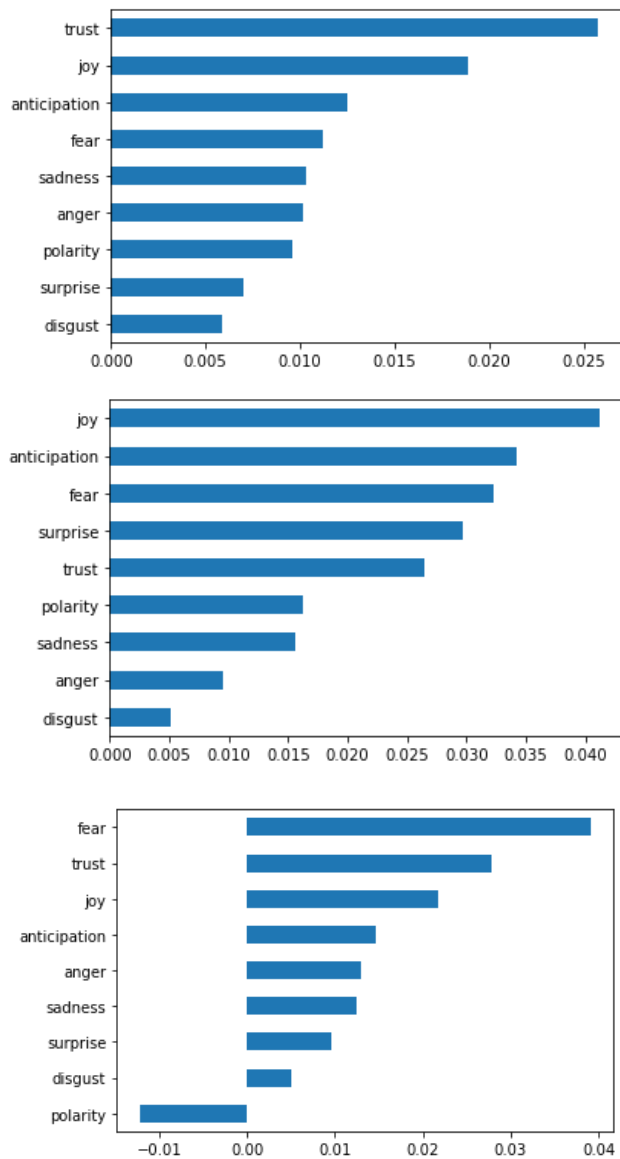
Fig V: Sentiment for Theaetetus, Statesman and Sophist respectively

The *Theaetetus* and *Statesman* are more similar to each other than they are to the *Sophist*. When compared with the *Theaetetus*, the *Statesman* has greater surprise, and lower trust. This is expected, since the *Statesman* is aimed at analyzing the differences between a person of the world and a philosopher. Something I wish to note here is polarity. Both the *Theaetetus* and *Statesman* have positive polarity, whilst the *Sophist* has negative polarity. This, again, is to be expected, as the *Sophist* attempts to distinguish between the sophist and a true philosopher. And since Plato does not hold the sophists in high regard, the polarity of the dialogue becomes something quite expected. Nevertheless, in terms of sentiment, there is still no way to reconcile the *Sophist* with the *Theaetetus/Statesman*.

Perhaps these works are closely related when it comes to PCA analysis. These were the results.

book_id	PC0	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
22 Theaetetus	-0.253645	0.056755	-0.192210	0.030813	0.010496	-0.222344	-0.148445	0.005242	0.103061	0.061912
19 Sophist	-0.156078	0.052774	-0.208607	0.054977	-0.077855	0.134631	-0.058339	0.129799	0.138576	0.129822
20 Statesman	-0.014703	-0.123959	-0.199870	-0.123955	0.029977	0.062724	0.212736	-0.001876	0.143775	0.066004

Fig V: PCA Analysis

If we only consider the first and second components, we see that the *Theaetetus* and the *Sophist* turn out to be very similar. Nevertheless, the *Statesman* differs. The words of the first component deal mostly with *certainty*, while the second component deals mainly with *lacking*. It seems that both the *Theaetetus* and *Sophist* score quite low in the first component and high in the second component. The *Statesman*, however, scores low in the first component and has a negative sign for the second component. Graphically, the *Statesman* differs more from the *Theaetetus* than the *Sophist* does. This is counter-intuitive, since the *Statesman* and the *Theaetetus* are considered to be quite similar works.

It may be so that these works do not have any similarities. Nevertheless, the results of my clustering do suggest otherwise.
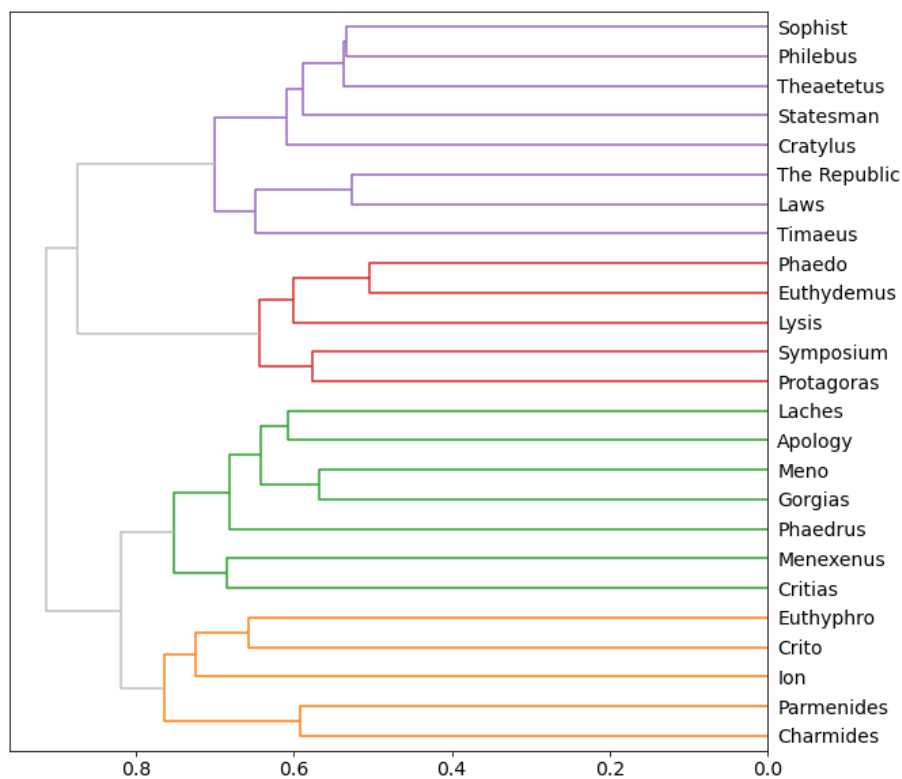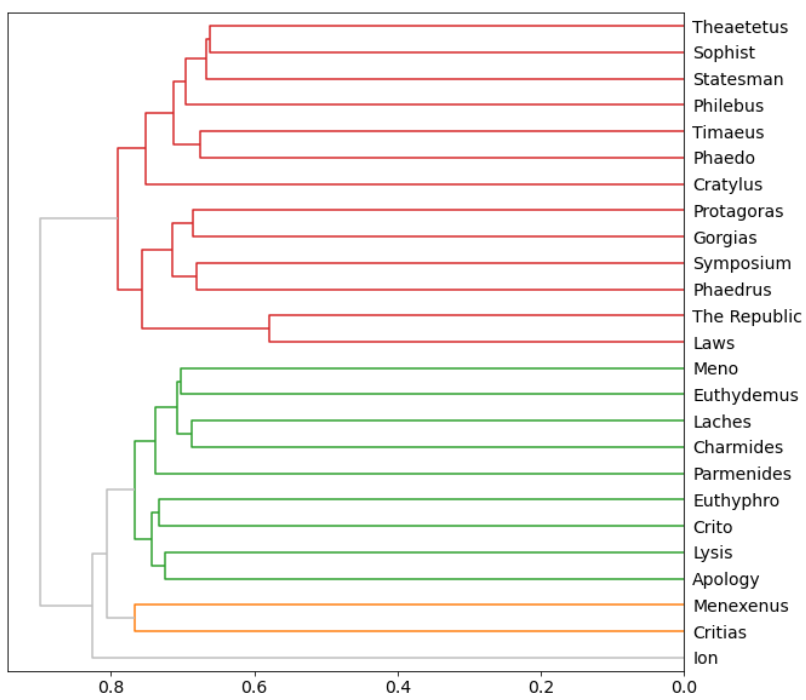
Fig VI: Cosine Clustering of Works

Fig VII: Jaccard Clustering of Works

In both the Jaccard and Cosine clusterings, we see that there exists some non-significant similarities between the three dialogues. Thus, in both high-frequency and same-frequency methods,[3] we see that the trilogy has some unifying features.

I then examined the topic models for all three works.

book_id	Theaetetus	book_id	Sophist	book_id	Statesman
37	0.172465	37	0.216663	6	0.355543
9	0.070521	32	0.138333	32	0.093429
6	0.060018	7	0.054440	35	0.076086
0	0.057496	35	0.047270	7	0.053783
34	0.042441	29	0.045821	1	0.022017

Fig IX: Top Five Topics IDs for Works

The best account I can give of the similarities between the three works can be seen in the fact that the top five topic IDs in all three works include either topic 7, 32 or 35. If we refer to our topic clustering analysis, we see that these three topics make up a cluster of topics. I shall call this cluster of topics the '*Warning'* cluster, for these clusters discuss how outsiders or 'strangers' to knowledge may corrupt philosophy. So, I postulate that the *Theaetetus, Statesman* and *Sophist* are all dialogues that attempt to define philosophical knowledge against outsiders. This is what unites the trilogy.

Conclusion and Future Extensions

I want to propose three preliminary answers to the questions raised in the Introduction.

As to the question of distinguishing between Plato's early, middle and later works, we see that Plato's early works were much less certain. As to the question of whether Plato's works converge, and if so on what topics, we see that Plato mainly talks about 'ontology.' Of course, this is related to the medieval interpretation of Plato as being an ethicist. Nevertheless, ethics was not the main focus of Plato. As to the question of the mysterious *trilogy,* I interpret them as being united in the need to properly define philosophical knowledge.

Future extensions to this project include the following:

- Linking OCHO to *Stephanus* pages, thus allowing for easy indexing for Plato scholars to chunk larger works
- Comparing Plato's corpus with Aristotle to see how each author takes a different approach, especially *vis a vis* topics

---

[3] Turney & Patel, 162

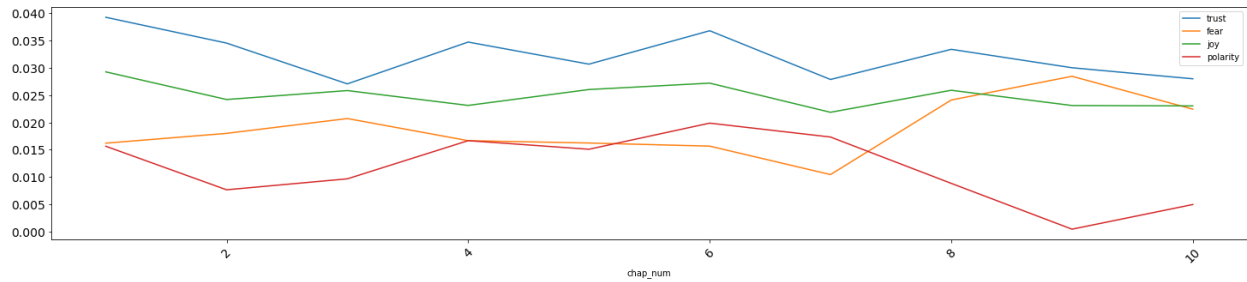**Appendix: Sentiment of *The Republic***

Fig X: *The Republic* Sentiment Analysis

The *Republic* is one of Plato's most consequential and influential works. Moreover, since it is divided into neat chapters, one can perform a book by book analysis. This is given in *Fig X,* which gives a semi-narrative picture of the varieties of sentiment within the *Republic* by book, in this case called 'chap_num.' From Book VII, we see that the sentiment of fear over takes the polarity. Interestingly, Book VII is where the cave allegory is presented. The famous allegory contributes to the negative polarity immensely, as should be expected. It introduces the doubt that the discussion the characters have been having so far might come under threat. As such, the polarity decreases, whilst the sentiment of fear increases.