

Comparing Capitalism and Communism - how are they similar and how are they different?

Exploring the similarities and differences between the philosophical schools for capitalism and communism based on texts written by Adam Smith, David Ricardo, John Maynard Keynes, Karl Marx, and Vladimir Lenin. These texts are compiled in the *History of Philosophy* dataset on [Kaggle](#).

Preliminary Data Exploration:

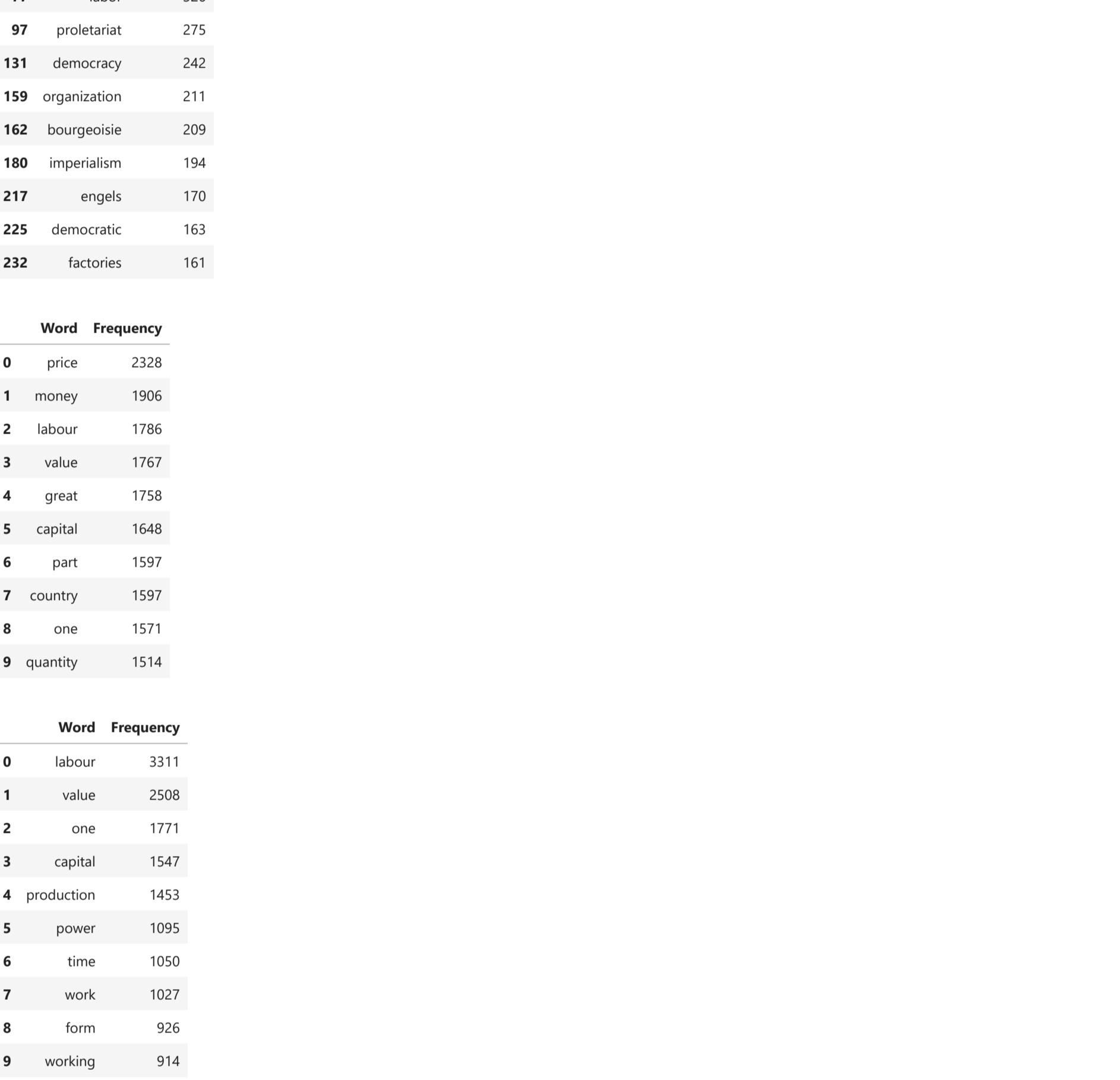
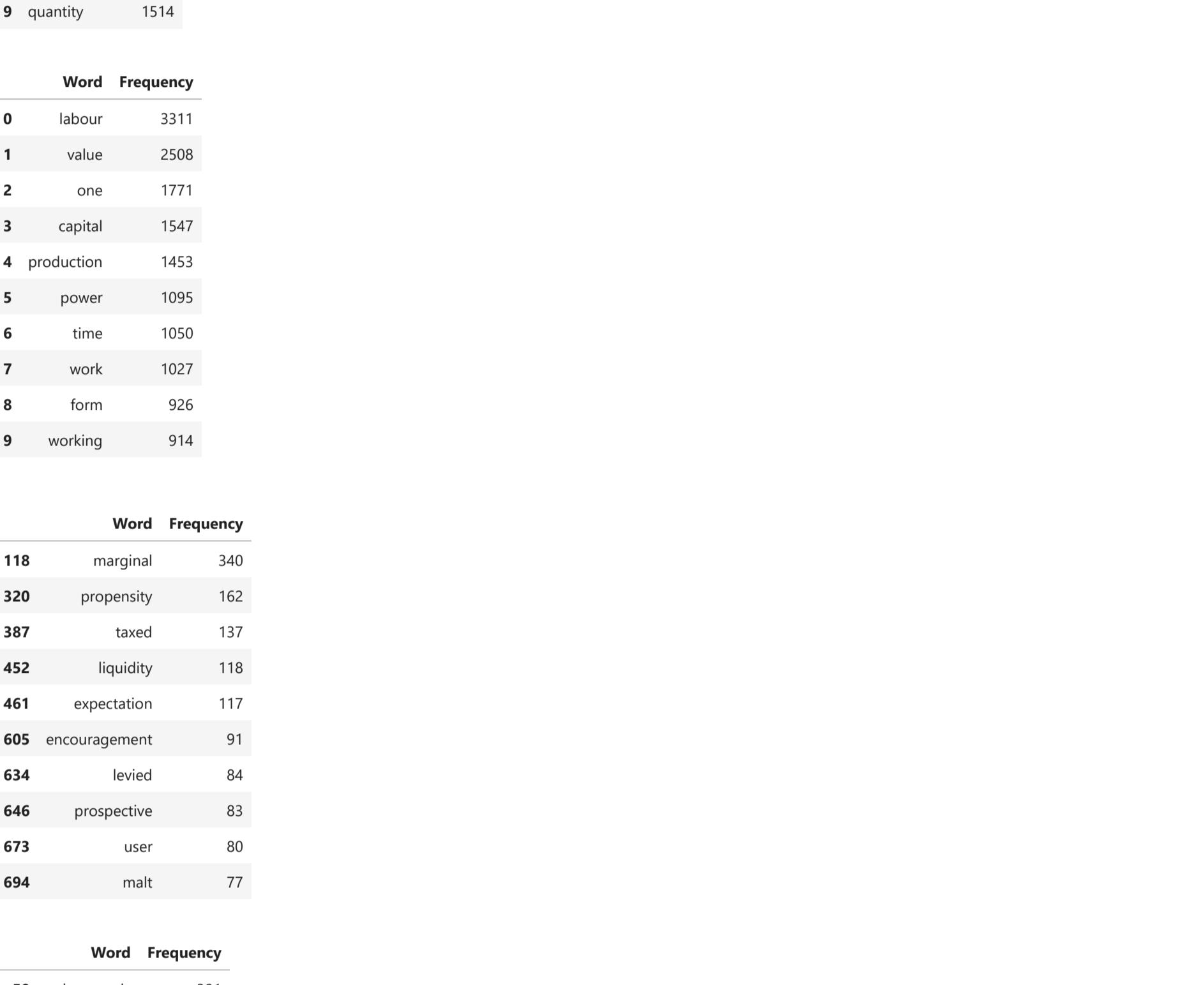
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 360808 entries, 0 to 360807
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
 0   title            360808 non-null   object  
 1   author           360808 non-null   object  
 2   school           360808 non-null   object  
 3   sentence_spacy  360808 non-null   object  
 4   sentence_str    360808 non-null   object  
 5   original_publication_date 360808 non-null   int64  
 6   corpus_edition_date 360808 non-null   int64  
 7   sentence_length  360808 non-null   int64  
 8   sentence_lowered 360808 non-null   object  
 9   tokenized_txt   360808 non-null   object  
 10  lemmatized_str 360808 non-null   object  
dtypes: int64(3), object(8)
memory usage: 30.3+ MB
```

	author	title
307896	Smith	The Wealth Of Nations
319589	Ricardo	On The Principles Of Political Economy And Tax...
322679	Keynes	A General Theory Of Employment, Interest, And ...
	author	title
289938	Marx	Capital
302934	Marx	The Communist Manifesto
303427	Lenin	Essential Works Of Lenin

Exploratory Plots:

Exploratory boxplots to show the distributions of sentence_length for capitalism and communism. They are quite similar in appearance with the IQRs of both distributions entirely below 250 tokens. There are a few very long sentences in both schools (>500 tokens).

Specifically amongst the capitalism authors, Keynes has the tightest distribution of sentence length. Otherwise their distributions appear similar. As for Marx vs Lenin, Marx has a slightly lower IQR than Lenin.



Data Preprocessing:

```
[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\Gabriel\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

Exploratory Analysis:

Top ten words used by frequency in capitalism and communism tokenized texts. Visually one can see many shared words (eg 'labour', 'one', 'value', 'capital', etc). Evidently all the top words appear somewhere in both texts (with >400 occurrences). The most commonly used word in capitalism texts that is not used in the communism texts is 'marginal'. For communism the most used word not used in capitalism is 'bourgeois'.

Word Frequency (Capitalism)

Word	Frequency
price	2328
money	1906
labour	1786
value	1767
great	1758
capital	1648
part	1597
country	1597
one	1571
quantity	1514

Word Frequency (Communism)

Word	Frequency
marginal	340
propensity	162
taxed	137
liquidity	118
expectation	117
encouragement	91
levied	84
prospective	83
user	80
malt	77

Word Frequency (Communism)

Word	Frequency
bourgeois	391
labor	326
proletariat	275
democracy	242
organization	211
bourgeoisie	209
imperialism	194
engels	170
democratic	163
factories	161

Word Frequency (Capitalism)

Word	Frequency
labour	3311
value	2508
one	1771
capital	1547
production	1453
power	1095
time	1050
work	1027
form	926
working	914

Word Frequency (Capitalism)

Word	Frequency
price	55.20%
value	23.78%
labour	21.02%

Word Frequency (Communism)

Word	Frequency
one	55.50%
value	25.66%
labour	18.83%

Word Frequency (Communism)

Word	Frequency
price	41.72%
value	39.42%
labour	18.86%

Word Frequency (Communism)

Word	Frequency
one	39.34%
value	30.50%
labour	30.05%

Sentiment Analysis:

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]   C:\Users\Gabriel\AppData\Roaming\nltk_data...
[nltk_data] Package vader_lexicon is already up-to-date!
```

Sentiment analysis of the three different capitalism authors reveals similar sentiment distributions using this +0.05 compound score threshold.

Sentiment Analysis for Keynes

Sentiment Analysis for Ricardo

Sentiment Analysis for Smith

Sentiment analysis of Marx and Lenin reveals a more significant difference. While they use similar percentage of "positive" sentiment words, Lenin uses ~10% fewer neutral words and ~12% more negative words (in absolute terms) relative to Marx. This indicates Marx may have used more critical language when describing capitalism.

Sentiment Analysis for Marx

Sentiment Analysis for Lenin

