An EDA of The Philosophy Project

In this project,I have tried to do a basic analysis of different Schools Of Thoughts and Philosophers/Authors in The Philosophy Data Project.Hope you enjoy reading!

In [38]:

```python
#Importing Packages
from google.colab import drive
drive.mount('/content/gdrive')
```

Drive already mounted at /content/gdrive; to attempt to forcibly remount, call drive.mount("/content/gdrive", force_remount=True).

In [25]:

```python
import pandas as pd
df=pd.read_csv('/content/gdrive/MyDrive/philosophy_data.csv')
```

In [26]:

```python
#Importing Packages

import numpy as np
import pandas as pd
import os
import matplotlib.pyplot as plt
import seaborn as sns
! pip install wordcloud
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
from PIL import Image
```

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: wordcloud in /usr/local/lib/python3.7/dist-packages (1.8.2.2)
Requirement already satisfied: numpy>=1.6.1 in /usr/local/lib/python3.7/dist-packages (from wordcloud) (1.21.6)
Requirement already satisfied: pillow in /usr/local/lib/python3.7/dist-packages (from wordcloud) (7.1.2)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from wordcloud) (3.2.2)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->wordcloud) (1.4.4)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packages (from kiwisolver>=1.0.1->matplotlib->wordcloud) (4.1.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from python-dateutil>=2.1->matplotlib->wordcloud) (1.15.0)

In [27]:

```python
display(df)
```

	title	author	school	sentence_spacy	sentence_str	original_publication_date	corpus_edition_date	sentenc
0	Plato - Complete Works	Plato	plato	What's new, Socrates, to make you leave your	What's new, Socrates, to make you leave your	-350	1997	

	title	author	school	your ... sentence_spacy	your ... sentence_str	original_publication_date	corpus_edition_date	sentence
1	Plato - Complete Works	Plato	plato	Surely you are not prosecuting anyone before t...	Surely you are not prosecuting anyone before t...	-350	1997	
2	Plato - Complete Works	Plato	plato	The Athenians do not call this a prosecution b...	The Athenians do not call this a prosecution b...	-350	1997	
3	Plato - Complete Works	Plato	plato	What is this you say?	What is this you say?	-350	1997	
4	Plato - Complete Works	Plato	plato	Someone must have indicted you, for you are no...	Someone must have indicted you, for you are no...	-350	1997	
...	...	...	...	...	...	...	...	
360803	Women, Race, And Class	Davis	feminism	But the socialization of housework including m...	But the socialization of housework including m...	1981	1981	
360804	Women, Race, And Class	Davis	feminism	The only significant steps toward endingdomest...	The only significant steps toward endingdomest...	1981	1981	
360805	Women, Race, And Class	Davis	feminism	Working women, therefore, have a special and v...	Working women, therefore, have a special and v...	1981	1981	
360806	Women, Race, And Class	Davis	feminism	Moreover, under capitalism, campaigns for jobs...	Moreover, under capitalism, campaigns for jobs...	1981	1981	
360807	Women, Race, And Class	Davis	feminism	This strategy calls into question the validity...	This strategy calls into question the validity...	1981	1981	

360808 rows × 11 columns

In [28]:

```
t=df.groupby("title")
t.describe()
```

Out[28]:

	original_publication_date								corpus_edition_date					sentence_le	
title	count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	max	count	me
A General Theory Of Employment, Interest, And Money	3411.0	1936.0	0.0	1936.0	1936.0	1936.0	1936.0	1936.0	3411.0	2003.0	...	2003.0	2003.0	3411.0	19
A Treatise Concerning The Principles Of Human Knowledge	1040.0	1710.0	0.0	1710.0	1710.0	1710.0	1710.0	1710.0	1040.0	2009.0	...	2009.0	2009.0	1040.0	18
A Treatise Of	7047.0	1739.0	0.0	1739.0	1739.0	1739.0	1739.0	1739.0	7047.0	2003.0	...	2003.0	2003.0	7047.0	18

title	original_publication_date								corpus_edition_date					sentence_le...	
	count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	max	count	m...
Human Nature															
Anti-Oedipus	6679.0	1972.0	0.0	1972.0	1972.0	1972.0	1972.0	1972.0	6679.0	1997.0	...	1997.0	1997.0	6679.0	16
Aristotle - Complete Works	48779.0	-320.0	0.0	-320.0	-320.0	-320.0	-320.0	-320.0	48779.0	1991.0	...	1991.0	1991.0	48779.0	15
Being And Time	8505.0	1927.0	0.0	1927.0	1927.0	1927.0	1927.0	1927.0	8505.0	1996.0	...	1996.0	1996.0	8505.0	12
Beyond Good And Evil	1906.0	1886.0	0.0	1886.0	1886.0	1886.0	1886.0	1886.0	1906.0	2003.0	...	2003.0	2003.0	1906.0	18
Capital	12996.0	1883.0	0.0	1883.0	1883.0	1883.0	1883.0	1883.0	12996.0	1887.0	...	1887.0	1887.0	12996.0	14
Critique Of Judgement	4204.0	1790.0	0.0	1790.0	1790.0	1790.0	1790.0	1790.0	4204.0	2007.0	...	2007.0	2007.0	4204.0	21
Critique Of Practical Reason	2452.0	1788.0	0.0	1788.0	1788.0	1788.0	1788.0	1788.0	2452.0	2002.0	...	2002.0	2002.0	2452.0	17
Critique Of Pure Reason	7472.0	1781.0	0.0	1781.0	1781.0	1781.0	1781.0	1781.0	7472.0	1998.0	...	1998.0	1998.0	7472.0	19
Dialogues Concerning Natural Religion	1265.0	1779.0	0.0	1779.0	1779.0	1779.0	1779.0	1779.0	1265.0	2009.0	...	2009.0	2009.0	1265.0	16
Difference And Repetition	5861.0	1968.0	0.0	1968.0	1968.0	1968.0	1968.0	1968.0	5861.0	1994.0	...	1994.0	1994.0	5861.0	16
Discourse On Method	340.0	1637.0	0.0	1637.0	1637.0	1637.0	1637.0	1637.0	340.0	2008.0	...	2008.0	2008.0	340.0	37
Ecce Homo	1504.0	1888.0	0.0	1888.0	1888.0	1888.0	1888.0	1888.0	1504.0	2016.0	...	2016.0	2016.0	1504.0	13
Elements Of The Philosophy Of Right	4923.0	1820.0	0.0	1820.0	1820.0	1820.0	1820.0	1820.0	4923.0	1991.0	...	1991.0	1991.0	4923.0	16
Enchiridion	323.0	125.0	0.0	125.0	125.0	125.0	125.0	125.0	323.0	2014.0	...	2014.0	2014.0	323.0	11
Essay Concerning Human Understanding	7742.0	1689.0	0.0	1689.0	1689.0	1689.0	1689.0	1689.0	7742.0	2004.0	...	2004.0	2004.0	7742.0	19
Essential Works Of Lenin	4469.0	1862.0	0.0	1862.0	1862.0	1862.0	1862.0	1862.0	4469.0	1966.0	...	1966.0	1966.0	4469.0	18
Ethics	3304.0	1677.0	0.0	1677.0	1677.0	1677.0	1677.0	1677.0	3304.0	2003.0	...	2003.0	2003.0	3304.0	14
History Of Madness	8033.0	1961.0	0.0	1961.0	1961.0	1961.0	1961.0	1961.0	8033.0	2006.0	...	2006.0	2006.0	8033.0	17
Lewis - Papers	13120.0	1985.0	0.0	1985.0	1985.0	1985.0	1985.0	1985.0	13120.0	2008.0	...	2008.0	2008.0	13120.0	10
Meditations	2212.0	170.0	0.0	170.0	170.0	170.0	170.0	170.0	2212.0	2008.0	...	2008.0	2008.0	2212.0	13
Meditations On First Philosophy	792.0	1641.0	0.0	1641.0	1641.0	1641.0	1641.0	1641.0	792.0	2008.0	...	2008.0	2008.0	792.0	19
Naming And Necessity	2681.0	1972.0	0.0	1972.0	1972.0	1972.0	1972.0	1972.0	2681.0	1990.0	...	1990.0	1990.0	2681.0	12
Off The Beaten Track	6734.0	1950.0	0.0	1950.0	1950.0	1950.0	1950.0	1950.0	6734.0	2001.0	...	2001.0	2001.0	6734.0	10
On Certainty	1984.0	1950.0	0.0	1950.0	1950.0	1950.0	1950.0	1950.0	1984.0	1969.0	...	1969.0	1969.0	1984.0	7
On The Improvement Of Understanding	489.0	1677.0	0.0	1677.0	1677.0	1677.0	1677.0	1677.0	489.0	1997.0	...	1997.0	1997.0	489.0	17
On The Principles Of Political Economy And Taxation	3090.0	1817.0	0.0	1817.0	1817.0	1817.0	1817.0	1817.0	3090.0	2010.0	...	2010.0	2010.0	3090.0	18
Philosophical Investigations	5838.0	1953.0	0.0	1953.0	1953.0	1953.0	1953.0	1953.0	5838.0	1986.0	...	1986.0	1986.0	5838.0	8
Philosophical	3668.0	1910.0	0.0	1910.0	1910.0	1910.0	1910.0	1910.0	3668.0	2015.0	...	2015.0	2015.0	3668.0	16

| title | original_publication_date | | | | | | | | corpus_edition_date | | ... | | | sentence_le | |
	count	mean	std	min	25%	50%	75%	max	count	mean	...	25%	max	count	m
Plato - Complete Works	38366.0	-350.0	0.0	-350.0	-350.0	-350.0	-350.0	-350.0	38366.0	1997.0	...	1997.0	1997.0	38366.0	11
Quintessence	7373.0	1950.0	0.0	1950.0	1950.0	1950.0	1950.0	1950.0	7373.0	2004.0	...	2004.0	2004.0	7373.0	12
Science Of Logic	10678.0	1817.0	0.0	1817.0	1817.0	1817.0	1817.0	1817.0	10678.0	2010.0	...	2010.0	2010.0	10678.0	18
Second Treatise On Government	1143.0	1689.0	0.0	1689.0	1689.0	1689.0	1689.0	1689.0	1143.0	2010.0	...	2010.0	2010.0	1143.0	26
The Analysis Of Mind	3513.0	1921.0	0.0	1921.0	1921.0	1921.0	1921.0	1921.0	3513.0	2008.0	...	2008.0	2008.0	3513.0	14
The Antichrist	1170.0	1888.0	0.0	1888.0	1888.0	1888.0	1888.0	1888.0	1170.0	2006.0	...	2006.0	2006.0	1170.0	13
The Birth Of The Clinic	2518.0	1963.0	0.0	1963.0	1963.0	1963.0	1963.0	1963.0	2518.0	2003.0	...	2003.0	2003.0	2518.0	18
The Communist Manifesto	493.0	1848.0	0.0	1848.0	1848.0	1848.0	1848.0	1848.0	493.0	1970.0	...	1970.0	1970.0	493.0	15
The Crisis Of The European Sciences And Phenomenology	4832.0	1936.0	0.0	1936.0	1936.0	1936.0	1936.0	1936.0	4832.0	1970.0	...	1970.0	1970.0	4832.0	19
The Idea Of Phenomenology	910.0	1907.0	0.0	1907.0	1907.0	1907.0	1907.0	1907.0	910.0	1999.0	...	1999.0	1999.0	910.0	15
The Logic Of Scientific Discovery	4678.0	1959.0	0.0	1959.0	1959.0	1959.0	1959.0	1959.0	4678.0	2002.0	...	2002.0	2002.0	4678.0	13
The Order Of Things	4689.0	1966.0	0.0	1966.0	1966.0	1966.0	1966.0	1966.0	4689.0	2002.0	...	2002.0	2002.0	4689.0	21
The Phenomenology Of Perception	7592.0	1945.0	0.0	1945.0	1945.0	1945.0	1945.0	1945.0	7592.0	2002.0	...	2002.0	2002.0	7592.0	17
The Phenomenology Of Spirit	7099.0	1807.0	0.0	1807.0	1807.0	1807.0	1807.0	1807.0	7099.0	1977.0	...	1977.0	1977.0	7099.0	16
The Problems Of Philosophy	1560.0	1912.0	0.0	1912.0	1912.0	1912.0	1912.0	1912.0	1560.0	2004.0	...	2004.0	2004.0	1560.0	15
The Search After Truth	12997.0	1674.0	0.0	1674.0	1674.0	1674.0	1674.0	1674.0	12997.0	1997.0	...	1997.0	1997.0	12997.0	16
The Second Sex	13017.0	1949.0	0.0	1949.0	1949.0	1949.0	1949.0	1949.0	13017.0	2009.0	...	2009.0	2009.0	13017.0	14
The System Of Ethics	5308.0	1798.0	0.0	1798.0	1798.0	1798.0	1798.0	1798.0	5308.0	2005.0	...	2005.0	2005.0	5308.0	15
The Wealth Of Nations	11693.0	1776.0	0.0	1776.0	1776.0	1776.0	1776.0	1776.0	11693.0	2009.0	...	2009.0	2009.0	11693.0	18
Theodicy	5027.0	1710.0	0.0	1710.0	1710.0	1710.0	1710.0	1710.0	5027.0	2005.0	...	2005.0	2005.0	5027.0	15
Three Dialogues	1694.0	1713.0	0.0	1713.0	1713.0	1713.0	1713.0	1713.0	1694.0	2009.0	...	2009.0	2009.0	1694.0	11
Thus Spake Zarathustra	5916.0	1887.0	0.0	1887.0	1887.0	1887.0	1887.0	1887.0	5916.0	2008.0	...	2008.0	2008.0	5916.0	8
Tractatus Logico-Philosophicus	1212.0	1921.0	0.0	1921.0	1921.0	1921.0	1921.0	1921.0	1212.0	2001.0	...	2001.0	2001.0	1212.0	10
Twilight Of The Idols	3052.0	1888.0	0.0	1888.0	1888.0	1888.0	1888.0	1888.0	3052.0	2016.0	...	2016.0	2016.0	3052.0	12
Vindication Of The Rights Of Woman	2559.0	1792.0	0.0	1792.0	1792.0	1792.0	1792.0	1792.0	2559.0	2001.0	...	2001.0	2001.0	2559.0	19
Women, Race,	3059.0	1981.0	0.0	1981.0	1981.0	1981.0	1981.0	1981.0	3059.0	1981.0	...	1981.0	1981.0	3059.0	13

59 rows × 24 columns

`In [29]:`

```
a=df.groupby("author")
a.describe()
```

`Out[29]:`

	original_publication_date								corpus_edition_date		...		
	count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	max
author													
Aristotle	48779.0	-320.000000	0.000000	-320.0	-320.0	-320.0	-320.0	-320.0	48779.0	1991.000000	...	1991.0	199
Beauvoir	13017.0	1949.000000	0.000000	1949.0	1949.0	1949.0	1949.0	1949.0	13017.0	2009.000000	...	2009.0	200
Berkeley	2734.0	1711.858815	1.456718	1710.0	1710.0	1713.0	1713.0	1713.0	2734.0	2009.000000	...	2009.0	200
Davis	3059.0	1981.000000	0.000000	1981.0	1981.0	1981.0	1981.0	1981.0	3059.0	1981.000000	...	1981.0	198
Deleuze	12540.0	1970.130463	1.995820	1968.0	1968.0	1972.0	1972.0	1972.0	12540.0	1995.597847	...	1997.0	199
Derrida	5999.0	1967.000000	0.000000	1967.0	1967.0	1967.0	1967.0	1967.0	5999.0	2001.000000	...	2001.0	200
Descartes	1132.0	1639.798587	1.834457	1637.0	1637.0	1641.0	1641.0	1641.0	1132.0	2008.000000	...	2008.0	200
Epictetus	323.0	125.000000	0.000000	125.0	125.0	125.0	125.0	125.0	323.0	2014.000000	...	2014.0	201
Fichte	5308.0	1798.000000	0.000000	1798.0	1798.0	1798.0	1798.0	1798.0	5308.0	2005.000000	...	2005.0	200
Foucault	15240.0	1962.868832	2.204679	1961.0	1961.0	1961.0	1966.0	1966.0	15240.0	2004.273622	...	2006.0	200
Hegel	22700.0	1814.523304	5.205012	1807.0	1807.0	1817.0	1817.0	1820.0	22700.0	1995.559295	...	2010.0	201
Heidegger	15239.0	1937.163528	11.422452	1927.0	1927.0	1927.0	1950.0	1950.0	15239.0	1998.209463	...	2001.0	200
Hume	8312.0	1745.087584	14.369048	1739.0	1739.0	1739.0	1739.0	1779.0	8312.0	2003.913138	...	2003.0	200
Husserl	5742.0	1931.404040	10.591483	1907.0	1936.0	1936.0	1936.0	1936.0	5742.0	1974.595960	...	1970.0	199
Kant	14128.0	1784.892978	4.177676	1781.0	1781.0	1781.0	1790.0	1790.0	14128.0	2001.372310	...	2007.0	200
Keynes	3411.0	1936.000000	0.000000	1936.0	1936.0	1936.0	1936.0	1936.0	3411.0	2003.000000	...	2003.0	200
Kripke	12479.0	1974.355477	1.232185	1972.0	1975.0	1975.0	1975.0	1975.0	12479.0	2006.488340	...	2011.0	201
Leibniz	5027.0	1710.000000	0.000000	1710.0	1710.0	1710.0	1710.0	1710.0	5027.0	2005.000000	...	2005.0	200
Lenin	4469.0	1862.000000	0.000000	1862.0	1862.0	1862.0	1862.0	1862.0	4469.0	1966.000000	...	1966.0	196
Lewis	13120.0	1985.000000	0.000000	1985.0	1985.0	1985.0	1985.0	1985.0	13120.0	2008.000000	...	2008.0	200
Locke	8885.0	1689.000000	0.000000	1689.0	1689.0	1689.0	1689.0	1689.0	8885.0	2004.771863	...	2004.0	201
Malebranche	12997.0	1674.000000	0.000000	1674.0	1674.0	1674.0	1674.0	1674.0	12997.0	1997.000000	...	1997.0	199
Marcus Aurelius	2212.0	170.000000	0.000000	170.0	170.0	170.0	170.0	170.0	2212.0	2008.000000	...	2008.0	200
Marx	13489.0	1881.720810	6.567993	1848.0	1883.0	1883.0	1883.0	1883.0	13489.0	1890.033509	...	1887.0	197
Merleau-Ponty	7592.0	1945.000000	0.000000	1945.0	1945.0	1945.0	1945.0	1945.0	7592.0	2002.000000	...	2002.0	200
Moore	3668.0	1910.000000	0.000000	1910.0	1910.0	1910.0	1910.0	1910.0	3668.0	2015.000000	...	2015.0	201
Nietzsche	13548.0	1887.281960	0.695604	1886.0	1887.0	1887.0	1888.0	1888.0	13548.0	2009.814142	...	2016.0	201
Plato	38366.0	-350.000000	0.000000	-350.0	-350.0	-350.0	-350.0	-350.0	38366.0	1997.000000	...	1997.0	199
Popper	4678.0	1959.000000	0.000000	1959.0	1959.0	1959.0	1959.0	1959.0	4678.0	2002.000000	...	2002.0	200
Quine	7373.0	1950.000000	0.000000	1950.0	1950.0	1950.0	1950.0	1950.0	7373.0	2004.000000	...	2004.0	200
Ricardo	3090.0	1817.000000	0.000000	1817.0	1817.0	1817.0	1817.0	1817.0	3090.0	2010.000000	...	2010.0	201

author	original_publication_date								corpus_edition_date				
	count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	max
Russell	5073.0	1918.232407	4.153573	1912.0	1912.0	1921.0	1921.0	1921.0	5073.0	2006.769959	...	2008.0	200
Smith	11693.0	1776.000000	0.000000	1776.0	1776.0	1776.0	1776.0	1776.0	11693.0	2009.000000	...	2009.0	200
Spinoza	3793.0	1677.000000	0.000000	1677.0	1677.0	1677.0	1677.0	1677.0	3793.0	2002.226470	...	2003.0	200
Wittgenstein	9034.0	1948.048041	10.716670	1921.0	1950.0	1953.0	1953.0	1953.0	9034.0	1984.278946	...	1986.0	200
Wollstonecraft	2559.0	1792.000000	0.000000	1792.0	1792.0	1792.0	1792.0	1792.0	2559.0	2001.000000	...	2001.0	200

36 rows × 24 columns

In [30]:

```python
s=df.groupby("school")
s.describe()
```

Out[30]:

| school	original_publication_date								corpus_edition_date			
	count	mean	std	min	25%	50%	75%	max	count	mean	...	75%
analytic	55425.0	1958.655354	23.978784	1910.0	1950.0	1959.0	1975.0	1985.0	55425.0	2003.105386	...	2008.0
aristotle	48779.0	-320.000000	0.000000	-320.0	-320.0	-320.0	-320.0	-320.0	48779.0	1991.000000	...	1991.0
capitalism	18194.0	1812.959987	60.984677	1776.0	1776.0	1776.0	1817.0	1936.0	18194.0	2008.044960	...	2009.0
communism	17958.0	1876.813120	10.252063	1848.0	1862.0	1883.0	1883.0	1883.0	17958.0	1908.938412	...	1966.0
continental	33779.0	1966.298292	3.810169	1961.0	1963.0	1967.0	1968.0	1972.0	33779.0	2000.471476	...	2003.0
empiricism	19931.0	1715.526316	27.683636	1689.0	1689.0	1713.0	1739.0	1779.0	19931.0	2004.993728	...	2009.0
feminism	18635.0	1932.693319	57.334263	1792.0	1949.0	1949.0	1949.0	1981.0	18635.0	2003.305125	...	2009.0
german_idealism	42136.0	1802.506906	14.311706	1781.0	1790.0	1807.0	1817.0	1820.0	42136.0	1998.697646	...	2010.0
nietzsche	13548.0	1887.281960	0.695604	1886.0	1887.0	1887.0	1888.0	1888.0	13548.0	2009.814142	...	2016.0
phenomenology	28573.0	1938.088300	10.688899	1907.0	1927.0	1945.0	1945.0	1950.0	28573.0	1994.471284	...	2002.0
plato	38366.0	-350.000000	0.000000	-350.0	-350.0	-350.0	-350.0	-350.0	38366.0	1997.000000	...	1997.0
rationalism	22949.0	1680.694627	17.275349	1637.0	1674.0	1674.0	1677.0	1710.0	22949.0	2000.158830	...	2005.0
stoicism	2535.0	164.266272	15.007697	125.0	170.0	170.0	170.0	170.0	2535.0	2008.764497	...	2008.0

13 rows × 24 columns

In [31]:

```python
#analyzing the different parameters author and school

parameters = ["author", "school"]

for p in parameters:
    plt.figure(figsize=(16,5))
    df[p].value_counts().plot(kind='bar')
    plt.title(p)
    plt.grid()
    plt.show()
```

school
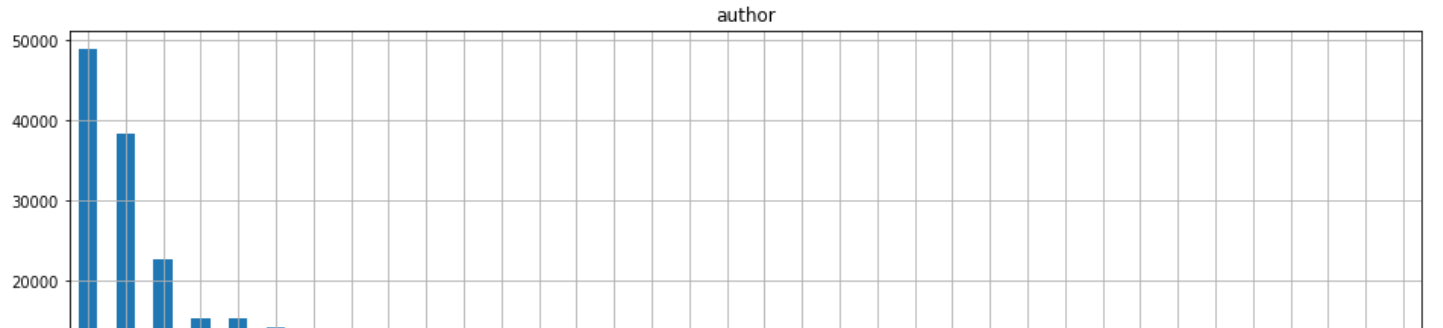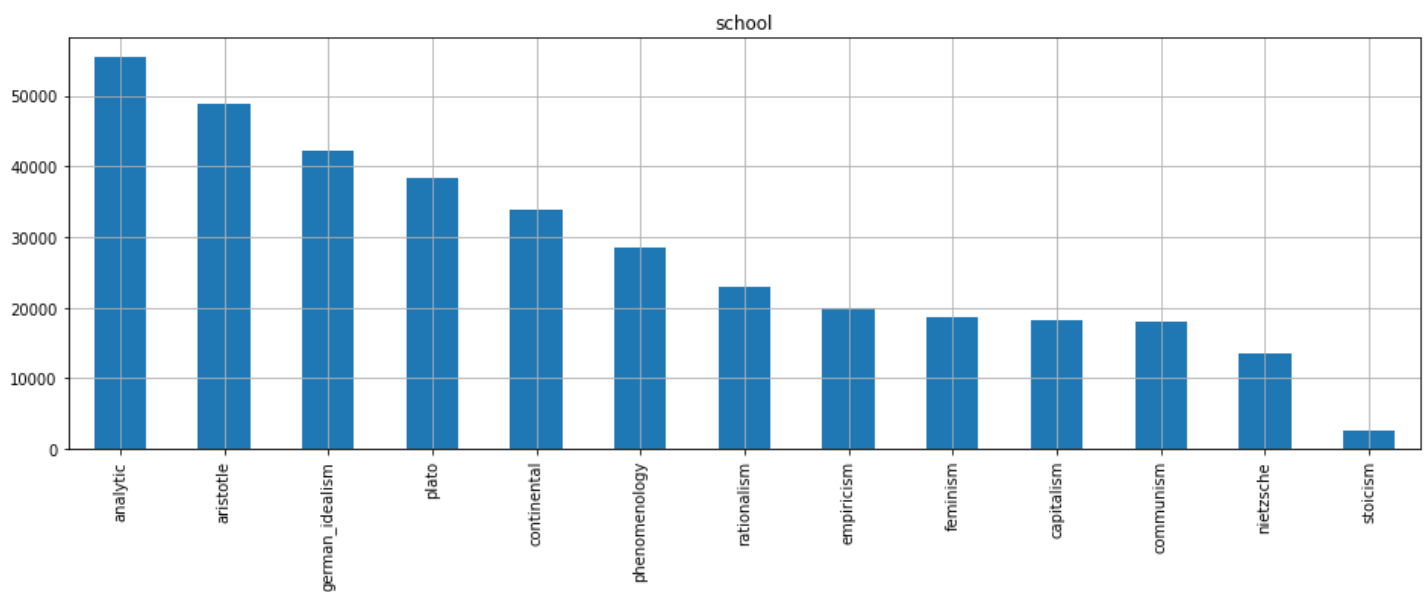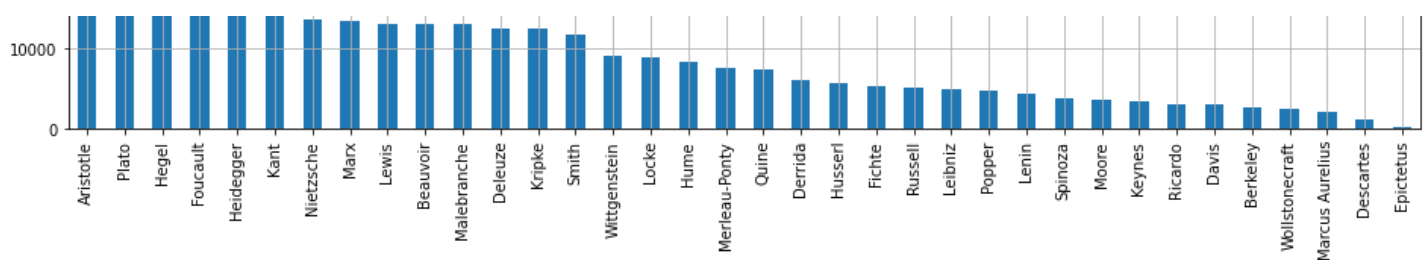
In [32]:

```
#creating heat map to see correlation between all the parameters
sns.heatmap(df.corr(), annot=True)
```

Out[32]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9d08cef310>
```
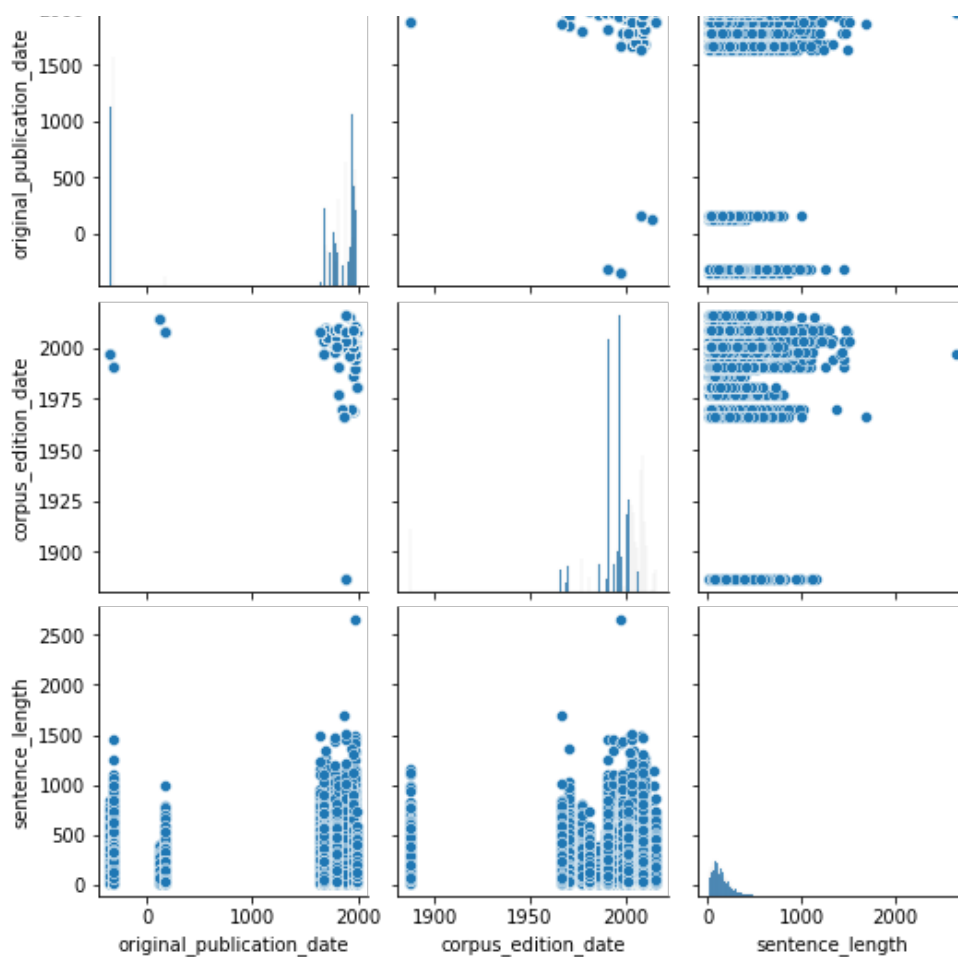

In [ ]:

```
#making a pairplot to see which parameters have an effect on each other
sns.pairplot(df)
```
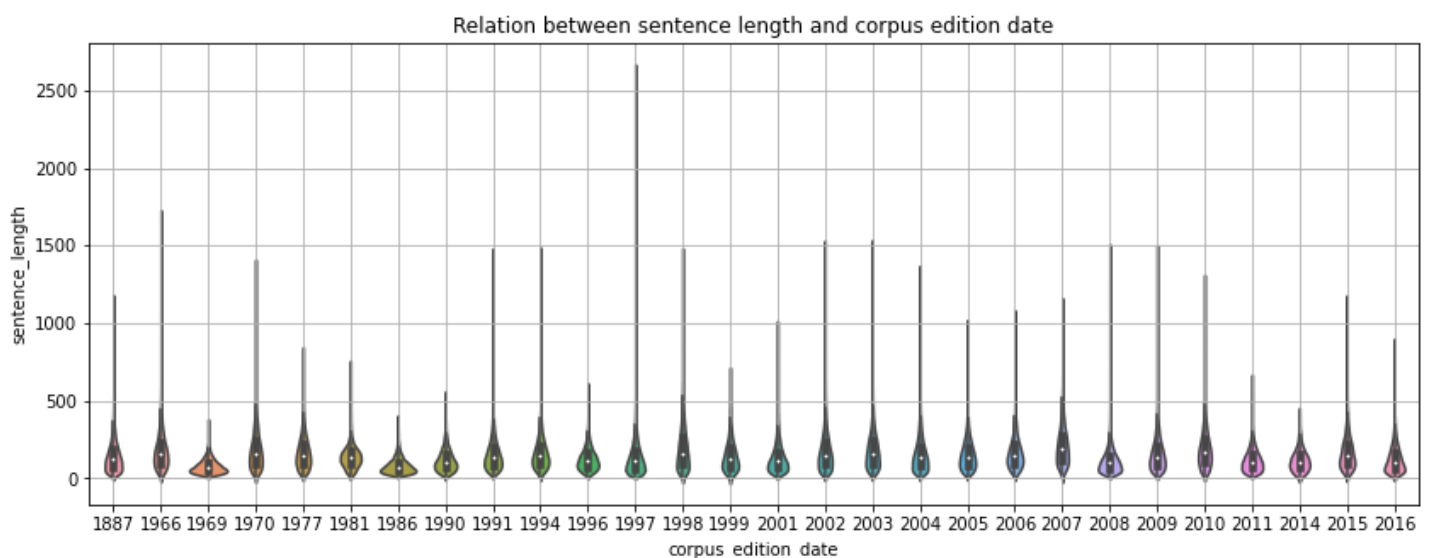
Out[ ]:

```
<seaborn.axisgrid.PairGrid at 0x7f9d09da8f90>
```

In [ ]:

```
#Violin plots show how sentence length corresponds to which year a particular school was
predominant in

plt.figure(figsize=(14,5))
sns.violinplot(x="corpus_edition_date", y="sentence_length", data=df)
plt.title("Relation between sentence length and corpus edition date")
plt.grid()
```


In [ ]:

```
plt.figure(figsize=(20,5))
sns.violinplot(x="school", y="sentence_length", data=df)
plt.title("Relation between sentence length and school")
plt.grid()
```
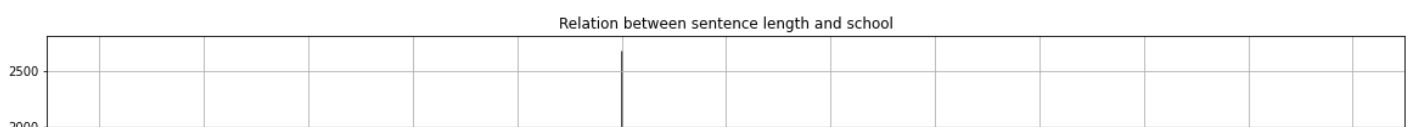
In [ ]:

```
#Boxplot confirms findings from before

plt.figure(figsize=(20,5))
sns.boxplot(x="school", y="corpus_edition_date", data=df)
plt.title("Corpus edition date by school")
plt.grid()
```


In [ ]:

```
#Comparison between violin plots show relation between sentence length and number of tokens

df["tokens"] = list(map(len,map(eval,df.tokenized_txt)))
plt.figure(figsize=(16,5))
sns.violinplot(x="corpus_edition_date", y="tokens", data=df)
plt.title("Relation between number of tokens and corpus edition date")
plt.grid()
```
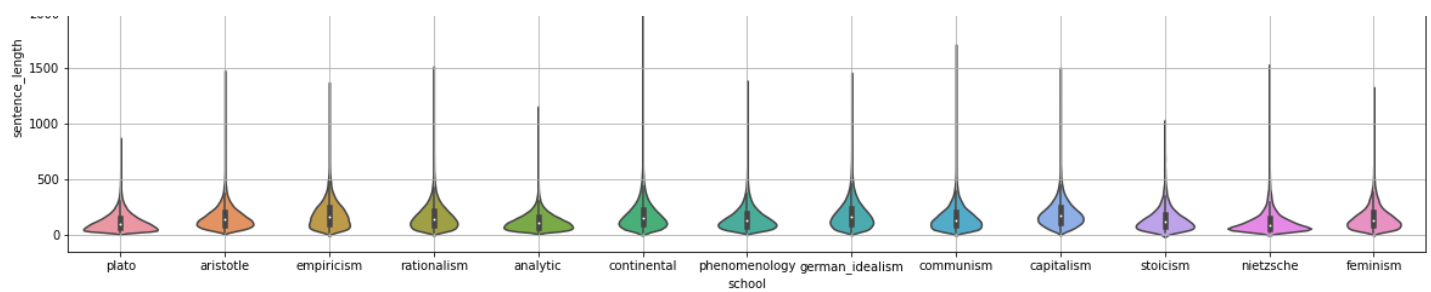

In [ ]:
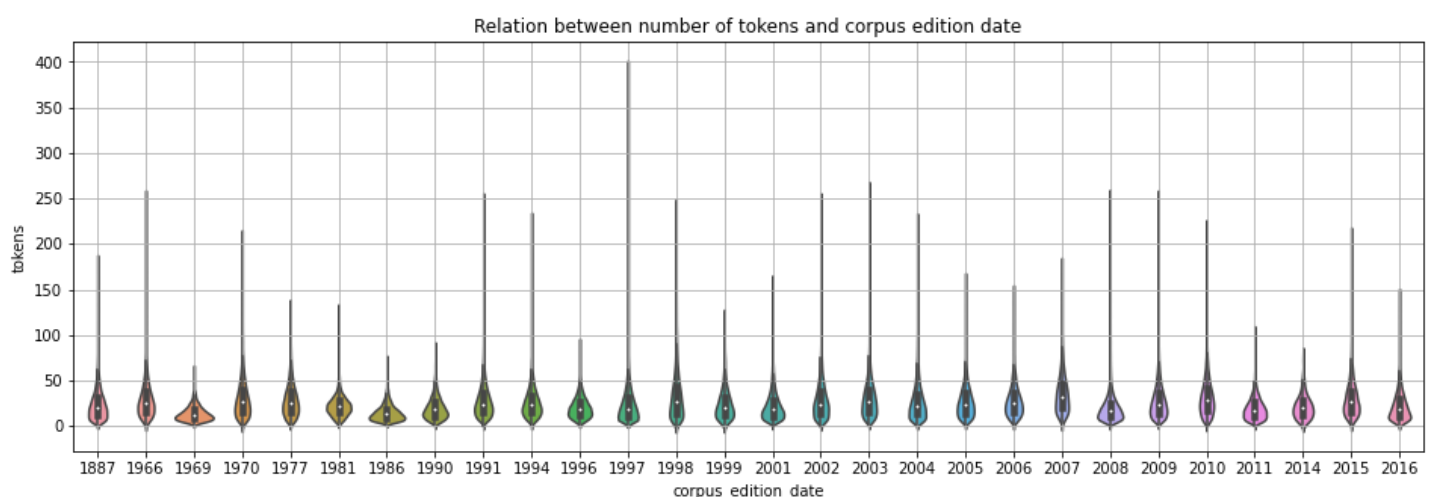
```
plt.figure(figsize=(20,5))
sns.violinplot(x="school", y="tokens", data=df)
plt.title("Relation between number of tokens and school")
plt.grid()
```
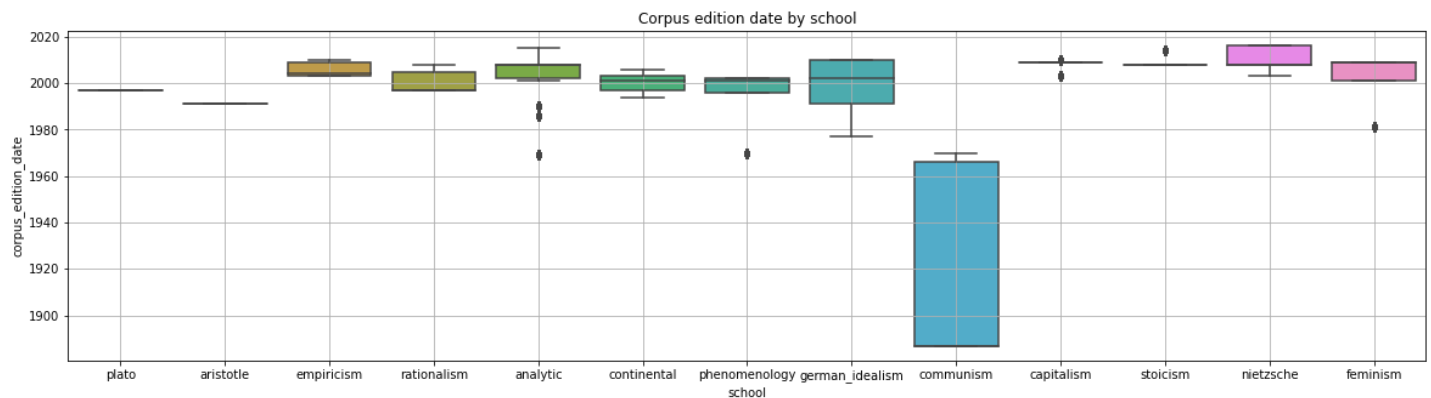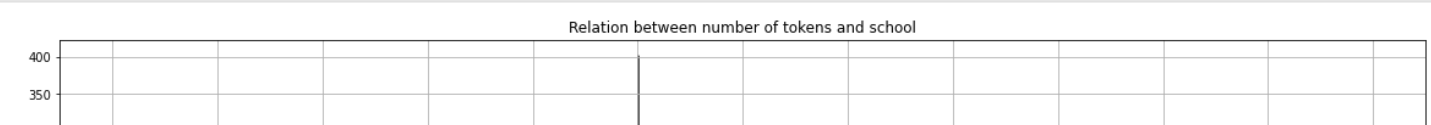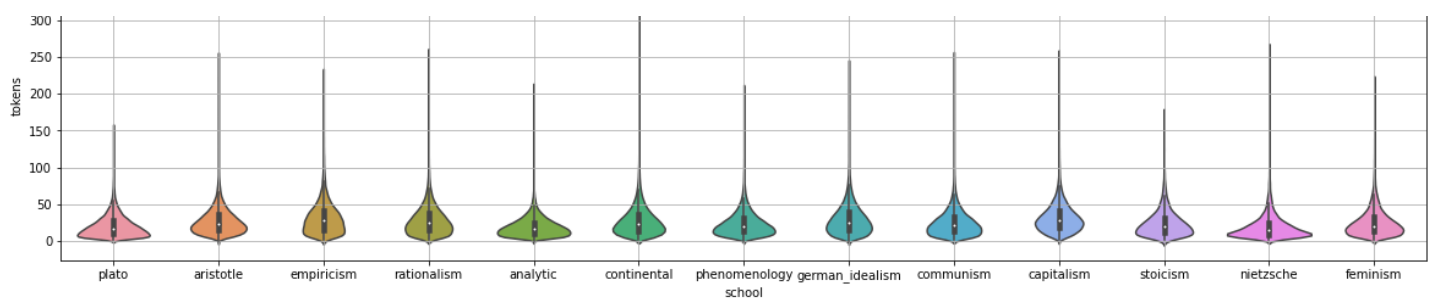
In [ ]:

```
#finding mean values corresponding to title, author and school
df.groupby(by=["title","author","school"]).mean().round(2)
```

Out[ ]:

title	author	school	original_publication_date	corpus_edition_date	sentence_length	tokens
A General Theory Of Employment, Interest, And Money	Keynes	capitalism	1936.0	2003.0	196.65	32.60
A Treatise Concerning The Principles Of Human Knowledge	Berkeley	empiricism	1710.0	2009.0	184.72	32.15
A Treatise Of Human Nature	Hume	empiricism	1739.0	2003.0	183.01	30.74
Anti-Oedipus	Deleuze	continental	1972.0	1997.0	165.51	26.64
Aristotle - Complete Works	Aristotle	aristotle	-320.0	1991.0	153.22	27.51
Being And Time	Heidegger	phenomenology	1927.0	1996.0	126.47	20.95
Beyond Good And Evil	Nietzsche	nietzsche	1886.0	2003.0	188.08	31.44
Capital	Marx	communism	1883.0	1887.0	142.97	23.89
Critique Of Judgement	Kant	german_idealism	1790.0	2007.0	211.98	35.34
Critique Of Practical Reason	Kant	german_idealism	1788.0	2002.0	175.38	28.97
Critique Of Pure Reason	Kant	german_idealism	1781.0	1998.0	197.86	32.47
Dialogues Concerning Natural Religion	Hume	empiricism	1779.0	2009.0	164.51	27.27
Difference And Repetition	Deleuze	continental	1968.0	1994.0	161.58	25.87
Discourse On Method	Descartes	rationalism	1637.0	2008.0	375.60	65.16
Ecce Homo	Nietzsche	nietzsche	1888.0	2016.0	133.98	23.09
Elements Of The Philosophy Of Right	Hegel	german_idealism	1820.0	1991.0	161.01	26.66
Enchiridion	Epictetus	stoicism	125.0	2014.0	118.43	21.41
Essay Concerning Human Understanding	Locke	empiricism	1689.0	2004.0	190.59	33.60
Essential Works Of Lenin	Lenin	communism	1862.0	1966.0	181.42	28.91
Ethics	Spinoza	rationalism	1677.0	2003.0	142.07	24.90

title	author	school	original_publication_date	corpus_edition_date	sentence_length	tokens
History Of Madness	Foucault	continental	1961.0	2006.0	174.42	28.92
Lewis - Papers	Lewis	analytic	1985.0	2008.0	109.72	18.19
Meditations	Marcus Aurelius	stoicism	170.0	2008.0	139.78	25.19
Meditations On First Philosophy	Descartes	rationalism	1641.0	2008.0	192.34	33.79
Naming And Necessity	Kripke	analytic	1972.0	1990.0	120.57	20.09
Off The Beaten Track	Heidegger	phenomenology	1950.0	2001.0	108.53	18.47
On Certainty	Wittgenstein	analytic	1950.0	1969.0	79.38	14.18
On The Improvement Of Understanding	Spinoza	rationalism	1677.0	1997.0	176.80	30.48
On The Principles Of Political Economy And Taxation	Ricardo	capitalism	1817.0	2010.0	186.25	32.27
Philosophical Investigations	Wittgenstein	analytic	1953.0	1986.0	83.58	15.03
Philosophical Studies	Moore	analytic	1910.0	2015.0	167.25	29.80
Philosophical Troubles	Kripke	analytic	1975.0	2011.0	118.60	19.42
Plato - Complete Works	Plato	plato	-350.0	1997.0	114.94	20.72
Quintessence	Quine	analytic	1950.0	2004.0	121.64	19.85
Science Of Logic	Hegel	german_idealism	1817.0	2010.0	187.17	30.73
Second Treatise On Government	Locke	empiricism	1689.0	2010.0	266.79	47.38
The Analysis Of Mind	Russell	analytic	1921.0	2008.0	142.64	24.11
The Antichrist	Nietzsche	nietzsche	1888.0	2006.0	133.34	22.82
The Birth Of The Clinic	Foucault	continental	1963.0	2003.0	184.99	29.47
The Communist Manifesto	Marx	communism	1848.0	1970.0	150.68	23.85
The Crisis Of The European Sciences And Phenomenology	Husserl	phenomenology	1936.0	1970.0	192.05	30.35
The Idea Of Phenomenology	Husserl	phenomenology	1907.0	1999.0	150.56	24.78
The Logic Of Scientific Discovery	Popper	analytic	1959.0	2002.0	139.55	22.28
The Order Of Things	Foucault	continental	1966.0	2002.0	218.20	35.85
The Phenomenology Of Perception	Merleau-Ponty	phenomenology	1945.0	2002.0	170.93	28.74
The Phenomenology Of Spirit	Hegel	german_idealism	1807.0	1977.0	168.70	28.31
The Problems Of Philosophy	Russell	analytic	1912.0	2004.0	154.54	26.82
The Search After Truth	Malebranche	rationalism	1674.0	1997.0	164.43	28.93

title	author	school	original_publication_date	corpus_edition_date	sentence_length	tokens
The Second Sex	Beauvoir	feminism	1949.0	2009.0	148.79	25.11
The System Of Ethics	Fichte	german_idealism	1798.0	2005.0	151.96	25.96
The Wealth Of Nations	Smith	capitalism	1776.0	2009.0	185.28	31.51
Theodicy	Leibniz	rationalism	1710.0	2005.0	157.09	27.23
Three Dialogues	Berkeley	empiricism	1713.0	2009.0	111.98	19.56
Thus Spake Zarathustra	Nietzsche	nietzsche	1887.0	2008.0	80.56	14.31
Tractatus Logico-Philosophicus	Wittgenstein	analytic	1921.0	2001.0	100.19	16.52
Twilight Of The Idols	Nietzsche	nietzsche	1888.0	2016.0	126.84	21.58
Vindication Of The Rights Of Woman	Wollstonecraft	feminism	1792.0	2001.0	190.96	31.98
Women, Race, And Class	Davis	feminism	1981.0	1981.0	139.67	22.41
Writing And Difference	Derrida	continental	1967.0	2001.0	143.43	23.40

In [33]:

```python
#Creating wordclouds of 5 authors by masking onto a picture of the philisopher

authors=["Aristotle"]
for sc in authors:
    df_temp = df[df.author==sc]

    print('Author = ', sc, ':')

    text = " ".join(txt for txt in df_temp.sentence_lowered)
    custom_mask = np.array(Image.open("/content/gdrive/MyDrive/a2.png"))
    wordcloud = WordCloud(max_font_size=50, max_words=500,
                    width = 600, height = 400,
                    background_color="white", mask=custom_mask).generate(text)
    image_colors = ImageColorGenerator(custom_mask)
    wordcloud.recolor(color_func=image_colors)
    plt.figure(figsize=(12,8))
    plt.imshow(wordcloud, interpolation="bilinear")
    plt.axis("off")
    plt.show()
```

Author = Aristotle :

```python
authors=["Hume"]
for sc in authors:
    df_temp = df[df.author==sc]

    print('Author = ', sc, ':')

    text = " ".join(txt for txt in df_temp.sentence_lowered)
    custom_mask = np.array(Image.open("/content/gdrive/MyDrive/hume.png"))
    wordcloud = WordCloud(max_font_size=50, max_words=500,
                          width = 600, height = 400,
                          background_color="white", mask=custom_mask).generate(text)
    image_colors = ImageColorGenerator(custom_mask)
    wordcloud.recolor(color_func=image_colors)
    plt.figure(figsize=(12,8))
    plt.imshow(wordcloud, interpolation="bilinear")
    plt.axis("off")
    plt.show()
```

Author = Hume :

```python
authors=["Kant"]
for sc in authors:
    df_temp = df[df.author==sc]

    print('Author = ', sc, ':')

    text = " ".join(txt for txt in df_temp.sentence_lowered)
    custom_mask = np.array(Image.open("/content/gdrive/MyDrive/kant.png"))
    wordcloud = WordCloud(max_font_size=50, max_words=500,
                          width = 600, height = 400,
                          background_color="white", mask=custom_mask).generate(text)
#     image_colors = ImageColorGenerator(custom_mask)
#     wordcloud.recolor(color_func=image_colors)
    plt.figure(figsize=(12,8))
```

```
    plt.imshow(wordcloud, interpolation="bilinear")
    plt.axis("off")
    plt.show()
```

Author = Kant :

In [36]:

```python
authors=["Marx"]
for sc in authors:
    df_temp = df[df.author==sc]

    print('Author = ', sc, ':')

    text = " ".join(txt for txt in df_temp.sentence_lowered)
    custom_mask = np.array(Image.open("/content/gdrive/MyDrive/kant.png"))
    wordcloud = WordCloud(max_font_size=50, max_words=500,
                          width = 600, height = 400,
                          background_color="white", mask=custom_mask).generate(text)
    image_colors = ImageColorGenerator(custom_mask)
    wordcloud.recolor(color_func=image_colors)
    plt.figure(figsize=(12,8))
    plt.imshow(wordcloud, interpolation="bilinear")
    plt.axis("off")
    plt.show()
```

Author = Marx :

```
authors=["Plato"]
for sc in authors:
    df_temp = df[df.author==sc]

    print('Author = ', sc, ':')

    text = " ".join(txt for txt in df_temp.sentence_lowered)
    custom_mask = np.array(Image.open("/content/gdrive/MyDrive/kant.png"))
    wordcloud = WordCloud(max_font_size=50, max_words=500,
                          width = 600, height = 400,
                          background_color="white", mask=custom_mask).generate(text)
    image_colors = ImageColorGenerator(custom_mask)
    wordcloud.recolor(color_func=image_colors)
    plt.figure(figsize=(12,8))
    plt.imshow(wordcloud, interpolation="bilinear")
    plt.axis("off")
    plt.show()
```

Author = Plato :

Thank You!