Introduction to Text Analysis

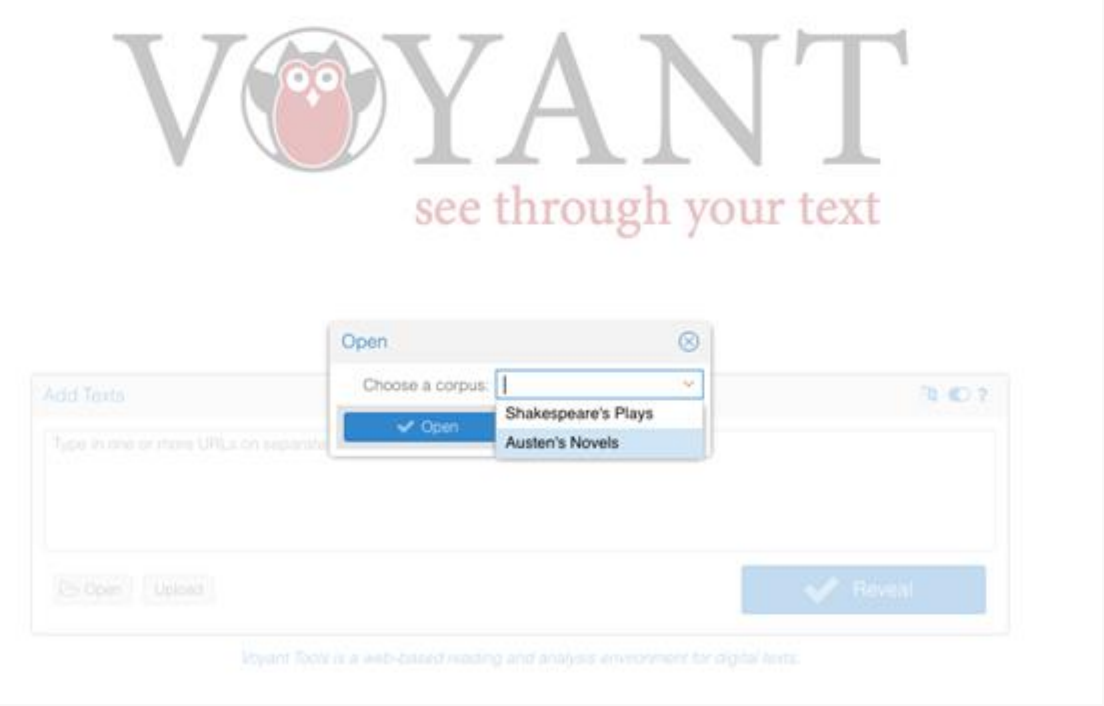
with Voyant

Sierra Eckert
Wesleyan University

An off-the-shelf alternative for exploratory data analysis:

Voyant is a basic dashboard for text analysis

voyant-tools.org

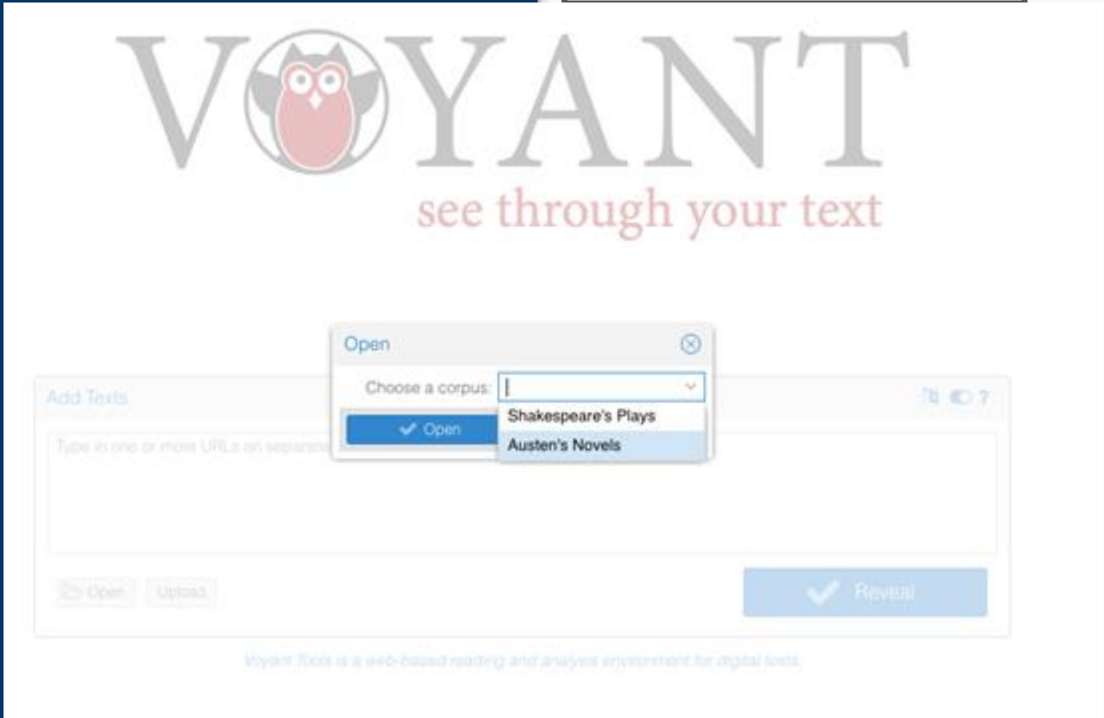


Voyant Tools is a web-based reading and analysis environment for digital texts.

Let's try some text analysis!

Please go to:

voyant-tools.org



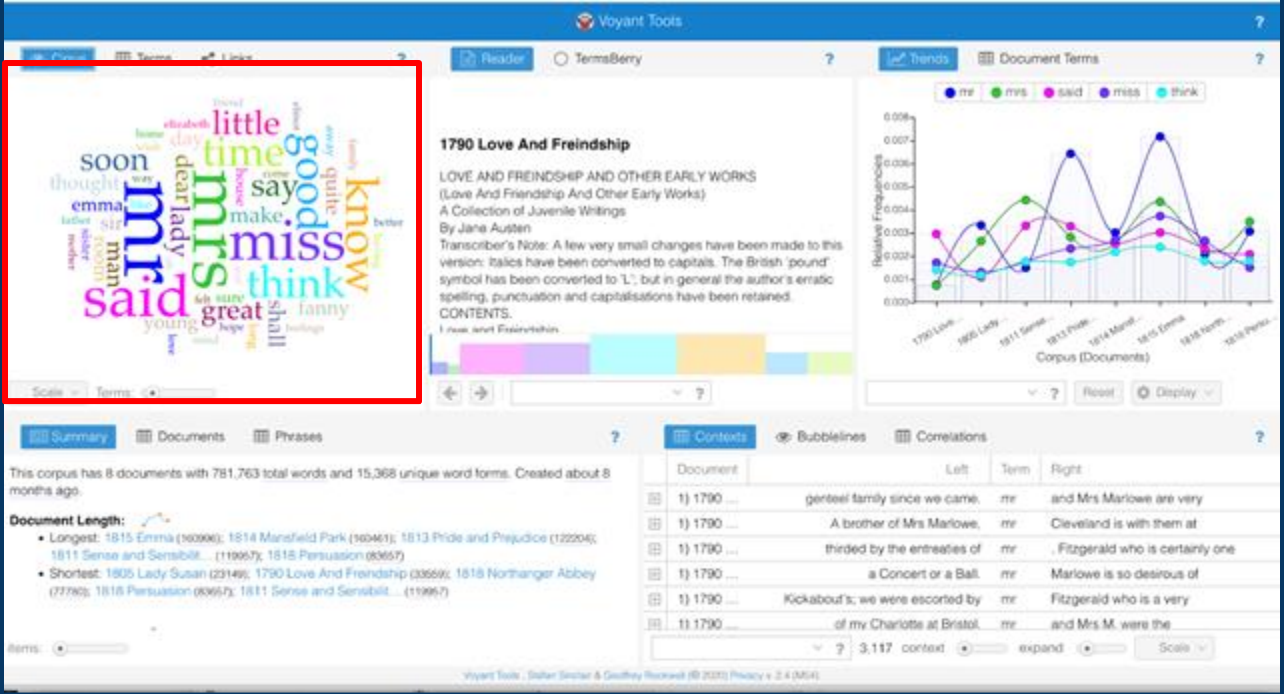
Voyant Tools is a web-based reading and analysis environment for digital texts.

From the **Open** menu, choose the corpus **Austen's Novels**.

Then press **Reveal**.

Voyant Tools

On the upper left is a WORD CLOUD.
This is a list of words sized by their frequency



1790 Love And Freindship

LOVE AND FREINDSHIP AND OTHER EARLY WORKS
(Love And Friendship And Other Early Works)
A Collection of Juvenile Writings
By Jane Austen
Transcriber's Note: A few very small changes have been made to this version: Italics have been converted to capitals. The British 'pound' symbol has been converted to 'L', but in general the author's erratic spelling, punctuation and capitalisations have been retained.
CONTENTS.
Love and Freindship

This corpus has 8 documents with 781,763 total words and 15,368 unique word forms. Created about 8 months ago.

Document Length:

- Longest: 1815 Emma (160996); 1814 Mansfield Park (160461); 1813 Pride and Prejudice (122204); 1811 Sense and Sensibilit... (119957); 1818 Persuasion (83657)
- Shortest: 1805 Lady Susan (23149); 1790 Love And Friendship (33569); 1818 Northanger Abbey (77780); 1818 Persuasion (83657); 1811 Sense and Sensibilit... (119957)

Items:

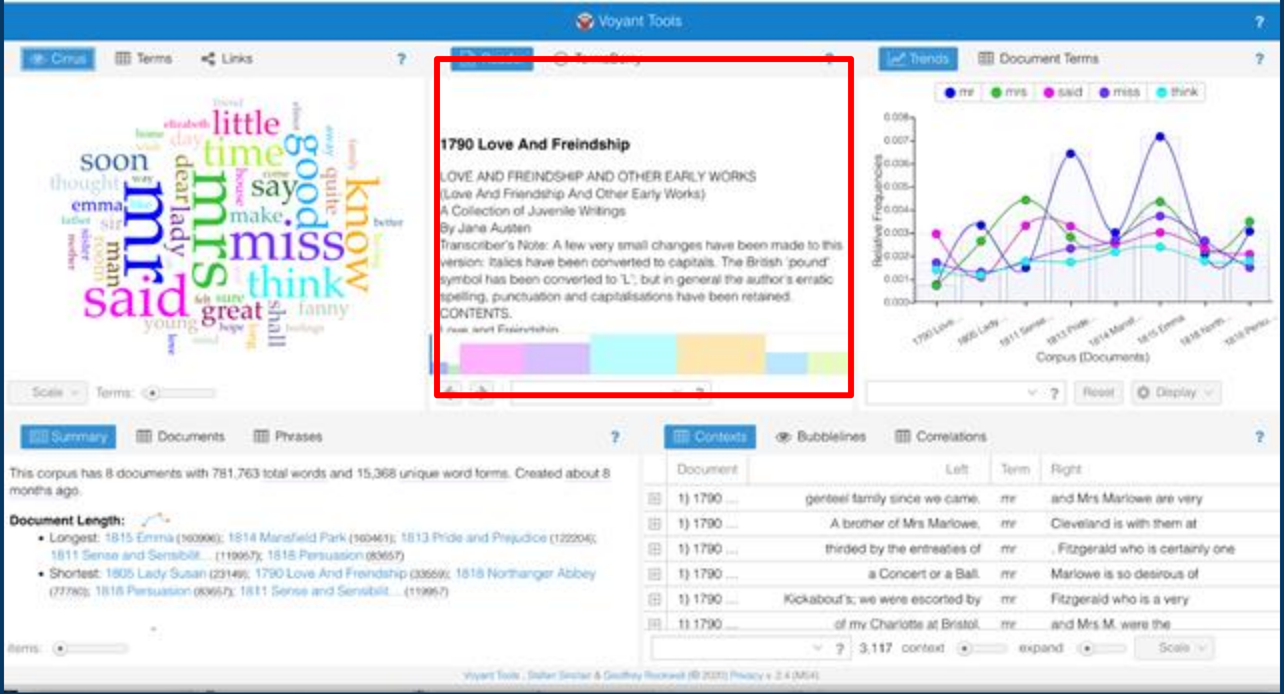
Document	Left	Term	Right
1) 1790 ...	genteel family since we came.	mr	and Mrs Marlowe are very
1) 1790 ...	A brother of Mrs Marlowe.	mr	Cleveland is with them at
1) 1790 ...	thirded by the entreaties of	mr	. Fitzgerald who is certainly one
1) 1790 ...	a Concert or a Ball.	mr	Marlowe is so desirous of
1) 1790 ...	Kickabout's; we were escorted by	mr	Fitzgerald who is a very
1) 1790 ...	of my Charlotte at Bristol.	mr	and Mrs M. were the

? 3,117 context expand Scale

Voyant Tools , Stéfan Sinclair & Geoffrey Rockwell (© 2020) Privacy v. 2.4 (M54)

In the upper center square is the **TEXT of your CORPUS**.
This gives you the full text that you're analyzing, listed in order that they are labeled (in this case, by date). If you hover over a word, it will tell you how many times it appears in the collection.
Take a look at the text in the box. What do you notice? What kinds of problems might it pose for our analysis?

Voyant Tools



1790 Love And Freindship

LOVE AND FREINDSHIP AND OTHER EARLY WORKS
(Love And Friendship And Other Early Works)
A Collection of Juvenile Writings
By Jane Austen
Transcriber's Note: A few very small changes have been made to this version: Italics have been converted to capitals. The British 'pound' symbol has been converted to 'L', but in general the author's erratic spelling, punctuation and capitalisations have been retained.
CONTENTS.
Love and Freindship

This corpus has 8 documents with 781,763 total words and 15,368 unique word forms. Created about 8 months ago.

Document Length:

- Longest: 1815 Emma (160996); 1814 Mansfield Park (160461); 1813 Pride and Prejudice (122204); 1811 Sense and Sensibilit... (119957); 1818 Persuasion (83657)
- Shortest: 1805 Lady Susan (23149); 1790 Love And Freindship (33569); 1818 Northanger Abbey (77780); 1818 Persuasion (83657); 1811 Sense and Sensibilit... (119957)

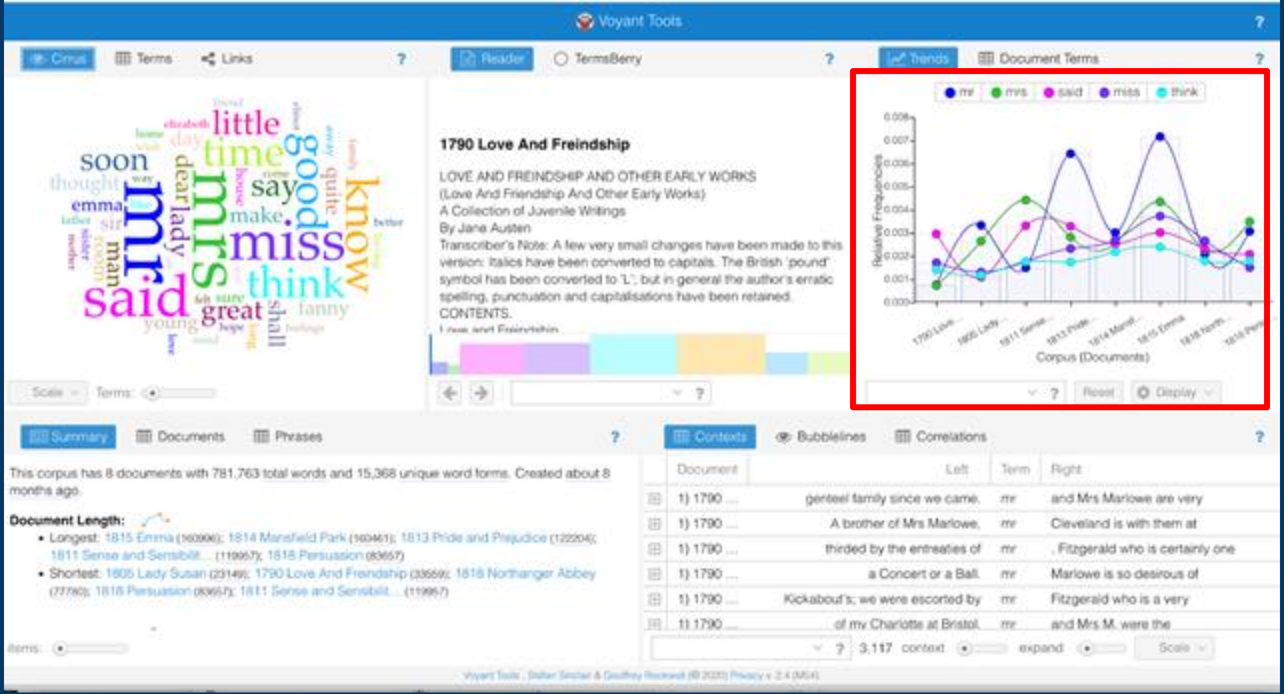
Document	Left	Term	Right
1) 1790 ...	genteel family since we came.	mr	and Mrs Marlowe are very
1) 1790 ...	A brother of Mrs Marlowe.	mr	Cleveland is with them at
1) 1790 ...	thirded by the entreaties of	mr	, Fitzgerald who is certainly one
1) 1790 ...	a Concert or a Ball.	mr	Marlowe is so desirous of
1) 1790 ...	Kickabout's; we were escorted by	mr	Fitzgerald who is a very
1) 1790 ...	of my Charlotte at Bristol.	mr	and Mrs M. were the

3,117 context expand Scale

Voyant Tools , Stéfan Sinclair & Geoffrey Rockwell (© 2020) Privacy v. 2.4 (M54)

In the upper right is the WORD FREQUENCY visualization.
This shows you how frequently a word appears in each of the documents in the corpus.
Try typing in “very” in the search bar at the bottom of this box. What do you notice?

Voyant Tools



1790 Love And Freindship

LOVE AND FREINDSHIP AND OTHER EARLY WORKS
(Love And Friendship And Other Early Works)
A Collection of Juvenile Writings
By Jane Austen
Transcriber's Note: A few very small changes have been made to this version: Italics have been converted to capitals. The British 'pound' symbol has been converted to 'L', but in general the author's erratic spelling, punctuation and capitalisations have been retained.
CONTENTS.
Love and Freindship

This corpus has 8 documents with 781,763 total words and 15,368 unique word forms. Created about 8 months ago.

Document Length:

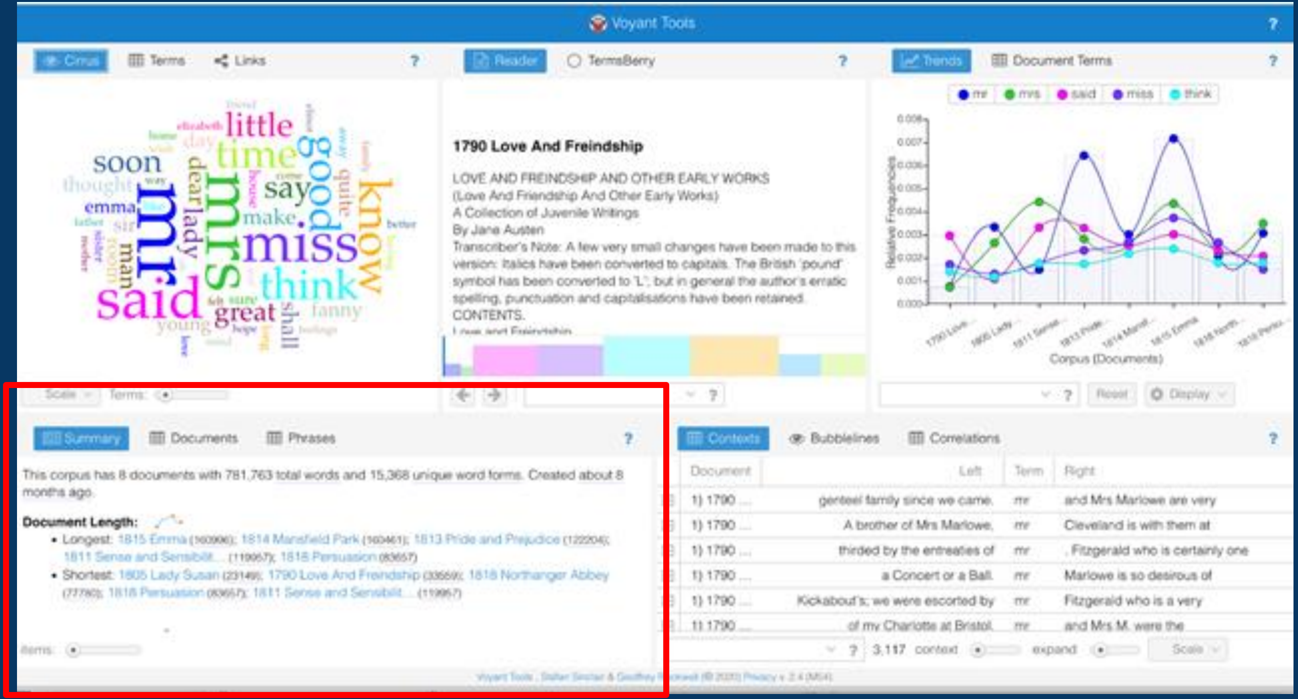
- Longest: 1815 Emma (160906); 1814 Mansfield Park (160461); 1813 Pride and Prejudice (122204); 1811 Sense and Sensibilit... (119957); 1818 Persuasion (83657)
- Shortest: 1805 Lady Susan (23149); 1790 Love And Freindship (33559); 1818 Northanger Abbey (77780); 1818 Persuasion (83657); 1811 Sense and Sensibilit... (119957)

Document	Left	Term	Right
1) 1790 ...	genteel family since we came.	mr	and Mrs Marlowe are very
1) 1790 ...	A brother of Mrs Marlowe,	mr	Cleveland is with them at
1) 1790 ...	thirded by the entreaties of	mr	, Fitzgerald who is certainly one
1) 1790 ...	a Concert or a Ball.	mr	Marlowe is so desirous of
1) 1790 ...	Kickabout's; we were escorted by	mr	Fitzgerald who is a very
1) 1790 ...	of my Charlotte at Bristol.	mr	and Mrs M. were the

3,117

In the bottom left are STATISTICS about your corpus.
Scrolling down in the “Summary” view gives a longer list of some of most distinctive words in each text, the average document and sentence length. “Phrases” allows you to sort by short phrases.

Voyant Tools



1790 Love And Freindship

LOVE AND FREINDSHIP AND OTHER EARLY WORKS
(Love And Friendship And Other Early Works)
A Collection of Juvenile Writings
By Jane Austen
Transcriber's Note: A few very small changes have been made to this version: Italics have been converted to capitals. The British 'pound' symbol has been converted to 'L', but in general the author's erratic spelling, punctuation and capitalisations have been retained.
CONTENTS.
Love and Freindship

This corpus has 8 documents with 781,763 total words and 15,368 unique word forms. Created about 8 months ago.

Document Length:

- Longest: 1815 Emma (160996); 1814 Mansfield Park (160461); 1813 Pride and Prejudice (122204); 1811 Sense and Sensibilit... (119957); 1818 Persuasion (83657)
- Shortest: 1805 Lady Susan (23149); 1790 Love And Freindship (33569); 1818 Northanger Abbey (77780); 1818 Persuasion (83657); 1811 Sense and Sensibilit... (119957)

Items:

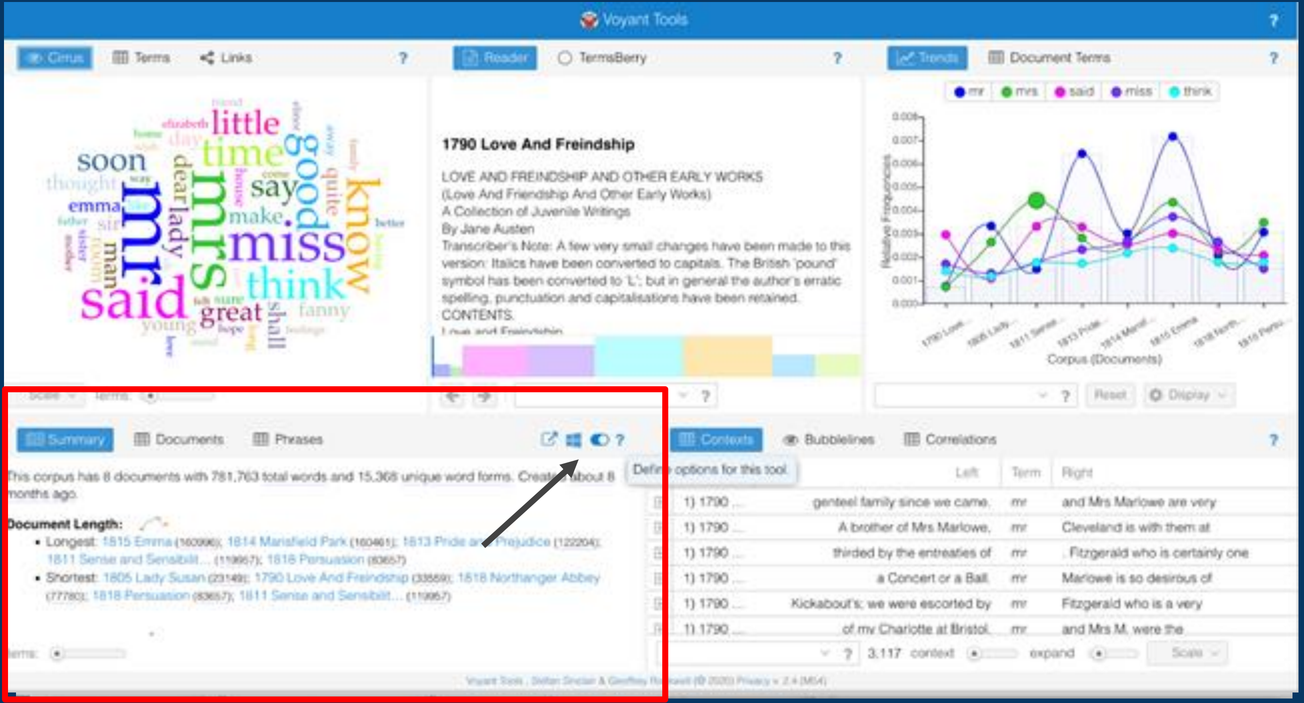
Document	Left	Term	Right
1) 1790 ...	genteel family since we came.	mr	and Mrs Marlowe are very
1) 1790 ...	A brother of Mrs Marlowe,	mr	Cleveland is with them at
1) 1790 ...	thirded by the entreaties of	mr	. Fitzgerald who is certainly one
1) 1790 ...	a Concert or a Ball.	mr	Marlowe is so desirous of
1) 1790 ...	Kickabout's; we were escorted by	mr	Fitzgerald who is a very
1) 1790 ...	of my Charlotte at Bristol.	mr	and Mrs M. were the

3,117 context expand Scale

Voyant Tools , Stéfan Sinclair & Geoffrey Rockwell (© 2020) Privacy v. 2.4 (M24)

This box also allows you to control the global filters for the toolset.
Hover over the upper right corner of this box and click on the “options” toggle

Voyant Tools



1790 Love And Freindship

LOVE AND FREINDSHIP AND OTHER EARLY WORKS
(Love And Friendship And Other Early Works)
A Collection of Juvenile Writings
By Jane Austen
Transcriber's Note: A few very small changes have been made to this version: Italics have been converted to capitals. The British 'pound' symbol has been converted to 'L', but in general the author's erratic spelling, punctuation and capitalisations have been retained.
CONTENTS.
Love and Freindship

This corpus has 8 documents with 781,763 total words and 15,368 unique word forms. Created about 8 months ago.

Document Length:

- Longest: 1815 Emma (160996); 1814 Mansfield Park (160461); 1813 Pride and Prejudice (122204); 1811 Sense and Sensibilit... (119957); 1818 Persuasion (83657)
- Shortest: 1805 Lady Susan (23149); 1790 Love And Freindship (33669); 1818 Northanger Abbey (77780); 1818 Persuasion (83657); 1811 Sense and Sensibilit... (119957)

	Left	Term	Right
1) 1790 ...	genteel family since we came.	mr	and Mrs Marlowe are very
1) 1790 ...	A brother of Mrs Marlowe.	mr	Cleveland is with them at
1) 1790 ...	thirded by the entreaties of	mr	. Fitzgerald who is certainly one
1) 1790 ...	a Concert or a Ball.	mr	Marlowe is so desirous of
1) 1790 ...	Kickabout's; we were escorted by	mr	Fitzgerald who is a very
1) 1790 ...	of my Charlotte at Bristol.	mr	and Mrs M. were the

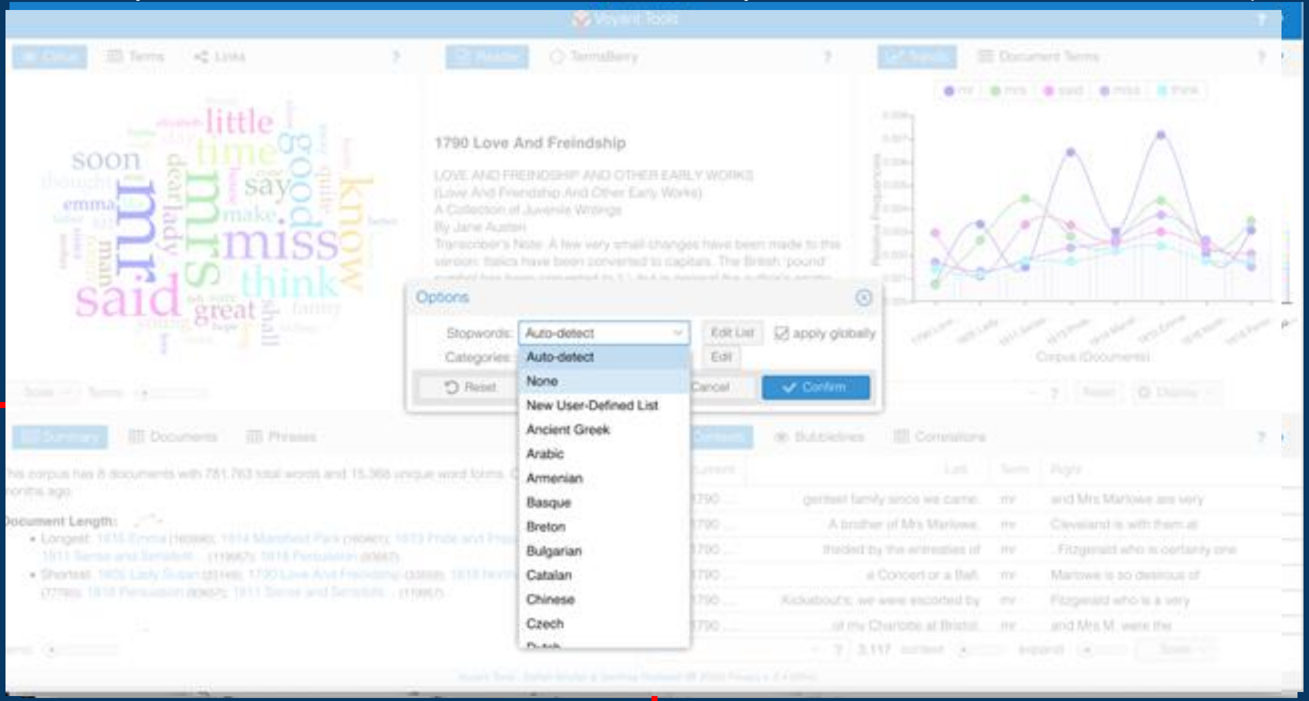
In the options box, click on “None.” Then click Confirm. What happened?

What you just removed was a “stop words” list.

Click on options again, and click on “auto-detect.” Then click on the Edit list button. What do you notice about the words?

When would you want to filter certain words out? When *wouldn’t* you want to remove them? What are the implications?

Voyant Tools



1790 Love And Freindship

LOVE AND FREINDSHIP AND OTHER EARLY WORKS
(Love And Friendship And Other Early Works)
A Collection of Juvenile Writings
By Jane Austen
Transcriber's Note: A few very small changes have been made to this version: Italics have been converted to capitals. The British 'pound'

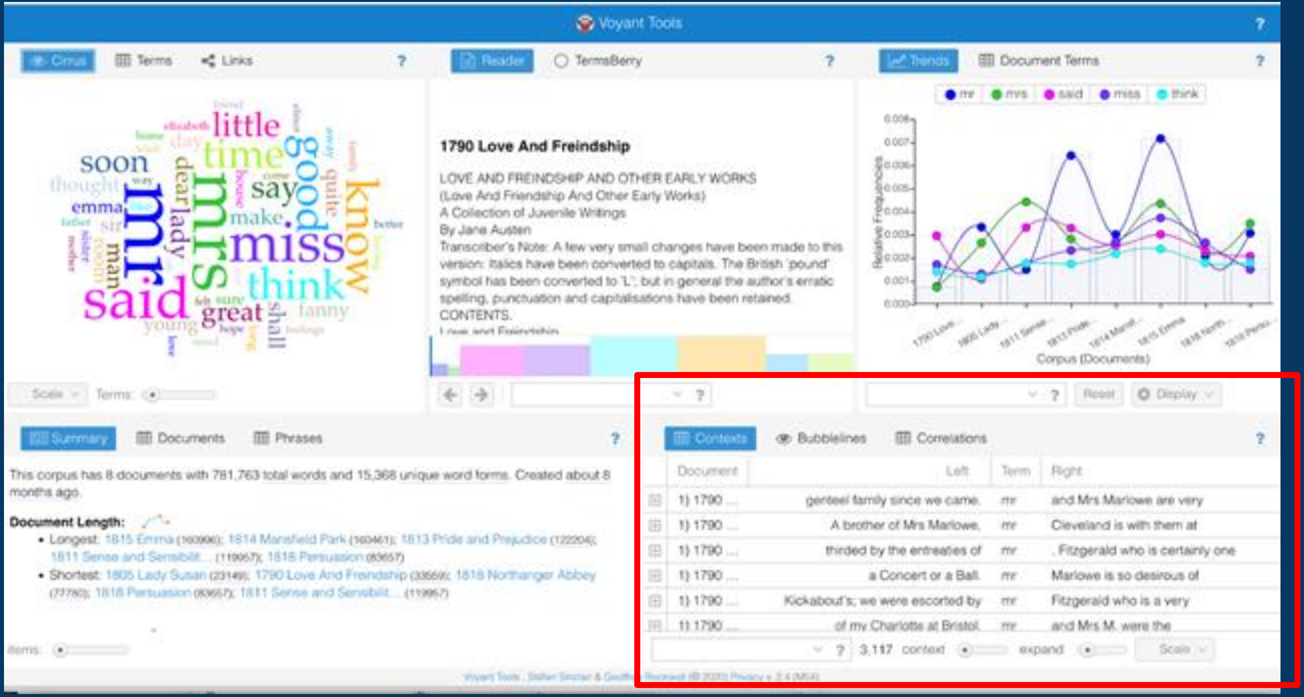
This corpus has 8 documents with 781,763 total words and 15,368 unique word forms.

Stopwords: Auto-detect

Remember our discussion about stopwords?? What implications do they have for our analysis here?

Voyant Tools

In the bottom right is a CONCORDANCE.
This gives you the context of words in your corpus as they appear in each document.
Try typing in “gentleman” and sorting by the words that appear on the left.
Toggle to the “Bubblelines” view. Type in “pounds,” “estate,” “money” and “inheritance.” What do you notice?



1790 Love And Freindship

LOVE AND FREINDSHIP AND OTHER EARLY WORKS
(Love And Friendship And Other Early Works)
A Collection of Juvenile Writings
By Jane Austen
Transcriber's Note: A few very small changes have been made to this version: Italics have been converted to capitals. The British 'pound' symbol has been converted to 'L', but in general the author's erratic spelling, punctuation and capitalisations have been retained.
CONTENTS.
Love and Freindship

This corpus has 8 documents with 781,763 total words and 15,368 unique word forms. Created about 8 months ago.

Document Length:

- Longest: 1815 Emma (160906); 1814 Mansfield Park (160461); 1813 Pride and Prejudice (122204); 1811 Sense and Sensibilit... (119957); 1818 Persuasion (83657)
- Shortest: 1805 Lady Susan (23149); 1790 Love And Freindship (33569); 1818 Northanger Abbey (77780); 1818 Persuasion (83657); 1811 Sense and Sensibilit... (119957)

Document	Left	Term	Right
1) 1790 ...	genteel family since we came.	mr	and Mrs Marlowe are very
1) 1790 ...	A brother of Mrs Marlowe,	mr	Cleveland is with them at
1) 1790 ...	thirded by the entreaties of	mr	, Fitzgerald who is certainly one
1) 1790 ...	a Concert or a Ball.	mr	Marlowe is so desirous of
1) 1790 ...	Kickabout's; we were escorted by	mr	Fitzgerald who is a very
1) 1790 ...	of my Charlotte at Bristol.	mr	and Mrs M. were the

3,117

Voyant Tools , Stéfan Sinclair & Geoffrey Rockwell (© 2020) Privacy v. 2.4 (M54)

Voyant Tools

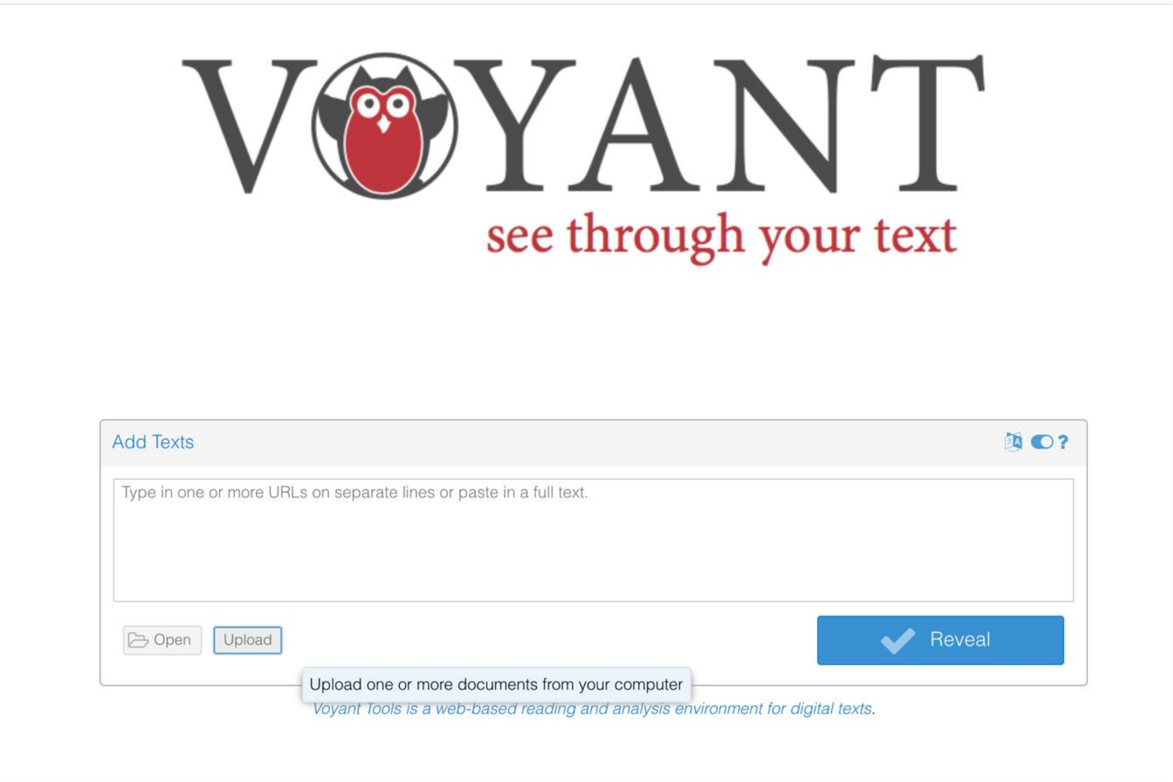
Take a minute to play around some of the features. Toggle the amount of words in the CONCORDANCE, or the “items” in the STATISTICS box.

Brainstorm a few questions that you could explore with this kind of interface.

What kind of questions could you ask?
What kind of questions could you not ask?

Uploading our own texts

We can also upload our own corpus of plain text files.
Let's try the corpus of U.S. Inaugural Addresses



Type in one or more URLs on separate lines or paste in a full text.

Upload one or more documents from your computer

Voyant Tools is a web-based reading and analysis environment for digital texts.

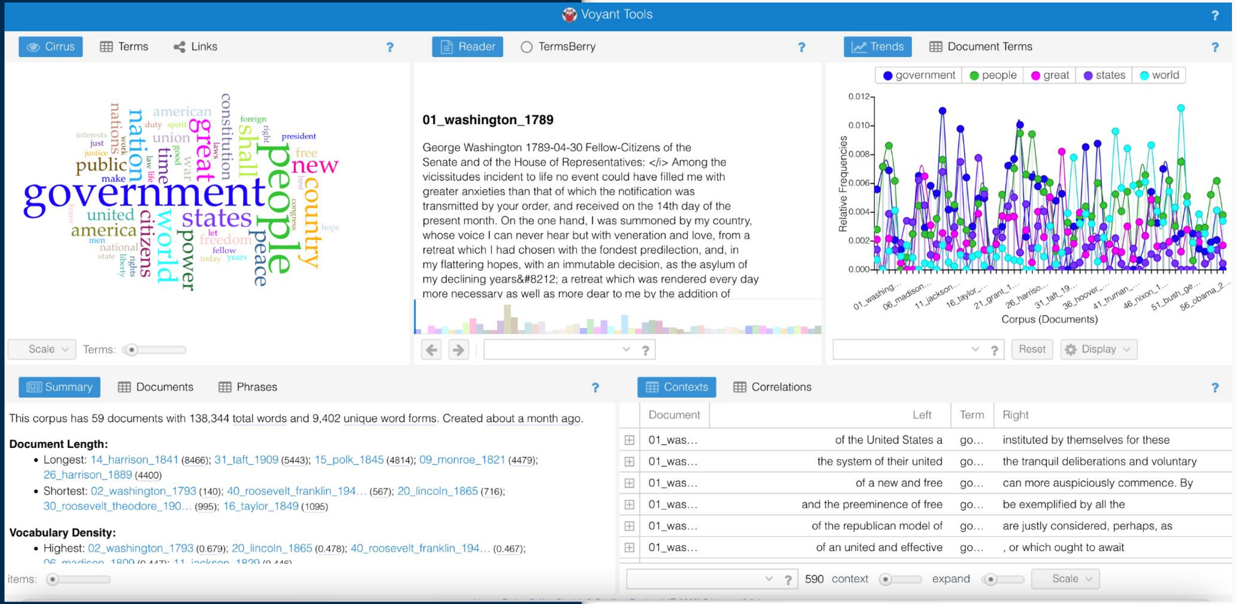
Select the **Upload** menu, and upload all the text files that we downloaded from our U.S. Inaugural Addresses dataset

Alternately, go to:
tinyurl.com/InauguralAddresses

Let's try some text analysis!

Please go to:

tinyurl.com/InauguralAddresses



01_washington_1789

George Washington 1789-04-30 Fellow-Citizens of the Senate and of the House of Representatives: </i> Among the vicissitudes incident to life no event could have filled me with greater anxieties than that of which the notification was transmitted by your order, and received on the 14th day of the present month. On the one hand, I was summoned by my country, whose voice I can never hear but with veneration and love, from a retreat which I had chosen with the fondest predilection, and, in my flattering hopes, with an immutable decision, as the asylum of my declining years&#8212; a retreat which was rendered every day more necessary as well as more dear to me by the addition of

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms. Created about a month ago.

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5443); 15_polk_1845 (4814); 09_monroe_1821 (4479); 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

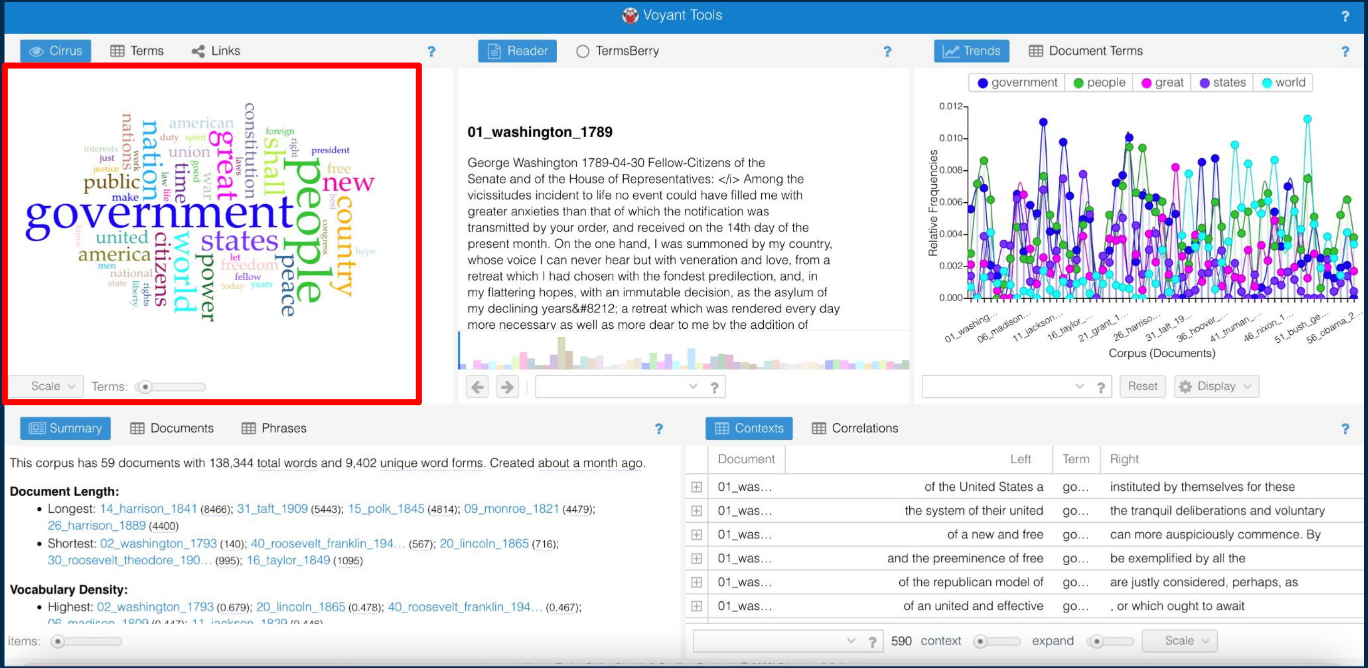
Vocabulary Density:

- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194… (0.467); 06_madison_1809 (0.447); 11_jackson_1829 (0.446)

	Document	Left	Term	Right
⊞	01_was…	of the United States a	go…	instituted by themselves for these
⊞	01_was…	the system of their united	go…	the tranquil deliberations and voluntary
⊞	01_was…	of a new and free	go…	can more auspiciously commence. By
⊞	01_was…	and the preeminence of free	go…	be exemplified by all the
⊞	01_was…	of the republican model of	go…	are justly considered, perhaps, as
⊞	01_was…	of an united and effective	go…	, or which ought to await

On the upper left is a WORD CLOUD.
This is a list of words sized by their frequency

Voyant Tools



01_washington_1789

George Washington 1789-04-30 Fellow-Citizens of the Senate and of the House of Representatives: </i> Among the vicissitudes incident to life no event could have filled me with greater anxieties than that of which the notification was transmitted by your order, and received on the 14th day of the present month. On the one hand, I was summoned by my country, whose voice I can never hear but with veneration and love, from a retreat which I had chosen with the fondest predilection, and, in my flattering hopes, with an immutable decision, as the asylum of my declining years&#8212; a retreat which was rendered every day more necessary as well as more dear to me by the addition of

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms. Created about a month ago.

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5443); 15_polk_1845 (4814); 09_monroe_1821 (4479); 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

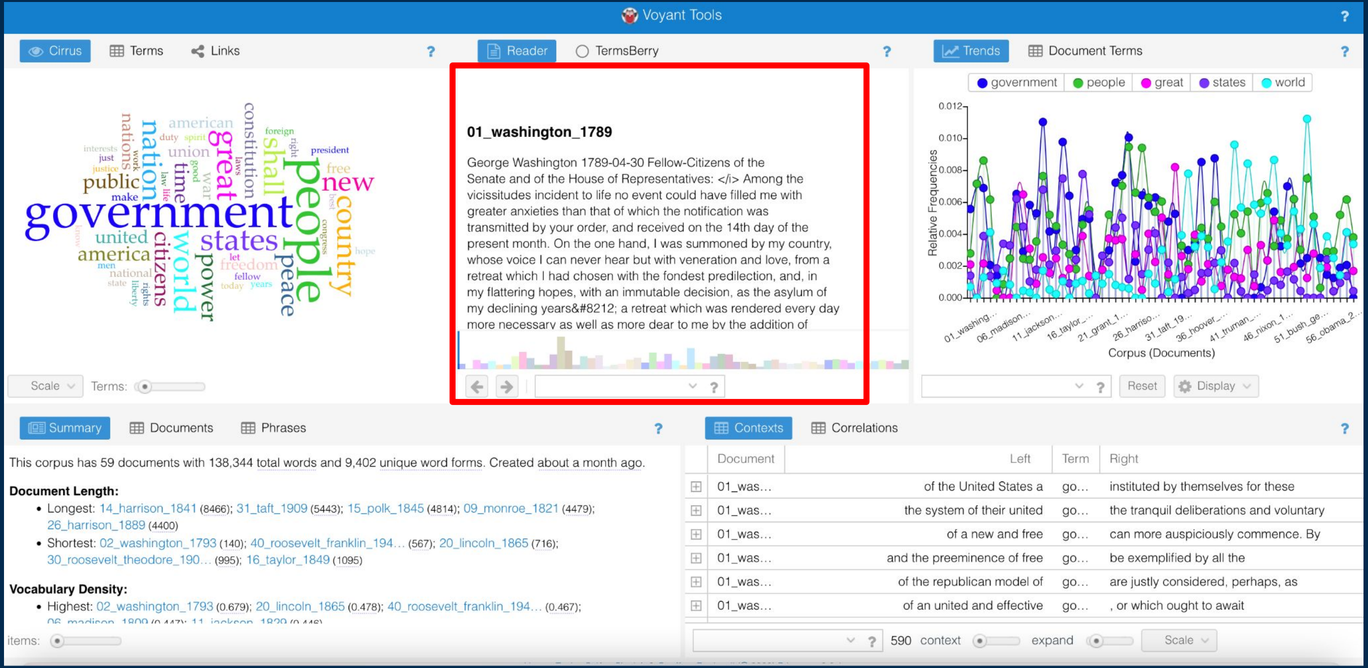
Vocabulary Density:

- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194… (0.467); 06_madison_1809 (0.447); 11_jackson_1829 (0.446)

	Document	Left	Term	Right
⊞	01_was…	of the United States a	go…	instituted by themselves for these
⊞	01_was…	the system of their united	go…	the tranquil deliberations and voluntary
⊞	01_was…	of a new and free	go…	can more auspiciously commence. By
⊞	01_was…	and the preeminence of free	go…	be exemplified by all the
⊞	01_was…	of the republican model of	go…	are justly considered, perhaps, as
⊞	01_was…	of an united and effective	go…	, or which ought to await

In the upper center square is the **TEXT of your CORPUS**.
This gives you the full text that you're analyzing, listed in order that they are labeled (in this case, by date). If you hover over a word, it will tell you how many times it appears in the collection.

Voyant Tools



01_washington_1789

George Washington 1789-04-30 Fellow-Citizens of the Senate and of the House of Representatives: </i> Among the vicissitudes incident to life no event could have filled me with greater anxieties than that of which the notification was transmitted by your order, and received on the 14th day of the present month. On the one hand, I was summoned by my country, whose voice I can never hear but with veneration and love, from a retreat which I had chosen with the fondest predilection, and, in my flattering hopes, with an immutable decision, as the asylum of my declining years&#8212; a retreat which was rendered every day more necessary as well as more dear to me by the addition of

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms. Created about a month ago.

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5443); 15_polk_1845 (4814); 09_monroe_1821 (4479); 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

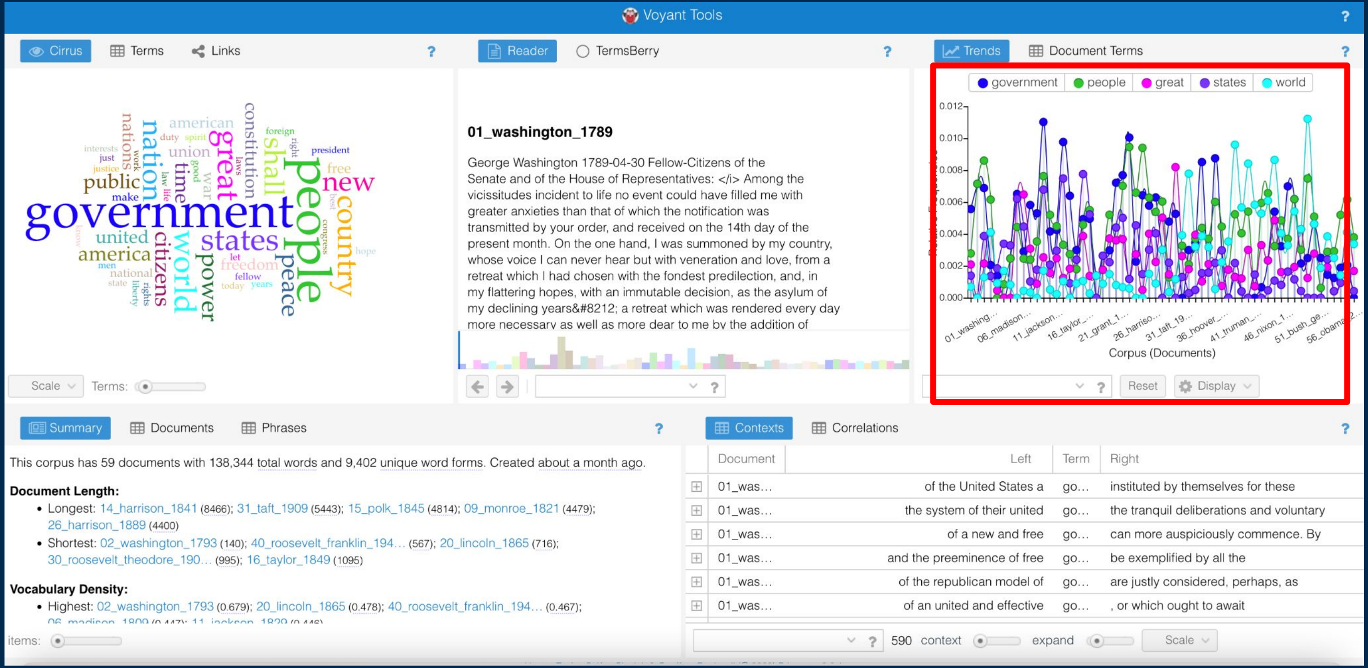
Vocabulary Density:

- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194… (0.467); 06_madison_1809 (0.447); 11_jackson_1829 (0.446)

	Document	Left	Term	Right
⊞	01_was…	of the United States a	go…	instituted by themselves for these
⊞	01_was…	the system of their united	go…	the tranquil deliberations and voluntary
⊞	01_was…	of a new and free	go…	can more auspiciously commence. By
⊞	01_was…	and the preeminence of free	go…	be exemplified by all the
⊞	01_was…	of the republican model of	go…	are justly considered, perhaps, as
⊞	01_was…	of an united and effective	go…	, or which ought to await

Voyant Tools

In the upper right is the WORD FREQUENCY visualization.
This shows you how frequently a word appears in each of the documents in the corpus.
Try typing in “America” in the search bar at the bottom of this box. What do you notice?



01_washington_1789

George Washington 1789-04-30 Fellow-Citizens of the Senate and of the House of Representatives: </i> Among the vicissitudes incident to life no event could have filled me with greater anxieties than that of which the notification was transmitted by your order, and received on the 14th day of the present month. On the one hand, I was summoned by my country, whose voice I can never hear but with veneration and love, from a retreat which I had chosen with the fondest predilection, and, in my flattering hopes, with an immutable decision, as the asylum of my declining years&#8212; a retreat which was rendered every day more necessary as well as more dear to me by the addition of

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms. Created about a month ago.

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5443); 15_polk_1845 (4814); 09_monroe_1821 (4479); 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

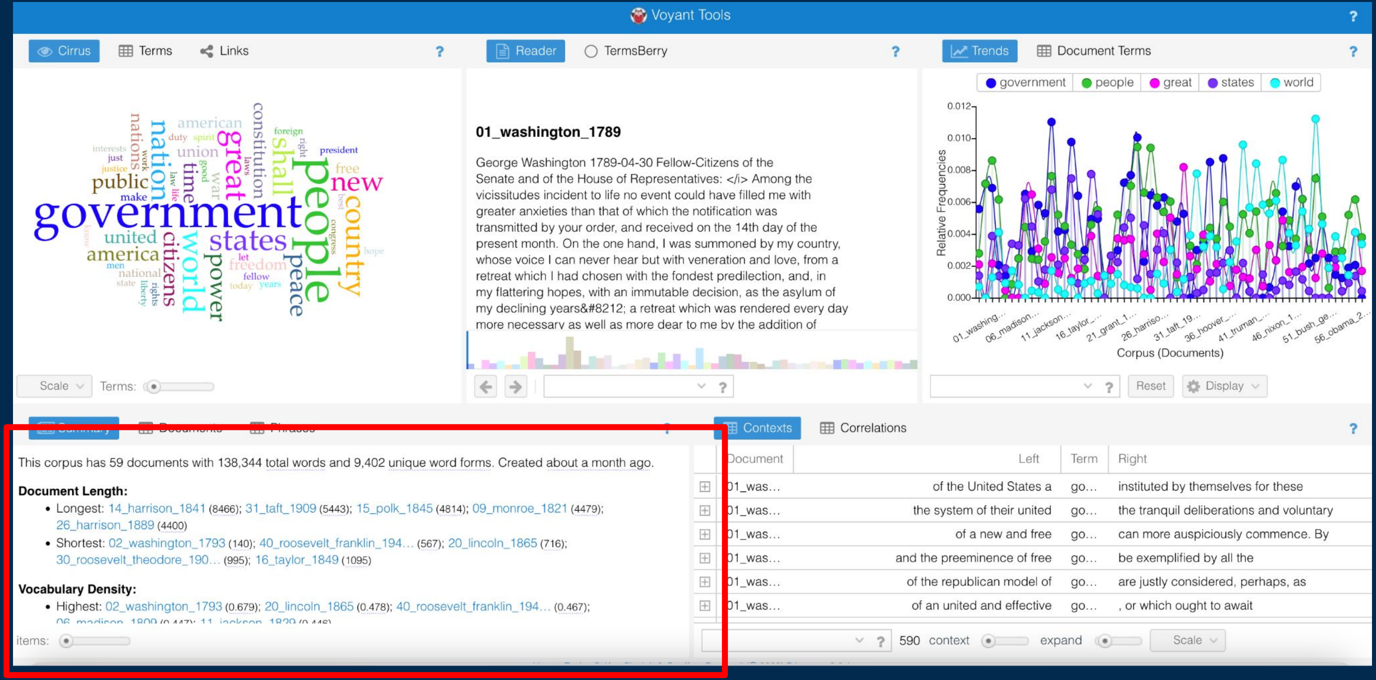
Vocabulary Density:

- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194… (0.467); 06_madison_1809 (0.447); 11_jackson_1829 (0.446)

	Document	Left	Term	Right
⊞	01_was…	of the United States a	go…	instituted by themselves for these
⊞	01_was…	the system of their united	go…	the tranquil deliberations and voluntary
⊞	01_was…	of a new and free	go…	can more auspiciously commence. By
⊞	01_was…	and the preeminence of free	go…	be exemplified by all the
⊞	01_was…	of the republican model of	go…	are justly considered, perhaps, as
⊞	01_was…	of an united and effective	go…	, or which ought to await

Voyant Tools

In the bottom left are STATISTICS about your corpus.
Scrolling down in the "Summary" view gives a longer list of some of most distinctive words in each text, the average document and sentence length. "Phrases" allows you to sort by short phrases.



01_washington_1789

George Washington 1789-04-30 Fellow-Citizens of the Senate and of the House of Representatives: </i> Among the vicissitudes incident to life no event could have filled me with greater anxieties than that of which the notification was transmitted by your order, and received on the 14th day of the present month. On the one hand, I was summoned by my country, whose voice I can never hear but with veneration and love, from a retreat which I had chosen with the fondest predilection, and, in my flattering hopes, with an immutable decision, as the asylum of my declining years&#8212; a retreat which was rendered every day more necessary as well as more dear to me by the addition of

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms. Created about a month ago.

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5443); 15_polk_1845 (4814); 09_monroe_1821 (4479); 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

Vocabulary Density:

- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194… (0.467); 06_madison_1809 (0.447); 11_jackson_1829 (0.446)

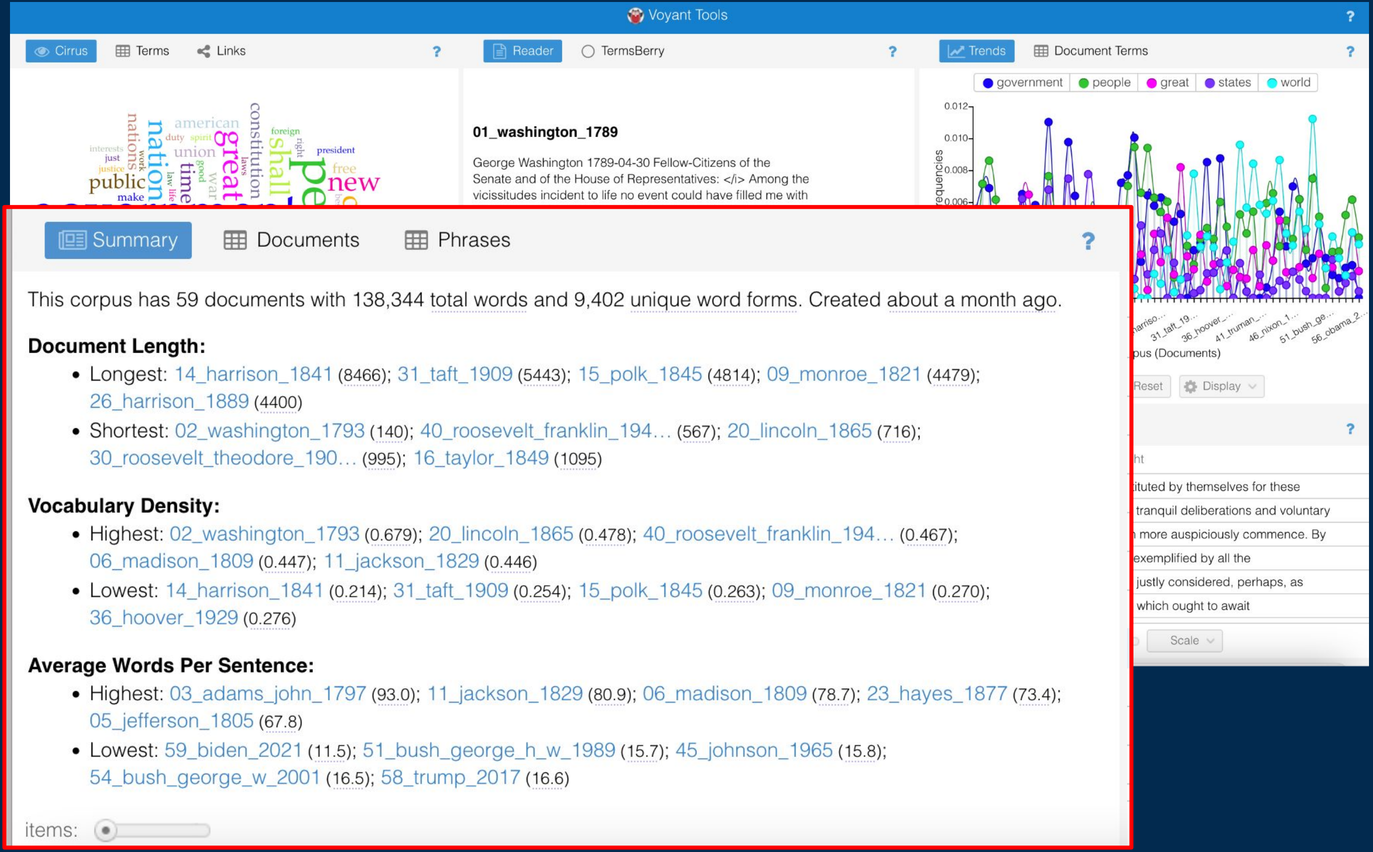
items:

Document	Left	Term	Right
01_was…	of the United States a	go…	instituted by themselves for these
01_was…	the system of their united	go…	the tranquil deliberations and voluntary
01_was…	of a new and free	go…	can more auspiciously commence. By
01_was…	and the preeminence of free	go…	be exemplified by all the
01_was…	of the republican model of	go…	are justly considered, perhaps, as
01_was…	of an united and effective	go…	, or which ought to await

590 context expand Scale

With this same “Summary” box, you’ll have descriptive statistics on the text or corpus (collection of texts) that you’re working with. These include “Document Length”, “Vocabulary Density”, “Average Words Per Sentence” You’ll also see “Most Frequent Words” in the corpus and Most Distinctive Words in each document. If you want to know what exactly these are measuring, click on the question mark in this box’s upper right corner.

Voyant Tools



01_washington_1789

George Washington 1789-04-30 Fellow-Citizens of the Senate and of the House of Representatives: </i> Among the vicissitudes incident to life no event could have filled me with

ituted by themselves for these
tranquil deliberations and voluntary
more auspiciously commence. By
exemplified by all the
justly considered, perhaps, as
which ought to await

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms. Created about a month ago.

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5443); 15_polk_1845 (4814); 09_monroe_1821 (4479); 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

Vocabulary Density:

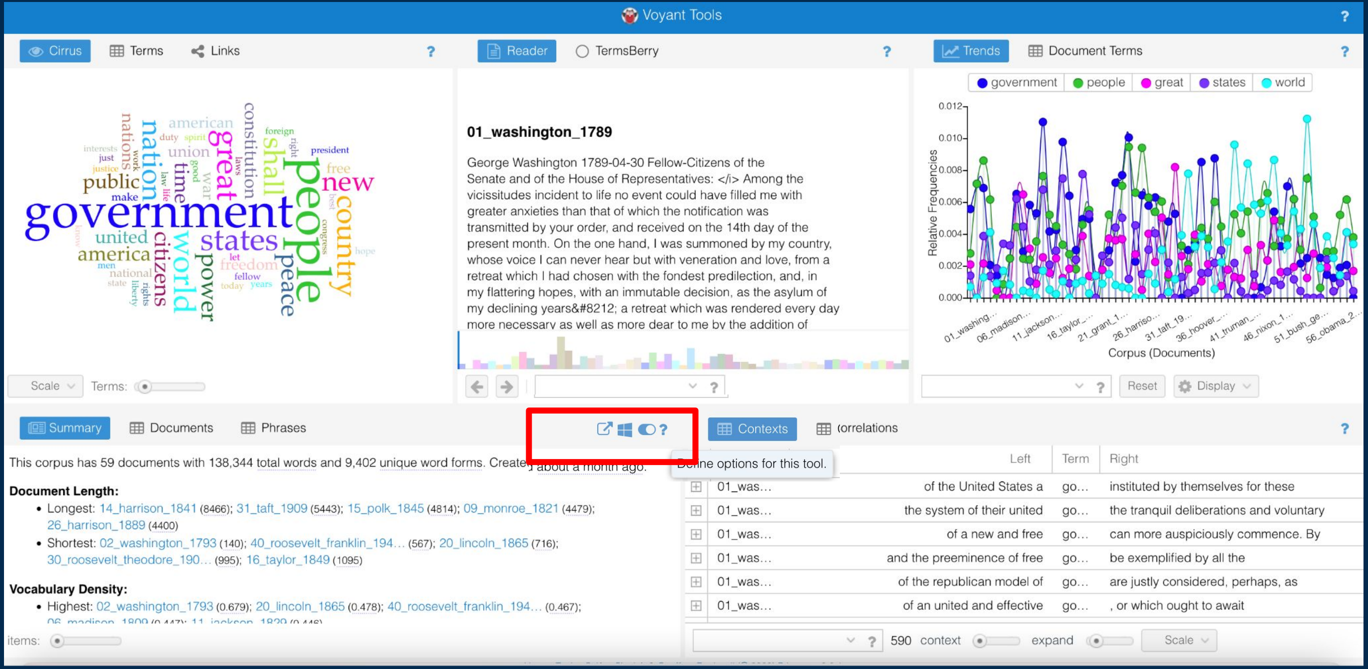
- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194… (0.467); 06_madison_1809 (0.447); 11_jackson_1829 (0.446)
- Lowest: 14_harrison_1841 (0.214); 31_taft_1909 (0.254); 15_polk_1845 (0.263); 09_monroe_1821 (0.270); 36_hoover_1929 (0.276)

Average Words Per Sentence:

- Highest: 03_adams_john_1797 (93.0); 11_jackson_1829 (80.9); 06_madison_1809 (78.7); 23_hayes_1877 (73.4); 05_jefferson_1805 (67.8)
- Lowest: 59_biden_2021 (11.5); 51_bush_george_h_w_1989 (15.7); 45_johnson_1965 (15.8); 54_bush_george_w_2001 (16.5); 58_trump_2017 (16.6)

This box also allows you to control the global filters for the toolset.
Hover over the upper right corner of this box and click on the "options" toggle

Voyant Tools



01_washington_1789

George Washington 1789-04-30 Fellow-Citizens of the Senate and of the House of Representatives: </i> Among the vicissitudes incident to life no event could have filled me with greater anxieties than that of which the notification was transmitted by your order, and received on the 14th day of the present month. On the one hand, I was summoned by my country, whose voice I can never hear but with veneration and love, from a retreat which I had chosen with the fondest predilection, and, in my flattering hopes, with an immutable decision, as the asylum of my declining years&#8212; a retreat which was rendered every day more necessary as well as more dear to me by the addition of

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms. Create[d] about a month ago.

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5443); 15_polk_1845 (4814); 09_monroe_1821 (4479); 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

Vocabulary Density:

- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194… (0.467); 06_madison_1809 (0.447); 11_jackson_1829 (0.446)

Define options for this tool.

		Left	Term	Right
⊞	01_was…	of the United States a	go…	instituted by themselves for these
⊞	01_was…	the system of their united	go…	the tranquil deliberations and voluntary
⊞	01_was…	of a new and free	go…	can more auspiciously commence. By
⊞	01_was…	and the preeminence of free	go…	be exemplified by all the
⊞	01_was…	of the republican model of	go…	are justly considered, perhaps, as
⊞	01_was…	of an united and effective	go…	, or which ought to await

590

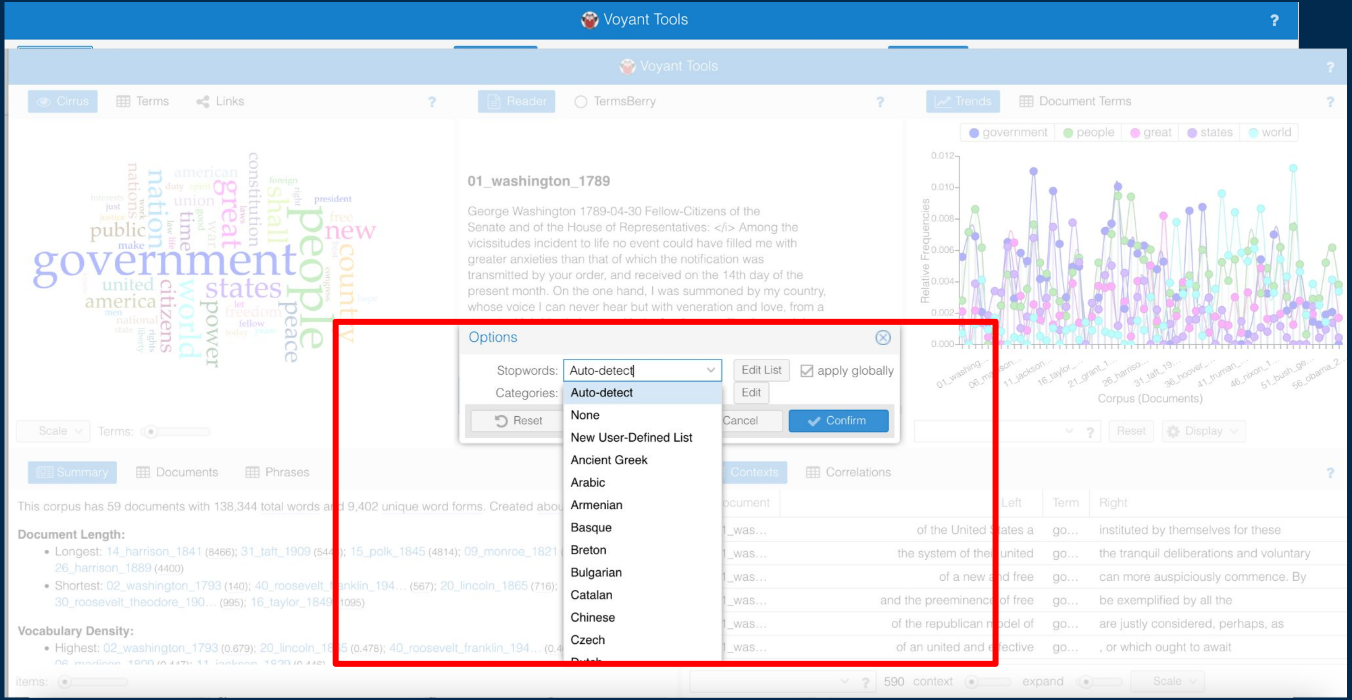
Voyant Tools

In the options box, click on "None." Then click Confirm. What happened?

What you just removed was a "stop words" list.

Click on options again, and click on "auto-detect." Then click on the Edit list button. What do you notice about the words?

When would you want to filter certain words out? When *wouldn't* you want to remove them? What are the implications?



George Washington 1789-04-30 Fellow-Citizens of the Senate and of the House of Representatives: Among the vicissitudes incident to life no event could have filled me with greater anxieties than that of which the notification was transmitted by your order, and received on the 14th day of the present month. On the one hand, I was summoned by my country, whose voice I can never hear but with veneration and love, from a

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms.

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5444); 15_polk_1845 (4814); 09_monroe_1821; 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

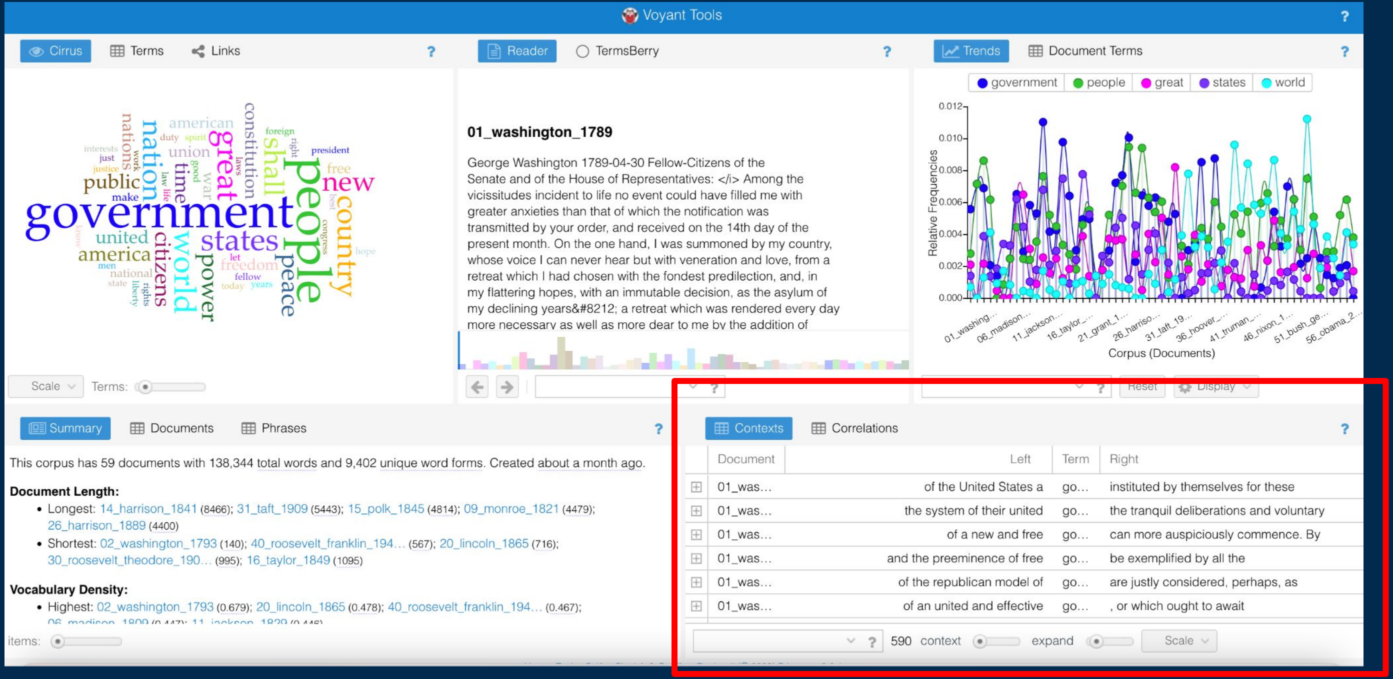
Vocabulary Density:

- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194…

For more about stopwords––their history and their role in computational analysis today–– see this article by Daniel Rosenberg, "Stop, Words." *Representations* 127, no. 1 (August 1, 2014): 83–92. https://doi.org/10.1525/rep.2014.127.1.83.

In the bottom right is a CONCORDANCE.
This gives you the context of words in your corpus as they appear in each document.
Try typing in “government” and sorting by the words that appear on the left.
Toggle to the “Bubblelines” view. Type in “America,” “government,” “liberty.” What do you notice?

Voyant Tools



01_washington_1789

George Washington 1789-04-30 Fellow-Citizens of the Senate and of the House of Representatives: </i> Among the vicissitudes incident to life no event could have filled me with greater anxieties than that of which the notification was transmitted by your order, and received on the 14th day of the present month. On the one hand, I was summoned by my country, whose voice I can never hear but with veneration and love, from a retreat which I had chosen with the fondest predilection, and, in my flattering hopes, with an immutable decision, as the asylum of my declining years&#8212; a retreat which was rendered every day more necessary as well as more dear to me by the addition of

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms. Created about a month ago.

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5443); 15_polk_1845 (4814); 09_monroe_1821 (4479); 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

Vocabulary Density:

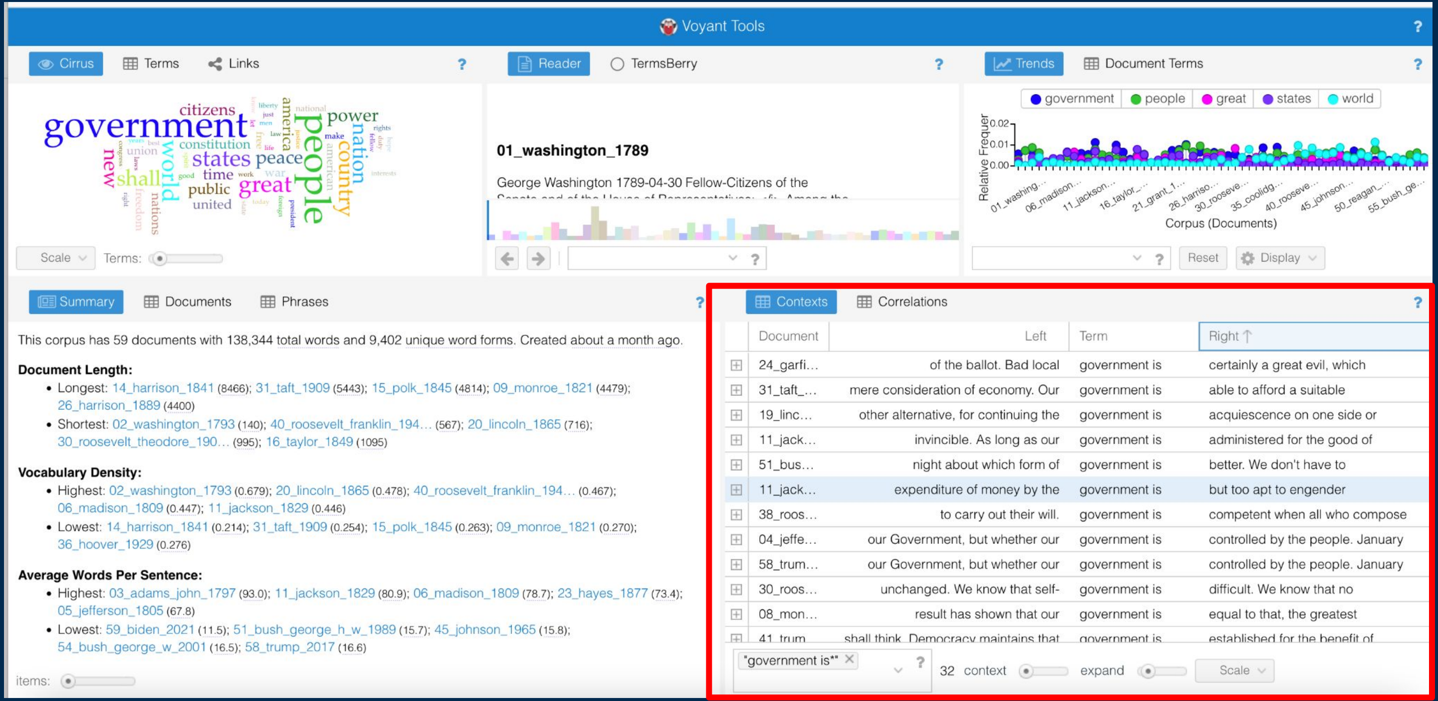
- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194… (0.467); 06_madison_1809 (0.447); 11_jackson_1829 (0.446)

	Document	Left	Term	Right
⊞	01_was…	of the United States a	go…	instituted by themselves for these
⊞	01_was…	the system of their united	go…	the tranquil deliberations and voluntary
⊞	01_was…	of a new and free	go…	can more auspiciously commence. By
⊞	01_was…	and the preeminence of free	go…	be exemplified by all the
⊞	01_was…	of the republican model of	go…	are justly considered, perhaps, as
⊞	01_was…	of an united and effective	go…	, or which ought to await

590 context expand Scale

Voyant Tools

You can also look for short phrases in context:
Try typing in “I think” and sorting by the words that appear on the right.
What do you notice?



01_washington_1789

George Washington 1789-04-30 Fellow-Citizens of the

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms. Created about a month ago.

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5443); 15_polk_1845 (4814); 09_monroe_1821 (4479); 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

Vocabulary Density:

- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194… (0.467); 06_madison_1809 (0.447); 11_jackson_1829 (0.446)
- Lowest: 14_harrison_1841 (0.214); 31_taft_1909 (0.254); 15_polk_1845 (0.263); 09_monroe_1821 (0.270); 36_hoover_1929 (0.276)

Average Words Per Sentence:

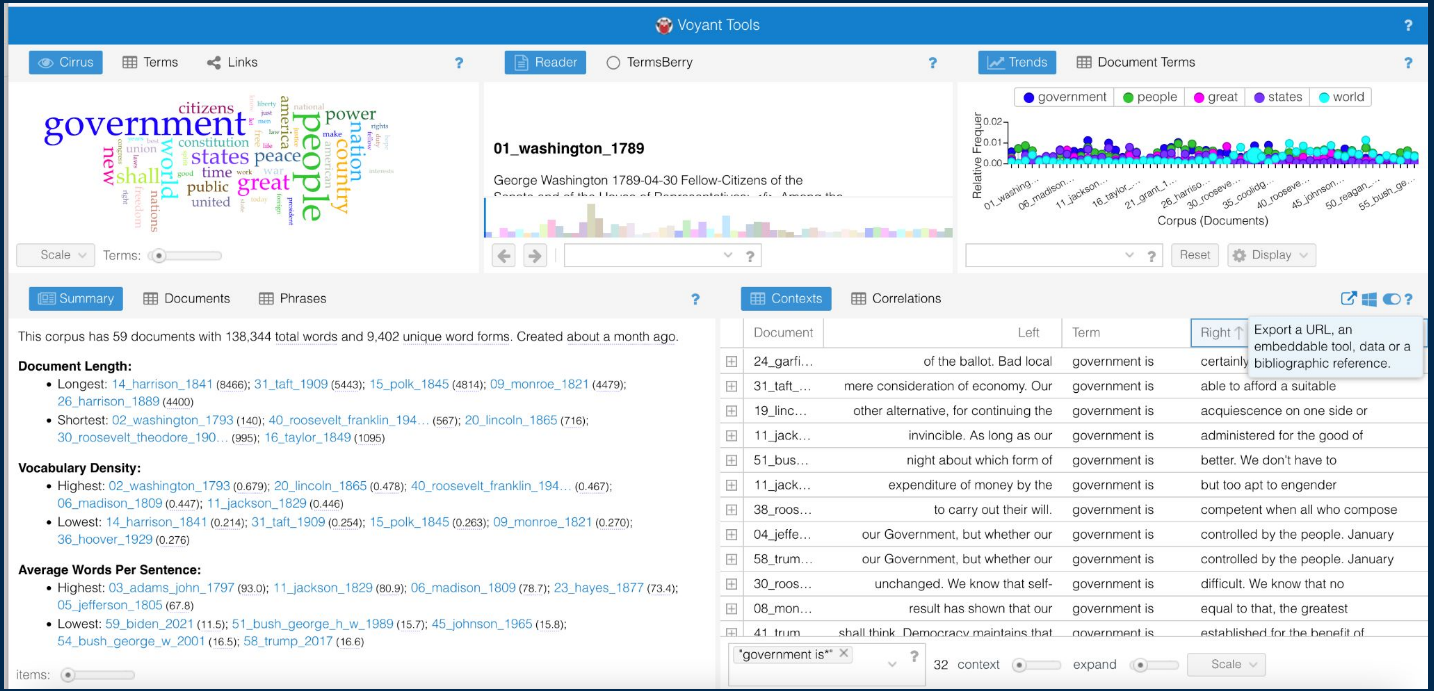
- Highest: 03_adams_john_1797 (93.0); 11_jackson_1829 (80.9); 06_madison_1809 (78.7); 23_hayes_1877 (73.4); 05_jefferson_1805 (67.8)
- Lowest: 59_biden_2021 (11.5); 51_bush_george_h_w_1989 (15.7); 45_johnson_1965 (15.8); 54_bush_george_w_2001 (16.5); 58_trump_2017 (16.6)

	Document	Left	Term	Right ↑
⊞	24_garfi…	of the ballot. Bad local	government is	certainly a great evil, which
⊞	31_taft_…	mere consideration of economy. Our	government is	able to afford a suitable
⊞	19_linc…	other alternative, for continuing the	government is	acquiescence on one side or
⊞	11_jack…	invincible. As long as our	government is	administered for the good of
⊞	51_bus…	night about which form of	government is	better. We don't have to
⊞	11_jack…	expenditure of money by the	government is	but too apt to engender
⊞	38_roos…	to carry out their will.	government is	competent when all who compose
⊞	04_jeffe…	our Government, but whether our	government is	controlled by the people. January
⊞	58_trum…	our Government, but whether our	government is	controlled by the people. January
⊞	30_roos…	unchanged. We know that self-	government is	difficult. We know that no
⊞	08_mon…	result has shown that our	government is	equal to that, the greatest
⊞	41_trum…	shall think. Democracy maintains that	government is	established for the benefit of

"government is*" ✕

Finally, Voyant will also allow you to download data and visualizations.
Hover over the upper right corner of the Concordance view and click on the arrow and box export view

Voyant Tools



01_washington_1789

George Washington 1789-04-30 Fellow-Citizens of the

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms. Created about a month ago.

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5443); 15_polk_1845 (4814); 09_monroe_1821 (4479); 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

Vocabulary Density:

- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194… (0.467); 06_madison_1809 (0.447); 11_jackson_1829 (0.446)
- Lowest: 14_harrison_1841 (0.214); 31_taft_1909 (0.254); 15_polk_1845 (0.263); 09_monroe_1821 (0.270); 36_hoover_1929 (0.276)

Average Words Per Sentence:

- Highest: 03_adams_john_1797 (93.0); 11_jackson_1829 (80.9); 06_madison_1809 (78.7); 23_hayes_1877 (73.4); 05_jefferson_1805 (67.8)
- Lowest: 59_biden_2021 (11.5); 51_bush_george_h_w_1989 (15.7); 45_johnson_1965 (15.8); 54_bush_george_w_2001 (16.5); 58_trump_2017 (16.6)

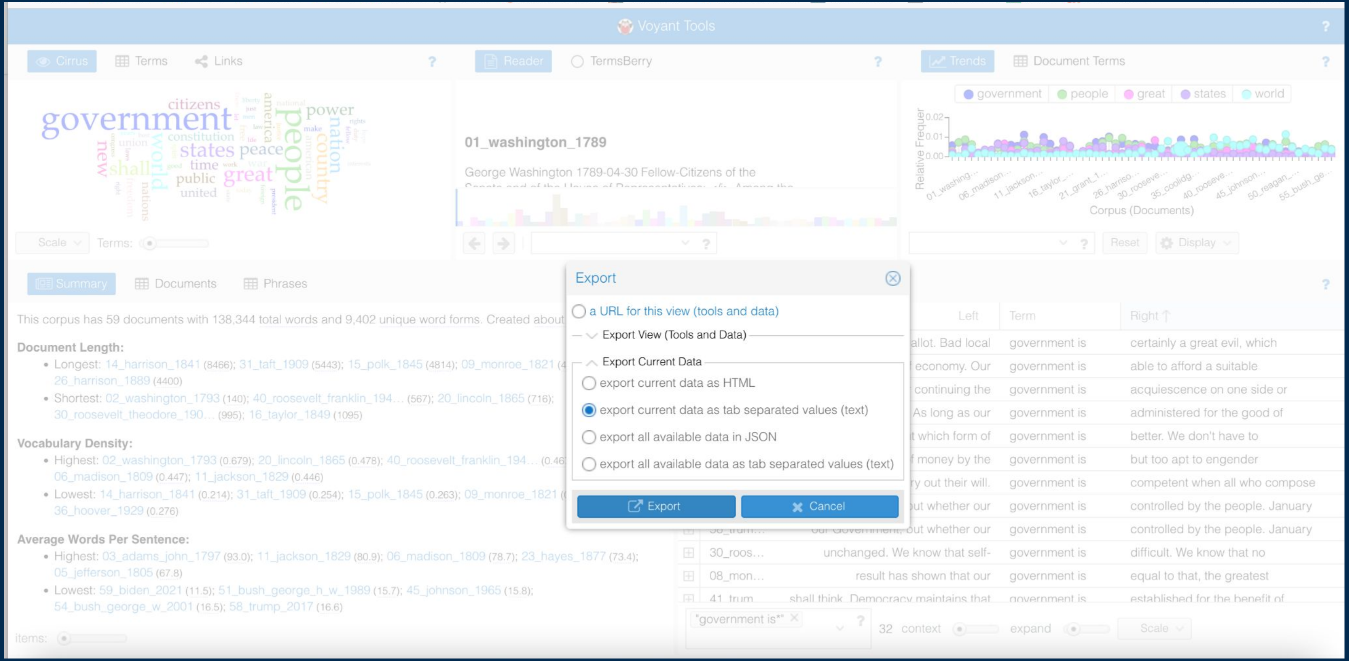
Export a URL, an embeddable tool, data or a bibliographic reference.

	Document	Left	Term	Right ↑
⊞	24_garfi…	of the ballot. Bad local	government is	certainly
⊞	31_taft_…	mere consideration of economy. Our	government is	able to afford a suitable
⊞	19_linc…	other alternative, for continuing the	government is	acquiescence on one side or
⊞	11_jack…	invincible. As long as our	government is	administered for the good of
⊞	51_bus…	night about which form of	government is	better. We don't have to
⊞	11_jack…	expenditure of money by the	government is	but too apt to engender
⊞	38_roos…	to carry out their will.	government is	competent when all who compose
⊞	04_jeffe…	our Government, but whether our	government is	controlled by the people. January
⊞	58_trum…	our Government, but whether our	government is	controlled by the people. January
⊞	30_roos…	unchanged. We know that self-	government is	difficult. We know that no
⊞	08_mon…	result has shown that our	government is	equal to that, the greatest
⊞	41_trum	shall think. Democracy maintains that	government is	established for the benefit of

"government is*" ✕

Finally, Voyant will also allow you to download data and visualizations.
Hover over the upper right corner of the Concordance view and click on the arrow and box export view
You should see a pop up menu.

Voyant Tools


Voyant Tools
Cirrus
Terms
Links
government citizens power people nation country america states peace great public united shall world new union constitution time
Scale
Terms:
Reader
TermsBerry
01_washington_1789
George Washington 1789-04-30 Fellow-Citizens of the
Trends
Document Terms
government
people
great
states
world
Relative Frequen
0.02
0.01
0.00
01_washing… 06_madison… 11_jackson… 16_taylor_… 21_grant_1… 26_harriso… 30_rooseve… 35_coolidg… 40_rooseve… 45_johnson… 50_reagan_… 55_bush_ge…
Corpus (Documents)
Reset
Display
Summary
Documents
Phrases
Items:
"government is*"
32
context
expand
Scale

This corpus has 59 documents with 138,344 total words and 9,402 unique word forms. Created about

Document Length:

- Longest: 14_harrison_1841 (8466); 31_taft_1909 (5443); 15_polk_1845 (4814); 09_monroe_1821 (4 26_harrison_1889 (4400)
- Shortest: 02_washington_1793 (140); 40_roosevelt_franklin_194… (567); 20_lincoln_1865 (716); 30_roosevelt_theodore_190… (995); 16_taylor_1849 (1095)

Vocabulary Density:

- Highest: 02_washington_1793 (0.679); 20_lincoln_1865 (0.478); 40_roosevelt_franklin_194… (0.46 06_madison_1809 (0.447); 11_jackson_1829 (0.446)
- Lowest: 14_harrison_1841 (0.214); 31_taft_1909 (0.254); 15_polk_1845 (0.263); 09_monroe_1821 (36_hoover_1929 (0.276)

Average Words Per Sentence:

- Highest: 03_adams_john_1797 (93.0); 11_jackson_1829 (80.9); 06_madison_1809 (78.7); 23_hayes_1877 (73.4); 05_jefferson_1805 (67.8)
- Lowest: 59_biden_2021 (11.5); 51_bush_george_h_w_1989 (15.7); 45_johnson_1965 (15.8); 54_bush_george_w_2001 (16.5); 58_trump_2017 (16.6)

Export

○ a URL for this view (tools and data)

Export View (Tools and Data)

Export Current Data

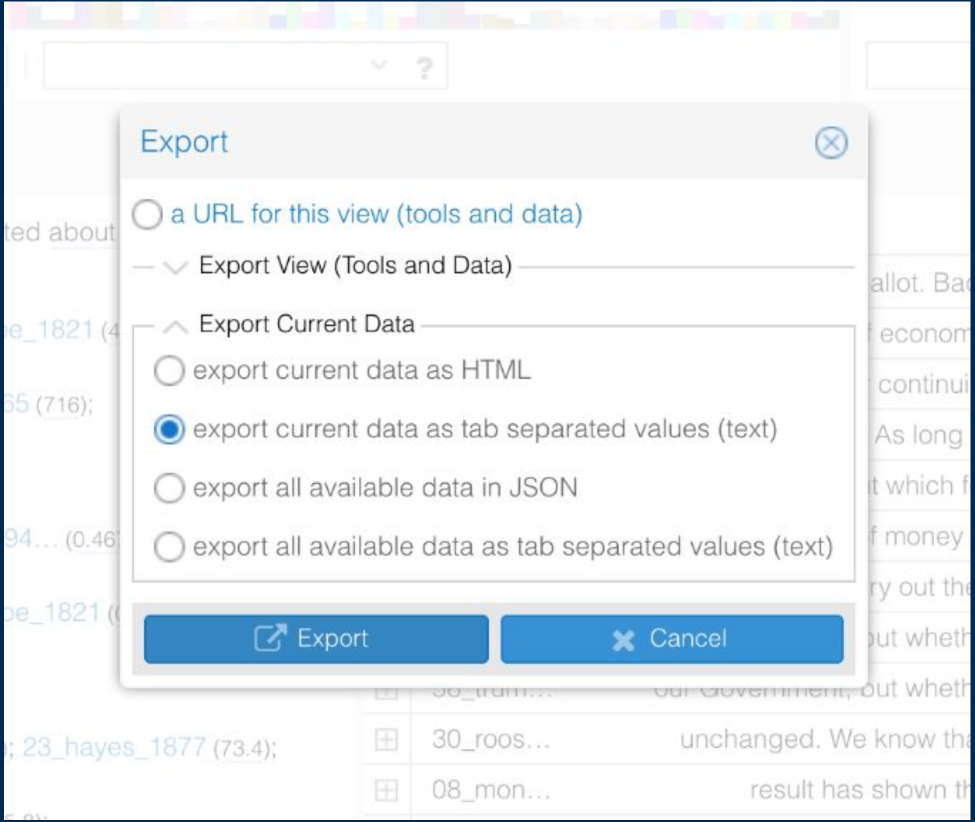
○ export current data as HTML
● export current data as tab separated values (text)
○ export all available data in JSON
○ export all available data as tab separated values (text)

Export | Cancel

	Left	Term	Right
	allot. Bad local	government is	certainly a great evil, which
	economy. Our	government is	able to afford a suitable
	continuing the	government is	acquiescence on one side or
	As long as our	government is	administered for the good of
	t which form of	government is	better. We don't have to
	f money by the	government is	but too apt to engender
	y out their will.	government is	competent when all who compose
	ut whether our	government is	controlled by the people. January
38_trum…	our Government, out whether our	government is	controlled by the people. January
30_roos…	unchanged. We know that self-	government is	difficult. We know that no
08_mon…	result has shown that our	government is	equal to that, the greatest
41_trum…	shall think. Democracy maintains that	government is	established for the benefit of

Finally, Voyant will also allow you to download data and visualizations.
Hover over the upper right corner of the Concordance view and click on the arrow and box export view
You should see a pop up menu. Click on the third option to “Export Current Data” and select tab sep. values

Voyant Tools



Export

○ a URL for this view (tools and data)

Export View (Tools and Data)

Export Current Data

○ export current data as HTML

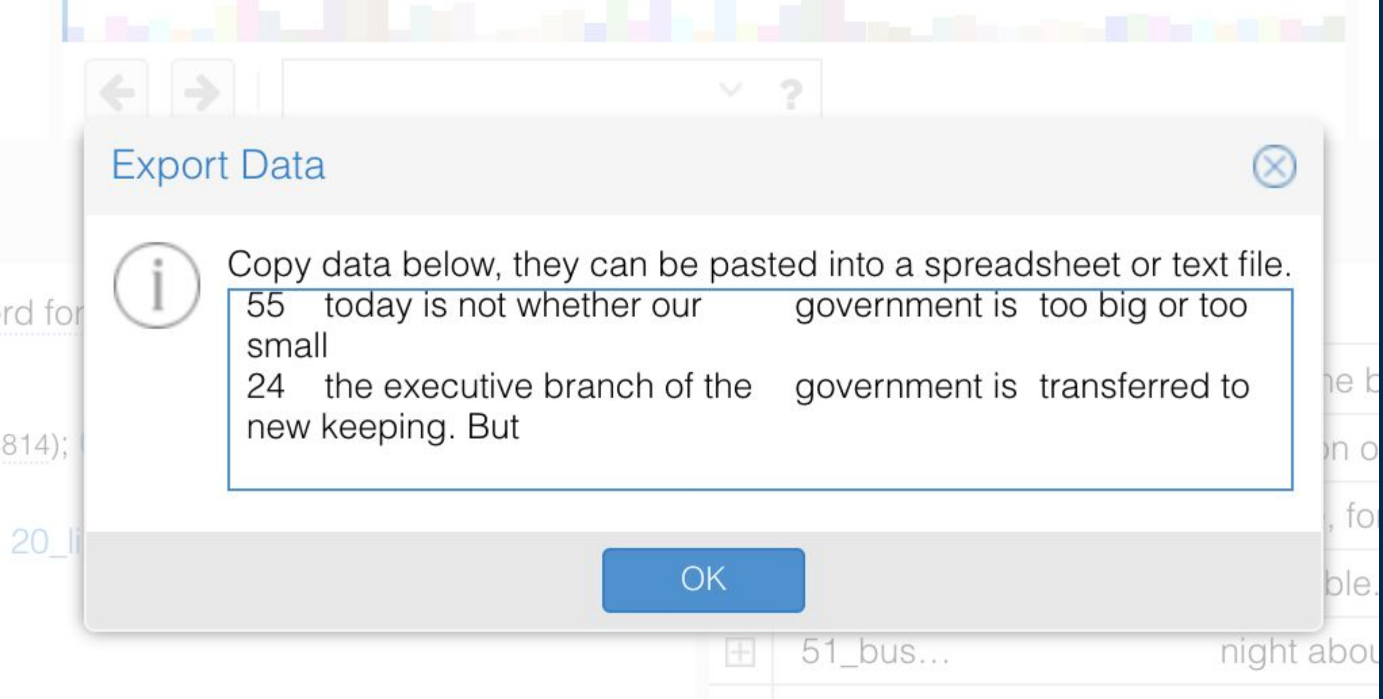
● export current data as tab separated values (text)

○ export all available data in JSON

○ export all available data as tab separated values (text)

Voyant Tools

Exporting the current data as tab separated values (text) will give you a second popup window with data formatted in TSV that can can be copied into a spreadsheet (like Excel or Googlesheets) or a simple text editor



Export Data

Copy data below, they can be pasted into a spreadsheet or text file.

```
55	today is not whether our	government is	too big or too small
24	the executive branch of the	government is	transferred to new keeping. But
```

Voyant Tools

Take a minute to play around some of the features. Toggle the amount of words in the CONCORDANCE, or the “items” in the STATISTICS box.

Brainstorm a few questions that you could explore with this kind of interface.

What kind of questions could you ask?
What kind of questions could you not ask?