Text Analysis
Introduction to Voyant

Dr. Sierra Eckert
Princeton University
sceckert@princeton.edu
An off-the-shelf alternative for exploratory data analysis:

Voyant - a basic dashboard for text analysis

1. Topic modeling browser
An off-the-shelf alternative for exploratory data analysis:

Voyant - a basic dashboard for text analysis

1. Topic modeling browser

If you’re interested, I’m happy to share a simple tutorial I developed using today’s corpus!
On the upper left is a WORD CLOUD. This is a list of words sized by their frequency.
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.
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?
In the bottom left are STATISTICS about your corpus. Scrolling down in the “Summary” view gives a longer list of some of the most distinctive words in each text, the average document and sentence length. “Phrases” allows you to sort by short phrases.
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.
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.
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

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

Voyant Tools

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
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?