

Text Analysis using Voyant and Sentiment Viz

What you'll learn:

- Textual data sources and formats
- How certain toolkits differently analyze and present the same data
- Advantages and disadvantages of text analysis

Part 1: Voyant Term Trends

1. We're going to explore some tools which identify sentimental tone in language, based off of categorizations of words. Open the [Digital Scholarship Hub Dataset folder](#) and download the dataset *tweets_cc_text*.
2. Navigate to <https://voyant-tools.org/>
3. Click on the **Upload** button. Find where you saved *tweets_cc_text* and click **Upload**.

The screenshot shows the 'Add Texts' interface in Voyant Tools. It features a text input field with the placeholder text 'Type in one or more URLs on separate lines or paste in a full text.' Below the input field are three buttons: 'Open', 'Upload', and 'Reveal'. A blue arrow points from the 'Upload' button to a file explorer window below. The file explorer shows a list of files:

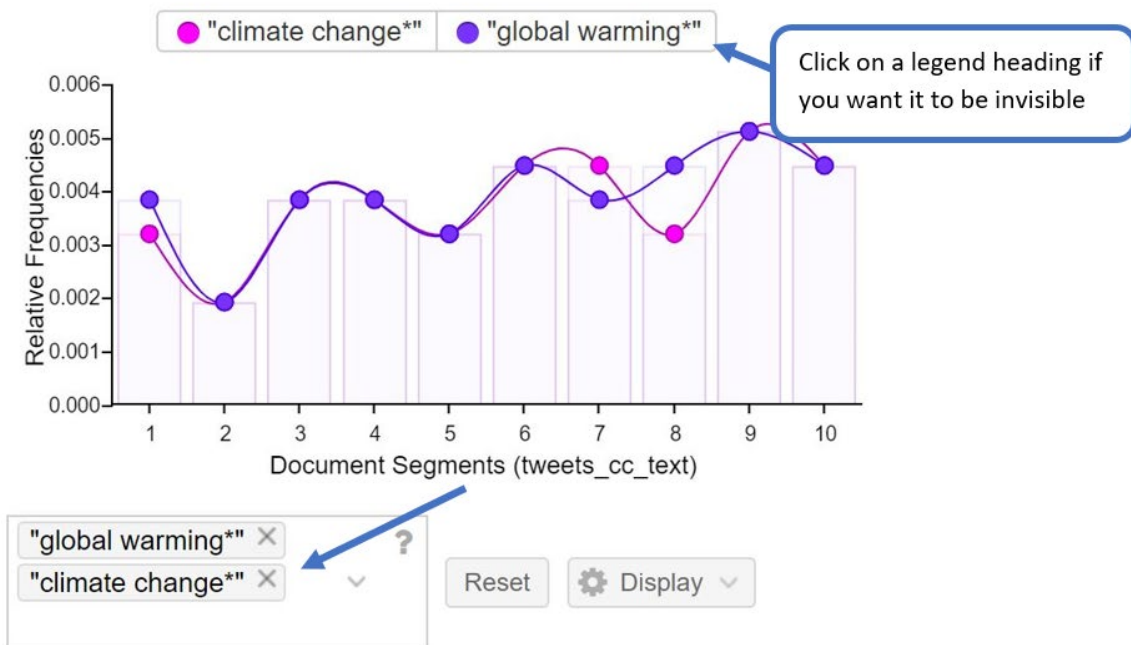
File Name	Modified	Type	Size
tweets_cc_orig	9/8/2022 8:48 AM	File folder	
tweets_cc_clean	9/8/2022 12:55 PM	Microsoft Excel Work...	27 KB
tweets_cc_text	9/8/2022 12:56 PM	Microsoft Excel Work...	19 KB

4. Take a few minutes exploring the different visualizations. Some features to notice:

The screenshot shows a web-based text analysis tool. On the left, a word cloud features terms like 'iphone', 'change', 'climate', and 'warming'. In the center, a list of tweets is displayed. On the right, a line graph plots 'Relative Frequencies' across 'Document Segments'. Three callout boxes provide instructions: 1. 'Some visualizations are hidden. Click to see them.' points to a 'Documents' tab. 2. '“Documents” = number of files you uploaded, not the # of Tweets.' points to a 'Documents' section in the left sidebar. 3. 'Pay attention to optional settings in each visualization. Usually at the bottom.' points to a settings area at the bottom of the graph.

6. Click on the **Trends Term Search** and type **Climate Change**. Next, select the “climate change*” term which appears beneath your cursor.
7. Repeat these steps, and search **Global Warming**. You should end up with a graph showing “climate change*” vs “global warming*”. The X axis defaults to 10 segments. There’s a total of 100 tweets in the dataset, so they’re being grouped into 10’s, presenting average frequencies.

This process is a little buggy, and you may get a duplicate. Just click the X to delete it.



Term Trends Tips:

- You can change the number of segments. Click on **Display** (to the right of where you typed in the terms). Slide the **Segments** bar all the way to the right, and you will have 100 segments. Click on **Palette** to change bar/line colors.

Options

Stopwords: Auto-detect apply globally

Categories: auto

Segments:

frequencies: Raw Relative

Palette: default

- Preserve your visualization as an image file by clicking on **Export** on the upper righthand side of the **Terms Trends Tool**.

Define options for this tool.

a URL for this view (tools and data)

— Export View (Tools and Data)

— Export Current Data

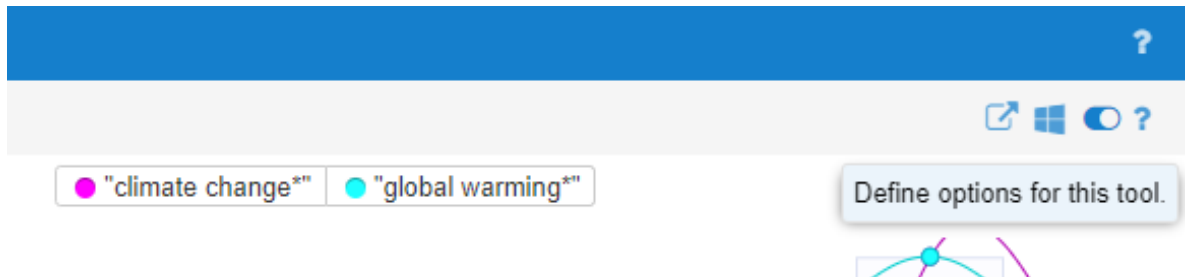
^ Export Visualization

export a PNG image of this visualization

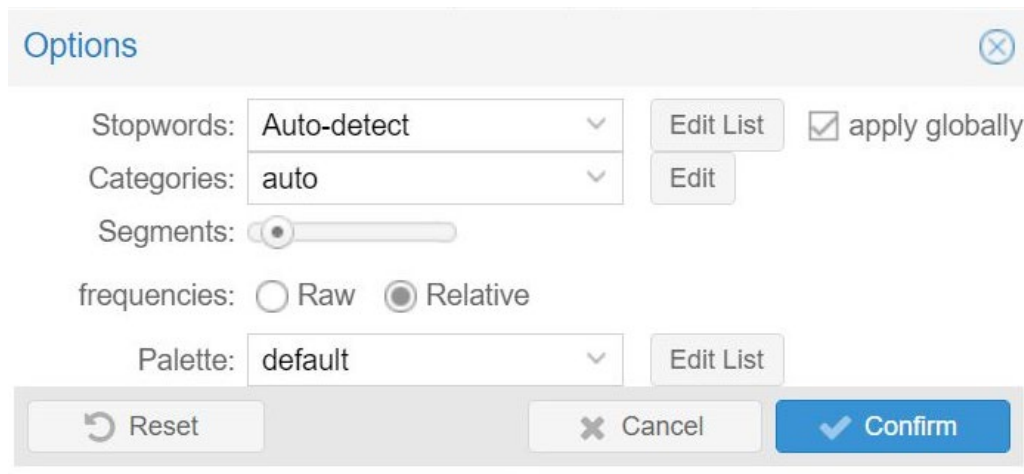
scaling (1):

Part 2: Term Trends Advanced Features

1. On the upper righthand corner of **Trends** click on the **on/off toggle**, which allows you to define options for the tool.



2. Drag the **Segments** bar all the way to the right. Doing so will make 100 categories on the X axis, meaning that each tweet will have its own data point. You can also change the color options, if so desired. Click **Confirm**.

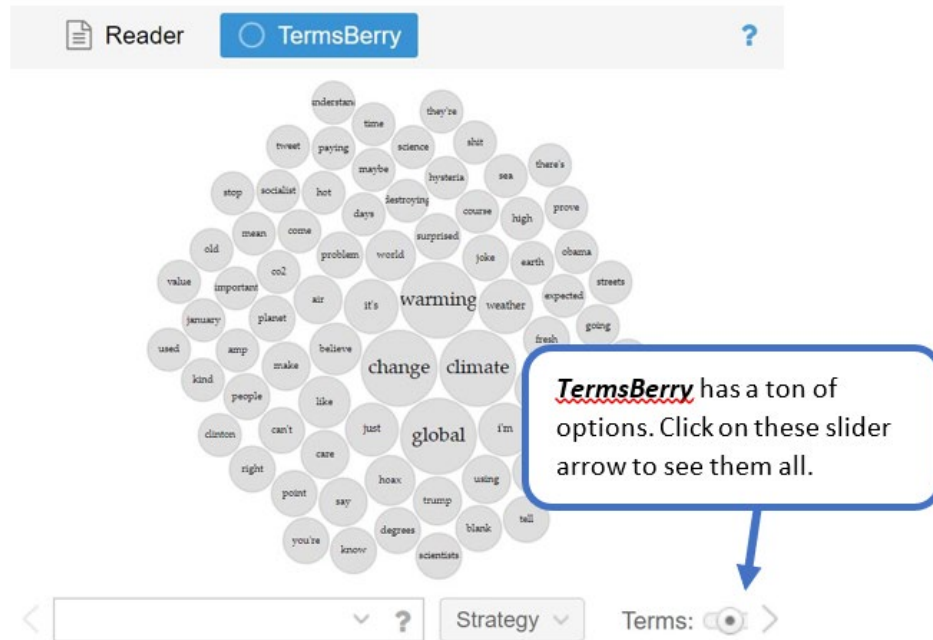


Part 3: Voyant TermsBerry

1. We're using the **TermsBerry** tool now, which shows relationships between high-frequency terms.

Find the **TermsBerry** tool. It usually defaults to the center of the screen. You may need to click on the tool name to see it.

2. Hover your mouse over a term. It should change color, indicating it's your **target**. Notice that other terms change color as well; there's a number beneath each one. **Numbered words** most frequently appear in context with the **target**; the number indicates the actual frequency.



3. At the bottom of the tool, there are some options:
 - **Strategy** impacts which term relationships are selected
 - **Top Terms** (default): the highest frequency terms in the corpus
 - **Distinct Terms**: a collection of distinct terms in each document (as measured by TF-IDF). FYI: TF-IDF calculation is more suitable when you upload more than one file.
 - **Terms**: a slider that determines how many terms to show (default is 75)
 - **Context**: a slider to show how many terms to consider in context on *each* side of each keyword for the collocates (default is 2).
 - **Scaling**: a slider which alters the size of the circles (default is 3).
4. With Voyant – or any visualization tool – it's critical for the analyst to consider whether the result is meaningful. Ask yourself:
 - Is Voyant useful? Trustworthy? Problematic? Why?
 - How should Voyant be used in conjunction with tools like Constellate or Scopus? Try thinking of specific research questions, alternative datasets (i.e. besides Twitter, etc.)

- Are the complexities of this digital methodology balanced with respect to the results you acquire? In other words: is Voyant worth the effort?

Your answers will change depending on the nature of your dataset and project.

Part 4: Exporting your Voyant Visualizations

Exporting your whole project:

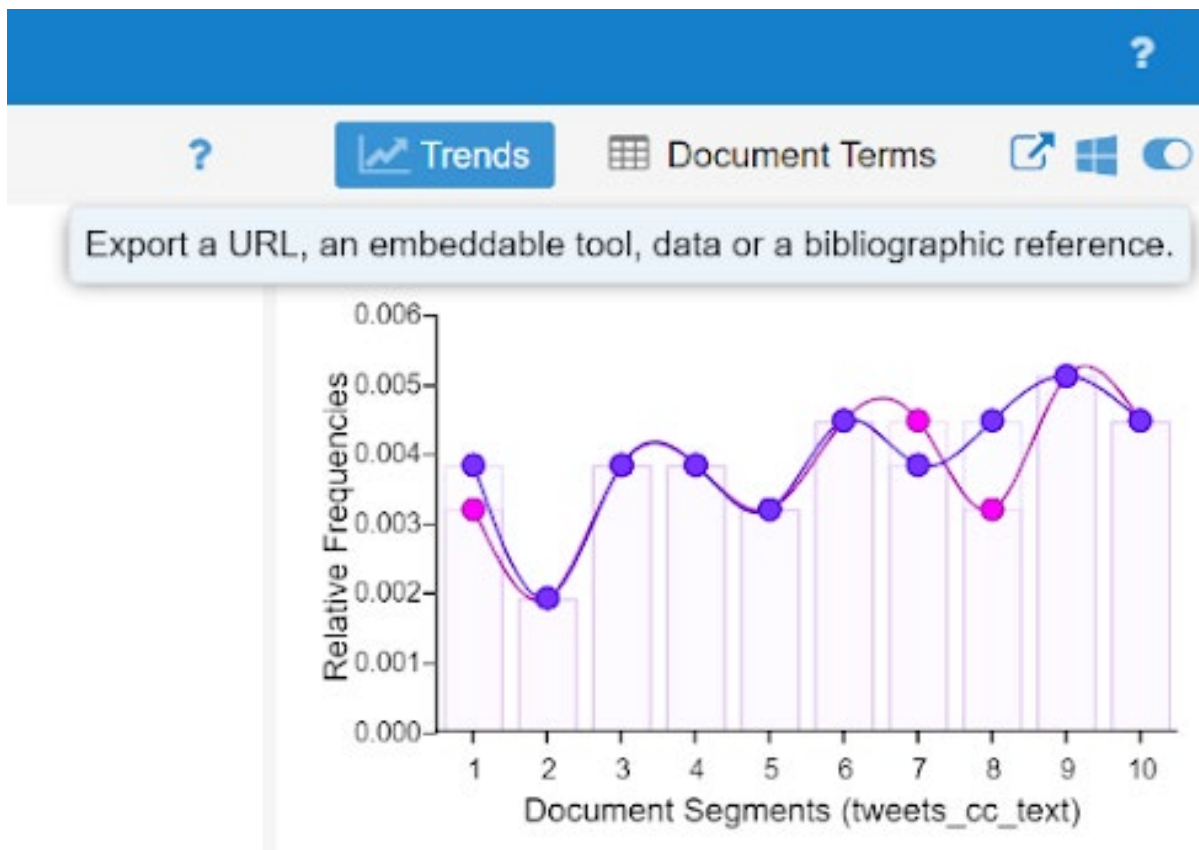
1. At the very top righthand of the page, hover the mouse near the big **?** and click on the **Export** button.



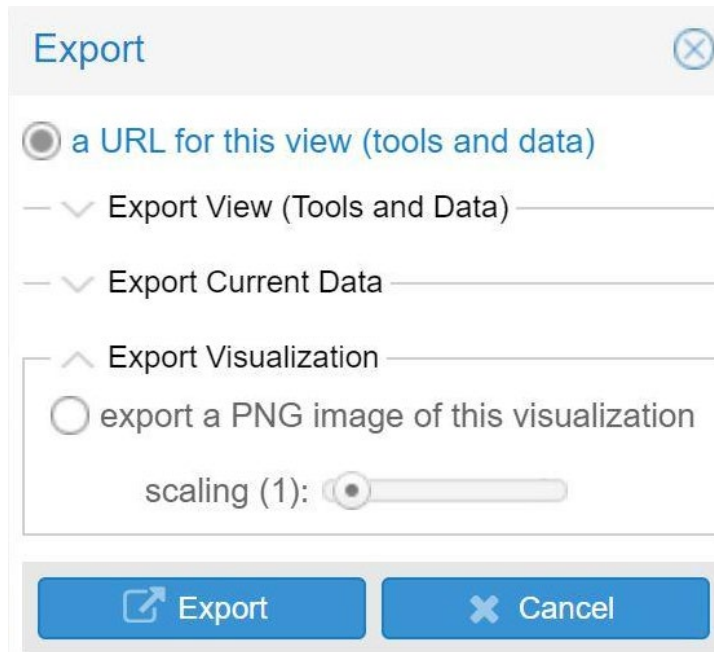
2. Click through the menu. This option is going to give you a shareable link that will share the entire project (i.e. all of the visualizations on your screen, with the data). Note that this is a copy and if someone adjusts settings, your own version will remain the same.

Exporting a specific visualization:

1. On the upper righthand corner of a tool, hover the mouse and click on the **Export** button.



2. Choose your option.



Part 5: Sentiment viz (optional)

This app will pull data directly from Twitter and produce visualizations about how often that keyword is discussed in a positive or negative way. There's a small key at the bottom of the page describing what each visualization is showing. Unfortunately, the Map option has been down for a few weeks. If you want a more in-depth description about their methods, see this [page](#).

1. Access the Sentiment Viz tool at https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/
2. Try typing “global warming” into the **Keywords** query.
3. Explore the different outputs.