Natalie M. Susmann, PhD
Brandeis University | Fall 2023

Text Analysis

Detecting Trends Using Constellate and Scopus

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

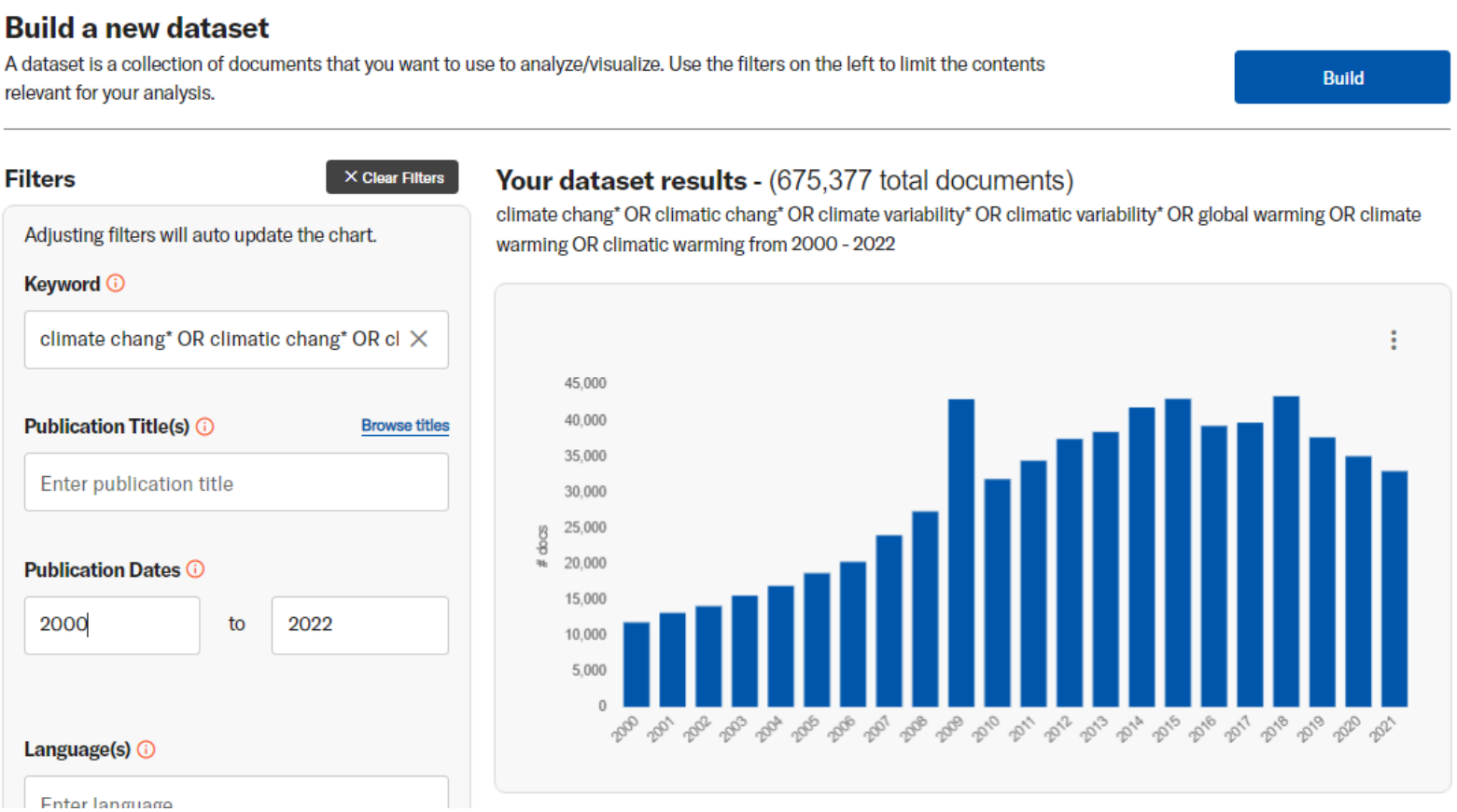
Part 1: Basic Text Analysis in Constellate

1. Create a shared folder on Google Drive.

2. Navigate to https://constellate.org. Login with Brandeis credentials.

3. Insert the following into the ***Keyword*** search bar:
   climate chang* OR climatic chang* OR climate variability* OR climatic variability* OR global warming OR climate warming OR climatic warming

4. Adjust the ***Publication Date*** range to 2000 – 2022.



5. Scroll down and look at the different filters on the left. Stop at the ***Document Type*** bar. Toggle on/off the various publication types. As you do, pay attention to the ***Bar Graph.***

6. Ask yourself: Which publication type(s) yield interesting results to you? What additional steps could you take (using Constellate or not!) to gain further context about those results?

Part 2: Term Frequency in Constellate

Understanding the Research Scenario:

Global warming refers to human-induced heating of the Earth's surface, and this has happened since 1850. **Climate change** refers to deep-time, average weather patterns (happening on a local and global level) caused by both natural processes and human activity.[1] Despite their distinctive meanings, there are many cases where experts *and* the general public use them interchangeably.[2]

Using Constellate to Analyze Global Warming:

7. Keep the ***Keyword*** search **blank**. Notice that if you don't search for any keywords, the ***Bar Graph*** is still showing you data – this is the total number of publications per year.

[1] Earth Sciences Communication Team at NASA's Jet Propulsion Laboratory. 2022. "Global Warming vs. Climate Change." *Global Climate Change: Vital Signs of the Planet.* Site Accessed on September 1, 2022. https://climate.nasa.gov/global-warming-vs-climate-change/

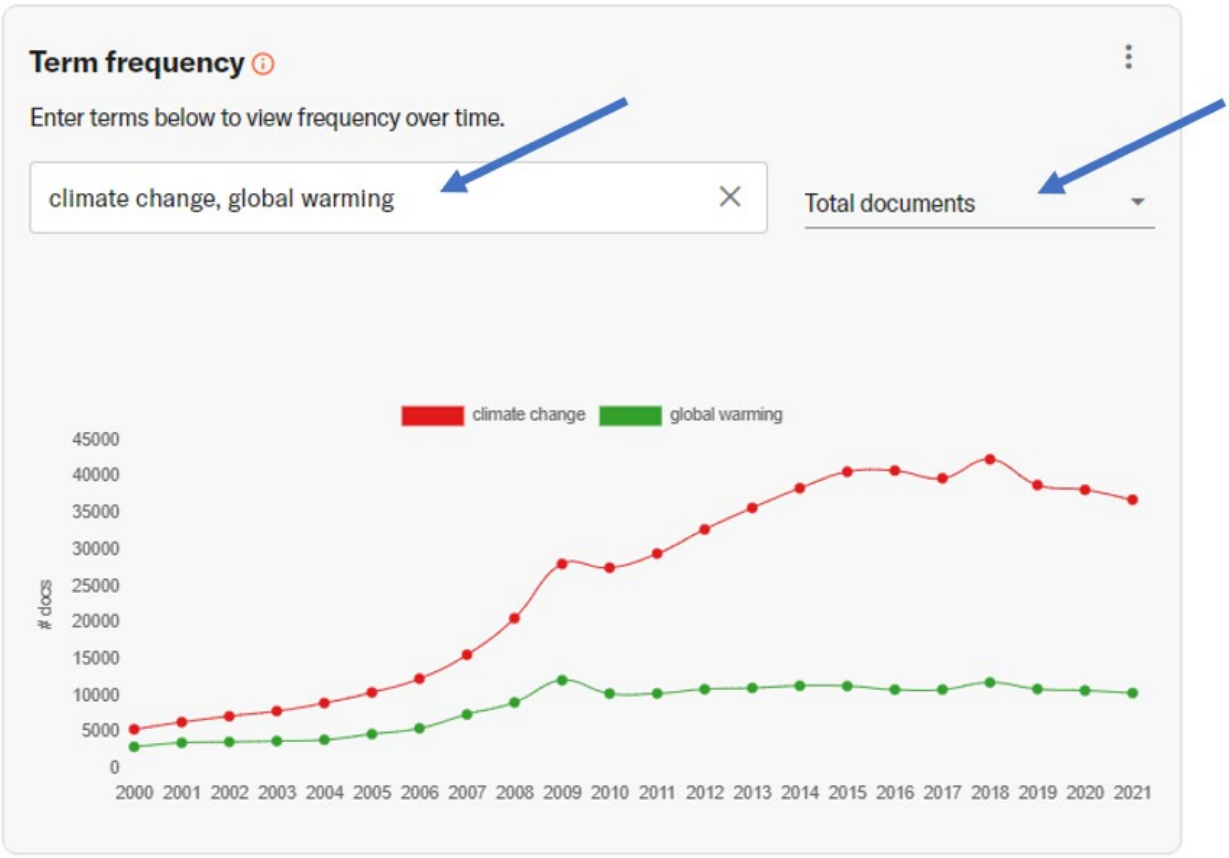
[2] Ming Liu and Jingyi Huang. "'Climate change' vs. 'global warming': A corpus-assisted discourse analysis of two popular terms in *The New York Times." Journal of World Languages* 8(1): 34-55.

8. Adjust ***Publication Date*** range to 2000 – 2022.

9. Scroll down to the ***Term Frequency*** box. Constellate might have some default terms already entered for you. In the ***Terms Frequency*** box, enter:

   global warming, climate change

10. Notice that you can choose between ***% of documents*** and ***Total documents*** in the adjacent dropdown.



11. Save the **Total Documents version** of your visualization as an image. Click on ***the 3-dot menu*** on the upper right-hand corner. Save the image to your shared Google Drive folder. Or, take a screen shot and share it to the folder.

12. Ask yourself these questions:

    - What is this visualization showing? Make sure you understand the difference between ***% of documents*** vs ***total documents***.

    - Thinking back to the confusion between "climate change" and "global warming", does this visualization offer any explanation or context? For example: who is making the error, why it's happening, if the error is more prevalent at certain points in time?

Part 3: Term Frequency in Scopus

13. We're going to explore that same "global warming" vs "climate change" trend using a different text analysis interface. Go to https://guides.library.brandeis.edu/az.php?a=s → find ***Scopus*** in the different databases listed.

14. Click the ***Scopus*** link and follow the prompts for logging in + Duo authentication.

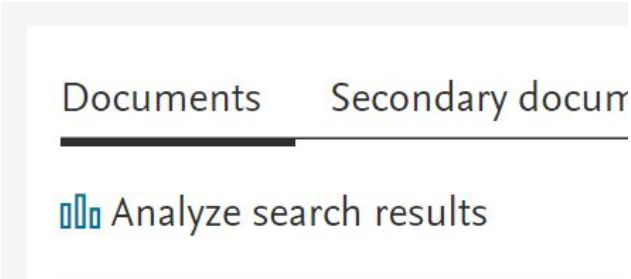
15. Notice the ***Search within*** dropdown menu. There are a variety of options which could impact the results of your search. Stick to the default: ***Abstract title, Abstract, Keywords.*** This will ensure your search is closely linked to the article contents.

16. Insert "global warming" (don't type " ") in ***Search documents*** and click ***Search*** (or type ***Enter***).

17. Explore the side bar. Notice that you can select ***Limit to*** or ***Exclude*** and then various filters beneath that.

18. Reset any filters you selected; you want to be looking at the same as in from step #10.

19. Click ***Analyze search results.***



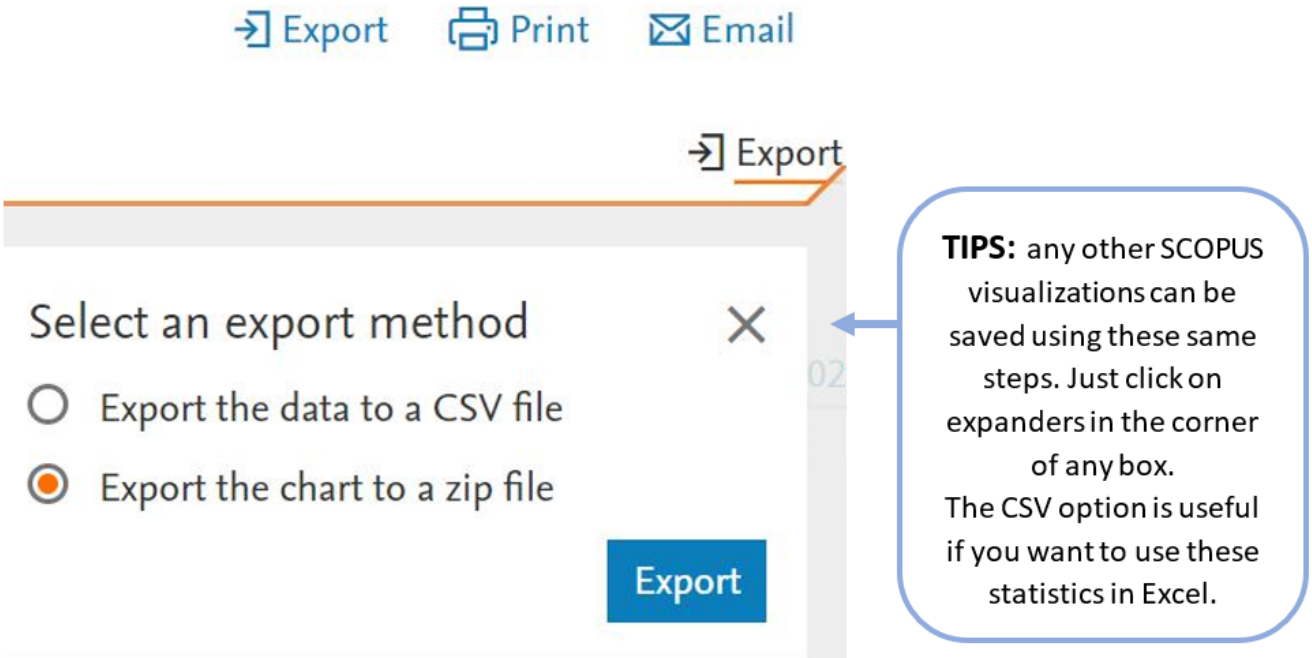
20. Now you'll see an interface similar to Constellate. Explore the different visualizations.

21. On the ***Select year range to analyze*** setting, adjust the years so that they match what we used in the Constellate analysis. Click ***Analyze.***

22. Look at the **Documents by year** visualization.

23. Export the **Documents by year visualization** as an image. There are two ways to do so.
    - **Quick:** take a screen shot and save the image to your shared Google Drive folder.
    - **Slower + Better Quality**: Look above the ***Documents by year*** label → ***Export*** → ***Export the chart to a zip file.*** A folder will save to your computer, and inside will be a jpeg and png version of the graph. Move this to your shared folder.

24. Regardless of which method you use, keep in mind that the visualizations do not have titles embedded in the graphic. I recommend changing the file name to indicate it's about Global Warming (or whatever your research topic is) so you know what the visualization is showing.



25. Now, use SCOPUS to replicate this search, this time with keywords “climate change”. Make sure you use the same date range as in step #19. Export the **Documents by Year** visualization to your Google Drive folder. Make sure “climate change” is somewhere in the file name.

26. Go to your Google Drive folder and open up the images you’ve made:
    - Constellate Term Frequency (global warming vs climate change) (from step #10)
    - SCOPUS “Global Warming” Frequency (from step #21)
    - SCOPUS “Climate Change” Frequency (from step #22)

27. The SCOPUS and Constellate visualizations are not identical. What factor(s) are contributing towards these differences? Which tool produces better results? Why?