



Data Visualization and Analytic Tools

Preamble

This group is aimed at students and researchers in the humanities. We need you to try out our search engine on your own project. The goal here is twofold: 1) to help you identify books that are relevant to your own research; and 2) to develop compelling use cases, so that, when we take our work to the press, we can demonstrate what this search engine is capable of!

Main Goals

Your goal is to experiment with various means of visualizing the results derived from our search engine. This could include online tools (Tableau, for example), Principal Component Analysis, enriched maps, timelines, etc. You should not feel constrained to use any particular set of tools, and you may use any you may already be comfortable using. If you're starting from scratch, we recommend visiting the following sites and creating accounts to use their free, web-based data tools, particularly Tableau:

- Google Sheets - <https://sheets.google.com>
- Microsoft Office Excel 365 - <https://login.microsoftonline.com/>
- RAWGraphs - <http://rawgraphs.io/>
- Tableau Public - <https://public.tableau.com/en-us/s/>
- ClustVis: <http://biit.cs.ut.ee/clustvis/>

Other goals / instructions (in order of priority)

Develop tools with the ability to compare concept use across multiple texts. (For this purpose, we recommend you read Mark Schiefsky's article on Arboreal, included in your working group's Github folder.

Performing your own search in advance

If you want to try out visualizing data relative to a project that you are working on, you may want to have us perform a search on your behalf before the hackathon. Stephen Osadetz will complete these in advance and bring your output files on a thumb drive. To have Stephen perform a search on your behalf, send him an e-mail (osadetz@fas.harvard.edu) with a few sentences describing your concept and a list of 10-20 words that you believe would cluster around that concept. For instance, a search for the concept *dog* might include such words as *dog*, *canine*, *kibble*, *tail*, etc. A few tips: we can't search for phrases, so restrict your request to single words; the search algorithm involves stemming words, so there's no need to include variants of a single words like *dog*, *dogs*, etc. The search engine treats these as the same word. Try to choose words that won't result in the same stem. Please keep in mind that at this point we are only able to search 18th-century English-language texts. At this point, search is slow – it takes one hour to do a single search – so it is unlikely we will be able to do a search for you on the day of the event.

GitHub Contents (may be added to before and during the event):

- Sample search results (to be used if you choose not to do your own search in advance)
- Schiefsky article on Arboreal
- Group Notebook
- Code Uploads

Other groups to consult

- **Research Exploration**
- **Corpus Enrichment** (this group will be enriching the ECCO metadata, which may have the ability to improve your visualizations)