

# Phase Two, Week A - Out of the Box Tools

- [Short Tutorial](#)
  - [Voyant](#)
  - [nGrams](#)
  - [Bookworm](#)
- [Group 1 - Voyant](#)
- [Group 2 - nGram Viewer](#)
- [Group 3 - Bookworm](#)
- [Bringing it back together](#)

## Short Tutorial

### Voyant

- <https://voyant-tools.org/?corpus=afbb8be9b5a1b4ed4bd5150715908a48>
- basic orientation to panes
- Basic text/navigation pane
- stopword control
- keyword control
- swapping out tools/visualizations

### nGrams

- ["Ebola" 1975-2008 in various languages](#)
- [data from WHO on Ebola outbreaks](#)
- [map of official languages of African countries](#)

### Bookworm

- [Does truth correlate with beauty?](#)
- facet by language
- constrain years

## Group 1 - Voyant

### Tips

- Suggest that you use Chrome if you are on Windows. It is most responsive and resilient against screen lockups for this tool.
- Your exploration will begin at <https://voyant-tools.org/?corpus=bd21a9fdb8a7e67556ede72ab51d10d>
- Help files for customizing the tool are at <http://voyant-tools.org/docs/#!/guide/start>
- Choose roles of **recorder** and **presenter**

### As a group explore and discuss

- Spend some time exploring the interface. What can you tell about the **underlying data** as you explore with the tool?
  - Is the text clean? Indexed? Filtered? Anything else interesting you note about the data?
- Explore the **functions** of the tool. Attempt to make **claims about the intellectual content of the text** based on the tool and its visualizations. (Feel free to reach a little.)
  - Try constraining the various tools to specific terms or phrases. Do any words trend together? Inversely? What sort of assertions can you tentatively make based on relationships?
  - Can you isolate phrases of interest for special scrutiny?
  - Identify and operate the sliders - in what ways could these be useful?
  - Adjust the stopwords to dampen "noise" and heighten "signal"; what are potentially positive and negative consequences of doing this?
    - How does this affect claims you might make about the text?
    - What strategies would you employ as a result?
  - Do any of the the visualizations help illustrate what you notice/assert regarding the text? Are some misleading? Confusing?
  - Swap out tools in the interface by using the individual toolbars for each pane (hover an the toolbar buttons will appear). Explore those other tools. Are there any that are helpful in exploring this text? Unhelpful?
  - If time - upload a text of interest and explore it as well - define/refine claims about the text by adjusting tools to help depict your assertions about the text.
- Consider the **value** of the tool
  - What can you manage to do? What is this tool good for?
  - What sorts of things did you want to do, but could not?
  - What can you infer from the interface about the text? What is still opaque?

After your exploration, be prepared to report your findings to the other team(s) and take their questions. Be prepared to talk about

- the underlying data
- the ways the tools can be customized
- the utility and potential of the tools, both positive and negative

## Group 2 - nGram Viewer

### Tips

- Any browser should work fine.
- Your exploration can begin [here](#), but you can also develop other interesting explorations. NOTE: you may have to hit the blue "Search Lots of Books" button to bring in the visualization.
- Help files for customizing the tool are at <https://books.google.com/ngrams/info>
- Discussion of underlying data is [here](#).
- Choose roles of **recorder** and **presenter**

### As a group explore and discuss

- Spend some time exploring the notes about the **underlying data**. What kind of text data is this?
  - Is the text clean? Indexed? Filtered? Structured? Anything else interesting you note about the data?
- Explore the **functions** of the tool. Attempt to make **claims about the intellectual content of the text** based on the tool and its visualizations. (Feel free to reach a little; definitely refine and make the input better.)
  - How comprehensive are these terms? Can you make them more comprehensive by making them case insensitive? What happens to your results? What does this mean about your assertions?
  - Can you make some phrases case sensitive while making others case insensitive? How?
  - Can you segment by language to indicate data from Great Britain as distinct from US English? What implications might that have for your developing ideas about these terms? What does it mean to find these terms in Spanish?
  - Can you add related terms? Do these terms show interesting trending, either coincident or inverse?
  - Add the coalesced term "(male+(chauvinism+chauvinist))". Note the way the frequency of this form obliterates the other waveforms. What can you do to temper this action?
  - What sorts of supplemental data would be helpful in making sense of these visualizations?
  - Operate the date parameters.
    - How far can you extend the visualization?
    - What happens to the representation of the data when you focus on a specific year span years? How does that affect the narrative you would tell about the trend of the frequency of a word?
  - Play around - try other searches on other themes and customize them. Observe and evaluate their effects. Try to shoot for something meaningful to share with the other group.
- Consider the **value** of the tool
  - What can you manage to do? What is this tool good for?
  - What sorts of things did you want to do, but could not?
  - What can you infer from the interface about the text? What is still opaque?

After your exploration, be prepared to report your findings to the other team(s) and take their questions. Be prepared to talk about

- the underlying data
- the ways the tools can be customized
- the utility and potential of the tools, both positive and negative

## Group 3 - Bookworm

### Tips

- Any browser should work fine.
- Your exploration will begin [here](#). If this doesn't render, open <https://bookworm.htcr.illinois.edu/develop/>, clear filtering, replace the two terms with "consumption" and "tuberculosis", and run the search.
- Discussion of the use of the tool is at <https://wiki.htcr.illinois.edu/pages/viewpage.action?pageId=26705922> (See the bottom of the page under the heading "Using HT+BW")
- (Extremely short) description of underlying data is [here](#), and greater documentation [here](#).
- Choose roles of **recorder** and **presenter**

### As a group explore and discuss

- Spend some time exploring the notes about the **underlying data**. What kind of text data is this?
  - Is the text clean? Indexed? Filtered? Structured? Anything else interesting you note about the data?
- Explore the **functions** of the tool. Attempt to make **claims about the intellectual content of the text** based on the tool and its visualizations. (Feel free to reach a little; definitely refine and make the input better.)
  - Add other communicable diseases of interest.
  - Try comparing "malaria" as found in publications of the United States or the United Kingdom.
    - Where does the data to facet in this way come from?
    - What narrative could you potentially make from this graph?
    - What sorts of supplemental data would be helpful in constructing this narrative? (Wave your magic wand here...)
  - Conjecture as to what "Class", "Subclass" and "Narrow class" might mean. Where would this faceting data come from?
  - Click on a spot on one of the plotted curves (and wait, rendering can take a little time). What is this data in the drop down? Explore it - how might it be useful?
  - Operate the date sliders.
    - How far can you extend the visualization?

- What happens to the representation of the data when you focus on a specific year span years? How does that affect the narrative you would tell about the trend of the frequency of a word?
  - What sorts of supplemental data would be helpful in making sense of these visualizations?
  - Play around - try other searches and customize them; observe and evaluate their effects.
- Consider the **value** of the tool
  - What can you manage to do? What is this tool good for?
  - What sorts of things did you want to do, but could not?
  - What can you infer from the interface about the text? What is still opaque?

**After your exploration, be prepared to report your findings to the other team(s) and take their questions. Be prepared to talk about**

- the underlying data
- the ways the tools can be customized
- the utility and potential of the tools, both positive and negative

## Bringing it back together

- What sort of conclusions/suspicions can we draw from canned tools in general?
  - In what types or phases of projects would these tools be useful?
  - In what types or phases of projects would we need more control over analysis?