Making a Data Set
http://bit.ly/1PoA0G8

Building a Corpus - An Intellectual Pursuit

From this activity, you will learn:

- How to identify what your data can tell you OR what data you need to answer a question
- Principles of selection

Open the Table of Contents file in the Slave Narratives data set.

Path: na-slave-narratives / data / toc.csv

You will see that you have information about the author, title, and date of each narrative. From this set of data (an example of metadata), please answer the following questions. Feel free to sort or search the data in any way you want.

1. Which texts would you use to answer:
    a. How do texts which identify themselves as autobiographies differ from biographies?
    b. How do constructions of domesticity evolve in slave narratives between 1830 and Reconstruction?

2. For questions 1.a and 1.b, what information would you need to answer this question?
    a. Identify what aspects of the text you'd need to get information from.
    b. Identify what aspects of context you'd need to know.

3. Develop one other question you could ask with a corpus from this set and identify the texts you would use.

Building a Corpus - A Techno-Social Pursuit

From this activity, you will learn:

- How to find texts for text analysis
- How to document your process for best data curation
- How to make the metadata you need

For this activity, you can choose between data sets:
Making a data set for a question you are asking
Making a data set about an animal in the 19th century (like chickens)

1. Find five texts that fit your corpus.
   To do this, you can look at:
   HathiTrust
   Project Gutenberg
   "the internet"
   Google Books
   Europeana

2. For these texts, do the following actions and answer the following questions:
   a. What format(s) can you get these texts in?

3. Make a folder with a title for your project.
   Download or get the texts into this folder.
   What format(s) can you get these texts in?
   Name the files with a metadata scheme that is helpful.
   What elements did you choose to name?
   Make a metadata spreadsheet.

**Bonus: Make a readme file. See example in UNC texts. Also see:
https://zenodo.org/record/13103/files/Readme.txt