

COS 126/CDH Humanities Data Workshop

This workshop will first provide context for three featured humanities-based datasets; students can then select which dataset and corresponding breakout room to join.

These breakout sessions will be an opportunity to define broad questions and hypotheses that emerge from the data, and work with peers to find methods to test their hypotheses computationally (e.g. how would you test for the largest vocabulary computationally).

By the end of the break-out session, students should be comfortable and enthusiastic about one dataset, with a good understanding of the fields it contains and what can and cannot be computed with the sole data available in the dataset.

This workshop is a collaboration between the Center for Digital Humanities and the COS 126 course staff within the Department of Computer Science. Please visit <https://cdh.princeton.edu> to learn more about the Center for Digital Humanities.

FEATURED DATASETS

1. The Endangered Languages Project

- Data (CSV): <https://www.endangeredlanguages.com/userquery/>

2. Shakespeare and Company: Library Lending Data

- Data (CSV): <https://shakespeareandco.princeton.edu/about/data/>

3. The Bechdel Test for Movies

- [Data](#) (spreadsheet/CSV)

Additional Datasets

For these datasets, we do not provide prepared Standard/Sprinkle/Sparkle project ideas, but please feel free to reach out to your preceptor to discuss ideas.

1. [List of Humanities Datasets](#)
2. [Modernist Journals Project](#)
 - a. [Data](#)
3. [Torn Apart/Separados](#)
 - a. [Data](#)
4. [What people had in their houses in the 1700s in rural Pennsylvania](#)
5. [A database of graphic novels](#)
6. [Prisoner records from the Eastern State Penitentiary](#)
7. [A database of Scottish witchcraft](#)
8. [Atlas of Surveillance in the U.S](#)

Schedule a consultation with the CDH
<https://cdh.princeton.edu/engage/consult/>
dhi@princeton.edu



THE CENTER
FOR DIGITAL
HUMANITIES
@PRINCETON