2024-02-07

Data in the Humanities

Wk. 02: Histories of the Humanities; CSVs

THIS WEEK

Workshops

– Thu 2/8 Citation Management with Zotero (PUL)
– Thu 2/8 Introduction to Python for Working with Text (DiScho)
– Thu 2/8 How to Create and Collect Geographic Data using QGIS (PUL)
– Thu 2/8 Getting Started with LaTeX (PUL)
– Mon 2/12 Select and Analyze Geographic Features and Data with QGIS (PUL)
– Tue 2/13 Introduction to Handwritten Text Recognition (HTR) (CDH)

Lectures & Events

– Tue 2/13 Scenario for a Past Future and Avant-Garde Immersive Worlds
(Lewis & CDH)
– Wed 2/14 Annual Douglass Day Transcribathon (w/ HTR) (PUL)

Today's plan

1. 📖 Discuss readings on history of the humanities
2. 🎓 Review breakdown of divisions at Princeton
3. 💾 Browse sources for humanities datasets

1. Readings 📖

The humanities are...

the study of the expressions of the human mind (Dilthey)

humanities = understanding; science = explaining (Dilthey)

the quest for patterns in humanistic material based on methodological principles (Bod)

"the cultural practices of reflection, argument, criticism, and speculative testing of ideas" (Small)

descriptive, appreciative, imaginative, provocative, speculative, critical (Small)

"careful observation applying rigorous skepticism about what is observed . . . focus[ing] focus on small, complex units of evidence" (Hayot)
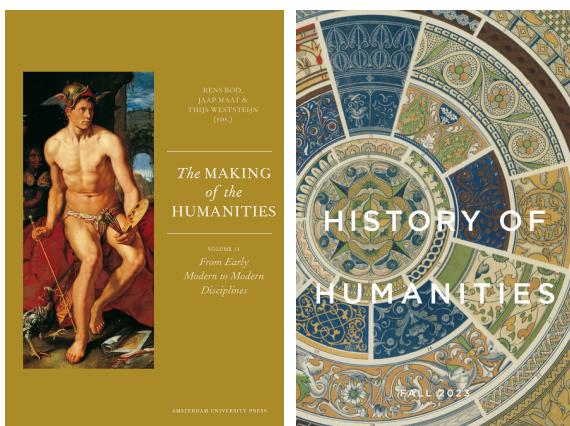
OED's "historical thesaurus"

Wk. 2 Readings

– Eric Hayot, *Humanist Reason: A History. An Argument. A Plan* (Columbia University Press, 2021), Chapters 1 & 4.

– Helen Small, *The Value of the Humanities,* (Oxford: Oxford Univ. Press, 2013). Excerpts.

– Rens Bod, *A New History of the Humanities* (Oxford: Oxford Univ. Press, 2014), Introduction: "The Quest for Principles and Patterns."

distinctive specificity of the humanities

"It is in such circumstances (offensive and defensive) that the qualities by which the humanities claim distinctive purpose become most sharply apparent:

Small, *Value of the Humanities,* p. 29-30

   – their tendency to value qualitative above quantitative reasoning;

   – their distrust of proceduralism;

   – their greater faith in interpretative than in positivistic thinking;

   – their orientation as much toward historical analysis as toward synchronic structural analysis;

   – and their attention to the role of the perceiver in ascertaining even the most philosophically secure of knowledge claims;

   – (relatedly) their interest in the specificity of the individual response (its content and its style) over and above the generalized or collective response, and

   – their concern with what can be known or understood even though it is incapable of empirical verification.

describing humanities practices

*The work of the humanities is frequently **descriptive**, or **appreciative**, or **imaginative**, or **provocative**, or **speculative**, more than it is critical. It includes **ways of attending to objects of study that are, variously, technical, aesthetically evaluative, curatorial**. Its public purposes can include **maintaining and reanimating knowledge of the cultural heritage**, explication of the products and process of culture, the **stimulation of public curiosity in new subjects**——again, not primarily critical activities. In the main the humanities are as concentrated on **the character of an individual response** as on any generalizable knowledge claims, and they typically have an interest in the colour and temper of that response that goes well beyond conveying knowledge or pursuing a critical interpretation of it. Not least, the humanities are rightly most **admired when their practice rises to an art**: a writing, or speaking or creative performance, that bears the stylistic and temperamental imprint of the individual voice. Their work is not always conducted from a position of high seriousness; some of its best expressions will seem **trivial, silly, even frankly irresponsible** judged from the position of the philosophically or politically committed critic.*

Small, *Value of the Humanities,* p. 26

the modern humanistic method

The modern humanistic method is an empirical method of acquiring knowledge that has characterized the development of the humanities since at least the tripartite division of the university into the humanities, social sciences, and sciences, but which draws on methods that go back thousands of years in other forms of humanist practice. It involves careful observation applying rigorous skepticism about what is observed, given that cognitive assumptions can distort how one interprets the observation. It tends to focus on small, complex

units of evidence, including historical events, cultural artifacts, and social pro-
cesses; and to formulate, on the basis of close interpretation and observation
of that evidence, both the individual qualities of that evidence and their poten-
tial generalizability into larger principles, which might be tested against other
complex units. This inductive-deductive looping produces further refinement
(or elimination) of generalizations. These are principles of the modern human-
istic method, as distinguished from a definitive series of steps applicable to all
humanistic enterprises. (Hayot 182)

arguments for the value of the humanities

1. that the Humanities study the meaning-making practices of culture, and have a distinctive understanding of what constitutes knowledge and understanding;
2. that, they are useful to society, but often at odds with or at a remove from the descriptions of 'usefulness' preferred by economists;
3. that they contribute to human happiness;
4. that they are a force for democracy; and
5. that they are a good in themselves, to be valued 'for their own sake'. (Small 2013)

2. What counts as

Humanities & Social Science here? 🎓

3. Finding & Understanding Data 💾

What is a CSV?
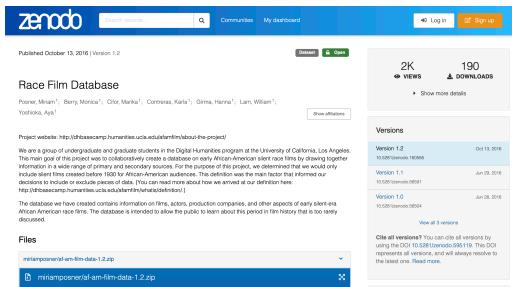
```
wikidata_ID, title, creator, released
Q29649659,Too Like the Lightning,Ada Palmer,2016
```

	wikidata_ID	title	creator	released
1				
2	Q29649659	Too Like the Lightning	Ada Palmer	2016

unstructured data

semi-structured data

structured data (CSV / JSON)

Race Film Database (UCLA) & data paper

Fields in Posner et al.'s dataset:

Title

Genre

Primary Source Verification

Production Company

Production date

Distribution date

Length (reels)

Actors

Directors

Producers

Camerawork
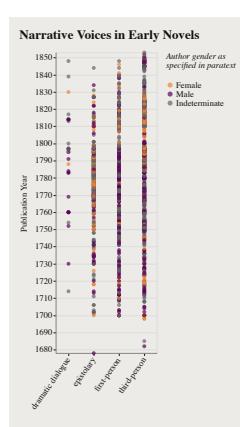
Source

Writer

Otherwise Involved
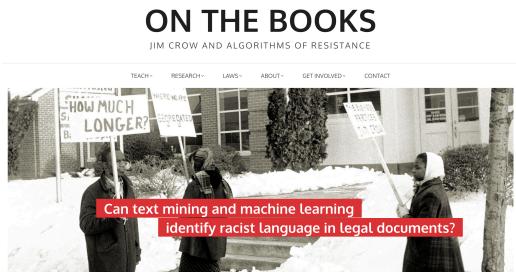
Link to media

Distributor

Early Novels Dataset

On the Books (UNC)

Can text mining and machine learning identify racist language in legal documents?

"Tidy Data"

Hadley Wickham, 2014
– one variable per column,

– one observation per row

– consistent data types

– unique identifiers for each object

– clear names for your fields

– *if you find you're trying to fit too much information into a column or row, use another table and cross-reference*

FAIR Data

articulated 2016

Findable

Aaccessible

Interoperable

Reusable

Linked open data

Data that is "structured using recognized standards so that computers interrogating the data can process it consistently," that refers to people / places / things using the same identifiers, and that is published openly.

Controlled vocabularies are used to:

☞ Internally link or identify records within datasets

☞ Link to institutionalized standards that other scholars / professionals might use ("Linked Open Data")

The Getty Vocabularies

Oxford Dictionary of National Biography

VIAF: The Virtual International Authority File

MIMO Thesaurus of Musical Instrument Names

The Programming Historian's intro to linked data

sources for humanities data 💾

Matt Lavin's Humanities Data

Melanie Walsh's list of datasets

Miriam Posner's list of datasets

Jeremy Singer-Vine's *Data Is Plural* archive

The Pudding's data repository

CDH-curated list of humanities datasets

do this:
  – Find a CSV, download it, open it in a spreadsheet program (Excel, Numbers, Google Sheets).

answer this:
  – What does your dataset contain?
  – How does it uniquely identify each record?
  – How does it format people / places / dates?
  – What categories and values are over-represented?
  – Do you see any initial patterns or trends?