

# Cultural Analytics

ENGL 64.05

Fall 2019

Prof. James E. Dobson



# Raw Data?

- All data requires some processing, some conversion to make it usable and comparable.
- Systems and standards *standardize* data. We might be distrustful of such efforts and consider them cooked and fetishize access to raw data.
- Yet the “raw” vs “cooked” binary reinforces the idea that data exist out there in the world, without collection, curation, and even imagination.



# Interpretation and Imagination

- We should question the separation between method and interpretation.
- The assumption of this split haunts all data-driven research.
- Humanities scholars question our objects all the time: where did this thing come from? Is it representative? How was it collected?
- Think about Special Collections: We have many materials from Dartmouth students in the past. Who decided to donate their materials? Who decided what to keep and what to toss in the trash? What was indexed?
- We cannot trust the archive to be representative of student life in the past, only an archive of what was thought worth archiving by those who gave, collected, and indexed the material.

# Gitelman and Jackson

They argue that data are

- *abstract* but ironically require material expression (6)
- *aggregative* and are collected in assortments of individual, homologous data *entries* and are accumulated into larger or small data *sets* (8).
- *mobilized graphically*, by which they mean that in order to be used as part of an explanation or argument data need to be represented and presented (12).

# Counting and Countlessness

- Laura Mandell explains that the project of recognizing women's contribution to literature is never done. She calls this the “cycle of forgetting women writers” (515).
- Literary history valorizes the book—both as the object of analysis and the product of scholarship.
- Books, especially edited editions, are monuments and decontextualized; the recovery of women writers requires contextualization.
- This has resulted in the de-valuation of women writers, who did not have equal access to publishing in earlier historical moments.

# Significance for the Analytics

- We need to ask what it is we are examining and what are our claims. Is some analytical project making claims about “the past” or “writing” or “literature” as such?
- If there is a lack of presence of women in the (digital archive), what about others who were explicitly denied access to literacy? How do we include their productions, their stories?
- Can statistical methods alone adjust for imbalance in the archive, in our data?

“When marshalled in huge numbers, women writers are not countless enough: in the absolute biggest datasets, the number of women is dwarfed in comparison to every man who ever wrote and becomes a small if not insignificant subset of the data stream” (Mandell 520).

DARTMOUTH Python-Lab-01 (autosaved)



Logout

Control Panel

File Edit View Insert Cell Kernel Help

Trusted

Python 3



DARTMOUTH Python-Lab-01 (autosaved)

File Edit View Insert Cell Kernel Help



```
In [ ]: # This is Jupyter code cell. This line is a comment; comments are not executed by the interpreter.

# Here we are assigning a variable 'var' the value of 'Cultural Analytics'
# this will automatically make var a string
var = 'Cultural Analytics'
```

## **Accessing Jupyter Hub:**

<https://jhub.dartmouth.edu>

## **Refreshing Data (Prior to each Lab):**

[http://dartgo.org/jupyterhub\\_engl64\\_o5](http://dartgo.org/jupyterhub_engl64_o5)



# Next Class

- Franco Moretti, “Style, Inc.”
  - Examining metadata for literary history.
- Jupyter Lab: Lists