Cultural Analytics

ENGL 64.05

Fall 2019

Prof. James E. Dobson

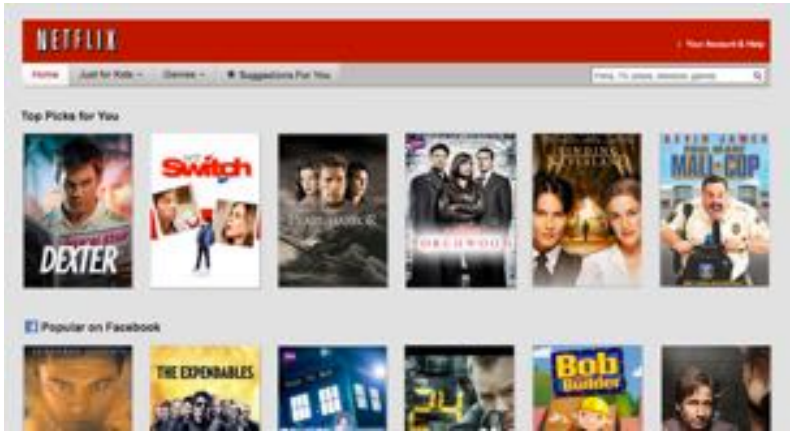
Computers + Culture

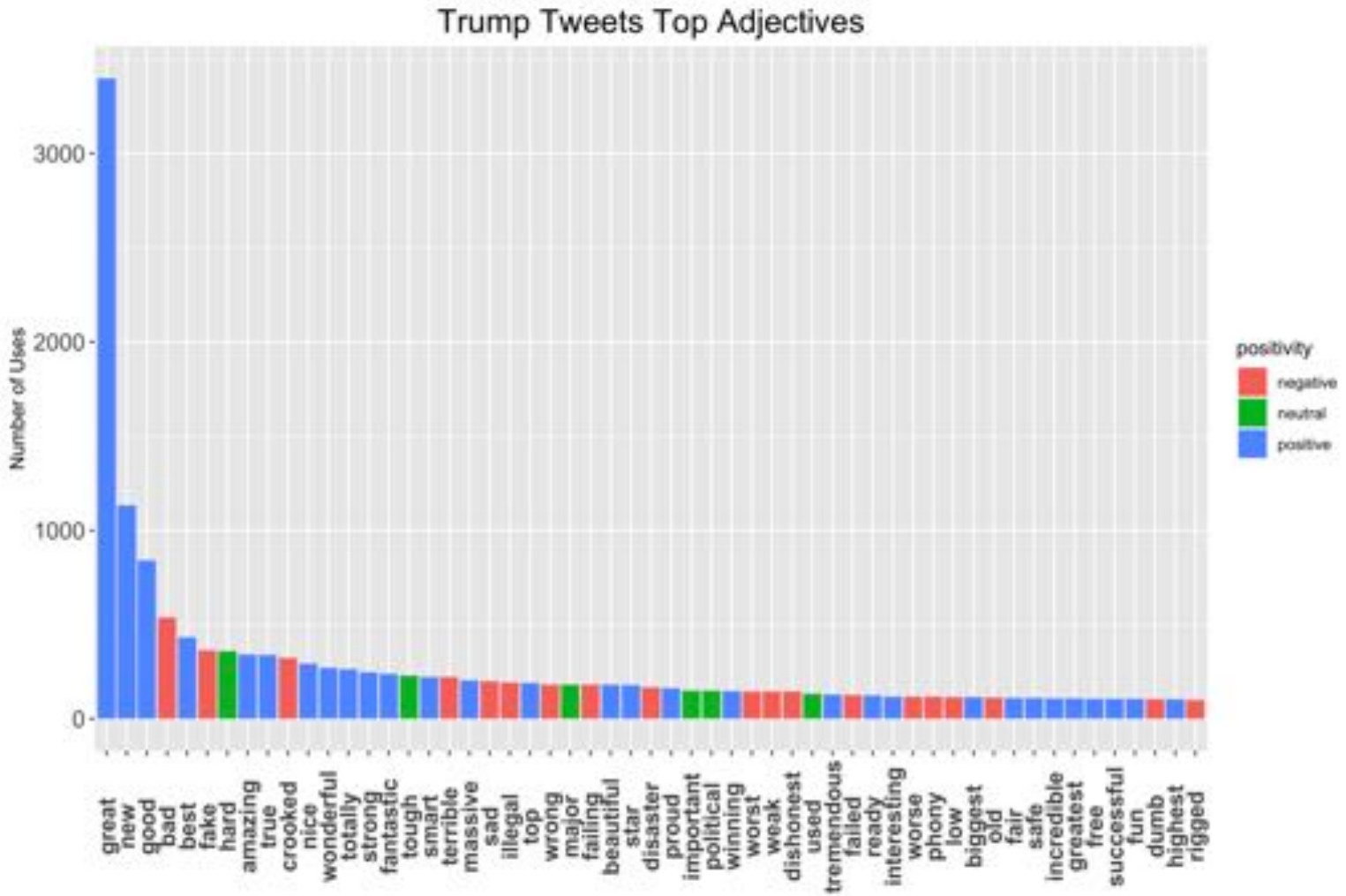
- QDS Distrib? In the Humanities? In English? Why?
- Critique + Skills
  - Data Analysis
  - Classification
  - Reproduction
  - Theory
  - Bullshit detection

Cultural Analytics

Lev Manovich defines *Cultural Analytics* as "the analysis of massive cultural data sets and flows using computational and visualization techniques" ("The Science of Culture? Social Computing, Digital Humanities, and Cultural Analytics" 2005).

Recommendation Engines





"Analyzing Trump's Tweets" (Tauberg, 2018)

Comparing Metadata (Moretti, 2009)

Topics or Themes (Jockers, 2013)

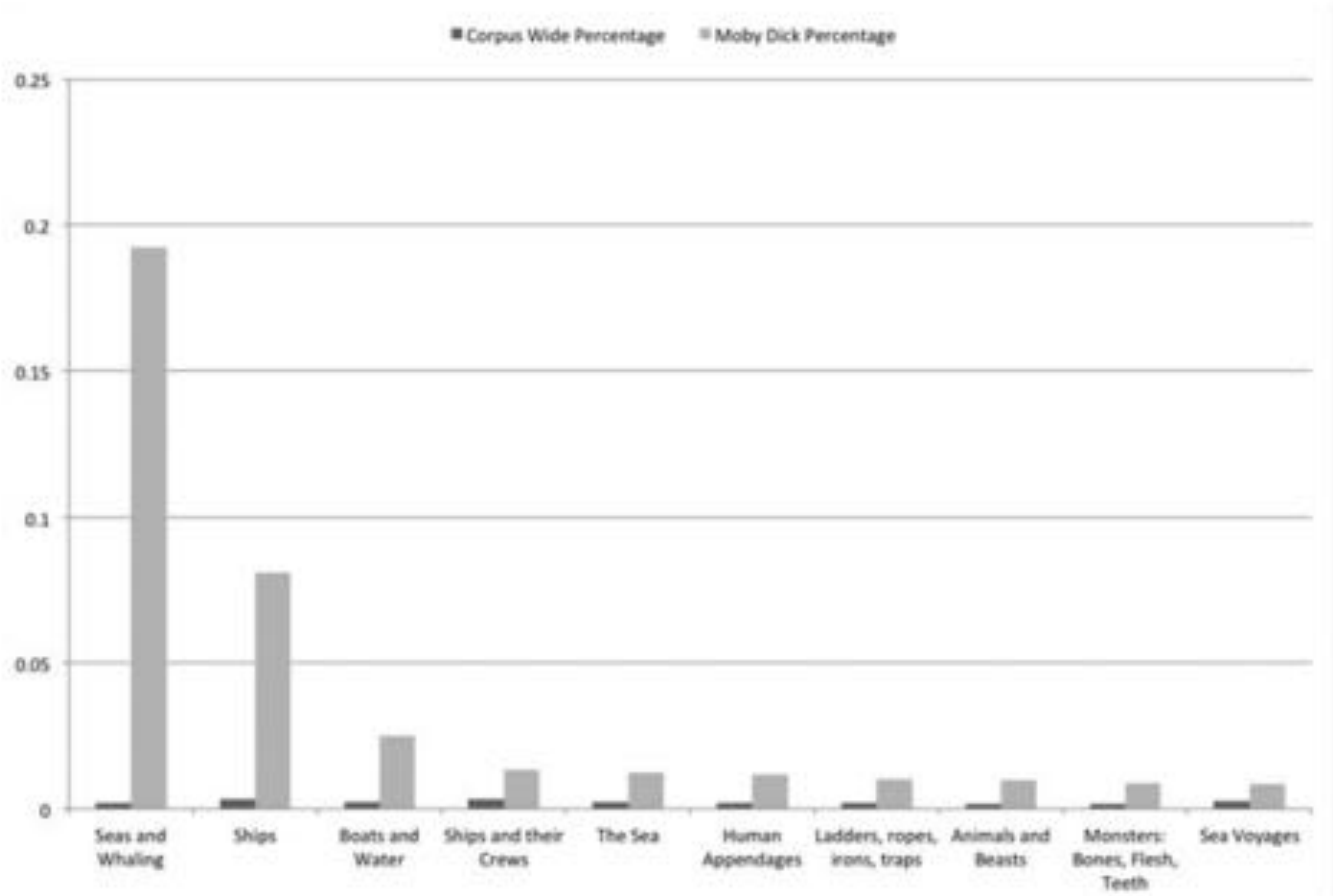
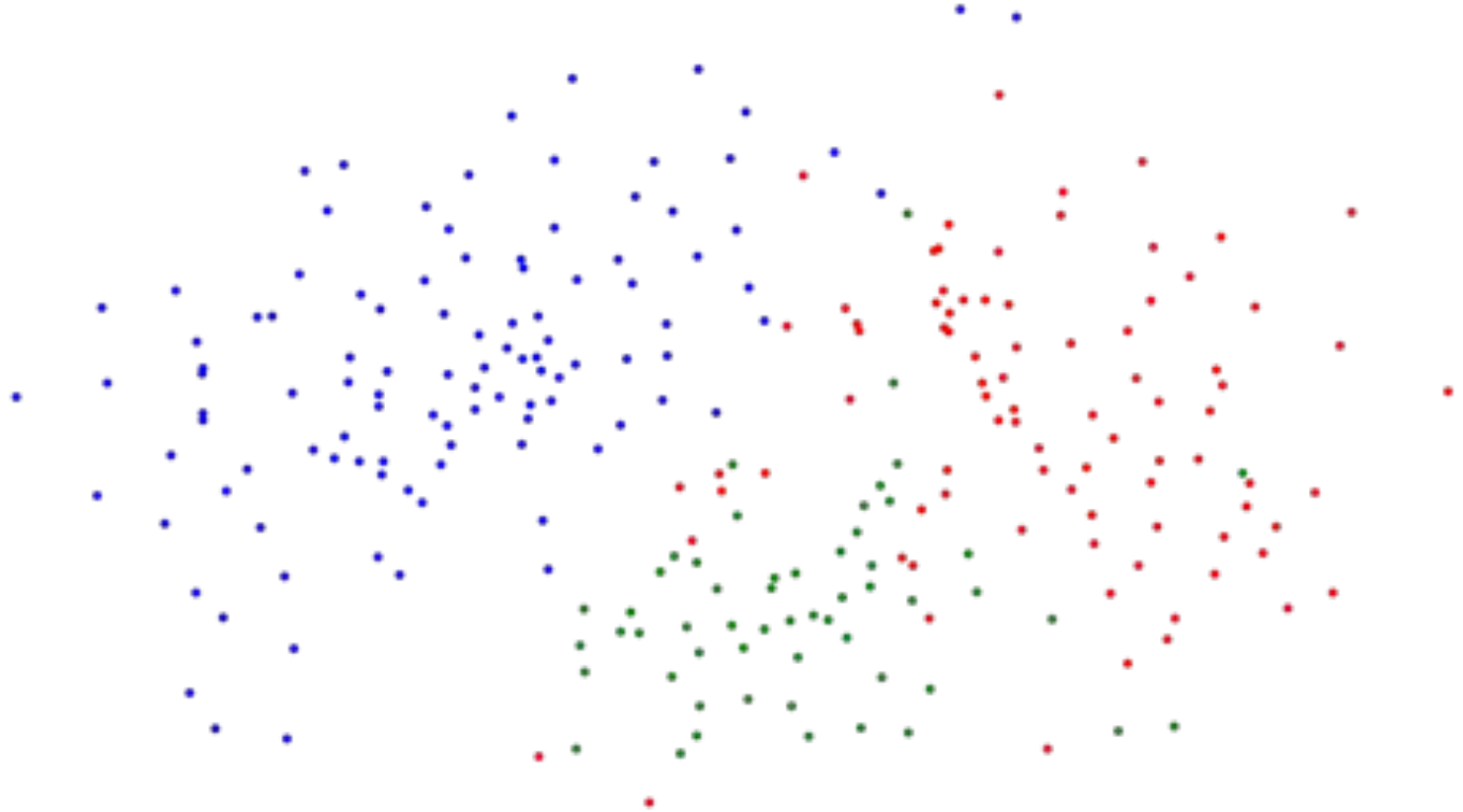


Figure 8.4. Top-ten topics in *Moby Dick*

Cultural Analytics



Mapping Genres

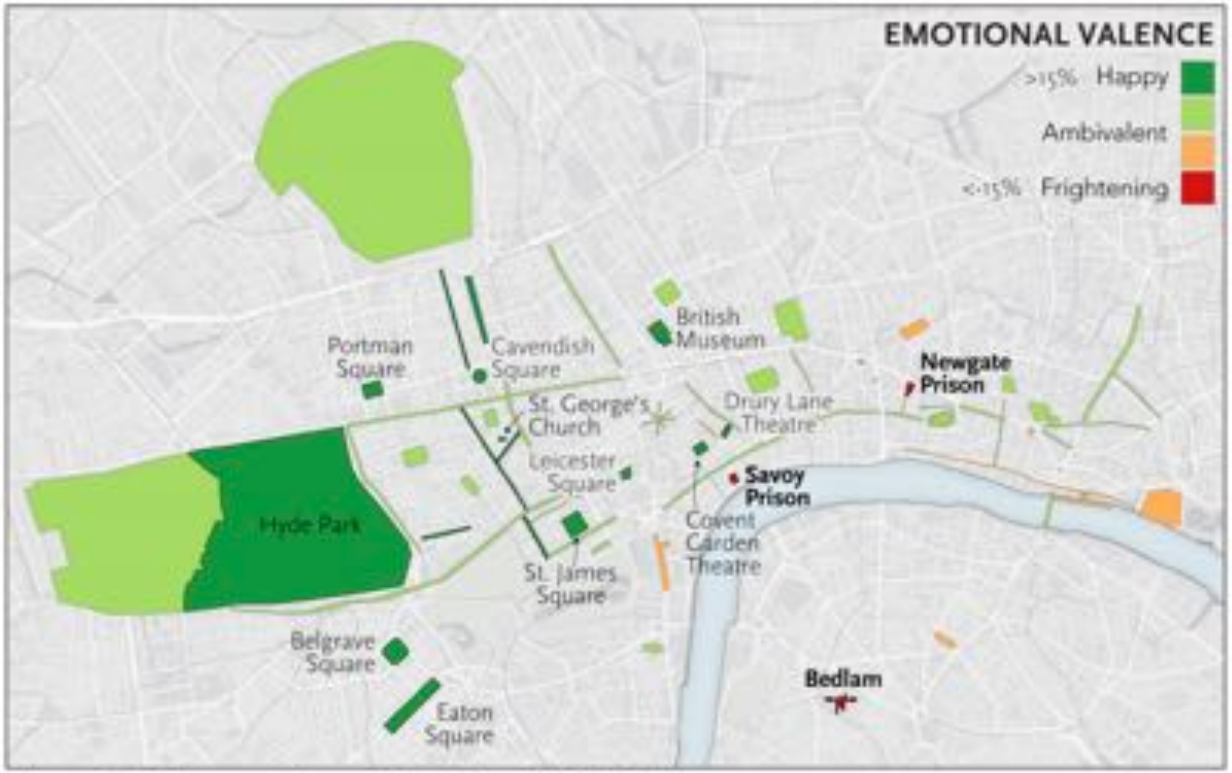


Figure 5.1 The emotions of London, 1700-1900

In this image, green is particularly prevalent in squares (the term that was also the most distinctive of the West End's lexicon), whereas passages where fear dominates are most often located in spaces of coercion and internment.

Emotions of London (Heuser et al 2016)

Cultural Analytics

"Who Cares About Literary Prizes?"
(Manshel et al 2019)

What are we doing when we do this?

- Computational Literary Studies
- Digital Humanities
- Quantitative Formalism
- Machine Reading
- Cultural Analytics
- ???

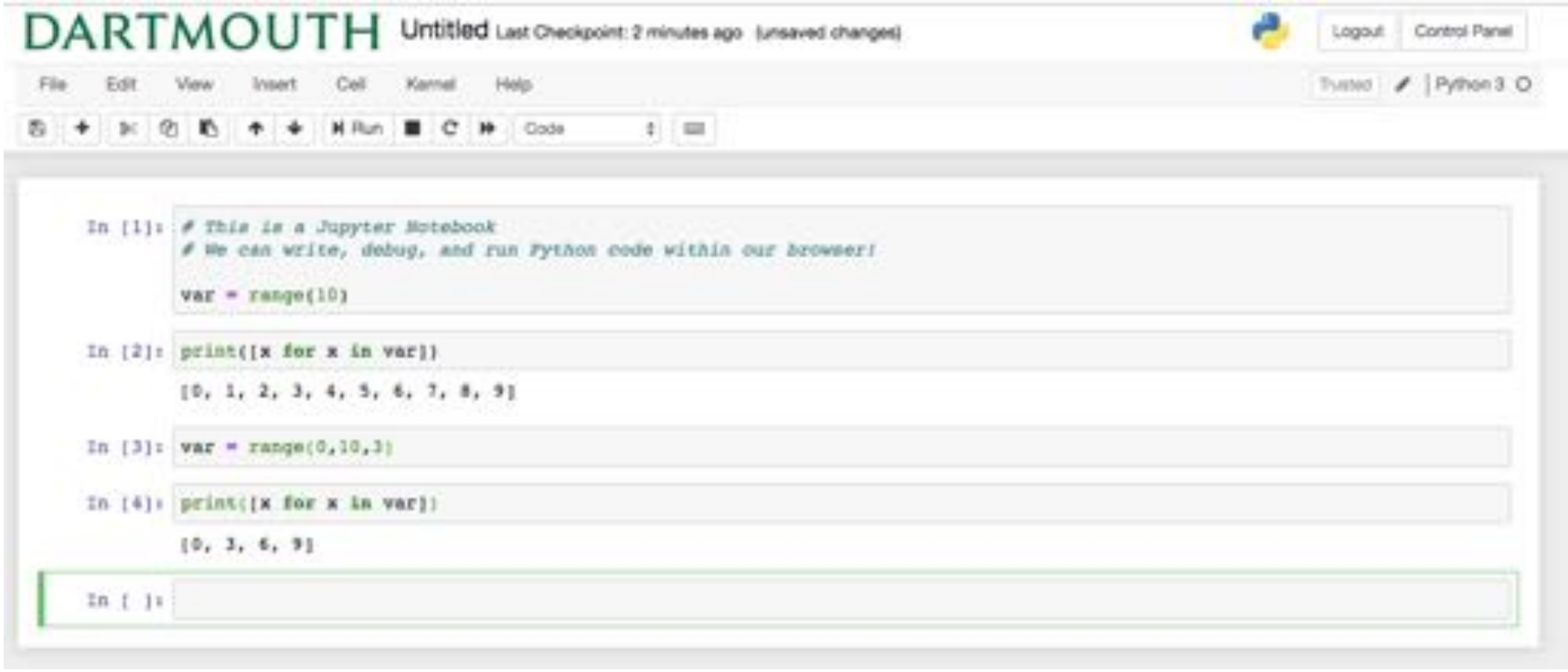
Information Science + Linguistics + Statistics + Literary Criticism + Theory

Course Overview

- Datafication and Critical Approaches to Data
- Experimental Results, Part I
- Measures of Similarity and Probability
- Experimental Results, Part II
- Designing Studies
- Topic Models and Other Methods
- Sentimental Analysis, Frequency Analysis, Data Smoothing
- Cultural Critique
- Presentation of Data and Critique

What We'll Learn

- Digital Theory
- Python Basics
- Text Mining and Natural Language Processing
  - Pattern matching
  - Extraction of features
  - Part-of-speech tagging
  - Cosine Distance
- Classification Basics (SVM, kNN)
- Visualization Basics (matplotlib)
- Multidimensional Scaling (MDS) and Principal Components Analysis (PCA)
- Sentiment Analysis
- Major recent critical work in computational literary studies


DARTMOUTH
Untitled Last Checkpoint: 2 minutes ago (unsaved changes)
Logout
Control Panel
File Edit View Insert Cell Kernel Help
Trusted
Python 3
Run
Code
In [1]:
# This is a Jupyter Notebook
# We can write, debug, and run Python code within our browser!
var = range(10)
In [2]:
print([x for x in var])
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
In [3]:
var = range(0,10,3)
In [4]:
print([x for x in var])
[0, 3, 6, 9]
In [ ]:

Jupyter

- Open Access textbook for learning Python basics.
- Maybe you already know Python, maybe it's all new.
- Success in course depends on sophisticated thinking, not programming.

"Homework"

- Short Jupyter Notebooks
- Two Types:
  - Execute and Submit
  - Revise and Complete and Submit
- Basic skills needed for the major projects
- Quickly discover limitations in methods

Two Projects

- Reproduce existing research
  - Evaluate the research of others
  - Critique methods and approach
- Design our own computational experiment
  - Explore new datasets and different methods
  - Is any of this stuff useful?

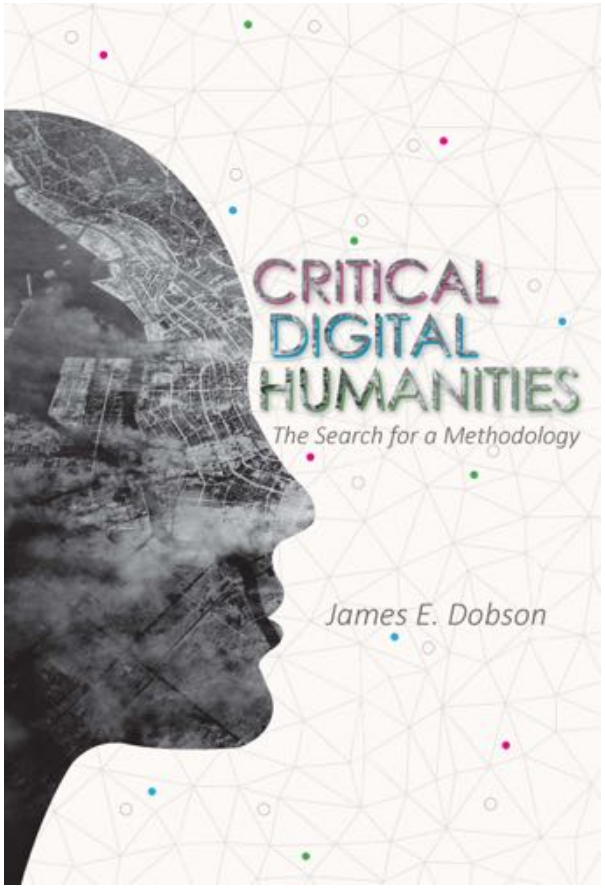
Replication vs. Reproduction

- **Replication** is an exact duplicate: re-running the prior experiment with same data and exact same methods. An an ideal but also limited.

- **Reproduction** means trying the experiment again. Perhaps with slightly different data or methods.

Cultural Analytics

Our Major Text



DARTMOUTH
VOX CLAMANTIS IN DESERTO
1769