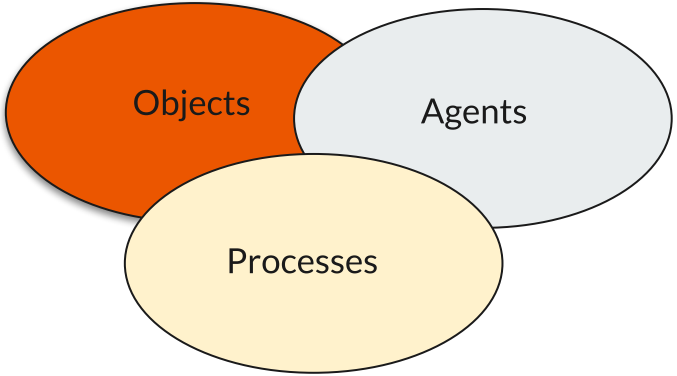
Foundations of Social and Cultural Data Analysis

Dr. Nanne van Noord & Dr. Melvin Wevers

Round of introductions

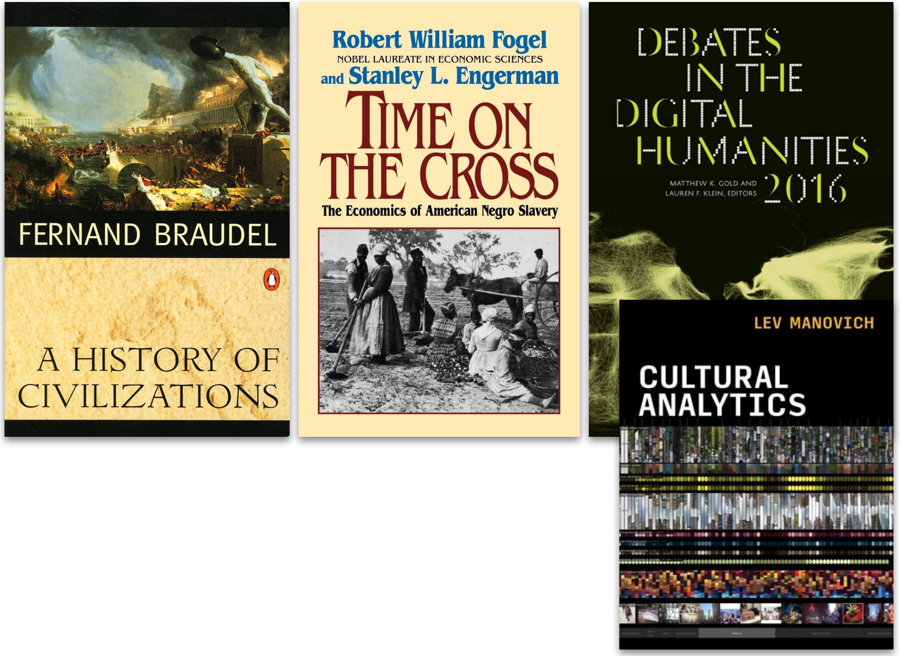
What is cultural data analysis?

- Using techniques and methods in data analysis to study cultural and social phenomena
- Getting information from data:
  - identify and explain patterns, trends and connections
  - make predictions and classifications
  - test hypotheses



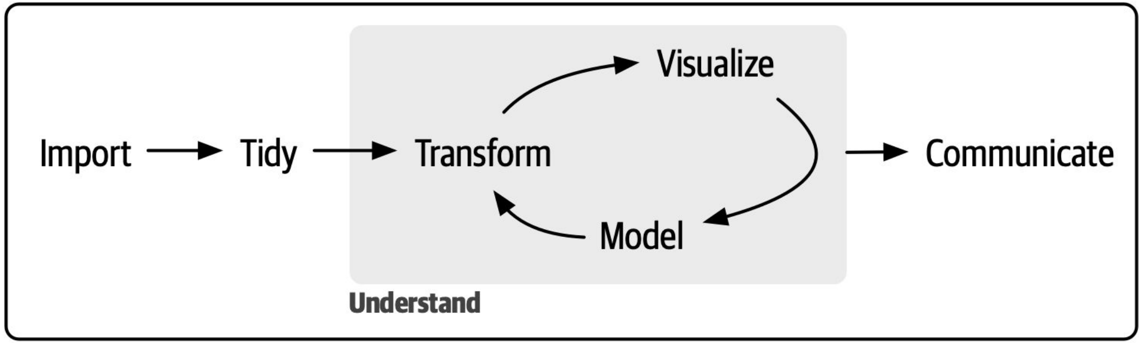
Tradition of Quantitative methods in the Humanities

- Goes back to anthropology, literature studies, and (social) history (50-70s)
- Backlash in the 1980s with the Cultural Turn
- Digital Humanities (early 200s) again put an emphasis on working with data, but not exclusively using quantitative methods. Greater role for hermeneutics and interpretation.
- Our focus here is on the use of computational methods / data science (computational humanities / cultural analytics) to better understand phenomena relevant to the humanities using cultural data


FERNAND BRAUDEL
A HISTORY OF CIVILIZATIONS
Robert William Fogel
NOBEL LAUREATE IN ECONOMIC SCIENCES
and Stanley L. Engerman
TIME ON THE CROSS
The Economics of American Negro Slavery
DEBATES IN THE DIGITAL HUMANITIES
MATTHEW K. GOLD AND LAUREN F. KLEIN, EDITORS
2016
LEV MANOVICH
CULTURAL ANALYTICS

Data Science Cycle

1. **Problem Formulation**
2. **Data Collection**
3. **Data Preparation**
4. **Exploratory Data Analysis (EDA)**: Explore data, visualize patterns, and identify trends.
5. **Modeling:** selecting/building machine learning/statistical models to solve a particular problem or achieve specific objectives
6. **Evaluation**: Assess model performance using appropriate metrics.
7. **Feedback and Iteration**



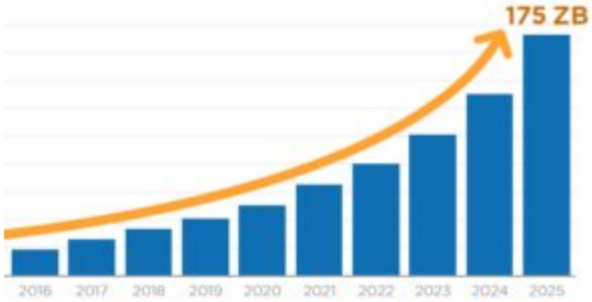
“The data may not contain the answer. The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.”

- John Tukey

Tukey, J. (1986). Sunset Salvo. The American Statistician.

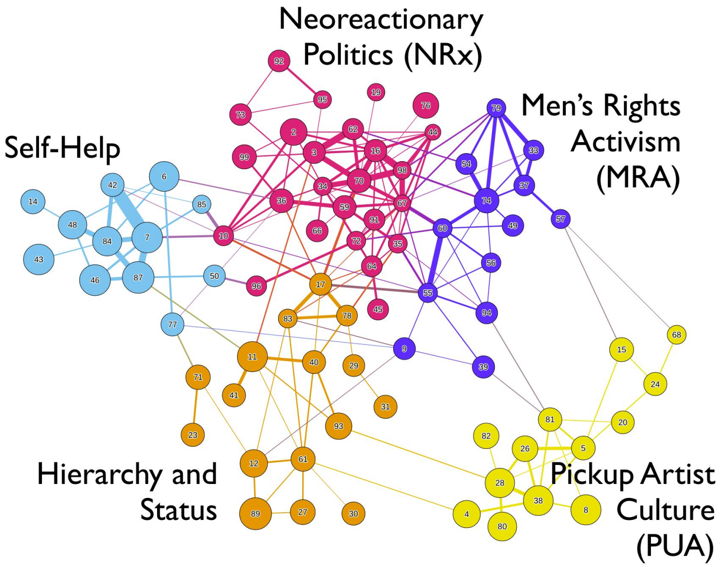
What is Cultural Data?

- (Social) media (Twitter/X, REddit, Newspapers, Magazines)
- User-generated content (web archives, fan fiction, memes, reviews)
- Digital cultural heritage (art collections, photo collections, books)

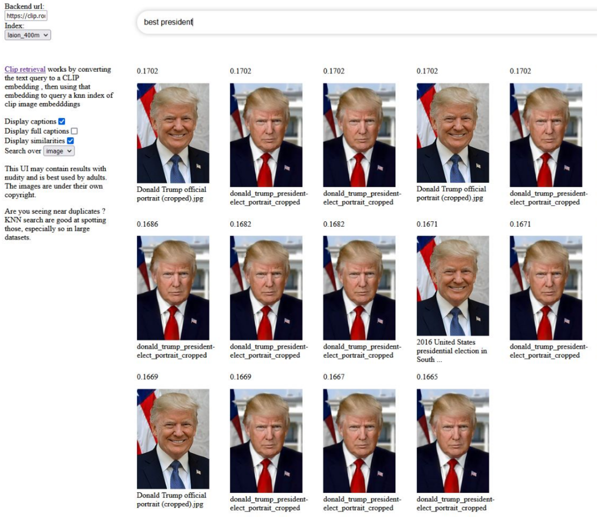


Modeling Culture

- Messy, Complex, Time-dependent
- Models are abstraction and focus on specific elements/features (reductionist)
- "All models are wrong, but some are useful." (George Box)
- Models are not just for prediction ("Why Model?" - Joshua Epstein)


Neoreactionary Politics (NRx)
Men's Rights Activism (MRA)
Self-Help
Hierarchy and Status
Pickup Artist Culture (PUA)

Data is Culture



Backend url: https://clip.roı

Index: laion_400m

best president

Clip retrieval works by converting the text query to a CLIP embedding , then using that embedding to query a knn index of clip image embedddings

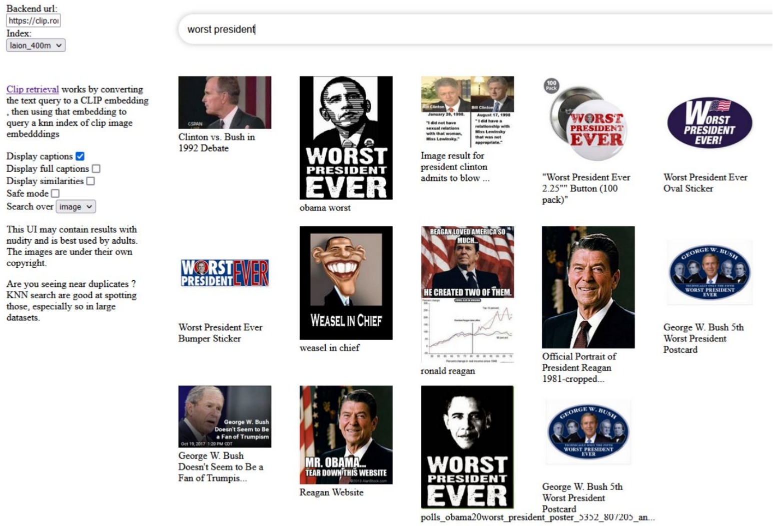
Display captions ☑
Display full captions □
Display similarities ☑
Search over image

This UI may contain results with nudity and is best used by adults. The images are under their own copyright.

Are you seeing near duplicates ? KNN search are good at spotting those, especially so in large datasets.

0.1702 Donald Trump official portrait (cropped).jpg
0.1702 donald_trump_president-elect_portrait_cropped
0.1702 donald_trump_president-elect_portrait_cropped
0.1702 Donald Trump official portrait (cropped).jpg
0.1702 donald_trump_president-elect_portrait_cropped
0.1686 donald_trump_president-elect_portrait_cropped
0.1682 donald_trump_president-elect_portrait_cropped
0.1682 donald_trump_president-elect_portrait_cropped
0.1671 2016 United States presidential election in South ...
0.1671 donald_trump_president-elect_portrait_cropped
0.1669 Donald Trump official portrait (cropped).jpg
0.1669 donald_trump_president-elect_portrait_cropped
0.1667 donald_trump_president-elect_portrait_cropped
0.1665 donald_trump_president-elect_portrait_cropped

(a) Best president



Backend url: https://clip.roı

Index: laion_400m

worst president

Clip retrieval works by converting the text query to a CLIP embedding , then using that embedding to query a knn index of clip image embedddings

Display captions ☑
Display full captions □
Display similarities □
Safe mode □
Search over image

This UI may contain results with nudity and is best used by adults. The images are under their own copyright.

Are you seeing near duplicates ? KNN search are good at spotting those, especially so in large datasets.

Clinton vs. Bush in 1992 Debate
obama worst
Image result for president clinton admits to blow ...
"Worst President Ever 2.25"" Button (100 pack)"
Worst President Ever Oval Sticker
Worst President Ever Bumper Sticker
weasel in chief
ronald reagan
Official Portrait of President Reagan 1981-cropped...
George W. Bush 5th Worst President Postcard
George W. Bush Doesn't Seem to Be a Fan of Trumpis...
Reagan Website
polls_obama20worst_president_poster_5352_807205_an...
George W. Bush 5th Worst President Postcard

WORST PRESIDENT EVER
WEASEL IN CHIEF
REAGAN LOVED AMERICA SO MUCH...
HE CREATED TWO OF THEM.
MR. OBAMA... TEAR DOWN THIS WEBSITE
George W. Bush Doesn't Seem to Be a Fan of Trumpism

Birhane et al. 2021

(b) Worst president

Cultural factors impact how we produce and interpret data. These (subjective) factors can vary over time and place.

Learning goals

After completing this course, the student is able to:

- **Code** in Python to perform a variety of practical tasks.

- **Formulate** a humanities research question that invites the use of data analysis.

- **Apply** data analysis tools and techniques on humanities data.

- **Relate** data analysis results to humanities research questions.

- **Explain** the surplus-value and limitations of data analysis from a humanities perspective.

- **Reflect** on the implications of the use of data analysis in studying historical and contemporary cultures.

Course logistics

Each Monday we meet for a **lecture** and a **laboratory.**

Lectures are on the topic at hand, they might contain some live coding too. **Laboratories** are mostly live coding with interaction and exercises for you.

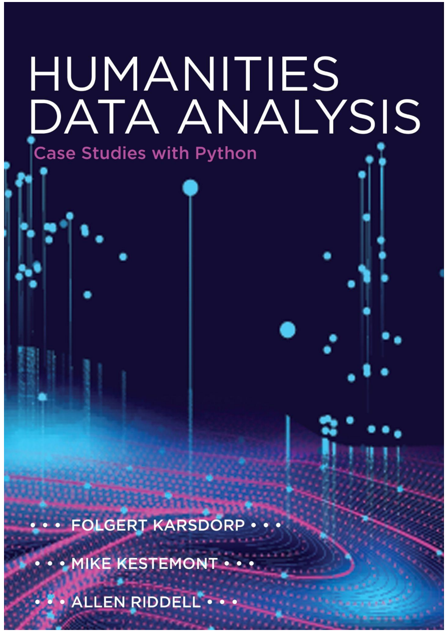
Every week, you will be handed out a **Coding Exercise** as a **group assignment** (check Canvas for the deadlines) (60%)

Towards the end of the course you will work on your **Project Proposal** (40%) also as a **group project**.

Execute project proposal during the next course ‘Applications of Cultural Data Analysis’

References and how to make the most of it

1. The main text is Karsdorp, F., Kestemont, M., & Riddell, A. (2021). Humanities Data Analysis: Case Studies with Python. Princeton University Press
2. Some weeks include additional reading. See the course manual for more information.
3. A reference for statistics: Canning, *Statistics for the Humanities*, 2014


HUMANITIES
DATA ANALYSIS
Case Studies with Python
• • • FOLGERT KARSDORP • • •
• • • MIKE KESTEMONT • • •
• • • ALLEN RIDDELL • • •

Python 101 Quiz

Link

Book

https://www.humanitiesdataanalysis.org/introduction-cook-books/notebook.html