

Seth Bernstein
Turlington 2328
T/Th: 8:30am-10:25am / 9:35am-10:25am
Course Site: <https://ufl.instructure.com/courses/485223>

sethbernstein@ufl.edu
Office Hours: KF 209
T/Th: 11am-1pm
(and by appointment)

EUH/HIS/WOH 3931
Digital Methods in History

Course Description:

Digital methods are becoming more important in history. It is true that history is about old things and cannot do without close reading. The increasing accessibility of large databases and technologies that can interpret them provide an opportunity to answer the questions that animate traditional histories in new ways.

1. Familiarize students with the basics of digital methods of analysis like database analysis, natural language processing, geographical information systems and (to a lesser extent) network analysis.
2. Learn methods of digital presentation of materials through visualizations and venues for online presentation (e.g., blogs or digital exhibits).
3. Analyze the benefits and problems of digital methods and the use of big data.
4. Practice working in small groups on a collective project.

The class is not connected to a specific region or time period. The main focus is on methodology. However, the major project for the course will be in the region of registration for the course. If you are registered for the WOH section, that means doing a final project on a topic outside of European or American history.

This course does not demand technical expertise as a prerequisite. Students will gain the necessary skills during the course.

Course Assignments:

Participation (100 points): Engagement in discussions that demonstrate a student's having read and understood assigned materials. Attendance is not a guarantor of a good participation grade but a good participation grade is not possible without attendance.

Weekly Laboratory (40 points x 10): During the first ten weeks of the class, we will have a weekly laboratory aimed at practicing or analyzing the uses of a technique or data set.

Group Proposal Presentation (70 points = 30 for presentation + 40 for written report): Students will work in small groups of 4-5 students with one data set or problem. In November, you will submit a collective written project proposal (700-1000 words) to be distributed to the entire class and participate in a feedback session with your classmates. The grade is collective.

Group Final Report and Presentation (100 points = 30 for presentation and 70 for written introduction): At the end of the term, the project groups will create a collective written project summary (1000-1500 words) that ties together the individual projects into a bigger idea. The report will be distributed to the class. The group will also present the project briefly as a whole before the class. The default platform for the project is Omeka, but the groups could consider other sites. The grade is collective.

Individual Project (200 points): Students will create a project using their group's data set. This project should analyze a historical problem using data and/or visualizations that are incorporated into a text of approximately 1500 words. For example, a project might be to create a time-lapse map and use the text to explore what the meaning of the map within a historical and historiographical context. The grade will depend, in order of importance, on skill at connecting data to historical problems and technical sophistication (computational, aesthetic), and writing.

Considered Reflection (100 points): In 1200 words, write about the use of computational methods in history. Were they necessary or helpful to tell the stories you wanted to tell? What could they do for you that close reading alone could not? Where did they fail? What areas would you consider exploring further and what seem like dead ends? Use concrete examples from your project and from readings.

Required Resources: The course will work from online resources that will be posted week-to-week. A text we will use often in the course is Shawn Graham, Ian Milligan, and Scott Weingart, *Exploring Big Historical Data: The Historian's Macroscope*. The first edition is available online here:

http://web.archive.org/web/20200618171130/http://www.themacroscope.org/?page_id=584. It is also available in second edition but the basic concepts and problems are the same.

A specific aspect of this class is that you will need access to a laptop computer for our in-class activities and to complete your class projects. Most of the programs are available online with a Google account. I recommend signing in to Google via your UFL account, possibly via an incognito tab to avoid a conflict with a non-UFL Google account. A non-exhaustive list of programs we will use include:

- Google Colabs (<https://colab.research.google.com/> Python platform. It might be worth downloading Python on its own but Colab is convenient because it means that everyone is working in the same programming environment.)
- Google Sheets (<https://sheets.google.com/> Spreadsheet)
- QGIS (<https://qgis.org/en/site/> Mapping)
- Voyant Tools (<https://voyant-tools.org/> Linguistic Analysis)
- Gephi (<https://gephi.org/> Network Analysis)

Department of Events and Propaganda: We have many events and opportunities at UF - ie, things that you can go to for enlightenment and profit. The Department of History is working to inform students about events and funding opportunities related to History overall and many are

connected to the kinds of digital methods we will encounter in this class. You can learn about these opportunities by following @UFHistory on [Instagram](#), [Twitter](#), or [Facebook](#).

Academic Integrity: The University of Florida holds its students to the highest standards, and we encourage students to read the University of Florida Student Honor Code and Student Conduct Code (Regulation 4.040), so they are aware of our standards. Any violation of the Student Honor Code will result in a referral the Student Conduct and Conflict Resolution and may result in academic sanctions and further student conduct action. The two greatest threats to the academic integrity of the University of Florida are cheating and plagiarism. Students should be aware of their faculty’s policy on collaboration, should understand how to properly cite sources, and should not give nor receive an improper academic advantage in any manner through any medium.

Lateness Policy: An important part of students’ work is meeting deadlines. Late assignments will be docked three percentage points per day overdue. Under extraordinary circumstances extensions will be granted but every effort should be made to avoid the need to take an extension.

Accommodations for Students with Disabilities: Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center by visiting <https://disability.ufl.edu/students/get-started/>. It is important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester.

Online Evaluations: Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at <http://gatorevals.aa.ufl.edu/students>. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via ufl.bluera.com/ufl/. Summaries of course evaluation results are available to students at <http://gatorevals.aa.ufl.edu/public-results/>.

Grading Policy:

You can find University of Florida’s grading policies here:

<http://catalog.ufl.edu/UGRD/academic-regulations/grades-gradingpolicies/>

Letter Grade	Percentage
A	93-100
A-	90-92
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76

C-	70-72
D+	67-69
D	63-66
D-	60-62

Attendance and Makeup Policy

Attendance is mandatory, and unexcused absences will adversely affect your grade. Requirements for class attendance and make-up assignments in this course are consistent with university policies that can be found in the online catalog at Attendance Policies <https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/> (Links to an external site.)>.

Course Schedule:

Week 0 (August 24): Introduction

Try this: *What Is Digital Humanities*: <https://whatisdigitalhumanities.com/>

Week 1 (August 29/31): What Is (Clean/Dirty) Data?

Readings:

Tuesday:

Reading:

Historian's Macroscope, chapter 1 (The Joys of Big Data for Historians, Big Data, Putting Big Data to Good Use: Historical Case Studies, Why this All Matters Now: The Third Wave of Computational History, Accessing the Third Wave Today, The Limits of Big Data, or Big Data and the Practice of the History)

Tutorial: Cleaning some data.

Thursday:

Reading: Rawson and Muñoz, "Against Cleaning," *Debates in the Digital Humanities 2019*.

<https://dhdebates.gc.cuny.edu/read/untitled-f2acf72c-a469-49d8-be35-67f9ac1e3a60/section/07154de9-4903-428e-9c61-7a92a6f22e51##ch23>.

Assignment Discussion

Assignment:

Step 1: Keep a journal (about life, a football team, classes, the news, exercise routines, a TV show, etc.) for three days. Think one or two paragraphs per day.

Step 2: Turn that journal into a spreadsheet with at least four fields entries in each.

Week 2 (September 5/7): Downloading Data Like a Computer

Tuesday:

Doings:

Using Google Colab: <https://colab.research.google.com/>

Python Programming Basics (starting here to lesson 15):

<https://programminghistorian.org/lessons/introduction-and-installation>

Historian's Macroscope, chapter 2 (The DH Moment, Intro to Several Key Digital History Terms, Delving into Big Data, Why We're All Digital Now, Building the Historian's Toolkit, Automatic Retrieval of Data, How to Become A Programming Historian, a Gentle Introduction, Basic Scraping: Getting Your Data, Normalizing and Tokenizing Your Data, Chapter Two Conclusion: Bringing It All Together: What's Ahead in the Great Unread)

Tutorial: Downloading from the *Old Bailey Proceedings*; Soviet Dissertations.

Thursday:

Reading:

Shoemaker, Robert B. "The Old Bailey Proceedings and the Representation of Crime and Criminal Justice in Eighteenth-Century London." *Journal of British Studies* 47, no. 3 (2008): 559–80. <http://www.jstor.org/stable/25482829>.

Assignment Discussion

Assignment:

Write a Python script that downloads and saves entries from the *Henry A. Wallace Police Crime Database* (<https://policecrime.bgsu.edu/>). Hint: The database is cataloged as a series that goes <https://policecrime.bgsu.edu/Cases/Details/XXXXXX> where XXXXX goes from 00001 to 12703. A less sophisticated version of the script will just download the pages. A more sophisticated version will also create a catalog from the pages.

Week 3 (September 12/14): Working with Big Data

Tuesday:

Readings:

Reading: James Scott, *Seeing Like a State*, 11-52, 87-102.

In-Class Tutorial: Soviet Dissertations as a Database

Thursday:

Heer, Nancy Whittier. "This Precious Legacy: Some Comments from the Perspective of Soviet Historiography." *Studies in Comparative Communism* 5, no. 2/3 (1972): 333–40.

<http://www.jstor.org/stable/45367575>.

Assignment Discussion

Assignment: Take the Soviet Dissertations database. Find how many dissertations were written in a single Soviet city by year. Turn in your code. Answer me this for extra credit: What kinds of dissertations were written in that city and how did they change over time?

Week 4 (September 19/21): Working with Words Pt. 1

Tuesday:

Reading: Moretti, "Style, Inc." in *Distant Reading*

Tutorial: Playing with the Enron database in Voyant Tools and Python.

Thursday:

Benke, Gavin. "From Green Fields to Narrating Nothingness: Neoliberal Logic and the Move Away from Environmental Responsibility in Enron's Rhetoric and Visual Style." *American Studies* 53, no. 2 (2014): 71–94. <http://www.jstor.org/stable/24589589>.

Assignment: Work with one person's emails. What are the most popular words they use? Are they using any words that deal with the environment? Turn in the code or explain the process by which you found it. Think about the kind of language you might find to think about the language of the environment (including euphemisms).

Week 5 (September 26/28): Working with Words Pt. 2

Tuesday:

Reading: Confino, Alon. "Collective Memory and Cultural History: Problems of Method." *The American Historical Review* 102, no. 5 (1997): 1386–1403. <https://doi.org/10.2307/2171069>.

Tutorial: *I Remember* Database, Entities

Thursday:

Bernstein, Seth. "Remembering War, Remaining Soviet: Digital Commemoration of World War II in Putin's Russia." *Memory Studies* 9, no. 4 (2016): 422-436.

Assignment: Find top 5 spatial entities in spaCy for one of the interviews (to be assigned individually). What questions does it raise about the document that you might investigate?

Week 6 (October 3/5): Mapping Pt. 1, Basics

Tuesday:

Reading: Jack Giesecking, Jen. “Where Are We? The Method of Mapping with GIS in Digital Humanities.” *American Quarterly* 70, no. 3 (2018): 641-648. [doi:10.1353/aq.2018.0047](https://doi.org/10.1353/aq.2018.0047).

Tutorial: Coordinate extraction/export with Python; basic GIS in ArcGIS Online.

Thursday:

Reading: Harvey, David. “The Sociological and Geographical Imaginations.” *International Journal of Politics, Culture, and Society* 18, no. 3/4 (2005): 211–55.

<http://www.jstor.org/stable/20059684>.

Assignment: Take the Enron database and write code that will find the spatial entities in a set of emails (your choice). Write code that will find the coordinates of those spatial entities. Export those places to a basic GIS in a CSV (comma separated values) file. Display the places in an ArcGIS Online map.

Week 7 (October 10/12): Mapping Pt. 2: Time Mapping

Tuesday:

Reading: Schvitz, Guy, Seraina Rügger, Luc Girardin, Lars-Erik Cederman, Nils Weidmann, and Kristian Skrede Gleditsch. "Mapping The International System, 1886-2017: The CShapes 2.0 Dataset." *Journal of Conflict Resolution* 66, no. 1 (2022): 144–61.
<https://journals.sagepub.com/doi/pdf/10.1177/00220027211013563>

Tutorial: QGIS and CShapes: Temporal Mapping (aka Time Maps); making and breaking borders.

Thursday:

Westerveld, Levi, and Anne Kelly Knowles. "Loosening the Grid: Topology as the Basis for a More Inclusive GIS." *International Journal of Geographic Information Science* 35, no. 10 (2021): 2108-2127.

Assignment: Design a GIS that could be used for temporal mapping with two or three lines of data (a spreadsheet with geographical data). Write a paragraph based on the Westerveld and Knowles reading: What would that time map capture? What would it miss?

Week 8 (October 17/19): Networks, Just the Basics

Tuesday:

Reading: *Historian's Macroscope*, "Networks in Historical Research"

Tutorial: Networking historians in Gephi.

Thursday:

Moretti, "Network Theory, Plot Analysis" <https://newleftreview.org/issues/ii68/articles/franco-moretti-network-theory-plot-analysis>

Assignment: Take the same network of historians from the AHA conference. Instead of connecting the historians, connect them to words used in the titles of their panels (i.e., with spaCy). Then use Gephi to generate a graph of that network.

Week 9 (October 24/26): Digitization and Digital Museums

Tuesday:

Readings:

Rozenzweig and Cohen, *Digital History* (“Designing for the History Web” and “Collecting the Past”)

Putnam, “The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast” *AHR*

Tutorial: Making a digital museum in Omeka with OCRed documents

Thursday:

Discussion of digital museums.

Assignment: Find a digital museum or archive online that uses a collection system like Omeka. Write a 250-500-word review covering the strengths and weaknesses of the format. If the institution has a physical location, what are the advantages and disadvantages of the digital collection? Does the digital collection complement the physical location or do something different? For collections without a physical location, would it make sense to have a physical location? For all collections, what does the online archive do (beyond accessibility) that a physical space cannot? Are there things you would do with it?

Week 10 (October 31/November 2): Visualization and Computer Vision

Tuesday:

Reading:

Milligan, Ian. "Learning to See the Past at Scale: Exploring Web Archives through Hundreds of Thousands of Images." In *Seeing the Past with Computers: Experiments with Augmented Reality and Computer Vision for History*, edited by Kevin Kee and Timothy Compeau, 116–36. University of Michigan Press, 2019. <http://www.jstor.org/stable/j.ctvnjbdr0.10>.

Tutorial:

A very basic overview of object recognition and/or some visualization tools.

Thursday:

Assignment discussion/looking at effective visualizations.

Assignment: Find a visualization with historical data that you admire. In 100-200 words, explain how it effectively tells a data story and link to the visual.

Week 11 (November 7/9): Project Proposals (In Person)

Each group will have approximately ten minutes to give an overview of their proposed project. The following issues should be addressed:

- What dataset will this project use?
- What historical questions can the use of this data answer?
- What will the work process look like? (How will you work together to wrangle the data? Are there parts of the project where you can divide the programming?)
- What kinds of techniques/technologies will you use?
- Are there problems that the group envisions?

We should have approximately ten minutes for feedback and questions.

Week 12 (November 14/16): Project Meetings (Scheduled at Keene-Flint Library)

I will meet for approximately 20 minutes with each group to check in on the progress of each group after the project proposal presentations. We will schedule meetings the week before. You only need to come to your group's meeting.

Week 13 (November 21): Troubleshooting (Optional, Remote and probably In Person)

There will be an optional class where groups/individuals can discuss issues they are encountering with their projects. I will probably be in person (although I will know closer to the date) but given that many students will be away, there will be a Zoom option as well.

Week 14 (November 28): Presentations (In Person)(November 30 – No Class)**

Week 15 (December 5): Presentations (In Person)

The groups will present their projects for the class. Each group should plan on having approximately 10 minutes to speak about their collective findings and on having 5-10 minutes for questions.