



01:090:101:18

Data Mining in the Humanities

Tuesdays, 2:00-3:20 p.m., Alexander Library, Room 413

Francesca Giannetti, Digital Humanities Librarian, Alexander Library

Office Hours: By appointment at <https://libcal.rutgers.edu/appointments/francesca>

Email: francesca.giannetti@rutgers.edu

Phone: 848-932-6097

Preferred contact: email or LibCal appointment

Course Description: Popular media often portray “big data” as the exclusive province of information scientists, but data collection in the humanities can swiftly exceed the capacity of the human brain to analyze. Increasingly, humanists turn to digital tools to conduct research on literary and historical texts, websites, tweets, images, video and sound recordings, among other formats. How does one create or reuse a humanities dataset? What tools are used to store, manipulate and process those data? How does one begin to analyze humanities data and share findings? This course will explore methodologies of both qualitative and quantitative analysis in the humanities, using digital tools to yield insights into data that might otherwise be difficult to obtain. Through lectures, discussion, labs, and a digital final project, students will familiarize themselves with several tools of digital humanities scholarship. In doing so, they will learn about the process of doing research digitally, from data collection, to analysis, to the presentation and sharing of results.

Course Website: We will use <https://rutgersdh.github.io/humdata22/> for the blogging assignments. The most up-to-date version of the course schedule, including reading assignments, is also kept here. Most of our readings are online. We have a Canvas site that I mainly use for emails to the class and to store readings that are not freely available online.

Prerequisite(s): None

Credit Hours: 1, pass/no credit

Course Objectives:

At the completion of this course, students will be able to

1. evaluate the use of humanities data through readings and by looking critically at several scholarly digital projects;
2. apply several methods for collecting, processing and presenting humanities data;
3. understand some of the technical, ethical, and social factors that influence which data gets produced and analyzed, and by whom.

Grade Distribution:

Discussion	30%
Blog posts	30%
Final project	40%

Course Policies:

- Attendance

- This course will involve active, engaged participation in class discussions and hands-on exercises. We won't have many lectures and there is no final exam. You should come to class having read the required reading and prepared to discuss it with your classmates. Maintaining an animated class discussion also requires that you be *present*, both physically and mentally. This means
 - * Arriving or logging in on time;
 - * Being attentive to the conversation;
 - * Preparing assigned readings before the start of class;
 - * Bringing assigned readings to class. If we are reading online articles, load them on your computer or device, or print them and bring that hard copy.
- Byrne seminars meet for only ten weeks. You may miss one class without penalty, but if you miss two, your discussion grade will be affected.
- It is your responsibility to find out what you missed and to make up any pertinent assignments. If you miss a lab due to an excused absence you should attempt to make up that work.

- COVID Caveat

- This seminar was originally developed to be taught in person during a very different time. I have tried to adjust the expectations to be a bit more forgiving of our strange, Omicron-laden circumstances. It is likely that I won't be able to foresee every twist and turn ahead of us, and will be reliant on your understanding as I teach this seminar in partially new ways. I hope to extend the same understanding to you for whatever range of situations you may find yourselves in this semester.

- Digital Etiquette

- Phones, tablets and other mobile devices should be muted and stowed away for the duration of class.
- The use of a laptop, on which you have administrative privileges (you are allowed to install software), is essential to this course. You will need it to participate in the digital labs and in-class exercises. We may move the class session to a computer lab, if necessary and feasible in COVID times.
- This course relies on access to computers, some fiddly software applications, and the Internet. At some point during the semester, just consider it a given that you will have a problem with technology. Just off the top of my head, your hard drive might fail, a server might go down, a piece of software will not behave the way you expect it to, or something else goes wrong. Try not to panic. While not panicking, consider learning some good habits to safeguard your work? Start assignments early and save often. Keep notes on what you did and where you left off. Always keep a backup copy of your work, stored some place secure, and ideally remotely accessible like in a Box or Google Drive folder. Hopefully we can have some good discussions as a class on digital workflows for success. Ultimately, though, it is your responsibility to ensure that your work is not irrecoverably lost. If your laptop crashes, **use a computer lab**. Most professors, myself included, are unlikely to grant extensions based on last-minute problems with technology.

- Grades
 - If you participate in class and turn in all assignments on time, you won't have difficulty passing this course. There will be one mid-semester grade check-in, allowing you to track your progress, at which time I'll report the assignments I have collected for you.
- Learning Support
 - Rutgers has **Learning Centers** on each campus where any student can get tutoring and other help. Rutgers also has **Writing Centers** where students can get assistance with writing skills and assignments.
- Students with Disabilities
 - To ensure that disability-related concerns are properly addressed from the beginning of the semester, I invite students with disabilities who require assistance to participate in this class to contact me as soon as possible to discuss their needs and any concerns they may have. In addition, the Rutgers Office of Disability Services may be able to provide additional resources. Phone: (848) 445-6800. Website: <https://ods.rutgers.edu>.

Academic Integrity

In addition to knowledge and skills, Rutgers University aims to teach students appropriate ethical and professional standards of conduct. The Academic Integrity Policy exists to inform students and faculty of their obligations in upholding the highest standards of professional and ethical integrity. All student work is subject to this policy. Professional and academic practice provides guidance about how to properly cite, reference, and attribute the intellectual property of others. A pertinent quote:

The principles of academic integrity require that a student:

- properly acknowledge and cite all use of the ideas, results, or words of others.
- properly acknowledge all contributors to a given piece of work.
- make sure that all work submitted as his or her own in a course or other academic activity is produced without the aid of impermissible materials or impermissible collaboration.
- obtain all data or results by ethical means and report them accurately without suppressing any results inconsistent with his or her interpretation or conclusions.
- treat all other students in an ethical manner, respecting their integrity and right to pursue their educational goals without interference. This requires that a student neither facilitate academic dishonesty by others nor obstruct their academic progress.
- uphold the canons of the ethical or professional code of the profession for which he or she is preparing.

The full text of the Rutgers Academic Integrity Policy may be found at <http://academicintegrity.rutgers.edu/>.

Course Schedule:

The weekly coverage may change depending on the progress of the class.

Acknowledgment:

In developing this seminar, I learned from several scholars and their courses, such as Miriam Posner's **Introduction to Digital Humanities**, Shawn Graham's **Introduction to Digital History**, Anastasia Salter's **Critical Making for Humanist Scholarship**, and Ted Underwood's **History and Theory of Digital Humanities**.

Week	Content
Week 1, Jan 18	<ul style="list-style-type: none"> • Topic: Introduction to the seminar • In-class exercise: Annotation
Week 2, Jan 25	<ul style="list-style-type: none"> • Topic: What is/are (digital) humanities? • Reading: Underwood, Ted. “Why an Age of Machine Learning Needs the Humanities.” <i>Public Books</i> (blog), December 5, 2018. https://www.publicbooks.org/why-an-age-of-machine-learning-needs-the-humanities/. • In-class exercise: Markdown and GitHub
Week 3, Feb 1	<ul style="list-style-type: none"> • Topic: What is/are data? • Reading: D’Ignazio, Catherine, and Lauren Klein. “Collect, Analyze, Imagine, Teach.” In <i>Data Feminism</i>. Cambridge, MA: MIT Press, 2020. • Assignment: blog post #1 due
Week 4, Feb 8	<ul style="list-style-type: none"> • Guest speaker: Caryn Radick, Digital Archivist, Special Collections and University Archives • Topic: Rutgers College War Service Bureau Records • Reading: Berry, Dorothy. “The House Archives Built.” <i>up//root</i> (blog), June 22, 2021. https://www.uproot.space/features/the-house-archives-built.
Week 5, Feb 15	<ul style="list-style-type: none"> • Topic: Markup • Reading: Flanders, Julia, Syd Bauman, and Sarah Connell. “Text Encoding.” In <i>Doing Digital Humanities: Practice, Training, Research</i>, 104–22. Abingdon, Oxon; New York, NY: Routledge, 2016. • Lab: Close Reading with the Text Encoding Initiative (TEI)
Week 6, Feb 22	<ul style="list-style-type: none"> • Topic: People and places • Reading: Horne, John. “Soldiers, Civilians and the Warfare of Attrition: Representations of Combat in France, 1914-1918.” In <i>Authority, Identity and the Social History of the Great War</i>, edited by Frans Coetzee and Marilyn Shevin-Coetzee, 325–55. Providence, RI: Berghahn Books, 1995. • Assignment: blog post #2 due
Week 7, Mar 1	<ul style="list-style-type: none"> • Topic: Maps • Reading: Turnbull, David. “Maps Are Territories: Science Is an Atlas: A Portfolio of Exhibits.” Chicago: University of Chicago Press, 1993. Read Exhibits 1-4. Optional: browse/read Exhibit 10. • Lab: mapping
Week 8, Mar 8	<ul style="list-style-type: none"> • Topic: Networks • Reading: D’Ignazio, Catherine, and Lauren Klein. “On Rational, Scientific, Objective Viewpoints from Mythical, Imaginary, Impossible Standpoints.” In <i>Data Feminism</i>. Cambridge, MA: MIT Press, 2020. • Assignment: blog post #3 due • Lab: network analysis
Week 9, Mar 22	<ul style="list-style-type: none"> • Topic: Edition • Reading: Earhart, Amy E. “An Editorial Turn: Reviving Print and Digital Editing of Black-Authoring Literary Texts.” In <i>The Digital Black Atlantic</i>, 30–48. Minneapolis: University of Minnesota Press, 2021.
Week 10, Mar 29	<ul style="list-style-type: none"> • Group projects; in-class discussion