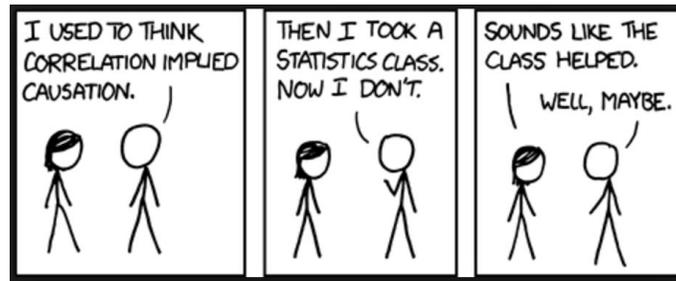


HIST/SOC/STS3604: “Data in Social Context: Big Data, Knowledge, and the Humanities” Online course syllabus concept (in progress)



(source with alt text: <https://xkcd.com/552/>)

About the Course

At present, digital technologies generate over 2.5 quintillion bytes of data each day. As data proliferate, they present novel opportunities for inquiry, creativity, and critique. “Data in Social Context: Big Data, Knowledge, and the Humanities” explores epistemological research and methodological approaches pertinent to Big Data and the humanities.

The first section of the course introduces theoretical perspectives on the relationship between data and knowledge. Here, we ask if Big Data merely extends our analytic and creative capacities, or, alternatively, if it is changing epistemological norms. In the second section, we review a suite of digital methods applicable to humanities scholars. Lectures and assigned materials impart technical knowledge and present scholarly reflections on specific tools. In the third section, we explore digital projects which address the human impact of Big Data. The course culminates with a final project: students have the option to submit a traditional paper or a digital humanities project proposal.

Good To Know

- ❖ All required materials are free and will be made available as PDFs or web links via this syllabus and Canvas.
- ❖ There’s a shortened version of the syllabus on Canvas. This lists only what you have to turn in to me and when it’s due, along with the date of the optional midterm quiz. It’s on Canvas under “Files” > “Syllabus” > “Syllabus_ShortVersion.”
- ❖ Everything is due by midnight in your time zone.
- ❖ Keep in mind that we may be in different time zones. I am on Eastern Standard Time.
- ❖ In addition to the required instructor meeting, I encourage one-on-one conversations. Please email me at stamm@vt.edu to set up an appointment via Zoom, Skype, or phone call.

How I give lectures

I make audio “podcasts” of my lectures and upload them to Canvas as mp3s. I post each week’s podcast on Monday before 5PM EST. The podcasts cover all of the assigned content for that week. They also provide reminders and course updates. I will upload podcast transcripts as Word doc files when I upload the mp3s.

Collaboration groups

By the end of week two, I will assign you to a group of three or four students. You will collaborate with group members on the assignments for grade components five and seven (see below). Groups will be determined by student time zone and major. You’ll be placed with classmates in or near your time zone, and each group will include students from a variety of majors. I will make introductions over email, and you will mutually decide on your methods and schedule for collaboration.

What you'll be doing as expressed in your final grade

Here's the breakdown by percentage. Please note that each percentage point corresponds to one point on your final grade on a scale of 1-100. For example, if an assignment is worth 5 points, that's 5% of your final grade.

Grade Component 1: The first assignment (pass/fail)

There are two parts to this assignment:

- 1) Fill out the class questionnaire, which I will send as a Google form on the first day. It should only take twenty minutes. Remember that the questionnaire is fully anonymous and that you may choose not to answer any question. This questionnaire will be used to generate the dataset for Grade Component 5.
- 2) Upload a document to Canvas under "Assignments" > "First Assignment." This document must include a brief autobiography (minimum one paragraph) and notification of whether you choose option A or B for Grade Component 5 (see below).

Note that the first assignment is pass/fail. If you do not turn it in on time, you will lose half a point on your final grade for every day late.

Grade component 2: Instructor meeting (pass/fail)

At some point between weeks 3 and 7, you will meet with me via video or traditional phone call. The point of this meeting is to discuss your final project, and we talk through any issues relating to the course.

Note that this is pass/fail, but if you do not participate in the instructor meeting, you will lose three points on your final grade.

Grade component 3: Discussion posts (5 points each; 20 points total)

Four times throughout the semester, you will make a three-hundred word discussion post on Canvas. The first discussion post asks you to introduce yourself to your fellow classmates. For the other discussion posts, you will respond to an open-ended prompt regarding recent content.

Grade component 4: Data mining exercises (2.5 points each; 5 points total)

At the end of week six, please complete any two of the five exercises on page 15 of *Discovering Knowledge in Data*. You will upload this as a document to Canvas under "Assignments" > "Data mining exercises."

Grade component 5: Group qualitative coding (20 points)

For this assignment, your collaboration group will come up with a qualitative coding scheme for the class dataset. You will use this schema to analyze the data. Submissions include multiple deliverables, including a written. More information will be provided early in the semester. One selected group member will upload your deliverables to Canvas under "Assignments" > "Group qualitative coding."

Grade component 6: Midterm exam or midterm paper (you choose) (25 points)

You may choose to A) take an exam or B) submit a short paper (minimum 1,200 words) at midterms. More information will be provided early in the semester.

Grade component 7: Final project (5 points for peer review; 25 points for project; total 30 points)

You may choose to A) write a traditional paper or B) submit a comprehensive proposal for a humanities research project that deploys digital methods. For both options, the submission will be no shorter than 2,000 words.

Two weeks before the final project is due, you will send a working draft to each of your collaboration group members. You will complete a peer-review sheet for each draft you receive and provide it to the author of the project one week later. The peer-review sheets for your projects — i.e., the sheets you receive from your peers, not those you fill out — are due to the instructor along with the final project submission.

Course schedule

Section I: Theory

Week 1

Reading assignments:

Lisa Gitelman and Virginia Jackson: "Introduction" from *Raw Data is an Oxymoron* (PDF on Canvas)
Alexander R. Galloway: ["From Data to Information"](#)

Week 2

Reading assignments:

Kate Crawford, Kate Miltner, and Mary L. Gray: ["Critiquing Big Data: Politics, Ethics, Epistemology"](#)
Rob Kitchin: ["Big Data, New Epistemologies and Paradigm Shifts"](#)

Week 3

Reading assignments:

Chris Anderson: ["The End of Theory: The Data Deluge Makes the Scientific Method Obsolete"](#)
Nathan Jurgenson: ["View from Nowhere"](#)
Geoffrey Bowker: ["The Theory/Data Thing"](#)

Video assignment:

Theorizing the Web 2018 Keynote (various participants): ["God View"](#)

Week 4

Reading assignments:

Alan Liu: ["Theses on the Epistemology of the Digital"](#)
Peter de Bolla: ["Digital Knowledge: Format, Scale, and the Information-knowledge Parallax at the Total Knowledge Horizon — A Reply to Alan Liu"](#)

Video assignment:

Timandra Harkness: [What is Knowledge in the Age of Big Data?](#)

Section II: Methodology

Week 5

Reading assignment:

Boris Lubarsky: [Re-Identification of Anonymized Data](#)

Video assignment:

Katharina Rasch: [What every data scientist should know about data anonymization](#)
Special guest on podcast: data consultant Nathaniel Porter on data privacy and anonymity practices

Week 6

Reading assignment:

[Qualitative Coding](#)

Video assignment:

[Beginner's Guide to Qualitative Coding](#)

Week 7

Reading assignment:

Daniel T. Larose and Chantal D. Larose: [Discovering Knowledge in Data: An Introduction to Data Mining](#) pp. 1-15

Week 8

Reading assignment:

Pedro Domingos: [A Few Useful Things to Know About Machine Learning](#)

Adrian Mackenzie: Introduction and Chapter 1, "Into the Data," from *Machine Learners: Archaeology of a Data Practice* (PDF on Canvas)

Special guest on podcast: machine learning specialist Stephen Paff on machine learning, ethnography, and demography amidst Covid-19

Week 9

Reading assignment:

Colin Allen and Jaimie Murdock: [LDA Topic Modeling: Contexts for the History and Philosophy of Science](#)

Section III: Project case studies

Week 10

Reading assignment:

Shaka McGlotten: "[Black Data](#)" (optional video behind link)

Oliver L. Haimson and Anna Lauren Hoffman: "[Constructing and Enforcing 'Authentic' Identity Online: Facebook, Real Names, and Non-Normative Identities](#)"

Project case studies:

Ben Grosser: [Instagram, Twitter, and Facebook Demetricators](#)

Week 11

Reading assignment:

Finn Brunton and Helen Nissenbaum: "[Vernacular Resistance to Data Collection and Analysis: a Political Theory of Obfuscation](#)"

Project case studies:

Daniel C. Howe, Mushon Zer-Aviv, and Helen Nissenbaum: [Ad Nauseam](#); Ben Grosser: [Go Rando](#); [ScareMail](#)

Week 12

Reading assignments:

Ingrid Burrington: "[Mapping the Amazon](#);" *ibid.*: "[Policing Is An Information Business](#)"

Project case studies:

Ingrid Burrington: [Networks of New York](#)

Week 13

Reading assignments:

John Cheney Lippold, "[Who we are as data might soon become more important than who we are as people](#);" Sun-ha Hong: "[Data's Intimacy: Machinic Sensibility and the Quantified Self](#)"

Project case study:

James Bridle: [Citizen Ex](#)

Week 14

Olivia Solon: "[The rise of 'pseudo-AI': how tech firms quietly use humans to do bots' work](#);" Lilly Irani: "[Justice For Data Janitors](#)"

Project case study:

Lilly Irani: [Turkopticon](#)

Week 15:

No assigned material. Ample time to work on your final projects!

Week 16:

Final project submission

