ILLINOIS
School of Information Sciences

Culture At Scale:
A Seminar on Computational Approaches to Studying Culture
IS597
Spring 2024
Mondays 1-4pm

This syllabus may be obtained in alternative formats upon request. Please contact the instructor.

Instructor: Zoe LeBlanc

Course Description

Cultural production is often thought of at the individual level - an artist, author, auteur - but with the turn of the century, new theories, methods, and fields have started to coalesce around the possibilities of producing *and* studying culture with computers, and specifically at larger and larger scales. Rather than reading a novel or watching a film, or even dozens, scholars are increasingly utilizing mass digitization and born digital materials to explore hundreds, if not thousands or millions of cultural objects; in essence undertaking what might be described as culture at scale. Whether labeled computational humanities or cultural analytics, or digital humanities or distant reading (to name just a few!), this new area of research is focused on understanding what Andrew Piper succinctly described as "computation plus culture." Often treated as largely a technical practice or as distinct fields that require interdisciplinary teams, this seminar is devoted to understanding how we might engage with these two practices constitutively and simultaneously – uncovering how computation can transform our understandings of culture and how in turn focusing on culture can impact how and when we use computation.

None of these terms (culture, computation, or scale) are self-evident. Culture for whom and by whom? Do we mean culture with a capital C or culture in the sociological sense? What is computation, and what is its relationship to statistics and programming? And even more oblique is scale. While Big Data has become ubiquitous, scale itself is not necessarily new, but the proliferation of relatively available digital storage and increasingly powerful processing power has profoundly accelerated and at times even transformed the horizons of possibilities. Rather than outright define these terms, this course will explore how scholars have determined and navigated these questions, as well as traversed and created new scales to make knowledge claims.

This course will therefore require both technical and theoretical expertise, as we engage with materials from multiple Humanities disciplines and sub-fields, use a variety of computational and statistical methodologies, and ultimately consider the very sociology of knowledge production and culture itself, which is increasingly happening at scale. Students will have the opportunity to develop

specific research projects, and also read broadly across disciplines and methodologies. Seminar will be devoted to discussing both assigned readings *and* exploring relevant code, methods, and datasets that were used to produce the readings. Students therefore should feel comfortable prior to enrolling the class with the foundations of research programming and working with data programmatically, as well as humanities-centric research (this does not mean humanities disciplines per se, but research concerned with culture). The goal in combining what has previously been called "yack and hack"[1] is that this seminar will explore how combining culture and computation at scale can "in an ideal world, [...] equal more than the sum of its parts."[2]

Pre- and Co-requisites

This course is inherently experimental and emergent, as many of the topics we will discuss are still being developed and refined by researchers – from statistical methods to novel model architectures to theoretical and intellectual frameworks.

Therefore, no one prerequisite is required but students should feel comfortable with programming and some area(s) of humanities research (again both broadly defined). Many of the course in the iSchool provide excellent foundation for this course, including:

- IS407 Introduction to Data Science
- IS417 Data Science in the Humanities
- IS430 Foundations of Information Processing
- IS509 History and Foundations of Information Science
- IS517 Methods of Data Science
- IS557 Applied Machine Learning Team Projects
- IS578 Introduction to Digital Humanities

However, this is not a definitive list. Interested students who have prior experiences that may be relevant, but not formal instruction, are also welcome to enroll.

Interested students should contact the instructor if they have any questions.

Credit hours

4

Student Learning Outcomes

Upon successful completion of the course, students will:

- Gain an overview of historic and current scholarly approaches to studying and producing culture at scale, as well as be exposed to how these approaches are being translated and applied in a variety of disciplines and domains.
- Develop a critical perspective on the possibilities and limitations of combining culture and computation to make knowledge claims, as well as the processes and norms that are developing around these practices.
- Have the opportunity to delve deeper into particular research questions and methods throughout the course that could become the basis for later published work.

---

[1] Bethany Nowviskie "On the Origin of "Hack" and "Yack"" *Debates in Digital Humanities* 2016

[2] Andrew Piper, "There Will Be Numbers," *Cultural Analytics* May 23, 2016. DOI: 10.22148/16.006 p. 2

- Ultimately, experience the razor's edge of research in culture at scale and be challenged to grow both intellectually and professionally.

Assignments and Methods of Assessment

Setting the Foundation: Flexibility, Commitments, and Intellectual Journeys

What does it mean to do "assignments" or be assessed as a graduate student? Ideally this question wouldn't matter, and you enrolled in this course with a deep passion for this research, so grading would be of minimal concern. In reality, we know that grades have enormous influence in both how we have been taught to value ourselves and in turn how society often values us. While I cannot wave a magic wand and restructure society, I do want to address this situation up front and honestly for a few reasons.

First, we are still in a global pandemic that has affected all of us, but not in equal ways. I address policies for what should happen if you or myself contract COVID-19 below under Attendance, but fundamentally no course is worth jeopardizing anyone's health over (whether physical or mental!). While having a hybrid option mitigates some of our risk, the fact is that teaching and learning is fundamentally different in a COVID world.

That being said, while this may all sound very depressing, I firmly believe we have an opportunity to build a more equitable and supportive learning environment even in these circumstances. But this requires some ground rules and (re)setting of our foundations when it comes to this course.

First, I hope that we can all approach this course as an **experiment in learning where flexibility and patience is prioritized.** Such a perspective does not mean that we don't have deadlines or expectations, but rather we acknowledge that we are all coming into this course with differing levels of expertise, responsibilities, and bandwidth (your instructor included!!). Rather than trying to erase these differences, we will strive to accommodate them as much as possible.

But to have this flexibility, we also want to make sure that we are committed to doing our best in this course. Otherwise, we lose this rare opportunity to think collectively and grow intellectually. So how to balance these goals: flexibility and commitment, in an increasingly difficult to predict world? Historically this is where grades and deadlines would come in. While I do have to submit grades for you, I want to emphasize that **this course is about your individual intellectual development and journey**. Therefore, I will try to make all expectations for grading as transparent as possible and also ask you to reflect on your intellectual journey during the course and remember an important adage throughout the course: **comparison is the thief of joy**.

The reason I want to emphasize this point is that rather than compare yourselves to each other, I hope that you can focus on your individual growth during the semester and see each other as colleagues to learn from, instead of competition. Some of you will already be in this mindset, but I find that this is one of the most important shifts from undergraduate to graduate education, so hopefully this reminder is helpful!

So, what does this all look like in practice?

The focus of the assignments in this course is to help you complete a draft of either a future dissertation chapter or article. While I do not expect you to have this completed in one semester, the purpose of this course is to help you either initiate a research project or build upon existing work, iteratively revise and improve upon this work, and ultimately, produce an extensive draft project by the end of the semester.

Students are welcome to work either individually or in groups and will develop their research project with input from the instructor and their peers. **The only requirement of this project is that students use computation, broadly defined, to study some aspect of culture, also broadly defined**. While there are few constraints on the research question per se, students will be encouraged to develop their project with an orientation towards the audiences for this research and to consider how this project can further their larger research agenda and scholarly goals.

Assignments

1. **Lead Seminar Discussion 25% (Dates Assigned First Week)**

Each student will be responsible for leading seminar discussion once over the course of the semester. In this capacity, they will be given the opportunity to develop a lesson plan and also delve deeper into the assigned materials with the goal of reproducing (as much as possible) the analysis of the assigned week (i.e. re-running the code and exploring relevant datasets). Students will be graded on their preparation for discussion, the questions they prepare for the course, and their ability to engage with the materials. Students are not responsible for how much their peers participate (though hopefully that will not be an issue). Students will consult with the instructor at least one week prior to leading discussion to confirm assigned readings and the topic focus for the week.

Students can co-lead discussion if there is more than one student interested in the topic, and the only extra requirement is that co-leading students will have to individually submit a short summary detailing how they divided the work. If for some reason a student(s) cannot lead the week they are scheduled, they must contact the instructor as soon as possible, and collectively we will work to reschedule.

Ideally, students should **select a weekly topic that furthers their engagement with their domain and selected methods for their research project**.

2. **Book & Code Review 25% Due March 11, 2024**
   (Extension Possible Until March 18, 2024)

Students will select one of the semester's readings (or suggest one to the instructor that fits the course themes) to review in two senses: first, as an exercise in writing a book review, and second, in performing a code review. Both of these activities are fundamental to culture at scale but often treated as distinct. We will experiment with what it looks like to combine these two activities; to consider both the interpretation and infrastructure simultaneously, and how that can inform both our critical perspectives and understanding of scholarship.

If there is no book that meets your interests, you are welcome to select a subset of articles (no more than four) to write a thematic review across them. You may also review more than one book in your

review if that would be useful to your research goals. Finally, you may also undertake a more methodological review, where you review either the use of a method in multiple articles or infrastructure for undertaking a method.

This review will be written individually, though students may review the same materials. The review should be 4-8 pages double spaced or no more than 2000 words (not counting references or footnotes). Students are welcome to use whatever style guide they prefer for citations and bibliographies.

The goal of this assignment is to help you engage with the secondary literature in your domain/methodology of interest. Ideally, you should select something that you will read for when you are leading seminar discussion. Students will be assessed on the quality of their writing, their ability to position their review materials in larger debates around culture at scale, and their assessment of both the strengths and weaknesses of their selected materials.

3. **Research Project 50% Due May 11, 2024**

Students working either individually or in groups will develop a research project with input from the instructor and their peers. Beginning from the first class, students will decide if they either plan to test and develop new methods for an existing domain and research question, or plan to work to apply methods they know already into a new domain of cultural production (other permutations exist as well). **The only requirement of this project is that students use computation, broadly defined, to study some aspect of culture, also broadly defined**. While there are few constraints on the research question per se, students will be encouraged to develop their project with an orientation towards the audiences for this research and to consider how this project can further their larger research agenda (whether as a dissertation chapter or a future article).

The final two weeks of the semester are currently allocated for student presentations of their research project, which will be between 20-30 minutes. Students will then submit their final paper, 15-30 pages double-spaced, along with their code and datasets. Students again can use any style guide they prefer, but please be consistent in your usage. If working in a group, students will also submit a short summary outlining the division of labor in their project. Students will be assessed on the quality of their writing, the formulation of their research questions, the implementation and suitability of their methods and data, and their ability to engage with relevant secondary research from both the course and beyond.

Students should plan to meet with the instructor at least once over the course of the semester (ideally in the first two weeks) to discuss their proposed project in-depth and to ensure that it fits the course remit.

<u>Some Caveats and Clarifications</u>

No participation grades?

I have intentionally not assigned any assessment to participation since this is a graduate seminar, which inherently expects you to engage with assigned materials through discussion. Furthermore, trying to assess what counts as good participation is always fraught. However, you may be wondering if this lack of participation grades means that you could theoretically not attend any

course meeting and still do well in the course. The answer is hypothetical yes, but there are few things to consider. Much of what we will discuss in this course cannot be gleaned from just reading the assigned materials. So, if you never attend, you will miss out on learning from your peers and Instructor, and furthermore, your submitted assignments will struggle to engage with these materials in a sufficiently rigorous manner. I have no interest in forcing anyone to attend a course, so the choice is ultimately yours.

This question of attendance is further discussed in the COVID-19 & Attendance section.

Using AI Tools?

You are welcome to use any AI tools that will help you in this course, whether that is tools like GitHub Co-Pilot or OpenAI's ChatGPT. I personally do not think these tools are going anywhere soon, and so learning to leverage them in your research is likely beneficial. However, I realize that many of these increasingly charge subscription fees, so please let me know if you would like to try a tool and are constrained for financial reasons, and I will try to advocate for some temporary funds from the iSchool.

Course Schedule

The schedule will be finalized after our first meeting, but it will involve combining three thematic threads:

- The experience of developing a research project around culture at scale: How do we develop computational narratives in the humanities? How do you develop a research question? How do you develop appropriate and relevant methods? How do you interpret these results and what data do you use? How do you integrate narrative with interpretation?
- Historical development of culture at scale: How has computation been deployed/leveraged in the humanities historically? How have various subfields theorized and worked with computation and data? How have these frameworks/theories shaped the types of knowledge claims that scholars have and can make? What are the limitations and criticisms of these approaches?
- Evaluating methodological and theoretical trends and practices for working with culture at scale: How have the developments across data and computation shifted the methods used in the humanities? Can we look across sub fields and disciplines to see meta level trends? How can we approach data and computational methods in ways that view them as holistically connected?

Notes on reading the schedule and assigned materials

This course is somewhat unique in that we are not focused on one domain or discipline, or one set of methods, which means we do not have a more traditional course schedule. I have initially selected materials around broad themes for each week, but these are mostly suggestions and will likely be altered depending on the interests of whomever is leading discussion and our collective interests. There are general framing questions that we will discuss each week that are listed below. These questions transcend the assigned materials and are intended to help us both work through larger meta-issues in working with culture at scale, and also consider how we will deal with these issues in our respective research projects. Many of these questions could be entire courses on their own, so

we will engage with them as much as possible, as well as our weekly case studies and theoretical readings.

When leading discussions, students are welcome to engage with these questions but are primarily responsible for discussions relating to the assigned materials. My tentative plan for seminar discussions is that we will spend most of the session discussing the assigned materials, but then reserve some time towards the end of the seminar for discussing these framing questions, as well as any methodological or project questions students may have. I want to emphasize that this is a *tentative* plan and will largely depend on student participation and input.

You will also notice that on our schedule we have different categories of assigned materials: core and background, as well as applied and theoretical. The first binary is my attempt to manage the scope of readings for this course, while at the same time pointing you to relevant materials for further research. You are only required to read **core materials** but are welcome to bring in any background material to discussions as well. The second binary is a bit hazier and is mostly to indicate to you the ways that you should expect to engage with the assigned readings. *Applied* readings will likely be examples of scholars using computation to make knowledge claims about culture, whereas *theoretical* readings are more likely to be arguments both for and against culture at scale. Some theoretical readings will include code and datasets, and some applied readings will also have theoretical and intellectual arguments as well.

Week to Week Schedule

1. Introductions and ~~First~~ Next Steps Jan 22, 2024

General Framing Questions:

What is culture at scale? How do we make knowledge claims about culture and the past? How have we in the past? Why does scale matter? What changed once we had scale? What are the origins of these practices, whether known as Cultural Analytics or something else?

Assigned Readings:

Core Theoretical Materials

- Piper, Andrew. "There Will Be Numbers." Journal of Cultural Analytics 1, no. 1 (May 23, 2016). https://doi.org/10.22148/16.006.
- "Introduction: In Pursuit of Theater's Digital Traces" Varela, Miguel Escobar. *Theater as Data: Computational Journeys into Theater Research.* University of Michigan Press, 2021. https://doi.org/10.3998/mpub.11667458.

Background Theoretical Materials

- Milic, Louis T. "The Next Step." Computers and the Humanities 1, no. 1 (1966): 3–6. (Available via the UIUC Library)
- Manovich, Lev. 2016. "The Science of Culture? Social Computing, Digital Humanities and Cultural Analytics." Journal of Cultural Analytics 1 (1). https://doi.org/10.22148/16.004

- Michael Piotrowski and Mateusz Fafinski (2020). "Nothing new under the sun? Computational humanities and the methodology of history". In: CHR2020: Proceedings of the Workshop on Computational Humanities Research (Amsterdam, Nov. 18–20, 2020). CEUR Workshop Proceedings, pp. 171–181. URL: http://ceur-ws.org/Vol-2723/short16.pdf

Core Applied Materials

- Sanderman, Erin, Deb Verhoeven, and Laura Mandell. "The 21°3'N of Separation of the Journal of Cultural Analytics: Mapping the First Five Years | Published by Journal of Cultural Analytics," 2021. https://culturalanalytics.org/post/1144-the-21-3-n-of-separation-of-the-journal-of-cultural-analytics-mapping-the-first-five-years
- Underwood, Ted. "Mapping the Latent Spaces of Culture." *Startwords,* no. 3 (August 2022). https://doi.org/10.5281/zenodo.6567481.

Background Applied Materials

- Robertson, Stephen, and Lincoln A. Mullen. "Arguing with Digital History: Patterns of Historical Interpretation," *Journal of Social History* 54, no. 4 (2021): 1005–1022, https://model-articles.rrchnm.org/articles/introduction/ (Make sure you use this link and not the version from the journal)

Assignments:

- Come prepared to both introduce themselves (your background, research interests, etc…) and also an area/question they hope to work with during the course. I realize this might be a bit disconcerting for the first week of class, but I do not expect you to have a fully polished research question. Rather, we just want to get started on the research project as soon as possible.

2. AI, Arguments, and Algorithms: Concepts Jan 29, 2024

General Framing Questions:

How has mass digitization and the digital age made culture at scale a possibility? How do we start to frame hypotheses and think with scale? Where do we find data and how do we create it? What sorts of arguments can we make about culture with data? How do we balance domain with method when developing a research question? How is AI impacting *all of this*?

Assigned Readings:

Core Theoretical Materials

- Chapter 1 "An Algorithmic Criticism" Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. University of Illinois Press, 2011. https://www.jstor.org/stable/10.5406/j.ctt1xcmrr
- Skim one of the following:

- Nguyen, Dong, Maria Liakata, Simon DeDeo, Jacob Eisenstein, David Mimno, Rebekah Tromble, and Jane Winters. "How We Do Things with Words: Analyzing Text as Social and Cultural Data." Frontiers in Artificial Intelligence 3 (August 25, 2020): 62. https://doi.org/10.3389/frai.2020.00062
- Ziems, Caleb, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. "Can Large Language Models Transform Computational Social Science?" arXiv, April 12, 2023. https://doi.org/10.48550/arXiv.2305.03514.

Background Theoretical Materials

- Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, 2013. https://www.jstor.org/stable/10.5406/j.ctt2jcc3m.

Core Applied Materials

- Kleymann, Rabea, Andreas Niekler, and Manuel Burghardt. "Conceptual Forays: A Corpus-Based Study of 'Theory' in Digital Humanities Journals." *Journal of Cultural Analytics* 7, no. 4 (December 19, 2022). https://doi.org/10.22148/001c.55507
  - Code for Conceptual Forays https://github.com/theory-in-dh/conceptual_forays
- Gavin, Michael, Colin Jennings, Lauren Kersey, and Brad Pasanek. "Spaces of Meaning: Conceptual History, Vector Semantics, and Close Reading." In *Debates in the Digital Humanities*, 2019. https://dhdebates.gc.cuny.edu/read/untitled-f2acf72c-a469-49d8-be35-67f9ac1e3a60/section/4ce82b33-120f-423f-ba4c-40620913b305#ch21

Background Applied Materials

- Spinaci, Gianmarco, Giovanni Colavizza, and Silvio Peroni. "A Map of Digital Humanities Research across Bibliographic Data Sources." *Digital Scholarship in the Humanities* 37, no. 4 (December 1, 2022): 1254–68. https://doi.org/10.1093/llc/fqac016.
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, THE GOOGLE BOOKS TEAM, Joseph P. Pickett, et al. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331, no. 6014 (January 14, 2011): 176–82. https://doi.org/10.1126/science.1199644.
- Underwood, Ted. "Do Humanists Get Their Ideas from Anything at All?" The Stone and the Shell (blog), January 24, 2012. https://tedunderwood.com/2012/01/24/discovery-and-hypothesis-testing/.
- Weingart, Scott. "Argument Clinic – Scottbot.Net," January 22, 2022. https://web.archive.org/web/20220122103133/http://scottbot.net/argument-clinic/

3. Close and Distant: Genre Feb 12, 2024

General Framing Questions:

How does scale complement and contrast non-scalable methods for working with cultural materials? How can we start to explore datasets but also operationalize hypotheses? How do we start to turn culture into machine-readable data and what is gained/lost in this process?

Assigned Readings:

Core Theoretical Materials

- Chapter 1 "Abstraction, Singularity, Textuality: The Equivalence of "Close" and "Distant" Reading" in Bode, Katherine. *A World of Fiction: Digital Collections and the Future of Literary History*. University of Michigan Press, 2018. https://www.jstor.org/stable/j.ctvdtpj1d.
- Gabi, Kirilloff. "Computation as Context: New Approaches to the Close/Distant Reading Debate." College Literature, vol. 49 no. 1, 2022, p. 1-25. https://doi.org/10.1353/lit.2022.0000.

Background Theoretical Materials

- Underwood, Ted. "A Genealogy of Distant Reading." Digital Humanities Quarterly 011, no. 2 (June 27, 2017). http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html
- Sinykin, Dan. "Distant Reading and Literary Knowledge." Post45 (blog), May 6, 2019. https://post45.org/2019/05/distant-reading-and-literary-knowledge/(and the entire cluster is fantastic).
- Primorac, Antonija, Rosario Arias, Roxana Patras, Eva Eglāja-Kristsone, Karina van Dalen-Oskam, Berenike Herrmann, Christof Schöch, and Pieter François. "Distant Reading Two Decades On: Reflections on the Digital Turn in the Study of Literature." Digital Studies / Le Champ Numérique 13, no. 1 (October 25, 2023). https://doi.org/10.16995/dscn.8855.

Core Applied Materials

- "Preface, Chapters 2 & 5." Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, 2019. (Available on Canvas, Chapter 1 optional)
    - Code for Chapter 2 https://github.com/tedunderwood/horizon/tree/master/chapter2

Background Applied Materials

- Underwood, Ted. "The Life Cycles of Genres." *Journal of Cultural Analytics* 2, no. 2 (May 23, 2016). https://doi.org/10.22148/16.005.
- Sharma, Aniruddha, Yuerong Hu, Peizhen Wu, Wenyi Shang, Shubhangi Singhal, and Ted Underwood. "The Rise and Fall of Genre Differentiation in English-Language Fiction: 1st Workshop on Computational Humanities Research, CHR 2020." CEUR Workshop Proceedings 2723 (2020): 97–114. https://ceur-ws.org/Vol-2723/long27.pdf
- Allison, Sarah, Matthew Jockers, Ryan Heuser, Franco Moretti, and Michael Witmore. "Quantitative Formalism: An Experiment." *Literary Lab*, 2011. https://litlab.stanford.edu/LiteraryLabPamphlet1.pdf.

4. Categories and Classifications: Gender Feb 19, 2024

General Framing Questions:

How can we categorize and classify culture? How can we use computation to understand cultural categories of the past and present? How can we replicate studies and what is the value of validating results? What are the limits of computation for studying culture?

Core Theoretical Materials

- Da, Nan Z. "The Computational Case against Computational Literary Studies." *Critical Inquiry* 45, no. 3 (March 2019): 601–39. https://doi.org/10.1086/702594.
- Schöch, Christof. "Repetitive Research: A Conceptual Space and Terminology of Replication, Reproduction, Revision, Reanalysis, Reinvestigation and Reuse in Digital Humanities." International Journal of Digital Humanities 5, no. 2 (November 1, 2023): 373–403. https://doi.org/10.1007/s42803-023-00073-y.

Background Theoretical Materials

- Mandell, Laura. "Gender and Cultural Analytics: Finding or Making Stereotypes?" In Debates in the Digital Humanities, 2019. https://dhdebates.gc.cuny.edu/read/untitled-f2acf72c-a469-49d8-be35-67f9ac1e3a60/section/5d9c1b63-7b60-42dd-8cda-bde837f638f4.
- "Computational Literary Studies: A Critical Inquiry Online Forum" March 2019 https://critinq.wordpress.com/2019/03/31/computational-literary-studies-a-critical-inquiry-online-forum/.

Core Applied Materials

- "Chapter 4 and Appendices" Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, 2019. (Available on Canvas)
    - Code for *Distant Horizons* https://github.com/tedunderwood/horizon/

Background Applied Materials

- Underwood, Ted, David Bamman, and Sabrina Lee. "The Transformation of Gender in English-Language Fiction." *Journal of Cultural Analytics* 3, no. 2 (February 13, 2018). https://doi.org/10.22148/16.019.
- Kraicer, Eve, and Andrew Piper. "Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction." *Journal of Cultural Analytics* 3, no. 2 (January 30, 2019). https://doi.org/10.22148/16.032.
- Lucy, Li, and David Bamman. "Gender and Representation Bias in GPT-3 Generated Stories." In Proceedings of the Third Workshop on Narrative Understanding, 48–55. Virtual: Association for Computational Linguistics, 2021. https://doi.org/10.18653/v1/2021.nuse-1.5.
- Blevins, Cameron, and Lincoln Mullen. "Jane, John … Leslie? A Historical Method for Algorithmic Gender Prediction." Digital Humanities Quarterly 009, no. 3 (December 23, 2015). http://www.digitalhumanities.org/dhq/vol/9/3/000223/000223.html.
- Cheng, Jonathan. "Fleshing Out Models of Gender in English-Language Novels (1850 – 2000)." *Journal of Cultural Analytics* 5, no. 1 (January 29, 2020). https://doi.org/10.22148/001c.11652.

- Cheng, Jonathan, 2020, "Replication Data for: Fleshing Out Models of Gender in English-Language Novels", https://doi.org/10.7910/DVN/QUGW8V, Harvard Dataverse, V1

5. Models and Measurements: Race Feb 26, 2024

General Framing Questions:

How can we further transform datasets to answer our questions? How can we turn our questions into models? What are models? How do we deal with minimizing information loss and maximizing algorithmic power? How do we use current methods but make them work for our purposes? How much do we care about statistics?

Core Theoretical Materials

- Levy-Eichel, Mordechai, and Daniel Scheinerman. "Digital Humanists Need to Learn How to Count." *Chronicle of Higher Education*, May 17, 2022. https://www.chronicle.com/article/digital-humanists-need-to-learn-how-to-count
- So, Richard Jean. "'All Models Are Wrong.'" *PMLA* 132, no. 3 (May 2017): 668–73. https://doi.org/10.1632/pmla.2017.132.3.668.

Background Theoretical Materials

- Breiman, Leo. "Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author)." *Statistical Science* 16, no. 3 (August 2001): 199–231. https://doi.org/10.1214/ss/1009213726.
- McGrath, Laura B. "Charisma (Embodiment): A Response to Tess McNulty." Post45 (blog), May 7, 2019. https://post45.org/2019/05/charisma-embodiment-a-response-to-tess-mcnulty

Core Applied Materials

- "Introduction and Chapters 1-2" So, Richard Jean. *Redlining Culture: A Data History of Racial Inequality and Postwar Fiction*. Columbia University Press, 2020. https://www.jstor.org/stable/10.7312/so--19772

Background Applied Materials

- Henley, Amanda, Matt Jansen, Lorin Bruckner, Neil Byers, and Rucha Dalwadi. "On the Books: Jim Crow and Algorithms of Resistance White Paper," August 31, 2020. https://doi.org/10.17615/hvz4-sr14.
- So, Richard Jean, Hoyt Long, and Yuancheng Zhu. "Race, Writing, and Computation: Racial Difference and the US Novel, 1880-2000." *Journal of Cultural Analytics* 3, no. 2 (January 11, 2019). https://doi.org/10.22148/16.031.
- Bamman, David. "LitBank: Born-Literary Natural Language Processing." In Computational Humanities Research, 2020. https://people.ischool.berkeley.edu/~dbamman/pubs/pdf/Bamman_DH_Debates_CompHum.pdf

6. Vectors and Clusters: Publishing Mar 4, 2024

General Framing Questions:

How can we represent culture in space? What are the benefits of high dimensional spaces and the challenges of dimensionality reduction? How well do unsupervised or off-the-shelf methods find patterns? How much validation work should we do with these methods?

Core Theoretical Materials

- Klein, Lauren F. "Dimensions of Scale: Invisible Labor, Editorial Work, and the Future of Quantitative Literary Studies." *PMLA* 135, no. 1 (January 2020): 23–39. https://doi.org/10.1632/pmla.2020.135.1.23.
- Rettberg, Jill Walker. "Algorithmic Failure as a Humanities Methodology: Machine Learning's Mispredictions Identify Rich Cases for Qualitative Analysis." *Big Data & Society* 9, no. 2 (July 1, 2022): 20539517221131290. https://doi.org/10.1177/20539517221131290.

Background Theoretical Materials

- Chang, Kent K., and Simon DeDeo. "Divergence and the Complexity of Difference in Text and Culture." Journal of Cultural Analytics 5, no. 2 (October 7, 2020). https://doi.org/10.22148/001c.17585.
- Manshel, Alexander, Laura B. McGrath, and J. D. Porter. "Who Cares about Literary Prizes?" Public Books (blog), September 3, 2019. https://www.publicbooks.org/who-cares-about-literary-prizes/.
- Walsh, Melanie. "Where Is All the Book Data?" Public Books (blog), October 4, 2022. https://www.publicbooks.org/where-is-all-the-book-data/.

Core Applied Materials

- Sinykin, Dan, and Edwin Roland. "Against Conglomeration." *Journal of Cultural Analytics* 6, no. 2 (April 20, 2021). https://doi.org/10.22148/001c.22331.
    - Sinykin, Daniel; Edwin Roland, 2021, "Replication Data for: Against Conglomeration", https://doi.org/10.7910/DVN/EUPMKL , Harvard Dataverse, V1
- "Chapters 3-4, and Conclusion" So, Richard Jean. *Redlining Culture: A Data History of Racial Inequality and Postwar Fiction*. Columbia University Press, 2020. https://www.jstor.org/stable/10.7312/so--19772

Background Applied Materials

- Schmidt, Benjamin. "Stable Random Projection: Lightweight, General-Purpose Dimensionality Reduction for Digitized Libraries." *Journal of Cultural Analytics* 3, no. 1 (October 3, 2018). https://doi.org/10.22148/16.025.
- Schmidt, Ben. "Sapping Attention: When You Have a MALLET, Everything Looks like a Nail." Sapping Attention (blog), November 2, 2012.

http://sappingattention.blogspot.com/2012/11/when-you-have-mallet-everything-looks.html.

- Underwood, Ted. "Topic Modeling Made Just Simple Enough." The Stone and the Shell (blog), April 7, 2012. https://tedunderwood.com/2012/04/07/topic-modeling-made-just-simple-enough/.

SPRING BREAK Mar 11, 2024

7. Similarities and Distances: Laws Mar 18, 2024

General Framing Questions:

How can we aggregate data to find patterns? How do we know certain patterns are meaningful? How can we find similarities and differences in our results? How do we define these terms and what do we consider statistically meaningful and at the same time interesting to our domain areas?

Core Theoretical Materials

- Schmidt, Ben. "Two Volumes: The Lessons of Time on the Cross." Ben Schmidt, December 5, 2019. https://benschmidt.org/post/2019-12-05-totc/2019-aha/.
- Ruggles, Steven. "The Revival of Quantification: Reflections on Old New Histories." Social Science History 45, no. 1 (2021): 1–25. https://doi.org/10.1017/ssh.2020.44.

Background Theoretical Materials

- Braudel, Fernand, and Immanuel Wallerstein. "History and the Social Sciences: The Longue Durée." Review (Fernand Braudel Center) 32, no. 2 (2009): 171–203. http://www.jstor.org/stable/40647704
- Guldi, Jo, and David Armitage. *The History Manifesto*. Cambridge: Cambridge University Press, 2014. https://www.cambridge.org/core/books/history-manifesto/AC1A1EC711AE91A4F9004E7582D79AFD.

Core Applied Materials

- Mullen, Lincoln A. *America's Public Bible: A Commentary*. Stanford University Press, 2023. https://americaspublicbible.supdigital.org.
    - Code for America's Public Bible https://github.com/lmullen/americas-public-bible
- Funk, Kellen, and Lincoln A. Mullen. "The Spine of American Law: Digital Text Analysis and U.S. Legal Practice (Annotated Version)," 2021. https://doi.org/10.31835/ma.2021.07.
    - Code for The Spine of American Law https://github.com/lmullen/civil-procedure-codes/

Background Applied Materials

- Tilton, Lauren, Taylor Arnold, and Courtney Rivard. *Layered Lives: Rhetoric and Representation in the Southern Life History Project*. Stanford University Press, 2022. https://doi.org/10.21627/2022ll.
- Beals, M. H. "Close Readings of Big Data: Triangulating Patterns of Textual Reappearance and Attribution in the Caledonian Mercury, 1820-1840 (Annotated Version)," 2021. https://doi.org/10.31835/ma.2021.05.

8. Speculation and Prediction: Networks Mar 25, 2024

General Framing Questions:

How can we study and predict the past with scale? How can speculation be used to understand culture? How can computation create and uncover connections?

Core Theoretical Materials

- "Viral Textuality" and "Textual Criticism as Language Modeling" in Cordell, Ryan, David A. Smith, Abby Mullen, Jonathan D. Fitzgerald, and Avery Blankenship. *Going the Rounds*. University of Minnesota Press, 2022. https://manifold.umn.edu/projects/going-the-rounds.

Background Theoretical Materials

- Ahnert, Ruth, Sebastian E. Ahnert, Catherine Nicole Coleman, and Scott B. Weingart. "The Network Turn: Changing Perspectives in the Humanities." *Elements in Publishing and Book Culture*, December 2020. https://doi.org/10.1017/9781108866804.

Core Applied Materials

- Soni, Sandeep, Lauren F. Klein, and Jacob Eisenstein. "Abolitionist Networks: Modeling Language Change in Nineteenth-Century Activist Newspapers." Journal of Cultural Analytics 6, no. 1 (January 18, 2021). https://doi.org/10.22148/001c.18841.
    - Code for Abolitionist Networks https://github.com/sandeepsoni/semantic-leadership-network
- Ahnert, Ruth, and Sebastian Ahnert. "Protestant Letter Networks in the Reign of Mary I: A Quantitative Approach (Annotated Version)," 2021. https://doi.org/10.31835/ma.2021.04.

Background Applied Materials

- Ladd, John R. "Imaginative Networks: Tracing Connections Among Early Modern Book Dedications." Journal of Cultural Analytics 6, no. 1 (March 30, 2021). https://doi.org/10.22148/001c.21993.
- Valeriola, Sébastien de. "Can Historians Trust Centrality? Historical Network Analysis and Centrality Metrics Robustness." Journal of Historical Network Research 6, no. 1 (October 25, 2021). https://doi.org/10.25517/jhnr.v6i1.105.

- Dueñas, Marco, and Antoine Mandel. "The Structure of Global Cultural Networks: Evidence from the Diffusion of Music Videos." PLOS ONE 18, no. 11 (November 13, 2023): e0294149. https://doi.org/10.1371/journal.pone.0294149.
- Nelson, Laura K. "The Inequality of Intersectionalities in Chicago's First-Wave Women's Movement." Signs: Journal of Women in Culture and Society 47, no. 4 (June 2022): 905–30. https://doi.org/10.1086/718866.

9. Influence and Embeddings: Ideas Apr 1, 2024

General Framing Questions:

How do we critically incorporate the latest technological developments into our research methods? How are these new infrastructures and architectures changing how we study culture at scale?

Core Theoretical Materials

- Nelson, Laura K. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research* 49, no. 1 (February 1, 2020): 3–42. https://doi.org/10.1177/0049124117729703.

Background Theoretical Materials

- Antoniak, Maria, and David Mimno. "Evaluating the Stability of Embedding-Based Word Similarities." Transactions of the Association for Computational Linguistics 6 (February 1, 2018): 107–19. https://doi.org/10.1162/tacl_a_00008.
- Underwood, Ted. "Do Humanists Need BERT?" The Stone and the Shell (blog), July 15, 2019. https://tedunderwood.com/2019/07/15/do-humanists-need-bert/.

Core Applied Materials

- Underwood, Ted, Kevin Kiley, Wenyi Shang, and Stephen Vaisey. "Cohort Succession Explains Most Change in Literary Culture." *Sociological Science* 9 (May 2, 2022): 184–205. https://doi.org/10.15195/v9.a8.
    - Code for Cohort Succession https://github.com/tedunderwood/period-cohort
- Vicinanza, Paul, Amir Goldberg, and Sameer B Srivastava. "A Deep-Learning Model of Prescient Ideas Demonstrates That They Emerge from the Periphery." *PNAS* Nexus 2, no. 1 (January 1, 2023): pgac275. https://doi.org/10.1093/pnasnexus/pgac275.
    - Code for Prescient Ideas https://github.com/pvicinanza/prescience

Background Applied Materials

- Nelson, Laura K. "Leveraging the Alignment between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century U.S. South." *Poetics, Measure Mohr Culture*, 88 (October 1, 2021): 101539. https://doi.org/10.1016/j.poetic.2021.101539
- Card, Dallas, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky. "Computational Analysis of 140 Years of US Political

Speeches Reveals More Positive but Increasingly Polarized Framing of Immigration." *Proceedings of the National Academy of Sciences* 119, no. 31 (August 2, 2022): e2120510119. https://doi.org/10.1073/pnas.2120510119.

- Barron, Alexander T. J., Jenny Huang, Rebecca L. Spang, and Simon DeDeo. "Individuals, Institutions, and Innovation in the Debates of the French Revolution." *Proceedings of the National Academy of Sciences* 115, no. 18 (May 2018): 4607–12. https://doi.org/10.1073/pnas.1717729115.

10. Power and Prestige: Society Apr 15, 2024

General Framing Questions:

How are new forms of social data providing new avenues for studying culture? What are the ethics of using these materials and how do they differ from other forms of cultural data?

Core Theoretical Materials

- Bail, Christopher A., D. Sunshine Hillygus, Alexander Volfovsky, Maxwell B. Allamong, Fatima Alqabandi, Diana Jordan, Graham Tierney, Tina Tucker, Andrew Trexler, and Austin van Loon. "Do We Need a Social Media Accelerator?," December 26, 2023. https://doi.org/10.31235/osf.io/ucfbk.
- Walsh, Melanie. "The Challenges and Possibilities of Social Media Data: New Directions in Literary Studies and the Digital Humanities." In *Debates in the Digital Humanities* 2023, edited by Matthew K. Gold and Lauren F. Klein, 275–94. University of Minnesota Press, 2023. https://dhdebates.gc.cuny.edu/read/f3f87448-138c-4d19-8ff8-b06acf40ddd1/section/a57b98ab-0f10-45d0-b205-3e563aab7ea8#ch18 .

Background Theoretical Materials

- Lavigne, Sam. "Scrapism: A Manifesto." *Critical AI* 1, no. 1–2 (October 1, 2023). https://doi.org/10.1215/2834703X-10734046.

Core Applied Materials

- Lucy, Li, Suchin Gururangan, Luca Soldaini, Emma Strubell, David Bamman, Lauren Klein, and Jesse Dodge. "AboutMe: Using Self-Descriptions in Webpages to Document the Effects of English Pretraining Data Filters." arXiv, January 12, 2024. https://doi.org/10.48550/arXiv.2401.06408.
- Antoniak, Maria, David Mimno, Rosamond Thalken, Melanie Walsh, Matthew Wilkens, and Gregory Yauney. "The Afterlives of Shakespeare and Company in Online Social Readership," January 14, 2024. https://arxiv.org/abs/2401.07340v1.

Background Applied Materials

- Walsh, Melanie, and Maria Antoniak. "The Goodreads 'Classics': A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism." *Journal of Cultural Analytics* 6, no. 2 (April 20, 2021). https://doi.org/10.22148/001c.22221.

11. Capitalism and Stability: Absence Apr 22, 2024

General Framing Questions:

What are the limitations of culture at scale? How can we deal with missing-ness in our data? How can this work be beneficial for society, and at the same time how can we critically understand the political economy of scale? Is it possible to have scale without capitalism? What might that look like?

Core Theoretical Materials

- Tsing, Anna Lowenhaupt. "On Nonscalability: The Living World Is Not Amenable to Precision-Nested Scales." Common Knowledge 18, no. 3 (August 1, 2012): 505–24. https://doi.org/10.1215/0961754X-1630424.
- Hanna, Alex, and Tina M. Park. "Against Scale: Provocations and Resistances to Scale Thinking," October 17, 2020. https://doi.org/10.48550/arXiv.2010.08850.

Background Theoretical Materials

- Birhane, Abeba, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. "The Values Encoded in Machine Learning Research." In 2022 ACM Conference on Fairness, Accountability, and Transparency, 173–84. FAccT '22. New York, NY, USA: Association for Computing Machinery, 2022. https://doi.org/10.1145/3531146.3533083.
- Pfotenhauer, Sebastian, Brice Laurent, Kyriaki Papageorgiou, and Jack Stilgoe. "The Politics of Scaling." Social Studies of Science 52, no. 1 (February 1, 2022): 3–34. https://doi.org/10.1177/03063127211048945.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜." In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–23. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021. https://doi.org/10.1145/3442188.3445922.
- Colavizza, Giovanni, Tobias Blanke, Charles Jeurgens, and Julia Noordegraaf. "Archives and AI: An Overview of Current Debates and Future Perspectives." Journal on Computing and Cultural Heritage 15, no. 1 (December 14, 2021): 4:1-4:15. https://doi.org/10.1145/3479010.

Core Applied Materials

- Koeser, Rebecca and Zoe LeBlanc "Missing Data, Speculative Reading" *Journal of Cultural Analytics* (pre-print)

Background Applied Materials

- Underwood, Ted, and Richard Jean So. "Can We Map Culture?" Journal of Cultural Analytics 6, no. 3 (June 17, 2021). https://doi.org/10.22148/001c.24911.
- Boykis, Vicki. "What's New with ML in Production." Vicki Boykis Tech Blog (blog), January 15, 2024. https://veekaybee.github.io/2024/01/15/whats-new-with-ml-in-production/.

- Ryan, Yann C., and Sebastian E. Ahnert. "The Measure of the Archive: The Robustness of Network Analysis in Early Modern Correspondence." Journal of Cultural Analytics 6, no. 3 (July 21, 2021). https://doi.org/10.22148/001c.25943.

12. Plots: Narratives and Storytellings Apr 29, 2024

Student Presentations

General Framing Questions:

How do we tell stories with our data? How do we give constructive feedback? How do we balance what we have achieved with what we hoped to achieve? How do we balance explaining knowledge claims with detailing methods and data selection?

No assigned materials

13. Presentations: Interpretations and ~~Final~~ Future Steps May 6, 2024

Student Presentations

General Framing Questions:

What are our next steps with our projects? How do we document what we have and explain future directions? What is the relationship between coding documentation and written publications?

No assigned materials

Archived Week: Background Viewing and Visible: Images

General Framing Questions:

How do we deal with non-textual data? What are generative methods, and what are the tradeoffs for dealing with accuracy of our phenomenon of interest? How do we balance considerations of ownership and ethics with data driven research?

Core Theoretical Materials

- "Introduction and Chapters 1-2" Arnold, Taylor, and Lauren Tilton. *Distant Viewing: Computational Exploration of Digital Images*. The MIT Press, 2023. https://doi.org/10.7551/mitpress/14046.001.0001.
    - Code for Distant Viewing https://github.com/distant-viewing/dvt

Background Theoretical Materials

- Offert, Fabian, and Peter Bell. "Generative Digital Humanities," 2020. https://www.semanticscholar.org/paper/Generative-Digital-Humanities-Offert-Bell/e5aadc6bd00c6518cf97c22be259c4c621370c7f.
- Lang, Sabine, and Bjorn Ommer. "Transforming Information Into Knowledge: How Computational Methods Reshape Art History." *Digital Humanities Quarterly* 15, no. 3 (2021). http://www.digitalhumanities.org/dhq/vol/15/3/000560/000560.html.

Core Applied Materials

- Thompson, Laure, and David Mimno. "Computational Cut-Ups: The Influence of Dada." *The Journal of Modern Periodical Studies* 8, no. 2 (July 1, 2017): 179–95. https://doi.org/10.5325/jmodeperistud.8.2.0179.
- Hosseini, Kasra, Katherine McDonough, Daniel van Strien, Olivia Vane, and Daniel C S Wilson. "Maps of a Nation? The Digitized Ordnance Survey for New Historical Research." Journal of Victorian Culture 26, no. 2 (April 1, 2021): 284–99. https://doi.org/10.1093/jvcult/vcab009.

Background Applied Materials

- Lee, Benjamin Charles Germain, Joshua Ortiz Baco, Sarah H. Salter, and Jim Casey. "Navigating the Mise-En-Page: Interpretive Machine Learning Approaches to the Visual Layouts of Multi-Ethnic Periodicals," September 3, 2021. https://doi.org/10.48550/arXiv.2109.01732.
- Fyfe, Paul, and Qian Ge. "Image Analytics and the Nineteenth-Century Illustrated Newspaper." *Journal of Cultural Analytics* 3, no. 1 (October 25, 2018). https://doi.org/10.22148/16.026.