# ENG 612/MLL 772
# Topics in DH: Humanities Data
# Spring 2022

Course site: http://lindsaythomas.net/eng612s22

Wednesdays, 2-4:30 pm
MB 205 (via Zoom until Jan 31)

Lindsay Thomas, lindsaythomas@miami.edu
Office hours: W, 11:30 am – 1:30 pm, and by appointment
In person (outside) or via Zoom

## Course Description

This class will provide an introduction to data – as a concept, as an object and method of study, and as a scholarly product – and how it operates in the humanities today. We will focus not only on how humanists understand the concept and history of data, but also on how they go about collecting, organizing, and analyzing it. We will discuss the place of humanistic data analysis within what is more widely known as the digital humanities; what constitutes "data" in the humanities and how to go about collecting it; the relationship between data and archival collections; the logic, practice, and problems associated with quantification; methods of data analysis; and what it means to understand datasets as scholarship. We will explore a variety of computational research projects in the humanities over the course of the semester, seeking to understand the decisions researchers have made in constructing, interpreting, and publishing their data and to articulate the consequences, both positive and negative, of these decisions.

This class is designed to introduce participants to the concepts and methods researchers employ when collecting and analyzing humanities data (for our purposes, this mainly means text). As such, it will include a significant hands-on component: participants will learn to explore and analyze existing humanities datasets and, by the end of the semester, construct their own scholarly dataset. This will entail developing basic familiarity with spreadsheets, regular expressions, and Python, a programming language. However, this class is not designed to teach programming. Rather, the goal will be to equip participants with some basic research skills that will make collecting, organizing, exploring, and analyzing data – and, crucially, understanding how other scholars have collected, organized, explored, and analyzed data – easier. Assignments will include lab reports, including a reflection on an existing scholarly dataset, and the creation of a scholarly dataset that reflects participants' research interests and the semester's discussions of humanities data.

The course is open to students across the humanities, although it will focus on literary and cultural studies. No experience in the digital humanities or with digital tools or methods is required. This course will count toward the completion of the Graduate Certificate in Digital Humanities (it will count as the practicum course for those students who need to fulfill that requirement this year).

## COVID Disclaimer

As we all know, times are tough. This is a demanding course, and while I have tried to adjust the course policies and expectations to account for the times, I'm certain I have not imagined every

situation that may arise or accounted fully for every challenge you may be facing this semester. Please know, without a doubt, that your health and the health of your friends and family will always be more important to me than this course. You do not have to apologize to me if we need to find an alternate path for you through this course, or if you need more time to complete something. I want to support you, and I want you to communicate with me about how I can do that better.

## Masks

For the first two weeks of the semester, we will meet remotely using Zoom. Assuming we return to in-person instruction after that, you will be required to wear a mask covering your mouth and nose at all times during class. You may briefly remove your mask to drink. Each class session, we will take a 10-minute break after the first 75-90 minutes of class. Please try to eat before or after class, or during the break (outside the room), so that you do not need to remove your mask for long periods of time during class.

## Course Digital Infrastructure and Creating your own Website

We will make use of several online systems and program in this course: a course site, a class Google drive folder, and Zoom. The "ENG 612 S22 Digital Infrastructure" doc in our shared class folder on Gdrive will contain the passwords and other information you need to use our course digital systems.

Course Site

We will use our course site to manage course information and our schedule (http://lindsaythomas.net/eng612s22). You will find an online version of our course calendar there (including the most up-to-date version of reading assignments and due dates), as well as information about all course assignments.

Google Drive

We will also use a shared class Google drive folder to distribute course readings and other materials. Every UM student, staff, and faculty member has free access to Google Drive, but you will need to sign in to Google Drive using your UM email and password. At the beginning of the semester, I will share you into this folder via your UM email address.

To access our class Google drive folder:

1. Go to drive.google.com and log in using your UM email address (the one given to you by UMIT, with the numbers in it, NOT an aliased email). If you are signed into Google drive via another account, you will need to sign out or select "add another account" by clicking on your account icon in the upper right corner of the screen.
2. Entering your UM email address will redirect you to the UM single sign on page, where you will enter your CaneID and password.
3. You will now be signed into Gdrive with your UM credentials. You will know you are signed in with your UM credentials because the U logo will appear in the upper right corner of your screen. If you do not see this logo, you are not signed in with your UM credentials. Signing in with your UM credentials is important because it protects your academic work behind UM's firewall and ensures your privacy.

4. Click on "Shared with me" in the left-hand menu, where you should see our class folder: "ENG 612 MLL 772 S22".

Zoom

We will use Zoom for the first two weeks of classes, as well as for office hours and one-on-one meetings (if not meeting outside). The "ENG 612 S22 Digital Infrastructure" document in our class Google drive folder will contain our class Zoom link and our office hours link.

We will not record class sessions or office hours. Our class meeting room will not have a waiting room, but I will enable a waiting room for office hours. You will enter the waiting room first when signing on via our office hours link. I will then let you into the office hours Zoom room. If I do not let you in right away, this is because I am meeting with another student.

Creating your Website

You will also create your own website to which you will post your lab notebook entries and the written portion of your final project. You will begin creating this website on the first day of class, in Lab 1. You will share the address of your website (and its attendant repository) with me so that I can check your progress. For those students who plan on completing the requirements of the Graduate Certificate in Digital Humanities, this website can serve as the foundation for your final portfolio (though it doesn't have to). You will have the option to make your website repository private, but if you choose to make it public, I will post the address in a document in our class Google drive folder so that other members of the class may read your posts, as well (note: because this document will only be available via our shared Google drive folder, people outside of this class will not have access to this list).

## Doing New Things, Like Coding, But Not Just Coding

In this class, you will be asked to do and to read new things, things with which you may not have much, or any, experience. This may include learning technical details or processes, including coding, but it will also include learning about concepts and methods in how to collect, curate, organize, analyze, store, and make reproducible humanities datasets. These tasks involve new ways of thinking and will sometimes feel frustrating and hard. There will be times when you get stuck, or when you are confused about what to do in a lab, or when you are wrong. All of that is fine and expected, and I want to encourage all of us to make this class safe for experimentation, error, and failure. I do not expect anyone to walk into or out of this class an expert in work with humanities data. What I do expect is a good-faith effort, a willingness to experiment, and the ability to keep an open mind. Most importantly, please be gentle on yourself without giving up; failure is not a reflection on who you are as a person, student, or scholar – it's just a reflection of the fact that you are learning something new.

We will be doing some coding in Python in this class, and we will also be working with the command line, GitHub, and Excel. We are doing this technical work for three reasons: 1) To introduce you to what it's like to work with data in a humanities research context and to communities of humanities scholars who collect and use data in their work; 2) To teach you how you can contribute to these communities (and to your other fields of expertise) by communicating your research and making it reproducible for others; 3) To help you to complete the final project in this class, which will involve creating your own humanities dataset.

About 10 years ago, there was some debate in the digital humanities about whether people "need" to code in order to do DH. You do not need to code to do DH, at all, mainly because "doing DH" involves a wide variety of practices, only a few of which this course will cover. Working with data – including collecting humanities datasets, curating online archives, and doing computational research and/or criticism – by no means constitutes the entirety of "doing DH." There are also many GUI (graphical-user interface) tools designed to accomplish some tasks in data management and/or analysis (including Excel, ArcGIS, Gephi, the GUI topic modeling tool, etc). Generally speaking, these tools require no coding. However, the argument of this class is that achieving some level of technical literacy, however basic, is advantageous because it gives you a baseline for understanding – and for potentially being able to implement – new and unfamiliar developments in humanities data work. It also (gently, I hope!) encourages you to "get your hands dirty" by trying things out and, inevitably, by experiencing and overcoming failure and frustration.

If you take away only one thing about the technical side of working with data in the humanities from this class, I hope it is that you can absolutely teach yourself the technical skills you need to do the research you want, even if you think you are not a "technical person." There is no way that I can teach you every technical skill you might need to work with data in the humanities; there are simply too many different things one can do, and technical skills, methods, and platforms change very quickly. However, my hope is that in introducing you to some basics, you will learn that there is nothing special about learning to work with data in the humanities, and that given time and effort, you can learn to do it, too.

## Consultative Grading

I have borrowed the philosophy and much of the language in this section from Ryan Cordell, whose experiments in contract and consultative grading I have been following for a couple of years now.

Grading, especially in a graduate course, is not an accurate reflection of your learning or your abilities. It also does not resemble how you will be evaluated in your careers, where you will define many of your own goals and be assessed by how effectively you achieve them. For these reasons, last semester, I tried contract grading for the first time in my undergraduate course. It worked well, and I want to adapt this system for our graduate course this semester; adopting Cordell's terminology, we will call it "consultative grading." My hope is that this system will encourage you to take ownership over your own work, ensuring that it meets professional (or, perhaps more accurately, professional-in-training) standards, while also allowing room for flexibility, experimentation, and even failure.

Throughout the semester, you will assess your work in the course in dialogue with me, as mentor rather than judge. Since the university will ultimately require me to assign a grade to your work at the end of the semester, you will grade your own work by assessing your effort and performance across the course assignments as they relate to the goals you set for yourself, your work to meet those goals, and your intellectual growth during the class. Formally, this means that I will ask you to draft a statement of your goals at the beginning of the semester and then to draft self-evaluations two times during the semester. This will include a final self-evaluation through which you will assign yourself a grade. Barring extreme circumstances (see the Adjustment Caveat below) this self-assessment will determine your grade for the semester. Ideally, knowing this process in advance will free you to do more ambitious work from the beginning of the semester.

My Commitments

To foster your progress this semester, I commit myself to:

1. Providing substantive and timely commentary on your assignments aimed at cultivating your research skills, analytical abilities, and scholarly voice.
2. Making myself available for in-person consultation and practical help during office hours and at other scheduled times, including virtual and in-person meetings (in-person meetings will be outside if we can manage it, after the first two weeks of classes).
3. Assuming no prior knowledge of course concepts and no technical expertise from students going into any of our labs. We will begin at the beginning, so that no one feels left behind.
4. Allowing students with expertise in particular technologies or skills, or with ongoing projects they wish to continue working on, to challenge themselves and craft their own laboratory experiences and/or final project assignment beyond our work in class.
5. Working with you to understand your goals and methods when you take intellectual risks in assignments, even if the final product does not turn out as expected.
6. Respecting you, your perspectives, and your intellectual commitments in class discussions and assignments. I may push you to consider other perspectives or ideas, but I will not dismiss your thoughts or take them lightly. If you feel I am doing either of those things, I will listen and adjust my responses as necessary.

Your Commitments

This system will only work, however, if you also commit to:

1. Holding yourself to the highest standards. You should work to the best of your abilities throughout the semester in your reading, in-class conversation, and assignments.
2. Taking intellectual risks when possible, pushing yourself to think, write, and create in new modes and grow as a scholar, while also ensuring our course is safe space for experimentation and failure. Push yourself beyond your comfort zone, but also allow yourself – and your class colleagues – to be wrong and to change your mind. This may prompt anxiety, but we can work through this together.
3. Seeking to understand first, and to question and critique second, only after you understand. This applies to everything we do in this class, from course readings, to discussing them in class, to completing assignments and sharing them with your colleagues.
4. Experimenting with new tools and learning new technical skills with an open mind.
5. Assisting your colleagues with our labs when you have prior expertise, or if you acquire it more quickly during the lab itself.
6. Respecting your colleagues and their perspectives and intellectual commitments in class discussions and assignments. You may push them to consider other perspectives, but you should not dismiss their thoughts or take them lightly. If someone feels you are doing either of those things, you should listen and adjust your responses as necessary.
7. Meeting with me, in person or via Zoom, at least once during the semester to discuss your work and ensure you are meeting expectations (mine and yours) for work in the class.

Adjustment Caveat

I reserve the right to adjust grades as appropriate if a student takes undue advantage of the consultative grading paradigm. However, I do not anticipate needing to exercise this right.

## Assignments

### Readings

All of our course readings are available online or through our class Google drive folder. I expect you to come to class having completed all of the assigned reading for that day.

### Attendance and Active Participation

As with any graduate course, you are expected to do the reading for, attend, and actively participate in every class period, barring emergencies or illness. If you are someone for whom participation in class is difficult, please talk to me about strategies for increasing your participation. If you know you will miss a particular class period, please let me know in advance. In case this needs to be said: Obviously, I do not want you to come to class if you feel sick, if you are experiencing even mild cold- or flu-like symptoms, if you have been in contact with someone who tests positive for COVID-19, or if you test positive for COVID-19. Just stay home, take care of yourself, and do not run the risk of infecting others. If you find that you are ill or will have to quarantine for an extended period (i.e., more than 1 class), please talk to me and we will come up with a way to catch you up. Your grade will not be negatively affected due to illness or quarantine.

### Lab Notebook

Due:

- Lab 1: Wed, Feb 2
- Lab 2: Web, Feb 9
- Lab 3: Wed, Feb 16 (Feedback on labs 1-3 from instructor)
- Lab 4: Wed, Feb 23
- Lab 5: Wed, March 2
- Lab 6: Fri, March 11 (Feedback on labs 4-6 from instructor)

During the first eight weeks of the semester, you will complete labs designed to introduce you to some basic technical skills and procedures that humanists who work with data might use. On most of these weeks (as marked in the syllabus), the last hour of each class period will be devoted to beginning the labs together in class. Two central goals of this course are to use our readings and discussions during these weeks to contextualize our applied work in labs, and to use our applied work in labs to enrich and enhance our understanding of concepts from our readings.

You will report on your completion of these labs using your lab notebook, which you will post to your website repository. Your lab notebook entries will vary quite a bit from lab to lab, but in each lab notebook entry, I will expect to find a brief description of the lab activities completed. For our coding lab, you will likely integrate code snippets directly into your Markdown file (more on this as that lab approaches), while for other labs you may instead reference external proof of your work, such as screenshots. Each lab notebook entry will also include a written discussion of/reflection on the lab. For most labs, I will include a prompt to help start your thinking; you should begin the reflective portion of the lab from there. In this portion of the lab, I will expect prose that reflects analytically on the work of the lab, putting it into conversation with one or two readings from the same week of class as the lab (you may also include readings from the larger class or beyond if appropriate). You should integrate the readings explicitly, through direct quotation if possible.

Your final lab notebook entry (lab 6) will be slightly more substantial. For this lab, you will select an existing scholarly dataset that you have not created; answer some questions about its composition,

organization, and scope; and discuss its contributions to a particular field(s) and/or subfield(s). You will need to select the dataset you plan to use for this assignment by class on Wednesday, Feb 23, as you will use this dataset in completing lab 5 as well.

During weeks 2-6, we will find a time each week to hold weekly work sessions outside of class. I will be available during this time for drop-in help with the labs and other challenges, and class members can help each other during this time as well.

You will post each lab notebook entry to your website repository as its own Markdown (.md) file. Lab notebook entries are due 1 week after beginning the lab in class (except for labs 1 and 6). I will check each week that you have completed each lab notebook entry, and I will offer feedback on your lab notebook entries as a whole after labs 3 and 6. **To emphasize: each lab notebook entry should be saved as a separate file in your GitHub repository.**

You should name your lab notebook entries using the following convention:

- YYYY-MM-DD-labreport-WHAT-WORDS-YOU-WANT.md

**Final Project**

Due:

- Abstract: Wed, April 6
- 4-5 minute presentation of final project in progress: Wed, April 27
- Proof-of-concept dataset, codebook, and critical introduction: finals week (date TBD by class)

Your final project in this class is a scholarly dataset and a critical introduction to that dataset. You may work with other members of the class to complete the final project. If working with others, your team will turn in 1 copy of your proof-of-concept dataset (i.e., you don't each have to turn it in to me). You will also have a choice about how you would like to write your critical introduction: each member of your team may write their own, or, alternatively, you may turn in a collaboratively written critical introduction (or some combination of these two options). If you opt to collaborate with others on the final project, it will be important for your team to discuss with me how you plan to organize your labor so that all members contribute equally to the final product.

If you are already working on a project of your own involving data and/or the creation of a scholarly dataset, you may continue that work for your final project. We will work together to determine what exactly this work will look like so that it fulfills the requirements of this assignment.

<u>Planning your Dataset and Writing your Abstract</u>

Your goal is to create a proof-of-concept scholarly dataset that could be used to answer research questions in a particular field/subfield of humanistic inquiry (ideally, in your own field(s) of research). For many of you, given the focus of this class, this dataset will likely consist of works of literature, films, social media posts, and/or cultural artifacts of various kinds (whether contemporary or historical). Alternatively, you may wish to create a dataset including ethnographic data of various kinds (survey results, interview transcriptions, etc.), though you should talk to me well in advance if you are interested in this option. There is no minimum number of records that your dataset needs to contain, nor is there a minimum number of metadata fields – these numbers will vary depending on the data being collected, the methods of collection, the information available, etc. – but you should

strive to make the theoretical version of your dataset as complete and as fully imagined as possible. However, the actual dataset you turn in to me will likely only be a small(ish) subset of this larger, fully imagined dataset. This is what I mean by the term "proof-of-concept dataset;" because you likely won't have time to collect all of the data that your (fully imagined) dataset contains, you may wish to focus on collecting (and organizing, describing, etc.) only a subset of this data by the end of the semester.

How you collect this data is up to you, but you should take considerations about data collection very seriously when deciding what your dataset will be. You should think hard about what kind of data it will be possible for you to collect in the time that you have to complete this assignment. When deciding what data you want to collect, consider the following criteria:

- This dataset should not already exist.
- Your dataset should be conceptually meaningful, meaning its entries should be grouped/collected logically and according to explicit criteria. You should keep the needs of two sometimes competing audiences in mind when creating your dataset: 1) Scholars in the particular field(s)/subfield(s) in which the dataset is located, i.e., content experts; and 2) Other humanities researchers who may wish to use your data in their own projects to answer questions you may not be fully aware of, i.e., general use experts.
- You should be able to collect this data ethically and transparently. This means you should be aware of copyright and/or fair use restrictions (if applicable), human subjects protocols (if applicable), and other potential barriers or complications to collection.
- You should have ideas about how you would scale up data collection efforts if you had the time (and/or the money) to collect the full dataset (instead of just a subset, like you are doing for this class).

You (and/or your team, if you are working collaboratively) will turn in an abstract describing your plans for your dataset by Wednesday, April 6. This abstract should include the following:

1. If you are working with a partner or a team, who your partner or teammates are.
2. The title of your dataset.
3. A brief description of the kind of data your dataset will contain, how you plan to collect it, and your dataset's boundaries/scope (~2-3 paragraphs). What is included or excluded from your dataset, and why (i.e., what are the inclusion/exclusion criteria)?
4. The titles and brief descriptions of each metadata field in your dataset (i.e., an initial draft of your dataset documentation/codebook). This will likely change (and expand) as you work on your dataset, but you should have an initial plan.
5. The audience(s) for your dataset. Who are you creating this dataset for? Again, you should think in terms of the overlapping (and often competing) audiences of content experts and general-use experts here.
6. Several questions your dataset could help these audiences answer.
7. A list of 2-3 already existing related scholarly datasets and an explanation of how your dataset offers a unique contribution/how it is different from these existing datasets (~2-3 paragraphs).

<u>Creating your Dataset</u>

You can create your dataset in whatever format makes the most sense for your data (my guess is that most of you will choose to present your data as an Excel/Google sheets spreadsheet, but this is by no means the only option). If you use a spreadsheet to organize your data, each record should be 1

row of your dataset, and your metadata fields should comprise the columns of your spreadsheet. As always, please let me know if you have questions about the best format for presenting your data.

When you turn in your dataset, you should also turn in a codebook, or documentation of each metadata field included in your dataset and a brief description of what that field means/the kind of information it records. You can turn this in as a separate list, as a separate tab in your spreadsheet, as an appendix to your critical introduction, or in whatever format suits your data/project the best. We will look at some examples in class.

Presenting your Dataset in Progress

Our class period on Wednesday, April 27 will be devoted to quick presentations of your datasets in progress. In these presentations, you should describe your dataset (including the kind of data it contains, its boundaries/scope, and its metadata fields) and quickly contextualize this dataset in relation to other existing scholarly datasets, the questions it allows researchers to answer, gaps in existing fields, etc. In brief, your presentation should answer the following 2 questions: 1) What is this dataset?; 2) Why is it significant?

If you are working on your final project individually, your presentation should be about 4 minutes long (I will warn you at 4 minutes and cut you off at 5 minutes). If you are working on your final project with others, talk to me in advance about the length of your presentation and its content, as team presentations will likely need more time.

Writing your Critical Introduction

After collecting your dataset, you will write a ~2500-3000-word critical introduction to your dataset. You may organize your critical introduction how you choose, but it should contain the following elements:

1. A description of your dataset and (brief) documentation of how you collected it so that your collection efforts are reproducible (to the extent this is possible). Depending on what your data collection process entailed, you may wish to include discussion of data collection as a technical appendix to your critical introduction.
2. An examination of the affordances and limits of your dataset, of the curatorial choices you made in creating your dataset, of the questions it allows researchers to ask, and/or of what other issues, questions, and/or data in its field(s) it is in conversation with. This may include some initial exploratory analysis of your dataset, though it need not. Basically, I am asking you here to contextualize your dataset in relation to existing scholarship, to discuss any unique features or affordances of your dataset, and to argue for its overall significance.
3. A reflection on 1-2 issues, problems, or larger concepts that creating this dataset helped you to understand or to think about more clearly. What did this process illuminate for you, either about the data you chose to collect specifically or the process of data collection more generally or the concept of data itself? You should relate this discussion to at least 1-2 readings from our class, though you may also include other texts as appropriate.

While this specific genre of paper may be new to you, what I am asking for here still involves research. This means your critical introduction should demonstrate knowledge of its field (i.e., post-1945 US literature, or what have you), and it should contribute to knowledge in this field. It should include a works cited page/bibliography. You may wish to explore the pieces published in the "Data Sets" section in the *Journal of Cultural Analytics* to get a sense of the range of things you can discuss in your critical introduction.

If you are working with a team and you and your team plan to co-author the critical introduction, you should speak to me in advance about your plans for writing this document to ensure equitable distribution of labor. Collaborative critical introductions will likely be longer in length.

Turning in your Final Project

You (and/or your team) will post your critical introduction to your (or a specific team member's) website as its own Markdown (.md) file. You can turn in your dataset in a way that makes sense for your data: you may want to send me a link to a Google sheets spreadsheet or a GitHub repo, share a folder with me on Dropbox, email me the dataset, etc.

You should name your critical introduction using the following convention:

- YYYY-MM-DD-finalproject-WHAT-WORDS-YOU-WANT.md

Final Self-Assessment

Please assess your effort and performance across in-class discussions and the course assignments as they relate to the goals you set for yourself, your work to meet those goals, and your intellectual growth during the semester. If you collaborated with others on your final project, please also include an assessment of your contributions to the final project. Finally, please include the letter grade you would give yourself for the semester. There is no specific length requirement, but you should be detailed and specific.

Each member of the class should individually email me their final self-assessment when they submit their final project (i.e., do not post your self-assessment to your website).

## Schedule

Readings are either linked below or located in our course Google drive folder (in the "Course Readings" subfolder). Readings are due – meaning they should be completed – on the dates indicated.

The most accurate and up-to-date version of this calendar can be found on our course site. Use the online calendar to check on reading assignments, rather than this print version, since the print version of this syllabus will not be updated throughout the semester.

I reserve the right to change the course calendar as needed; adequate advance notice will always be given of any changes.

**Week 1: Introductions**
**Wednesday, January 19**

- Matthew K. Gold, "The Digital Humanities Moment," from *Debates in the Digital Humanities 2012* (2012)
- Lauren F. Klein and Matthew K. Gold, "Digital Humanities: The Expanded Field," from *Debates in the Digital Humanities 2016* (2016)
- Matthew K. Gold and Lauren F. Klein, "A DH That Matters," from *Debates in the Digital Humanities 2019* (2019)
- Kelly Baker Josephs and Roopika Risam, "The Digital Black Atlantic," from *The Digital Black Atlantic* (2021)
- Lab 1: GitHub Pages and Markdown

**Week 2: What is Data? What is Digital Scholarship?**
**Wednesday, January 26**

- Daniel Rosenberg, "Data Before the Fact," from *"Raw Data" is an Oxymoron* (2013)
- Catherine D'Ignazio and Lauren F. Klein, "Introduction: Why Data Science Needs Feminism," from *Data Feminism* (2020)
- Catherine D'Ignazio and Lauren F. Klein, "Ch 2: Collect, Analyze, Imagine, Teach," from *Data Feminism* (2020)
- Julia Flanders and Trevor Muñoz, "An Introduction to Humanities Data Curation," from *DH Data Curation*
- Lara Putnam, "The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast," *The American Historical Review* (2016)
- Lab 1, con't: GitHub Pages and Markdown

**Week 3: An Impossible View from Nowhere (or, on objectivity and quantification)**
**Wednesday, February 2**

- **Lab 1 notebook entry due by class**
- Catherine D'Ignazio and Lauren F. Klein, "Ch 3: On Rational, Scientific, Objective Viewpoints from Mythical, Imaginary, Impossible Standpoints," from *Data Feminism* (2020)
- Catherine D'Ignazio and Lauren F. Klein, "Ch 4: What Gets Counted Counts," from *Data Feminism* (2020)

- Sarah Wilson, "Black Folk by the Numbers: Quantification in Du Bois," *American Literary History* (2016)
- Jessica Marie Johnson, "Markup Bodies: Black [Life] Studies and Slavery [Death] Studies at the Digital Crossroads," *Social Text* (2018)
    - Familiarize yourself with the Trans-Atlantic Slave Trade Database: https://www.slavevoyages.org/voyage/database
- Lab 2: Spreadsheets

**Week 4: Collecting, Organizing, and Cleaning Data**
**Wednesday, February 9**

- **Lab 2 notebook entry due by class**
- Katie Rawson and Trevor Muñoz, "Against Cleaning," from *Debates in the Digital Humanities 2019* (2019)
- Catherine D'Ignazio and Lauren F. Klein, "Ch. 5: Unicorns, Janitors, Ninjas, Wizards, and Rock Stars" from *Data Feminism* (2020)
- Ryan Cordell, "Why You...Should Care About OCR" (Jan 19, 2019)
    - Skim report: David A. Smith and Ryan Cordell, "A Research Agenda for Historical and Multilingual Optical Character Recognition" (2018)
- Quinn Dombrowski, "DSC #1: Quinn's Great Idea," from *The Data-Sitters Club* (Nov 7, 2019)
- Katherine Bowers and Quinn Dombrowski, "DSC #2: Katia and the Phantom Corpus," from *The Data-Sitters Club* (Dec 12, 2019)
- Lab 3: RegEx

**Week 5: Understanding Data I**
**Wednesday, February 16**

- **Lab 3 notebook entry due by class (Lab notebook check: Lindsay will give individual feedback on labs 1-3 after this class)**
- Julia Flanders and Fotis Jannidis, "Data modeling in a digital humanities context" from *The Shape of Data in Digital Humanities* (2019)
- Sarah Allison, "Other People's Data: Humanities Edition," *Journal of Cultural Analytics* (2016)
- Catherine D'Ignazio and Lauren F. Klein, "Ch. 6: The Numbers Don't Speak for Themselves" from *Data Feminism* (2020)
- Maria Sachiko Cecire, "DSC #3: The Truth About Digital Humanities Collaborations (and Textual Variants!)", from *The Data-Sitters Club* (Jan 10, 2020)
- Anouk Lang, "DSC #4: AntConc Saves the Day," from *The Data-Sitters Club* (April 10, 2020)
- Katherine Bowers, "DSC #6: Voyant's Big Day," from *The Data-Sitters Club* (Sept 15, 2020)
- Lab 4: Exploratory Data Analysis I

**Week 6: Understanding Data II**
**Wednesday, February 23**

- **Lab 4 notebook entry due**
- Taylor Arnold and Lauren Tilton, "New Data? The Role of Statistics in DH," from *Debates in the Digital Humanities 2019*

- Richard Jean So, "'All Models are Wrong,'" *PMLA* (2017)
- Benjamin M. Schmidt, "Do Digital Humanists Need to Understand Algorithms?" from *Debates in the Digital Humanities 2016*
- Jo Guldi, "Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora," *Journal of Cultural Analytics* (2018)
- Laura Nelson, "Computational Grounded Theory: A Methodological Framework," *Sociological Methods & Research* (2017)
- Lab 5: Exploratory Data Analysis II
    - Come to class having selected the scholarly dataset you will use for labs 5 and 6 (see the last section of Lab 4)

**Week 7: Digital Archives**
**Wednesday, March 2**

- **Lab 5 notebook entry due**
- Trevor Owens, "Introduction: Beyond Digital Hype and Digital Anxiety," from *The Theory and Craft of Digital Preservation* (2018), focus especially on the "Sixteen Guiding Digital Preservation Axioms" (pgs 4-9)
- Andrew Prescott and Lorna Hughes, "Why Do We Digitize?: The Case for Slow Digitization," *Archive Journal* (2018)
- Thomas Padilla, "On A Collections As Data Imperative," Library of Congress (2017)
- Jennifer Guliano and Carolyn Heitman, "Difficult Heritage and The Complexities of Indigenous Data," *Journal of Cultural Analytics* (2019)
- Miriam Posner and Marika Cifor, "Generative Tensions: Building a Digital Project on Early African American Race Film," *American Quarterly* (2018)
    - Familiarize yourself with the Race Film Database: https://zenodo.org/record/160585#.YcHs8SxOmLc
- The Colored Conventions Project:
    - P. Gabrielle Foreman, Sarah Patterson, and Jim Casey, "Introduction to the Colored Conventions Movement"
    - "CCP Principles"
    - "The Colored Conventions Project in Three Videos" (watch the 3 videos)
    - Familiarize yourself with the Digital Records section of the site
    - Familiarize yourself with the CCP Corpus (download and explore)
    - Familiarize yourself with the Exhibits section of the site (select at least 1 exhibit to skim/explore)
- Discuss Lab 6

**Week 8: Encounters with Digital Archives**
**Wednesday, March 9**

- Matthew Kirschenbaum, "Ch. 1: Archives Without Dust," from *Bitstreams: The Future of Digital Literary Heritage* (2021)
- Mike Ashenfelder, "The NEH 'Chronicling America' Challenge: Using Big Data to Ask Big Questions," *The Signal: A Library of Congress Blog* (Aug 4, 2016)
- Roopika Risam, "Ch 2: Colonial Violence and the Postcolonial Digital Archive" from *New Digital Worlds* (2019)

    - Familiarize yourself with the Early Caribbean Digital Archive
- Lauren F. Klein, "The Image of Absence: Archival Silence, Data Visualization, and James Hemings," *American Literature* (2013)
    - Familiarize yourself with The Papers of Thomas Jefferson archive
- Michelle Caswell, Marika Cifor, and Mario H. Ramirez, "'To Suddenly Discover Yourself Existing': Uncovering the Impact of Community Archives," *The American Archivist* (2016)
    - Familiarize yourself with the South Asian American Digital Archive

**Friday, March 11: Lab 6 due (Lab notebook check: Lindsay will give individual feedback on labs 4-6 over break)**

**Week 9**
**Wednesday, March 16: NO CLASS – SPRING BREAK**

**Week 10: Data and Computation in Post-45 US Literary Studies**
**Wednesday, March 23**

- Discussion of final project assignment
- Ted Underwood, "A Genealogy of Distant Reading," *Digital Humanities Quarterly* (2017)
- Richard Jean So, "Introduction," from *Redlining Culture: A Data History of Racial Inequality and Postwar Fiction* (2020)
- Richard Jean So, "Ch. 1: Production: On White Publishing," from *Redlining Culture: A Data History of Racial Inequality and Postwar Fiction* (2020)
- Laura B. McGrath, "America's Next Top Novel," *Post45* (2020)
- Richard Jean So, "Contemporary Culture After Data Science," *Journal of Cultural Analytics* (2021) (introduction to special issue that includes Sinykin and Roland's article below)
- Dan Sinykin and Edwin Roland, "Against Conglomeration," *Journal of Cultural Analytics* (2021)
- Contribute 1 reading to class Google doc of readings for next week (can contribute a digital project/archive, but be specific about what you want people to read/explore)

**Week 11: Data and Computation in the Humanities**
**Wednesday, March 30**

- Readings set by class
    - Select 3 readings from class Google doc to read for today (including the one you contributed)
- After break: Discuss preliminary final project ideas with class (come to class ready to discuss this)

**Week 12: Proceed with Care**
**Wednesday, April 6**

- **Final project abstract due**
- Nan Z. Da, "The Computational Case against Computational Literary Studies," *Critical Inquiry* (2019)

    - Also skim the Appendix, especially section 9, "Suggested Guidelines for Reviewing CLS Manuscripts," pg 25. This is in our course Box folder as a zip file, titled "Da - Appendix.zip." Download the zip file and unzip to read.
- "Computational Literary Studies: A Critical Inquiry Online Forum" (March 2019) (these are short responses to Da's article that were published on *Critical Inquiry*'s blog shortly after the publication of her article):
    - Katherine Bode, Day 1 response
    - Lauren F. Klein, Day 1 response
    - Hoyt Long and Richard Jean So, Day 1 response
    - Katherine Bode, Day 2 response
    - Lauren F. Klein, Day 3 response
    - Nan Z. Da, Final comments
- Recommended: Katherine Bowers and Quinn Dombrowski, "Katia and the Sentiment Snobs," from *The Data-Sitters Club* (Oct 25, 2021)

**Week 13: Futures of Humanities Data Work**
**Wednesday, April 13**

- Safiya Umoja Noble, "Toward a Critical Black Digital Humanities," from *Debates in the Digital Humanities 2019* (2019)
- Ted Underwood, "Digital Humanities as a Semi-Normal Thing," from *Debates in the Digital Humanities 2019* (2019)
- Kaiama L. Glover and Alex Gil, "On the Interpretation of Digital Caribbean Dreams," from *The Digital Black Atlantic* (2021)
- Rachel Mann, "Paid to Do but Not to Think: Reevaluating the Role of Graduate Student Collaborators," from *Debates in the Digital Humanities 2019* (2019)
- Christina Boyles, Anne Cong-Huyen, Carrie Johnston, Jim McGrath, and Amanda Phillips, "Precarious Labor and the Digital Humanities," *American Quarterly* (2018)
- Data Sets section of *JCA*: https://culturalanalytics.org/section/1579-data-sets
    - Select a recent article to skim
- Post45 Data Collective: https://data.post45.org/our-data/
    - Skim available datasets

**Week 14: NO CLASS**
**Wednesday, April 20**

- Lindsay will be available during class time for individual/team meetings about final projects

**Week 15: Final Projects**
**Wednesday, April 27**

- Final project presentations and discussion

**Final project due finals week: Date TBD by class**

## Acknowledgments

This syllabus has benefitted enormously from the labor of others. In particular, I have borrowed and/or adapted ideas, readings, assignments, and conceptual framing from Ryan Cordell's Intro to DH (F20) and BookLab: Print to Programming (S22) courses; from Miriam Posner's Museums in the Digital Age (F20) course; and from Jim McGrath's Digital Archives and Digital Publics (S20) course.