

# DS 2001

## Data Science Programming Practicum: Social Sciences and Humanities

Fall 2021

### What is this class about?

Digital traces of our daily lives are increasingly recorded, aggregated, analyzed, and used to shape our experiences, and large-scale computational methods and resources for understanding human behavior are accessible like never before. These data and methods offer the potential for rich insights into society, while simultaneously introducing new ethical and infrastructural challenges. In this practicum we will (1) practice the skills you learn in DS2000 using applied examples drawn from the social sciences, (2) read about how data science is impacting society, (3) develop an intuition for computational social science, and (4) interrogate the ethical and political ramifications of data science. The practicum will meet once a week, where we'll discuss short readings and practice programming through hands-on tutorials and assignments. Your grade will be based on reading annotations, programming exercises, a project proposal, and a final project and presentation.

### Who is the instructor?

My name is Ryan Gallagher and I'm a Ph.D. candidate in Network Science here at Northeastern. You can call me Ryan and use he/him or they/them pronouns. *Please do not refer to me as Mr. Gallagher*, just Ryan is perfectly fine. I research online communication on social media platforms like Twitter, Facebook, Reddit, and TikTok. I'm particularly interested in how people amplify their voices online to speak against oppression and marginalization. I enjoy teaching and mentoring and I'm excited to work with you this semester!

If you need to get in touch, here's how:

Email	<a href="mailto:gallagher.r@northeastern.edu">gallagher.r@northeastern.edu</a>
Virtual Office Hours	Mondays 2-4pm, or by appointment
Office Hours Link	See syllabus on Canvas for link
Office Hours Passcode	See syllabus on Canvas for link
Course Website	<a href="https://ryanjgallagher.github.io/classes/ds2001_fall2021">https://ryanjgallagher.github.io/classes/ds2001_fall2021</a>
Canvas	<a href="https://northeastern.instructure.com/courses/86786">https://northeastern.instructure.com/courses/86786</a>

### What do I need to purchase for this class?

In short: probably nothing.

### Books and Readings

All of the readings will be available on Canvas. We will read several chapters from *Bit by Bit: Social Research in the Digital Age* by Matthew Salganik, *Data Feminism* by Catherine D'Ignazio and Lauren Klein, and *Algorithms of Oppression: How Search Engines Reinforce Racism* by Safiya Noble. **All of these books are freely available online**, and I highly recommend them if you are interested in any of the topics we cover in class.

*Bit by Bit*: <https://www.bitbybitbook.com/en/1st-ed/preface/>

*Data Feminism*: <https://data-feminism.mitpress.mit.edu/>

*Algorithms of Oppression*: <https://www.jstor.org/stable/j.ctt1pwt9w5>

## Software

***Students who have access to a laptop should bring it to class every day.*** If you do not have a laptop, please get in touch with me as soon as possible so we can find a suitable arrangement.

This class will be taught in the open source programming language Python 3. This course will use the Anaconda distribution of Python, which is available on Windows, macOS, and Linux, and includes many of the code libraries that we will be using in the course. We will spend our first class setting up Anaconda, which you can download at <https://www.anaconda.com/download/>. We will write and run Python code using the (free) text editor Spyder. You are welcome to use a different text editor if you would like, but any instructions or demonstrations I provide will use Spyder. Part way through the course we will introduce the programming environment Jupyter, which is included in the Anaconda distribution.

## How does this class work?

This is a hands-on course that will apply programming to questions in social science. It will consist of in-class programming exercises, assigned readings, class discussions, and a final project. Each class, we will discuss the reading for the week and complete a programming exercise assigned within the scheduled time. Instructors and TAs will be on-hand to go over concepts, answer questions, and lend a hand.

## Practicum Assignments

Learning Python is like learning a foreign language—we'll learn by using it as much as possible in hands-on assignments. Each class you will be given a programming exercise that covers topics you learned in DS2000 (often those from the previous week). **The practicum assignments are designed to be completed within the class time, but the deadline for practicum assignments will be each Friday at 11:59pm EST.** If you are healthy and able, I encourage you to come to class so that you can work on the assignment in real-time, hand it in at the end of class, and not have to worry about submitting anything later.

After class on Wednesday, the practicum assignment will be posted on Canvas so that those who are sick or not able to attend class can still work on it for credit. If COVID-19 or any other circumstance prohibits your ability to complete the assignment, please reach out to me so that we can make alternative arrangements. I do not want you to feel like you need to sacrifice your health for a programming assignment.

## Reading and Discussions

The assigned readings explore real-world applications of data science with a particular relevance to social science, both in their impact on society and on how we study society. You are expected to complete readings for each week prior to the class session where they are listed (e.g. readings listed for September 15th should be completed before class on September 15th), and class sessions will be used to build upon and illustrate class readings. Which is to say, we will not be covering everything in the readings during class, nor will everything we discuss in class be covered in the readings. **All of the readings are available on Canvas or otherwise freely available.** If you have trouble accessing any of the materials, please let me know as soon as possible.

To facilitate discussion, we will be using a group annotation tool called [Hypothesis](#), which will be available directly through Canvas. Every week you will collaboratively annotate the readings with questions, thoughts, connections, ideas, and examples. **For each reading, you must post at least 1 annotation.** If there are multiple readings in one week, then you should do 1 annotation for each one (i.e. 3 total if there are 3 readings). Annotations should demonstrate that you are engaging with the material. For example, an annotation can be any of the following:

- A question about a specific important point in the reading that you didn't understand
- A question about the implications this reading has for some social phenomenon
- A discussion of how the reading relates to readings from prior weeks
- A thoughtful reply to one of your peer's annotations
- An example of how this reading applies to some other social phenomenon
- A critique respectfully explaining why you disagree with a point being made

These are just some examples; here are some other [tips](#) for annotations. An annotation should **not** just be a summary of what was said in the reading: you should be adding to our understanding of the readings, not restating them. Annotations should be posted by the time class starts that week. Avoid annotations that are the same as another student's annotation, and avoid only annotating the beginning of articles. If you choose to reply to another student's annotation, respect their thoughts, give everybody space to talk, and address your comments towards their ideas and not the person themselves.

## Final Project and Presentation

The goal of the final project is to merge your data analysis skills with research questions related to social science. It will also allow you to gain hands-on experience with finding, importing, analyzing, visualizing, and presenting a dataset of your choosing. Through this project you should show that you understand (a) what types of questions are interesting or important to social scientists, (b) what types of questions can be best answered using computational or digital techniques, (c) what types of techniques and evidence are appropriate to best answer your question, and (d) that you can think about how to present your findings and analysis in a reproducible way and in a way that supports, and persuades others of, your (preliminary)

conclusion. The final project consists of a project proposal, Jupyter notebook, and final presentation.

The final project will be graded in terms of three components: a proposal for the project, the analysis and writeup of the project itself, and a final presentation. I will distribute more materials about the expectations of the project as we progress in the course.

### **What happens if I miss a deadline?**

Missed or late assignments create scheduling conflicts and they make it difficult for me to issue grades in a timely manner. However, I realize students are often balancing multiple deadlines and this semester may present some unique challenges. **For that reason, there is flexibility built into the course.**

- Your lowest in-class practicum assignment will be dropped
- Your three lowest reading annotations will each be dropped

In the event that you become seriously ill or are otherwise unable to submit work for a week or more, please reach out so that we can make alternate arrangements. I have no interest in failing students who are unable to complete coursework due to physical, mental, or community health challenges.

### **How is this course affected by the pandemic?**

It is difficult to anticipate how this semester will unfold. At Northeastern, we are fortunate to have more resources to cope with the pandemic than most. Nevertheless, we are all working and learning under extremely stressful circumstances.

I will wear a face covering during class and you should do the same. When you are able, please place distance (ideally at least six feet) between yourself and others; I recognize that this may not always be possible since the university is operating at full capacity. If you are able, please try to eat before or after class so that you are able to keep your mask for the entire duration of the class; I recognize that you all need water to stay hydrated, our class is during lunch time, and that you may not have another opportunity to eat before or after, so use your best judgement to keep yourself and your classmates safe. **If you test positive for COVID-19 or are experiencing symptoms of it**, I expect you to follow the university's protocol and **not come to class**. Please instead follow the university's guidance for getting medical help.

I will not be taking attendance, **so not showing up to class will not directly impact your grade**. If you are sick or having symptoms of COVID-19, I would much rather prefer you keep me and your peers safe by not coming to the classroom. That said, each week we will develop skills and knowledge that build on previous weeks. **If you are healthy and able**, I strongly encourage you to attend class because I will be available directly to help you work on that week's practicum assignment. You may reach out to me with questions if you are not able to attend class, but my capacity to help and explain things will be more limited than working through it with you in

person. Because our classes consist mainly of in-class exercises, they will not be streamed or recorded.

All this said, we will put people first this semester. Course rules and guidelines are helpful, until they are not. I pledge to be flexible and adaptable as things change, and I ask that you do the same. We do not know what this semester will bring, but we will get through it together.

### **How can I communicate with the instructor?**

I will generally respond to emails ([gallagher.r@northeastern.edu](mailto:gallagher.r@northeastern.edu)) and Canvas posts within one business day, e.g. between the hours of 9am–5pm, Monday through Friday. I will try to respond to messages the same day, but if you send a message near the end of the day I will most likely respond to it the following day, and if you email me or post questions on a Friday afternoon or the weekend, I likely will not respond until the following Monday. If you email me, please include “[DS 2001]” at the beginning of your email’s subject heading.

You are encouraged to come to my virtual office hours on Mondays from 2-4pm (<https://northeastern.zoom.us/j/93148874932?pwd=NEtqVUVnS0pvWHhTa1o3U2hIRHdvQT09>). Office hours are designated times that you can always count on me to be available to answer questions and discuss the course material. You do *not* need to email in advance to attend office hours and I am always happy to provide help, including general programming questions you may have. If you have a regularly occurring meeting that conflicts with my office hours and you need to discuss anything regarding the course with me, please email me so that we can make alternative arrangements.

If you come across errors as you run code, post them to the discussion board on Canvas. You may also post questions or comments about the readings or about your final project. I encourage everyone to answer each other's questions, as this is the best way to learn complicated material. Often many people will get the same error or will have similar questions, so check the discussion board for answers before posting your error or question. Keep your comments civil and directed towards not the course material, not a particular person.

### **How will I be graded?**

Practicum assignments will be graded on a scale of 1 (no effort) to 5 (practicum time was used wisely and the submitted code shows significant effort). Reading annotations will be graded on a scale from 1 (little effort) to 3 (the annotation demonstrates the readings were read closely and significant effort has gone into engaging with them). The final project will be graded in terms of three components: the proposal, the project itself, and the final presentation. You will receive more details about the expectations for these assignments as the course progresses.

The course grade will be weighted as follows:

Weekly coding exercises	40%
-------------------------	-----

Reading annotations	20%
Final Project - Proposal	10%
Final Project - Notebook	20%
Final Project - Presentation	10%

*Note:* you will receive a separate grade for DS2000 based on that course's syllabus.

### **How does that translate into a letter grade?**

Final letter grades will be assigned as follows. The cutoffs are firm and there will be no rounding. If your grade score falls exactly on the border of two grades, you will receive the higher one (e.g. if you get a 93, you will get A; if you get a 92.7, you will receive an A-).

A	93-100
A-	90-93
B+	87-90
B	83-87
B-	80-83
C+	77-80
C	73-77
C-	70-73
D+	67-70
D	63-67
D-	60-63
F	< 60

### **What happens if I plagiarize or cheat?**

In this course, as in any course you take, academic integrity is paramount. You are expected to comply with university regulations regarding academic integrity. Please familiarize yourself with Northeastern's policies on academic integrity, and ask questions if you are confused:

[www.northeastern.edu/osccr/academicintegrity/](http://www.northeastern.edu/osccr/academicintegrity/)

Any and all forms of cheating or plagiarism are absolutely unacceptable. This includes turning in work that is not your own, turning in collaborative work without appropriate credit, and copying someone else's code, including code you may find online. If you find code online that helps answer your question, you should adapt it and cite it in a comment by that line of code. Violations of academic integrity will result in failing grades.

This policy is not flexible or negotiable. However I believe that most instances of cheating and plagiarism arise when students become overwhelmed and see no other options. **If you are thinking about cheating because you are overwhelmed by the coursework, please stop what you are doing and contact me immediately.** I would much rather make accommodations that will allow you to complete your assignments ethically than deal with instances of cheating and plagiarism.

### **Are there any other rules or guidelines?**

When you join DS2001, you are joining a learning community dedicated to advancing our knowledge of programming and computational social science. We are all part of the learning community and we will all support each other's learning this semester. Learning will happen individually and collaboratively, with peers and with the instructor.

### **Civility**

I aim to build a classroom climate that is comfortable for all. I expect that students will 1) display respect for all members of the classroom---including the instructor and fellow classmates; 2) pay attention to and participate in class sessions and activities; 3) avoid unnecessary disruption during class time (e.g. having private conversations, using social media, doing work for other classes, sending text messages, etc.); 4) avoid negative language that may unnecessarily exclude members of our campus and classroom; and 5) keep class conversations in confidence. Constructive disagreement and debate are encouraged; derogatory, marginalizing, and threatening comments and actions are not. Civility is a cooperative and evolving process; if you ever feel uncomfortable or excluded in our classroom, please let me know right away so that we can work together to co-create a comfortable and inclusive classroom.

### **Accommodations**

I am happy to discuss any academic accommodations that will help ensure your success in class. This policy applies to students with documented disabilities as well as those whose needs are not formally documented. This includes students with chronic illnesses, those who are food or housing insecure, who have extracurricular commitments that disrupt school time, etc. In short, if there is something I can reasonably do to help you learn, I am happy to work with you to do it. It is helpful when requests for academic accommodations are made in the first week of the semester. However, if a need arises in the middle of the term, do not hesitate to ask for the support you need.

### **Recording and Sharing Course Materials**

Unless otherwise specified, course materials and interactions should be kept in confidence. There may be circumstances where recording or screen capture makes sense, however surreptitious recording is not allowed. Please meet with me if you would like to record or share course materials or interactions so we can create a plan that is transparent and consistent with local laws, policies, and conventions.

### Zoom Etiquette

Students are expected to respectfully attend any Zoom office hours or other meetings, just as you would attend in-person class meetings on campus. This includes arriving sober, dressed, and prepared to engage. I hope students will leave cameras on whenever possible, though I realize the realities of home learning setups may not allow this consistently. I expect that various roommates, siblings, children, pets, etc. will appear on screen from time to time. Do not feel as if you need to leave your camera off because your Zoom space is not completely private or pristine.

### What is the schedule for this course?

I reserve the right to alter topics, readings, and due dates, but we will likely follow the following without any changes. Remember, all readings are on Canvas and that all readings are due on the day for which they are assigned (e.g. readings assigned for September 15th should be read in preparation for September 15th's class).

Date	Topic	Readings	Due
9/8	Introduction to the course and getting started with Python		
9/15	Big Data	<p>Lazer et al. (2014). "The Parable of Google Flu: Traps in Big Data Analysis." <i>Science</i>.</p> <p>boyd &amp; Crawford. (2011). "Critical Questions for Big Data." <i>Information, Communication, &amp; Society</i>.</p> <p>Foucault Welles. (2014). "On Minorities and Outliers: The Case for Making Big Data Small." <i>Big Data &amp; Society</i>.</p>	
9/22	Data Visualization	<p>D'Ignazio &amp; Klein. (2020). "Chapter 3: On Rational, Scientific, Objective Viewpoints from Mythical, Imaginary, Impossible Standpoints" in <i>Data Feminism</i>.  <a href="https://data-feminism.mitpress.mit.edu/pub/5evfe9yd/release/5">https://data-feminism.mitpress.mit.edu/pub/5evfe9yd/release/5</a></p> <p>Yau. (2018). "Visualizing Incomplete and Missing</p>	



		Data.” <a href="https://flowingdata.com/2018/01/30/visualizing-incomplete-and-missing-data/">https://flowingdata.com/2018/01/30/visualizing-incomplete-and-missing-data/</a>	
9/29	Data Ethics	Salganik. (2018). “Chapter 6: Ethics” in <i>Bit by Bit: Social Research in the Digital Age</i> . <a href="https://www.bitbybitbook.com/en/1st-ed/ethics/">https://www.bitbybitbook.com/en/1st-ed/ethics/</a>	
10/6	Categorization	D’Ignazio & Klein. (2020). “Chapter 4: ‘What Gets Counted Counts’” in <i>Data Feminism</i> . <a href="https://data-feminism.mitpress.mit.edu/pub/h1w0nbqp/release/3">https://data-feminism.mitpress.mit.edu/pub/h1w0nbqp/release/3</a>  Brown (2020). “The changing categories the U.S. census has used to measure race” from Pew Research Center. <a href="https://www.pewresearch.org/fact-tank/2020/02/25/the-changing-categories-the-u-s-has-used-to-measure-race/">https://www.pewresearch.org/fact-tank/2020/02/25/the-changing-categories-the-u-s-has-used-to-measure-race/</a>	
10/13	Location Data	Thompson & Warzel. (2019). “Twelve Million Phones, One Dataset, Zero Privacy” in <i>The New York Times</i> . <a href="https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html">https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html</a> (Recommended to view online for interactive graphics)  Boorstein et al. (2021). “Top U.S. Catholic Church official resigns after cellphone data used to track him on Grindr and to gay bars” in <i>The Washington Post</i> . <a href="https://www.washingtonpost.com/religion/2021/07/20/bishop-misconduct-resign-burrill/">https://www.washingtonpost.com/religion/2021/07/20/bishop-misconduct-resign-burrill/</a>  Wang. (2021). “For the U.S. Census, Keeping Your Data Anonymous and Useful is a Tricky Balance” in <i>NPR</i> . <a href="https://www.npr.org/2021/05/19/993247101/for-the-u-s-census-keeping-your-data-anonymous-and-useful-is-a-tricky-balance">https://www.npr.org/2021/05/19/993247101/for-the-u-s-census-keeping-your-data-anonymous-and-useful-is-a-tricky-balance</a>	
10/20	Algorithmic Oppression	Noble. (2018). “Introduction: The Power of Algorithms” in <i>Algorithms of Oppression: How Search Engines Reinforce Racism</i> .  Angwin et al. (2016). “Machine Bias” in	<b>Project:</b> Team proposals (if working in a team)

		<p><i>ProPublica</i>.  <a href="https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing">https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing</a></p>	
10/27	Algorithm Audits	<p>Keegan. (2021). "Facebook Got Rid of Racial Ad Categories. Or Did It?" in <i>The Markup</i>.  <a href="https://themarkup.org/citizen-browser/2021/07/09/facebook-got-rid-of-racial-ad-categories-or-did-it">https://themarkup.org/citizen-browser/2021/07/09/facebook-got-rid-of-racial-ad-categories-or-did-it</a></p> <p>Johnson. (2021). "Twitter's Photo Crop Algorithm Favors White Faces and Women" in <i>Wired</i>.  <a href="https://www.wired.com/story/twitter-photo-crop-algorithm-favors-white-faces-women/">https://www.wired.com/story/twitter-photo-crop-algorithm-favors-white-faces-women/</a></p> <p>Wall Street Journal. (2021). "Investigation: How Tiktok's Algorithm Figures Out Your Deepest Desires" in <i>The Wall Street Journal</i>.  <a href="https://www.wsj.com/video/series/inside-tiktoks-highly-secretive-algorithm/investigation-how-tiktok-algorithm-figures-out-your-deepest-desires/6C0C2040-FF25-4827-8528-2BD6612E3796">https://www.wsj.com/video/series/inside-tiktoks-highly-secretive-algorithm/investigation-how-tiktok-algorithm-figures-out-your-deepest-desires/6C0C2040-FF25-4827-8528-2BD6612E3796</a> (Video)</p>	<b>Project:</b> Project proposal (1 page)
11/3	Conducting Computational Social Science	<p>Salganik. (2018). "Chapter 2: Observing Behavior" in <i>Bit by Bit: Social Research in the Digital Age</i>. <b>Only need to read Sections 2.4-2.6.</b>  <a href="https://www.bitbybitbook.com/en/1st-ed/observing-behavior/strategies/">https://www.bitbybitbook.com/en/1st-ed/observing-behavior/strategies/</a></p> <p>Hoffman et al. (2021). "Integrating Explanation and Prediction in Computational Social Science." <i>Nature</i>.</p>	
11/10	Replication	<p>Yong. (2021). "Replication Studies: Bad copy." <i>Nature</i>.  <a href="https://www.nature.com/articles/485298a">https://www.nature.com/articles/485298a</a></p> <p>King (1995). "Replication, Replication." <i>Political Science and Politics</i>.</p> <p>Anonymous. (2021). "Evidence of Fraud in an Influential Experiment about Dishonesty."  <a href="https://datacolada.org/98">https://datacolada.org/98</a></p>	
11/17	Social Media Data	<p>Freelon. (2018). "Computational Research in the Post-API age." <i>Political Communication</i>.</p>	

		Pasquetto et al. (2020). "Tackling Misinformation: What Researchers Could Do with Social Media Data." <i>Harvard Kennedy School Misinformation Review</i> .	
11/24	Thanksgiving Break		
12/1	Social Network Data	Robins. (2015). "The difference with social networks research" in <i>Doing Social Network Research: Network-based Research Design for Social Scientists</i> .	
12/8	The Pandemic	Buckee et al. (2021). "Thinking Clearly About Social Aspects of Infectious Disease Transmission." <i>Nature</i> .	<b>Project:</b> Jupyter notebook writing methods, analysis, and interpretation