

DS2001 | Spring 2019
Programming with Data: Social Sciences and Humanities Practicum

Instructor

Sarah Shugars

Pronouns: she/her/hers

Email: shugars.s@northeastern.edu

Twitter: @Shugars

Office hours: Wednesdays, 1:30-3pm or by appointment

TA

Magnus Frennberg

Office hours: Thursday 6-9pm. KA202

(See course website for all TA office hours)

Course Description

Meetings: Wednesdays, 11:45am-1:25pm, West Village H 210

Piazza: <https://piazza.com/northeastern/spring2019/ds2000>

Course website: <https://course.ccs.neu.edu/ds2000/index.html>

Readings

All of the readings are available online and the links are provided in the syllabus. If you have trouble accessing any of the readings, please let me know as soon as possible.

Overview

Large-scale data and computationally complex methods for understanding human behavior are accessible like never before with the emergence of vast archives of passive data collection, online experimentation, and innovative uses of simulation. Digital traces of our daily lives are increasingly recorded, aggregated, analyzed, and used to shape our future experience. 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 the DS2000 lecture using applied examples drawn from the social sciences and humanities, and (2) will read about how these methods are impacting the social science and humanities disciplines, to encourage a computational social science imagination.

The practicum will meet once a week for 1 hour 40 minutes. Class time will be a combination of discussion and reactions to short readings, and hands-on tutorials that practice the skills you learned in lecture. Your grade will be based on your reading reflections, programming exercises, a project proposal, and a final project and presentation.

Course Requirements

Students who have access to a laptop should bring it to class every day. This workshop will be taught in the open source programming language Python 3. The room

scheduled for practicum has desktops with Python 3 installed on them for those who do not have access to a laptop.

We will begin by writing Python code in the open source text editor IDLE, Python's own Integrated Development Environment (IDE). An IDE combines the Python interpreter with an editor for your code, which makes it easy to work on your code and test/run your software.

Download Python 3.7.1 from <https://www.python.org/downloads/release/python-371/>. It's available for Windows, Mac OSX, and Linux. Once installed, click on IDLE to open it up. You can use Python's interactive environment, or you can write and save a file with a .py extension. Part way through the course we will introduce the programming environment Jupyter.

Course Schedule

Date	Week	Topic	Reading
Jan. 9	1	Introduction and ethics in data science	
Jan. 16	2	Variables and functions Exercise: practice with functions	<ul style="list-style-type: none">• Virginia Eubanks (2018). <i>Automating inequality: How high-tech tools profile, police, and punish the poor</i>: "Introduction: Red Flags." St. Martin's Press. https://us.macmillan.com/excerpt?isbn=9781250074317• Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner (2016). <i>Machine Bias</i>, ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
Jan. 23	3	Conditionals and Boolean expressions Exercise: more practice with functions	<ul style="list-style-type: none">• Kelsey Campbell (2018), <i>Plot Me Genderfluid</i>, Gayta Science. Blog post. http://www.gaytascience.com/plot-me-genderfluid/• Brooke Foucault Welles (2014). <i>On minorities and outliers: The case for making Big Data small</i>. Big Data & Society. https://journals.sagepub.com/doi/pdf/10.1177/2053951714540613
Jan. 30	4	Functions and Strings Exercise: functions for string/list manipulation	<ul style="list-style-type: none">• Matt Salganik (2017). Introduction. From <i>Bit by Bit: Social Research in the Digital Age</i>. Princeton, NJ: Princeton University Press. https://www.bitbybitbook.com/en/1sted/introduction/

Feb. 6	5	Iteration, loops, and lists Exercise: loop over data, aggregate results	<ul style="list-style-type: none"> • Matt Salganik (2017). Chapter 2: Observing Behavior, Parts 2.1 – 2.3. https://www.bitbybitbook.com/en/1st-ed/observing-behavior/
Feb. 13	6	Strings and file processing Exercise: read and process a file of data, output results	<ul style="list-style-type: none"> • Matt Salganik (2017). Chapter 2: Observing Behavior, Parts 2.4 – 2.5. https://www.bitbybitbook.com/en/1st-ed/observing-behavior/
Feb. 20	7	Dictionaries Exercise: read a file, categorize, output results	<ul style="list-style-type: none"> • Matt Salganik (2017). Chapter 3: Asking Questions, Parts 3.1-3.4; 3.7 up until “Mathematical Notes. https://www.bitbybitbook.com/en/1st-ed/asking-questions/
Feb. 27	8	Introduction to Notebooks & Mid-semester review Exercise: Hello, notebook!	<ul style="list-style-type: none"> • Matt Salganik (2017). Chapter 6: Ethics. Parts 6.1 – 6.3. https://www.bitbybitbook.com/en/1st-ed/ethics/
Mar. 6	SPRING BREAK		
Mar. 13	9	Object oriented programming Project pitches/discussion Exercise: design and implement a data object	<p>No reading.</p> <p>DUE: Submit 1-page project proposal to shugars.s@northeastern.edu by 5 pm on Friday, March 15.</p> <p>*** Each person must submit their own proposal document. Teams must also submit 1-page listing team members and describing responsibilities.</p>
Mar. 20	10	Matplotlib and data visualization Exercise: analyze and visualizes data	<ul style="list-style-type: none"> • Kieran Healy and James Moody (2014). Data Visualization in Sociology. <i>American Review of Sociology</i>. 40: 105-28. Available at https://kieranhealy.org/files/papers/data-visualization.pdf • Catherine D’Ignazio (2015). <i>What would feminist data visualization look like?</i> Available on the MIT Center for Civic Media’s blog at

			https://civic.mit.edu/2015/12/01/feminist-data-visualization/
Mar. 27	11	APIs Exercise: API activity	No reading, work on your projects
Apr. 3	12	Projects	No reading. Projects due April 5
Apr. 10	13	Project presentations Presentations will take place in Lecture or Practicum this week	

Course Structure

This is a hands-on course that will introduce students to the practical application of computational methods to questions important to humanists and social scientists. In this course you will learn what types of questions humanists and social scientists are answering using computational techniques, what types of evidence these techniques produce, how you can use this evidence to draw conclusions about the social world, and you will learn the programming language Python and the developing environment Jupyter. To facilitate learning both programming and domain knowledge, the course will consist of applied exercises aimed at getting you processing and analyzing cultural material via Python, as well as discussions about assigned readings that explore a practical question or issue about the social sciences or humanities. It is important that you complete the readings before each class and come prepared to discuss the material. During these discussions there will be space to critique the material and these methods. It is important that we respect one another's thoughts, give everybody the space to talk, and address our comments at the ideas and not the person.

Grading

The practicum grade will be based on:

Factor	Number	Weight
Coding exercises	weekly	25%
Reading reflections	6	25%
Class + Piazza discussion	weekly	25%
Final Project + Presentation	1	25%
	TOTAL:	100%

Assignments

Each week will consist of in-class activities that should be submitted at the end of the practicum. See the first [Practicum Handout](#) for how to test and submit practicum assignments.

Additionally, over the course of the semester, you must post at least 6 reading reflections to Piazza. These reflections should be a 250-500 words and should be posted before class (by 11am on Wednesday) the week that reading is assigned. You are welcome to post your reflections as a new thread or as respectful commentary on someone else's reflection. Note that readings are assigned for 8 weeks of the class, so choose your 6 weeks of reading reflection wisely!

Attendance and Participation and Exercises

In each week we will learn skills and develop knowledge that build on previous skills learned, so it is important to attend every class. Learning Python is like learning a foreign language. The best way to learn it is to use it all the time. While I will not keep attendance, each class period will include in-class exercises. If you miss a class you will lose the opportunity to get those points. If you know you are going to miss a class you should notify me at least four days in advance and we can arrange a way to make up the material. If you notify me fewer than four days in advance you might not be able to make up those points.

Reading Assignments

This course will teach you Python, but I also want you to learn how to apply it to social science and humanities questions. I will thus assign short readings each week to get you familiar with this space, and how computational methods about being used in these disciplines. In class I will have you answer brief questions about the readings. This will contribute to your in-class exercise points for the day. Most of the readings will come from Matt Salganik's book *Bit by Bit*.

You can read it online here: <https://www.bitbybitbook.com/en/1st-ed/preface/>. I also encourage you to buy it, and read the whole book when you have a chance: <https://www.bitbybitbook.com/>

Project

The goal of the final project is to creatively combine the techniques you learned in the course to explore a question related to the humanities or social sciences that has either not been addressed before, or explore an old question in new ways.

Through this project you should show that you understand (a) what types of questions are interesting or important to humanists and/or 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 project *proposal*, (due by 5 pm on Friday, March 15) will be a 1-page document detailing a preliminary plan for your final project. You should include the following in your project proposal:

1. Identify a general question related to the humanities or the social sciences that you plan to address in your final project. You should outline why this is an interesting or important question and describe why computational methods are necessary and/or helpful in exploring this question. If possible, explain how others have answered/attempted to answer this question using different methods.
2. Identify the data or collection of material you will use to explore this question, and briefly describe why the data/material is appropriate. Additionally describe how you will collect the data/material and whether or not it will need to be cleaned prior to analysis. If possible, import your data and provide a glimpse of its format.
3. Describe the techniques you expect, or would like, to use to analyze the data/material and explore your question. Why these techniques and not others? What kind of evidence will these techniques produce, and how will this help you answer the question and persuade others of your answer? If you already have some preliminary analyses, include these as well.
4. Briefly discuss any data visualization or interpretive techniques you will, or would like to, use to present your findings and convince others of your interpretation.

If working in a team

Every member of the team should submit their **own** 1-page proposal articulating their vision of the project. Please note the names of your team members on your proposal. Additionally, each team should **submit up to one additional page** detailing your planned division of labor and a shared contract about your collective expectations. This page only needs to be submitted once per team. I encourage you to revisit this document well before the final project is due, to check-in with each about whether the shared contract is working. The best teams will include people from different disciplinary backgrounds so you can leverage each other's specialized knowledge.

Questions? Discussion Board, Office Hours, and Email

If you come across errors as you run code post them to the discussion board on Piazza (start a new thread for new errors). 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. This is not the comments section on YouTube, so keep your comments respectful. Disrespect will absolutely not be tolerated.

You are also encouraged to come to my office hours. Please feel free to email me at shugars.s@northeastern.edu if you would like to meet outside of scheduled office hours

or need to inform me of a planned absence.

I will respond to emails and Piazza posts within one business day; eg, between the hours of 9am – 5pm, Monday through Friday. I will typically respond to messages the same day, but if you send a message near the end of the day I will most likely respond the next morning, and if you email me or post questions on a Friday afternoon or a weekend, I may not respond until the following Monday.

Statement of Non-Discrimination

As the instructor of the course, I am committed to maintaining a positive learning environment based upon communication, mutual learning, and respect. Any suggestions as to how to further such a positive and open environment in this class will be appreciated and given serious consideration. The university does not discriminate on the basis of race, sex, age, disability, religion, sexual orientation, color, or national origin. If you are a person with a disability and anticipate needing any type of accommodation in order to participate in this class, please advise me and make appropriate arrangement with Disability Resource Center (617) 373-4428. If you need accommodation for any topic not listed here, please let me know.

Consulting Resources

I encourage you to take advantage of the Digital Scholarship Group at Northeastern. They offer a wealth of services – including digital data collections – and can offer advise on collecting and structuring digital data. They also offer a quiet space to work.

Note on Plagiarism

I encourage you to work together to help each other review the readings and to learn the coding. However, *all written and coding work must be your own*. I take academic honesty seriously, and you should too. The Northeastern University Policy on Academic Integrity can be found at: <http://www.northeastern.edu/osccr/academichonesty.html>)

Since students in this course are often encouraged to work in teams, some specific remarks are in order:

It is not considered cheating if you:

- Work together on homework assignments, as long as you each work out and submit your own final answers
- Get help from professors, physics workshop, tutors, etc. on the homework assignments
- Work together on preparing for quizzes and exams

It is considered cheating if you:

- Submit work done by others (without your participation) as your own
- Copy work on quizzes and exams

Final thoughts

If you are unsure about anything related to the rules guiding this course, please ask!