INSH 1500, Summer II

Digital Methods for Social Sciences and Humanities

Instructor
Jeffrey Sternberg
Email: sternberg.je@northeastern.edu
Twitter: @JeffSternberg
Office hours: TBD (or by appointment), Online via Zoom

Course Description

Meetings: Asynchronous Online Blackboard Instruction, Videos Available Mondays, Wednesdays, and Fridays

Readings:

All of the readings are available online or on Blackboard. I provide the link to most of the readings in the syllabus, and most of the readings are also available on Blackboard. If you are having problems accessing these readings let me know as soon as possible.

In addition to readings from the social sciences and humanities, we will be using two main Python textbooks, both are under an open license and are available online.

- Charles R. Severance, [Python for Everybody: Exploring Data Using Python 3](#).
- Jean Mark Gawron, [Python for Social Science](#).

Overview:

Increasingly, humanity's cultural material is being captured and stored in the form of electronic text, images, and other material. From historical documents, literature and poems, diaries, political speeches, and government documents, to emails, text messages, and social media, students from the humanities and social sciences now have access to immense amounts of rich, and diverse, digitized cultural material. Scholars and practitioners in industry, non-profits, and government, are increasingly using computational methods to analyze this new material in order to ask, and answer, a diverse array of questions about the social world: Does social media reflect public political opinion, or drive it? What determines trust in online communities? Do taxi drivers work more or fewer hours when it is raining? What types of blog posts get censored in China and why? Are diurnal and seasonal mood cycles cross-cultural? What was the form of cultural and institutional change through the "civilizing process" in England between the 16th and 20th centuries? What is the life cycle of a literary genre? What was the geographic imagination of Civil War era literature? What are textual allusions in Classical Latin poetry? Can the FBI really analyze 650,000 emails in 3 days? (Spoiler: Yes, they can!)

Digitized texts, images, and other material, and computational methods used to access that material, are impacting virtually every field within the humanities and social sciences. This is opening up exciting and creative possibilities for any student who wants to explore questions about the social world around them. In this course we will explore how digital material and digital methods are impacting the humanities and social sciences and we will learn practical skills to collect, store, and analyze digitized cultural material.

This course is NOT a computer science course. It is not even a course on computer programming and no prior programming experience is required or assumed. It is primarily a humanities and social science course, with an eye toward digital technologies. We will not have computers analyze data or cultural material for us. Instead, we will harness the superior ability for computers to count and extract patterns from cultural material, and use this output to enhance our own critical thinking and interpretive analyses. To implement these methods we will use the open source (and free!) programming language Python and the Jupyter platform.

Specific skills covered include collecting digitized data, text analysis, network analysis, and geographic analysis. The ultimate goal is to encourage you to think about novel and creative ways you can apply these techniques to your own area of study. By the end of the course you will have a better understanding of the range of ways computers can help you answer questions, what kind of evidence the different techniques produce, how this evidence can be used to help you better understand the social world, and how to use this evidence to persuade others of your interpretation.

The skills required to ask and answer a question about the social world using computational methods, and present the findings in a concise and persuasive way, are widely used in academia, industry, non-profit organizations, think tanks, and government. This course will facilitate career paths in all of these domains.

This course is divided into three sections:
1. Introduction to Computational Thinking and Computational Tools
2. Collecting, Storing, and Visualizing Digital Data
3. Overview of Subfields: Text Analysis, Network Analysis, and Geographic Analysis

The course will meet virtually each week, with the Exercise Solution and Week Agenda Lectures posted on Mondays, and In Class Coding Tutorial Videos posted before Wednesday. Reading and Discussion Lecture Videos will be posted on Fridays. In terms of Assignments, Reading Responses will be due Thursday by 8pm, and Programming Exercises and In-Class Tutorial Notebooks will be due on Sunday by 12pm the week they are due. Class time will be a combination of lectures, discussion, and hands-on tutorials. Students' grades will be based on class participation, reading assignments, programming exercises, a coding Check-In, and a project proposal.

Course Requirements

Technology Requirements

Students must have access to a laptop, and you must bring it to class every day. If you do not have a laptop contact me and we can try to work something out.

This workshop will be taught in the open source programming language Python and the programming environment Jupyter. **Participants should install** Anaconda for Python 3.7 using the 64-bit graphical installer **on their laptops prior to the first day of class**. Anaconda includes Python, the necessary Python packages, and Jupyter.

Grading and Assignments

10% Attendance and Participation:
      5% - Attendance (Watching the Video Tutorials & Lectures)
      5% - Active participation in class discussion and tutorials (Turning in Completed In-Class Tutorial Notebooks)

25% Reading assignments

25% Programming exercises

15% Peer-reviewed Check-In
      10% - Check-In Notebook Submission
      5% - Peer Review Discussion Post

25% Research Project Proposal:
      10% - Initial Research Proposal Idea Memo
      5 % - Responses to Peers' Project Memos
      10% - Final Research Proposal

Course Schedule

June 29 - Week 1: Course Introduction, Overview, and Introduction to Computational Thinking

July 6 - Week 2: Complexity, Computation, Data & Programming Fundamentals

July 13 - Week 3: Operationalizing and Measurement

July 20 - Week 4: Check-In Assignment & an Overview of Subfield Techniques
- Check-In Assignment Due
- Check-In Peer Feedback Due

July 27 - Week 5: Introduction to Basic Statistics with Python Data Structures

August 3 - Week 6: Collecting and Cleaning Data
- Initial Project Proposal Memos Due
- Peer Feedback on Proposal Memos Due

August 10 - <u>Week 7: Data Visualization and Computational Subfields</u>

August 13 - Last Day of Class
August 18 - Final Project Proposals Due by 8pm

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 through the applied use of these techniques. To facilitate learning both programming and domain knowledge, the course will consist of practical tutorials 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 using these techniques.

Assignments

You are allowed to skip two (2) of the reading assignments or programming exercises before your grade is affected. (Note this is two total, not two for each category). If you complete all the reading assignments, or programming exercises, your lowest grade on either will be dropped for the corresponding category.

The in-class check-in is meant to make sure you are learning the material (and that I am sufficiently teaching it to you!) and help you build toward your final project proposal, and will allow you to work through any challenges you have in a supportive environment. In lieu of a final exam, you will be required to ask and explore a question relevant to the humanities and/or social sciences using some form of digitized data or material and computational techniques. The final project proposal is designed to encourage you to creatively combine the knowledge and skills built through the semester to explore a question about the social world that has not yet been answered, or to explore an old question in a new way. The form of the final project proposal is a Jupyter notebook. I will hand out a detailed rubric closer to the due date for the check-in and the project proposal.

Attendance and Participation

In each week we will learn skills and develop knowledge that build on previous skills learned, so it is important to watch every lecture, in class tutorial, and exercise solution video. I will track attendance by checking whether you have been watching the video lectures and tutorials posted **cumulatively at the end of the semester**. I will track participation on the In-Class Tutorials by requiring you to submit the accompanying Jupyter Notebook weekly filled out with the code included in the Tutorial Lecture Videos **due Sundays by 8pm**. Learning Python is like learning a foreign language. The best way to learn it is to use it all the time. To encourage the continual use

of the skills you are learning there will be short assignments (reading assignments, programming exercises, and tutorial notebooks) due over the course of each week to make sure you are progressing overtime, as opposed to trying to complete all the assignments at once. If you know you are not going to be able to complete an assignment on time, you should notify me at least one day in advance and we can come up with a reasonable submission extension. I understand that Summer Courses are abbreviated and condensed, with your weekly schedules made unstable by Quarantine and remote asynchronous course delivery, and I am willing to make reasonable accommodations. I just ask you let me know in advance and not after you miss a deadline.

Reading Assignments

In this course humanities and social science questions are central, and computational methods are used to answer those questions. There will be assigned readings that discuss either the theory behind these methods, or their practical application to real-world questions. Every week, you will be asked to submit a short (two-paragraphs to one-page) reading response to that week's material in the form of a discussion board post, due **Thursdays by 8pm**. I will post questions to Blackboard that you should cover in your response. These short responses will help you understand and evaluate the applied use of these methods, they will help you get used to writing and talking about computational methods and the different types of evidence these methods produce, and they will help frame the structure of Reading Lectures posted Friday Afternoons, my lectures acting to summarize and react to your reading responses. I encourage everyone to read each other's responses to help you unpack the readings and substantive content at the end of each week.

Programming Exercises

You will also be required to complete programming exercises most weeks. These will typically be in the form of Jupyter notebooks. As with learning any foreign language, the best way to learn Python is to write a lot of code. The programming assignments are meant to give you practice writing code, and they will help me gauge the speed at which we're going through the programming material.

Mid-Term Check-In

There will be one, peer-reviewed, mid-term check-in, on **Wednesday, July 22**. The check-in is meant to ensure that you are understanding the material (and that I am successfully teaching the material!), and to help you work toward your final project.

For the check-in **(Wednesday, July 22)**, you will be asked to upload a short programming assignment to blackboard, in the form of a Jupyter notebook, with comments throughout explaining your code snippet by snippet, You will then be assigned to peer review a peer's notebook. You will write a discussion post that 1) will describe your thought process and why you did what you did to solve the problem, 2) explain how your peer did it differently and the benefits/drawbacks of each approach, and 3) come up with two substantive questions that might be addressed by the technique(s) mentioned in the assignment. By the end of the week **(Sunday, July 26)** you will be required to submit the discussion post to blackboard for grading.

Project Proposal

Being that this is an abbreviated Summer Course, and it takes more time to get proficient at coding than the course allows, we will be stopping just short of completing a final project; however, the final for this course will be for you to write, design, and propose a computational research project that you could potentially carry out in the future. The goal of conceiving of a final project is to creatively combine the techniques you learned in the course to preliminarily 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. Your final project proposal may also explore putting together an original dataset, and thinking through research questions you could ask of the dataset and techniques that could be applied to it. Note that this is an introductory course, and I will present a lot of material throughout the semester, so your final project proposal will by necessity be only a preliminary exploration of a research question.

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 conclusion.

You will submit an initial two-to-three paragraph memo brainstorming ideas for your final project proposal (**August 5**) to blackboard. This two-to-three paragraph memo should include 2-3 project ideas, a paragraph per idea, that lists potential research questions, possible data sources and types, potential analysis techniques, and the project's broader potential relevance. You will then review all of your classmate's project memos on the discussion board, evaluating the project ideas on the appropriateness of the technique to the chosen question. The goal here is not to grade your classmate (I will be giving the grade), but to practice evaluating the application of computational techniques to real-world questions in order to improve your own project and help your peer improve theirs. By **August 9,** you will be required to submit your feedback on each of your classmate's project memos, posting comments to each project memo post on the discussion board.

The final project proposal will be a Jupyter notebook detailing a preliminary plan for a final Project, hopefully incorporating the feedback you received from your classmates on your project idea memo. **The Final Project Proposal will be due on August 18 by 12pm**. 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.

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.

5. Briefly discuss any obstacles you think you might hit when carrying out your project, any project or data limitations, and hypotheses about what you think you might find.

Questions? Discussion Board, Office Hours, and Email

If you come across errors as you run code that you can not solve through Googling, post them to the discussion board on Blackboard (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 which I would like to be a general drop-in zoom conference weekly where multiple people can attend at once, ask questions, and see how to approach different coding problems, bugs, and errors in each other's respective assignments. I also will be available for individual office hour appointments by request, just send me an email to set up a time. Email should be used mainly for quick logistical questions, scheduling office hour appointments, or if you need to inform me of a planned absence. I will get back to emails within 16 working-hours, so plan ahead.

Consulting Resources

I encourage you to take advantage of tutoring services offered by Northeastern. They can help you with assignments and with your final project. For information on these services visit http://www.cps.neu.edu/student-resources/tutoring-services.php.

Note on Plagiarism

I encourage you to work together to help each other review the readings and to learn the coding. However, all written work must be your own. I take academic honesty seriously, and you should too.

For more information on your rights and responsibilities as a student see:
http://www.northeastern.edu/osccr/code-of-student-conduct/

Readings and Schedule

Note: Text rendered in italics indicate programming lessons

Part 1: Introduction to Computational Thinking and Computational Tools

In this part of the course we will do two things: 1) Think through how social scientists and humanities scholars study society, and how computation and technical tools are changing these fields, and 2) we will introduce some Python basics. The computational tools learned in Part 1 will provide the foundation that you will use later in the course, and hopefully into the future.

Week 1: Course Introduction and Overview, and an Introduction to Computational Thinking

This week we will focus mostly on the humanities and social sciences, thinking about how people in these fields study society and how digital technologies are changing these fields. We will also begin to get comfortable with what Python is, what Jupyter Notebooks are, and how we can access this software.

Monday, June 29: First day of class
- Course introduction and overview
- Course expectations
- Learning goals

How do we study society?
- The social science and humanities approach to inquiry
- *Technology overview*
  - *Installation check*
  - *Jupyter notebooks*
  - *File structures*

There will be Numbers
- How are digital data and technologies changing the way we study society?
- *Explore Anaconda*
  - *Jupyter*
  - *Spyder IDE*
  - *Interpreter*
- *Python basics: variables, expressions, and statements*
- *User input, choosing variable names, and debugging*

In-Class Tutorials
00-InstallationCheck.ipynb
01.00-VariablesAndConditionals_InClassTutorial.ipynb

Exercises Due
00.01_InstallationCheck_Exercises.ipynb
01.02-VariablesAndConditionals_Exercises.ipynb

Required Readings:
- Matt Salganik (2017). "Introduction" from Bit by Bit: Social Research in the Digital Age. Princeton, NJ: Princeton University Press.
- Andrew Piper (2016). "There Will be Numbers." Journal of Cultural Analytics.
- Python for Everybody, Chapter 1 and Chapter 2

Week 2: Complexity, Computation, Data & Programming Fundamentals

This week we'll dig deeper into Python basics, but we'll pair this with thinking more deeply about why, as social scientists and humanities scholars, we might want to learn programming and toward what end. You will learn important programming concepts that you will use throughout the rest of this course, so this is a crucial week. Don't worry if not everything makes sense right now. We'll continue to serve these vegetables throughout the rest of the course.

Monday, July 6:
- Data analysis and the goals of inquiry
- Why should you learn to program?
- Thinking with data: Different Approaches
- Digital Data Ethics Module
- *Conditionals and functions*
- *Iteration*

In-Class Tutorials
02.00_FunctionsAndIterations_InClassTutorial.ipynb

Exercises Due
02.03_FunctionsAndIterations_Exercises.ipynb

Required Readings:
- Lev Manovich (2016). The Science of Culture? Social Computing, Digital Humanities and Cultural Analytics. Journal of Cultural Analytics.
- Torching the Modern-Day Library of Alexandria, James Somers, *The Atlantic*, 2017-04-20.
- Gitelman *"Raw Data" Is an Oxymoron*, Introduction
- Python for Everybody, Chapters 3, 4 and 5

Week 3: Operationalizing and Measurement

Social scientists and humanities scholars deal with complex, unsettled, and changing concepts. Computation, on the other hand – as we just learned – requires precision and absolute clarity. How do we reconcile these two cultures? How might we turn theoretically interesting concepts into data, and use this to build new and interesting theory? We'll explore these issues in this week's readings, while continuing to dig deeper into Python by learning two additional data types: strings and lists.

Monday, July 13
- Strings and the Digital Humanities
- How might we turn theoretically interesting concepts into data?
- *The string data type in Python*
- Connecting concepts and theory in empirical research
- *Python: Lists*

In-Class Tutorials
03.00_StringsAndLists_InClassTutorial.ipynb

Exercises Due
None

Readings:
- Franco Moretti (2013). "Operationalizing". New Left Review.
- Tressie McMillan Cottom, "More Scale, More Questions: Observations from Sociology"
- Robert K. Merton (1948). "The Bearing of Empirical Research Upon the Development of Social Theory." American Sociological Review 13(5): 505-515.
- Python for Everybody, Chapter 6 and 8

Week 4: Check-In Assignment and an Overview of Subfield Techniques
Time for a check-in!

Monday, July 20

Take a break from learning new programming material and work on your assignment. Read up on three subfields in Digital Methods and the types of analysis each offer. I will introduce some of these methods to you in programming tutorials in Week 7.

Wednesday, July 22: Check-In, Upload in your completed Notebook to the Discussion Board

Readings:
- Text Analysis and Causality

- o Witmore "Text: A Massively Addressable Object"
- o Drury & Abinesh "Tomorrow's News Today" SAGE Ocean Blog Post on Sentiment Analysis and Topic Modelling
- Network Analysis
  - o Weingart "Demystifying Networks, Parts I & II."
- Geo-Spatial Analysis
  - o Blevins "Space, Nation, and the Triumph of Region."

Part 2: Collecting, Exploring, and Visualizing Digital Data

In part two of this course we will get our hands dirty by exploring real data, collecting web-based data using an API, and by doing some basic visualizations. This is where things get interesting!

Week 5: Introduction to Basic Statistics with Python Data Structures

With any large amount of data or cultural material you will likely want to summarize it in an interesting or useful way. This week we'll walk through how to do so using Python. We'll also walk through some basic statistics, which will not only help you in your own research, but will help you intelligently read all the numbers thrown at you in the news. Use this knowledge to cut straight through all the fake news out there!

Monday, July 27
- How to read a quantitative research paper
- Basics of statistics
- Populations and samples
- *Python Pandas*

In-Class Tutorials
04.00_DictionariesAndDataframes_InClassTutorial.ipynb

Exercises Due
04.02_DictionariesAndDatafames_Exercises.ipynb

Readings:
- Jeremy Freese, Brian Powell, and Lala Carr Steelman (1999). "Rebel Without Cause or Effect: Birth Order and Social Attitudes." American Sociological Review 64: 207-231.
- Greta Krippner (2000). "How to Read a (Quantitative) Journal Article." American Sociological Association.
- Python for Social Science - Sections 6.5 and 6.6.

Week 6: Collecting and Cleaning Data

Data are everywhere, especially on the web. Luckily, web designers often make it easy to collect that data by building APIs. On Monday we'll learn about APIs and how to access them using Python. On Wednesday we'll think about what to do with that data once you've collected it.

Monday, August 3
- Ethics of collecting found data
- Gathering data from the web
- *Accessing Web APIs using Python*
- *Intro to XML and JSON*
- Introduction to data cleaning
- Importance of reproducibility
- "Other People's Data"
- *Regular expressions*

Initial Project Proposal Due Wednesday, August 5
Peer Comments Due Friday, August 7

In-Class Tutorials
05.00_XML-JSON-APIS_InClassTutorial.ipynb
07.00_TextAnalysis_InClassTutorial.ipynb

Exercises Due
05.02_XML-JSON-APIS-ProgrammingExercise.ipynb

Readings:
- Danyel Fisher, David W. McDonald, Andrew L. Brooks, and Elizabeth F. Churchill (2010). "Terms of Service, Ethics, and Bias: Tapping the Social Web for CSCW Research." CSCW, ACM, Savannah, Georgia. (Reading on Blackboard)
- Sarah Allison (2016). "Other People's Data: Humanities Edition." Journal of Cultural Analytics.
- Microsoft Excel blamed for gene study errors. BBC News.

Part 3: Overview of Subfields

In this part of the course we will briefly be exposed to three subfields: text analysis, network analysis, and geospatial analysis. The goal of this section of the course is to introduce you to various ways you can put your programming skills toward a more complex and sophisticated research project. If any of these subfields interest you, you can take more advanced classes in each one at Northeastern.

Week 7: Data Visualization and Computational Subfields

Data visualization is incredibly important at every step of your project: at the beginning, to make sure there's nothing off about your data, in the middle, to make sure your analysis makes sense, and at the end, to convey your findings to others and impart your knowledge. We'll get a quick and dirty introduction to visualizations in Python, and an overview of Topic Modelling, Network Analysis and Geo-Spatial Analysis and we will think through how you might want to use them in

your own project. You will choose one of three tutorials to watch and complete on these techniques.

There will be no reading or reading responses due this week. Instead, I will provide sandbox time for you to work on your final project proposals, have multiple drop-in office hours, and be available for questions and help. I'll provide a brief course wrap up at the end of the week.

Monday, August 10
- Persuasive visualizations
- *Visualization in Python*
- *Text Analysis – Topic Modelling*
- *Network Analysis – Intro to Python for NetworkX*
- *Geospatial Analysis – Intro to Tableau*

In-Class Tutorials
06.00_PracticingPandas_InClassTutorial(1).ipynb

Choose One of the Following Tutorials to Complete and Submit Based on Interest
08.00_TextToNetworks_InClassTutorial.ipynb
Topic Modelling Tutorial
Mapping Spatial Data with Tableau Tutorial

Last Day of Class August 13th

Tuesday, August 18: Final Project Proposal Due by 8pm