Introduction to Computational Methods in the Social Sciences Syllabus (Sociology 476 Fall 2025)

Class Meetings: Mondays, 9:00-11:50, Parkes Hall 222.

Instructor: Doron Shiffer-Sebba, Department of Sociology, 1808 Chicago Ave., Room 103 doron@northwestern.edu

Office hours: By appointment or Mondays 3:30-5pm. Sign up here. (https://calendar.app.google/gTb8WR2KcMiPMwuo7)

Please use email to ask any questions or schedule an appointment if you cannot make regular office hours. I will do my best to respond within 24 hours M-F 9am-5pm.

Synopsis

Social scientists use computers for every step of the research process, but seldom take advantage of their full potential. This course equips social scientists with computational principles and habits for making the research process more efficient, flexible, and reproducible. It also provides a solid foundation for pursuing advanced computational methods. The course is designed for early stage graduate students using *any* research method (statistical, ethnographic, historical comparative, etc.) with no specialized background in computational methods. Most techniques taught in this course require concentrated time and effort (learning, installing, and customizing) but offer exponential gains in efficiency and peace of mind over one's research career. This course aims to provide structured time for this initial adoption.

Course Objectives

Recognize fundamental computational concepts for future learning and application of computational techniques

Understand current debates on reproducibility to do better science

Develop computational problem-solving skills to overcome future challenges

Establish your own reproducible and efficient research pipeline for peace of mind, efficiency, and reproducibility

Schedule Overview

Date	Topic	Class Format	Assignment Due
Week 1 09/22/2025	Overthrowing our Software Overlords: Plain Text	Lecture	–
Week 2 09/29/2025	Reproducibility: Structure and the Shell	Lecture - Lab	Capture Screen, Customize Editor
Week 3 10/06/2025	Reproducibility: History and Technical Debt	Lecture - Lab	Directory Strategy and The Shell
Week 4 10/13/2025	Fundamentals of Coding I: Languages, IDEs, Variables, and Data Types	Lecture - Lab	Technical Debt Strategy and Git
Week 5 10/20/2025	Fundamentals of Coding II: Packages, Methods, and Functions	Lecture - Lab	IDE & Problem Set
Week 6 10/27/2025	Qual vs. Quant: APIs, RegEx, and Virtual Environments	Lab - Lecture	Reproducibility Function
Week 7 11/03/2025	Complex Code Pipelines	Lecture - Lab	Set up Environment/ Search Corpus
Week 8 11/10/2025	Writing Pipeline	Lecture	Custom Shell Command
Week 9 11/17/2025	AI and HPCs	Lecture - Lab	–
Week 10 11/24/2025	Student Pipeline Showcase	Presentations	Data-to-Paper Pipeline

Assessment

All assignments take the form of short narrated (either orally or in writing) screen-capture videos, where students will showcase and explain their new computational techniques. All assignments are due on Canvas by 5pm the day before class. You may (and are encouraged to) use GenAI, so long as you disclose it and/ or it is captured in the video submission. Readings are due by start of each class.

1) Data-to-Paper Pipeline (due 11/24/2025): **20 Points**

The "Data-to-Paper Pipeline" is the synthesis of the quarter's work. It will encapsulate all the new computational techniques the student chose to incorporate

into their research pipeline, an explanation of why they chose these techniques, and an exploration of what techniques might be missing or which principles are lacking. A short final presentation on 11/24/2025, which will be explained on 11/17/2025, will be considered part of this final assignment. Late submissions will be penalized 2 points/ day.

2) All Other Assignments (due 9/29/2025 - 11/10/2025): **60 Points** (10 each)

All assignment will be posted on Canvas. Serious, on-time submissions will receive either full points or "requires resubmission". Resubmission deadlines are one week after original deadlines. Resubmissions will receive either full or partial credit. Late submissions will be penalized 1 point/ day. Late resubmissions will not be accepted. One lowest scoring assignment will be dropped from the final grade calculation.

3) Attendance: **20 Points**

While you will get out of this course what you put in, at a minimum it is designed to give you class time to execute new computational techniques. Therefore, attendance is mandatory. 3 points will be deducted for every missed class. One absence without a formal NU note is permitted. More than one absence without a formal note will be excused only in unusual circumstances (please do not attend if you are feeling sick, sick notes are easily obtained). Students should notify of conflicts due to religious holidays well in advance of any anticipated absence. For accessibility-related accommodations, please contact AccessibleNU.

Full Schedule

Week #1 09/22/2025 Overthrowing our Software Overlords: Plain Text

No readings or assignments due. Please bring a fully-charged computer to every class.

Week #2 09/29/2025 Reproducibility: Structure and Shell

Assignments: (due on Canvas by noon the day before class)

(Read the Healy chapters before starting the assignments).

1. Familiarize yourself with recording your screen and voice simultaneously. Mac and Windows 11 have this capability natively: Mac; Windows. But there are a variety of alternatives out there.
2. Follow along with Chapter 1 of Software Carpentry (Introducing the Shell).
3. Download and set up VS Code if you have not already done so. If you have strong preferences for an alternative I am open to it. Figure out how to open the shell from within the code editor.

4. Find and install *at least 7* code editor extensions that are useful or desirable for you (hint: search Google and Youtube for extensions for particular purposes).
5. Create a Markdown (`.md`) file with at least a header, sub-header, **bold text**, *italic text*, numbered list, link or image, and a footnote.
6. Finally, record a short narrated video capturing your screen where you give a tour of your new extensions in action (e.g., what they do for plain text files like `.txt` or `.md` or what they do in the editor shell) and explain why you chose these extensions. Upload the video to Canvas.

Readings: (to be completed before and for this class)

- Chapters 1 & 3 of Kieran Healy, 2019. The Plain Person's Guide to Plain Text Social Science, https://plain-text.co/.
- Patrick Ball. 2016. Principled Data Processing (video).[1]
- Jenny Bryan. 2022. Naming Things (video).
- Chapter 4 of Gentzkow & Shapiro, 2014. Code and Data for the Social Sciences: A Practitioner's Guide.

Week #3 10/06/2025 Reproducibility: History and Technical Debt

Assignments:

1. Follow along with Chapters 2-5 of Software Carpentry. (not part of this week's video submission)
2. Come up with an informed and informative general-purpose folder/ file structure that you can use for multiple projects. Record a short narrated video capturing your screen where you display and explain your structure, and how it suites your research methods/ preferences, with reference to last week's "readings" on the topic. Upload the video to Canvas.

Readings:

- Chapter 2 of Kieran Healy. 2019. The Plain Person's Guide to Plain Text Social Science, https://plain-text.co/.
- MIT Missing Semester: Version Control.
- Toor, Skye. 2019. Avoiding Technical Debt in Social Science Research.

Week #4 10/13/2025 Fundamentals of Coding I: Languages, IDEs, Variables, and Data Types

Note: If you have a Windows machine you may need to install Python on your computer (see here).

Assignments:

1. Chapter 6 of Software Carpentry.
2. Come up with a strategy to keep track of your technical debt.

[1] For this and under useful sources I thank Alex Hanna.

3. Create an account with a Git hosting service (e.g., Github, GitKraken). Set up a project with Git – either create a brand new project (locally or directly on Git host website) or select an existing project – and initiate it as a Git repository.
4. Screen capture yourself marking technical debt in your project, committing the change to your Git repository, and pushing the change to the Git hosting service (both via the terminal, not a GUI).

Reading:

- Official Python documentation that introduces its data types.

Week #5 10/20/2025 Fundamentals of Coding II: Packages, Methods, and Functions

Assignments:

1. Finish setting up a Python IDE, including any useful VSCode extensions.
2. Screen capture solving the problem set from Canvas in your IDE.
3. Paste the solution into an external script and run it through the IDE or shell.
4. Upload the video to Canvas.

Readings: *None*

Week #6 10/27/2025 Qual vs. Quant: APIs, RegEx, and Virtual Environments

Assignments:

1. Identify an external package/ library that may be useful to you and install it on your computer (see here for help with Python package installation).
2. In week #3 you came up with a general purpose naming convention for your files and folders. Write two functions in your IDE: one that creates folder names consistent with your naming system and one that creates file names consistent with your naming system, given a set of parameters/ inputs. For example, my convention is for each project to have a series of folders corresponding to different steps, and for each step to include an `"input"`, `"output"`, and `"src"` folder. So I would write a function that, given a number (e.g., "5"), would manufacture the folder name `"step_5"`, the string `"input"`, the string `"output"`, and the string `"src"`. It would also manufacture the string `"step_5.R"`, the name of the R script that would live in the `"src"` folder. The main thrust of this assignment, however, is to practice writing functions with inputs and outputs. Screen capture your functions in action and upload to Canvas.

Readings:

- **Quantitative**: Watch NUIT's Singularity 1 Video for an introduction to containers.

- **Both**: Read MIT Missing Semester section under "Regular Expressions" here.
- Read MIT Missing Semester section under "APIs" here.

Week #7 11/03/2025 Complex Code Pipelines

Assignment:

1. **Qualitative**: place a corpus of data in one folder (collection of interviews, primary documents, etc.).
    1. Open that folder in VS Code and run 5 searches using regular expressions for theoretically informed words or phrases.
    2. Conduct the same searches using a shell.
    3. Conduct a shell search for the file you created in last week's assignment #2 using a regular expression.
    4. Screen capture all searches and upload to Canvas.
    5. Try to understand what extensions like Dendron and Foam do. Think about how you might use these programs for qualitative research.
2. **Quantitative**: create a virtual environment for a project in a language of your choosing. Read documentation or watch videos to learn how to activate that virtual environment with the least effort.
    1. initiate the virtual environment for your project.
    2. Install some packages.
    3. Screen capture listing the packages through the virtual environment and upload to Canvas.
3. If do **both**: do both (you chose this life buster. . .)

Reading:

- MIT Missing Semester: Meta-programming.

Week #8 11/10/2025 Writing Pipeline

Assignments:

1. Automating a Reproducible Structure:
    1. Translate both assignment #2 functions (that created folder/ file names) from two weeks ago into a shell script.
    2. Enhance that script so that it creates your entire folder structure from week #2, including multiple folders and at least one file.
    3. Find a way to execute that script with either a keyboard shortcut or a command that does not require the full script path (hint: this may involve custom key bindings in your code editor).
2. **Quant Only:** Make a Make file that includes at least three files (they can be any combination of bash scripts, R, Python, or only one of those), where each file uses as input the output from the previous file.
3. Think: what research tasks do you do repeatedly on your computer. What would it take to automate them?

4. Screen capture #1 (and #2) above and upload to Canvas.

Readings:

- Chapter 5 of Kieran Healy. 2019. The Plain Person's Guide to Plain Text Social Science.
- Read about YAML on the World Wide Web and decide whether it is useful for you.

Week #9 11/17/2025 AI and HPCs

Assignments: *None*

Readings:

- Gibney, Elizabeth. 2025. "What are the Best AI Tools for Research? *Nature*'s Guide".
- Github Copilot on VS Code.
- Bal, Christopher. 2024. "Can Generative AI Improve Social Science".
- **Qualitative**: Hayes, Adam 2025. "Conversing" With Qualitative Data: Enhancing Qualitative Research Through Large Language Models (LLMs).
- **Quantitative**: Davidson, Thomas & Daniel Karell 2025. Integrating Generative Artificial Intelligence into Social Science Research: Measurement, Prompting, and Simulation.

Week #10 11/24/2025 Student Pipeline Showcase

Assignment:

Create the ideal "Data-to-Paper" computational pipeline for you (or as close to it as you can get). What makes it ideal? Is there anything you wish you could do differently? Did you purposefully exclude potentially relevant elements? Beyond the pipeline itself, please include your complete reasoning for various choices.

You will present your pipeline to the class during this final class.

Policies

This course follows the Northwestern University Syllabus Standards unless stated otherwise. Students are responsible for familiarizing themselves with this information.

Letter Grading Scale

93+	A	90-92	A-
87-89	B+	83-86	B
77-79	C+	73-76	C
67-69	D+	60-66	D
0-59	F		

Grade Disputes

If you are not satisfied with a grade on a given assignment, you may request a re-grade. For a given assignment, all submissions that have requested a re-grade are graded anonymously by someone who has not graded them before (either a TA, instructor, or external expert). Re-grade requests may only be submitted over email or privately in office hours at least three days after the grade has been given. NOTE: *a re-grade may result in a lower grade*, it is not guaranteed to stay at or above the original grade.

Well-Being

Students can find useful resources for safety and security, academic support, and mental and physical health and well-being at the NUhelp website and app.

Slides

Slides will be posted to Canvas by the end of each week.