GESIS Fall Seminar in Computational Social Science 2023

Syllabus for week 1:
“Introduction to Computational Social Science with Python”

Lecturers: Milena Tsvetkova  Patrick Gildersleve
Affiliation: London School of Economics London School of Economics
Email: m.tsvetkova@lse.ac.uk p.gildersleve@lse.ac.uk

Date: September 11-15, 2023
Time: 09:00-12:00 and 13:00-16:00 (including one 15 min break per session)

About the Lecturers
Milena Tsvetkova is Assistant Professor of Computational Social Science at the Department of Methodology at the London School of Economics and Political Science. She completed her PhD in Sociology at Cornell University and postdoctoral training at the Oxford Internet Institute. In her research, she uses large-scale web-based experiments, network analysis of online data, and agent-based modeling to investigate fundamental social phenomena such as cooperation, social contagion, segregation, and inequality.

Patrick Gildersleve is an LSE Fellow in Computational Social Science in the Department of Methodology. Patrick graduated with a Masters in Physics from the University of Oxford, before completing his PhD at the Oxford Internet Institute in 2021. In his PhD research, he worked on studying the intersection of news media and Wikipedia. Patrick analyzed how current events are recorded and accessed on the online collaborative encyclopedia as well as its implications for theories of news values, newsworthiness, and collective attention dynamics. He has continued this work with an expanded research agenda around popularity and collective memory across platforms online.

Course Description
The course provides an introduction to the basic computational tools, skills, and methods in Computational Social Science using Python. Python is the most popular programming language for data science, used widely in both academia and the industry. Students will learn to use common workflow and collaboration tools, design, write, and debug simple computer programs, and manage, summarize, and visualize data with common Python libraries. The course will employ interactive tutorials and hands-on exercises using real social science data. Participants will work independently and in groups with guidance and support from the lecturers. The practical exercises are designed to demand more autonomy and initiative as the course progresses over the five days, culminating in an open-ended group project in the last afternoon session.

Keywords
Python, programming, data analysis

Course Prerequisites
This is an introductory course, and no prior experience with programming is required. Basic understanding of statistics and some scripting experience (e.g., from building web pages or statistical analysis programs such as Stata or R) will be helpful but not needed.
For those who would like a primer or refresher in Python, we recommend taking the online workshop “Introduction to Python” that takes place from 04-06 September 2023.
**Target Group**

Participants will find the course useful if they:

- have no or limited technical and computational background
- have a background in one of the social sciences (sociology, political science, psychology, etc.)
- would like to pursue research or a professional career in computational social science or social data science (e.g., in academia, think tanks, government, NGOs, social media companies, tech startups)

**Course and Learning Objectives**

By the end of the course participants will:

- possess an understanding of the tools, methods, tasks, and goals of Computational Social Science
- design procedures and algorithms to solve data analysis tasks
- write simple programs in Python
- work confidently with pandas, matplotlib, seaborn, and other popular Python modules and libraries for data science
- use bash, Jupyter Notebook, and GitHub to write, run, collaborate on, and share programming code

**Organizational Structure of the Course**

The course will consist of two three-hour-long sessions each day. The morning session will use interactive instruction to introduce participants to the topic, demonstrate the new methods, and facilitate discussion. The afternoon session will make use of guided hands-on exercises with real-world data to practice the new material.

Participants will work individually, in pairs, and in groups, and the lecturers will be available throughout both sessions for consultation and support.

**Software and Hardware Requirements**

Participants require a laptop computer with Anaconda and git installed. Some time will be allocated on the first day of the course to install Anaconda and git.

**Recommended Literature to Look at in Advance**


**Day-to-day Schedule and Literature**

**Day 1: Computation for Social Science**

- What is CSS?
  - Data, methods, and questions
  - Accountability, reproducibility, and ethics
- Setting up your workflow
  - Installing Python with Anaconda
  - Introduction to Jupyter Notebooks
  - Introduction to Bash and GitHub
- Introduction to programming with Python
  - Scalar data types, operators, and expressions
  - Variable assignment, printing, and comments
  - Non-scalar data types, indexing, and slicing
List and string methods

Recommended Literature:
Matthes, Eric. Python Crash Course Cheat Sheet. 
https://github.com/ehmatthes/pcc/releases/download/v1.0.0/beginners_python_cheat_sheet_pcc_all.pdf


GitHub Tutorials. https://docs.github.com/en

Day 2: Writing Legible, Modular, and Optimized Code
- Understanding control flows
  - Conditionals
  - Iteration
  - List comprehensions
- Abstraction and decomposition with functions
  - Function specifications
  - Function arguments
  - Variable scope
- Object-oriented programming with classes
- Modules and libraries

Recommended Literature:
Python Documentation. https://docs.python.org/3/

Day 3: Obtaining Data
- Handling social data
  - Ethics of data access
  - Reading and writing common file types
  - More complex data types: time and dates, Unicode, etc.
- Accessing web data
  - Inspecting webpages
  - Parsing static HTML with BeautifulSoup
  - JSON and working with APIs
- Text analysis with nltk

Recommended Literature:
Selenium Documentation. https://selenium-python.readthedocs.io

Day 4: Analysing Rectangular Data
- Introduction to pandas
  - Creating DataFrames
  - Accessing and filtering data
  - Computing summary statistics
  - Reading and writing data
- Manipulating pandas DataFrames
  - Handling different data types
  - Combining data from different tables
Applying functions to DataFrames
 Creating basic plots using pandas
 Machine learning with scikit-learn

Recommended Literature:
Pandas Documentation. https://pandas.pydata.org/docs/

Day 5: Visualising Data and Analysing Non-Rectangular Data
 Basics of visualization
  Understanding plot elements
  Choosing the right chart
  Principles of color
 Plotting data with Matplotlib and Seaborn
  Basic plotting in Python
  Pyplot vs the object-oriented approach
  Customising plots and figures
  Attractive plots with Seaborn
 Network analysis with networkx

Recommended Literature:

Additional Recommended Literature