

GESIS Fall Seminar in Computational Social Science 2021

Syllabus for week 1:

"Introduction to Computational Social Science with Python"

Lecturers: Dr. Orsolya Vásárhelyi
Email: orsolya.vasarhelyi@gmail.com

Luis Natera
natera@hey.com

Date: September 13-17, 2021

Time:

10:00-12:15 morning session (with a 15 mins. break)
13:30-15:00/16:00 afternoon sessions, depending on preparation workload
(with a 15 mins. break)

About the Lecturers:

Orsolya is a postdoctoral fellow at the University of Warwick, Center for Interdisciplinary Research. Her research focuses on the gender differences in career development in project-based environments. She has been teaching python for social scientists for 4 years.

Luis is a Ph.D. candidate of network science at the Department of Network and Data Science at the Central European University, in Budapest, Hungary. His research revolves around urban mobility in multiplex urban networks, bridging urban planning and complex systems. He has experience in interdisciplinary research and has worked in academic, government and private sectors.

Course Description:

This course provides a highly interactive introduction to Computational Social Science using Python. Python is the most popular programming language of data science, used in natural language processing, machine learning, and artificial intelligence. This five-day course is designed for social scientists who would like to conduct data collection, analysis, and modelling with Python and attend further courses within the GESIS Fall Seminar in Computational Social Science. Classes focus on hands-on exercises, therefore participants are required to watch pre-recorded videos about the theoretical background that is needed to start your journey in computational social science (e.g.: software installation, data ethics, tips on how to visualize data).

Keywords:

Python; Programming; Data Analysis; Web Scraping; APIs; Intro to Coding

Course Prerequisites:

No programming experience needed. A basic understanding of statistics and quantitative data analysis will be helpful. Prior experience with writing code (e.g., syntax for statistical analysis programs) is a plus.

Target Group:

- Class is specifically designed for participants with no (or limited) technical background who would like to start their journey to analyze data using Python.
- Participants have various backgrounds usually within the social sciences (e.g.: sociology, political science, psychology)
- Course is recommended but not limited to professionals working/planning to work with data in NGOs, research institutes and universities in any level (Ph.D. candidates, postdocs, professors)

Course and Learning Objectives:

By the end of the course participants will:

- Understanding of what Computational Social Science is
- Introductory skills to programming in Python
- Basic data collection methods: scraping, APIs
- Basic data analysis using pandas (e.g.: data frames)
- Data visualization
- Ethical concerns of working with data

Organisational Structure of the Course:

The course is structured around 30 minutes to 1.5 hours of preparation time, which includes following tutorials and listening to pre-recorded lectures, and 3 to 4 hours of lab sessions depending on the preparation load. Total hours of the course do not exceed 5 hours daily, including preparation time, however we expect participants to arrive to the class prepared.

Lab sessions are structured the following way: At the morning sessions, the instructor showcases an example of the given topic by writing code live with the help of participants. During this time, the other instructor is available online for debugging, and helping participants to follow the class. At the afternoon sessions, participants work on projects in groups and later demo their solution to the class. Lecturers will be available for individual consultations on participants' projects outside of class hours.

Software requirements:

Participants should install Python on their computers prior to the course. Participants need to have a Google account, because we will use [Google Colab](#) and Google Drive throughout the course. For communication purposes a specific Slack channel will be established, where participants are invited to join. Classes will take place via Zoom.

Long Course Description:

The course focuses on preparing participants to start analyzing data and conducting research in computational social science (CSS) using the popular programming language Python. This is an introductory course; therefore, no prior programming experience is required, and the focus is on hands-on exercises. The class starts with a gentle introduction to what is computational social science, then builds the basic technical skills that are needed to start learning about data collection techniques and analysis. This course does not cover advanced data analysis, such as multivariate statistics, Machine Learning or Natural Language Processing. One of the main goals of the course is to provide good programming basics, and tools for participants to continue their journey on data collection and introductory data analysis. We believe that motivated participants after taking this introductory course will be able to learn new skills independently to conduct CSS research using Python. Although it is important to note they will not exit the course as senior programmers, but they will have the required basics to keep going.

Day-to-day schedule and literature:

Day 1: Introduction to Computational Social Science and Python.

- What is CSS?
- Setting up your working environment (Google Colab, Bash, Drive)
- Introduction to the syntax of Python
- Group works focusing on a set of programming games and challenges

Day 2: Data Collection I. - Scraping

- HTML structures
- Identifying tags in HTML
- Python for scraping websites
- Group work to scrape data from a webpage

Recommended Literature:

Mitchell, Ryan. Web Scraping with Python, 2nd Edition. O'Reilly 2018.

Day 3: Data Collection II. – APIs

- What is an API?
- Python and APIs
- Group work using the Spotify API to analyze music

Day 4: Data Analysis I. – Data visualization and manipulation in pandas

- Intro to pandas
- Intro to data visualization
- Data visualization with matplotlib and seaborn

Recommended Literature:

Tufte, Edward R. The Visual Display of Quantitative Information. Graphics Press 2001.

Knaflic, Cole Nussbaumer. Storytelling with data: A data visualization guide for business professionals. John Wiley & Sons, 2015.

Day 5: Data Analysis II. – Intro to data analysis with Python & Data Ethics

- Data analytical group work (using pandas, seaborn and descriptive statistics)
- Discussion on the ethical consequences of working with data in CSS
- Recommendation on how to continue your journey as a computational social scientist

Recommended Literature:

O'neil, Cathy. Weapons of math destruction: How big data increases inequality and threatens democracy. Crown, 2016.