University of Zurich[UZH]

Seminar "Foundations of Data Science for Social Scientists"

University of Zurich, Spring/Summer, 2020

Lecturer: Prof. Dr. Karsten Donnay (donnay@ipz.uzh.ch)
Assistant: Philipp Kling (kling@ipz.uzh.ch)
Contact: donnay@ipz.uzh.ch and OLAT
Office Hours: By appointment (online)
Course number: 615b539a Seminar

Overview and Objectives

This course introduces foundational concepts from computer science that are fundamental to data science approaches in the social sciences. Topics covered range from basic principles of information coding, data types and structures, algorithms and their complexity, to efficient approaches for automated data collection, storage, processing and analysis. For each topic, the course first covers the conceptual foundations from computer science before illustrating them using hands-on examples from the social sciences. Throughout the class we also cover general best practices (e.g., reproducibility, version control using Git) for data science. The course uses the R statistical programming language as teaching language and a basic familiarity is assumed. *If you are not familiar with R but want to attend the course, please prepare yourself with basic tutorials that are, for example, available on datacamp.com.*

Course Objectives and Key Skills

- Students will develop a good understanding of technical and conceptual foundations of data science approaches in the social sciences.
- They will learn to apply them in relevant research settings while adhering to best practices and standards of data quality and reproducibility.
- They will familiarize themselves with relevant tools and approaches and learn to apply them to their own research questions.

Course Times

The course takes place in the form of a block seminar in the week of June 15 to 19, 2020 (Mon. to Fri) each day with two sessions from 10:15 to 12:00 and from 14:00 to 15:45. The final coding project will be due one week after the class on June 26, 2020 (Fri.). Note that given the current restrictions due to COVID-19 the course will be taught completely online using *BigBlueButton* and the OLAT platform.

Course Assessment

The course grade is based on written exercises (40%) and a written assignment (60%):

- There will be four written exercises on Mon. through Thu., each due before the afternoon session of the following day. These exercise sheets deepen the topics covered during each day's exercise session.
- There will be a final coding project due one week after the class, i.e., on Friday June 26, 2020. The project is meant to focus on a practical social science research question leveraging techniques for data collection and analysis acquired during class.

Communication

Email addresses of the instructors and all participants in the course are to be treated confidentially. The (virtual) classroom is the best place to raise questions which are relevant for everybody in the class. The best time to ask short questions that might not be relevant for everybody is after class. The office hours should be dedicated to discuss more in- depth questions. I strongly discourage emails about class content that could be solved in class, after class or during office hours. I would like to encourage you to contact me via email or during office hours if there is something that makes you feel uncomfortable in class or about the course.

Course Outline

Part 1: Foundations

Day 1: Mon. 15.06.2020

10:15 – 12:00 Lecture 1: Information Coding & Data Structures
14:00 – 15:45 Exercise Session 1

Day 2: Tue. 16.06.2020

10:15 – 12:00 Lecture 2: Programming & Algorithms
14:00 – 15:45 Exercise Session 2

Day 3: Wed. 17.06.2020

10:15 – 12:00 Lecture 3: Complexity & Efficiency
14:00 – 15:45 Exercise Session 3

Part 2: Applications

Day 4: Thu. 18.06.2020

10:15 – 12:00 Lecture 4: Data Collection & Quality
14:00 – 15:45 Exercise Session 4

Day 5: Fri. 19.06.2020
10:15 – 12:00 Lecture 5: Research on Digital Media
14:00 – 15:45 Exercise Session 5