COMPUTATIONAL SOCIAL SCIENCE METHODS IN R

Introduction

Taehee Kim

Summer Semester 2022

ABOUT ME

Taehee Kim

- Postdoc research associate in Political System of Germany team
- Email: taehee.kim@uol.de
- Office: A06 3-303a, Phone: 4435
- Office hour: **contact me via email**

BACK GROUND, RESEARCH INTEREST

- Ph.D Political Science, Keio University, Japan
- M.Sc.Social and Economic Data Analysis, University of Konstanz
- Political attitudes, behavior
- Social media and democracy
  - Hate speech spread on Twitter: SocArXiv
  - Conspiracy theory spread and party support in Germany

COURSE INTRODUCTION

Github repository
https://github.com/thkim0321/ol2022-css

For example, you can find syllabus in ol2022-css/Slides.

COMPUTATIONAL SOCIAL SCIENCE

- applying computational methodologies in social science
- computational approaches to substantial social science problems
- collection, representation, processing, storage and access to social science data

Schelling (1971)

- early study of agent-based model
- desire to being majority among community or want to congregate with his own color in small number are sufficient to generate segregation at the macro level –“inferences about individual motives can usually not be drawn from aggregate patterns” (p.143)

WHY WE NEED CSS?

Data revolution in social science

- very large, diverse source of data is available
- digital traces, e.g., social media data
- potential to fundamentally change our understanding of social interaction

Needs for computational methods

- to collect, store, process, and analyze massive, diverse type of data
- automated data collection
- design efficient and correct algorithms

BIG DATA AND MACHINE LEARNING

- Big data
  - too large to adequately handle
  - related to machine learning: we can find some patterns so that make a prediction (e.g., trend, human behavior)
- Machine learning
  - learn from large data without relying on rules-based programming
  - make a prediction based on the learning

READINGS

- Lazer et al. 2009. “Computational Social Science.” Science 323:721–723. link
- Ruths and Pfeffer. 2014. “Social Media for Large Studies of Behavior.” Science 192:59–60.link
- Giles, Jim. 2012. “Computational Social Science: Making the Links.” Nature 488:448–450. link
- Golder and Macy. 2011. “Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures.” Science 333:1878 – 1881. link

READINGS

- Grimmer et al. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” Political Analysis 21(3):267–297. link
- Schich et al. 2014. “A Network Framework of Cultural History.” Science 345:558–562. link
- Watts, Duncan J. 2004. “The “New” Science of Networks.” Annual Review of Sociology 30(1):243–270. link

GOAL OF THE COURSE

- Introduce the concepts and techniques of computational social science, focusing on **machine learning**.
- Supervised machine learning algorithms
  - text
  - image
- Google Machine Learning APIs

PROGRAMMING LANGUAGE

- **R** for data wrangling
- **Python** for machine learning
  - one of the popular versatile script languages; a large set of available machine learning packages

HOW DO WE PROCEED

Go to the syllabus

PREPERATION

R

- R > 4.0
- RStudio (most recent version)

PREPERATION

Python

- Installation: Anaconda distribution
- Jupyter Notebook
- We go through from installation to launching Jupyter Notebook together step by step. We use Anaconda Navigator which make you easily manage environments.

Let’s install!