

GV918: Data for Social Data Science Course organisation

Akitaka Matsuo

About me: Akitaka Matsuo

- Lecturer in the Department of Government
- Ex-research fellow in Institute for Analytics and Data Science (IADS)
- Ph.D in Political Science, Rice University
- Research Interest
 - Political methodology (natural language processing, scaling)
 - Legislative politics
 - Social media analysis
 - UK Politics / Japan v Korea
- Member of quanteda project

Plan for Week 02

- Administration and logistics
- Data
- Why python?
- Jupyter notebook and google colab
- git and github
- Introduction to python

GV918

- This is an introductory course for data science
- We will learn:
 - The lifecycle of the data in social data science
 - data acquisition, pre-processing, storing for analysis
 - Knowledge of collaborative working space
 - Cloud computing
 - Basic principles of machine learning
- This is not a course for:
 - Statistical analysis (GV903 or GV900)
 - Machine learning methods

Course outline

2. Data in social science
3. Data manipulation
4. Data visualisation
5. Cloud computing
6. Using the data from the Internet
7. Working with APIs
8. Working with databases I
9. Working with databases II
10. Introductory machine learning I
11. Introductory machine learning II

Prerequisites and software

- Prerequisites:
 - Expect some familiarity with programming
 - R, Stata, SPSS etc
- Computer:
 - We will do all course work on the cloud
- Software:
 - Python 3.x
 - Jupyter notebook
 - Google colab
 - Get a gmail account, if you don't have
 - github

Reading

- The majority of reading is textbooks
 - Should be available electronically or online
- Some substantive reading on the specific topics on each week
- Reading list is on talis
 - <https://rl.talis.com/3/essex/lists/69885F11-38FA-F804-D958-1FBE3898003E.html>

Lecture videos and class meetings

- Lectures: CTC.3.02, Tuesdays, 10am-12pm
 - Can be watched again
- Classes: IT Lab K, Thursdays, 16pm-18pm
- Office hours
 - Office hours: Wednesdays, 3-5pm, Room: 5A.135 (temporarily)

Assessment

- The course will be assessed with coursework
 - Three assignments:

Assignment Title	Due Date	Coursework Weighting*	Feedback Due
Assignment 1	Week 5	30%	Week 8
Assignment 2	Week 8	35%	Week 11
Assignment 3	Week 16	35%	Week 19

- All assignments are:
 - Due on Wednesday, 9:45 of the week
 - Distributed by github classroom
 - Submit on Faser
 - Python programming + Write-up (explanation and interpretation)