Linked Open Data and Computational Social Science Methods

This course introduces methodologies for linked open data and computational social science. The first part of this course is theory-oriented and covers concepts in linking data and open data policies. You will also learn how to use high-performance cloud computing resources (https://www.tacc.utexas.edu/systems/chameleon). The second part is analysis-oriented and covers network analysis, text analysis, and text classification using neural networks. Meanwhile, we will introduce Awesome Public Datasets (https://github.com/awesomedata/awesome-public-datasets) according to the class’s interests. Each of you will present a dataset by your selection. The final part is challenge-oriented. You will form groups to complete a challenge as your final project.

Although programming is an essential part of this course, the course schedule and reading materials are framed within a social science context. We will be coding for social good.

Prerequisites:

1. College level statistics. For example, you are confident to use probability for hypothesis testing, you can run and understand OLS and multivariate regression.
2. Comfortable with programming using Python. The class is Python based, but you can use R or any other programming language as long as you can complete the assignments and final challenge. If you haven't used Python for a while or not familiar with it, please complete an online tutorial before taking this class. Example Python packages used in this course: Pandas, Requests, regular expression, NetworkX, NLTK, TensorFlow, Keras, etc. We will introduce these packages in class, but you should be familiar with Python programming in general before class.

Recommended online tutorials:

1. If you don't have any Python programming experience, take this course first: <https://www.coursera.org/learn/python> (you can audit this course for free).
2. You should be familiar with all the topics covered by this tutorial before class: <https://www.learnpython.org/> .

Grading: 40% assignments, 20% presentation of datasets, and 40% final project.