University of California, San Diego
Department of Economics
**ECON 267: Data and Methods**
Spring 2015

**Time:** Tue/Thu 3:30 -4:50 pm
**Location:** SEQUO 244
**Instructor:** Junjie Zhang, RBC 1303, (858) 822-5733, junjiezhang@ucsd.edu
**Office Hour:** by appointment

**Course Description:** This course introduces data sources and computing tools that are useful in empirical studies. In recent years, the emerging data sources, especially big data, have stimulated new research questions and provided new opportunities to test economic theories. Many of these data are not well structured and tend to be in large size. Using these data requires new computing tools that can obtain the information on the Internet and improve the efficiency of data manipulation. For this purpose, this class introduces Python and other computing tools (such as R and SQL) to empirical researchers. All these tools are open source and can fully substitute proprietary software such as Stata, MATLAB, SAS, and ArcGIS.

This course is beginner friendly by removing unnecessary learning hurdles. The study goal is pragmatic such that we will learn just enough for the next empirical project. In addition, we will work on real research questions, which means the example codes developed in the class can be directly used in other projects. The cases in the class draw from the following topics in economics: dynamic system, optimal control, regression and data analysis, geostatistical analysis, natural resource management, energy and climate change.

This course is originally designed for PhD students in environmental and resources economics. However, those students in other fields of economics may find the applied tools useful in their research. The students in other social sciences such as political science may also benefit from the data sources and computing tools. There is no prerequisite for this class.

**Course Format:** The course is a combination of lectures and computer experiments. Thus, attendance at class meetings is mandatory. Please bring your computer to the class so you can play with the codes during the lectures. There are 5 homework assignments for students to practice programming skills. The final project should propose to apply the data sources and computing tools to an economic question. The final project report is due on June 7, 2015.

**Grading:** The course grade will be based 20% on classroom participation, 40% on homework assignment, and 40% on a final project report.

**Academic Integrity:** Student conduct related to this course is governed by the UCSD Policy on Integrity of Scholarship: "Students are expected to complete the course in compliance with the instructor's standards. No student shall engage in any activity that involves attempting to receive a grade by means other than honest effort."

<div align="center">**Course Outline**</div>

1. **Python programming**

   - introduction and advanced topics

2. **Using Python in economics**

   - applied theory: nonlinear dynamics
   - applied econometrics: data analysis and regressions

3. **Visualization**

   - Python + Matplotlib, R + ggplot2, Asymptote

4. **Database management**

   - SQL and its implementation in Python

5. **Data applications in environmental economics**

   - reduced form versus structural approaches
   - applications to myopia in energy use

6. **Disaggregate data and heterogeneity**

   - heterogeneity and efficiency of environmental policy
   - aggregating up and matching

7. **Text as data**

   - Regular expression and its implementation in Python

8. **Web scraping**

   - applications: AEA JOE, cars.com, Twitter, NOAA FTP

9. **Geostatistical analysis**

   - satellite data: nightlight data, land cover, AOD, NO2
   - Python, R, (and maybe ArcGIS)

10. **Final project presentations**