This version: January 8, 2018

POLI/ECON 5D
Introduction to Social Data Analytics

Winter Quarter 2018
SOLIS 104, Mon/Wed 12-12:50PM

Yiqing Xu
SSB 377
yiqingxu@ucsd.edu
Office Hours: Tuesday 3:00–5:00PM

Teaching Fellows:
Yu-Chang Chen
Office: Econ 124
yuc391@ucsd.edu
Office Hours: Monday 4PM-6PM

Brian Tsay
Office: SSB 323
brtsay@ucsd.edu
Office Hours: Wednesday 3-5PM

Overview

As data about individuals, organizations, and governments become increasingly available, social data analytics are transforming the way we think about the economy, politics and society. This course will teach skills necessary to navigate the world of social data. We will learn basic principles of coding through the lens of popular social science data analytics softwares Excel, Stata, and R. While learning coding fundamentals, we will shed light on big social science questions and grapple with larger societal questions that the era of a society governed by data presents us.

Assessment

Your grade will be based on a combination of:

- **Homeworks (40%):** Four problem sets will be given throughout the quarter. Problem sets will contain analytical, computational, and data analysis questions. Each problem set will be counted equally toward the calculation of the final grade. The following instructions will apply to all problem sets unless otherwise noted.

- Problem sets will be due at the beginning of class. Late submission will be accepted but 1 point (out of 10) will be subtracted from your score for each day of late submission.
- Hard-copy of the homework write-up should be turned in in class, and a copy of the homework write-up and accompanying code should be turned in electronically via Blackboard by the start of class.
- Working in groups is encouraged for conceptual and sometimes technical discussion, but each student must submit their own writeup of the solutions that shows their independent work on the assignment. In particular, you should not copy someone else's answers or computer code. We also ask you to write down the names of the other students with whom you solved the problems together on the first sheet of your solutions.
- For analytical questions, you should include your intermediate steps, as well as comments on those steps when appropriate. For data analysis questions, include annotated code as part of your answers. You will lose points on your problem set if your code and write-up is not properly formatted and documented. All results should be presented so that they can be easily understood and code should run easily without errors.

- **Midterm (20%):** A midterm will be given in class on February 12 (Monday) covering the material in the first half of the class.
- **Final (25%):** The final for this class is tentatively scheduled on Wednesday, March 21 (Wednesday) from 11:30-2:30PM.
- **Participation (15%):** Attendance in lecture and section is mandatory. Quizzes will be given in *every* lecture on assigned readings using i>Clickers. Section attendance will be taken. Students are strongly encouraged to ask questions and actively participate in discussions during lectures and sections.

* **Important Note:** Midterm and final exams are mandatory. A student will not receive a passing grade without taking both exams.

Academic Honesty and Plagiarism

All of your graded work must be done by you. If you are unfamiliar with the University's policy on academic integrity, please see

http://students.ucsd.edu/academics/academic-integrity/policy.html.

We take this very seriously. If we observe signs of violations of academic integrity, we will report to the university *immediately*.

Course Website and Piazza Forum

Syllabus and course materials. The syllabus will be updated and posted on Piazza periodically throughout the course, so that we can keep with the cadence of the class. Assignment, solutions, and other course materials will be posted on Piazza, too.

Online Q&A. Throughout this class we will use the Piazza online discussion board. This is a question-and-answer platform that is easy to use and designed to get you answers to questions quickly. It supports code formatting, embedding of images, and attaching of files. We encourage you to ask questions on the Piazza forum for clarifications, questions about concepts, or about your projects in addition to attending recitation sessions and office hours. You can sign up to the Piazza course page either directly from the below address (there are also free Piazza apps for the iPhone and iPad):

https://piazza.com/ucsd/winter2018/poli5d

Using Piazza will allow you to see and learn from other students' questions. The TAs and the instructor will regularly check the board and answer questions posted, although everyone else is also encouraged to contribute to the discussion. A student's respectful and constructive participation on the forum will count toward his/her class participation grade. *Do not email your questions directly to the instructors or TAs* (unless they are of personal nature) — we will not be answering your questions regarding course materials or problem sets through email.

Course Materials

Since we will be learning Excel, Stata, and R, we will draw on a number of different resources. Many of these resources will be videos from YouTube, blogs, and some will be traditional textbooks. All are freely available online or have been provided by the authors. A few of the primary sources are listed below:

- Principles of Coding: We will rely on videos and exercises from the Hour of Code: https://code.org/learn
- Excel Easy Tutorial: http://www.excel-easy.com/
- Princeton Stata Tutorial: http://data.princeton.edu/stata
- UCLA Stata Resources: http://www.ats.ucla.edu/stat/stata/
- **Textbook** (for R): *A First Course in Quantitative Social Science*, by Kosuke Imai, Princeton University Press.

Participation (with i>Clickers)

Evidence-based research on teaching and learning has documented a strong causal relationship between active participation/discussion and student learning. The risk of large courses like ours is that students miss out on the opportunity to meaningfully discuss course materials, and thus learn less. For this reason, I will use clickers.

1. Official counting period. We will begin experimenting with clickers during the first three lectures, but the "official" counting period will not begin until Week 3 (Jan 24, Wednesday). This should give you time to find a clicker to borrow or purchase.

2. Type of questions. In general, we will ask two types of questions: (1) factual questions and (2) discussion questions. Factual questions focus on a central point from your readings, or a point covered in lectures. Discussion questions ask that you take a stand on a particular problem or issue using course materials as evidence.

3. Grading.

   - Factual questions. One point is given for correct answers, and .5 for participating.
   - Discussion questions. You will receive full points (1 point) simply for participating.
   - One-time exemption. Throughout this quarter, you will have *one* chance of not participating using your iClicker, either because you cannot physically come to class that day or because you forget to bring your iClicker or it does not function properly – in case that happens, please fill choose the data of class on this doodle pool:

     https://doodle.com/poll/fynycfw5nziwvuds

     Your score for the quiz taken on that day will be the average score of the entire class. We will not accept any other excuses or complaints.
   - Reporting. You will be find the record of your responses on TritonEd throughout the quarter (there may be lags occasionally).
   - Below is the formula we will use to calculate your final participation grade:

     Participation Grade = Your Total iClicker Points / Total #Questions Asked * 10
     + Section Attendance Rate * 5

Software

This course will consist of three different statistical software programs commonly used by social scientists.

- Excel: All students will need to have purchased access to Excel. Excel is also available in UCSD computer labs.
- Stata: Instructions for getting Stata through the Virtual Computing Lab are available on the TritonEd Website.
- `R`: an open-source statistical package. You can download it from the web here: http://cran.r-project.org/ RStudio is a useful tool for coding in R. You can download it from the web here: https://www.rstudio.com/

COURSE SCHEDULE

January 8: Course Introduction and Why Data Analytics?

Course Materials

- "Getting Started with Data," Hilary Mason. https://www.youtube.com/watch?v=GXjjMSn2Nws
- "Big data in the service of humanity: Jake Porway" https://www.youtube.com/watch?v=fZ3xXXeVrIQ

January 11: Data Format and Introduction to Excel

Course Materials

- "Why the New Research on Mobility Matters: An Economist's View," Justin Wolfers. http://www.nytimes.com/2015/05/05/upshot/why-the-new-research-on-mobility-matters-an-econ html
- *Statistical Modeling: A Fresh Approach*, Daniel Kaplan. Chapter 2, 2.1-2.4. http://www.mosaic-web.org/go/StatisticalModeling/Chapters/Chapter-02.pdf

January 15: Martin Luther King Day: No Class

January 17: Functions in Excel

Course Materials

- "Introduction to Functions and Formulas" http://www.excel-easy.com/functions.html
- "Cell References" http://www.excel-easy.com/functions/cell-references.html
- "Logical" http://www.excel-easy.com/functions/logical-functions.html
- "Count and Sum" http://www.excel-easy.com/functions/count-sum-functions.html
- "Statistical Functions" http://www.excel-easy.com/functions/statistical-functions. html
- "Lookup and Reference Functions" http://www.excel-easy.com/functions/lookup-reference-functio html
- "Function Errors" http://www.excel-easy.com/functions/formula-errors.html

January 22: Introduction to Stata and Reproducibility

Problem Set 1 Due

Course Materials

- "Stata Tutorial: Introduction" http://data.princeton.edu/stata/
- "Introduction to the Stata Interface," Alan Neustadtl, 15 minutes. https://www.youtube.com/watch?v=KkCKEK7lwuo&index=1&list=PLRYSxJ3XjgQM342QrBkzek8clHa5ue4Sd
- "Using the Stata Program Editor," Alan Neustadtl, 15 minutes. https://www.youtube.com/watch?v=XmvWydFD2Y0&index=6&list=PLRYSxJ3XjgQM342QrBkzek8clHa5ue4Sd

January 24: Data Management and Description in Stata

Course Materials

- Data Management in Stata, http://data.princeton.edu/stata/dataManagement.html

January 29: IF Statements

Course Materials

- Bill Gates Explains If Statements, Hour of Code, https://www.youtube.com/watch?v=m2Ux2PnJe6E
- Complete Bee Conditional Puzzles https://studio.code.org/s/course3/stage/7/puzzle/1

January 31: Graphics in Stata

Course Materials

- "The Beauty of Data Visualization," David McCandless TED Talk, 20 minutes. https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization?language=en
- "Stata Graphics", http://data.princeton.edu/stata/graphics.html

February 5–7: Regression in Stata

Problem Set 2 Due

Course Materials

- "The Easiest Introduction to Linear Regression," Quant Concepts, https://www.youtube.com/watch?v=k_OB1tWX9PM, 15 minutes.
- "Simple and Multiple Regression in Stata," Section 1.0 and 1.3 http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter1/statareg1.htm

February 12: Midterm

February 14: Introduction to R

Course Materials

- "R You Ready for R?" *The New York Times* http://bits.blogs.nytimes.com/2009/01/08/r-you-ready-for-r/
- Imai, 1.3.1-1.3.3

February 19: President's Day, No Class

February 21: Analysis of Experiments by Subsetting Data in R

Course Materials

- "Research: How Subtle Class Cues Can Backfire on Your Resume," *Harvard Business Review* https://hbr.org/2016/12/research-how-subtle-class-cues-can-backfire-on-your-resume
- Imai, 2.1-2.2

February 26: Visualizing Data in R

Problem Set 3 Due

Course Materials

- "Visualizing Ourselves with CrowdSourced Data," Aaron Koblin https://www.ted.com/talks/aaron_koblin
- Imai, 3.3, 3.6

February 28: IF Statements and FOR Loops in R

Course Materials

- Mark Zuckerberg on For Loops, https://www.youtube.com/watch?v=mgooqyWMTxk&
- Trina Roy on Counters, Hour of Code, https://www.youtube.com/watch?v=gxc9RCMvky0&
- Complete all 11 Artist Puzzles https://studio.code.org/s/20-hour/stage/11/puzzle/6

March 5: IF Statements and FOR Loops in R

Course Materials

- Imai, 4.1

March 7: Regression in R

Course Materials

- Imai, 4.2

March 12: Functions in R

Problem Set 4 Due

Course Materials

- Chris Bosh on Functions, https://www.youtube.com/watch?v=0eo0ESEX9DE
- Imai, 1.3.4

March 14: Putting it All Together and Some Big Questions

Course Materials

- "Program or be Programmed," Excerpts and Video. Douglas Rushkoff. http://www.shareable.net/blog/program-or-be-programmed
- "Big Data, Inequality, and the Law" Latanya Sweeny (first 15 minutes ONLY) https://vimeo.com/146814921

March 21: Final Exam