This version: March 30, 2018

Political Science 5D/Economics 5
Introduction to Social Data Analytics

Instructor Name: Arushi Kaushik
email id: arkaushi@ucsd.edu
Office: Econ 124
Office Hours: Mondays, 11:00AM- 1:00PM

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 on Monday's at 9AM. Late submission will not be accepted under any circumstances. Each student will be allowed to drop one problem set grade to accommodate for special circumstances.
  - Copies of the homework write-up and accompanying code should be turned in electronically via TritonEd by the due date.
  - 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 **May 2, 2018** covering the material in the first half of the class.
- **Final (30%):**The final for this class is **June 13, 2018 from 8:00AM-11:00AM**.
- **Reading Quizzes (5%):** Short, take home reading quizzes will be given once a week, due Wednesday and will cover that week's readings. Quizzes will be released on Friday and will be due on Wednesday morning.
- **Participation (10%):** Attendance in lecture and section is mandatory. Attendance will be taken. Students are strongly encouraged to ask questions and actively participate in discussions during lectures and sections.

1 First Assignment

Please fill out this survey:

`https://goo.gl/forms/2PbDjjXJCi3kCK2P2`

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.

Course Website and Piazza Forum

Syllabus and course materials. The syllabus, assignments, solutions, and other course materials will be posted on Piazza. Assignments will be turned in via TritonEd.

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 directly from the below address (there are also free Piazza apps for the iPhone and iPad):

`https://piazza.com/ucsd/spring2018/poli5decon5`

Using Piazza will allow you to see and learn from other students' questions. The instructors will regularly check the board and answer questions posted, although everyone else is also encouraged to contribute to the discussion. You can opt to send a question only to your class, or to send it to the wider group of students in Poli5D/Econ5. 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* (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: *Quantitative Social Science- An Introduction*, by Kosuke Imai (Princeton University Press)

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/`

Dates to Remember

- April 11 9AM: Reading Quiz 1 Due: Covers Week 1 and 2 Readings.
- April 16 9AM: Problem Set 1 Due
- April 18 9AM: Reading Quiz 2 Due: Covers Week 3 Readings.
- April 23 9AM: Problem Set 2 Due
- April 25 9AM: Reading Quiz 3 Due: Covers Week 4 Readings.
- **May 2 9:00 AM-9:50AM**: Midterm
- May 9 9AM: Reading Quiz 4 Due: Covers Week 6 Readings.
- May 14 9AM: Problem Set 3 Due
- May 16 9AM: Reading Quiz 5 Due: Covers Week 7 Readings.
- May 23 9AM: Reading Quiz 6 Due: Covers Week 8 Readings.
- May 30 9AM: Reading Quiz 7 Due: Covers Week 9 Readings.
- June 4 9AM: Problem Set 4 Due
- June 6 9AM: Reading Quiz 8 Due: Covers Week 10 Readings.
- **June 13 8:00AM- 11:00AM**: Final

COURSE SCHEDULE

2 April 2: 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`

3 April 4: Data Format and Intro to Excel

Course Materials

- *Statistical Modeling: A Fresh Approach*, Daniel Kaplan. Chapter 2, 2.1-2.4. `http://www.mosaic-web.org/go/StatisticalModeling/Chapters/Chapter-02.pdf`

4 April 9: Functions in Excel

Course Materials

- "Introduction to Functions and Formulas" http://www.excel-easy.com/introduction/formulas-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

5 April 11: Functions in Excel

Course Materials

- "Introduction to Functions and Formulas" http://www.excel-easy.com/introduction/formulas-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

6 April 16: Introduction to Stata and Reproducibility

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, first 7 minutes. https://www.youtube.com/watch?v=XmvWydFD2Y0&index=6&list=PLRYSxJ3XjgQM342QrBkzek8clHa5ue4Sd

7 April 18: Description in Stata and If Statements

Course Materials

- Bill Gates Explains If Statements, Hour of Code, https://www.youtube.com/watch?v=m2Ux2PnJe6E
- Data Management in Stata, http://data.princeton.edu/stata/dataManagement.html

8 April 23: 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

9 April 25: Regression in Stata

Course Materials

- "Introduction to Residuals and Least Squares Regression," Khan Academy, https://www.youtube.com/watch?v=yMgFHbjbAW8, 7 minutes.
- "Simple and Multiple Regression in Stata," Section 1.0 and 1.3 https://stats.idre.ucla.edu/stata/webbooks/reg/chapter1/regressionwith-statachapter-1-simple-and-multiple-regress

10 April 30: Midterm Review

11 May 2: Midterm

12 May 7: Introduction to R

Course Materials

- "Data Analysts Captivated by R's Power" *The New York Times* http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html
- Imai, 1.3.1-1.3.3

13 May 9: Analysis of Experiments by Subsetting Data in R

Course Materials

- Imai, 2.1-2.2

14 May 14: Visualizing Data in R

Course Materials

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

15 May 16: For Loops and If Statements 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&`

16 May 21: For Loops and If Statements in R

Course Materials

- Imai, 4.1

17 May 23: Regression in R

Course Materials

- Imai, 4.2

18 May 28: Memorial Day, No Class

19 May 30: Functions in R

Course Materials

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

20 June 4: Functions in R

Course Materials

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

21 June 6: Putting it All Together and Final Review

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`