Economics 5/Political Science 5D
Introduction to Social Data Analytics
UC San Diego
Winter 2021 Online[1]

Instructional Team

Instructor	Virtual Office Hours	Zoom Meeting ID
David Arnold	Wednesday 9-10 AM	930-2420-3620

Teaching Assistant	Virtual Office Hours	Zoom Meeting ID
Anjali Pai	Wednesday 1-2 PM	TBA

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.

Normally, I consider the syllabus my contract with the class. However, this quarter some flexibility might be desirable due to all of the uncertainty related to the pandemic. The syllabus that follows is my current prediction of how the class will be structured and assessed. While I will do what I can to keep to the predicted assessments for this course, the evolving situation may make it necessary for me to make changes. If that happens, I will make sure to inform you as early as possible, and to explain as best as I can the rationale behind the change.

Lectures and Labs

There will be two classes per week. The lecture will be recorded and made accessible on Canvas. You do not need to attend lecture to participate in the course, but are encouraged to in order to ask questions. Many lectures will will be accompanied by an activity or dataset that you should download from Canvas before beginning the lecture. While the lecture is running, you should follow along and complete the activity.

In addition to lecture, each week there will be a lab. Each lab will involve completing an Excel workbook, Stata Do-file, or R script. During the virtual lab session, students will be assigned to "breakout rooms" in small groups with whom they may collaborate on completing the lab exercises.

---

[1]This course will be held entirely online in response to the university's efforts to limit the spread of COVID-19.

Participating in lab is part of your grade. If you attend a virtual lab meeting, then you receive full credit for that lab. If you cannot attend a virtual lab for a given week, then you can still get lab credit by turning in the answers to the lab on Canvas. The labs will be graded based on completeness. Please "show up" (virtually) to the lab session on-time to facilitate assignment to breakout rooms. The lab will be due on Sunday at 11:59PM every week for those who are unable to attend the lab session. If you attend lab each week, then you do not need to turn in any lab materials. There is no lab for week one.

Assessment

Your grade will be based on a combination of:

- **Homeworks (40%):** Three 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:

  - Homeworks will be due one week after they are posted. Each day a homework is late will reduce the grade by 10 percentage points. Homework will not be accepted if it is more than three days late.
  - Copies of the homework write-up and accompanying code should be turned in via the Gradescope tab on Canvas by the due date.
  - Although it is permissible to discuss conceptual questions with other students enrolled in class, each student must submit their own writeup of the solutions that shows their independent work on the assignment. In particular, one should neither copy someone else's answers or code nor share their answers or code with anyone. We also ask you to write down the names of the other students with whom you solved the problems together at the top of your solutions submission. Solutions/code that appear overly similar between students will be reported to the Academic Integrity Office.

- **Take-home Final Project (40%):** Students will complete an independent project that demonstrates mastery of the material taught during the quarter. The project will be due on Thursday, March 18th at 6:00PM, but updates will be due throughout the quarter with homework submissions. See the final project prompt for specific details on the final project. Late submissions will lose a letter grade for every day (or part thereof) late. No submissions more than three days late will be accepted.

- **Lab and Quizzes (20%):**

  - *Lab (10%)*: If you attend lab each week then you will receive full credit for lab participation. If you can't attend lab, but make sure to turn in a lab assignment each week that is complete and shows effort, then you will receive full lab participation. Partial credit will be given for labs that are incomplete or show a lack of effort.
  - *Quizzes (10%)*: There will be 5 short quizzes throughout the quarter. You will be able to drop the grade of the lowest quiz. Quizzes will be posted early in the week and are due on Friday. In general, they will be around 5-10 questions and you can take each quiz multiple times to improve your grade.

Academic Honesty and Plagiarism

All of your graded work must be done by you. To be explicit, *sharing solutions or code are violations of academic integrity* and will be reported. If you are unfamiliar with the University's policy on academic integrity, please see `http://senate.ucsd.edu/Operating-Procedures/Senate-Manual/Appendices/2`.

Course Website and Piazza Forum

Syllabus and course materials. The syllabus, assignments, solutions, and other course materials will be posted on Canvas. All assignments will be turned in via the Gradescope tab on Canvas. If you are not familiar with Gradescope, take some time to familiarize yourself before the first assignment is due (`https://www.youtube.com/watch?v=u-pK4GzpId0&feature=youtu.be`).

Online Q&A. Given the online format of this class, it will be immensely helpful to share questions and answers on a platform that all students may access. Throughout this class we will use the Piazza online discussion board. Piazza 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 images, attaching files, and customized email frequencies. We encourage you to ask questions on the Piazza forum for clarifications, conceptual questions, or figuring out methods related to your project. Please do NOT share code that would give away a solution to a homework assignment that has not yet been graded. You may join the Piazza page for our course directly from the below address (there are also free Piazza apps for the iPhone and iPad):

<div align="center">

`https://piazza.com/class/kigkwbmq9z946x`

</div>

Please do not email questions directly to the instructional team. Others likely share similar questions, so it is preferred that you share your question publicly on Piazza. If your question is of personal nature (e.g. related to a grade), you may set the visibility of your post to be seen only by instructors and yourself.

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: *A First Course in Quantitative Social Science*, by Kosuke Imai (Princeton University Press)

Software

This course will consist of three different statistical software programs commonly used by social scientists: Excel, Stata, and R. Excel and Stata both require licenses that are available for free to UCSD students. R is open-source and is free to everyone. Instructions on how to install the three software packages will be shared on Canvas.

A note on recordings. All lectures, labs, and office hours will be delivered online through Zoom. You can have access to the meeting links through the "Zoom LTI PRO" navigation tab in Canvas. You do not need to be physically on campus for any of the activities pertaining to this course. All lectures will be recorded and posted on Canvas. Office hours and lab sessions will *not* be recorded.

Important Due Dates

- If you do not attend lab, then weekly lab due Sundays at 11:59PM

- Homework 1 (Excel) assigned Friday January 15th, due Friday January 22nd at 11:59PM

- Homework 2 (Stata) assigned Friday February 5th, due Friday February 12th at 11:59PM

- Homework 3 (R) assigned Friday February 26th, due Friday March 5th at 11:59PM

- Final Project due Thursday, March 18th at 6:00PM

COURSE SCHEDULE–Subject to Change

The schedule below lays out what is covered each week in terms of coding and software usage skills. Each class will also be accompanied by an empirical application. A number of these applications will be on topics that social scientists are currently working on, aided by the explosion of data in recent years. These topics will including voting patterns, opioid usage, unemployment, intergenerational mobility, discrimination, the impacts of education, among others.

Week 1: Introduction to Data Analysis and Excel

- Discuss content of course and how data can be used to adress important social problems

- Introduce Excel: spreadsheets, types of variables, and functions

Week 2: Excel and Data Visualization

- Introduce topic of sampling and representative populations

- Pivot tables and pivot charts

- More complicated functions (VLOOKUP)

- Constructing graphics in Excel

Week 3: Conclude Excel and Introduce Stata

- Plot figures in excel

- Introduce how to write, save and edit do-files in Stata

- Basic functions and data manipulation

Week 4: Data Visualization/Manipulation and Data Wrangling

- Data manipulation: how to generate new variables and manipulate current variables

- Data visualization: how to interpret and plot histograms

- Data wrangling: how to merge and append datasets

Week 5: Finish Data Wrangling and Begin Regression in Stata

- More functions to reshape data

- Univariate regression

- Forming predictions and plotting results

Week 6: Finish Regression in Stata

- Hypothesis Testing

- Interpreting Regression output: causality vs. correlation

Week 7: Introduce R and Data Wrangling

- Locate and identify the essential parts of the Rstudio interface

- How to download packages

- Write and interpret logic statements in R

- Introduce the package "dplyr" for data wrangling

Week 8: Data Visualization in R

- Introduction to "ggplot2" package and how to construct barplots, histograms, boxplots, line plots and scatter plots

Week 9: Loops and Regression in R

- Automating code using loops

- How to use iteration counters to simplify your code

- How to perform linear regression in R

Week 10: Conclude R and Overview Class

- How to write a function in order to automate your code

- Concluding thoughts and discuss the frontier of data analysis in social sciences