



**DEPARTMENT OF ECONOMICS / DEPARTMENT OF POLITICAL SCIENCE
ECON 5 / POLI 5D (C00): DATA ANALYTICS FOR THE SOCIAL SCIENCES**

TIME(S) / LOCATION(S): TUESDAYS & THURSDAYS, 5PM-5:50PM / PODEM 1A20

**WINTER QUARTER 2024
SYLLABUS**

Instructor: Dr. Steve Levkoff, PhD, CAP®

E-mail: slevkoff@ucsd.edu

Course Webpage: <http://canvas.ucsd.edu>

Office: Atkinson Hall 4115

Office Hours: Tuesdays & Thursdays, 2:15PM-3:15PM, open door, and by appointment (no office hours in-person until week 2). If you could email me in advance to let me know you plan on showing up (and even what you'd like to address in office hours), that would be appreciated!

Course Description: This course has three main goals. The first goal is to introduce you to interesting and important social science questions. Each chapter will highlight a different application, often highlighting research from faculty at UCSD. We will cover a wide range of topics, including how colleges promote intergenerational mobility, what motivates people to vote, how do we identify discrimination in labor markets, among others. The second goal is to show you how data can be used to inform our discussion of these topics. We will be using a lot of different datasets in this course. Some come from experiments that have been run by researchers. Others come from administrative datasets collected by governments. For many of the datasets, we will only have time to scratch the surface of what is possible with this data. The third goal is to give you the tools to perform data analysis. We will focus on three popular software: Excel (1 chapter), Stata (4 chapters) and R (5 chapters). While learning coding fundamentals in each of these programs, we will shed light on big social science questions.

Prerequisites: You should be familiar with basic algebra and arithmetic in being able to solve equations, compute fractions & percentages, and interpret growth rates. Otherwise, there are no formal course prerequisites and no programming expectations.

Readings:

Required:

- [1] The textbook for the course was written in-house (Department of Economics) and can be found free online here:

<https://daarnolducsd.github.io/ep5book/>. The course will follow the text very closely.

Technology Prerequisites: It is your responsibility to make sure you are enrolled in the online course (Canvas) by registering for the class formally through the university. I will not provide access to the Canvas course if you are not already enrolled or on the wait list to avoid complications with the course roster and to routinely check it and your email for announcements and to access newly distributed material. *You must also make sure that your browser is compatible with the Canvas quiz assessment system.* In the past, some students had browsers that did not display the math text. Any questions regarding Canvas issues or general technology issues should be directed to ITS@ucsd.edu. We will also be using Excel, Stata, and R, and will provide instruction on Canvas in regards to how to install and access the software for the course (top module on Canvas). There are computers in the library that have these programs installed. The Data and GIS lab: <https://library.ucsd.edu/computing-and-technology/data-and-gis-lab/index.html> has computers that have these programs installed. However, if you attend lab in person, most of the lab will be using your laptop to complete a coding assignment. If you do not have a laptop that is capable of downloading R and Stata, there is also a possibility for a long-term loaner laptop from the University. Please go here: <https://library.ucsd.edu/computing-and-technology/computers-and-laptops/index.html> to receive more information.

Lectures, Podcasts, and Labs: There will be two lectures per week. All in-person lectures will be recorded and posted after the lecture via podcast. In addition, all the lecture material, as well as bonus material, is covered in a series of videos on Canvas. In certain cases, I will ask you to watch a video before arriving at class. Make sure you watch this video so that you are prepared for the material in lecture. Additionally, there will be reading material that accompanies each lecture. 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. If you attend in person, you will work on sections of the lab together with your classmates that attend in person. Each week you will need to answer a short quiz related to the lab. If you attend in person, the answers to the quiz will be discussed during the lab. If you cannot attend in person, you can still get credit for the lab by completing the lab and then completing the quiz related to that lab.

Lab Sections: Lab sections will be primarily focused on going through the weekly programming assignment related to the weekly lab quiz:

Times & Locations:

Fridays in WLH 2112 from 2-2:50PM

Teaching Assistant: Artur Bayramyan (abayramy@ucsd.edu)

COURSE DELIVERABLES

#FinAid Survey This is the first item in the first Canvas module. While not part of the course grade, you should submit a response to this survey as the university uses it for reporting the commencement of academic activity for the term for consideration of financial aid. Failure to do so may result in loss of consideration for future financial aid opportunities.

Quizzes: There will be weekly (roughly) concept check quizzes administered remotely (and asynchronously) via Canvas. The quizzes will involve applications and problem solving related to recently covered topics and will have 10 questions of a variety of formats. Each student will effectively have a different quiz using randomized question variants from a question pool. The quizzes will be available Wednesday at 12:00am (PST) and will remain open until 11:59pm (PST) on Friday. You will have only a single attempt at the quiz, so be careful when submitting your answers. You may use your notes and digital resources to complete the quiz. Your lowest quiz score will automatically be dropped at the end of the term. In the event you miss a quiz, it will count as your dropped quiz.

Lab Quizzes: There will also be weekly (roughly) lab quizzes on Canvas, which are typically shorter (3-5 questions) about the weekly lab, which can be done remotely / asynchronously by following the instructions carefully. If you attend the lab session live in-person, the TA will work through the lab and the answers will be revealed. Lab quizzes will be available on Tuesdays at 12:00am (PST) and will remain open until 11:59pm (PST) on Sunday. You will have only a single attempt at the quiz, so be careful when submitting your answers. You may use your notes and digital resources to complete the quiz.

Examinations: There will be two exams: a midterm (Thursday, 2/8/24 in lecture) and a final (scheduled by the registrar on Thursday, 3/21/24 from 7-10pm). The first exam focuses on the content introduced the first four weeks of the course. The final exam, while cumulative, will focus more on the R material introduced in the second part of the course.

Summary of Deliverable Deadlines:

- Quizzes – Available from 12:00am on Wednesdays through 11:59pm on Fridays (see the Canvas calendar for more details)
- Lab Quizzes – Available from 12:00am on Tuesdays through 11:59pm on Sundays (see the Canvas calendar for more details)
- Midterm Exam is on 2/8/24 in lecture
- Final Exam 3/21/24 from 7PM-10PM (C00) (NOTE: You must take the exam in the section of the course in which you are registered)

Grading:

Lab Quizzes	10%
Quizzes	15%
Midterm Exam (2/8/24)	25%

Final Exam (3/21/24)	50%
Total	100%

In the *past*, a student could typically guarantee themselves some type of A by ranking in the top 25% of students in the course and some type of B by ranking in the top 60% of students in the course. The median grade for the course typically winds up being around a B-. In the past, scoring a course composite score of 90% or above typically puts students in the A range. Similarly, scoring a course composite score of at least 80% tends to put students somewhere in the B range. In the past, scoring at least a 60% composite score for the course was the required threshold to pass the class (with C-). The instructor reserves the final judgement in all grade assessment matters. ***It is important to note that the Canvas grade center does not automatically apply these weights – you will need to compute this manually as outlined above.***

Midterm Exam Re-grade Policy: It should be known that there may be some questions on the exams that have no right or wrong answer, so how credit is awarded depends crucially on *how* you defended your answer. Accordingly, there is a *BIG* difference between an answer being a *technically correct answer* and an answer being ***the best answer***. In these cases, credit is awarded (according to the grading rubric) for how close your answer comes to being the *best answer*. That is, an answer, while being technically correct, may not necessarily have been the best answer and hence, wouldn't necessarily receive full credit despite technical correctness.

After your exams are graded, we will release your graded exams on *Gradescope* so you can view a .pdf of your graded exam. If you find that there was a minor grading issue (ie: points were added up incorrectly), let us know and we will remunerate accordingly. If there is a major issue (ie: you don't think points were awarded correctly according to the grading rubric), submit a re-grade request – be advised that in the event your request is rejected, the TA's have authority to deduct an additional penalty for asking for a regrade (only in the case where the request is denied or viewed as excessive).

You have one week from the announcement of the grade distribution to resolve grading issues. Otherwise, grades are NON-NEGOTIABLE and any requests that are determined to be *excessive* in scope may warrant further point deductions unless sound arguments are used to justify the request. The instructor and TAs reserve the right to refuse any request believed to be *excessive*.

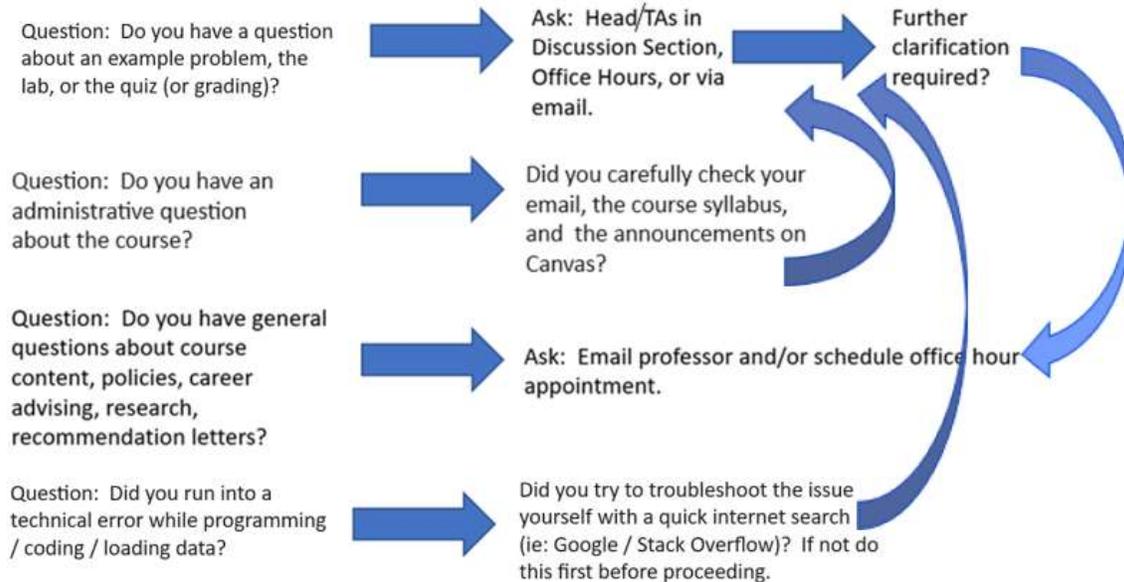
Absences & Attendance: *The asynchronous Canvas deliverables (Quizzes and Labs) cannot be made up for any reason.* Otherwise, absence from an exam assessment will be awarded a score of zero. The instructor reserves the right to re-weight the composite grading structure above to accommodate valid excused absences (ie: medical emergency, etc.). This could involve an oral exam or alternate format. To hedge risk, it is in the students' best interest to complete as many of the course assessments as possible.

Supplemental Material & Slides: Throughout the course, the instructor may post supplemental readings and slides via Canvas. These materials are meant to be used in

addition to the lecture and are not to be used as a substitute for going to lecture, reading the textbook, or watching required video content. The instructor reserves the right to remove access to this material if he feels that it has adversely affected attendance in the lecture. Students that perform the best in the course have historically taken maximal advantage of supplemental materials (especially the videos).

Classroom Decorum & Email: To avoid distracting others in the classroom, please arrive on time. When class is in session. I use a “call-and-response” type lecture style, and you should feel encouraged to participate when queried by the instructor with a leading question. Please restrict the use of email to the minimally necessary volume, put the course number (ECON 5 / POLI 5D) in the subject of the email – I teach other classes and this is the fastest way for me to know who I’m responding to. Put your *full name* at the end of email messages as well. Email questions regarding *how to do a particular homework /quiz / lab question* should be first directed to your TAs (so they have an opportunity to work within their purview). Your email decorum should be professional. You should make sure you address the email formally and properly and use college level grammar. The way to get the fastest response via email is to email me directly at slevkoff@ucsd.edu. Messaging me through Canvas requires that I log into Canvas to reply, and this is more cumbersome. I also spend about 2 hours per day responding to emails, so if you don’t get a reply immediately (I answer them in the order I get to them), please either follow up or wait. I should get back to you within no longer than three days at the most (especially if you ask over the weekend).

Who Should I Ask?: Please reference the flow chart below to efficiently utilize your course resources when making inquiries:



Statement of Academic Integrity: Integrity of scholarship is essential for an academic community. The University expects that both faculty and students will honor this principle

and in so doing protect the validity of university intellectual work. For students, this means that all academic work will be done by the individual to whom it is assigned, without unauthorized aid of any kind.

Examination Policies: Consistent with the University's mission to preserve academic integrity, there are several policies and procedures that must be adhered to by students during exams.

- 1) In order to be allowed into the exam, students must have:
 - A BASIC or SCIENTIFIC calculator (no graphing calculators, cell phones, or other mobile devices unless given prior approval by the instructor)
 - Your UCSD student ID
- 2) During the exam, the following policies will be enforced:
 - Your seat will be randomized for each exam. When you enter the lecture on exam day, find your name and assigned seat number on the projector and quickly and quietly sit. Once everyone is seated, the exam will be handed out.
 - NO BATHROOM BREAKS after the first 45 minutes following the exam start. Be sure to use the restrooms before the exam begins. Exams are less than an hour and a half long! You can make it!!!
 - No hats, hoodies, or sunglasses during the exam.
 - Turn cell phones off during the exam and leave them in your bag.
- 3) Violations of academic integrity will not be tolerated. For this course in particular, violations include, but are not limited to anything that may be perceived as the following actions:
 - looking at or copying from other students' exams
 - talking during an exam while exams are still out
 - looking at notes during an exam
 - taking the wrong version of an exam
 - removing an exam from the examination room
 - removing pages from an exam
 - falsifying identification or an exam book during or after the exam
 - sitting in the wrong seat during an exam (if applicable)
 - using an unapproved device/item during an exam (ie: programmable calculator, cell phone, etc. - see above list)

Violation (or perceived violation) of any of the abovementioned policies will be enforced via zero tolerance, will result in an automatic zero on the exam/assessment in question, an automatic failing grade for the course, and may be referred to the student conduct process through the AIO office, so don't do anything that would even come close to something that an observer would potentially interpret as academic dishonesty. NO EXCEPTIONS.

Tentative Schedule of Topics (Subject to Change):

Week 1: Introduction to Excel

- Empirical Application: Instructor Incentives and Student Performance, by Andy Brownback and Sally Sadoff (2020)
- Data tables
- Functions
- Pivot tables

Weeks 2: Introduction to Stata

- Empirical Application: Intergenerational Mobility Rates by College (Opportunity Insights)
- The Stata Graphical User Interface (GUI)
- Do-files
- Basic data analysis commands
- Interpret and constructing histograms

Week 3: Data Wrangling in Stata

- Empirical Application: Racial Discrimination in Traffic Stops (Stanford Open Policing Project)
- Introduce the concept of data wrangling
- Learn the append, merge, and collapse commands
- Bar charts in Stata
- Ways to improve data visualization in Stata

Week 4: Regression in Stata

- Empirical Application: Disrupting Education using Technology, by Muralidharan, Sing, and Ganimian (2019)
- Estimate and interpret linear regression in Stata
- Introduce concept of fitted values and residuals
- Visualize and plot the results of regressions in Stata

Week 5: Binned Scatter Plots (MIDTERM WEEK)

- Empirical Application: The Legacy of Colonial Medicine by Lowes and Montero (2021)
- Binned Scatter Plots
- Missing values and value labels

Week 6: Introduction to R

- Empirical Application: Resume Experiments, by Bertrand and Mullainathan (2004)
- Objects and variables in R
- Introduction to data frames
- Subsetting data frames in R

Week 7: Data Wrangling in R

- Empirical Application: The Rug Rat Race, by Garey Ramey and Valerie Ramey
- If statements
- For loops
- Introduction to the tidyverse package

Week 8: Data Visualization in R

- Empirical Application: China's War on Air Pollution by Greenstone He, Jai, and Liu
- Histograms, Scatter Plots, Box Plots

Week 9: Linear Regression in R

- Empirical Application: The Butterfly Ballot
- Linear Regression in R
- Plotting the regression line
- Predicted values and residuals

Week 10: Functions in R

- Empirical Application: The Impact of Unconditional Cash Transfers by Haushofer and Shapiro
- Building your own functions in R