Data Wrangling for Economics

Syllabus

Richard Startz

Fall 2020

Real economic data is often messy. It can require significant filtering and re-arranging before it can be analyzed. In this course we learn techniques to accomplish this. We focus on the what to do questions, such as the importance of converting data from “wide” to “tall.” Along the way, we learn some answers to the ‘how to do it’ questions. We will learn some basic use of the R computer environment, although once you understand the issues you will be able to apply concepts in other environments as well. We will learn some very basic techniques of data description, analysis, and presentation. But this course is not a substitute for a computer science course nor for courses in econometrics/statistics. It’s all about the nitty-gritty of dealing with real data in economics.

Data Wrangling for Economics is a learning by doing course. To be successful in this course you should like to learn concepts and you should like to solve puzzles. Pretty much everything is open book, so memorization is of very little importance. This course is open to individuals of all skill sets. No prior R experience is required! In fact, this course is designed for individuals with no prior R experience.

Student Learning Objectives and Purpose of Class

Students will learn how to:

a) Organize data for computer analysis.

b) Display and summarize data in order to answer substantive questions.

c) Communicate findings clearly and succinctly.
Course Organization

The course has a number of moving parts.

1. Asynchronous lectures are on GauchoSpace. Watch them when convenient.

2. Section “meetings” are scheduled for Wednesday and Thursdays at 6:00 pm Pacific time. These will be synchronous zoom meetings. Attendance is not required and no new material will be covered in section meetings. Sections will begin the week of Oct. 5.

3. Relaxing beverage hour is a synchronous zoom meeting each Tuesday at 5:00pm with Professor Startz to discuss anything you’d like to discuss...except R and the specifics of the homework (those belong on Nectir). Concepts in data wrangling, questions about the lecture material, advice on careers, and the state of the world are all fair game. Think of this as zoomed office hours. BYOB.

4. Guided exercises are on GauchoSpace. Each one walks you through material relevant to the course and to assignments. We strongly suggest that you code along with the examples that we go through, although there will be times when copying and pasting can be useful (e.g. loading in a data set). Spending time on the guided exercises is strongly recommended. You do not turn in anything from the guided exercises.

5. Homeworks: You will receive an email with data for a homework assignment weekly. Along with each homework you will receive a task checklist. For each homework (unless otherwise specified) you will:
   a. Upload R code to GradeScope. Gradescope will grade your code and give you a little bit of feedback. You can turn code in to GradeScope as often as you like, only the highest scoring submission counts for a grade.
   b. Create a written analysis. You upload your draft analysis to Eli Review through GauchoSpace and then receive constructive comments from classmates. Then you turn in a final draft through GradeScope.
   c. Give two classmates constructive advice on their drafts on Eli Review. The idea is to be as helpful as possible to your classmates. Comments are anonymous and meant to develop writing skills in a safe, nonjudgmental environment. Before submitting your final write-up version, you might find it helpful to consider any suggestions that your peers had on your writing. It is also true that reading the writing of others and reflecting on what you like and what don’t like is an excellent way to improve your own writing.

6. Simulation week. During the week of November 16 - 22, you will receive multiple analysis assignments. As is often true in organizations, these assignments will be short-fuse, time-critical. Be sure to set aside time for the needed tasks. These assignments count for more than one homework. You will use Eli Review and turn in assignments to Gradescope.

7. Online implicit bias test. You receive participation credit for this assignment. You are not graded on your answers.
8. Final project. The final project is an extended homework. In addition to turning it in for a grade, think of your final project as giving you a writing sample and some talking points you can use as part of a job search. This assignment counts for more than one homework. You will turn in the assignment (e.g. coding + write-up) to Gradescope. There is no Eli Review for the final project.

Necessary equipment

Basically, you need an internet connection and some kind of computer. You need something with a keyboard and a bigger screen is better. (A smart phone won’t do.) You need software to create a pdf document, but almost any word processor will do that.

If you do not already have suitable equipment and/or a useable internet connection go to https://keeplearning.id.ucsb.edu/2020/03/22/internet-access-technology-help/ for assistance.

Collaborating and getting help

- You are strongly encouraged to collaborate on assignments with classmates. Having discussions on Nectir is one way to do this. However, you need in the end to write your own code and your own analysis. See “Academic integrity,” below.

- All questions are fair game for you to post on Nectir. (Do keep it PG, which can sometimes be hard while coding.) And drop in when you can to help out a classmate! Course staff members will monitor Nectir as regularly as the budget allows. So we hope that between classmates and staff that response rates will be relatively quick.

- For administrative questions about the course—not about help with homework or R—send email to econ-econ145@ucsb.edu. We will try to get you a response within 24 hours.

Academic integrity

- Copying a couple of lines of code is fine...copying more is not. When you copy code, it is considered a professional courtesy to include a comment citing the source.

- The system we use checks for plagiarism. If you cheat, consequences include failing the course and being suspended from the University—for a first offense.

- All course materials (class lectures and discussions, handouts, examinations, web materials) and the intellectual content of the course itself are protected by United States Federal Copyright Law, the California Civil Code. The UC Policy 102.23 expressly prohibits students (and all other persons) from recording lectures or discussions and from distributing or selling lectures notes and all other course materials without the prior written permission of the instructor (See http://policy.ucop.edu/doc/2710530/PACAO5-100). Students are permitted to make notes solely for their own private educational use. Exceptions to accommodate students with disabilities may be granted with appropriate
documentation. To be clear, in this class students are forbidden from completing study guides and selling them to any person or organization. This text has been approved by UC General Counsel.

- We know perfectly well that the vast majority of students conduct themselves with integrity without being lectured about it. Apologies to you. If it is any consolation, the economics department is surprisingly good at catching cheaters.

**Topics**

- Introduction to R
- Summarizing and cleaning data
- String fiddling
- Reproducible results
- Scaling and weights
- Making data tall
- Dealing with heterogeneity
- Visualization of data
- Debugging
- Programming techniques
- Relational databases

**Grade points**

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homeworks</td>
<td>100 (x7)</td>
</tr>
<tr>
<td>Eli Review Comments</td>
<td>100</td>
</tr>
<tr>
<td>Implicit Bias Test Participation</td>
<td>100</td>
</tr>
<tr>
<td><strong>Simulation Week</strong></td>
<td>200</td>
</tr>
<tr>
<td>Final</td>
<td>300</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1400</td>
</tr>
</tbody>
</table>

Each assignment will consist of two parts: a coding portion and a written portion. The coding portion may be submitted to gradescope an unlimited amount of times. The last submission will be counted. The written portion will be graded based on addressing the prompt, clarity of writing, and organization.

**Late work**

Generally, it is a bad idea to ever turn anything in late. However, the *coding* portion of each homework may be turned in up to 1 week late for half credit. For instance, suppose Homework 1 (coding portion) is worth 60 points. If you turn this in late and score a 44/60, you will get a total of 22/60 points.
Insurance policy

Do not wait for the last minute to do assignments. The internet goes down. Computers crash. People get sick. (Given the way 2020 has been going, locusts may land and carry off your laptop.) The grades above come with an automatic no-request-needed insurance policy. Basically, if you miss a homework, we add back in 100 points. More specifically, up to 100 points missed anywhere in the course for any reason will automatically be restored. (Obviously, you don’t want to use up your insurance policy frivolously, as something might go wrong for real later in the term.)

Textbooks

The following textbooks are highly recommended, but not required:

*R for Data Science, 2nd edition.* 2016. Hadley Wickham and Garrett Grolemund, O’Reilly Media or https://r4ds.had.co.nz/.


The following books are also useful:


Useful cheat sheets can be found at https://rstudio.com/resources/cheatsheets/

Weekly Schedule

Every Sunday evening, you will receive an email from econ-econ145@ucsb.edu discussing the assignments for the week. This includes a list of lectures to watch, guided exercises to go through, and assignments due including your individually crafted datasets. The coding portion and written portion of homework assignments will be uploaded to Gradescope no later than Friday of that week. The written portion will also be uploaded to Eli Review no later than Friday of that week. Comments on Eli Review are due the following Wednesday. Your final written analysis is due Friday, two days after comments are on Eli Review.

The overall course schedule is provided below. Note that while you will receive data for assignments throughout the term, you are free to work through lectures and guided exercises as early as you like.
• Week 0 (Sept. 28 - Oct. 4):
  □ Watch lecture 0.

• Week 1 (Oct. 5 - 11):
  □ Watch lectures 1-2.
  □ Complete Guided Exercises 1.
  □ Complete the Implicit Bias Test link on Gauchospace (**due date: Friday, October 9**).
  □ Submit Homework 1 (coding portion) to Gradescope (**due date: Friday, October 9**).
  □ Submit Homework 1 (write-up) to Eli Review (**due date: Friday, October 9**).

• Week 2 (Oct. 12 - 18):
  □ Watch lectures 3-6.
  □ Complete Guided Exercises 2.
  □ Complete the peer review of Homework 1 write-up via Eli Review (**due date: Wednesday, October 14**)
  □ Submit Homework 2 (coding portion) to Gradescope (**due date: Friday, October 16**).
  □ Submit Homework 2 (write-up) to Eli Review (**due date: Friday, October 16**).
  □ Make edits on Homework 1 write-up based on Eli Review peer review and submit to Gradescope (**due date: Friday, October 16**).

• Week 3 (Oct. 19 - 25):
  □ Watch lectures 7-9 which are linked on the Gauchospace page.
  □ Complete Guided Exercises 3.
  □ Complete the peer review of Homework 2 write-up via Eli Review (**due date: Wednesday, October 21**)
  □ Submit Homework 3 (coding portion) to Gradescope (**due date: Friday, October 23**).
  □ Submit Homework 3 (write-up) to Eli Review (**due date: Friday, October 23**).
  □ Make edits on Homework 2 write-up based on Eli Review peer review and submit to Gradescope (**due date: Friday, October 23**).

• Week 4 (Oct. 26 - Nov. 1):
  □ Watch lectures 10-15.
  □ Complete Guided Exercises 4.
  □ Complete the peer review of Homework 3 write-up via Eli Review (**due date: Wednesday, October 28**)
  □ Submit Homework 4 (coding portion) to Gradescope (**due date: Friday, October 30**).
  □ Submit Homework 4 (write-up) to Eli Review (**due date: Friday, October 30**).
  □ Make edits on Homework 3 write-up based on Eli Review peer review and submit to Gradescope (**due date: Friday, October 30**).
• Week 5 (Nov. 2 - 8):
  □ Watch lectures 16-19.
  □ Go vote (due date: Tuesday, November 3)!!
  □ Complete Guided Exercises 5.
  □ Complete the peer review of Homework 4 write-up via Eli Review (due date: Wednesday, November 4).
  □ Submit Homework 5 (coding portion) to Gradescope (due date: Friday, November 6).
  □ Submit Homework 5 (write-up) to Eli Review (due date: Friday, November 6).
  □ Make edits on Homework 4 write-up based on Eli Review peer review and submit to Gradescope (due date: Friday, November 6).

• Week 6 (Nov. 9 - 15):
  □ Watch lectures 20-22.
  □ Complete Guided Exercises 6.
  □ Complete the peer review of Homework 5 write-up via Eli Review (due date: Wednesday, November 11)
  □ Complete Homework 6.
  □ Submit Homework 6 (coding portion) to Gradescope (due date: Friday, November 13).
  □ Submit Homework 6 (write-up) to Gradescope!! There is no Homework 6 Eli Review! (due date: Friday, November 13).
  □ Make edits on Homework 5 write-up based on Eli Review peer review and submit to Gradescope (due date: Friday, November 13).

• Week 7 (Nov. 16 - 22):
  □ Submit Simulation Week Assignment Part 1 to Gradescope (due date: Wednesday, November 18).
  □ Submit Simulation Week Assignment Part 2 to Gradescope (due date: Friday, November 20).

• Week 8 (Nov. 23 - 29):
  □ THANKSGIVING WEEK

• Week 9 (Nov. 30 - Dec. 6):
  □ Watch lecture 23.
  □ Complete Guided Exercises 7.
  □ Submit Homework 7 (coding portion) to Gradescope (due date: Friday, December 4).
  □ There is no write-up for Homework 7.

• Week 10 (Dec. 7 - 13):
  □ Choose Final Project and begin.
• Finals Week (Dec. 14 - 20)
  □ Turn in Final Project to Gradescope (due date: Wednesday December 16).
  □ There is no Eli Review for the Final Project.