Econ 5/Poli 5D
Data Analytics/ Social Sciences

Shubhro Bhattacharya
UC San Diego

Lab-2
January 15, 2026

Today's Agenda & Motivation

- What kinds of questions economists answer using large administrative data
- How raw data gets translated into meaningful economic objects
- Why distributions often matter more than averages
- How Stata fits into the empirical workflow
- How today's lab is structured and what you should focus on

Who I Am

- PhD student in Economics at UC San Diego
- Work in applied micro and development economics
- Most of my research involves working with large datasets and writing Stata code

Why I care about this lab

- Learning Stata is mostly about practice, not memorization
- My goal is to make the software feel less intimidating and more usable

Why Opportunity Insights Data?

- Opportunity Insights uses linked administrative tax data
- Allows researchers to study intergenerational mobility at scale
- Lets us compare outcomes across colleges in a systematic way
- Forces us to think carefully about measurement, not just results
- Widely used in modern empirical economics research

Quintiles: Concept and Construction

- A **quintile** splits the data into five equal-sized groups (20% each)
- Individuals are ranked by income (or another variable)
- Quintiles capture relative position in the distribution
- In Stata, quintiles are created by ranking and binning observations
- Example command: `xtile income_q = income, nq(5)`

Five Important Stata Principles

- Always understand what data you have loaded before doing anything else
- Use do-files so your work is reproducible
- Do-files should be self-contained
- Very important to organize folders early on
- Expect errors — debugging is part of the process

Next: we now jump straight into the do-files and start working in Stata.

Quick Survey

What skills do we focus on in this class?