

# ECON 5/POLI 5D DATA ANALYTICS/ SOCIAL SCIENCES

SHUBHRO BHATTACHARYA  
UC SAN DIEGO

LAB-3

JANUARY 22, 2026

# QUICK SURVEY

What did we learn from Lab-2?



# TODAY'S AGENDA & MOTIVATION

- How economists use large administrative data to study policing
- Why **merging datasets** is essential in real empirical work
- What today's lab emphasizes: workflow, inspection, merges, and checks
- Then we dive straight into the do-file and fill in the **XX**'s together

# FROM RAW STOPS TO ANALYSIS VARIABLES

- Raw data is not analysis-ready: variables live across multiple files
- Typical structure:
  - File A: stop details (who, when, where)
  - File B: search outcomes (searched? contraband found?)
- Research questions require **constructing outcomes** (e.g., search rates)
- Main skills today:
  1. Merging datasets
  2. Creating analysis-ready variables

# SKILL 1: MERGING DATASETS

- Goal: join information from multiple files into one analysis dataset
- **Key step: identify a unique ID** that matches observations across files
- Use `isid <id>` to test if an ID uniquely identifies rows
- Decide merge type: 1:1, 1:m, or m:1 based on uniqueness
- Always inspect `_merge` after merging (what matched vs. didn't)

## SKILL 2: MAKING VARIABLES ANALYSIS-READY (STRINGS $\rightarrow$ NUMBERS)

- Many datasets store key variables as **strings** (e.g., "TRUE", "FALSE")
- For analysis, create **numeric indicators**: 0/1 (or labeled categories)
- Always verify: `tab <var>` and `summ <var>` after recoding
- Rule: **analysis happens on numeric variables**, not raw strings

# HOW TO ACCESS COURSE MATERIALS?

Course materials on my website

