Introduction and Causality

Introduction to Quantitative Social Science

Kosuke Imai
Harvard University / University of Tokyo

Summer 2022

Today's Quantitative Social Science

- Data, Data, and Data
- Past: government data, national survey data
- Today: more of old types of data and lots of new data
    - Randomized experiments and surveys conducted by researchers
    - Administration records: voter files, contributions, lobbying, ...
    - Economic data: trade, company information, finance, ...
    - Military data: casualty, insurgent attacks, ...
    - Social media data: websites, blogs, tweets, cell phones, ...
    - GIS data: satellite, climate, natural resource discoveries, ...
    - Text, images, sounds: news, speeches, bills, commercials, ...

- The problem is not about data
- Need for new substantive ideas
- Need for new data analysis tools

What You Need for Good Quantitative Research

- Theoretically motivated questions and puzzles
- Data nobody else analyzed
- Innovative methods that get most out of the data

- Some recent examples from successful dissertation:
  - Texts of court opinions and rulings
  - Lobbying reports, firm-level data, and product-level trade data
  - GIS data about the boundaries of European states
  - Campaign contribution data and a survey of contributors
- My own ongoing research:
  - Airstrikes, insurgency attacks, economic aid in Iraq and Afghanistan
  - Impacts of machine learning algorithms on human decisions in US criminal justice system
  - Analysis of television and online political campaign commercials
  - Methods for assessing legislative redistricting plans
  - Experimental evaluation of India's national health insurance program

What This Course is About

- **Introduction** to quantitative social science
  - no pre-requisite
  - little math
  - focus on data analysis

- Emulate the undergraduate introductory course at Princeton/Harvard
  - 6 weeks instead of 12 weeks
  - It moves at a fast pace. You must keep up!

- Review of the syllabus

What (I Hope) You Have Done Before the Class

1. **Pre-class Exercise** via **qsslearnr**:
   - Installed **R** and **RStudio** on your computer
   - If unsuccessful, use **RStudio Cloud**
   - Read QSS Chapter 1 and tried the code in it
   - Completed **qsslearnr** lesson `Tutorial0: Introduction to` **R**
   - Uploaded your answer in a PDF file to **GradeScope**

2. Also, try **qsslearnr-tidy**

3. **Persall**
   - Announcements, course materials, Q&As
   - Check previous posts (and Google!) before asking questions
   - Help each other

Agenda

1. In-class Exercise: Bias in Self-reported Turnout
   - Measuring turnout (QSS Section 1.4.1)
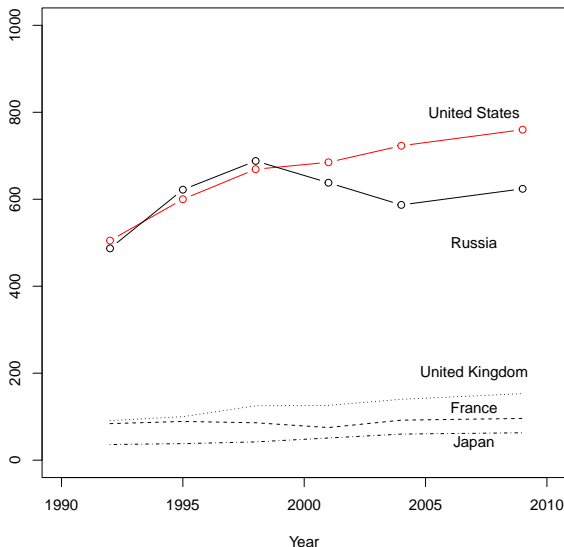   - Data set `turnout.csv` available at Perusall

2. Causality Preview
   - Counterfactuals
   - Fundamental problem of causal inference

Measuring the Turnout in the US Elections

- Question: How do you measure turnout rate?

- Numerator: Total votes cast
- Denominator:
  1. Registered voters
  2. VAP (voting-age population) from Census
  3. VEP (voting-eligible population)

- VEP = VAP + overseas voters − ineligible voters
  - overseas voters: military personnel and civilians
  - ineligible voters: non-citizens, disenfranchised felons, those who failed to meet states' residency requirement, etc.
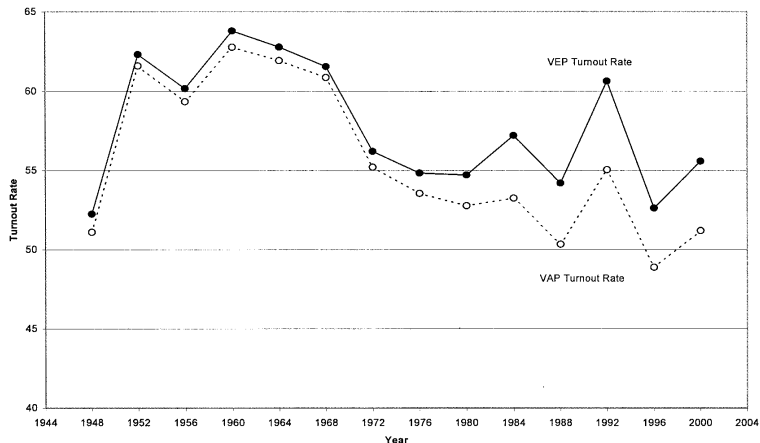
Growing Prison Population

Prison population per 100,000 inhabitants (OECD)

VAP and VEP Turnout Rates are Different

FIGURE 1. National VAP and VEP Presidential Turnout Rates, 1948–2000

McDonald and Popkin (2001) *American Political Science Review*

Bias in Self-reported Turnout

- Measuring individual turnout:
  - voter file: registered voters only
  - survey: American National Election Survey (ANES)

- Social desirability bias: Did you vote? uhhh... yes.

- Data set: `turnout.csv`

Variables	Description
`year`	election year
`ANES`	ANES estimated turnout rate
`VEP`	Voting Eligible Population (in thousands)
`VAP`	Voting Age Population (in thousands)
`total`	total ballots cast for highest office (in thousands)
`felons`	total ineligible felons (in thousands)
`noncitizens`	total non-citizens (in thousands)
`overseas`	total eligible overseas voters (in thousands)
`osvoters`	total ballots counted by overseas voters (in thousands)

Causal Questions in Social Science Research

- Does the minimum wage increase the unemployment rate?
  - Unemployment rate went up after the minimum wage increased
  - Would the unemployment rate have gone up, had the minimum wage increase not occurred?

- Does race affect one's job prospect?
  - Jamal applied for a job but did not get it
  - Would Jamal have gotten a job if he were white?

- Comparison between factual and counterfactual
- Fundamental problem of causal inference:
  We must infer counterfactual outcomes

- No causation without manipulation: immutable characteristics

Assignments

1. Finish the in-class exercise and submit Q5 and Q6 to **Gradescope**

2. Sections 2.1–2.4 of QSS Chapter 2
   - Learn about causality and randomized experiments
   - Learn how to subset data
   - Don't just read. Try all the commands on your own

3. Pre-class Exercise: **qsslearnr** `Tutorial1: Causality I`
4. Also, try **qsslearnr-tidy**