Final Project Data Sources

By Julian W. Oolman ([jjpwade2@illinois.edu](mailto:jjpwade2@illinois.edu))

Start with your research question first. Then come to this document for data resources. The datasets below are merely suggestions. You are welcome to choose data from a source not listed. You must have at least 1,000 observations and 8 covariates.

You will need to create a free account at some websites to download their data.

Data Libraries

1. [IPUMS (Integrated Public Use Microdata)](https://ipums.org/)
   * Census & American Community Survey
   * Current Population Survey (CPS) - used to calculate unemployment statistics, similar to above
   * Time Use (respondents can be linked to the CPS)
   * International
   * Development
   * Environment
   * Health
   * STEM Higher Education
2. [Kaggle](https://www.kaggle.com/)
   * A data science website with data, projects, and competitions

Macro

The one country-level data set that I think is okay to use:

1. [Macro History](http://www.macrohistory.net/data/)
   * Yearly macro data from 1870 onwards for 17 countries
2. Bureau of Economic Research: [Regional Economic Accounts](https://apps.bea.gov/regional/downloadzip.cfm)
   * GDP – MSA, County, State
   * Personal consumption expenditures by categories

Macro/Industrial Organization – Firm Dynamics

1. [Census Business Dynamics Statistics](https://www.census.gov/programs-surveys/bds.html)
   * Firm/establishment exit and entry rates
     1. By industry, size, age, geography
   * Use legacy tables for more data
2. [County Business Patterns](https://www.census.gov/programs-surveys/cbp/data/datasets.html)
   * State, County, Industry (NAICS)
   * Employment, Payroll, Establishments
3. [Manufacturing Industry Database](https://www.nber.org/research/data/nber-ces-manufacturing-industry-database)
   * 1958 – 2011
   * Many industries
   * Value of products, capital, labor, wages, energy, inventories, etc.
4. [U.S. Patents](https://www.nber.org/research/data/us-patents-1975-1999)
   * 1963 – 1999
   * Patents, citations, classifications
   * Alternative data sources available

Finance

1. [Yahoo Finance](https://finance.yahoo.com/)

* Daily exchange rates for many countries to USD
* Daily stock market indexes for various countries
* Daily commodity prices

International Trade

* 1. [USA Trade](https://usatrade.census.gov/index.php)
     + Country and state origin/destination
     + Commodity, values
  2. [Commodity Flow Survey](https://www.census.gov/programs-surveys/cfs/data/tables.html)
     + Domestic trade of commodities
     + Origin/destination regions/state
     + Value, weight, distance, mode of transportation, temperature controlled

Micro

Labor

1. IPUMS - USA, CPS, ATUS, Health Surveys, Higher ED
2. [The National Longitudinal Survey of Youth](https://www.nlsinfo.org/investigator/pages/login) (NLSY)
   * ~10,000 individuals tracked from youth through adulthood
   * A **plethora** of variables
   * This one is really cool
3. [Opportunity Insights](https://opportunityinsights.org/data/?geographic_level=0&topic=0&paper_id=1652#resource-listing)

* Intergenerational mobility
* Tract level child neighborhood characteristics on their adult outcomes

1. [National Center for Education Insights](https://nces.ed.gov/programs/digest/current_tables.asp)
   * Elementary – university
   * Enrollment rates, graduation rates, demographics, outcomes
   * Faculty, staff, student-teacher ratios, funding

Crime

1. [Stanford Open Policing Project](https://openpolicing.stanford.edu/)
   * Detailed traffic stops data
   * Race, why stopped, whether arrested, etc.
2. Arrests data [NYPD](https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u/data) [LAPD](https://data.lacity.org/A-Safe-City/Arrest-Data-from-2010-to-2019/yru6-6re4/data) [SeattlePD](https://data.seattle.gov/Public-Safety/SPD-Crime-Data-2008-Present/tazs-3rd5/data)
   * Data on those arrested
   * Arrest categories
   * Demographic, geographic data

Natural Resources

1. [U.S. Energy Information Administration](https://www.eia.gov/)
   * Oil prices, coal prices, renewable prices, consumption, energy output, etc.

Mortgages/Housing

1. [Consumer Financial Protection Bureau](https://www.consumerfinance.gov/data-research/hmda/historic-data/)
   * Yearly macro data from 1870 onwards for 17 countries
   * Loan approved or denied
   * Demographics
2. [Fannie Mae](https://loanperformancedata.fanniemae.com/lppub/index.html#Portfolio)
   * BIG DATA – Caution!
   * Loan performance (delinquent, foreclosed)
   * Loan characteristics, credit score, etc.
3. [American Housing Survey](https://www.census.gov/programs-surveys/ahs/data.html)
   * House values, house characteristics
   * Owner economic and demographic features