Sources of Data

This is a list of the data sources Professor Duflo includes in her lecture slides. We are continuing to update the links and add other data sources so please let us know if you encounter any broken links!

Existing data libraries:

- A great resource for MIT students and others: https://libguides.mit.edu/socscidata

- Popular sources of data

  - Data.gov: Datasets generated by the executive branch of the US government http://www.data.gov/
  - IPUMS: censuses from the US and many more1 https://www.ipums.org/
  - International IPUMS https://international.ipums.org/international/
  - ICPSR http://www.icpsr.umich.edu/icpsrweb/ICPSR/: A data repository with many data sets on lots of subjects
  - Harvard-MIT Data center and Harvard Dataverse https://dataverse.harvard.edu/ where many researchers archive their data
  - Amazon dataverse http://aws.amazon.com/public-data-sets/

International household survey data

- Demographic and Health surveys http://www.dhsprogram.com/

- World Bank http://data.worldbank.org/

- LSMS http://surveys.worldbank.org/lsms

- RAND public-use databases https://www.rand.org/well-being/social-and-behavioral-policy/data.html

Replication data from researchers

- Randomized control trials

  - https://dataverse.harvard.edu/dataverse/jpal
  - https://dataverse.harvard.edu/dataverse/socialsciencercts

- The American Economic Association journals require posting of any data used for research: there is lots of data on the AEA website

The internet!

- Many websites that are data intensive are making that data directly available to people

  - 538
  - Yahoo data dump: "a sample of anonymized user interactions on the newsfeeds of several Yahoo properties" https://webscope.sandbox.yahoo.com/?guc
  - Uber movement https://movement.uber.com/cities?lang=en-US

- Some sites are specializing in aggregating data

  - Sports datasets http://nbasavant.com/
  - Web pages: Way back machine https://archive.org/web/
  - Google's dataset search tools: https://toolbox.google.com/datasetsearch https://www.blog.google/products/search/making-it-easier-discover-datasets/

- There is much more... search in library catalog, Google, etc.