Potential Data Sources for Econometrics Project

Prof. Tyler Ransom, Univ. of Oklahoma

September 6, 2018

This document lists publicly available data sets on a variety of topics. The purpose of the document is to help you see two things: *(i)* there is a data set on almost any topic imaginable; and *(ii)* you can freely access these. In some cases, there are existing packages in R that will further ease your process in downloading and using the data. To find the data set you're intersted in, simply enter the full name into your favorite internet search engine. The top result will likely take you to a page where you can download the data. For many of these data sets, you can visit http://asdfree.com and find the name of the data set in the lefthand panel. Once you click on it, it will pop up with additional information.

Dataset Search Engines

- Google dataset search: https://toolbox.google.com/datasetsearch

- US Government dataset search: https://www.data.gov

- Blog post describing R packages that will automatically scrape data from a variety of online sources: https://www.computerworld.com/article/3109890/data-analytics/these-r-packages-import-sports-weather-stock-data-and-more.html

Datasets by topic

Social

- Behavioral Risk Factor Surveillance System (BRFSS)

- General Social Survey (GSS)

- Pew Research Center (PEW)

Children & Family

- Early Childhood Longitudinal Study, Birth (ECLS-B)

- Early Childhood Longitudinal Study, Kindergarten (ECLS-K)

- National Survey of Children's Health (NSCH)

- National Survey of Family Growth (NSFG)

Adolescents

- National Longitudinal Study of Adolescent to Adult Health (Add Health)

- National Longitudinal Surveys (NLS)

- Youth Risk Behavior Surveillance System (YRBSS)

- National Educational Longitudinal Survey 1988 (NELS:88)

- Education Longitudinal Survey 2002 (ELS)

Consumption / Household Finance

- Consumer Expenditure Survey (CEX)

- Survey of Consumer Finances (SCF)

Business

- Survey of Business Owners (SBO)

- `forbesListR` (R package)

- `edgarWebR` (R package)

Finance / Macro

- FDIC Failed Bank List

- `Quandl` (R package)

- `Quantmod` (R package)

- `FredR` (R package)

- `eu.us.opendata` (R package)

Labor / Education

- American Community Survey (ACS); has R packages `ipumsr` and `censusapi` and `tidycensus`

- Current Population Survey - Annual Social and Economic Supplement (CPS-ASEC); has R package `ipumsr`

- Current Population Survey - Basic Monthly (CPS); has R package `ipumsr`

- Integrated Public Use Microdata System (IPUMS); has R package `ipumsr`

- Panel Study of Income Dynamics (PSID); has R package `psidR`

- Survey of Income and Program Participation (SIPP)

US Education System

- Integrated Postsecondary Education Data System (IPEDS)

- Common Core of Data (CCD)

- Schools and Staffing Survey (SASS)

- National Teacher and Principal Survey (NTPS)

- College Scorecard

- National Student Loan Data System (NSLDS)

Time use

- American Time Use Survey (ATUS)

Climate / Weather

- `rnoaa` (R package)

Housing

- American Housing Survey (AHS)

- Home Mortgage Disclosure Act (HMDA)

- Housing Affordability Data System (HADS)

- New York City Housing and Vacancy Survey (NYCHVS)

Health

- Basic Stand Alone Public Use File (BSAPUF)

- California Health Interview Survey (CHIS)

- National Beneficiary Survey (NBS)

- National Immunization Survey (NIS)

- National Health and Nutrition Examination Survey (NHANES)

- National Health Interview Survey (NHIS)

- National Study on Drug Use and Health (NSDUH)

- Surveillance Epidemiology and End Results (SEER)

Medical Claims / Expenditures

- FDA Adverse Event Reporting System (FAERS)

- Medical Expenditure Panel Survey (MEPS)

- Medical Large Claims Experience Study (MLCES)

- National Plan and Provider Enumeration System (NPPES)

Retirement / Ageing

- Health and Retirement Study (HRS)

- National Survey of OAA Participants (NPS)

- Social Security Administration Public Use Microdata (SSA)

Political Participation

- American National Election Study (ANES)

- `pollstR` (R package)

Foreign Countries

- Brazilian Censo Demografico (CENSO)

- Exame Nacional de Desempenho de Estudantes (ENADE)

- Exame Nacional do Ensino Medio (ENEM)

- Pesquisa Mensal de Emprego (PME)

- Pesquisa Nacional por Amostra de Domicilios (PNAD)

- Pesquisa Nacional por Amostra de Domicilios - Continua (PNADC)

- Pesquisa Nacional de Saude (PNS)

- Pesquisa de Orcamentos Familiares (POF)

Cross-National Comparative Surveys

- Demographic and Health Surveys (DHS)

- European Social Survey (ESS)

- Programme for the International Assessment of Adult Competencies (PIAAC)

- Progress in International Reading Literacy Study (PIRLS)

- Program for International Student Assessment (PISA)

- Survey of Health, Ageing and Retirement in Europe (SHARE)

- Trends in International Mathematics and Science Study (TIMSS)

- World Bank Living Standard Measurement Study (LSMS)

- World Values Survey (WVS)

Municipal / Regional Analysis

- Area Health Resource File (AHRF)

- Public Libraries Survey (PLS)

- `RSocrata` (R package)

- Bureau of Economic Analysis (BEA) — has R package `bea`

- Bureau of Labor Statistics Local Area Unemployment Statistics (LAUS) — has R packages `blscrapeR` and `blsAPI`