

Data Sources for Econ 123 Empirical Project

Your projects will be more exciting if you work on a topic that interests you. There are a lot of great datasets in the world, though they are not as easy to find as they should be. This document briefly lists a few places where you can find social science data. There are many others, and you are not restricted to social science topics, but this list might help you get started if you don't know where to look.

Consumer Expenditure Survey (CEX)

<https://www.bls.gov/ce/>

The Consumer Expenditure Survey (CEX) is collected every quarter by the US Bureau of Labor Statistics from a large nationally representative sample of American households. It includes numerous measures of income, purchases, household characteristics, and individual characteristics. This data is used to by the government to compute poverty thresholds for social programs and revise the Consumer Price Index (CPI) market basket of goods and services. Businesses and academic researchers use the data to learn about consumer spending habits and trends.

General Social Survey (GSS)

<http://gss.norc.oregon.edu/>

The General Social Survey (GSS) has been collecting data on American knowledge, opinions, and behavior every two years since 1972. It asks many of the same questions every round giving us our best source of data for examining trends in these factors. It also asks new questions on timely issues every round.

Integrated Public Use Microdata Series (IPUMS).

<https://www.ipums.org/>

Hosted by the University of Minnesota's Population Center, this site contains census and survey data on the United States and many other countries around the world. Very clean and well-documented data sets used intensively by researchers. Also includes data from: Demographic and Health Surveys, Time-Use Surveys, Health Surveys and the like.

Harvard Dataverse

https://dataverse.harvard.edu/?q=&fq0=subject_ss%3A%22Social+Sciences%22&types=dataverses%3Adatasets&sort=dateSort&order=desc

Replication data for about 6,000 research studies in the social sciences. Includes contributions from the International Food Policy Research Institute, the Gates Foundation and many other contributors.

AidData

<http://aiddata.org/datasets>

A group of datasets on flows of international aid expenditures, both at the country-level and within countries to particular sectors and projects.

World Bank

<http://data.worldbank.org/>

A large compendium of micro-data on developing countries. Demographic surveys, business surveys, exports and imports, etc. Also macro-level data on education, GDP, trade and the like.

World Bank – Living Standards Measurement Survey

<http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:21610833~pagePK:64168427~piPK:64168435~theSitePK:3358997,00.html>

A set of World Bank surveys dedicated to measuring poverty.

Abdul Latif Jameel Poverty Action Lab Dataverse.

<https://dataverse.harvard.edu/dataverse/jpal>

Hosted at Harvard's dataverse, this contains datasets from recent randomized experiments in development economics conducted by MIT's Poverty Action Lab and its researchers around the world.

Energy Information Administration

<https://www.eia.gov/tools/models/datatools.cfm>

Data on prices, quantities, imports and exports for many different fuels in the United States and abroad.

International Monetary Fund

<http://www.imf.org/en/Data#data>

Global country-level data on prices, financial access, debt and trade.

American Economic Review

<https://www.aeaweb.org/articles.php?doi=10.1257%2Faer.106.2>

Many papers at this premier journal include the data set on which the analysis is based and replication code. If you find an article of interest, and it was published recently enough, chances are you could find the data here by going to the right issue.

Journal of Applied Econometrics

<https://onlinelibrary.wiley.com/journal/10991255>

[The JAE Data Archive](#) contains data for all papers accepted after 1994 unless the data are confidential. For most recent papers, the Data Archive also contains programs and supplementary material, such as technical appendices and additional graphs. JAE also has a regular section focusing on replication of published studies, where you can find numerous [examples](#) on how to write a replication study.

Demographic and Health Surveys

<http://www.dhsprogram.com/Data/>

The Demographic and Health Surveys (DHS) Program has collected, analyzed, and disseminated accurate and representative data on population, health, HIV, and nutrition through more than 300 surveys in over 90 countries.

FED NY Center for Microeconomic Data

<https://www.newyorkfed.org/microeconomics/databank.html>

The Center for Microeconomic Data serves to centralize the collection, acquisition, and analysis of microeconomic data at the New York Fed and act as a catalyst for microeconomic research by promoting engagement with the larger academic community. The wide-ranging data, research, and analysis produced by CMD researchers provide insight into individual-level financial and nonfinancial economic conditions, expectations, and behavior in the United States.

Pew Research Center Data bank

<http://www.pewresearch.org/download-datasets/>

Pew Research Center is a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping the world. They conduct public opinion polling, demographic research, content analysis and other data-driven social science research and make their data available to the public for secondary analysis after a period of time.