

Where to find data? A Primer

As a beginner (and even for more advanced users), finding data may not be an easy task. The degree of difficulty varies greatly based on the topic, geography, and other dimensions. The following is a document that can help ease this process.

Data Search Engines

The following websites can be thought of as "Data Search Engines"; that is, you enter a query and the website will show you a set of options of possible data available (either publicly or privately) related to that query.

In some cases, the data is not readily available for the public. However, do not get discouraged! Sometimes the barriers to entry are as small as sending an e-mail, while in other situations, you need only to fill out a form or a data request. Always ask an organization or institution how you can gain access to their data.

- [ICPSR](#)
A service of the Institute for Social Research at the University of Michigan. This data-search engine is a great resource and will provide you with many good options. While you can find information on other countries, this source is skewed heavily toward domestic information.
- [Dataverse](#)
This page is a data repository for papers that already exist. In many cases, if you are interested in the data sources used by a specific paper, Dataverse can provide you with this information, as well do-files for that data. Like ICPSR, you can just type a search word and try your luck! (You should be aware that the journal itself may have asked the authors to provide a replication kit, which would be available in the online page of the paper.)
- [Quand](#)
This website is most likely provided by a private entity. If you are looking for financial variables, this can provide a good start.
- [FRED](#)
A service of the St. Louis Fed. This is a good website for findings data related to macroeconomic variables (unemployment, interest rates, inequality indexes, etc.). While it has information on other countries for certain variables, it is mostly U.S.-focused.
- [Google Public Data](#)
Google's take on data. Most of the information here is aggregated at the country level, but it is a great way to see what's available. Most of the data comes from organizations like UN, World Bank, etc. This is also another great resource to explore trends and create interesting graphs. By nature, it contains a lot of aggregate information from other countries.

Data lists

The following are websites that have *a list* of datasets. The benefit from most of the data found on these sites is that the data is relatively "cleaner" than what you would find in the search engines.

- [NBER](#) - This is data provided by the National Bureau of Economic Research. It covers a wide range of topics: Macro, Industry and Productivity, International Trade, Household Surveys, Health Care Data, Demographics, Patents, and Others.
- [IPUMS](#) - This fantastic institution provides around 10 surveys (mostly U.S.) that have been cleaned and harmonized. The data might be challenging to use in the beginning, but once you get

the hang of it, it is very useful (especially when trying to get a quick estimate on a variable). I tend to frequently use the IPUMS-CPS (which is the cleaned version of the Current Population Survey) and IHIS (which is the cleaned version of the National Health Interview Survey). Note that not *all* variables found in the original dataset are in the IPUMS data. Hence, if you don't find a variable you are looking through IPUMS for a specific year, it doesn't mean that it doesn't exist in the original data.

- [CDC Wonder](#) - This site from the CDC has a list of dataset that they sponsor. It is mostly on health outcomes.
- [CDC Health Data \(Others\)](#) - This is another list on other health-related datasets.
- [BEA](#) - This is a data list from the datasets offered by the Bureau of Economic Analysis.
- [AEA Data List](#) - This is a data list from the American Economic Association.
- [Non-filter extensive list](#) - This is a data set that someone else has composed.
- [Historical Statistics](#) - Historical data may be hard to find, but this website provides is a great first step.
- [Data is plural](#) is a mailing list that will send you a newsletter with some odd datasets. Subscribe! But also find “weird” data on this [repository](#)

Others

- [Health Policy Tracking](#) - This is not a dataset list per se. This is a database in itself of changes in policies by state. It not comprehensive, so there might be a policy change that is not recorded in this database, but it is a good place to start.
- [Health Policy Research Dataset](#) - This is a very handy dataset that has a lot of information on states' policy changes and the year in which they occurred. It provides lots of quick “controls” at the state-year level.

Tips:

- If you find reports or papers where you see a statistic or a “number” that you need, look in the footnote or reference of the document to see where the authors obtained the data. At some point you may need to contact or email someone. Be sure to be polite and use personal judgement (i.e. do not constantly bombard an individual with lots of emails!) when composing e-mails. Some of those papers or reports may have a replication-file that is publicly accessible.
- Ask around for data sources. If you have people in your network who are experts on a topic, ask them for their advice.
- Get creative. If you cannot find the variable that you are looking for, what is another variable that would be a good proxy? Can you find that?
- Validate your data source. If you find data that you liked, see if anyone else has used it, and for what purpose?