DAV 5400

Where to Find Data?

The web is literally awash with freely available data. A simple web search can yield a huge amount of prospective data sets that can be used for different types of data analysis. There are also many, many sites that either maintain their own collections of data sets or try to act as an aggregator of information on freely available data sets.

If you are struggling to find data that you are genuinely interested in working with, here's a list of some sites that might be of help. Please note that this list is by no means exhaustive and the inclusion of any particular web page within the list is not meant to serve as an endorsement of its content. You can certainly find many other sources of data with a little effort on your own part, but feel free to let this list serve as a starting point if needed.

** Please note that some of these sites might require you to register with them before you are allowed to download their otherwise freely available data.

OpenDataSoft: Open Data Portals (Worldwide): https://www.opendatasoft.com/a-comprehensive-list-of-all-open-dataportals-around-the-world/#/united-states

"Awesome Public Data Sets"(maintained on GitHub): https://github.com/awesomedata/awesome-public-datasets

Earth Science data sets: http://www.ciesin.org/datasets/dataset-home.html

Vanderbilt University Biostatistics: http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets

University of California – Irvine: https://archive.ics.uci.edu/ml/datasets.html

University of Pittsburgh "World Historical Dataverse": http://dataverse.pitt.edu/external/datasets.php

R Data Sets: https://vincentarelbundock.github.io/Rdatasets/datasets.html

US Bureau of Labor Statistics: https://www.bls.gov/data/

US Dept. of Agriculture – County Level Data Sets: https://www.ers.usda.gov/data-products/county-level-datasets/download-data/

Fed. Reserve Bank of St Louis – Economic Time Series Data Sets: https://research.stlouisfed.org/pdl/

World Bank Economic Indicators: https://data.worldbank.org/indicator

UN Human Development Reports: http://hdr.undp.org/en/2018-update

World Health Organization: http://apps.who.int/gho/data/node.home

Data.gov: https://catalog.data.gov/dataset

Microsoft Anonymous Web Data: https://kdd.ics.uci.edu/databases/msweb/msweb.html

Kaggle Data Sets: https://www.kaggle.com/datasets

DrivenData.org (website hosts data analysis competitions): https://www.drivendata.org/competitions/

538.com: https://github.com/fivethirtyeight/data

New York City Open Data: https://opendata.cityofnewyork.us/