Dataset Sources List

Seungjun (Josh) Kim
Last Updated 2020/11/28

• Kaggle
www.kaggle.com
➔ Has all kinds of datasets. Pros – you can look at what kind of analyses (python notebooks) people did on specific datasets / Cons – Anyone can upload datasets, so the accuracy and credibility of datasets may not be totally reliable for econ research. Recommend using it as a sub-dataset to link with a more reliable and official dataset you use as the main dataset for research (or use it just for practicing python and Machine Learning)

• Google Dataset Search
https://datasetsearch.research.google.com/
➔ Similar to Google Scholars (for papers) for dataset

• US Government statistics
https://catalog.data.gov/dataset

• Integrated Public Use Microdata Series (IPUMS)
https://usa.ipums.org/usa/

• ICPSR
https://www.icpsr.umich.edu/web/pages/ICPSR/index.html

• National Health and Nutrition Examination Survey
https://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm

• National Oceanic and Atmospheric Administration
https://www.ncdc.noaa.gov/

• World Bank Dataset
https://databank.worldbank.org/home.aspx

• Food and Agriculture Organization (FAO) of the UN
http://www.fao.org/statistics/databases/en/

• Gapminder (global data)
https://www.gapminder.org/data/

• UK Government Data
https://data.gov.uk/search

• European Statistics
https://ec.europa.eu/eurostat/data/database

• Environmental Protection Agency (EPA)
https://www.epa.gov/fueleconomy

• KDNuggets Dataset Curated List
https://www.kdnuggets.com/datasets/index.html
➔ One of the famous Data Science Blog websites; on top of this dataset list, tutorials and blog posts themselves are pretty good to check out

• Github Awesome Public Dataset Compilation
https://github.com/awesomedata/awesome-public-datasets
➔ It's always a good idea to make a github account and showcase your analytics / data projects on it if you are serious about econometrics / data science / CS etc.

• UCI Machine Learning Repository

https://archive.ics.uci.edu/ml/index.php
➔ Good classic datasets to use for practicing ML.

• Knoema
https://knoema.com/atlas/sources

• Data World
https://data.world/datasets/open-data

• Reddit Dataset Thread
https://www.reddit.com/r/datasets/
➔ People can't directly upload datasets here but users usually comment links or urls of the datasets of interest

• FiveThirtyEight Data
https://github.com/fivethirtyeight/data
➔ FiveThirtyEight – famous for its data visualization and it uploads datasets it used to create those visualizations

• BuzzfeedNews Data
https://github.com/BuzzFeedNews

• NASA data
https://earthdata.nasa.gov/
➔ Earth and Space related datasets

• Amazon AWS Open Data
https://registry.opendata.aws/

• Google cloud Open Data
https://cloud.google.com/bigquery/public-data/
➔ You need to make a Google Cloud account though

• Quandl
https://www.quandl.com/search

• Academic Torrents
https://academictorrents.com/browse.php?cat=6

• Kevin Chai's Dataset (Blog)
http://kevinchai.net/datasets
➔ 이렇게 open source로 tutorial, information 등 공유해주는 사람 = 참 고마운 사람