

Dataset Sources

- Awesomedata: Awesome Public Datasets: <https://github.com/awesomedata/awesome-public-datasets>
- Datasetlist: Machine learning datasets: <https://www.datasetlist.com>
- UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/index.php>
- Data is Plural Newsletter: <https://tinyletter.com/data-is-plural>
- Harvard Dataverse: <https://dataverse.harvard.edu/>
- Google Dataset Search: <https://datasetsearch.research.google.com/>
- Centers for Disease Control and Prevention: <https://data.cdc.gov/>
- United States Census: <https://data.census.gov/cedsci/> or <https://censusreporter.org/> or <https://github.com/hrecht/censusapi> or <https://walker-data.com/tidycensus/index.html>
- Zillow Data: <https://www.zillow.com/research/data/>
- Austin R. Benson's Datasets (including several graph datasets): <https://www.cs.cornell.edu/~arb/data/index.html>
- Registry of Open Data on AWS: <https://registry.opendata.aws/>
- Google BigQuery Public Datasets: <https://cloud.google.com/bigquery/public-data/>
- Grouplens: <https://grouplens.org/datasets/movielens/>
- IMDb: <https://www.imdb.com/interfaces/>
- Amazon Reviews: <http://jmcauley.ucsd.edu/data/amazon/>
- AidData: <https://www.aiddata.org/datasets>
- UNCTADStat: <https://unctadstat.unctad.org/>
- Interesting Datasets: <https://github.com/curran/data>
- The World Bank: <https://data.worldbank.org/> or <https://databank.worldbank.org/databases>
- Datahub: <https://datahub.io/collections>
- 18 Places to Find Free Datasets: <https://www.dataquest.io/blog/free-datasets-for-projects/>
- FDA: <https://open.fda.gov/>
- USDA: <https://www.ers.usda.gov/data-products/>
- World Health Organization: <https://www.who.int/data/collections>
- Global Health Data Exchange: <http://ghdx.healthdata.org/>
- Bureau of Economic Analysis: <https://www.bea.gov/data>
- Centers of Medicare and Medicaid Services: <https://data.cms.gov/> or <https://developer.cms.gov/data-medicare>
- Environmental Protection Agency: <https://edg.epa.gov/metadata/catalog/main/home.page>
- National Center for Education Statistics: <https://nces.ed.gov/>
- World Inequality Database: <https://wid.world/data/>
- reddit/r/datasets: <https://www.reddit.com/r/datasets/>
- Spotify: <https://github.com/charlie86/spotifyr>
- Genius: <https://github.com/ewenme/geniusr>
- BallR and BaseballR: <https://github.com/toddwschneider/ballr> and <http://billpetti.github.io/baseballr/>

- Twitter: <https://developer.twitter.com/en/docs>
- FiveThirtyEight: <https://data.fivethirtyeight.com> and <https://github.com/fivethirtyeight/data>
- BuzzFeedNews: <https://github.com/BuzzFeedNews/everything>
- ProPublica: <https://www.propublica.org/datastore/datasets>
- data.world: <https://data.world/>
- U.S. Open Data: <https://www.data.gov/>
- Nashville Open Data: <https://data.nashville.gov/>
- Academic Torrents: <https://academictorrents.com/> (requires a bittorrent client like deluge)
- HealthData: <https://healthdata.gov>