DATA RESOURCES

Dataset	Details
Healthcare Cost and Utilization Project (HCUP) http://www.hcup-us.ahrq.gov	Includes US longitudinal hospital care data with databases, software and online tools
Health Data.Gov http://www.healthdataov	Users can search by data category (8) and format(.csv,.xls, zip, PDF, rdf, JSON, html, txt and API
Centers for Disease Control and Prevention http://www.cdc.gov/nchs/data_access/ftp_data.hm	Includes public use files (PUFs) from surveys from multiple government agencies
Centers for Disease Control and Prevention https://www.cdc.gov/DataStatistics/	Various data reports, charts and downloads related to US Health
Centers for Disease Control and Prevention https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html	Tracks the prevalence of diabetes by state. Provides a variety of filters, geo-codes, charts, and tables
Centers for Disease Control and Prevention (CDC) Wonder https://wonder.cdc.gov/Welcome.html	Public health-oriented databases
Data.gov https://www.data.gov/	Government repository of open data. Has almost 200,000 datasets that include medical data available to download in various formats
World Health Organization https://www.who.int/data/gho/data/indiors	A myriad of health indicators with visualizations and downloadable datasets
Expert Health Data Programming http://www.ehdp.com/vitalnet/datasets.htm	Host links to about 45 large datasets
Health Services Research Information Central https://hsric.nlm.nih.gov/hsric_public/display_links/722	Has extensive health datasets, statistics, international data and data tools
Vanderbilt Biostatistics Database http://biostat.mc.vanderbilt.edu/wiki/Main/datasets	Multiple health-related datasets are available to download as Excel, ASCII, R, and S- Plus files. Also includes links to international datasets
MIMIC III Critical Care database https://mimic.physionet.org	Repository of more than 40,000 de- identified critical care patient-level inpatient data

CMS Data Navigator https://dnav.cms.gov/Default.aspx	Expedites the search for Medicare and Medicaid data
Gapminder https://www.gapminder.org	Hosts lots of international datasets in .csv format
County Health Rankings and Roadmap http://www.countyhealthrankings.org/	Ranks multiple health measures for US counties. Able to compare counties and download raw data. Data is from 2015-2017 time period
Medicare Data https://data.medicare.gov	Provides downloadable (csv) data to compare hospitals, physicians, nursing homes, hospice, dialysis centers, home health, inpatient rehab, and long term inpatient care. Each category has multiple sub-categories. For example, Hospital Compare has 77 files
National Health and Nutrition Examination Survey (NHANES) https://www.cdc.gov/nchs/nhanes/index.htm	The long-standing project that conducts interviews, physical and lab exams on US citizens. Data is available for download with complete data dictionaries
Health Information National Trends Survey (HINTS) https://hints.cancer.gov/about-hints/default.aspx .	Part of the National Cancer Institute. Surveys have been conducted since 2003 on cancer and multiple other topics. The data format for SPSS, SAS, and Stata

National Health Interview Survey (NHIS) https://www.cdc.gov/nchs/nhis/index.htm	Another US national health trend survey tool. Includes an interview but no physical exam or lab data.
University of California, Irvine Repository: https://archive.ics.uci.edu/ml/datasets.html	The site includes 325 validated datasets covering many domains, different sizes, and data types, and different analytical methods, such as classification and regression. These datasets are commonly used for machine learning exercises
KDNuggets: www.kdnuggets.com	The site includes 71 datasets available for free download, from various industries.
The Datahub: https://datahub.io/dataset	Managed by the Open Knowledge Foundation, this site hosts more than 10,000 datasets from most industries.

Kaggle: www.kaggle.com	Provides free, interesting datasets for various user interests and analysis.
Data.World https://data.world	New site for creating collaborative data projects with the ability to host data and analyze with embedded SQL. Tools available to link to Tableau, R, and Python languages, and machine learning. Repository for many medical and non-medical datasets
OpenML https://openml.og	Hosts a variety of datasets and ML workflows. Includes the results of multiple ML algorithms run
Sklearn datasets https://sciit-learn.org/stable/datasets/index.html	Python package includes six datasets for analyses
Keras datasets https://keras.io/api/datasets/	Includes seven datasets appropriate for deep learning
Google Dataset Search https://toolbox.google.com/datasetsearch	Google search feature for a variety of datasets, to include health-related. You can search by last updated, format, usage rights and topic. You can also search by task e.g., regression
Google datasets https://ai.google/tools/datasets/	Wide offering generally used to train machine learning and AI
Microsoft Open Datasets https://msropendata.com	Covers categories of computer science, information science, physics, and social science

Open Datasets on GitHub https://github.com/awesomedata/awesome-public-datasets	Extensive list in multiple categories
PyCaret https://pycaret.org	This unique Python package comes with 50 datasets for supervised and unsupervised learning
Synthea https://synthea.mitre.org/downloads	Data on 1,000 synthetic patients can be downloaded that includes conditions, vital signs, medications, encounters, etc. Also, has a 10K COVID-19 dataset
Dataset list https://www.datasetlist.com/	Extensive machine learning datasets that are non- medical
Fast.ai. Datasets https://course.fast.ai/datasets.html	Image, NLP and image localization libraries

NLP datasets https://github.com/niderhoff/nlp-datasets	Primarily raw data for NLP

Kaggle healthcare datasets https://www.kaggle.com/tags/healthcare	Variety of datasets used for Kaggle competitions
SEER datasets https://seer.cancer.gov/explorer/	National Cancer datasets
Oasis imaging datasets http://www.oasis-brains.org/	Open-source neuroimaging data
BigML datasets https://bigml.com/gallery/datasets/healthcare	Healthcare data with spreadsheet-like functionality
Aylward Image Repository http://www.aylward.org/home	Image repository created by Stephen Aylward PhD