

Data Set Sites for Statistics Courses

Thanks to the AMATYC Statistics FLC (Faculty Learning Community) Group, which contributed many of these data sources. Revised: 5/15/2021

- **Data.gov** - <https://www.data.gov>

Site Description: The purpose of Data.gov is to increase public access to high value, machine readable data sets generated by the Executive Branch of the Federal Government. It offers numerous free data sets in a searchable database.

- **American National Election Studies (ANES)** - <http://www.electionstudies.org>

Site Description: To serve the research needs of social scientists, teachers, students, policy makers and journalists, the ANES produces high quality data from its own surveys on voting, public opinion, and political participation. Central to this mission is the active involvement of the ANES research community in all phases of the project.

- **DataWeb (USITC)** - <https://dataweb.usitc.gov>

Site Description: The United States International Trade Commission (USITC) Interactive Tariff and Trade DataWeb provides U.S. international trade statistics and U.S. tariff data to the public full-time and free of charge. All trade data are compiled from official data retrieved from the U.S. Bureau of the Census (an agency within the U.S. Department of Commerce).

- **CAUSEweb** - https://causeweb.org/cause/resources/library?field_material_type_tid=104

Site Description: The Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) is a national organization whose mission is to support and advance undergraduate statistics education in the four target areas of resources, professional development, outreach, and research. Data are provided by members and come from a variety of sources.

- **National Climatic Data Center (NOAA)** - <https://www.ncdc.noaa.gov>

Site Description: The National Oceanic and Atmospheric Administration (NOAA)'s National Centers for Environmental Information (NCEI) is responsible for preserving, monitoring, assessing, and providing public access to the Nation's treasure of climate and historical weather data and information.

- **Data is Plural** - <https://www.data-is-plural.com/archive/>

Site Description: This site is more of a weekly newsletter of curious/useful data sets. In each weekly article one can find a few short descriptions about interesting data. Each data set description contains links one can go to in order to peruse/download data.

- **Awesome Data** - <https://github.com/awesomedata/awesome-public-datasets>

Site Description: Awesome Data provides links to real world data sets which are organized by topic. The topic-centric list includes Agriculture, Biology, Education, Government, Economics, and Cybersecurity just to name a few. The creator of this list states that most of the data sets are free, but there are a few that are not.

- **Worldometer** - <https://www.worldometers.info/>

Site Description: The site Worldometer is essentially an ever-evolving data hub that showcases population, government, economic, social, media, environmental, food, water, energy, and health data. The Worldometer homepage does provide links to other sites, but one may have to look for the data sets provided at these links as they are not direct links to the data sets.

- **The City of Chicago Portal** - <https://data.cityofchicago.org/>

Site Description: The City of Chicago Portal contains data on salaries, taxi rides, red light cameras for traffic violations, graffiti reports, and much more. Some of the data may require some cleaning for missed observations, typos and strange results.

It should be noted that most major cities have a data portal that is open to the public. The City of Los Angeles has such a portal, <https://data.lacity.org/>. The City of Baltimore also has a portal through the **BNIA** (Baltimore Neighborhood Indicator Alliance). The link to this site is <https://vital-signs-bniajfi.hub.arcgis.com/>.

- **Teaching of Statistics in the Health Sciences Resources Portal -**

<https://www.causeweb.org/tshs/category/dataset/>

Site Description: The TSHS portal is a great resource for finding data sets specific to health-related studies. The data sets for each of the studies may be downloaded for your preferred statistical software package. The site also does a great job of explaining the background for each study as well as being informative as to which statistical methods were used.

- **Inter-university Consortium for Political and Social Research (ICPSR) -**

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

Site Description: The ICPSR advances and expands social and behavioral research, acting as a global leader in data stewardship and providing rich data resources and responsive educational opportunities for present and future generations. This site has downloadable data sets in various formats. Users will need to create an account in order to download data sets from this site.

- **Georgia Department of Education SAT and ACT Results -**

<https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Curriculum-and-Instruction/Pages/SAT-and-ACT-Results.aspx>

Site Description: The Georgia Department of Education site has SAT and ACT results broken down by School or District for the 2019 and 2020 school years. These files are downloadable excel files.

- **US News and World Report - 100 Best Jobs in America -**

<https://money.usnews.com/careers/best-jobs/rankings/the-100-best-jobs>

Site Description: This site does not contain a data set to download and work with. The site simply shows the projected number of jobs, median salaries, and education level for the top 100 jobs for 2021. The site does provide a dropdown menu allowing the viewer to change the rankings to view other areas of interest such as Best STEM Jobs, Best Paying Jobs, Highest paying jobs without a degree, and many more.

- **Gap Minder** - <https://www.gapminder.org/>

Site Description: [Gapminder.org](https://www.gapminder.org/) is a popular site which many individuals may be familiar with. They have animated scatter plots which can compare a wide range of socio-economic issues. The animated features might be good for a survey statistics course. The site also has several hundred data sets covering all countries over various time periods. The data is downloadable. One might consider downloading data and creating custom data sets that could be used in an intro course. One could compare, for example, employment rates, foreign aid given or alcohol consumption in the US with those of the UK & Germany for the last 30 years.

- **Tuva Labs** - <https://tuvalabs.com/>

Site Description: You need to sign up to get access to over 15 free data sets. The data sets cover a wide range of topics – science, government, environment, social, etc. You can use the site’s tools to explore the data, or you can download the data as .csv. For each set, the site suggests an appropriate grade level and provides activities that others have created for that topic. It also lists what looks like common core objectives associated with the topic the data covers. As an example, one can download a file comparing Dreamworks vs. Pixar animated movies.

- **Census at School (ASA)** - <http://ww2.amstat.org/censusatschool/>

Site Description: The American Statistical Association’s “Census at School” has readily downloadable data sets that you can create from a pool of K-12 surveys, both US and foreign. Go to the site and click on “Random Sampler”. You select the sample size, state, grade level, etc. and download as a .csv. There are 40 variables covering demographics, preferences, etc. The survey questions are in the “Student” section. As an example, one can create a file of New York 12th graders. College freshmen might be interested in comparing similar data collected at their school with data from the census.

- **UN Data** - <http://data.un.org/>

Site Description: The United Nations database has a lot of good information and many data sets to peruse through. It also has filters that will allow you to make the data sets smaller.

- **Organization for Economic Co-Operation and Development -**
<http://stats.oecd.org/Index.aspx>

Site Description: This site has tons of data sets broken down topic. It appears to be fairly current in the data it has included. Some of these may be usable in an introductory course. Some data sets may have to be fine-tuned in order to make them more manageable.

- **Journal of Statistics Education (JSE) -**
http://ww2.amstat.org/publications/jse/jse_data_archive.htm

Site Description: This link from the American Statistical Association (ASA) has many data sets, most in text format. Many of these sets have articles describing how the data was used in activities for students. Some of the data is old, but interesting nonetheless. These are relatively small data sets, but have several variables.

- **Stats 101: Resources for Intro Stats Instructors -**
<http://community.amstat.org/stats101/home>

Site Description: This American Statistical Association (ASA) link was put together by Dick DeVaux and others. It currently has 5 investigations involving large data sets. There is a detailed description of how each investigation is conducted.