



Statistics and Datasets

Melissa Mallon

To cite this article: Melissa Mallon (2019) Statistics and Datasets, Public Services Quarterly, 15:1, 24-33, DOI: [10.1080/15228959.2018.1556147](https://doi.org/10.1080/15228959.2018.1556147)

To link to this article: <https://doi.org/10.1080/15228959.2018.1556147>



Published online: 28 May 2019.



Submit your article to this journal [↗](#)



Article views: 14



View Crossmark data [↗](#)

INTERNET RESOURCES

Melissa Mallon, Column Editor



Column description. *The Public Services Quarterly* Internet Resources Column is designed to be a clearinghouse for free, online websites; each column will focus on themes relevant to current issues in academic libraries and feature resources selected to make the lives of public services librarians easier. Any comments about the column, including suggestions for themes or recommendations of web resources, can be directed to Melissa Mallon at mallon.melissa@gmail.com.

Statistics and Datasets

MELISSA MALLON

Vanderbilt University, Jean & Alexander Heard Library, Nashville, TN, USA

There are few research projects that do not require some sort of reliance on data, statistics, or other types of numerical support. Academic librarians are often on front lines when it comes to identifying, accessing, and interpreting this data. From a general reference question related to census data, to an undergraduate student that needs help locating a dataset on gender and pay, to a PhD student uncovering statistical trends for their dissertation, quick and easy access to data sources is a staple of academic librarianship.

The resources in this column include datasets and statistics covering both North America (United States and Canada) as well as international countries; some are more specific, such as KIDS Count data (which tracks the well-being of U.S. children), while others are broader in nature.

© Melissa Mallon

Comments and suggestions should be sent to the Column Editor: Melissa Mallon, Vanderbilt University, Vanderbilt Libraries, 419 21st Avenue South, Nashville, TN, 37203, USA. E-mail: mallon.melissa@gmail.com

Regardless of their geographic and topical scope, the resources included in this column provide researchers and librarians with access to a copious amount of freely available data that can be downloaded, analyzed, and used in any number of research projects.

Data.gov

<https://www.data.gov/>

Visited: Winter 2018

Reviewer: Laureen P. Cantwell, Reference & Distance Services Librarian, John U. Tomlinson Library, Colorado Mesa University

With over 300,000 searchable datasets on a wide range of topics, an *Internet Resources* column about sources for statistics and data would be incomplete without discussing *Data.gov*. The Technology Transformation Service of the U.S. General Services Administration manages and hosts this website and follows the Project Open Data schema with standard fields for each dataset. Given the connection to the Federal Open Data Policy (2013), the intention of *Data.gov* is clearly to provide data information to the general public. That said, librarians will find data relevant to the needs of students, faculty, and broader community members alike—particularly if local government employees or nonprofit organizations are among their patrons.

Datasets are searchable by keyword, tags, types, formats, groups, organization types, organizations, and categories, and in combinations of the aforementioned limits. For example, at the time of this review, a keyword search for [wildfire AND California] results in 34 datasets from state, federal, and university organization types. Datasets from this search include:

- Combined wildfire dataset for the United States and certain territories, 1870-2015 (Department of the Interior, data pub. Jan. 2017)
- Fuel treatment and fire history within the Rim Fire in California (Department of Agriculture, Forest Service, data pub. Apr. 2018)
- Historical Fire Perimeters - Southern California (State of California, data pub. Jan. 2005)

Alternatively, a keyword search for income with a limit to the topic of Education produces a 46-item result set including:

- Annual Survey of School System Finances (Department of Commerce, data last updated June 2016)
- Free Application for Federal Student Aid, 2015-2016 (Department of Education, Office of Federal Student Aid, data last updated June 2016)
- National Longitudinal Transition Study-2 (National Center for Special Education Research, data last updated June 2016)

Datasets might also include additional metadata such as the last time the data and/or the metadata was updated, who is responsible for the data, and contact information for the responsible organization—and this metadata is also searchable. Phrase searching with quotes functions well in the keyword search box, including for contact email addresses. Users may find datasets in .zip files, Excel files, rasters, shapefiles, and PDFs, and resource links to websites may accompany the data.

Additionally useful to know is the nifty, map-based geospatial search option. Drawing a box on the map will result in the identification of datasets tagged for that region. For example, the application of a rectangle over western Colorado leads to over 8,000 datasets addressing topics like climate, ecosystems, and local government; geospatial and non-geospatial data; tags of “earth sciences”, “u.s. department of commerce”, “colorado plateau”, and altitude; organizations such as the Department of the Interior, the Census Bureau, the USGS, FEMA, and others; and a variety of publishers.

Users can also navigate the *Data.gov* site by exploring topics from the home page. Clicking “Public Safety” leads to datasets regarding product recalls, an article on launching the police data initiative, and more. Topic pages include tabs for data, apps, updates, challenges, resources, and contact. Lastly, librarians working with faculty may be interested to know that users can submit a “data story” that “describes an application or solution using a specific dataset on *Data.gov*” and that topic-based challenges might include app creation, data visualizations, food safety detection methods, and more. This could provide exciting fodder for campus contests, course projects, and more.

In Short: *Data.gov* is relatively easy to use, usefully integrates searchable dataset content submitted by a wide variety of university, state, and federal organizations, and allows users to filter search results with useful limiting facets, including a geospatial option. Users in most need of this source may include post-secondary students and faculty in a variety of disciplines, but also members of the broader community who may have personal- or business-related uses for the data.

Highly Recommended.

World Bank Open Data

<https://data.worldbank.org>

Visited: Winter 2018

Reviewer: Margot Hanson, Instruction & Outreach Librarian, CSU Maritime Academy

The *World Bank Open Data* site is a data-lover’s dream. It affords easy access to massive amounts of specific international development data, such

as the literacy rates of young men or the percentage of pregnant women receiving prenatal care. Before “open data” started becoming common practice, the World Bank was at the vanguard of making their datasets open and accessible to the public. When the *Open Data* platform launched in 2010, it not only provided downloadable access to datasets, it also provided a robust and enjoyable user interface complete with interactive visualization tools. The resulting heatmaps or longitudinal line graphs could even be embedded via dynamic widgets on a user’s personal website. Other international agencies such as International Monetary Fund (IMF) and Organisation for Economic Co-operation and Development (OECD) have followed suit, now providing access to their data in interactive ways.

This reviewer has been using the *World Bank Open Data* site since 2012, when it earned an Innovator Award from the Scholarly Publishing and Academic Resources Coalition (SPARC). The platform was updated in 2017 with some adjustments to layout, hierarchy and colors; however, users can view the original version on the Internet Archive Wayback Machine if curious about the changes. *World Bank Open Data* has been invaluable as a resource for reference questions, classroom instruction, information literacy assignments, and recommended data sources included on this reviewer’s library’s subject guides. For example the site is useful for business students comparing country-level economic development over the past several decades. It is a convenient tool to introduce the concepts of data sets and data collection, and provides a low barrier to entry for students learning to create their own Excel charts and graphs for beginning analysis. This resource is also invaluable to the advanced researcher, with advanced data queries and various capabilities for slicing and dicing the data directly within the interface.

The World Bank Group comprises a global partnership of five institutions with offices in over 130 locations, and the data they collect and disseminate is focused on development-related indicators. Datasets are updated on an ongoing basis, with some time series going back as far as 1960 (subject to availability). As with any international data collection, some country data is spottier than others, and obviously some datasets, such as mobile cellular subscriptions, only cover more recent years. Data is available to use and download for free through a Creative Commons Attribution 4.0 International License. A Data API is also available for developers to access data directly for custom visualizations in combination with other data sources. For many of the data charts or maps, pre-built widget code snippets are available that include automatic updating when the time series are updated.

World Bank Open Data is accessible via three main interfaces: The World Development Indicators (the main scope of this review), the DataBank (extensive time series data with custom query and visualization

functions), and the Data Catalog (a search interface for all World Bank datasets, plus additional sources). The World Development Indicators is the primary interface for users to browse by country or indicator, allowing researchers to view the overall scope of data available. Country browsing is also broken down into pre-defined subsets such as regions, income levels, lending groups, “small states,” OECD members, and many more. Broad topics for the Indicators section include Aid Effectiveness, Education, Gender, Health, and Poverty, among many others. Scope notes, metadata, and methodology are available for each indicator.

In Short: The *World Bank Open Data* interface is useful for novice and expert researchers looking for statistics and visualizations of international development data.

Highly Recommended.

KIDS COUNT Data Center

<https://datacenter.kidscount.org/>

Visited: Winter 2018

Reviewer: Annis Lee Adams, E-Resources Librarian, California State University, East Bay

KIDS COUNT is a project of the Annie E. Casey Foundation that tracks the well-being of children in the United States. Its *Data Center* provides access to a variety of child-related data as well as trend analyses in order to highlight children’s issues. The KIDS COUNT *Data Center* receives data from national sources as well as state organizations and thus offers a wide breadth of statistics at the local, state, and national levels.

The *Data Center*’s home page provides different pathways to dig into the data based upon the specific indicators of interest. Users can either start by selecting a geographical area or by selecting a topic. Either path will lead to the same datasets. When starting by geography, a user can select a U.S. state on the map presented or select national data. When choosing a state, the selection can be further refined by city, county, or congressional district. Once a geography is chosen, the available indicators appear. The indicators can be browsed from the following categories: demographics, economic well-being, education, family and community, health, safety and risky behaviors, or race and ethnicity. For example, after clicking on a state, a user could display the following data by county: 3-4 year olds who attend preschool; children with dental insurance; children who had a parent who was ever incarcerated by race and ethnicity; or the rate of children in foster care, among many other options. Selecting an indicator produces a table with the data. From there, users can generate a heat map, a line graph, or bar chart. Additionally, if a user has selected

the city, county, or congressional district options, they can add in the state data to compare the localities to the state.

If one prefers to start by selecting indicators, they should use the horizontal menu bar. The “By Topic” option presents the same categories as listed above, and “By Characteristic” offers the following options: race and ethnicity, age, and family nativity. From there, a user can choose from the available indicators in those areas. Eventually, they will be prompted to identify a specific geographical area or use national data.

The source of the data is clearly identified for each indicator along with helpful notes or definitions clarifying the data. Features of the *Data Center* include the ability to download raw data, as well as to save, print, email, share on social media, or generate an embed code for the table, chart, map, or graph generated.

The last menu option on the horizontal bar is “Publications.” KIDS COUNT publications compile the data into reports to provide trend analyses. All publications are available for free to read online or download.

In Short: KIDS COUNT *Data Center* provides high quality data related to child well-being in the U.S. at the national, state, and local levels. The interface is clean and easy to use. One can generate tables, heat maps, graphs, and charts that can be exported, as well as downloading the raw data or the KIDS COUNT publications.

Highly Recommended.

Statistics Canada

<https://www.statcan.gc.ca/>

Visited: Winter 2018

Reviewer: Rhys Stevens, Librarian, University of Lethbridge

After hearing the terms consumer price index (CPI), unemployment rate, retail/wholesale sales, population estimates, life expectancy, and total farm area, one might wonder how these values are calculated. Who provides them for a particular country? Where can they be accessed online? Typically, a national statistical office collects indicators such as these for use by government departments, the business sector, academic researchers, and the general public. In Canada, this central statistical office is named *Statistics Canada*. It is a federal organization responsible for providing and disseminating official statistics via its website for the country of Canada, including its ten provinces and three territories. *Statistics Canada's* website is a rich source of up-to-date statistical data about Canada's population, resources, economy, society, and culture.

Statistics and data available from the *Statistics Canada* website are collected from 350+ active surveys which it is responsible for administering to Canadians. Examples of major recurring surveys include the Census of

Population, Census of Agriculture, Canadian Community Health Survey (CCHS), Labour Force Survey, and General Social Survey (GSS). A selection of “Key indicators” important to Canada and its provinces are derived from these surveys and are displayed prominently on the right-hand side of the home page.

Raw data from all *Statistics Canada* surveys is first cleaned and anonymized before being loaded into the “data portal” (until mid-2018 the data portal database was known as the Canadian Socio-Economic Information Management System (CANSIM)), which is accessed by clicking “Data” in the website’s top navigation bar. The data portal contains over 7,300 tables of statistics gathered from questions asked on active and archived *Statistics Canada* surveys. To find specific statistical tables of interest within the data portal, users filter results by performing keyword searches across table content and browsing by subject category and survey name. Subject category browsing is also accessible separately by clicking “Subjects.” Statistics captured from iterations of particular survey questions are displayed once an identified table is selected. The default table view includes statistics for the entire country over the five most recent reference periods. Tables can normally be customized to display statistics for specific geographic locations or different temporal reference periods. Entire tables or selected data can also be downloaded as Excel-compatible CSV files. The video “How to use the data tables” (<https://www.statcan.gc.ca/eng/sc/video/howto>) provides more information.

The “Analysis” section of the website contains articles and reports produced by analysts at *Statistics Canada* that provide added value from the interpretation of survey statistics and data. In the “Reference” section, there are details about key concepts, definitions, data sources, and survey methodologies. The “Geography” section allows users to find statistics and data by place name and also map geographic regions (e.g., census division, health regions, etc.). Lastly, the “Census program” section includes both the Census of Population and the Census of Agriculture, which are Canada’s most comprehensive surveys and are undertaken every five years. Within this section are specialized data products (e.g., Census Profile, Census Data Viewer, etc.) produced specifically from census program data. Other census program content includes infographics and thematic maps, which serve to visualize results in relevant and insightful ways.

In Short: The *Statistics Canada* website provides official statistical information relating to the commercial, industrial, financial, social, economic, and general activities and condition of the people of Canada. Data tables can be searched, manipulated and viewed online from the “Data” portal section of the site. To identify tables and analytical reports of interest, users perform keywords searches and filter results by subject,

geography or survey name. This site is most relevant for those seeking current or historical Canadian statistics and data sources.

Highly Recommended.

Uniform Crime Report

<https://ucr.fbi.gov/ucr>

Visited: Winter 2018

Reviewer: Karen Evans, Instruction and Research Librarian, Indiana State University

The *Uniform Crime Report (UCR)* was developed in 1929 by the International Association of Chiefs of Police to create a dependable source of crime statistics in the United States. In 1930, the Federal Bureau of Investigation (FBI) took over the process. The *UCR* publishes four yearly publications: Crime in the United States (CIUS), a National Incident-Based Reporting System (NIBRS), Law Enforcement Officers Killed and Assaulted (LEOKA), and Hate Crime Statistics. The publications are created from data voluntarily submitted by more than 18,000 city, university and college, county, state, federal, and tribal law enforcement organizations. The *UCR* also publishes topical reports; recent subjects include cargo theft and human trafficking. The data for all publications is in table format with many publications having a user guide and an overview or summary on their data.

Crime in the United States (CIUS) details the volume and rate of violent and property crime in the United States as a country and for each state. The report provides arrest information on twenty-eight offenses, including murder, rape, and robbery; clearance (cases closed); trends; and law enforcement employee data (law enforcement officers and civilians).

A National Incident-Based Reporting System (NIBRS) was created to improve the crime data collected by law enforcement agencies. The goal was to create more detailed information on crimes committed. For example, NIBRS provides additional information on victims, known offenders, and the relationship between the two. Data is available from 2011 onward; the system (as of 2016) also reports animal cruelty and incidents of identity theft fraud and computer hacking.

The Law Enforcement Officers Killed and Assaulted (LEOKA) publication provides information on officers assaulted as well as those killed feloniously or accidentally. Summaries of officers killed provides detailed information on the situations that led to their murders. A Topic Guide provides additional facts on the incident; topics include weapon information, body armor use, and lighting and environmental conditions. Information is available on methodology and criteria for inclusion in LEOKA.

Hate Crime data provides information on incidents and offenses, victims and offenders, and hate crimes by jurisdiction. Data is reported for single and multi-bias hate crimes. The Hate Crimes publication also includes incidents due to bias against gender, race, religion, sexual orientation, or disability.

Resources are easy to locate on the site; the “Publications” section houses the yearly and topical items. “General Resources” provides information on the proper use of *UCR* data, quality guidelines, and program contacts. “Program Information” houses a FAQ, fact sheets, and definitions (e.g., historical and current definitions of rape). The “Latest Publications” page provides access to the most current data on LEOKA, Hate Crimes, and NIBRS.

In Short: The *Uniform Crime Report* is an excellent, up-to-date resource for anyone looking for information on crime data in the United States.

Highly Recommended.

Data USA

<https://datausa.io/>

Visited: Winter 2018

Reviewer: Dianna L. Kim, Assistant Professor and Research and Instruction Librarian, Newton Gresham Library, Sam Houston State University

The value of information sharing in academic scholarship is undeniable. Rising rates of scholarly publication and the advancement of research has made information access more important than ever before. Academic libraries have an active role in supporting both student educational endeavors as well as the professional teaching and disciplinary research of faculty. Providing recommendations of tools, such as datasets, that assist in meeting research needs is one mechanism of support.

Although the United States government has made numerous datasets available for public use, they are oftentimes difficult to locate or understand. Utilizing an open platform, *Data USA* has transformed several of these datasets to facilitate access, collection, visualization, and analysis.

The result of a partnership between Deloitte, a professional services network, Datawheel, a web platform design company, and the Massachusetts Institute of Technology Media Lab’s Collective Learning Group, *Data USA* is a free web resource that provides curated datasets, maps, and charts using data from several sources including the Bureau of Labor Statistics, O*Net Skills occupational database, Integrated Postsecondary Education Data System, American Community Survey, and Bureau of Economic Analysis.

Data USA's homepage features a single search box. Users also have the option of selecting from categories including locations, industries, occupations, and universities directly from the homepage or via a left-hand drop down menu. *Data USA* organizes data and converts it into visual layouts such as bar graphs, tables, charts, and maps. Visualizations can be saved, shared, or embedded, viewed in plain text lists, or added to a shopping cart allowing the user to make data comparisons. Comparison data tables can then be downloaded into a CSV file for later use. The shopping cart feature is not inherently intuitive, however, and items are sometimes difficult to remove from the cart once placed within it. For this reason, the site would benefit from explanatory materials or a FAQ section.

Data USA also includes a "Stories" section containing articles covering topics such as Medicare, opioid addiction, and education. While the inclusion of articles covering topics of national importance is advantageous, one cause for concern is that the most recent entries (at the time of this review) are dated June 2017, which is somewhat disconcerting.

In Short: *Data USA* spans a broad range of topics including higher education, health care, transportation, job and labor markets, and regional demographics. The site allows users to gain access to a wide array of governmental data quickly and relatively easily and provides visually appealing arrangements of information. Although there are a few caveats, *Data USA* is nevertheless valuable for anyone seeking U.S. government information and statistics and would be a useful tool for academic librarians to recommend to researchers.

Recommended.