

5-2019

Using Data and Statistics

Randy L. Miller

Liberty University, rlmiller5@liberty.edu

Follow this and additional works at: https://digitalcommons.liberty.edu/lib_fac_pubs



Part of the [Higher Education Commons](#), and the [Information Literacy Commons](#)

Recommended Citation

Miller, Randy L., "Using Data and Statistics" (2019). *Faculty Publications and Presentations*. 191.
https://digitalcommons.liberty.edu/lib_fac_pubs/191

This Miscellaneous is brought to you for free and open access by the Jerry Falwell Library at Scholars Crossing. It has been accepted for inclusion in Faculty Publications and Presentations by an authorized administrator of Scholars Crossing. For more information, please contact scholarlycommunications@liberty.edu.

If you are viewing this handout directly in Dropbox, the links to Liberty's subscription databases might not work. Download the handout and open it in Word and then the links should work. You may have to click Ctl-Click rather than just click on them.

What is the difference between Data and Statistics?

In regular conversation, both words are often used interchangeably. In the world of libraries, academia and research there is an important distinction between data and statistics. Data is the raw information from which statistics are created. Put in the reverse, statistics provide an interpretation and summary of data.

Statistics

- Statistical tables, charts, and graphs
- Reported numbers and percentages in an article

If you're looking for a quick number, you want a statistic. A statistic will answer "how much" or "how many". A statistic repeats a pre-defined observation about reality.

Statistics are the results of data analysis. It usually comes in the form of a table or chart. This is what a statistical table looks like:

Table 1206. Adult Attendance at Sports Events by Frequency: 2007

[In thousands (2,343 represents 2,343,000), except percent. For fall 2007. Based on survey and subject to sampling error; see source]

Event	Attend one or more times a month		Attend less than once a month		Event	Attend one or more times a month		Attend less than once a month	
	Number	Percent	Number	Percent		Number	Percent	Number	Percent
Auto racing—NASCAR . . .	2,343	1.1	10,209	4.6	Weekend professional games . . .	4,007	1.8	11,787	5.3
Auto racing—Other	2,384	1.1	7,443	3.4	Golf	1,499	0.7	6,122	2.8
Baseball	7,591	3.4	20,664	9.4	High school sports	10,850	4.9	10,557	4.8
Basketball:					Horse racing:				
College games	3,812	1.7	9,830	4.5	Flats, runners	1,279	0.6	5,860	2.7
Professional games	3,280	1.5	10,996	5.0	Trotters/harness	629	0.3	4,906	2.2
Bowling	1,602	0.7	5,460	2.5	Ice hockey	1,872	0.9	8,499	3.9
Boxing	990	0.5	5,012	2.3	Motorcycle racing	854	0.4	5,127	2.3
Equestrian events	475	0.2	5,177	2.3	Pro beach volleyball	403	0.2	4,729	2.1
Figure skating	391	0.2	5,044	2.3	Rodeo/bull riding	744	0.3	6,333	2.9
Fishing tournaments	740	0.3	4,933	2.2	Soccer	3,437	1.6	6,497	2.9
Football:					Tennis	901	0.4	5,527	2.5
College games	5,759	2.6	12,705	5.8	Truck and tractor pull/ mud racing	904	0.4	5,895	2.7
Monday night professional games	2,165	1.0	6,821	3.1	Wrestling—professional	943	0.4	5,562	2.5

Source: Mediamark Research, Inc., New York, NY, *Top-line Reports* (copyright). See also <<http://www.mediamark.com/mri/docs/TopLineReports.html>>.

Source: [Statistical Abstract of the United States](#)

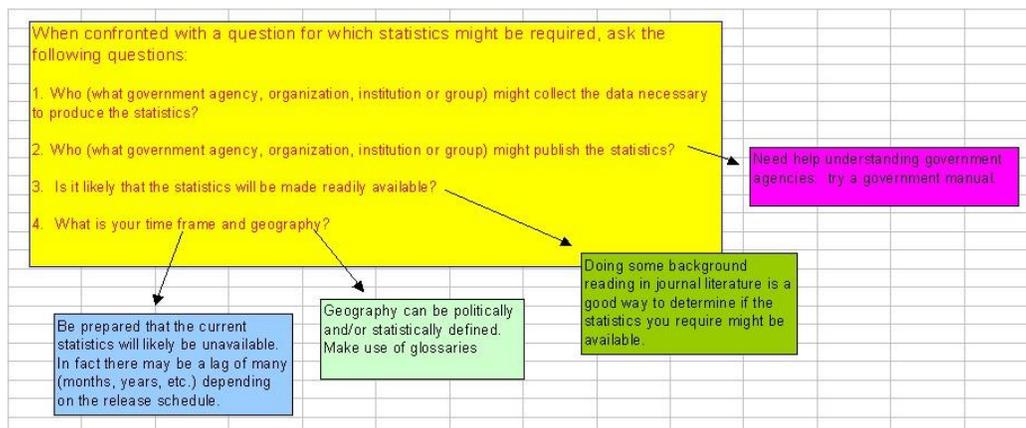
Data

- Datasets
- Machine-readable data files, data files for statistical software programs

If you want to understand a phenomenon, you want data. Data can be analyzed and interpreted using statistical procedures to answer “why” or “how.” Data is used to create new information and knowledge.

Raw data is the direct result of research that was conducted as part of a study or survey. It is a primary source. It usually comes in the form of a digital data set that can be analyzed using software such as Excel, SPSS, SAS, and so on. This is what a data set looks like:

	year	id	wrkstat	hrs1	hrs2	ework	occ
1	2006	1	1	35	-1	0	0
2	2006	2	1	40	-1	0	0
3	2006	3	5	-1	-1	1	0
4	2006	4	2	24	-1	0	0
5	2006	5	6	-1	-1	2	0
6	2006	6	1	37	-1	0	0
7	2006	7	1	40	-1	0	0
8	2006	8	4	-1	-1	0	0
9	2006	9	1	38	-1	0	0
10	2006	10	1	35	-1	0	0
11	2006	11	5	-1	-1	1	0
12	2006	12	8	-1	-1	1	0
13	2006	13	6	-1	-1	1	0
14	2006	14	1	43	-1	0	0
15	2006	15	7	-1	-1	1	0



“Tips and Tricks”

- Try doing a “domain” search (site:.gov or site:.edu)
- Are you looking at actual data figures or statistics based on sampling?
- When you find applicable statistics, look for the source for the statistics and go directly there if possible.
- Try a Google search for “Infographic” along with your search terms
- Realize that data is usually not available immediately. The latest “data” may be several years old. The U.S. Census is only taken every ten years, for instance.
- SPSS is statistical analysis software package. PSPP is a free software product that works much the same way.
- Use the Sage Research Methods database (available as one of our subscription databases at <http://ezproxy.liberty.edu/login?url=http://methods.sagepub.com/>) for reviewing various kinds of statistical analysis and why you would use a certain type for various kinds of research.
- Don't take statistics at face value. Consider the source and method used to create the statistic. Be a critical information consumer!
- Sometimes scholarly journal articles on your topic contain relevant statistics. The article should give a source for their figures.
- Remember that statistics can often lead you back to data.
- Check out the book *Damned lies and statistics: untangling numbers from the media, politicians, and activists* by Joel Best for more information about reading statistics with a critical eye. The website StatLit.org also has great information and resources about statistical literacy. Link for the Best book: <https://ebookcentral-proquest-com.ezproxy.liberty.edu/lib/liberty/detail.action?docID=1021173>

Identify potential producers

Ask yourself: Who might collect or publish this type of information? Then visit the organization’s website and see if you're right! These are some of the main types of producers of statistical information:

Government Agencies

The government collects data to aid in policy decisions and is the largest producer of statistics overall. For example, the U.S. Census Bureau, Federal Election Commission, Federal Highway Administration and many other agencies collect and publish data. To better understand the structure of government agencies read the U.S. Government Manual and browse FedStats. Government statistics are free and publicly available, but may require access through library resources.

Non-Government Organizations

Many independent non-commercial and nonprofit organizations collect and publish statistics that support their social platform. For example, the International Monetary Fund, United

Nations, World Health Organization, and many others collect and publish statistics. For more information about NGOs, visit Duke Libraries NGO Research Guide.

Academic Institutions

Academic research projects funded by public and private foundations create a wealth of data. Try a site:.edu search in Google.

Private Sector

Commercial firms collect and publish data and statistics as a paid service to clients or to sell broadly. Examples include marketing firms, pollsters, trade organizations, and business information. This information is almost always fee-based and may not always be available for public release. The library does subscribe to some commercial data services, particularly through the Business Research Guide.

Search in a Data Archive

Look within a data archive that collects within the general subject area that you are searching for.

- [Inter-University Consortium for Political and Social Research](#)
The world's largest social science data archive. It is one of the best places to start looking for a data set.
- [Data Repositories \(Open Access Directory\)](#)
Open data repositories from multiple academic disciplines.
- [re3data.org: Registry of Research Data Repositories](#)
re3data.org is a global registry of research data repositories that covers research data repositories from different academic disciplines.

Various formats for data and statistics

- .xls is an Excel spreadsheet
- *.csv stands for Comma Separated Values. It can be read by any spreadsheet program (like Rich Text can be read by any word processor)
- *.pdf is like a picture of a document. It can't be sorted, etc. like the above formats.

Government sources for statistics and data

- US Census <http://www.census.gov/>
- American Fact Finder <http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml> (gives information for particular cities and counties including median wages for a variety of professions)

- Statistical Abstract of the United States (no longer published, but the data that was used to create these volumes continues to be published) https://www.census.gov/library/publications/time-series/statistical_abstracts.html (see also the section on Historical Statistics of the United States, Colonial Times to 1970)
- Digest of Education Statistics <http://nces.ed.gov/programs/digest/>
- Occupational Outlook Handbook <http://www.bls.gov/ooh/>
- Statistic sources for various U.S. agencies <https://www.usa.gov/statistics>
- Crime in the United States/Uniform Crime Report <https://www.fbi.gov/services/cjis/ucr>
- United States datasets <http://www.data.gov/>
- Open government datasets <https://www.data.gov/open-gov/>
- CIA World Factbook <https://www.cia.gov/library/publications/the-world-factbook/>

Other data and statistic sources

- NGO Sources http://library.duke.edu/research/subject/guides/ngo_guide/ngo_database
- Industry reports <http://www.valuationresources.com/IndustryReport.htm> (We have a subscription version of this information)
- Television viewing <http://tvbythenumbers.zap2it.com/>
- Nielsen ratings <http://www.nielsen.com/us/en/top10s.html?ranking=snacks>
- Does Wikipedia have a list on your topic? https://en.wikipedia.org/wiki/List_of_best-selling_fiction_authors (If so, look up the sources and go directly to the source)

Public opinion sources

- Christian polling research <https://www.barna.com/> (use for an example of a site:barna.com search in Google)
- Lifeway Research <http://www.lifewayresearch.com/>
- Secular polling research <http://www.gallup.com/home.aspx>

Liberty Subscription resources that include statistical information

- CQ Researcher <http://ezproxy.liberty.edu:2048/login?url=http://library.cqpress.com/cqresearcher>
- Facts on File:Issues and controversies <http://ezproxy.liberty.edu:2048/login?url=http://online.infobaselearning.com/Direct.aspx?pid=WES7&aid=96753>
- Global Issues in Context http://ezproxy.liberty.edu:2048/login?url=http://find.galegroup.com/menu/start?prod=GIC&userGroupoupName=vic_liberty
- World Christian Database (contains information on world's religions, missions, denominations, etc. <http://ezproxy.liberty.edu:2048/login?url=http://www.worldchristiandatabase.org/wcd/>
- Many of our business databases include statistical information – Business Research Guide <http://libguides.liberty.edu/content.php?pid=544017&sid=4475347>
 - Sports Market Analytics <http://ezproxy.liberty.edu/login?url=http://sportsmarketanalytics.com>
 - Country Watch <http://ezproxy.liberty.edu/login?url=http://www.countrywatch.com/ip>

- Reference USA (allows you to create mailing or contact lists)
<http://ezproxy.liberty.edu:2048/login?url=http://www.referenceusa.com/Home/Home>
- Dismal Scientist
<http://ezproxy.liberty.edu:2048/login?url=http://www.economy.com/dismal/>
- Other Library subscription statistics sources
 - [American Mathematical Society](#) Reviews, abstracts, and bibliographic information for mathematical sciences literature, including bibliographic data for articles dating back to the early 1800s
 - [Annual Reviews \(Statistics and Its Application\)](#)
 - [CQ Press Vital Statistics on American Politics](#)
 - [CQ Press Vital Statistics on the Presidency](#)
 - [Data-Planet](#) The largest repository of standardized and structured statistical data
[Video Tutorial](#) on using Data-Planet
 - [PolicyMap](#) Data application with mapping capabilities that provides demographics, real estate, health, jobs, and more for communities across the U.S. Data is available in interactive maps, tables, charts, and reports. The data is downloadable as comma-separated value (.csv) files, and the maps can be printed, emailed, or saved as image files.
 - [Polling the Nations](#) Polling the Nations is a compilation of more than 14,000 surveys conducted by over 1,000 polling organizations in the United States and 100 other countries from 1986 to the present.

Analyzing your own statistical information

- Use SPSS (or PSPP) for quantitative data
- See “Which Stats Test?” under Research Tools in Sage Research Methods (and you can see further explanations of each statistical test).
<http://ezproxy.liberty.edu/login?url=http://methods.sagepub.com/>
- Use NVivo for analyzing qualitative data. (Available for free download by Liberty University.)
<https://www.liberty.edu/informationsservices/index.cfm?PID=41536>