What’s the difference?

Statistics: The dictionary definition of "statistics" refers to numeric indicators of nations. Popular usage of the term points to numeric summaries that condense information, or numbers that are used to make comparisons, or numbers that portray relationships or associations. Statistics are produced from data; data that has been analyzed or processed in some fashion.

Aggregate Data: data that have been aggregated or totaled from smaller units into a large unit. Example: "The Census Bureau aggregates data to preserve the confidentiality of individuals."

Data: the raw material out of which social and economic statistics are produced. Social science data originate from social research methodologies or administrative records, while statistics are produced from data. Data are the information collected and stored at the level at which the unit of analysis was observed. Summaries of these data are usually statistics. Data must be processed to be of practical use. This compilation is accomplished with statistical software, which reads the raw data from a computer file.

(Definitions taken from Glossary of Computing and Social Science Terms, http://3stages.org/glossary/)

Resources at UCLA

UCLA Library, http://guides.library.ucla.edu/publicpolicy
Provides research assistance to identify data and statistical sources.

ISR Data Archives, http://www.sscnet.ucla.edu/issr/da/
Provides support for quantitative research including identifying, acquiring and archiving data and assistance in using data with various statistical packages.

Provides a wealth of online tutorials, guides, etc. for major statistical packages. They feature a lending library as well as offer walk-in consulting.

Statistical Consulting Center, UCLA Department of Statistics, http://scc.stat.ucla.edu/
Provides free walk-in and online assistance with statistical and data analysis.

GIS @ UCLA, MapShare, http://gis.ats.ucla.edu/
Provides training and assistance with GIS. Online access to a wide variety of geospatial data sets; and regular updates about ArcGIS and other products.

KK Winter 2010
## Defining Your Research

<table>
<thead>
<tr>
<th>Geography</th>
<th>Physical areas</th>
<th>Government units</th>
<th>Political boundaries</th>
<th>Other – school districts, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time period</td>
<td>Current</td>
<td>Historical</td>
<td>Over time</td>
<td></td>
</tr>
<tr>
<td>Population &amp; Units of analysis</td>
<td>Gender, age, etc.</td>
<td>Race, ethnicity, ancestry, etc.</td>
<td>Persons</td>
<td>Households</td>
</tr>
<tr>
<td>Questions or topics</td>
<td>What questions do you want to answer? How are they covered in different data or statistical sources?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variables</td>
<td>What variables or item of data do you need to answer your research question?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td>What is your hypothesis? How does the data fit into your overall study?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Geography:** Since most data is geo-coded in some way it is important to define the level of geography you are studying. Keep in mind, not all data and statistics may be collected at the geography you want. Examples: Los Angeles city vs. county; a neighborhood; a state; a metropolitan statistical area; zip codes.

**Time period:** Let’s say we ask how many immigrants live in Los Angeles – do we want to know how many live here now as opposed to ten years ago? Do we want to measure any change in the number living here and/or moving away?

**Population/Units of Analysis:** Are you looking at a specific group or subgroup of a population, e.g. females over 65. Units of analysis are the observable entity of the study – this could be an individual, household, family, housing structures, schools, etc.

**Questions/Topics:** Think about the information you want find – what about your population do you want study or investigate? Voting habits? Political participation? Socialization? Access to services?

**Variables/Data points:** Do you want to know the age, gender, educational attainment, interest rates, etc. Based on your question/topic, what variables will provide you with the information to analyze and test your hypothesis. Keep in mind, that you may not be able to find every variable in just one source.

**Context:** Finally, what is the hypothesis in your work? How will the data points you select help you test that hypothesis?
Data Discovery Continuum

**News Articles**: You can look for information in the popular press discussing recent studies. You can find news articles using a resource like LexisNexis.

**Scholarly Articles**: Use the article databases, like the Social Sciences Citation Index, to locate analyses and find research about studies discussed in the news.

**Statistical Sources**: Government agencies and NGOs, at the local, state, national, international level, provide a great deal of statistical information.

**Data Sources**: This is the raw data that are arranged by rows and columns in flat files. We have access to a wealth of data sets through ICPSR, the UCLA ISR Data Archive or a government site such as the Census Bureau.

**Study Details**: In order to use the data well, you need to understand how the data were collected and who participated in the survey. These can be found in study documentation to describe sampling and survey methodology and in scholarly articles.

**Questionnaires**: You also want to know what questions were asked and in what order to determine if a data set is going to contain the variables you want. Sometimes the questionnaire is included in an appendix of a scholarly article or as part of the codebook.

**Codebooks**: A codebook is a document that describes the format and arrangement of the data file.

**Data Analysis**: Sometimes you won’t have the data points you seek until you actually analyze your data with a software tool or statistical package.

**Visualization**: We can often learn more about a data source by visualizing it as a chart, table, graph, map or other kind of graphical display. See sites such as UCLA MapShare, ArcGIS, Swivel, or Many Eyes.
Evaluating scholarly articles containing statistical information

1. Scholarly vs. popular
   a. Using popular opinions and beliefs in research: a social science approach
   b. Popular articles and reports vs. Scholarly articles about data

2. Data producers
   a. Government agencies (Census Bureau, National Center for Health Statistics)
   b. Academic institutions (research centers, individual researchers)
   c. Non-profit organizations (be careful here)

3. Components of reports, articles on data collection:
   a. Describes the organization, the study and why the data were collected
   b. Describes the population or sample
   c. Provides major findings; includes margin of error
   d. Provides list of questions asked
   e. Sometimes provides actual raw data
   f. Includes contact person(s), name of organization
   g. Caveat: analysis of findings provided by journalist not scholar

4. Components of a scholarly article containing statistical information:
   a. Author’s institutional affiliation
   b. Abstract describing article
   c. Introduction – outline of theory, history of topic, literature review
   d. Questions or hypotheses
   e. Methods, data, analysis
   f. Results, conclusions, tables
   g. Notes
   h. Bibliography
   i. Acknowledgements
   j. Caveat: May be a working paper or conference presentation instead of a publication in a journal or book.
Government Collected and Produced Statistics

Starting Points:
- Statistical Programs of the United States Government: Annual publication from the OMB that describes in detail the federal agencies and programs that have a data component [http://www.fedstats.gov/policy/](http://www.fedstats.gov/policy/)
- ICPSR [http://www.icpsr.umich.edu/icpsrweb/ICPSR/](http://www.icpsr.umich.edu/icpsrweb/ICPSR/)

Principal Statistical Agencies – their sole purpose/mission is to gather data and publish statistics on the American people, structures, programs, etc.
- Bureau of Economic Analysis
- Bureau of Justice Statistics
- Bureau of Labor Statistics
- Bureau of Transportation Statistics
- Census Bureau
- Economic Research Service, Dept of Agriculture
- Energy Information Administration
- Environmental Protection Agency
- Internal Revenue Service, Statistics of Income
- National Agricultural Statistics Service
- National Center for Education Statistics
- National Center for Health Statistics
- National Science Foundation, Science Resources Statistics
- Office of Management and Budget
- Social Security Administration, Office of Policy

Federal Agencies with Statistical Programs – support program planning and evaluation or are an outgrowth of their administrative responsibilities. Statistical program is not always executed by the agency that sponsors it, i.e. an agency may conduct the work for another agency. (for a full list, go to FedStats [http://www.fedstats.gov/agencies/index.html](http://www.fedstats.gov/agencies/index.html))
- Drug Enforcement Administration [Dept of Justice]
- Federal Highway Administration [Dept of Transportation]
- Environmental Protection Agency [Independent agency]
- Fish and Wildlife Service [Dept of Interior]
- National Aeronautics and Space Administration [Independent agency]
- National Oceanic and Atmospheric Administration [Dept of Commerce]
- Occupational Safety and Health Administration [Dept of Labor]
- Office of Federal Housing Enterprise Oversight [Dept of Housing and Urban Development]
- Patent and Trademark Office [Dept of Commerce]
- Substance Abuse and Mental Health Services Administration [Dept of Health and Human Services]
- United States Geological Survey [Dept of Interior]
Components of a table

Statistical information is presented in tables in a variety of ways; some are more explanatory than others. Look at the table below and note the bibliographic and informational components.

<table>
<thead>
<tr>
<th>Adults &amp; teens use online networks to say in touch with pre-existing friends</th>
<th>Adults</th>
<th>Teens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you use your online profile to…?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stay in touch with friends*</td>
<td>89%</td>
<td>91%</td>
</tr>
<tr>
<td>Make plans with friends</td>
<td>57</td>
<td>72</td>
</tr>
<tr>
<td>Make new friends</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>Organize with others for an event, issue or cause</td>
<td>43</td>
<td>n/a</td>
</tr>
<tr>
<td>Make new business or professional contacts</td>
<td>28</td>
<td>n/a</td>
</tr>
<tr>
<td>Promote yourself or your work</td>
<td>28</td>
<td>n/a</td>
</tr>
<tr>
<td>Flirt</td>
<td>20</td>
<td>17</td>
</tr>
</tbody>
</table>

Source: Adult internet user data comes from the Pew Internet & American Life Project May 2008 Tracking Survey. For internet users is 1,063 and the margin of error is ±3%. Teen data is from the Teen & Parent survey fielded in Oct-December 2006. Margin of error is ± 4% for online teens.

*This question was posed slightly differently to teens: “Do you ever use [social networking] sites to stay in touch with friends you see a lot?” There was also a complementary question asked of teens: “Do you ever use those sites to stay in touch with friends you rarely see in person?” and 22% of teens who use social networks responded yes.

In general, the following apply:

**RULE 1.** Read the data with a specific purpose in mind.

**RULE 2.** Find out the source of the data and the purpose for which they were collected.

**RULE 3.** Any number that is interesting is probably wrong.

**RULE 4.** Pay particular attention to summaries of summary measures.

**RULE 5.** Do not analyze the data until you know why and how they were collected. (See RULE 2.)

**RULE 6.** Test the data against common sense. If they conflict with intuition, trust the intuition and check the numbers.”

**RULE 7.** Always read a table beginning with the general and then move to the specific.

Title: “Adults & teens use online networks to stay in touch with pre-existing friends”

Source of Table: Pew Internet & American Life Project May 2008 Tracking Survey.

Columns: Adults, Teens

Rows: Uses of online profile

Rules for Reading Tables


Accessed: November 12, 2004 5.49 pm.

“Having got a basic feel for the table, we can begin to make sense of the data and their organization. In summary, the evaluation of data depends upon the perspective and the purpose for which they are intended. It is useful, at the outset to establish a goal for the data. Once that objective is clear, it is easier to answer questions regarding their use and to evaluate their value."
Evaluation Tool for Public Opinion Polls and Surveys
-Stephen Woods, Social Science Librarian, The Pennsylvania State University Libraries

1. What is the purpose of the poll/survey?
2. Who sponsored the poll/survey?
3. What polling organization conducted the poll/survey?
4. What questions were asked?
5. What was the order of the questions?
6. Who was polled/surveyed?
7. How were the interviews conducted?
8. What was the date the poll/survey was conducted?
9. What statistics offered substantiate accuracy? For example, was a standard error range provided?

Sample Poll Question taken from the Roper Center for Public Opinion Research

Which one of the following issues is most important in determining your vote for president this year (2008)...the economy and jobs, taxes and government spending, terrorism and national security, health care, the Iraq war, issues like abortion, guns, and same-sex marriage, or energy policy and gas prices?

<table>
<thead>
<tr>
<th>Issue</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>The economy and jobs</td>
<td>44</td>
</tr>
<tr>
<td>Taxes and government spending</td>
<td>14</td>
</tr>
<tr>
<td>Terrorism and national security</td>
<td>12</td>
</tr>
<tr>
<td>Health care</td>
<td>8</td>
</tr>
<tr>
<td>The Iraq war</td>
<td>7</td>
</tr>
<tr>
<td>Issues like abortion, guns, and same-sex marriage</td>
<td>6</td>
</tr>
<tr>
<td>Energy policy and gas prices</td>
<td>5</td>
</tr>
<tr>
<td>Other/None of these (Vol.)</td>
<td>1</td>
</tr>
<tr>
<td>Don't know</td>
<td>3</td>
</tr>
</tbody>
</table>

Subpopulation/Note: Registered voters
Topics: Vote For President Problems

Source: Survey by Newsweek.
Methodology: Conducted by Princeton Survey Research Associates International, October 22 - October 23, 2000 and based on telephone interviews with a national adult sample of 1,092.
Data provided by The Roper Center for Public Opinion Research, University of Connecticut.
[USPSRNEW.102408.R09]

KK Winter 2010
Citing Data Sets

Citations of machine-readable data files contain:

- **Author(s)** of the original study, whether individual(s) or institution(s)
- **Title** of the original study
- **[Computer file]** to indicate use of an electronic data set
- **City and state** of the data *producer* - usually the institution sponsoring the research (like the information about a publisher in a book citation)
- **Name** of the data *producer* with the designation [*producer*]
- **Year** the data was produced

Examples:


Resources:

A Style Manual for Machine Readable Data Files and Their Documentation

How to Cite a Data Set, North Carolina State University Library
http://www.lib.ncsu.edu/data/citingdatasets.html

How to Cite Roper Center Data Sets (Opinion Polls)
http://www.ropercenter.uconn.edu/data_access/data/how_to_cite.html

How to Cite a Data Set, International Polar Year Data and Information Service
http://ipydis.org/data/citations.html