



IN THIS ISSUE

The Flowering of Online Data
Analysis 1
Latinobarómetro 2
Country Statistical
Yearbooks 2
International Data at ICPSR . 2
New Studies at DISC 3
Crossroads Corner 4

.....

DISC News contains articles about local, national and international data issues. It is published twice a semester by the library staff and sent to faculty, graduate students, and librarians, with a focus on the Social Science departments.

Editor: Joanne Juhnke, Special Librarian
Staff contributors: Lu Chou, Senior Special Librarian

.....

University of Wisconsin-Madison
1180 University Drive
3308 Sewell Social Sciences Building
Madison, Wisconsin 53706 USA

Phone: (608) 262-0750
E-mail: disc@mailplus.wisc.edu
Web: <http://www.disc.wisc.edu/>
Hours: Monday-Friday 8:30-4:30

.....

This issue of DISC News is online at:
<http://www.disc.wisc.edu/pubs/Newsletters/nov07news.html>

Please Note

DISC will be closed:
Mon.-Tue. Dec. 24-25 for Christmas
Mon.-Tue. Jan. 1-2 for New Year's
Mon. Jan. 15 - Martin Luther King, Jr. Day

THE FLOWERING OF ONLINE DATA ANALYSIS

by Joanne Juhnke

Since at least the mid-1990s, projects have been underway to use the World Wide Web to allow users to analyze social science data online. Many sites have for years been using online technology to generate and display tables. However, the past few years have seen a dramatic increase in data available via online systems that include statistical analysis functions. Online data analysis, it seems, is flowering worldwide. This article looks at two such systems: SDA and Nesstar. Several additional programs in use at individual sites are mentioned briefly at the end of the article.

The past few years have seen a dramatic increase in data available via online systems that include statistical analysis functions.

The data analysis package SDA, which stands for Survey Documentation and Analysis, was

developed by the Computer-Assisted Survey Methods Program (CSM) at the University of California, Berkeley. Users of data that has been made available via SDA can do frequencies and crosstabulations, comparisons of means, correlation matrix, and regressions. SDA also allows users to re-code variables and create new variables. One particular strength of SDA is its ability to make and download user-specified subsets together with the documentation specific to those subsets. Search functionality is still in development; variable searching in SDA is only available on a few datasets so far.

The archive of SDA data at the Berkeley site may be found at <http://sda.berkeley.edu/archive.htm>, linking to the General Social Survey (GSS), the American National Election Study (ANES), and more. Other organizations that have opted to use SDA include:

- CPANDA, the Cultural Policy and the Arts National Data Archive at Princeton University, at <http://www.cpanda.org/>. CPANDA includes a growing collection of studies relevant to the arts and cultural policy in the United States, and all of the datasets in the collection are usable with SDA.
- Three major projects from the Minnesota Population Center: IPUMS USA at <http://usa.ipums.org/usa/>, IPUMS CPS at <http://cps.ipums.org/cps/>, and NAPP, the North Atlantic Population Project at <http://www.nappdata.org/napp/>. The two IPUMS projects cover multiple years of harmonized public use microdata, IPUMS USA from the US Decennial Census, and IPUMS CPS from the Current Population Survey. NAPP covers census data from the late 1800s from five North Atlantic Countries (see Crossroads Corner, p. 4, for more!)

Continued on p. 3

LATINOBARÓMETRO

As announced in the list of *New Studies at DISC* in the September 2007 issue of this publication, DISC has recently acquired data for the ten rounds of the Latinobarómetro conducted between 1995 and 2005.

The Latinobarómetro (or Latino-barometer) is an annual public opinion survey carried out in 18 Latin American countries representing a total population of 400 million. The survey first started in 8 countries in 1995, growing to 17 countries in 1996 with one additional country joining in 2004. Built on the model of the long-standing Eurobarometer, the Latinobarómetro seeks to provide insight into public opinion across countries on topics such as the economy, trade, politics, social participation, and the environment, among others.

The Latinobarómetro includes basic questions that are asked from year to year, providing opportunity for comparisons across time. The survey also takes on a different thematic issue every year, in addition to questions pertaining to current events. The principal theme for 2000 was poverty; 2002 focused on perceptions of democracy and the market; and 2003 introduced questions on taxes and on corporate social responsibility.

For more information search on “Latinobarometer” in the DISC online catalog, or visit the Latinobarómetro web site at <http://www.latinobarometro.org/>.

COUNTRY STATISTICAL YEARBOOKS

If you’ve ever looked up a table in the Statistical Abstract of the United States, you know what a helpful tool a statistical yearbook can be, whether for looking up a quick factoid or seeking sources for a larger data search.

Visitors to the DISC web site can now take that strategy and make it international. It turns out that many countries not only create statistical yearbooks but also make them available online. A unique new feature on the DISC web site called Country Statistical Yearbooks brings these resources together in one place, at <http://www.disc.wisc.edu/yearbooks/>.

At 84 countries and counting, Country Statistical Yearbooks is a growing resource. Countries are both listed alphabetically and grouped by continent. DISC has found and linked to multi-country statistical yearbooks as well, such as the Nordic Statistical Yearbook and the Statistical Yearbook of Latin America and the Caribbean. We look forward to adding new countries to the total; please let us know at disc@mailplus.wisc.edu if you discover one that we’ve missed!

INTERNATIONAL DATA AT ICPSR

ICPSR has recently made available two new options for accessing international data in the ICPSR archive (<http://www.icpsr.umich.edu/>).

In October, ICPSR announced a metadata revision, introducing a new tag that specifies whether or not a dataset has coverage extending beyond the United States. On the advanced search page, users can restrict their searches to international data by including the term “global” in the geographic coverage field.

In November, ICPSR announced the launch of the International Data Resource Center (IDRC) web site, at <http://www.icpsr.umich.edu/IDRC/>. Researchers can use the IDRC as a gateway to ICPSR’s international data holdings. International data available through the IDRC includes conflict data, economic data, data on electoral systems and political behavior, environmental data, health data, data pertaining to the human dimension of international relations, public opinion data, and data on international organizations.

Searches for international data in the IDRC can be conducted using several different methods, including subject searches, series data searches, or the IDRC’s interactive map interface. The site also includes instructional resources and links to core datasets and related citations in the field of international studies.

THE FLOWERING OF ONLINE DATA ANALYSIS

Continued from p. 1

- The Roper Center for Public Opinion Research, at http://www.ropercenter.uconn.edu/data_access/ideas/dataanalysistool.html. Roper introduced their SDA-based data analysis tool, called IDEAS, very recently, and only six of the many datasets in the Roper collection are currently available for online analysis.

Nesstar

Developed as a joint project between Norwegian Social Science Data Services (NSD), UK Data Archive and the Danish Data Archive (DDA), the Nesstar product was designed not only to make it possible to analyze and subset data online, but also to facilitate variable searching across multiple surveys and data catalogs. As part of this design, Nesstar also lets the organizations that use its product set varying levels of security and passwording for user access. Nesstar allows users to run frequencies and cross-tabulations, run correlation and regression analyses, create new variables and recode existing variables. While Nesstar allows analysis and downloading of subsets, documentation download is not customized for those subsets. The Nesstar home page, at <http://www.nesstar.com/>, provides documentation and description for the product, but the company does not itself archive data. Nesstar-based archives include:

- CESSDA, the Council of European Social Science Data Archives, at <http://www.nsd.uib.no/cessda/home.html>. This massive collaboration between 13 European social science data archives comprises 5287 studies to date. The entire holdings are searchable, but different archives provide different levels of user access (some entail a simple free registration, others have more stringent applications and criteria).
- The BADGIR archive here at UW-Madison, at <http://nesstar.ssc.wisc.edu/index.html>. The BADGIR archive contains the National Survey of Families & Households (NSFH) and the Health, Wellbeing and Aging in Latin America and the Caribbean (SABE), along with 25 studies from the DISC archive; free registration is required to analyze or download data.

Nesstar and SDA

Some data is available for analysis via both Nesstar and SDA:

- ICPSR, at <http://www.icpsr.umich.edu/ICPSR/access/sda.html>, now uses both SDA and Nesstar to make selected datasets from its archive available for online data analysis. Several hundred datasets are available through each of the two programs.
- The General Social Survey (GSS). The GSS 1972-2006 is available in SDA at Berkeley at <http://sda.berkeley.edu/archive.htm>, and in Nesstar at the National Opinion Research Center at http://publicdata.norc.org:41000/gssbeta/gss_nesstar.html.

Beyond SDA and Nesstar, here are some other links to explore for a sampling of other interfaces used solely by the organizations that developed them:

- NAEP Data Explorer for the National Assessment of Educational Progress at the National Center for Education Statistics (NCES), at <http://nces.ed.gov/nationsreportcard/naepdata/>.
- The Data Analysis System (DAS), <http://nces.ed.gov/das/>, also at NCES, covering many NCES surveys.
- The NLS Web-Investigator at Ohio State University, at <http://www.nlsinfo.org/web-investigator/index.php>, covering all cohorts of the National Longitudinal Study.

NEW STUDIES AT DISC

- American Housing Surveys (AHS), 1997-present.
- Capturing Campaign Dynamics: National Annenberg Election Surveys, 2000 and 2004.
- Common Core of Data (CCD), 1986-present.
- E-STAT.
- National Longitudinal Survey of Youth, Main File and Event History File for Rounds 1-9: 1997-2006.
- National Longitudinal Surveys of Labor Market Experience, Youth Cohort, 1979: Young Adults, 2006, early release.
- Survey of Consumer Finances, 1962-current.

Crossroads Corner highlights web sites recently added to the searchable Internet Crossroads in Social Science Data, available on the DISC web site at <http://www.disc.wisc.edu/newcrossroads/index.asp>.

North Atlantic Population Project (NAPP)

The North Atlantic Population Project (NAPP) was inspired by the fact that five North Atlantic countries have individual-level digitized census data for the late 1800s: Canada, Great Britain, Iceland, Norway, and the United States. These five countries shared economic ties and migration flows during this time, when many social, economic, and demographic changes were afoot. Hosted by the Minnesota Population Center, NAPP is harmonizing and providing online access to both complete-count and sample census data. Users can create and download extracts, or analyze data online using the SDA interface. A free registration is required: NAPP staff reviews the registrations, and potential users must provide a description of their proposed research, and agree not to redistribute the data or use it for genealogical research.

The NAPP site is online at <http://www.nappdata.org/napp/>.

Center for Population Health and Health Disparities Data Core

The Center for Population Health and Health Disparities (CPHHD) of the RAND Corporation has made available online a collection of public use datasets designed for analyzing disparities in cost-of-living, disability, pollution, population and housing characteristics, segregation, street connectivity, and neighborhood socioeconomic status in the United States. The datasets are derived from public-use data from the U.S. Census, the American Chamber of Commerce Research Association, and the U.S. Environmental Protection Agency.

Most of the data covers the 1990-2000 time period. The data is available for various geographic summarization areas including census tract, county, and MSA, and has also been put into both 1990 and 2000 geographical definitions. Data formats include SAS, Stata, and CSV. Free registration is required, along with a description of the research and who else is collaborating on the project, and registrations are reviewed before access is granted.

The CPHHD Data Core may be found at <http://www.rand.org/health/centers/pophealth/data.html>.

Swivel

Swivel, online at <http://www.swivel.com/>, offers a data sharing utility in the collaborative spirit of Web 2.0. As a self-proclaimed purveyor of “tasty data goodies,” Swivel's ambitious mission is “to liberate the world's data and make it useful so new insights can be discovered and shared.” Anyone can upload data to the Swivel site, either via copy-and-paste or uploading from CSV, Excel or Google Spreadsheets. Those who upload can sort and filter, map geographical data, plot charts or graphs, describe and categorize and tag the dataset, cite the data source, and make it available for other users to comment and download. Swivel also provides an opportunity for organizations to register as “official” entities, to show that the data is coming from the organization that collected it. Official sources that have signed up so far include the OECD, the World Health Organization, the U.S. Department of Commerce, and the Newspaper Association of America.

The current Swivel site is still billed as a preview edition. While the public edition is free, Swivel is also working on a fee-based private edition that will enable users to collaborate in a more secure environment and compare data without making it openly available.