



IN THIS ISSUE

What is (are) Data? 1
 10 Questions, 10 Minutes, No
 Changes: Census 2010 2
 News from ICPSR 2
 New Studies at DISC 3
 Finding Data Sets in the Social
 Sciences 3
 Public Comment on Public
 Data 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: Cindy Severt,
 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/nov09news.html>

Note:

DISC will be closed

- Thu. Nov. 27: Thanksgiving
- Fri. Nov. 28: State Furlough Day
- Thu. Dec. 24: Christmas Eve
- Fri. Dec. 25: Christmas Day
- Wed. Dec. 30: State Furlough Day
- Thu. Dec. 31: New Year's Eve
- Fri. Jan. 1: New Year's Day
- Mon. Jan. 18: Martin Luther King Jr. Day

WHAT IS (ARE) DATA?

by Cindy Severt

If you are reading this, and you believe the correct verb in the title should be “are,” you probably cleave to the axiom that data (derived from the Latin singular *datum* meaning *a thing given or granted; something known or assumed as fact, OED*) are coded numeric responses to questions asked in a survey, and anything else—especially anything born digital—is simply *not* data. You would not be wrong, but you may be feeling increasingly lonely in your position!

Now that nearly every academic department is producing (and sometimes drowning in) massive quantities of digital material as a result of research, the very word data has been co-opted by every department to fit its ontology. Digitized text, microscopy images, scanned photographs, characters on a page, sound recordings, even physical artifacts are considered data. There are now as many definitions for data as there are people to define them.

Data are the irrefutable building blocks from which information, knowledge, and wisdom are derived.

How is one to get around these multitudinous meanings and still communicate clearly, especially if one is serving on a campus-wide data curation committee? Put simply, data are the irrefutable building blocks from which information,

knowledge, and wisdom are derived. For example the various conspiracy theories (knowledge) about the assassination of President John F. Kennedy all hinge on evidence (information) gathered as a result of analyzing the 486 frames (data) of the Zapruder film.

In another example, *data* are the individual answers given to the questions asked on the monthly Current Population Surveys. *Information* is the monthly unemployment rate by age group and sex, extracted from the survey. *Knowledge* is the observation that most single mothers who had their first child in their teens have usually not completed high school, and have high rates of unemployment. *Wisdom* is implementing a policy to establish the infrastructure to enable single mothers who are high-school dropouts to graduate from high school and get some career training with the social objective of getting single mothers into the labor force and off welfare.

What both of these examples have in common is secondary analysis: different conclusions drawn from analysis of the same data. Hence the varied conspiracy theories based on the Zapruder film, and the many different policies that are the potential outcome of the Current Population Survey.

Continued on p. 3

10 QUESTIONS, 10 MINUTES, NO CHANGES: CENSUS 2010

by Joanne Jubnke

The US decennial census, coming up on April 1, 2010, will move forward with the 10-question "short form" as originally planned, despite a recent political challenge.

As required by law, the Bureau had brought the topics of the Census to Congress for approval three years ago, and the wording of the questions two years ago. At the time, there were no objections. However, in October 2009, Sen. David Vitter (R-LA) and Sen. Robert Bennett (R-UT) filed an amendment to the Fiscal Year 2010 Commerce, Justice, and Science Appropriations bill (H.R. 2847). The amendment proposed to withhold federal funding for Census 2010 unless the questionnaire was modified to ask about U.S. citizenship and immigration status.

On November 5th, the Senate voted 60-39 to end debate on the bill, blocking the amendment from further consideration and freeing Census 2010 to proceed as planned.

According to both current and former Census Bureau directors, adding an untested question would have put the accuracy of the enumeration in jeopardy. The cost of changing the already-printed questionnaire and the accompanying informational campaign would have been immense, and the Census and all its products, including apportionment numbers, would

Continued in the next column

NEWS FROM ICPSR

by Joanne Jubnke

2010 Undergraduate Summer Internship

ICPSR is now accepting applications for its ten-week 2010 summer undergraduate internship program in Ann Arbor, Michigan. The 2010 internship runs from June 7 to August 13. Interns will gain experience using statistical programs such as SAS, SPSS, and Stata for data checking and processing. Interns also attend courses in the ICPSR Summer Program in Quantitative Methods of Social Research, in addition to a weekly Lunch and Lecture series. Applicants must have completed their sophomore year in a social science major by the time of the internship. Compensation includes a stipend, room and board in university housing, and a scholarship covering the ICPSR Summer Program participation. Full details are available at <http://icpsr.blogspot.com/2009/10/icpsr-summer-undergraduate-internship.html>; application deadline is February 8, 2010.

2008 SETUPS Module

ICPSR recently launched its 2008 SETUPS web site, an online tool for investigating how factors like income, religion, and race affected the way people voted in the 2008 elections. *Voting Behavior: The 2008 Election*, online at <http://www.icpsr.umich.edu/SETUPS2008/>, is the latest in a series of instructional modules known as SETUPS (Supplementary Empirical Teaching Units in Political Science). The 2008 SETUPS presents a data set drawn from the 2008 American National Election Study (ANES). The core of the module is a set of analysis exercises designed to develop students' ability to understand and analyze survey data. The entire module can be accessed online, including the data analysis components.

Charles Prysby, professor of political science at the University of North Carolina at Greensboro, and Carmine Scavo, associate professor of political science at East Carolina University in Greenville, North Carolina, have coauthored the voting behavior SETUPS since 1984. The SETUPS series began with the 1972 election, and were published in print form until 2004, when the first online module was produced.

10 QUESTIONS, 10 MINUTES, NO CHANGES

Continued from previous column

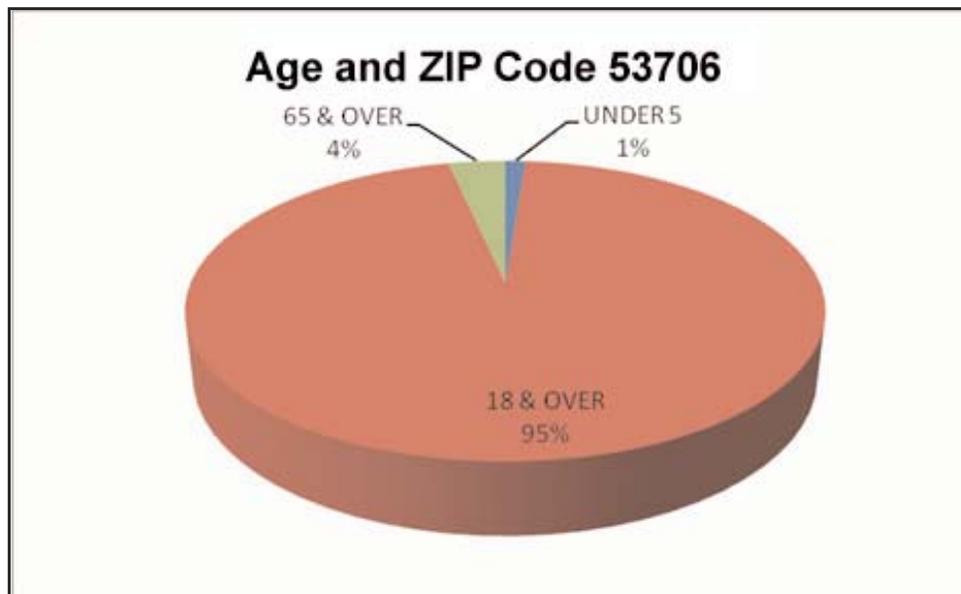
surely have been delayed. (Statement of Former Census Directors on Adding a New Question to the 2010 Census, <http://www.thecensusproject.org/letters/cp-formerdirs-16oct2009.pdf>)

Although the ongoing American Community Survey does ask a question regarding citizenship, the decennial census has never done so. Citizenship has never been a factor in determining Congressional representation, in accordance with the US Constitution. For further information, visit The Census Project web site at <http://www.thecensusproject.org/>.

WHAT IS (ARE) DATA?

Continued from p. 1

Now consider the following pie chart:



The pie chart above is a graphical representation of the categorical variable, age for ZIP code 53706. The population of this ZIP code is overwhelmingly 18 years and older, but neither this piece of information nor the pie chart itself is data; the data are the self-reported ages of individuals living in ZIP code 53706 as collected for the 2000 decennial census. All those various ages have been summarized into three categories, and made retrievable with a mouse click from the Census Bureau's website which I entered into an Excel spreadsheet and subsequently created the pie chart. The pie chart is static; data are not. I could have just as easily sliced the pie differently using different age categories. And with other data ingredients I could have made an entirely different flavored pie altogether.

NEW STUDIES AT DISC

- China Health and Retirement Longitudinal Study (CHARLS)
- English Longitudinal Study of Ageing (ELSA)
- General Social Surveys, 1972-2008 Cumulative File and 2008 Individual Year Survey
- Health and Retirement Study (HRS)
- Korean Longitudinal Study of Ageing (KLoSA)
- Political Constraint Index (POLCON) dataset
- Study of Global Ageing and Adult Health (SAGE)
- Survey of Health, Ageing and Retirement in Europe (SHARE)

FINDING DATA SETS IN THE SOCIAL SCIENCES

On Tuesday November 24, 5-6:30 pm, DISC librarian Cindy Severt will be presenting a workshop on Finding Data Sets in the Social Sciences, through the UW-Madison Libraries' Graduate Support Series. The workshop will cover strategies, search tools and resources for finding social science numbers, data, and data sets available through DISC and the UW-Madison libraries. Registration is free and open to the public:

<http://library.wisc.edu/workshops/workshop-date.html#nov>

PUBLIC COMMENT ON PUBLIC DATA

The Association of Public Data Users (APDU) is a national organization whose members use, produce, and distribute government statistical data. Advocacy is part of APDU's mission; to that end the organization monitors and publicizes opportunities for public comment on federal data collections.

APDU's advocacy web page at <http://www.apdu.org/advocacy.asp> links to a frequently-updated spreadsheet which lists comment opportunities grouped by the agency that collects the data. The site also includes an explanation of the public comment process.

APDU members receive regular e-mail updates of public-comment opportunities—but anyone can come and check the spreadsheet at the APDU web site.

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>.

Data.gov

Launched in May 2009 by the U.S. Office of Management & Budget, Data.gov is an initiative of the Obama administration designed to “increase public access to high value, machine readable datasets generated by the Executive Branch of the Federal Government.” The interface features three searchable catalogs: a “raw data” catalog (currently over 700 entries), a tool catalog (currently over 300 entries), and a geodata catalog (by far the largest catalog, at over 110,000 data sets).

Searches can be done by category, agency, or keyword. The project is also working with state and local agencies to launch similar initiatives, providing links to such catalogs via a clickable map. Users may comment on or give a 1-to-5-star rating on entries in the Data.gov catalog, and may also make suggestions for additions to the site.

An alternate place for discussion and recommendations for Data.gov is a wiki hosted by Wired, at http://howto.wired.com/wiki/Open_Up_Government_Data

Adversity Index

MSNBC’s Adversity Index, also known as the Elkhart Project, is “a measure of the economic health of 381 metro areas and the 50 states.” The project takes the form of an interactive map which displays monthly snapshots since June of 1994, with values for employment, single-family housing starts, housing costs, and industrial production, each expressed in terms

of percentage change from the previous year. Those four numbers are then used to label the economy of each state or metro area as expansion, at risk, recovery, or in recession. Unfortunately the site’s claim that it seeks to provide “the hard numbers around these hard times” does not extend to allowing downloads of the actual data. However, the display and the accompanying news analysis are useful for identifying trends.

The Adversity Index is online at http://www.msnbc.msn.com/id/29896874/ns/us_news-the_elkhart_project.

Madison Neighborhood Indicators

The Madison Neighborhood Indicators Project program, funded by the City of Madison and hosted at <http://madison.apl.wisc.edu/> by UW-Madison’s Applied Population Lab, offers a carefully-chosen selection of indicators that relate to the quality of life in Madison at the neighborhood level.

Currently the site carries a single year of selected data indicators plus mapping capability, covering the city of Madison as a whole and 70 neighborhoods, also organized as 57 planning districts. Indicators for each neighborhood include a basic area and population profile, public safety, health and well-being, community action & involvement, economic vitality, and housing. Mapping and neighborhood-comparison tools are available on the site. Note that some indicators, particularly relating to health and family well-being, are suppressed at the neighborhood level due to privacy concerns.

The project launched as a pilot in 2008 with 5 neighborhoods, and went city-wide in October 2009.