Appendix C

Secondary Data Sources

Many quantitative studies and some qualitative investigations use data available from previous research or government agencies. In the United States, the U.S. Census Bureau and many other government agencies make data available for general use. Data are also available for research purposes from many other countries as well as from world bodies like the United Nations and the World Bank. Academic researchers and students can draw on a very large data archive at the Inter-University Consortium for Political and Social Research (ICPSR). Qualitative researchers can also use information in published histories or other secondary sources, such as documents found in archival collections. This appendix identifies important sources of such *secondary data.*

U.S. Census Bureau

The U.S. government has conducted a census of the population every 10 years since 1790; since 1940, this census also has included a census of housing (see also Chapter 4). This decennial Census of Population and Housing is a rich source of social science data (Lavin 1994). The Census Bureau's monthly *Current Population Survey (CPS)* provides basic data on labor force activity that is then used in Bureau of Labor Statistics reports. The Census Bureau also collects data on agriculture, manufacturers, construction and other business, foreign countries, and foreign trade.

The U.S. Census of Population and Housing aims to survey an adult in every household in the United States. The basic "complete-count" census contains questions about household composition as well as ethnicity and income. More questions are asked in a longer form of the census that is administered to a sample of the households. A separate census of housing characteristics is conducted at the same time (Rives & Serow 1988:15). Participation in the census is required by law, and confidentiality of the information obtained is mandated by law for 72 years after collection. Census data are reported for geographic units, including states, metropolitan areas, counties, census tracts (small, relatively permanent areas within counties), and even blocks (Exhibit C.1). These different units allow units of analysis to be tailored to research questions.
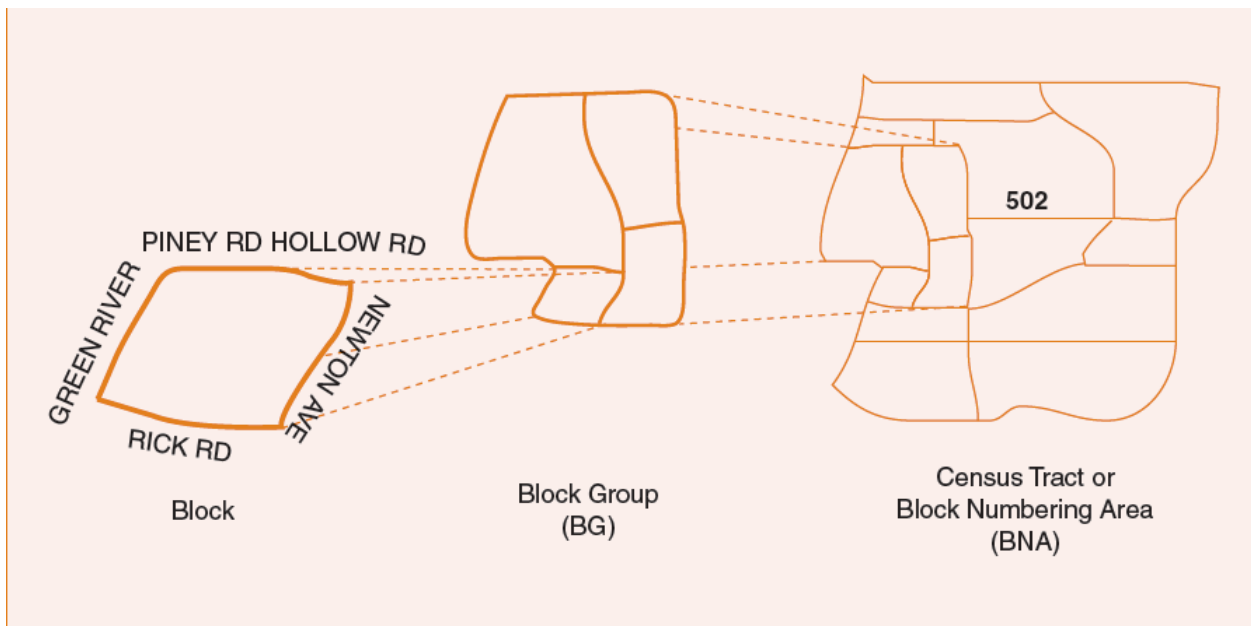
Exhibit C.1 Census Small-Area Geography

Census data are used to apportion seats in the U.S. House of Representatives and to determine federal and state legislative district boundaries, as well as to inform other decisions by government agencies. An interactive data retrieval system, American FactFinder, is the primary

means for distributing results from the 2000 Census: You can review it at

http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml

The catalog of the ICPSR (www.icpsr.umich.edu) also lists many census reports. Many

census files containing microdata—records from persons, households, or housing units—are

available online, while others can be purchased on CD-ROM or DVD from the Customer

Services Center at (301) 763-INFO (4636); census data can also be inspected online or

downloaded for various geographic levels, including counties, cities, census tracts, and even

blocks using the DataFerrett application (Federated Electronic Research, Review, Extract, and

Tabulation Tool). You can download, install, and use this tool at http://dataferrett.census.gov.

This tool also provides access to data sets collected by other federal agencies.

States also maintain census bureaus and may have additional resources. Some contain the

original census data collected in the state 100 or more years ago. The ambitious historical

researcher can use these returns to conduct detailed comparative studies at the county or state

level (Lathrop 1968:79).

Integrated Public Use Microdata Series

Individual-level samples from U.S. Census data for the years 1850 to 1990, as well as

historical census files from several other countries, are available through the Integrated Public

Use Microdata Series (IPUMS) at the University of Minnesota's Minnesota Population Center

(MPC). These data are prepared in an easy-to-use format that provides consistent codes and

names for all the different samples.

This exceptional resource offers 25 samples of the American population selected from 13

federal censuses, with at least 100,000 persons in each sample; in recent years the samples

contained more than 1 million persons. Each sample is independently selected, so that

individuals are not linked between samples. In addition to basic demographic measures, variables

in the U.S. samples include educational, occupational, and work indicators; respondent income;

disability status; immigration status; veteran status; and various household characteristics,

including family composition and dwelling characteristics. The international samples include

detailed characteristics from hundreds of thousands of individuals in countries ranging from

France and Mexico to Kenya and Vietnam. You can view these resources at

www.ipums.umn.edu.

Many other government agencies provide data for social scientists.

Bureau of Labor Statistics (BLS)

Another good source of data is the Bureau of Labor Statistics of the U.S. Department of Labor,

which collects and analyzes data on employment, earnings, prices, living conditions, industrial

relations, productivity and technology, and occupational safety and health (U.S. Bureau of Labor

Statistics 1991, 1997b). Some of these data are collected by the Census Bureau in the monthly

Current Population Survey (CPS); other data are collected through surveys of establishments

(U.S. Bureau of Labor Statistics 1997a).

The *CPS* provides a monthly employment and unemployment record for the United

States, classified by age, sex, race, and other characteristics. The *CPS* uses a stratified random

sample of about 60,000 households (with separate forms for about 120,000 individuals). Detailed

questions are included to determine the precise labor force status (whether they are currently

working or not) of each household member over the age of 16. Statistical reports are published

each month in the BLS's *Monthly Labor Review* and can also be inspected at its website

(http://stats.bls.gov). Datasets are available on computer tapes and disks from the BLS and

services like the Inter-University Consortium for Political and Social Research (ICPSR).

Other U.S. Government Sources

Many more datasets useful for historical and comparative research have been collected by

federal agencies and other organizations. The National Technical Information Service (NTIS) of

the U.S. Department of Commerce maintains a Federal Computer Products Center that collects

and catalogs many of these datasets and related software.

By 2008, more than 2,000,000 datasets and reports were described in the NTIS *Directory*.

The *Directory* is the essential source of information about the datasets and can be purchased

from the U.S. Department of Commerce (National Technical Information Service, 1993). Dataset

summaries can be searched in the *Directory* by either subject or agency. Government research

reports cataloged by NTIS can be searched online at the NTIS website (www.fedworld.gov) and

in a CD-ROM catalog available in some libraries.

International Data Sources

Comparative researchers can find datasets on population characteristics, economic and political

features, and political events in many nations. Some of these are available from U.S. government

agencies. For example, the Social Security Administration reports on the characteristics of social

security throughout the world (Wheeler 1995). This comprehensive report classifies nations in

terms of their type of social security program and provides detailed summaries of the

characteristics of each nation's programs. Recent data are available on the Internet at

www.socialsecurity.gov/policy/docs/progdesc/ssptw/. More recent data are organized by region.

A broader range of data is available in the *World Handbook of Political and Social Indicators,*

with political events and political, economic, and social data coded from 1948 to 1982

(www.icpsr.umich.edu, study no. 7761) (Taylor & Jodice 1986).

The European Commission administers the Eurobarometer Survey Series at least twice

yearly across all the member states of the European Union. The survey monitors social and

political attitudes and reports are published regularly online:

http://ec.europa.eu/public_opinion/index_en.htmCase level Eurobarometer survey data are stored

at the ICPSR.

ICPSR

The University of Michigan's Inter-University Consortium for Political and Social Research

(ICPSR) is the premier source of secondary data useful to social science researchers. ICPSR was

founded in 1962 and now includes more than 640 colleges and universities in North America and

hundreds of institutions on other continents. ICPSR archives the most extensive collection of

social science datasets in the United States outside of the federal government: More than 7,243

studies are represented in 450,000 files from 130 countries and from sources that range from

U.S. government agencies such as the Census Bureau to international organizations like the

United Nations, social research organizations like the National Opinion Research Organization,

and individual social scientists who have completed funded research projects.

The datasets archived by ICPSR are available for downloading directly from the ICPSR

website, www.icpsr.umich.edu/ICPSR/help/datausers. ICPSR makes datasets obtained from

government sources available directly to the general public, but many other datasets are available

only to individuals at the more than 500 colleges and universities around the world that have paid the fees required to join ICPSR. The availability of some datasets is restricted due to confidentiality issues; in order to use them, researchers must sign a contract and agree to certain conditions (www.icpsr.umich.edu/).

Survey datasets obtained in the United States and in many other countries that are stored at the ICPSR provide data on topics ranging from elite attitudes to consumer expectations. For example, data collected in the British Social Attitudes Survey in 1998, designed by the University of Chicago's National Opinion Research Center, are available through the ICPSR (go to the ICPSR website, www.icpsr.umich.edu, and search for study no. 3101). Data collected in a monthly survey of Spaniards' attitudes, by the Center for Research on Social Reality [Spain] Survey, are also available (study no. 17). Survey data from Russia, Germany, and other countries can also be found in the ICPSR collection.

Do you have an interest in events and interactions between nations, such as threats of military force? A dataset collected by Charles McClelland includes characteristics of 91,240 such events (study no. 5211). The history of military interventions in nations around the world between 1946 and 1988 is coded in a dataset developed by Frederic Pearson and Robert Baumann (study no. 6035).

This dataset identifies the intervener and target countries, the starting and ending dates of military intervention, and a range of potential motives (such as foreign policies, related domestic disputes, and pursuit of rebels across borders).

Census data from other nations are also available through the ICPSR, as well as directly through the Internet. In the ICPSR archives, you can find a dataset from the Statistical Office of the United Nations on the 1966 to 1974 population of 220 nations throughout the world (study

no. 7623). More current international population data are provided by the Center for

International Research and the U.S. Census Bureau (study no. 8490). See also the preceding

description of the Eurobarometer Survey Series.

Obtaining Data From ICPSR

You begin a search for data in the ICPSR archives at

www.icpsr.umich.edu/access/index.html.

Exhibit C.2 shows the search screen as it would appear when you might start a search for

data from studies involving the subject domestic violence. You can also see in this screen that

you can search the data archives for specific studies, identified by study number or title, as well

as for studies by specific investigators (this would be a quick way to find the dataset contributed

by Lawrence Sherman from his research, discussed in Chapter 2, on the police response to

domestic violence).
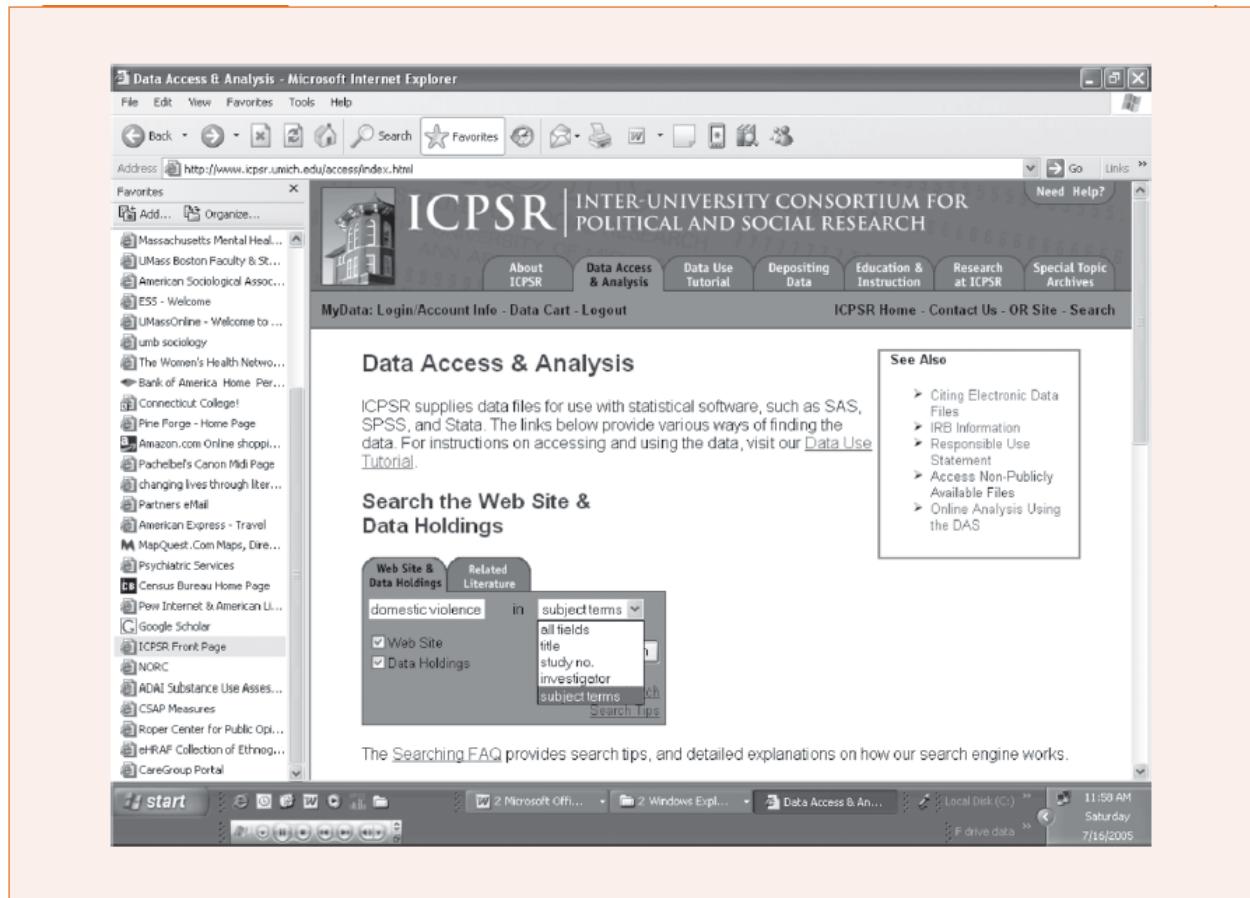
**Exhibit C.2**   Search Screen: Domestic Violence

Exhibit C.3 displays the results of my search: a list of 63 datasets that involved research on domestic violence and that are available through ICPSR. For most datasets, you can obtain a description, the files that are available for downloading, and a list of "related literature"—that is, reports and articles that use the listed dataset. Some datasets are made available in collections on a CD-ROM; the CD-ROM's contents are described in detail on the ICPSR site, but you have to place an order to receive the CD-ROM itself.
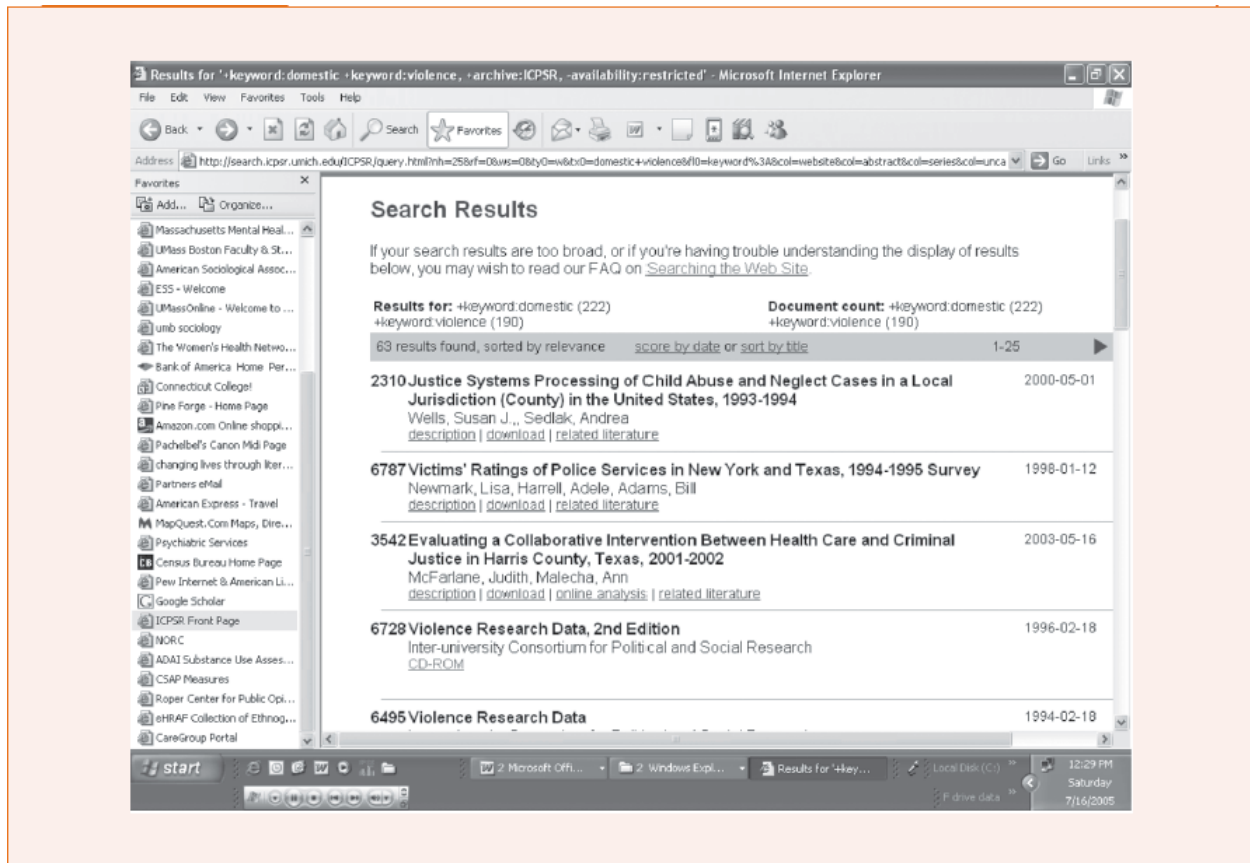
Exhibit C.3   Search Screen: Domestic Violence Results

When you click on the "Download" option, you are first asked to enter your e-mail

address and password. What you enter will determine which datasets you can access; if you are

not at an ICPSR member institution, you will be able to download only a limited portion of the

datasets—mostly those from government sources. If you are a student at a member institution,

you will be able to download most of the datasets directly, although you may have to be using a

computer that is physically on your campus to do so. Exhibit C.4 displays the ICPSR download

screen after selected files were downloaded from the study by Lisa Newmark, Adele Harrell, and

Bill Adams of victim ratings of police response in New York and Texas. To analyze the data

with the SPSS statistical package the dataset was downloaded in the form of an "SPSS Portable

File." The files downloaded in a zip file, so the WinZip © program was used to unzip them.

After unzipping the SPSS portable file, data analysis with the SPSS program could begin. If

you'd like to learn how to analyze data with the SPSS statistical program, review Appendix D.
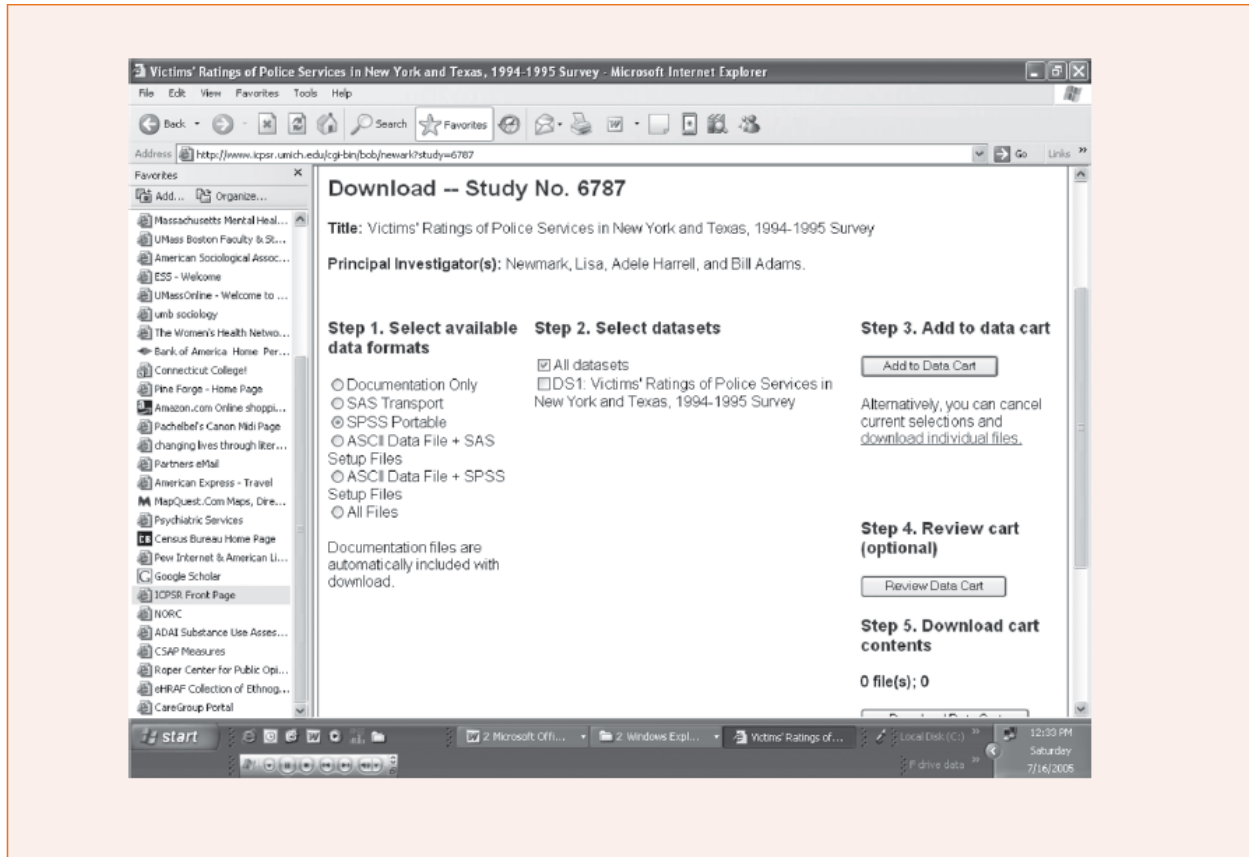
Exhibit C.4   ICPSR Download Screen

If you prepare your own paper based on an analysis of ICPSR data, be sure to include a

proper citation. Here's an example from the ICPSR itself

(www.icpsr.umich.edu/org/citation.html):

Reif, Karlheinz, and Anna Melich. *Euro-Barometer 39.0: European Community*

Policies and Family Life, March-April 1993 [Computer file]. Conducted by

INRA (Europe), Brussels. ICPSR06195-v4. Ann Arbor, MI: Inter-University

Consortium for Political and Social Research [producer], 1995. Koeln,

Germany: Zentralarchiv fuer Empirische Sozialforschung/Ann Arbor, MI: Inter-

University Consortium for Political and Social Research [distributors], 1997.

Some of the datasets are also offered with the option of "online analysis." If you have this option, you can immediately inspect the distributions of responses to each question in a survey and examine the relation between variables, without having any special statistical programs of your own. At the bottom of Exhibit C.5, you'll find the wording reported in the study "codebook" for a question used in the study of a collaborative health care and criminal justice intervention in Texas, as well as, in the top portion, the available statistical options. After choosing one or more variables from the codebook, you can request the analysis.
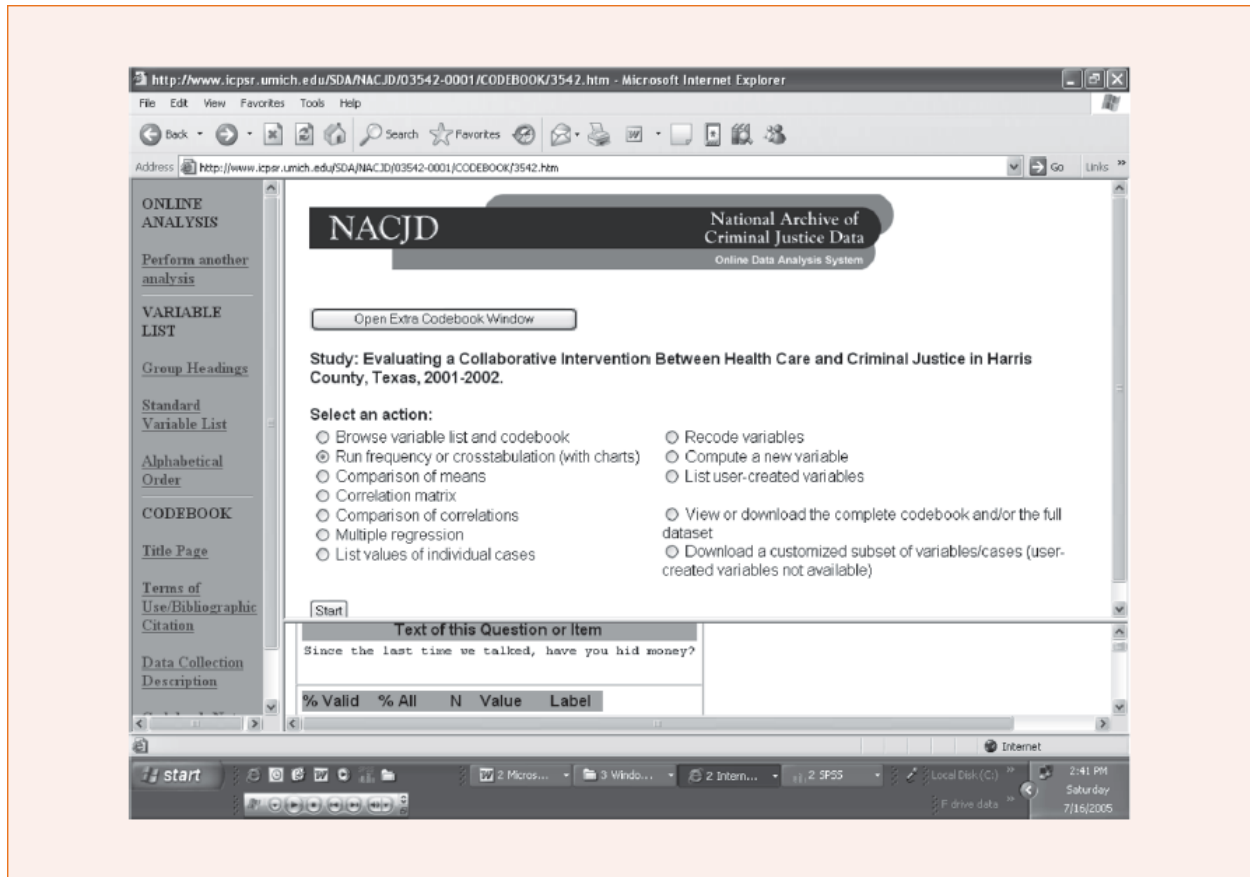
Exhibit C.5 ICSPR Online Analysis: Codebook Information and Statistical Options

The analysis begins with a chart of the distribution of victims' responses to a question about their current relationship with the abuser. As you can see in Exhibit C.6, about half had left the relationship, but half were still married or living as married with the abuser. This approach to analysis with secondary data can get you jump-started in your work. An online analysis option is also starting to appear at other websites that offer secondary data.
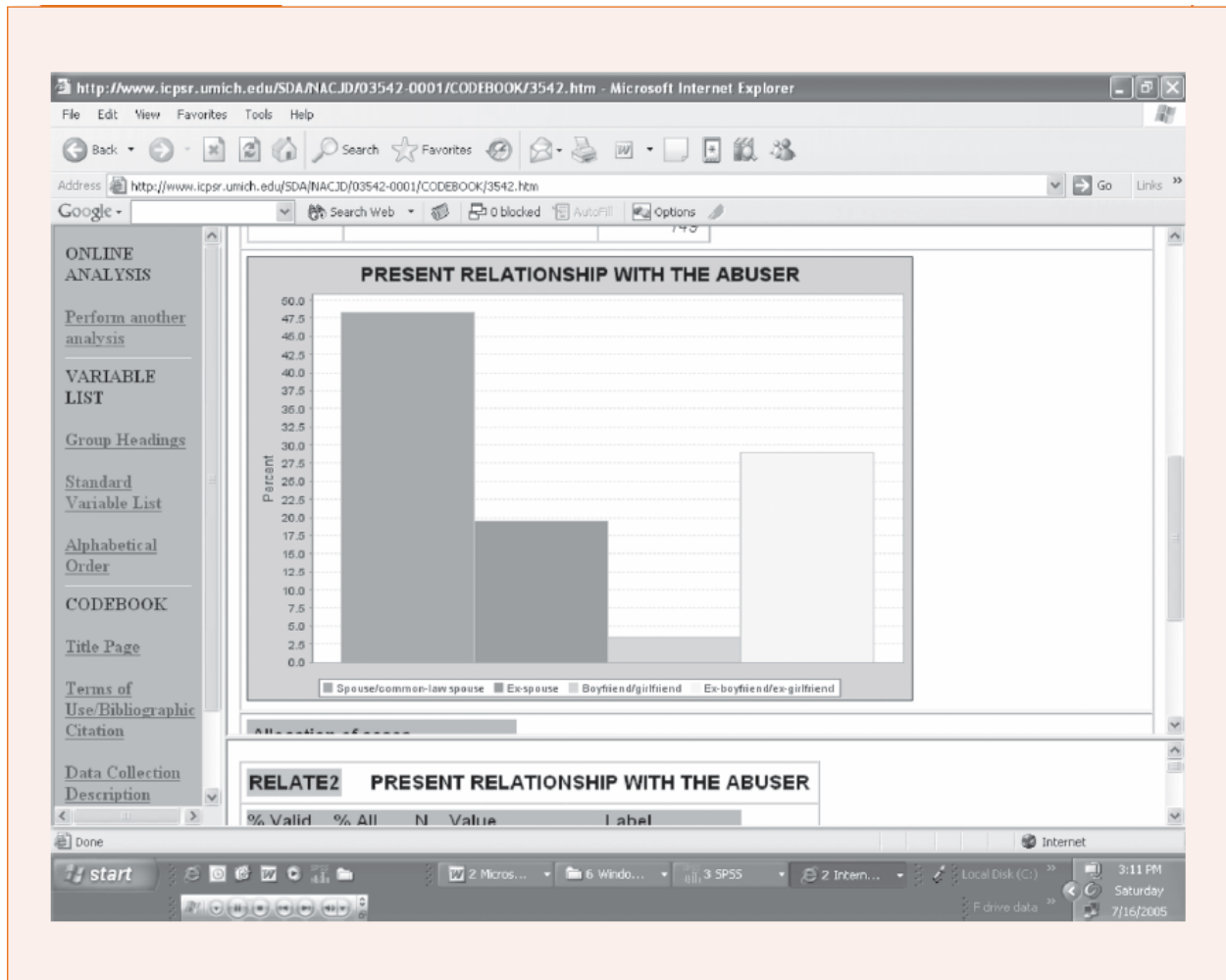
Exhibit C.6   ICPSR Online Analysis Bar Chart

ICPSR also catalogs reports and publications containing analyses that have used ICPSR

datasets since 1962—more than 7,990 studies are represented in more than 500,000 files from

130 countries and from sources that range from U.S. government agencies such as the Census

Bureau to international organizations such as the United Nations, social research organizations

such as the National Opinion Research Center, and individual social scientists who have

completed funded research projects. . This superb resource provides an excellent starting point

for the literature search that should precede a secondary data analysis. In most cases, you can

learn from detailed study reports a great deal about the study methodology, including the rate of response in a sample survey and the reliability of any indexes constructed. Published articles provide examples of how others have described the study methodology, let you know what research questions have already been studied with the dataset, and outline what issues remain to be resolved. You can search this literature at the ICPSR site simply by entering the same search terms that you used to find datasets, or else by entering the specific study number of the dataset on which you have focused (Exhibit C.7). Don't start a secondary analysis without reviewing such reports and publications.
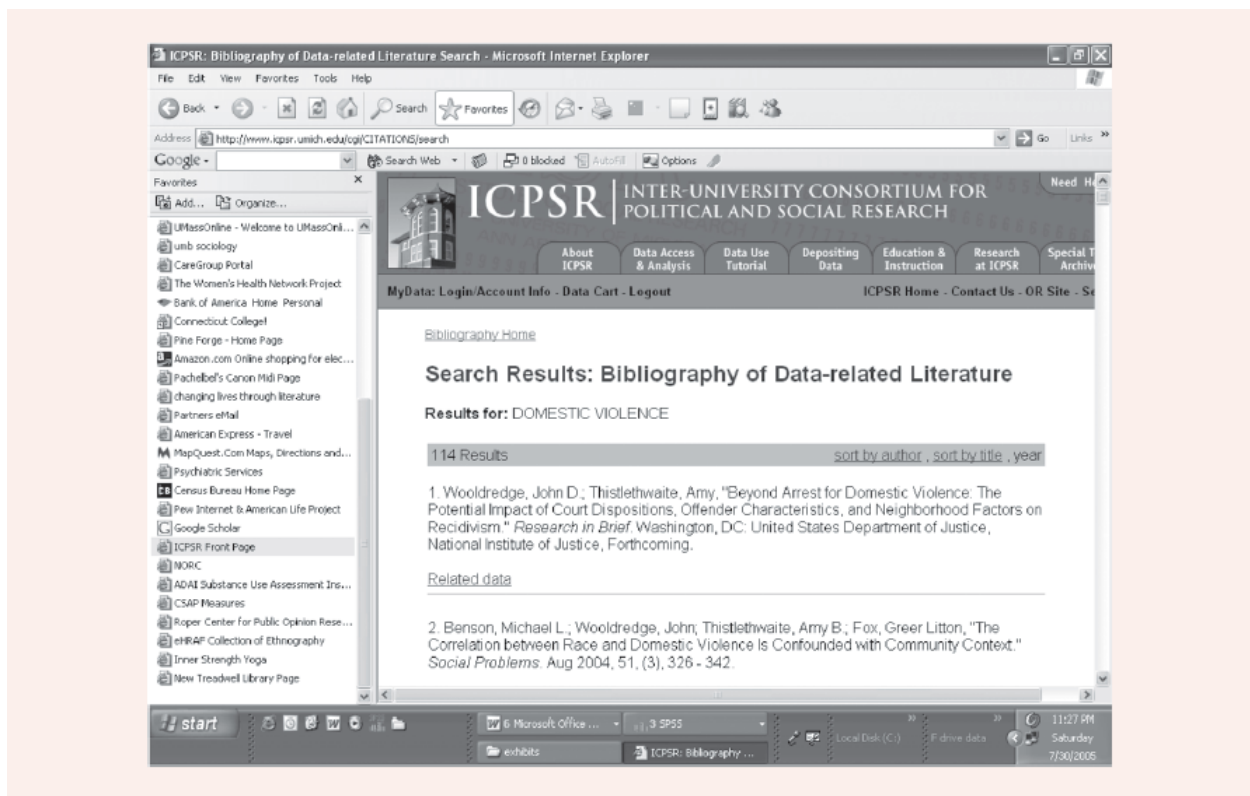

Exhibit C.7   ICPSR's Search of "Related Literature" on Domestic Violence

Qualitative Data Sources

Far fewer qualitative datasets are available for secondary analysis. By far the richest source, if you are interested in cross-cultural research, is the Human Relations Area Files (HRAF) at Yale University. The HRAF has made anthropological field data available for international cross cultural research since 1949 and currently contains over 800,000 pages of information on more than 365 different groups (HRAF 2005). The Human Relations Area Files Collection of Ethnography provides an extraordinary resource for qualitative comparative cross-sectional research) (Ember & Ember 2011). The HRAF was founded in 1949 as a corporation designed to facilitate cross-cultural research. The HRAF ethnography collection now contains over 1,000,000 pages of information on more than 400 different cultural, ethnic, religious, and national groups. The information is indexed by topic, in 710 categories, and now made available electronically (if your school pays to maintain access to the HRAF). Exhibit C.8 is an example of a page from an HRAF document that has been indexed for easy retrieval.
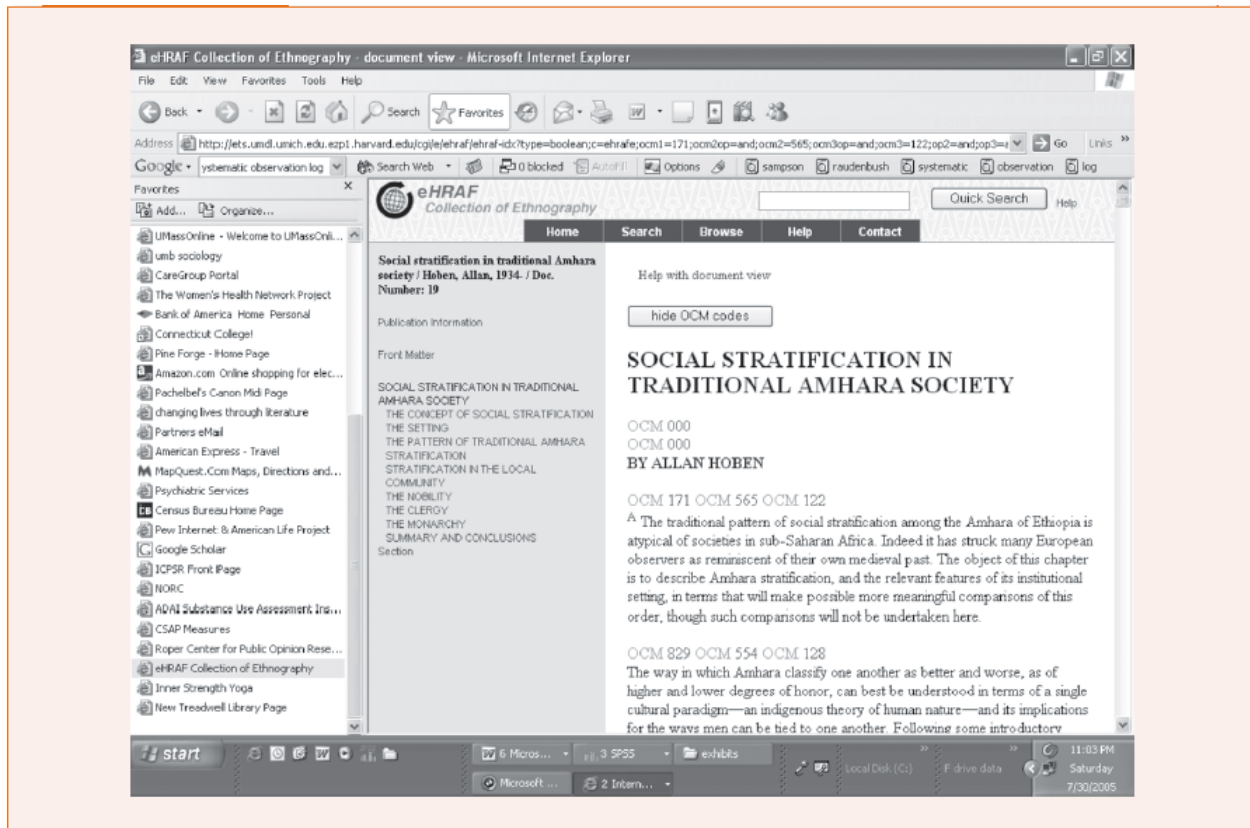
Exhibit C.8   HRAF Indexed Document

The ICPSR collection includes a limited number of studies containing at least some qualitative data or measures coded from qualitative data (494 such studies as of May 2011), but these include some very rich data. Studies range from transcriptions of original handwritten and published materials relating to infant and child care, from the turn of the century to World War II (LaRossa 1995) to Daniel Lockwood's (1996) transcripts of open-ended interviews with high school students involved in violent incidents.

Several other university-based centers have developed qualitative archive projects, although access is often limited. The Murray Research Center at Harvard's Radcliffe Institute for Advanced Study (http://murraydata.hmdc.harard.eduATX/) focuses on studies of lives over time, with special interest in issues of concern to women. Although the Murray Center's collection of

about 300 studies contains both quantitative and qualitative datasets, it has had a special interest in qualitative data, including videotape and audiotape collections, case study data, and transcripts from intensive interview studies and data from surveys that included many open-ended questions. The Murray Research Center has now been merged with the larger Harvard-MIT Data Center, which makes available a much larger set of datasets from quantitative studies, although only to members of Harvard-Radcliffe. Access to archived datasets requires submission of an online application.

Holdings of the Economic and Social Data Service of the Universities of Sussex and Manchester in England (www.esds.ac.uk/qualitdata/online/), ESDS Qualidata, include interview transcripts and other materials from several qualitative studies, including Paul Thompson's "100 Families: Families, Social Mobility and Aging, an Intergenerational Approach." Subsets of the interviews can be browsed or searched directly online, but access is restricted to those at member institutions. Although a great many universities in Britain and throughout the world have joined the ESDS, they include very few colleges and universities in the United States.

The University of Southern Maine's Center for the Study of Lives (www.usm.maine.edu/cehd/csl) collects interview transcripts that record the life stories of people of diverse ages and backgrounds. As of March 2011, their collection included transcripts from over 400 life stories, representing more than 35 different ethnic groups, experiences of historical events ranging from the Great Depression to the Vietnam War, and including reports on dealing with health problems such as HIV/AIDS. These qualitative data are available directly online without any registration or fee.