Dataset tools for SOCI 5200
About Datasets and the Library

• There are a few dataset tools and since you are searching studies based on your topic, keep in mind there may not be a study to match (or perfectly match your study) – discuss topics with professor OR search early before declaring a topic. Select studies can be “analyzed online” with SDA, etc.

• Please consult RSS (Research & Statistical Support) for SPSS assistance. For a list of general access labs with statistical software, consult the GACL Software List. The library/library staff cannot assist in analyzing your data. They can assist in searching for articles, books, etc. Your professor/TA is the most valuable here.
About SDA:
“SDA is a set of programs for the documentation and Web-based analysis of survey data. SDA is developed and maintained by the Computer-assisted Survey Methods Program (CSM) at the University of California, Berkeley. CSM also develops the CASES software package.

To see how it all works, test-drive SDA at our demonstration SDA Archive. Browse the documentation for a survey and get fast data analysis results. The SDA Archive includes several datasets, including the General Social Survey (GSS) and the American National Election Study (ANES). You can also look at some other archives that use SDA software.”

More @: http://sda.berkeley.edu/
Berkley Archive – includes GSS

- Berkley: Uses/Developmed SDA tool
- Archive (Not all PUMS, focus on California)
- http://sda.berkeley.edu/archive.htm

- Includes General Social Survey (GSS) Cumulative Datafile 1972-2008
Example Search (depression) then View (MNTLHLTH)

1. Search “depression” and see MNTLHLTH.
2. Click “View” button and See pop up of results
Adding Variables to Rows (Row), Columns (Col) or as the Control (Ctrl)
Recoding Variables: How to do basic coding

http://sda.berkeley.edu/HELPDOCS/helpan.htm#recode

The "Recoding Variables" link will give you this information.

A numeric variable can be transformed temporarily, for purposes of running the current analysis. There are five types of temporary transformations.

- Recode a variable
- Collapse a variable into fewer categories
- Create a dummy variable (a dichotomy coded 0 or 1)
- Create multiple dummy variables from a single variable (for regression)
- Create a product variable (for correlation and regression)

Temporarily Recode a Variable

Temporary recodes are created by specifying groups of codes that are to be combined into a single category. This type of transformation can be very simple, but certain options can make it a little more complex. These are the possibilities:

- Basic recoding
  - Assigning particular new code values
  - Assigning labels to the new code values
  - Open ranges (with " ", or " **")
  - Overlapping ranges
  - Multiple specifications for one recoded group
  - Treatment of missing data

Basic recoding

For example, to combine the categories of 'age' into three groups, you can specify the variable as:

age = (r: 18-30; 31-50; 51-96)

Notice that the name of the variable ('age') is followed by parentheses, then the instruction 'r' (or 'R') followed by a colon (':'), and then the groupings of codes. Those groupings can consist of single code values, ranges, or a combination of many values and/or ranges. Each group is separated from the other by a semicolon (';'). Spaces are optional, but are added here for readability.

Using this basic method of recoding, the new groupings of codes are given the default code values 1, 2, 3, and so forth. The default label for each group is the range of original codes that constitute that group ('18-30', for example).

Any categories of 'age' not included in the specified groupings will become missing-data on the recoded version, and they will be excluded from the analysis in the table.

On the other hand, any original missing-data categories of 'age' that are explicitly mentioned in the recode, will be included. For instance, if the value '90' for 'age' were flagged as a missing-data code, but included as in the example above, it would become part of the third recoded category. This is discussed in more detail in the section on "Treatment of missing data."

Assigning particular new code values

It is possible to assign new code values that are different from the default 1, 2, 3, and so forth. To do this, give the new code value, then an equal sign, then the grouping. (The new code value must be a whole number, and decimal places will be ignored. If you want the new code value to include decimal places, use the regular SDA RECODE program.)

For example, the variable 'age' can be recoded into the same three groups as above, but with the new code values 1, 5, and 10, by specifying the recode as follows:

age = (r: 18-30=1; 31-50=5; 51-96=10)
Example of Recoded chart with Control

Coded variables (optional)
Run the table >
ICPSR

Interuniversity Consortium for Political and Social Research (ICPSR)
Searching ICPSR

Note: Not all your topics will have datasets – be flexible or search sources first.

• **Search**: Searching ICPSR can be accomplished in several different ways.
• By entering a search term in the textbox and clicking "Search for Data".
• By **topic**
• By **geographic location**
• By **investigator** (the institution(s) conducting the study).
• Browse by **series** of studies.
• Browse the most recent updates and editions to a study.
See dataset tab of your class page for ICPSR notes. Databases & More Help page has my ICPSR Tutorial. Note: Interface just changed. Browse ICPSR on YouTube (example video:). Note: the library does not have software for SPSS. Please consult RSS (Research & Statistical Support) for SPSS assistance.
IPUMS USA (http://usa.ipums.org/usa/)

Not All Studies can be Analyzed Online

"Examples and screenshots are available in our short instructions page."
http://usa.ipums.org/usa/resources/sda/sdainstructions.pdf

Quick Links: IPUMS USA Short Instructions PDF