

Extracting and Downloading the General Social Survey (GSS), 1972-2004 from the SSCC

```
[your_netid@seldon your_netid]$ cd /datalib/4000s/4295  
[your_netid@seldon 4295]$ ll
```

```
total 95300  
-rwxrwxr-x 1 datalib datalib 30721778 Aug 31 13:06 4295_codebook.pdf.gz  
-rwxrwxr-x 1 datalib datalib 11320258 Aug 31 13:06 4295_codebook_appendix.pdf.gz  
-rwxrwxr-x 1 datalib datalib 17148970 Aug 31 13:07 4295_data.txt.gz  
-rwxrwxr-x 1 datalib datalib 190762 Aug 31 13:07 4295_errata.pdf.gz  
-rwxrwxr-x 1 datalib datalib 186939 Aug 31 13:07 4295_spss_setup.sps.gz  
-rwxrwxr-x 1 datalib datalib 2632 Oct 5 14:16 readme.txt  
-rwxrwxr-x 1 datalib datalib 37872194 Aug 31 13:07 sp4295.por.gz
```

```
[your_netid@seldon 4295]$ cd
```

* This returns you to your home directory.

```
[your_netid@seldon your_netid]$ gunzip -c /datalib/4000s/4295/sp4295.por.gz > sp4295.por
```

*This unzips the cumulative SPSS portable file and saves a copy on your directory

```
[your_netid@seldon your_netid]$ dos2unix sp4295.por
```

* This convert the SPSS portable file from DOS to UNIX/LINUX format.

```
[your_netid@seldon your_netid]$ nano
```

* This opens an editing program, which allows you to create a file of SAS commands. You should receive a blank screen, with a single bar of commands across the top. Type the following commands and save the file as "gssdata.com":

```
libname gss "~your_netid";  
filename spssgss "sp4295.por";  
  
proc convert spss=spssgss out=new;  
run;  
  
data gss.gss_subset;  
set new;  
if year in (1972 1976);  
keep year age class;  
run;  
  
proc freq data=gss.gss_subset;  
tables age class;  
run;
```

* These commands convert the SPSS file into a SAS dataset, keeping all variables listed after the "keep" command. List all the variables that you need.

* Because the GSS is a cumulative file, always retrieve YEAR, so that you can limit your study to the time period that you are interested in, or eliminate years in which your survey questions were not asked. Use the "if" command to select the years that you want.

* The "proc freq" command is an optional command that will help you to check your data to make sure there are no errors. It will record the frequencies for the variables listed.

* The gss subset named "gss_subset" is saved as a new, smaller file on your home directory. This is an SAS UNIX version file and can be converted into SAS transport file or even a dataset in other format like STATA by StatTransfer (st) on SSCC's seldon. Please note that StatTransfer is available only on seldon (seldon.it.northwestern.edu).

* Exit the editor program by typing <Ctrl> x. The editor will ask you to name your command file. Enter "gssdata.com" Now, run the command file you have generated with the following command. This step will also take a few minutes, probably longer, depending on how many variables you request, and amount of use on SSCC:

```
[your_netid@seldon your_netid]$ sas gssdata.com
[your_netid@seldon your_netid]$ more gssdata.com.lst
```

*The "more" command will allow you to view the SAS output (list) file. Scroll down and view the frequencies of the variables you requested. Make sure that the data appears correct. One way to do this is to compare the total N for each year with the figures listed in the GSS codebook.

```
[your_netid@seldon your_netid]$ st gss_subset.sas7bdat gss_subset.dta
```

* This command converts the output SAS subset into a STATA dataset.

* Next, check to make sure the file is there, and logout:

```
[your_netid@seldon your_netid]$ ll gss*
-rw-----  1 your_netid  users      13050 May  9 14:40 gss_subset.dta
-rw-----  1 your_netid  users     90112 May  9 14:38 gss_subset.sas7bdat
-rw-----  1 your_netid  users        224 May  9 14:35 gssdata.com
-rw-----  1 your_netid  users      2766 May  9 14:38 gssdata.com.log
-rw-----  1 your_netid  users      8903 May  9 14:38 gssdata.com.lst

[your_netid@seldon your_netid]$ logout
```

Retrieving Data from your the SSCC account

To retrieve your file for use on STATA/Windows, use the FTP software (SSH, Absolute FTP, etc.) at the library lab or on your own PC. Select "seldon.it.northwestern.edu" or "hardin2.it.northwestern.edu" as the profile name, and enter your user name and password. When the connection is made, you should see the files listed in your local PC directory on the left, and the files in your SSCC home directory on the right. Highlight the "gss_subset.dta" file in your home directory and click the arrow to direct the file into the local PC directory. Close the connection.