NAME:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**STATA Example Codebook**

The General Social Survey (GSS) conducts face-to-face interviews with about 2,000 Americans every other year. Sample members are selected at random from the population of Americans age 18 and older who live in households (i.e., not jails or college dorms or hospitals). The data file “STATA Example Datafile.dat” includes five variables measured for 1,974 GSS respondents who were interviewed in 2012. You can learn more about the GSS at [http://www3.norc.org/Gss+website/](http://www3.norc.org/Gss%2Bwebsite/)).

Below I include information about each of the five variables, including (1) the name of the variable; (2) the variable’s column locations in the data file; (3) a brief description of the variable; and (4) the label associated with each numeric value.

The example STATA syntax file on the course web site (“STATA Example Syntax.do”) reads the data file into STATA; labels the variables; labels the values of each variable; declares certain values to be missing; produces frequency and percentage distributions (for the whole sample and separately by sub-group); produces histograms depicting distributions of continuous variables; and produces statistics summarizing distributions of continuous variables (for the whole sample and separately by subgroup). The syntax shows how to save an output file (which records your results) and how to save graphs as pictures (so they might be inserted into homework assignments, for example).

*Begin with this syntax for future assignments. Modify it as necessary.*

**DESCRIPTION OF VARIABLES IN “STATA EXAMPLE DATAFILE.DAT”**

**AGE** (Columns 1-2)

*Respondent's age in years*

 18=18 Years old

 ...

 89=89 Years old

 -9 = Value is missing

**EDUCATION** (Columns 3-4)

*Respondent's years of schooling completed*

 1=1 Year

 2=2 Years

 ...

 20=20 Years

 -9 = Value is missing

**HEALTH** (Columns 5-6)

*Respondent's self-reported overall health*

 1=Excellent

 2=Good

 3=Fair

 4=Poor

 -9 = Value is missing

**PARTYID** (Columns 7-8)

*Respondent's political party identification*

 0=Strong Democrat

 1=Not Strong Democrat

 2=Independent, Near Democrat

 3=Independent

 4=Independent, Near Republican

 5=Not Strong Republican

 6=Strong Republican

 7=Other Party

 -9 = Value is missing

**SEX** (Columns 9-10)

*Respondent's sex*

 1=Male

 2=Female

 -9 = Value is missing