Community Learning Resource
Preserving, Reshaping and Transferring Data in Stata
Andrew P. Davis
3/2/2016

This blog-post will focus on preserving, reshaping and transferring data in Stata – skills that are quite common and make data manipulation more efficient and less prone to human error.

I will begin with reviewing the logic behind reshaping data and the Stata commands to achieve this. I will then move to provide information on transferring data using StatTransfer. Finally, as with my previous blogpost http://apdsociology.blogspot.com/2016/02/regular-expressions-in-stata.html I will conclude with several exercises to help you get the hang of preserving and reshaping data.

Preserve and Restore

In Stata, a "preserve" command saves a pre-manipulated version of the data that will remain unadulterated after you conclude your programming.
-For instance if you write a do-file program that includes data transformations, you could use a "preserve" command that would return you data to its original version post-manipulation.
-Intuitively, "restore" commands call back the previously "preserved" data

These commands are called as follows
Syntax:
preserve
(syntax that changes the data in some way)
restore

And are essential to reshaping datasets.

Reshaping Datasets

Data exist in "long" or "wide" formats. An example of a "wide" dataset would be the following matrix:

	countryid	gdp90	gdp91	gdp92
1.	3	95000	96000	77000
2.	1	60000	50500	51000
3.	2	75000	45400	65800

These data are considered to be "wide" as each variable-year is given its own column. It actually is "wide."

When we transfer these data to "long" or "narrow" format they look like this:

***Continued on following page

	countryid	year	gdp
1.	1	90	60000
2.	1	91	50500
3.	1	92	51000
4.	2	90	75000
5.	2	91	45400
6.	2	92	65800
7.	3	90	95000
8.	3	91	96000
9.	3	92	77000

So that there is variable-year, long format data that is produced in which the collated measures from the wide format are broken down by year.
-Year is simply a convenient example, you could still theoretically transform wide data into long data and vice-versa if the data are not listed by year

*Okay. But how do you do *reshapin*g in Stata?

Examples using reshape wide and reshape long:
**These examples will use fabricated data*

Syntax:
List

```
. list
```

	Country	Year	GDP
1.	Canada	2010	119
2.	Canada	2011	110
3.	Canada	2012	123
4.	Jamaica	2010	97
5.	Jamaica	2011	90
6.	Jamaica	2012	102
7.	Mexico	2010	78
8.	Mexico	2011	65
9.	Mexico	2012	45
10.	USA	2010	170
11.	USA	2011	165
12.	USA	2012	140

*As you can see, the data are already in "long" format, that is, data are in country-year format

Syntax:
Ex.: reshape long var1, i(id) j(year)
reshape wide GDP, i(Country) j (Year)

```
. reshape wide GDP, i(Country) j (Year)
(note: j = 2010 2011 2012)

Data                               long   ->   wide
-----------------------------------------------------------------------------
> ——
Number of obs.                       12   ->       4
Number of variables                   3   ->       4
j variable (3 values)              Year   ->   (dropped)
xij variables:
                                    GDP   ->   GDP2010 GDP2011 GDP2012
-----------------------------------------------------------------------------
> ——
```

list

```
. list
```

	Country	GDP2010	GDP2011	GDP2012
1.	Canada	119	110	123
2.	Jamaica	97	90	102
3.	Mexico	78	65	45
4.	USA	170	165	140

*As you can see, our reshape has worked, data are in “wide” format.
*But what if we wanted to move our data back into “long” format?

Syntax:

List

```
. list
```

	Country	GDP2010	GDP2011	GDP2012
1.	Canada	119	110	123
2.	Jamaica	97	90	102
3.	Mexico	78	65	45
4.	USA	170	165	140

Ex.: reshape long var1, i(id) j(year)
reshape long GDP, i(Country) j(Year)

```
. reshape long GDP, i(Country) j(Year)
(note: j = 2010 2011 2012)

Data                               wide   ->   long
-----------------------------------------------------------------------------
> ——
Number of obs.                        4   ->      12
Number of variables                   4   ->       3
j variable (3 values)                     ->   Year
xij variables:
                GDP2010 GDP2011 GDP2012   ->   GDP
-----------------------------------------------------------------------------
> ——
```

list

```
. list
```

	Country	Year	GDP
1.	Canada	2010	119
2.	Canada	2011	110
3.	Canada	2012	123
4.	Jamaica	2010	97
5.	Jamaica	2011	90
6.	Jamaica	2012	102
7.	Mexico	2010	78
8.	Mexico	2011	65
9.	Mexico	2012	45
10.	USA	2010	170
11.	USA	2011	165
12.	USA	2012	140

```
.
```

Transferring data:

Using StatTransfer products
*At times you will work with colleagues who work with data in a different format that you do (and different than you would like to work with)
*Notably, moving data between two of the more popular statistical packages used in the social sciences, SPSS and Stata is difficult if the data is in SPSS format or in Stata format

StatTransfer allows for the easy movement of data between formats commonly used in major statistical packages

StatTransfer version 13 will move data among the following programs

1-2-3
Microsoft Access (Versions 2.0 through Office XP version)
dBASE (all versions)
Delimited ASCII
Delimited ASCII with a Stat/Transfer SCHEMA file
Data Documentation Initiative (DDI) Schemas
Epi Info
EViewsNew
Excel worksheets (all versions, including Excel 2013)
Fixed format ASCII
FoxPro
GAUSS (Windows and Unix)
GenstatNew
gretl
JMP 10
LIMDEP
MATLAB
MATLAB Seven Datasets
Mineset
Minitab 14 (read only)
MPLUS (write only)
NLOGIT
ODBC data sources (Oracle, Sybase, Informix, etc.)
Open Document Spreadsheets
OSIRIS (read-only)
Paradox
Quattro Pro for DOS and Windows
R
RATS
S-PLUS (now supported through version 7)
SAS CPORT datasets and catalogs (read only)
SAS for Unix—HP, IBM, Sun
SAS for Unix—DEC Alpha
SAS for Windows and OS/2
SAS PC/DOS 6.04 (read only)
SAS Transport
SAS Value Labels
SAS version 7–9
SPSS through version 21
SPSS Datafiles (Windows and Unix)
SPSS Portable Files
Stata (all versions, including 14)
Statistica versions 7–8 (Windows only)
SYSTAT 13
Triple-S Survey Interchange Format

http://www.stata.com/products/stat-transfer/

Stat/Transfer has a very user-friendly interface, details on transferring data can be found here:

http://www.ats.ucla.edu/stat/st/default.htm

Exercises

Using the "City Temperature" dataset (sysuse citytemp) please complete the following tasks and provide evidence using commands that you have reshaped the data.

1.) "List" the data. Is it in long or wide format?
2.) Preserve the data
3.) Reshape the data from long to wide format
4.) Re-reshape the data from wide to long format

Open your own dataset, play around with this a bit but be sure to use the "preserve" command before reshaping data.

-Is your data in long or wide format? How do you know?
-Reshape the data in the direction you'd like

-What would be the benefit or reshaping the data?