Introduction to Geographic Information Science

Week 05 Stata Commands

Christopher G. Prener, Ph.D.

Spring, 2016

Contents

1 Renaming Variables

1.1 Description

The rename command is used to reassign variable names.

1.2 Syntax

```
rename oldvar_name newvar_name
```

1.3 Example

The following example uses the census.dta dataset that comes pre-installed with Stata. The following example renames the variable state2 to stateAbbrev:

```
. sysuse census.dta
(1980 Census data by state)

. rename state2 stateAbbrev

. describe stateAbbrev

              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
stateAbbrev     str2    %-2s                  Two-letter state abbreviation
```

1.4 Notes

Additional details for the rename command can be found in the Stata documentation for that command (link).

2 Recoding a Variable

2.1 Description

The recode command is used to reassign numeric values within a variable. It is most often used with categorical or ordinal-level data.

2.2 Syntax

```
recode varname (rule) [(rule) ...]  [, generate(newvar)]
```

2.3 Examples

The following example uses the `census.dta` dataset that comes pre-installed with Stata. It creates a binary (`0/1`) indicator for being located in the U.S. Census Bureau's New England region where `1` represents states in New England:

```
.  sysuse census.dta
(1980 Census data by state)

.  generate newEngland = region

.  recode newEngland 1=1 2/4=0
(newEngland:  41 changes made)

.  tabulate region newEngland
```

Census region	newEngland 0	1	Total
NE	0	9	9
N Cntrl	12	0	12
South	16	0	16
West	13	0	13
Total	41	9	50

The recode command in this example explicitly states that the value `1` should remain `1`. This is not required by Stata, but it is good practice to make code as explicit as possible. Including the `1=1` statement in the recode command, it is unambiguous that the author of the code intended for no change to be made to `1` values. Not including it leaves open the door that it was omitted by mistake.

2.4 Notes

Additional details for the recode command can be found in the Stata documentation for that command (link).

3 Variable Order (Part 2)

3.1 Description

The `order` command is used to re-order the variables within a dataset. When combined with the `drop` and/or `keep` commands, the `order` command gives you full control over the variable structure of a particular dataset.

3.2 Syntax

`order` *varlist* [, *options*]

3.3 Options

There are six options for the `order` command, four of which are particularly useful to know. The first is the `first` option, which moves the *varlist* to the beginning of the dataset. The second is the `last` option, which moves the *varlist* to the end of the dataset.

The third and fourth options allow you to move the *varlist* to the middle of a dataset. They do so by specifying the variable that the *varlist* should come before (the `before` option) or after (the `after` option). Both of these options require the specification of a `(`*varname*`)`.

3.4 Examples: `first` and `last`

The following examples each use the `census.dta` dataset that comes pre-installed with Stata. The following example moves the variable `state2` to the beginning of the dataset:

```
. sysuse census.dta
(1980 Census data by state)

. describe

Contains data from /Applications/Stata/ado/base/c/census.dta
  obs:            50                          1980 Census data by state
 vars:            13                          6 Apr 2014 15:43
 size:         2,900
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
state           str14   %-14s                 State
state2          str2    %-2s                  Two-letter state abbreviation
region          int     %-8.0g     cenreg     Census region
```

(*Output omitted*)

```
. order state2, first
```

(*Continued on next page*)

```
. describe

Contains data from /Applications/Stata/ado/base/c/census.dta
  obs:            50                          1980 Census data by state
 vars:            13                          6 Apr 2014 15:43
 size:         2,900
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
state2          str2    %-2s                  Two-letter state abbreviation
state           str14   %-14s                 State
region          int     %-8.0g     cenreg     Census region
```

(*Output omitted*)

Similarly, we can move the `state` and `region` variables in the `census.dta` dataset to the end of the dataset using the `last` option:

```
. order state region, last
. describe

Contains data from /Applications/Stata/ado/base/c/census.dta
  obs:            50                          1980 Census data by state
 vars:            13                          6 Apr 2014 15:43
 size:         2,900
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
state2          str2    %-2s                  Two-letter state abbreviation
pop             long    %12.0gc               Population
```

(*Output omitted*)

```
divorce         long    %12.0gc               Number of divorces
state           str14   %-14s                 State
region          int     %-8.0g     cenreg     Census region
-------------------------------------------------------------------------------
Sorted by:
```

3.5 Examples: `before` and `after`

The following examples each use the `census.dta` dataset that comes pre-installed with Stata. The following example moves the variables `death, marriage,` and `divorce` to the middle of the dataset, before the variable `pop`:

```
. sysuse census.dta
(1980 Census data by state)

. describe

Contains data from /Applications/Stata/ado/base/c/census.dta
  obs:            50                          1980 Census data by state
 vars:            13                          6 Apr 2014 15:43
 size:         2,900
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
state           str14   %-14s                 State
state2          str2    %-2s                  Two-letter state abbreviation
region          int     %-8.0g     cenreg     Census region
pop             long    %12.0gc               Population
```

(*Output omitted*)

```
medage          float   %9.2f                 Median age
death           long    %12.0gc               Number of deaths
marriage        long    %12.0gc               Number of marriages
divorce         long    %12.0gc               Number of divorces
-------------------------------------------------------------------------------
Sorted by:

. order death marriage divorce, before(pop)
```

(*Continued on next page*)

```
. describe

Contains data from /Applications/Stata/ado/base/c/census.dta
  obs:            50                          1980 Census data by state
 vars:            13                          6 Apr 2014 15:43
 size:         2,900
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
state           str14   %-14s                 State
state2          str2    %-2s                  Two-letter state abbreviation
region          int     %-8.0g     cenreg     Census region
death           long    %12.0gc               Number of deaths
marriage        long    %12.0gc               Number of marriages
divorce         long    %12.0gc               Number of divorces
pop             long    %12.0gc               Population
```

(*Output omitted*)

The same change to the variable order could have been accomplished with the `after` option. Since the variable `region` comes right before the variable `pop`, we could use the following syntax to achieve the same result as the previous example:

```
. order death marriage divorce, after(region)
```

3.6 Notes

Additional details for the `order` command can be found in the Stata documentation for that command (link).

4 Creating a Codebook

4.1 Description

The codebook command produces documentation about your dataset.

4.2 Syntax

```
codebook [, header notes]
```

4.3 Example

To create a codebook for the census.dta dataset that comes pre-installed with Stata, use the following command:

```
. sysuse census.dta
(1980 Census data by state)

. codebook, header notes

                Dataset:  /Applications/Stata/ado/base/c/census.dta
           Last saved:   6 Apr 2014 15:43

                Label:  1980 Census data by state
   Number of variables:  13
Number of observations:  50
                 Size:  2,900 bytes ignoring labels, etc.

-------------------------------------------------------------------------------
state                                                                   State
-------------------------------------------------------------------------------

                type:  string (str14), but longest is str13

        unique values:  50                       missing "":  0/50

             examples:  "Georgia"
                        "Maryland"
                        "Nevada"
                        "S. Carolina"

             warning:  variable has embedded blanks
```

(*Output omitted*)

4.4 Notes

Additional details for the codebook command can be found in the Stata documentation for that command (link).

5 Common Errors

5.1 Error 111: Variable Not Found

When referencing a variable using any command that includes a *varlist*, you may receive this error:

```
. sysuse census.dta
(1980 Census data by state)

. list state3 in 1/5
variable state3 not found
r(111);
```

This error indicates one of two conditions: (1) the variable does not exist in the dataset or (2) there is a misspelling in the variable name.

5.2 Error 198: Option Not Allowed

If an option is misspelled or is not actually a valid option for a command, you will see the following error. In this case, the option for the `summarize` command has been misspelled - it is `detail`, not `detailed`:

```
. summarize pop, detailed
option detailed not allowed
r(198);
```

5.3 Error 199: Command Not Recognized

If a command is misspelled, it will generate the following error. In this case, the command `describe` has been misspelled:

```
. dscribe type
command dscribe is unrecognized
r(199);
```

If the command was written as part of a user-installed package, this error could also indicate that package has not been installed locally.