spcompress — Compress Stata-format shapefile

Description | Quick start | Menu | Syntax
Option | Remarks and examples | Stored results | Also see

Description

`spcompress` creates a new Stata-format shapefile omitting places (geographical units) that do not appear in the Sp data in memory. The new shapefile will be named after the data in memory.

Quick start

Create new file `new_shp.dta` containing only cases identified by `mysample` from `old_shp.dta`

```
use old
keep if mysample
save new
spcompress
```

Menu

Statistics > Spatial autoregressive models

Syntax

```
spcompress [ , force ]
```

`collect` is allowed; see **[U] 11.1.10 Prefix commands**.

Option

`force` allows replacing an existing shapefile. `force` is the option name StataCorp uses when you should think twice before specifying it. In most cases, you want to create a new shapefile.

Remarks and examples

Remarks are presented under the following headings:

Introduction
Using the force option

Introduction

In [SP] **Intro 4** and [SP] **Intro 7**, we discussed how to find and prepare the analysis dataset, `tl_2016_us_county.dta`, and the shapefile dataset, `tl_2016_us_county_shp.dta`. We again use those datasets here.

You sometimes want to analyze a subset of the data. In those cases, you might type

```
. use tl_2016_us_county            // use all the data
. keep if STATEFP == "48"          // keep the subset of interest
. save texas                       // save under a different name
```

All will work fine. File `texas.dta` is linked to `tl_2016_us_county_shp.dta`, which contains a lot of unnecessary information, but that will cause Sp no difficulty.

Next, you can type

```
. spcompress
```

Now, files `tl_2016_us_county.dta` and `tl_2016_us_county_shp.dta` remain unchanged, and file `texas_shp.dta` was created. `texas.dta` was resaved so that the copy on disk would reflect that it is now linked to `texas_shp.dta` instead of `tl_2016_us_county_shp.dta`.

Sp will run a little faster if we compress the shapefile. We say a little because only `grmap` will run faster.

Using the force option

Above, we showed an example. Here is what would have happened had we omitted the line `save texas`:

```
. use tl_2016_us_county
. keep if STATEFP == "48"
(2,979 observations deleted)
. * save texas              // save texas intentionally commented out
. spcompress
file tl_2016_us_county_shp.dta already exists
r(602);
```

Whether you type `save texas` makes all the difference. Do you really want to replace `tl_2016_us_county_shp.dta`? If so, specify `force`.

The option is called `force` because Stata wonders whether you really meant to type

```
. use tl_2016_us_county, clear
. keep if STATEFP == "48"
(2,979 observations deleted)
. save texas
file texas.dta saved
. spcompress
  (texas_shp.dta created with 254 spatial units, 2,979 fewer than previously)
  (texas_shp.dta saved)
  (texas.dta saved)
```

Even if you intended to discard all but Texas from `tl_2016_us_county.dta` and `tl_2016_us_county_shp.dta`, we would recommend that you type

```
. use tl_2016_us_county
. keep if STATEFP == "48"
. save texas
. spcompress
. erase tl_2016_us_county.dta
. erase tl_2016_us_county_shp.dta
```

Stored results

`spcompress` stores the following in `r()`:

Scalars

`r(num_drop_ids)`	# of spatial units dropped
`r(num_ids)`	# of spatial units remaining

Also see

[SP] **Intro** — Introduction to spatial data and SAR models

[D] **compress** — Compress data in memory

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.

For suggested citations, see the FAQ on citing Stata documentation.