

*The A to Z of how to create thematic maps
of Italy using spmap*

Maurizio Pisati

Department of Sociology and Social Research
University of Milano-Bicocca (Italy)
maurizio.pisati@unimib.it

2014 Italian Stata Users Group meeting
Milano
November 13-14, 2014

Outline

- ① Tools of the trade
- ② Getting Italy into shape
- ③ Making spmaps
- ④ Spatial reference systems

TOOLS OF THE TRADE

shp2dta

Type User-written Stata command

Purpose Converts shapefiles into pairs of Stata datasets (*attribute* and *coordinate* datasets) ready to be used with `spmap`

Author Kevin Crow (StataCorp LP)

Availability Freely downloadable from the SSC Archive

mergepoly

Type User-written Stata command

Purpose Dissolves boundaries between adjacent polygons of a Stata coordinate dataset, possibly by attribute values

Authors Robert Picard and Michael Stepner

Availability Freely downloadable from the SSC Archive

spmap

Type User-written Stata command

Purpose Visualizes spatial data of various kinds

Author Maurizio Pisati (University of Milano-Bicocca)

Availability Freely downloadable from the SSC Archive

ogr2ogr

- Type** Command-line application included in the Geospatial Data Abstraction Library (GDAL)
- Purpose** Converts one OGR defined data source into another OGR data source, performing various operations during the process, such as reprojection
- Authors** Open Source Geospatial Foundation
- Availability** Released under an X/MIT style Open Source license. Links to binaries for several platforms are available at <http://trac.osgeo.org/gdal/wiki/DownloadingGdalBinaries>

GETTING ITALY INTO SHAPE

Shapefiles

- To draw a geographical map, you need one or more files defining the boundaries of the geographical area of interest and, possibly, of its administrative subdivisions
- The **Esri shapefile** is one of the most common vector formats for storing geospatial data of this kind
- A *shapefile* is actually a set of three mandatory files, plus one or more optional files
- Mandatory files:
 - **.shp**: stores the coordinates of the spatial objects
 - **.dbf**: stores the attributes of the spatial objects
 - **.shx**: indexes the spatial objects
- Relevant optional file:
 - **.prj**: defines the spatial reference system

Shapefiles of Italy

- The most complete collection of shapefiles of Italy is made publicly available by the Italian National Institute of Statistics (ISTAT)
- The ISTAT website has a section dedicated to cartographic resources with links to shapefiles of Italy at various levels of administrative units: regions, provinces, municipalities, sub-municipalities, census areas, and census tracts
- ISTAT also makes available other shapefiles representing functional subdivisions of Italy – such as court districts, health districts, urban areas, dioceses – and several kinds of infrastructures – such as airports, railways, highways
- Check it out at
<http://www.istat.it/it/strumenti/cartografia>

Other sources of shapefiles

- Other good sources of shapefiles relevant to Italy are:
 - <http://www.datiopen.it>: general collection of open data regarding Italy, including geospatial data
 - <http://www.dati.gov.it>: public data archive made available by the Italian government
 - <http://www.cartografia.regione.lombardia.it>: public archive of geospatial data regarding Regione Lombardia
 - <http://dati.comune.milano.it>: open data archive of the municipality of Milano

MAKING SPMAPS

Example 1

Regional boundaries (Italy 2011)

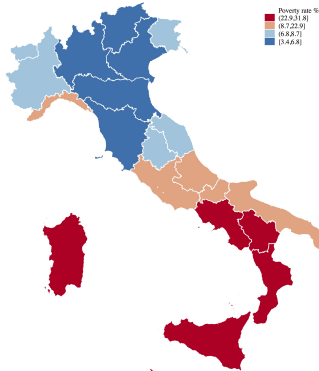
```
/* 1. Get shapefile */  
copy "http://www.istat.it/it/files/2011/04/reg2011.zip" "reg.zip"  
unzipfile "reg.zip", replace  
  
/* 2. Convert shapefile to Stata attribute and coordinate datasets */  
shp2dta using reg2011, data("reg-attr.dta") coord("reg-coord.dta") ///  
    genid(stid) gcentroids(cc) replace  
  
/* 3. Draw map */  
use "reg-attr.dta", clear  
spmap using "reg-coord.dta", id(stid) ocolor(black) osize(vthin)
```



Example 2

Poverty rate by region (Italy 2007)

```
/* 1. Get Excel dataset */  
copy "http://www.istat.it/it/files/2011/02/tavole_131.zip" "data.zip"  
unzipfile "data.zip", replace  
  
/* 2. Import data of interest into Stata */  
import excel using "Capitolo 13.xls", sheet("13.8 2007") ///  
    cellrange(G7:G28) clear  
drop in 5/6  
rename G poor  
format poor %4.1f  
save "poor.dta", replace  
  
/* 3. Merge data with attribute dataset */  
use "reg-attr.dta", clear  
merge 1:1 _n using "poor.dta"  
  
/* 4. Draw map */  
spmap poor using "reg-coord.dta", id(stid) fcolor(BuRd) ///  
    ocolor(white ..) osize(thin ..) legend(position(2)) ///  
    legtitle("Poverty rate %")
```



Example 3

Dissolving regional boundaries to generate Italy outline

```
/* 1. Dissolve polygons stored in existing coordinate dataset */  
use "reg-attr.dta", clear  
mergepoly stid using "reg-coord.dta", coord("ita-coord.dta") replace  
  
/* 2. Draw map using resulting coordinate dataset */  
spmap using "ita-coord.dta", id(stid) ocolor(black) osize(vthin)
```



Example 3

Dissolving regional boundaries to generate Italy outline

```
/* 1. Dissolve polygons stored in existing coordinate dataset */  
use "reg-attr.dta", clear  
mergepoly stid using "reg-coord.dta", coord("ita-coord.dta") replace  
  
/* 2. Draw map using resulting coordinate dataset */  
spmap using "ita-coord.dta", id(stid) ocolor(black) osize(vthin)
```



Example 3

Dissolving regional boundaries to generate Italy outline

```
/* 1. Dissolve polygons stored in existing coordinate dataset */  
use "reg-attr.dta", clear  
mergepoly stid using "reg-coord.dta", coord("ita-coord.dta") replace  
  
/* 2. Draw map using resulting coordinate dataset */  
spmap using "ita-coord.dta", id(stid) ocolor(black) osize(vthin)
```



Example 4

Mapping line spatial objects : Lombardia's regional rail network

```
/* 1. Get shapefile */
local URL "https://www.dati.lombardia.it/download"
copy "URL"/wkc8-2hf7/application/zip" "railnet.zip"
unzipfile "railnet.zip", replace

/* 2. Convert shapefile to Stata attribute and coordinate datasets */
shp2dta using Rete_ferroviaria_10000_CT10_line, ///
data("railnet-attr.dta") coord("railnet-coord.dta") ///
genid(stid) replace

/* 3. Create layer variable */
use "railnet-attr.dta", clear
generate type = 3
replace type = 1 if DN_BIN=="Non valutabile"
replace type = 1 if DN_BIN=="Non valutato"
replace type = 2 if DN_BIN=="Un solo binario"
label define type 1 "Undefined", modify
label define type 2 "1 track", modify
label define type 3 "2+ tracks", modify
label values type type
keep stid type
rename stid _ID
save "type.dta", replace

/* 4. Merge layer variable with coordinate dataset */
use "railnet-coord.dta", clear
merge n:1 _ID using "type.dta"
save "railnet-coord.dta", replace

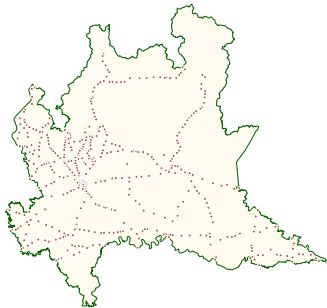
/* 5. Draw map */
use "reg-attr.dta", clear
spmap using "reg-coord.dta" if stid==3, id(stid) ///
fcolor(eggshell) ocolor(dkgreen) osize(thin) ///
line(data("railnet-coord.dta") by(type) ///
color(gs12 sand maroon) size(medium ..) ///
legenda(on) legend(position(2))
```



Example 5

Mapping point spatial objects : Lombardia's railway junctions

```
/* 1. Get shapefile */  
local URL "https://www.dati.lombardia.it/download"  
copy "`URL'/v8yp-qek6/application/zip" "railjun.zip"  
unzipfile "railjun.zip", replace  
  
/* 2. Convert shapefile to Stata attribute and coordinate datasets */  
shp2dta using Nodi_della_rete_ferroviaria_10000_CT10_point, ///  
data("railjun-attr.dta") coord("railjun-coord.dta") ///  
genid(stid) replace  
  
/* 3. Draw map */  
use "reg-attr.dta", clear  
spmap using "reg-coord.dta" if stid==3, id(stid) ///  
fcolor(eggshell) ocolor(dkgreen) osize(thin) ///  
point(data("railjun-coord.dta") x(_X) y(_Y) ///  
size(*0.6) fcolor(sienna) ocolor(white) ///  
osize(vvthin))
```



SPATIAL REFERENCE SYSTEMS

Spatial reference systems for Italy

- A **spatial reference system** (SRS) is the system of coordinates used to represent the shape and/or the location of the spatial objects of interest
- The shapefiles of Italy distributed by ISTAT – as well as by most of the other providers of cartographic resources relevant to Italy – use either the “ED50/UTM zone 32N” or the “WGS84/UTM zone 32N” spatial reference system
- Both of these systems are *projected* SRS, i.e., they represent spatial objects on a two-dimensional *planar* surface
- On most scales, the “ED50/UTM zone 32N” and the “WGS84/UTM zone 32N” SRS are practically indistinguishable from each other, so we can safely mix them in the same map

Changing the spatial reference system

- Sometimes, the spatial data of interest are expressed in a spatial reference system different from “ED50/UTM zone 32N” or “WGS84/UTM zone 32N”
- Typically, we might have spatial data expressed in a *spherical* coordinate system, where each point is defined by a latitude/longitude coordinate pair
- For example, suppose we would like to draw a map showing the spatial distribution of historical street markets in Lombardia

Example 6

Mapping unprojected point data onto a projected base map



```
/* 1. Get dataset in CSV format */
local URL "https://www.dati.lombardia.it/api/views"
copy "`URL'/jayn-vsw3/rows.csv" "markets.csv"

/* 2. Import dataset into Stata */
insheet using "markets.csv", comma names clear

/* 3. Generate Stata point dataset */
keep idmerc coord_lng
replace coord_lng = substr(coord_lng,1,10)
replace coord_lng = substr(coord_lng,11,10)
split coord_lng, generate(coord) parse(",") destring
rename coord1 ycoord
rename coord2 xcoord
drop coord_lng
save "markets.dta", replace

/* 4. Draw map */
use "reg-attr.dta", clear
spmap using "reg-coord.dta" if stid==3, id(stid) ///
fcolor(eggshell) ocolor(dkgreen) osize(vthin) ///
point(data("markets.dta") x(xcoord) y(ycoord) ///
fcolor(orange))
```

Changing the spatial reference system

- Unfortunately, in the Lombardia's historical street markets dataset, locations are expressed in lat/lon format, while our base map is expressed in the “ED50/UTM zone 32N” spatial reference system
- Thus, we need to convert the unprojected point data into projected point data expressed in the “ED50/UTM zone 32N” SRS
- To this end, the `ogr2ogr` command-line application comes in handy

Example 7

Lombardia's historical street markets

```
/* 1. Export Stata dataset into CSV dataset */
use "markets.dta", clear
outsheet using "hsm0.csv", comma replace

/* 2. Generate VRT file */
file open VRT using "hsm0.vrt", write text replace
file write VRT "<OGVRTDataSource>" _n
file write VRT "" <OGVRTLayer name="hsm0">" _n
file write VRT "" <SrcDataSource>hsm0.csv</SrcDataSource>" _n
file write VRT "" <GeometryType>wkbPoint</GeometryType>" _n
file write VRT "" <LayerSRS>WGS84</LayerSRS>" _n
file write VRT "" <GeometryField encoding="PointFromColumns" x="xcoord" y="ycoord"/>" _n
file write VRT "" <Field name="idmerc" src="idmerc" type="String" />" _n
file write VRT "" </OGVRTLayer>" _n
file write VRT "" </OGVRTDataSource>" _n
file close VRT

/* 3. Convert VRT file to shapefile */
shell ogr2ogr -f 'ESRI Shapefile' "hsm0.shp" "hsm0.vrt"

/* 4. Convert unprojected shapefile to projected shapefile */
shell ogr2ogr -t_srs EPSG:23032 "hsm.shp" "hsm0.shp"
```

Example 7

Lombardia's historical street markets

```
/* 5. Convert shapefile to Stata attribute and coordinate datasets */  
shp2dta using hsm, data("hsm-attrib.dta") coord("hsm-coord.dta") ///  
genid(stid) replace  
  
/* 6. Draw map */  
spmap using "reg-coord.dta" if stid==3, id(stid) ///  
fcolor(eggshell) ocolor(dkgreen) osize(thin) ///  
point(data("hsm-coord.dta") x(_X) y(_Y) ///  
size(*0.8) fcolor(orange) ocolor(white) ///  
osize(vvthin))
```

