

Simple Thematic Mapping in Stata

1st Italian Stata Users
Group Meeting
Roma, 25 October 2004

Maurizio Pisati
University of Milano Bicocca – Italy
maurizio.pisati@unimib.it

Thematic maps

- Thematic maps represent the spatial distribution of one or more variables of interest within a given geographical unit

Thematic maps

- Examples:
 - A sociologist could use a choropleth map (a.k.a. shaded map) to show how the percentage of families below the poverty line varies across the states or the provinces of a given country
 - A police officer could be interested in analyzing a dot map showing the locations of drug markets within a given city

Software for thematic mapping

- Usually, to produce state-of-the-art thematic maps one has to resort to specialized software (e.g., ArcView, MapInfo)
- In some cases, however, it is possible to exploit the graphical engine of a general-purpose statistical package to draw simple but effective thematic maps

Stata's mapping capabilities

- Up until version 7, Stata offered very limited mapping capabilities
- On the other hand, the graphical engine introduced in Stata 8 is quite flexible and makes it possible to draw several kinds of maps in a relatively simple manner

The `tmap` package

- The `tmap` package is a suite of Stata programs designed to draw five kinds of thematic map:
 - Choropleth maps
 - Proportional symbol maps
 - Deviation maps
 - Dot maps
 - Label maps

The `tmap` package

- Choropleth, proportional symbol, and deviation maps are intended to depict area data
- Dot maps are suitable for representing point data
- Label maps can be used to show data of both types

The `tmap` package

- The `tmap` package exploits the possibility – offered by the new Stata graphical engine – to overlay a large number of different graphs, each of which is used to create a distinct element of the desired map

The tmap package

- Specifically:
 - `graph twoway area` is used to draw the outlines of the geographical areas of interest and to fill them with the appropriate colors
 - `graph twoway scatter` is used to plot the proper symbols or labels when required

tmap choropleth

- `tmap choropleth` represents the spatial distribution of area data by means of choropleth maps, i.e., maps where each of n sub-areas is colored (or shaded) according to a discrete scale based on the value taken on by a quantitative variable of interest in that sub-area

`tmap choropleth`

- The number of classes that make up the discrete scale must be between 2 and 9
- The corresponding class breaks can be based on four different criteria:
 - Quantiles
 - Equal intervals
 - Standard deviates
 - Custom

`tmap propsymbol`

- `tmap propsymbol` represents the spatial distribution of area data by means of proportional symbol maps, i.e., maps where the value taken on by a quantitative variable of interest in each of n sub-areas is represented by a symbol whose size is proportional to the value itself

tmap deviation

- **tmap deviation** represents the spatial distribution of area data by means of deviation maps, a particular kind of proportional symbol maps where:
 - symbol **size** expresses the absolute deviation of the quantitative variable of interest from its mean or median
 - symbol **fill** expresses the sign of the deviation (positive or negative)

tmap dot

- **tmap dot** represents the spatial distribution of point data by means of dot maps, i.e., maps where the locations at which some “events” of interest have occurred are indicated by symbols whose color or shape can vary according to the type of “event”

tmap label

- **tmap label** is an auxiliary program that allows the user to superimpose onto a base map the values taken on by a numeric or string variable at different locations
- This program can be used, for example, to plot sub-area names or to represent the spatial distribution of a given quantitative variable of interest in numeric form

Base maps

- All the programs included in the **tmap** package require that the boundaries of the whole geographical unit of interest R or of its sub-areas be stored in an external file

Base maps

- This file must always include the following three variables:
 - `_ID`, which contains the numeric identifier of the polygon(s) representing R or its sub-areas
 - `_x`, which contains the x -coordinates of the polygon(s)
 - `_y`, which contains the y -coordinates of the polygon(s)

Base maps

- The coordinates of each polygon must be arranged so as to correspond to consecutive nodes
- Each polygon must be “closed”, i.e., the last pair of coordinates of each polygon must be equal to the first pair

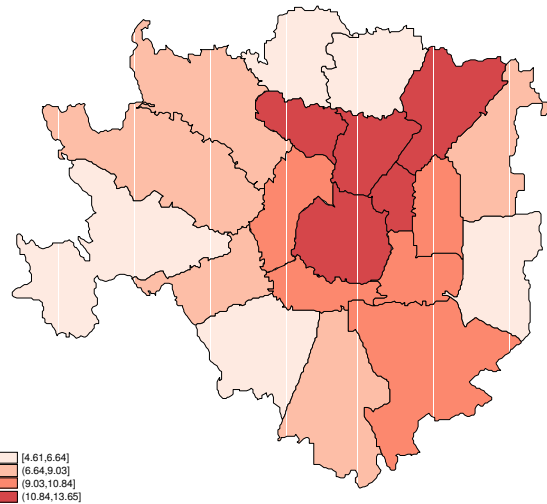
Base maps: example

<code>_ID</code>	<code>_X</code>	<code>_Y</code>
1	514396	5036528
1	514393	5036473
...
1	514396	5036528
2	516654	5039135
2	516726	5039090
...
2	516654	5039135

Choropleth maps: examples

```
tmap choropleth foreign01,           ///
  id(id)                             ///
  map("Milano-AreaMap.dta")         ///
  clmethod(quantile)                 ///
  clnumber(4)                         ///
  palette(Reds)
```

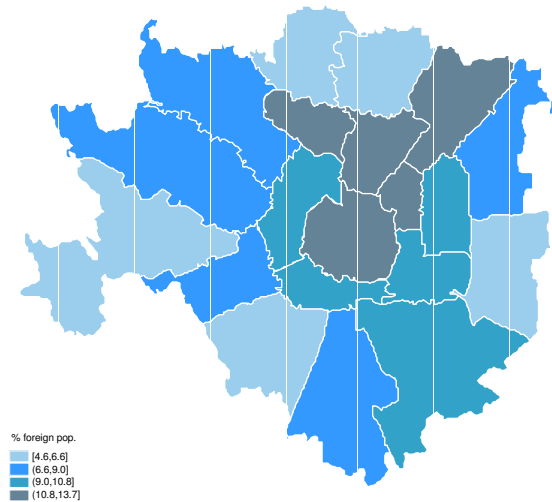
Percentage foreign population, by administrative district.
Milano, 2001



Choropleth maps: examples

```
tmap choropleth foreign01,      ///  
id(id)                        ///  
map("Milano-AreaMap.dta")    ///  
clmethod(quantile)           ///  
clnumber(4)                  ///  
palette(Custom)              ///  
colors(eltblue midblue      ///  
       ebblue edkblue)      ///  
ocolor(white)                ///  
osize(medium)                ///  
legtitle("% foreign pop.")   ///  
legformat("%4.1f")
```

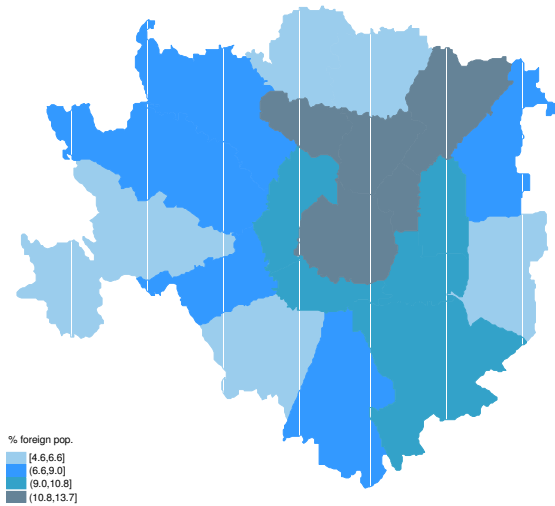
Percentage foreign population, by administrative district.
Milano, 2001



Choropleth maps: examples

```
tmap choropleth foreign01,      ///
id(id)                        ///
map("Milano-AreaMap.dta")    ///
clmethod(quantile)           ///
clnumber(4)                   ///
palette(Custom)               ///
colors(eltblue midblue       ///
       ebblue edkblue)       ///
ocolor(none)                  ///
osize(none)                   ///
legtitle("% foreign pop.")    ///
legformat("%4.1f")
```

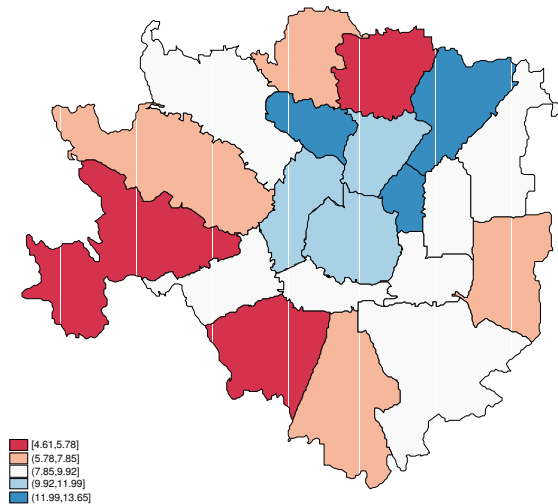
Percentage foreign population, by administrative district.
Milano, 2001



Choropleth maps: examples

```
tmap choropleth foreign01,      ///
      id(id)                    ///
      map("Milano-AreaMap.dta") ///
      clmethod(stdev)           ///
      clnumber(5)
```

Percentage foreign population, by administrative district.
Milano, 2001

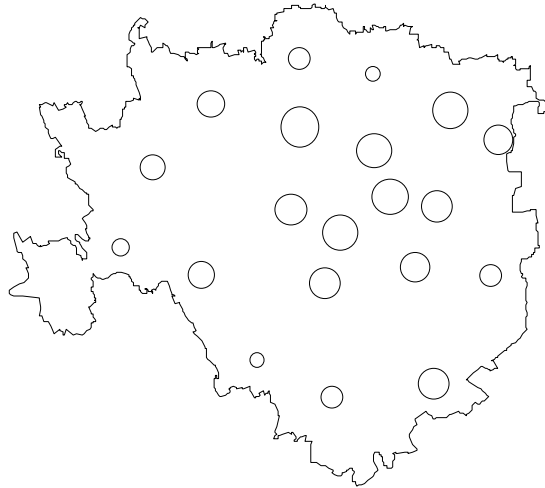


Prop. symbol maps: examples

```
tmap propsymbol foreign01,           ///  
  x(x) y(y)                          ///  
  map("Milano-GeneralMap.dta")
```

Percentage foreign population, by administrative district.
Milano, 2001

Symbol size proportional to variable value

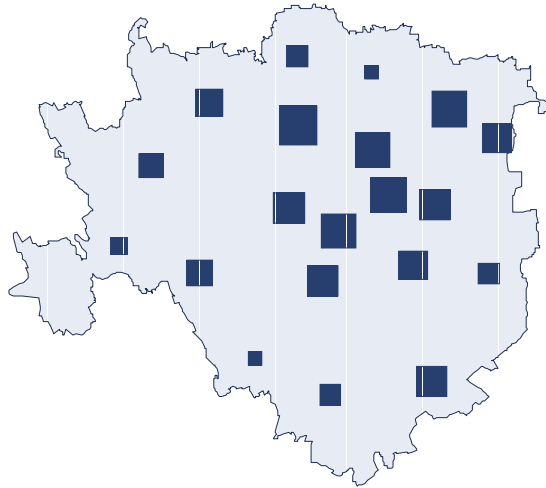


Prop. symbol maps: examples

```
tmap propsymbol foreign01,      ///  
  x(x) y(y)                    ///  
  map("Milano-GeneralMap.dta") ///  
  sshape(S)                    ///  
  scolor(navy)                 ///  
  ocolor(dknavy)              ///  
  fcolor(ltbluishgray)
```

Percentage foreign population, by administrative district.
Milano, 2001

Symbol size proportional to variable value

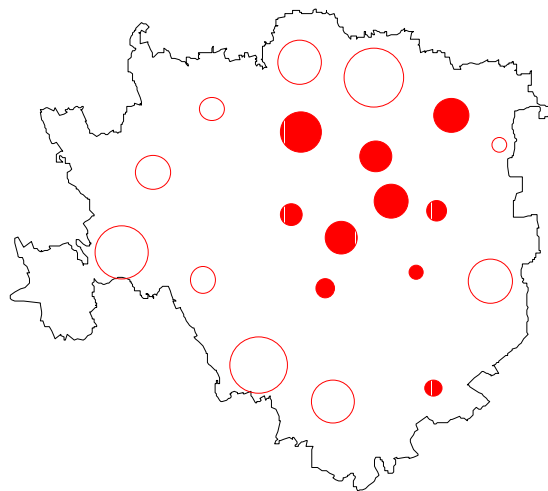


Deviation maps: example

```
tmap deviation foreign01,      ///  
  x(x) y(y)                   ///  
  map("Milano-GeneralMap.dta") ///  
  scolor(red)
```


Percentage foreign population, by administrative district:
Deviations from overall mean. Milano 2001

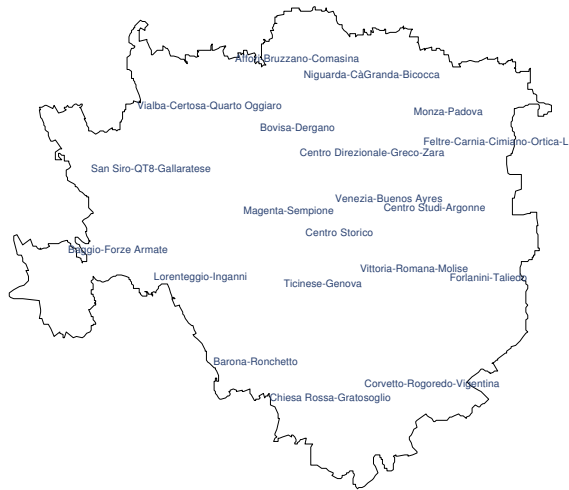
Solid symbols=positive deviations - Hollow symbols=negative deviations
Symbol size proportional to absolute value of deviations



Label maps: example

```
tmap label district,          ///  
  x(x) y(y)                  ///  
  map("Milano-GeneralMap.dta")  ///  
  llength(30)                 ///  
  lsize(0.8)                   ///  
  lcolor(navy)
```

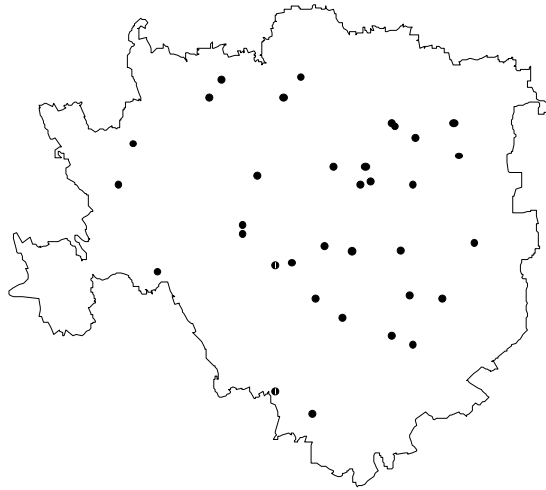
Administrative districts of Milano



Point maps: examples

```
tmap dot, //  
  x(x) y(y) //  
  map("Milano-GeneralMap.dta")
```

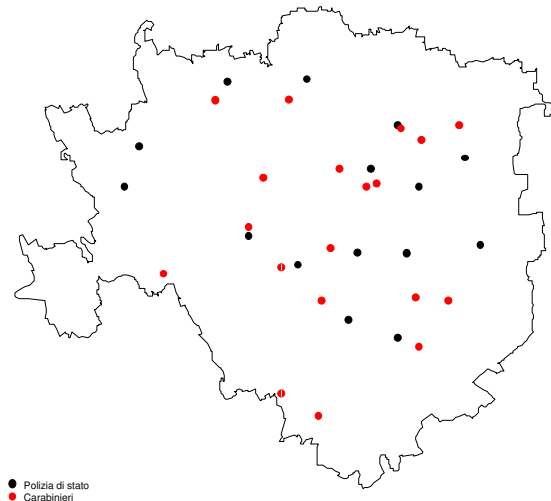
Location of police stations.
Milano, 2004



Point maps: examples

```
tmap dot, ///  
  x(x) y(y) ///  
  map("Milano-GeneralMap.dta") ///  
  by(type)
```

Location of police stations, by police force.
Milano, 2004



Acknowledgments

- The color schemes used in **tmap choropleth** were designed by Dr. Cynthia A. Brewer, Department of Geography, The Pennsylvania State University, University Park, Pennsylvania, USA. The color schemes are used with Dr. Brewer's permission and are from the ColorBrewer map design tool available at ColorBrewer.org

Acknowledgments

- I wish to thank an anonymous reviewer and Nick Cox for their valuable suggestions and for helping improve the `tmap` package
- Any remaining errors are mine