2018 Canadian Stata Conference
Morris J. Wosk Centre for Dialogue, Vancouver, BC

geotools: Exporting cartography data from Stata to GIS systems

Sergiy Radyakin
sradyakin@worldbank.org

Development Economics Data (DECDG),
The World Bank

July 27, 2018

Introduction

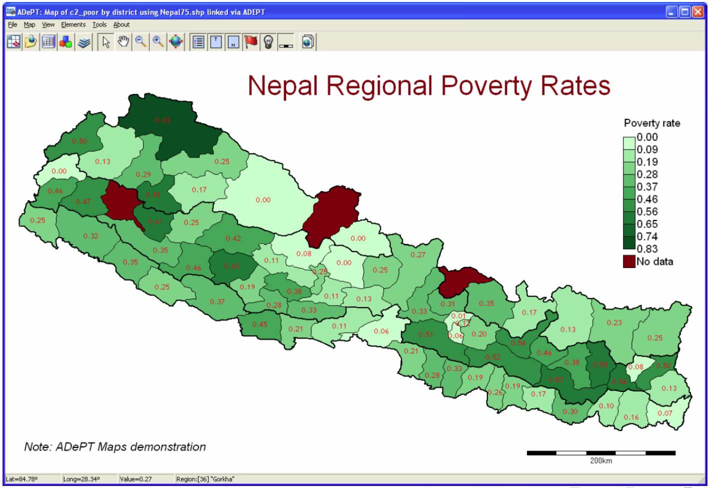
"I am told there are people who do not care for maps, and I find it hard to believe."

- Robert Louis Stevenson, August 1894

`https://ebooks.adelaide.edu.au/s/stevenson/robert_louis/s848aw/part5.html`

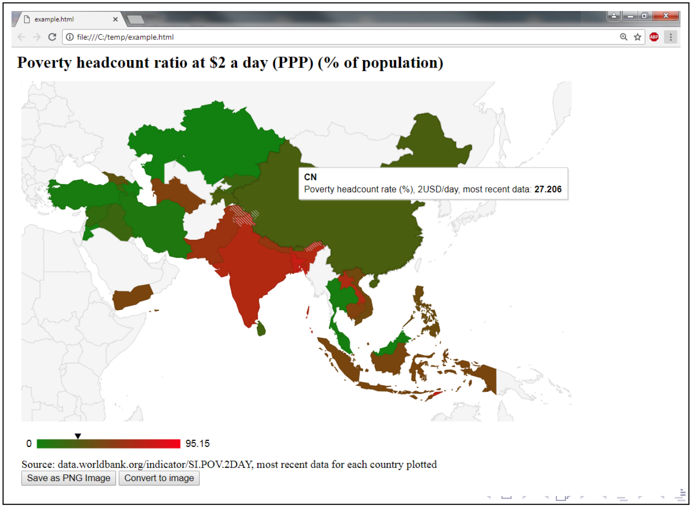
ADePT_maps (ca. 2007)

ADePT Maps was a command (**amap**) allowing to build and manipulate maps interactively (in Windows only), utilizing shapefiles directly:


ADePT: Map of c2_poor by district using Nepal75.shp linked via ADEPT
File Map View Elements Tools About
Nepal Regional Poverty Rates
Poverty rate
0.00
0.09
0.19
0.28
0.37
0.46
0.56
0.65
0.74
0.83
No data
Note: ADePT Maps demonstration
200km
Lat=84.78° Long=28.34° Value=0.27 Region:[36] "Gorkha"

GEOCHART (ca. 2013)

geochart produces an HTML file which utilizes Google's GeoChart component (google.visualization.GeoChart) showing a basic choropleth map from user's data:



Poverty headcount ratio at $2 a day (PPP) (% of population)

CN
Poverty headcount rate (%), 2USD/day, most recent data: **27.206**

Source: data.worldbank.org/indicator/SI.POV.2DAY, most recent data for each country plotted

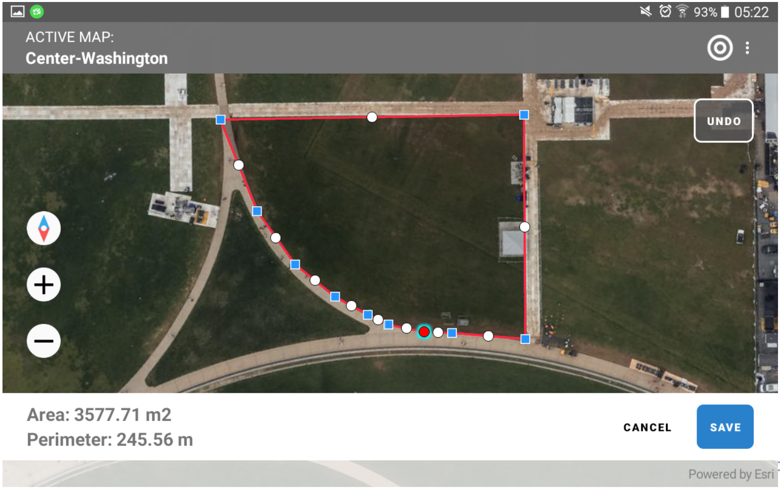
Data sources (Survey Solutions)

GIS data is obtained from direct measurements on location:

Wakiso district, Uganda, 2018

Data sources (Survey Solutions)

GIS data is also obtained from offline satellite images by pointing relevant features interactively on a tablet's screen:



Area: 3577.71 m2
Perimeter: 245.56 m

Data sources (Survey Solutions)

Survey Solutions exports data directly in Stata 14 formatted binary data files (*.dta) with cartography data stored in geography format:

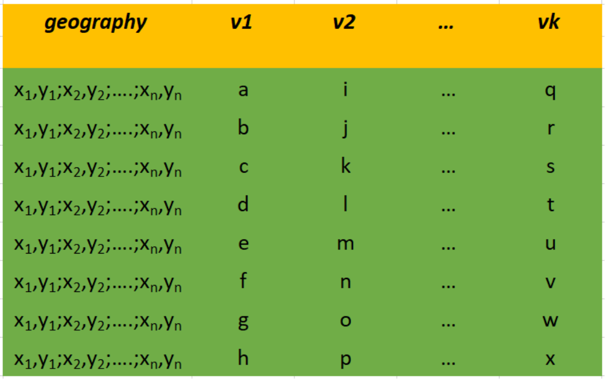
Geography format

$$x_1,y_1;\ x_2,y_2;\ \ldots;\ x_n,y_n$$

Note, this is different from e.g. **shp2dta** (by Kevin Crow) output as shp2dta creates a separate dataset and stores coordinates in separate variables. But shp2dta may also be a source of the coordinates data from existing shape files before their processing/transformation in Stata.

Data sources (Survey Solutions)

Features stored in geography format are exported as part of the main dataset. Each data file may include multiple features alongside other variables.



geography	*v1*	*v2*	*...*	*vk*
$x_1,y_1;x_2,y_2;....;x_n,y_n$	a	i	...	q
$x_1,y_1;x_2,y_2;....;x_n,y_n$	b	j	...	r
$x_1,y_1;x_2,y_2;....;x_n,y_n$	c	k	...	s
$x_1,y_1;x_2,y_2;....;x_n,y_n$	d	l	...	t
$x_1,y_1;x_2,y_2;....;x_n,y_n$	e	m	...	u
$x_1,y_1;x_2,y_2;....;x_n,y_n$	f	n	...	v
$x_1,y_1;x_2,y_2;....;x_n,y_n$	g	o	...	w
$x_1,y_1;x_2,y_2;....;x_n,y_n$	h	p	...	x

GIS features

Survey Solutions collects the following GIS features:

- individual points;
- groups of points;
- paths;
- areas.

Data transformations

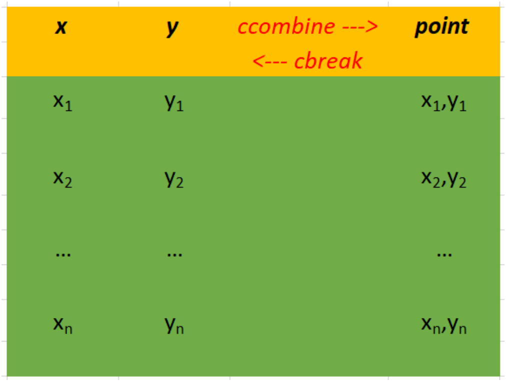
Different data storage conventions may be applied to store data more comfortably for analysis:

- **coordinates** – individual coordinates of points are stored in different variables (e.g. X and Y, Lat and Lon, etc);
- **points** – individual coordinates of points are stored together (string var "X,Y"), but points forming a feature are stored separately;
- **geography** – all coordinates data related to one feature are stored together in one [potentially very long] string variable.

Hence we need utilities to convert between these different formats.

Data transformations

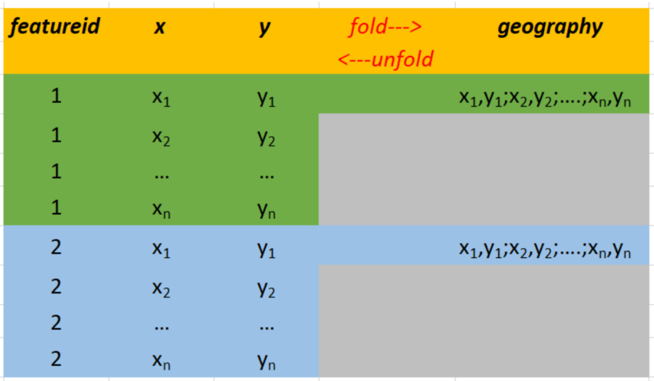
Individual coordinates can be transformed to points and back with the **geoutils ccombine** and **geoutils cbreak** commands:



x	*y*	*ccombine --->* *<--- cbreak*	*point*
x_1	y_1		x_1,y_1
x_2	y_2		x_2,y_2
...	...		...
x_n	y_n		x_n,y_n

Data transformations

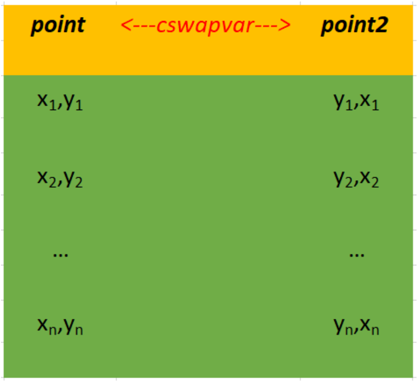
Individual coordinates can be transformed to geography and back with the **geoutils fold** and **geoutils unfold** commands:



featureid	*x*	*y*	*fold---> <---unfold*	*geography*
1	x_1	y_1		$x_1,y_1;x_2,y_2;....;x_n,y_n$
1	x_2	y_2		
1	...	...		
1	x_n	y_n		
2	x_1	y_1		$x_1,y_1;x_2,y_2;....;x_n,y_n$
2	x_2	y_2		
2	...	...		
2	x_n	y_n		

Data transformations

Individual coordinates can be swapped within a geography (or point) variable with **geoutils cswapvar** command:



point	*<---cswapvar--->*	***point2***
x_1,y_1		y_1,x_1
x_2,y_2		y_2,x_2
...		...
x_n,y_n		y_n,x_n

There is also an immediate version of this command to swap coordinates in a string: **geoutils cswapstr** and return the result in an r(coords) result.

Export to GIS

Stata can build maps by representing them as twoway plots (see spmap by Maurizio Pisati). But we often need to export data to GIS for further analysis, transformation, and visualization:

- Shapefiles (*.shp) are dominant with offline systems (ESRI ArcGIS, QGIS, etc).
  - binary format, with open description of file format;
  - different types of layers are stored separately (in separate files);
  - stores coordinates data separately from data attributes;
  - coordinates data does not store any styling info.
- GeoJSON (*.geojson) format is popular with online visualizers/interactive maps (Leaflet, Mapbox);
  - text format (variation of JSON) with open description of file format;
  - different types of layers may be stored jointly (in a single file);
  - stores coordinates data jointly with data attributes;
  - coordinates data may store feature styling info.

Modern GIS systems often support conversion from one format to the other.

Formal output specifications

Commands comprising the geotools package output the data following these formal file format specifications:

GeoJSON

RFC 7946, as was current in 2018/07, accessible at https://tools.ietf.org/html/rfc7946 ;

Shapefile

ESRI Shapefile Technical Description, An ESRI White Paper - July 1998, accessible at https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

Within each format only some features are implemented for export.

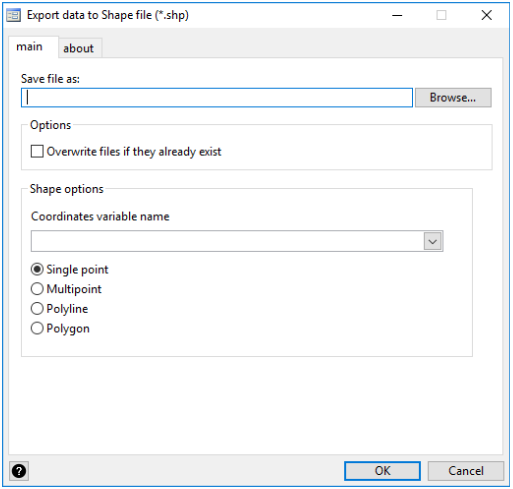
Feature types

Feature	**Shapefile***	**GeoJSON***
Separate points	*point*	*Point*
Multiple points	*multipoint*	*MultiPoint*
Lines	*polyline*	*LineString*
Polygons	*polygon*	*Polygon*

*Spelling exactly as expected by respective export commands. Survey Solutions' geography question follows ESRI Shapefile notation.

Exporting Shapefiles

Type **db shp_save** to export to a shape file GIS format:


Export data to Shape file (*.shp)

main | about

Save file as:

Options

☐ Overwrite files if they already exist

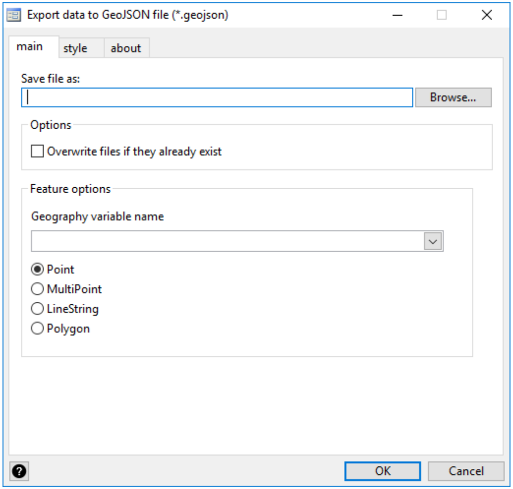
Shape options

Coordinates variable name

(● Single point) ○ Multipoint ○ Polyline ○ Polygon

Exporting GeoJSON files

Type **db gj_save** to export to a file in GeoJSON format:



Export data to GeoJSON file (*.geojson)

main | style | about

Save file as:

Browse...

Options

[ ] Overwrite files if they already exist

Feature options

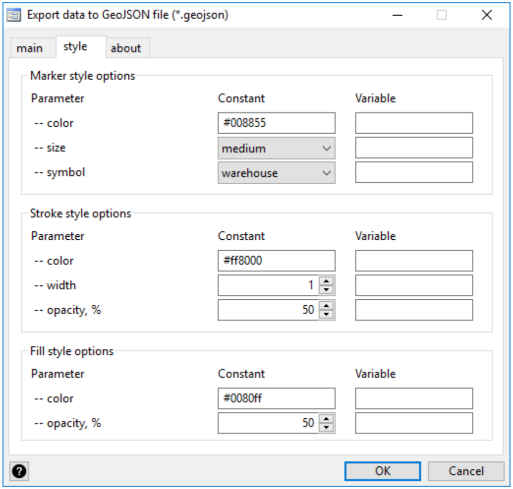
Geography variable name

(●) Point
( ) MultiPoint
( ) LineString
( ) Polygon

OK Cancel

Exporting GeoJSON files (formatting)

Type **db gj_save** to export to a file in GeoJSON format:



Marker style options

Parameter	Constant	Variable
-- color	#008855	
-- size	medium	
-- symbol	warehouse	

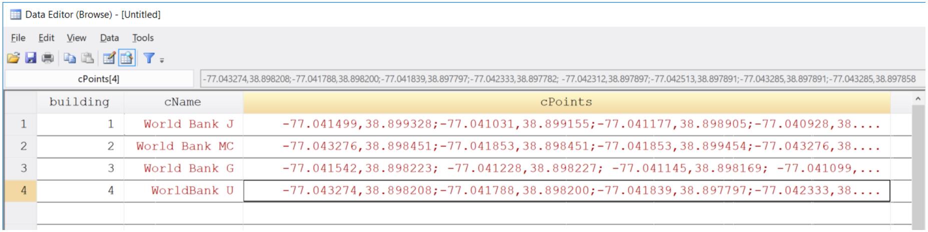
Stroke style options

Parameter	Constant	Variable
-- color	#ff8000	
-- width	1	
-- opacity, %	50	

Fill style options

Parameter	Constant	Variable
-- color	#0080ff	
-- opacity, %	50	

Syntax example: output to a polygon layer



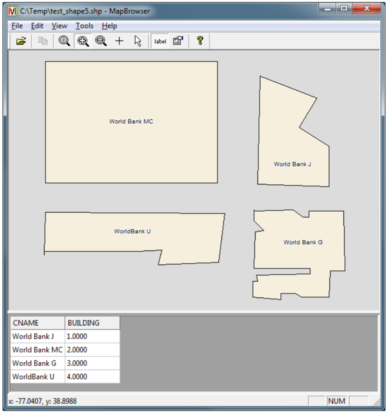
cPoints[4] -77.043274,38.898208;-77.041788,38.898200;-77.041839,38.897797;-77.042333,38.897782; -77.042312,38.897897;-77.042513,38.897891;-77.043285,38.897891;-77.043285,38.897858

	building	cName	cPoints
1	1	World Bank J	-77.041499,38.899328;-77.041031,38.899155;-77.041177,38.898905;-77.040928,38....
2	2	World Bank MC	-77.043276,38.898451;-77.041853,38.898451;-77.041853,38.899454;-77.043276,38....
3	3	World Bank G	-77.041542,38.898223; -77.041228,38.898227; -77.041145,38.898169; -77.041099,...
4	4	WorldBank U	-77.043274,38.898208;-77.041788,38.898200;-77.041839,38.897797;-77.042333,38....

```

clear
do "example\wb_buildings.do"

shapefile save cPoints using "C:\temp\wb_buildings", ///
      replace type(polygon)

geojson save cPoints using "C:\temp\wb_buildings.geojson", ///
      replace features(Polygon)

```

Output (Shapefile)

Example of output to a polygon layer of shapefile format:


C:\Temp\test_shape5.shp - MapBrowser
File Edit View Tools Help
World Bank MC
World Bank J
WorldBank U
World Bank G

CNAME	BUILDING
World Bank J	1.0000
World Bank MC	2.0000
World Bank G	3.0000
WorldBank U	4.0000

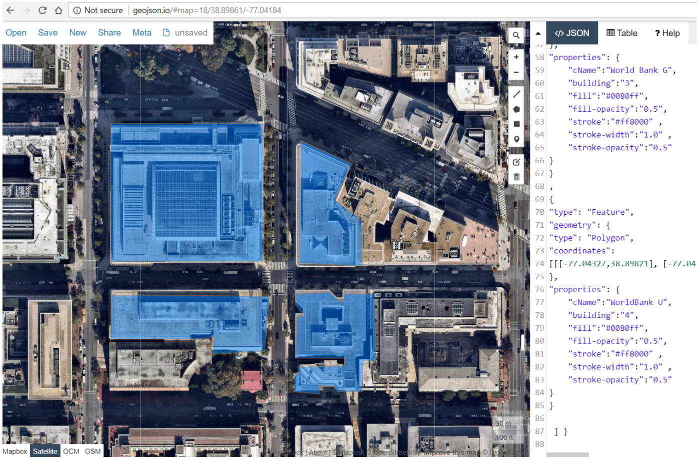
x: -77.0407, y: 38.8988 NUM

as rendered in MapBrowser (free) from VDS technologies:

http://www.vdsgeo.com/mapbrowser.aspx

Output (GeoJSON)

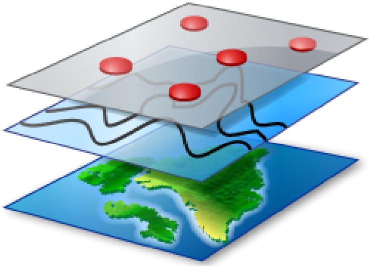
Example of output (polygons layer) to GeoJSON format:



```
"properties": {
    "cName":"World Bank G",
    "building":"3",
    "fill":"#0080ff",
    "fill-opacity":"0.5",
    "stroke":"#ff8000" ,
    "stroke-width":"1.0" ,
    "stroke-opacity":"0.5"
}
}
,
{
"type": "Feature",
"geometry": {
"type": "Polygon",
"coordinates":
[[[-77.04327,38.89821], [-77.04
},
"properties": {
    "cName":"WorldBank U",
    "building":"4",
    "fill":"#0080ff",
    "fill-opacity":"0.5",
    "stroke":"#ff8000" ,
    "stroke-width":"1.0" ,
    "stroke-opacity":"0.5"
}
}

 ] }
```

as rendered in geojsonio (free, online) at: http://www.geojson.io

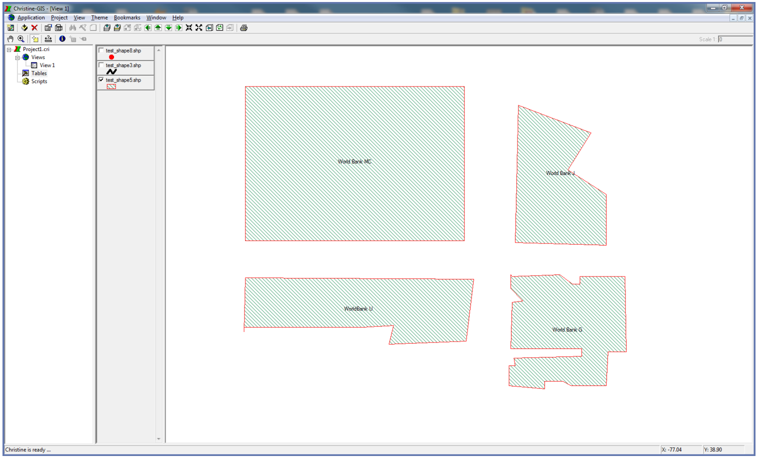
Output (GeoJSON)



Example: create multi-layer files with features of different types

Output (Shapefile)

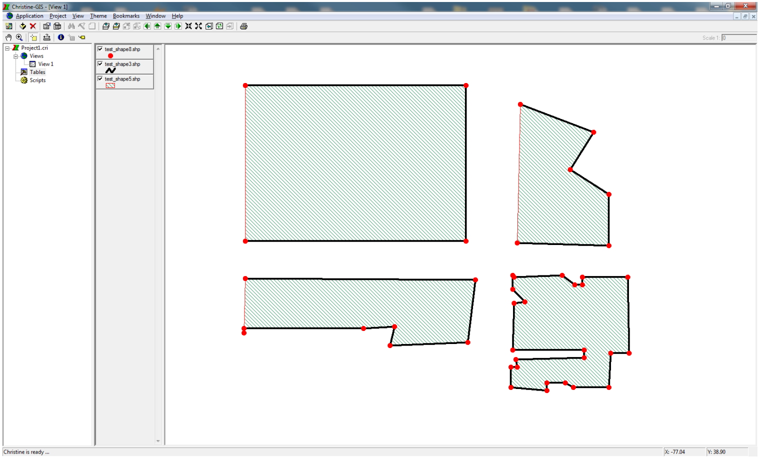
Example of output to a polygon layer of shapefile format:



as rendered in Christine-GIS Viewer (free) from Christine-GIS: http://www.christine-gis.com

Output (Shapefile)

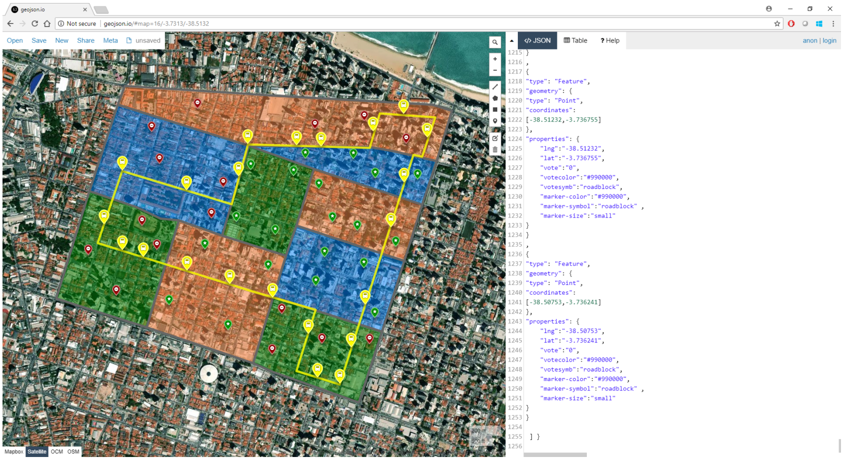
Individual shapefiles may be combined into a multilayer structure in a GIS system (example):



as rendered in Christine-GIS Viewer (free) from Christine-GIS: http://www.christine-gis.com

Output (GeoJSON)

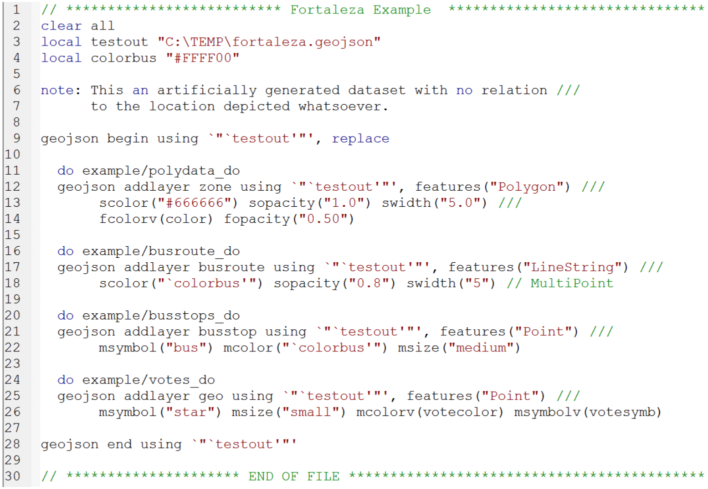
GeoJSON output example with features of various types:



```
}
,
{
"type": "Feature",
"geometry": {
"type": "Point",
"coordinates":
[-38.51232,-3.736755]
},
"properties": {
    "lng":"-38.51232",
    "lat":"-3.736755",
    "vote":"0",
    "votecolor":"#990000",
    "votesymb":"roadblock",
    "marker-color":"#990000",
    "marker-symbol":"roadblock" ,
    "marker-size":"small"
}
}
,
{
"type": "Feature",
"geometry": {
"type": "Point",
"coordinates":
[-38.50753,-3.736241]
},
"properties": {
    "lng":"-38.50753",
    "lat":"-3.736241",
    "vote":"0",
    "votecolor":"#990000",
    "votesymb":"roadblock",
    "marker-color":"#990000",
    "marker-symbol":"roadblock" ,
    "marker-size":"small"
}
}

 ] }
```

as rendered in geojsonio (free, online) at: http://www.geojson.io

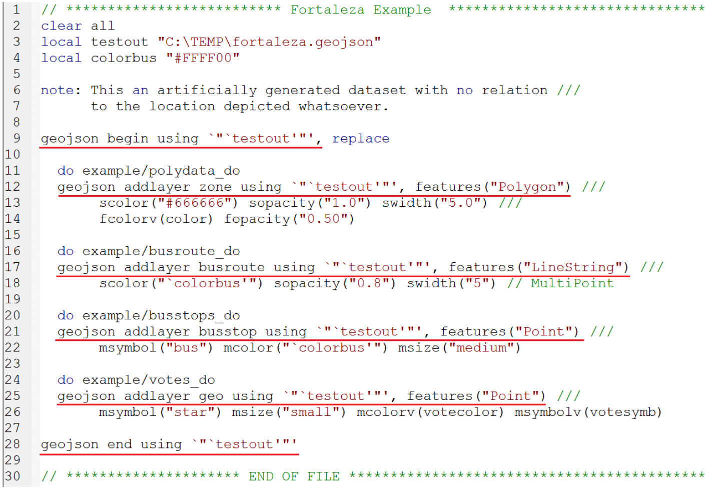
Output (GeoJSON)



```
// ************************** Fortaleza Example  ********************************
clear all
local testout "C:\TEMP\fortaleza.geojson"
local colorbus "#FFFF00"

note: This an artificially generated dataset with no relation ///
      to the location depicted whatsoever.

geojson begin using `"`testout'"', replace

  do example/polydata_do
  geojson addlayer zone using `"`testout'"', features("Polygon") ///
       scolor("#666666") sopacity("1.0") swidth("5.0") ///
       fcolorv(color) fopacity("0.50")

  do example/busroute_do
  geojson addlayer busroute using `"`testout'"', features("LineString") ///
       scolor("`colorbus'") sopacity("0.8") swidth("5") // MultiPoint

  do example/busstops_do
  geojson addlayer busstop using `"`testout'"', features("Point") ///
       msymbol("bus") mcolor("`colorbus'") msize("medium")

  do example/votes_do
  geojson addlayer geo using `"`testout'"', features("Point") ///
       msymbol("star") msize("small") mcolorv(votecolor) msymbolv(votesymb)

geojson end using `"`testout'"'

// ******************** END OF FILE ******************************************
```

Example: outline of the Stata code used to create this image

Output (GeoJSON)

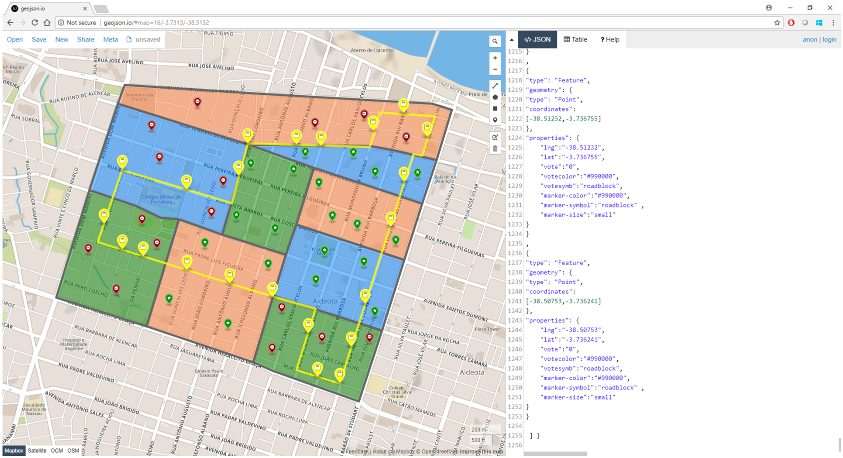


```
// ************************** Fortaleza Example  ********************************
clear all
local testout "C:\TEMP\fortaleza.geojson"
local colorbus "#FFFF00"

note: This an artificially generated dataset with no relation ///
      to the location depicted whatsoever.

geojson begin using `"`testout'"', replace

  do example/polydata_do
  geojson addlayer zone using `"`testout'"', features("Polygon") ///
       scolor("#666666") sopacity("1.0") swidth("5.0") ///
       fcolorv(color) fopacity("0.50")

  do example/busroute_do
  geojson addlayer busroute using `"`testout'"', features("LineString") ///
       scolor("`colorbus'") sopacity("0.8") swidth("5") // MultiPoint

  do example/busstops_do
  geojson addlayer busstop using `"`testout'"', features("Point") ///
       msymbol("bus") mcolor("`colorbus'") msize("medium")

  do example/votes_do
  geojson addlayer geo using `"`testout'"', features("Point") ///
       msymbol("star") msize("small") mcolorv(votecolor) msymbolv(votesymb)

geojson end using `"`testout'"'

// ******************** END OF FILE *******************************************
```

Example: outline of the Stata code used to create this image

Output (GeoJSON)

Online tools allow switching the underlying maps really simply, here MapBox is used:

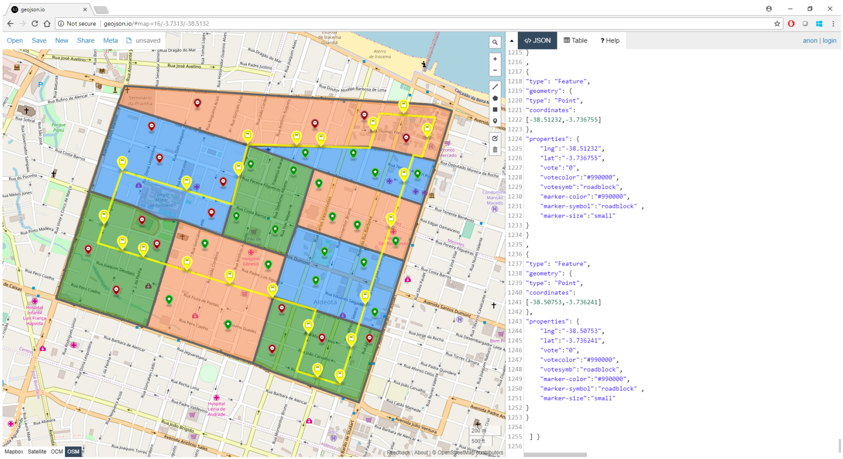

geojson.io
Not secure | geojson.io/#map=16/-3.7313/-38.5132
Open Save New Share Meta unsaved
JSON Table Help
anon | login
Aldeota
Colégio Militar de Fortaleza
Mapbox Satellite OCM OSM

```
}
,
{
"type": "Feature",
"geometry": {
"type": "Point",
"coordinates":
[-38.51232,-3.736755]
},
"properties": {
    "lng":"-38.51232",
    "lat":"-3.736755",
    "vote":"0",
    "votecolor":"#990000",
    "votesymb":"roadblock",
    "marker-color":"#990000",
    "marker-symbol":"roadblock" ,
    "marker-size":"small"
}
}
,
{
"type": "Feature",
"geometry": {
"type": "Point",
"coordinates":
[-38.50753,-3.736241]
},
"properties": {
    "lng":"-38.50753",
    "lat":"-3.736241",
    "vote":"0",
    "votecolor":"#990000",
    "votesymb":"roadblock",
    "marker-color":"#990000",
    "marker-symbol":"roadblock" ,
    "marker-size":"small"
}
}

 ] }
```

as rendered in geojsonio (free, online) at: http://www.geojson.io

Output (GeoJSON)

Online tools allow switching the underlying maps really simply, here OpenStreetMaps is used:



geojson.io
Not secure | geojson.io/#map=16/-3.7313/-38.5132
Open Save New Share Meta unsaved
JSON Table Help
anon | login
Aldeota
Mapbox Satellite OCM OSM

```
}
,
{
"type": "Feature",
"geometry": {
"type": "Point",
"coordinates":
[-38.51232,-3.736755]
},
"properties": {
    "lng":"-38.51232",
    "lat":"-3.736755",
    "vote":"0",
    "votecolor":"#990000",
    "votesymb":"roadblock",
    "marker-color":"#990000",
    "marker-symbol":"roadblock" ,
    "marker-size":"small"
}
}
,
{
"type": "Feature",
"geometry": {
"type": "Point",
"coordinates":
[-38.50753,-3.736241]
},
"properties": {
    "lng":"-38.50753",
    "lat":"-3.736241",
    "vote":"0",
    "votecolor":"#990000",
    "votesymb":"roadblock",
    "marker-color":"#990000",
    "marker-symbol":"roadblock" ,
    "marker-size":"small"
}
}

 ] }
```

as rendered in geojsonio (free, online) at: http://www.geojson.io

Validation

Output files have been verified:

- GeoJSON format – by GeoJSONLint http://geojsonlint.com.
- Shapefiles – by ESRI ArcMap 10.1 build 3035 = ArcGIS for Desktop 10.1 final for Windows.

Most of the viewers that I've tried didn't mind for polygon vertices sequence to be clockwise or counterclockwise when there is only one polygon per feature, but to pass validation and comply with the files' format specifications the **polygon direction detection and reversal** procedures were implemented.

Other notes

- Both Shapefile and GeoJSON formats allows for other types of features, which are not supported by these exporting tools, such as polygons with holes, or lines with breaks;
- Shapefile stores data attributes in a DBF file, which is a DOS-era database format with plenty of its own limitations (on number of variables, length of strings, use of characters, etc); modern GIS systems have taken this format beyond its original specification, for example permitting storing utf-8 unicode characters, but it's up to the individual system to decide how to interpret the data if it deviates from standard ASCII; **geotools** writes additional markers for ESRI ArcGIS and QGIS systems to let them know the content is in utf-8 unicode;
- Stata supports exporting to DBF format from version 15; I have written own DBF writer earlier for this project and keep using it to maintain compatibility with Stata 14;

Other notes

- data is assumed to be unprojected latitude and longitude (and this is how Survey Solutions produces it) as this is the only format officially supported by GeoJSON format;
- Shapefiles may be represented in various projections; *.prj file is created indicating *WGS_1984* projection by this exporter automatically by default; if this is not correct the user should replace this file with a different specification after exporting (it is a text file);
- if you need to do any reprojections of GIS data in Stata, consider using **geo2xy** by Robert Picard.

Custom properties

- All of the notes associated with a Stata dataset in memory are exported in the GeoJSON file as properties of a special null-shape along with some meta-information (production date, Stata version, module name and version);
- All of the variables of the dataset in memory except the geography variable defining features are exported as properties of those features;
    - This may be a good thing, some visualizers automatically pick properties like *title* or *name* as feature names (and show as labels or upon hover/click);
    - This may be a bad thing, e.g. if the *id* variable exists in your data and is not unique, it may confuse the visualizer;
- It is best to write only what you do need, leaving (**keep**-ing) only the features and attributes necessary for your output map.

Future development

- utilize Stata's own DBF writer;
- add controls over formatting of the output data;
- website generation for maps based on GeoJSON format;
- generation of project files for popular GIS systems for combining multiple layers of shapefiles, for example, QGIS (project is an XML file);
- choropleth maps export in GeoJSON and Shapefile formats.
- **fieldarea** and **geochart** (currently separate commands) to become part of the package **geotools**.

Disclaimer

The boundaries, colors, denominations and any other information shown on the maps generated with these programs or contained in this presentation do not imply, on the part of The World Bank Group, any judgment on the legal status of any territory, or any endorsement or acceptance of such boundaries.

2nd disclaimer

Any examples shown in this presentation are to demonstrate the features of the program and not to single out any residence, dwelling, or business.

Homepage

The homepage of GEOTOOLS package is:

http://www.radyakin.org/stata/geotools