

Intro to Programming for Public Policy Week 10

Geographic Information Systems (GIS)

Eric Potash

May 29, 2018

Introduction to GIS

What is GIS?

Software for working with geographic data including creation, analysis, and visualization.

GIS Data

- ▶ There are two basic types of GIS data:
 - ▶ Vector, stored in Geodatabase, Shapefile, KML, etc.
 - ▶ Raster stored in GeoTIFF files, etc.

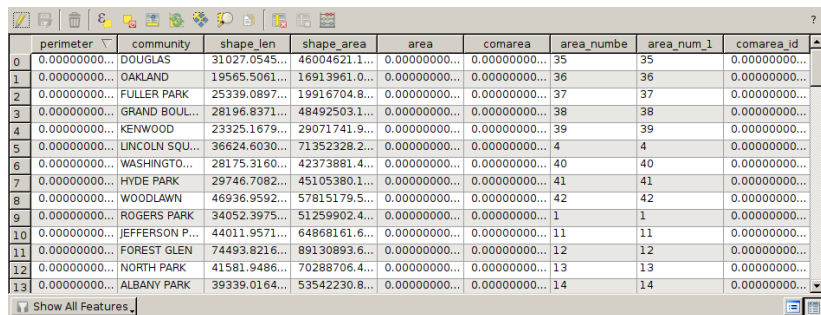
Vector data

Vector data are shapes:

- ▶ Points (e.g. **crime locations**)
- ▶ Lines (e.g. **streets**)
- ▶ Polygons (e.g. **community areas**)

Vector attributes

In addition to storing the shapes (e.g. community areas) themselves, *shapefiles* commonly also have tabular data for each shape.



The screenshot shows a GIS application window with a table of vector attributes. The table has 10 columns: an index column, 'perimeter', 'community', 'shape_len', 'shape_area', 'area', 'comarea', 'area_numbe', 'area_num_1', and 'comarea_id'. The rows represent different community areas, with the first 13 rows visible. The 'community' column contains names like DOUGLAS, OAKLAND, FULLER PARK, GRAND BOUL..., KENWOOD, LINCOLN SQU..., WASHINGTO..., HYDE PARK, WOODLAWN, ROGERS PARK, JEFFERSON P..., FOREST GLEN, NORTH PARK, and ALBANY PARK. The other columns contain numerical values representing various attributes of these shapes.

	perimeter	community	shape_len	shape_area	area	comarea	area_numbe	area_num_1	comarea_id
0	0.00000000...	DOUGLAS	31027.0545...	46004621.1...	0.00000000...	0.00000000...	35	35	0.00000000...
1	0.00000000...	OAKLAND	19565.5061...	16913961.0...	0.00000000...	0.00000000...	36	36	0.00000000...
2	0.00000000...	FULLER PARK	25339.0897...	19916704.8...	0.00000000...	0.00000000...	37	37	0.00000000...
3	0.00000000...	GRAND BOUL...	28196.8371...	48492503.1...	0.00000000...	0.00000000...	38	38	0.00000000...
4	0.00000000...	KENWOOD	23325.1679...	29071741.9...	0.00000000...	0.00000000...	39	39	0.00000000...
5	0.00000000...	LINCOLN SQU...	36624.6030...	71352328.2...	0.00000000...	0.00000000...	4	4	0.00000000...
6	0.00000000...	WASHINGTO...	28175.3160...	42373881.4...	0.00000000...	0.00000000...	40	40	0.00000000...
7	0.00000000...	HYDE PARK	29746.7082...	45105380.1...	0.00000000...	0.00000000...	41	41	0.00000000...
8	0.00000000...	WOODLAWN	46936.9592...	57815179.5...	0.00000000...	0.00000000...	42	42	0.00000000...
9	0.00000000...	ROGERS PARK	34052.3975...	51259902.4...	0.00000000...	0.00000000...	1	1	0.00000000...
10	0.00000000...	JEFFERSON P...	44011.9571...	64868161.6...	0.00000000...	0.00000000...	11	11	0.00000000...
11	0.00000000...	FOREST GLEN	74493.8216...	89130893.6...	0.00000000...	0.00000000...	12	12	0.00000000...
12	0.00000000...	NORTH PARK	41581.9486...	70288706.4...	0.00000000...	0.00000000...	13	13	0.00000000...
13	0.00000000...	ALBANY PARK	39339.0164...	53542230.8...	0.00000000...	0.00000000...	14	14	0.00000000...

Show All Features

Raster data

Raster data store a value for each cell in a geographic grid. The value could represent color, elevation, land use, etc.

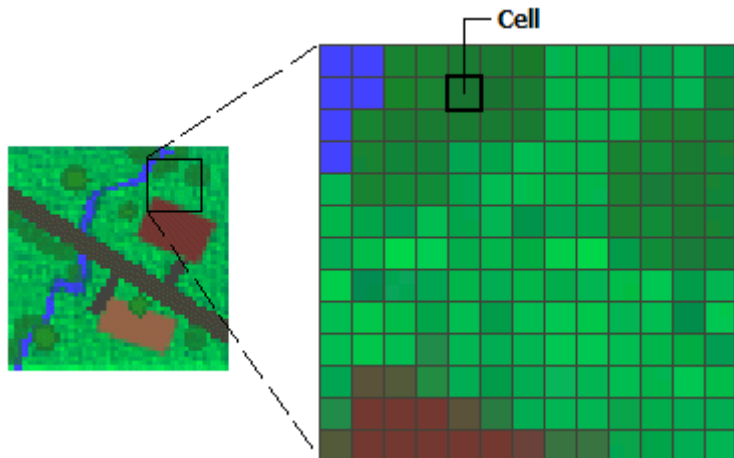


Figure 1: Source: arccgis.com

Layers

A typical GIS problem may involve many kinds of GIS data that are stacked on top of each other.

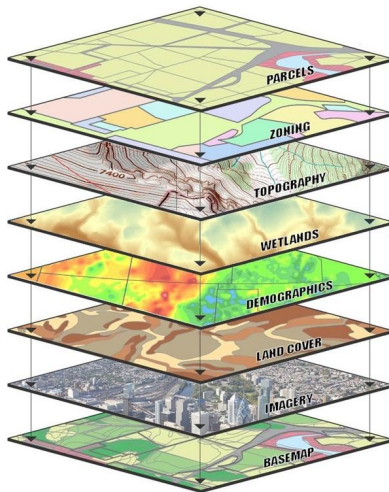


Figure 2: Source: USGS

Latitude-longitude

Chicago center has latitude 41.8781° N and longitude 87.6298° W.
What does that mean?

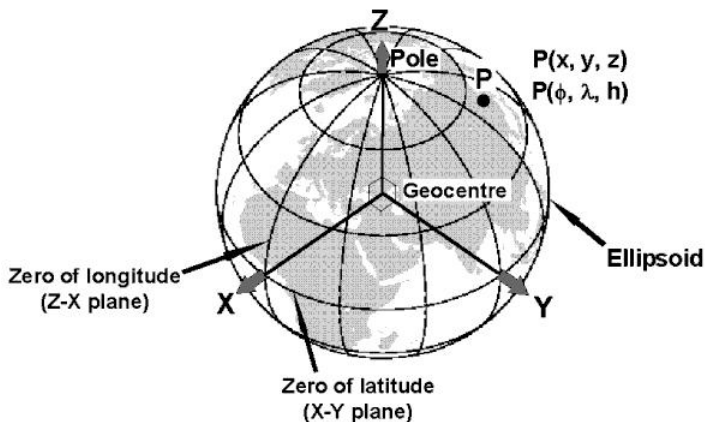


Figure 3: Source: EA4EOZ

WGS84

The above coordinates for Chicago use a *datum* called WGS84 that was developed by the U.S. Department of Defense and used in its Global Positioning Satellites (GPS).

Projections

Latitude/longitude are not a consistent measure, e.g. one degree of latitude is 68.7 miles long at the equator but 69.4 miles at the poles.

Latitude-longitude is a good system for storing data but less good for analyzing and visualizing it. So we use *projections*.

Projection trade-offs

Every projection distorts at least one of the following:

- ▶ Area
- ▶ Direction
- ▶ Shape
- ▶ Distance

Projection types

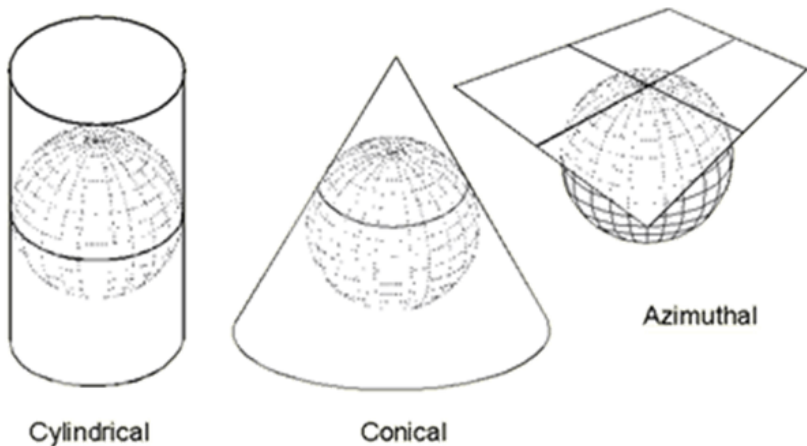
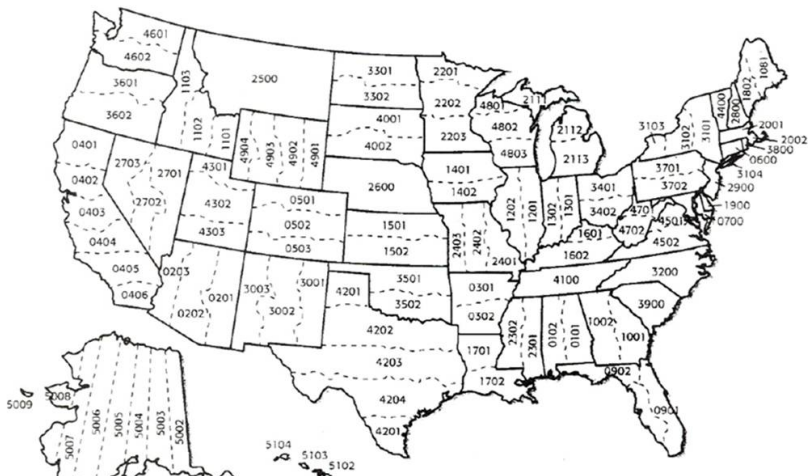


Figure 4: Source: humboldt.edu

Local projections

The *State Plane Coordinate System* is a collection of 124 coordinate systems that cover the U.S.

E.g. Illinois State Plane East, this is the coordinates that the Chicago crime data were in.



Operations

The most common GIS operations are:

- ▶ Distance: the distance between points
- ▶ Length: the length of a line segment (or sequence of lines)
- ▶ Area: the area of a polygon
- ▶ Intersection: the intersection of multiple polygons

Spatial Join

Just like with non-spatial data, it can be useful to join spatial data. Examples of join conditions include:

- ▶ Containment: find the community area containing each crime
- ▶ Distance: find the hospital closest to each accident
- ▶ etc.

GIS Software

Standalone software

- ▶ Desktop
 - ▶ QGIS
 - ▶ ESRI ArcGIS
- ▶ Web
 - ▶ Carto
 - ▶ Mapbox

Homicide rate map

In the next few slides we will use QGIS to make a map of Chicago where the community areas are color-coded by homicide rate.

Homicide rate map: export CSV

First, export homicide rates to csv from python:

```
$ head homicide_rates.csv
Community Area Number,Homicide Rate
1,7.27391755014
2,2.78001723611
3,8.87122529364
4,2.53209429519
5,0.0
6,0.0
7,1.55967309252
8,6.21241488992
9,0.0
```

Homicide rate map: import CSV

- ▶ Select Layer > Add Layer > Add Delimited Text Layer from the QGIS menu.
- ▶ Then select the homicide_rates.csv file
- ▶ Select No geometry (attribute only table)

File Name

Layer name Encoding

File format ☒ CSV (comma separated values) ☐ Custom delimiters ☐ Regular expression delimiter

Record options Number of header lines to discard ☒ First record has field names

Field options ☐ Trim fields ☐ Discard empty fields ☐ Decimal separator is comma

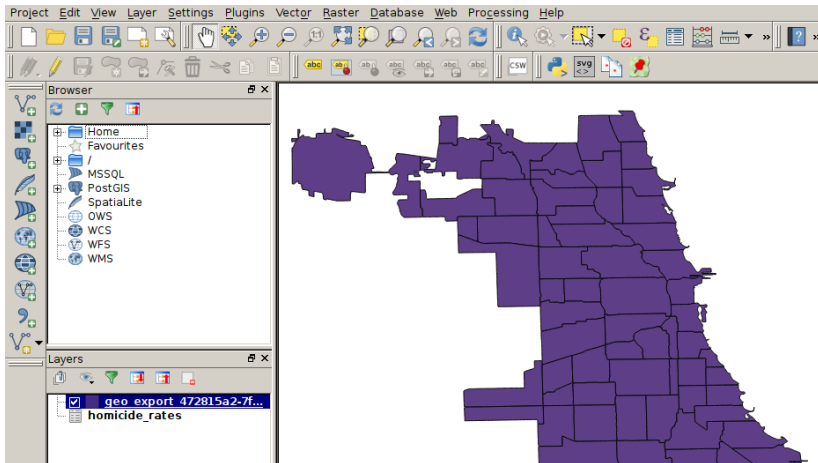
Geometry definition ☐ Point coordinates ☐ Well known text (WKT) ☒ No geometry (attribute only table)

Layer settings ☐ Use spatial index ☐ Use subset index ☐ Watch file

	Community Area Number	Homicide Rate
1	1.0	7.27391755014
2	2.0	2.78001723611
3	3.0	8.87122529364
4	4.0	2.53209429519
5	5.0	0.0
6	6.0	0.0
7	7.0	1.55967309252
8	8.0	6.21241488992
9	9.0	0.0
10	10.0	2.70102368798

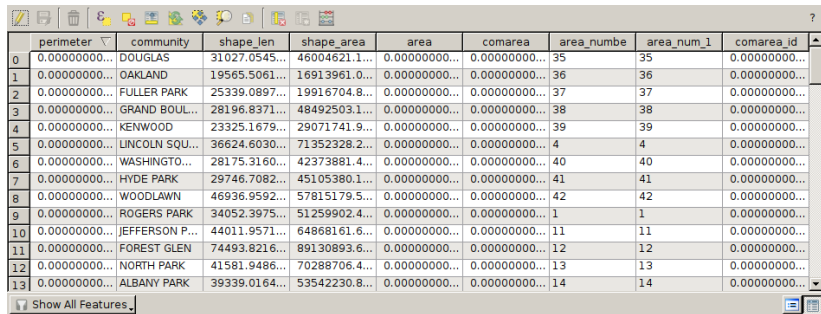
Homicide rate map: import shapefile

- ▶ Select Layer > Add Layer > Add Vector Layer
- ▶ Then select the community areas shapefile (geo_export_472815a2-7f5d-40c8-be47-2f8ab88a4466.shp)
- ▶ After importing, right click the shapefile layer and click Zoom to layer



Homicide rate map: attribute table

Right click the shapefile layer and click Attribute Open Attribute Table to see the community area attributes.

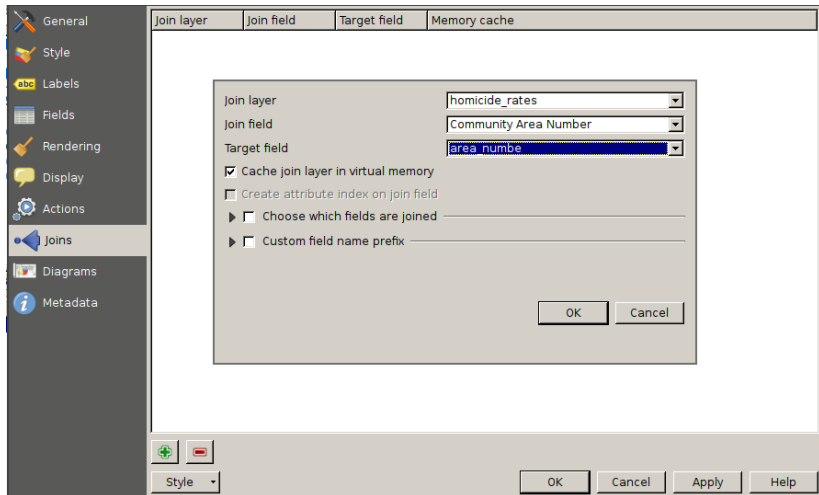


	perimeter	community	shape_len	shape_area	area	comarea	area_numbe	area_num_1	comarea_id
0	0.00000000...	DOUGLAS	31027.0545...	46004621.1...	0.00000000...	0.00000000...	35	35	0.00000000...
1	0.00000000...	OAKLAND	19565.5061...	16913961.0...	0.00000000...	0.00000000...	36	36	0.00000000...
2	0.00000000...	FULLER PARK	25339.0897...	19916704.8...	0.00000000...	0.00000000...	37	37	0.00000000...
3	0.00000000...	GRAND BOUL...	28196.8371...	48492503.1...	0.00000000...	0.00000000...	38	38	0.00000000...
4	0.00000000...	KENWOOD	23325.1679...	29071741.9...	0.00000000...	0.00000000...	39	39	0.00000000...
5	0.00000000...	LINCOLN SQU...	36624.6030...	71352328.2...	0.00000000...	0.00000000...	4	4	0.00000000...
6	0.00000000...	WASHINGTO...	28175.3160...	42373881.4...	0.00000000...	0.00000000...	40	40	0.00000000...
7	0.00000000...	HYDE PARK	29746.7082...	45105380.1...	0.00000000...	0.00000000...	41	41	0.00000000...
8	0.00000000...	WOODLAWN	46936.9592...	57815179.5...	0.00000000...	0.00000000...	42	42	0.00000000...
9	0.00000000...	ROGERS PARK	34052.3975...	51259902.4...	0.00000000...	0.00000000...	1	1	0.00000000...
10	0.00000000...	JEFFERSON P...	44011.9571...	64868161.6...	0.00000000...	0.00000000...	11	11	0.00000000...
11	0.00000000...	FOREST GLEN	74493.8216...	89130893.6...	0.00000000...	0.00000000...	12	12	0.00000000...
12	0.00000000...	NORTH PARK	41581.9486...	70288706.4...	0.00000000...	0.00000000...	13	13	0.00000000...
13	0.00000000...	ALBANY PARK	39339.0164...	53542230.8...	0.00000000...	0.00000000...	14	14	0.00000000...

Show All Features

Homicide rate map: join

- ▶ Right click the shapefile layer and click Properties
- ▶ Click on the Joins panel and click the green +
- ▶ Select homicide_rates as the Join layer, Community Area Number as the Join field, and area_numbe as the Target field.



Homicide rate map: style

- ▶ If you re-open the attribute table, you'll see a new field from the join: homicide_rates_Homicide Rate
- ▶ You can make the style of the map a function of this field in the Properties > Style panel.

The screenshot shows the QGIS Style panel for a map layer. The 'Style' tab is selected. The 'Column' is set to 'homicide_rates_Homicide Rate'. The 'Symbol' is set to 'Graduated'. The 'Color ramp' is set to '[source]'. The 'Legend Format' is set to '%1 - %2'. The 'Classes' are set to 5, and the 'Mode' is set to 'Equal Interval'. The 'Precision' is set to 0. The 'Legend' table is visible, showing the color ramp and the corresponding value ranges.

Symbol	Values	Legend
<input checked="" type="checkbox"/>	0.00 - 23.95	0.0000 - 23.9474
<input checked="" type="checkbox"/>	23.95 - 47.89	23.9474 - 47.8949
<input checked="" type="checkbox"/>	47.89 - 71.84	47.8949 - 71.8423
<input checked="" type="checkbox"/>	71.84 - 95.79	71.8423 - 95.7897
<input checked="" type="checkbox"/>	95.79 - 119.74	95.7897 - 119.7371

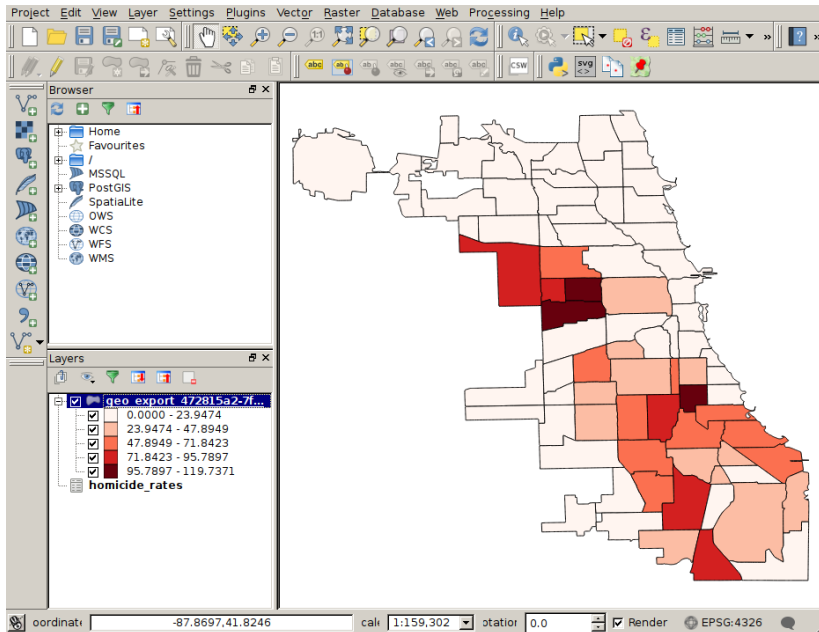
Buttons: Classify, Add class, Delete, Delete all, ☒ Link class boundaries, Advanced

Layer rendering section:

- Layer transparency: 0
- Layer blending mode: Normal
- Feature blending mode: Normal

Buttons: Style, OK, Cancel, Apply, Help

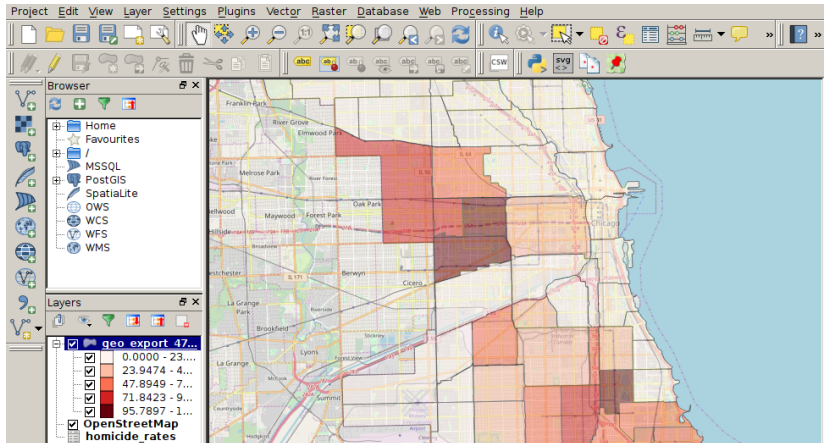
Homicide rate map



Homicide rate map: street layer

It can be useful to add a layer of streets and place names to a map.

- ▶ Install the OpenLayers Plugin from Plugins > Manage and Install Plugins
- ▶ Select Web > OpenLayers plugin and one of the maps (e.g. OpenStreetMap).



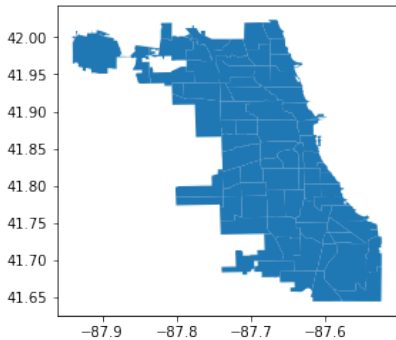
Geopandas

Geopandas is a python module that extends pandas to support geographic data. Geopandas provides two main objects:

- ▶ `GeoSeries` is a series of geographic data (e.g. points, lines, or shapes)
- ▶ `GeoDataFrame` is a `DataFrame` that contains a `GeoSeries`

plot()

```
gdf.plot()
```



dtypes

```
>>> gdf.dtypes
perimeter      float64
community      object
shape_len      float64
shape_area     float64
area           float64
comarea        float64
area_numbe     object
area_num_1     object
comarea_id     float64
geometry       object
dtype:object
```

The community area number columns are text here but numbers in the homicide rate data, so we'll need to convert one to the other before merging.

merge

```
gdf['area_num_1'] = gdf['area_num_1'].astype(int)
gdf_merged = gdf.merge(df,
                        left_on='area_num_1',
                        right_on='Community Area Number')
gdf_merged.plot(column='Homicide Rate', cmap='OrRd')
```

