

API-231 / GIS-PubPol

Meeting 04 (Spatial Analysis and Geoprocessing)

Yuri M. Zhukov
Visiting Associate Professor of Public Policy
Harvard Kennedy School

February 6, 2024

Outline

1. Spatial Analysis

- a) spatial queries
- b) measurement

2. Geoprocessing

- a) joining datasets
- b) Modifiable Areal Unit Problem

Spatial Analysis

Definition: spatial analysis

- any method that uses data on objects'
locations and attributes

Types

- queries
- measurements
- transformations
- spatial joins
- descriptive summaries
- hypothesis tests

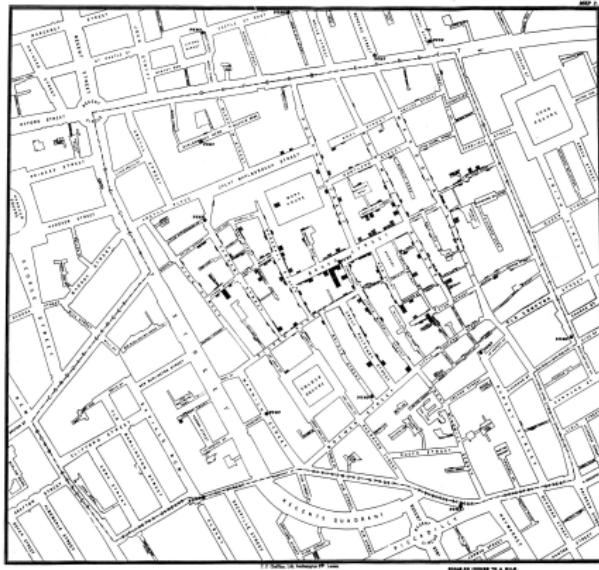


Figure 1: For example, this

Spatial Queries

Spatial Queries

- simplest type of spatial analysis
- focus on descriptive information
- requires no changes to dataset

Examples

- feature identification
("which county is this?")
- feature selection
("which counties are located in the Midwest?")
- selection by attributes
("which are most populous?")
- selection by location
("which are closest to the Canadian border?")

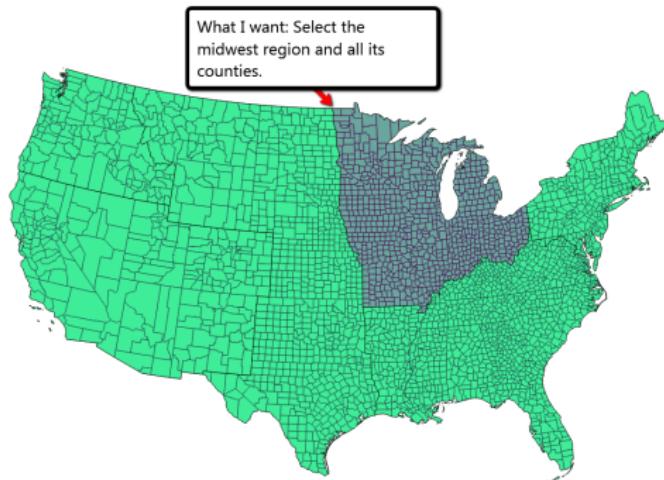
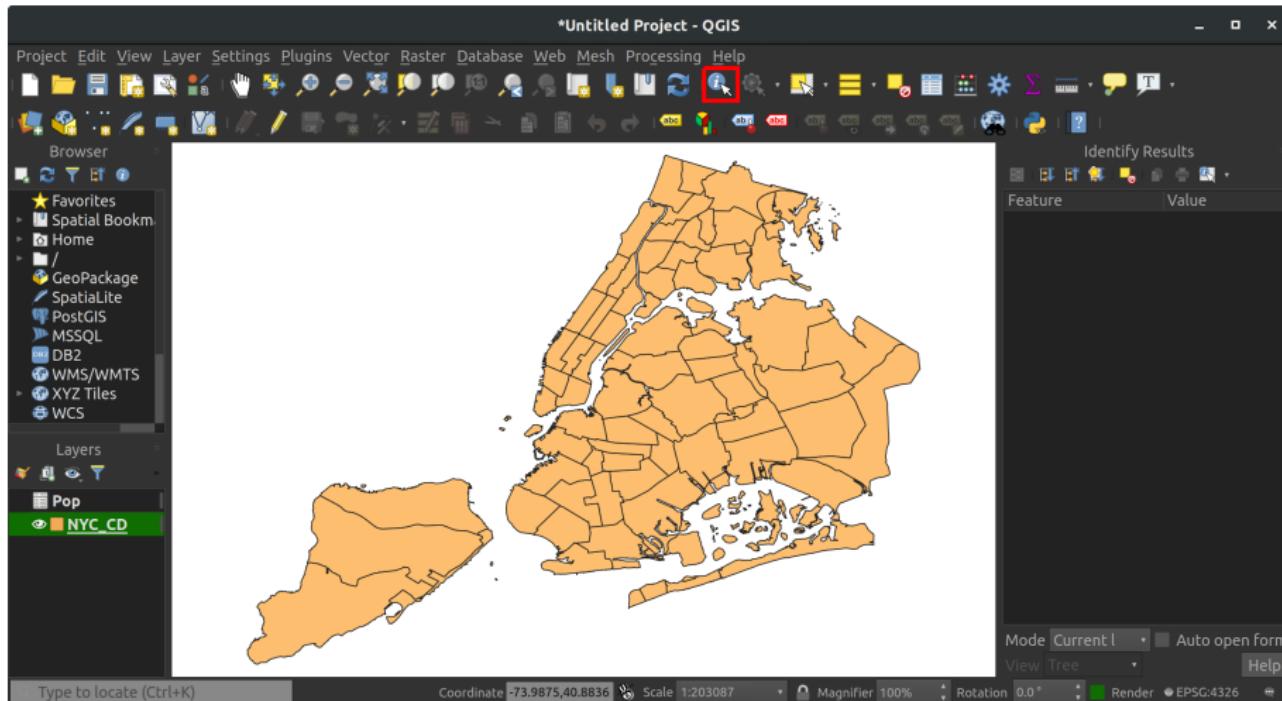
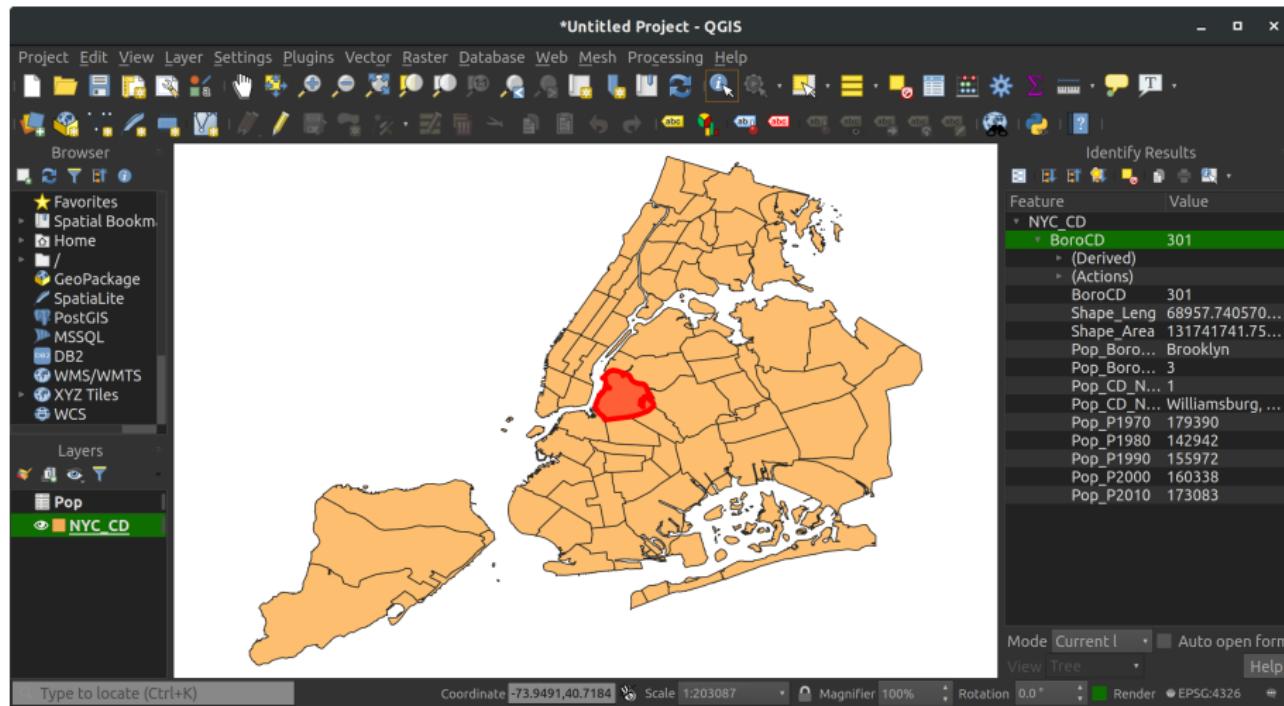


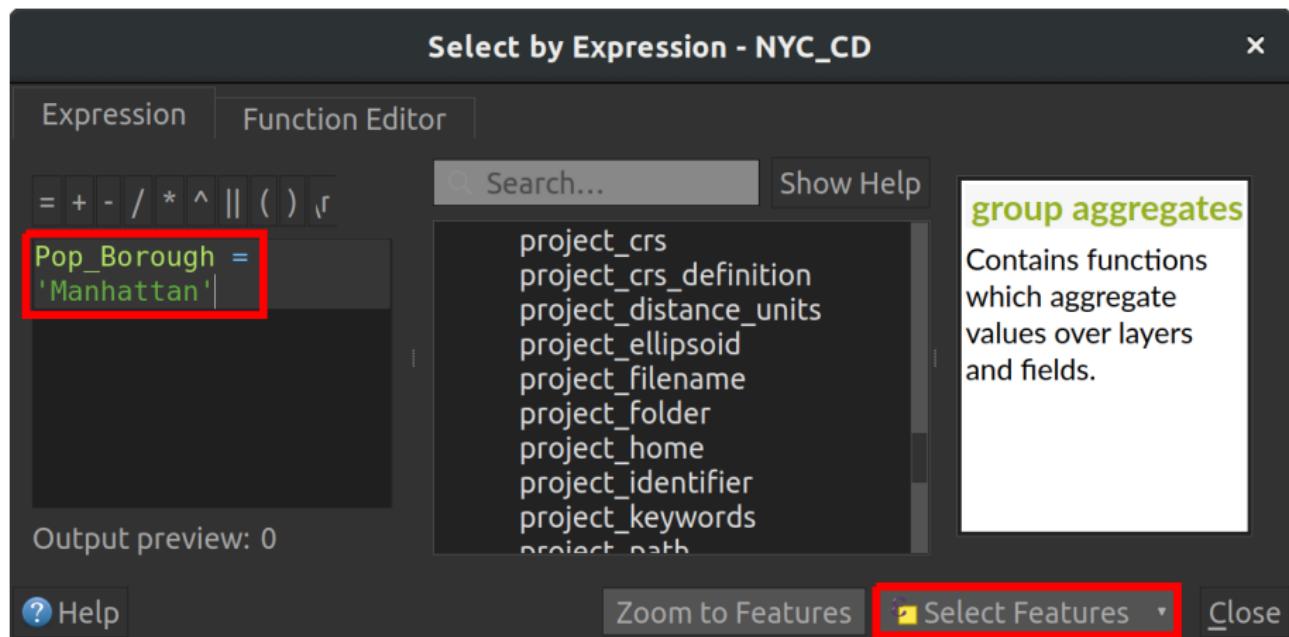
Figure 2: Example of a spatial query



In QGIS, you can do a simple query by clicking the Identify Features button

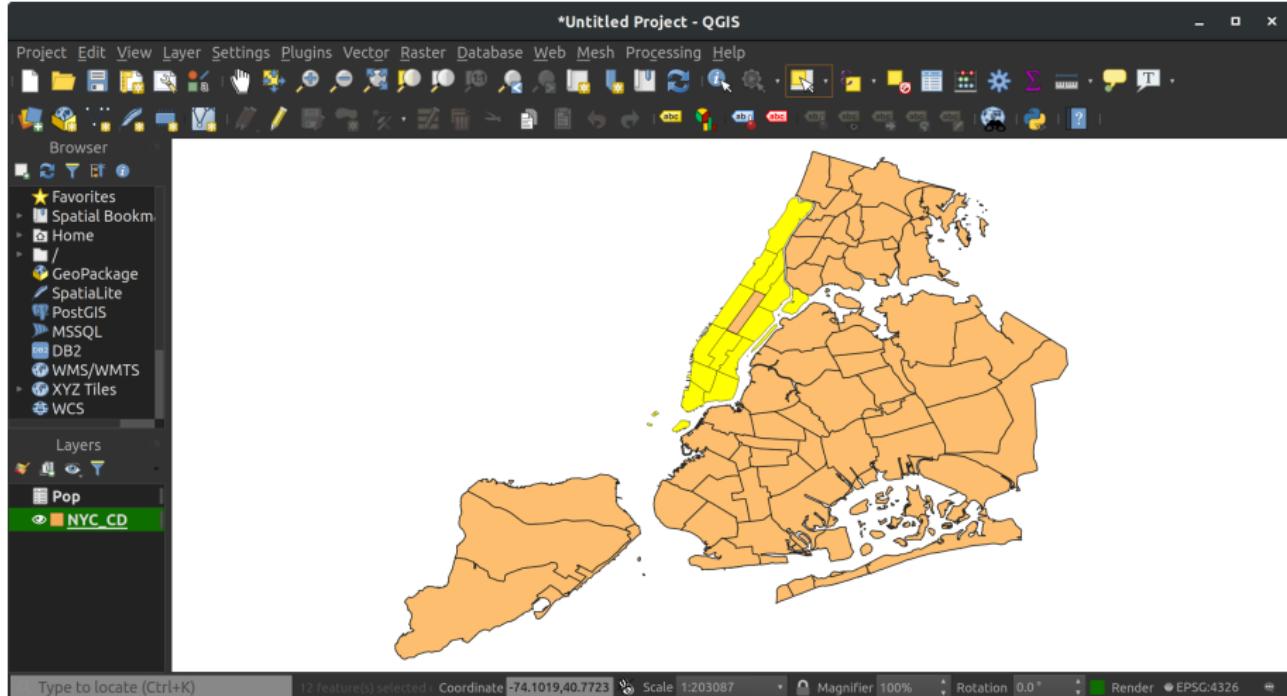


... and clicking on a polygon (or point, line, grid cell)
This polygon represents the Williamsburg, Brooklyn community district

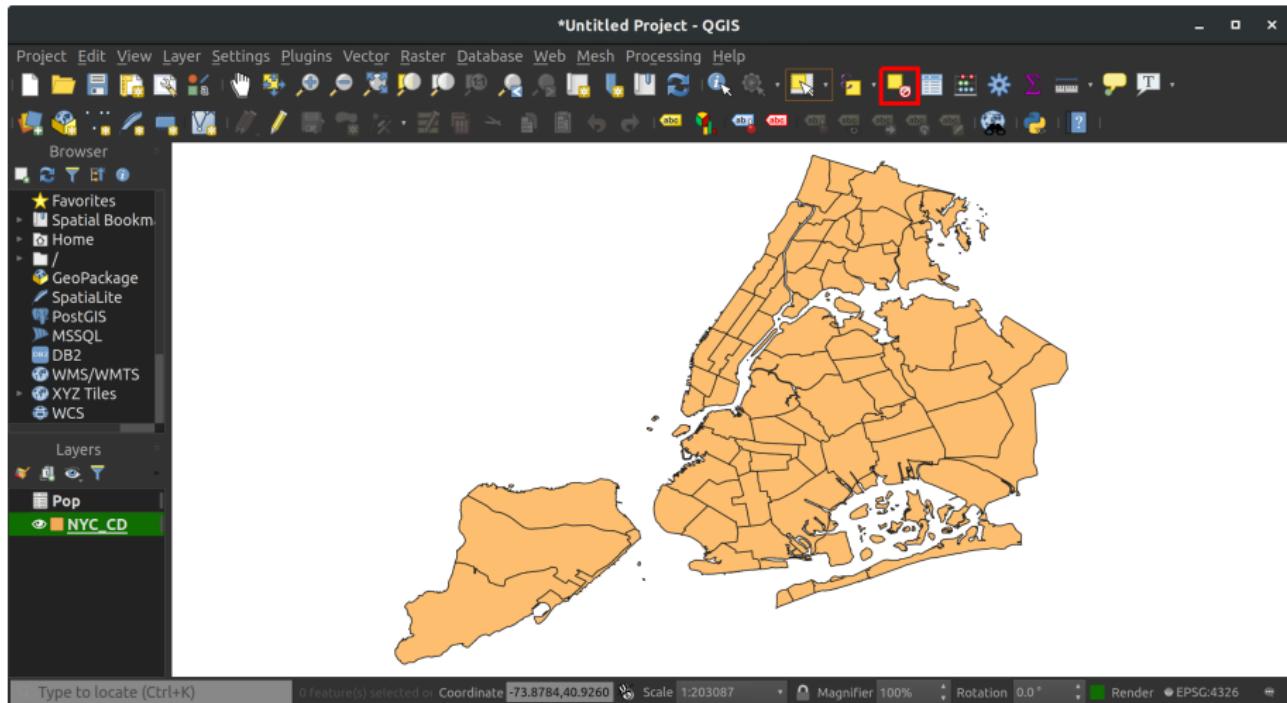


What if we wanted to select all districts in a borough?

- we can do so through the Select Features by Expression tool
(Edit menu → Select submenu → Select Features by Expression...)
- here we are selecting all districts with Pop_Borough = Manhattan



Here are the features we just selected.



Clear the selection by pressing Deselect Features from All Layers button

Measurement

Measurement

- computation of statistics, based on the (relative) locations of features

Examples

- **distance**
("how far is each county from the nearest Army base?")
 - **length**
("how many miles of paved roads are in each county?")
 - **area**
("how large is the jurisdiction for each police precinct?")
 - **perimeter**
("how much coastline?")



Figure 3: Distance between point a and ...

Measures of Length

1. *Euclidean distance* (2-dimensions)
 - straight-line distance between two points on Cartesian plane
 - from Pythagorean Theorem:
$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

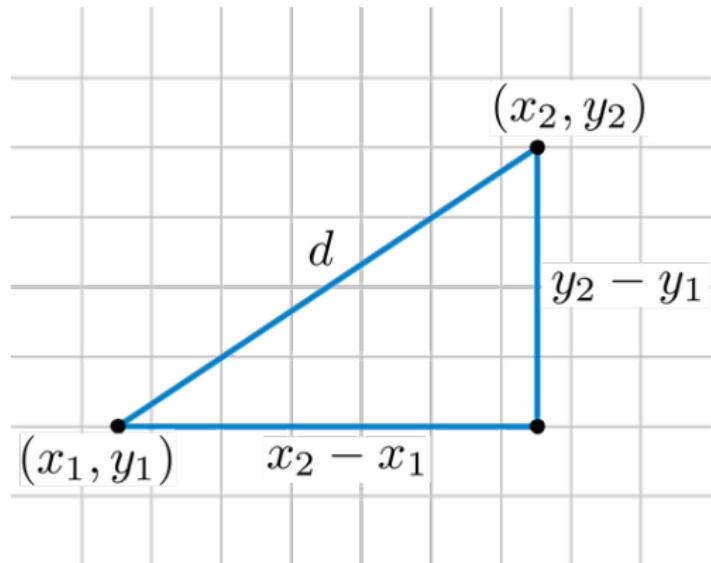


Figure 4: Oldie, but goodie

2. Great Circle distance

- shortest distance between two points on a sphere
- straight lines are replaced by curves (geodesics)
- proportional to central angle

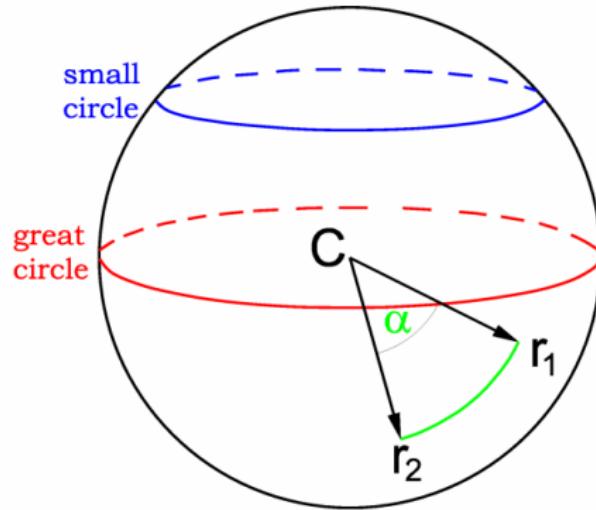


Figure 5: Circles are great

3. Rhumb distance

- arc of constant bearing
- appears as straight line on Mercator projection

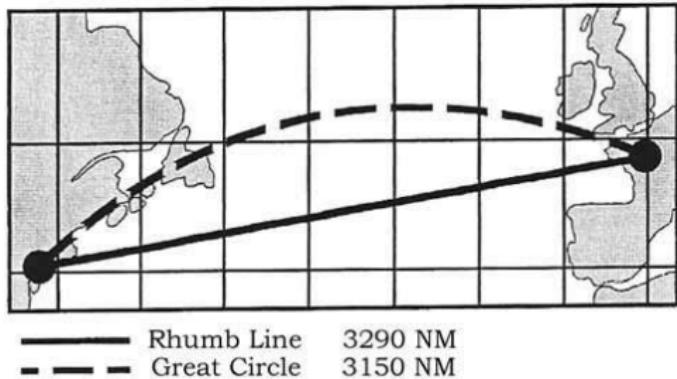
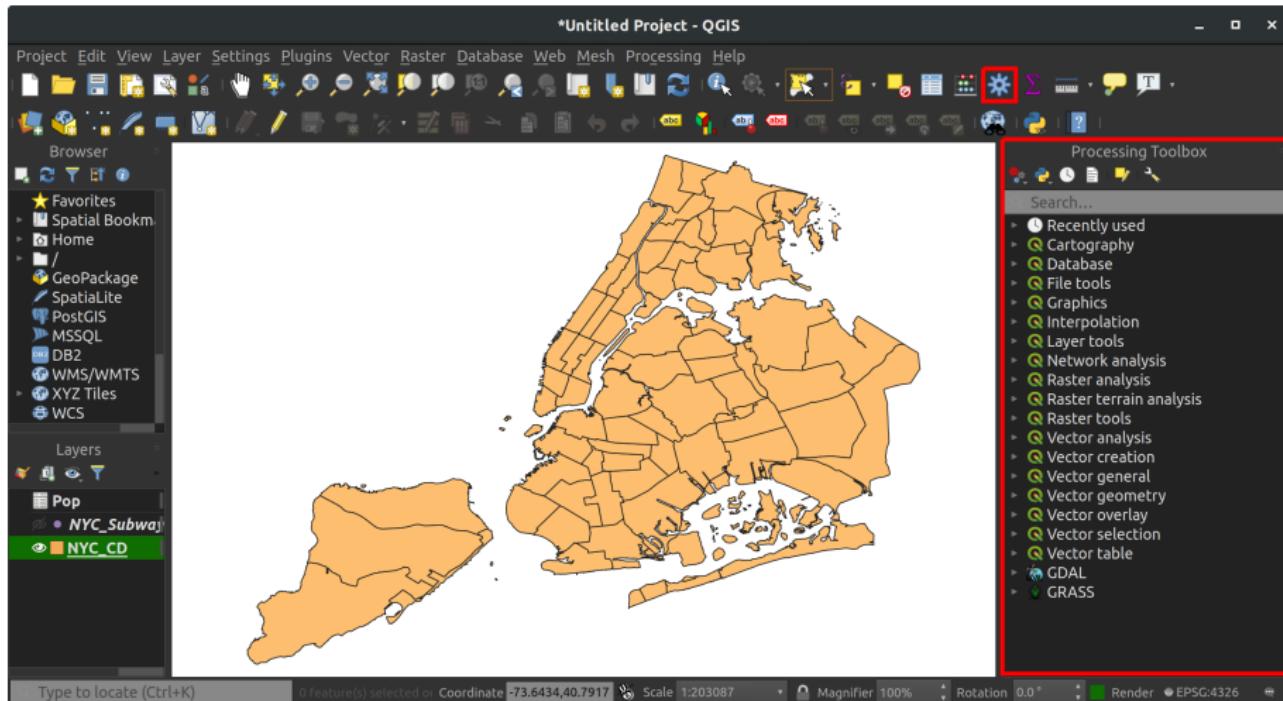
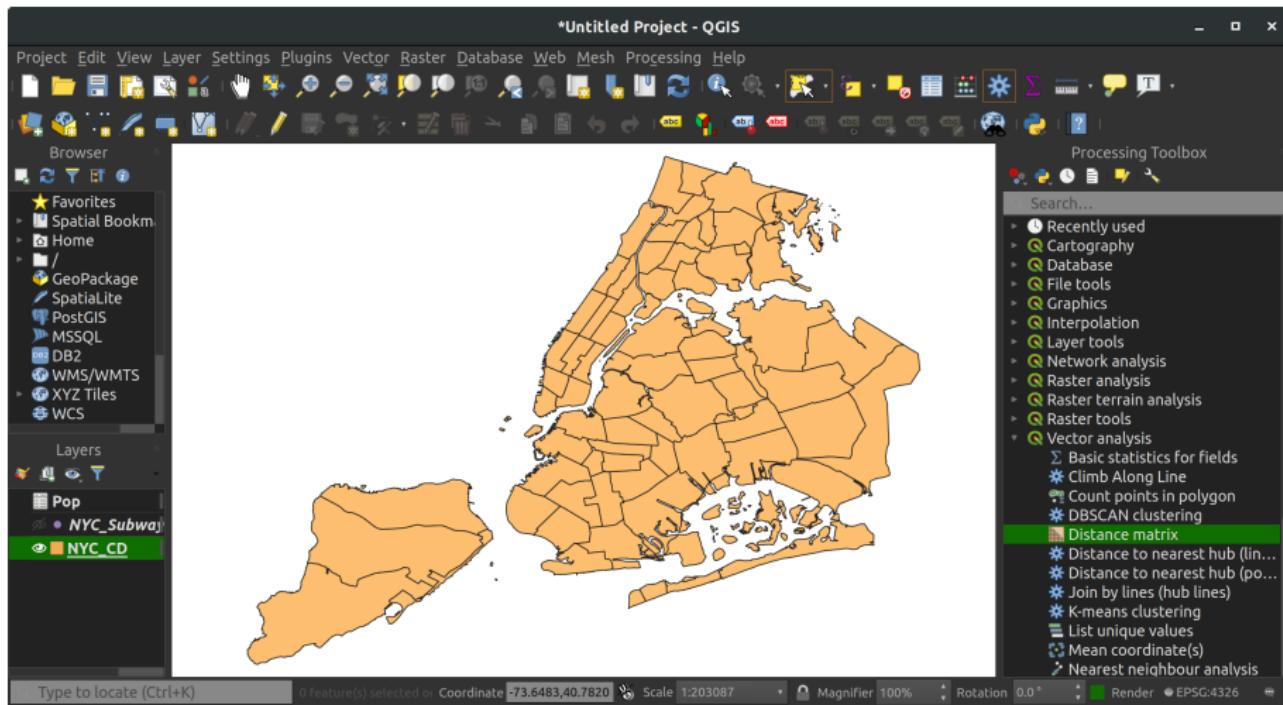


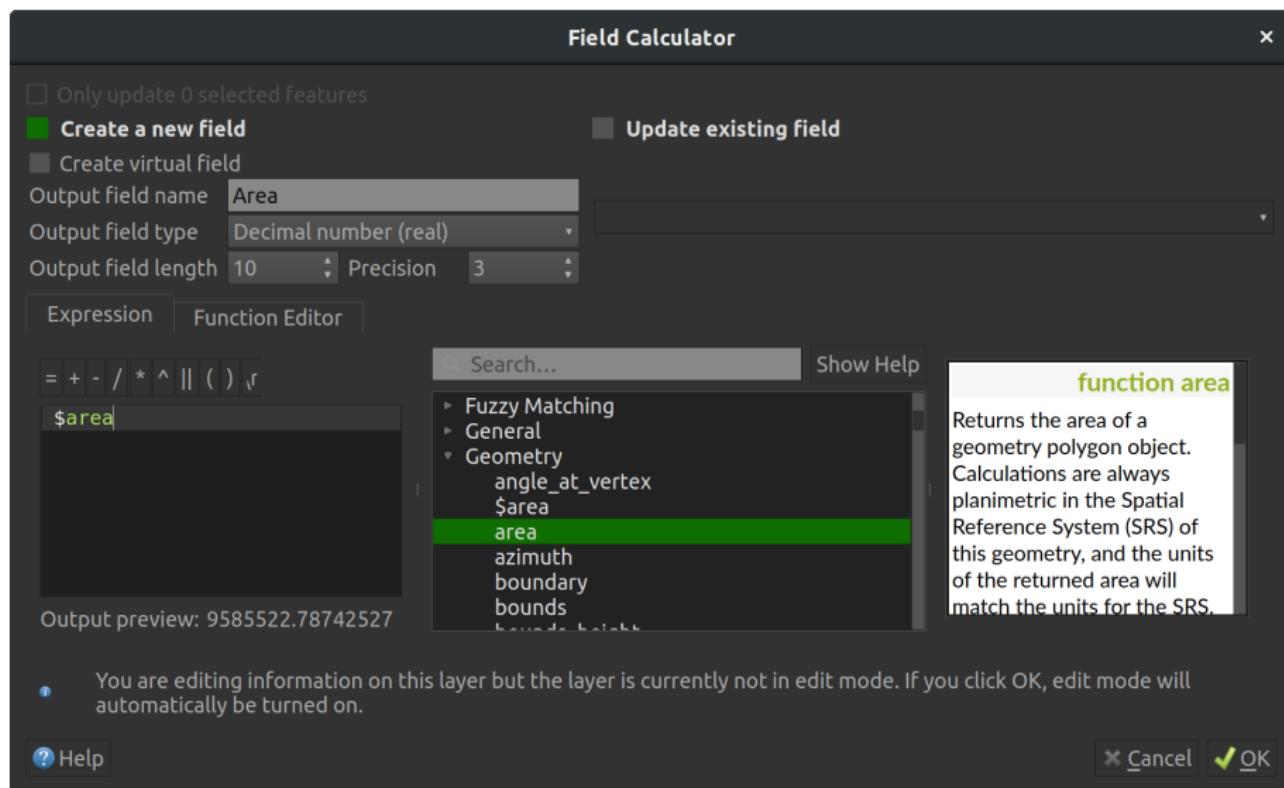
Figure 6: Shortest distance \neq straight line



In QGIS, many of these capabilities are embedded in the Processing Toolbox



For example, here is where you find the tool to create a distance matrix.



... some basic geometry measurements are also in the Field Calculator

Measures of Length

- measurements depend on map projection, distance type
- true length of curve > length of line or perimeter of polygon
- estimated length & area on 2D projection (usually) < true length and area on 3D surface

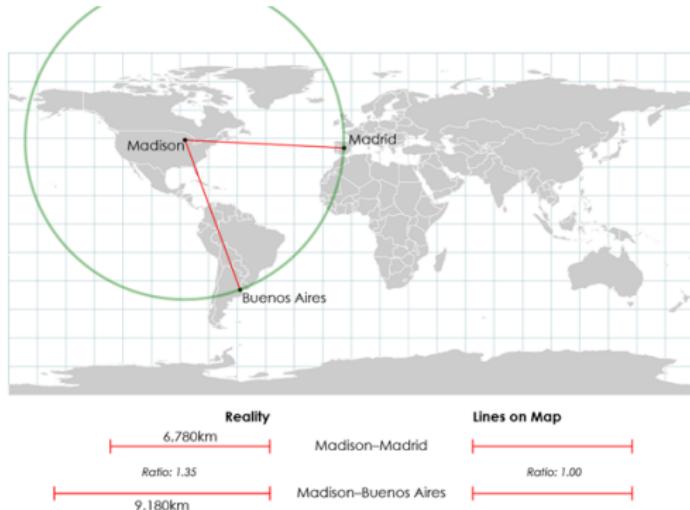


Figure 7: Why projections matter

Geoprocessing

Transformations

Transformations

- creation of new spatial objects, based on locations, shapes and attributes of existing objects

Examples

1. *Point-in-polygon*

("how many crimes in each police precinct?")

- input: points + polygons
- output: polygons, with new attribute (e.g. 'nCrimes')

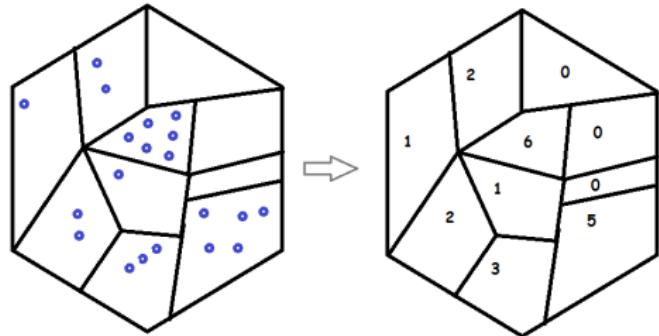


Figure 8: Point-in-polygon

2. *Line-in-polygon*

("how many miles of paved road in each district?")

- input: polylines + polygons
- output: polygons, with new attribute (e.g. 'pvdroadlength')

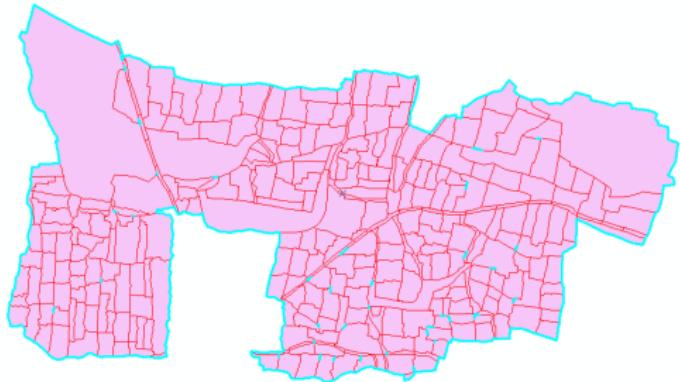


Figure 9: Line-in-polygon

3. *Buffers*

("which areas are within 5 miles of toxic waste site?")

- input: points + buffer distance
- or lines + buffer distance
- or polygons + buffer distance
- output: polygons

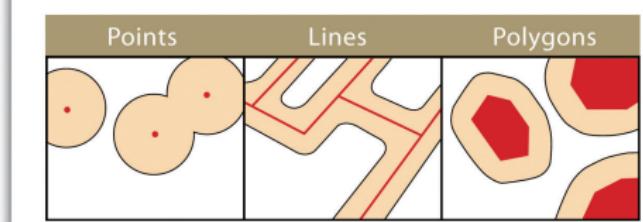


Figure 10: Buffers

3. Kernel density

("where are crime hotspots?")

- input: points
- output: raster, where cell values are estimated local density of points

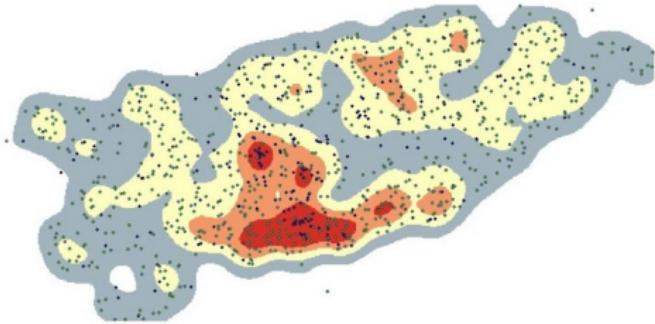


Figure 11: Kernel density

Point-in-polygon transformation

- whether a given point lies inside/outside of a polygon

Examples

1. Generalization

- calculate number of points in each polygon
- can be broken down by type of point (e.g. violent vs. non-violence crimes)

2. Assignment

- assign attributes of polygon to overlapping points ("in which precinct did a particular crime occur?")

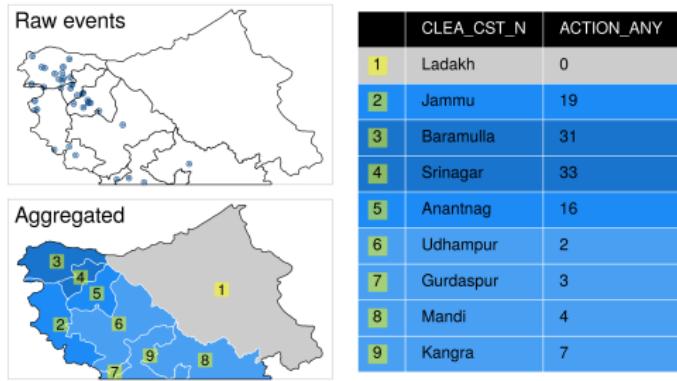


Figure 12: Generalization example

Joining Datasets

1. Joining by attribute

- add data in tabular format (e.g. .dbf, .xls, .csv) to the attributes of a geographic layer (e.g., polygon, line, or point)
- requires an attribute/field common to both tables ("key")

The diagram illustrates the process of joining two datasets based on a common key field. It shows two tables: the Target Layer and the Join Layer, with arrows indicating the fields to be matched.

Target Layer:

Type	Key
Apartment	A
House	B
Commercial	C

Join Layer:

Building_Type_ID	Building_Name	Occupants
A	Silverbirch Estates	130
A	Pine Ridge	8
A	Lake View	250
B	Yellow	3

Join one to one - First record (default):

Type	Key	Building_Name	Occupants
Apartment	A	Silverbirch Estates	130
House	B	Yellow	3

Join one to one - Order by Occupants (Largest):

Type	Key	Join Count	Building_Name	Occupants
Apartment	A	3	Lake View	250
House	B	1	Yellow	3

Join one to one - Add statistics Sum of Occupants

Type	Key	Join Count	Sum
Apartment	A	3	388
House	B	1	3

Join one to many:

Type	Key	Building_Name	Occupants
Apartment	A	Silverbirch Estates	130
Apartment	A	Pine Ridge	8
Apartment	A	Lake View	250
House	B	Yellow	3

Figure 13: Join-by-attribute example

2. *Joining by location*

- add data from attributes of one geographic layer, to the attributes of another geographic layer
- join based on relative location, rather than common values in attribute table

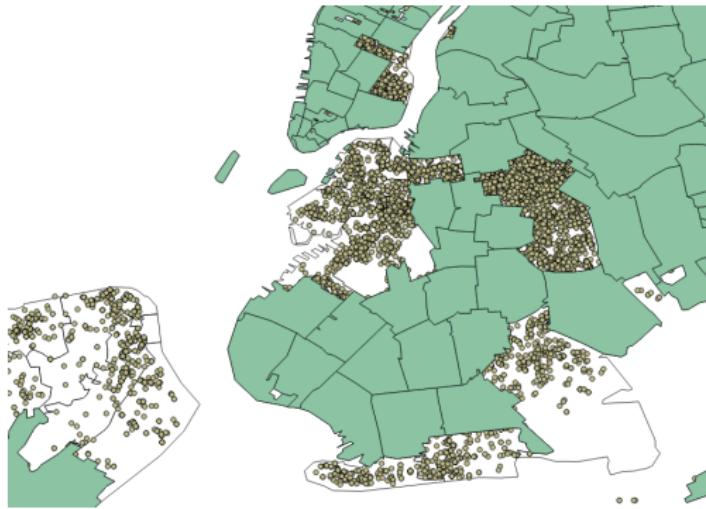


Figure 14: Join-by-location example

Modifiable Areal Unit Problem

Modifiable Areal Unit Problem

- source of statistical bias
- occurs when point-based measures (e.g. events, people) are aggregated into zonal units (e.g. districts, countries)
- number, size, shape, precision of zonal units affect results
- (often) no objective criteria for selecting units
- different boundaries → different distributions

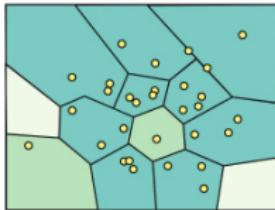
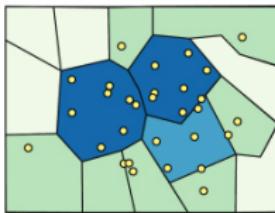
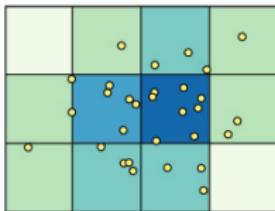


Figure 15: MAUP

Example

- legislative redistricting is a Modifiable Areal Unit Problem

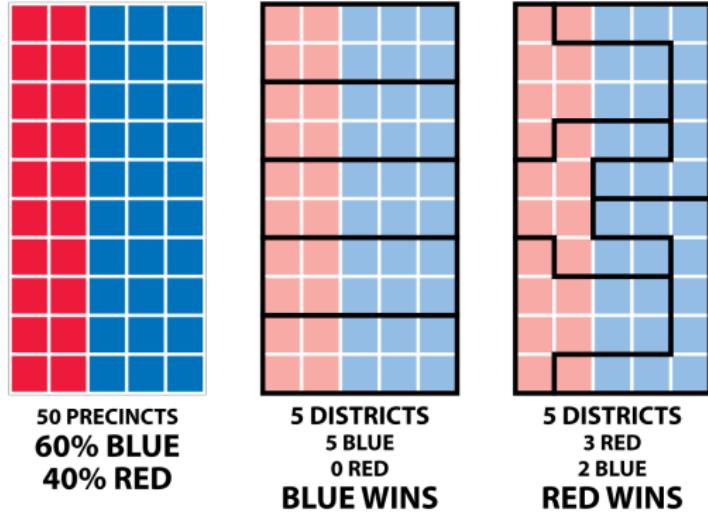


Figure 16: Which map is best?