

advanced qgis

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Tuesday 1st September, 2020 18:04

outline

geocoding

SQL

spatial join

>probably finish first part here>

geo-processing

outline

geocoding

SQL

spatial join

>probably finish first part here>

geo-processing

geocoding: address \rightarrow (lat,lon)

- say that we have some addresses and we want to geocode them
- https:
//sites.google.com/site/adamokuliczkozaryn/
gis_int/apartments-for-rent.xls
- open, looks reasonably clean, save as csv

MMQGIS-Geocode

- MMQGIS-Geocode-Geocode CSV with Web Service
- Input CSV, and make sure Address Field, City Field, State Field are right; best if you give more info
- Web Service: OpenStreetMap/Nominatim
- make sure notfound.csv (and output shp) saved where you can write!
- ref: <https://mangomap.com/blog/how-to-make-a-web-map-from-a-list-of-addresses-in-a-spreadsheet/>
-
- note as of qgis 3.5, it seems to work fine to have everything just under address!
-
- btw, if already got X/Y lat/lon:

important to check!

- add basemap to check location
 - does it make sense? houses in river or park?
 - zoom-in to street, click some points with “identify tool”: pop-up address—does it match with the street?
 - usually some miscodings, say few percent
 - usually because the address is misspelled or incomplete
- do see notfound.csv: mostly those with a range of street numbers
- need to fix them/adjust them:
 - to check can just google them and see if you get a clean hit, and adjust accordingly to get it clean, and then change in csv

outline

geocoding

SQL

spatial join

>probably finish first part here>

geo-processing

SQL: Structured Query Language

- full blown (not in qgis) SQL is only little more complicated
 - very much English-like, just with some strict syntax rules
- also a job market skill: put 'basic SQL' on your linkedIN next to 'gis' skill
 - very easy to master in no time
- <https://www.youtube.com/watch?v=jJeae7PJVv4>

advanced filter (expression): SQL

- nj_counties-Open Attribute Table
- bottom left box- “Advanced Filter (Expression)”
 - Fields and Values “REGION”
 - and on the right Load values: “all unique”
 - then we can type
- “REGION” = 'CENTRAL' and hit OK
 - now easy to modify at the bottom of table, say:
- “REGION” = 'CENTRAL' OR “REGION” = 'SOUTHERN'
- “REGION” = 'CENTRAL' AND “POP2010” > 598349

regular expressions

- can also match part of a string:
- `regexp_match("COUNTY", 'C.*N')`
- `regexp_match("COUNTY", '^C.*N')` must start with 'C'
- `regexp_match("COUNTY", '^C.*N$')` and end with 'N'
- then can hit ctrl-a to select all data
- right click layer, save as (check "selection")

saving selection often necessary

- keep in mind simplicity principle!
- drop all unnecessary clutter
- do not map things that you don't care about

outline

geocoding

SQL

spatial join

>probably finish first part here>

geo-processing

doing it commonsensically

- you can actually spatial join with regular join we've covered
- the idea is that you have non-matching geographical levels
- say hospitals in excel and zip-codes in shapefile,
- you want to map sum of patients in hospitals per zipcode
- you can do it in qgis (next slides)
- but you can do it by hand:
 - use stata, excel, sas, spss, etc
 - just add patients within each zipcode and
 - merge zipcode patient sums with gis file at zipcode level

a proper spatial merge

- as above: things do not fit geographically...
- say zip codes in one data, and counties in another data
- can map both and merge based on location
- so called “spatial join”
- have to pick: mean, sum, or first

join counties with universities

- `nj_counties` `https://docs.google.com/uc?id=1xJDhcRCkgv7k4tNCa720og5bohV6dTB2&export=download`
- `and universities`
`https://sites.google.com/site/adamokuliczkozaryn/gis_int/hsip_colleges.zip?attredirects=0&d=1`
- `and first make ENROLL numeric: text-to-float`

thinking

- as always, think what your are doing and what does it mean
 - and double check
- here there are some institutions with 0 enrollment
 - they were missing ("") before the text-to-float
 - ideally, you should find out what these enrollments are:
eg call the institution

dropping cases

- universities-Open Attribute Table
- select features using an expression
 - "ENROLL" > 0
- layer-Export selected features as

MMQGIS-Combine-Spatial Join

- Output: nj counties
- Spatial Operator: Contains
- Data (Join) Layer: universities
- Attribute Operation: Sum
- Fields: COUNTY_LAB AND ENROLL
- (make sure ENROLL is float)
- note: it often crashes, yet it saves the joined layer
 - so if crashed try opening first the saved layer :)

double check

- double check: say Atlantic has 2 and sum of 13,880
- use identify tool: $7035 + 6845 = 13,880$
- click on ENROLL col header to sort and we see that
 - Essex County wins with COUNT of 9 and ENROLL of 65k
- now could do graduated map of ENROLL for counties

more about spatial join

- matching census tracts with towns <http://trendct.org/2015/05/29/tutorial-how-to-merge-data-from-two-different-maps-using-qgis/>

outline

geocoding

SQL

spatial join

>probably finish first part here>

geo-processing

outline

geocoding

SQL

spatial join

>probably finish first part here>

geo-processing

this is a whole bag of tools

- we switch gears a little and discuss
 - more advanced topics beyond mapping
 - more like typical GIS/IT stuff
- we will just cover few tools
- there are dozens of them
- you may present some of those for extra credit
 - let me know— some may not be useful for this class
- those that i think are especially useful are covered below
- most are under 'Vector', and also 'Plugins'

dissolve

- nj_counties <https://docs.google.com/uc?id=1xJDhcRCkgv7k4tNCa720og5bohV6dTB2&export=download>
- dissolve into a larger area
 - (get rid of inside boundaries)
- Vector-Geoprocessing Tools-Dissolve
 - nj_counties
- uncheck “Dissolve all”
 - otherwise it will dissolve all
- “dissolve field:” REGION

dissolve your way

- can dissolve into your own categories/definitions
- let's take regions and dissolve into south and north jersey
- Open attribute table-toggle editing-New column-integer:
'southNorth'
- sort on REGION and mark southern regions with 1, and the rest with 0
- may also highlight the row to see which county is where
- Vector-Geoprocessing tools-Dissolve
- "Dissolve field:" southNorth
- often you will have to do something like this
- no way to find a shapefile for South Jersey online!

simplify polygons

- remember from principles: simplify as much as possible
- simplifying polygons means dropping vertexes, so that polygons are defined by fewer coordinates **draw**
- it reduces size of a file
- Vector-Geometry tools-Simplify geometries
 - Input: 'nj_counties'
- play with “tolerance” to achieve desired simplicity
 - try 1000—turn off/on to compare to original: see the difference?
 - can also simplify lines (fewer nodes)
- and i guess you can also simplify points (fewer dec points)

○ reference <http://gis.stackexchange.com/questions/25914/>

centroids

- calculate a center of a polygon or turn polygon into a point
 - useful when merging non-overlapping polygons—say congressional districts and counties
 - then you can calculate centroid of one of those and merge with polygons of the other layer if a centroid is in that polygon using spatial merge
- draw a picture
- Vector-Geometry tools-Polygon centroids
 - Input: nj counties

centroids

- note: the new shapefile will have the same data
- can now map another variable and overlay on another variable
- can map both points and polygons with some symbology
- let's map population for polygons
 - and population density for points
 - note: make points bigger to see symbology well
- this solves the problem of showing 2 vars in one map

buffering

- kind of opposite of centroids:
- buffer (circle) around a point or poly or line; eg:
 - 'dry zone' around schools
 - waste processing plants and houses
 - 2-mile heavy pollution around hwy
 - walkability to healthy stores, etc
- load nj universities
 - https://sites.google.com/site/adamokuliczkozaryn/gis_int/hsip_colleges.zip
- Vector-Geoprocessing Tools-Fixed Distance Buffer
- use 20,000 feet (buffer size is in map units)
- Properties-Metadata or even -General: unit is US ft
- note: buffer is a new layer and then can spatially merge it with another layer

measure line tool

- note different measuring units
- measure distances—how far from a point to point
- measure radius to make sure it is right
- and measure say jogging route in segments from RU to 8th and market
- note we will do the same in google maps

example: environmental problems around univ

- download and add to qgis
- https://docs.google.com/uc?id=1T_n1y_Mj5yQiWpZwrbuuFFwmIVJ2QWFZ&export=download
- make smaller, say size of .4 so can better see

MMQGIS-Combine-Spatial Join

- Output: Buffer
- Spatial Operator: Contains
- Data (Join) Layer: NJ contaminated sites
- Attribute Operation: Sum
- Fields: NAME

investigate

- open attr table of merged shapefile
- go to last column 'COUNT' and click 2x to sort descending
- under 'NAME' we find that 'NEW JERSEY MEDICAL SCHOOL'
 - has biggest problem! over thousand contaminated sites
- select say 3 rows at top
- click at the top 'zoom map to selected features'
 - a lot of overlap there
- but from the table can select schools with greatest problems
 - and take some measures to help with the situation

related: select by location (say id problematic ones)

- say select polluted sites within 1000 ft from a school
- Vector-Geoprocessing Tools-Fixed Distance Buffer
- Vector-Research Tools-Select by location
- Layer to select from: NJ Contaminated Sites
- Additional layer (intersection layer): Buffer
- Gemetric predicate: within
- and then: NJ contaminated sites-Save As
 - check 'Save only selected features'
 - and save as csv
 - got 80 places we can call and ask to clean up