# advanced qgis

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Friday 23$^{rd}$ March, 2018    00:42

**<u>outline</u>**

geocoding

SQL

spatial join

probably finish first part here

geo-processing

Extra/bonus

example: apts close to episcopal church in Philly

example: hwys in NJ

**extra credit opportunities; present:**

◇something we did not cover (has to be GIS, of course)

◇alternative way of doing something that we have covered

## **outline**

geocoding

SQL

spatial join

probably finish first part here

geo-processing

Extra/bonus

example: apts close to episcopal church in Philly

example: hwys in NJ

**geocoding: address→(lat,lon)**

◇ let's 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, and for simplicity just keep first 10!

◇ looks reasonably clean, and save as csv

### MMQGIS-Geocode

· MMQGIS-Geocode-Geocode CSV with Google/OpenStreetMap

◇ it works better if you specify more information

◇ make sure Address Field, City Field, State Field are right

· make sure notfound.csv is saved where you want

◇ let's hit ok, it takes like 10sec

◇ https://mangomap.com/blog/
how-to-make-a-web-map-from-a-list-of-addresses-in-a-spreadsheet/

· if goog complains, try the other one, or get goog API key, cheap

·

◇ btw, if already got X/Y lat/lon:
just add your csv with "Add Delimited Text Layer" tool

### important to check!

◇ see notfound.csv: mostly those with a range of street numbers (if you geocode everything)

◇ need to fix them/adjust them:

· to check can just google them and see if you get a clean hit

◇ check location on OpenLayers

· 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

## **outline**

geocoding

SQL

spatial join

probably finish first part here

geo-processing

Extra/bonus

example: apts close to episcopal church in Philly

example: hwys in NJ

## 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 it on your linkedIN next to 'gis' skill

· very easy to master in no time

◇ https://www.youtube.com/watch?v=afPL7-QfHr4

◇ https://www.youtube.com/watch?v=jJeae7PJVv4

### advanced filter (expression): sql/regexp

⋄ 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

**cont**

⋄ 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

## **outline**

## 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 <u>location</u>

◇ so called "spatial join"

· have to pick: mean, sum, etc!

**join counties with universities**

⋄ nj_counties

https://docs.google.com/uc?id=1xJDhcRCkgv7k4tNCa72Oog5bohV6dTB2&export=download

⋄ and universities

https://sites.google.com/site/adamokuliczkozaryn/gis_int/hsip_colleges.zip?
attredirects=0&d=1

⋄ and first make $ENROLL$ numeric

### thinking

$\diamond$ as always, think what your are doing and what does it mean

$\cdot$ and double check

$\diamond$ here there are some institutions with enrollment of zero

$\cdot$ (they were missing "" before the text to float)

$\cdot$ and that affects of course total(sum) enrollment for a county

$\cdot$ and it is unlikely that an institution has zero enrollment

$\cdot$ so ideally, you should find out what these enrollments are...e.g. call the institution

### drooping cases

- at the very minimum, acknowledge the problem by saying that totals have negative bias (say which ones and how many schools missing)
- or maybe better replace with avg; but for now just drop them

◇ hsip_universities-Open Attribute Table

◇ bottom left box-"Advanced Filter (Expression)"

- "ENROLL" >0

◇ then in table without zeros, ctrl-a to select all

◇ layer-save as, ('Save only selected features')

- and 'Add saved file to map'

### MMQGIS-Combine-Spatial Join

◇ Output: nj counties

◇ Spatial Operator: Contains

◇ Data (Join) Layer: universities

◇ Attribute Operation: Sum

◇ Fields: COUNTY AND ENROLL

◇ double check: say Atlantic has 2 and sum of 13,880

◇ use identify tool: 7035+6845=13,880

◇ Essex County is the winner with COUNT of 9 and ENROLL of 65k

◇ now could map ENROLL for counties

## more about spatial join

◇ 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

Extra/bonus

example: apts close to episcopal church in Philly

example: hwys in NJ

## **outline**

### 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

· do let me know which one(s)!– some may not be very
   useful for this class

◇ those that i think are especially useful are covered below

◇ most are under 'Vector', 'Processing', 'MMQGIS', and also
   'Plugins'

## dissolve

◇ dissolve into a larger area

· (get rid of inside boundaries)

◇ Vector-Geopocessing Tools-Dissolve

· nj_counties

◇ "dissolve field:" region

## dissolve your way

◇ can dissolve into your own catgories/definitions

◇ let's take regions and dissolve into south and north jersey

◇ create new variable 'southNorth':

· Open attribute table-toggle editing-New column-integer

◇ mark southern regions with 1, and the rest with 0

· highlight the row to see which county is where

◇ Vector-Geopocessing tools-Dissolve

◇ "Dissolve field:" southNorth

**dissolve your way**

$\diamond$ and now we have a shapefile for south an north jersey

$\diamond$ ofen you will have to do something like this

$\diamond$ there is no way you'll find a shapefile for south jersey online!

$\diamond$ so this tool, like other geoprocessing tools discussed here, is very useful!

### simplify polygons

◇ remember from graphing principles: simplify as much as possible

◇ simplifying polygons means dropping vertices, so that polygons are defined by fewer coordinates

◇ it reduces size of a file

◇ Vector-Geometry tools-Simplify Geometries

· Input: 'nj_counties'

◇ you can play with "tolerance" to simplify it to the point that is needed

· let's try 1000–see the difference?

· for tolerance value, just play with different numbers

## simplify polygons

◇ it is useful if you email things to people, or upload say to google maps

· your data cannot be too big (gmail<10M or so)

· also, you can simplify lines (fewer nodes)

◇ and i guess you can also simplify points (fewer dec points)

· reference http://gis.stackexchange.com/questions/25914/how-to-smooth-generalize-a-polygon-in-qgis

http://stackoverflow.com/questions/1849928/

how-to-intelligently-degrade-or-smooth-gis-data-simplifying-polygons

**centroids**

◇ calculate a center of a polygon

· 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**

$\diamond$ note: the new shapefile will have the same data

$\diamond$ can now map another variable and overlay on another variable

$\diamond$ can map both points and polygons with some symbology

$\diamond$ let's map population for polygons

$\cdot$ and population density for points

$\cdot$ note: make points bigger to see symbology well

$\diamond$ this solves the problem of showing 2 vars in one map

### buffering

◇ kind of opposite of centroids:

· make a buffer (circle) around a point

◇ say, need a 'dry zone' around schools

◇ load $2007\_11\_30\_NJ\_COLL\_UNIV\_{NJSP}$

◇ Vector-Geoprocessing tools-Buffer

◇ use 20,000 feet (buffer size is in map units)

◇ save as 'colBuf'

◇ Properties-Metadata or even -General

· unit is us feet

◇ note: buffer is a new layer and then can spatially merge it with another layer

**example: environemntal problems around univ**

◇ download and add to qgis

· `http://www.nj.gov/dep/gis/digidownload/zips/statewide/Envr_mon_gw_KCSL.zip`

◇ Vector-Data Management Tools-Join Attributes By Location

◇ Target: colBuf

◇ Join: Envr_mon_gw_KCSL

◇ Take summary of intersecting features

· say 'mean'; but we only care about counts, which is automatic

◇ Keep all records

do here 'select by location tool !'

**investigate**

◇ open attr table of merged shaefile

◇ 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

## buffering: applications

◇ why would you do buffering?

◇ sex offenders and schools

◇ liquor stores and schools

◇ waste processing plants and houses

◇ 2-mile heavy pollution around hwy

◇ walkability to healthy stores, etc

◇ many applications!

TODO: update to 2.x

**references**

◇ http://maps.cga.harvard.edu/qgis/wkshop/buffer.php

◇ can select by location:

· $(1.7!)$ http://qgis.spatialthoughts.com/2011/12/tutorial-performing-spatial-queries-in.html

· $(1.8!)$ http://gis.stackexchange.com/questions/61753/how-to-select-points-within-a-polygon-from-another-layer

· more towards bottom:

http://www.qgistutorials.com/en/docs/performing_spatial_queries.html

## **outline**

geocoding

SQL

spatial join

probably finish first part here

geo-processing

Extra/bonus

example: apts close to episcopal church in Philly

example: hwys in NJ

**other things/later**

◇ analysis tools contains many useful tools

◇ can calculate line lengths: e.g. railroads

· http://qgis.spatialthoughts.com/2010/10/
  calculating-line-lengths-and-statistics.html

◇ spatial queries–e.g. select objects within a distance
  http://qgis.spatialthoughts.com/2011/12/
  tutorial-performing-spatial-queries-in.html

◇ calculate X,Y http://maps.cga.harvard.edu/qgis/
  wkshop/x_y_field.php

### and there are many more

◇ mostly under vector menu

◇ but also using plugins

◇ you are more than welcomed to use things we did not cover
  in ps or final project

· also you can have a presentation about some useful tool

· just explore them and google them

◇ in any case it will be extra credit

## next week is the last qgis class

◇ what would you like to cover ?

· anything new ?

· cover anything again?

◇ maybe use some new data for examples?

· I have an impression that we should go to lower level

· title of this class is also (cross-listed) "urban mapping"

· maybe do tracts/blocks in Philly or Camden ?

· maybe zoning or public transportation?

· other ideas?

TODO: guess mv somewhere!

### **outline**

geocoding

SQL

spatial join

probably finish first part here

geo-processing

Extra/bonus

example: apts close to episcopal church in Philly

example: hwys in NJ

## example: apts close to episcopal church in Philly

◇ google 'list of philadelphia episcopal churches'

· https://en.wikipedia.org/wiki/List_of_church_buildings_in_Philadelphia

◇ copy table, put into excel, clean up a bit,

· drop pics, add cols with 'philadelphia' and 'pa'

· save as csv and geocode with MMQGIS

◇ then geocode apt:

· http://philadelphia.apartmenthomeliving.com/apartments-for-rent.xls

◇ make buffers: vector-geoprocessing tools-buffer; say on apt .01

◇ analysis tools-points in polygon

· input polygon: apt buffer; input point: churches

## **outline**

geocoding

SQL

spatial join

probably finish first part here

geo-processing

Extra/bonus

example: apts close to episcopal church in Philly

example: hwys in NJ

## a question

⋄ say we want to find out which county has the longest
interstate hwy network...

### nj roads

◇ http://www.state.nj.us/transportation/gis/data.shtm

◇ and get

◇ NJ Roadway Network

◇ http://www.state.nj.us/transportation/gis/zip/NJ_Roads_shp.zip

◇ SQL ROUTE_SUBT=1 (interstate hwys)

◇ and save selection as hwy.shp

◇ VECTOR-ANALYSIS TOOLS-SUM LINE LENGTH

· Input polygon: 'nj_counties'

· Input line vector: 'hwy'

◇ and the winner is... Morris county

## getting creative with lines

◇ say you work for a local govt...

· and want to lobby state to build more roads

· may produce a map showing miles of reads per capita

◇ or say you want to build more bike lanes

· calculate length of them per capita

· and compare to other leading cities, say Portland, OR