

advanced qgis1

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Thursday 15th February, 2018 12:16

outline

ps comments

geocoding

sql

spatial join

extra credit opportunities

- ◊ present your final project early
 - in addition to extra credit you will get feedback how to improve it
 - and you have to do it anyway later
- ◊ present something we did not cover (has to be GIS, of course)
- ◊ present alternative way of doing something that we have covered

key to success

- ◇ is to start early!!

outline

ps comments

geocoding

sql

spatial join

ps3

- ◊ this was probably the most difficult ps
- ◊ read instructions...
- ◊ if in doubt email listserv

merge again

- ◊ it appears that many had troubles
- ◊ do you have specific questions?
- ◊ again? – see few key slides from data.pdf
- ◊ “Example: New Jersey Home Values”

tips

- ◊ cannot find right data? email listserv !
- ◊ use excel functions to change upper case etc
- ◊ name files for merging with a short word, say 3 characters
- ◊ beware of white space ! say you have “camden county”
- ◊ and you delete county and make it “camden ” –but there is white space!

q

- ◊ how to change color for NULL (missing)
- ◊ there may be other ways...
- ◊ but one way is to replace it with say '-9'
- ◊ a common code for missing in svy research
- ◊ and then when you classify for thematic mapping:
 - add class; make it from -9 to -9 and label as NULL or missing or sth
- ◊ important! record it in journal or somewhere ! so that you know that -9 is NULL

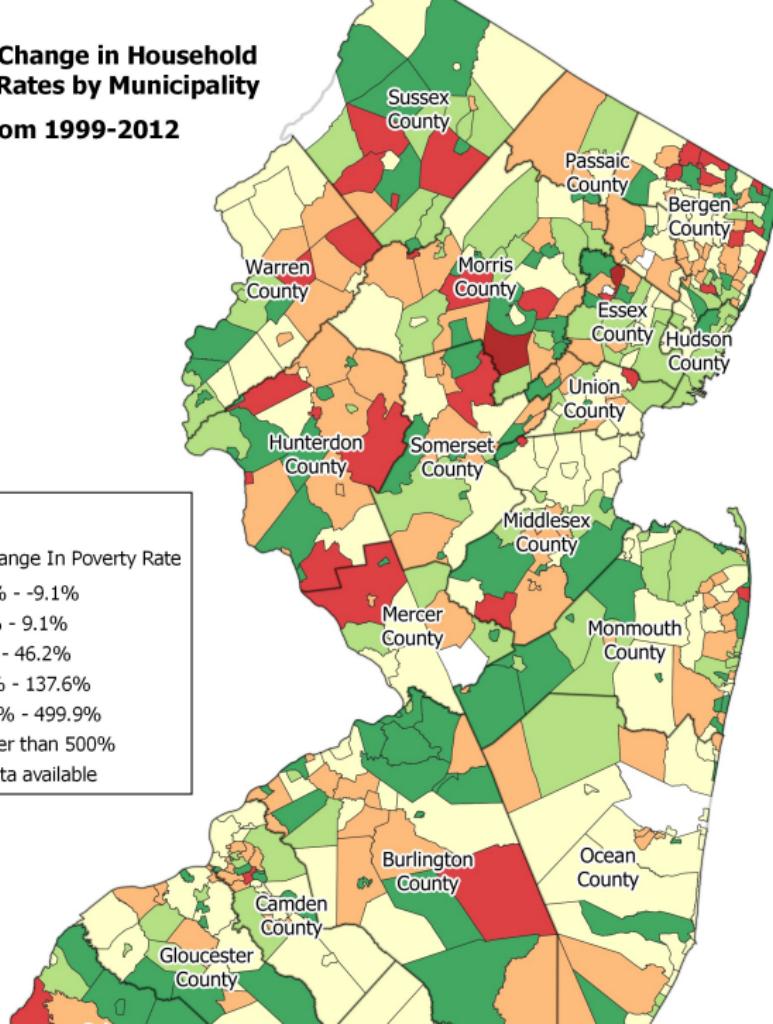
Percent Change in Household Poverty Rates by Municipality

From 1999-2012

Legend

Percent Change In Poverty Rate

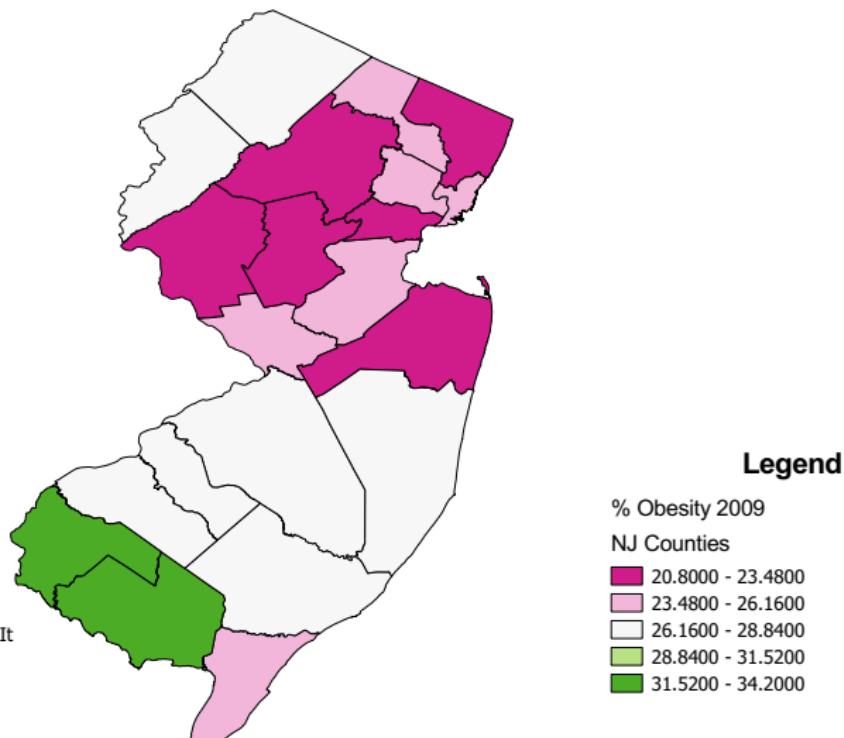
- -100% - -9.1%
- -9.0% - 9.1%
- 9.2% - 46.2%
- 46.3% - 137.6%
- 137.6% - 499.9%
- Greater than 500%
- No data available



poverty change

- ◊ readable, informative, great!
- ◊ great variation ! quite striking
- ◊ wonder why some places do quite well and some so bad...
- ◊ future: change from 1% to 5% and from 10% to 50% are both 500%
 - yet they mean different things
- ◊ can use it for policy analysis...
- ◊ say could thing why such great divergence...
- ◊ any policies that could have worked?

Percentage Of Obesity Prevelance in NJ in 2009



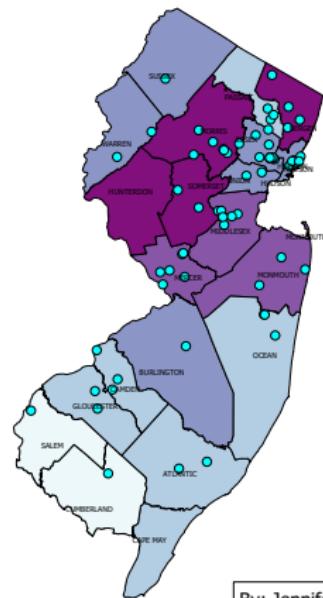
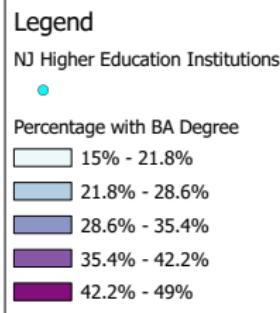
*All data is provided by the CDC. It is county level data

obesity

- ◊ interesting !
- ◊ fit North v not so fit South
- ◊ future research—can map gyms/pc
 - someone did it last year...
- ◊ cut decimal points
- ◊ could make it bigger, use space better

BA degrees

New Jersey Percentage of People with a BA Degree by County



BA degrees

- ◊ this is great ! a very meaningful map
- great idea to overlap NJ higher ed institutions with percent BA/BS and higher
- just expand it a little bit and you have your final project ready !
- ◊ future:
 - go over time—e.g. how it changed over past 50 years
 - scale size or color of NJ institutions by enrollment—some are huge, some are tiny...
 - maybe also see HS and or master's etc...
 - can also do it by census tract...

BA degrees

- ◊ minor points:
 - ugly map ! unreasonably huge font for title and legend; and map itself too small
 - is it BA or BS as well?
 - what is the year ?

how can you get similar data?

- ◊ get data
- ◊ google “education by county”

<https://www.google.com/search?q=education+by+county&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:en-US:official&client=firefox-a>

- ◊ first hit is usda [http://www.ers.usda.gov/
data-products/county-level-data-sets/
education.aspx#.VEU_EfEpAUE](http://www.ers.usda.gov/data-products/county-level-data-sets/education.aspx#.VEU_EfEpAUE)

- just select new jersey in first box

- ◊ then just extract what needed into csv file:

County,Percent_bachelors_degree

Atlantic County,22.7

Bergen County,46

Burlington County,34.1

Camden County,28.7

outline

ps comments

geocoding

sql

spatial join

geocoding

- ◊ assigning coordinates to addresses
- ◊ you may have addresses:
 - school address, hospital address, etc
- ◊ you need to geocode them (turn address into lat/lon and save as gis file)
- ◊ note that you can reverse-geocode, too

geocoding example

- ◊ let's do an example
- ◊ let's say that we have some addresses and we want to geocode them
- ◊ [http://philadelphia.apartmenthomeliving.com/
apartments-for-rent.xls](http://philadelphia.apartmenthomeliving.com/apartments-for-rent.xls)
- and save as csv

MMQGIS-Geocode

- ◊ instal MMQGIS plugin
 - 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
- ◊ let's hit ok, it takes like a minute or two
- ◊ <http://mangomap.com/blog/post/74368997570/>
how-to-make-a-web-map-from-a-list-of-addresses-in-a

important to check!

- ◊ bring in googlemaps (openlayers plugin)
 - does it make sense? houses in river or park?
 - zoom-in to street, click some points to pop-up address—does it match with the street?
 - usually, there will be some miscodings, say less than 1%
 - and it usually happens because the address is misspelled or incomplete
- ◊ also, later we will use google fusion tables that can do it, too

outline

ps comments

geocoding

sql

spatial join

a reference

- ◊ <http://maps.cga.harvard.edu/qgis/wkshop/query.php>
- ◊ maybe also the following:
 - <https://infogeoblog.wordpress.com/2013/01/13/queries-in-qgis-pt-1-attribute-queries/>
 - <http://gis.stackexchange.com/questions/7295/a-sql-console-in-qgis>
 - <https://freecity.commons.gc.cuny.edu/2012/05/07/qgis-fast-sql-layer>

SQL

- ◊ SQL: Structured Query Language
- ◊ quite straightforward, almost like regular language
- ◊ it is also a job market skill...
- ◊ put it on your linkedIN next to 'gis' skill

data (same as in the past)

- ◊ let's get some NJ data

- ◊ it all comes from

https://njgin.state.nj.us/NJ_NGINExplorer/DataDownloads.jsp

- (and there's more than we use here—hospitals, satellite pictures etc)

- ◊ I re-posted them on my website:

http://people.hmdc.harvard.edu/~akozaryn/myweb/bounds_nj_shp.zip

http://people.hmdc.harvard.edu/~akozaryn/myweb/hsip_colleges.zip

search

- ◊ layers-nj_counties-open attribute table-bottom left box
 - column filter: "COUNTY_LAB", enter 'camden', hit "Apply"
 - note it produced query:
 -
 -
- "COUNTY_LAB" ILIKE '%camden%'
 - now easy to modify it, say:
 -
 - "COUNTY_LAB" ILIKE '%bur%' gets 'Burlington'
 - '%mo%': Monmouth, Morris; etc etc
- ◊ then you can save selection as new shapefile

advanced filter (expression): sql/regexp

◊ layers-nj_counties-open attribute table-bottom left box

◊ select “advanced filter”



`regexp_match("COUNTY", 'C.*N')`

· there is 'C', some chars '.', '*' , 'N'



`regexp_match("COUNTY", '^C.*N')`

· must start with 'C'



`regexp_match("COUNTY", '^C.*N$')`

· must start with 'C' and end with 'N'

advanced filter (expression): sql

- ◊ layers-nj_counties-open attribute table-bottom left box
- ◊ select “advanced filter”
- ◊ “REGION” = ‘CENTRAL’ (can do “Load values” “all unique”
 - “REGION” = ‘CENTRAL’ AND “POP2010” > 598349
- ◊ then hit ctrl-a to select all data, close table
- ◊ right click layer, save as, and check “selection”

more about calculator

- ◊ reference for operations (note 1.8!) http://docs.qgis.org/1.8/en/docs/user_manual/working_with_vector/field_calculator.html
- ◊ $a||b$: combines two strings (i think!), e.g. 'Camden' + 'NJ'
- ◊ create categorical var
- ◊

CASE

```
    WHEN "POPDEN2010" > 0 AND "POPDEN2010" < 5000 THEN 1  
    WHEN "POPDEN2010" > 5000 AND "POPDEN2010" < 10000 THEN 2  
END
```

- ◊ note, unspecified val are NULL!; let's check in attr tab

sql in general

- ◊ full blown (not in qgis) sql is only little more complicated
 - it is Structured Query Language
 - very much English-like, just with some strict syntax rules
 - very easy to master in no time
 - interested in learning more? see ref 1st slide in this sec

outline

ps comments

geocoding

sql

spatial join

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, etc!

join counties with universities

- ◊ they are 2 different geographies of course
- ◊ bring in nj_counties.shp
 - and 2007_11_30_NJ_COLL_UNIV_njsp.shp
 - and first make ENROLL numeric
- ◊ calculate a new int field: “enrN”, “to_int(ENROLL)”

thinking

- ◊ as always, think what you are doing and what does it mean
- ◊ for instance, here we are calculating sum
 - so turning NULL to 0 on “enrN” is a problem
- ◊ there are some institutions with enrollment of zero
 - and that affects course total(sum) enrollment for a county]
 - and it is unlikely that an institution has zero enrollment
 - 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
 - ◊ select "enN" >0
 - ◊ in query builder
 - ◊ then in table without zeros, ctrl-a to select all
 - ◊ close table, layer-save as, 'save only selected features'
 - and 'add saved file to map'
 - ◊ Remove the original colleges shapefile

clean up table

- ◊ let's only keep fields that we need
- ◊ usually a good idea to keep it simple and clean
- ◊ there seem to be problems if you don't do that !
- ◊ properties-fields; toggle editing and ctrl-select all fields
- ◊ and drop everything but ID and enrN

joining

- ◊ VECTOR-DATA MGMT TOOLS-JOIN ATTRIBUTES BY LOCATION
 - target: nj_counties
 - join vector layer: 2007_11_30_NJ_COLL_UNIV_njsp (one without zeros)
 - TAKE SUMMARY OF INTERSECTING FEATURES
 - this is important !
 - think what does it mean, what is meaningful
 - we do sum: want to know tot enrollment for each county
 - “keep all records”
 - and save a new shapefile somewhere... say merged.shp

open attribute table

- ◊ have a new field 'SUMenrN'
- ◊ note it also created field counting points in polygon!
- ◊ if no univ in a county, then it's "NULL"
- and we can make a thematic map of 'SUMenrN'

more about spatial join

- ◊ http://maps.cga.harvard.edu/qgis/wkshop/join_spatial.php
- ◊ [http://trendct.org/2015/05/29/
tutorial-how-to-merge-data-from-two-different-maps-using-qgis/](http://trendct.org/2015/05/29/tutorial-how-to-merge-data-from-two-different-maps-using-qgis/)
- ◊ http://www.qgistutorials.com/en/docs/performing_spatial_joins.html