# misc: rules, tips, tricks, ethics

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Thursday 18th October, 2018 12:49

# outline

data

tips and tricks

some rules

ethics

an example of my research

research design again

# outline

# outline

# have a big screen

- again, i cannot overemphasize, that
- a big screen is key for gis work
- (it's inexpensive, too)
-
- and of course if you dont have it yet, get a mouse

# useful tools

- zoom to layer extent
- un-select features if a tool behaves unpredictably
- use identify features tool
- explore plugins

# workflow

- save the whole project (with many layers) and next time just open it
- can move layer around
- can have many layers with say different symbology of the same shapefile

# misbehaving software

- most of the software sometimes misbehaves...
  - it crashes; refuses to do something, etc
- troubleshooting:
  - email me
  - do what you are doing in a different way-e.g try different dataset; different var; different approach etc (usually can do same thing in many ways)
  - shut it down and fire it up again
  - reinstall (last resort)

## google it

- depressing, but whatever you are mapping, someone has already done it
- accept it, and make use of it!
- google and see images, say: 'nj counties contamination sites' `https://www.google.com/search?q=nj+counties+contamination+sites&tbm=isch`
- or "Philadelphia healthy stores map" (sometimes need word 'map' otherwise get pics of healthy food)
  - `https://www.google.com/search?q=philadelphia+healthy+stores+map&tbm=isch`
- get ideas, inspiration from these googled maps
- try to make your map better than the competition
- still, the key to be innovative is to join data!

## google

- cannot overestimate the usefulness of google for finding data
- e.g. "what you are looking for, shapefile"
- e.g. "new jersey public schools, shapefile"
- tips:
  - may need to look for a higher level; e.g. NJ schools instead of Depford Twshp schools
- if you cannot find it, contact govt; e.g. city of Camden, state of NJ, etc–they will be happy that you want to use their data
- again, may find only traditional data and need to join with gis data

# google it

- likewise, if you want to map it, and not sure how
- or perhaps just want to visualize it, google it
  - say 'census regions or divisions'
  - instead of reading about what this could be
  - a map is worth 1,000 words!
  - `https://www.google.com/search?q=us+census+divisions&client=firefox-a&hs=VPH&rls=org.mozilla:en-US:official&source=lnms&tbm=isch&sa=X&ei=sgUzVLSeOoeoyQTVh4LIBQ&ved=0CAgQ_AUoAQ&biw=1147&bih=1261`

## join data

- the real value comes from joining data!
- again, a map about any single var was already made
  - but 2 given vars in a map or in 2 maps are rare
- there are so many data and variables out there
- use your creativity and imagination
- and you'll easily come up with something that no one did
- then join the data and make a map
- eg `http://people.hmdc.harvard.edu/~akozaryn/myweb/rel_inn.pdf`
  - see 2 maps at the end
- sure, gis is mostly a technical skill; but there's some art here too!

# outline

# quality

- GIGO: Garbage In, Garbage Out
- 'Cos it's in the computer, don't mean it's right
  - double, triple check
  - ask yourself if it makes sense... (Camden richer than Cherry Hill?)
  - use several datasources and or several variables to measure the same thing
  - are you getting similar results? why not?

## unknowns by Rumsfeld (be humble in your findings)

- There are known unknowns.
- That is to say there are things that we now know we don't know.
  - (these are benign, but be explicit about them)
- But there are also unknown unknowns. There are things we do not know we don't know.
  - (these are tricky: you can't do anything about unknown unknowns other that acknowledge that they exist; and never say you "proved" something)
  - (your statements are valid until disproved: all Swans are white, only until you see one day a black Swan)

# outline

# integrity/honesty

- be explicit about problems in your data
  - eg non-joins, missing data, miscodings
- be explicit about problems in your models:
  - eg don't throw away variables from maps just because they contradict your story
  - discuss it: how, why; ask audience to comment/criticize
- instead of forcing data to tell your story, listen carefully; let data tell you her story !
- if you work for somebody: e.g. a bank or NGO: they will ask you to find something; use a disclaimer saying that

# ethics

- everybody wants to sell something
- we academics or thinkers or students, too!
- we try to sell some idea or point of view
- rarely if ever anyone is 100% objective
- keep that in mind !
- and always try to present alternative/opposite points of view
- present the whole picture

# ethics: bad examples

- cherry picking of vars or samples or timeframes, etc
- eg using only variables or operationalizations that fit your story
- eg using year in which your find what you wanted to find
- classification bins: playing with bins will always emphasize your story
- in short: force yourself to be objective, because by default humans aren't
- see fascinating `https://righteousmind.com/`

# outline

# happiness in Europe

- have a look at `https://sites.google.com/site/adamokuliczkozaryn/pubs/gesis3.pdf`
- first, on p.5 I show a histogram of happiness
  - again qgis has native histograms
  - or can use plugins; search for 'stat'
- and then on p. 6 and 7 two maps using quantiles and natural breaks/jenks
- note, that you can be creative, and calculate other interesting quantities sch as variation e.g. p. 11

# outline

# important for paper

- think (and address them) about those things below
- when working on a paper

# error of measurement

- keep in mind that measurement is always imprecise...and ask yourself how imprecise
- see literature; e.g. happiness has been cross-validates: PET scans, opinions of friends etc

## error of measurement

- who produced data ? e.g. Chinese data are less reliable than US data
  - e.g. national govt reports lower pollution to look good internationally
  - local govt reports higher pollution to show that it develops and produces a lot...
  - disclaimer: i have read it somewhere, not sure how true is that
  - the point is to always think about quality of data
  - for Chinese case it would be good to triangulate it with some intl data, say satellite images...

# think about incentives

- who is producing that data?
- again, you can measure a concept in many different ways
- people have an incentive to measure it in a way that benefits them

# construct validity

- are you measuring what you say you are measuring?
- say you want measure ability, or IQ, but you only have data about education
- http://www.socialresearchmethods.net/kb/constval.php
- seven sins map

  http://2.bp.blogspot.com/_R3SXJVojagU/SwLzZJL1E2I/AAAAAAAAAIE/7GbMzcZPDDk/s1600/sevendeadlysins.bmp

# external validity

- are your data representative ?
- how big is the sample ?
- e.g. I was geocoding WVS at province level only to find out it was unrepresentative

# triangulate

- triangulation=use different measures for the same concept
- e.g. education:
  - years of schooling
  - highest degree obtained
  - avg SAT score
  - avg ranking of schools in the area
  - etc etc

## time matters, too

- we are exploring spatial variation
- but there is also time variation
- usually it is nice to show time changes in your maps
- e.g. can display a variable as a difference say
  - $POP10 - POP00$–which county gained most population (let's do it with nj_counties)
- other time issue is that things fluctuate over time, say due to business cycle
  - if you want to show a more reliable estimate. take an average
  - say avg. 5-yr unemployment rate

# go places

- when you make maps and find things, go and visit that place–i drove through MI from TX to NJ