misc: rules, tips, tricks, ethics

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Tuesday 10th October, 2023 12:31

outline

data

tips and tricks

some rules

ethics

an example from my research

research design again: important from now on

outline

data

tips and tricks

some rules

ethics

an example from my research

research design again: important from now on

outline

data

tips and tricks

some rules

ethics

an example from my research

research design again: important from now on

have a big screen

- again, i cannot overemphasize, that
- a big screen is key for gis work
- (it's inexpensive, too)
-
- and get a mouse (folium leaflet later)

google it

- depressing, but whatever you are mapping, someone has already done 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, make your map better
- still, the key to be innovative is to join data!

google it

- cant overestimate usefulness of goog for finding dat
- eg "what you are looking for, shapefile"
- eg "new jersey public schools, shapefile"
- tips:
  - may need to look for a higher level; eg NJ schools instead of Depford Twshp schools
  - if you cant find it, contact govt; eg city of Camden, state of NJ, etc–they'll be happy you're using their data
  - again, may find only traditional data and need to join

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 set of 2 maps: 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

outline

data

tips and tricks

some rules

ethics

an example from my research

research design again: important from now on

quality

- GIGO: Garbage In, Garbage Out
  - 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 (triangulation)

outline

data

tips and tricks

some rules

ethics

an example from my research

research design again: important from now on

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 hide maps bc 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!

ethics

- everybody wants to sell something
- we academics or thinkers or students, too!
- we try to sell some idea or point of view
- nobody 100% objective
- always try to present alternative/opposite points of view
- present the whole picture
- force yourself to be objective, because humans aren't
- see fascinating `https://righteousmind.com/`

ethics: bad examples

- cherry picking of vars or samples or timeframes, etc
- eg using only vars/operationalizations that fit your story
- eg using year in which you find what you wanted to find
- classification: playing with bins to fit your story

outline

data

tips and tricks

some rules

ethics

an example from my research

research design again: important from now on

happiness in Europe

- have a look at `https://theaok.github.io/docs/gesis3.pdf`
- first, on p.5 I show a histogram of happiness
  - do histograms!
- 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 such as variation eg p. 11

outline

data

tips and tricks

some rules

ethics

an example from my research

research design again: important from now on

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 ?
- eg I was geocoding WVS at province level only to find out it was unrepresentative

triangulate

- triangulation=use different measures for the same concept
- eg 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
- eg 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