thematic maps

adam okulicz-kozaryn
adam.okulicz.kozaryn@gmail.com

this version: Friday 13\textsuperscript{th} November, 2020 12:33
outline

misc

basics again

classification methods: 2 useful references

thematic mapping

heatmaps

layers-properties: labels and metadata
outline

misc

basics again

classification methods: 2 useful references

thematic mapping

heatmaps

layers-properties: labels and metadata
how’s ps2?

- any quick questions?
- we’ll try to flip the ending of the class and work on it
how is qgis so far?

• what doesn’t work?
• what shall i cover more/again?
outline

misc

basics again

classification methods: 2 useful references

thematic mapping

heatmaps

layers-properties: labels and metadata
variable definitions

- be very clear about what you are measuring
  - put on the map, in description, or into appendix, but have to have it somewhere!
  - eg do we have small breweries that are at some bars? how exactly is a brewery defined?
  - eg what is exactly a bike lane—do we include paths in parks?
    - does it have to be designated for bikes only?
    - and paths not for bikes but used by bikes?
  - ideally map them all!
map labeling: clarity and simplicity!

- always have a self explanatory title/caption and legend
- self-explanatory means a random person will understand what it’s about
- in other words it will pass “a grandma test”
  - give it to your grandma and she’ll get it
  - if she doesn’t, then it isn’t clear enough
classification methods: 2 useful references
Properties-Style-histogram tab; skew

A
B
C
D

classification methods: 2 useful references
references: very useful!

- let’s open both and do 2nd pdf: 7,8: creating classes
- and then do each classification type one by one from BOTH docs; and s15 from 2nd on counts v ratios
- http://www.geo.umass.edu/courses/geo494a/thematic_map_design.pdf

  [*] afficionados may do value-by-area

misc

basics again

classification methods: 2 useful references

thematic mapping

heatmaps

layers-properties: labels and metadata
always think abt the meaning; interpret!

- always interpret the map, think about what it means
  - usually want to standardize to achieve meaningfullness
- standardize by area (“per sq km”) or by pop (“per capita”)
- or even: specific (eg habitable) area; specific (eg disadvantaged) pop
  - eg much of area may be water or forest, so hydrants/inhabited sq km
  - similar with populations-they may only work or sleep in some area, (Cherry Hill is a bedroom city) etc
  - eg Cape May has many liquor stores per capita (just because nobody lives there)
standardize: gen a new variable

• nj counties
  https://drive.google.com/open?id=1xJDhcRCKgv7k4tNCa720og5bohV6dTB2

• map POP2010

• duplicate the layer so can easily compare

  • “Open Field Calculator”

  • “Output filed name”: “pd10” [qgis doesn’t like long var nam]

  • “Output field type”: “Decimal number (real)
    and bump up precision to say 10 (decimal points)

• calc \( \frac{POP2010}{SQ\_MILES} \) (can select from variables drop-down)

• map it: equal interval, and compare to the original

• big difference—the county next to NYC is much more dense than everything else
what do we see? (distribution, skew)

• but wait! this map is not very useful because there is not much variability in it

• this happens when data are skewed—the county next to NYC is much more dense than anything else (right-skewed, draw distribution)

• Properties-Style, “Histogram” tab, hit “Load values”

• try more classes and see how distr changes

• but even if we have 10 classes it doesn’t help much

• better pick some other classification technique

• try NATURAL BREAKS (JENKS)

• note! almost always have to move cutoff lines manually so that clusters are colored same col!!!
level of analysis

• remember i was repeating myself over and over again that the level matters
• and that usually the lower (finer) the better
• and that the higher, the more information you lose
• here’s an example
level of analysis: example

- load NJ_MUNIS
- and map with 5 quantiles POP_DEN2010
  - a huge difference! [and same data!!]
  - note many areas next to Philadelphia, NYC and some coastal areas
- the previous map did not showed that at all!
  - Only one county next to NYC showed up because it were small and ALL densely populated
- but the rest of the counties were densely populated only in few subareas
classification methods

- always understand the distribution—use hist!
  - have a hist in ps (at least of main var)
  - think about it, discuss and motivate classification method
  - (I’ll cut points)
- I like NATURAL BREAKS/JENKS or QUANTILES
- usually more “truthful” than equal intervals
- start with many, say 10, then shrink it to say 5 or 3 without losing too much detail
- make it as parsimonious, clean, and simple as possible
choice of classification method is critical

- be as objective as possible
- never choose classification forcing your story
  - let the data speak, listen carefully, don’t force it
- scientist must be objective
- play with it: explore the distribution and categorize differently
- then pick the most parsimonious AND best representing the pattern
  - (put the alternative ones into appendix, so can always compare)
- let the data speak! do not force your story
let the data speak, but you pick the story!

- data have always many stories to tell
  - and you choose which one you want to present
- say may emphasize extremes with dramatic colors
  - eg purple for values way different from everything else
    - (for intervention, disaster response, etc)
- or paint the gradient, where values raise and level off etc
  - like my urban-rural happiness gradient
- also in space: clusters of happiness: https://link.springer.com/content/pdf/10.1007/s11205-010-9671-y.pdf
  - (still using alt classifications for robustness)
  - (and std dev in addition to levels)
categorized symbology

- good for categorical data
- what are categorical data?
- examples?
- continuous vs ordinal, nominal (multinomial and binary)

- categorized symbology—how it works?
  - pick colors (or symbols) for levels of a variable
bring in universities

- load https://sites.google.com/site/adamokuliczkozaryn/gis_int/hsip_colleges.zip?attredirects=0&d=1
  - layer-Properties-Style; select “Categorized”
- do CATEGORIZED classify by NAICSDESCR and pick some big symbol for “universities” level
- then can easily see there are only 2 univ in SJ
- use IDENTIFY TOOL (arrow with i) to identify
- Aha! RU and Rowan–maybe then should merge them
more than one var: dots, hashed lines

- map additional var with empty fill as hashed lines or dots
- let's try it: colored pop and hashed/dotted pop den
- nj counties

  https://docs.google.com/uc?id=1xJDhcRCkgv7k4tNCa720og5bohV6dTB2&export=download

- btw can just click symbol under main layers in main window
misc

basics again

classification methods: 2 useful references

thematic mapping

heatmaps

layers-properties: labels and metadata
too many points? heatmap! or clusterer!

- https://docs.google.com/uc?id=1T_n1y_Mj5yQfWpZwrbuuFFwmlVJ2QWFZ&export=download

- we got a map, but mess! make them smaller:
  - under style, change size to say .4

- better a heatmap:
  - right click layer-Properties-Style: Heatmap
  - play with Radius to achieve desired heat
  - (at home: overlay with county bounds etc to locate better)

- or clusterer: increase clustering distance to 10mm
  - make symbol bigger and font smaller

- refs:
  - https://docs.qgis.org/2.8/en/docs/user_manual/plugins/plugins_heatmap.html
misc
basics again
classification methods: 2 useful references
thematic mapping
heatmaps
layers-properties: labels and metadata
what else under layers-properties?

- we’ve covered *STYLE*...
- let’s stick in some *LABELS*
- can pick ANY text you get when you use IDENTIFY FEATURES TOOL, ie any text from properties table
- from `NJ_COUNTIES` display `COUNTYLAB`
- select a “buffer” to have nice outline—easier to read
- note: can put as label any var, incl numeric, letter, etc!
  - so it is a way of having 2 vars in one map: thematic+label
label only certain features

- can subset a shapefile, that is select features of interest and save them and load again and then label,
  - let's do it say with South Jersey
- or there is also another way: [http://anitagraser.com/2015/12/04/how-to-label-only-selected-features-in-qgis-2-8-and-up/](http://anitagraser.com/2015/12/04/how-to-label-only-selected-features-in-qgis-2-8-and-up/)