

thematic maps

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Thursday 15th February, 2018 17:11

outline

misc

basics again

basic descriptive statistics

classification methods

thematic mapping

heatmaps

layers-properties: labels and metadata

outline

misc

basics again

basic descriptive statistics

classification methods

thematic mapping

heatmaps

layers-properties: labels and metadata

how is qgis so far?

- ◇ what doesn't work?
- ◇ what shall i cover more/again?

outline

misc

basics again

basic descriptive statistics

classification methods

thematic mapping

heatmaps

layers-properties: labels and metadata

variable definitions...

- ◇ be very clear about what you are measuring
 - put it either on the map, or into metadata, or into “codebook” 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 ?

map labeling

- ◇ must have a legend
- ◇ must have a self explanatory title/caption
- ◇ self-explanatory means that if I give it to a random person that person will understand what is it about
- ◇ in other words it will pass “a grandma test”
 - give it to your grandma and she must be able to understand it
 - if she doesn't, then it isn't clear enough

questions

- ◇ a question was how to deselect features:
 - there is a tool with red color for deselecting
 - let's select and deselect something

outline

misc

basics again

basic descriptive statistics

classification methods

thematic mapping

heatmaps

layers-properties: labels and metadata

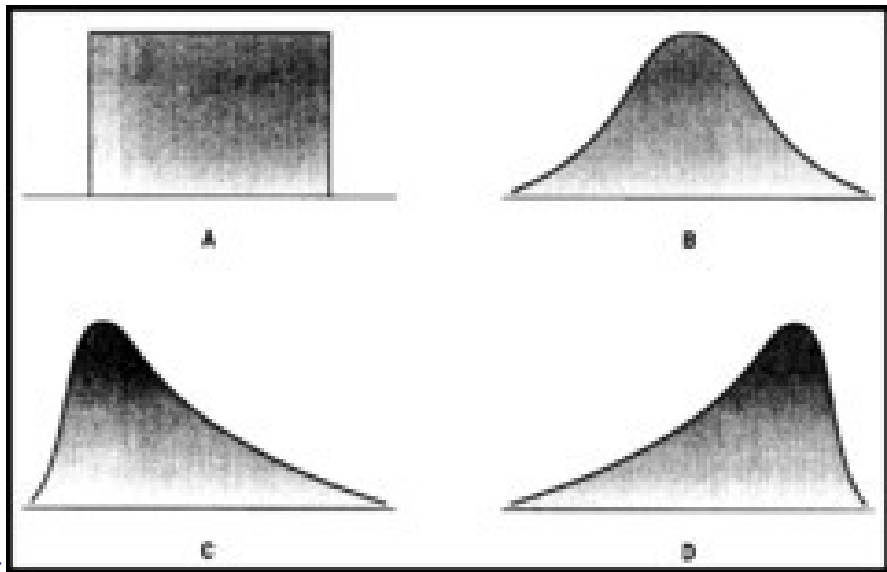
why? it's a gis class

- ◇ important to know a little for understanding thematic mapping
- ◇ again, thematic mapping is about classifying values into bins
- ◇ it all depends on how the vales are distributed
- ◇ you need to know something about distributions
- again: Properties-Style-histogram tab

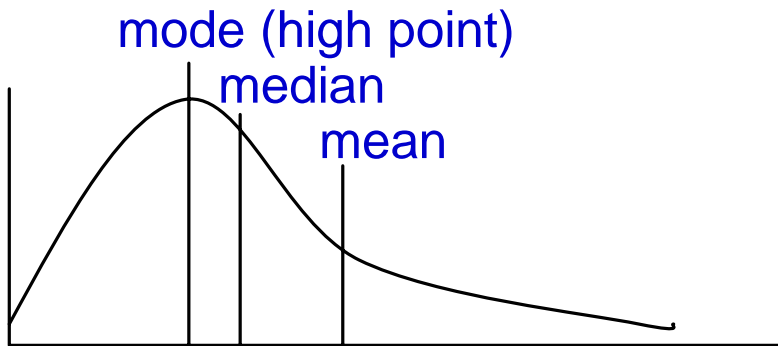
central tendency and skewness

- ◇ eg: 1, 2, 2, 3, 12 (right skewed)
- ◇ median: middle value 2 (if even take the mean of the middle two)
- ◇ mean $\frac{1+2+2+3+12}{5}=4$ (affected by extremes)
- ◇ left skewed: $\mu < M$
- ◇ right skewed: $\mu > M$

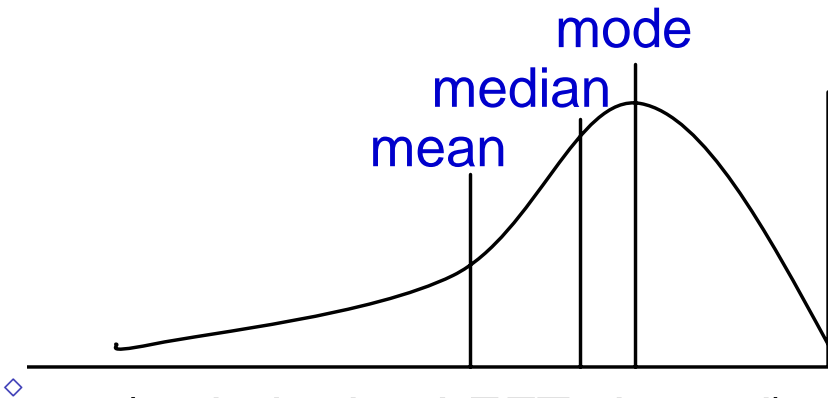
skew



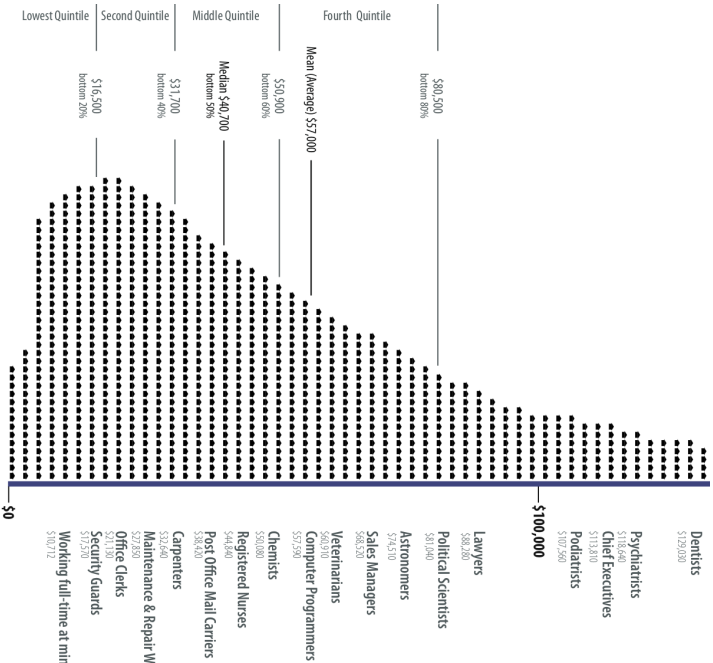
right skew



left skew



income distribution



outline

misc

basics again

basic descriptive statistics

classification methods

thematic mapping

heatmaps

layers-properties: labels and metadata

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
- ◇ http://www.gitta.info/Statistics/en/html/StandClass_learningObject2.html
- ◇ http://www.ttu.ee/public/e/ehitusteaduskond/Instituudid/Teedeinstituut/Geodeesia_oppetool/oppematerjalid/thematic_map_design.pdf

outline

misc

basics again

basic descriptive statistics

classification methods

thematic mapping

heatmaps

layers-properties: labels and metadata

standardization-always think about the meaning

- ◇ nj counties <https://drive.google.com/open?id=1xJDhcRCkgv7k4tNCa720og5bohV6dTB2>
- ◇ it's not meaningful (for most purposes) to rank U/As by population given the fact they differ in size
- ◇ most of the time you want to standardize by area ("per sq km") or by population ("per capita")
- ◇ or by specific area and by specific population
 - eg much of some area may be water or forest
 - similar with populations-they may only work or sleep in some area, (Cherry Hill is a bedroom city) etc etc
 - eg Cape May has many liquor stores per capita (just because nobody lives there)

generate a new variable

- ◇ “Open Field Calculator”
- ◇ “Output filed name”: “pd10” [qgis doesn’t like long var names!]
- ◇ “Output field type”: “Decimal number (real)”
 - and bump up precision to say 10 (decimal points)
- ◇ calculate *POP2010/SQ_MILES* (can select from variables drop-down)
- ◇ map it and compare to the original
- ◇ big difference—the county next to NYC is much more dense than everything else

what do we see ? (the distribution)

- ◇ 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” (have window big or wont open)
- ◇ try more classes (draw eq. size bins on the distr)
- ◇ or better yet pick some other classification technique
- ◇ even if we have 10 classes it doesn't help much
- ◇ let's try NATURAL BREAKS (JENKS)

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 loose
- ◇ here's an example

level of analysis: example

- ◇ load NJ_MUNIS
- ◇ and map with 5 quantiles POP_DEN2010
 - a huge difference!
 - 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

what do we see ? (the distribution)

- ◇ 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)
- ◇ may try more classes... (draw eq. size bins on the distribution)
- ◇ or better yet pick some other classification technique
- ◇ even if we have 10 classes it doesn't help much
- ◇ let's try NATURAL BREAKS/JENKS

classification methods

- ◇ again, always think hard about the distribution of a variable that you are mapping—histogram is one of the best tools
- ◇ i like NATURAL BREAKS/JENKS or QUANTILES
- ◇ they usually show the data better than equal intervals
- ◇ start with many, say 10, and then see if you can shrink it to say 5 or 3 without losing too much information
- keep in mind graphing principles we covered last week: clarity and parsimony

choice of classification method is critical

- ◇ try to be as objective as possible
- ◇ never choose a method that shows something that fits your story
- ◇ you are a scientist, you have to be objective
- ◇ explore the distribution; look at different ways of categorizing the values
- ◇ pick the one that is most parsimonious, yet it does represent what is going on
- ◇ let the data speak! do not force your story

categorized symbology

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

categorized symbology—how it works?

- ◇ you can specify your own symbols and/or colors for levels of a variable

bring in universities

- ◇ load 2007_11_30_NJ_COLL_UNIV_njsp.shp
 - layer-Properties-Style; select “Categorized”
- ◇ do CATEGORIZED classify by NAICSDESCR and pick some big symbol for “universities” level
- ◇ then we can easily see that there are only 2 universities in South Jersey...
- ◇ use the IDENTIFY TOOL (arrow with i) to see what they are
- ◇ Aha ! RU-Camden and Rowan—maybe then we should merge them...

outline

misc

basics again

basic descriptive statistics

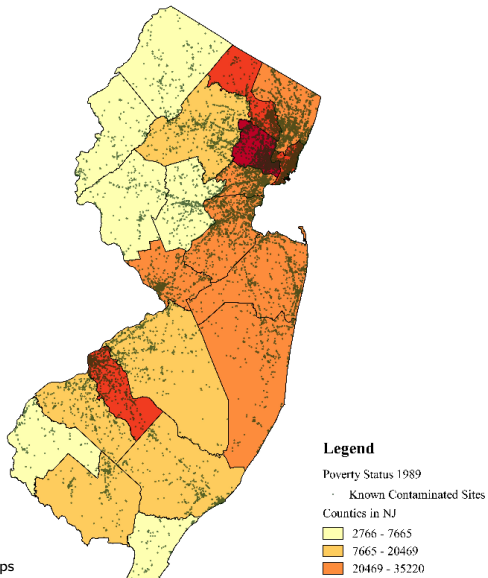
classification methods

thematic mapping

heatmaps

layers-properties: labels and metadata

Contaminations Sites in New Jersey 1992



contaminations

- ◇ this is a pretty good map!
- ◇ perfect size and color for contaminated sites!
 - doesn't overlap much but big enough to see
- ◇ so you could just do something like that and you are fine!
- ◇ but you can do something little more fancy
- ◇ and sometimes you probably have to do something little more fancy
 - that is when there are way too many points, like thousands...
 - (well you could zoom in, but if you want to show the whole thing)
- ◇ then do a heatmap!

contaminations: too many points? heatmap!

◇ get

https://docs.google.com/uc?id=1T_n1y_Mj5yQiWpZwrbuuFFwmIVJ2QWFZ&export=download

◇ load it and...we got a map

- but lots of points! make them smaller:
- under style, change size to say .4

◇ but can also do 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)

◇ reference:

- http://www.qgistutorials.com/en/docs/creating_heatmaps.html
- https://docs.qgis.org/2.8/en/docs/user_manual/plugins/plugins_heatmap.html
- <https://www.mapbox.com/tilemill/docs/guides/designing-heat-maps/>
- http://www.digital-geography.com/create-point-density-raster-in-qgis/#.VrtsS_F0kUE

outline

misc

basics again

basic descriptive statistics

classification methods

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 some of the text you get when you use IDENTIFY FEATURES TOOL
- ◇ from NJ_COUNTIES display COUNTY_LABEL
- ◇ select a “buffer” to have nice outline—easier to read

label only certain features

- ◇ can subset a shapefile, that is select features of interest and save them and load again and then label,
- ◇ or there is also another way: <http://anitagraser.com/2015/12/04/how-to-label-only-selected-features-in-qgis-2-8-and-up/>

layers-properties-metadata

- ◇ remember i was stressing this is important
- ◇ metadata=data about data
- ◇ and for now we'll skip the other tabs...