thematic maps

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Thursday 1st October, 2020 08:34

outline

misc

basics again

basic descriptive statistics

classification methods: 2 useful references

thematic mapping

more than var

heatmaps

layers-properties: labels and metadata

outline

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

variable definitions...

- be very clear about what you are measuring
  - put it either on the map, or 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 ?

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

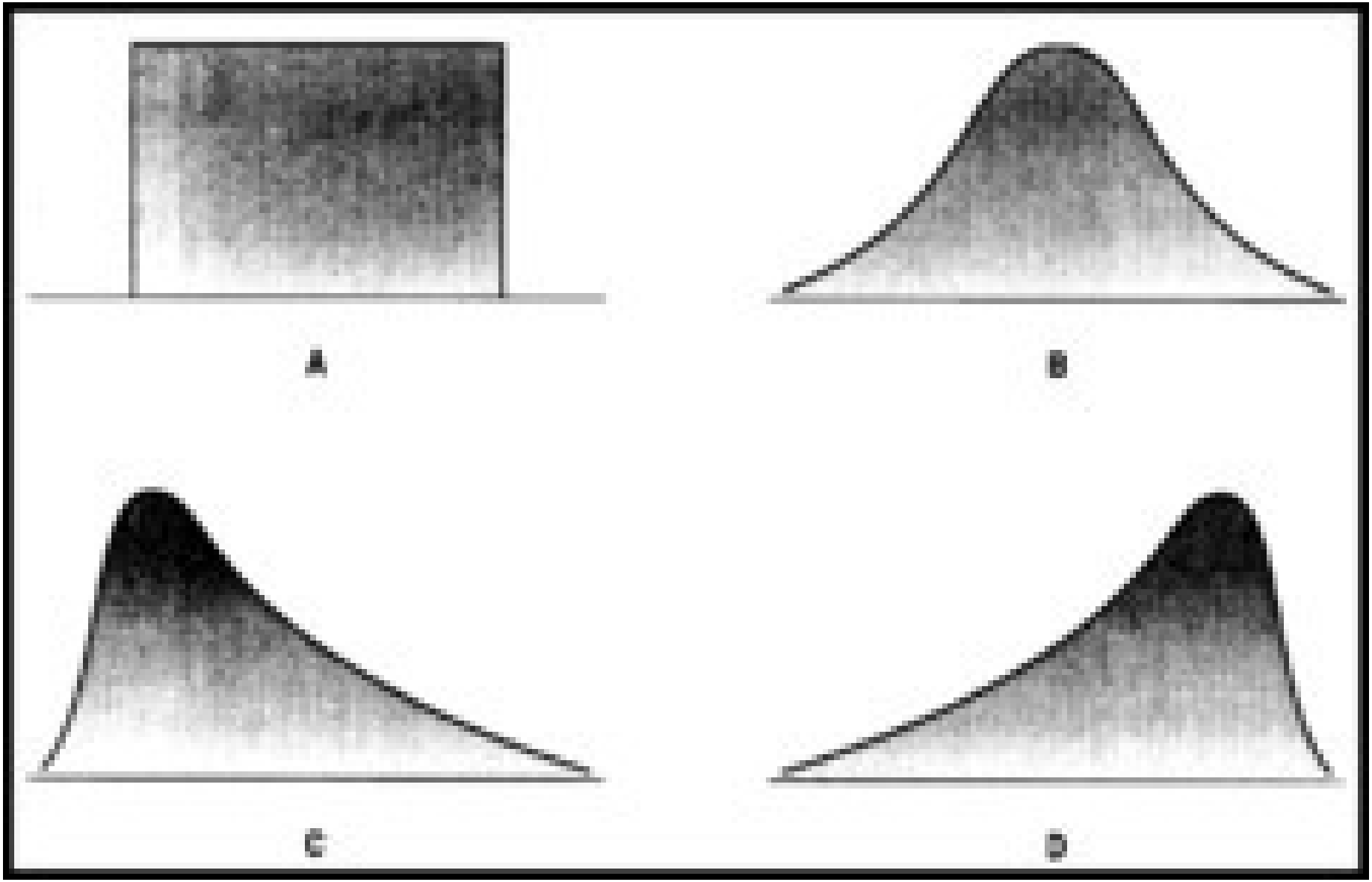
- a question was how to deselect features:
  - there is a tool with red color for deselecting
  - let's select and deselect something

outline

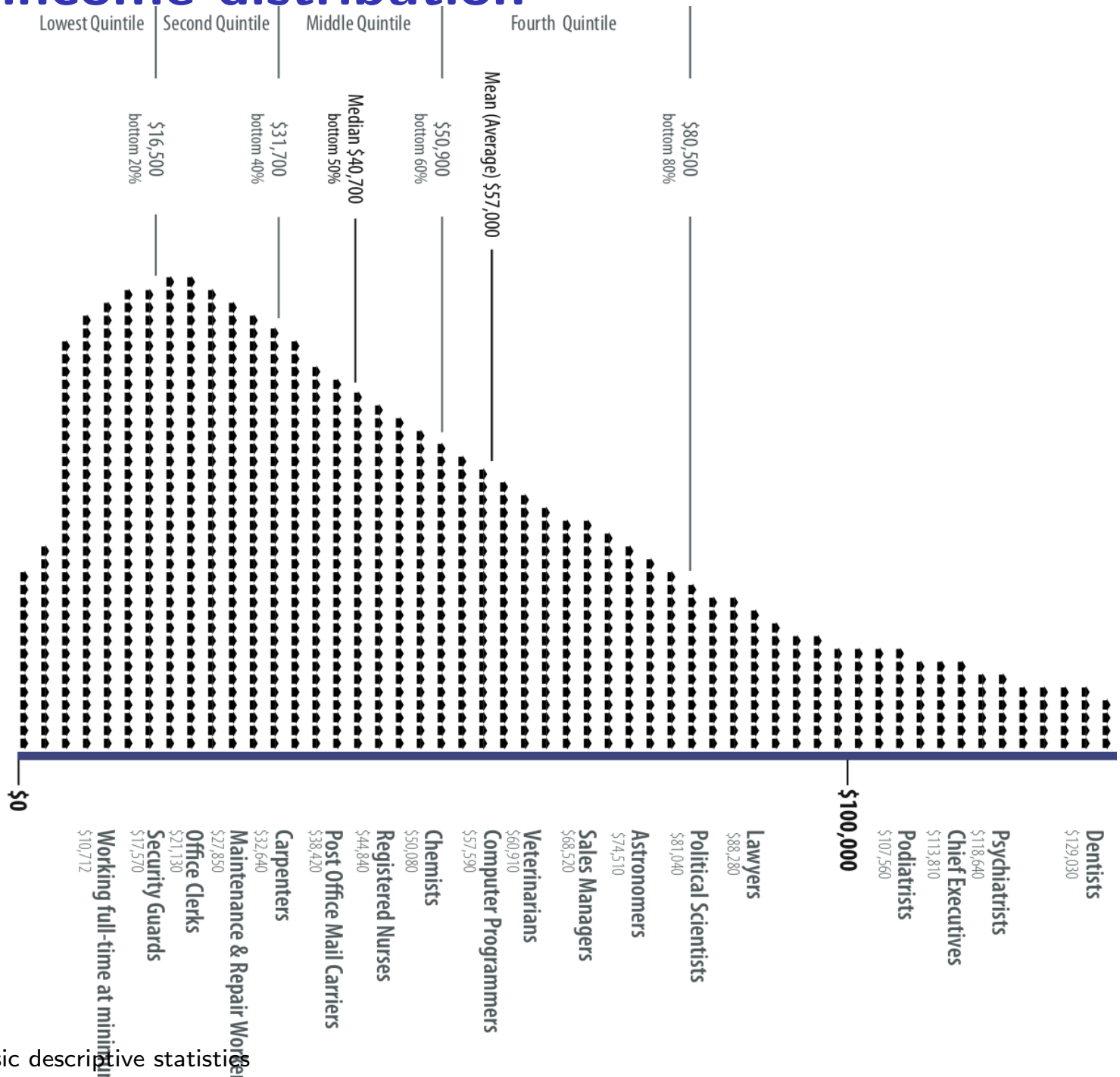
why? it's a gis class

- 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
- show side by side histogram with map in qgis

skew



income distribution



outline

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.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

standardization-always think about the meaning

- nj counties
  https://drive.google.com/open?id=1xJDhcRCkgv7k4tNCa72Oog5bohV6dTB2
- map POP2010: 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

- first duplicate the layer
- "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: 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 ? (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"
- try more classes and see how distr changes
- but even if we have 10 classes it doesn't help much
- better yet pick some other classification technique
- 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! [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

- again, always think hard about the distribution of a variable that you are mapping–histogram is one of the best tools
  - should have the histogram in presentation/paper
  - think about it hard, discuss, and do motivate classification technique!
  - if not, i will cut off points!
- 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 loosing 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 `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 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

centroids

- we will see in advQ.pdf
- that we can generate centoids
- and color them

dots, hashed lines

- but for now can just duplicate the layer
- and express additional var with empty fill
- as hashed lines or dots
- of various colors
- lets try it pop and pop den
- nj counties

`https://docs.google.com/uc?id=1xJDhcRCkgv7k4tNCa72Oog5bohV6dTB2&export=download`

- note that can click symbol under main layers in main window
  - and can right-click there and change style right away that affects color of hashed lines

outline

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 better 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)
- (note can also do point cluster; increase distance to 10mm
- reference:
  - http://www.qgistutorials.com/en/docs/creating_heatmaps.html
  - https://docs.qgis.org/2.8/en/docs/user_manual/plugins/plugins_heatmap.html
  - http:

outline

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_LAB
- 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,
  - lets 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/

layers-properties-metadata

- remember i was stressing this is important
- metadata=data about data
  - U/A, num of obs, etc
- and for now we'll skip the other tabs...