

thematic maps

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Saturday 6th January, 2018 17:22

outline

misc

nj counties and colleges data

basic thematic mapping

old ps2 comments

>>>[2nd class starts here:] basic descriptive statistics

classification methods

advanced thematic mapping

heatmaps

layers-properties; what else is there?

outline

misc

nj counties and colleges data

basic thematic mapping

old ps2 comments

>>>[2nd class starts here:] basic descriptive statistics

classification methods

advanced thematic mapping

heatmaps

layers-properties; what else is there?

how is qgis so far?

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

outline

misc

nj counties and colleges data

basic thematic mapping

old ps2 comments

>>>[2nd class starts here:] basic descriptive statistics

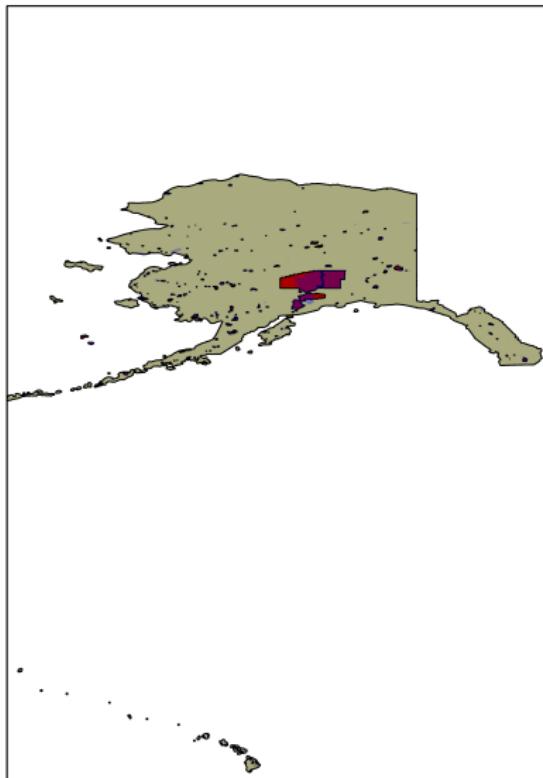
classification methods

advanced thematic mapping

heatmaps

layers-properties; what else is there?

Indians



Legend

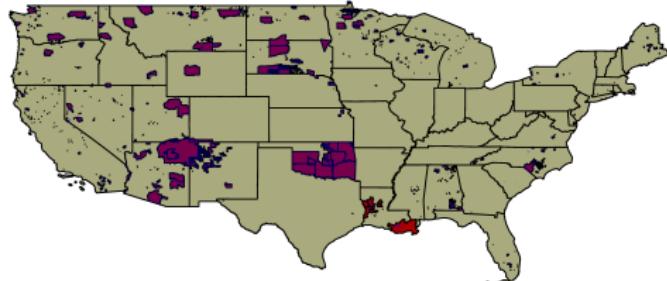
2012 American Indian/Alaska Native/Native Hawaiian Reserved Area



2007 American Indian/Alaska Native/Native Hawaiian Reserved Area



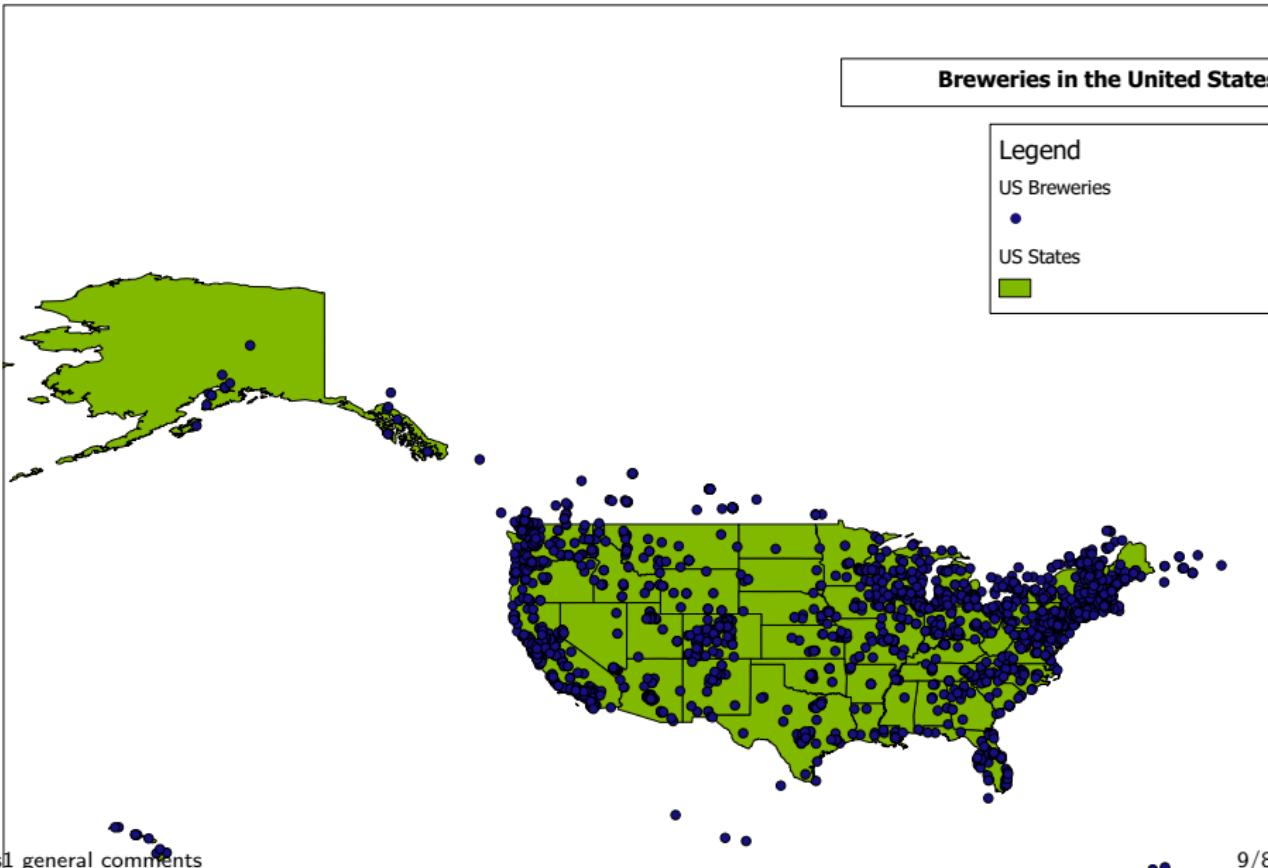
State Boundaries



Indians

- ◊ nice map ! can improve:
 - smaller font
 - transparent colors to see time difference
 - zoom in more
- ◊ future research:
 - longer time span, say hundreds of years
 - also show population
 - and other sociodemographics

breweries

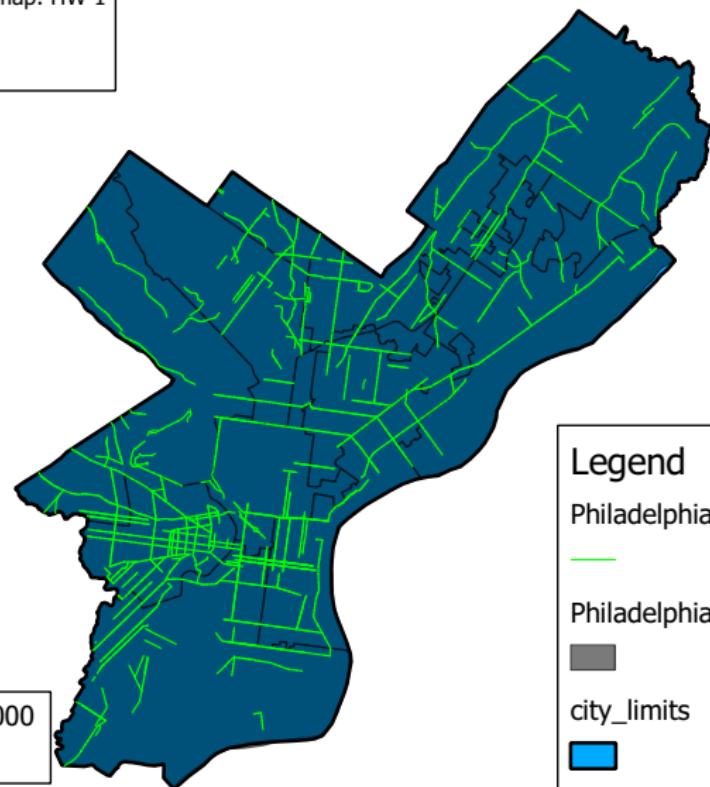


breweries

- ◊ I like the map ! can improve:
 - smaller points so that they don't overlap
 - drop non-US breweries
- ◊ future research: e.g. alcohol consumption and proximity to breweries
 - there are alcohol consumption data at CDC
 - do people consume more alcohol close to breweries ?
- ◊ maybe religion... (ARDA)
 - are there fewer breweries in religious counties ?

bike lanes

Philadelphia bike map: HW 1



Legend

PhiladelphiaBikeNetwork201204



PhiladelphiaCouncilDistricts_2000



city_limits



bike lanes

- ◊ very nice !
- ◊ can improve: beautify fonts, colors...
- ◊ future research:
 - add bike traffic (not sure if available, but can measure...)
 - add sociodemographics by tract
 - e.g. do young or rich or educated people bike ?
 - compare with other cities, e.g. Boston, Portland...

ps1

- ◊ for ps1 the idea was just to load the data
- ◊ and display it on a map
- ◊ we'll be doing it throughout the course
- ◊ but i will be more picky...

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 !
 - e.g. do we have small breweries that are at some bars ?
how exactly is a brewery defined ?
 - e.g. 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

map must be pretty

- ◊ what is a pretty map? subjective, but...
- ◊ zoom in properly, so that I can see the features
 - if they are too small to see, it does not make sense
- ◊ use space efficiently
- ◊ use lines and fonts of proper size

PA roads

- ◊ have legend that says that black lines are tracts and blue are roads
- ◊ should probably drop green background color for readability
- ◊ great! :
- ◊ “sought out a reliable url with a preference of using data from a .gov url.”

questions

- ◊ a question was how to deselect features:
 - there is a tool with red color for deselecting
 - let's select and deselect something
- ◊ some people struggled finding gis data
 - again, email listserv! and see data_sources.csv under today's class

other comments

- ◊ do not use exclusively variables from datasets that i have done in class
- ◊ that is, you cannot just submit something that i have done in class

outline

misc

nj counties and colleges data

basic thematic mapping

old ps2 comments

>>>[2nd class starts here:] basic descriptive statistics

classification methods

advanced thematic mapping

heatmaps

layers-properties; what else is there?

data

- ◊ let's get some NJ data
- ◊ it all comes from

https://njgin.state.nj.us/NJ_NJGINExplorer/DataDownloads.jsp

- (and there's more than we use here—hospitals, satellite pictures etc)

- ◊ I reposted them on my website:

http://people.hmdc.harvard.edu/~akozaryn/myweb/bounds_nj_shp.zip

http://people.hmdc.harvard.edu/~akozaryn/myweb/hsip_colleges.zip

- ◊ load all datasets (counties, municipalities, and colleges)
 - finest resolution (coll) at the top (move in layers window)
 - nj_munis-properties-style and increase “Layer Transparency”
 - now you can see darker county lines from nj_counties

opening/saving

- ◊ project-open or -save would save the project (.qgis)
 - important: this does not have any data
 - it just points to data on your drive
 - hence, if you email that to me
 - it won't work because it points to your hard drive

outline

misc

nj counties and colleges data

basic thematic mapping

old ps2 comments

>>>[2nd class starts here:] basic descriptive statistics

classification methods

advanced thematic mapping

heatmaps

layers-properties; what else is there?

thematic mapping

- ◊ this is really where fun begins
- ◊ and this is what this class is mostly about
- ◊ we use colors and/or symbols to display levels of variables' values on a map
- ◊ examples include red-blue state voting maps
 - we will be doing much more fancy examples

first, sorting the table

- ◊ note that you can sort the data in the table
- ◊ turn other layers off; nj_counties-open attribute table
- ◊ just click on the variable: POP2010
 - and highlight first 3; then last 3
- ◊ that way you'll easily see what's at the top and what's at the bottom
- ◊ that's the idea behind thematic mapping
 - to show U/As that are high and low on some var

your first thematic map

- ◊ nj_counties-Properties-Style and from drop-down:
“Graduated”
- ◊ Column: “POP2010”
- ◊ Color ramp: can just leave Blues
- ◊ and hit “Classify” button
- ◊ there is lots of ways to do the classification
 - we'll talk about it later
 - if you're curious—just poke around with classifications
 - easy to experiment and figure things out there
- ◊ and hit “OK” to see the map—viola!

print the map

- ◊ project-new print composer; first draw rectangle with “Add new map”
- ◊ draw rectangle with “Add new legend”
- ◊ again, you always need the legend

editing the legend

- ◊ (make sure you selected it (clicked it))
- ◊ you must uncheck “Auto update”
- ◊ and you can now highlight “nj_counties”
 - and click NOTEPAD ICON to edit
- ◊ likewise, we can edit out excessive zeroes for each color
- ◊ and, put commas between thousands—I will be picky about it
 - (keep in mind clarity principle from the previous class)
- ◊ use space efficiently
- ◊ make the map as big as possible
- ◊ stick the legend somewhere efficiently

standardization-always think about the meaning

- ◊ this map is not very interesting, or it's even somewhat misleading
- ◊ 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
 - e.g. 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 load the same data again, so that we have 2 layers
 - and rename it to say county2 (otherwise we'll an error later)
- ◊ “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)
- ◊ counties2-Properties-Style, “Histogram” tab, hit “Load values” button
 - [qgis>2.10]
- ◊ 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

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 show 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
 - and they're averaged about across the county...

outline

misc

nj counties and colleges data

basic thematic mapping

old ps2 comments

>>>[2nd class starts here:] basic descriptive statistics

classification methods

advanced thematic mapping

heatmaps

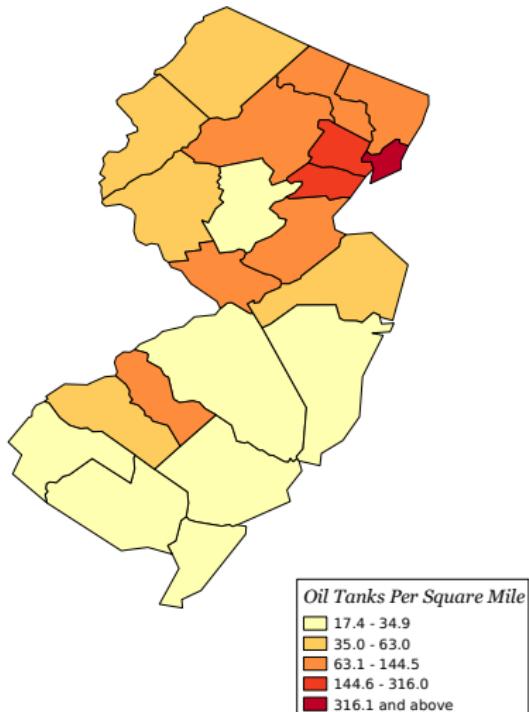
layers-properties; what else is there?

print composer again

- ◊ on the left green square button “move item content” move map within composer
- ◊ to resize it: on the right under item properties: just change scale or use a scroll button on mouse
- ◊ [http://gis.stackexchange.com/questions/11213/
how-to-zoom-in-qgis-print-composer-without-a-scroll-wheel](http://gis.stackexchange.com/questions/11213/how-to-zoom-in-qgis-print-composer-without-a-scroll-wheel)

oil heating

*Household Oil Consumption
to Heat Homes 2012
New Jersey (By County)*

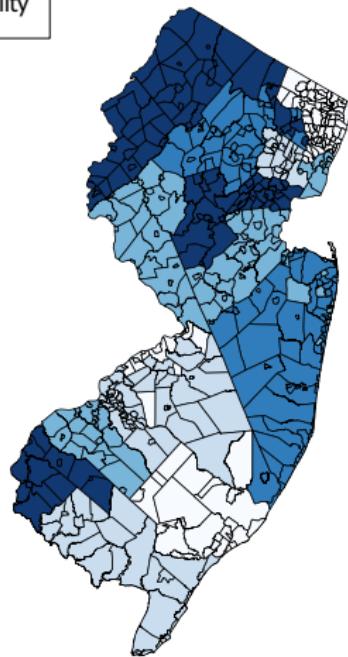


oil heating

- ◊ meaningful title; year included
- ◊ meaningful legend
- ◊ zoom the map as much as possible
 - if few u/a like here, less important; but always do it!
- ◊ it is interesting per sq mile (say fire threat)
 - yet it would be more meaningful per capita in general
- ◊ “There are five classifications, distributed in natural breaks.
If there were fewer classifications there would not be as
much distinction shown between the counties’ oil
consumption.”
- ◊ 5 natural breaks is usually a good choice

droughts

New Jersey Drought: By Municipality



Drought Frequency 1999-2004

- 0 - 110
- 110- 228
- 228- 349
- 349- 468
- 468- 587



droughts

- ◊ "natural breaks (jenks)" [...] minimizing the average standard deviation within each class and maximizes the standard deviation between each class. This allows your individual classes to contain relatively close values but the overall five classes are distinctively separated by their assigned values.
- ◊ What do we see in the map and what does it mean? [...] Clearly the north west is the region most affected. There are also areas in south west and central New Jersey. Areas in south central New Jersey are home to large areas of forests which would explain the reduced frequency of drought. Forest provide elaborate root systems which store and maintain groundwater sources.
- ◊ can overlay forest cover, population density, water basins, etc etc
- ◊ very important topic! we have global warming!

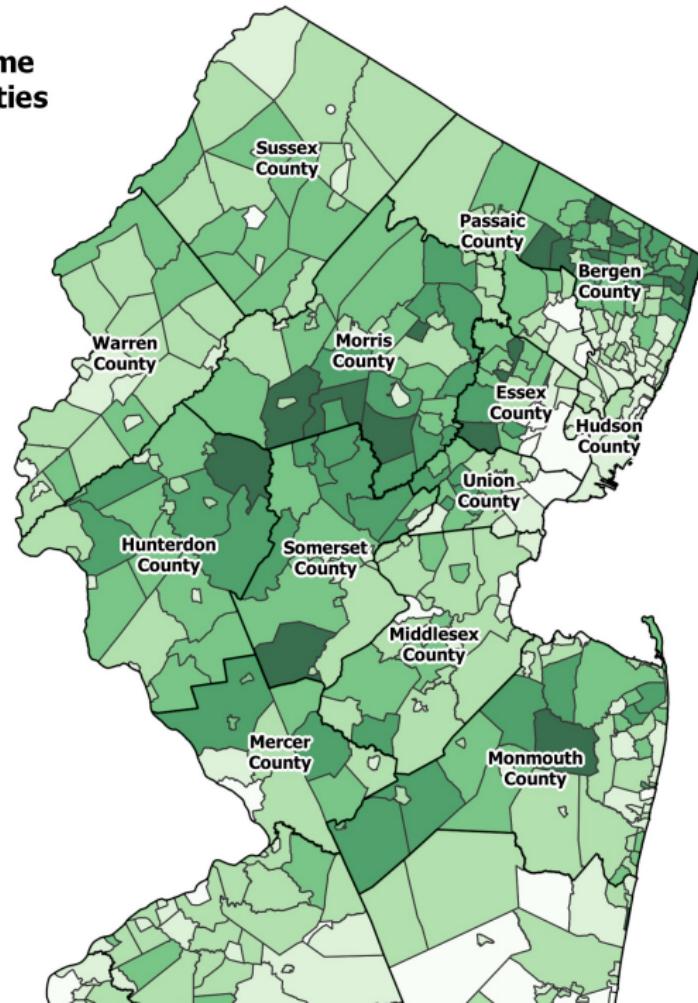
Median Household Income in New Jersey Municipalities

For 2006-2010
Median Income: \$67,681

Legend

Median Income Range

- \$25,682 - \$48,702
- \$48,703 - \$69,915
- \$69,916 - \$90,411
- \$90,412 - \$113,542
- \$113,542 - \$144,299
- \$144,300 - \$250,001
- No data available



income in NJ

- ◊ note, map cut to display detail
- ◊ nice title, we know timeframe
- ◊ legend: note decimal points; love yellow for missing data
- ◊ really nothing to fix; 2 things below *personal* preference:
 - use white borders for municipalities—little cleaner
 - use a 2 color ramp (say red-green) to signal poverty-affluence
- ◊ interesting: such big disparities so close to each other
- ◊ a note on level of analysis:
 - if county level data: N dark green; S: bright green
 - it would cover up deep poverty pockets in NYC metro!
- ◊ aggregation does matter! results are totally different

outline

misc

nj counties and colleges data

basic thematic mapping

old ps2 comments

>>>[2nd class starts here:] basic descriptive statistics

classification methods

advanced thematic mapping

heatmaps

layers-properties; what else is there?

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 values are distributed
- ◊ you need to know something about distributions
- ◊ you can get basics in qgis (use Statist)
 - or in qgis> 2.10 (eg apps.rutgers.edu) there is
 - histogram tab: layer-style- "graduated"

definitions

- ◊ 1, 2, 2, 3, 12 (right skewed)
- ◊ mode: most frequent value 2
- ◊ 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)

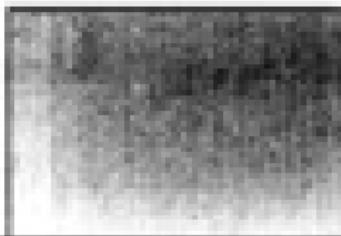
central tendency and skewness

- ◊ left skewed: $\mu < M$
- ◊ right skewed: $\mu > M$

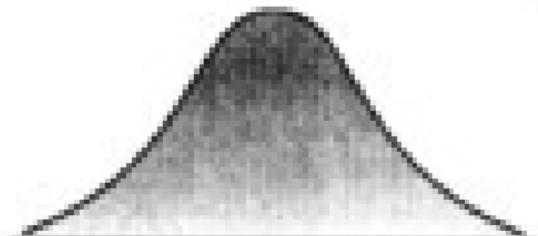
distribution types

- ◊ uniform
- ◊ normal symmetrical unimodal
- ◊ left skewed
- ◊ right skewed (income)
- ◊ bimodal

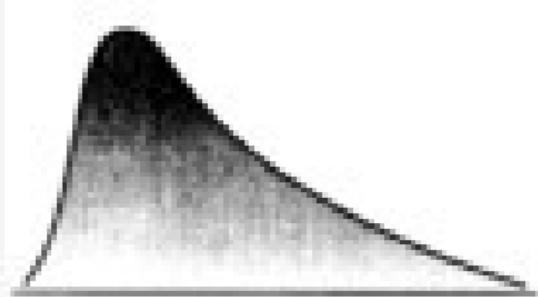
skew



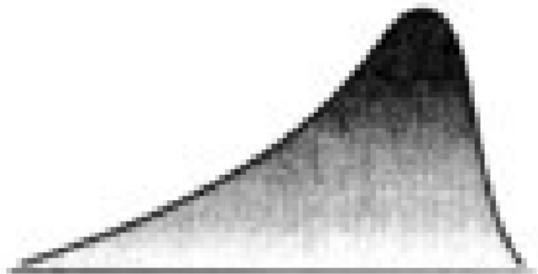
A



B



C



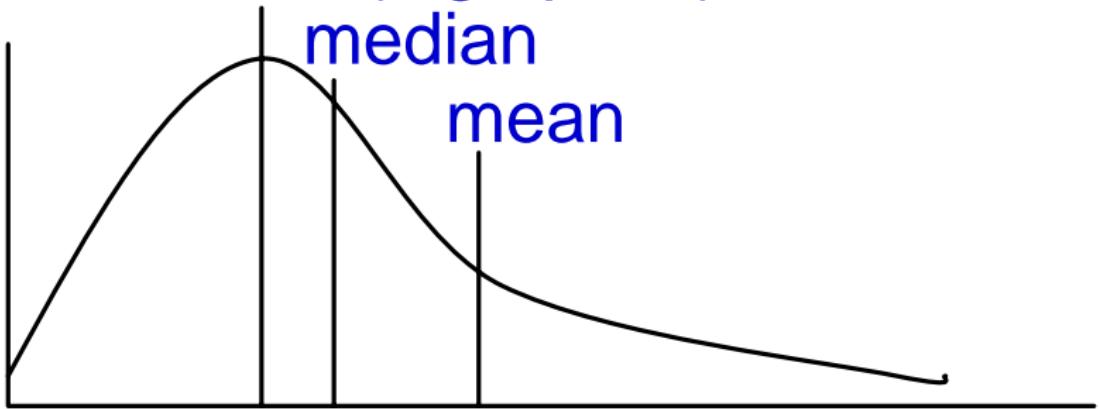
D

right skew

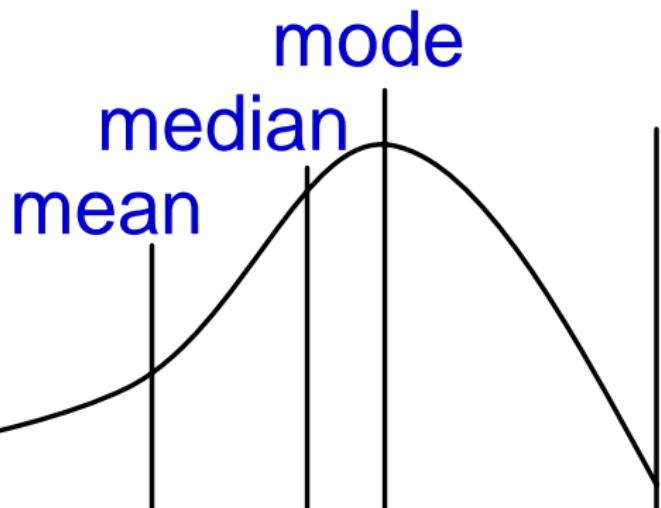
mode (high point)

median

mean



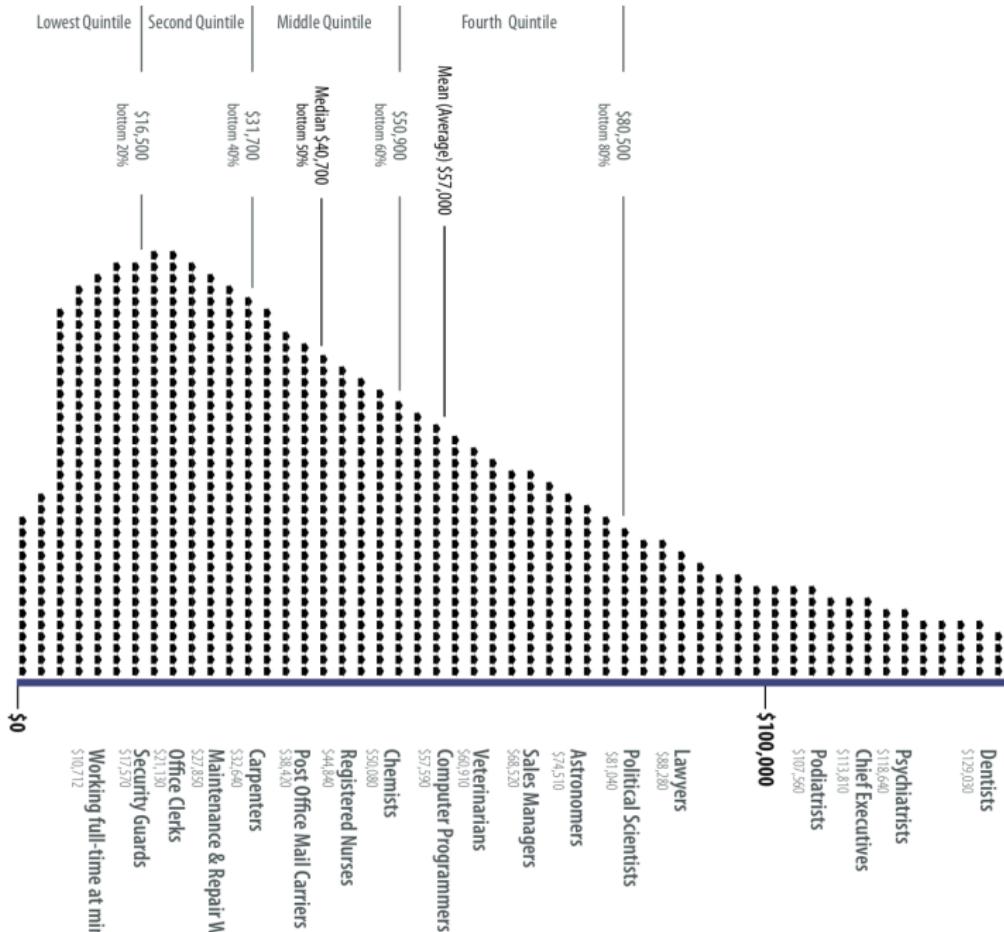
left skew



where are you on the distribution?

- ◊ http://en.wikipedia.org/wiki/Household_income_in_the_United_States#Household_income

income distribution



outline

misc

nj counties and colleges data

basic thematic mapping

old ps2 comments

>>>[2nd class starts here:] basic descriptive statistics

classification methods

advanced thematic mapping

heatmaps

layers-properties; what else is there?

references: very useful!

- ◊ let's open both and do first 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

nj counties and colleges data

basic thematic mapping

old ps2 comments

>>>[2nd class starts here:] basic descriptive statistics

classification methods

advanced thematic mapping

heatmaps

layers-properties; what else is there?

reference

- ◊ http://www.qgistutorials.com/en/docs/basic_vector_styling.html
- ◊ http://maps.cga.harvard.edu/qgis/wkshop/export_map_print.php
- ◊ <http://maps.cga.harvard.edu/qgis/wkshop/symbolize.php>

data

- ◊ our favorite NJ data
- ◊ all from https://njgin.state.nj.us/NJ_NJGINExplorer/DataDownloads.jsp
- (and there's more than we use here—hospitals, satellite pictures etc)
- ◊ I reposted them on my website:
http://people.hmdc.harvard.edu/~akozaryn/myweb/bounds_nj_shp.zip
http://people.hmdc.harvard.edu/~akozaryn/myweb/hsip_colleges.zip

classifying popden2010 again

- ◊ map popden2010 with
 - EQUAL INTERVAL
 - QUINTILES
 - NATURAL BREAKS/JENKS

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

histograms: Statist

- ◊ again, always think hard about the distribution of a variable that you are mapping
- ◊ histogram is one of the best tools

classification methods

- ◊ 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 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

nj counties and colleges data

basic thematic mapping

old ps2 comments

>>>[2nd class starts here:] basic descriptive statistics

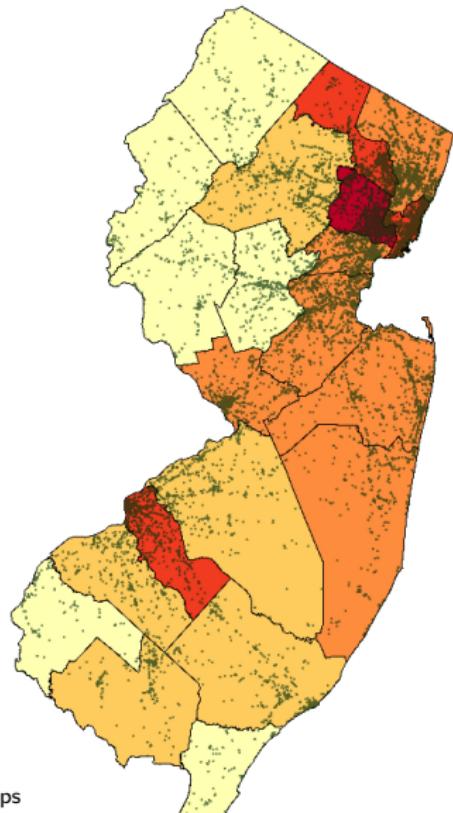
classification methods

advanced thematic mapping

heatmaps

layers-properties; what else is there?

Contaminations Sites in New Jersey 1992



Legend

- Poverty Status 1989
- Known Contaminated Sites
 - Counties in NJ
- | | |
|--------------|---------------|
| Light Yellow | 2766 - 7665 |
| Orange | 7665 - 20469 |
| Dark Red | 20469 - 35220 |

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!

do it at home: see links

- ◊ let's just scroll down through these screenshots:
- ◊ http://www.qgistutorials.com/en/docs/creating_heatmaps.html
- ◊ http://docs.qgis.org/1.8/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

nj counties and colleges data

basic thematic mapping

old ps2 comments

>>>[2nd class starts here:] basic descriptive statistics

classification methods

advanced thematic mapping

heatmaps

layers-properties; what else is there?

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

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...

have your own color

- ◊ load nj_munis, LAYERS-PROPERTIES
and zoom in on Jersey City
select POP_DEN10
STYLE-GRADUATED
- ◊ SYMBOL: transparency 50%
- ◊ COLOR RAMP-NEW COLOR RAMP-GRADIENT
i like yellow-red; or red-green is my new favorite
- ◊ these 2 opposite color show nicely bad-good say with
red-green

killing 2 birds with one stone

- ◊ again, thematic maps are much more exciting if you map things that interest you
- ◊ that way you research the topic of your interest
 - (you can use it for other classes or towards your thesis)
 - and you practice the gis software for this class