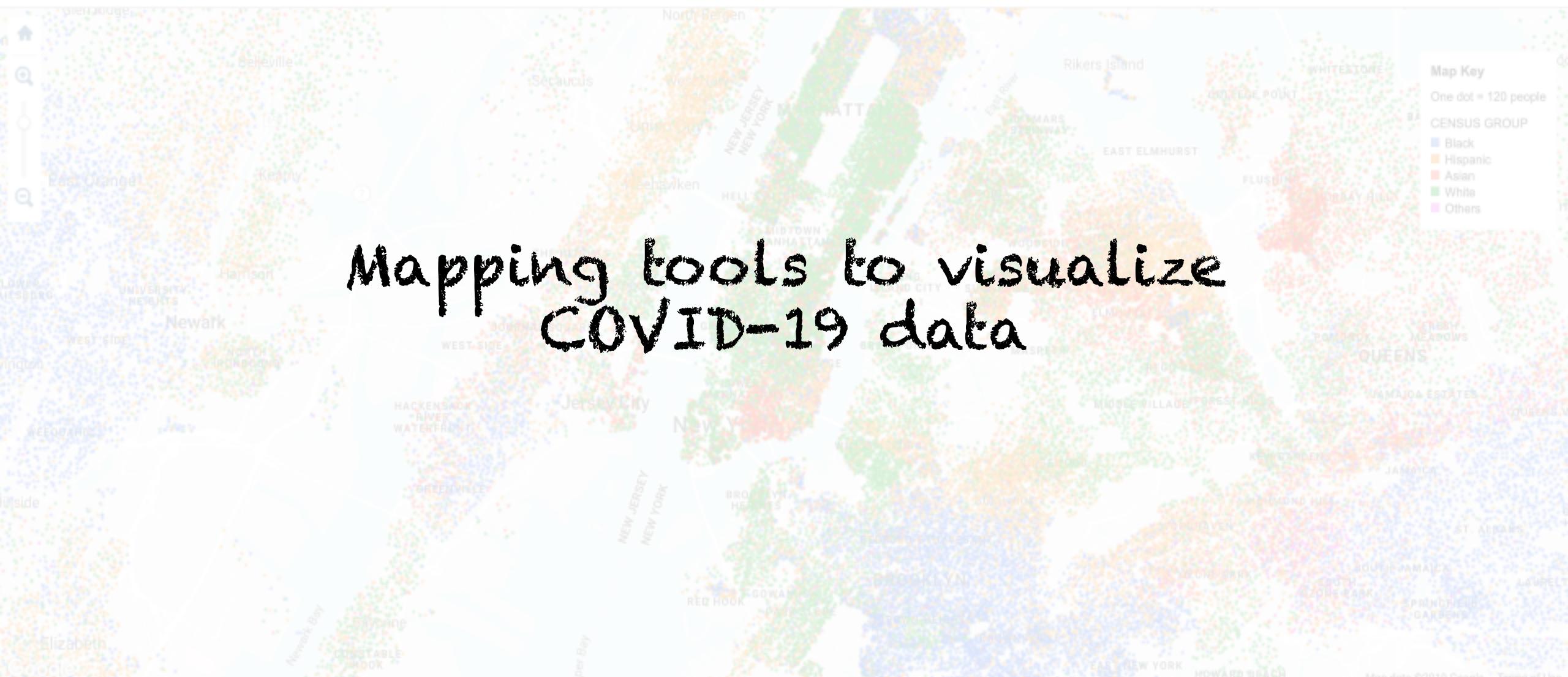


Mapping Segregation

New government rules will require all cities and towns receiving federal housing funds to assess patterns of segregation.



Mapping tools to visualize COVID-19 data



Map data ©2019 Google Terms of Use

Prerequisites

Familiar with vector data GIS

Know R and QGIS

Have R (3.6 or later), RStudio (1.2 or later) and QGIS (3.1 or later)

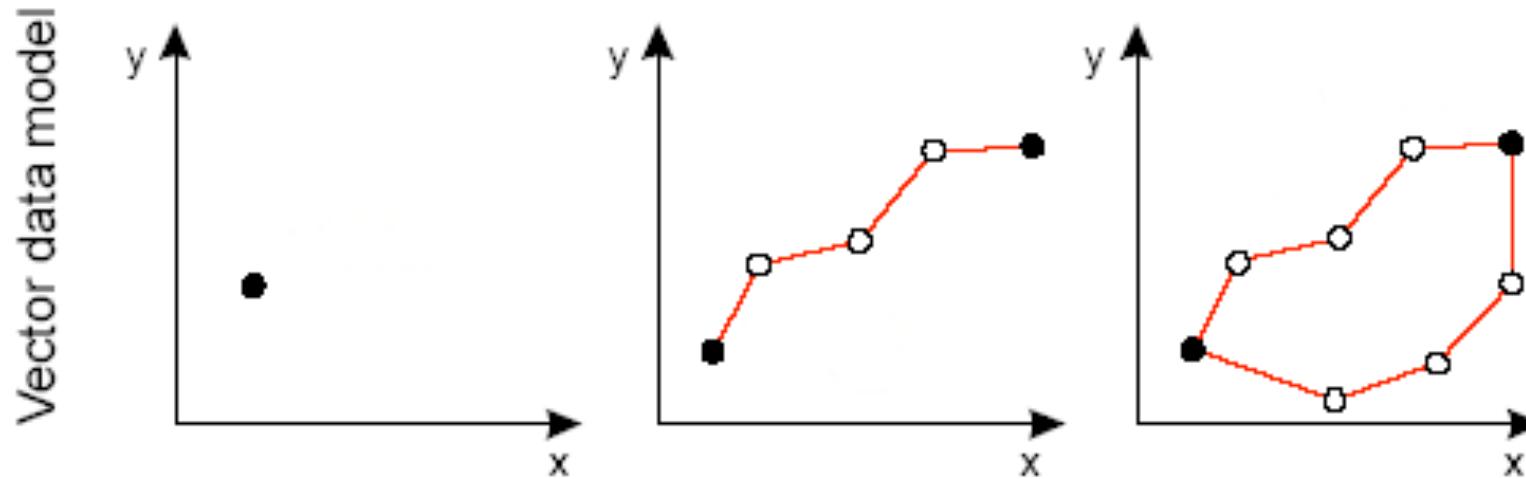
Have a Census API key

You have downloaded the data and the code that I shared

Vector data model

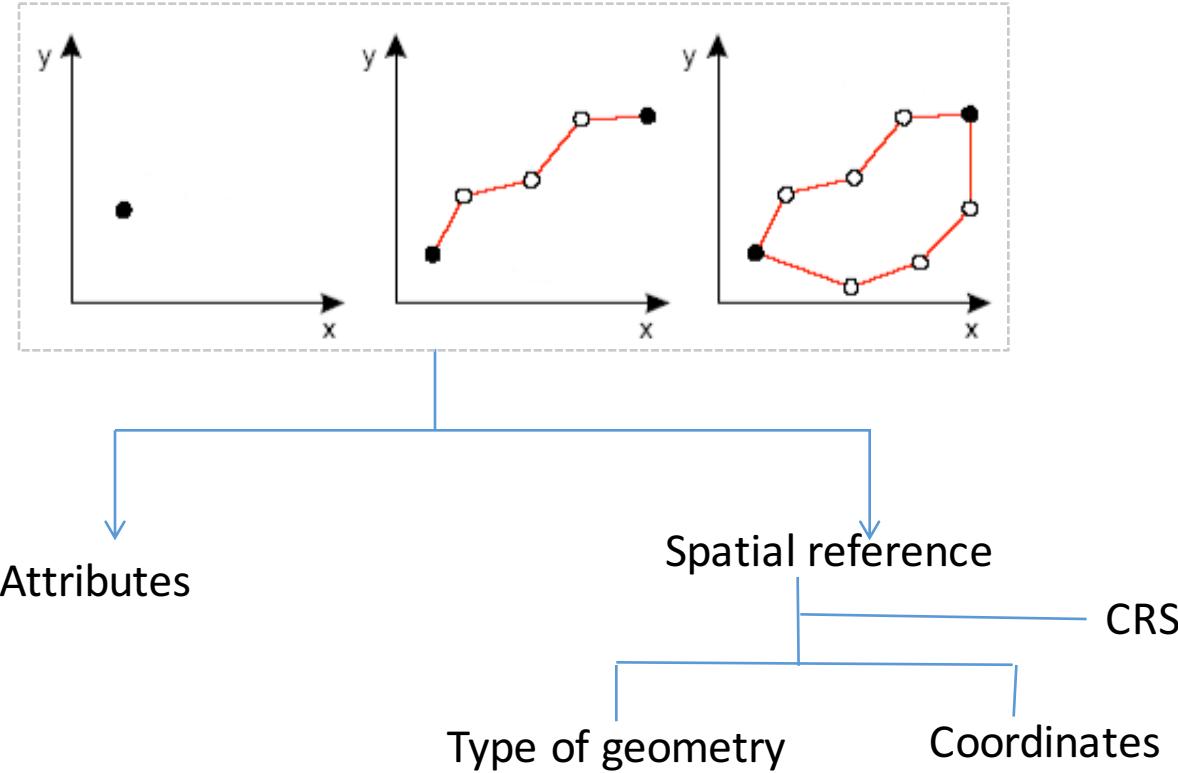
The world is made of discrete, mutually exclusive objects

All objects can be represented with three primitives - points, lines, and polygons



Objects have attributes and spatial reference

Geographic objects: geometry + CRS + attributes



No	Name	Sport	Country	Address	Type	Latitude	Longitude
1	Lionel Messi	Soccer	Argentina	Arroyo 841, CABA, Cdad. Autónoma de Buenos Aires, Argentina	Point	34.603° S	58.381° W

Vector data: SF data frame in R

Simple feature collection with 49 features and 5 fields

geometry type: MULTIPOLYGON

dimension: XY

bbox: xmin: -124.7332 ymin: 24.51496 xmax: -66.9499 ymax: 49.38436

epsg (SRID): 4269

proj4string: +proj=longlat +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +no_defs

First 10 features:

	GEOID	ALAND	cases	deaths	state_abbr	geometry
1	31	198956658395	10348	129	NE	MULTIPOLYGON (((-104.0535 4...
2	53	172112588220	19598	1016	WA	MULTIPOLYGON (((-122.3283 4...
3	35	314196306401	5978	265	NM	MULTIPOLYGON (((-109.0502 3...
4					
5					





R and QGIS



QGIS is bundled with the core functionalities of SAGA and GRASS, two other open source GIS software

Similar to ArcGIS, lags behind a bit (one to two years) in functionalities

You can use Python in QGIS

R is very powerful for advanced spatial analysis, especially for vector data, but may lack algorithms required for geometric analysis and manipulation

You can call QGIS functions in R via RQGIS (need 'reticulate')

My preference: Use R for data manipulation, analysis; QGIS for exploratory visualization; R for final maps once I have settled down on Ideas

Geo-visualization tools

Map is the quintessential tool for spatial and geo-visualization

Geographical network are also becoming popular

You can combine map and network to get greater flexibility

Combine space and time to show temporal changes over space

Our focus in this workshop is maps

If you are interested in 3D visualization check out CSCAR workshops offered by Andrew Hlynka



Paul Butler, Facebook



2012-2016

Maps

For a given data and objective, does an optimum map exist?

COVID-19 data

We will work with the COVID-19 data available from the GitHub site of the New York Times

We will use state and county level data for the US – there are different data bundles, we will use the one that is updated daily

I downloaded the data about three weeks back – we will use this data, available in the box folder, for the workshop so that all of us have the same data and there are no surprises

COVID-19 cases at state level: covid_state_level.R

Read state level COVID-19 data

Examine the structure of the data – understand what is recorded

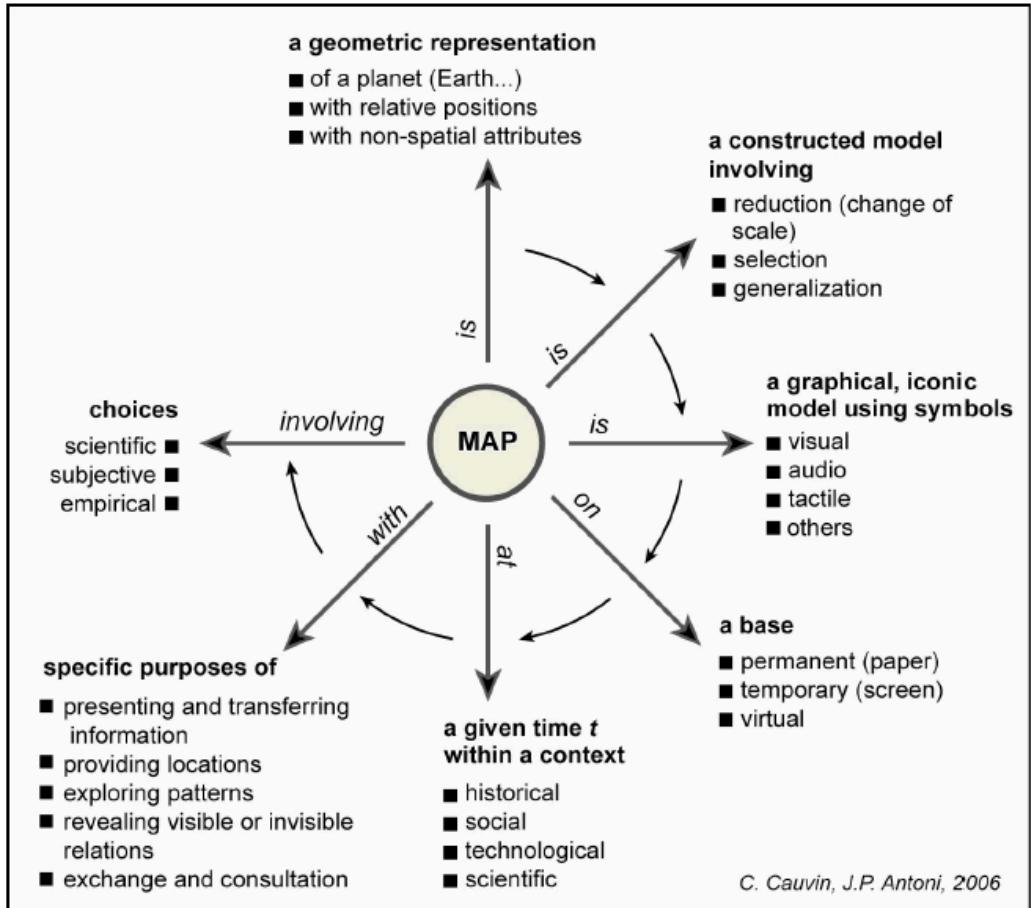
Do some sanity checks

Aggregate at state level and attach state level boundaries – make sure you know the CRS

Remove island territories, Alaska and Hawaii

Plot the geometry

What is a map



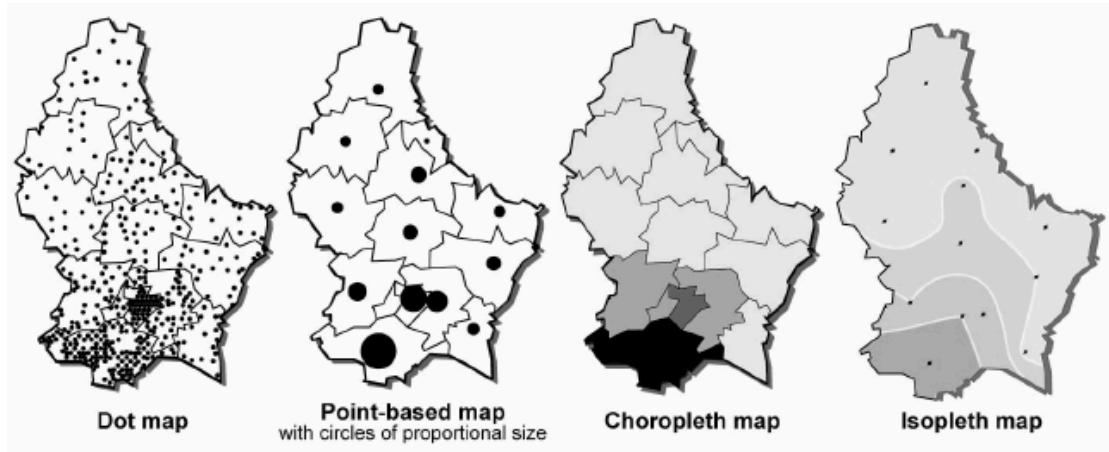
I like to think of map as a spatial model

This perspective forces you to reflect upon at least a few important elements:

- (1) What is the form of your 'model' and how it matches your objective
- (2) How do you interpret it
- (3) Developing a good model needs back and forth amongst data, objective, model, and interpretation
- (4) There may be more than one way to present the same data

Maps: Multiple representation of the same data

Two main components: space and attribute



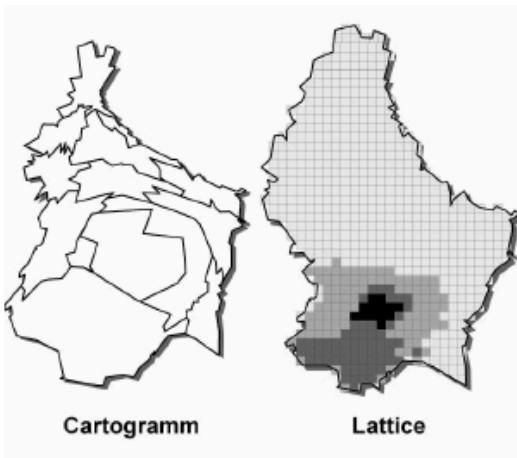
In the first three, the representation of the space is the same

Compare the dot map with the choropleth map

If you want to show local clusters, which one you would choose

We will cover Cartograms later

Suppose each dot represent a COVID case, where do think the risk is more or less?

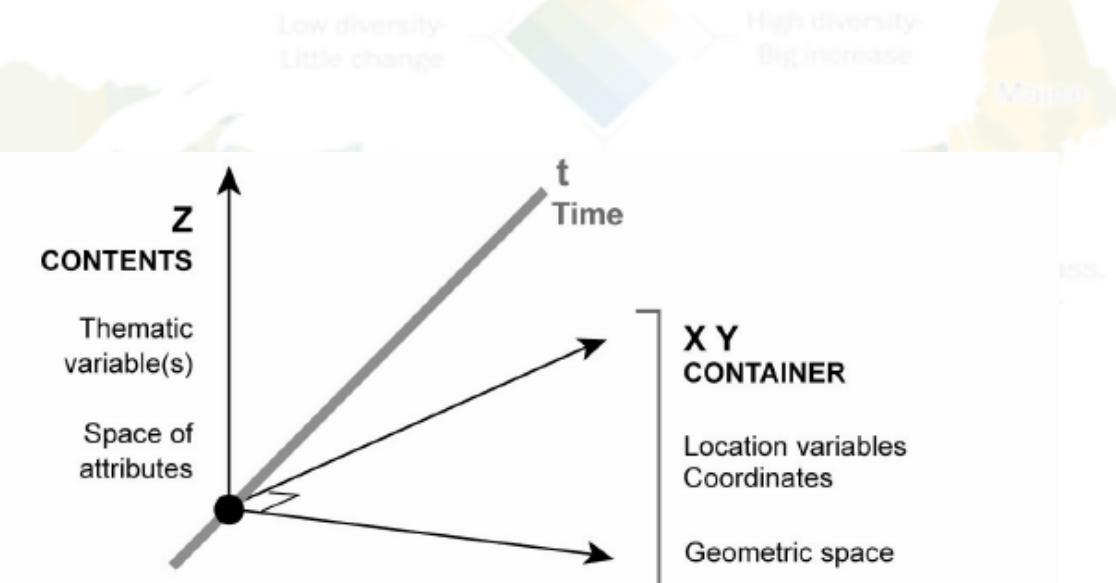


Single variable choropleth map

You have a spatially distributed variable 'P'

You want to show continuous or discrete variation in P over space

Map variation in P (continuous or discrete) to a color scheme



Decisions 1: Continuous or discrete (in digital world everything is discrete, but) variation. How to discretize, several options. Try to have an objective justification of the choice you make

Decision 2: Map the values of the variable to a color scheme

Single variable choropleth map: Optimal no. of classes

Higher number of classes, better representation: extreme one color for each unique value, but this may not be very informative

Discretization

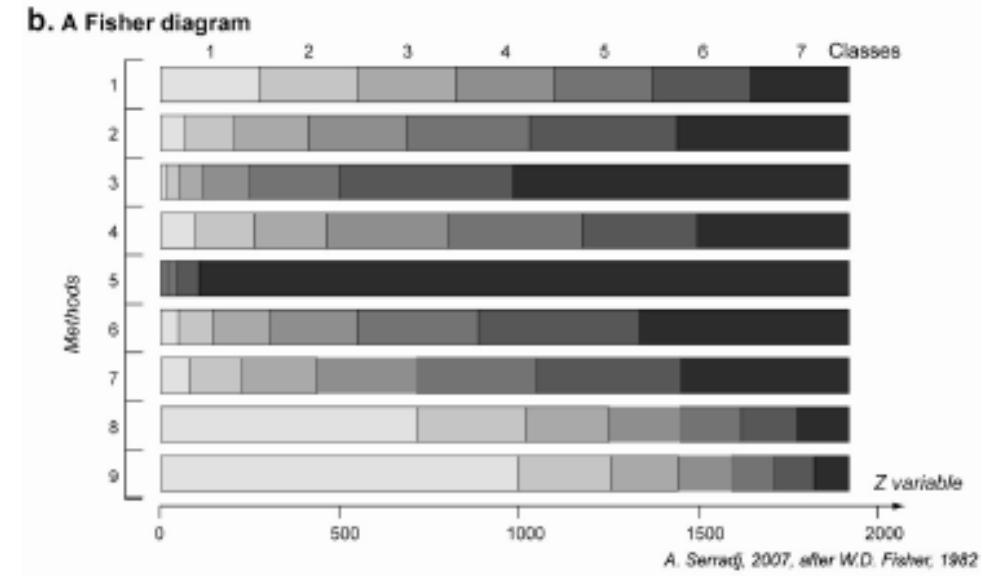
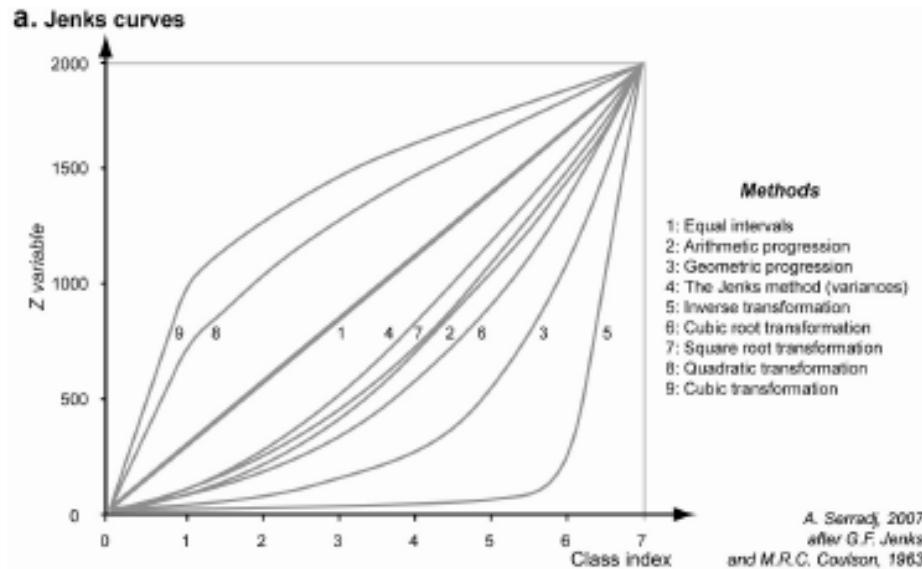
- Intuitive
- Exogenous categories (e.g. children, adults, old)
- Mathematical
- Statistical
- Experimental (*There is always some experimentation involved*)

Sometimes, you may need to transform the original variable to get better results, but the transformed variable should have an independent and desired interpretation

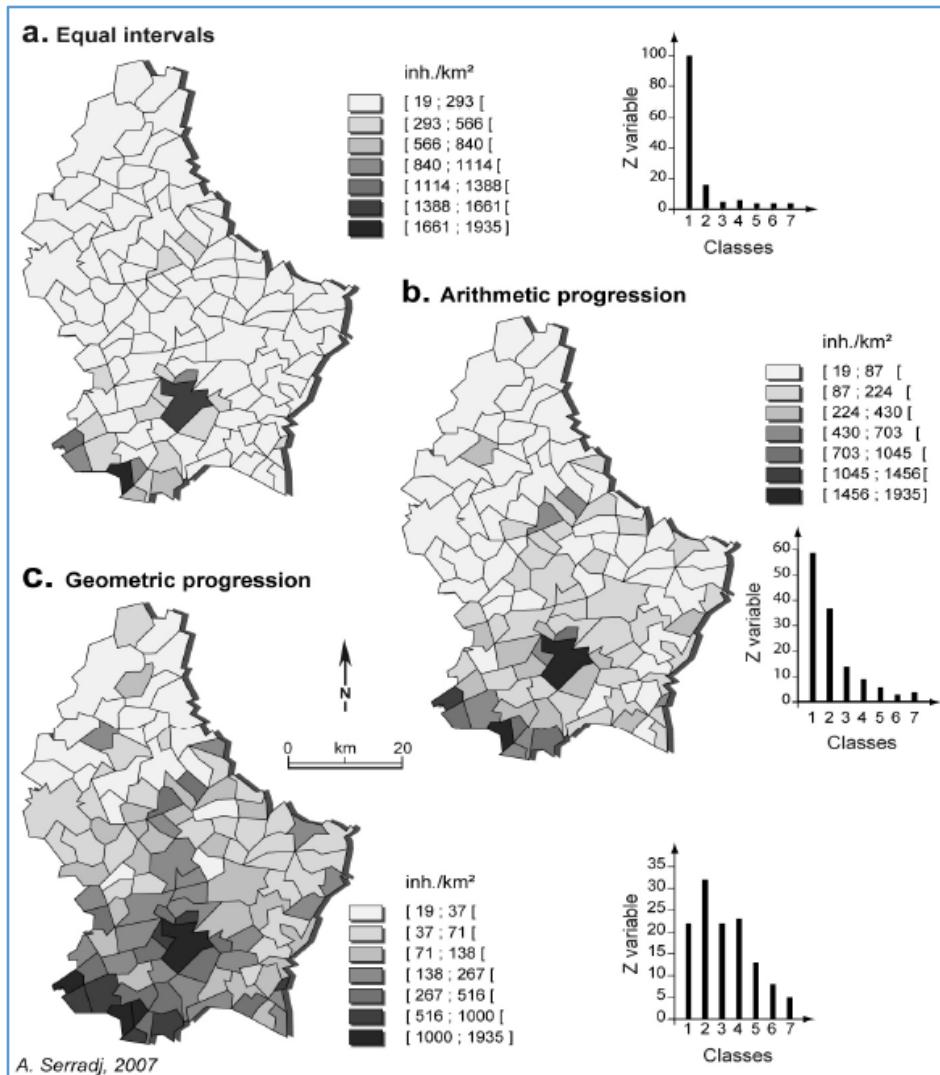
Single variable choropleth map: Discretization

Optimize the discretization and the associated color mapping, given your data and geometry

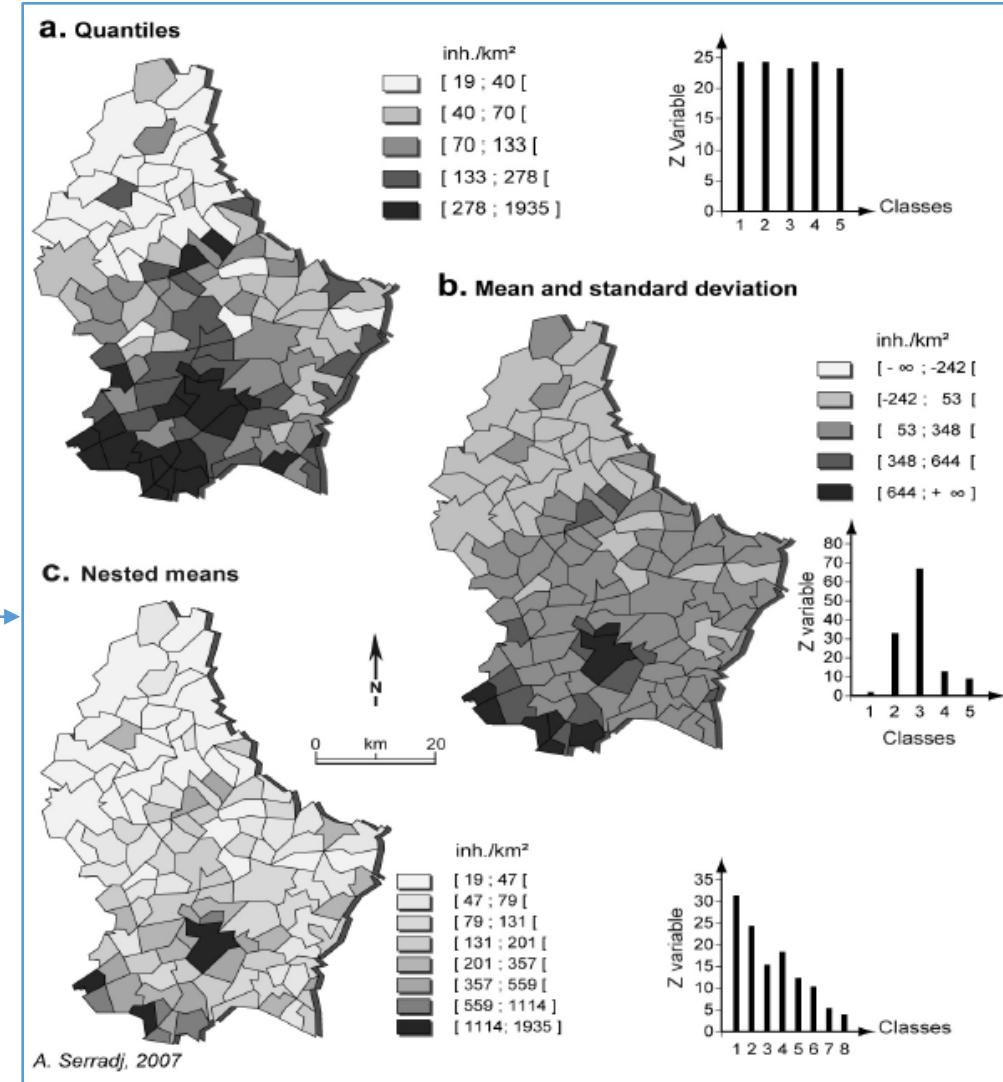
Suppose you have 'n' spatial units, the attribute value ranges between 0 and 2000, and you want 7 categories



Single variable choropleth map: Discretization



Matheamtical



Statistical

Mapping categories to a color scheme

Think of what you want to highlight, what should ‘pop out’ from your map to a reader

Think of mutual contrast, what do people associate a color with (may be cultural)

I do not know color science and especially how it is perceived by humans, but check out the references in the R Markdown files and the workshop box folder e.g.

Sequential Range: High to Low



Diverging Schemes: Distance from mean, +ve and -ve



Qualitative Scheme: Nominal categories, no order



Mapping categories to a color scheme

ColorBrewer

(<https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3>), is a nice package (Harrower M & Brewer CA; The Cartographic Journal)

R and QGIS packages

They have suggestions for color blind people

Original paper is a bit dated, but might still be useful

Sequential Range: High to Low



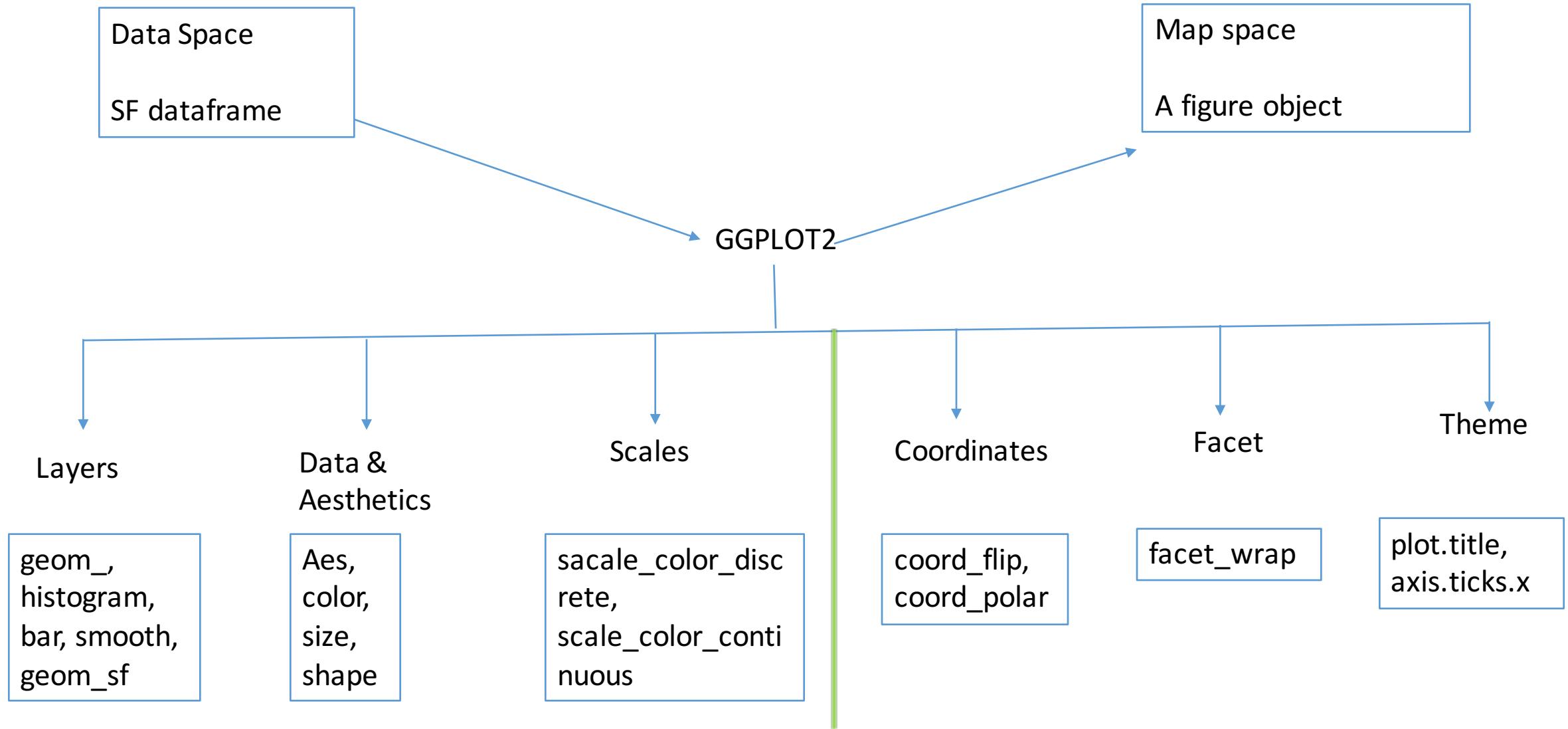
Diverging Schemes: Distance from mean, +ve and -ve



Qualitative Scheme: Nominal categories, no order



Mapping in R



Mapping in R

Make a map of case count and case ratio in R

(1) map_state_level_2.R

Map in QGIS

Load covid_state.gpkg

Style the map

Add a base map (from QuickMapServices – OSM Standard) – change hue and saturation of the base map

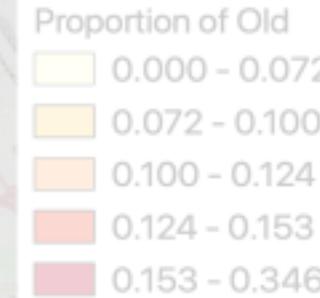
Try Google Imagery from HCM GIS as a base map – add label

Add a print composer

Add map, add legend

Finalize your map and save it as a .png file

Does your map convey what you intend it to do?



Make a webmap in QGIS

Load covid_state.gpkg

Add a column of the location of the time series graph in your sf dataframe

Give your layers human-friendly names in the Layers Panel

Change fieldnames in Layer > Properties > Fields > Alias

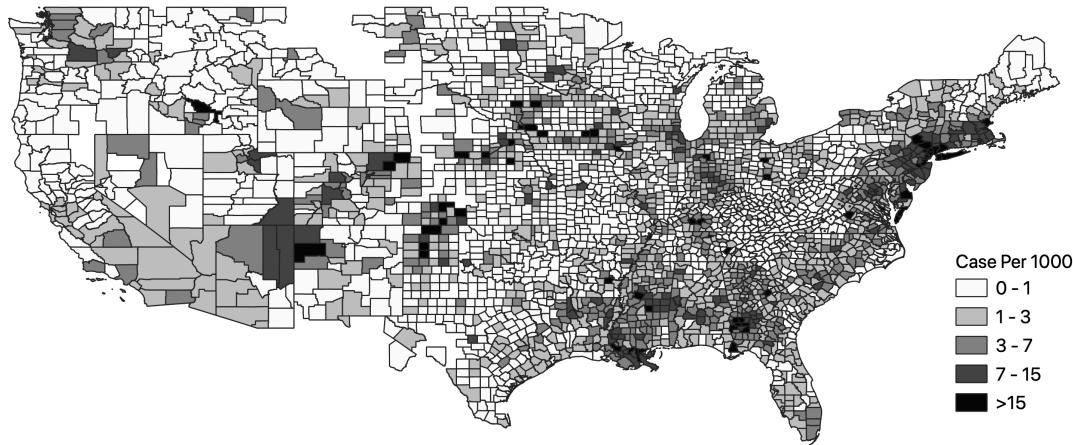
Hide the columns we do not want in popup set Edit widget to "Hidden" or change in attribute table

Change the widget for the graphs to "Attachment" and set the path

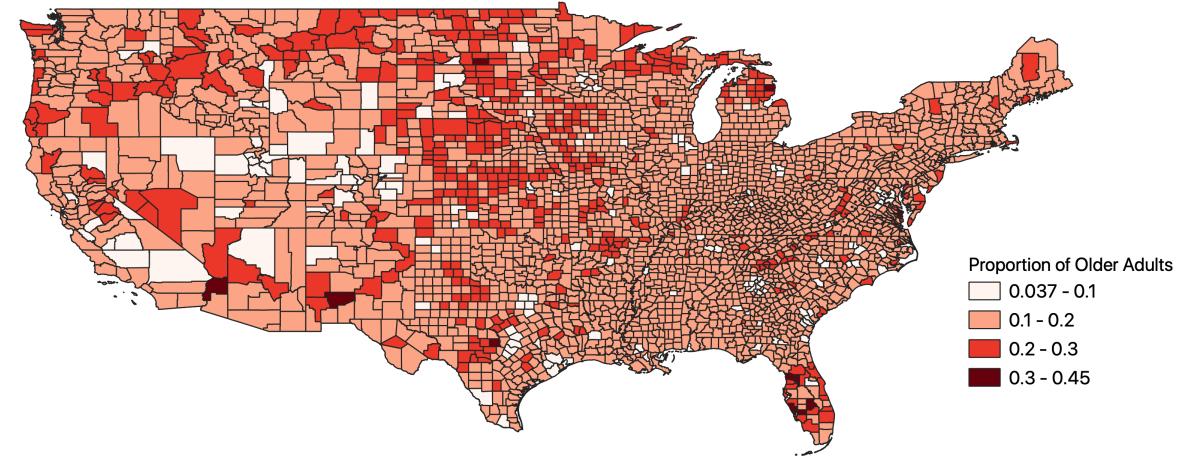
Map visualization: Bivariate Choropleth Map

Bivariate Choropleth Map

Case per 1000



Proportion of older adults



Combine the two to show joint variation in Case Per 1000 and Proportion of Older Adults

Bivariate choropleth map

You have two spatially distributed variable 'P' and 'Q'

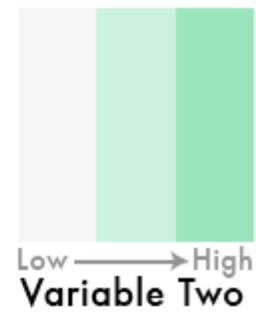
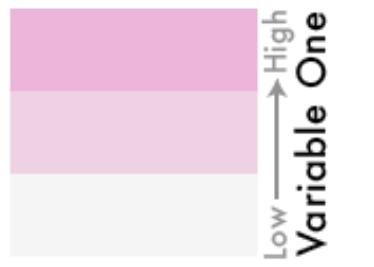
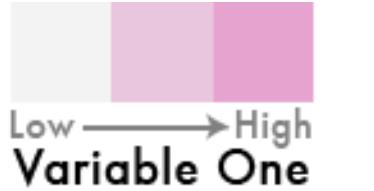
You want to show joint variation in P and Q over space

Discretize the two variables

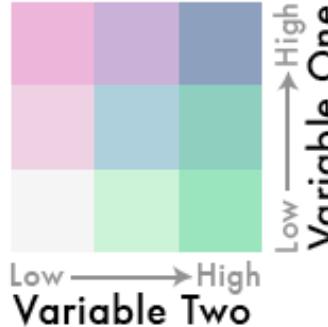
Case per 1000/ Proportion	P1 (older adults $p \leq 0.1$)	P2 (older adults $p > 0.1$ & < 0.2)	P3 (older adults $p > 0.2$)
Q1 (Low Cases)	P1 & Q1	P2 & Q1	P3 & Q1
Q2 (Average Cases)	P1 & Q2	P2 & Q2	P3 & Q2
Q3 (High Cases)	P1 & Q3	P2 & Q3	P3 & Q3

We can now map each category to a unique color and prepare a choropleth map

Bivariate choropleth map: Discretization and color



+



Choose two color gradients for each variable – preferably something light

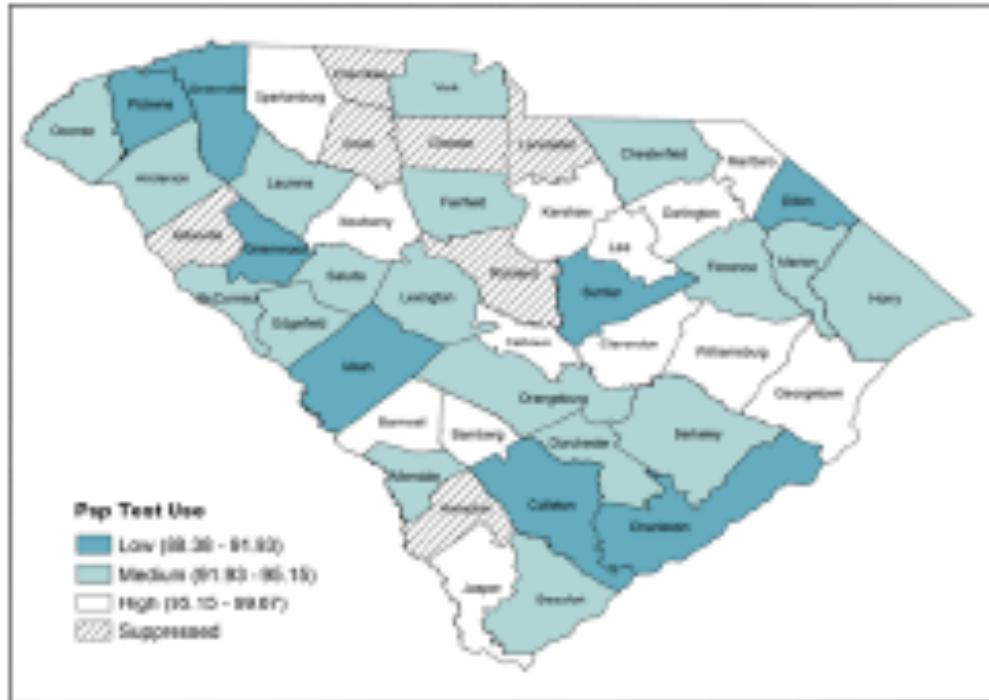
Combine the two to create the color scheme

It takes a good deal of time and effort to get the colors right

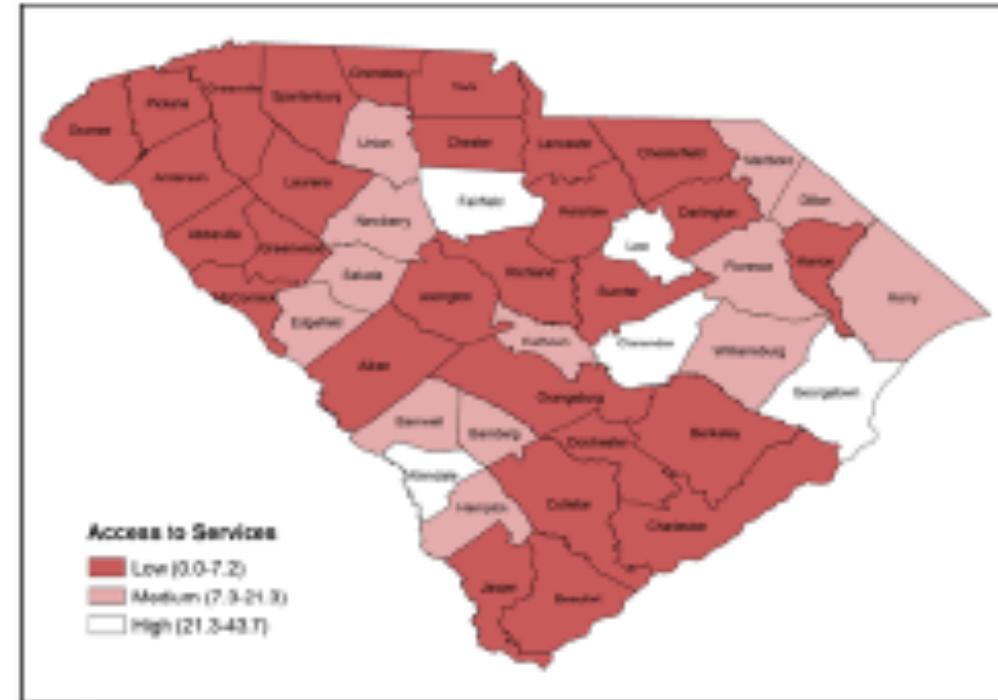
Based on <https://www.joshuastevens.net/cartography/make-a-bivariate-choropleth-map/>

Bivariate choropleth map: Example

Pap test use

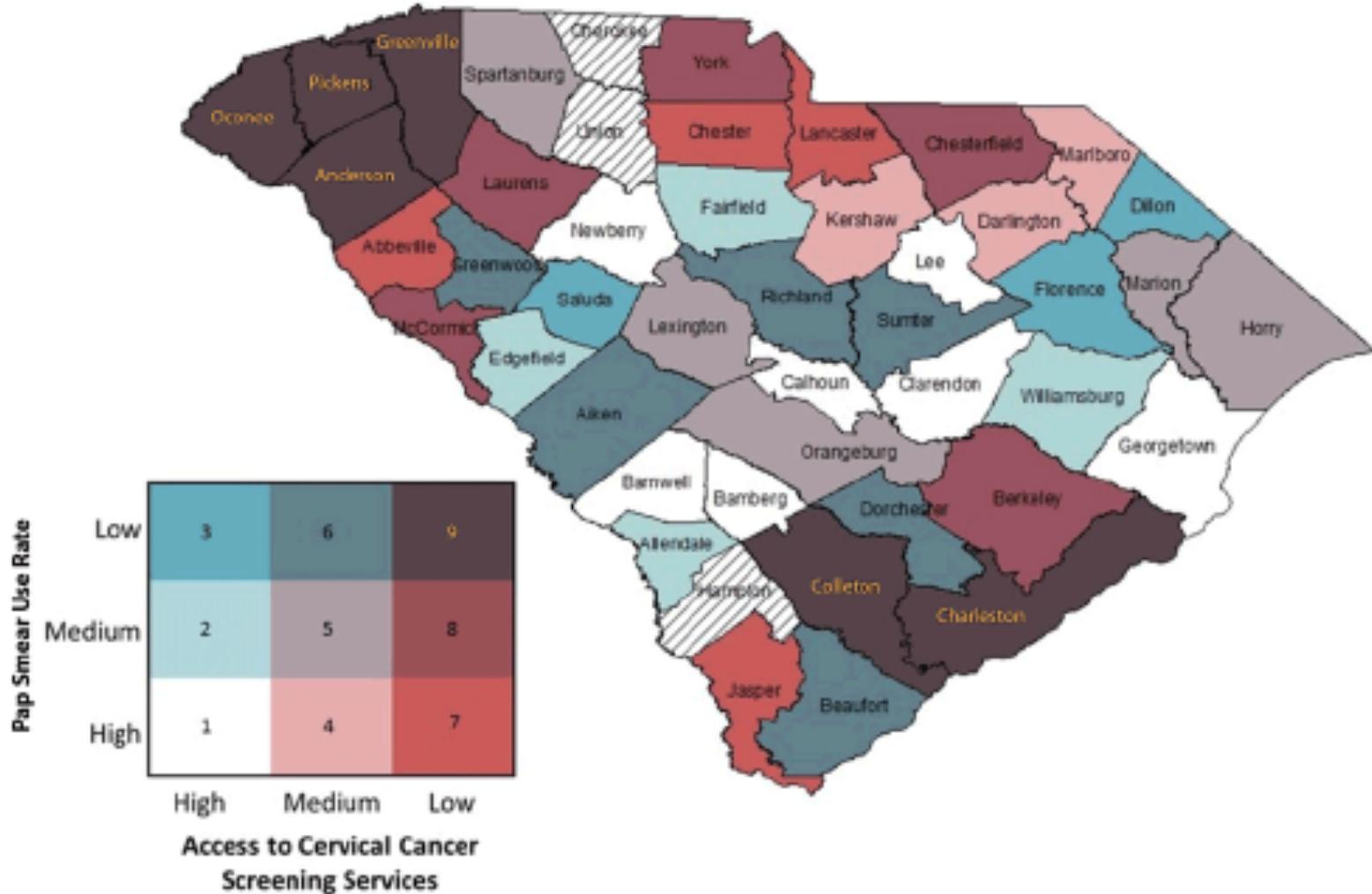


Access to service



Biesecker et al., 2020, CDC

Bivariate choropleth map: Example



Counties with
stripes had
insufficient data

Trumbo's principles

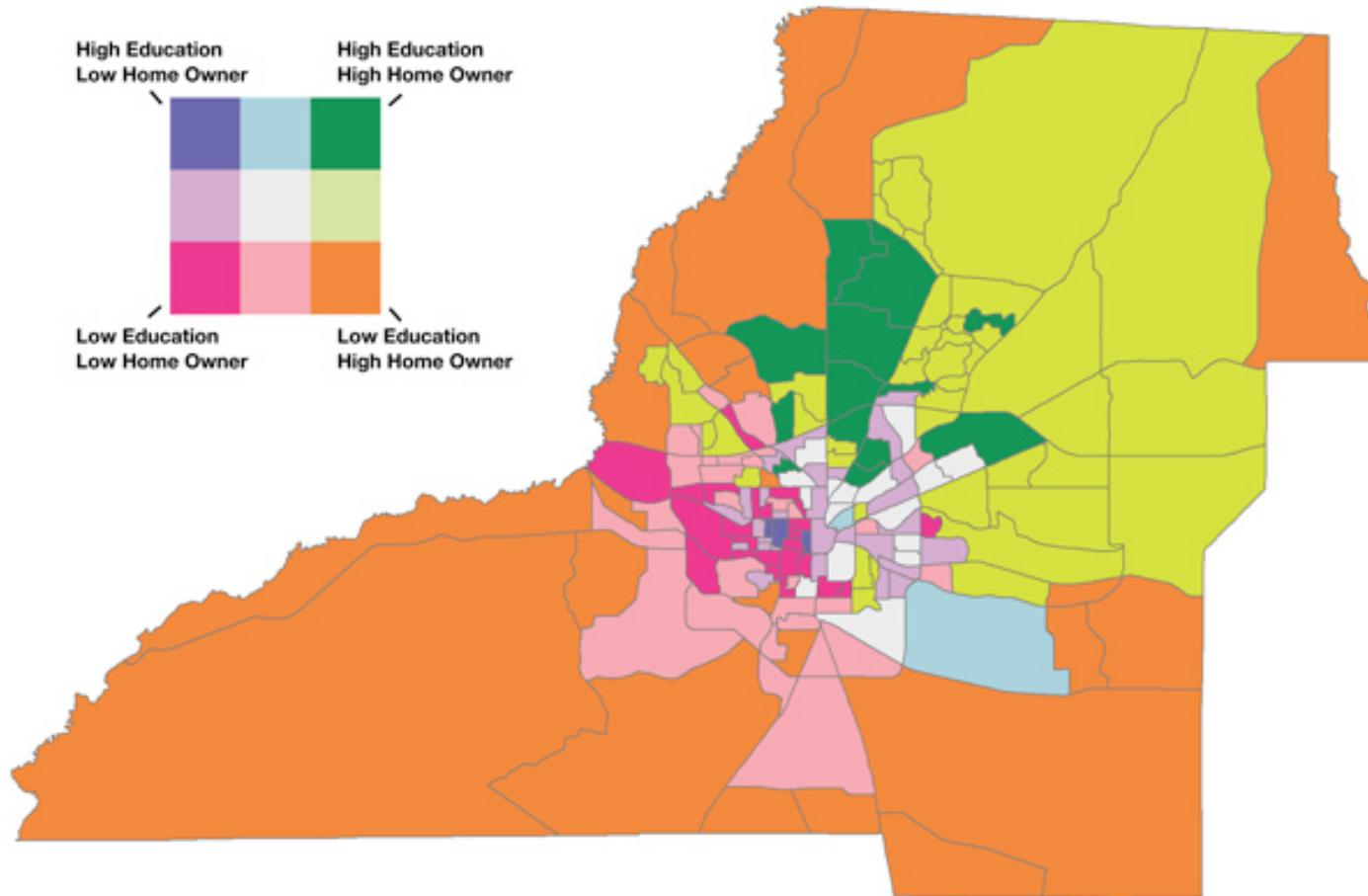
Type of Question	Inquiry Formula	Sample Questions
Inverse relationship	low/high of __ and/or low/high of __	Where are high elderly populations with low income?
Range of one variable within a category of another variable	range of __ within low/high of __	What is the range of education among high earners?
Direct relationship	relationship of __ and __	What is the relationship between income and education?

In a bi-variate legend each row and column should preferably be a visual sequence

If the focus is on the interaction between the two variables, the diagonal of the legend should be visually distinct from the rest of the legend

Let us go back and see the pep test figure

Trumbo's principles: Emphasis on corner values

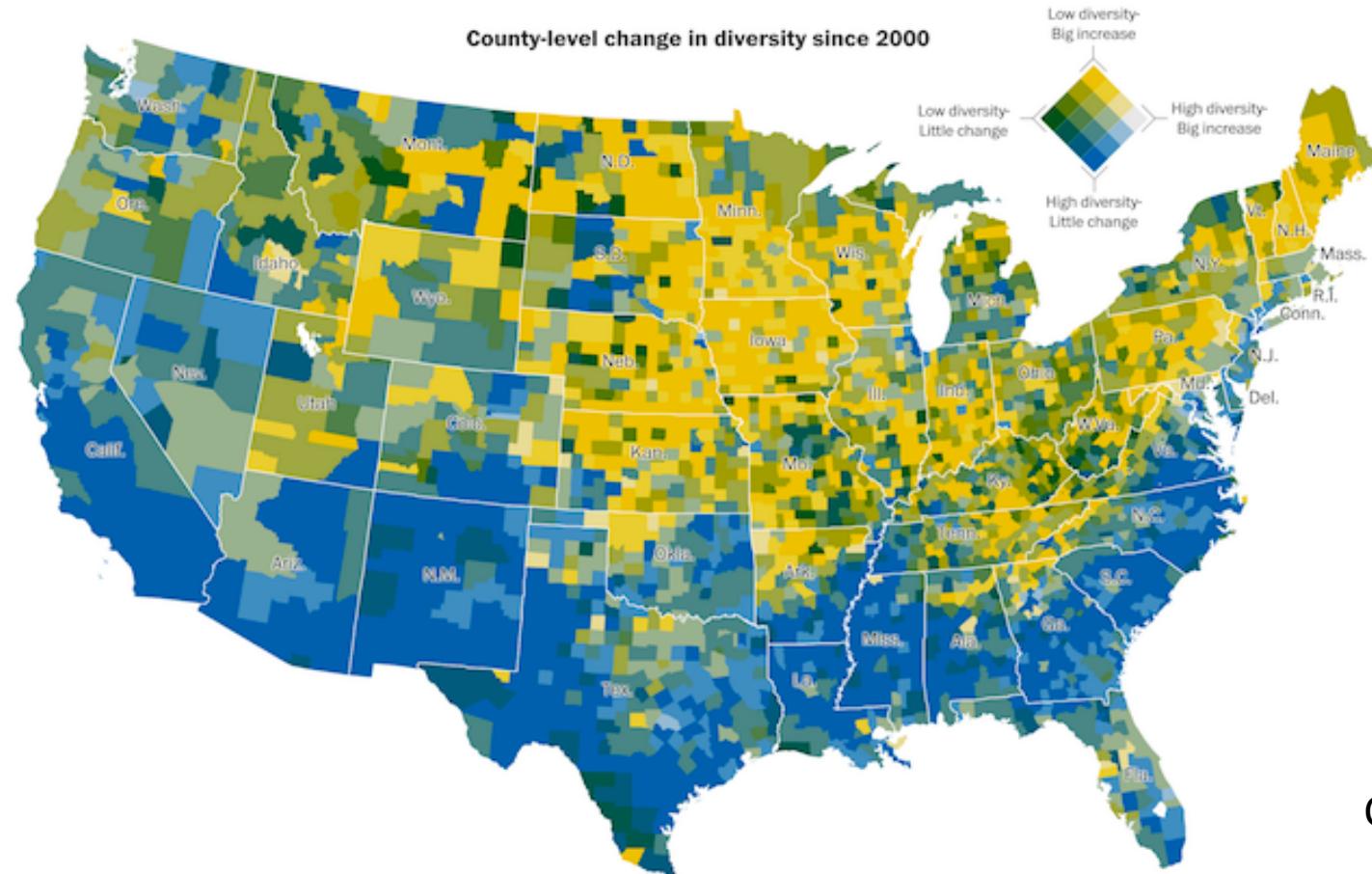


Read about the other two models - range model and diagonal model - in the paper at the url below

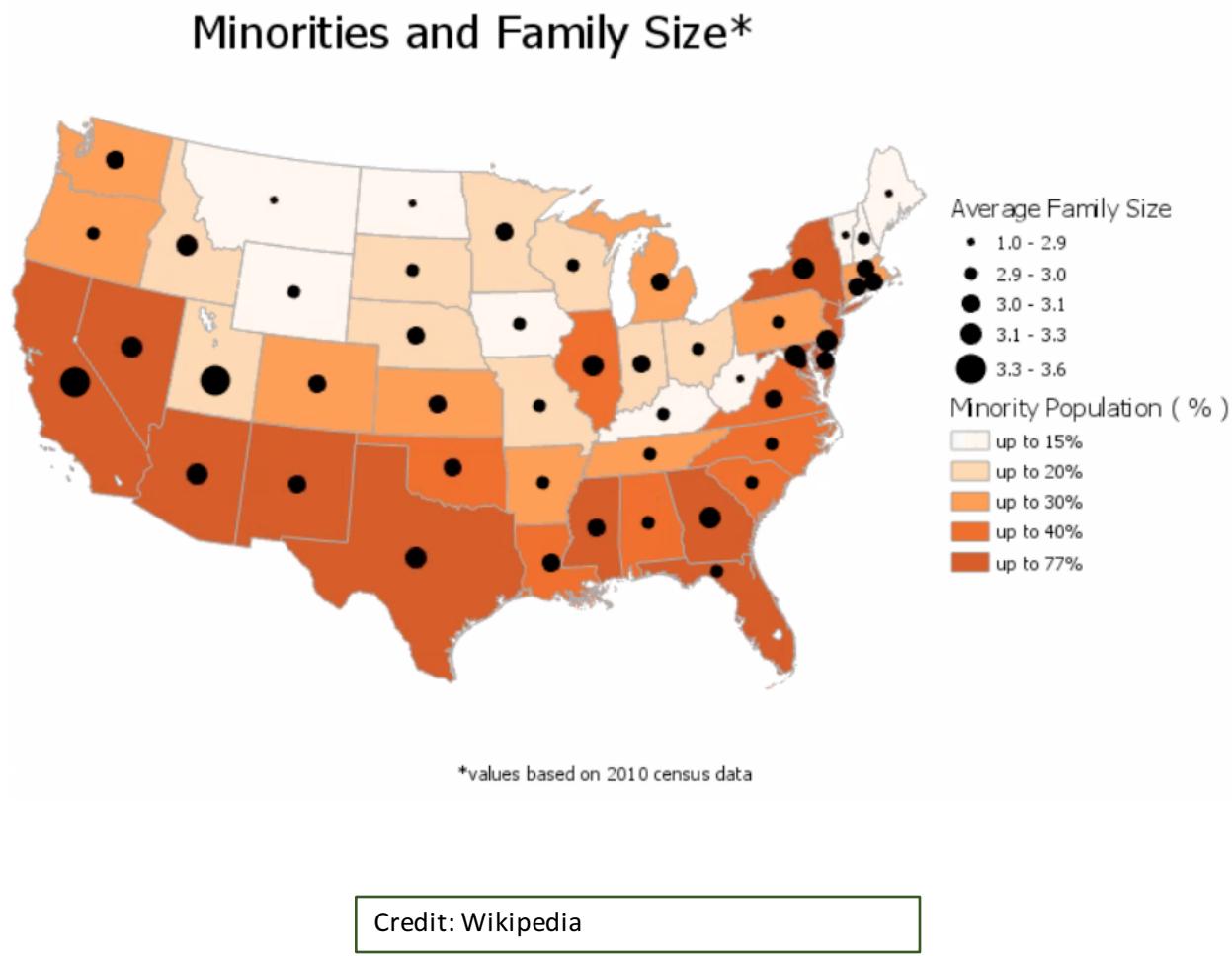
Bivariate choropleth map

The increasingly diverse United States of America

The racial and ethnic diversity of communities varies greatly across the country, but rapid change is coming to many of the least-diverse areas.



Bivariate map: Alternatives



One variable is mapped to color,
and the other variable to the size
of a symbol

Also check out value-by-alpha

<https://gistbok.ucgis.org/bok-topics/multivariate-mapping>

https://gistbok.ucgis.org/all-topics?term_node_tid_depth=97

Cartograms

I do not like Cartograms and am baffled by their popularity, but they have their utility

Value by alpha maps – alternative to cartograms

Both QGIS and R have libraries for creating cartograms

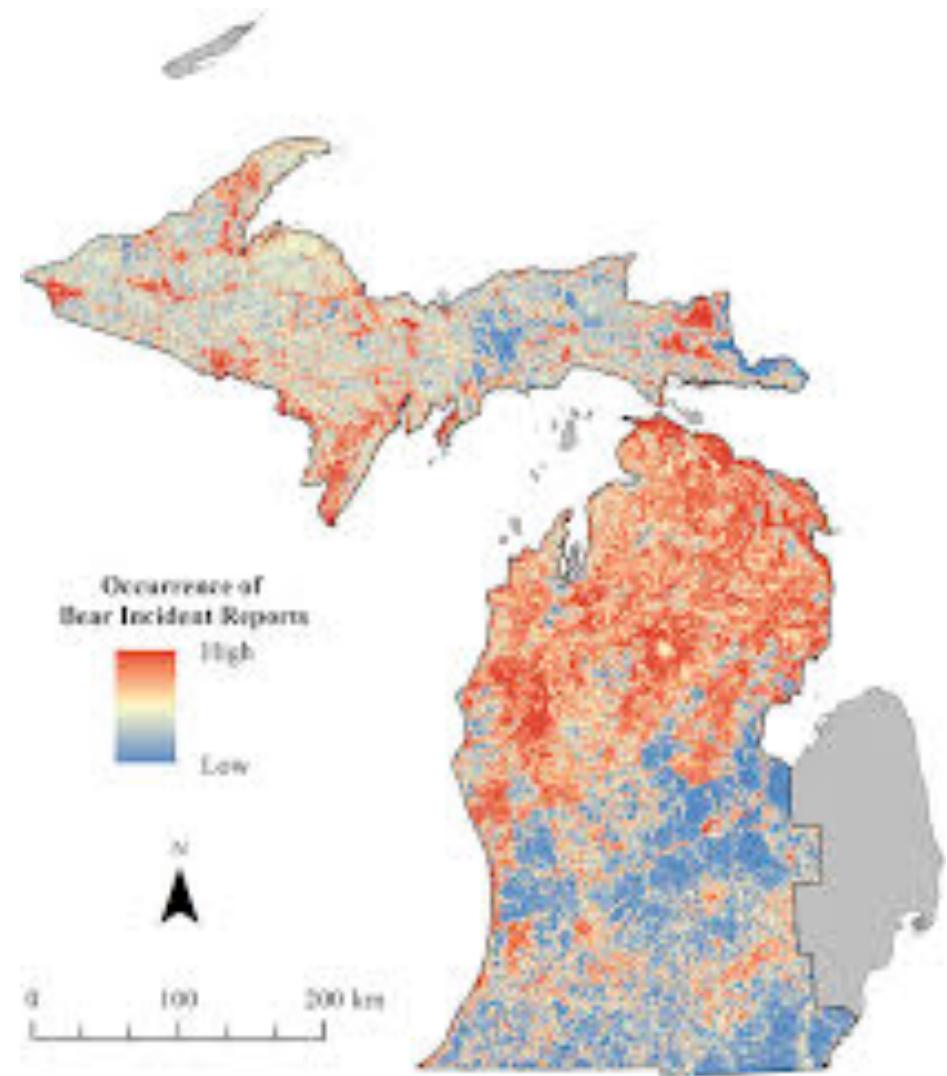


Density map or heat map

This is a raster representation of your data – like a continuous field

Bi-variate kernel density estimate

Not relevant really with this data, unless you are willing to make certain assumptions about how the data is distributed over space



[Image credit: anotherview.info](http://anotherview.info)