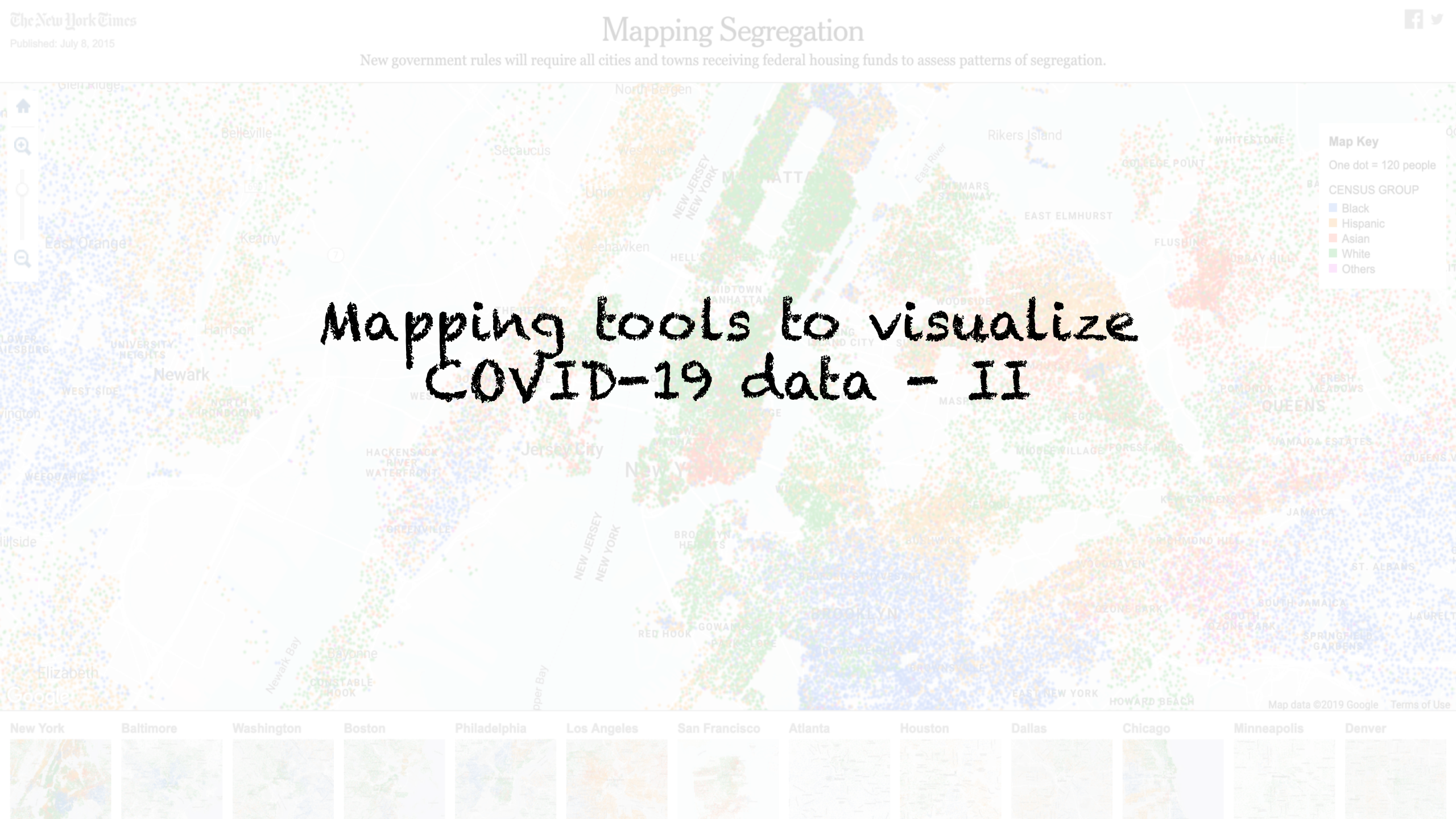


Mapping tools to visualize COVID-19 data - II

Prerequisites

Familiar with vector data GIS

Know R and QGIS

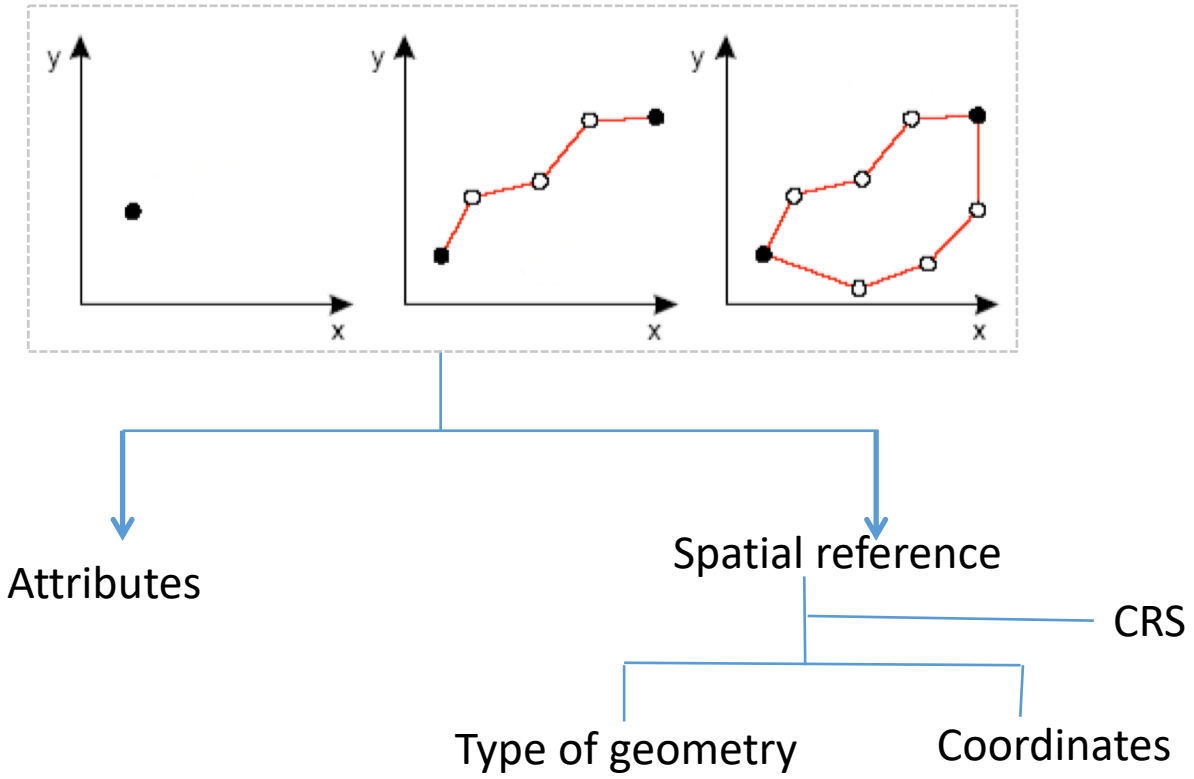
Have R (3.6 or later), RStudio (1.2 or later) and QGIS (3.1 or later)

Have a Census API key

You have downloaded the data and the code that I shared

You have a basic familiarity with singe variable choropleth map

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: -124.73 ymin: 24.514 xmax: -66.949 ymax: 49.384
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	.....					

If you simplify these geometries to component polygons – which of the following states will have more than one polygons: Nebraska, California, Florida, Colorado

Maps

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

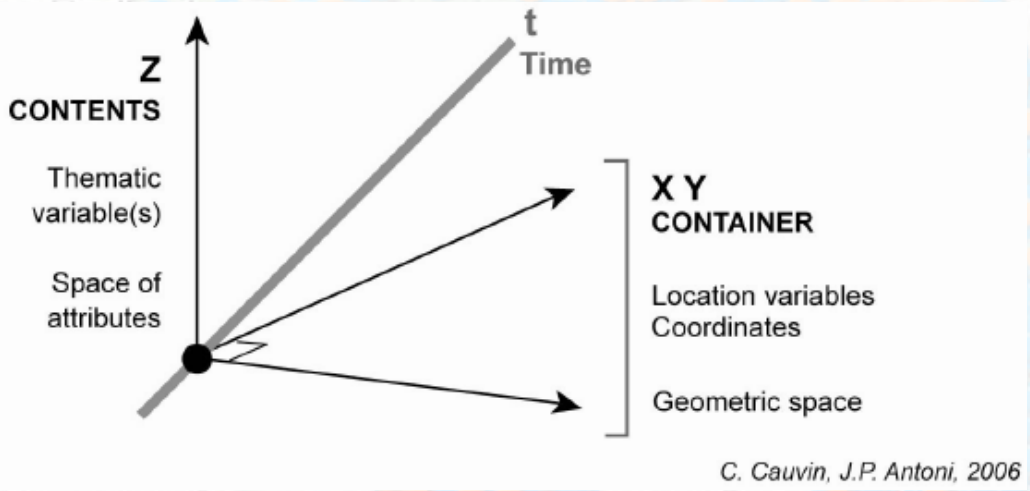
COVID-19 data

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 four days back (on July 1st or 2nd) – 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

What is a map



Thematic Cartography and Transformations; C. Cauvin, F. Escobar, and A Serradj

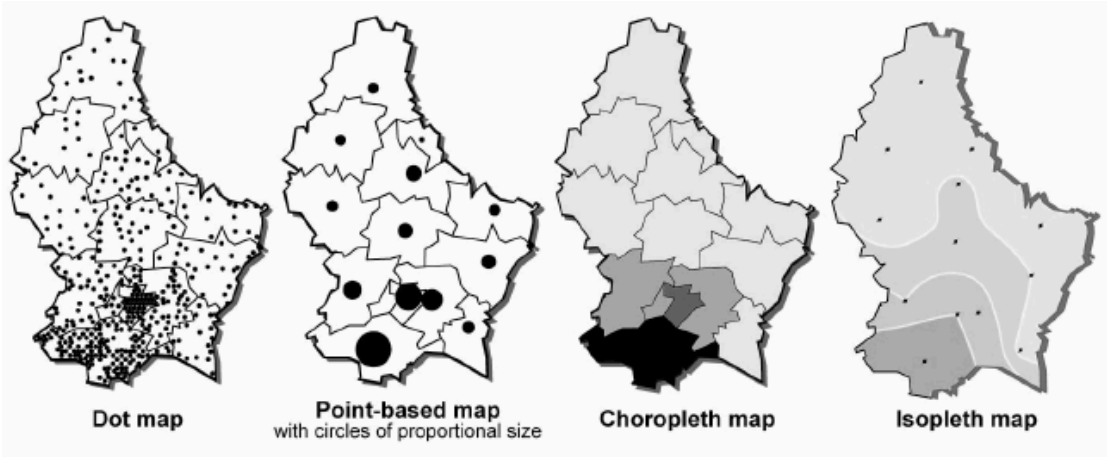
I like to think of map as a spatial model

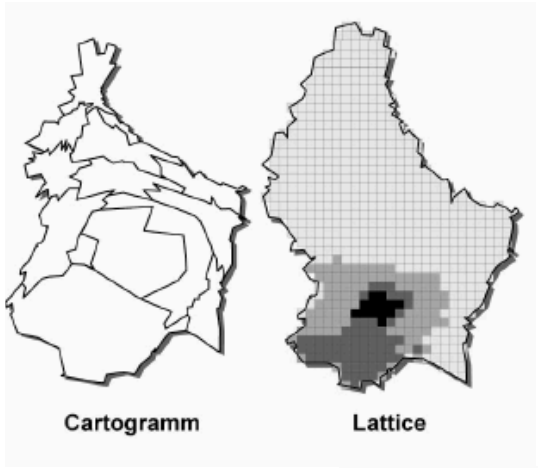
This perspective forces you to reflect upon at least a few important elements:

(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





C. Cauvin, J.P. Antoni, 2006
Software: Carte & Données 5.0, Surfer 8, ArcGis 8.2

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

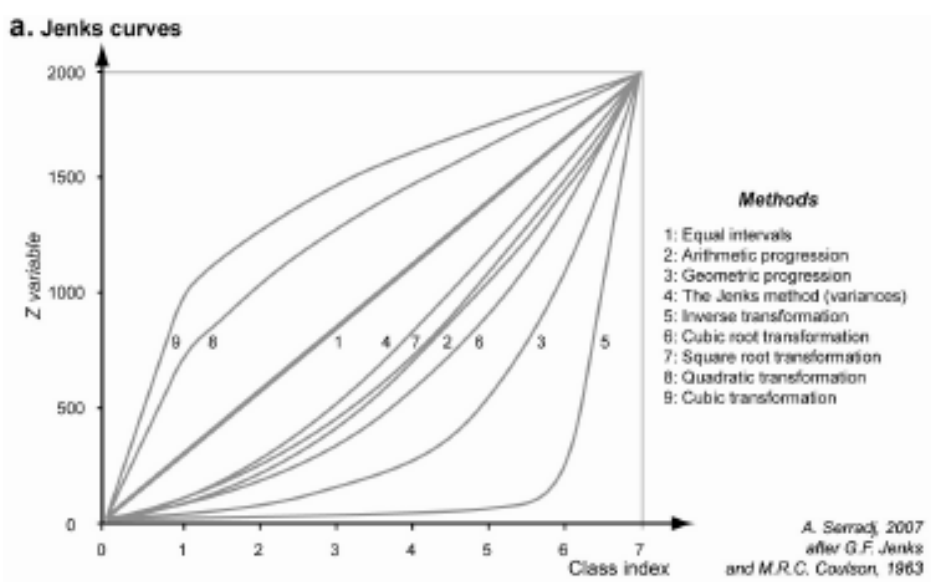
We will cover Cartograms later

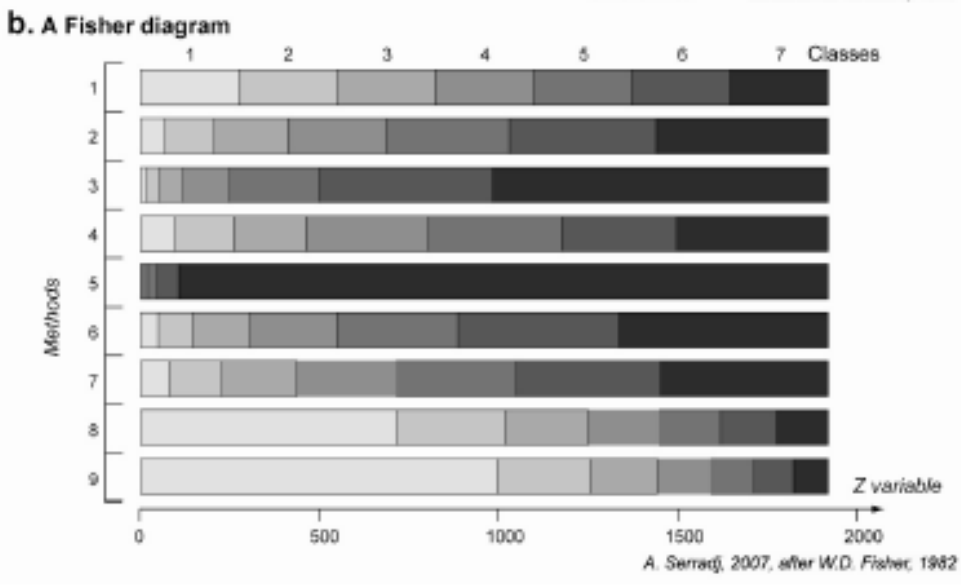
Suppose each dot represent a COVID case, where do think the risk is more or less?

Single variable choropleth map: 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





Cartography and the Impact of the Quantitative Revolution; C. Cauvin, F. Escobar, and A Serradj

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)

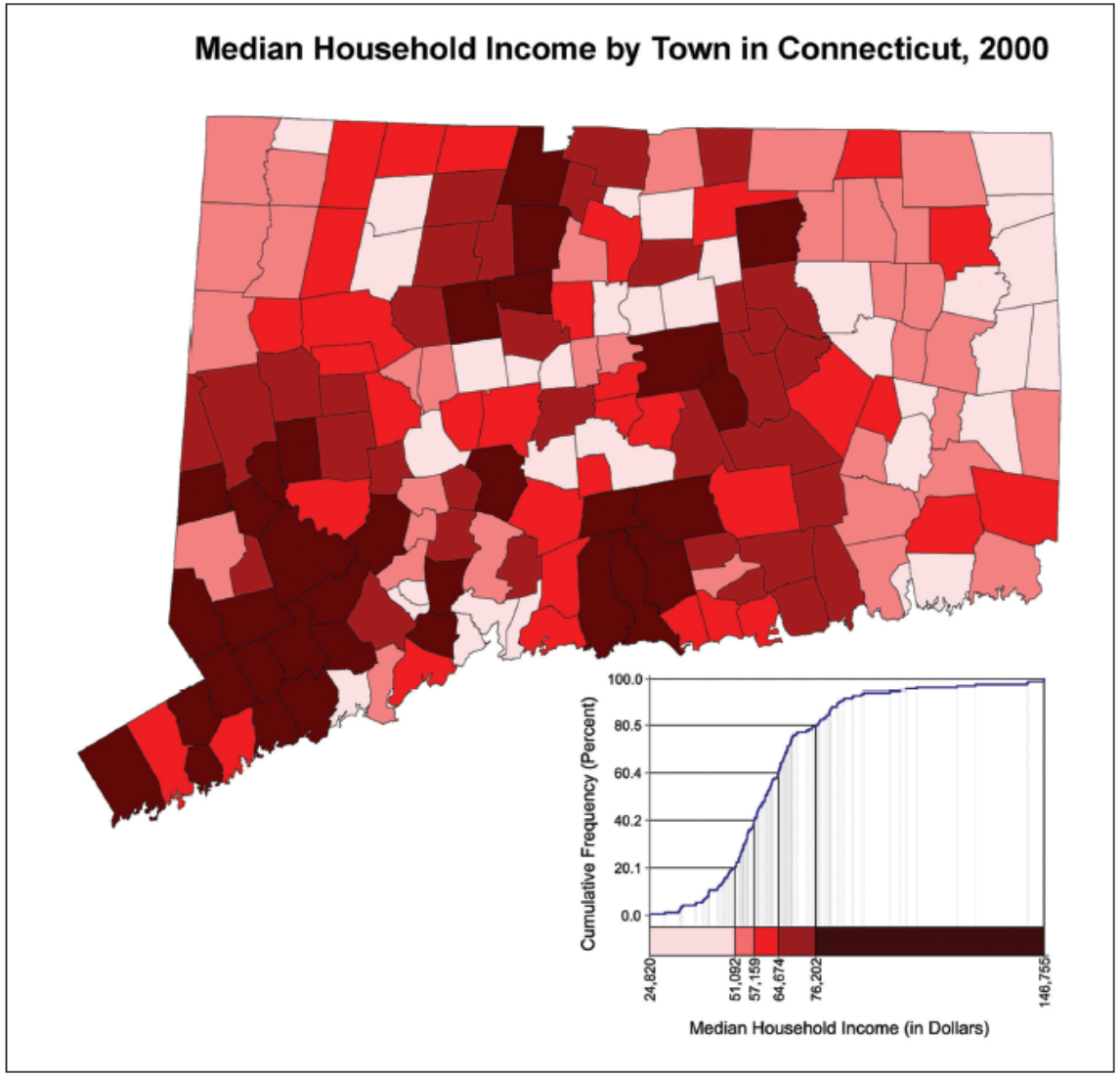
I do not know color science and especially how it is perceived by humans, but check out the references in the R Markdown files and the workshop box folder e.g.

Sequential Range: High to Low

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

Qualitative Scheme: Nominal categories, no order

More informative legend



24820 - 51092
51247 - 57159
57308 - 64674
64680 - 76202
76843 - 146755

Cromley and Ye, 2006, Ogive-based legends for choropleth maps

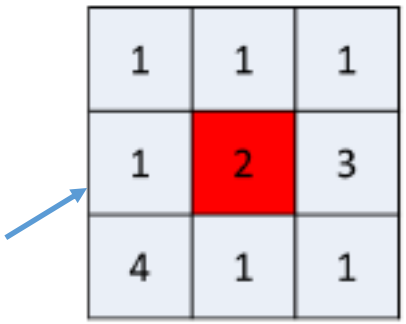
Sensitivity of a map to classification boundaries

Zhang and Maciejewski, 2016

How you cluster data can have a significant impact on the visual attributes of a map, especially spatial clustering.

How sensitive is your map to minor changes in cluster boundaries; especially spatial clustering or autocorrelation in your map

Sensitivity of a map to classification boundaries



Zhang and Maciejewski, 2016

Four categories classification

Suppose the center pixel is at the boundary between category 1 and 2,

If a small change in boundary classifies the center pixel as 1, the area would look like a cluster of 1

But now suppose you shift the boundary to the other side, which makes cell 4 (with the arrow) go from category 1 to category 2, the area will look fragmented

Effect of careful classification

You should strike a balance between simple and accurate

Err towards simplicity for general audience

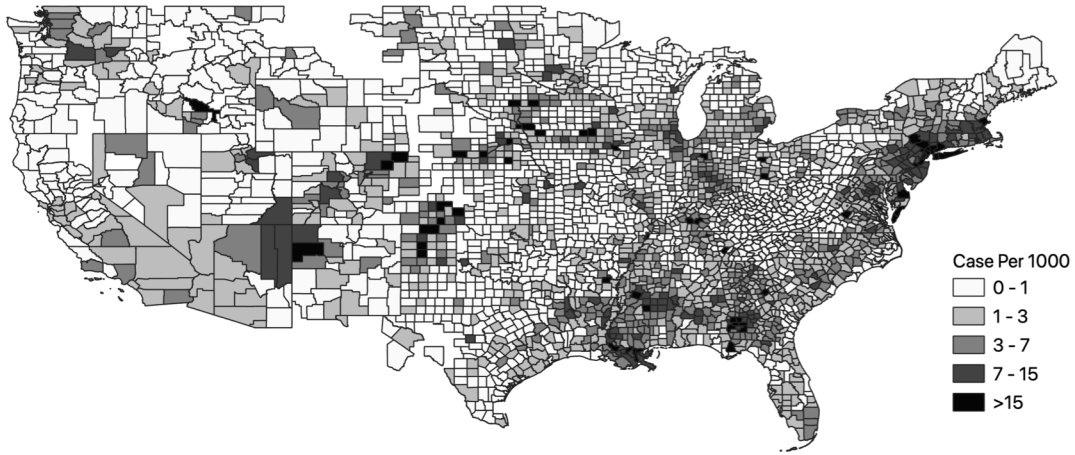
Go for more accuracy for selected, trained audience

Careful optimization procedures in choropleth classification may not offer any benefit over the simpler method for the general map-reading tasks .. (*Brewer and Pickle, 2002, Annals of the Association of American Geographers*)

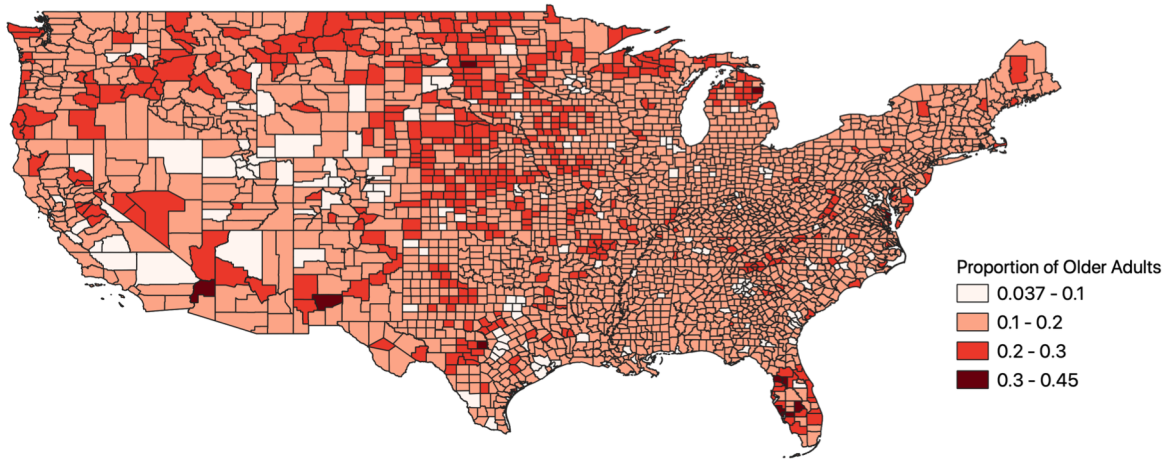
Map visualization: Bivariate Choropleth Map

Bivariate Choropleth Map

Case per 1000

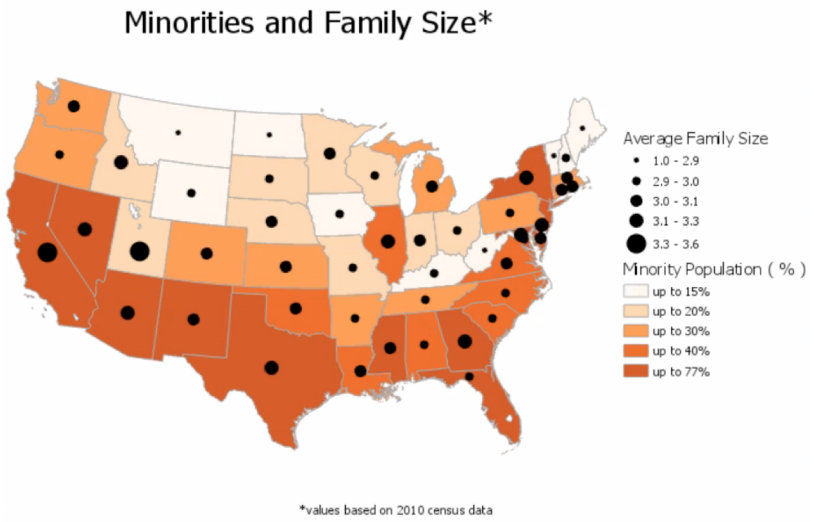


Proportion of older adults



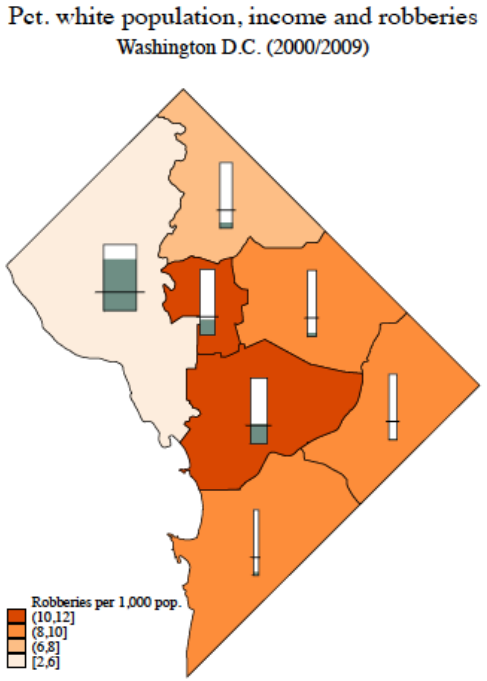
Combine the two to show the joint variation in Cases Per 1000 and Proportion of Older Adults

Bivariate map: Options



Credit: Wikipedia

Three variables:
(1) Color,
(2)Width of the glass,
(3) How filled it is



Credit: Pisati M., https://www.stata.com/meeting/italy12/abstracts/materials/it12_pisati.pdf

Many other examples, e.g. pie chart in each polygon

You need sufficiently large polygons to clearly display other information. We do not have this for county level maps.

Bivariate choropleth map

You have two spatially distributed variable ‘P’ and ‘Q’

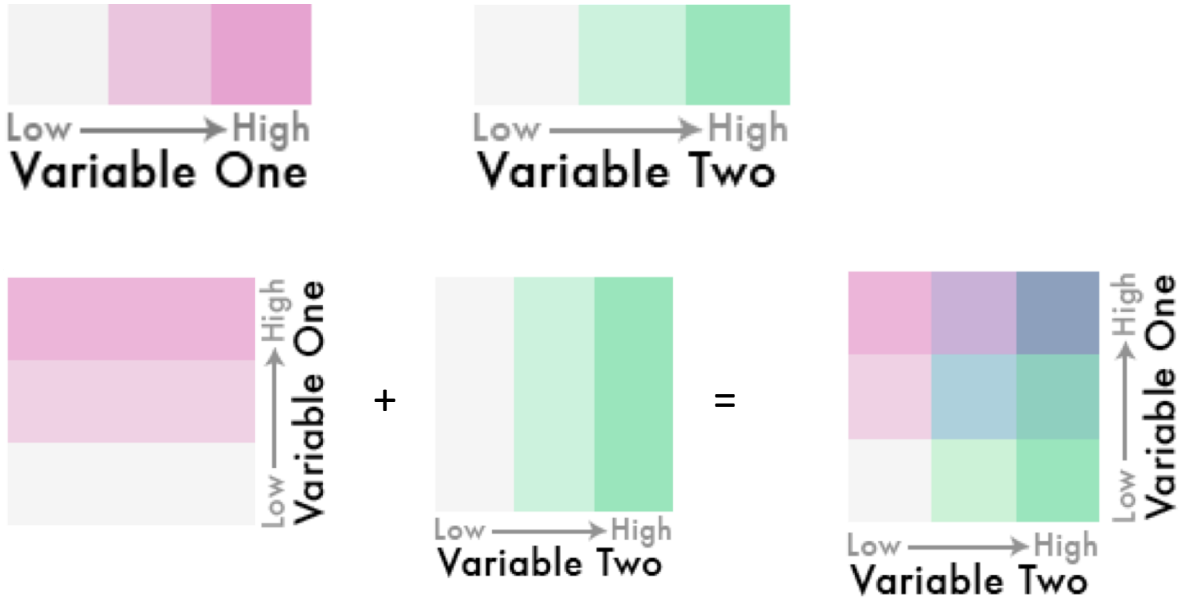
You want to show joint variation in P and Q over space

Discretize the two variables

Case per 1000/ Proportion	P1 (older adults p<=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



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

Assume the two variables are numerical

Choose a color gradient for each variable – preferably something light

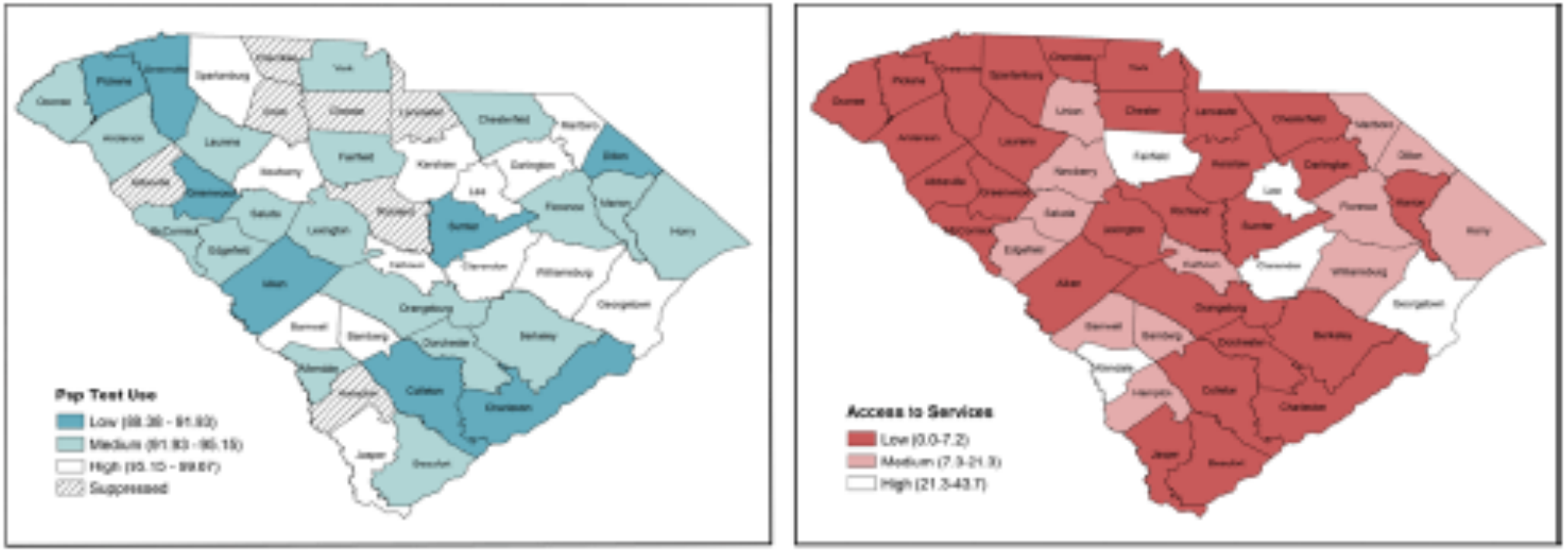
Combine the two to create the color scheme

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

Bivariate choropleth map: Example

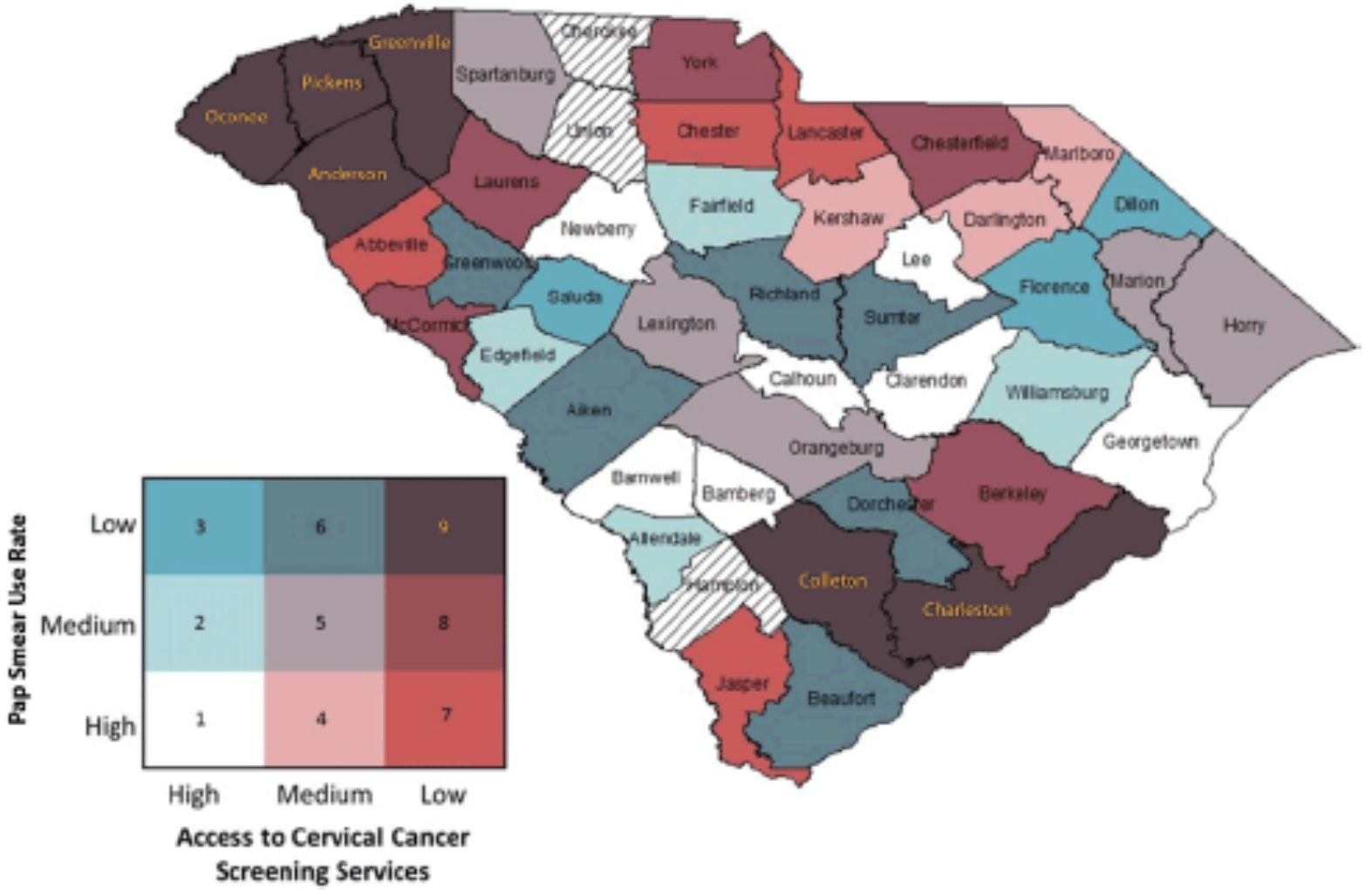
Pap test use

Access to service



Biesecker et al., 2020, CDC

Bivariate choropleth map: Example



Counties with stripes had insufficient data

Biesecker et al., 2020, CDC

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?

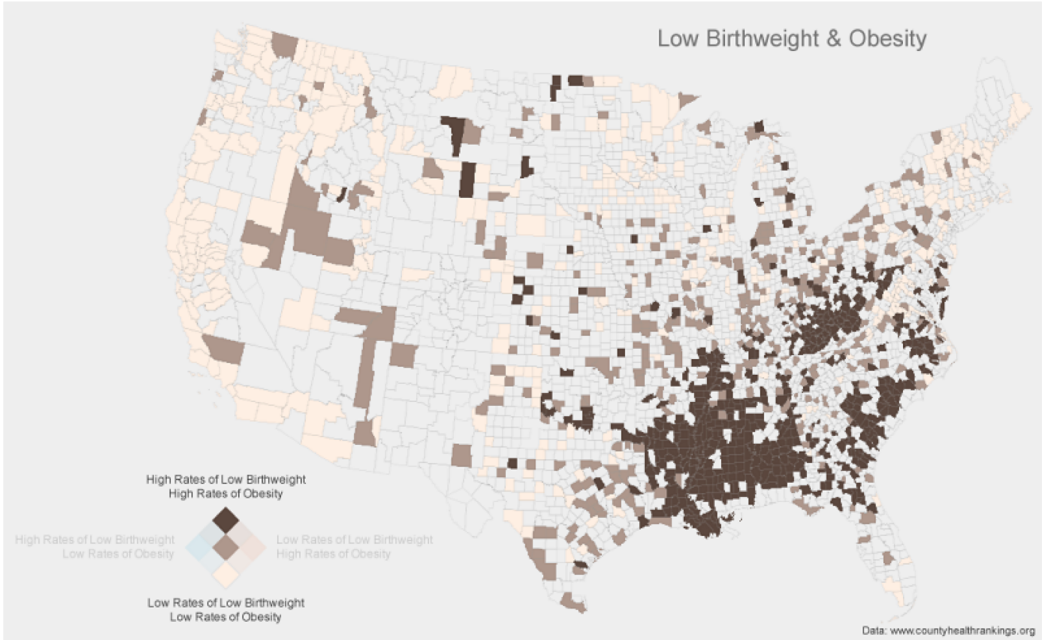
In a bi-variate legend each row and column should preferably be a visual sequence

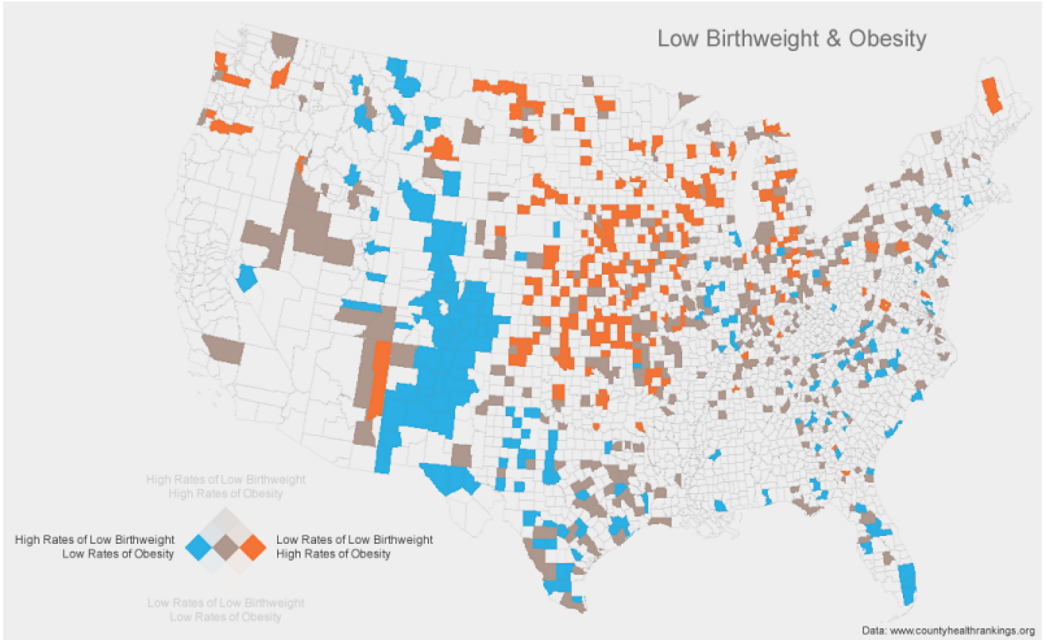
If the focus is on the interaction between the two variables, the diagonal of the legend should be visually distinct from the rest of the legend

Let us go back and see the pep test figure

https://cartographicperspectives.org/index.php/journal/article/download/1538/1819?inline=1

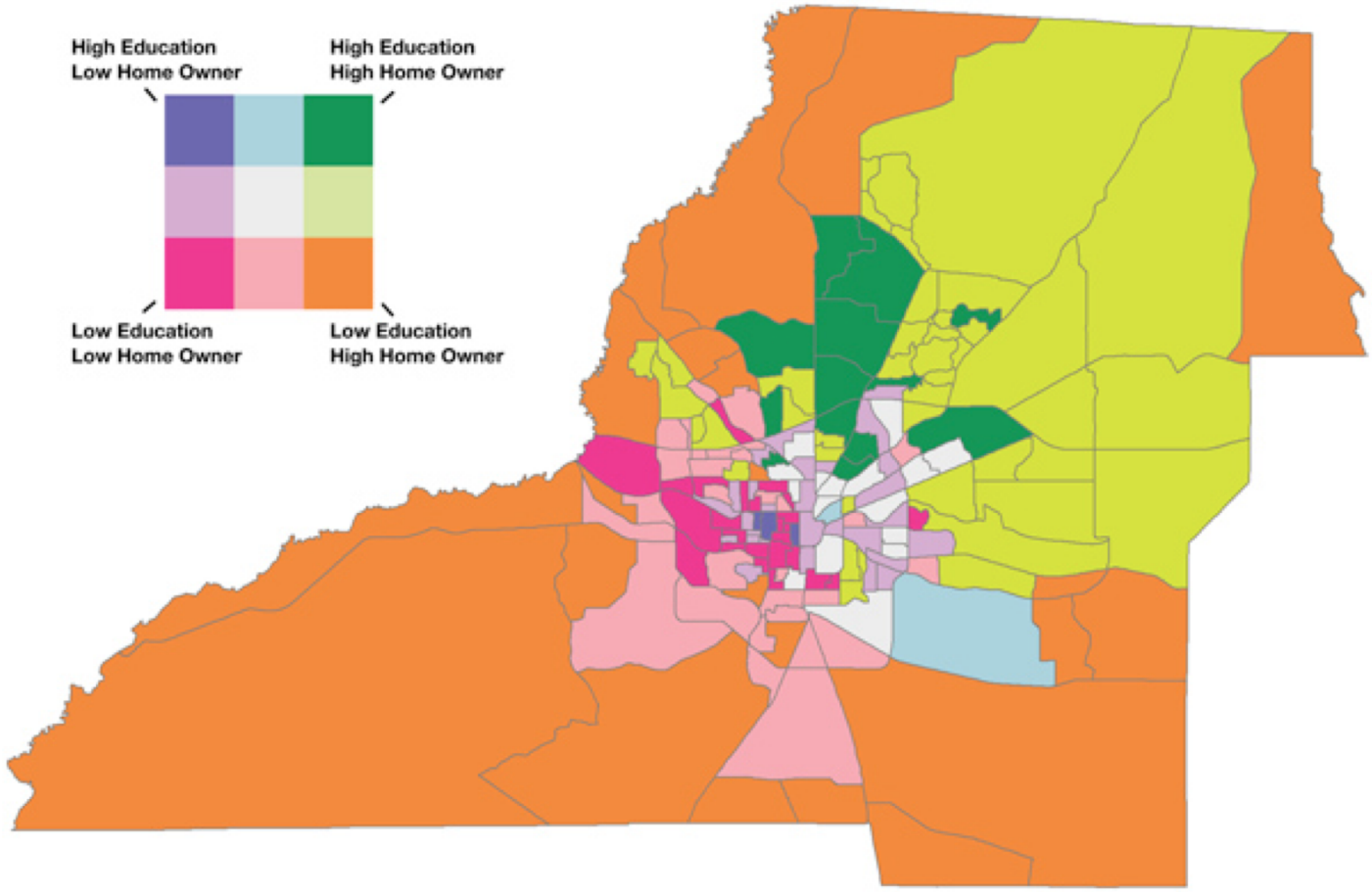
Trumbo's principle





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

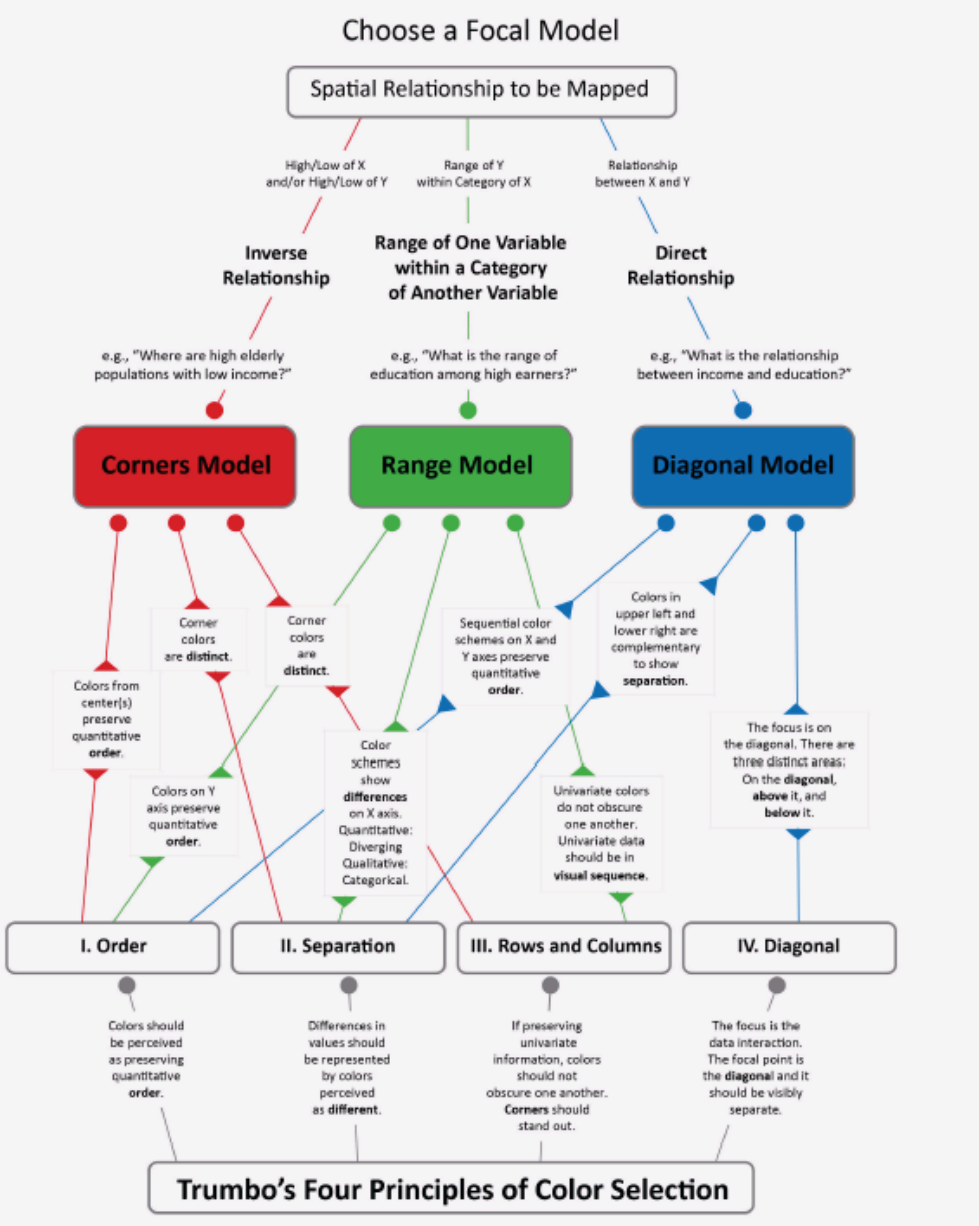
Trumbo's principles: Emphasis on corner values



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

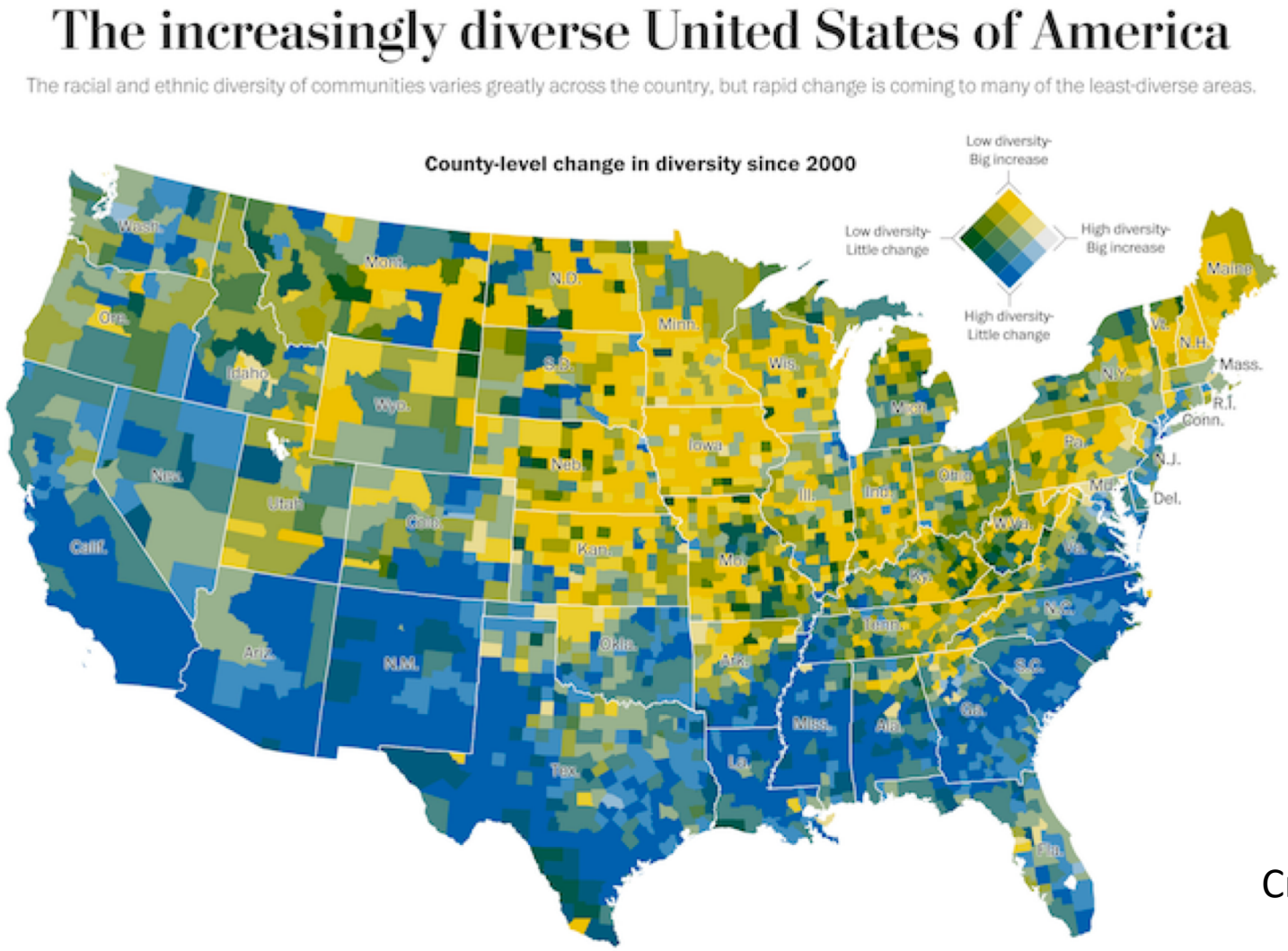
https://cartographicperspectives.org/index.php/journal/article/download/1538/1819?inline=1

Trumbo's principles: Color selection



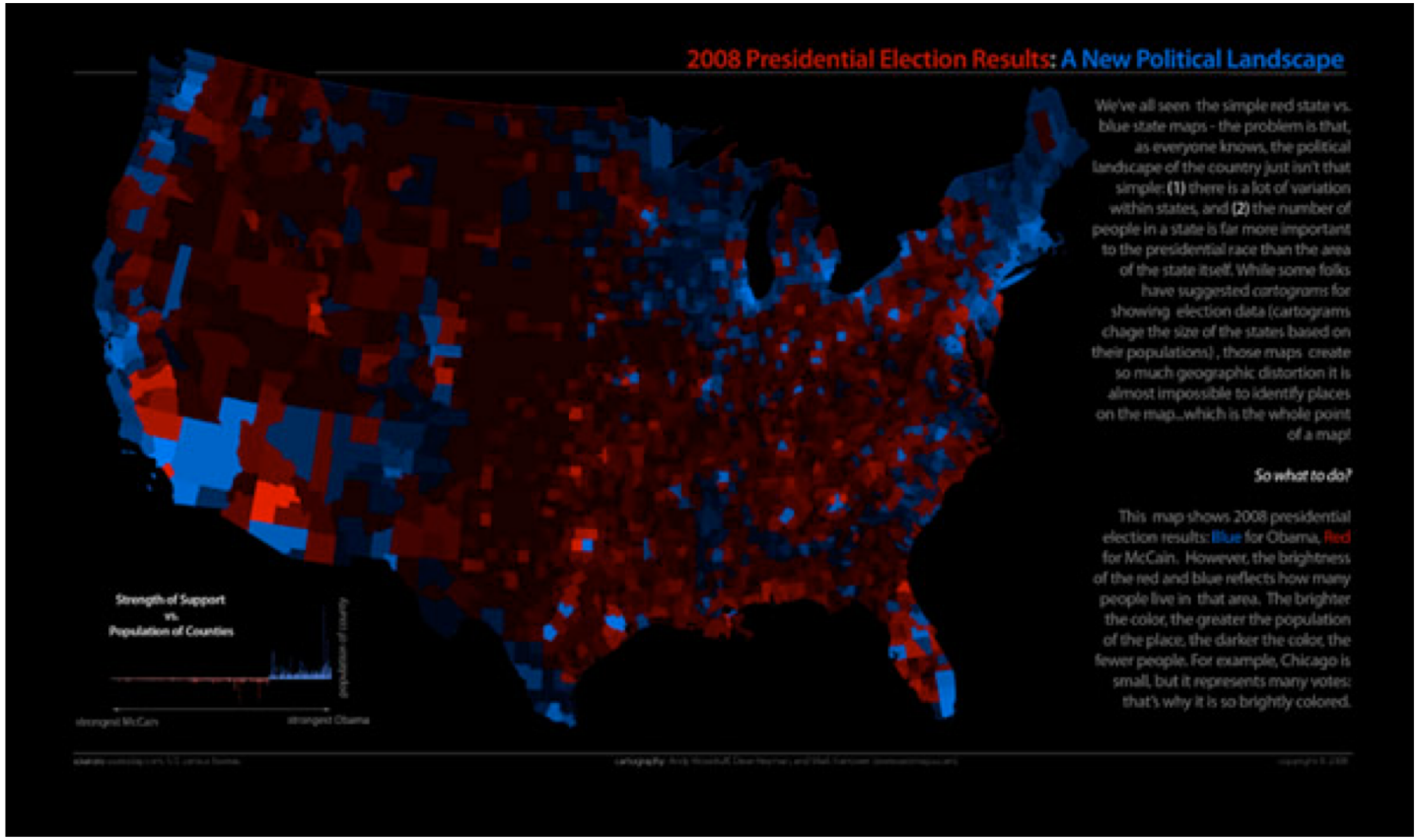
Strode G, Morgan JD, Thornton B, Mesev V, Rau E, Shortes S, et al. Operationalizing Trumbo's Principles of Bivariate Choropleth Map Design. Cartographic Perspectives. 2020.

Bivariate choropleth map



Credit: The Washington Post

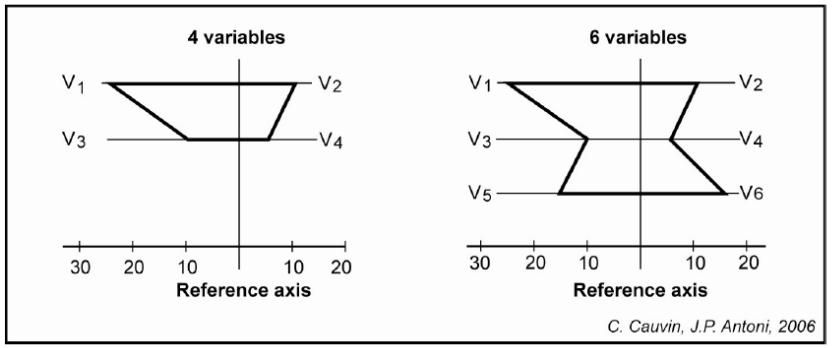
Bivariate: value by alpha map



http://andywoodruff.com/blog/value-by-alpha-maps/

Multiple variables

Profile method



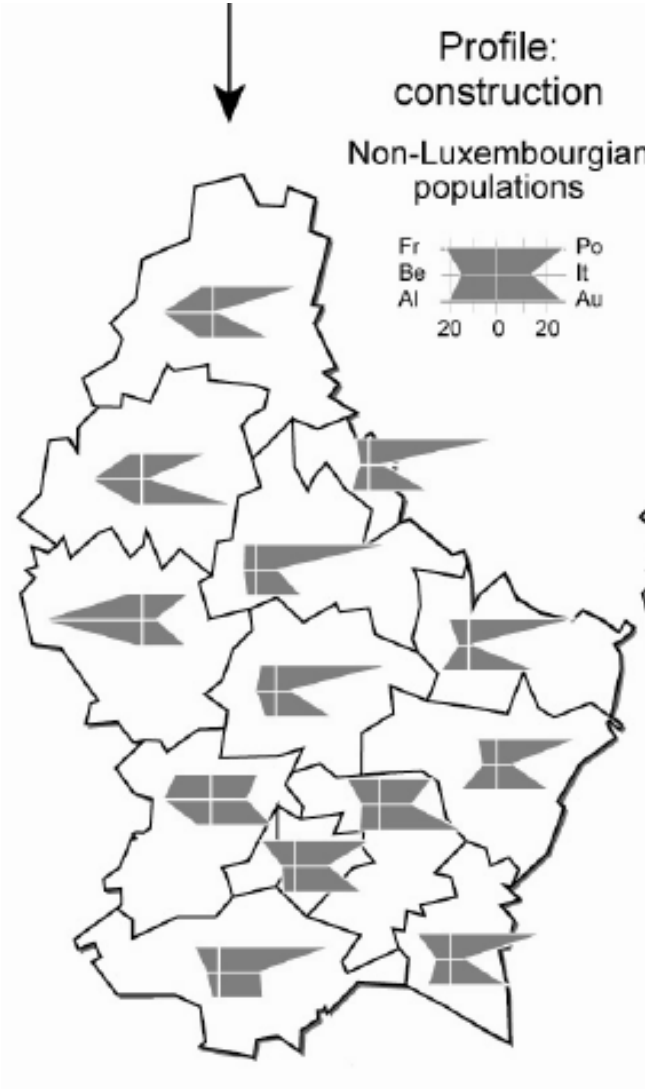
Suppose there are k variables describing n spatial units. For each spatial unit a characteristic profile is constructed:

– Draw the vertical Y axis which serves as the axis of origin (= axis of reference).

– Draw k semi-axes on both sides of the Y axis, perpendicular to the Y axis and parallel to each other (k = number of variables). The point of origin on each of these axes is the point of intersection with the axis of reference.

– Mark the value of each variable in the spatial unit considered on the semi-axis which corresponds to this variable.

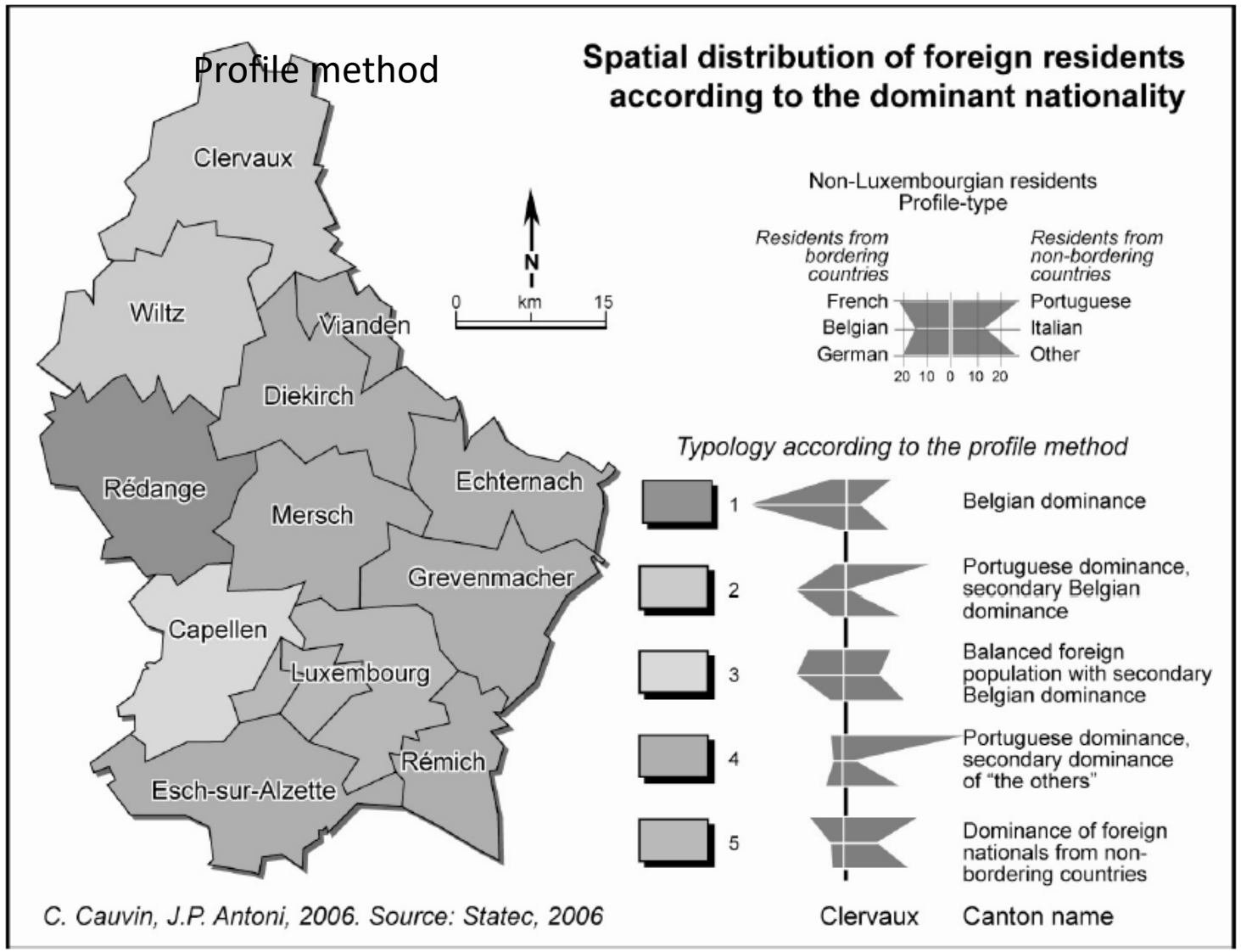
– Make the plot by joining all the marked points on the semi-axes without intersections with the axis of reference.

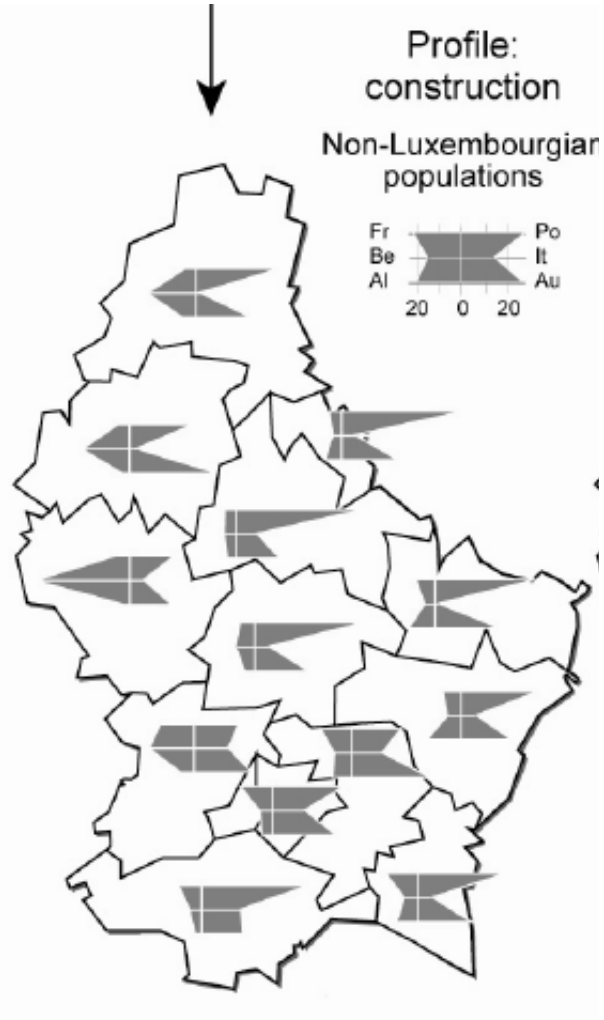


Thematic Cartography and Transformations; C. Cauvin, F. Escobar, and A Serradj

Multiple variables

Profile method





Which approach do you think is better, why?

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

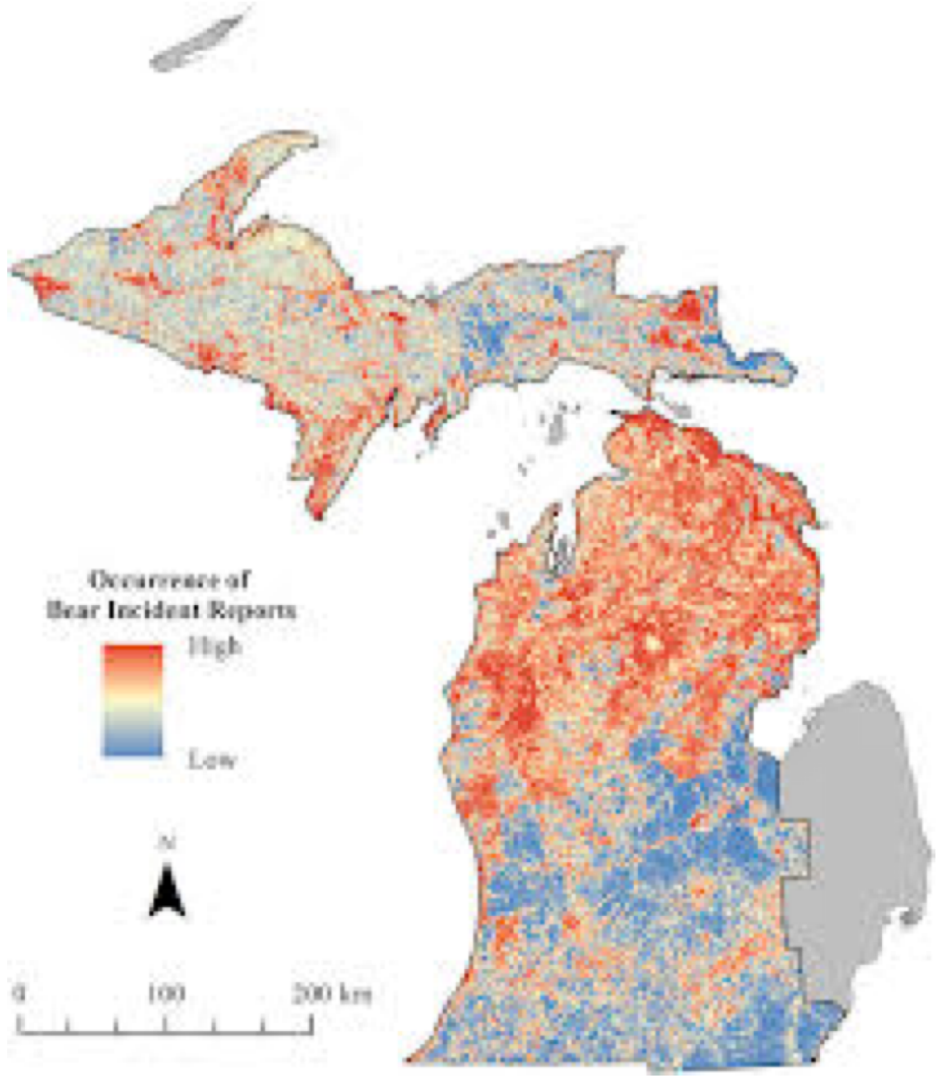


Image credit: anotherview.info

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