NYU

CS-GY 6313 B: Information Visualization

10/03/2024

Logistics

- Assignment 1 presentations
- Lecture
- Final project: look for partners, brainstorm ideas, consult with me

2D Viz: Spatial Data

Geographical data

- Any data that contains geographical data

- Spatial objects
  - Counties
  - Regions
  - Buildings
  - Rivers, lakes
  - Etc.
- Non-spatial geolocated objects
  - Cars
  - People
  - Animals
  - Weather stations
  - Etc.

Spatial data

- Use the given spatial position
- When?
  - Dataset contains spatial attributes and they have primary importance
  - Central tasks revolve around understanding spatial relationships
- Eg:
  - Geographical/cartographic data
  - Sensor/simulation data

Geographic maps

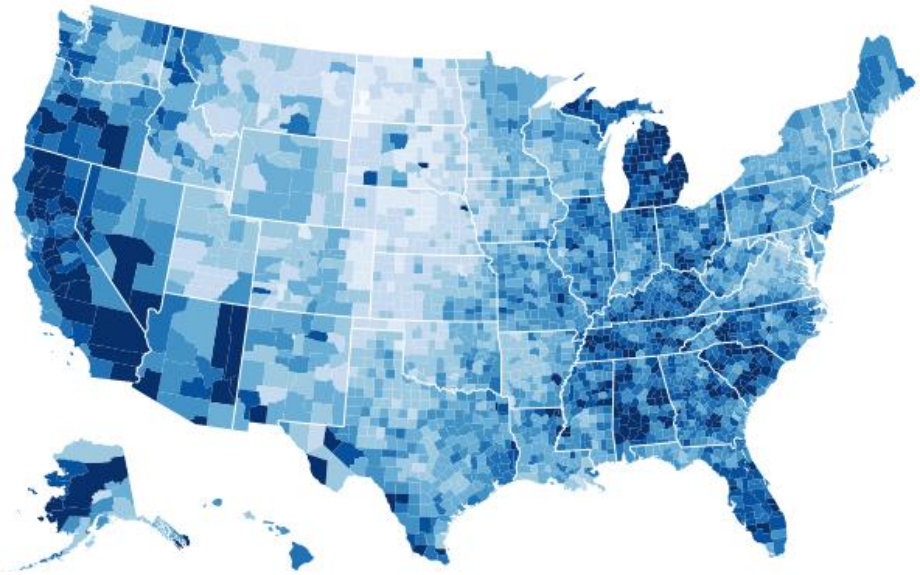
- Interlocking marks!
  - Shape coded
  - Area coded
  - Position coded
- Cannot encode another attribute with these channels. They are taken!

Thematic maps

- Shows the spatial variability in an attribute (the “theme”)
  - Combine geographic / reference map with (simple, flat) tabular data
  - “Region” has 2 meanings here:
    - Interlocking area marks (provinces, countries with outline shapes)
    - Categorical key attribute in a table
- Major techniques:
  - Choropleth
  - Symbol map
  - Cartogram
  - Dot density map

Choropleth map

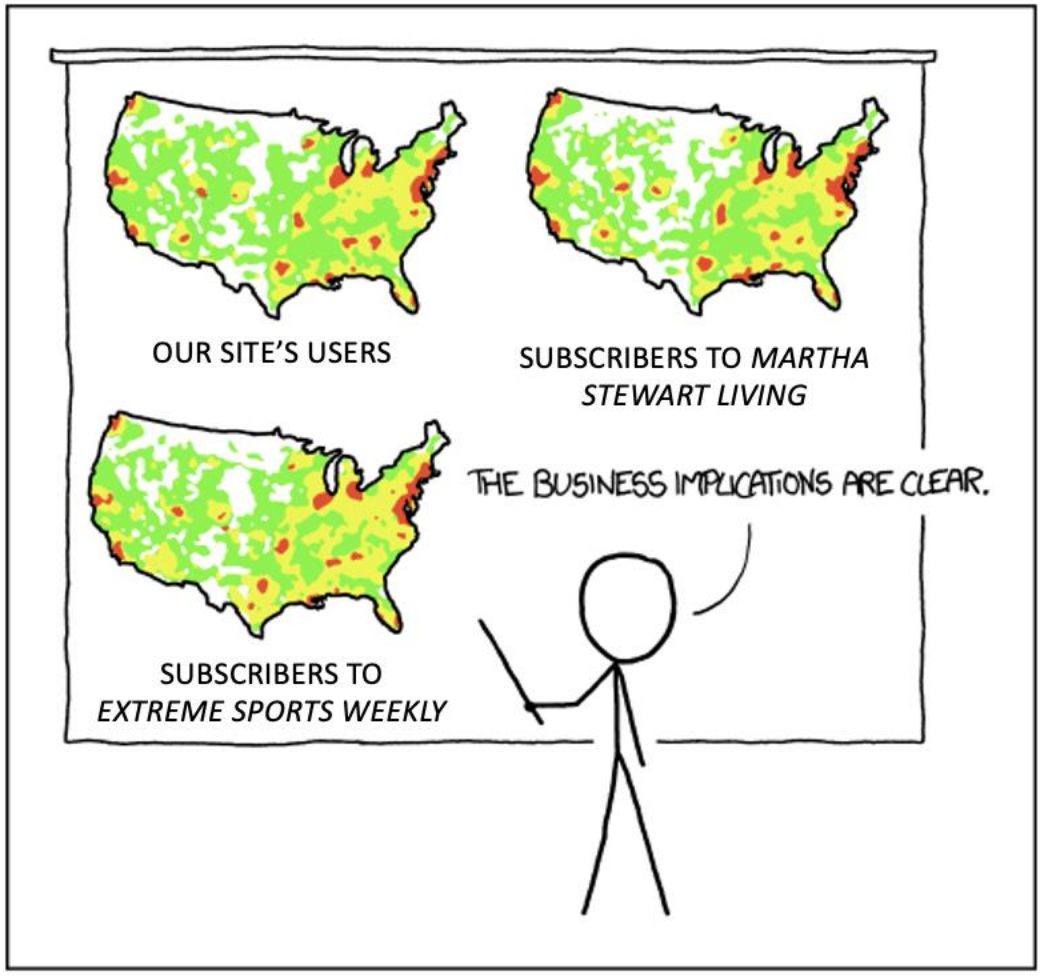
- Use when the central task is understanding spatial relationships
- Data:
  - Geographic geometry
  - Table with 1 quant. attribute per region
- Encoding:
  - Position: use given geometry for area mark boundaries
  - Color: sequential segmented color map



http://bl.ocks.org/mbostock/4060606

Beware: population map trickiness!

- Spurious correlations: most attributes just show where people live
- Consider when to normalize by population density
  - Encode raw data values → tied to underlying population
  - Use normalized values → account for population density
    - E.g. unemployed people per 100 citizens
- Not restricted to choropleth maps!
  - General issue in viz
  - Failure to normalize is a common error



PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE BASICALLY JUST POPULATION MAPS

Choropleth map: recommendations

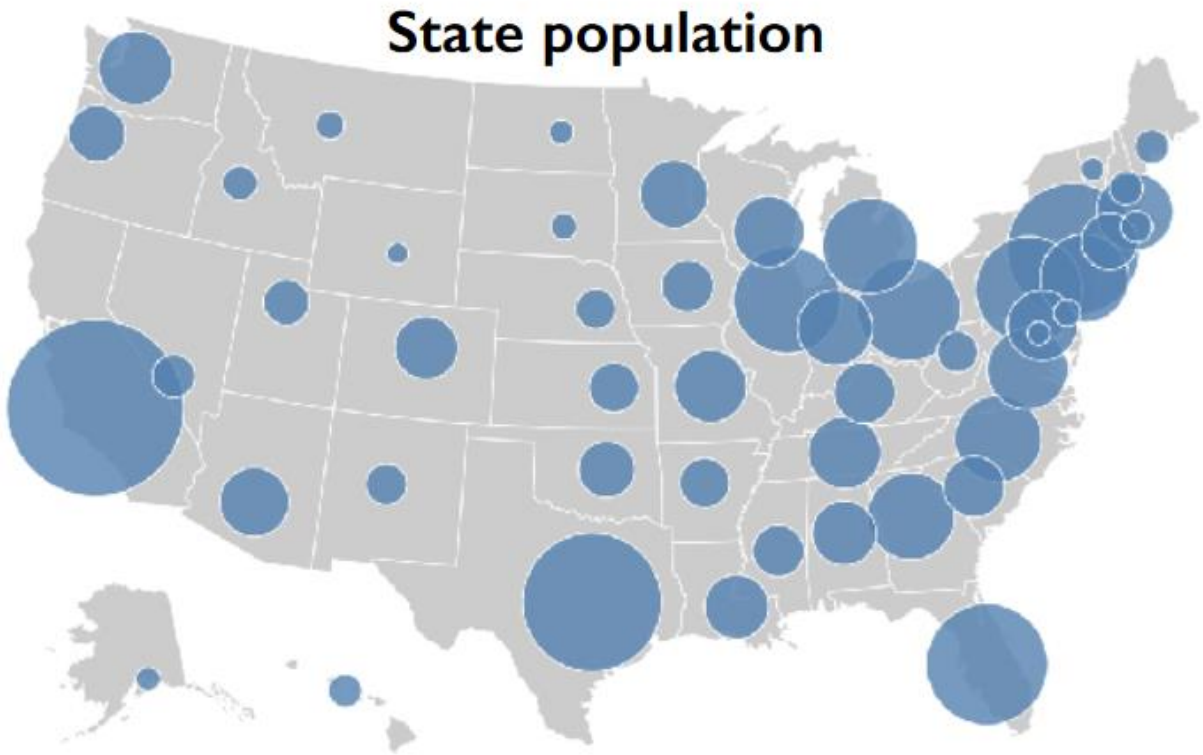
- Only use when central task is understanding spatial relationships
- Show only one variable at a time
- Normalize when appropriate
- Be careful when choosing colors & bins
- Best case: regions are roughly equal sized

Choropleth map: pros and cons

- Pros:
  - Easy to read and understand
  - Well established visualization (no learning curve)
  - Data is often collected and aggregated by geographical regions
- Cons:
  - Most effective visual variable used for geographic location
  - Visual salience depends on region size, not true importance w.r.t. attribute value
    - Large regions appear more important than small ones
  - Color palette choice has a huge influence on the result

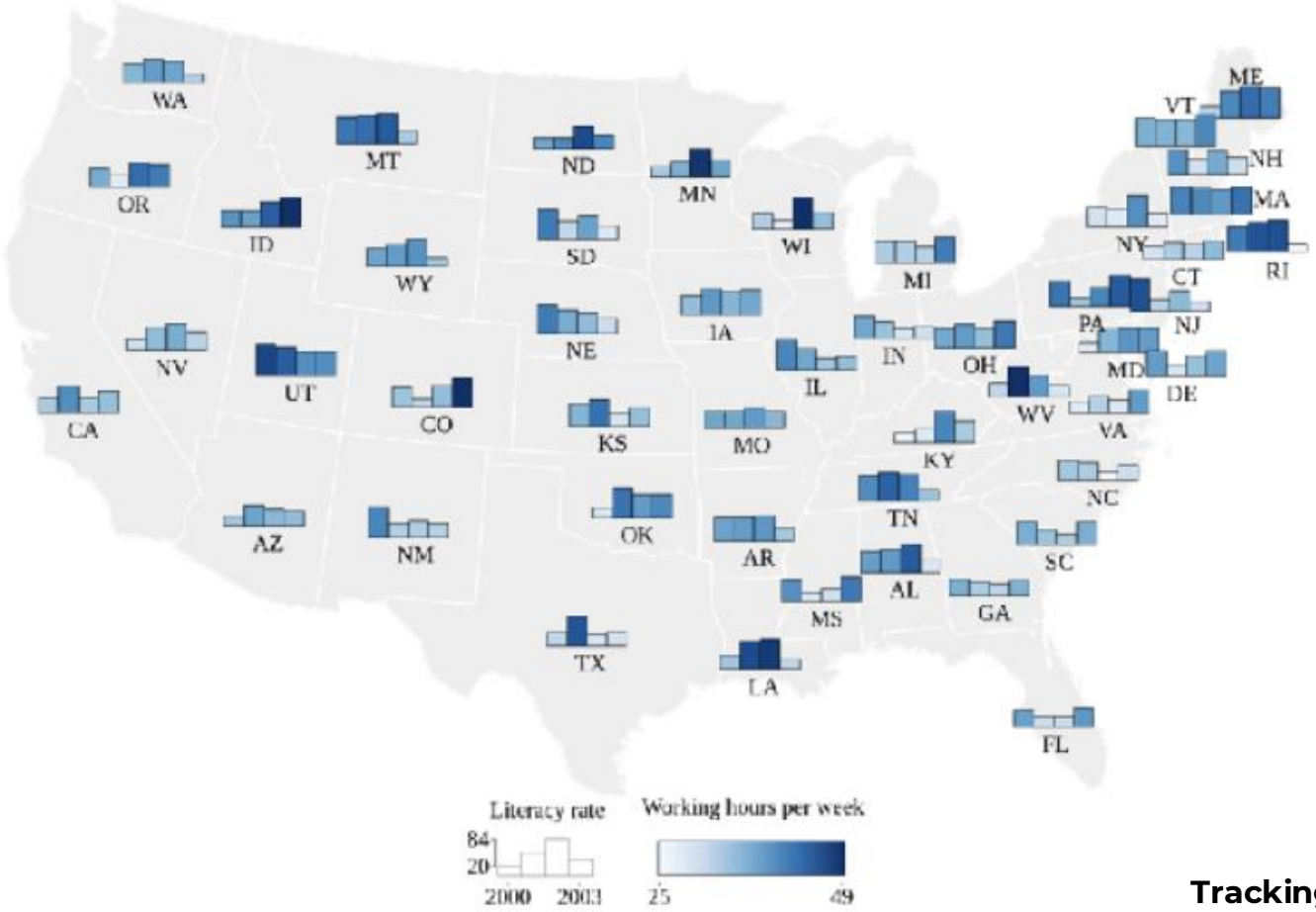
Symbol maps

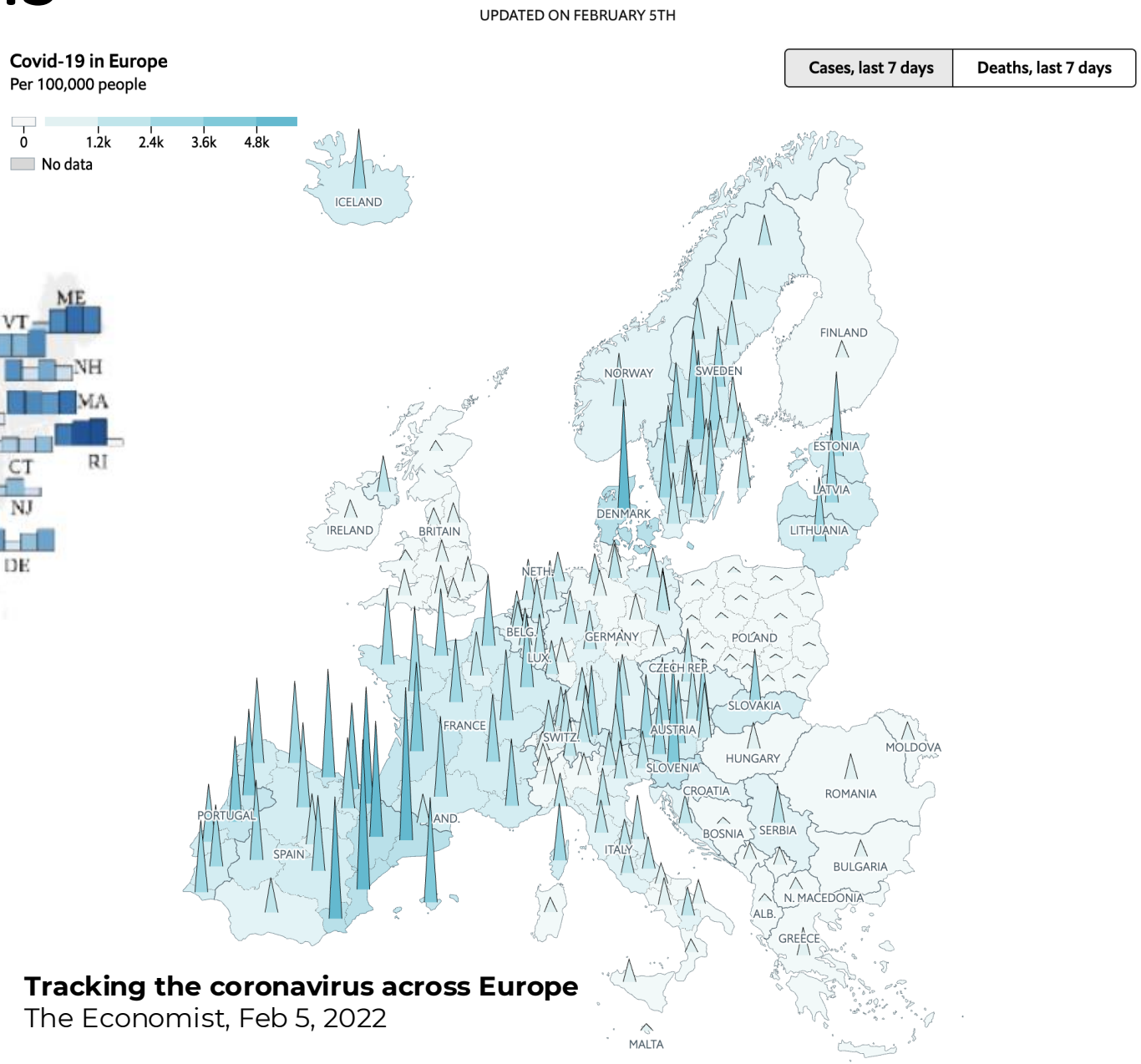
- A symbol is used to represent aggregated data (marks or glyphs)
  - Allows the use of size and shape and color channels
    - AKA proportional symbol maps, graduated symbol maps
- Keep original spatial geometry in the background
- Often a good alternative to choropleth maps



Symbol maps with glyphs

- Glyph = complex marks





Tracking the coronavirus across Europe
The Economist, Feb 5, 2022

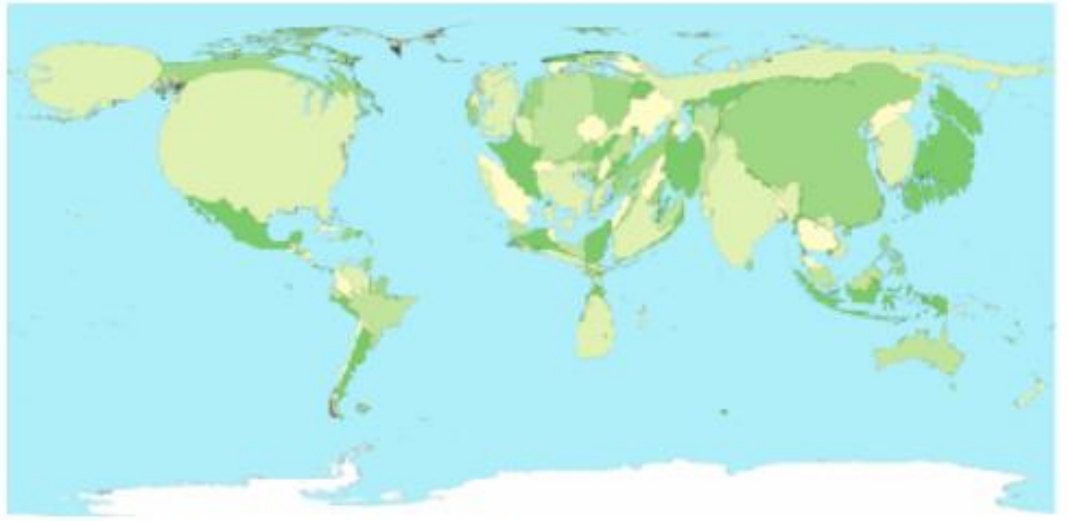
https://www.economist.com/graphic-detail/tracking-coronavirus-across-europe

Symbol map: Pros & cons

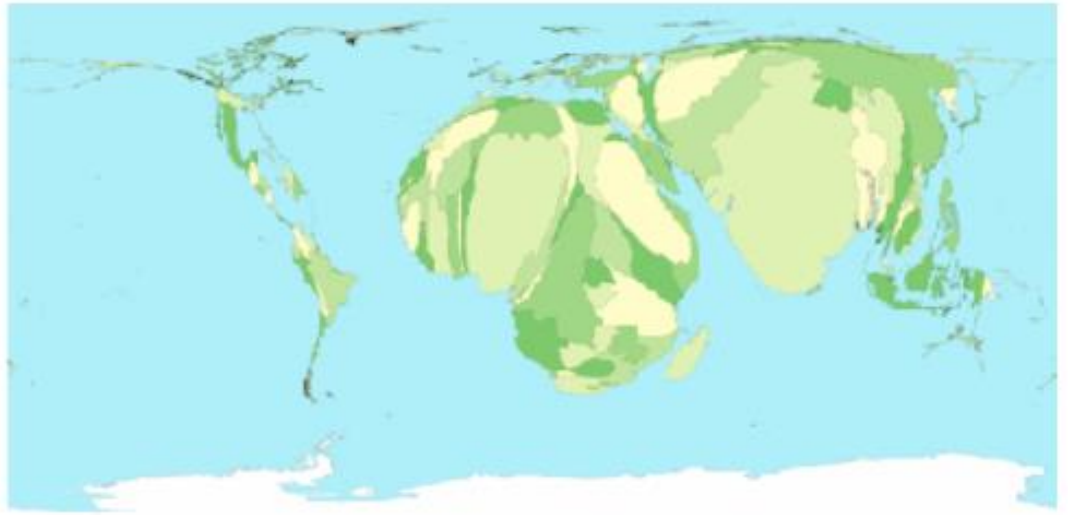
- Pros:
  - Somewhat intuitive to read and understand
  - Mitigate problems with region size vs data salience
    - Marks: symbol size follows attribute value
    - Glyphs: symbol size can be uniform
- Cons:
  - Possible occlusion / overlap
    - Symbols could overlap each other
    - Symbols could occlude region boundaries
  - Complex glyphs may require explanation / training

Contiguous cartogram

- Interlocking marks: shape, area, and position coded
- Derive new interlocking marks
  - Based on combination of original interlocking marks and new quantitative attribute
- Need an algorithm to create new marks
  - Input: target size
  - Goal: shape as close to the original as possible
  - Requirement: maintain constraints
    - Relative position
    - Contiguous boundaries with their neighbors



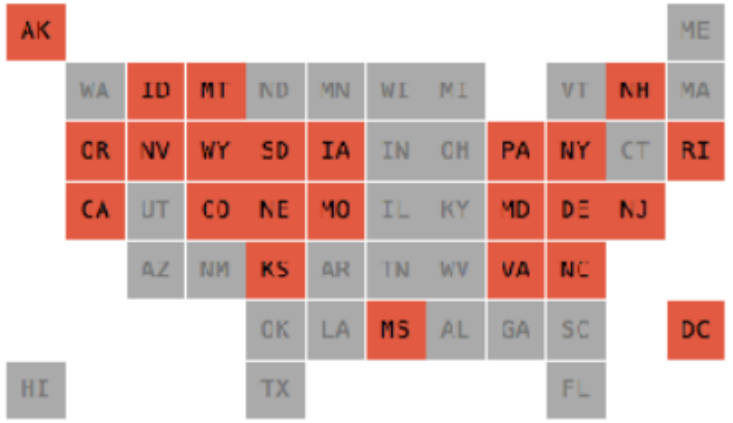
Greenhouse Emissions

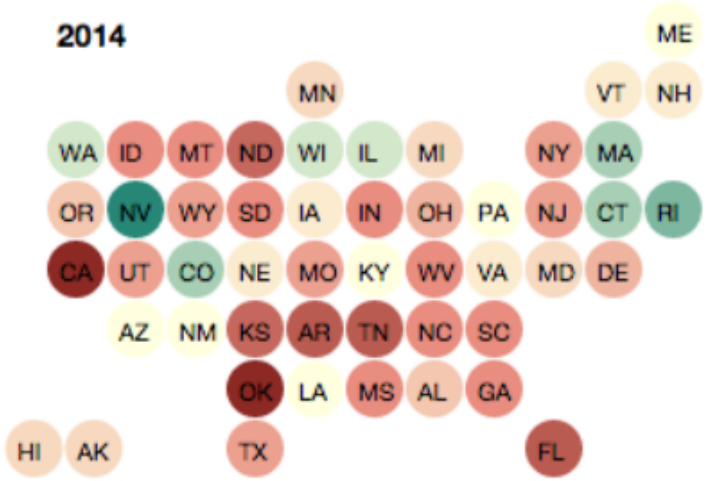


Child Mortality

Grid Cartogram

- Uniform-sized shapes arranged in rectilinear grid
- Maintain approximate spatial position and arrangement



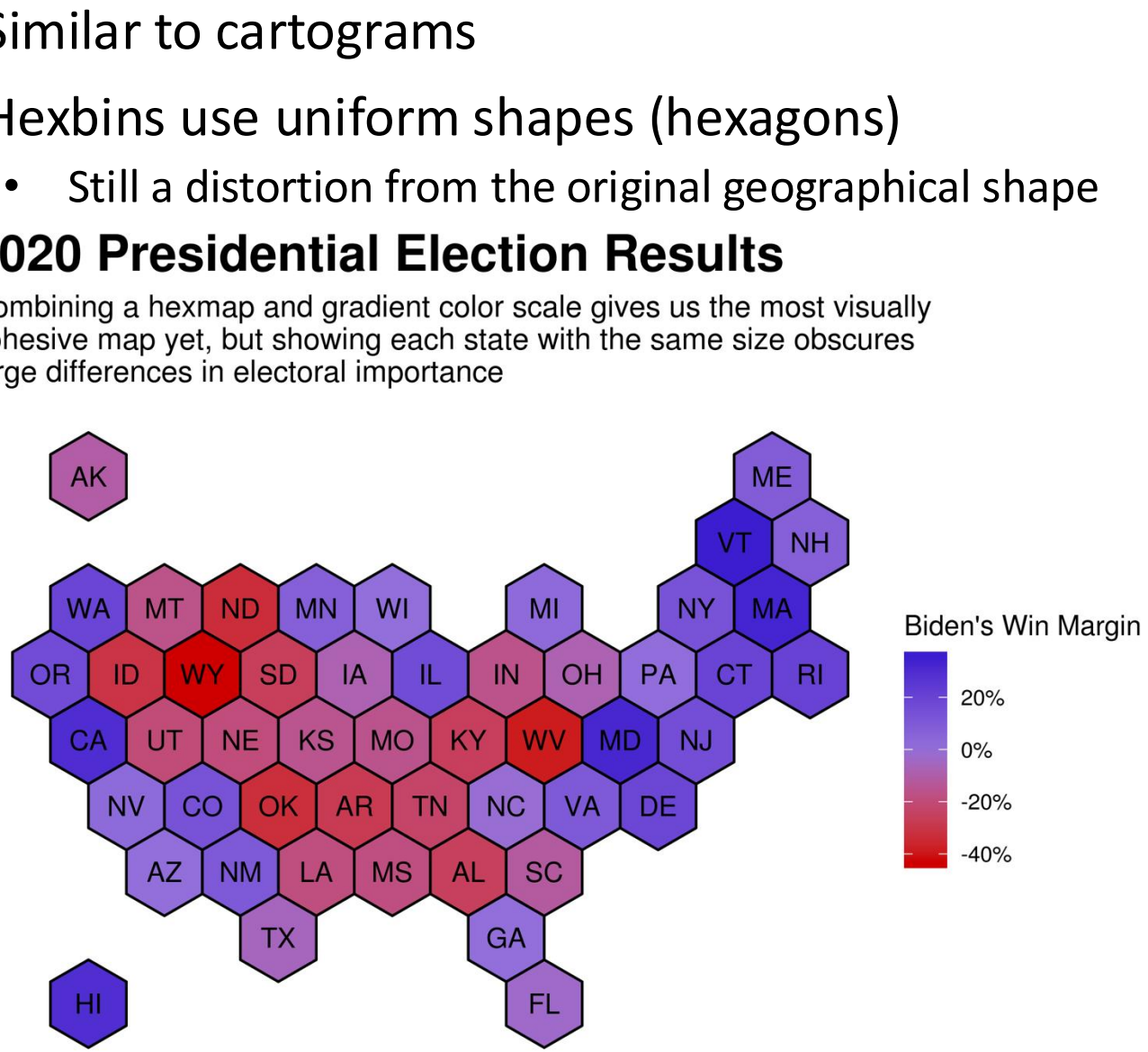


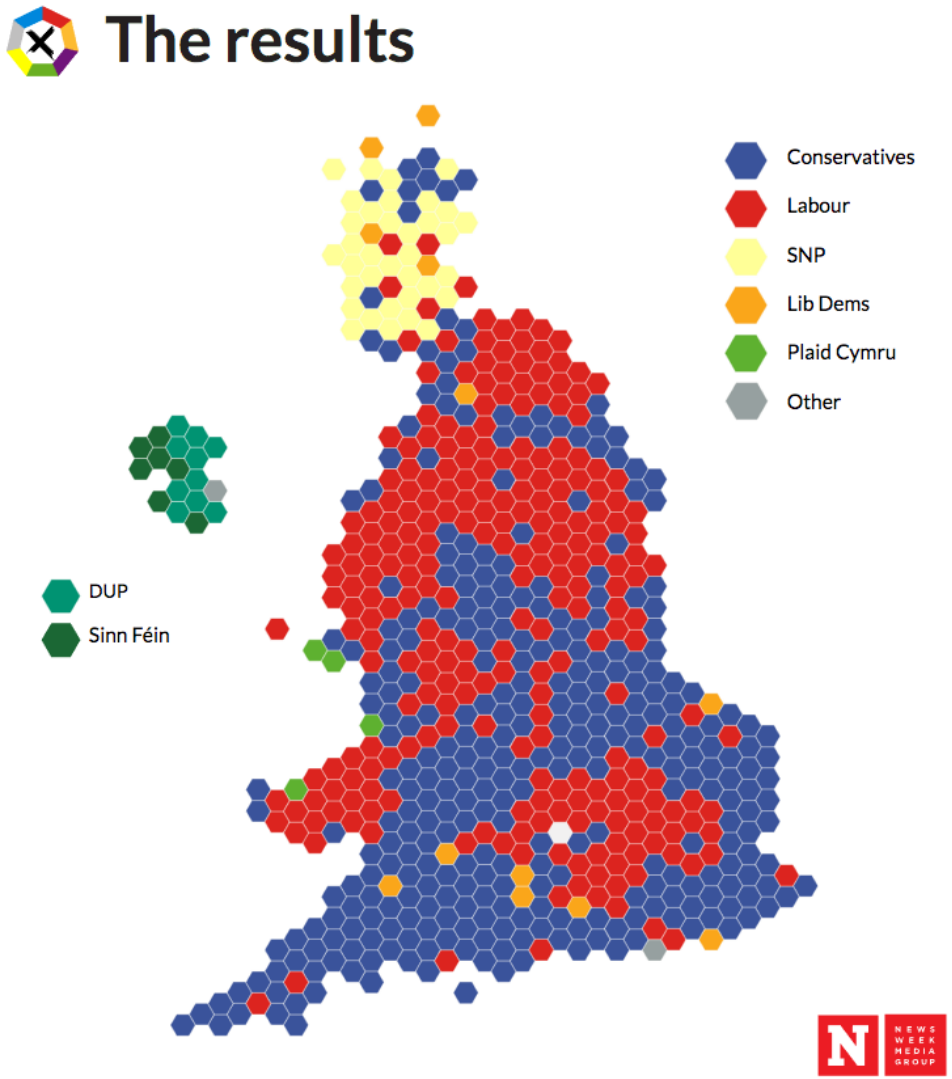
Cartogram: Pros & cons

- Pros:
  - Can be intriguing and engaging
  - Best case: strong and surprising size disparities
  - Non-contiguous cartograms often easier to understand
- Cons:
  - Require substantial familiarity with original dataset & use of memory
    - Compare distorted marks to memory of original marks
    - Mitigation strategies: transitions or side by side views
  - Major distortion is problematic
    - May be aesthetically displeasing
    - May result in unrecognizable marks
  - Difficult to extract exact quantities

Hexbin map

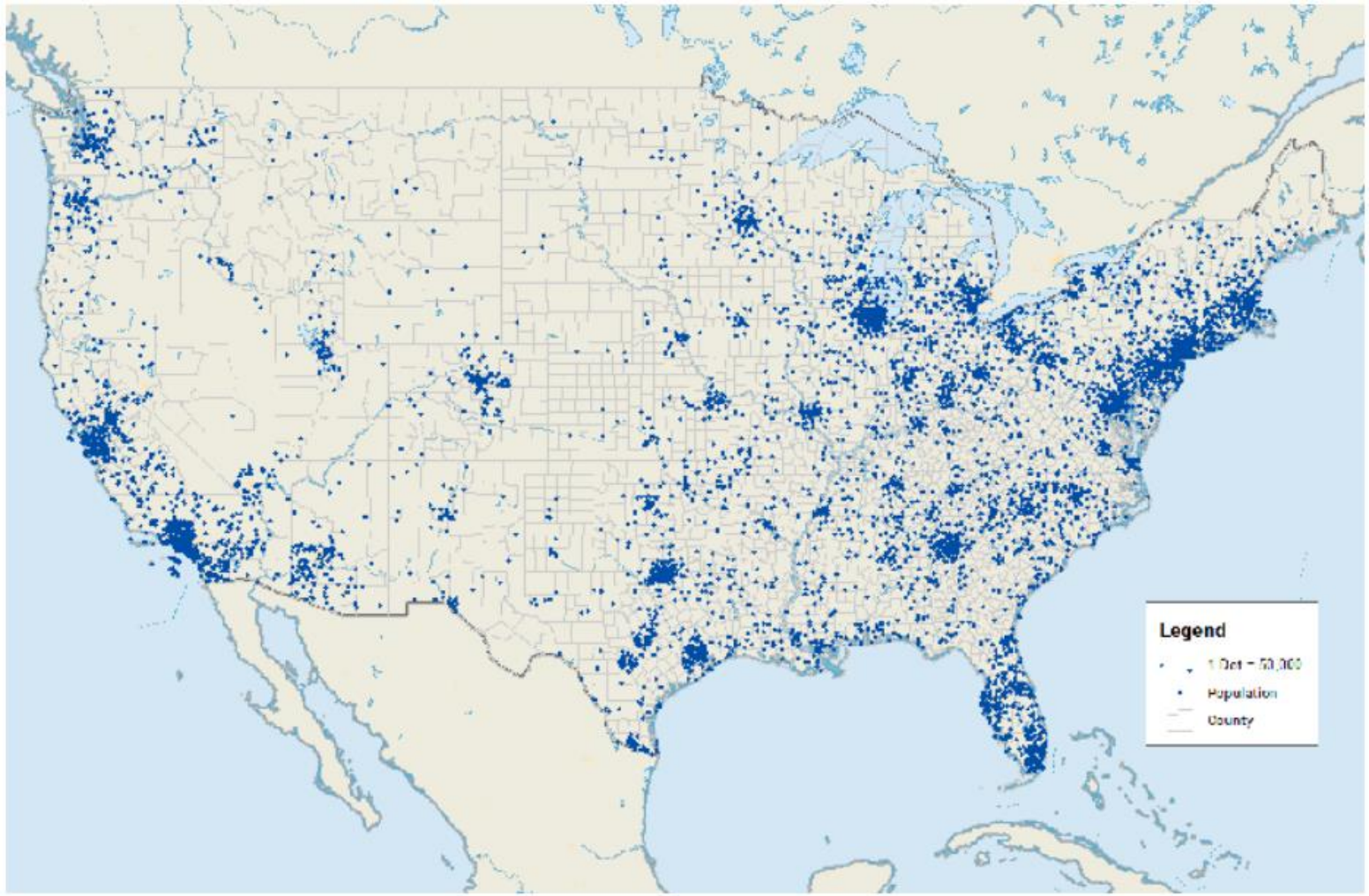
- Similar to cartograms
- Hexbins use uniform shapes (hexagons)
  - Still a distortion from the original geographical shape





Dot density maps

- Visualize distribution of a phenomenon by placing dots
- One symbol represents a constant number of items
  - Dots have uniform size & shape
  - Allows use of color channel
- Task: show spatial patterns, clusters

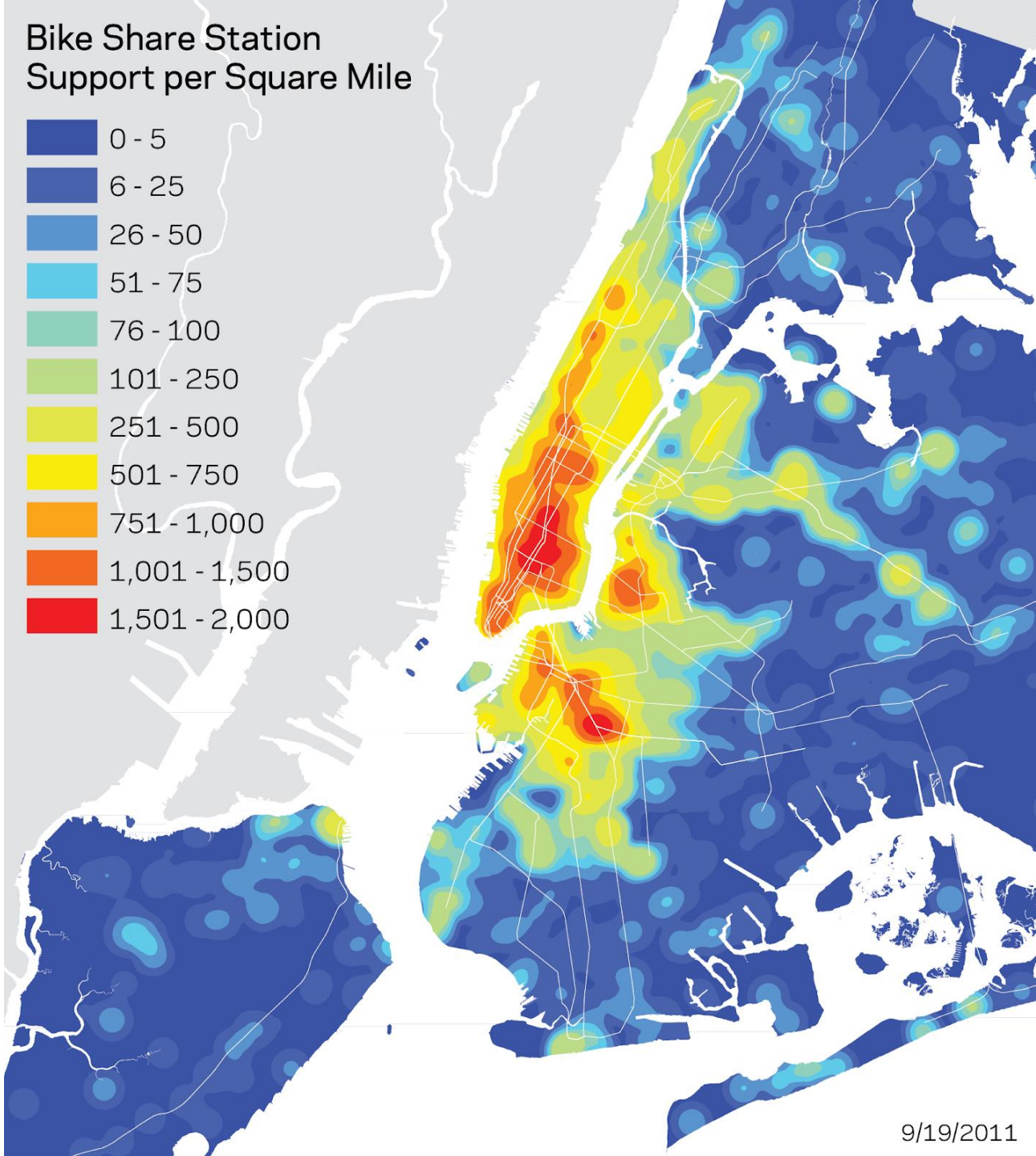


Dot density maps: Pros and cons

- Pros:
  - Straightforward to understand
  - Avoids choropleth non-uniform region size problems
- Cons:
  - Challenge: normalization, just like choropleths
    - Show population density (correlated with attribute), not effect of interest
  - Perceptual disadvantage: difficult to extract quantities
  - Performance disadvantage: rendering many dots can be slow

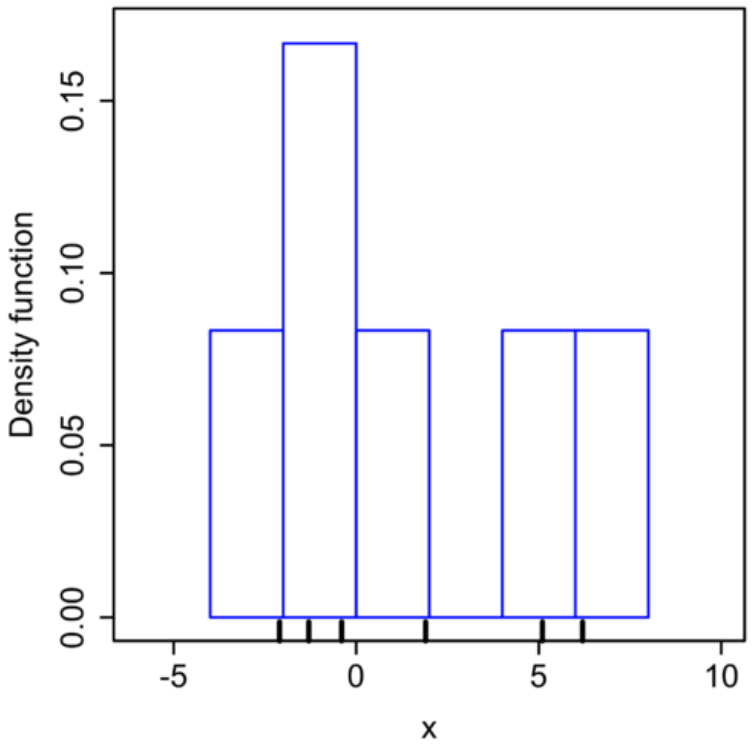
Heat maps

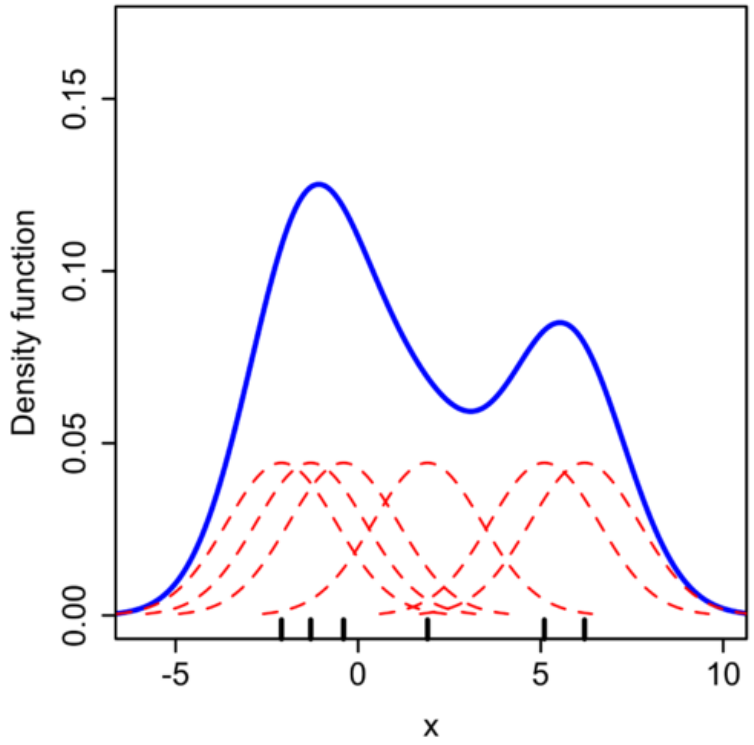
- Basically the same as dot density maps
- More continuous
  - Note: is an **estimate** of discrete data (dots)



Heat maps

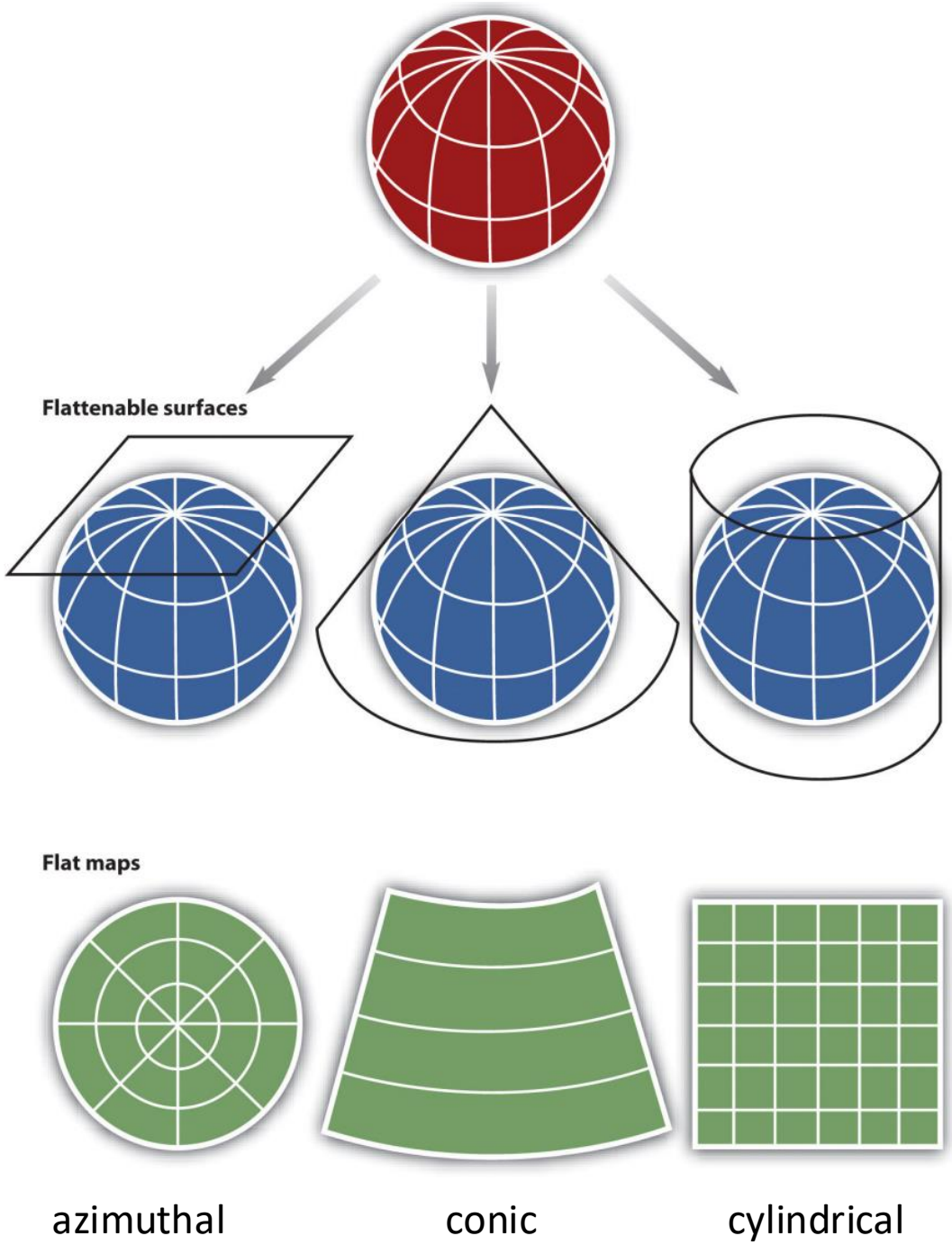
- Basically the same as dot density maps
- More continuous
  - Note: is an **estimate** of discrete data (dots)
- Typically, use a “density estimation” method to estimate a continuous distribution



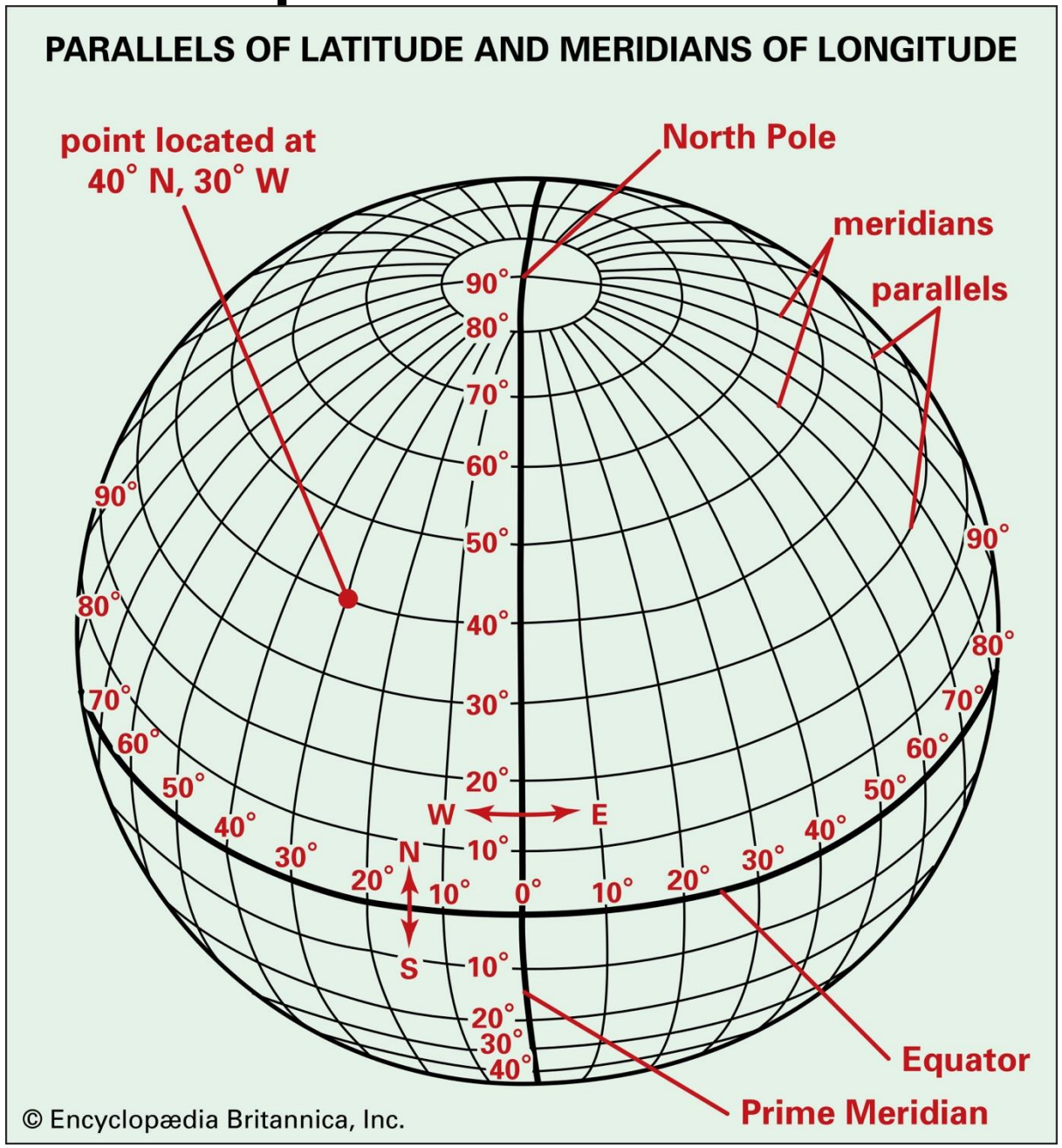


Map Projections

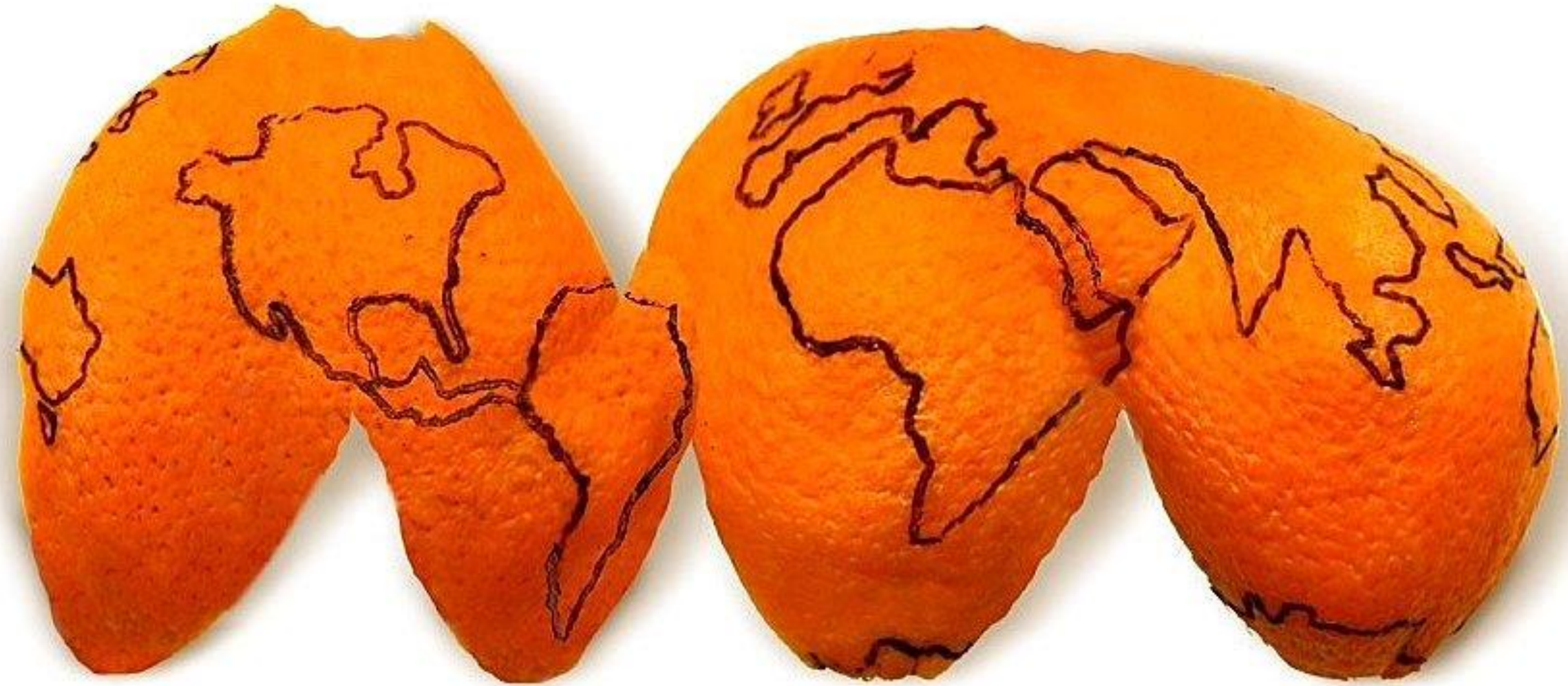
- Dimensionality reduction!
  - Loss of data
- Mathematical functions that map 3D surface geometry of the Earth to 2D maps
- All projections of sphere on plane necessarily distort surface in some way



Earth coordinates - spherical

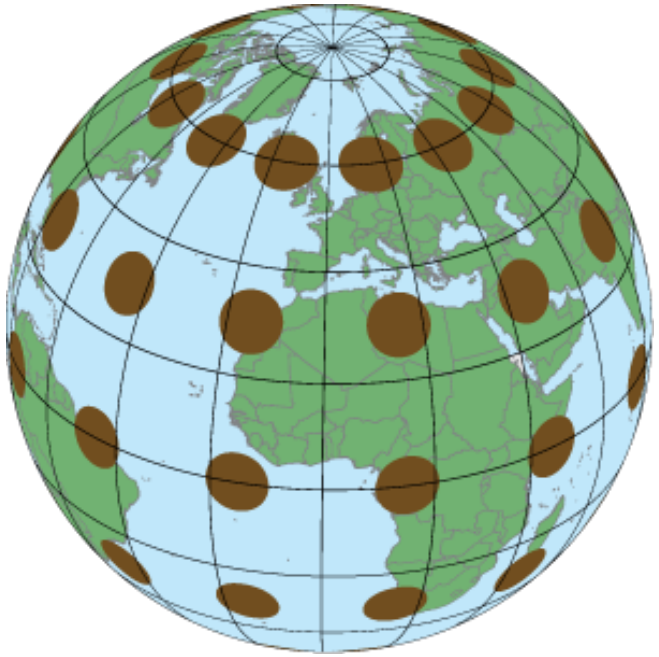


Earth coordinates - flattened

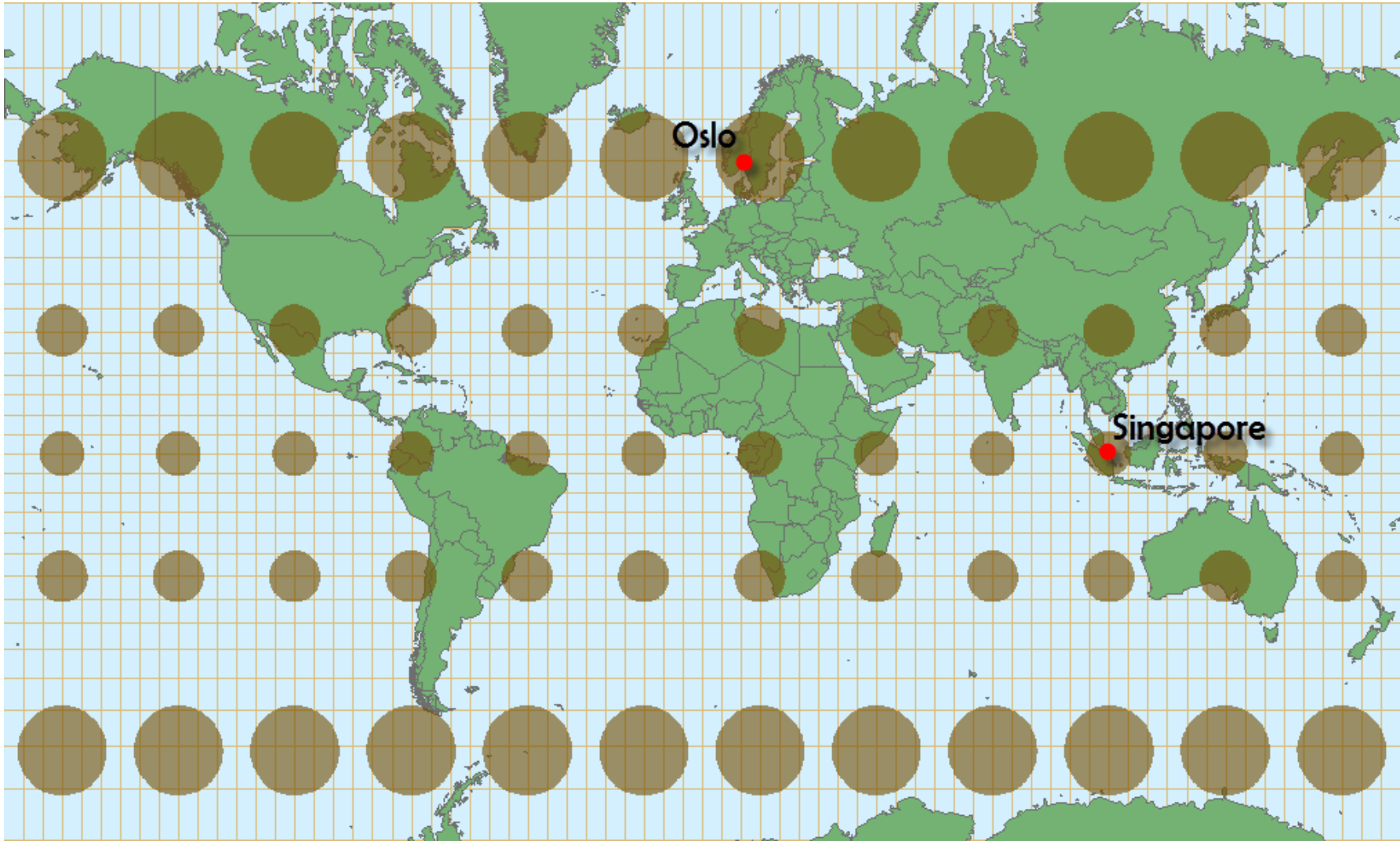


Projection

- Data is distorted
  - Need to preserve *something* (distance, area, angle).
  - Must choose

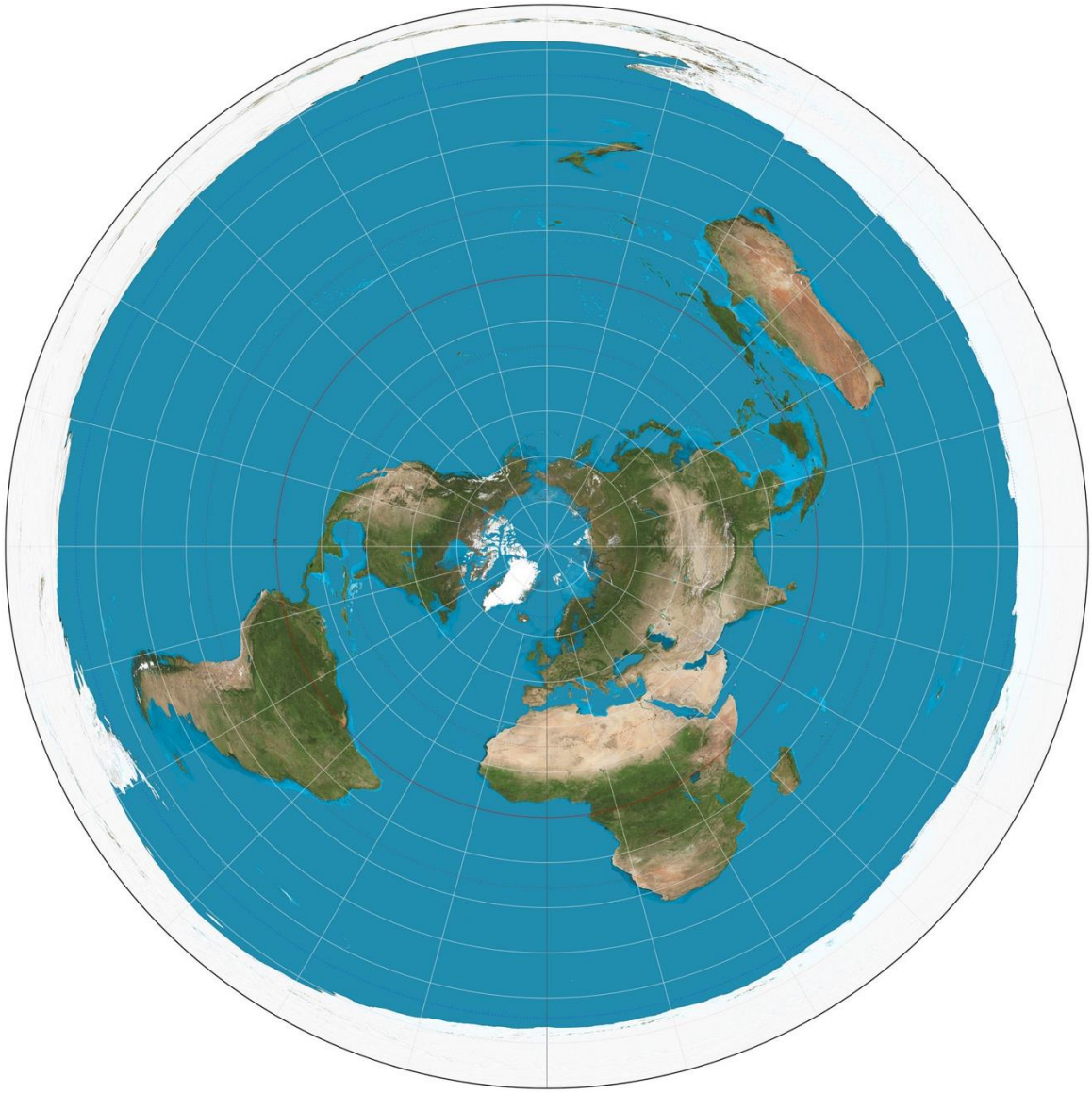


Tissot's indicatrix



Azimuthal Equidistant

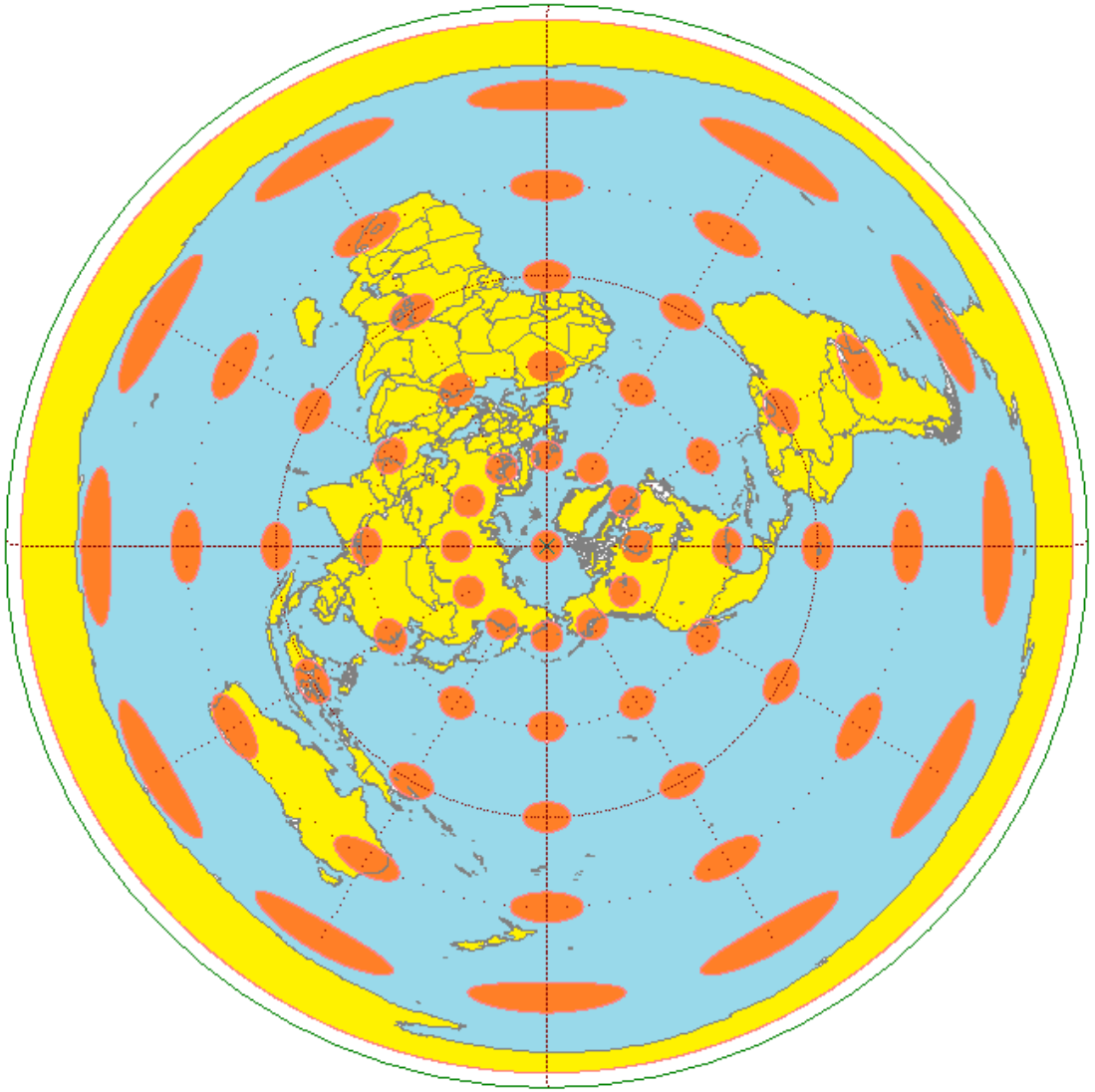
- Preserves distance and direction from center point.
- Used for travel, propagation from center point.



Azimuthal Equidistant

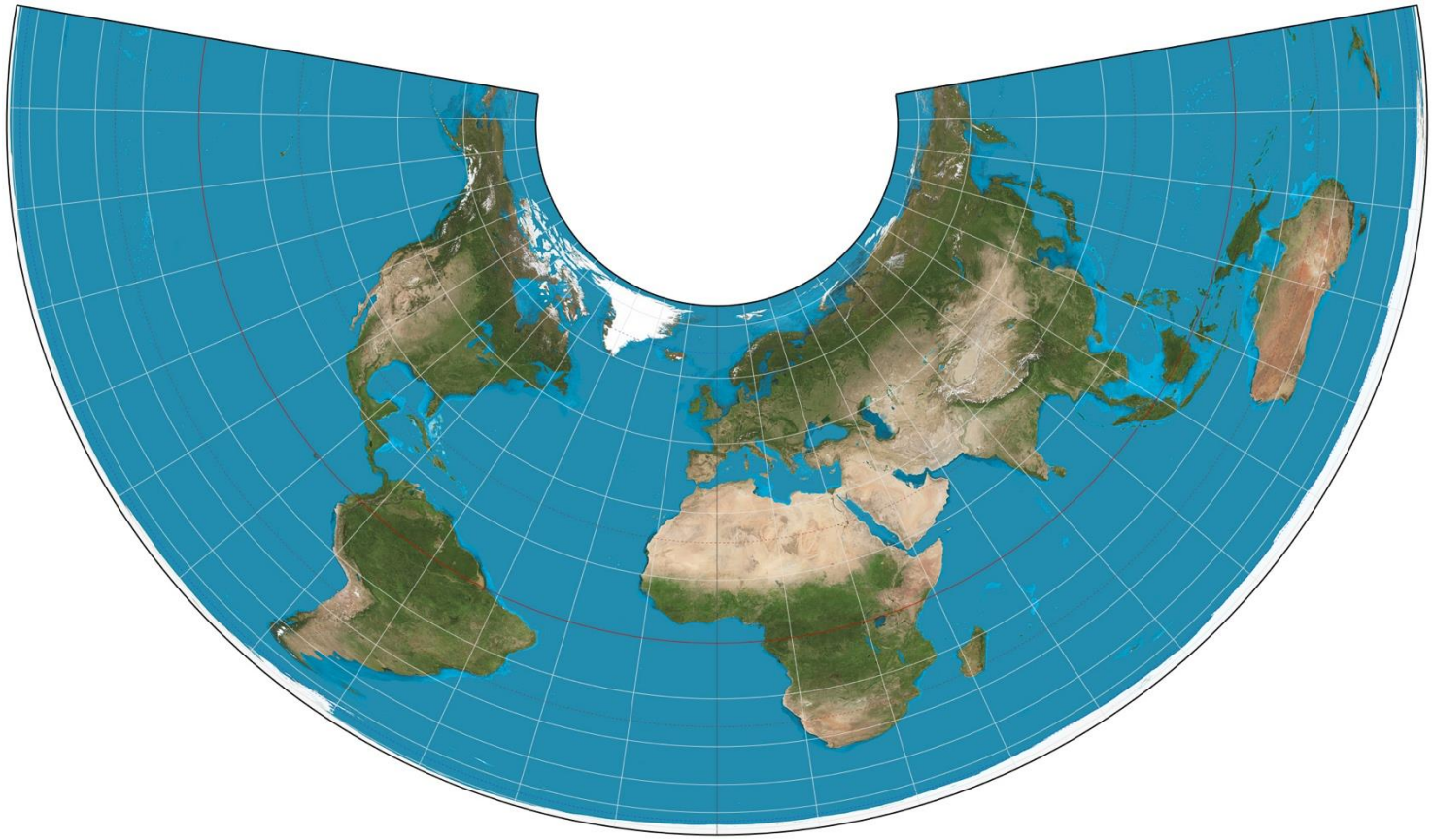
- Preserves distance and direction from center point.
- Used for travel, propagation from center point.

Tissot's indicatrix



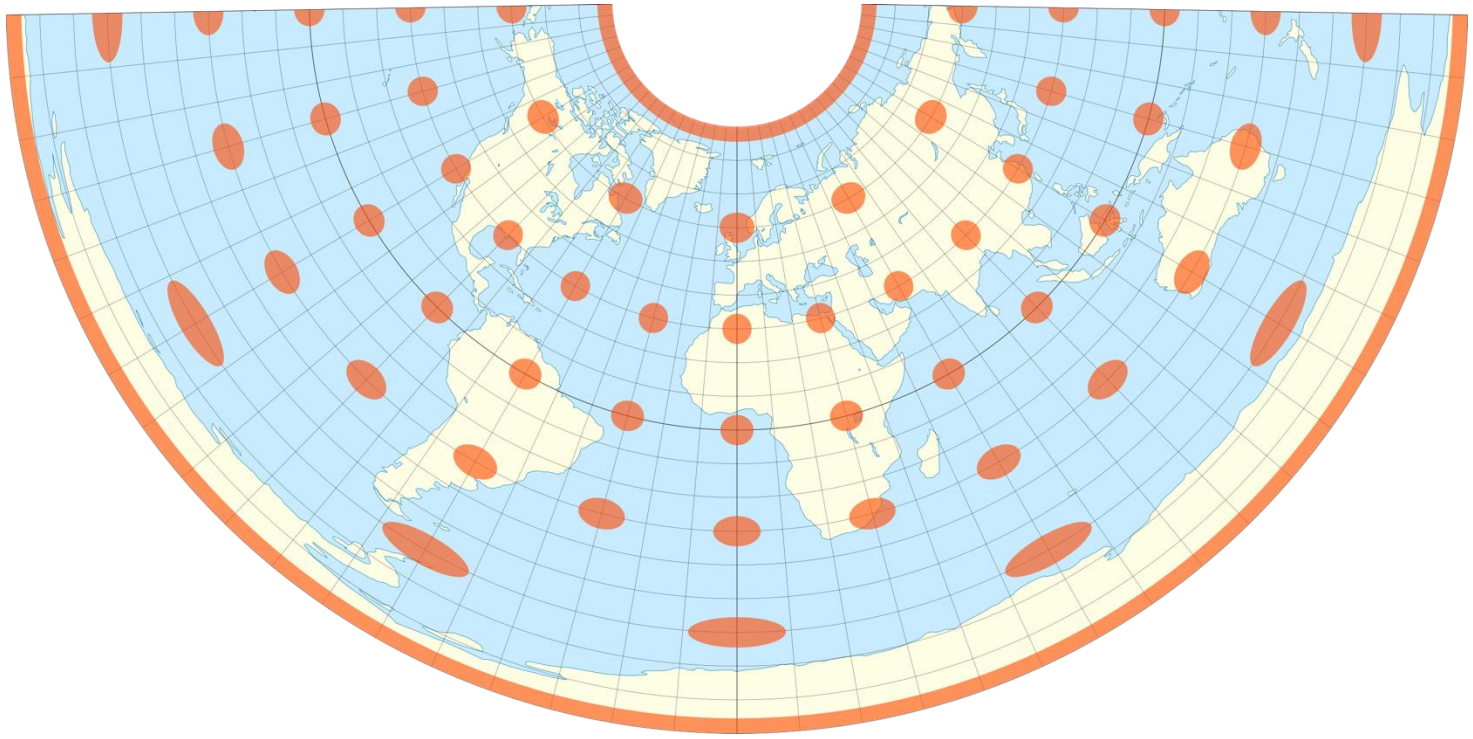
Equidistant conic

- Preserves proportional area of regions.
- Used for land survey



Equidistant conic

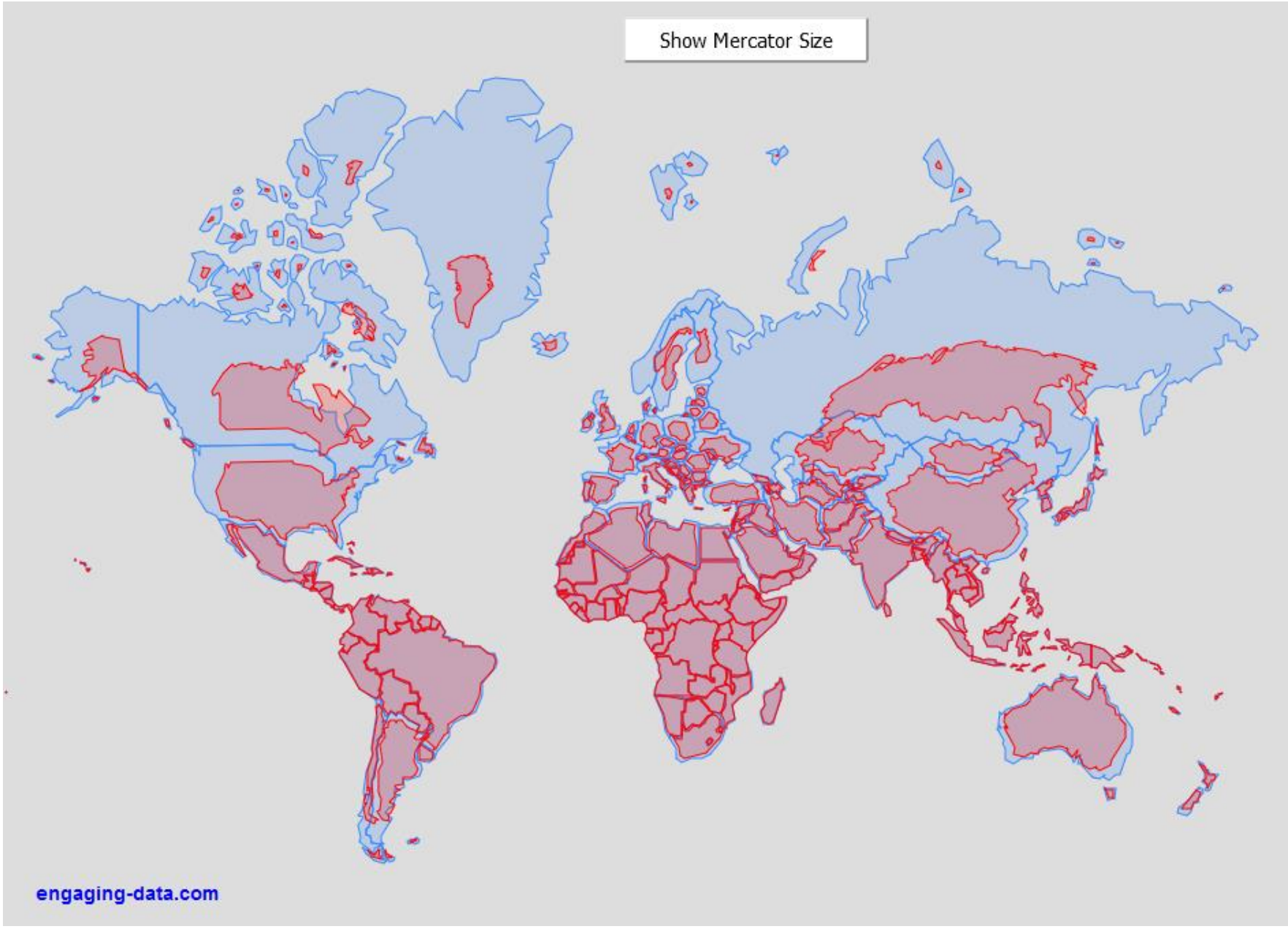
- Preserves proportional area of regions.
- Used for land survey



Tissot's indicatrix

Mercator Projection

- Preserves angle
- Used for navigation (and almost everything else)



Lots of others!

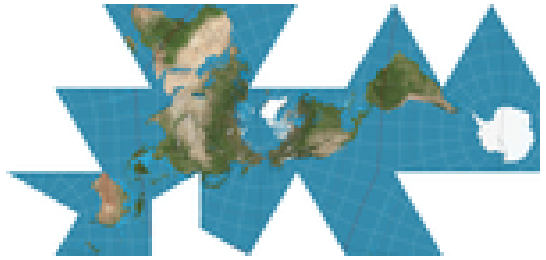
- https://philogb.github.io/page/myriahedral/ and https://www.jasondavies.com/maps/

Robinson

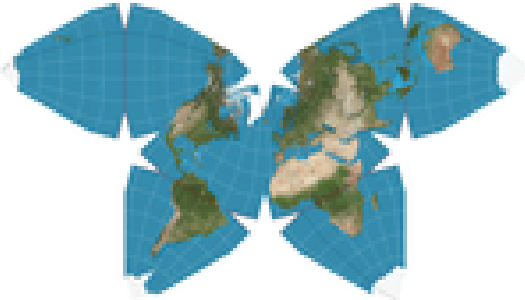
Goode homolosine

Gall-Peters

Orthographic

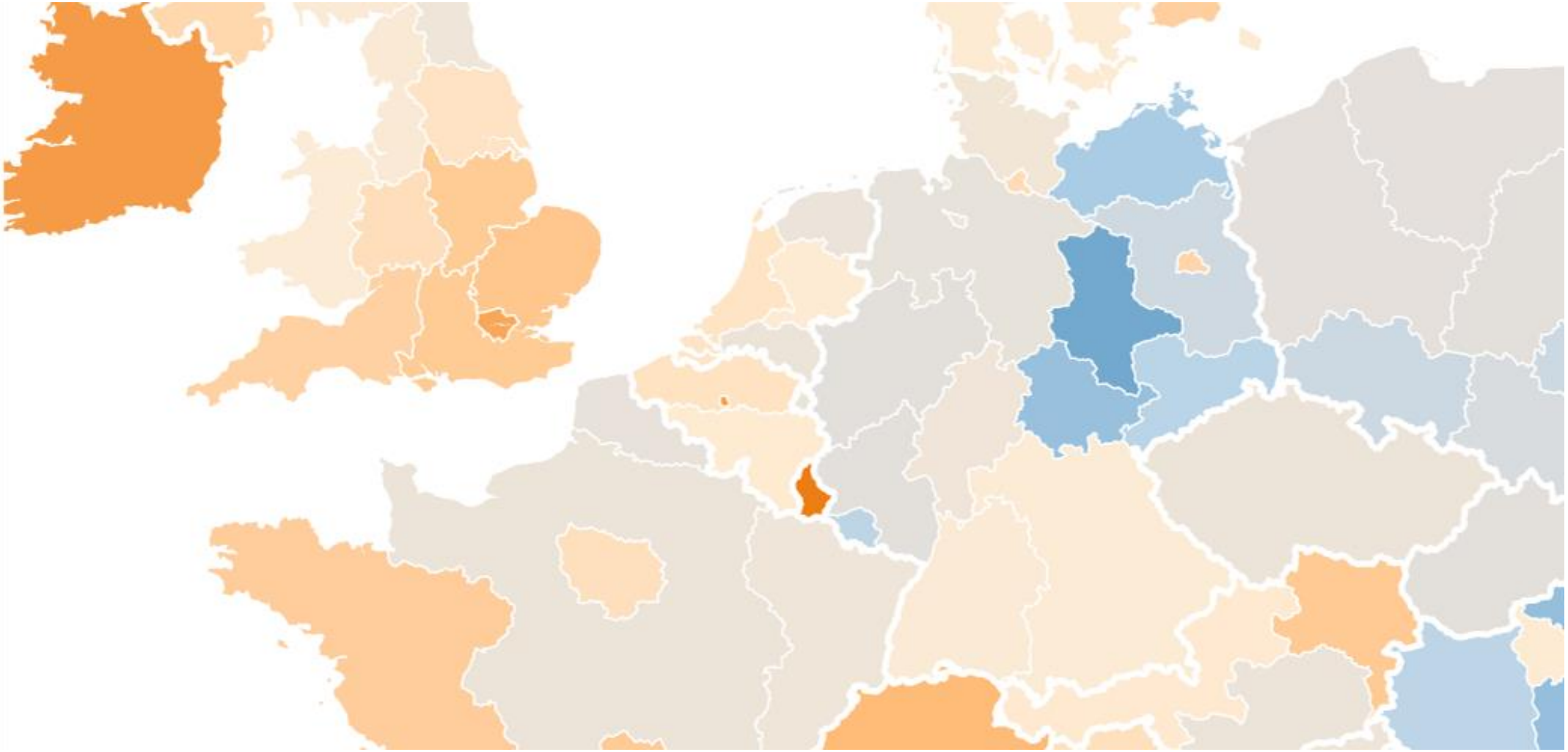


Dymaxion map



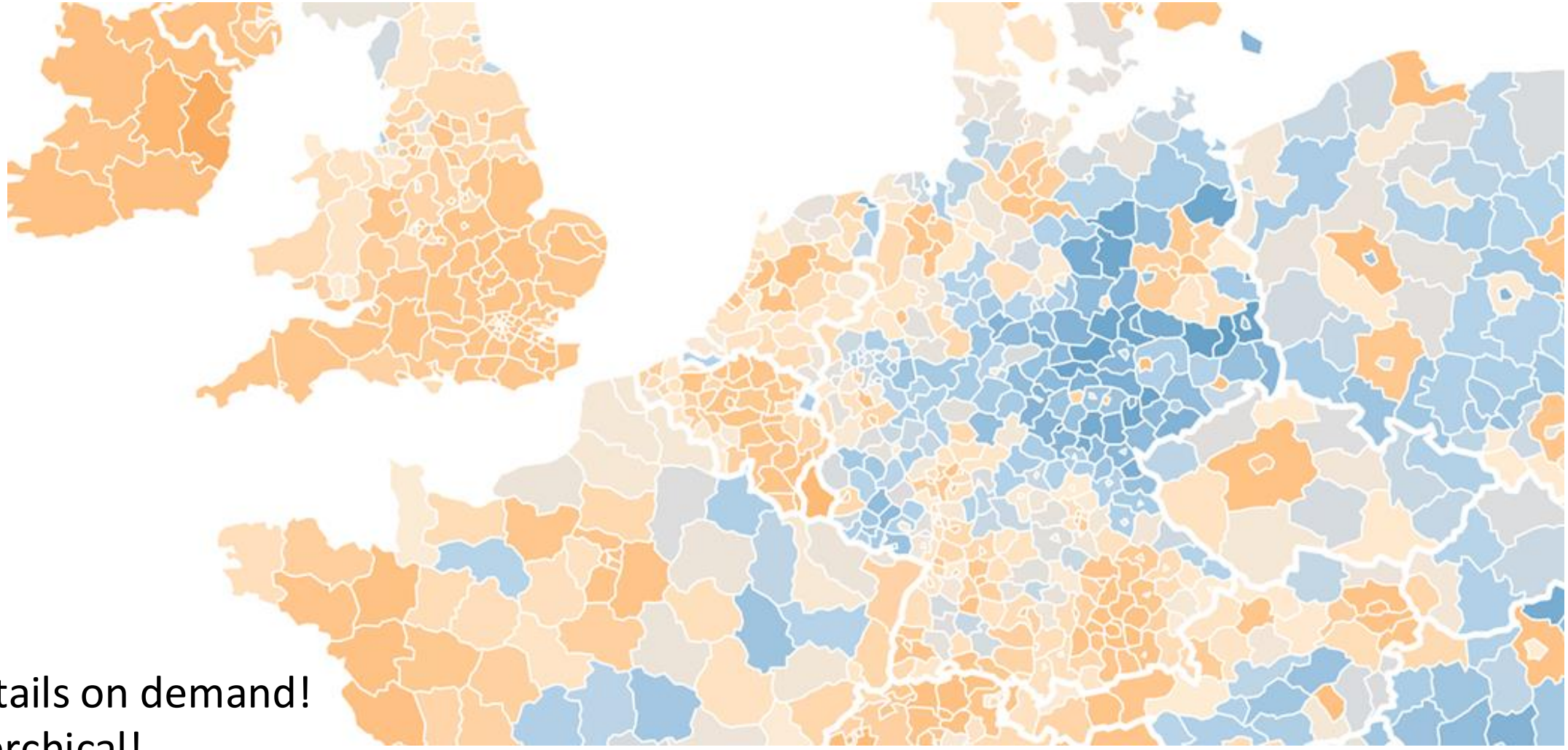
Waterman butterfly

Level of detail



Source: https://blog.datawrapper.de/weekly-chart-europegrowth/

Level of detail



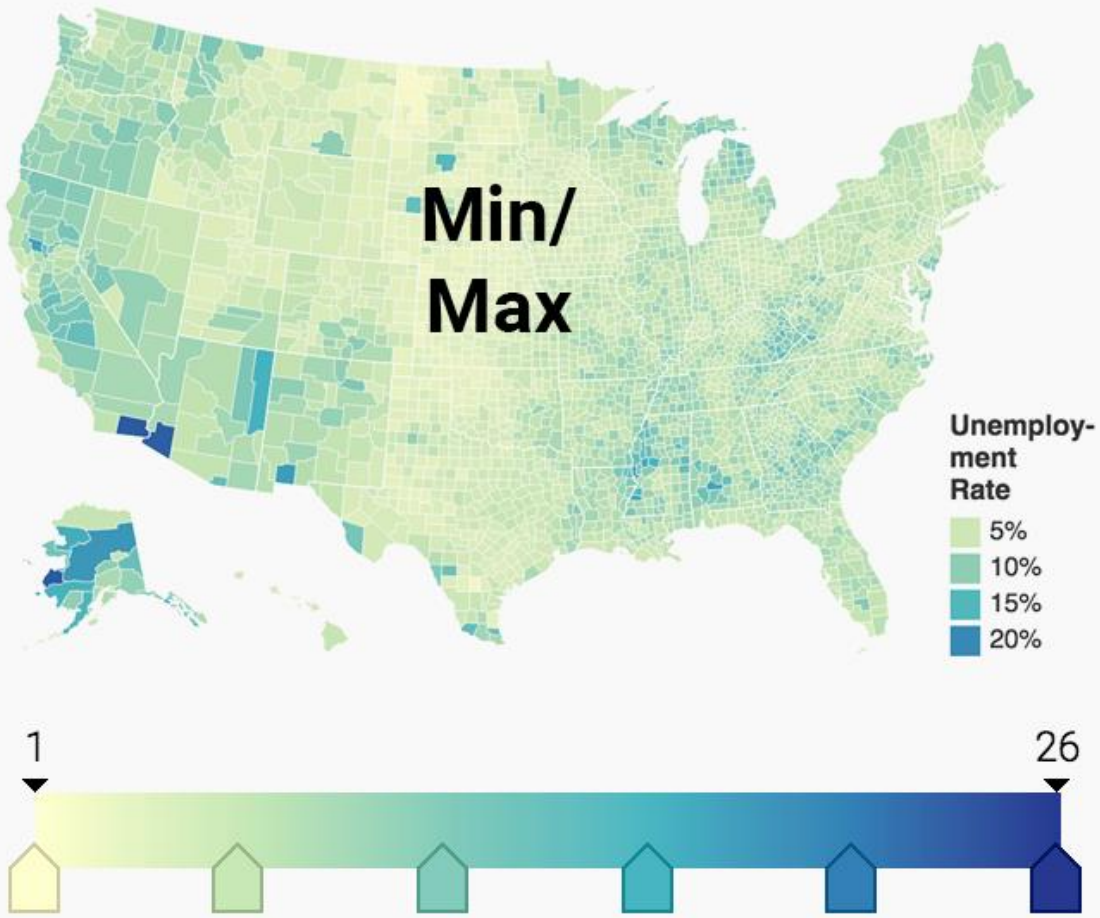
- Details on demand!
- Hierchical!

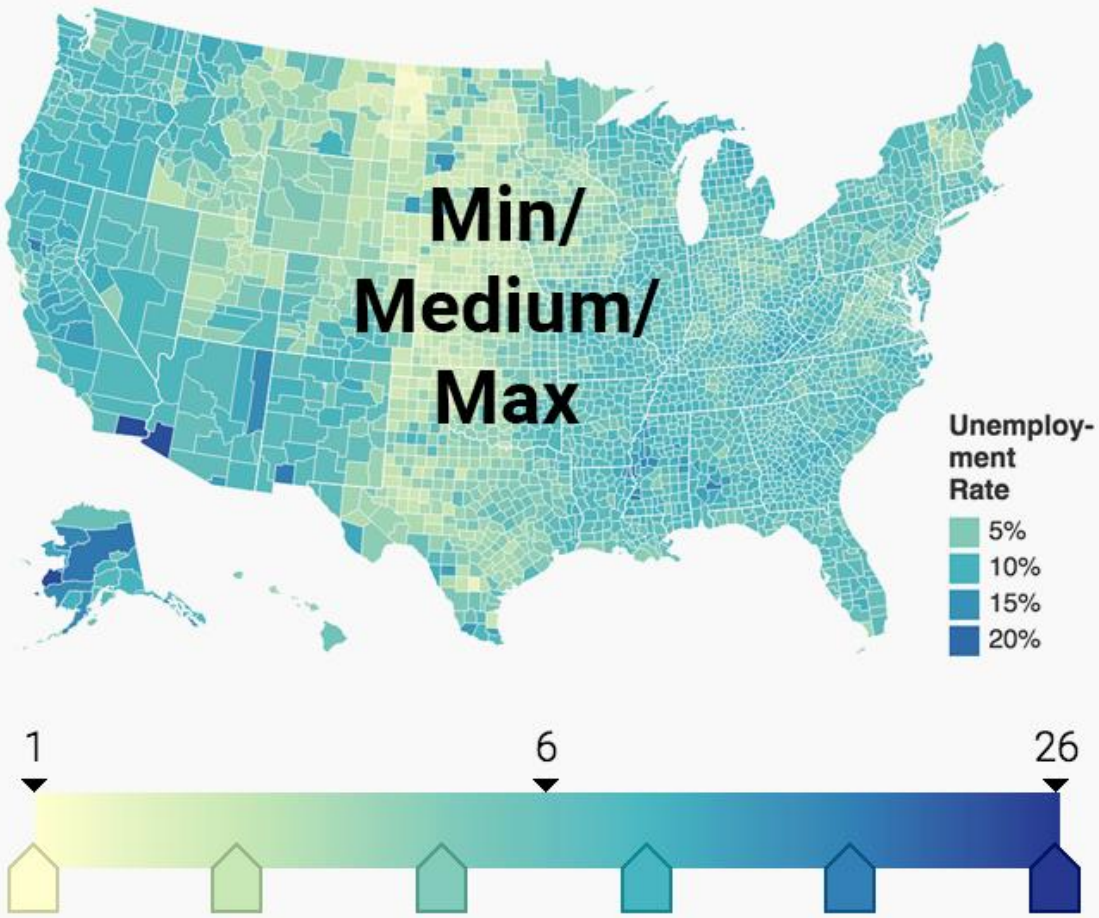
Source: https://blog.datawrapper.de/weekly-chart-europegrowth/

Common Issues

Color matters!

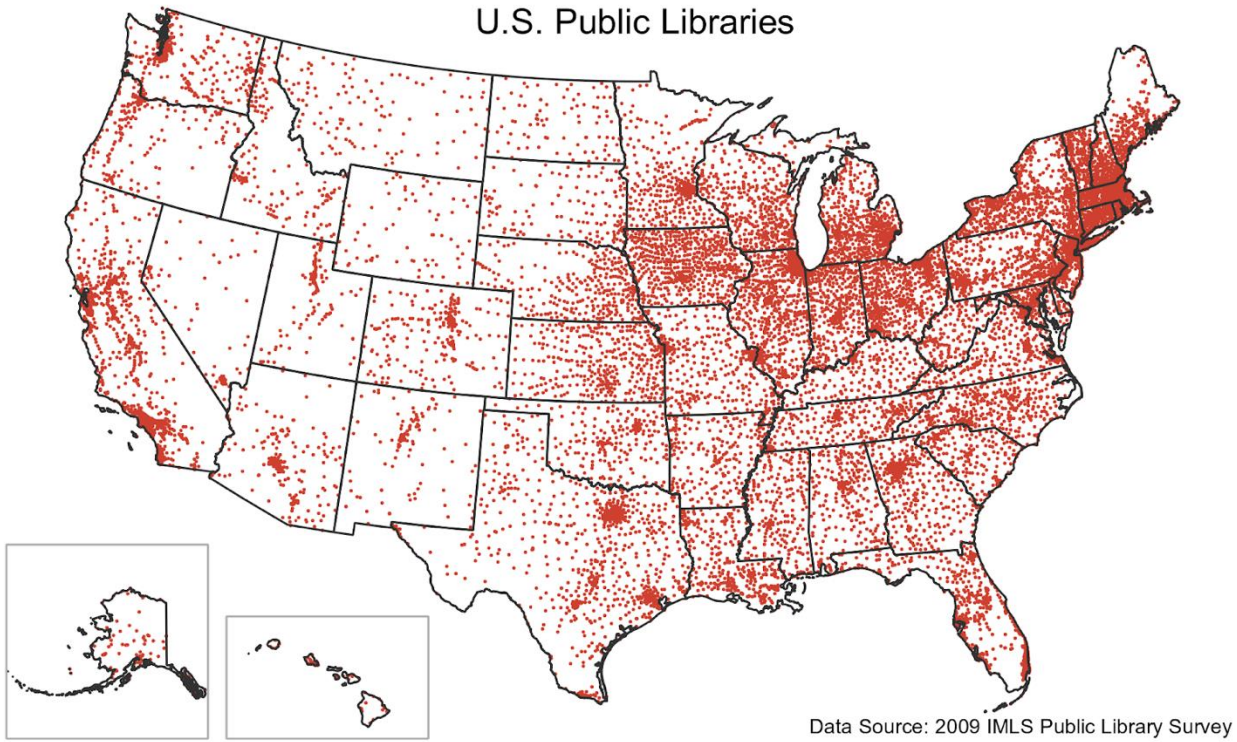
- Color map and data range are important

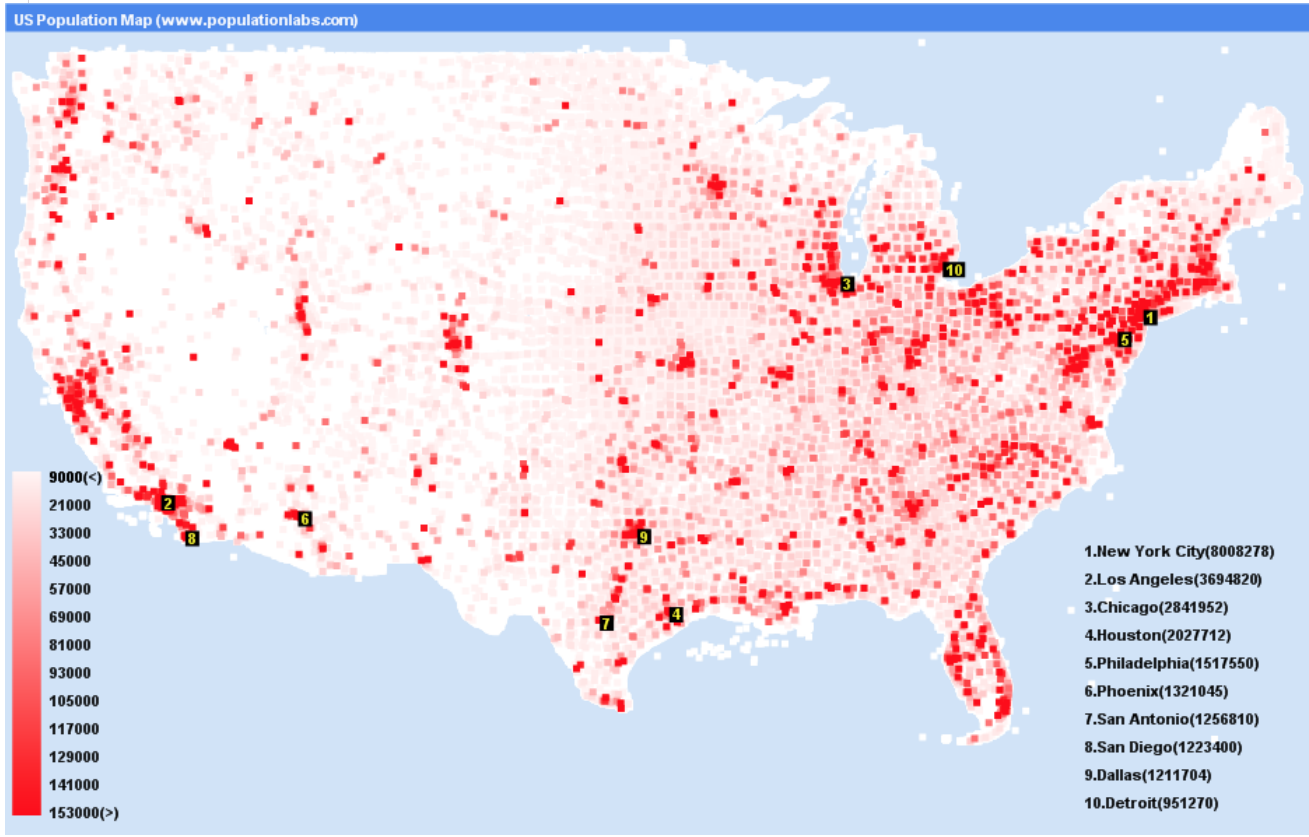




Source: https://academy.datawrapper.de/article/117-color-palette-for-your-map

Correlation with population density



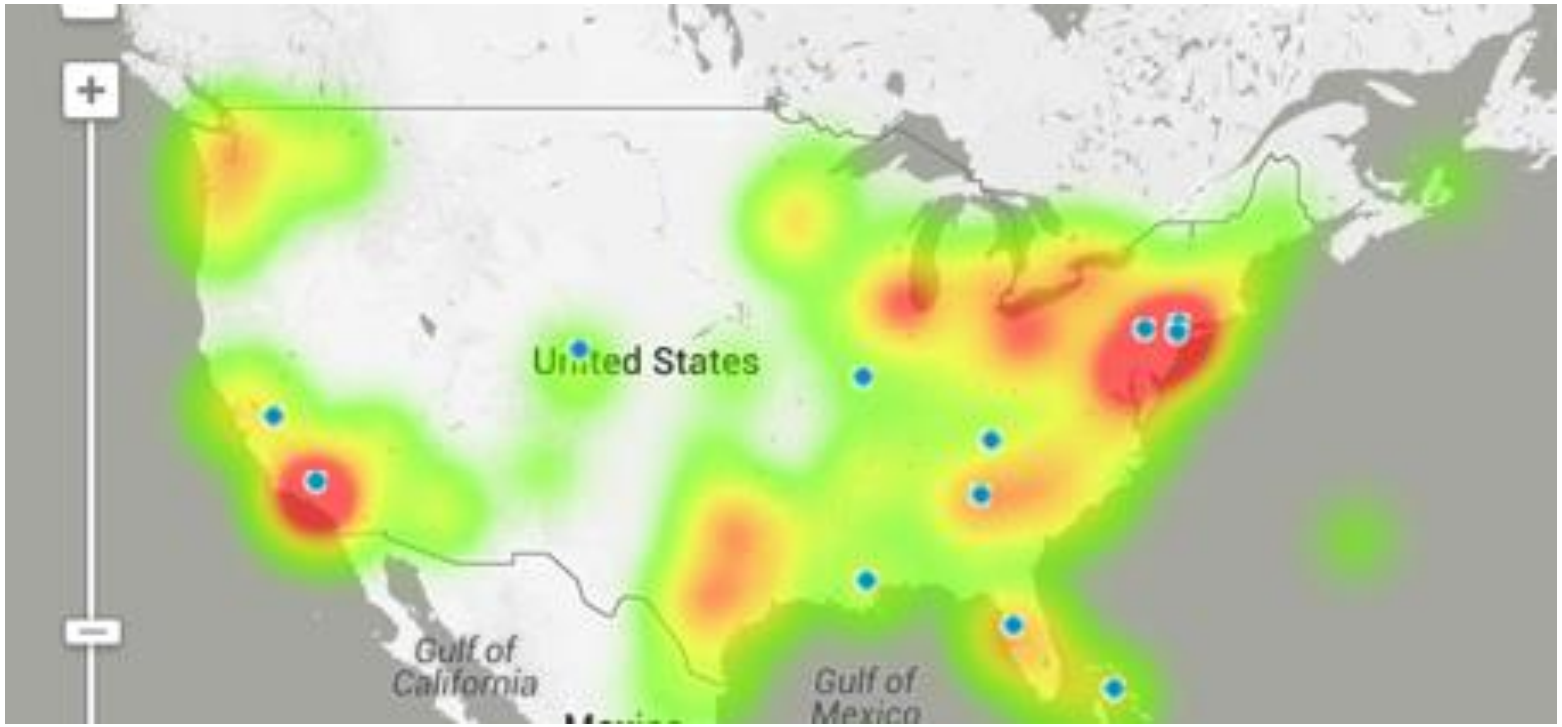


Correlation with population density

True or False?

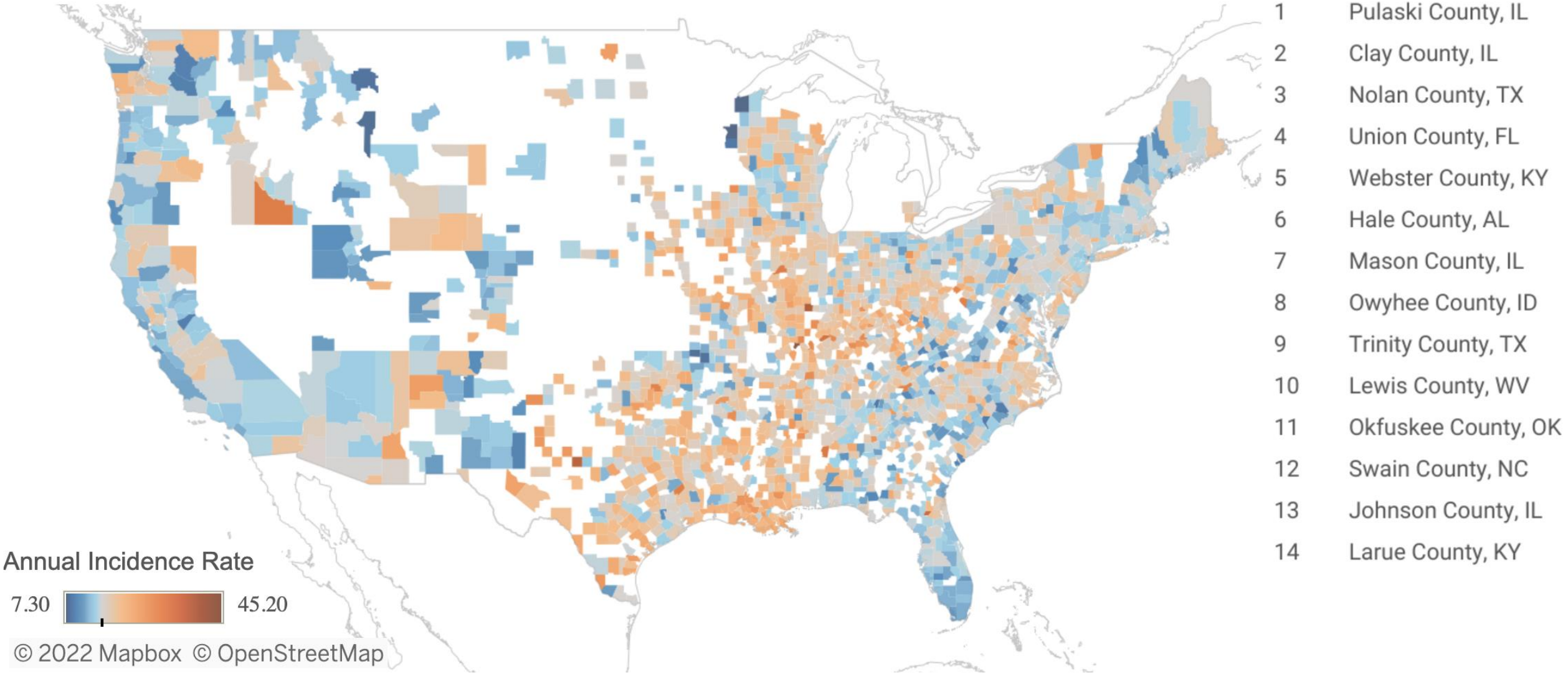
People in the west and east coast use Twitter more often than in the other regions.

Average Volume of Twitter Messages Posted in a Day



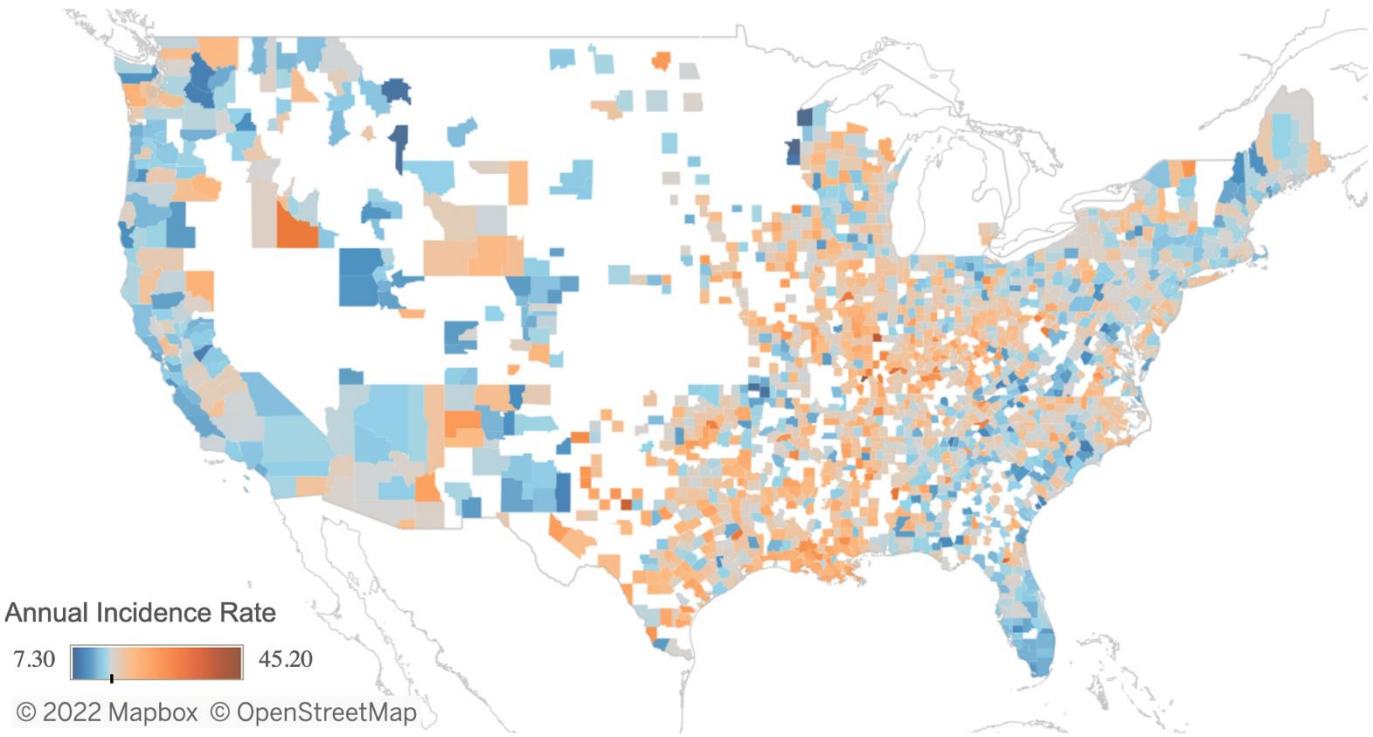
Insensitivity to Sample Size

Kidney Cancer and Insensitivity to Sample Size



Source: https://dataremixed.com/2015/01/avoiding-data-pitfalls-part-2/

Kidney Cancer and Insensitivity to Sample Size



1	Pulaski County, IL
2	Clay County, IL
3	Nolan County, TX
4	Union County, FL
5	Webster County, KY
6	Hale County, AL
7	Mason County, IL
8	Owyhee County, ID
9	Trinity County, TX
10	Lewis County, WV
11	Okfuskee County, OK
12	Swain County, NC
13	Johnson County, IL
14	Larue County, KY

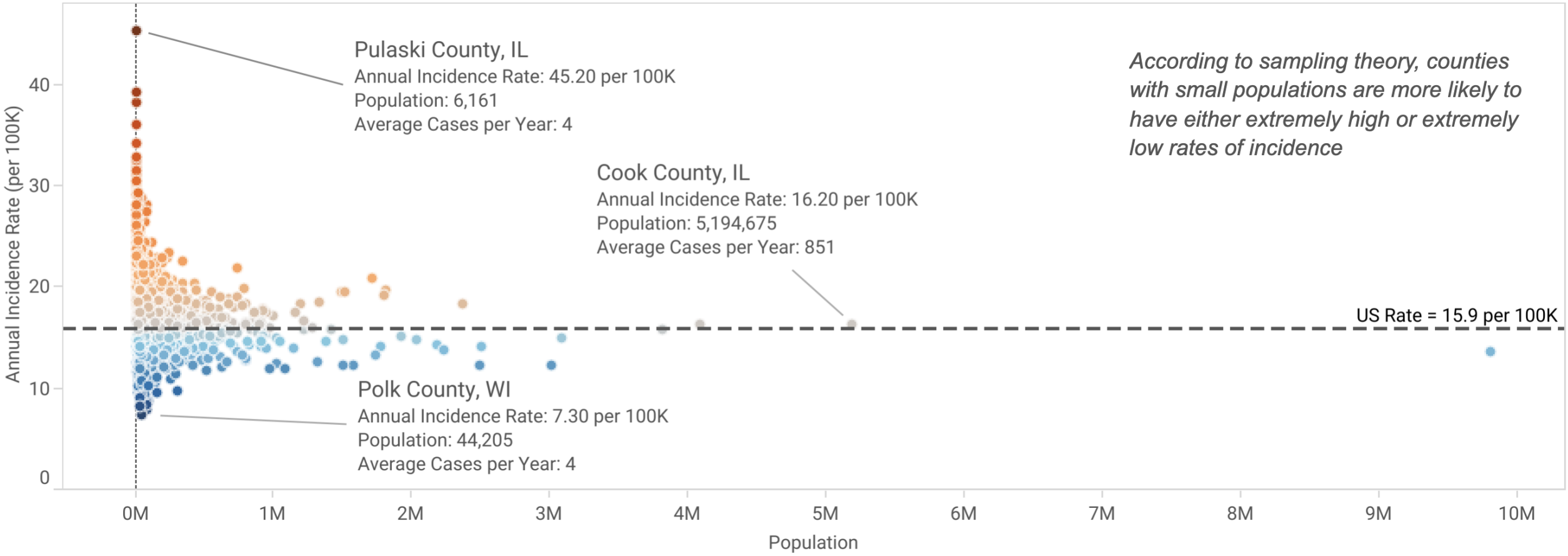
Wainer, Howard, and Harris L. Zwerling. "**Evidence that smaller schools do not improve student achievement**." *Phi Delta Kappan* 88.4 (2006): 300-303.

*"The counties in which the incidence of kidney cancer is **lowest** are mostly rural, sparsely populated, and located in traditionally Republican states in the Midwest, the South, and the West." p.109*

*"Now consider the counties in which the incidence of kidney cancer is **highest**. These ailing counties tend to be mostly rural, sparsely populated, and located in traditionally Republican states in the Midwest, the South, and the West."*

Source: https://dataremixed.com/2015/01/avoiding-data-pitfalls-part-2/

Insensitivity to Sample Size

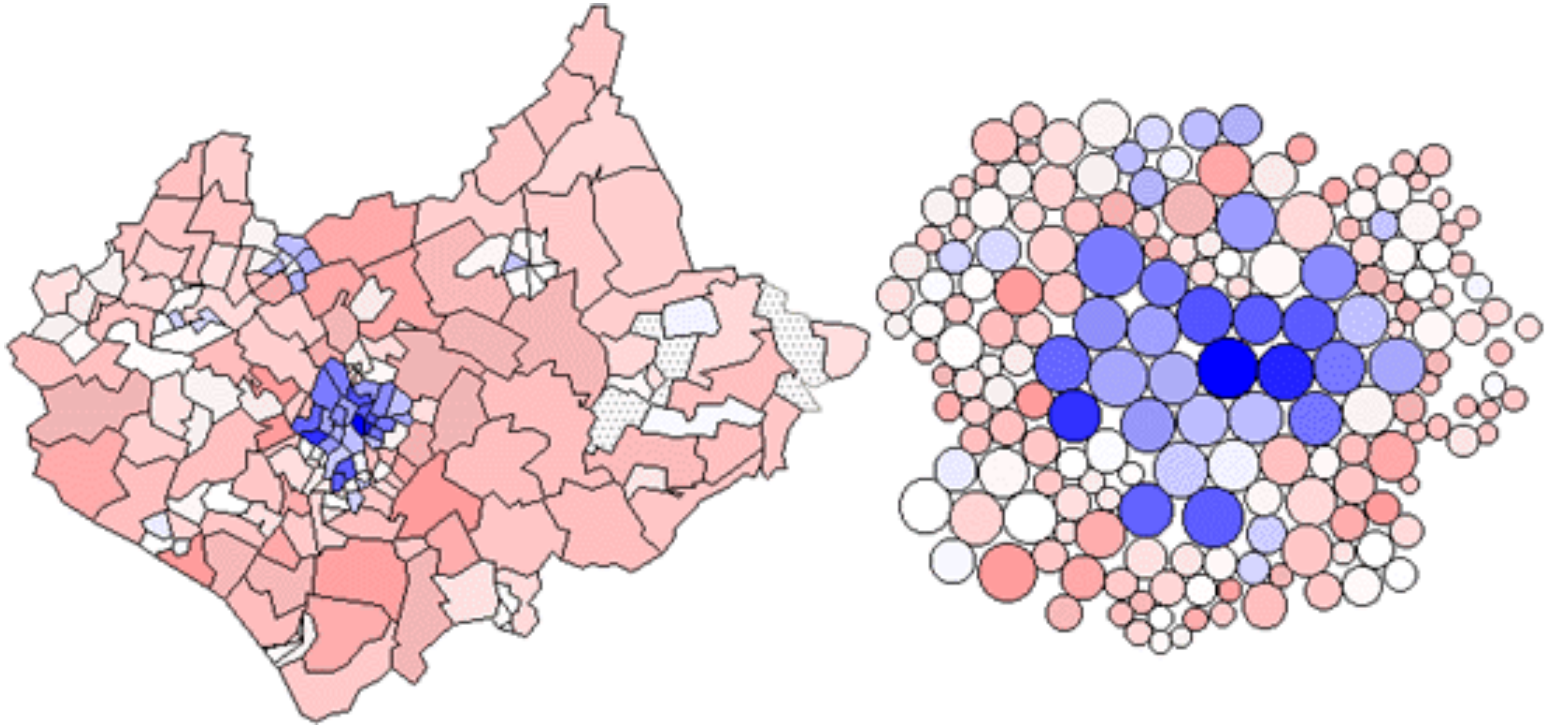


Source: https://dataremixed.com/2015/01/avoiding-data-pitfalls-part-2/

Skewed Spatial Distributions

Skewed Spatial Distributions

- Dorling Cartogram
  - Regions are represented by symbols instead of region borders.
  - Symbol size is not proportional to region size.
  - Topological relationship is preserved approximately.

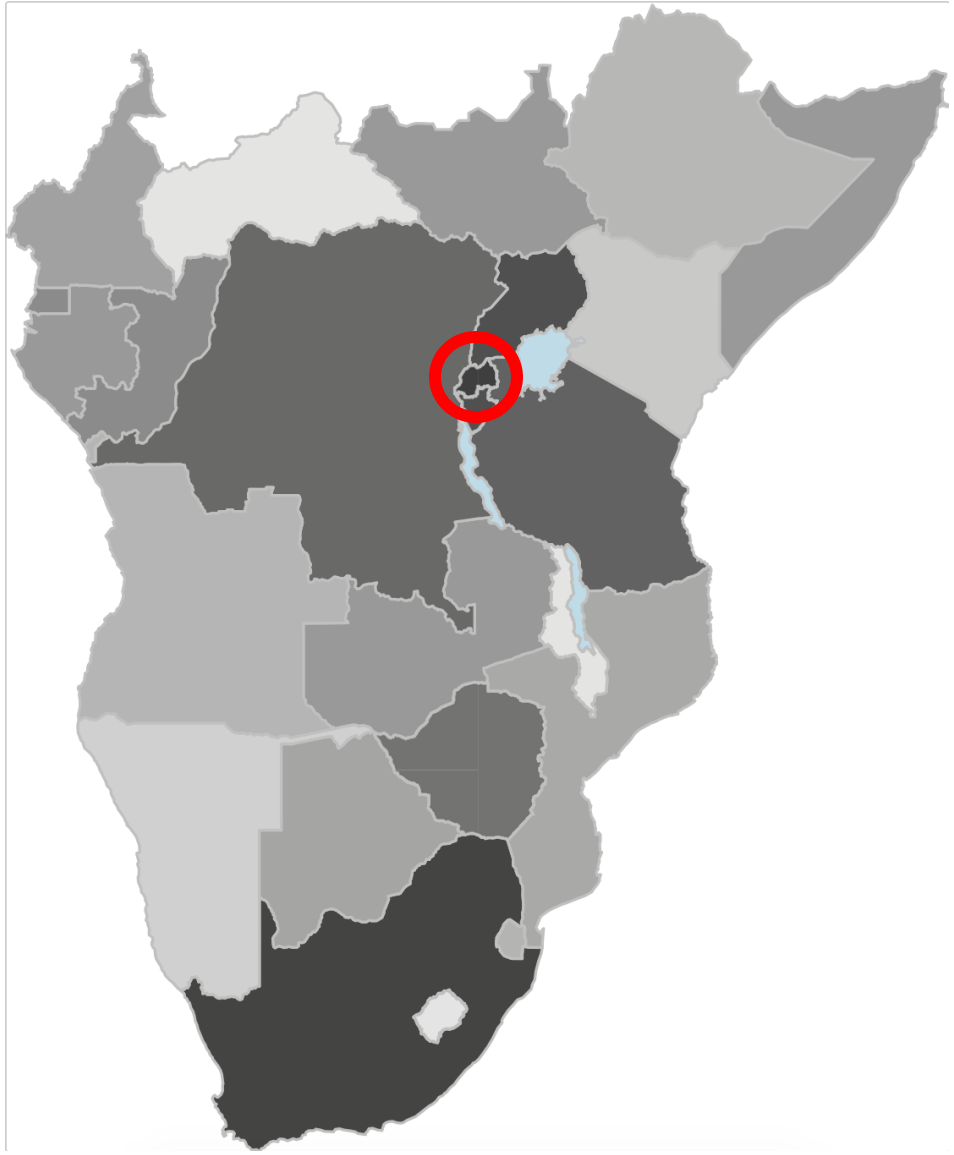


http://www.agocg.ac.uk/reports/visual/casestud/dykes/issue2_1.htm

Perception issues

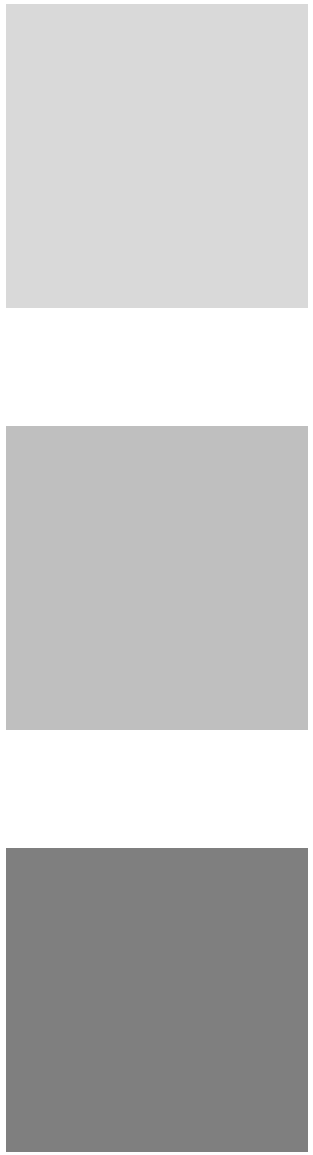
- Which country in Africa has the highest value?

- Why?
  - Color perception is highly affected by size.
  - Color of the surrounding regions affects the way we perceive color.

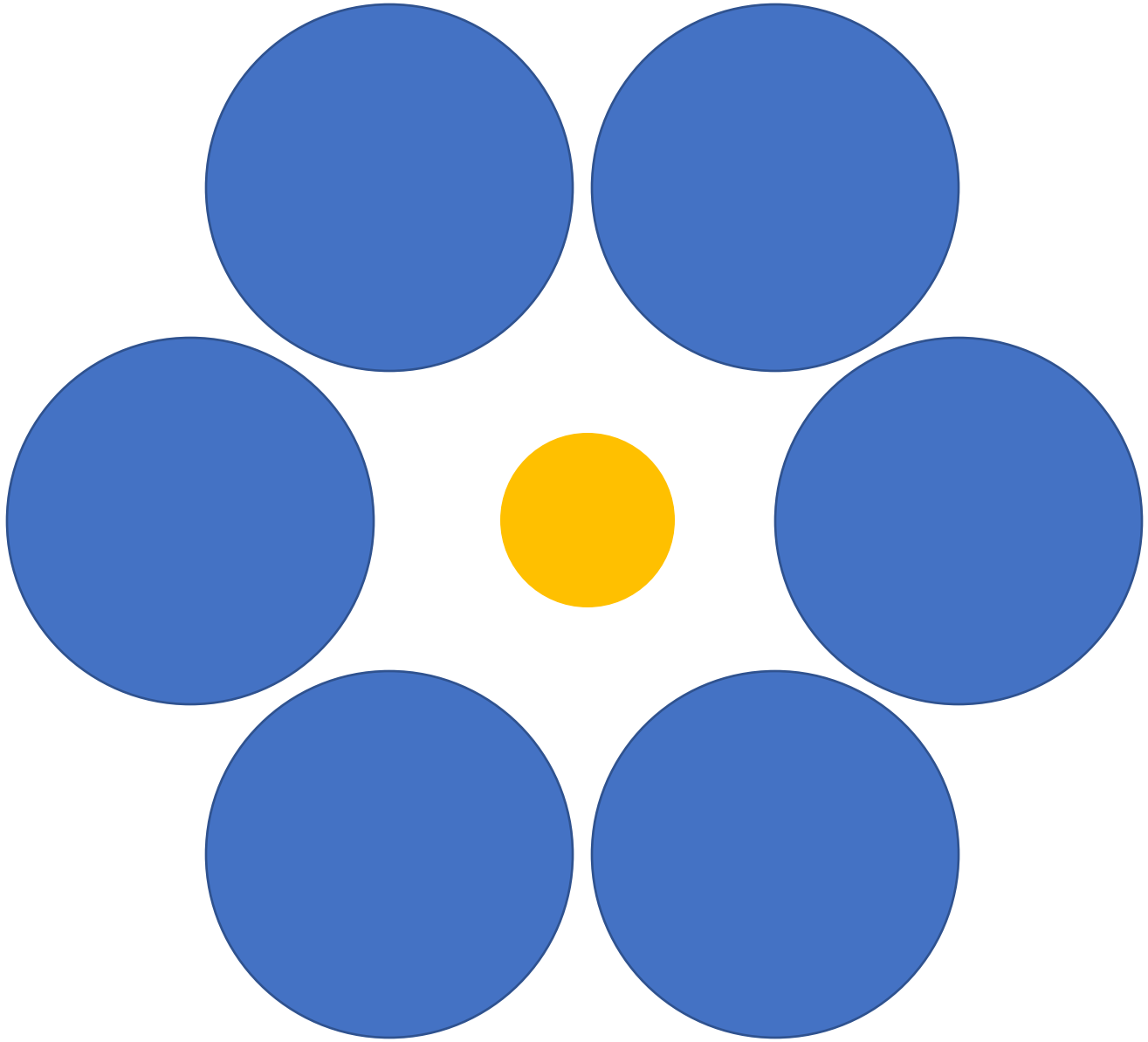


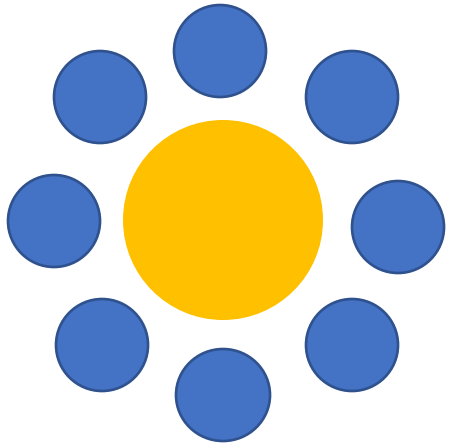
STEPHEN. FEW. *Signal: Understanding what matters in a world of noise*. Routledge, 2015.

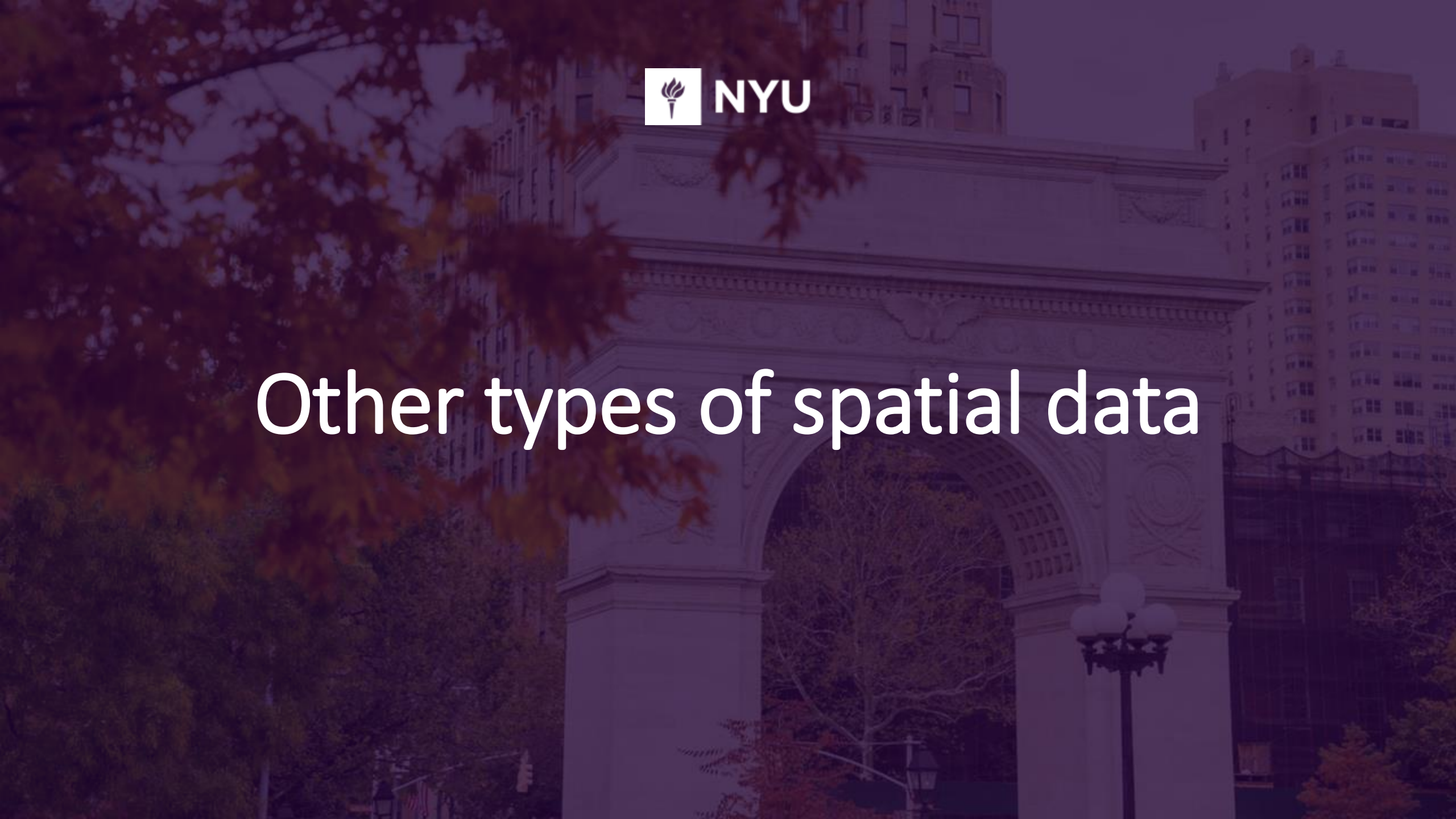
Perception issues



Perception issues







Other types of spatial data

Topographic map

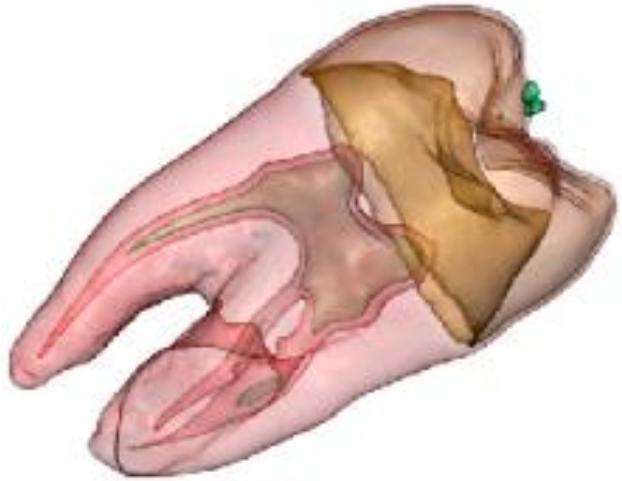
- Data format is:
  - Geographic geometry
  - Scalar spatial field
    - 1 quant attribute per grid cell
- Visualize derived data
  - Isoline geometry
    - Isocontours computed for specific levels of scalar values
- Task:
  - Understanding terrain shape
    - Densely lined regions = steep

Topographic map: pros and cons

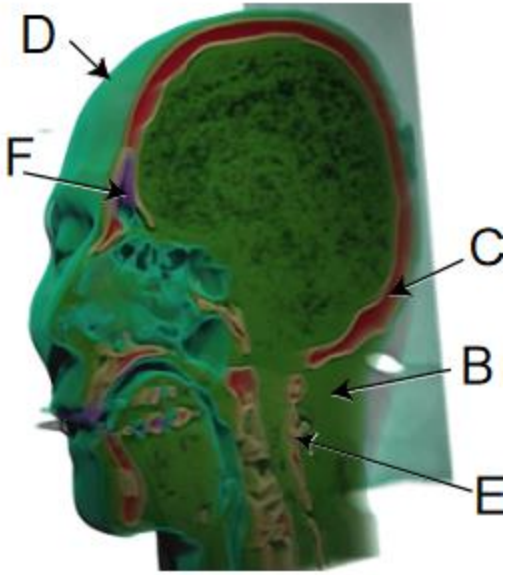
- Pros:
  - use only 2D position, avoid 3D challenges
  - color channel available for other attributes
- Cons:
  - significant clutter from additional lines

Isosurfaces & direct volume rendering

- Data format is:
  - Scalar spatial field (3D volume)
    - 1 quant attribute per grid cell
- Task:
  - Shape understanding & spatial relationship



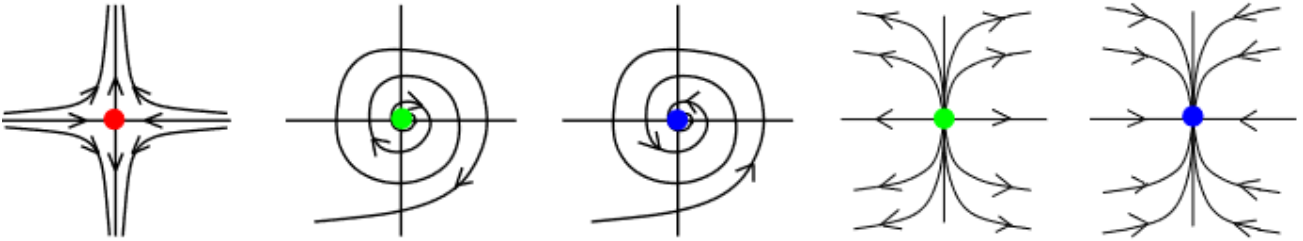
- Isosurface
  - Derived data: isocontours computed for specific levels of scalar values



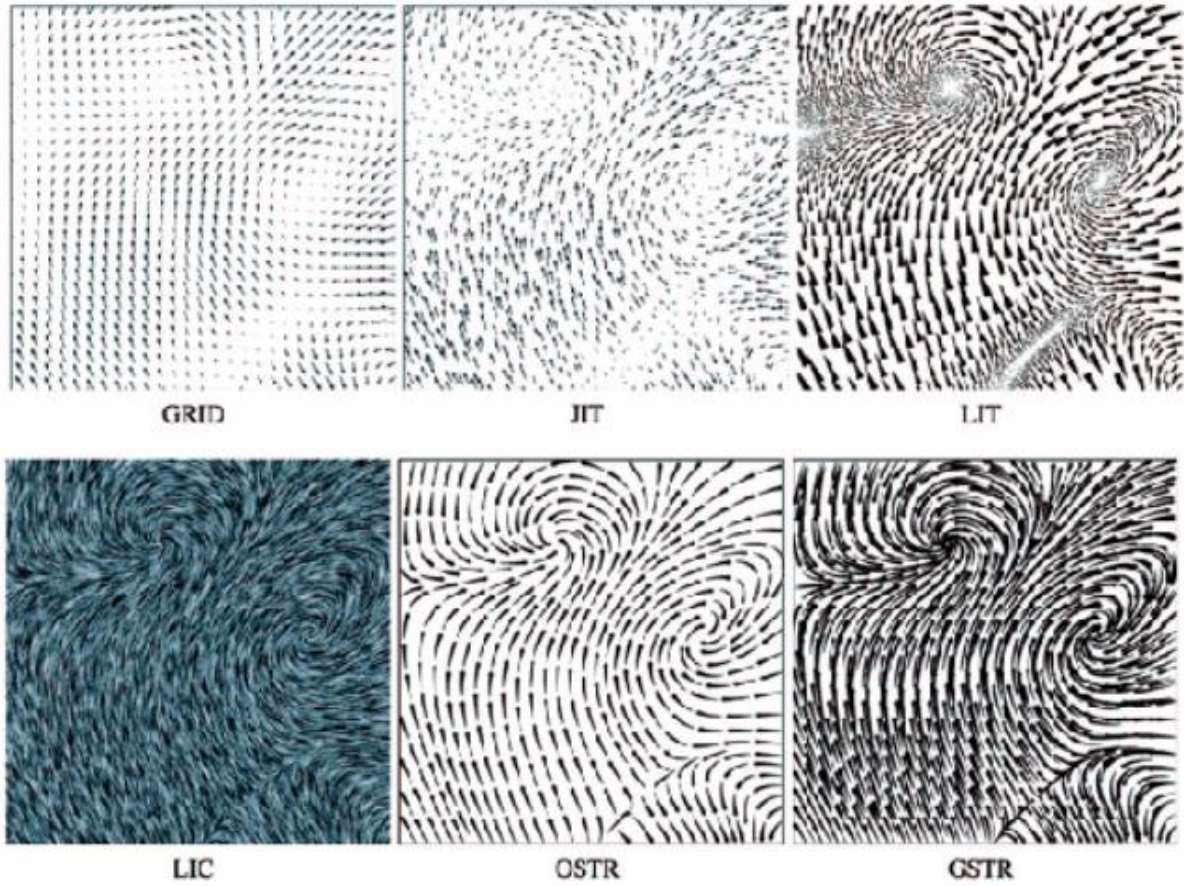
- Direct volume rendering
  - Transfer function maps scalar values to color, opacity
    - No derived geometry → comes from the data

Vector and tensor fields

- Data format:
  - Multiple attributes per cell (vector → 2 attributes)
- Tasks:
  - Find critical points, identify their types
  - Identify what type of critical point at a specific location
  - Predicting where a particle will end up based on starting position



Topology tracking for the visualization of time-dependent two-dimensional flows. Tricoche,Wischgoll, Scheuermann, and Hagen. Computers & Graphics 26:2 (2002), 249–257



Comparing 2D vector field visualization methods: A user study. Laidlaw et al. IEEE Trans. Visualization and Computer Graphics (TVCG) 11:1 (2005), 59–70.

Break

Final Project Time

- Find partners
- Brainstorm ideas
- Ask me questions