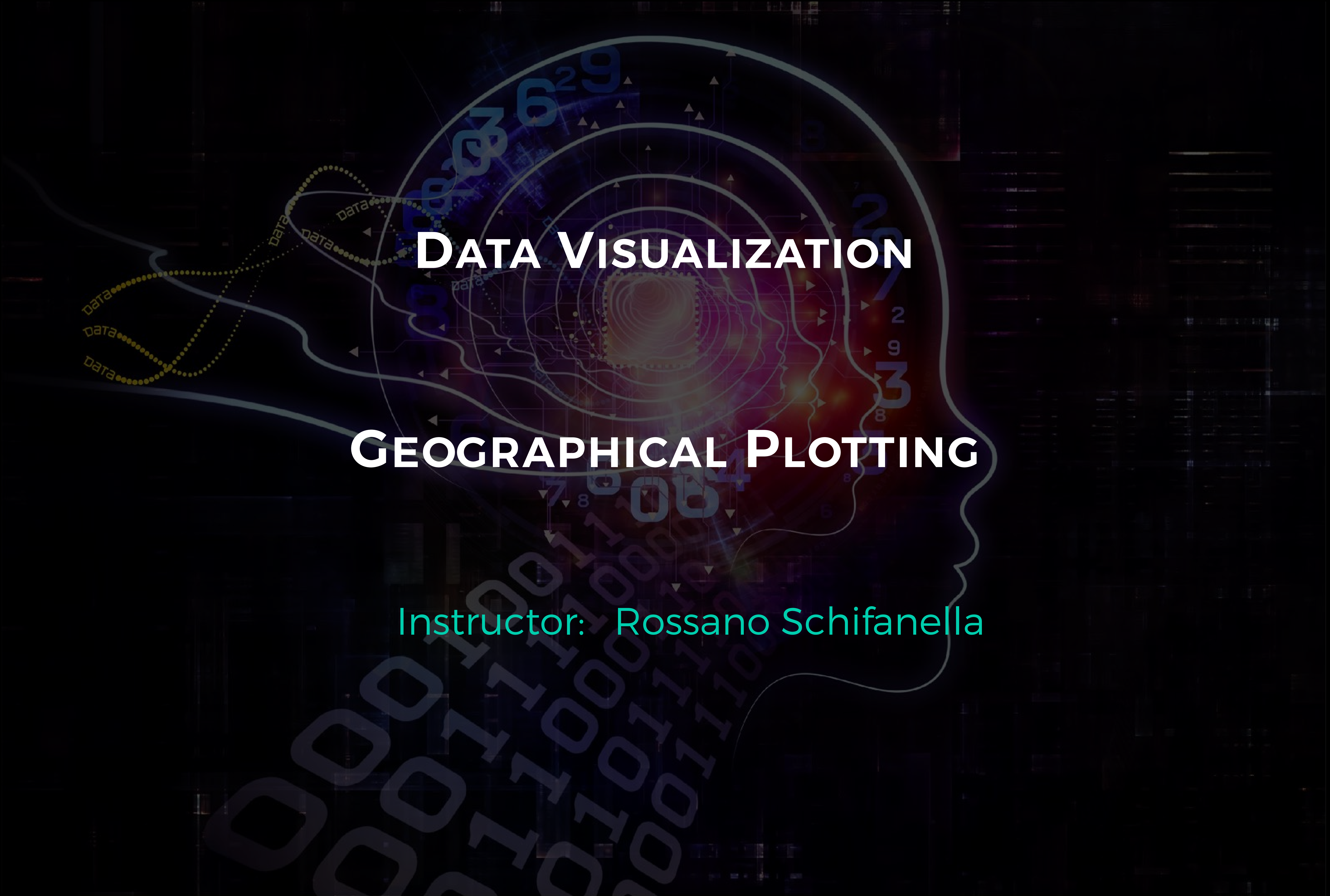


Data Visualization

Geographical Plotting

Instructor: Rossano Schifanella

learning objectives

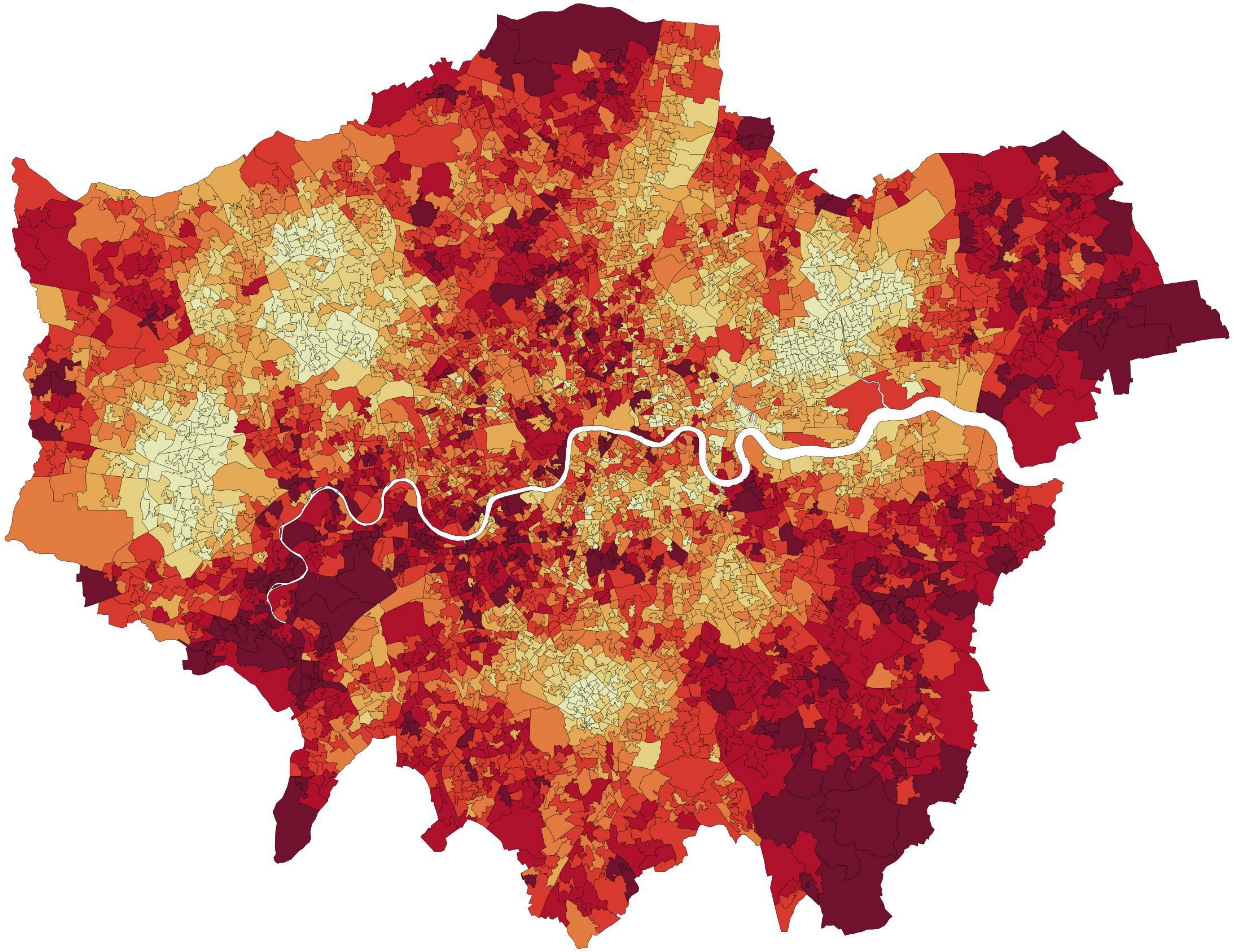
- **The importance of spatial thinking**
- **Visually exploring spatial phenomena**
    - Learn the basic steps to create an informative thematic cartography
    - Know the main thematic cartography types
    - Know some basic thematic cartography rules of thumb
- **Pitfalls of using spatial data in computational disciplines**

The importance of spatial thinking

Spatial patterns matter

Percentage of white people in London (LSOA, census 2011)

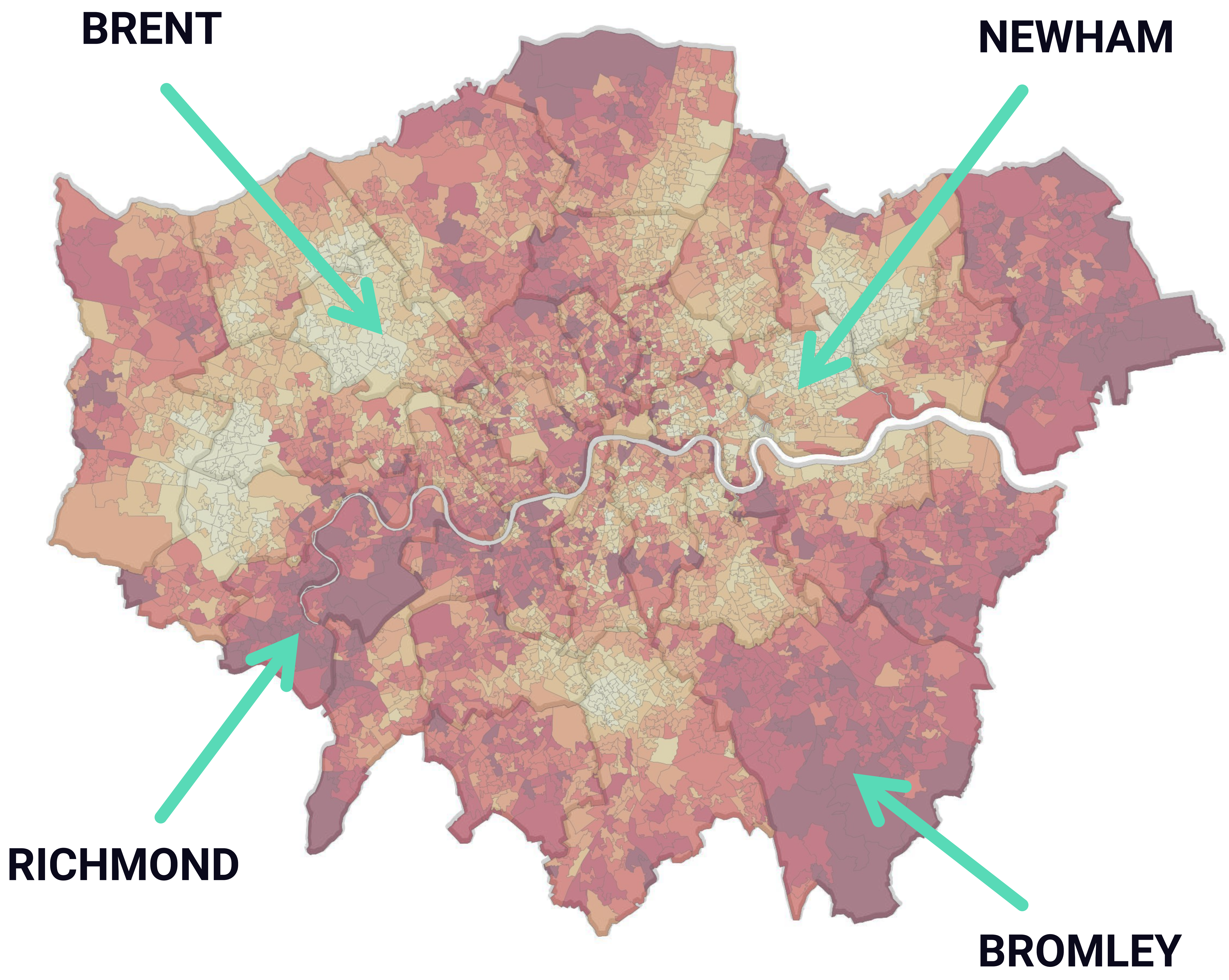
darker red means higher concentration



Spatial patterns matter

Percentage of white people in London (LSOA, census 2011)

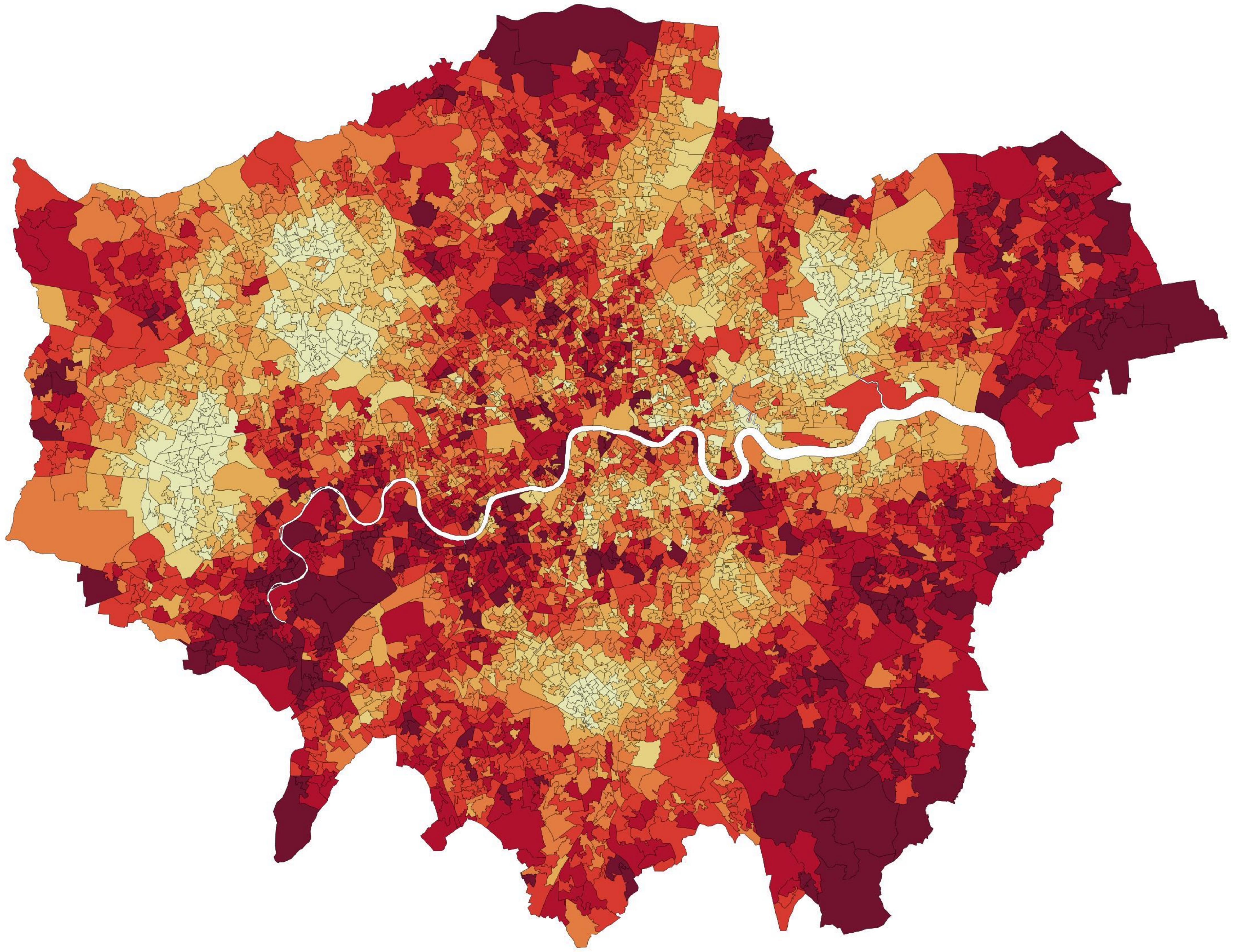
darker red means higher concentration

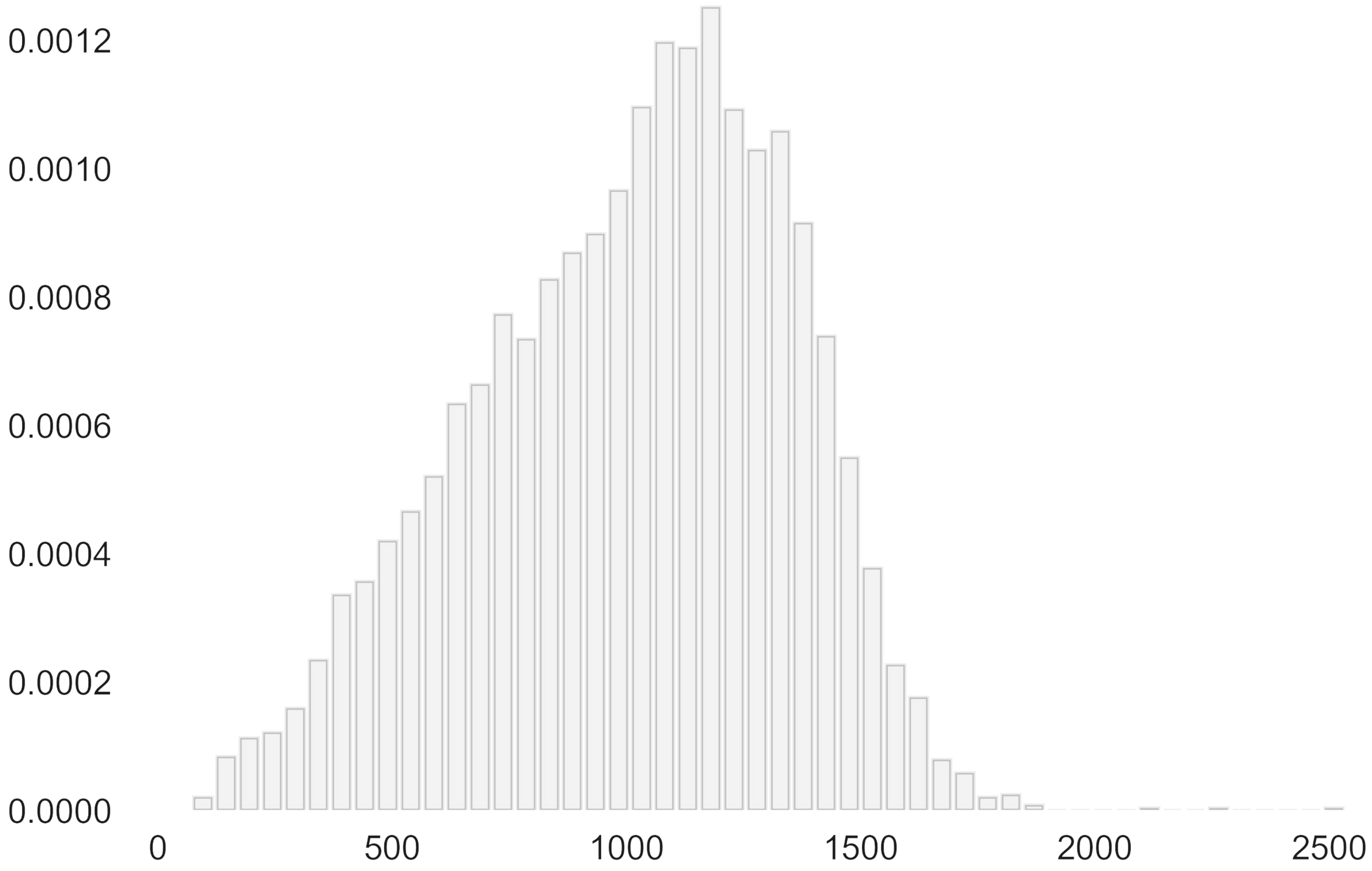


Spatial patterns matter

Percentage of white people in London (LSOA, census 2011)

darker red means higher concentration

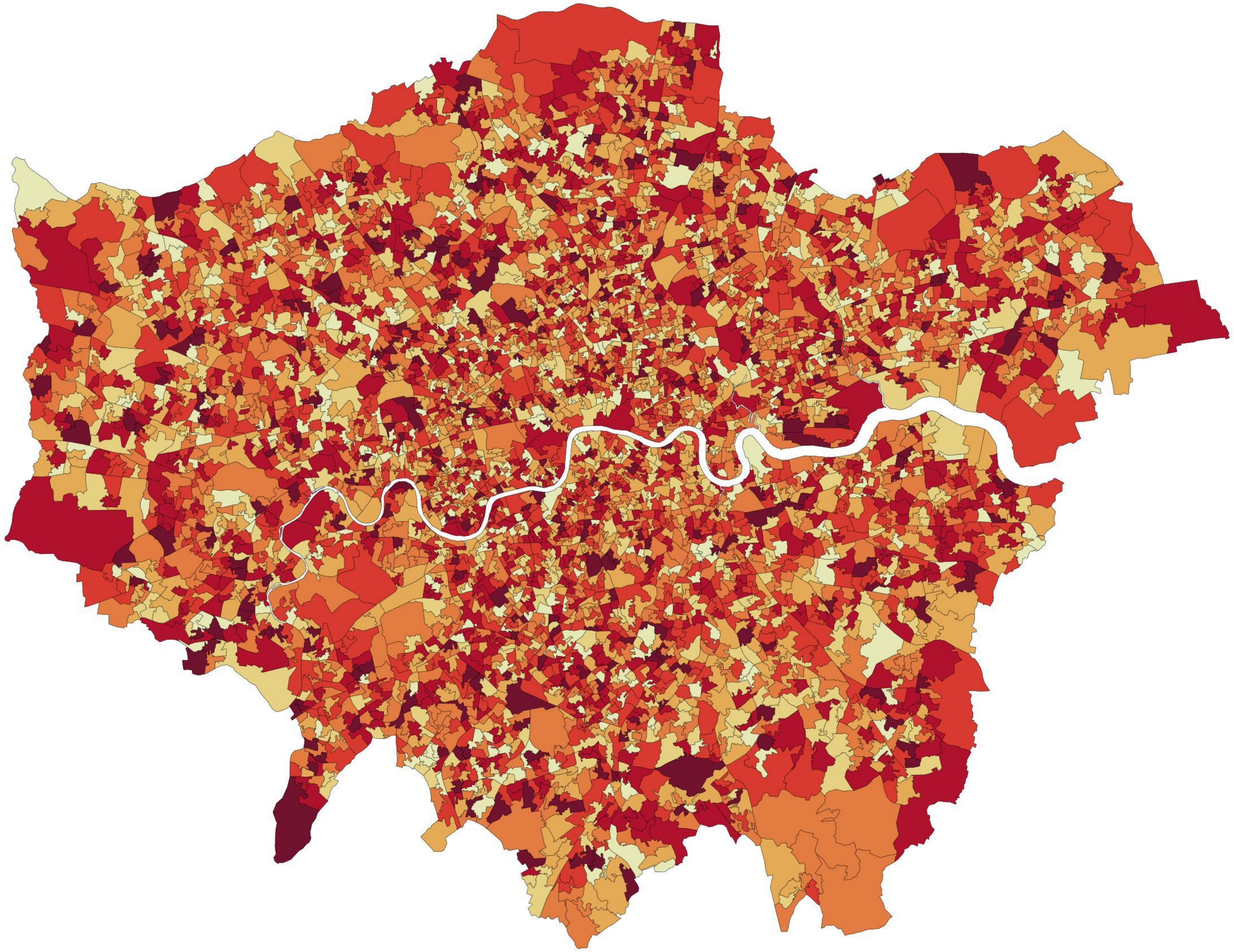




Spatial patterns matter

Percentage of white people in London (LSOA, census 2011)

darker red means higher concentration

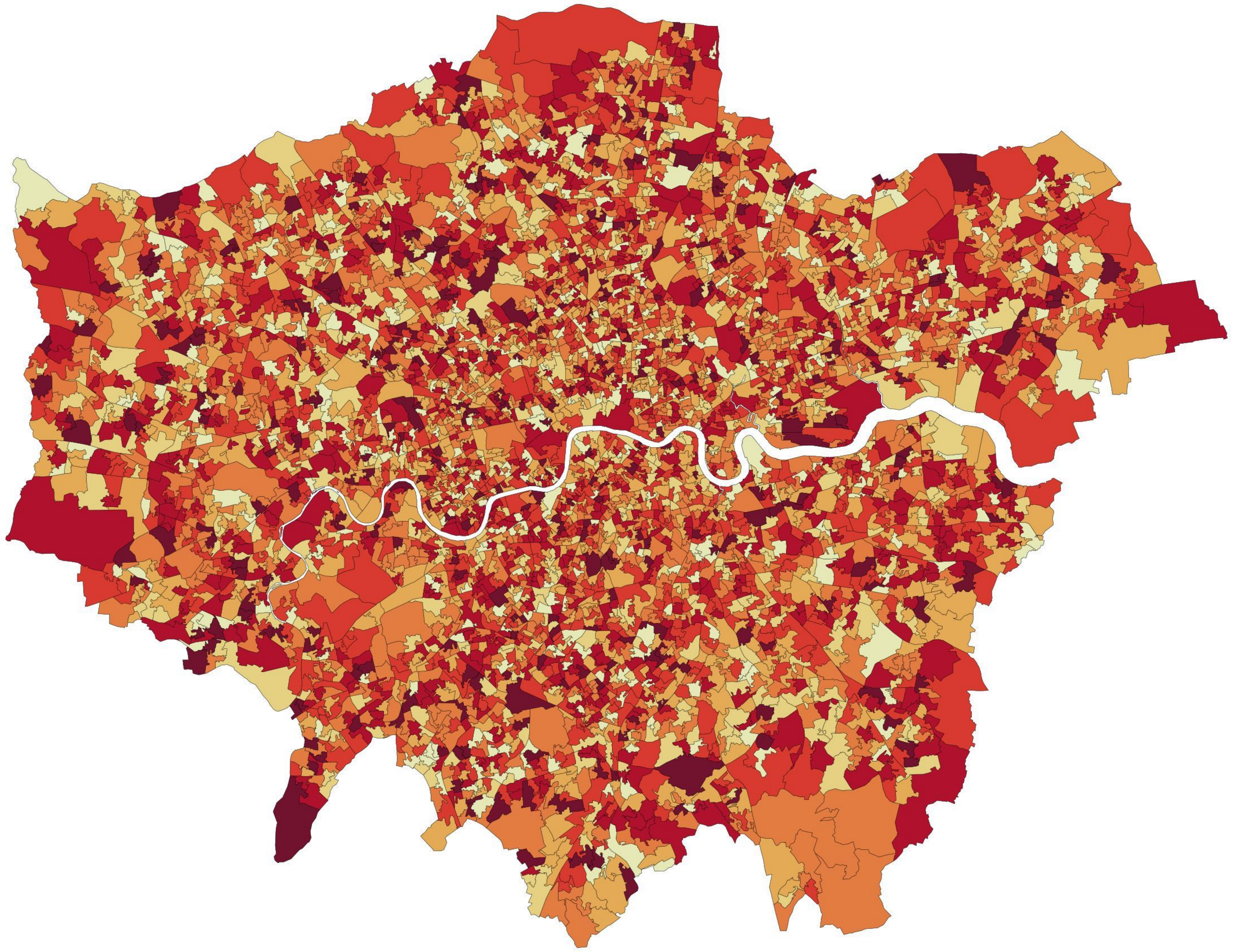


RANDOMLY RESHUFFLED

Spatial patterns
matter

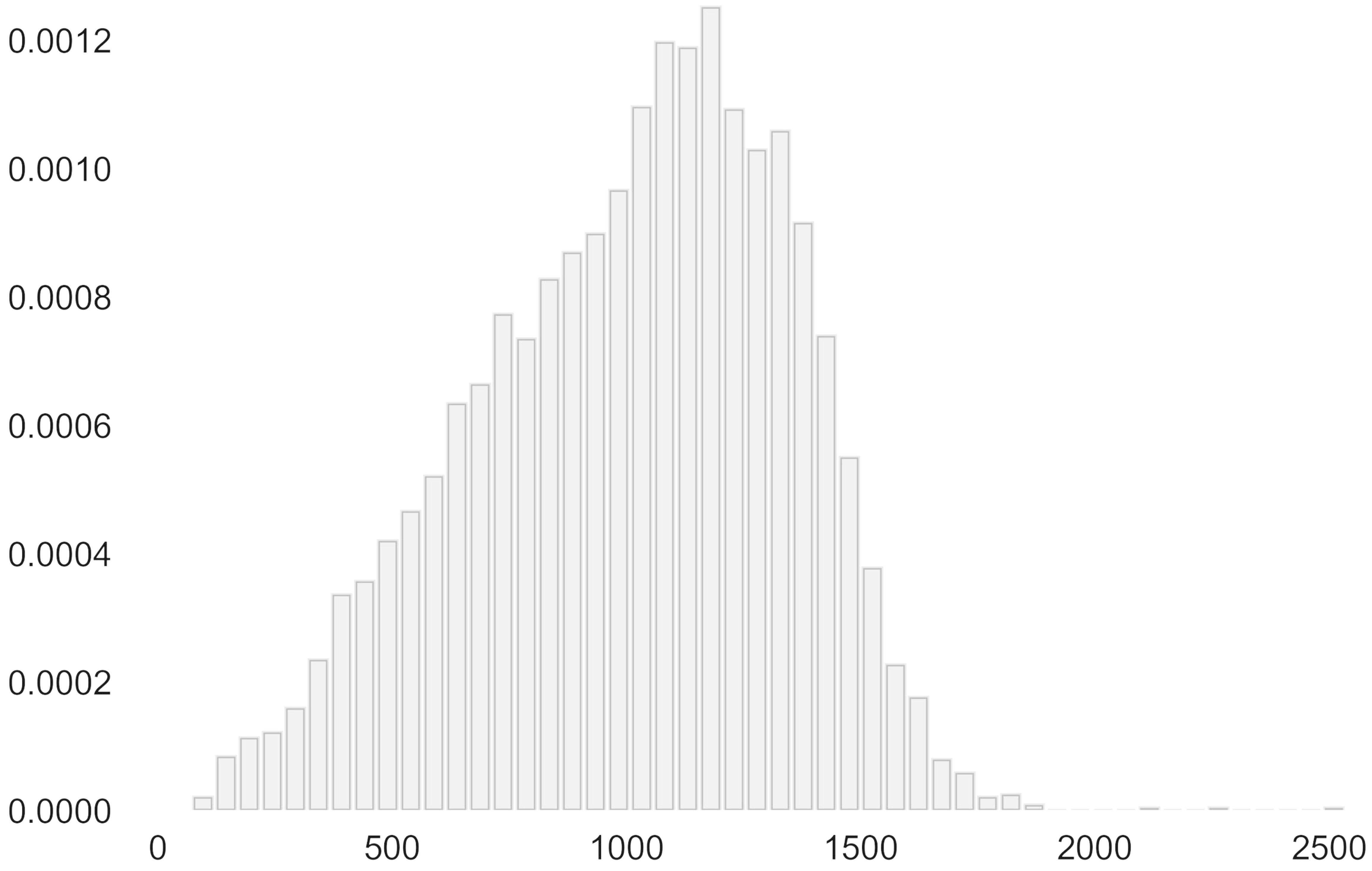
Percentage of white people in London (LSOA, census 2011)

darker red means higher concentration

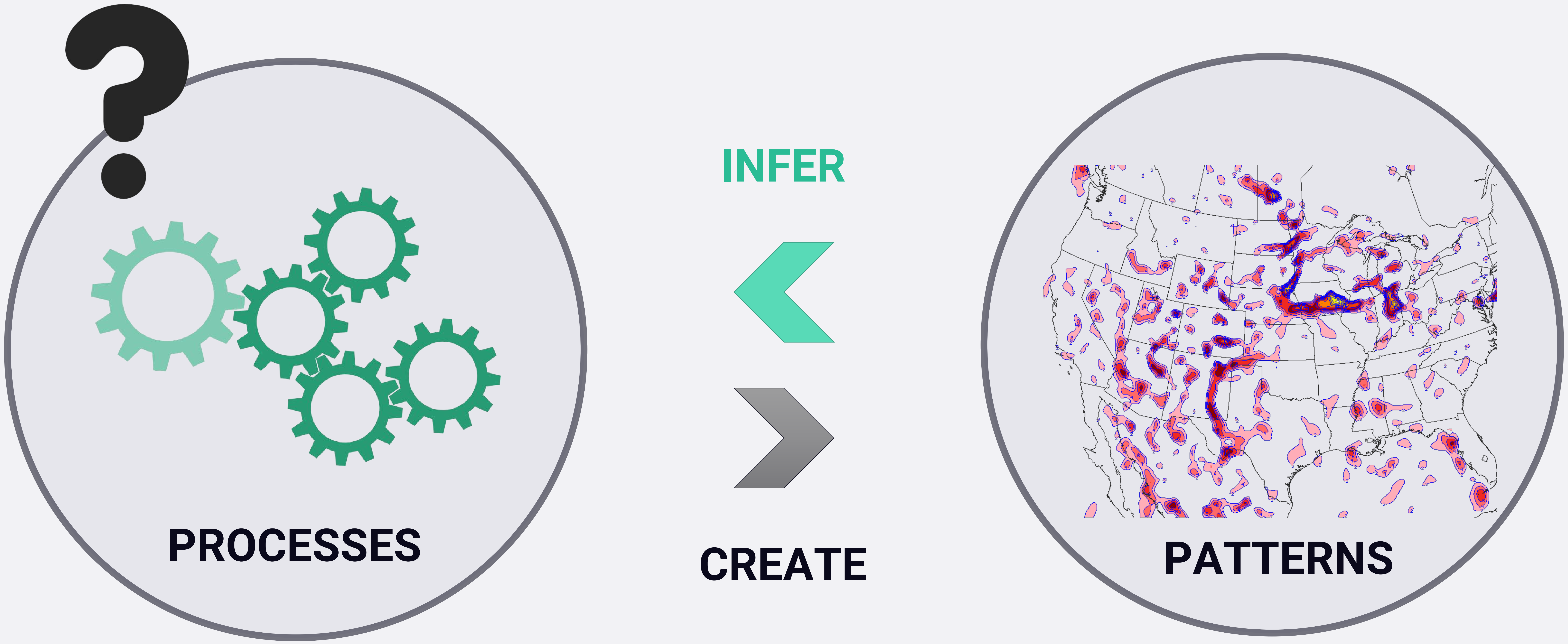


RANDOMLY RESHUFFLED

Same distribution



- **Processes operating in space create patterns**
- **Spatial Data Analysis is aimed at:**
  - Identifying and describing the **patterns**
  - Identifying and understanding the **processes**



PROCESSES

INFER

CREATE

PATTERNS

Often not directly observable

Often a sample of the data

spatial data analysis:
successive levels of sophistication

- **Spatial Data Description**
  - Focus is on describing the spatial data and representing it in a digital format (maps, databases)
  - Uses classic GIS capabilities (buffering, map layer overlay, spatial queries, measurement)
- **Exploratory Spatial Data Analysis (ESDA)**
  - Showing and discovering interesting patterns
- **Spatial statistical analysis and hypothesis testing**
  - An extension of traditional statistics into a spatial settings to determine whether or not data are typical or unexpected
- **Spatial modeling**
  - Explaining interesting patterns
  - Optimization, simulation, prediction
  - Involves constructing models to predict spatial outcomes

cartography is

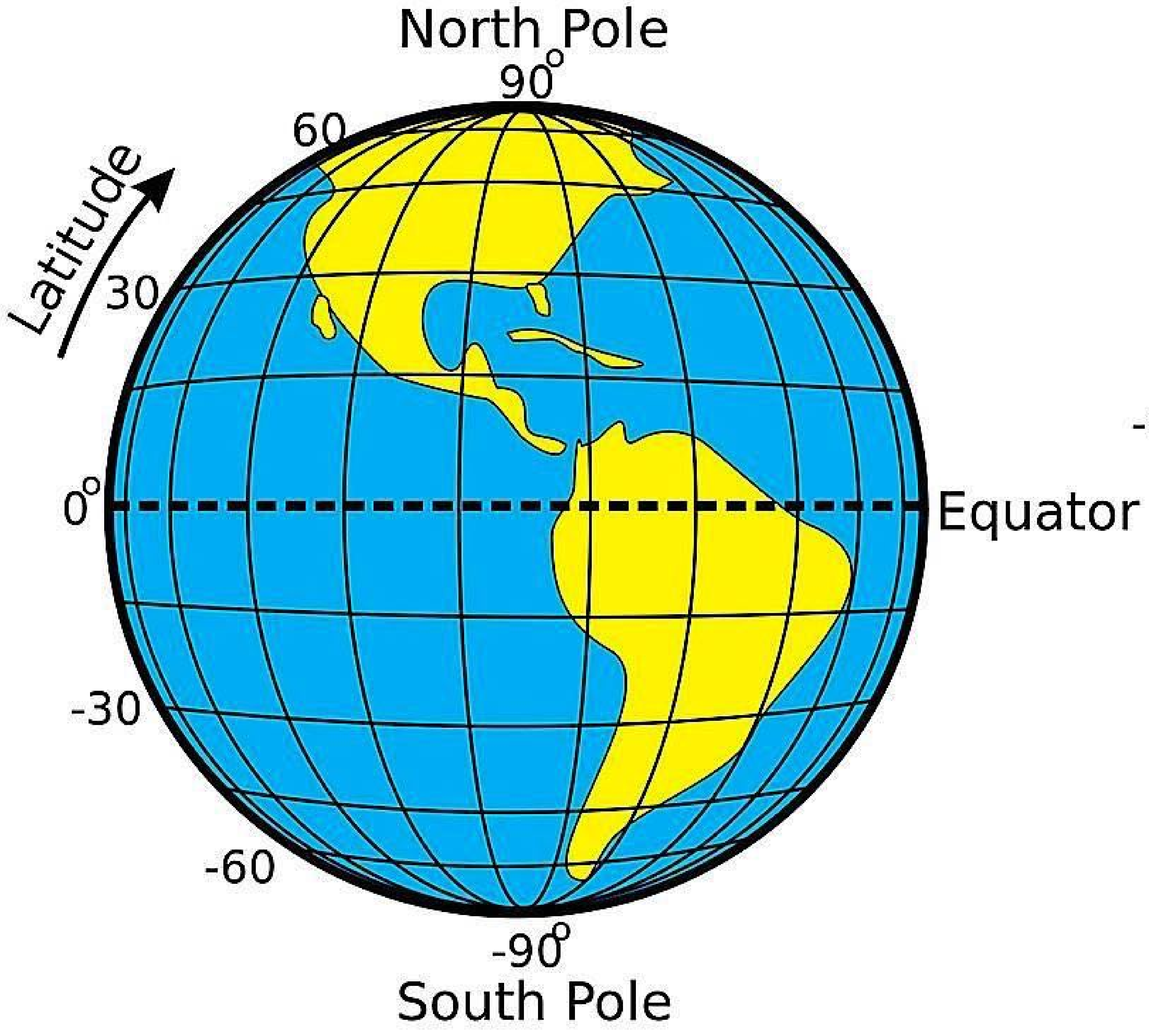
the study and practice of making maps

Everyone uses maps

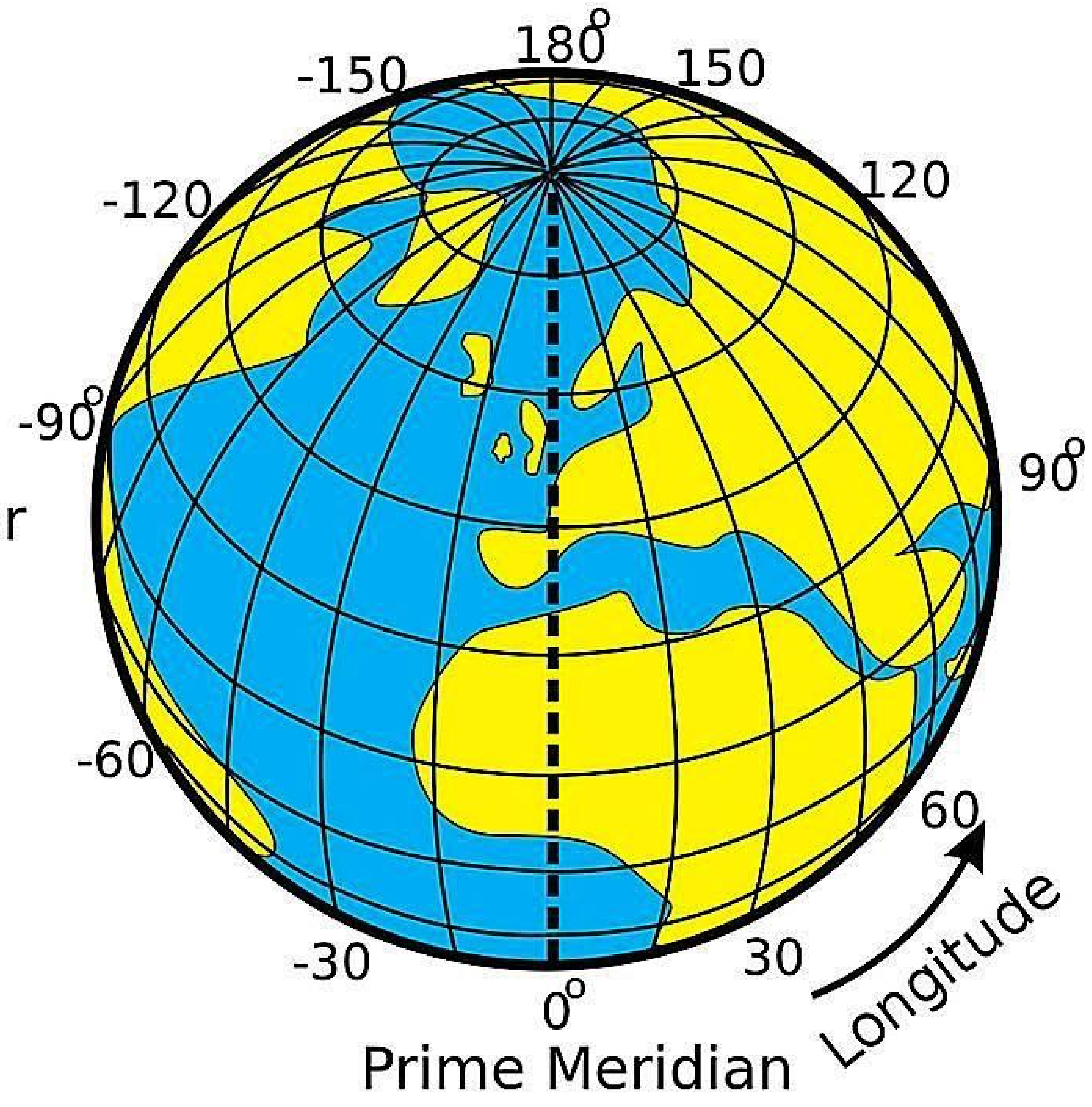
https://www.youtube.com/watch?v=AQ45fO1uiOQ

georeferenced data:
coordinates

- **(longitude, latitude) can be associated with**
  - altitude, accuracy, timestamp
- **can be reversed geocoded to a readable address**



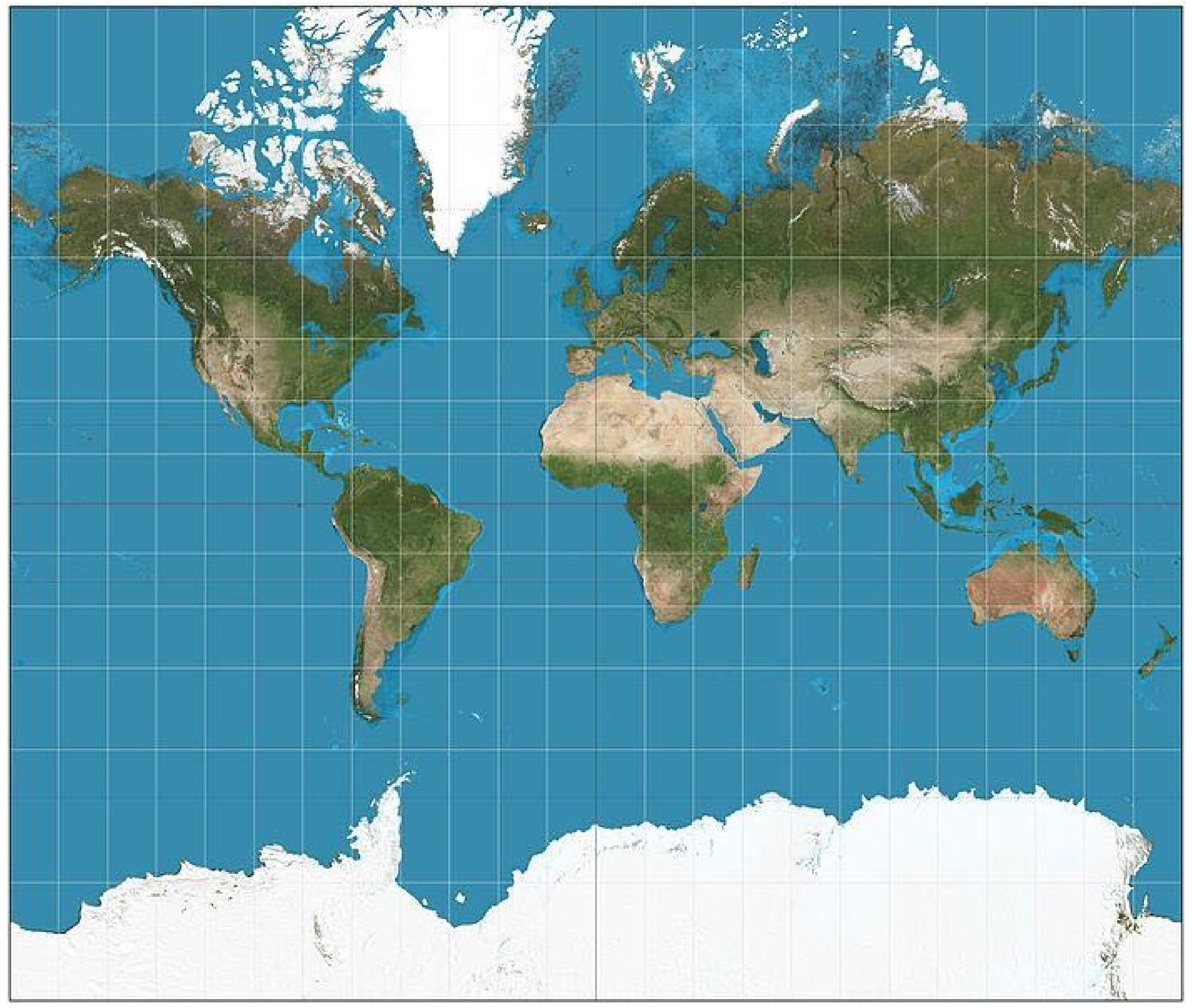
VERTICAL



HORINZONTAL

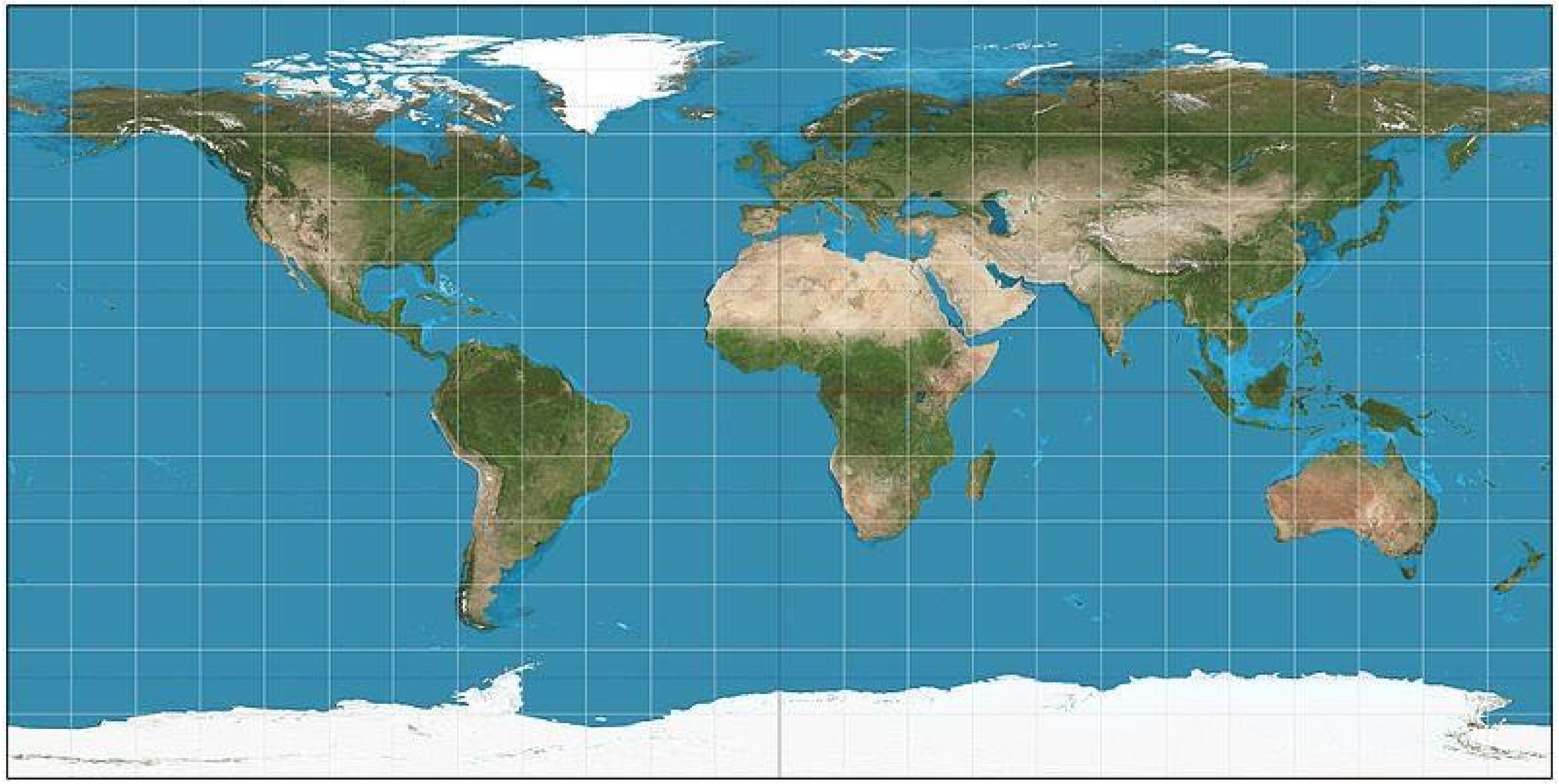
Earth is 3D, maps no!
map projections

- a **projection** is used to transform the geographic coordinates from the curved surface of our planet to the flat surface of a plane

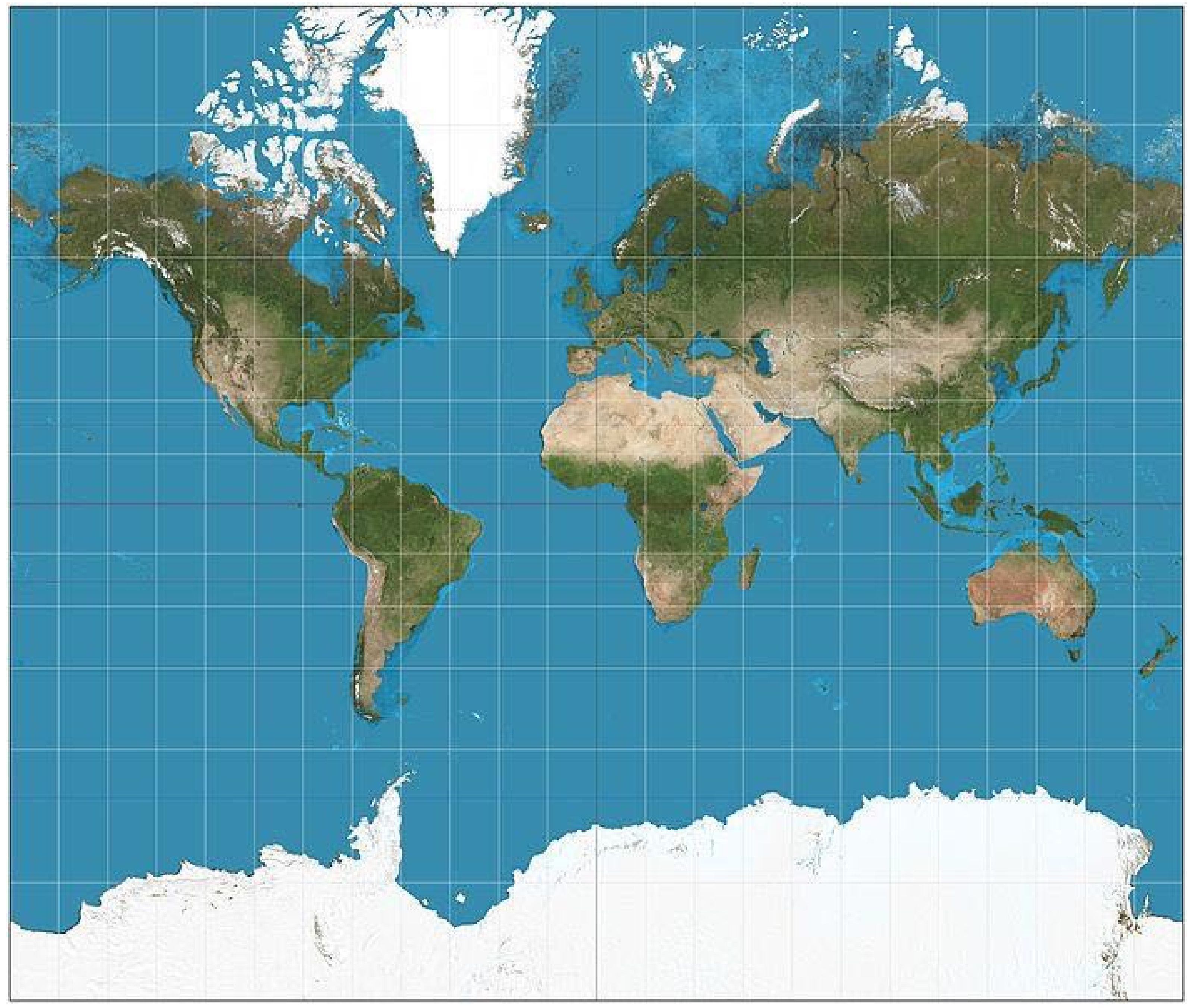


map projections

- over the years a variety of map projections have been proposed



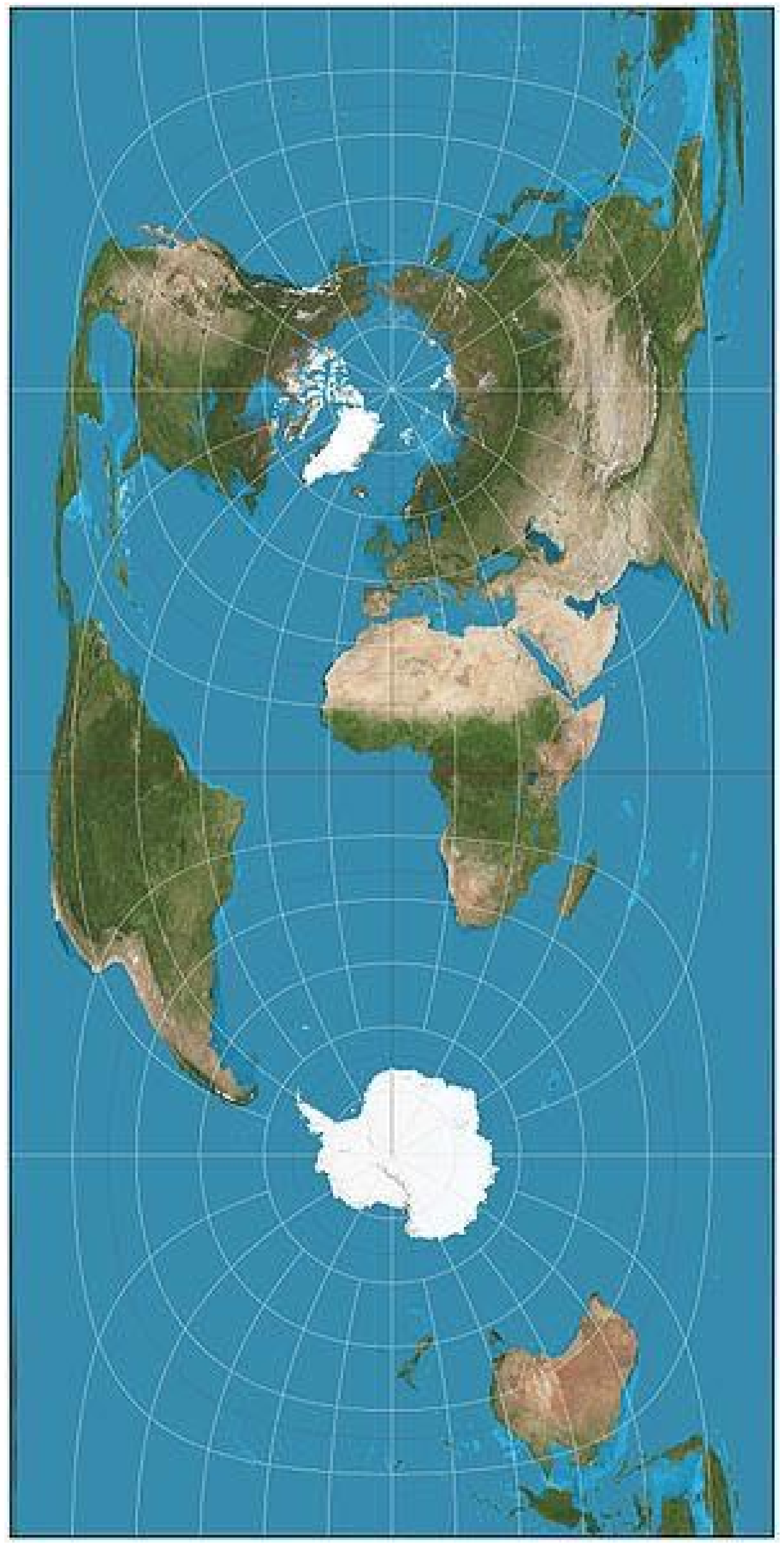
Equirectangular



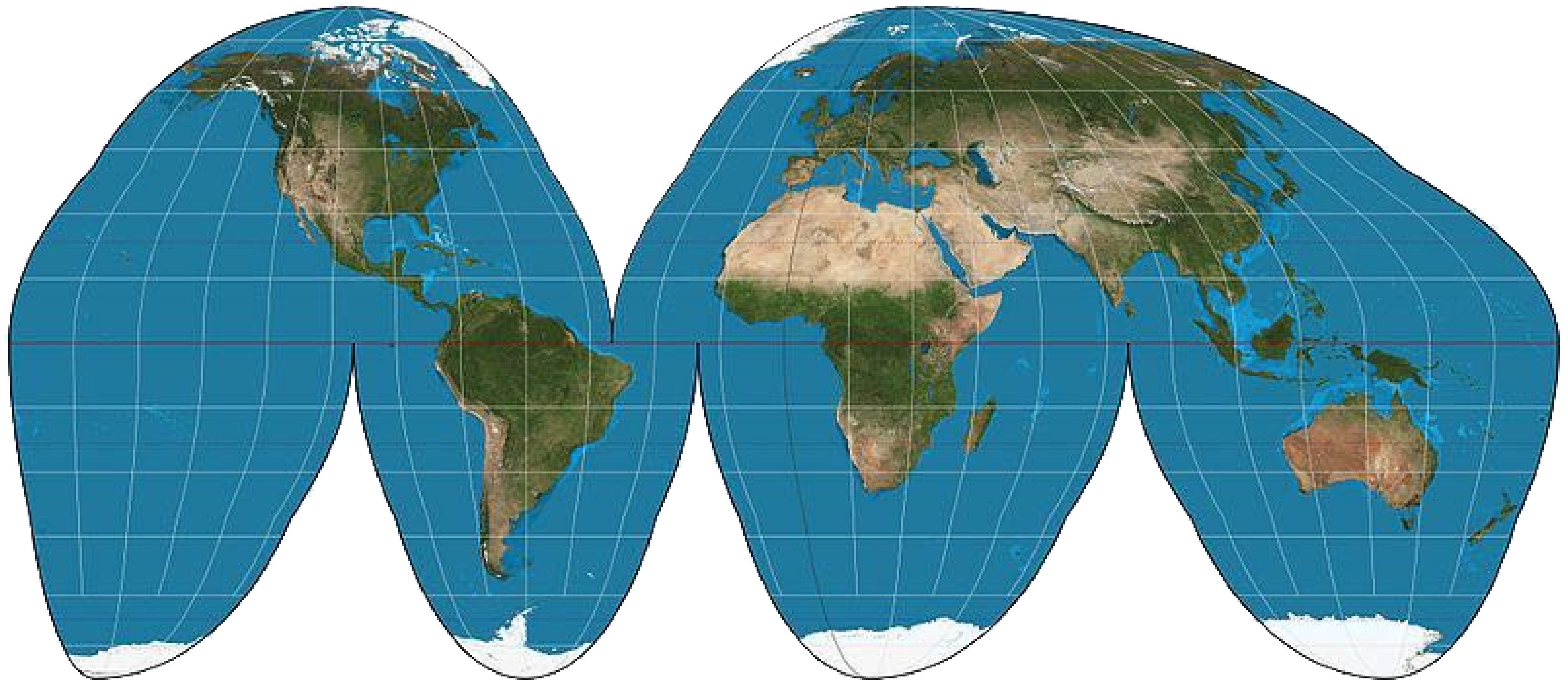
Mercator

All map images by Strebe, CC BY-SA 3.0

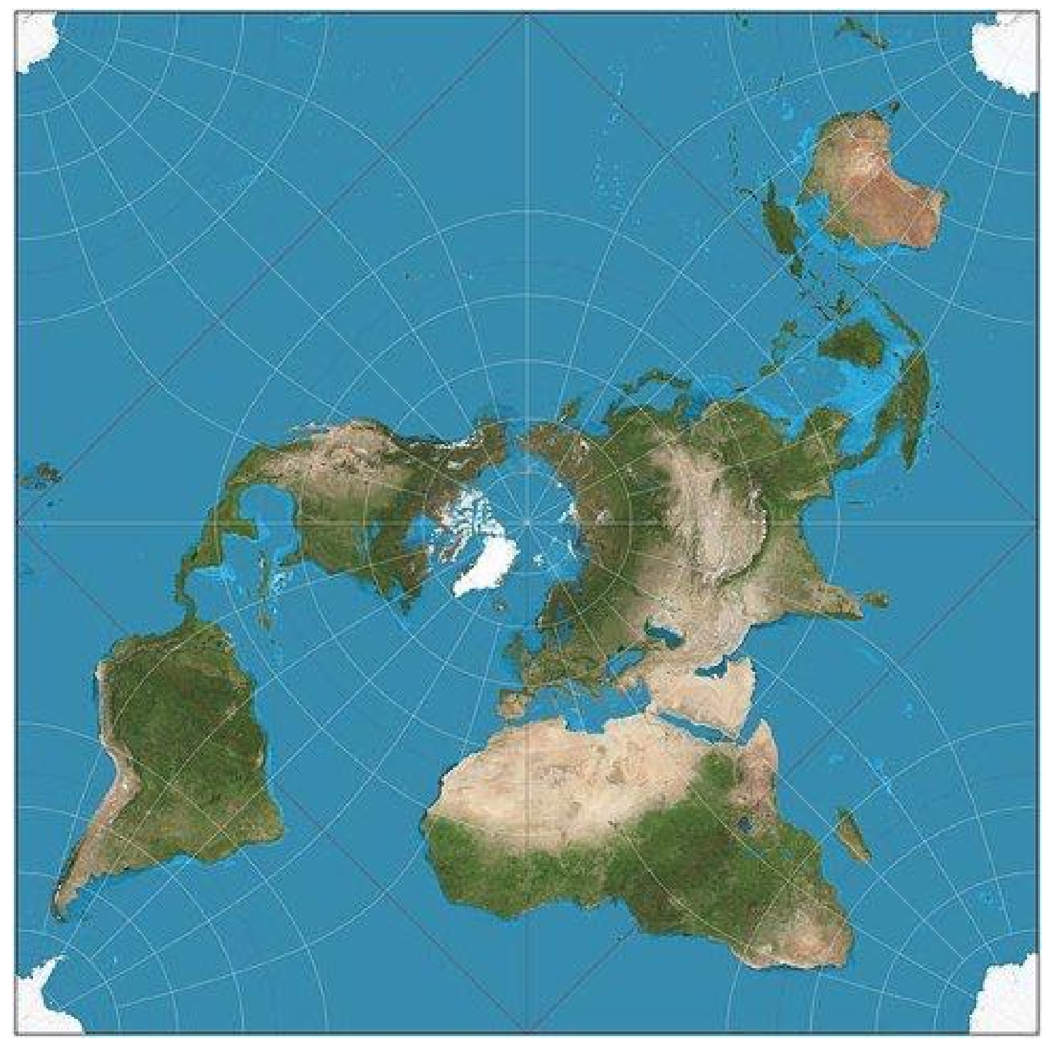
map projections



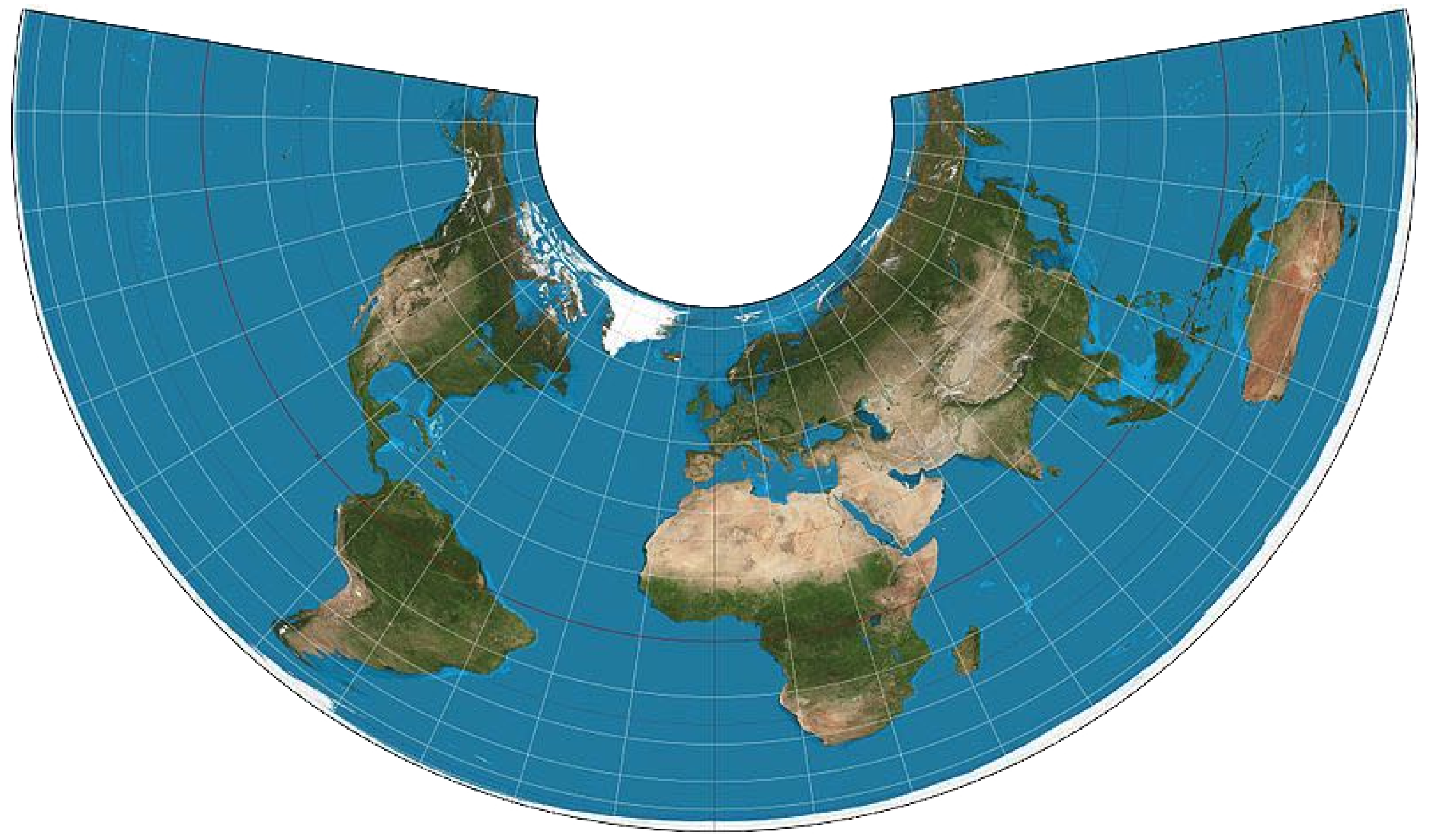
Cassini



Goode homolosine

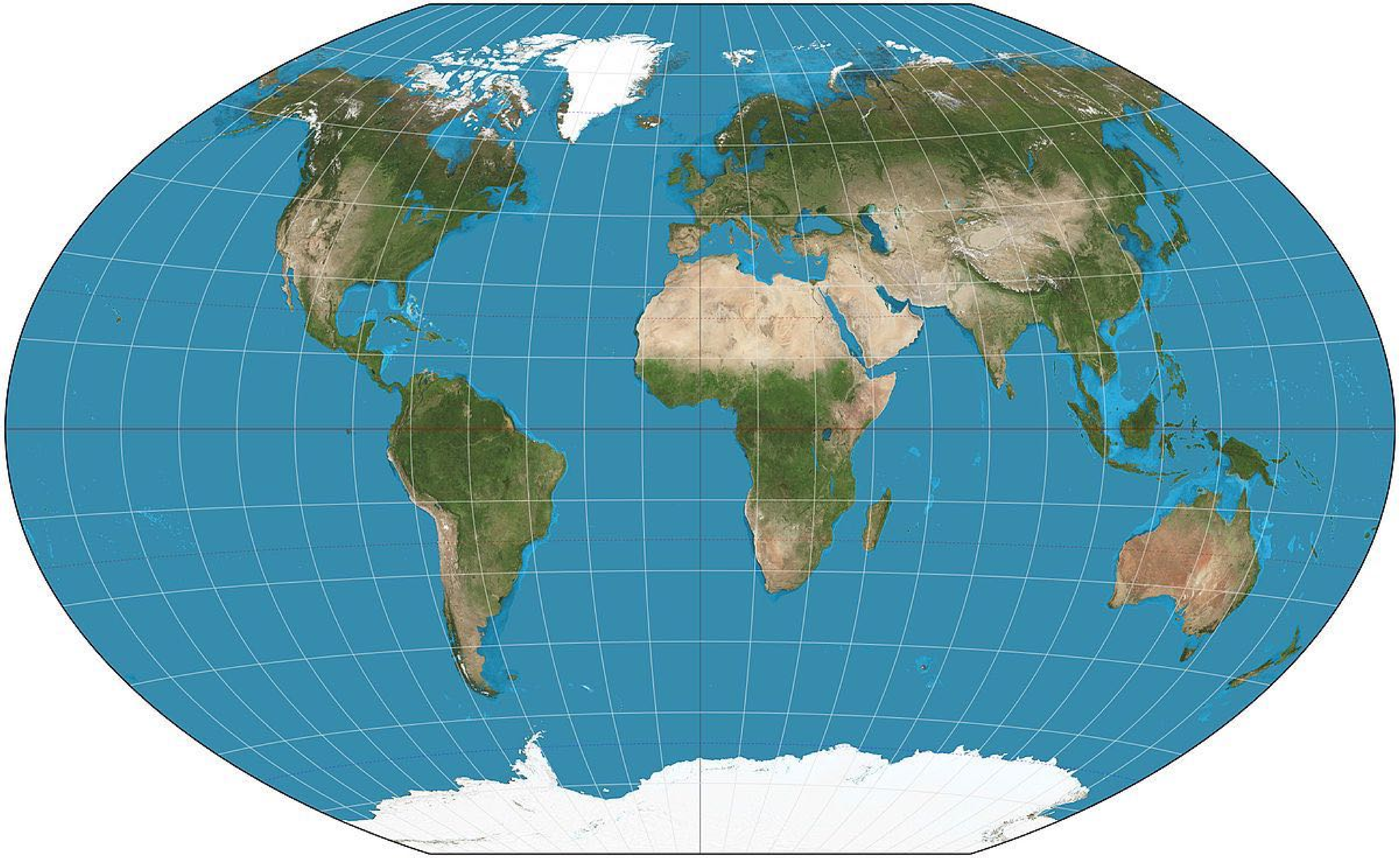


Peirce quincuncial



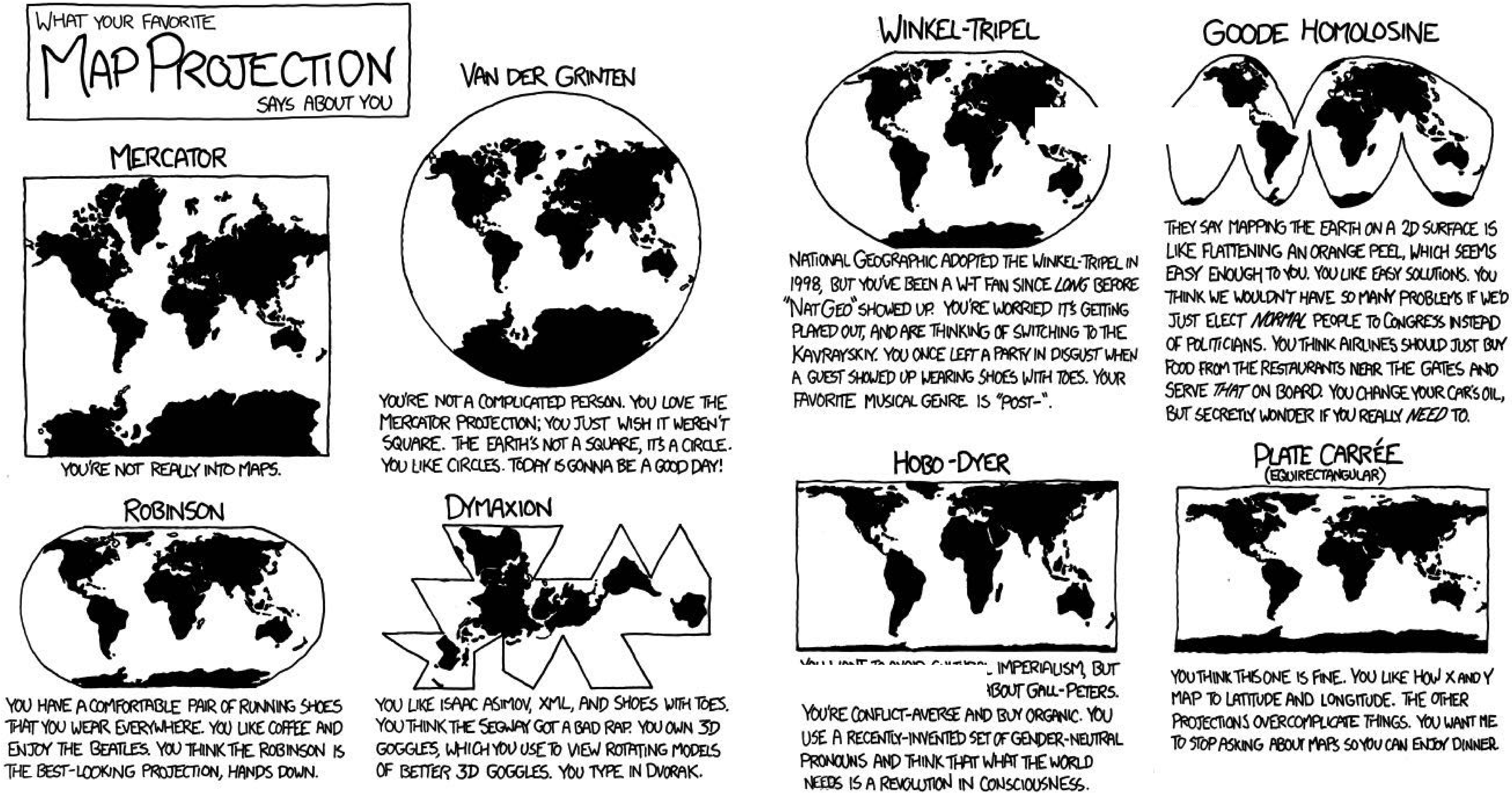
Albers conic

map projections



Winkel Triple

adopted by National Geographic



WHAT YOUR FAVORITE MAP PROJECTION SAYS ABOUT YOU

MERCATOR

YOU'RE NOT REALLY INTO MAPS.

VAN DER GRINTEN

YOU'RE NOT A COMPLICATED PERSON. YOU LOVE THE MERCATOR PROJECTION; YOU JUST WISH IT WEREN'T SQUARE. THE EARTH'S NOT A SQUARE, IT'S A CIRCLE. YOU LIKE CIRCLES. TODAY IS GONNA BE A GOOD DAY!

ROBINSON

YOU HAVE A COMFORTABLE PAIR OF RUNNING SHOES THAT YOU WEAR EVERYWHERE. YOU LIKE COFFEE AND ENJOY THE BEATLES. YOU THINK THE ROBINSON IS THE BEST-LOOKING PROJECTION, HANDS DOWN.

DYMAXION

YOU LIKE ISAAC ASIMOV, XML, AND SHOES WITH TOES. YOU THINK THE SEGWAY GOT A BAD RAP. YOU OWN 3D GOGGLES, WHICH YOU USE TO VIEW ROTATING MODELS OF BETTER 3D GOGGLES. YOU TYPE IN DVORAK.

WINKEL-TRIPEL

NATIONAL GEOGRAPHIC ADOPTED THE WINKEL-TRIPEL IN 1998, BUT YOU'VE BEEN A W-T FAN SINCE *LONG* BEFORE "NAT GEO" SHOWED UP. YOU'RE WORRIED IT'S GETTING PLAYED OUT, AND ARE THINKING OF SWITCHING TO THE KAVRAYSKIY. YOU ONCE LEFT A PARTY IN DISGUST WHEN A GUEST SHOWED UP WEARING SHOES WITH TOES. YOUR FAVORITE MUSICAL GENRE IS "POST-".

GOODE HOMOLOSINE

THEY SAY MAPPING THE EARTH ON A 2D SURFACE IS LIKE FLATTENING AN ORANGE PEEL, WHICH SEEMS EASY ENOUGH TO YOU. YOU LIKE EASY SOLUTIONS. YOU THINK WE WOULDN'T HAVE SO MANY PROBLEMS IF WE'D JUST ELECT *NORMAL* PEOPLE TO CONGRESS INSTEAD OF POLITICIANS. YOU THINK AIRLINES SHOULD JUST BUY FOOD FROM THE RESTAURANTS NEAR THE GATES AND SERVE *THAT* ON BOARD. YOU CHANGE YOUR CAR'S OIL, BUT SECRETLY WONDER IF YOU REALLY *NEED* TO.

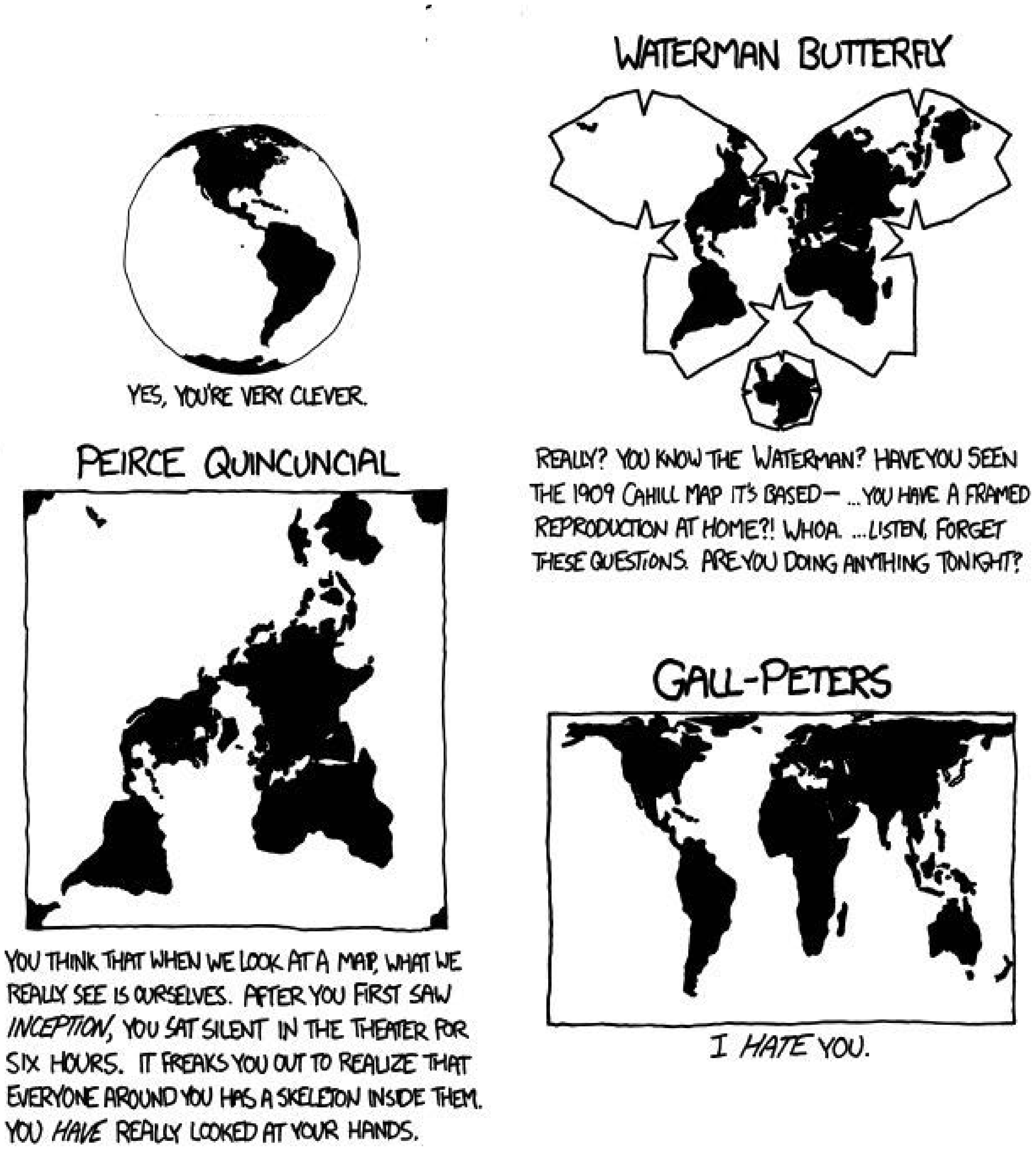
HOBO-DYER

[illegible] IMPERIALISM, BUT [illegible] ABOUT GALL-PETERS. YOU'RE CONFLICT-AVERSE AND BUY ORGANIC. YOU USE A RECENTLY-INVENTED SET OF GENDER-NEUTRAL PRONOUNS AND THINK THAT WHAT THE WORLD NEEDS IS A REVOLUTION IN CONSCIOUSNESS.

PLATE CARRÉE (EQUIRECTANGULAR)

YOU THINK THIS ONE IS FINE. YOU LIKE HOW X AND Y MAP TO LATITUDE AND LONGITUDE. THE OTHER PROJECTIONS OVERCOMPLICATE THINGS. YOU WANT ME TO STOP ASKING ABOUT MAPS SO YOU CAN ENJOY DINNER.

from xkcd

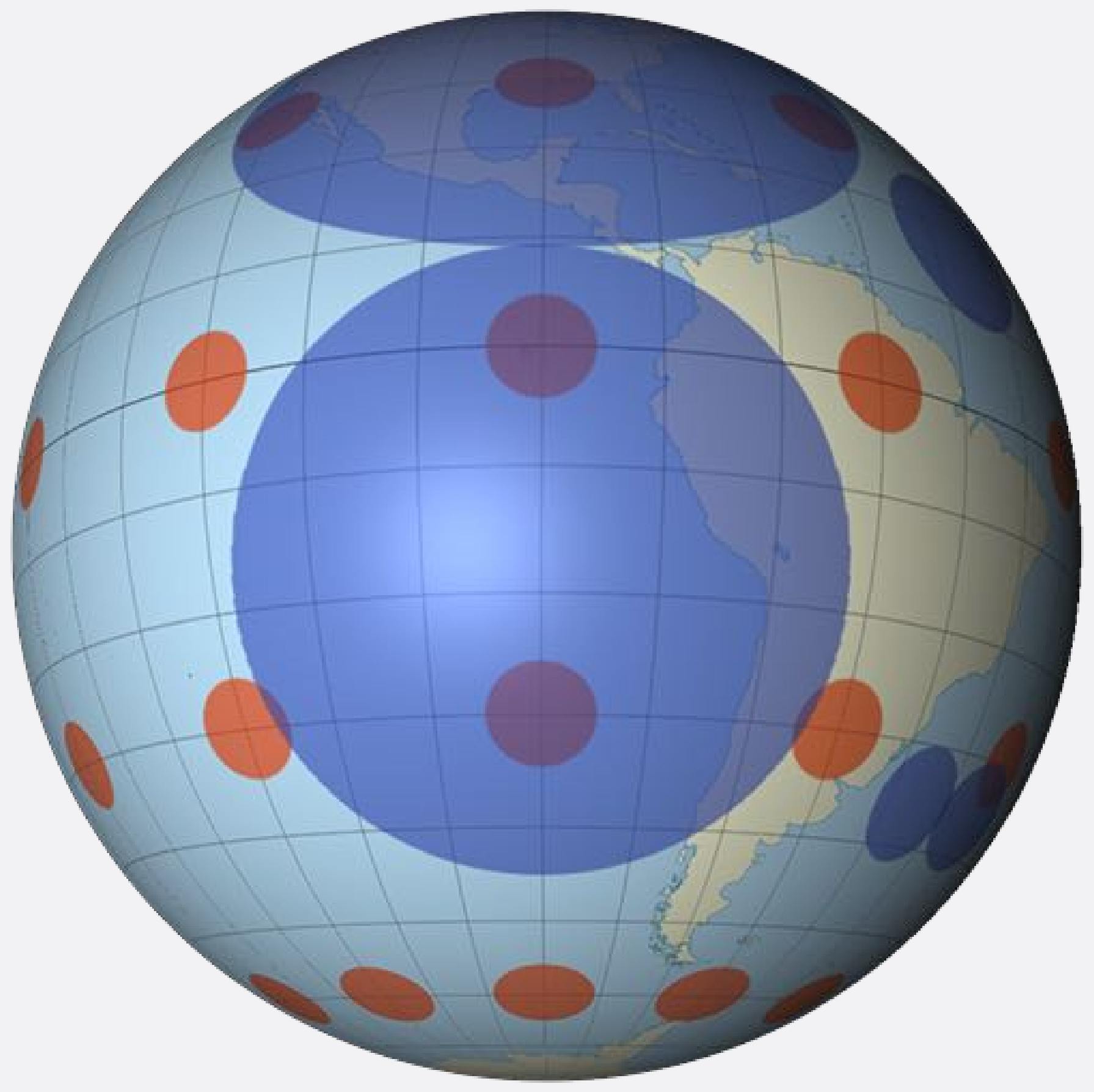

WATERMAN BUTTERFLY
YES, YOU'RE VERY CLEVER.
REALLY? YOU KNOW THE WATERMAN? HAVE YOU SEEN THE 1909 CAHILL MAP IT'S BASED— ...YOU HAVE A FRAMED REPRODUCTION AT HOME?! WHOA. ...LISTEN, FORGET THESE QUESTIONS. ARE YOU DOING ANYTHING TONIGHT?
PEIRCE QUINCUNCIAL
YOU THINK THAT WHEN WE LOOK AT A MAP, WHAT WE REALLY SEE IS OURSELVES. AFTER YOU FIRST SAW INCEPTION, YOU SAT SILENT IN THE THEATER FOR SIX HOURS. IT FREAKS YOU OUT TO REALIZE THAT EVERYONE AROUND YOU HAS A SKELETON INSIDE THEM. YOU HAVE REALLY LOOKED AT YOUR HANDS.
GALL-PETERS
I HATE YOU.

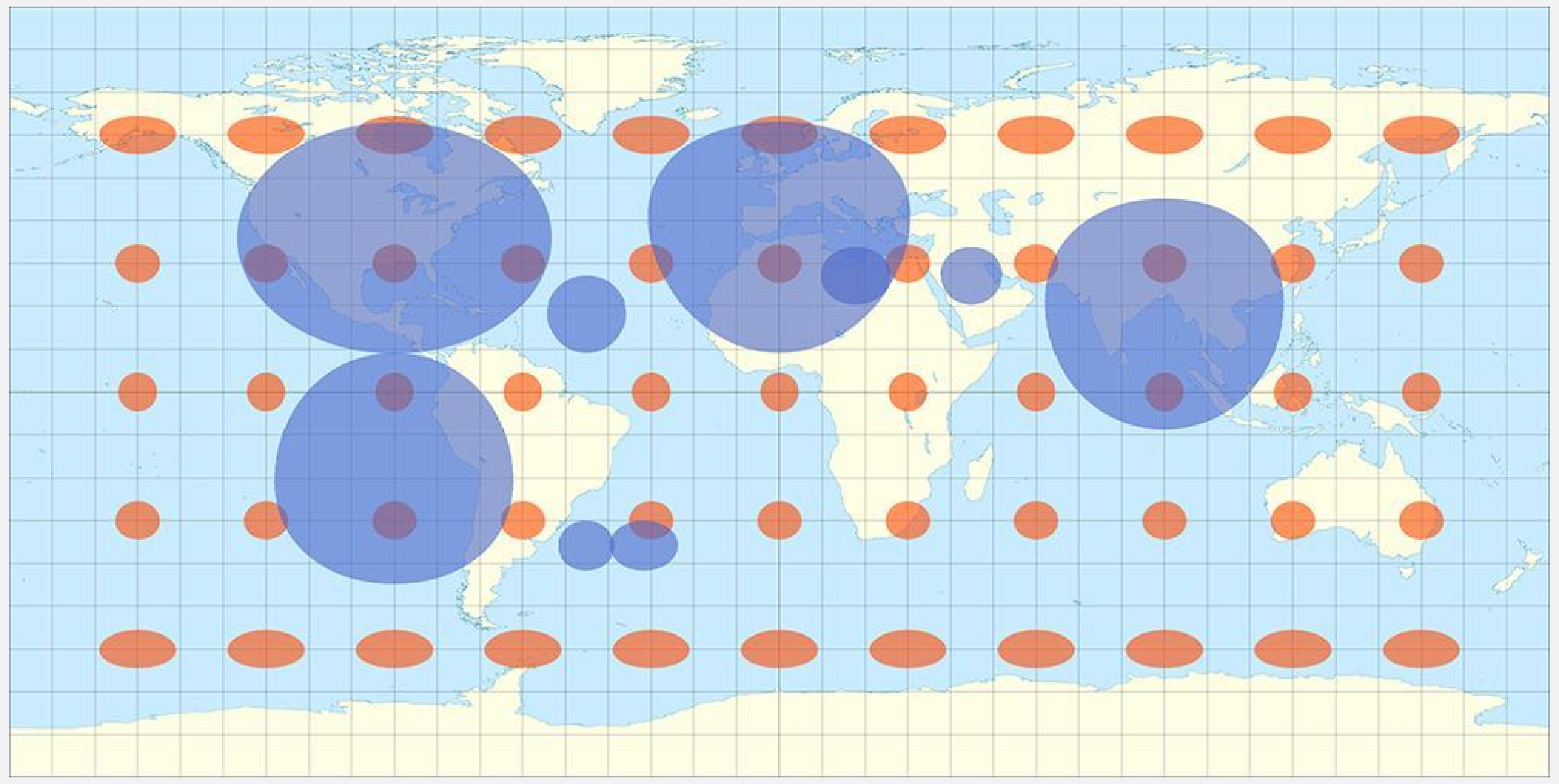
Which one should I use?

There is not a perfect projection!

map projections
distortions

- **projections cause distortions**
  - **shape, area, distance, direction**
  - depending on the application, some projections may be more suitable than others

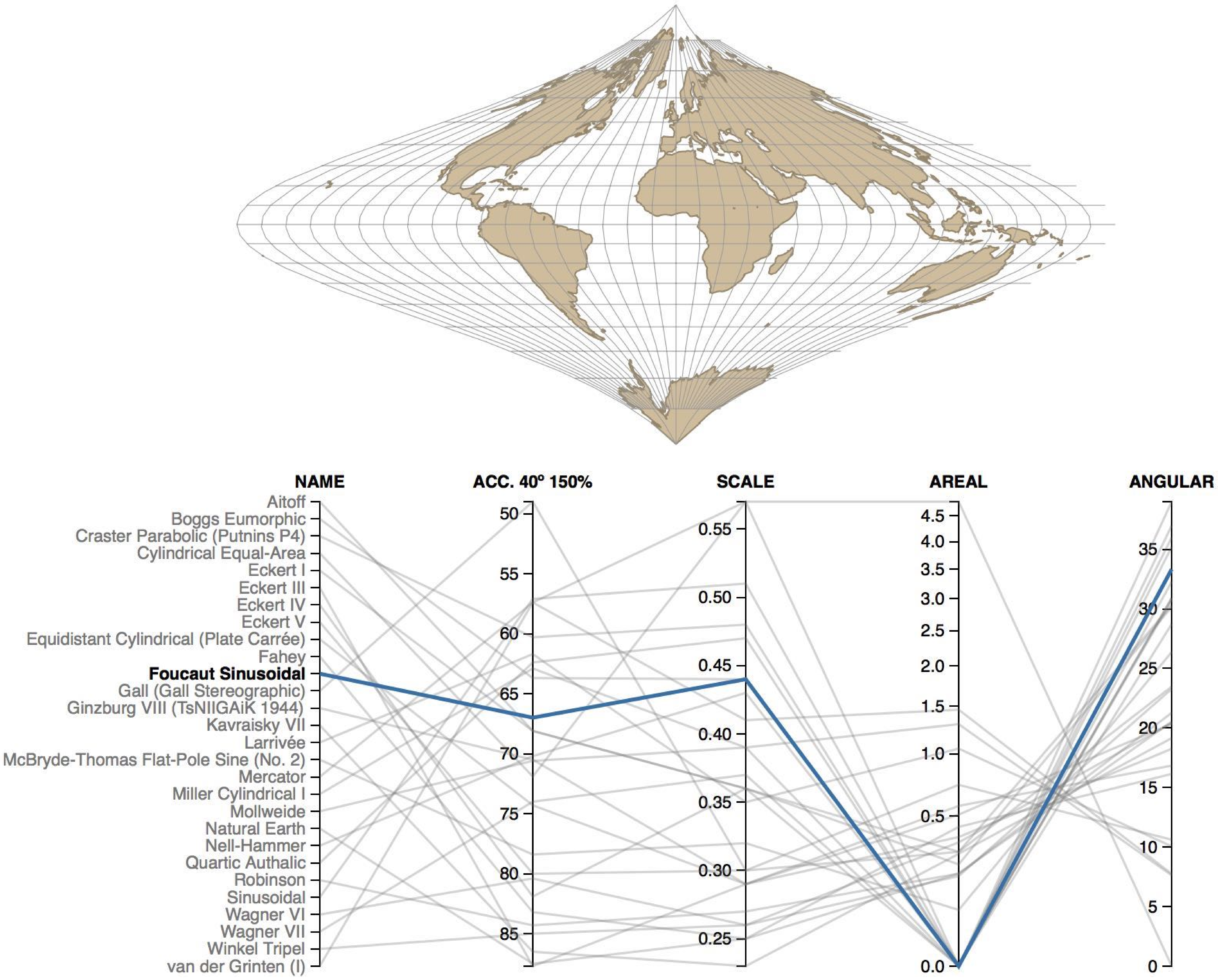




types of projections

- **azimuthal**
  - preserves the azimuth (direction) from center
- **conformal**
  - local angles are correct, preserving small shapes
- **equal-Area**
  - equal-area maps preserve area measure, generally distorting shapes
- **equidistant**
  - distances from center (or along certain lines, like along meridians) are correct

compare projections on d3.js



projections can produce societal biases

- **See video from "The West Wing" Season 2 Episode 16**
  - https://www.youtube.com/watch?v=vVX-PrBRtTY&t
- Other useful references:
  - https://www.youtube.com/watch?v=KUF_Ckv8HbE
  - https://www.youtube.com/watch?v=kIID5FDi2JQ

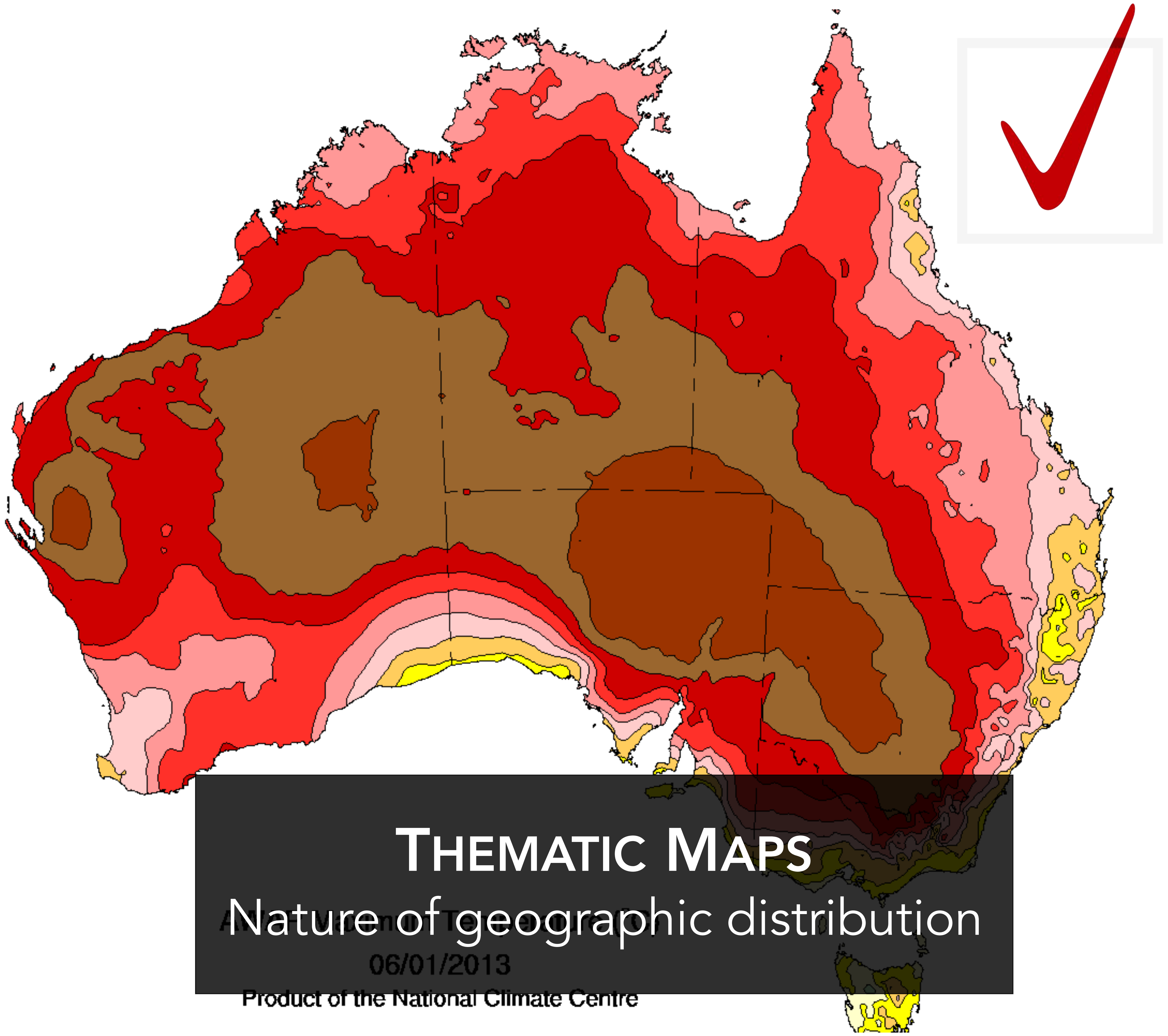
mapping

two (overlapping) categories

Angler Reef (submerged)
submerged rocks
Lee Point
Dariba Rock
CASUARINA COASTAL RESERVE
tower (67m)
submerged
Hope Inlet
numerous masts
sewage plant
radio masts
salt evaporator
SHOAL BAY
Casuarina
numerous masts
pools
Korebum Lagn
Nightcliff
Leanyer Swamp
COASTAL RESERVE
pool
Koolpinyah
SHOAL BAY PENINSULA
Noogoo Swamp
golf course
sports complex
Micket Creek Shooting Complex
rifle range
East Point
NTS 221
NTS 220
DARWIN INTERNATIONAL AIRPORT
towers (75m)
Milners Swamp
Howard River
HOWARD SPRINGS HUNTING RESERVE
Fisher Lagoon
Lyons Lagn
SHOAL BAY COASTAL RES
POINT
ROAD
Fannie Bay
NTS 216
racecourse
Winnellie
oil tks
masts
quarries
DARWIN
Knuckey Lagoons
numerous sandscrapes
GUNN
Berrimah
breakwater
Emery Point
NTS 217
tower (61m)
CHARLES DARWIN NP
golf course
jetty
racetrack
pools
Howard Springs
water tower
numerous
pool
sandscrapes
Larrakeyah
fuel tanks
substation
breakwater
Frances Bay
quarry
masts
Darwin
East Arm
towers (62m)
Plater Rock
Stevens Rock
wharves
quarry
pool
PORT
causeways
PALMERSTON
golf course
HOWARD SPRINGS
Weed Reef
wharf
Old Man Rk
North Shell Is
STUART
silos
East Arm
Wickham Point
quarries
numerous
DARWIN
Girraween Lagoon
Howard
aband
Benhams Lagoon
PEAK HILL
sewage plant
Coolalinga
pools
quarries
natural gas treatment plant
towers (82m)
mast (82m)
pool
River
extraction pit
extraction pits
golf course
Channel Island
tower (54m)
gas tk
causeways
BLACKMORE HILL
power station
KINGS TABLE
pool
McMINNS LAGOON
MIDDLE
ARM
prawn aquaculture farm
substation
tower (82m)
HAYCOCK HILL
crocodile farm
quarries
pool
scrapes
scrape
scrape
pool
Noonamah
quarry
tower
aquaculture

Reference Maps

Orientation and Navigation



06/01/2013

Product of the National Climate Centre

Thematic Maps

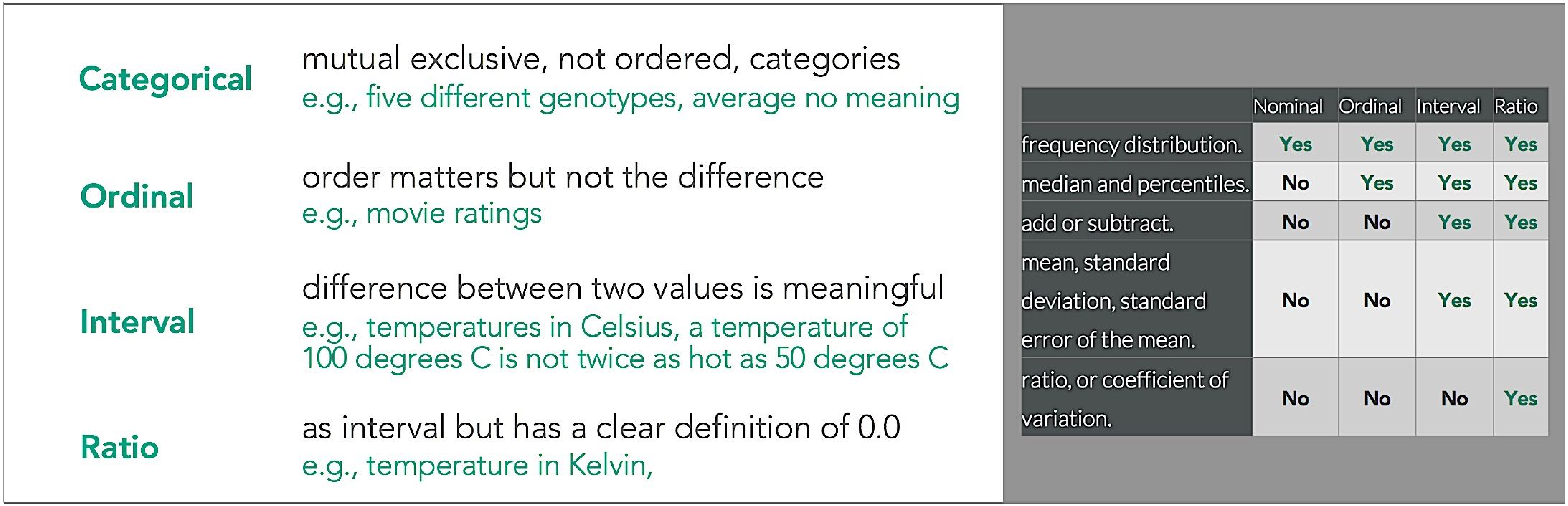
Nature of geographic distribution

Thematic maps

- **Visualize spatial distributions of data, e.g., population density**
- **Thematic maps serve three primary purposes.**
  - 1. They provide specific information about particular locations.
  - 2. They provide general information about spatial patterns.
  - 3. They can be used to compare patterns on two or more maps.

Design is driven by

- **Data**
  - Categorical, ordinal, interval, ratio



Categorical mutual exclusive, not ordered, categories
e.g., five different genotypes, average no meaning

Ordinal order matters but not the difference
e.g., movie ratings

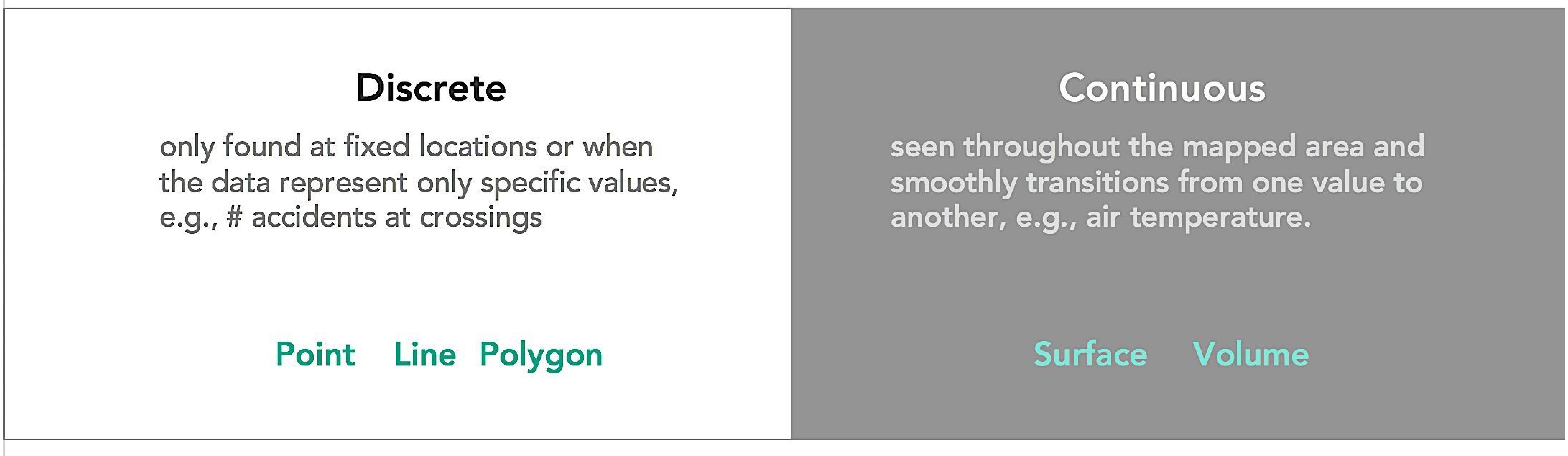
Interval difference between two values is meaningful
e.g., temperatures in Celsius, a temperature of 100 degrees C is not twice as hot as 50 degrees C

Ratio as interval but has a clear definition of 0.0
e.g., temperature in Kelvin,

	Nominal	Ordinal	Interval	Ratio
frequency distribution.	Yes	Yes	Yes	Yes
median and percentiles.	No	Yes	Yes	Yes
add or subtract.	No	No	Yes	Yes
mean, standard deviation, standard error of the mean.	No	No	Yes	Yes
ratio, or coefficient of variation.	No	No	No	Yes

Design is driven by

- **Data**
  - Categorical, ordinal, interval, ratio
- **Spatial scale and granularity**
  - discrete vs continuous



Discrete

only found at fixed locations or when the data represent only specific values, e.g., # accidents at crossings

Point Line Polygon

Continuous

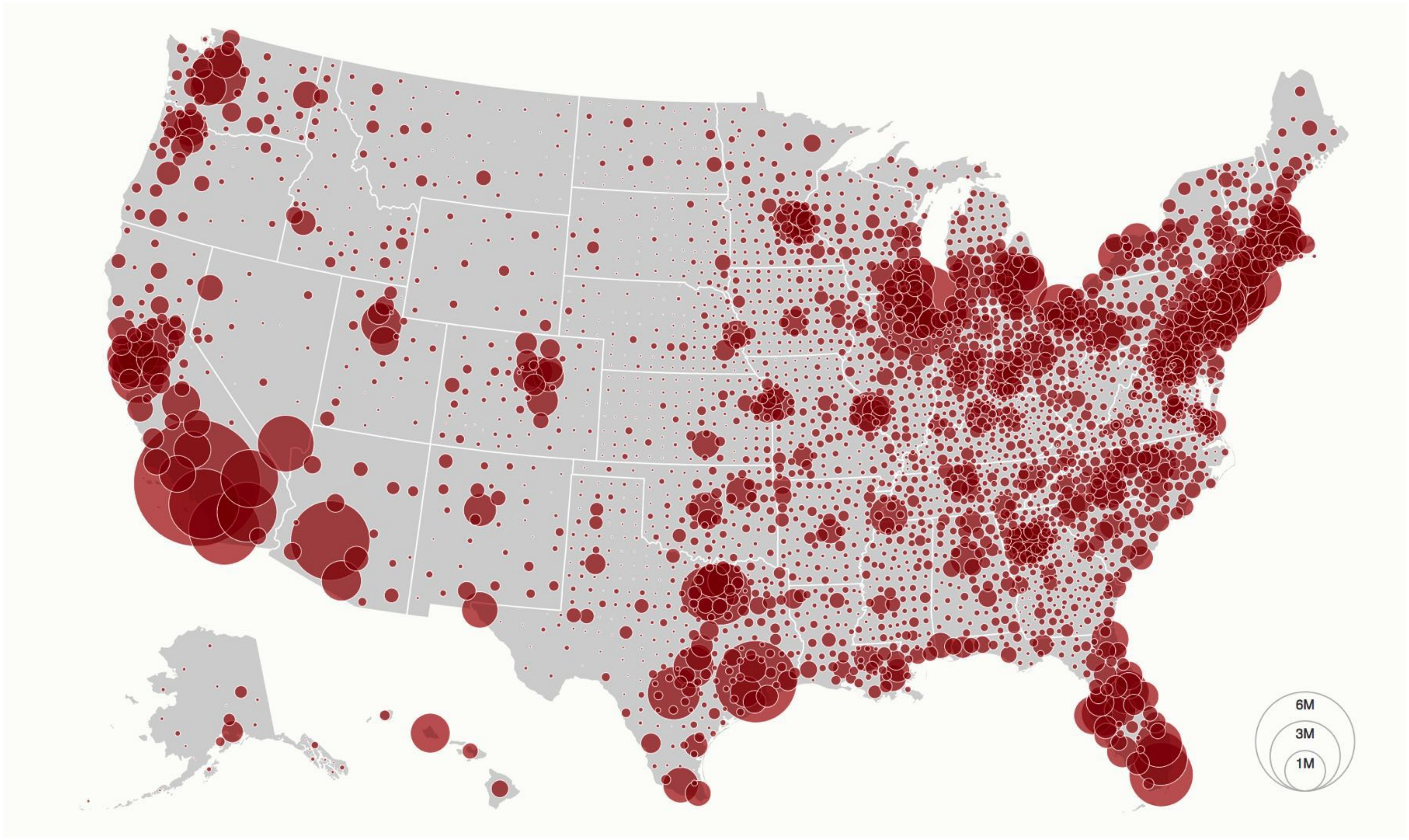
seen throughout the mapped area and smoothly transitions from one value to another, e.g., air temperature.

Surface Volume

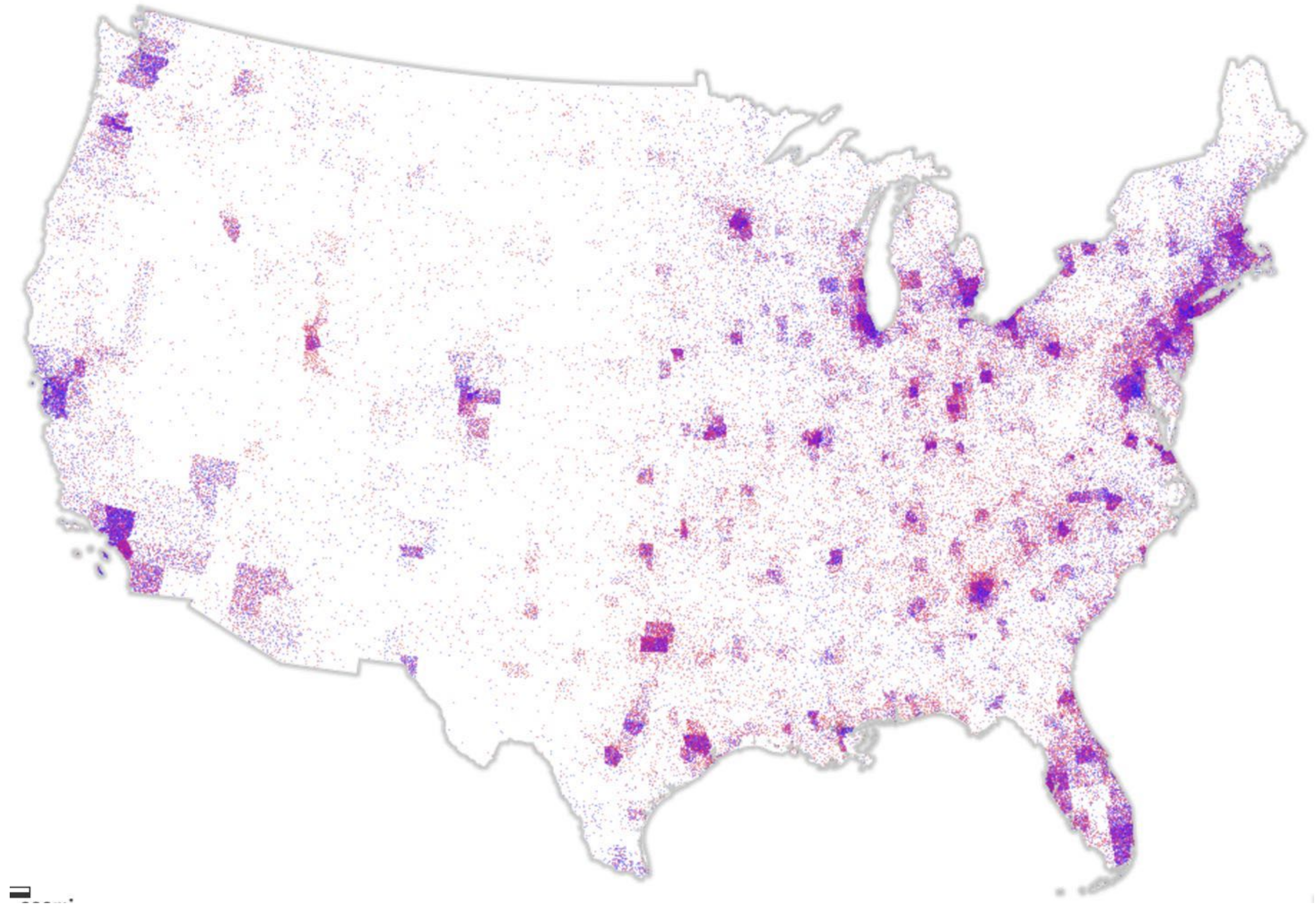
Design is driven by

- **Data**
  - Categorical, ordinal, interval, ratio
- **Spatial scale and granularity**
  - discrete vs continuous
- **Human visual perception and aesthetics**
  - choosing the correct visual variables, e.g., symbols, colors
- **Audience**
  - knowing who will read the thematic map and for what purpose helps define how it should be designed
    - political scientist vs biologist

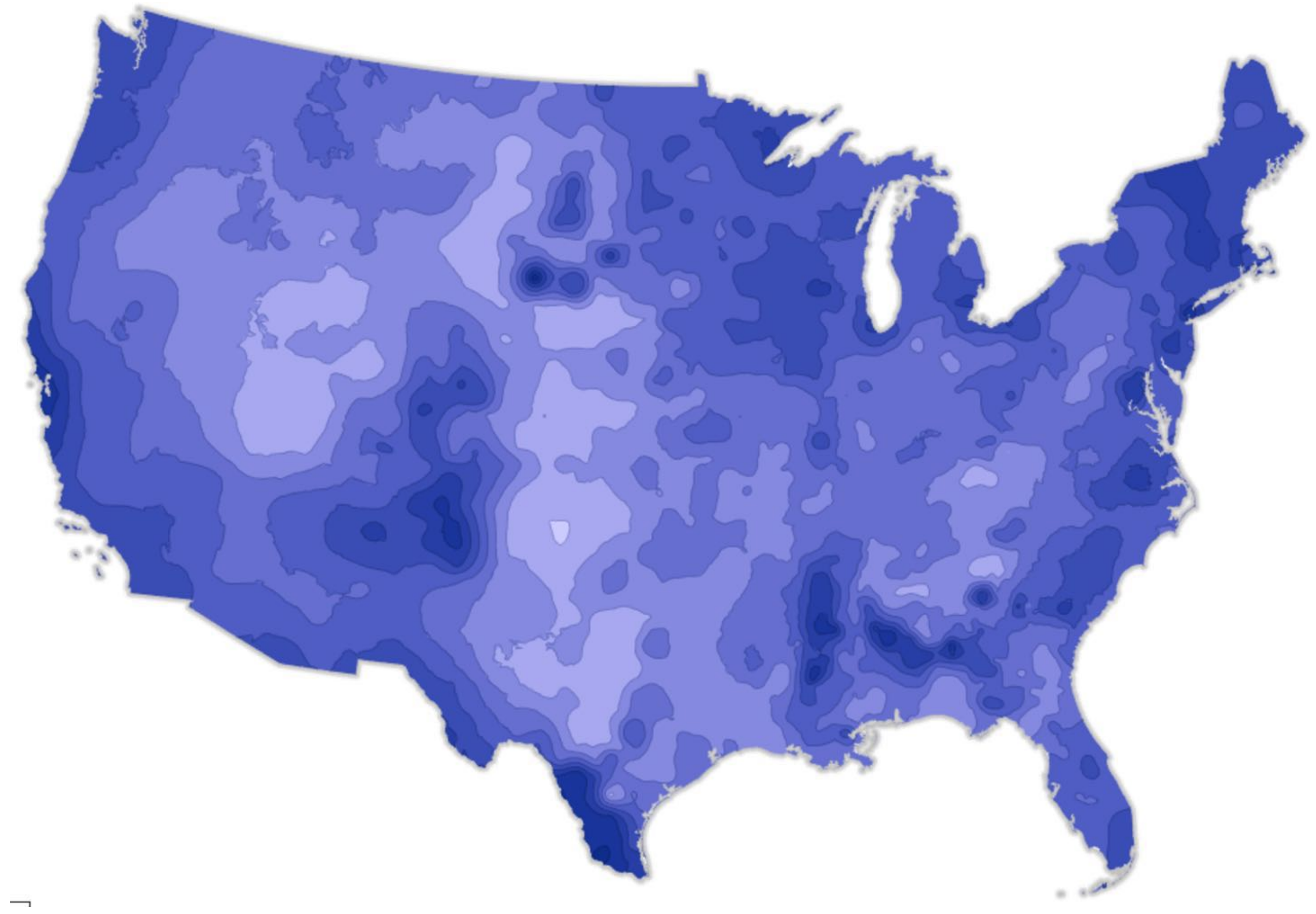
Proportional symbols



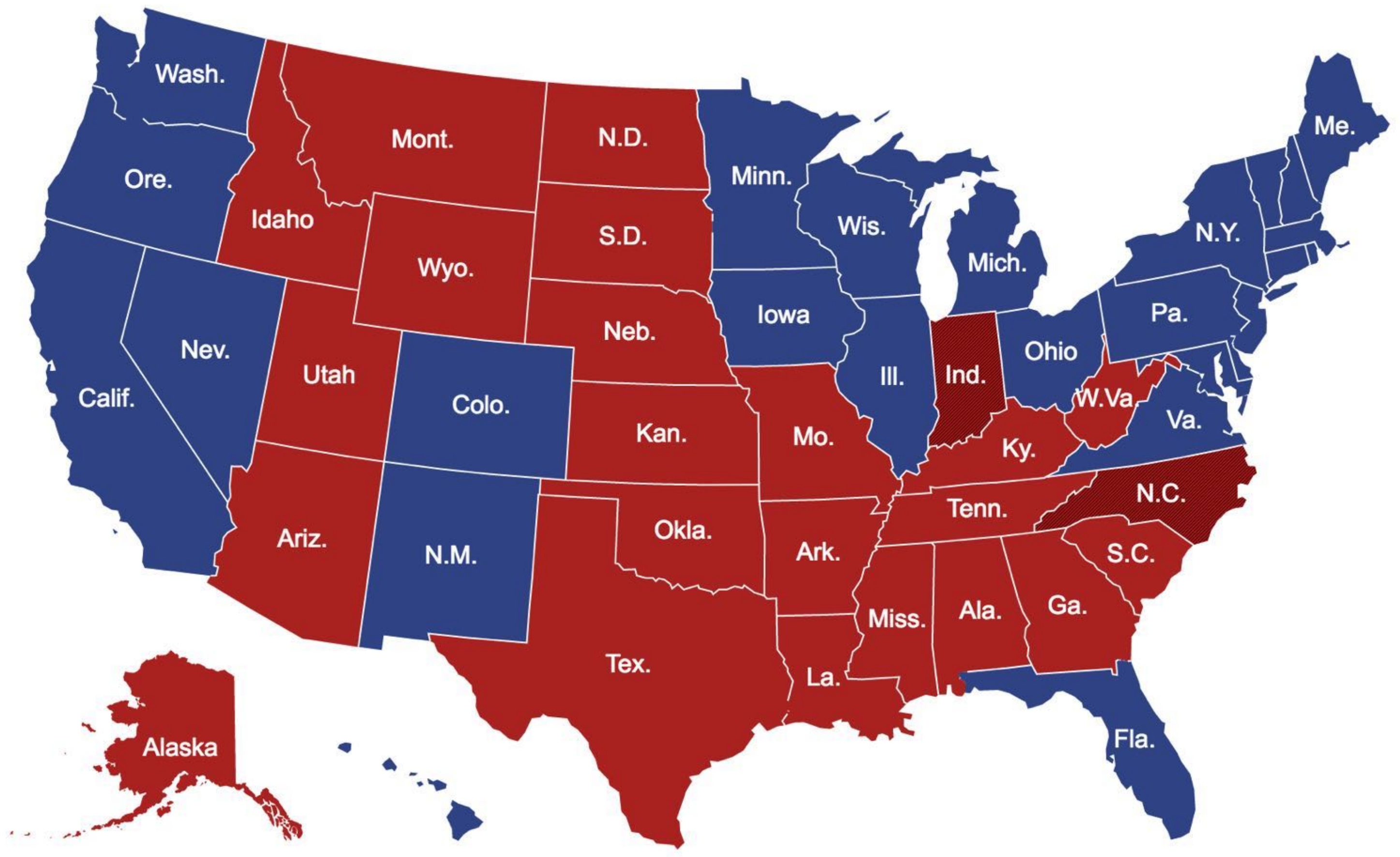
Dot distribution



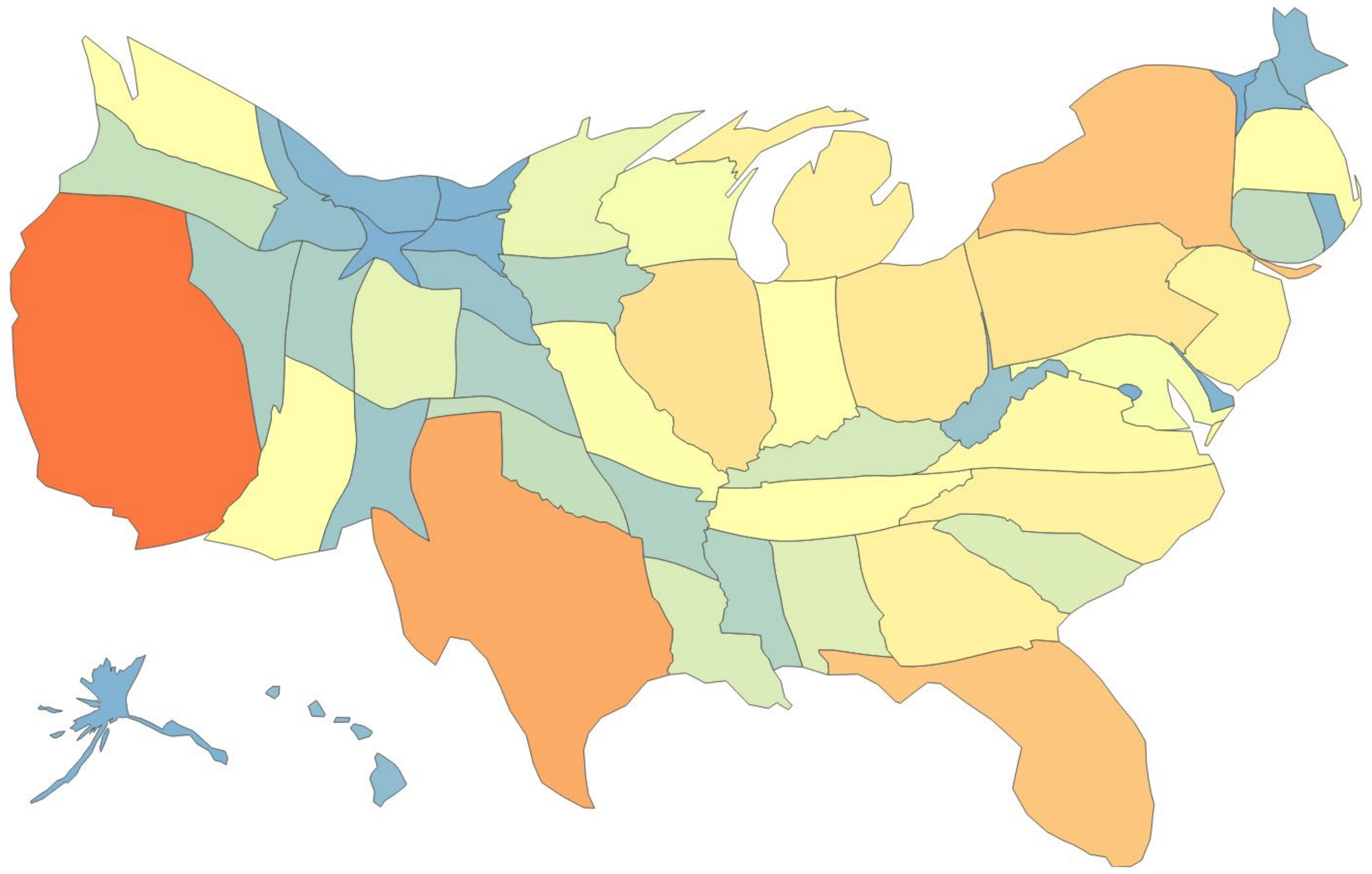
Isopleth



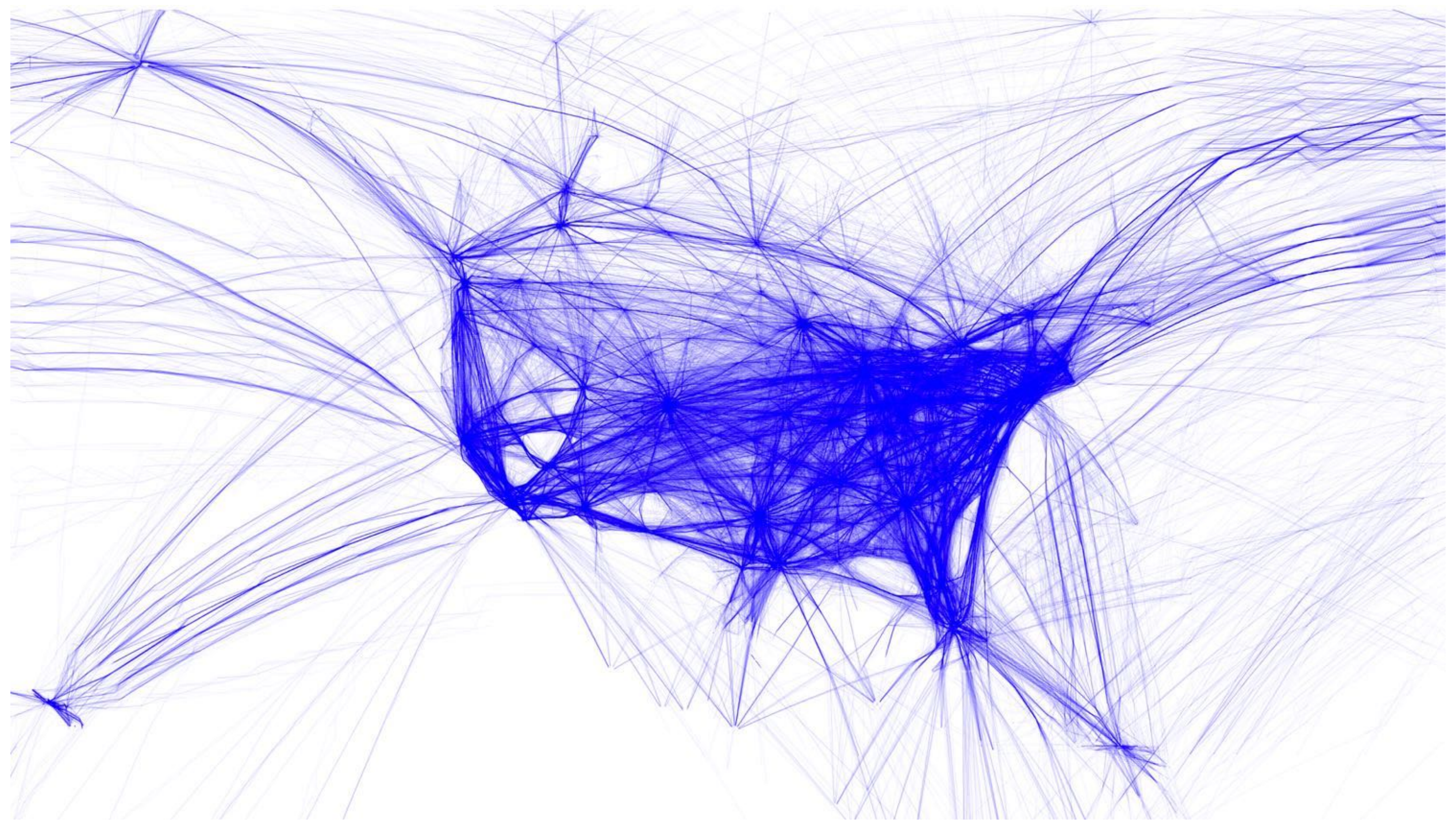
Choropleth



Cartograms

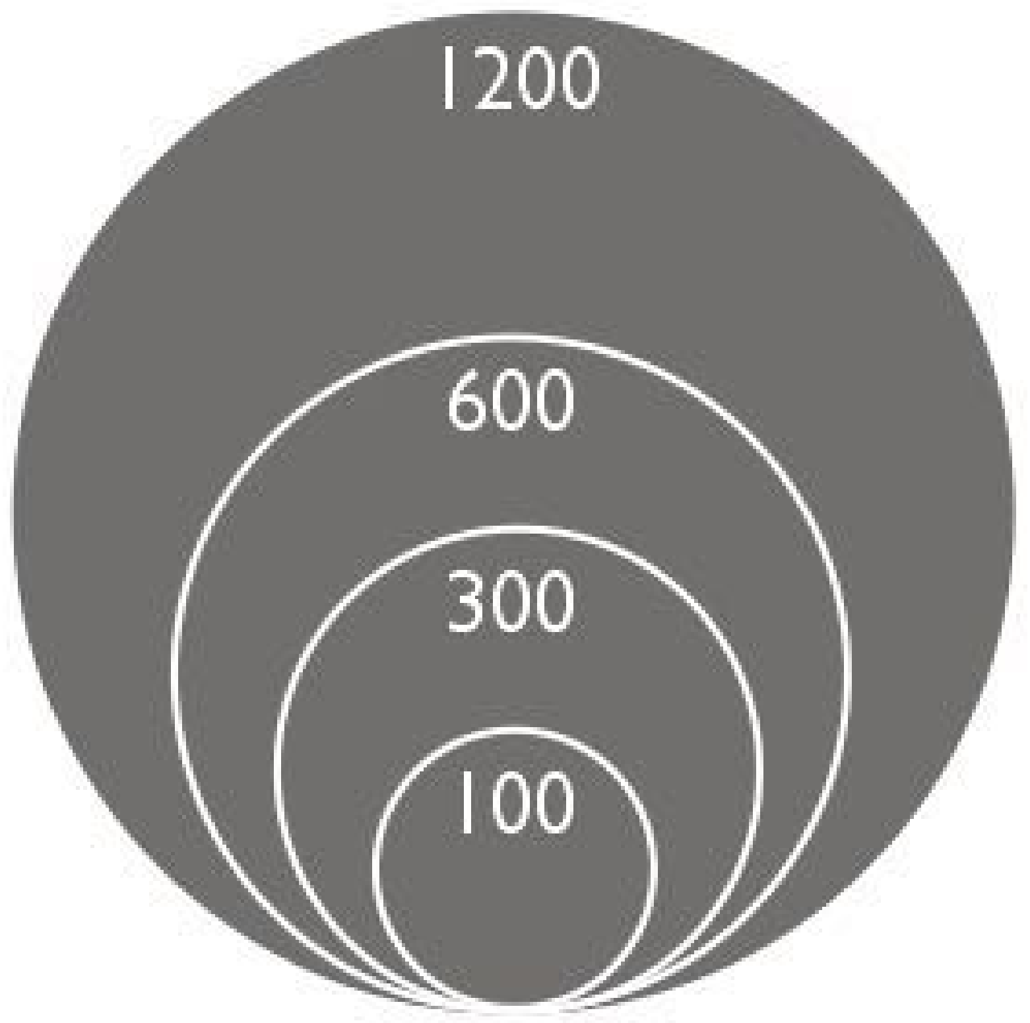


Flow Maps

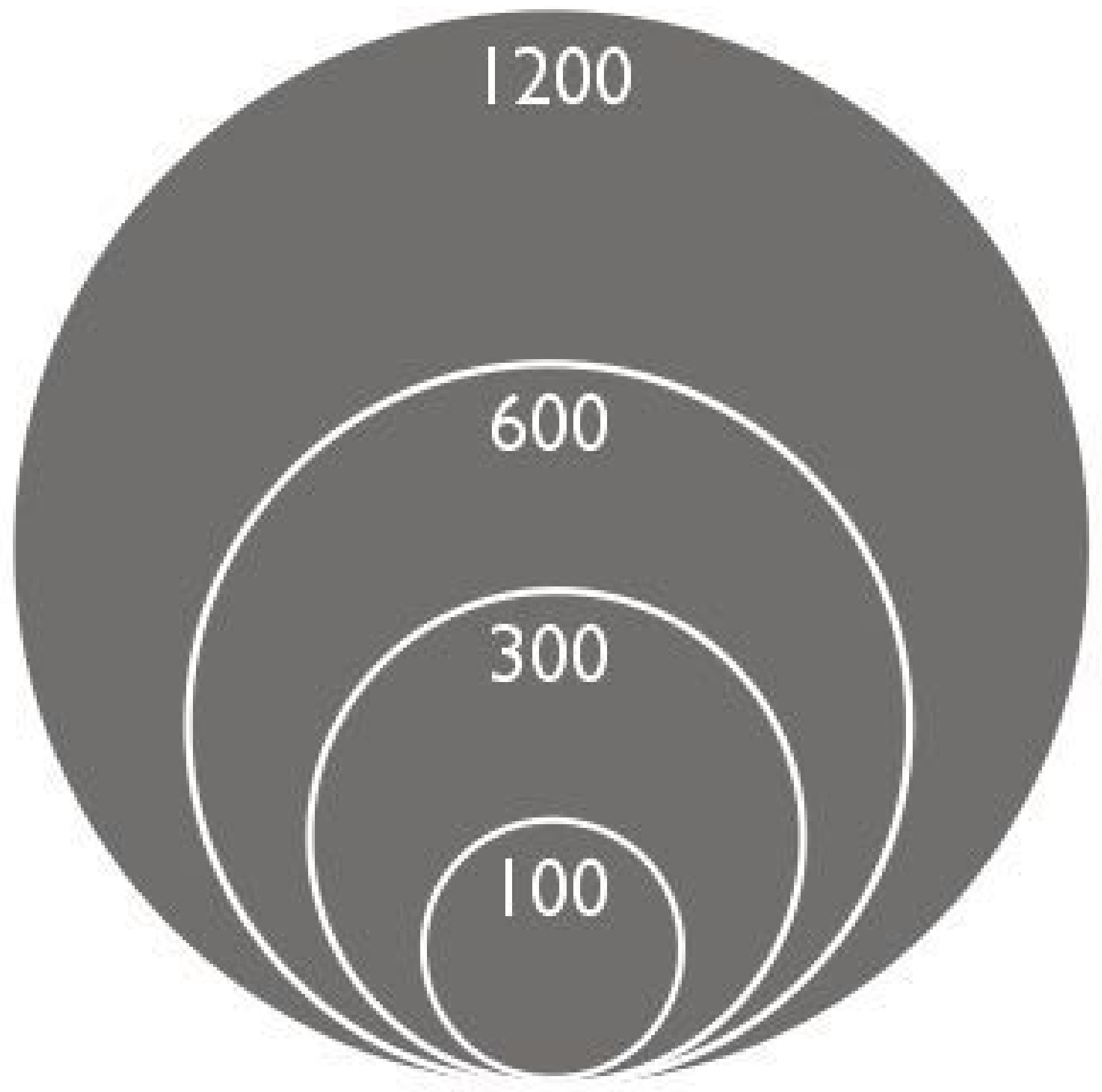


Proportional symbol maps

- Represent data variables by symbols that are sized, colored according to their amount or type.
- Data is (or can be) aggregated at points within areas.
- Three methods for setting symbol size:
  - absolute scaling
  - apparent magnitude (perceptual) scaling
- psychophysical research revealed that people tend to correctly estimate lengths, and to underestimate areas and volumes.
  - range grading



Absolute Scaling



Apparent Scaling
(Flannery's Compensation)

Proportional Symbol

2012 US Presidential election results by County, by total votes

Map type

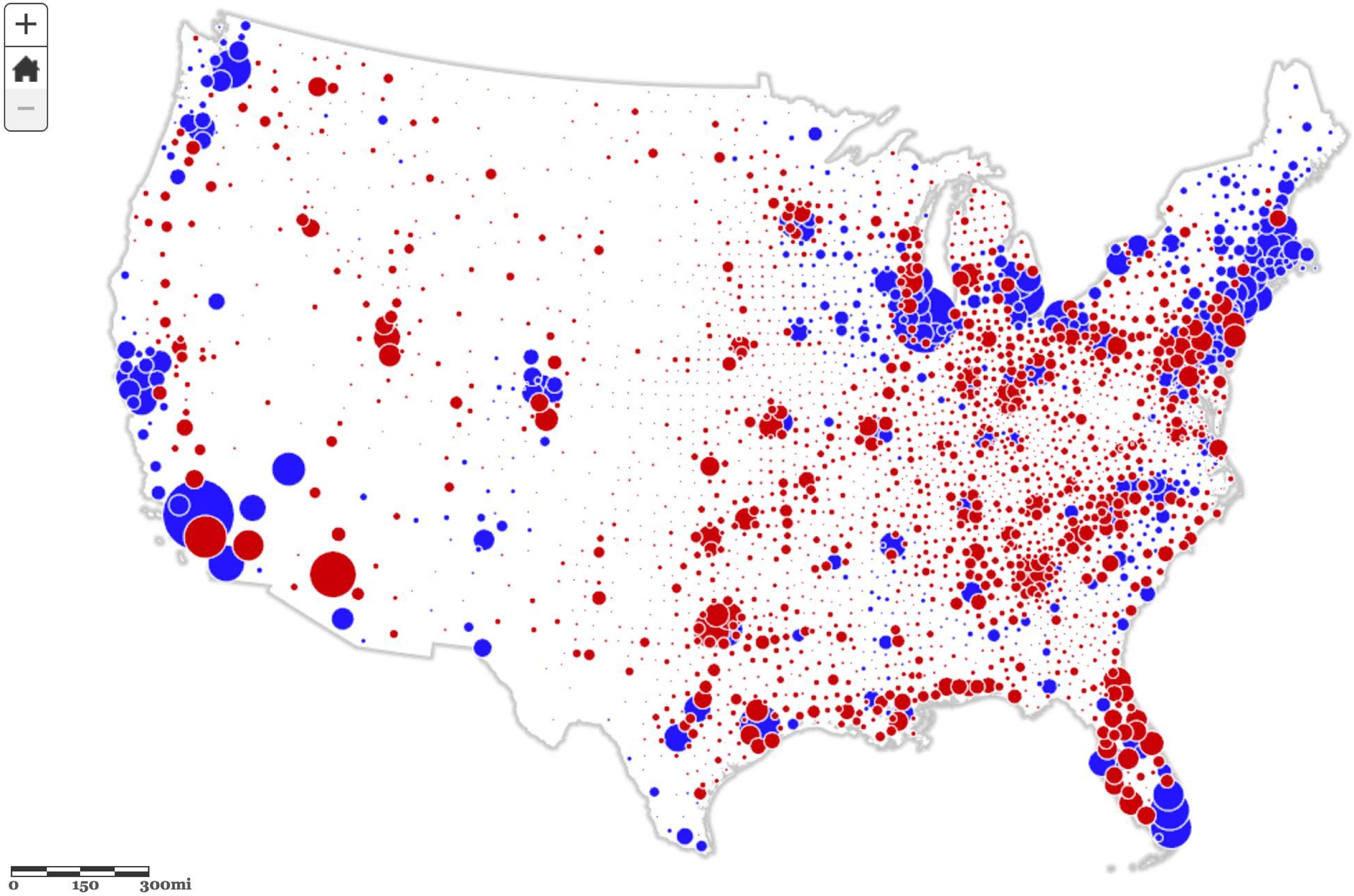
The purpose of a **proportional symbol** thematic map is to show how features differ in quantity for the theme being mapped. In this example of the 2012 Presidential election, the map is designed to show the number of votes cast for the predominant party in each County.

Data

For the theme being mapped, the data should
be **numerical (quantitative)** and represent differences between features on an **interval or ratio** scale of measurement. The map type requires data to be absolute, as totals. Here, the vote totals are augmented by symbols that define a second **categorical** characteristic of the data, namely 'Republican' or 'Democrat'.

Symbols

Symbols are scaled to the data values and should be designed so that different magnitudes of data can be easily distinguished from one another through variation in the **size** of the symbol, used as an **ordering visual variable**. Symbols should be scaled so that the smallest are visible and the largest do not overly smother the map



Kenneth Field, politico.com

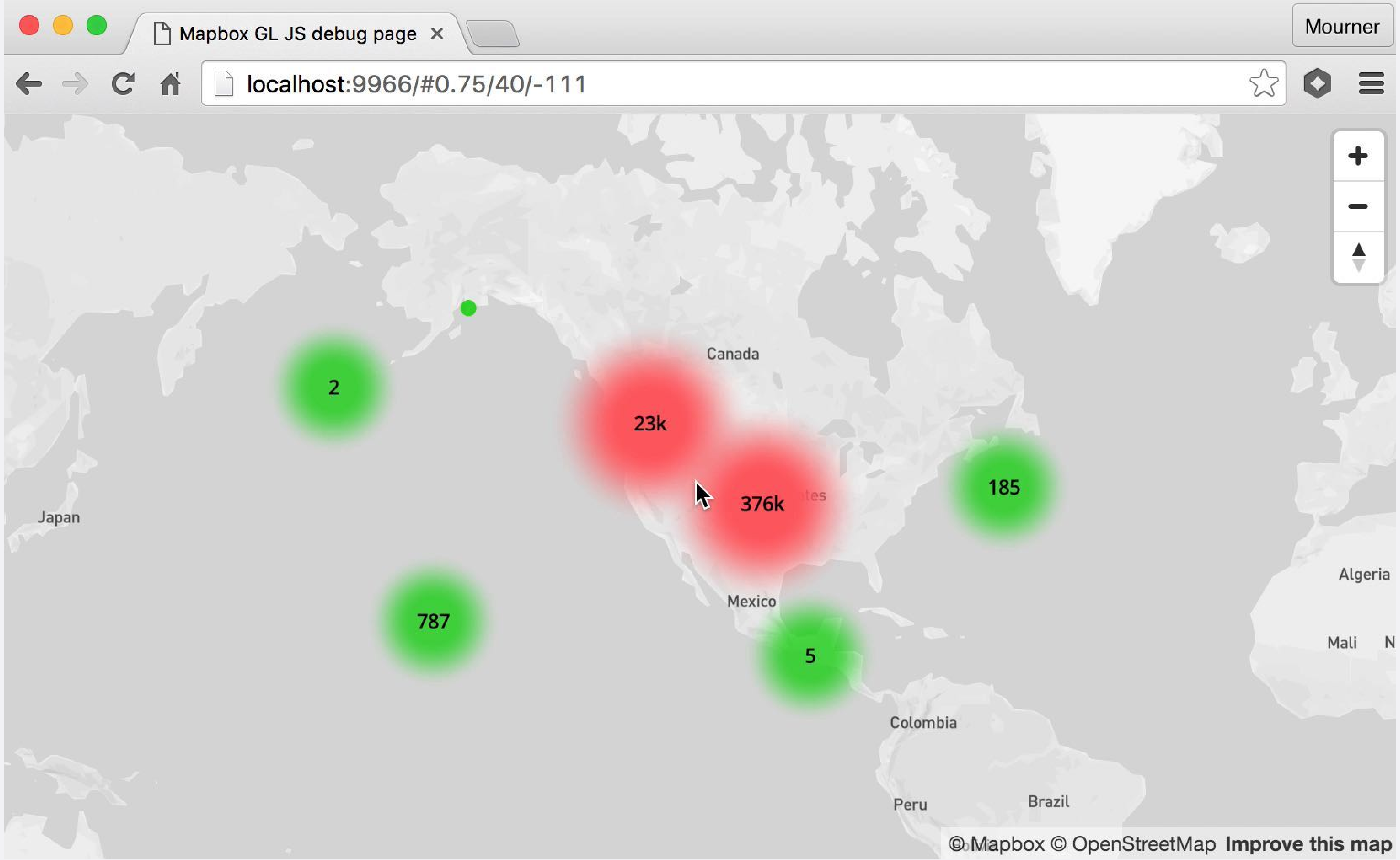
dot distribution

Twitter and Flickr Activity

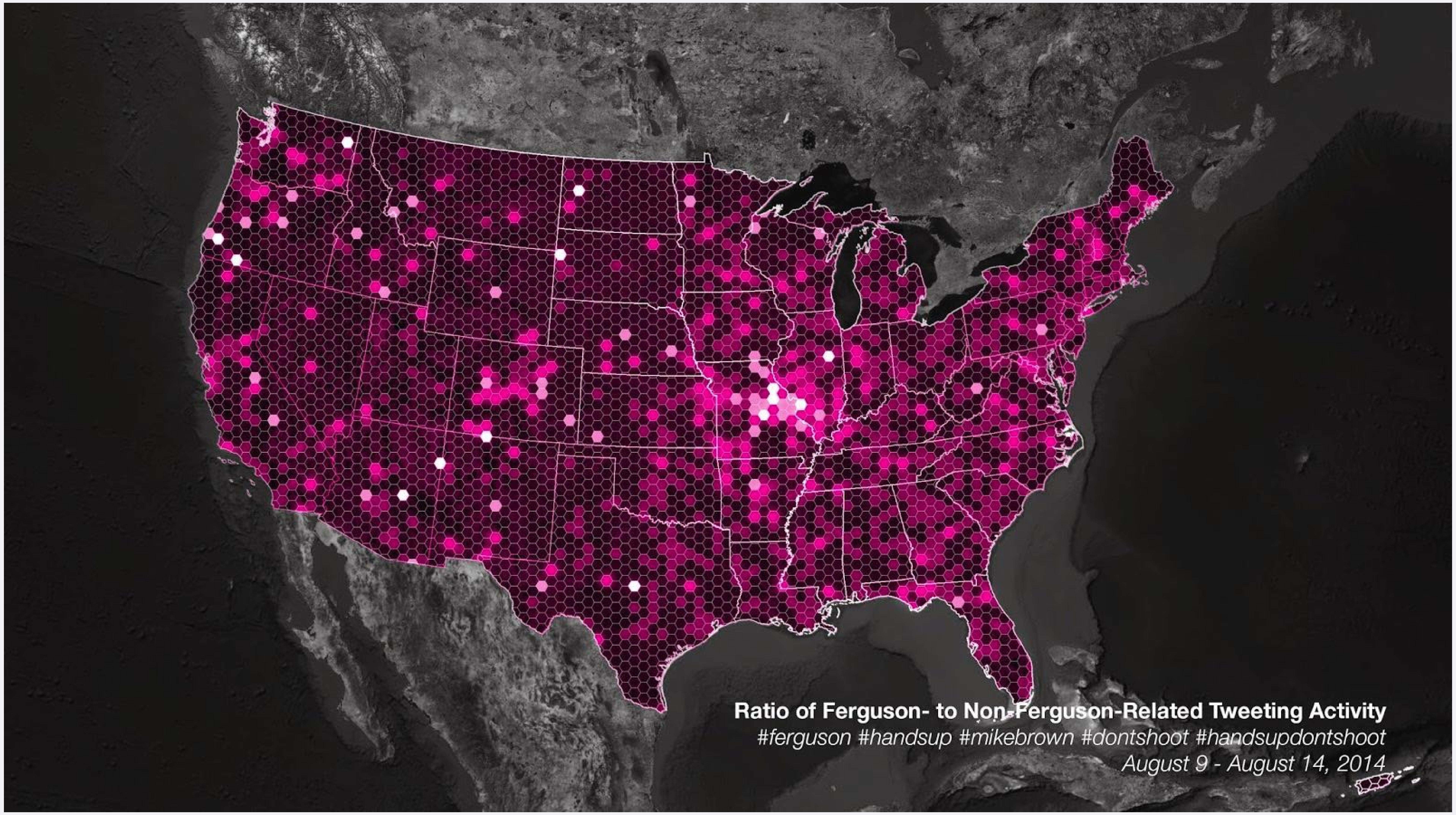
New York

San Francisco

Clustering (dealing with a lot of points)



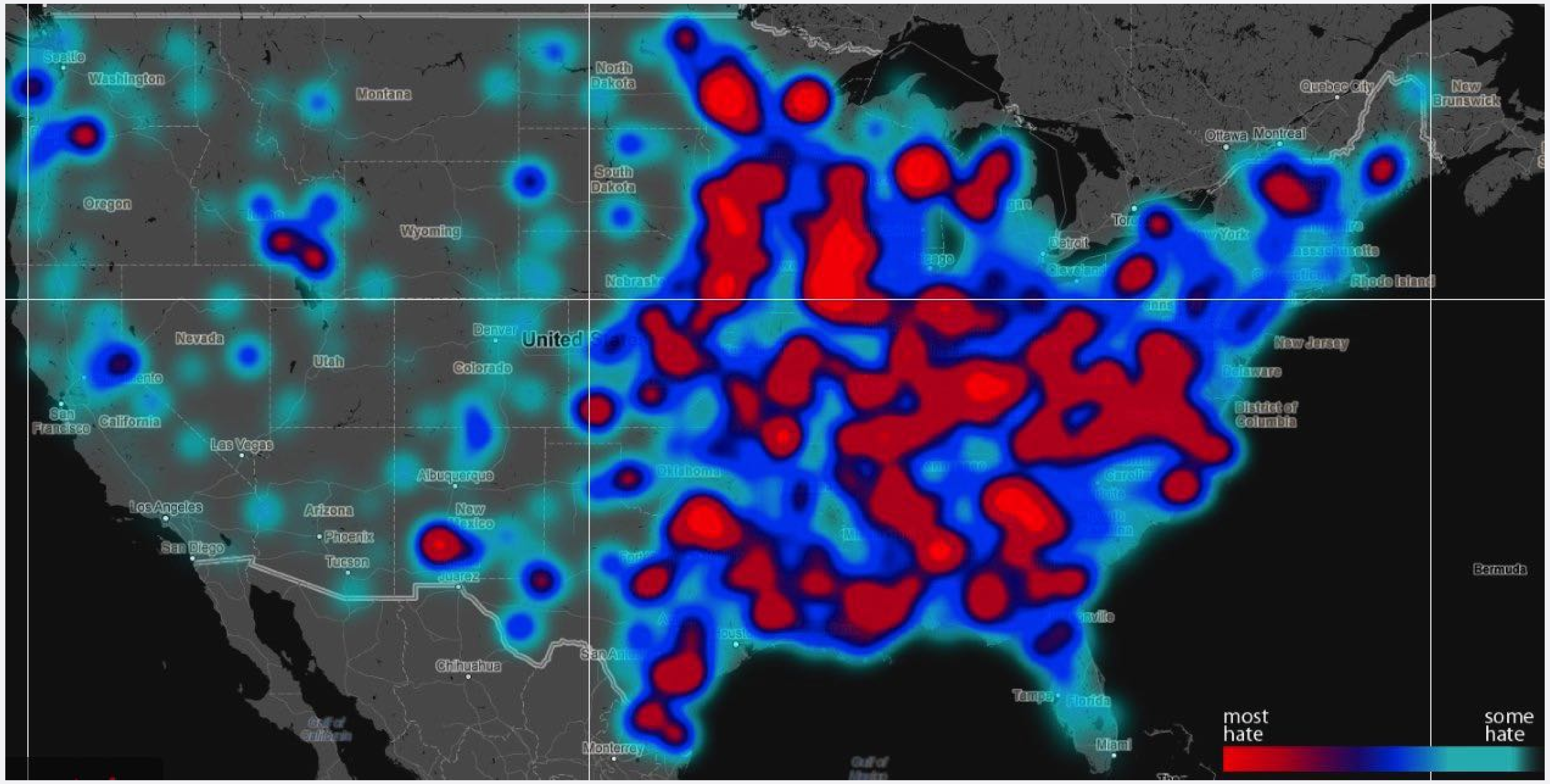
Hexbins

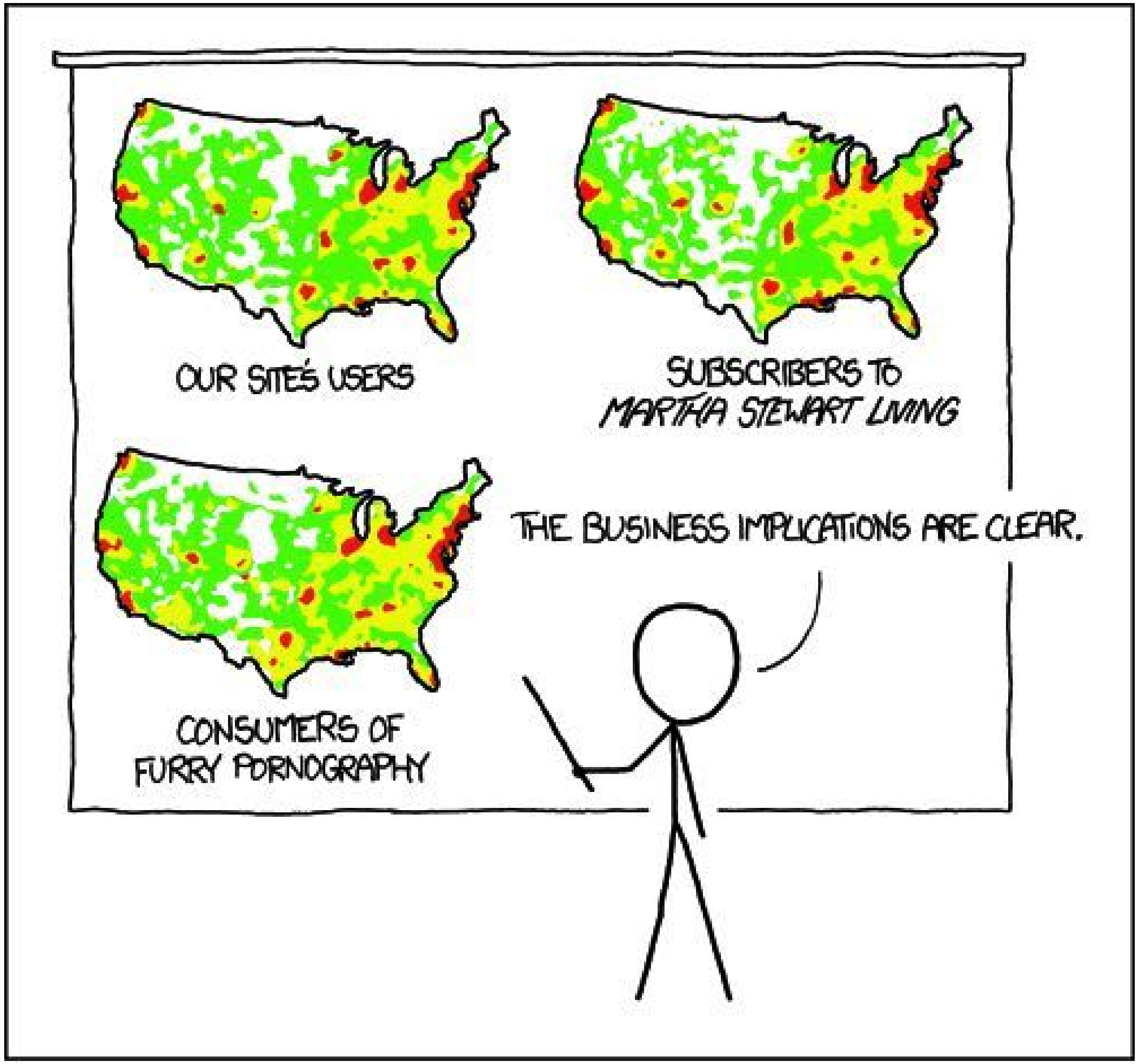


Ratio of Ferguson- to Non-Ferguson-Related Tweeting Activity
#ferguson #handsup #mikebrown #dontshoot #handsupdontshoot
August 9 - August 14, 2014

Heatmaps

- Used to identify clusters where there is a high concentration of activity (attribute under analysis)
- They can ben also useful for doing hotspot analysis.




OUR SITE'S USERS
SUBSCRIBERS TO *MARTHA STEWART LIVING*
CONSUMERS OF FURRY PORNOGRAPHY
THE BUSINESS IMPLICATIONS ARE CLEAR.

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

Isopleth: Filled Contours

2012 US Presidential election results: Democrat share of vote

Map type

An **isarithmic** map is a two-dimensional representation of a three-dimensional volume. Two types exist: an **isometric** form that is constructed from data at points and an **isoplethic** form constructed from data that occur over geographic areas. The purpose of an **isopleth** thematic map is to show how features differ in quantity as a surface. This can be achieved through representing the volume using **contour lines** or by using **filled contours** that are shaded according to the quantitative value being mapped. In this example of the 2012 Presidential election, the map is designed to show the share of the vote gained by the Democrat party based on County level data.

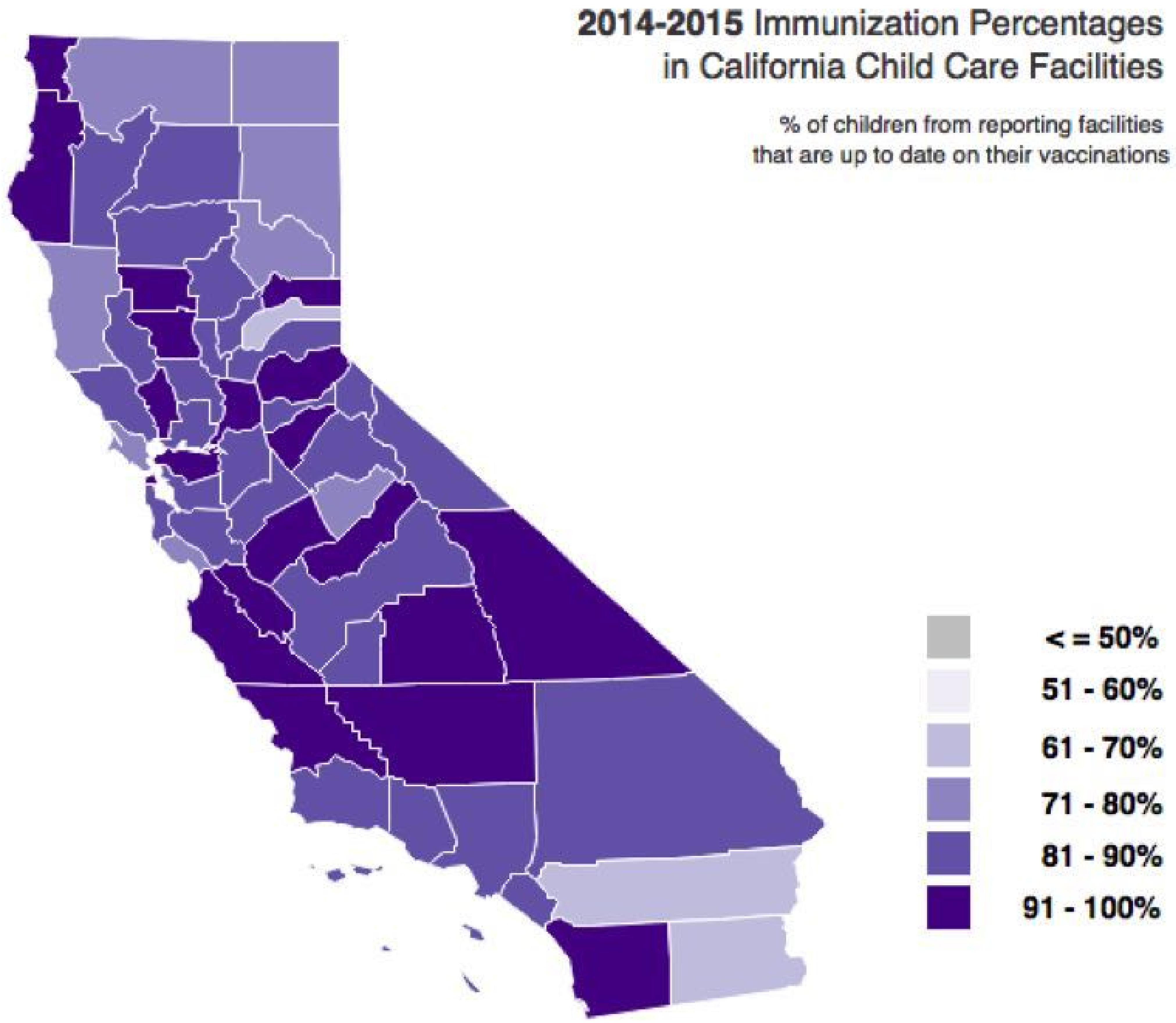
Data

Isopleth maps are generated from data that occur over geographical areas and values represent **numerical (quantitative)** diffe between features on an **interval or ratio** scale of measurement. Absolute values cannot be illustrated isoplethically due to the inherent problems of using totals for areas that might vary in size or which contain an unequal denominator of the data being mapped. This is the issue that prevents **choropleths** from being used to map totals and the same occurs for

Kenneth Field, politico.com | Kenneth Field

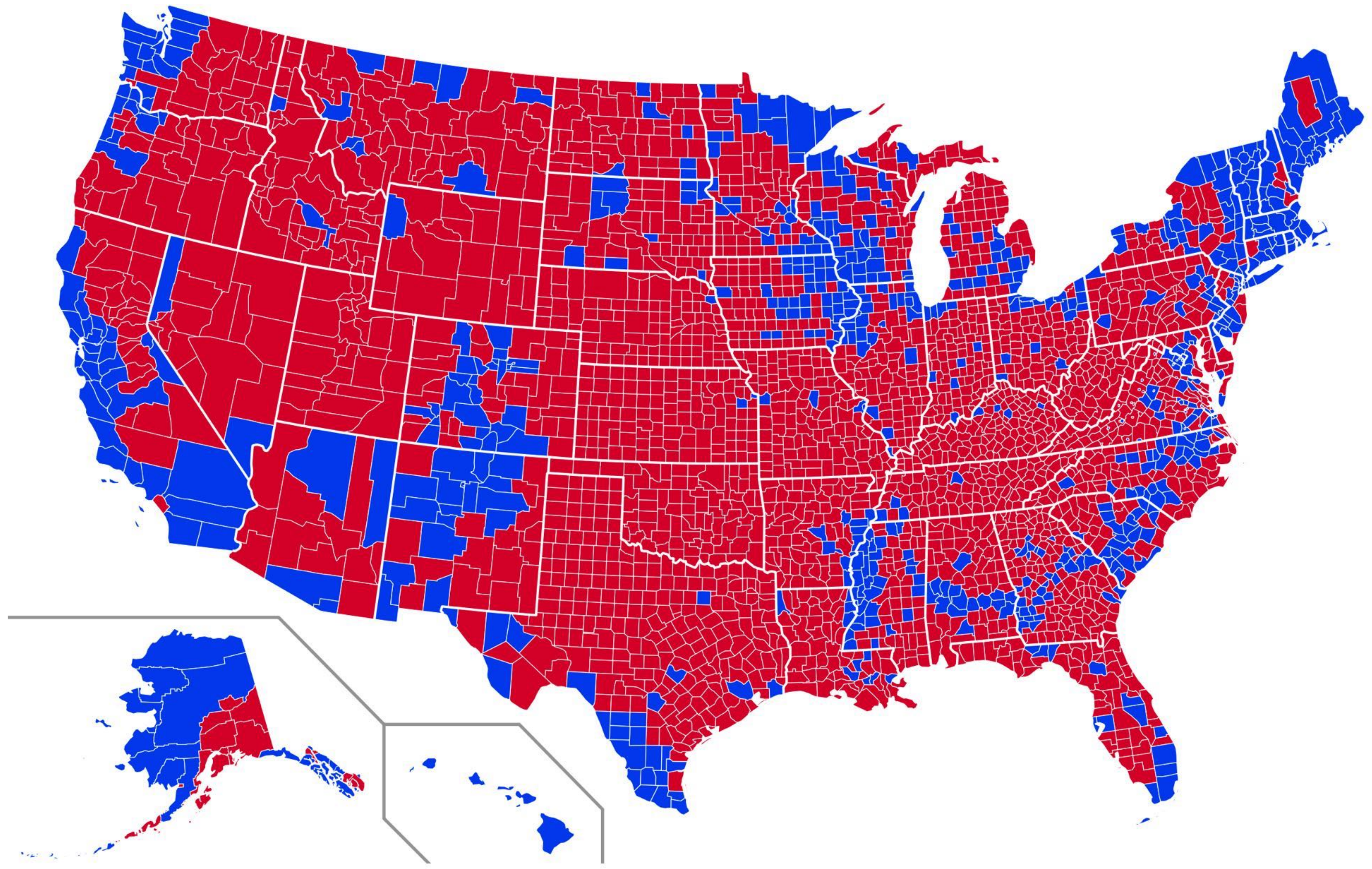
Choropleth
(from Greek χῶρος ("area/region") + πλῆθος (“multitude"))

- **Areas are shaded or colored in proportion to the measurement of the statistical variable being displayed on the map.**
- **Key factors:**
  - Resolution of the base map
  - Data
    - source and processing
    - classification
    - MAUP
    - legend
  - Symbolization

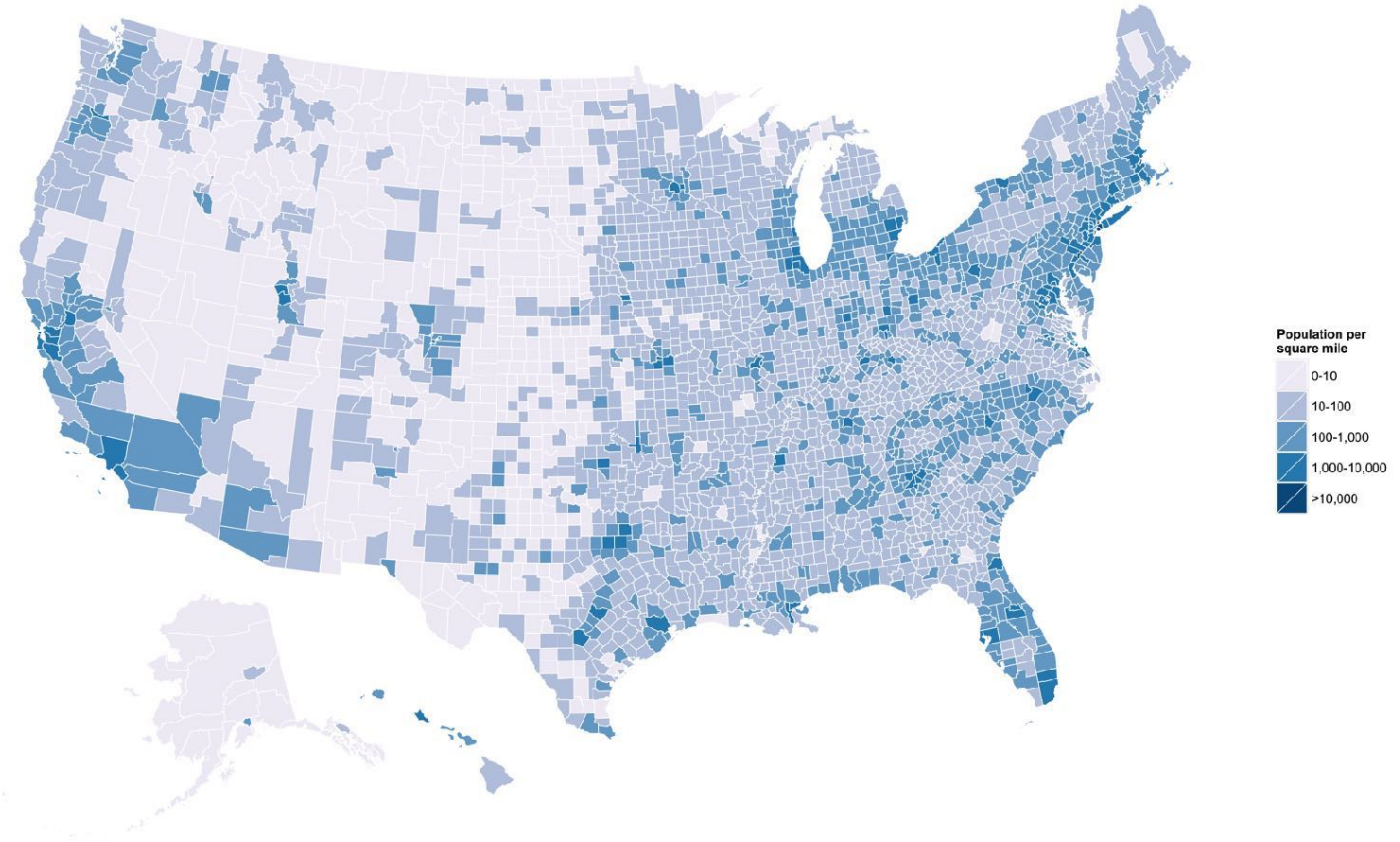

2014-2015 Immunization Percentages
in California Child Care Facilities
% of children from reporting facilities
that are up to date on their vaccinations
< = 50%
51 - 60%
61 - 70%
71 - 80%
81 - 90%
91 - 100%

Data type

Presidential election 2008



Population density 2014



CATEGORICAL

Obama or Romney

CONTINUOUS

interval [0, 1]

Data Type

Continuous

(sequential)

!=

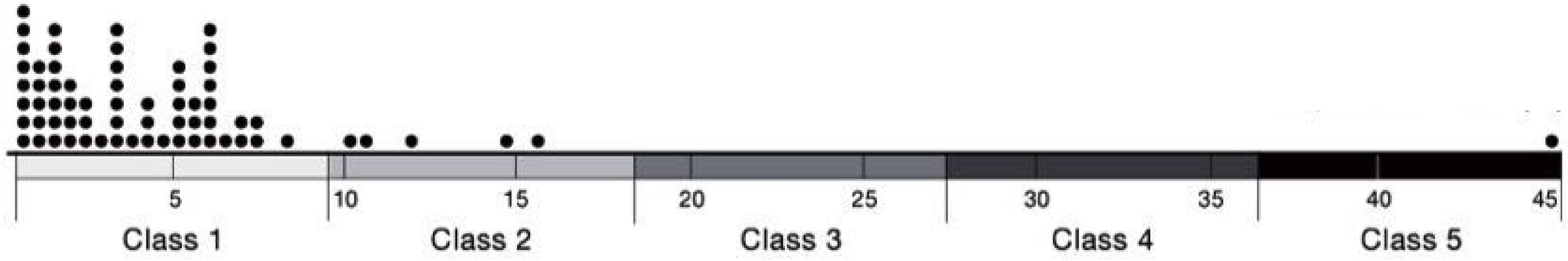
Color Scheme

Categorical

(qualitative)

Classification

- **Take observations and group them into data ranges or classes**



How many classes?

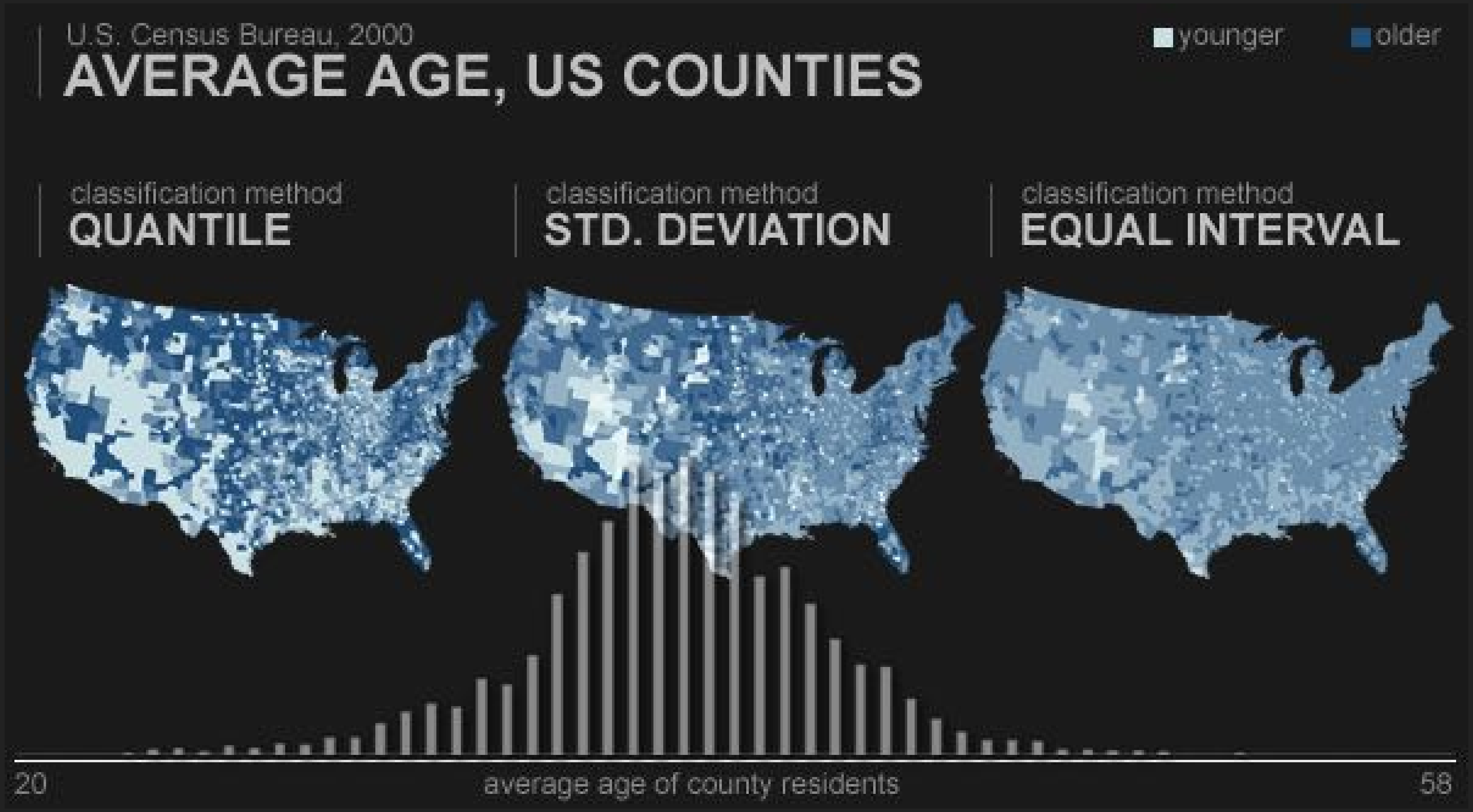
5-7 ± 2

George Miller (1956)
short term memory capacity

Which method?

Classification methods

- **Natural breaks**
- **Equal intervals**
  - not valid if your data is skewed or in presence of outliers.
- **Quantiles**
  - can position elements in a class even if being closer to the adjacent
- **Standard deviation**
- **Fisher-Jenks: reduce the variance within classes and maximize the variance between classes**
  - unique classification, hard to compare between maps.
- **Python PySAL library implementation**
  - http://pysal.readthedocs.io/en/latest/library/esda/mapclassify.html



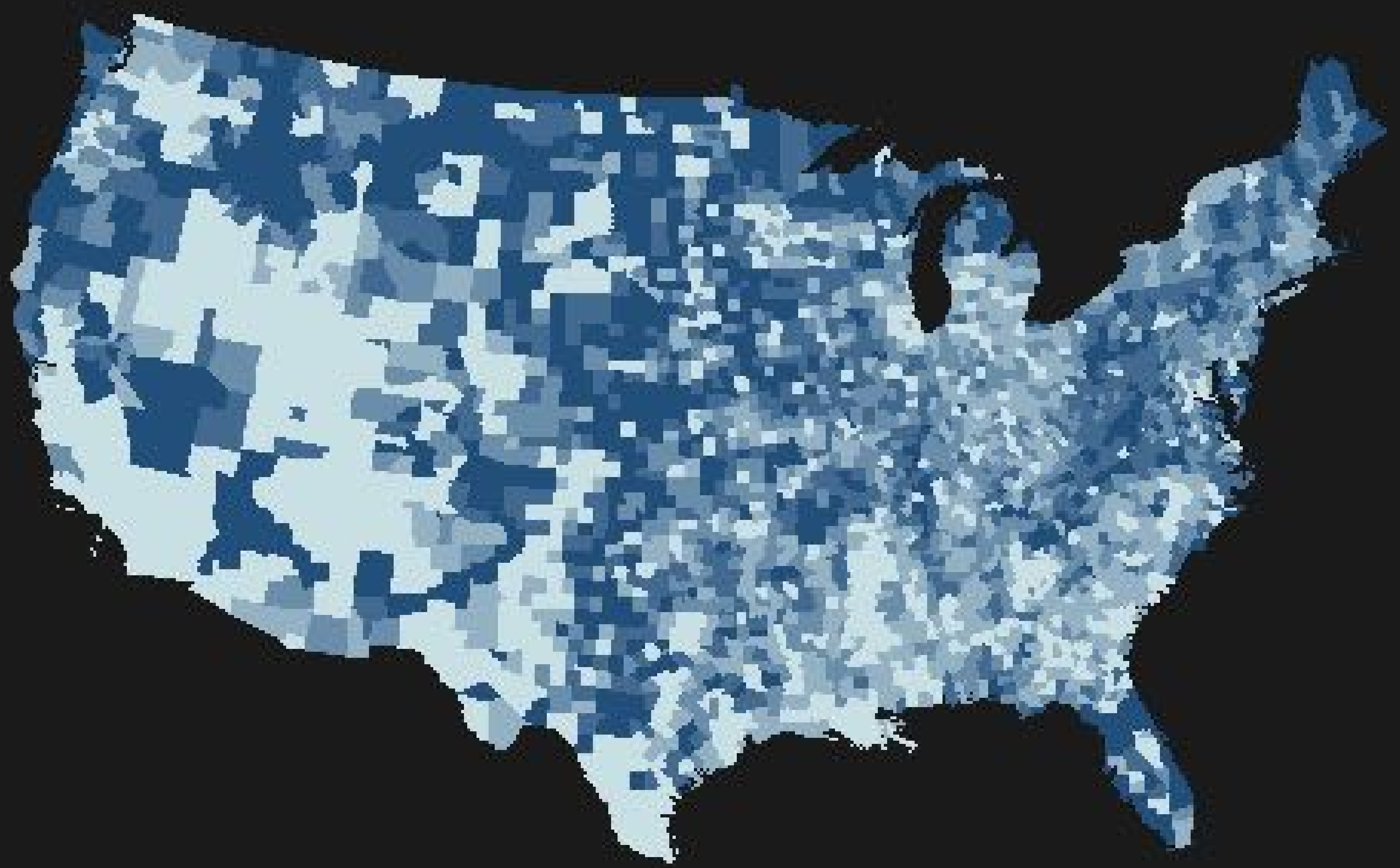
http://uxblog.idvsolutions.com/2011/10/telling-truth.html

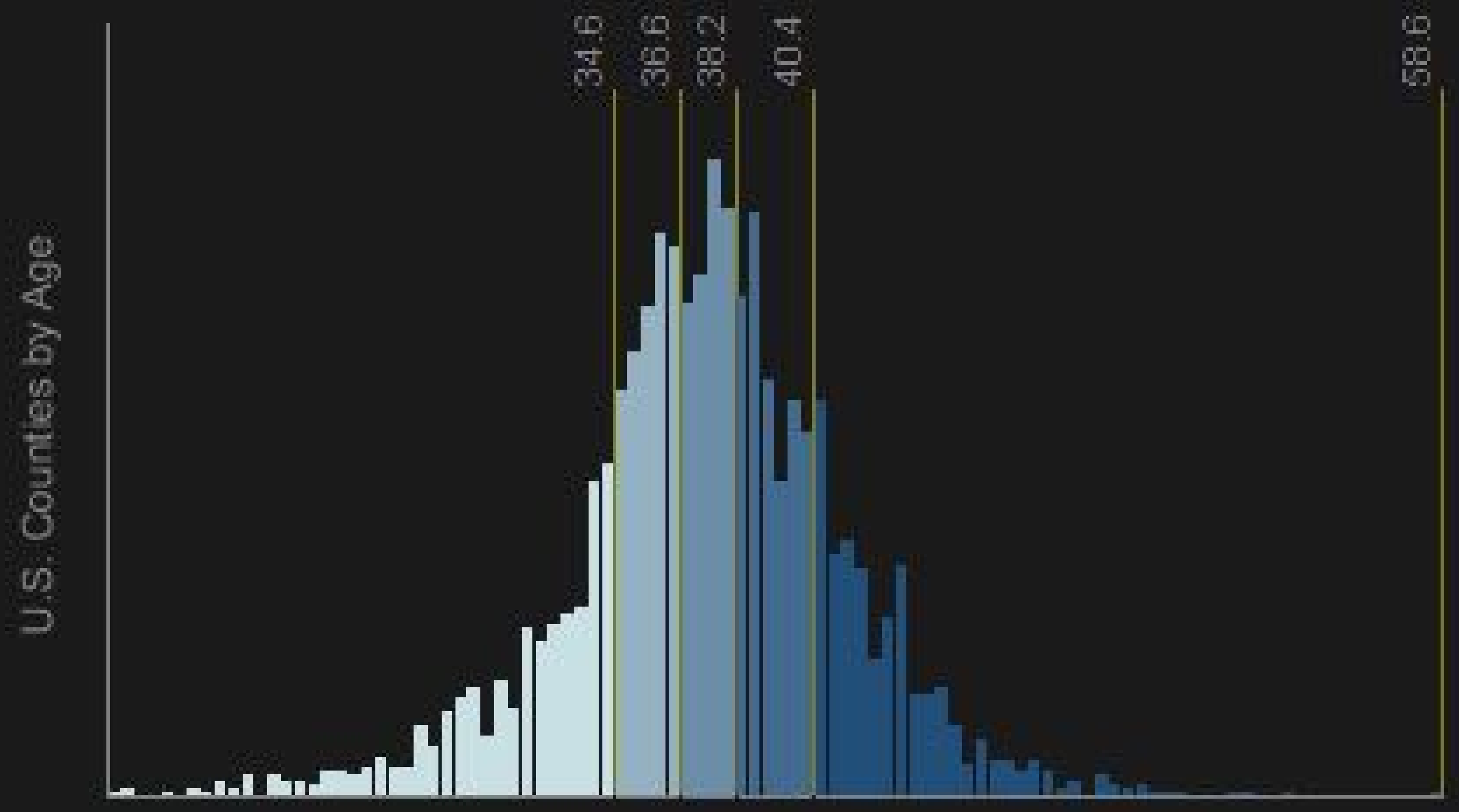
U.S. Census Bureau, 2000

MEDIAN AGE

classification

QUANTILE



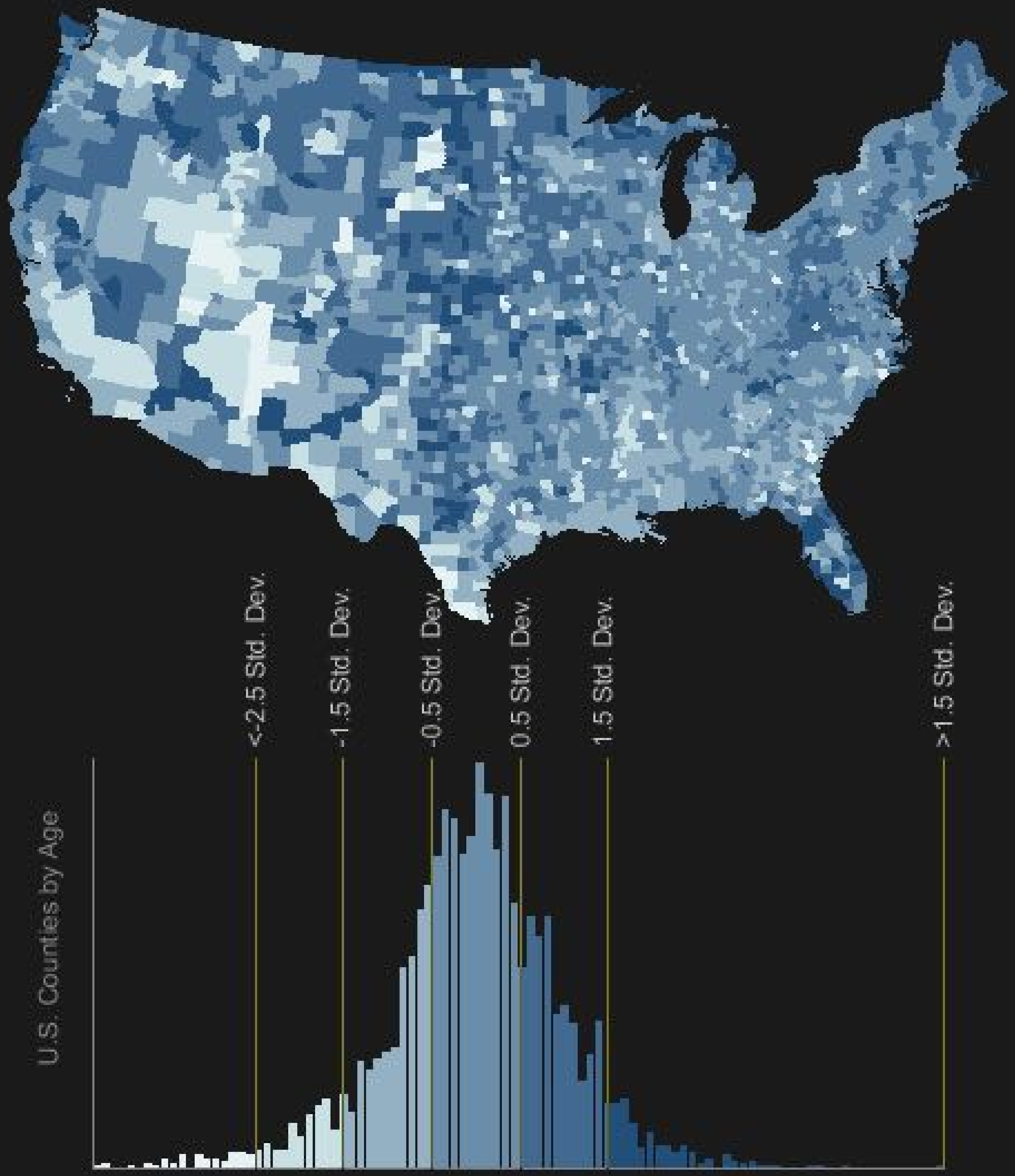


U.S. Census Bureau, 2000

MEDIAN AGE

classification

STD. DEVIATION

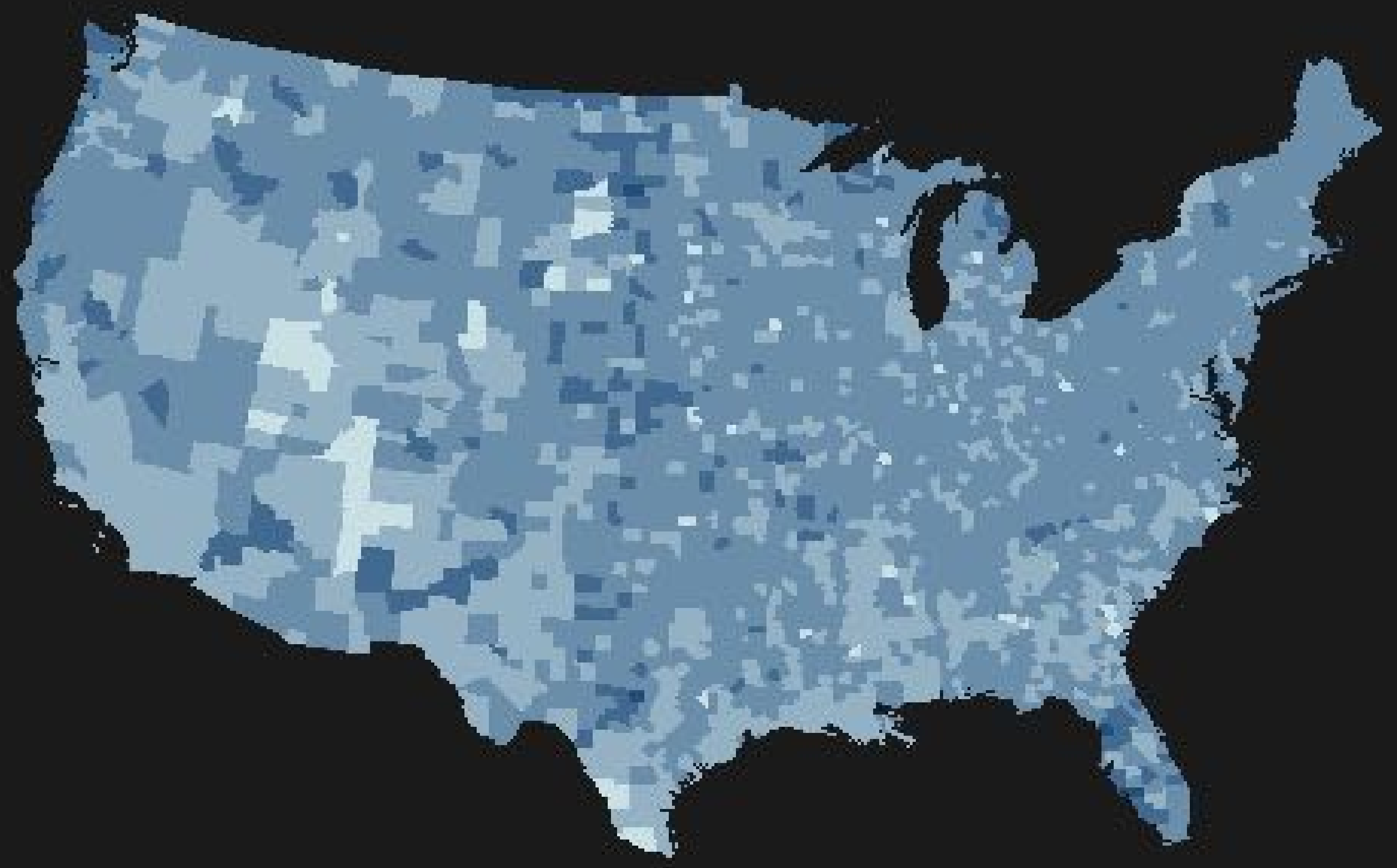


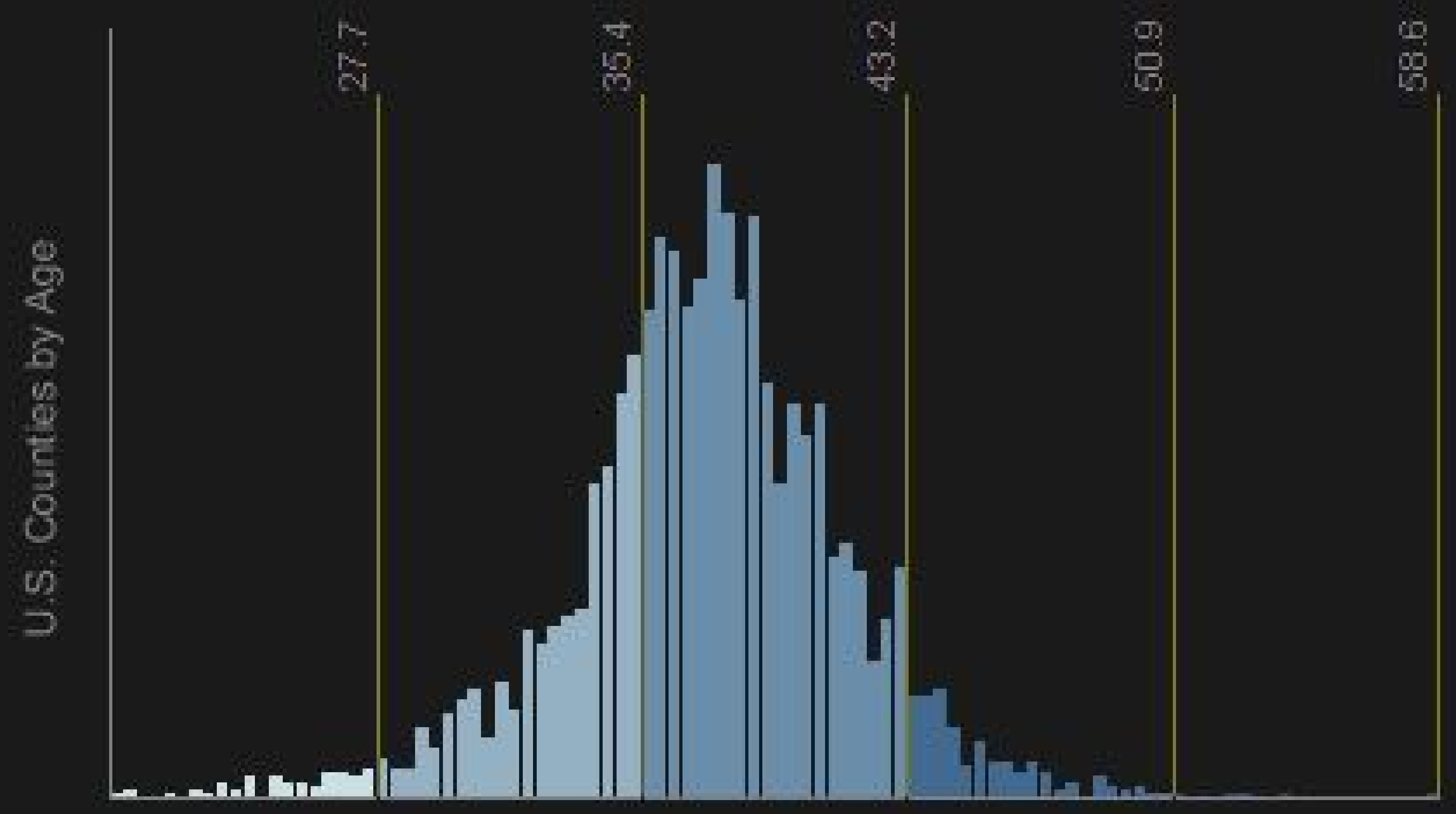
U.S. Census Bureau, 2000

MEDIAN AGE

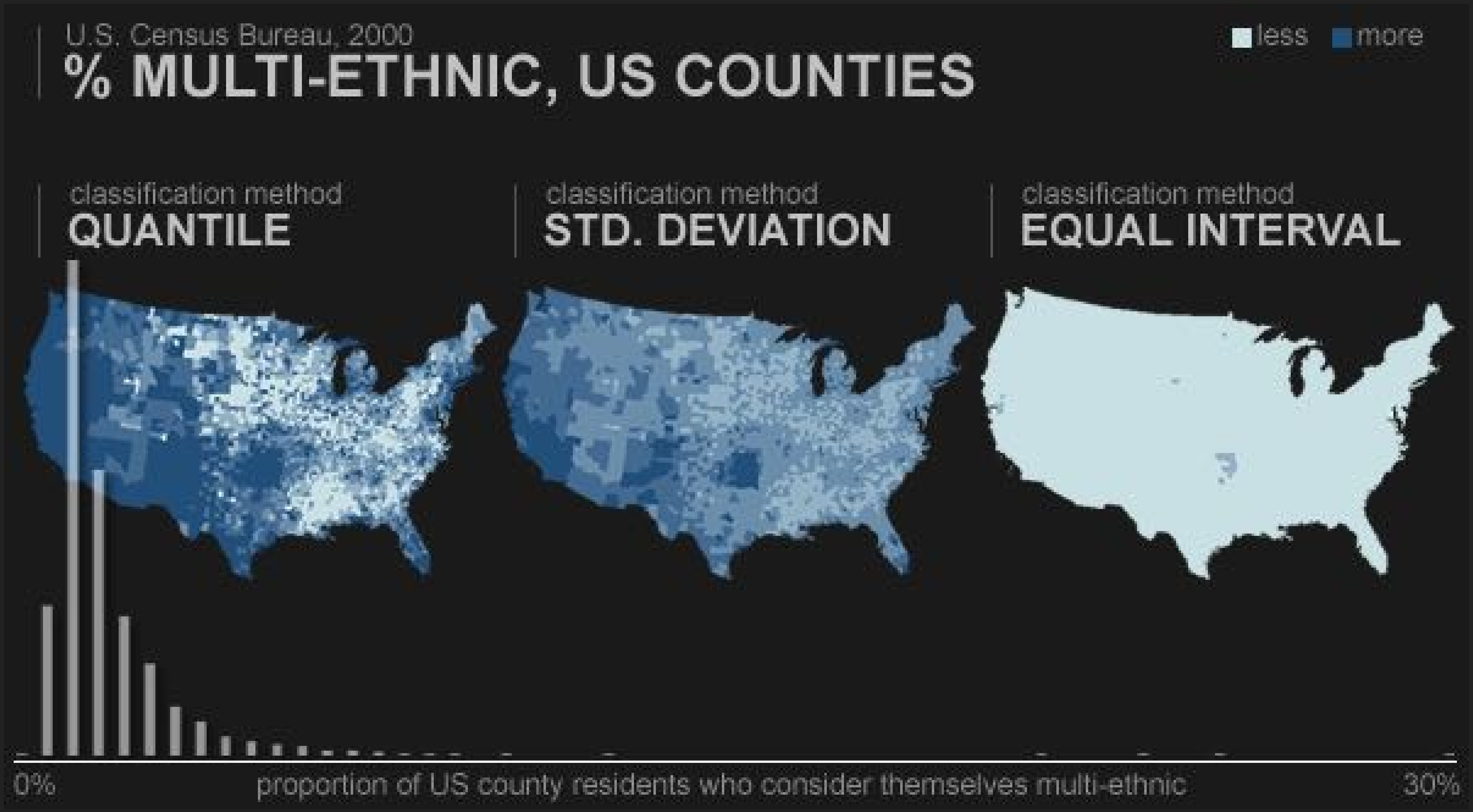
classification

EQUAL INTERVAL





Proportion of US county residents who consider themselves multiethnic

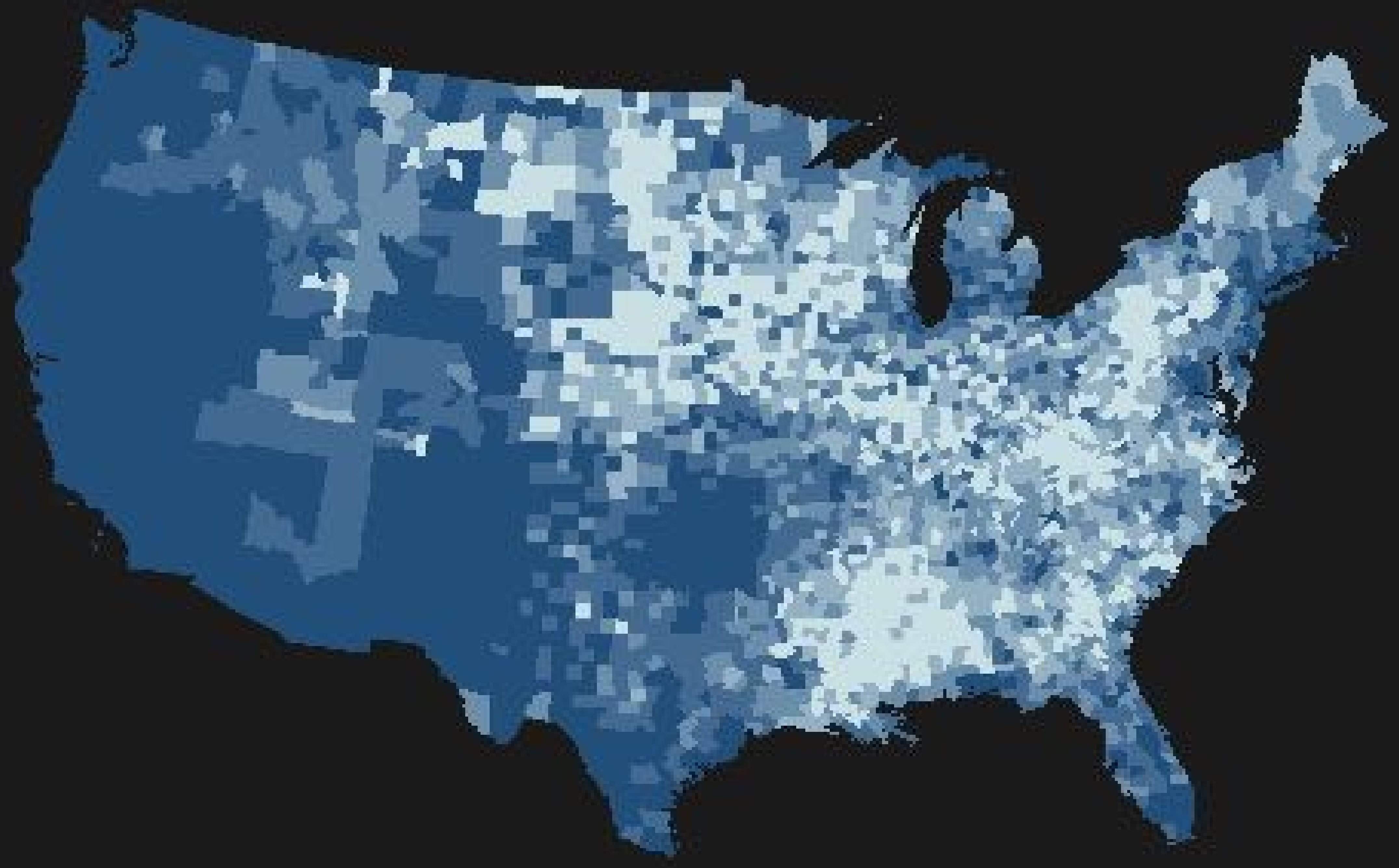


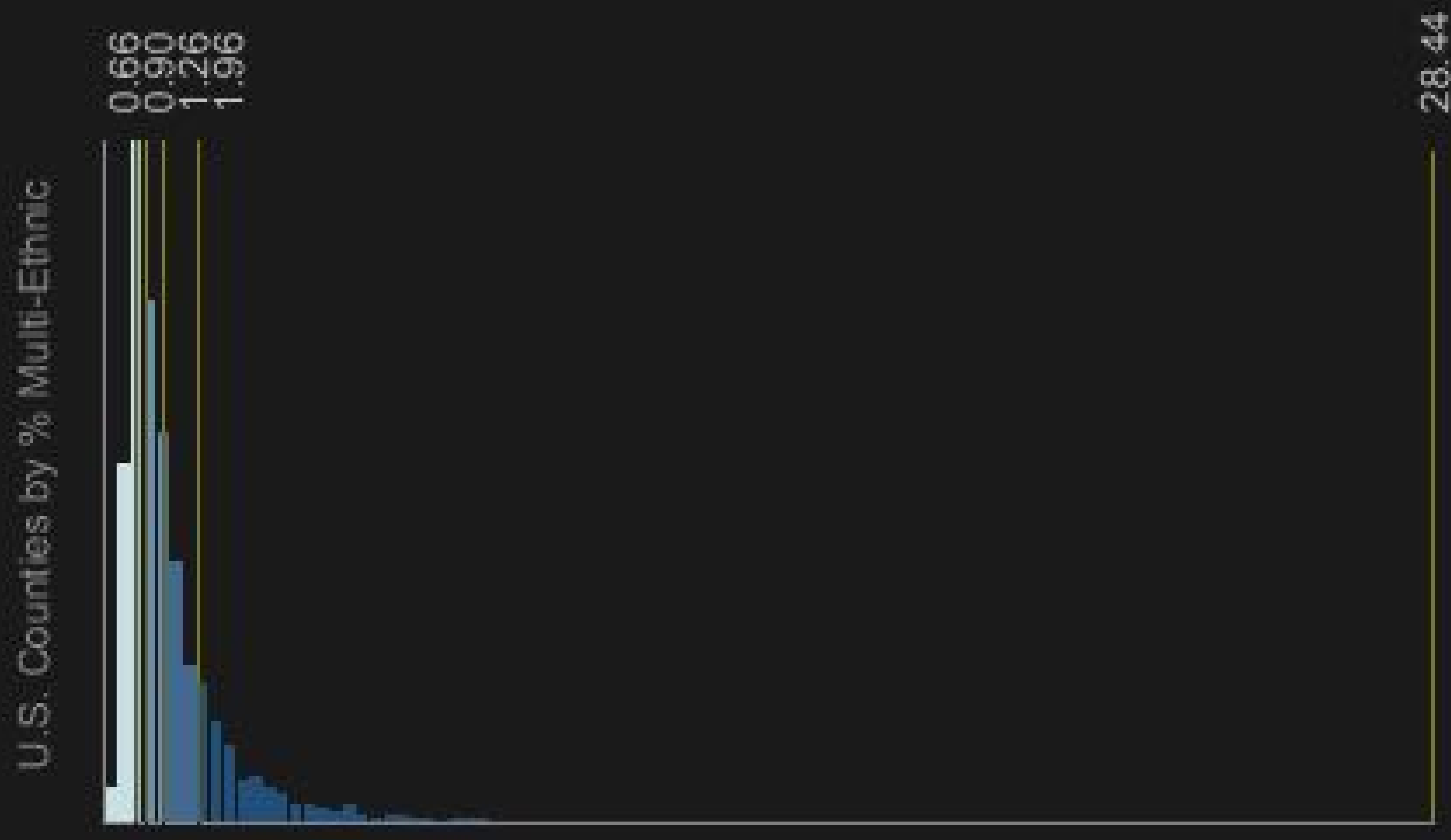
U.S. Census Bureau, 2000

% MULTI-ETHNIC

classification

QUANTILE



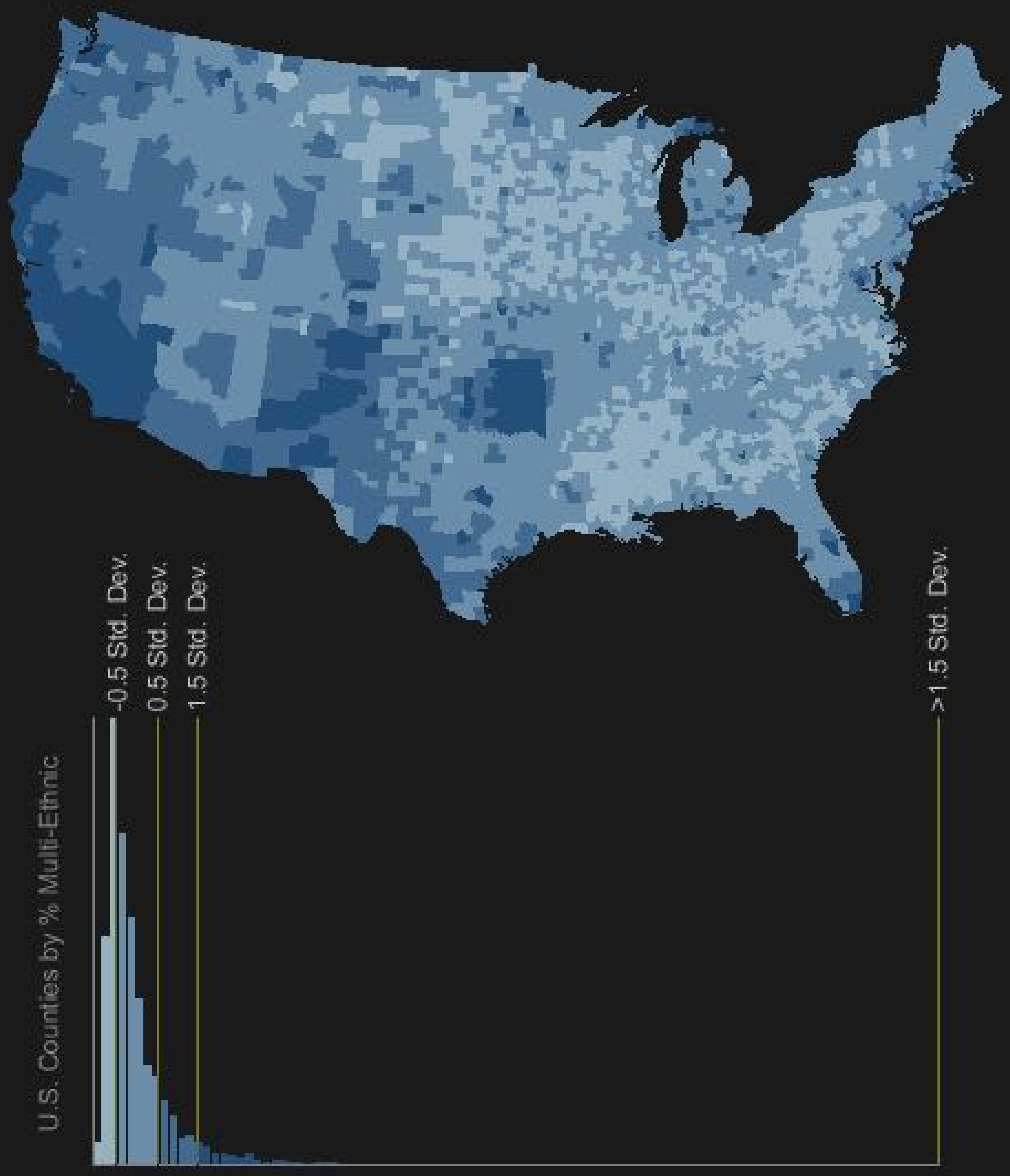


U.S. Census Bureau, 2000

% MULTI-ETHNIC

classification

STD. DEVIATION

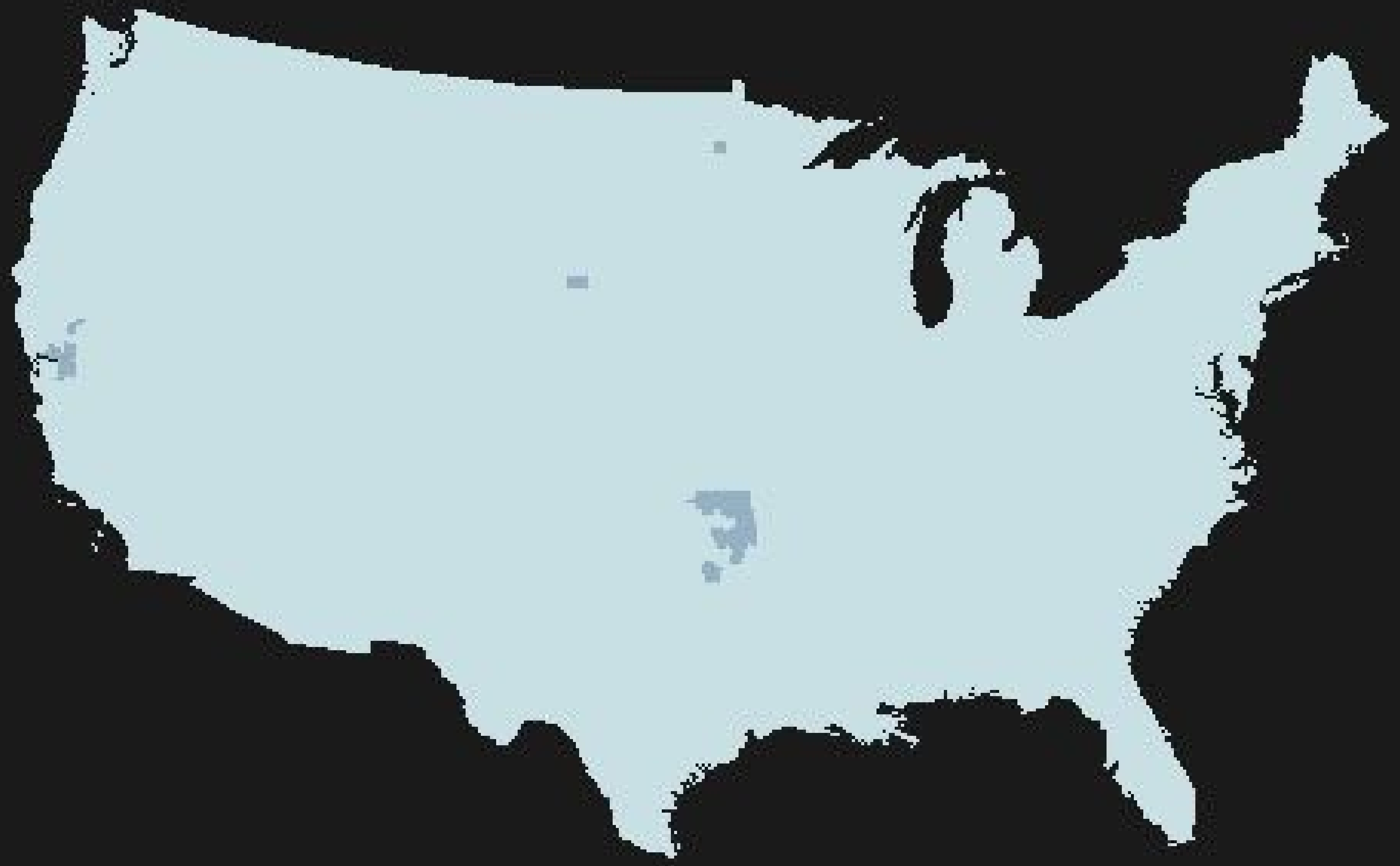


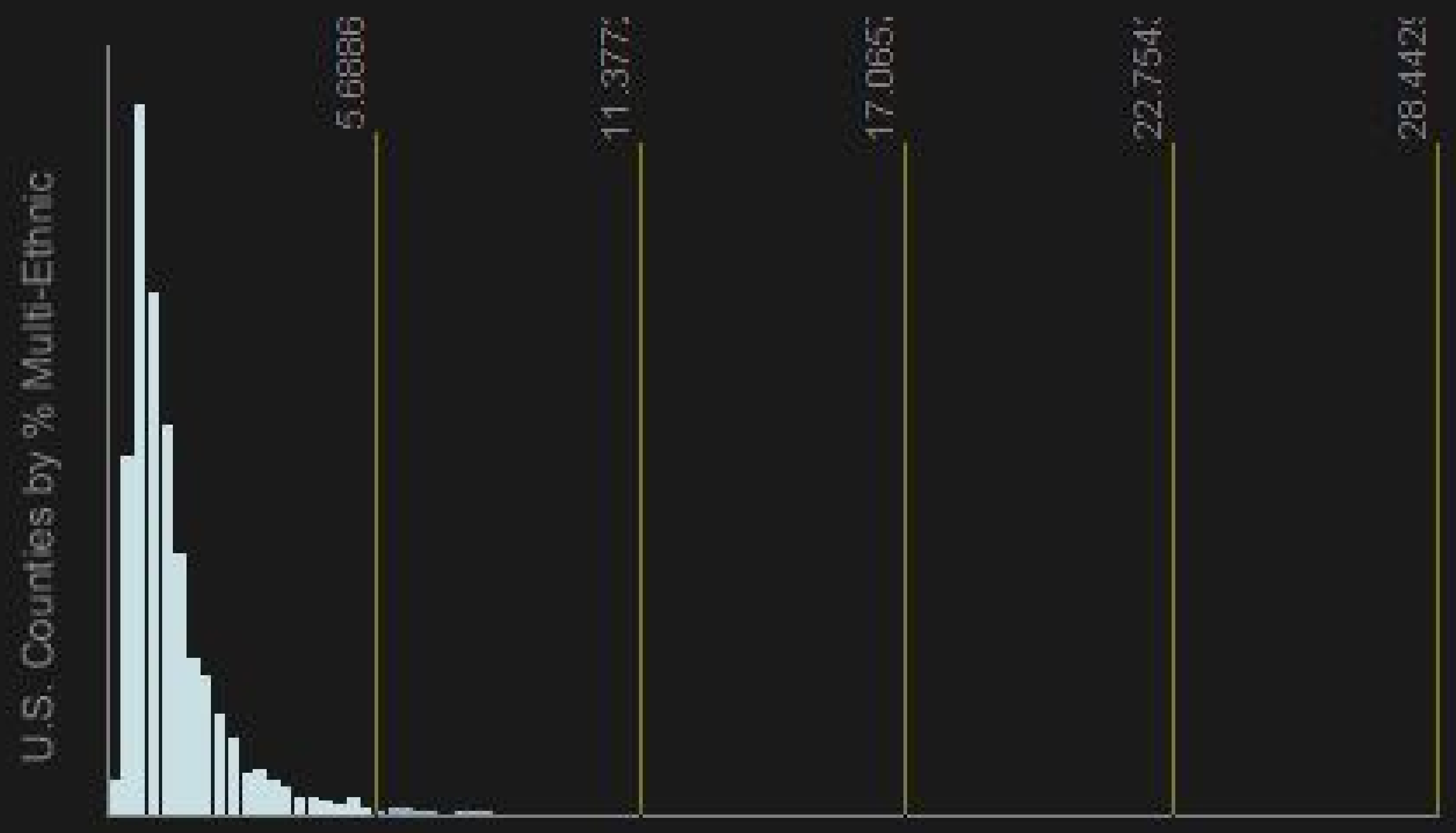
U.S. Census Bureau, 2000

% MULTI-ETHNIC

classification

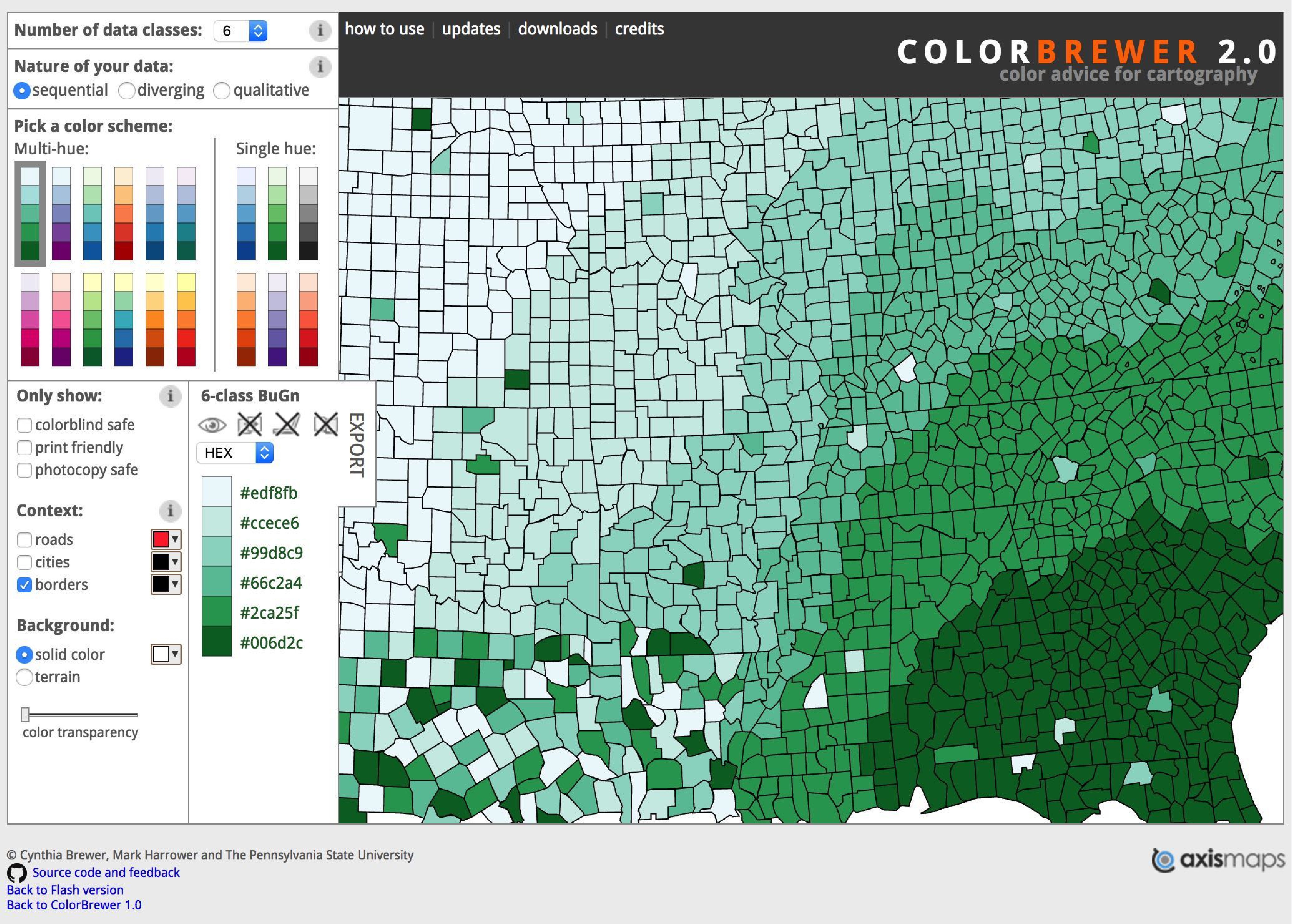
EQUAL INTERVAL





Symbolizing the choropleths

- **Select colors wisely!**
- **Monochrome shading**
  - darker is more!
  - vary density
  - different schemas: Munsell vs Stevens
- **Color shading**
  - hue is not always a good variable, unless bipolar distribution
  - use saturation or intensity



Number of data classes: 6

Nature of your data: ● sequential ○ diverging ○ qualitative

Pick a color scheme:

Multi-hue: Single hue:

Only show:

□ colorblind safe
□ print friendly
□ photocopy safe

Context:

□ roads
□ cities
☑ borders

Background:

● solid color
○ terrain

color transparency

6-class BuGn

EXPORT

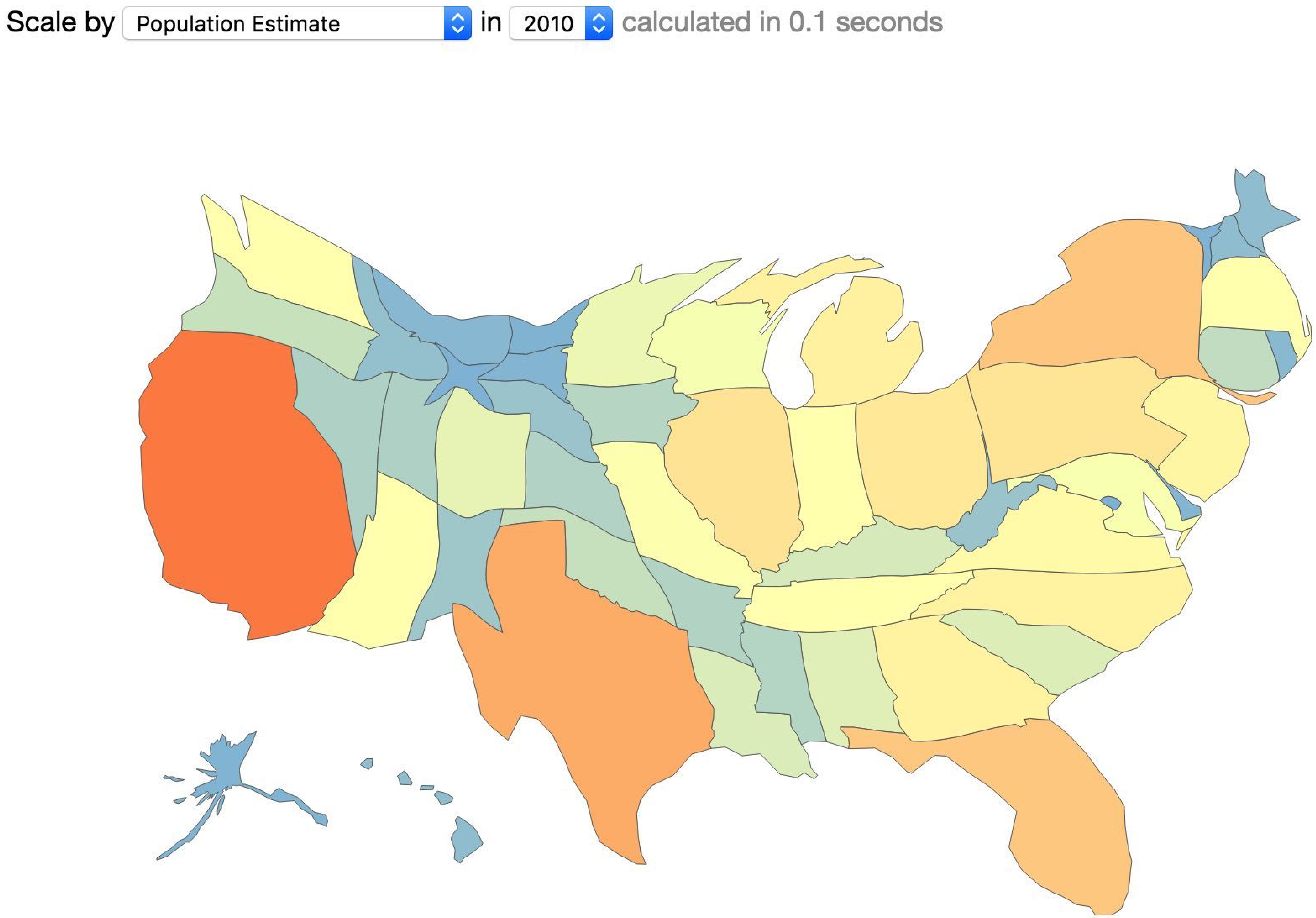
HEX

🟩	#edf8fb
🟩	#ccece6
🟩	#99d8c9
🟩	#66c2a4
🟩	#2ca25f
🟩	#006d2c

how to use | updates | downloads | credits
COLORBREWER 2.0
color advice for cartography
axismaps

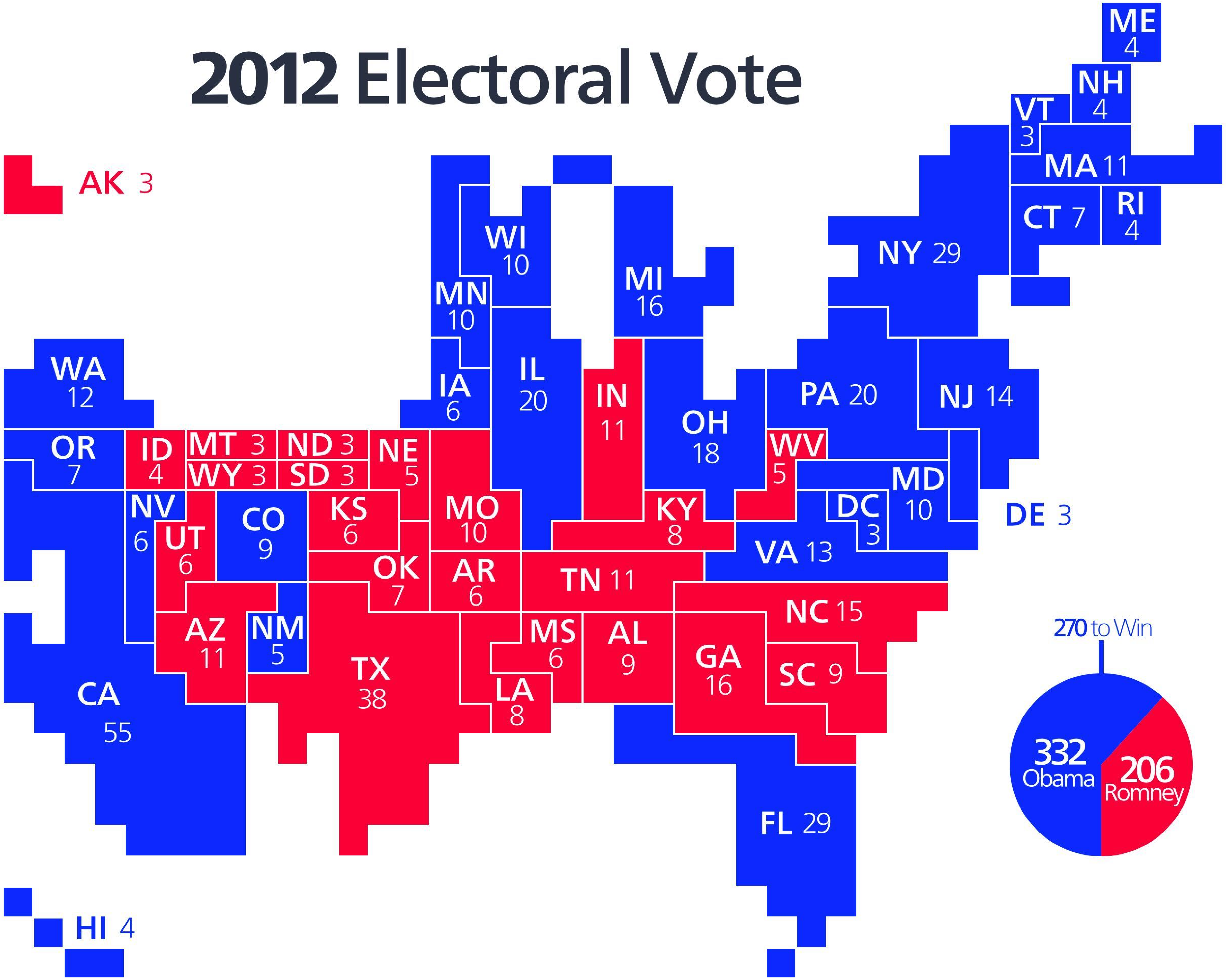
Source code and feedback
Back to Flash version
Back to ColorBrewer 1.0

cartograms



Scale by Population Estimate in 2010 calculated in 0.1 seconds

2012 Electoral Vote



Flow Maps

- **Point pairs (one/two ways and symbol) trajectories**
- **Encoding**
- **Edge between two locations indicates flow between those locations**
  - Width of edge proportional to flow
  - Usually wider end of edge is source of flow
- **Limitations**
  - Can get difficult to compare flows
  - Best flow maps are done by hand

OPTIONS
Show
Colours
Filters
+
−
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

DECK.GL
EXAMPLES
DOCUMENTATION
BLOG
GITHUB
Overview
CORE LAYERS
LineLayer
HexagonLayer
IconLayer
GeoJsonLayer
ScreenGridLayer
ArcLayer
ScatterplotLayer
CUSTOM LAYERS
Brushing Layer
Trip Routes
Wind Map
BEYOND MAPS
3D Surface Explorer
3D Indoor Scan

Flights To And From London Heathrow Airport

Flight paths in a 6-hour window

From 08:32:43 GMT to 14:32:43 GMT on March 28th, 2017

Flight path data source: The OpenSky Network
Airport location data source: Natural Earth

NO. OF LINE SEGMENTS

141.3K

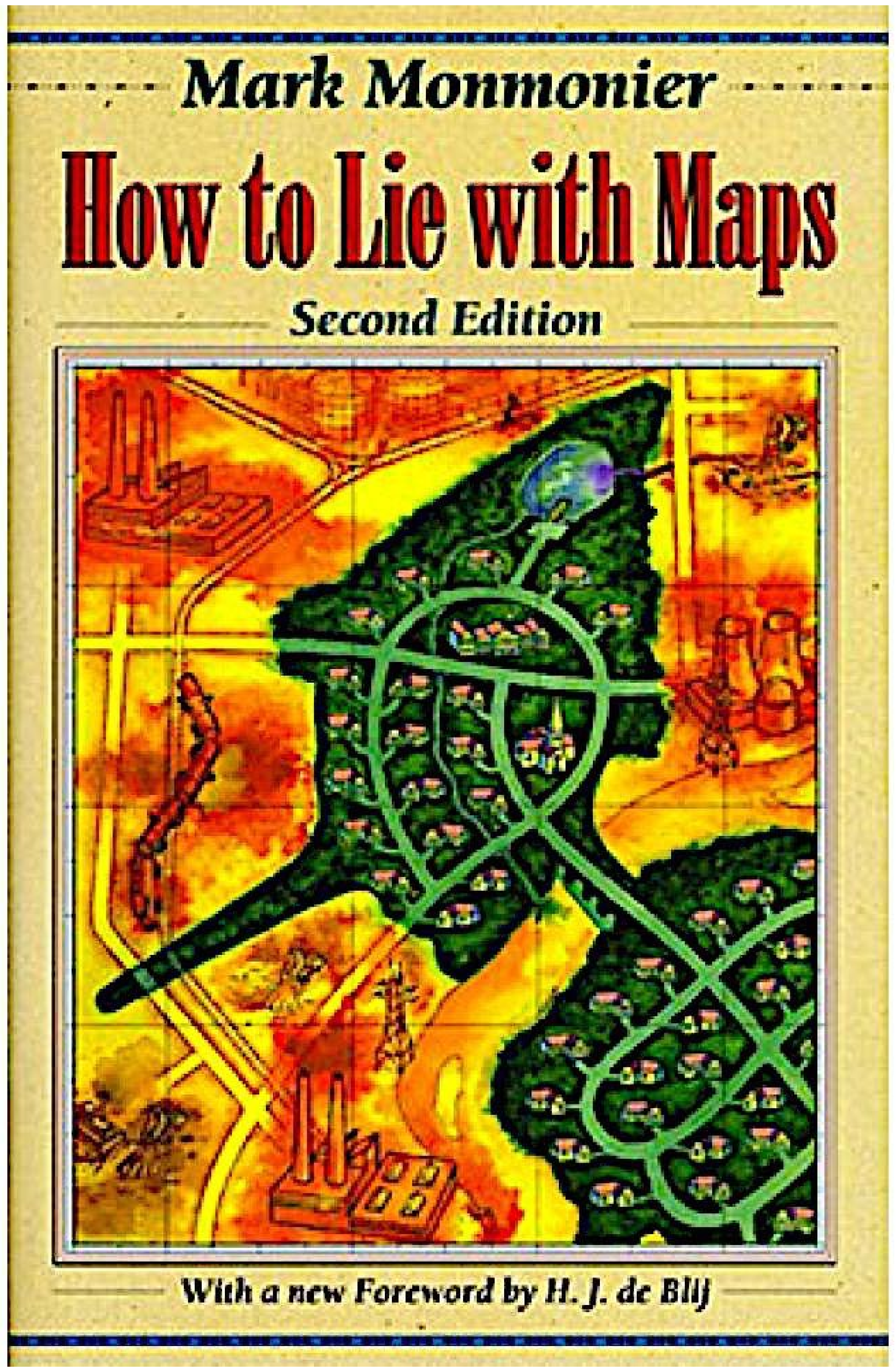
Stroke Width

View Code ↗
Hold down shift to rotate
© Mapbox © OpenStreetMap Improve this map

pitfalls

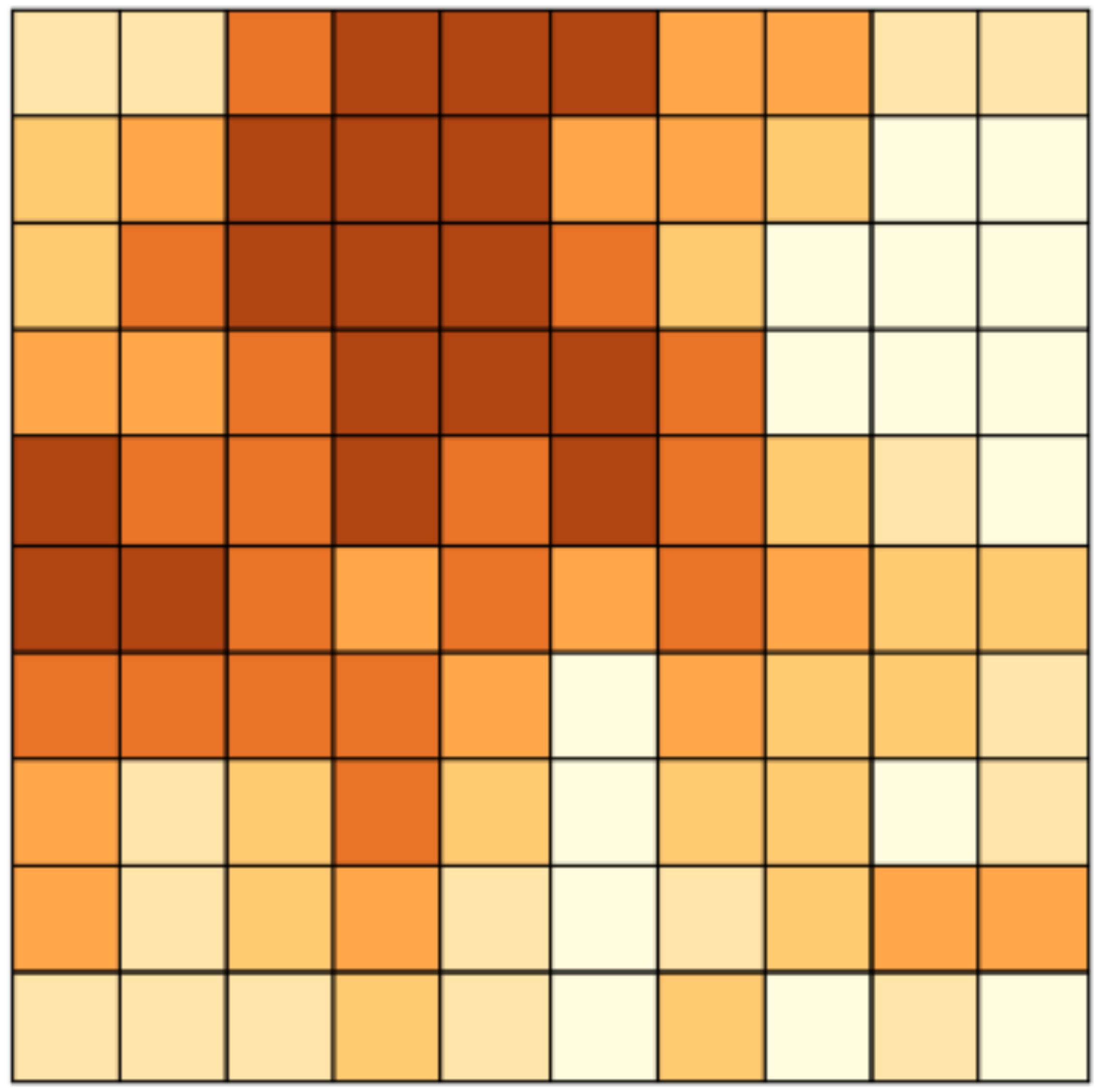
How to lie with maps

- Visual inspection is not enough
- Visual inspection sometimes could lead to wrong conclusions
- We must test rigorously using spatial analysis methods.



visual experiment
apophenia

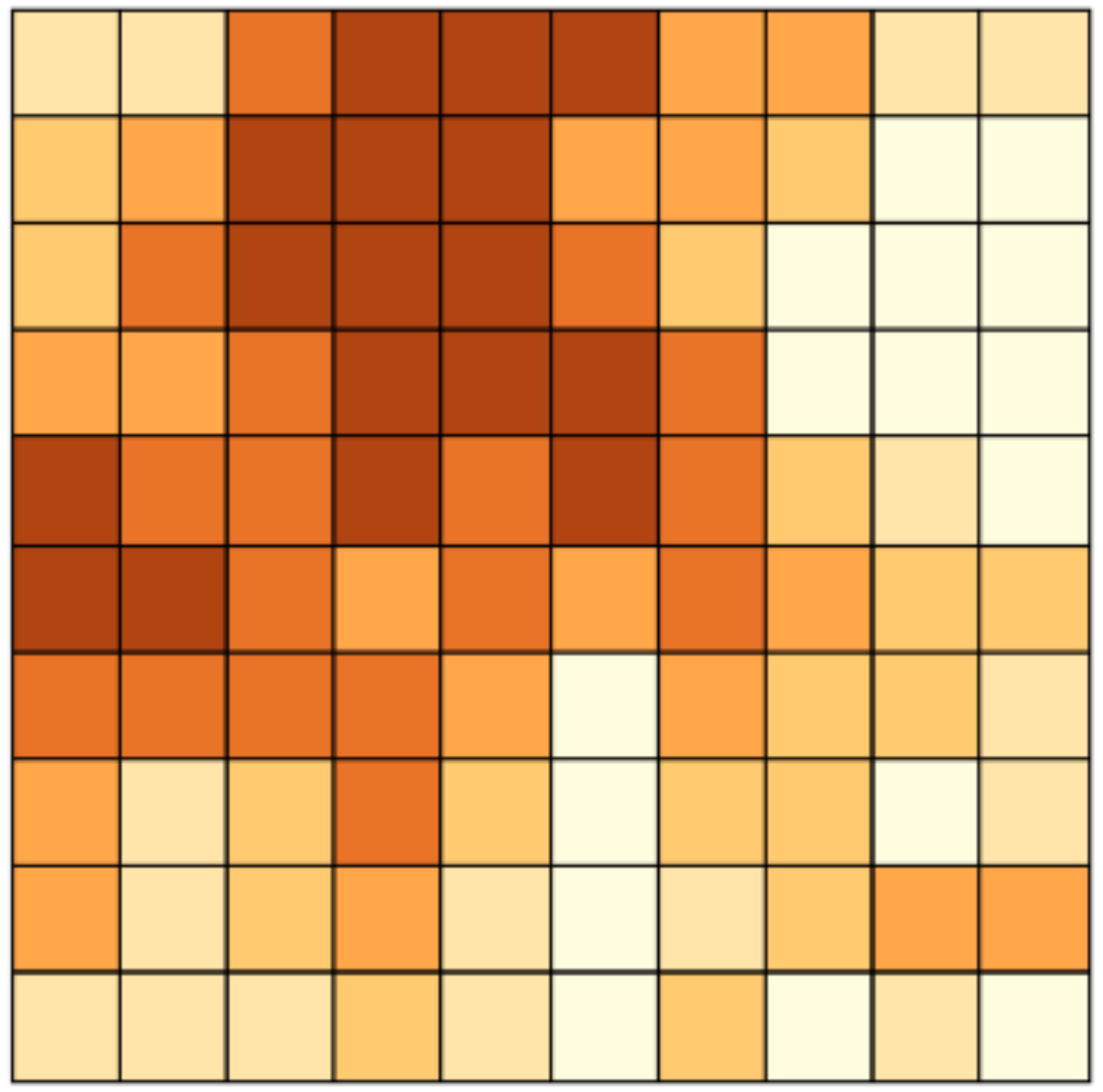
A B C



random pattern?

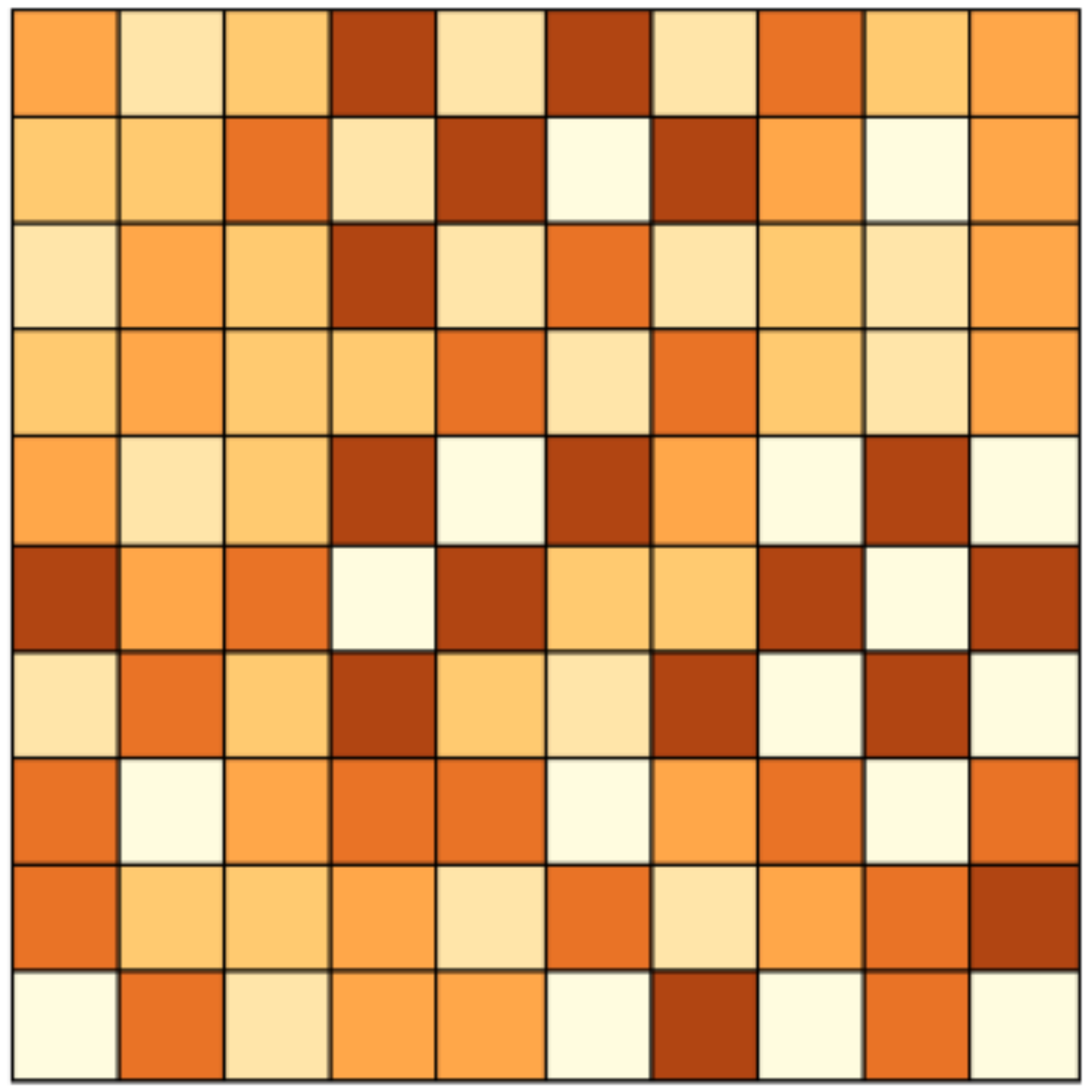
visual experiment
apophenia

A



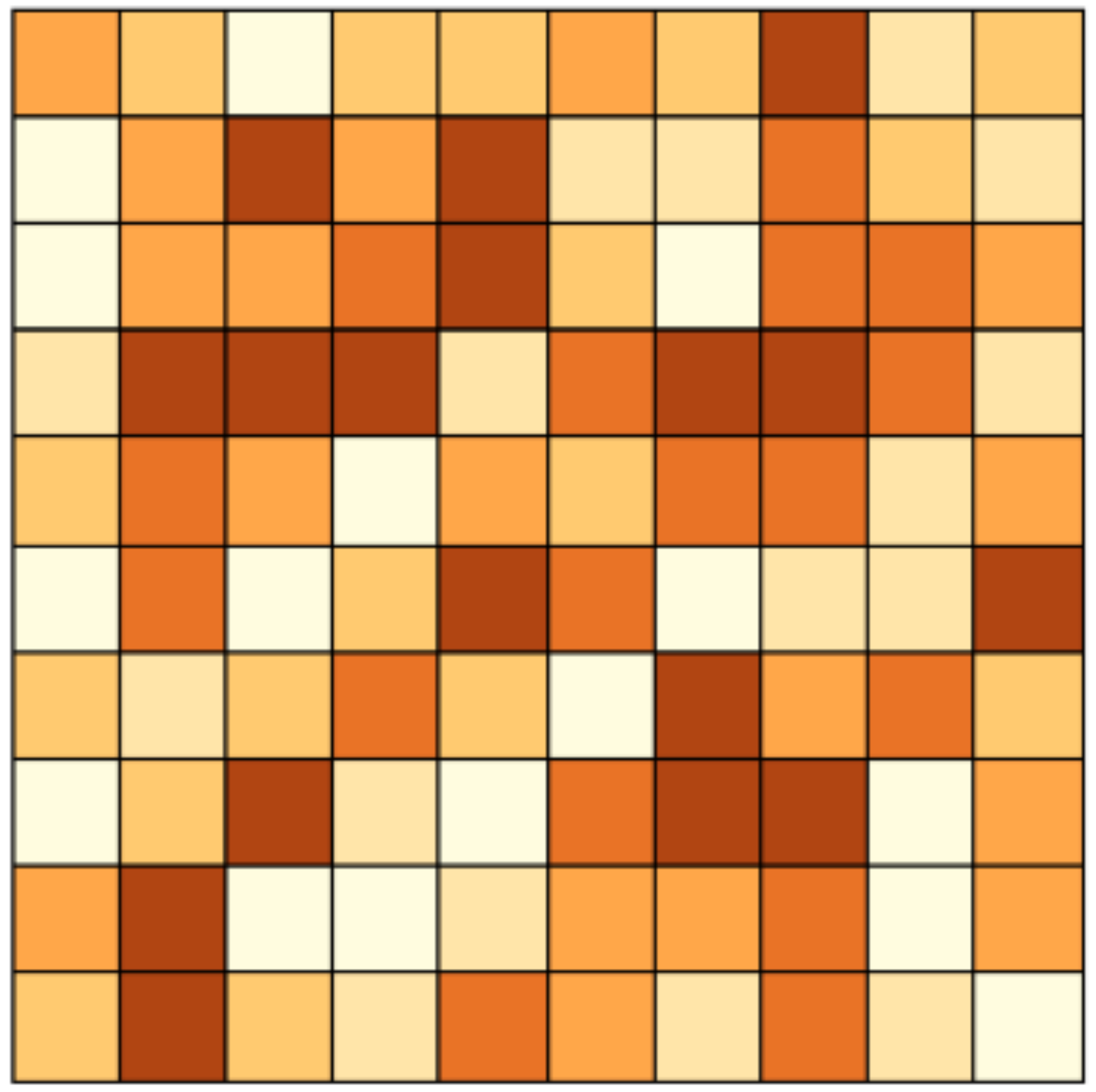
clustered

B



dispersed

C



random

Patternicity
Michael Shermer (2008)

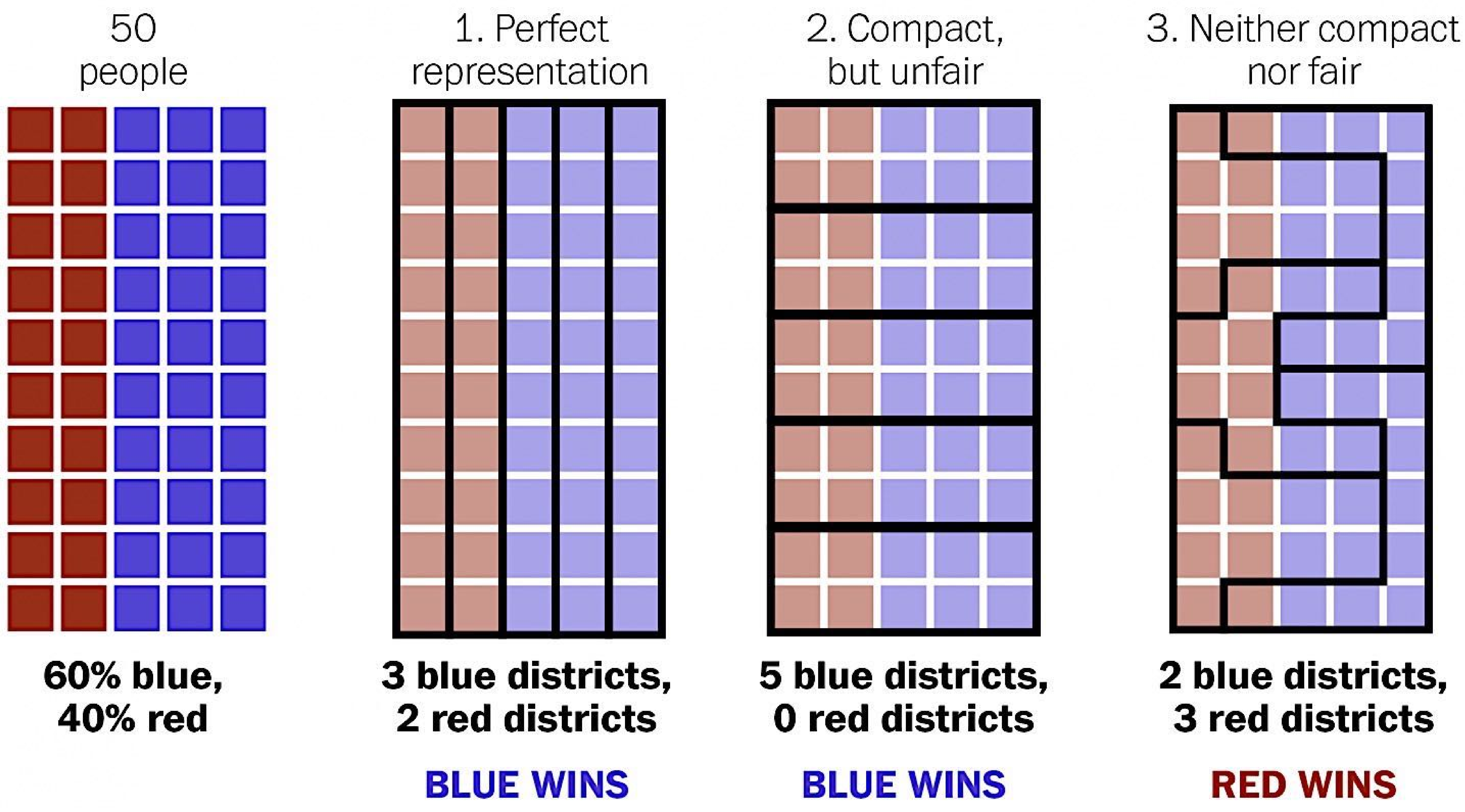
- **The tendency to find meaningful patterns in both meaningful and meaningless noise**
  - Type I error (false positive)
  - Type II error (false negative)
- **Humans are pattern-seeking primates and this behavior is hardly-coded in how our brain works**
- **Related to survival skills**
- **https://www.ted.com/talks/michael_shermer_the_pattern_behind_self_deception**

MAUP
Modifiable Area Unit Problem

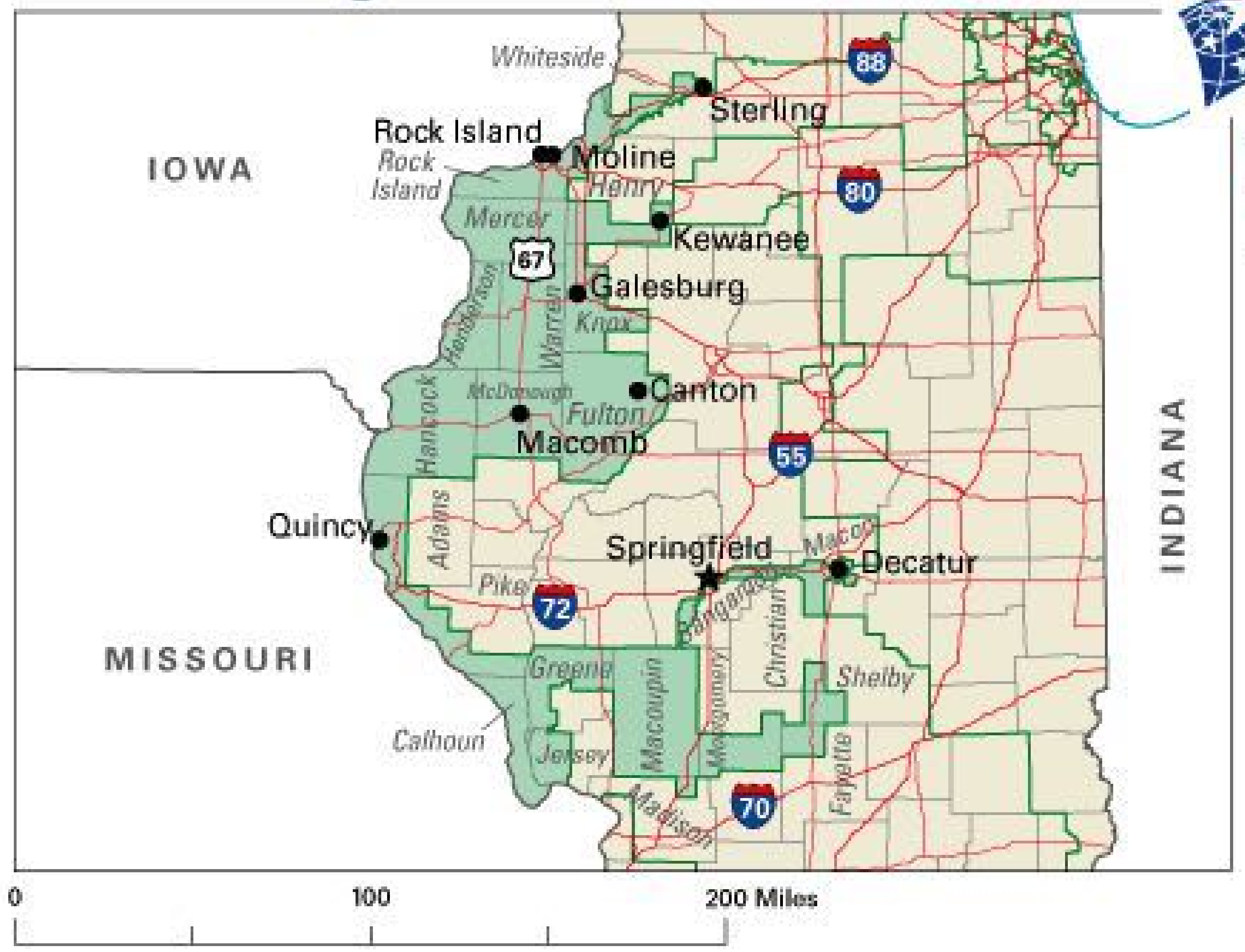
- **The same basic data yield different results when aggregated in different ways**
  - Nice read: “A million or so correlation coefficient: three experiments on the modifiable area unit problem” (Openshaw and Taylor, 1979)
- **Zonal effect**
  - Similar size and number of units, but different boundaries
    - Zip codes versus census tracts, postal zones versus city neighborhoods
- **Scale effect**
  - Increases size and decreases number of units
    - US counties versus states
  - Global model might be inconsistent with local models
- **The take home message is that how we aggregate the input units will impact the values of the output units**

a first real example
gerrymandering

- In the process of setting electoral districts, intended to establish a political advantage for a particular party or group by manipulating district boundaries


50 people
1. Perfect representation
2. Compact, but unfair
3. Neither compact nor fair
60% blue, 40% red
3 blue districts, 2 red districts
BLUE WINS
5 blue districts, 0 red districts
BLUE WINS
2 blue districts, 3 red districts
RED WINS

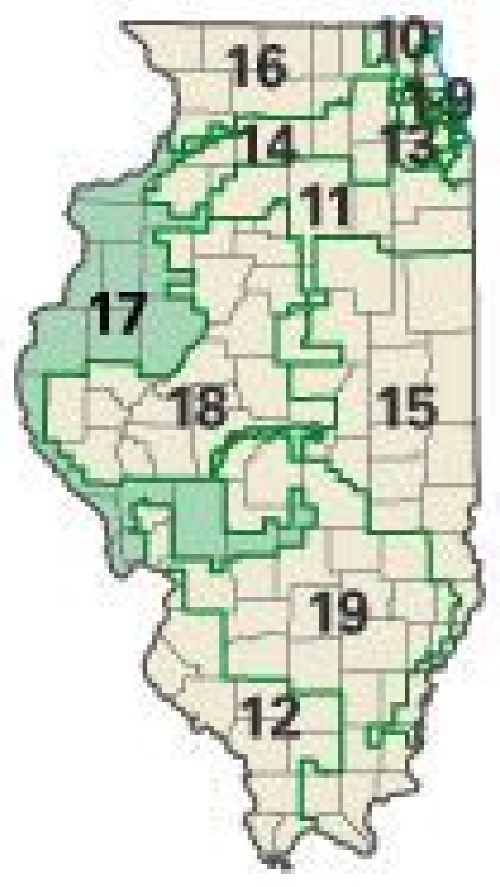
Congressional District 17



nationalatlas.gov™

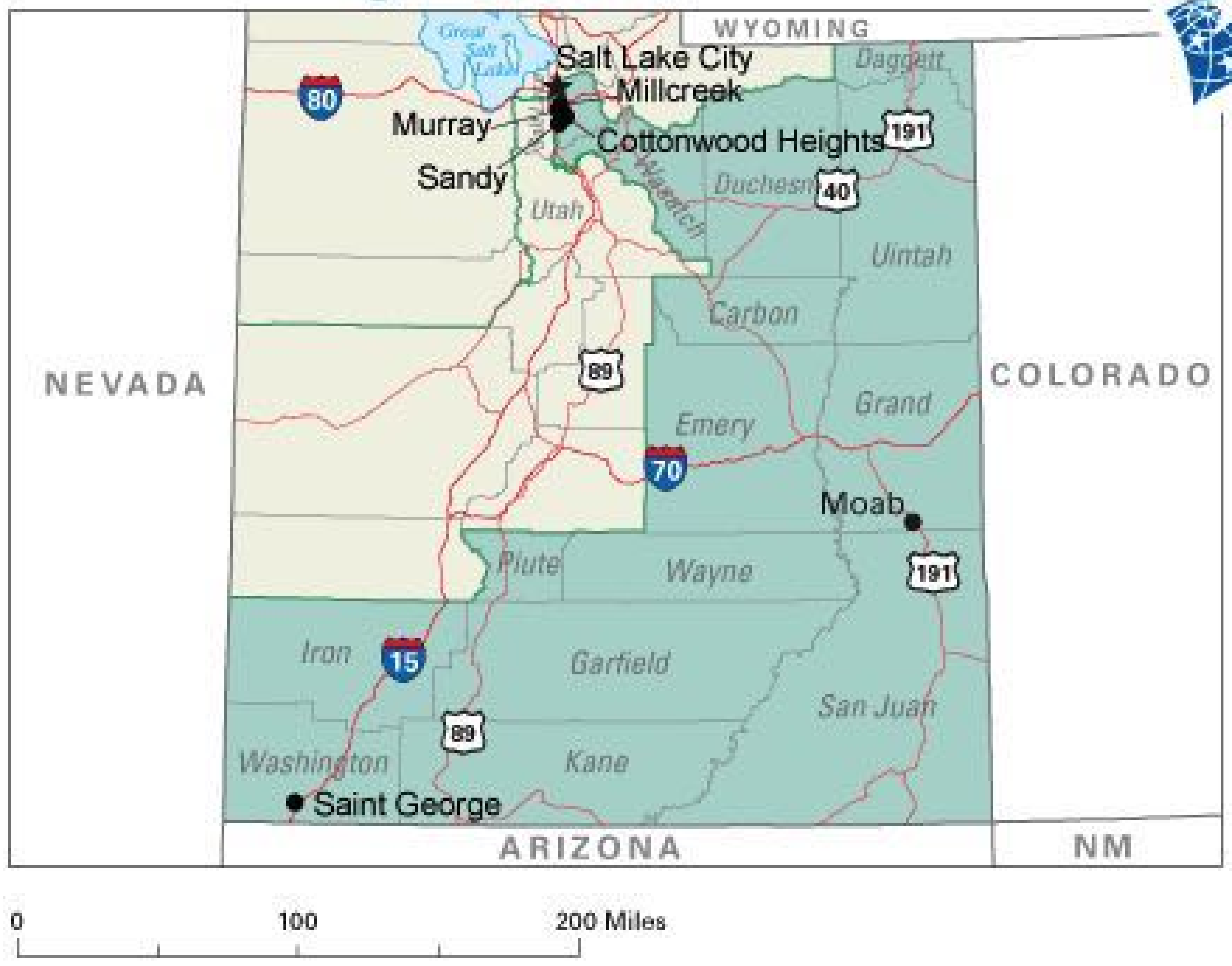
🟩 17 Congressional District

Fulton County



Illinois (19 Districts)

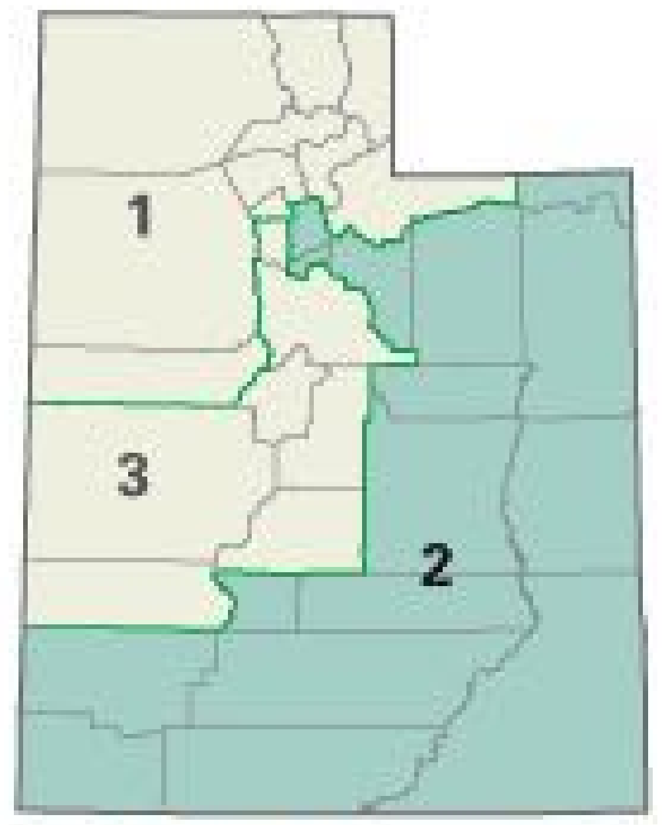
Congressional District 2



nationalatlas.gov™

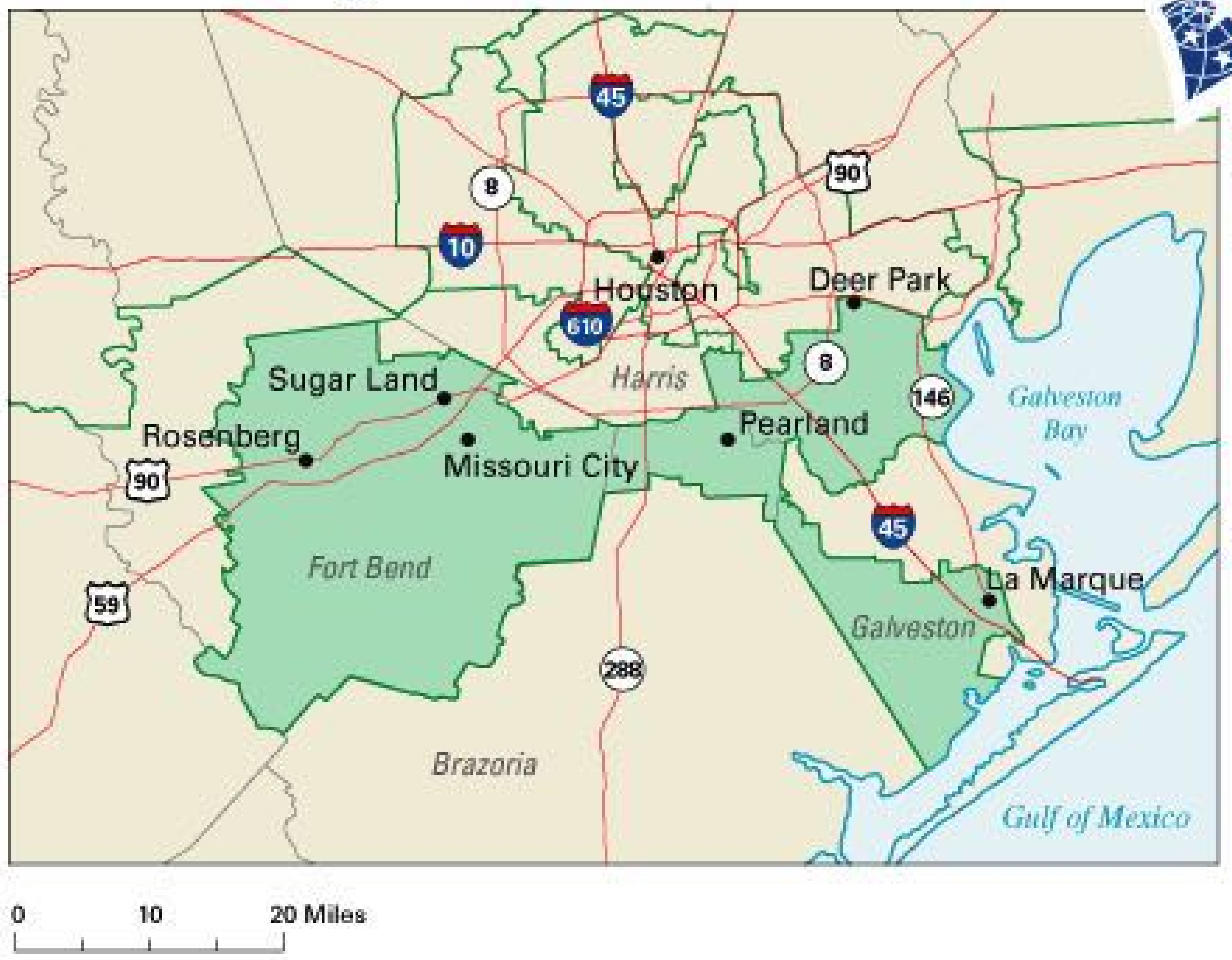
🟩 2 Congressional District

Grand County



Utah (3 Districts)

Congressional District 22



nationalatlas.gov™

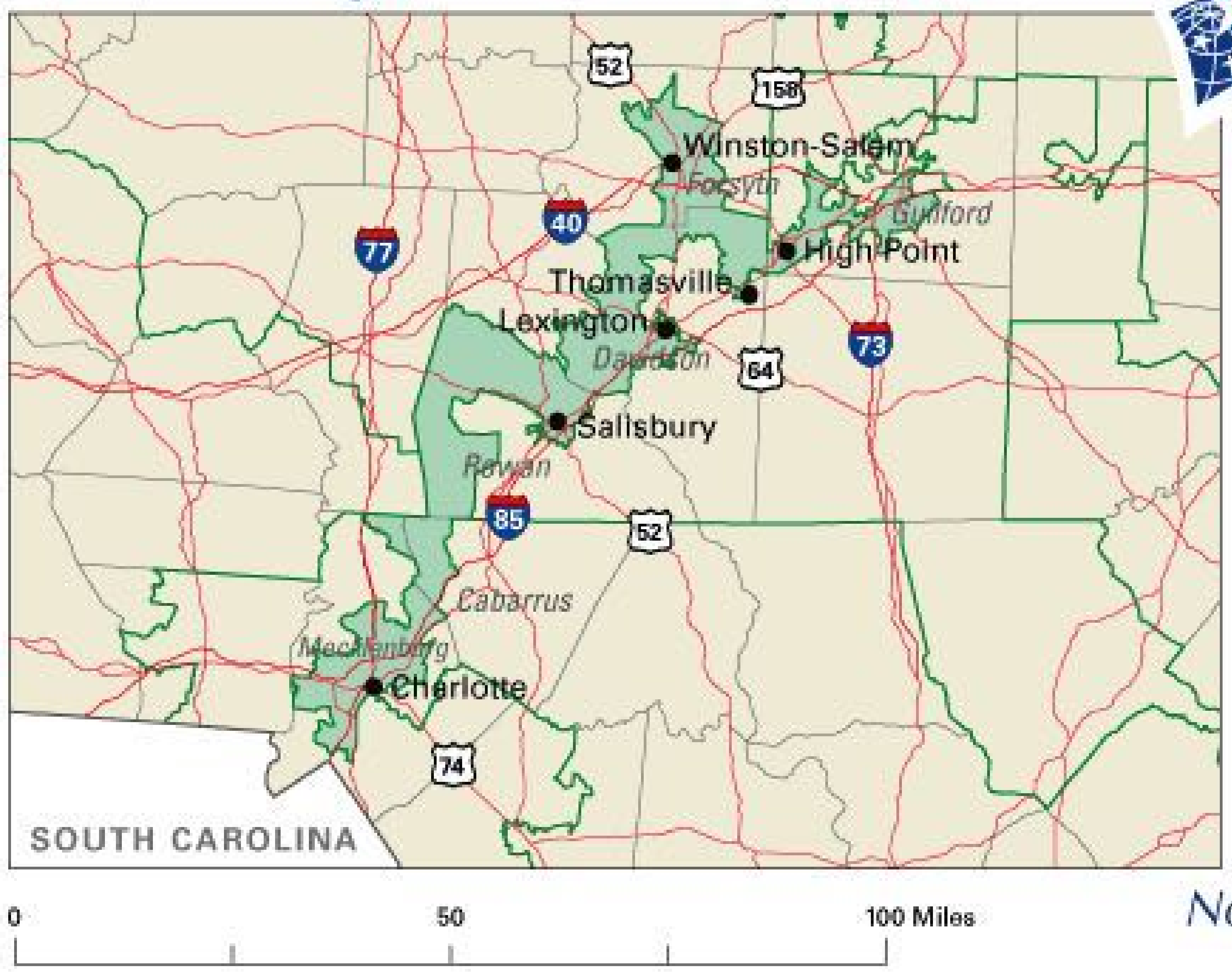
🟩 22 Congressional District

Harris County

3, 24, 26
30, 32
13
19
4
16
11
23
21
10
28
7, 9,
18, 29
25
15
27
14
2
22
12
17
5
6
1
31
20
8

Texas (32 Districts)

Congressional District 12



nationalatlas.gov™

🟩 12 Congressional District

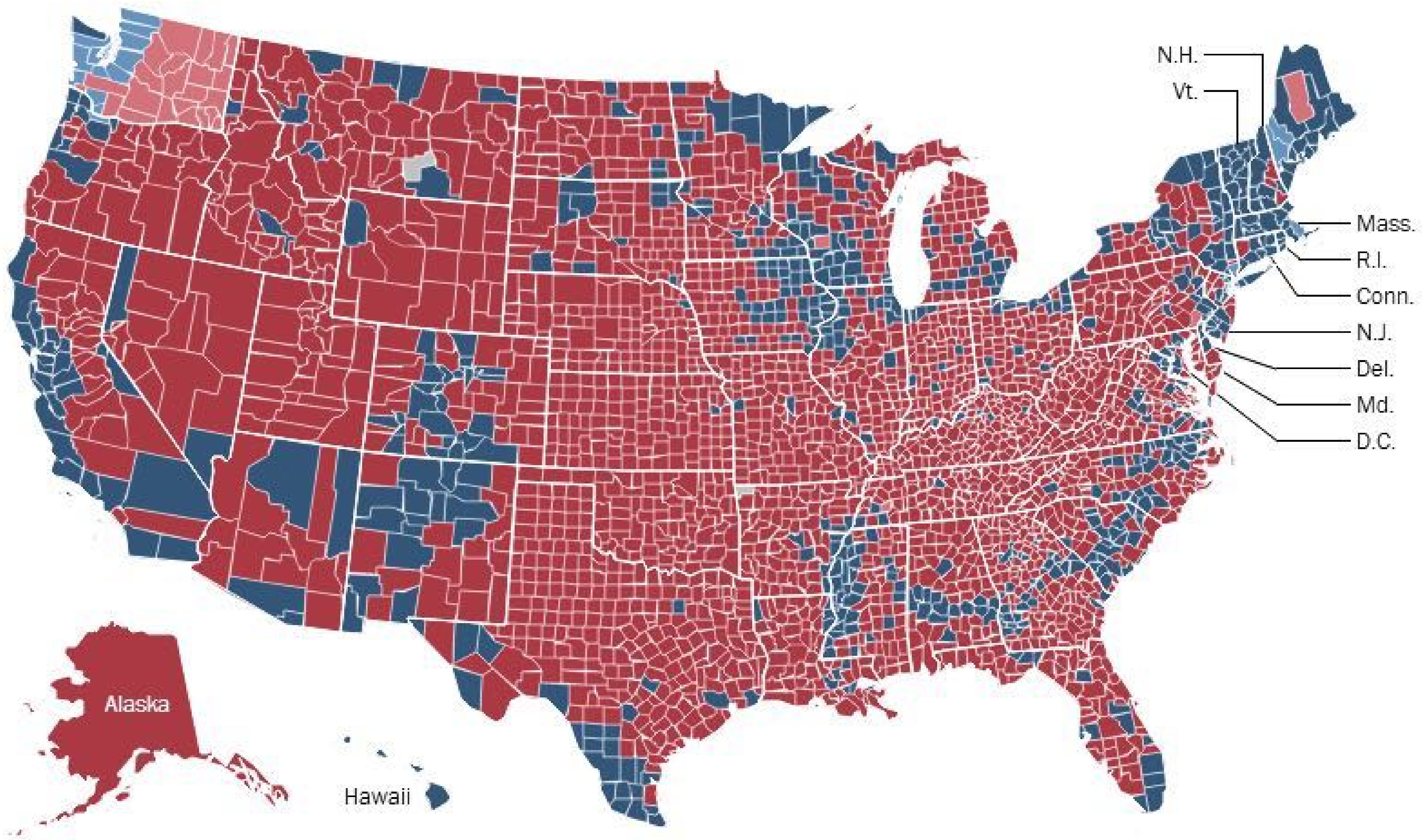
Rowan County

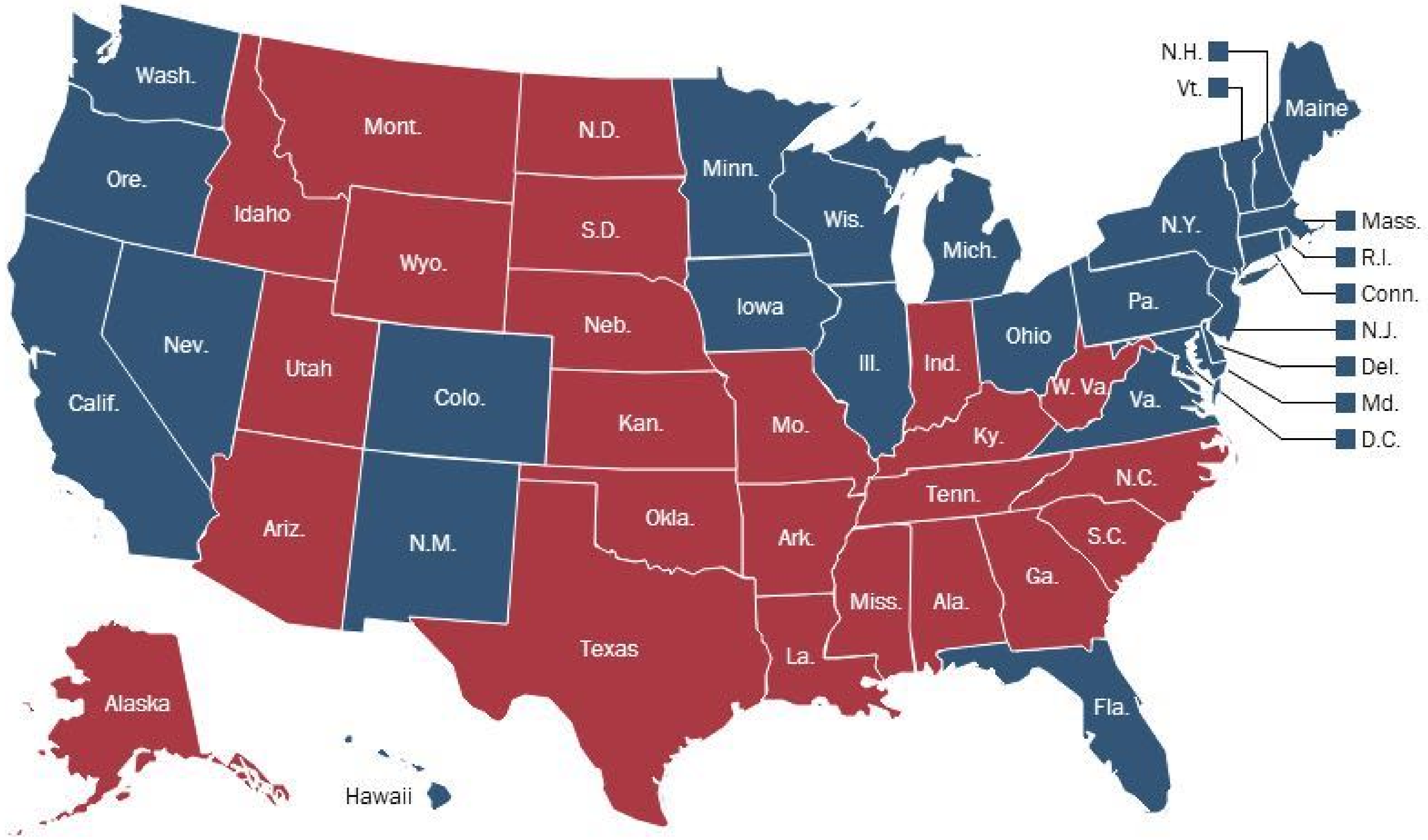
12
13
1
5
4
6
11
10
2
9
8
3
7

North Carolina (13 Districts)

US election 2012

counties versus states





Ecological Fallacy

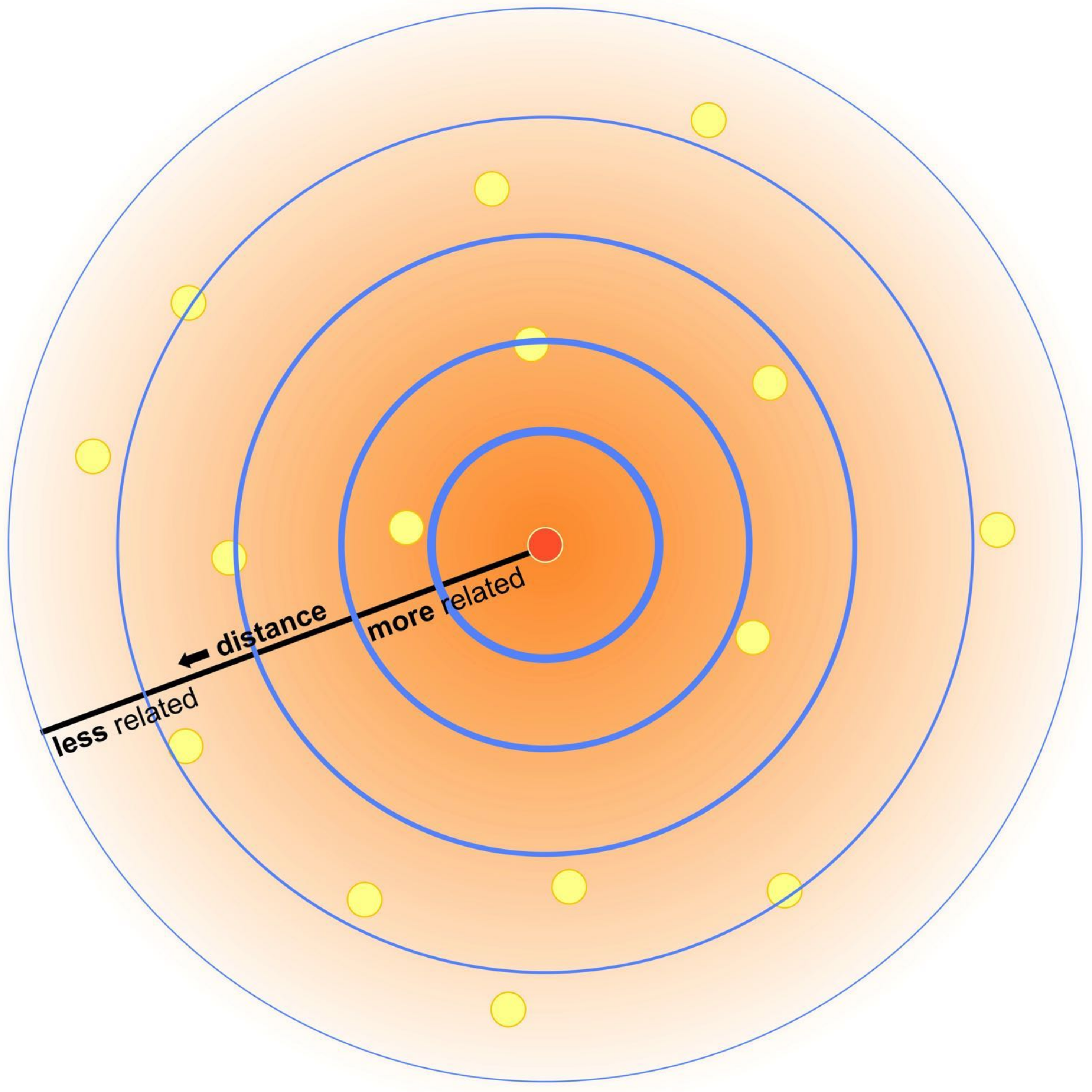
- **Individual behavior cannot be explained at the aggregate level**
- **Issue of interpretation**
  - e.g., county homicide rates do not explain individual criminal behavior
  - model aggregate dependent variables with aggregate explanatory variables
  - alternative: multilevel modeling

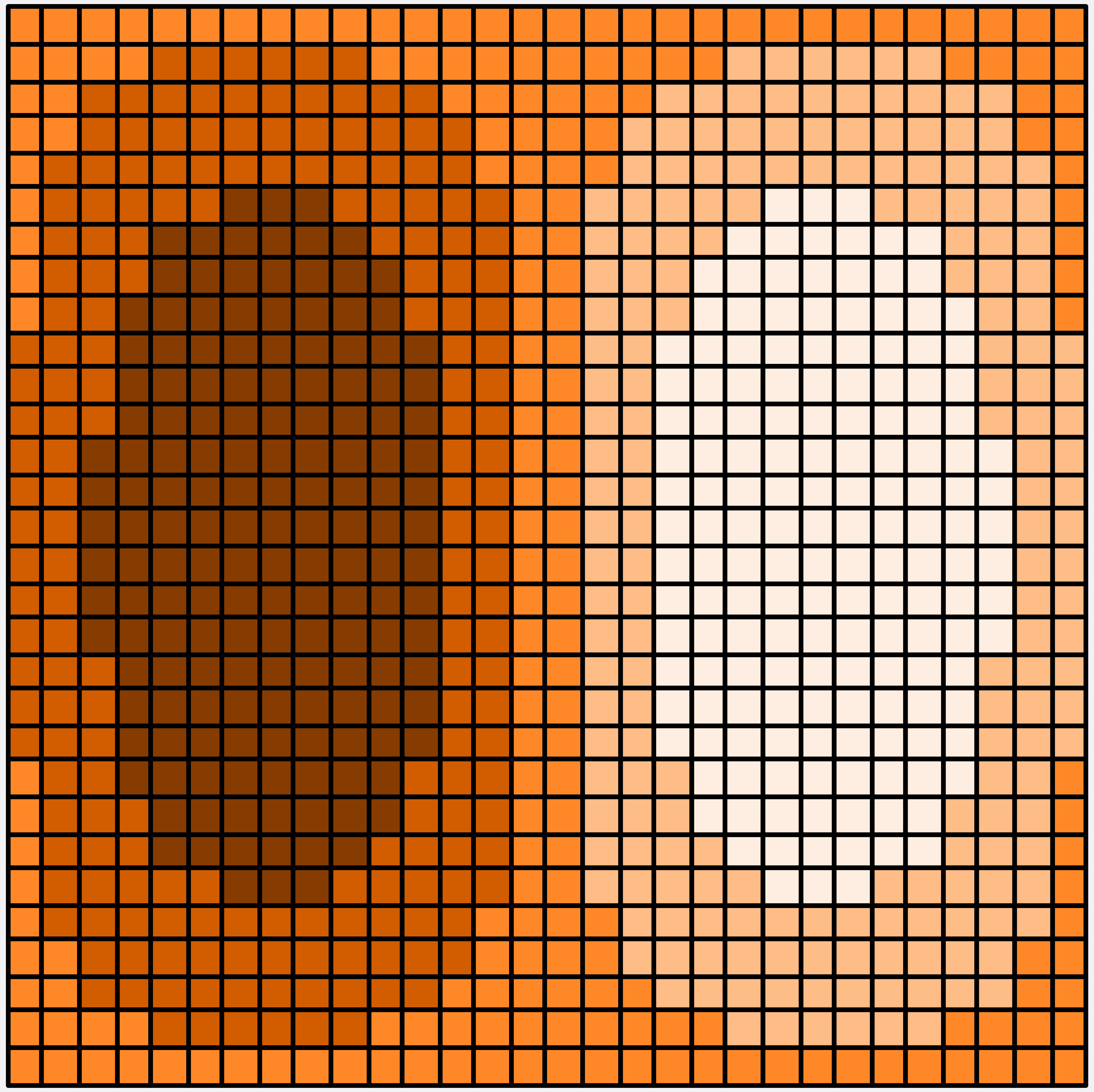
COSP
Change of Support Problem

- **Variables measured at different spatial scales**
- **Spatial misalignment**
  - we collected the data on one scale, but need to make inferences on a different scale.
    - How do we change from one spatial scale to another?
  - have different spatial datasets that come to us on different spatial scales.
    - How do we combine data sources?
- **Aggregate up to a common scale (the finest possible)**

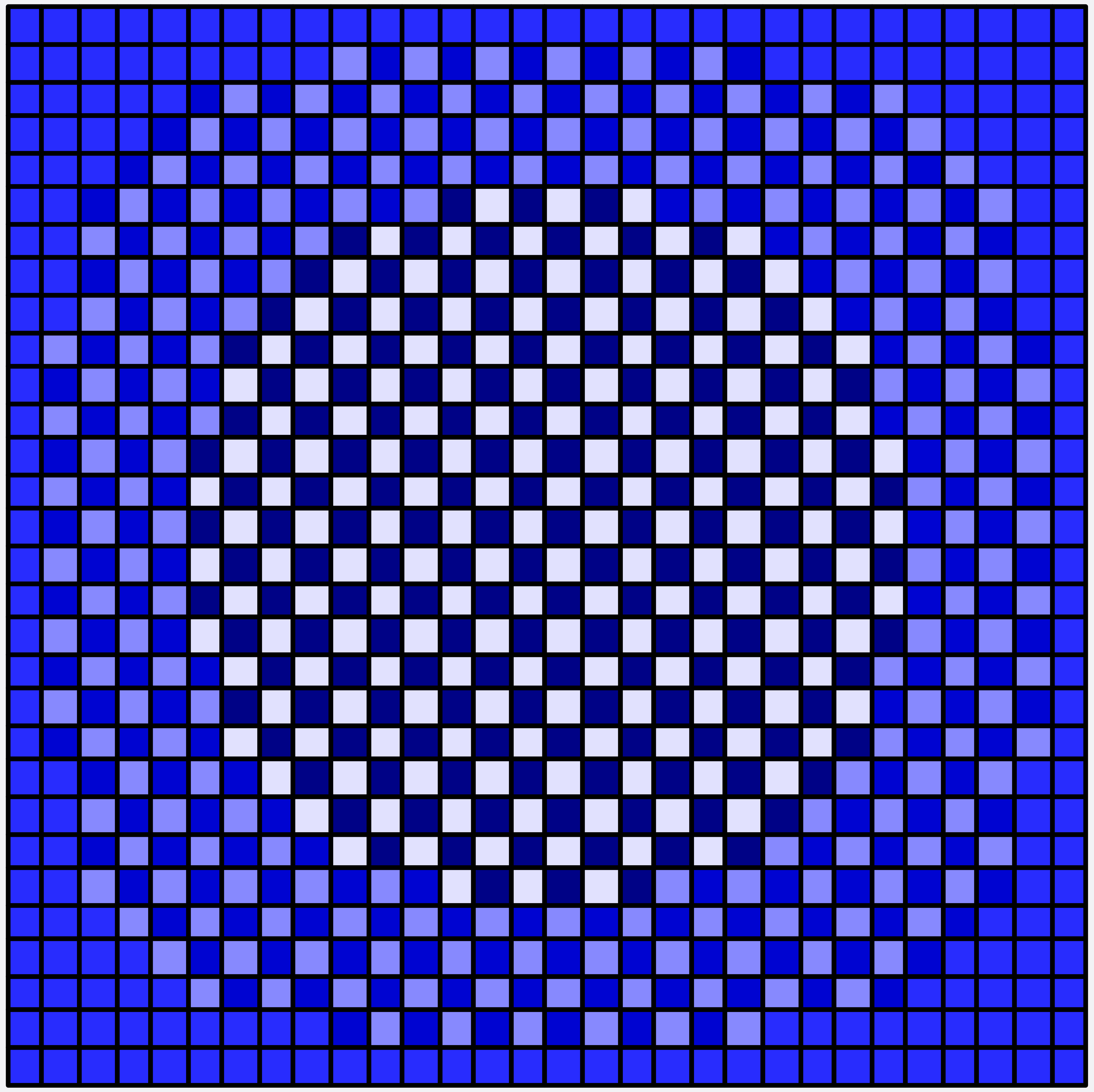
other critical issues

- **Spatial autocorrelation**
  - **Measures the correlation of a variable with itself through space**
  - Related to Tobler's first law of geography
    - Everything is related to everything else, but near things are more related than distant things.





positive = clustered



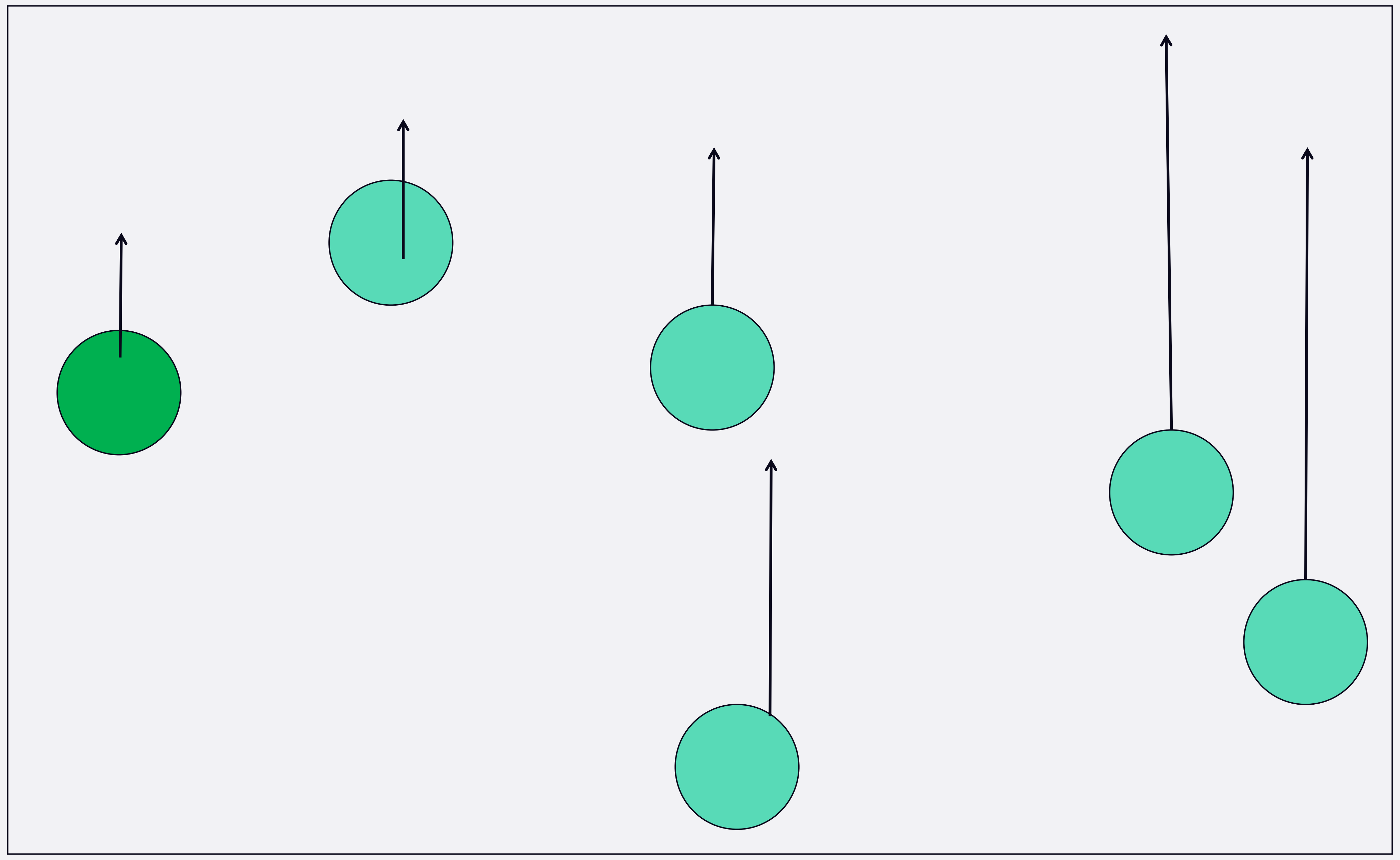
negative = dispersed

why is spatial autocorrelation important?

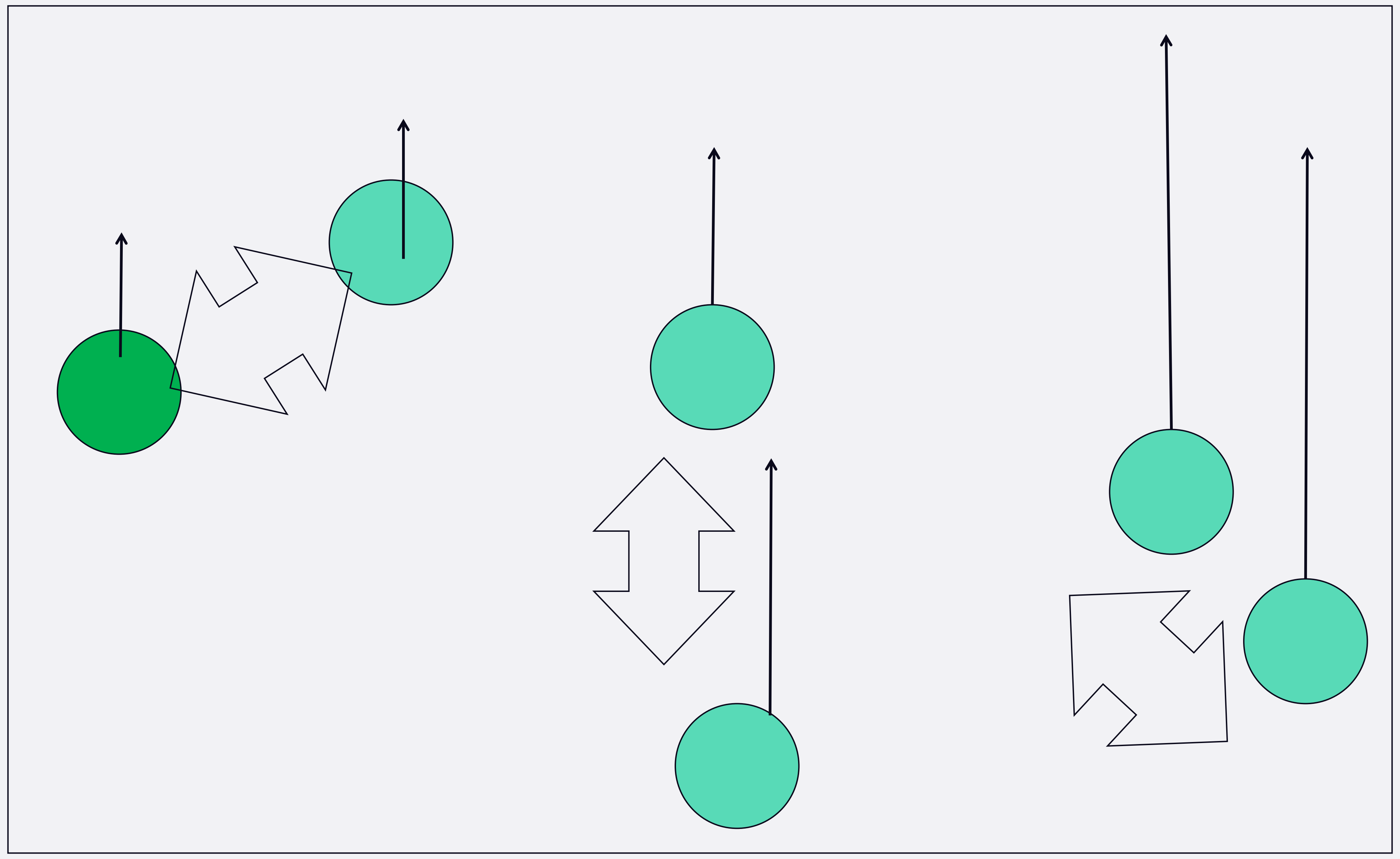
- **It implies the existence of a spatial process**
  - Why are near-by areas similar to each other?
  - Why do high income people live close each other?
  - These are geographical questions.
    - They are about location
- **It invalidates most traditional statistical inference tests**
  - If spatial autocorrelation exists, the results of standard statistical inference tests may be incorrect
  - We need to use spatial statistical inference tests
- **For example**
  - You are more likely to incorrectly conclude a relationship exists when it does not
  - You believe that the relationship is stronger than it really is

Why are standard statistical tests wrong?

- **Statistical tests are based on the assumption that the values of observations in each sample are independent of one another**
- **spatial autocorrelation violates this**
    - samples taken from nearby areas are related to each other and are not independent



Values near each other are similar in magnitude.



Implies a relationship between nearby observations

QUESTIONS?