

Visualization for Data Science

Representing Spatial Data

Geographic Maps



Project Updates

PM2 due October 31st at 6pm. Now moved to November 7th at 6pm.

- EDA analysis

- One for each member of the team. Jupyter Notebook
- In your report state what each person EDA covers, this falls under general

Team Member	Numerical Attributes	Temporal Attributes	Ordinal Attributes	Nominal	Type of Analysis
Frodo Baggins	(2) Cost, Emotion,	Age of Middle Earth	Feet size		Univariate on all mentioned. Bivariate on (feet size and cost)
Gandalf	(1) Length of beard,	Age of Universe		(2) Color of beard, Weapons	

- Then for each person in your report you should include summaries of what you learned from your EDA. Feel free to include figures from your jupyter notebook. Note that we don't expect code in your report.

- INQUIRY FCOUS

- Theme, questions, and analytic tasks – SEE UPDATED MILESTONE 2

Adminstrivia

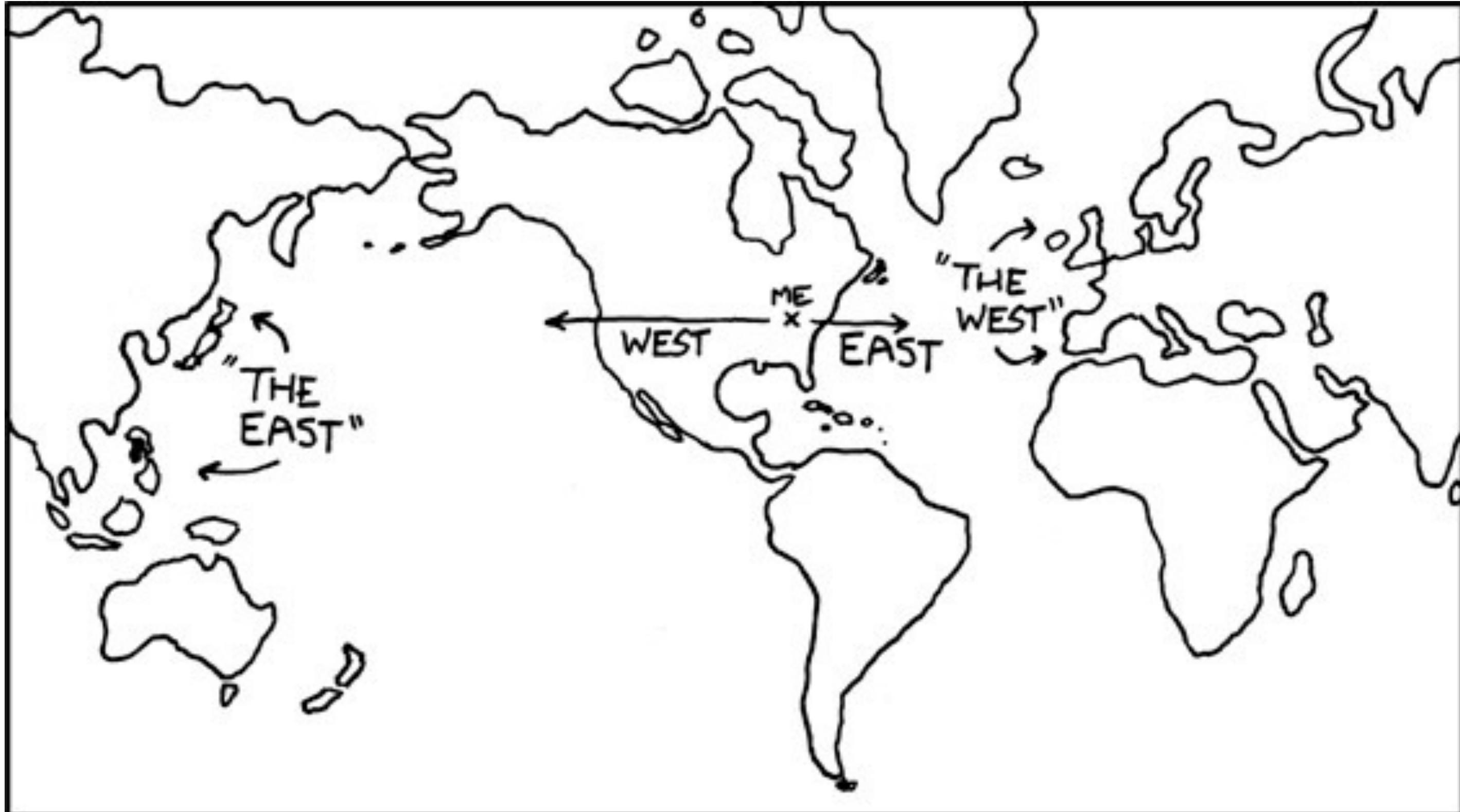
Final Exam is set by the University

- December 10th at 3:30pm in CBTF for most students
- CFA

For those who typically write with CBTF, you have received a Ed Chat message from me. Please respond.

Tutorial 10 (Maps) and Tutorial 11 (Color) have been posted

- Work through them and complete the associated exercises.



THIS ALWAYS BUGGED ME.

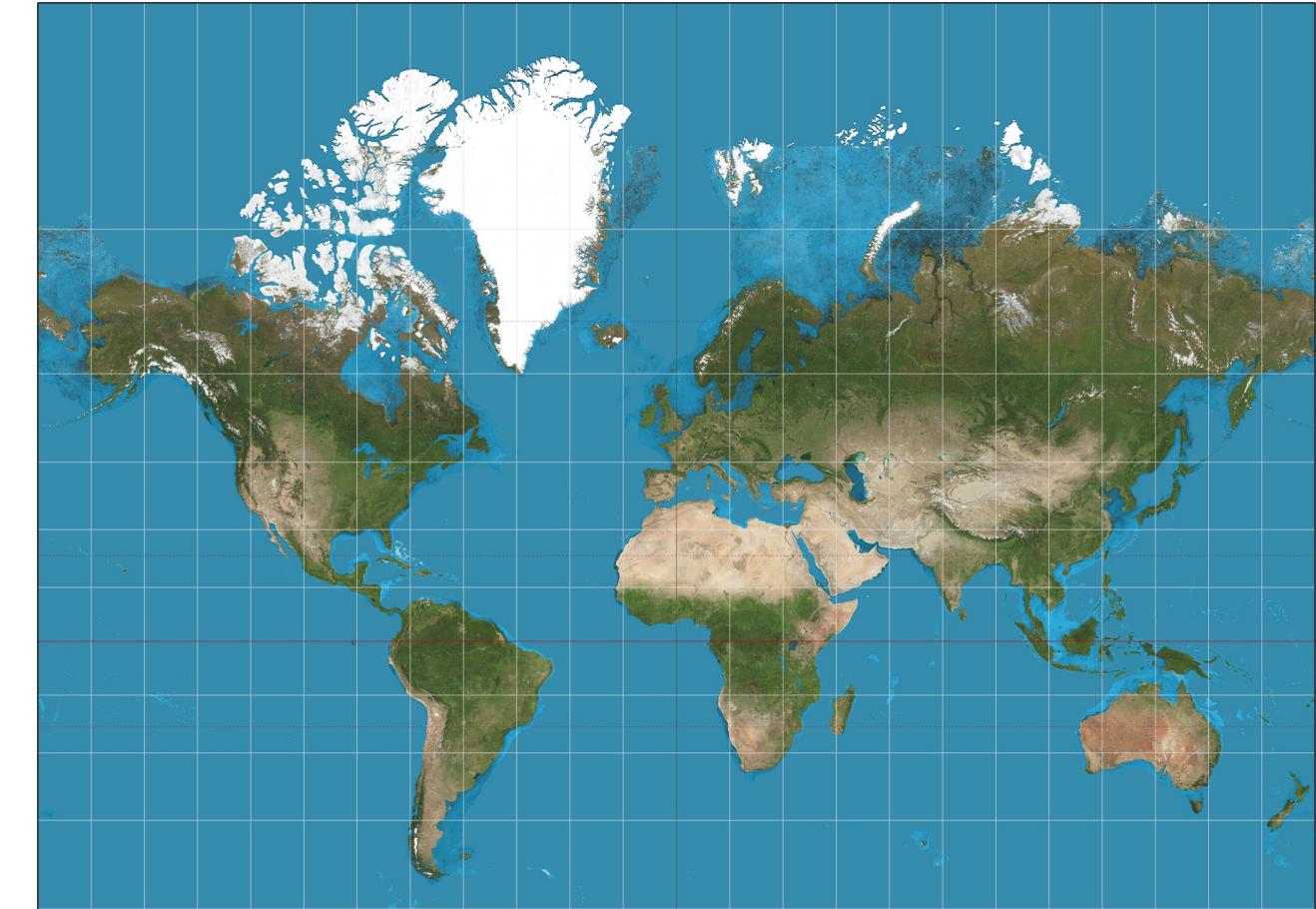
Learning Outcomes

- List some common map projections, describe their characteristics and understand the (dis)advantages of each.
- Describe the different map types and what types of tasks they are used for
- Describe the map visualization pipeline and explain how different errors can occur in each stage.
- Learn what elements visually encode data in maps.
- Critique map visualizations
- Select the visualization technique best suited for a given task

Clicker Question

If Greenland and Africa were compared by size, which would be true?

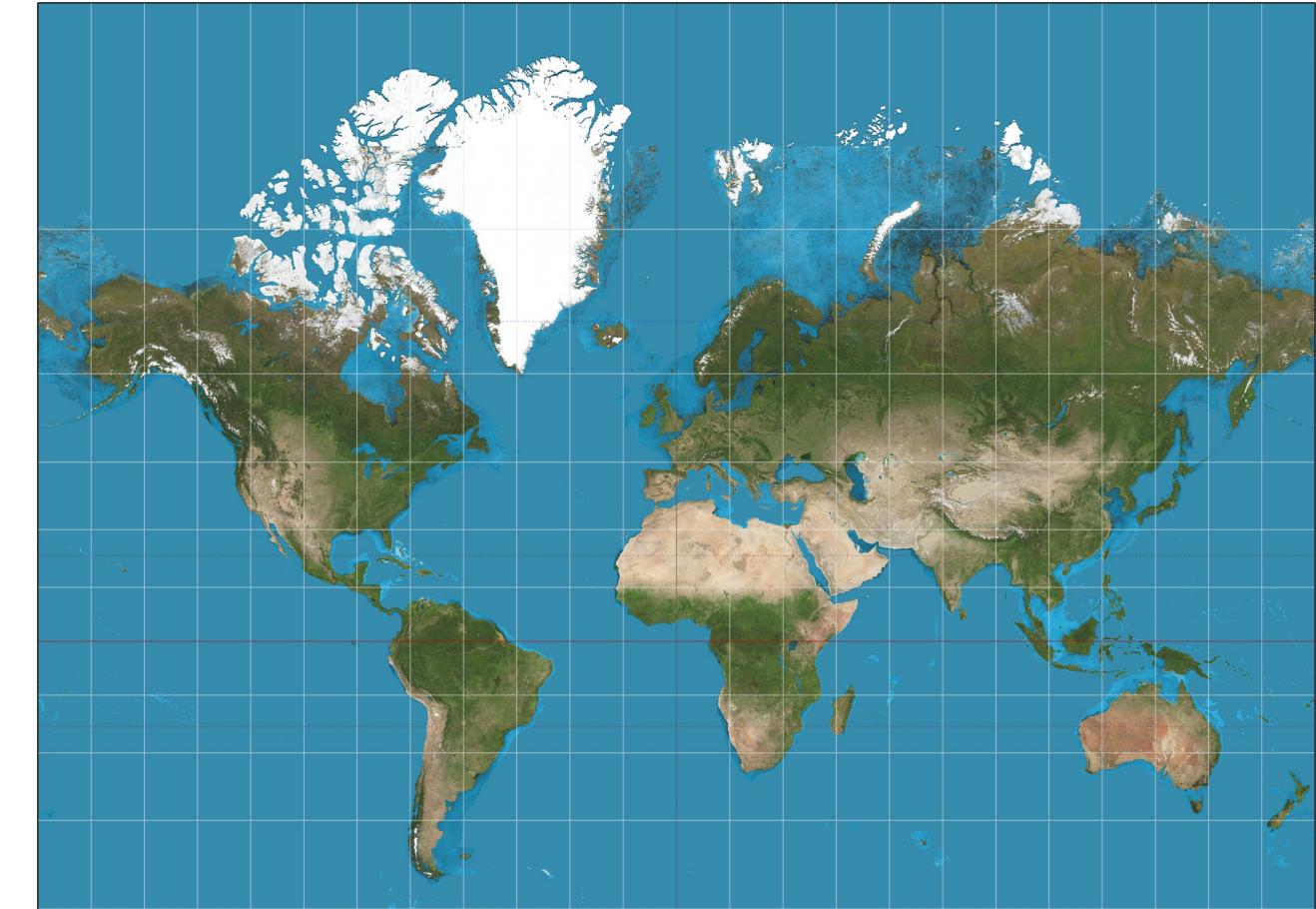
- A) Greenland is nearly twice the size of Africa.
- B) Africa is significantly larger than Greenland.
- C) Greenland and Africa are roughly the same size.
- D) Greenland is slightly larger than Africa.



Clicker Question

China, when superimposed on Africa, would:

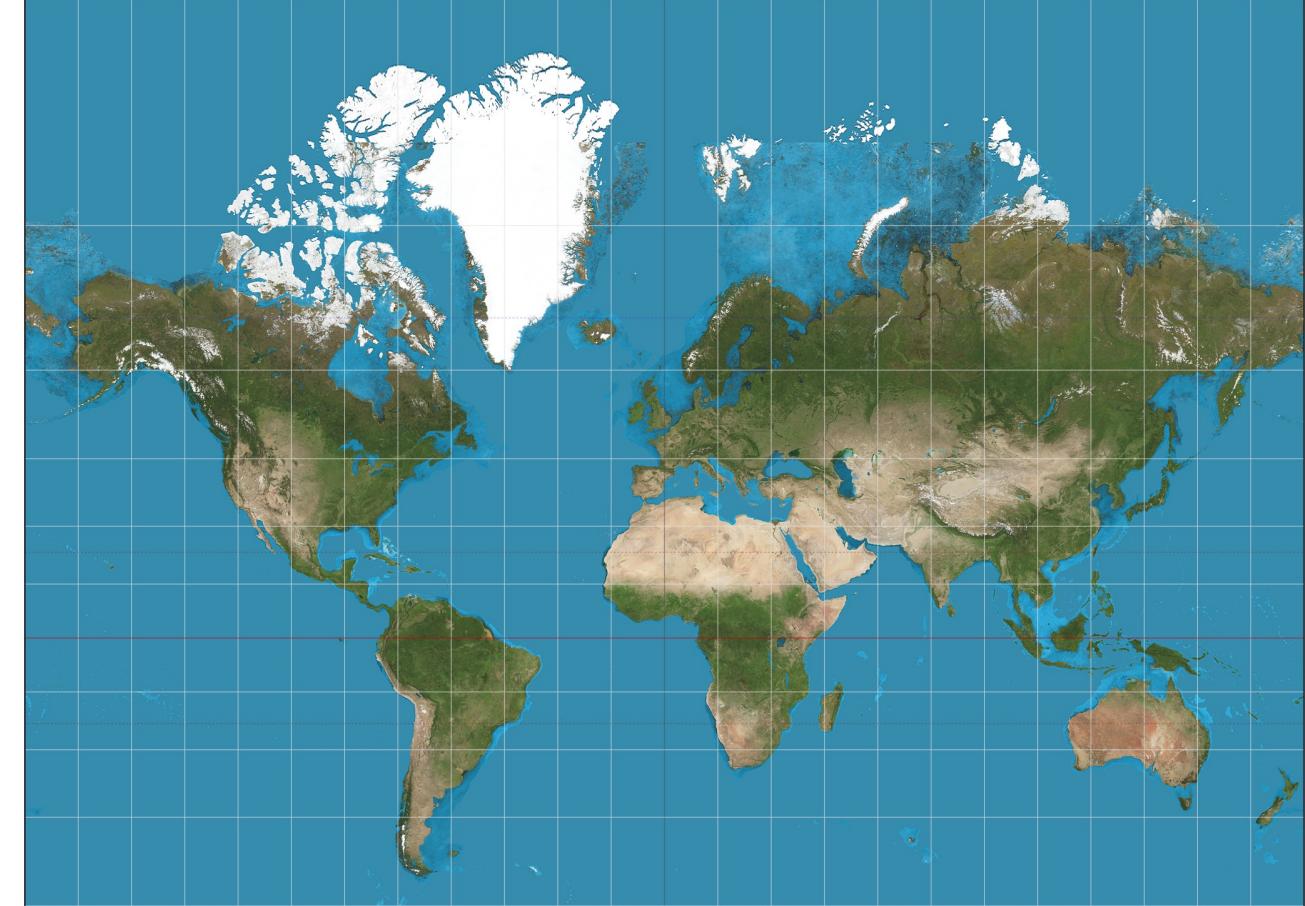
- A) Fit comfortably within the borders of East Africa.
- B) Cover a significant portion, but not the majority, of Africa.
- C) Extend from the north to the south of Africa.
- D) Be nearly identical in size to Africa.



Clicker Question

If Brazil were placed on top of Africa, how would it compare in size?

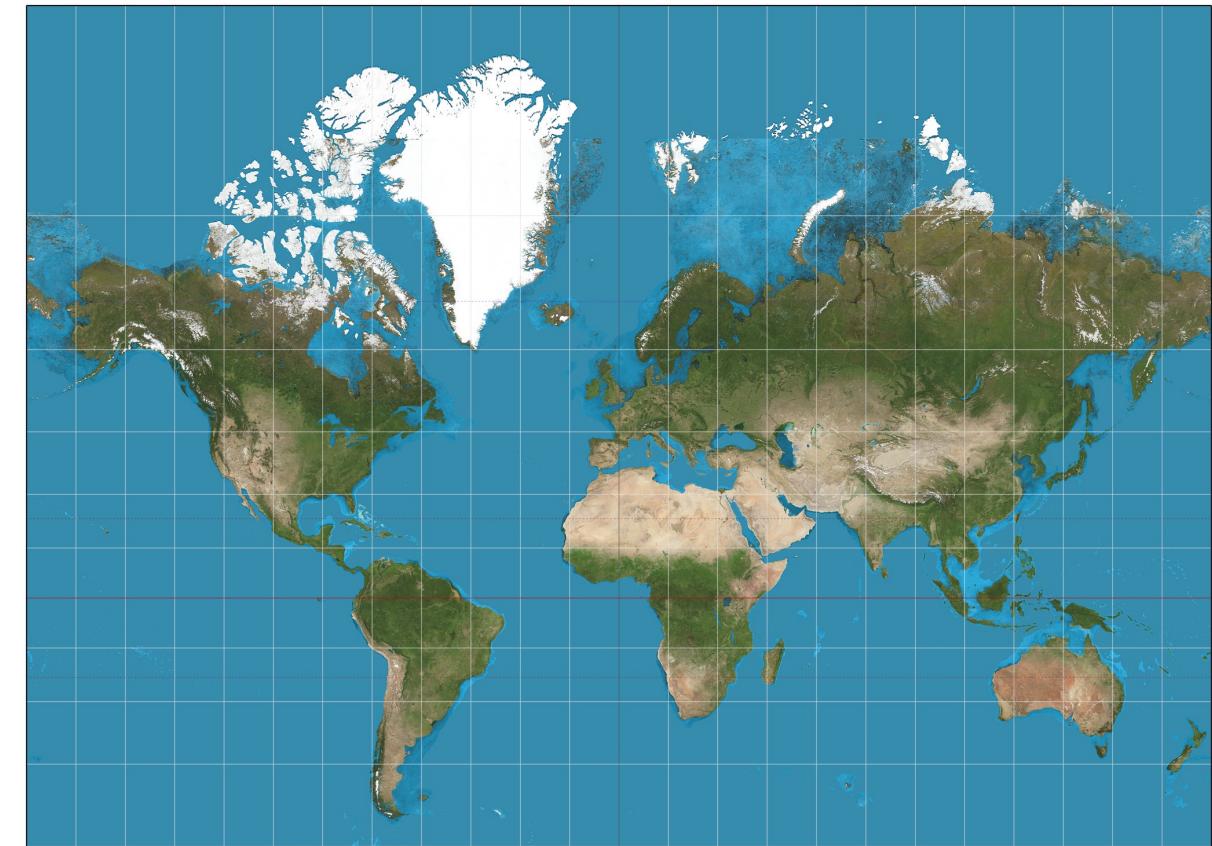
- A) It would be roughly the same size as the Congo Basin.
- B) It would be similar in size to West Africa.
- C) It would occupy the majority of Southern Africa.
- D) It would cover a comparable area to the Sahara Desert.



Clicker Question

When comparing the actual size of Greenland to that of Brazil, how does it compare?

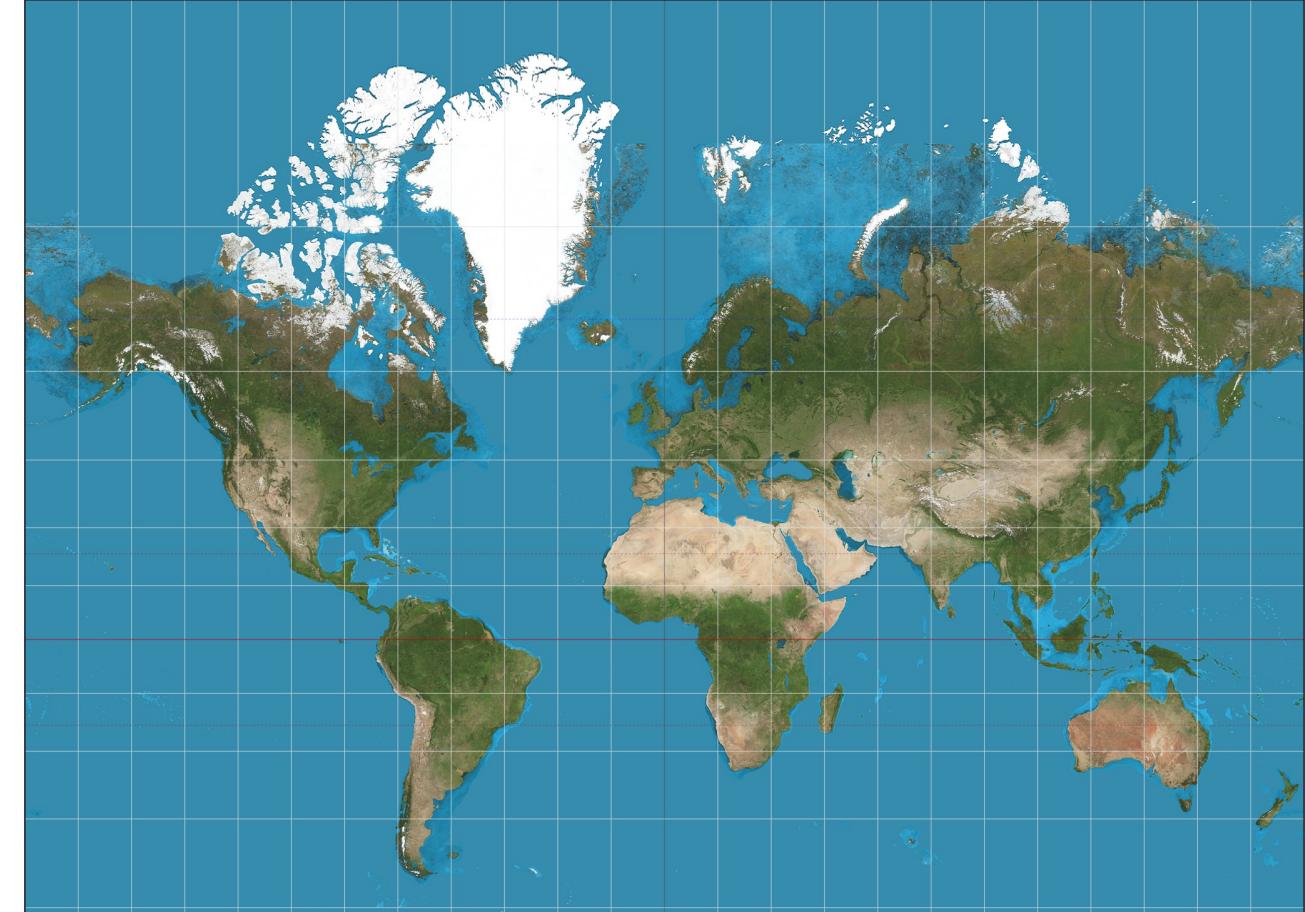
- A) Greenland is nearly as large as Brazil.
- B) Greenland is about half the size of Brazil.
- C) Greenland is comparable in size to the Brazilian state of Amazonas.
- D) Greenland is only slightly smaller than Brazil.



Clicker Question

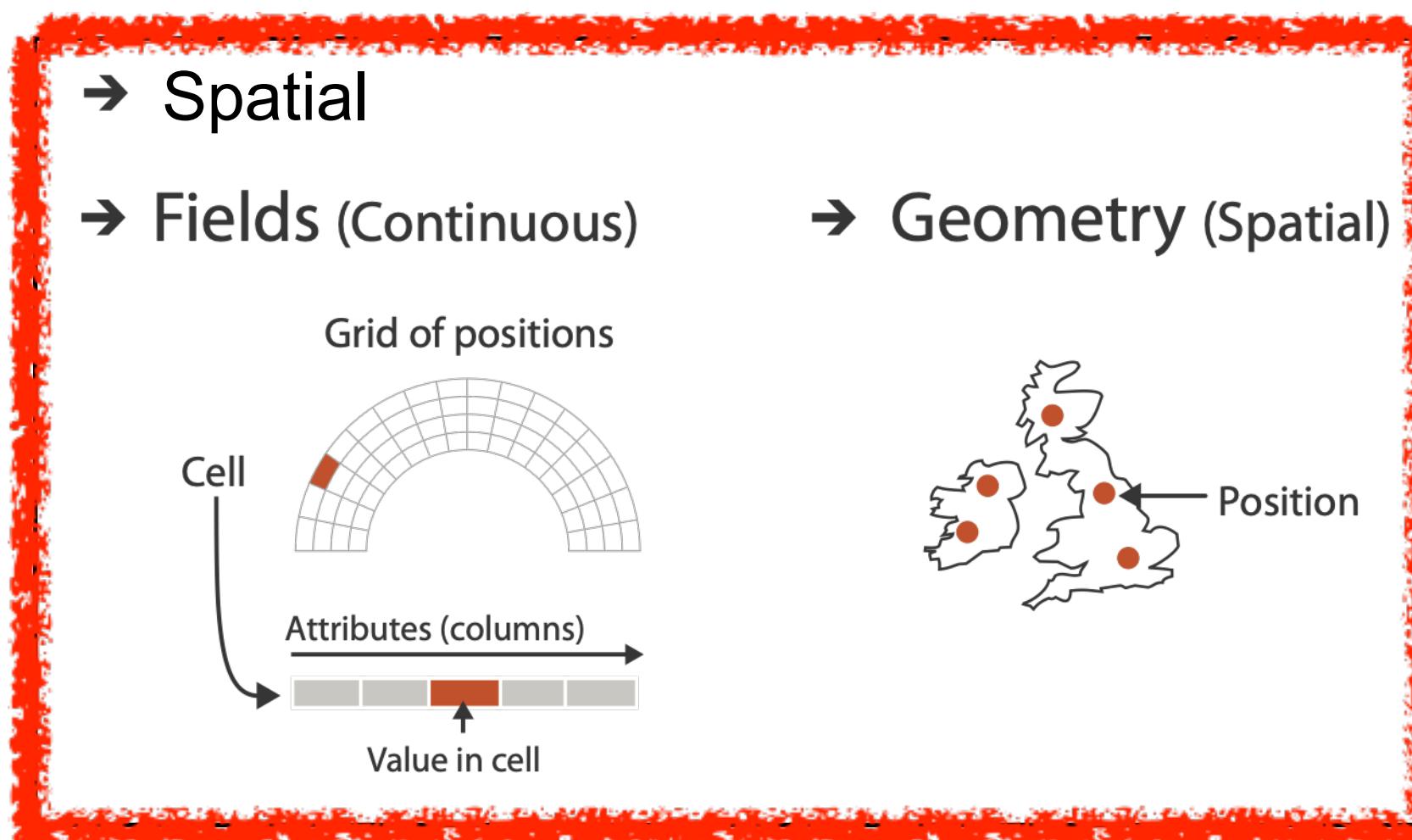
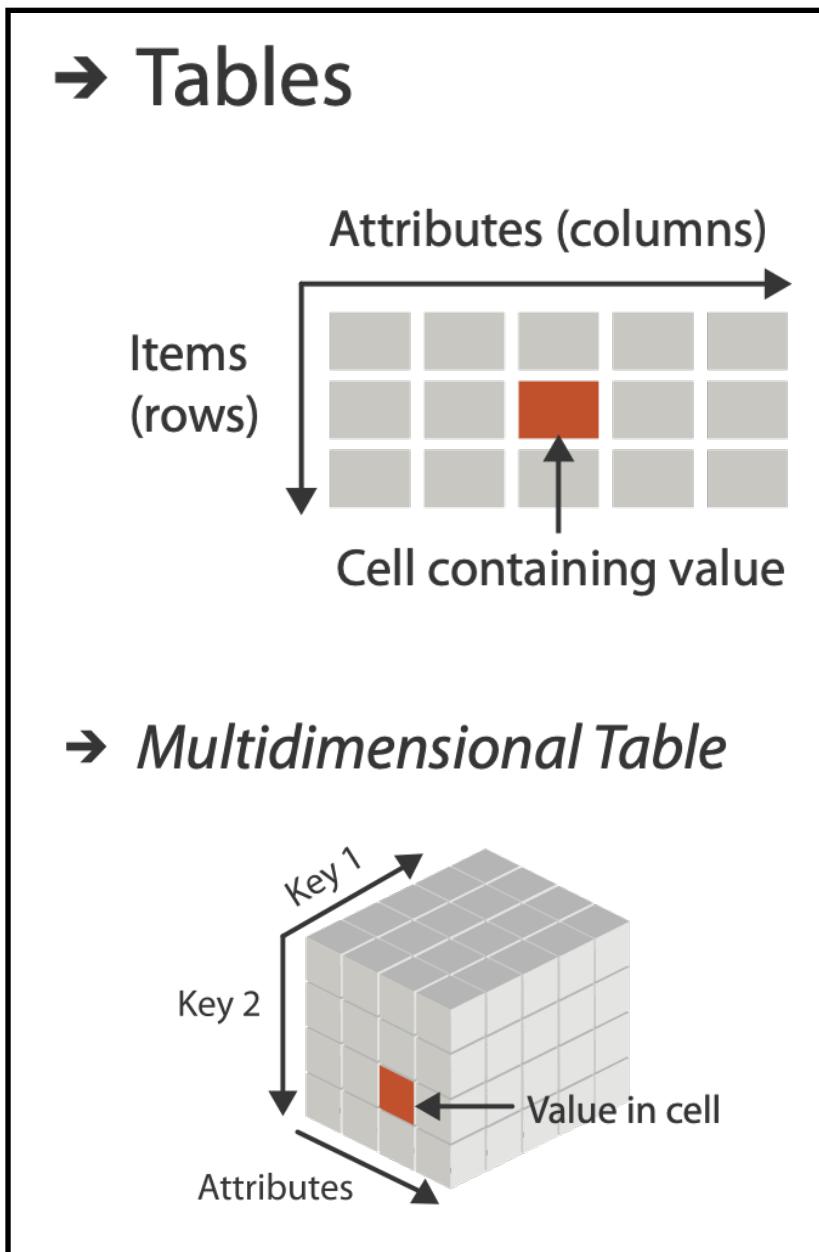
If Greenland and India were compared by size, which would be true?

- A) Greenland is nearly twice the size of India.
- B) Greenland and India are roughly the same size.
- C) India is significantly larger than Greenland.
- D) Greenland is slightly larger than India.



Focus on Spatial

→ Dataset Types



How?

Encode

→ Arrange

→ Express



→ Separate



→ Order



→ Use



What?

Why?

How?

→ Map

from categorical and ordered attributes

→ Color

→ Hue → Saturation → Luminance

→ Size, Angle, Curvature, ...



→ Shape



→ Motion

Direction, Rate, Frequency, ...



Manipulate

→ Change



→ Select



→ Navigate



Facet

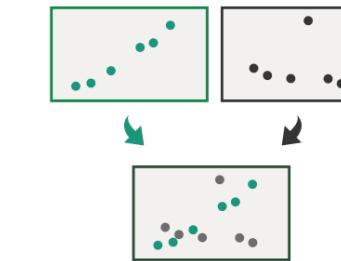
→ Juxtapose



→ Partition



→ Superimpose

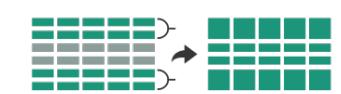


Reduce

→ Filter



→ Aggregate



→ Embed



Geographic Map



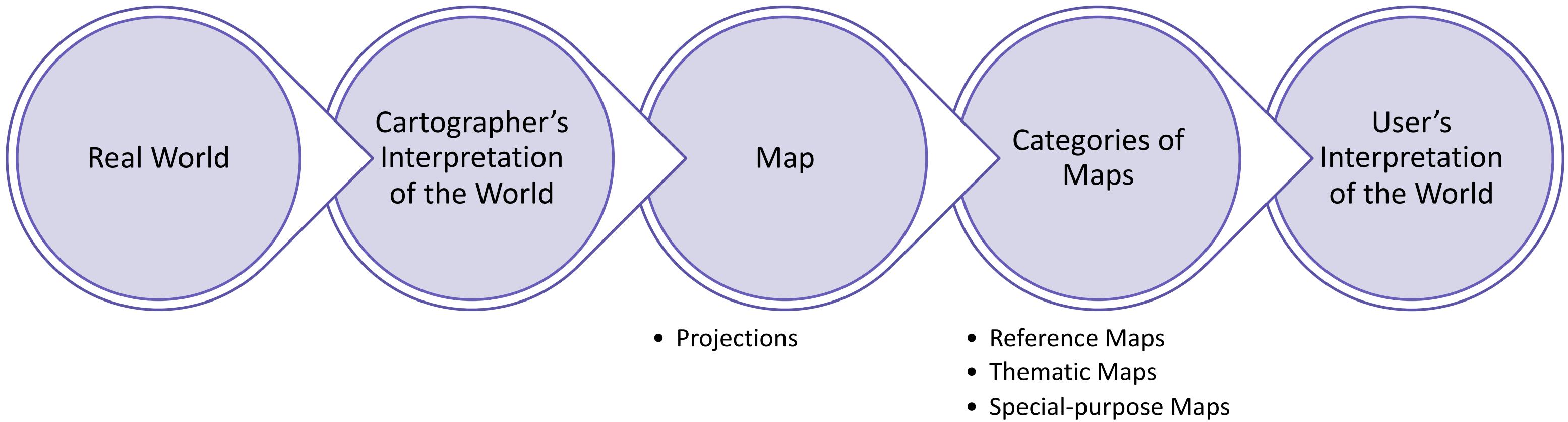
Interlocking marks

- **shape** coded
- **area** coded
- **position** coded
- cannot encode another attribute with these channels, they're "taken"

Spatial data

- use given spatial position
- when?
 - dataset contains spatial attributes and they have primary importance
 - central tasks revolve around understanding spatial relationships
- examples
 - geographical/cartographic data
 - sensor/simulation data

Geographic Map Pipeline



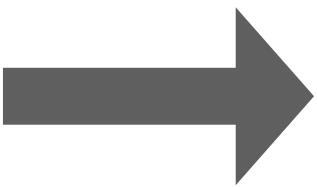
Simple Solution: Use Globe



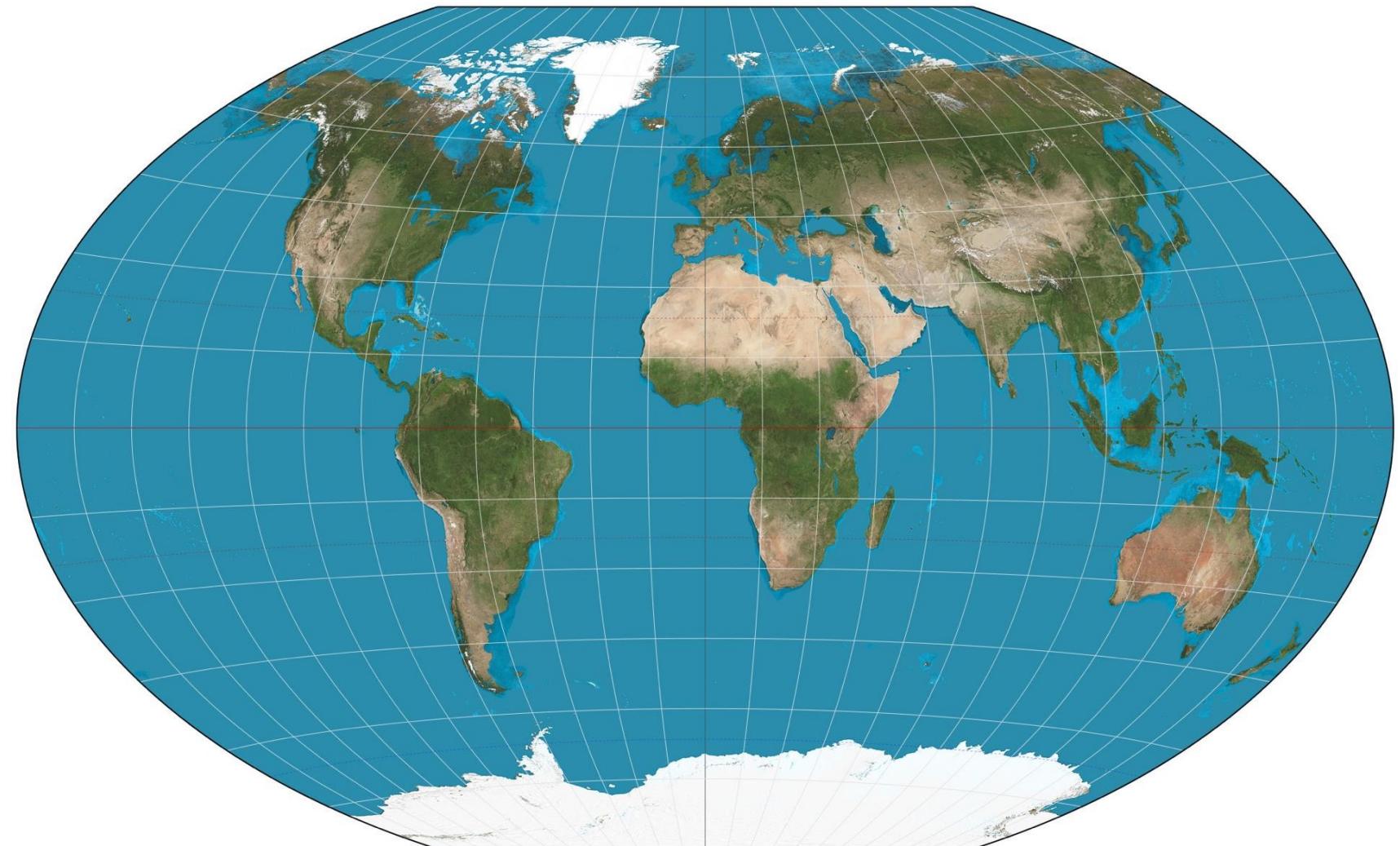
Geographic Map Projections

Dimensionality reduction

3D

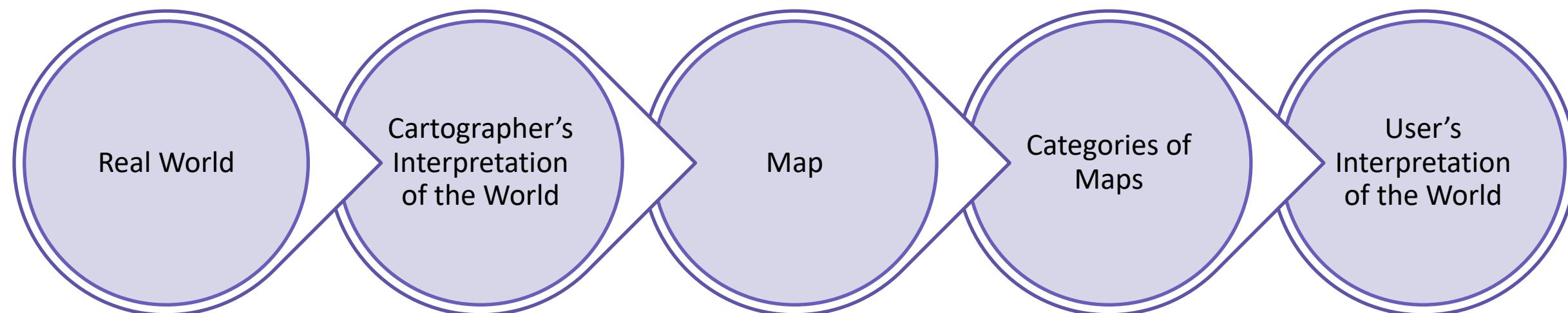


2D



Why projections?

- Earth is a (flattened) Sphere
- Need to project or “unfold” the hull of the sphere to fit onto paper/screens
- Relevant attributes: Area, Shape, Direction, Bearing, Distance, Scale



Projections

The West Wing – Mercator vs. Gall Peters

<https://www.youtube.com/watch?v=vVX-PrBRtTY>

Pay attention to

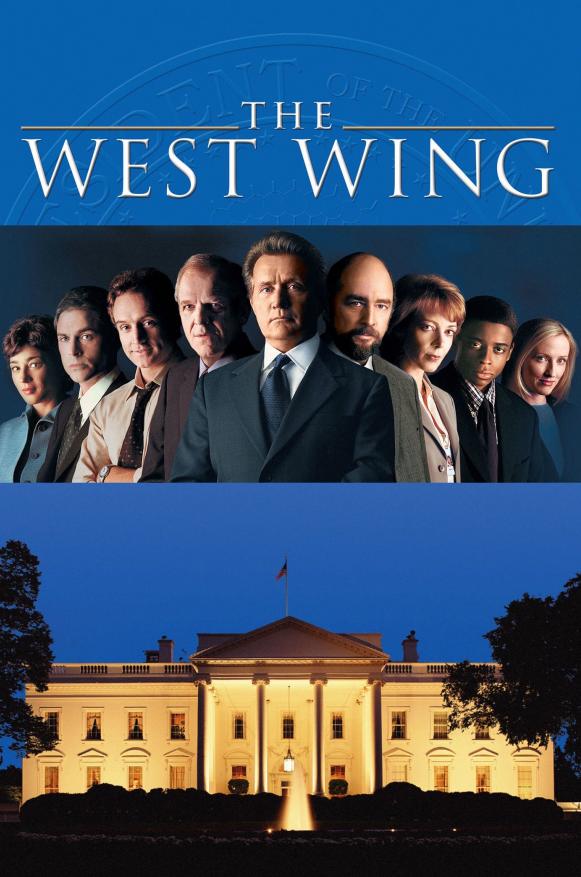
1. The strengths and limitation of the Mercator Projection
2. The impact a projection has on society

Vox – Why All Maps Are Wrong?

<https://www.youtube.com/watch?v=kIID5FDi2JQ>

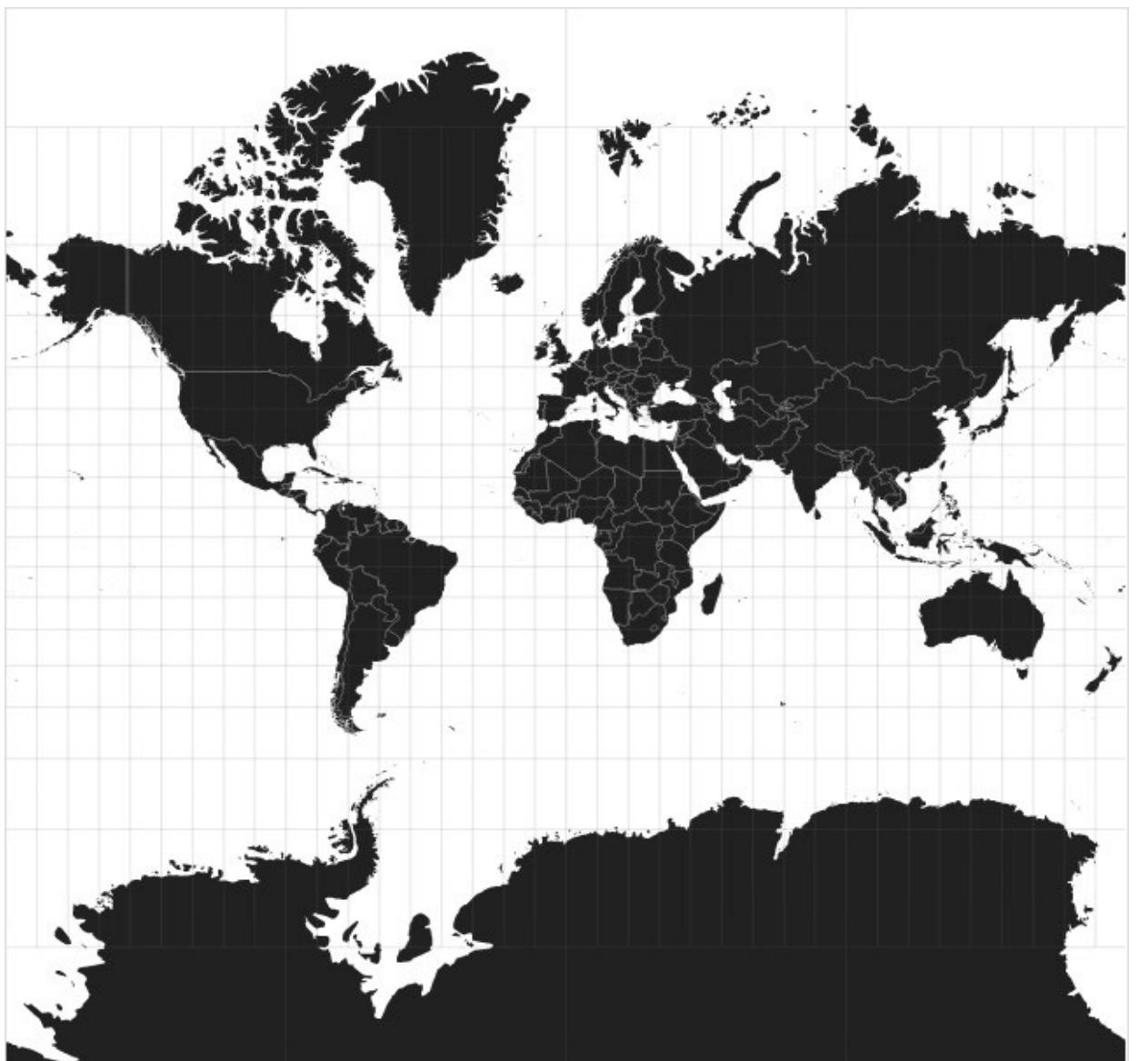
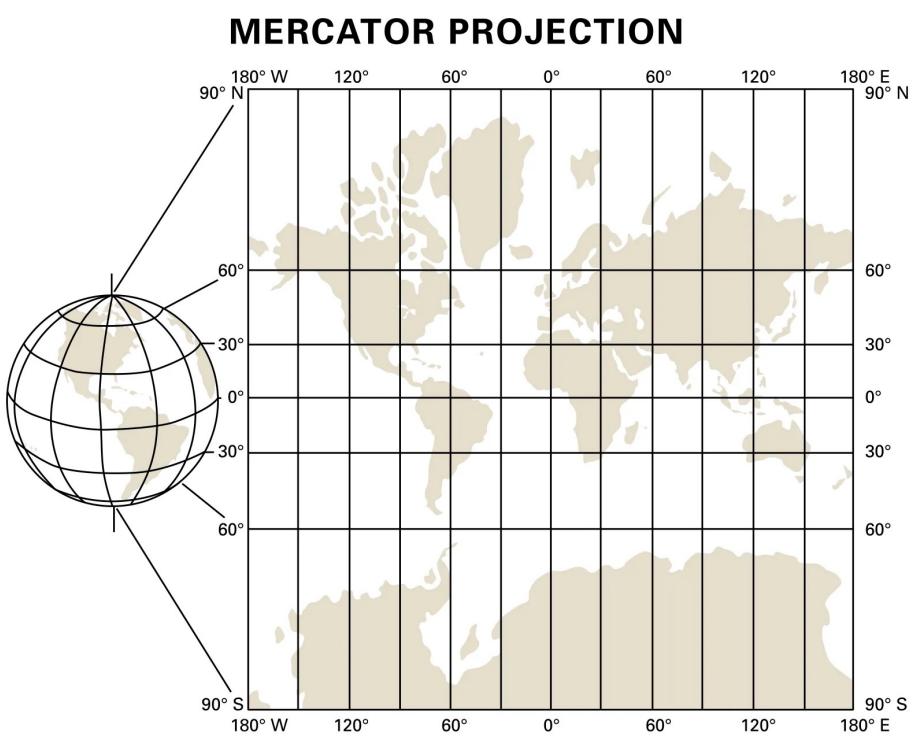
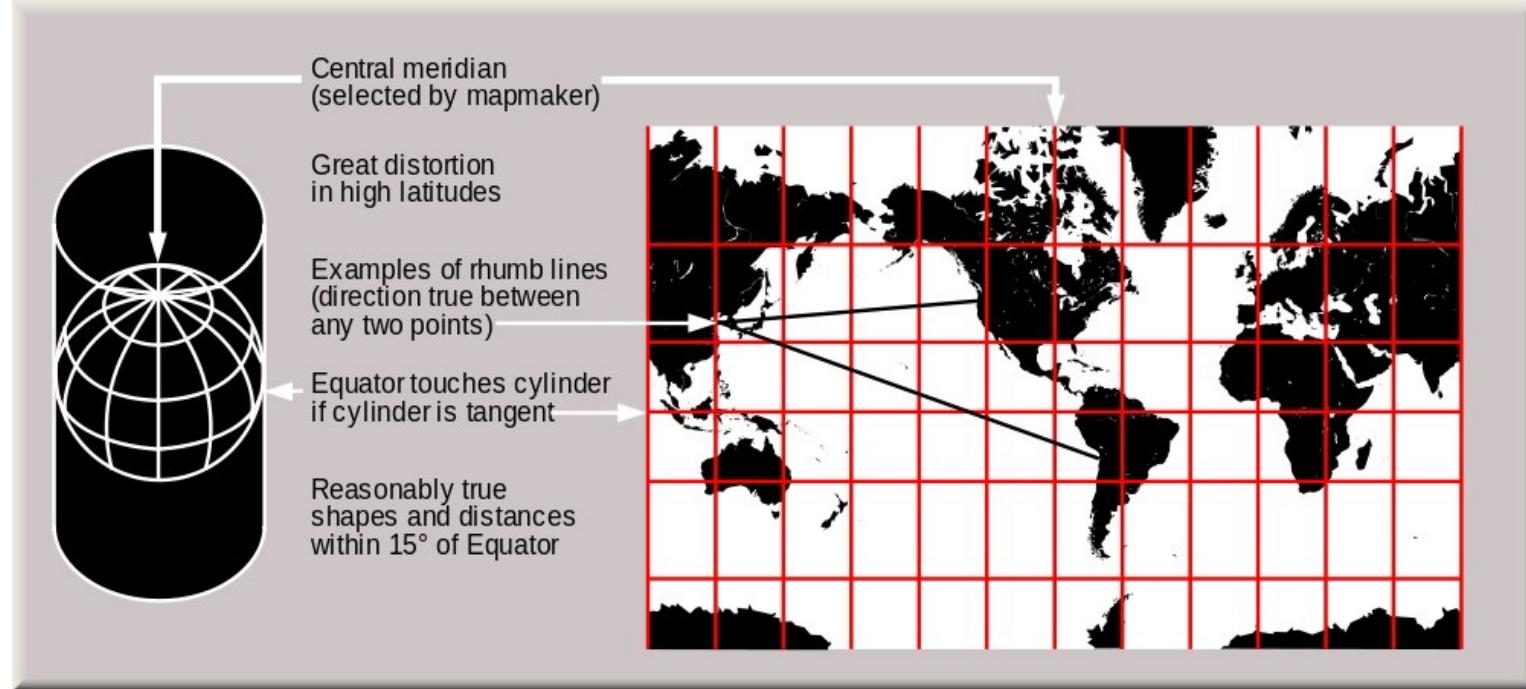
Watch after class

Pay attention to the description of various projections and why all are inherently flawed

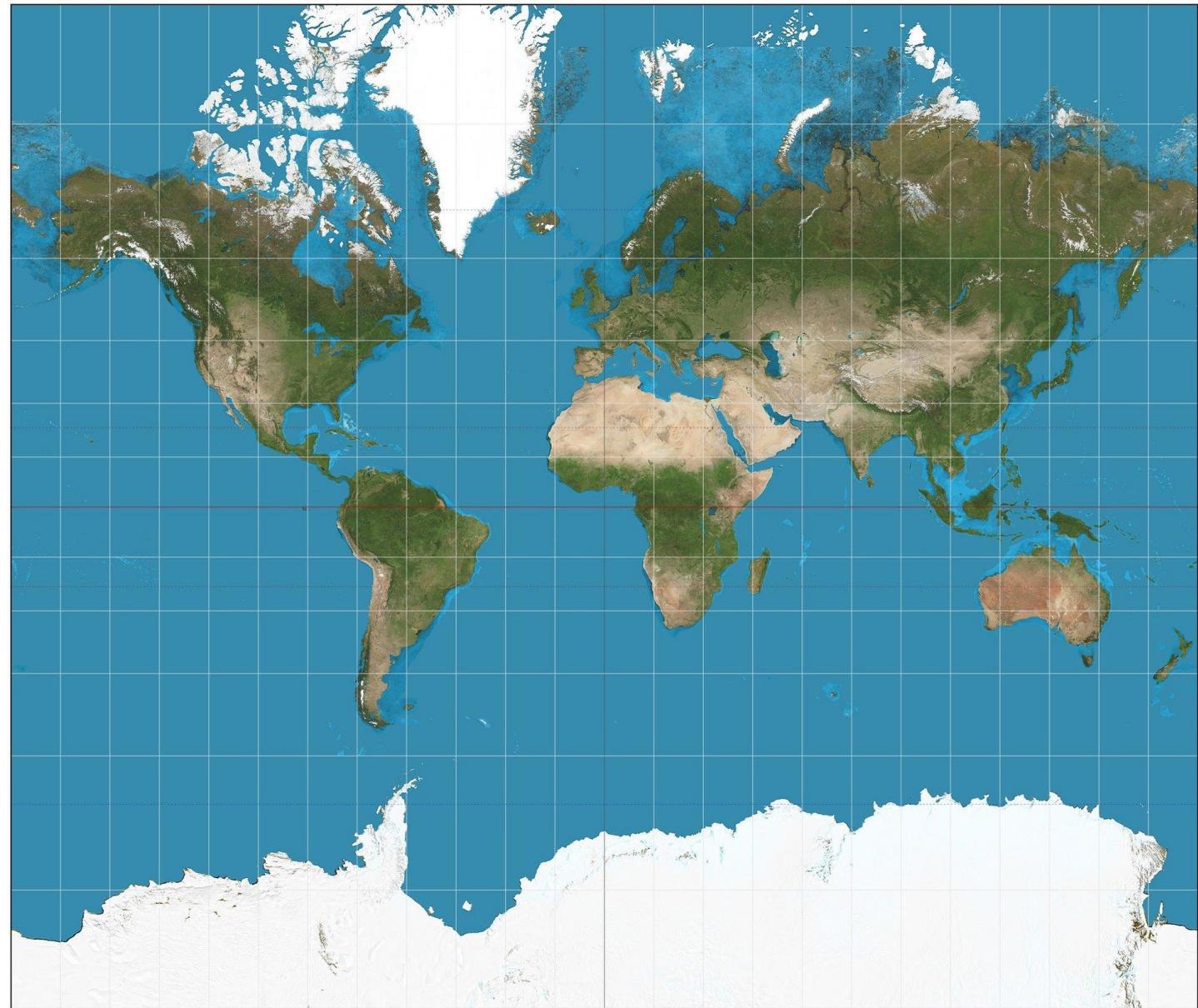


Mercator Projection (1569)

- Projection onto a cylinder wrapped around the globe Angles are preserved.
- Lines of constant bearing are straight lines.
- Constant bearing means constant compass heading – developed for sailors

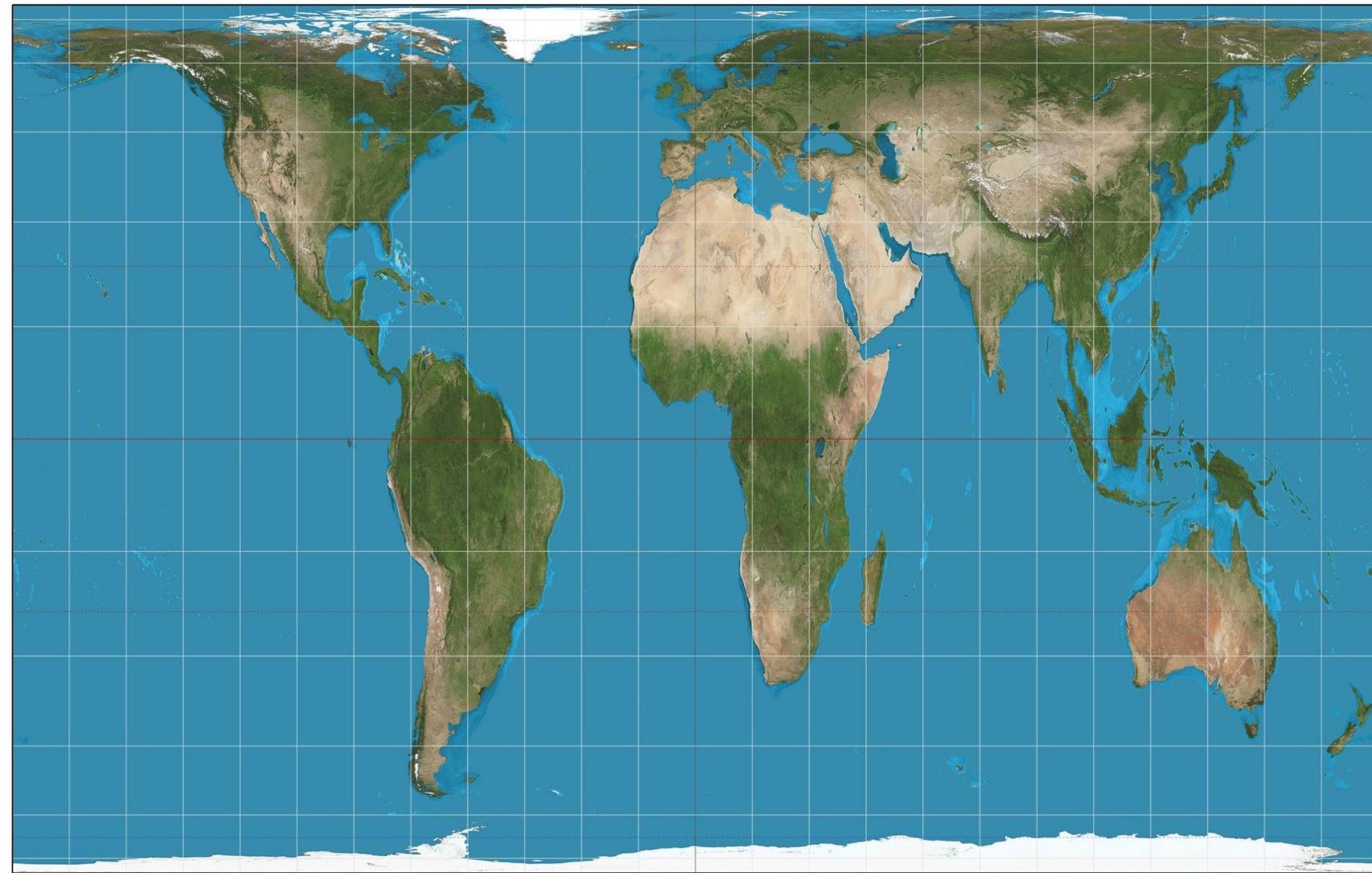


Mercator Projection



Great for ocean navigation,
but dramatically exaggerates poles.

Gall-Peters Projection



More accurate land areas.
(Officially endorsed by the UN.)

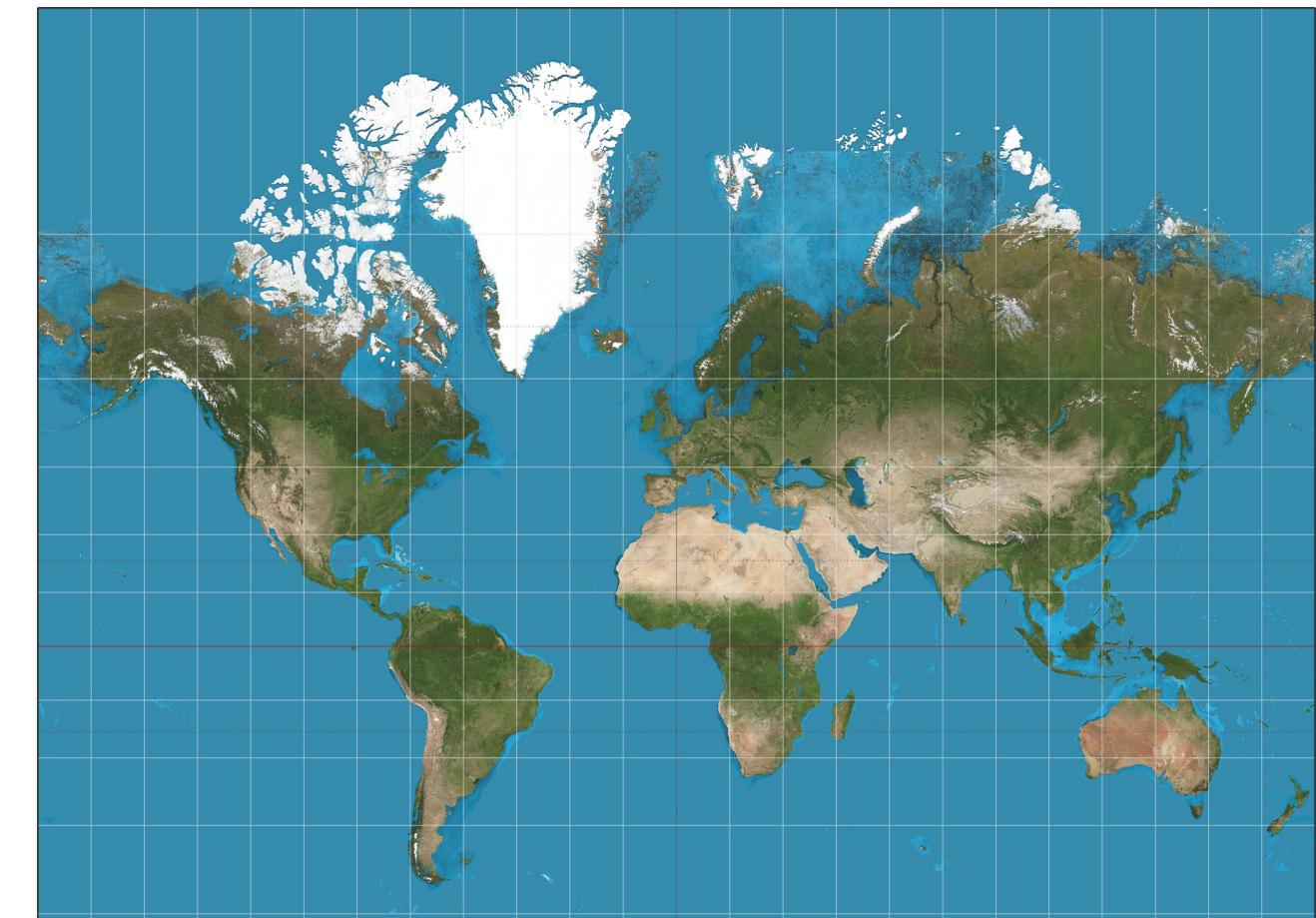
Clicker Question

If Greenland and Africa were compared by size, which would be true?

- A) Greenland is nearly twice the size of Africa.
- B) Africa is significantly larger than Greenland.**
- C) Greenland and Africa are roughly the same size.
- D) Greenland is slightly larger than Africa.

Africa is 30.3 mil km²

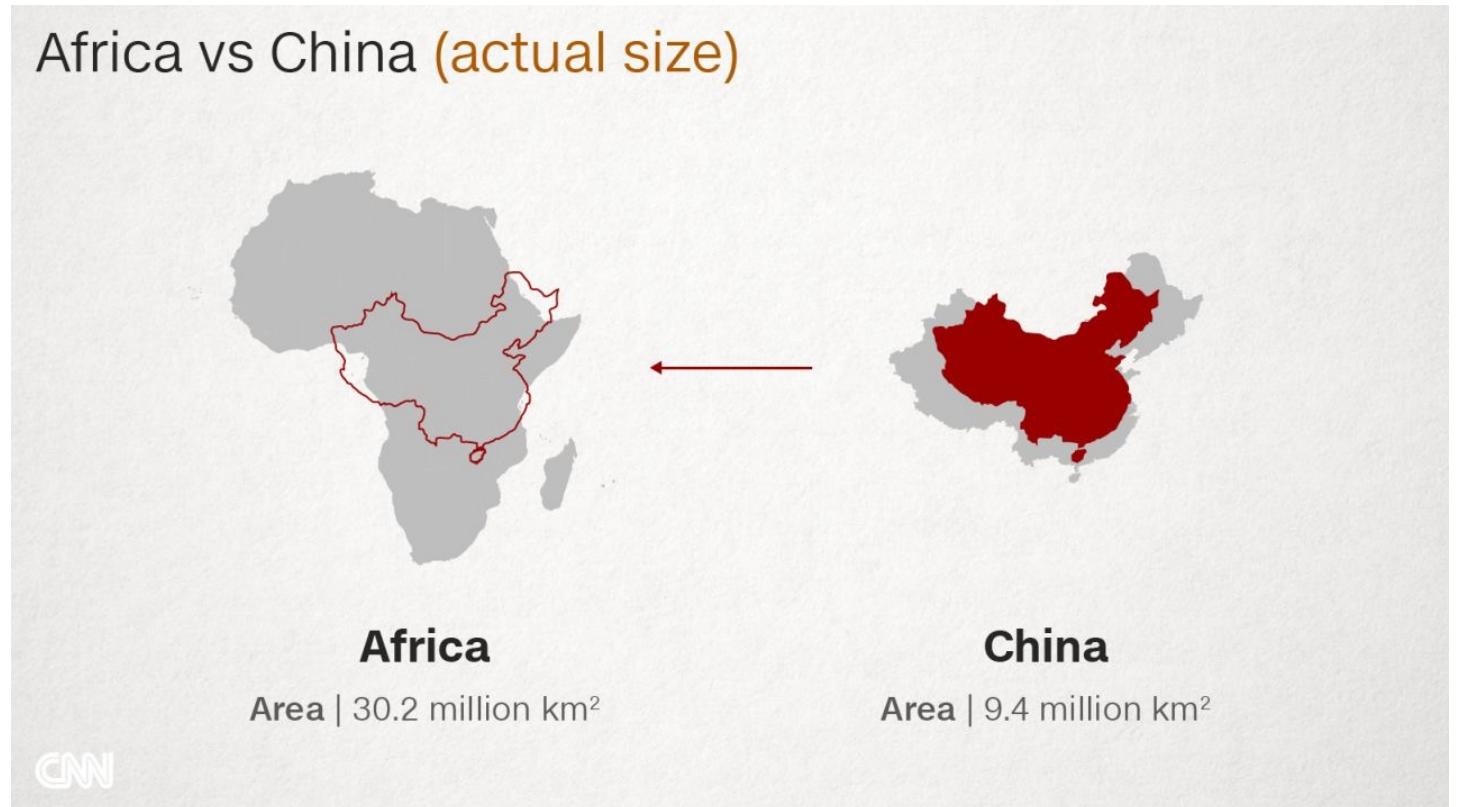
Greenland is 2.1 mil km²



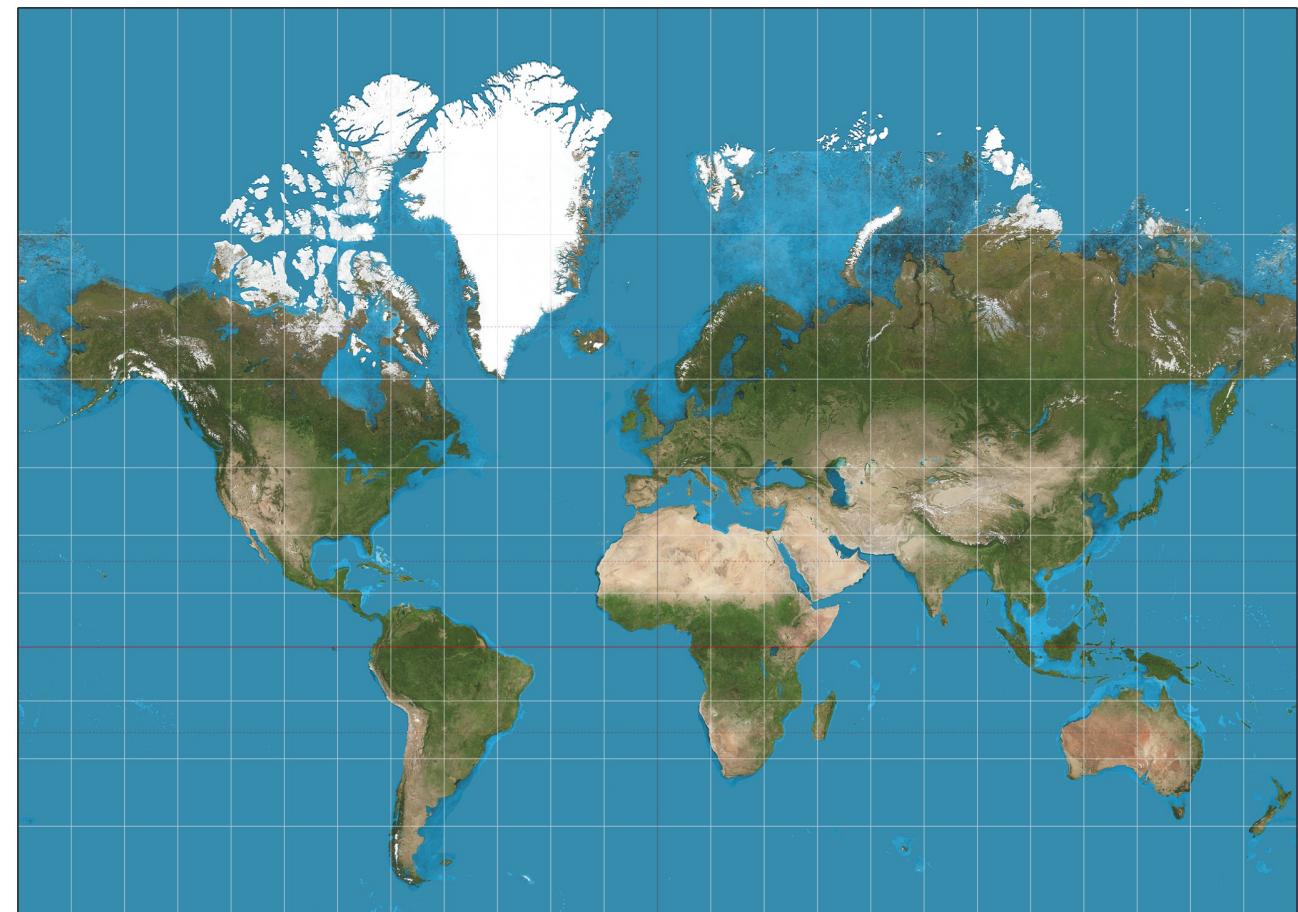
Clicker Question

China, when superimposed on Africa, would:

- A) Fit comfortably within the borders of East Africa.
- B) Cover a significant portion, but not the majority, of Africa.**
- C) Extend from the north to the south of Africa.
- D) Be nearly identical in size to Africa.



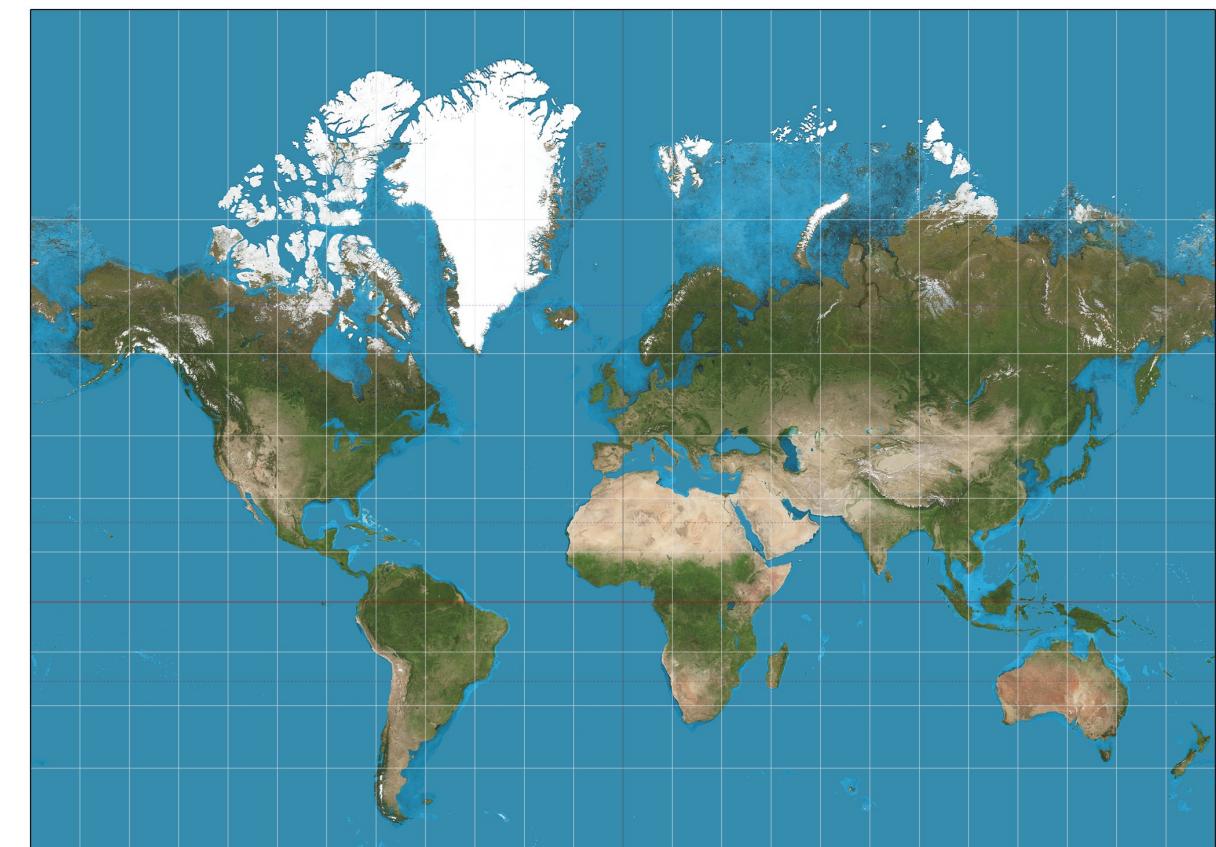
<https://www.cnn.com/2016/08/18/africa/real-size-of-africa>



Clicker Question

When comparing the actual size of Greenland to that of Brazil, how does it compare?

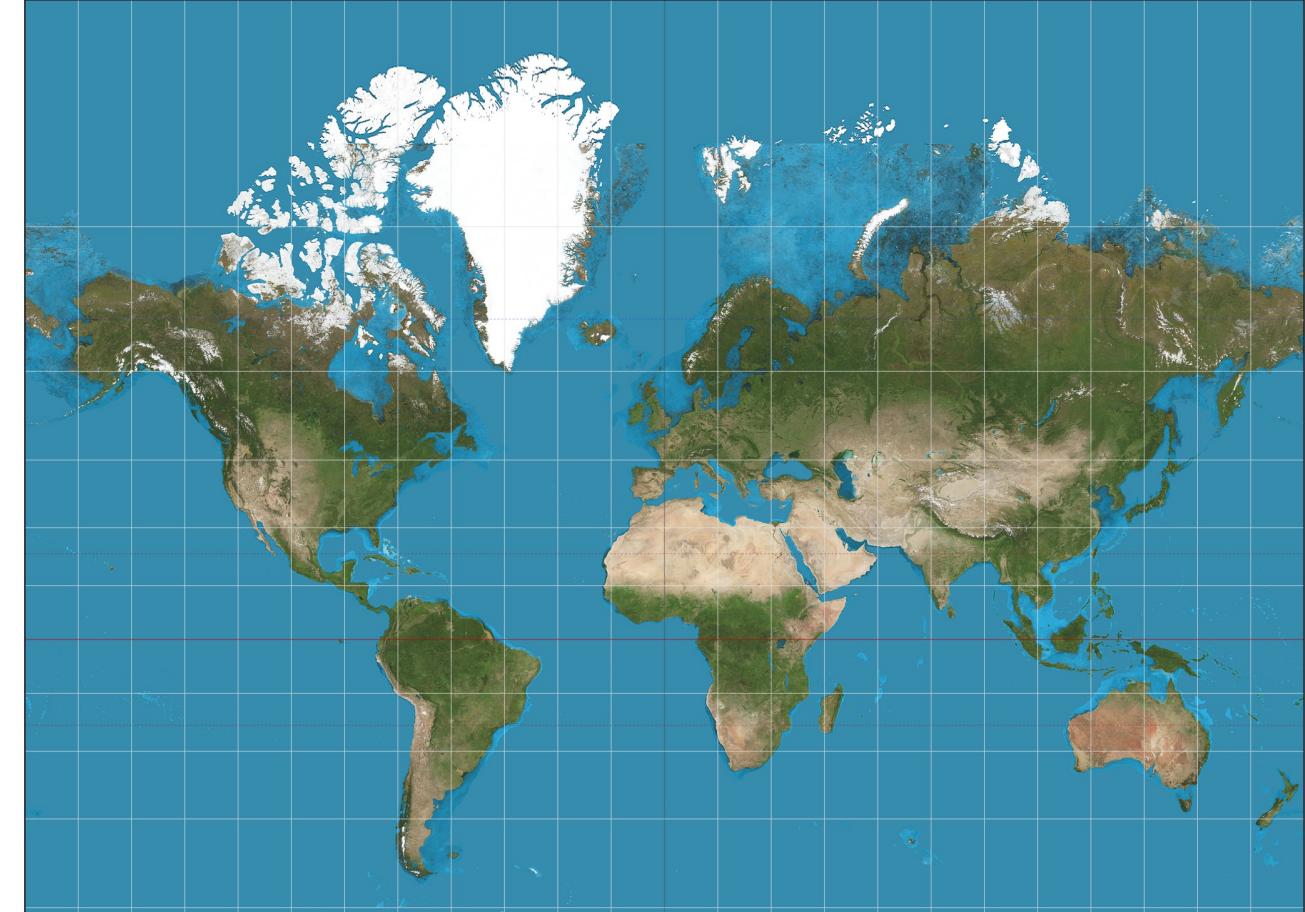
- A) Greenland is nearly as large as Brazil.
- B) Greenland is about half the size of Brazil.
- C) Greenland is comparable in size to the Brazilian state of Amazonas.
- D) Greenland is only slightly smaller than Brazil.



Clicker Question

If Greenland and India were compared by size, which would be true?

- A) Greenland is nearly twice the size of India.
- B) Greenland and India are roughly the same size.
- C) India is significantly larger than Greenland.
- D) Greenland is slightly larger than India.



In-Class Exercise: The True Size (5 min)

Go to <https://thetruesize.com>

Clear the map.

Find at least two countries using the search bar.

Position them at the equator near each other

Does this match your perception of the shape and size of that country

Try putting your countries at other locations on the map closer to the poles. How does this affect the shape & size?

See if you can find the worst possible distortion you can between the original map and a more accurate view at the equator. Post on Ed

Discussion

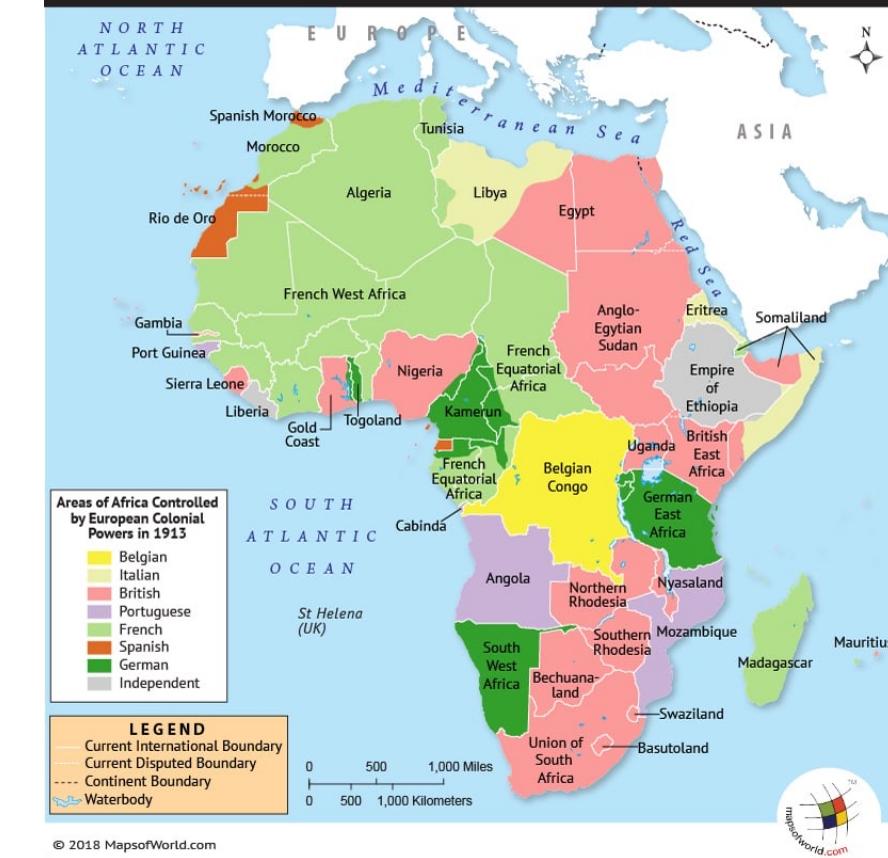
World Mercator projection with true country size added



@neilrkaye

Why Mercator was is Problematic

- Traditional map, was used to teach geography AND history
- Massive distortion (i.e., appear larger) of area distant from equator
- “unfair to the global South, making places that are mostly trees, snow, and better-off white people look huge, and the places where most of the world’s population lives look puny”



© 2018 MapsofWorld.com



The True Size of Africa

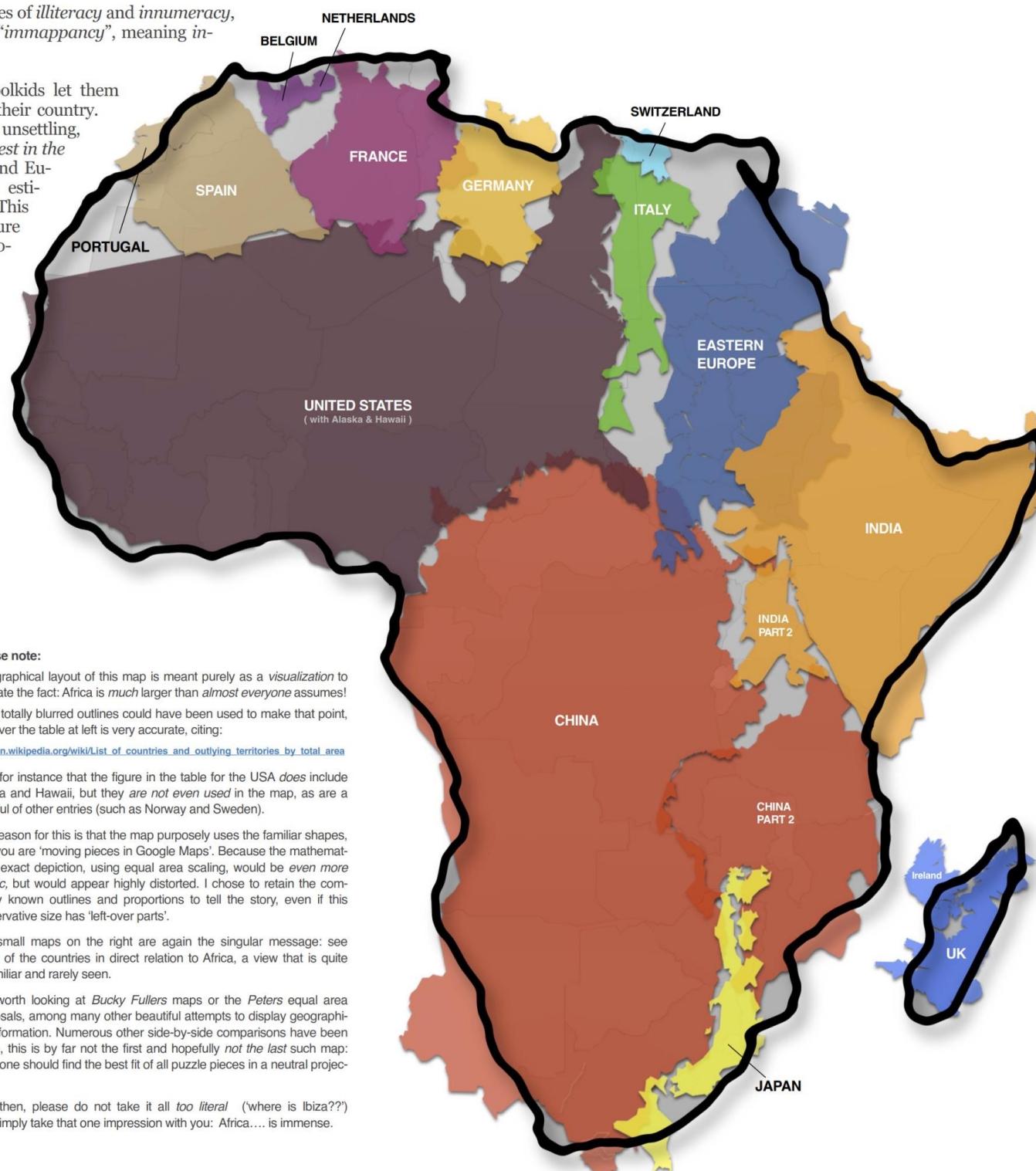
A small contribution in the fight against rampant *Immappancy*, by Kai Krause

In addition to the well known social issues of *illiteracy* and *innumeracy*, there also should be such a concept as "*immappancy*", meaning *in-sufficient geographical knowledge*.

A survey with random American schoolkids let them guess the population and land area of their country. Not entirely unexpected, but still rather unsettling, the majority chose "*1-2 billion*" and "*largest in the world*", respectively. Even with Asian and European college students, geographical estimates were often off by factors of 2-3. This is partly due to the highly distorted nature of the predominantly used mapping projections (such as *Mercator*).

A particularly extreme example is the worldwide misjudgement of the true size of Africa. This single image tries to embody the massive scale, which is larger than the *USA*, *China*, *India*, *Japan* and *all of Europe - combined!*

COUNTRY	AREA x 1000 km ²
USA	9.629
China	9.573
India	3.287
Mexico	1.964
Peru	1.285
France	633
Spain	506
Papua New Guinea	462
Sweden	441
Japan	378
Germany	357
Norway	324
Italy	301
New Zealand	270
United Kingdom	243
Nepal	147
Bangladesh	144
Greece	132
TOTAL	30.102
AFRICA	30.221
Just for Reference: The Surface of the MOON	37.930



No Rights Reserved

This work is placed in the Public Domain

Top 100 Countries

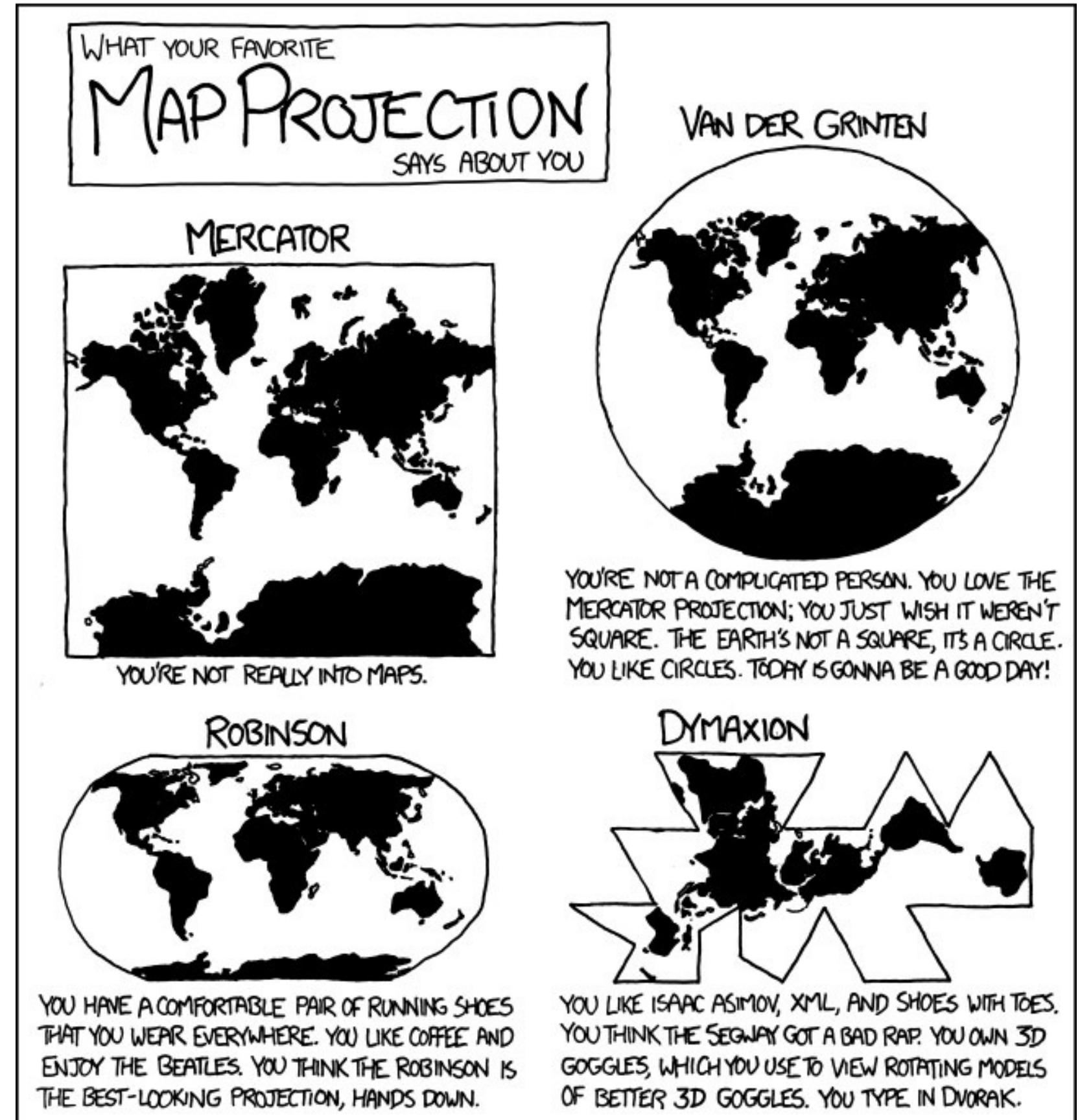
Area in square kilometers, Percentage of World Total
 Sources: Britannica, Wikipedia, Almanac 2010

	AREA km ²	%
1	Russia	17.098.242
2	Canada	9.984.670
3	China	9.596.961
4	United States	9.629.091
5	Brazil	8.514.877
6	Australia	7.692.024
7	India	3.287.263
8	Argentina	2.780.400
9	Kazakhstan	2.724.900
10	Sudan	2.505.813
11	Algeria	2.381.741
12	Congo	2.344.856
13	Greenland	2.166.086
14	Saudi Arabia	2.149.690
15	Mexico	1.964.375
16	Indonesia	1.860.360
17	Libya	1.759.540
18	Iran	1.628.750
19	Mongolia	1.564.100
20	Peru	1.285.216
21	Chad	1.284.000
22	Niger	1.267.000
23	Angola	1.246.700
24	Mali	1.240.192
25	South Africa	1.221.037
26	Colombia	1.141.748
27	Ethiopia	1.104.300
28	Bolivia	1.098.581
29	Mauritania	1.025.520
30	Egypt	1.002.000
31	Tanzania	945.087
32	Nigeria	923.768
33	Venezuela	912.050
34	Namibia	824.116
35	Mozambique	801.590
36	Pakistan	796.095
37	Turkey	783.562
38	Chile	756.102
39	Zambia	752.612
40	Myanmar	676.578
41	Afghanistan	652.090
42	Somalia	637.657
43	France	632.834
44	C. African Rep	622.984
45	Ukraine	603.500
46	Madagascar	587.041
47	Botswana	582.000
48	Kenya	580.367
49	Yemen	527.968
50	Thailand	513.120
51	Spain	505.992
52	Turkmenistan	488.100
53	Cameroon	475.442
54	Papua New Guinea	462.840
55	Uzbekistan	447.400
56	Morocco	446.550
57	Sweden	441.370
58	Iraq	438.317
59	Paraguay	406.752
60	Zimbabwe	390.757
61	Japan	377.930
62	Germany	357.114
63	Rep o.t. Congo	342.000
64	Finland	338.419
65	Vietnam	331.212
66	Malaysia	330.803
67	Norway	323.802
68	Côte d'Ivoire	322.463
69	Poland	312.685
70	Oman	309.500
71	Italy	301.336
72	Philippines	300.000
73	Burkina Faso	274.222
74	New Zealand	270.467
75	Gabon	267.668
76	Western Sahara	266.000
77	Ecuador	256.369
78	Guinea	245.857
79	United Kingdom	242.900
80	Uganda	241.038
81	Ghana	238.539
82	Romania	238.391
83	Laos	236.800
84	Guyana	214.969
85	Belarus	207.600
86	Kyrgyzstan	199.951
87	Senegal	196.722
88	Syria	185.180
89	Cambodia	181.035
90	Uruguay	176.215
91	Suriname	163.820
92	Tunisia	163.610
93	Nepal	147.181
94	Bangladesh	143.998
95	Tajikistan	143.100
96	Greece	131.957
97	Nicaragua	130.373
98	North Korea	120.538
99	Malawi	118.484
100	Eritrea	117.600
	TOP 100 TOTAL	132.632.524 89.34

Map Projections

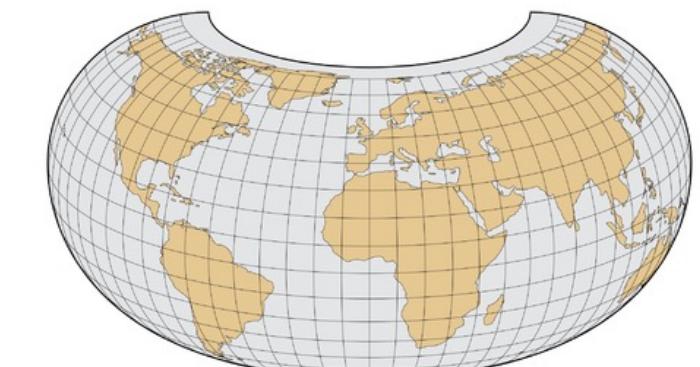
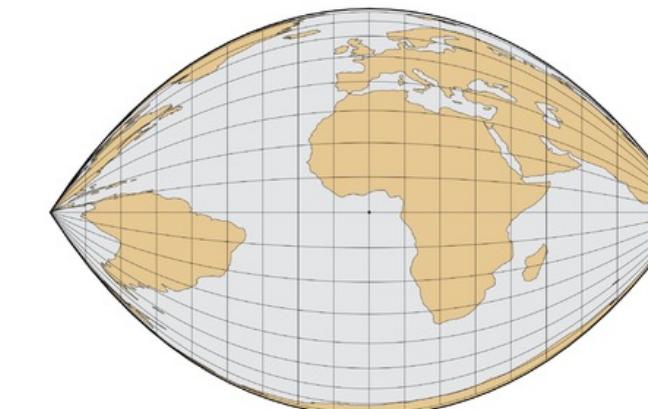
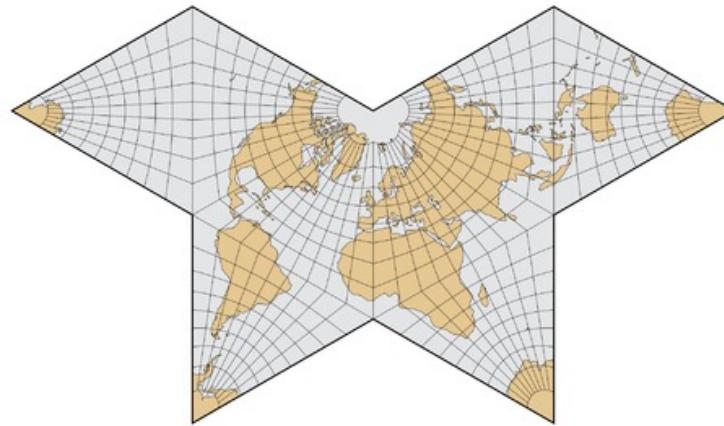
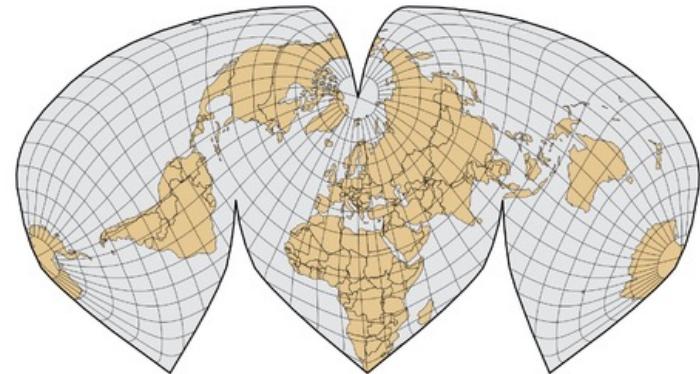
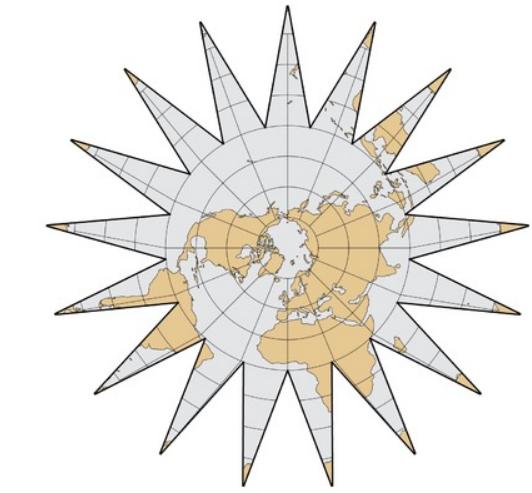
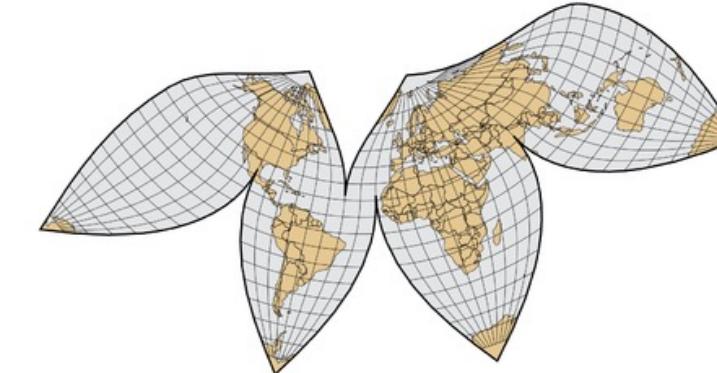
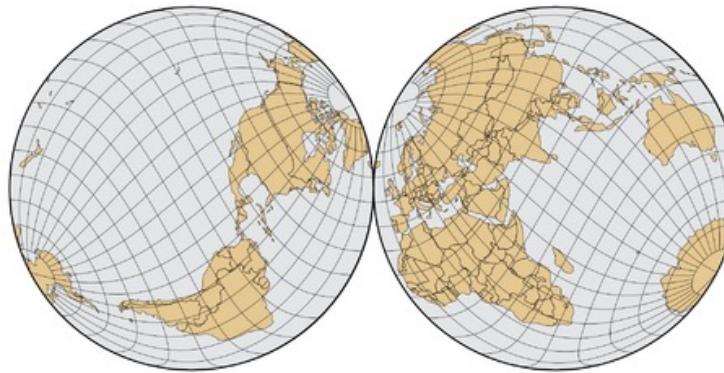
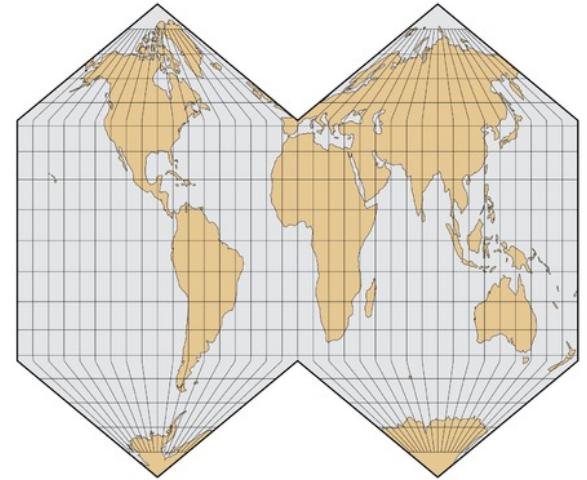
Projection Transitions -

[https://observablehq.com/@d3/
projection-transitions](https://observablehq.com/@d3/projection-transitions)

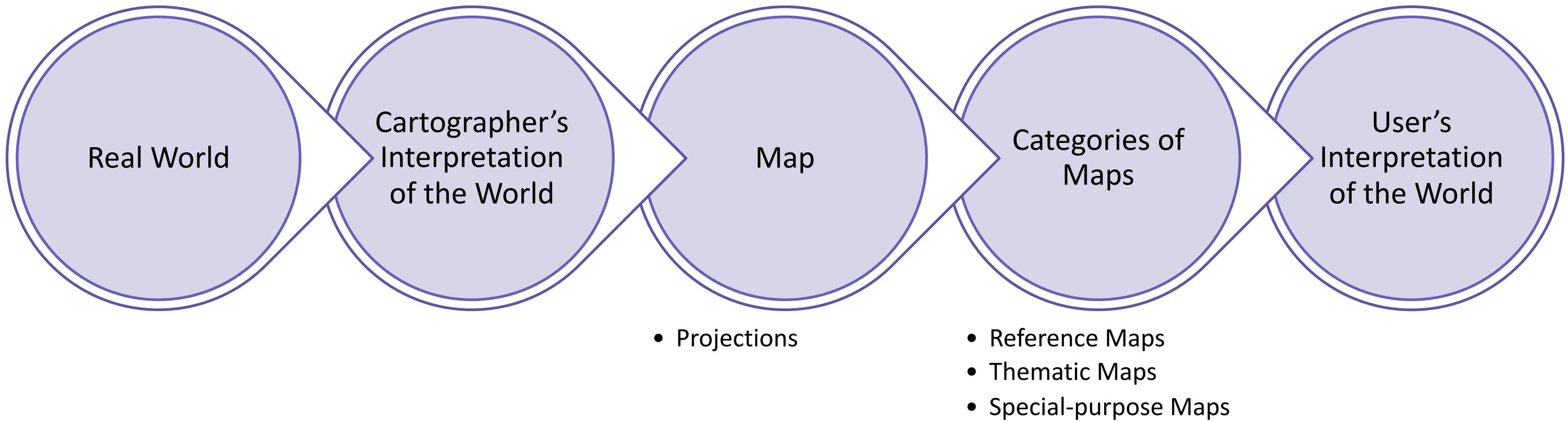


Map Projections

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



Geographic Map Pipeline

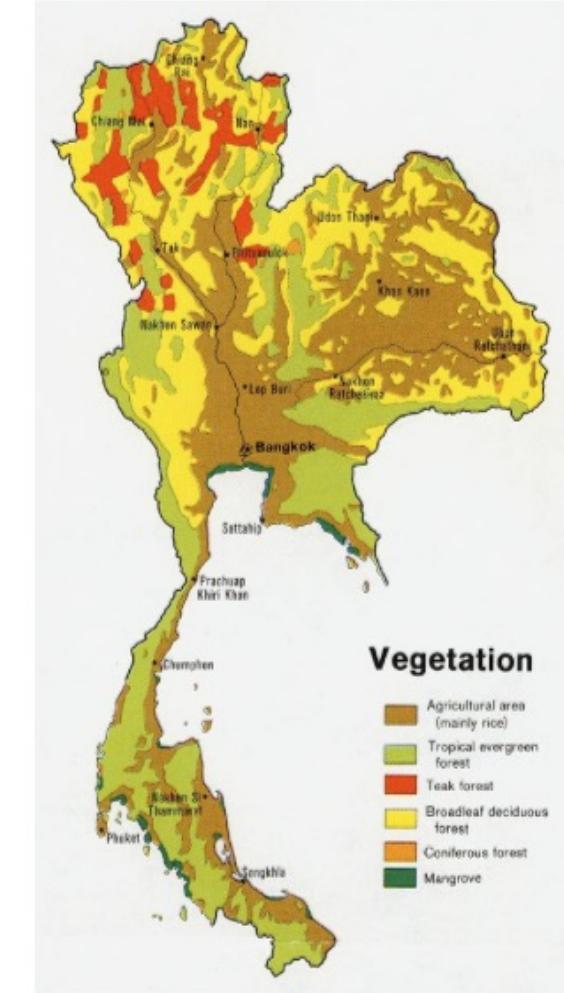


Categories of Maps

Reference: show several types of spatial data without specific emphasis on one type over another

Thematic: have a specific focus, typically display attributes of features that vary spatially in a qualitative or nominal way

Special-purpose: typically thematic maps but are task/user specific (i.e., used like reference maps but for specific tasks or specific types of data).



LEFT: Reference map of Thailand.

CENTER: Thematic map classes of vegetation in 1974

RIGHT: Special purpose map – Thailand rail map

Thematic maps

show spatial variability of attribute ("theme")

- combine geographic / reference map with (simple, flat) tabular data
- join together
 - region: interlocking area marks (provinces, countries with outline shapes)
 - also could have point marks (cities, locations with 2D lat/lon coordinates)
 - region: categorical key attribute in table
 - use to look up value attributes
- major idioms
 - choropleth
 - symbol maps
 - cartograms
 - dot density maps

Principles

Use maps when spatial relationships are paramount

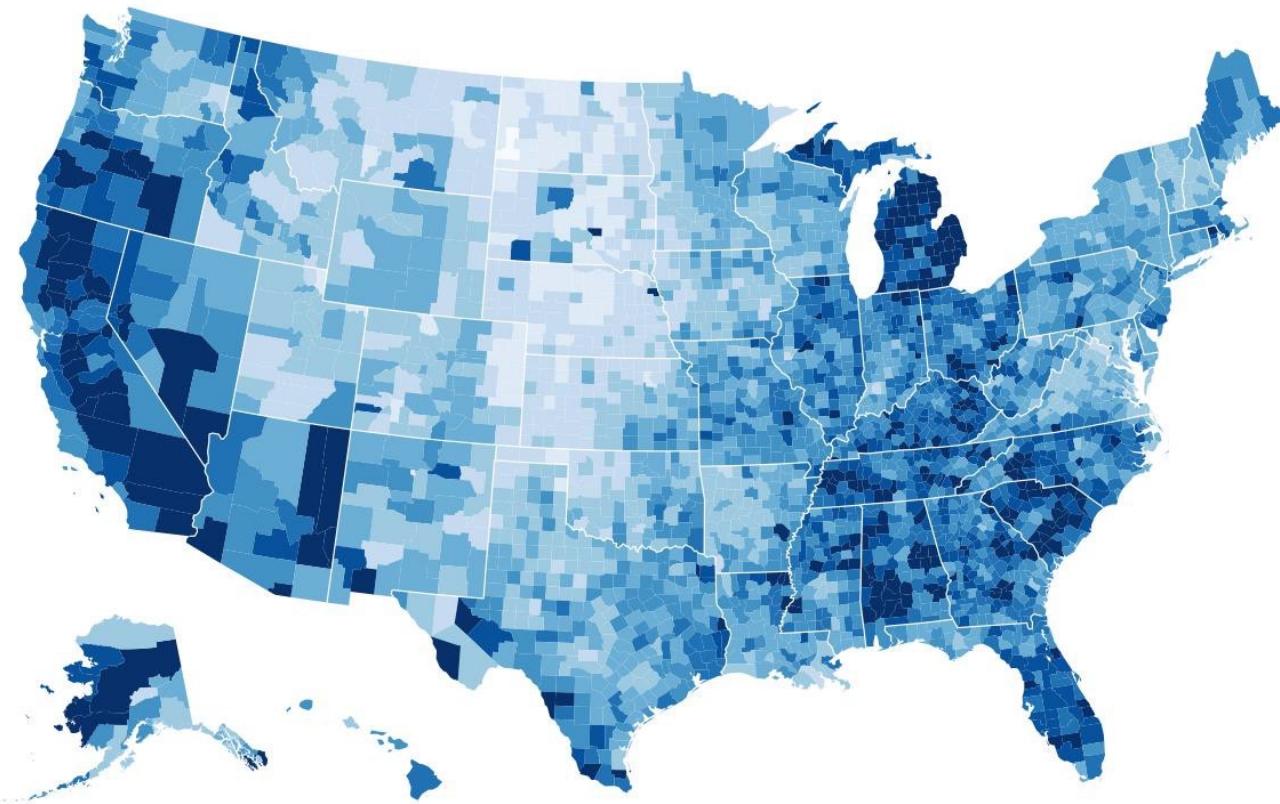
Map Tasks:

- Find location / feature (county, country, city, street)
- Find route
- Identify attribute associated with location (elevation, land/water, GDP)
- Compare attributes between locations/features

CAVEAT: Thematic maps are misleading for large areas (e.g. at country level) but distortion is not a problem at (city or county level)

Idiom: **choropleth map**

- use given spatial data
 - when 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 colormap



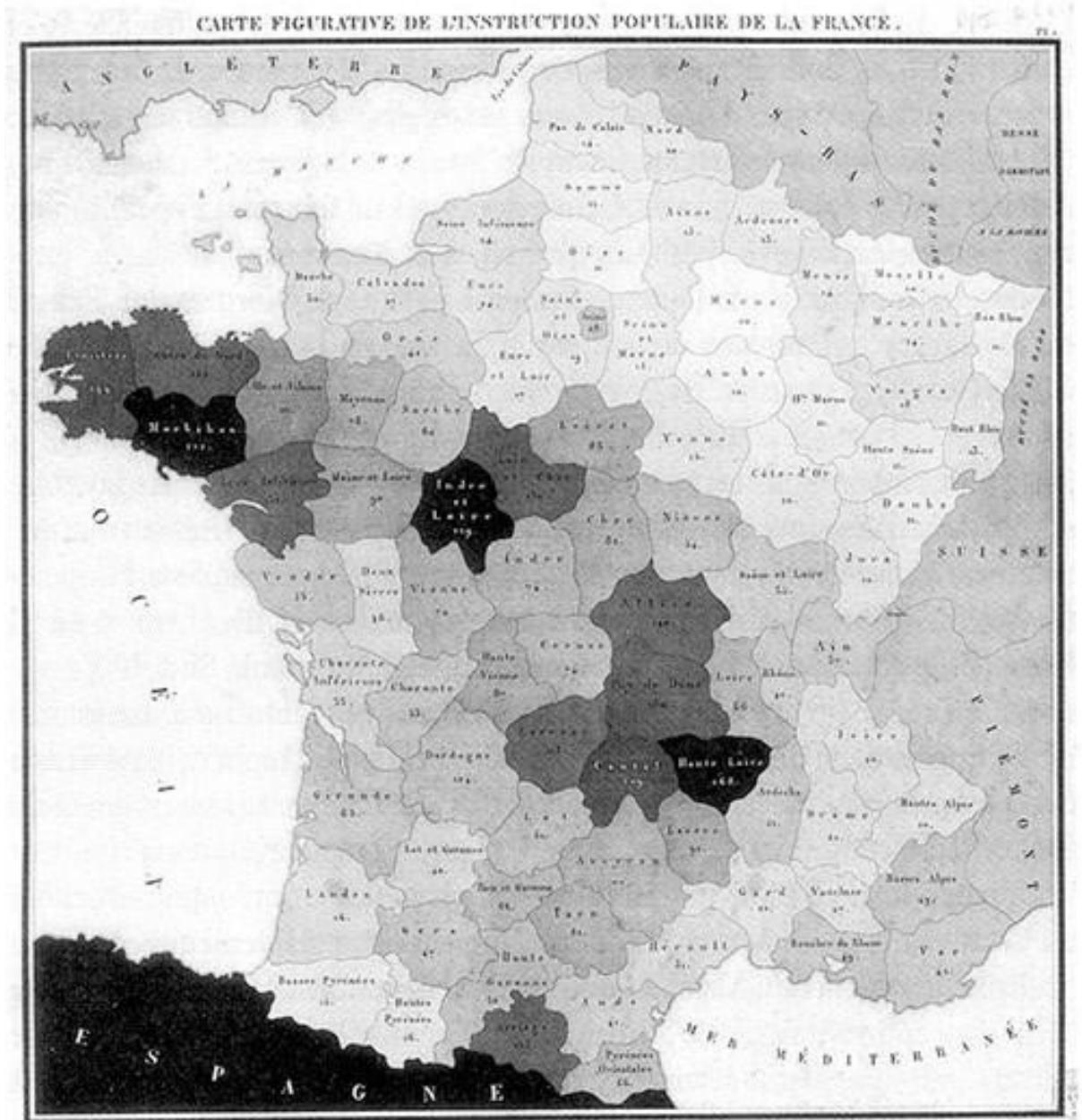
<http://bl.ocks.org/mbostock/4060606>

Idiom: **choropleth map**

Areas are shaded or patterned in proportion to measurement

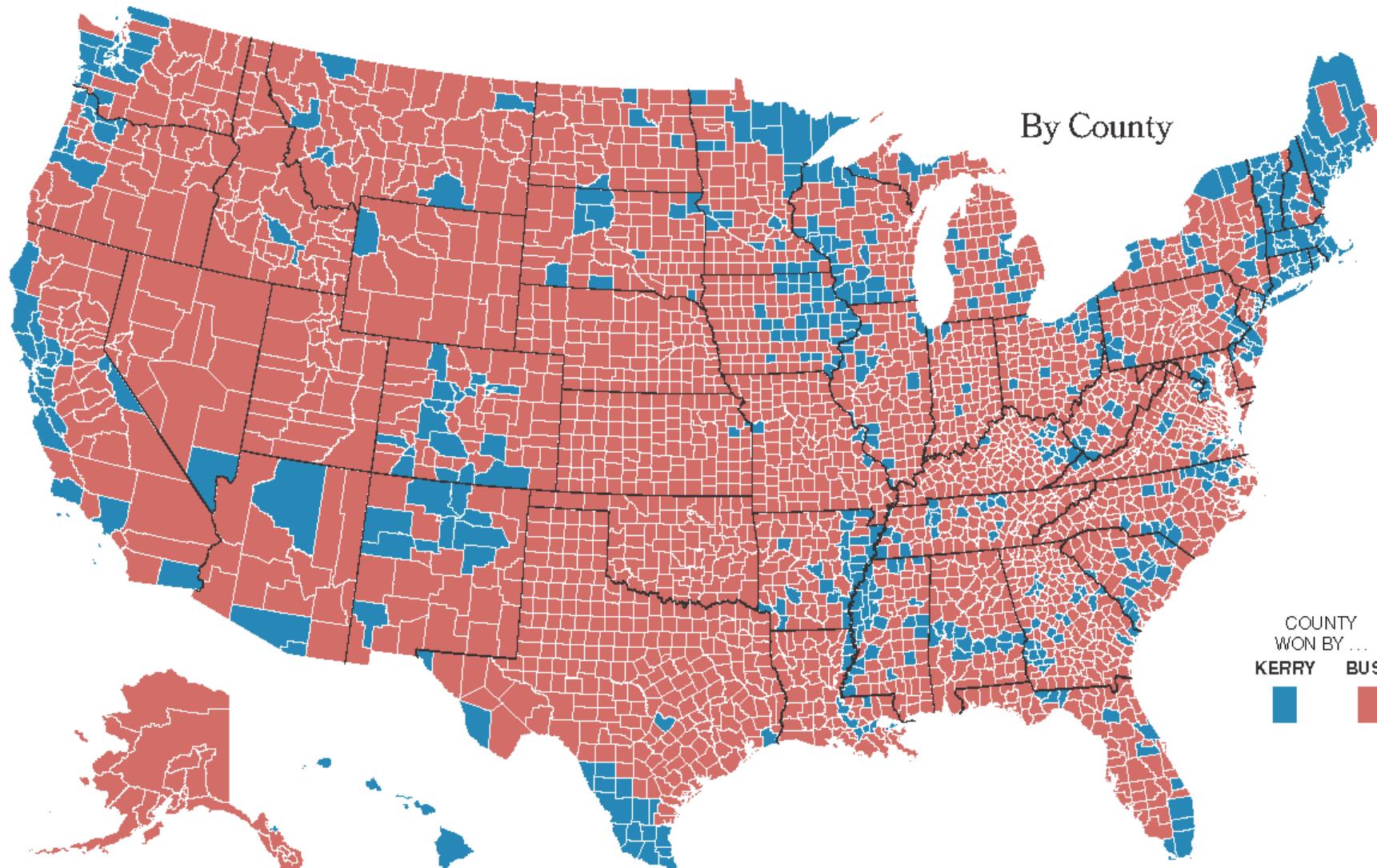
Each spatial unit is filled with a uniform color or pattern

Illiteracy in France



Charles Dupin, 1826

Presidential Elections in United States



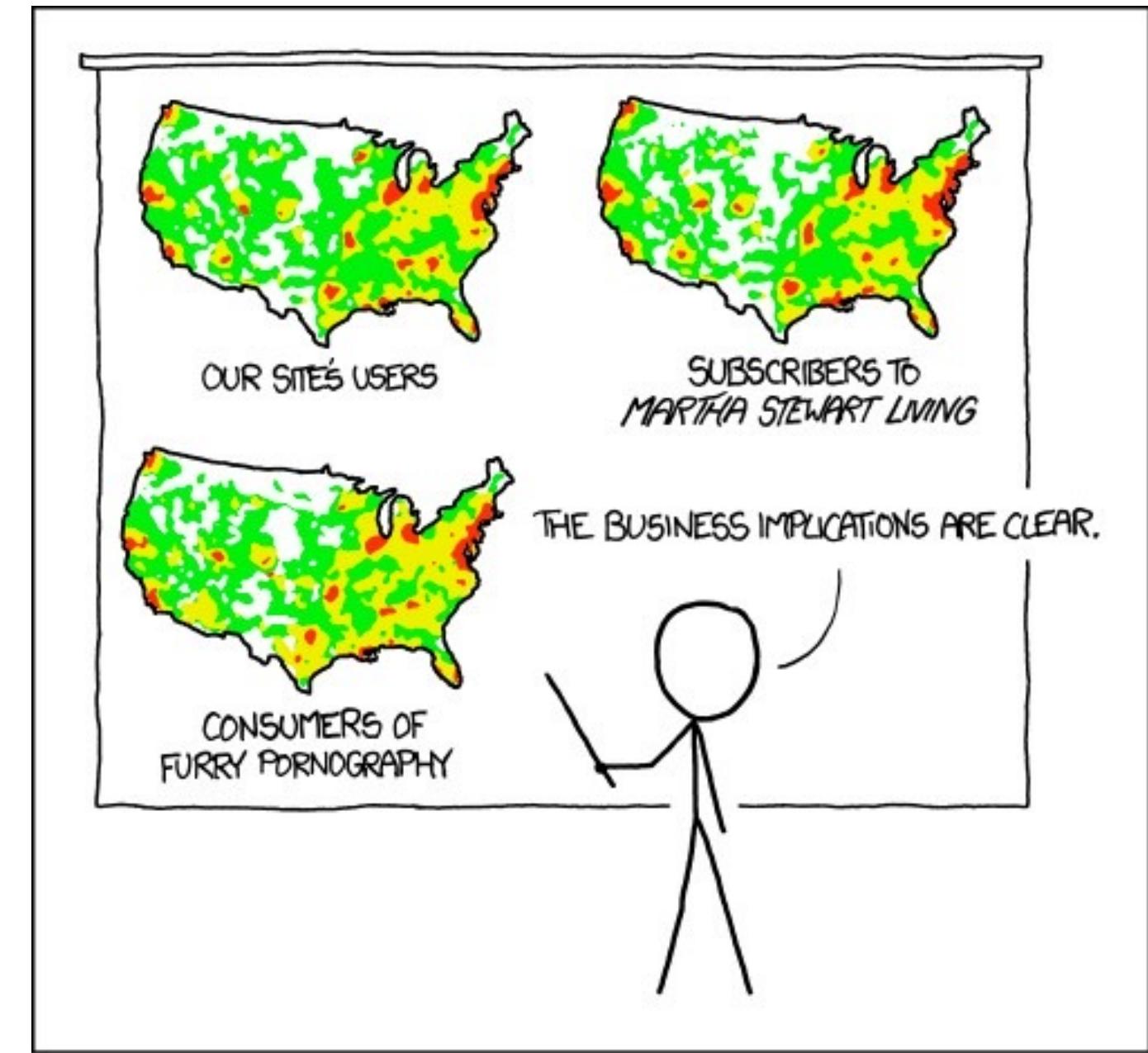
Based on this map, what percentage of the population do you think voted Blue (for Kerry) vs voted for Red (for Bush). *Choose the one that you think is closest to what the map shows.*

- A. 80% Red, 20% Blue
- B. 70% Red, 30 % Blue
- C. 60% Red, 40% Blue
- D. 50% Red, 50% Blue

Beware: Population maps 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
 - but should use normalized values
 - unemployed people per 100 citizens, mean family income
- general issue
 - absolute counts vs relative/normalized data
 - failure to normalize is common error



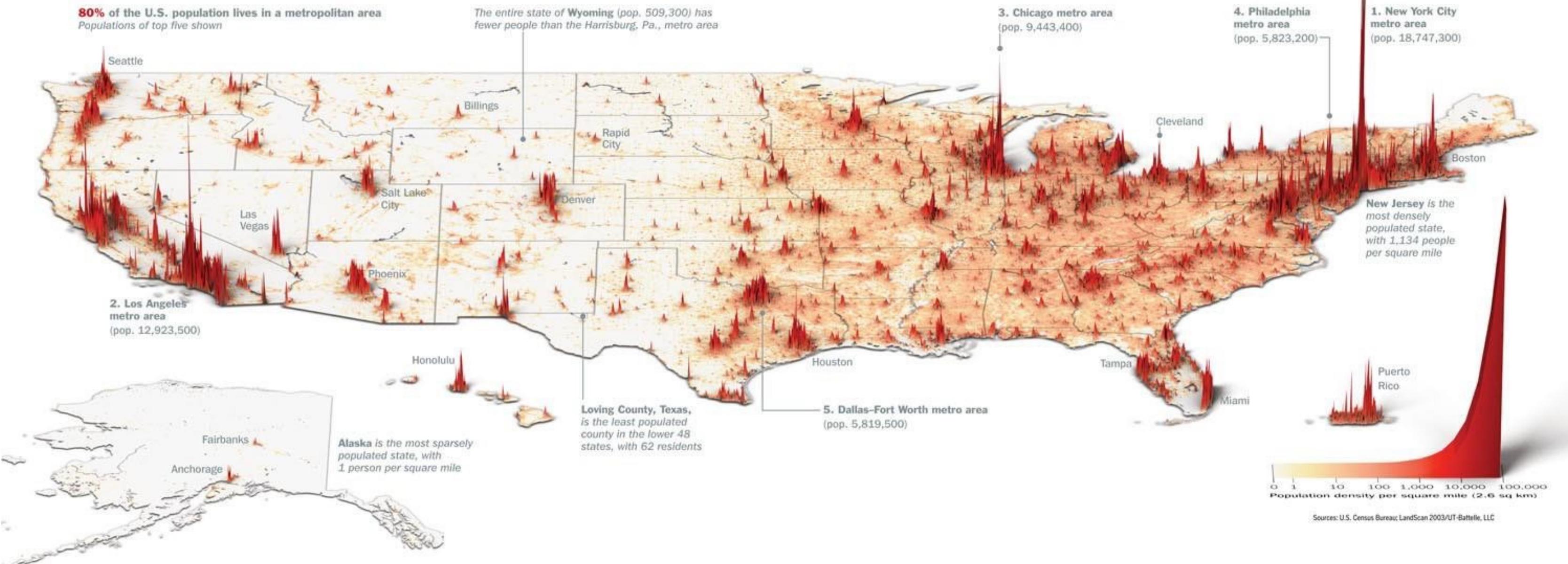
[<https://xkcd.com/1138/>]

Where We Live...

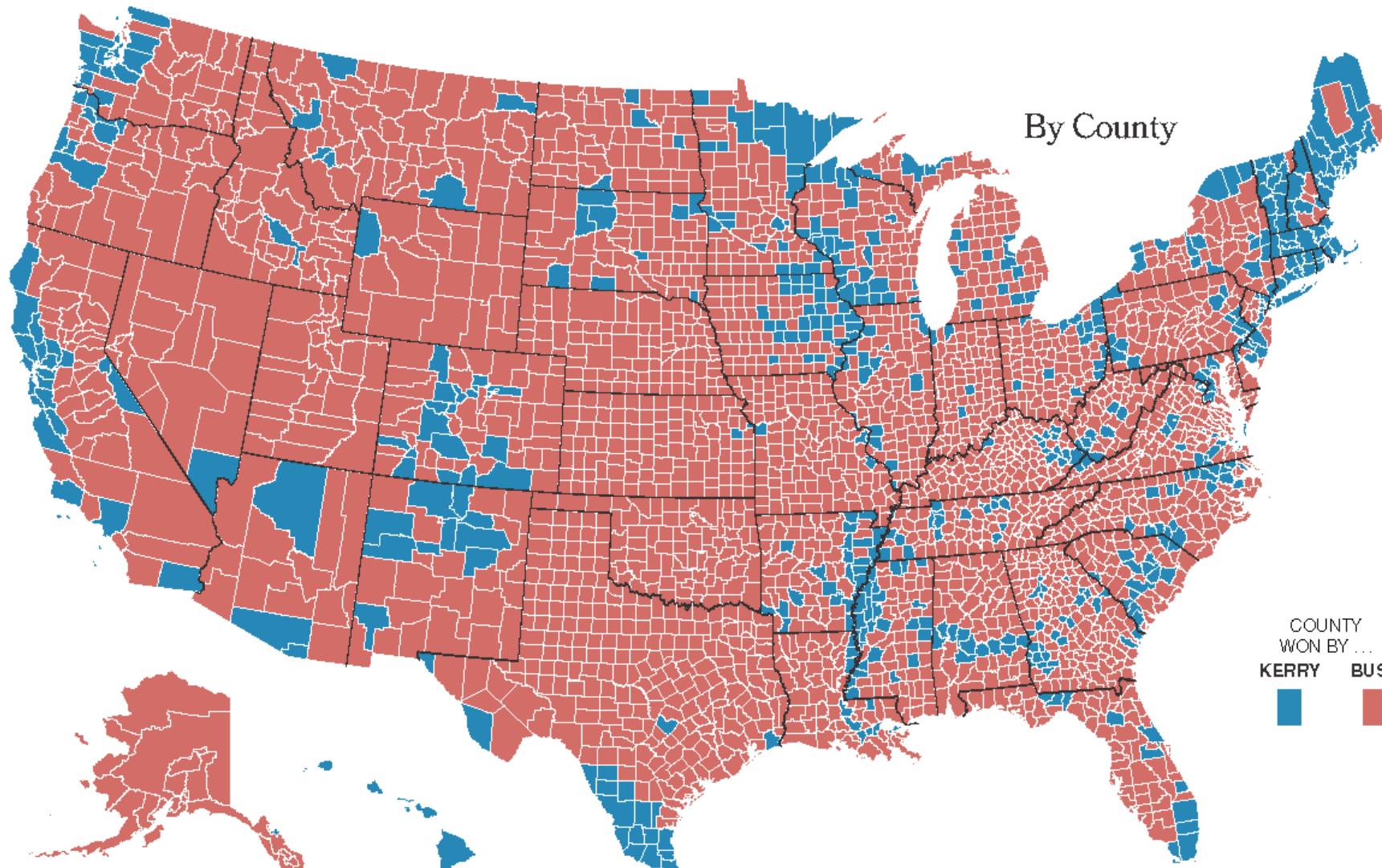
Unlike many developed countries, the U.S. keeps growing. We are also moving south and west. But compared with China or India, the nation is a vast prairie

Our families are getting smaller—with one vital exception. Compared with those of Europe and Japan, the U.S. population is younger and more colorful because of the continued arrival of immigrants and their higher-than-average birthrates. Of the 100 million Americans who will join us in the next 37 years, half will be immigrants or their children. In the next few decades, 97% of the world's population growth will occur in the developing world; the U.S. is the largest developed country in the world that is still growing at a healthy clip. That matters, strategically, economical-

Ala.; Possum Trot, Ky.; or Lonelyville, N.Y. But they are all probably close to someone's idea of paradise. —By Nancy Gibbs



Presidential Elections in United States



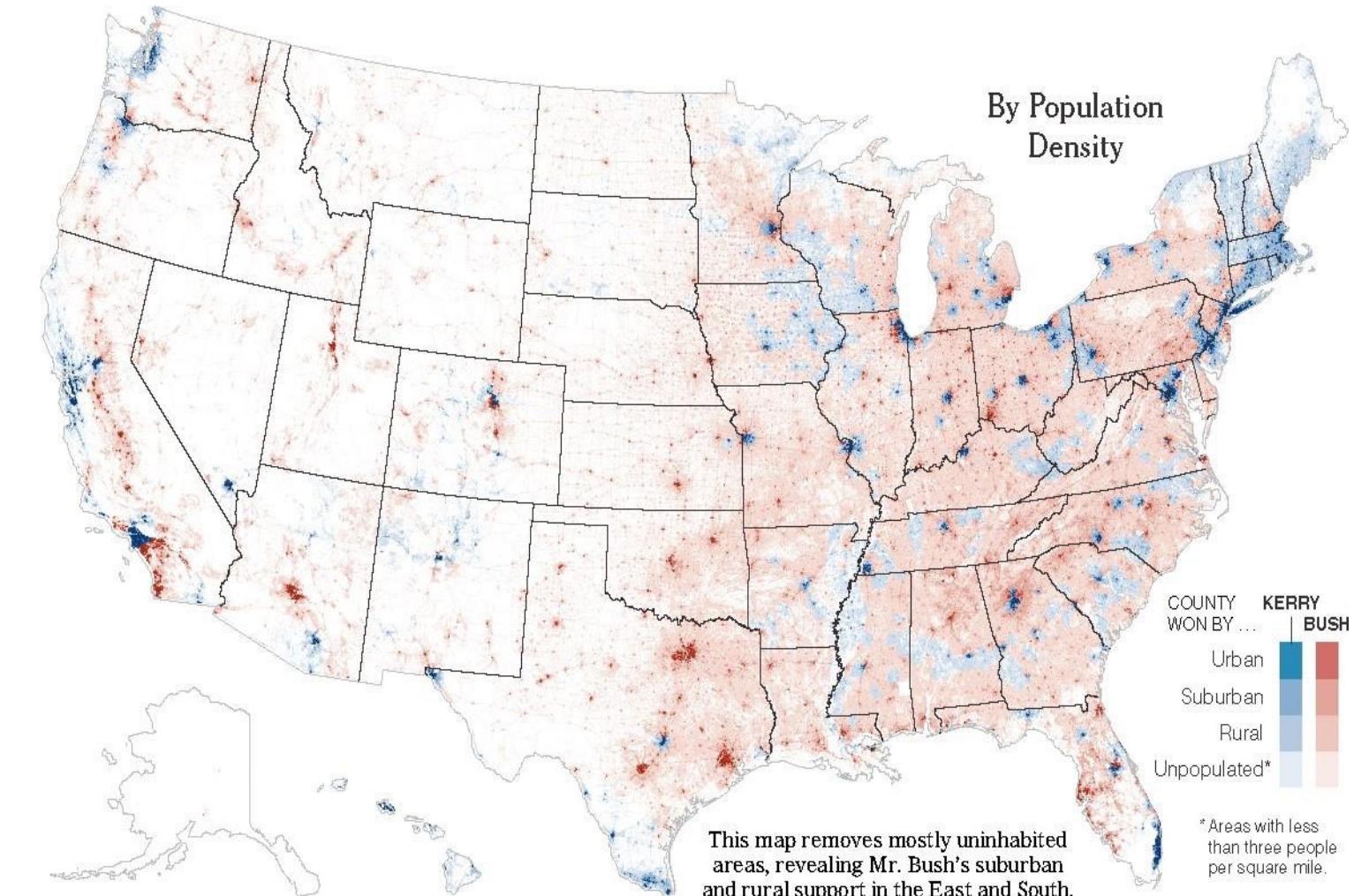
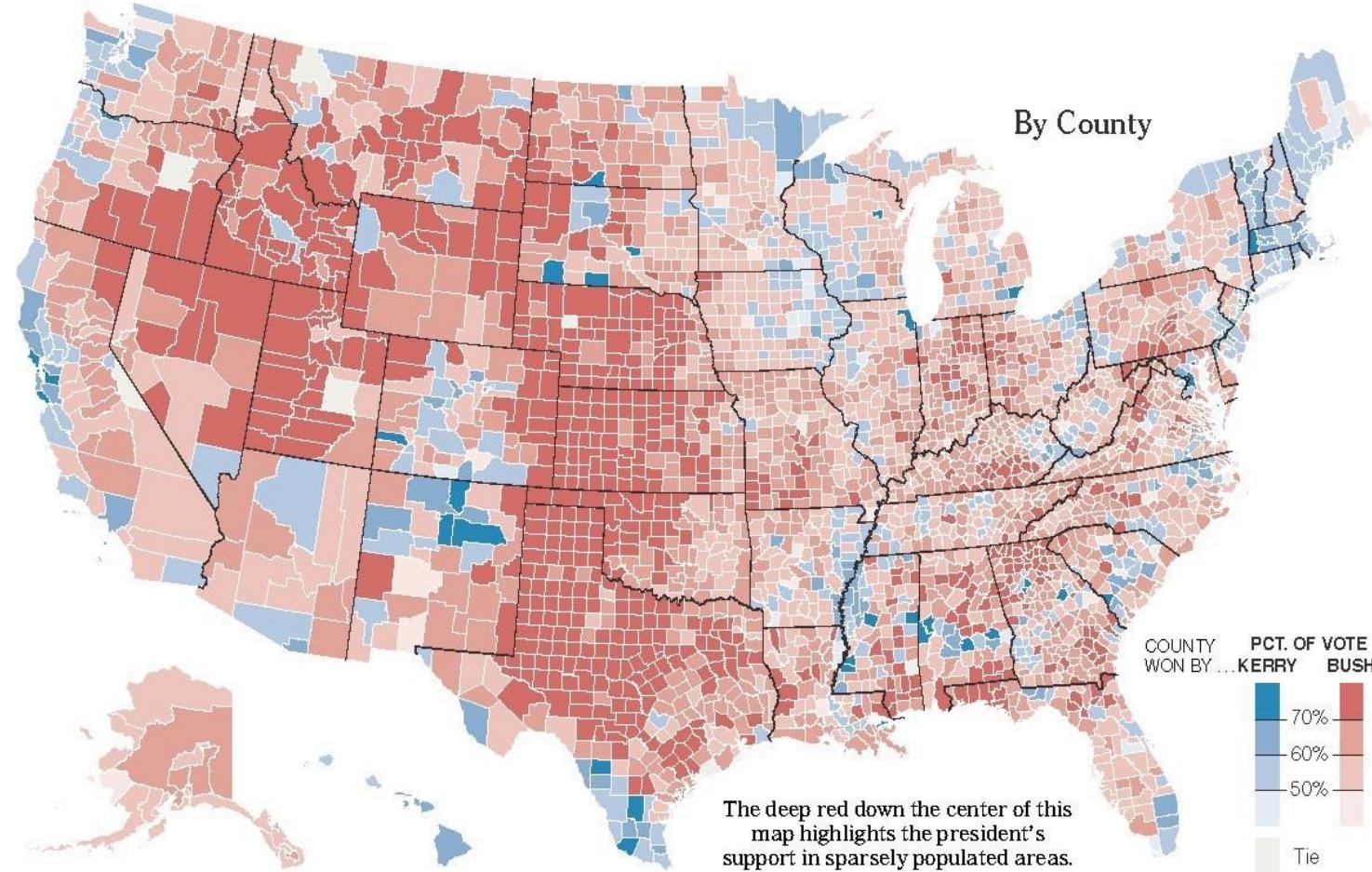
Magnitude of Effect vs Perceived Effect

2004 Popular Vote



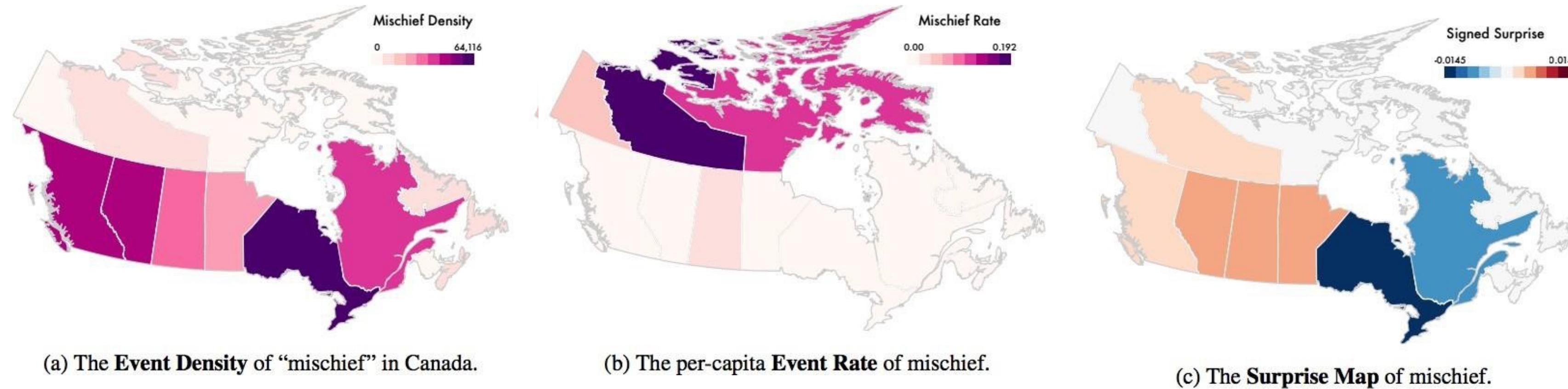
Amount of red and blue shown on map





Approach: Use a Prior, show difference. Which province is safest?

mischief "is a category of property crime: it includes things like vandalism and graffiti, where the intent is neither to steal anything nor hurt anybody."



Provinces with either unusually high/low per capita rates, or large populations with significant deviations from the average.

Choropleth maps: 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 & 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 wrt attribute value
 - large regions appear more important than small ones
 - color palette choice has a huge influence on the result

Idiom: **Symbol maps**

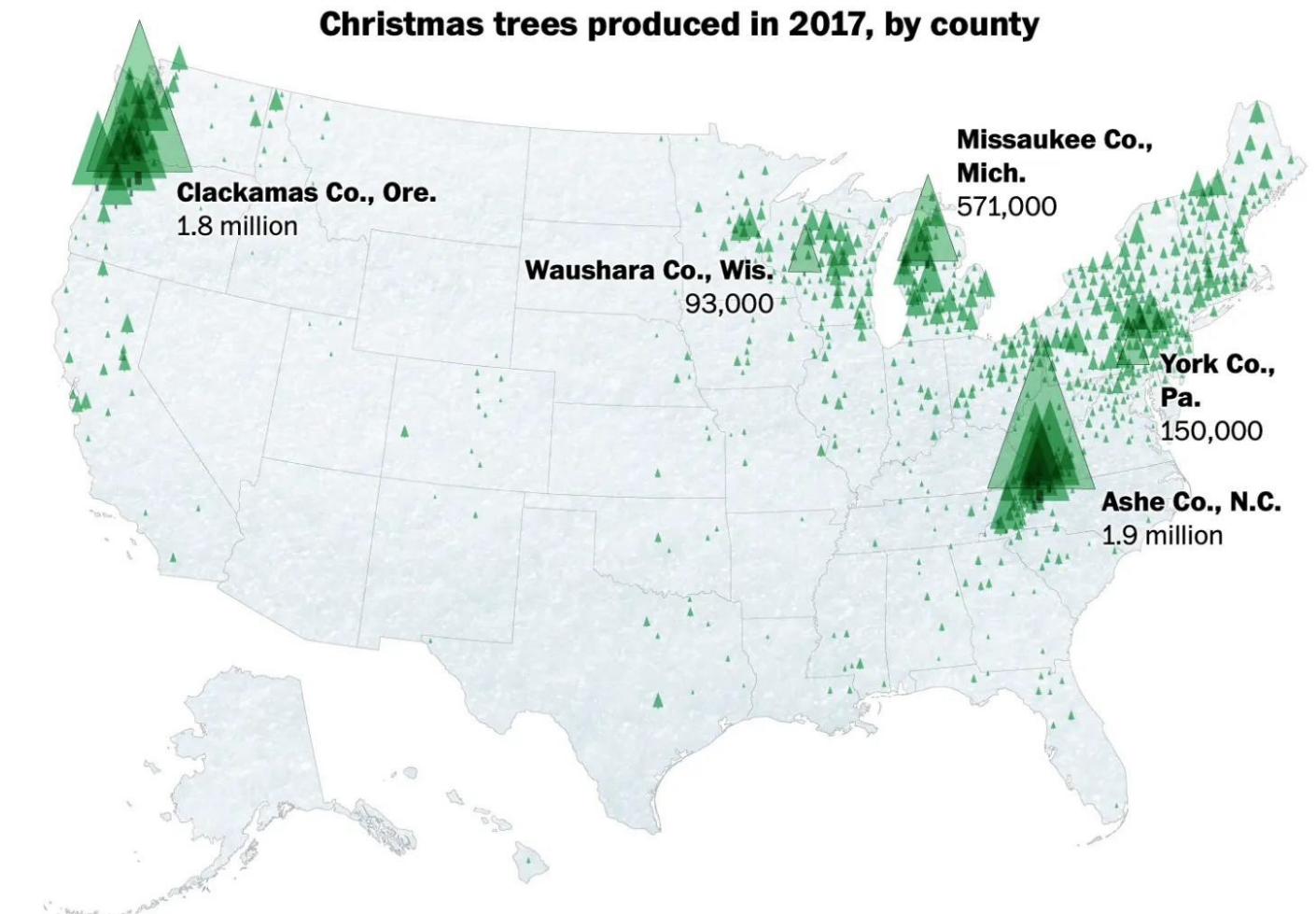
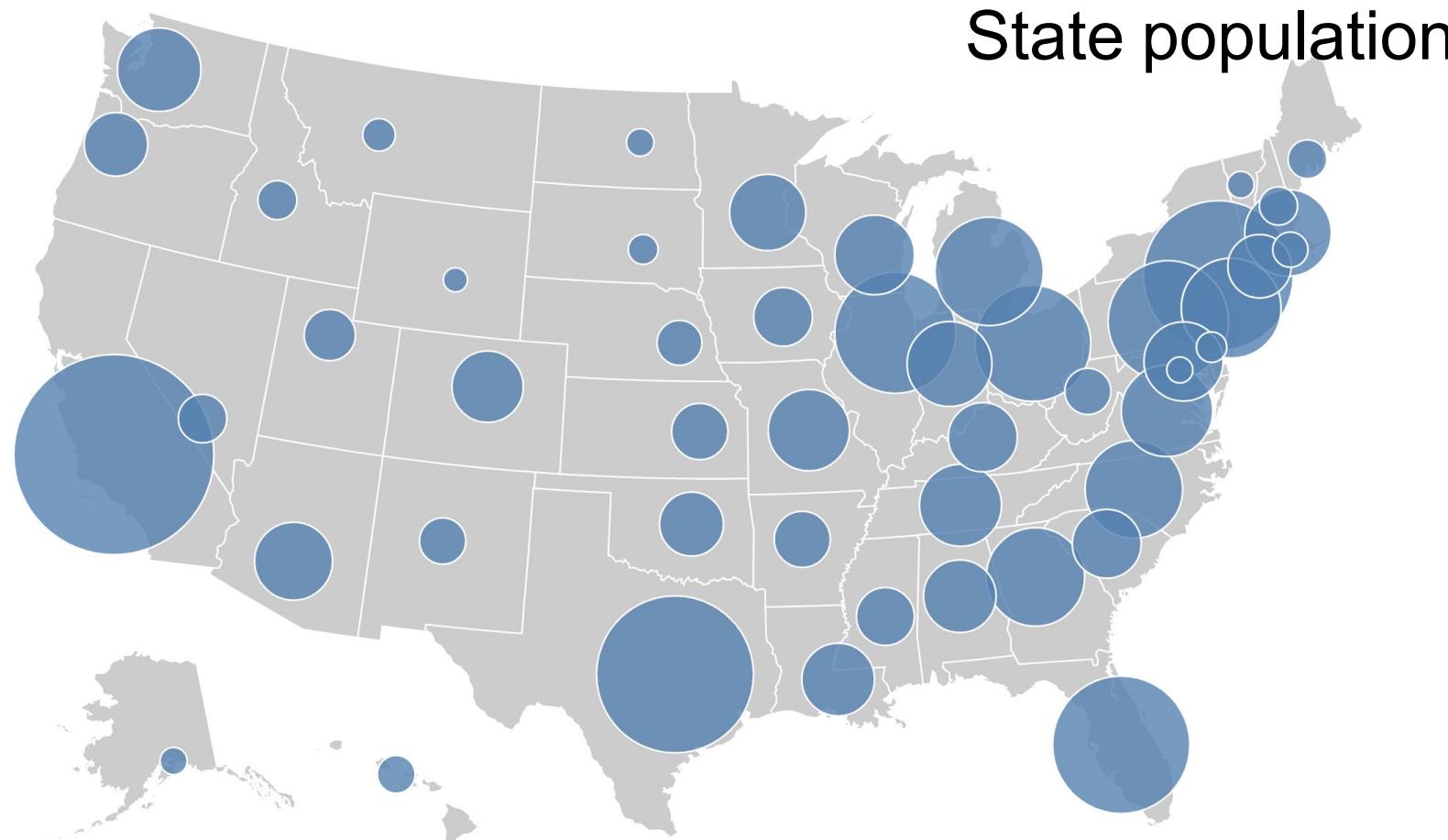
symbol is used to represent aggregated data (mark or glyph)

allows use of size and shape and color channels

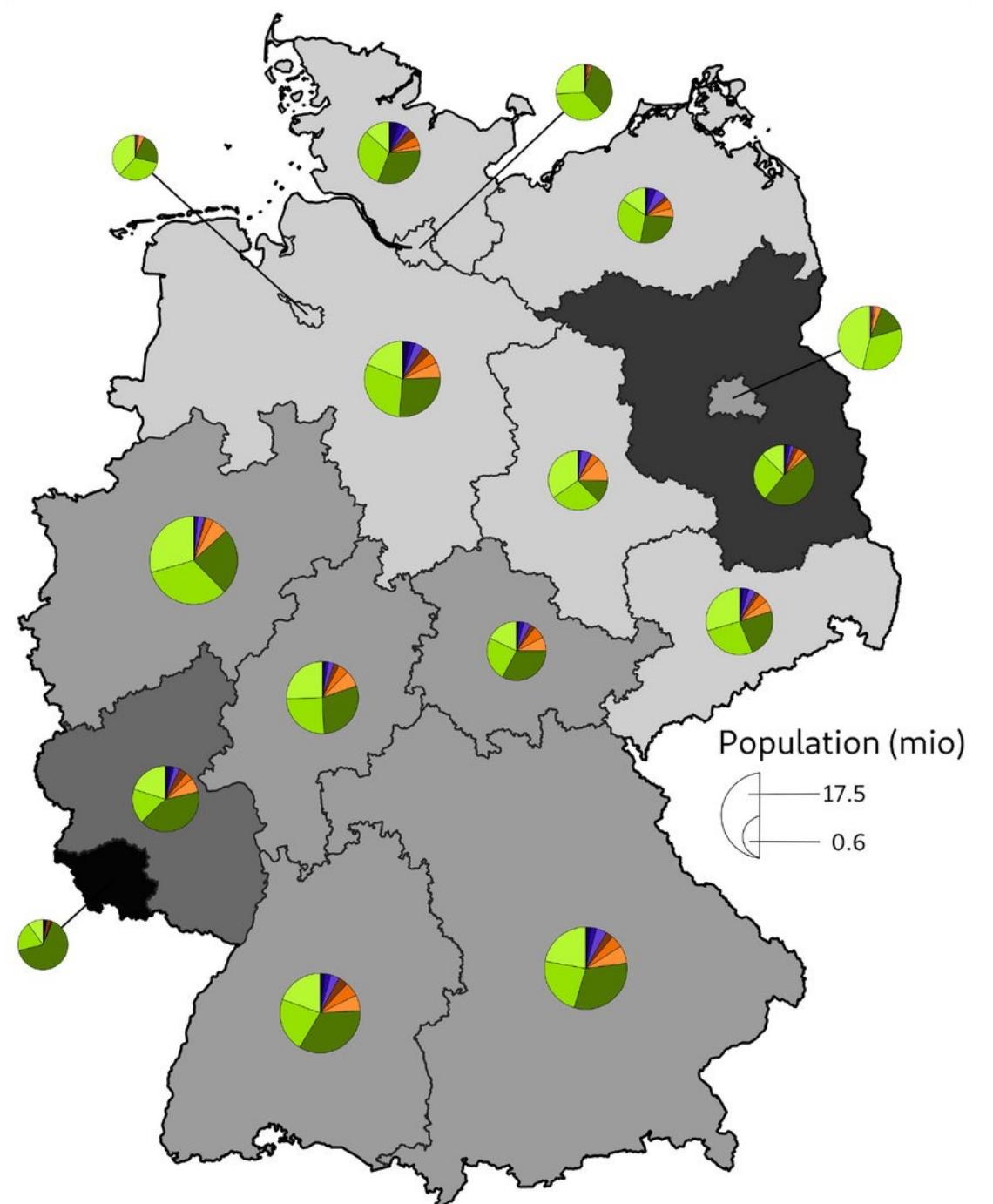
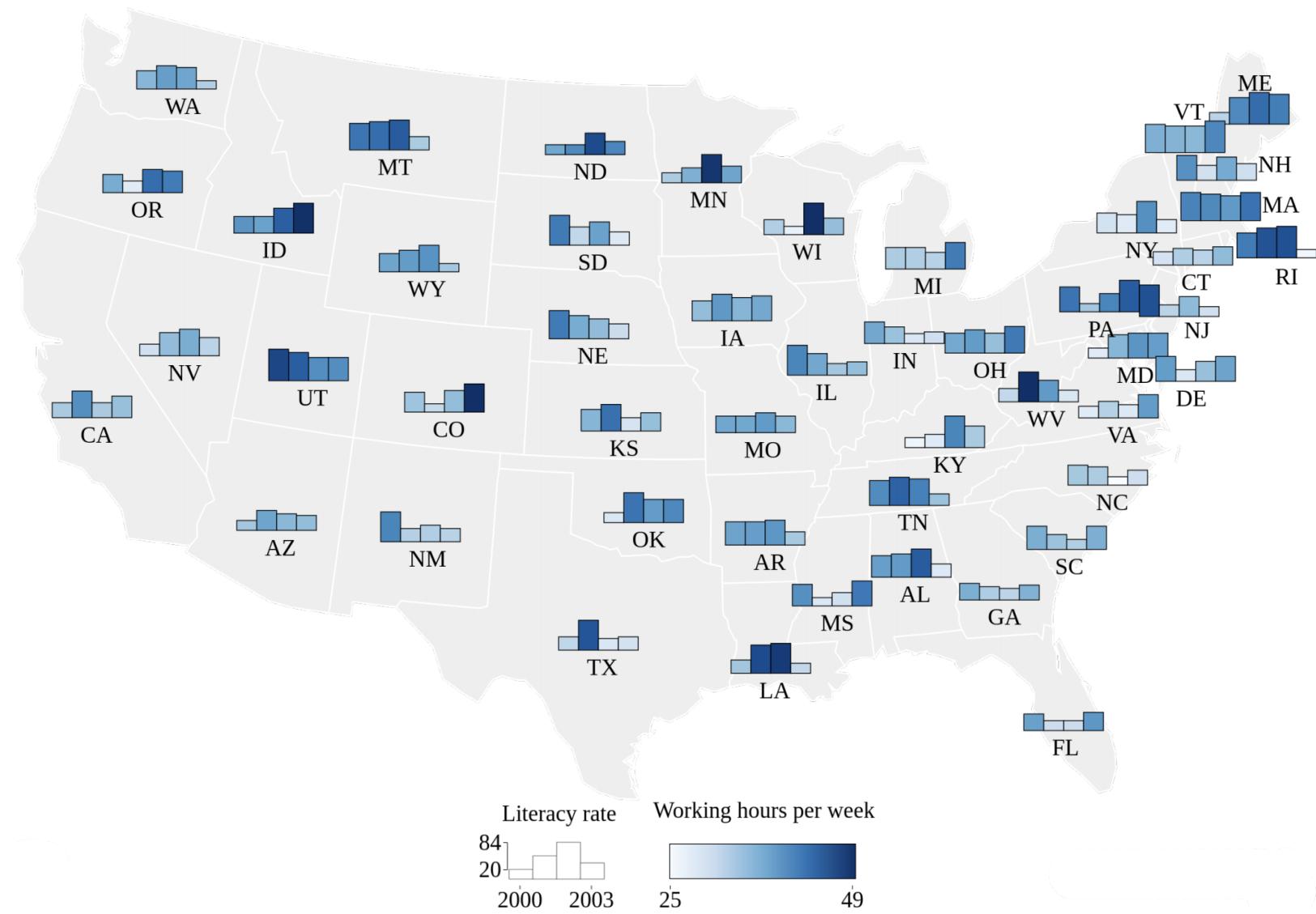
aka proportional symbol maps, graduated symbol maps

keep original spatial geometry in the background

often a good alternative to choropleth maps

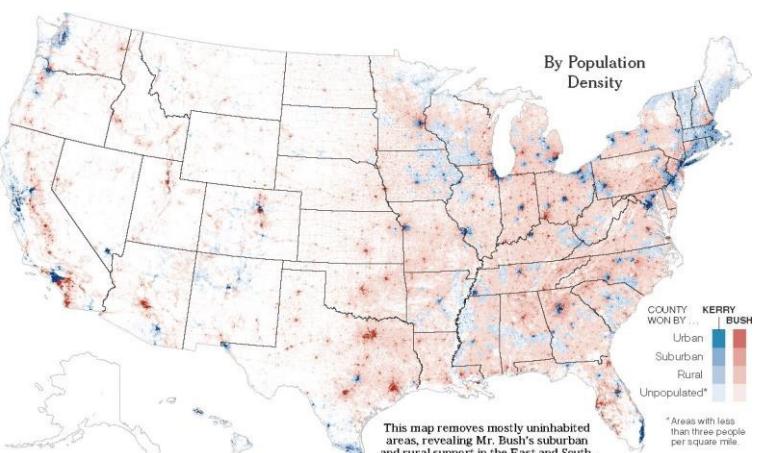
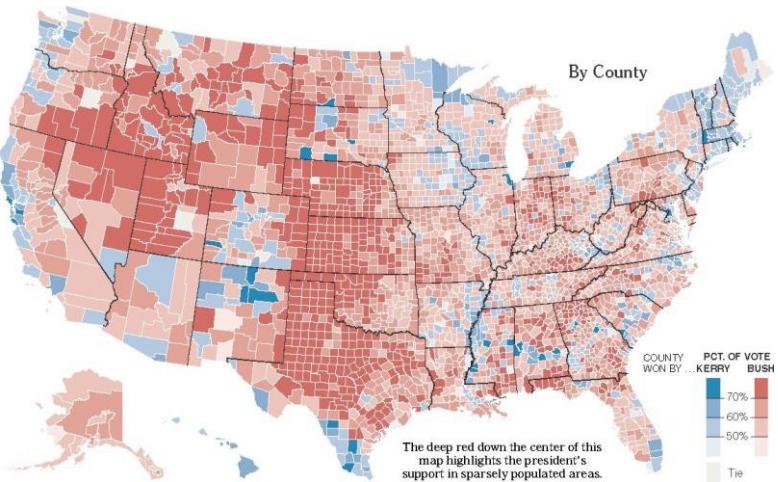
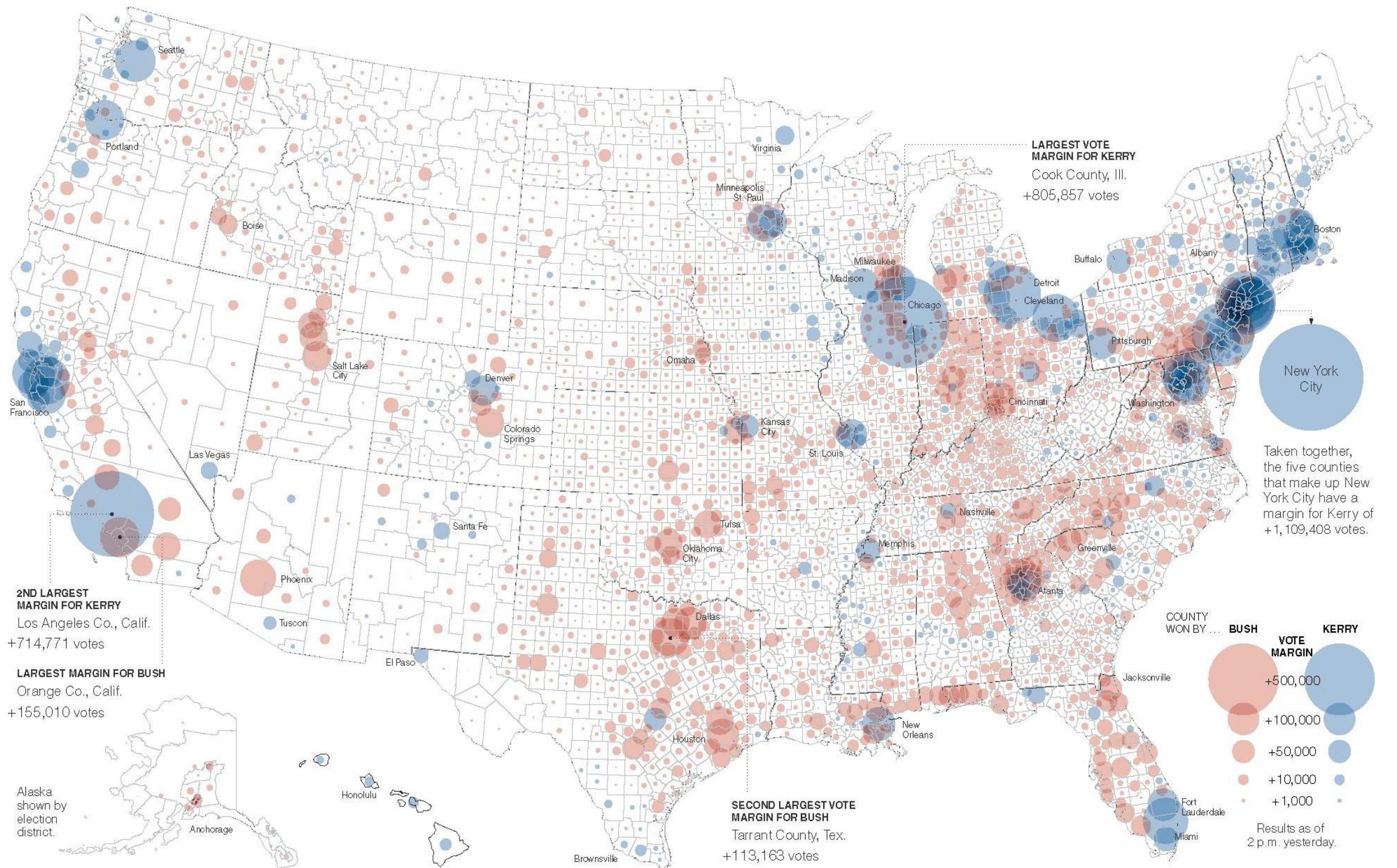


<https://www.washingtonpost.com/nation/2019/12/12/where-christmas-trees-come/>



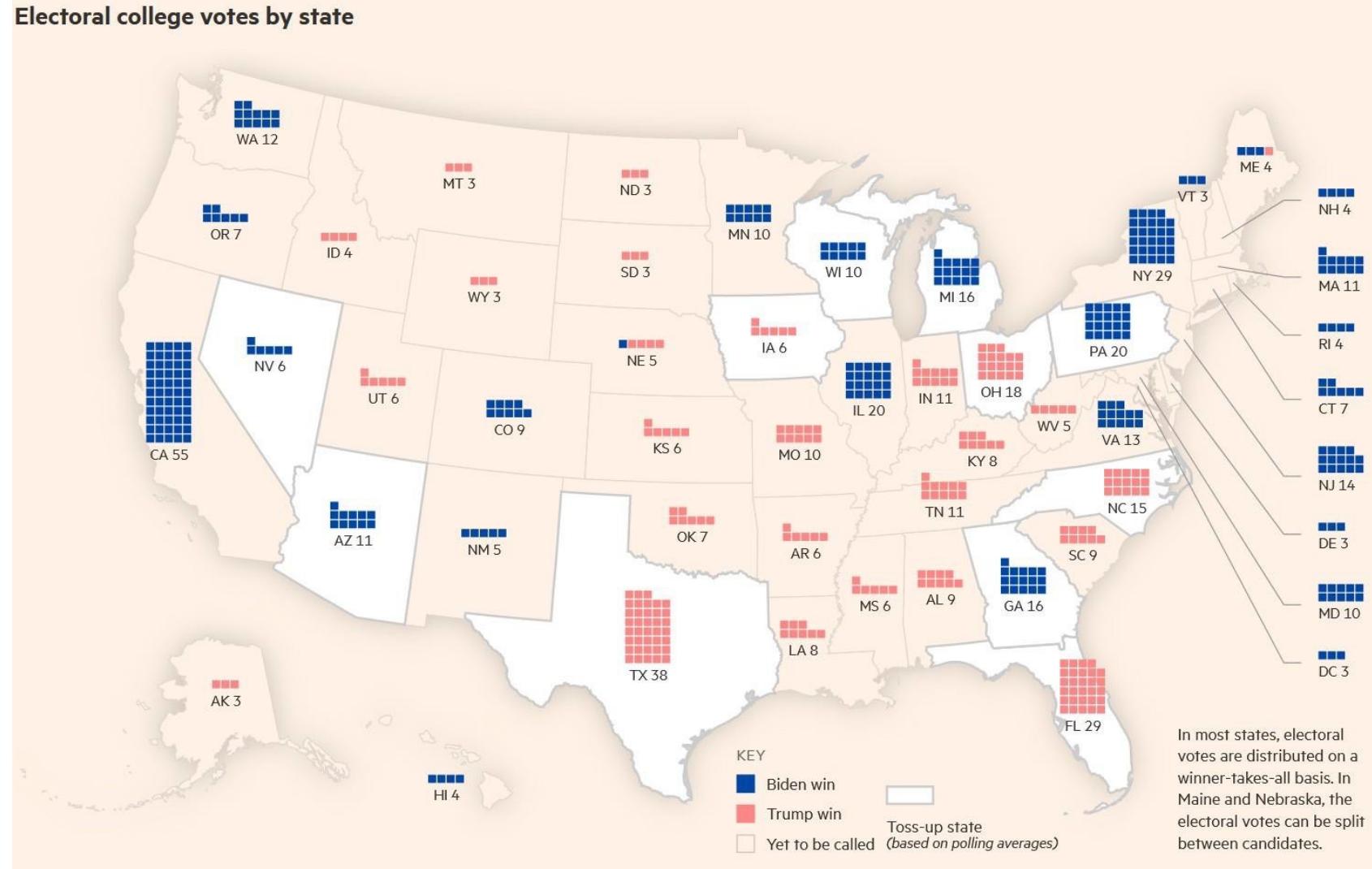
<https://hal.science/hal-02320617v1/file/main.pdf>

https://www.researchgate.net/figure/Green-space-availability-in-Germany-by-federal-states-States-are-shaded-by-the-fig3_375816455

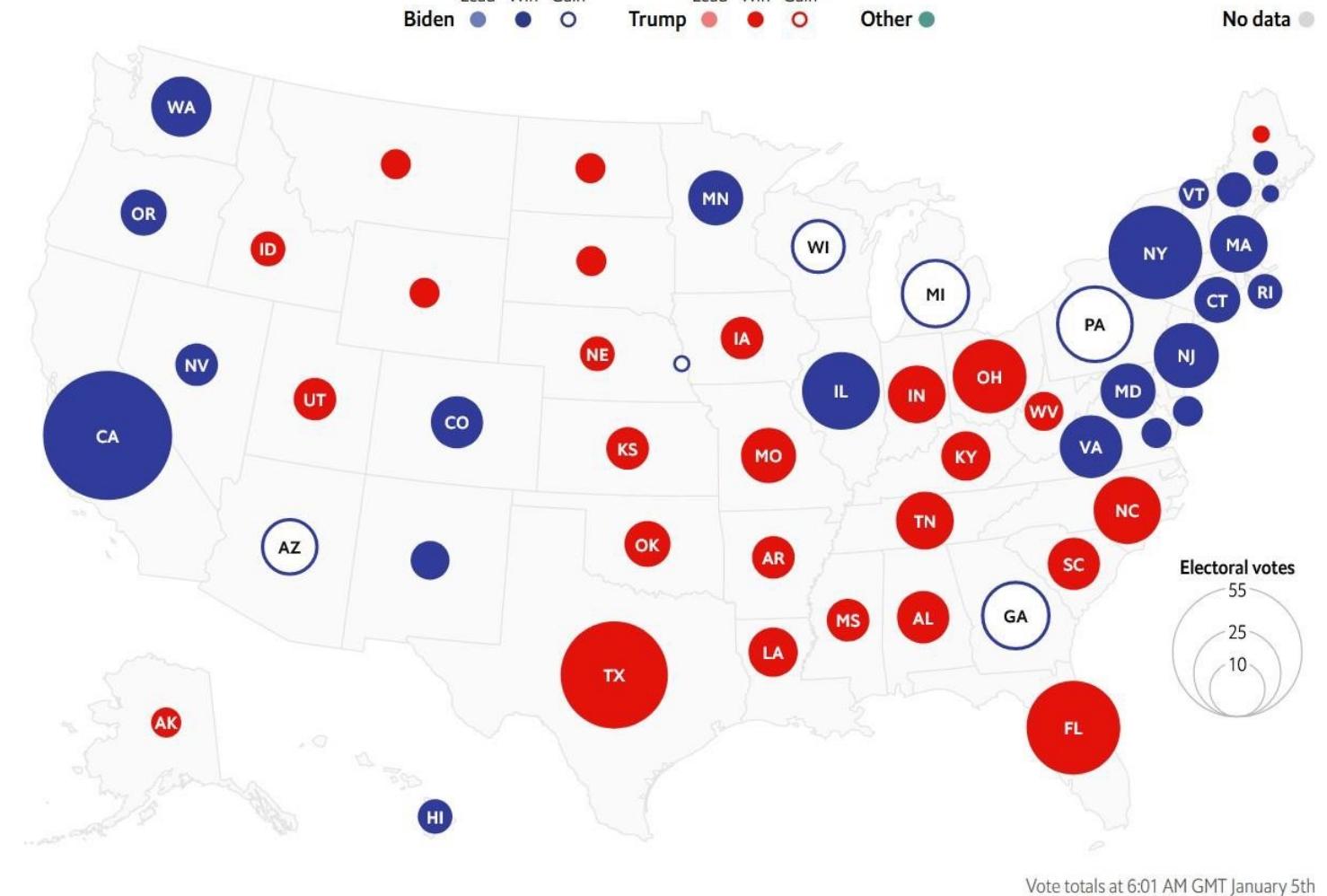


Matthew Ericson, NY Times

Electoral college votes by state



Biden Lead Win Gain
Trump Lead Win Gain
Other ● No data ●

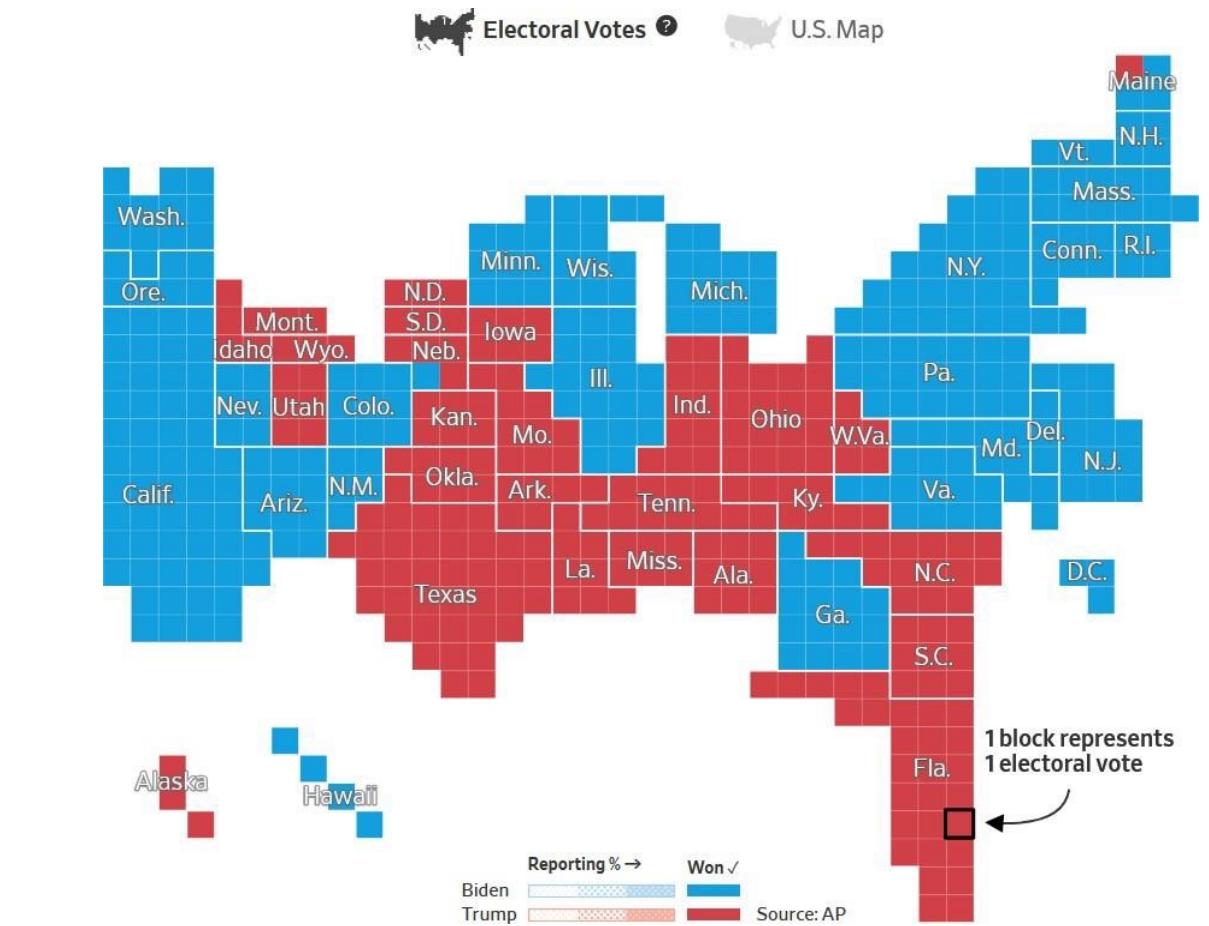
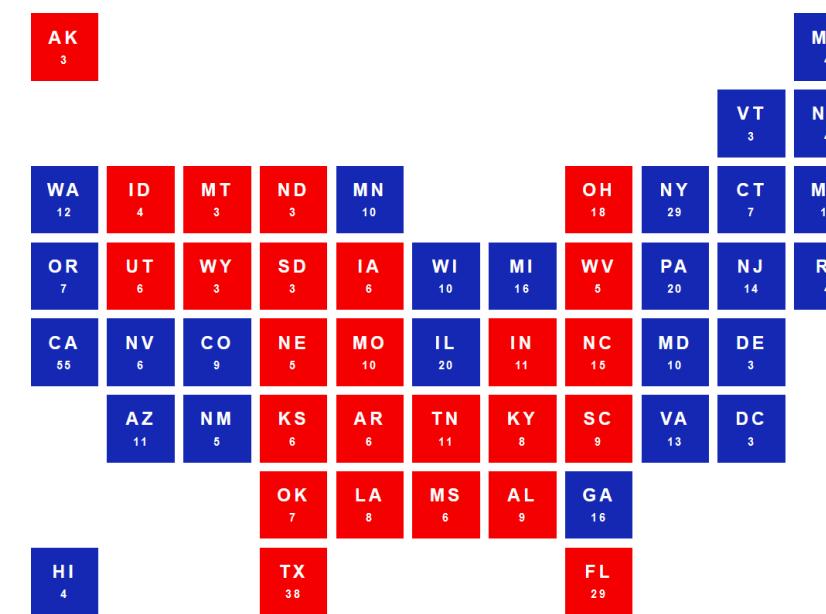
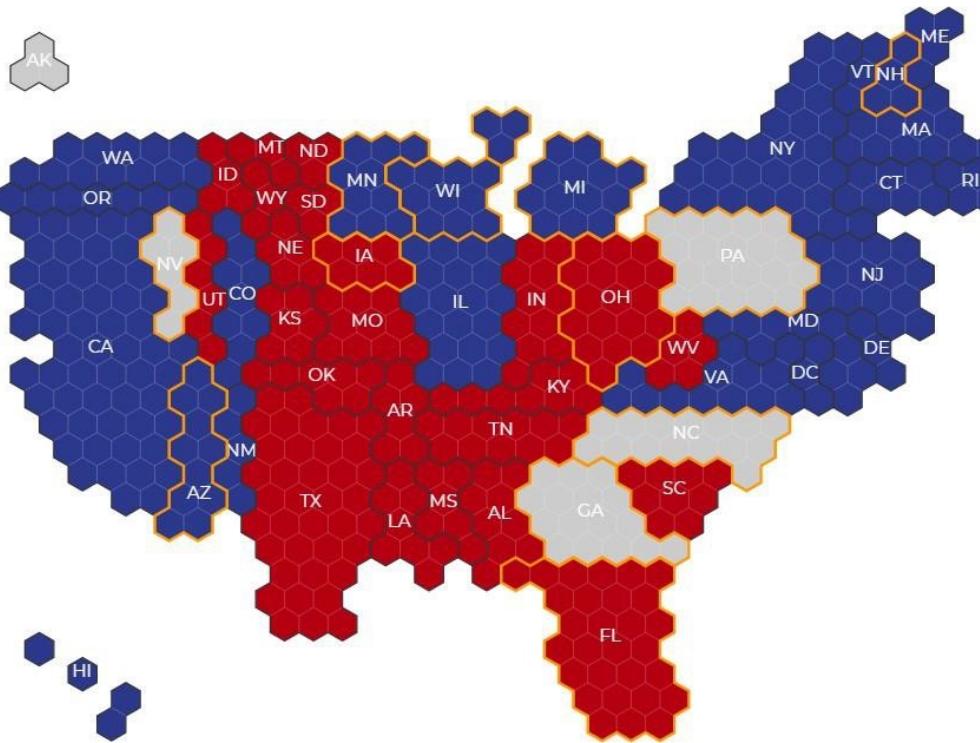


Symbol map: Pros & cons

- 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

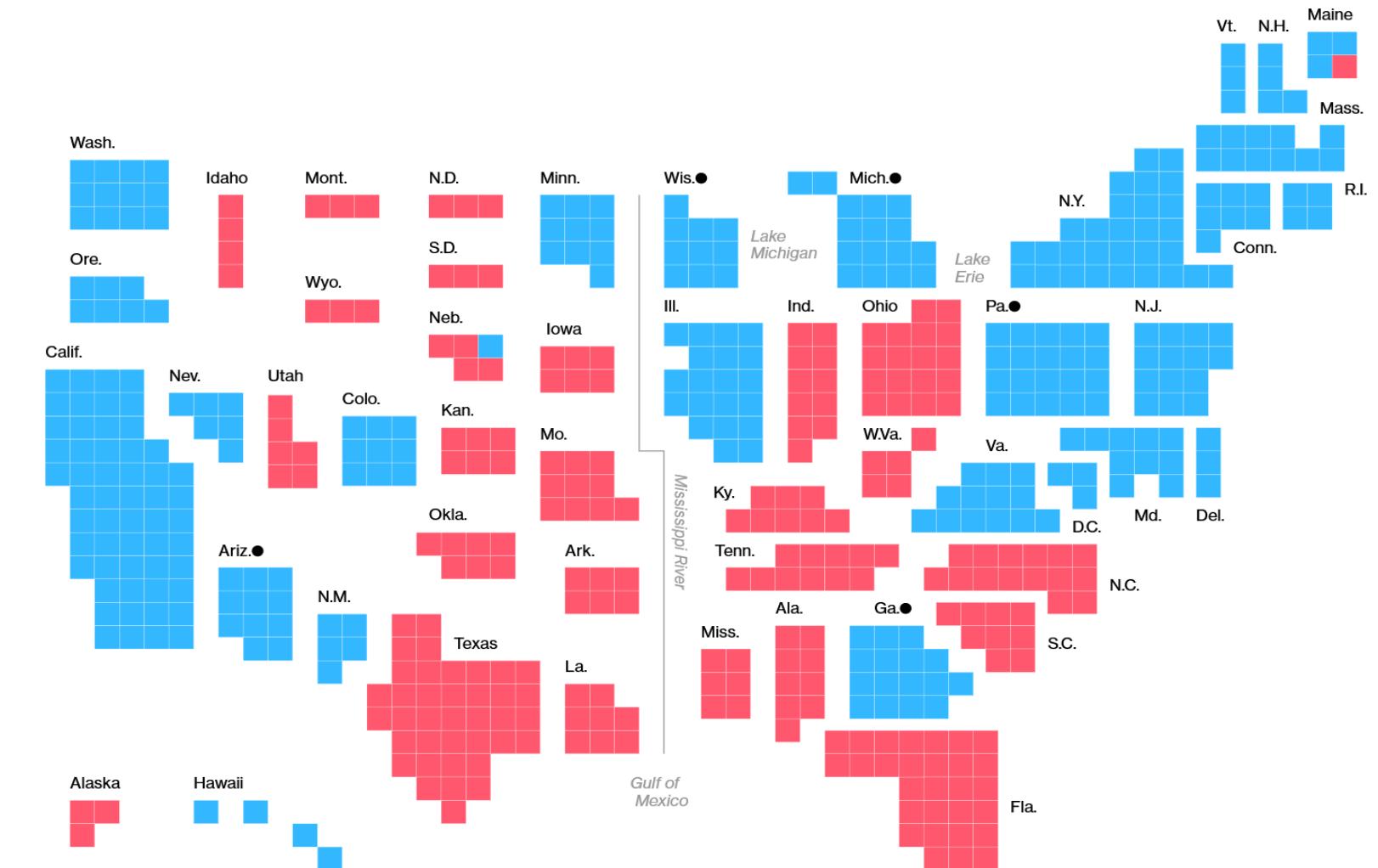
Idiom: Grid Cartogram

- uniform-sized shapes arranged in rectilinear grid
- maintain approximate spatial position and arrangement



Idiom: **Contiguous cartogram**

- We are changing the size of regions on the map
- Compromise between geospatial accuracy and quality of data encoding.



<https://www.bloomberg.com/graphics/2020-us-election-results/>

Idiom: **Contiguous cartogram**

Also known as anamorphic map

interlocking marks:

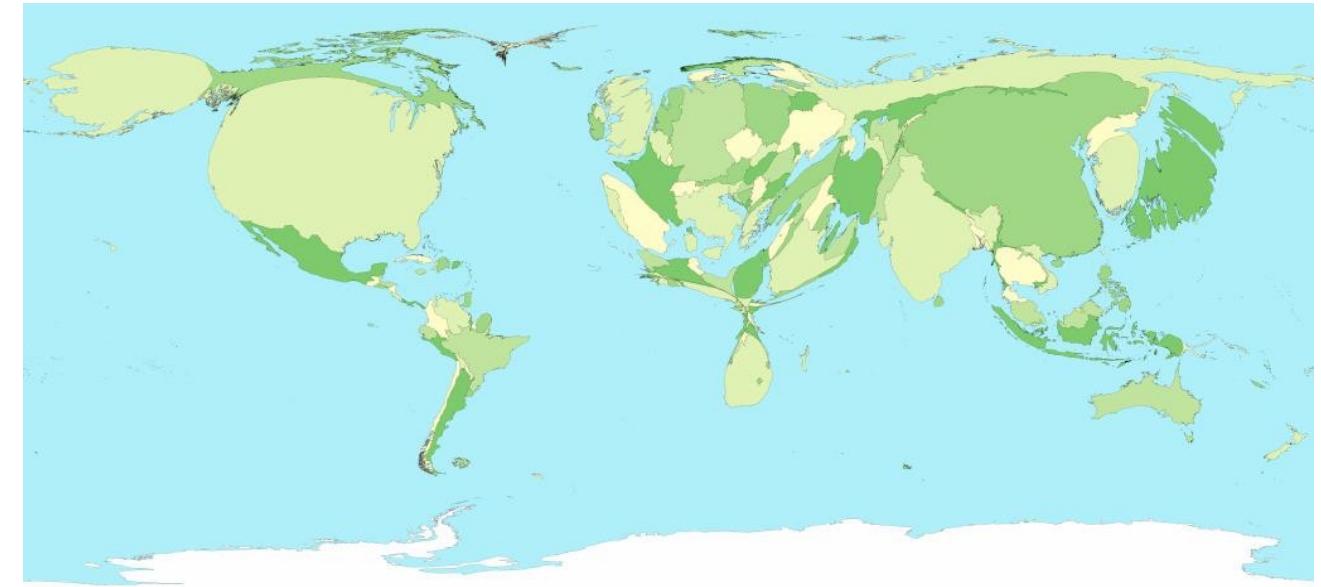
- shape, area, and position coded

- derive new interlocking marks

- based on combination of original interlocking marks and new quantitative attribute

- 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 neighbours

Greenhouse Emissions



Population



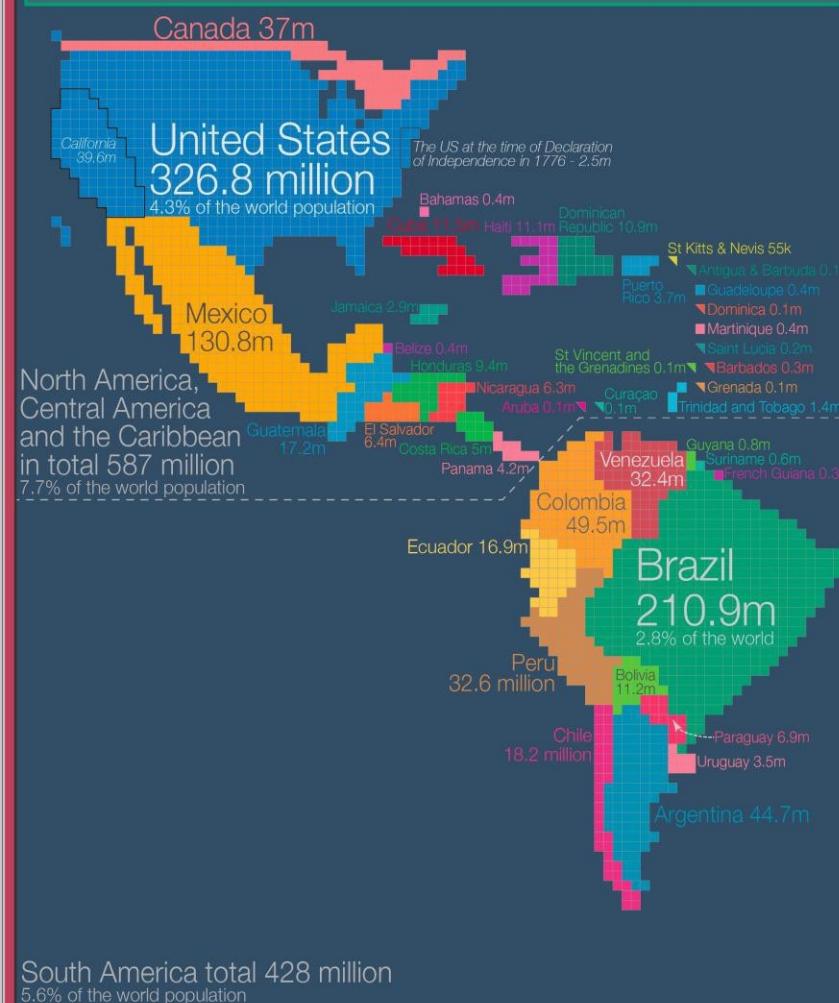
World Population in 2018

The country's size in this map represents the size of the population.
Each square [■] represents 500,000 people.
All 15,266 squares show where the world's 7.633 billion people live.

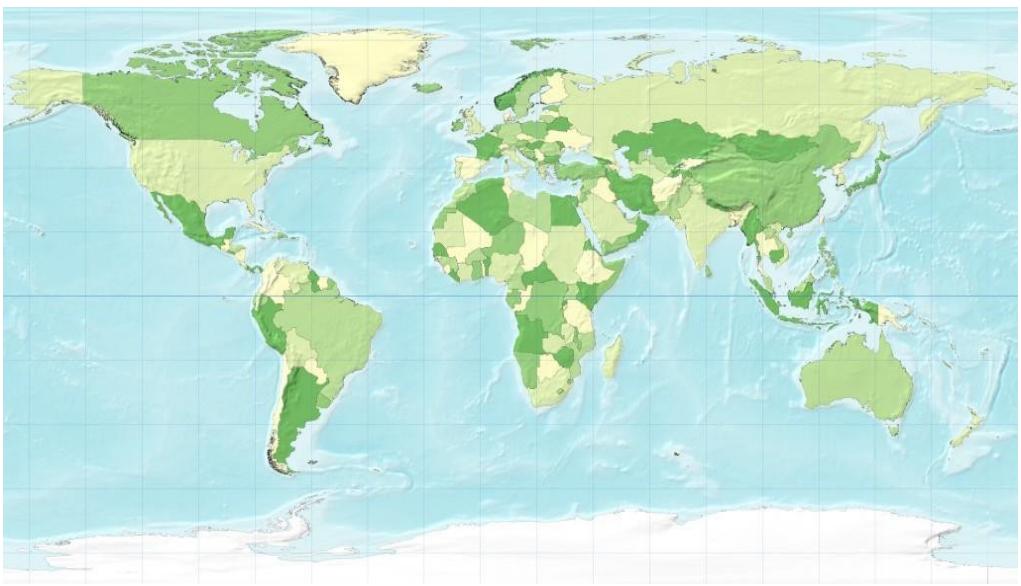
by Max Roser for OurWorldinData.org – the free online publication that presents the data and research on how the world is changing.

Population data from the *United Nations Population Division*
Version 3 (October 2018).

Licensed under CC-BY-SA

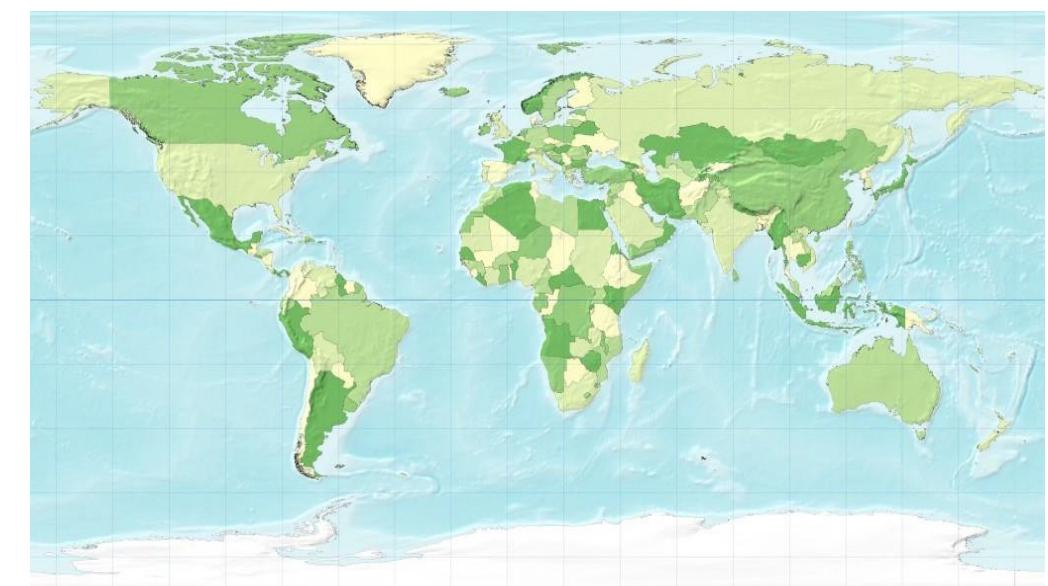
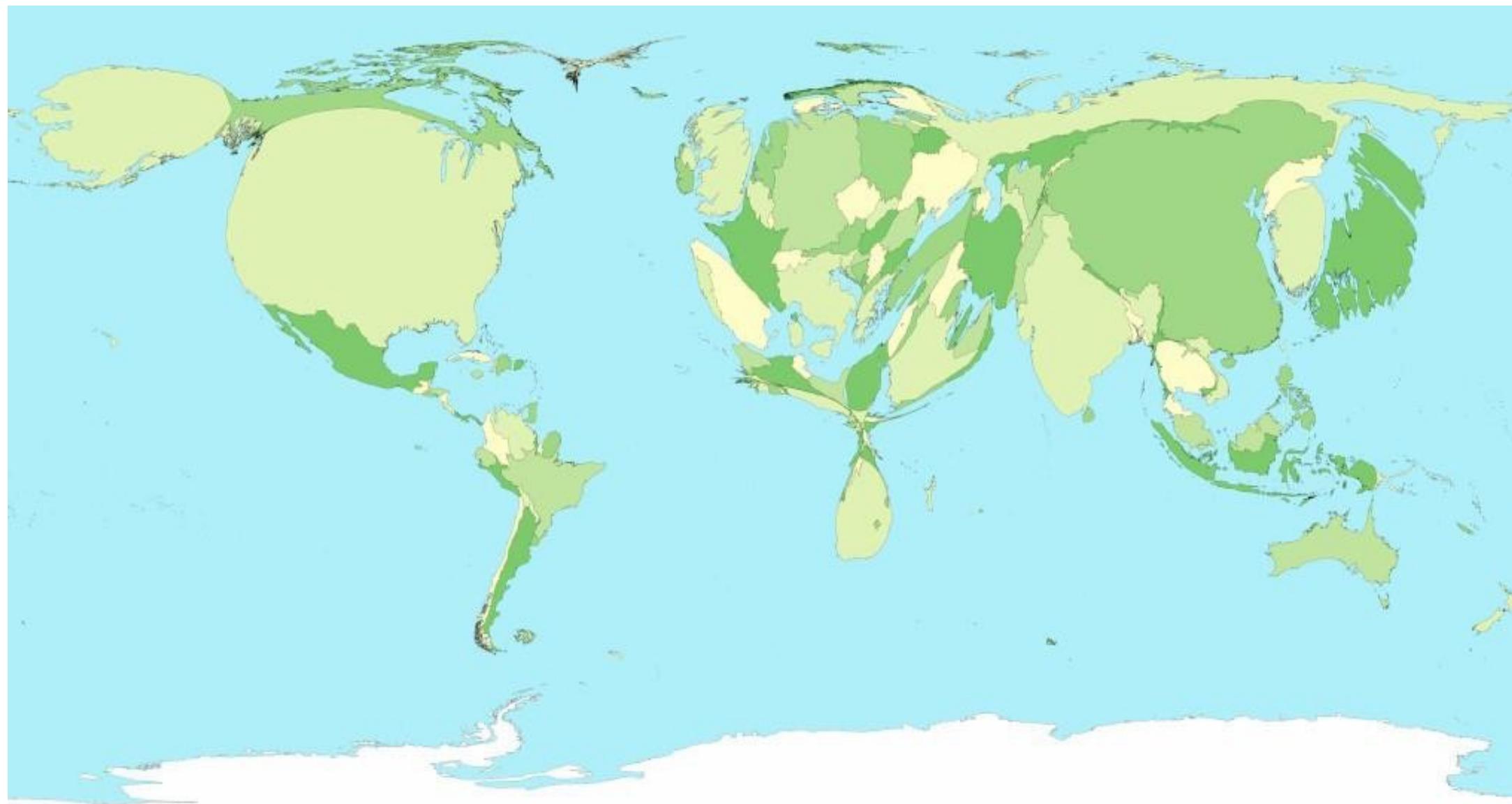


Population



Mark Newman, Univ.
Michigan

Greenhouse Emissions



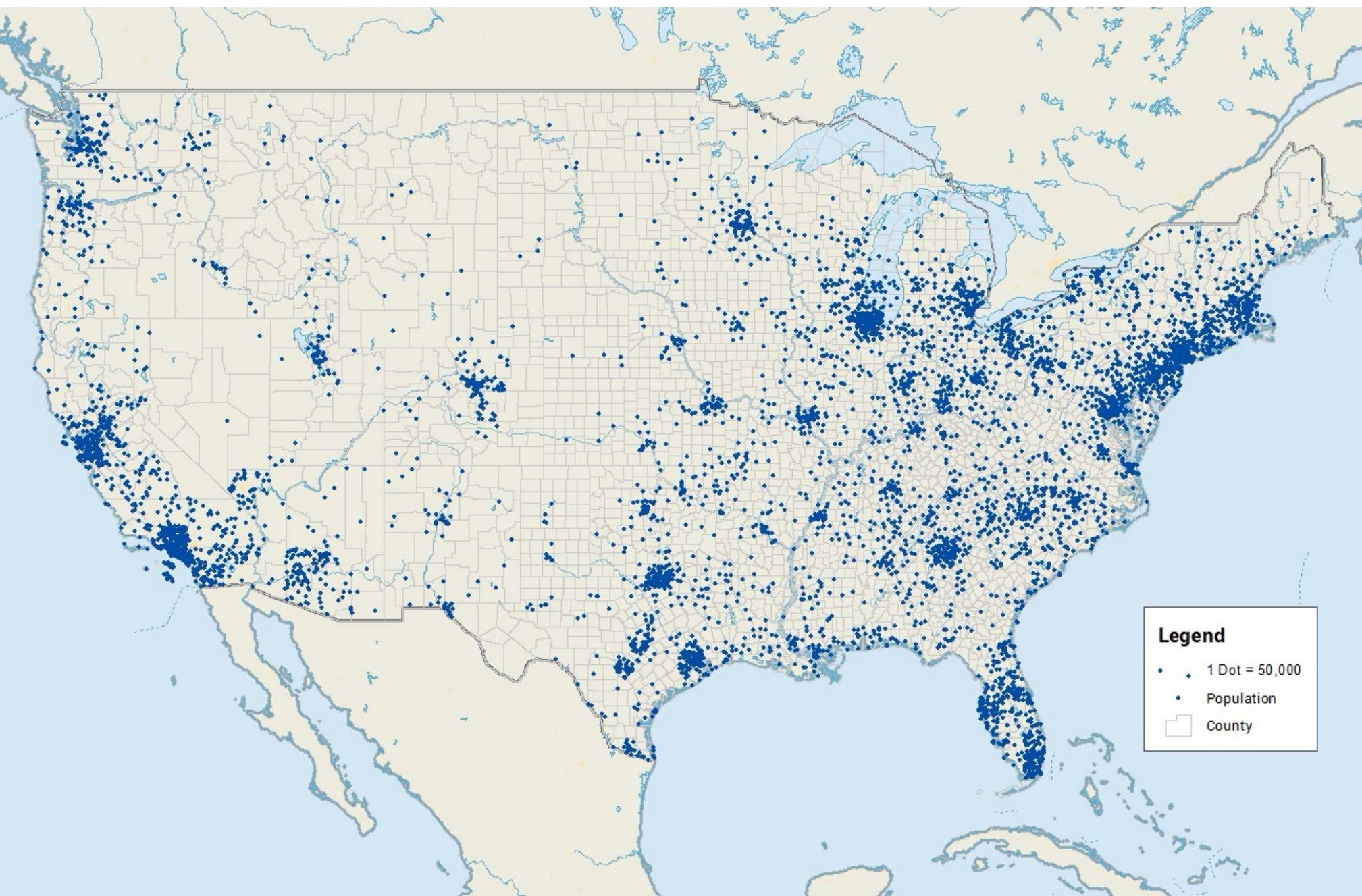
Mark Newman, Univ.
Michigan

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

Idiom: **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