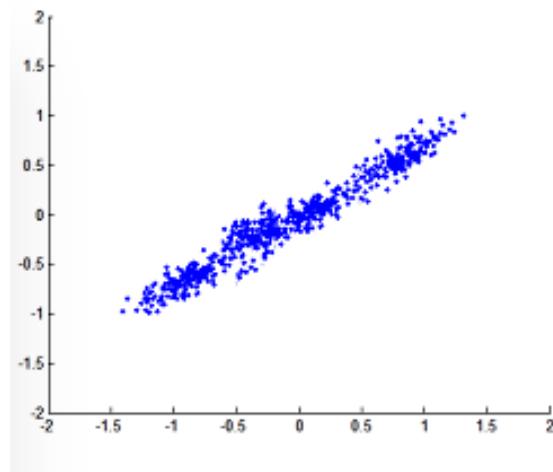


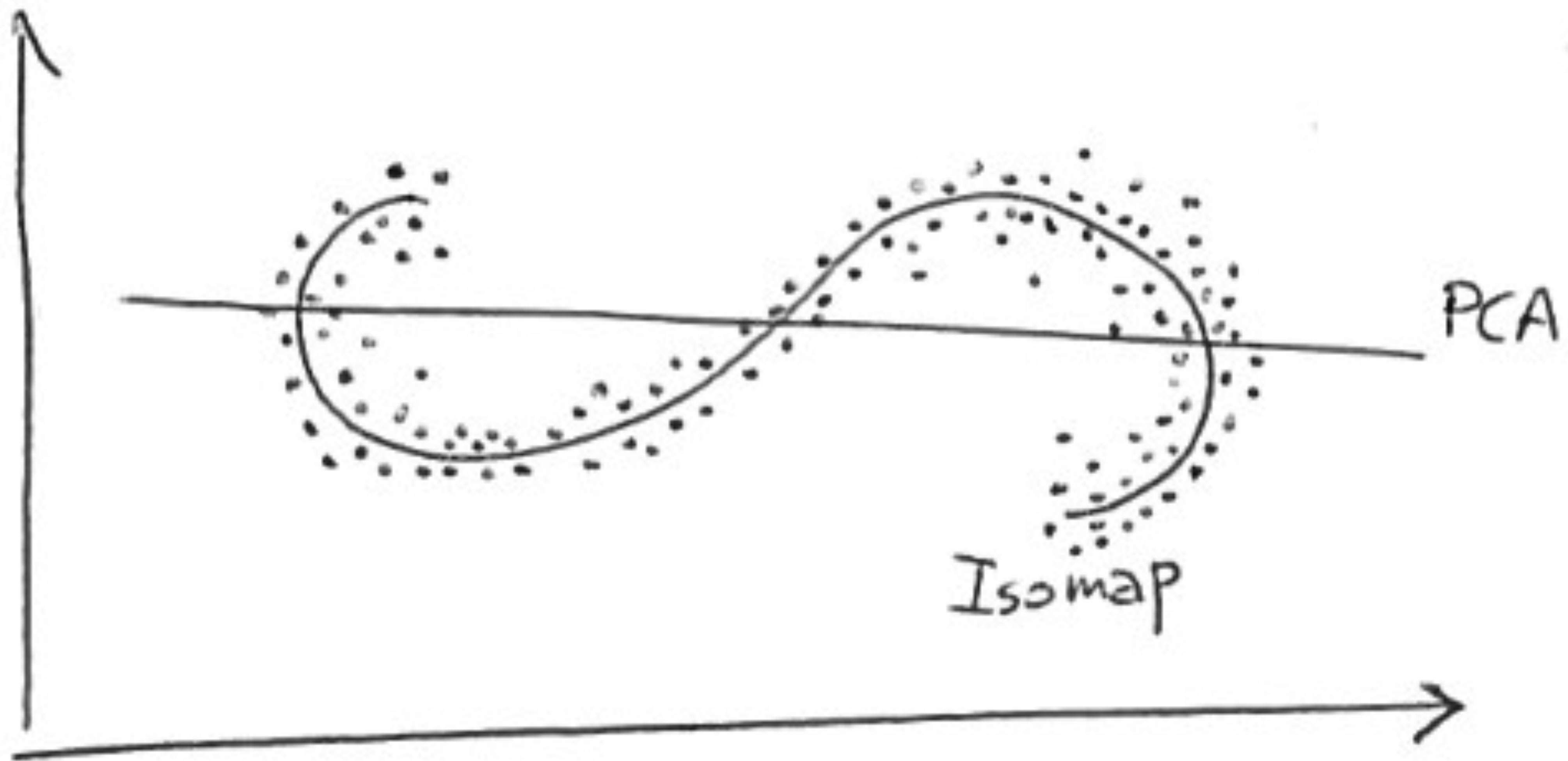
# Quiz

- To obtain the principal components (in PCA), you first center the data, calculate \_\_\_\_\_ matrix, and then calculate the \_\_\_\_\_ vectors.
- The first principal component is the \_\_\_\_\_ vector that corresponds to the \_\_\_\_\_ \_\_\_\_\_ value.
- We have a bivariate data like this. Can you sketch it and annotate the two principal components on it?

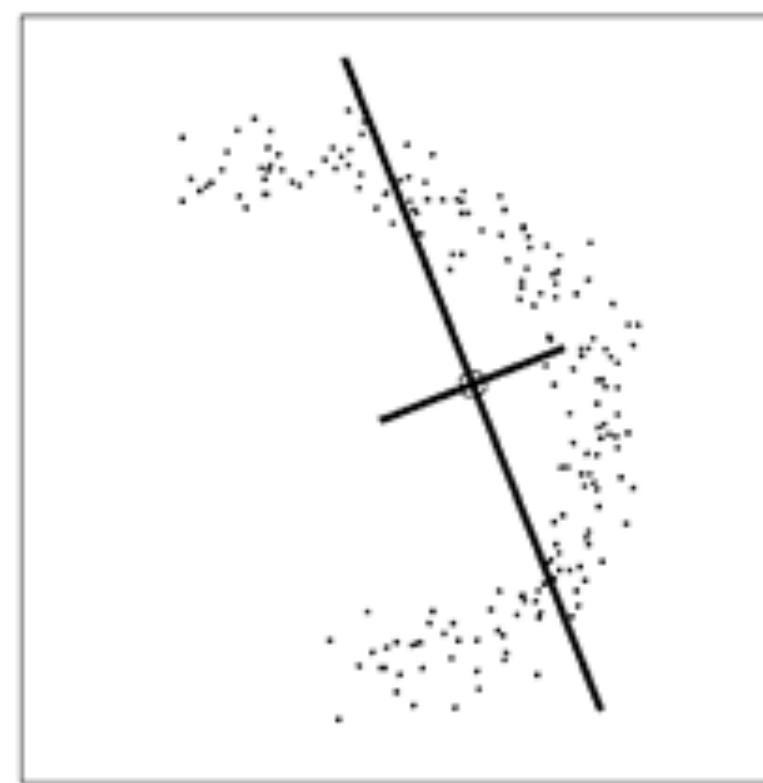


- When would PCA fail to capture the structure of data (while some other methods may work perfectly find)?

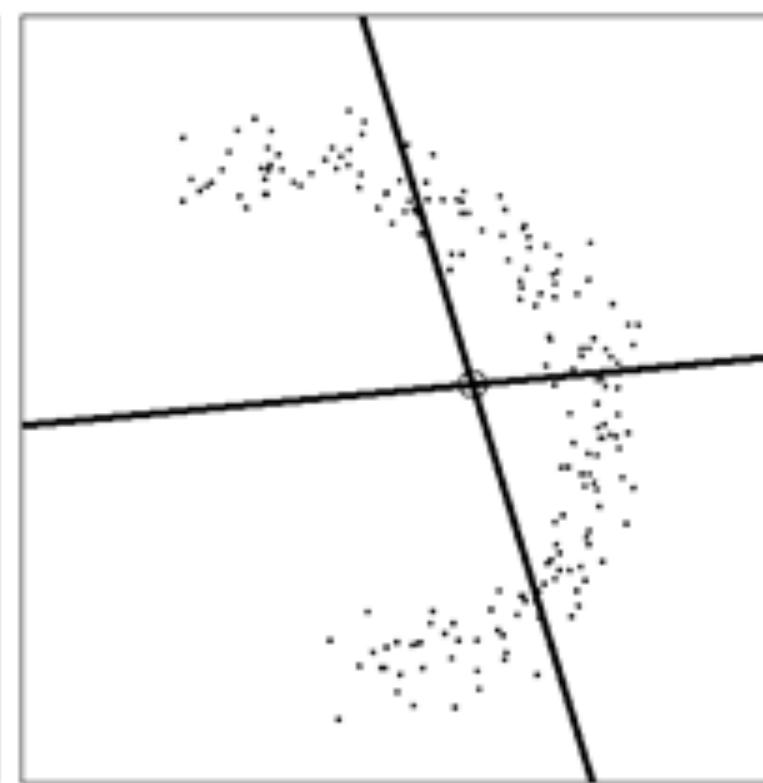
# Dimensionality Reduction



Sometimes the “axis” may not be linear. (it’s called “manifold”)



(a)

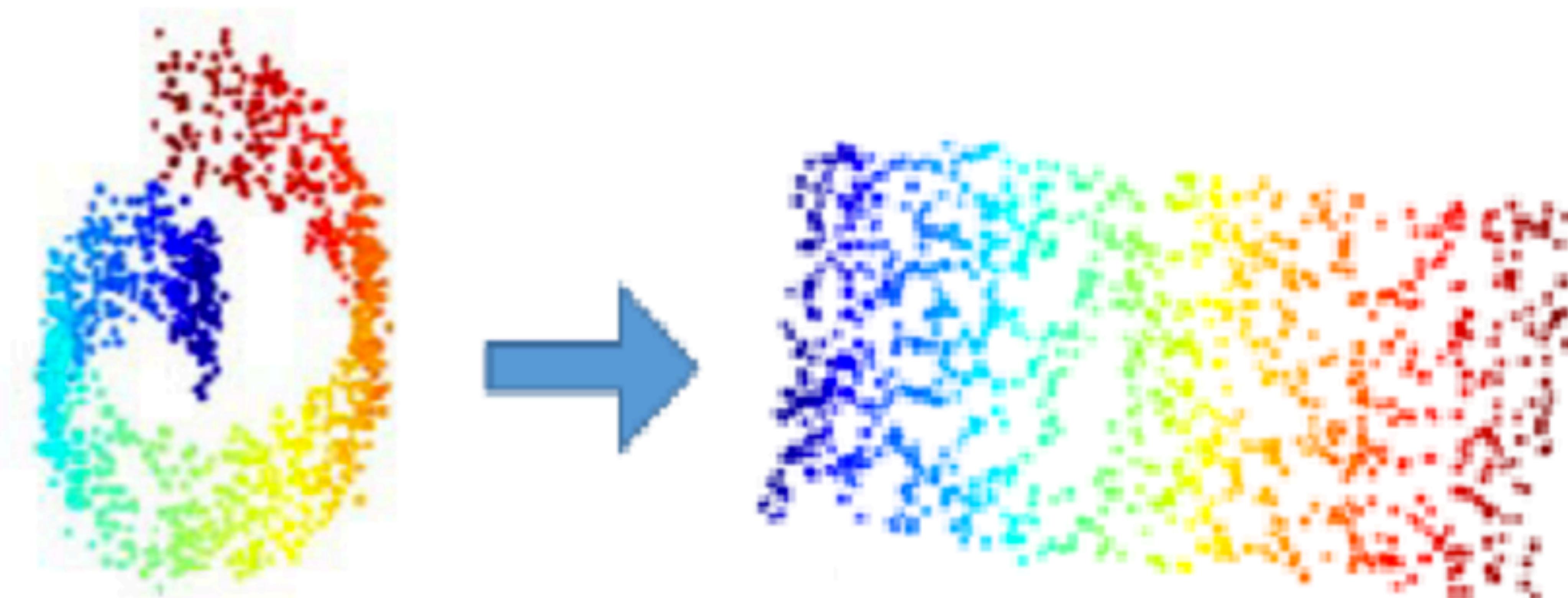


(b)



(c)

a **Manifold**: a topological space  
that locally resemble Euclidean  
space (imagine curved space)



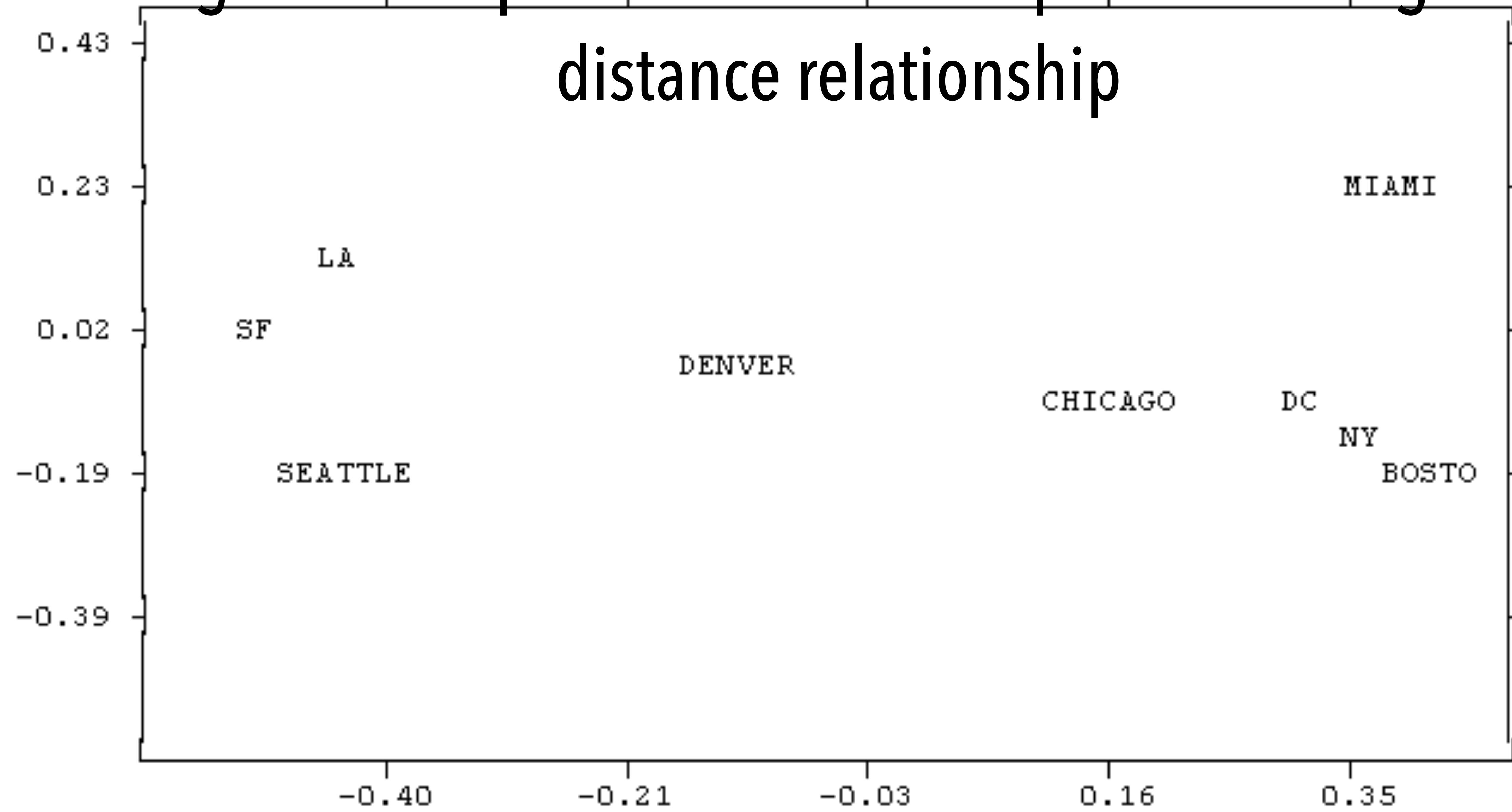
**Other methods:  
MDS, Isomap, t-SNE, and UMAP**

# Multi-Dimensional Scaling (MDS)

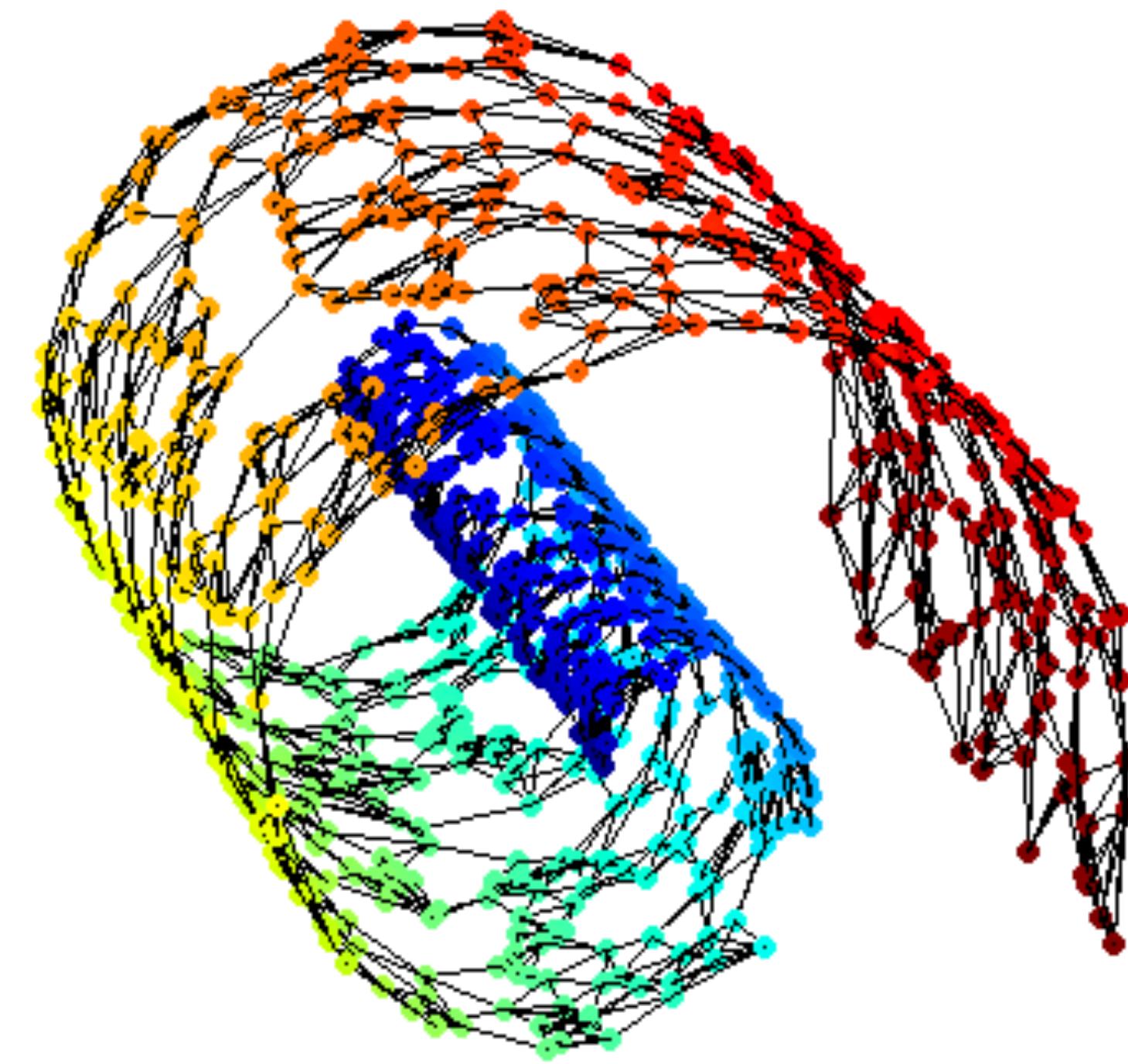
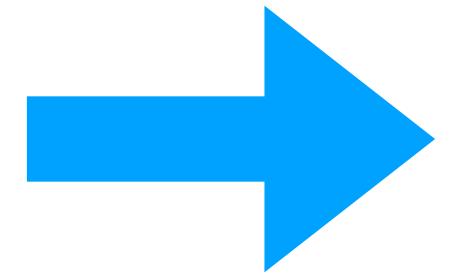
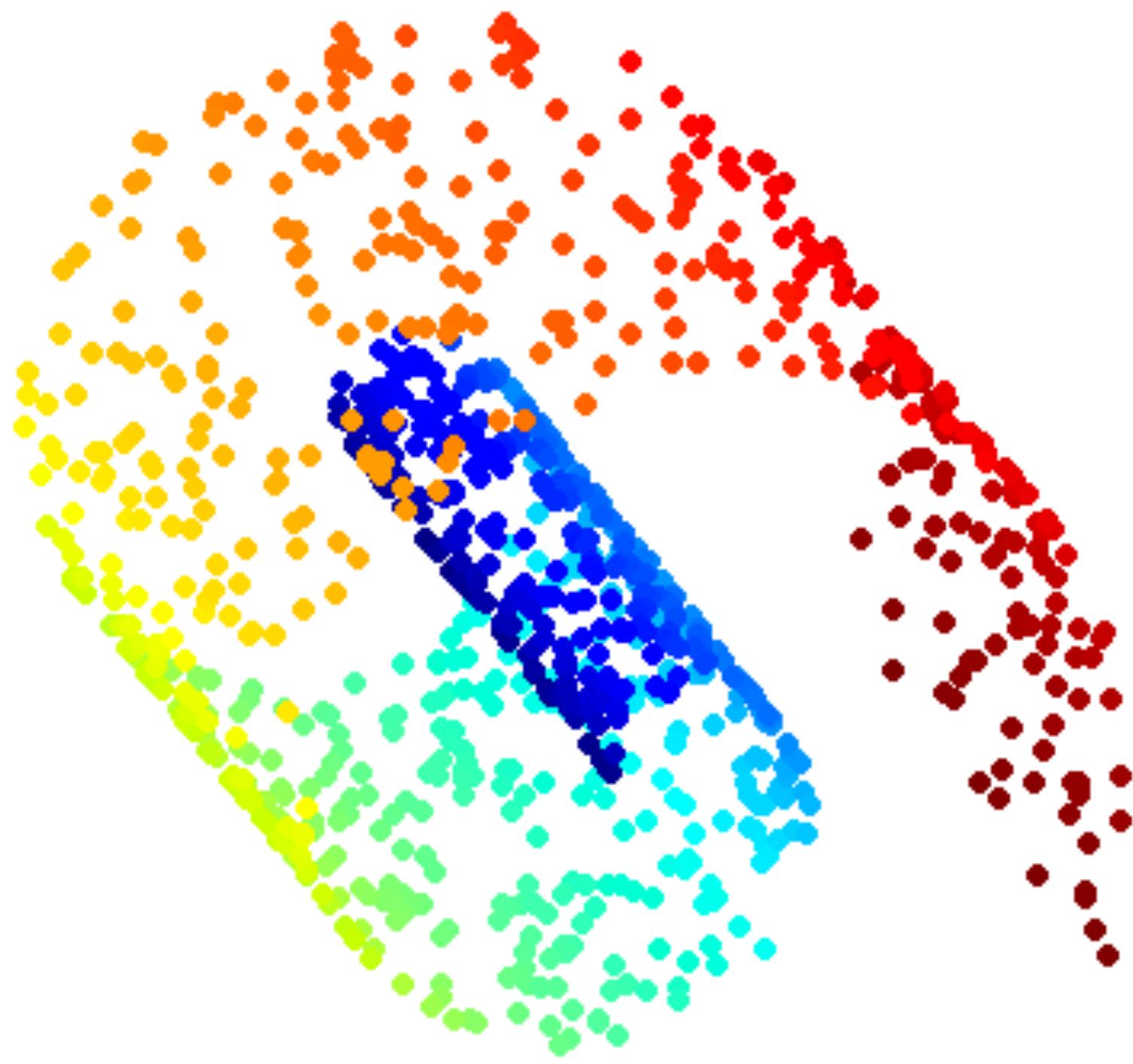


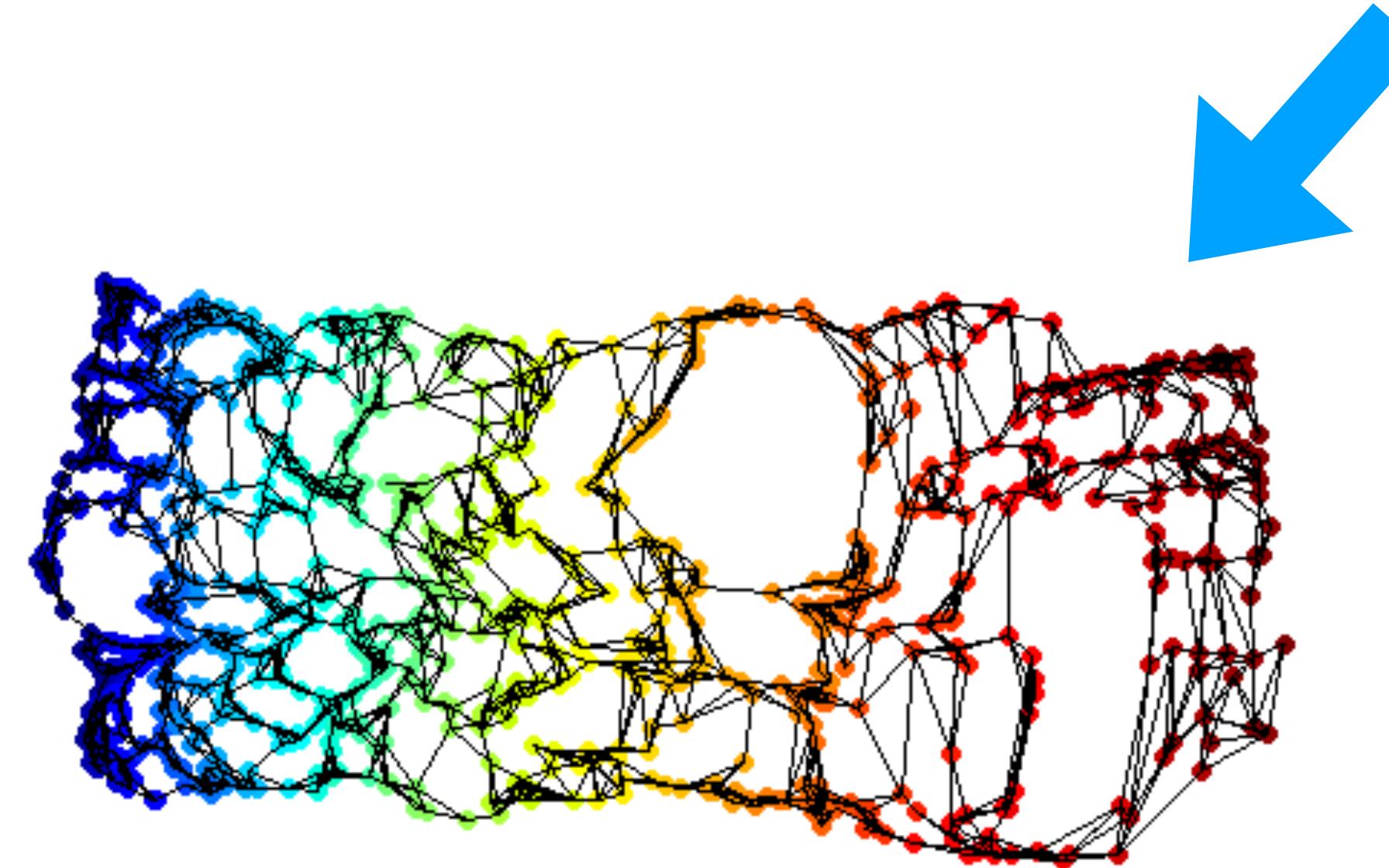
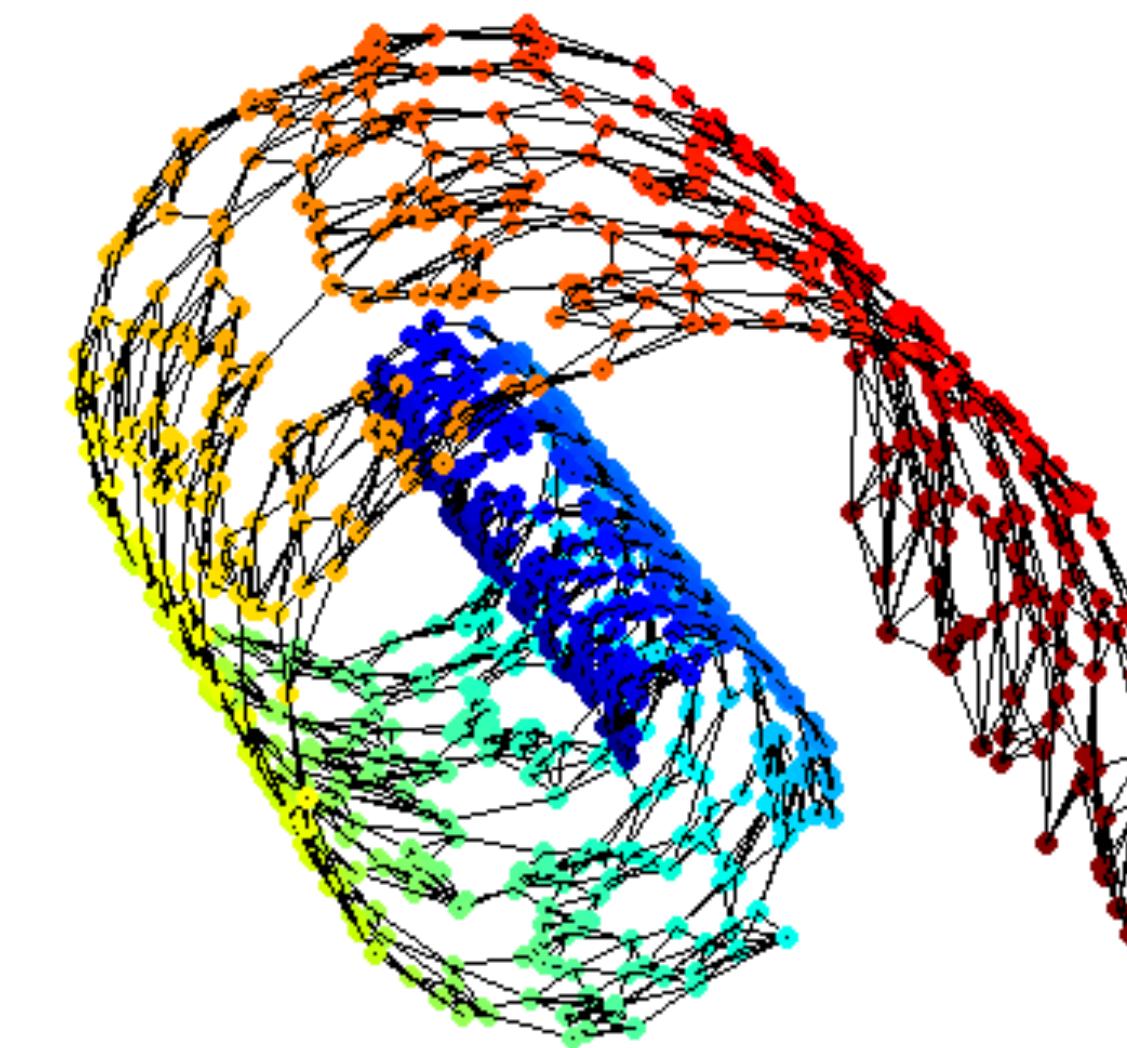
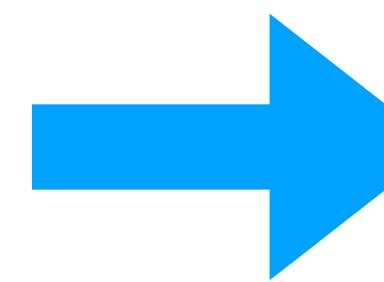
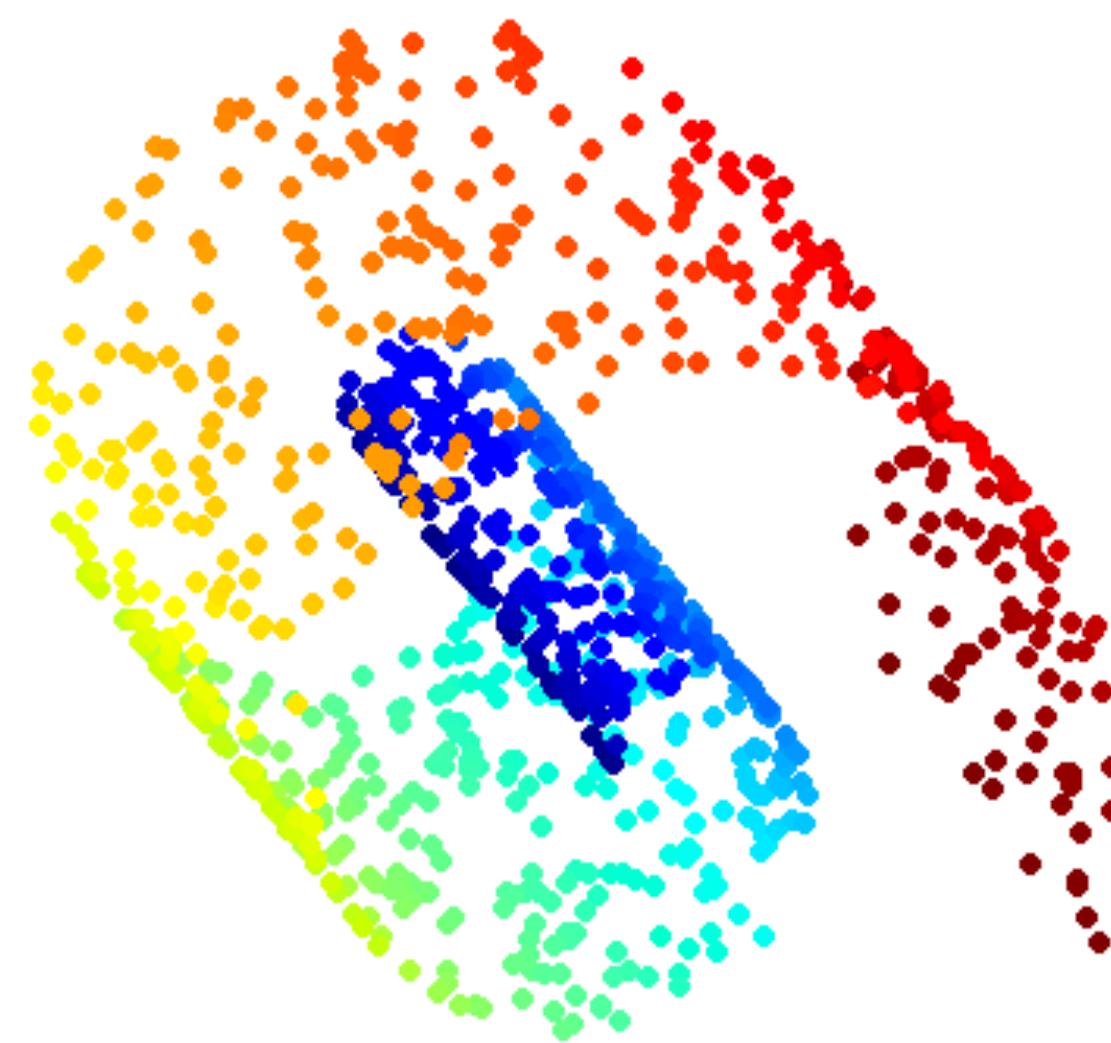
e.g. Distances between cities in 3D

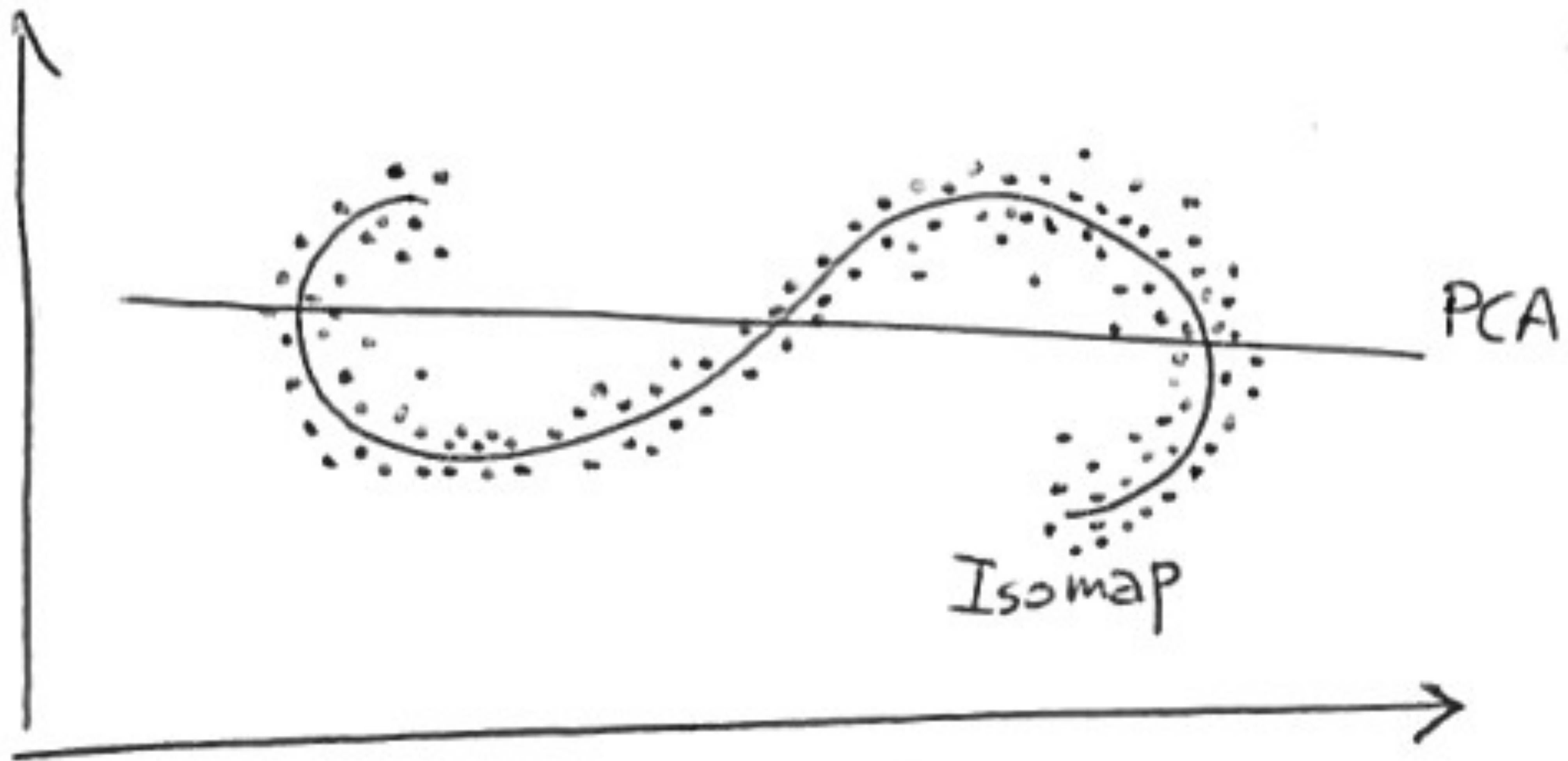
A good 2D representation that captures the original  
distance relationship



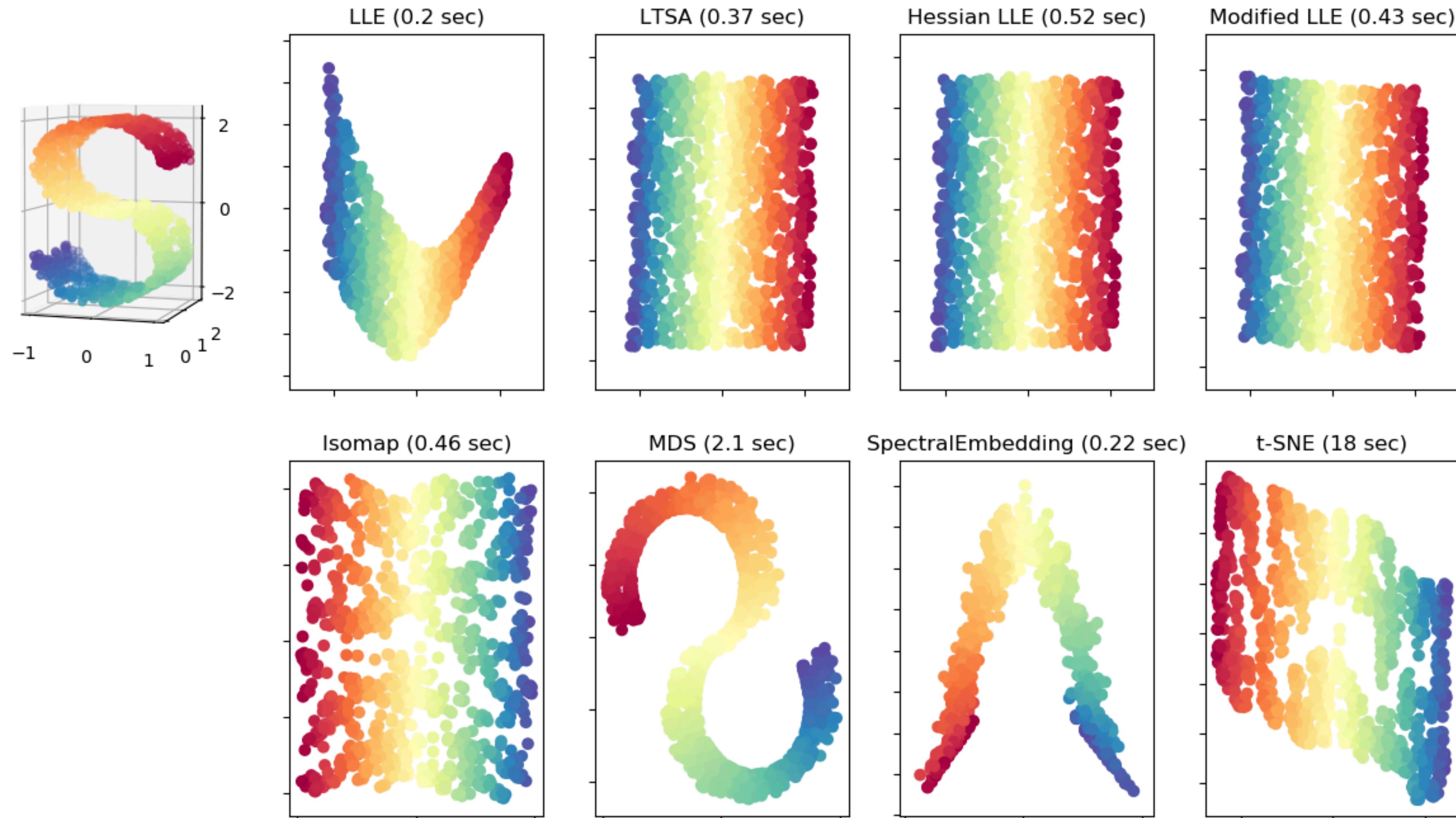
# Isomap



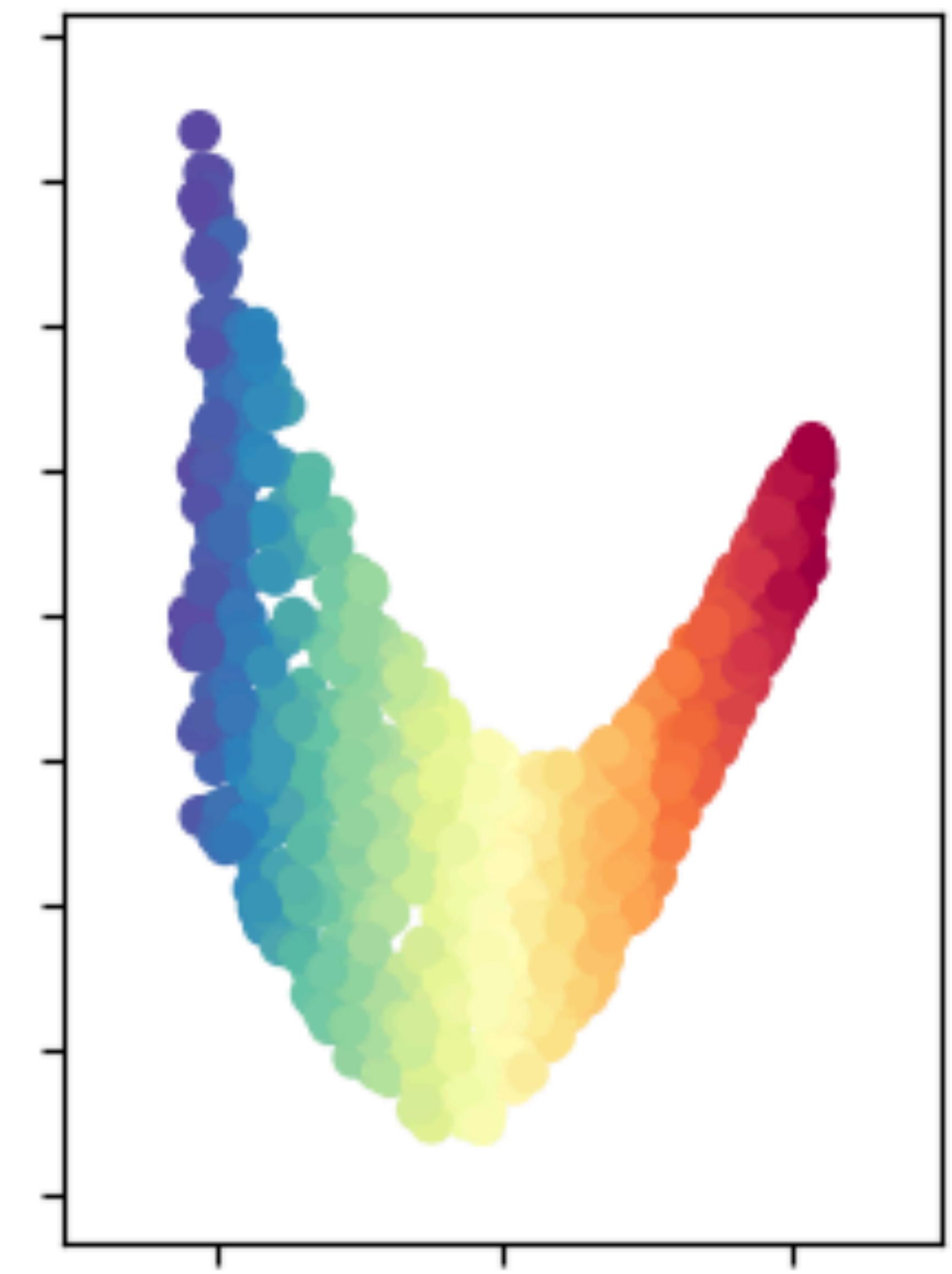
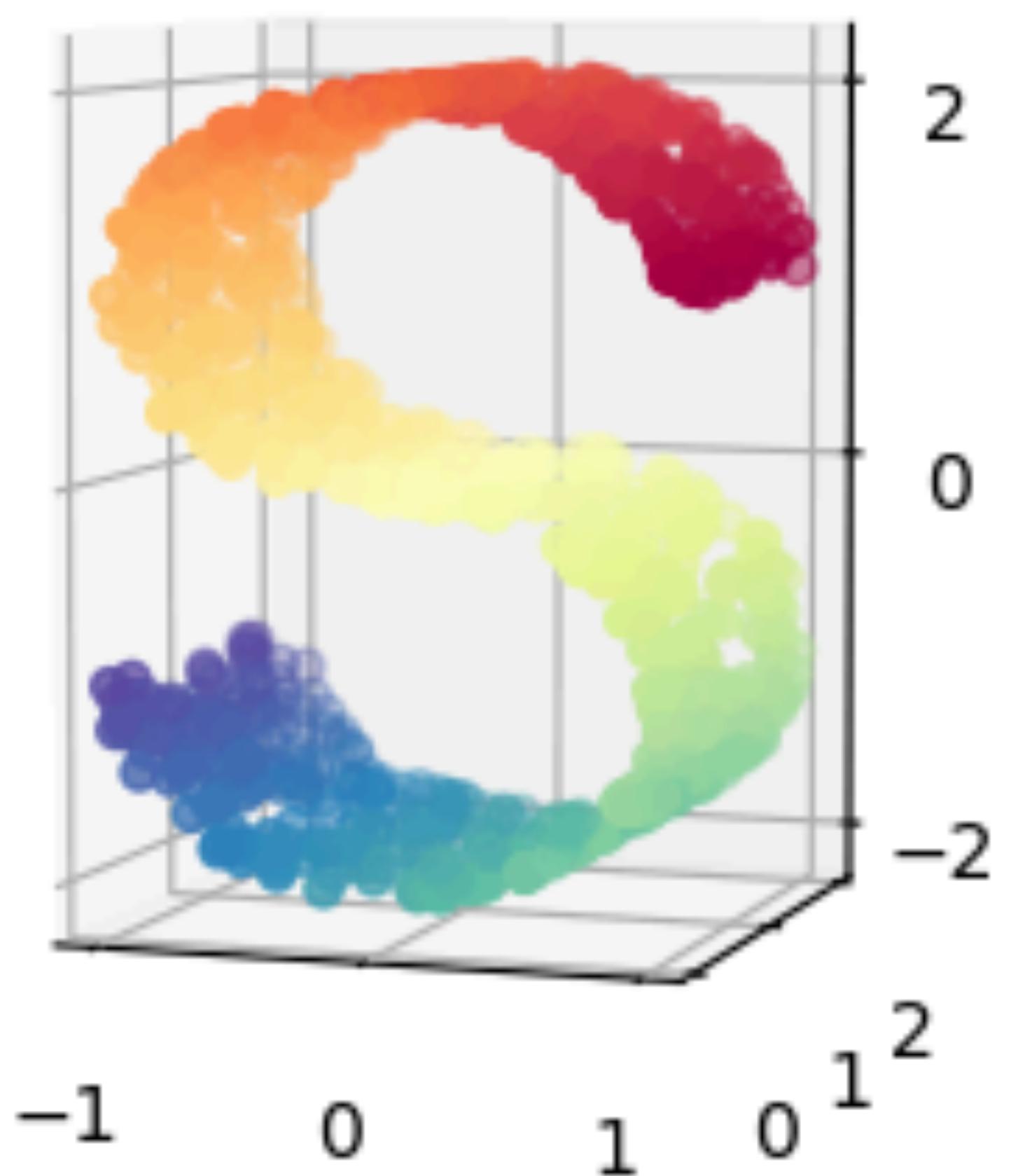




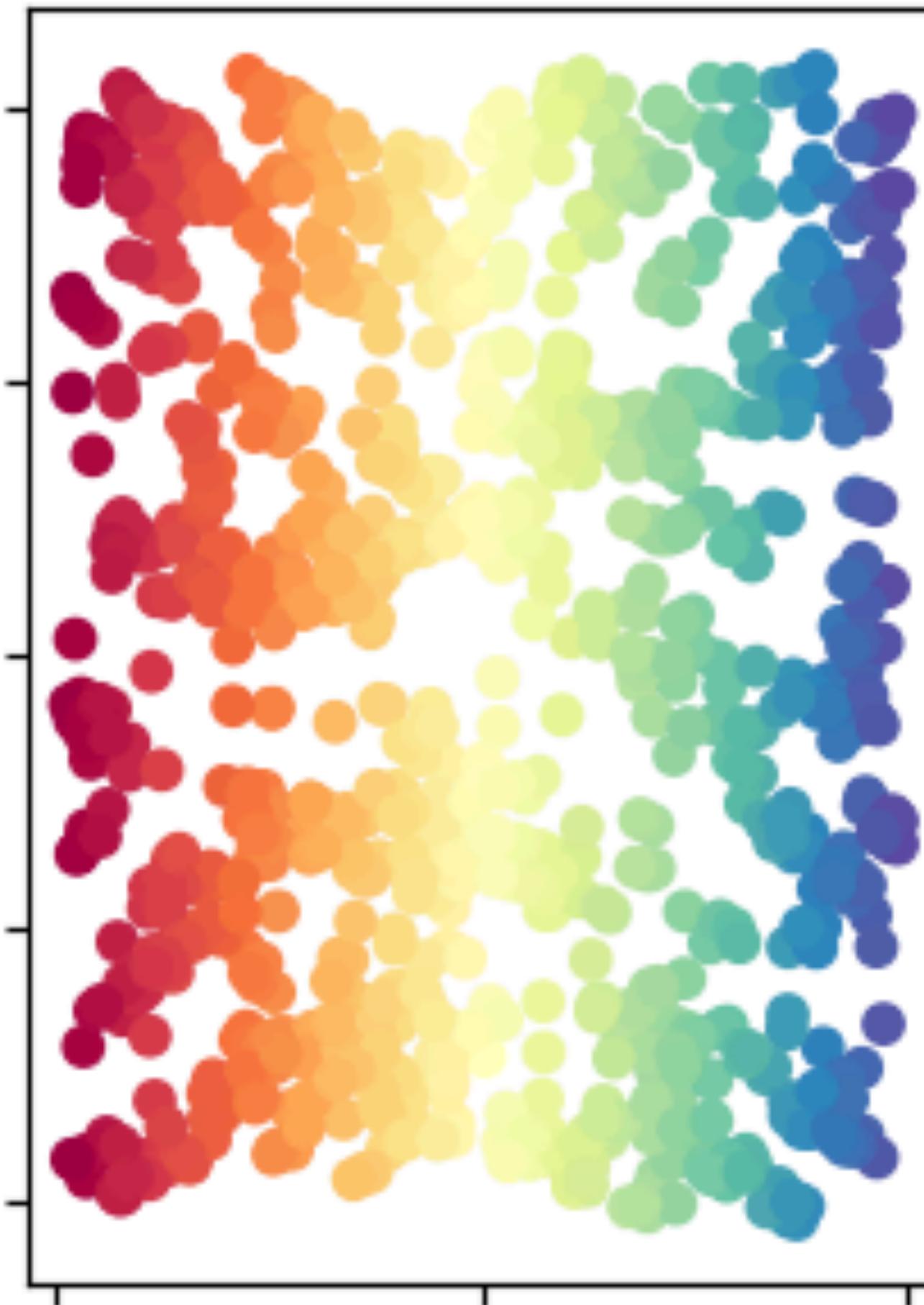
## Manifold Learning with 1000 points, 10 neighbors



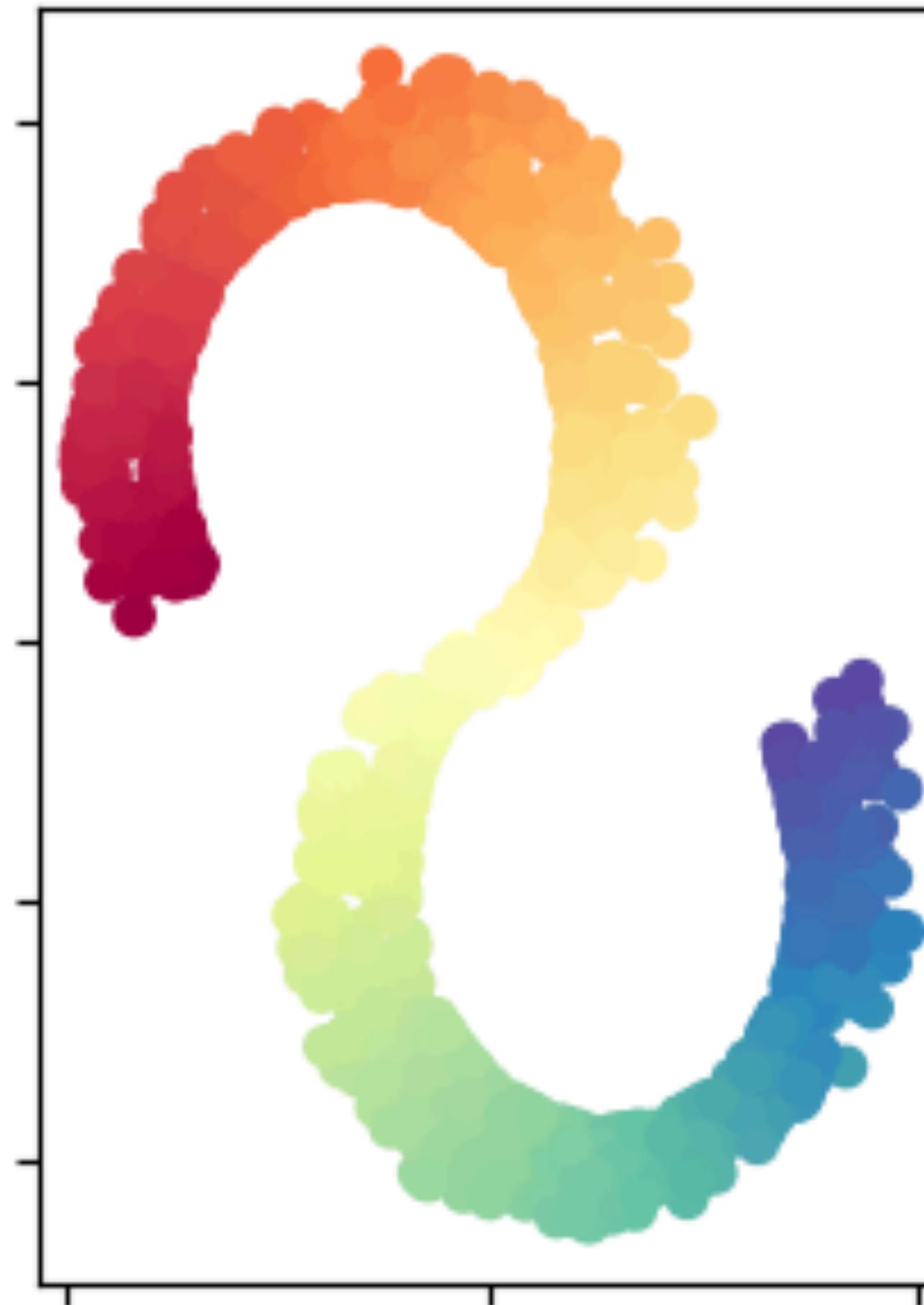
LLE (0.2 sec)



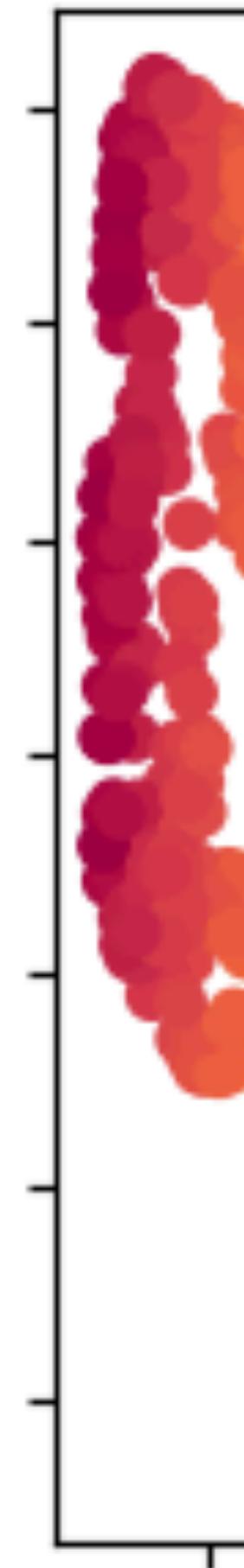
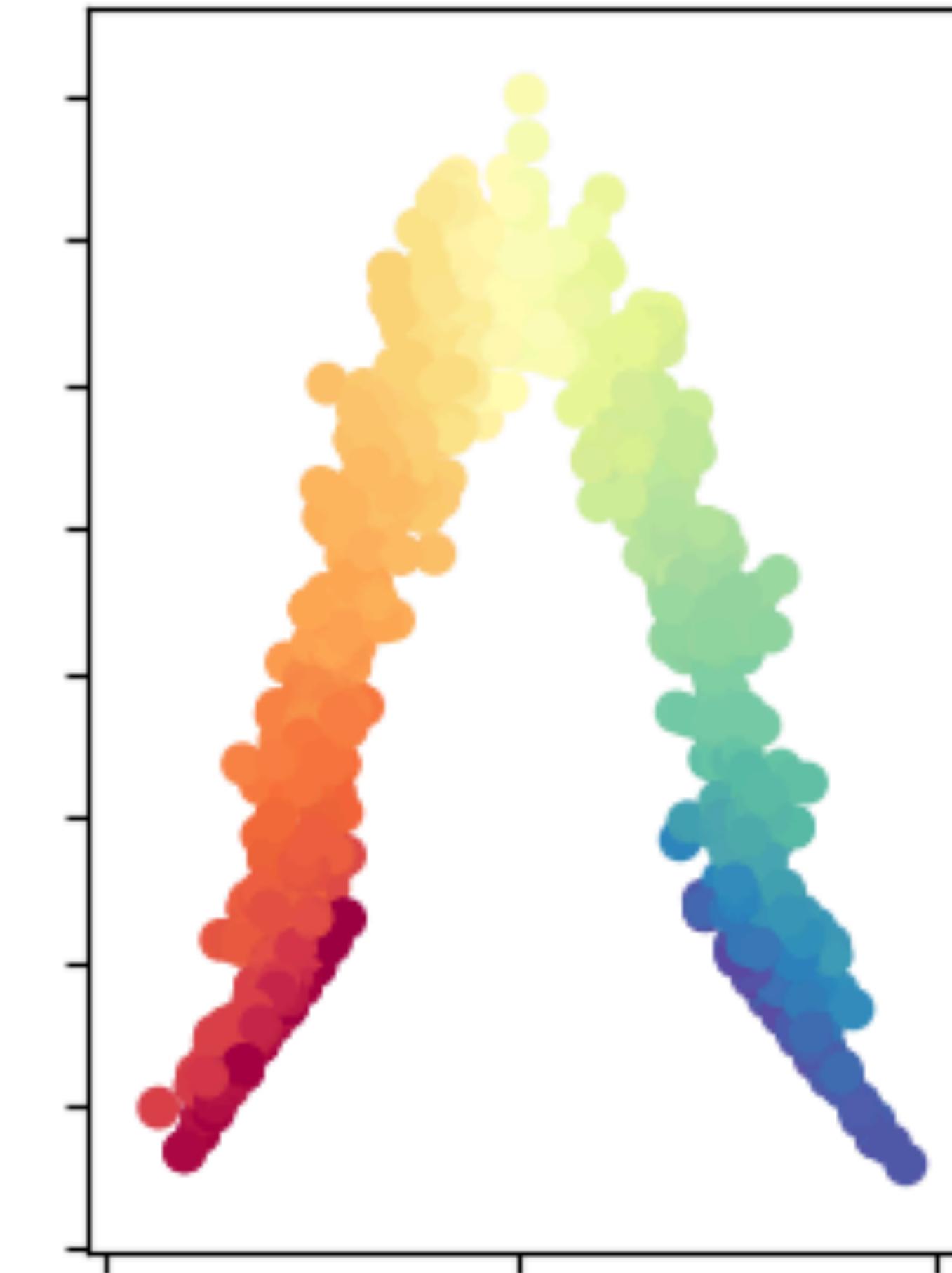
Isomap (0.46 sec)



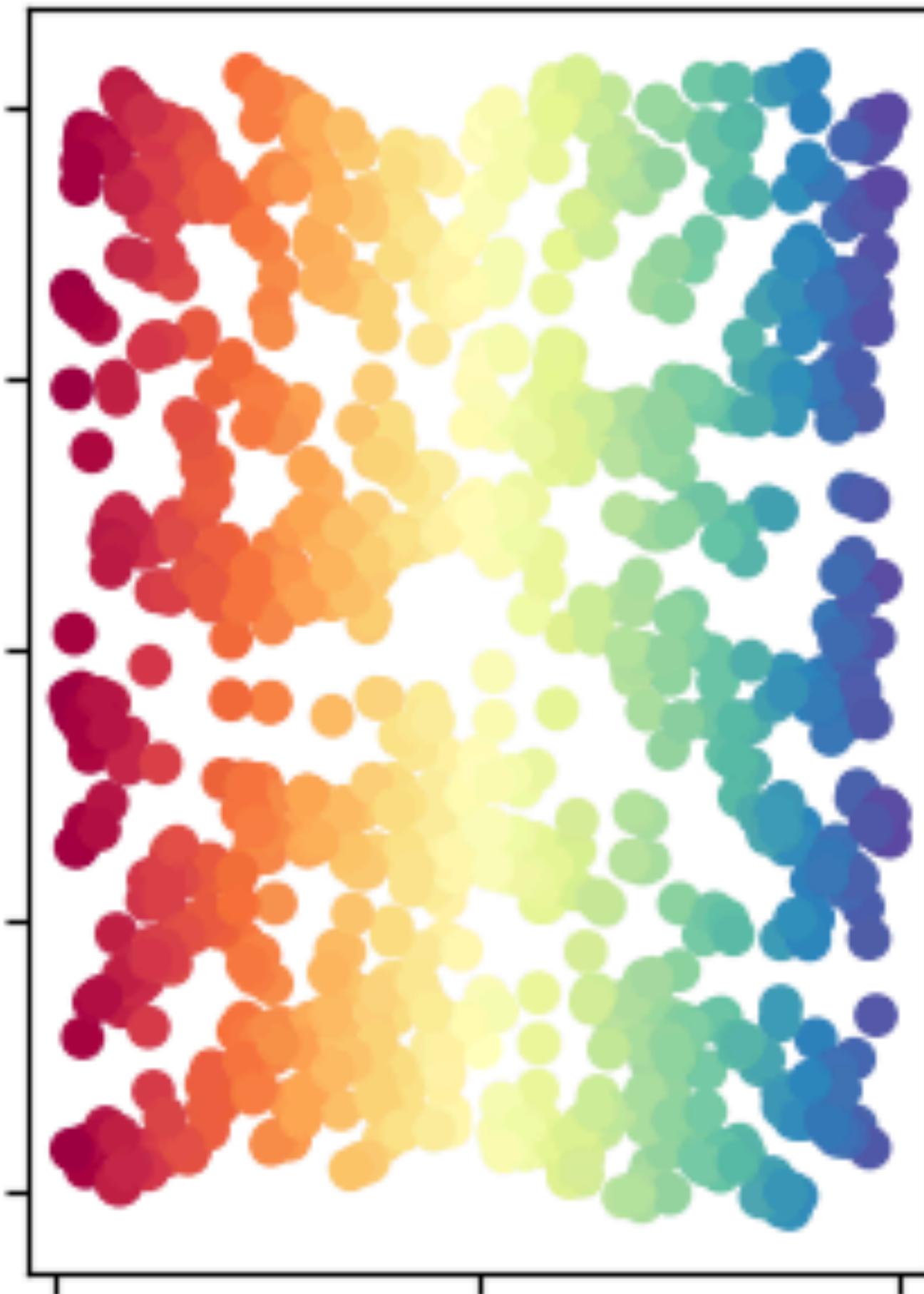
MDS (2.1 sec)



SpectralEmbedding (0.22 sec)



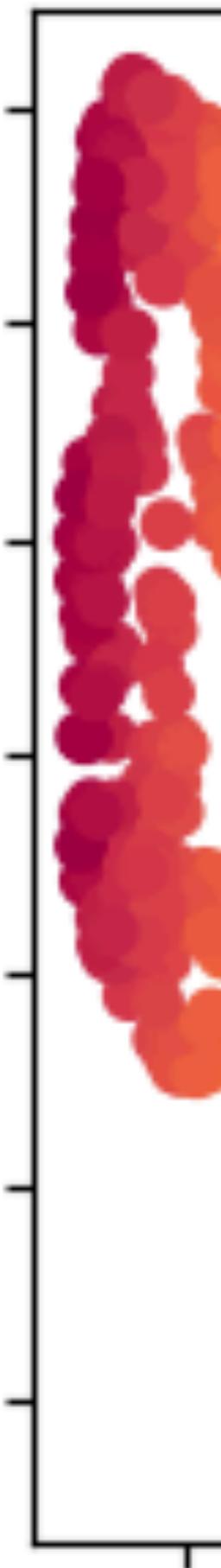
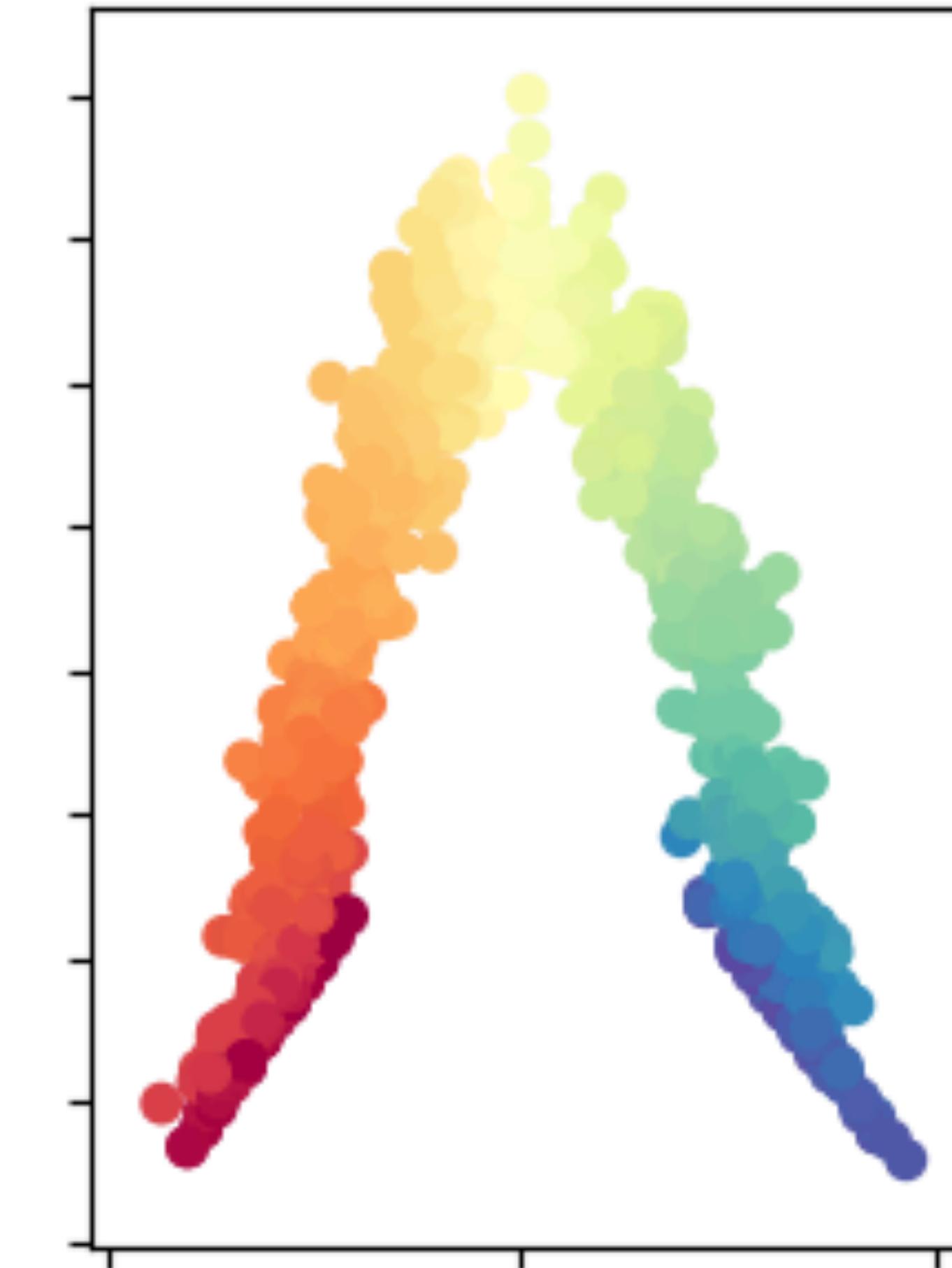
Isomap (0.46 sec)



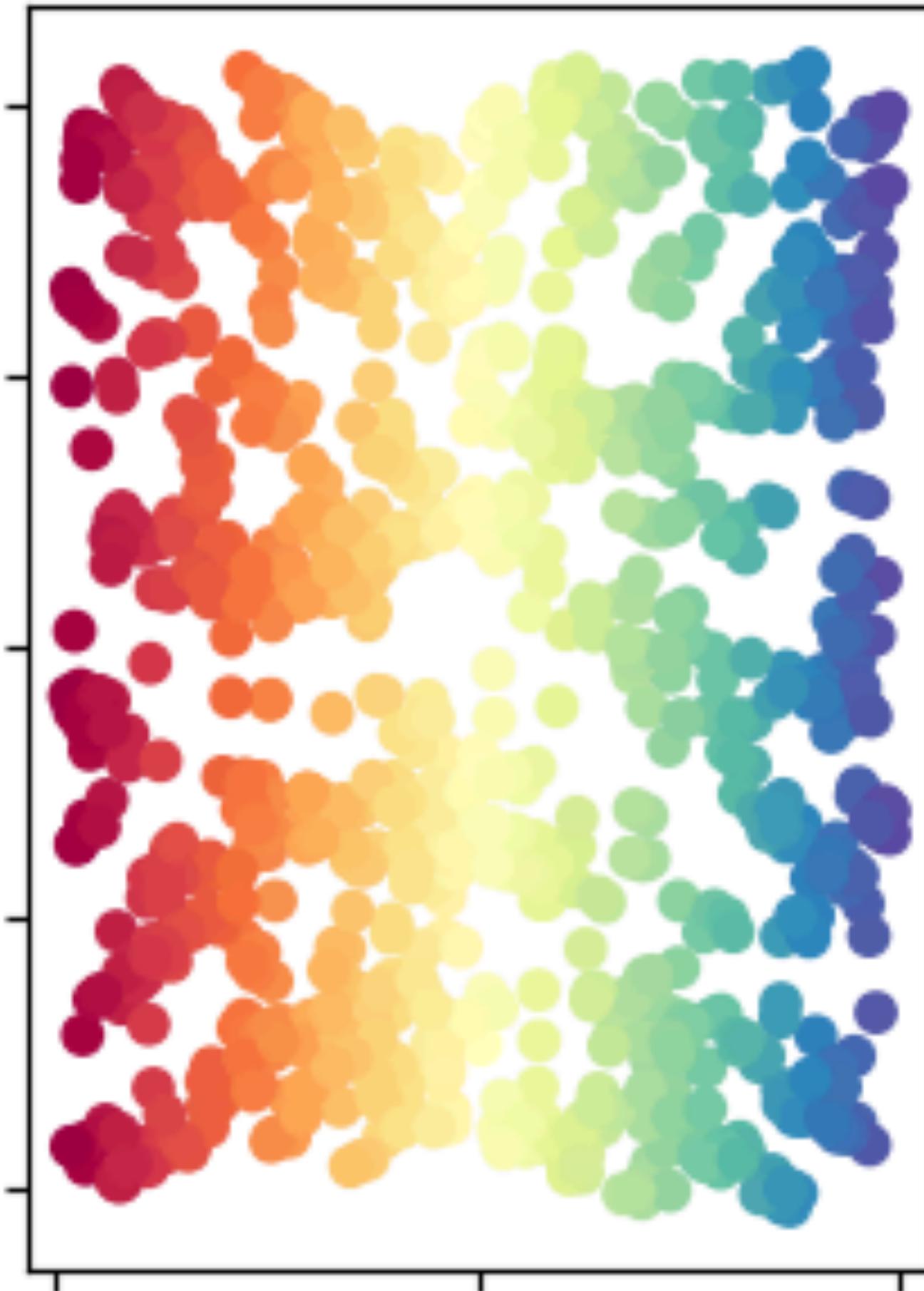
MDS (2.1 sec)



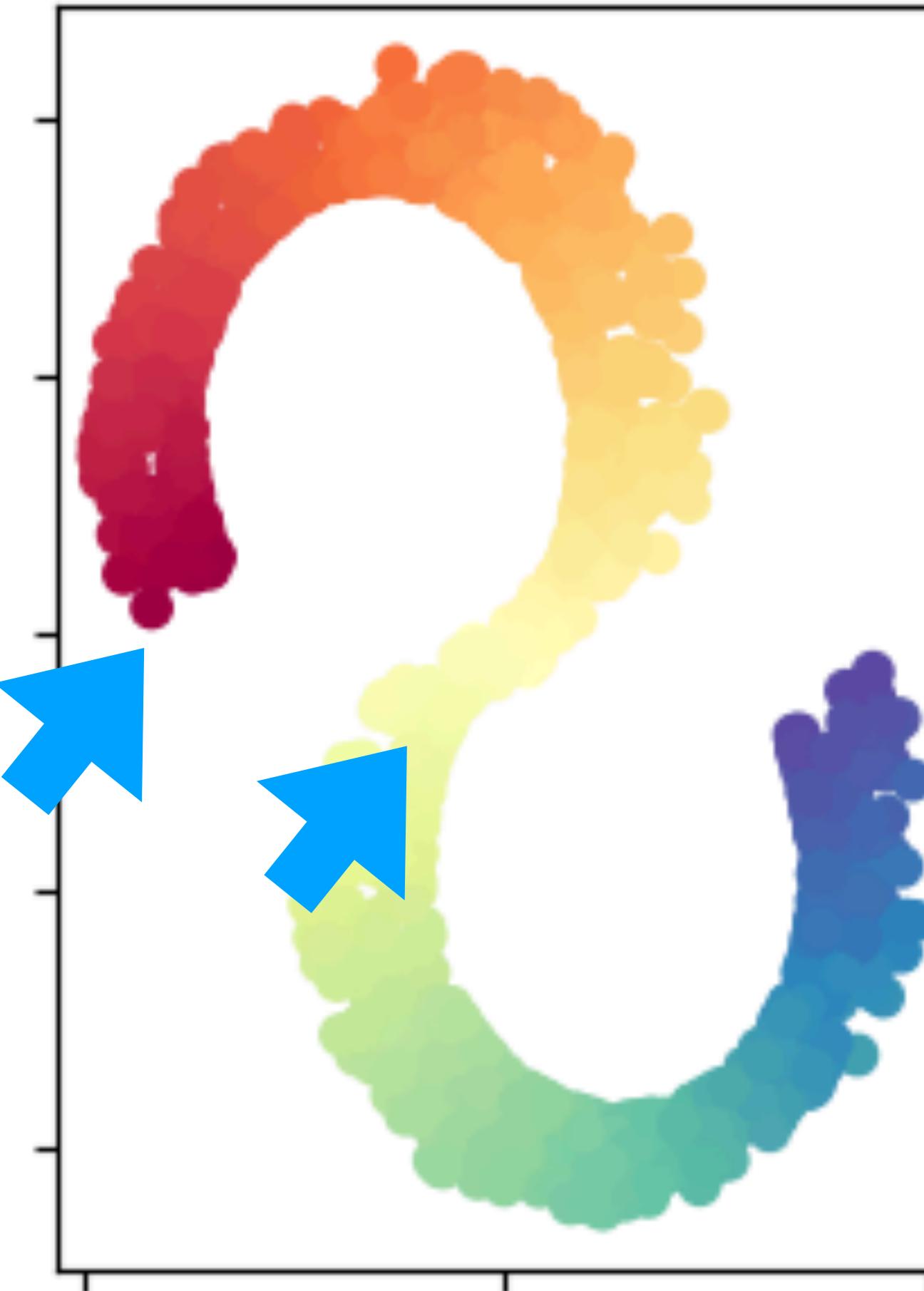
SpectralEmbedding (0.22 sec)



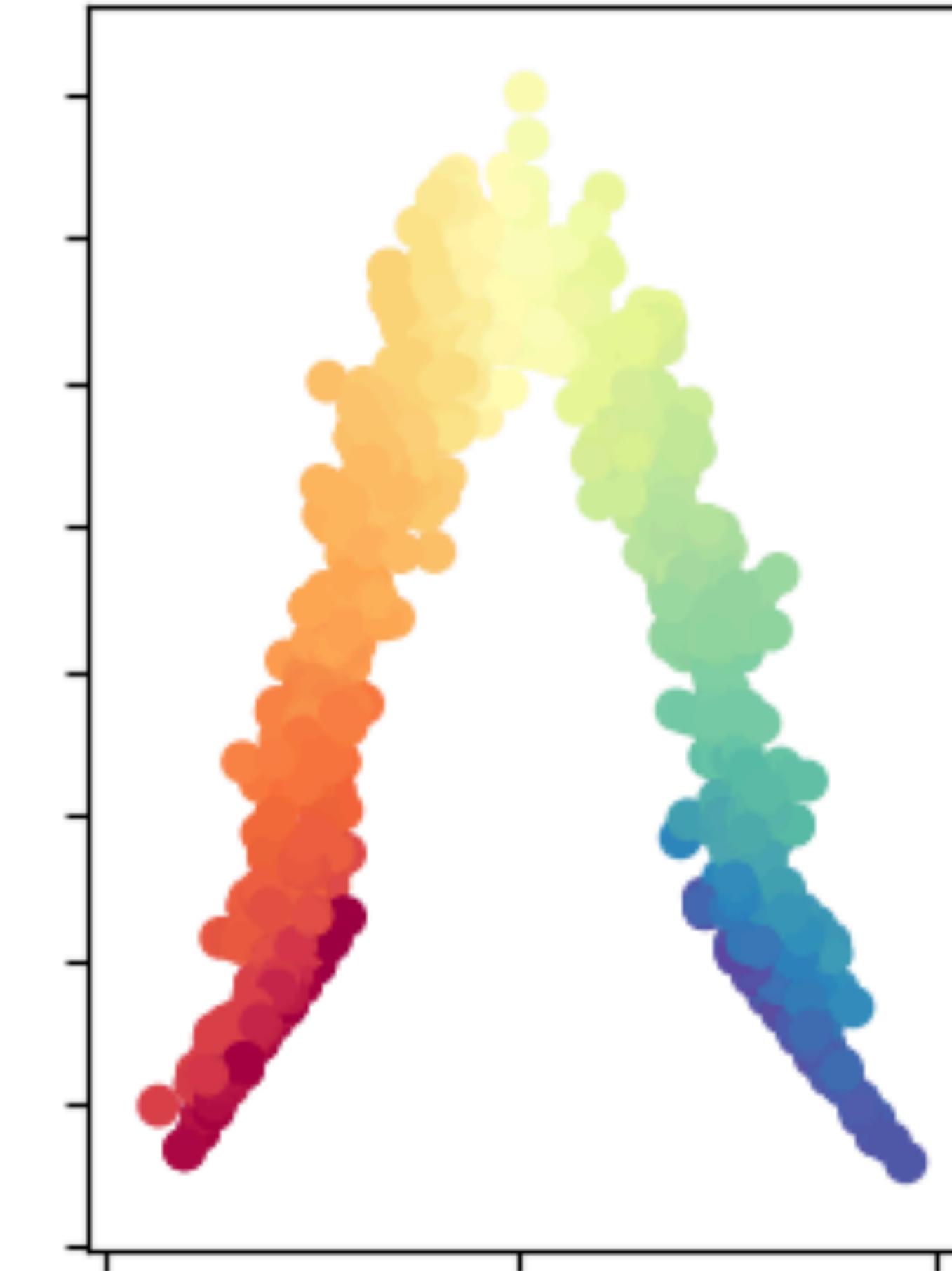
Isomap (0.46 sec)



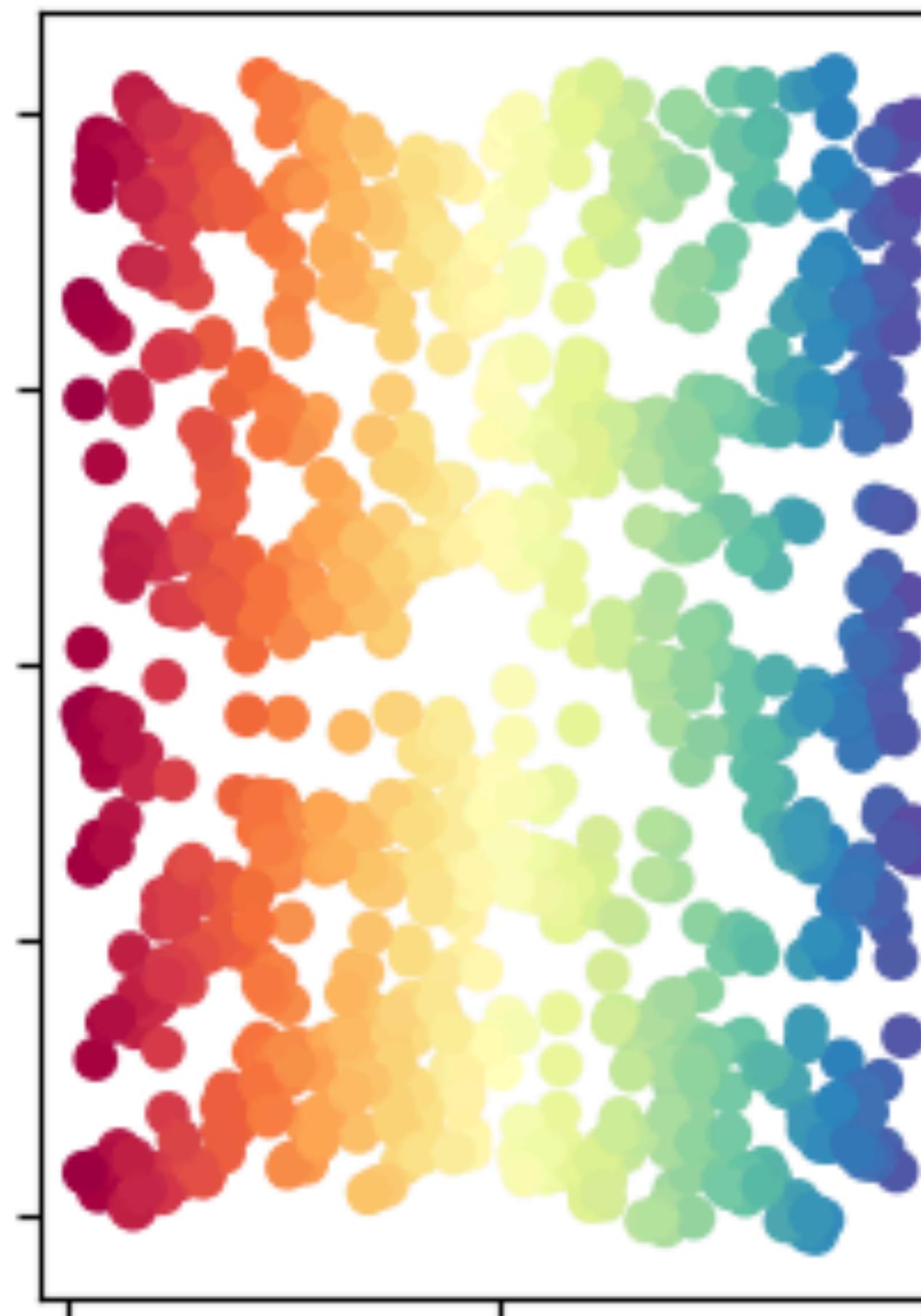
MDS (2.1 sec)



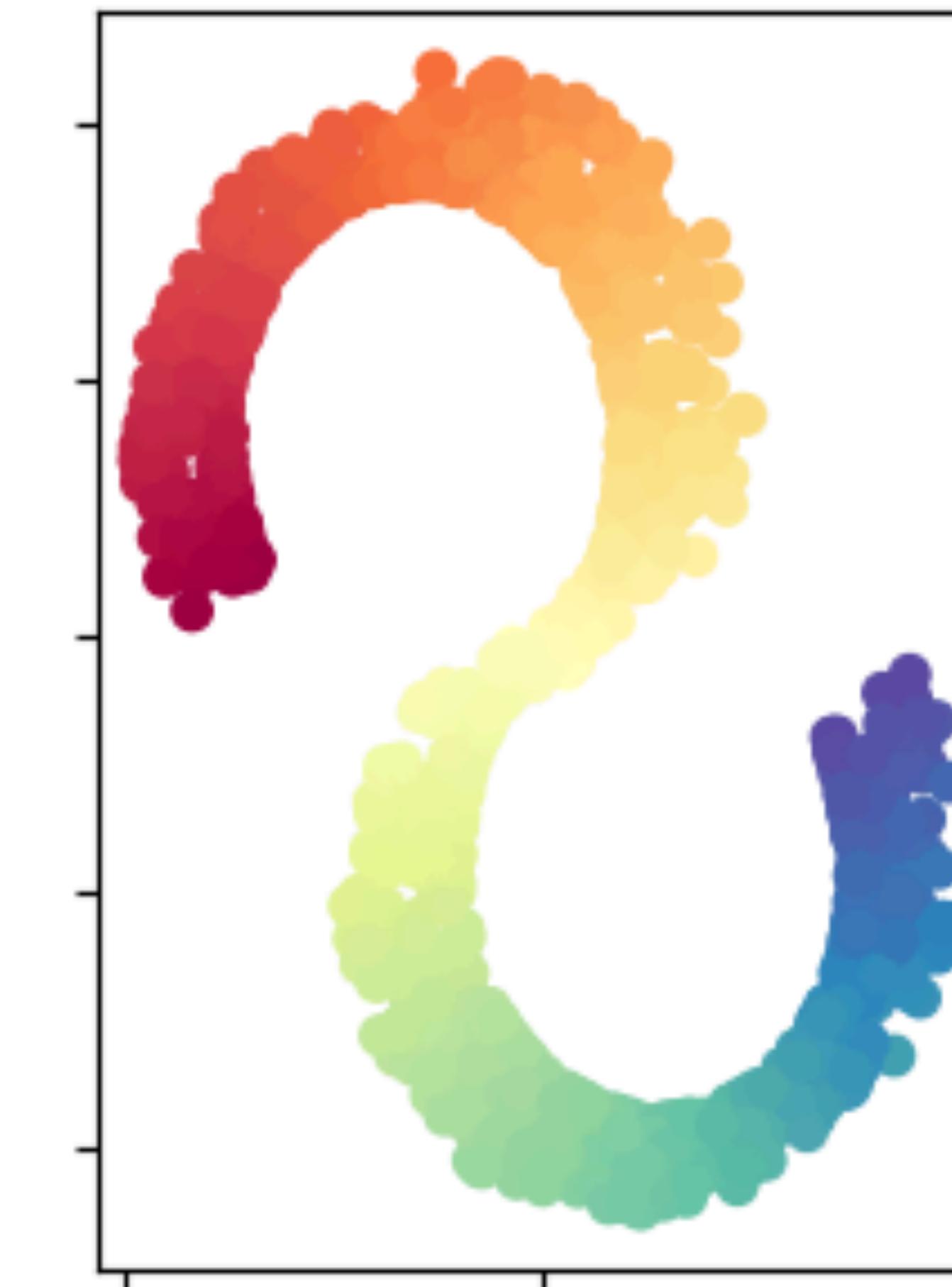
SpectralEmbedding (0.22 sec)



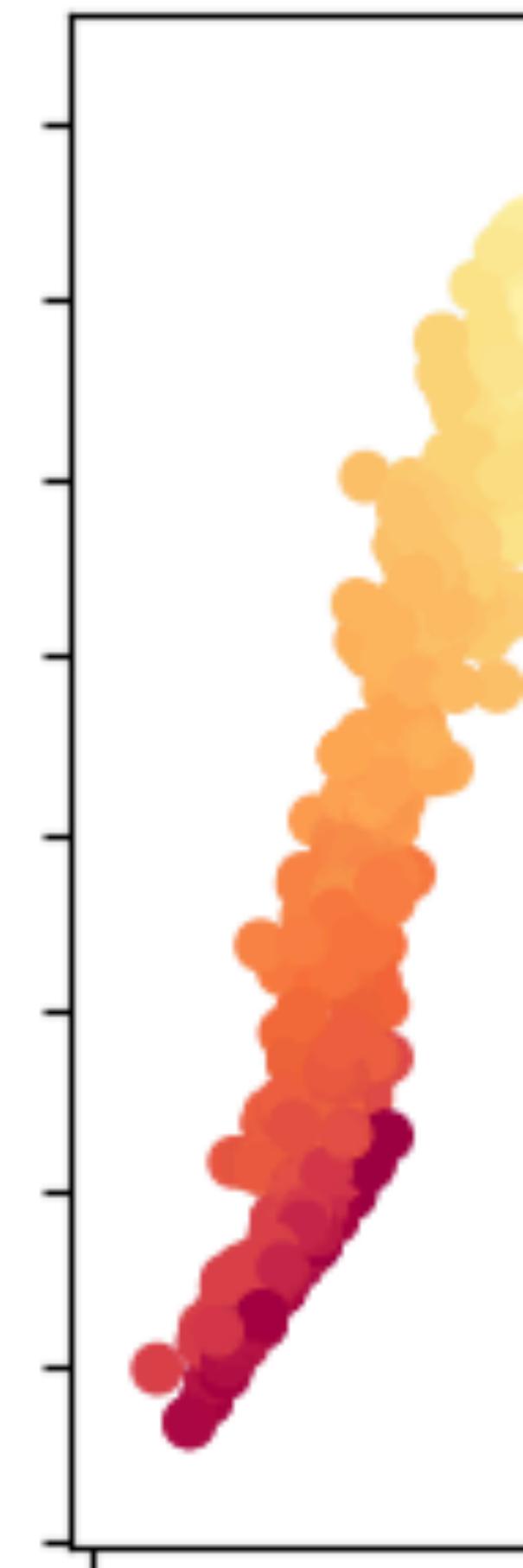
Isomap (0.46 sec)

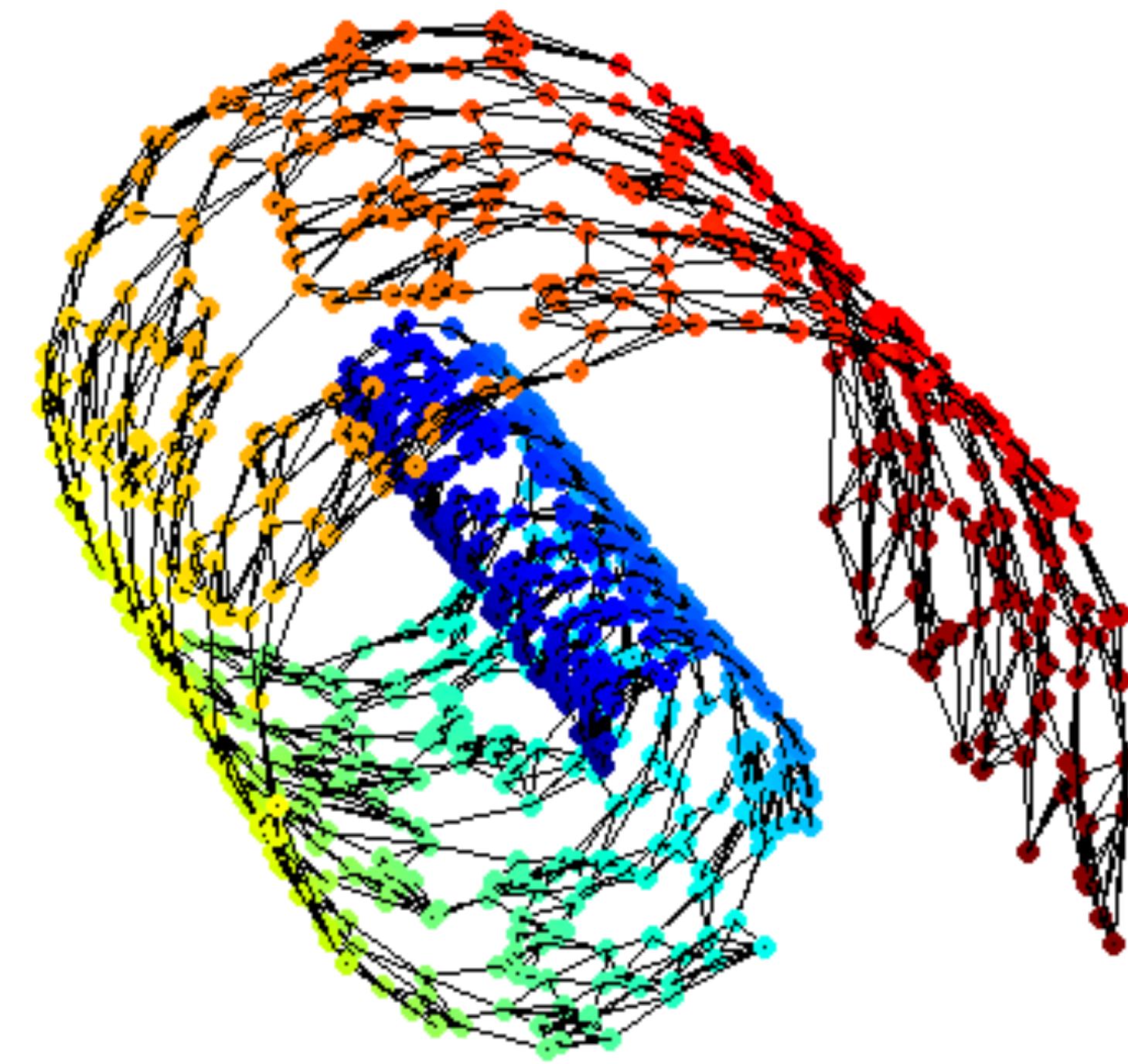
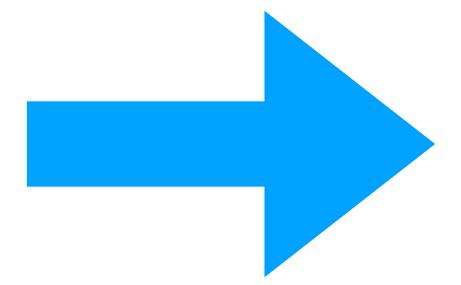
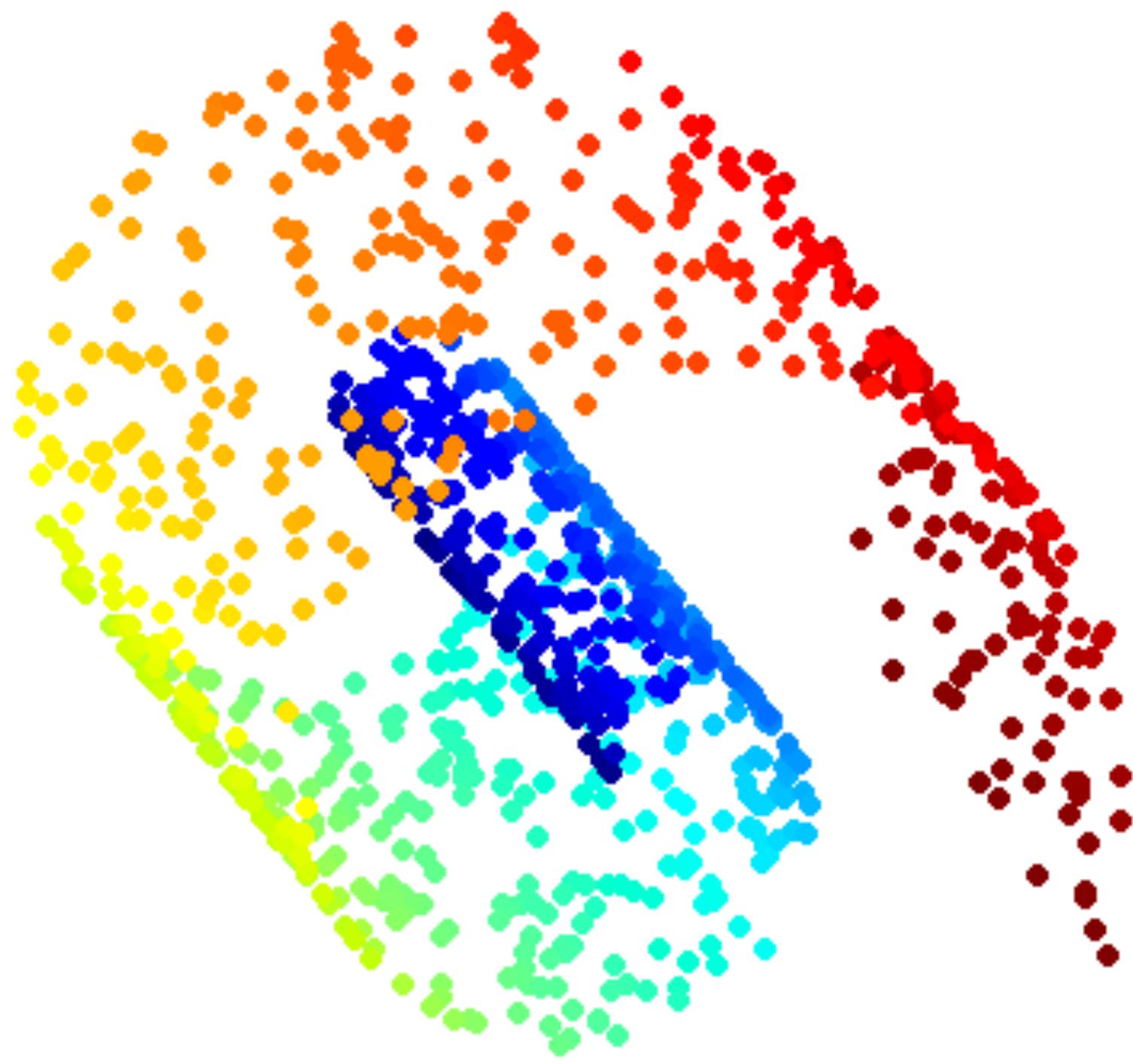


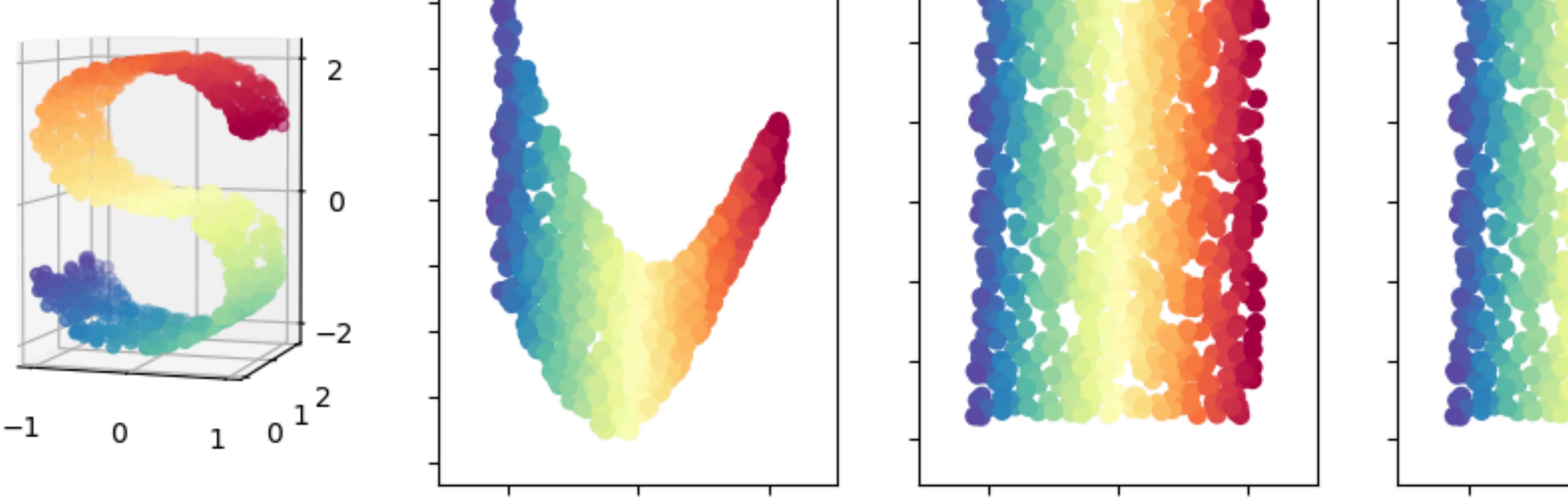
MDS (2.1 sec)



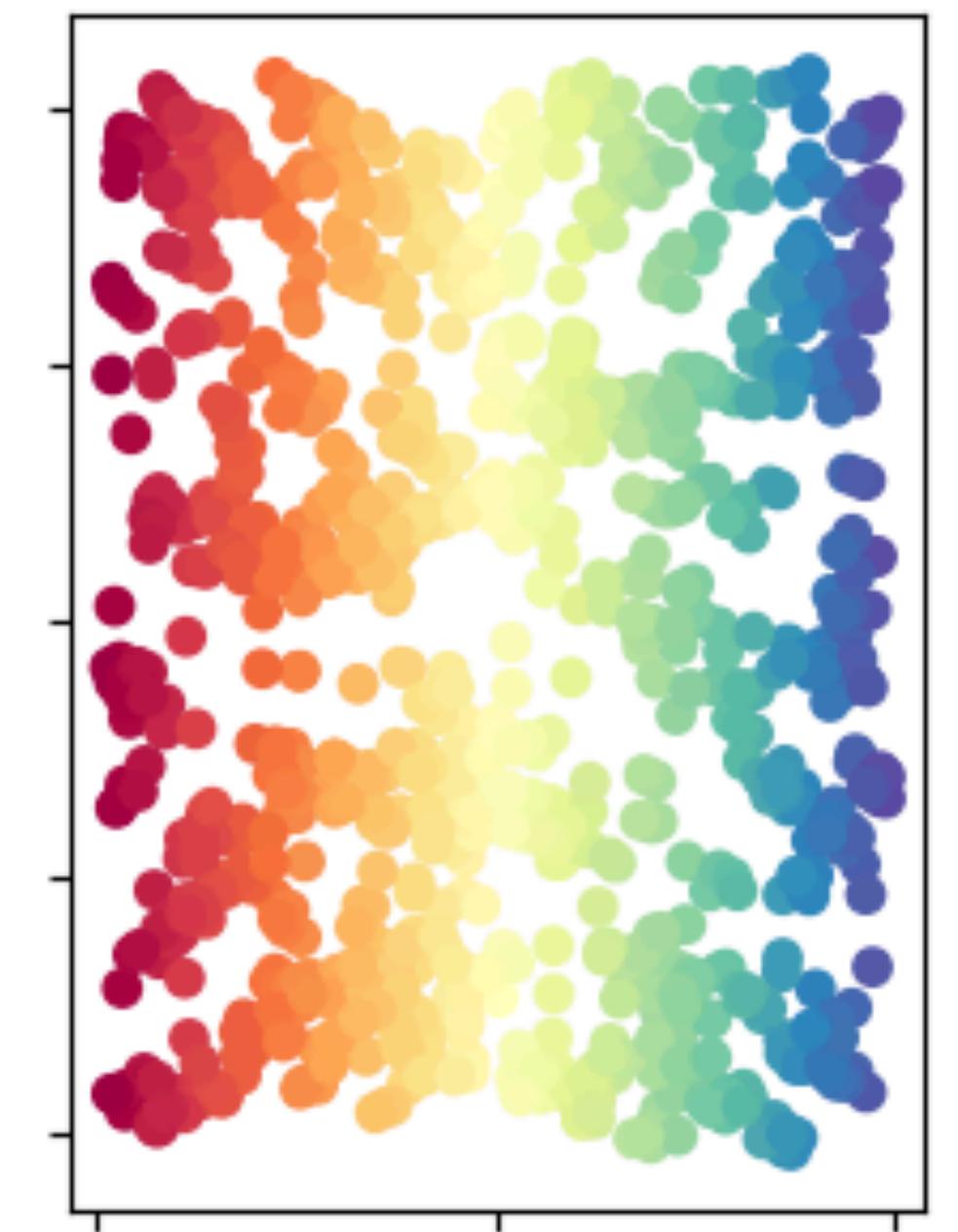
SpectralEmbed



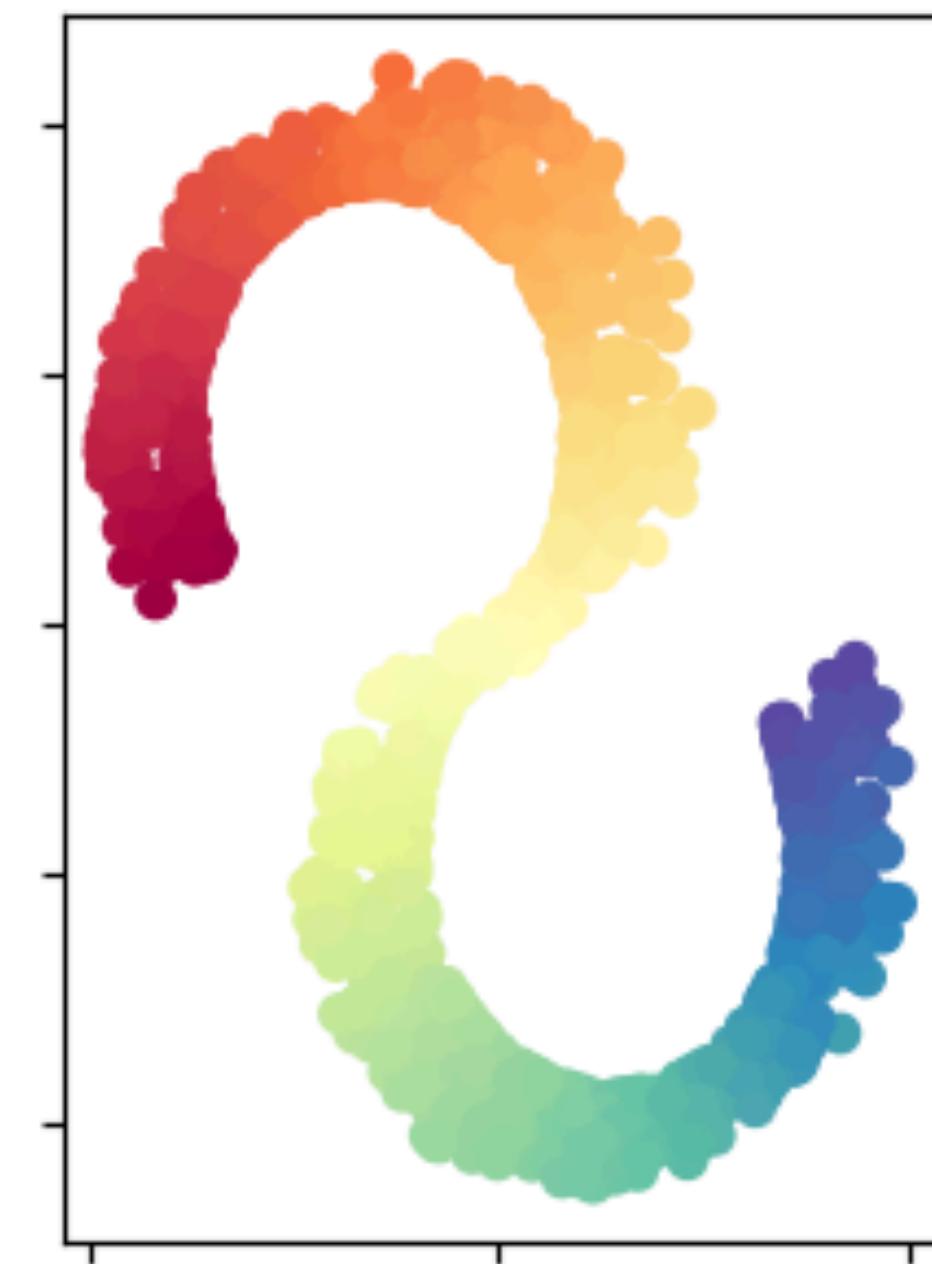




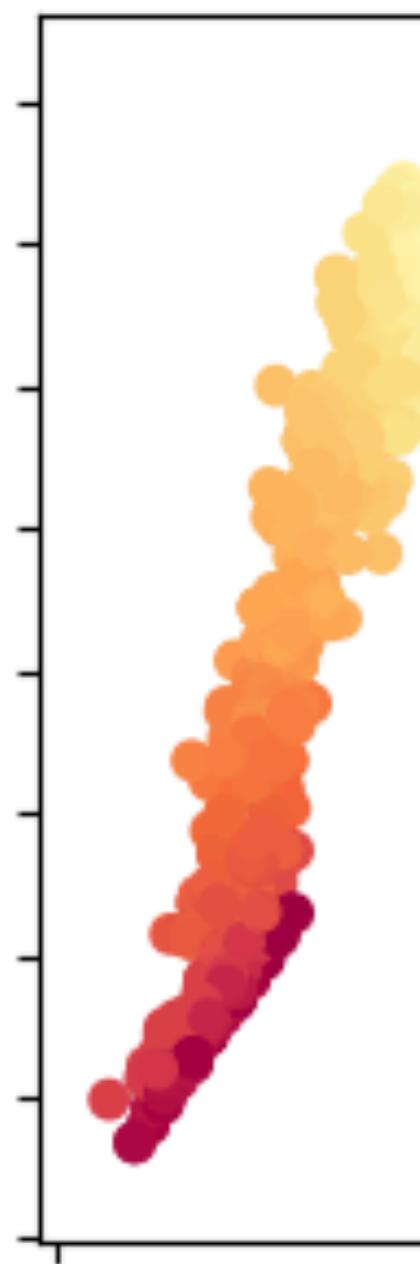
Isomap (0.46 sec)



MDS (2.1 sec)



SpectralEmbed



MDS: global

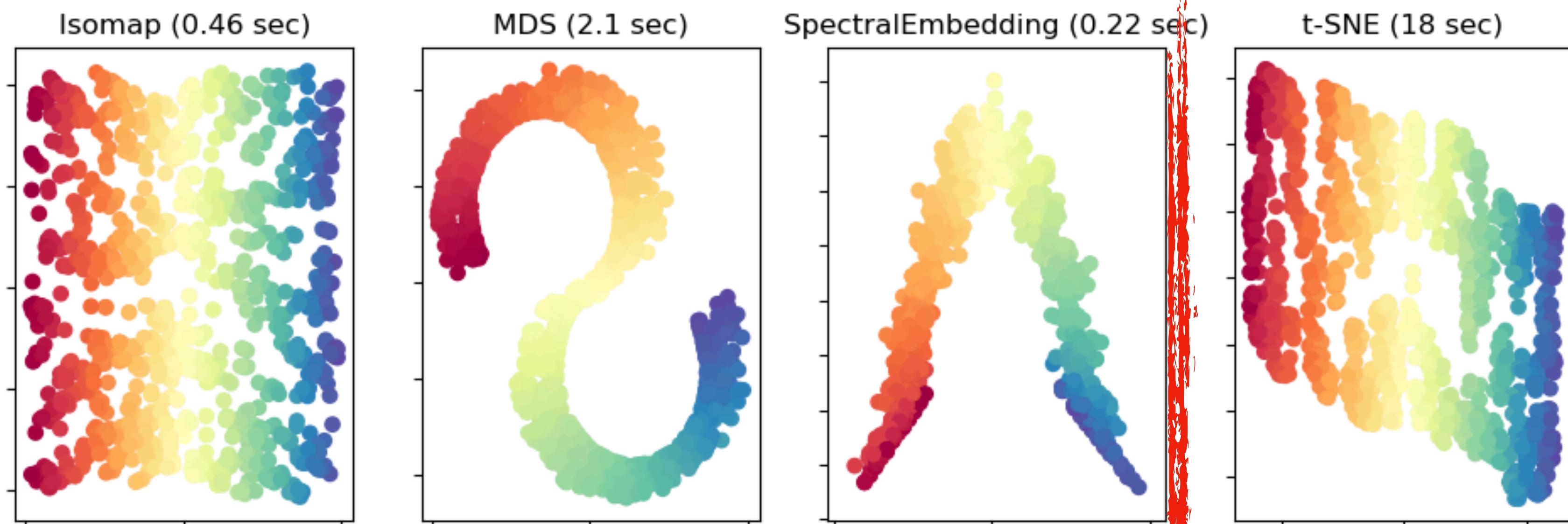
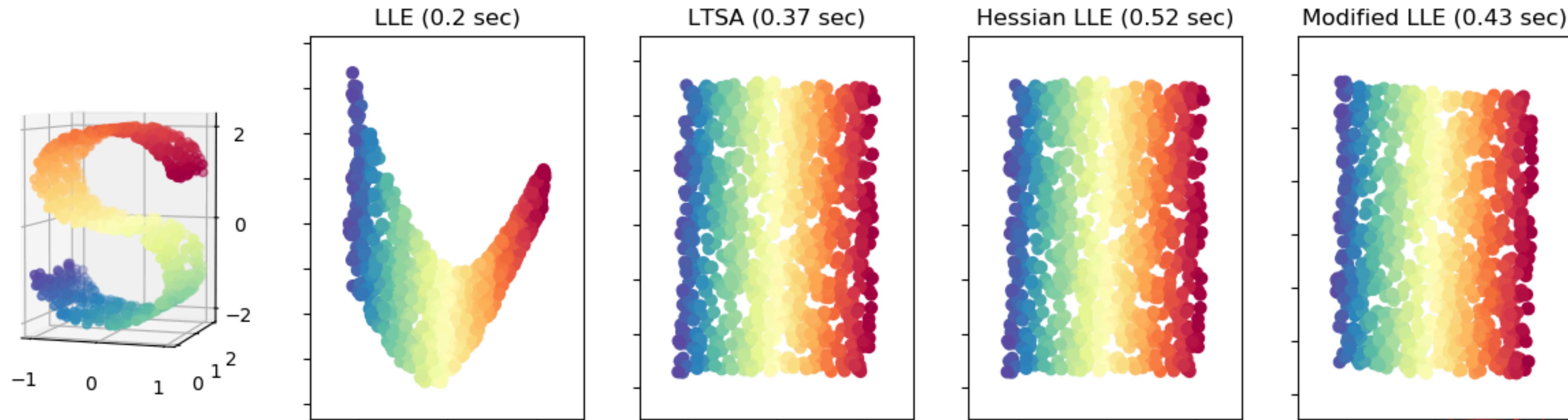
?

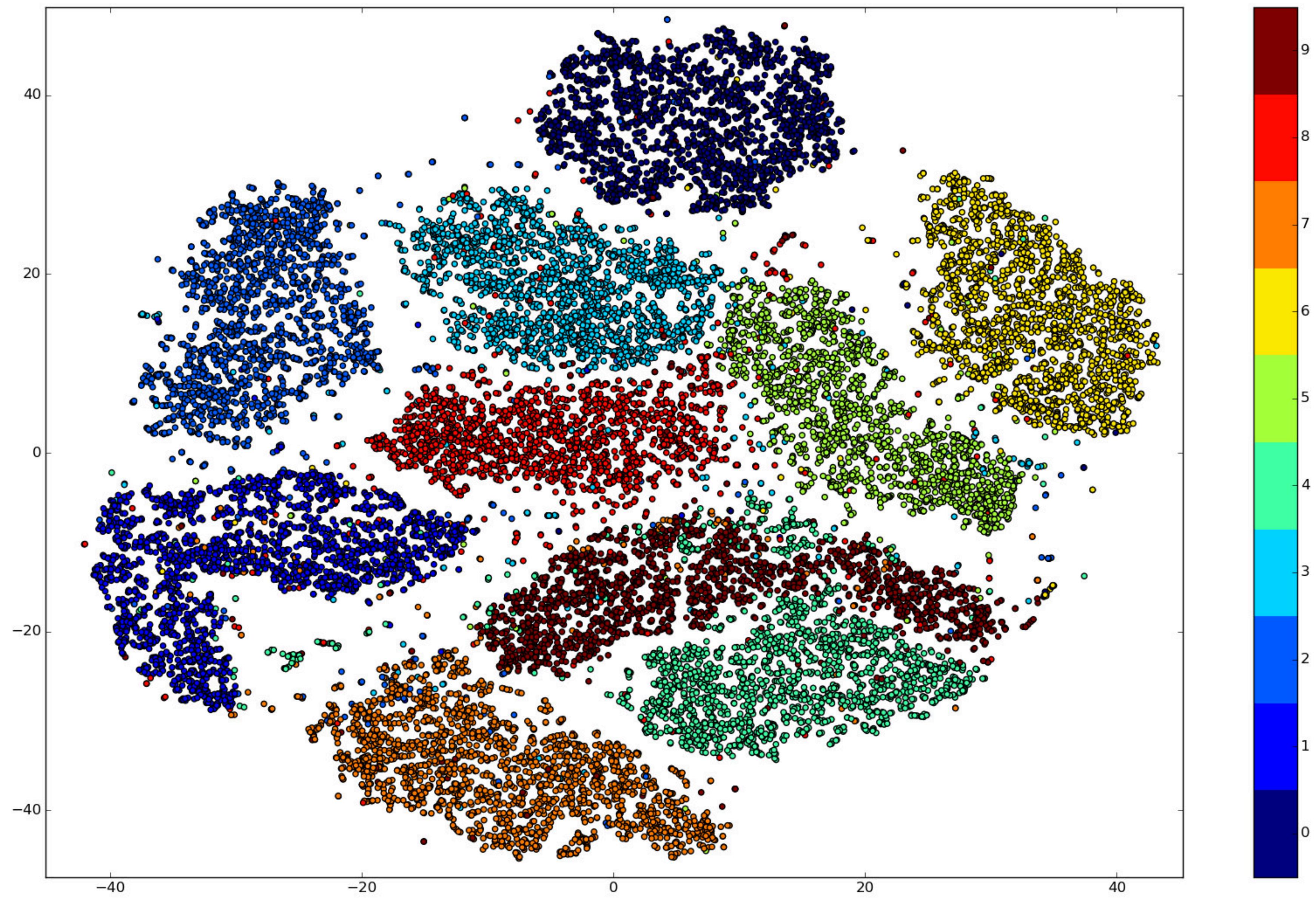
Isomap: local

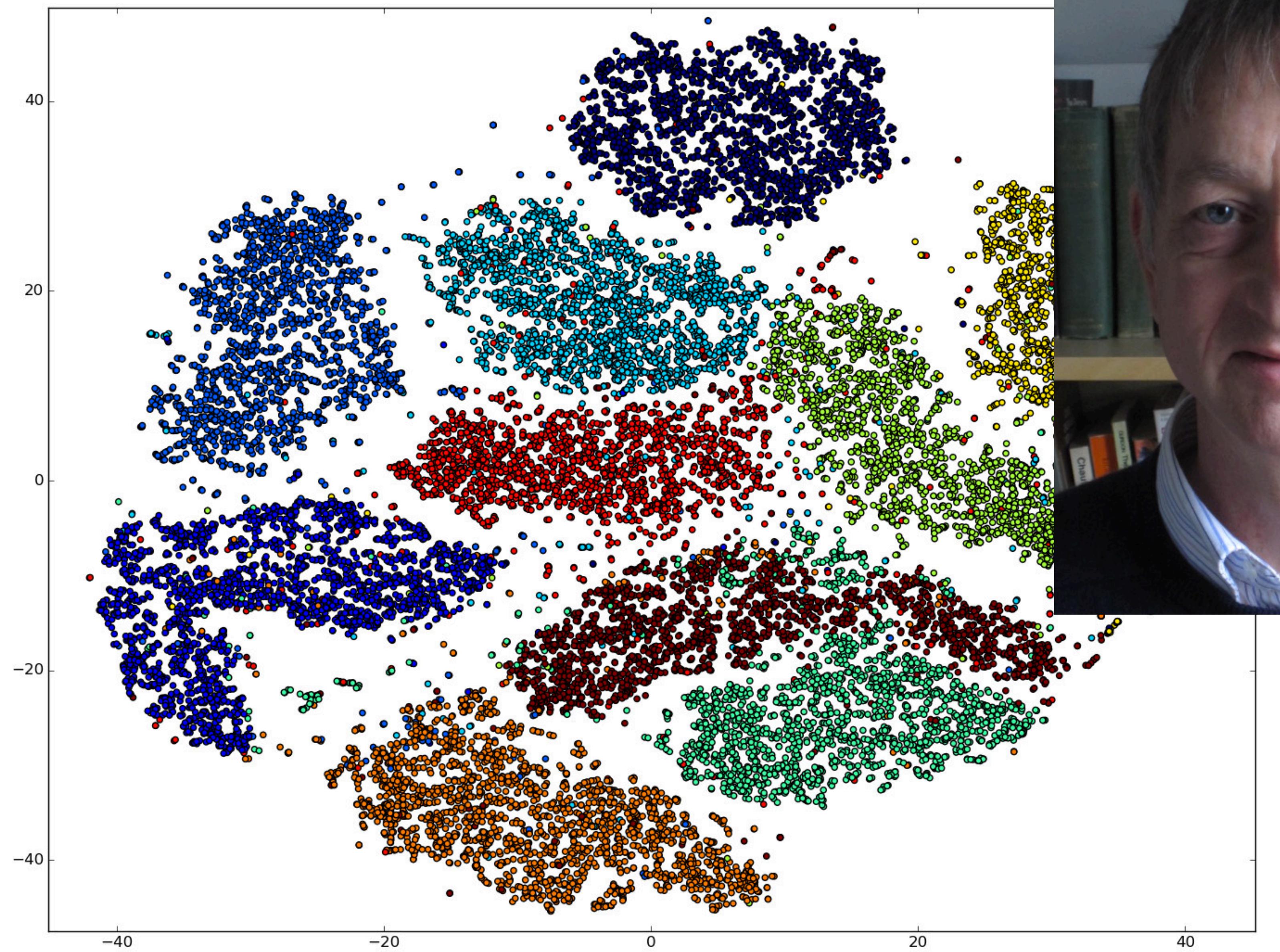
**t-SNE**

**t-distributed Stochastic  
Neighbor Embedding**

# Manifold Learning with 1000 points, 10 neighbors

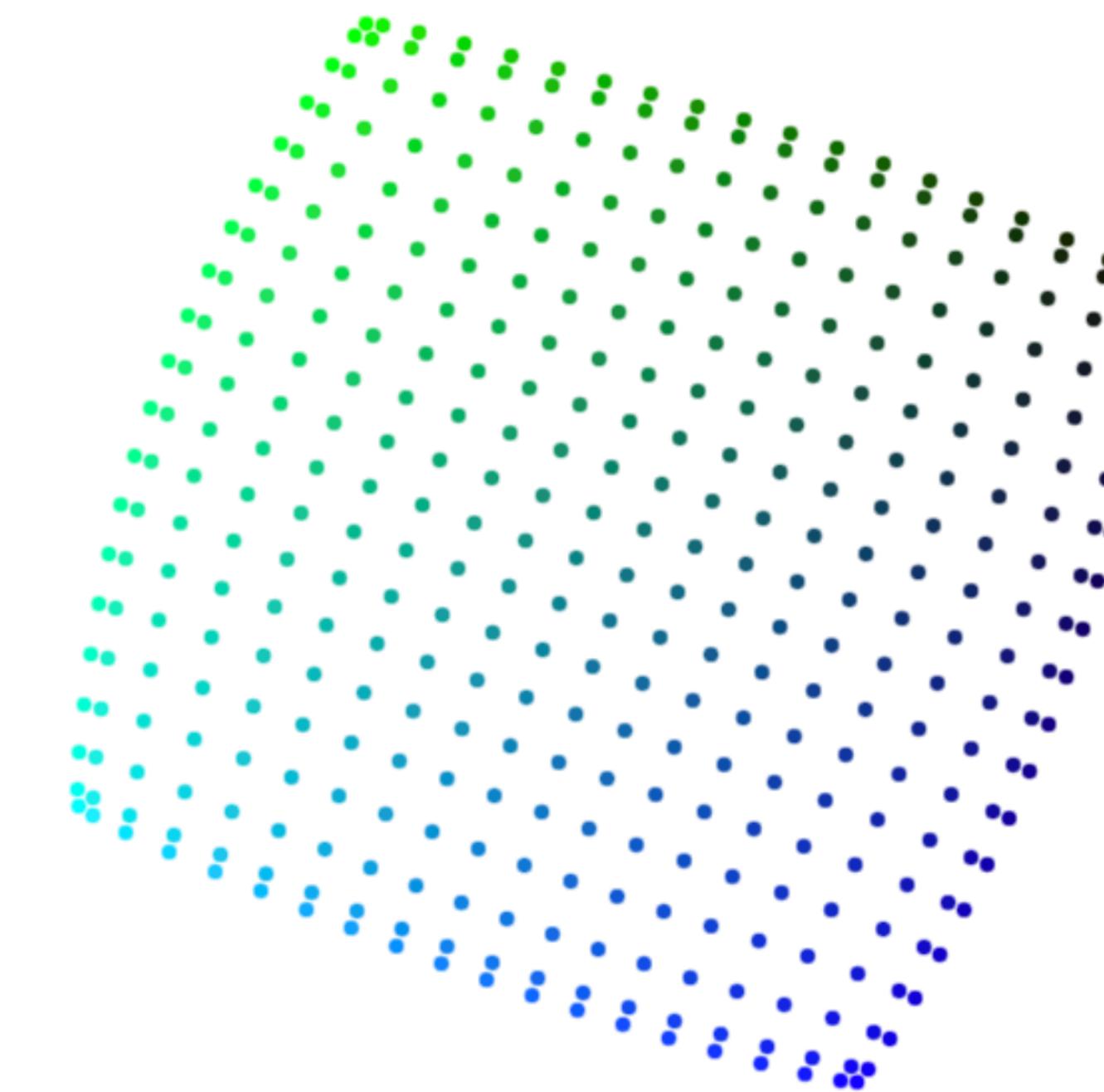






# How to Use t-SNE Effectively

Although extremely useful for visualizing high-dimensional data, t-SNE plots can sometimes be mysterious or misleading. By exploring how it behaves in simple cases, we can learn to use it more effectively.



Step  
187

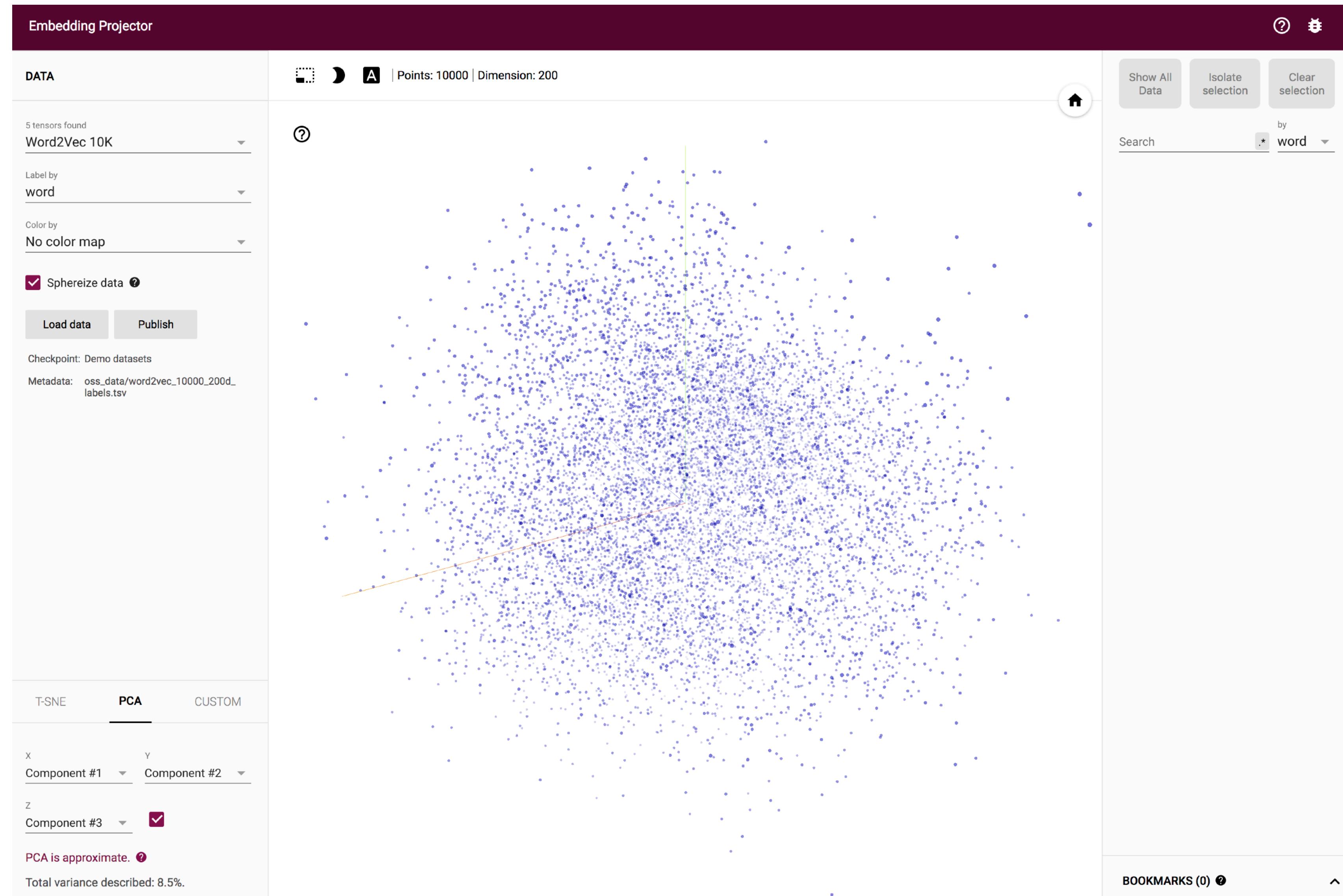
Points Per Side 20

Perplexity 10

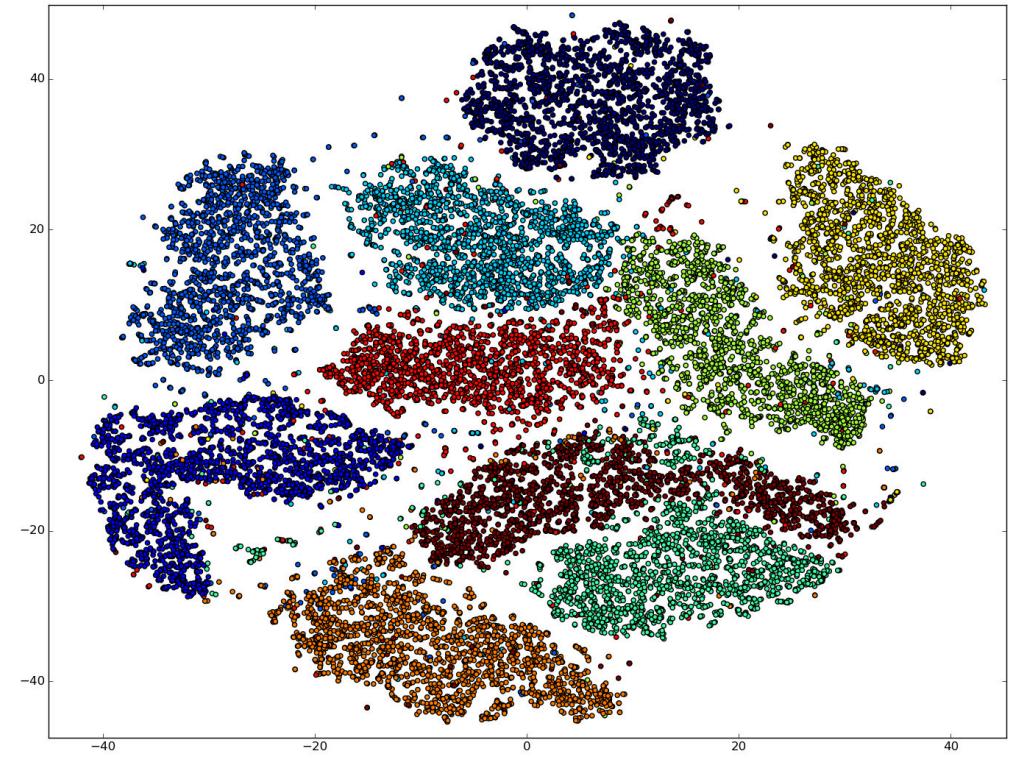
Epsilon 5

A square grid with equal spacing between points.  
Try convergence at different sizes.

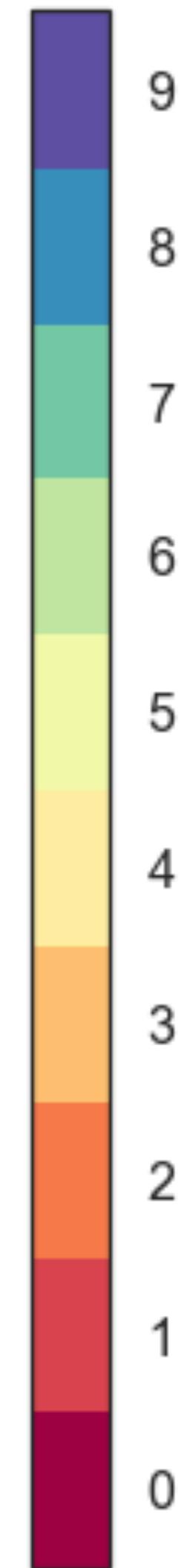
<https://projector.tensorflow.org/>



# UMAP



MNIST Digits Embedded via UMAP



- PCA: linear, basic, simple, and transparent
- MDS – t-SNE/UMAP – Isomap

# Maps

So what's the challenges in  
drawing maps?

# **What is your favorite map projection?**

1. Pick one and read about it.
2. Explain:
  1. What is the objective of the projection? What does it preserve?
  2. What does it ignore?
  3. Why is it cool?

[www.jasondavies.com/maps/transition/](http://www.jasondavies.com/maps/transition/)

WHAT YOUR FAVORITE  
**MAP PROJECTION**  
SAYS ABOUT YOU

MERCATOR

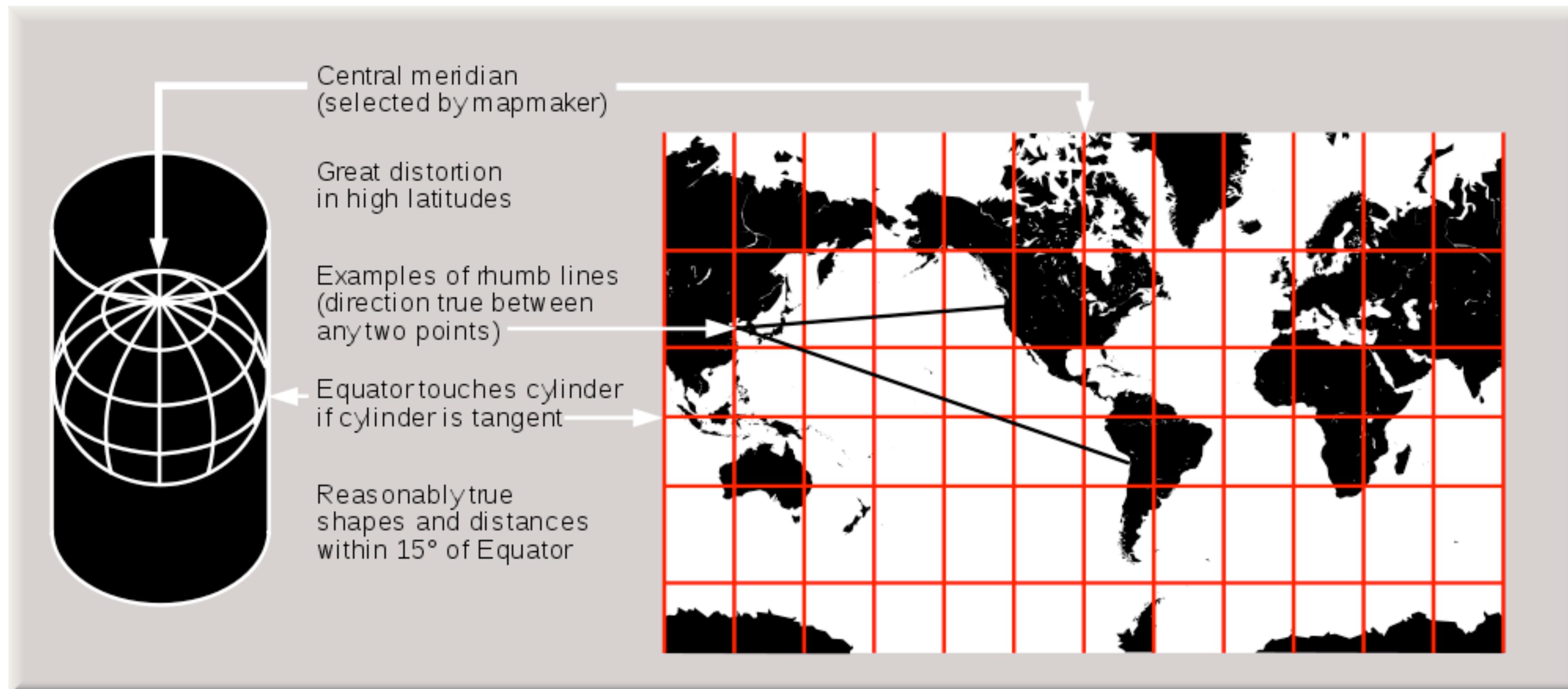


YOU'RE NOT REALLY INTO MAPS.



YOU'RE IN  
MERCATOR  
SQUARE.  
YOU LIKE

# How is it created?



[http://en.wikipedia.org/wiki/Mercator\\_projection](http://en.wikipedia.org/wiki/Mercator_projection)

# How do you navigate?



[http://en.wikipedia.org/wiki/Age\\_of\\_Discovery](http://en.wikipedia.org/wiki/Age_of_Discovery)



(SE, 1000km)



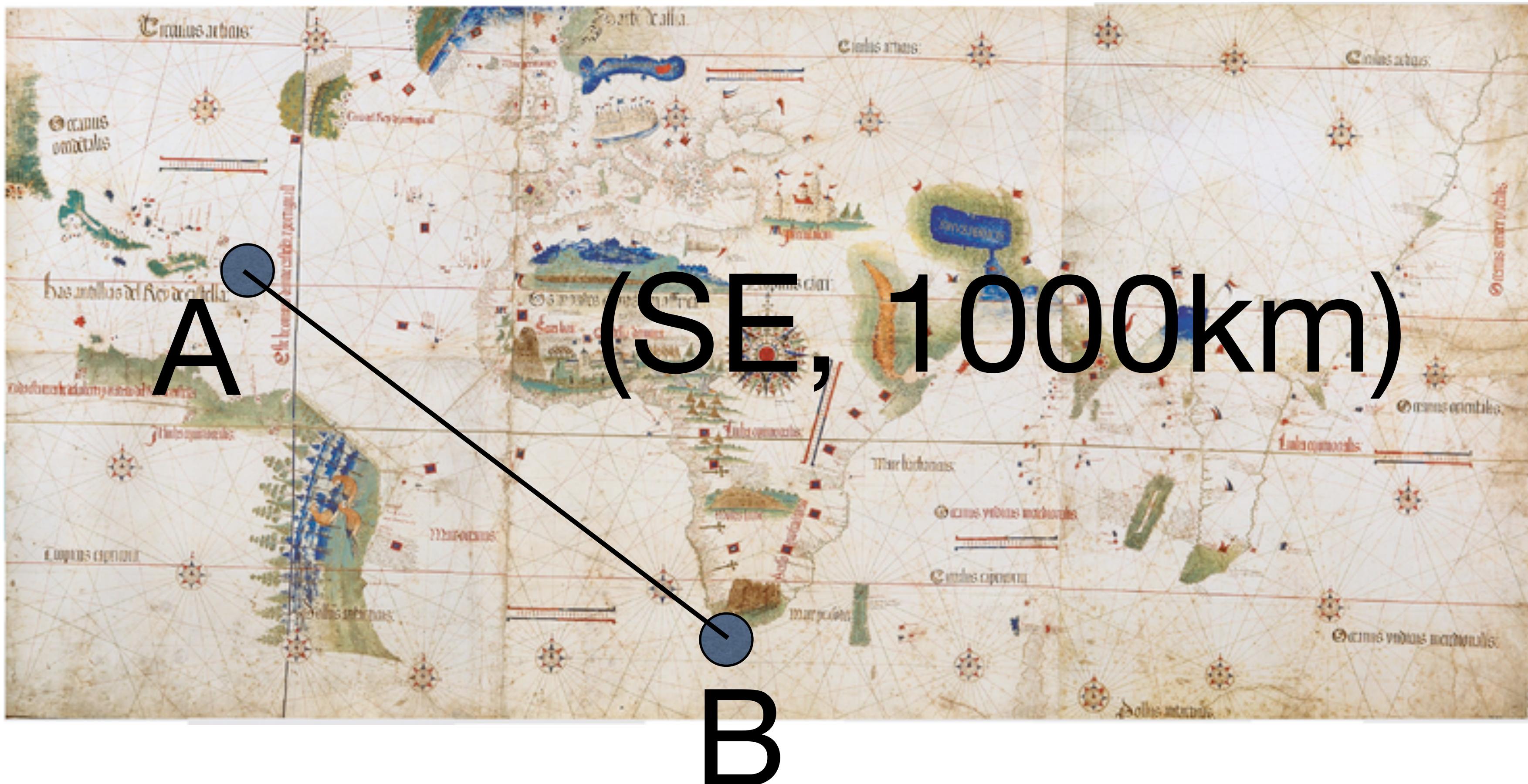


(SE, 1000km)

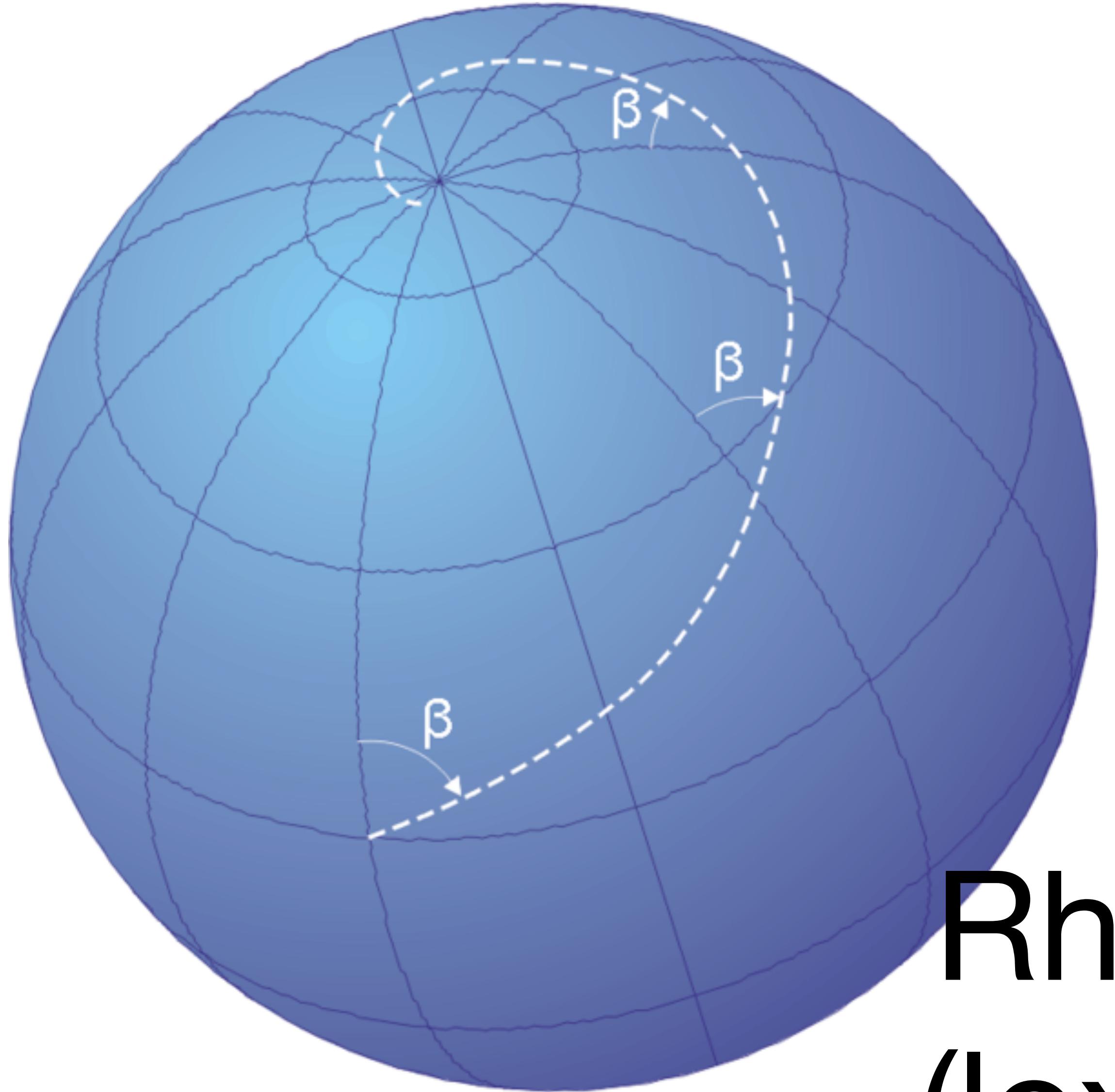
“Bearing”



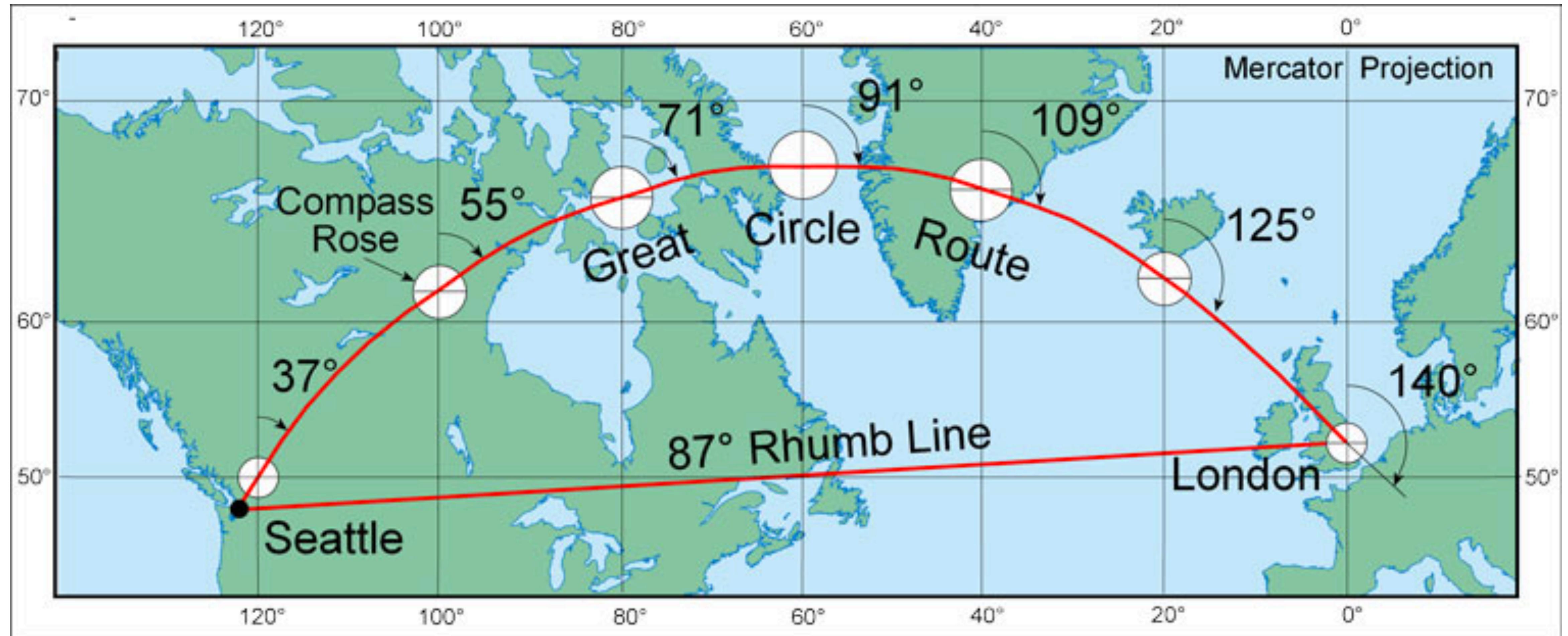
# Best map for navigation?



Is this still the best for  
air-routes?



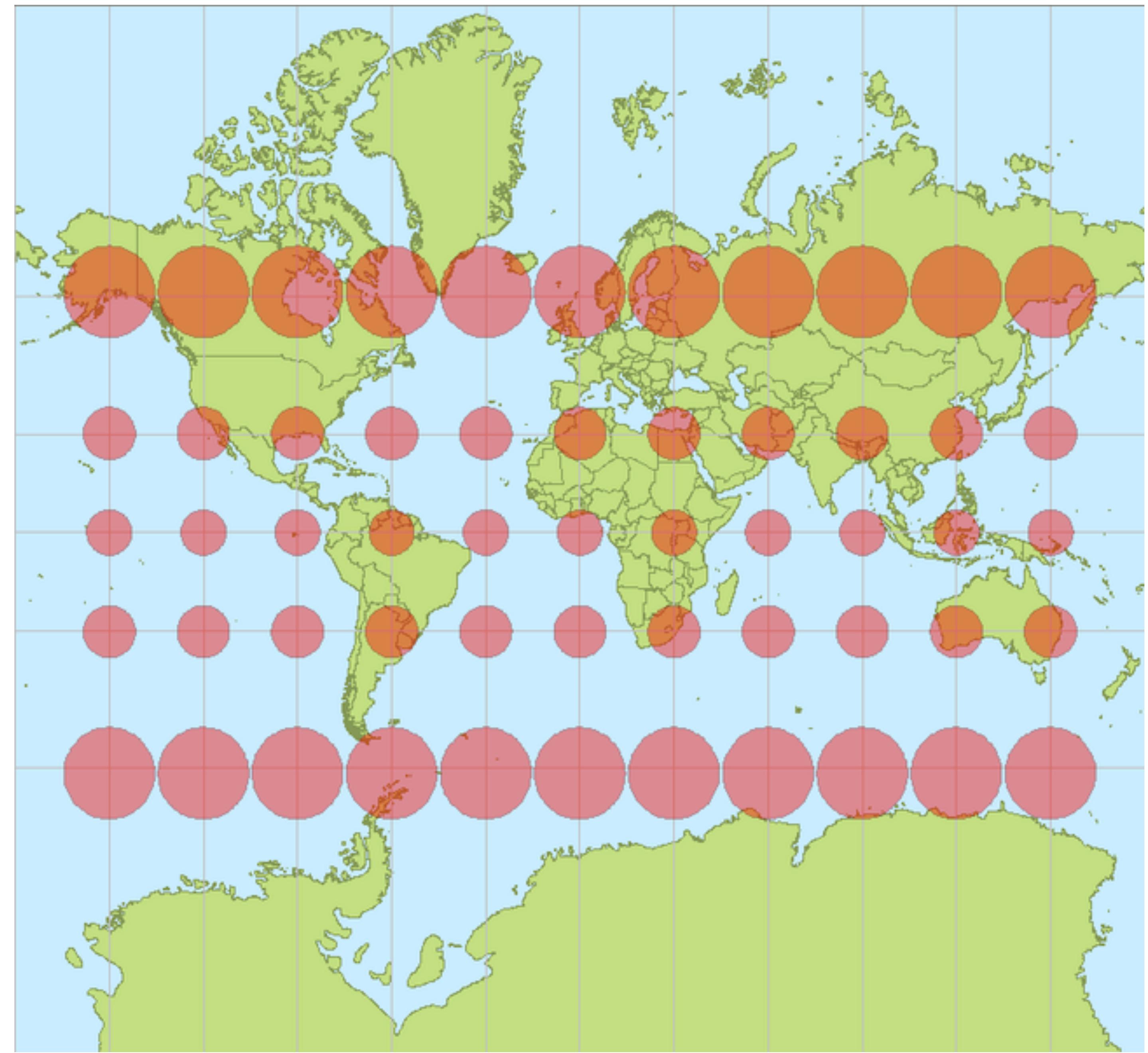
**Rhumb line  
(loxodrome)**

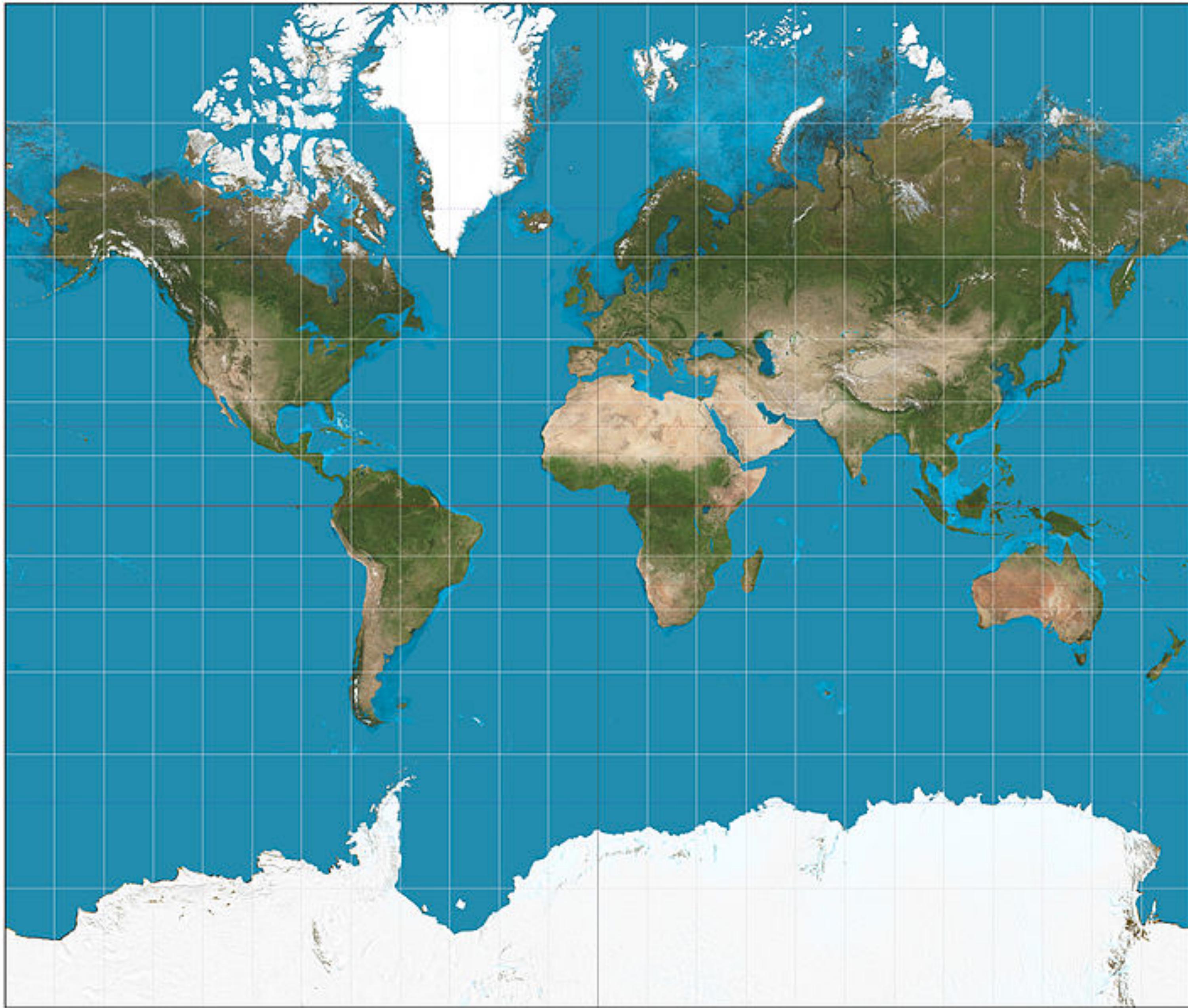


Ok. Mercator projection is  
very good for navigation.

Any drawbacks?

It doesn't preserve  
area.







# The True Size of Africa

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

Graphic layout for visualization only ( some countries are cut and rotated )  
But the conclusions are very accurate: refer to table below for exact data

COUNTRY	AREA x 1000 km <sup>2</sup>
China	9.597
USA	9.629
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
<b>TOTAL</b>	<b>30.102</b>
<b>AFRICA</b>	<b>30.221</b>

In addition to the well known social issues of *illiteracy* and *innumeracy*, there also should be such a concept as "*immappancy*", meaning insufficient 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!



## Top 100 Countries

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

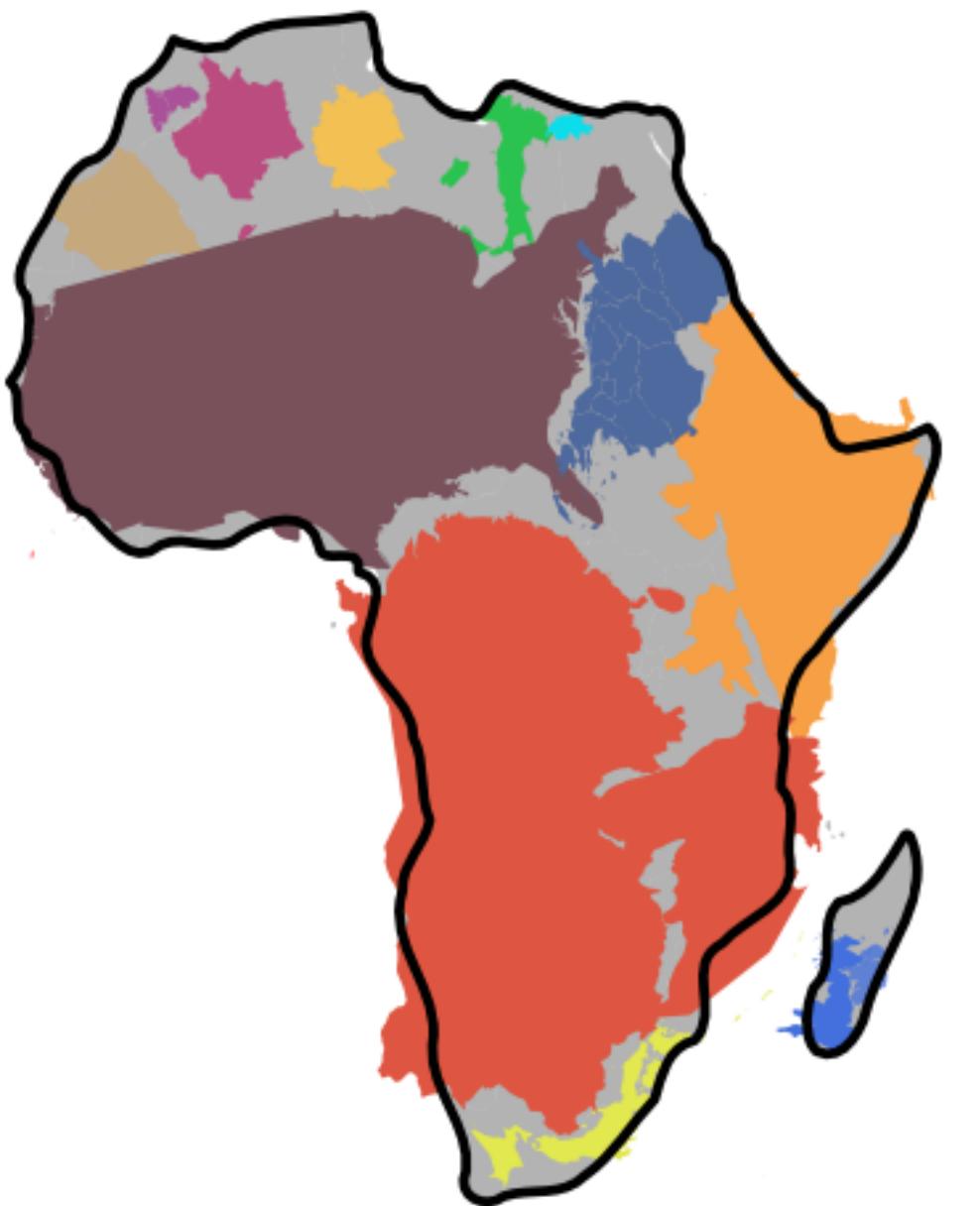
	AREA km <sup>2</sup>	%
1 Russia	17.098.242	11,50
2 Canada	9.984.670	6,70
3 China	9.596.961	6,40
4 United States	9.629.091	6,40
5 Brazil	8.514.877	5,70
6 Australia	7.692.024	5,20
7 India	3.987.263	2,30
8 Argentina	2.780.400	2,00
9 Kazakhstan	2.724.900	1,80
10 Sudan	2.505.813	1,70
11 Algeria	2.381.741	1,60
12 Congo	2.344.858	1,60
13 Greenland	2.166.086	1,50
14 Saudi Arabia	2.149.690	1,40
15 Mexico	1.964.375	1,30
16 Indonesia	1.860.360	1,30
17 Libya	1.759.540	1,20
18 Iran	1.628.750	1,10
19 Mongolia	1.564.100	1,10
20 Peru	1.285.216	0,86
21 Chad	1.284.000	0,86
22 Niger	1.287.000	0,85
23 Angola	1.246.700	0,85
24 Mali	1.240.192	0,83
25 South Africa	1.221.037	0,82
26 Colombia	1.141.748	0,74
27 Ethiopia	1.104.300	0,74
28 Bolivia	1.098.581	0,69
29 Mauritania	1.025.520	0,69
30 Egypt	1.002.000	0,67
31 Tanzania	945.087	0,63
32 Nigeria	923.768	0,62
33 Venezuela	912.050	0,61
34 Namibia	824.156	0,55
35 Mozambique	801.590	0,54
36 Pakistan	796.095	0,53
37 Turkey	783.562	0,53
38 Chile	756.102	0,51
39 Zambia	752.612	0,51
40 Myanmar	676.578	0,45
41 Afghanistan	652.090	0,44
42 Somalia	637.657	0,43
43 France	632.834	0,43
44 C. African Rep	622.984	0,42
45 Ukraine	603.500	0,41
46 Madagascar	587.041	0,39
47 Botswana	582.000	0,39
48 Kenya	580.367	0,39
49 Yemen	527.968	0,35
50 Thailand	513.120	0,34
51 Spain	505.992	0,34
52 Turkmenistan	488.100	0,33
53 Cameroon	475.442	0,32
54 Papua New Guinea	462.840	0,31
55 Uzbekistan	447.400	0,30
56 Morocco	446.550	0,30
57 Sweden	441.370	0,30
58 Iraq	438.317	0,29
59 Paraguay	406.752	0,27
60 Zimbabwe	390.757	0,26
61 Japan	377.930	0,25
62 Germany	357.154	0,24
63 Rep. of Congo	342.000	0,23
64 Finland	338.419	0,23
65 Vietnam	331.212	0,22
66 Malaysia	330.863	0,22
67 Norway	323.862	0,22
68 Côte d'Ivoire	322.463	0,22
69 Poland	312.685	0,21
70 Oman	309.500	0,21
71 Italy	301.336	0,20
72 Philippines	300.000	0,20
73 Burkina Faso	274.222	0,18
74 New Zealand	270.467	0,18
75 Gabon	267.668	0,18
76 Western Sahara	266.000	0,18
77 Ecuador	256.369	0,20
78 Guinea	245.857	0,17
79 United Kingdom	242.900	0,16
80 Uganda	241.038	0,16
81 Ghana	238.539	0,16
82 Romania	238.391	0,16
83 Laos	236.800	0,16
84 Guyana	214.969	0,14
85 Belarus	207.600	0,14
86 Kyrgyzstan	199.951	0,13
87 Senegal	196.722	0,13
88 Syria	185.180	0,12
89 Cambodia	181.035	0,12
90 Uruguay	176.215	0,12
91 Suriname	163.820	0,11
92 Tunisia	163.610	0,11
93 Nepal	147.181	0,10
94 Bangladesh	143.998	0,10
95 Tajikistan	143.100	0,10
96 Greece	131.957	0,09
97 Nicaragua	130.373	0,09
98 North Korea	120.538	0,08
99 Malawi	118.484	0,08
100 Eritrea	117.600	0,08
<b>TOP 100 TOTAL</b>	<b>132.632.524</b>	<b>89,34</b>



"The True Size of Africa"  
infographic map



The *real* true size of these  
areas, relative to Africa



How I might redo  
the map



Every projection  
**preserves something**  
(area, distance from  
center, Rhumb line, etc.)  
  
**at the cost of others.**

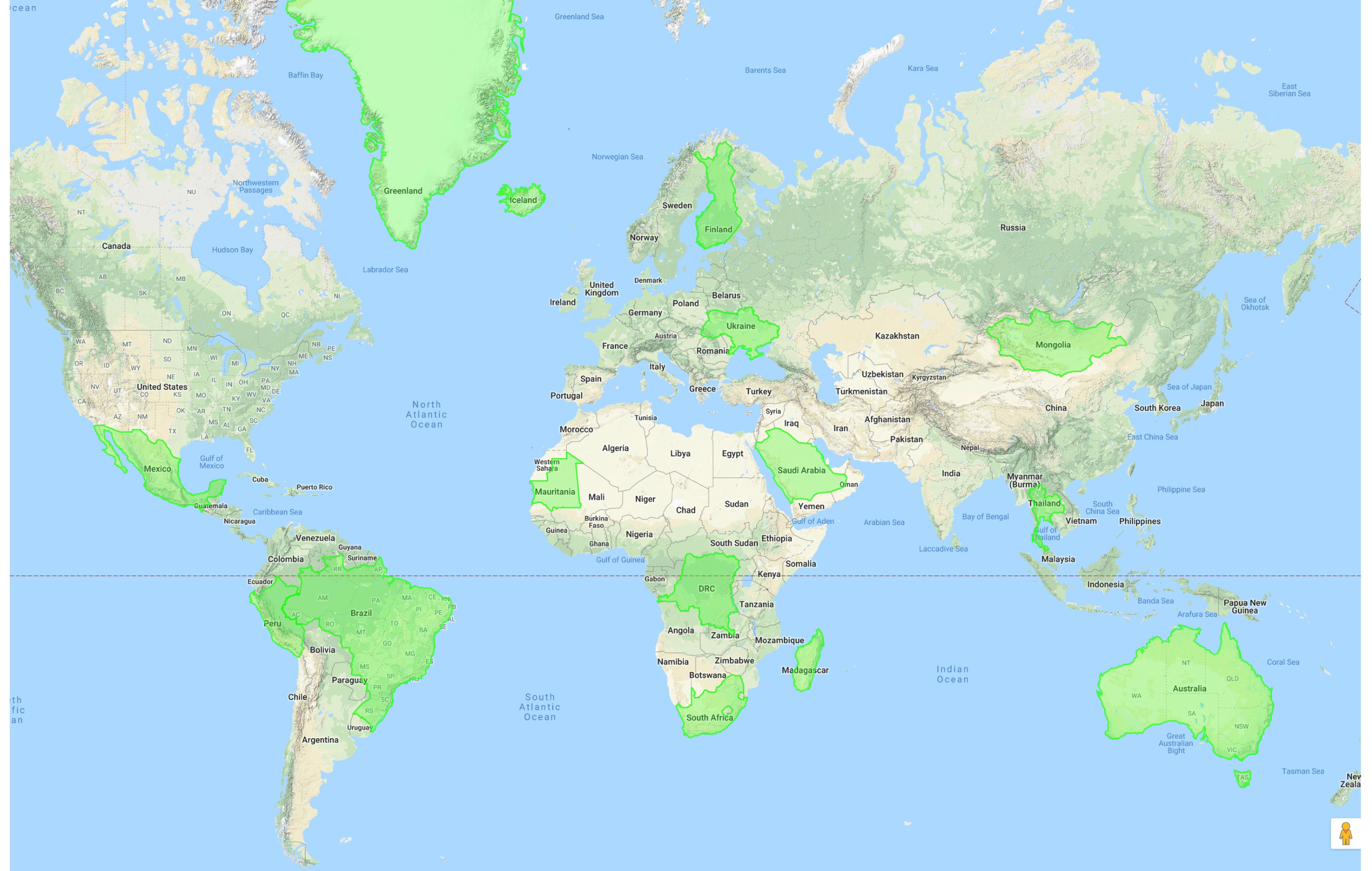
# Mercator puzzle

<http://hive.sewanee.edu/ldale/maps/10/06-LOCAL.html>

<https://bramus.github.io/mercator-puzzle-redux/>

or google

“Mercator puzzle redux”



[thetruesize.com](http://thetruesize.com)

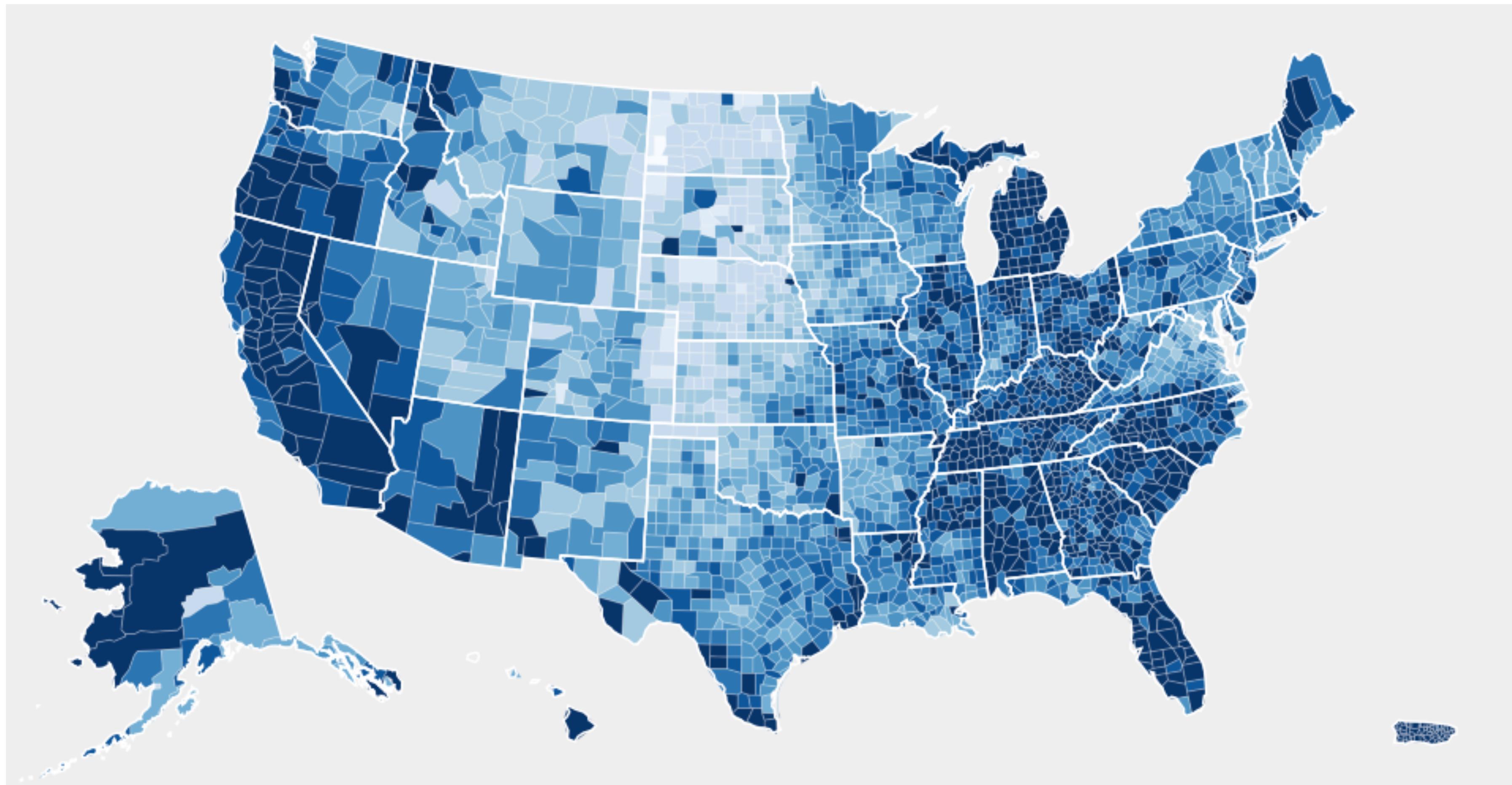
What is the most interesting example?  
(you can check out individual states in US)

So, how to visualize  
data on the map?

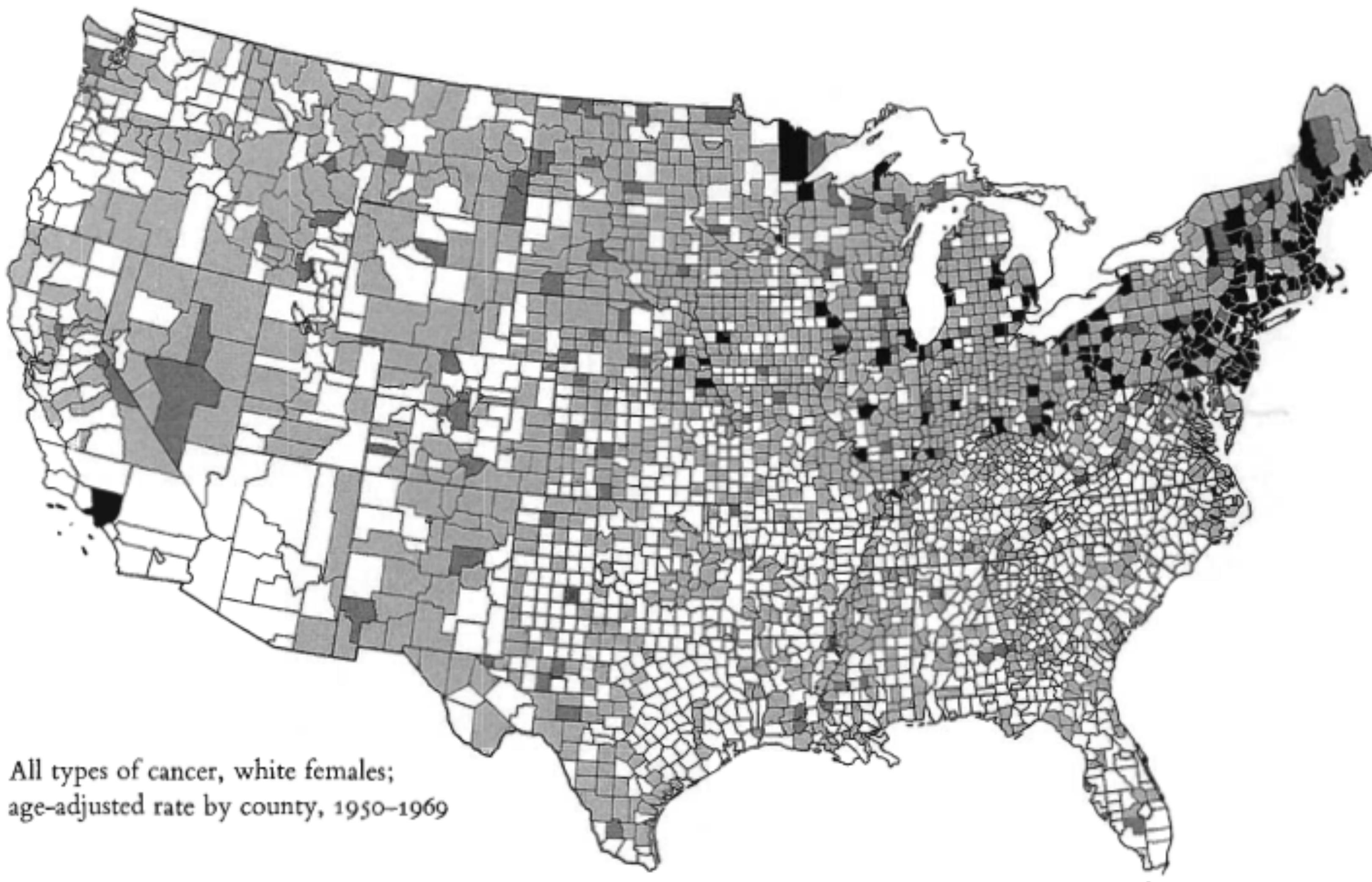
# Choropleth Maps

χώρο – + πλήθ

“region” + “multitude”



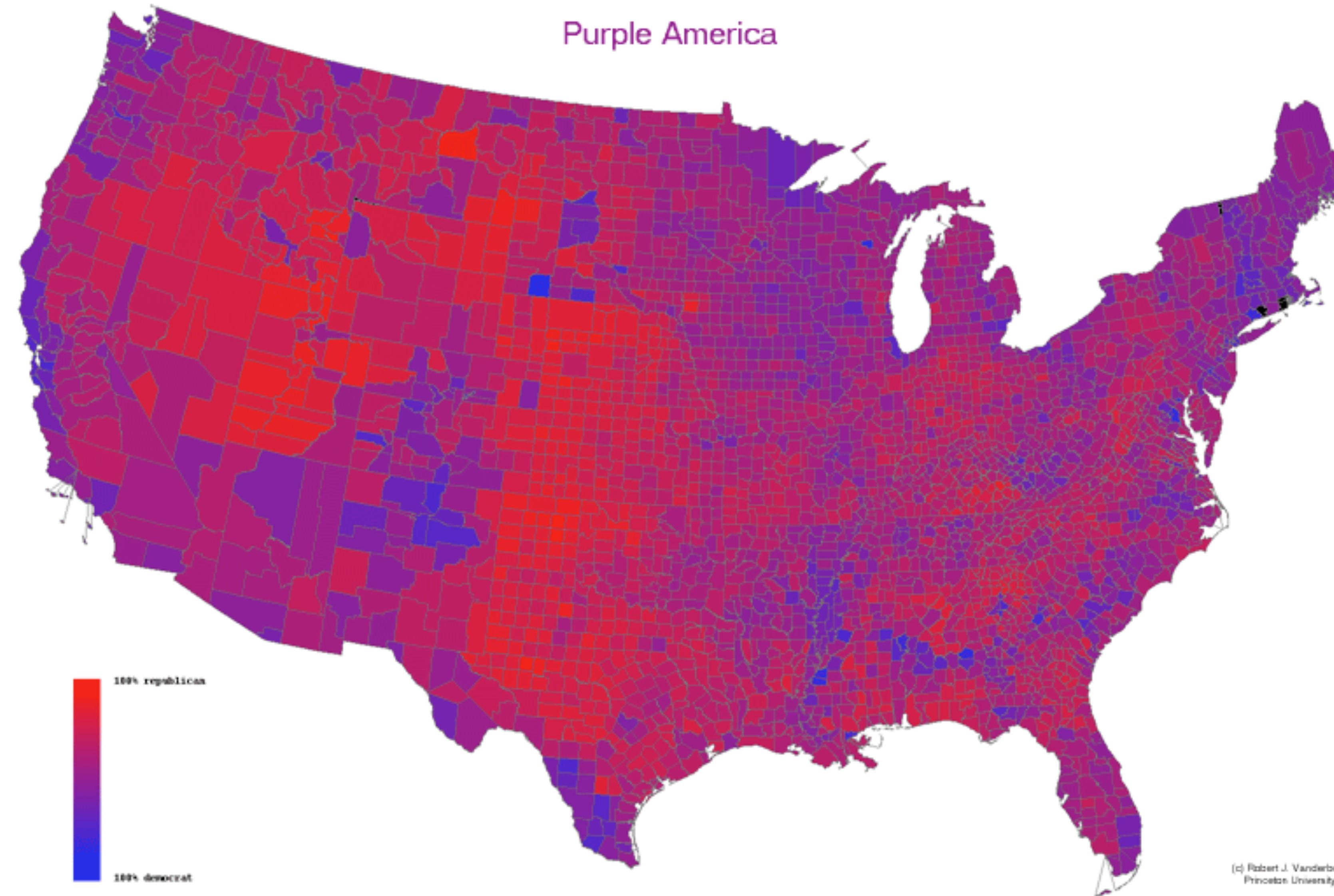
<http://mbostock.github.com/d3/ex/choropleth.html>



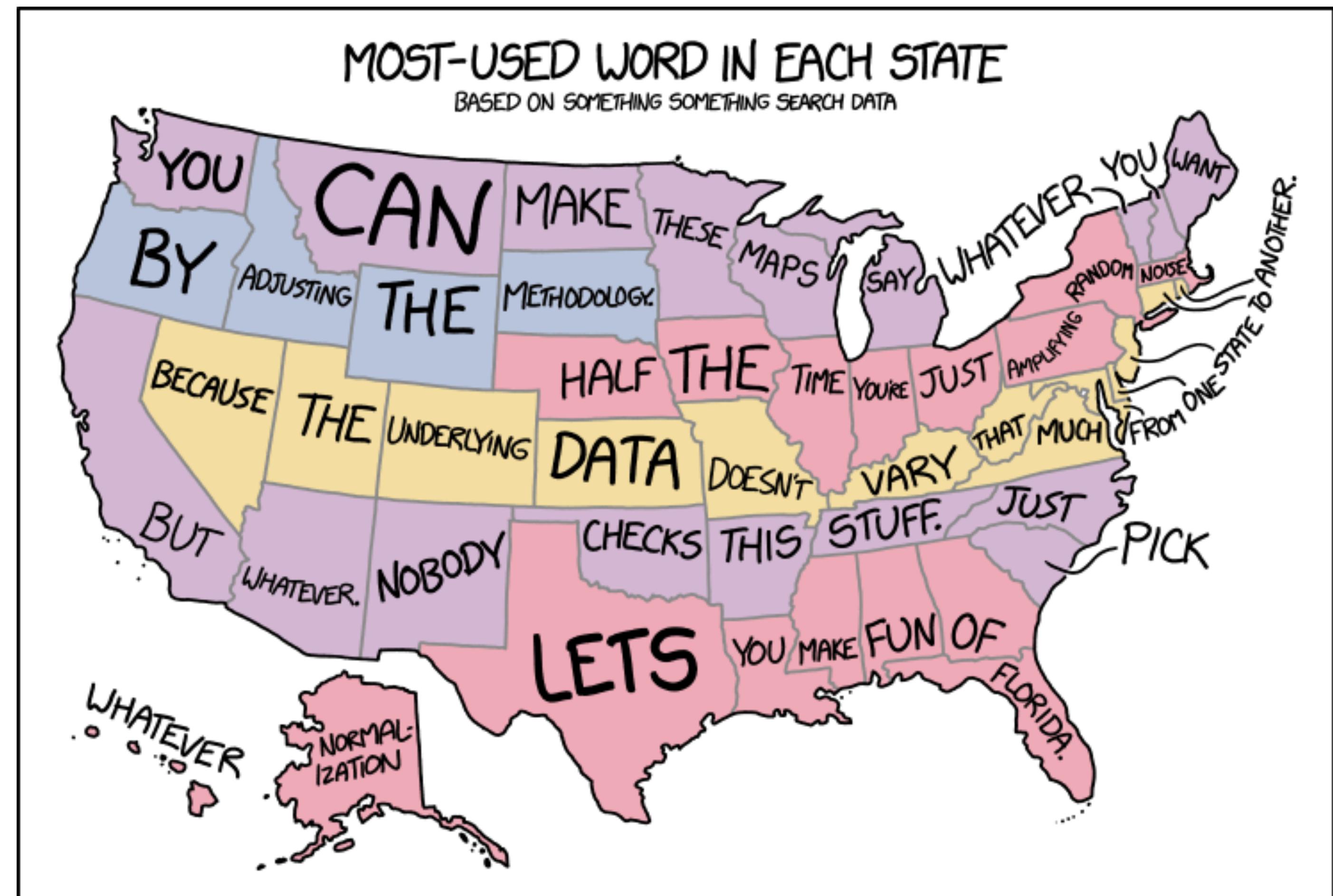
All types of cancer, white females;  
age-adjusted rate by county, 1950–1969

# 2004 Presidential Election

Purple America

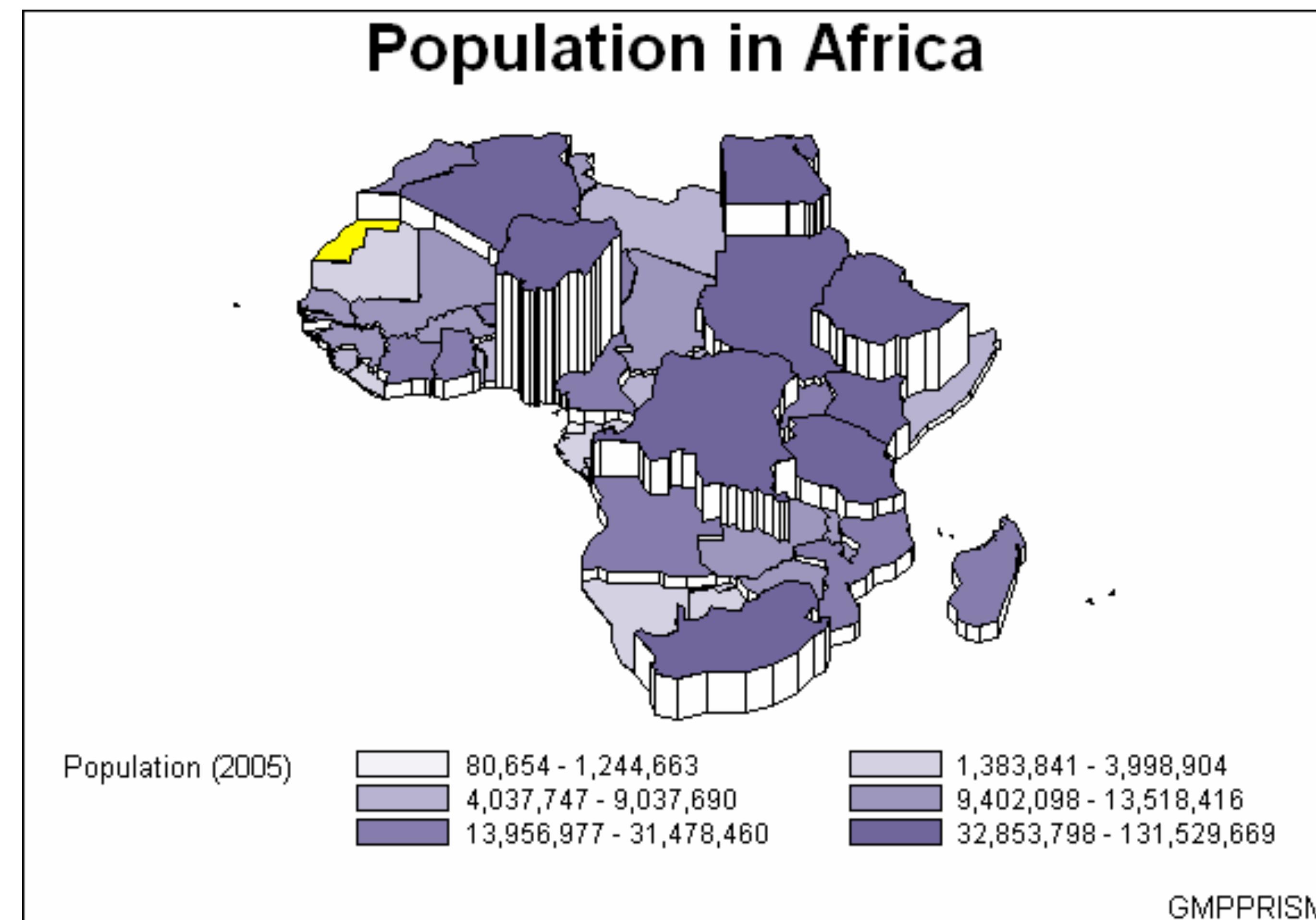


(c) Robert J. Vanderbei  
Princeton University

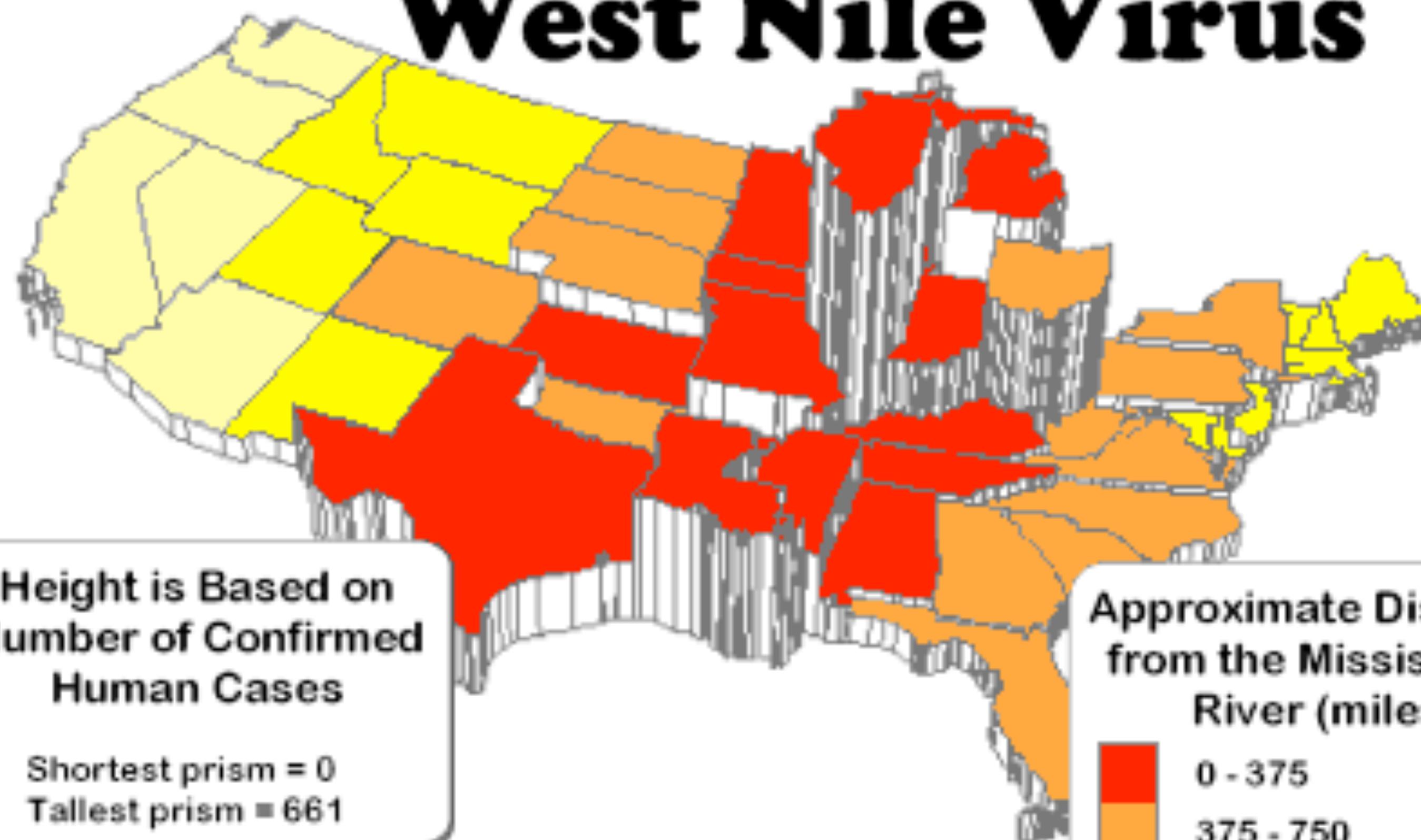


<https://xkcd.com/1845/>

# Prism map



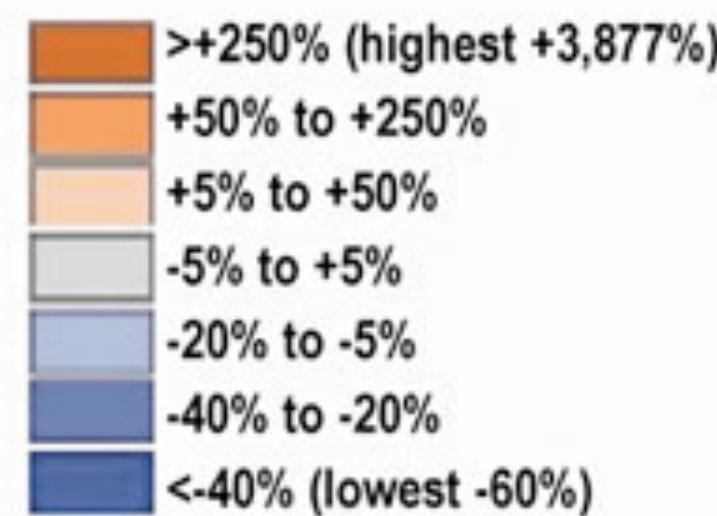
# Tracking the West Nile Virus



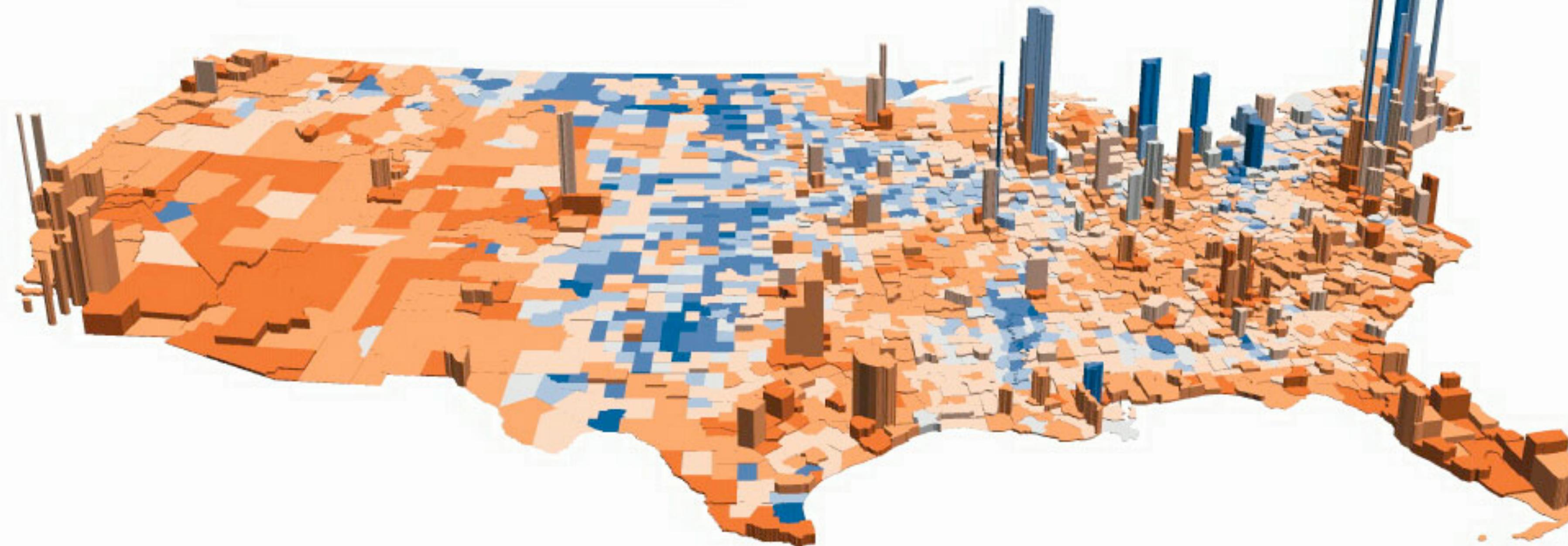
## US Population and Growth Trends

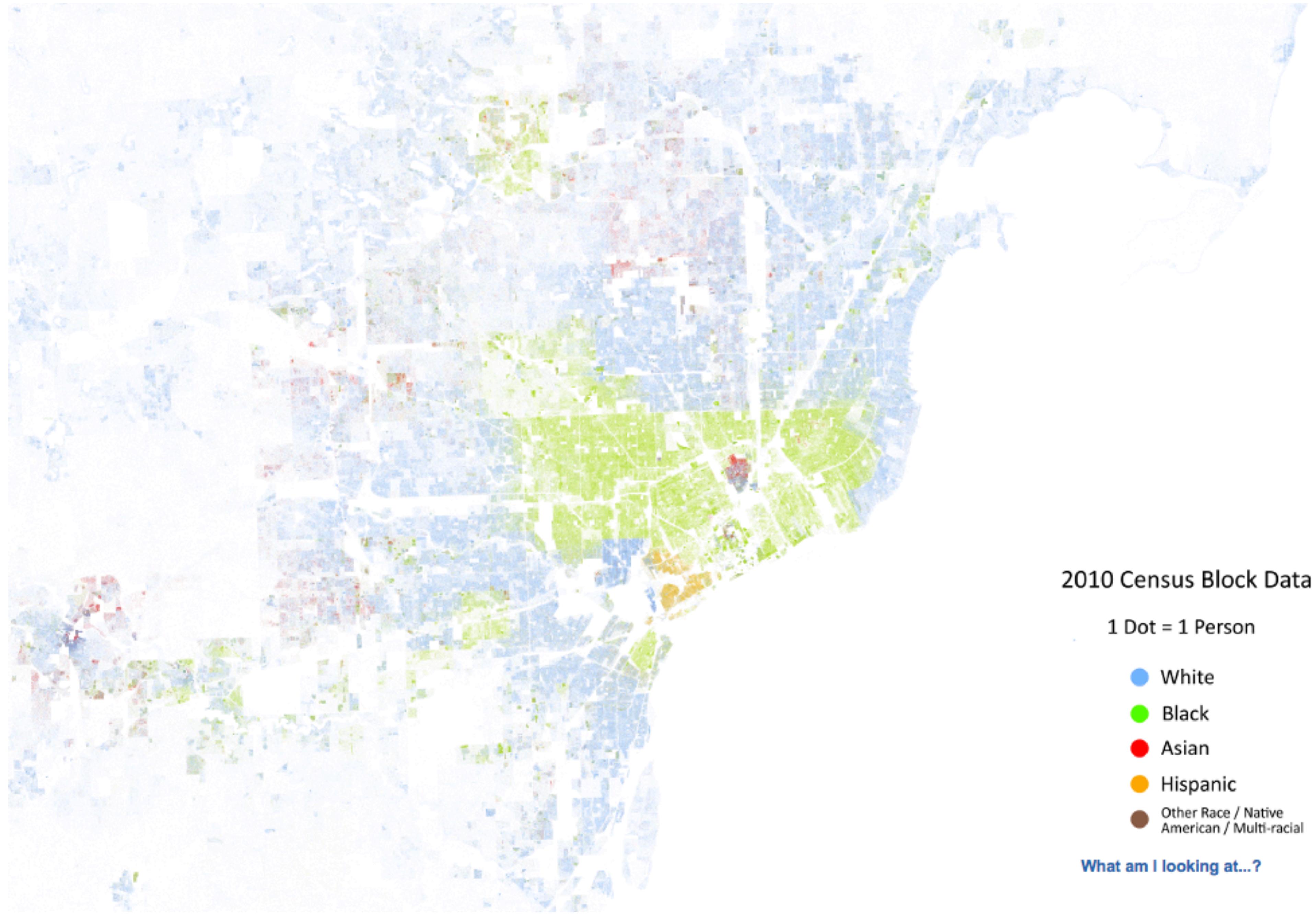
### Change in county population, 1970-2030

Projected change in county population (percent), 1970 to 2030



Each block on the map illustrates one county in the US. The height of each block is proportional to that county's population density in the year 2000, so the volume of the block is proportional to the county's total population. The color of each block shows the county's projected change in population between 1970 and 2030, with shades of orange denoting increases and blue denoting decreases. The patterns of recent population change, with growth concentrated along the coasts, in cities, and in the South and West, are projected to continue.





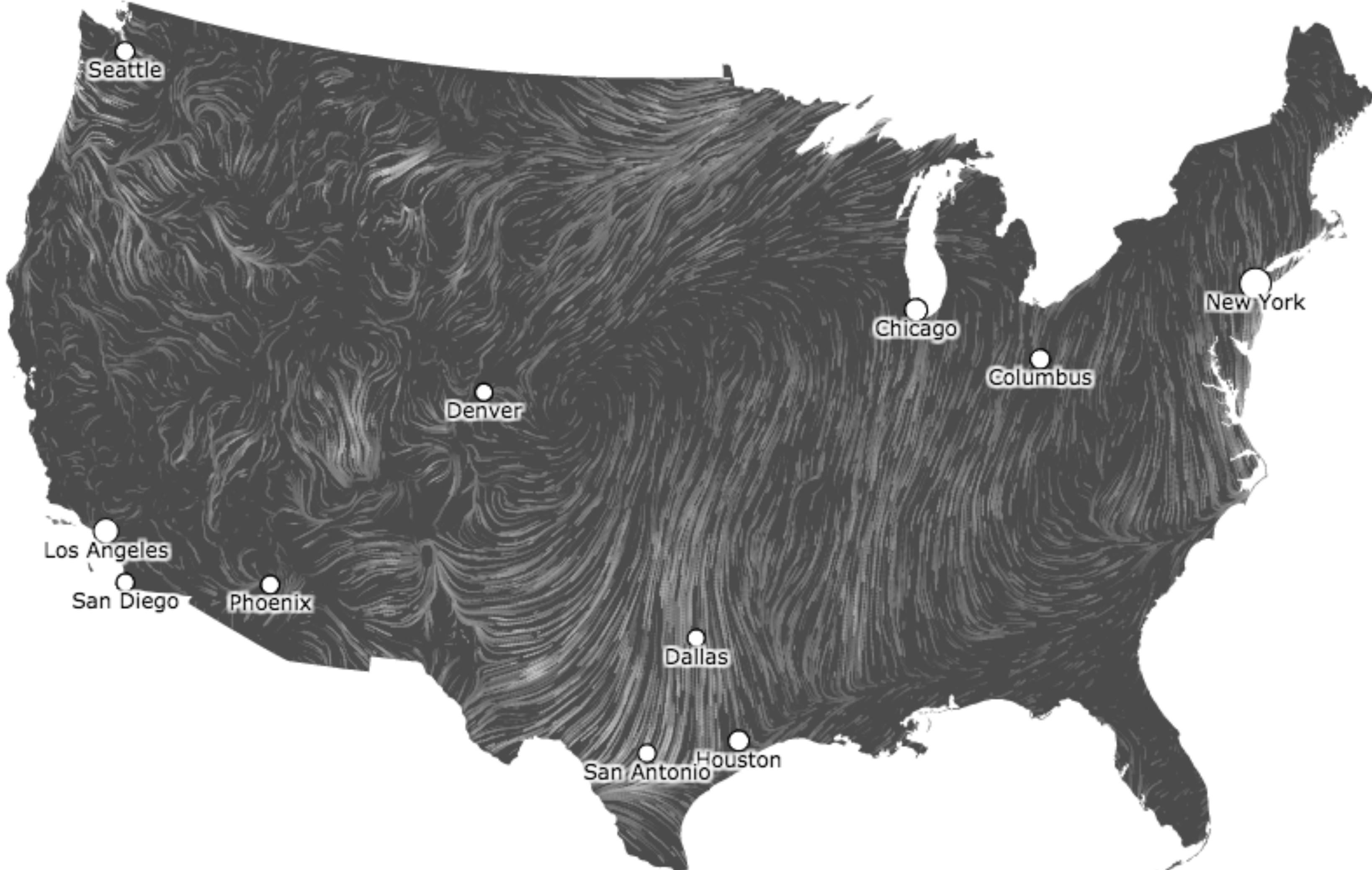
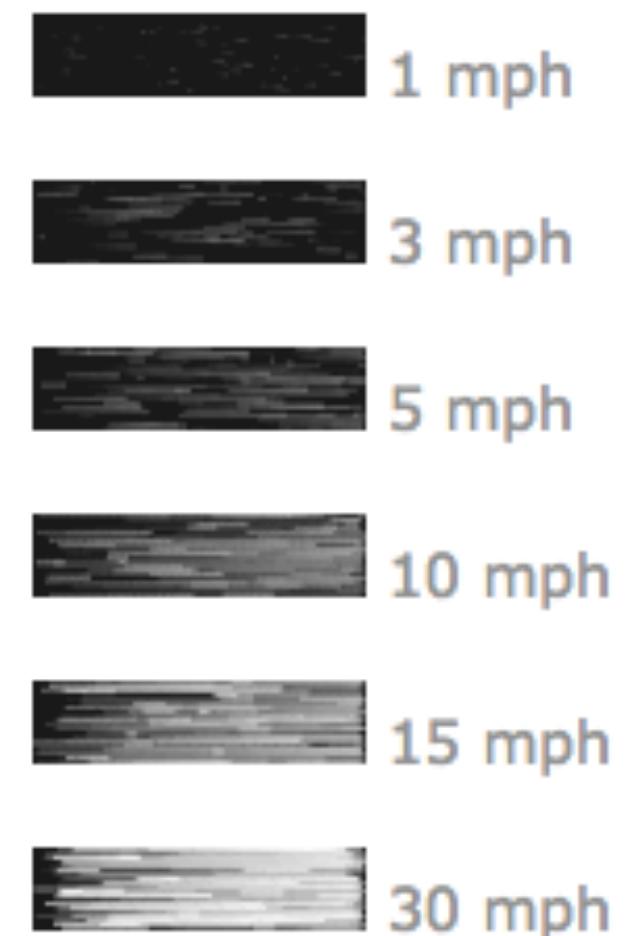
# wind map

Nov. 7, 2017

10:36 pm EST

(time of forecast download)

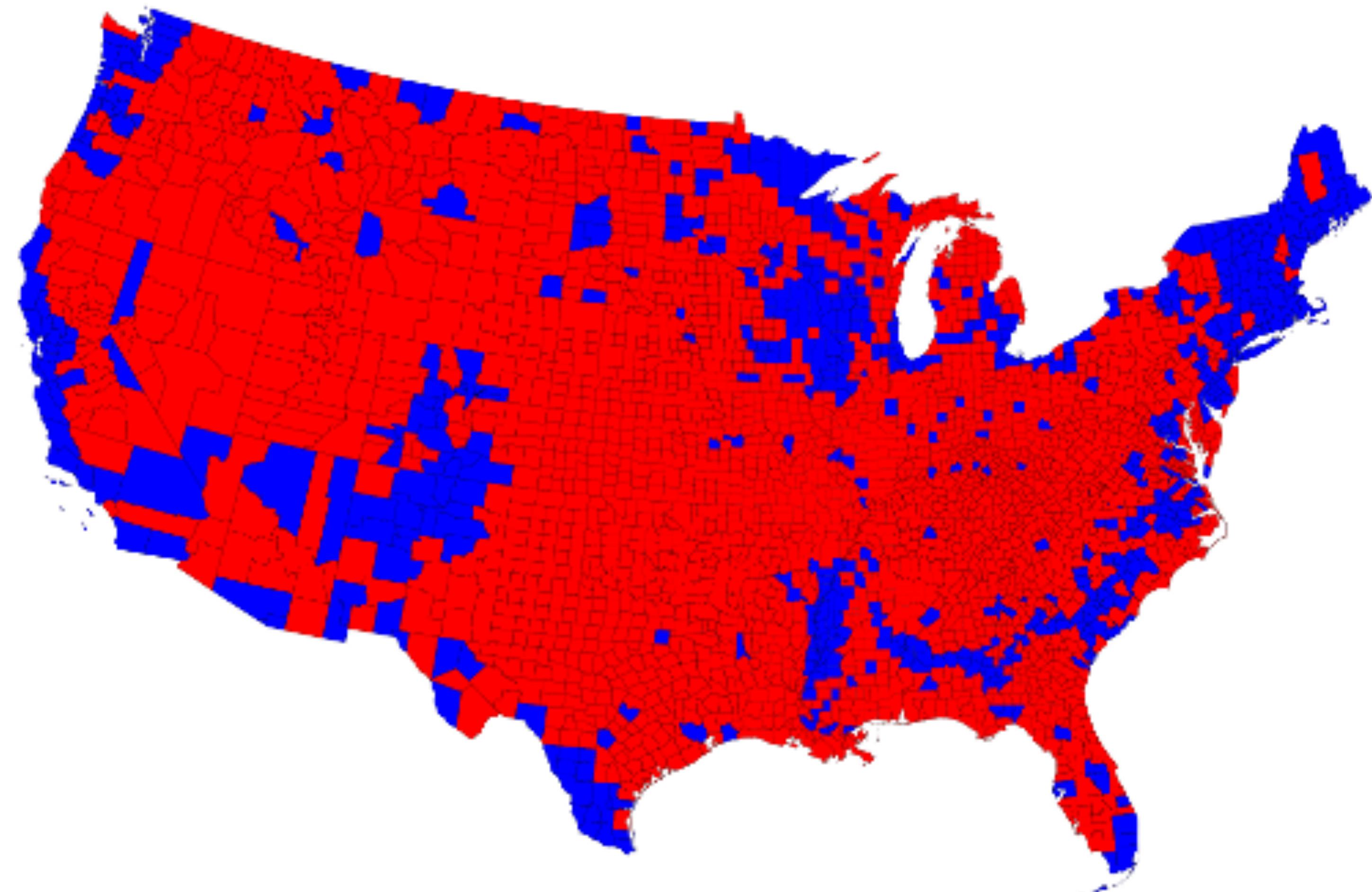
top speed: **36.1 mph**  
average: **7.1 mph**



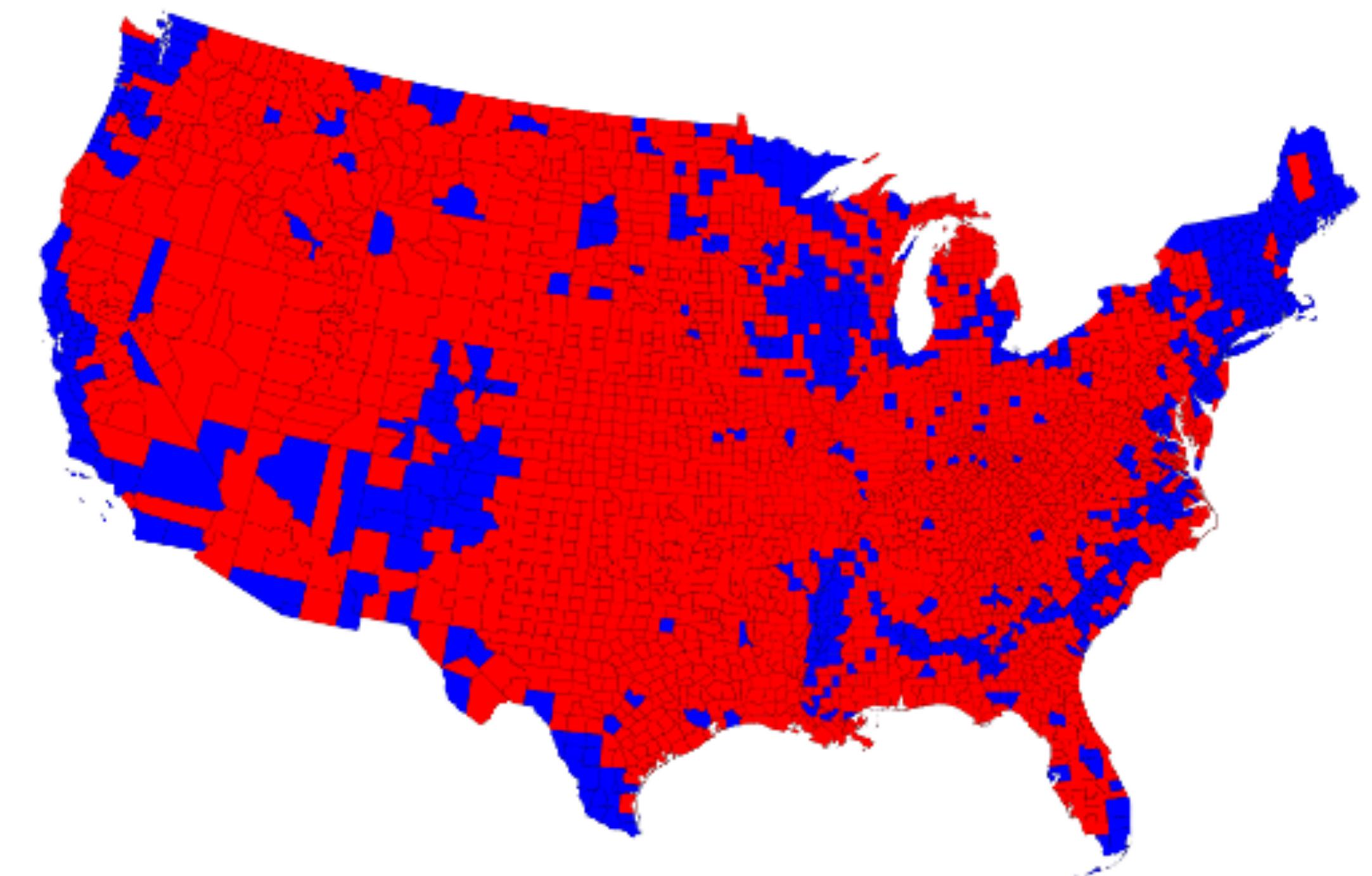


Lung cancer cases 1993-1997  
(each dot represents 10 cases)

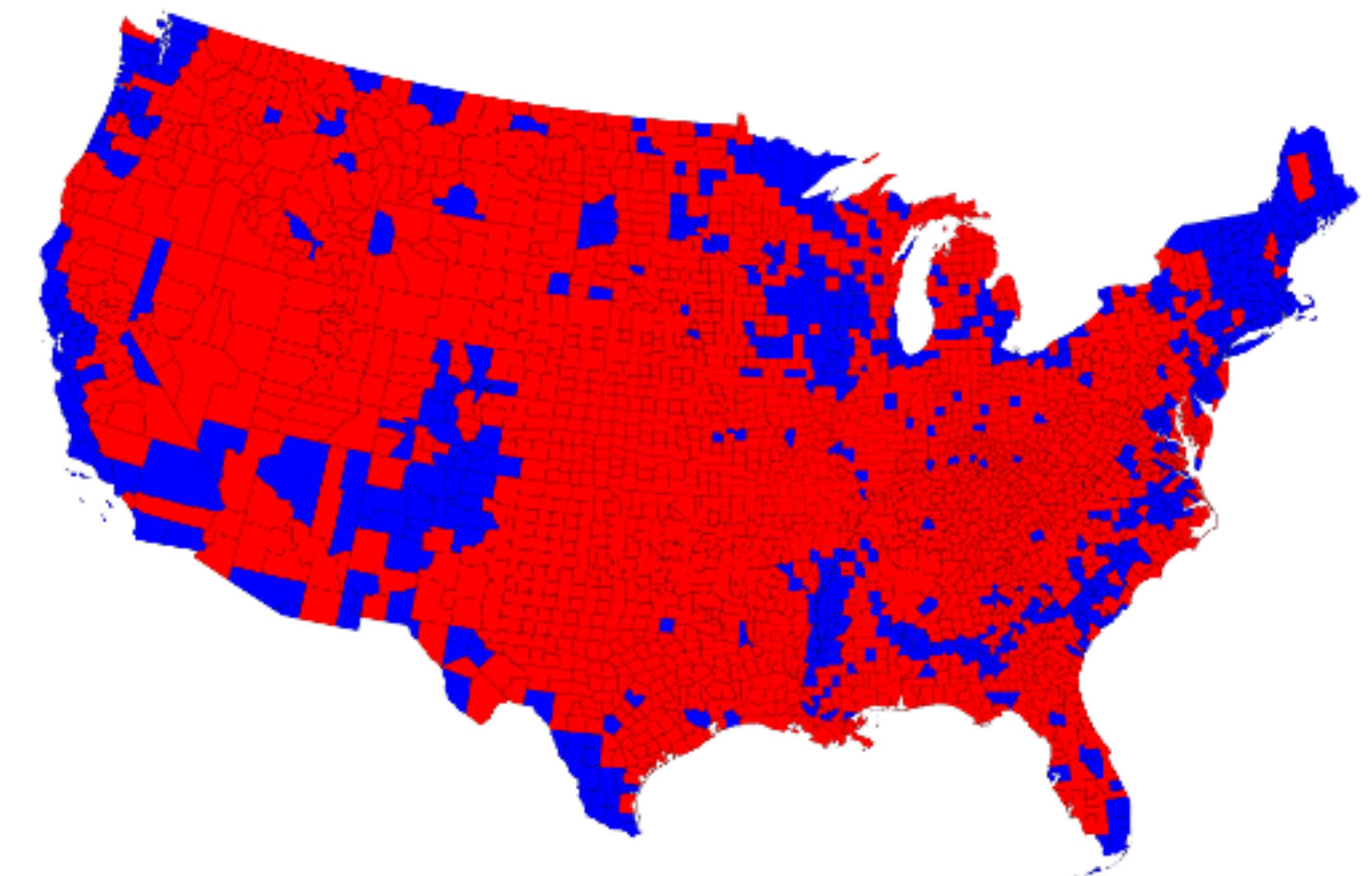
# 2012 Election result by counties



# Any problem?



# How can we mitigate the issue?

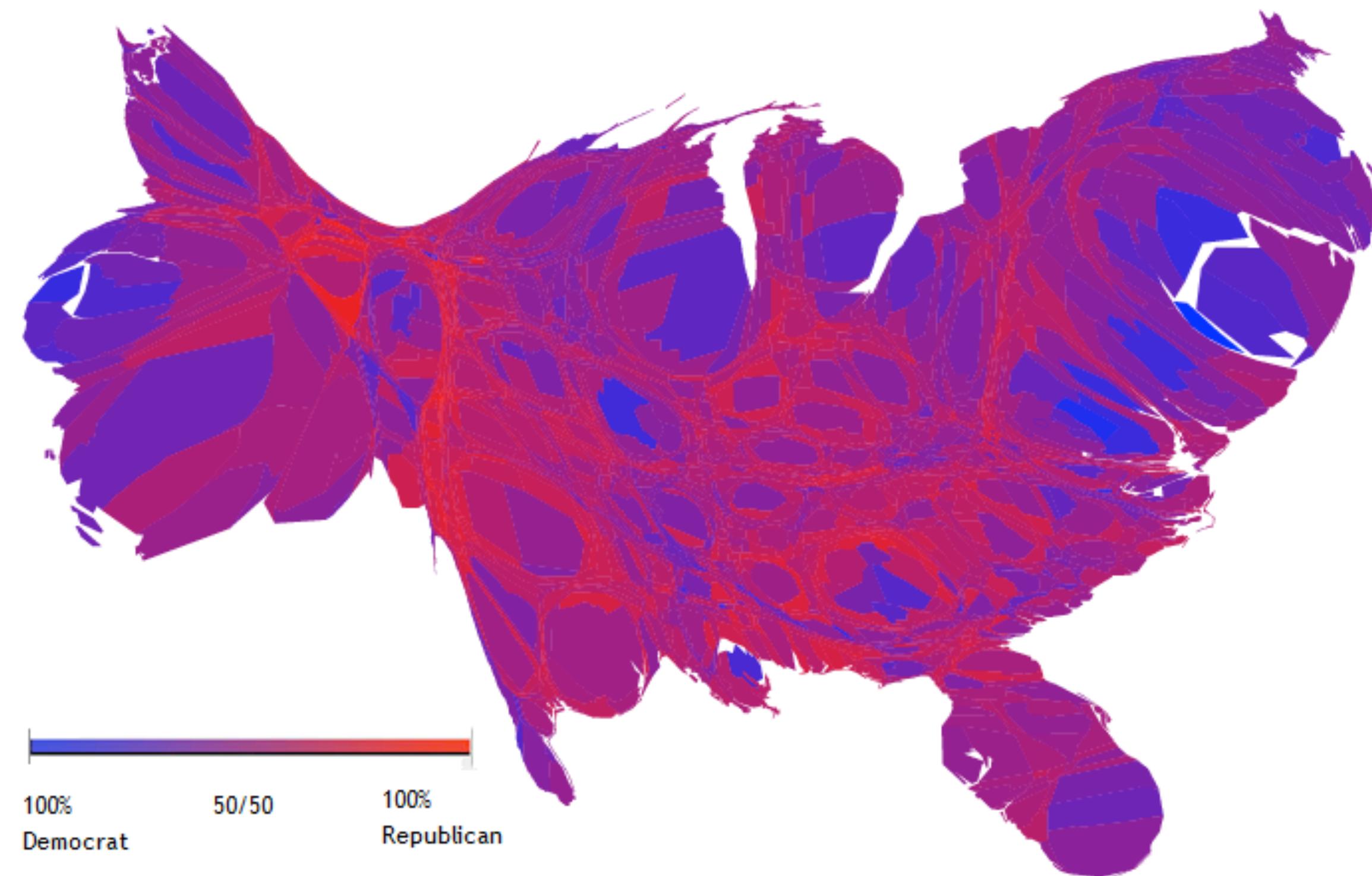


Many solutions!

# Cartogram

“A **cartogram** is a **map** in which some thematic mapping variable – such as travel time, population, or **Gross National Product** – is substituted for land area or distance”

# Area cartogram

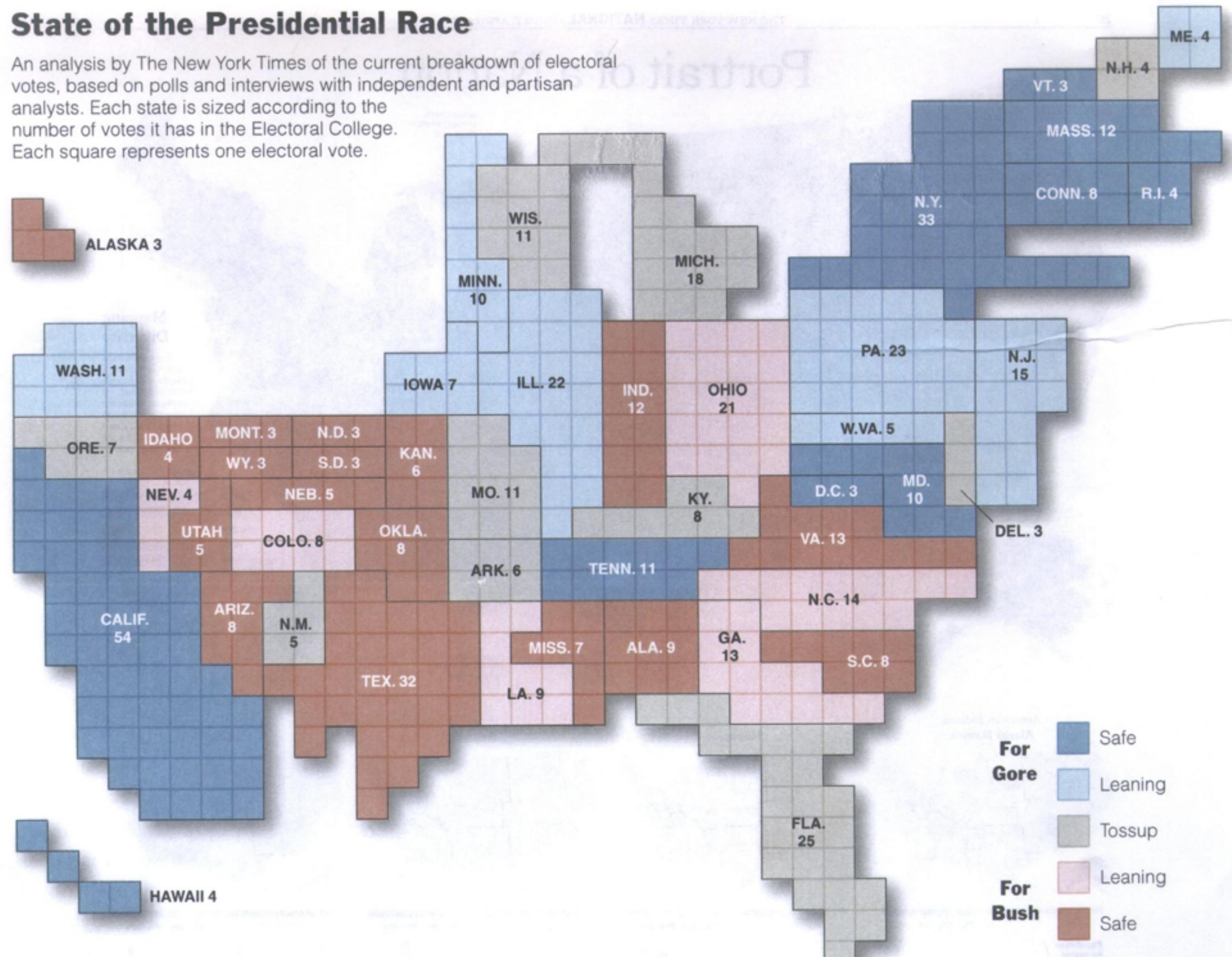


population -> area

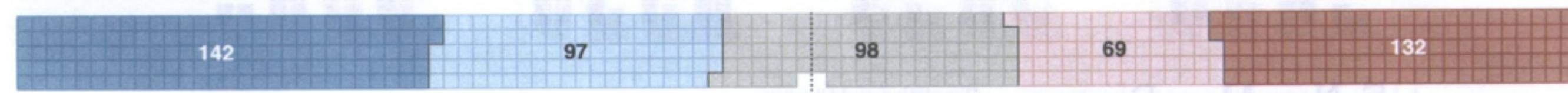
<http://en.wikipedia.org/wiki/File:Cartlinearlarge.png>

## State of the Presidential Race

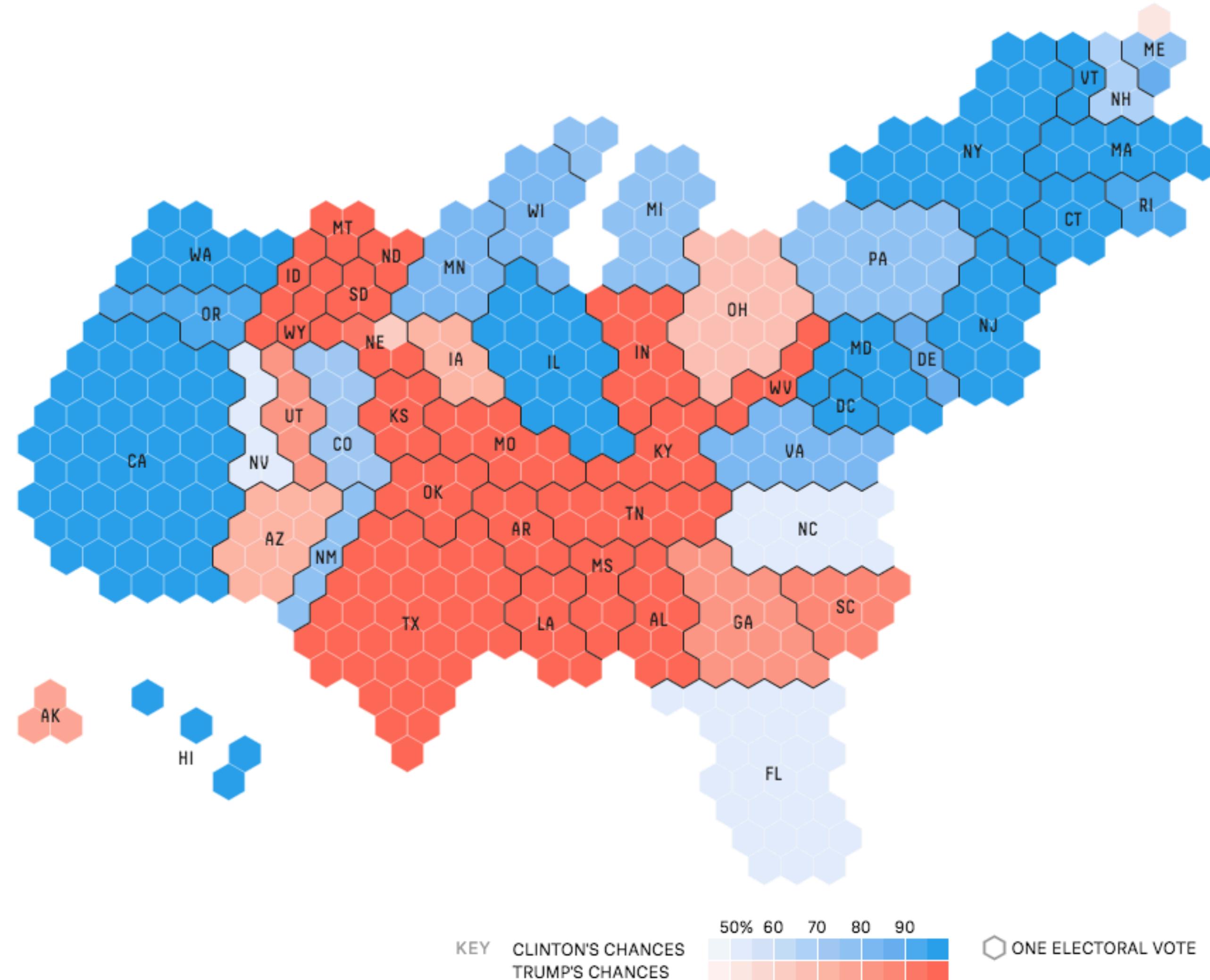
An analysis by The New York Times of the current breakdown of electoral votes, based on polls and interviews with independent and partisan analysts. Each state is sized according to the number of votes it has in the Electoral College. Each square represents one electoral vote.

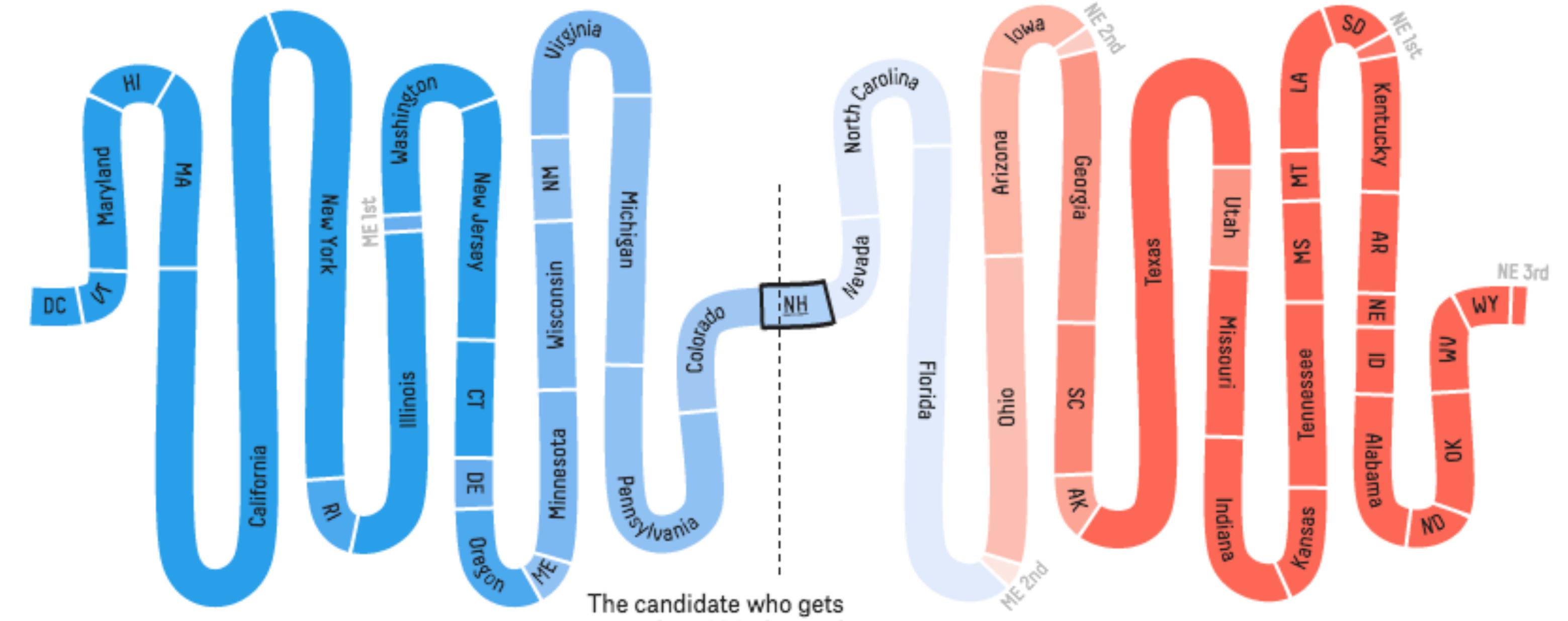


**The electoral vote breakdown** There are a total of 538 electoral votes.



2000 election, nytimes





KEY  
— ONE ELECTORAL VOTE

## Dorling and Dorling-like Cartograms

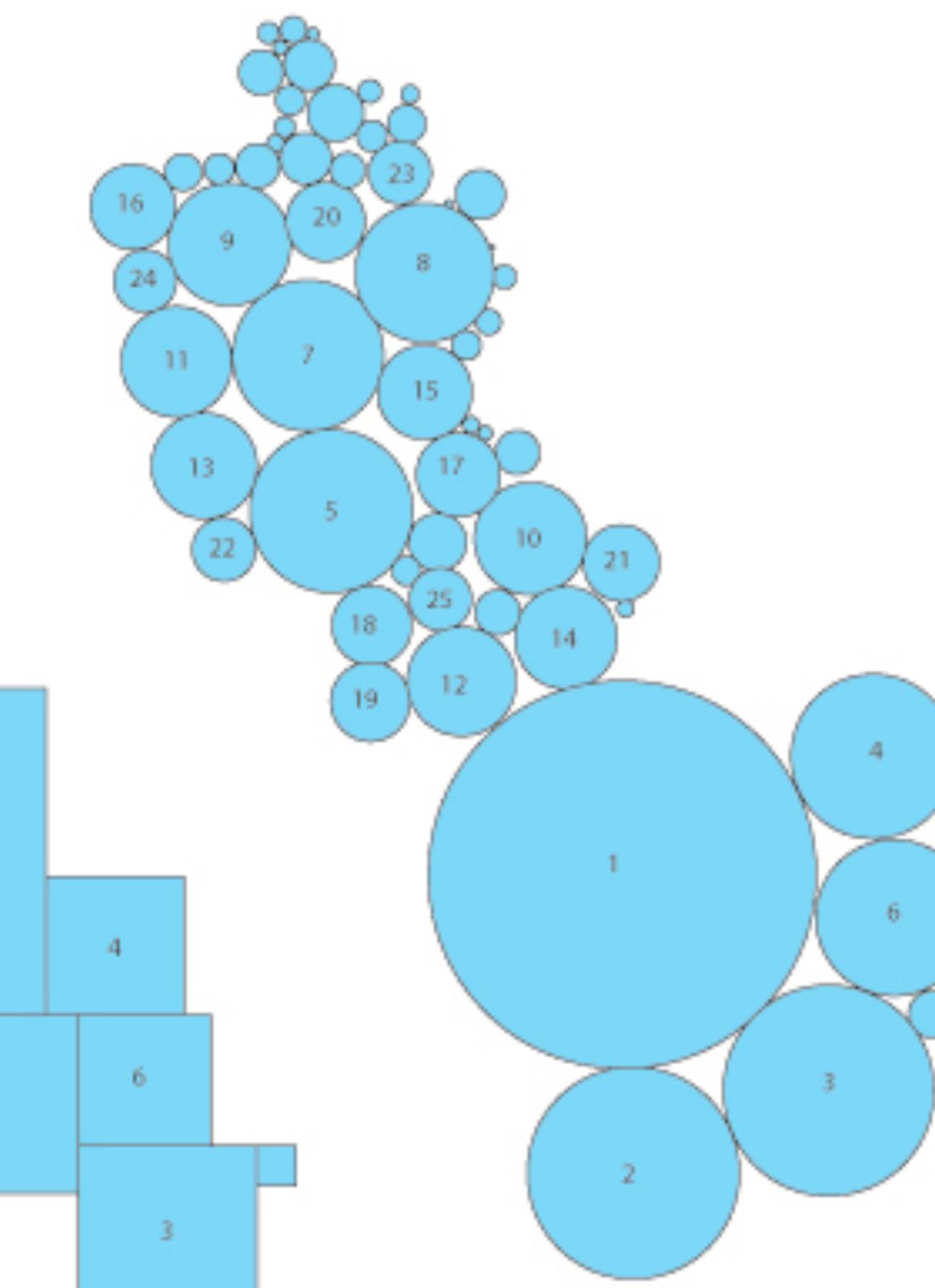
Graduated Symbol Map



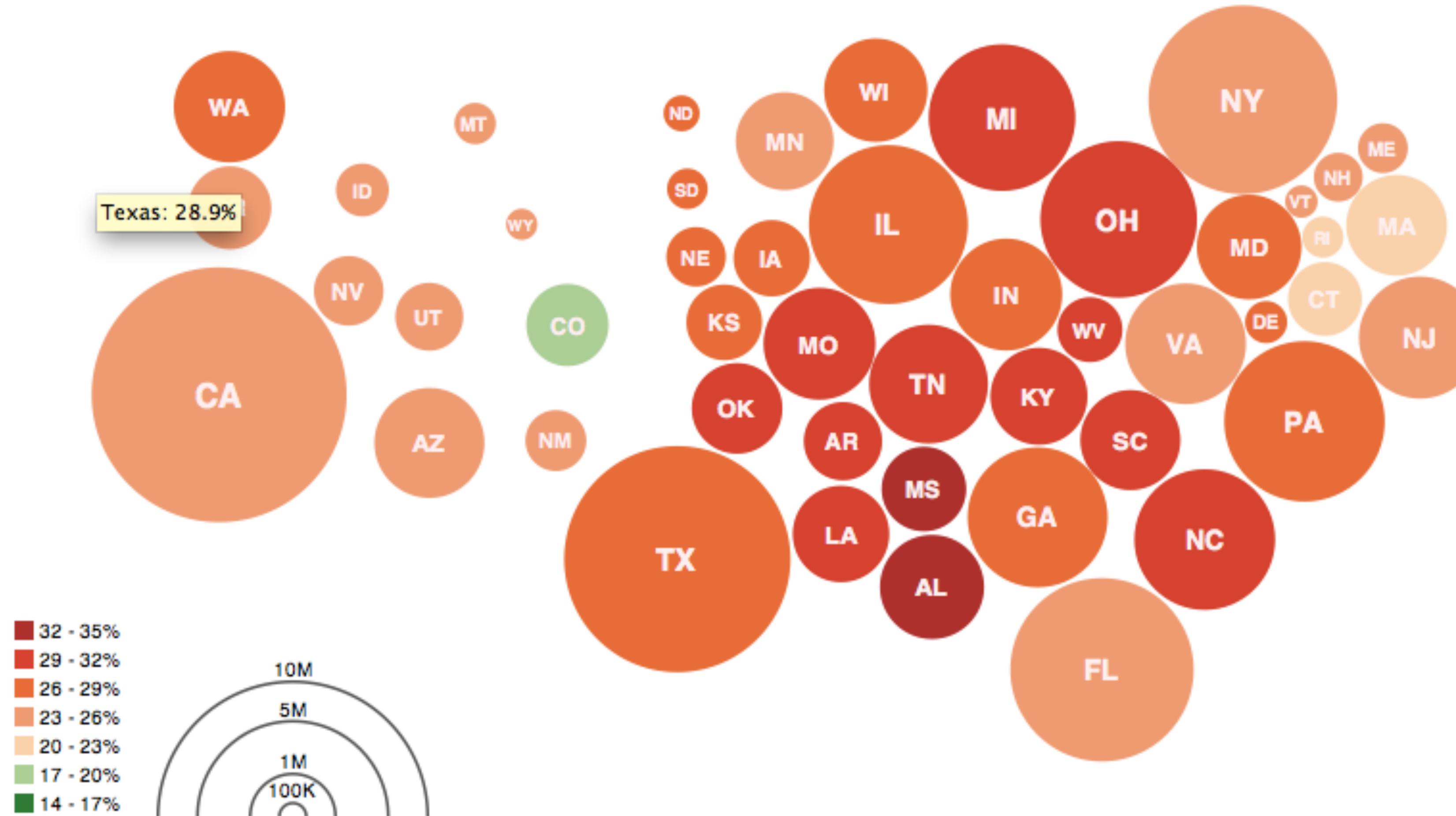
Demers Cartogram



Dorling Cartogram

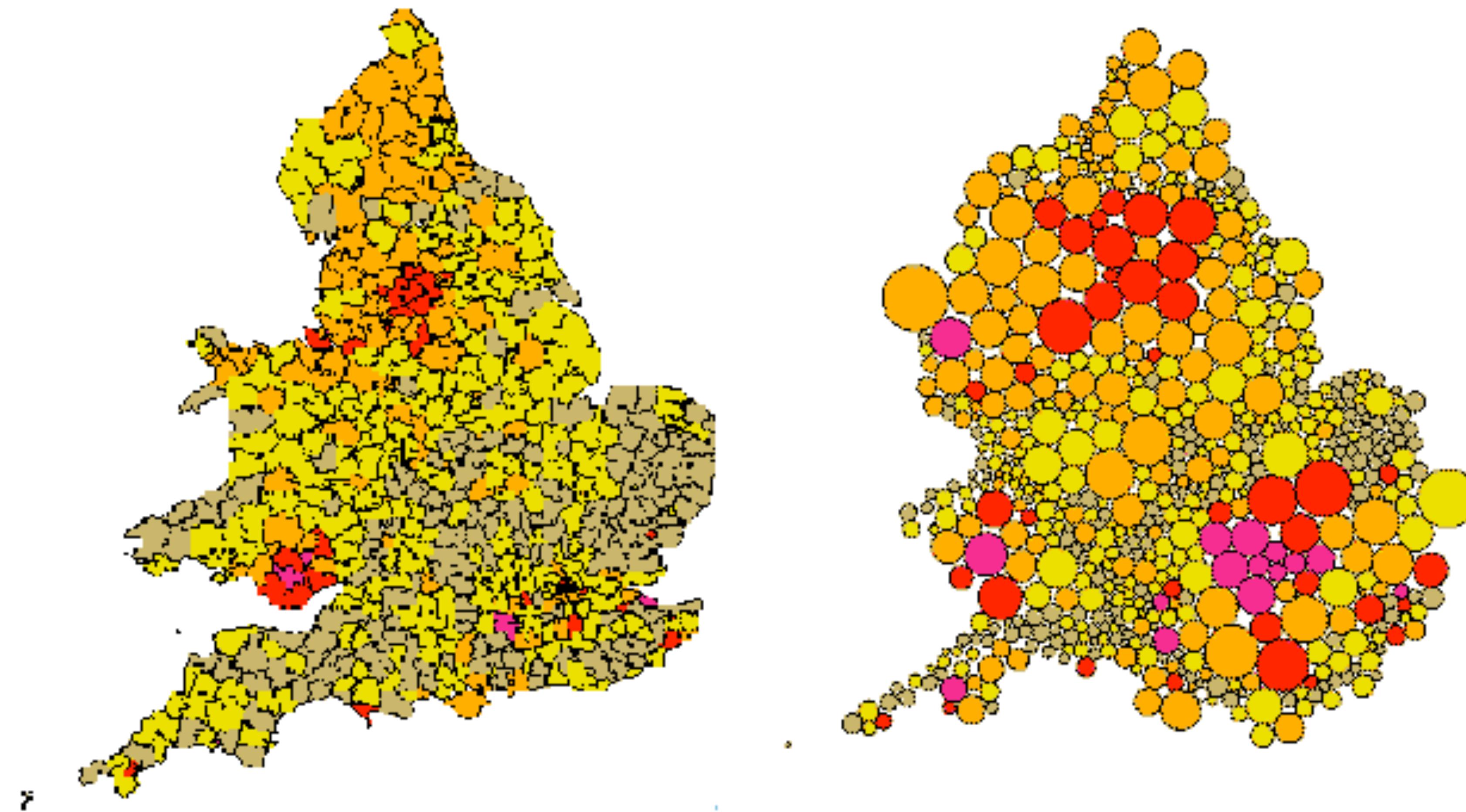


# Dorling cartogram



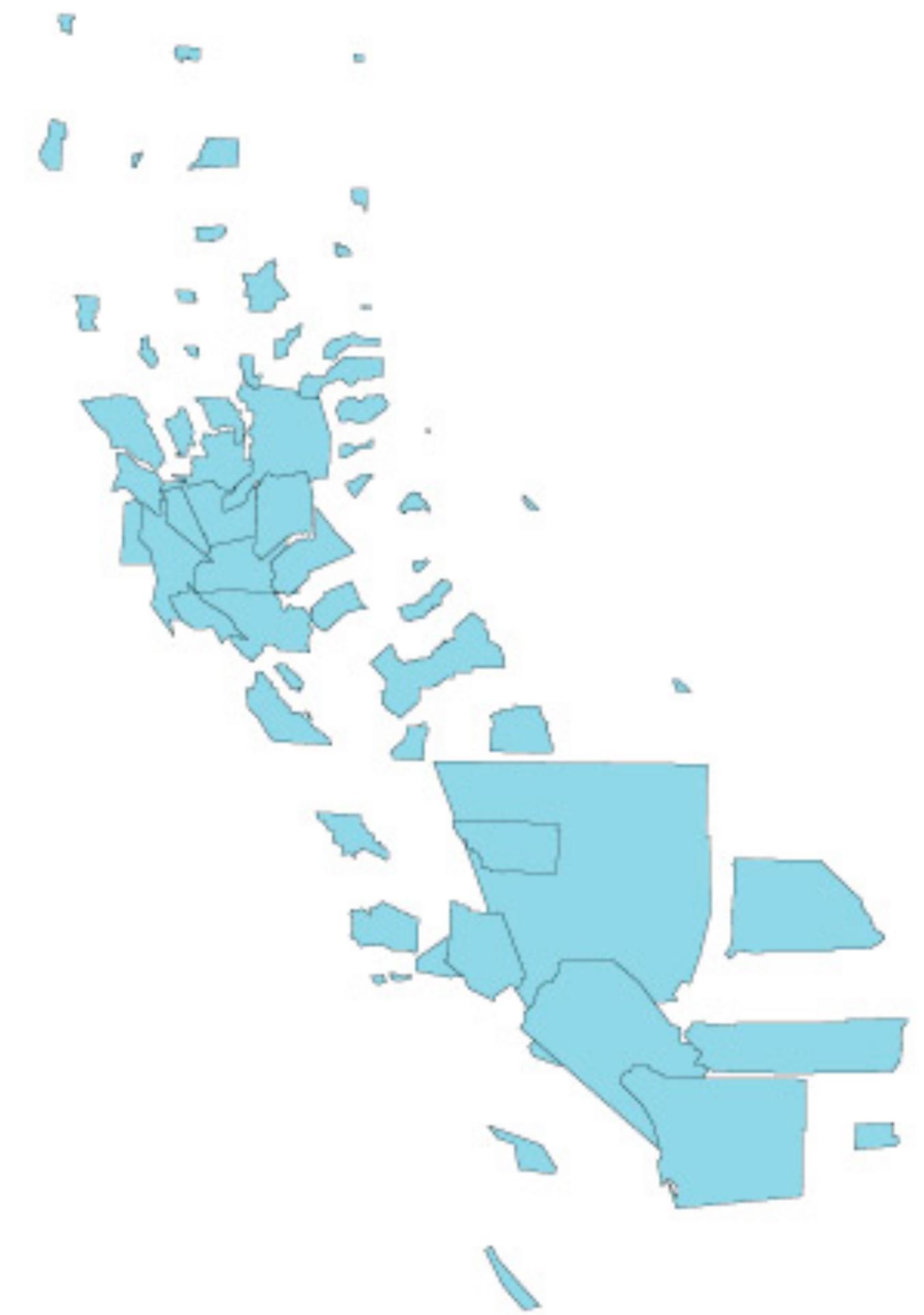
<http://mbostock.github.com/protovis/ex/cartogram.html>

COMPARISON OF A TRADITIONAL MAP AND CARTOGRAM REPRESENTATIONS  
OF THE PERCENTAGE OF THE MALE POPULATION OF WORKING AGE IN 1891



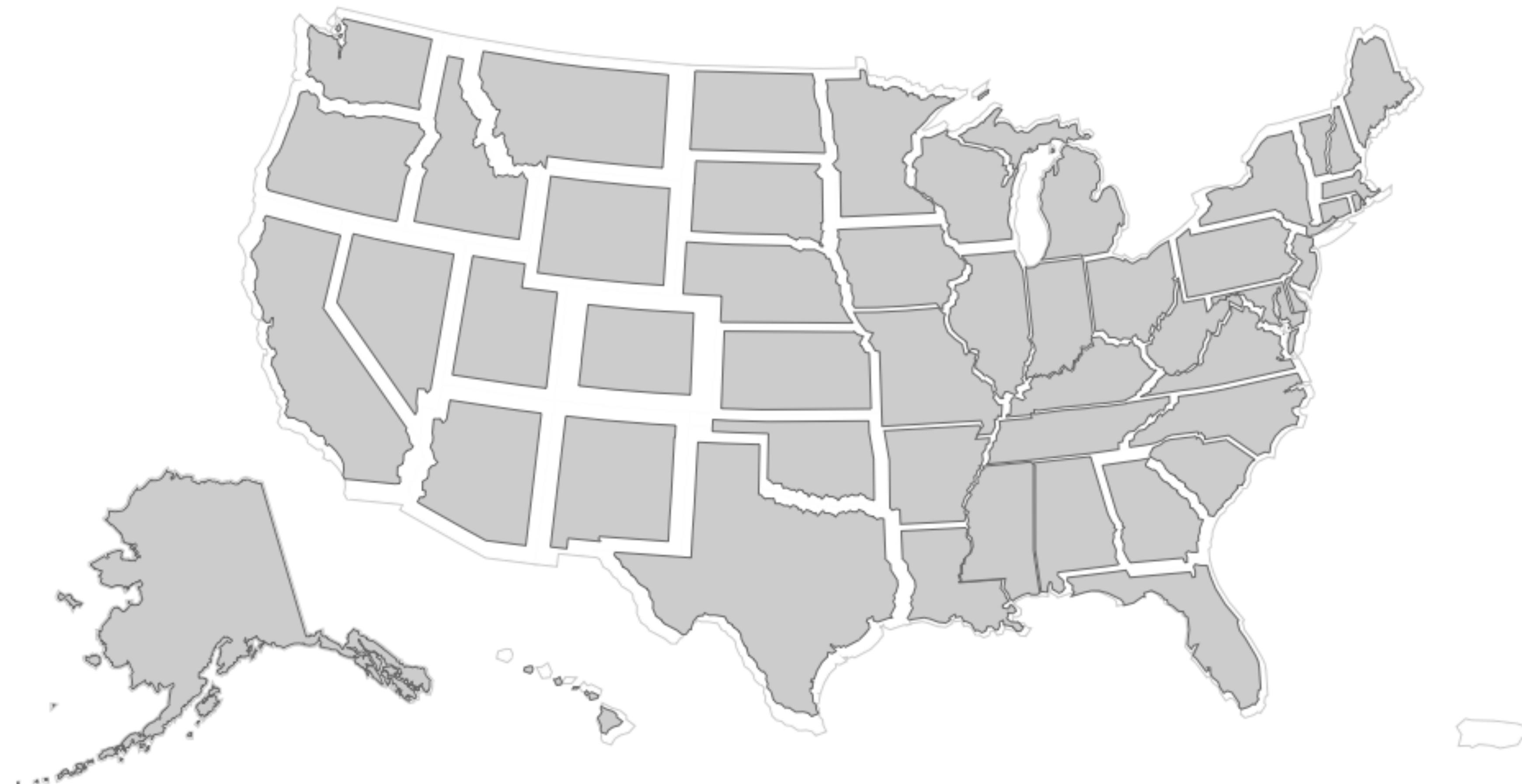
## Non-Contiguous Cartograms

Overlapping



Non-Overlapping



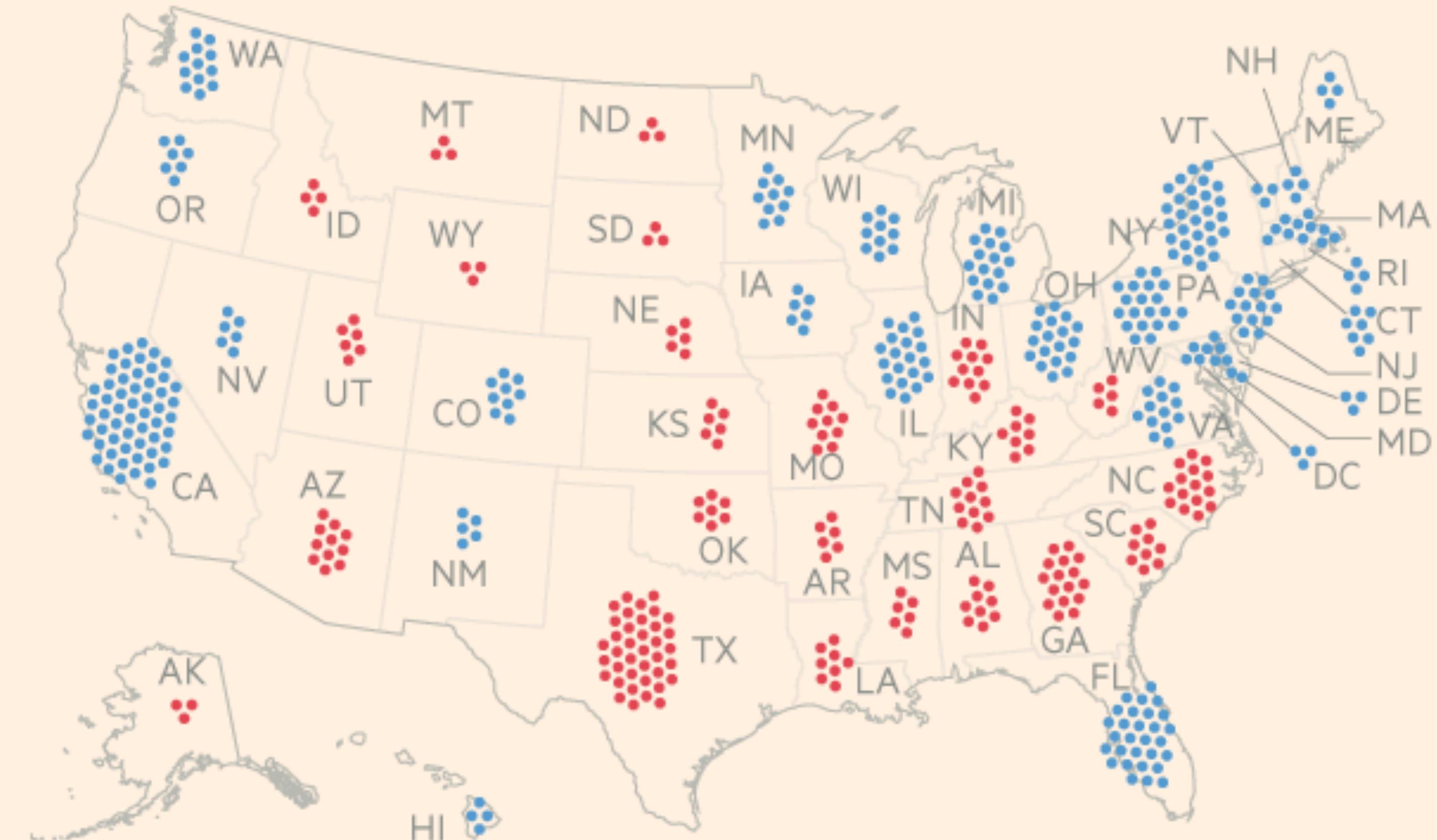


<http://mbostock.github.com/d3/ex/cartogram.html>

## Electoral college votes by state

2012 presidential election results\*

Obama      Romney



\* Each state's vote allocation remains the same in 2016 as in 2012.

FT

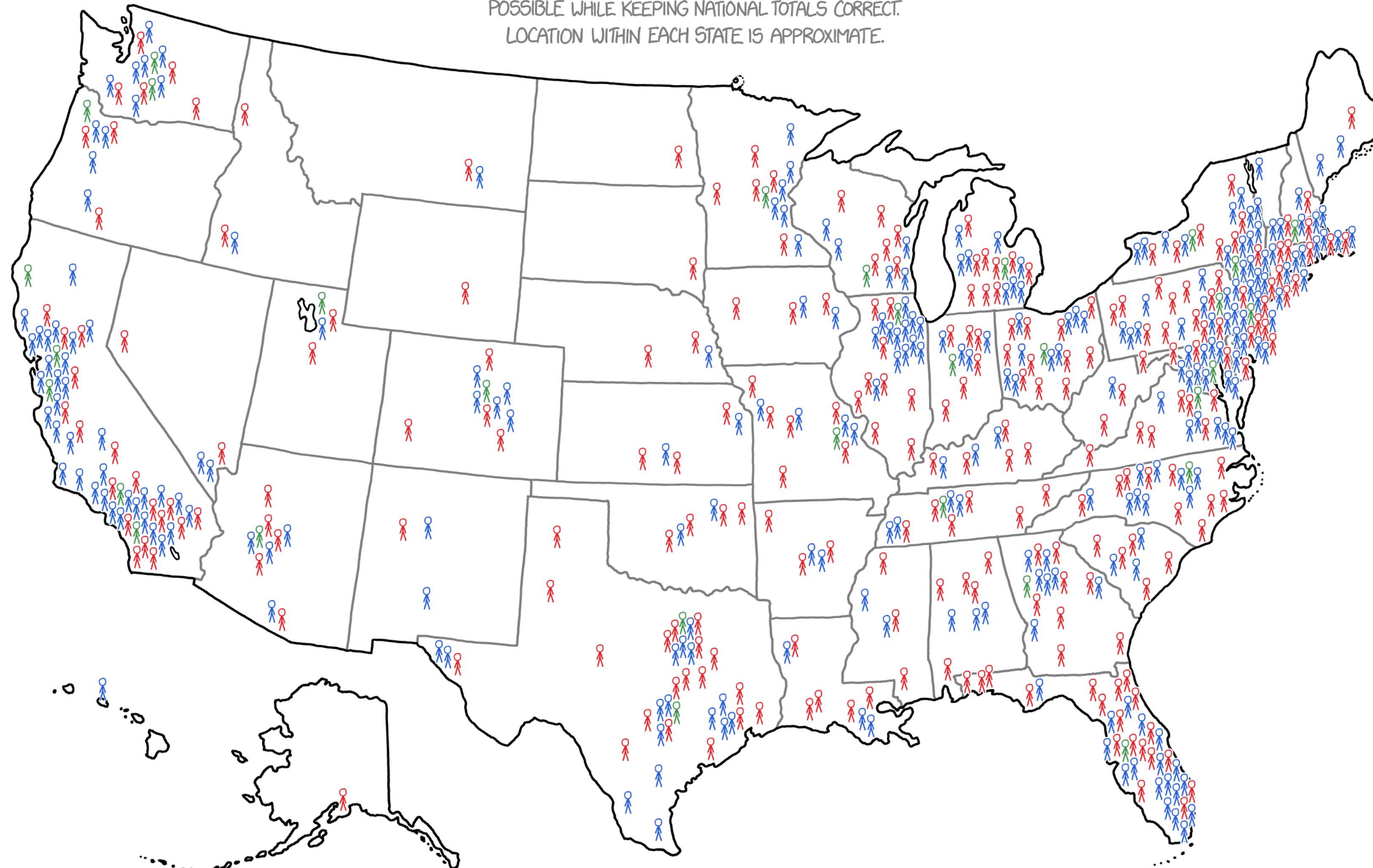
# 2016 ELECTION MAP

EACH FIGURE REPRESENTS 250,000 VOTES

TRUMP CLINTON OTHER

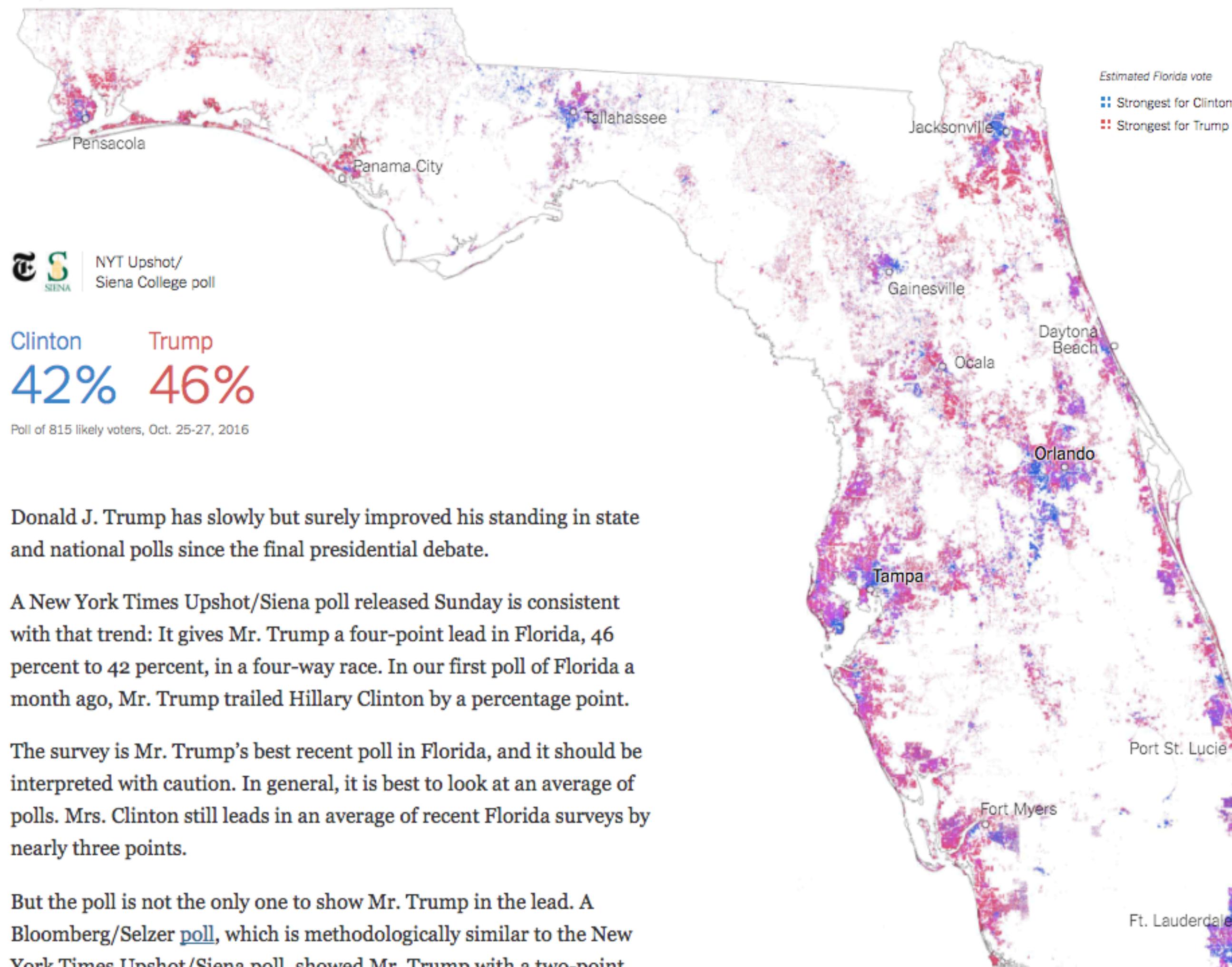
VOTES ARE DISTRIBUTED BY STATE AS ACCURATELY AS  
POSSIBLE WHILE KEEPING NATIONAL TOTALS CORRECT.

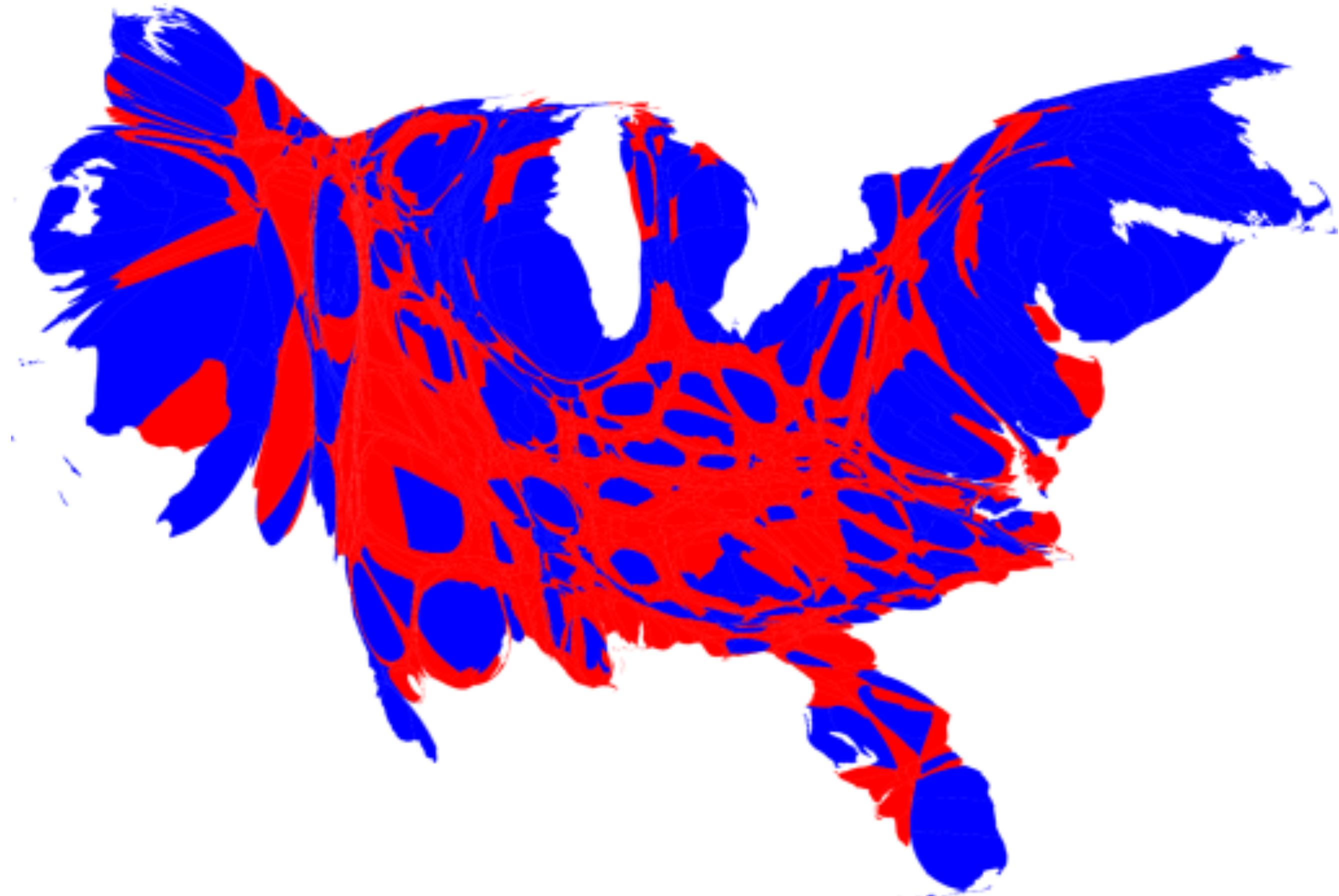
LOCATION WITHIN EACH STATE IS APPROXIMATE.



# Latest Upshot Poll Shows Trump With a Lead in Florida

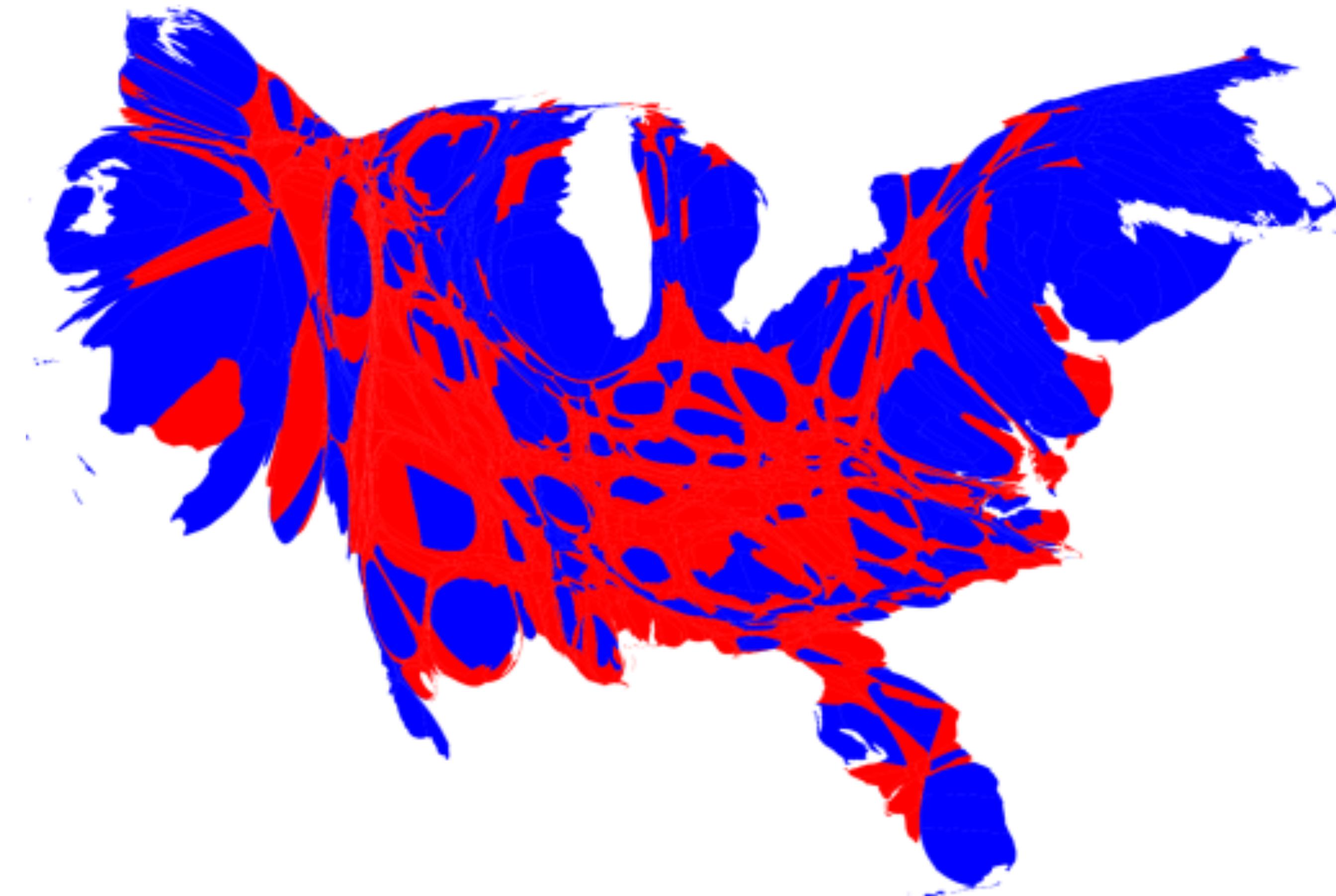
By NATE COHN OCT. 30, 2016



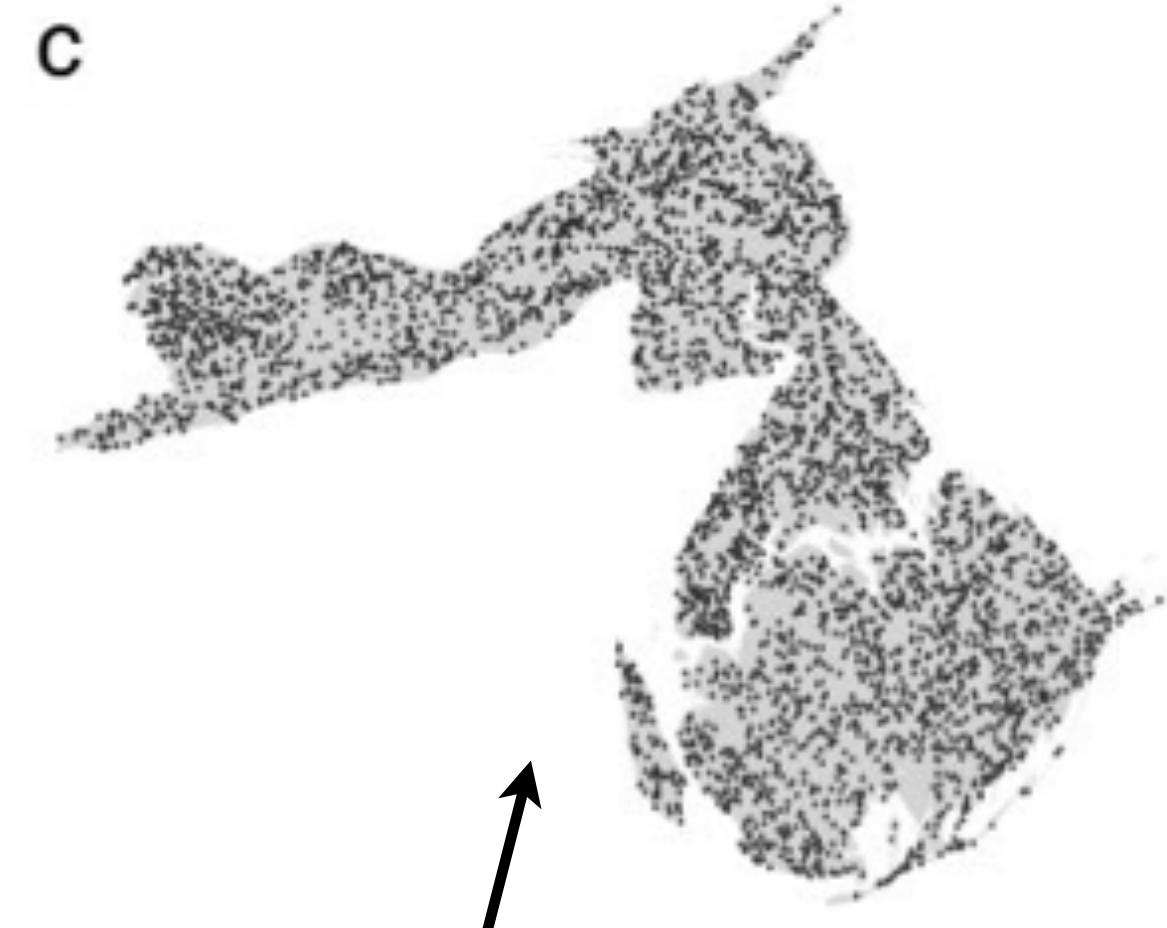
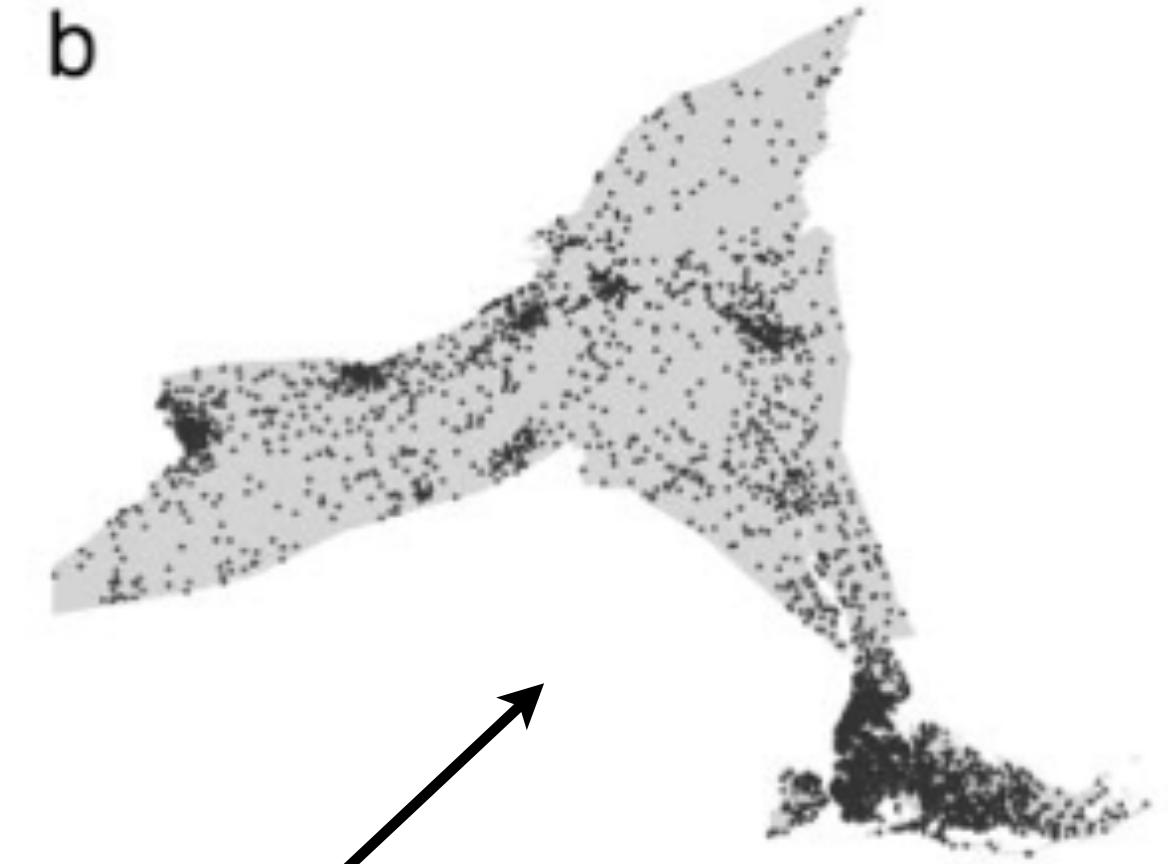
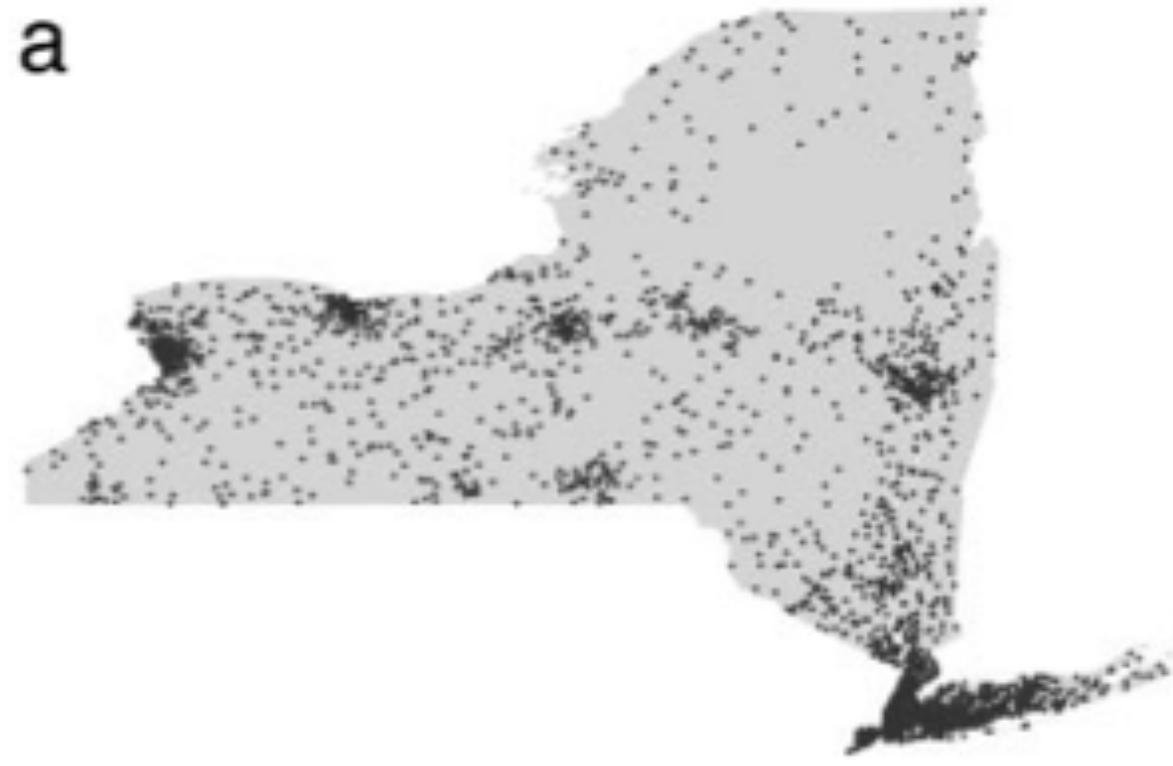


Michael T. Gastner and M. E. J. Newman, Diffusion-based method  
for producing density-equalizing maps  
<http://www.pnas.org/content/101/20/7499.abstract>

# Density-equalizing cartogram



Michael T. Gastner and M. E. J. Newman, Diffusion-based method  
for producing density-equalizing maps  
<http://www.pnas.org/content/101/20/7499.abstract>



Cartogram with  
**coarse-grained**  
density data

Cartogram with  
**fine-grained**  
density data

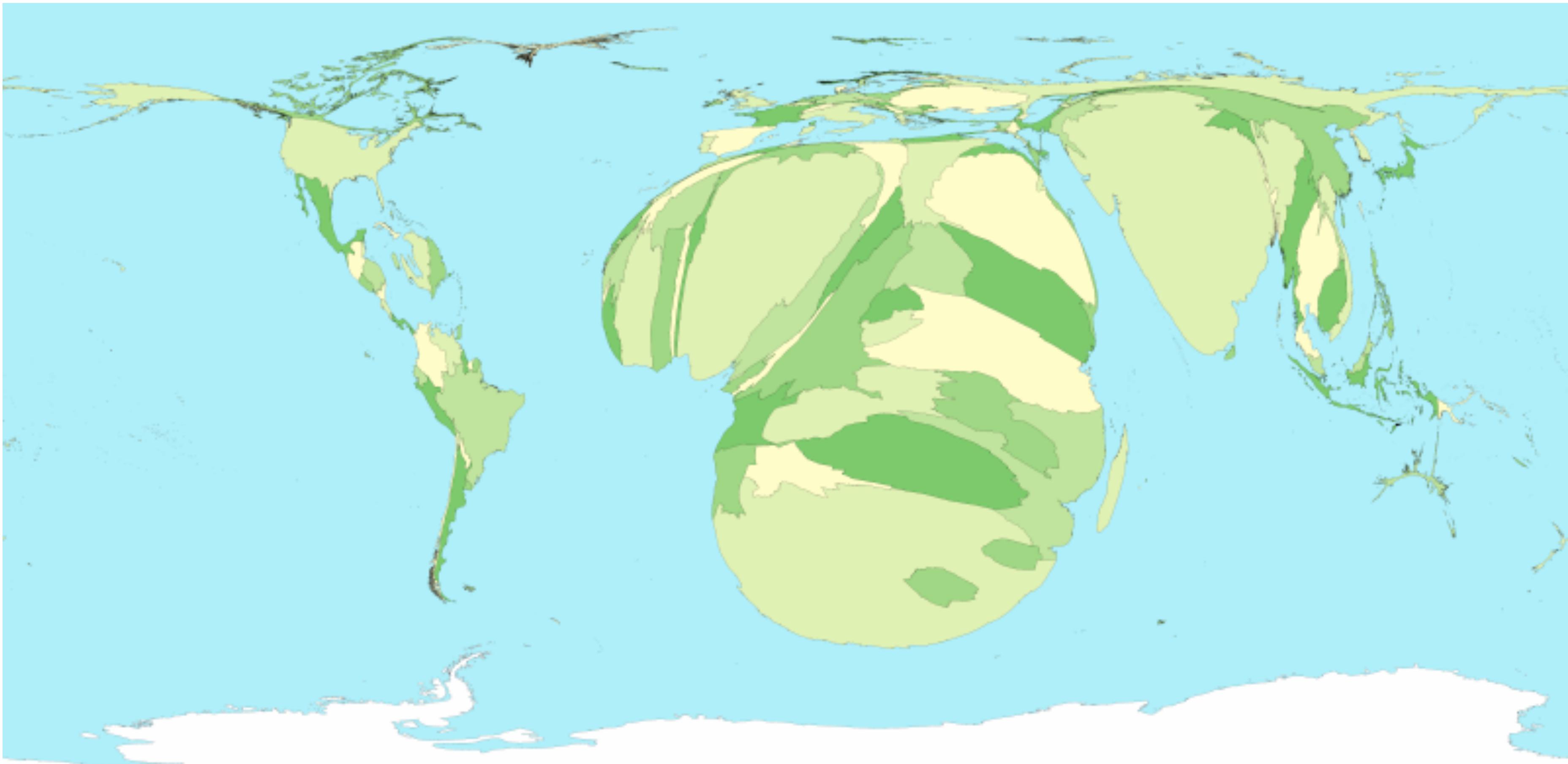
C

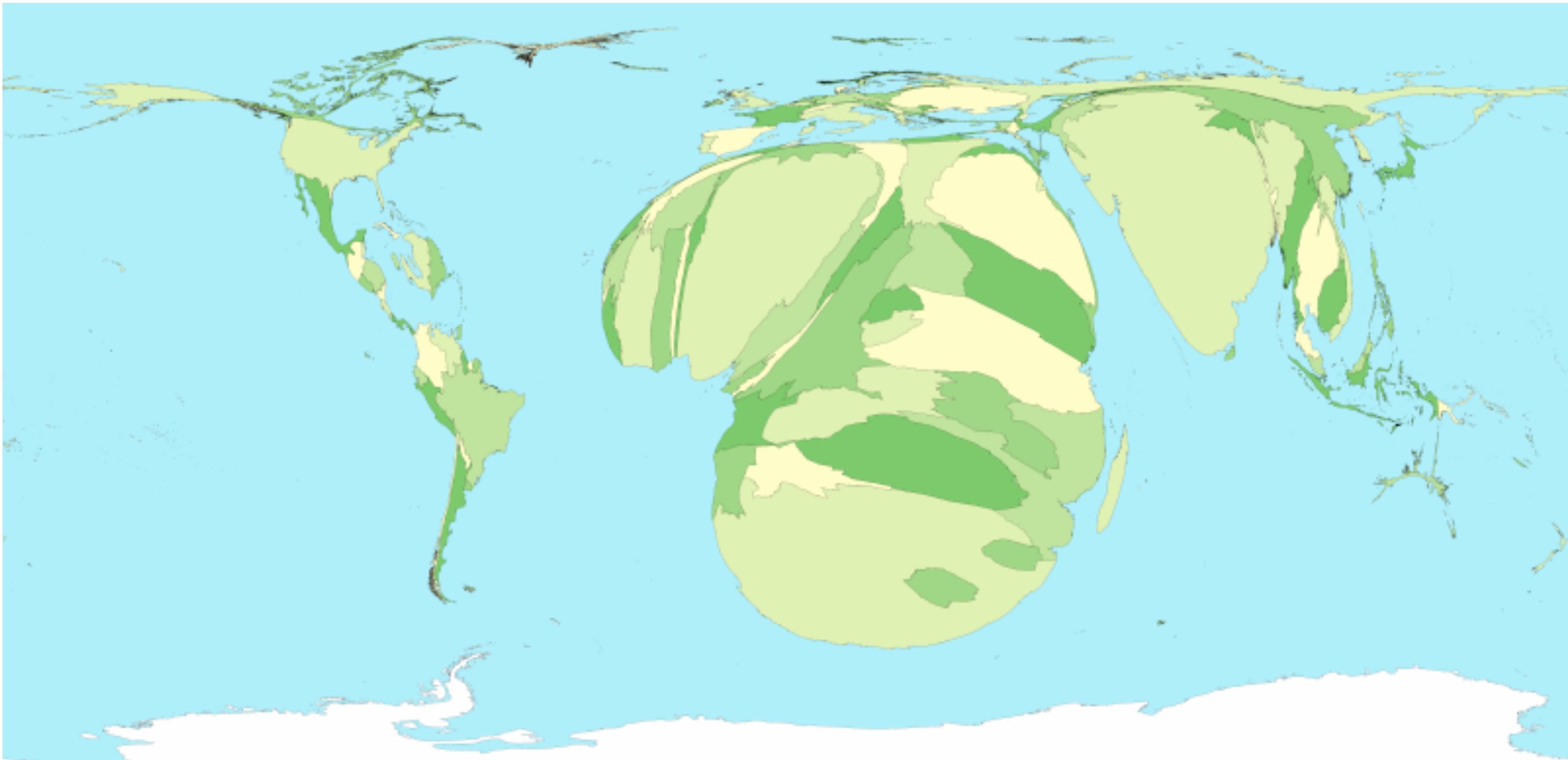




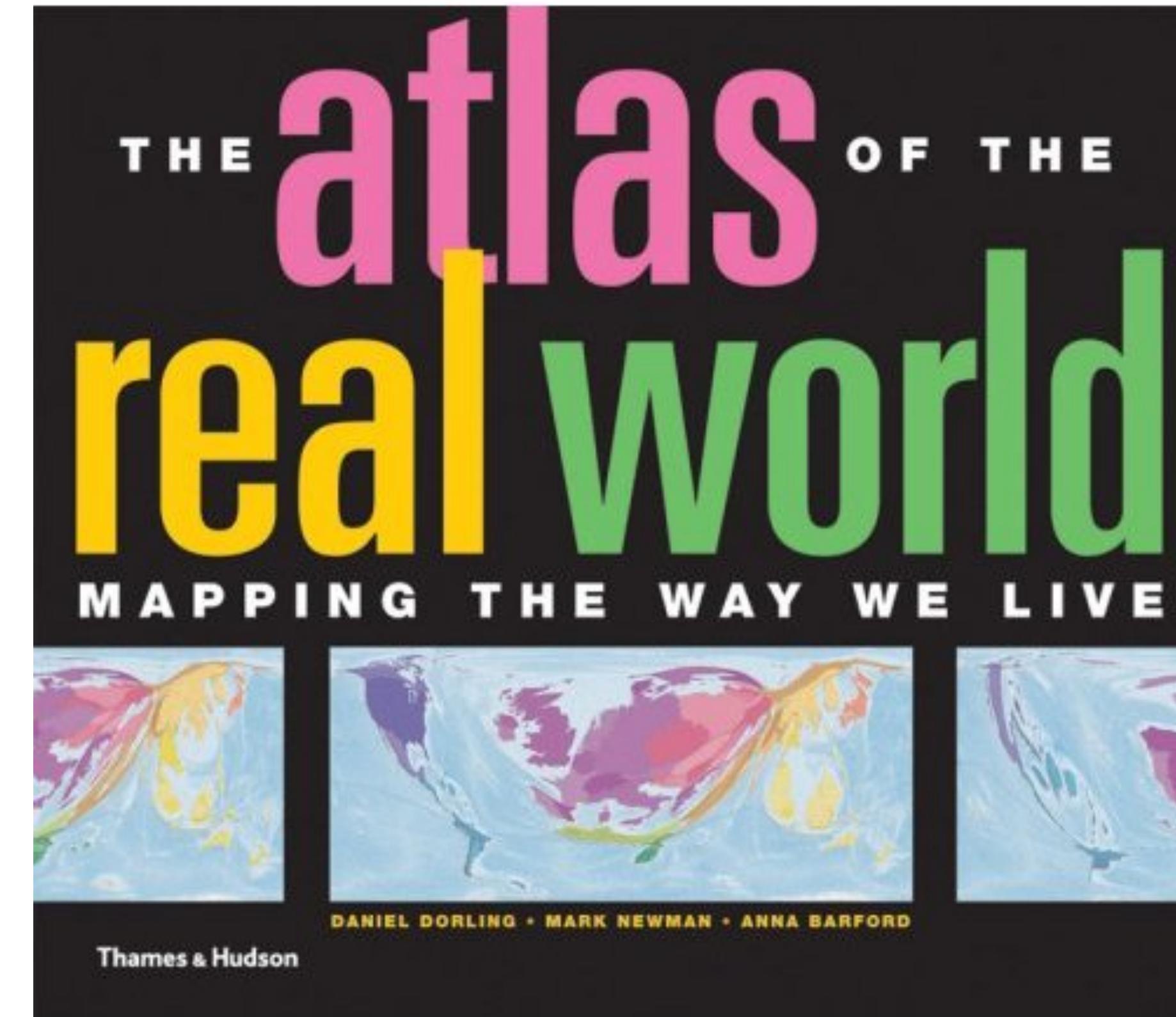


# Population



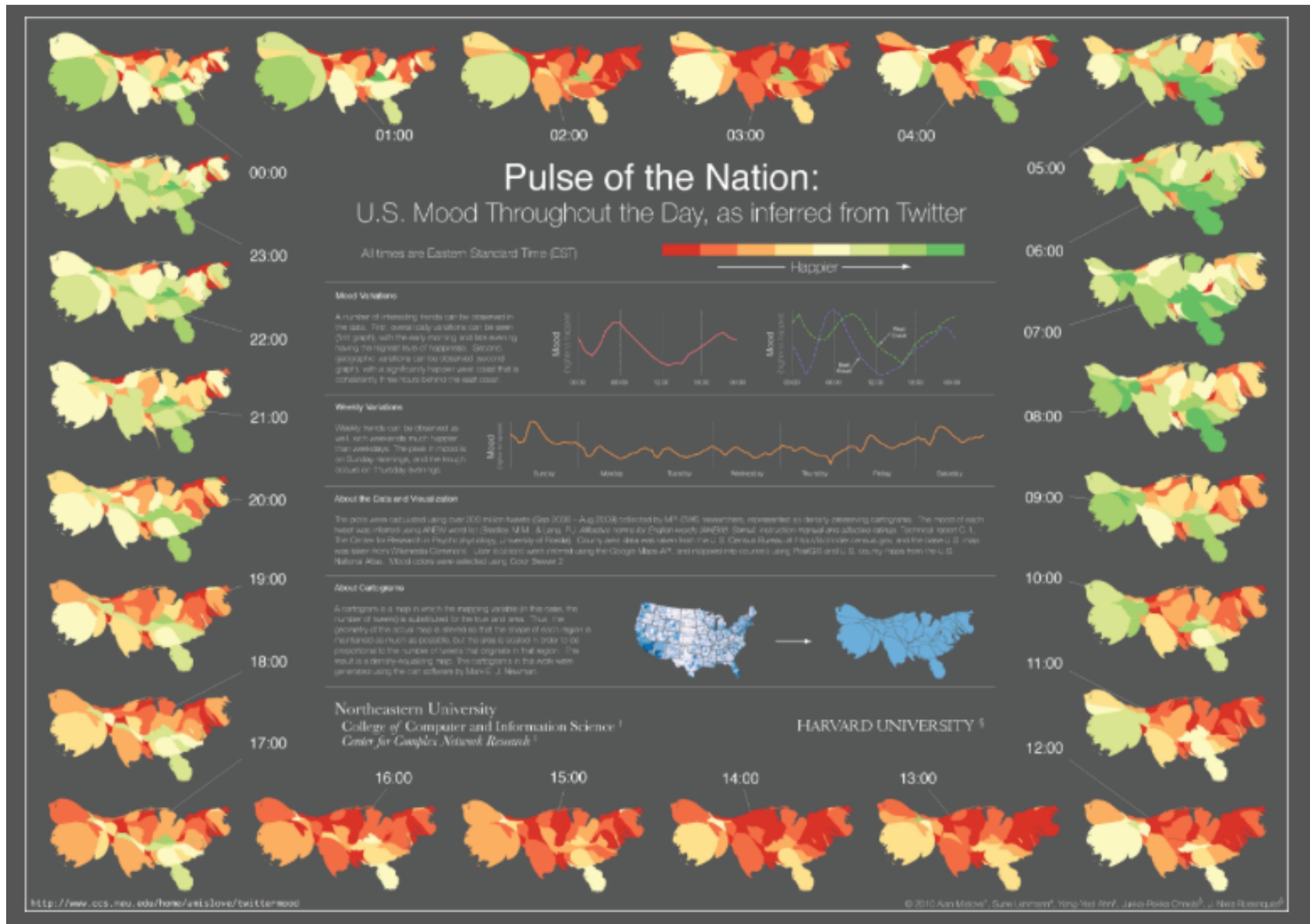


HIV



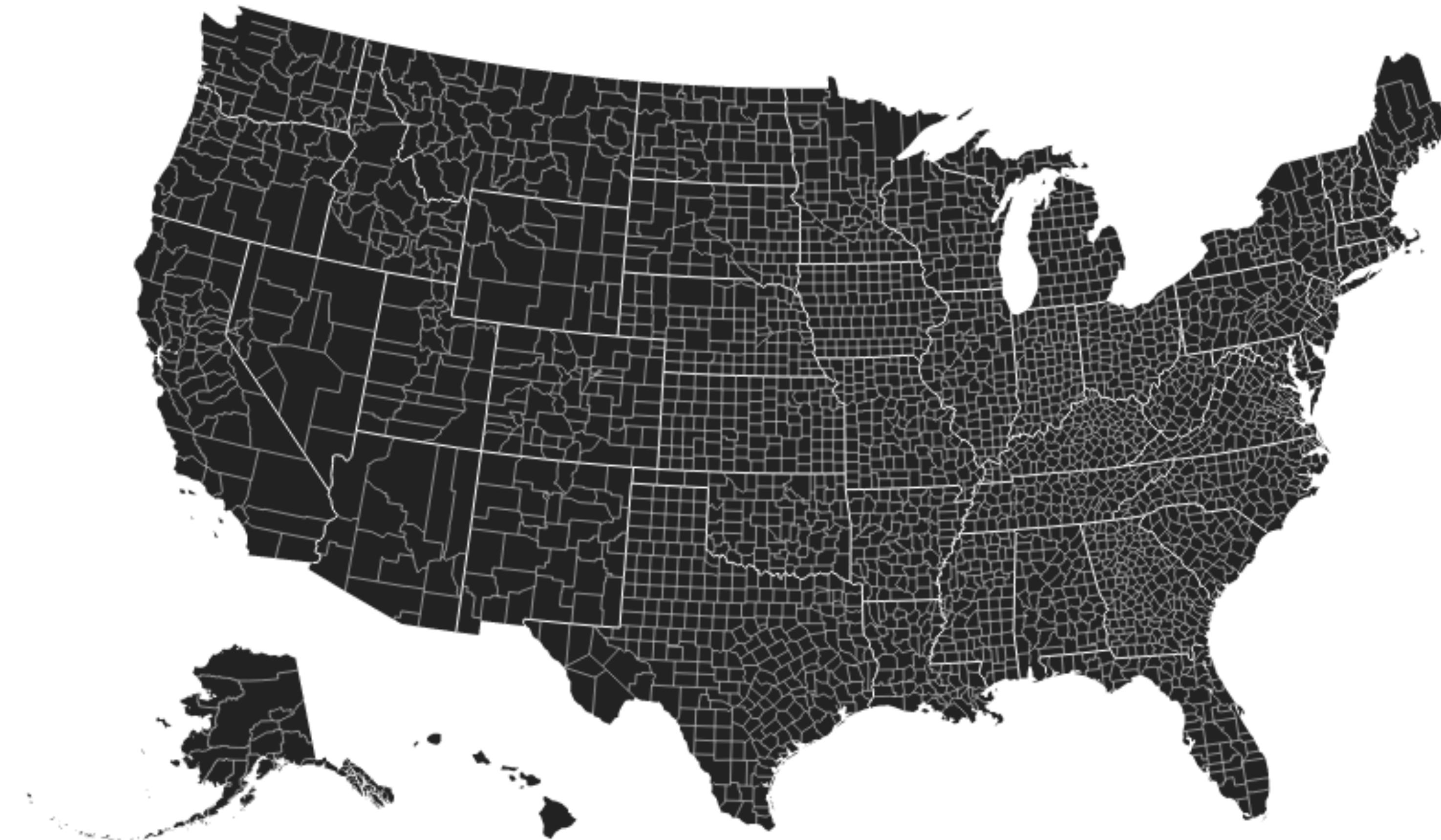
<http://www-personal.umich.edu/~mejn/cartograms/>

<https://www.google.com/search?q=The+Atlas+of+the+Real+World>



<http://www.youtube.com/watch?v=ujcrJZRSGkg>

# How do you practically store & draw a map like this?



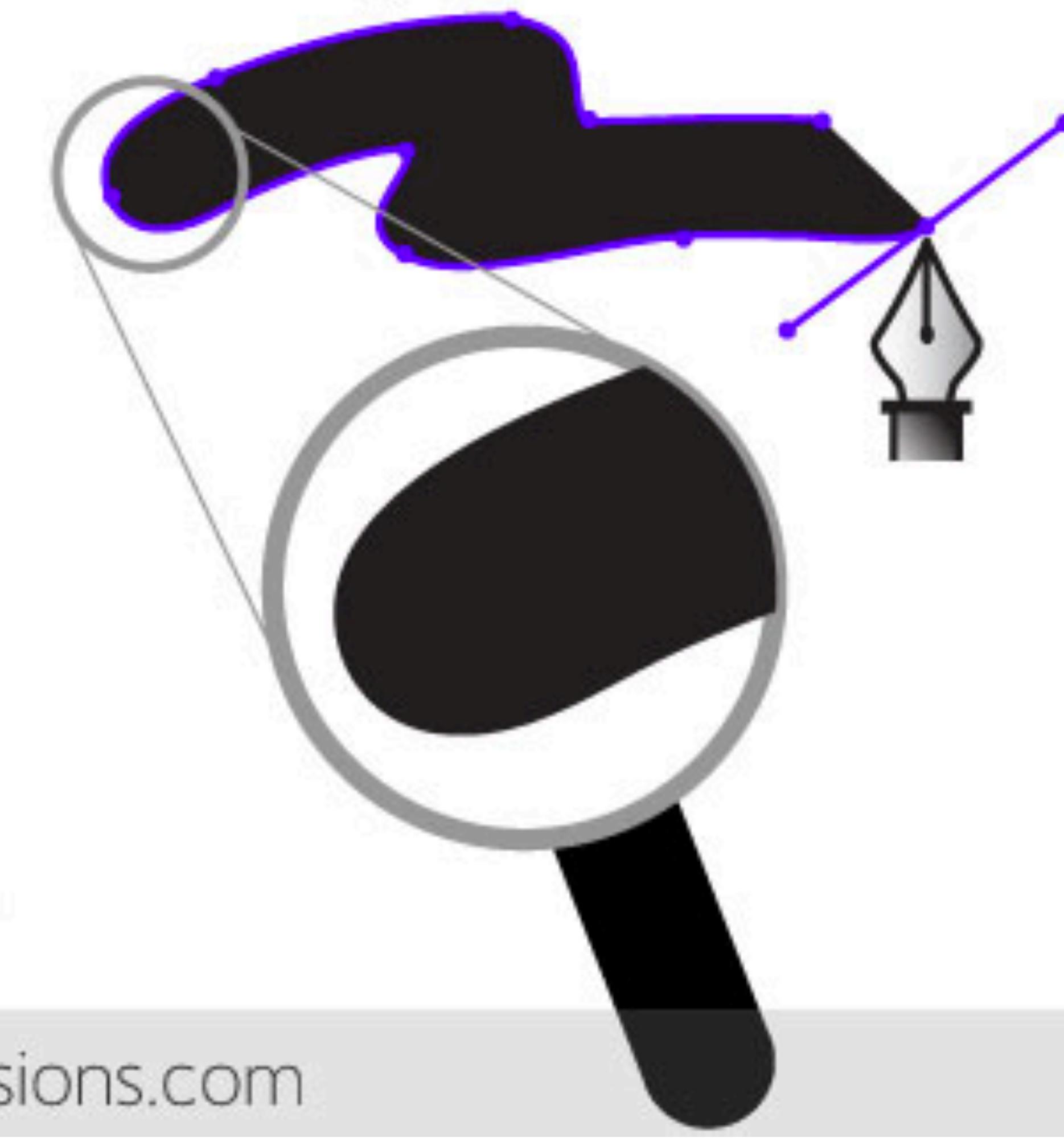
# Raster vs. Vector

# Image vs. Drawing Instruction

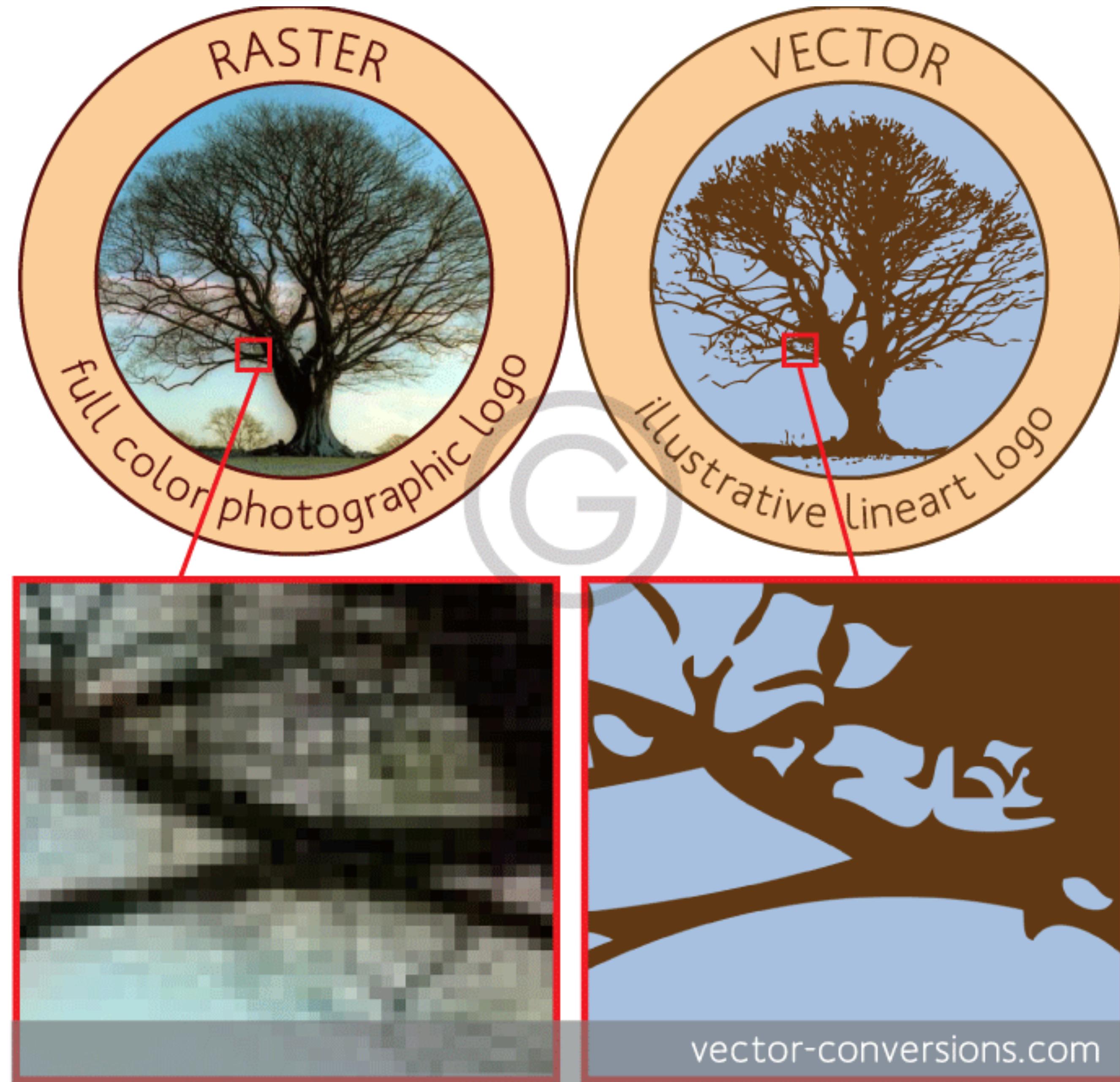
painting with pixels



drawing with vectors



[vector-conversions.com](http://vector-conversions.com)



# TopoJSON

<https://github.com/mbostock/topojson/wiki>

```
{  
  "type": "Topology",  
  "transform": {  
    "scale": [0.036003600360036005, 0.017361589674592462],  
    "translate": [-180, -89.99892578124998]  
  },  
  "objects": {  
    "aruba": {  
      "type": "Polygon",  
      "arcs": [[0]],  
      "id": 533  
    }  
  },  
  "arcs": [  
    [[3058, 5901], [0, -2], [-2, 1], [-1, 3], [-2, 3], [0, 3], [1, 1], [1, -3], [2, -5], [1, -1]]  
  ]  
}
```

Origine\_Point  
"coordinates": [0,0]

Under\_Point  
"coordinates": [0,-1]

Under\_LineString  
"arcs": [3]

Left\_Polygon  
"arcs": [[0,1]]

Right\_Polygon  
"arcs": [[2,-1]]

"arcs":  
[[1,2],[0,-2]],  
[[1,0],[-1,0],[0,2][1,0]],  
[[1,2],[1,0],[0,-2],[-1,0]],  
[[0,-1],[2,0]]

1

