

Welcome (Again!) to **MATH 4100/COMP 5360– Introduction to Data Science**

(Some) Principles of Data Visualization, cont'd

February 28, 2023

*Based on prior lectures from
Alex Lex and Bei Wang Phillips
+ others where noted*



Announcements / Reminders / Schedule

HW 6 – Due Friday!

Thursday: Classification, Decision Trees!

Friday: Team formation check-in!

...and then Spring Break!

...and then... project proposal!

Team Formation

Please fill out Google Form by Friday:

https://docs.google.com/forms/d/e/1FAIpQLSc0foXMyjCu3haQHHEVjyN2dBnar9dswjkVLCinh0luHBnZEA/viewform?usp=sf_link

Really quick! Just a topic + team members, emails, UID, and programming experience.

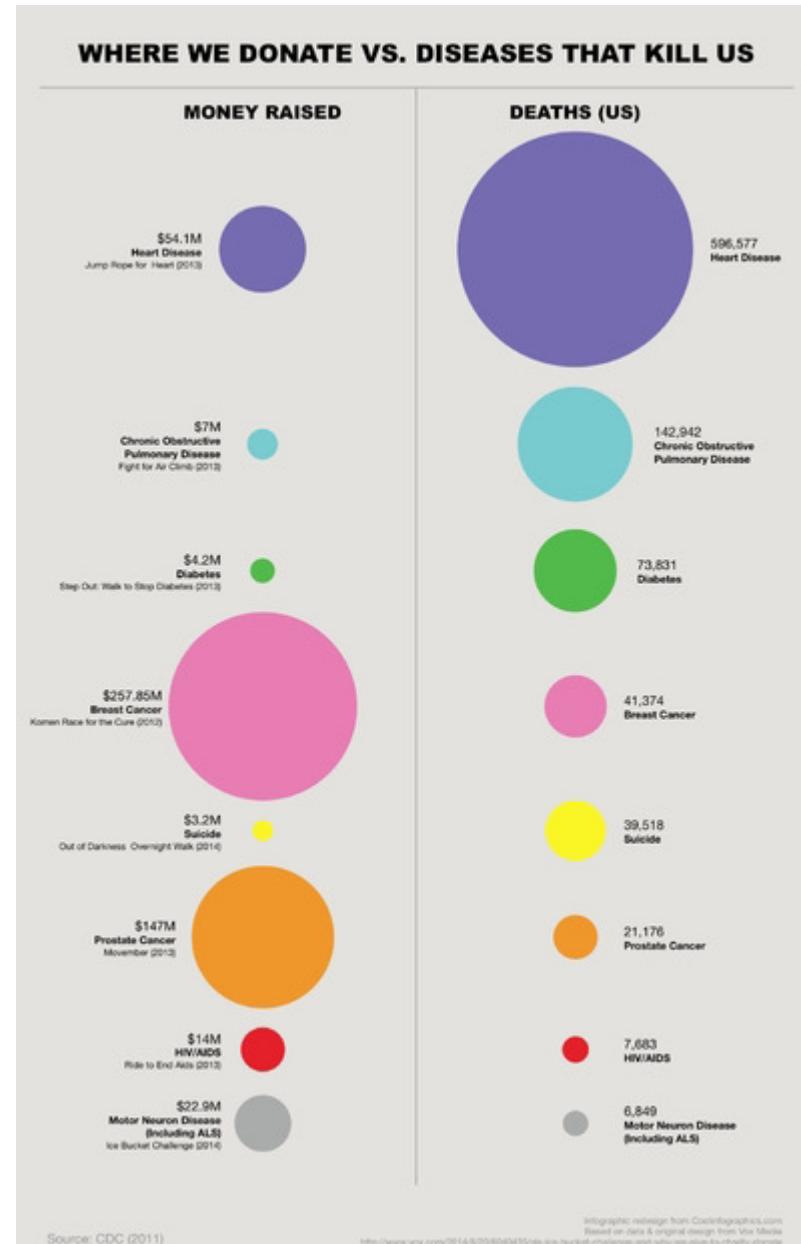
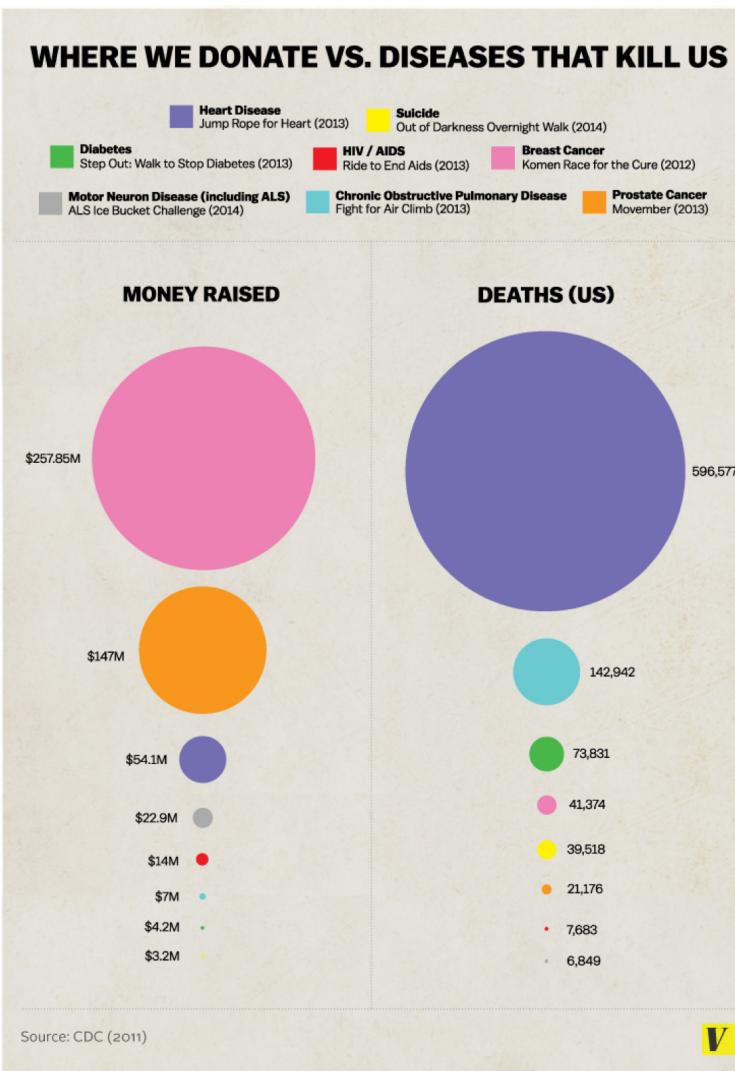
Tuesday Office Hours – Changing after Spring Break

Today 6:30 Over Zoom

AFTER Spring Break, 5 - 6PM (Right after Class) in-person

- We'll try them here until we get in trouble ☺
- If the classroom doesn't work, WEB 4660

Recall -- Circles: Encode by Area not Radius



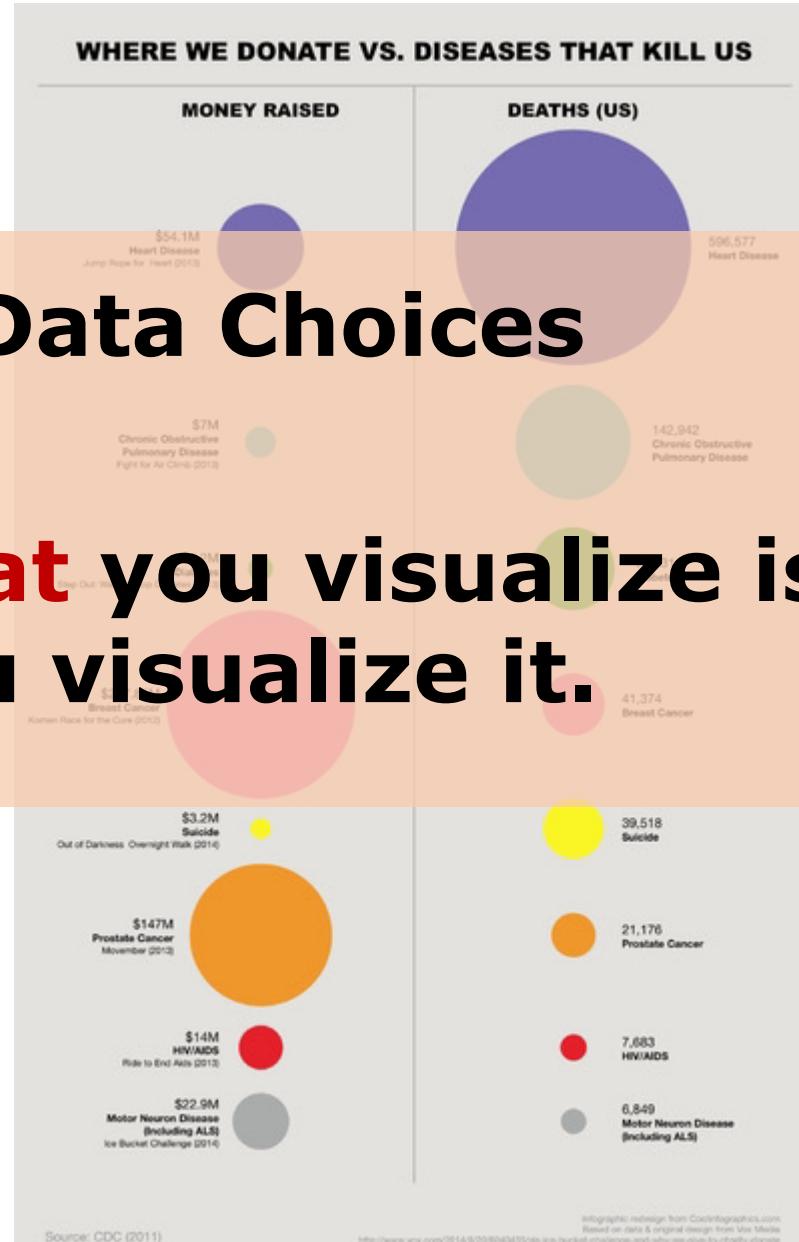
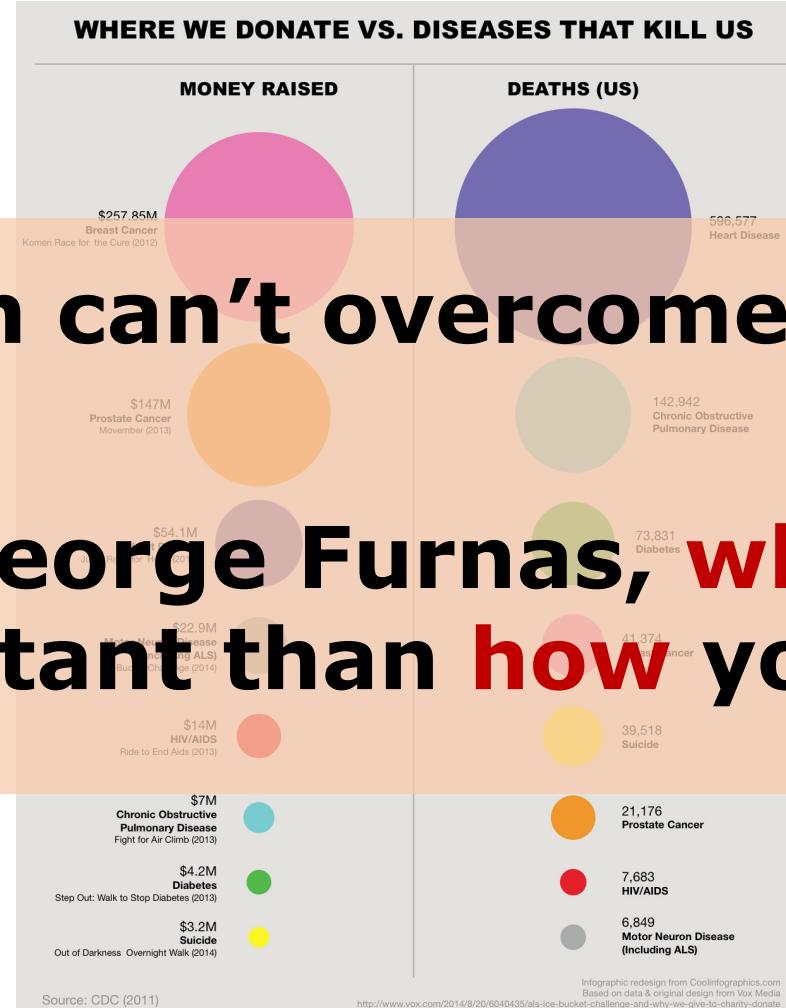
Images from Vox and
<http://coolinfographics.com/blog/2014/8/29/false-visualizations-sizing-circles-in-infographics.html>

Circles: Encode by Area not Radius



Visualization can't overcome Data Choices

To paraphrase George Furnas, **what you visualize is more important than how you visualize it.**



Images from Vox and
<http://coolinfographics.com/blog/2014/8/29/false-visualizations-sizing-circles-in-infographics.html>

What chart should I use?

**What are you trying to
understand?**

What is the nature of my data?

What is the nature of my task?

Some sources offer “visual vocabularies” for common abstract goals

Visual Vocabulary Deviation Correlation Ranking Distribution Change over Time Part-to-Whole Magnitude Spatial Flow

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationship is causal (one causes the other).

Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

A scatterplot showing the relationship between the percentage of people with a BA degree (Y-axis) and the percentage of obese people (X-axis). The X-axis ranges from 20 to 36, and the Y-axis ranges from 10 to 50. Data points are represented by bubbles of varying sizes, with state abbreviations labeled. A vertical line at approximately 27.5% on the X-axis indicates the US obesity average (27.0%). A horizontal line at approximately 27.2% on the Y-axis indicates the US average for people with a BA. A dashed regression line shows a negative correlation. The chart includes a legend on the left with color-coded squares for different themes: Visual Vocabulary (blue), Deviation (orange), Correlation (green), Ranking (red), Distribution (purple), Change over Time (pink), Part-to-Whole (yellow), Magnitude (teal), Spatial (light blue), and Flow (dark blue).

Line + Column

A good way of showing the relationship between an amount (columns) and a rate (line).

A dual-axis chart showing Sales (left Y-axis, orange bars) and Profit (right Y-axis, blue line) over time. The X-axis represents the Quarter of Order Date from 2015 Q3 to 2018 Q3. The left Y-axis for Sales ranges from 0K to 400K. The right Y-axis for Profit ranges from 0K to 80K. The profit line shows significant fluctuations, peaking around 2017 Q3 and 2018 Q3. Sales bars are shown in orange.

Connected Scatter

Usually used to show a relationship over time.

A line chart showing the top 0.01% of sales over time. The Y-axis ranges from 0% to 10%. The X-axis shows dates from 2015 to 2018. The line shows a general upward trend with some fluctuations, starting around 5% in 2015 and reaching about 8% in 2018. The chart includes a legend on the right with color-coded squares for different themes: Visual Vocabulary (blue), Deviation (orange), Correlation (green), Ranking (red), Distribution (purple), Change over Time (pink), Part-to-Whole (yellow), Magnitude (teal), Spatial (light blue), and Flow (dark blue).

Bubble

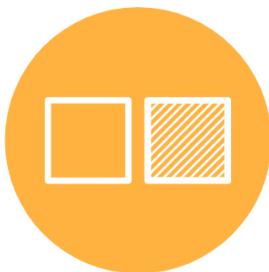
Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

XY Heatmap

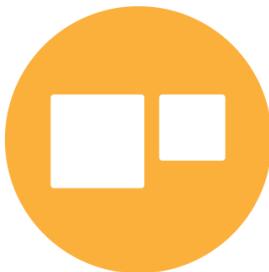
A good way of showing the patterns between 2 categories of data, less good at showing fine differences.

What do you want to show?

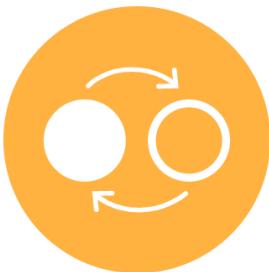
Here you can find a list of charts categorised by their data visualization functions or by what you want a chart to communicate to an audience. While the allocation of each chart into specific functions isn't a perfect system, it still works as a useful guide for selecting chart based on your analysis or communication needs.



Comparisons



Proportions



Relationships



Hierarchy



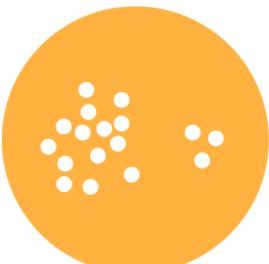
Concepts



Location



Part-to-a-whole



Distribution



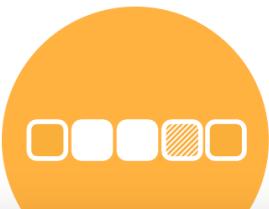
Data Sources



Tools



Flow



Progress

Visual Vocabulary

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it could be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Example FT uses
Trade surplus/deficit, climate change

Diverging bar

A simple standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar

Perfect for presenting survey results which involve sentiment (eg disagree/neural/agree).

Spine

Splits a single value into two contrasting components (eg male/female).

Surplus/deficit filled line

The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.

Correlation

Show the relationship between two or more variables. Be careful with this as you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses
Inflation and unemployment, income and life expectancy

Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

Column + line timeline

A good way of showing the relationship between an amount (columns) and a rate (line).

Connected scatterplot

Usually used to show the relationship between 2 variables has changed over time.

Bubble

Like a scatterplot but adds additional detail by sizing the circles according to a third variable.

XY heatmap

A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

Slope

Perfect for showing how ranks have changed over time or vary between categories.

Lollipop

Lollipops draw more attention to the data value than standard bar/column and can also show rank and value effectively.

Bump

Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using color.

Cumulative curve

A good way of showing how unequal a distribution is; it's always cumulative frequency, x axis is always a measure.

Frequency polygons

For displaying multiple distributions of data. Like a regular line chart, best limited to a maximum of 3 or 4 datasets.

Beeswarm

Use to emphasize individual points in a distribution. Points can be sized to an additional variable. Best with medium

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results

Ordered bar

Standard bar charts display the ranks of values much easier when sorted into order.

Ordered column

See above.

Ordered proportional symbol

Use when there are big differences between values and/or seeing fine differences between data is not so important.

Dot strip plot

Dot placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.

Barcode plot

Like dot strip plot, good for displaying the data in a tabular form when highlighting individual values.

Boxplot

Summarise multiple distributions by showing the median (centre) and range (the data)

Violin plot

Similar to a box plot, more effective with complex distributions (data that cannot be summarised with averages).

Population pyramid

A standard way of showing the age breakdown of a population distribution, effectively back histograms.

Cumulative curve

A good way of showing how unequal a distribution is; it's always cumulative frequency, x axis is always a measure.

Frequency polygons

For displaying multiple distributions of data. Like a regular line chart, best limited to a maximum of 3 or 4 datasets.

Beeswarm

Use to emphasize individual points in a distribution. Points can be sized to an additional variable. Best with medium

Show values in a dataset and how they occur. The shape ('or 'skew') of distributions can be memorable when highlighting the lack of uniformity in equality in the data.

Example FT uses
Income distribution, population ($\alpha/\beta/\gamma$) sex distribution, revealing inequality

Histogram

The standard way to show a statistical distribution - key gaps between categories small to highlight 'shape' of the data

Dot plot

A simple way of showing the change (range) of data across multiple categories.

Dot strip plot

Good for showing individual values distribution, can be problem when too many dots have same value.

Barcode plot

Like dot strip plot, good for displaying the data in a tabular form when highlighting individual values.

Boxplot

Summarise multiple distributions by showing the median (centre) and range (the data)

Violin plot

Similar to a box plot, more effective with complex distributions (data that cannot be summarised with averages).

Population pyramid

A standard way of showing the age breakdown of a population distribution, effectively back histograms.

Cumulative curve

A good way of showing how unequal a distribution is; it's always cumulative frequency, x axis is always a measure.

Frequency polygons

For displaying multiple distributions of data. Like a regular line chart, best limited to a maximum of 3 or 4 datasets.

Beeswarm

Use to emphasize individual points in a distribution. Points can be sized to an additional variable. Best with medium

What if I have tabular data?

...like a DataFrame or CSV

Tasks for tabular data...

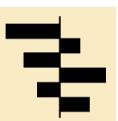
Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Example FT uses

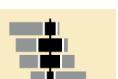
Trade surplus/deficit, climate data

Diverging bar



A simple standard chart that can show both negative and positive magnitude values.

Diverging stacked bar



Perfect for positive and negative survey results that involve sentiment (disagree/neutral/agree).

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Example FT uses

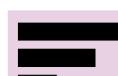
Commodity production, market capitalisation, volumes in general

Column



The standard way to compare the size of things. Must always start at 0 on the axis.

Bar



See above. Good when the data are not time series and labels have long category names.

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Example FT uses

Fiscal budgets, company structures, national election results

Stacked column/bar



A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Marimekko



A good way of showing the size and proportion of data at the same time – as

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses

Wealth, deprivation, league tables,

Basic choropleth (rate/ratio)



The standard approach for putting data on a map – should always be rates rather than totals and use a sensible base geography.

Proportional symbol (count/magnitude)



Use for totals rather than rates – be wary that small differences in data will be hard to

Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses

Its, economic time series in a market

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses

Movement of funds, trade, migrants, lawsuits, information; relationship graphs.

Sankey



Shows changes in flows from one condition to at least one other; good for tracing the eventual outcome of a complex process.

Waterfall



Designed to show the sequencing of data through a flow process, typically

Histograms

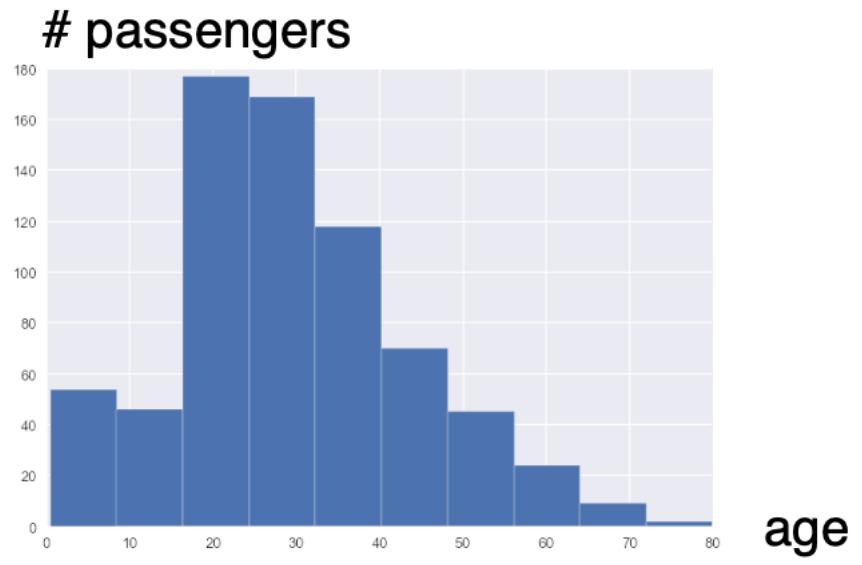
Hard to predict how many bins to use.

...possibly solve with interactivity!

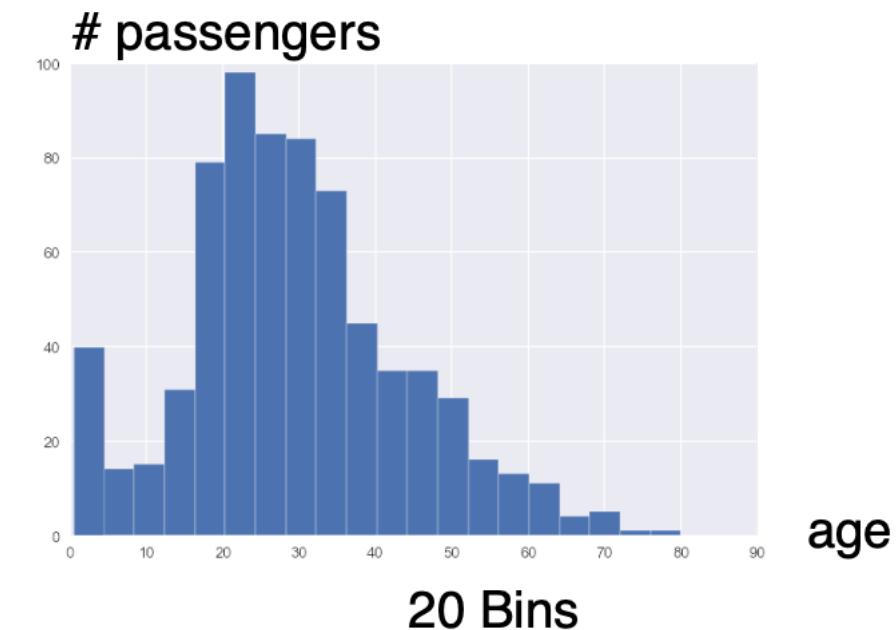
Some rules of thumb:

$$\# \text{bins} = \sqrt{n}$$

$$\# \text{bins} = \log_2(n) + 1$$



10 Bins



20 Bins

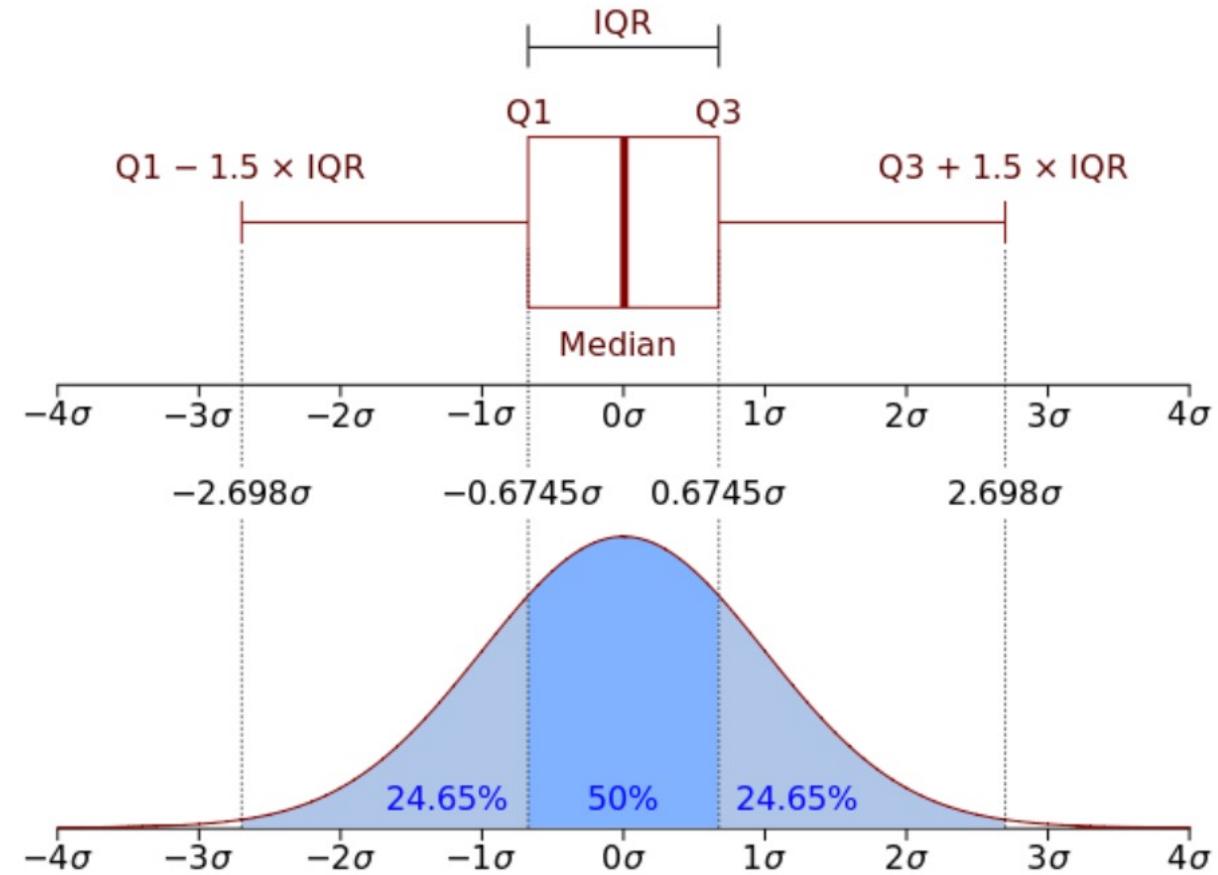
Box plots

a.k.a. “Box-and-whisker plots”

Helpful to show outliers as points

Less useful for data not following normal distribution

Especially bad for bimodal or multi-modal distributions.

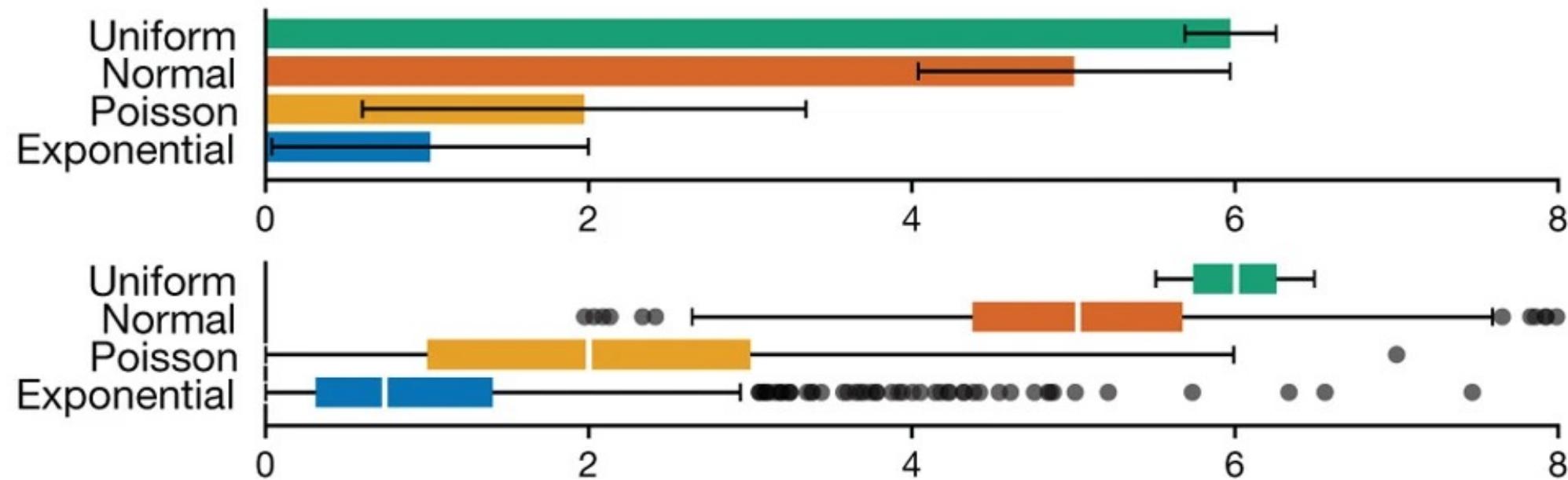


Wikipedia

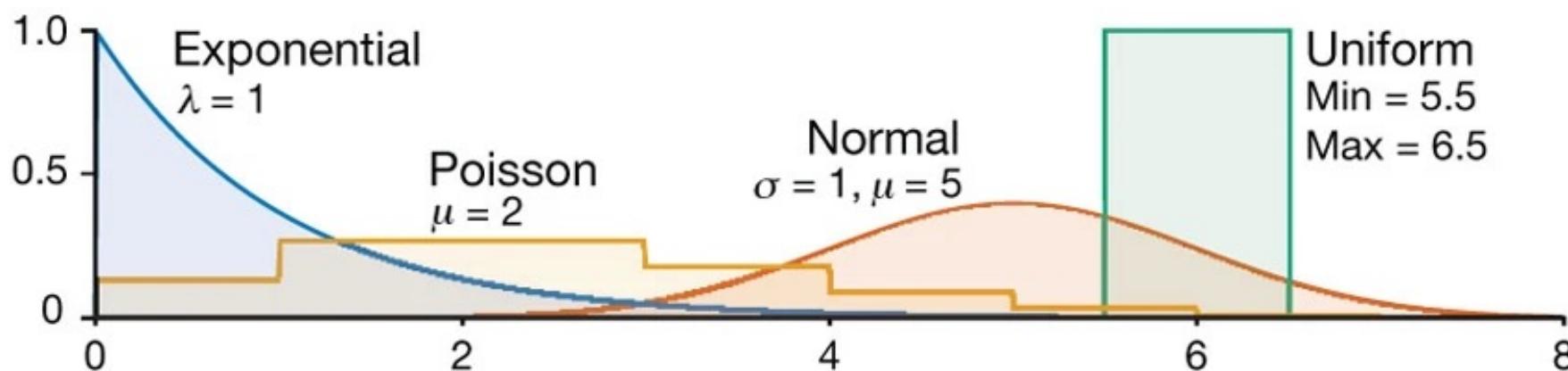
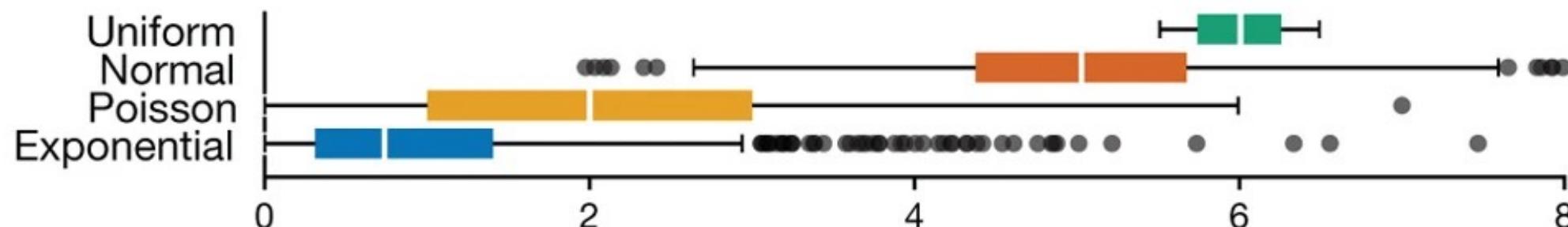
Error bars vs. Boxplots on some test datasets



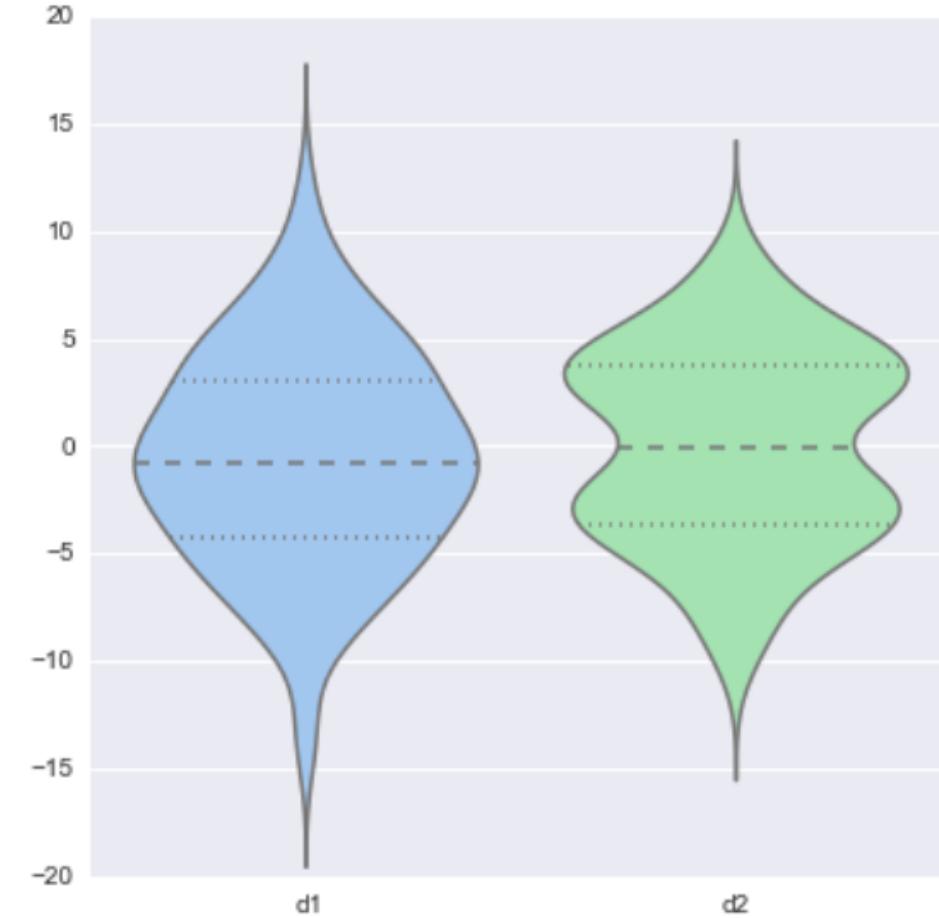
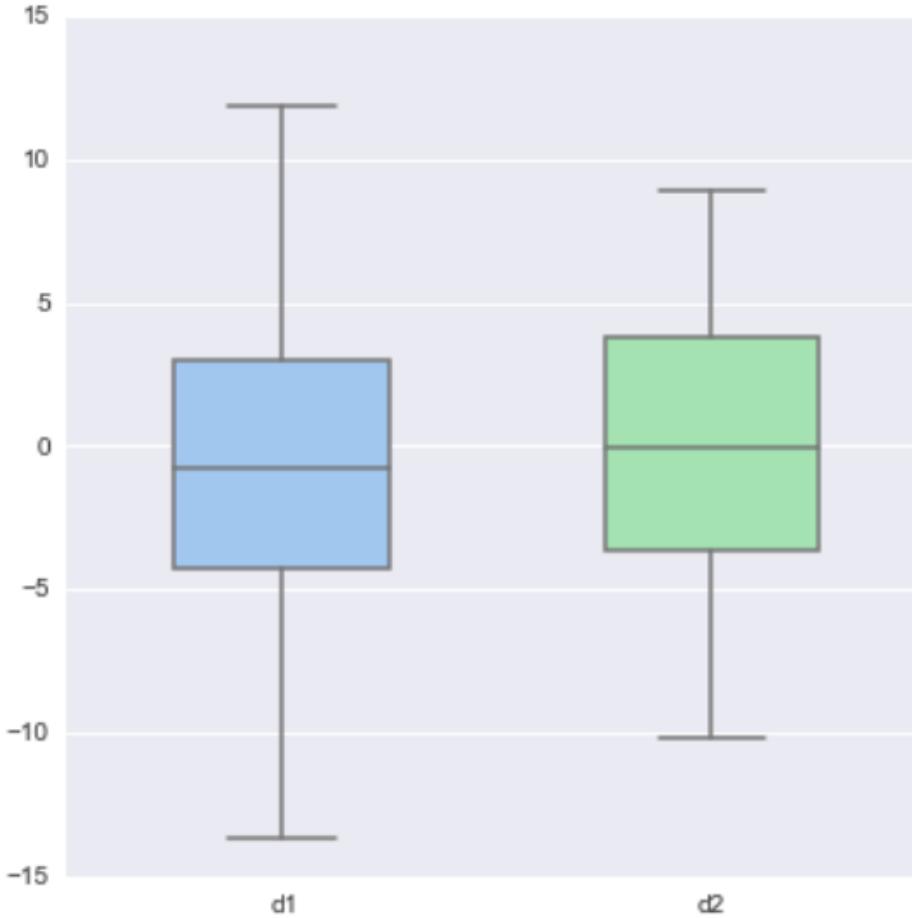
Error bars vs. Boxplots on some test datasets



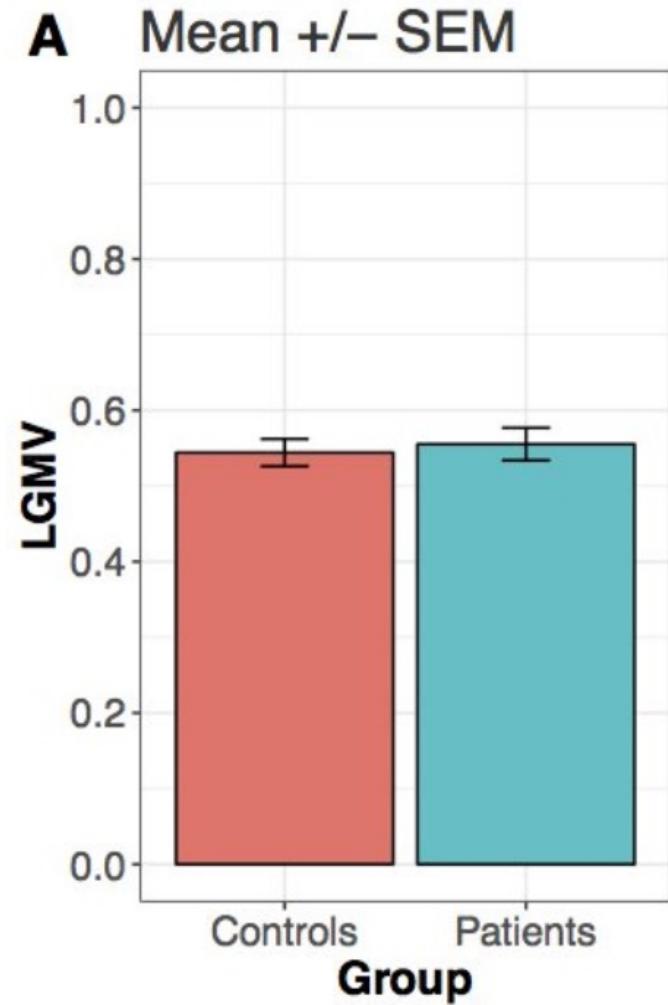
Error bars vs. Boxplots on some test datasets



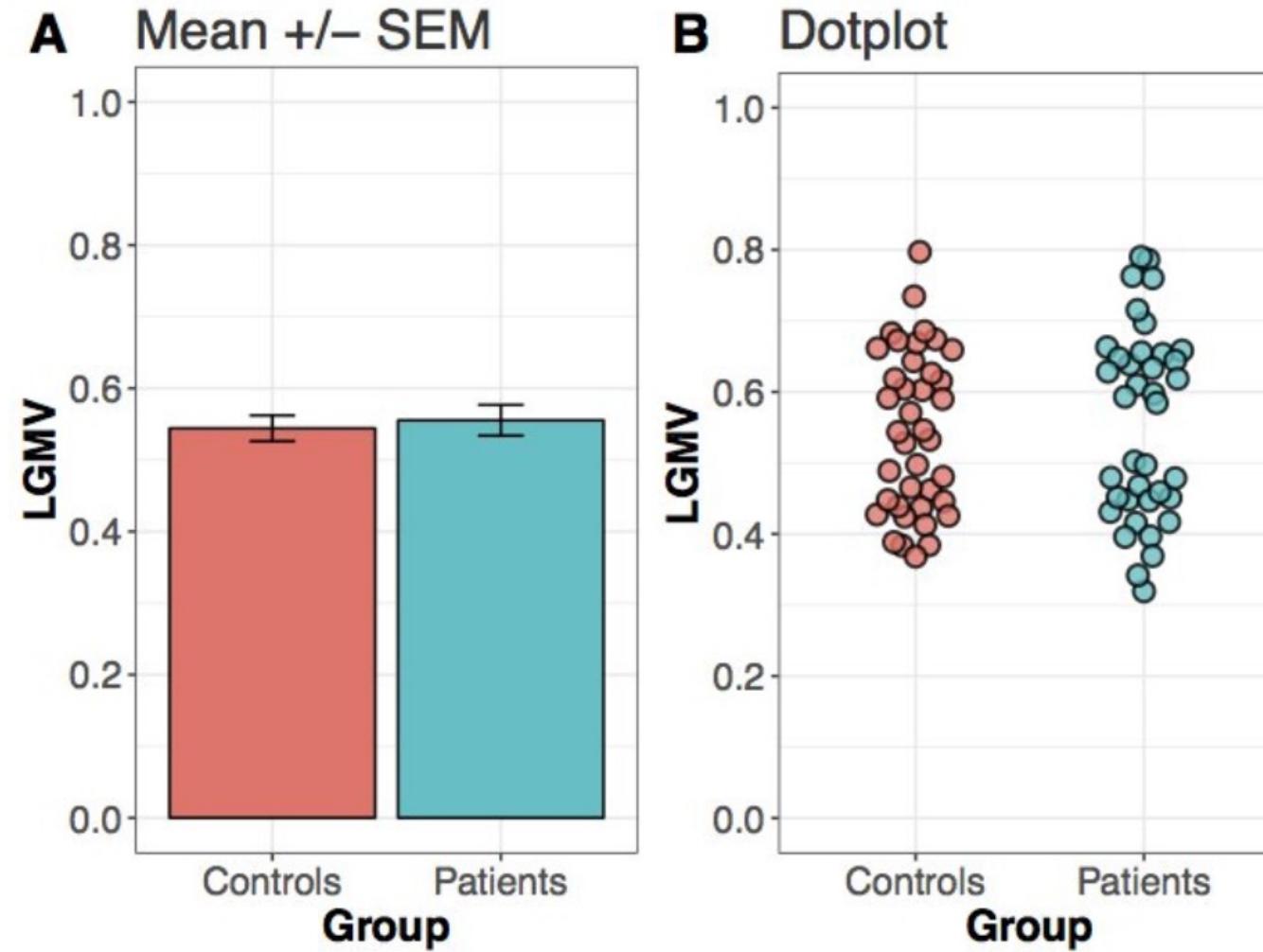
Violin plots are box plots with probability density functions



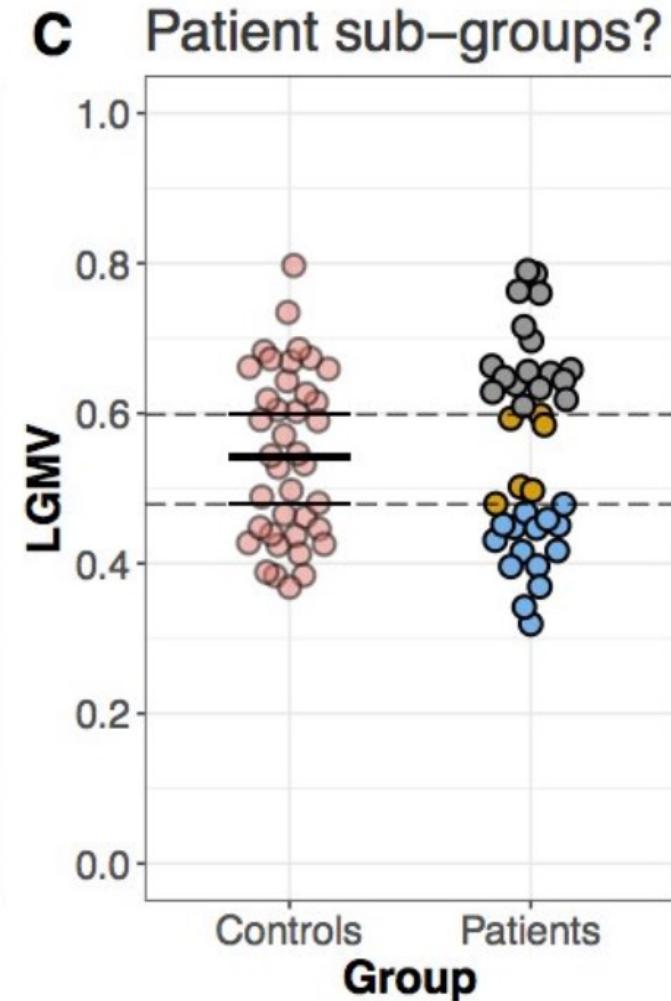
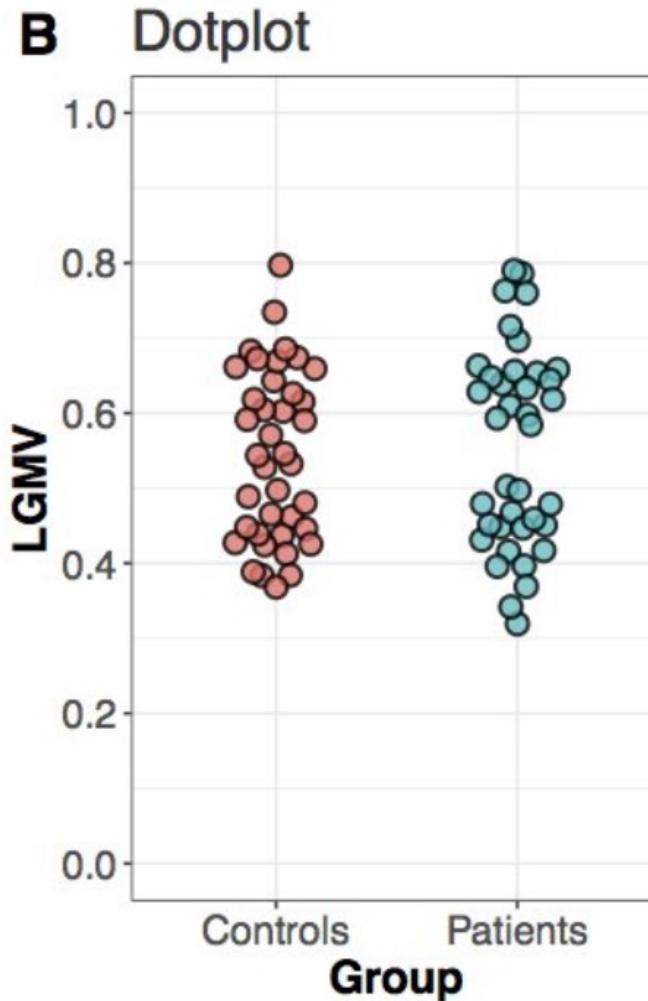
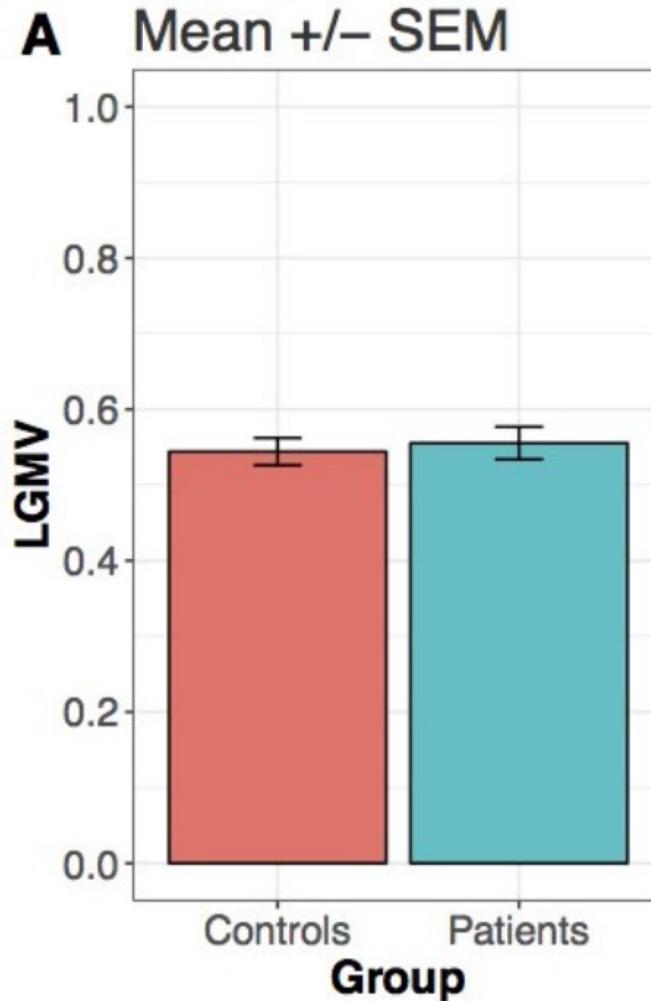
Box plots vs. Dot plots



Box plots vs. Dot plots

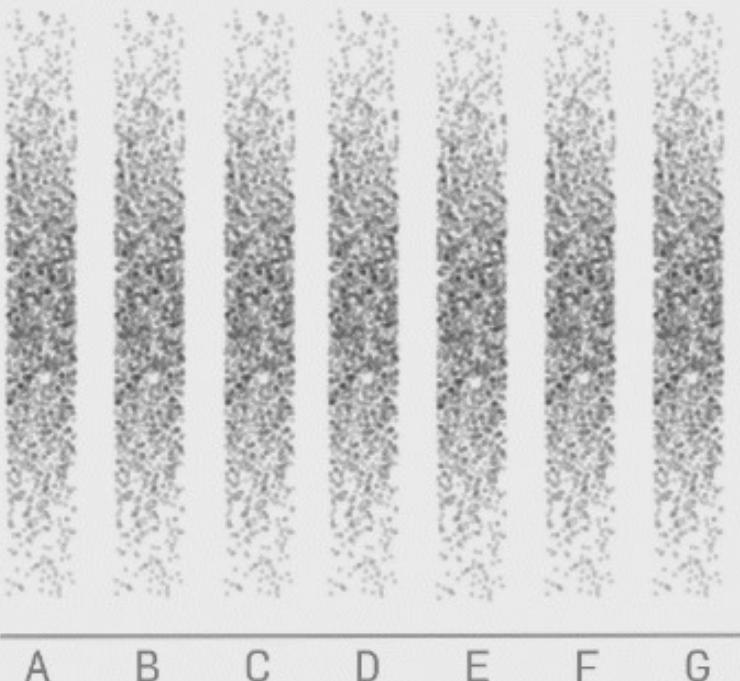


Box plots vs. Dot plots

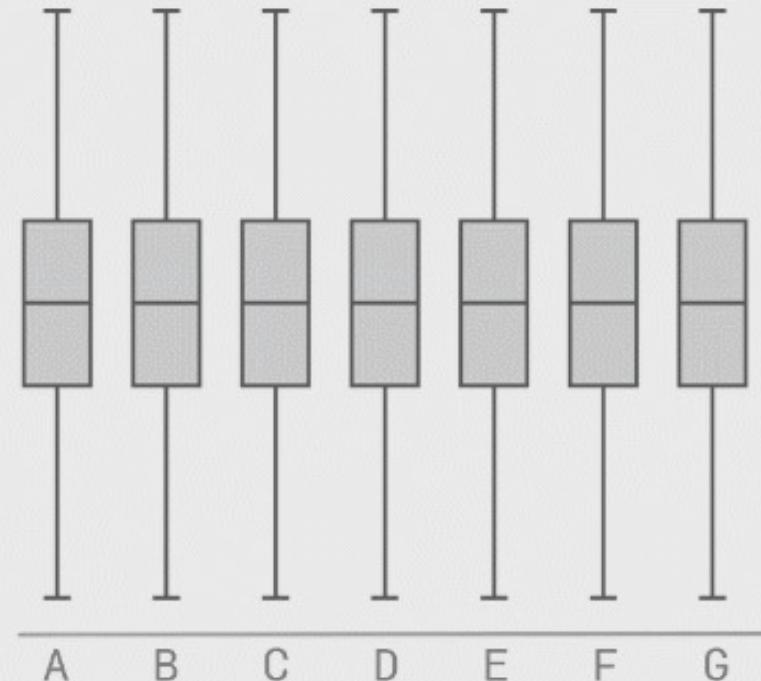


Different distributions, same box plot

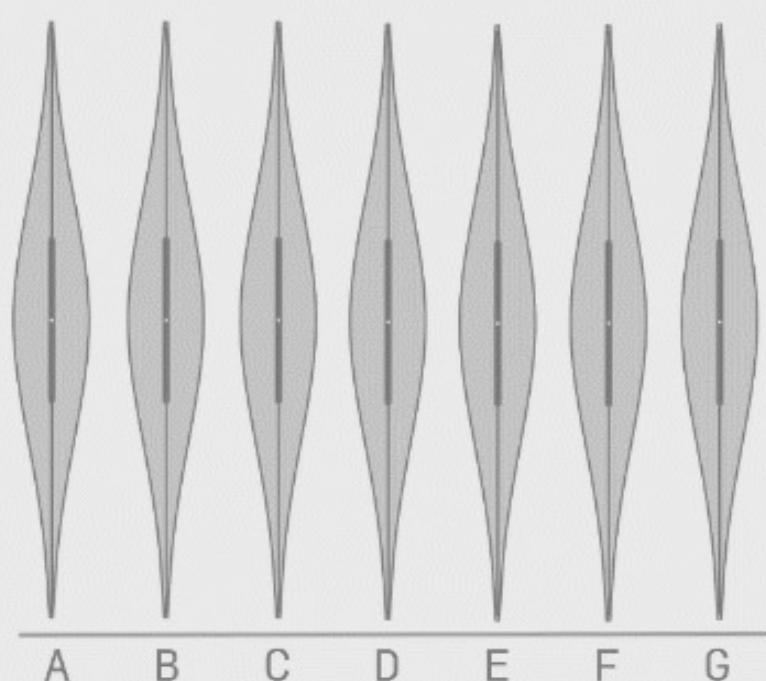
Raw Data



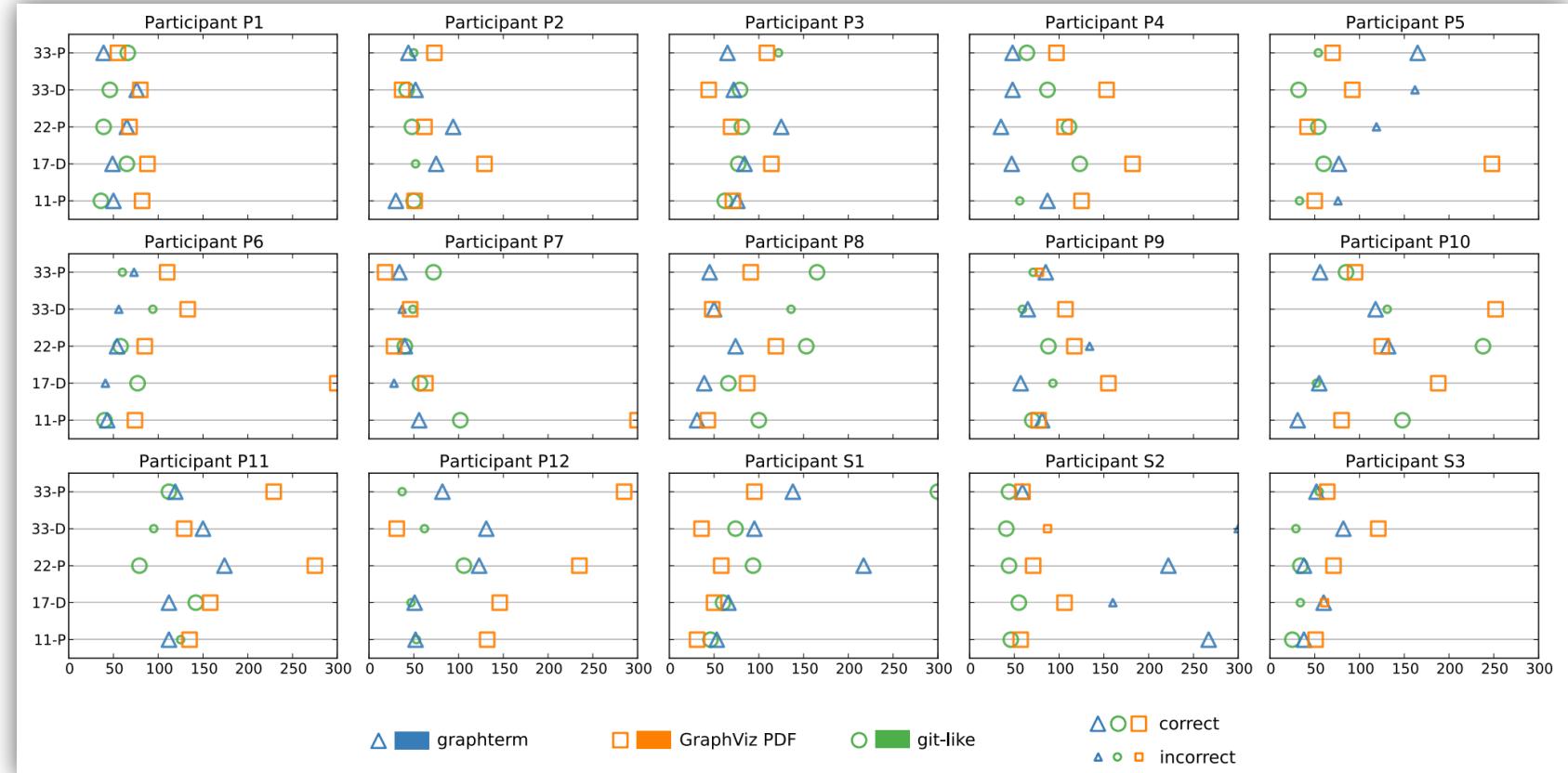
Box-plot of the Data



Violin-plot of the Data



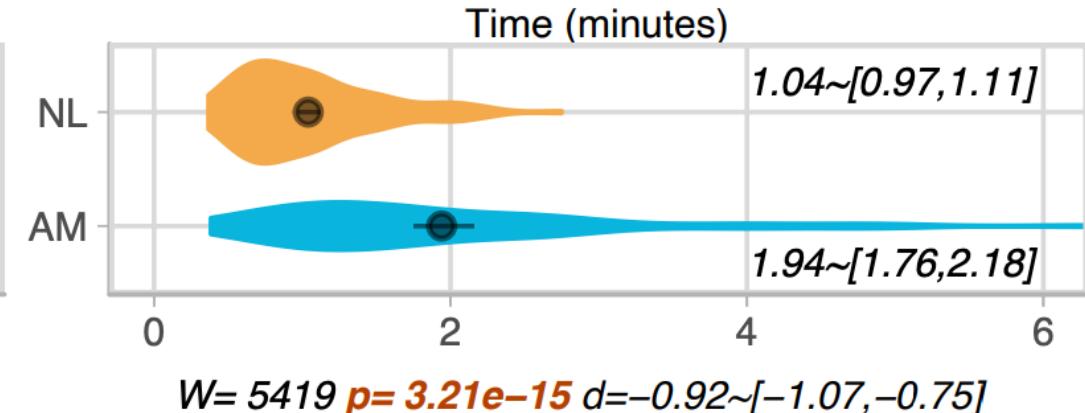
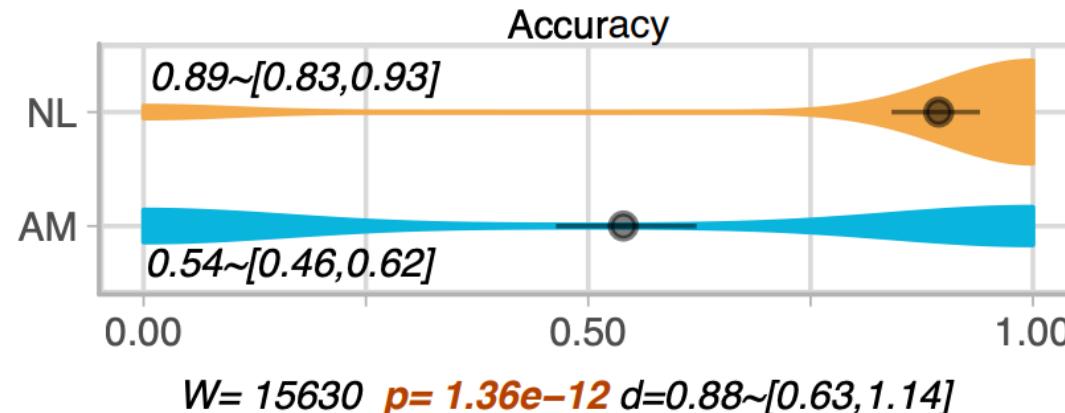
**When possible,
show all the
data**



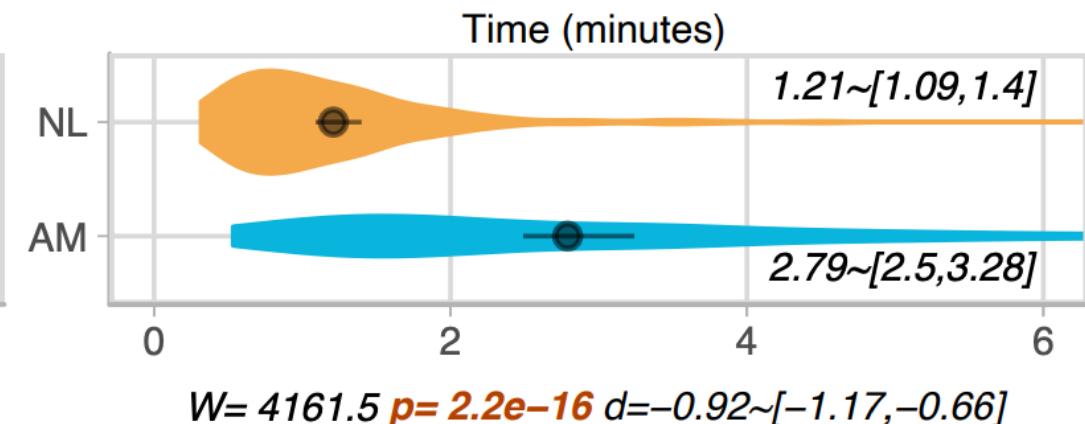
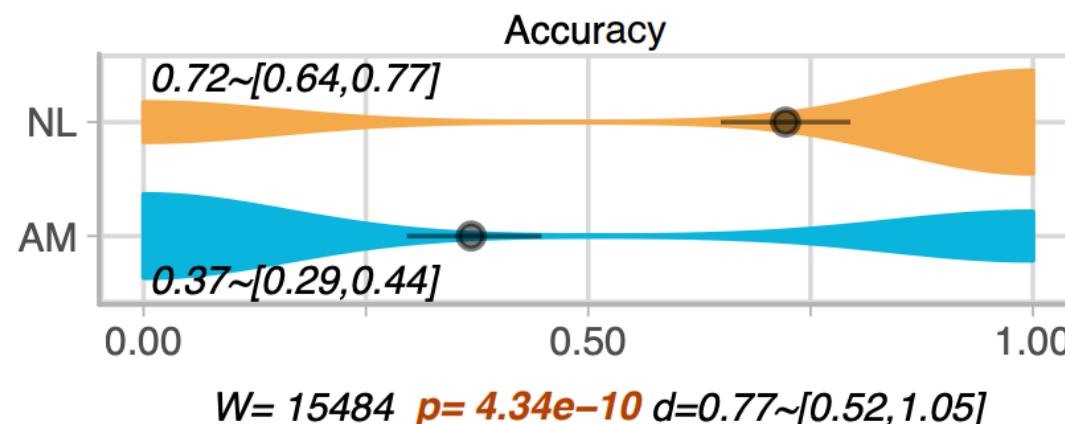
**Can pair with
summary**

It's not always possible to show all the data, so show confidence interval in context

T13 – Attribute along Shortest Path

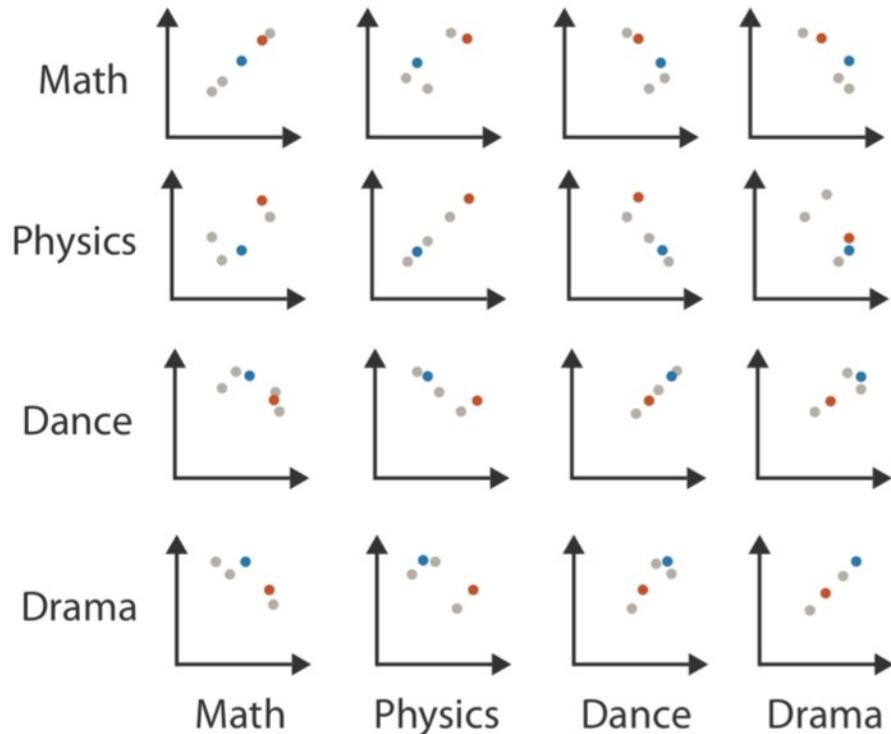


T15 – Attribute on Multiple Paths

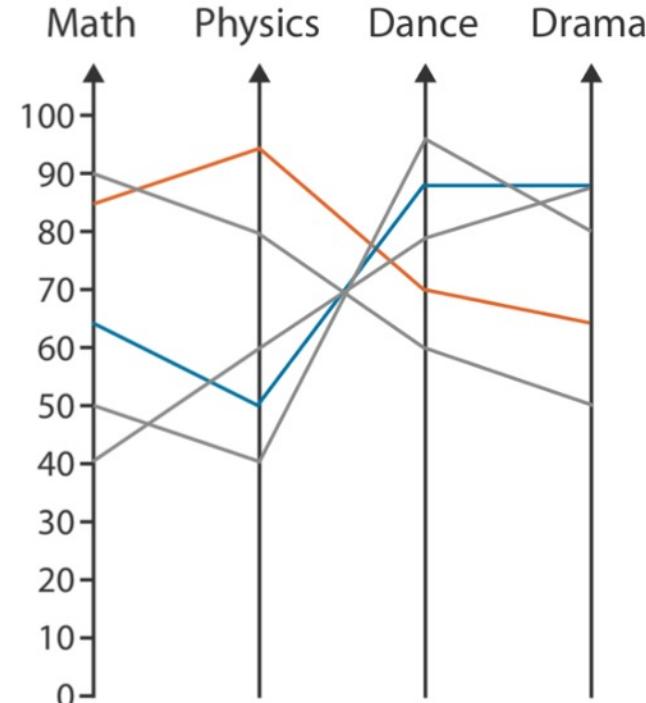


Compare many attributes...

Scatterplot Matrix



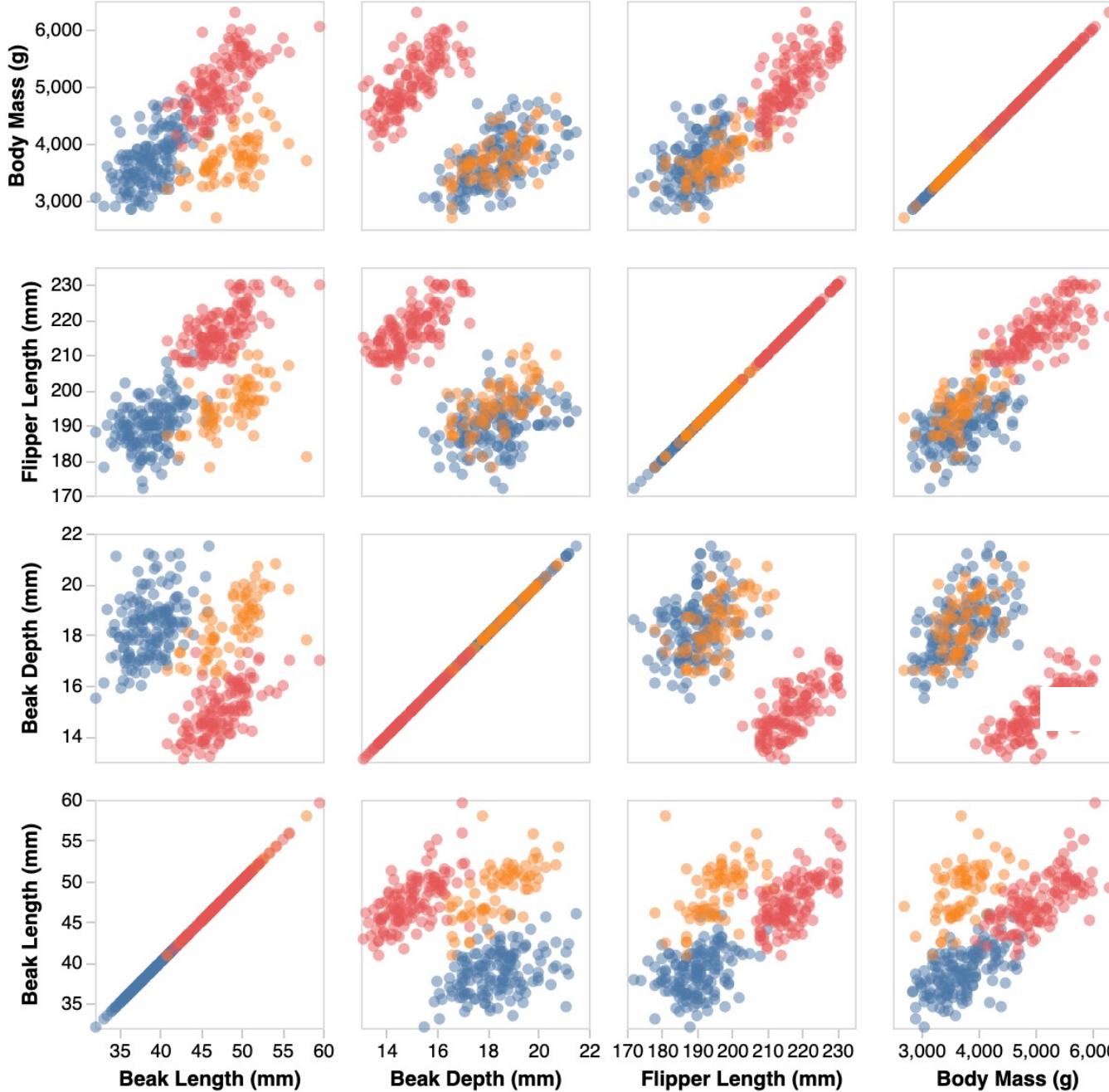
Parallel Coordinates



Table

	Math	Physics	Dance	Drama
85	95	70	65	
90	80	60	50	
65	50	90	90	
50	40	95	80	
40	60	80	90	

Scatterplot Matrix (SPLOM)



Matrix of size $d \times d$

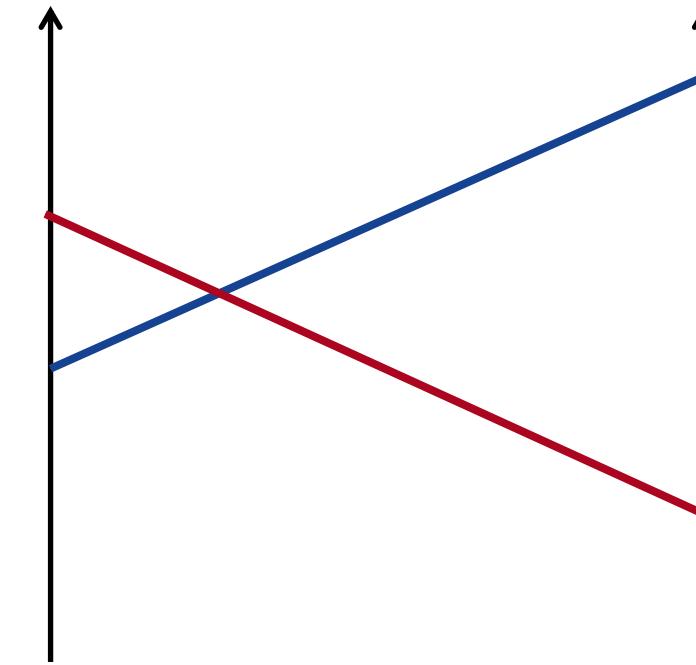
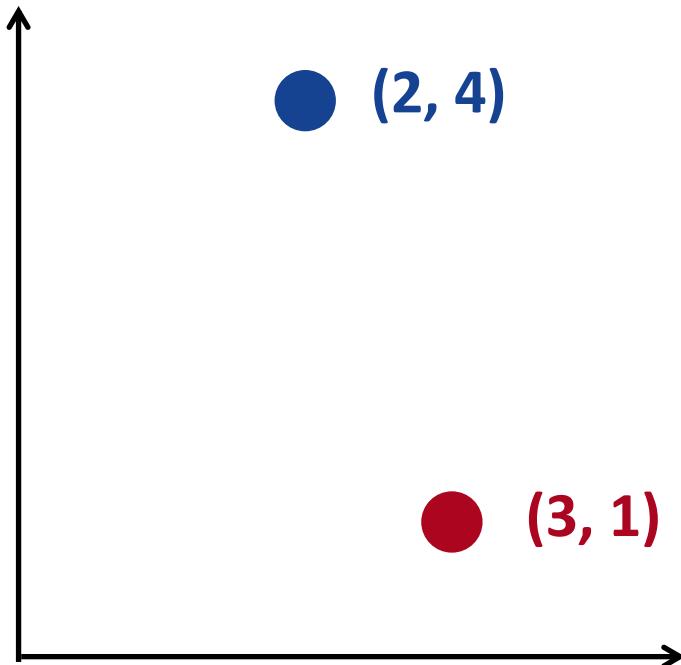
Each row/column is one dimension

Each plot is two dimensions

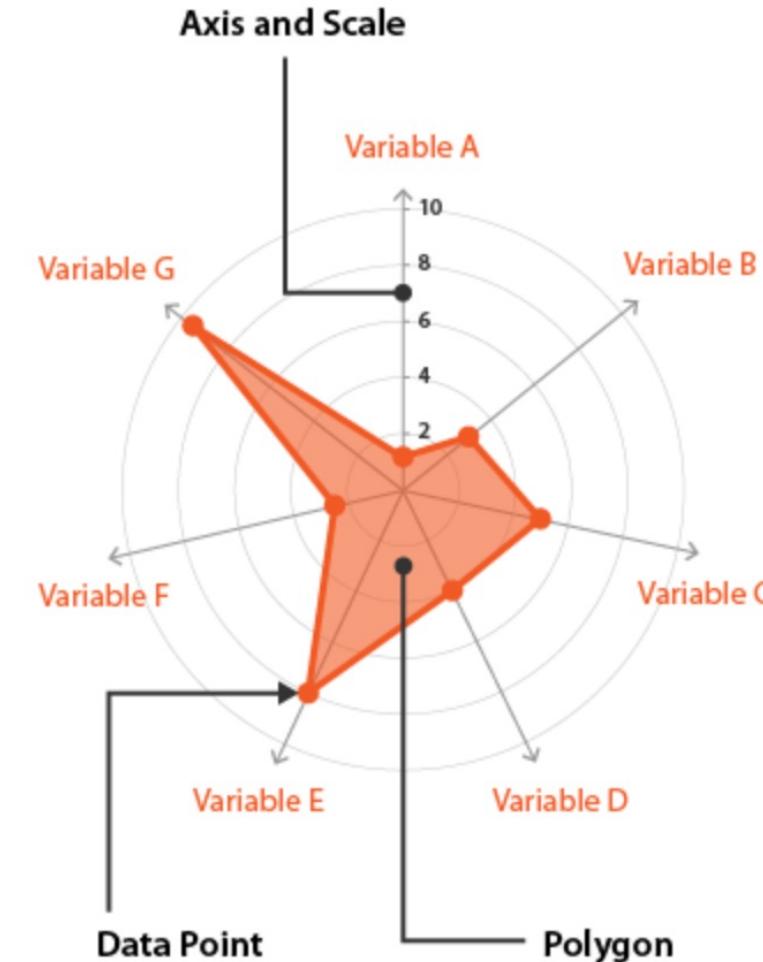
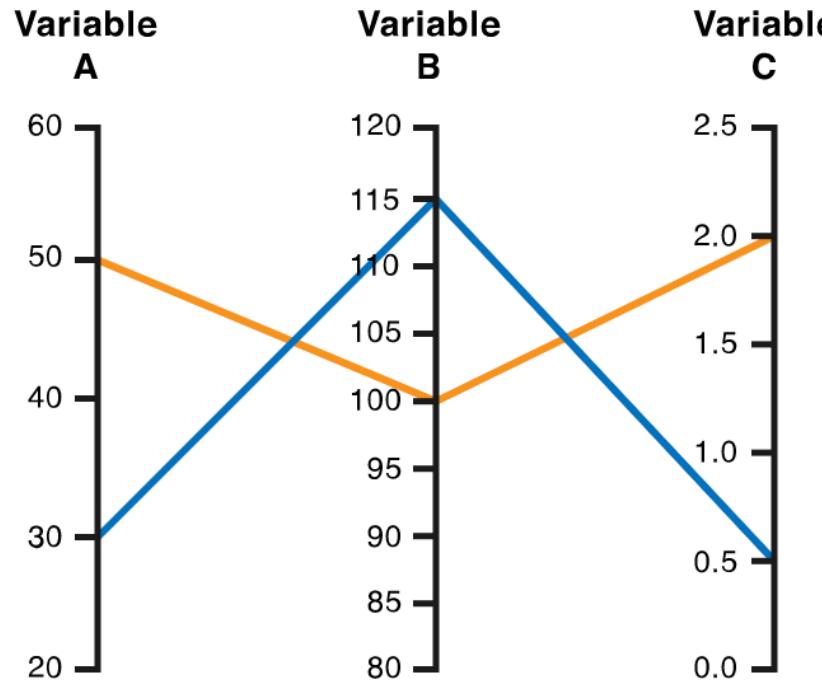
Parallel coordinates

Each dimension gets one parallel axis

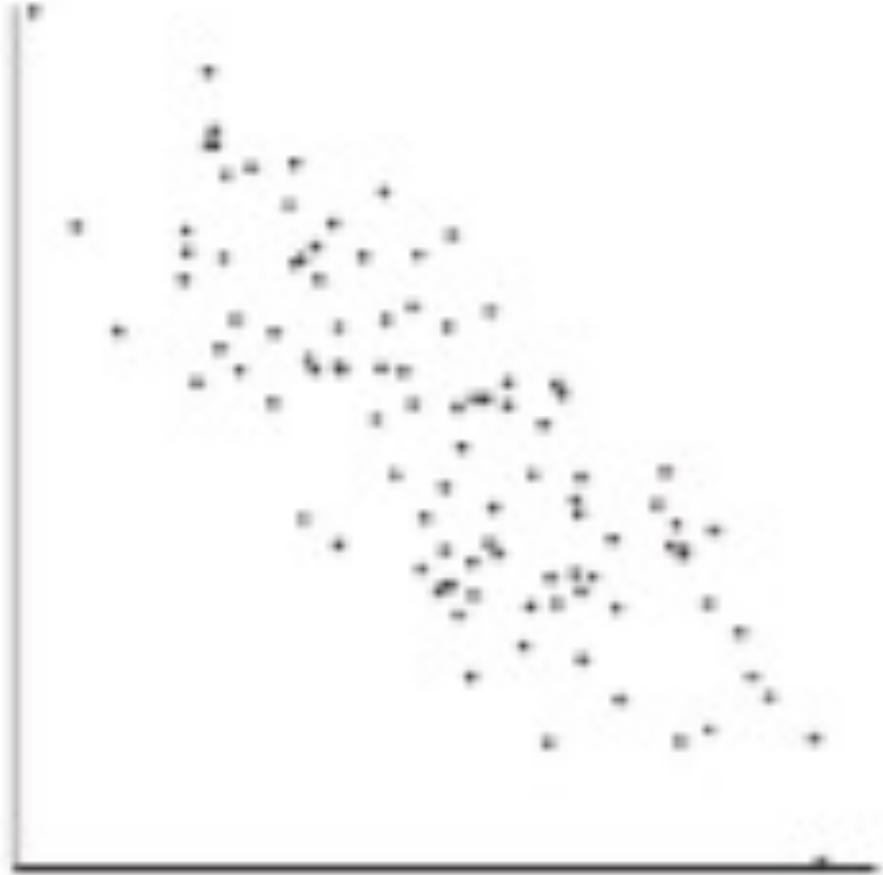
Items in dataset represented with a (poly-)line



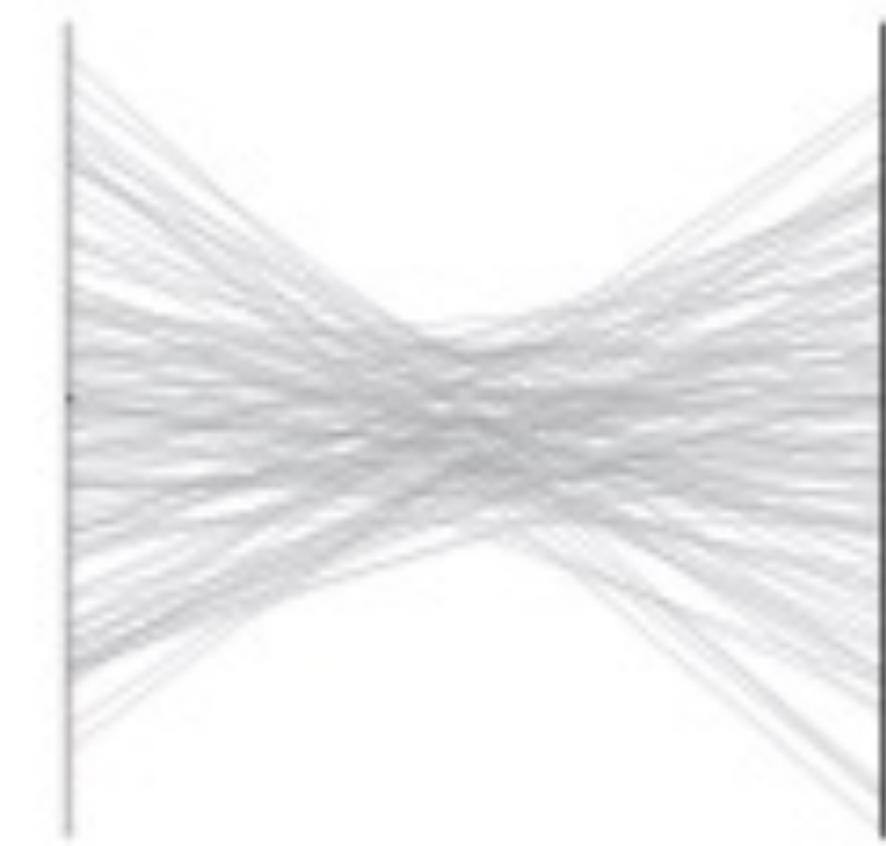
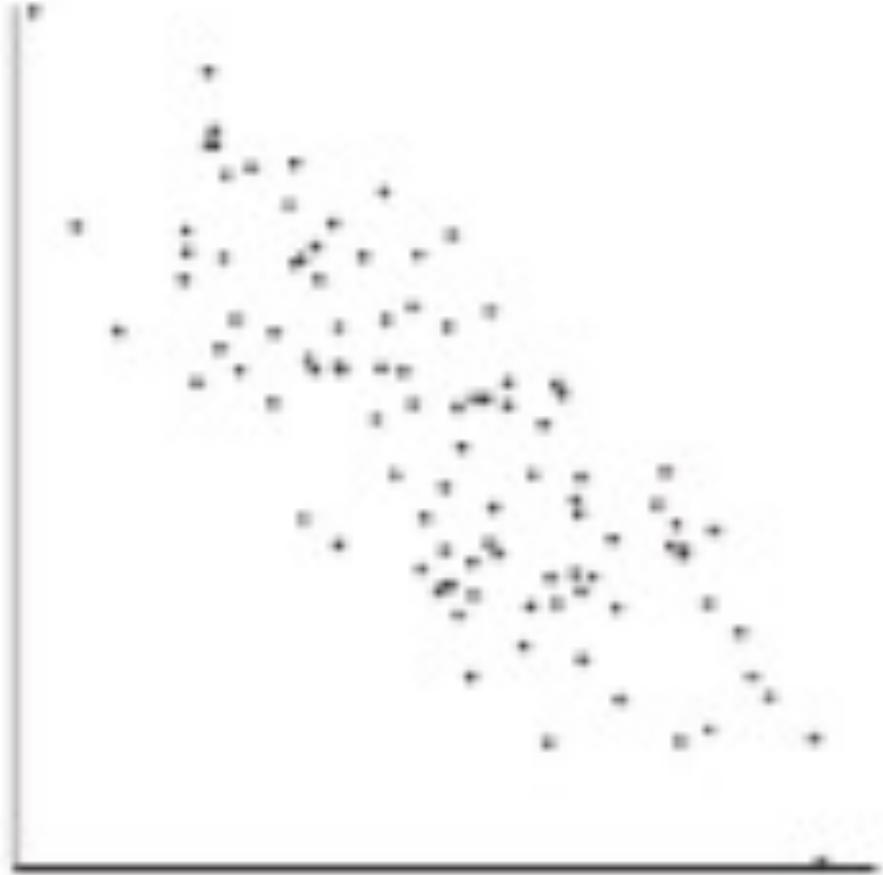
Parallel coordinate plots are similar to Radar/Star charts but have common orientation.



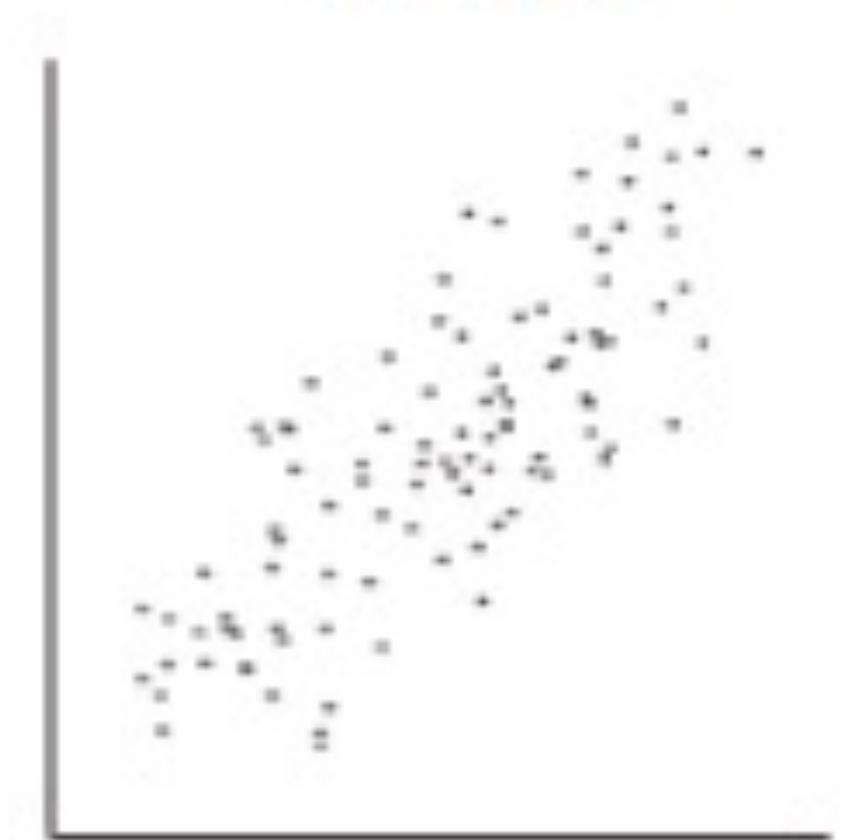
Scatterplots versus Parallel Coordinates



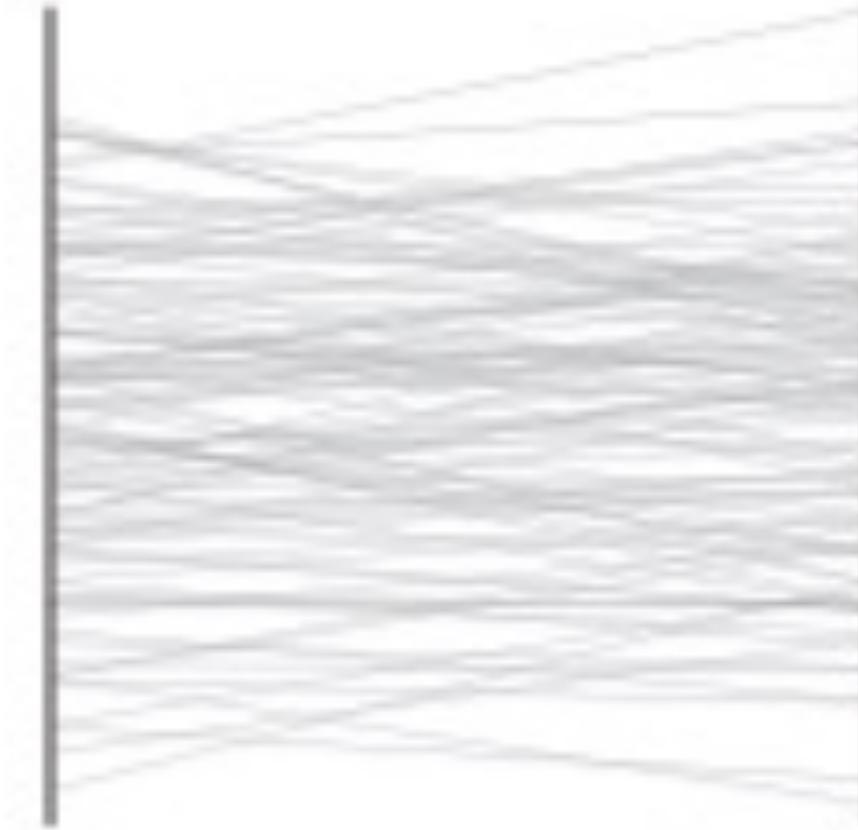
Scatterplots versus Parallel Coordinates



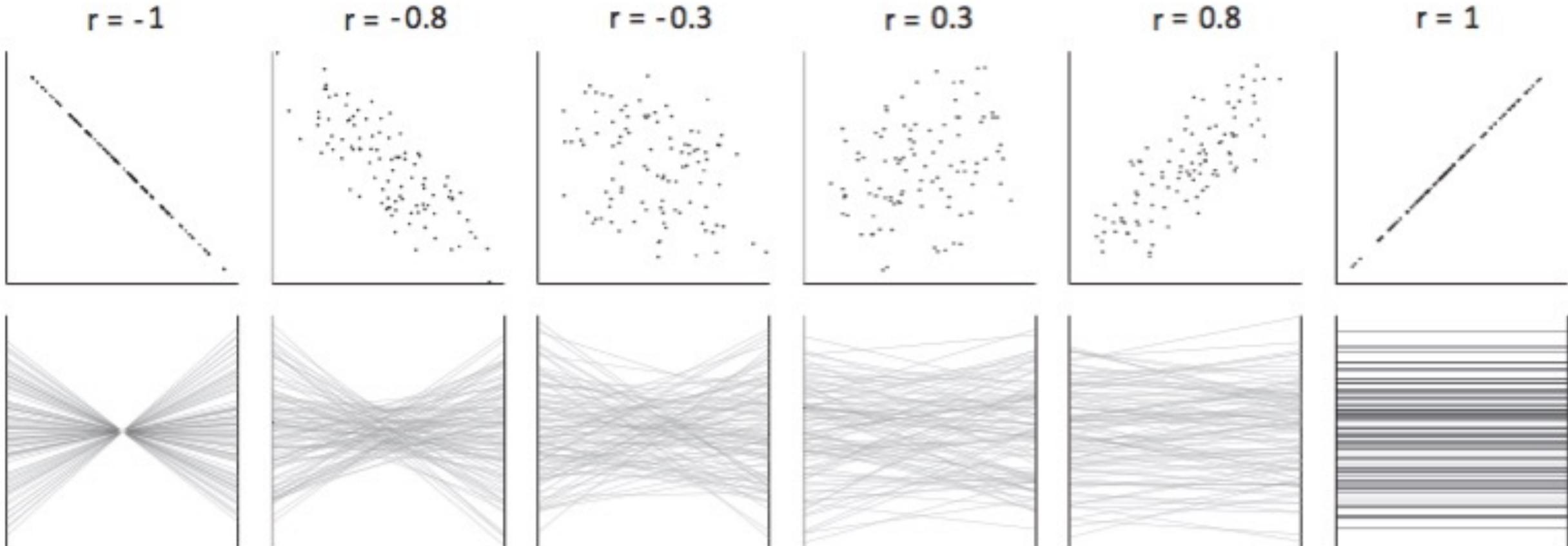
Scatterplots versus Parallel Coordinates



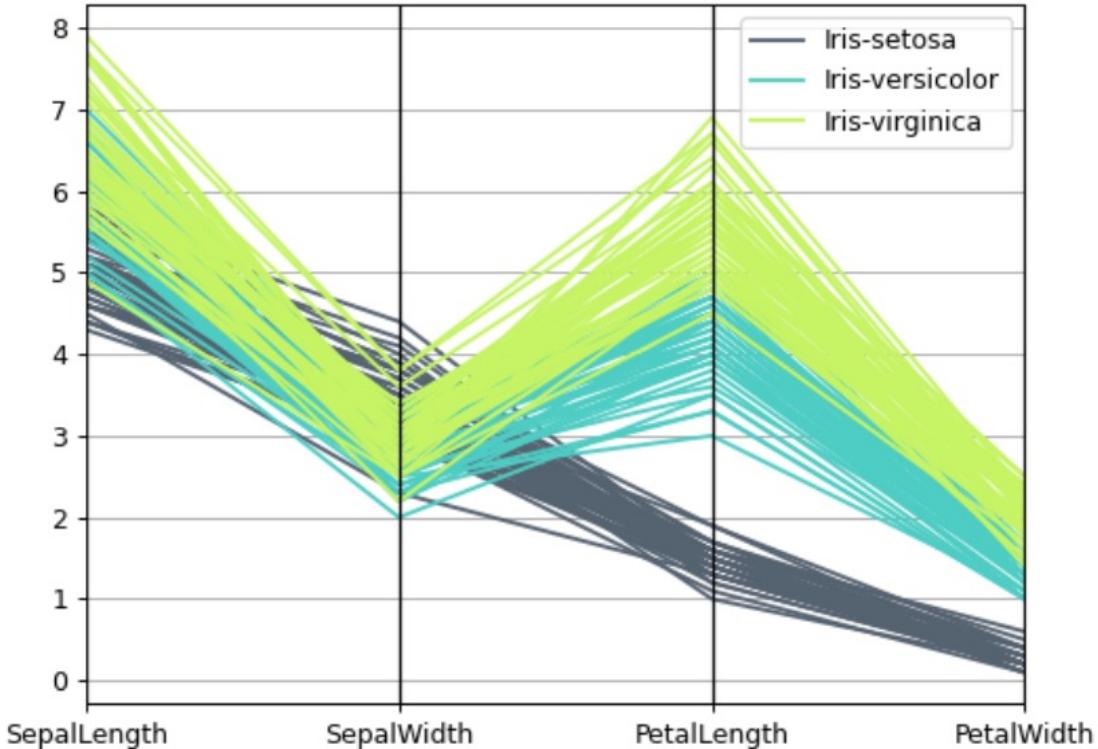
Scatterplots versus Parallel Coordinates



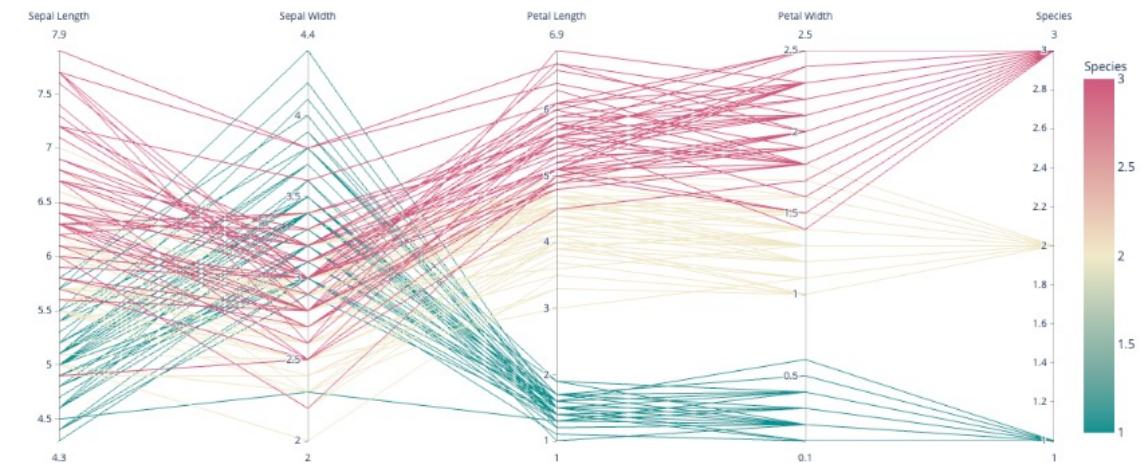
Scatterplots versus Parallel Coordinates



Pandas and Plot.ly support



https://pandas.pydata.org/docs/reference/api/pandas.plotting.parallel_coordinates.html

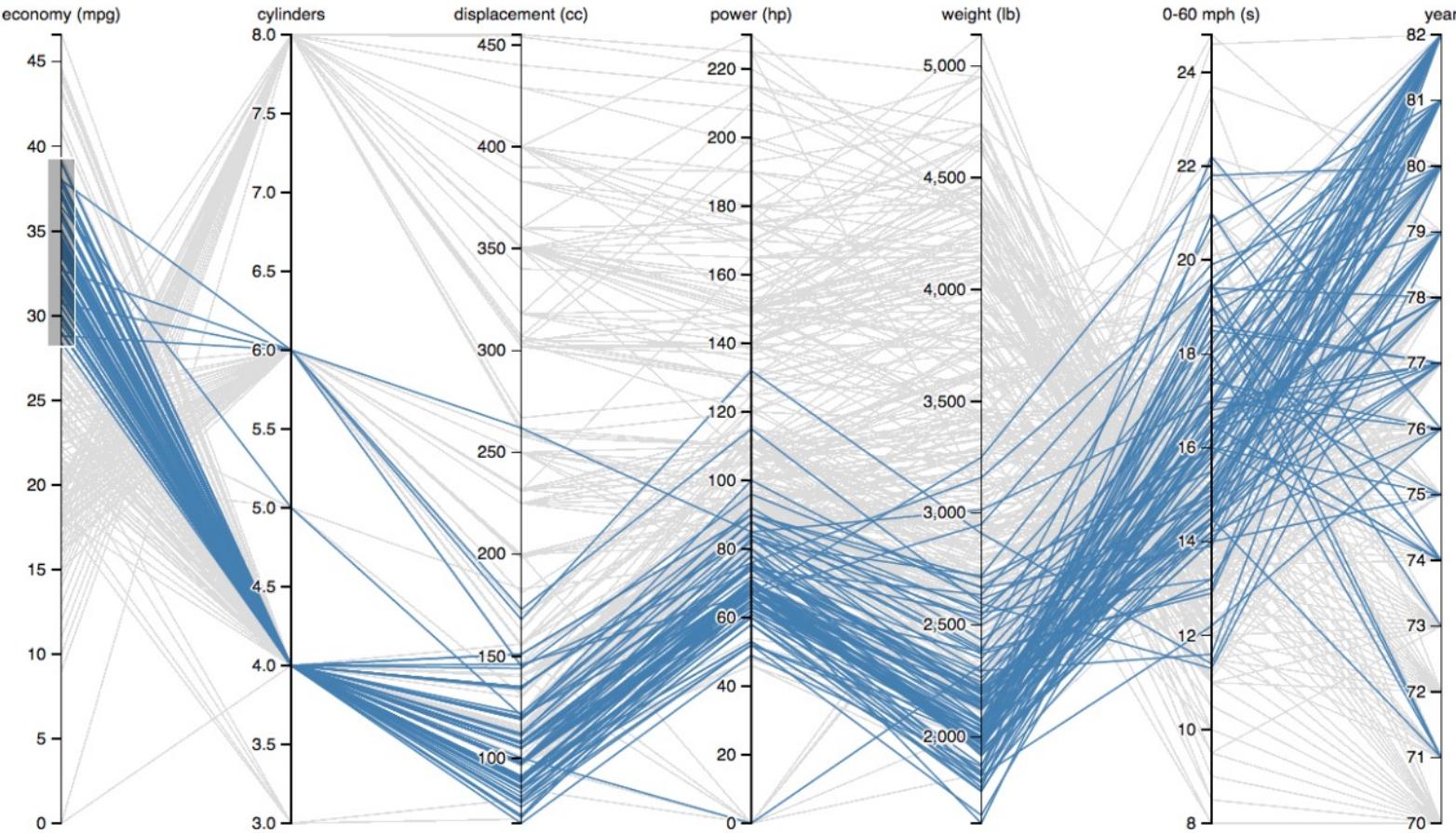


<https://plotly.com/python/parallel-coordinates-plot/>

Parallel coordinates

Brushable across many axes at once

Axes may be re-ordered



Pixel-based displays

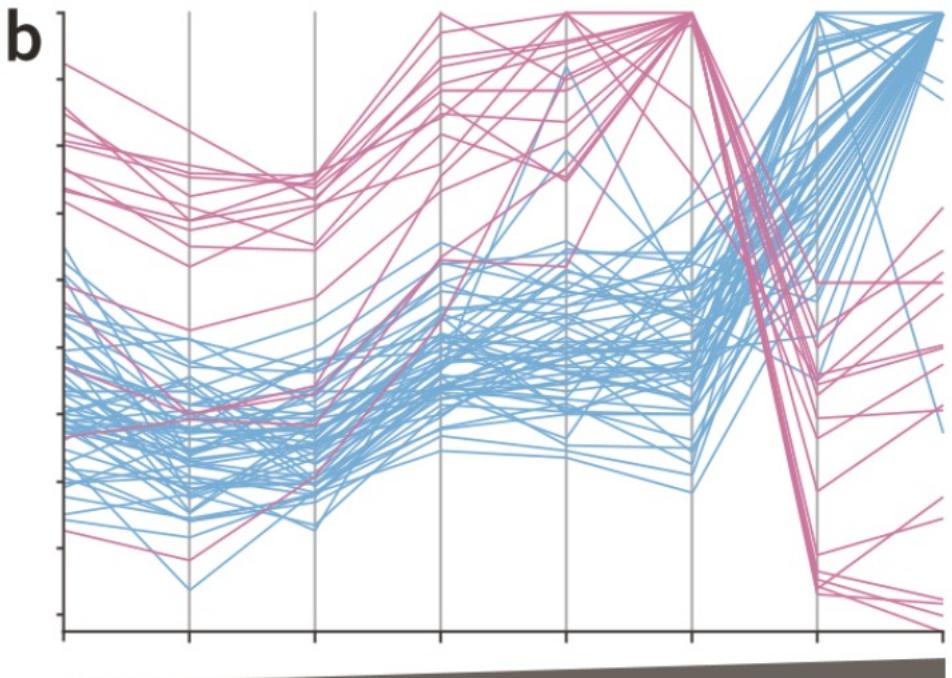
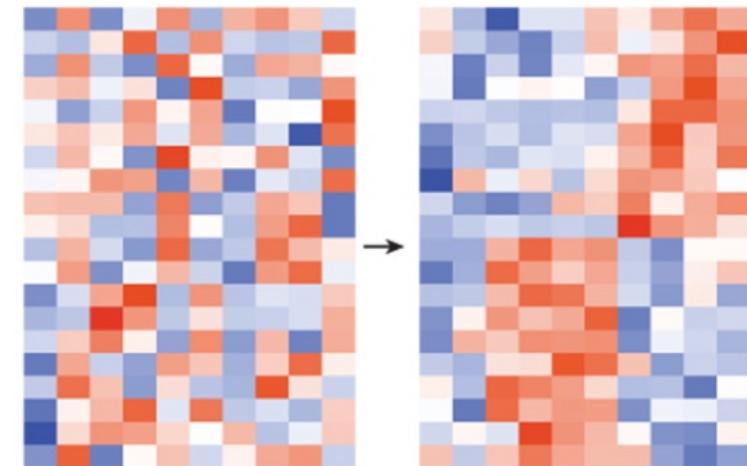
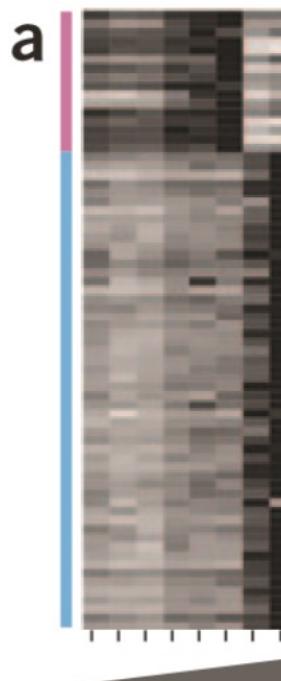
Each cell is a “pixel”
Value encoded with color

Relies on ordering for
interpretability

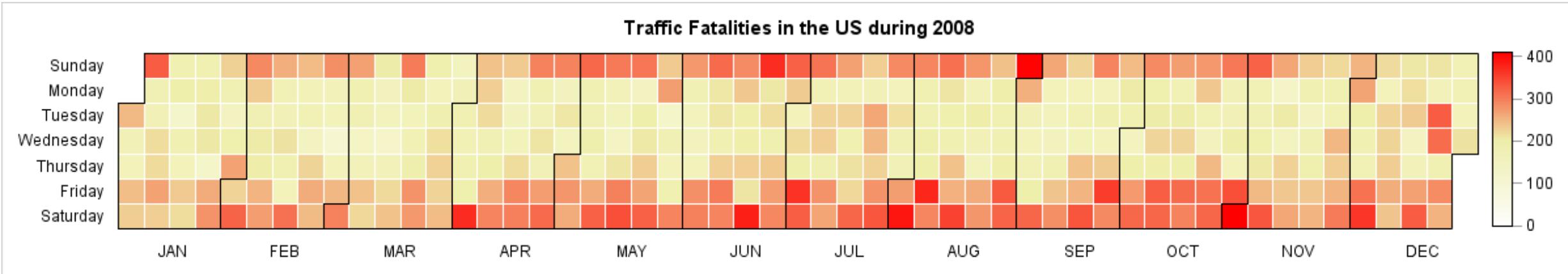
If order isn’t obvious, use
clustering

May scale to 1px / item

Good for data with same
scale and type

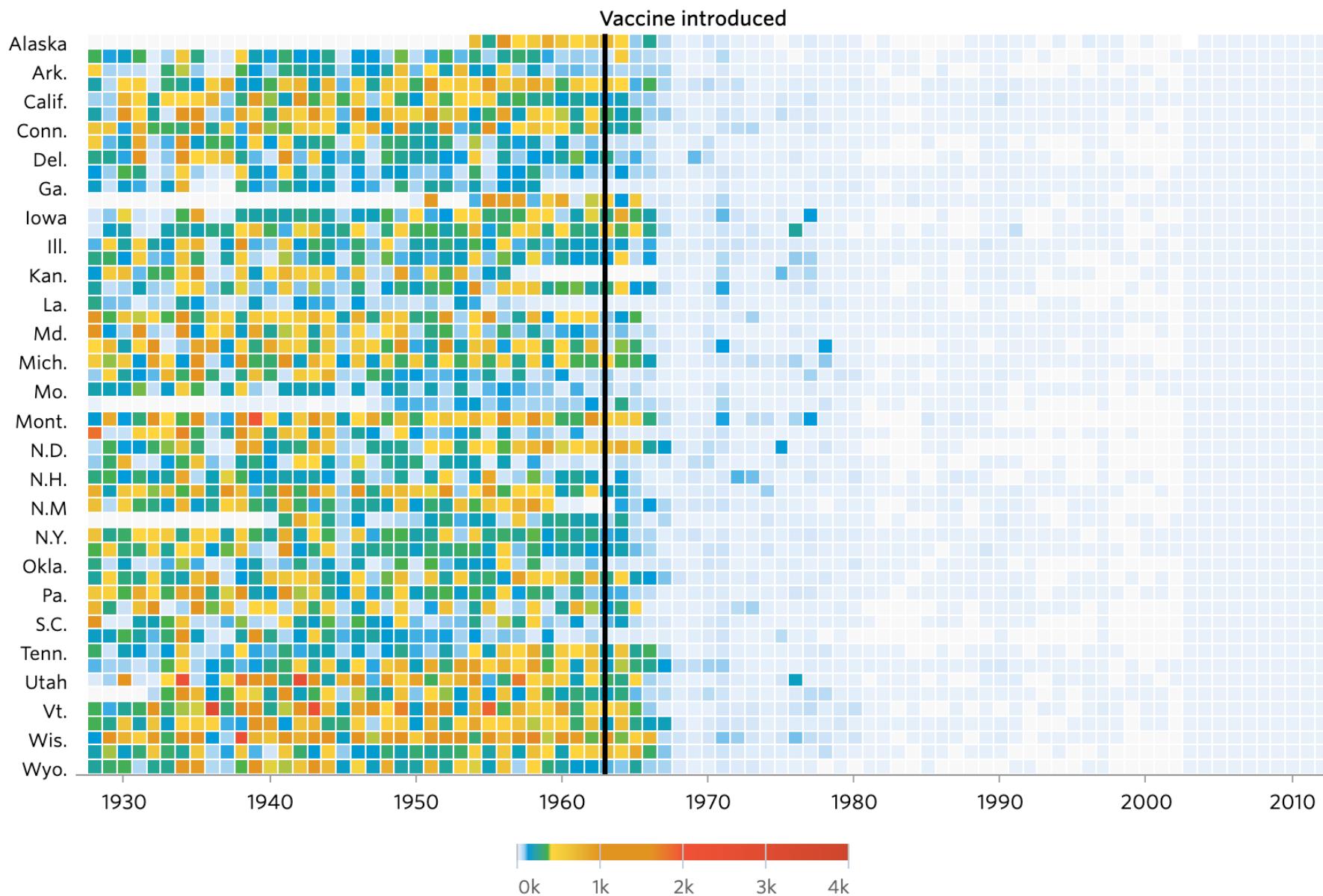


Heat map & calendar heat map



Seen something similar elsewhere?

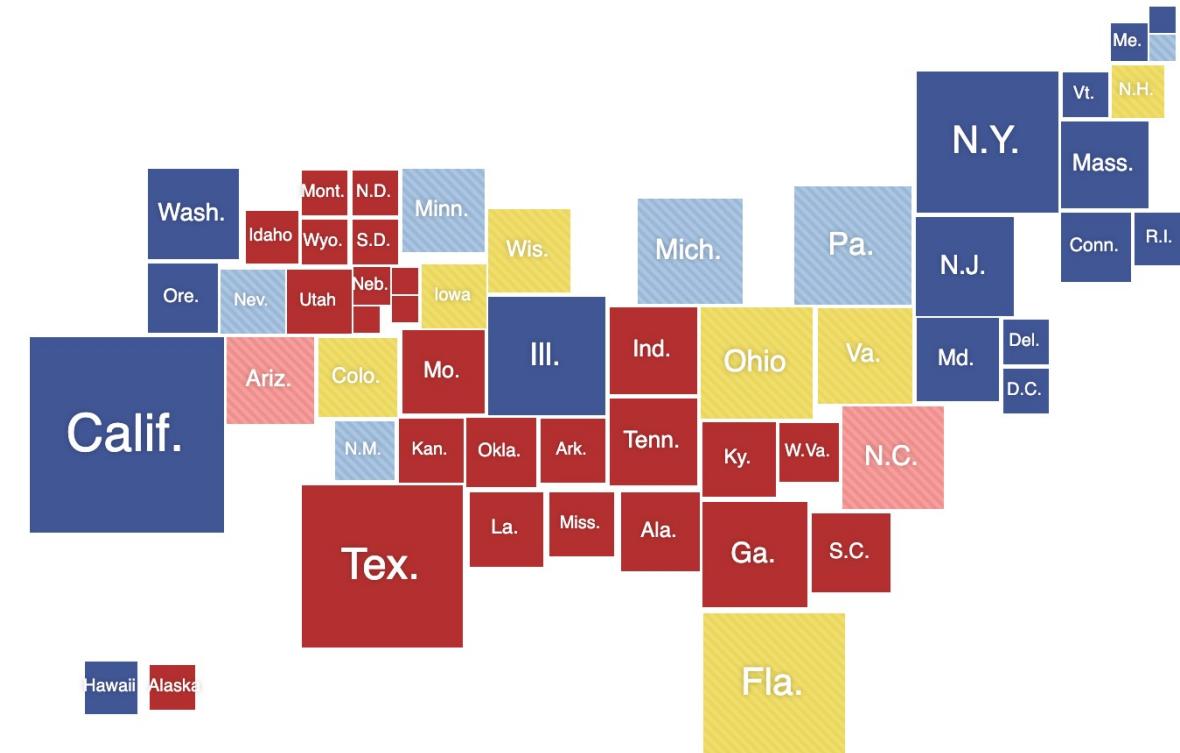
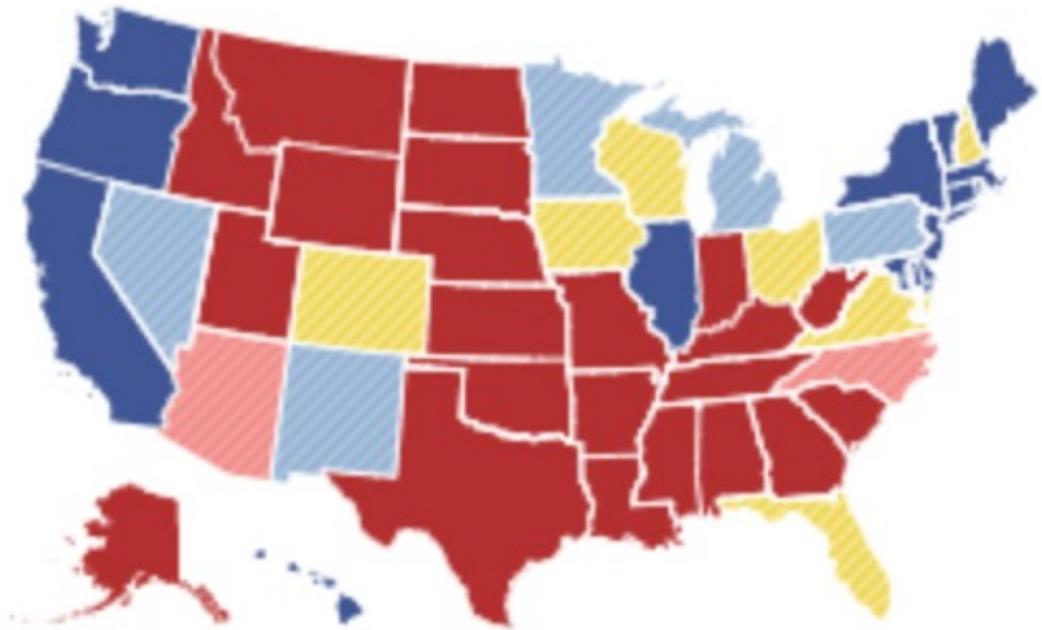
Measles



What if I have geospatial data?

...like lat/long

Maps & Cartograms

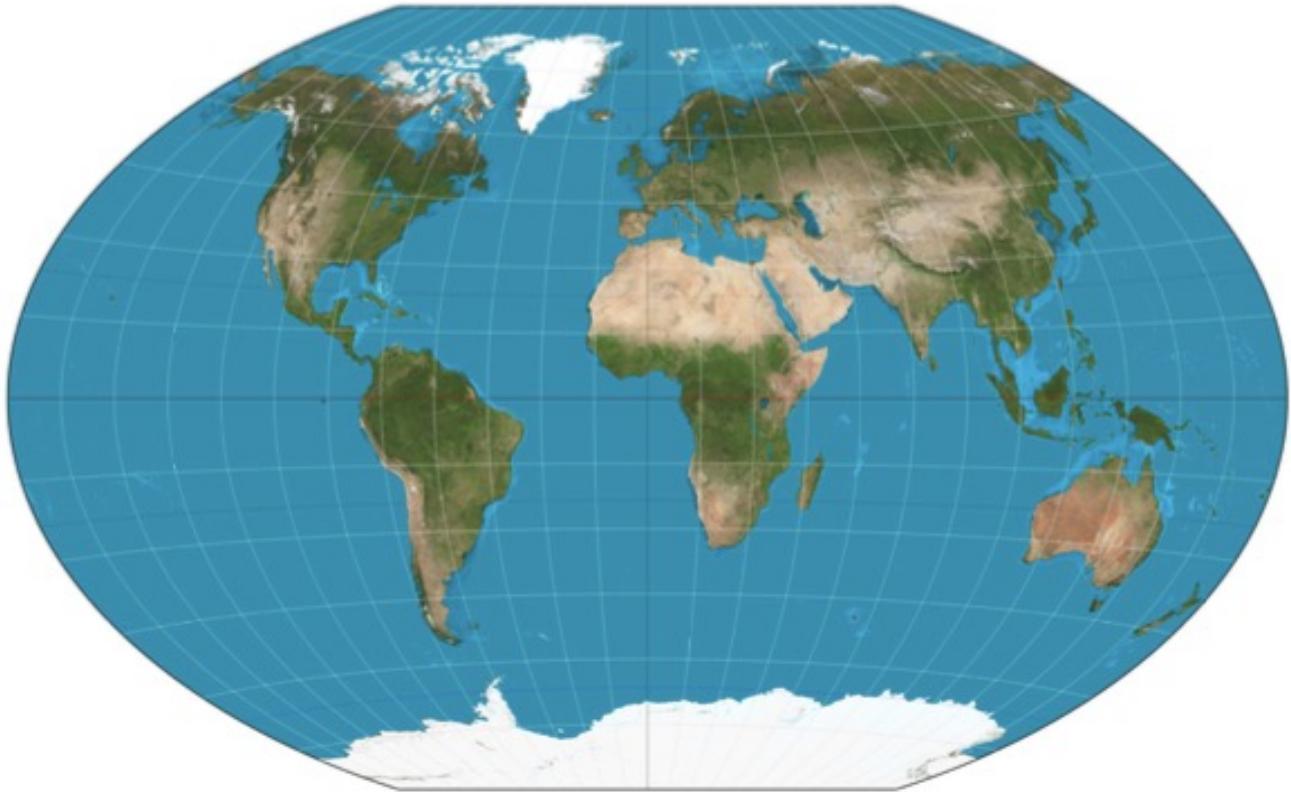


Map projection is a form of dimensionality reduction

3D

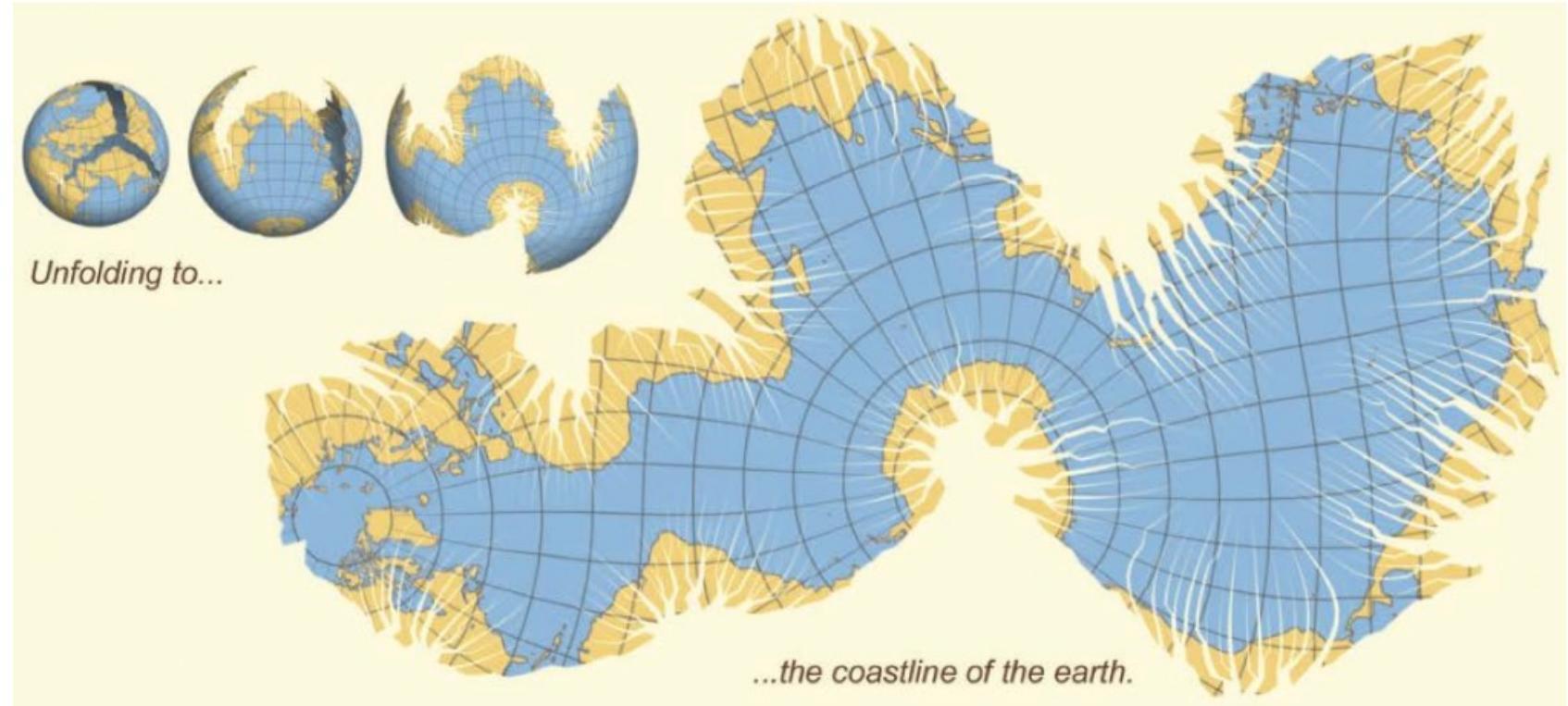


2D



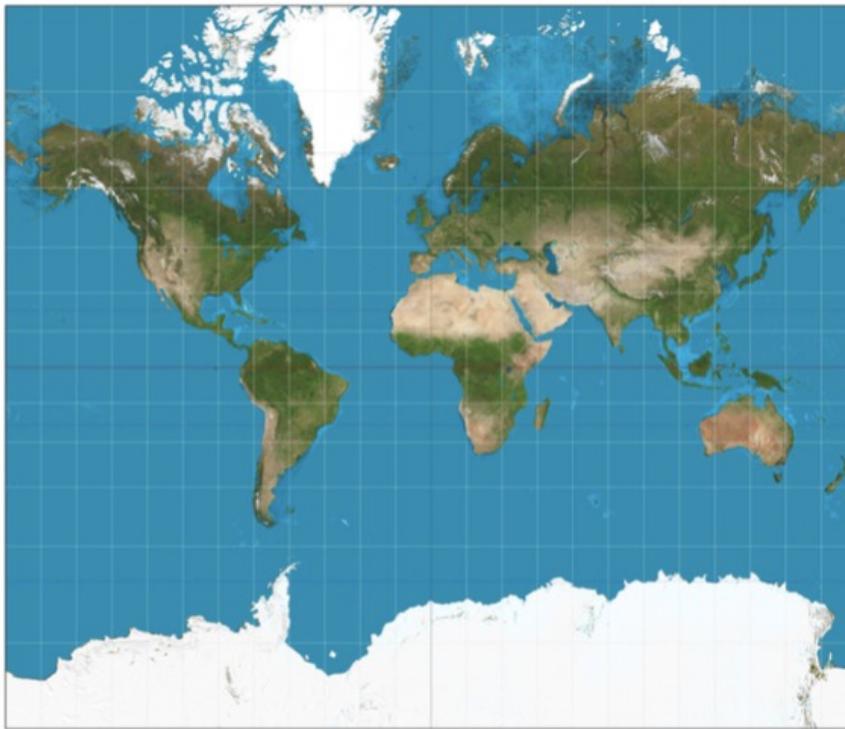
Different projections preserve different things

- Areas
- Shapes
- Directions
- Bearings
- Distances
- Scales

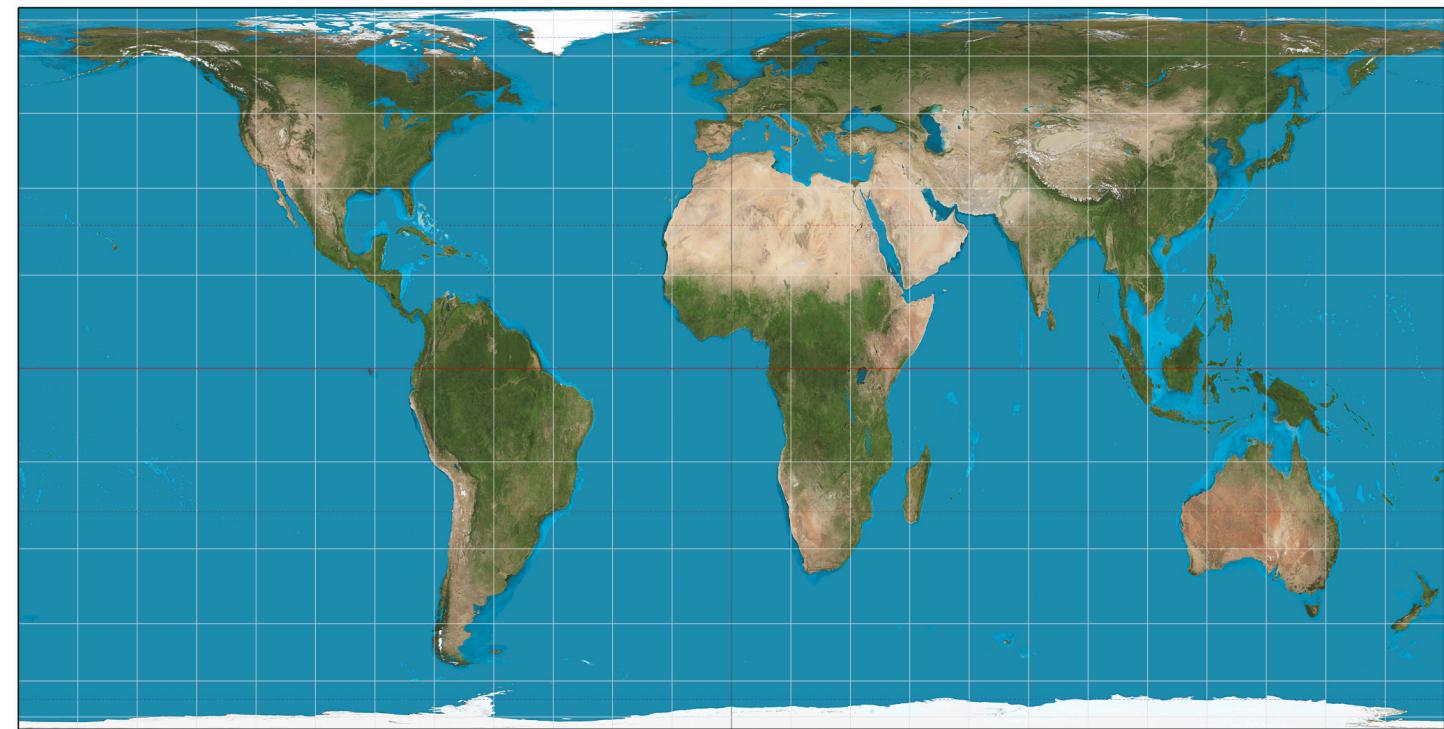


Which projection should I use?

Mercator

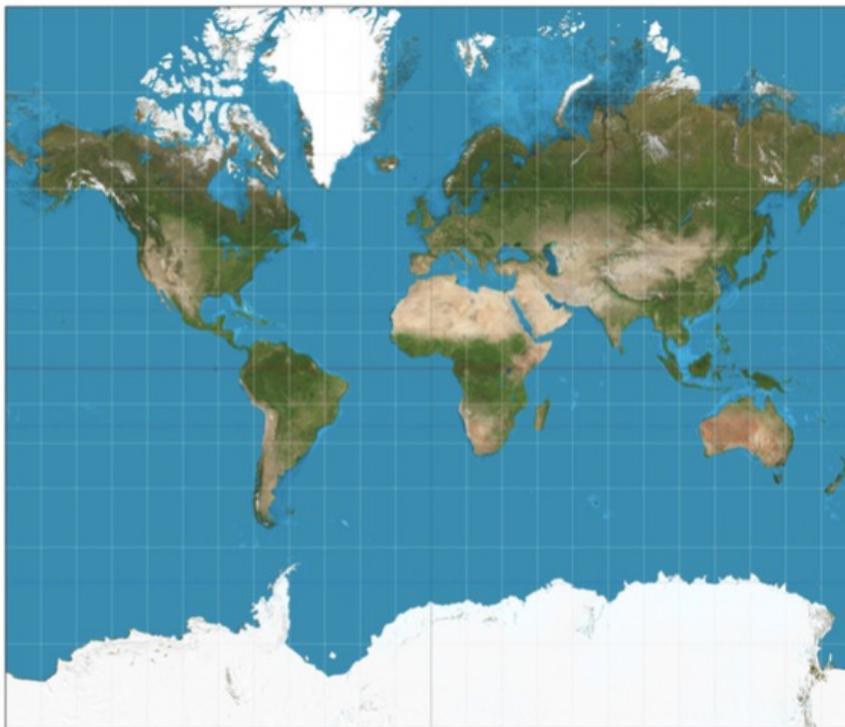


Hobo-Dyer



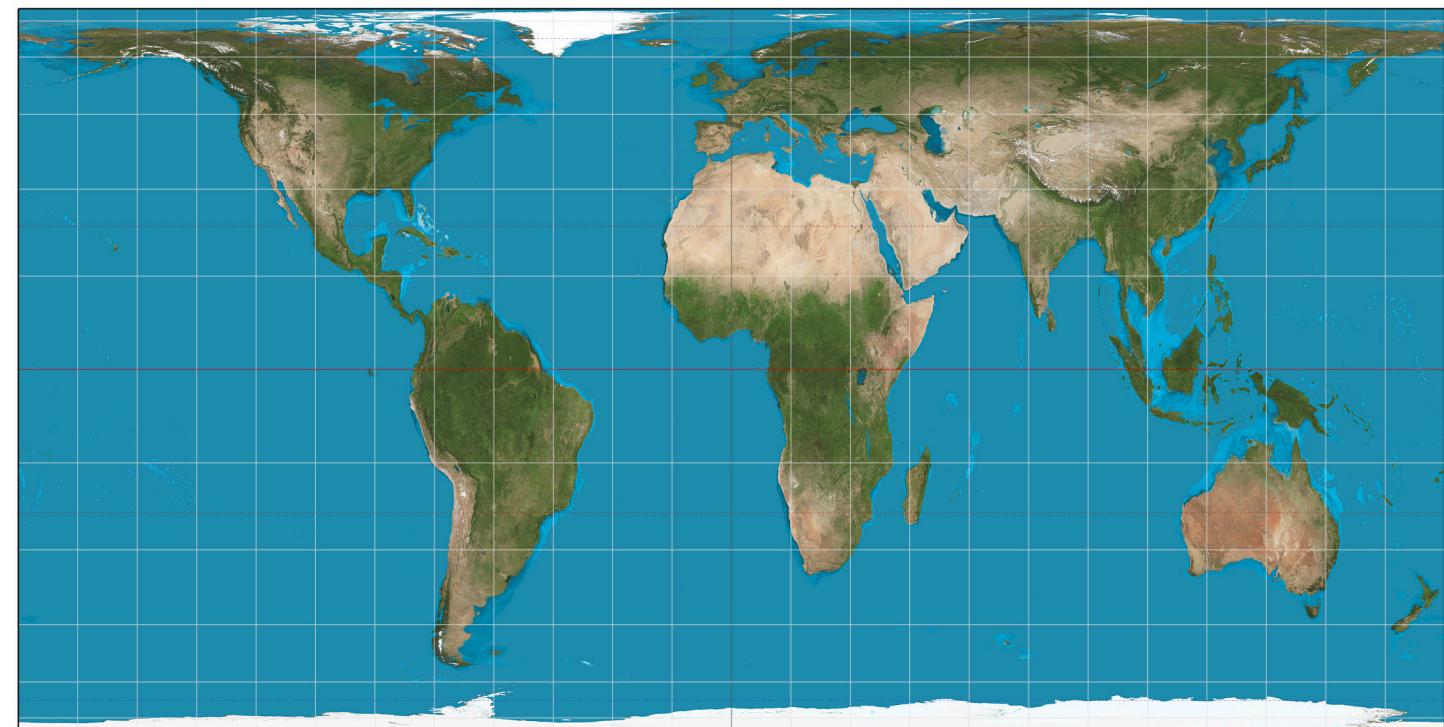
Which projection should I use?

Mercator



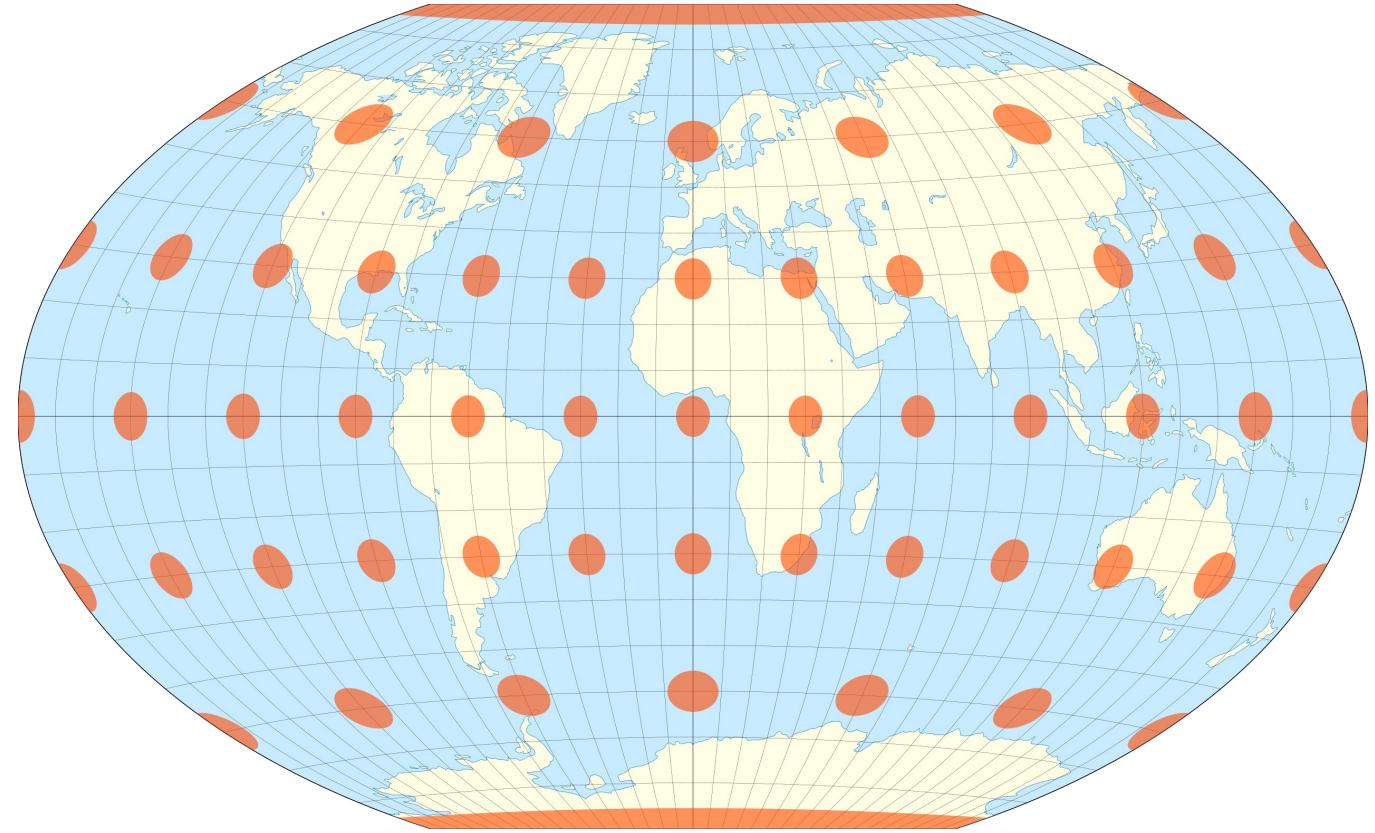
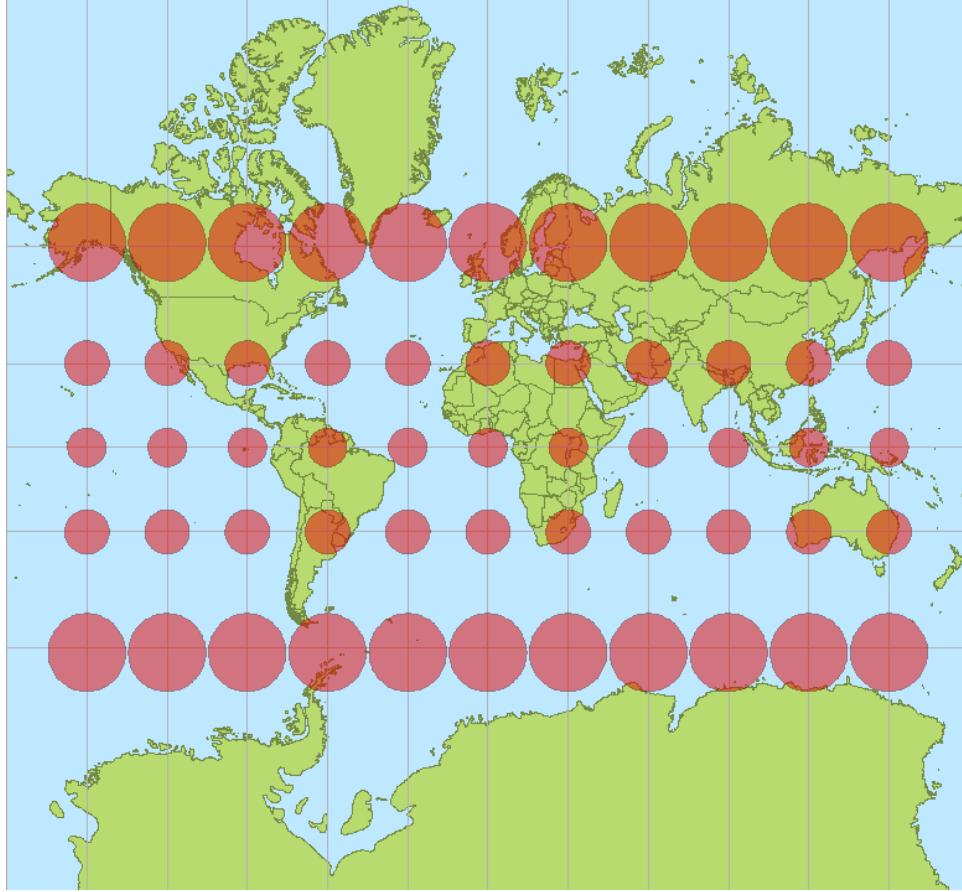
Navigation

Hobo-Dyer



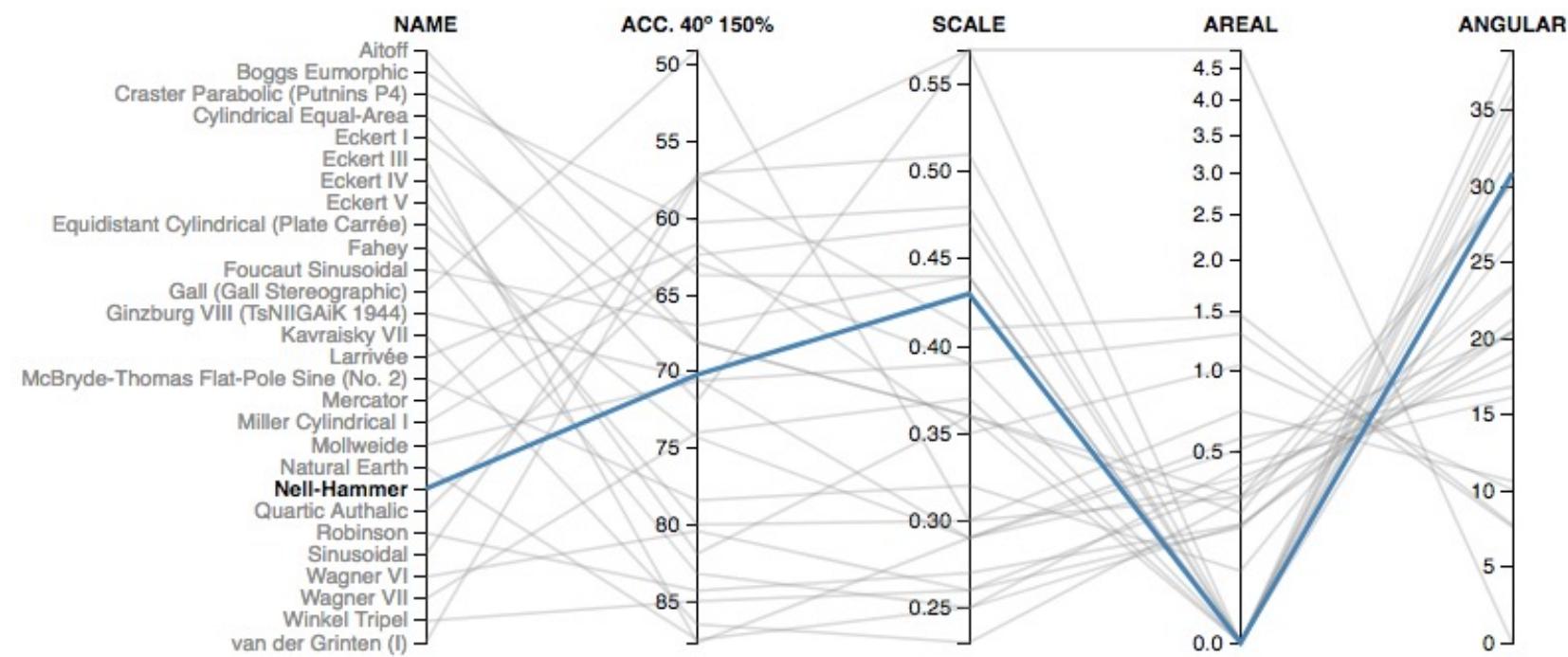
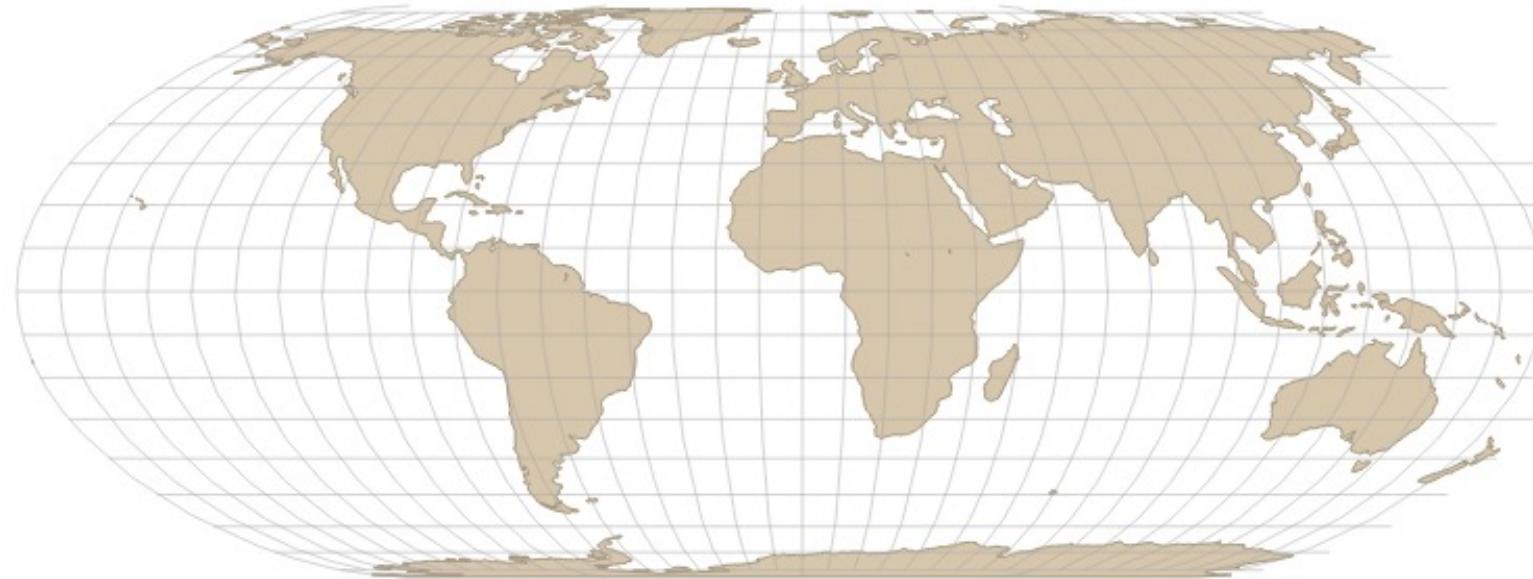
Area Comparison

Comparing Map Projections – Tissot's Indicatrix



See also: <https://observablehq.com/@d3/tissots-indicatrix>

Comparing Map Projections – Distortions



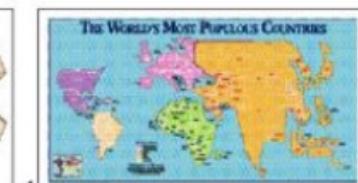
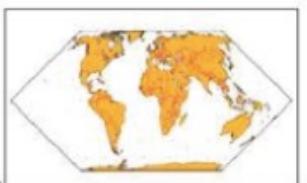
Projection Choices & Orientation Change What We See



The Hobo-Dyer Equal Area projection (above) shows each country's area at true proportion. This makes it fair to all countries. Other ways of seeing the world include:

- (1) Buckminster Fuller's Dymaxion Map,
- (2) the Eckert II projection,
- (3) Leonardo da Vinci's *mappamundi*, and
- (4) a population cartogram.

For more information on these images visit www.odt.org/hdp



ODT, Inc.
1-800-736-1293
www.odt.org
ISBN 1-881047-00-1

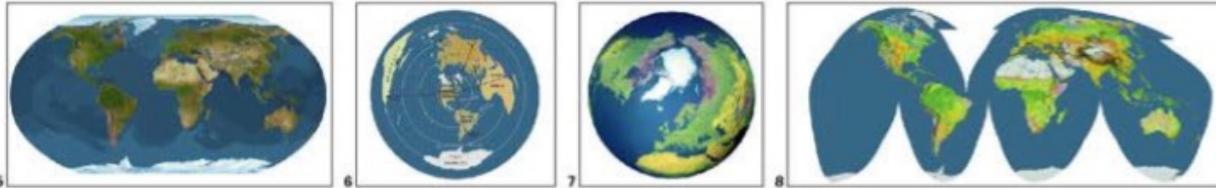
ID 781931 057004

Adapted from slides
of Josh Levine &
Miriah Meyer.
47

Projection Choices & Orientation Change What We See

Take the quiz! Compare country size.

Which of the images on both sides of this placemat are "area accurate?" How is the Hobo-Dyer projection below different from the one on the reverse side? Answers and details about all the images are at www.odt.org/hdp. To the right:
(5) Van Sant's Geosphere,
(6) Guelke's Toronto-centered projection,
(7) the Oxford Globe, and
(8) Goode's Homolosine



ISBN 1-931057-11-7

To order: ODT, Inc. 1-800-736-1293

Int'l Calls: 1-413-549-1293

www.odt.org

E-mail: odtstore@aol.com

Fax: 1-413-549-3503

Box 134, Amherst, MA 01004 USA

© 2002, ODT, Incorporated. All rights reserved.

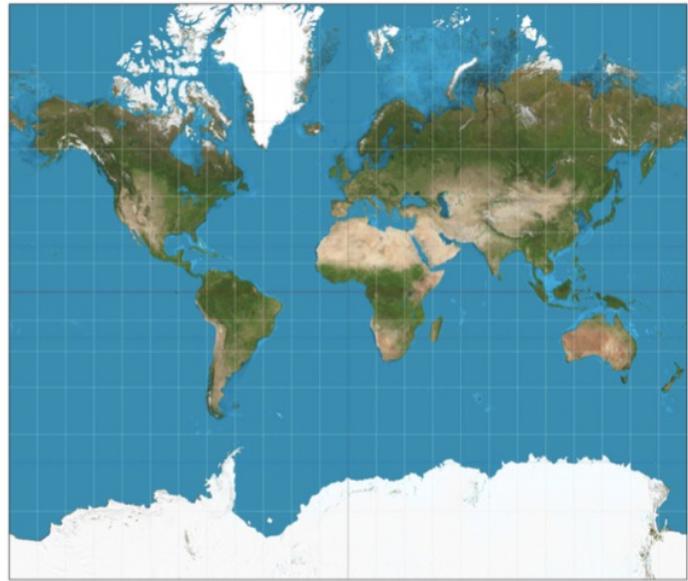


Adapted from slides
of Josh Levine &
Miriah Meyer.

How big is Africa?



How big is Africa?



The True Size of Africa

A small contribution in the fight against rampant Amnesia 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 <small>x 1000 km²</small>
China	9.587
USA	9.029
India	3.287
Mexico	1.964
Peru	1.265
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

In addition to the well known social issues of illiteracy and innumeracy, there also should be such a concept as "*inmappancy*", 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 "3-5 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!



Rank	Country	Area (sq km)	% of World Total
1	Russia	17,082,462	11.9%
2	Canada	9,980,470	6.7%
3	China	9,597,561	6.4%
4	United States	9,491,351	6.4%
5	Brazil	8,843,977	5.7%
6	Australia	7,970,524	5.2%
7	India	3,297,285	2.3%
8	Argentina	2,790,460	1.9%
9	Kazakhstan	2,754,849	1.8%
10	Ukraine	2,731,575	1.7%
11	Algeria	2,511,741	1.6%
12	Greece	2,344,050	1.5%
13	Greenland	2,165,080	1.4%
14	Saudi Arabia	2,140,050	1.4%
15	Russia (Asia)	1,924,375	1.3%
16	Indonesia	1,865,360	1.3%
17	Uganda	1,750,549	1.2%
18	Iran	1,620,769	1.1%
19	Kenya	1,580,000	1.1%
20	France	1,520,219	1.0%
21	China	1,391,200	0.9%
22	Niger	1,292,000	0.8%
23	Angola	1,280,760	0.8%
24	Mali	1,200,182	0.8%
25	South Africa	1,151,037	0.8%
26	Cambodia	1,151,204	0.7%
27	Malta	1,024,869	0.7%
28	Bolivia	1,014,474	0.7%
29	Argentina (Asia)	1,008,000	0.6%
30	Morocco	970,060	0.6%
31	Tunisia	970,067	0.6%
32	Nigeria	923,788	0.6%
33	Honduras	924,259	0.6%
34	Kenya (Asia)	884,118	0.5%
35	Morocco (Africa)	801,559	0.54%
36	Pakistan	758,059	0.53%
37	Uganda (Africa)	758,192	0.51%
38	Zambia	758,412	0.51%
39	Myanmar	758,278	0.51%
40	Egypt	757,229	0.51%
41	Albania	752,019	0.44%
42	Venezuela	747,447	0.48%
43	France (Africa)	733,050	0.45%
44	G. African Rep.	662,259	0.42%
45	Uruguay	653,919	0.41%
46	Maldives	657,011	0.40%
47	Guatemala	640,220	0.38%
48	Kosovo	639,287	0.38%
49	Yemen	621,058	0.35%
50	Thailand	612,729	0.38%
51	Yemen (Asia)	608,052	0.38%
52	Indonesia (Asia)	598,118	0.33%
53	Cameroon	576,412	0.32%
54	New Caledonia	482,463	0.31%
55	Uzbekistan	511,019	0.31%
56	Malta (Europe)	508,224	0.31%
57	Barbados	507,279	0.30%
58	Ang	506,217	0.30%
59	Panama	498,750	0.27%
60	Zimbabwe	506,757	0.25%
61	Japan	507,838	0.25%
62	Germany	507,114	0.24%
63	Rep. of Congo	242,059	0.22%
64	Poland	326,118	0.19%
65	Ukraine	301,212	0.20%
66	Malta (Asia)	300,000	0.19%
67	Norway	325,007	0.17%
68	Ethiopia	322,493	0.22%
69	Poland (Asia)	274,959	0.21%
70	Dominican Republic	318,008	0.21%
71	Italy	257,259	0.20%
72	Philippines	320,000	0.22%
73	Switzerland	204,072	0.18%
74	New Zealand	225,467	0.18%
75	Uganda (Europe)	204,000	0.18%
76	Western Sahara	256,000	0.18%
77	Russia	256,299	0.20%
78	Qatar	268,897	0.17%
79	United Kingdom	252,955	0.18%
80	Uganda (Africa)	267,028	0.18%
81	Shane	258,558	0.18%
82	Monaco	258,291	0.18%
83	Lux	258,000	0.18%
84	Qatar (Asia)	257,000	0.18%
85	Malta (Europe)	257,000	0.18%
86	Kyrgyzstan	258,251	0.18%
87	Senegal	136,722	0.13%
88	Sierra Leone	136,188	0.17%
89	Centralville	151,055	0.12%
90	Uruguay (Asia)	176,225	0.12%
91	Suriname	105,879	0.11%
92	Russia (Asia)	105,078	0.11%
93	Honduras	145,059	0.11%
94	Georgian Republic	145,059	0.11%
95	Taiwan	105,159	0.11%
96	Greece	121,087	0.09%
97	Nicaragua	136,079	0.09%
98	Iran & Korea	128,038	0.08%
99	Malta (Asia)	136,001	0.08%
100	Greece (Asia)	117,029	0.06%
	TOTAL	133,659,814	99.24%



United States



Europe



India

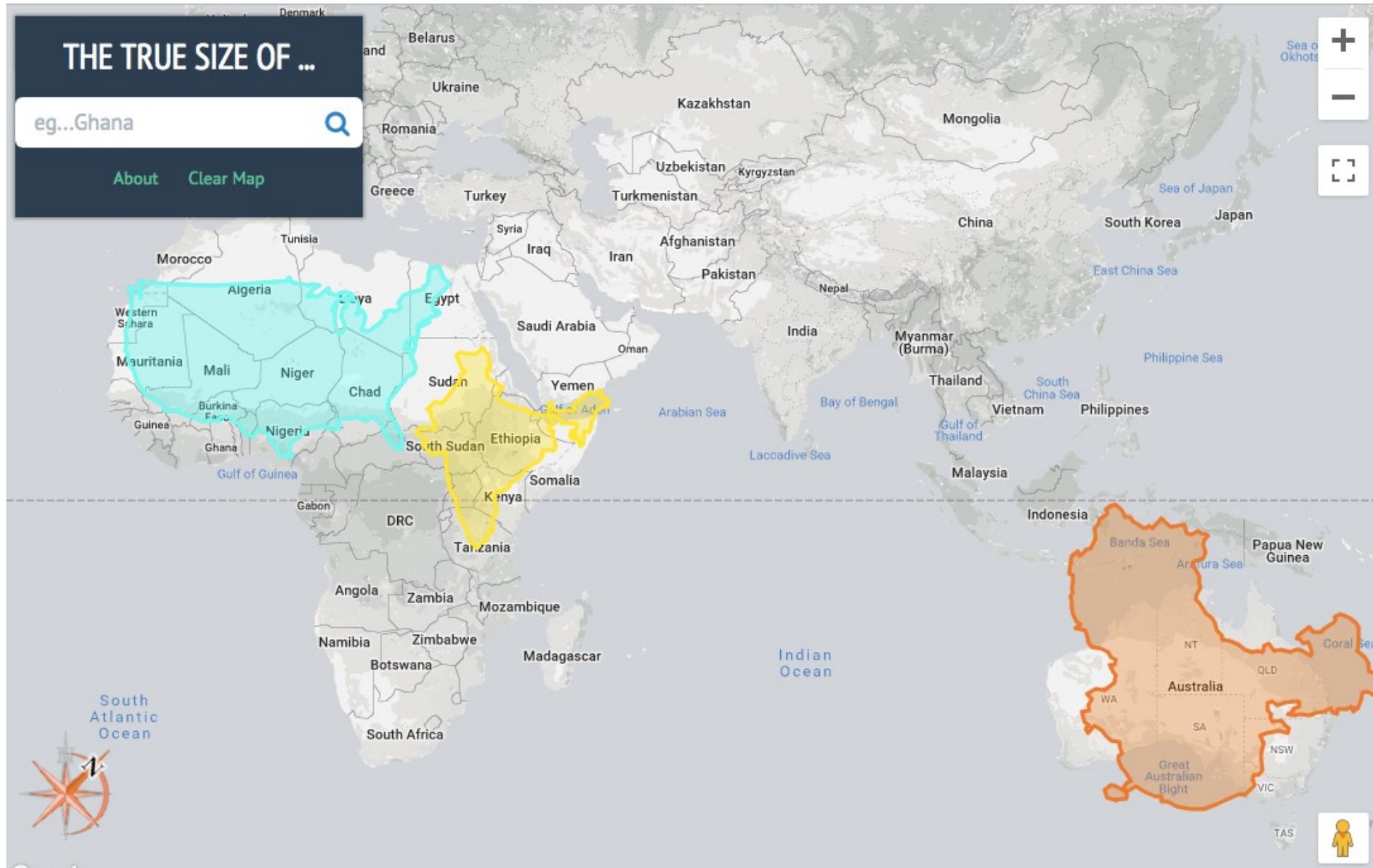


Japan

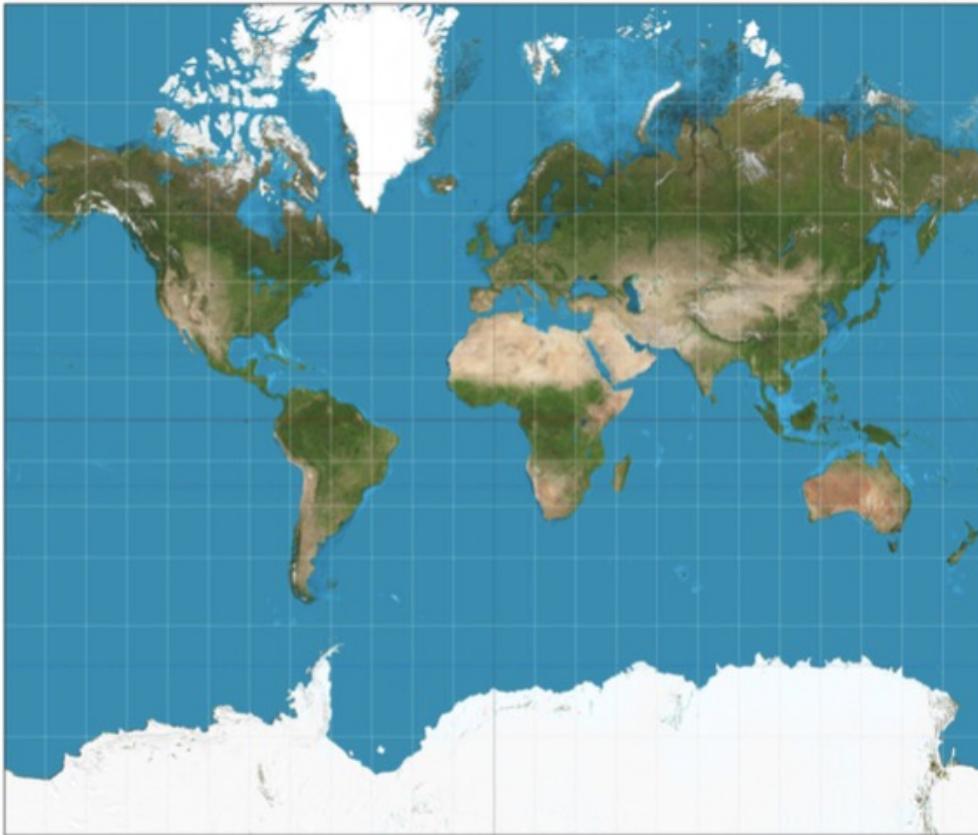


China

Adapted from slides of
Michelle Borkin,
<https://www.sciencealert.com/this-is-the-true-size-of-africa>



How big is everything else?



World Mercator projection with true country size and shape added



Visualizing Data on Maps – Points

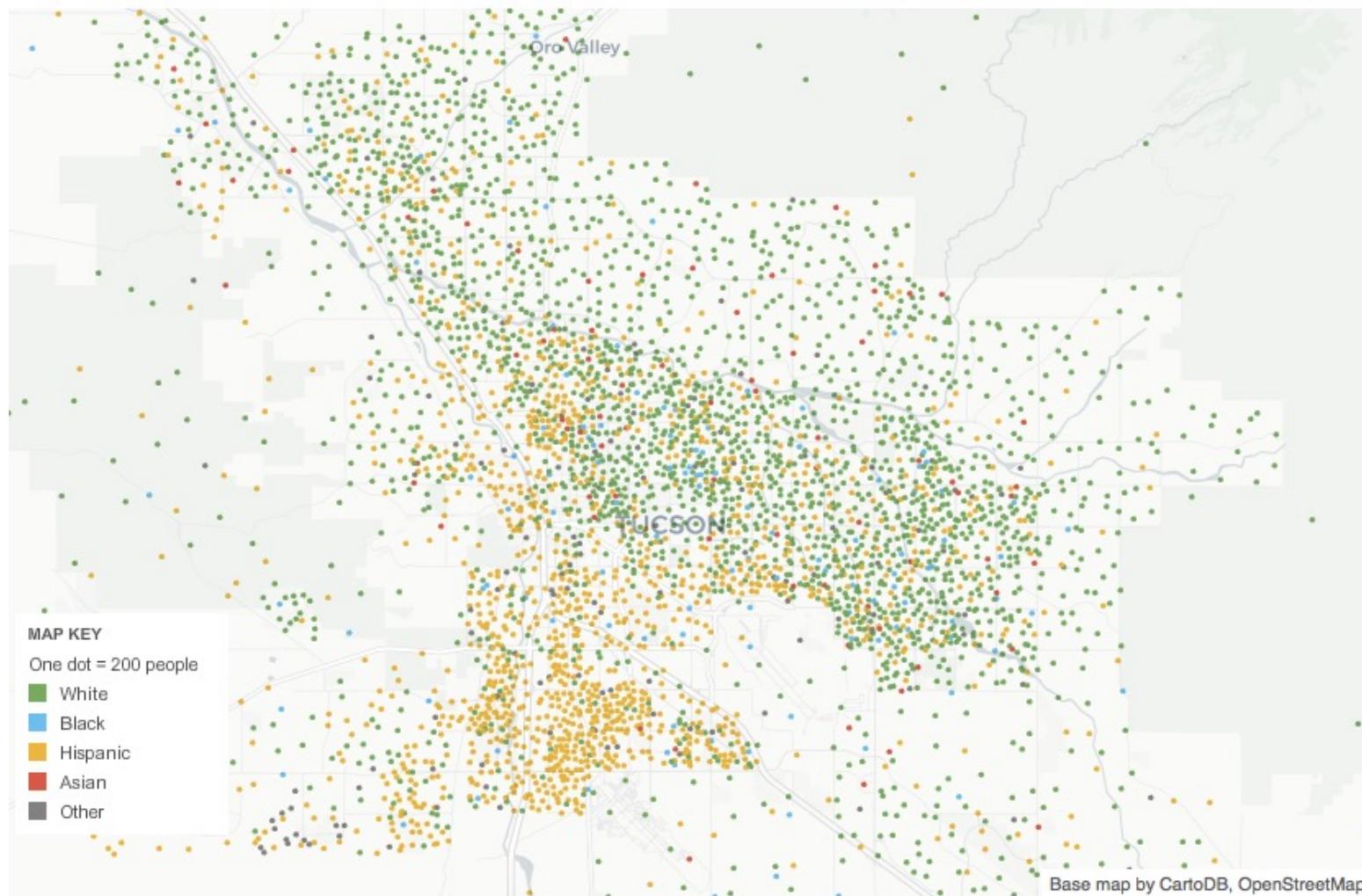
The New York Times

Mapping America: Every City, Every Block

Browse local data from the Census Bureau's American Community Survey, based on samples from 2005 to 2009. Because these are samples, they are subject to a margin of error, particularly in places with a low population, and are best regarded as estimates.

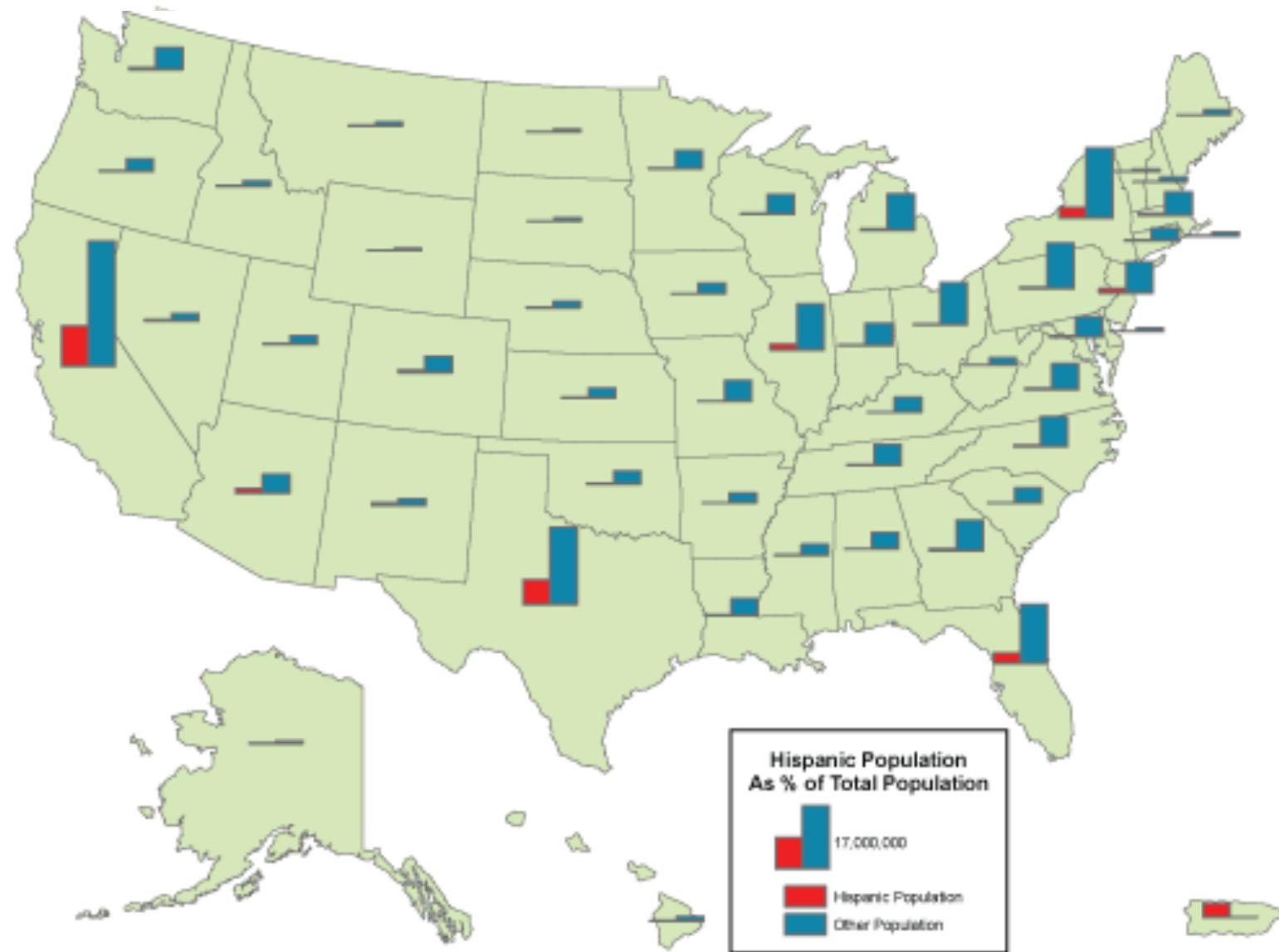
Distribution of racial and ethnic groups

[View More Maps](#)

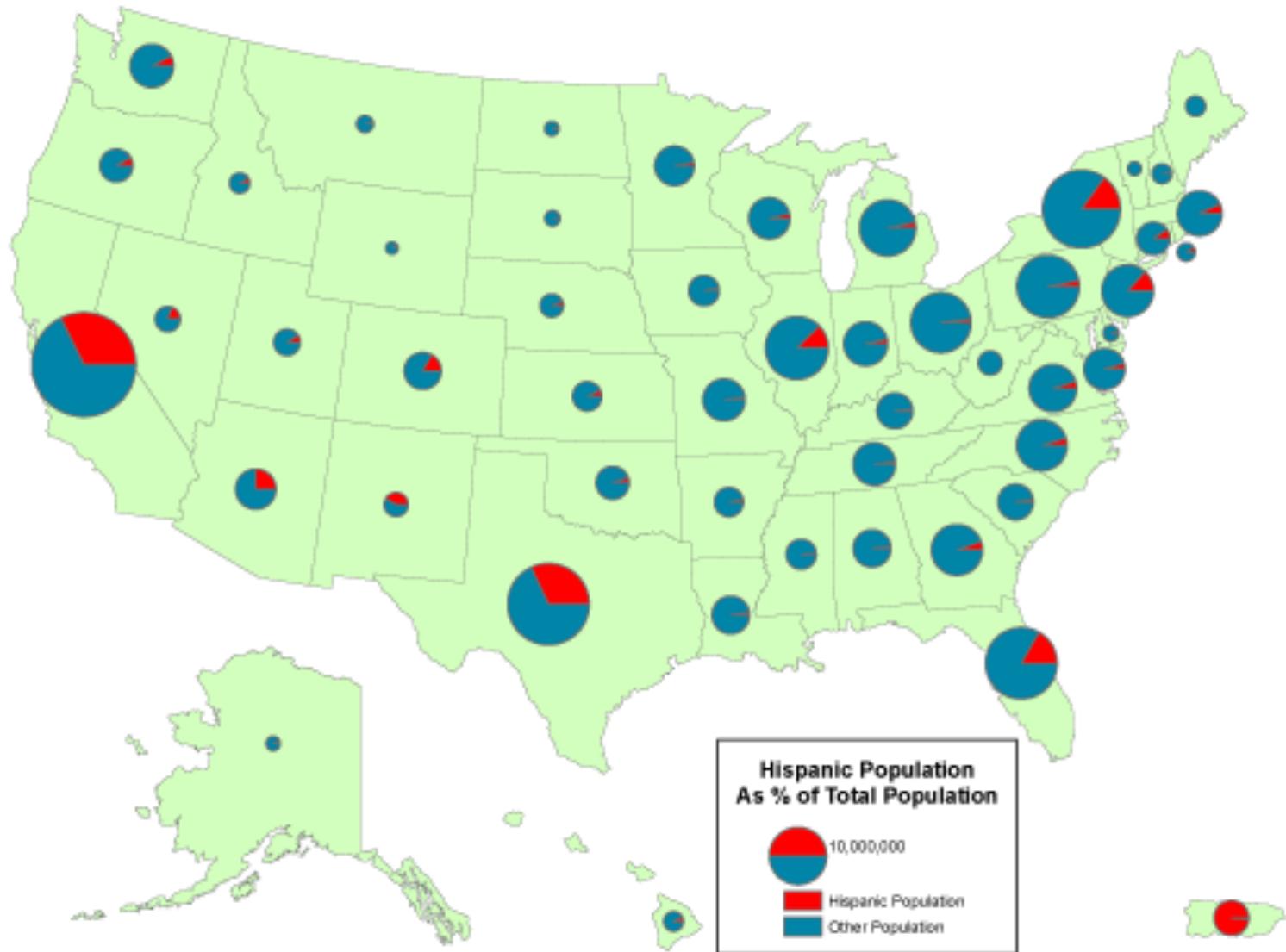


Adapted from slides of Meyer and Levine, See also
<http://racialdotmap.demographics.coopercenter.org/>

Visualizing Data on Maps – Charts

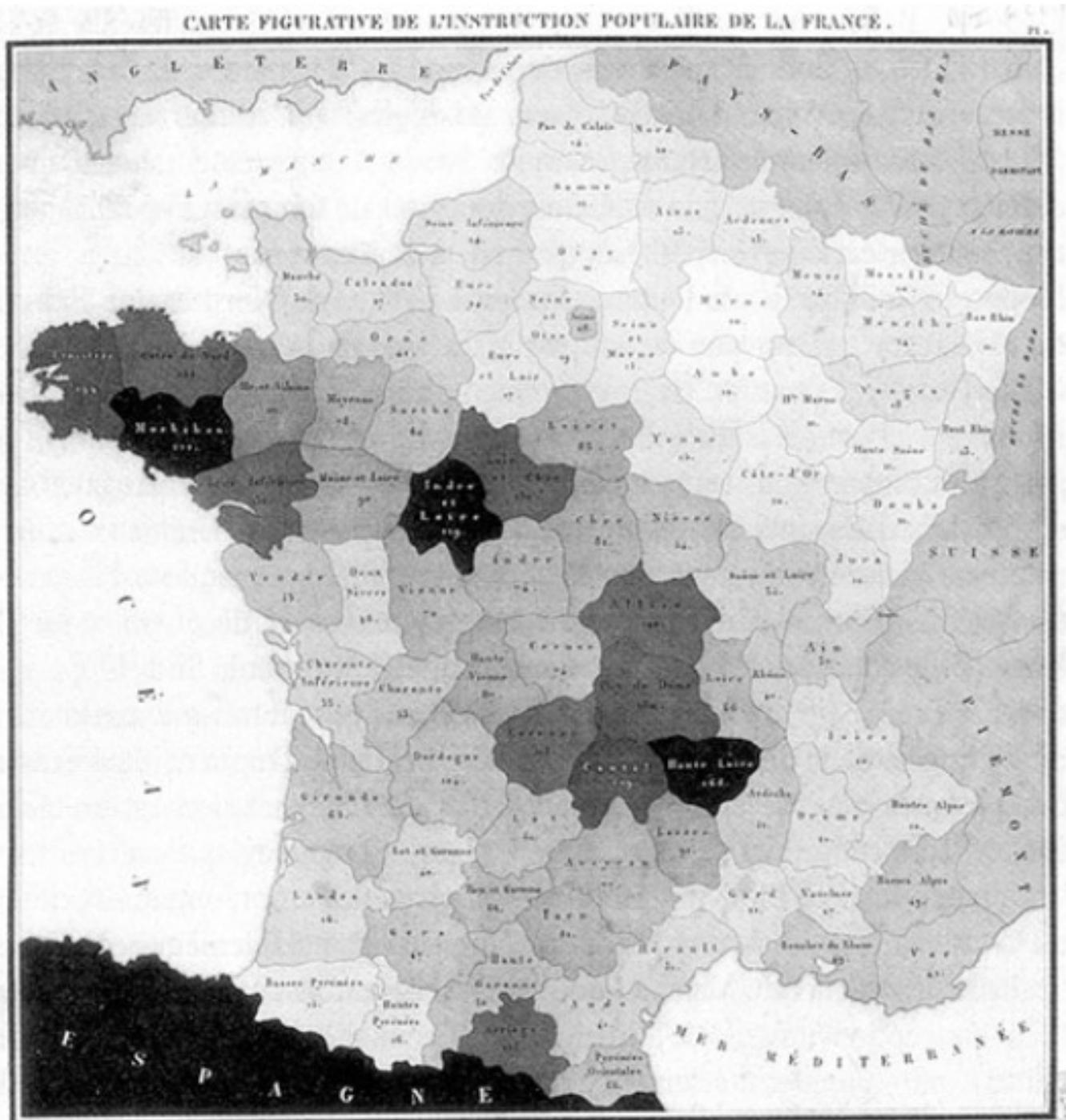


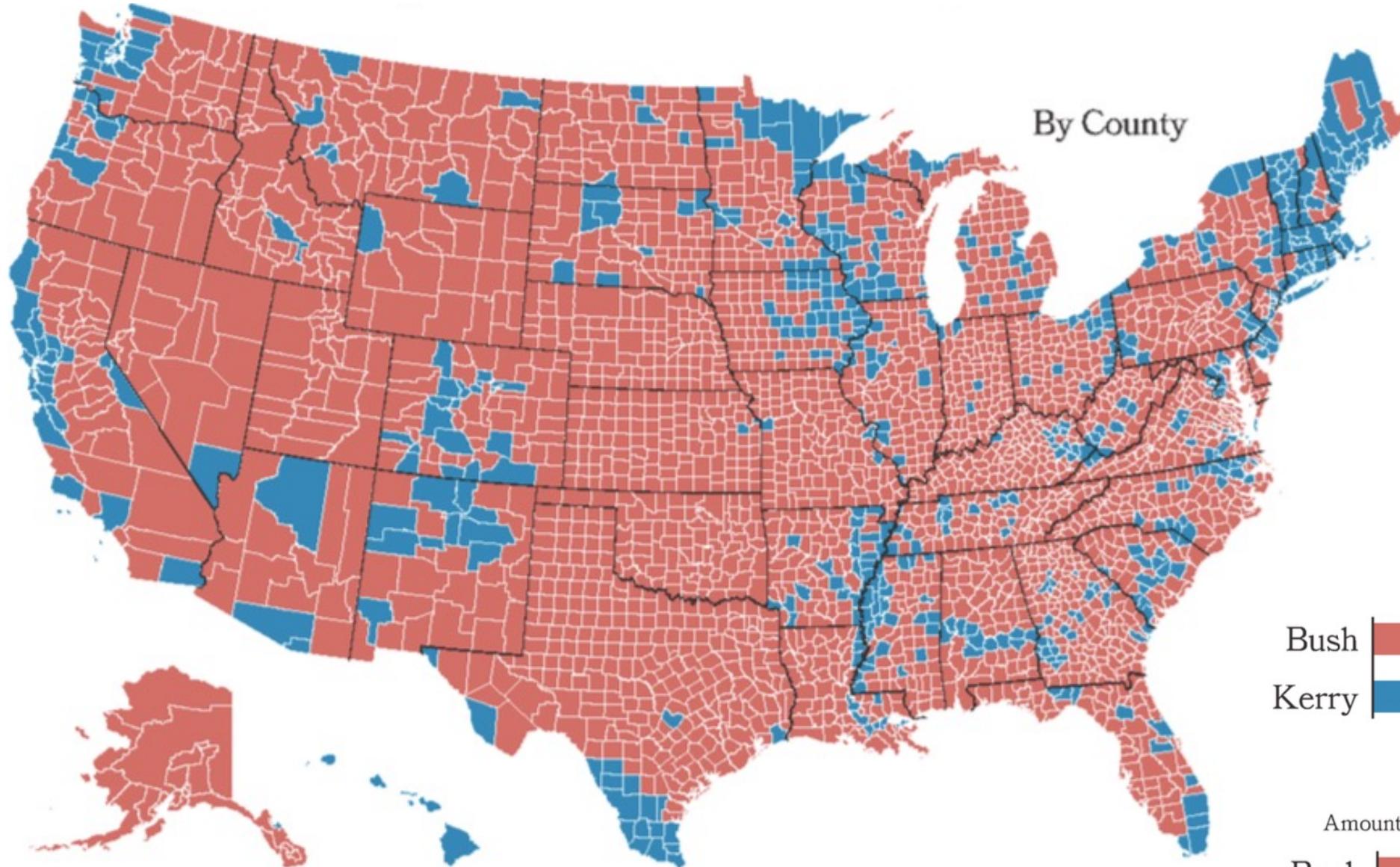
Visualizing Data on Maps – Charts



Choropleth maps color, shade, or fill areas to represent some (aggregated) data attribute

Illiteracy in France,
Dupin 1826





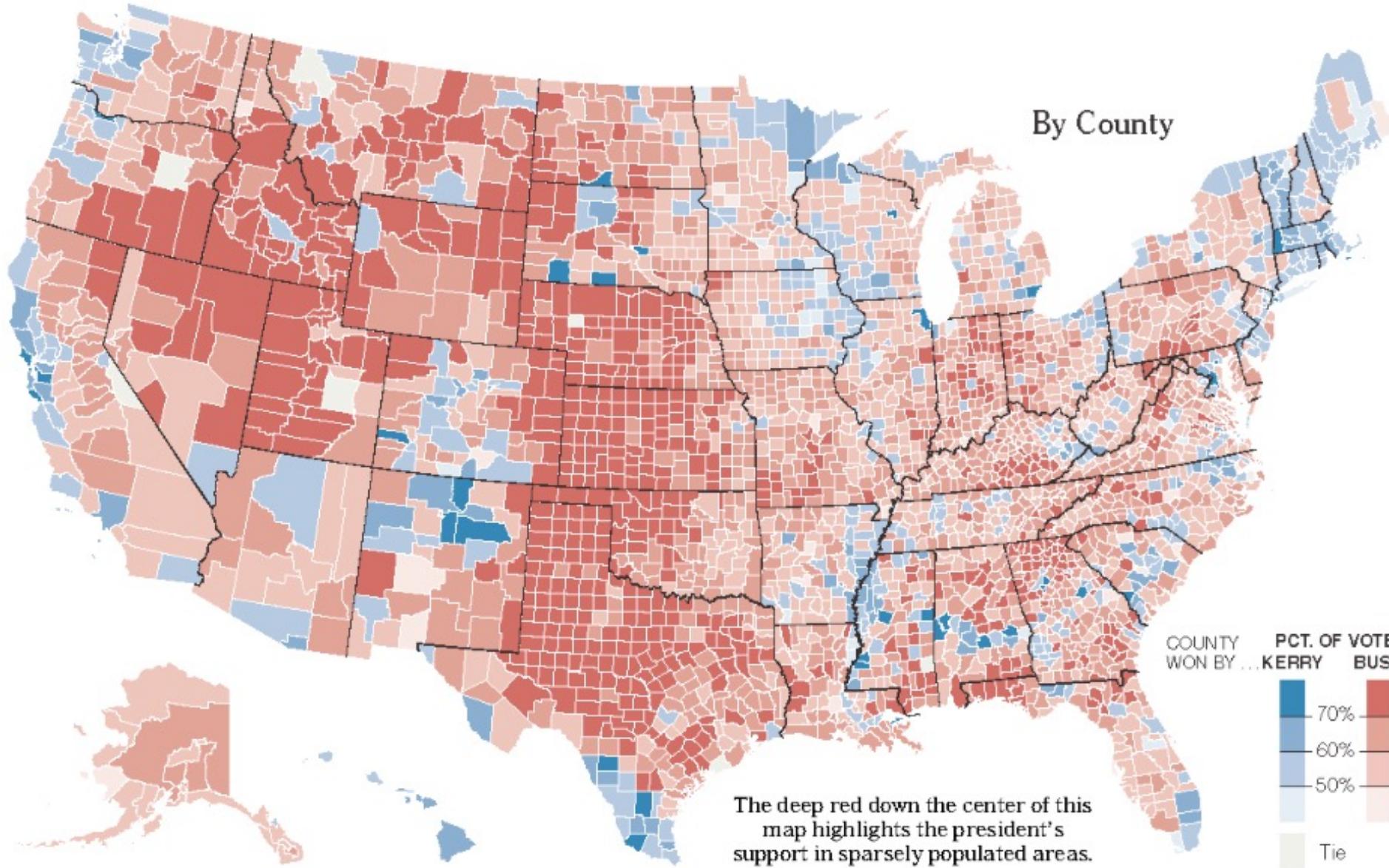
Bush v. Kerry, 2004 Election

2004 Popular Vote

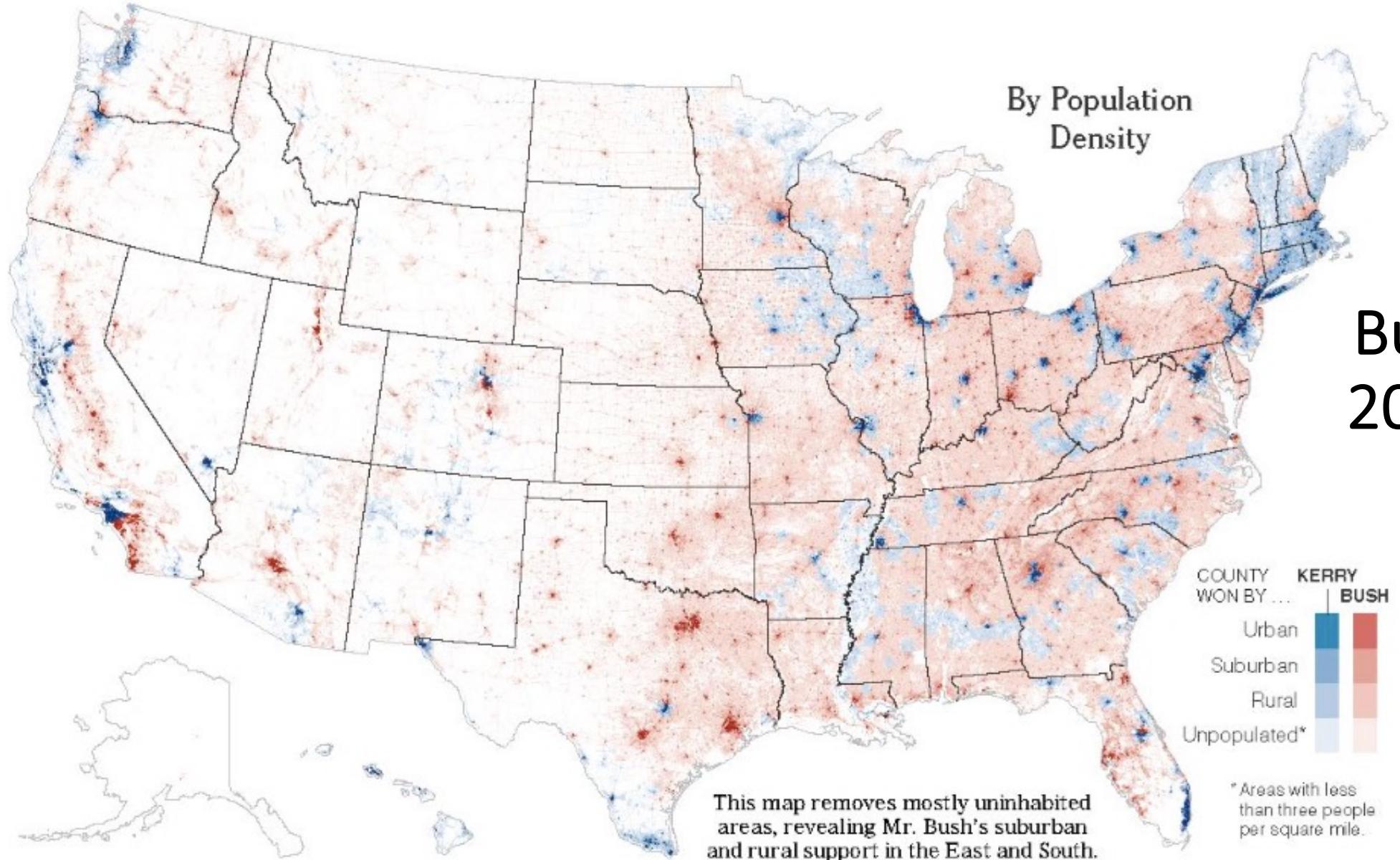


Amount of red and blue shown on map

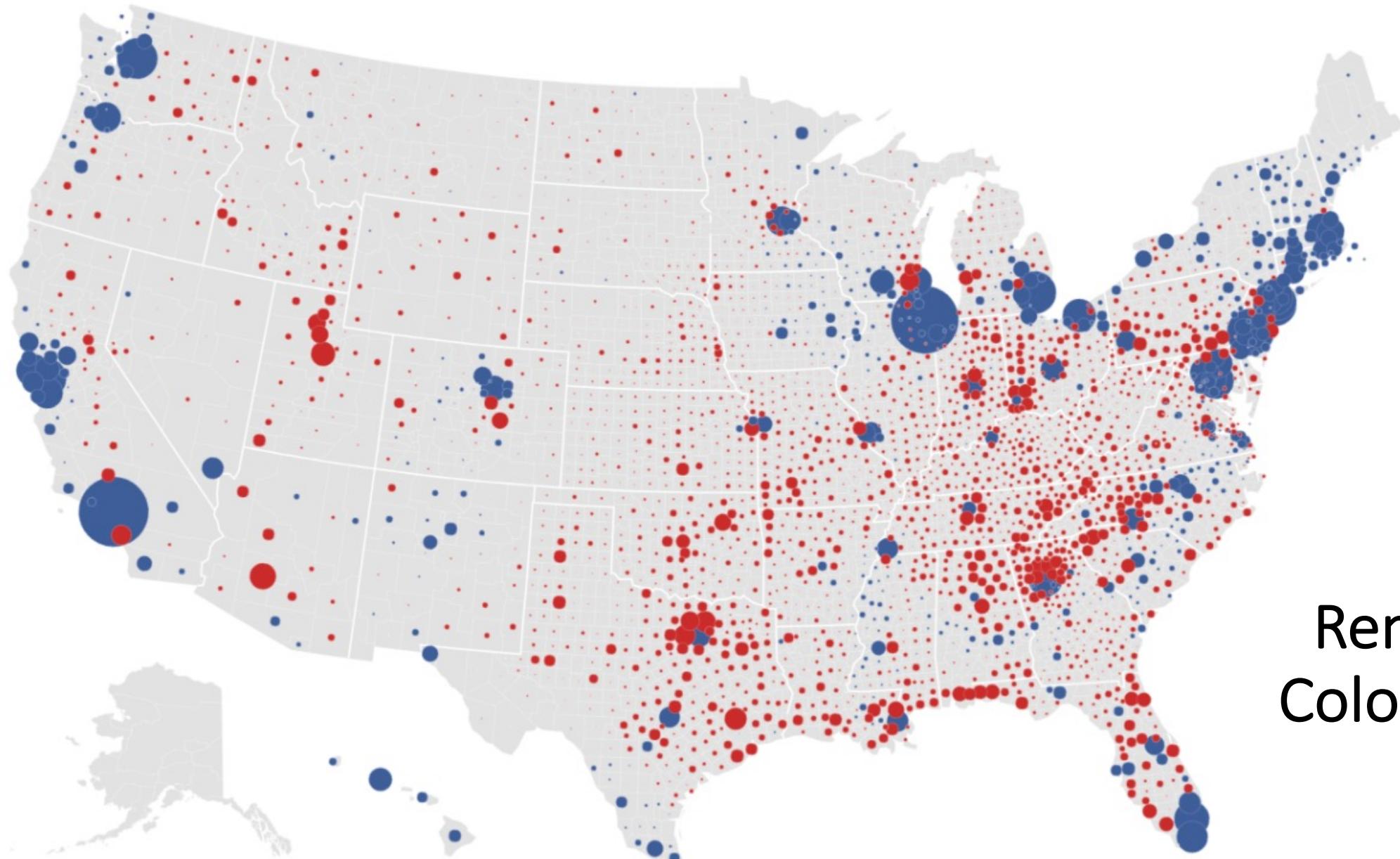




Bush v. Kerry, 2004 Election



Bush v. Kerry, 2004 Election



Remember:
Color is hard...

Cartograms scale and distort areas to encode an attribute

Land areas of countries,
Emile Levasseur 1868,
Rectangular cartogram

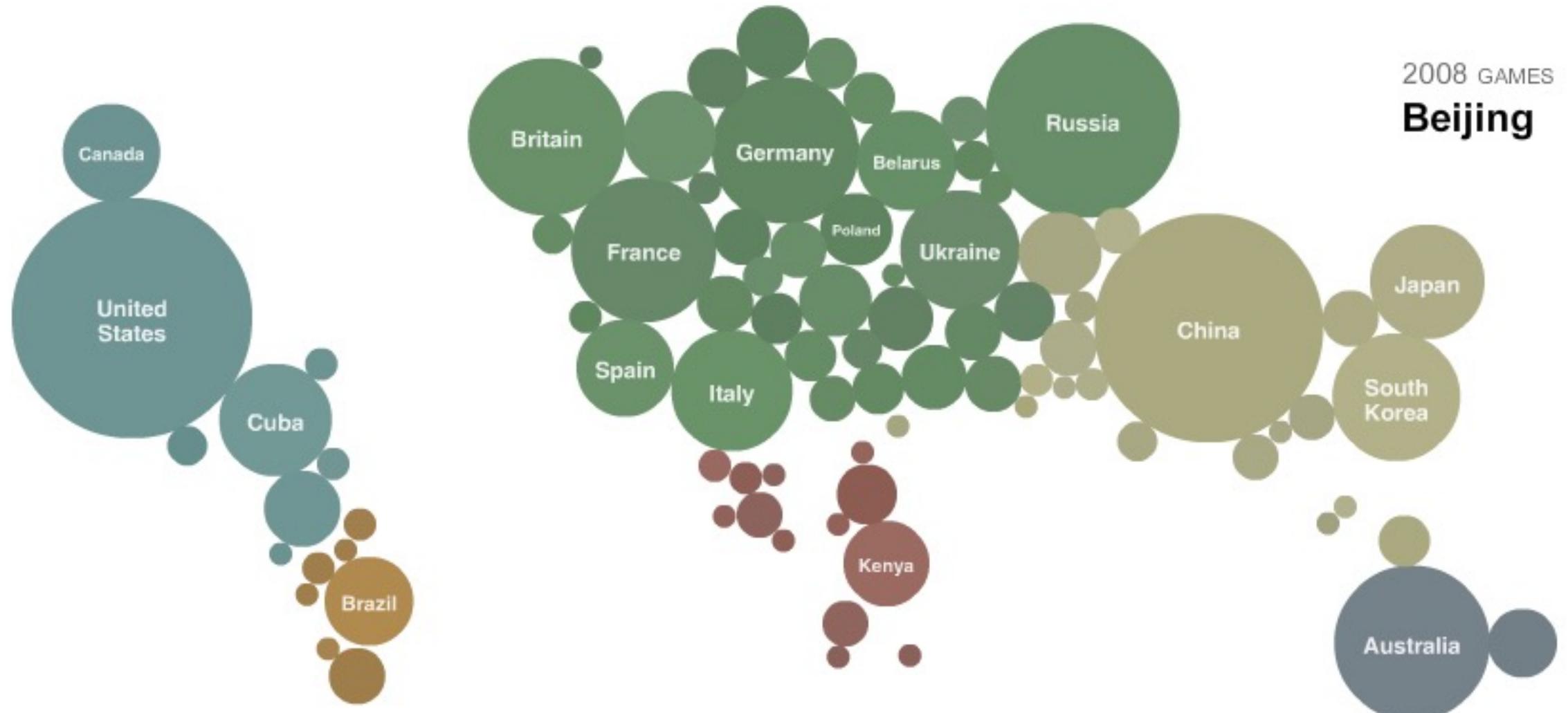
STATISTIQUE FIGURATIVE

SUPERFICIE

1 millimètre carré représente
955 kilomètres carrés



Dorling cartogram of Olympic medal count



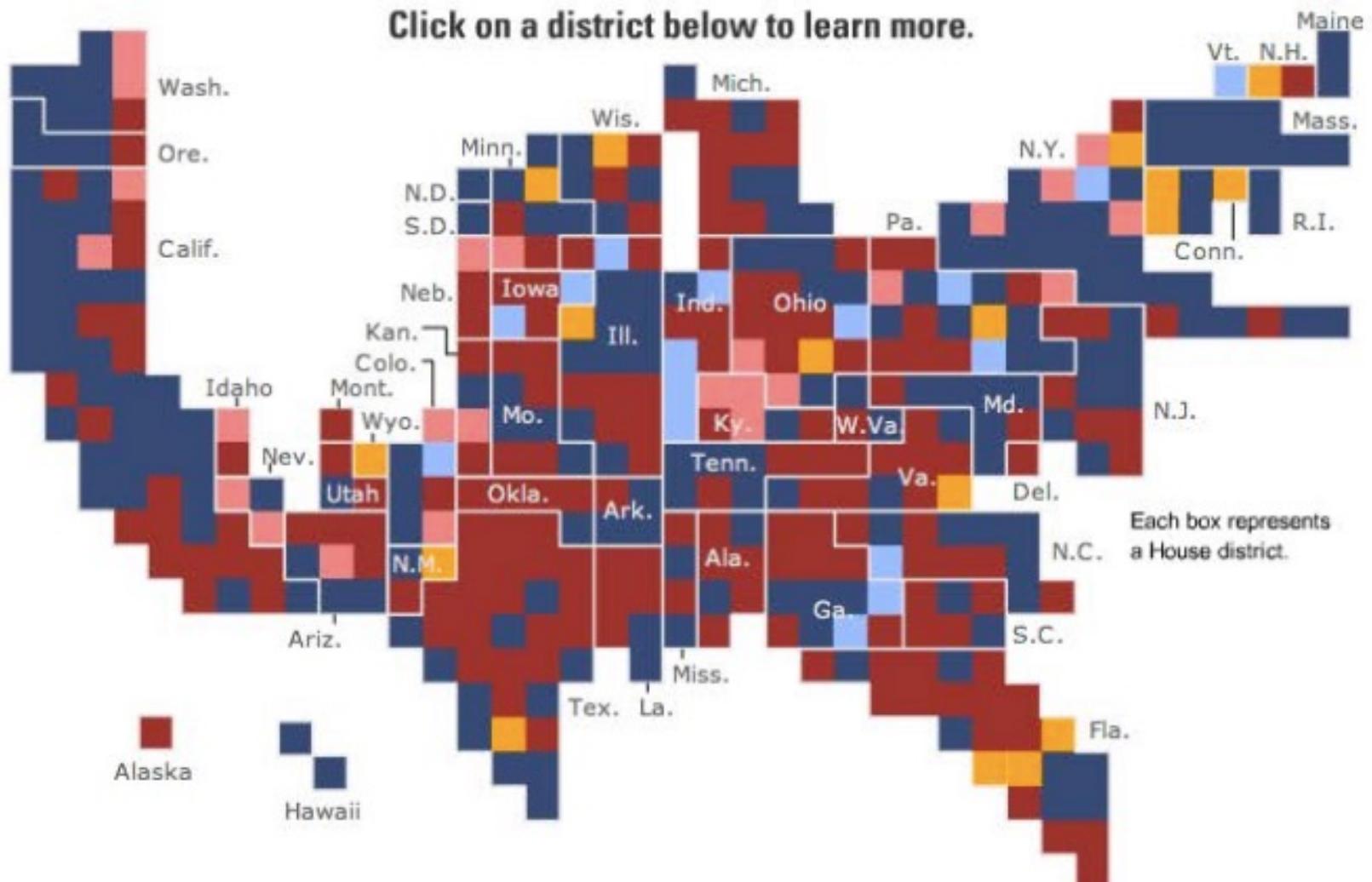
Rectangular cartogram of by-county election predictions

2006 ELECTION GUIDE

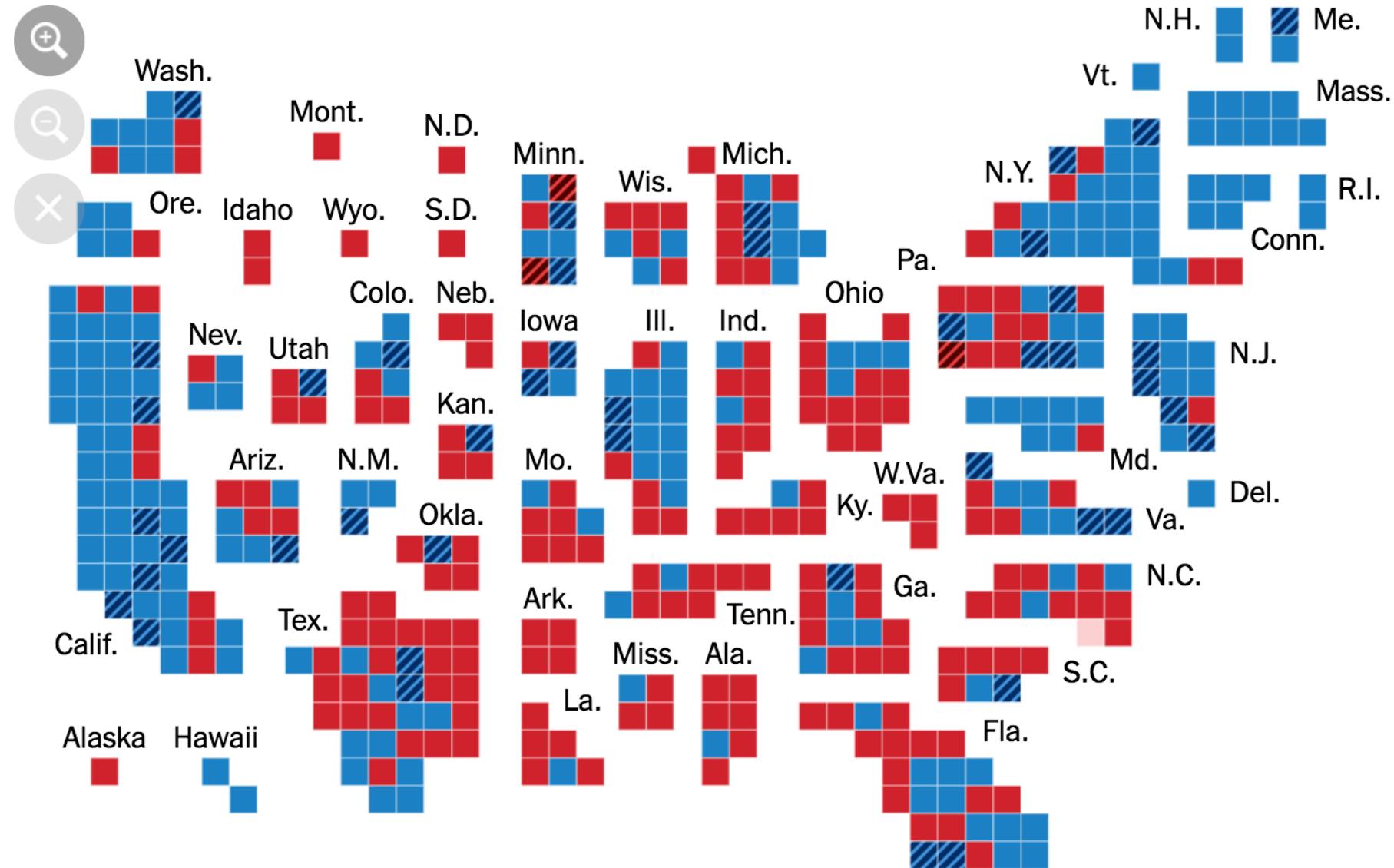
SENATE RACES

HOUSE RACES

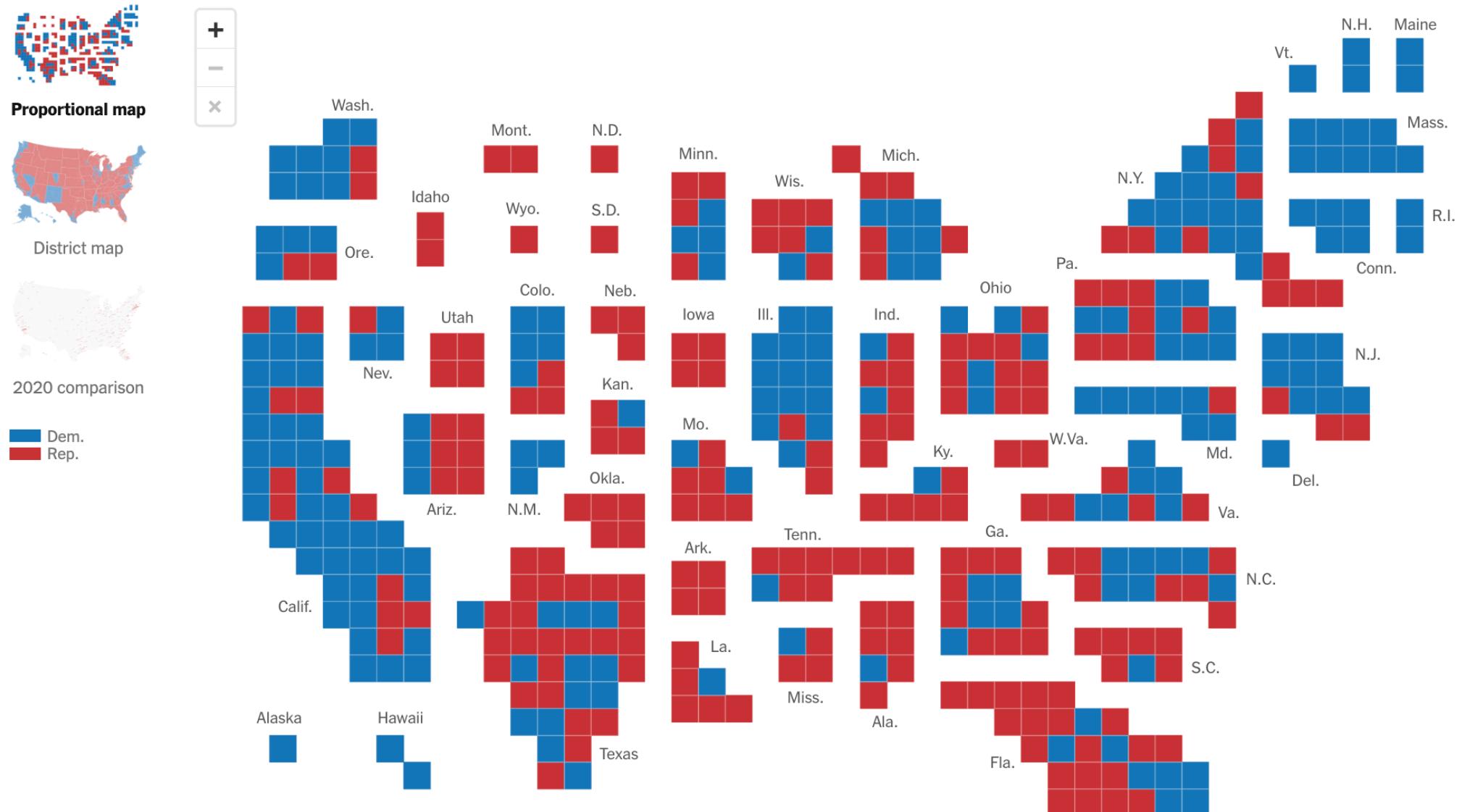
New York Times ratings **198** Safe Dem. **16** Leaning Dem. **17** Toss up **24** Leaning Rep. **180** Safe Rep.



12 years later...



And in 2022...



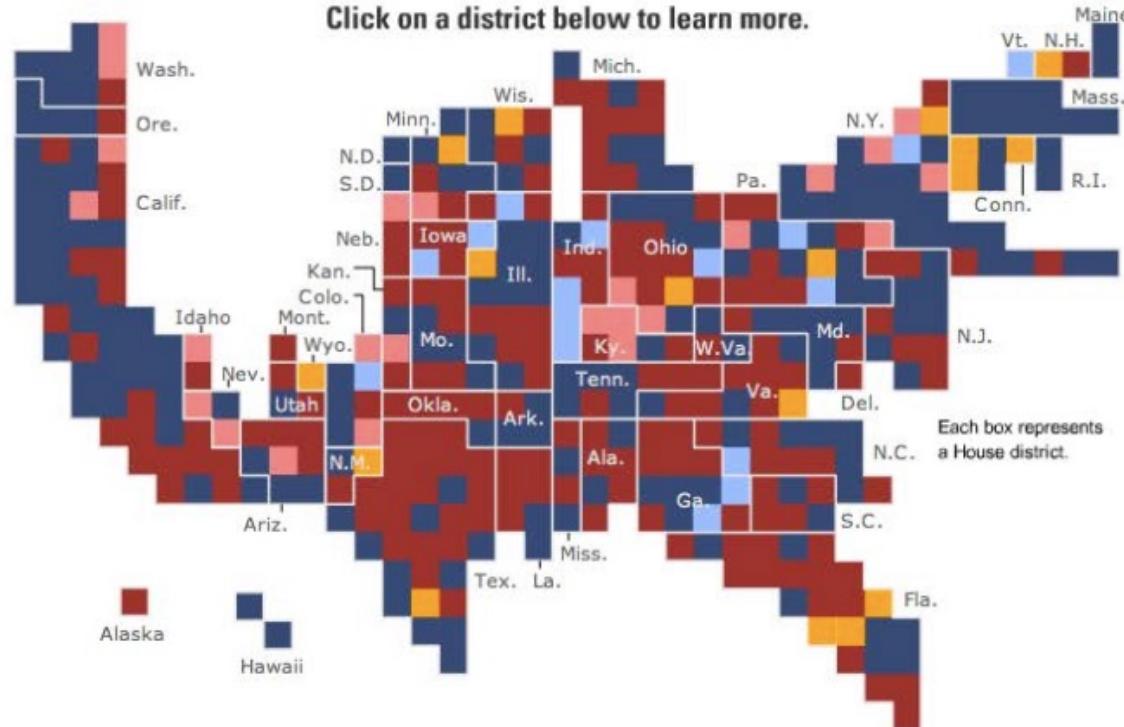
Which do you prefer?

2006 ELECTION GUIDE

**New York Times
ratings**

SENATE RACES
HOUSE RACES

Click on a district below to learn more.



Wash.



Mont.



N.D.



Minn.



Wis.



Mich.



Pa.



Ohio



Ind.



Iowa



III.



Ill.



Mo.



Kan.



Okl.



Colo.



Neb.



Wyoming



Utah



Ariz.



N.M.



Oka.



Tex.



Ark.



Miss.



Ala.



La.



Ga.



Tenn.



S.C.



Vt.



Mass.



R.I.



Conn.



N.J.



Md.



Del.



Va.



N.C.



S.C.



Fla.



Pa.



Ohio



Ind.



Ill.



Mo.



Kan.



Oka.



Tex.



Ark.



Miss.



Ala.



La.



Ga.



Tenn.



Wyo.



S.D.



N.D.



Wash. D.C.



Mont.



Wyo.



Colo.



Neb.



Iowa



III.



Ind.



Ohio



Pa.



Mich.



Wis.



Conn.



N.J.



Md.



Del.



Va.



N.C.



S.C.



Fla.



Ala.



La.



Ga.



Tenn.



S.C.



Fla.



N.C.



S.C.



Fla.



Pa.



Ohio



Ind.



Ill.



Mo.



Kan.



Oka.



Tex.



Ark.



Miss.



Ala.



La.



Ga.



Tenn.



S.C.



Fla.



N.C.



S.C.



Fla.



N.C.



S.C.



Fla.



N.C.



S.C.



Fla.



Conn.



R.I.



Mass.

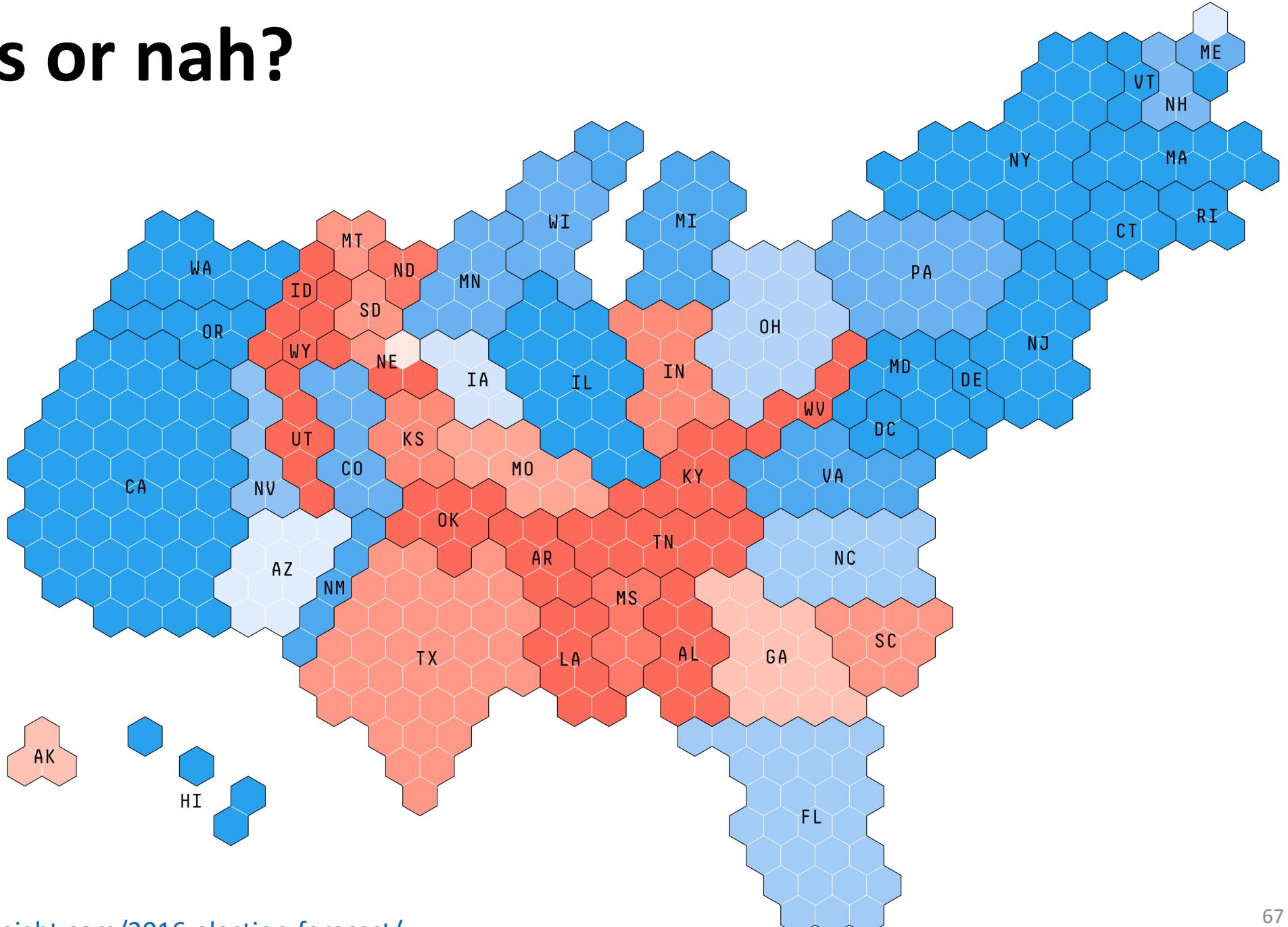


N.H.



Me.

Hexagons or nah?

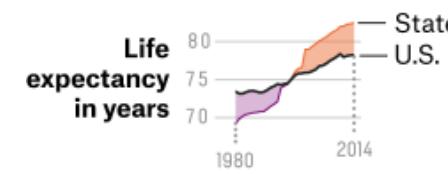


You could preserve shape



Geo-grid + small multiple of charts

Life expectancy in each state vs. U.S. average, 1980-2014



<https://fivethirtyeight.com/features/as-u-s-life-expectancies-climb-people-in-a-few-places-are-dying-younger/>