

Data 227

Data Visualization & Communication

No additional requests may be approved at this time. The number of reserved seats (enrolled students plus unclaimed seats) has reached the section's room capacity. If a student drops a seat or an unclaimed seat expires, you may then approve additional requests.

Week 3 readings

- D3: Data-Driven Documents. Michael Bostock, Vadim Ogievetsky & Jeffrey Heer. InfoVis 2011.
- Chapter 3: Sources of Graphical Integrity, In The Visual Display of Quantitative Information. Tufte. (easy)

Logarithmic scales: why?

- Compress wide range of values onto one graph
[Dynamic Range: the ratio of the largest to the smallest values a tool of yours can handle]
- Convention (Earthquake, RF signal power, sound level, information theory)
- When net effect is the result of small multiplicative effects (microbiology, finance) = sum of log-effects
- Diligence: Log scales should usually be marked with the numbers, not their logarithms. Otherwise, base of logarithm must be specified (2, e, 10, and $\sqrt[10]{10}$ are common, field-specific choices)

The big picture

- task
 - find outliers, diagnose model fit, make decision
- data
 - physical type (int, float...)
 - abstract type (nominal, ordinal, geographic, temporal...)
- domain
 - metadata
 - semantics
 - conceptual model
- mapping
 - visual encoding
 - visual metaphor
- image
 - visual channels
 -

2012

2013

2014

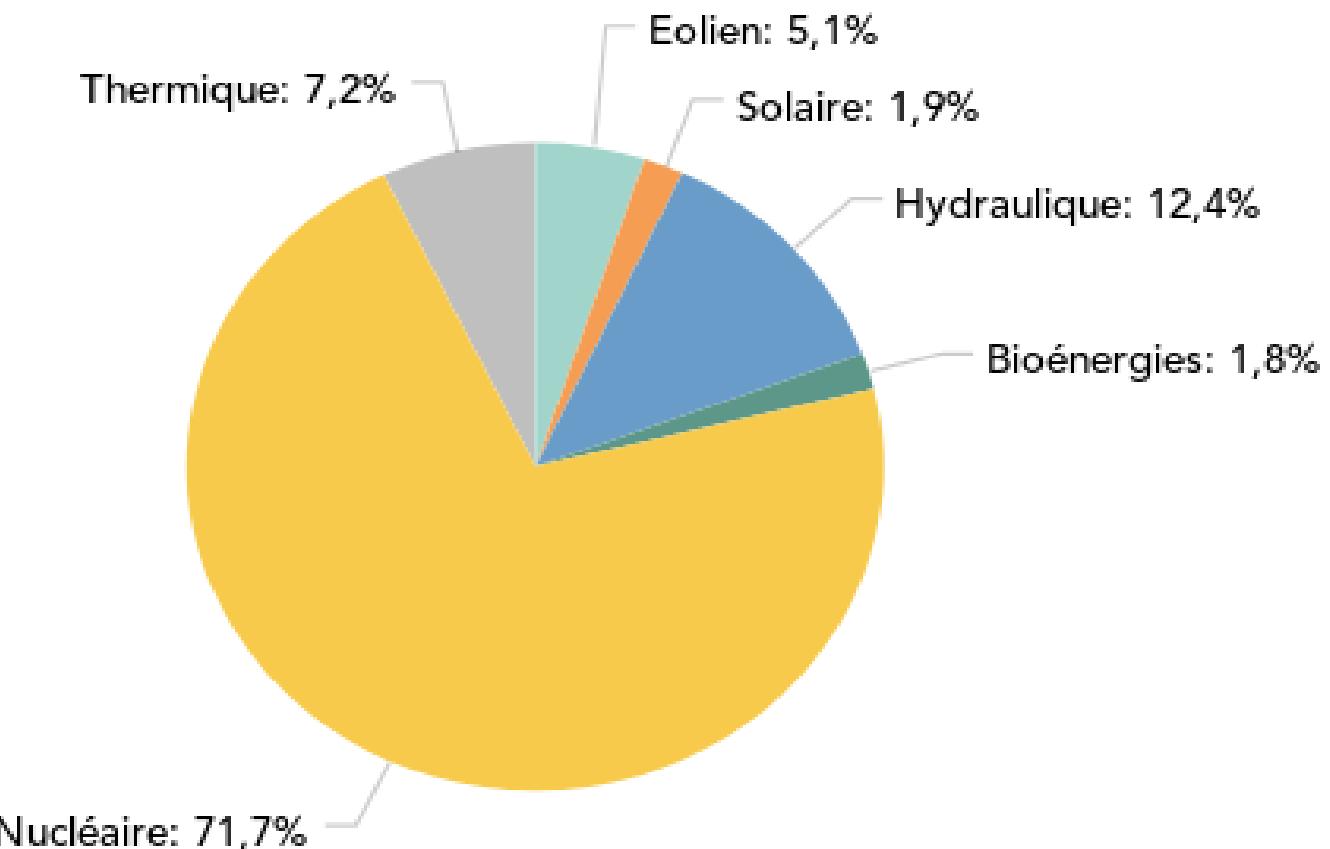
2015

2016

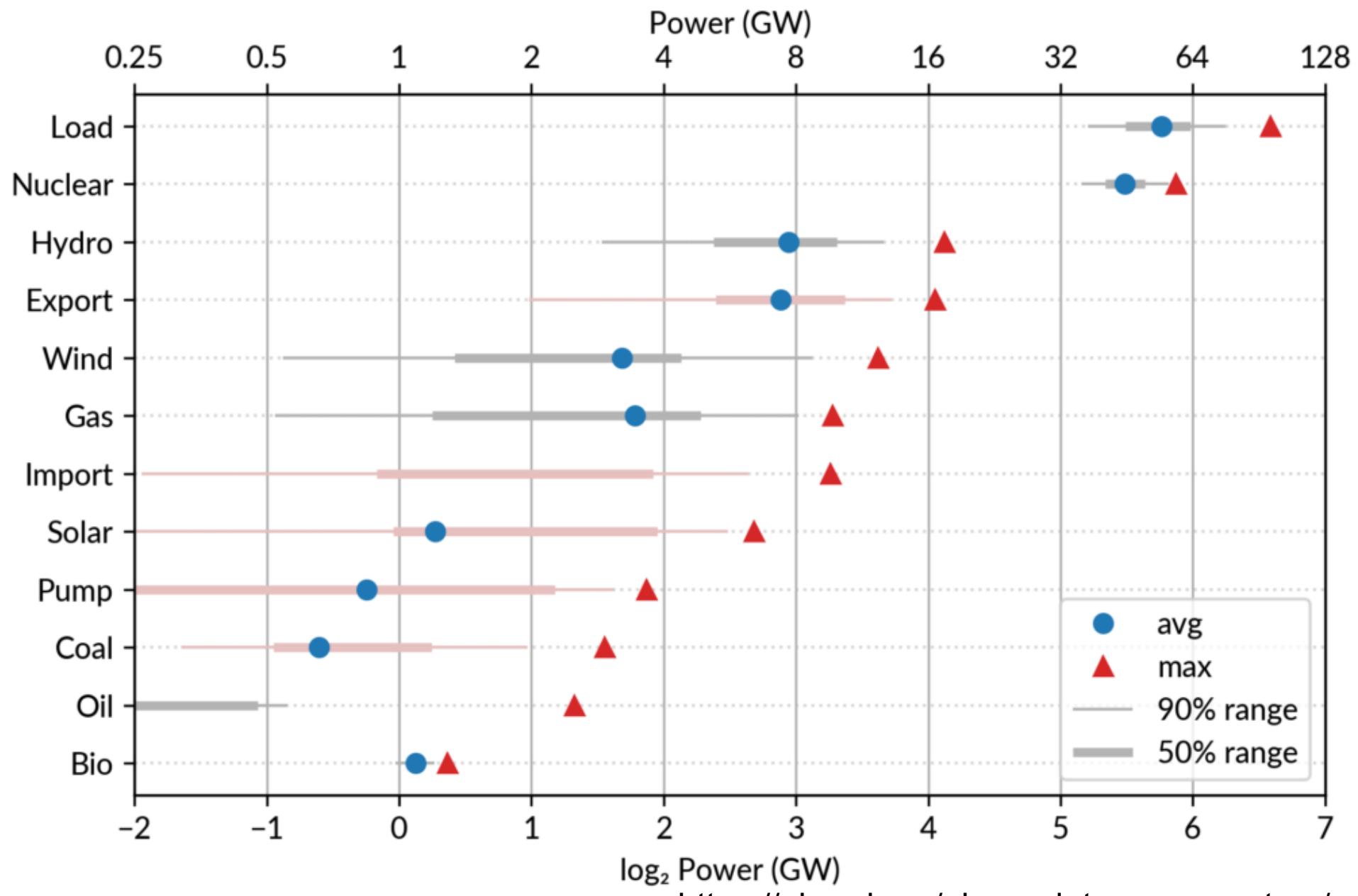
2017

2018

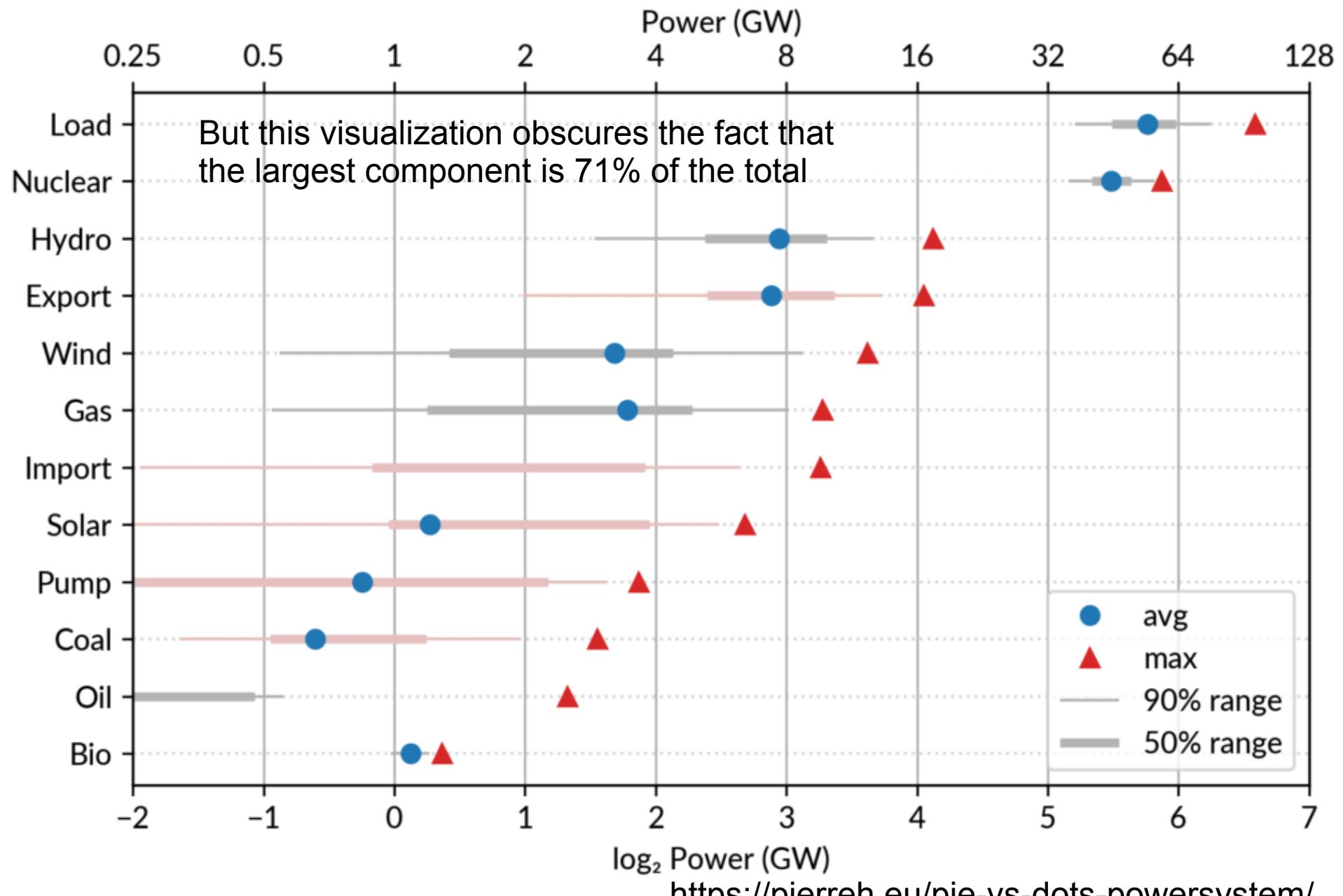
Energie produite



French power system data for 2018 (RTE éCO₂mix)



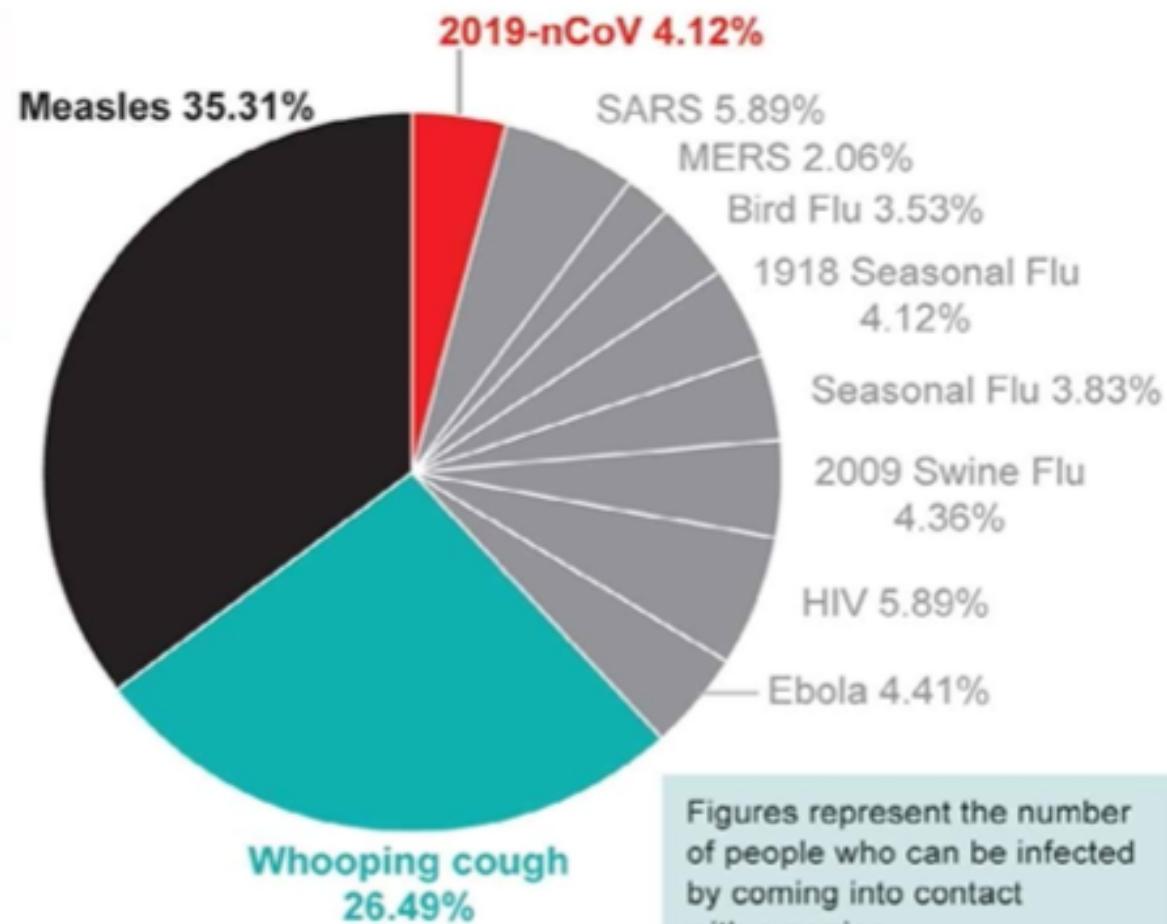
French power system data for 2018 (RTE éCO₂mix)



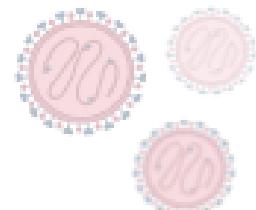
The Bad

How contagious is coronavirus?

The numbers for Covid-19 remain a rough estimate because the long incubation period means we still have no idea how many people have been infected.



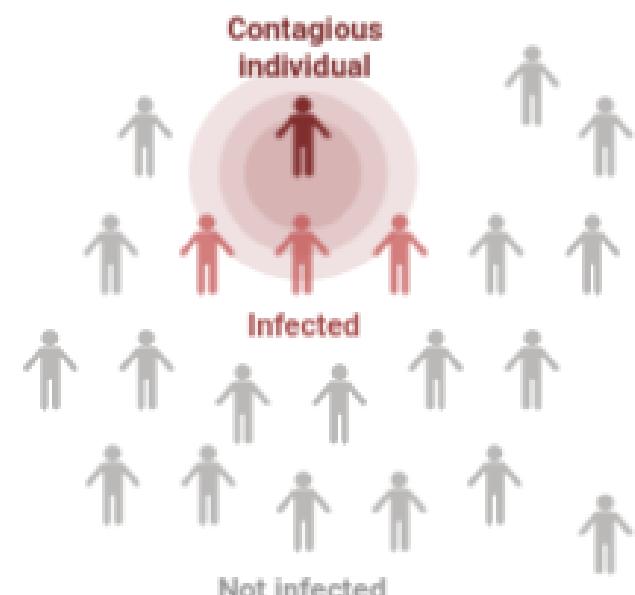
Average Basic Reproduction Number (R_0) of common viral infections



Infectious illness

	R_0
*MERS	0.5
*Seasonal flu	1.0
*Bird flu	1.2
*Swine flu	1.5
Ebola	1.9
*COVID-19	2.2
*Spanish flu	2.2
*SARS	2.8
Polio	3.5
Zika	4.2
Yellow fever	4.3
HIV	6.0
*Mumps	7.0
*Chickenpox	8.5
*Measles	9.0

*Mode of transmission:
air droplets

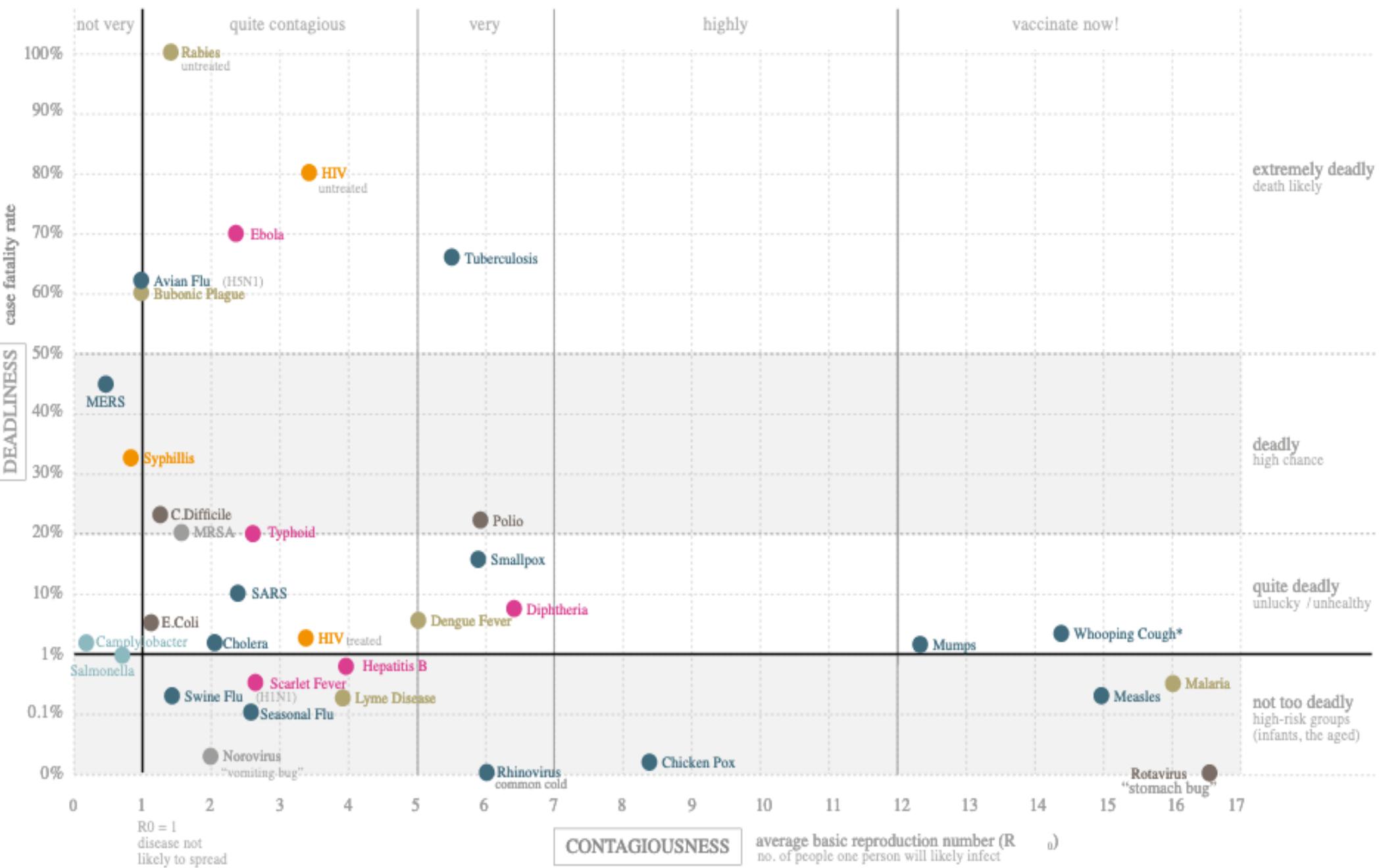


The average basic reproduction number (R_0) is an epidemiologic metric that describes the transmissibility of infectious agents. R_0 measures the expected number of secondary infections produced by a single infectious individual in a susceptible population during the mean infectious period.

The Microbe-scope

PRIMARY TRANSMISSION METHOD

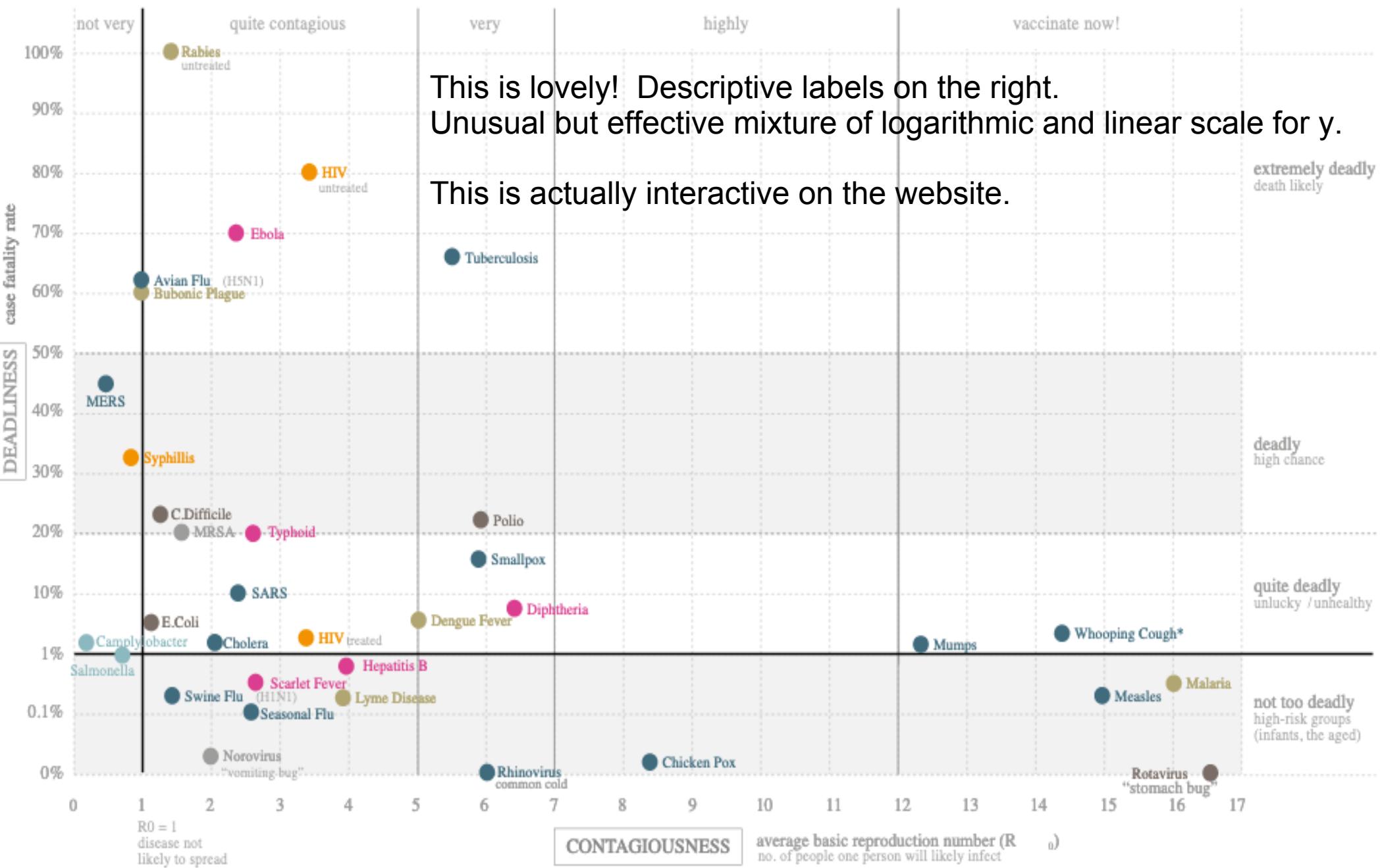
airborne bites body fluids fecal-oral food sexual contact surfaces



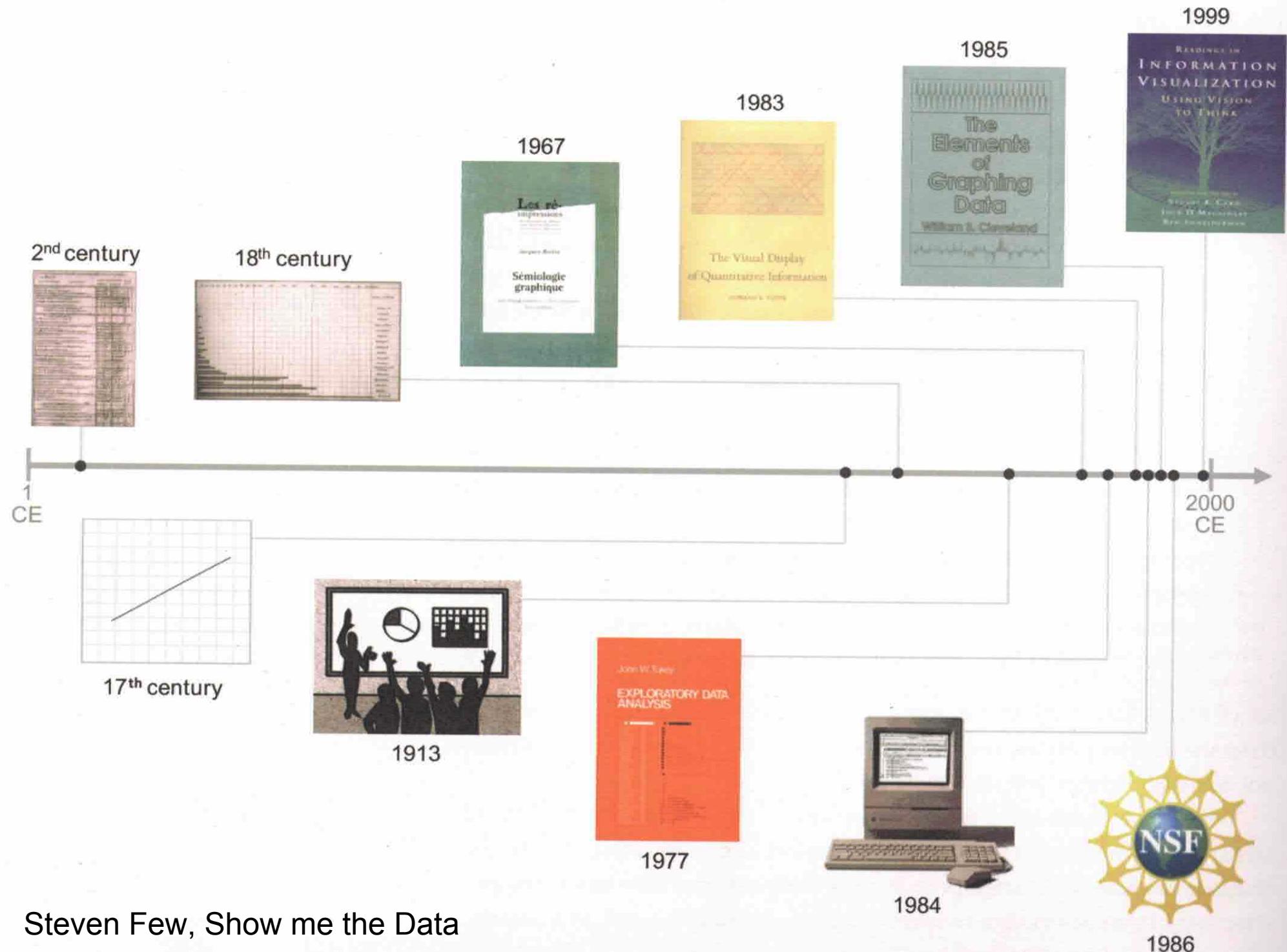
<https://www.informationisbeautiful.net/visualizations/the-microbescope-infectious-diseases-in-context/>

The Microbe-scope

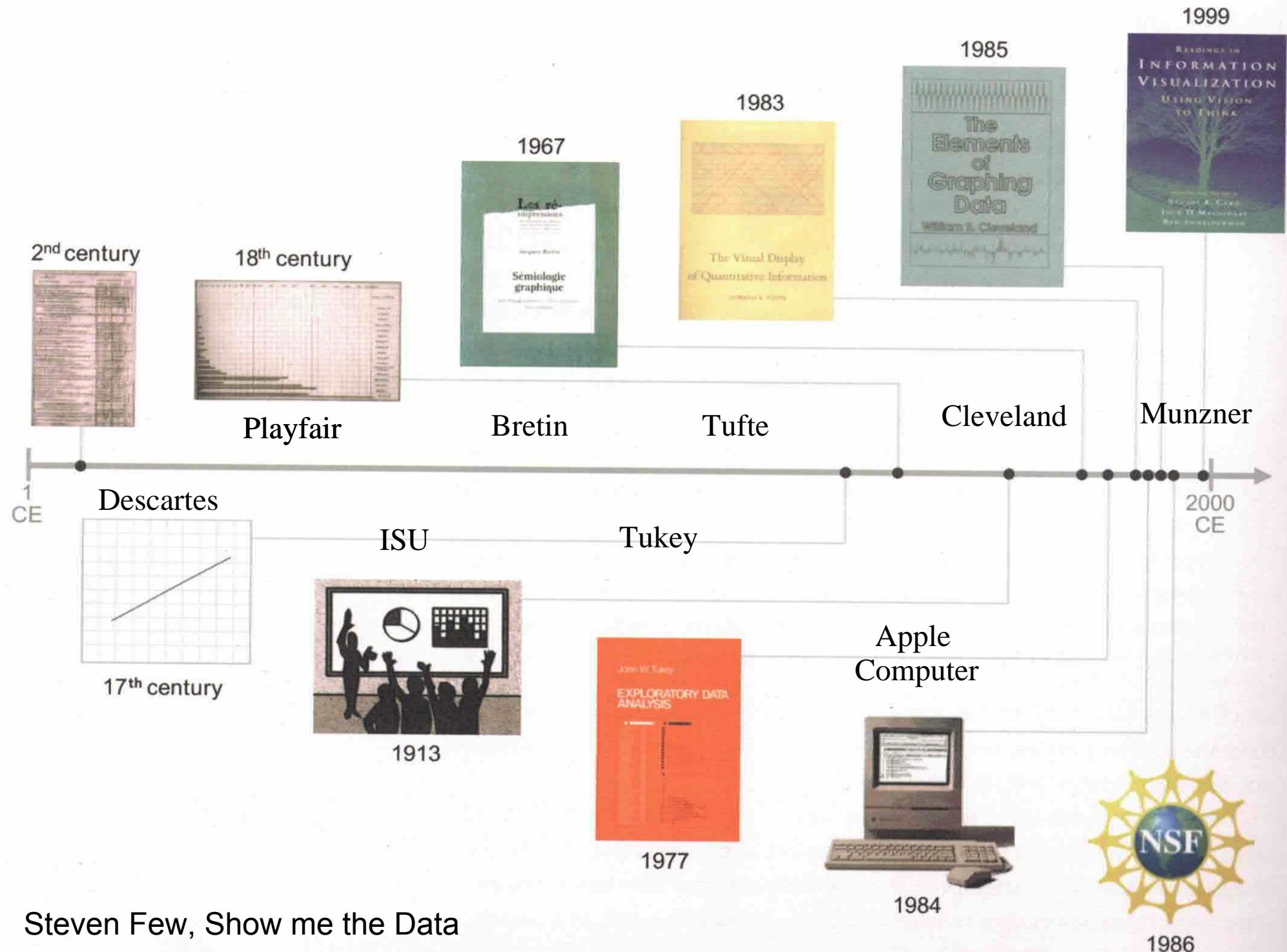
PRIMARY TRANSMISSION METHOD airborne bites body fluids fecal-oral food sexual contact surfaces



<https://www.informationisbeautiful.net/visualizations/the-microbescope-infectious-diseases-in-context/>



Steven Few, Show me the Data





BETTER DATA VISUALIZATIONS

A Guide for Scholars, Researchers, and Wonks



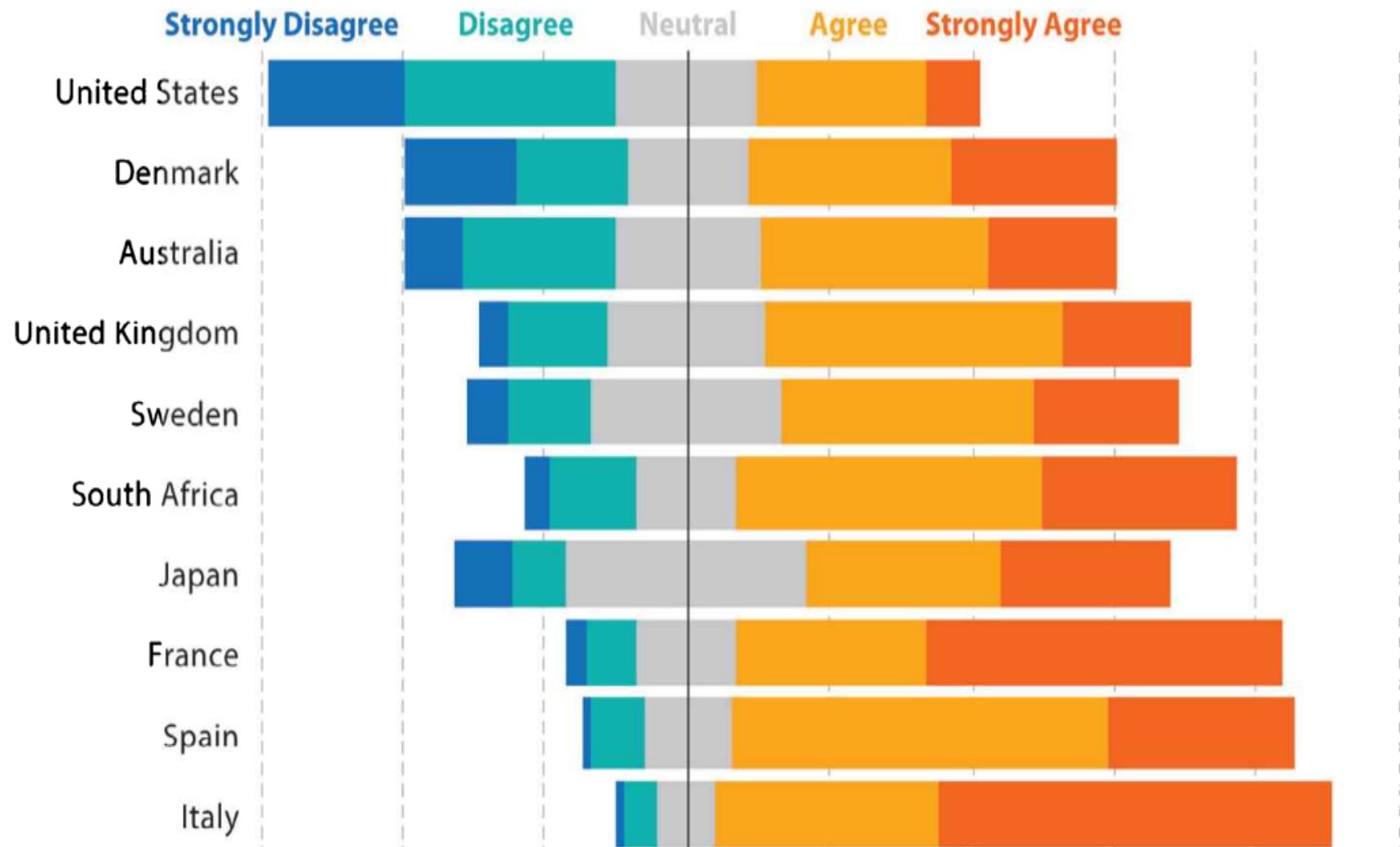
Jonathan Schwabish

From a political/policy perspective, this is a pretty good nuts-and-bolts book on making pretty good graphs.

Likert scale -- diverging bar chart

It is the responsibility of government to reduce differences in income between people with high & low incomes

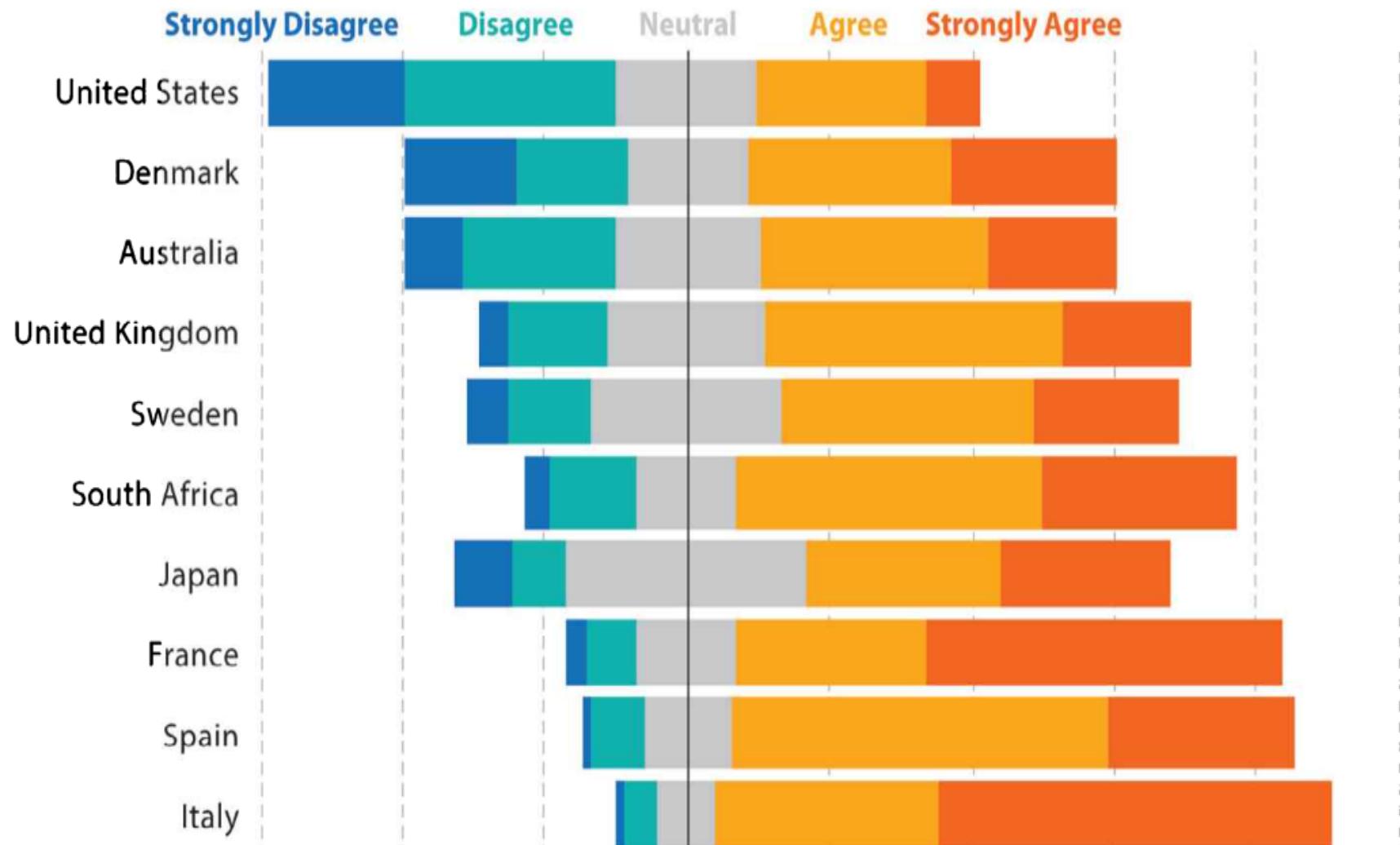
(Percent)



Likert scale -- diverging bar chart

It is the responsibility of government to reduce differences in income between people with high & low incomes

(Percent)

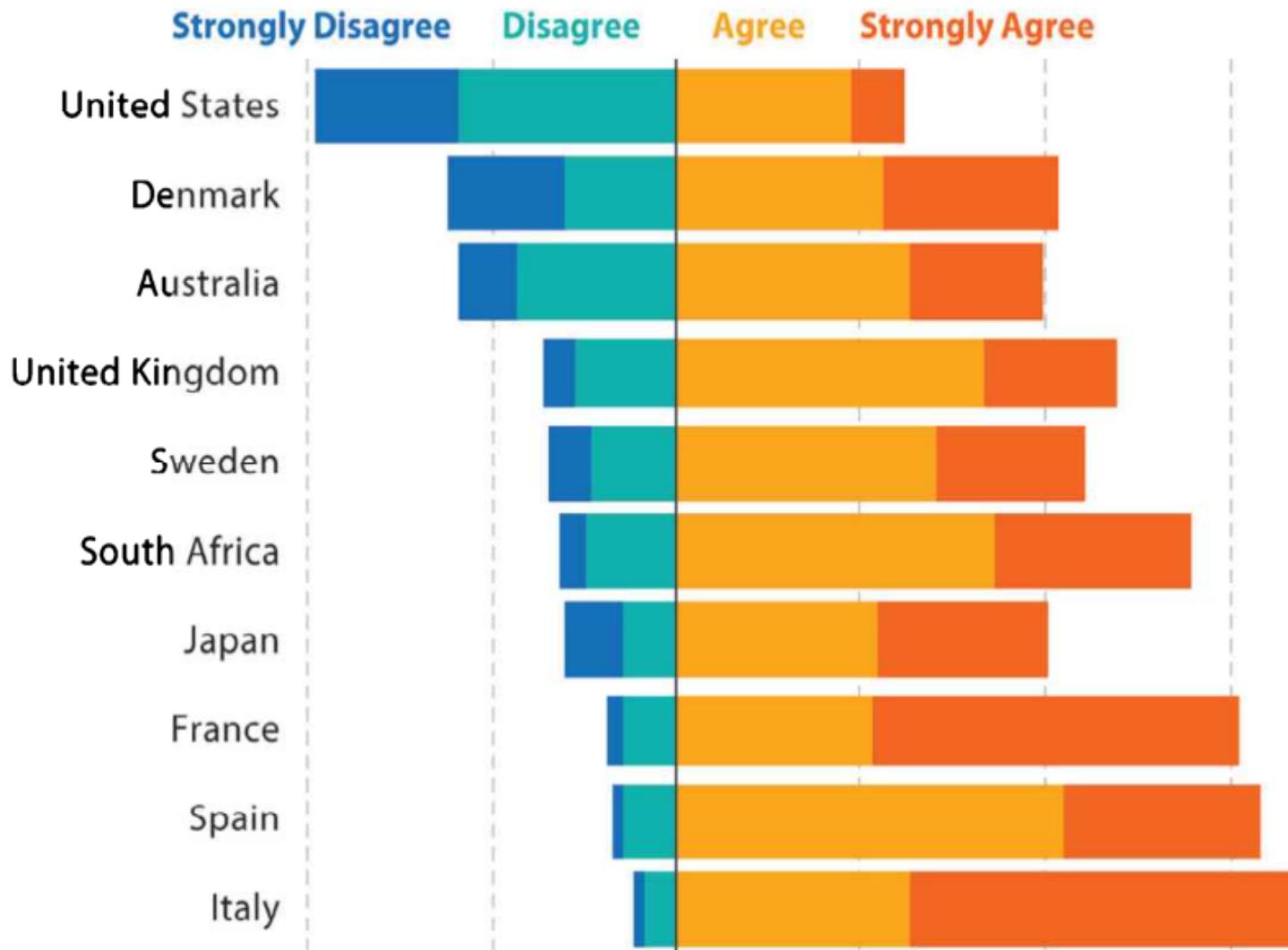


Likert scale -- diverging bar chart

It is the responsibility of government to reduce differences in income between people with high & low incomes

(Percent)

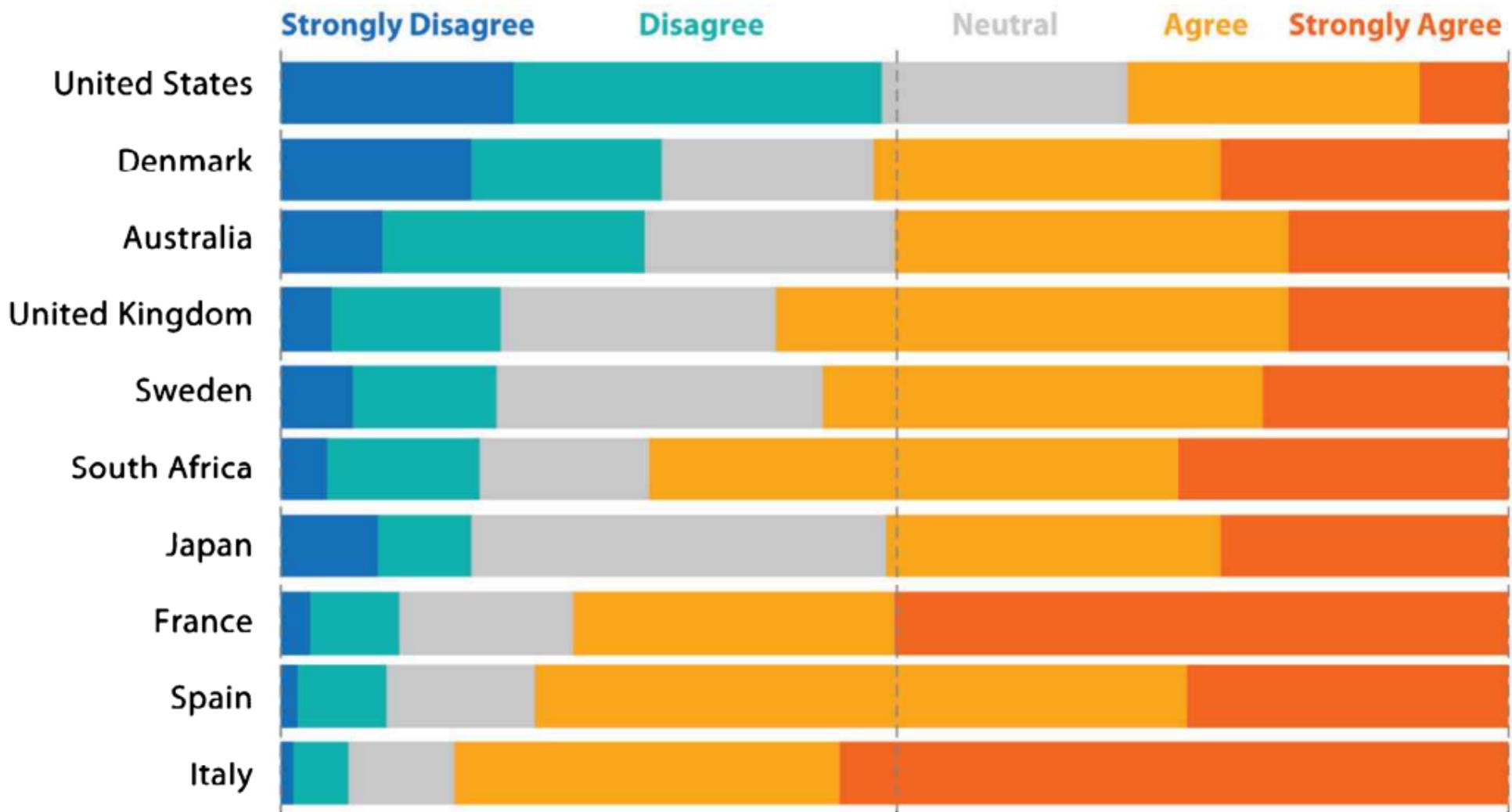
Neutral has not been plotted.



Likert scale -- stacked bar

It is the responsibility of government to reduce differences in income between people with high & low incomes

(Percent)



Source: International Social Survey Programme, 2009

Jonathan Schwabish, Better Data Visualizations

Likert scale - small multiples

It is the responsibility of government to reduce differences in income between people with high & low incomes

(Percent)

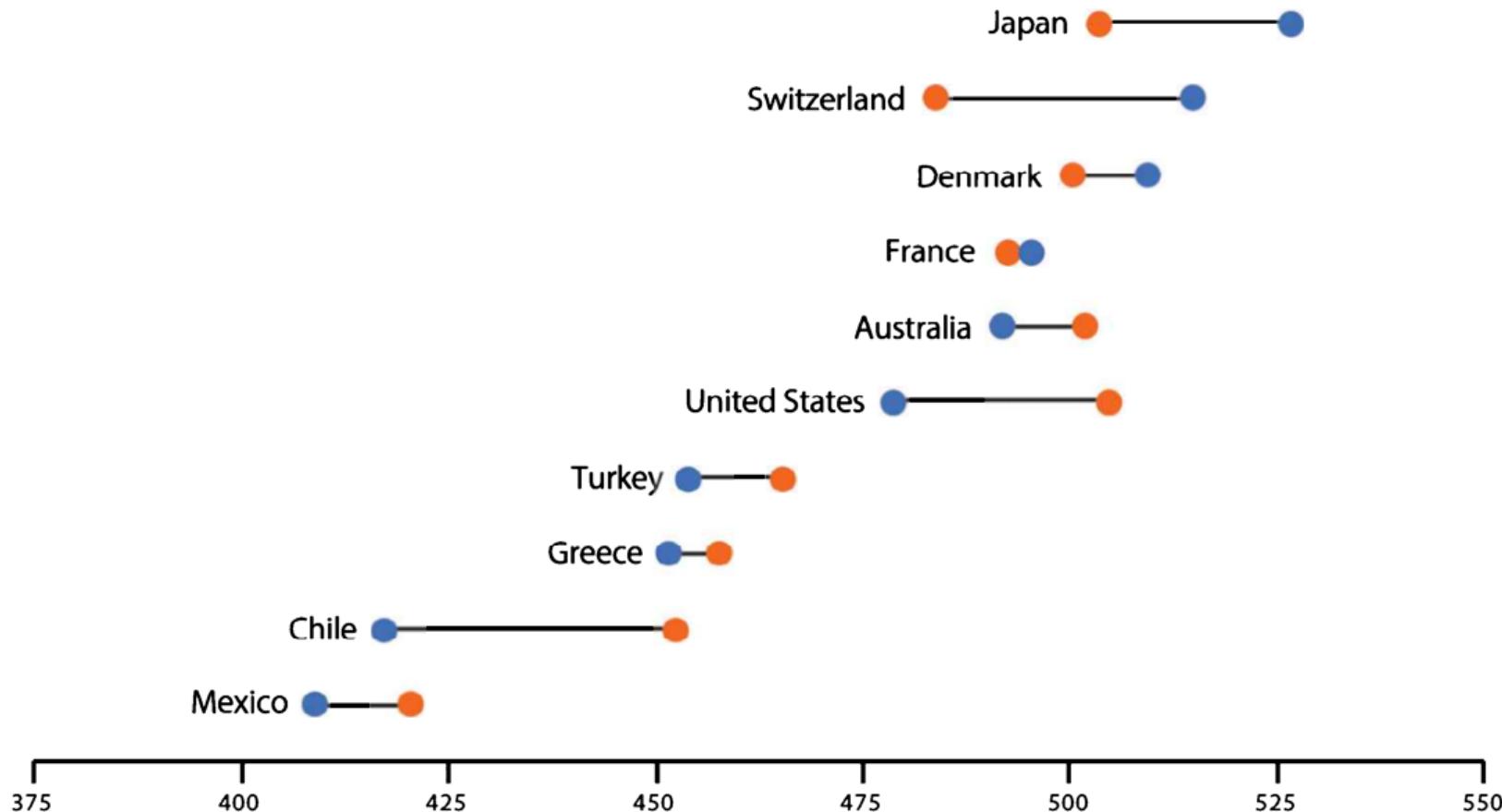


Source: International Social Survey Programme, 2009

Jonathan Schwabish, Better Data Visualizations

“Dot plot”

PISA scores for **math** and **reading** among 10 OECD countries

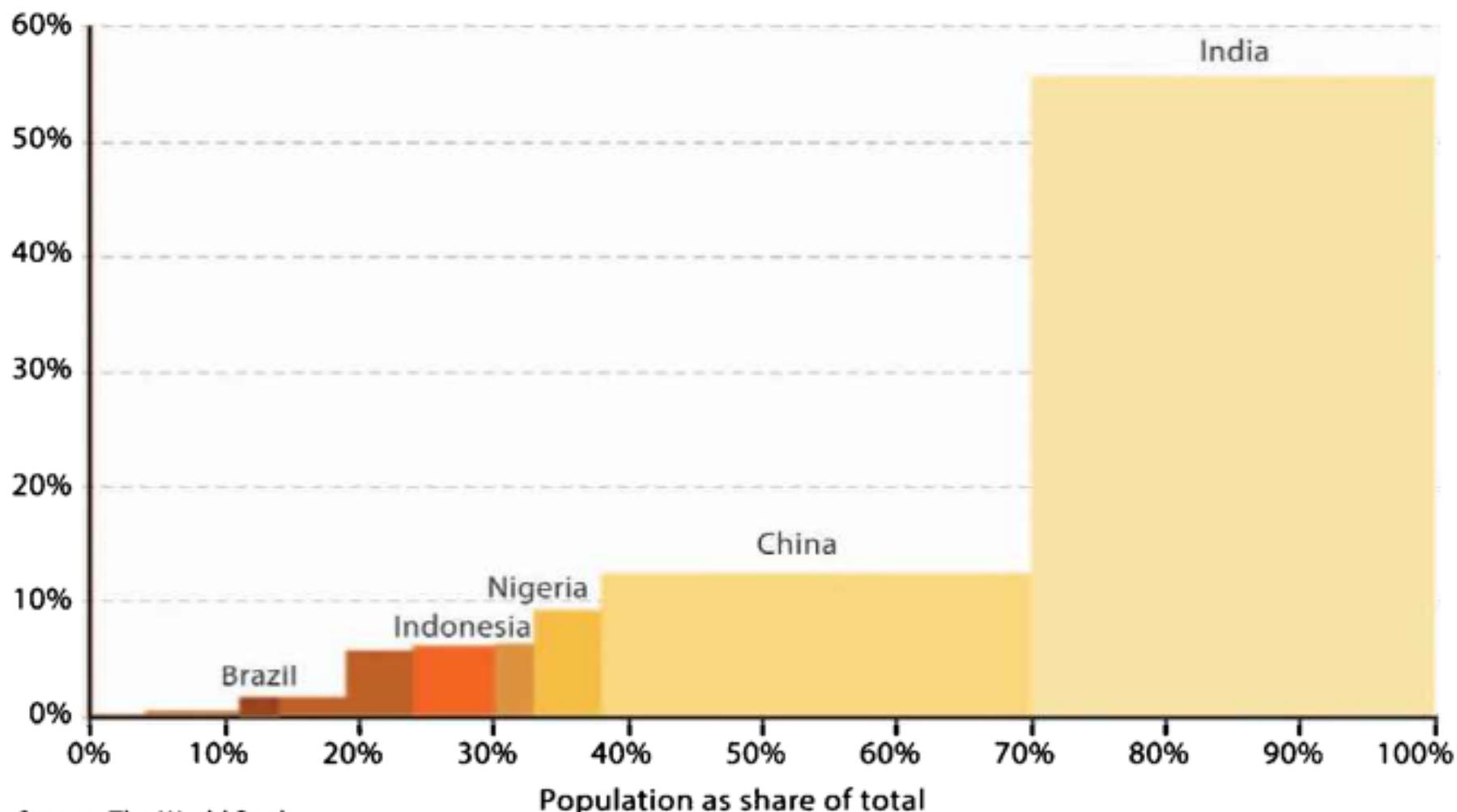


Source: Programme for International Student Assessment

Marimekko graphs: Bar Charts with more variables

High population, high poverty

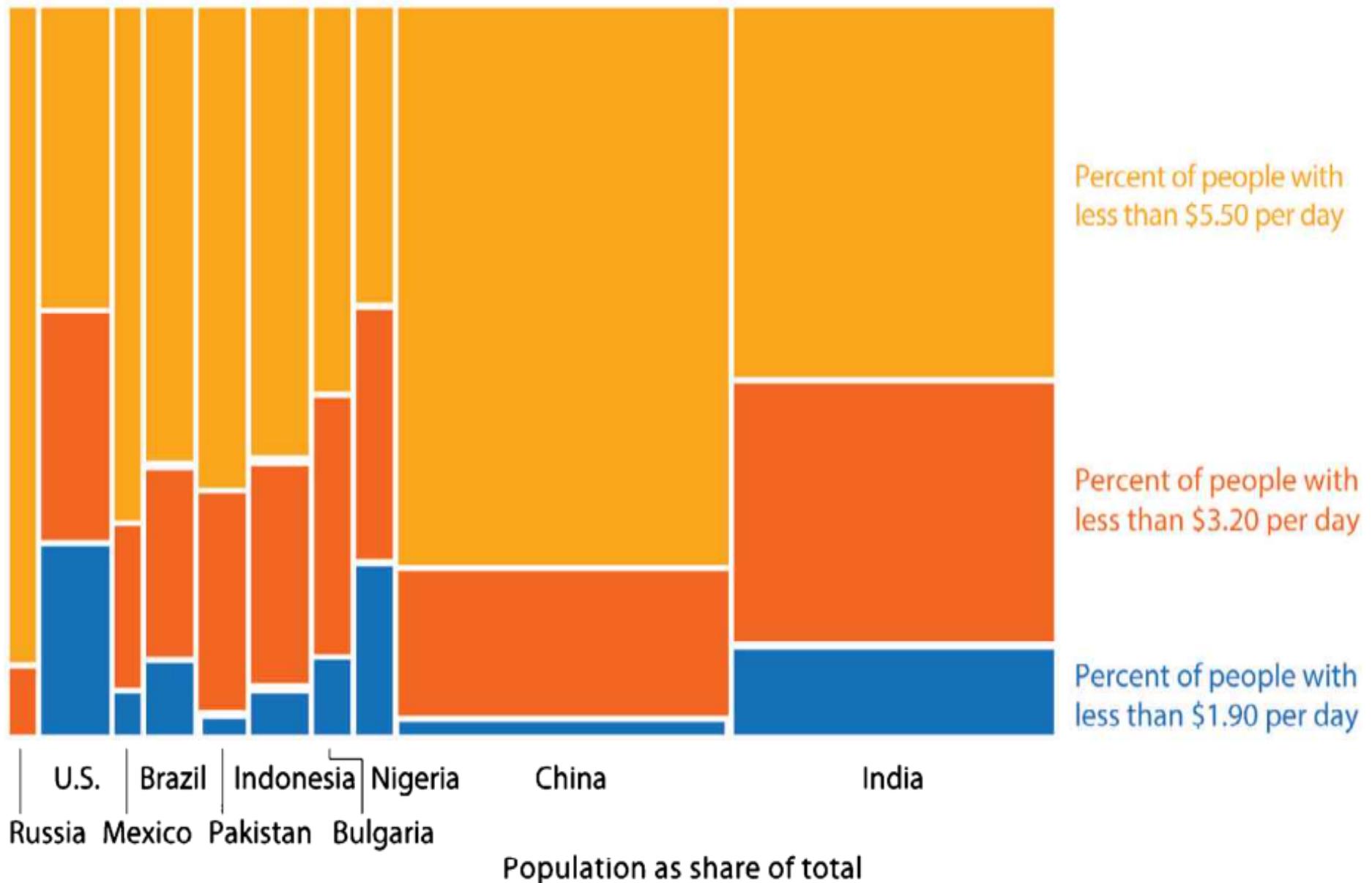
(Percent of people with less than \$5.20/day)



Source: The World Bank

Marimekko graphs

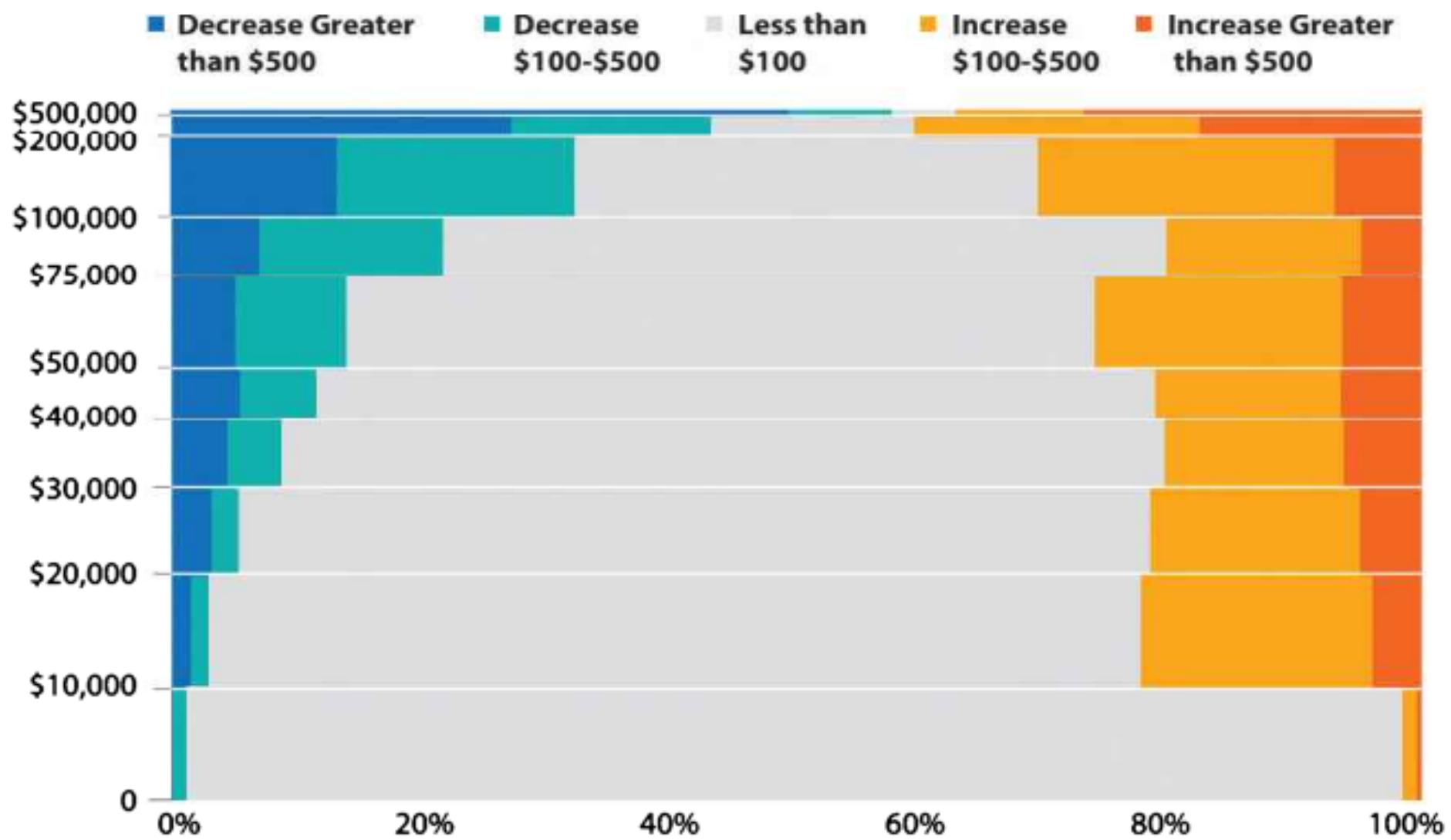
High population, high poverty



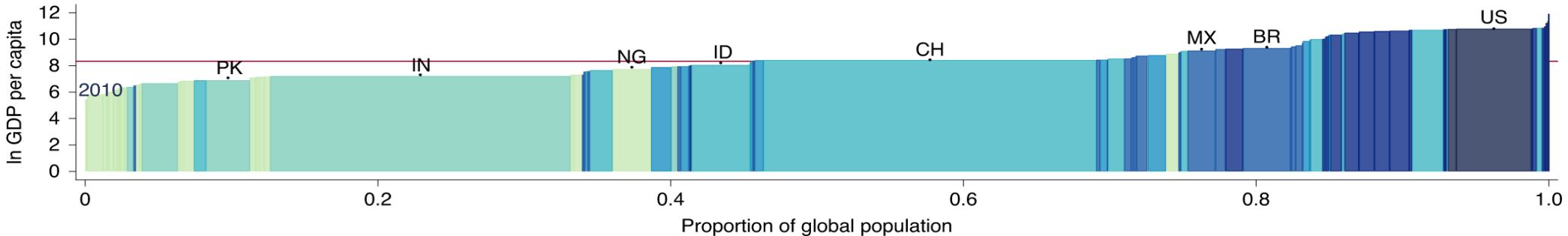
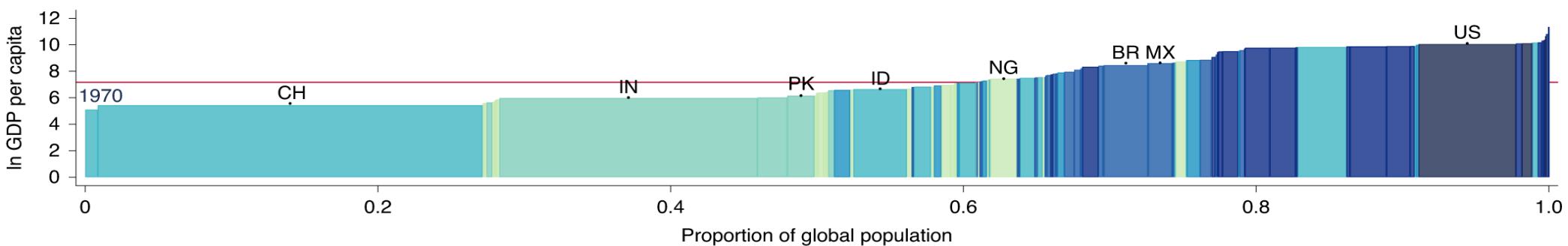
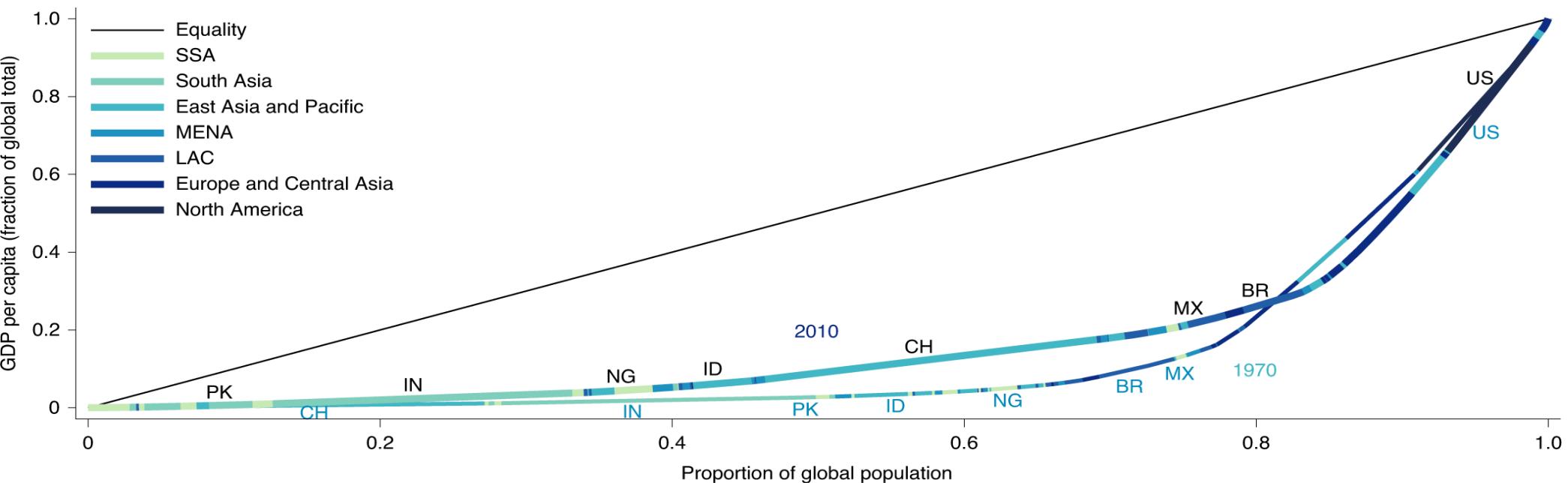
Source: The World Bank

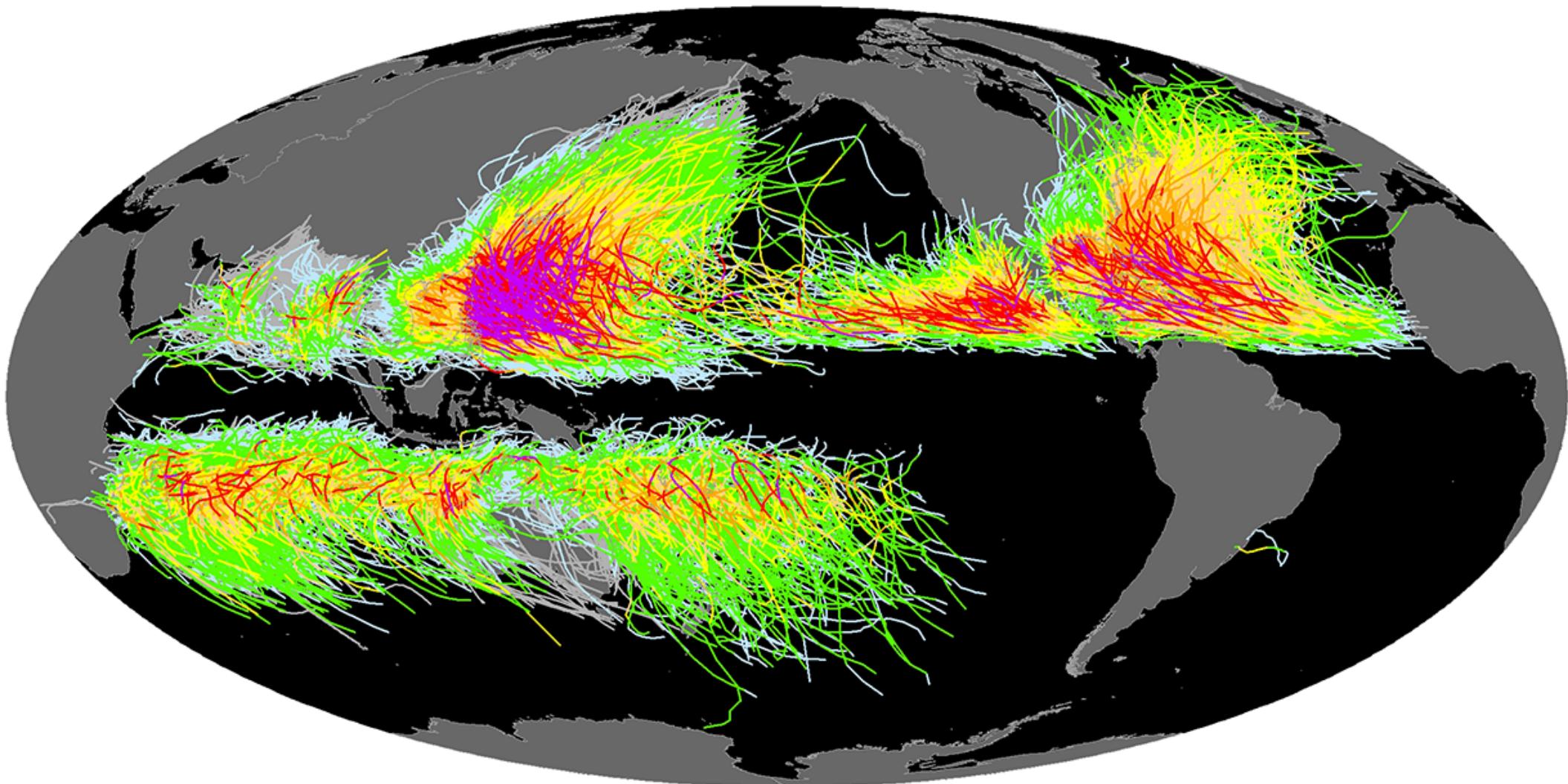
Marimekko graphs

Distribution of tax returns by size of tax change for the “Tax Cuts and Jobs Act”



Source: Joint Committee on Taxation





The International Best Track Archive for Climate Stewardship (IBTrACS) stores global tropical cyclone information.

Saffir-Simpson Hurricane Wind Scale

Intensity Missing	—	Category 1
Tropical Depression	—	Category 2
Tropical Storm	—	Category 3
	—	Category 4
	—	Category 5

HW2: Hurricanes and lake temperature

- Read the wikipedia page on tropical cyclones, ok?
- You can just make a scatterplot for problem 1
BUT
 - Map-like scatterplots can't be dramatically different from 1:1 scale. Do not make a square plot that covers 1 degree of latitude and 16 degrees of longitude.
 - If you don't use a map projection, you need to label latitude and longitude
 - If you use size-of-symbol to indicate diameter of sustained winds, **you should make the diameter match the scale of the map.**

HW2: Hurricanes and lake temperature

- Play with the data, find something, and show it.
- What can you do with weather data?
 - When might one swim in the lake without pain?
 - When does the lake produce heavy fog?
 - Why does the water temperature sometimes drop suddenly in the summer?
 - How long does it take the lake to heat up ? How much later does it heat up compared to the air?
 - Why did my son get sunburn in October?

Back to the matrix

- 04-SALARY-OPEN.ipynb Loads in the Chicago salary database, makes a histogram with matplotlib, and makes a somewhat nicer-looking histogram with altair
- altair_data_transformation.ipynb: math, log transforms, binning and aggregation in altair
- Charley_track.ipynb
Draws a map:

