CMPU 100 · Programming with Data

Advanced Visualization

Class 15

Join us for the 11th annual

DataFest VASSAR 2026

Friday, April 17 - Sunday, April 19, 2026

5 pm Friday - 5 pm Sunday
Rockefeller Hall

Your challenge for the weekend: Ask questions, and draw insights from a surprise dataset. To participate in this friendly competition, you only need enthusiasm for data. Winners will receive prizes and there will be food and swag for everyone.

To register and find out more information, visit http://pages.vassar.edu/datafest/

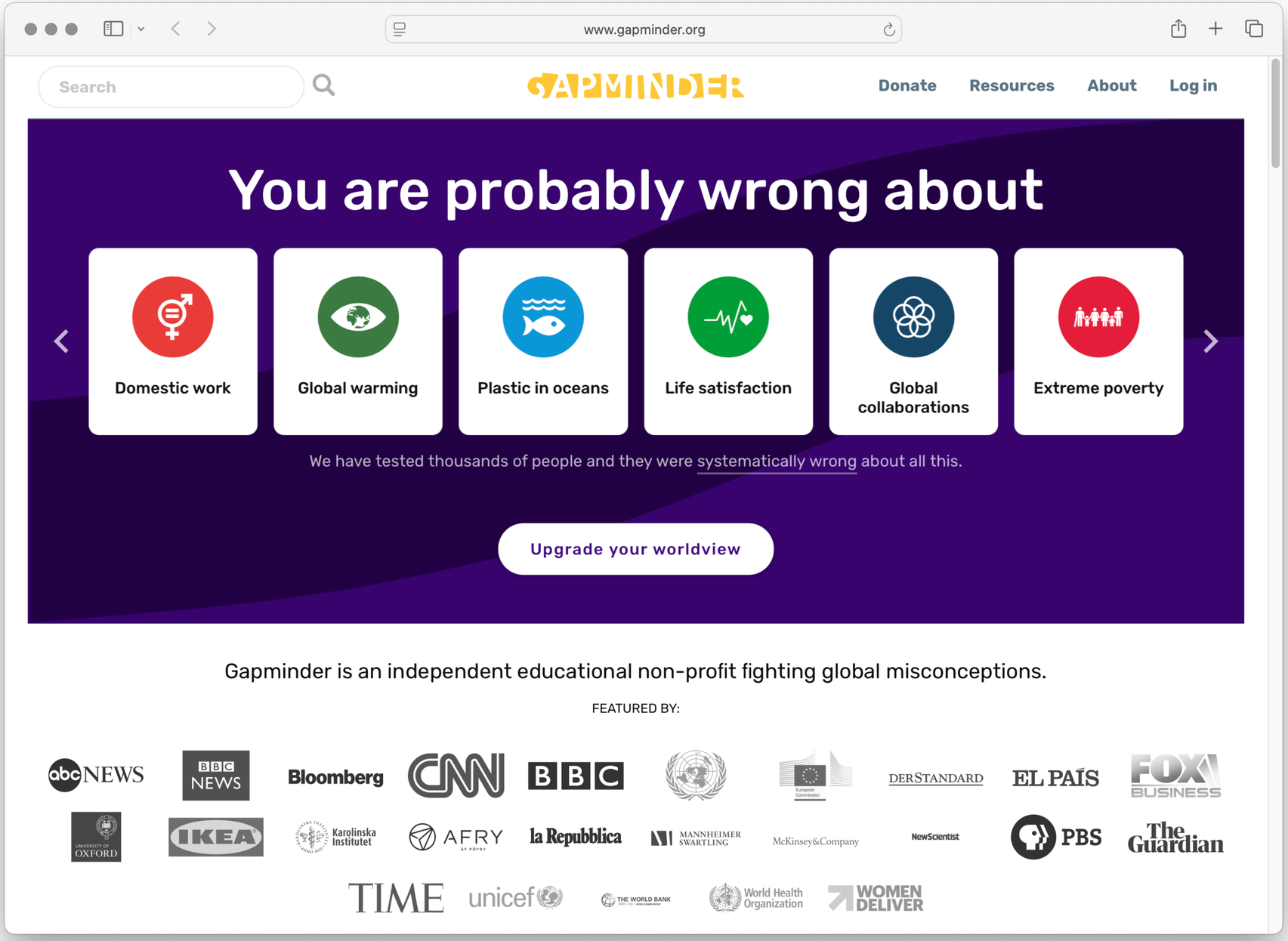
Register by 3/27

Sponsored by: Art, Biology, CCE, CIS, CogSci, CS, DSS, Earth Science & Geography, Economics, Education, Environmental Studies Program, ERI, Math & Stats, NBP, Physics & Astronomy, Poli Sci, President's Office, Psychological Science, and STS

Individuals with disabilities requiring accommodations or information on accessibility should contact the Campus Activities Office, (845) 437-5370. Without sufficient notice, appropriate space and/or assistance may not be available.

Plotly is a popular Python library for visualization. It has a lot of options, which makes it powerful – but also a bit confusing when you're just starting out.

> You're not required to know Plotly for this class – specifics about it won't be on next week's exam – but we want to give you a taste of what it can do!



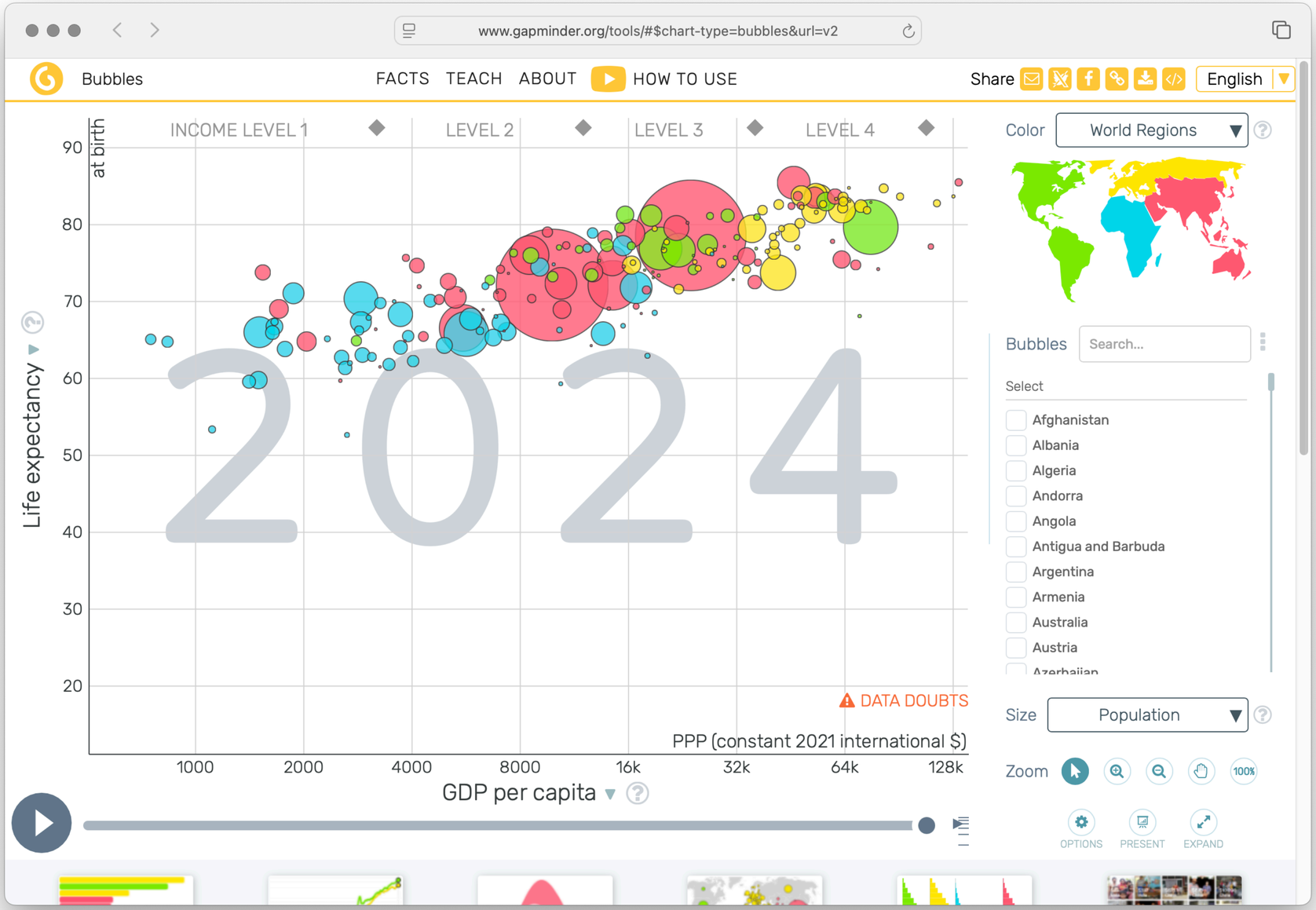
gapminder.org

Notebook: *Gapminder*

youtu.be/Z8t4k0Q8e8Y

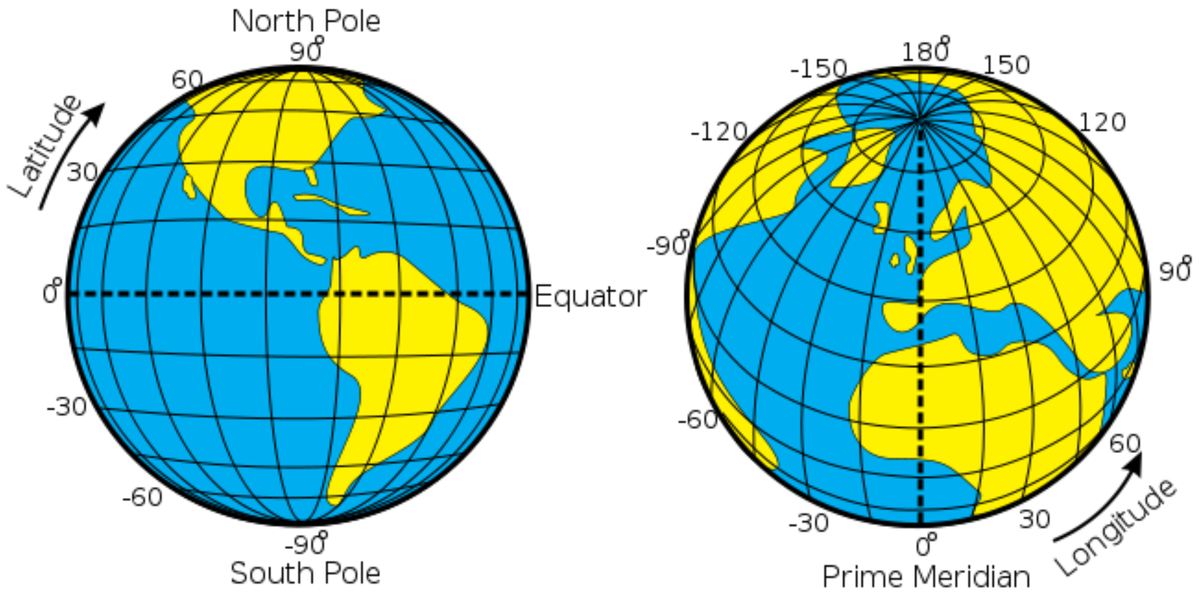
Scatter plots revisited

Notebook: *Scatter plots revisited*



gapminder.org/tools

Scatter plot maps revisited

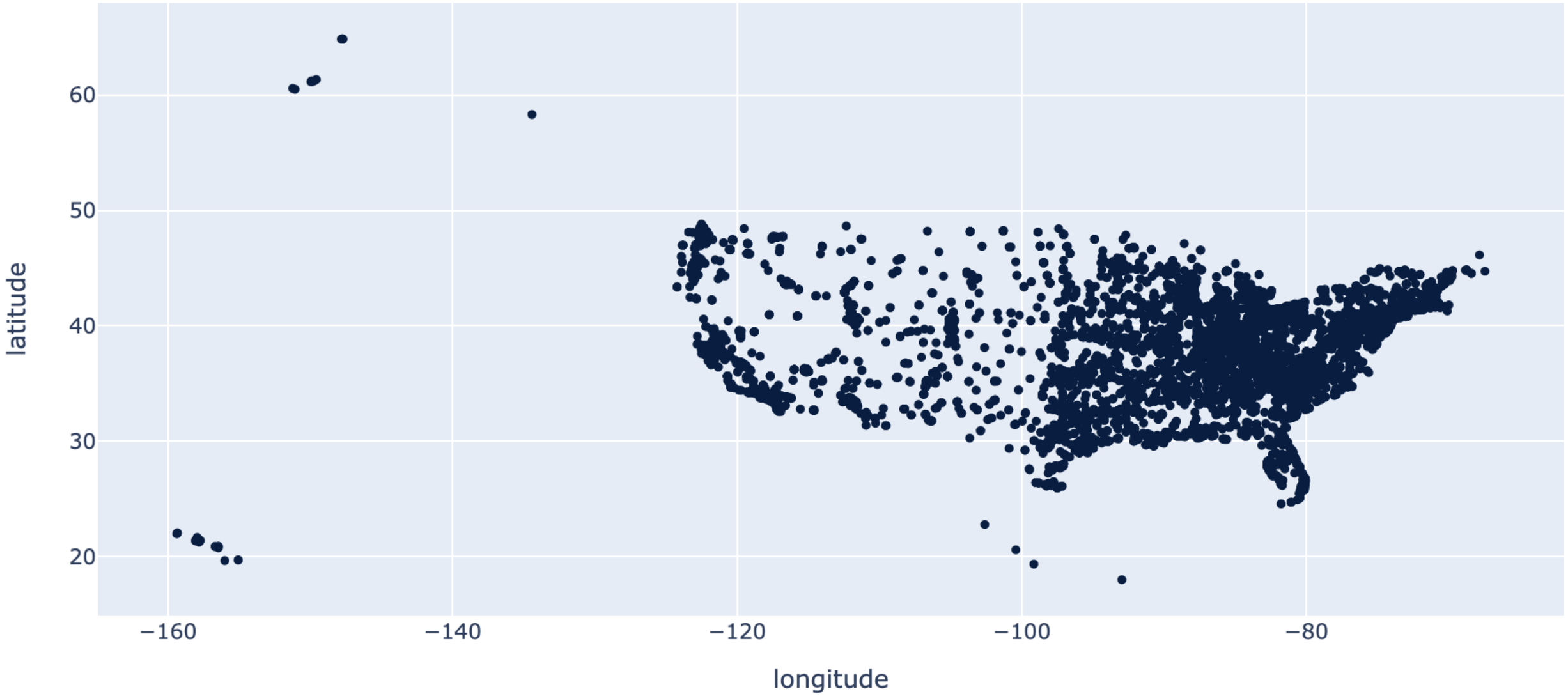


Any point on Earth can be described by its

latitude – the vertical (*y*) position – and

longitude – the horizontal (*x*) position.

What's wrong with just plotting latitude and longitude using a scatter plot?



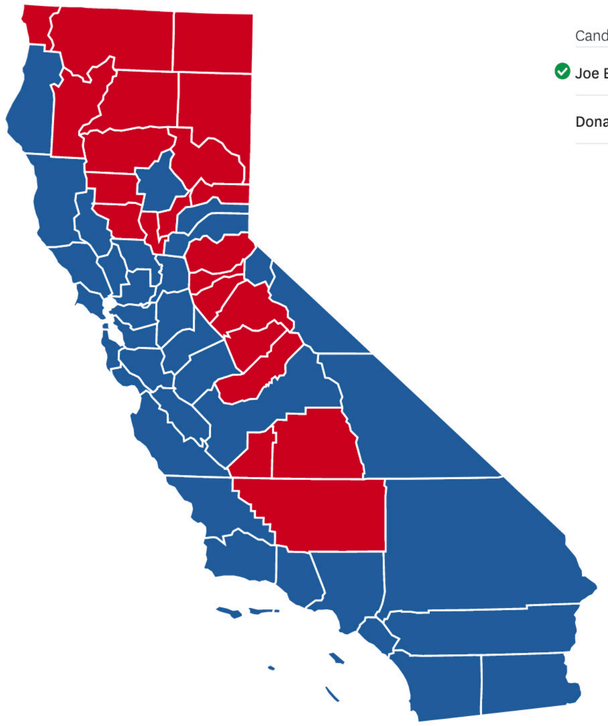
Notebook: *Scatter plot maps revisited*

Choropleth maps

Choropleth maps are useful for visualizing numerical variables across different regions (e.g., municipalities, counties, states, countries).

In this sense, they're analogous to bar charts since they encode one categorical variable (the region) and one numerical variable.

Example: Election mapping



Candidate	Votes	Pct.
Joe Biden D	8,281,504	65.1%
Donald Trump R	4,204,244	33.0%

County	Joe Biden	Donald Trump
Alameda	355,743	69,871
Alpine	473	238
Amador	8,011	13,427
Butte	44,455	41,182
Calaveras	8,804	13,275
Colusa	2,926	3,953
Contra Costa	265,942	86,253
Del Norte	4,324	5,647
El Dorado	47,635	54,177
Fresno	147,683	123,941
Glenn	3,315	6,139

SHOW FEWER

San Francisco Chronicle

Notebook: *Choropleth maps*

Acknowledgments

This class incorporates material from:

Data 6, University of California, Berkeley, CC BY-NC-SA