



LECTURE 20

Maps

Adding another tool to our visualization toolkit.

Data 6 Summer 2022

Developed by students and faculty at UC Berkeley and Tuskegee University

data6.org/su22/syllabus/#acknowledgements-

Week 4

- **Homework 3** has been released and will be due on **7/28 @ 11 PM**
- **Homework 2** grades have been released on Gradescope. Regrade requests are due by **7/27 @ 6 PM**
- Remember that you can get **3%** extra credit if you have a **mid-semester check-in** with a course staff member
 - See **Ed** for more details

Announcements!

Today's Roadmap

Lecture 20, Data 6 Summer 2022

1. Motivation
2. Scatter Plot Maps
3. Choropleth Maps

➤ **1. Motivation**

- 2. Scatter Plot Maps
- 3. Choropleth Maps



Motivation



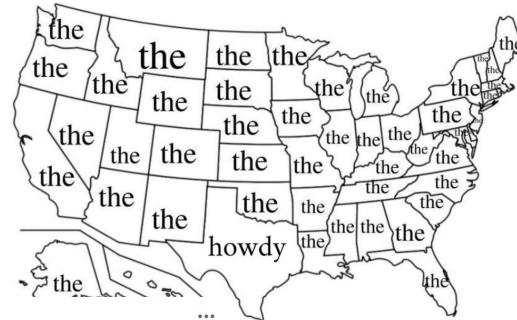
Tanzania

$$\frac{\sin(zania)}{\cos(zania)} =$$

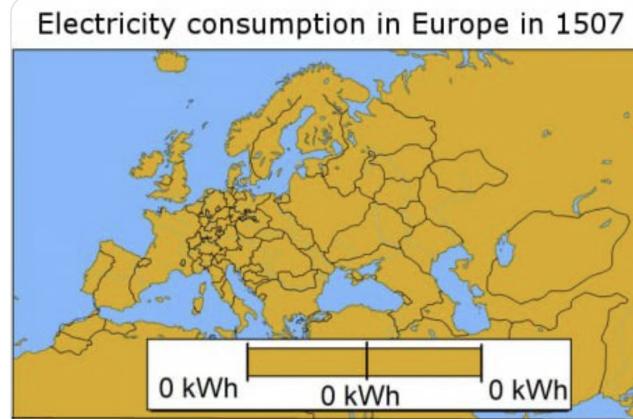


...

The most popular word in each state



Electricity consumption in Europe in 1507



Today's Data

Over 10,000 fast food restaurants from across the US, with over 500 unique fast food chains.



name	city	latitude	longitude	province	country
McDonald's	Massena	44.9213	-74.8902	NY	US
Wendy's	Washington Court House	39.5326	-83.4453	OH	US
Frisch's Big Boy	Maysville	38.6274	-83.7914	KY	US
McDonald's	Massena	44.9501	-74.8455	NY	US
OMG! Rotisserie	Athens	39.3516	-82.0973	OH	US
Domino's Pizza	Hamilton	39.4176	-84.4764	OH	US
Domino's Pizza	Englewood	39.8697	-84.2936	OH	US
McDonald's	Saluda	34.006	-81.7704	SC	US
Wendy's	Batesburg	33.9134	-81.5333	SC	US
Pizza Hut	Paragould	36.0611	-90.5233	AR	US

... (9990 rows omitted)

- 1. Motivation
- **2. Scatter Plot Maps**
- 3. Choropleth Maps



Quick Review

Scatter Plots

The method

`t.scatter(column_for_x, column_for_y)`

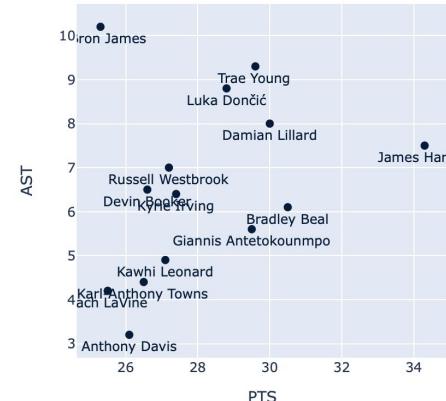
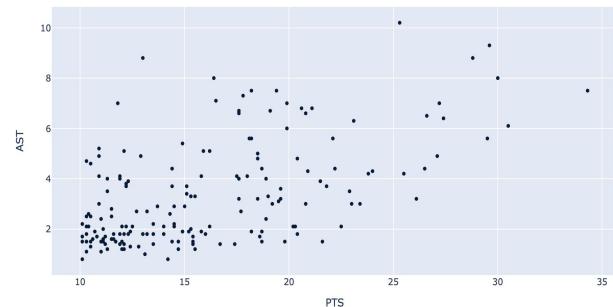
creates a **scatter plot** using the specified columns. Both columns must contain **numerical values**.

Optional arguments, in addition to color `column_for_x`,

`column_for_y`:

- **group** (str): points will be colored according to category in this **categorical column**.
- **labels** (str): points will be labeled according to their value in this column.

```
nba.scatter('PTS', 'AST')
```



• **group()**

The term “**group**” in data science is most commonly associated with **data aggregation** and **disaggregation**.

Aggregation: A process in which information is gathered and expressed in **collective** or **summary** form.

Disaggregation (aka **disentanglement**): A process of taking aggregated data and **breaking it down** into smaller information units.

The method **t.group(column)** counts the number of rows for each unique value in **column**, and returns a two-column table with the results.

In [4]: `restaurants.groupby('name').sort('count', descending=True)`

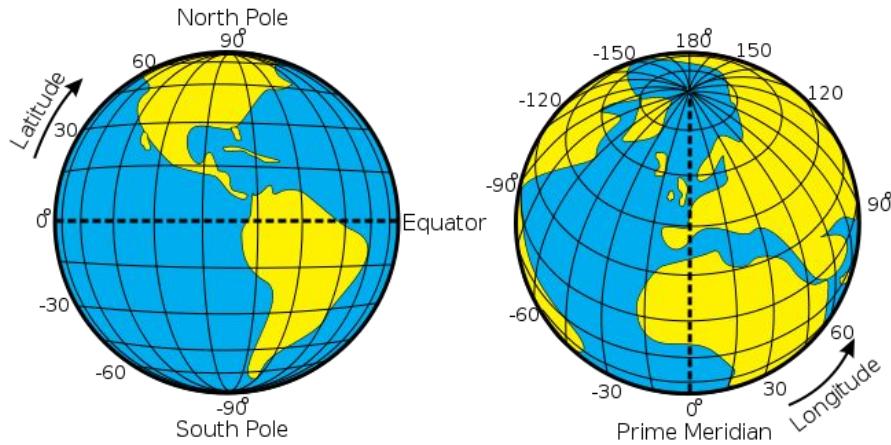
Out[4]:

	name	count
1	McDonald's	1886
2	Burger King	1154
3	Taco Bell	873
4	Wendy's	731
5	Arby's	518

- 1. Motivation
- **2. Scatter Plot Maps**
- 3. Choropleth Maps

Scatter Plot Maps

Latitude and Longitude



Any point on Earth can be described by its latitude and longitude.

- **Latitude** can be thought of as the “y” or “vertical” position.
- **Longitude** can be thought of as the “x” or “horizontal” position.
- When describing a location, latitude always comes before longitude.

What's Wrong with a Scatter Plot?



Scatter Plot Maps

When we want to visualize the **geographic locations** of a lot of data points, it's often helpful to start with a **scatter plot map**.

- Scatter plots with geographic maps
- Help you visualize geographic locations in relation to cities, states, and countries.

Use `px.scatter_geo(df, lat, lon)`
data frame, latitude, longitude

```
In [4]: # Just run me
px.scatter_geo(top_5_restaurants.to_df(),
              lat = 'latitude',
              lon = 'longitude',
              color = 'name',
              locationmode='USA-states',
              scope = 'usa',
              title = "Top 5 Fast Food Chain Locations"
            )
```

Scatters Plot + Map = ❤

Top 5 Fast Food Chain Locations



A Side Note

For our maps we will use a Python library called **Plotly** (`px` in your notebooks). **Plotly** is a commonly used **visualization library** and is really useful for creating maps.

- **Plotly** can be quite confusing and difficult to understand
- The [documentation](#) is really helpful

We **do not** expect you to memorize/remember any **Plotly** syntax. **This is purely for fun.**



Plotly

Optional Arguments

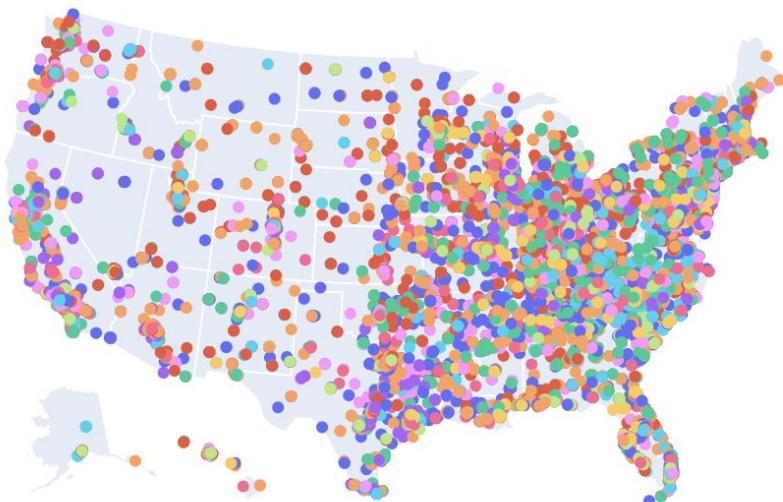
We can **customize** our scatter plot maps by specifying the following arguments:

Argument	Example	Behavior
color	color = 'name'	The colors of points on the map are assigned based on their category in the name column of the data frame
locationmode	locationmode = 'USA-states'	Specifies which map to display. We usually set this to 'USA-states' to show a state map of the US
scope	scope = 'usa'	Specifies the scope of the map (i.e. what is visible). Setting scope to 'usa' zooms the map in just to the US
title	title = 'My Map'	Sets the title of the map

Questions?

Example: All Restaurants

All Restaurants in Our Dataset



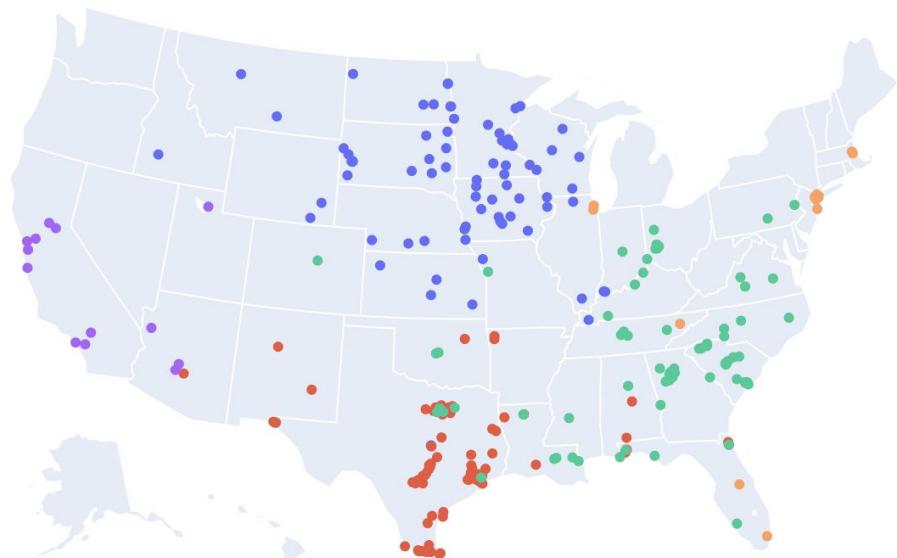
Just because we can plot all restaurants in our dataset doesn't mean we *should*.

name

- McDonald's
- Wendy's
- Frisch's Big Boy
- OMG! Rotisserie
- Domino's Pizza
- Pizza Hut
- SONIC Drive In
- Long John Silver's
- Taco Bell
- Qdoba Mexican Grill
- Sonic Drive-In
- Taco John's
- Whataburger
- Hamburger King
- Burger King
- McDonalds
- Taco Time

Example: Regional Chains

Regional Fast Food Chains Across the US



name

- Taco John's
- Whataburger
- Waffle House
- In-N-Out Burger
- Dunkin' Donuts



- 1. Motivation
- 2. Scatter Plot Maps
- **3. Choropleth Maps**



Choropleth Maps

Choropleth Maps

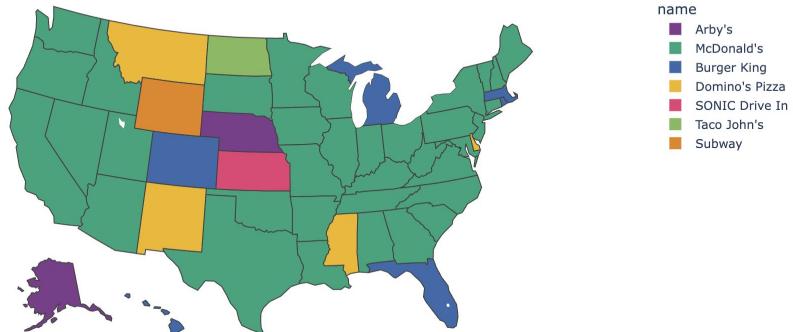
Choropleth maps are useful for visualizing numerical variables across different states or countries. In this sense they are analogous to bar charts, since they encode one categorical variable (state or country) and one numerical variable.

Use `px.choropleth(df, locations)`
data frame, state abbreviations

```
In [21]: px.choropleth(favorite_restaurant.to_df(),
                     locations = 'province',
                     color = 'name',
                     color_discrete_sequence = px.colors.qualitative.Bold,
                     locationmode='USA-states',
                     scope="usa",
                     title = "Most Common Fast Food Chain by State"
                 )
```

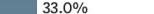
Aggregation!

Most Common Fast Food Chain by State



Example: Election Mapping



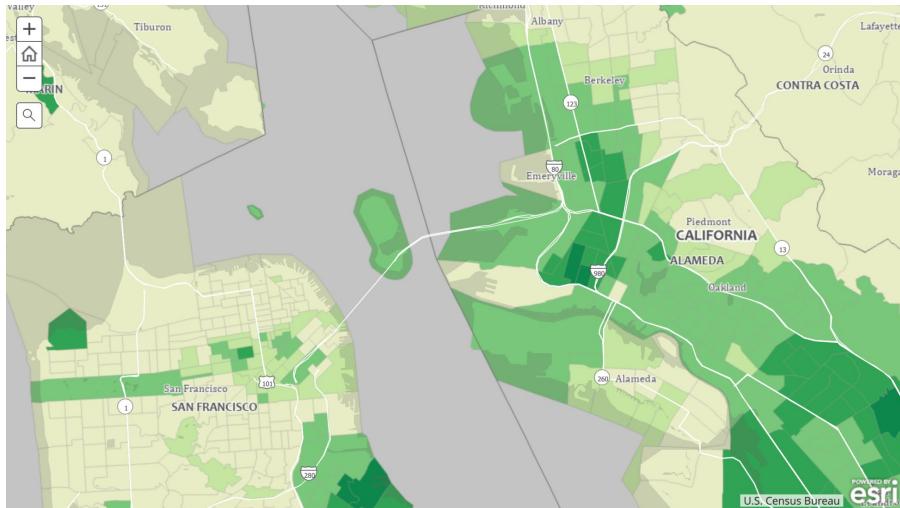
Candidate	Votes	Pct.
Joe Biden 	8,281,504	 65.1%
Donald Trump 	4,204,244	 33.0%

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

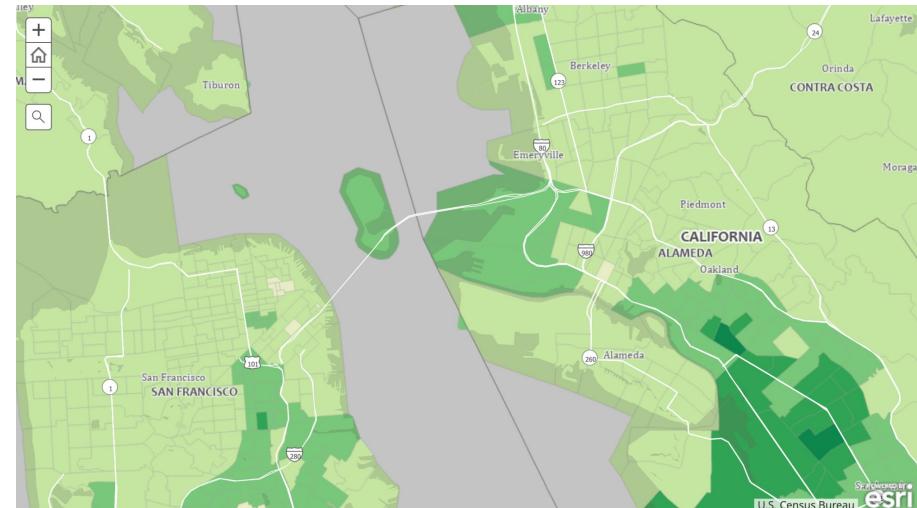
 [SHOW FEWER](#)

Example: Census Data

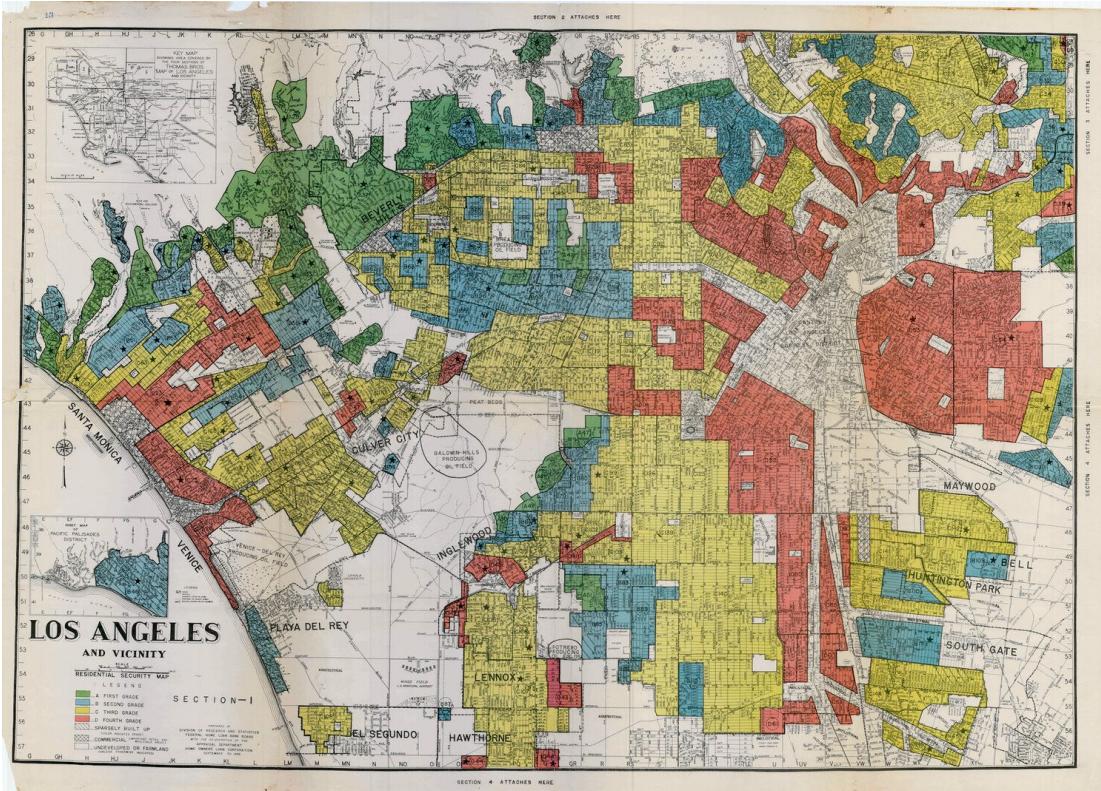
Percent Black or African American



Percent Hispanic or Latino



Example: Redlining



LA "Residential Security Map" from *The Color of Law* (Rothstein)

Plotly Choropleth Maps

We can **customize** our choropleth maps by specifying the following arguments:

Argument	Example	Behavior
<code>color</code>	<code>color = 'name'</code>	The colors of points on the map are assigned based on their category in the <code>name</code> column of the data frame
<code>color_discrete_sequence</code>	<code>color_discrete_sequence = px.colors.qualitative.Bold</code>	Specifies the color palette to use for coloring the categories.
<code>locationmode</code>	<code>locationmode = 'USA-states'</code>	Specifies which map to display. We usually set this to <code>'USA-states'</code> to show a state map of the US
<code>scope</code>	<code>scope = 'usa'</code>	Specifies the scope of the map (i.e. what is visible). Setting scope to <code>'usa'</code> zooms the map in just to the US
<code>title</code>	<code>title = 'My Map'</code>	Sets the title of the map

Colors

There are a lot of options to choose from for **color palettes**.

`px.colors.qualitative.D3`

`px.colors.qualitative.Set2`

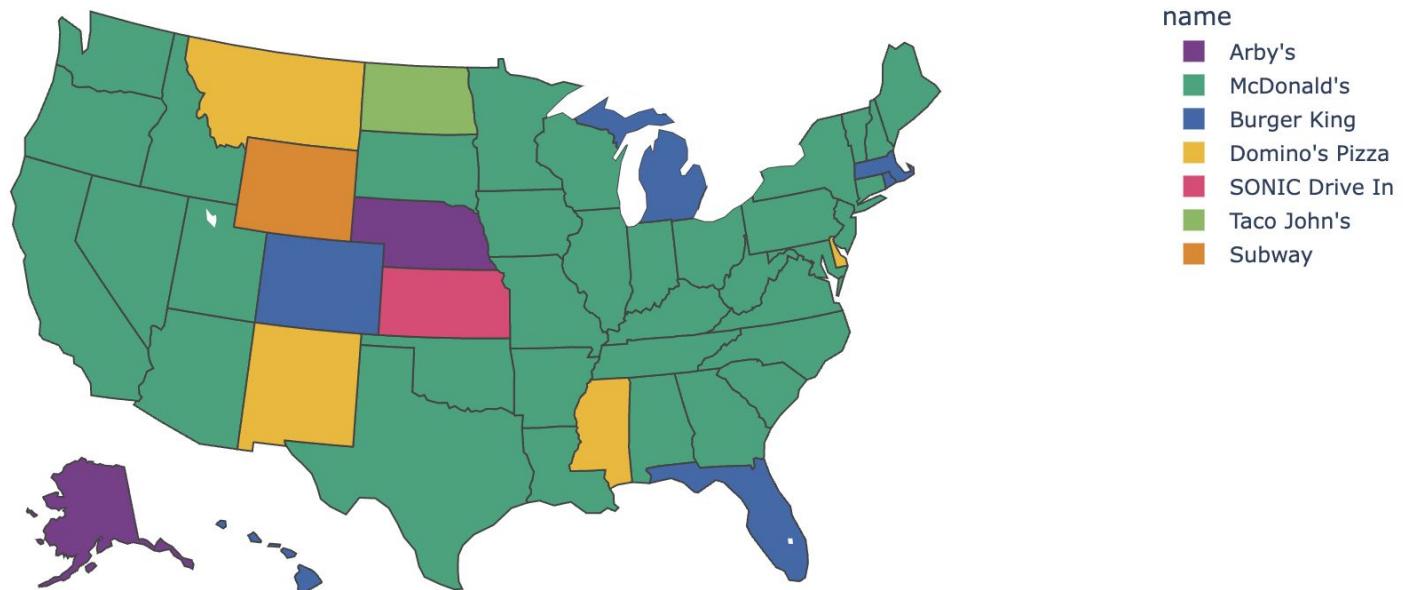
...

What are some considerations we should keep in mind when choosing a color palette?



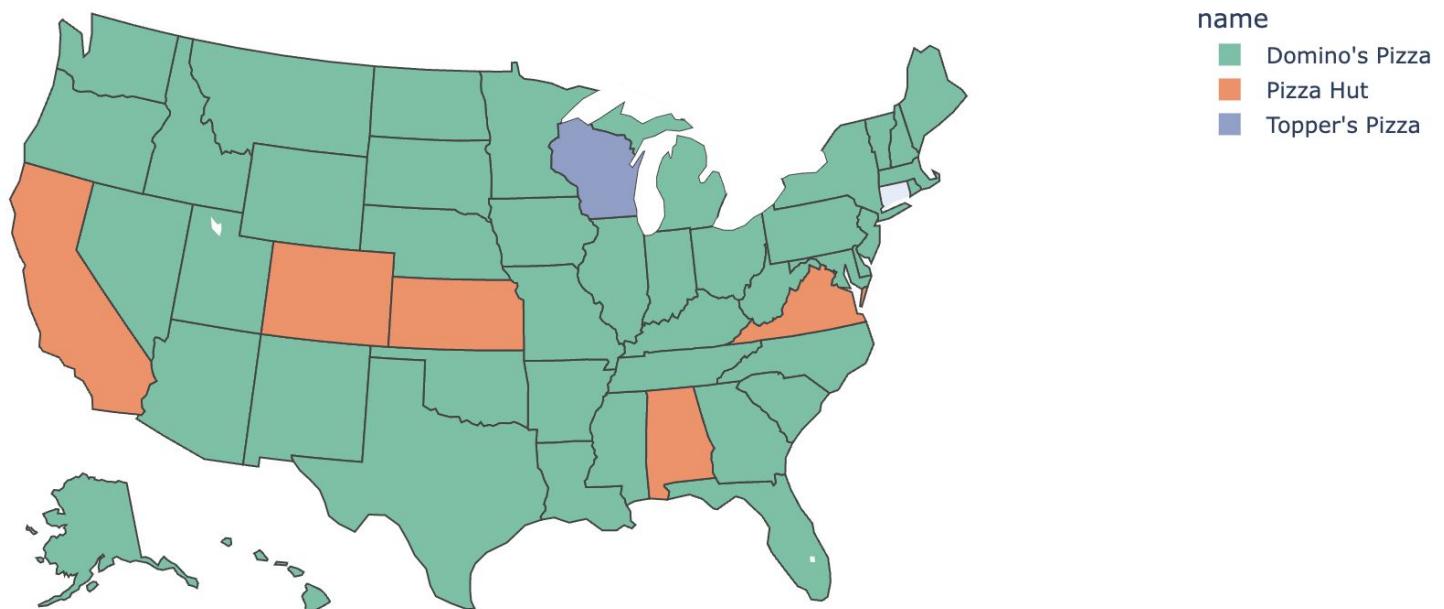
Example: Favorite Chains

Most Common Fast Food Chain by State



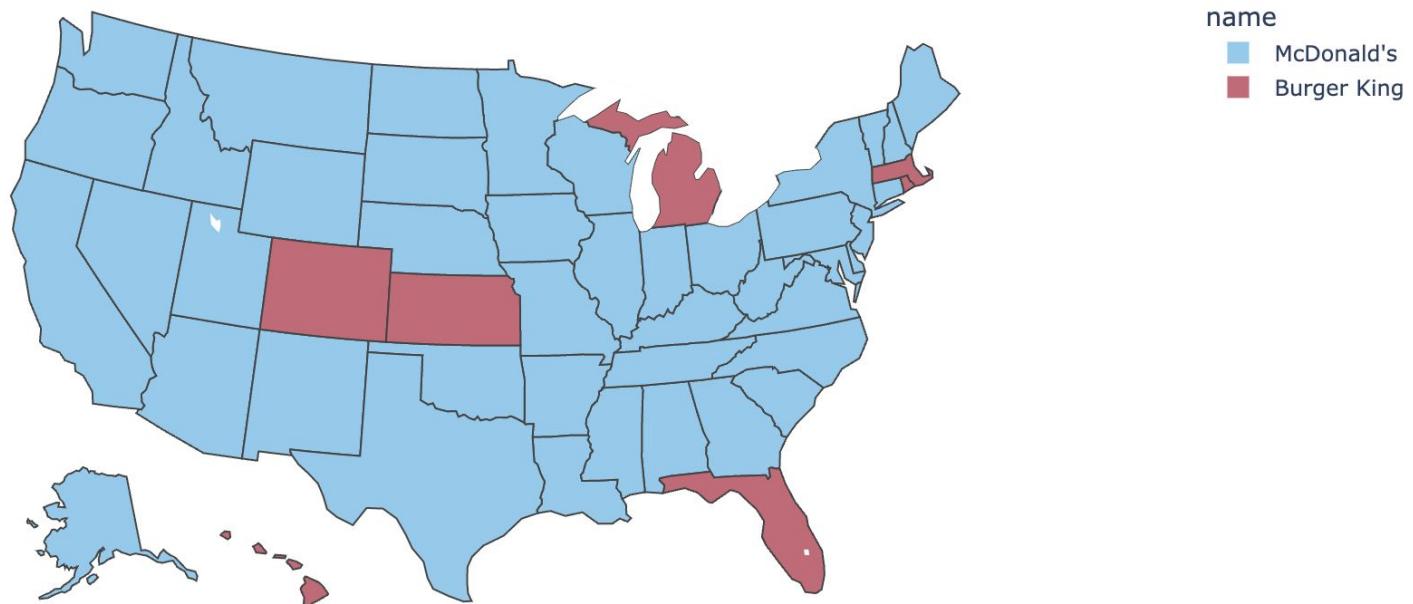
Example: Pizza Chains

Most Common Pizza Chain by State



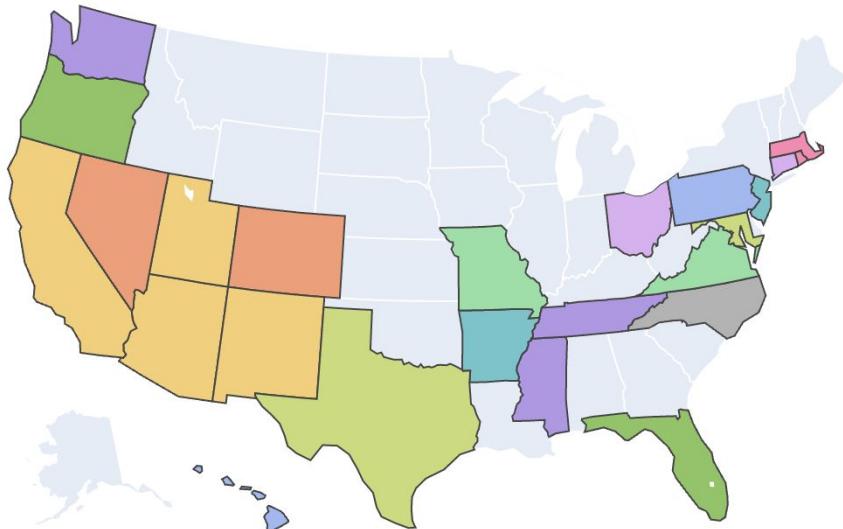
Example: McDonald's vs. Burger King

Does Your State Have More McDonald's or Burger Kings?



Example: Local Burger Chains

Other Burger Chains in the US (Excludes McDonald's, Burger King, and Whataburger)



name

- Backyard Burgers
- In-N-Out Burger
- Good Times Burgers & Frozen Custard
- MOOYAH Burgers, Fries & Shakes
- Flip Burger Bar
- Teddy's Bigger Burgers
- Bobby's Burger Palace
- BGR - The Burger Joint
- Fallguys Burger Company
- Back Yard Burgers
- Bad Daddys Burger Bar
- Rahway Chicken & Burger Inc
- Blake's LotaBurger
- Ernie's All American Burger
- Burger King Salou
- Burgerville
- Checkers Burgers and Fries

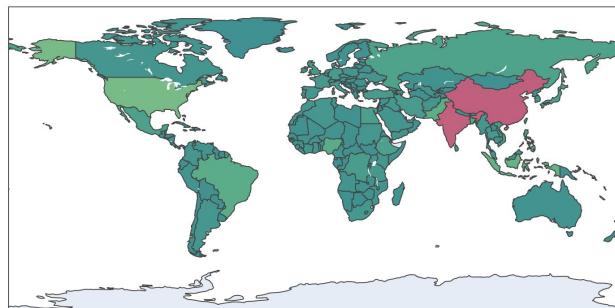
Questions?

In Conclusion...

Summary

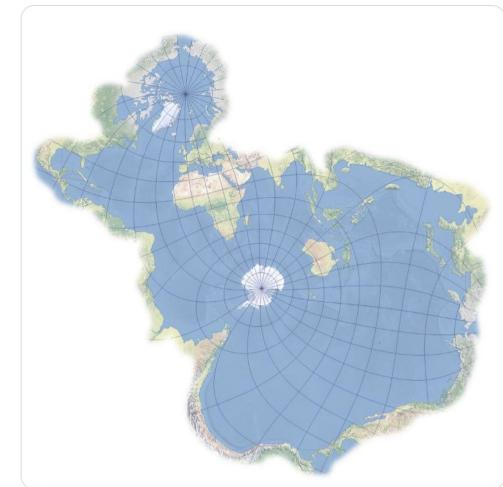
- **Scatter plot maps** are useful when we want to visualize the **geographic locations** of a lot of data points, but it is really easy to **overcrowd** our maps
- **Choropleth maps** allow us to visualize data **aggregated** by county, state, or country
- **Plotly** is a very powerful mapping/visualization library, but you're **not** expected to be an expert at it.

World Population (2022) in Thousands



Terrible Maps
@TerribleMaps

World map according to fish



Map wisely!

Recap

- Maps
 - Scatter maps
 - Choropleth maps

Next Time

- Web Development
 - Making your own website!

Week 4

- **Homework 3** has been released and will be due on **7/28 @ 11 PM**
- **Homework 2** grades have been released on Gradescope. Regrade requests are due by **7/27 @ 6 PM**
- Remember that you can get **3%** extra credit if you have a **mid-semester check-in** with a course staff member
 - See **Ed** for more details

Announcements!