# Data Driven Insights into Ideal Places to Live
DU Project 1 – Team 1 (Jeff Flachman, Pedro Zurita & Thomas Brown)

- **Team:**
  - Jeff Flachman, Pedro Zurita, Thomas Brown

- We want to answer a basic question:
  - Where would I like to live based on the weather in that region?
- People enjoy different temperature ranges
  - Affording activities that fit their favored temperatures
- Identified regions by three categories: cold, hot and temperate climates
  - ChatGPT & Claude.io were used to identify parameters for these categories
  - Our analysis provides a means for others to easily investigate our results and refine for their own preferences


Days 60-80
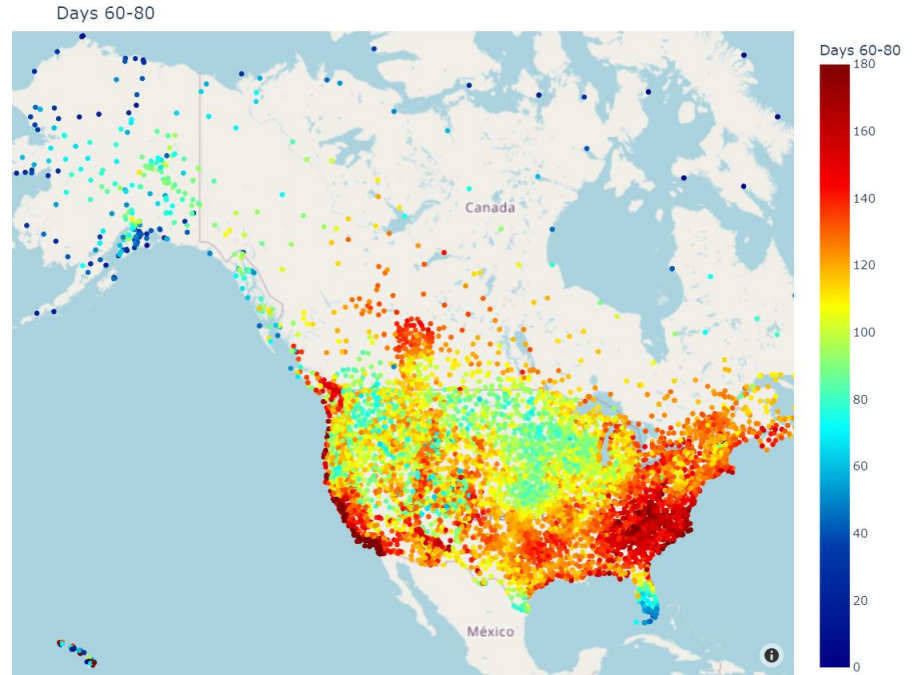
- Goal: **Where would I like to live based on the weather in that region?**

- **Questions** were refined by querying ChatGPT and Claude.ai to define the temperature ranges

- Q1: What are the different climate categories from arid to wet?
  - Arid to Semi-arid: < 100 days of precipitation
  - Humid Continental climates: 100-200 days of precipitation or less

- Q2: For people who like cold weather, what is the ideal temperature?
  - 25°F to 55°F was selected as the desired temperature for people who like colder temperatures
  - An alternate range of 32°F to 65°F was also analyzed.

- Q3: For people who like hot weather, what is the ideal temperature?
  - 75°F to 95°F was selected as the desired temperature for people who like hotter temperatures.

- Q4: For people who like temperate weather, what is the ideal temperature?
  - 60°F to 80°F was selected as the desired temperature for people who like hotter temperature

- Question 5: What temperatures would be considered extreme that people who typically like hot or cold climates?
  - 20°F was selected as the lower comfortable temperature range.
  - 95°F was selected as the upper comfortable temperature range.

- NOAA Global Daily temperatures for 42027 stations were used (37M records)
  - https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/
- Built analysis on counts of daily max temperature for each station
  - Outdoor activity determination often predicated on the days high and chances for rain/snow
- Cleaned data
  - Reduced data to US and Canada
  - Binned data into ranges
  - Performed analysis of days > X, days < X, and X < days < Y
- Plotted all the result on a dynamic map using plotly (plotly.com)
- Analyzed for regions of temperature dynamics
- Assumptions:
  - Missing Precipitation data set to zero.
  - Eliminated data with missing TMAX.
  - Eliminated stations with TMAX count < 300
  - TMIN, TAVG cleaned but not needed for analysis
  - Precipitation is used to inform the idea of rain days or potential cloudy day. This is an imperfect assumption
- Normalized all weather station TMAX (daily highs) to 365 days

- Defined goals and question were refined with ChatGPT

- Cleaned and analyzed data

- Created visualizations

  - Dynamic Geographic visualizations provide the best insight into the data

- Analyzed the visualizations to make our conclusions
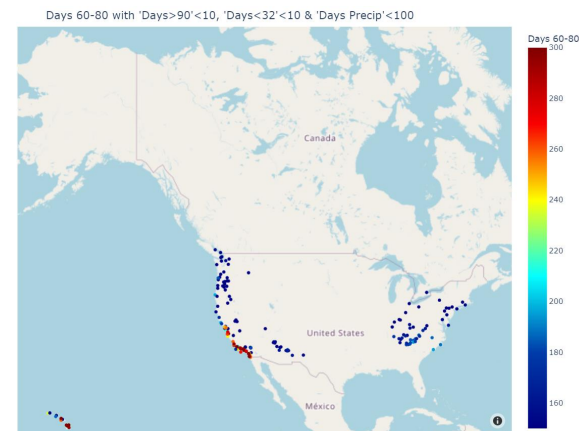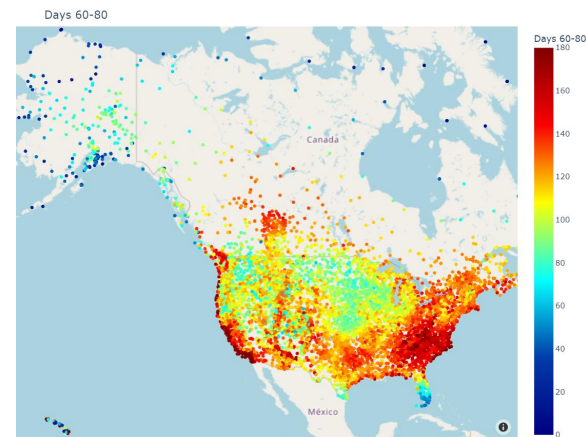
- **Really too cold: Days < 20**
  - Canada, north central US
  - Rocky Mountains in WY, CO
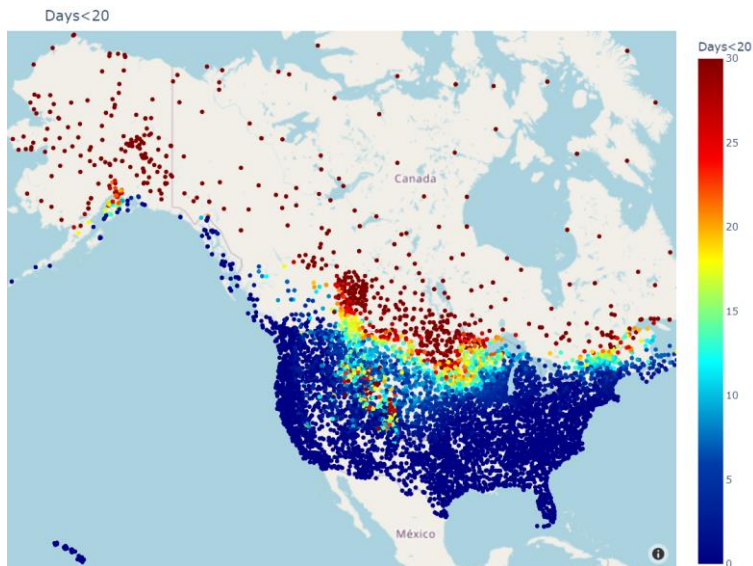
- **Really too hot: Days > 95**
  - Most regions below Lat: ~34
 - See Brown regions
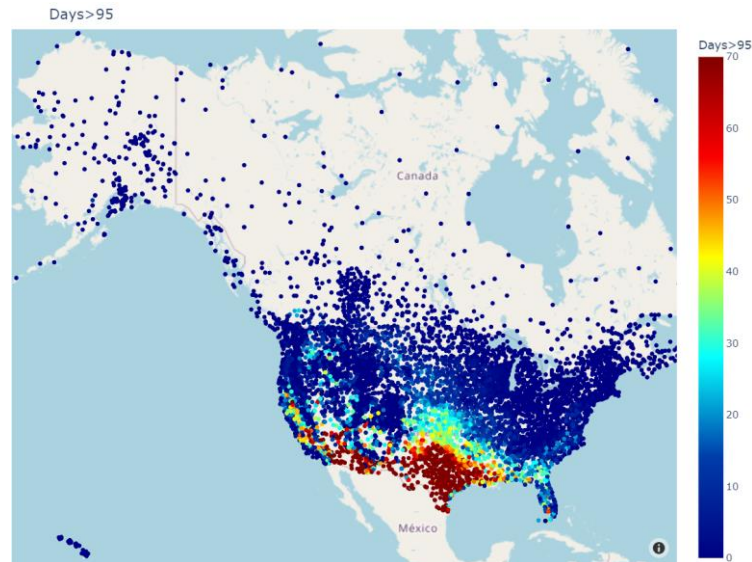
- **Arid Regions:**
  - Southwester US

- **Wet Regions:**
  - Northwestern US and Canada
  - US and Canada East Coast toward the Midwest

- **Median Precipitation**
  - Midwest- Dakotas to East Texas)

Days of Precipitation
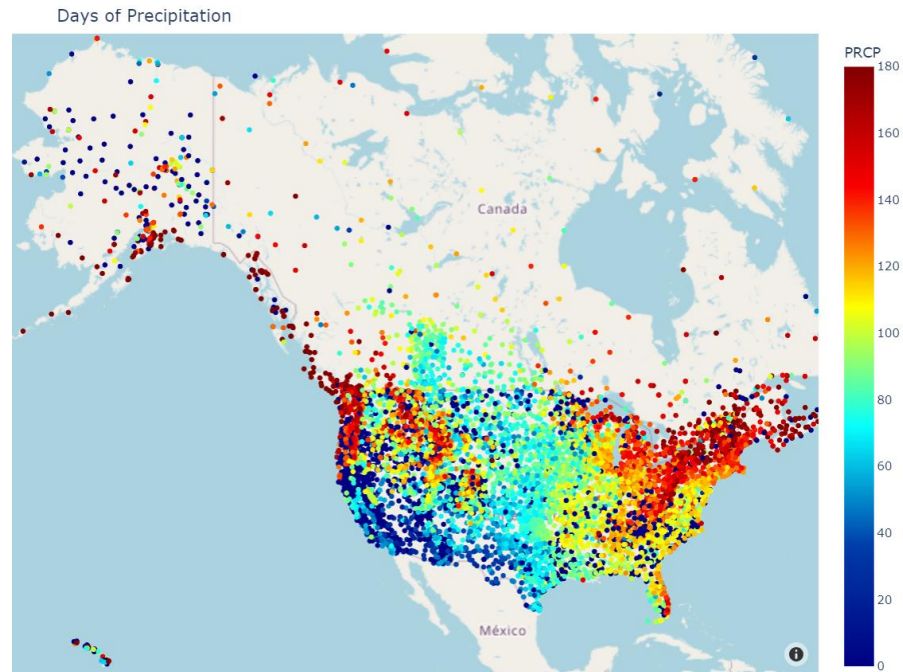
- **Comfort Zones for people who like it <u>cold</u> 25°F to 50°F**
  - Northern US and all of Canada
  - North of Latitude 40
  - Area includes southern states in the west

- **Ideal Cold zones 25°F to 50°F**
  - Fewer than 10 days < 10°F
  - Fewer than 10 days > 80°F
  - Fewer than 200 days of Precipitation
- **Regions, NW US, AK, NE US**

Days 25-55

Days 25-55 with 'Days>80'<10, 'Days<10'<10 & 'Days Precip'<200
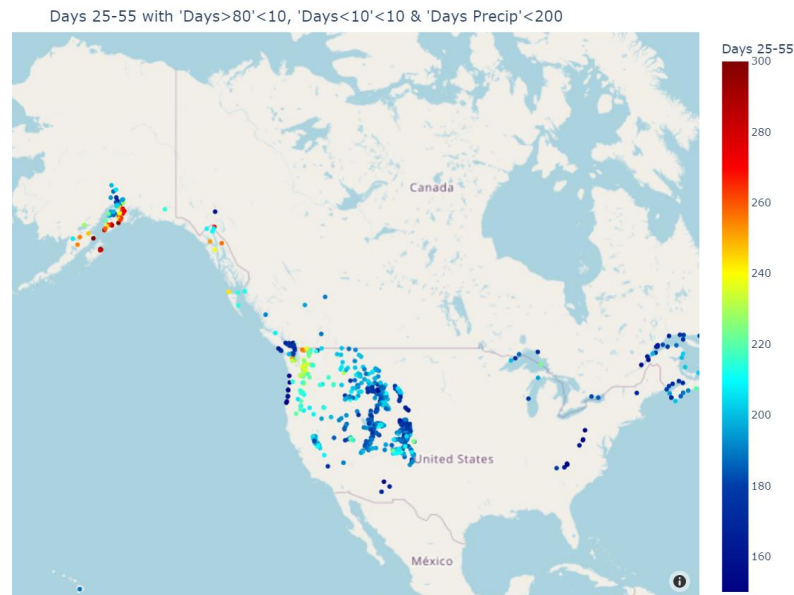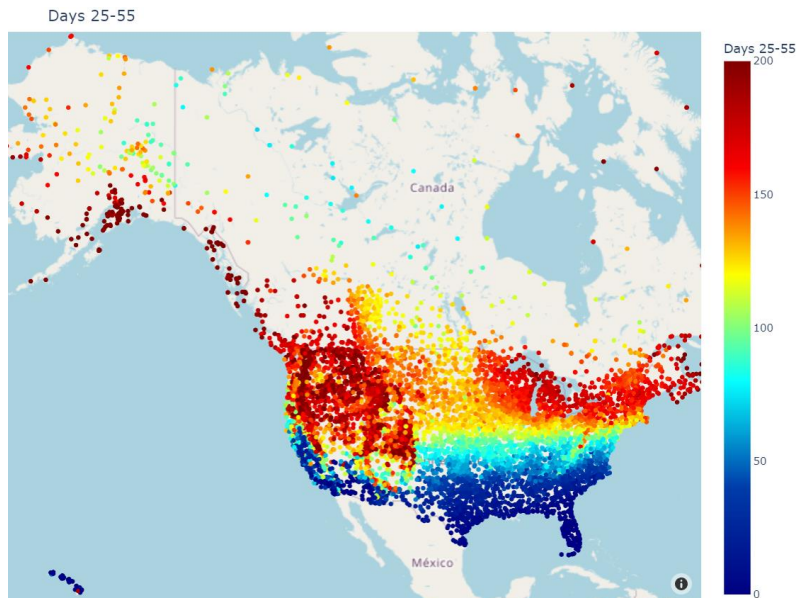
- **Comfort Zones for people who like it <u>hot</u> 75°F to 95°F :**
  - Southeast US (FL, LA, AL, GA, SC, NC, TN)
  - Central Southern CA
  - South TX
  - Hawaii

- **Ideal Hot zones 75°F to 95°F**
  - Fewer than 10 days < 10°F
  - Fewer than 10 days > 95°F
 - ARID - Fewer than 100 days of Precipitation
- **Regions, Southwest coast, Mid SE US, Hawaii**

Days 75-95

Days 75-95 with 'Days>95'<10, 'Days<40'<10 & 'Days Precip'<100
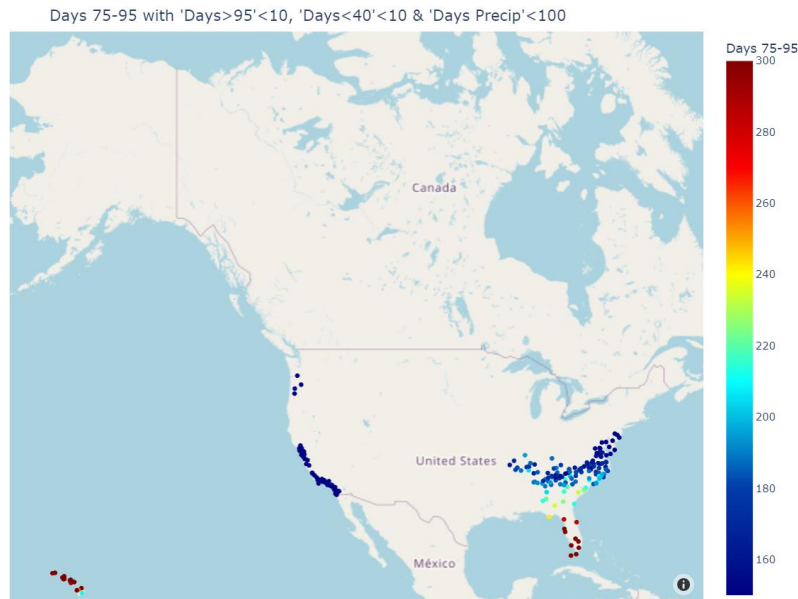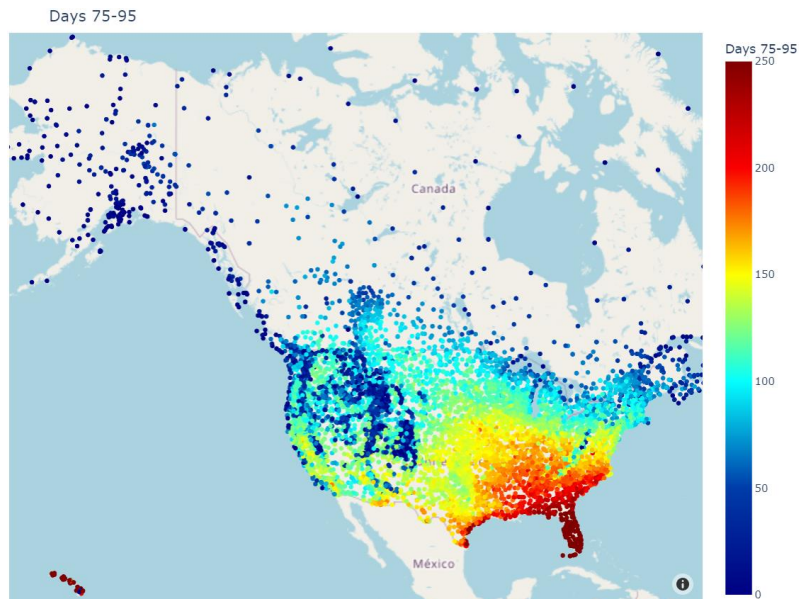
- **Our analysis widened the comfort zone to 60°F to 80°F**
- **Comfort Zones for people who like temperate climates:**
  - Westcoast near the coastline CA, OR, WA
  - Mid-southeast US (VA, WV, NC, SC, TN, GA and N AL)
  - Arizona at elevation (Flagstaff on to SE)
  - HI
  - Canada BC – Vancouver, Edmonton

- Ideal Temperate zones 60°F to 80°F
  - Fewer than 10 days < 32°F
  - Fewer than 10 days > 90°F
  - ARID - Fewer than 100 days of Precipitation
- **Regions: West Coast, HI, North East**

Days 60-80
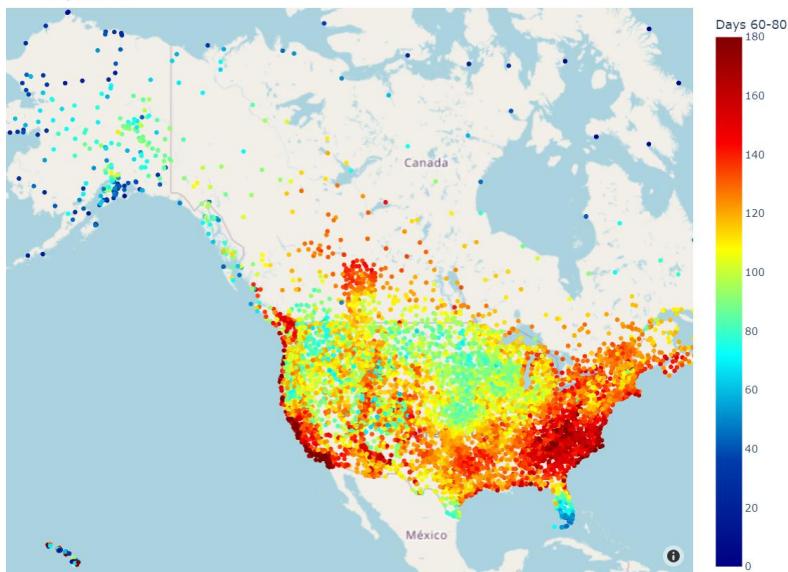
Days 60-80 with 'Days>90'<10, 'Days<32'<10 & 'Days Precip'<100
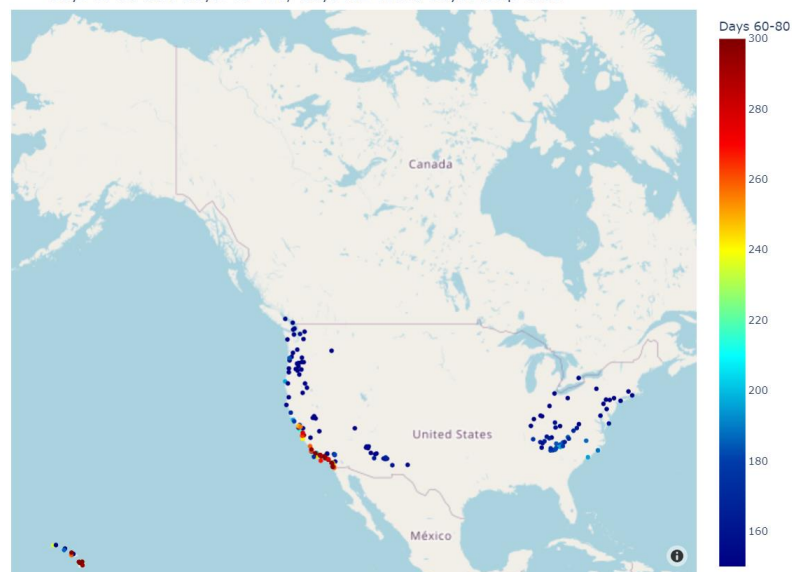
- **Comfort Zones for people who like it hot:**
    - Southeast US (FL, LA, AL, GA, SC, NC, TN)
    - Central Southern CA
    - South TX
- **Comfort Zones for people who like it cold:**
    - Northern US and all of Canada
    - Area includes southern states in the west (NV, UT, CO)
- **Comfort Zones for people who like temperate climates:**
    - Westcoast within a short range from the coastline (CA, OR, W
        - Canada BC – Vancouver, Edmonton
    - Mid-southeast US (VA, WV, NC, SC, TN, GA and N AL)
    - Arizona at elevation (Flagstaff on to SE)
- **Extremes:**
    - Too hot: large number of days >90: Southern US
    - Too cold: large number of days <20: Norther US and Canada

PROBLEMS ENCOUNTERED |

- **Problems with the data we used:**
  - Some bad data was in the dataset (extreme temp outliers)
  - Some weather stations provided limited TMAX data
    - We chose to use only weather stations with at least 300 days of TMAX data
    - We normalized data to 365 days (this is the reason that TMAX days is a float)

- **Problems adding to our analysis**
  - Data by County or Zip Code was limited or costly
  - Spent way too much time searching for additional data to add in that was freely available
  - Ran out of time to integrate mean house price from US Census with zip code Geojson file to create a Choropleth layer coloring counties by median house value
    - See example figure on next slide: plotly choropleth example from plotly.com with counties and unemployment rate

FUTURE CONSIDERATIONS |

- **Future statistical analysis should include NOAA climate normals for cross-year statistical analysis.**
  - Climate normals are three-decade averages of climatological variables, such as temperature and precipitation.
- **Create an interactive tool where users enter temperature range, and a map is generated**
- **Add housing median price by zip code (Choropleth) and overlay weather on top**
  - See example figure on next slide: plotly choropleth example from plotly.com with counties and unemployment rate
- **Add a precipitation overlay (correlation to cloudy)**
- **Show additional visualizations for min or average temperature**
- **Find other sources for environmental impacts by geography: tornados, earthquakes, humidity, pests, etc.**

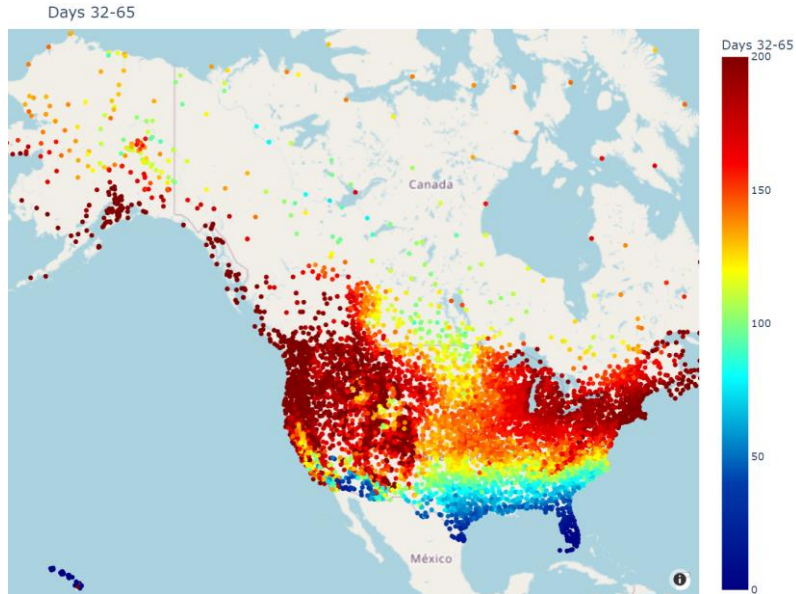Plotly choropleth example from plotly.com with counties and unemployment rate
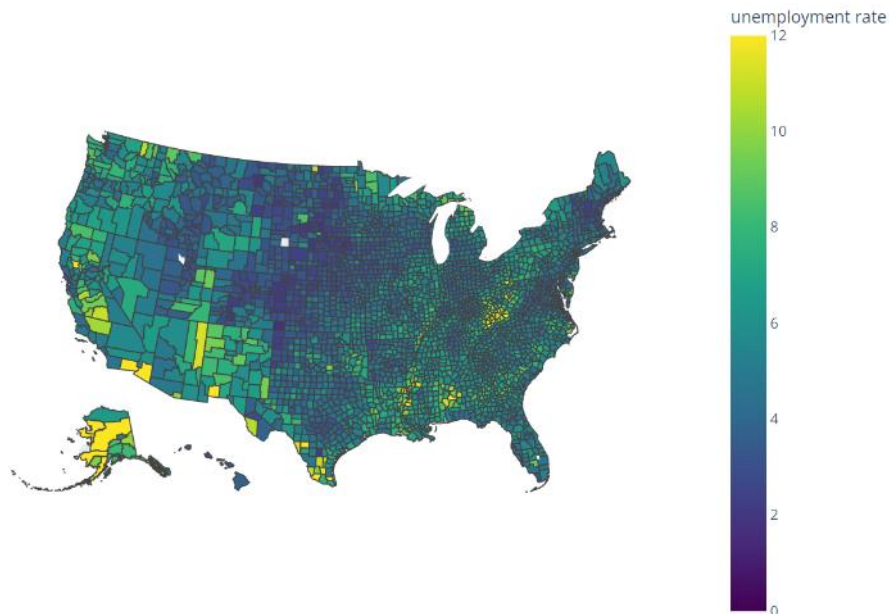
Questions

Backup

- **Comfort Zones for people who like it <u>cold</u> 32°F to 65°F**
  - Northern US and all of Canada
  - North of Lat 35
  - Area includes southern states in the west

- **Ideal Cold zones 32°F to 65°F**
  - Fewer than 10 days < 10°F
  - Fewer than 10 days > 80°F
  - Fewer than 200 days of Precipitation

- **NW US, NE US**

Days 32-65

Ideal Days 32-65 with 'Days>80'<10, 'Days<10'<10 & 'Days Precip'<200