Global Temperature Changes

Process Book
DS 4630/CS 5630
Dr. Paul Rosen

Github Repository

Katelyn Abraham - u0797823@utah.edu
James Crawford - u1220541@utah.edu
Max Terranova - u1416738@utah.edu

Table of Contents:

Background:

Climate change is one of the largest challenges facing the human race in the 21st century[1]. Rising global temperatures, food scarcity, and impacts on infrastructure are changing the world we live in. Though the issue affects all of us, there is still much debate around the subject, preventing actionable steps from being taken to address it. For the last half century[2], the conversation around climate change and global surface temperatures has been growing, but with that growth, public skepticism has grown as well. The public's skepticism can be tied to a misunderstanding of how the data is collected and analyzed, misinformation being pushed for political gain, and/or a general ignorance of the details surrounding the climate change issue[3].

By creating data-based, easy-to-understand visualizations of surface temperatures from across the country and the globe, we hope to make the data clearer and more accessible. The U.S. is both a leading emitter of greenhouse gases and highly vulnerable to climate-related risks, including heatwaves, hurricanes, and droughts. Looking at U.S. climate trends gives a closer view of how global warming shows up at a national scale. By analyzing the raw temperature data, we can see whether long-term changes match what mainstream claims about climate change suggest, while also highlighting the impacts that communities in the U.S. are already facing[4].

Objectives:

Our first objective is to track how average temperatures in the United States have changed from the 19th century through today. Looking this far back lets us contextualize recent warming in the broader scope of history, rather than just the last few decades. By extending the timeline to the early 1800s, we capture key historical periods: the Industrial Revolution, the Second Industrial Revolution, and even the tail end of the Little Ice Age, when cooler conditions still shaped much of the Northern Hemisphere[5]. This longer view makes it possible to see whether the shifts of the past century stand apart from natural variations or fit into a broader global trend. Using the Berkeley Earth dataset, which brings together millions of historical temperature records, we can trace those changes with enough resolution to show how today's climate compares with that of the earlier baseline.

The second objective is to take what we find and make it understandable to people outside of climate science. A lot of climate graphs are either overly technical or stripped down to the point of being unclear. We want to avoid both of those extremes. Our plan is to build straightforward visuals such as line charts, heat maps, and trend plots that show the main story without overwhelming the viewer. The idea is that someone looking at our charts should walk away with a clearer picture of how U.S. temperatures have shifted without needing to dig through a scientific paper to get there.

---

[1] Intergovernmental Panel on Climate Change, Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report (Cambridge: Cambridge University Press, 2021), https://www.ipcc.ch/report/ar6/wg1/

[2] Kieran Mulvaney, "4 Key Moments That Forced Americans to Confront Climate Change," HISTORY, A&E Television Networks, April 19, 2022, updated May 28, 2025, https://www.history.com/articles/climate-change-global-warming-events

[3] John Cook et al., "Consensus on Consensus: A Synthesis of Consensus Estimates on Human-Caused Global Warming," Environmental Research Letters 11, no. 4 (2016): 048002, IOP Publishing, https://repository.library.noaa.gov/view/noaa/62111

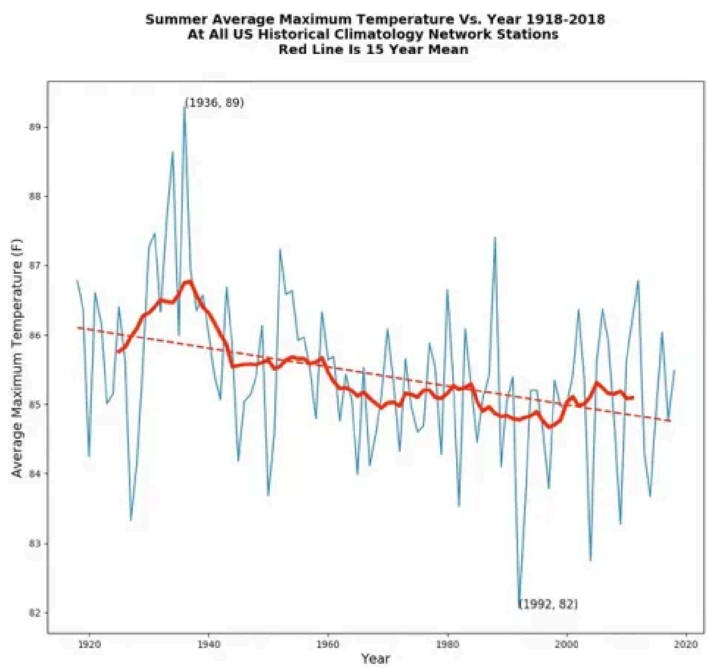
[4] U.S. Global Change Research Program, Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II (Washington, DC: U.S. Government Publishing Office, 2018), https://repository.library.noaa.gov/view/noaa/19487

[5] Emmanuel Le Roy Ladurie, Times of Feast, Times of Famine: A History of Climate Since the Year 1000, trans. Barbara Bray (Garden City, NY: Doubleday, 1971).

Related Work:

USA Temperature: can I sucker you?- This article was included in one of the optional readings for the course. It highlights how you can manipulate data to make it look like the overall temperature in the US is going down instead of increasing with global warning. It was motivating for us because it shows how misinformation about climate change continues to persist due to misleading visualizations. From the article:

> There's a graph going around the internet from Steve Goddard a.k.a. Tony Heller, claiming to show that temperature in the U.S. has been declining, using only high temperatures, using only summertime temperatures, using only data since 1918, based on a simple average without taking into account new stations coming online or old stations retiring or area-weighting or any of that "expert" stuff.



Data:

- Data set sources:
  - Kaggle: Climate Change: Earth Surface Temperature Data
    - This data set contains monthly data of Earth temperatures. It offers two csv files that will be used. One is for state data and the other for city.
  - Kaggle: Average Monthly Temperature by US State
    - Covers the average monthly temperatures of each US state from Jan 1950 to August 2022.
    - Might use to fill in any missing data points

For the data visualizations we need three different data sets outputted as csv files:

1. State Monthly Averages (1800-2013)

a. Use: This data set is going to be used on the national map for the US when the user wants to filter the map by the monthly temperature average instead of the yearly average.

b. Missing data & Cleanup, data set: Climate Change: Earth Surface Temp

   i. From 1850 to 2013, only Alaska and Hawaii are missing monthly temperatures.

   ii. All of the NaN values are in the columns for average temp and average temp uncertainty. Most of those values and the most incomplete states are from western states that were sparsely populated in the early 1800s

      1.
```
na_counts = us_modern_ds['AverageTemperature'].isna().groupby(us_modern_ds['State']).sum()
print(na_counts)
```
```
State
Alabama                   0
Alaska                   76
Arizona                   1
Arkansas                  0
California                0
Colorado                  1
Connecticut               0
Delaware                  0
District Of Columbia      0
Florida                   0
Georgia (State)           0
Hawaii                    2
Idaho                    25
Illinois                  0
Indiana                   0
Iowa                      0
Kansas                  166
Kentucky                  0
Louisiana               180
Maine                     0
Maryland                  0
Massachusetts             0
Michigan                  0
Minnesota                 0
Mississippi               0
Missouri                  0
Montana                 166
Nebraska                166
Nevada                   59
New Hampshire             0
New Jersey                0
New Mexico               10
New York                  0
North Carolina            0
North Dakota            166
Ohio                      0
Oklahoma                240
Oregon                   47
Pennsylvania              0
Rhode Island              0
South Carolina            0
South Dakota            166
Tennessee                 0
Texas                     0
Utah                     95
Vermont                   0
Virginia                  0
Washington               47
West Virginia             0
Wisconsin                 0
Wyoming                 221
Name: AverageTemperature, dtype: int64
```

   iii. After trimming the original data set to just temperature for the United States from 1800-2013, we have a remaining 126,213 data points and 1,824 null values for the average temperatures and average temperature uncertainty. For the 214 years in the range we're using for 51 states, there should be 2568 months for each location and 130,968 rows overall. To fill in those missing points see the code below and the confirmation that after running it, we had the correct number of rows.

```
#Create full monthly date range from Jan 1800 to Dec 2013
full_dates = pd.date_range(
    start='1800-01-01',
    end='2013-12-31',
    freq='MS'
)

#Reindex for each state, without the warning
stateds_filled = (
    us_modern_ds
    .groupby('State', group_keys=False)
    .apply(lambda g: (
        g.set_index('dt')
         .reindex(full_dates)
         .assign(State=g['State'].iloc[0])
         .reset_index()
         .rename(columns={'index': 'dt'})
    ))
)
```

Show hidden output

```
print(stateds_filled.count())
```

```
dt                               130968
AverageTemperature               124379
AverageTemperatureUncertainty    124379
State                            130968
Country                          126213
dtype: int64
```

1.

iv. The average temperature in the original data was recorded in Celcius, so we added a new column with the temperatures in Fahrenheit.

```
stateds_filled['AverageTemperature_F']=stateds_filled['AverageTemperature']*(9/5)+32
```

1.

2. State Yearly Averages (1800-2013)
    a. Use: This data set will be used to generate the national map view of the US on our main page. In that visualization each state will be color coded by the average temperature for a whole year.
    b. Missing data, data set: Climate Change: Earth Surface Temp
        i. Starting where we left off on the data set for monthly averages by state, we extracted the monthly averages to get a yearly average for each state.

```
# Step 1: make sure we have a 'Year' column
stateds_filled['Year'] = stateds_filled['dt'].dt.year

# Step 2: compute the yearly average temperature per state
yearly_avg = (
    stateds_filled
    .groupby(['State', 'Year'], as_index=False)['AverageTemperature_F']
    .mean()
    .rename(columns={'AverageTemperature_F': 'AvgTemp_F_Yearly'})
)

# Step 3: merge that yearly average back to your main DataFrame
stateds_filled = stateds_filled.merge(yearly_avg, on=['State', 'Year'], how='left')
```

1.

ii. After getting a yearly average, created a new data set from the old one that only contains the year, yearly average in F and C, and the state.

```
state_year_avg = stateds_filled[['State', 'Year', 'AvgTemp_F_Yearly']].drop_duplicates()
state_year_avg['AvgTemp_C_Yearly'] = (state_year_avg['AvgTemp_F_Yearly'] - 32) * (5/9)
state_year_avg = state_year_avg.sort_values(['State', 'Year']).reset_index(drop=True)
state_year_avg.head()

#new DS for just the state yearly average
```

	State	Year	AvgTemp_F_Yearly	AvgTemp_C_Yearly
0	Alabama	1800	62.75210	17.084500
1	Alabama	1801	63.35000	17.416667
2	Alabama	1802	63.67145	17.595250
3	Alabama	1803	63.41330	17.451833
4	Alabama	1804	63.47480	17.486000

1.

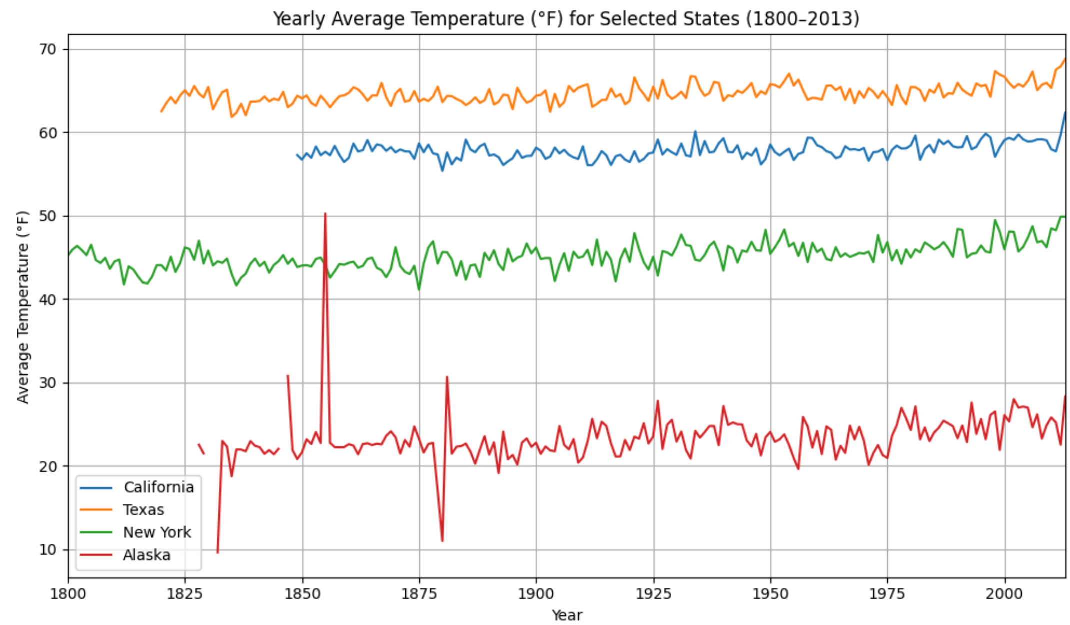
iii. Manually added the yearly average for Alaska and Hawaii for 2013 using data from the National Centers for Environmental Information: ncei.noaa.gov/

```
state_year_avg.loc[
    (state_year_avg['State'] == 'Alaska') & (state_year_avg['Year'] == 2013),
    'AvgTemp_F_Yearly'
] = 26.2

state_year_avg.loc[
    (state_year_avg['State'] == 'Hawaii') & (state_year_avg['Year'] == 2013),
    'AvgTemp_F_Yearly'
] = 65.7
```

1.

c. Initial Results:



i. As expected, some western states did not have yearly data until much later, but there were some anomalies in Alaska. Looking back through how the yearly averages were calculated, we realized that the function .mean() ignores NaN values, so for 1855 Alaska's yearly average was based on the only two values recorded for that year.

ii. To fix the issue, updated the code for calculating the yearly average to set the yearly average to NaN if there's missing data for that year

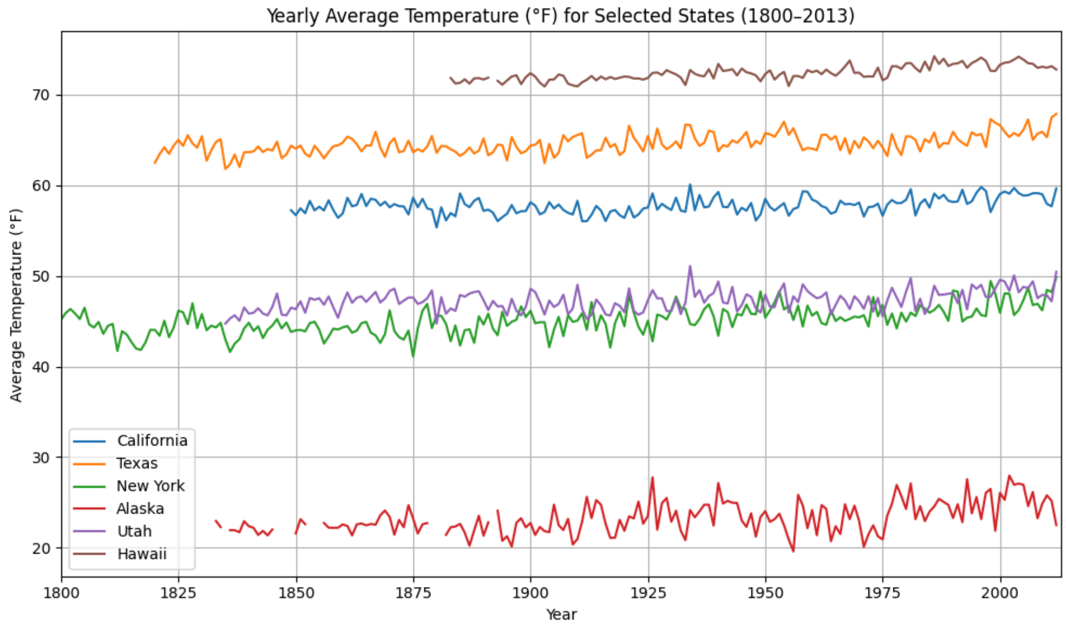
1.

```
# Step 1: make sure we have a 'Year' column
stateds_filled['Year'] = stateds_filled['dt'].dt.year

# Step 2: compute the yearly average temperature per state
yearly_avg_ds = (
    stateds_filled.groupby(['State', 'Year'], as_index=False)
    .agg({
        'AverageTemperature_F': lambda x: x.mean() if x.notna().all() else float('nan')
    })
    .rename(columns={'AverageTemperature_F': 'AvgTemp_F_Yearly'})
)

stateds_filled = stateds_filled.merge(yearly_avg_ds, on=['State', 'Year'], how='left')
```

d. Final results:



i.

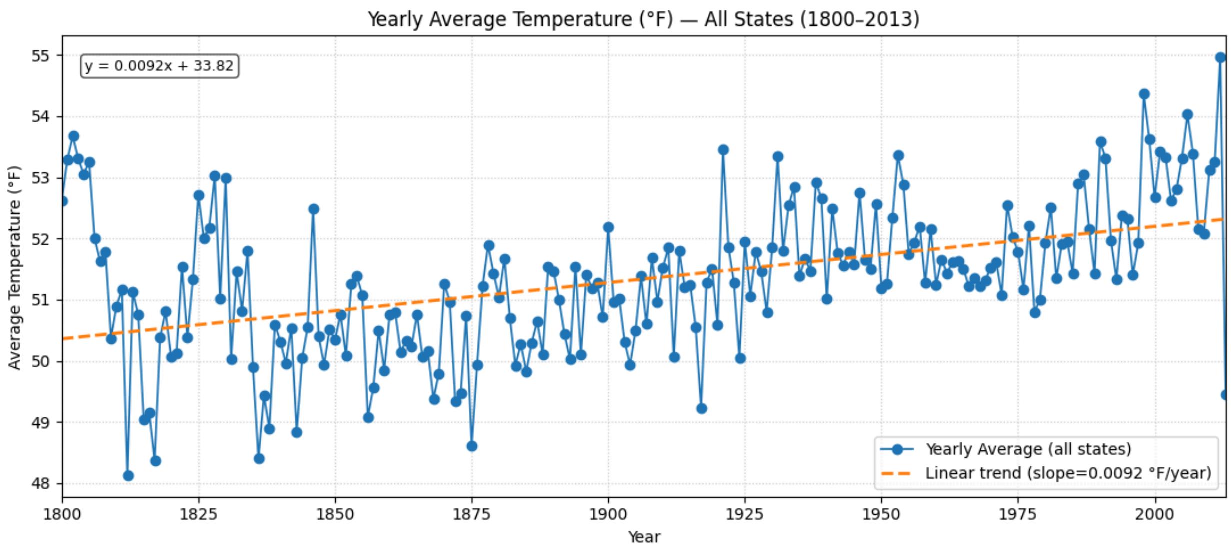
3. City Monthly Averages (1800-2013)
    a. Use: For use in the State page, shows the average temperatures of different cities throughout the US. Using the lat/ long data, map them directly to their location in the state.
    b. Missing data, Data cleaning: The dataset initially contained data for all major cities (ans some smaller suburbs) globally, constituting over 8.5 million different temperature readings. As this project only contains analysis of US values after the year 1800, all non-US cities and readings dating prior to 1800 were culled. This produced a dataset of approximately 700k temperature readings.

       The dataset did not contain state information, only date, city, country, latitude, and longitude, the latter two being approximations, where the values were rounded to two significant figures. This meant the state names and abbreviations needed to be mapped to each of the values. Initially, the state names mapping was attempted using a geospatial approach, where the longitude and latitude were used to assign state values. After this approach was implemented, errors were found, due to the aforementioned rounding of the geographic coordinates and the mapping method was scrapped. Next, a manual mapping option was chosen. Since there were less than 300 US cities represented in the dataset, with no duplicate cities (cities with the same name as other cities in different states), each city was mapped to its corresponding states, with the help of OpenAI's ChatGPT.

Exploratory Data Analysis:

Initially, we used a line graph to look at the average yearly and monthly temperatures. From the line graphs of different states, we learned when records for temperatures started being recorded. Some states had data all the way back to 1800, but California didn't start until the 1850s. Hawaii didn't start until after 1875. For our design, this gave us some insight into what additional information we needed to include in the state pages.

After looking at the averages for the individual states, we created a line graph for the average yearly temperatures for the whole country. By adding a trend line to the data, it was easier to see the overall trend in average temperatures. From this, we learned the slope was a 0.0092 ℉ increase per year. Since this visualization made the trend much easier to see, we decided to add a trend line to all of the line graphs on the finished visualizations.



Design Evolution:

What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?

Our design had two main components: A national map view of the whole U.S. and individual state views.

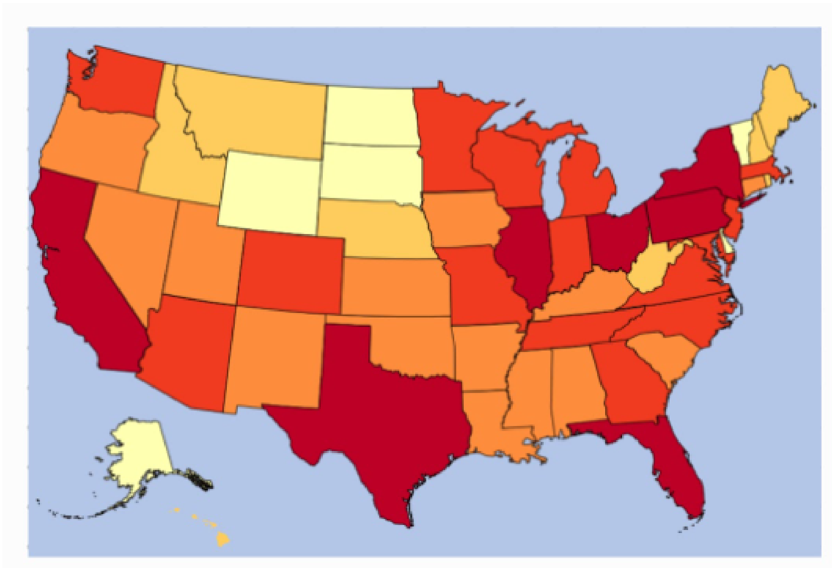
National View:

For the national view, we proposed a heat map of the whole U.S. with each state color coded for the average temperature. The purpose of this project was to show changes in average temperature across the US, so using a heat map was the most natural approach. This method allows users to quickly compare different states using a sequential colorscale. To cut down on the amount of color and to create sufficient luminance contrast we placed the map on a solid, neutral background. The colorscale we used was also limited to a smaller range to make each color more distinguishable.

We proposed a few interactive features to give the user the ability to filter between yearly and monthly data using a sliding scale that updates the map in real time. Giving the user more freedom to explore the visualization helps with both engagement and comprehension of the data being shown. To provide more granular information on the temperatures we planned to include a tool-tip, so users can hover over a state to get the exact average. This will also make the map more accessible for people with degrees of color blindness.

NOTE: Update throughout the project as we deviate from the original proposal

Proposed Map:

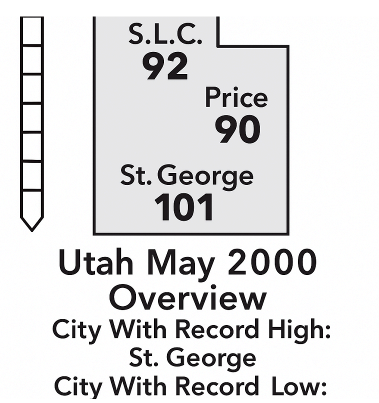


Final Map:

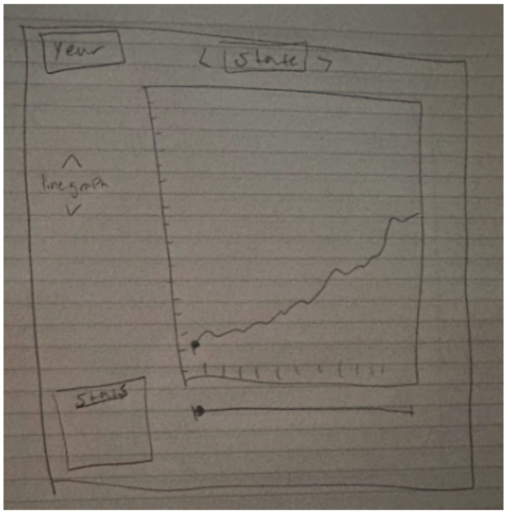
State View:

The proposed state view provides a mini-map for an individual state. The colors are consistent with the national map, so the colors have the same meaning for the two pages. We plan to include extra data on the state maps with average temperature for major cities and their locations on the map. The cities will not be color coded to cut down on the amount of color used on the map. In addition to the map, we plan to include a line graph to provide an easy way for the user to see how the state's monthly or yearly averages change over time. A line graph was ideal because it allows for a clear representation of changes over time. An additional trend line in the graph also shows the user how the temperatures have been steadily increasing.

Proposed Map:



Proposed Line Graph:

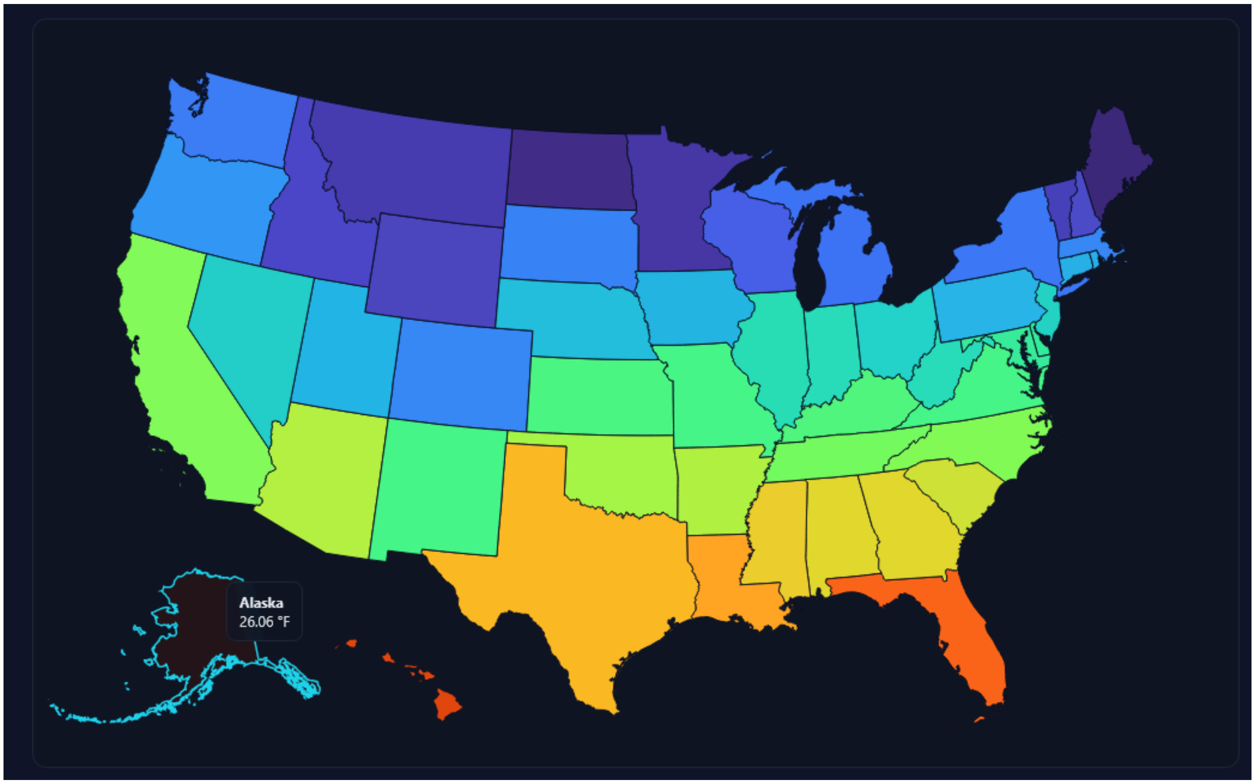


Implementation:

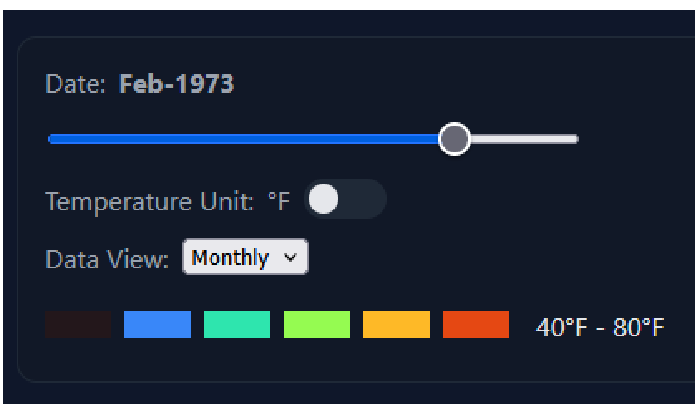
Landing Page: National View:

- 10/19/2025 - 10/25/2025

For our initial set up we started with the landing page. The landing page shows temperature changes for the whole U.S. in one map. Before worrying about the data sets, we began with a rough outline of the map with a few made up data points. To draw the map, we used geographic from D3, geoAlbersUsa and geoPath, to render each state as an SVG path. Once the map was working correctly, we linked it to the actual data sets. For more information on how the data sets were acquired and cleaned, see the data section. For our initial findings, see the exploratory analysis section. While linking the data to the map, we found that the data for Georgia was not being found. After inspecting the csv files we learned that the original data set had labeled the state of Georgia as Georgia (state), likely as a means of differentiating it from the country of Georgia. In the VS code, we use find all and replace to fix all the entries for Georgia in the monthly and yearly averages data sets. The final version is pictured below:



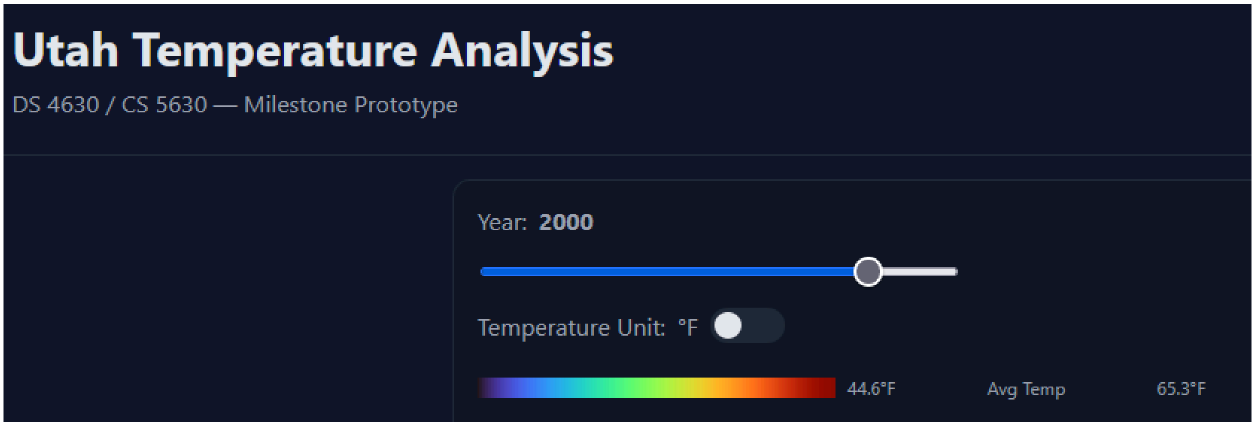
To give users more information about the temperature, we added a tool tip, so they can see the exact average for the state when they hover over the map. Above the map we provided a legend with the color scale. We included a slider to allow users to easily navigate the map over time, from 1800 to 2013. There is a drop down to switch between a map should the average temperature for a whole year or the average temperature for each month. An additional toggle allows them to switch from Fahrenheit and Celsius.

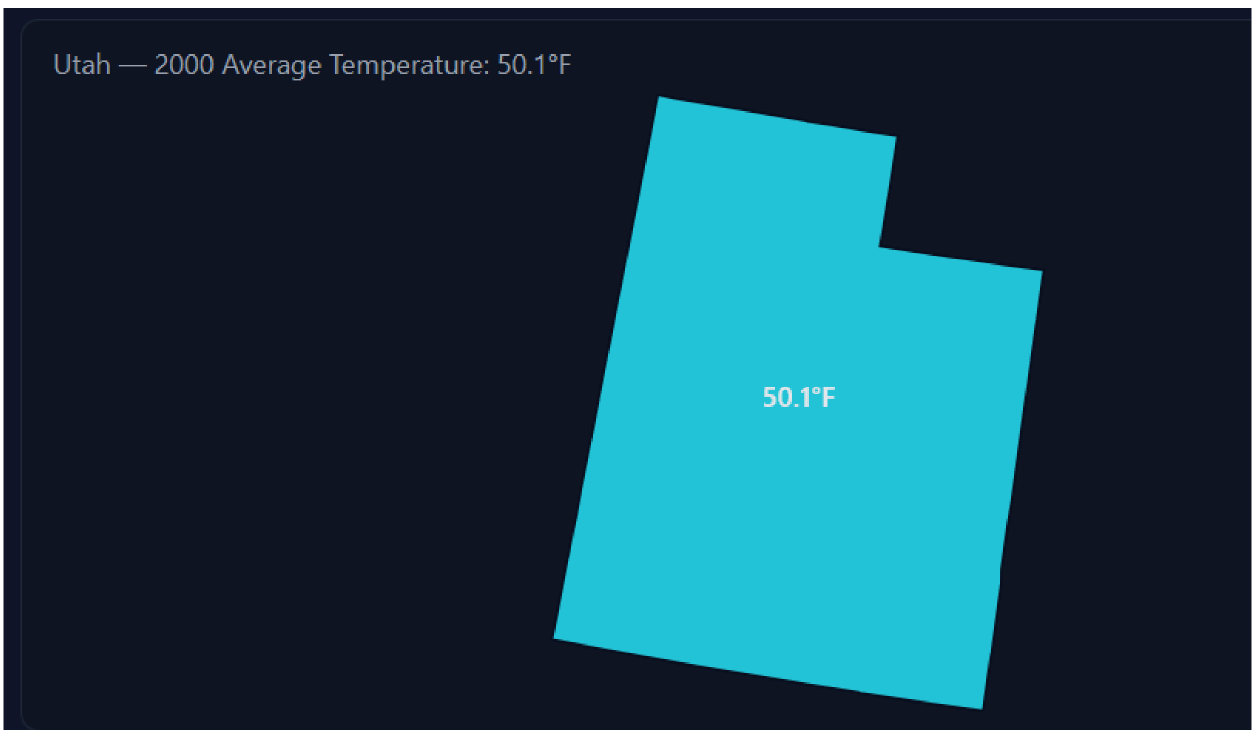


State View:

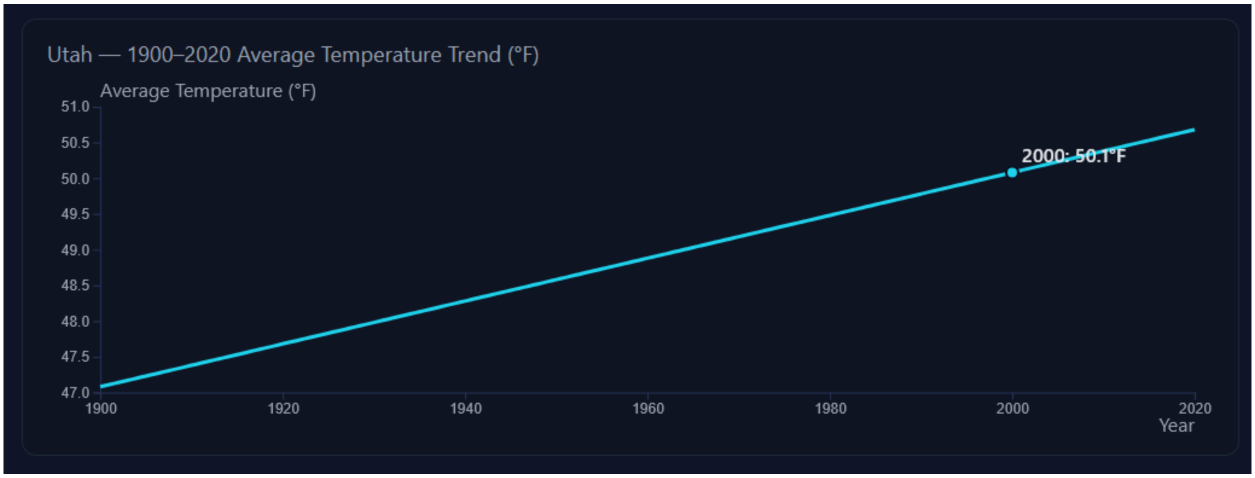
- 10/19/2025 - 10/25/2025

At first we had the state view on the same page as the national, but then we split it into its own page. Our next step is to link the two pages, so that when a user clicks on a state they are redirected to the state view page. We also need to link it to the data set for the city averages and then add those cities to the map. At this stage, the state page highlights the selected state into a map view of just that state. Above the map, we've included a slider for the user to change the year and a toggle to go between Fahrenheit and Celsius.





Below the map view, we've started with a simple line graph that shows the trend in temperatures for the selected state. On the line, we've included a point of interest, so the user can see the selected year.



<u>Evaluation:</u>

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

NOTE: The evaluation will be completed once we've finished the project.