

DATA DRIVEN INSIGHTS INTO IDEAL PLACES TO LIVE

DU AI Bootcamp: Project 1 Report

Abstract

An examination of daily highs and precipitation days across 7800 weather stations in the US and Canada to determine regions where people who like cold, hot, or temperate climates might want to live

Team: Jeff Flachman, Pedro Zurita, Thomas Brown

Table of Contents

1	Project Overview.....	1
1.1	Executive Summary	1
1.2	Goals & Questions	1
1.2.1	Goals.....	1
1.2.2	Questions	1
1.3	Approach	2
1.4	Data Collection, Cleanup & Exploration.....	3
1.4.1	Jupyter files:	4
1.4.2	Data Cleaning	4
1.4.3	Analysis.....	5
1.4.4	Assumptions:.....	5
2	Results & Conclusions	5
2.1	Precipitation Regions (ARID vs Wet, Dry vs Humid).....	6
2.2	Cold Climates	7
2.2.1	Cold Regions (25°F to 55°F)	7
2.2.2	Alternate Cold Regions (32°F to 65°F)	9
2.3	Hot Climates (75°F to 95°F).....	11
2.4	Temperate Climates (60°F to 80°F)	13
2.5	Areas of Extreme Temperatures	15
3	Summary.....	17
3.1	Analysis Summary.....	17
3.1.1	Question 1: What are the different climate categories from arid to wet?	17
3.1.2	Question 2: For people who like cold weather, what is the ideal temperature?	17
3.1.3	Question 3: For people who like hot weather, what is the ideal temperature?.....	18
3.1.4	Question 4: For people who like temperate weather, what is the ideal temperature?	18
3.1.5	Question 5: What temperatures would be considered extreme that people who typically like hot or cold climates would not like?.....	18
3.2	Overall Conclusions	18
4	Problems Encountered.....	19
4.1	General Problems	19
4.2	Problems with the data used:	19
4.3	Problems adding sources to the analysis.	20

5	Future Considerations.....	21
6	References	21
6.1	NOAA Data used	21
6.2	Question 1: Arid vs Wet Climates	22
6.3	Question 2: Cold Climates.....	24
6.4	Question 3: Hot Climates	26
6.5	Question 4: Temperate Climates	27
6.6	Question 5: Extreme Temperatures.....	29

Table of Figures

Figure 1: #Days between 60°F to 80°F	3
Figure 2: Ideal #Days between 60°F to 80°F	3
Figure 3: Days of Precipitation	6
Figure 4: # Days between 25°F to 55°F	7
Figure 5: Ideal # Days between 25°F to 55°F	8
Figure 6: # Days between 32°F to 65°F	9
Figure 7: Ideal # Days between 32°F to 65°F	10
Figure 8: # Days between 75°F to 95°F	11
Figure 9: Ideal # Days between 75°F to 95°F	12
Figure 10: # Days between 60°F to 80°F	13
Figure 11: Ideal # Days between 60°F to 80°F	14
Figure 12: 3 Days below 20°F	15
Figure 13: # Days above 95°F	16
Figure 14: Choropleth example (source https://plotly.com/python/choropleth-maps/)	20

1 Project Overview

1.1 Executive Summary

This team for DU Project 1 was: Jeff Flachman, Pedro Zurita, Thomas Brown

People like different temperature ranges that allow them to enjoy where they live and the types of activities they do. For this project weather data was examined to identify areas where diverse types of people may choose to live based on their preference for cold, hot, or temperate climates.

ChatGPT and Claude were queried to refine parameters for cold, hot, and temperate climates. These recommendations were discussed, and a range of temperatures was defined for each of the climates. Global weather station data for the US and Canada was pulled, cleaned, and organized.

Because weather is a geographic centric statistic, it was decided to analyze and present the data visually in a geographic format on dynamic plotly maps. The maps in the report contain links to dynamic plotly maps saved on the project github pages at <http://iflachman.github.io>.

1.2 Goals & Questions

1.2.1 Goals

The project goal was to answer a basic question: **Where would I like to live based on the weather in that region?**

There are many sources that talk about cities with great weather, culture, and activities. However, most of their analysis is based on Average temperature information. A location with a low of 70 and a high of 100 might report an average temperature of 85 degrees.

Most people do not decide what you are going to do today based on the average monthly or annual temperature. Activity decisions are based on the high for the day, the max temperature.

Therefore, using the high or max temperature for the day provided the best statistic the analysis. Several questions were defined to drive the analysis.

1.2.2 Questions

To define parameters to support the following questions, two AI sources were consulted: ChatGPT (chatgpt.com) and Claude (claude.ai). ChatGPT provided different answers when asked a second time to confirm the data for this report.

The use of these sources, and the chosen application of the recommendations is subjective. Others may rerun the analysis, altering these parameters to refine the result to better meet their interpretation.

The text of the ChatGPT and Claude interactions and the chosen parameter summaries is included in Appendix A (section 6).

The project questions and the summary of the parameter selections is listed below:

Question 1: What are the different climate categories from arid to wet?

The Summary: For Arid and Semi-Arid climates, 100 days of precipitation or less was used. This is good for hot and temperate climates.

For Humid Continental climates 100-200 days of precipitation or less was used. This may (a subjective answer) be better for colder climates to support winter activities.

Question 2: For people who like cold weather, what is the ideal temperature?

The Summary: **25°F to 55°F** was selected as the desired temperature for people who like colder temperatures based on the initial ChatGPT and Claude responses.

An alternate range of **32°F to 65°F** was selected as the desired temperature for people who like colder temperatures based on the initial ChatGPT and Claude responses.

For the Ideal weather, a more Arid to humid continental climates with less than 200 days of precipitation was used. This allows for snow to support enjoyable cold weather activities.

Question 3: For people who like hot weather, what is the ideal temperature?

The Summary: **75°F to 95°F** was selected as the desired temperature for people who like hotter temperatures based on the initial ChatGPT and Claude responses.

For the Ideal weather, a more Arid to Semi-arid climate with less than 100 days of precipitation was used.

Question 4: For people who like temperate weather, what is the ideal temperature?

The Summary: **60°F to 80°F** was selected as the desired temperature for people who like hotter temperatures based on the initial ChatGPT and Claude responses.

For the Ideal weather, a more Arid to Semi-arid climate with less than 100 days of precipitation was used.

Question 5: What temperatures would be considered extreme that people who typically like hot or cold climates would not like?

The Summary: **20°F** was selected as the point for people who like colder temperatures and will find the cold to be extreme. **95°F** was selected as the point for people who like hotter temperatures and will find the heat to be extreme.

1.3 Approach

The project started by determining an interesting problem to solve. Addressing the following goal was selected: **Where would I like to live based on the weather in that region?**

Questions were defined. Two AI sources, ChatGPT and Claude, were queried to refine the questions into parameters that could be analyzed. Raw weather data was found on the NOAA site. Because weather is a geographic centric statistic, it was determined best to analyze and present the data visually on dynamic plotly maps.

Additional data and analysis were considered such as:

- Grouping weather by county or zip code
- Combining information with median house prices
 - Additional question: where is the most affordable location(s) for a particular climate?

- Overlaying weather data on housing data on a single plotly map visualization

These additional investigations along with others were tabled for future work. Median home prices by zip code are available on the US Census website. However, APIs to convert weather stations to zip codes were expensive or had significant bandwidth limitations. A full list of data sources investigated can be found here: <https://github.com/jflachman/DU-project-1/blob/main/resources.md>

Therefore, the scope defined temperature categories. A count of the number of days a category was met and the number of days of precipitation were computed. These were then plotted with on the map with color defined by the number of days for the given category.

These plots were analyzed visually to identify areas that best addressed the questions (fig 1). In addition, plots were created with more targeted constraints to define subjective ideal locations. For example, Days between 60°F to 80°F, with fewer than 10 days > 90, fewer than 10 days < 32, & fewer than 10 days of precipitation (rain, snow or equivalent) (fig 2).

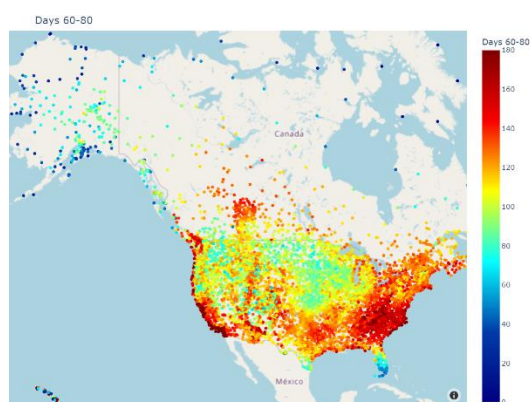


Figure 1: #Days between 60°F to 80°F.

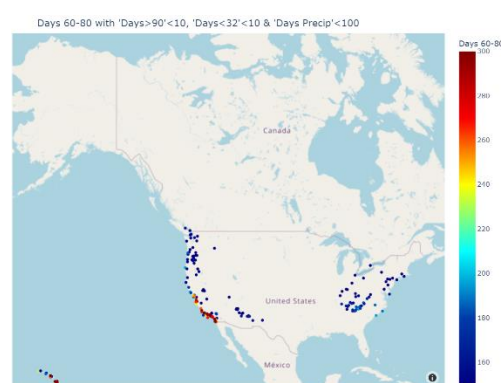


Figure 2: Ideal #Days between 60°F to 80°F.

1.4 Data Collection, Cleanup & Exploration

Daily weather data was found on the National Oceanic and Atmospheric Administration (NOAA) website: <https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/>. The 2023.csv.gz file in the by_year folder was used for all the weather information. The ghcn-stations.txt file was used for the locations (lat/lon) and name of each station. The 2023 data contained more than 37 million records for 42027 ground stations worldwide.

Analysis and visualization were broken into steps with a Jupyter file for each step and the result written to a parquet file in the results_folder. The reason for this was twofold. First, some of the calculations occasionally crashed Jupyter notebook. It was important to complete those calculations and save the work for the next step. Second, separate Jupyter files allowed us to separate work packages for team members.

An examination of file formats was performed to define the best method to save the intermediate work. A compressed parquet file was found to be among the best options. This information was found

in this article: <https://towardsdatascience.com/the-best-format-to-save-pandas-data-414dca023e0d>
written Ilia Zaitsev.

1.4.1 Jupyter files:

Data cleaning and analysis was broken into steps using a number of Jupyter files:

- `step1_clean_noaa_weather_data.ipynb`
- `step2_clean_ground_stations.ipynb`
- `step3_add_summaries.ipynb`
- `step4_combine_weather_locations.ipynb`
- `step5_plot_us_canada_pedro.ipynb`
- `step6_plot_your_own_ranges.ipynb`

1.4.2 Data Cleaning

The following steps were taken to in collecting, cleaning, and exploring the data:

- Intermediate progress DataFrames for the work below was saved in compressed Parquet files to keep file size to minimum.

step1_clean_noaa_weather_data.ipynb

- Pulled in 37M daily weather records from 42K weather stations across the globe.
 - https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/2023.csv.gz
- Reduced the data to US and Canada weather stations.
- Transformed the data by consolidating (groupby) for Precipitation & min, max and average temperature.
- Resulted in 8687 stations across US and Canada

step2_clean_ground_stations.ipynb

- Pulled in station locations from NOAA website in column separated file:
 - <https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt>
- Cleaned station locations and column names.

step3_add_summaries.ipynb

- Continued with parquet result file from step 1.
- Binned all the data within ~20 temperature ranges for TMAX (count of days in each range)
- Eliminated stations with fewer than 300 days of data and normalized to 365 days.
 - This allowed a reasonably accurate count of days in the analysis.
 - **Assumption:** missing days are evenly distributed across the bins. This is an imperfect assumption.
- Resulted in 7767 stations across US and Canada
- Performed analysis of `#days > X`, `#days < X`, and `X < #days < Y` for a range of temperatures X & Y and stored the results in the associated columns.

step4_combine_weather_locations.ipynb

- Added Station Lat/Lon (Merge) to binned station information.

step5_plot_us_canada_pedro.ipynb

- Plotted all the result on multiple dynamic maps using plotly (plotly.com)
- All the resulting plots are located on the github pages: <https://jflachman.github.io/>

step6_plot_your_own_ranges.ipynb

1.4.3 Analysis

- Visually analyzed regions for temperature / climate statistics:
 - Because weather is a geographic centric statistic, it was best to analyze and present the data visually on dynamic plotly maps.
- Provided a screen capture of the plotly files and a summary of the analysis results in section 2: Results & Conclusions

1.4.4 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 rainy days or potential cloudy days. This is an imperfect assumption.

2 Results & Conclusions

To refine the questions, two AI sources were consulted: ChatGPT (chatgpt.com) and Claude (claude.ai). ChatGPT provided different answers when asked a second time to confirm the data for this report.

The text of the ChatGPT and Claude interactions and the chosen parameter summaries are included in Appendix A (section 6). The questions and summaries are provided in the following sections.

Data analysis typically focuses on definitive, computational results with bar charts, graphs, and tables. However, weather is a geographic centric statistic. Therefore, it was best to analyze and present the data visually on dynamic plotly maps. A future effort could transform the data to zip codes, counties, towns, and cities to provide alternate meaningful presentations. Regardless, the visual map presentation is much more interesting and provides greater value than alternate options.

It is important to note that a methodology was used to define the desired cold, hot and temperate climates. However, these parameters remain subjective and may not fit everyone's definition of the perfect temperatures for their desired climate. For this reason, an additional Jupyter file (step6_plot_your_own_ranges.ipynb) was created to allow others to create maps for their definition of ideal climate.

2.1 Precipitation Regions (ARID vs Wet, Dry vs Humid)

What are the different climate categories from arid to wet?

One widely used classification system is the Köppen climate classification system, which divides climates into several main categories and subcategories (ref ChatGPT). For additional information, see 0 Appendix A.

For Arid and Semi-Arid climates, 100 days of precipitation or less was used. This is good for hot and temperate climates. For Humid Continental climates 100-200 days of precipitation or less was used. This may be better for colder climates to support winter activities (a subjective answer).

From the data the days of precipitation were derived across the US and Canada. See figure 3.

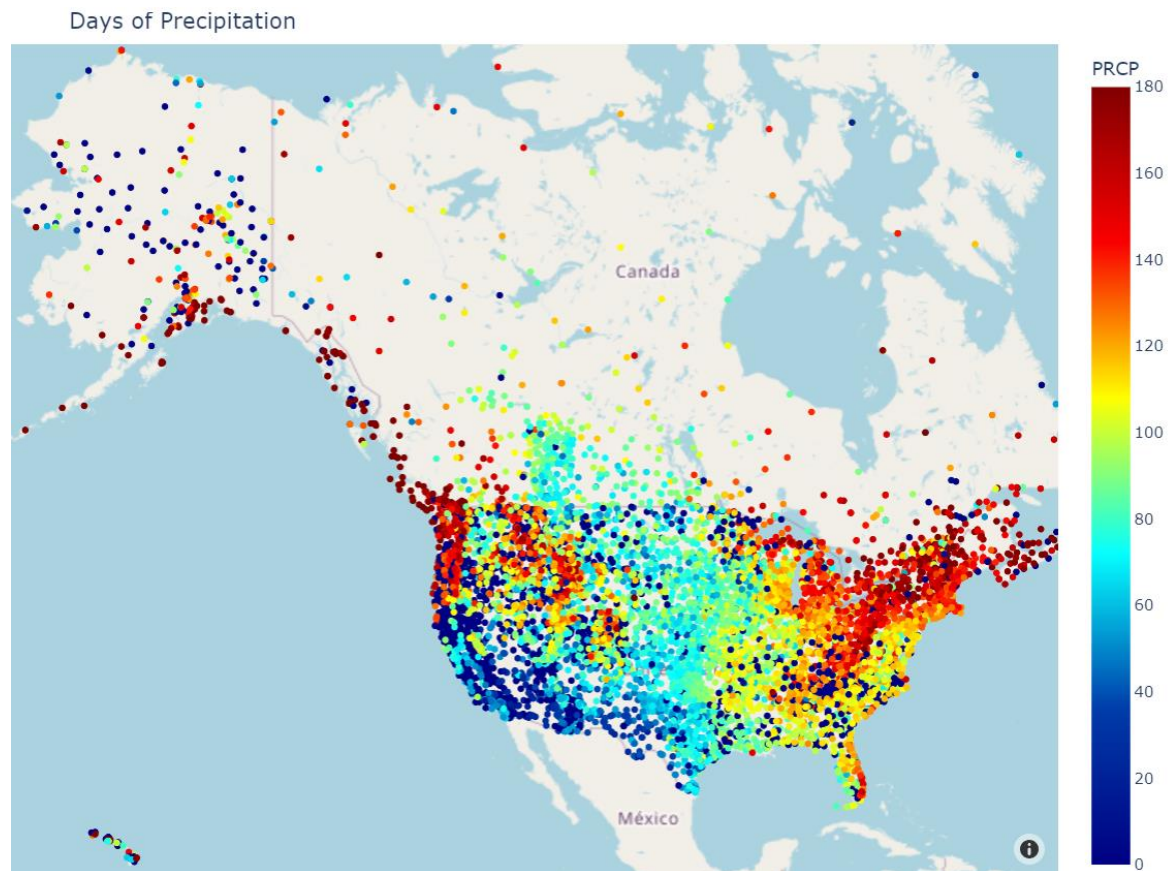


Figure 3: Days of Precipitation

Conclusions:

Arid to Semi-Arid Regions:

- Southwester US

Wetter or Humid Regions:

- Northwestern US and Canada
- US and Canada East Coast toward the Midwest

Median Precipitation

- Midwest- Dakotas south to East Texas)

2.2 Cold Climates

For people who like cold weather, what is the ideal temperature?

ChatGPT and Claude provided several ranges for people who like cold weather. These results were simplified to two ranges which are discussed below. These ranges are **25°F to 55°F** and **32°F to 65°F**.

2.2.1 Cold Regions (25°F to 55°F)

25°F to 55°F was selected as the desired range for people who like colder temperatures based on the initial ChatGPT and Claude responses. For more information see section 0 Appendix A.

From the data regions with a higher number of days of temperatures between 25°F & 55°F were derived across the US and Canada.

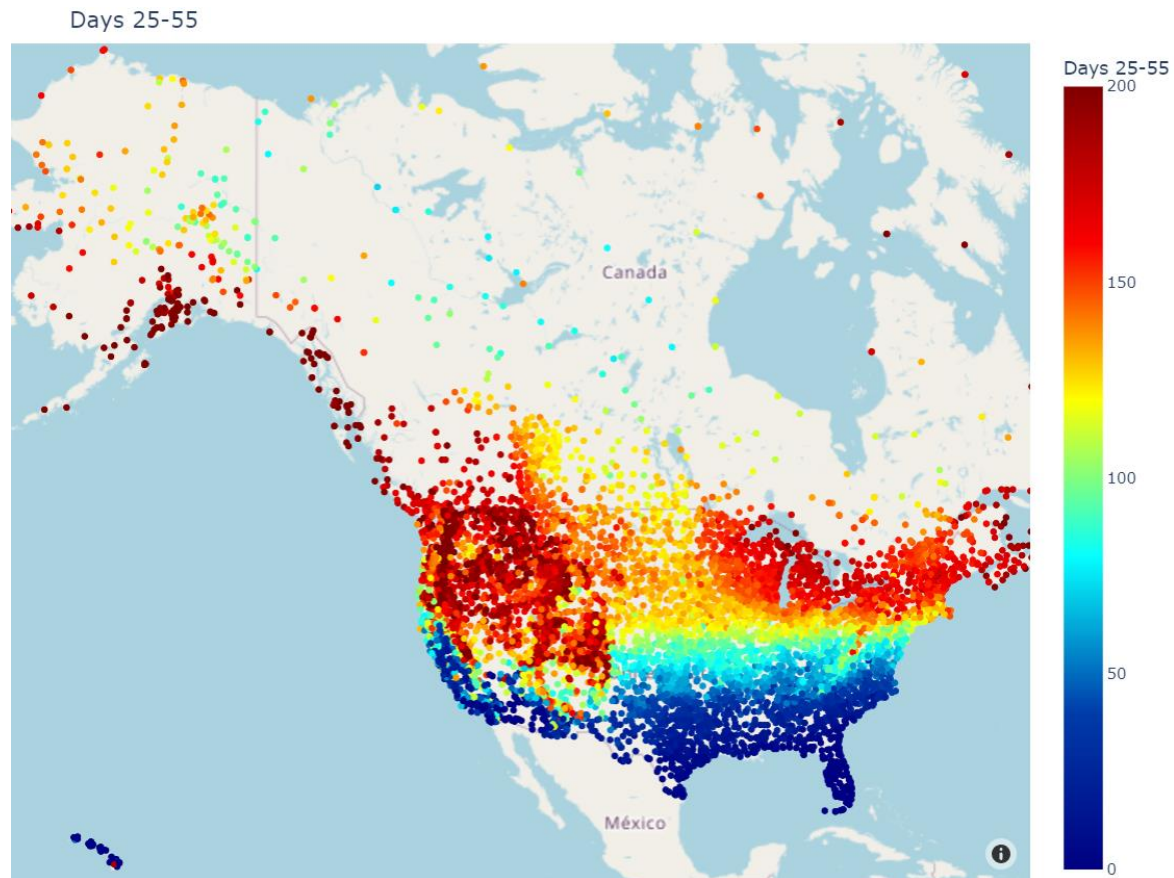


Figure 4: # Days between 25°F to 55°F.

Conclusions:

Based on the Coloring across the map, Comfort Zones for people who like it cold:

- Northern US and all of Canada
- The area includes southern states in the west.

Ideal Regions for Cold climates with moderate humidity and few extreme days

These colder climates can be further narrowed to a smaller “ideal” region by adding additional parameters.

For the Ideal weather, an Arid to humid continental climates were selected with less than 200 days of precipitation. This allows for snow to support enjoyable cold weather activities.

For example, the figure below shows:

- # Days between 25°F & 55°F
- Limit to stations that are not too hot with fewer than 10 Days > 80°F.
- Limit to stations that are not too cold with fewer than 10 Days < 10°F.
- Limit to stations with fewer than 200 days of precipitation.

Days 25-55 with 'Days>80'<10, 'Days<10'<10 & 'Days Precip'<200

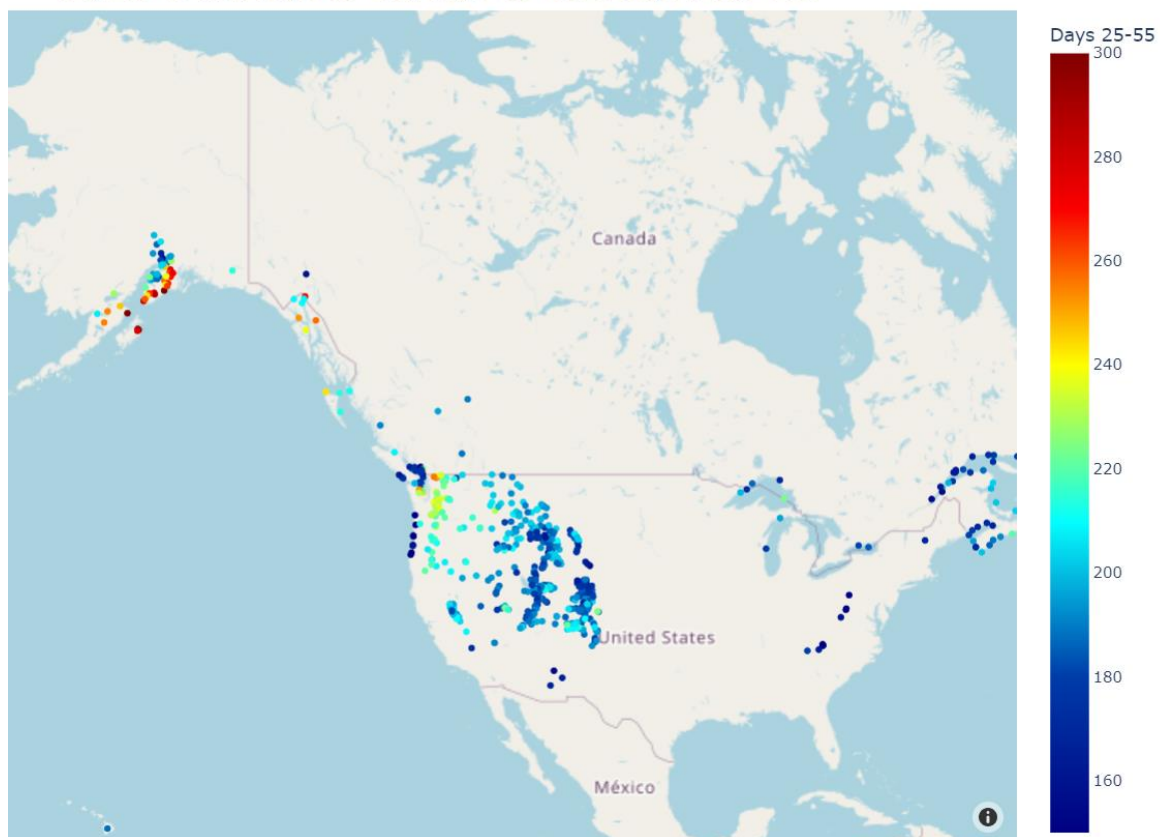


Figure 5: Ideal # Days between 25°F to 55°F.

2.2.2 Alternate Cold Regions (32°F to 65°F)

32°F to 65°F was selected as an alternate definition for the desired range for people who like colder temperatures based on the initial ChatGPT and Claude responses. For more information see section 0 Appendix A.

From the data regions with a higher number of days of temperatures between 32°F & 65°F across the US and Canada were graphically derived.

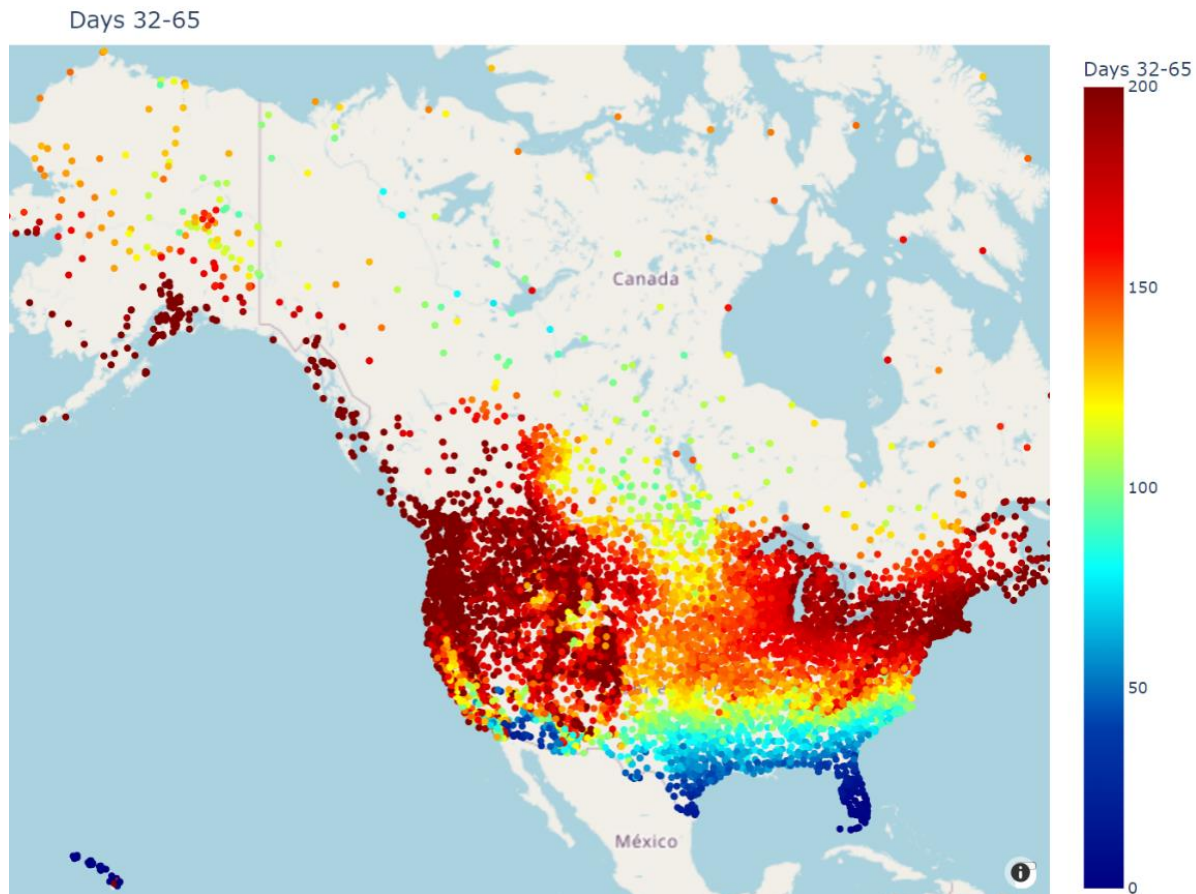


Figure 6: # Days between 32°F to 65°F.

Conclusions:

Based on the Coloring across the map, Comfort Zones for people who like it cold:

- Mid latitudes of US and all of Canada

Ideal Regions for Cold climates with moderate humidity and few extreme days

These colder climates can be further narrowed to a smaller “ideal” region by adding additional parameters.

For the Ideal weather, an Arid to humid continental climates were selected with less than 200 days of precipitation. This allows for snow to support enjoyable cold weather activities.

For example, the figure below shows:

- # Days between 32°F & 65°F
- Limit to stations that are not too hot with fewer than 10 Days > 80°F.
- Limit to stations that are not too cold with fewer than 10 Days < 10°F.
- Limit to stations with fewer than 200 days of precipitation.

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

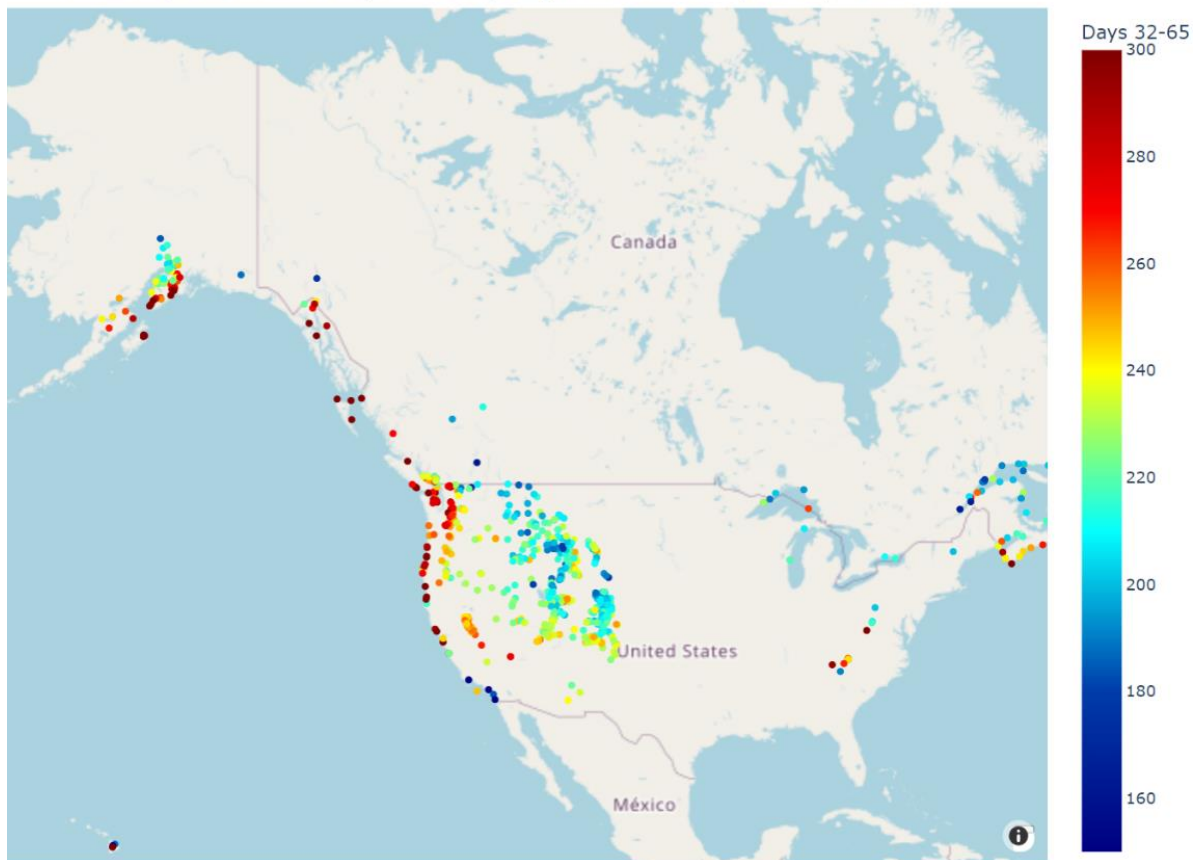


Figure 7: Ideal # Days between 32°F to 65°F.

2.3 Hot Climates (75°F to 95°F)

For people who like hot weather, what is the ideal temperature?

75°F to 95°F was selected as the desired range for people who like hotter temperatures based on the initial ChatGPT and Claude responses. For more details see section 0 Appendix A.

From the data, regions with a higher number of days with temperatures between 75°F & 95°F across the US and Canada were graphically derived.

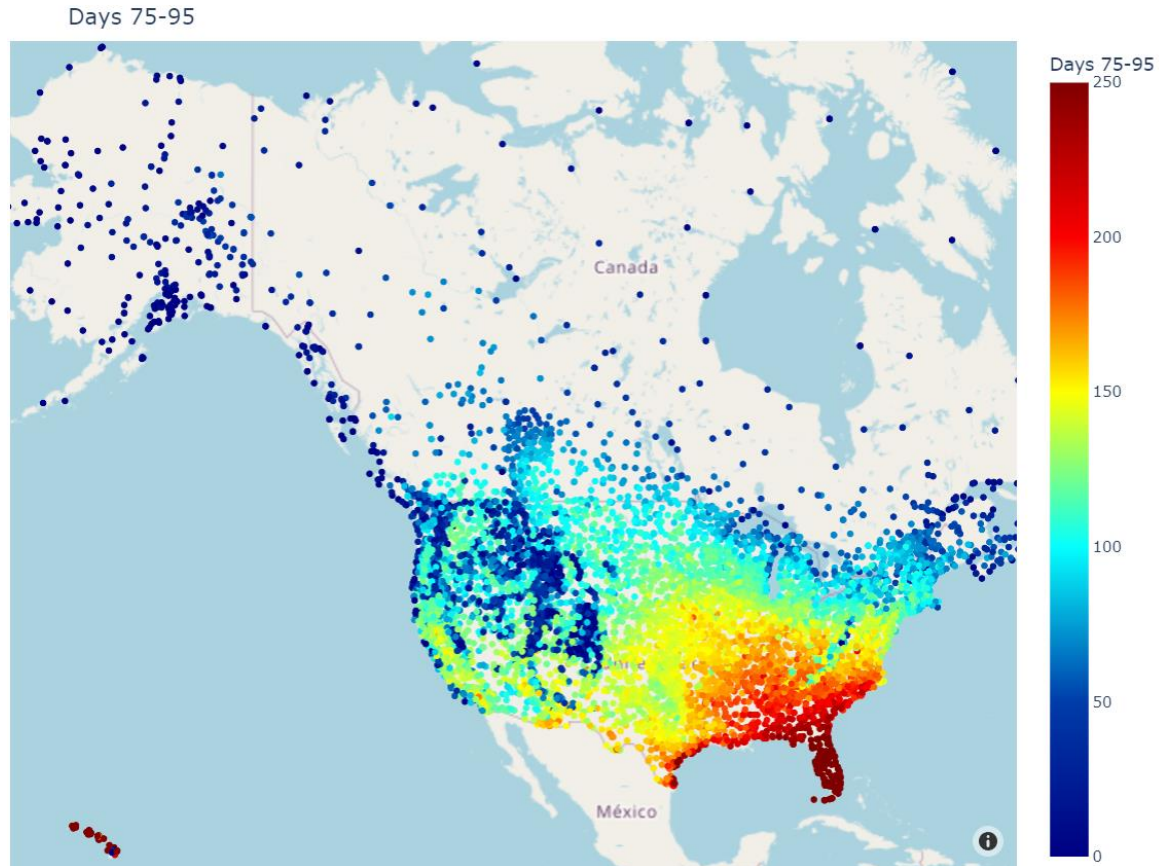


Figure 8: # Days between 75°F to 95°F.

Conclusions:

Based on the Coloring across the map, Comfort Zones for people who like it hot:

- Southeast US (FL, LA, AL, GA, SC, NC, TN)
- Central Southern CA
- South TX & AZ

Ideal Regions for Hot climates with low humidity and few extreme days

These hotter climates can be further narrowed to a smaller “ideal” region by adding additional parameters.

For the Ideal weather, a more Arid to Semi-Arid climates was chosen with less than 100 days of precipitation. This emphasized a dry heat which is much more comfortable at hot temperatures than a humid heat.

For example, the figure below shows:

- # Days between 75°F & 95°F
- Limit to stations that are not too hot with fewer than 10 Days > 95°F.
- Limit to stations that are not too cold with fewer than 10 Days < 10°F.
- Limit to stations with fewer than 100 days of precipitation.

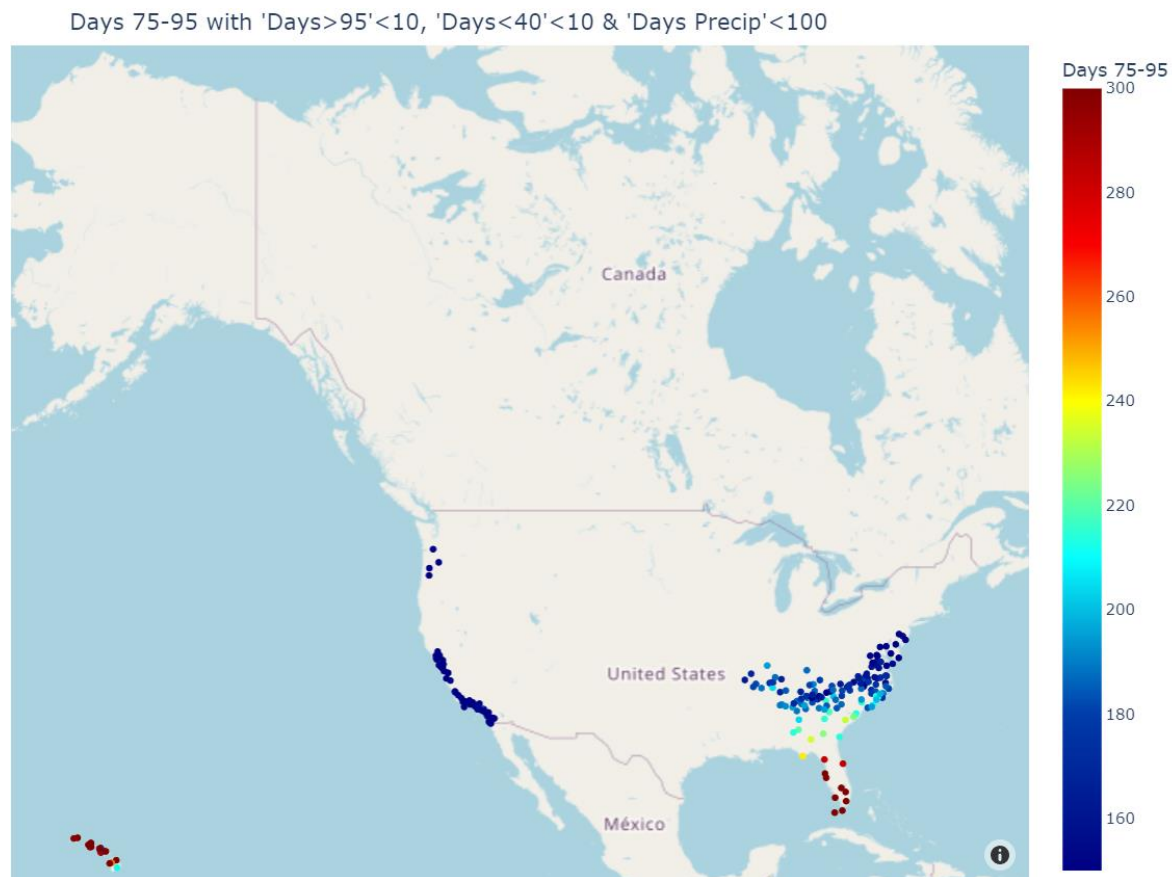


Figure 9: Ideal # Days between 75°F to 95°F.

2.4 Temperate Climates (60°F to 80°F)

For people who like temperate weather, what is the ideal temperature?

60°F to 80°F was selected as the desired range for people who like temperate temperatures based on the initial ChatGPT and Claude responses. For more details see section 0 Appendix A.

From the data regions with a higher number of days with temperatures between 60°F & 80°F across the US and Canada were graphically derived.

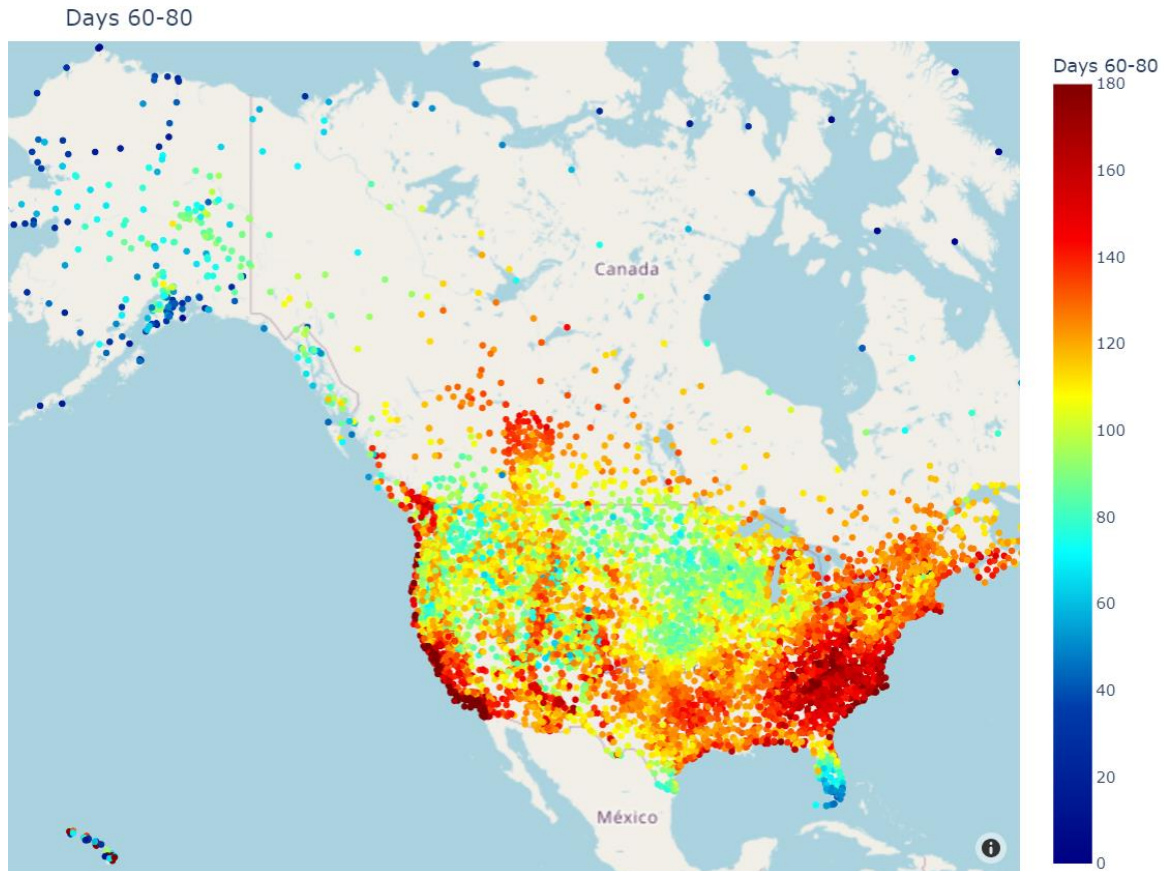


Figure 10: # Days between 60°F to 80°F.

Conclusions:

Based on the Coloring across the map, Comfort Zones for people who like temperate weather include:

- Westcoast within a short range from the coastline (CA, OR WA)
- Canada BC – Vancouver, Edmonton
- Mid-southeast US (VA, WV, NC, SC, TN, GA, and N AL)
- Arizona at elevation (Flagstaff on to SE AZ)

Ideal Regions for Temperate climates with low humidity and few extreme days

These temperate climates can be further narrowed to a smaller “ideal” region by adding additional parameters.

For the Ideal weather, a more Arid to Semi-Arid climate was chosen with less than 100 days of precipitation. This emphasizes a dry climate which is more comfortable at higher temperatures than humid heat. It also emphasizes a climate with a higher percentage of clear days to support outdoor activities.

For example, the figure below shows:

- # Days between 60°F & 80°F
- Limit to stations that are not too hot with fewer than 10 Days > 95°F.
- Limit to stations that are not too cold with fewer than 10 Days < 10°F.
- Limit to stations with fewer than 100 days of precipitation.

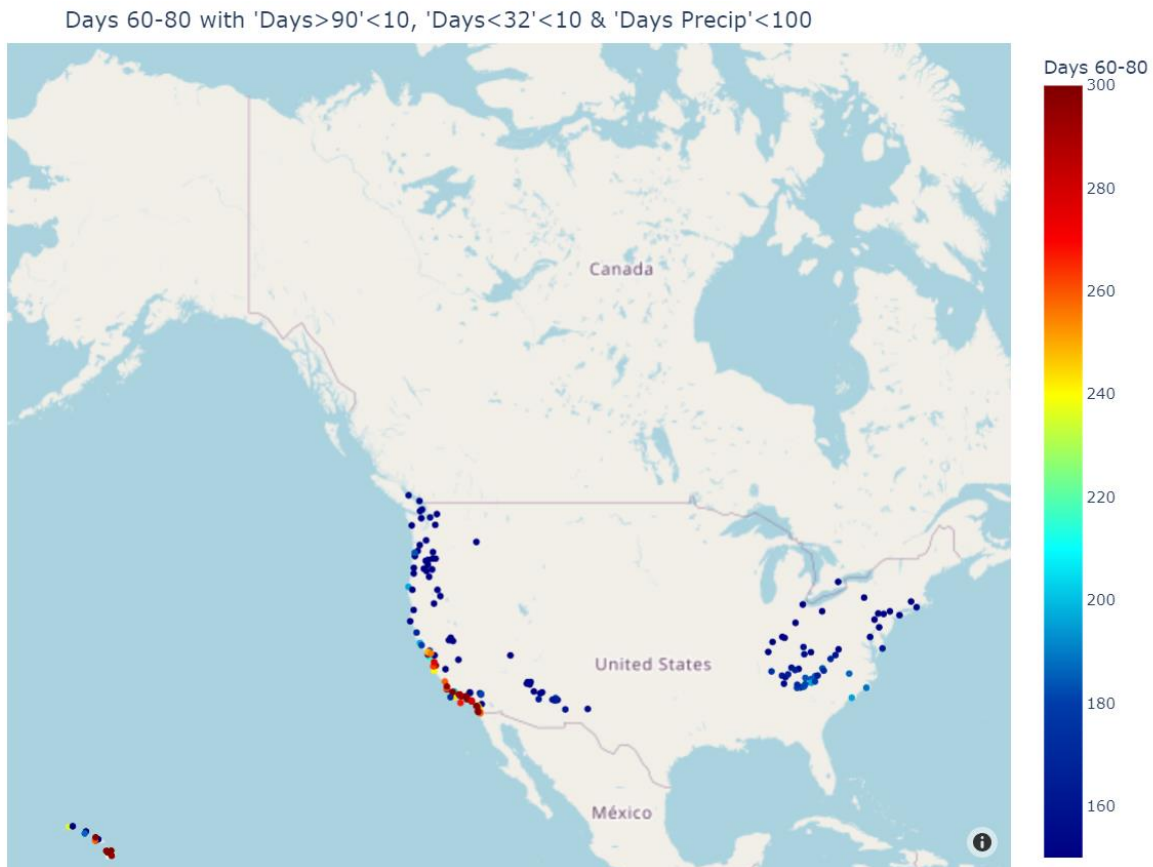


Figure 11: Ideal # Days between 60°F to 80°F.

2.5 Areas of Extreme Temperatures

What temperatures would be considered extreme even for people who typically like hot or cold climates?

Temperatures less than **20°F** were selected as the range where people who like colder temperatures will find the cold to be extreme. Temperatures greater than **95°F** were selected as the point where people who like hotter temperatures will find the heat to be extreme.

The following figures graphically highlights areas of extreme temperatures either below 20°F or above 95°F:

Extreme Days < 20

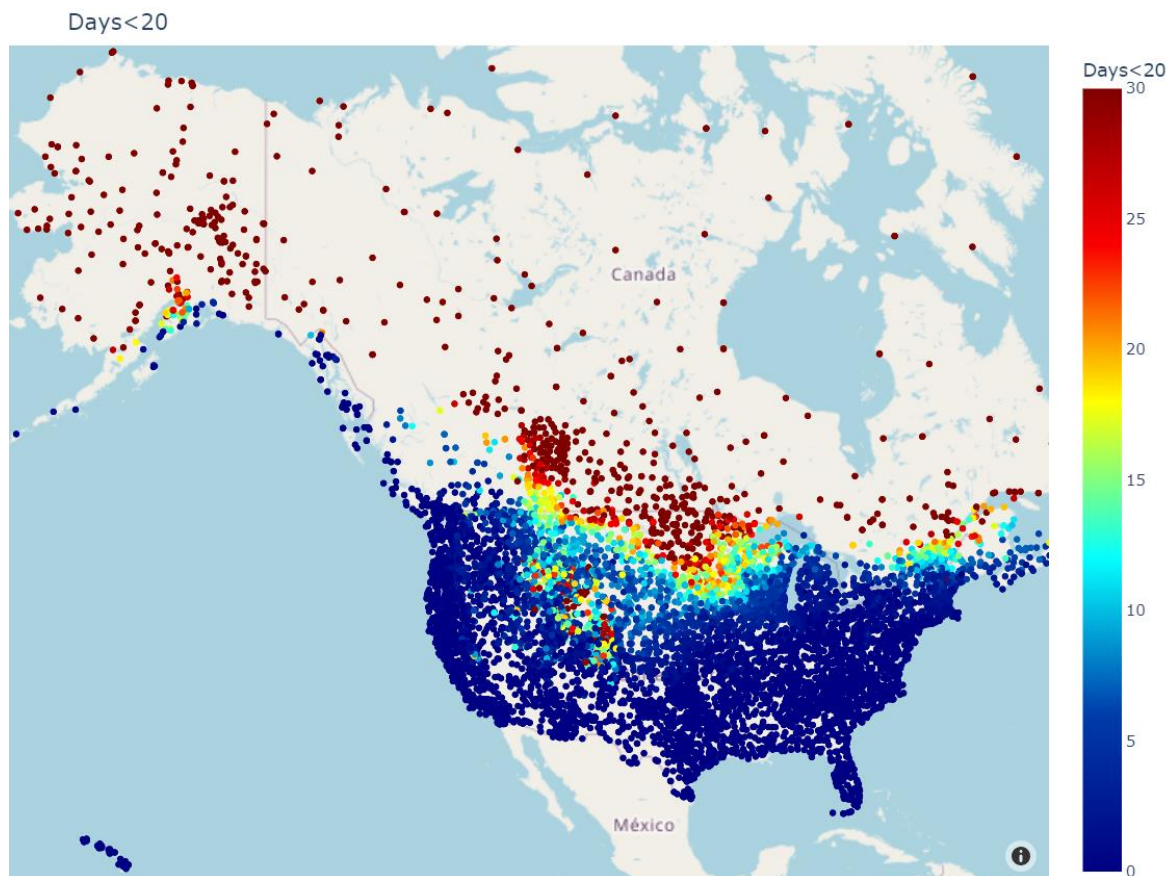


Figure 12: 3 Days below 20°F.

Conclusions:

Even people who prefer cold climates will find the weather to be extreme in these areas:

Really too cold: Days < 20°F

- Most regions above Lat: 51 (Canada, Alaska)
 - Coastal areas in Alaska and Canada are less extreme due to the stability of the water temperatures.
- Also dipping into the northern central US (ND, SD)
- And the Central west US at altitude (WY, CO)

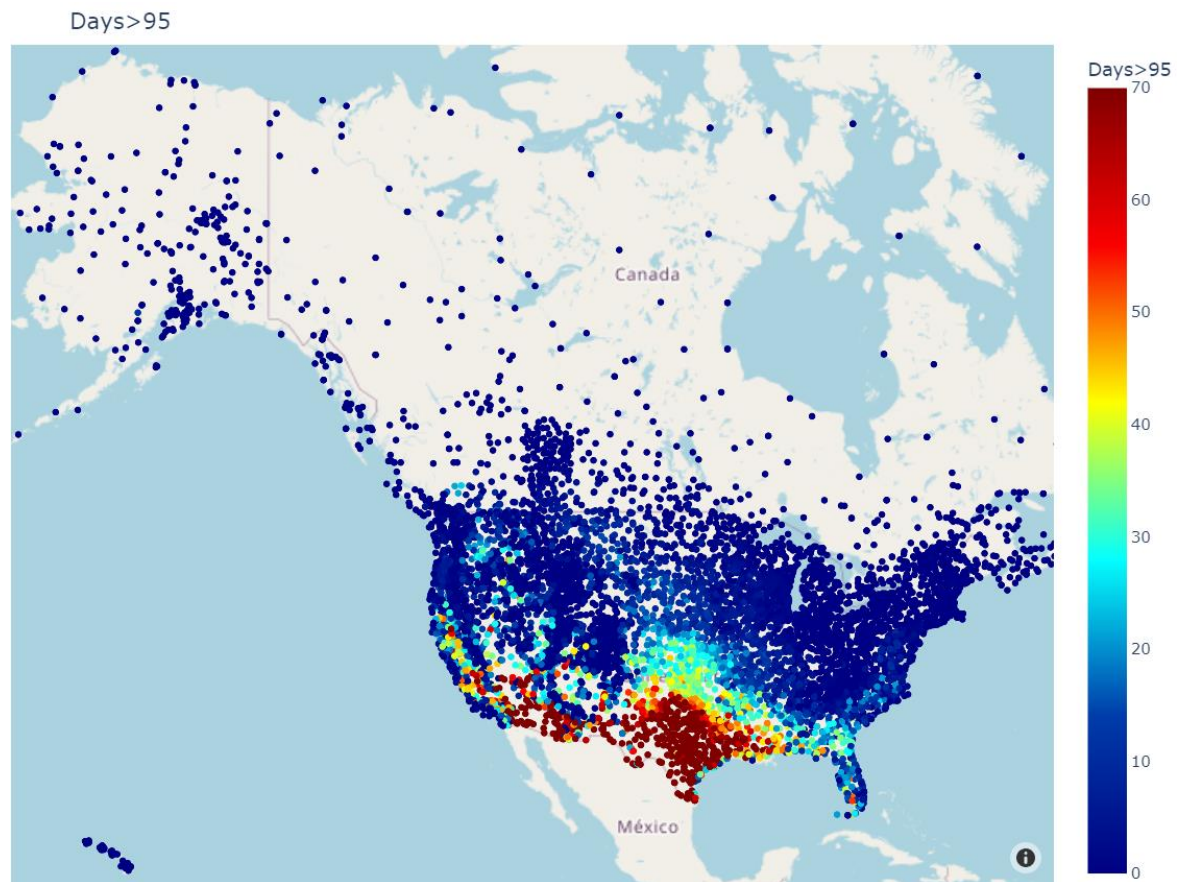
Extreme Days > 95

Figure 13: # Days above 95°F.

Conclusions:

Even people who prefer hot climates will find the weather to be extreme in these areas:

Really too hot: Days > 90°F

- Most regions below Lat: ~36 for Texas to the west. Also ranges up into Kansas.
- See Brown & Cyan regions.

3 Summary

Parameters were defined to support the following questions: two AI sources were consulted: ChatGPT (chatgpt.com) and Claude (claude.ai). ChatGPT provided different answers when asked a second time to confirm the data for this report.

The use of these sources, and the chosen application of the recommendations is subjective. Others may rerun this analysis, altering these parameters to refine the result to better meet their interpretation.

The text of the ChatGPT and Claude interactions and the chosen parameter summaries are included in Appendix A (section 6).

The project questions and the summary of the parameter selections are listed below:

3.1 Analysis Summary

3.1.1 Question 1: What are the different climate categories from arid to wet?

Summary:

For Arid and Semi-Arid climates are characterized by 100 days of precipitation or less. This is good for hot and temperate climates. For Humid Continental climates are characterized by 100-200 days of precipitation or less. This may (a subjective answer) be better for colder climates to support winter activities.

Arid to Semi-Arid Regions:

- Southwester US

Wetter or Humid Regions:

- Northwestern US and Canada
- US and Canada East Coast toward the Midwest

Median Precipitation Regions

- Midwest- Dakotas south to East Texas)

3.1.2 Question 2: For people who like cold weather, what is the ideal temperature?

Summary:

25°F to 55°F was selected as the desired temperature for people who like colder temperatures based on the initial ChatGPT and Claude responses.

Comfort Zones for people who like it cold **25°F to 55°F**:

- Northern US and all of Canada
- The area includes southern states in the west.

32°F to 65°F was selected as an alternate definition for the desired range for people who like colder temperatures based on the initial ChatGPT and Claude responses.

Comfort Zones for people who like it cold **32°F to 65°F**:

- Mid latitudes of US and all of Canada

3.1.3 Question 3: For people who like hot weather, what is the ideal temperature?

Summary:

75°F to 95°F was selected as the desired temperature for people who like hotter temperatures based on the initial ChatGPT and Claude responses.

Comfort Zones for people who like it hot:

- Southeast US (FL, LA, AL, GA, SC, NC, TN)
- Central Southern CA
- South TX & AZ

3.1.4 Question 4: For people who like temperate weather, what is the ideal temperature?

Summary:

60°F to 80°F was selected as the desired temperature for people who like hotter temperatures based on the initial ChatGPT and Claude responses.

Comfort Zones for people who like temperate weather include:

- Westcoast within a short range from the coastline (CA, OR WA)
- Canada BC – Vancouver, Edmonton
- Mid-southeast US (VA, WV, NC, SC, TN, GA, and N AL)
- Arizona at elevation (Flagstaff on to SE AZ)

3.1.5 Question 5: What temperatures would be considered extreme that people who typically like hot or cold climates would not like?

Summary:

Temperatures less than **20°F** were selected as the range where people who like colder temperatures will find the cold to be extreme. Temperatures greater than **95°F** were selected as the point where people who like hotter temperatures will find the heat to be extreme.

Really too cold: Days < 20°F

- Most regions above Lat: 51 (Canada, Alaska)
 - Coastal areas in Alaska and Canada are less extreme due to the stability of the water temperatures.
- Also dipping into the northern central US (ND, SD)
- And the Central west US at altitude (WY, CO)

Really too hot: Days > 90°F

- Most regions below Lat: ~36 for Texas to the west. Also ranges up into Kansas.
- See Brown & Cyan regions.

3.2 Overall Conclusions

There is no definitive answer as to the best region where any individual may choose to live. People have a variety of preferred climates. However, given knowledge that the west coast, Hawaii and the eastern US have higher home prices it would lead to the understanding the temperate climates are preferred by most people.

4 Problems Encountered

The original project concept ideas were overly aggressive. A great deal of time was spent searching for sources to convert weather station locations into zip code, counties, or provinces. This would have allowed correlation of weather statistics with other statistics such as median home prices, unemployment, crime, and other environmental attributes such as earthquakes, tornadoes, weather normal, etc. A good deal of time (way too much time) was spent searching for additional sources that could be correlated with the weather data.

The discussion below will address problems with the data used as well as problems extending the analysis with additional data.

Additionally, data analysis typically focuses on definitive, computational results with bar charts, graphs, and tables. However, weather is a geographic centric statistic. Therefore, it was determined best to analyze and present the data visually on dynamic plotly maps. A future effort could transform the data to zip codes, counties, towns, and cities to provide alternate meaningful presentations. Regardless, the visual map presentation is much more interesting and provides greater value than alternate options.

In this section impediments to integrating this data are discussed.

4.1 General Problems

The project team was guided to use Git Large File Storage (LFS). Several types of binary files were added to the lfs records in the `.gitattributes` file. However, after pushing several source and results files to the project repo, the free limit for lfs files were exceeded on github. At that point, github prevented adding or updating any files to the project repo. To resolve this issue, all sources and result files were removed from the repo. The Jupyter files were updated to pull source files directly from NOAA websites. The result_files are stored locally. Cloning the repo and creating a local result_files directory will allow anyone to recreate the work.

The data sources investigated are recorded in the project github here:

<https://github.com/jflachman/DU-project-1/blob/main/resources.md>

4.2 Problems with the data used:

Below is a bulleted summary of issues encountered with the NOAA data used and the assumptions made to resolve these issues.

- Some bad data was in the dataset (extreme temperature outliers)
 - Temperature bins between -50 and 125 were used, eliminating other outliers.
- Some weather stations provided limited TMAX data.
 - Assumption:
 - 300 days of data should be sufficient for quality analysis.
 - Weather stations with at least 300 days of TMAX data were chosen.
- The data was normalized to 365 days (this is the reason that TMAX days is a float)

The resolution to the data issue is not perfect. However, examination of the data visually in various regions shows consistency of result with some outliers remaining.

4.3 Problems adding sources to the analysis.

Weather Normals: Weather normal data with 92 attributes across 10 years is available. This source was discovered late in the project. It also requires downloading a > 10k files, which is not realistic for this project.

Weather by county or zip code: It is advantageous to group weather station data by zip code, county, or province to support correlation with other data sources. APIs to convert from lat/lon known for the weather station to county or zip code were costly or provided extremely limited bandwidth per their license. It is possible to install and configure a local Nominatum server with zip code and county geojson files. This would allow for unlimited local conversion from lat lon to regions of interest.

<https://nominatim.org/release-docs/latest/admin/Installation/>

Median Home Prices: Median home prices were found on the US Census Website using dataset B25077. This data provides median home prices and a margin of error for that estimate by zip-code.

See:

[https://data.census.gov/table/ACSDT5Y2022.B25077?t=Financial%20Characteristics:Housing%20Value%20and%20Purchase%20Price&g=010XX00US\\$8600000&y=2022](https://data.census.gov/table/ACSDT5Y2022.B25077?t=Financial%20Characteristics:Housing%20Value%20and%20Purchase%20Price&g=010XX00US$8600000&y=2022)

Integrating Additional Data as a Choropleth layer with weather data: A solution was found to provide a graphical underlay of other data with the weather data using plotly choropleth. This simple implementation was not used due to issues collecting other data sources and implementing zip codes.

See: <https://plotly.com/python/choropleth-maps/>

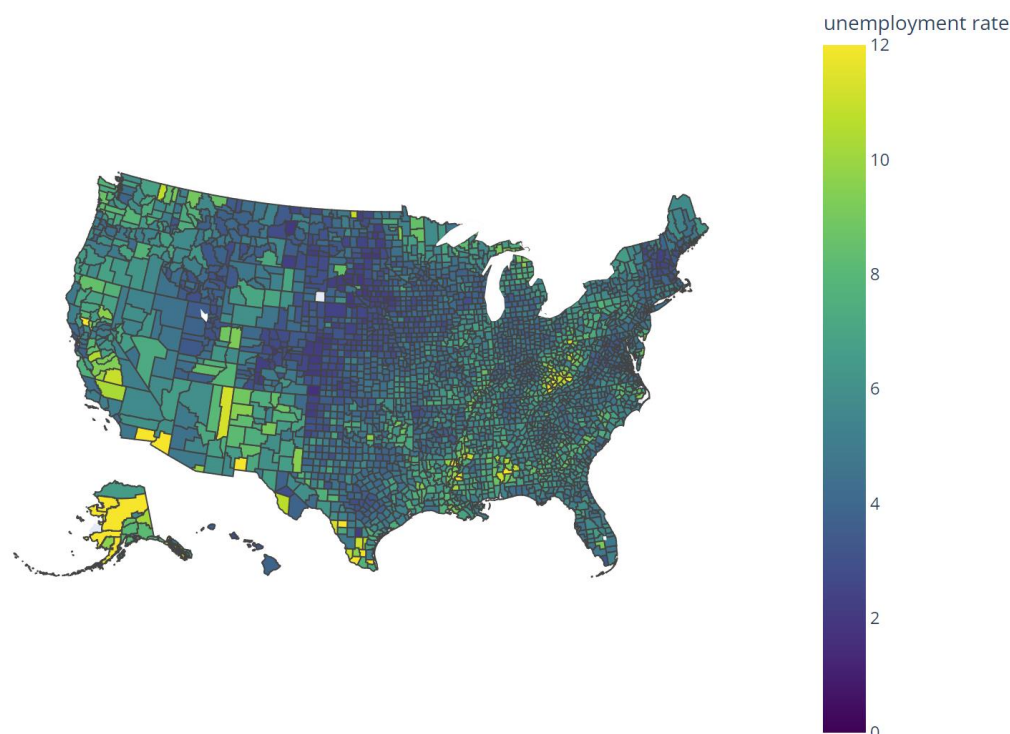


Figure 14: Choropleth example (source <https://plotly.com/python/choropleth-maps/>)

5 Future Considerations

Some of these future considerations are discussed in section 4 Problems Encountered. Many individuals spend their careers collecting, analyzing, and presenting weather and other geographical data. Regardless of the breadth of potential analysis, a follow-on project of 2-3 weeks could extend this analysis in a meaningful way with some of the following work:

Add median housing data by zip-code as a Choropleth underlay to the existing data.

Install and setup a Nominatum DB: This would allow the conversion of lat / lon data to regions such as counties, zip codes or provinces.

Incorporate weather normal: Future statistical analysis should include NOAA Climate Normals for cross-year statistical analysis. Research a method to download and collate weather normal data for US and Canada. This research has not been completed to define the probability of success for this option.

Find other sources for environmental impacts by geography: tornados, earthquakes, humidity, pests, etc.

Create a simple python application: Allow users to enter parameters for Ideal climates and return a dynamic plotly map for them.

6 References

6.1 NOAA Data used.

See: <https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/readme.txt>

Menne, M.J., I. Durre, B. Korzeniewski, S. McNeill, K. Thomas, X. Yin, S. Anthony, R. Ray, R.S. Vose, B.E. Gleason, and T.G. Houston, 2012: Global Historical Climatology Network - Daily (GHCN-Daily), Version 3. [indicate subset used following decimal, e.g. Version 3.12]. NOAA National Climatic Data Center. <http://doi.org/10.7289/V5D21VHZ> [access date].

Appendix A

To refine the questions, two AI sources: ChatGPT (chatgpt.com) and Claude (claude.ai) were consulted.

The detailed result of these queries is included below:

6.2 Question 1: Arid vs Wet Climates

Question:

What are the different climate categories from arid to wet?

The Summary:

- For Arid and Semi-Arid climates, 100 days of precipitation or less were used.
- For Humid Continental climates, 100-200 days of precipitation or less were used.

ChatGPT:

Climate classification systems categorize climates into different types based on factors such as temperature, precipitation, and vegetation. One widely used classification system is the Köppen climate classification system, which divides climates into several main categories and subcategories. Here is an overview of the different climate categories from arid to wet in the Köppen classification system:

1. Arid Climates (Group B):

- Deserts (BW): Very low precipitation, often hot temperatures. Examples include the Sahara Desert and the Arabian Desert.
- Steppe or Semi-Arid (BS): Low precipitation, typically characterized by grasslands with sparse vegetation. Examples include parts of central Asia and the Great Plains of North America.

2. Temperate Climates (Group C):

- Temperate (C): Moderate precipitation, with distinct seasons including warm summers and cool winters. Subcategories include:
 - Mediterranean (Csa, Csb): Warm, dry summers and mild, wet winters. Found around the Mediterranean Sea and in parts of California, Chile, and South Africa.
 - Humid Subtropical (Cfa, Cwa): Hot, humid summers and mild to cool winters. Examples include the southeastern United States, eastern China, and Japan.

3. Continental Climates (Group D):

- Continental (D): Moderate precipitation, with more extreme seasonal temperature variations. Subcategories include:
 - Warm Summer (Dfa, Dwa): Warm to hot summers and cold winters. Found in parts of central and eastern Europe, central Asia, and the central United States.
 - Cold Winter (Dfb, Dwb): Cool to cold summers and very cold winters. Examples include parts of Canada, Russia, and northern China.

4. Polar Climates (Group E):

- Tundra (ET): Cold climates with very little precipitation and short, cool summers. Vegetation is typically low-growing plants like mosses and lichens. Found in northern parts of North America, Europe, and Asia.
- Ice Cap (EF): Extremely cold and dry climates, with temperatures below freezing year-round. Ice caps are found in Antarctica and Greenland.

5. Tropical Climates (Group A):

- Tropical Rainforest (Af): High precipitation and consistently warm temperatures year-round. Dense vegetation, including tall trees and diverse plant life. Examples include the Amazon Rainforest and the Congo Basin.
- Tropical Monsoon (Am): High precipitation, with distinct wet and dry seasons. Found in regions such as parts of Southeast Asia and northern Australia.

These are just a few examples, and there are variations and subcategories within each climate type. The Köppen system is widely used, but there are other classification systems as well, each with its own set of categories and criteria.

Claude:

Here are the main climate categories ranging from arid to wet, based on precipitation amounts and patterns:

1. Arid Climates:

- Hyper-Arid (desert) - Less than 5 precipitation days per year
- Arid - 5 to 30 precipitation days per year

2. Semi-Arid/Moderately Dry Climates:

- 30 to 100 precipitation days per year

3. Dry Subtropical Climates:

- 30 to 90 precipitation days per year, with distinct wet and dry seasons

4. Mediterranean Climates:

- 30 to 90 precipitation days per year, with hot, dry summers and mild, wet winters

5. Humid Subtropical Climates:

- 90 to 150 precipitation days per year, with hot, humid summers

6. Humid Continental Climates:

- 100 to 200 precipitation days per year, with four distinct seasons

7. Oceanic/Marine West Coast Climates:

- 150 to 250 precipitation days per year, with cool summers

8. Wet Tropical Climates:

- Over 150 precipitation days per year, hot and humid year-round

9. Monsoon Climates:

- Over 200 precipitation days per year, extreme wet and dry seasons

10. Tropical Rainforest Climates:

- Over 200 precipitation days per year, hot and humid with frequent rainfall

So, in summary, the categories go from hyper-arid (under 5 rain days), to arid (5-30), semi-arid (30-100), then increasing precipitation for dry subtropical, Mediterranean, humid subtropical/continental, marine west coast, wet tropical, monsoon, and finally the wettest tropical rainforest climates.

6.3 Question 2: Cold Climates

Question:

For people who like cold weather, what is the ideal temperature?

The Summary:

25°F to 55°F was selected as the desired temperature for people who like colder temperatures based on the initial ChatGPT and Claude responses.

Alternate Cold climate range may be **30°F to 65°F**. Both interpretations will be evaluated.

For the Ideal weather, a more Arid to humid continental climates with less than 200 days of precipitation was chosen. This allows for snow to support enjoyable cold weather activities.

ChatGPT:

ChatGPT provided different answers when asked a second time to confirm the data for this report.

Answer 1:

For those who enjoy cold weather, the ideal temperature typically ranges **23 to 50** degrees Fahrenheit. However, preferences can vary widely among individuals, with some preferring even colder temperatures, especially if they enjoy winter sports like skiing or snowboarding. Additionally, factors such as humidity, wind chill, and personal tolerance to cold also influence what someone considers their ideal cold weather temperature.

Answer 2:

The ideal temperature for people who enjoy cold weather can vary significantly based on personal preferences, activities, and physiological factors. However, temperatures in the range of 10°C to 20°C (**50°F to 68°F**) are often considered comfortable for those who prefer cooler climates. Here are a few reasons why this temperature range might be ideal for cold weather enthusiasts:

1. **Comfortable for Outdoor Activities:** Temperatures between 10°C to 20°C are often comfortable for outdoor activities such as hiking, skiing, or simply taking a

leisurely walk. It is cool enough to be refreshing but not so cold as to be uncomfortable.

2. **Enjoyable for Layering Clothing:** Cooler temperatures allow for comfortable layering of clothing, which many cold weather enthusiasts enjoy. This allows for flexibility in adapting to changing weather conditions and outdoor activities.
3. **Maintains Energy Levels:** For some people, cooler temperatures can help maintain energy levels and alertness, especially during physical activities. The brisk air can feel invigorating and refreshing.
4. **Better Sleep Environment:** Many people find it easier to sleep in cooler environments. Sleeping in slightly cooler temperatures can promote better sleep quality and may be more conducive to restful sleep.
5. **Reduced Allergens:** Cooler temperatures can sometimes be associated with lower humidity levels, which may reduce the presence of allergens like mold and dust mites. This can be beneficial for individuals with allergies or respiratory issues.

Of course, individual preferences can vary widely, and what one person considers ideal, another may find too cold or too warm. Some individuals may even prefer much colder temperatures and thrive in climates with temperatures below freezing. The ideal temperature for cold weather enthusiasts depends on personal comfort levels, lifestyle, and the activities they enjoy.

Claude:

For people who enjoy colder weather, the ideal temperature range tends to be:

- **50°F (10°C) to 65°F (18°C)** - This range is considered cool to mild for most people. At the lower end around 50°F (10°C), it starts feeling quite crisp and refreshing for cold weather fans.
- **40°F (4°C) to 55°F (13°C)** - Many who love cold temperatures revel in this brisk, chilly range. Around 45°F (7°C) is an ideal cool temperature for some cold devotees.
- **30°F (-1°C) to 45°F (7°C)** - The colder part of this ranges from freezing (32°F/0°C) down to around 30°F (-1°C) is absolutely perfect cold weather for extreme cold lovers.

Of course, preferences can vary based on the individual, clothing worn, wind chill, and humidity levels. But in general, most people who appreciate cold conditions seem to idealize temperatures from about 30°F to 55°F (-1°C to 13°C).

Below freezing from 25°F (-4°C) and lower becomes too frigid for most cold weather fans to truly enjoy being outdoors for extended periods. And above 65°F (18°C) starts feeling too warm for diehard cold aficionados.

The crisp 40s Fahrenheit (around 5°C to 9°C) seem to be the sweet spot for maximum cold weather enjoyment for many people who love brisk, wintry conditions.

6.4 Question 3: Hot Climates

Question:

For people who like hot weather, what is the ideal temperature?

The Summary:

75°F to 95°F were selected as the desired temperature for people who like hotter temperatures based on the initial ChatGPT and Claude responses.

For the Ideal weather, a more Arid to Semi-arid climate with less than 100 days of precipitation was chosen.

ChatGPT:

ChatGPT provided different answers when asked a second time to confirm the data for this report.

Answer 1:

The ideal temperature for people who enjoy hot weather can vary based on personal preference, but it ranges between **77 to 95 degrees Fahrenheit**. Some might even prefer temperatures higher than that, especially if they live in regions with consistently warm climates. However, it is essential to note that **humidity levels also play a significant role in how comfortable or enjoyable hot weather feels** to individuals.

Answer 2:

For individuals who enjoy hot weather, the ideal temperature typically falls within a warmer range compared to those who prefer cooler climates. However, just like with cold weather enthusiasts, the ideal temperature for hot weather lovers can vary based on personal preferences, activities, and other factors. Here are some temperature ranges that might be considered ideal for people who enjoy hot weather:

1. **Warm and Sunny Days (25°C to 35°C / 77°F to 95°F):** Many individuals who enjoy hot weather find temperatures between 25°C and 35°C (77°F to 95°F) to be ideal. These temperatures are warm enough to feel comfortably hot without being excessively sweltering.
2. **Beach Weather (28°C to 32°C / 82°F to 90°F):** For those who love spending time at the beach or engaging in water activities, temperatures ranging from 28°C to 32°C (82°F to 90°F) are often considered ideal. These temperatures provide the perfect balance for sunbathing, swimming, and enjoying outdoor water sports.
3. **Dry Heat (30°C to 40°C / 86°F to 104°F):** In regions with dry climates, such as desert areas, temperatures ranging from 30°C to 40°C (86°F to 104°F) can be preferred by individuals who enjoy hot weather. The lack of humidity makes these temperatures feel more comfortable compared to humid conditions at similar temperatures.
4. **Tropical Climates (25°C to 30°C / 77°F to 86°F):** In tropical regions, temperatures ranging from 25°C to 30°C (77°F to 86°F) are common and are often considered ideal by individuals who enjoy hot and humid weather. These

temperatures are typically accompanied by lush vegetation and frequent rainfall.

5. **Hot Summer Days (30°C to 35°C / 86°F to 95°F):** During the summer months, temperatures between 30°C and 35°C (86°F to 95°F) are often appreciated by hot weather enthusiasts. These temperatures provide ample opportunity for outdoor activities such as hiking, picnicking, and barbecuing.

It is important to note that individual preferences for hot weather can vary greatly, and what one person finds enjoyable, another may find uncomfortably hot. Factors such as humidity levels, wind speed, and personal tolerance for heat also play a significant role in determining the ideal temperature for hot weather lovers.

Claude:

There is no definitive "ideal" temperature that will be perfect for everyone who enjoys hot weather. Personal preferences and factors like humidity can influence what feels comfortable. However, some general guidelines for temperatures that many people who like hot weather would find pleasant:

- **80°F (27°C) to 95°F (35°C)** - This range tends to be considered very warm to hot by most people. The hotter end around 90°F (32°C) or above is when heat starts to become oppressive for some.
- **85°F (29°C) to 90°F (32°C)** - This could be considered an ideal hot temperature range for many heat-lovers, though the humidity level makes a big difference in how comfortable these temps feel.
- **75°F (24°C) to 85°F (29°C)** - Some may find the hotter part of this range ideal, while others who really crave heat may want it a bit warmer still.

Personal acclimatization also plays a role. Those who live in very hot climates year-round tend to prefer even hotter temperatures than visitors from cooler areas. But in general, most people who like it hot seem to ideal hot temperature in the mid-80s to low 90s Fahrenheit (29°C to 34°C) with low humidity. Anything much above 100°F (38°C) can feel uncomfortably hot for extended periods even for heat-lovers.

6.5 Question 4: Temperate Climates

Question:

For people who like temperate weather, what is the ideal temperature?

The Summary:

60°F to 80°F were selected as the desired temperature for people who like hotter temperatures based on the initial ChatGPT and Claude responses.

For the Ideal weather, a more Arid to Semi-arid climate with less than 100 days of precipitation was chosen.

ChatGPT:

ChatGPT provided different answers when asked a second time to confirm the data for this report.

Answer 1:

People who prefer moderate temperatures typically find comfort in a range of **64 to 75** degrees Fahrenheit. This range provides a pleasant balance between warmth and coolness, allowing for comfortable indoor and outdoor activities without feeling too hot or too cold. Moderate temperatures like these are often considered ideal for various daily activities, such as outdoor exercise, leisurely walks, or simply enjoying time outdoors without discomfort.

Answer 2:

For individuals who enjoy temperate weather, the ideal temperature typically falls within a moderate range that is neither too hot nor too cold. Temperate weather enthusiasts often appreciate a comfortable balance that allows for outdoor activities and enjoyment without extreme temperatures. Here are some temperature ranges that might be considered ideal for people who like temperate weather:

- **Mild and Pleasant Days (20°C to 25°C / 68°F to 77°F):** Many individuals who prefer temperate weather find temperatures between 20°C and 25°C (68°F to 77°F) to be ideal. These temperatures are neither too hot nor too cold, providing a pleasant and comfortable environment for outdoor activities, such as hiking, biking, or simply enjoying nature.
- **Spring and Fall Weather (15°C to 20°C / 59°F to 68°F):** Temperatures ranging from 15°C to 20°C (59°F to 68°F) are often appreciated by those who enjoy temperate climates during the spring and fall seasons. These temperatures are mild and comfortable, allowing for enjoyable outdoor activities like picnics, gardening, or leisurely walks.
- **Cool Summer Evenings (18°C to 22°C / 64°F to 72°F):** During the summer months, temperatures between 18°C and 22°C (64°F to 72°F) are often preferred by individuals who enjoy temperate weather. These temperatures provide relief from the heat of the day while still allowing for outdoor gatherings, barbecues, or stargazing.
- **Moderate Humidity (20°C to 25°C / 68°F to 77°F with Low Humidity):** Temperatures between 20°C and 25°C (68°F to 77°F) with low humidity levels are often considered ideal by temperate weather enthusiasts. Moderate humidity levels contribute to a comfortable and enjoyable outdoor environment without feeling too sticky or uncomfortable.
- **Balanced Climate (No Extreme Temperatures):** Overall, individuals who prefer temperate weather appreciate climates that avoid extreme temperatures, whether hot or cold. A balanced climate with moderate temperatures year-round is often considered ideal for those who enjoy temperate weather.

It is important to note that individual preferences for temperate weather can vary, and what one person finds ideal, another may find too warm or too cool. Additionally, factors such as wind speed, sunshine, and personal tolerance for temperature fluctuations also influence what individuals consider to be the ideal temperate weather.

Claude:

For people who prefer temperate, moderate weather conditions, the ideal temperature range tends to be:

- **65°F (18°C) to 75°F (24°C)** - This range is often considered perfect temperate weather by many people. Not too hot, not too cold, just pleasantly mild.
- **60°F (16°C) to 70°F (21°C)** - Some may find the cooler end of this range ideal for temperate conditions, while others lean toward the warmer end.
- **70°F (21°C) to 80°F (27°C)** - As you get up into the 70s Fahrenheit, it starts feeling warm rather than just mild for some temperate weather fans. But low 70s can still be appreciated.

A few key factors that make certain temperatures feel ideally temperate include:

- **Low humidity** - Dry air feels more comfortable than muggy conditions at the same temperature.
- **Ample cloud cover** - Sunny days make temperatures feel warmer than overcast ones.
- **Little wind** - Breezy conditions can make a mild temperature feel colder.

In general, though, temperatures from about 65°F to 75°F (18°C to 24°C) combined with low humidity, partial cloud cover, and light winds tend to be considered a perfect temperate weather zone for most who dislike extremes of hot or cold.

The mid to upper 60s Fahrenheit (around 18°C to 21°C) often hits that idealized sweet spot for temperate weather enjoyment. Not too cool, not too warm - just right in the middle.

6.6 Question 5: Extreme Temperatures

Question:

What temperatures would be considered extreme that people who typically like hot or cold climates would not like?

The Summary:

20°F was selected as the point for people who like colder temperatures and will find the cold to be extreme. **95°F** was selected as the point for people who like hotter temperatures and will find the heat to be extreme.

ChatGPT:

For people who typically enjoy hot or cold climates, there are temperature extremes that can become uncomfortable or even dangerous:

Extreme temperatures that individuals who typically enjoy hot or cold climates might find uncomfortable typically fall outside of their preferred temperature ranges. Here are some examples:

1. **Extreme Heat (Above 40°C / 104°F):** Temperatures above 40°C (104°F) are considered extreme and can be uncomfortable, even for individuals who enjoy hot weather. High temperatures can lead to heat-related illnesses such as heat exhaustion and heatstroke, and prolonged exposure to extreme heat can be dangerous.
2. **Extreme Cold (Below -10°C / 14°F):** Temperatures below -10°C (14°F) are considered extreme cold and can be challenging, even for individuals who typically enjoy cold climates. Extreme cold temperatures can increase the risk of frostbite, hypothermia, and other cold-related health issues.
3. **Heatwaves:** Heatwaves, which are prolonged periods of excessively hot weather, can be particularly uncomfortable and dangerous for individuals who enjoy hot climates. Heatwaves often bring high temperatures, humidity, and poor air quality, increasing the risk of heat-related illnesses and discomfort.
4. **Cold Snaps:** Cold snaps, which are sudden and prolonged periods of extremely cold weather, can be challenging for individuals who enjoy cold climates. Cold snaps often bring frigid temperatures, strong winds, and heavy snowfall, making outdoor activities difficult and increasing the risk of cold-related health issues.
5. **Temperature Extremes in Combination with Other Factors:** Extreme temperatures can become even more uncomfortable and dangerous when combined with other factors such as high humidity, low humidity (resulting in dry air), strong winds, or poor air quality. These conditions can exacerbate the effects of extreme heat or cold and increase the risk of health problems.

It is important to note that individual tolerance for extreme temperatures varies, and what one person considers extreme, another may find manageable. However, extreme temperatures outside of the typical range for hot or cold climates can pose health risks and discomfort for individuals who enjoy those climates.

Claude:

For people who typically enjoy hot or cold climates, there are temperature extremes that can become uncomfortable or even dangerous:

For Hot Weather Lovers:

- Temperatures above 105°F (41°C) are considered extremely hot, even for heat enthusiasts. The risk of heat illness increases significantly.
- In dry desert climates, temperatures of 115°F (46°C) or higher can be life-threatening with prolonged exposure.
- With high humidity levels, temperatures as "low" as 95°F (35°C) can feel oppressively hot and unbearable.

For Cold Weather Lovers:

- Temperatures below 0°F (-18°C) are considered extremely cold, even for cold aficionados. Risk of frostbite and hypothermia increases.
- Wind chills below -20°F (-29°C) make being outdoors very difficult and dangerous without proper protection.
- In polar regions, temperatures of -40°F (-40°C) or lower are considered extraordinarily cold extremes.

So, in general:

- Temperatures over 105-115°F (41-46°C) are extremely hot, even for hot climate fans.
- Temperatures below 0°F (-18°C), especially with wind chill, are extremely cold even for cold lovers.

At these extreme ends of the temperature range, risks increase significantly regardless of hot/cold preferences. Most people find prolonged exposure very uncomfortable and potentially hazardous without proper preparation and precautions.

Moderation becomes more important as temperatures veer toward these thresholds considered hazardous extremes by meteorologists and health officials. Acclimation helps but does not eliminate all danger at these limits.