

**Climate Visualizer
Process Book**

Jacob Blomquist u0843402

William Ludwig u0905218

Overview and Motivation:

While searching for ideas for our visualization project, we stumbled upon an extensive dataset provided by NOAA (National Oceanic and Atmospheric Administration). This dataset contains historical climate data as well as several other large data sets. We are focusing on the GHCN (Global Historical Climatology Network) which contains daily temperature, precipitation, wind, pressure, and other useful and interesting data from climate stations around the world.

We are making a visualization using this data that shows trends in climate. This data helps to show climate trends over periods of time and how the climate may be changing in particular areas. This dataset has so much data that we needed to use multiple visualizations in order to display all of the data. We hope our visualization will allow a user to quickly see trends from any station they choose.

Related Work:

There really wasn't any related work that inspired us to do this project. We were looking for ideas of projects to do and we came across a dataset that had a lot of really good information. We were looking for a dataset set that had enough interesting data that we would be able to make a good visualization out of it. After finding and looking through the dataset provided by NOAA we decided that we would be able to make an interesting and helpful visualization for that data.

Questions to Answer:

With our visualization a user can visually see the climate data for the selected area/weather station. They are able to interact with the visualizations by selecting the area/weather station they want to see data for. They will also be able to select from a drop down what specific area they want to see data for such as: temperature, rainfall, etc. All of these visualizations will be linked together and when you interact with one

view the other ones will update as well.

With our project we want users to be able to easily see climate data from anywhere in the USA so that they can easily select and view whatever category they are interested in. We also want them to be able to see how these categories are changing and/or have changed over time.

The dataset that we are pulling data from has a lot more information than what we are actually needing for our visualization.

The benefits of our project include the ability to easily and effectively view current and past data for a certain area/weather station. We are also planning to include a comparison view where you can select multiple weather stations at once and compare them side by side. This visualization could be very beneficial to people who are wanting to see how climate has changed over the years, i.e. people who study global warming, etc. It would also be interesting to anyone who wants to see how climate change has affected the area in which they live.

Data:

Our data is taken from the Global Historical Climatology Network (NOAA). The link for our data is <ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/>. This data shows high/low temperatures, precipitation, pressure, wind, and more for every day going back to ~1900 for weather stations across the whole world. This is a ton of data so we will be filtering it down and focusing mainly on the United States and on certain categories. The full dataset is roughly 28.3 GB in size. So as we mentioned we will be filtering the data and only using a certain portion of it. We ended up filtering the data to only weather stations within the continental United States which were part of the Historical Climatology Network. There are other station networks such as the Integrated Surface Database, the Global Climate Observation System, and the National Weather Surface which were also included in the dataset which we decided not to use. Even after filtering out all of these, we were still left with approximately 1000 weather stations in the Historical Climatology Network. Once we filtered out the stations we then needed to

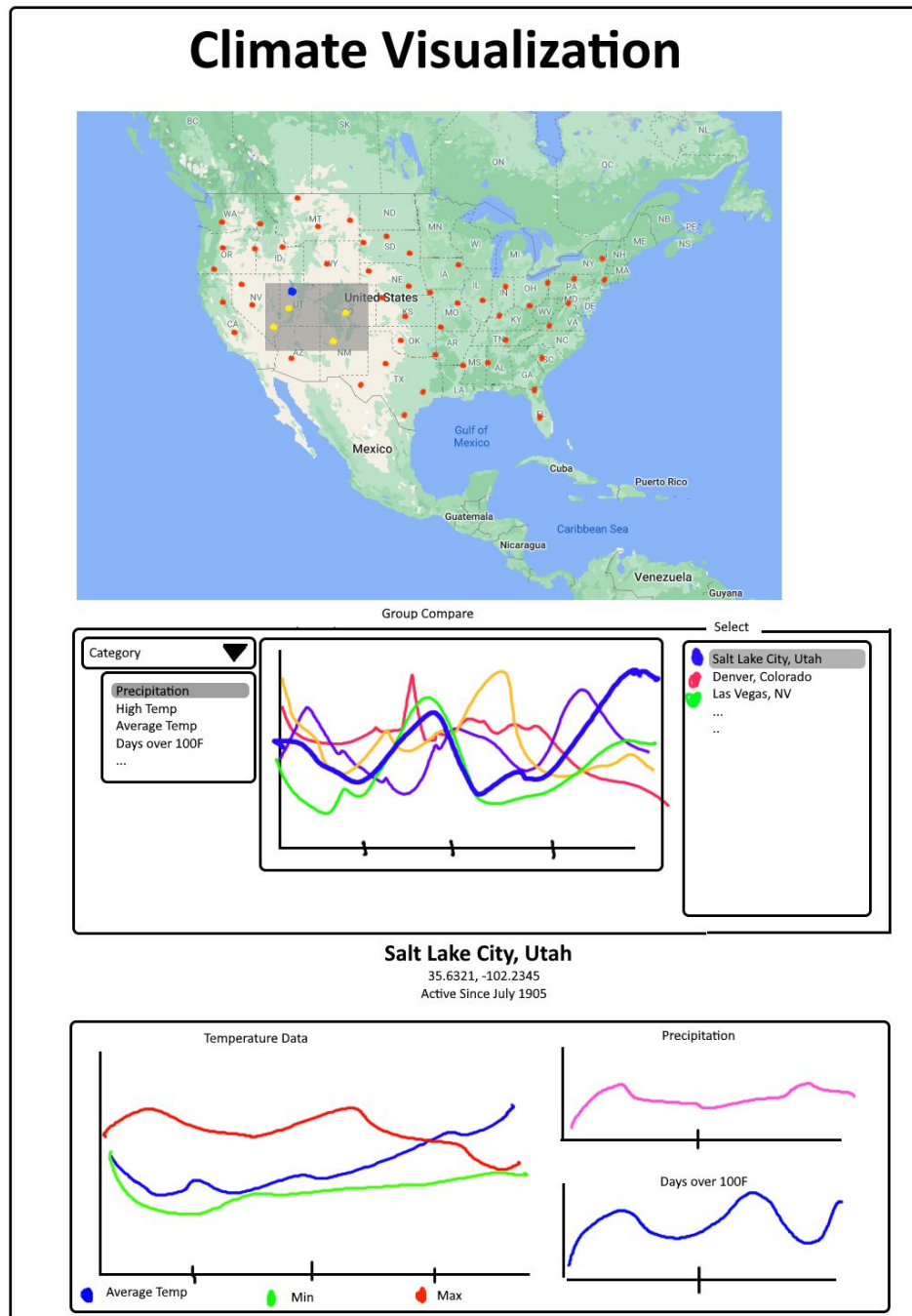
convert the data from the proprietary file format that NOAA uses into JSON. This involved creating python scripts which parsed each file and built up a JSON structure for our needs. We also filtered out a bunch of unneeded data here as well such as soil temperature, cloudiness, wind direction, etc and focused more on temperature. We thought that Python was the best option for us to pull the data out of that huge dataset due to its ease of use.

Exploratory Data Analysis:

When we were exploring this project and the data set that we found we looked into maps and how to use those in a visualization. We were deciding between using a D3 map similar to what we did in one of our assignments and between a google map. After playing around with both map options and researching them we decided that google maps would be better for our purposes because it is more user friendly and can easily zoom in. This allows users to easily pinpoint the location that they want to see data for. Since we are aiming to allow anyone to use this tool to find weather data for their immediate neighborhood/city, google maps is the favorable choice.

Design Evolution:

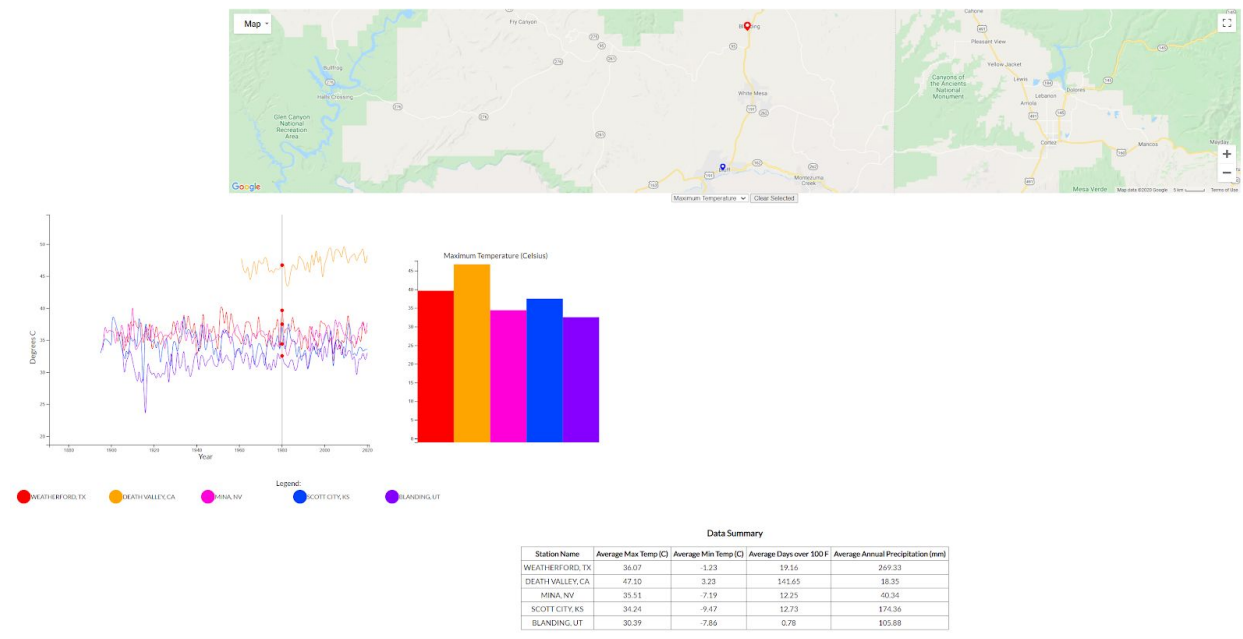
Our Vision:



Final Version:

Climate Visualizer

Names: Jacob Blomquist & William Ludwig E-Mails: u0843402@utah.edu & u0905218@utah.edu UIDs: u0843402 & u0905218



© Jacob Blomquist | William Ludwig 2020

Our original design only had the map and some charts, however, we decided to add a table as well along with the other things that we had planned. This table shows easy to read numerical data. This was also an easy way for us to make sure that the visualization/interactivity was working correctly and that the correct data was being displayed.

The TA recommended that instead of starting out with a marker for every single weather station across the US that we cluster them together and that you could click on a cluster and it would zoom into that area (The above screenshot is zoomed in really far). We took her advice and implemented our map that way.

We originally planned on having a more detailed section of all of the line charts of the different categories for one selected station, however, we decided that this was

redundant and so we decided to allow for a hover feature on the line chart that created a Bar chart that shows a direct comparison for the current category, the selected year, and up to 5 selected stations.

We uploaded an image of our visualization to an online color blindness tester and all of the colors that we are using are differentiable between each other.

Implementation:

11/4/2020

- Created the folder layout and initial files for the project.
- Added the google map to our visualization

11/6/2020

- Began processing data.
 - Filtered out unwanted weather stations.
 - Parsed weather station location and elevation data.

11/10/2020

- Processed the data
 - Processed all daily weather data and condensed into monthly data.
- Associated the data with the appropriate markers on the map.
- Loaded the data in from the JSON files
- Displayed the data in a table

11/19/2020

- Created initial graph
 - Scales based on dates and temperatures
 - Basic hovering selection
 - Tooltip

11/25/2020

- Added a vertical brush for selecting a column
 - Gets the value at the point specified
- Added map clustering to reduce clutter

11/27/2020

- Polished the brushing
- Added bar chart comparison
- Added color support
- Made it look better on mobile.

11/29/20

- Created a Drop Down menu for selecting the category that you want to view
- Created a clear selected button to be able to clear all the currently selected stations
- Updated the Line Graphs and Bar Chart so that they update based on the currently selected category.

11/30/20

- Checked our visualization for color blindness deficiency
- Created the legend that show which color corresponds to the lines on the line chart and the corresponding color in the bar chart
- Updated the table to show detailed average numerical values for all categories since the stations were created
- Played with the formatting of everything to make it look better

Evaluation:

The data was a lot noisier than we thought and there isn't complete data for every month so this messed up some of our visualizations and we had to figure out a way around this. We successfully visualized the data that we wanted to. We have successfully shown the max temperatures, min temperatures, days over 100 degrees Fahrenheit, and the amount of precipitation. We have also shown the average yearly value for all of these categories since the weather stations were created.

If we were to improve this visualization we would include more categories that the user could select from. The dataset that we got our data from has so many categories that it is kind of overwhelming. We would like to choose a few more that would be interesting to include.