Visualization of Utah's Air Quality

Data Visualization

Alex Yang (u1444424, u1444424@utah.edu)
Noah Carlson (u1254179, noahrk14@gmail.com)

https://github.com/xxsyang/CS5630_Project

- **Background and Motivation**

  At the beginning of this semester, we noticed that the air quality had been quite poor here in Salt Lake. Due to fires and pollution, everyone in Utah could smell the bad air beneath the smokey, dusk-colored sky. On top of the current bad air quality, Salt Lake has a couple of major patterns of decreasing air quality. The first major pattern is a weather inversion every winter season. This weather inversion traps pollution in the valley causing extremely harsh air quality conditions. The second pattern is the decreasing health of the Great Salt Lake. With lower water levels in the lake, more pollutants and particles are being released into the air furthering the deterioration of the air quality. As a result, we decided to create a visualization tool to display the air quality across every county in Utah over the years 2005 - 2022. In this way, we may be able to recognize some new patterns of deterioration or reasons behind increasingly poor air quality.

  As aspiring software engineers, we want to practice our skills in creating intuitive and coherent products. We aim to provide a tool that visualizes air quality, along with two to three interactive features. For instance, when hovering over each county on the map, users will be able to view detailed information about the air quality in that area. This project also allows us to put into practice the theories we have learned about how humans interact and perceive data. For example, color theory will be useful in the heat map for displaying pollutant density. These skills are something engineers do not always focus on but are one of the most important aspects when creating a project for users.

- **Project Objectives**

  The primary goal is to identify patterns and trends in air quality across Utah's counties over time, focusing particularly on PM 2.5 levels. By analyzing data across multiple years, we are trying to to:

  1. Detect any significant changes or tendencies in air quality for specific regions.
  2. Identify periods (yearly, monthly, or daily) when air quality either improved or worsened.
  3. Determine which Utah counties experience the worst PM 2.5 pollution compared to others and at what time

- **Data.**

  PM 2.5 data.

  Utah counties' boundary geojson data.

- **Data Processing.**

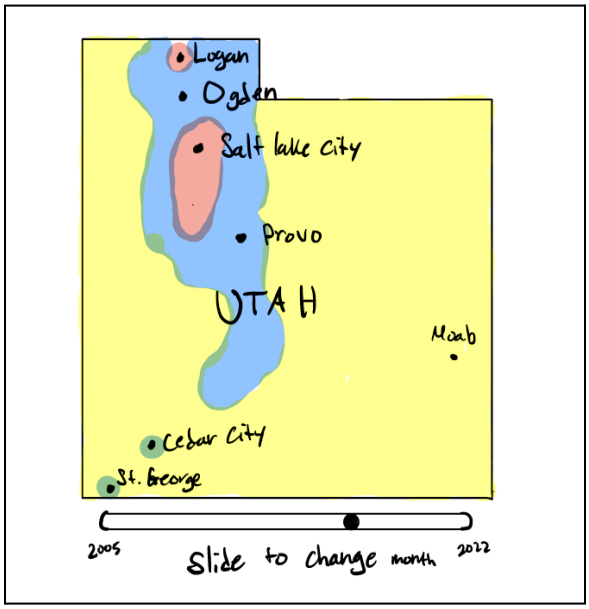
  The dataset we chose has a comprehensive list of many Utah PM stations starting in 2004 to 2022. We found that this data is hourly data points of the pollutant levels so we will have to aggregate the data into daily, monthly, and yearly summaries for each station. For the aggregations of data, we will find minimum, maximum, and averages for each time period per location. This allows for variable amounts of precision of the information ranging from visualizing trends between morning to night to between seasons.

  Also under closer inspection, we found that 2004 is not complete as well as the 2022 CSV file has some missing data. This leads us to believe more CSVs will be missing data points for which we will drop those columns. Another special case is that one station is missing an entire month of data. In this case, instead of dropping the station's data entirely, we can consider using interpolation or other imputation techniques to estimate the missing values based on the available data from nearby stations or from previous months' trends. This will allow us to retain as much information as possible for our analysis without significantly impacting the overall results.

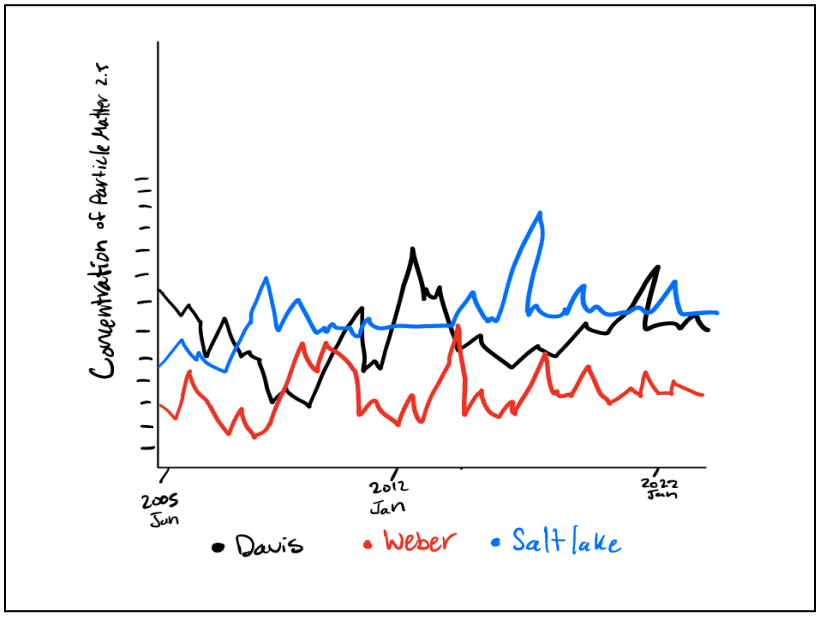
  To process this data as we stated above we will use Python Pandas to convert the original dataset into the dataset we need. Then we will use javascript to read these files and parse them into the right objects for visualization. Finally, to visualize the data we will use a mix of html, javascript, css, and d3.

- **Visualization Design.**

1. **Interactive heat map of Utah** along with a slider that changes the time period shown. The colors on this heat map will also represent the density of the air pollutants.



2. **Line graph** showing over-time data per each location



3. **Ordered table list** sorted by minimum, maximum, or average over given timespan

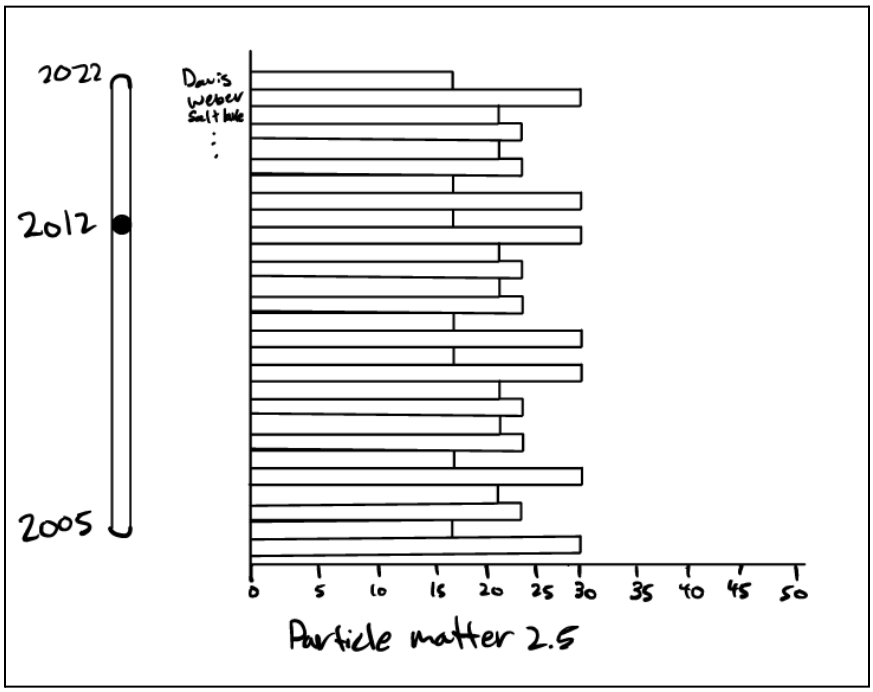
Particle Matter Per Location

Location	Minimum	Maximum	Average
Davis	2	29	11
Weber	.5	24	9
Salt Lake	1	38	20
.	.	.	.
.	.	.	.
.	.	.	.

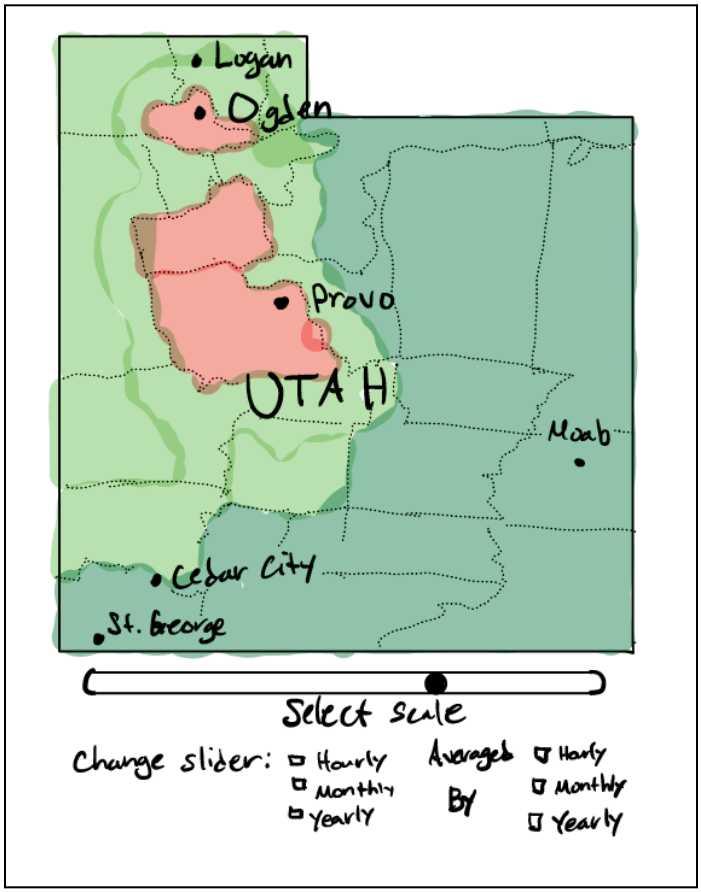
Particle Matter Per Location Over Time

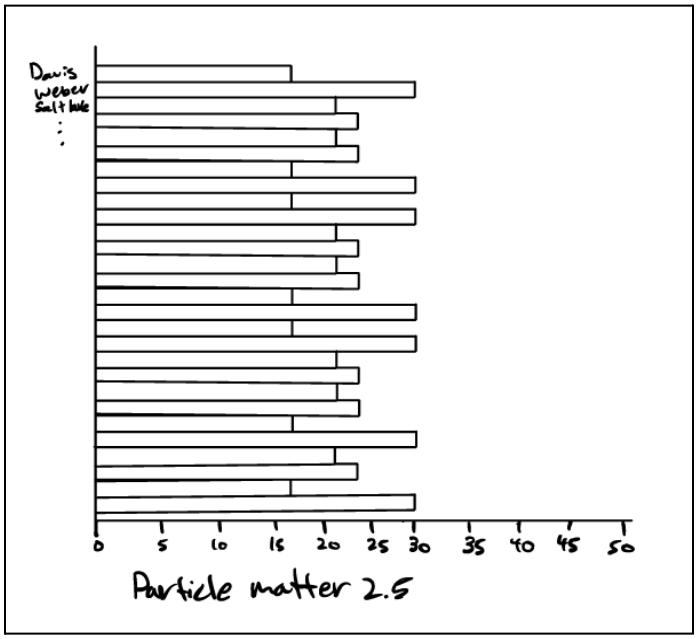
Time	Location	Minimum ▲	Maximum	Average
2005	Davis	.5	29	11
2005	Weber	2	24	9
2005	Salt Lake	1	38	20
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
2022	Davis	2	27	16
2022	Weber	3	28	18
2022	Saltlake	4	42	24

4. **Bar graph** of data displayed by location and value at a certain time



5. Final Design





- **Must-Have Features**
    - Heat map of air pollutants density
    - When the mouse hovers over one area of the map, it will show the information about PM 2.5 under this area.
    - Hovering over an area of the map will highlight the corresponding bar on the bar chart.
    - A slider or dropdown to traverse the overtime data
    - Color to differentiate between low and extreme values
    - Bar chart to display more in depth information

- **Optional Features**
    - Animate the change over time data for the heat map
    - Using the three.js library for rendering our plot into 3D.
    - More specific information on air quality representation of Counties. (The geojson data we have separates Utah into counties. For some counties there are multiple stations therefore we have decided to average the levels of these stations to gain a county wide value. This loses some information but does mean there is room for more specificity in our visualization as an optional feature.)
    - Combine weather data with the air quality data
    - Combine traffic data with the air quality data
    - Track snowfall / snow melt in comparison to the air quality data
    - Using a more detailed map api

- **Project Schedule**

  From now to the Fall break: We will finish the data exploration and clean up. Then use javascript to read these files and parse them into the right objects based on the D3 library.

  From Fall break to November 1st: We will work on creating basic data representations in d3. We will also make the basic setup of html and css on the github website.

  From November 1st to December: We will work on the d3.js library for plotting our graphs, and use bootstrap or other recently popular stylesheets used in the industry to make our UI more complete.

  Weekly: We will have a team meeting for reporting each other's process, problems so far, and the planning for the following week. After sharing our daily schedule, we decided to set up this meeting for after Thursday's lecture.