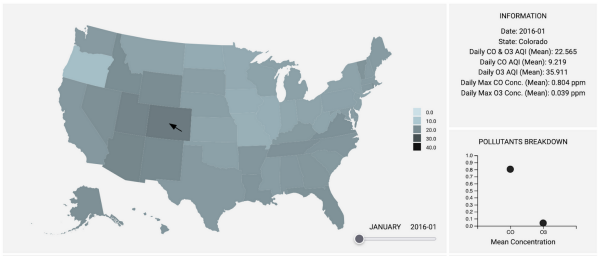
Milestone 5: Final Report

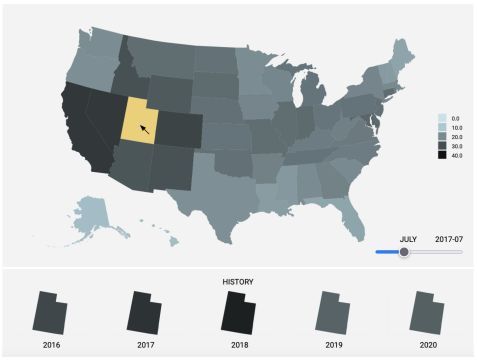
Data Set Characteristics

After a few iterations of data sets due to realizations that they did not meet our initial expectations and from narrowing our scope, we settled on data collected from the United States Environmental Protection Agency. From their portal, we manually downloaded "Outdoor Air Quality Data" for each state by year between 2016 and 2020 and focused only on the Carbon Monoxide (CO) and Ozone (O_3) pollutants. Before wrangling, the data set had 1940310 rows and 22 columns. Many of these columns are specific to their research, so they are irrelevant to our visualization project. We kept the 'Date', 'Daily Max 8-hour CO Concentration', 'Daily Max 8-hour Ozone Concentration', 'DAILY_AQI_VALUE', and 'STATE' features. Then after wrangling, the data set was truncated to 3060 rows and 7 columns. These columns include 'Date', 'Date_Keys', 'DM8O3_Concentration', 'State_Fixed', 'DM8CO_Concentration', 'O3_AQI', and 'CO_AQI.' These features are preprocessed from the raw data. The preprocessing step first involved slicing the 'Date' values into Year-Month pairs, and then grouping all of the rows by same Year-Month pair and taking the mean of the DAILY_AQI_VALUE for each pollutant. This means we created a 'monthly mean' by taking the mean of all the recorded daily AQI means and respective pollutant concentrations. We can confirm that our data is ready for visualization because (51 states) (12 months)(5 years) = 3060, which is equal to the number of rows in the data set.

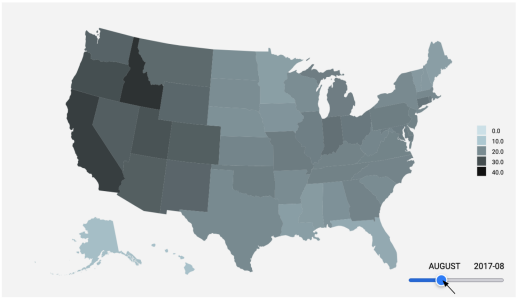
Interface Images

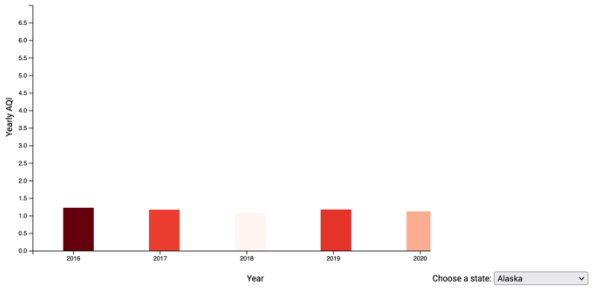


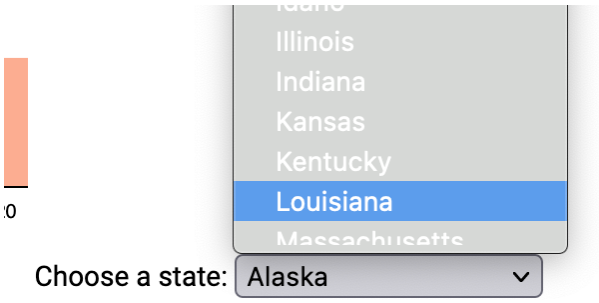
This view shows on-hover details when mousing over a state in the INFORMATION and POLLUTANTS BREAKDOWN panels on the left. The INFORMATION panel informs the user about general statistics, such as DATE, STATE, AQI, and Concentration values.

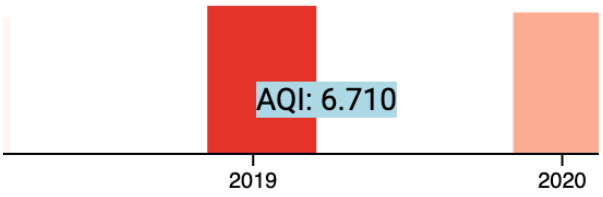


This view shows an updated HISTORY panel when mousing down on a specific state; the state highlights in yellow. The HISTORY panel is populated with multiple of the mouse-down state colored by AQI across all of the years in the data set for the current month.

<table>
<tr>
<td> 

This view shows the slidable time range from 2016 to 2020, which changes and updates the whole map based on data for the selected year-month pair.</td>
<td> 

This view shows an overview of the bar graph visualization. The colors of the bars are mapped to its AQI values. A darker color means a higher AQI while a lighter color means a lower AQI.</td>
</tr>
<tr>
<td> 

A dropdown is available to select the state that the bar chart will show data for.</td>
<td> 

Hovering over a bar with a cursor shows a popup with the specific AQI value rounded to three decimal places.</td>
</tr>
</table>

Group member contributions.

Abner Benitez	Soonyoung Hwang	Randy Truong
- Contributed to M2 (Domain and Data, Tasks) - Proposed interactive bar chart linked to map with brushing - Created first three design sheets in M3 - Contributed current progress image to M4 - Data wrangled the necessary	- Worked on sources and bibliography for M2. - Otherwise, N/A.	- Proposed visualizing human byproducts, such as waste or pollution. - Guided meeting conversations, brainstorming, design thinking - Wrote user stories, did iteration document, and helped with sources and bibliography for M2 - Contributed and presented M3

data for bar chart - Coded and developed interactive bar chart in abner.js - Presented M5 Final Project (bar chart visualization) to the class - Contributed to M5 Final Report		design document - Wrote M4 status update - Coded and developed visualization system in randy.js, which includes the various panels, temporal choropleth map with slider, mouseover details, mousedown interactions in HISTORY, breakdown chart, and color and design. - Presented M5 final project; wrote M5 paper

Reflections

The implementation mostly went according to plan. Although we were behind our anticipated timeline for implementation, we created a system that showed our intended features, including a temporal bar chart, a choropleth map, and breakdown chart. Our initial idea hoped to implement a contiguous cartogram to better represent the data through area/size as a visual encoding; however, this was later scrapped due to its technical complexity and lack of time. Initially our project was about visualizing municipal solid waste generated in the US, but we could only find datasets that aggregated data for all states. We were told the only way to obtain state level data is to request that data for each individual state's environmental department. We decided to pivot and visualize the air pollution since there are state and county level datasets on this topic and is semi-related to our original plan of visualizing harmful emissions that humans create. Originally, we wanted to visualize the global air pollution but then we shifted from global data to the United States because the number of countries represented in global data sets compared to the total number of countries in the world was very low and those that were in the dataset sporadically lacked yearly data. One major roadblock was finding and collecting the right data that had relevant features, such as dates and mappable values with an appropriate scale. The data collection was very manual and laborious and data preprocessing was also challenging. The full dataset starts at 1990's meanwhile we were only able to wrangle data starting from 2016.

Aside from the discontinued cartogram, shift from visualizing waste to visualizing air pollution, and shift from a global view to a domestic view, our design did not diverge much. We were flexible and

open to new implementations and designs. For example, the HISTORY panel was initially not intended to be part of our system, but it was implemented spontaneously after seeing that memorability could be improved. Due to using different versions of D3 and lack of time and coordination, we were unable to synchronize the temporal bar chart and the choropleth map. Some other roadblocks include having off-schedule system development and faltering communication between group members. Finally, the team did not work well together. One team member disappeared in the middle of the semester and thus did not contribute to many of the important aspects of the project. The two remaining team members did not have ideas or schedules that aligned well with each other. Most of the group meetings through Zoom were productive and the meeting recording was helpful for the members that couldn't attend the meeting or to play back at a later date. Project development and technical implementations were put off until the last minute.