PROJECT 3 - GROUP 8

Hannah Lashway
Alex Moore
Jessica Price
Andrew Vick

GLOBAL AIR QUALITY INDEX

Introduction

Our group chose an Air Quality Index (AQI) dataset for this project. The U.S. AQI was developed by the Environmental Protection Agency as an index for measuring, reporting and recording air quality. The AQI measures the presence of five specific pollutants (carbon monoxide, ground-level ozone, nitrogen dioxide, particulate matter (PM2.5 and PM10) and sulfur dioxide) in the air and assigns a score between 0 and 500 based on the prominence of those pollutants. The AQI is divided into six color-coded categories; each category has a corresponding health warning. The higher the AQI, the higher the health hazard.

AQI Basics for Ozone and Particle Pollution

Daily AQI Color	Levels of Concern	Values of Index	Description of Air Quality
Green	Good	0 to 50	Air quality is satisfactory, and air pollution poses little or no risk.
Yellow	Moderate	51 to 100	Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive to air pollution.
Orange	Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is less likely to be affected.
Red	Unhealthy	151 to 200	Some members of the general public may experience health effects; members of sensitive groups may experience more serious health effects.
Purple	Very Unhealthy	201 to 300	Health alert: The risk of health effects is increased for everyone.
Maroon	Hazardous	301 and higher	Health warning of emergency conditions: everyone is more likely to be affected.

The purpose of our project was to analyze the AQI values of different cities and countries to gain a better understanding of the state of air quality around the world. Through this, we hope to raise awareness about air pollution and its impact on health and the environment.

We believe that by breaking down air quality data into intuitive, interactive data visualizations, we can empower individuals and communities to make more informed decisions about their health and wellbeing.

Data Cleaning/Database creation

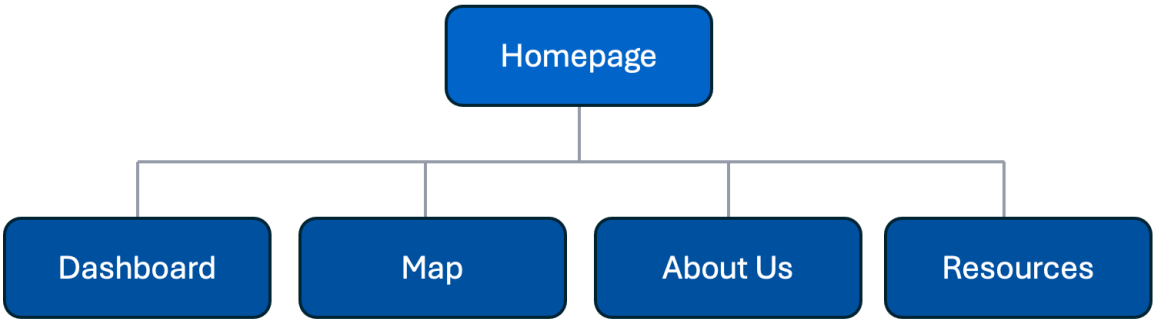
The dataset, *World Air Quality Index by City and Coordinates*, was pulled from Kaggle.com: https://www.kaggle.com/datasets/adityaramachandran27/world-air-quality-index-by-city-and-coordinates. The file was a .csv and included everything we needed to meet this project's requirements, including: global data, coordinates for plotting locations on a map and enough rows of usable data to analyze and draw conclusions about the state of global air quality. Fortunately, the original dataset did not need much cleaning. The original .csv file came with 16,695 rows of data with some null values in the “Country” column. After dropping the null values, we were left with 16,393 rows. This was more than enough data for our group to create the necessary bar charts and heat map. Our Next step was converting the .csv into a .sqlite file using Python in a Jupyter Notebook and creating a database in pgAdmin 4. Once we created our .sqlite file, we were ready to build our web app in VS Code.

Color design considerations

For our color design, we decided to follow what the industry colors are around the world. We matched each category with that found online. Hazardous is a maroon color, very unhealthy is purple, unhealthy is red, unhealthy for sensitive groups is orange, moderate is yellow, and good is green. We applied these colors to both our bar charts as well as on ouer map as the different colors for the overlay heat map. We used the common logo for aqi category as our logo/home page on the website itself to match all the colors together as well.

Website architecture

The website backend was designed using Flask, a backend micro-framework for Python and SQLite, a serverless database engine. The website frontend was built using Javascript, HTML and CSS.



Homepage

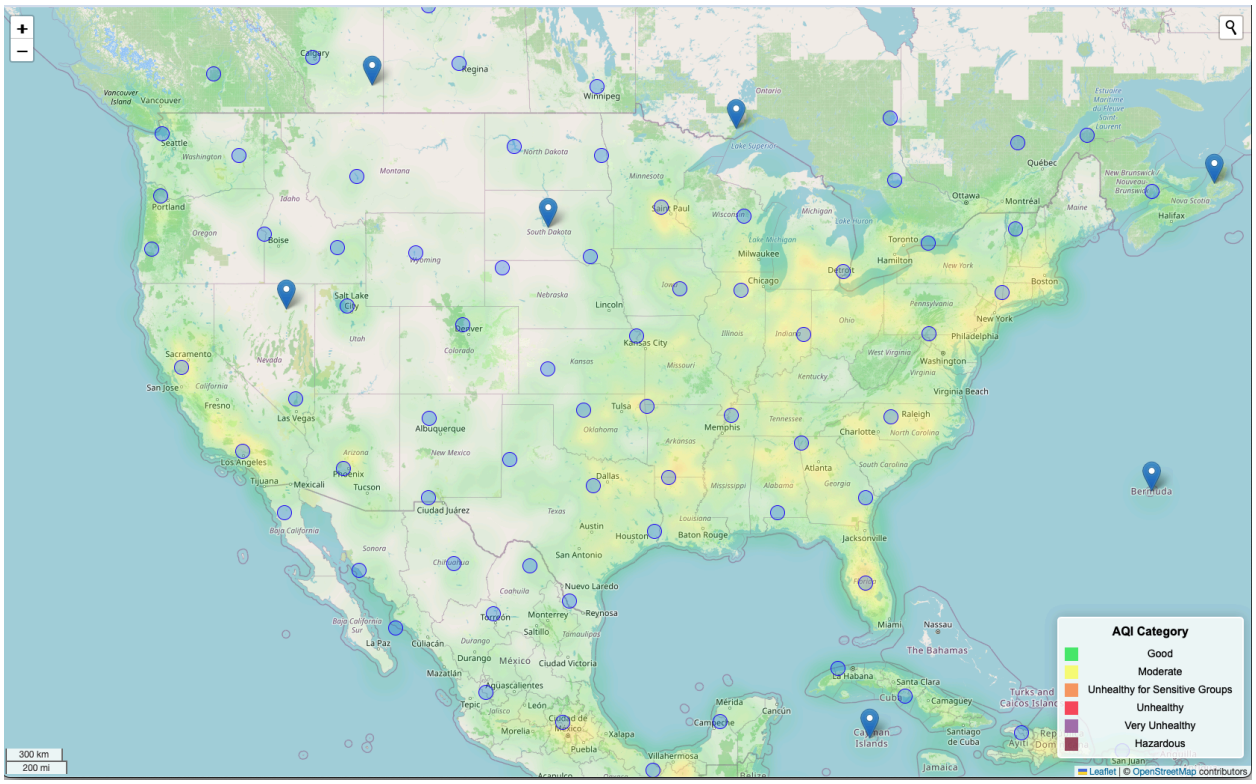
The Homepage served as the website landing page. It included general information explaining what AQI is and what the categories are. It also included a section on current AQI data in the real world.

Dashboard

The Dashboard page was designed with the user in mind. It offers an interactive dropdown menu with a list of countries the user can choose from. Once a user selects a country and clicks the “Filter!” button, the data visualizations update to show the AQI data in the selected country. See additional information in the Dashboard Design Concepts section below.

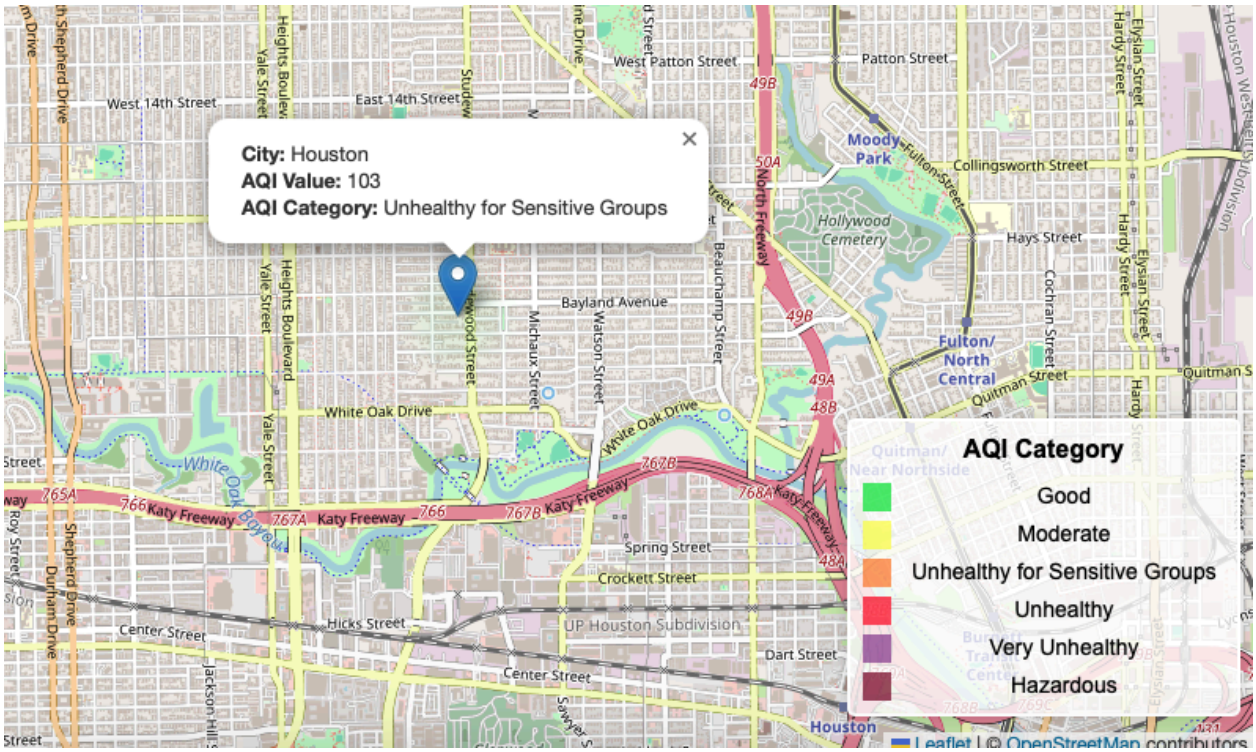
Map

For our heatmap, we wanted to use the same color coding system as the official Air Quality Index. We first based the colors of the heatmap on the AQI Value, but the map was not popping the way we wanted. The colors were not bleeding as much as we would have liked and looked a little boring. We attained a much better visual result by switching from the AQI value to the AQI Category for our heat layer. The “Moderate” layers bled out more from the “Good” layers and became more prominent for the user. The user could also choose how to view the map by checking the filter boxes to display only the markers or the heat layers. We were now very happy with the aesthetic presentation of our AQI map.

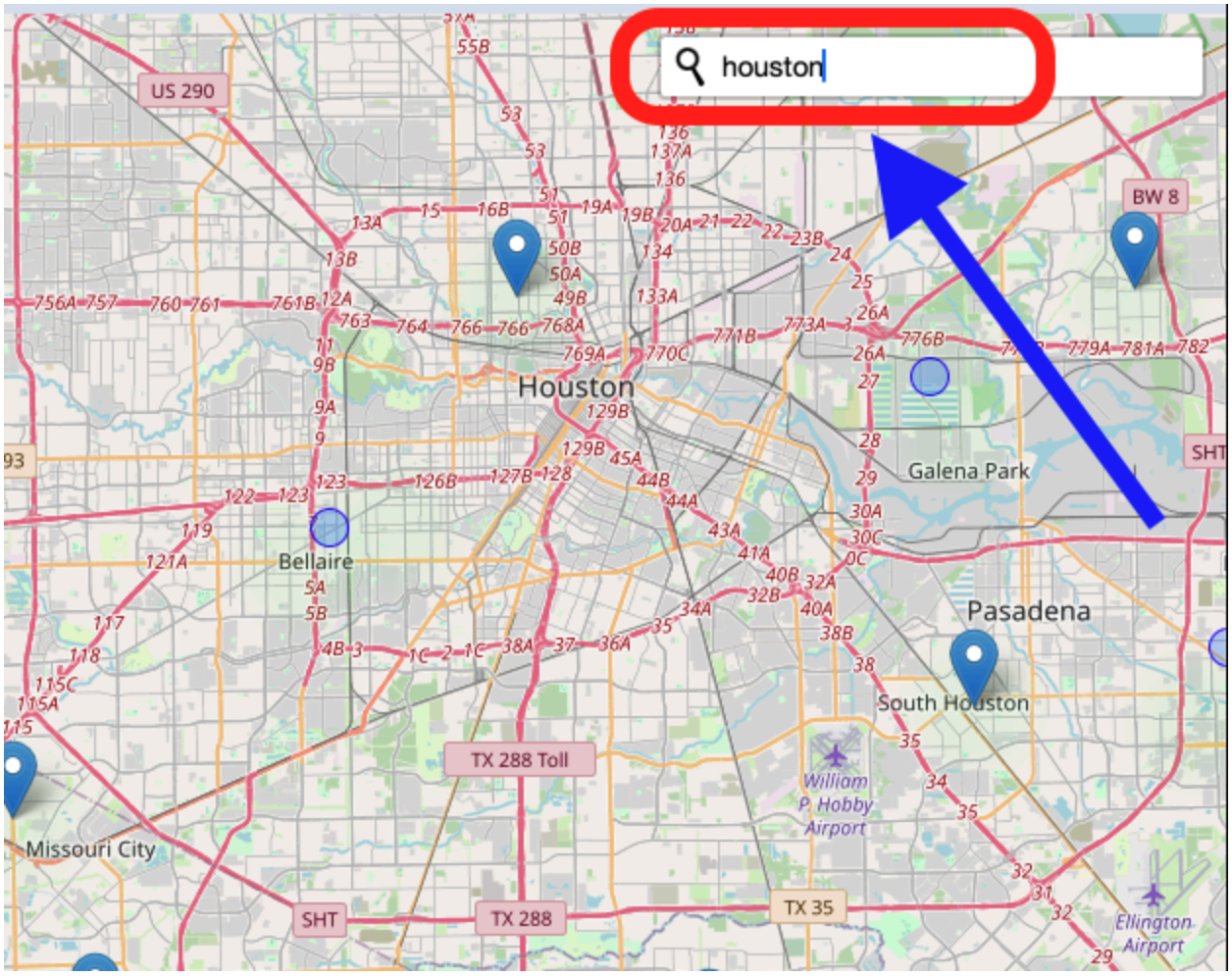


Next, we planned out how to best represent the data by applying markers to the map. There are over 16,000 entries in our dataset, so we had to apply marker clusters so as to not inundate the map with clutter. Then, we worked on what each marker needed to display when the user clicked on it. Originally, we had the City, Country, AQI Value, and AQI Category listed in the marker popup window, but then we noticed a problem with our dataset. In most cases, cities that have the same name in other countries would sometimes have the wrong country listed. For example, when the user clicked on the marker for Paris, France, the user would see the country labeled as the United States of America. There is a Paris, TX in the United States, thus the mixup. This was the case for a large number of cities; every other data point was correct except for the country label. Our group assumed that whoever created the original dataset used some code to identify the countries based on

the cities rather than the latitude and longitude coordinates. So, we took out the “Country” label to make it look like there was nothing incorrect with our map or dataset.



As a final touch, we added a search bar in the top right corner. This gives the user the ability to search for any location in the world regardless of whether or not we had data for that particular city. You could search a specific country or city and the map would automatically refresh to that location. This made our map very user friendly and intuitive. There would be no tedious clicking of marker clusters or scrolling to navigate the map for our users. We wanted to give them the ability to search for any location that would pop up in their head, and it could not have come out any better. The search function was our favorite aspect of the entire map; its ease of use was unparalleled. The heatmap was now ready for launch.



About Us

The About Us page gives a brief introduction to the team (including photos), features our project mission statement and provides contact information for those who want to reach out to the team.

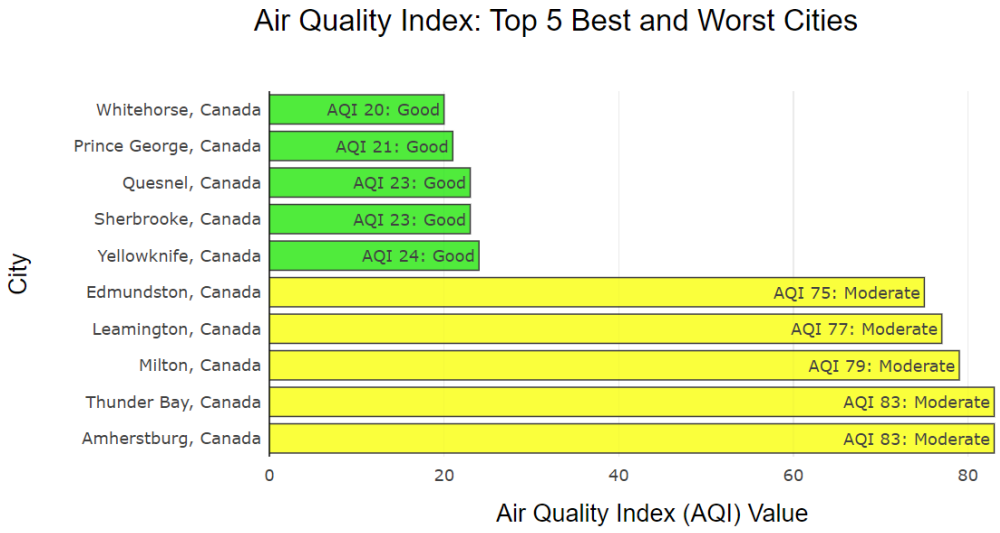
Resources

The Resources page provides the user with all of the resources used to produce this project.

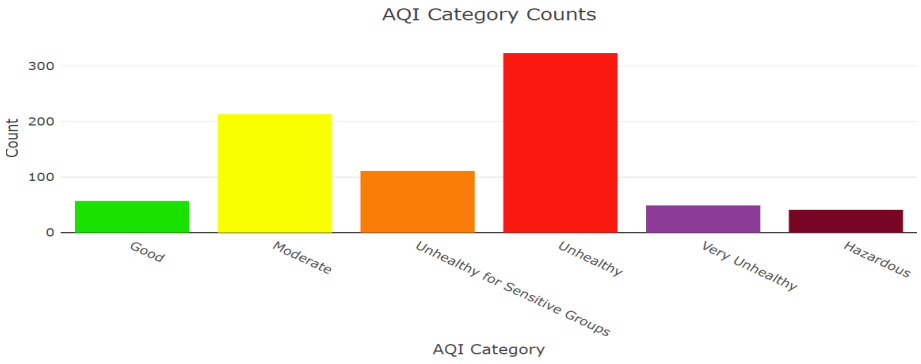
Dashboard Design Concepts

We decided to use two bar charts and a table to display our data and help answer the research questions we posed at the beginning of this project. Our first bar chart displayed the top 5 cities with the best aqi value and bottom 5 cities that had the worst aqi values in

the selected country. We decided on doing the top 5 and bottom 5 instead of just the top 10 or bottom 10 because this could give the viewer a better idea of what the country as a whole might be rather than to give the wrong impression if we were to just show the best or worst cities and could lead to a misrepresentation of the data.



Our second bar chart that we created, we used the aqi category column in our dataset to display the count of cities in the selected country with the corresponding aqi category. The possibilities for aqi category in our dataset were good, moderate, Unhealthy for Sensitive Groups, unhealthy, very unhealthy, and hazardous. These categories are another representation of the aqi value except in a little more understandable terms for a viewer who might not know all about aqi values and what they represent. This bar chart also gives a more accurate representation of the country that was selected as a whole and where; on average, the country sits in terms of its aqi category. Our chart below is of the country India. This is a perfect example of why having the count of aqi category is a great chart to have because if we only had the first chart no one would be able to tell if India is super polluted, or super clean because they have cities in both the good category and hazardous category. However, with this chart you're able to see every city in India and can take away that they are around the middle of clean and polluted.



For our last addition to our dashboard we included a table that would change when the user selected a country with the other two charts. This table displays the first 10 cities in alphabetical order of the user's selected country. If there are more than 10 cities with data in the selected country, multiple pages will be created for the user to flip through. This table includes the city name, aqi value, aqi category, co aqi value, no2 aqi value, and the latitude and longitude of the city. This table is able to be sorted ascending or descending for any of the columns; this includes sorting alphabetically for the city names. It also includes a search bar in case the user would like to search for a specific city by name within the selected country.

AQI Value by Country

10 entries per page

Search:

City	Aqi Value	Co Aqi Value	No2 Aqi Value	Latitude	Longitude
Auckland	69	0	0	-36.8406	174.74
Avarua	28	0	0	-21.207	-159.771
Gisborne	27	0	0	-38.6625	178.0178
Gisborne	27	0	0	-37.49	144.5889
Hawera	26	0	0	-39.5933	174.2783
Invercargill	14	0	1	-46.4131	168.3475
Levin	13	0	1	-40.6219	175.2867
Lower Hutt	65	0	0	-41.2167	174.9167
Masterton	13	0	2	-40.97	175.65
Nelson	14	0	1	-41.2708	173.2839

Showing 1 to 10 of 28 entries

« ‹ 1 2 3 › »

How does your dashboard answer your research questions?

Our dashboard was able to conclusively answer our first two research questions and our dashboard and map provided valuable insight to our third question.

Addressing our first question, which looked to answer which cities ranked worst overall for air quality, we found that the 5 worst cities, all with the worst possible AQI values of 500, as

indicated by our dashboard were in countries like Pakistan, Russia, South Africa. We did have what we believed to be an outlier here and that was Durango, Colorado. While Durango did also have an API score of 500, we noted that unlike the other cities, whose poor air quality could be linked to industrial pollution, Durango's poor air quality was more-so a result of it's geographic location. Winds often blow wildfire smoke in from the North and West, and get trapped by The Rockies and settle over Durango which sits at the foothills of those same mountains.

Our second question looked to answer, rather than the worst quality, where was the best air quality located globally. Based on the data on our dashboard, we could conclusively say that Ecuador held the title of best air quality globally. Three of the top five cities in our dashboard were located within Ecuador. When we filtered our dashboard to just look at Ecuador's data, we observed that 55% of their AQI ratings were 'good', the best possible rating, and the other 45% were all rated as 'moderate', the second best possible rating.

The reason that we stated our dashboard and map provided valuable insight to our third question, but not a conclusive answer, is because the question, which looked to see if there were similarities in the AQI, CO and Ozone levels between countries, is relative and changes dependent on which countries we are looking at. Take Europe for example, the majority of the cities within European countries have their AQI values classified as 'good' or 'moderate' so if we are comparing countries like Germany and France to each other, then yes there is a lot of similarity between their values. However, if you look at cities in China or India, they have much lower counts of cities with 'good' and 'moderate' AQI ratings and many more counts of 'unhealthy' and 'hazardous' AQI values. This may be due to their higher populations, and thus higher transportation pollution, and more notably their high levels of industrial pollution.

Bias/Limitations

As common with most projects, we became more aware of the biases and limits of our dataset, the further along we got into our project. The first, and most obvious one we noticed, was that the country labels weren't always correct in our dataset. We knew this because the latitude and longitude values provided would correspond to a city in one country, however when we hovered over that point on our map, a different country would be shown, and this affected our bar charts because of the country filter we used in our dropdown selection on our dashboard. An example of this was San Antonio, located in Texas within the United States, however when we hovered over it, the country displayed was Costa Rica.

Another limitation of our dataset was its size - we had quite a large dataset which was beneficial for our dashboard, however, with so many individual values, it actually took away

from the impact of our heat map. There were so many values that it was difficult to clearly distinguish areas with significantly better or worse AQI values.

While doing research into the AQI value, in terms of what it means and how it is measured, we discovered that many countries have their own, different standards for measuring the air quality. Some countries measure all of the same pollutants but many measure less, and additionally have different rating systems for the air quality. What was not clear in our dataset, is if all of this data has been collected and measured following the EPA's guidelines for measuring the AQI or if other countries' data has been measured using their respective systems and converted to the American equivalent quality. The significance of this is, had conversions been the case with international data, we can't accurately compare countries because some countries do not measure as many pollutants and thus, their AQI ratings don't take those pollutants into consideration like we do in the United States.
Lastly, as we noted in addressing our first question, while researching why Durango, Colorado, was an outlier in the worst air quality rating, we discovered the impact of natural disasters on the air quality in not just affected but surrounding areas as well. Given that information, and the fact that we do not have more information regarding natural disasters like volcanic eruptions and wildfires across the globe, we do not know if there are any other AQI values that have been affected and skewed as a result and thus, consider that a limit in our dataset.

Conclusions/Reflection

In the two or so weeks we spent as a group working on this project and creating a webpage to display different ways to answer questions we originally posed, we not only learned a lot of great information on aqi values, aqi categories, and much more relating to it, we also got a better understanding in javascript, html, and css. This project has helped teach us how to work as a team and how to complete different tasks on the same project to get the project done sooner than if we all worked on everything together. In conclusion this project was a great way for all four of us to learn more about coding, but also what working on a project at a real job will be like with multiple different collaborators.

Works Cited

AirNow. “AirNow.” Airnow.gov, 2010, www.airnow.gov/.

“Air Quality Data Visualization for Environmental Research.” Atmotube.com, atmotube.com/blog/air-quality-data-visualization-for-environmental-research.

Kirk, Alexander. “Denver's Air Quality Is among the Worst in the World on Tuesday.” 9News, 9News, 23 July 2024, www.9news.com/article/weather/weather-colorado/denver-air-quality-wildfires-smoke-haze-rank-world/73-2ed93eaa-335f-4d04-8299-c17a33354727. Accessed 13 Aug. 2024.

“Standards for Air Quality Indices in Different Countries (AQI).” Atmotube.com, atmotube.com/blog/standards-for-air-quality-indices-in-different-countries-aqi.

“World Air Quality Index by City and Coordinates.” www.kaggle.com, www.kaggle.com/datasets/adityaramachandran27/world-air-quality-index-by-city-and-coordinates.