ABSTRACT OF THE PROJECT

Interactive Pollution Map

The main objective of the project is to implement an interactive map visualization that enables the users to view the pollution data effectively and allows them to interact directly with the map in order to construct their own understandings of the information. The map allows user input to facilitate interaction. The input to the pollution map is in the form of filters. These filters provide the ability to select multiple pollutants and time frames to visualize on the map. The implementation of the project is a web application which displays the interactive map and air pollutant information along with multiple statistics. Geospatial data is used for mapping pollution in each state in the United States.

The interactive map helps the users to determine patterns among different pollutants, change of pollution over time and intensity of pollution at a given time and place. This information is useful in determining the cause of pollution and predicting the health hazards related to the pollutants.

INTRODUCTION

Interactive visualization is a field of computer science and programming that is focused on graphic visualizations and improving the way we can access and interact with information. For visualizations to be considered interactive, they must have a feature of human input. The input can be taken by clicking on a button, moving a slider, selecting option from dropdown, text in search text area. The response time of visualization must be quick enough to show a real relation between input and output. Visualizations enables the users to explore, manipulate, and interact with data by using dynamic charts, changing colors and shapes based on queries or interactions.

Interactive visualization is emerging as a vibrant new form of communication, providing compelling presentations that allow viewers to interact directly with information in order to construct their own understandings of it. The power of interactive visualization is used in the this project to view the air quality index for each state in the U.S. The numeric value of air quality index is not much useful for analysis if they are not compared against each other. The comparison of pollution in different states helps to analyze many useful factors like industrial impact on the environment and health of the people.

DATA

The Dataset is in the form of CSV which contains Air Quality Index of four pollutants namely Nitrogen Dioxide, Ground level Ozone, Sulphur Dioxide and Carbon Monoxide. The AQI for all these pollutants are recorded thrice per day for all states. The following figure shows the raw dataset and its fields.

State Code	Site Num	Address	State	Date Local	NO2 AQI	O3 AQI	SO2 AQI	CO AQI
4	3002	1645 E ROOSEVELT ST	Arizona	1/1/2008	46	34	13	24
4	3002	1645 E ROOSEVELT ST	Arizona	1/1/2008	46	34	13	25
4	3002	1645 E ROOSEVELT ST	Arizona	1/1/2008	46	34	13	26
4	3002	1645 E ROOSEVELT ST	Arizona	1/1/2008	46	34	13	25
4	3002	1645 E ROOSEVELT ST	Arizona	1/2/2008	34	27	4	24
4	3002	1645 E ROOSEVELT ST	Arizona	1/2/2008	34	27	4	26
4	3002	1645 E ROOSEVELT ST	Arizona	1/2/2008	34	27	4	21
4	3002	1645 E ROOSEVELT ST	Arizona	1/2/2008	34	27	4	26
4	3002	1645 E ROOSEVELT ST	Arizona	1/3/2008	48	14	16	25
4	3002	1645 E ROOSEVELT ST	Arizona	1/3/2008	48	14	16	28
4	3002	1645 E ROOSEVELT ST	Arizona	1/3/2008	48	14	16	30
4	3002	1645 E ROOSEVELT ST	Arizona	1/3/2008	48	14	16	28
4	3002	1645 E ROOSEVELT ST	Arizona	1/4/2008	72	28	23	27
4	3002	1645 E ROOSEVELT ST	Arizona	1/4/2008	72	28	23	34
4	3002	1645 E ROOSEVELT ST	Arizona	1/4/2008	72	28	23	36
4	3002	1645 E ROOSEVELT ST	Arizona	1/4/2008	72	28	23	34
4	3002	1645 E ROOSEVELT ST	Arizona	1/5/2008	58	10	21	34
4	3002	1645 E ROOSEVELT ST	Arizona	1/5/2008	58	10	21	42
4	3002	1645 E ROOSEVELT ST	Arizona	1/5/2008	58	10	21	42
4	3002	1645 E ROOSEVELT ST	Arizona	1/5/2008	58	10	21	42
4	3002	1645 E ROOSEVELT ST	Arizona	1/6/2008	71	21	24	41
4	3002	1645 E ROOSEVELT ST	Arizona	1/6/2008	71	21	24	41
4	3002	1645 E ROOSEVELT ST	Arizona	1/6/2008	71	21	24	41
4	3002	1645 E ROOSEVELT ST	Arizona	1/6/2008	71	21	24	41
4	3002	1645 E ROOSEVELT ST	Arizona	1/7/2008	41	20	30	40
4	3002	1645 E ROOSEVELT ST	Arizona	1/7/2008	41	20	30	40
4	3002	1645 E ROOSEVELT ST	Arizona	1/7/2008	41	20	30	40

FIGURE 1: POLLUTION DATASET

Data Fields:

State Code: This field denotes the code for the state.

Site Number: This field denotes the site number where the AQI is recorded

Address: This field provides the address of site where the AQI is recorded

Date Local: This field denotes the date when the AQI is recorded

NO2 AQI: Air Quality Index of Nitrogen Dioxide

O3 AQI: Air Quality Index of Ground level Ozone

SO2 AQI: Air Quality Index of Sulphur Dioxide

CO AQI: Air Quality index of Carbon Monoxide

Data Preprocessing

The most important fields in the raw dataset are state, NO AQI, O3 AQI, SO2 AQI, CO AQI and date. Although all the information is provided in the dataset, it is difficult to analyze and determine any pattern because of numerous entries caused by 3 records of AQI per day. Hence, for better visualization and analysis the daily records are averaged into monthly records. For this change, the local date field is broken down into new_date, month and year. The new_date field is deleted as it is of no further use. The AQI for all four pollutants are averaged by month. By the end of this process, we have monthly AQI dataset which is more concise and easier to analyze.

To view the AQI on map, the coordinates of each state are necessary. Each state is associated with its longitude and latitude field. These coordinates are necessary for visualizing the pollutions on the map of United States. Some of the other changes were to delete State code, Site num and Address field as they were not used in visualization. Below is the screenshot of dataset after data cleaning and preprocessing.

State	Latitude	Longitude	NO2	O3	SO2	CO	Month	Year
Arizona	34.048927	-111.43122	48.22472	21.29213	9.865169	20.52809	1	2010
California	36.116203	-119.68156	34.39236	19.58565	5.806713	16.75145	1	2010
Colorado	39.059811	-105.3111	43.21739	15.82609	18.6087	16.47826	1	2010
District Of Colur	38.897438	-77.026817	41.03226	14.41935	23.41935	21.41935	1	2010
Florida	27.766279	-81.686783	29	22.67742	5.387097	14.6129	1	2010
Illinois	40.349457	-88.986137	27.88136	14.13559	59.0678	8.633333	1	2010
Kansas	38.5266	-96.726486	17.3587	23.03261	6.097826	5.086957	1	2010
Louisiana	31.169546	-91.867805	38.35484	19.64516	21.54839	11.77419	1	2010
Massachusetts	42.230171	-71.530106	43.70833	11.04167	17.04167	9.583333	1	2010
Nevada	38.313515	-117.05537	16.48387	28.22581	1.709677	1.806452	1	2010
New Jersey	40.298904	-74.521011	42.15517	13.84483	25.67241	11.81034	1	2010
New York	42.165726	-74.948051	39.12088	13.27473	35.03297	8.725275	1	2010
Pennsylvania	40.590752	-77.209755	41.22581	8	25.77419	13.22581	1	2010
Texas	31.054487	-97.563461	38.704	24.472	5.096	12.51613	1	2010
Arizona	34.048927	-111.43122	50.03614	31.37349	9.168675	18.68675	2	2010
California	36.116203	-119.68156	37.24242	27.17576	5.91129	15.87677	2	2010
Colorado	39.059811	-105.3111	44.25	23.07143	24.14286	11.67857	2	2010
District Of Colur	38.897438	-77.026817	42.64286	21.67857	16.60714	20.46429	2	2010
Florida	27.766279	-81.686783	29.71429	28.14286	5.75	6.285714	2	2010
Illinois	40.349457	-88.986137	28	22.5	45.17857	9.464286	2	2010
Kansas	38.5266	-96.726486	16.11538	30.58974	5.717949	5.525641	2	2010
Kentucky	37.66814	-84.670067	27.33333	10.33333	30.33333	3.333333	2	2010
Louisiana	31.169546	-91.867805	39.82143	29.67857	21.5	9.642857	2	2010
Massachusetts	42.230171	-71.530106	48.39286	20.21429	19.17857	8.5	2	2010
Nevada	38.313515	-117.05537	16.67857	34.67857	1.214286	2.107143	2	2010
New Jersey	40.298904	-74.521011	43.05357	22.60714	19.25	10.16071	2	2010
New York	42.165726	-74.948051	40.95302	20.67785	29.10067	8.053333	2	2010

FIGURE 2: POLLUTION DATASET AFTER DATA CLEANING AND PREPROCESSING

The dataset is still in csv format. For visualization on map, the most suitable format is to use a GeoJSON format which is a way of representing geographic data. GeoJSON consists of the following different parts:

- Geometry object: This is either the point, line, or polygon described earlier. Basically the location information.
- Feature object: This is the geometry object and the associated data. GeoJSON doesn't care what data you associate with the location information.
- FeatureCollection: Basically just a list of feature objects.

  So typically, one GeoJSON file (or dataset) will consist of a FeatureCollection containing a list of your data.

To convert CSV to GeoJson format, an online csv to geoJson converter is used. The following figure shows the GeoJSON dataset.


```
{
   "type": "FeatureCollection",
   "features": [
  {
    "type": "Feature",
    "geometry": {
       "type": "Point",
       "coordinates":  [ -111.431221,34.048927 ]
    },
    "properties": {
    "State":"Arizona",
    "NO2":48.2247191,
    "O3":21.29213483,
    "SO2":9.865168539,
    "CO":20.52808989,
    "Month":1,
    "Year":2010
    }
  },
  {
    "type": "Feature",
    "geometry": {
       "type": "Point",
       "coordinates":  [ -119.681564,36.116203 ]
    },
    "properties": {
    "State":"California",
    "NO2":34.39236111,
    "O3":19.58564815,
    "SO2":5.806712963,
    "CO":16.75144509,
    "Month":1,
    "Year":2010
    }
  },
```


FIGURE 3: GEOJSON DATASET

At the top level, a FeatureCollection is defined which is a list of data. Features is an array of Feature objects. Each Feature has a geometry object and a properties object. The geometry object is of type point and the coordinates are given. Inside the properties are the remaining fields which describe the coordinate point. User input and filtering are done using different properties.

TECHNOLOGIES

Deciding the technology to be used for implementation was challenging as there are multiple libraries available for visualization. To choose which one is suitable for this implementation, I explored various libraries including Leaflet, D3 and their compatibility with different front end frameworks including AngularJs and ReactJs. After careful consideration, I have used the following technologies for implementation.

- Leaflet: For map visualization.
- JavaScript: Enabling user input
- HTML: To create web page
- CSS: To add styles
- CSV to GeoJSON Converter: For Data preprocessing

IMPLEMENTATIONS

The main goal of the project is to build an interactive map which enables the users to input filters and view/ analyze the air pollution in United States. For this requirement a web application is built using JavaScript, HTML, CSS and Leaflet library. This web application provides an interactive map to view air quality index of the major four pollutants in each state.

Air Quality Index

An air quality index (AQI) is used by government agencies to communicate to the public how polluted the air currently is. Public health risks increase as the AQI rises. The AQI can increase due to an increase of air emissions (for example, during rush hour traffic or when there is an upwind forest fire) or from a lack of dilution of air pollutants. During a period of very poor air quality, such as an air pollution episode, when the AQI indicates that acute exposure may cause significant harm to the public health.

EPA establishes an AQI for five major air pollutants regulated by the Clean Air Act. Each of these pollutants has a national air quality standard set by EPA to protect public health:

- ground-level ozone
- particle pollution (also known as particulate matter, including PM2.5 and PM10)
- carbon monoxide

- sulfur dioxide
- nitrogen dioxide

In this project, the major four pollutants (NO2, O3, SO2, CO) are displayed. The user has the ability to select any air pollutant to view the amount of pollution caused by that particular pollutant.

Application Description

The web application uses AQI as a measure to visualize pollution in United States. The web application is divided into 3 section.

Section I: Interactive Air Pollution Map

This section displays the interactive air pollution map with filters. There are 6 filters that can be applied to the map.

1. Ground level Ozone Pollutant
2. Nitrogen Dioxide Pollutant
3. Sulphur Dixode Pollutant
4. Carbon Monoxide Pollutant
5. Month
6. Year (Ranging from 2008 to 2018)

The following figure shows the interactive air pollution map with the filters. The radius of the circle markers denotes the amount of AQI/ air pollution. A bigger circle marker means more air pollution. Using these circle markers, easy comparison can be done between the states. The color of circle marker denotes the type of pollutant causing air pollution. Green denotes O3 pollutant. Yellow denotes NO2, Red denotes SO2 and Purple denotes CO.

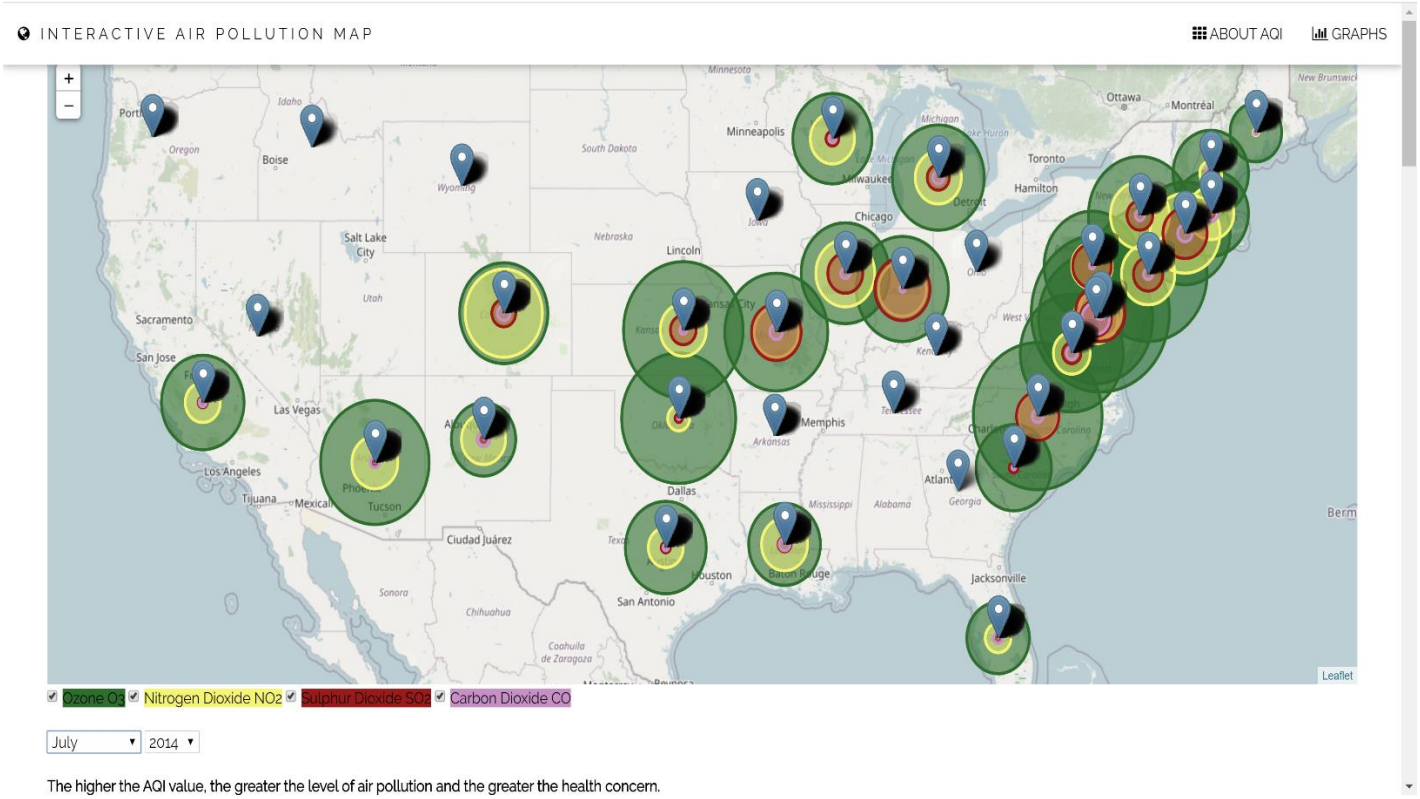


FIGURE 4: INTERACTIVE AIR POLLUTION MAP

The functions provided by the interactive map are as follows

- *Interaction/ Filters*

As mentioned above, the map provides 6 filters. These filters can be divided into two groups.

1. Pollutant selection: Consists of NO2, SO2, O3 and CO filters
2. Time selection: Consists of Month and Year Filters

Following figures are the outputs for selection of different filters.

Selection of Month and Year with all pollutants. Here the pollution is shown for March 2016

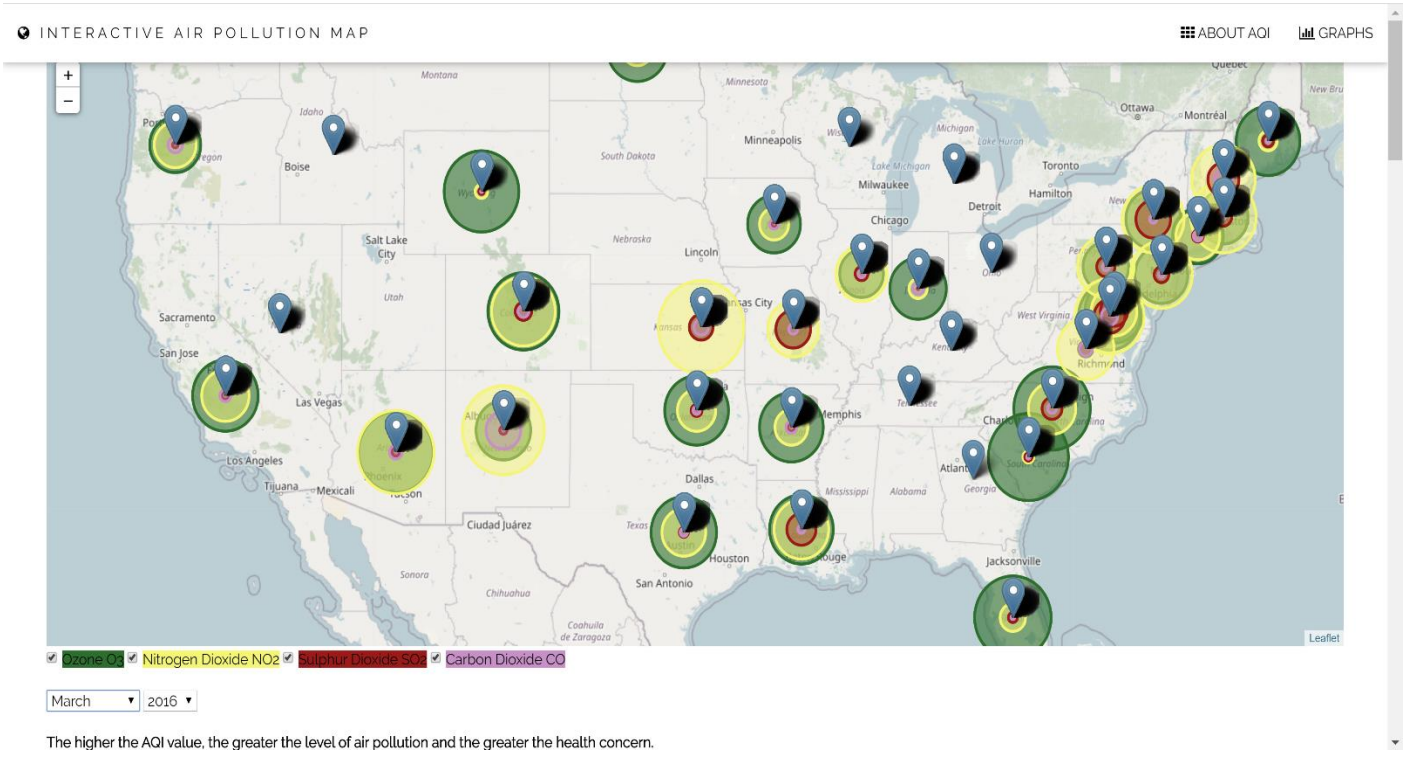


FIGURE 5: MAP WITH ALL POLLUTANTS FOR MARCH 2016

Selection of only O3 Pollutant for April 2018

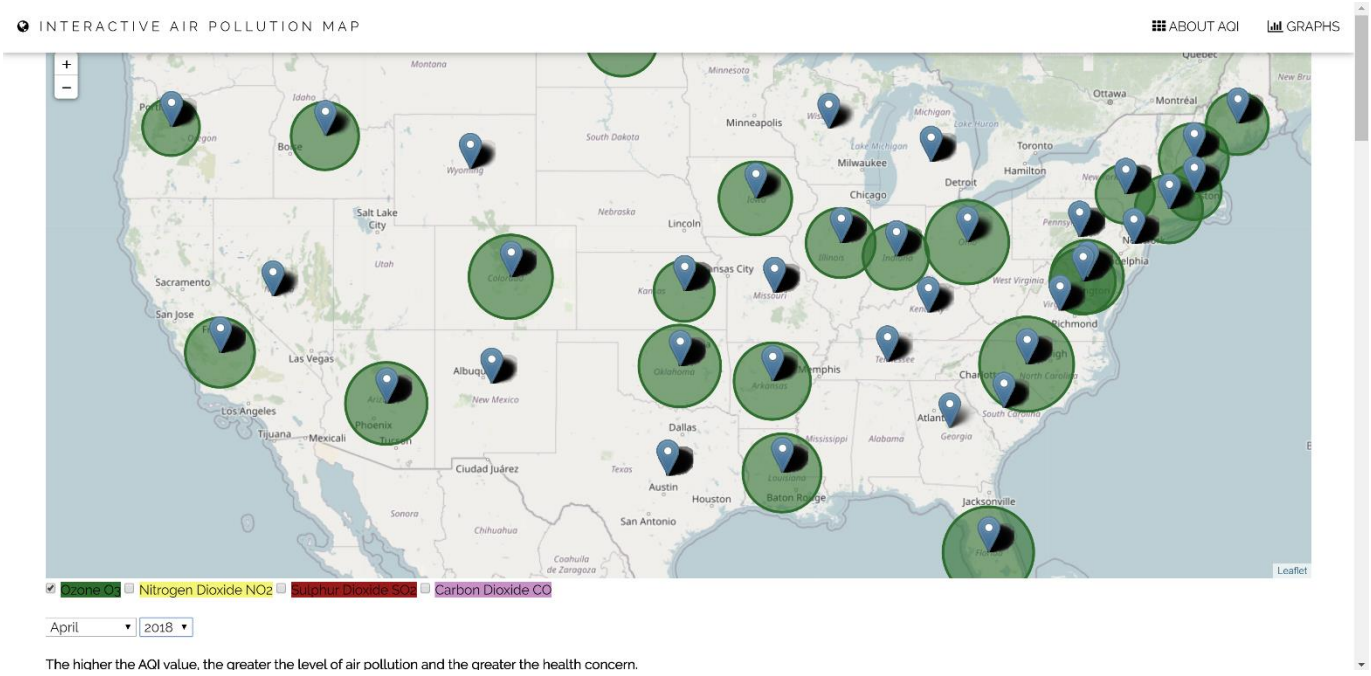


FIGURE 6: MAP WITH O3 POLLUTANT

Selection of only NO2 pollutant for March 2016

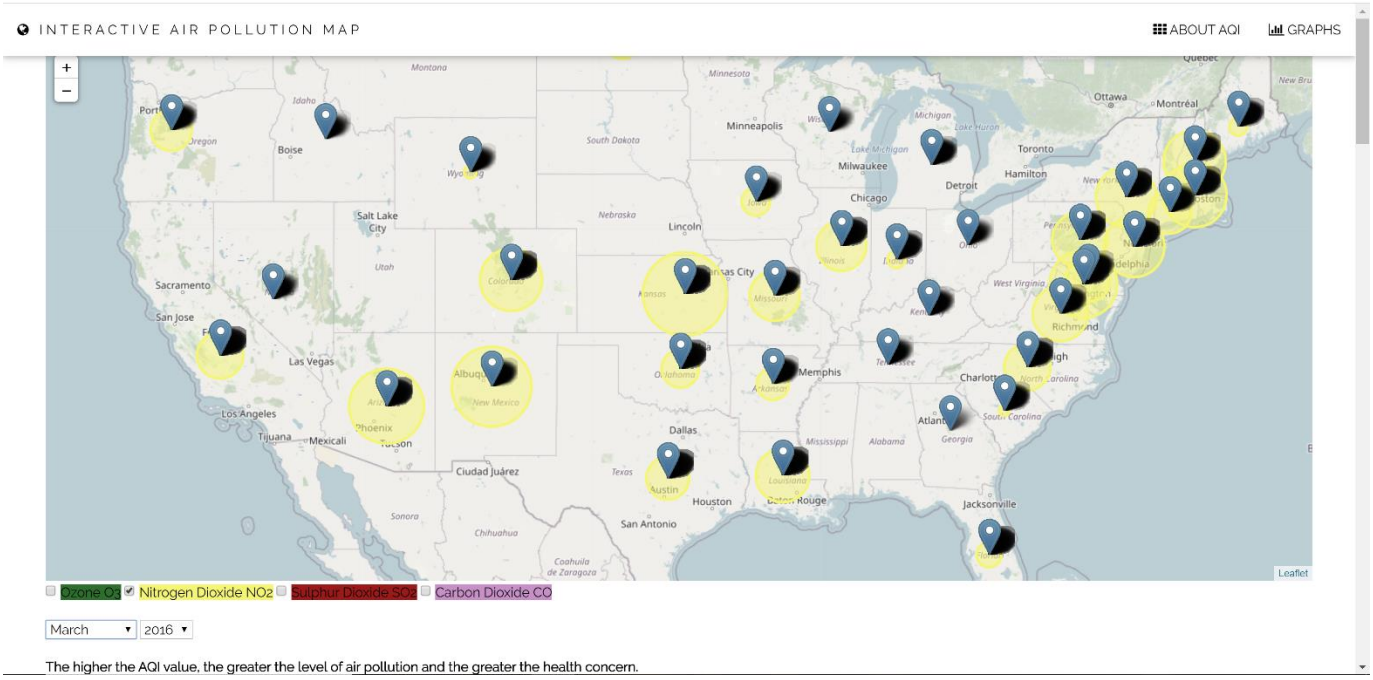


FIGURE 7: MAP WITH NO2 POLLUTANT

Selection of only SO2 pollutant for September 2008

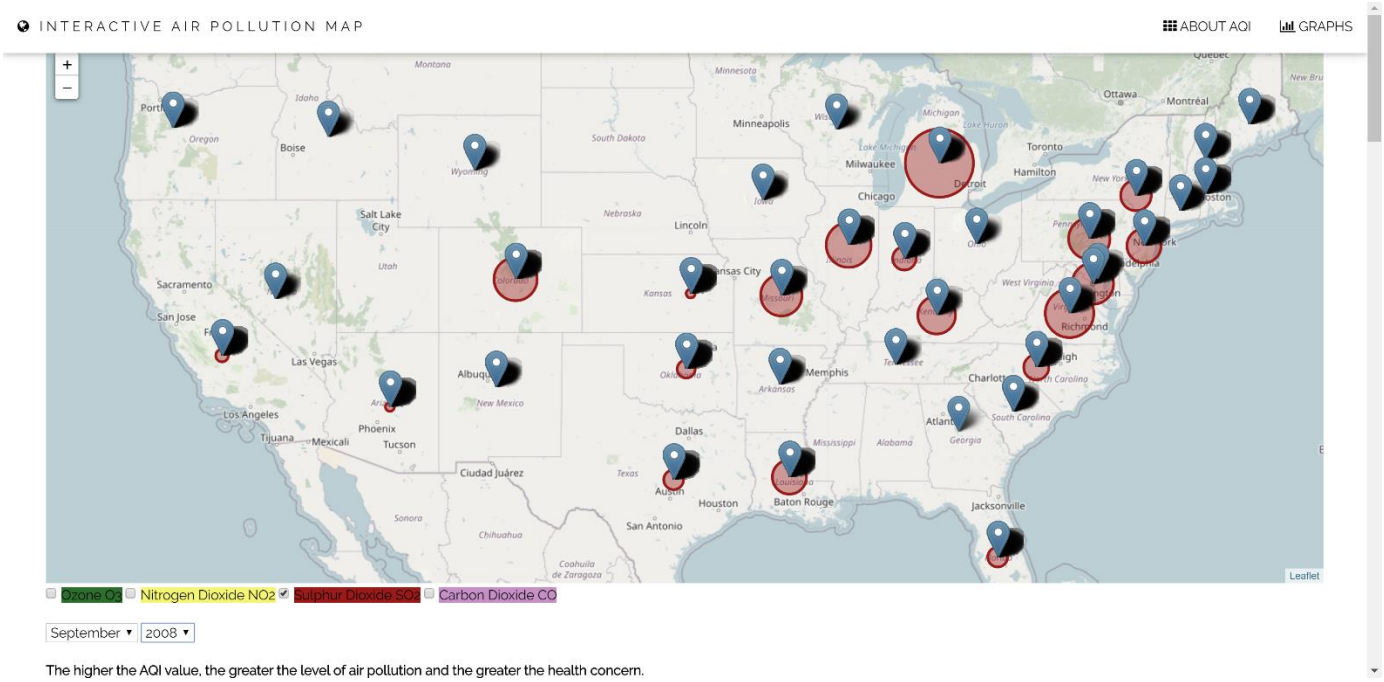


FIGURE 8: MAP WITH SO2 POLLUTANT

Selection of only CO pollutant for January 2014

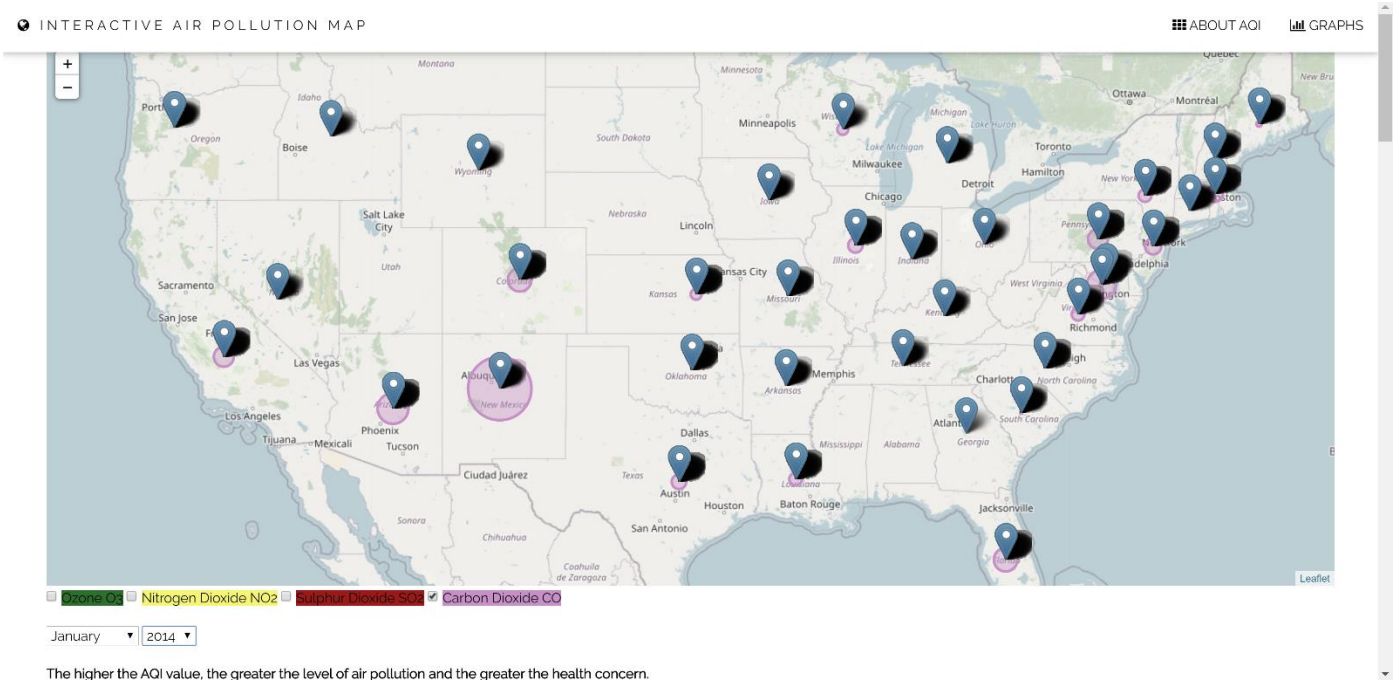


FIGURE 9: MAP WITH CO POLLUTANT

Combination of pollutant selection: O3 and NO2 for April 2015

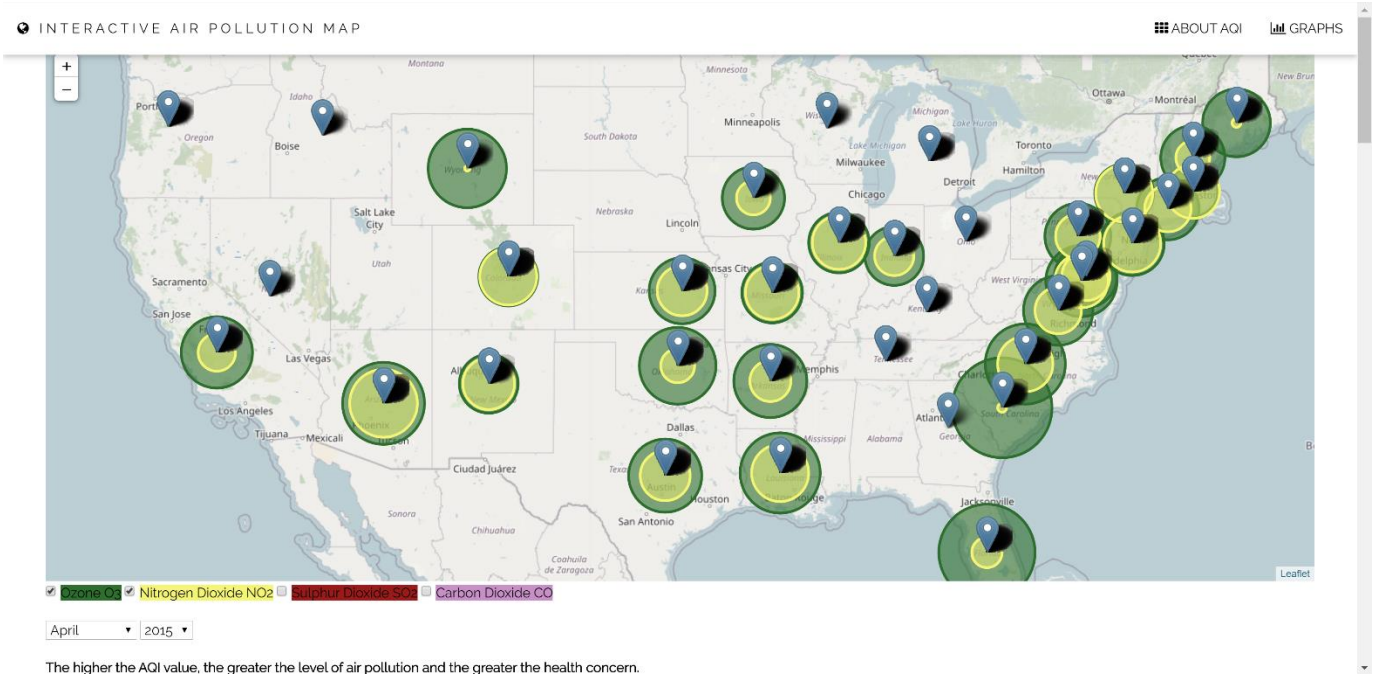


FIGURE 10: MAP WITH O3 AND NO2 POLLUTANTS

- *PopUp Information on click*

On clicking any circle marker the user can view the AQI for all pollutants. This is useful when the circle markers overlap each other. In the following figure, the overall AQI information is given for the state of Virginia after clicking the respective circle marker.

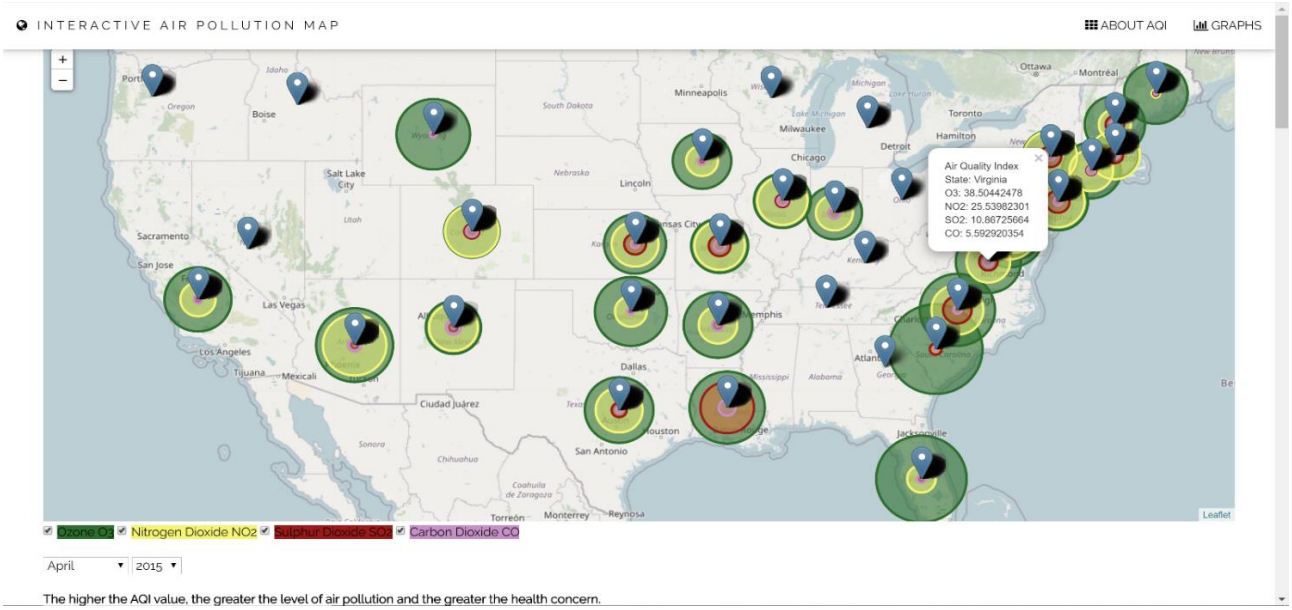


FIGURE 11: MAP WITH ON-CLICK POPUP

- *Zoom Out / Zoom In*

The user can zoom out and zoom in the map.

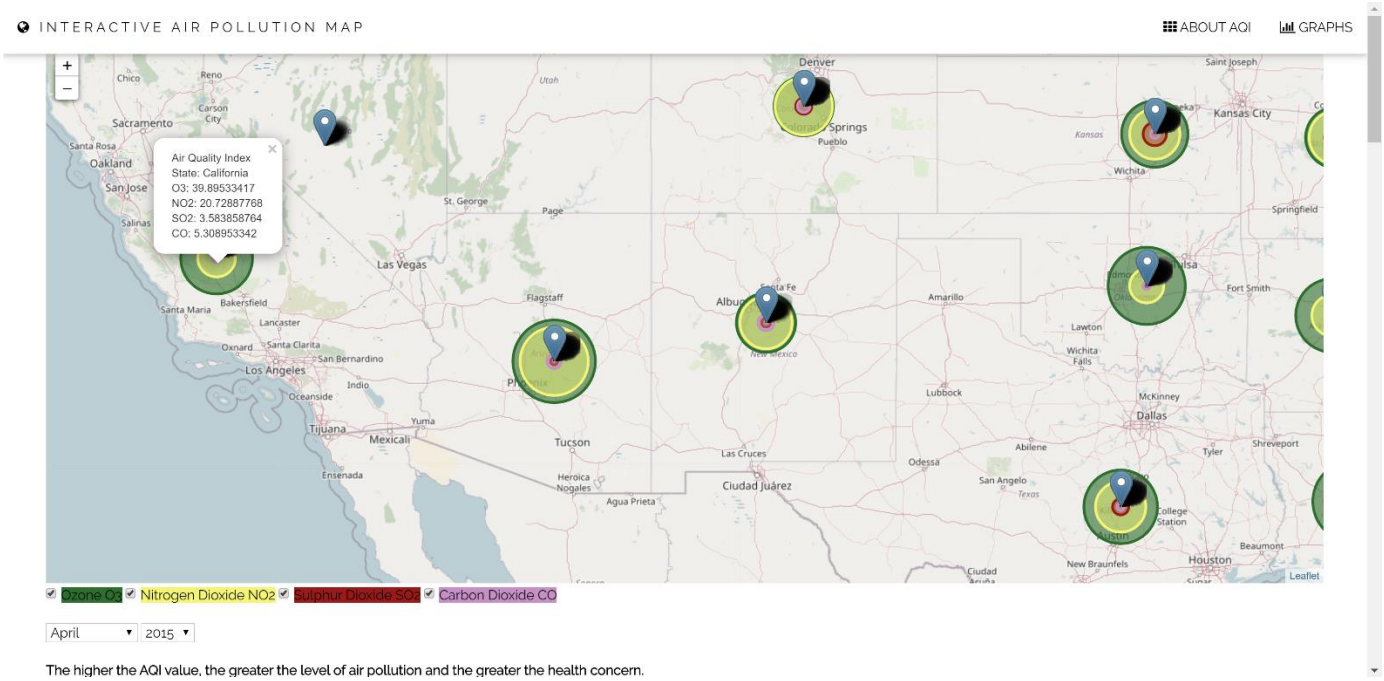


FIGURE 12: MAP ZOOM IN

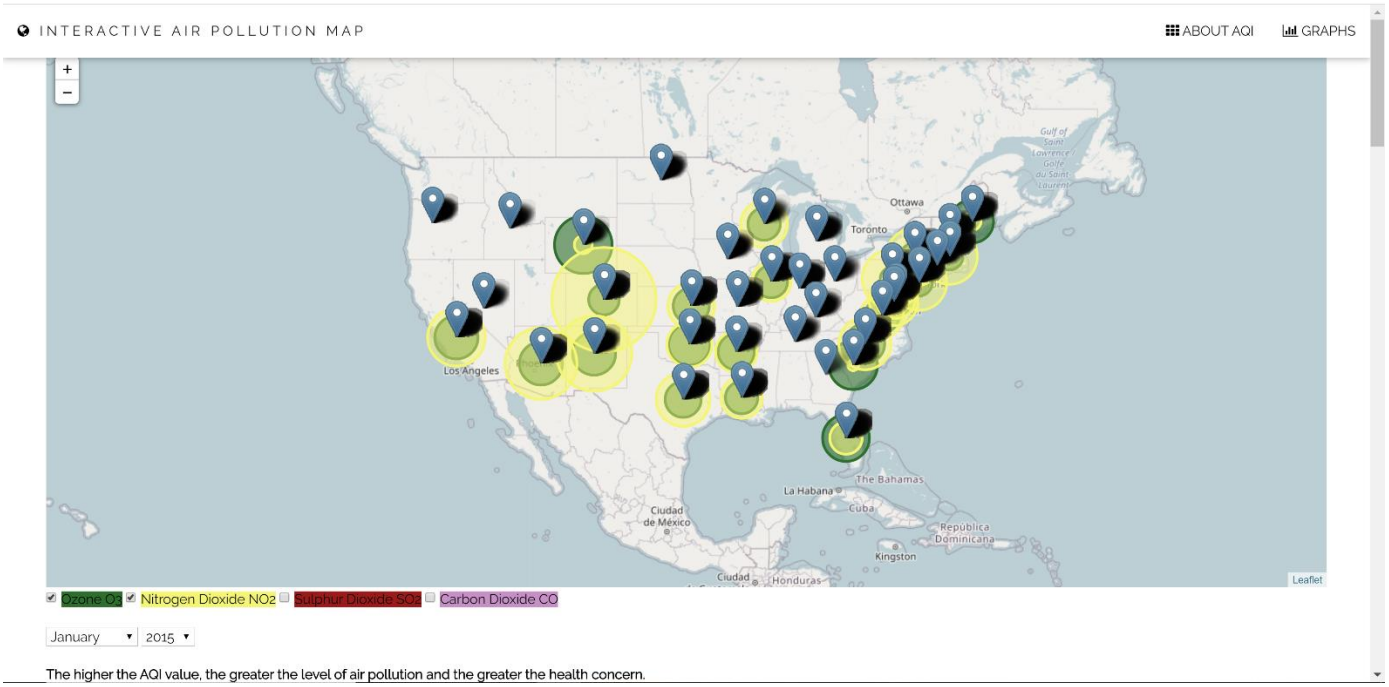
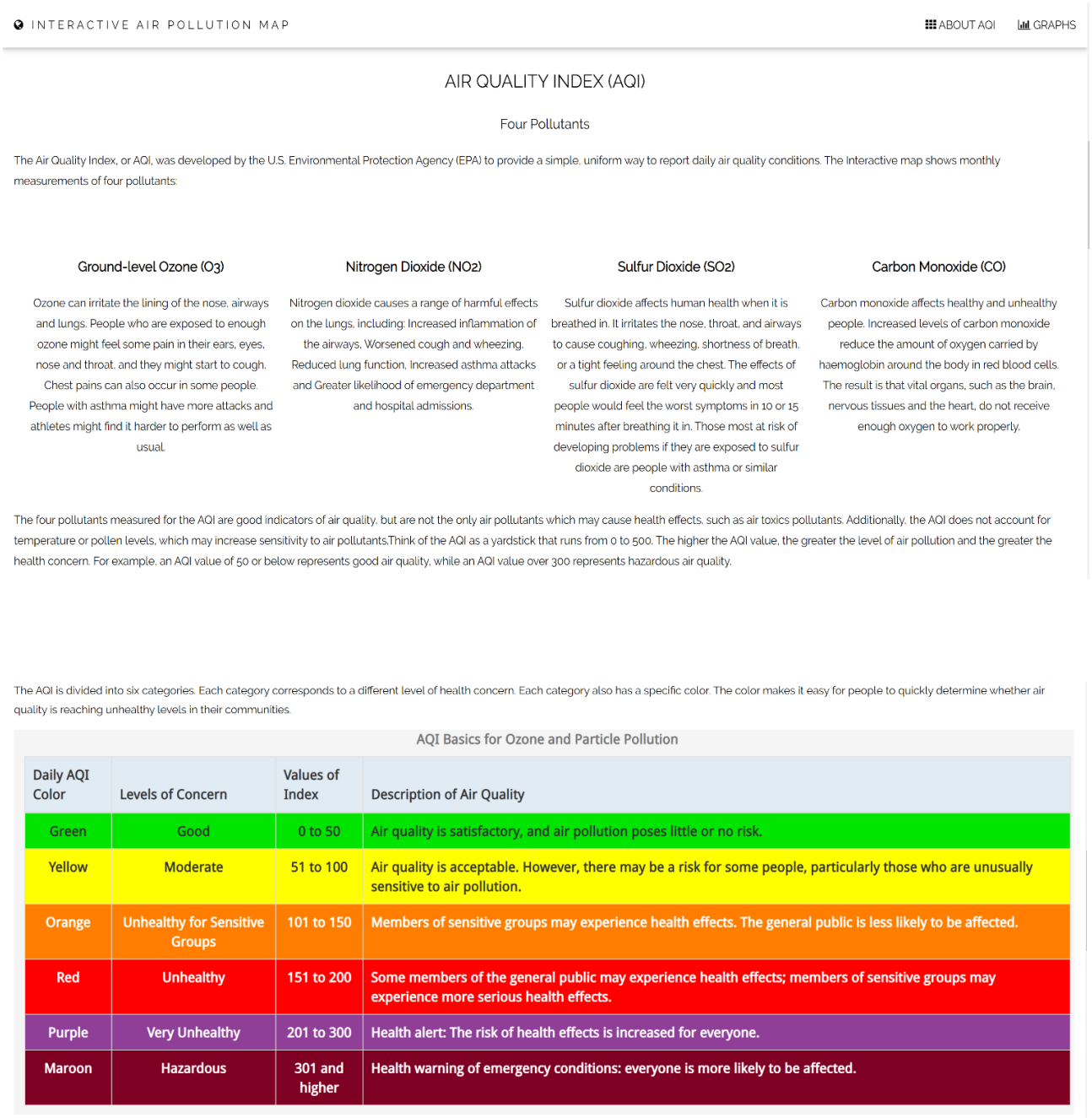


FIGURE 13: MAP ZOOM OUT

Section II: About AQI

This section is below the interactive air pollution map. It gives information about AQI and various pollutants. It also mentions the health hazards related to each air pollutant.



INTERACTIVE AIR POLLUTION MAP

ABOUT AQI GRAPHS

AIR QUALITY INDEX (AQI)

Four Pollutants

The Air Quality Index, or AQI, was developed by the U.S. Environmental Protection Agency (EPA) to provide a simple, uniform way to report daily air quality conditions. The Interactive map shows monthly measurements of four pollutants:

Ground-level Ozone (O3)

Ozone can irritate the lining of the nose, airways and lungs. People who are exposed to enough ozone might feel some pain in their ears, eyes, nose and throat, and they might start to cough. Chest pains can also occur in some people. People with asthma might have more attacks and athletes might find it harder to perform as well as usual.

Nitrogen Dioxide (NO2)

Nitrogen dioxide causes a range of harmful effects on the lungs, including: Increased inflammation of the airways, Worsened cough and wheezing, Reduced lung function, Increased asthma attacks and Greater likelihood of emergency department and hospital admissions.

Sulfur Dioxide (SO2)

Sulfur dioxide affects human health when it is breathed in. It irritates the nose, throat, and airways to cause coughing, wheezing, shortness of breath, or a tight feeling around the chest. The effects of sulfur dioxide are felt very quickly and most people would feel the worst symptoms in 10 or 15 minutes after breathing it in. Those most at risk of developing problems if they are exposed to sulfur dioxide are people with asthma or similar conditions.

Carbon Monoxide (CO)

Carbon monoxide affects healthy and unhealthy people. Increased levels of carbon monoxide reduce the amount of oxygen carried by haemoglobin around the body in red blood cells. The result is that vital organs, such as the brain, nervous tissues and the heart, do not receive enough oxygen to work properly.

The four pollutants measured for the AQI are good indicators of air quality, but are not the only air pollutants which may cause health effects, such as air toxics pollutants. Additionally, the AQI does not account for temperature or pollen levels, which may increase sensitivity to air pollutants.Think of the AQI as a yardstick that runs from 0 to 500. The higher the AQI value, the greater the level of air pollution and the greater the health concern. For example, an AQI value of 50 or below represents good air quality, while an AQI value over 300 represents hazardous air quality.

The AQI is divided into six categories. Each category corresponds to a different level of health concern. Each category also has a specific color. The color makes it easy for people to quickly determine whether air quality is reaching unhealthy levels in their communities.

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.

FIGURE 14: AQI INFORMATION SECTION

Section III: Graphs

This section is below "About AQI" section. It gives a static map which provides a bar chart visualization of AQI for each state over the period of 10 years between 2008 to 2018.

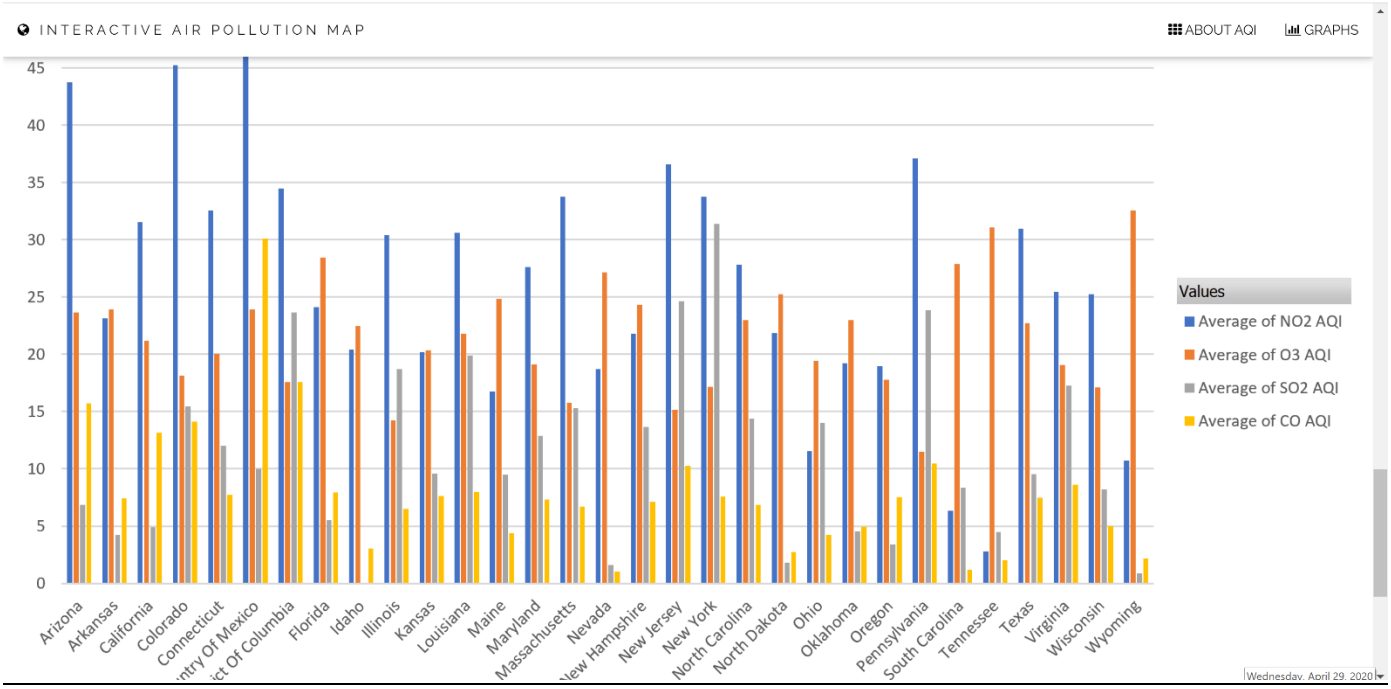


FIGURE 15: POLLUTION BAR CHART

All the sections of the web application combined provides complete information of air pollution in United States.

SUMMARY

The major components of implementation were data gathering, data preprocessing, designing frontend with user input, embedding map responding to the user inputs. The interactive map has given the users an ability to select the pollutant type and view its air quality index during a specific time period.

Using this interactive visualization, the users can easily view the difference in air quality in different states in the United States. It also shows the comparison of the major

four air pollutants. The intensities of each air pollutant can be helpful in determining the source of pollution. It can be combined be other demographics data to understand additional information about each state. This is useful to alert the public about any health risks associated with the pollutant.

FUTURE WORK

The major four pollutants visualized in the map give important information and statistics but adding additional layers and features can improve the overall analysis.

The future enhancements would be:

1. To add more pollutant filters like Particulate matter (PM10 and PM2.5).
2. Each state can be linked with demographics and industrial data to learn more about the cause of pollution in a specific state.
3. Real time data polling feature can be added to get real time air quality data and visualize it.