# The Ability of Vegetation Absorbs Vehicle Pollution in US

Garett Fogel, BoYang, Linn Xia

**ABSTRACT —** Pollution control is always one of the important and popular debate topics around the world. In recent years, air pollution has wreaked havoc around the world, often causing panic and being a major topic of discussion among the public and air pollution experts in government and academia. Therefore, it is especially important to use an effective visualization method to directly, quickly, and represent the Spatio-temporal information contained in air pollution data [1, 4]. Visual design can test hypotheses and allow users to make useful discoveries [4]. Visualization methods make data easily understood by the public and inspire or aid further research in other fields. In our data, automobile exhaust contains six pollutants: CO, SO2, NO2, O3, PM10, PM2.5. Data is available on the total amount of six pollutants emitted by cars per city in the United States, as well as the capacity of vegetation in each city to remove those six pollutants from the atmosphere. By using the D3 model to implement what our client wants to achieve, we got a multi-view for the vehicle pollution and relationships between them.

**Index Terms—** virtual worlds, virtual environments, vehicle pollutants, visualization, the United States

## Introduction and motivation

Global climate change is undeniable, in substantial portions due to human-created pollution of the environment. Automobile exhaust pollution makes up a sizable portion of this pollution. Cars account for about 40 percent of all emissions of pollutants. Therefore, it is necessary to formulate the right environmental protection policies to reduce the emission of pollutants by cars. Through various processes such as photosynthesis, plants can remove these pollutant compounds from the atmosphere. On average, more than 90 percent of the pollutants emitted by vehicles are preferentially removed by local vegetation. It has been found that the ability of vegetation to remove different pollutants is different in different areas. Therefore, not all counties necessarily need the same policies. Policies should be dependent on how much pollution the local vegetation can remove, what types of pollution the local vegetation is efficient at removing, and the amount of car pollution in the area. For example, an area may have a lot of vegetation and few cars, so there is little pollution from cars. Vegetation in some areas may be poor at removing SO2 but good at removing other pollutants, so it would be rational to promote gasoline with lower amounts of sulfur. For areas where vehicle pollution exceeds vegetation removal capacity, car travel should be limited, and public transportation should be promoted. Then, we want to apply the dataset to create a visualization that can help the government create policy and help to decrease the pollutants from vehicles.


- Garett Fogel: fogel.45@osu.edu, OHIO STATE UNIVERSITY
- Bo Yang: yang.6113@osu.edu, OHIO STATE UNIVERSITY
- Linn Xia: xia.537@osu.edu, OHIO STATE UNIVERSITY


date of current version 12 09. 2021.

## 1 Related work

Our goal for the project is to create a tool that will allow users to interactively view and understand the pollution data. Also, our client wants the project to allow readers to analyze vehicle pollution in each county and allow them to utilize the tool to make environmental policies tailored to each county.

### 1.1 Research questions

We have mainly three research questions as follows.

1. Where is the amount of vehicle pollution greater than the ability of local vegetation to remove it?

2. Is the removal efficiency of different pollutants by vegetation related to latitude and longitude?

3. What kind of environmental protection policies should be implemented in different regions? Such as restricting car travel, imposing different taxes on cars with different engines and promoting different types of petrol.

### 1.2 Hypothesis

We have few hypotheses which would help us conduct our product. The types of vegetation will vary depending on geographical location, and therefore the ability of local vegetation to remove pollutants will also depend on geographic location. Different regions' vegetation will have a different makeup of what pollutants they are good or bad at removing, and therefore should promote different fuels. In general, areas with a lot of vegetation such as the forests of Maine and Florida everglades are likely to have a higher threshold for pollution removal. Areas with sparse vegetation such as the plains of North Dakota or the deserts of Texas are likely to have a lower threshold for pollution removal. Areas with highly populated cities such as New York City or Los

Angeles are likely to have more vehicle pollution. Rural areas with lower populations are likely to have lower vehicle pollution.

### 1.3 Implementation and environment

We intend to use Python to pre-process the data and create a REST API. Pandas and NumPy will be used to manipulate the data into a shape that is useable by our regression algorithm as well as for our front end visualizations. Sklearn will be used to run a regression on our dataset. Flask will be used to create our REST API to share this data with our front end in a way that can be easily used to generate our visualizations. Our front end will be HTML, CSS, and JavaScript. jQuery AJAX will be used to make requests to our REST API. D3 and Mapbox will be used to create our visualizations. Bootstrap will be used to make our web pages look clean without spending loads of time writing our own CSS and JavaScript. For hardware, we will be able to run this locally on our laptops by simply running the API script and starting a HTTP server for the front end via the python command. If we decide to publish this, we will need to rent a server such as AWS LightSail or EC2. After setting up the environment we need, we implemented the code and get the ideal result we want. Our team in the end use D3.js as our coding tool to implement our visualization designs with interactions and output a web page for the users.

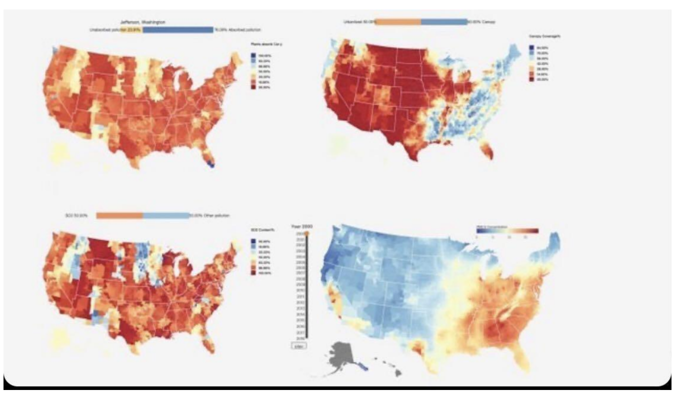

Fig. 1. Overview of our visualization

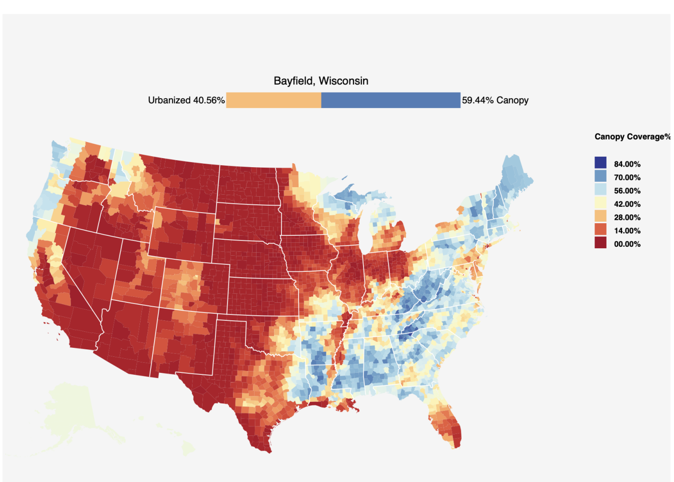


Fig. 2. Canopy coverage for each county

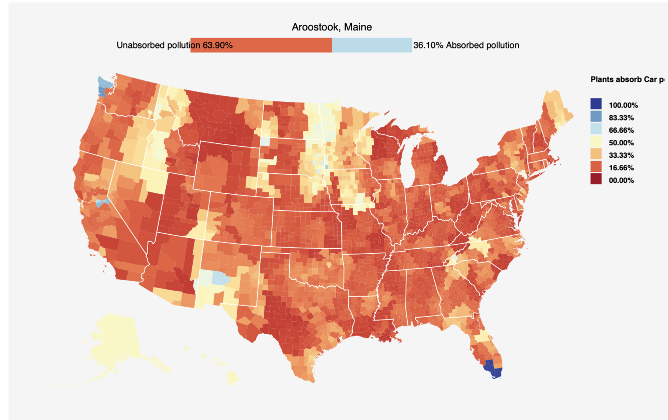


Fig. 3. Plants absorb for each county

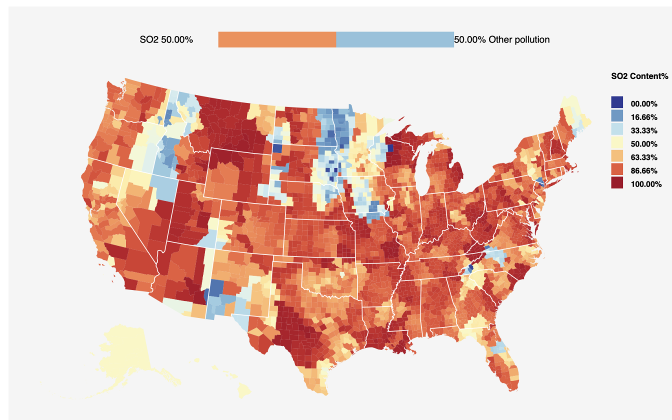


Fig. 4. SO2 contents for each county

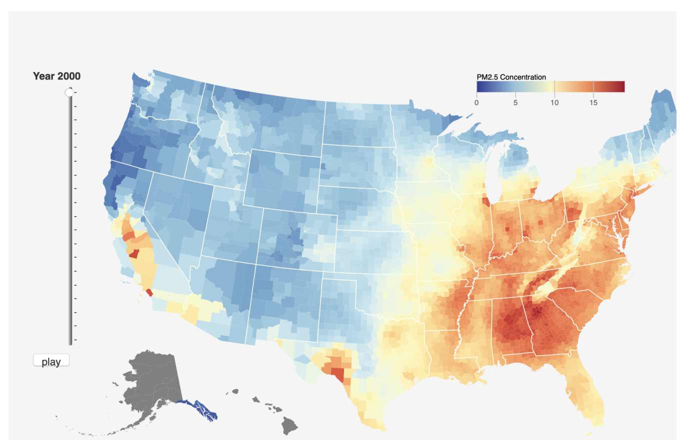


Fig. 5. PM2.5 concentration map

We chose D3 rather than other libraries because D3 has excellent control ability to view results, and it provided multiple ways of visualizing data quickly. Another important reason was that it had a huge community backing it, so it was easy to find resources to support our design implementation and our web page. Users can easily interact with the visualization we provided by doing this method. They can simply have a mouse on the value to read the data.

### 1.4 Exploratory study

In Fig.1. Users can directly get an overview of what they can do on this website. We spend plenty of time combining all four visualizations on the same page, which is worth it for us to do that. Also, we think our visualization tool can be used for much professional research because we have a holistic dataset and a clear visualization corresponding to these most important values in the dataset. To be more specific, during the process of dealing with data, we found there are mainly four pollutants emitted from vehicles. All counties can absorb 100% of CO and NO2, 98% of PM2.5. But only 8% of SO2 is absorbed. In Fig.4., there is an important visualization that mainly shows SO2 contents for each county. And PM2.5' s content for each county shows in Fig.5. that we keep this value because the absorb rate for PM2.5 cannot be a hundred percent.

## 2 Design method

### 2.1 Data abstraction

Our data is directly from our client, so we do not need to spend any time on collecting our data. But we spend some time cleaning data, because the dataset given was huge and contains a lot of information that we do not need to use. Within our data, it shows amounts emitted and absorbed for each pollutant for counties in the United States. It is also giving canopy coverage for each county. We plan to present this data in a way that is useful for policy makers and environmental scientists to leverage while creating rational environmental protection policies around cars specific to their location and the surrounding vegetation.

### 2.2 Task abstraction

We plan to use the regression algorithm of machine learning to perform regression fitting merging of the data and feature chi-square analysis to verify the correlation of variables. We will use scatter plots to display these correlations (or lack thereof).

We also plan to create a map that shows how much vehicle pollution there is, how much pollution the vegetation can remove and the difference between the two. We plan to make this map interactive so that users can choose which data they see and filter to show only certain pollutants. Users will also be able to zoom in and out and move around the map to get a better view of the areas that interest them.

### 2.3 Design considerations

A method that can enhance the reader's understanding and the memory of the data is crucial for pollutant data visualization [6]. In human cognition, understanding and memory are intertwined. The key to pollutant data visualization is to enhance readers' memory of image information due to the variety and a large amount of pollutant data. Using Amazon's Mechanical Turk, the authors obtained general metrics for quantifying the effectiveness of information memorization, which are critical in determining how to design effective visualizations.

Another way [2, 3] to optimize the visualization of pollutant data is to incorporate interactive design into the visualization. This work [3] implements a Web-based geographic information system (Web GIS) for Marine pollution monitoring and forecasting. This system allows users to view simulation results in a geographic environment. It provides end-users with comprehensive and integrated spatial and temporal environment information through a remotely customized user-friendly graphical interface. This work [2] presents a tool that supports smooth and flexible use visualization which can be used for interactive dynamic visual analysis.

The focus of the work mentioned above focuses on the visualization of pollutant data. With our work, we hope not only to effectively visualize pollutant data but also to convey effective information from the visualization to policymakers. In this paper [4], a prediction system based on pollutant visualization and deep learning are proposed to predict future pollution. The entire system is contained in a fast, easy-to-use Web service and a client that visually renders the system response. The system was built and tested based on data from the city of Skopje. The goal of visualization systems is to facilitate data-driven insight discovery, but what if these insights are false? Features or patterns in visualization can be treated as relevant insights, even though they may come from noise. Zgraggen et al proposed, a method to evaluate MCP in visualization tools [5]. The paper shows how a confirmatory analysis method can account for all visual comparisons, insightful and non-insightful. Based on this research, we choose to use four maps to present our data in this way, which we regard can let users easier to memory and understand.

## 3 Implementation method

We used Jupyter Notebook in our data processing. Compared to other ideas, Jupyter Notebook allows us to directly observe the processing data at each step of the data processing stage. We used PANDAS to edit the data because pandas' function helps us quickly sort through the data. In the data processing stage, we observed a lot of interesting phenomena. For example, the absorption rate of each pollutant by vegetation in different areas was the same, which helped us adjust the visual design in time. The project mainly uses JavaScript, CSS, Html three languages. Visualization relies on JavaScript D3, which is a very powerful resource library. We input data in JSON files and CSV files. Json files contain map information, and CSV files contain environment information. We used different VERSIONS of D3, versions 4 and 5 respectively. The method of providing data is d3. queue (), which is only supported in version 4 D3. Visualizations can be viewed in various browsers (e.g., Safari, Chrome, Firefox, etc.)

## 4 Result and analysis of results

After generating our visualizations, we found that SO2 is the least absorbed pollutant in automobile pollution. The plot of SO2 content corresponds to the plot of plant absorption of pollution. Also, by comparing the vegetation coverage map and PM2.5 map, it is found that the areas with the highest vegetation coverage are also the areas that produce the most pollution. The United States has researched a balance between pollution and treatment.

## 5 Discussion of results

Our design is driven by a dynamic United States map to help improve the environment, which can help policymakers to see how pollutants emitted by vehicles changed by years on some specific chemical materials. Our aim is to reduce complex work and time effort on year-related studies. The advantage of our design is that it draws people's attention as they look at it and see how three different rates change by years and connect between these values.

## 6 Conclusions and future work

To sum up, we reached our goal to make a tool that can show the pollutants of vehicles for each county. And, our visualizations can assist our users see the change directly. The future work we can do is to make our tool more memorable and more professional, which means to add more values in, such as from month to month, from day to day, and so on. Besides, we can import a button that users can directly import the environment database they have and our website can directly generate the visualization based on that. We want to do this function because some governments may have some environment dataset securely and do not want to share with the public. By using our tool, they can generate visualizations by themselves, which can help them make some policies within the specific area.

## References

[1] Li H, Fan H, Mao F. A visualization approach to air pollution data exploration—a case study of air quality index (PM2. 5) in Beijing, China[J]. Atmosphere, 2016, 7(3): 35.

[2] Heer J, Shneiderman B. Interactive dynamics for visual analysis[J]. Communications of the ACM, 2012, 55(4): 45-54.

[3] Kulawiak M, Prospathopoulos A, Perivoliotis L, et al. Interactive visualization of marine pollution monitoring and forecasting data via a Web-based GIS[J]. Computers & Geosciences, 2010, 36(8): 1069-1080.

[4] Korunoski M, Stojkoska B R, Trivodaliev K. Internet of things solution for intelligent air pollution prediction and visualization[C]//IEEE EUROCON 2019-18th International Conference on Smart Technologies. IEEE, 2019: 1-6.

[5] Zgraggen E, Zhao Z, Zeleznik R, et al. Investigating the effect of the multiple comparisons problem in visual analysis[C]//Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018: 1-12.

[6] Borkin M A, Vo A A, Bylinskii Z, et al. What makes a visualization memorable?[J]. IEEE transactions on visualization and computer graphics, 2013, 19(12): 2306-2315.