

## Abstract

Air pollution has negative impact on human health and leads to many chronic diseases. U.S. Environmental Protection Agency (EPA) has been closely monitoring the air pollution using its ground stations in various locations around the nation. The collected data has been included in its air pollution database and made publically available in its website. The detailed daily air pollutant concentrations (e.g. PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, CO, Pb, NO<sub>2</sub>, Ozone) can be downloaded in Excel format. In this poster, we visualize and analyze the air pollution in the US East Coast in the past years using Tableau software. Such visualization allows us to observe the trend of air pollution and its transmission pattern in major cities of the east coast. The correlations between air pollution and various conditions (e.g. traffic, season, location) are discussed. The influence of various terrain conditions to the PM<sub>2.5</sub> pollutant diffusion is explored. The visualization and analysis of air pollution data helps better understand its mechanism and distribution pattern.

## Introduction

Air pollution has negative impact on human health, leading to increased respiratory diseases, including lung cancers. Air pollution resulted diseases cause significant healthcare cost and shorten human life expectancy. United States Environmental Protection Agency (EPA) closely monitors the air pollution in different states around the country. Such data is included in their air pollution database and made publically available in their website [1]. The EPA has accumulated huge amount of air pollution monitoring data. However, how to effectively analyze these data and find the trends of air pollution as well as its correlation with various environmental/temporal/geographical factors remain very challenging. A visual methodology is needed for efficient and reliable exploration, particularly in the case of air pollution data, to improve the depth, readability, and accuracy of data analysis. In this poster, we proposed the visualization exploration method that realizes the process of observation – hypothesis – verification. This method was tested and validated in a year-long case study of the air quality index (AQI) of different air pollution and daily PM<sub>2.5</sub> concentration in US east coast. The useful findings of air pollution in US east coast based on the visualization and analysis of EPA air pollution data could be used for various applications, such as knowledge-based decision making and finding ways to effectively reduce air pollution.

## Method and Data

Workflow for visualizing air pollution data is shown in Figure 1 and consists of three blocks, namely, Database, Hypotheses and Visualization. We get the data from U.S. Environmental Protection Agency (EPA) website [1]. Then we use software to combine those CSV of different states together. Finally we use Tableau software to visualize those data. Data are checked and analyzed by the Visualization block through various basic visual methods, such as bar charts, heat tables, heat map, and line charts. Based on these process, hypotheses can be verified and we can analysis the data to draw conclusions on the trend and correlation of air pollution to various factors.

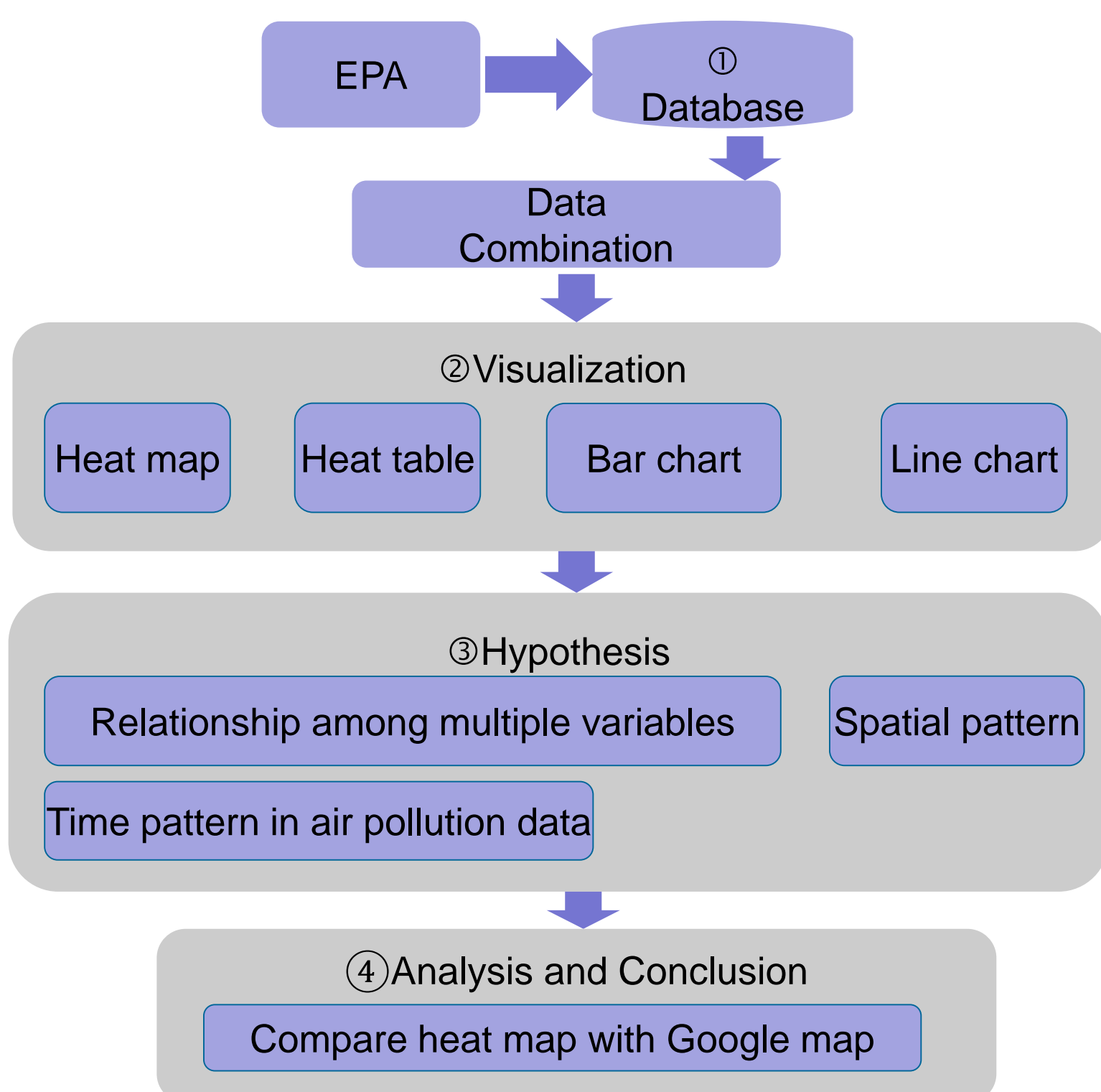


Fig 1. Visualization workflow for air pollution data

Data accessing and storage are displayed as Step (1) in Figure 1. We download the air pollution data from EPA website [1]. Air pollution data often contain time and geographic location and daily concentration of air pollution. We focus on the daily concentration of PM<sub>2.5</sub>. These data are in the same format, so we can combine them together and directly upload them to Tableau. Visualization is shown in Step (2), which is used to analysis the data. The main processes in this block include visualization design, data preprocessing, and visual analysis to achieve visualization flexibility with various methods, such as heat maps, heat chart, bar chart and line chart. The next block shows the Hypotheses. We assume there exists relationship between air pollution and multiple variables, such as time and location. In step (4), we analyze the visualized data and compare the heat map with Google population density map and Google plant distribution map to get our observation and conclusions.

## Visualization and Analysis

Based on the EPA data, we analyzed the daily average PM<sub>2.5</sub> concentration in Connecticut in 2016. The heat chart is shown in Table 1. From the results in Table 1, it indicates that the mean value of PM<sub>2.5</sub> concentration in weekdays are generally higher than that in weekends. Based on the result, we can get some possible hypotheses that the higher PM<sub>2.5</sub> air pollution in weekdays compared to weekend is due to the increased traffic for people going to/returning from work. People go out to work on weekdays and the vehicles they drive would produce more PM<sub>2.5</sub> air pollution.

Table 1. Heat table of week and week mean PM<sub>2.5</sub> concentration  
2016 WEEKDAY PM<sub>2.5</sub> IN CONNECTICUT

Month of D.	Sunday	Monday	Tuesday	Wednes.	Thursday	Friday	Saturday
January	7.177	6.242	6.871	6.734	11.693	8.037	6.250
February	9.434	9.959	6.186	5.796	3.708	4.959	7.386
March	5.836	9.025	5.409	7.839	8.060	5.540	3.471
April	3.943	6.949	5.507	4.539	5.903	6.250	5.049
May	5.114	4.772	5.191	6.616	8.854	8.285	7.918
June	5.854	5.494	4.906	5.455	7.406	6.150	8.614
July	7.093	6.278	4.539	6.113	8.943	9.598	6.649
August	4.502	4.060	6.447	6.620	7.108	7.403	7.352
September	6.835	4.551	4.055	5.964	6.396	6.924	6.209
October	6.900	8.060	10.403	8.448	7.249	7.077	6.814
November	5.500	12.689	9.162	8.576	10.923	9.137	7.757
December	8.794	7.174	8.663	11.402	8.544	5.615	8.810

We plotted the monthly average of PM<sub>2.5</sub> concentration in Connecticut during 2011-2014, the curves are shown in Figure 2. We see that the higher value of PM<sub>2.5</sub> concentration is in winter months (e.g. February) and the PM<sub>2.5</sub> concentration decreases with years. Based on these results, we can make some possible hypothesis that this is because people burn firewood or oil for heating in the cold months of winter and those heating process would produce more PM<sub>2.5</sub>.

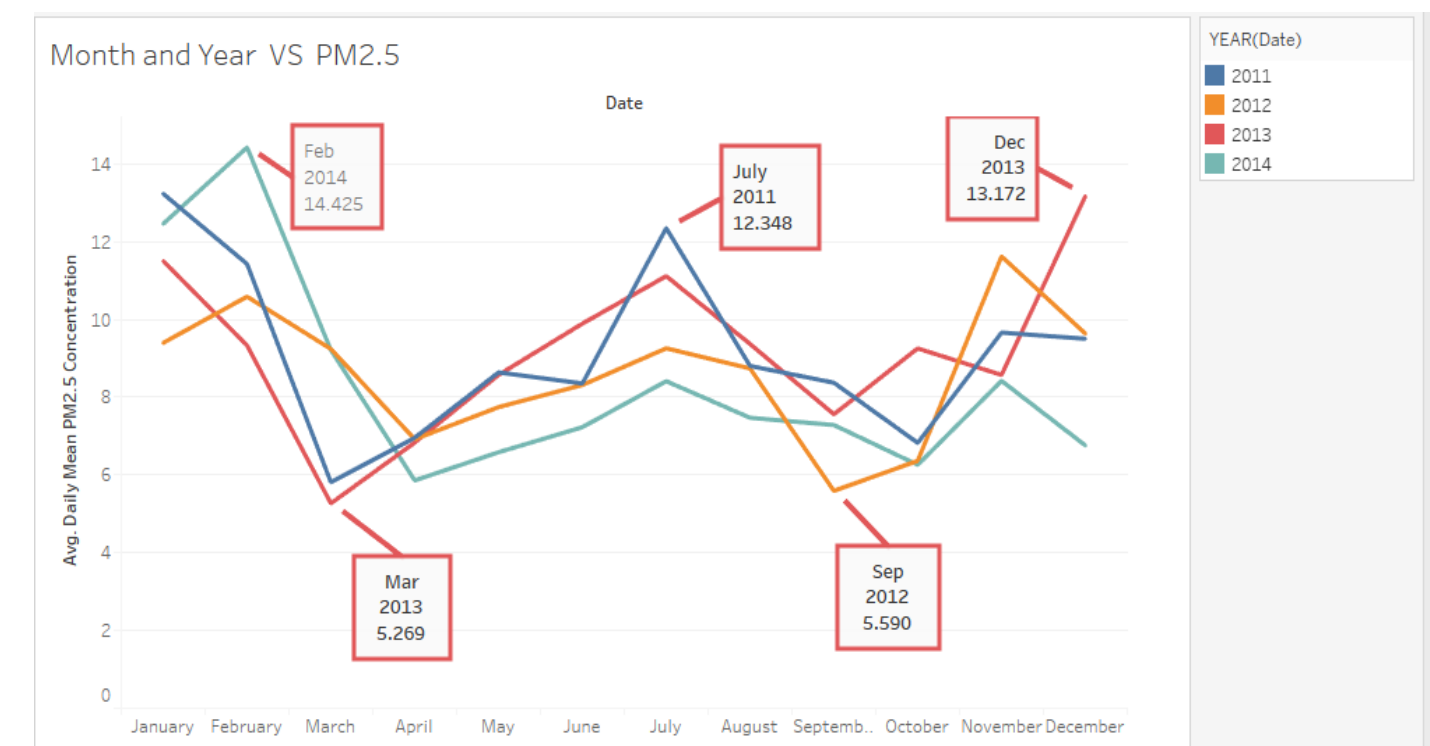


Figure 2. Monthly mean values of PM<sub>2.5</sub> concentration in Connecticut (2011~2014) We also analyzed the percentage of different ingredients in air pollutants during 2012~2106. The results are shown as pie charts in Figure 3.

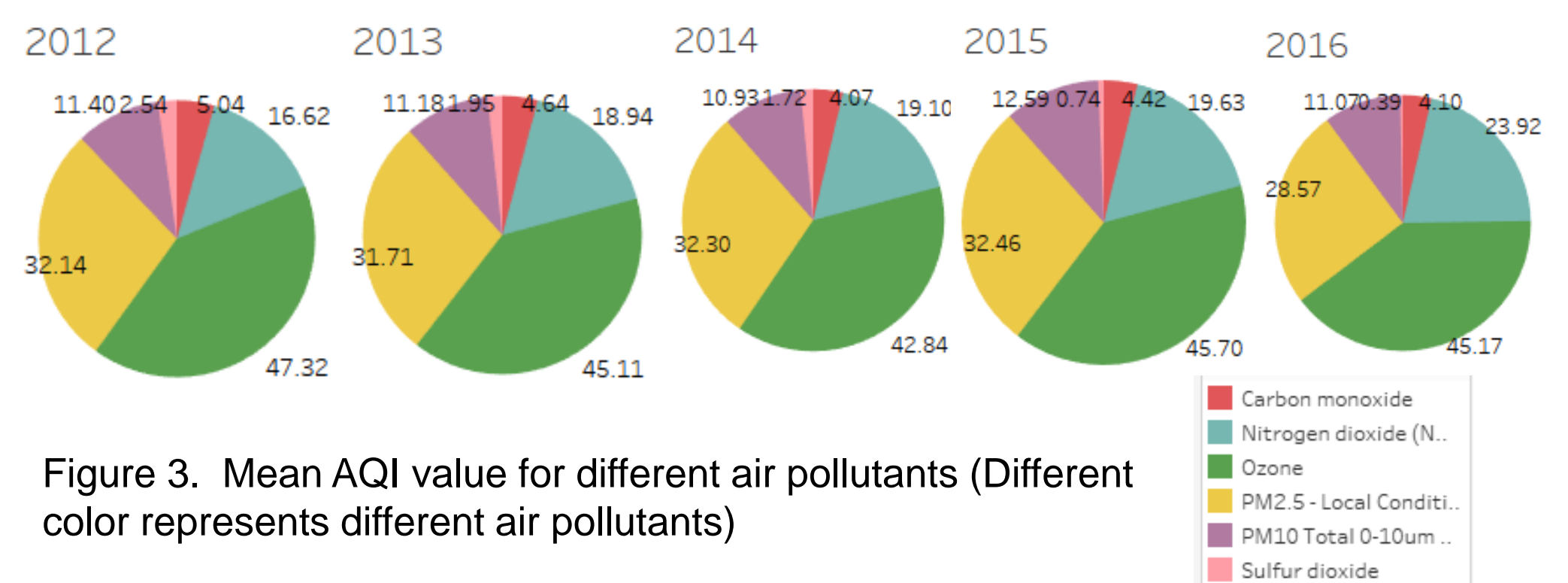


Figure 3. Mean AQI value for different air pollutants (Different color represents different air pollutants)

Based on the result, we found that the major air pollution is Ozone, PM<sub>2.5</sub>, NO<sub>2</sub> and PM<sub>10</sub>. The mean AQI value of SO<sub>2</sub> and PM<sub>2.5</sub> decrease with years. The mean AQI value of NO<sub>2</sub> increase with years. NO<sub>2</sub> comes mainly from automobile exhaust, so it indicates number of cars increases with years. As a poisonous pollutant, SO<sub>2</sub> has decreased in recent years, which is a good sign.

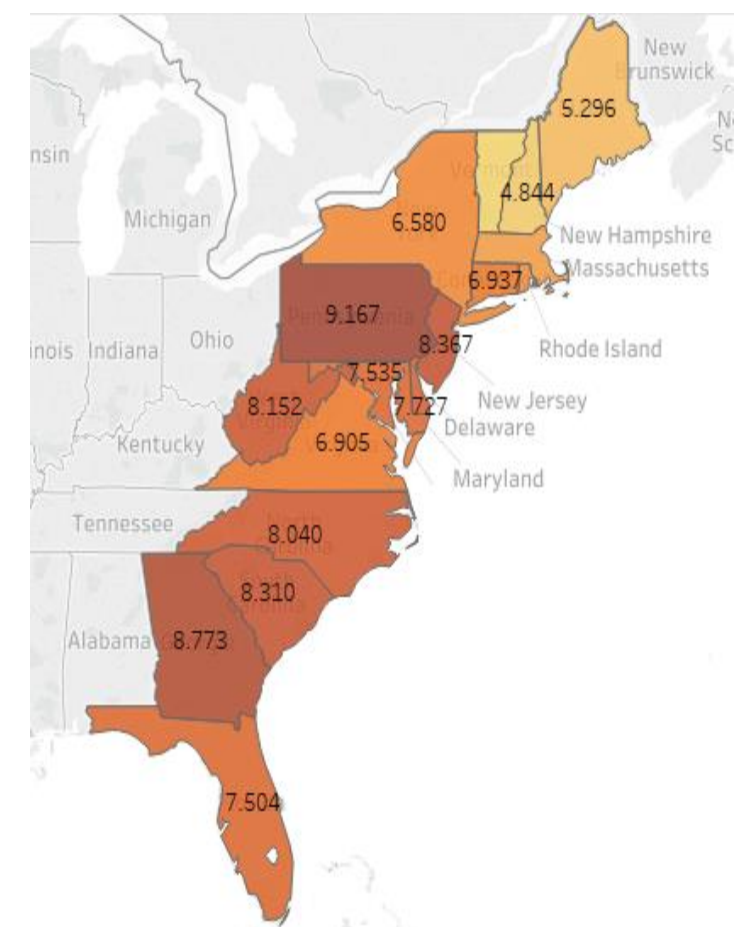


Figure 4. Daily mean PM<sub>2.5</sub> concentration in US east coast in 2016



Figure 5. Google map of plant distribution in US east coast

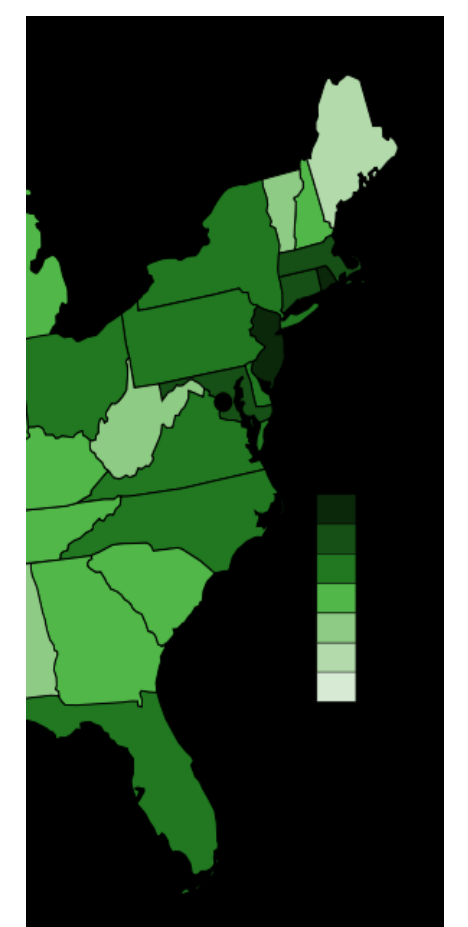


Figure 6. Google map of population density

The heat map shown in Figure 4 indicates the daily mean PM<sub>2.5</sub> concentration in US east coast in 2016. Based on the visualization, we can see the value of PM<sub>2.5</sub> concentration for majority of coastal states is lower than those inland cities. For those states near the ocean, the PM<sub>2.5</sub> can diffuse quickly due to the wind from the ocean. Furthermore, the shoreline cities are close to the open water, which produces less air pollutants due to human activities. Thus the daily mean value of PM<sub>2.5</sub> concentration is lower for those coastal states. We download the google map of plant distribution in US east coast and the population density in US east coast. By comparing those two maps, it shows that the coastal cities have better plant coverage. This also helps reduce the air pollution because green plants can absorb air pollutants and purify air. Population density may also have influence on the value of PM<sub>2.5</sub> concentration. Both Florida and Maine are surrounded by sea, but PM<sub>2.5</sub> concentration for Maine is lower than Florida, partially because population density is lower in Maine. PM<sub>2.5</sub> pollution is very complex and may be affected by many factors. The visualization of the air pollution data helps to uncover the potential influence factors for PM<sub>2.5</sub> air pollution.

## Conclusions and Future Work

In this poster, visualization analysis was performed on the PM<sub>2.5</sub> air pollution data collected by EPA for US east coast cities. This enables rapid processing and multi-perspective exploration of air pollution data to reveal potential spatio-temporal correlation patterns. It helps find the trend and distribution pattern of PM<sub>2.5</sub> pollution, so that a better understanding about the air pollution mechanism can be achieved. This will be helpful in finding ways to reduce air pollution for better human health.

## References

[1]. United States Environmental Protection Agency, Air Data: Air Quality Data Collected at Outdoor Monitors Across the US, URL: <https://www.epa.gov/outdoor-air-quality-data>.