

Project Title

Now You See Me

Link: <https://github.com/ar-ambuj23/dataviz-project2019>

Members

Ambuj Arora
UID: u1265867
Email: ambuj.arora@utah.edu

Shaurya Sahai
UID: u1266148
Email: u1266148@utah.edu

Sushmitha S N
UID: u1265043
Email: sushmitha.sn@utah.edu

Background and Motivation

Our motivation for the project comes from the fact that we wanted to build something which could be used by the general public and/or experts to analyse trends in something that affects all of us.

Air pollution, as most of us are aware, is one of the most serious problems in this age and time. It refers to the contamination of the atmosphere by toxic chemicals or organic materials. Polluted air has an adverse effect on the ecological system. It's important to study the statistics of air pollution because it shows how the quality of air is changing over time. Generally, the statistics reflect the levels of different pollutants such as ozone, nitrogen dioxide, sulfur dioxide, carbon monoxide, etc. There is no denying the fact that reducing the pollutants in the air is crucial for human health and environment. Therefore, the study of air pollution is very important.

We have taken US pollution dataset from 2000 to 2016. Our tool displays trends of major air pollutants such as Ozone, Nitrogen Dioxide, Sulphur Dioxide, Carbon Monoxide in the United States. We are providing both temporal and spatial views for clear visualization of the data. Using our tool, we aim to let users observe the trend in air pollution over a period of time in various states. Also, by studying the existing pattern we plan to extrapolate and make future predictions.

Project Objectives

Our dataset contains air pollutant levels for four major pollutants including Nitrous Oxide, Carbon Monoxide, Sulphur Dioxide and Ozone for the states/counties of the United States of America, for the period 2000-2016.

The primary questions we are aiming to answer with our project are the following:

- How has pollution varied over time and how have the individual pollutant levels varied over time in the United States?
- Which are the highest and lowest polluted states with respect to the pollutants present in our dataset?
- Are there any possible explanation for spikes in pollution level (if found)?
- Is there a trend in pollution level and can it be used to predict levels in the future?

We believe the answers to these questions would help us analyse how has the air quality degraded or improved over time. The environmental issues seem to be at an extremely concerning level in this day and age. With our project, we aim to build a tool which can efficiently give people an idea of how the air quality has changed over time and maybe analyse how it will change in the future. The latter is still a secondary objective of our project and would be incorporated based on time availability and difficulty level.

Data

The dataset deals with pollution in the U.S. Pollution in the U.S. has been well documented by the U.S. EPA. We aim at visualizing the distribution and trends of pollutant levels across the whole US. There are four pollutants that will be visualised, namely:

- NO2
- SO2
- O3
- CO

Apart from that, the data is already in its best form with no need of cleanup.

Data Source: <https://www.kaggle.com/sogun3/uspollution>

Data Processing

The dataset has each row representing pollution data about a county in a particular state. Thus, the dataset has a huge number of rows. To make things simpler, we are going to rollup the data to state level and then subsequently use it for our visualization.

We also plan to have a zooming-on-state feature as one of the optional features. If time allows us to do that, we will have to rollup the original data to city level.

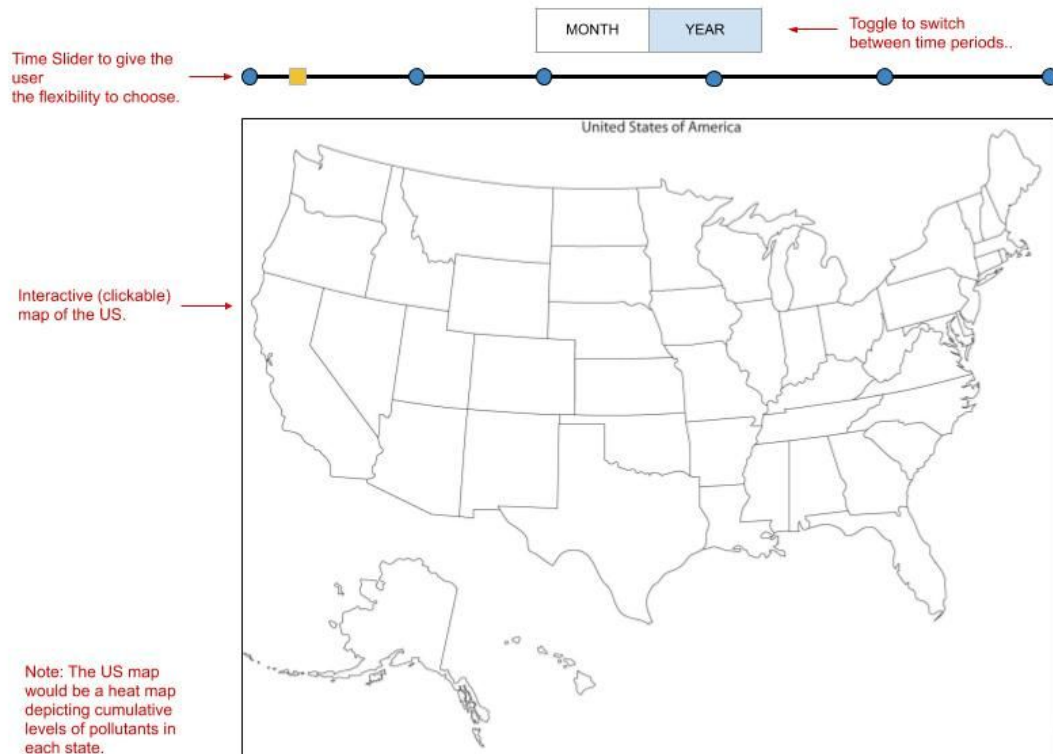
Also, the data is on a daily level. We plan on rolling that up to monthly level according to our formulated designs. All the pre-processing of the data will be done using Python(Pandas).

Further rolling up of the data, i.e., from monthly to yearly level will be done using Javascript as and when required.

Visualization Designs

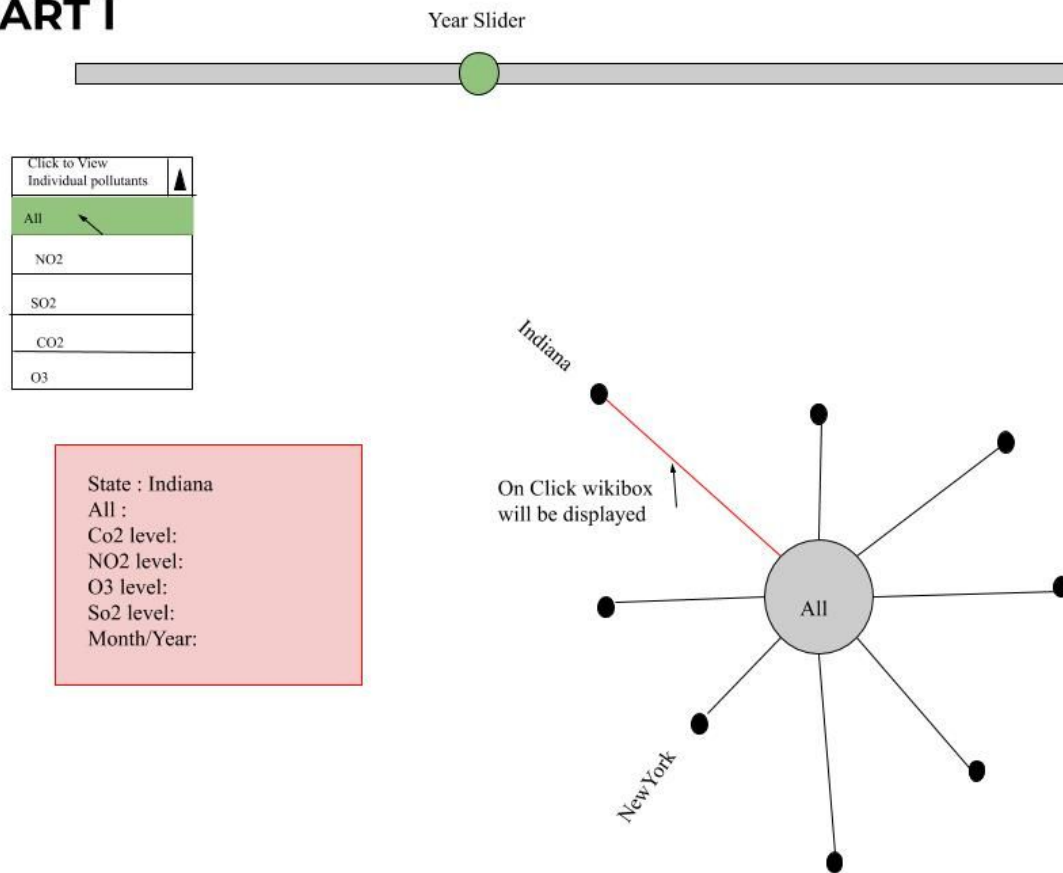
We have come up with three different visualizations to visualize the data which will be connected together and the user can navigate through them as shown below(Part 1, Part II and Part III). We have three variants for the *home screen*(Part I). Below we discuss about the pros and cons of each design. Finally, we decide to use design-3 for Part I, followed Part I and Part III.

PART I



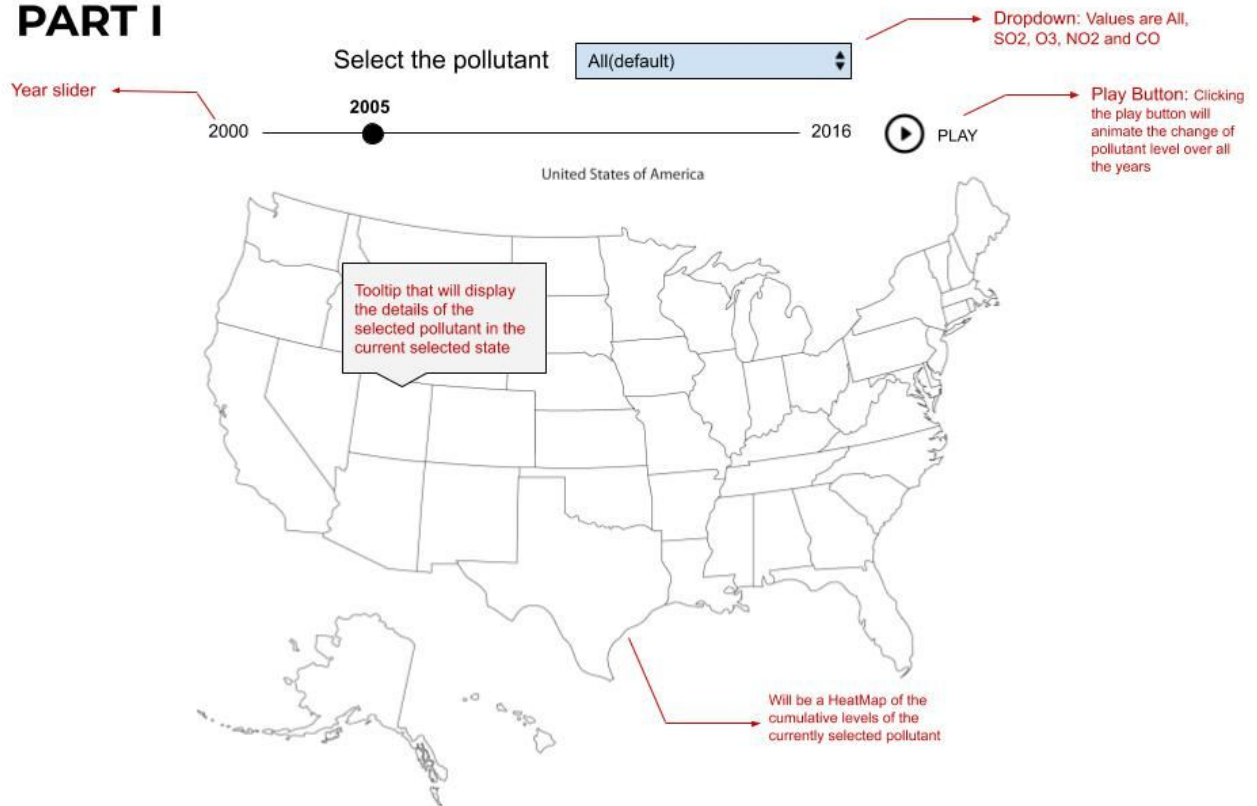
- This design shows the Map view of the pollutant level, as a whole, with a slider which allows the user to change the current year or the current month.
- The user can also toggle between monthly and yearly views.
- The map shows overall pollution level of selected year/month as a Heatmap for each state. We plan on extending this, if time allows, to city level.
- When the user clicks on a particular state, the visualisation is changed to Part 2 which is a zoomed view of the selected state with all the analysis graphs.

PART I



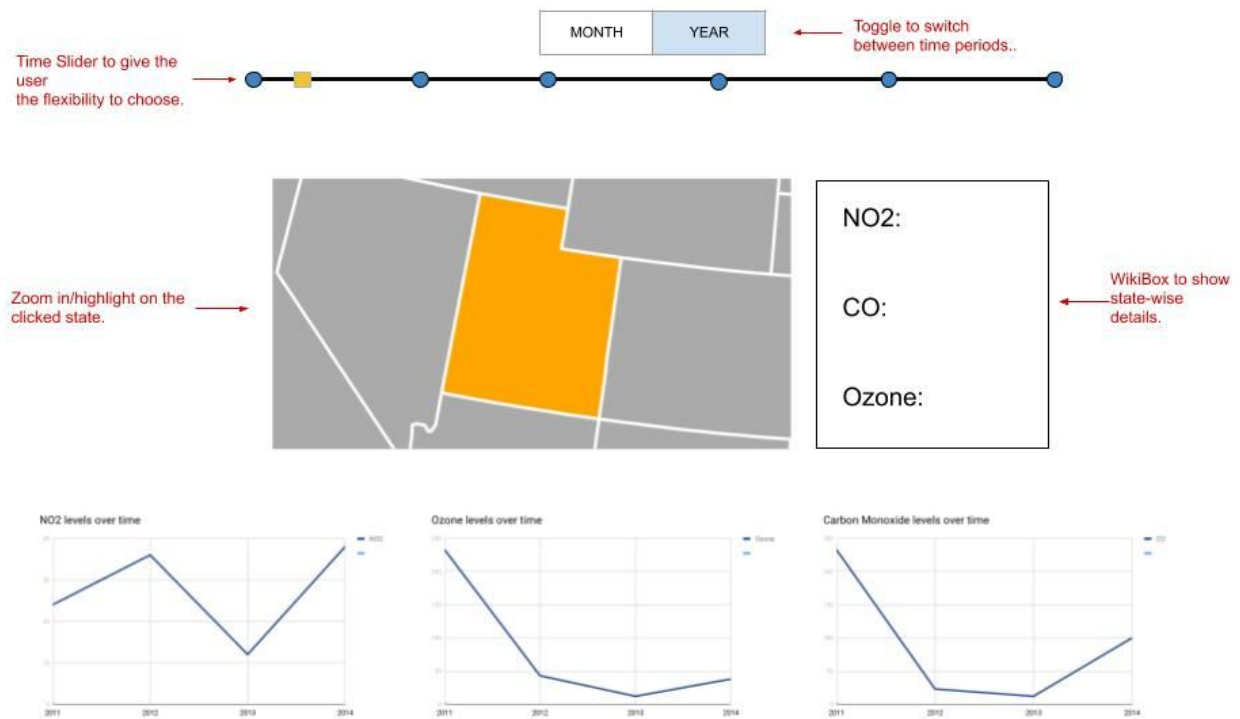
- This variant of Part I focuses on the distribution of the selected pollutant level for all the states.
- A dropdown menu allows the users to visualise data about a particular or all pollutants.
- This has been achieved by using a radar display which has all the states as its edges.
- The length of an edge signifies the pollution level in that state.
- When an edge is selected, an infobox is displayed having all the details for that particular state such as individual pollutant levels, and month/year of selected time period
- Similar to design 1, this variant also gives user the flexibility to change the current year being displayed.

PART I



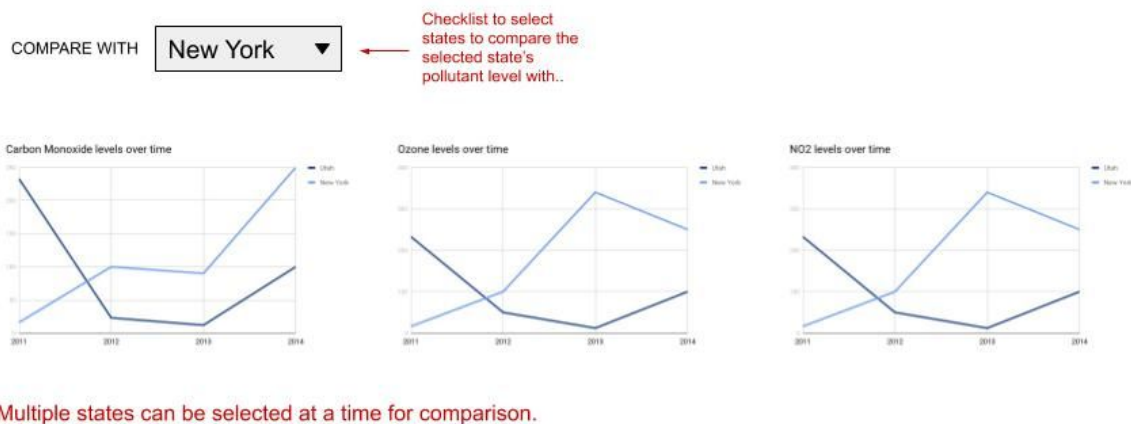
- This design is an iteratively improved version of design 1.
- In addition to the year slider and US map in the design 1, it has a play button which displays pollution level from 2000 - 2016 with animation. We plan on improving this animation from the current year to 2016, instead of 2000, as an optional feature.
- This design also provides a drop box to choose between pollutants.
- There is also an addition of a tooltip that displays the details of the state, on hovering over that state.
- Similar to design 1, clicking on a state takes the user to the zoom view of the state to part 2.

PART II



- This design view is a zoomed view of the state selected in part 1.
- As a starter, we plan on highlighting the selected state with all the analysis line graphs but later we will change this functionality to be a zoomed view of the state.
- Wikibox will appear to show pollutants levels of selected state.
- Line charts to display trend in each pollutant.
- Each time a new state is selected, the old selection will be cleared. We will use animation to have better visual effects.

PART III



- This view is an extension of the 2nd design and allows the users to compare the pollution trend between any 2 (possibly many) states using the “compare with” dropdown menu.
- As a must have feature, we will allow the user to compare two states but later we will increase this capability to 5 states.

Must-have features

- Show state-wise pollution level across the United States of America.
- Allow the user to dig into individual states and show the pollutant levels for each state - clicking on a state highlights the state.
- Give the user the ability to choose the time and time period (year, month) to view data.
- Compare the pollutant levels in different states, limited to 2.

Optional Features

- The play button animates the change in the level of pollutants over the time period of 2000 to 2016.
- Additional feature of the play button which makes the animation starts from the current year and not from 2000.
- Clicking on a state zooms the state.
- Increase the comparison capability to 5 states.

Project Schedule

Nov 1 - Nov 7

- Load the data, create the map, create the year slider
- Create the toggle switch
- Tooltip to display information about selected states
- Process Book: Overview, Motivation and Related Work

Nov 8 - Nov 14

- Dropdown to select pollutants, toggle, between month and year
- Create Wikibox to show pollutants details
- Complete Part 1 design of the project
- Process Book: Questions, Data, Exploratory analysis,

Nov 15 - Nov 21

- Create the line chart of individual pollutants
- Upon clicking on a different state, clear the line chart and wikibox of previous state and display current states wikibox and line chart
- Process Book: Design evolution, Implementation

Nov 22 - Nov 26

- Create the dropdown to compare with. Update line charts on each selection. Provide a functionality to clear the selection
- By now we must have completed all the must have features
- Work on optional features.
- Fix all the bugs, if any.
- Process Book: Evaluation.
- Complete the process book.

Nov 27

- Make sure project runs without errors.
- Final submission.