Lilian Sao de Rivera, DATS 6401, Summer 2020

Progress Report

## Trends in Pollution in the United States from 2000 to 2016

The pollution information was collected from the U.S. EPA: <https://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html>  and It is hosted in the Kaggle dataset repository. The information was downloaded and processed using phyton 3.7. The data contains some issues; one of them is the repletion of observation. An observation contains the information of a station in a city in the U.S.; however, this information may be repeated four times per city with the only difference in the columns which describe mean. There is no apparent explanation of this situation; thus, we are using the max values for each observation with a primary key composed by the fields city, county, and state.

The resulting information was uploaded into spreadsheets to be processed. This information was rearranged to comply with the requirements for each graphic and plot that is being developed.

The following step was to design the layout of the application. The flow of the site is important to instruct the user on how to use the graphics. This instruction should be intuitive. Thus the website must be consistent. This fact implies that each graph developed in the project has a uniform format. The overall layout is developed using Bootstrap 4 with a couple of high-quality images as a background. Two pages are important: the main page and the graph pages which will include dashboards with the information.

The final step is creating the graphs. These graphs are being developed using D3.js, which will provide some animation to show the increasing or decreasing pollution in each state. Some graphs may include other tools like Tableau, but at this stage, we are only considering using D3.js. The final report will address the necessity of using another type of tool and why it was used.