Visualization of mobile-monitored air quality data and its distribution onto the local community at Austin

Jing Wu

CE 394K

Fall 2018

12/07/2018

Table of Contents.

Figures.

Introduction

Air pollution has direct effects on human health – a journal article suggested that "The acute and chronic health impacts of short and long-term exposures to particulate matter (PM) are well-established in the literature". Different pollutants have different impacts on human respiratory system and cardiovascular system. According to a recently published journal article, 'population exposures to ambient air pollutants such as PM2.5 are also often highly correlated with adverse neighborhood-built environment features' (Malecki et al. 2018, pg. 2). Chemical precursors (such as air pollution) and non-chemical stressors both play important roles on human health. Air quality monitoring, therefore, is critical to protect local community from over-exposure of those pollutants. Our research group partnered with Google Street view cars for this summer to collect air quality data around central Austin. The dataset recorded is able to let us examine how air quality distribute around the city and how social-economic factors play a role on air pollution exposure. Massive amount of raw data will be processed by ArcGIS to give visualization of the dataset, and statistical analysis will be performed as a preliminary result for future use.

Data source

For this project, air quality data collected by Google Street view cars over the past summer was investigated. Central Austin area was break up to sub-polygons for driving, and two cars were distributed to different polygons each day to cover more study area. Google cars were equipped with many air pollution monitors – various gas types were observed; for example, ozone, ultrafine particles, and NOx are monitors together with other types. In this project, I will study both $PM_{2.5}$ and black carbon (BC) concentrations and how they distribute at local community here at Austin. Figure 1 shows the raw data that we collected from Google – which contains GPS coordinates, gas concentrations in each line. Note that there are 19534 lines in total – each line represents one point on the map with its unique air quality data. Raw data itself does not convey a lot of useful information – therefore, data visualization is the next important step to display information effectively. ArcGIS pro is a great tool in doing so, especially GPS coordinates are available for our data. It can also perform various statistical calculations, which is helpful for future studies. Figure 2 presents the GPS coordinates that Google cars cover over the project time. Two major highways – Mopac and I-35 are included, major neighborhoods are examined as well. Population density is shown in the bottom layer for preliminary analysis purpose.


```
-97.6956417363,30.2908318251,7,286.5,14.4525,0.4,399.6075,14.4525,6206.66666666667
-97.6955973534,30.2944142522,7,358.75,12.9575,0.9,401.116666666667,12.9575,4045
-97.6955830563,30.2915956841,7,230.482954545455,14.599537037037,0.471794871794871,401.151,14.599537037037,6651.25
-97.6955805993,30.2937183333,7,199.25,15.6620833333333,0.3,401.555,15.6620833333333,6320
-97.6955784588,30.2901629136,6,449.666666666667,12.6175,1.45,401.48,12.6175,5630
-97.6954553312,30.2930144795,7,393.2,12.404,0.62,400.46875,12.404,6639
-97.6953577868,30.2909412948,7,254,14.952,0.383333333333333,399.226,14.952,6870
-97.6952855917,30.2944227173,7,174.5,13.4023333333333,0.6,401.078333333333,13.4023333333333,5277.75
-97.6952773902,30.2916463288,7,181.958333333333,13.667,0.7,400.71875,13.667,6654.5
-97.6952688559,30.2937272389,7,236.75,13.815,0.4,401.044,13.815,5970
-97.6952671764,30.290179943,6,-32.6,12.534,0.8,400.225,12.534,5660
-97.6951438683,30.2930271052,7,360.333333333333,13.7225,0.333333333333333,399.88,13.7225,5656.66666666667
-97.6951246009,30.2945439158,2,256.75,22.1,NaN,427.8,22.1,21200
-97.6951133657,30.2941817775,7,-50.5,13.0075,0.45,400.566,13.0075,4392.5
-97.6951043784,30.2939112175,7,263,13.4433333333333,0.325,400.544,13.4433333333333,4756.66666666667
-97.6950924276,30.2935487814,7,269.5,12.7233333333333,0.275,399.640416666667,12.7233333333333,4275.41666666667
-97.6950843515,30.293278198,7,302.5,13.54,0.2,399.31375,13.54,4421.66666666667
-97.6950713117,30.2928293391,7,224.666666666667,13.97875,0.566666666666667,399.475,13.97875,5680
-97.6950634873,30.2925587509,7,212.041666666667,13.09625,0.583333333333333,399.790833333333,13.09625,5851.25
-97.6950518044,30.2922222713,7,160.75,14.19,0.466666666666666,400.240833333333,14.19,6063.75
-97.6950475363,30.2909652455,7,167.2,14.452,0.56,398.941666666667,14.452,6688
-97.6950403541,30.2919517805,7,209.833333333333,12.972,0.2,399.6235,12.972,6213.33333333333
-97.695028935,30.2916812882,7,297,13.6866666666667,0.5,399.946666666667,13.6866666666667,6747.14285714286
-97.6950215312,30.2914769415,7,210.416666666667,13.0866666666667,0.5,399.59,13.0866666666667,7046.66666666667
-97.6950119741,30.2912063956,7,40,14.3533333333333,0.616666666666667,399.311666666667,14.3533333333333,6770
-97.6949935133,30.2907068866,6,112,13.8866666666667,0.533333333333333,399.636666666667,13.8866666666667,6643.33333333333
-97.6949831065,30.2904363638,6,415.466666666667,13.2549305555556,0.798809523809524,399.498888888889,13.2549305555556,6145.55555555556
-97.6949760992,30.2937345402,3,317,18.52,0.8,399.46,18.52,6400
-97.694959507,30.2944315053,2,327.25,15.975,0.4,439.515,15.975,10077.5
-97.6949467666,30.2930316869,2,168,14.94,0.7,398.14,14.94,7680
```

Ln 19534, Col 63

Figure 1. Raw data collected by Google Street view cars.

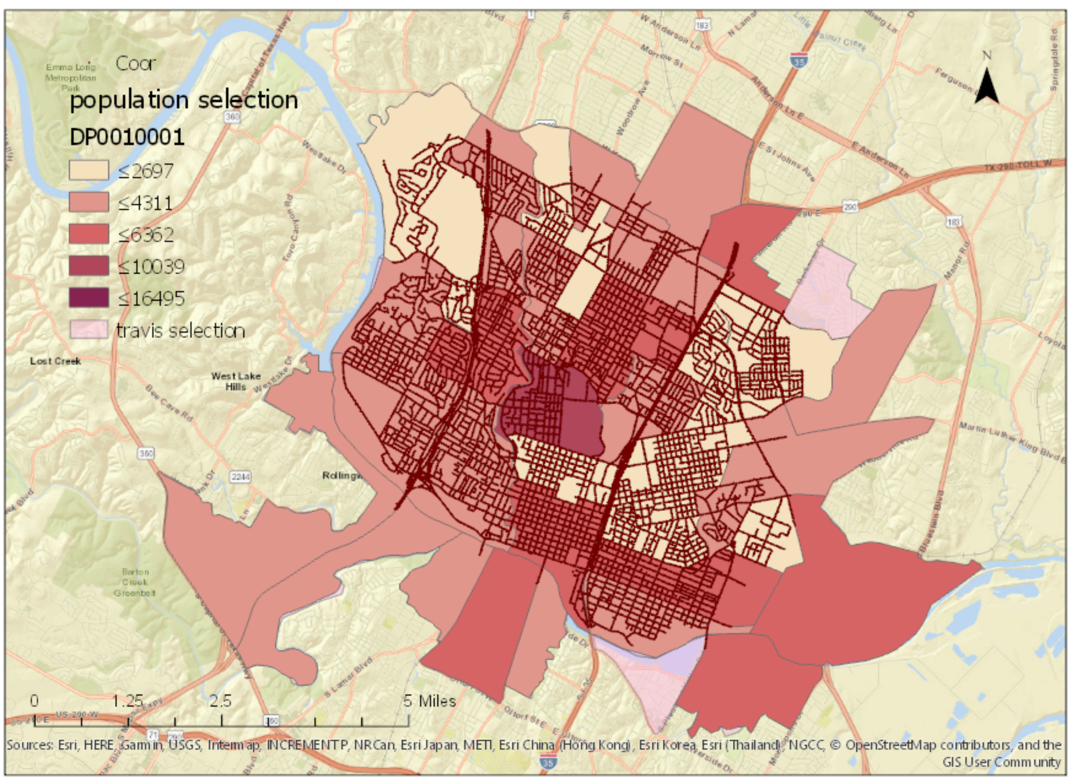

Figure 2. Driving map of two Google cars.

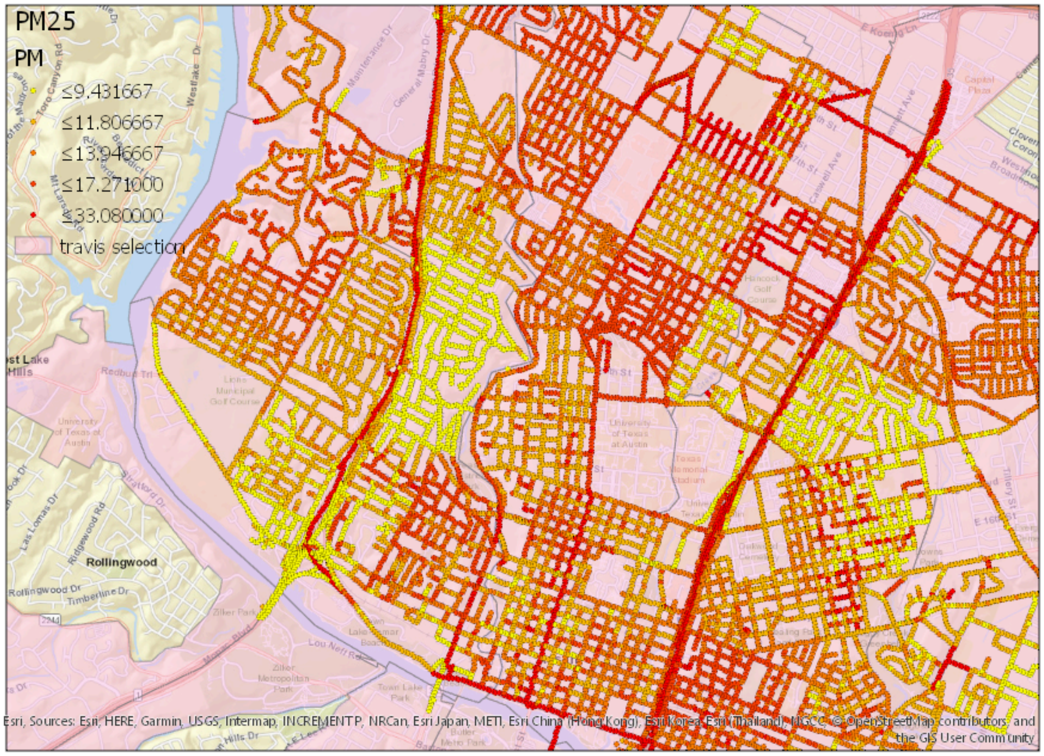

Figure 3. $PM_{2.5}$ distributions.

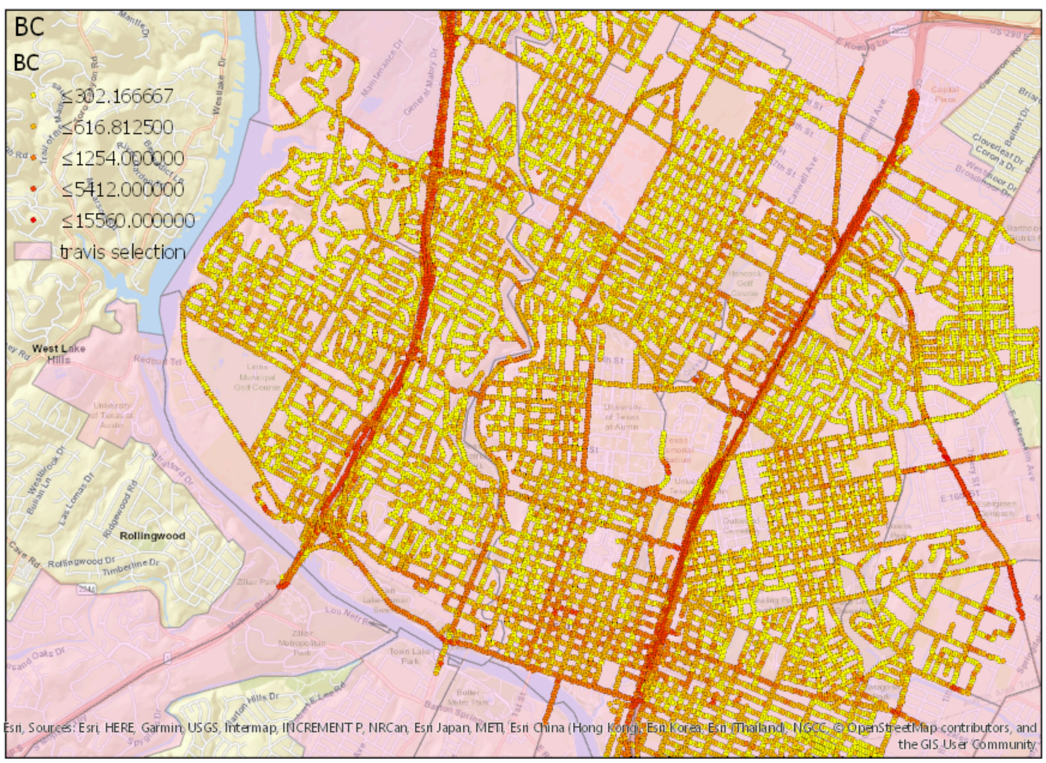

Figure 4. BC distributions.

After linking PM2.5 and BC with GPS coordinates, air quality can be easily shown on the map. PM2.5 and BC data has its own variation, but similarities can be observed spatially (Figure 3 and Figure 4 below). Note there are several limitations in our data collection. First, not all polygons were driven at the same time – due to the limitations of work hours. Second, each coordinate on map is showing the median value of 7-8 measurements – this might not be a great representation for the annual concentration. The driving map was designed by Google map, the value collected at each coordinate may represent the air quality at different time of the day. Two layers of demographic information are used in the project, both from ACS (American Community Survey, 2016 dataset). One is health care spending by tract, the other one is individual income by tract (Figure 5 and Figure 6 below). Regions that overlap with air quality data are selected, and both layers are shown by color scale. As present in the figure, health care spending across the city is not various whereas income level varies a lot; in general, community at western Austin has a higher annual income compared with community at eastern Austin. UT-Austin campus is located at the middle tracts – income level is lower than the surrounding tracts since students represent the major population in those tracts. Demographic information can help us better understand how local communities get exposed by air pollution ($PM_{2.5}$ and BC in the project) and can help policy makers to identify air quality hot spots – policies or regulations can be implemented to better protect human health.

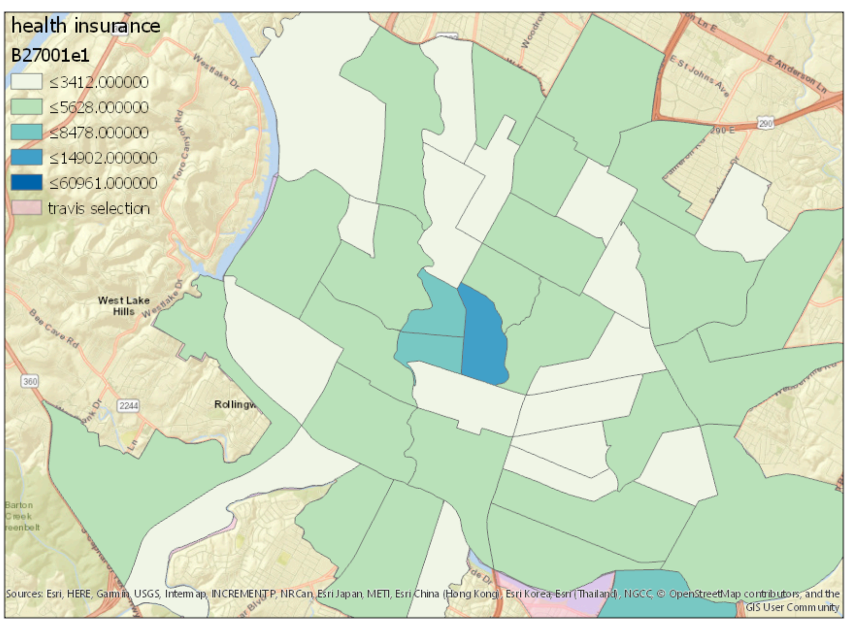

Figure 5. Health insurance by tract.

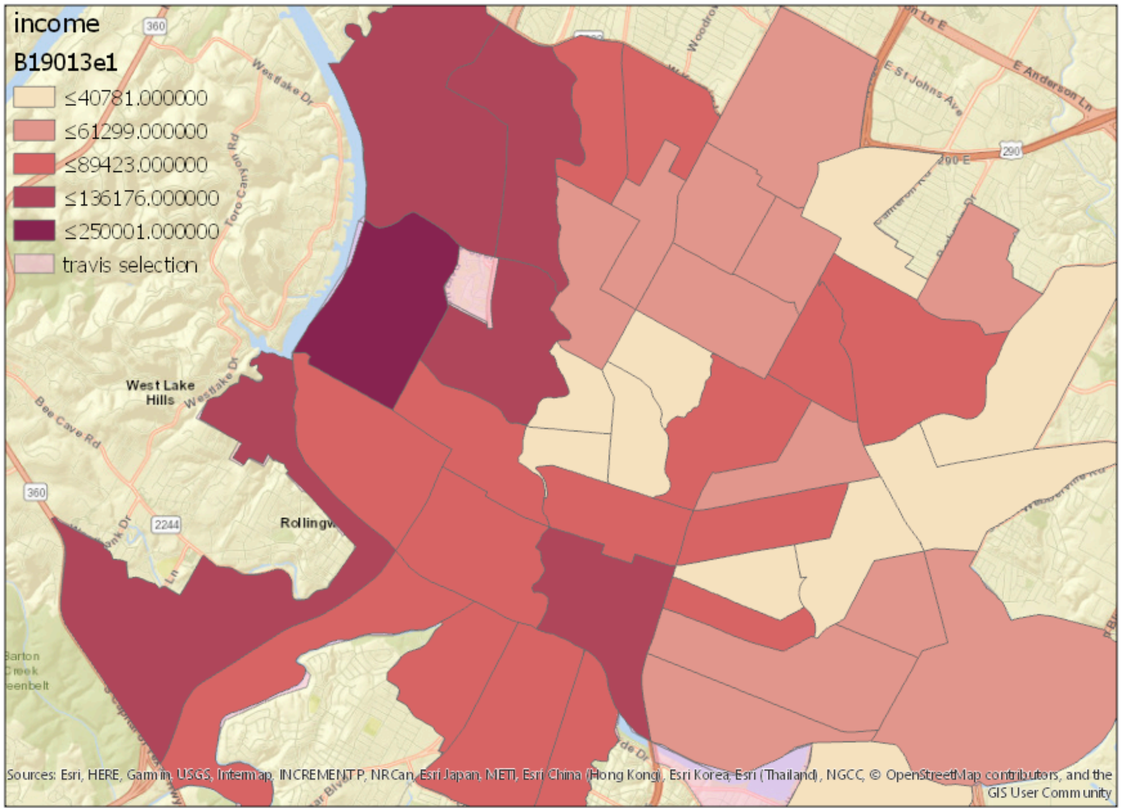

Figure 6. Individual income by tract.

Methods and Results

In order to have a better visualization of where the locations are mostly polluted, one of the methods used for the gas data is called hot spot analysis - which is a built-in function in ArcGIS pro. It conducts analysis for the given data and creates a new feature class where it identifies statistically spatial clusters of high values (hot spots) and low values (cold spots). $PM_{2.5}$ and BC data are processed by hot spot analysis and are projected onto both health care spending and income level layer (Figure 7, 8, 9 and 10).

Figure 7. BC hot spot analysis projected onto health care spending.

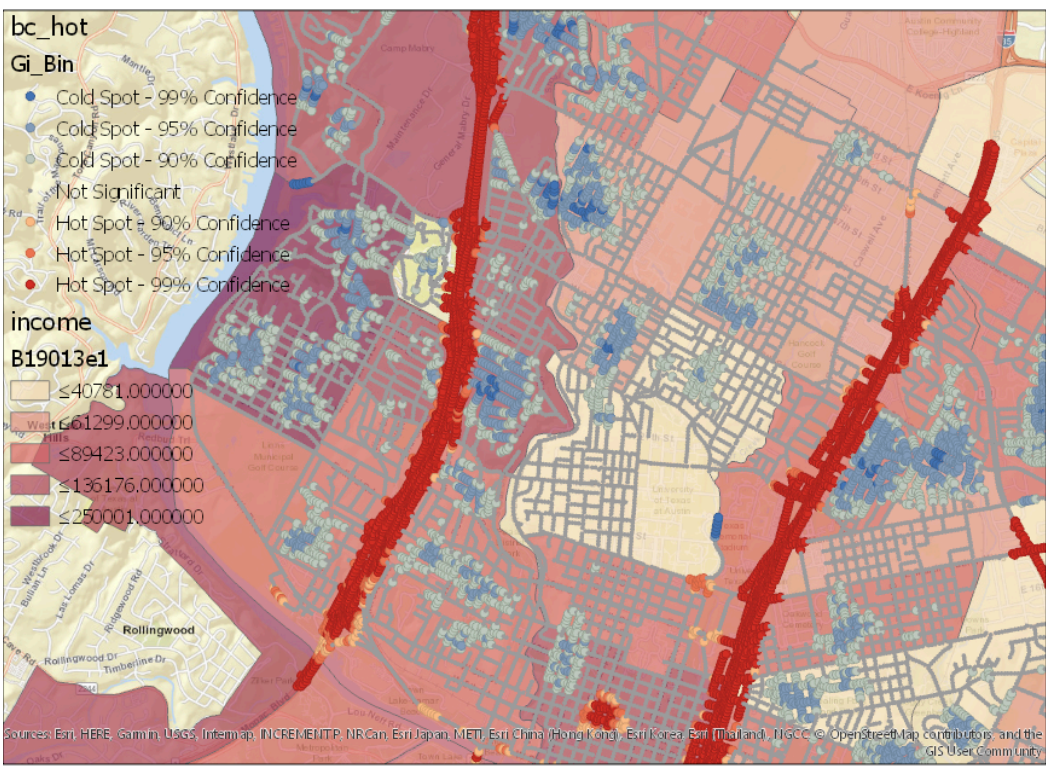

Figure 8. BC hot spot analysis projected onto income level.

As shown in the above figures, BC hot spots are most likely gathered around two major highways (Mopac and I-35) during the observation period. Note that data at UT campus is not availabel, this is due to logistics of data collection process. It would be very interesting to see how student community exposes to air pollution during the summer. Another hot spot region is downtown Austin, which compacted with plenty of resurants and commertial buildings. Again, black carbon sources are mostly contributed by traffic and cooking sources – the collected data is also suggesting this common fact. Figure 9 and 10 show the hot spot analysis for $PM_{2.5}$ concentration over the study area. Compare to BC level, $PM_{2.5}$ has a more spread-out distribution for central Austin. Highways and busy traffic still has high impacts on $PM_{2.5}$ level, and other neighborhoods (hyde park region, triangle region) suffers from high $PM_{2.5}$ as well. Again, time is another factor that will influence air quality throughout the day and the way data was collected introduces certain level of uncertainty to the analysis.

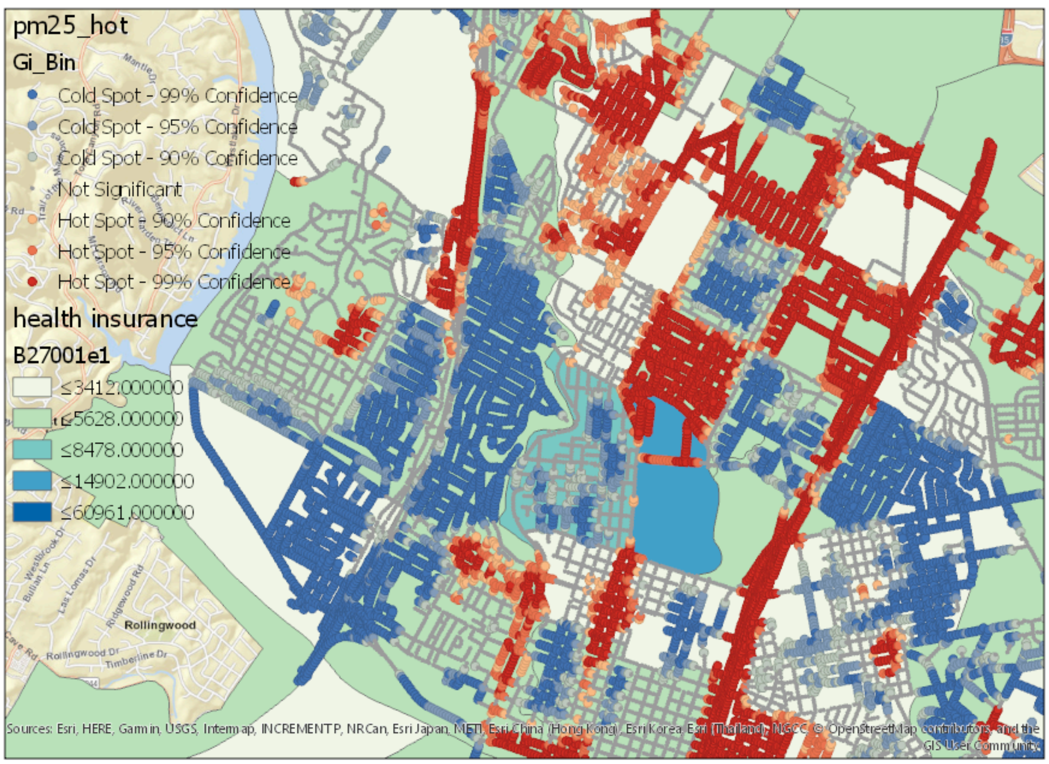

Figure 9. $PM_{2.5}$ hot spot analysis projected onto health care spending.

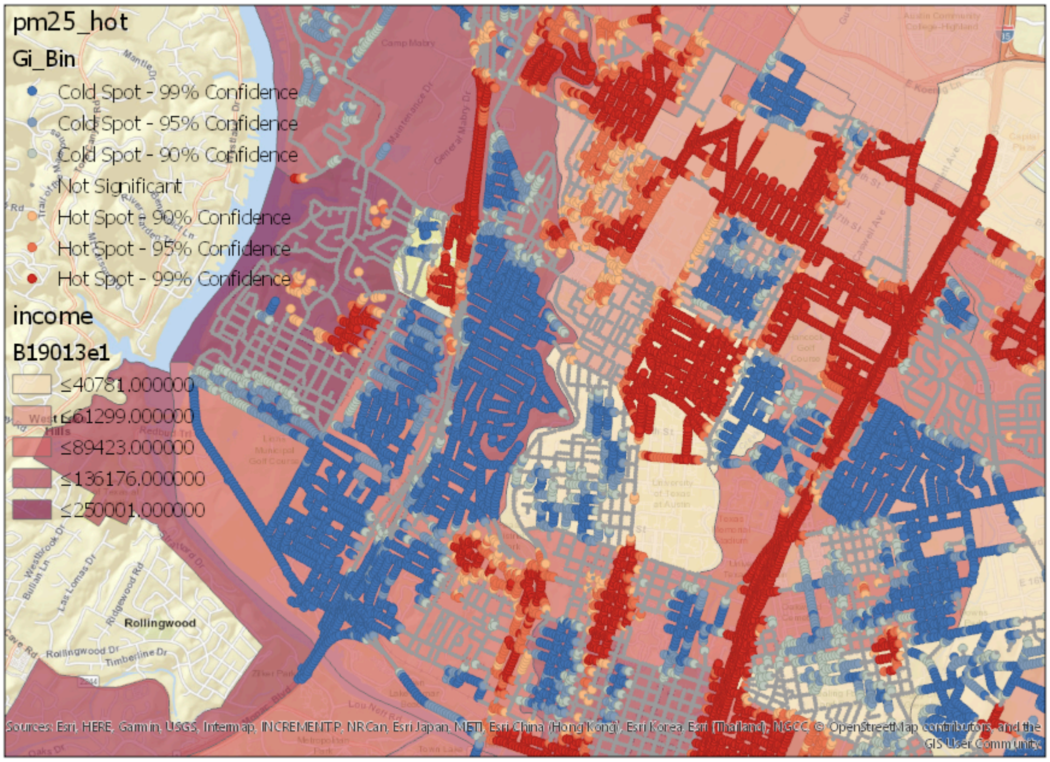

Figure 10. $PM_{2.5}$ hot spot analysis projected onto income level.

Data visualization via ArcGIS pro is very useful to present data, however, more statistical analysis can be done by other tools in GIS. Points are groups by tract on the map, and the mean value for each tract is summarized by GIS – next, the value is displayed by color scale. Histogram of the study area is constructed for both $PM_{2.5}$ and BC and the mean values is displayed by a vertical red line.

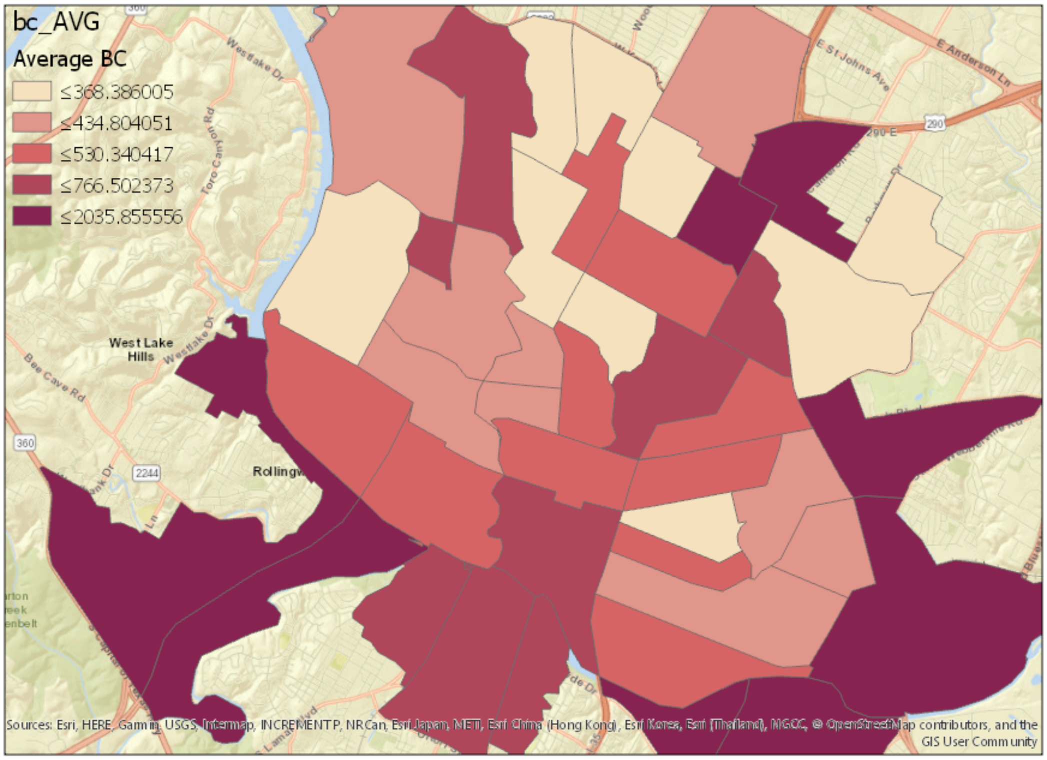

Figure 11. Group average of BC by tract.

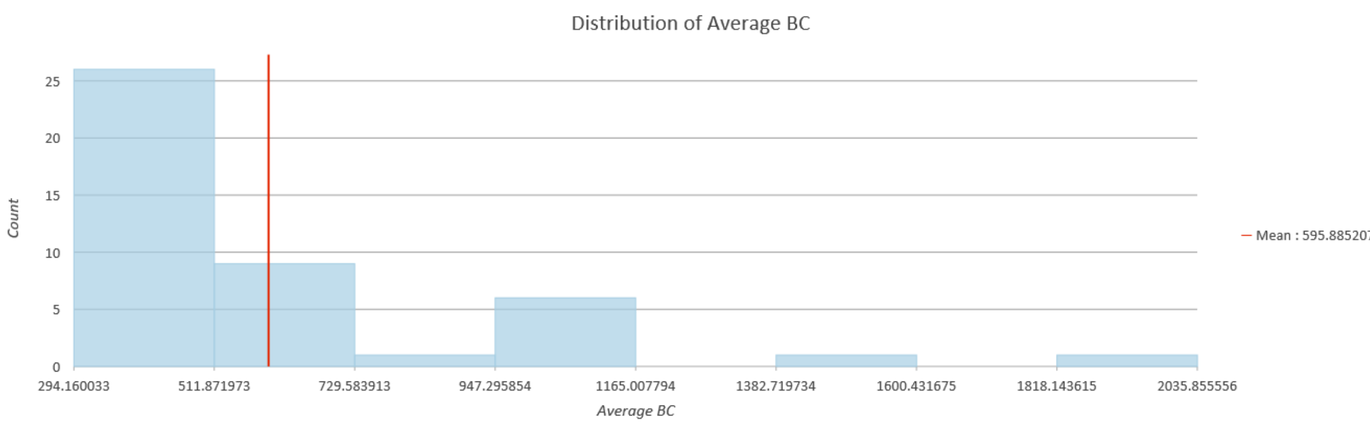

Figure 12. Histogram of group average of BC.

For black carbon concentrations, most tracts fall in the region of 294 – 512 ug/m^3 and mean value is around 596 ug/m^3. Again, from the hot spot analysis, most of the black carbon are surround by highways and heavy traffic region – tracts that contain those busy roads tend to have a higher BC mean concentrations where as other neighborhood has a relatively lower mean value. As mentioned before, black carbon has long term health impacts on human and the residents living near traffic should pay more attention to respiratory protection. For those who live in relatively lower income tracts, more health benefits or subsidies are encouraged.

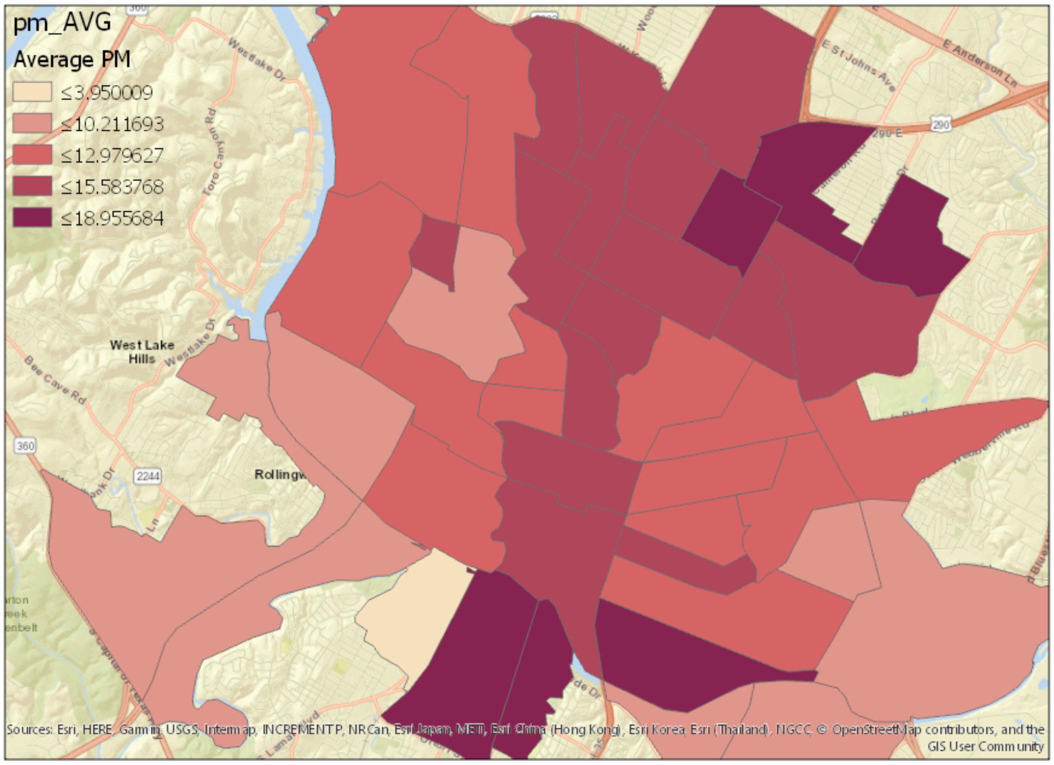

Figure 13. Group average of $PM_{2.5}$ by tract.

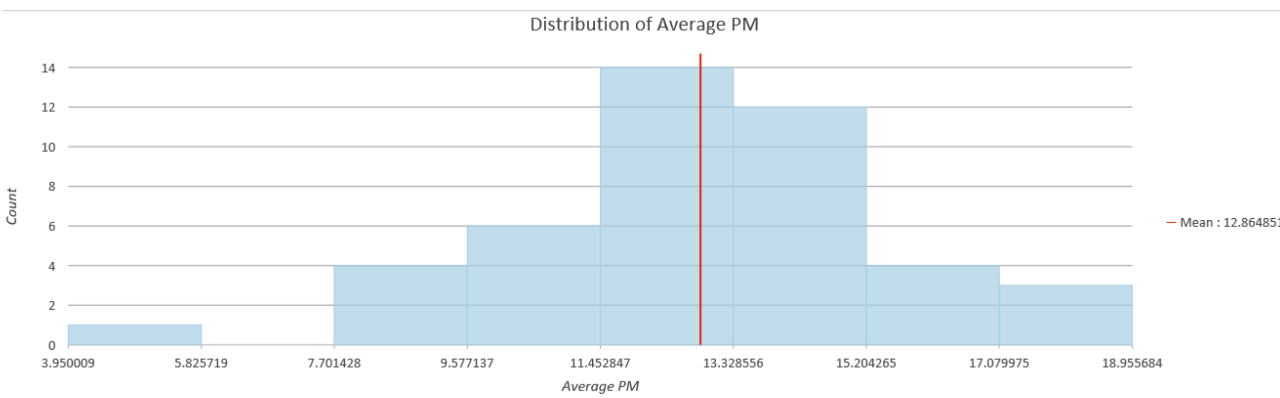

Figure 14. Histogram of group average of $PM_{2.5}$.

Another studied pollutant, $PM_{2.5}$, seems to have a normal distribution by the histogram shown above. Since $PM_{2.5}$ is one of the criteria air pollutants, it must follow the NAAQS standards by EPA. According to EPA website, primary $PM_{2.5}$ annual mean (averaged over 3 years) should not exceed 12 ug/m^3, secondary $PM_{2.5}$ annual mean (averaged over 3 years) should not exceed 15 ug/m^3. By the data collected, the mean value is about 12.9 ug/m^3, which is slightly above the primary standards, but below secondary standards. As mentioned before, there are limitations of the data collecting process – therefore, a more comprehensive dataset should be used to give a more accurate estimate of annual average. People who work or live at those dark red tracts should be aware of the potential $PM_{2.5}$ pollution that might appear for the rest of the year and regulations should be enforced stronger in order to reduce $PM_{2.5}$ level for those tracts. In addition, for some outside tracts, not all coordinates are covered, which might lead to an over/under estimation of the group average.

Conclusion and future works

ArcGIS provide a very easy and efficient way to visualize our raw dataset. Pollution concentration can be shown by color scale and is very reader-friendly – GIS provide a powerful tool to deliver air quality data to the general public. It also enables to combine the air quality data with demographic information (ACS data in this project) to make calculations of air quality status, and more importantly, provide predictions and suggestions for policy implementation to improve air environment for the local community. In the future, analysis over block groups, instead of tract, can be performed in order to give a more precise calculation – coverage rate of coordinates per units can be much higher in this case, which will make the group mean more representative. Demographic data can be more targeted as well – since elder population tends to have a more severe response to air quality problems, health care spending of elder generations can be extracted from the database. More plots can be made by the data given, for example, a scatter plot of exposure concentration vs. age and exposure concentration vs. social income would be interesting to see – those potential results can better help policy making process become more effective and evidential.

Work Cited

https://www.epa.gov/criteria-air-pollutants/naaqs-table (NAAQS standards)

Malecki, K., Schultz, A., Bergmans, R. (2018). Neighborhood perceptions and Cumulative Impacts of Low Level Chronic Expoture to Fine Particular Matter (PM2.5) on Cardiopulmonary Health. *International Journal of Environmental Research and Public Health*. 15(84), 1-19.

Maji, K., Dikshit, A., Arora, M., Deshpande, A. (2017). Estimating premature mortality attributatble to PM2.5 Exposure and benefit of air pollution control policies in China for 2020. Science of the Total Environmentla. 683-693.