Report: Geospatial Relationships of Air Quality and Asthma Rates in New Haven, CT

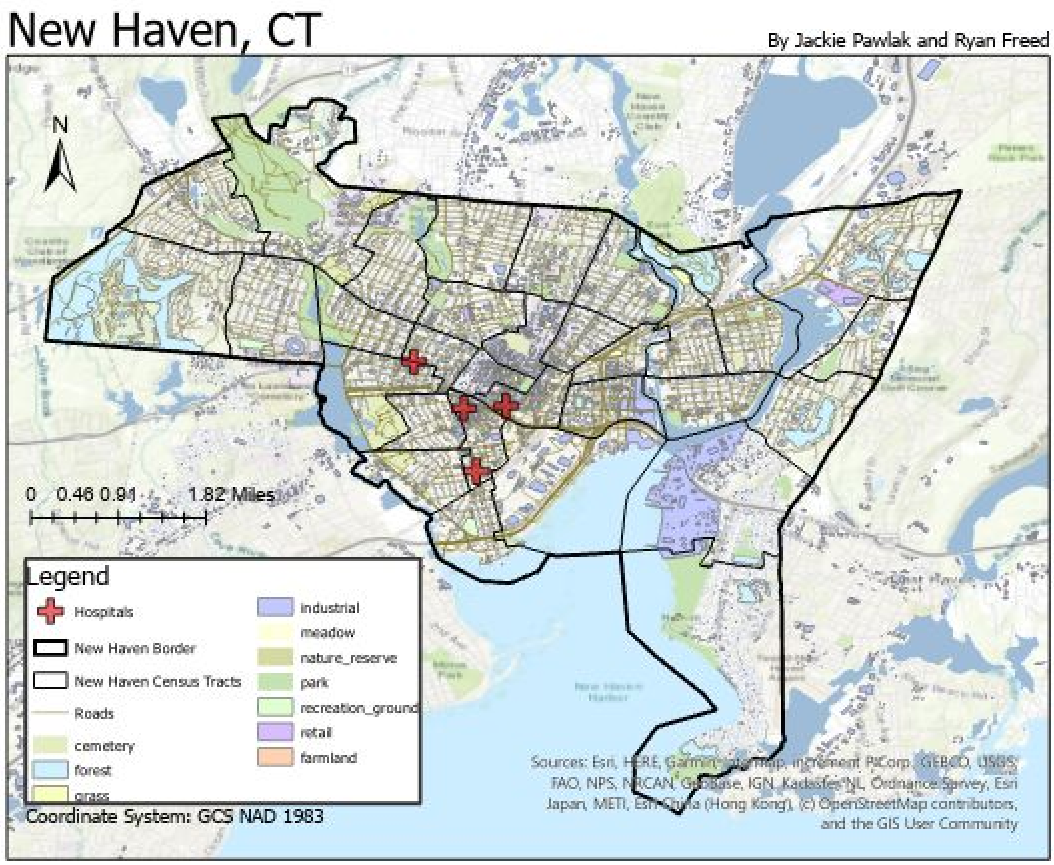
Ryan Freed (rfreed@clarku.edu), Jackie Pawlak (jpawlak@clarku.edu)

Abstract:

Asthma exposure in urban environments is typically a multifaceted public health issue. Asthma rates have only increased in United States urban geographies, where in New Haven, CT is no exception. This study aims to identify geographic areas of interest where asthma rates are outside citywide norms, evaluate the phenomenon over a 10 year study period, and identify the relationship between rates of asthma in the city and different socioeconomic cohorts. In our analysis, we found that census tracts that are lower income and higher percent black population incurred higher asthma crude prevalence in the city. We found no relationship between PM 2.5 levels and asthma rates in the study area from the 2010 data, but some positive correlation in the 2019 data. While no definitive correlations can be made between these evaluative factors and asthma rates due to a small sample size in the study area, these results suggest some positive relationship between race, air quality, and income to asthma rates.

Problem Statement and Objectives:

Our project was to identify vulnerable populations in New Haven, CT, ones that were most affected by asthma and poor air quality. We suspected a positive correlation between unfavorable socioeconomic status, air quality and asthma prevalence. We also aimed to analyze these relationships over time to see if the experience of vulnerable populations in New Haven was improving. We chose New Haven as a study area due to its high asthma rates relative to other metropolitan areas in the United States (Weir 2018).



Map 1: Study Area of New Haven CT, with land use featured for spatial context

GIS was the proper tool in this analysis to identify these spatial relationships of interest. The final product as a map was also the most accessible mode of reading and visualizing our findings. When considered the following determinants of health as high risk for asthma: air quality, socioeconomic status, and spatial relationship to health resources, GIS tools helped us evaluate the factors that seem to be more influential on asthma in the city and display these results.

Previous research on the geospatial relationship between air pollution and asthma in New Haven was conducted by the New Haven Green Fund which we referenced when developing our research objectives. The New Haven Green fund conducted a study related to walkability of the region and also investigated asthma rates as a way to understand environmental issues (Ankrah 2019). They observed hospitalizations due to asthma by age and race across the state, and found demographic disparities in the hospitalization numbers. The authors argued that areas of high hospitalization rates at the census tract level were also areas with low-income individuals and communities of color (Ankrah 2019). We found this report very insightful for our objectives, however the visualization of the data could be improved upon and further analysis could be done to identify the census tract areas that were most at risk for developing asthma according to their income level, race, and exposure to poor quality air.

Research done on asthma and air quality has found it difficult to exclusively link air quality to asthma instances due to the other socioeconomic factors that also play a role in a person's susceptibility to asthma. In their study done on Detroit, Lemke et. Al. indicate strong neighborhood-level spatial relationships between asthma-induced hospitalizations and air quality in Detroit, with statistically significant relationships between pollutant presence and asthma hospitalization rates at the ZIP code scale of analysis (2013).

Data:

We received our data from the US Census Bureau, the 500 Cities Project, the US EPA OAQPS Air Data provided by the ArcGIS hub and the Connecticut Department of Public Health. The Census Data gave us the shapefiles for the census tracts in New Haven. The rest of the data we were able to download as .csv files or excel tables that could then be joined to these census tracts. Table 1 below shows each data source, format, resolution, and date.

DATA LINK	DATA NAME	YEAR	DESCRIPTION	FILE FORMAT
Link	Asthema and Hospitalization Data	2010 - 14	Combined Asthema Allergy and Hospitalizations by Census Tract	.json
Link	Connecticut GIS Data	2010 Census	Road Data, Census Tract Shape files for append, boundaries and building indices	.shp
Link	EMP2.5 Levels GIS Data	2010 - 14	EMP2.5 measurments by census tract	.csv
Link	NewHavenFinal.pdf	2010 - 2014	Combined Asthma ED and	.xlsx

			Hospitalizations by Census Tract, New Haven, Connecticut, 2010 - 2014	
Link	DataHaven		Community Wellbeing Survey	.pdf
Link	500 Cities Project	2019	500 Cities Project: Local Data for Better Health	.csv

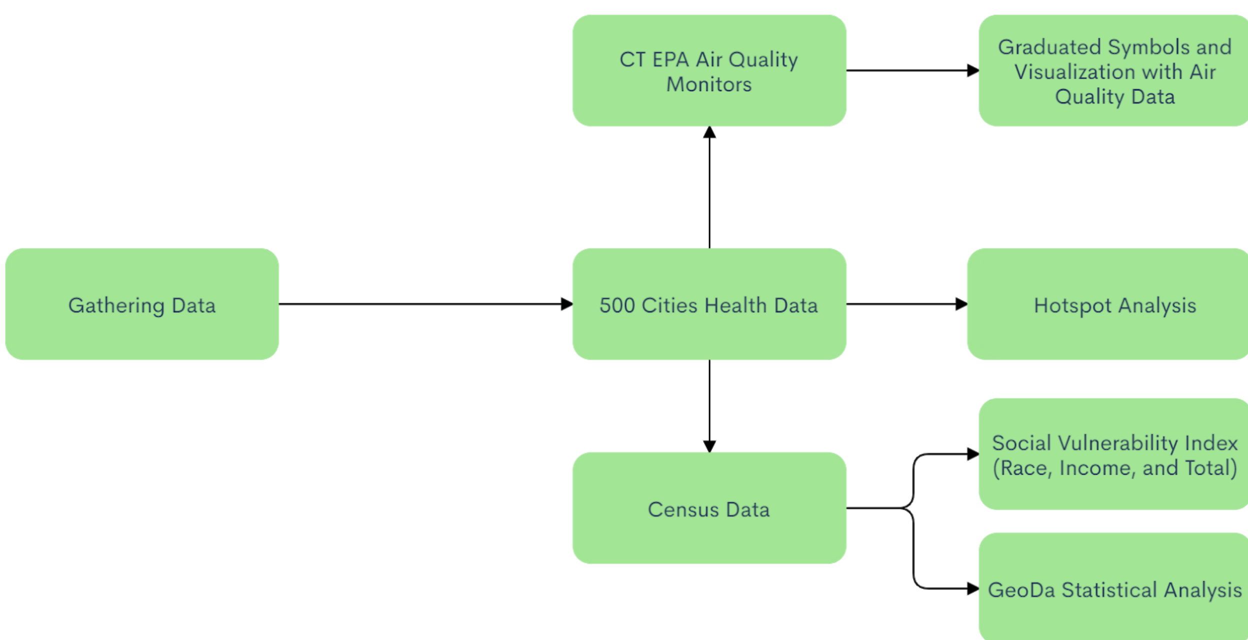
Table 1: Data Sources

Methodology:

We utilized three geospatial tools for our study, the first of which was a hot spot analysis. A hot spot analysis was chosen to evaluate if there were areas outside the crude prevalence norm, as we were able to compare these tracts to the crude prevalence standard deviation supplied in the dataset. Further, a hot spot analysis would determine cold spots that could be used for comparison in the analysis. The 500 Cities Project Public Health Data on asthma crude prevalence was appended to New, Haven CT census tracts, where then a hot spot analysis with false discovery rate correction was produced. The spatial relationship used was a queen-based contiguity due to the relatively small sample size of the district, and the perceived spatial compactness of asthma rates across the study area. Similar hotspot analyses have been conducted with public health data. Tuluri et. Al. (2007) evaluated PM2.5 levels by census tract in New York state, using a kernel density estimation to aid in their analysis. While their study utilized a K nearest neighbor conceptualization of spatial relationships, the principle of neighborhood relationships for this type of data is an appropriate one. Our second geospatial tool was to create and map a vulnerability index of the relevant socioeconomic factors. This was a valuable way to quantify socioeconomic factors that we were interested in studying, rather than the guesswork of a visual analysis of simple choropleth mapping. We applied the following function with the Field Calculation tool onto three variables- average annual income, percent black population by census tract, and crude prevalence of asthma

$$Standardized\ Value\ (SV)\ =\ Standardization\ Size\ *\ (Raw\ Value\ -\ min)/(max\ - min)$$

where the standardization size was the range of the index scale. We decided on median annual income as one indicator because it encapsulated the most data in urban geography, surveying renters and homeowners alike. We chose percent black population as another indicator because it was the only racial group in the city with a statistically significant relationship between PM2.5 and asthma crude prevalence levels, which also aligned with previous literature on health outcomes and race in New Haven (Ankrah 2019). From there, we took the mean of these standardized values as our index. Vulnerability indices are common in public health research, with the CDC using multitudes of socioeconomic factors to create a Social Vulnerability Index (SVI) widely cited by researchers. Kolling et. Al. (2012) evaluated the strength of the SVI in determining geospatial asthma outcome. They found that these social indices were strong indicators of county-level variance in asthma rates as well as tracking PM2.5 rates. Lastly, we used GeoDa Spatial analytics tools to visualize our findings. This was a crucial component of the analysis, as it granted tests for statistical significance (p values) and evaluating how effectively the line of best fit mapped onto the data (R^2 value). To do so, we exported the relevant indices, PM2.5 levels and asthma rates into their own spatial feature files and used the GeoDa scatter plot tools and standardizations to visualize the data.



Data workflow and spatial analysis done with each dataset.

Results:

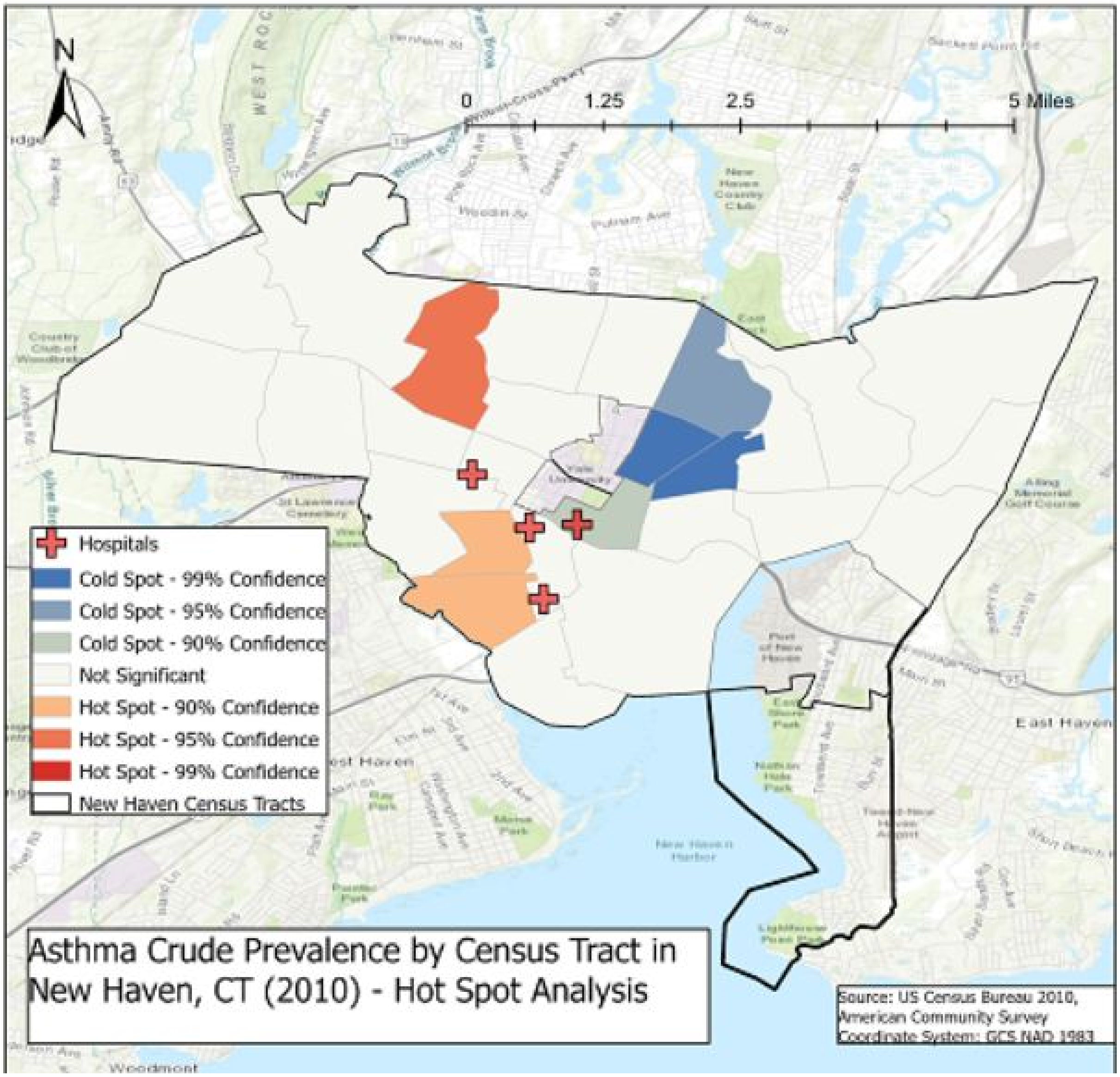
Our findings suggest that, generally, asthma has increased in prevalence from 2010 to 2019. First, our hotspot analysis showed consistent hot spots of asthma cases over time. Newhallville, Dixwell, and West River neighborhoods were consistent hot spots (see Map 2 and Map 3). East Rock, Fair Haven, and Webster Square neighborhoods were consistent cold spots. These became areas of interest in our study.

The vulnerability index helped us identify areas with relatively high asthma and socioeconomic risk. Maps 4 and 5 helped visualize the data between the two study years. There was consistency in the vulnerable areas between the two study years. These neighborhoods were Dixwell, Newhallville, West River, West Rock, and Edgewood, the same neighborhoods of interest produced from the hotspot analysis.

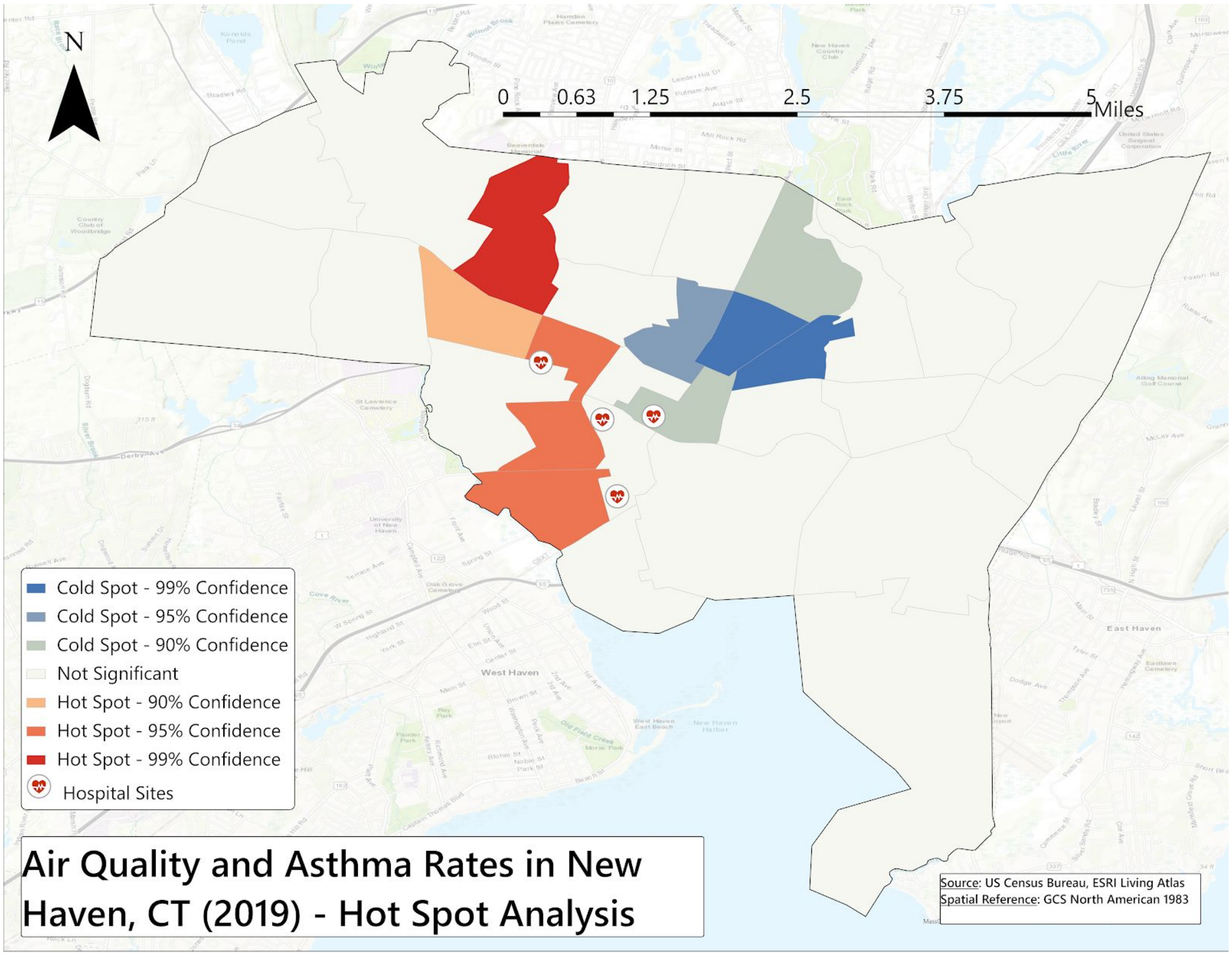
Our statistical analysis using GeoDa software allowed us to examine the trends in our data through scatterplots. Figure 1 shows the scatterplot for percent Asian population versus the asthma rate in 2019, where we saw the data concentrated in the low numbers for prevalence regardless of the percent of the population that was Asian. This data was consistent in 2010. Figure 2, which shows the percent Black population and asthma rate in 2019, demonstrates that the percent Black population is more likely correlated with asthma rate than other racial demographics. We also created a scatterplot for PM levels against asthma crude prevalence (Figure 4) which did not show a statistically significant relationship in the 2010 data. However, there was a positive correlation between both the vulnerability index and asthma crude prevalence in the 2019 data. This lack of strong trends in the 2010 data was surprising but may be due to the way weather and air quality affect large areas and cannot be confined to individual census tracts.

Finally, Figure 3 maps the vulnerability index vs the air quality data for both years. The plots generally show data consistent to the vulnerable populations that we identified with our index, but the 2010 data does demonstrate a slightly negative trend in the data. There is one area that is identified as having a high asthma rate to the West, called Westville, but is mostly recreational land and water, a statistical outlier that is present in the scatterplots but not representative of the entire dataset. We were also missing data from a few of the census tracts in

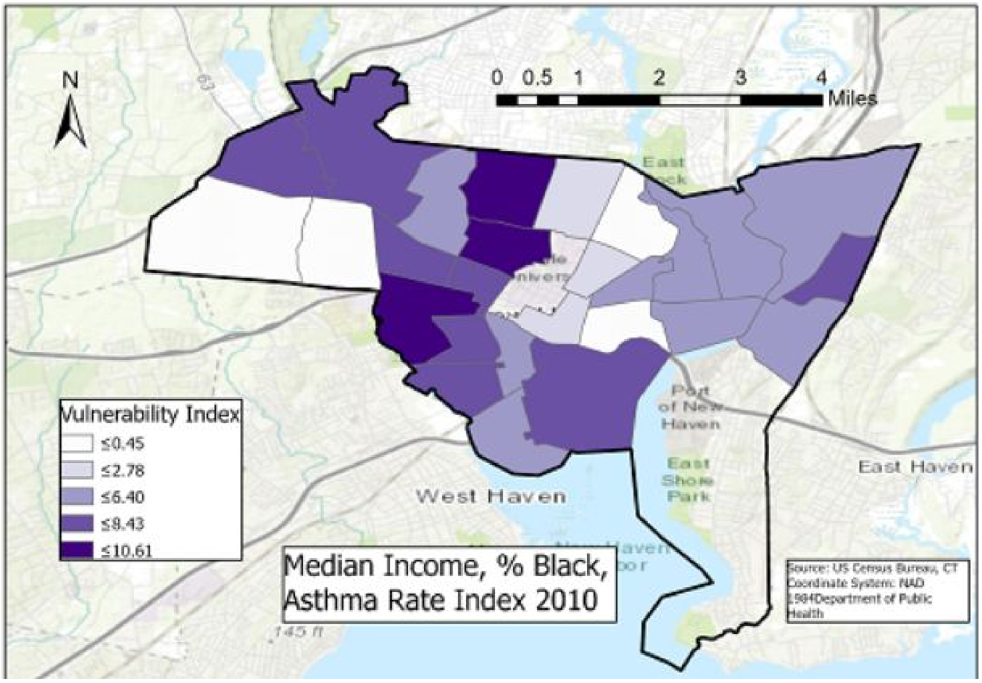
2010 due to differences in data collection between these periods. Due to the small sample size, this may have exacerbated our linear regression and overrepresented trends in the data, as is the case in Figure 1. For this reason, it is important to consider spatial context in the analysis and use the scatterplots as a companion to our maps.



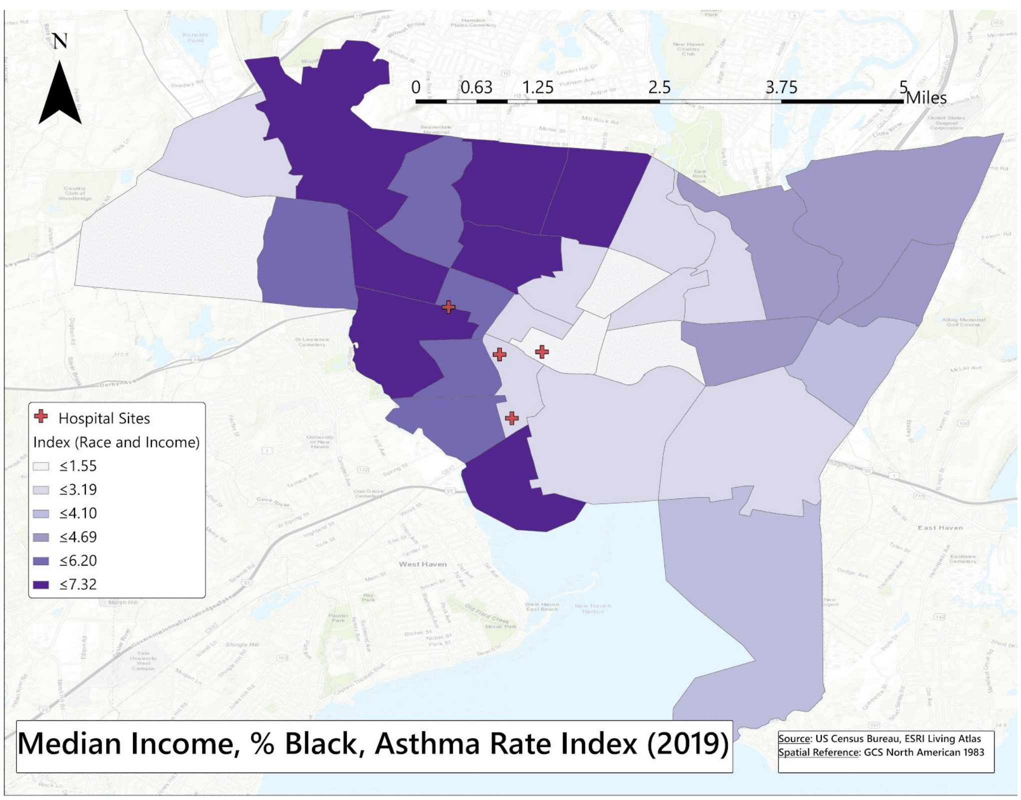
Map 2: Asthma Rate Hot Spots in 2010



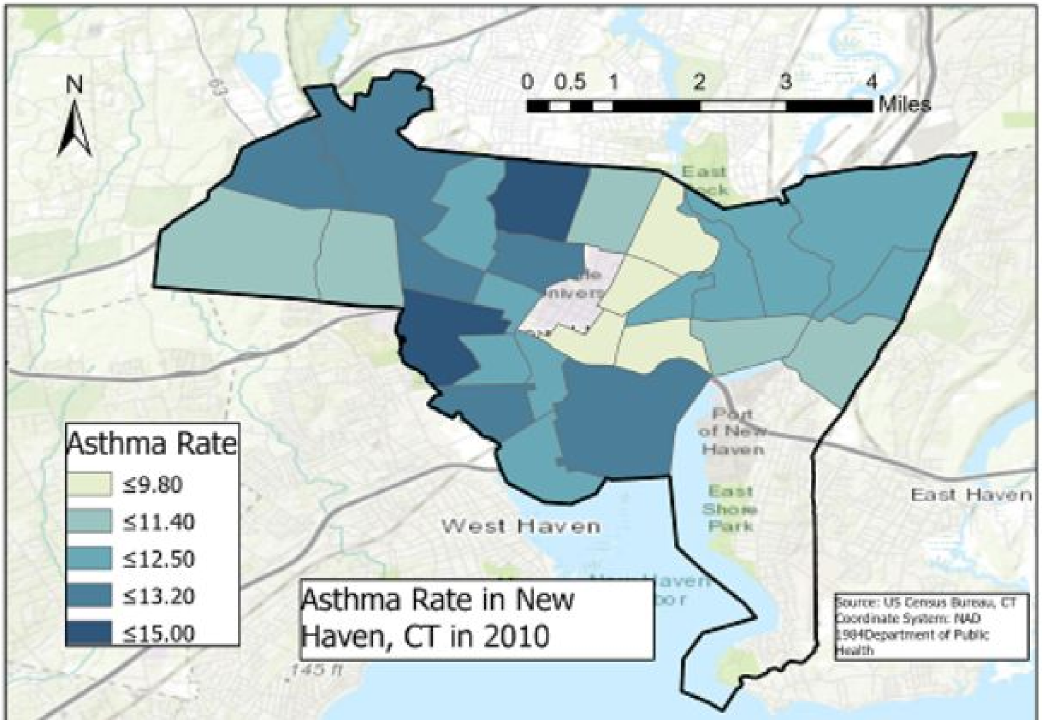
Map 3: Asthma Rate Hot Spots in 2019



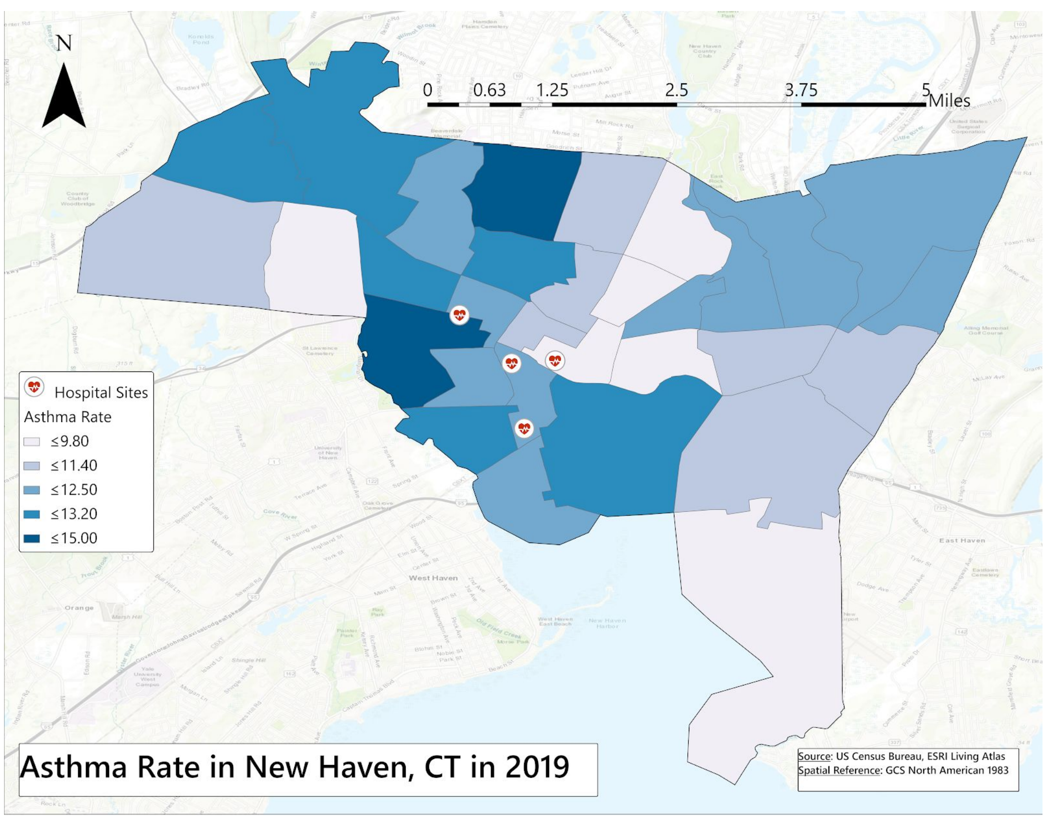
Map 4: Vulnerability Index Values 2010



Map 5: Vulnerability Index Values and Hospital Sites in 2019



Map 6: Asthma Crude Prevalence New Haven, CT 2010

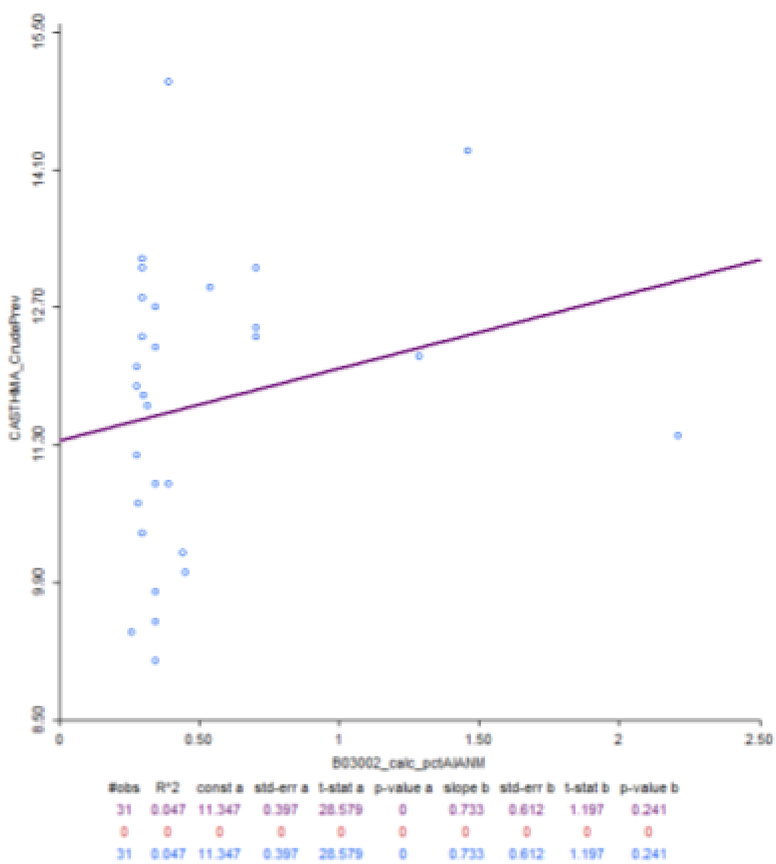


Map 7: Asthma Crude Prevalence in New Haven, CT 2019

Conclusion:

While we were looking to identify areas where people were more vulnerable to asthma based on the identified health determinants of race and household income as well as air quality, we found more evidence for links between the socioeconomic determinants and asthma than with air quality. We were able to identify census tracts that are lower income and higher percent black population and concluded that they experienced higher asthma crude prevalence in the city consistently between 2010 and 2019, a likely positive correlation across time periods that seemed to worsen in 2019. The lack of relationship between PM 2.5 levels and asthma rates in the 2010 study area disproved our hypothesis, but a slight increase in positive correlation from the 2019 data provoked new questions. The difference between the 2010 and 2019 data is most likely on account of an unknown intervening variable (e.g., climate change, increase in pollutants, industrialization, etc.) in the study area, which policymakers and future works could evaluate for useful information as to how and why asthma rates and air quality are worsening in New Haven.

Figures and Tables:



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
31	0.047	11.347	0.397	28.579	0	0.733	0.612	1.197	0.241
0	0	0	0	0	0	0	0	0	0
31	0.047	11.347	0.397	28.579	0	0.733	0.612	1.197	0.241

Figure 1: % Asian by Census Tract vs. Asthma Crude Prevalence

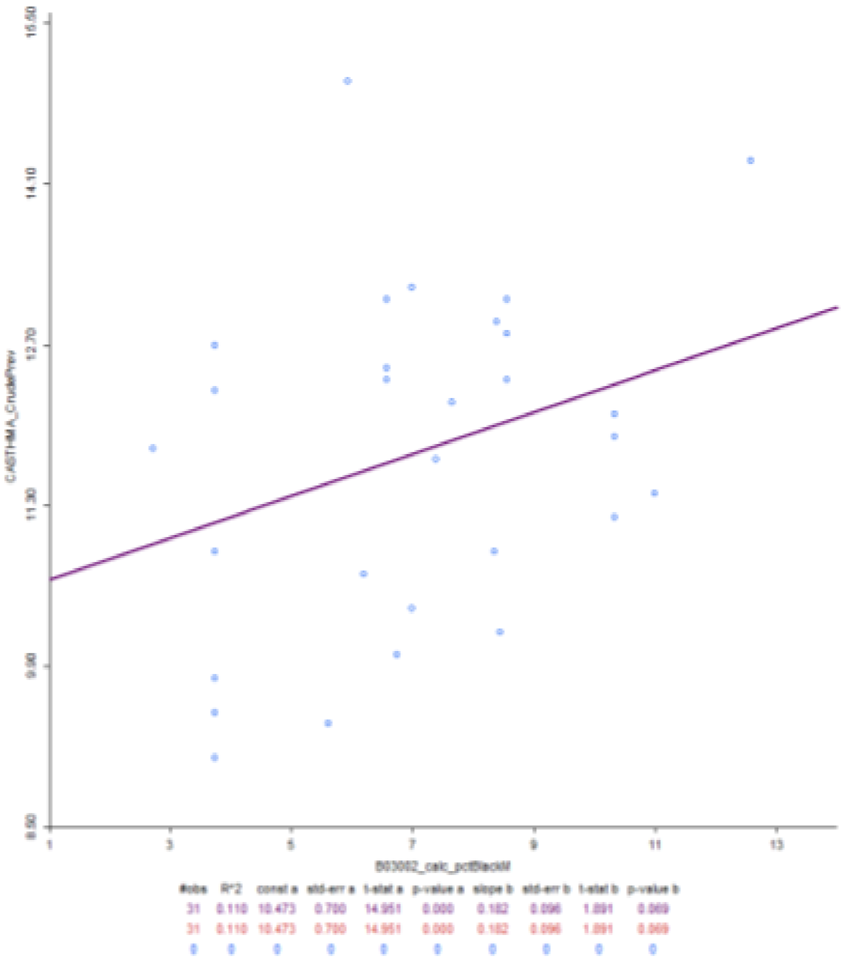
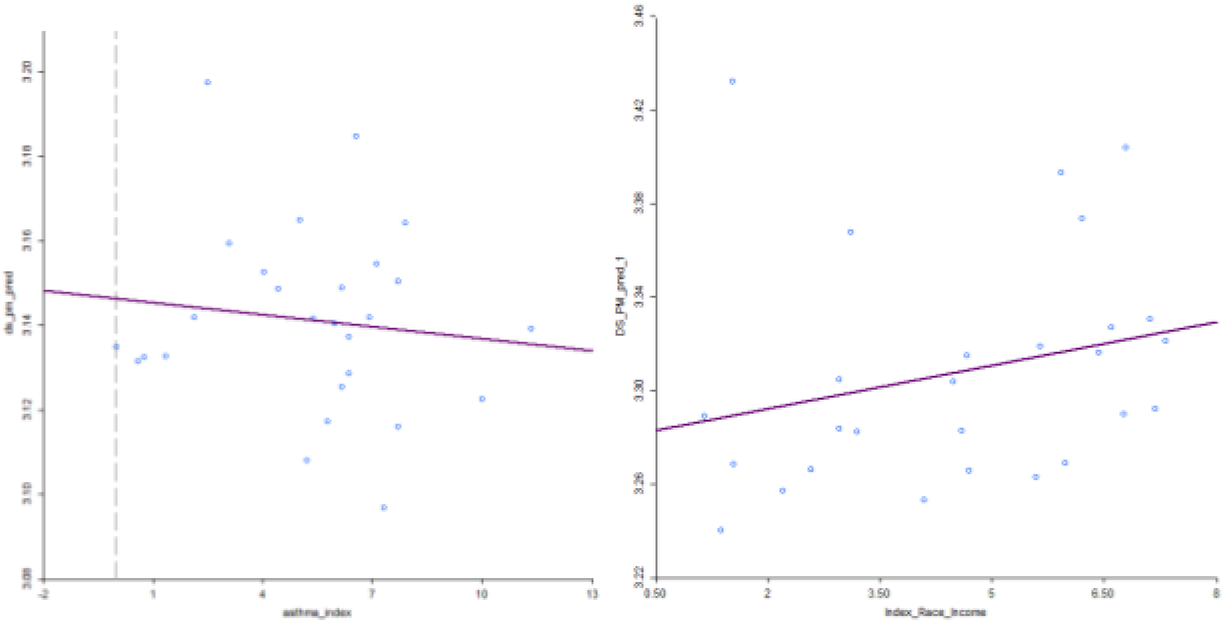


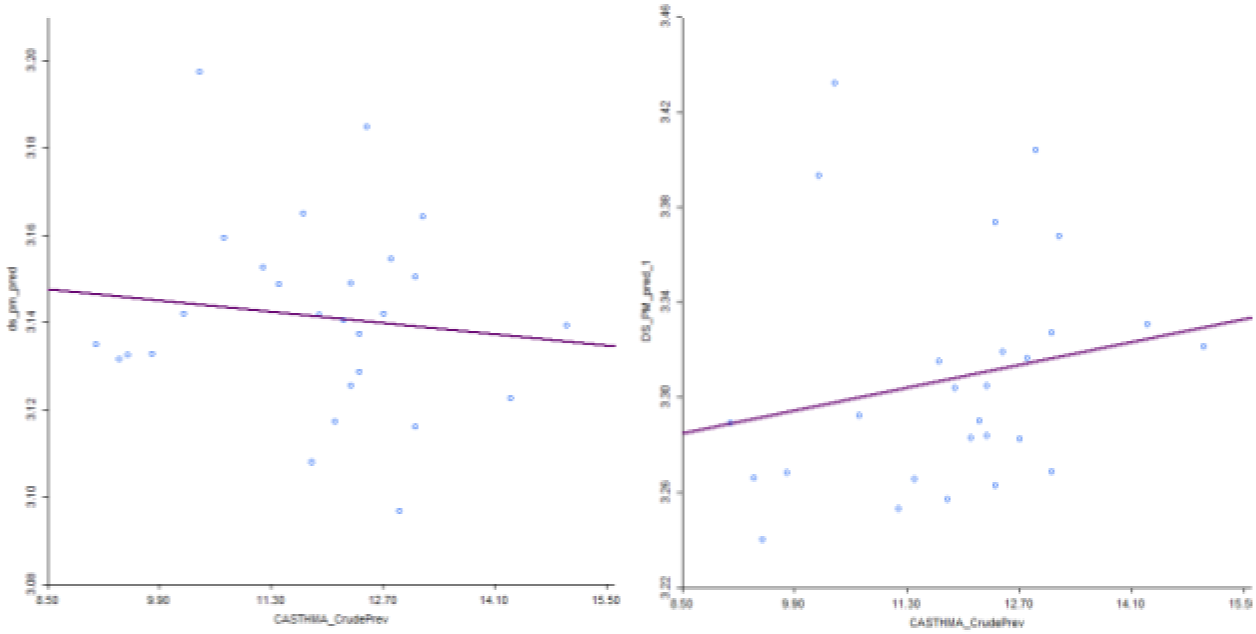
Figure 2: % Black Population by Census Tract vs. Asthma Crude Prevalence



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
27	0.015	3.146	0.009	340.705	0	-0.001	0.002	-0.624	0.539
0	0	0	0	0	0	0	0	0	0
27	0.015	3.146	0.009	340.705	0	-0.001	0.002	-0.624	0.539

#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
27	0.064	3.280	0.023	140.698	0	0.006	0.005	1.311	0.202
0	0	0	0	0	0	0	0	0	0
27	0.064	3.280	0.023	140.698	0	0.006	0.005	1.311	0.202

Figure 3: 2010 (Left) and 2019 (Right) Social Vulnerability Index vs Air Quality



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
27	0.015	3.163	0.035	89.284	0	-0.002	0.003	-0.624	0.539
0	0	0	0	0	0	0	0	0	0
27	0.015	3.163	0.035	89.284	0	-0.002	0.003	-0.624	0.539

#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
27	0.042	3.226	0.079	40.938	0	0.007	0.007	1.043	0.307
0	0	0	0	0	0	0	0	0	0
27	0.042	3.226	0.079	40.938	0	0.007	0.007	1.043	0.307

Figure 4: 2010 (Left) and 2019 (Right) Asthma Crude Prevalence by Census Tract vs. Air Quality

References:

Ankrah, J. (2019). Asthma in Greater New Haven:Local context and the role of walkability, bikeability, parks, and safety. *New Haven Green Fund*. Retrieved from http://www.gnhgreenfund.org/uploads/4/5/7/0/45701091/final_report_062819_data_haven.pdf

Department of Public Health. (2016, June 27). Combined Asthma ED and Hospitalizations by Census Tract, New Haven, Connecticut, 2010 - 2014. *Connecticut Asthma Program.* Retrieved from https://portal.ct.gov/-/media/Departments-and-Agencies/DPH/dph/hems/asthma/gis/NewHavenFinalpdf.pdf

Kolling, J., Strosnider, H. (2013). Social and environmental risk factors associated with county-level asthma emergency department visits. Center of Disease Control Agency for Toxic Substances and Disease Registry. Retrieved from https://svi.cdc.gov/Documents/Publications/CDC_ATSDR_SVI_Materials/APHAposterV7_TOPRINT.pdf

Lemke et. Al., (2013, November 13). Geospatial relationships of air pollution and acute asthma events across the Detroit–Windsor international border: Study design and preliminary results. *Journal of Exposure Science & Environmental Epidemiology.* Retrieved from https://www.nature.com/articles/jes201378

Tuluri, F., Gorai, A.K. (2015). Hotspot analysis for examining the association between spatial air pollutants and asthma in New York State, USA using kernel density estimation (KDE). *Journal of the Mississippi Academy of Sciences.* 60. 220. Retrieved from https://proceedings.esri.com/library/userconf/proc17/papers/2255_12.pdf

Weir, W. (2018, October 18). Report: CT Asthma Rate Higher Than US. *Hartford Courant*. Retrieved from https://www.courant.com/health/hc-xpm-2012-12-06-hc-asthma-in-ct-1207-20121206-story.html

Team Effort:

Both members worked together on the GIS Analysis components of the project. We divided the maps and analyses by the year (Jackie 2010, Ryan 2019). Ryan worked with GeoDa and performed the statistical analyses. Jackie attempted to use a Network Analysis to include hospital travel times in our project, but was unsuccessful and this portion had to be eliminated from the project. Both members worked together in writing the paper and composing the presentation.