City of Boston: Transit & Air Quality Team C

Team Members:

Matias Ou, Class of 2024, matiasou@bu.edu - Team Representative

Ziping (Peter) Wang, Class of 2024, pwang78@bu.edu

Yuchen Lu, Class of 2024, luyc@bu.edu

Eric Wang, Class of 2024, eawang@bu.edu

Zhiyi (Steve) Yu, Class of 2024, steveyu@bu.edu

Introduction

Project Goal and Overview

In the face of escalating environmental challenges and public health concerns, this project embarks on a critical investigation to unravel the intricate link between transportation infrastructure and air quality in Boston, Massachusetts. We aim to delve into the varied effects of transportation infrastructure across different neighborhoods, shedding light on how these elements intertwine to shape the air quality and, consequently, the quality of life in urban areas.

By meticulously analyzing how different neighborhoods bear disproportionate environmental burdens, this project aims to bring to the fore the socio-economic disparities that pervade our urban landscapes.The findings of this study are not just numbers and graphs; they are powerful tools for urban planners, environmental scientists, and policy-makers. With these insights, stakeholders are better equipped to make informed decisions, strategize effective interventions, and ultimately pioneer the transformation towards cleaner air and more equitable urban environments.

Potential for Long-Term Impact

The implications of this study are far-reaching. Improving air quality is not just about meeting environmental benchmarks; it's about fostering healthier communities, reducing the strain on our healthcare systems, and building sustainable cities that thrive both ecologically and socially. By addressing these issues, we contribute to a broader vision – a vision of urban spaces where quality of life is uplifted, and environmental justice is a reality, not just an aspiration.

Target Audience and Stakeholders

Our findings are crafted with urban planners, environmental scientists, and local policymakers in mind. We envision our study as a beacon guiding policy decisions and infrastructural changes. Through our research, we aspire to ignite a conversation about urban development, one that is grounded in scientific inquiry and geared towards tangible, positive change in our communities. This report begins with a comprehensive methodology, followed by a detailed analysis of our findings.

Base Project Analysis and Conclusions

Exploratory Data Analysis

We were provided with a number of valuable data sources to guide our investigation and begin our research. Our first tasks were to analyze these data sources and perform preliminary analysis on each of them to better understand the nature of the data we were collecting. The following is a brief summary of the primary data sources we used throughout the project:

AirNow API: Of the various public APIs we were provided in order to guide our exploration of Boston's air quality, we ultimately decided to primarily use the AirNow API, as it ended up being the most accessible and intuitive source to fetch data from. We collected hourly air quality data over the span of multiple years (2019-2023) before aggregating them to provide a measure of the overall trends in air quality over the past few years in Boston. Ultimately, we were able to gather information from three different sites in Boston for our project.

PurpleAir API: We also examined multiple other sources and APIs for air quality, including PurpleAir. However, there were multiple issues that ultimately prevented us from collecting useful data to analyze from these sources. For example, PurpleAir's air quality collection sites are privately owned, and are inaccessible for public use. Of the ones that are publicly available, the periods of time they were active for are very sporadic, unpredictable, and often located outside of Boston, leading us to eventually conclude that PurpleAir's air quality data was insufficiently complete to draw any meaningful conclusions from.

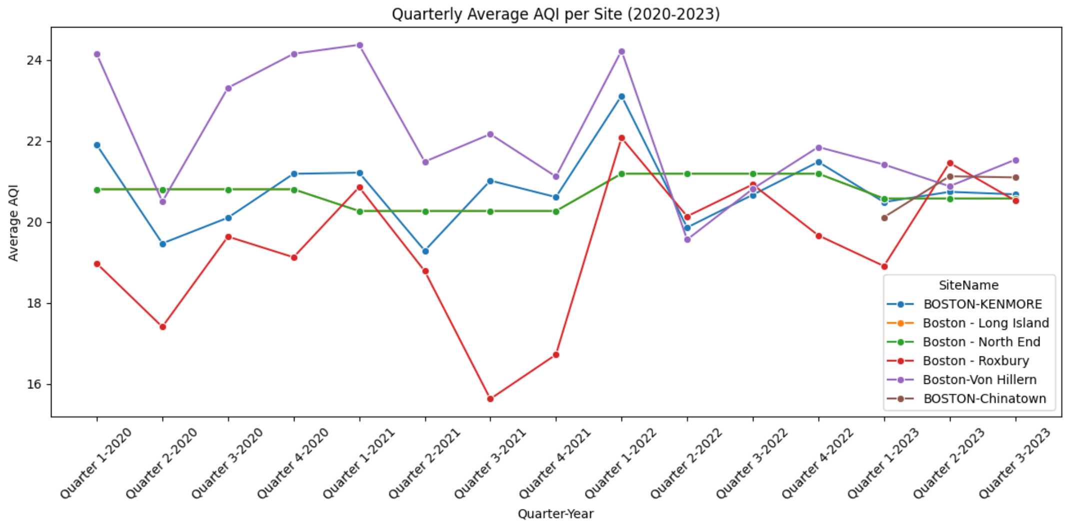
Social Vulnerability Index: We collected data from Climate Ready Boston, focusing on geographic and social vulnerability information for each location. The social vulnerability index is used to aggregate the proportion of Boston populations in various vulnerable demographics, including young children, older adults, people of color, people with limited English proficiency, and people with low yearly income. This data was used to calculate the percentage of the population in each socially vulnerable category and the housing and population densities. We aggregated this data to create a social vulnerability index for each neighborhood, applying equal weights to all categories.

Proximity to Roads: In a study of the MAPC region, researchers examined where residents live in relation to high-pollution roads, focusing on households within 150 meters and introduced the Pollution Proximity Index (PPI) that measures nearby traffic factors and proximity to pollution sources. We analyzed and visualized this PPI data for Boston, conducting a basic inspection, outputting the top five rows and summary statistics to a text output file.

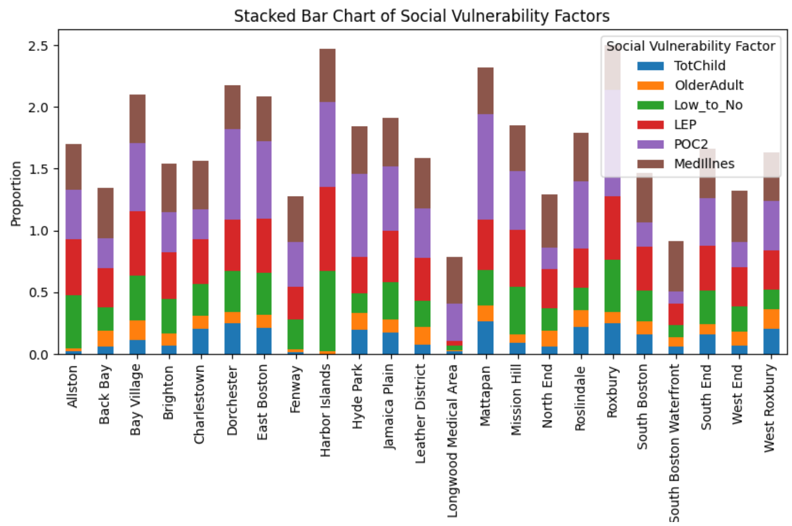
Census Demographics: We collected census data for all zip codes in Boston all the way back to 2011 from Census.gov and the American Community Survey, including gender, age, race, and citizenship status. We initially gathered these data in hopes to draw conclusions between areas with better air quality and poor quality based on these characteristics. However, due to issues with the AirNow API and AQI data collection, we were unable to fully utilize this data either. Rather in the extension section of the project, we concentrated on analyzing two neighborhoods in Boston: Kenmore and Roxbury.

Income Distribution: We gathered data on the median income and distribution of the population across various income brackets for all 23 Boston neighborhoods from BostonPlans.org and Census.gov for 2021. Our initial plan was to extend this data collection to 2020-2023 for a comprehensive analysis alongside the AQI data. However, due to issues with the AirNow API and changes in the project scope, this extension was not realized. Similarly to our analysis of the Census Demographics, we decided to concentrate on Kenmore and Roxbury in our results.

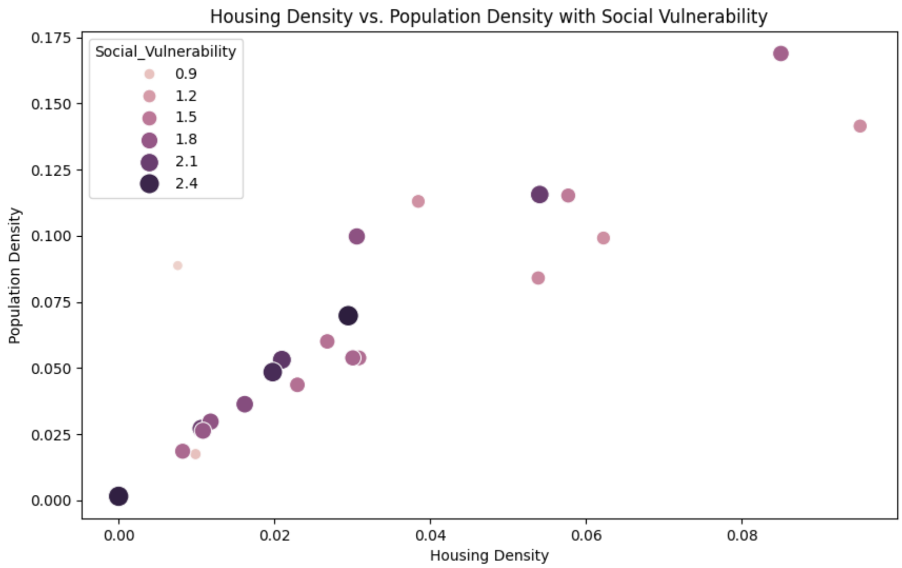
Data Visualization and Results



- The above graph depicts the average AQI for all available sites for each quarter from 2020 to 2023. However, there are some instances of missing records from the dataset. In particular, the orange line (Long Island) is missing because the site is inactive for all four years, and the site at Chinatown only began recording data at the start of 2023.
- Overall, the AQI values collected at North End are pretty steady in the past four years while the values collected at Von Hillern and Roxbury fluctuate a lot compared to the other sites. Areas such as Von Hillern have consistently worse air quality than other sites such as Roxbury, but the differences in AQI are not very large.
- Note that while the variations in the graph may appear drastic, the entire range of AQI only ranges from ~16 at best to ~24 at worst, which overall exhibits very small variance in air quality over these four years. From this, we can conclude that while there are some small differences in air quality from year to year, the overall air quality in Boston has not changed significantly over recent years, and has been consistently good.



- The graph above displays every neighborhood in Boston along with the proportion of the population in each of the six social vulnerability factors: young children, older adults, those with low to no income, those with limited English proficiency, people of color, and those with medical illnesses.
- In every neighborhood, POC2 (people of color) and MedIllness (people with medical illness) take about the same proportion of the index. In comparison to the total population overall, people of color statistically have lower income, which means a high proportion of POC2 is correlated with a high proportion of Low_to_No incomes.
- Individuals with lower income may be restricted to which neighborhoods they can choose. This means that people may be unable to afford housing in a more excellent neighborhood or air purifiers. Additionally, if the neighborhood has lousy air quality, people with medical issues will be more likely to suffer.



- The scatter plot above maps housing density against population density for neighborhoods in Boston, with the size of each data point reflecting the Social Vulnerability Index (SVI). A clear positive correlation is depicted between housing and population density, indicating that as the number of housing units per area increases, the population density tends to rise correspondingly.
- The SVI, represented by the varying sizes of the plot points, seems to initially increase with density, suggesting that denser areas may experience higher social vulnerability. However, the analysis indicates a turning point around (0.03, 0.075) on the graph where the SVI begins to decrease as the density continues to increase. This could suggest that beyond a certain density threshold, neighborhoods may see a decline in social vulnerability, potentially due to better access to resources, infrastructure, and community resilience measures.
- The presence of an outlier near the origin suggests that there are neighborhoods with low density but disproportionately high social vulnerability, an outlier that would need further investigation. The graph suggests a complex relationship between housing density, population density, and social vulnerability, with implications for urban development and social policy.

Conclusions

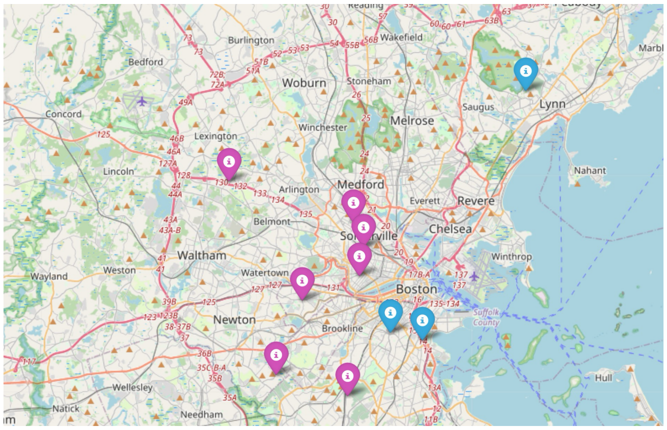
From our initial examination of the dataset, we were able to draw various conclusions and create multiple hypotheses based on our findings and the above analysis:

- **Higher population density strongly correlates with higher housing density**, and appears to be loosely correlated with a higher social vulnerability index. We predict that higher population/housing density will be correlated with worse air quality in general due to increased density causing worse traffic and congestion for all transportation systems.
- There appears to be **little to no correlation between social vulnerability and air quality of neighborhoods**. Neighborhoods with better air quality appear to have a greater variance in terms of social vulnerability, but beyond that it is hard to draw a strong conclusion between these two factors.
- Overall, **the air quality of Boston neighborhoods is consistently good**[1] from the years of 2020-2023. With the exception of Roxbury and Von Hillern, the air quality index of all recorded neighborhoods has remained very consistent with little variation.

Limitations

The overarching problem and limitation of the data collected from the AirNow API is the lack of consistency and granularity of the collection sites. When further investigating how the AQI data was collected, we realized that all the data was collected from only five different collections sites, some of which weren't within Boston or even consistently active. When accounting for these factors, we were left with only two data collection sites that were consistently collecting data within Boston–Kenmore and Roxbury. In addition, other collection services such as PurpleAir had similar issues with active consistency over a long period of time. While PurpleAir had many more collection sites than AirNow, a majority of them were privately owned, and the remaining sites were either located outside of Boston and/or were only active for a short amount of time, making them unusable for a long-term study.

[1] According to the United States Environmental Protection Agency (EPA) "Good" AQI is 0 - 50. Air quality is considered satisfactory, and air pollution poses little or no risk.



Purple Air API Collection Sites

AirNow API Collection Sites

Here is a mapping of the geolocations of both PurpleNow and AirNow collection sites.

Extension Project and Further Research

Extension Proposal and Direction

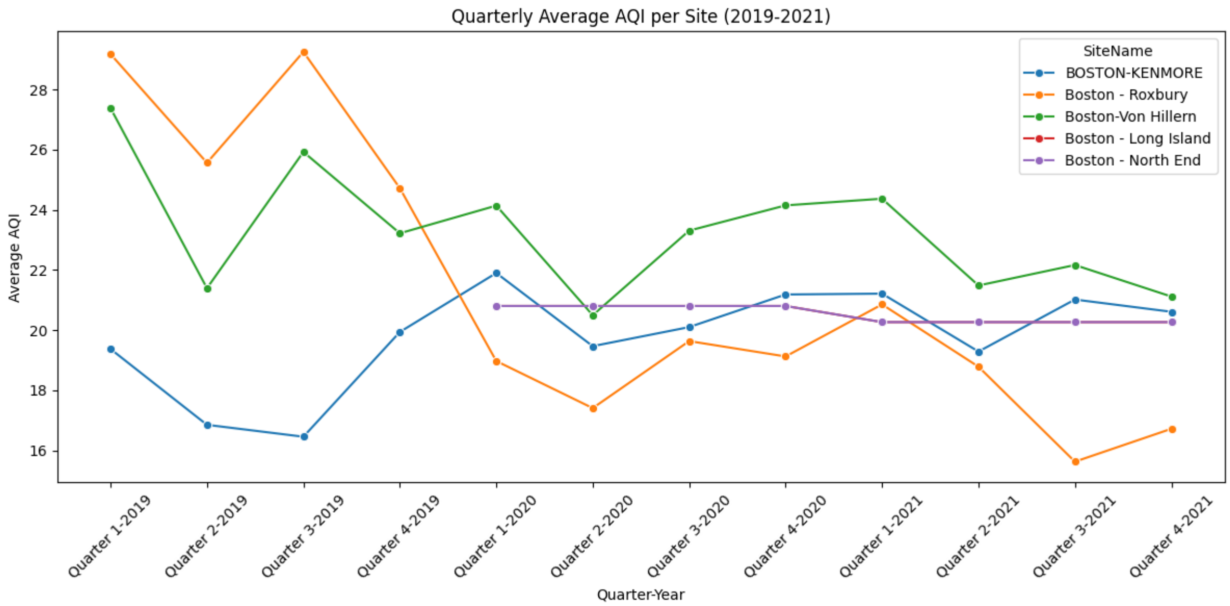
As a result of the aforementioned issues with gathering data on the air quality of Boston, we decided to shift our project's focus to areas near the sensors, concentrating our efforts on analyzing and addressing air quality concerns in immediate proximity to the monitoring devices. This approach involved a more localized and granular examination of air quality, allowing us to uncover nuances that may not be evident when looking at broader geographic areas.

By honing in on sensor proximity, we can:

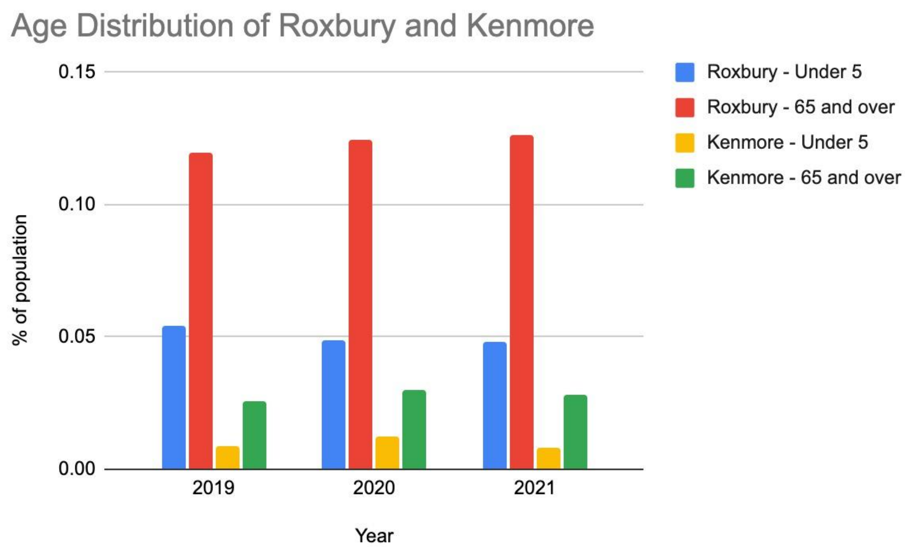
- Identify and target specific areas with air quality issues, potentially leading to more efficient interventions.
- Provide hyper-localized air quality information, which could be valuable for communities, businesses, and policymakers.
- Explore trends and patterns in air quality variations at a finer scale, offering a deeper understanding of environmental factors affecting specific regions. In particular, we examined the economic and demographic differences between the regions to examine if there was any correlation between air quality and these differences.

This modification resulted in a more targeted and actionable project, focusing on the intricacies of air quality within close reach of the sensors. In essence, we narrowed the scope of the project to focus on the specific differences in air quality between Kenomore and Roxbury.

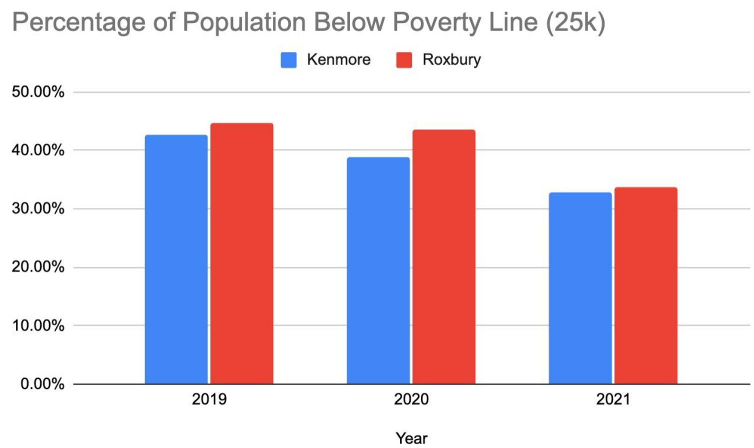
Data Visualization and Analysis



- This graph illustrates the change in AQI over time for every quarter year from 2019 to 2021. We notice that throughout 2019, Kenmore's AQI was lower than Roxbury's AQI (meaning that Kenmore had better air quality). However, for 2020 and onwards, Roxbury's AQI was better than Kenmore. Additionally, we observe that Roxbury's AQI exhibited far more variance over time than Kenmore.
- One possible explanation for this is that Roxbury contains many residential areas, while Kenmore has more commercial areas (including Boston University), so during the pandemic period of 2020 and 2021, the air quality of these two areas may be affected differently.

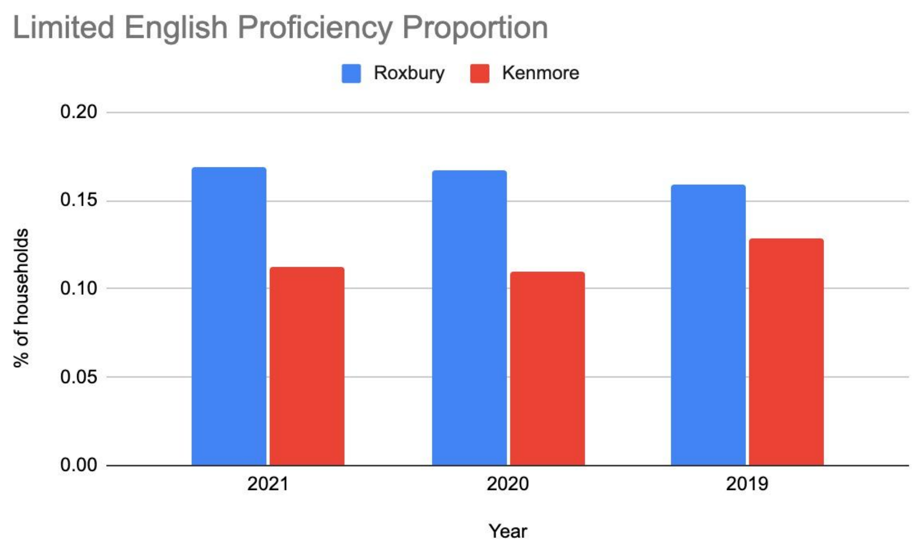


- The above graph provides a visual representation of the distribution of different age groups in Roxbury and Kenmore from 2019 to 2021. In particular, we compiled the age groups considered to be socially vulnerable from Climate Ready Boston: children under the age of 5 and adults over the age of 65.
- The age proportions for both areas seem relatively stable over the three years. Comparing the proportions between two areas, Roxbury has a higher percentage of both under 5 and 65 and over population than Kenmore, which is understandable given that Roxbury is a residential neighborhood in Boston, while a large proportion of Kenmore residents are Boston University students.
- A high aging population and a high proportion of under 5 population can relate to other social vulnerability factors such as low median income, high poverty rates, and high limited English proficiency rates.



- This graph visualizes the percentage of households below the **poverty line**[2] for Kenmore and Roxbury from 2019 to 2021. Both areas show a decreasing trend in the percentage of households below the poverty line over the three years. The rate of decrease for Kenmore was more rapid between 2019 and 2020, which has a steeper slope. Between 2020 and 2021, both areas show a continued decrease with a similar rate. However, during this three year period, the line for Roxbury remains above Kenmore, which indicates that Kenmore had a lower prevalence of households below the poverty line throughout this period.
- The percentage of the population under the poverty line appears very high at first glance, but considering that the Massachusetts minimum wage increased from $12.00 to $13.50 an hour over these three years, it is understandable that a very large percentage of the population would be below the poverty line, and that this proportion would gradually decrease from 2019 to 2021.

[2] Yearly income less than $25K



- This graph shows the proportion of households with limited English proficiency in Roxbury and Kenmore ranging from 2019 to 2021. The graph indicates that the proportion of limited English proficiency has remained stable in both areas across the three years.
- During this period, Roxbury has a higher percentage of limited English proficiency than Kenmore. Combining these three graphs, it seems that areas with better English proficiency are correlated with higher median income and lower poverty rates.

<u>Conclusions</u>

By limiting the scope of the extension project to better focus on Kenmore and Roxbury, we are able to draw a few more significant conclusions from our data analysis:

- **In 2019, Kenmore's air quality was better than Roxbury's, but this changed from 2020 onwards**. This is likely due to the effects of the COVID-19 pandemic and the resulting quarantine and other measures taken to isolate and keep the population safe. The overall transportation and traffic present in Roxbury, a residential area, may have been more heavily impacted by the quarantine than Kenmore, a more commercial area.
- Overall, there appears to be **little to no correlation between the social vulnerability of an area and its air quality** from what we were able to observe. Of the two neighborhoods we examined, Kenmore and Roxbury, many social vulnerability metrics remained very consistent throughout the observed time period, while the air quality of these two areas fluctuated much more than can be expected just from social vulnerability.

Limitations

Similarly to our base project, we encountered issues with gathering enough usable data to analyze, both with air quality data and census data.

- Moving forward we decided to solely focus on the years 2019-2021, as the Census data for 2022 was not broken down into individual zip codes as previous years were, and the 2023 data has yet to be collected.
- Additionally, certain aspects of social vulnerability, such as the proportion of the population with disabilities or medical illnesses (including heart disease, respiratory issues, cancer, etc.), were lacking sufficient data of a fine enough granularity to be used for our analysis.

In addition, even while focusing on just the neighborhoods of Roxbury and Kenmore, this presents a new limitation in that these neighborhoods are very similar to each other, both demographically, geographically, and in terms of air quality. As such, analyzing the differences between these neighborhoods is extremely difficult, and finding any correlation between social vulnerability and air quality was inconclusive.

Future Project Questions

While the current project as it stands is overall inconclusive with the current data, there are still potential areas to explore further in a future project which may prove to be fruitful and provide more conclusive results.

- *What areas/neighborhoods of Boston would be best to compare if the underlying air quality data was present?*

Due to our limited air quality data, we were required to compare Roxbury and Kenmore, two very similar neighborhoods with similar demographics and air quality. If we did not have this restriction, we may have been able to compare two very different neighborhoods, which may have provided more insight into the possible relationship between social vulnerability and air quality.

- *What other factors may be related to air quality?*

Our investigation focused primarily on social vulnerability and other demographic factors, but the data we have is too limited to conclusively establish a relationship between social vulnerability and air quality. Instead, another potential avenue of research could lie in a different aspect, such as the density of different public transportation systems (bus, trains, subway), private transportation systems (cars), and other personal transportation systems (walking, bikes).

- *Longer period of time to research? (2016-2018 vs 2019-2021)*

The timescale we focused on encompassed 2019-2021, which included the COVID-19 pandemic and the ensuing transportation restrictions. This obviously had a very large impact on the overall movement of the population and the resulting air quality from transportation systems. To get a better understanding of the overall impact of COVID-19 travel restrictions, it would be extremely valuable to examine the air quality and trends of years before 2019, and compare them with our findings after 2019 to understand the bigger picture of how air quality fluctuates over a longer time frame.

Individual Contributions

a. Matias Ou: For this project, I helped fetch the Air Quality index hourly data from 2019-2023 using the bash script written by our team investing over 50 hours in this task. I also collected the data from the Proximity to Roads paper and wrote code to analyze it by writing code to find the median, average, and standard deviation. I also wrote code in order to visualize the data in the form of graphs. Additionally I wrote part of the report for Deliverable 1. I then developed code to analyze and visualize the yearly change in the Air Quality from the different collection sites. Eric and I collaborated on crafting most of checkpoint A. I led discussions on data limitations and challenges, addressing core and extension questions. I spearheaded the changes that needed to be made for the new overall scope of the Air Quality and Transit project. I explored and analyzed new datasets, seeking ways to adapt our project for the semester's remainder. I actively communicated updates and outcomes via Slack to TAs and other teams working on Transit and Air Quality. I also collected the necessary data for the creation of the social vulnerability index for both Kenmore and Roxbury. Furthermore, I contributed significantly to the Deliverable 2 and 4 report, summarizing our progress in addressing project questions, adapting to scope changes, outlining future structural needs, and coming to conclusion.

b. Yuchen Lu: For this project, I contributed to data collecting, cleaning, processing and analyzing. For data collection, I wrote bash scripts to automate the data fetching process to minimize the time as much as possible. Using the data from the project manager and teammate, I explored the different datasets, cleaned unnecessary columns and did feature engineering to create columns as needed. For example, I created a social vulnerability index by assigning different weights to all factors of social vulnerability so that we can better analyze the data with air quality. For air quality related data, since we were initially focusing on comparing across different neighborhoods, data formats were modified to suit the needs. I created plots for easier data analysis. For extended project, I explored the new data sources (e.g. PurpleAir, Google air quality and CDC health data) and found that Google source does seem to be a good fit for our purpose but according to the documentation, it may not include the data that is older than 30 days. I then worked on reshaping the format of 2019 AQI data to be the same as the other years, grouping hourly data for 2019-2021 by quarters and sites, to create necessary plots for analysis. I also explored median income and poverty distribution data to create graphs for cross-site comparison.

c. Eric Wang: For this project, I contributed by processing and analyzing much of the AirNow air quality index data, census data, and demographic information. I created a Jupyter notebook to convert all the data into a consistent format to make it easier to process and create visualizations for analysis for Deliverable 1. For Deliverable 2, I participated in the ongoing discussions regarding the Transit + Air Quality extension project and the limitations of the current datasets in answering the base/extension questions, and helped develop our new extension project and helped analyze demographic data and created many of the graphs necessary for the extension. For Deliverables 3 and onwards, I helped collect and analyze the census data needed for our analysis of Kenmore and Roxbury, specifically the age distribution and the limited English proficiency proportion of the populations. I then helped compile the semester's findings into the Deliverable 4 report.
d. Peter Wang: For this project, I mainly focused on data collecting. I helped fetch the air quality index data from AirNow and conducted the necessary analysis for Deliverable 1. I then worked on fetching the MBTA transit data from ArcGIS. However, I was not able to extract the data due to lack of permission. For Deliverable 2, I helped formulate our extension project and new direction. Then I worked on extracting data from google maps api and other credible sources including aqicn.org, mapc.org and other cities. Then, I fetched census data census.gov to find new and usable data relating to our extension project. For Deliverable 4, I worked on analyzing the census data from census.gov. They were fetched earlier but were not included in previous deliverables. For each report, I worked on analyzing graphs and coming to conclusions related to our base and extension project. Lastly, I worked with the rest of the group to complete presentations needed for each deliverable.
e. Steve Yu: For this project, I fetched the Air Quality Index data using a bash script. I conducted exploratory data analysis, particularly focusing on the relationship between the Air Quality Group and the social vulnerability score. Given the limitations of our current data, we were unable to fully address the questions, but I also contributed to formulating

the hypothesis for the questions. I worked on data and graph analysis first. I prepared and added slides for the initial insights presentation. Additionally, I engaged in identifying and addressing limitations within our current datasets. Although I explored the CDC's Local Data for Better Health datasets to establish a correlation between air quality and health conditions, I found the relationship was not strongly supported by the data. I also participated in discussions about the extension project. Lastly, I collaborated with the rest of the team to complete and finalize the Deliverable 2 report. In the project's extension phase, I conducted a graphical analysis of two locations, Roxbury and Kenmore, focusing on several socio-economic indicators: the percentage of the population living below the poverty line, median income levels, the proportion of residents with limited English proficiency, and the age distribution in both Roxbury and Kenmore.