CAS CS 506 Final Report - Boston University

Livable Streets: Bike Infrastructure

Team C

William Huang, CAS '25 - wh@bu.edu
Jacob Stein, CAS '25 - jmstein@bu.edu
Marco Chin Fung Lam, CAS '24 - lcfmarco@bu.edu
Nyx Zhao - CAS '24 - zhaolili@bu.edu

I. Project Introduction

LivableStreets Alliance is a non-profit organization based in the Metro Boston area that seeks to transform streets into safe and accessible public spaces connecting people and the places they frequent. The group explores the effectiveness of different transportation solutions, with an emphasis on convenience and accessibility. These solutions often manifest themselves as bike lanes, open spaces, and other public amenities that promote sustainability in urban areas. Our group was tasked with **quantifying the extent of displacement inflicted on local communities by new and existing transportation infrastructure.**

The intersection of displacement and transportation infrastructure is a complex issue influenced by socio-economic factors and urban planning (or a lack thereof). Several metrics, including rent prices, property values, and homeownership rates, help quantify displacement, though establishing a direct causal relationship with public infrastructure is challenging. For instance, while greenways are often associated with gentrification and displacement, the extent of their direct impact on rent prices in Boston is difficult to isolate. Throughout the semester, we have attempted to circumvent this challenge by establishing as many correlative relationships as possible between transportation infrastructure and population displacement that may provide valuable insights for our client.

We approached this project by broadly exploring various transportation methods and displacement metrics, guided by our interests, and gradually narrowed our focus based on iterative analysis. We isolated trains, buses, bikes, and income fluctuations as our four original distinct indicators or precipitators of displacement. These topics then transformed into more targeted areas, like bike lane connectivity and the Green Line Extension project in response to our client's need for pertinent, current data. We then found and collected relevant data, and used processing and analysis techniques from class to derive applicable findings and convert them to a helpful format for the client. The following sections of this paper delve more deeply into our focus areas and extension projects.

The findings of our project and related future work can shed important insights into the relationship between public infrastructure and displacement. By establishing a more clear understanding of how transportation development affects urban communities, **LivableStreets can inform and influence public policies that mitigate or potentially exacerbate displacement**. Future urban planning projects— specifically in vulnerable neighborhoods like Roxbury, Dorchester, and Mattapan— can conscientiously mitigate displacement and set a high standard for sustainable transportation developments. Beyond the infrastructure developers, this project can also educate the public on the tangible effects of transportation developments, which may shift what communities and voters choose to focus on in their advocacy. The ultimate goal would be a clear cut set of policies that reconcile the need for sustainable public infrastructure while mitigating the resultant displacement. These standards would then hopefully be commonplace among development projects in Boston and beyond.

II. Our Data Source - The American Community Survey

Finding correlations at the intersection of public infrastructure and displacement requires highly granular data to ensure the findings are precise and relevant. The City of Boston and the Massachusetts Bay Transportation Authority provide extensive data on the operations of Boston's public infrastructure. Analyzing displacement on a nuanced level, however, also requires highly specific information on individual neighborhoods or demographic groups. For example, determining the impact of a new bike lane or train line on local housing prices requires geographically granular data over a certain period of time. Furthermore, we need to be able to isolate the effect of individual variables, despite exogenous factors that can influence relevant metrics, such as simultaneous infrastructure developments.

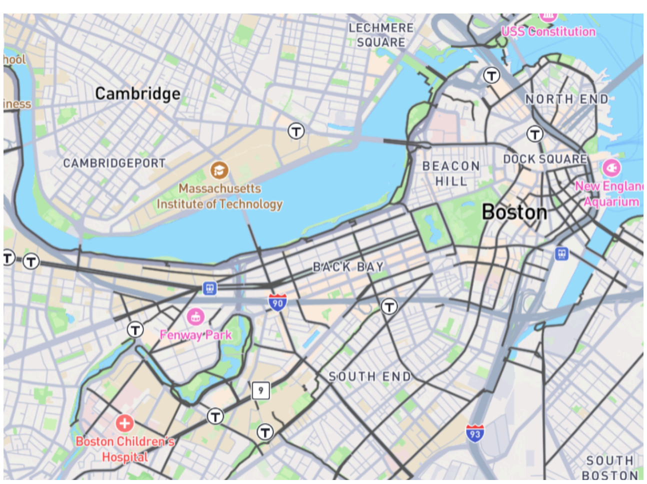
Few organizations beyond the U.S. Census Bureau have the capacity to collect data on a street-by-street basis. The American Community Survey is annually conducted to gather data previously only available in the decennial census. The Census Bureau aggregates their data to produce ACS 5-year estimates, which provide granular data down to a census-tract level. Whereas the decennial census is more focused on counting every individual, the ACS seeks metrics on how people live— such as homeownership rates, rent prices, and transportation methods— organized by distinguishing characteristics, like socioeconomic status and race. The greatest advantage of the ACS is the ability to analyze individual census tracts, rather than having to extrapolate relevant information from larger datasets.

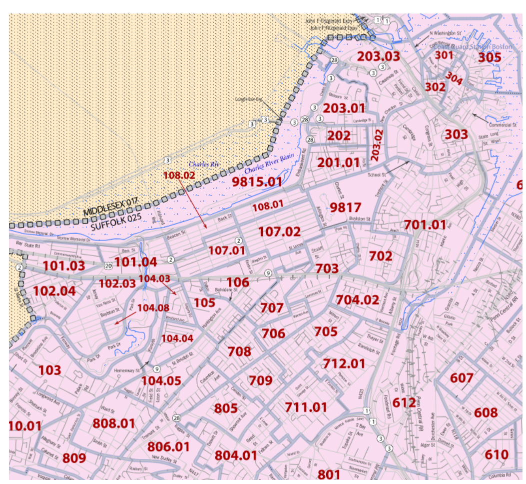
The ACS survey was a clear choice for our source of data for several reasons. By providing data at various geographic levels, we can extensively analyze the impact of transportation projects within or beyond a localized level. The ACS also contains incredibly relevant surveys like DP04 (Selected Housing Characteristics) and SO804 (Means of Transportation to Work). The availability of annual survey data, as opposed to the decennial census, also enables us to track changes in metrics over time. This is important because displacement is oftentimes a gradual process— so understanding its progression requires chronologically linear data. Lastly, the ACS's breadth of demographic detail enables us to understand which groups are most impacted by displacement, which is pivotal knowledge in developing mitigation strategies. Please see our Jupyter Notebooks for more information on the specific surveys we used.

III. Original Findings - Bike Lane Connectivity

One of the facets of displacement we focused our analysis on was the relationship between it and the placement of bike lanes – more specifically, seeing if bike lanes that were "more connected" to the Boston bike network had a more significant impact on displacement than more "isolated" bike lanes. Coming into the project, we knew that the installation of bike lanes would raise bike usage for commuters in the surrounding areas, but not all bike lanes are equal – some connect to downtown Boston, or connect to other bike lanes with no gaps in infrastructure, or serve for a longer distance than others.

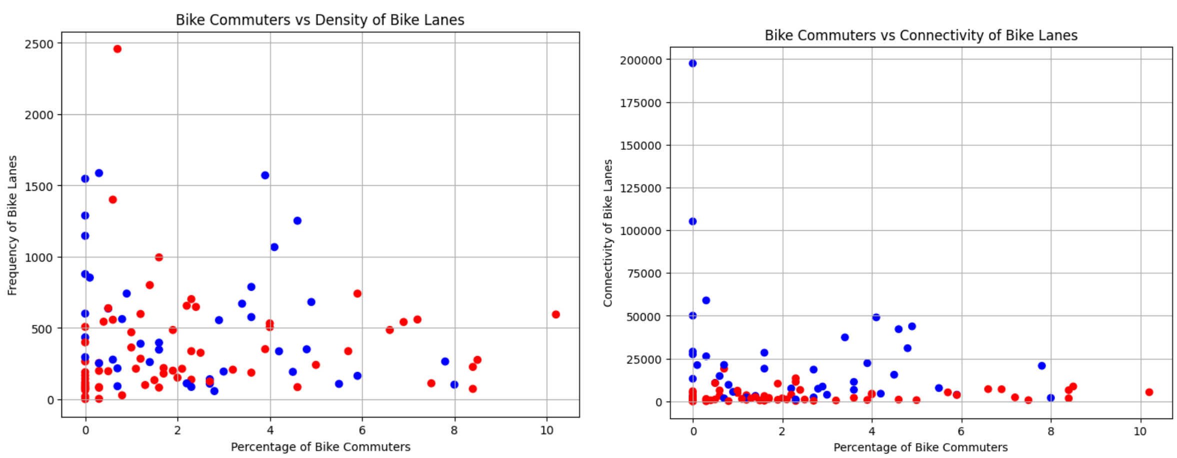
Thus, to test the effectiveness of bike lanes in different levels of connectivity, we used data collected from the ACS survey (as mentioned earlier) of individuals living in **Suffolk County**, sectioned by **census tract** to track the relationship between **bike usage** and **bike lane connectivity**. We used **bike lane density** as a form of comparison against bike usage. Bike lane data was taken from the [Existing Bike Network 2023 Dataset] from Analyze Boston, in which we then used the [Geocoder API] from the US government to find the census tract each individual bike lane belonged to.





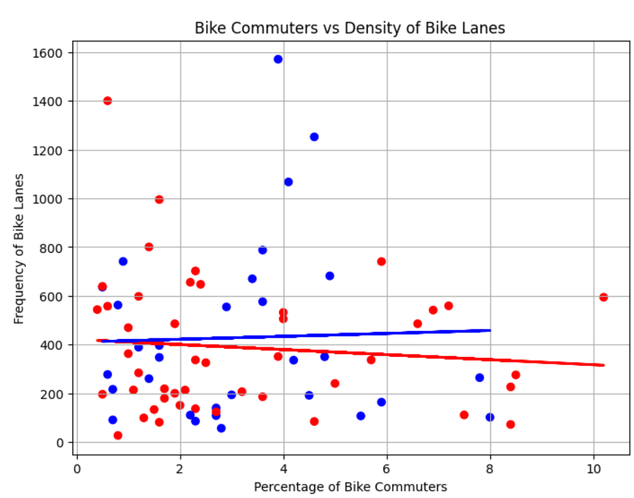
We then take the newly made dataset of bike lanes in each census tract and calculate the **density of bike lanes** and **connectivity of bike lanes** to store in two separate datasets to test. Bike lane density is calculated by dividing the **sum of lengths of all bike lanes** in a census tract by the **landmass** of the census tract. Bike lane connectivity is calculated by **adding up all density values** of each census tract that **each bike lane in a given census tract passes through**. The resulting value doesn't have a unit, but is a useful heuristic to compare how well connected a census tract's bike lanes are relative to other bike lanes in Suffolk County. We are now left with datasets where each census tract is assigned a density and a connectivity value.

We also took the population density of each census tract and classified each census tract into two categories: above and below the average population density, which allowed us to observe further trends based on factors outside just the connectivity and density of bike lanes within a census tract. Population data was also taken from the ACS survey. Finally, we land on the following scatterplots:

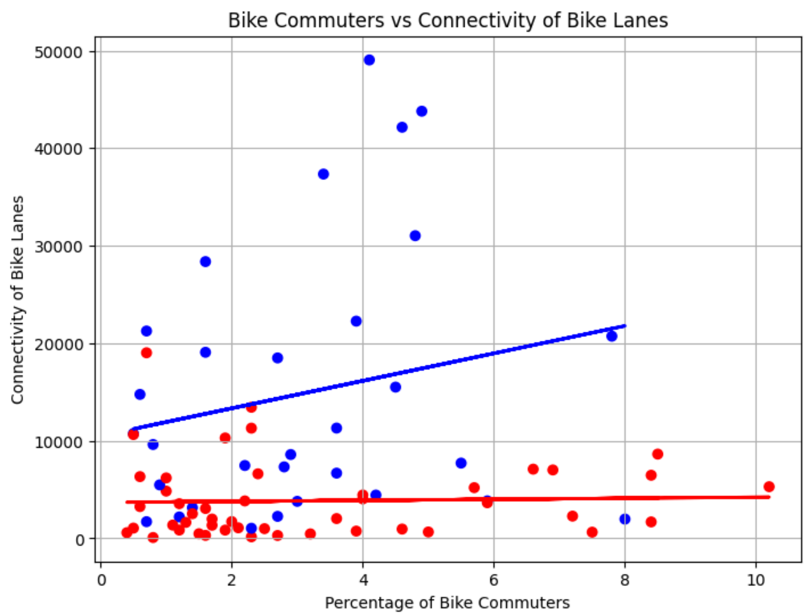


X Axis = Percentage of Commuters Who Bike. Y Axis = Density, Connectivity of Census Tract. A point represents a census tract; red = lower than average population density, blue = opposite.

The scatterplots initially showed no trends due to a number of outliers and a large number of census tracts having 0% bike commuters, which can be due to a number of factors – inadequate survey response numbers, tract not consisting of land (watermass), etc. To account for this, we gave a threshold of 0.3% bike usage for a census tract to be considered in the scatterplot, which resulted in the graphs and trends



shown above. For lane density, the tracts have too much variance to draw a concrete conclusion – the relationship between bike lane density and bike usage is weak. Perhaps other features like the quality of bike lanes (protected lanes vs. shared lanes) or location of census tracts could have marked a more noticeable trend.



On the other hand, there is a noticeable trend in bike lane connectivity with bike usage – especially in census tracts with above average population density. This verifies the hypothesis that we suspected in the first place. However, a more interesting trend popped up here: the fact that in less dense census tracts, there is almost no correlation between connectivity and bike commuter percentage. This could be for a few reasons, like less dense tracts being generally further away from city centers and thus being too far for cycling despite good connectivity, or the populations being so committed to cycling that they would cycle regardless of infrastructure quality.

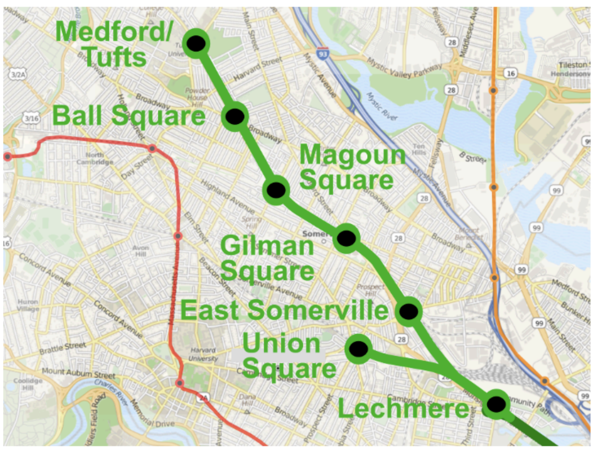
There are a few limitations to our testing here, most notable being the fact that the data is blind to factors outside of population density of census tracts, such as distance from city centers, location of tract in general, and the demographics of the people who live in the tracts (income probably plays a role in bike usage!). Additionally, this research was conducted within Suffolk County, and results may be different in more suburban or car-centric areas of Greater Boston.

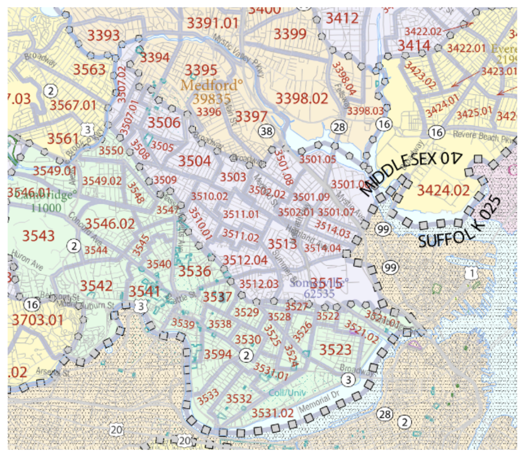
In the end, though, we still found a very clear correlation between bike lane connectivity and bike usage, which shows the importance of planning and establishing a proper, connected bike network. Just placing bike lanes in areas that currently lack them could result in less than stellar bike usage rates if said bike lanes are not connected to the places people need to go, or at least to other bike lanes that go to places people need to go – it is the existence of and connectivity to a bike network that will ultimately raise cycle rates.

IV. Original Findings - Green Line Extensions and Displacement

Another component of our based analysis centered around displacement caused by the Green Line Extension (GLX) project. Since 2005, the MBTA had been planning the project to extend the Green Line into Somerville and Medford. The MBTA has estimated that the line will support 45,000 one-way trips by 2030. The line has been intermittently under construction since 2018, finally being completed in December 2022. This makes the stops and lines along the extension the most recent additions to be added to Boston's train infrastructure.

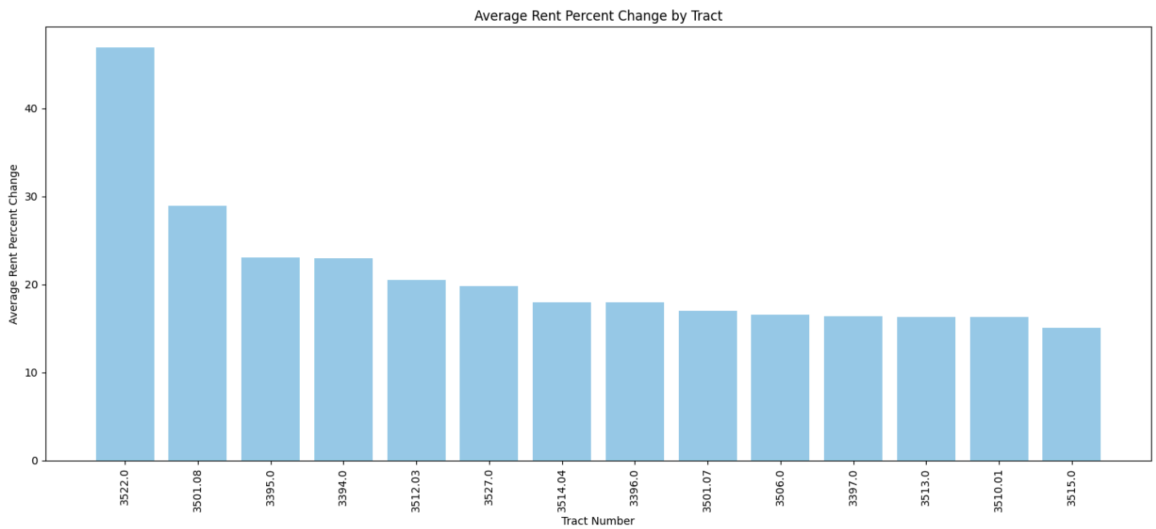
Observing displacement over time caused by the GLX project requires data specifically on the census tracts surrounding the extension project. We once again used the ACS estimates in the tracts within which the project was built, seen in the side-by-side comparison below:





We manually isolated the tracts, and used the filters on the ACS website to retrieve data from surveys SO802 (Means of Transportation to Work), SO601 (General Demographic Characteristics of the Population), and DP04 (Selected Housing Characteristics). The ACS only offers CSVs by survey and by year, meaning we needed three surveys across five years, resulting in fifteen total surveys. We then combined the annual surveys into one dataset per survey by grouping tracts in ascending order of year. See an example of the data below:

We immediately began analysis by looking for changes in rent, property value, and mortgage medians— key indicators of socioeconomic displacement— during the construction of the GLX. The graph below shows the average percentage increase in median rent prices between 2018 and 2022 by tract. Property value and mortgage medians followed similar trends:



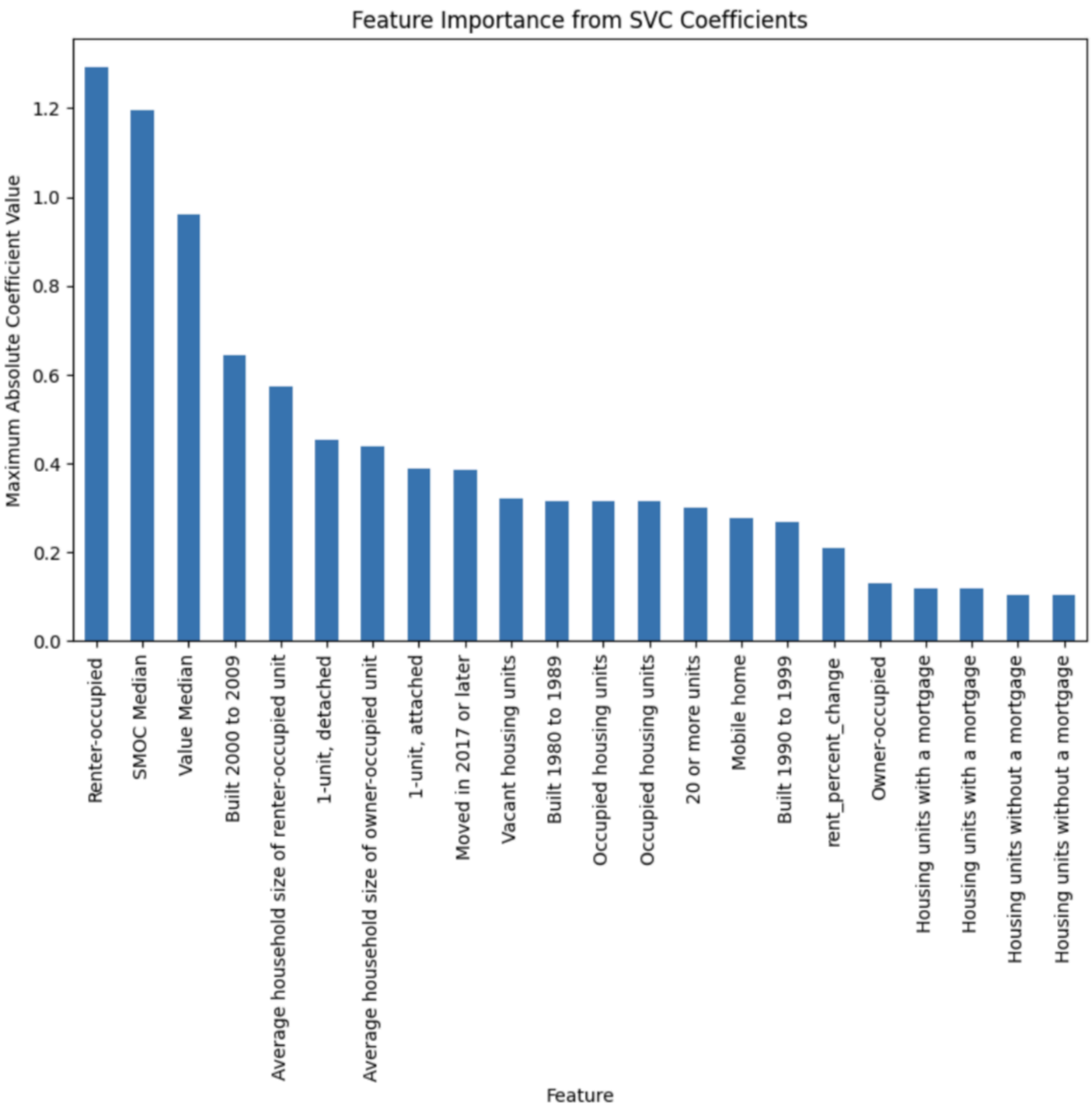
This simple analysis makes it immediately apparent that the GLX project resulted in egregious increases in property value and rent medians, making it very difficult for low-income households to sustain their residencies. With this in mind, we proceeded with more complex analysis using a Support Vector Machine to develop a notion of rent affordability.

We first created a weighted affordability score for each tract and year as a function of the percentage of income spent on rent. Next, we set thresholds to categorize tracts as a high, medium or low affordability score. We then standardized the data using a standard scaler, and trained an SVM using a linear kernel function on 80% of the source data. As the classification report shows, the model exhibited high precision and recall for the high and low affordability classes, yet struggled with the middle affordability class. Overall, the model exhibited a high degree of effectiveness in being able to make these classifications, which may inform the practicality of new housing units being built in these tracts.

	precision	recall	f1-score
High Affordability	1.00	1.00	1.00
Low Affordability	0.86	1.00	0.92
Medium Affordability	1.00	0.78	0.88
accuracy			0.91
macro avg	0.95	0.93	0.93
weighted avg	0.93	0.91	0.91

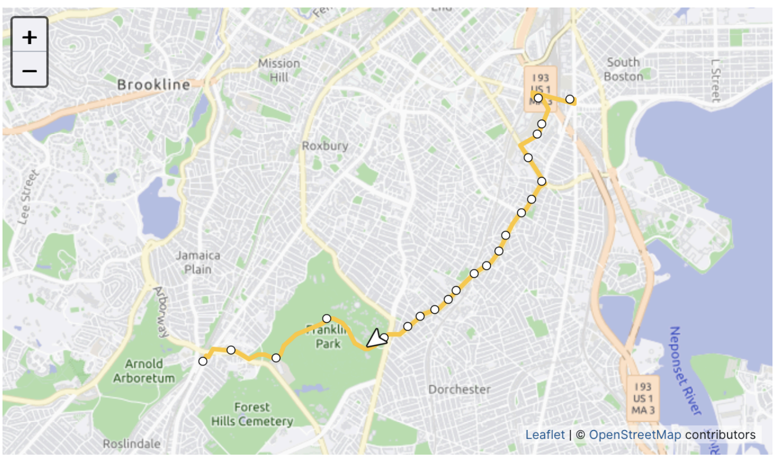
To derive more

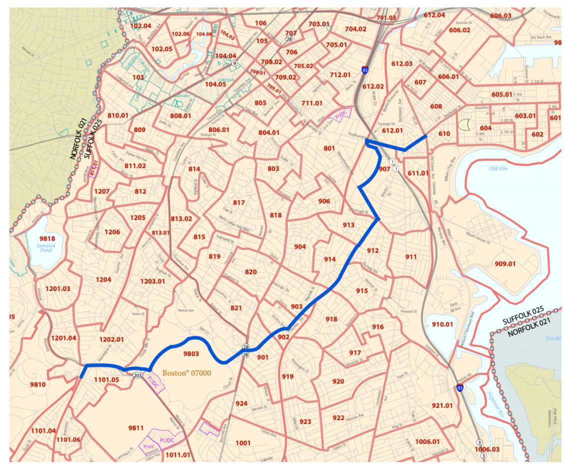
applicable insights from the SVM, we observed the feature importance graph produced by the model. Sorted in descending order, the graph shows that certain features— like the SMOC median and age of the building— are more important in determining rent affordability. The importance of certain features can inform policymakers and developers as they continue building infrastructure in Somerville and Medford. For example, local policymakers may introduce a rent cap to ensure residents living in older units are not displaced by the large increase in median rent prices.



V. Extension Project - Bus Correlations

Another component we explored for our extension project is possible bus correlations with displacement. We began this extension project with a very vague and general direction, trying to look at all MBTA bus routes in Boston and seeing if we could find any relationship of the distribution of ridership in terms of load, boardings, and alightings across Boston neighborhoods with displacement (please see mid-semester bus analysis in GitHub repository for more or the mid-semester presentation for a brief overview). However, after the mid-semester presentation, we decided to focus specifically on one of the Columbia Road bus routes, route 16, per the client's request.

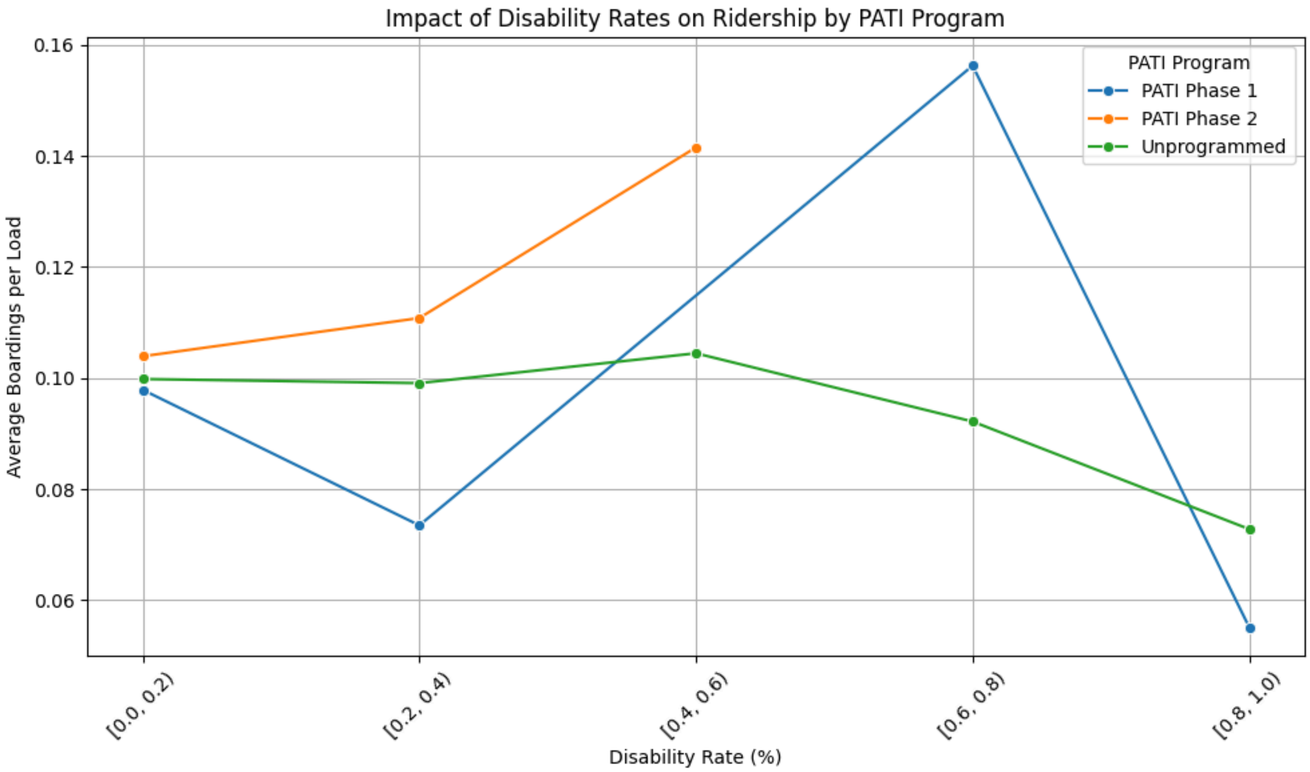




Source: Route 16

Route 16 goes from Andrew Station or Harbor Point in South Boston, through Columbia Road in Dorchester, Franklin Park, to Forest Hills Station. We once again used the ACS estimates in the tracts surrounding and within the path of route 16, seen in the picture to the right. Like with the GLX project, we manually isolated the tracts, and used the filters on the ACS website to retrieve data from the survey S1810: Disability Characteristics for those particular tracts for the year 2022. Additionally, we used the 2023 Census Tract TIGER/Line Shapefile as well as MBTA data on the Plan for Accessible Transit Infrastructure (PATI) bus stops and ridership characteristics.

Although we began with a general focus on route 16, the focus of this analysis shifted into seeing if there existed correlations between accessibility and displacement. We started with examining the distribution of bus stops among different PATI programs on route 16, which demonstrated that the PATI program is still in the works as the majority of stops are unprogrammed, with only 6 unique stops on route 16 being part of the PATI program phases 1 or 2. Next, we want to see how this unproportional distribution affects ridership and areas with varying disability rates.

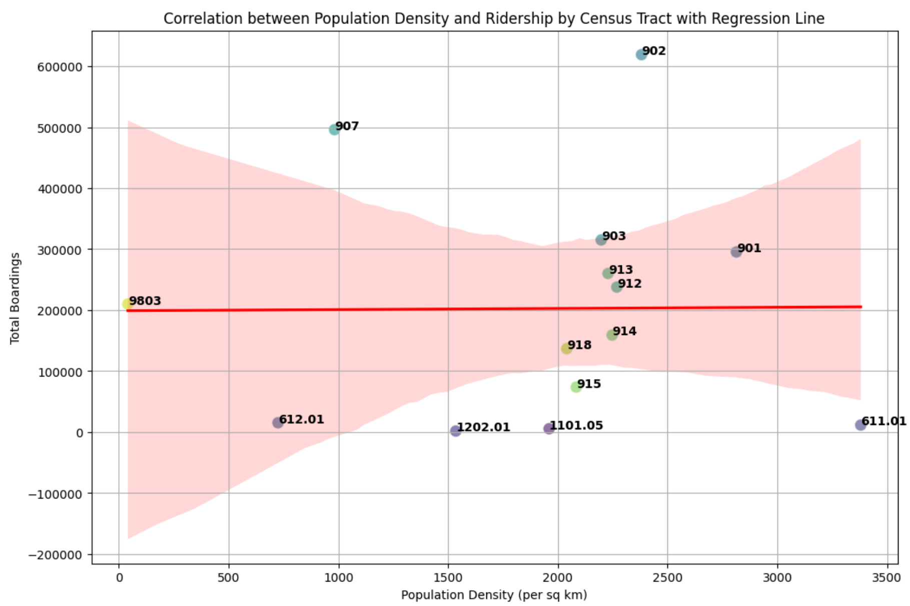


Graph showing the relationship between average boardings per load (y-axis) and disability rate across route 16 tracts (x-axis) for the three different phases of the PATI program.

We created a line plot graph showing a trend line for each PATI program category to demonstrate how the PATI program might influence ridership differently in contexts with varying disability rates. The disability rates were placed into bins to allow for better visualization of the data: 0.0-0.2%, 0.2-0.4%, 0.4-0.6%, and 0.6-0.8%. This graph suggests that structured PATI phases programs may be better at accommodating or attracting riders with disabilities at varying rates compared to unprogrammed services with the steep decline in ridership for high disability rate tracts.

The data suggests that Phase 2 is particularly effective up to a certain threshold of disability prevalence, beyond which additional support or program adjustments may be necessary to maintain high service utilization. However, the correlation between the relationship between disability rates and average boardings per load under different phases of the PATI program is not fully obvious especially with our smaller sample data for a particular route. Understanding these trends further is crucial for assessing the impact of transportation programs on accessibility and for planning future services to better meet the needs of disabled passengers.

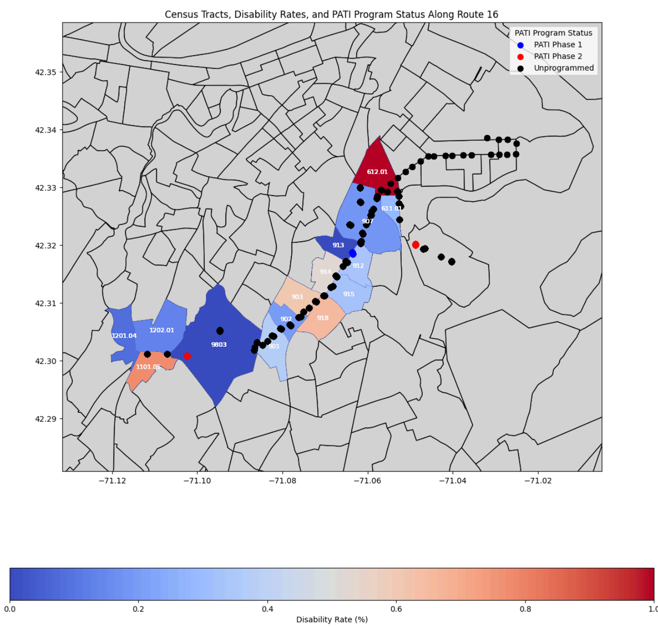
As a result, we decided to pivot and do a similar analysis on population density and ridership to see if accessibility to the bus would be affected by the amount of people living in a particular tract along route 16 (i.e. being unable to use the bus due to overcrowding). To visualize this, we directly plotted a scatterplot to see the correlation between population density and ridership by census tract to see if high population density would result in high ridership and vice versa.



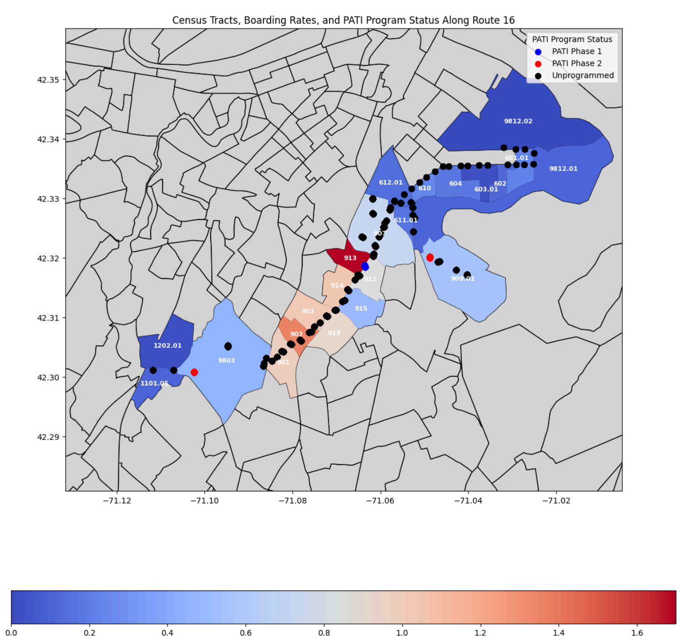
Graph showing the relationship between total boardings per load (y-axis) and population density across route 16 tracts (x-axis) for each census tract (points) with linear regression and a confidence interval.

We decided to use total boardings over average boardings to assess the overall demand and usage, rather than passenger density per trip. The graph above, with a mostly flat regression line, suggests that there is little to no correlation between population density and total boardings. In other words, the accessibility of buses nor the usage of buses along route 16 are not correlated with population density. This lack of correlation suggests that simply increasing population density might not directly increase public transit ridership, emphasizing the need to consider other factors such as transit accessibility, service quality, and socioeconomic variables. The variability in ridership despite similar population densities could indicate areas where public transit is either underutilized or overburdened. This information can help in prioritizing areas for transit development or enhancements to meet actual demand more effectively, especially for particular tracts for route 16.

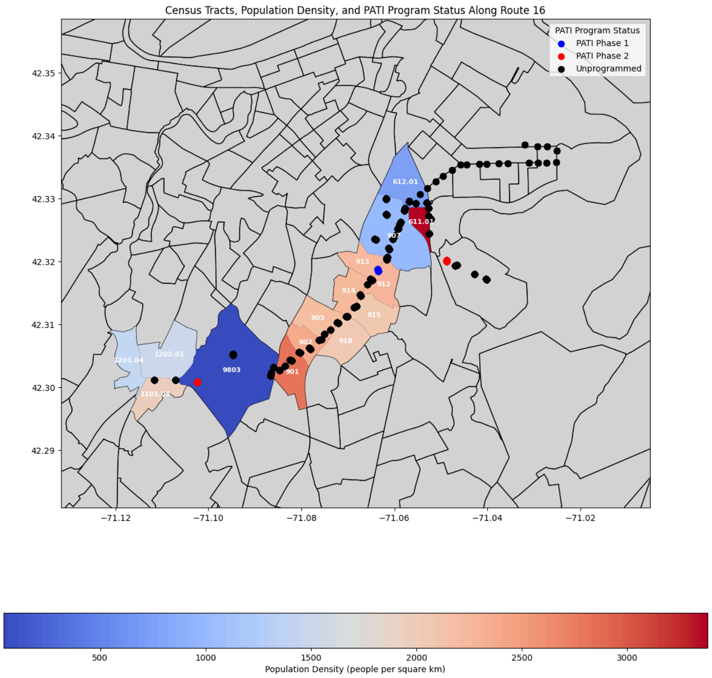
Next, we decided to create maps to better visualize each factor in terms of the census tracts, since the above graphs were not promising in showing any correlations. We looked at disability rate, average boardings, and population density separately in relationship to the census tracts that route 16 travels through.



Map showing disability rate by route 16 census tracts and PATI bus stop program statuses.



Map showing average boardings by route 16 census tracts and PATI bus stop program statuses.



Map showing population density by route 16 census tracts and PATI bus stop program statuses.

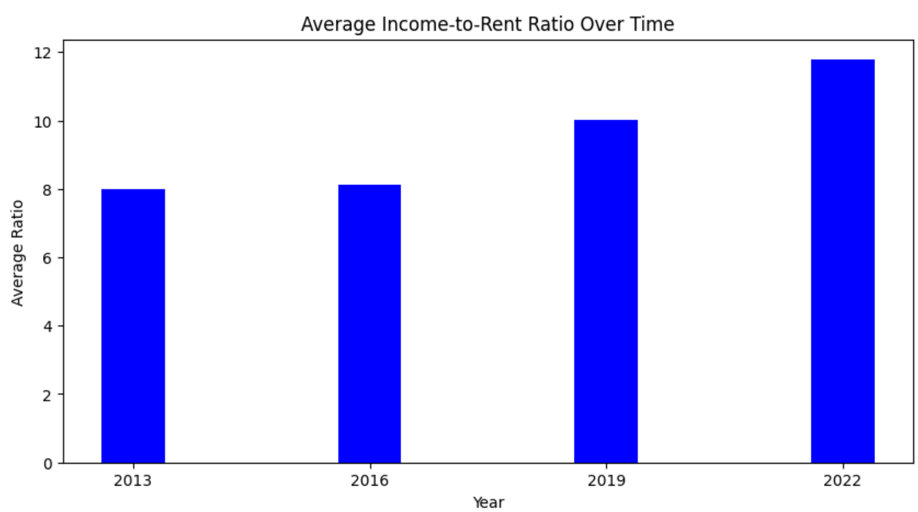
By mapping out each of the features, disability rate, average boarding rate, and population density onto the route 16 tracts, we can begin to paint a picture by cross-analyzing these three graphs. We can see some correlation between the map for average boardings and the map with population density, with similar tracts. For example, tract 9803 (which is a greenway) is low in both population density and disability rates but average for boardings. This indicates that people are still boarding that stop fairly frequently considering that there are little to no people living in that tract. This can be due to people migrating from neighboring tracts. Namely, people may be moving from tract 1101.05, which has a relatively high disability rate to access the PATI accessible bus stop. Additionally, we can see tract 913 has very high boardings but relatively average population density meaning that tract 913 is a transit hub area or a tract that contains people who need, prefer, or have greater access to public transportation.

Although the three graphs do not show strong correlations in terms of disability rates, average boardings, and population density (in relation to each other and in relation to the PATI program's efforts on their own), there are stories and finding that can be found within the graphs when we cross analyze them. In other words, although disability rates, average boardings, and population density alone do not have correlations with displacement, these features combined do paint a picture of the importance of accessibility in terms of displacement.

VI. Extension Project - Income-Based Correlations

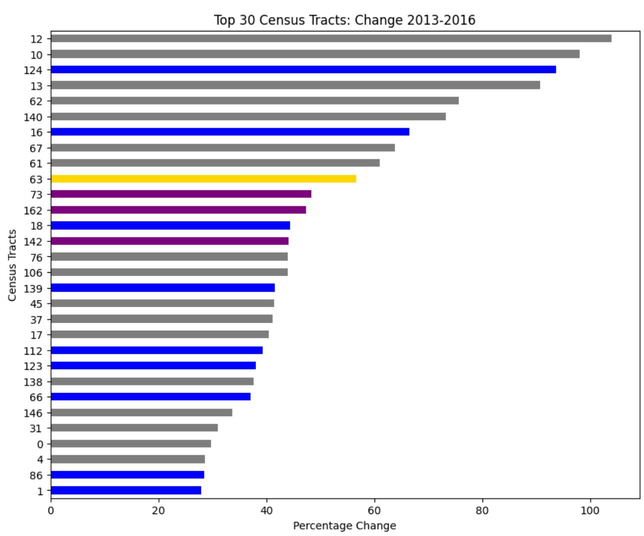
One of the components we explored for our extension project was the relationship income had on displacement. The topic was intentionally vague because there are so many correlations we can analyze from income alone. In particular, the focus of this analysis is to identify any relationships between income and housing data, and also income and transportation data. The rationale behind this approach is to see if there is any sign of gentrification that may cause displacement of lower income households or individuals. Significant shifts in income, housing prices or transportation habits of an area are often indicators of some form of gentrification, thus we chose to focus on using these three variables for our exploration and analysis. If any relationship exists or new relationships are discovered, we can then try to make sense of them to come up with possible hunches as to why these correlations are happening and what it means to displacement.

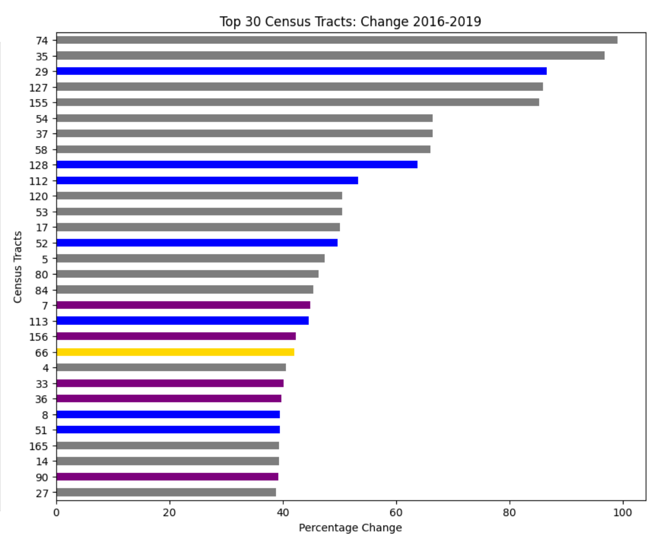
To explore these relationships, we once again used data collected from the ACS survey of individuals and households living in the Suffolk County of Massachusetts, sectioned by census tracts to easily navigate which areas of the Suffolk County are experiencing economical, financial and demographic shifts. These shifts should be highlighted as it indicates that there is something going on. In order to highlight shifts over time, I took ACS survey data in 2013, 2016, 2019 and 2022 and started by looking into the relationship income has on rent.

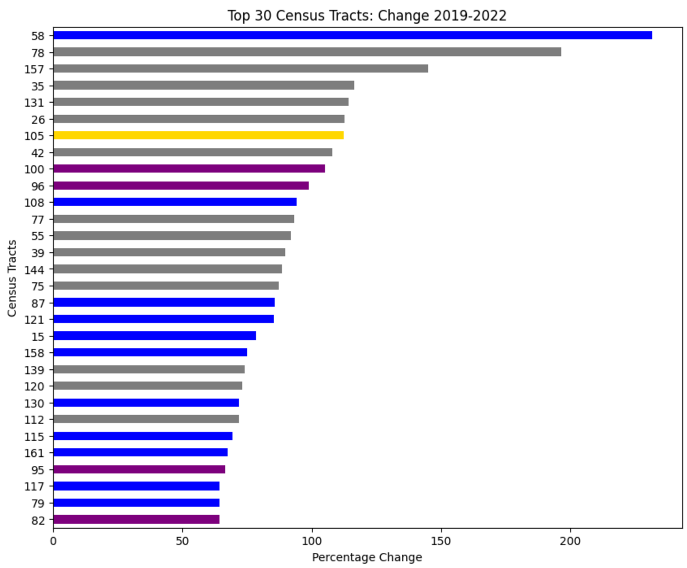


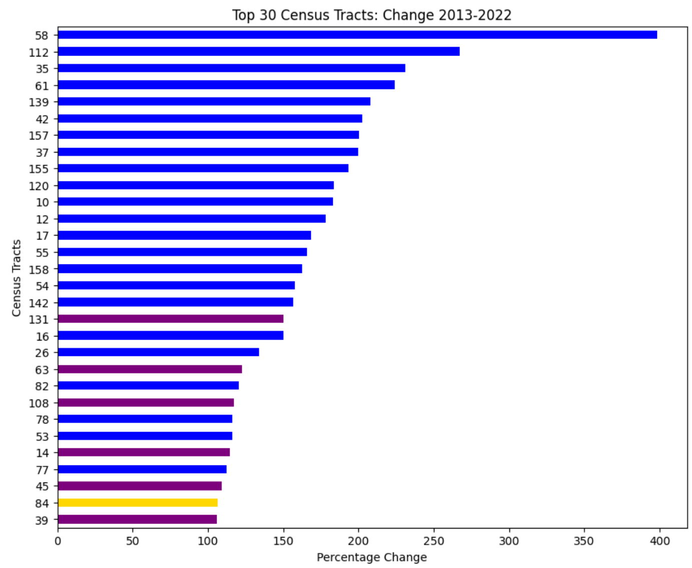
Above is a bar graph showing the shift in income to rent ratio from 2013 to 2022. I selected the income to rent ratio as a metric because it provides crucial insights into housing affordability or broader economic positions. A lower ratio likely means that a higher percentage of a household's income is spent on rent, which can indicate financial stress and burden on households. Monitoring these changes can help us assess whether housing is becoming more or less affordable over time. Currently, the income-to-rent ratio of all census tracts is averaged to identify overall trends. As we can see, the average ratio has been on an upward trend from 2013 to 2022, which indicates that housing is becoming more affordable overtime, or that the overall income for most households is increasing.

However, examining Suffolk County as a whole may obscure significant local dynamics that drive these shifts. Although the average income to rent ratio is increasing, we have to recognize that the broader view of this can mask underlying disparities. Just because the average is increasing, it does not mean that all areas are experiencing sharp increases in affordability. Some areas may see stagnant or declining ratios, where incomes are not keeping pace with rising rents, which leads to increased financial pressure on residents. The variation in economic conditions can lead to displacement, particularly in tracts where rent increase surpasses the growth of household income. Displacement often occurs when long-standing residents, particularly those with lower income, find themselves unable to afford increasing living costs or are forced out by newcomers to the area, which is often associated with gentrification. To understand and address the complexities of displacement, we must delve deeper into the census tract level, and identify areas where income to rent ratio changes are more significant. This will help identify areas that are at risk of displacement, and potentially even find common factors that may contribute to this.



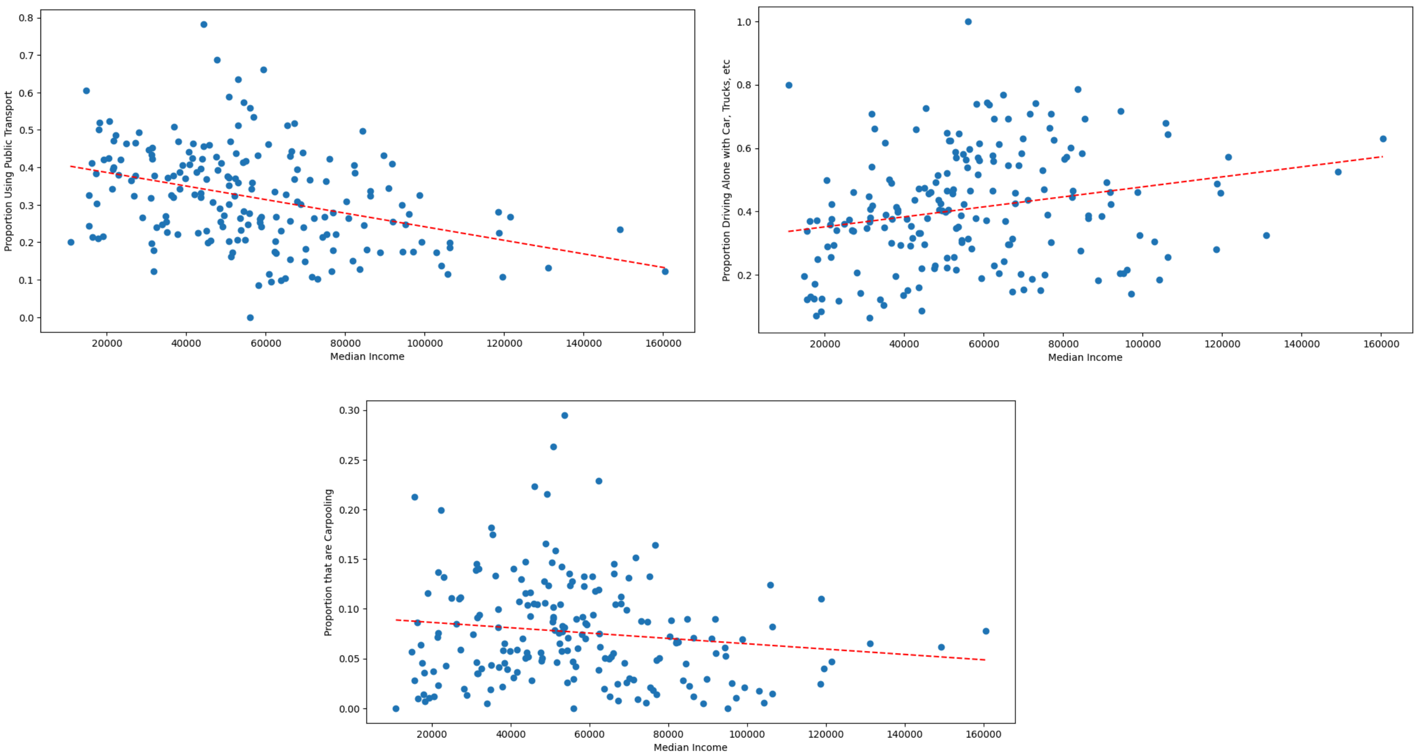






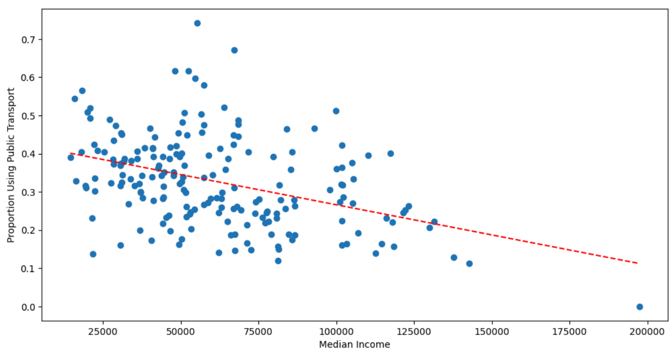
The four graphs above show the top 30 census tracts that have had the most drastic change in income to rent ratio within a given time period. The degree of change for each census tract is measured by percentage change from one year to another. In this case, we have graphs of four different ranges of years: 2013-2016, 2016-2019, 2019-2022 and 2013-2022. The census tracts are also color coded, such that tracts in gold mean that the tract has shown up in all 4 year ranges, purple means it has shown up in 3 year ranges, blue means it has shown up in 2 year ranges, and gray means it has shown up in 1 year range. The purpose of color coding these tracts is to easily identify areas with persistent changes in economic conditions over time. For example, tracts in gold may require additional attention for further detailed study, as the consistent appearance suggests ongoing economic transitions. Tracts in blue or purple also hold the same effect at a smaller scale, and can still indicate specific events or developments in those periods where they have shown up. Gray tracts may reflect changes attributed to one-off events or non-systematic shifts. We have built a foundational understanding of the diverse economic experiences by highlighting tracts with significant changes.

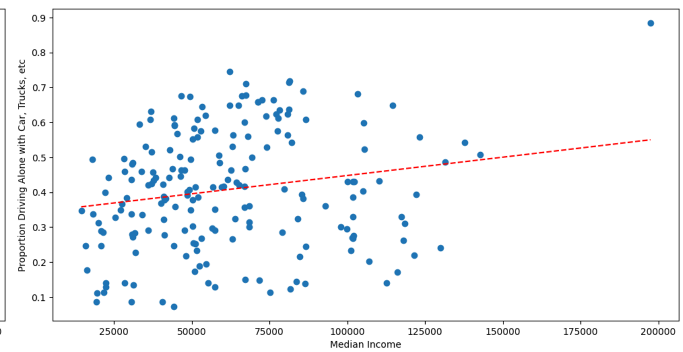
As we transition from examining housing affordability to exploring the relationship between income levels and transportation usage, we need to consider how impactful these elements are to the daily lives of individuals. For example, transportation availability and affordability can significantly impact a community’s access to employment opportunities, which influences income stability and housing affordability. There is interconnectedness here and it would be beneficial to understand transportation habits and usage in these tracts.

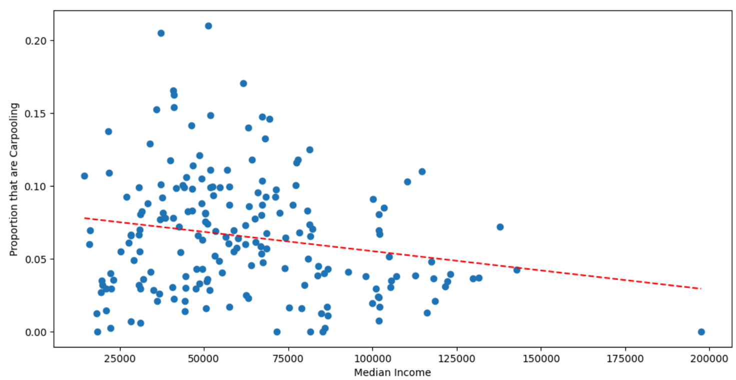


2013 Data for 3 Modes of Transportation

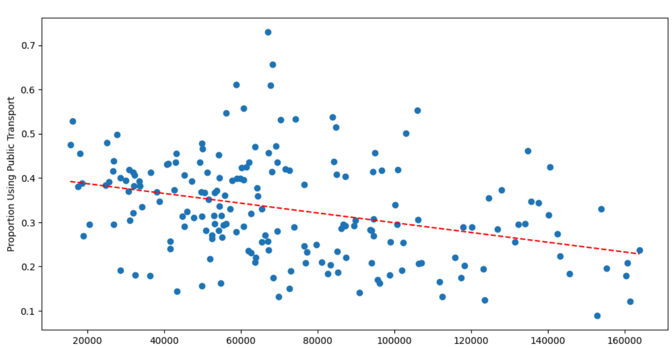
The three graphs above are the relationship between means of transportation to the median income of an area, such that every sample point represents a census tract. This trio of graphs represent surveys from 2013. The combination of these graphs provide a foundational understanding of how income influences transportation choices within various communities. For public transportation, there is a clear negative correlation between public transit usage and median household income. As median income increases, the proportion of the population using public transportation decreases. This trend likely indicates that higher-income individuals have more access to private vehicles or other convenient modes of transportation and therefore rely less on public transit. It could also reflect the availability and quality of public transportation services in higher-income areas, which may not meet the needs or preferences of wealthier residents. For people who chose to drive alone, it shows a positive correlation between the proportion of people driving alone to work and their median income. Similarly to the public transit relationship, this relationship suggests that individuals in higher-income brackets are more likely to own personal vehicles and opt in for the convenience of driving alone. This might also reflect the quality of urban planning, such that road infrastructure is better designed and more spread out in higher income regions, making it a more feasible option compared to public transport. Finally, the graph about carpooling shows a slight negative correlation, where carpooling slightly decreases as median income increases. Carpooling is a cost-effective transportation solution, so low-income groups may opt for this mode of transport more.

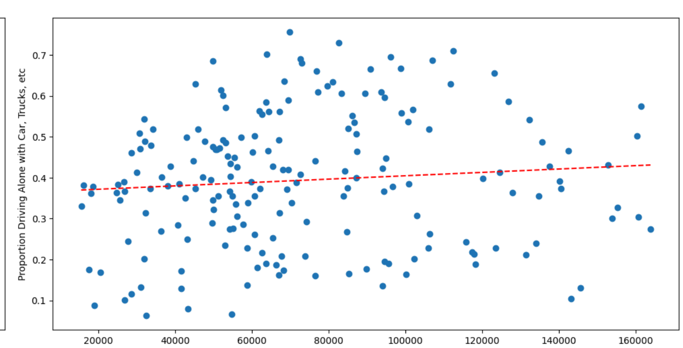


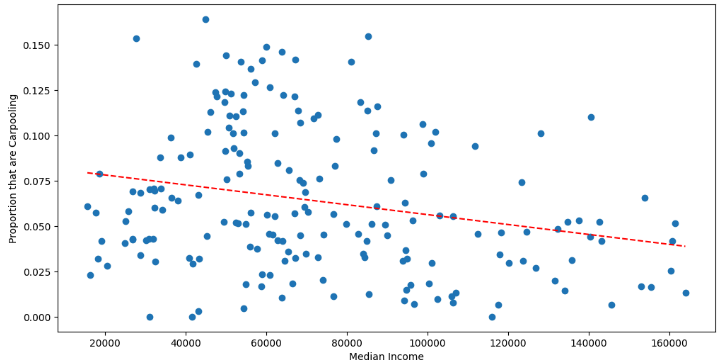




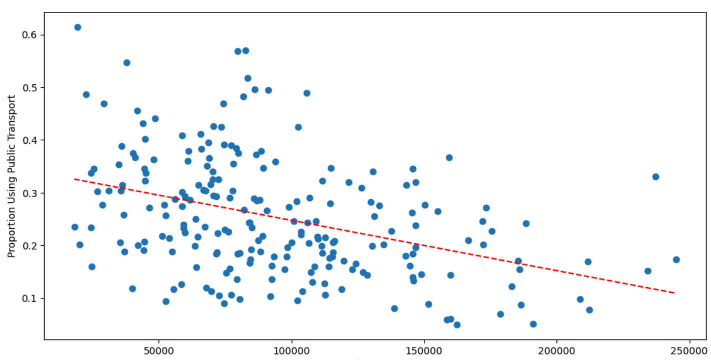
Just like for 2013, I have made the same graphs for 2016, 2019 and 2022. This trio of graphs is for 2016. Since the behavior of these graphs are incredibly similar to 2013, the trends describe the same behavior and suggest the same things for a given mode of transportation.

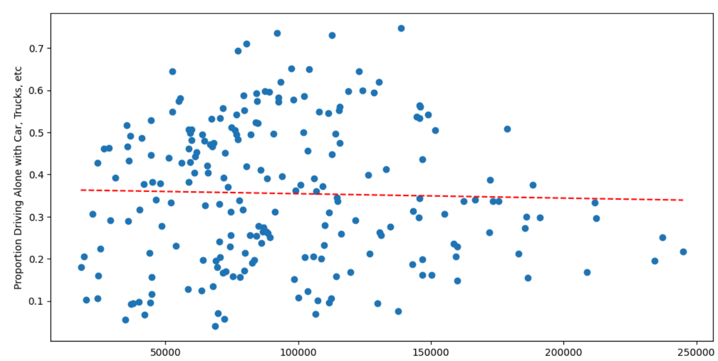


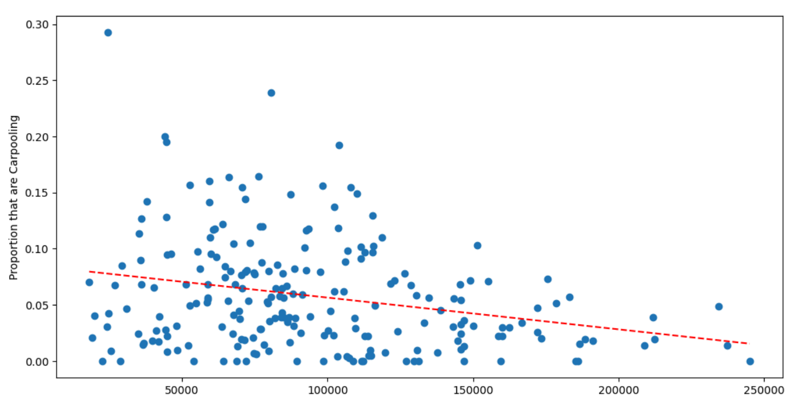




These are the graphs for data from 2019. Once again, the trends and transportation habits remain the same across different income groups, thus suggesting the same things as the previous graphs. However, a slight difference here is the steepness of the slope for choosing to drive alone compared to 2013 and 2016. The previous two years had a clearer positive slope, but the slope for 2019 has started to flatten out, suggesting that an individual's choice to drive alone has become more widespread across the span of different income groups. We cannot be certain why this is the case, but some hunches include increased affordability of private vehicles, residents in newer suburban areas may find themselves with fewer public transit options, or the development of rideshare services. Nonetheless, this slight change is quite interesting.

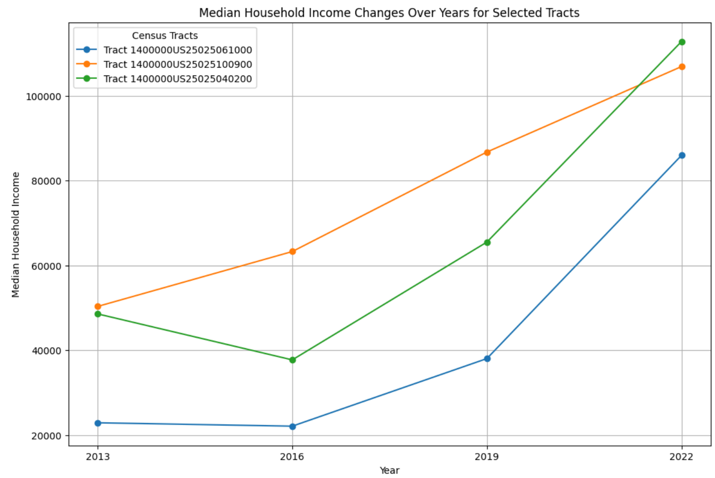




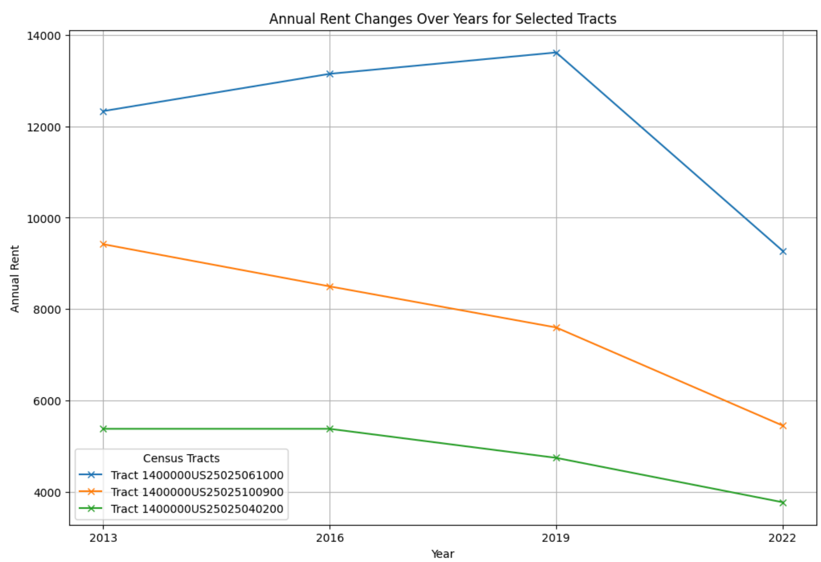


Finally, the trio of graphs above here is data from 2022. Once again, the trends are similar to previous years, except for the relationship between driving alone and median income. In fact, it appears that the trend line is essentially flat, but slightly sloping down. As opposed to the other 3 years, this is no longer a positive relationship. While this is likely due to the limitations and flaws of the survey data (perhaps fewer people with vehicles didn't take the survey), we can still argue that the usage of private vehicles is still more widespread among the lower income groups. Since this is not a coincidence and a growing trend starting from 2019, we can link this to suggestions and implications about displacement. An overall rise in car ownership (which might be what the data suggests) can facilitate a greater suburban sprawl because people are able to live further away from where they work. This can lead to urban displacement such that low income groups are pushed towards the outskirts due to the rising costs of living in the central areas as they become more car dependent. This might be more applicable to city centers, such that road infrastructure or parking systems may have improved overtime. This incentivizes higher income residents, which can potentially lead to gentrification over the years, and causes the displacement of lower income groups not being able to afford the rising rents and property prices. However, this is all just a hunch derived from a change in trend behaviors over time and is not conclusive at all.

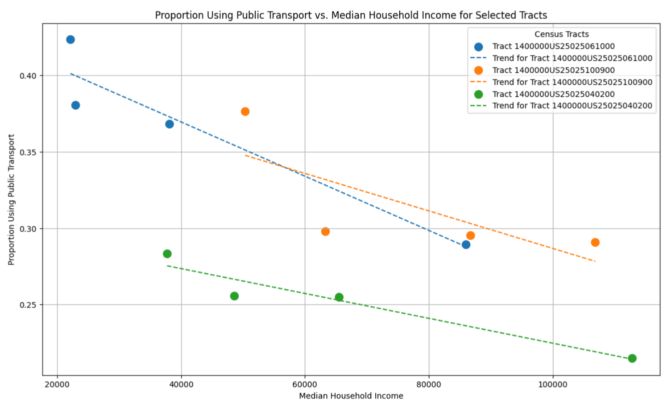
Now that we have looked at more general trends of income to housing data, and income to transportation data, it might be valuable to also perform similar analyses on a small group of census tracts to better understand and observe the nuances in the social and economic shifts. For simplicity purposes, I opted in to choose three census tracts, and I selected the census tracts with these IDs: '1400000US25025061000', '1400000US25025100900', and '1400000US25025040200'. These tracts were selected based on the top 3 tracts with the largest percentage change of income to rent ratio from 2013 to 2022 (refer to the previous analysis about top 30 census tracts). We know that these tracts have had substantial change at some point over the past decade, and the goal is to see the specific economical and social shifts over time in these tracts.

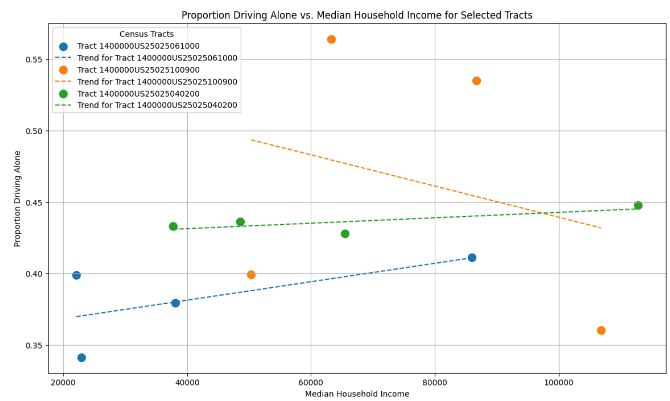


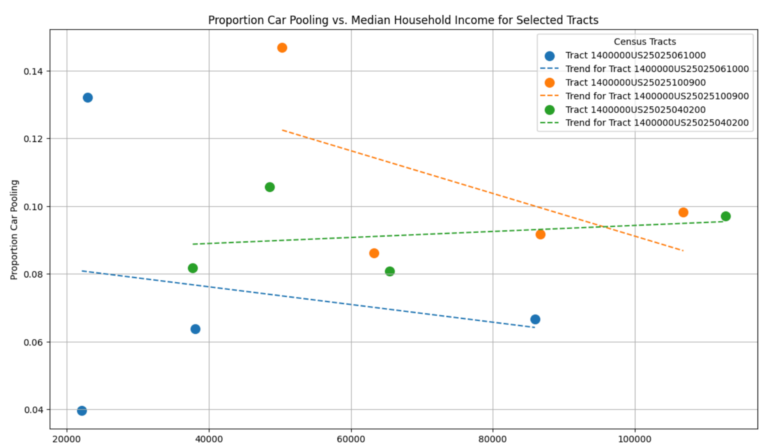
The graph above illustrates the changes in median household income from 2013 to 2022 for the three census tracts selected. All tracts here show an increase in median income over time, which tracks as these were selected for the largest percentage changes. The sharp increase in median income over the span of a decade suggests that there are profound changes happening in these tracts, perhaps some redevelopment happening, gentrification, or changes in the housing market that possibly lead to an influx of high income residents to reside in these tracts.



The graph above illustrates the changes in annual rent for the three tracts from 2013 to 2022. In comparison to the graph illustrating the changes in income, it makes sense that the percentage change of these tracts were so large. The median annual rent of all these tracts has been on a steep decline over time, yet the residents are becoming richer. Tracing back to what was said earlier, perhaps this change in income to rent ratio does imply that houses are becoming more affordable in these tracts, however, one thing we haven't considered is the factor of house ownership. It is more common for residents in higher income groups to be house owners. While this analysis wasn't paired with house ownership data, it might explain why rent is decreasing while income is increasing. The shift of renting to owning would decrease the demand for rental housing, thus potentially decreasing the rent prices. Additionally, as housing markets become more competitive when the community transitions to house ownership, landlords might be more incentivized to lower rents to retain existing tenants. While this insight is not supported by house ownership data, it suggests a potential shift in the rental market dynamics that could impact overall housing affordability. With such shifts in housing dynamics, the increase in property values can lead to economic displacement. Even though rents are decreasing, residents may still find it difficult to afford other living costs in the area or may feel more pressured to move elsewhere as the area becomes more gentrified and upscale.

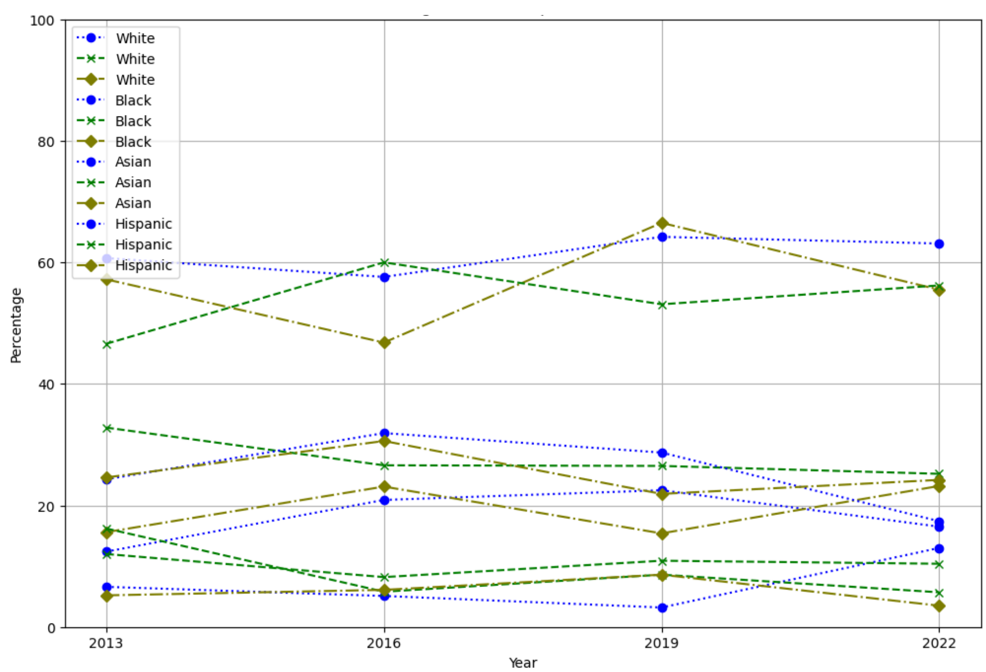






In addition to looking at income and rent data, I also looked more into transportation usage habits in these census tracts. The graph may look a little confusing, but here is a breakdown of the graph. Each sample point on the graph represents a year, either 2013, 2016, 2019 or 2022. Since I am also connecting this to household income, that is the x-axis of the graph. But because median household income has shown to increase overtime for these three census tracts, we can also see how means of transportation habits have changed over time relative (as the points are on the right, that shows the time shift). Overall, these trends are quite similar to the transportation usage findings earlier with all the census tracts, such that there is a clear decline in usage of public transit and carpooling over time as income increases, and the usage of private vehicles tend to show a more positive trend over time as income increases. One interesting observation we can make, however, is to compare those that remain relatively consistent over time, and those that do not. For example, the green census tract in the graph with driving alone either has slight positive or flat trends. This suggests that despite the median household income, using a private vehicle is still more preferred in these areas. Rather than implying that people would drive more as income increases, this flatness of the trends for some of these tracts can be related to road infrastructure or public transit infrastructure.

Finally, to explain these drastic growth in median income for these tracts, I wanted to compare it with demographic data, in particular with the race distribution of the census tract. Based on past evidence, certain race groups have a history of being more financially well off, in particular people that are white. Observing demographic shifts over the year may help explain if high income individuals are entering these tracts, and if so, low income residents being displaced?



The graph above illustrates the percentage of race groups in each three tracts. The races included are white, black, asian, and hispanic. As observed, there are not any significant shifts in demographic data. The race distribution in these areas remained relatively the same throughout the decade, despite the increase in median household income. While the illustration of this graph can be a little confusing, we can observe that there are not any drastic changes as the lines for the data points remain relatively intact to each other. Although there is no correlation between demographic shifts over time in these census tracts, it doesn't mean that low income groups are not being displaced in these census tracts as overall income increases. It just means that we cannot identify the influx and outflow of high and low income individuals based on race data alone.

The exploration of relationships between income, rent, and transportation has unveiled some of the more complex and nuanced dynamics that impact urban/suburban living. While income as a component is quite vague as it can be paired up with some many other aspects in many different ways, it is still a critical part in understanding the shifts and changes that occur in an area. Based on our observations of shifts in income, shifts in transportation usage, and shifts in rental costs, the many patterns make two implications. First, it implies that the higher income levels are associated with a decreased usage of public transportation, and typically an increased usage of private vehicles – however, this can still be influenced by public transit infrastructure and road infrastructure as we have noticed a weaker trend on this pattern over time. Additionally, the interplay between rising incomes and rental market dynamics may contribute to gentrification and displacement of long-term residents with lower incomes. Although the findings are not very definitive due to the vagueness of income, it paves a road for more targeted investigations and inquiries in the future. There were many suggestions, hints and patterns from our analysis that points towards potential social displacement, and economical displacement.

The most significant limitation to these analysis is the difficulty it is to measure data over time. Since the data I am using are surveys conducted during separate years, there may be a large margin of error from many of these data values due to inconsistencies of who is being surveyed. While there are some prominent trends and patterns observed from our analysis, the variability in the survey methodologies over different years can introduce uncertainties. Unfortunately, this may just be a part of the nature of trying to capture data shifts and changes over time.

VII. Future Scope

Expanding this project to future scopes and followup questions involves challenging many of the assumptions that were made when conducting our original data collection and analysis. More specifically, a future project could compare our analysis and visualizations with ACS census metrics five or ten years from now. Another project could explore the indirect impact of transportation infrastructure on communities that are geographically distanced from the development. Lastly, our findings could be compared with similar analysis done on infrastructure in other cities like Boston.

A general limitation of our analysis is that Boston's transportation infrastructure is quite old. The train and bus lines were largely established in the first half of the twentieth century, and some of Boston's green infrastructure is several centuries old. This means there is a lack of data on the immediate displacement impact caused by new transportation infrastructure. This is why we chose to focus on one of Boston's newer transportation developments, the GLX project. In the next few years, we should get more precise data on GLX ridership details and the demographics of the local community. This paves the way for more comprehensive analysis that may inform any planned transportation developments.

We also operated under the assumption that the impact of new transportation line is geographically immediate. While the impact on the local community is most apparent, it is worth observing the impact on exogenous communities or infrastructure. For example, the addition of new bike lanes may reduce the need for local Uber drivers and train operators, thereby forcing them to find a new job in another area. Thus, it is worth looking into the less immediate impact of new infrastructure, perhaps visualized on a geographic heat map.

Lastly, another followup project should examine to what extent Boston's urban planning mitigates or exacerbates the effect of infrastructure-based displacement. For example, New York City has a far more expansive subway system, intuitively designed in a grid-like manner for ease of use. Therefore, the addition of a new bike lane may have less of a displacement effect in New York City than in Boston. Likewise, adding a new bike lane in Los Angeles may also have a negligible displacement effect because of the city's car-dependent infrastructure. Comparing Boston's infrastructure with other cities would allow a future project to isolate the confounding factors unique to Boston. Moreover, a project could examine how other cities mitigate the displacement effect caused by new infrastructure, and bring it to Boston.

VIII. Individual Contributions

See below individual contributions:

William Huang	Marco Chin Fung Lam
1. Acted as representative for the team for meetings with client, project managers, and relayed communications to team from higher-ups. 2. Use ACS 5-year estimates to estimate early insights and find correlations for mid-semester findings between the 2023 Boston Bike Network and percentage of commuters who cycled in Suffolk County. 3. Presented both Early Insights and Mid-Semester deliverables to clients. 4. Contributed to Demo Day poster and Final Report.	1. Did early insights for general trends and relationships of income with rent data and means of transportation data as an extension project. 2. Use ACS 5-year estimates (2013, 2016, 2019, 2022 data) to find any evidence of displacement based on income, rent and transportation trends – in particular, any drastic demographic, economical/financial and housing shifts. 3. Presented early income insights to the client for the mid-semester deliverables. 4. Contributed to Demo Day poster and Final Report.

Jacob Stein	Nyx Zhao
1. Did early insights analysis on three Boston greenways: Rose Kennedy, Riverway, and East Boston Greenway. 2. Use ACS 5-year estimates to find correlations between the 2018-2022 Green Line Extension and displacement in the local community for mid-semester deliverable. 3. Presented GLX project findings to client for mid-semester deliverable. 4. Contributed to Demo Day poster and Final Report.	1. Did early insights for general trends of bus displacement correlation extension project. 2. Use ACS 2022 census tract, surveys, and MBTA characteristic data to explore possible correlations between accessibility, ridership, and displacement for Columbia Road bus route 16. 3. Presented general bus early insights to the client for mid-semester deliverables. 4. Contributed to Demo Day poster and Final Report.