Old Streets, New Bike Lanes: Bicycle Facilities and Neighborhood Change in Boston, MA

A thesis submitted by

Phoebe J. Whitwell

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Urban and Environmental Policy and Planning

Tufts University

August 2024

Adviser: Shomon Shamsuddin

Reader: Sumeeta Srinivasan

Abstract

Like many cities in the US and internationally, Boston has invested heavily in its bicycle network in recent years. While this investment is intended to improve sustainability and liveability, community members are concerned that there may be unintended and unwanted results, such as rising housing costs and displacement through green gentrification. Prior literature on this topic mainly focused on equitability of network distribution. The studies that did address gentrification found that higher rates of facility installation in a neighborhood were correlated with greater socioeconomic change but that there was only weak evidence for a causal relationship. This thesis therefore used a panel regression model to evaluate the relationship between bicycle facility presence and assessed housing values on the parcel level in Boston between 2008 and 2019, controlling for other characteristics of the parcel and location. While the final model was a strong fit for the data, bicycle facility coverage was not statistically significantly associated with assessed housing value. Rather, variance in assessed property values between parcels was mainly determined by lot and building size, while change in assessed property values for the same parcel over time was driven by unobserved regional forces that were included as time fixed effects in the model. While bicycle facility installation does not appear to be a direct driver of increases in assessed property values, it is nonetheless essential for planners to consider the wider context of ongoing inequity and gentrification in Boston and place the bicycle network expansion within a larger strategy of equity-oriented planning policies.

Acknowledgements

I am deeply grateful for the support of my thesis advisor, Professor Shomon Shamsuddin, who guided this project conceptually and practically. In particular, whatever rigor I managed to achieve in the construction of the dataset and the formulation of the panel regression (not to mention my understanding of panel regressions as a concept) are entirely due to his helpful questions and insights. I am similarly grateful to Professor Sumeeta Srinivasan, who read for this thesis and provided much-needed additional support on the GIS components. Thank you both for so generously offering your time and energy to this project and to my personal and professional progress as a UEP student and an urban planner.

I would also like to acknowledge the UEP program and community for being an excellent place to learn and grow over the past two years and for enabling me to leave UEP as not just a more knowledgeable but a more thoughtful and confident researcher and practitioner than I entered. Finally, I am grateful for the support of my friends and family, particularly Nora, who patiently and sweetly assured me countless times I could and would complete this project, no matter how fervently I claimed I could not.

Table of Contents

List of Tables

List of Figures

Old Streets, New Bike Lanes: Bicycle Facilities and Neighborhood Change in Boston, MA

Chapter 1: Introduction

In 2010, then-Boston Mayor Thomas Menino declared, “The car is no longer king in Boston.” Since then, Boston has been a leader in North American cities in expanding its bicycle network, installing close to 100 miles in bicycle lanes and off-street paths between 2010 and 2020 and effectively doubling the size of the network. The stated goals of the network expansion are to provide “comfortable and safe bicycle paths, lanes, and shared streets that appeal to new and existing riders” (Tinlin et al 2013). Community members of neighborhoods that have been historically overlooked for infrastructure improvements say that the investment is very needed to improve safety for bicyclists and make bicycling more accessible for residents who don’t currently feel comfortable biking in lanes unseparated from car traffic (Woodard 2022).

Other community members oppose bike lanes, citing concerns of green gentrification. Following the announcement of a project to redesign Blue Hill Avenue, a major corridor in the historically marginalized neighborhoods of Dorchester and Mattapan, to include expanded infrastructure for bicycling and transit, residents shared concerns that the bicycle lanes were intended to “attract young white professionals to buy in the area” and “push [current Black and Latino residents] out of the city” (Woodard 2022). Aligning with an overall perception that most cyclists are young, white, and higher income, residents worried that the benefits of bike lanes, such as enabling community members to avoid sitting in car traffic and to make more environmentally friendly transportation choices, would be felt mostly by higher-income people moving into the area. Shavel’le Olivier, the executive director of Mattapan Food & Fitness Coalition, summed up her analysis of these concerns by saying, “Until we resolve...issues around housing, there will continue to be resistance to bike lanes and it will continue to be seen as a negative thing” (Olivier 2022).

Green gentrification is an emerging field of study and an important area of focus for urban planners, as it joins transit development, art installations, zoning, highway construction, and much more on the long list of ways that planning (intentionally or inadvertently) has been a force for neighborhood change. Neighborhood change refers simply to the changes in the physical or socioeconomic characteristics of an area over time. It is typically considered in a negative light, such as in cases of gentrification. Gentrification is defined by many anti-displacement activists as “a profit-driven, race, and class change of a historically disinvested neighborhood” (NLIHC 2019). A common understanding of gentrification is that investment in the physical infrastructure or landscape of amenities (such as businesses or public resources like libraries) in a neighborhood contributes to the neighborhood becoming more desirable. New interest in the neighborhood from people able to pay more for housing allows landlords to increase rents and homeowners or developers to increase housing sale prices. This in turn displaces the lower-income residents who had previously been living in the neighborhood, either directly due to their inability to pay increased housing costs or indirectly due to a new lack of affordable housing options. Even if low-income residents are able to stay, they may be forced into higher levels of housing cost burden and housing instability (NLIHC 2019).

Green gentrification is most clearly understood as a cyclical process. It does not apply similarly in all communities and circumstances. In some cases, neighborhood change may follow greening. In others, municipalities may invest more in green infrastructure or active transportation in neighborhoods that are already changing. Greening neighborhoods may also experience no demographic changes at all (Collins and Rigolon 2022). The green gentrification cycle is not yet well understood. Planners and policymakers will and must continue attempting to address an ongoing climate crisis and to improve liveability and transportation in urban areas.

Active transportation facilities will continue to be an essential piece of that puzzle; however, considering who will benefit from these facilities and how that will occur is an equally essential piece. Causing neighborhoods to become unaffordable to low-income residents results in the same inequitable distribution of resources as the historic practice of investing only in wealthy neighborhoods. It is therefore important for planners to evaluate both the equitability of infrastructure investment distribution and how those investments are impacting surrounding communities.

Research Questions and Goals

This thesis seeks to answer the following research question:

- How does the installation of bicycle facilities in Boston relate spatially and temporally to rising housing values and changing neighborhood demographics?

The goal of the analysis is to develop an understanding of the relationship between bicycle facilities and neighborhood characteristics through quantitative analysis of spatial data. My working hypothesis is that construction of new bicycle facilities is associated with increases in assessed property values of surrounding parcels and with demographic changes typically associated with gentrification.

Chapter 2: Literature Review

Overall, the literature is inconclusive on the questions of whether bicycle networks are distributed equitably (and for what groups) and what the relationship is between bicycle facilities and neighborhood change. Many studies found that there was at least some correlation between low levels of access to high-quality bicycle facilities and high presence of marginalized populations, particularly low-income people and people of color. However, other studies found that people of color and low-income people had better access to bicycle facilities compared to other groups. It's clear that whether the bicycle facility network in a given city or region is distributed equitably is highly dependent on the specific characteristics of that area. While it's likely that this is also true for the relationship between bicycle facilities and neighborhood change, there have not been enough studies conducted on this topic to make any strong conclusions. Of these studies, all three found at least a weak correlation between greater increases in median household income and greater investment in bicycle facilities. However, the studies either did not address or could not conclude whether bicycle facility installation was driving socioeconomic changes or neighborhoods experiencing socioeconomic change were prioritized for bicycle facility projects (Flanagan et al 2016, Hirsch et al 2017, Ferenchek and Marshall 2021). Similarly, the three studies I cited that evaluated the relationship between bicycle facilities and property values were equally inconclusive. Two studies found that shared use paths were positively associated with increased property sale prices while on-street facilities were either negatively associated or neutral (except in suburban areas, where all bicycle facilities were negatively associated with sale prices) (Krizek 2006, Welch et al 2016). Meanwhile, a third study found that all bicycle facilities were positively associated with increased sale prices (Liu and Shi 2017).

Authors	Year	Geographic area	Method	Spatial unit	Demographic variables/groups of interest	Bicycle facility metrics
Krizek	2006	Minneapolis/ St. Paul, MN	OLS Regression	Parcel	Property sale prices	Distance to nearest facility
Dill and Hagerty	2009	Portland, OR	Observational comparison	Block group	Poverty Race and ethnicity Youth Older adults	Bikeway length per area
Cradock et al	2009	United States	OLS regression	County	Poverty Educational attainment	Project initiation (y/n) Project funding
Deka and Connolly	2011	New Jersey	ANOVA	Block group	Median income Race and ethnicity Non-English- speakers Immigrants	Facility located within one mile (y/n)
Teunissen et al	2013	Bogotá, Colombia	Observational comparison	Neigh-borhood	Income level	Distance to bicycle network Bicycle mode share
Flanagan et al	2016	Portland, OR and Chicago, IL	OLS regression	Census tract	Race and ethnicity Homeownership Educational attainment Median household income Age Median home value	Cycling infrastructure investment index
Welch et al	2016	Portland, OR	Random effects spatial panel regression	Parcel	Property sale prices	Distance to nearest facility
Fuller and Winters	2017	8 cities in Canada	OLS regression	Census tract	Income quintile	Bike Score
Wang and Lindsay	2017	Minneapolis, MN	Observational comparison	Block group	Access to a vehicle Race and ethnicity Poverty Youth Older adults	Gini coefficient Three-mile service area via bikeway
Hirsch et al	2017	4 cities in the United States	Local Moran's I spatial cluster analysis	Neigh-borhood	Poverty Unemployment Median household income Housing occupancy Race and ethnicity	Bikeway length
Liu and Shi	2017	Portland, OR	Pooled OLS hedonic price regression Spatial hedonic regression	Parcel	Property sale prices	Distance to nearest facility Bikeway length per area

Winters et al	2018	3 cities in Canada	ANOVA Bivariate Moran's I spatial cluster analysis	DA	Median household income	Bikeway length per area
Parra et al	2018	Bogotá, Colombia	Observational comparison	Neigh-borhood	Income level	Distance to bicycle network
Tucker and Manaugh	2018	Rio de Janeiro and Curitiba, Brazil	OLS regression	Neigh-borhood	Income quintile	Bikeway length Distance to amenities via bikeway
Houde et al	2018	Montreal, Canada	OLS regression Logistic regression	Census tract	Youth Older adults Household income Immigrants	Bikeway length per area Bikeway length per roadway length Connectivity (link/node) metrics Distance to nearest bikeway
Kent and Karner	2019	Baltimore, MD	Observational comparison	Neigh-borhood	Race and ethnicity Poverty Access to a vehicle	Service area of low-stress bikeways Amenities within service area
Braun et al	2019	22 cities in the United States	Multilevel mixed-effect regression	Block group	Race and ethnicity Educational attainment Composite socioeconomic status index	Bikeway length per roadway length Distance reachable via bikeway Distance to nearest bikeway
Standen et al	2021	Sydney, Australia	Observational comparison	N/A	Gender Age Household income Educational attainment	Accessibility from predictive model
Ferenchek and Marshall	2021	29 cities in the United States	Kruskal-Wallis test Logistic regression Difference-in-difference analysis	Block group	Median household income Race and ethnicity	Bikeway length per roadway length

Table 1. *Existing literature on bicycle facility network distributions*

Bicycle facility distribution equitability

One of the first studies on this topic evaluated Portland, Oregon's current and proposed bicycle network (Dill and Hagerty 2009). Dill and Hagerty calculated a service metric of bikeway miles per square mile for each block group and within calculated service areas for schools, employment centers, parks, grocery stores, and transit stations. They also classified areas as below or above average for selected demographic metrics, including poverty, percentage of non-white residents, percentage of children, and percentage of older adults. They then identified areas for each demographic metric and service area type which were high on the scale of the demographic metric and poorly served by the bicycle network. They found that low-income block groups and service areas were generally better served by the bicycle network than average, while youth were generally poorly served. Race and age were not statistically significant to the level of network service.

Another early study (conducted in 2009 by Cradock et al) examined federal funding for bicycle and pedestrian projects at the county level from 1990 to 2004. They analyzed data for more than 3,000 counties in the United States, including almost 2,000 counties which had initiated a bicycle- or pedestrian-related project during the study period, and evaluated both the likelihood that a county had initiated a project and the funding that county had received as outcomes. They tested for differences in outcomes by county size, region, and demographics. The authors found that there were significant disparities in project initiation and funding. While non-metropolitan counties were less likely to have initiated a project, per-capita funding was higher for those counties that did have a project compared to urban counties. Similarly, counties in the Northeast region were more likely to have a project, but counties in the Western region had the highest per-capita funding on average. Most relevantly, Cradock et al found that there were disparities associated with social and economic factors. Counties with higher poverty levels and lower rates of educational attainment were less likely to have initiated a project during the study period, and low educational attainment was further associated with lower levels of per capita funding in counties that had initiated a project (Cradock et al 2009).

Similar to the 2009 study by Dill and Hagerty, a 2011 study by Deka and Connolly examined the distribution of bicycle facilities in the state of New Jersey. They used a one-way analysis of variance to examine the demographic characteristics of block groups located within one mile of on- or off-road bicycle facilities. This study found that block groups located close to facilities had, on average, populations with lower median incomes and higher proportions of nonwhite and non-English-speaking people and immigrants.

A study by Kent and Karner focused on the city of Baltimore, Maryland and evaluated the area from each neighborhood reachable using low-stress bikeways only (2019). Notably, unlike most of the literature, this study used calculated levels of traffic stress rather than bicycle facilities to evaluate access to high-comfort bicycling. Under this metric, a low-traffic local road would be considered low-stress, even if it did not have a dedicated bicycle facility, while a separated bike lane on a busy multi-lane road would be considered high-stress due to the higher volume of car travel. They analyzed the number of supermarkets, banks, pharmacies, and public libraries available within the service area for each neighborhood and developed an overall metric of accessibility for each neighborhood which they then compared to demographic distributions of neighborhoods with high concentrations of residents who were Black, living in poverty, or did not have access to a vehicle.

Kent and Karner found that while Black residents were more likely to live in neighborhoods with larger service areas accessible by low-stress bikeways, this did not translate to having access to larger numbers of businesses or more diverse business types. While this result indicates inequity in business accessibility in Baltimore, the use of level of bicycle traffic stress and business distributions in this study means that it does not meaningfully represent anything about investment in bicycling facilities, as the results are driven mostly by existing land use and roadway patterns rather than purposeful design of the cycling network. Despite these limitations, they were able to effectively use the results to suggest potential bicycle facility projects that would increase accessibility and equity in Baltimore, providing a method for project prioritization that can serve as a model for similar work in other cities (Kent and Karner 2019).

In the most geographically extensive study to date, Braun et al examined the relationship between socioeconomic factors and characteristics of bicycle facility access in 22 US cities using multilevel mixed-effects regression models (2019). They used three metrics of bicycle facilities as their dependent variables: coverage, or the meters of bike lane per kilometer of roadway in a block group; the distance reachable only via bike lane from each block group; and the distance to the nearest bike lane. They used sociodemographic characteristics such as race and ethnicity, educational attainment, and a composite socioeconomic status index as their independent variables, controlling for other variables such as population density, distance to downtown, and bike commuting levels and accounting for differences between cities by clustering standard errors at the city level (Braun et al 2019).

This analysis found that both higher educational attainment and socioeconomic status were significantly associated with higher access to bike lanes, while higher Latino populations were significantly associated with lower access. The authors concluded that "while the disparities observed in this analysis could reflect directly discriminatory practices, it is perhaps more likely that they result from institutional issues in planning and advocacy" (Braun et al 2019 p. 8). They suggest that planners should more intentionally incorporate equity considerations into bicycle facility projects going forward (Braun et al 2019).

International studies

The first study located outside the United States focused on Bogotá, Colombia's extensive "Cicloruta" bicycle network (Teunissen et al 2013). This study evaluated access to the Cicloruta and to the Ciclovía, Bogotá's temporary recreational street closure program, for neighborhoods classified as part of different socio-economic strata (SES; evaluated by income level). They found that access to the Cicloruta and Ciclovía was significantly higher for residents of high and middle SES, while bicycle mode share was significantly greater for low SES residents. The authors concluded that bicycle mode share was driven by necessity rather than access to cycling infrastructure, as lower-income residents were less likely to own a car. They suggested that expanding the Cicloruta and the Ciclovía program to lower-income areas of the city was therefore key to improving equity of access and serving the greatest possible number of bicycle users (Teunissen et al 2013).

In Canada, Fuller and Winters examined income inequalities in census tract-level Bike Score (of which cycling infrastructure, quantified as "Bike Lane Score", is a significant part) (2017). They developed a linear regression model with Bike Score as the outcome and income quintiles, city-level fixed effects, and covariates including unemployment rate, number of homeowners, number of recent immigrants, and population as predictors. Their analysis demonstrated consistent income inequality in Bike Lane Score, with the three highest income quintiles all showing a statistically significant increase compared to the lowest income quintile and census tracts in the highest income quintile having on average the highest Bike Lane Score (Fuller and Winters 2017).

In a subsequent study focused on three mid-sized Canadian cities, Winters et al examined the relationship between density of bicycle infrastructure and median household income (2018).

Bicycle infrastructure density was calculated as bikeway length divided by the total area of each "dissemination area". They calculated the density for DAs within each income quintile and found that in two out of three of the cities, lower-income quintiles had statistically significantly higher densities of bicycle infrastructure. They also conducted spatial analysis, performing a Bivariate Local Moran's I test to identify clusters of DAs that both were low income and had low densities of bicycle infrastructure (Winters et al 2018).

Parra et al conducted a study of neighborhoods in Bogotá, Colombia (2018). Using the Network Analysis tool in ArcGIS Pro, they calculated the shortest path from the centroid of each neighborhood to the nearest access point of Bogotá's cycling network. They found that the neighborhoods identified as the most disadvantaged by socioeconomic status were on average more than twice as far from the cycling network as the median distance for all neighborhoods and that the average distance from the cycling network decreased consistently as socioeconomic status increased (Parra et al 2018).

In Brazil, Tucker and Manaugh examined the extent of the cycling networks in Rio de Janeiro and Curitiba and how well the cycling networks in these two cities provided access to commercial destinations (2018). They used these two metrics as dependent variables for their regression analysis, with income quintiles in each city as the independent variable. They calculated accessibility to commercial areas using the Network Analysis tool in ArcGIS Pro and added a weighted cost to each segment based on the type of cycling infrastructure and roadway type, then compared the calculated weighted least cost path and shortest least cost path. They found that in both cities, neighborhoods in higher income quintiles had more bicycle facilities. Similarly, they found that increases in income at the neighborhood level were associated with greater access to commercial destinations along paths utilizing higher quality cycling infrastructure.

Longitudinal studies

Wang and Lindsey assessed the equitability of Minneapolis's bikeway network in 2010 and 2014 using Lorenz curves and Gini coefficients (2017). They found that the overall Gini coefficient for bikeway distribution in Minneapolis was 0.73 in 2010 and 0.55 in 2014, indicating that it had become more equitable during that time. They calculated the Gini coefficients for the bikeway distribution for selected study populations including: households without a vehicle, people of color, households with incomes below the poverty level, youth, and older adults. They found that bikeway distributions were more equitable than the average for people of color and households without a vehicle and less equitable than average for youth and older adults. Wang and Lindsey also used the Network Analysis tool in ArcGIS Pro to evaluate the areas reachable in a three mile trip from the centroid of each block group using low-stress bikeways and using all roadways. By comparing the two areas, they evaluated the accessibility by bikeway for each block group.

Houde et al studied the expansion of cycling networks in three urban regions in the Montreal area between 1991 and 2016 and the equity of network distribution (2018). They calculated multiple metrics for each of density, connectivity, and accessibility at five year intervals within that time period for each region. Density was calculated as the length of the cycling network divided by total surface area and the length of the cycling network divided by total roadway length. The connectivity metrics they calculated included the number of network links divided by the number of network nodes and the number of intersections in the network divided by total surface area, along with two other calculated connectivity metrics. They calculated accessibility using the Network Analyst tool in ArcGIS Pro. For each year in the study period, they conducted both classical and logistical regression models to evaluate the relationship between the calculated accessibility and the percentage of selected demographic groups in the population of the area: children, seniors, low-income people, and recent immigrants.

Houde et al found that while the cycling network in the Montreal area expanded by 162% over the study period, there was little increase in connectivity. From the regression analysis, they found that children experienced the greatest inequity in access to the network while low-income people had better access to the network compared to the population overall. This study did not examine the expansion of the cycling network as a possible driver of neighborhood change. However, they did note that the positive relationship between a greater presence of low-income people and greater access to the cycling network declined in strength over time. The authors offered possible gentrification of those areas as one hypothesis to explain that trend (Houde et al 2018, p. 18).

Standen et al took a different approach (2021). They used the results of a travel mode survey distributed to residents of Sydney, Australia to train a predictive transport mode choice model. They then used that model to estimate the impacts of two scenarios for future cycling infrastructure -- a single protected cycleway versus a complete network -- on accessibility and physical activity for different groups. They found that in the scenario of a single cycleway, gender and income were related to statistically significant differences in increases to accessibility, with accessibility increasing by greater amounts for men and people with higher incomes. However, in the scenario of a complete network, these disparities disappeared. The authors concluded that the accessibility benefits of bicycle infrastructure projects were not equitably distributed across population groups, that benefits were more evenly distributed when projects were part of a complete network, and therefore that “failure to provide a connected, low-stress bicycle network is an example of structural discrimination” (Standen et al 2021 p. 13).

Relationship to neighborhood change

Departing from the majority of the literature on this topic, which concerned the question of whether cycling infrastructure is distributed equitably rather than considering the possibility of cycling network expansion as a driver of neighborhood change, a 2016 paper by Flanagan et al was the first to associate bicycle facilities with gentrification. They analyzed cycling networks in Portland, Oregon and Chicago. They used the percentage white population, homeownership, educational attainment, median household income, age, and median home value as their metrics for neighborhood change. They conducted regression analysis using cycling infrastructure indices as their dependent variable and the demographic change metrics as their independent variables, controlling for population density, distance to downtown, and distance to transit. Flanagan et al found that cycling infrastructure investment was most strongly correlated with the distance to downtown and population density, but was also strongly correlated with an increase in median household income. Overall, they state that “in both cities, lower income or lower home value neighborhoods experiencing incoming populations with greater relative wealth are correlated with cycling infrastructure investment” (Flanagan et al 2016, p. 22).

Around the same time, Hirsch et al examined spatial patterns of bicycle facility installation and neighborhood demographic changes (2017). They evaluated changes in facilities and demographics between 1985 and 2010 in Birmingham, Alabama; Chicago, Illinois; Minneapolis, Minnesota; and Oakland, California. Bicycle lanes and off-street trails were evaluated using the total length within neighborhood boundaries. The socio-demographic variables they analyzed included poverty levels, unemployment, median income, housing occupancy rates, and percent white. They found that access to bicycle facilities increased during the study period, and that those increases were spatially clustered. They further found that facility installation was related to neighborhood change, with neighborhoods experiencing greater increases in facilities also experiencing greater increases in socioeconomic affluence (Hirsh et al 2017). These studies both suggest that there is evidence that installation of bicycle facilities may mirror where neighborhoods are experiencing socioeconomic changes. However, they do not consider the question of whether facility installation precedes neighborhood change or vice versa.

A 2021 paper by Ferenchek and Marshall was the first to quantitatively consider this causality dilemma. They examined longitudinal data from 29 U.S. cities between 2010 and 2019. Ferenchek and Marshall used income and percentage of the population that was non-white as their metrics for neighborhood change. They used miles of bicycle lane facilities divided by total roadway miles as their bicycle facility variable and census block group as their spatial unit. They first compared the distribution of bicycle facilities for block groups which were above and below the median income and median percentage POC for each city using Kruskal-Wallis test, to evaluate whether facilities were distributed equitably. They then used logistic regression models to examine the question of causality, using changes in race/income as the dependent variable and installation of bicycle facilities as the independent variable and then conducting the model again with the variables switched. In both cases they controlled for distance from downtown.

Ferenchek and Marshall found that lower-income white areas had a large increase in bicycle facilities over the study time period, while POC areas had a statistically significantly lower increase in bicycle facilities compared to white areas, regardless of income. However, these findings were not consistent across all cities and bicycle facility types. The regression analysis showed that distance to downtown was the variable most strongly associated with the installation of bicycle facilities. There were positive correlations between increases in income and increases in bicycle facilities which were weak but statistically significant when income was the independent variable; however, this correlation was not statistically significant when increase in bicycle facilities was the independent variable. Ferenchek and Marshall concluded that these results suggested that while bicycle facilities do appear to be inequitably distributed, there was only weak evidence for a causal relationship between installation of bicycle facilities and demographic changes within an area. They suggested that planners should question, "whether bike facilities are causing the SED changes, or whether planners are simply cognizant of changing SED characteristics in a neighborhood and prioritize those areas for installation" (Ferenchek and Marshall 2021, p. 13).

Use of property values as a metric for gentrification

The use of a hedonic pricing model to quantify the desirability of bicycle facilities in housing choice has been used in studies dating back to 2006, though not with the goal of evaluating the effect of bicycle facilities on neighborhood change. The first such study was conducted using data from the urban areas of Minneapolis and St. Paul and the surrounding suburbs (Krizek 2006). The goal of this study was to assess which type of bicycle facility was most preferred between on-street facilities, roadside shared use paths, and off-street facilities. Krizek calculated an OLS regression using the distance to the nearest bicycle facility of each type as the independent variable, property sale prices from 2001 as the dependent variable, and covariates including a neighborhood-level fixed effect. Krizek also used interaction terms to evaluate the effect of the location variables for urban and suburban areas. This study found that in the city, off-street facilities were positively associated with sale prices, while roadside facilities were negatively associated with sale prices and on-street facilities had no effect. In the suburbs, all bicycle facilities were associated with lower property sale prices (Krizek 2006).

Welch et al studied property sales in Portland, Oregon in an attempt to demonstrate that bicycle facilities were adding property value for the city (2016). They conducted a random effects spatial panel model incorporating both spatial lag and spatial error effects as well as temporal effects. The data used in this study was from sales between 2003 and 2013. They used the Network Analyst module in ArcGIS to calculate the distance from each property to the nearest bike lane, shared use path, and transit station as the independent variables of the study and included covariates of property and location characteristics. The authors found that local and regional shared use paths were positively associated with increased sale prices, with regional shared use paths having a substantial positive effect. However, they found that bike lanes were negatively associated with sale prices. The study found that for every one foot of increased distance to the nearest bike lane, the average sale price rose by almost $2.50 (Welch et al 2016).

Also focusing on Portland, Oregon, Liu and Shi examined property sale data between 2011 and 2013 (2017). They calculated pooled OLS hedonic price regressions for both single-family residential and multi-family residential properties using two key metrics: proximity to advanced bicycle facilities (defined as separated bike lanes and off-street paths) and density of advanced bicycle facilities within a half mile buffer, along with covariates of relevant property and neighborhood characteristics. They found that both of these key metrics were positively associated with property sale prices. The authors found that, as expected, there was spatial autocorrelation in the property sale data, and they ran a spatial lag model for single-family residential properties and a spatial error model for multi-family residential properties to account for this. Through these spatial regression models, they found that both proximity to and density of bicycle facilities were associated with higher sale prices for single-family residential properties, while only density of bicycle facilities was associated with higher sale prices for multi-family residential properties (Liu and Shi 2017).

In addition to the literature related to bicycle facilities, property values were a metric used by studies on green gentrification generally. One notable such study examined the relationship between urban trees and gentrification in Portland, Oregon (Donovan et al 2021). The authors examined data on housing sales between 1990 and 2019 and estimated a hedonic pricing model to isolate the effects of location on sale price. They then regressed neighborhood desirability against tree plantings over the 30 year period. This study found that tree plantings were associated with an increase in neighborhood desirability, with a time lag of six years following the tree planting. To evaluate their calculated metric for neighborhood desirability, Donovan et al calculated the correlation between that metric and changes to race, income, education level, and owner-occupancy during the time period. They found a high level of correlation, particularly with the increase in percentage of white residents and the increase of median household income. This is consistent with metrics used by other studies of neighborhood change and suggests that property values are effective as a proxy metric for neighborhood change.

Another study examined the relationship between house prices and proximity to Chicago's 606 trail, a "rails to trails" project on an old elevated rail line completed in 2015 (Smith et al 2016). Researchers from the Institute for Housing Studies at DePaul University used parcel-level administrative data to evaluate the change in house prices before and after trail construction. They performed a regression analysis using distance from the 606 trail as the independent variable. The researchers also compared the neighborhoods of 606 East and 606 West. 606 East started with higher housing prices and a more stable, stronger market, while 606 West started with lower housing prices and a more volatile market. They found that while housing prices in the 606 East neighborhood did not change substantially following the opening of the trail, the trail opening was associated with a $100,000 increase in housing prices in the lower income neighborhood of 606 West. This study indicates that not only was active transportation infrastructure associated with increases in housing costs in this instance, it was specifically associated with housing cost increases in a lower income area more vulnerable to gentrification (Smith et al 2016).

A similar study focused on the Atlanta Beltline. The authors used data on property sales obtained from Fulton County. They examined the increase in median sale prices between 2011 and 2015 in areas within a half mile of the Beltline and calculated the cumulative appreciation rate over the study period. They found that the cumulative expected change in price for homes within a half mile of the Beltline was between 50.1% and 58.8% (depending on Beltline segment), compared to a 32.2% expected appreciation rate for homes located farther than a half mile from the Beltline. In their regression model, they included interaction terms representing the combined effects of Beltline proximity and the year which indicated a statistically significant relationship between Beltline proximity and cumulative appreciation rate. This study concluded that large-scale adaptive reuse projects like the Atlanta Beltline can quickly lead to increases in housing costs in surrounding neighborhoods and that planners should therefore address housing affordability concerns in the early stages of project development (Immergluck and Balan 2017).

Noting that housing sale prices are significantly spatially autocorrelated, some studies used spatial lag models or geographically weighted regressions to evaluate the relationship between green amenities and housing prices. One study, conducted in 2012 by Wenjie Wu, focused on housing prices in Beijing, China. Wu compared the results of an ordinary least squares regression with a geographically weighted regression model. The study used proximity to parkland and the size of the nearest park as the independent variables and controlled for proximity to schools and transit along with neighborhood demographics and crime rates. The results of the study demonstrate that proximity to parkland (particularly if the size of the parkland is greater) is associated with higher housing prices. Wu highlights the benefits of using a geographically weighted regression approach to account for local variability in the spatial autocorrelation of housing prices. However, Wu also states that even if a GWR approach results in a better fit for the model, this is not the same thing as a model that better represents the actual conditions in the study area (Wu 2012).

A 2014 study examining housing prices in the Jönköping region of Sweden used a similar geographically weighted regression approach (Nilsson 2014). The author used proximity to the nearest area of open space and the size of that open space parcel as the explanatory variables in their regression model and controlled for parcel characteristics such as lot size, distance to the nearest urban center, and neighborhood characteristics including median income, population density, and crime. Nilsson found that proximity to open space was an important determinant of housing prices, but that this relationship was not geographically uniform. The association between open space proximity and increased housing values was significantly higher in areas where open space was relatively scarce and population density was high (Nilsson 2014).

Key Takeaways

The existing literature on this topic illustrates the emergent nature of the field of study. Notably, there is no one metric for bicycle facility coverage, connectivity, or accessibility that is considered to be the standard. Each individual study chooses the metric that aligns best with their specific research question and methods. Similarly, there were different metrics for neighborhood characteristics and covariates chosen across the literature. The use of different methods highlights the difficulty of trying to model a complex real world system. For instance, studies that used spatial methods typically relied on observational comparisons rather than evaluations of statistical significance. Finally, there was not one clear conclusion across the literature when it came to assessing the relationship between bicycle facilities and neighborhood change or even neighborhood demographics.

This thesis contributes to the existing literature in the following ways. First, it focuses on the bicycle facility network in Boston, MA, which has experienced substantial growth in recent years, is now one of the most comprehensive networks in the United States, and has not yet been quantitatively evaluated. Second, it uses assessed property values as a proxy for neighborhood change. Only a few prior studies have evaluated hedonic pricing models incorporating metrics of bicycle facilities. Finally, in considering the direct relationship between bicycle facility installation and neighborhood change, my research joins only three prior studies addressing this question.

Chapter 3: Data and Methods

This thesis relies primarily upon a panel regression analysis of bicycle facility lane miles and assessed property values to evaluate the relationship between bicycle facility installation and neighborhood change, as represented by increasing property values. The scope of the study covers the City of Boston geographically and a period of 12 years (from 2008 to 2019) temporally. This chapter first discusses the data and methods used to assemble the dataset used for the analysis, then walks through the process of completing the panel regression model, then reviews supplementary analysis methods.

Data Sources

Data	**Format**	**Source**	**Link**
Existing Bike Network 2023	Shapefile (lines)	City of Boston	https://bostonopendata-boston.opendata.arcgis.com/datasets/5cf9533364da4033b0b3bcfa55a1fee5_0/explore
Parcels	Shapefile (polygons)	City of Boston	https://bostonopendata-boston.opendata.arcgis.com/datasets/6a76c94092574188b08a4d7f9ee5bd81_0/explore
Census Tracts	Shapefile (polygons)	City of Boston	https://hub.arcgis.com/datasets/boston::2020-census-tracts-in-boston/about
Property Assessment	Tabular (CSV)	City of Boston	https://data.boston.gov/dataset/property-assessment
American Community Survey Data	Shapefiles (polygons)	US Census Bureau	https://www-census-gov.ezproxy.library.tufts.edu/geographies/mapping-files/time-series/geo/tiger-data.2010.html#list-tab-1656998034
Open Space	Shapefiles (polygons)	City of Boston	https://bostonopendata-boston.opendata.arcgis.com/datasets/boston::open-space-5/about
MBTA Stations	Shapefile (points)	MassGIS	https://www.mass.gov/info-details/massgis-data-mbta-rapid-transit
Schools	Shapefiles (points)	City of Boston	https://bostonopendata-boston.opendata.arcgis.com/datasets/public-schools/explore https://bostonopendata-boston.opendata.arcgis.com/datasets/non-public-schools/explore

Table 2. *Data sources with format and hyperlinks*

The first data source used was a series of datasets of assessed property values from the City of Boston's Assessing Department. Staff in the Assessing Department are required by law to assess each parcel in Boston by January 1 each year. They do this by evaluating recent property sales in the area, drive-by assessments of the physical characteristics of the property (including size and condition), and any additions or renovations to the property in the last year. In Boston, assessors also inspect properties which have not been inspected or sold in the prior ten years ("Assessing" 2024). The Assessing Department makes the resulting assessed values for each parcel available online through the "Analyze Boston" open data initiative. Datasets are available for each year as .csv files. There are approximately 99,000 parcels in the City of Boston. The number of parcels listed for each year varied within a deviation of 500 parcels but did not decrease or increase over time. All parcels in the City are included in property tax assessments, including privately-owned residential and commercial properties, tax-exempt properties owned by non-profits (including churches), and publicly-owned parcels.

The bicycle facility dataset used in this thesis was available as a shapefile from the City of Boston. This dataset contained information on all bicycle facilities in Boston as of January 2023 and included the year of installation and facility type for each segment. Segments represented unbroken line features; if a bike lane crossed an intersection, the sections of bike lane on either side of the intersection were considered to be different segments. This meant that bike lanes on a single roadway were often made up of multiple segments. In 2023, there were 3,243 bicycle facility segments, ranging in length from 2.3 feet to 14,315.5 feet. The facilities included in the dataset included high comfort bike facilities such as shared use paths and buffered bike lanes as well as low comfort facilities like shared lane markings.

This dataset was first compiled in 2013 based on prior data from the Boston Bikes team. Boston Bikes, formerly known as the Boston Bicycle Advisory Committee, was a group of local cycling enthusiasts who advocated for the construction of more bicycle facilities and for the implementation of a bikeshare system, worked with the City to prioritize maintenance projects, and hosted events to encourage and teach Boston residents how to bike (Bike Attorney 2024). Boston Bikes is now a subset of the City of Boston's Transportation Planning department. The Transportation Department updates the dataset every year to address errors and add newly constructed projects. The dataset is cross-referenced with other city planners and the City's Construction team for accuracy (Fink 2024).

Installation year was not available for all segments (514 segments out of 3,243 were missing installation year). I filled in the missing data points as much as possible using Google Maps street view. Street View provides photographic snapshots of roadway conditions at given points in time. In Boston, there are multiple Street View snapshots available for any given point, sometimes with as great a frequency as every year. I reviewed the Street View snapshots available for locations with bicycle facilities without installation years. The majority of missing installation years were for facilities that had been in place before 2007 (the first installation year included in the dataset). Therefore, if I observed that the bicycle facility was visible on a Street View image from 2005, for instance, I was able to confidently list an installation year of "before 2007" for that facility. As the earliest year of my study period was 2008, this was sufficient for my analysis. In less than 100 cases, the facility was not present in the 2007 Street View snapshot. By reviewing the snapshots available for later years in that location, I could pinpoint when a facility first appeared on a Street View snapshot and therefore when it was constructed.

The City of Boston's Open Data portal was the data source for other shapefiles used in this analysis as well. These included a shapefile of polygons of parcel boundaries within the City of Boston. This shapefile was used for spatial calculations of bicycle facility coverage, distance to downtown, and distance to transit and for spatial joins to Census block group and Census tract level information, detailed below. It was available from the Assessing Department through the "Analyze Boston" initiative. Both this shapefile and the .csv files of assessed property values included the variable "Parcel ID", which allowed me to join the two. I also used a shapefile of 2020 Census tract boundaries provided by the City of Boston, which I used for summary calculations of neighborhood characteristics.

The "Analyze Boston" initiative was also the source of a polygon shapefile of open space in Boston and a point shapefile of Boston schools (both public and private). The open space dataset was compiled and is maintained with yearly updates by the City of Boston Parks and Recreation Department. It includes both recreational open space such as plazas, parks, reservations, playgrounds and athletic fields, beaches, community gardens, and cemeteries and conservation open space preserved due to its ecological significance and open to the public for limited use, such as marshes, rivers, and forested areas. The schools dataset was compiled based on information from the Massachusetts Department of Elementary and Secondary Education. It was most recently updated in 2018. This dataset includes listings for schools serving all grades from pre-Kindergarten to high school. It includes public, private, and charter schools along with approved special education programs.

I used information on MBTA rapid transit stations available from the Boston Region Metropolitan Planning Organization's Central Transportation Planning Staff through MassGIS. This data was available as a point shapefile with points for each station on the Red, Green, Orange, Blue, and Silver lines. I used five-year estimates from the American Community Survey for the demographic variables included in my analysis. This data was available from the US Census as polygon shapefiles corresponding to block group boundaries and including the ACS data in accompanying tables. For each year, I used the five-year estimate that included that year as the middle year in the estimate. For example, I used the American Community Survey five-year estimate for 2015-2019 as my data for 2017. As five-year estimates were only available through 2021, this limited my dataset to the years 2008 to 2019. The ACS data I used corresponded to questions B02, B19, and B25. B02 asks respondents to specify their race. B19 asks about household income. B25 asks about housing characteristics including occupancy and home ownership status.

Variables

The dependent variable for the regression model was the total assessed property value of the parcel. This variable was included in the datasets from the Assessing Department. I generated a metric of bicycle facility coverage to serve as the independent variable in the model. This was a continuous metric from zero to one that represented the percentage of roadway miles within a half-mile buffer of each parcel that had bicycle facilities. For parcels with bicycle facilities available on all or almost all of the nearby roadways, this metric would theoretically be close to a value of one, indicating that close to 100% of roads within a half mile buffer of the parcel had bicycle facilities. Similarly, a calculated metric of zero indicated that no roads within a half mile buffer of the parcel had bicycle facilities. I considered using the Network Analysis tool in ArcGIS Pro to generate a metric for the distance a bicyclist could reach using the bicycle network from each parcel but chose to go with the coverage metric to better answer my research question, which concerned the visible presence of bicycle facilities in an area rather than the useability of the network.

For the purposes of this analysis, I chose to include all types of bicycle facilities in my calculation of facility coverage. This included both high-comfort bicycle facilities such as separated and buffered bicycle lanes and shared use paths and low-comfort bicycle facilities such as sharrows (road markings indicating that bicycles can ride with car traffic). This choice was again driven by the use of a coverage metric rather than a useability metric. Though sharrows provide less connectivity and accessibility for bicyclists of all ages and abilities, they are nonetheless considered to be part of the bicycle network and contribute to the presence of bicycle facilities in the area. It therefore made sense to include them in a metric assessing that presence.

As property values are influenced by many different factors, I included a wide set of covariates in my analysis at the parcel, block group, and Census tract levels. I chose these covariates based on the characteristics that were found to relate to housing costs and/or bicycle facility distribution in my literature review. Parcel-level variables from the Assessing Department datasets included parcel area in square feet, building age (calculated from the "year built" variable in the Assessing Department datasets), building square footage, the number of floors in the building, the number of bedrooms in the building, the number of bathrooms in the building, the number of units in the building, and whether the building was owner-occupied. Other parcel-level variables included distance to downtown and distance to the nearest transit station. These were straight-line distances calculated in ArcGIS Pro. I chose to use straight-line distances rather than roadway distances calculated through Network Analyst due to limitations on the calculations possible for this analysis due to the size of the dataset. The use of a straight-line distance is also consistent with the use of presence rather than usability for the bicycle facility coverage metric. The distance to downtown used a point in Boston Common close to the Park Street MBTA station as “downtown”. This was chosen as Park Street is the station MBTA trains consider to be “inbound” and trains traveling away from Park Street in any direction are considered to be heading “outbound”. Boston Common also functions as a central park for the city and is close to both the Massachusetts State House and Boston City Hall.

Block group-level variables included the percentage of the population that identified as white alone, median household income, the percentage of housing units that were renter-occupied, and the percentage of housing units that were vacant. All of these variables were from the American Community Survey 5-year estimates and were normalized based on the values for total population or total housing units given for each question. These variables were frequently chosen as indicators of neighborhood change in the literature. As neighborhood change is a cyclical process and demographic changes can in turn be drivers of increasing housing costs, it was important both to include demographic variables in my regression model and to relate my analysis to any demonstrated changes in neighborhood demographics within my study area.

Census tract-level variables included the population density, number of schools, and square miles of open space. Population density was calculated using the value for Census tract population from the American Community Survey and the land area from the Census tract shapefile. Number of schools and square miles of open space were summarized by Census tract in ArcGIS Pro to represent neighborhood characteristics. Finally, Census tract ID was included in the final dataset for the regression analysis and used to cluster error terms by neighborhood.

Methods

I first cleaned the datasets from the Assessing Department to ensure that the formatting was consistent from year to year. I then joined the data for each year to a shapefile of parcels accessed from MassGIS using Parcel ID. This process created fifteen shapefiles, one for each year between 2008 and 2023, each containing assessed property values and other parcel-level characteristics for all Boston parcels (such as year of construction and building size). From the bicycle facility dataset, I used the "Select by Attribute" and "Export Feature" tools in ArcGIS to generate distinct shapefiles for each year with only the facilities that had installation years earlier than that year and therefore were assumed to be present at the beginning of that year. This ensured that the bicycle facility data for each year was appropriately aligned with the assessed property values for that year, which were published in January.

To develop the bicycle facility coverage metric, I generated a half-mile buffer around each parcel. The half-mile extent was chosen as it represents the distance approximately within a 15-minute walk of each parcel, which is the extent typically considered in the literature as the area a typical person is likely to cover walking to complete everyday activities. For earlier years in my study period, I was able to use the ArcGIS tool "Summarize Nearby" to calculate the length of bicycle facilities within each buffer area. However, the quantity of both parcels and facilities in later years caused the tool to fail. I therefore wrote a Python script (included in the Appendix) to perform the calculation outside of ArcGIS Pro for later years, then joined the Python output to the ArcGIS parcel datasets using parcel ID. I used a similar process to calculate the roadway miles within the half mile buffer of each parcel. I divided bicycle facility miles by roadway miles to create the continuous metric of bicycle facility coverage for each parcel.

I then calculated a straight-line distance from each parcel to my "downtown" point using the "Euclidean Distance" tool in ArcGIS. I similarly calculated a straight-line distance from each parcel to the closest MBTA rapid transit station. I used the "Summarize Within" tool in ArcGIS to summarize the area of green space and the number of schools within each Census tract then used a spatial join to join the characteristics of the Census tract shapefile to the parcel shapefiles. I joined the block group-level data from the American Community Survey to the parcel shapefiles using spatial joins in ArcGIS. As many of these processes were repetitive over the series of parcel datasets, I used ModelBuilder in ArcGIS to ensure consistency.

I exported the data to Stata to perform the statistical analysis. I first encoded any categorical variables such as the land use code and status of owner occupancy from the property assessment database. I then removed any publicly owned parcels and any parcels that did not contain any buildings. The starting dataset included 98,920 parcels with data for at least one year. After excluding these parcels, the dataset contained records for 93,419 parcels. I further excluded any parcels that were not zoned residential use only. This included parcels zoned for single-family residential, multi-family residential, condominium, and apartment uses but not those zoned for mixed uses. As factors influencing the valuation of commercial parcels may not be the same as those affecting residential parcels, this ensured that the regression model compared parcels which were as similar as possible. Additionally, my research question concerns neighborhood change driven by increasing housing costs, so it made sense to focus my analysis on residential parcels. This resulted in records for 72,285 residential parcels with assessment data for at least one year. The most frequent other land uses were tax-exempt use (8,154 parcels), commercial use (4,078 parcels), and industrial use (914 parcels).

I then addressed duplicates in the dataset. I removed duplicates in terms of all variables using the “duplicates drop” command. There were 3,941 identical duplicates in the database. I further evaluated the duplicates in the database by examining records for the same parcel ID in the same year. I observed that a common duplicate pattern was to show records for the same parcel in the same year split over multiple lines, so that fields that were blank in the first instance of the record included data in the second instance of the record and vice versa. Additionally, starting in 2012, the assessor’s office began including listings for both entire buildings and individual units, all under the same parcel ID number. This resulted in some parcels with up to 60 duplicates in a year for different units in the same building with large variability in assessed property value, square footage, and other variables. Failing to address these duplicates would have introduced inaccuracies into the regression model, as it would have resulted in entire buildings and single units being compared to each other. To address both of these issues, I collapsed the dataset using the “max” function. This kept only the maximum value for any given field in a set of duplicates. As the values for the listings for the full building were by necessity greater than the listings for a single unit, this kept only the record for the full building. It also replaced any blank values with existing values when combining a listing that was split over two lines. This resulted in a dataset including 65,318 parcels with data for at least one year.

I identified anomalous values in the data, including instances of typos and unusual increases from year to year, then filled in corrected values where possible or removed obviously incorrect values. This resulted in changes to almost a thousand data points but did not change the number of parcels included in the analysis. As many parcels included in the dataset were missing listings for one or more years, the dataset was considered to be unbalanced. Using unbalanced data in a panel regression increases the error caused by the differences between units (in this case, parcels). As the panel regression assesses the effects of the independent variable on the dependent variable over time within each unit, if some units only have a few data points, the regression results may be less accurate. I therefore limited the dataset to only those parcels that had complete listings from the Assessing Department for each year between 2008 and 2019. The final dataset used in this analysis consisted of assessed property values and parcel characteristics for 60,433 parcels geographically distributed across the City of Boston. Each parcel had data for each of the twelve years in the study period.

Panel Regression Model

Panel data, also known as cross-sectional time-series data, consists of data points for the same entities over time. In the case of this project, the entities are City of Boston parcels, with data points for each year. The purpose of a panel regression model is to control for variables not otherwise accounted for in the analysis related to the differences between entities (entity fixed effects) and the passage of time (time fixed effects). The fixed effects model accounts for characteristics of individual entities that are not random and may be related to the dependent variable but are not available to be included as a control variable in the regression model (such as the condition of a house: a house in better condition is likely assessed at a higher value, but condition data is subjective and largely unavailable). Including time fixed effects accounts for variables which change over time and affect all entities equivalently but may similarly be difficult to include in a regression model, such as relative desirability of property in the City of Boston.

A different model of panel regression includes entity random effects rather than entity fixed effects. This means that the unobserved individual effects included as entity-specific errors in the model are assumed to be uncorrelated with any other independent variables. Logically, it makes more sense to use a fixed effect model for this study, as I would expect additional characteristics of parcels to be related to the characteristics already included in my model. For example, the condition of a house may be correlated with the age of the house. I used the Hausman test to evaluate this. This test compares the coefficients from a fixed effect model and a random effect model of the same variables and tests whether the differences between the coefficients are systematic. The null hypothesis is that they are not. Conducting this test for my model resulted in a p-value close to zero, indicating that I should reject the null and use a fixed effects model, as the individual characteristics are correlated with the regressors.

I also used a Wald test to evaluate whether I needed to include time fixed effects in my model. This is a joint F-test which evaluates the likelihood that the coefficients for all time variables are jointly equal to zero. The null hypothesis is that they are, which would indicate that there are no unaccounted-for effects of the passage of time. I used the "testparm" command in Stata to conduct this test on the coefficients for my year variables. The test resulted in a p-value close to zero, indicating that I should reject the null and include time fixed effects in my analysis.

The equation for a panel regression model including both entity and time fixed effects is: $y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + \dots + \beta_n x_{nit} + \delta_t + u_i + \varepsilon_{it}$, where y is the outcome, α is an entity-specific constant, β is the coefficient for each variable, x is the variable (for each entity and time), δ is the coefficient for the time regressors representing the time fixed effects, u is the within-entity error term representing the entity fixed effects, and ε is the overall error term. The subscript i indicates the entity and the subscript t indicates the year. Including the fixed effects in the equation means that the interpretation of the coefficient is that it represents a common effect across entities controlling for all control variables and individual and time heterogeneity.

By default, the fixed effects model assumes that entity characteristics captured in the error term are correlated with the independent variables but that entity fixed effects are not correlated between entities. If this is not true, then the model needs to be adjusted to be robust to heteroskedasticity and autocorrelation. As the entity fixed effects included in my model were in part considered to be related to the location of the parcel, they are likely to be correlated between entities. This was confirmed with a Breusch-Pagan test. I therefore clustered the error terms included in my model by Census tract in an attempt to account for the spatial autocorrelation. Finally, as normally distributed variables are an assumption of regression models, I used the log form of several of the variables to normalize the distributions and establish a stronger model. These variables included the total assessed property value (the dependent variable in the model), bicycle facility coverage (the independent variable in the model), building square footage, number of floors, nearby open space, and percentage of vacant housing.

Spatial Clustering

I used LISA, or local indicators of spatial autocorrelation, to evaluate clusters of similar values for facility coverage, assessed property values, and demographic variables including median household income and the percentage of the population that identified as white alone. Local Moran's I is a statistic of the level of local spatial autocorrelation of a given variable. It evaluates the value at a given location along with the values of its neighbors and calculates a measure of spatial autocorrelation. This measure describes whether the spatial distribution of values is statistically significantly different from a random distribution, which would be expected if there is no spatial autocorrelation. If the Moran's I value is positive, this indicates that values at that location are clustered -- a high value at the location is surrounded by high values at neighboring locations or a low value at the location is surrounded by low values at neighboring locations. If the Moran's I value is negative, this indicates that values at that location are different from values at neighboring locations to a statistically significant extent -- a high value at the location is surrounded by low values at neighboring locations or vice versa. In practice, this analysis results in a map showing the extent to which the values of a given variable are spatially autocorrelated and statistically significant hot spots and cold spots. This allowed me to identify areas with more or less bicycle facility coverage compared to the overall facility coverage for Boston, for instance.

Bivariate Local Moran's I uses a similar method to evaluate spatial autocorrelation between two variables. Clusters identified through this method show locations where high values of the first variable are surrounded by high values of the second variable and low values of the first variable are surrounded by low values of the second variable. I used Bivariate Local Morans' I clustering to identify locations where low values of facility coverage were surrounded by low values for the percentage of the population that identified as white alone. These are locations of equity concerns, as they represent areas of the city with both high proportions of people of color and less access to bicycle facilities. As the parcels included in my dataset were not uniformly distributed across the geographical area of the city, I chose to use a distance band of a quarter mile for both the Univariate and Bivariate Local Moran's I calculations. This meant that each parcel was compared to all other parcels within a quarter mile in all directions, regardless of the number of neighbors.

Chapter 4: Results

Descriptive Statistics

The dependent variable for this analysis is the assessed property value from the City of Boston's Assessing Office. Descriptive statistics for this variable for the time period of the study are shown in the table below.

AV_TOTAL	Mean	Median	Standard Deviation	Minimum	Maximum
Overall	***$525,614.30***	***$384,000***	***$765,458***	***$8,767***	***$98,100,000***
2008	$492,410.40	$412,200	$522,509	$9,212	$37,500,000
2009	$450,734.40	$352,100	$551,156.20	$9,212	$54,400,000
2010	$422,934.10	$322,700	$544,407.40	$8,874	$56,700,000
2011	$417,786.70	$316,100	$546,478.40	$8,767	$56,000,000
2012	$425,507.10	$317,500	$582,405.90	$8,767	$55,200,000
2013	$445,599.20	$327,400	$638,073.40	$8,767	$60,700,000
2014	$469,391.20	$337,900	$704,821.10	$10,040	$66,300,000
2015	$517,169.40	$371,400	$779,018.60	$10,556	$77,900,000
2016	$597,444.10	$429,500	$889,656.10	$12,857	$87,500,000
2017	$651,745	$467,400	$982,756.40	$14,336	$97,400,000
2018	$683,378.30	$490,600	$998,258.10	$13,711	$90,900,000
2019	$733,281.50	$532,200	$1,057,352	$13,712	$98,100,000

Table 3. *Descriptive statistics for assessed property values throughout the study period*

On average, assessed property values fell between 2008 and 2011 and have steadily increased since. The average percentage change from the prior year for each year supports this analysis, showing that the average percentage change was negative until 2012. It was highest in 2016 but has remained above zero since 2012. Over the time period of the study, the mean of the percentage change was 3.4%, and the median was 3%. Over all years, there were many positive outliers, with some parcels seeing percentage changes much higher than the average.

As the average assessed property value has increased, so has the standard deviation. This indicates that there is a wider spread of values in later years. Some of this increase is due to a disproportionately greater increase in value over time for large luxury apartment buildings such as The Colonnade Residences at 118 Huntington Avenue in Boston's Back Bay neighborhood, which was consistently the parcel with the maximum value. (It is worth noting that while The Colonnade and other luxury apartment buildings like it were consistently the parcels with the greatest total assessed values, they were not the parcels with the greatest assessed values by area. The parcel with the greatest assessed value per square foot for all years was the Brandywyne Village Apartment Complex, a mixed-income development in East Boston. Besides the Brandywyne Complex, the parcels with the greatest assessed value per square foot were historic single-family row houses in Beacon Hill. These parcels similarly experienced disproportionately greater increases in value over time compared to the average.) This disparity is illustrated by the increasing positive skew of the data, as the difference between the mean and median for each year increases over time.

facilitycoverage	**Mean**	**Median**	**Standard Deviation**	**Minimum**	**Maximum**
Overall	***7.3%***	***6.9%***	***5.8%***	***0%***	***38.0%***
2008	2.5%	0.9%	3.2%	0%	22.0%
2009	2.7%	0.9%	3.5%	0%	26.4%
2010	3.2%	1.7%	3.7%	0%	29.8%
2011	5.4%	4.8%	4.4%	0%	33.3%
2012	6.6%	6.0%	4.8%	0%	33.3%
2013	7.3%	6.9%	5.2%	0%	33.3%
2014	8.9%	8.8%	5.2%	0%	34.5%
2015	9.1%	9.1%	5.3%	0%	34.5%
2016	9.4%	9.4%	5.4%	0%	34.5%
2017	10.0%	9.8%	5.9%	0%	38.0%
2018	10.5%	10.4%	6.0%	0%	38.0%
2019	11.1%	10.9%	6.1%	0%	38.0%

Table 4. *Descriptive statistics for the calculated facility coverage metric throughout the study period*

The independent variable for this analysis was bicycle facility coverage. Descriptive statistics for this variable for the time period of the study are shown in Table 4. The average facility coverage has been increasing over the study period. The data is positively skewed, with some parcels having significantly higher facility coverage than the average. The number of parcels with zero percent facility coverage is also decreasing, from 22,390 parcels in 2008 to 1,967 parcels in 2019. The largest increases in average facility coverage occurred between 2010 and 2011 and 2013 and 2014.

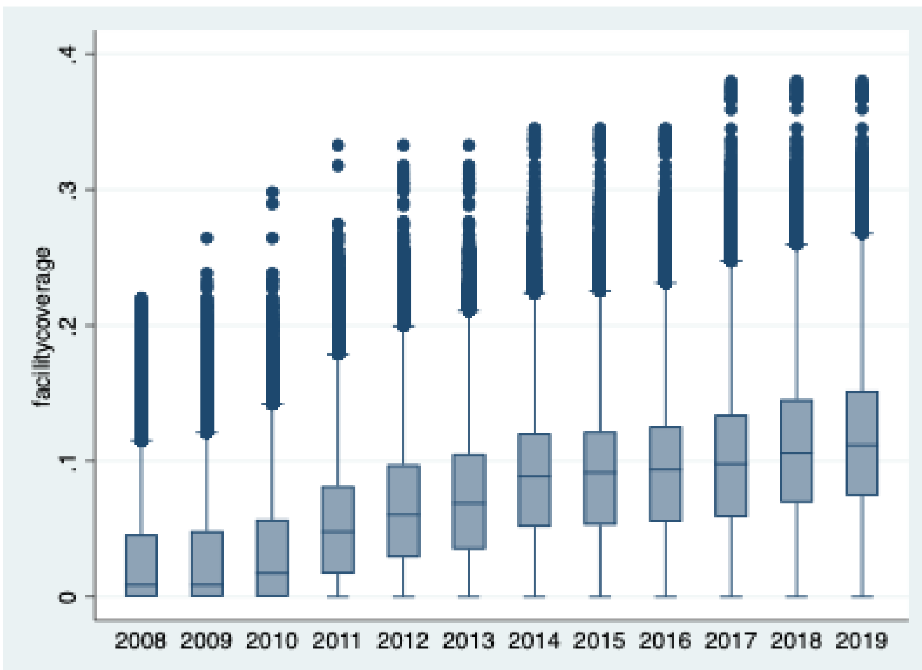

Figure 1. *Boxplots of the calculated facility coverage metric, 2008-2019*

The descriptive statistics for the control variables in this analysis are included in the table below. There were a total of 725,196 observations in the dataset, corresponding to 60,433 parcels over twelve years. Variables showing numbers of observations less than 725,196 in the table were missing observations for some parcels for some years. All of the control variables except for the demographic variables are time invariant. The average percentage of vacant units in the Census tract decreased between 2008 and 2015, from 8.467% in 2008 to 6.871% in 2019. The average Census tract median income increased over the study period, from $62,318.69 in 2008 to $101,504.30 in 2019. There were no significant trends in percentage of rental units, percentage of the population that identified as white alone, or population density.

Variable Name	**Variable Description**	**Mean**	**Median**	**Standard Deviation**	**Minimum**	**Maximum**	**# of Observations**
LAND_SF (evaluated as natural log)	Lot size in square feet	4,908.24 ft^2	4,379 ft^2	11,019.86 ft^2	375 ft^2	1,603,642 ft^2	719,517
AGE	Age of building	101.45 years	109 years	29.20 years	3 years	319 years	718,457
LIVING_AREA (evaluated as natural log)	Building area in square feet	2,734.67 ft^2	2,265 ft^2	3,752.58 ft^2	0 ft^2	475,102 ft^2	724,607
NUM_FLOORS (evaluated as natural log)	Building height in stories	2.26	2	0.73	0	34	722,423
NEAR_DIST	Straight line distance to Boston Common in miles	4.47 mi	4.44 mi	2.07 mi	0.15 mi	9.20 mi	725,196
MEDIAN INCOME	Block group median income from ACS	$75,855.81	$68,750.00	$36,834.76	$2,499.00	$250,001.00	696,450
NUMSCHOOLS	Number of schools within the Census tract	1.62	1	1.35	0	6	725,196
OPENSPACE_ SQMI (evaluated as natural log)	Area of open space (conservation or recreational) within the Census tract	0.04 mi^2	0.01 mi^2	0.07 mi^2	0 mi^2	0.39 mi^2	725,196
TRANSIT_DIST	Straight-line distance from the nearest MBTA rapid transit station in miles	0.98 mi	0.65 mi	0.87 mi	0.01 mi	3.81 mi	725,196
LU_RESSF	Dummy variable indicating if the land use code is single-family residential	0.53	1	0.50	0	1	725,196
OWNEROCC	Dummy variable indicating if the property is occupied by the owner	0.68	1	0.47	0	1	725,196

TOTAL_RMS	Total rooms in the building	9.94	9	3.85	0	80	680,611
BDRMS	Total bedrooms in the building	4.57	4	1.97	0	18	680,611
BTHRMS	Total bathrooms in the building	2.33	2	0.92	0	21	680,611
UNITS	Total number of units on the parcel	1.94	2	1.27	0	7	725,196
PERCWHITE	Percentage of block group population that identifies as white alone from ACS	53.0%	58.0%	32.7%	0%	100%	725,162
PERCRENTAL	Percentage of housing units in block group that are rental units	51.2%	52.8%	23.3%	0%	100%	725,162
PERCVACANT (evaluated as natural log)	Percentage of housing units in block group that are vacant	7.1%	6.1%	6.7%	0%	50.9%	725,162
POPDENSITY	Census tract population density	20,072 people/mi^2	15,479 people/mi^2	15,600 people/mi^2	0 people/mi^2	167,008 people/mi^2	721,072

Table 5. *Control variables included in the linear and panel regression models*

Spatial Clustering

Comparing the average facility coverage for quintiles of demographic variables in 2008 and 2019 shows that there is a trend for average facility coverage by quintile of percentage white alone from the beginning and the end of the study period. In 2008, the bottom two quintiles of the percentage of the population that identified as white alone had significantly lower average facility coverage than the top three quintiles. In 2019, that disparity had disappeared. A similar trend is evident in comparisons of percentage white alone by quintiles of facility coverage in 2008 and 2019. In 2008, the percentage of the Census tract population that identified as white alone increased significantly by quintile of facility coverage, while in 2019, there was no longer an evident correlation.

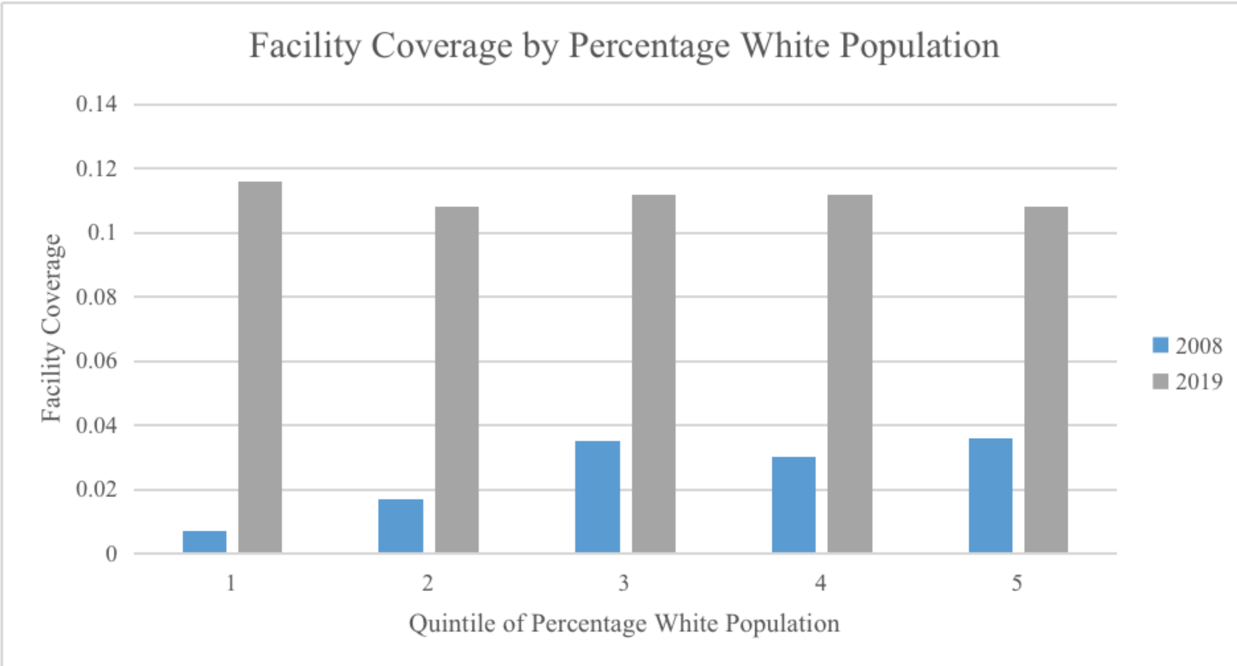

Figure 2. *Mean facility coverage for quintiles of percent white, 2008 and 2019*

Examining the Local Moran's I cluster maps for bicycle facility coverage and percentage white population in 2008 and 2019 shows low-low clusters of bicycle facility coverage in Dorchester and Mattapan in 2008, in the same area as low-low clusters of percentage white alone. This is confirmed by the Bivariate Local Moran's I cluster analysis, which shows clusters of low facility coverage surrounded by low percentage white alone in that area. In 2019, those same parcels did not show a statistically significant difference in bicycle facility coverage compared to the average. The Bivariate Local Moran's I cluster map for 2019 shows outliers of high facility coverage surrounded by low percentage white alone in Dorchester and Mattapan.

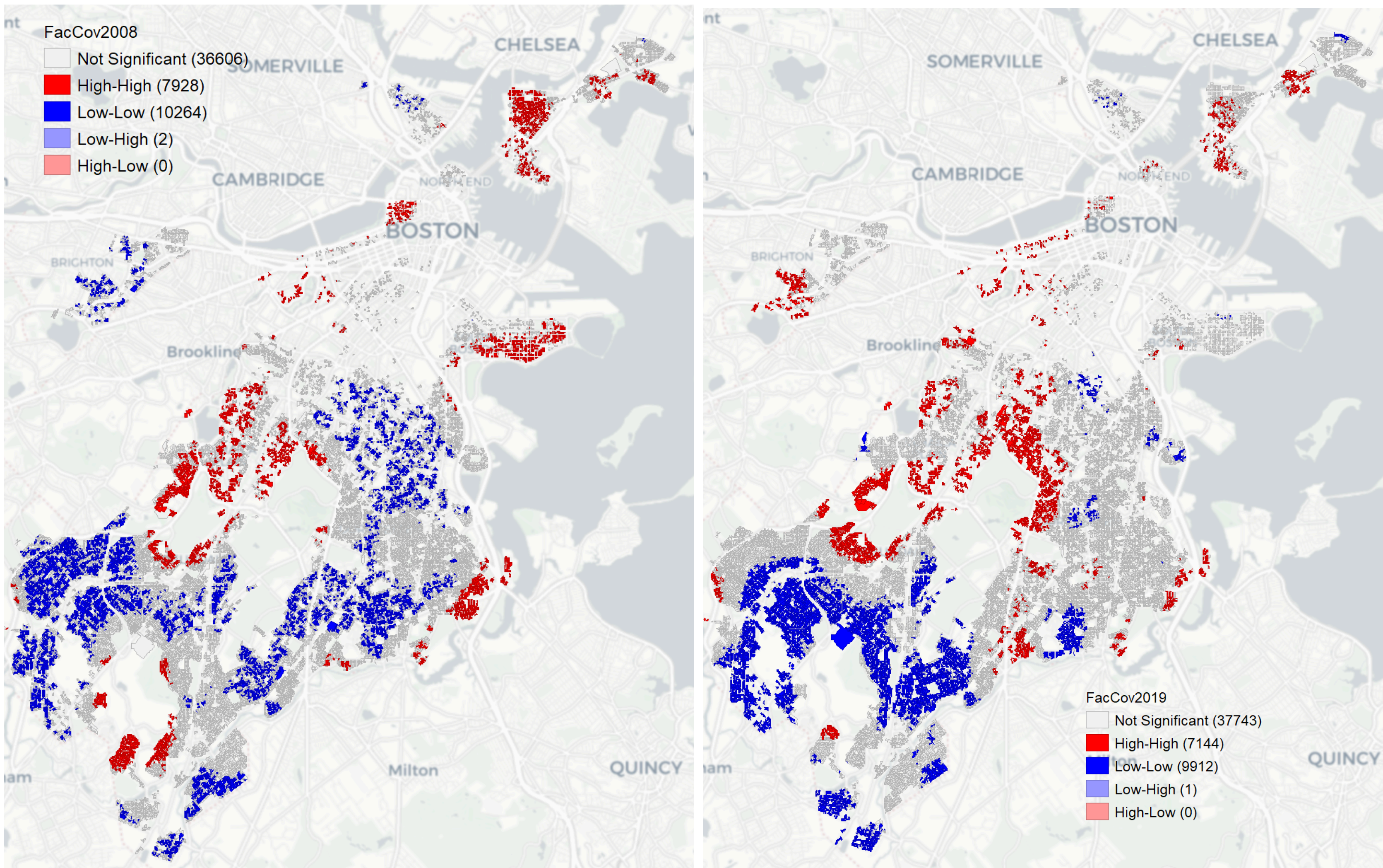

Figure 3. *Clusters of Facility Coverage, 2008 (left) and 2019 (right)*

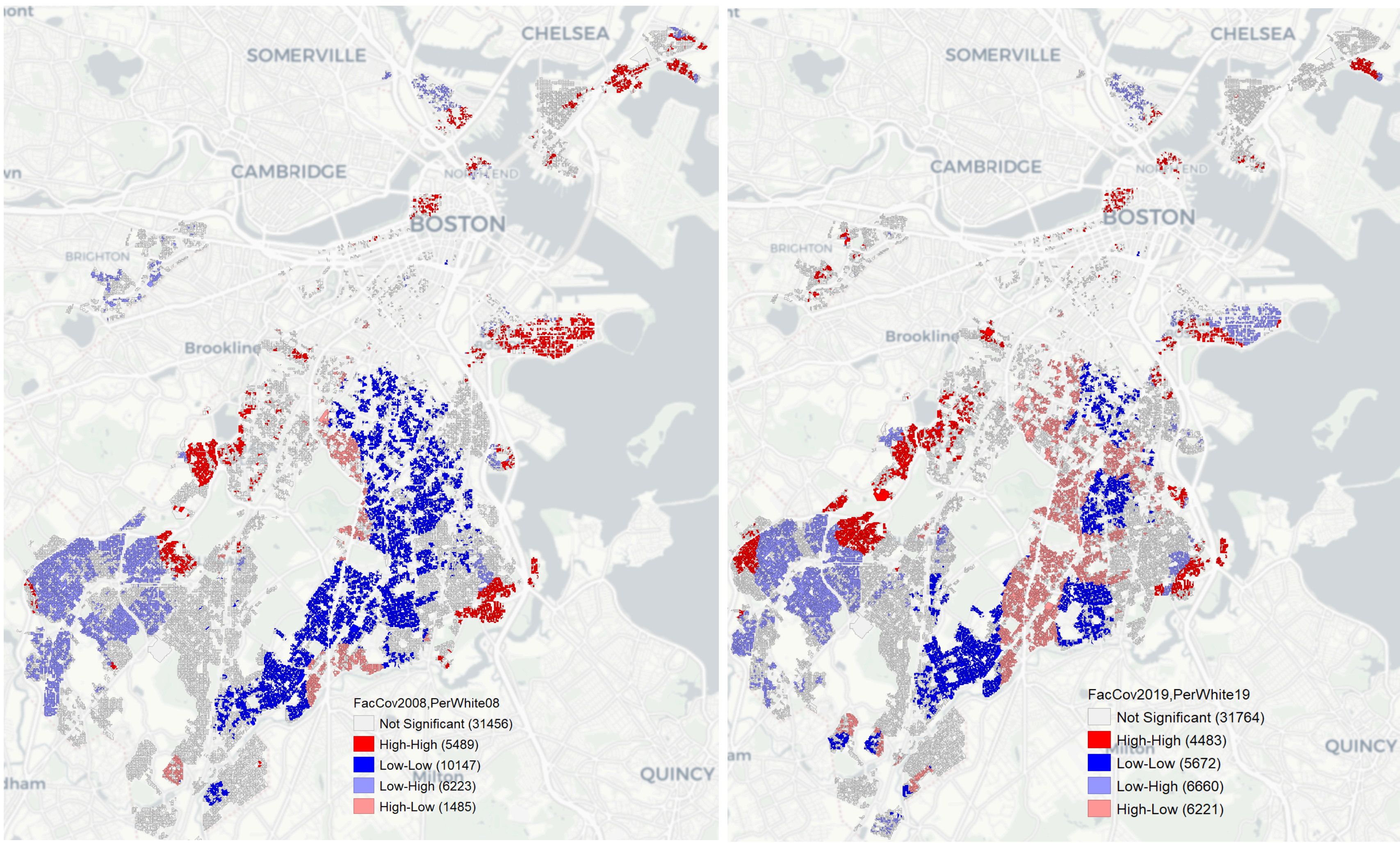

Figure 4. *Bivariate Clusters of Facility Coverage and Percentage White Population, 2008 (left) and 2019 (right)*

Comparing trends of average assessed housing value to quintiles of facility coverage shows an increasing disparity. In both 2008 and 2019, parcels with greater facility coverage had higher assessed housing values. In 2019, this inequity was greater. Finally, examining the parcels with zero facility coverage for each year also shows some trends. Between 2008 and 2019, the percentage of rental units, percentage vacant units, and percentage white population all decreased for parcels with zero facility coverage. These changes are more than likely driven by the expansion of bicycle facilities from downtown, meaning that parcels with zero facility coverage are increasingly located in areas away from the city center. In 2008, parcels with zero facility coverage were located on average 5.3 miles away from downtown. In 2019, they were located on average 7.5 miles away from downtown.

Facility Coverage Quintile	**Mean Assessed Value, 2008**	**Mean Assessed Value, 2019**
1	$419,409.10	$504,075.80
2	$408,891.70	$598,086.50
3	$456,443.40	$695,291.60
4	$539,622.70	$832,803.30
5	$629,071.30	$1,078,021

Table 6. *Mean assessed property value for quintiles of facility coverage, 2008 and 2019*

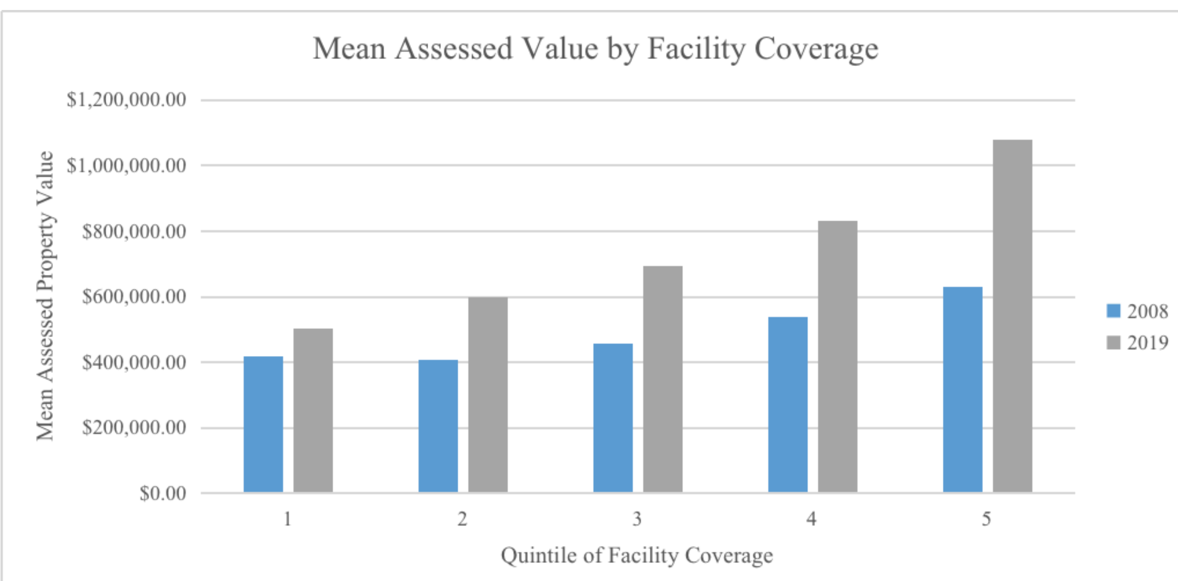

Figure 5. *Mean assessed housing value by quintile of facility coverage, 2008 and 2019*

Linear Regression

OLS Regression						
	Time effects not included			Time effects included		
R^2	0.646			0.713		
Prob > *F*	0.000			0.000		
RMSE	0.292			0.263		
Variable	**Coefficient**	**Standard Error**	***p*-value**	**Coefficient**	**Standard Error**	***p*-value**
ln_facilitycoverage	0.862	0.166	0.000	-	-	-
ln_land_sf	0.077	0.015	0.000	0.077	0.015	0.000
age	-	-	-	-	-	-
ln_living_area	0.543	0.040	0.000	0.564	0.042	0.000
ln_num_floors	0.065	0.023	0.004	0.076	0.024	0.001
near_dist	-	-	-	-	-	-
medianincome	0.000	0.000	0.000	0.000	0.000	0.000
numschools	-	-	-	-	-	-
ln_openspace	-0.022	0.009	0.018	-0.021	0.009	0.025
transit_dist	-	-	-	-	-	-
lu_ressf	-	-	-	-	-	-
ownerocc	-0.021	0.006	0.001	-0.019	0.006	0.002
total_rms	-0.013	0.002	0.000	-0.013	0.002	0.000
bdrms	-	-	-	-	-	-
bthrms	0.134	0.018	0.000	0.132	0.019	0.000
units	-0.114	0.022	0.000	-0.123	0.022	0.000
percwhite	0.430	0.043	0.000	0.537	0.043	0.000
percrental	0.216	0.068	0.002	-	-	-
ln_percvacant	-	-	-	-	-	-
popdensity	-	-	-	-	-	-
2009				-0.128	0.009	0.000
2010				-0.226	0.013	0.000
2011				-0.252	0.015	0.000
2012				-0.243	0.017	0.000
2013				-0.230	0.019	0.000
2014				-0.197	0.022	0.000
2015				-0.125	0.023	0.000
2016				-	-	-
2017				0.065	0.024	0.008
2018				0.094	0.025	0.000
2019				0.160	0.027	0.000

Table 7. *OLS regression model results*

The results of an ordinary-least-squares linear regression on the entire dataset are presented in Table 7. The results were sensitive to the inclusion of time effects, which accounted for 10% of the total variance in assessed property values. When time effects were included, facility coverage was not statistically significant. Errors were clustered by census tracts in an attempt to account for spatial autocorrelation. The model was run for a time-lagged value of facility coverage (with time lags ranging from one to five years), but the time lag did not significantly affect the results.

Besides facility coverage, most other variables were not sensitive to the inclusion of the time effects, which makes sense as most of the variables included in the model were time invariant. Percentage of rentals was the only other variable to change status of statistical significance between the models. To consider the fit of the OLS regression model across the years without the time effects, I ran the regression for each year. The values for the goodness of fit metrics (R^2, RMSE, and *F*-statistic) did not vary greatly from year to year from the values listed for the time-effects-included model above. Facility coverage was not statistically significant for any year.

Panel Regression

As the model includes both time and entity fixed effects, it supports the use of a panel regression model. The results of the panel regression model are below. In this model, distance to downtown, distance to transit, number of schools, and nearby open space are omitted as they are time invariant entity fixed effects and thus captured by the entity fixed effect variable. Facility coverage is not statistically significant in this panel regression model. If the model is run without the time fixed effects, facility coverage is statistically significant, but the within-R^2 value decreases to 0.330. This indicates that a large proportion of the change in total assessed value for each parcel over time is due to time effects which are not captured by the model.

Fixed Effects Panel Regression Model			
R^2	Within: 0.787	Between: 0.169	Overall: 0.296
Prob > *F*	0.000		
Rho	0.958		
Variable	**Coefficient**	**Robust std. err.**	***p*-value**
ln_facilitycoverage	-	-	-
ln_land_sf	0.178	0.060	0.003
age	-	-	-
ln_living_area	0.236	0.019	0.000
ln_num_floors	0.129	0.026	0.000
medianincome	0.000	0.000	0.000
lu_ressf	-	-	-
ownerocc	-0.005	0.002	0.031
total_rms	-0.006	0.002	0.001
bdrms	0.024	0.002	0.000
bthrms	0.070	0.005	0.000
units	-	-	-
percwhite	-	-	-
percrental	-	-	-
ln_percvacant	-	-	-
popdensity	-	-	-
2009	-0.139	0.008	0.000
2010	-0.232	0.012	0.000
2011	-0.259	0.015	0.000
2012	-0.254	0.017	0.000
2013	-0.229	0.019	0.000
2014	-0.120	0.020	0.000
2015	-0.105	0.021	0.000
2016	-	-	-
2017	0.112	0.022	0.000
2018	0.160	0.023	0.000
2019	0.231	0.023	0.000

Table 8. *Panel regression model results*

In an attempt to evaluate the model for similar parcels over time, I divided the dataset into quintiles of 2008 assessed property values and ran the panel regression model for each group. Each quintile contained between 11,100 and 11,500 parcels. The R^2 value for the model progressively increased from 0.732 for the bottom quintile to around 0.84 for the top two quintiles. Facility coverage continued to not be statistically significant except for the bottom quintile of 2008 assessed property values. For the bottom quintile, a 1% increase in facility coverage is associated with a 0.8% increase in the total assessed property value, controlling for all covariates and entity and time fixed effects. I also ran the model with facility coverage as the dependent variable and total assessed property value as the predictor value to evaluate whether assessed property values were statistically significantly related to facility coverage, which might suggest that facility coverage is higher in areas with greater property values. Assessed property value was not statistically significant in this model.

Chapter 5: Discussion

The analysis in this thesis demonstrates that assessed property values increased for parcels in Boston between 2008 and 2019, mainly in the last five years of the study period. Bicycle facility coverage also steadily increased during this twelve-year span, as the city worked to bring the vision of a complete bicycle network to fruition. This increase in coverage was not uniformly felt across all parcels. In 2008, Dorchester and Mattapan, both areas of the city with lower than average percentages of residents who identified as white alone, had statistically lower than average bicycle facility coverage. These neighborhoods have been historically under-resourced and de-prioritized when it came to infrastructure investments of all types. This long-time neglect, driven by systemic racism, is likely the cause of the inequitable distribution of the bicycle network in 2008. By 2019, focused investment in the bicycle network had greatly reduced this disparity. Though there were still pockets of lower-than-average facility coverage in these neighborhoods, for the most part, levels of facility coverage matched those in other areas of the city.

The linear regression model showed that the increase in assessed property values in Boston over the study period was largely driven by unobserved variables associated with the passage of time. These time fixed effects represent the forces driving property values and housing costs throughout metro-Boston. All years except 2016 were statistically significantly related to assessed value. Years before 2016 were associated with lower assessed property values, while years after 2016 were associated with higher assessed property values, controlling for all other variables. With the time effects included, the model was a strong fit for the data. Most of the variables which were statistically significantly related to assessed property values were parcel characteristics, particularly the area of the lot, the square footage of the building(s), and the numbers of rooms and units. This makes sense, as it is expected for the size of a building to be strongly correlated with its value.

Controlling for these characteristics, percentage white alone and median income were both statistically significant; however, facility coverage was not. The significance of percentage white alone and median income is likely due to the neighborhood in which the parcel is located affecting the assessed value. Again due to systemic racism and historic disinvestment, neighborhoods with lower median incomes and higher percentages of people of color tend to have lower property values. If an increase in facility coverage for the same parcel over time was strongly associated with an increase in the assessed property value, I would have expected facility coverage to be statistically significant and positively correlated with total assessed value. As it is not, I conclude from the linear regression alone that property values in Boston are largely driven by parcel and building size along with neighborhood location and are increasing due to the unquantifiable regional forces associated with increasing property values over time.

The panel regression model supports this interpretation. Controlling for both time and parcel fixed effects resulted in a model that was strongly fitted to the data; the included predictor and control variables explained approximately 79% of the variance in assessed property values within panels. 96% of the variance between panels was explained by unobserved parcel fixed effects, which in this case could be location or other parcel characteristics that were not included in the analysis. All years except 2016 were statistically significantly related to assessed property value. As in the linear regression, years before 2016 were negatively related to assessed value while years after 2016 were positively related to assessed value, indicating that property values changed with the passage of time. The other variables which were statistically significant were parcel characteristics related to size such as lot size, building area, and total number of rooms, median household income for the block group, and whether or not the parcel was owner-occupied. Though median household income was statistically significant, the coefficient was close to zero. This, along with the non-statistical significance of the other included demographic variables, suggests that changes in neighborhood demographics in Boston during this time period were not associated with changes in assessed property values. In this model as well, bicycle facility coverage was not statistically significant. This suggests that an increase in facility coverage was not associated with any changes in assessed property values.

These results are consistent with those in the existing literature. Most studies evaluating bicycle network distributions found that there were inequities in the distribution of existing networks but that it varied greatly by area. In Boston, the network was distributed inequitably at the beginning of the study period, but that inequity has been reduced through focused investment. Studies addressing gentrification found that there was some evidence that neighborhoods experiencing greater levels of investment in bicycle facilities also experienced greater demographic changes, but that the question of causality was inconclusive. Likewise, the conclusions of studies conducting hedonic pricing models varied greatly between area and time period studied and precise study method. I found that neighborhood changes in Boston were mainly related to larger regional trends. The stronger impact of time-fixed effects in this study and weak relationship between neighborhood change and bicycle facility investment in the literature explains the disagreement with existing literature on the question of green gentrification.

Limitations & Recommended Future Studies

There were a number of sources of uncertainty in this analysis, starting with the choice of variables. Most hedonic pricing models in the literature used data from property sales rather than property assessments. This is because sale data is a more accurate indicator of consumer desires, while assessment data in part represents the judgment of the assessor. As property assessments are in part based on prior property sales in the area, trends in assessed values tend to lag behind trends in sale values. Assessed values also tend to be lower than sale values and may be artificially low due to political pressures to reduce property taxes. I chose assessed property values as they were available for each year of my study area and allowed for the use of a panel regression model. However, all of these factors could contribute to inaccuracies in my results. A similar study in the future might be better served by the use of property sale data and a linear regression model.

There is also the question of whether property values are an accurate representation of neighborhood demographic change in Boston during this time period. While common knowledge and my literature review suggest that increasing property values are linked to rising housing costs, and rising housing costs are linked to the displacement of lower-income people, particularly people of color, each of these assumptions may not hold for this particular place and time. I attempted to test the use of assessed property values as a proxy for neighborhood change by regressing total assessed value against median income, percentage white alone, and percentage of rentals without including parcel characteristics. All of the demographic variables were statistically significant in this model, but only 36% of the variance in assessed values was explained. It is therefore recommended that further studies focus on evaluating the relationship between bicycle facilities and change in neighborhood demographic variables more directly.

Additional uncertainty was introduced by the design of the facility coverage metric used in this study. First, all bicycle facilities were included in the calculation of the facility coverage metric. This included sharrows, which were typically excluded in the literature as they do not provide mode separation from cars and are not high-comfort facilities. Recommended further research therefore includes recreating the analysis using high-comfort facilities only and comparing the results. Next, this analysis used the proportion of facility miles to roadway miles as the coverage metric. While this was a common coverage metric used in the literature, facility miles divided by calculation area (i.e. block group) and total facility miles within the calculation area were also commonly used. As I was performing this analysis on the parcel level, I chose to calculate this metric for a half mile buffer around each parcel. Most of the studies using this metric calculated it for each block group. The choice of a half-mile buffer rather than a quarter-mile or other distance therefore introduced uncertainty into the analysis. It is possible that the study would have resulted in different results had I chosen a different distance for the buffer. Repeating the analysis with a different distance buffer and comparing the results is therefore one recommendation for further research.

Another limitation of the facility miles to roadway miles proportion is that it is sensitive to variance in the length of the roadway miles within the calculation area. A calculation area containing one road with a sharrow would be considered to have 100% facility coverage, which may not be an accurate representation of the sense of bikeway presence within that area. Similarly, a calculation area with a dense network of roadways where most of the roadways were quiet local streets and the few high-traffic roadways had bike lanes would likely have a low value for facility coverage despite being a very bike-friendly area. The facility coverage metric cannot capture the real-world context of a parcel. One suggestion for further research would be to incorporate variables such as level of bike traffic stress, vehicle traffic counts, roadway condition, and grade of any hills to generate a composite index of cycling friendliness or bikeway presence for each parcel.

As demonstrated by the significant effects of time and parcel fixed effects in the various models, the forces driving property values are complex, regional, and often undefined. The inability to incorporate these variables beyond unobserved time effects introduced uncertainty into the analysis, as it was impossible to isolate the effects of these variables between different parcels. Similarly, there may have been parcel characteristics that would have been meaningful to include in the analysis that were not included in the available datasets. There were also some inaccuracies in the data used, particularly in the datasets accessed through the Assessing Department. I attempted to clean these datasets as best as possible, but the data inaccuracies as well as the assumptions made during the cleaning process introduced significant uncertainty into the analysis. It was particularly difficult to include the effects of location in the linear and panel regression models. Location has a significant effect on property values; one only needs to think of the common saying "Location, location, location" used in the home-buying process. While it was to some extent included in the parcel fixed effects term, this was not the most accurate possible way to account for the spatial autocorrelation of the data. Datasets with distinct values for each location can be analyzed using spatial regressions which include spatial lag or spatial error terms or using geographic weighted regressions which allow for variations in the fit of the model across the geographic area of the study. In future studies, I am highly recommending pursuing a multilevel mixed-effect model that includes time fixed effects, parcel fixed effects, and location effects through spatial statistics tools.

Finally, while many of the variables used in this analysis used parcel-level data, other variables were evaluated at the block group or Census tract levels. This introduced a great deal of uncertainty and possible error to the data. I chose to keep the analysis at the parcel level rather than aggregating to the smallest spatial unit available for all variables (Census tract) due to the high variance in parcel-level characteristics within each Census tract, which would be lost in a higher-level analysis. Ideally, all of the variables included in the analysis would be evaluated on the parcel level. While this may not be possible for demographic data accessed through the US Census, variables like nearby green space or number of schools could have been aggregated separately for each parcel using the same half-mile buffer used in the calculation of facility coverage. I recommend evaluating these variables at the greatest level of detail possible in future studies.

Chapter 6: Conclusions and Recommendations

My research question concerned the relationship between the installation of bicycle facilities and neighborhood change as represented by change in assessed property values. To do this, I examined changes in parcel characteristics over time using descriptive statistics and spatial clustering and completed linear and panel regression models. Overall, I did not find enough evidence to conclude that bicycle facilities are associated with higher property values. Variance in assessed property values between parcels was mainly determined by lot and building size, while change in assessed property values for the same parcel over time was driven by unobserved regional trends.

At the same time, the equity of the bicycle network in terms of low-income neighborhoods and neighborhoods with high proportions of people of color increased significantly between 2008 and 2019. This was due in large part to the City's investment in bicycle facilities in the historically underserved neighborhoods of Dorchester and Mattapan. While there are still some neighborhoods of the city with no or little facility coverage, these neighborhoods tend to be wealthier and whiter and therefore not areas of concern for the equity of the bicycle network.

Boston's bicycle network expansion is far from over. Among other projects, the city is planning large-scale corridor makeovers of Blue Hill Avenue and Columbia Road in Dorchester and Mattapan, in part to include bicycle facilities. These projects represent significant investments of city resources (including financial) and large amounts of energy and time on behalf of city staff and community organizers. Beyond those projects, there are almost one hundred and fifty miles of bicycle facilities to be installed on Boston streets as part of the network expansion. How can this thesis support these projects and access to bicycle infrastructure for all city residents?

Residents of low-income neighborhoods targeted for bicycle infrastructure improvements are concerned that the installation of these facilities will contribute to the rising housing costs they are already experiencing. This ongoing context of gentrification in Dorchester and Mattapan is critical in interpreting the results of this analysis. Though there is little evidence that bicycle facilities specifically contribute to higher property values, the historic lack of investment and care for marginalized community members in low-income neighborhoods gives the impression that any project is intended to benefit higher-income white people moving into the neighborhood. Additionally, as increasing property values are in large part driven by unknown regional forces, it is not possible to definitively dismiss concerns that the overall trend of greater investment in active transportation is contributing to rising housing costs across the entire metro area. Furthermore, whether or not they are correlated with bicycle facility installations, housing costs are still rising in Boston, and low-income residents are at risk of being pushed out of the city. It is entirely possible that the proposed bicycle facility expansions will end up serving mostly higher-income and white people due to larger ongoing patterns of displacement, even if the network expansion is not a specific cause.

It is therefore essential that city planners and policy makers take care to involve community members in developing hyper-local plans for bicycle facility expansion rather than solely relying on overarching goals, and that they collaborate with efforts to combat larger trends of gentrification. Residents are the experts on the needs of their communities, and many residents of low-income neighborhoods want and welcome bike lanes. Engaging both community members who are enthusiastic and those with concerns will strengthen the overall implementation of the network expansion by more precisely responding to community needs. It is also critical to recognize that bicycling is not the answer for all transportation needs and to continue pursuing the network expansion in tandem with improving transit options, pedestrian facilities, and traffic flow. Finally, transportation officials should support the city's efforts to stabilize low-cost housing options and address issues of displacement and extreme housing cost burdens. Policies focused on addressing one area of inequity cannot succeed in their goals without being placed in the context of the larger system.

If the City of Boston follows this recommendation, there is good cause to be cautiously optimistic. The work done on the bicycle network expansion so far has greatly increased the equity of the network distribution and thus far has not been shown to directly contribute to gentrification. Transportation officials can continue working to make low-stress biking accessible in all areas of the city with a reasonable degree of confidence that they are not doing unintended harm. Bicycle lanes are an important tool for environmental and transportation justice. By placing this tool within a larger strategy of equity-oriented planning policies, the City of Boston can ensure that they will truly serve cyclists of *all* ages, abilities, and backgrounds.

Appendix

Python code used to calculate facility coverage

```python
import geopandas as gpd


# Load the polygons and lines shapefiles

polygons = gpd.read_file(r'H:\Thesis\HalfMiBuffers.shp', usecols=['MAP_PAR_ID', 'geometry'])

lines = gpd.read_file(r'H:\Thesis\BikeNetwork2019.shp', usecols=['geometry'])


# Create spatial index for faster spatial operations

polygons_sindex = polygons.sindex

lines_sindex = lines.sindex


# Find the intersection between polygons and lines

intersection = gpd.overlay(polygons, lines, how='intersection', keep_geom_type=False)


# Calculate the length of the intersecting lines within each polygon

intersection['length'] = intersection.geometry.length


# Group by polygon ID and sum the lengths of intersecting lines

polygon_lengths = intersection.groupby('MAP_PAR_ID')['length'].sum().reset_index()


# Save polygon_lengths as a CSV file

polygon_lengths.to_csv('bikenetwork2019_lengths.csv', header=True, index=False)
```

Bibliography

"Assessing | Boston.Gov." 2016. January 27, 2016. https://www.boston.gov/departments/assessing.

"Bike Data | Boston.Gov." 2018. January 22, 2018. https://www.boston.gov/departments/boston-bikes/bike-data.

Blok, Anders. 2020. "Urban Green Gentrification in an Unequal World of Climate Change." Urban Studies 57 (14): 2803–16. https://doi.org/10.1177/0042098019891050.

Boone, Christopher, Geoffrey Buckley, J. Morgan Grove, and Chona Sister. n.d. "Parks and People: An Environmental Justice Inquiry in Baltimore, Maryland." Annals of the Association of American Geographers. Accessed November 20, 2023. https://www.tandfonline.com/doi/abs/10.1080/00045600903102949.

Braun, Lindsay M., Daniel A. Rodriguez, and Penny Gordon-Larsen. 2019. "Social (in)Equity in Access to Cycling Infrastructure: Cross-Sectional Associations between Bike Lanes and Area-Level Sociodemographic Characteristics in 22 Large U.S. Cities." Journal of Transport Geography 80 (October):102544. https://doi.org/10.1016/j.jtrangeo.2019.102544.

Collins, Alessandro and Rigolon, Timothy. 2022. "The Green Gentrification Cycle." Urban Studies, September. https://journals.sagepub.com/doi/full/10.1177/00420980221114952.

Cradock, Angie L, Philip J Troped, Billy Fields, Steven J Melly, Shannon V Simms, Franz Gimmler, and Marianne Fowler. 2009. "Factors Associated with Federal Transportation Funding for Local Pedestrian and Bicycle Programming and Facilities." Journal of Public Health Policy 30 (1): S38–72. https://doi.org/10.1057/jphp.2008.60.

"Cycling In Boston: An Overview | Suffolk County Bicycle Injury Lawyer." n.d. Accessed May 31, 2024. https://www.bikeattorney.com/cycling-in-boston-an-overview.html.

Deka, Devajyoti, and Mary Connelly. 2011. "Does Proximity to Activity-Inducing Facilities Explain Lower Rates of Physical Activity by Low-Income and Minority Populations?" Transportation Research Record 2264 (1): 83–91. https://doi.org/10.3141/2264-10.

Dill, Jennifer, and Brendon Haggerty. 2009. "Equity Analysis of Portland's Draft Bicycle Master Plan – Findings." Urban Studies and Planning Faculty Publications and Presentations, September. https://pdxscholar.library.pdx.edu/usp_fac/135.

Donovan, Geoffrey H., Jeffrey P. Prestemon, David T. Butry, Abigail R. Kaminski, and Vicente J. Monleon. 2021. "The Politics of Urban Trees: Tree Planting Is Associated with Gentrification in Portland, Oregon." Forest Policy and Economics 124 (March):102387. https://doi.org/10.1016/j.forpol.2020.102387.

Ferenchak, Nicholas N., and Wesley E. Marshall. 2021. “Bicycling Facility Inequalities and the Causality Dilemma with Socioeconomic/Sociodemographic Change.” Transportation Research Part D: Transport and Environment 97 (August):102920. https://doi.org/10.1016/j.trd.2021.102920.

Fink, Nathaniel. 2024. Interview. By author. January 30, 2024.

Flanagan, Elizabeth, Ugo Lachapelle, and Ahmed El-Geneidy. 2016. “Riding Tandem: Does Cycling Infrastructure Investment Mirror Gentrification and Privilege in Portland, OR and Chicago, IL?” Research in Transportation Economics, Transportation and Land Development: A Global View, 60 (December):14–24. https://doi.org/10.1016/j.retrec.2016.07.027.

Fotheringham, Alexander, Chris Brunsdon, and Martin Charlton. 2002. “Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.” John Wiley & Sons 13 (January).

Fuller, Daniel, and Meghan Winters. 2017. “Income Inequalities in Bike Score and Bicycling to Work in Canada.” Journal of Transport & Health 7 (December):264–68. https://doi.org/10.1016/j.jth.2017.09.005.

Hirsch, Jana A., Geoffrey F. Green, Marc Peterson, Daniel A. Rodriguez, and Penny Gordon-Larsen. 2017. “Neighborhood Sociodemographics and Change in Built Infrastructure.” Journal of Urbanism: International Research on Placemaking and Urban Sustainability 10 (2): 181–97. https://doi.org/10.1080/17549175.2016.1212914.

Houde, Maxime, Philippe Apparicio, and Anne-Marie Séguin. 2018. “A Ride for Whom: Has Cycling Network Expansion Reduced Inequities in Accessibility in Montreal, Canada?” Journal of Transport Geography 68 (April):9–21. https://doi.org/10.1016/j.jtrangeo.2018.02.005.

Immergluck, Dan, and Tharunya Balan. 2018. “Sustainable for Whom? Green Urban Development, Environmental Gentrification, and the Atlanta Beltline.” Urban Geography 39 (4): 546–62. https://doi.org/10.1080/02723638.2017.1360041.

Kent, Margaret, and Alex Karner. 2018. “Prioritizing Low-Stress and Equitable Bicycle Networks Using Neighborhood-Based Accessibility Measures.” International Journal of Sustainable Transportation 13 (March):1–11. https://doi.org/10.1080/15568318.2018.1443177.

Krizek, Kevin. 2006. “Two Approaches to Valuing Some of Bicycle Facilities’ Presumed Benefits.” Journal of American Planning Association 72 (3): 309-320. https://brucefreemanrailtrail.org/pdf/krizek2.pdf.

Lee, Richard, and Ipek Sener. 2016a. “P24⁎ - A Review of Equity in Active Transportation.” Journal of Transport & Health 3 (2, Supplement): S76–77. https://doi.org/10.1016/j.jth.2016.05.027.

———. 2016b. “Transportation Planning and Quality of Life: Where Do They Intersect?” Transport Policy 48 (May): 146-155. https://www-sciencedirect-com.ezproxy.library.tufts.edu/science/article/pii/S0967070X16300968?via%3Dihub.

Liu, Jenny and Wei Shi. 2017. “Impact of Bike Facilities on Residential Property Prices.” Transportation Research Record: Journal of the Transportation Research Board 2662: 50-58. http://dx.doi.org/10.3141/2662-06.

“Luxury High-Rise Apartments in Boston’s Back Bay.” n.d. The Colonnade Residences. Accessed April 28, 2024. https://www.colonnaderesidences.com/.

Martínez, Jessica. 2023. “‘ARE WE JUST KILLING PEOPLE?’: Centering Racial Capitalism in the Green Gentrification of the Atlanta BeltLine.” International Journal of Urban and Regional Research 47 (3): 444–60. https://doi.org/10.1111/1468-2427.13154.

Mora, Rodrigo, Ricardo Truffello, and Gabriel Oyarzún. 2021. “Equity and Accessibility of Cycling Infrastructure: An Analysis of Santiago de Chile.” Journal of Transport Geography 91 (February):102964. https://doi.org/10.1016/j.jtrangeo.2021.102964.

National Low Income Housing Coalition. 2024. “Gentrification and Neighborhood Revitalization: WHAT’S THE DIFFERENCE?.” April 25, 2024. https://nlihc.org/resource/gentrification-and-neighborhood-revitalization-whats-difference

Nilsson, Pia. 2014. “Natural Amenities in Urban Space – A Geographically Weighted Regression Approach.” Landscape and Urban Planning 121 (January):45–54. https://doi.org/10.1016/j.landurbplan.2013.08.017.

Olivier, Shavel’le. 2022. “Opinion: Meet Residents Where They Are on Bike Lane Discussions.” Dorchester Reporter. September 8, 2022. https://www.dotnews.com/2022/meet-residents-where-they-are-bike-lane-discussions.

Oscilowicz, Emilia, Isabelle Anguelovski, Melissa García-Lamarca, Helen V. S. Cole, Galia Shokry, Carmen Perez-del-Pulgar, Lucía Argüelles, and James J. T. Connolly. 2023. “Grassroots Mobilization for a Just, Green Urban Future: Building Community Infrastructure against Green Gentrification and Displacement.” Journal of Urban Affairs 0 (0): 1–34. https://doi.org/10.1080/07352166.2023.2180381.

Parra, Diana C., Luis F. Gomez, Jose D. Pinzon, Ross C. Brownson, and Christopher Millett. 2018. “Equity in Cycle Lane Networks: Examination of the Distribution of the Cycle Lane Network by Socioeconomic Index in Bogotá, Colombia.” Cities & Health 2 (1): 60–68. https://doi.org/10.1080/23748834.2018.1507068.

Pistoll, Chance, and Anna Goodman. 2014. “The Link between Socioeconomic Position, Access to Cycling Infrastructure and Cycling Participation Rates: An Ecological Study in Melbourne, Australia.” Journal of Transport & Health, Walking & Cycling: The contributions of health and transport geography, 1 (4): 251–59. https://doi.org/10.1016/j.jth.2014.09.011.

Prelog, Rachel. 2015. “GIS Methods for Investigating the Equity of Access to Bike Infrastructure.” Commissioned by the League of American Bicyclists. September 2015.

Rigolon, Alessandro, and Jeremy Németh. 2018. “‘We’re Not in the Business of Housing:’ Environmental Gentrification and the Nonprofitization of Green Infrastructure Projects.” Cities 81 (November):71–80. https://doi.org/10.1016/j.cities.2018.03.016.

Standen, Christopher, Melanie Crane, Stephen Greaves, Andrew T. Collins, and Chris Rissel. 2021. “How Equitable Are the Distributions of the Physical Activity and Accessibility Benefits of Bicycle Infrastructure?” International Journal for Equity in Health 20 (September):208. https://doi.org/10.1186/s12939-021-01543-x.

Teunissen, Thijs, Olga Sarmiento, Mark Zuidgeest, and Mark Brussel. 2015. “Mapping Equality in Access: The Case of Bogotá’s Sustainable Transportation Initiatives.” International Journal of Sustainable Transportation 9 (7): 457–67. https://doi.org/10.1080/15568318.2013.808388.

Tinlin, Commissioner Thomas, Vineet Gupta, Alice Brown, Kris Carter, Nicole Freedman, Michelle Danila, Nick Jackson, Laurie Pessah, and Pete Robie. 2013. “Boston Bike Network Plan.” City of Boston Transportation Department. https://www.cityofboston.gov/images_documents/Boston%20Bike%20Network%20Plan%2C%20Fall%202013_FINAL_tcm3-40525.pdf.

Tucker, Bronwen, and Kevin Manaugh. 2018. “Bicycle Equity in Brazil: Access to Safe Cycling Routes across Neighborhoods in Rio de Janeiro and Curitiba.” International Journal of Sustainable Transportation 12 (1): 29–38. https://doi.org/10.1080/15568318.2017.1324585.

Wang, Jueyu, and Greg Lindsey. 2017. “Equity of Bikeway Distribution in Minneapolis, Minnesota.” Transportation Research Record 2605 (1): 18–31. https://doi.org/10.3141/2605-02.

Welch, Timothy, Steven Gehrke, and Fangru Wang. 2016. “Long-term impact of network access to bike facilities and public transit stations on housing sales prices in Portland, Oregon.” Journal of Transport Geography 54 (June): 264-272. https://doi.org/10.1016/j.jtrangeo.2016.06.016.

Winters, Meghan, Jaimy Fischer, Trisalyn Nelson, Daniel Fuller, and David G. T. Whitehurst. 2018. “Equity in Spatial Access to Bicycling Infrastructure in Mid-Sized Canadian Cities.” Transportation Research Record 2672 (36): 24–32. https://doi.org/10.1177/0361198118791630.

Woodard, Tiana. 2022. “In Mattapan, Bike Lanes Divide the Community: ‘They’re Just Trying to Push Us Out’.” The Boston Globe. August 1, 2022. https://www.bostonglobe.com/2022/08/01/metro/mattapan-bike-lanes-divide-community-theyre-just-trying-push-us-out/.

Wu, Wenjie. 2012. “Spatial Variations in Amenity Values: New Evidence from Beijing, China.” SERC Discussion Paper 113, June.