# The effect of bike lane infrastructure on urban housing markets

Bracken Woolley

The Effect of Bike Lane Infrastructure on Urban Housing Markets

by

Bracken Woolley

Honors Economics Thesis

Submitted to:

Economics Department
University of Richmond
Richmond, VA

April 27, 2018

Advisor: Dr. Timothy Hamilton

# The Effect of Bike Lane Infrastructure on Urban Housing Markets

Bracken Woolley

University of Richmond

April 2018

## Abstract

Across the United States, cities are increasing their level of investment in bicycle infrastructure. The environmental and health benefits of this infrastructure are clear, but less research has been conducted on its economic impacts. This study examines the effect of bicycle infrastructure, specifically bike lanes, on New York City housing markets. Specifically, I look at the impact of bike lane length on median rent and percent vacancy in a given census tract. In addition to the independent variable of focus, bike lane length, census-based data was used to control for other economic and demographic factors that could impact property values. To control for endogenous factors, model tracts with and without bike lanes were matched using a propensity score matching method. In the preferred model, results suggest that the addition of one standard deviation of bike lane meters to a census tract decreases median rent values by $29.97 in addition to raising vacancy rates, meaning that bike lane infrastructure has a negative effect on urban neighborhoods.

## Section 1: Introduction

The collective and individual benefits of biking abound. The prospect of improved health, reduced air pollution and eased congestion has encouraged cities and suburbs across America to increase their investment in biking, often by building more bike lanes. In over 60% of the 70 largest cities in the United States, biking is on the rise due to policies that cultivate bike-friendly streets. Atlanta, Portland, Baltimore and Austin are just a sample of the cities that have recently committed to pro-bike plans (Sisson, 2017).

New York City (NYC) is the United States' forerunner in bicycle-friendly policies and investments. With over 1,000 miles of bike routes, NYC has the largest network of bicycle lanes in North America and intends to continue installing at least 50 lane-miles of bicycle facilities a year ("Bicyclists: Building the Network," 2017). NYC also pioneered the model of constructing lanes between curbs and parking spots, which has since been copied across the country. The number of daily bike trips within the city has increased faster than population and employment, rising from 170,000 in 2005 to more than 450,000 today, instigated by better bicycle infrastructure and potentially the overcrowding of public transit. However, barriers still face the expanding network, mostly in the form of complaints about bicycle safety, decreased space for parking, and lost road space for cars (Hu, 2017).

Overshadowed by health and environmental benefits, the economic impact of biking is commonly overlooked in public debate. Linking bike lanes to economic factors is an opportunity to expand the conversation around biking as a method to lift up the "economic vitality" of neighborhoods across all income levels (New York City Department of Transportation, 2014). In addition, the connection between bicycle infrastructure and residential markets may indicate its role as a valuable positive externality. As the density of bike lanes in NYC and across the U.S.

increases, it is also important to investigate the possibility that bike lanes have a negative economic impact, potentially because of increased congestion, risk to safety or the attraction of different demographics to the area. If so, research should begin to explore methods to mitigate negative economic outcomes.

This study analyzes the relationship between the total length of bike lanes within a census tract and the census tract's economic development by focusing on housing markets. Indicators of economic development in this study will include two main proxies for housing demand – vacancy rates and median rental rates. The goal of this analysis is to determine how bike lanes impact the quality of neighborhoods by making them more or less attractive for homeowners, renters and businesses as indicated by demand for housing and property. Based on past research analyzing the relationship between bicycle infrastructure and housing markets, particularly Racca and Dhanju (2006) and Pelechrinis et al. (2017), and the fact that NYC seems to embrace bicycle infrastructure through increased investment, it can be hypothesized that bike lanes lead to improved economic indicators such as median rent and vacancy rates within NYC housing markets.

## Section 2: Literature Review

When studying the economic impact of bicycle lanes, researchers have generally focused on two measures of economic growth, retail sales and property values. While each have been explored quantitatively, accessing quantitative retail data has proved more challenging, leading most business-focused studies to utilize qualitative data. The lack of retail data is the key reason this study will focus primarily on the other most common measure of economic strength, property values. Other relevant areas of research around this topic include factors that influence the placement of bike lane infrastructure in a given area and whether bike lanes increase the level of biking within their area. In addition, economic analyses utilizing propensity score matching methods will be examined to guide the model of this study.

**Section 2.1: Retail Sales**

In one example of a successful quantitative retail study, The New York City Department of Public Transportation (DOT) quantitatively measured the economic impact of street improvements by analyzing business performance in the report "The Economic Benefits of Sustainable Streets" (2014). The NYC DOT conducted the study to determine the connection between street improvement projects throughout New York City and economic development. These improvement projects typically include multiple elements that work together to calm traffic and make the street more pedestrian-friendly, such as tree-lined medians, pedestrian safety islands and dedicated bike lanes. To measure the economic effects of these kinds of projects, the DOT used sales tax data gathered from the New York City Department of Finance (DOF). The study analyzed commercial sales on improved streets pre- and post-improvement, comparing them to the borough's overall sales. The study also compared the sales levels to sales on "comparison" streets, or streets within the same borough with a similar commercial and residential composition as determined by the DOT. Although this methodology does not show causality, researchers gathered that street improvements are correlated with overall positive sales performance of the improved streets when compared to unimproved streets. The use of DOF sales tax data from local businesses makes this one of the first studies to use quantitative data to assess the impact street changes can have on businesses (New York City Department of Transportation, 2014).

Rowe (2013) used a very similar methodology to study the impact of "road diets" on major downtown streets in Seattle. Road diets are a technique used in transportation planning that reduces the effective width of a road. In this case, each road diet involved installing bike lanes. Rowe's research utilized taxable retail sales data provided by the Washington State Department of Revenue to examine two projects in Seattle. In both cases, he found a major improvement in sales compared to neighborhood-wide data and comparison-site data after the bike lanes were installed. Both "Sustainable Streets" (2014) and Rowe (2013) utilized comparison streets in their analysis. This study will build on this approach quantitatively by utilizing propensity score matching methods to compare treated and untreated tracts, to be discussed further below.

Qualitative survey data has been used most often to investigate the business impacts of bike lanes. In 2011, Stantec Consulting Ltd. used qualitative data to assess the business impacts of two newly constructed separated two-way bike lanes in downtown Vancouver. The study was conducted on behalf of the Vancouver Economic Development Commission in response to concerns expressed by downtown businesses that the bike lanes had negatively impacted sales. Due to difficulties accessing financial data, the researchers conducted stakeholder surveys of business owners and managers, customers, and employees to acquire necessary business information such as sales and profit. They also surveyed customers about how the bike lanes had affected their shopping habits and their chosen method of transportation. Based on this information, they found an overall decrease in sales due to the bike lanes. The study emphasized that its results would be subject to response bias since people with strong feelings about the bike lanes would be more likely to send in survey responses ("Vancouver Separated Bike Lane Business Impact Study," 2011). Similar survey studies have been conducted in cities around the world, such as San Francisco and Toronto, with mixed results as to the economic impact of bike lanes (Drennen, 2003; Sztabinski, 2009).

Focusing on business performance represents a strong method for measuring a relationship between bike lanes and economic development. However, these studies have shown the difficulty of acquiring quantitative sales data and how response bias and mixed results can make qualitative studies unreliable. Therefore, this analysis will take a quantitative approach and focus on measuring housing demand as an indicator of economic vitality.

**Section 2.2: Property Values**

A common way to assess a bike lane's impact on economic development in the form of housing demand is to consider the values of commercial or residential properties within proximity of that bike lane. Most notably, Racca and Dhanju (2006) applied this method to residential properties in Delaware. This study combined Geographical Information Systems data (GIS) with a hedonic pricing model, using geographic data on bike trails, tax parcels and property sale values to run regression models. The study found that properties within 50 meters of bike trails were valued at least $8,800 higher, controlling for other variables impacting housing prices such as number of acres, land assessment, building assessment, total number of rooms, and number of bedrooms.

Pelechrinis et al. (2017) utilized a similar research method to analyze the effects of Pittsburgh's shared bike system on housing prices. This study examined the impact of the bike share program at the microscopic and macroscopic levels on both rental and sale prices. First, it compared Pittsburgh zip codes containing shared bike stations to zip codes without any stations to see if the presence of stations impacted home values. Next, it compared Pittsburgh real estate prices as whole to comparison city prices that did not have a bike share system. It gathered real estate value data from Zillow and observational data on where the bike stations were located throughout the city, using the difference-in-differences method to make inferences from observational data. Ultimately the researchers found that the shared bike system led to increased sale and rental values within zip codes that contained stations and within Pittsburgh as a whole compared to areas without shared bike systems.

Departing from the conclusions of most bike route analyses, Krizek (2006) found that bike facilities decreased home values in urban and suburban Minneapolis-St. Paul. This study examined roadside, non-roadside and on street bicycle facilities separately and in both urban and suburban settings. It used home sale data from Regional Multiple Listing Services (RMLS) of Minnesota, Inc. and mapped the address of each home using GIS. It also calculated the distance to the nearest roadside trail, non-roadside trail and on street bike lane for each address. Krizek found that in the city, off-street facilities had a positive effect on home prices, while roadside facilities had a negative effect, and on-street facilities had no effect. In the suburbs all types of bike facilities had a negative effect. This study reveals the importance of analyzing different types of facilities separately, since they may have different impacts. It also presents the possibility of dividing areas of studies based on urban concentration since bike lanes could serve different purposes in different environments.

**Section 2.3: Endogenous Improvements**

Multiple studies have found that causality also occurs in the opposite direction of this analysis and the other studies mentioned. While bicycle infrastructure may instigate economic growth, strong economic indicators can also lead to the installation of bicycle infrastructure in a particular area. Flanagan, Lachapelle and El-Geneidy (2016) assessed the relationship between cycling infrastructure and socioeconomic wealth using census and municipal cycling infrastructure data in Chicago and Portland. Opinions from poor and minority communities that bicycle culture comes with rising living costs that displace established cultures inspired the study. Independent variables included percent white population, median household income and median home value throughout the years 1990 to 2010 in order to capture community demographics associated with gentrification and privilege. A regression found that areas of existing privilege or areas with increasing privilege are more likely to see increased bicycle infrastructure investment. The study controlled for other factors that may lead to investment, like population density and distance to downtown.

Cradock et al. (2009) conducted a larger scale study on the same subject across 3,140 counties in the United States to see what leads to federal transportation funding for pedestrian and bicycle facilities. They found that counties with high levels of poverty or low educational status were less likely to receive funding. From 1992 to 2004, the federal government spent $3.17 billion on these projects. Based on the results from the study, it is clear that most of this funding went to high-income areas.

While a relationship occurs in both directions between bicycle infrastructure and economic vitality, my analysis will focus on whether bike lanes spark economic development. This question has not been explored as extensively on a quantitative level and it will help assess how bike lanes impact urban neighborhoods. Given the reciprocal relationship between the level of bike routes in a certain area and economic vitality indicators as demonstrated in these two studies this analysis will address endogeneity in its empirical model.

**Section 2.4: Bike Lanes and Commuting Level**

While this analysis will focus on economic impacts of bicycle infrastructure, a study by Buehler and Pucher (2011) provides a basis for the assumption that increasing bike lanes leads to increased biking. When looking at economic development data, this finding connects any positive impact of bike lanes to biking itself. The study involved the largest multiple regression analysis to date involving bike lines to analyze the role of bike paths and lanes on variation in bike commuting rates in 90 of the 100 largest U.S. cities. The researchers used data on 'centerline miles,' or roads with bike lanes, collected from planners, transportation experts, and government officials to measure the relationship bike path and lanes have with cycling levels, controlling for factors including cycling safety, gasoline price, and public transport supply. The study found the greater a city's supply of bike paths and lanes, the higher its rate of bike commutes. This allows a connection to be made between the results of any study on bike infrastructure and the impact they have on commuting levels, the real attraction of having this infrastructure within a city.

**Section 2.5: Propensity Score Matching**

To account for endogeneity, this analysis will utilize propensity score matching, an econometric method used to generate an observational sample that mimics a random experimental distribution. Propensity score matching is often used in geographic analysis to explore the effects of a treatment within a certain area compared to a similar untreated area. A study by O'Keefe (2003) on the effect of California's enterprise zone (EZ) program on employment growth used propensity score matching to assign conditional probabilities to California census tracts based on their chance of being designated as an EZ. O'Keefe based this matching on census demographic characteristics in 1990 and employment data from 1992, allowing matches to be used more than once to improve the estimate of the treatment effect. Duncombe, Yinger and Zhang (2016) also used propensity score matching in their study of the impact of school district consolidation on New York City property values. In addition to reducing coviariate bias between treated and untreated districts, they also used propensity score matching to prevent sensitivity to incorrect assumptions in the model.

# Section 3: Data and Basic Analysis

## Section 3.1: Data Sources

This study will primarily analyze data on bike routes gathered from the NYC Department of Transportation (DOT) and demographic data based on the 2010 U.S. Census. All data points correspond with a U.S. census tract in a certain year, including the total length of bike lanes within the tract, demographic information and economic indicators. The total number of U.S. Census tracts included in the study is 2,168 over the course of eight years, resulting in over 17,000 observations.

A dataset titled "New York City Bike Routes" (2017) from the NYC DOT's *Bicycle and Greenway Program* maps the location of every bike lane throughout New York City. This "Bike Route" dataset breaks the NYC bicycle network into small segments and provides a set of information about each segment, including its location, the type of facility and the date of its installation. Geographic information systems (GIS) software is designed to analyze such geographic or spatial data. Through GIS, it was possible to match the "Bike Route" dataset with a geographic dataset of NYC census tracts, labeling each route segment with its corresponding tract. A map of this data is included in Figure 1 for reference. GIS also has a function to measure the exact length of each segment based on geographic and spatial data, assigning a length to each route segment. With these added variables, the "Bike Route" dataset was used to calculate the total length of bike lanes within each NYC census tracts for the years 2010 to 2017, the main variable in this analysis.

SimplyAnalytics, a web-based analytics and data visualization application, provides demographic and population data for each U.S. census tract. The data used from SimplyAnalytics can be categorized as interpolated data, because the platform uses an algorithm to connect 2010 census baseline data to data collected during the years between each census.[1] Measures of economic vitality, vacancy rates and median rent, in addition to demographic characteristics of each census tract were gathered from this application for the years 2010 to 2017. Tables 1 and 2 organize the summary statistics for each variable. Percent of the population with less than a high school degree, median income, percent of the population that is white, population and number of housing units for each census tract were all included as controls since they could impact the property values and vacancy rates within an area, in line with previous studies included in the literature review. In addition to serving as controls, demographic variables include the changing composition of census tracts in the analysis. In particular, the housing units variable measures the concentration of residential property compared to commercial with in a tract.

**Section 3.2: Summary Statistics**

Summary statistics were created for independent variables aggregated for all tract-year observations. The summary statistics for each independent variable show that New York City is a city of extremes, with some tracts disproportionally wealthy, white or bikeable compared to others. As indicated in Table 1, the 1st and 3rd quadrants in median income, percentage white and lane length see large differences, reflecting census tracts with wide disparities in these characteristics. The maximum values for each dependent variable also reveal this difference, showing that there are clear outliers that far exceed the median values for each variable. These wide disparities between tracts are important to consider in this analysis, especially because bike lane length is one of the most starkly divided variables and possibly connected to other outliers. In each case the minimum value is 0, reflecting about 15 census tracts in New York City without

[1] Estimates are projections for market research by Easy Analytic Software, Inc. (EASI) based on census data from 2010. The margin of error for these projections is not provided.

any population. These census tracts were included in the analysis in the chance that they may have contained bike lanes.

In Table 2, the dependent variable summary statistics show the change over time in both median rent and percent vacancy. There was little change over time in either variable, with percent vacancy showing little trend at all. Overall, there was a pretty consistent increase in mean median rental values across the city, as to be expected.

Because not every census tract in NYC includes a bike lane, it was important to analyze this data conditional on the presence of a bike lane to see this trend over time as well. Table 4 shows that the percentage of tracts containing bike lanes steadily increased throughout the study period. This provides a solid basis for the analysis of the impact of new lanes in this study. The standard deviations in Table 3 reveal that the total number of bike lanes within a tract varies widely, with some tracts containing significantly more lanes than others. The minimum values in Table 3 also reveal that the data includes tracts containing very little lane lengths. These lengths are so small that it is most likely due to a measurement error, where a segment of the bicycle network that crossed over slightly into another census tract and was therefore measured by GIS.

**Section 4: Methods and Model**

The effect of bicycle lanes on urban neighborhoods was analyzed through multiple linear regressions. Two different dependent variables were tested, median rent value of a tract and the percent vacancy rate of a tract. Independent variables included a range of demographic and geographic factors that could impact these dependent variables apart from bike lane length. These included population, the percentage of the population that is white, the percentage of the population with less than a high school degree, the median income of the tract and also the number of housing units within the tract to control for commercial versus residential balance. To goal of this model was to determine the impact of bike lane length on each dependent variable to get a sense for how bicycle infrastructure influences urban development. The equation for this model is included below, with i indicating tract i and t indicating year t. The Y symbol represents each dependent variable, median rent or percent vacancy for each census tract year.

$$Y_{it} = \beta_0 + \beta_1(\text{Lane Length}_{it}) + \beta_2(\text{Median Income}_{it}) + \beta_3(\%\ \text{White}_{it})$$
$$+ \beta_4(\%\text{Less than HS degree}_{it}) + \beta_5(\text{Population}_{it}) + \beta_6(\text{Housing Units}_{it})$$

As stated previously, the coefficient on lane length is expected to be positive for median rent and negative for percent vacancy. The coefficients on median income and percent white are expected to also have these signs because these are likely indicators of a more prosperous area that would therefore have higher rent values and less vacancy rates because it is more desirable to live in. Percent of the population with less than a high school degree is expected to have a negative relationship with median rent and a positive relationship with vacancy, although the relationship with vacancy is harder to predict since a lower rent area my attract more underprivileged tenants. Population and housing units have indeterminate relationships with median rent and vacancy. These variables are included as proxy controls for density and residential makeup of a census tract.

However, bike lanes are not randomly distributed throughout neighborhoods. As discussed earlier, past studies have found that economically prosperous areas, as indicated by median income, educational background and race, tend to receive more funding for bike lanes than economically weaker areas (Cradock et al. 2009). In New York City, the Department of Transportation has placed a majority of their investment into Manhattan and Brooklyn, the city's wealthiest and whites boroughs, although bicycle infrastructure projects do occur in other boroughs as well (Bliss 2017). Therefore, endogenous siting of bike lanes in more affluent neighborhoods may lead to biased coefficients. In addition, it is likely that that when expanding New York City's bicycle network, areas closer to the current network receive greater investment than areas far from current bicycle infrastructure. Therefore, census tracts already containing lanes would be more likely to receive additional infrastructure, again leading to biased coefficients.

To solve this problem, this analysis employs a propensity score matching method to shape the data to mimic an experimental study. Propensity score matching creates treatment and control groups with very similar probabilities of being treated, where probabilities are estimated using a logistic regression.

In order to create a treatment dataset and a control dataset, a probit was used to calculate the probability of treatment for each tract based on census data from 2000. In this case, treatment refers to a one-time decision to install bike lanes or not. Since the variation in bike lane additions was small, I chose not to match on individual year data. Instead, I approach this as a treatment effects model and focus on census data before bike lane construction reached a significant level. The variables used in the probit to calculate treatment probability for each tract included population, median age, the percentage of the population that was white, the percentage of the population under 25, education demographics, median income and number of housing units. In addition to these variables, which closely mirrored the final linear regression, the distance between the edge of a census tract and the nearest established bike lane in 2000 was included as a variable in the probit model, labeled "neardist." If a tract contained a bike lane in 2000, *neardist* was 0. Because the boundaries of census tracts change over time, these variables from 2000 were roughly matched with the census tract boundaries of the years 2010 to 2017.

Based on these probit values, or propensity scores, each treated tract was matched with an untreated tract that had a very similar probability of being treated. This match was conducted using the nearest neighbor method with replacement in order to create observationally similar groups. The nearest neighbor method matched each treated tract with an untreated tract with the closest propensity score. If multiple untreated tracts had the same differences, one was matched at random. The nearest neighbor method also does not place restrictions on the maximum difference between treated and untreated tracts. This method was chosen because it is the most commonly used and most straightforward matching method. The matching process also allowed replacement, meaning that multiple treated tracts could be matched to the same control, in order to get the closest possible match (Austin 2011). Any untreated tracts that did not match with treated tracts were removed from the dataset.

A covariate balance table was initially created based on the unmatched data to see if treated and untreated groups were significantly different in any of the variables used in the probit model. A covariate balance analysis shows whether the mean of each variable used to match the treated and untreated dataset, in this case the demographic data from 2000, is significantly different through hypothesis testing. The goal of propensity score matching is to ultimately create two dataset sets of treated and untreated observations that are observationally similar, so the covariate balance should fail to reject the null hypothesis, meaning the means are observationally similar. The unmatched dataset covariate balance table showed that the variables "neardist" and "HS," or the number of high school graduates, were significantly different between the treated and untreated groups, as indicated by a t-value higher than the critical t-value of about 2.12. However, all other variables were observationally similar even before matching, meaning that treated and untreated tracts were not very different to being with. This covariate balance table and all other covariate balance tables are included in Tables 5-8 for reference, with the unmatched covariate balance included in Table 5.

The covariance balance analysis on the matched dataset, detailed in Table 6, saw decreased t-values for each of the significantly different variables, but they were each still above the critical t-value of 2.11, indicating the two groups were still significantly different in regard to *neardist* and number of high school graduates. Therefore, property values in treated and untreated tracts would not only be impacted by bicycle lane length but also these variables in a linear regression, leading to biased coefficients.

To further correct for this difference, a cutoff to *neardist* values was added to the matching method. The *neardist* variable was chosen for restriction for both the quantitative reason that treated and controls groups still saw a significant difference in this variable and the intuitive reasoning that it is very unlikely for new bike lanes to be added to tracts far away from the existing network. Although the "HS" variable was also significantly different, cutting off two variables would have made the dataset too small, and *neardist* proved to be more intuitively significant. The first restriction ensured that only tracts with *neardist* values below the 75% percentile of *neardist* among the treated tracts could be included in the dataset, or below 3,174 meters. Adding this restriction decreased the t-value of the difference in means for *neardist* below the critical t value of roughly 2.11, as shown in Table 7. A second dataset with a stricter cutoff was also created to use in linear regression tests, shown in Table 8. This cutoff required all tracts to be less than 1,025 meters away from the nearest bike lane, which was the treated group's *neardist* median value. Covariant analysis of this data set showed that the t-value of the difference in means for *neardist* was significantly decreased, signaling strong observational similarity between the two groups. Each of these respective cut-off values was chosen in order to exclude extreme values without making the dataset too small, using the treated group as a guide. However, using the *neardist* treated median as the restriction cut the dataset down to a very small size of under 1,000 observations, which was significant given that the dataset began at over 17,000 observations.

**Section 5: Results**

To determine the relationship between bike lanes and property values, linear regressions were run on the unmatched dataset, the matched dataset and each respective matched dataset with a restriction. Coefficient values and standard errors for each regression are shown in Table 9. Each model includes year fixed effects. Additional models used spatial fixed effects, but it resulted in too much explanation because there was little change within the tracts across years.

Ultimately, each regression showed a significant negative relationship between total bike lane length and median rent, the opposite of the hypothesized direction. Before any matching or *neardist* restrictions, lane length had a significant negative relationship with median rent with a coefficient of about - 0.031. This means that for every one standard deviation in lane length, or 957.52 meters, the median rent value was estimated to decrease by about $29.68. The Adjusted R-squared for this regression was 0.5846. The regression on the matched data set saw an increased Adjusted R-squared up to 0.6234 and a slightly increased negative relationship between bike lane and median rent. With each restriction, the Adjusted R-squared increased as well, ultimately measuring 0.7179 in the dataset restricted to *neardist*'s treated median. The regression results from the final regressions, both with matching and restriction, are considered the preferred models since the restrictions best solved the problem of endogeneity as indicated by the covariate balance. In particular, the third regression with a restricted *neardist* of only 3174 meters is the best model because using the *neardist* treated median restriction cut the dataset down to a very small size of under 1,000 observations. Using this regression data, an increase in one standard deviation in bike lane leads to an average rent decrease of about $29.97, which is the opposite direction than expected but still quantitatively small.

In the regressions on median rent, the percent of people with less than a high school degree also had the opposite of the expected sign. The sign was expected to be negative since the more uneducated the population, the less desirable the area would likely be to new tenants and the lower the socioeconomic status of the neighborhood, bringing down median rent. In this regression, the relationship is positive, possibly indicating an edogeneity issue that was not addressed.

Although demonstrating a weaker relationship, regressions on percentage vacancy for each tract consistently indicated a significant positive relationship between lane length and vacancy rates, meaning bicycle lanes lead to increased vacancy. These results for each dataset, unmatched, matched and restricted, are included in Table 10. Fixed year affects were also included in these regressions. Therefore, the ultimate conclusion of this analysis is that bike lanes have a negative effect on housing markets in New York City.

The expected sign of the Lane Length coefficient was positive for the median rent regression and positive for percent vacancy, reflecting past research into bicycle infrastructure and housing markets by Racca and Dhanju (2006) and Pelechrinis et al. (2017). Each of these studies utilized direct real estate data sources to include specific details about properties like total number of rooms and land assessment, either on an individual basis, on the zip code level or on an aggregate city level. I pulled my data from a census source rather than real estate data, so these kinds of data points were not controlled for in my analysis, which could have led to a negative relationship. Additionally, my analysis did not control for type of bicycle lane.

Unprotected lanes likely have a different effect on neighborhoods compared to protected lanes, as Krizek (2006) found. These results would not account for that and could have led to a negative coefficient, especially because the majority of bike lanes in NYC are unprotected. Bike lanes likely also had different effects on individual streets within a census tract, so there is the possibility that streets with a negative relationship outweighed streets with a positive relationship.

Overall, this negative relationship suggests that bike lanes are not valued by people when making decisions about where to live and actually likely negatively impact these types of decisions. This could mean that bikes create greater congestion on streets, disrupt traffic or cause general chaos on the roads that they are added to, bringing down property values within a tract. It also could reflect a lack of regulation of NYC bike infrastructure and bikers themselves, allowing bikers to break traffic laws. This would increase the disruptive effect of bicycle lanes and make areas that they are a part of less desirable to live in.

## Section 6: Conclusion

While this analysis shows that bike lane length has a negative effect on urban neighborhoods, it is important to consider that bicycle infrastructure comes in many different forms. There is a significant difference between protected and unprotected bike lanes in the level of safety and structure they provide for bicycle paths, leading to different levels of congestion. A further study could run separate analysis on different kinds of bicycle infrastructure to isolate their different effects on urban neighborhoods.

Bike lanes also have very different natures depending on the area they are located in. A bike lane along a major avenue likely has a very different effect compared to a bike lane in a quiet residential area or even along one of New York City’s many bridges. All of these factors merit further investigation to determine where bike lanes have negative and positive economic impacts on their surrounding areas, especially because New York City census tracts vary widely in characteristics like housing density and population. For example, it may be true that bicycle lanes are often added to already-busy commuter paths, like main Avenues in Manhattan. Therefore bike lanes could congest these roads further, making an area less desirable to live in. A future study could use GIS software to more accurately determine population and commuter density of census tracts to directly control for these factors. A future study could also investigate this same question but in regard to suburban areas since bicycle lanes serve a very different purpose in areas that are more spread out.

New York City is just one example of a city where bicycle infrastructure is expanding. It does not necessarily represent the effect of bike lanes in other urban areas, perhaps without strong public transportation or with more room to expand roads. Further studies could investigate the effect of bike lane infrastructure on other urban housing markets across the U.S. individually or explore a similar question on an aggregate level, looking at multiple cities as samples to determine an overall effect.

References

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, *46*(3), 399–424. http://doi.org/10.1080/00273171.2011.568786.

Bicyclists: Building the Network. (n.d.). Retrieved October 12, 2017, from http://www.nyc.gov/html/dot/html/bicyclists/bicyclists.shtml.

Bliss, L. (2017). Here's How Godawful NYC Bike Lanes Used to Be. Retrieved from https://www.citylab.com/transportation/2017/02/heres-how-godawful-nyc-bike-lanes-used-to-be/515457/.

Buehler, R., & Pucher, J. (2012). Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. *Transportation*, *39*(2), 409-432. doi:https://doi.org/10.1007/s11116-011-9355-8.

Cradock, A. L., Troped, P. J., Fields, B., Melly, S. J., Simms, S. V., Gimmler, F., & Fowler, M. (2009). Factors Associated with Federal Transportation Funding for Local Pedestrian and Bicycle Programming and Facilities. *Journal of Public Health Policy*, *30*, S38-S72. doi:https://doi.org/10.1057/jphp.2008.60.

Drennen, E. (2003). *Economic Effects of Traffic Calming on Urban Small Businesses*. Retrieved from https://www.sfbike.org/download/bikeplan/bikelanes.pdf.

Duncombe, W. D., Yinger, J., & Zhang, P. (2014). How Does School District Consolidation Affect Property Values? A Case Study of New York. *Public Finance Review, 44*(1), 52-79. doi:https://doi.org/10.1177/1091142114524617.

Easy Analytic Software, Inc. (EASI), *Census US* (2010), distributed by SimplyAnalytics. Retrieved from http://simplyanalytics.com/.

Flanagan, E., Lachapelle, U., & El-Geneidy, A. (2016). Riding tandem: Does cycling infrastructure investment mirror gentrification and privilege in Portland, OR and Chicago, IL? *Research in Transportation Economics*, *60*, 14-24. doi:https://doi.org/10.1016/j.retrec.2016.07.027.

Hu, W. (2017, July 30). More New Yorkers Opting for Life in the Bike Lane. *The New York Times*. Retrieved from https://www.nytimes.com/2017/07/30/nyregion/new-yorkers-bike-lanes-commuting.html?_r=0.

Krizek, K. J. (2006). Two Approaches to Valuing Some of Bicycle Facilities' Presumed Benefits. *Journal of the American Planning Association, 72*(3). Retrieved from http://atfiles.org/files/pdf/krizek-valuing-bicycle%20facilities-benefits.pdf.

New York City Department of Transportation. (2014). *The Economic Benefits of Sustainable Streets*. Retrieved from https://www.ssti.us/wp/wp-content/uploads/2014/01/dot-economic-benefits-of-sustainable-streets.pdf.

New York City Department of Transportation, Bicycle and Greenway Program. (2017). *NYC Bike Routes* [Data set]. Retrieved from http://www.nyc.gov/html/dot/html/about/datafeeds.shtml#bikes.

O'Keefe, S. (2004). Job Creation in California's Enterprise Zones: A Comparison Using a Propensity Score Matching Model. *Journal Of Urban Economics*, 55(1), 131-150. doi:http://dx.doi.org/10.1016/j.jue.2003.08.002.

Pelechrinis , K., Zacharias, C., Kokkodis, M., & Lappas, T. (2017). Economic impact and policy implications from urban shared transportation: The case of Pittsburgh's shared bike system. *Plos One*, 12(8). doi:https://doi.org/10.1371/journal.pone.0184092.

Racca, D. P., & Dhanju, A. (2006, November). *Property Value/Desirability Effects of Bike Paths*

*Adjacent to Residential Areas* (Rep.). Retrieved from http://udspace.udel.edu/bitstream/handle/19716/13246/bikepathfinal.pdf?sequence=1.

Rowe, K. (2013). *Bikenomics: Measuring the Economic Impact of Bicycle Facilities on Neighborhood Business Districts*. University of Washington College of Built Environments. Retrieved from https://docs.google.com/file/d/0B0xHj6OM3QVWMUxScjZuMndxVkk/edit.

Sisson, P. (2017, April 18). Cycling success: 10 U.S. cities pushing biking forward. Retrieved from https://www.curbed.com/2017/4/18/15333796/best-cities-bike-commute-us-cycling.

Sztabinski, F. (2017, October 12). *Bike Lanes, On-Street Parking and Business: A Study of Bloor Street in Toronto's Annex Neighbourhood* (Rep.). Retrieved https://grist.files.wordpress.com/2011/04/bike-lanes-parking.pdf.

*Vancouver Separated Bike Lane Business Impact Study* (Rep.). (2011, July 20). Retrieved http://council.vancouver.ca/20110728/documents/penv3-BusinessImpactStudyReportDowntownSeparatedBicycleLanes-StantecReport.pdf.

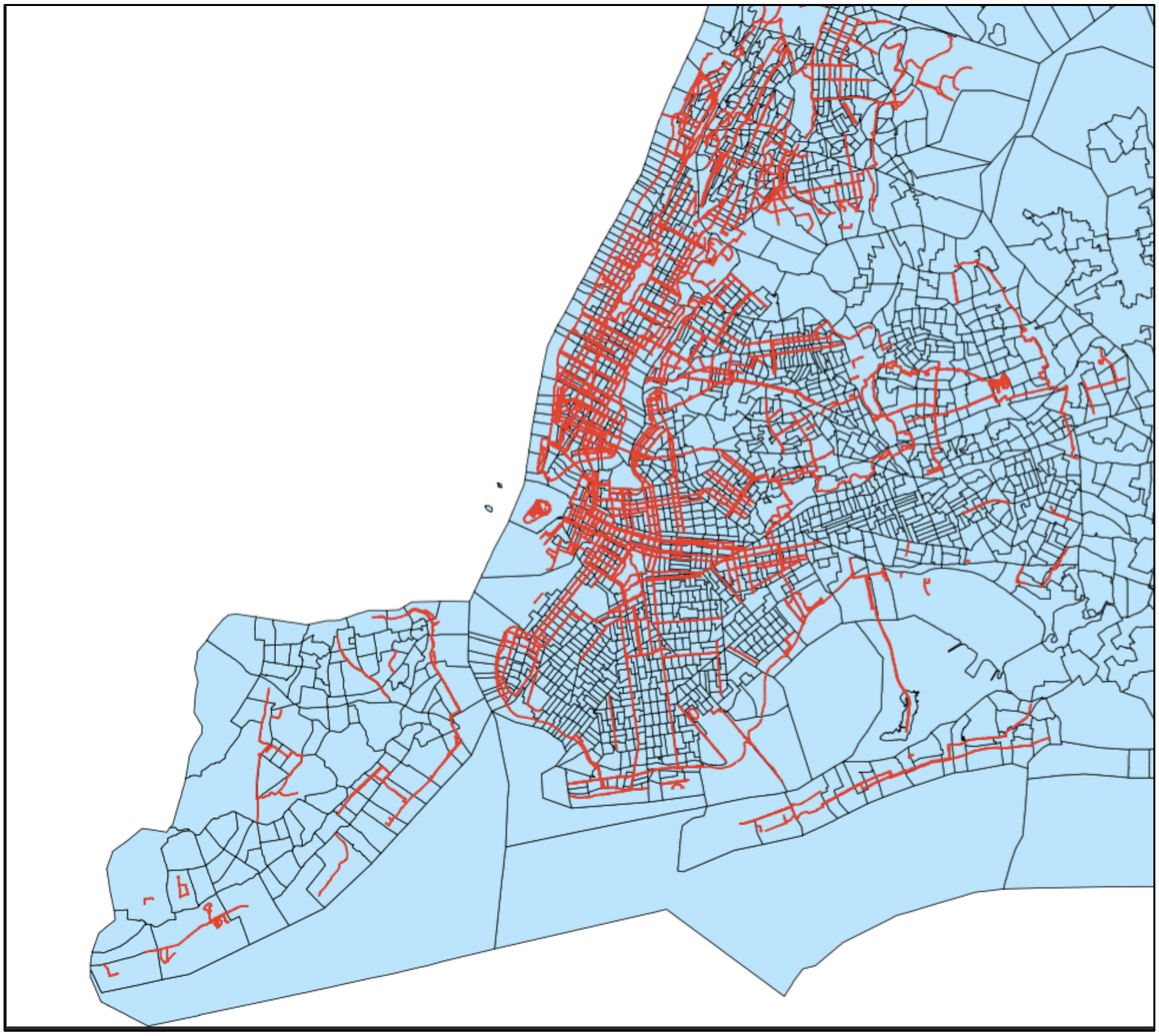

Figure 1
Screenshot of Layered Map in ArcGIS
*Red - Bike Lanes; Blue – Census Tracts*

Table 1
*Summary Statistics, Independent Variables*

| | Median | Mean | Standard Deviation | Minimum | 1st Quadrant | 3rd Quadrant | Maximum |
|---|---|---|---|---|---|---|---|
| Lane Length (meters) | 0.0 | 400.1 | 957.52 | 0 | 0.0 | 487.5 | 17197.5 |
| Population | 3513 | 3860 | 2223.73 | 0 | 2315 | 4933 | 27978 |
| % White | 37.00 | 41.81 | 29.37 | 0 | 16.05 | 68.51 | 100.00 |
| Housing Units | 1372 | 1592 | 1126.27 | 0 | 868 | 1978 | 13316 |
| % Less than High School Degree | 18.20 | 20.23 | 13.06 | 0 | 10.43 | 28.60 | 100.00 |
| Median Income | 59205 | 63221 | 31411.21 | 0 | 41689 | 80134 | 283344 |

Table 2
*Summary Statistics by Year, Dependent Variables*

| | Year | Median | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|---|---|
| %Vacancy | | | | | | |
| | 2010 | 6.45 | 7.81 | 6.53 | 0 | 100 |
| | 2011 | 6.68 | 8.04 | 6.62 | 0 | 100 |
| | 2012 | 6.86 | 8.30 | 6.75 | 0 | 100 |
| | 2013 | 7.06 | 8.39 | 6.70 | 0 | 100 |
| | 2014 | 6.40 | 7.76 | 6.49 | 0 | 100 |
| | 2015 | 6.21 | 7.64 | 6.49 | 0 | 100 |
| | 2016 | 6.30 | 7.71 | 6.54 | 0 | 100 |
| | 2017 | 6.26 | 7.61 | 6.48 | 0 | 100 |
| Median Rent | | | | | | |
| | 2010 | 944.00 | 976.11 | 342.44 | 0 | 2000 |
| | 2011 | 951.00 | 982.71 | 344.44 | 0 | 2000 |
| | 2012 | 1063.50 | 1074.91 | 358.34 | 0 | 2000 |
| | 2013 | 1098.50 | 1127.11 | 379.44 | 0 | 2237 |
| | 2014 | 1135.00 | 1154.34 | 378.65 | 0 | 2250 |
| | 2015 | 1133.00 | 1152.39 | 376.14 | 0 | 2242 |
| | 2016 | 1132.00 | 1153.26 | 377.63 | 0 | 2242 |
| | 2017 | 1141.00 | 1161.44 | 379.82 | 0 | 2243 |

Table 3

*Summary Statistics for New York City Census Tracts Conditional on Containing a Bike Lane*

| | Year | N | Median | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|---|---|---|
| Length of Bike Lanes within a Given Tract (meters) | 2010 | 751 | 607.74 | 1113.8 | 1579.03 | 0.002515 | 17135.65 |
| | 2011 | 753 | 622.38 | 1134.11 | 1604.88 | 0.003 | 17696.38 |
| | 2012 | 771 | 657.71 | 1163.09 | 1626.52 | 0.003 | 17966.87 |
| | 2013 | 811 | 680.03 | 1209.11 | 1685.87 | 0.003 | 18283.77 |
| | 2014 | 833 | 685.29 | 1234.75 | 1710.49 | 0.009 | 18283.78 |
| | 2015 | 864 | 711.39 | 1264.27 | 1723.42 | 0.009 | 18399.63 |
| | 2016 | 910 | 753.74 | 1285.77 | 1714.69 | 0.009 | 18399.71 |
| | 2017 | 914 | 756.37 | 1289.38 | 1713.33 | 0.009 | 18399.71 |

Table 4

*Portion of New York City Census Tracts Containing Bike Lanes*

| Year | Percent Containing Bike Lanes |
|---|---|
| 2010 | 34.65% |
| 2011 | 34.75% |
| 2012 | 35.57% |
| 2013 | 37.42% |
| 2014 | 38.44% |
| 2015 | 39.87% |
| 2016 | 41.99% |
| 2017 | 42.17% |

Table 5

*Covariate Balance - Unmatched Data, Unrestricted*

| | Mean Treatment | Mean Control | t.diff |
|---|---|---|---|
| Population | 3349.6894 | 2910.7081 | 0.4589 |
| White Population | 1675.1999 | 1585.2037 | 0.1512 |
| Neardist | 1984.5317 | 4117.3648 | -2.2830 |
| Housing Units | 1542.3099 | 1212.6832 | 0.9163 |
| High School degree | 605.7635 | 644.3578 | -3.9674 |
| College degree | 428.2283 | 309.2596 | 0.7680 |
| Graduate degree | 314.8316 | 199.8485 | 0.8377 |
| No High School degree | 748.1145 | 608.5342 | 0.6967 |
| Some College | 517.1849 | 508.5901 | 0.0671 |
| Median Income | 34956.6737 | 42449.7317 | -1.1767 |
| Median Age | 24.7774 | 27.0693 | -0.4230 |
| Population under 25 | 2614.1228 | 2270.5901 | 0.5915 |

Critical t-value: 2.12

Table 6

*Covariate Balance - Matched Data, Unrestricted*

| | Mean Treatment | Mean Control | t.diff |
|---|---|---|---|
| Population | 3349.6894 | 2669.1874 | 0.7504 |
| White Population | 1675.1999 | 1368.4442 | 0.5695 |
| Neardist | 1984.5317 | 3923.8513 | -2.1532 |
| Housing Units | 1542.3099 | 1147.9284 | 1.1608 |
| High School degree | 605.7635 | 588.9558 | -3.9355 |
| College degree | 428.2283 | 286.4000 | 0.9628 |
| Graduate degree | 314.8316 | 185.6547 | 0.9706 |
| No High School degree | 748.1145 | 610.4147 | 0.7065 |
| Some College | 517.1849 | 466.1390 | 0.4377 |
| Median Income | 34956.6737 | 41481.9705 | -1.0500 |
| Median Age | 24.7774 | 25.9004 | -0.2113 |
| Population under 25 | 2614.1228 | 2137.5642 | 0.8756 |

Critical t-value: 2.11

Table 7

*Covariate Balance - Matched Data, Restricted to 3rd Quadrant Treated Neardist of 3174 meters*

| | Mean Treatment | Mean Control | t.diff |
|---|---|---|---|
| Population | 3455.3493 | 3100.4868 | 0.3833 |
| White Population | 1804.4810 | 1806.5450 | -0.0035 |
| Neardist | 802.7053 | 1425.0492 | -2.0926 |
| Housing Units | 1626.9291 | 1283.0370 | 0.9593 |
| High School degree | 599.3483 | 609.7989 | -4.7397 |
| College degree | 481.7395 | 289.9577 | 1.2315 |
| Graduate degree | 369.0998 | 207.5926 | 1.1092 |
| No High School degree | 729.3204 | 659.1640 | 0.3354 |
| Some College | 528.8313 | 475.9577 | 0.4428 |
| Median Income | 34986.7006 | 37293.7937 | -0.3443 |
| Median Age | 25.1905 | 29.1180 | -0.7763 |
| Population under 25 | 2708.3393 | 2242.4709 | 0.8264 |

Critical t-value: 2.11

Table 8

*Covariate Balance - Matched Data, Restricted to Median Treated Neardist of 1025 meters*

| | Mean Treatment | Mean Control | t.diff |
|---|---|---|---|
| Population | 3612.0612 | 3770.5278 | -0.1592 |
| White Population | 2006.5403 | 1895.2500 | 0.1746 |
| Neardist | 215.9556 | 338.7662 | -1.2556 |
| Housing Units | 1742.7164 | 1553.8056 | 0.4707 |
| High School degree | 618.7552 | 760.2500 | -5.7473 |
| College degree | 547.7179 | 307.4444 | 1.3722 |
| Graduate degree | 437.4269 | 183.1667 | 1.6726 |
| No High School degree | 723.9060 | 837.1389 | -0.5132 |
| Some College | 558.1597 | 569.1944 | -0.0827 |
| Median Income | 36418.1687 | 31377.8889 | 0.8019 |
| Median Age | 25.8819 | 28.0889 | -0.4401 |
| Population under 25 | 2885.9657 | 2657.1944 | 0.3630 |

Critical t-value: 2.11

Table 9

*Linear Regressions on Median Rent by New York City Census Tract-Year, Fixed Years*

| | Regression 1 *Unmatched Data, Unrestricted* | Regression 2 *Matched Data, Unrestricted* | Regression 3 *Matched Data, 3rd Quadrant Restriction* | Regression 4 *Matched Data, Median Restriction* |
|---|---|---|---|---|
| (Intercept) | 453.7625*** | 446.0021*** | 450.1193*** | 462.9606*** |
| | (10.0678) | (10.7647) | (11.0980) | (16.9904) |
| Lane Length | -0.0309*** | -0.0329*** | -0.0313*** | -0.0362*** |
| | (0.0018) | (0.0019) | (0.0019) | (0.0022) |
| Median Income | 0.0082*** | 0.0084*** | 0.0085*** | 0.0091*** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| % White | 1.1538*** | 1.4860*** | 1.3434*** | 0.7125*** |
| | (0.0733) | (0.0818) | (0.0857) | (0.1537) |
| % Less than a High School Degree | 2.6795*** | 2.1128*** | 1.9623*** | 0.6748 |
| | (0.1970) | (0.2101) | (0.2161) | (0.3484) |
| Population | -0.0256*** | -0.0173*** | -0.0242*** | -0.0218*** |
| | (0.0022) | (0.0025) | (0.0025) | (0.0037) |
| Housing Units | 0.0907*** | 0.0728*** | 0.0854*** | 0.0695** |
| | (0.0045) | (0.0049) | (0.0048) | (0.0069) |
| $R^2$ | 0.5849 | 0.6238 | 0.6515 | 0.7185 |
| Adjusted $R^2$ | 0.5846 | 0.6234 | 0.6512 | 0.7179 |

Standard error in parentheses    * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10

*Linear Regressions on Percent Vacancy by New York City Census Tract-Year, Fixed Years*

| | Regression 1 *Unmatched Data, Unrestricted* | Regression 2 *Matched Data, Unrestricted* | Regression 3 *Matched Data, 3rd Quadrant Restriction* | Regression 4 *Matched Data, Median Restriction* |
|---|---|---|---|---|
| (Intercept) | 5.9182*** (0.2547) | 6.0826*** (0.2886) | 6.3334*** (0.3126) | 8.2314*** (0.5519) |
| Lane Length | 0.0003*** (0.0000) | 0.0002*** (0.0001) | 0.0001* (0.0001) | 0.0000 (0.0001) |
| Median Income | 0.0000*** (0.0000) | 0.0000*** (0.0000) | 0.0000*** (0.0000) | 0.0000*** (0.0000) |
| % White | 0.0268*** (0.0019) | 0.0283*** (0.0022) | 0.0276*** (0.0024) | 0.0120* (0.0050) |
| Less than a High School Degree | 0.0753*** (0.0050) | 0.0718*** (0.0056) | 0.0640*** (0.0061) | 0.0461*** (0.0113) |
| Population | -0.0021*** (0.0001) | -0.0022*** (0.0001) | -0.0022*** (0.0001) | -0.0024*** (0.0001) |
| Housing Units | 0.0035*** (0.0001) | 0.0035*** 0.0001 | 0.0033*** (0.0001) | 0.0035*** (0.0002) |
| $R^2$ | 0.1391 | 0.1479 | 0.1524 | 0.1417 |
| Adjusted $R^2$ | 0.1385 | 0.1471 | 0.1516 | 0.1397 |

Standard error in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$