Does Rail Transit Induce Displacement?
A Longitudinal Study of 24 US Metro Areas From 2000-2017

Eric Ronald Clute

A thesis
submitted in partial fulfillment of the
requirements for the degrees of

Master of Urban Planning

University of Washington
Spring 2020

Committee:

Rachel Berney
Jacob Vigdor

Programs Authorized to Offer Degree:

Urban Design and Planning

University of Washington

Abstract

Does Rail Transit Induce Displacement?
A Longitudinal Study of 24 US Metro Areas From 2000-2017

Eric Ronald Clute

Chair of the Supervisory Committee:

Rachel Berney

Department of Urban Design and Planning

ABSTRACT

After decades of decline, large US cities have begun reinvesting in urban neighborhoods including the construction of rail transit. Concerns over rail transit's role in driving displacement has become hotly debated. Community groups voice their distress over displacement while much of the current research fails to find significant evidence. This study uses recently published US Census data (2000-2017) to answer questions, including whether the introduction of rail transit is associated with changes in indicators of displacement in the receiving census tracts, such as log mean rent, educational attainment, age, and changes in the racial makeup. Utilizing a difference-in-difference regression model and a propensity score matching system, I select census tracts that received rail transit and test the relationship between the construction of rail transit and changes in these indicators of displacement. Specifically, my research sheds

significant light on how rail transit affects low income communities and people of color, who are more likely to be affected by displacement. My results indicate that displacement is occurring in census tracts near new rail stations across the country. Even with the introduction of control variables, communities of color and census tracts with low educational attainment are seeing the most change. Public officials and urban planners must craft policy so that large public investments do not place undue hard on communities that already struggle with disenfranchisement.

ACKNOWLEDGEMENTS

The support I have received from my family, friends, classmates, advisors, and committee members has been overwhelming, and to those I owe my heartfelt appreciation. I would not have survived three years of graduate coursework without all of your unwavering support. To my parents and siblings, I owe a special thanks for encouraging me to always follow my passions. To my friends, thank you for listening and helping me de-stress through many months of hard work. To my classmates Quinn Cordo and Tom McQueeny, I would not have completed this work without your support in laying the foundation for the analysis. And finally, thank you to Rachel Berney Ph.D. and Jake Vigdor Ph.D. for your support at various stages of my time at the UW. This project would not have happened without your assistance.

PREFACE

My research explores the relationship between space and identity, touching on the histories of urban policy and its effects on real people. As an academic and as a person of significant privilege, it is easy for me discuss issues of race, changing urban neighborhoods, and the effects of large public investments in ways that are divorced from the lived experiences of those affected. This privilege I hold as a white person allows me to discuss displacement while at the same time never worrying that my childhood neighborhood will feel those same pressures. I have the power to step back, to disengage, where many communities of color and low-income people do not. While I strive to reduce the influence of my privilege on the research, that may be an impossible task. In that light, I find myself with an imperfect, yet acceptable answer – ensuring that all research be grounded in quality data and guided by the experiences of those who are most impacted.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Since the 1990s, large US cities have seen significant reinvestment – neighborhoods formerly ignored because of the glamour of suburban cul-de-sacs have become the new darlings of young urban dwellers. Cities have spent billions of dollars in connecting these urban spaces with city centers and surrounding suburbs via rail transit (Kahn, 2007). Transit oriented development near these stations has become a standard planning practice, with the aim to increase density, reduce auto-dependency, decrease emissions, and revitalize urban communities (Rayle, 2015).

Even with these benefits, many studies detail both the concern of displacement [1](Rayle, 2015) as well as how housing costs near transit can be out of reach for low-income communities (Luckey et al., 2018). Others have concluded that displacement and gentrification are not universally occurring around transit stations, indicating that further analysis is necessary to better our understanding of the phenomenon (Kahn, 2007; Rayle, 2015). Most of these studies have focused on property values as an indication of housing costs, while my research sheds light on rental prices specifically, as low-income communities are more likely to be renters.

As mentioned in Kahn's work, a key indicator of gentrification is a change in educational attainment within a census tract (Kahn, 2007). A change in income, housing costs (such as rent and property values) and changes in the census tracts racial makeup are also indicators of gentrification, which can lead to displacement. Although there is disagreement within the

[1] I use the definition of displacement written by the researchers Grier & Grier, "the replacement of lower - income residents by more affluent households and the renovation of rundown housing in central urban neighborhoods" (Grier & Grier, 1978)

literature as to whether rail transit investments lead to displacement of low-income people, there is support that displacement is occurring due to larger economic forces which could be further exacerbated by the introduction of rail.

Urban planners and city officials looking to create dense, walkable, and transit accessible spaces must incorporate an anti-displacement lens into their work. To inform this perspective, research focused on measuring displacement is vital to ensuring large public investments, like rail transit, do not produce negative effects for low-income communities. The current research often has not found significant evidence of displacement due to rail transit investments; however, using the most up to date Census data and focusing on rental prices instead of property values is an opportunity to provide further clarity. This research follows these approaches.

I examine the relationship between transit investment and factors that increase the likelihood of displacing lower-income communities and communities of color. In particular, I am interested whether the introduction of rail transit in census tracts across the US led to changes in log mean rent, educational attainment, age, and the neighborhood's racial makeup. Using multiple multivariate regression models, data from the 2000 Decennial Census is compared to data from the American Community Survey (ACS) from 2013-2017. For each census tract that received rail, control census tracts are selected by a propensity score matching system.

Propensity scores are generated to predict the likelihood of a census tract receiving the treatment, based on variables similar among treated tracts. For example, I pose the questions, is there something similar among treated tracts, and if so, are there control tracts that also seem similar? Propensity scores are generated for treated and control tracts. A treated tract

that looks similar to a control tract, via a similar propensity score, are matched together for further analysis. I use the following variables to generate the propensity scores: log mean rent; percent female; percent Non-Hispanic white; percent of those over the age of 65; percent of those with a bachelor's degree or higher, percent use of public transit; distance to the nearest central business district (CBD) and percent of rental units.

A census tract that received rail is defined as "treated" when its population centroid, the center of the tract where the most people live, is within one mile of the new station. Control tracts are defined as a census tract located more than two miles from the station and also not treated by a different station. I define census tracts with a population centroid between one and two miles from a new station as the "messy middle" and remove them from the analysis (see Figure 1). As described by other researchers, displacement pressures do not lose their power one mile away from the station, and are likely causing "spatial spillover effects" on neighborhoods further away (Bardaka et al., 2018).

Some researchers focus on comparing messy middle tracts to treated tracts closer to the station, in an attempt to understand potential spillover effects. That said, I remove messy middle tracts from my analysis for three reasons: (1) time and resource limitations, (2) choosing to focus on tracts that are more likely to feel the effects of displacement, and (3) removing messy middle tracts ensures that they are not selected as control tracts by the model. The third point is worth emphasizing – if spillover effects are occurring, messy middle tracts cannot fully function as control tracts as they too are feeling the effects of the station. However, removing the messy middle does risk removing tracts that would make ideal treatment or controls, which could limit the results of my research.

The first model begins testing whether there is a relationship between the dependent variables (log mean rent, educational attainment, age, and racial makeup) *without* control variables. The second model introduces the following control variables: log mean rent; percent female; percent Non-Hispanic white; percent of those over the age of 65; percent of those with

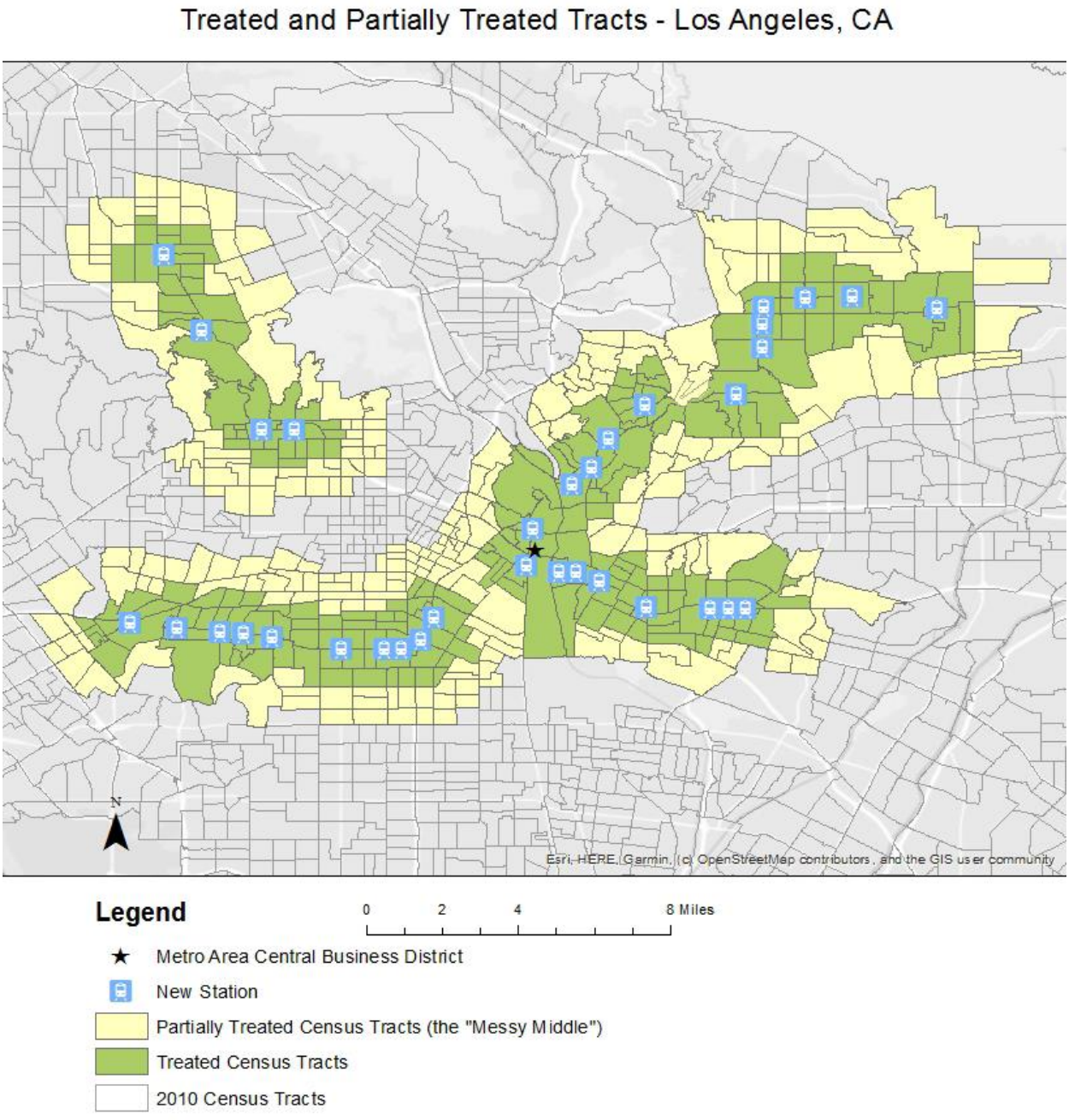

Figure 1: Treated and Partially Treated Tracts - Los Angeles, CA

a bachelor's degree or higher, percent use of public transit; and percent of rental units. The third model adds a control variable for the distance to the nearest central business district (CBD). This variable is then multiplied by the year variable to create a gradient. The gradient, interacting with each dependent variable, describes how the variable expresses itself through physical space. For example, the log mean rent gradient describes how rental prices change from the CBD out to the surrounding neighborhoods along the route of the rail line.

Finally, the fourth model utilizes a technique called effect heterogeneity, which attempts to understand whether the results of the model are falling on certain census tracts over others. Specifically, I test whether the effects of displacement are occurring in census tracts that have larger communities of color and those with lower educational attainment. I chose to focus on these two groups as the literature has also taken this approach, but also due to the discussions around displacement more generally. Advocacy groups representing communities of color and those with lower educational attainment are also discussed extensively in displacement literature (Cho, 2019; Feinstein & Allen, 2011; Sandoval, 2018).

Further analysis attempts to understand how and why certain census tracts experienced gentrification and displacement after the installation of rail transit. As mentioned by Kahn, stations considered "walk and rides" near higher density and without significant parking, experienced greater gentrification. Stations considered "park and rides" where little housing is built, experienced an increase in poverty rates (Kahn, 2007). This could indicate larger economic trends driven by land use decisions that may influence the level of displacement.

The following chapters discuss the current literature, my methods, results, and discussion. Finally, the conclusion summarizes my findings and expands on the limitations as well as opportunities for future research.

2.0 LITERATURE REVIEW

The existing research surrounding rail transit induced displacement is far from conclusive. Some researchers have found little evidence of the phenomenon while others have detailed significant displacement correlating with the installation of rail transit. In the following chapter I discuss the current research in the following four parts (1) framing theories, (2) current trends in urban reinvestment, (3) community voices, and (4) the current debate tying displacement to rail transit investments.

2.1 Framing Theories

Three authors set the stage for how I approached my analysis. Rayle's 2015 article details the gap in understanding between community groups and the research community. Much of the current literature does not find evidence of displacement occurring near rail transit; however, for years, community groups have claimed displacement is happening (Rayle, 2015). The second and third authors are Revington and Elliot-Cooper, who both discuss access to transit as both a class struggle and frame denying access to housing near transit as a form of violence (Elliott-Cooper et al., 2019; Revington, 2015).

Rayle's 2015 article was foundational in understanding the current disconnect within the literature regarding rail transit's effect on displacement. As Rayle discusses, much of the literature has been unable to directly tie rail transit, and the ensuing transit-oriented development (TOD), to displacement (Rayle, 2015). On the other hand, researchers have noted significant community voices that say displacement is occurring and ripping neighborhoods apart. The aim of my research is to shed light on these seemingly contradictory perspectives.

Rayle details four reasons for this disconnect between the literature and community voices. First, there are limitations to the methods typically employed for this type of study, most notably, the lack of data tracking displacement itself. The American Housing Survey (AHS) and the American Community Survey (ACS), both curated by the US Census Bureau, are often used in the research since they shed light on how households and communities are changing over time. However, neither are designed to measure displacement – put simply, no one is asking residents why they left a neighborhood or where they moved to. Due to these limitations, researchers are left to measure changes in variables that *may* indicate displacement is occurring, such as changes in property values or the racial composition of a census tract over time. As discussed in later chapters, I still utilize ACS data in my analysis due to its availability and comparability over time, even with the drawbacks. My analysis is not immune to Rayle's concerns.

Low-income people and communities of color are more likely to be renters and face significant challenges making ends meet as compared to homeowners (Scally & Gonzalez, 2018). Therefore, I choose to focus part of my analysis on the relationship between rental prices and rail transit installations, instead of using property values, as one indicator of displacement. It can be argued that property values are linked to rental prices, however the relationship is imperfect, and warrants further study (Gallin, 2008). Since low-income people and communities of color are most at risk, my analysis works to center their experiences and work to unpack any potential racial disparities regarding transit-led displacement. More on this is detailed in further chapters.

Second, Rayle proposes that the true number of displaced residents may be small, but the psychological effects of seeing your community transform may be significant. Residents may feel distressed seeing their community change and report that displacement is occurring to community groups and local government officials regardless of actual residential turnover. As mentioned above, bolstering these accounts becomes difficult given that existing data sources are not designed to measure displacement specifically.

Third, Rayle claims that TOD developments provide additional housing units while also allowing residents to lower their transportation costs by taking transit instead of driving. It is then argued that, while housing costs may increase, transportation costs may decrease. Since housing costs are often used as an indicator of displacement, there is an opportunity for this metric to lead research astray. I do not support this line of thinking in my analysis since oftentimes new units are significantly more expensive than older ones, leaving the benefit of lower transportation costs to wealthier residents who are less likely to be displaced in the first place. Furthermore, increases in TOD development likely does not completely explain demographic shifts associated with displacement throughout a census tract, such as a significant increase in the percentage of non-Hispanic white residents.

Fourth, Rayle argues that community groups may be leveraging discussions of displacement to claim development benefits for their community, regardless of the significance of the displacement threat. Stopping short of accusing community groups of lying about displacement pressures, Rayle argues there is a possibility that the reported displacement is overblown. While I believe it is important to keep the concerns of community members in perspective, choosing to soften their call for help threatens to understate the true needs of the

community. Those in the public sector must heed this inclination, instead balancing community concerns with relevant data. I aim to take this approach.

The second and third authors that frame my approach to the research are Revington and Elliot-Cooper, who both focus on class struggles surrounding access to transit (Elliott-Cooper et al., 2019; Revington, 2015). Revington sets gentrification and displacement into a Marxist perspective, stating that access to transit (and thus access to the neighborhoods near transit) are a class struggle between the rich and the poor. Inside an unregulated capitalist system, only the wealthy will have access to highly demanded public amenities like rail transit. Since renters are oftentimes lower income, this perspective further bolsters the importance of studying rental prices as one driver of displacement.

Elliot-Cooper categorizes displacement into direct and indirect, where direct refers to the physical removal of residents from a community, and indirect refers to the rapid change within a neighborhood that leads original residents to no longer feel at home. Relating to other forms of involuntary movement, "gentrification-induced displacement as a form of un-homing" (p.12) and ultimately a distinct form of violence (Elliott-Cooper et al., 2019). This understanding is also crucial to my approach, as I seek to understand how the installation of rail transit may contribute to these direct and indirect displacement pressures.

2.2 Current Trends in Urban Reinvestment

This section of the literature review helps set the stage for understanding displacement within the larger context of urban reinvestment seen in US cities over the last several decades. The connection between rail transit investments and rising property values is well documented (Cao & Lou, 2018; Heilmann, 2018; Mckenzie, 2013), with multiple authors detailing evidence

that property values increase near proposed rail stations, even before they are officially announced and built. This indicates strong market speculation that can lead to displacement even in neighborhoods that did not receive a rail transit station (Heilmann, 2018).

Dawkins finds that requiring new housing to include affordable units has a larger impact on providing space for low-income people to live near transit than the use of housing vouchers or attempts to increase the supply of housing near transit (Dawkins & Moeckel, 2016). I suspect that this may be true since increasing the supply of housing near transit takes considerable time, and the effects of increasing supply on prices takes even longer. If a housing shortage exists in an entire region, adding new housing near the station will likely not be enough to fulfill the demand, ensuring that the new units are not affordable to those with lower incomes. This is not to argue against increasing housing production to meet current housing shortages, but to acknowledge that the market cannot deliver affordability until supply comes closer to demand. In the meantime, low income people will find it difficult to afford market rate housing. These findings provide an understanding of how housing market pressures can present themselves near rail transit (or proposed) stations and highlight the importance of focusing on low-income renters specifically.

2.3 Community Voices

The perspective of the communities feeling displacement pressures near new rail transit stations is also crucial to understanding the divide between them and researchers studying these issues. Multiple researchers discuss community mobilizations to address displacement throughout the US (Cho, 2019; Feinstein & Allen, 2011; Sandoval, 2018). Representing mostly non-white neighborhoods, communities came together to form organizations such as the

United Neighbors in Defense Against Displacement (UNIDAD) based in south Los Angeles. Organizations like UNIDAD are looking to bring benefits to their communities by capitalizing on the need for additional housing. For example, UNIDAD has advocated for the Los Angeles city council to require developers to build affordable units in new residential buildings. Other groups have gone around local governments to draft community benefits agreements with individual developers to ensure the community's perspective is reflected in the final development.

Another important component to this research was conducted by Luckey et al., in which they discovered that residents of all incomes and races were interested in the benefits of transit-oriented development near rail transit and the lifestyle it can provide, however the largest concern was over housing costs (Luckey et al., 2018). Low-income people are likely unable to afford rent near transit, thus reinforcing my approach to study rental prices specifically[2].

2.4 Current Debate Tying Displacement to Rail Transit Investments

The current debate tying displacement to rail transit investments can be categorized into three categories of studies: studies that found little evidence of displacement, studies that found some evidence of displacement, and studies that found significant evidence of displacement. Studies focusing on where low-income renters live found no evidence that proximity to a rail station increased their likelihood of moving (Delmelle & Nilsson, 2019). As mentioned in the article, low-income renters often move more frequently than their higher-

[2] Some precedent exists for the construction of affordable ownership near transit. In southeast Seattle, affordable ownership opportunities are being built near the Othello Light Rail station. More information can be found here: https://living-future.org/affordable-housing-othello-square/.

income counterparts. Other studies have found no statistically significant displacement near rail transit stations in 14 metro areas in the US (Baker & Lee, 2019; Dong, 2017). Dong mentions, however, notable demographic changes near some stations in Denver and San Francisco but could not tie that change to the station itself.

Other studies have mixed results tying displacement to rail transit, including a comparison of similar 14 metro areas in the US (Kahn, 2007). This study found evidence of displacement near some "walk and ride" stations, while "park and ride" stations saw increases in poverty which may indicate the suburbanization of poverty[3]. A study of large Canadian cities found evidence of displacement, but not in Vancouver, BC, a city known for significant increases in housing costs (Grube-Cavers & Patterson, 2015). Another study of displacement in Denver found no evidence of displacement near rail transit stations but did find significant changes in demographic indicators usually associated with displacement like educational attainment (Bardaka et al., 2018). These studies display the wide array of perspectives in the literature surrounding rail transit induced displacement. They also offer an opportunity to understand how indicators of displacement may not be uniform across all cities and expecting them to be so could mask the truth – displacement is difficult to track, usually measured through related variables like educational attainment, rental prices, and other demographic information.

Finally, there are multiple studies that find significant relationships between rail transit and displacement. A study of 9 metro areas in the US found significant indicators of displacement for low-income communities while wealthier communities are less likely to feel

[3] I define the suburbanization of poverty as the decentralization of poverty from inner city neighborhoods out to the suburbs. This movement is often associated with the gentrification of inner city neighborhoods (Kavanagh et al., 2016).

these effects (Nilsson & Delmelle, 2018). Wealthier communities near transit are likely to experience increased densification after the introduction of rail transit, as one would suspect from changes in land use and market demand in a neighborhood. This study is limited, however, by its restriction to only light rail installations in nine US metro areas. My study focuses on light rail, heavy rail, and most commuter rail in 24 US metro areas, greatly increasing generalizability. More on the selection of rail transit is detailed in the methods section.

Another study found a significant relationship between the installation of rail transit and increases in the Non-Hispanic white population in south Seattle, as well as significant increases in racial diversity in Seattle suburbs (Hess, 2018). While this study shows gentrification and displacement occurring, it fails to distinguish if these changes are strictly due to the rail station or if larger economic trends were at play. Additionally, the study's focus on Seattle exclusively limits the generalizability of the results. Regardless, the study adds to the literature's understanding of how displacement is occurring.

Table 1: Evidence of Displacement Found in Prior Studies

	Little to no evidence of displacement	Some evidence of displacement at "walk & ride" stations	Significant evidence of displacement correlated with rail transit
Dong, 2018	x		
Delmelle, 2019	x		
Baker, 2019	x		
Kahn, 2007		x	
Grube-Cavers, 2015		x	
Bardaka, 2018		x	
Nilsson, 2018			x
Hess, 2018			x

As detailed in Table 1, the existing literature is mixed – some studies show evidence of displacement and others do not. All of them use secondary variables as evidence of displacement and gentrification, such as race, educational attainment, and age. The literature is still growing as our understanding of these pressures increases. Given what researchers hear from community groups, it may be easy to write off studies that do not find evidence of displacement. I chose to focus my research on the gap between reported study findings and the community's lived experience. As extensively discussed by Zuk et al., the current research may not accurately represent what is occurring in many urban neighborhoods, thus warrants further analysis (Zuk et al., 2018).

3.0 METHODS

My research focuses on the possible connection between the installation of a rail station and indicators of displacement in nearby census tracts. Treated tracts are defined as tracts with a population centroid that is within one mile of the new station. Tracts are eligible to be a control tract if their population centroid is further than two miles from the new station. Tracts with a population centroid one to two miles away from the station are considered a part of the "messy middle" and removed from the analysis. Using a difference-in-difference regression analysis, I test whether there is a significant change of indicators of displacement in treated tracts as compared to the control tracts, after a rail station is installed.

3.1 Census Data Collection & Interpretation

I select stations that were installed between 2000 and 2012 and use data from the 2000 Decennial Census and the 2013-2017 American Community Survey (ACS) to test whether the station contributed to measures the literature links to displacement. These measures include mean rent, educational attainment, age, and the census tract's racial makeup. I am interested in this time period since there is little literature that incorporates rail projects installed nationwide. These data also represent the most contemporary data sources released from the U.S. Census Bureau. Since the 1990s gentrification, displacement, and increased income inequality have become the hallmarks of America's urban neighborhoods, thus calling for more thorough and up to date analyses.

All census data is downloaded from SocialExplorer.com,[4] a repository for census data in its original form as well as modified for various research uses. In this case, I choose 2000 census data that has been normalized to 2010 geographies. Comparing data across census datasets that fall on either side of the decennial census is complicated. Every ten years the US Census Bureau redraws census tract boundaries, mostly driven by changes in population density. For example, if a census tract experiences significant population growth, the Bureau will split the tract in two.

The Bureau attempts to keep as many census tracts intact and splits or combines tracts where necessary. While these changes make comparability between census tracts easier, comparing over time is more complex. To combat this, Social Explorer created a version of the 2000 census data that has been normalized to the boundaries drawn during the 2010 decennial census. During the conversion, most variables can be salvaged, however any variable with a median value cannot be converted. Simply put, if one census tract is divided evenly in two, a median cut in half does not make two new medians. For example, median rent was forced into a mean rent, acknowledging that a mean is more susceptible to outliers.

The variables needed to complete each step of the analysis are defined in Table 2. The mean rent variable is converted into log mean rent to reduce the influence of outliers after being converted from a medial; however, these attempts may fail to reduce the unintended effect of outliers. Log mean rent, percent bachelor's degree or higher, percent non-Hispanic

[4] A data source for census data, in original form or modified. The organization was started by researchers at the City University of New York. Source: https://www.socialexplorer.com/

white, percent 65 plus are used as dependent variables in the models described below. They also act as control variables when not representing a dependent variable.

Table 2: Variable Definitions

Variable Name	Definition
Renter Occupied Tenure	Percentage of renter-occupied housing units
Percent Female	Percentage of female residents
Percent Non-Hispanic White	Percentage of Non-Hispanic white residents
Percent Age 65 Plus	Percentage of residents ages 65 and up
Percent Bachelors	Percentage of residents with a bachelor's degree or higher
Percent Public Transit	Percentage of workers who take public transit to work
Percent Rentals	Percentage of housing units that are renter-occupied
Log Mean Rent	The log of mean rental prices, in 2017 dollars
Distance 2 CBD	The distance (in miles) from a census tract to the nearest CBD

3.2 Identifying Line Segments

My next steps include collecting line segment data from the National Transit Database (NTD) and selecting segments that were opened during the study period. Line segments represent the train line and its corresponding stations, as it opened for service. Typically, a line segment is a section of rail that was funded as one piece. For example, the Eastlink Light Rail Project in Washington State is a new line segment from downtown Seattle to Downtown Bellevue. If the segment opened during the study period, each new station along the line would be selected.

Next I use the shapefiles for all transit lines in the United States from thetransportpolitic.com[5] selecting line segments that were opened between 2000 and 2012

[5] A collection of transit data collated by Yonah Freemark (Ph.D. candidate at MIT) and Steven Vance.

and identifying each new station (see the Appendix for a full list of metro areas with line segments included in the analysis). There may be segments that are not included in the data since the NTD only tracks train lines that received federal funding. For example, commuter rail that uses privately-owned tracks shared with freight rail traffic (such as the Sounder in Seattle, WA) is not included. However, the rail line in Austin, TX was built on former freight right of way, but freight no longer uses the route, and federal funding was used in its construction, so the line is included in the analysis. Regardless, these choices could lead to internal validity problems as the stations not included in the data may not be missing randomly. Further analysis should be done to understand their status.

3.3 Identifying Treated and Partially Treated Tracts (the "Messy Middle")

I use 2010 census tract data and their corresponding population centroids from the US Census Bureau, uploading into ArcMap along with the station location data mentioned above. Using a spatial analysis tool, I select the census tracts that surround each station in the data. A census tract will be considered "treated", or received a train station, if the tract's population centroid is located within a one-mile radius from the station. Station locations are often along large roads or existing rail lines which are frequently used to divide census tracts, so there is an opportunity for a treated tract to be missed by the spatial analysis tool.

Census tracts with a population centroid between one and two miles away from the station are removed from the analysis. The current literature tying displacement with rail investments argues that displacement pressures do not end one mile from the station, and are likely causing "spatial spillover effects" on tracts further away (Bardaka et al., 2018). I refer to these tracts as the "messy middle," as they may be feeling the effects of the rail investment,

however they are further from the station than would be considered walkable according to contemporary urban planning literature. Additionally, I removed the messy middle tracts for three specific reasons; (1) time and resource limit my ability to study their relationship with rail investment, (2) choosing to focus on tracts that are more likely to feel the effects of displacement, and finally (3) removing the messy middle tracts to ensure that they are not selected as control tracts. This third point is worth emphasizing – if there are spillover effects, these tracts cannot fully function as controls as they *too* are feeling the effects of the station. Simply stated, they do not make quality control tracts if they are influenced by the installation of the station.

This problem could also exist at the two or three mile mark out from the new station, where my model assumes that there is a significant difference between tracts on either side of this line. As before, I am relying on existing urban planning literature that states that a one mile walk to a station is considered the furthest a pedestrian will travel (Bardaka et al., 2018; Cervero, 2001). However, this assumption could limit the validity of my results and future research should test whether the results are sensitive to the removal or addition of tracts at this distance, similar to the approach taken by Bardaka and their colleagues (Bardaka et al., 2018).

3.4 Identifying Control Tracts – Propensity Score Matching

Identifying the control census tracts will be conducted using a regression analysis and propensity score matching. The regression is designed to identify census tracts within a US Current Metropolitan Statistical Area (CMSA) that are very similar to the census tracts that received a rail station, but themselves did not. In effect, I select census tracts that look quite

similar to the treated tract, and perhaps could have been considered for a rail station. Each tract receives a propensity score, which is its likelihood of being treated given the below variables. As with many difference-in-difference studies, there are many treated and control tracts that have similar propensity scores, which is expected. This means I have found tracts that are similar along the below variables, except one received the treatment and the other did not.

Table 3: Propensity Score Matching Variable

Variable Name
Percent Female
Percent Non-Hispanic White
Percent Age 65 Plus
Percent Bachelors
Percent Public Transit
Percent Rentals
Log Mean Rent
Distance 2 CBD

Once the treated and control census tracts are properly identified, I merge these data with the 2000 census data and the ACS data from 2013-2017. The difference-in-difference model structure requires merging these data into one spreadsheet so changes can be measured overtime for each census tract. Census data identifying every Current Metropolitan Statistical Areas (CMSA) throughout the US is downloaded and matched with the location of its Central

Business District (CBD).[6] The CMSA and CBD data are merged into the treated/control data allowing the Distance to CBD variable to be generated for each census tract.

3.5 Dependent Variables & Structure of Models

Each dependent variable, seen in Table 4, is subject to continually more complex models to test their relationship with the treatment, receiving rail. All models include year fixed effects and location fixed effects using the year and geo_fips variables, respectively. These approaches control for time and location in the regression. Year fixed effects are included so the model controls for time as it passes. Rent, for example, tends to increase over time so controlling for this is crucial. Additionally, location fixed effects are included to account for regional effects. For example, if one station is causing significant displacement pressures on the surrounding census tracts, the effect is not random; there is a grouping aspect to the results. Finally, every model is weighted using the renter-occupied tenure, in an attempt to ensure that census tracts with similar populations are used in each comparison.

Table 4: Dependent Variables

Dependent Variable
Log Mean Rent
Percent Non-Hispanic White
Percent Bachelors
Percent Age 65 Plus

[6] CBD location data curated by the economist Matt Holian. Sourced: http://mattholian.blogspot.com/2013/05/central-business-district-geocodes.html

3.6 Creating Model #1 – Simple Regressions

The first model is simple, testing whether any relationship exists between the dependent variable and the treatment while only controlling for year fixed effects. Additionally, I test whether this relationship is different inside and outside the common support region. The common support region identifies the area of overlap between propensity scores for all tracts, treated and control. Testing whether there is a difference in the model's outcome along these lines is not required for the difference-in-difference analysis, however it can check for consistency as the models progress in complexity. If, for example, the results of the model were significantly different inside and outside the common support region, it would indicate that my matching system is not working sufficiently.

3.7 Creating Model #2 – The Introduction of Controls

The second model builds on the first and includes the following control variables: percent female; percent Non-Hispanic white; percent ages 65+; percent bachelor's degree or higher; percent use of public transit; percent of rental units; and log mean rent. These variables are selected since they often appear in the existing literature, and, most importantly, they can ensure that the treated and control tracts make a quality match.

The introduction of the controls does provide an opportunity for omitted variable bias to skew the results. As with any regression, one must assume that there is no relationship between the dependent variable and the control variables other than through the treatment. For example, there must be no relationship between log mean rent and educational attainment, other than through the installation of a rail station. Now, of course, one could

think of many instances where rent would increase due to the educational attainment also increasing regardless of the distance from the rail station. The opposite may also be true.

Omitted variable bias is a significant problem within any regression analysis and cannot be avoided, merely mitigated. Techniques to reduce the power of omitted variable bias, such as the use of other control variables and testing inside and outside of the common support region, are implemented here. Additionally, these tests work to limit the effect of "torturing" the data to produce results one would like to see. Given the confidence intervals of 0.05, one can expect a significant result (a false negative) every 20 regressions. If my results remain consistent after time, this outcome becomes less likely. Regardless, these problems plague all social science analyses and future research must be done to limit its effects.

3.8 Creating Model #3 – The Introduction of Controls and Distance to CBD Variable

The third model includes the distance to the nearest CBD variable as an additional control, however only after it has been converted into a gradient by multiplying it by the year variable. The gradient is interpreted as the dependent variable over the distance to the nearest CBD. For example, the distance to the CBD control variable for log mean rent is interpreted as the change in the gradient of rents as one moves further out from the CBD. Foundational literature in urban economics has focused on the rent gradient, as it provides researchers an understanding of how rental prices are related to distance to a CBD and related transportation costs (Coulson & Engle, 1987). Going beyond log mean rent, I then create gradients for the other dependent variables of race, educational attainment, and age, as they may shed further light on the distributional changes observed over time and distance. For example, much of the existing research on displacement note that urban neighborhoods near the CBD have

transformed into whiter, younger, more highly educated areas. Using these three dependent variables in a gradient function has not been used in the literature, to my knowledge, and may carry risks of misinterpretation or skewed results.

3.9 Creating Model #4 – Tests for Effect Heterogeneity

The fourth model tests for effect heterogeneity, which could indicate whether a specific subset of tracts is more or less susceptible to displacement pressures brought by the new station. For example, research has found that communities of color are more likely to be displaced over their white counterparts (Hess, 2018). To test this hypothesis, I multiply the race control variable (prntnhwhite) by the treatment, creating a dummy variable called "treatedwhite". This new variable will test whether a tract with varying levels of Non-Hispanic white residents is feeling the pressures of displacement over others. This approach is crucial, since the previous models may be hiding disparities between groups.

I also utilize effect heterogeneity to test whether the educational attainment of a treated tract has any impact on its susceptibility to displacement pressures. For example, are tracts with lower educational attainment more likely to see displacement? As detailed in Kahn's research, changes in educational attainment are a key indicator that displacement is occurring (Kahn, 2007). It would follow that communities with lower levels of educational attainment are likely to be of lower income, and thus could not weather significant changes in housing costs that are another hallmark of displacement. To test this hypothesis, I multiply the educational attainment control variable (prntbachelors) by the treatment, creating a dummy variable called treatedbachelors. This new variable will test whether a tract with differing levels of educational attainment is feeling the pressures of displacement.

Throughout each model, I test whether each of these results differ both inside and outside the common support region, again to test the accuracy of the propensity score matching. Additionally, I employ a clustering method to group census tracts together that were treated by the same station. This has the effect of reducing the standard errors in my analysis, further refining the results. All results are evaluated for a p-value of less than 0.05. The null hypothesis is that there is no difference between treated and control tracts. The use of nation-wide data was strategic as it allows for generalizability across the US, providing further substance to the results.

3.10 Method Limitations

My research is focused on the significant gaps in the literature between what the research shows and the dire calls from community groups. Given this, there is a chance that my analysis is influenced by the same pressures that keep the existing research from aligning with the experience of people living in the treated tracts. Due to the limitations of my research and data, I am unable to determine whether any displacement pressure is *completely* due to the rail station, and not an omitted variable such as larger economic trends, poor data collection, community preferences, changes in government policy, and so forth. These pressures may mask the true impact rail stations have on treated census tracts. All regression analyses, especially those in social sciences, must face these threats to validity and acknowledge where they may lead researchers astray.

4.0 RESULTS

My research is focused on understanding the relationship between large, public rail investments and possible displacement effects on the surrounding community. The existing literature on transit-led displacement is far from conclusive, oftentimes claiming displacement is not occurring. A gap exists between what the research shows and dire calls from communities near new rail stations (Rayle, 2015; Zuk et al., 2018). This study uses recently published US Census data (2000-2017) to answer questions, including whether the introduction of rail transit is associated with changes in indicators of displacement in the receiving census tracts, such as log mean rent, educational attainment, age, and changes in the racial makeup. Utilizing a difference-in-difference regression model and a propensity score matching system, I select census tracts that received rail transit and test the relationship between the construction of rail transit and changes in these indicators of displacement. Specifically, my research sheds significant light on how rail transit affects low income communities and people of color, who are more likely to be affected by displacement.

4.1 Identifying Treated & Partially Treated Tracts (the "Messy Middle")

My analysis identified 1,125 treated tracts, 1,452 partially treated tracts, and 57,162 eligible control tracts, a robust sample size for the analysis (see Table 5 & Table 6). Identifying the treated tracts was simple, selecting their corresponding population centroids that are within one mile of the new stations. There were a few mistakes, however, where the census tract that received the rail station was not selected as treated since the centroid was far from the station. This typically occurred in census tracts that are low in population, and the station was built in commercial centers or rural areas.

Table 5: Summary Statistics for Treated Tracts

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
treated	1,125	1	0	1	1
renteroccupiedtenure	1,125	1,144	590.8	0.193	3,117
prcntfemale	1,125	0.497	0.0503	0.0203	1
prcntnhlwhite	1,125	0.387	0.289	0	1
prcnt65plus	1,125	0.103	0.0592	0	0.429
prcntbachelors	1,125	0.263	0.189	0	1
prcntpublictransit	1,125	0.151	0.137	0	0.867
prcntrentals	1,125	0.691	0.197	0.0171	1
logmeanrent	1,125	6.446	0.330	4.381	7.712
year	1,125	2,000	0	2,000	2,000
distance2cbd	1,125	6.085	5.769	0.0351	35.73
ptreated	1,125	0.0675	0.0317	-0.186	0.325

Table 6: Summary Statistics for Control Tracts

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
treated	58,491	0	0	0	0
renteroccupiedtenure	58,491	882.2	634.5	1.34e-06	8,556
prcntfemale	58,491	0.514	0.0347	0	1
prcntnhlwhite	58,491	0.595	0.310	0	1
prcnt65plus	58,491	0.120	0.0681	0	0.942
prcntbachelors	58,490	0.252	0.177	0	1
prcntpublictransit	58,487	0.0941	0.158	0	1
prcntrentals	58,491	0.495	0.240	1.61e-07	1
logmeanrent	58,448	6.398	0.370	-6.857	8.132
year	58,491	2,000	0	2,000	2,000
distance2cbd	57,208	12.39	11.98	0.0283	177.8
ptreated	57,162	0.0302	0.0331	-0.210	0.335

Selecting the partially treated tracts, the "messy middle", was also largely successful in identifying tracts that are one to two miles from the new stations. Messy middle tracts are defined as tracts that are too far from the station to be considered treated, however may still be influenced by spillover effects from the station. These tracts are removed from the analysis so that they are not considered eligible as a control tract. Simply put, they cannot fully function

as a control tract since they may "feel" the effects of the station opening. This technique is not without flaws. As seen in Figure 2, some tracts that received rail are considered a messy middle, since the population centroid was too far away to be considered treated. In other

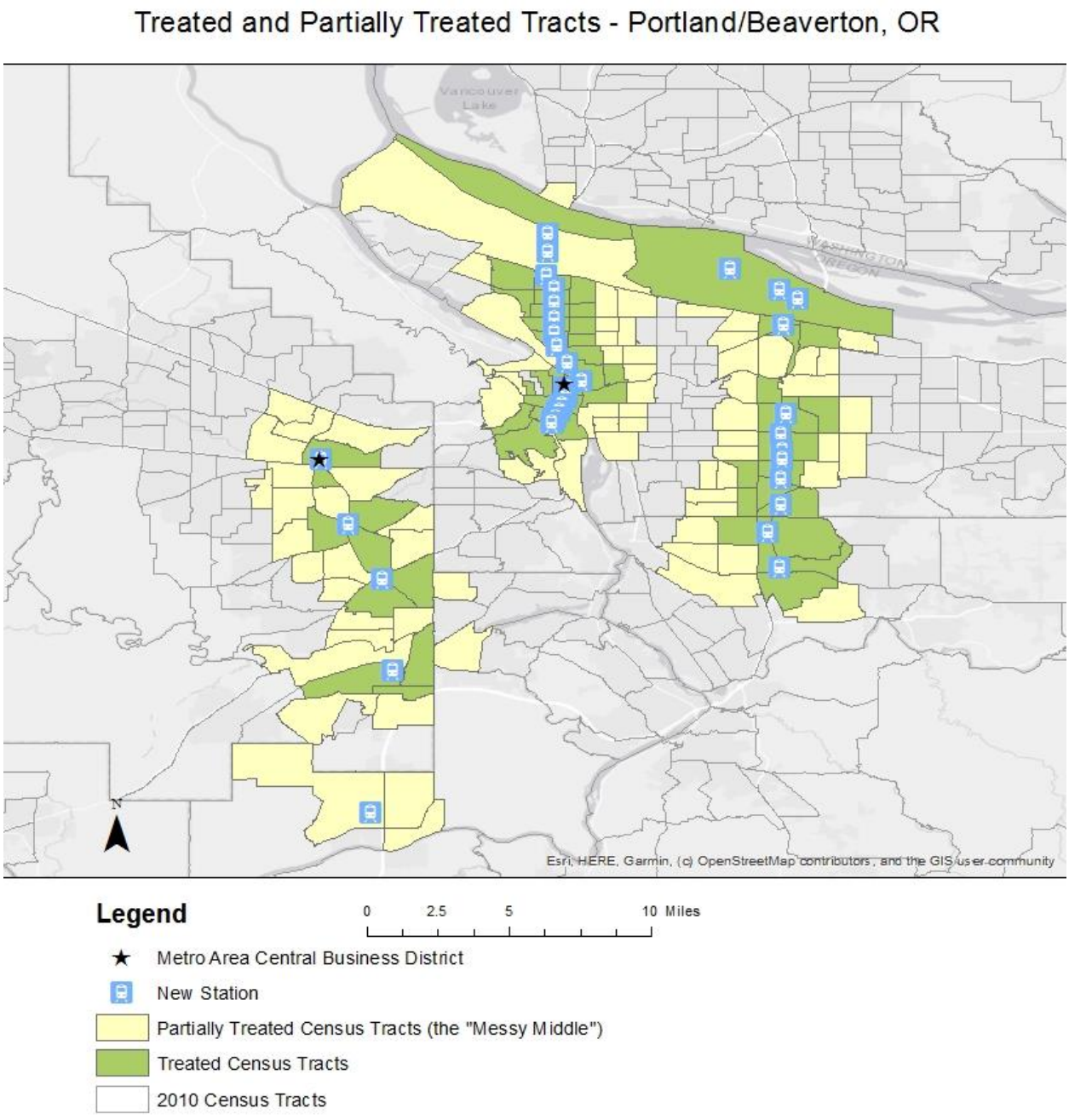

Figure 2: Treated and Partially Treated Tracts - Portland/Beaverton, OR

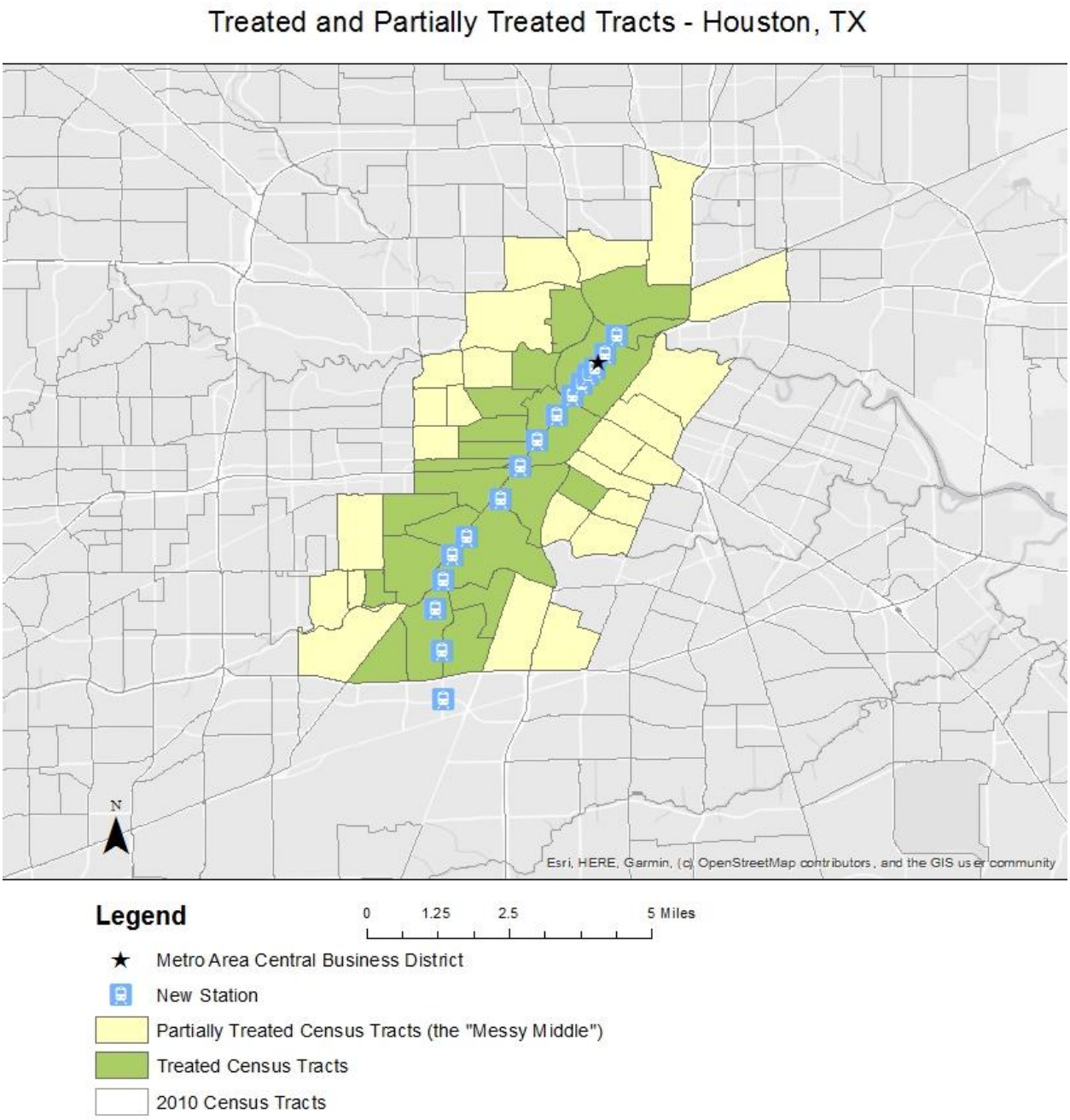

Figure 3: Treated and Partially Treated Tracts - Houston, TX

cases, a tract that received rail was not considered treated or a messy middle tract, since the population centroid was even further away from the station (see Figure 3). These errors are important and future researchers should work to rectify them. However, they represent a

small fraction of my data and will likely not skew the results. The results for all other metro areas included in the study did not have these errors, and can be found in the Appendix A. A list of included metropolitan areas and station counts can be found in Appendix C.

4.2 Identifying Control Tracts – Propensity Score Matching

Eligible control tracts are defined as census tracts within a Current Metropolitan Statistical Area (CMSA) and has a population centroid further than two miles from any new rail station. Of the 73,000 census tracts in the US, 57,162 were identified as control tracts. As seen in Table 6, not all variables have the same "N" or number of tracts, with a few hundred tracts missing from the last few variables. This is likely due to missing data from the Census Bureau. While a pattern for their omission was not found, future researchers should investigate why these tracts may be missing data, and thus could not be included in the analysis.

As seen in Table 5 and Table 6, the "ptreated" variables display the minimum and maximum likelihood that a tract would be treated with a train station, with the treated and control tracts expressing a 32.5% and 33.5% likelihood respectively. This overlap means there are tracts in both groups that look quite like each other, based on the matching variables, except one was treated and the other was not. This variety bolsters confidence in the quality of the matching between treated and control tracts. It is important that the treated and control tracts seem like each other, for my results to be plausible. The regression results that created the ptreated variables can be found in Table 7.

4.3 Model #1 – Simple Regressions

The results of the first model indicate a strong relationship between the treatment and log mean rent, educational attainment, race, and age in census tracts near the stations. As

seen in Table 8, all variables are significant with the large majority indicating a p-value of less than 0.001 with the exception of age which indicates a p-value of less than 0.05. Testing both inside and outside of the common support region produces similar results, increasing confidence in the relationships between the treatment and the dependent variables.

4.4 Model #2 – The Introduction of Controls

Taking it a step further with the second model, I add control variables for sex, race, age, educational attainment, use of public transit, percentage of rental units, and log mean rent (see Table 9). The inclusion of control variables removes these variables from potentially influencing the relationship between the treatment and the dependent variable; however, this only works when we are sure there is no relationship between the control variable and the dependent variable other than through the treatment. This is likely not the case. For example, rental prices may be influenced by the scarcity of rental units, such as the "percent of rental units" control variable, regardless of proximity to a rail station. This type of omitted variable bias is an unfortunate reality in social science research. Future research must be done to understand its potential effects on the results.

As seen in Table 9, the introduction of control variables erases the relationship between the treatment and log mean rent, which has a p-value of greater than 0.05. This result could indicate that the treatment is not a great indicator of where log mean rent will change. Rental prices seem to be driven by variables not in this model, such as economic conditions, access to other types of transportation, community preferences, or land use decisions. I imagine there are circumstances where rents are high in census tracts that are far from a train station, either due to the local municipality's lack of investment in rail or the community's preference to be

disconnected by public transit. A comparison inside and outside of the common support region does not yield differing results, adding confidence to the conclusions regarding log mean rent.

The relationship between the treatment and educational attainment *does* hold after the introduction of the control variables (see Table 9), with all control variables expressing a p-value of less than 0.05. Simply put, treated tracts do see a significant increase in educational attainment as compared to the control tracts, with these results holding both inside and outside the common support region. Treated tracts saw an increase of 2.2% in residents with a bachelor's degree or higher.

Interestingly, the coefficient for percent of rental units is negative, indicating that treated tracts that saw an increase in educational attainment had fewer rental units than the controls. This may reflect the fact that many rail systems are built through suburban contexts which may lack rental units near the station. Planning departments around the country looking to alleviate housing affordability crises as well as limit auto-dependency may find this result to justify transit-oriented development (TOD). TOD is one policy tool that can provide an opportunity to couple affordable rental housing and transit investments for those who could not afford to purchase property near the station.

As in the simple regression, the relationship between race (percent Non-Hispanic white) and the treatment remained significant after the introduction of the control variables, both inside and outside of the common support region. With a p-value of less than 0.001, treated tracts saw a 2.8% increase in Non-Hispanic white residents after the station was built. The results also indicate that treated tracts tend to gain older residents as they become whiter, which does not fit into the traditional understanding of displacement which typically attracts

younger residents. Future research should be conducted to understand this relationship further.

Adding the control variables to the regression changed the relationship between age (percent over 65 years) and the treatment. Outside the common support region stayed significant, with a p-value of less than 0.001, and inside the common support region is not significant at all. This indicates that the distribution of age groups is likely influenced by other factors not included in the model; however, the sign of the coefficient has stayed negative through the successive models, which does fit our traditional understanding of displacement which attracts younger residents to neighborhoods near transit investments.

4.5 Model #3 – The Introduction of Controls and Distance to CBD Variable

The next third model (see Table 10) incorporates the same control variables as indicated above; however, I also include the distance to the nearest central business district (CBD) using the variable "distance2CBD". This variable is designed to measure the 'gradient' of change in the dependent variable over the distance between a tract and its nearest CBD. I apply this gradient to each dependent variable, testing whether a rental gradient, educational attainment gradient, race gradient, and age gradient change over the time period. This variable may shed light on different land use and contexts found in many American cities as you move outward from the CBD. For example, many American cities have a relatively dense urban core that is susceptible to displacement, and relatively less dense suburban neighborhoods forming a ring around the metro area. Changes in the gradient of each dependent variable contributes to the story of change in American metro areas, and allows for an additional control variable, so that a

treated tract is compared to a control tract that is at a similar distance from their respective CBDs.

While the relationship between the treatment and log mean rent continues to be insignificant, the rental gradient is positive, albeit small. This result holds both inside and outside the common support region. A positive rental gradient implies that rental prices are flattening, becoming more uniform across the distance from the CBD. One could speculate that rental prices in urban neighborhoods were lower than in suburban contexts and have since risen. It is also possible that rental prices are increasing in the suburbs to match rental prices in neighborhoods near the CBD. Unfortunately, none of these models cannot infer causality or the direction of the relationship. Future research should be done to expand our understanding of the connection between rental prices and distance to the CBD.

The introduction of the distance2CBD variable does not hinder the relationship between educational attainment and the treatment, maintaining a p-value of less than 0.001, both inside and outside the common support region (see Table 10). The distance2CBD variable is not significant, indicating that the educational attainment gradient has not changed, which does not fit with the typical understanding of displacement. I suspect that the distance from the CBD may not be a great way to understand the distribution of educational attainment. As a remnant of redlining, pockets of wealthy, educated white residents exist near the CBD which could skew the analysis. Put more simply, no American city displays an identical educational attainment gradient emanating from the CBD and changing equally through space.

As in the previous models, the relationship between race and the treatment continues to be strong, with a p-value of less than 0.001 inside and outside the common support region

(see Table 10). The Distance2CBD variable is significantly negative (p-value < 0.001) indicating that central neighborhoods are becoming whiter as compared to the peripheral neighborhoods. This may provide evidence of displacement as well as the suburbanization of poverty, as communities of color are pushed further from the CBD. It may also indicate that racial segregation is increasing over the time period.

The introduction of the distance2CBD variable has partially removed the relationship between age (percent of age 65 and over) and the treatment. With a p-value of greater than 0.05 within the common support region, there is no significant relationship (see Table 10). However, the distance2CBD variable is significantly positive, meaning that the age gradient has flattened. This could be explained by the movement of retirees into urban areas as well as the diversification of the suburbs. The relationship between age and the treatment does stay significant outside the common support region with a p-value of less than 0.001. This could be the result of poor matching between the treated and control tracts, caused by another an omitted variable that has significant influence on the relationship between the age and the treatment. Further research should explore this relationship and, until it's understood, the results here should be thoroughly questioned.

As seen in Table 10, only educational attainment and race dependent variables were significant within the common support region. In both cases, the percent of rental units control variable were both negative with a p-value of less than 0.001. This indicates that the percentage of rental units in a treated tract decreased after the construction of the rail station. A decrease in the percentage of rental units could be caused by an increase in owner-occupied

units, or a decrease in the number of rental units. Either way, these changes could make it more difficult for low income people to have access to neighborhoods with transit investments.

4.6 Model #4 – Rent & Educational attainment DV's with Tests for Effect Heterogeneity

The fourth and final set of models test for effect heterogeneity, specifically looking for variations in the effect of the treatment on tracts that differ on each dependent variable. For example, as seen in Table 11, I test whether the relationship between educational attainment and the treatment is affecting tracts with differing racial makeups. I also tested whether the relationship between log mean rent and the treatment was heterogeneous; however, the coefficient on the treatment was not significant, with a p-value of greater than 0.05.

My results indicate that tracts with more non-Hispanic white people see a slight decrease in educational attainment after treatment (p-value less than 0.001). The opposite is also true – tracts with higher percentages of people of color are seeing increases in educational attainment after treatment, which fits into the current understanding of displacement. In terms of educational attainment, non-Hispanic white communities are not seeing the same displacement pressures as communities of color.

4.7 Model #4 – Race & Age DV's with Tests for Effect Heterogeneity

As seen in Table 12, the dependent variables for race and age are tested for effect heterogeneity. As before, I test for whether the relationship between the dependent variable and the treatment differs across educational attainment and race. The relationship between the treatment and the race dependent variable continues to stay strongly significant, with a p-value of less than 0.001. The introduction of the effect heterogeneity term, "treatedbachelors" provides more evidence of displacement occurring. Tracts that have lower educational

attainment are seeing significant increases in non-Hispanic white residents after treatment, as compared to neighborhoods with higher educational attainment.

As was true in the third model, the relationship between the treatment and the age dependent variable is not significant, except for outside the common support region (see Table 2). As before, the result being significant (p-value less than 0.05) outside the common support region calls into question the effectiveness of the propensity score matching. That said, the coefficient is negative on the treatment, and positive on "treatedbachelors" variable, which tests for effect heterogeneity. Simply put, treated tracts with higher educational attainment are seeing decreases in age after treatment. This does not fit into the general understanding of displacement which states that areas of lower educational attainment would be more susceptible to displacement pressures. However, it could be speculated that all areas that receive rail are drawing younger people, including areas with high educational attainment. More research should be conducted to further understand this relationship.

4.8 Summary of Results

The process of selecting of treated, partially treated, and control tracts was largely successful. With over 1,100 treated tracts and over 58,000 eligible control tracts, the regression analyses had ample data to create significant results that are generalizable across the country. The regression analyses were intended to test the relationships between the treatment and dependent variables over progressively more complex set of models. This particular structure was chosen since it provides further confidence in my results if the relationship between variables can hold together as the models increase in complexity.

Even so, there were several surprises in the results. The initial intent of this research was to shed light on how rental prices could be an indicator of displacement. Current research often used property values instead of rental prices. I hypothesized that focusing on renters would be crucial since the communities most at risk of displacement are more likely to be renters. The results indicate that rental costs are loosely correlated with the installation of rail stations, however that relationship fell apart with the introduction of control variables. I suspect this is due to the many omitted variables that could be influencing the relationship. Additionally, there are likely many factors influencing rental prices other than the installation of a rail station. It does not take long for one to think of a neighborhood where rental prices have risen significantly without the introduction of rail, like the Central District in central Seattle, WA.

Another surprise result was the relationship between age and the treatment. Oftentimes displacement is blamed on younger, wealthier residents moving into an urban neighborhood after the community has suffered from years of disinvestment. With this trend in mind, one would suspect that wherever a rail station is built, the surrounding census tracts would trend younger. My results do not indicate a relationship between age and the treatment after the introduction of controls. There is also a competing theory that retirees are beginning to move into urban neighborhoods at higher rates than in the past. There is a chance that these two demographic shifts are cancelling each other out.

Other interesting results came from the relationship between the remaining dependent variables, educational attainment and race, and the treatment. As seen in Table 10, the coefficient on the treatment variable for both dependent variables were about 0.02 (with p-values of less than 0.001). Across the entire country, wherever rail stations were built, the

surrounding census tracts saw a two percent increase in people with a bachelor's degree or higher or a two percent increase in non-Hispanic white residents. These results are significant since they occurred in any American context, regardless of land use, station planning, local economic trends, government policy, etc. Within the data's time period, local variations did not matter – if a station was constructed, it caused significant changes in indicators of displacement.

The effect heterogeneity models, seen in Table 11 and Table 12, draw a number of important conclusions. While I included all four dependent variables in this model, only educational attainment and race are significant, so I focus on those. I hypothesize that the results seen in the previous model is not uniformly felt across all treated tracts. For example, are the increases in educational attainment seen in all tracts, or are some seeing more substantial change? The results indicate that treated tracts with lower educational attainment saw changes in their racial makeup after the station was built, with the percentage of Non-Hispanic white residents increasing the most significantly. The results also indicate that treated tracts with a larger percentage of communities of color saw changes in the educational attainment after the station was built, with the percentage of residents with a bachelor's degree of higher increasing the most.

As with all regression modeling in social sciences, it would be a mistake to not mention the limitations of the results. Perhaps most importantly, there is a significant chance that omitted variable bias could be influencing the results. The regression model structure requires that there be no relationship between a dependent variable and an independent variable other than through the treatment. For example, there must be no relationship between log mean

rent and race, other than through the installation of a rail station. Now, one can imagine many instances where race, influenced by class, wealth, or privilege, could influence the rental prices someone can afford regardless of the distance to a rail station. These types of threats to validity exist throughout all social science research and are hard to avoid. In any case, the best a researcher can do is address them and work to limit their affects.

Regardless, the results point to significant demographic shifts occurring near rail investments, regardless of location, with the most change occurring in communities of color and communities with lower educational attainment. Due to systemic oppression and chronic lack of access to opportunity, these communities have the fewest resources to manage significant changes in their neighborhoods. Policy makers and urban planners looking to make large public investments like rail transit must confront these unintended consequences if they are to work towards increasing access and creating more equitable communities.

5.0 DISCUSSION

Much of the literature fails to tie displacement to rail investments, while at the same time the literature notes significant outcry from communities watching their neighborhoods transform (Rayle, 2015; Zuk et al., 2018). My research aims to address this gap by approaching the analysis in a slightly different fashion: by using the most recent data from the Census Bureau, focusing on rental prices which are more likely to impact communities of color, and using national data to enhance the generalizability of the results.

Displacement itself is not measured in Census data, however researchers can use other variables, or indicators, as a proxy for displacement. I chose to study changes in the following four indicators: log mean rent, educational attainment, race, and age. Focusing on indicators is similar to measuring someone's pulse and making assumptions on the condition of their heart – it's not an incorrect approach, but the tools used are woefully underprepared to measure the condition. However, much of social science faces these concerns as well as the threat of omitted variable bias. Unfortunately for researchers, the world is too complex to be accurately represented in the data or a statistical model. In the absence of a perfect answer, I have structured my models in ways to limit these negative effects.

The results of my analysis indicate that there is significant evidence of displacement occurring near new rail stations throughout the United States. This result holds even when controlling for numerous variables. Of the four dependent variables, educational attainment and race were positively related with the treatment. On average, census tracts that receive rail see an increase of 2.1% of residents with a bachelor's degree or higher. Additionally, census tracts that receive rail see an increase of 2.2% of non-Hispanic white residents. This result is

consistent across the entire country, regardless of the state or territory, government policy or land use, spanning 1,125 treated census tracts (see Table 5).

Through an effect heterogeneity analysis (see Table 11) I find that the effects on educational attainment are not consistent across all treated tracts. Census tracts with higher percentages of communities of color are seeing disproportionate increases in educational attainment, a classic sign of displacement. Communities of color are being affected at higher rates than their white counterparts, even when controlling for numerous variables. The reason why is harder to answer. One could speculate that, due to redlining and racial covenants, communities of color were forced into neighborhoods near the CBD, placing them right in the path of soon-to-be rail systems.

Through the second effect heterogeneity analysis (see Table 12) I find changes in the racial makeup of a census tract is not consistent across all treated tracts. Census tracts with low educational attainment are seeing significant increases in the percentage of non-Hispanic white residents after the station is constructed. Demographic shifts like these are not occurring in tracts with higher percentage of non-Hispanic white residents, or tracts where more residents with bachelor's degrees or higher. The evidence from both effect heterogeneity models conclude that communities of color, as well as communities with lower educational attainment, are shouldered with significant neighborhood change after the construction of a rail station.

The other two dependent variables, log mean rent and age, were not significant after the addition of various controls. This result suggests that rental prices and changes in age may have happened regardless, even if the construction of the rail station was canceled. I suspect

this could be due to larger socio-economic trends that affected both the treated and control groups. This result may also be due to the many omitted variables that could influence the relationship between these variables and the treatment. Difference-in-difference regression models assume that there is a strong relationship between the dependent variable and the treatment, and that the effect is not covered or diminished by the control variables. As is with any social science analysis, there is no way to be completely confident that this is not occurring.

The long mean rent dependent variable may be influenced by the control variable "percent of rental units". Since rail stations can increase housing production in nearby census tracts, controlling for the percentage of rental units may not allow the model to adjust for changes caused by the station. Put more simply, a control variable holds the variable constant, so it cannot fluctuate with the treatment. Of course, this example also shows that an omitted variable bias likely exists, where a relationship between log mean rent and the percentage of rental units is being influenced by something else other than the treatment. This results in a control variable that cannot be considered independent, which threatens the validity of my results. Future research should be completed to understand this and how it could influence the relationship between the dependent variables and the treatment.

The age dependent variable was found to be significant after the control variables were added in the Third Model, however only *outside* the common support region. The relationship between age and the treatment was no longer significant *inside* the common support region. This result questions the robustness of the propensity-score matching system used. Simply put, the matching did no good, as the census tracts that were deemed too dissimilar to each other were a better predictor of displacement based on age. More likely, the distribution of ages

throughout American metropolitan areas is influenced by many things other than the construction of rail transit. As seen with the education and race dependent variables, treated tracts often shift younger, however it represents one small piece of the age distribution.

After completing the analysis, and in discussion with my committee, additional questions were raised regarding shifts in treated census tracts. Future research could focus on changes in the number of housing units near the stations. If there are significant increases in housing units, this may suggest that demographic shifts are due to new residents occupying new units, and thus not fueling displacement. However, new residents may also move into existing units which could spur displacement. This dispersion of new residents will be key to understanding the nuances of displacement. Further research could also focus on changes in the renter/owner ratio near new stations, as this will provide additional perspectives on how neighborhoods may be transforming regardless of whether new units are being built. Since communities of color are more likely to be renters, studying changes in the renter/owner ratio will provide a unique understanding of their experience.

Situated within the current literature, my results are similar to other studies. While it is nearly impossible to conclusively tie a rail investment to displacement, I find evidence of it across the entire country, similar to two researchers mentioned in Table 1 (Hess, 2018; Nilsson & Delmelle, 2018). My research did not distinguish “walk & ride” stations from “park & ride” stations, as other researchers did, however my focus on census tracts within one mile of the station was intended to select neighborhoods within walking or biking distance (Bardaka et al., 2018; Grube-Cavers & Patterson, 2015; Kahn, 2007). These studies found some evidence of

displacement near this type of station, and my results may bolster this result. Further research should be done to distinguish station types.

Other researchers found little to no evidence of displacement, and in some ways, I ran into similar results (Baker & Lee, 2019; Dong, 2017; Nilsson & Delmelle, 2018). The log mean rent and age indicators did not have a significant relationship with the treatment in my results, which could be the consequence of omitted variable bias or data limitations. I suspect that the other researchers have ran into similar barriers.

Additionally, it must be noted that the structure of my analysis does represent a limited approach, as I have used dependent variables as control variables. For example, I test the relationship between race and the treatment, but then use race as a control variable for the relationship between rent and the treatment. This breaks the assumption that the control variables are independent from the dependent variable, other than through the treatment. Regression analyses require that there is no relationship between rent and race other than through the construction of a rail station. This is obviously false as there are clear disparities between races in regard to income which would influence their ability to pay market rate rent. Communities of color have been historically barred from building generational wealth and currently face barriers to employment, education, and certain neighborhoods, among others. All of which influence the racial makeup of a census tract long before a rail station is built.

The installation of new rail lines amplifies these disparities, providing another way for new residents and speculators alike to benefit from rising housing costs. Original residents who own their home can take advantage of increased housing values, however selling their property could mean leaving the neighborhood entirely. The remaining residents may choose to leave

voluntarily as culturally significant businesses, churches, and community groups have also left. Even if not displaced, residents may feel that they are no longer welcome in their community.

Ultimately, large public investments such as rail expansions can sometimes work to inflame existing inequality and social problems. Systemic racism upholds wealth and income inequality, which keeps communities vulnerable to displacement. While a large challenge, transit agencies and governments have the opportunity to address racial and class inequities through rail investments. This could be accomplished through providing direct support to the most vulnerable while constructing the rail system – a form of reparations intended to help communities of color and those with low educational attainment from being disproportionally impacted. This could express itself as simple payments to residents or the creation of community land trusts, among others. Additionally, all levels of government must work to reduce forms of racism that keep communities of color from generating wealth in the first place. Increasing access to homeownership and cracking down on racial disparities in employment would be places to start. In the end, government aid is needed to address the root causes of the vulnerabilities that rail installations, or any public investment, can exacerbate.

Future research should further discuss the intersection of inequality and racism with large public investments such as rail transit. The benefits of providing rail are significant; increases in access, reduced CO2 emissions, alleviating income inequality, and rising property values can provide benefits to existing residents and the country as a whole. However, we must not forget that these benefits may not be available to communities of color and those with lower educational attainment. Rail investments must be made, but not at the expense of

those who cannot afford to stay in their community. We need to reevaluate how we provide public investments, and ask ourselves who is truly benefiting?

6.0 CONCLUSION

My research focused on understanding the relationship between large, public rail investments and possible effects on the surrounding community. The existing literature on transit-led displacement is far from conclusive, oftentimes claiming displacement is not occurring. A gap exists between what the research shows and dire calls from communities near new rail stations (Rayle, 2015; Zuk et al., 2018). I address this gap by approaching the analysis in a slightly different fashion: by using the most recent data from the Census Bureau, focusing on rental prices which are more likely to impact communities of color, and using national data to enhance the generalizability of the results.

Using a difference-in-difference regression analysis, my results show that significant changes are occurring in census tracts near new rail stations across the country. Spikes in the percentage of Non-Hispanic white residents as well as residents with higher educational attainment, are seen in the years following the opening of a new rail station. These changes are occurring more substantially in census tracts with higher percentages of people of color as well as census tracts with low educational attainment. These results bolster our understanding of the connection between public investments and neighborhood change that could indicate displacement is occurring.

The benefits of providing rail are significant; increases in access, reduced CO2 emissions, alleviating income inequality, and rising property values can provide benefits to existing residents and the country as a whole. That said, the results of the analysis suggest that large public transportation investments are amplifying displacement. Due to systemic oppression and chronic lack of access to opportunity, communities of color and people with low

educational attainment have the fewest resources to manage significant changes in their neighborhoods. Policy makers and urban planners looking to make large public investments like rail transit must confront these unintended consequences if they are to work towards increasing access and creating more equitable communities.

REFERENCES

Baker, D. M., & Lee, B. (2019). How Does Light Rail Transit (LRT) Impact Gentrification? Evidence from Fourteen US Urbanized Areas. *Journal of Planning Education and Research*, *39*(1), 35–49. https://doi.org/10.1177/0739456X17713619

Bardaka, E., Delgado, M. S., & Florax, R. J. G. M. (2018). Causal identification of transit-induced gentrification and spatial spillover effects: The case of the Denver light rail. *Journal of Transport Geography*, *71*(June), 15–31. https://doi.org/10.1016/j.jtrangeo.2018.06.025

Cao, X. (Jason), & Lou, S. (2018). When and How Much Did the Green Line LRT Increase Single-Family Housing Values in St. Paul, Minnesota? *Journal of Planning Education and Research*, *38*(4), 427–436. https://doi.org/10.1177/0739456X17707811

Cervero, R. (2001). Walk-and-Ride: Factors Influencing Pedestrian Access to Transit. *Journal of Public Transportation*, *3*(4), 1–23. https://doi.org/10.5038/2375-0901.3.4.1

Cho, S. E. (2019). The People's Plan: A Model For Community-Led Community Planning and The Fight Against Displacement Amidst Transit-Oriented Development. *University of Pennsylvania Journal of Law and Social Change*, *22*(3), 149–183.

Coulson, N. E., & Engle, R. F. (1987). Transportation costs and the rent gradient. *Journal of Urban Economics*, *21*(3), 287–297. https://doi.org/10.1016/0094-1190(87)90003-9

Dawkins, C., & Moeckel, R. (2016). Transit-Induced Gentrification: Who Will Stay, and Who Will Go? *Housing Policy Debate*, *26*(4–5), 801–818. https://doi.org/10.1080/10511482.2016.1138986

Delmelle, E., & Nilsson, I. (2019). New rail transit stations and the out-migration of low-income residents. *Urban Studies*. https://doi.org/10.1177/0042098019836631

Dong, H. (2017). Rail-transit-induced gentrification and the affordability paradox of TOD. *Journal of Transport Geography*, *63*, 1–10. https://doi.org/10.1016/j.jtrangeo.2017.07.001

Elliott-Cooper, A., Hubbard, P., & Lees, L. (2019). Moving beyond Marcuse: Gentrification, displacement and the violence of un-homing. *Progress in Human Geography*, 1–18. https://doi.org/10.1177/0309132519830511

Feinstein, B. D., & Allen, A. (2011). Community Benefits Agreements with Transit Agencies-Neighborhood Change Along Boston's Rail Lines and a Legal Strategy for Addressing Gentrification. *Transportatoin Law Journal*, *38*, 85.

Gallin, J. (2008). The Long-Run Relationship Between House Prices and Rents. *Real Estate Economics*, *36*(4), 635–658. https://doi.org/10.1111/j.1540-6229.2008.00225.x

Grier, G. (The G. P., & Grier, E. (The G. P. (1978). Urban Displacement: A Reconnaissance. *U.S. Department of Housing and Urban Development*. https://www.huduser.gov/portal/searchbiblio/Bibliography/50456

Grube-Cavers, A., & Patterson, Z. (2015). Urban rapid rail transit and gentrification in Canadian urban centres: A survival analysis approach. *Urban Studies*, *52*(1), 178–194. https://doi.org/10.1177/0042098014524287

Heilmann, K. (2018). Transit access and neighborhood segregation. Evidence from the Dallas light rail system. *Regional Science and Urban Economics*, *73*, 237–250. https://doi.org/10.1016/j.regsciurbeco.2018.10.007

Hess, C. L. (2018). Light-Rail Investment in Seattle: Gentrification Pressures and Trends in Neighborhood Ethnoracial Composition. In *Urban Affairs Review*. https://doi.org/10.1177/1078087418758959

Kahn, M. E. (2007). Gentrification trends in new transit-oriented communities: Evidence from 14 cities that expanded and built rail transit systems. *Real Estate Economics*, *35*(2), 155–182. https://doi.org/10.1111/j.1540-6229.2007.00186.x

Kavanagh, L., Lee, D., & Pryce, G. (2016). Is Poverty Decentralizing? Quantifying Uncertainty in the Decentralization of Urban Poverty. *Annals of the American Association of Geographers*, *106*(6), 1286–1298. https://doi.org/10.1080/24694452.2016.1213156

Luckey, K. S., Marshall, W. E., Durso, C., & Atkinson-Palombo, C. (2018). Residential preferences, transit accessibility and social equity: insights from the Denver region. *Journal of Urbanism*, *11*(2), 149–174. https://doi.org/10.1080/17549175.2017.1422531

Mckenzie, B. S. (2013). Neighborhood Access to Transit by Race, Ethnicity, and Poverty in Portland, OR. *City and Community*, *12*(2), 134–155. https://doi.org/10.1111/cico.12022

Nilsson, I., & Delmelle, E. (2018). Transit investments and neighborhood change: On the likelihood of change. *Journal of Transport Geography*, *66*, 167–179. https://doi.org/10.1016/j.jtrangeo.2017.12.001

Rayle, L. (2015). Investigating the Connection Between Transit-Oriented Development and Displacement: Four Hypotheses. *Housing Policy Debate*, *25*(3), 531–548. https://doi.org/10.1080/10511482.2014.951674

Revington, N. (2015). Gentrification, transit, and land use: Moving beyond neoclassical theory. *Geography Compass*, *9*(3), 152–163. https://doi.org/10.1111/gec3.12203

Sandoval, G. F. (2018). Planning the Barrio: Ethnic Identity and Struggles over Transit-Oriented, Development-Induced Gentrification. *Journal of Planning Education and Research*, 1–15. https://doi.org/10.1177/0739456X18793714

Scally, C. P., & Gonzalez, D. (2018). *Renters are more likely than homeowners to struggle with paying for basic needs | Urban Institute*. Urban Wire: Housing and Housing Finance. https://www.urban.org/urban-wire/renters-are-more-likely-homeowners-struggle-paying-basic-needs

Zuk, M., Bierbaum, A. H., Chapple, K., Gorska, K., & Loukaitou-Sideris, A. (2018). Gentrification, Displacement, and the Role of Public Investment. *Journal of Planning Literature*, *33*(1), 31–44. https://doi.org/10.1177/0885412217716439

APPENDIX A

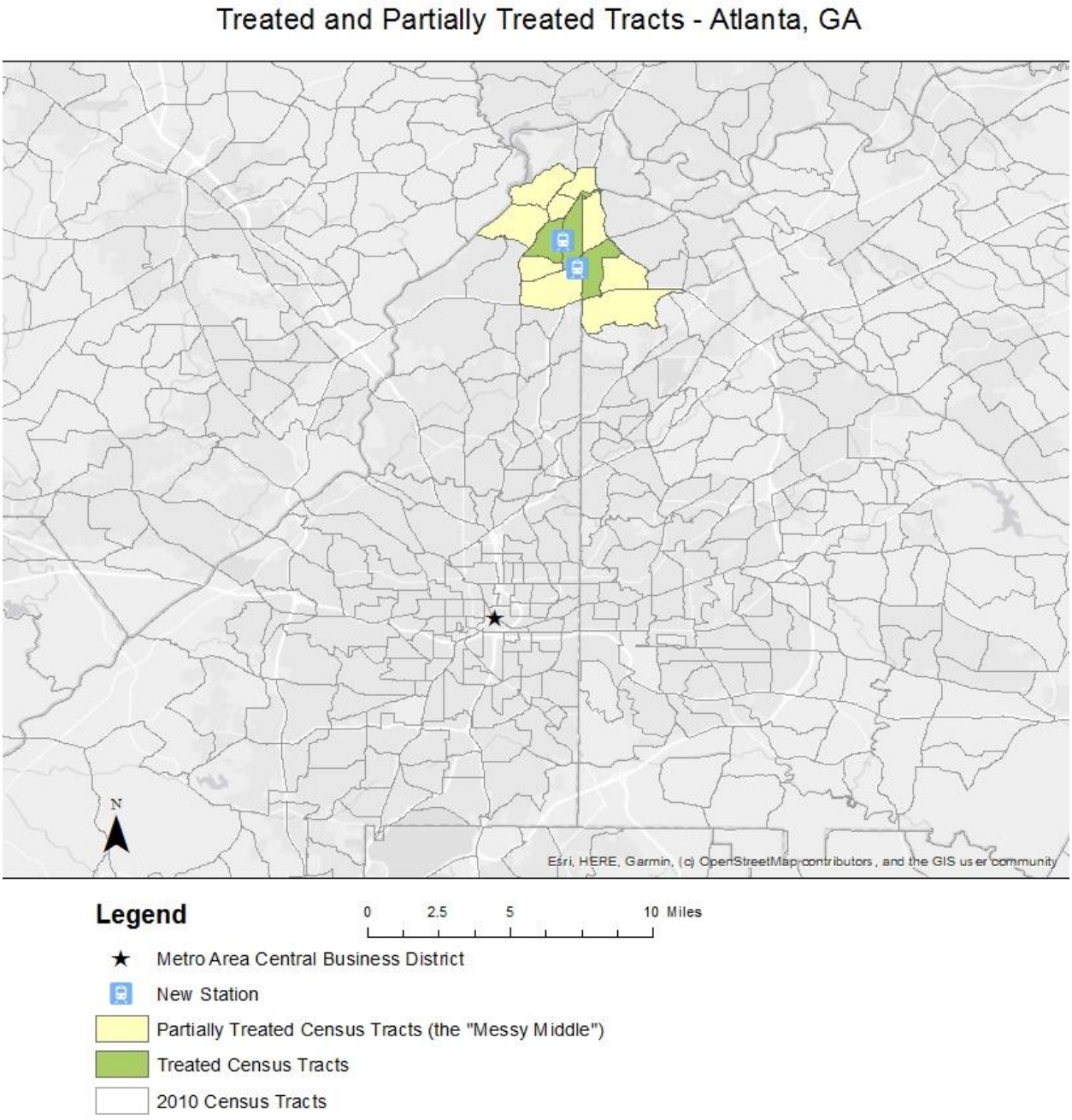

Figure 4: Treated and Partially Treated Tracts - Atlanta, GA

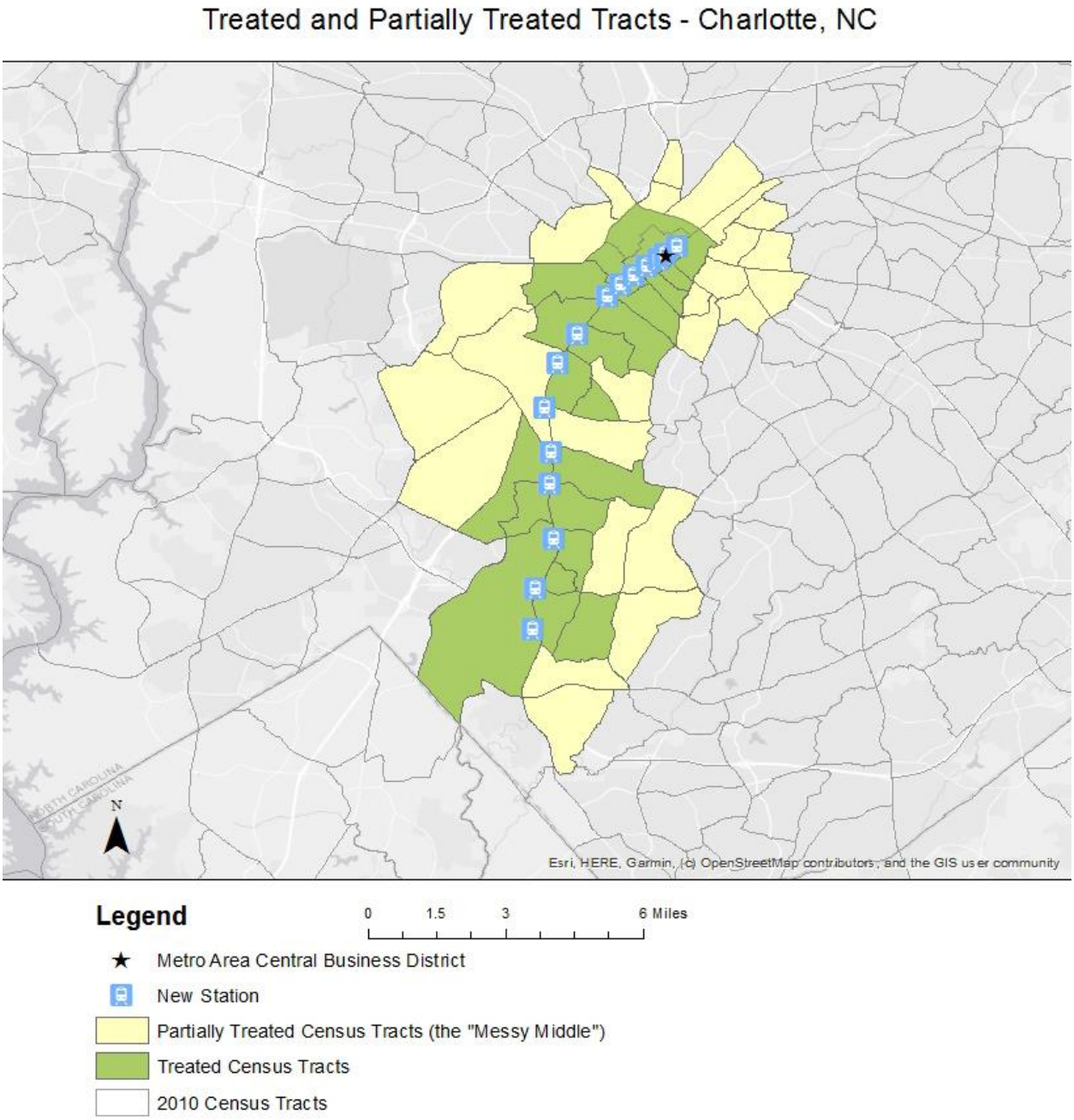

Figure 5: Treated and Partially Treated Tracts - Charlotte, NC

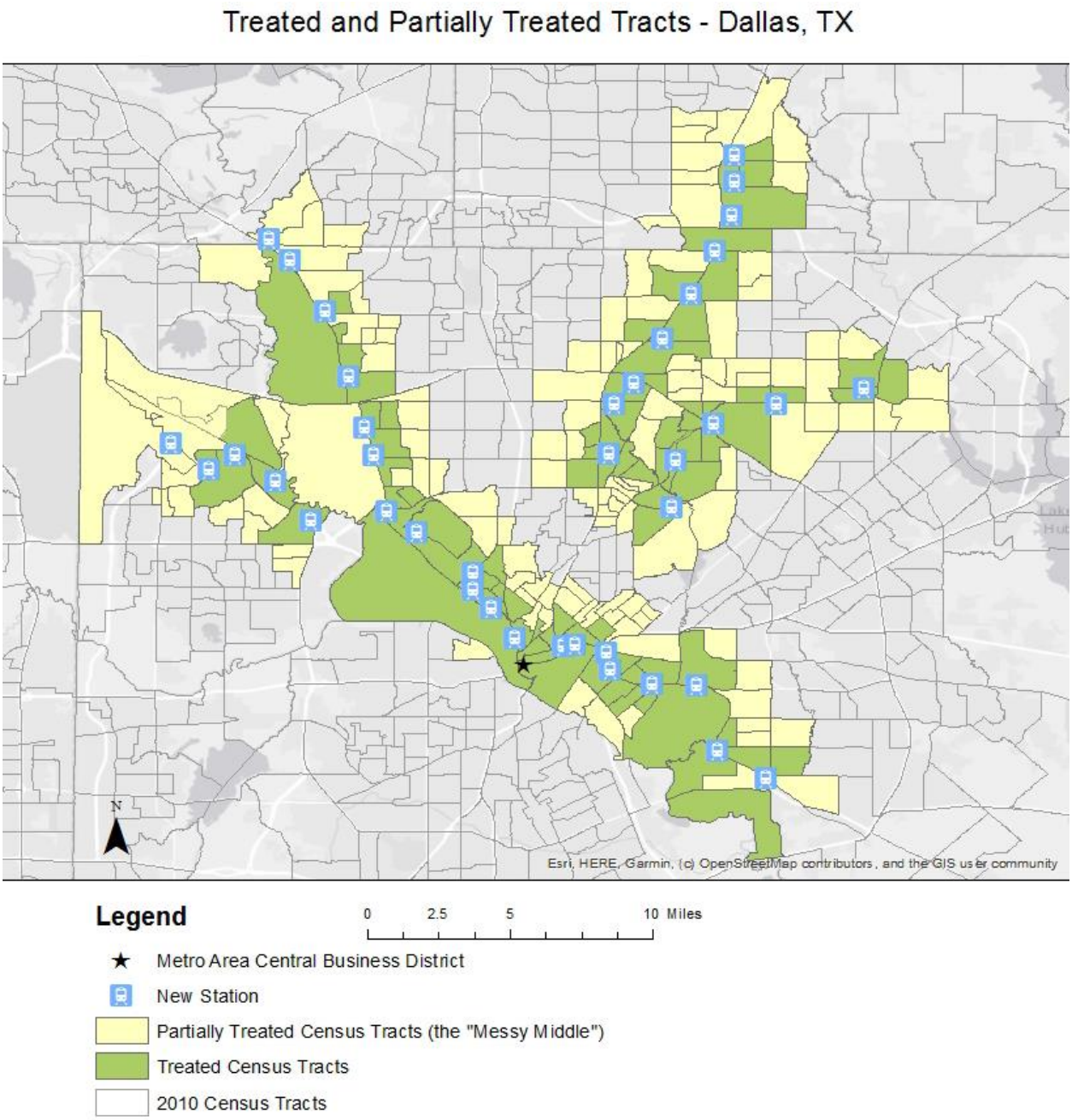

Figure 6: Treated and Partially Treated Tracts - Dallas, TX

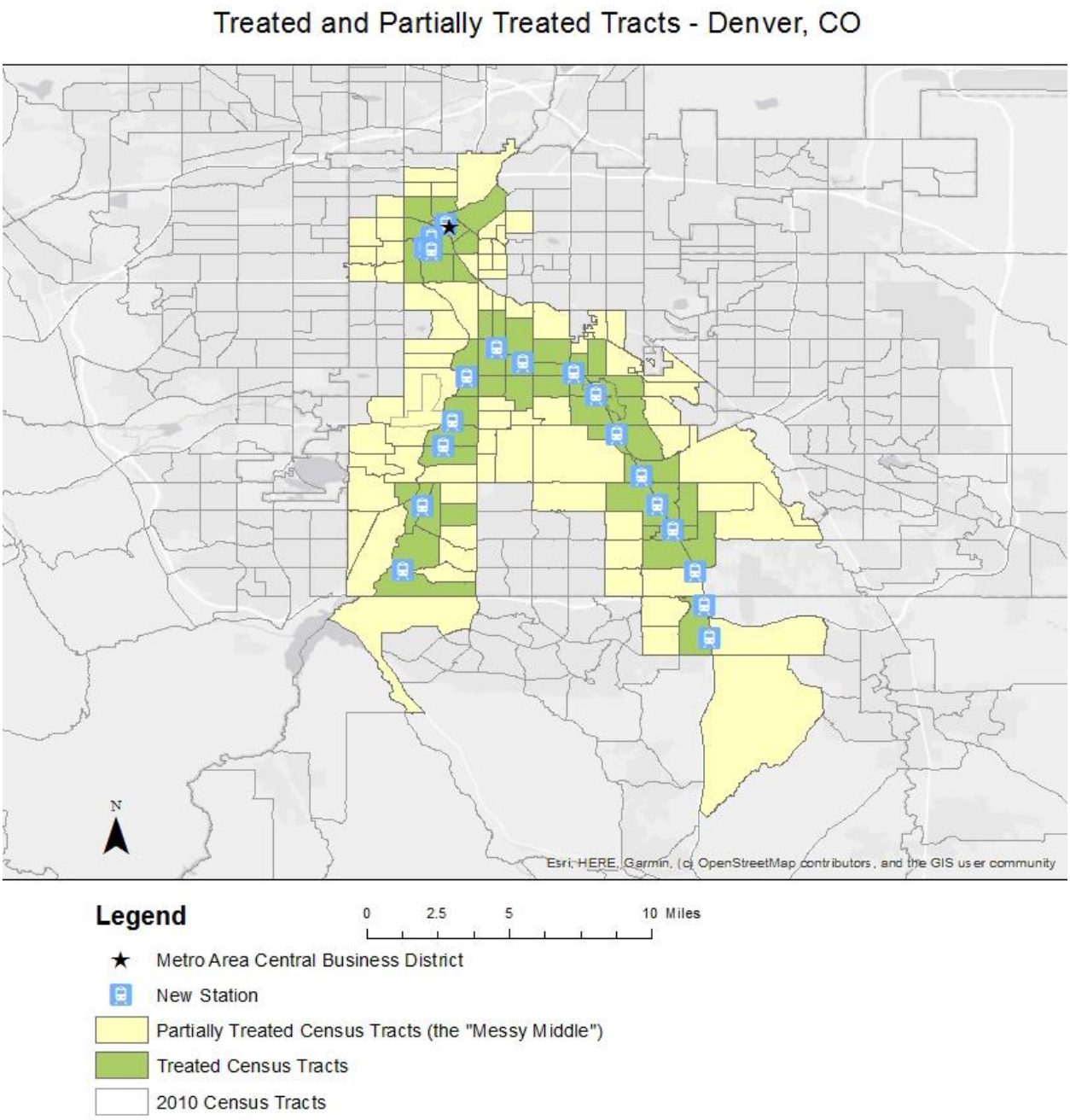

Figure 7: Treated and Partially Treated Tracts - Denver, CO

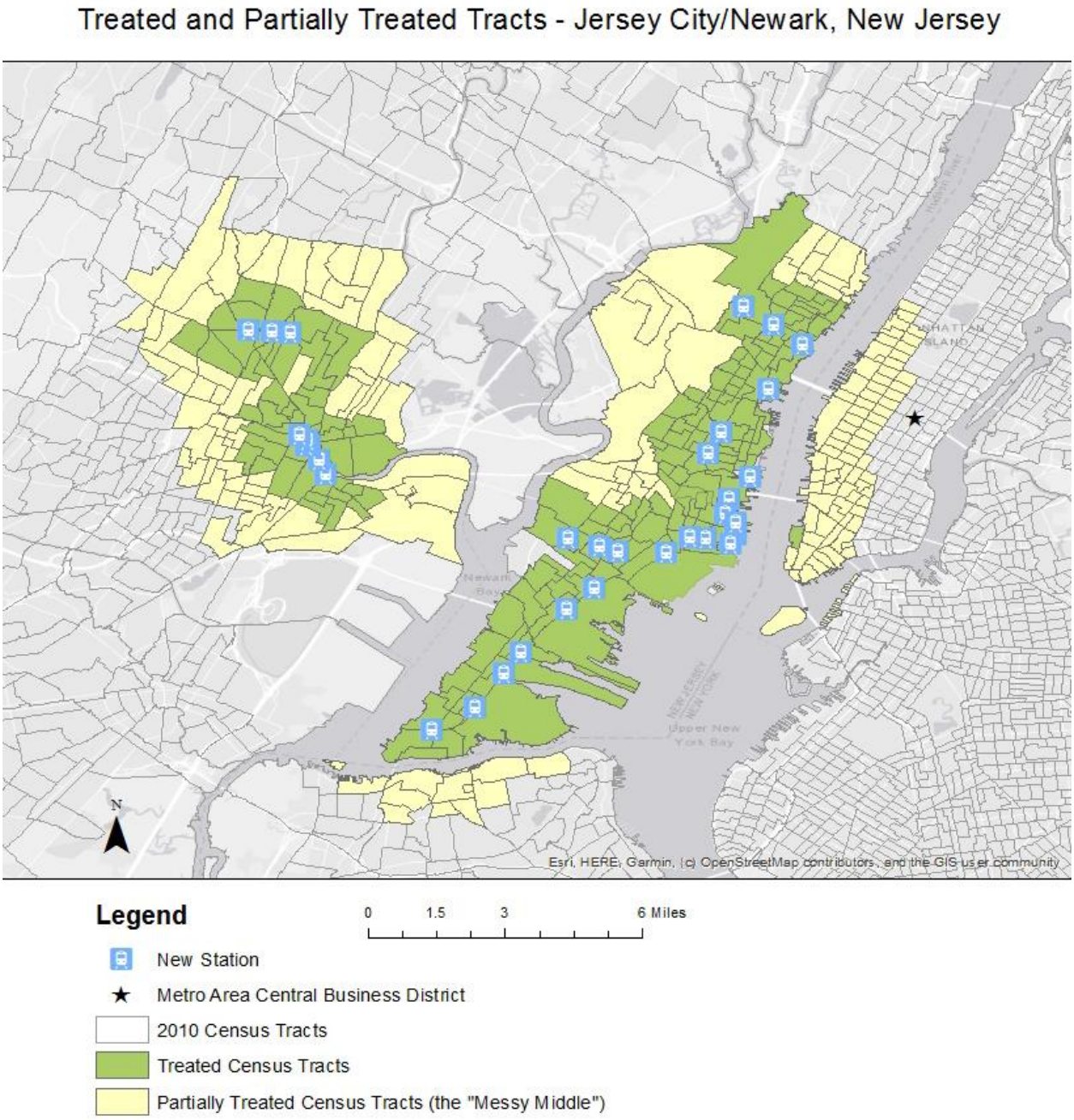

Figure 8: Treated and Partially Treated Tracts - Jersey City/Newark, NJ

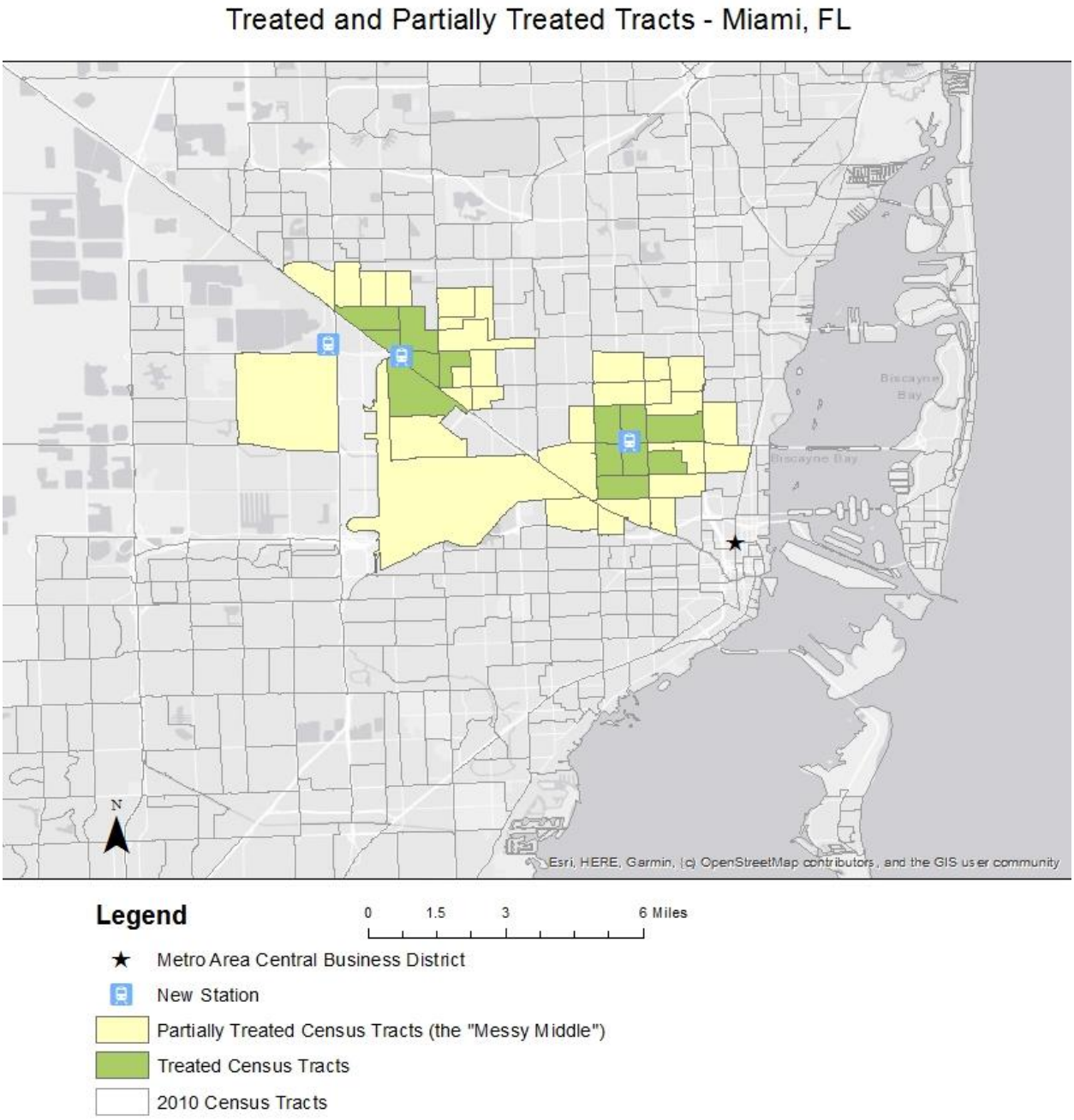

Figure 9: Treated and Partially Treated Tracts - Miami, FL

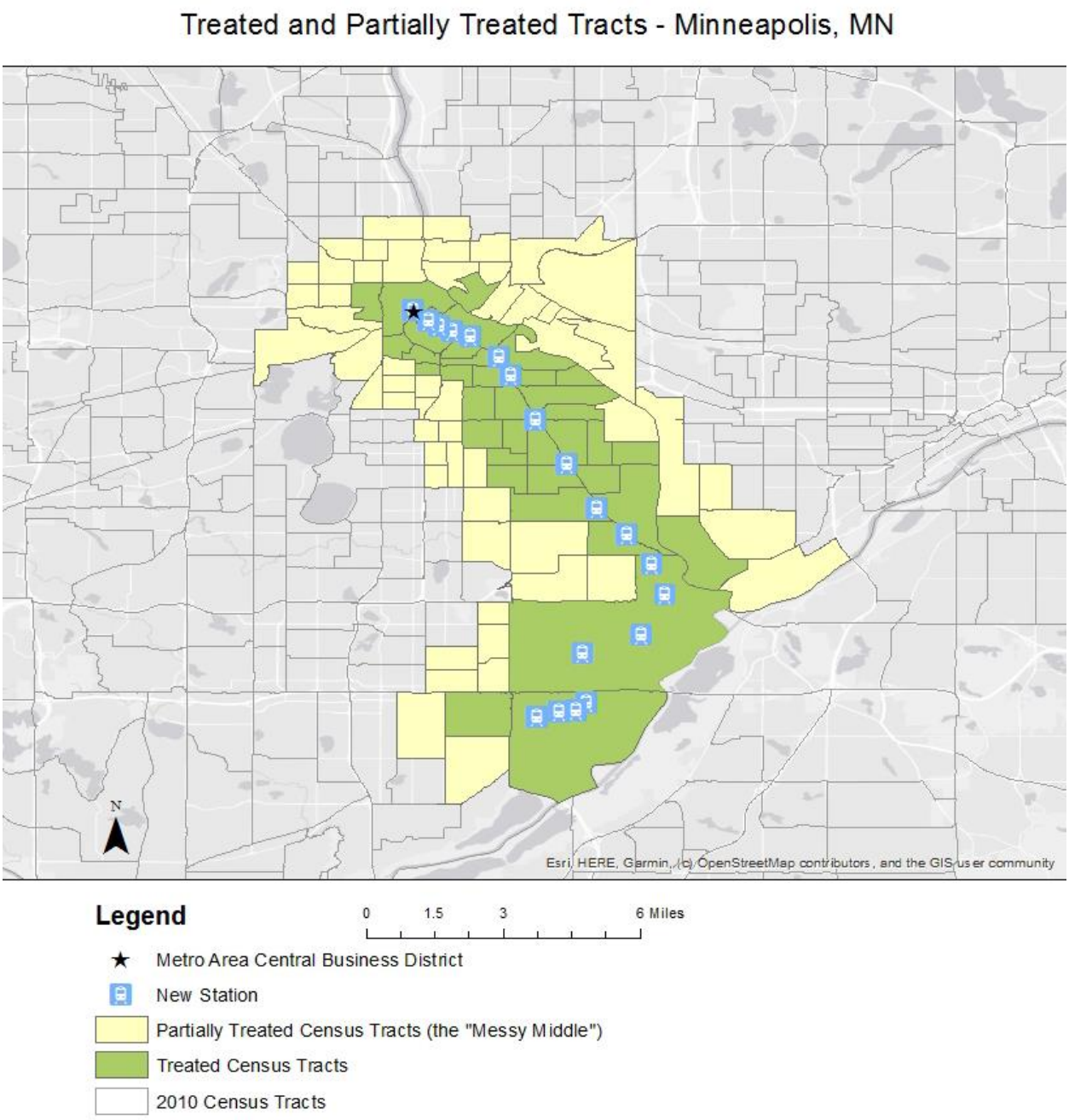

Figure 10: Treated and Partially Treated Tracts - Minneapolis, MN

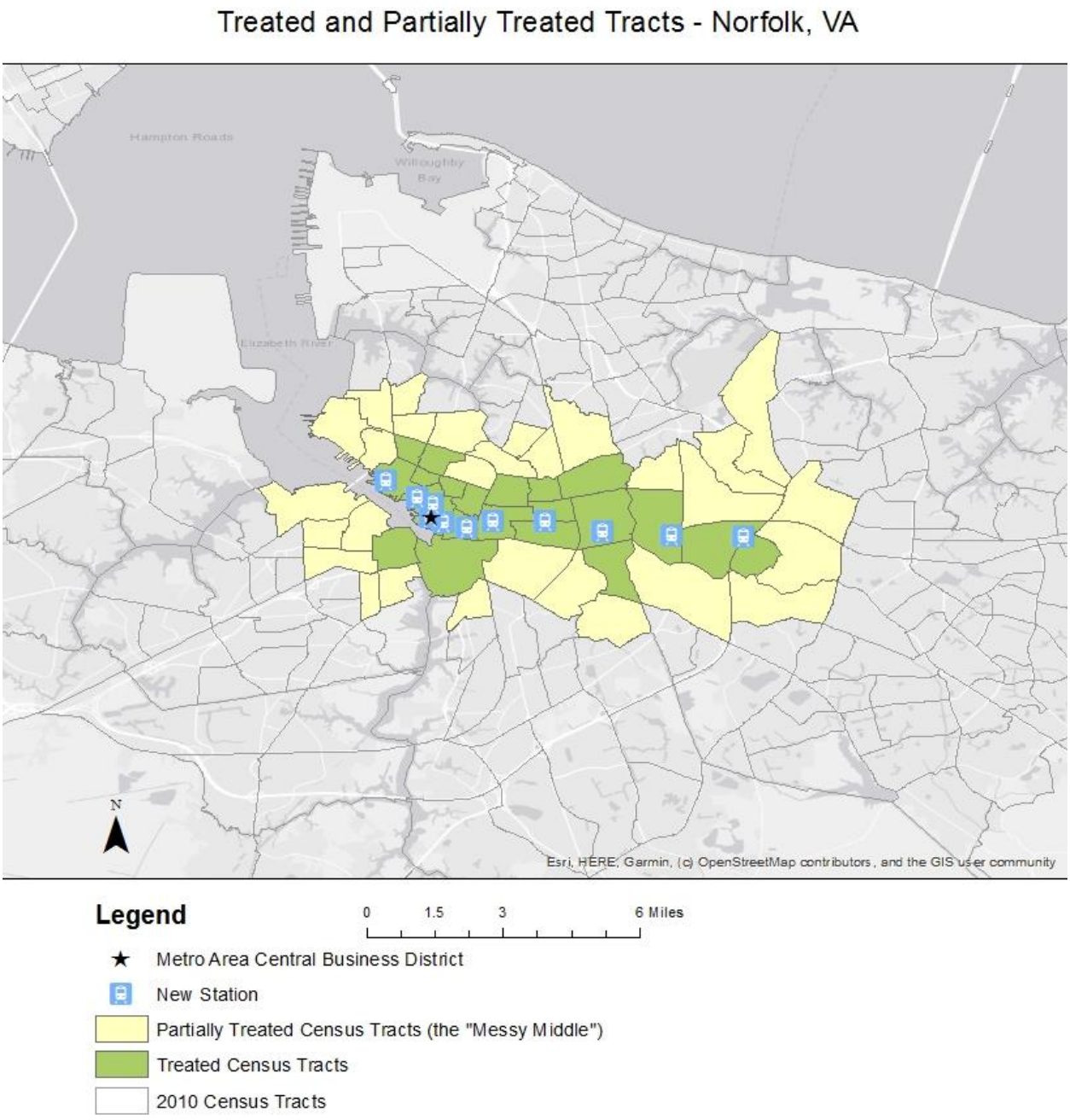

Figure 11: Treated and Partially Treated Tracts - Norfolk, VA

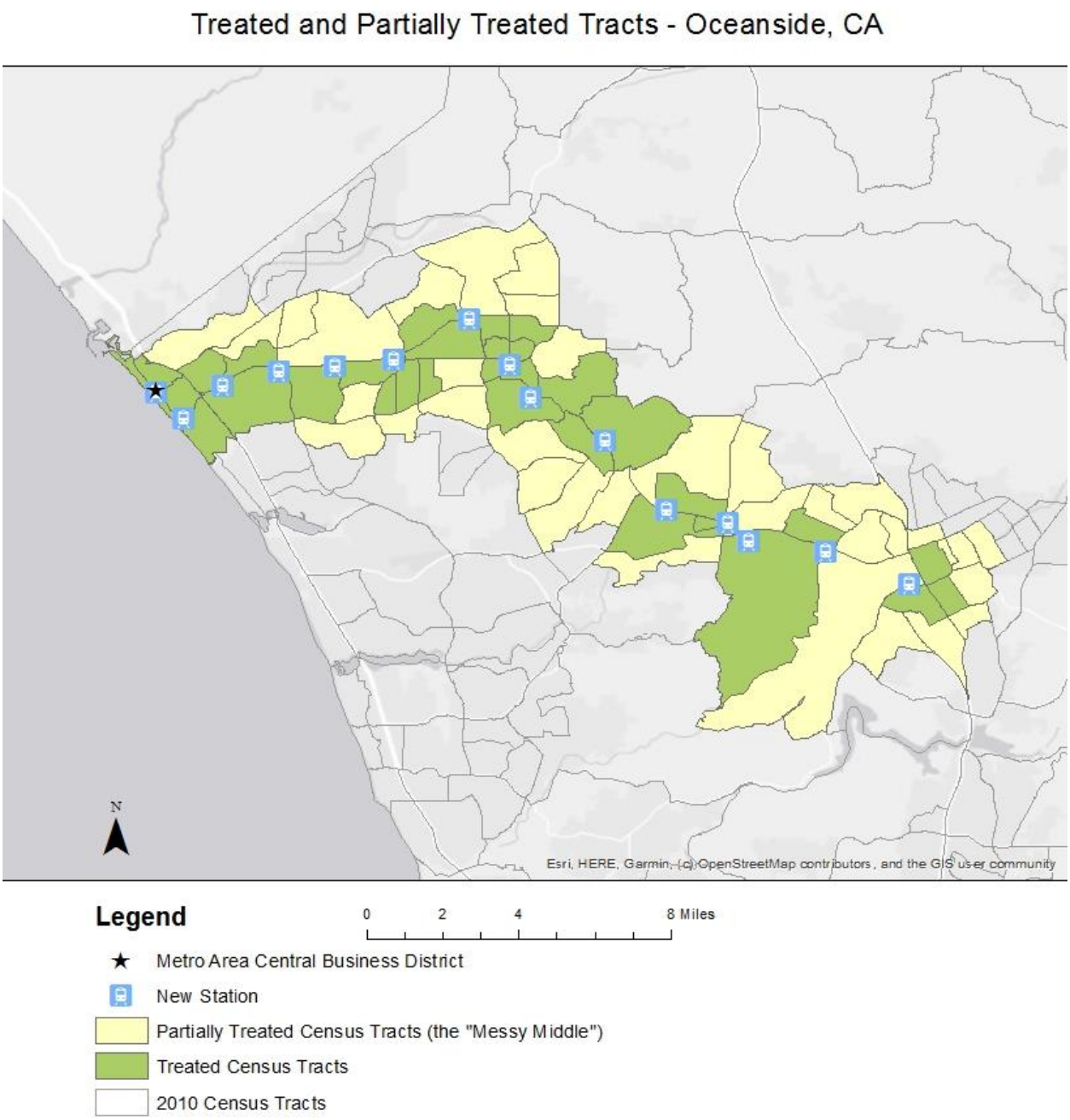

Figure 12: Treated and Partially Treated Tracts - Oceanside, CA

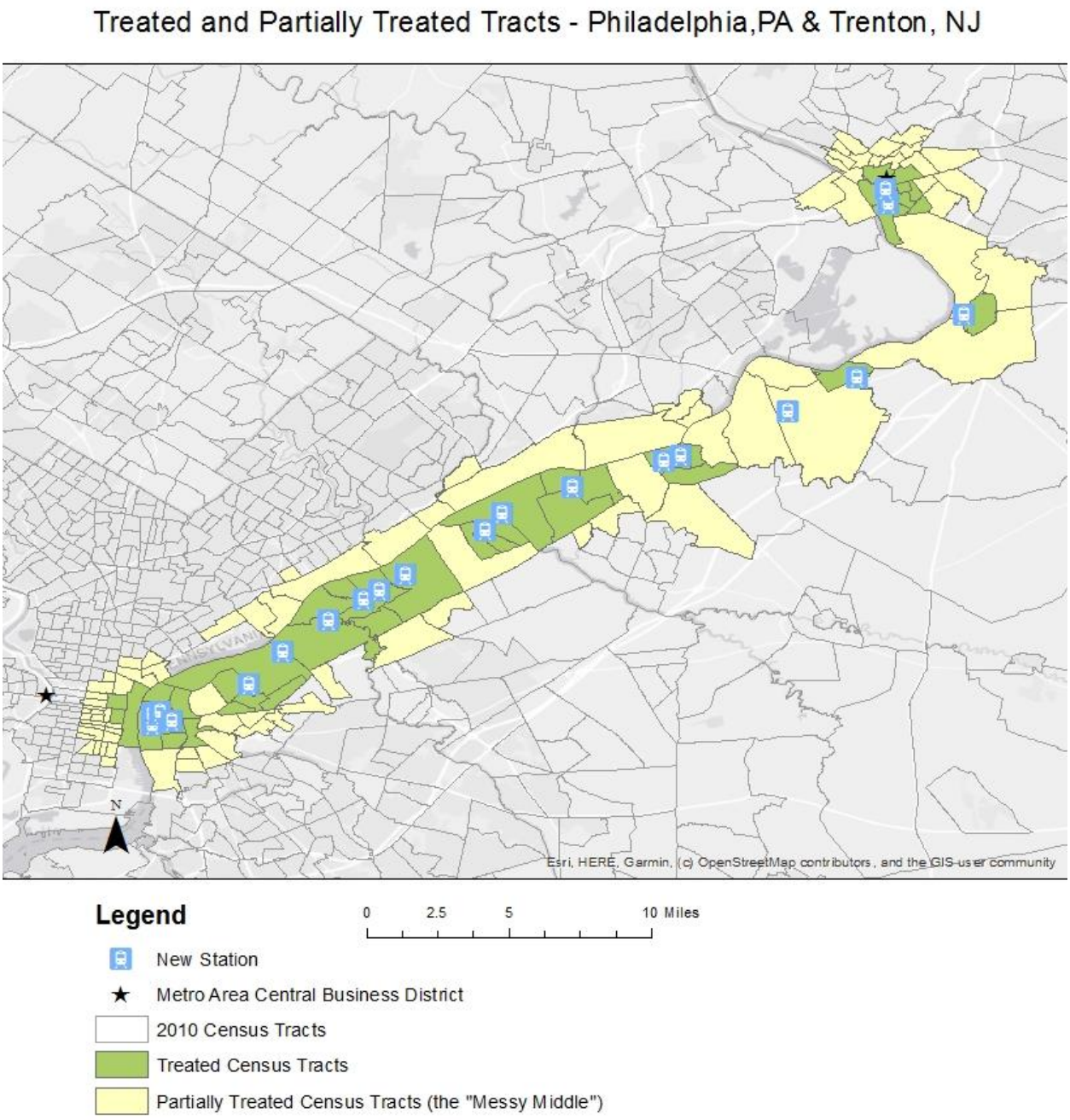

Figure 13: Treated and Partially Treated Tracts - Philadelphia, PA & Trenton, NJ

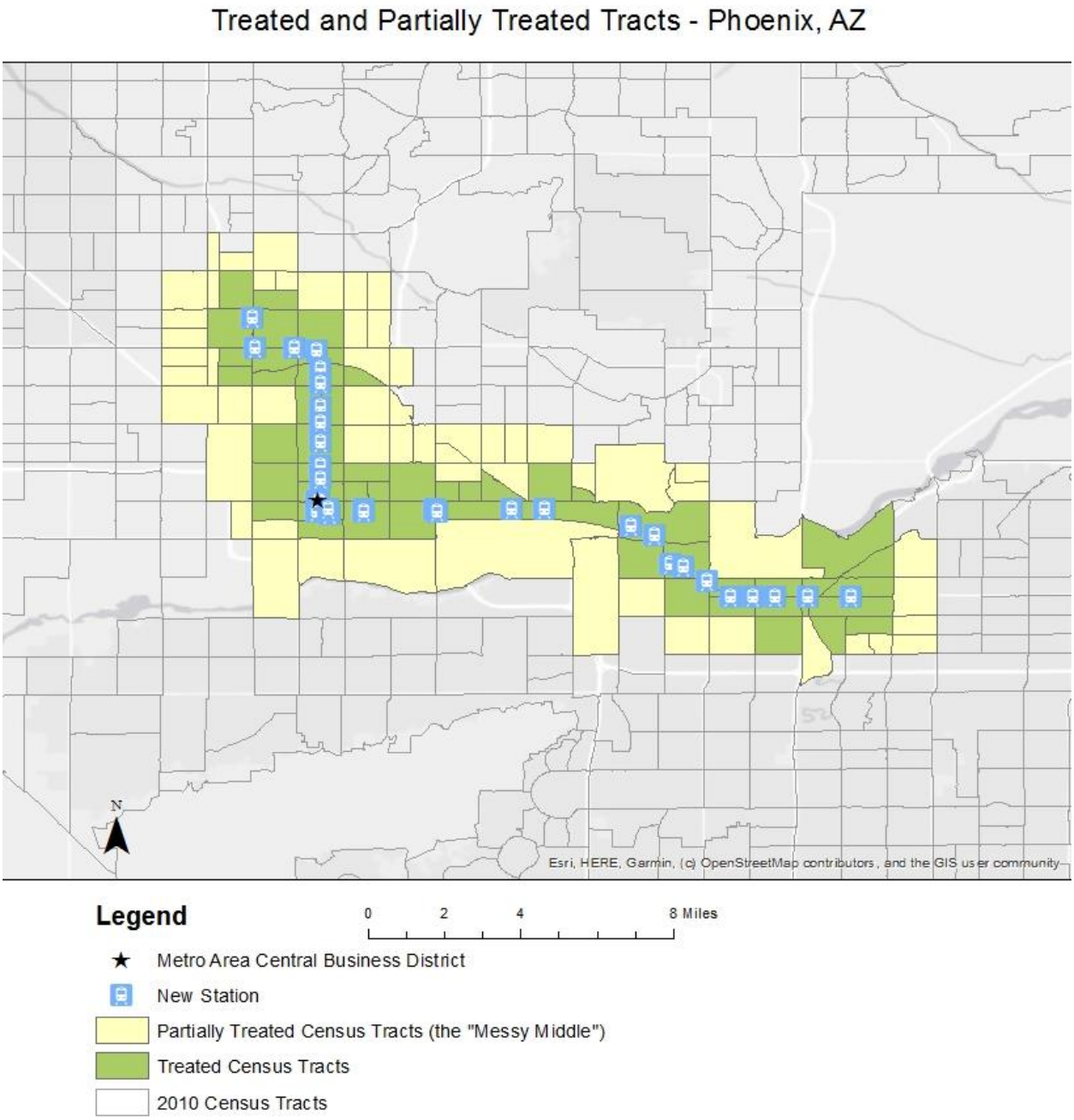

Figure 14: Treated and Partially Treated Tracts - Phoenix, AZ

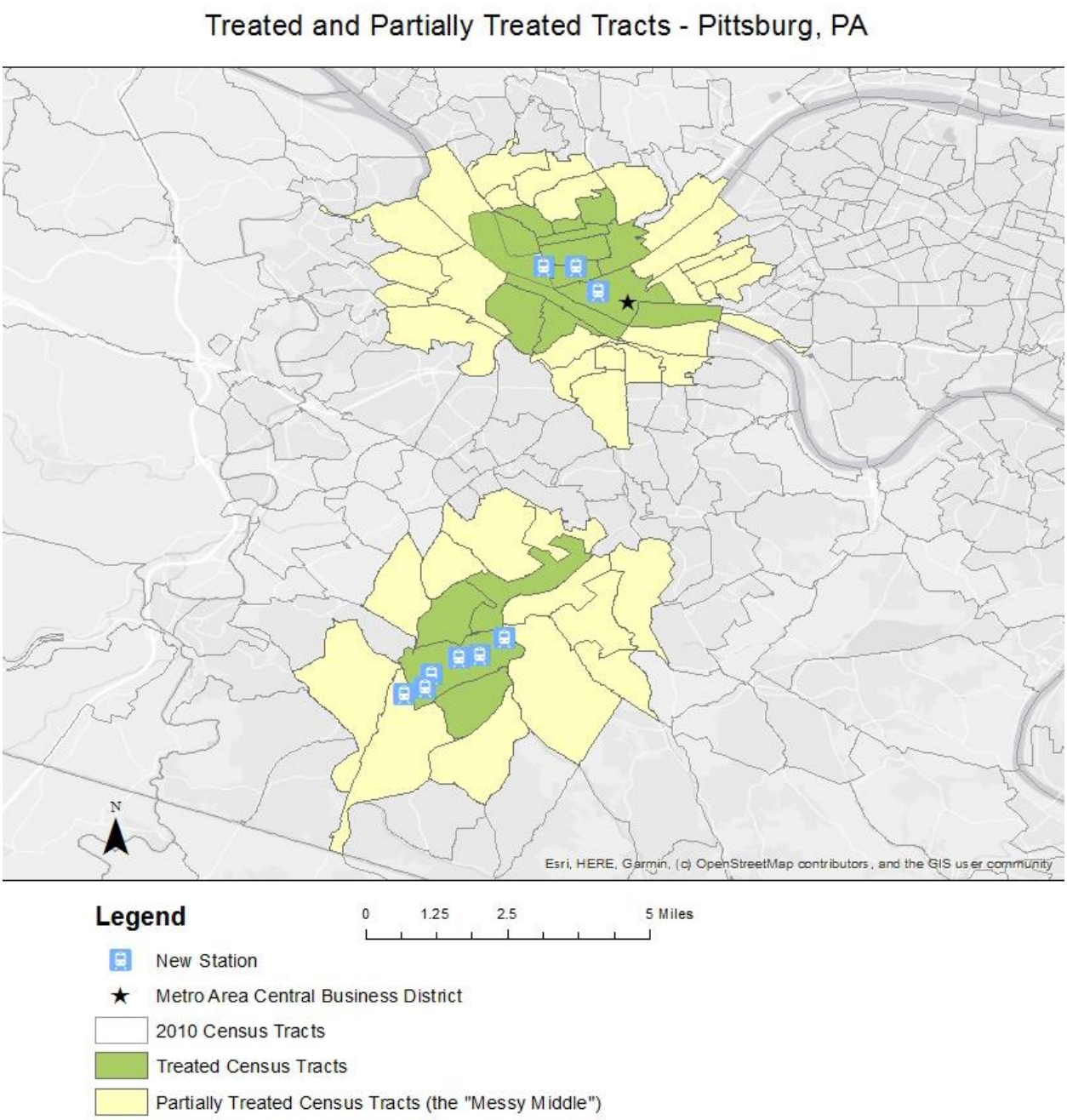

Figure 15: Treated and Partially Treated Tracts - Pittsburg, PA

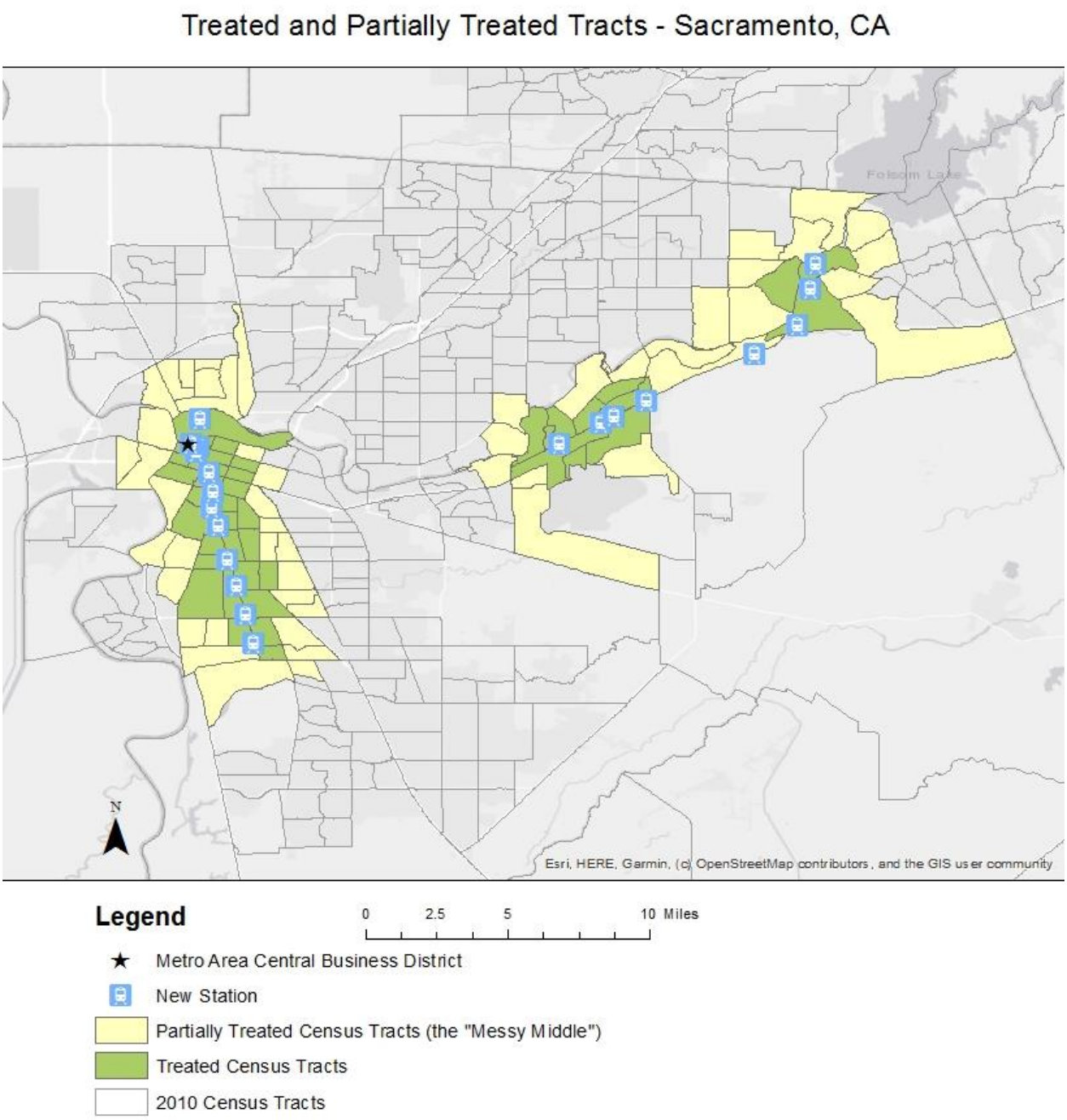

Figure 16: Treated and Partially Treated Tracts - Sacramento, CA

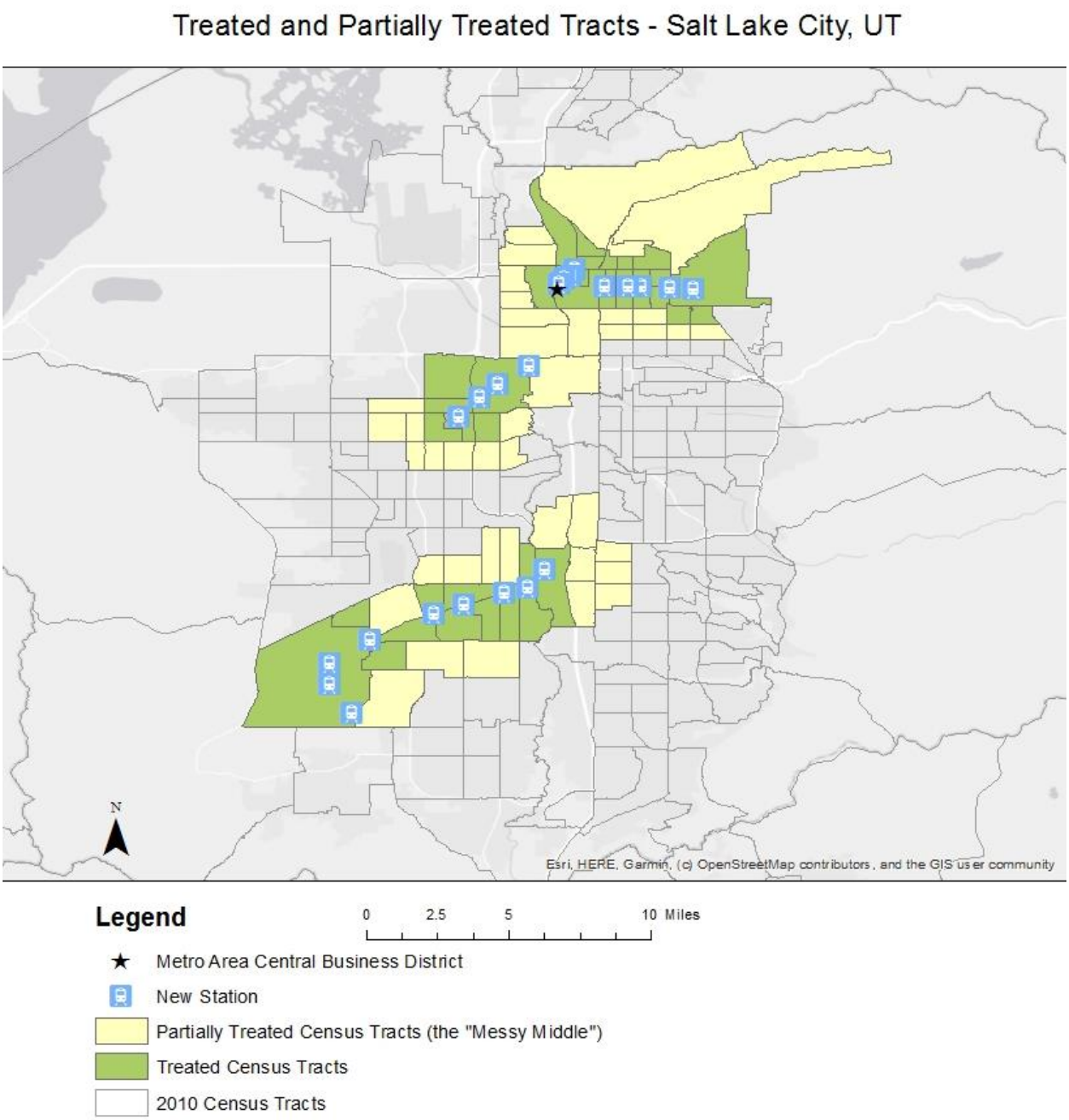

Figure 17: Treated and Partially Treated Tracts - Salt Lake City, UT

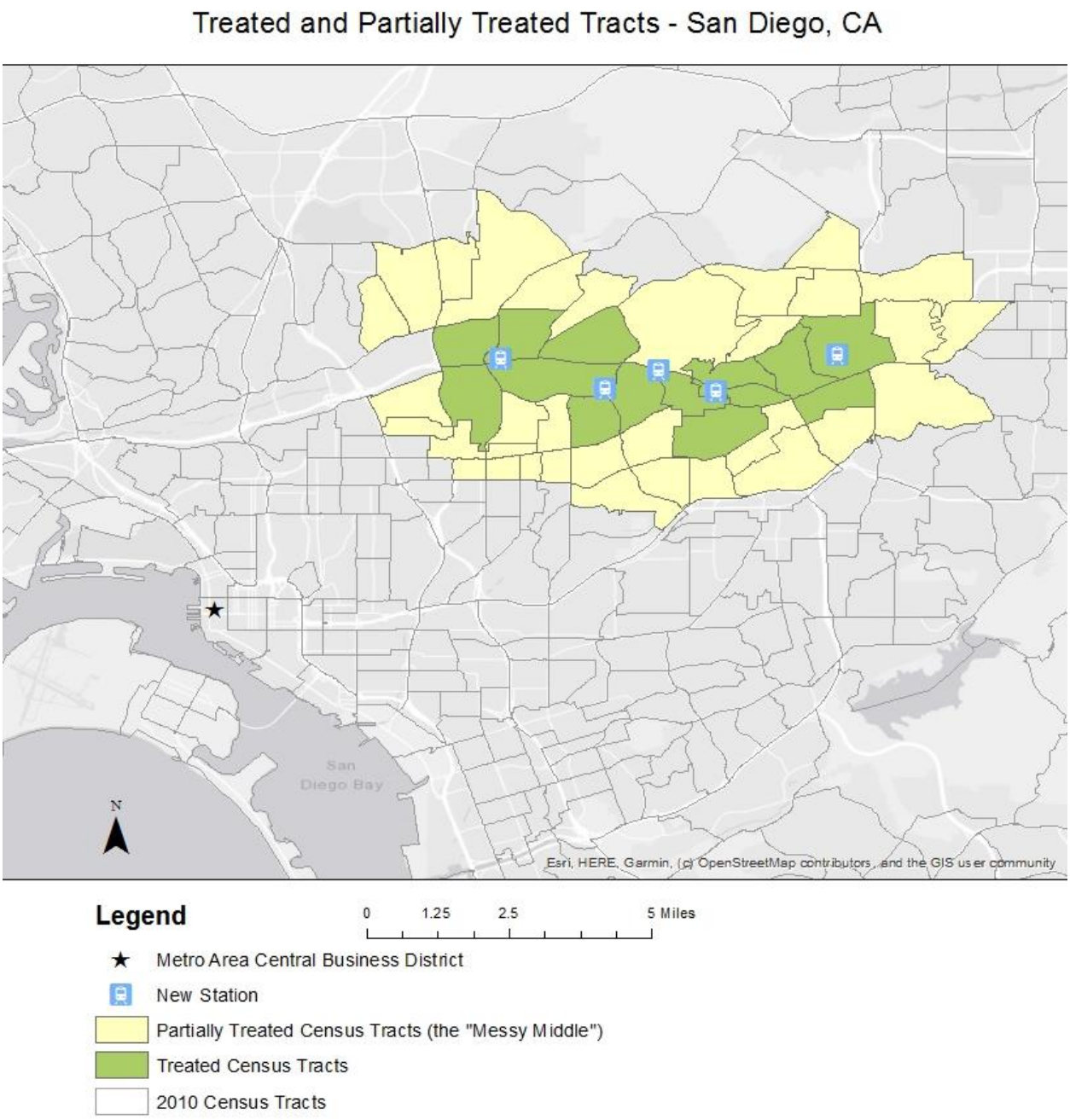

Figure 18: Treated and Partially Treated Tracts - San Diego, CA

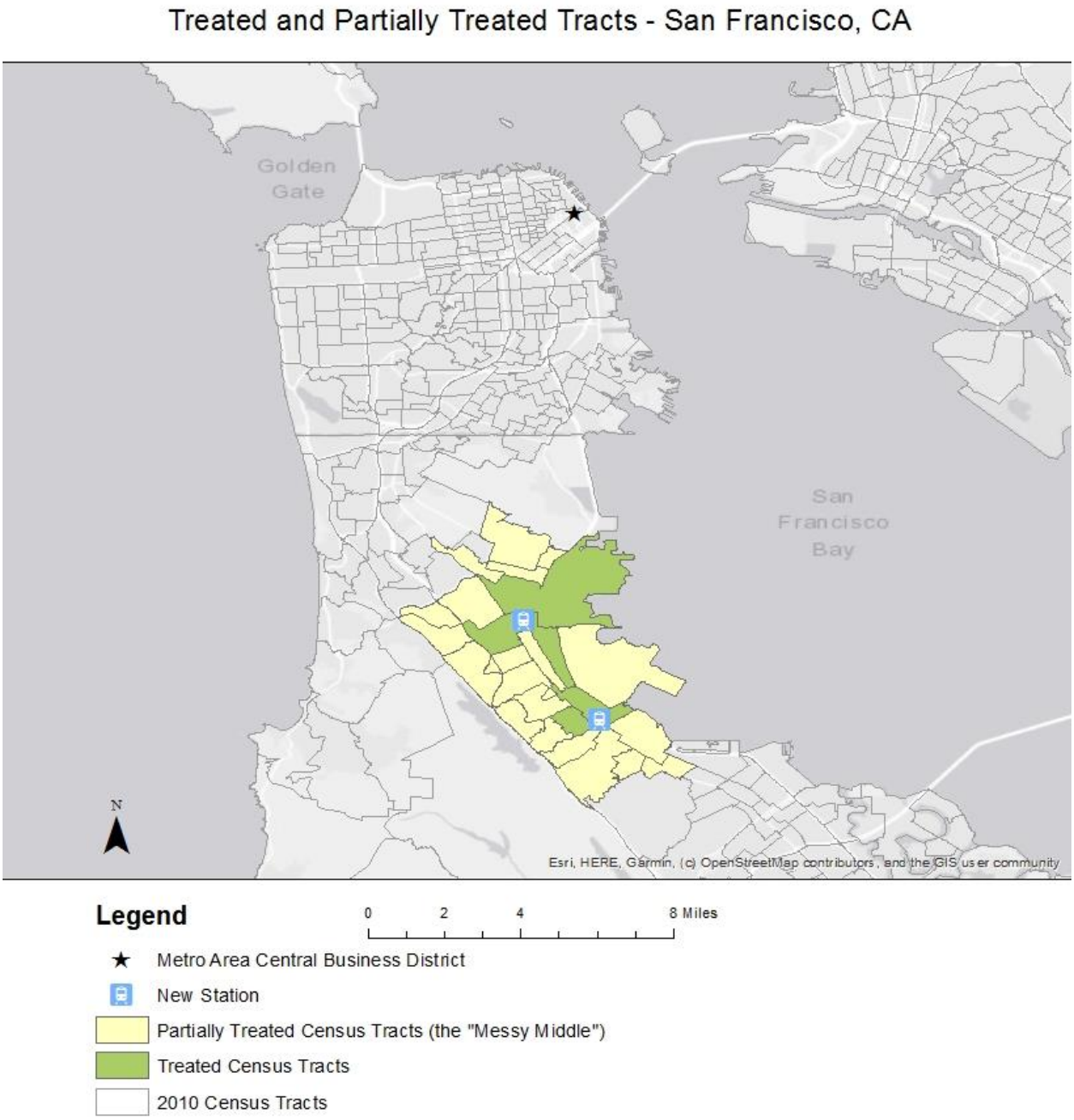

Figure 19: Treated and Partially Treated Tracts - San Francisco, CA

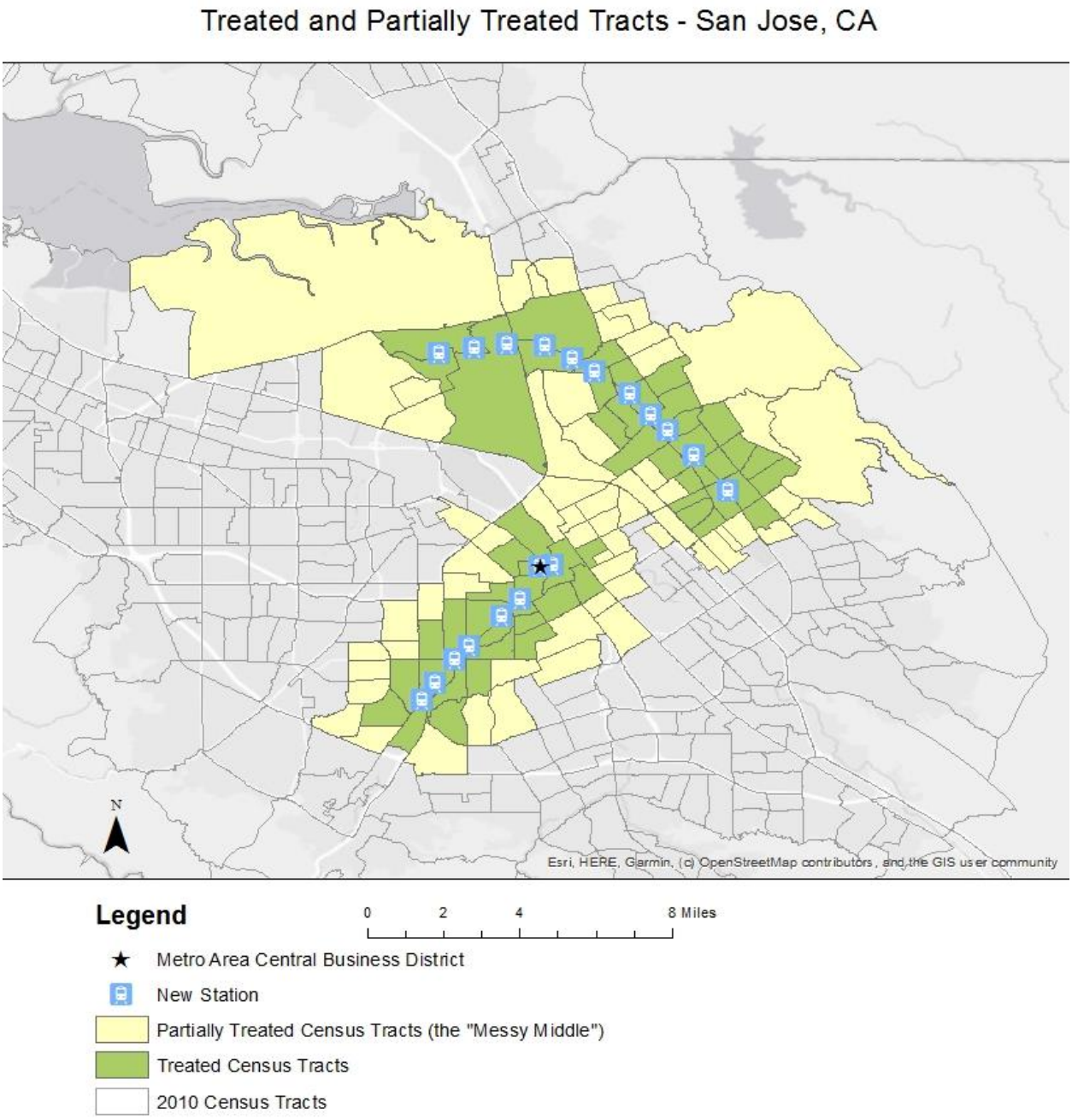

Figure 20: Treated and Partially Treated Tracts - San Jose, CA

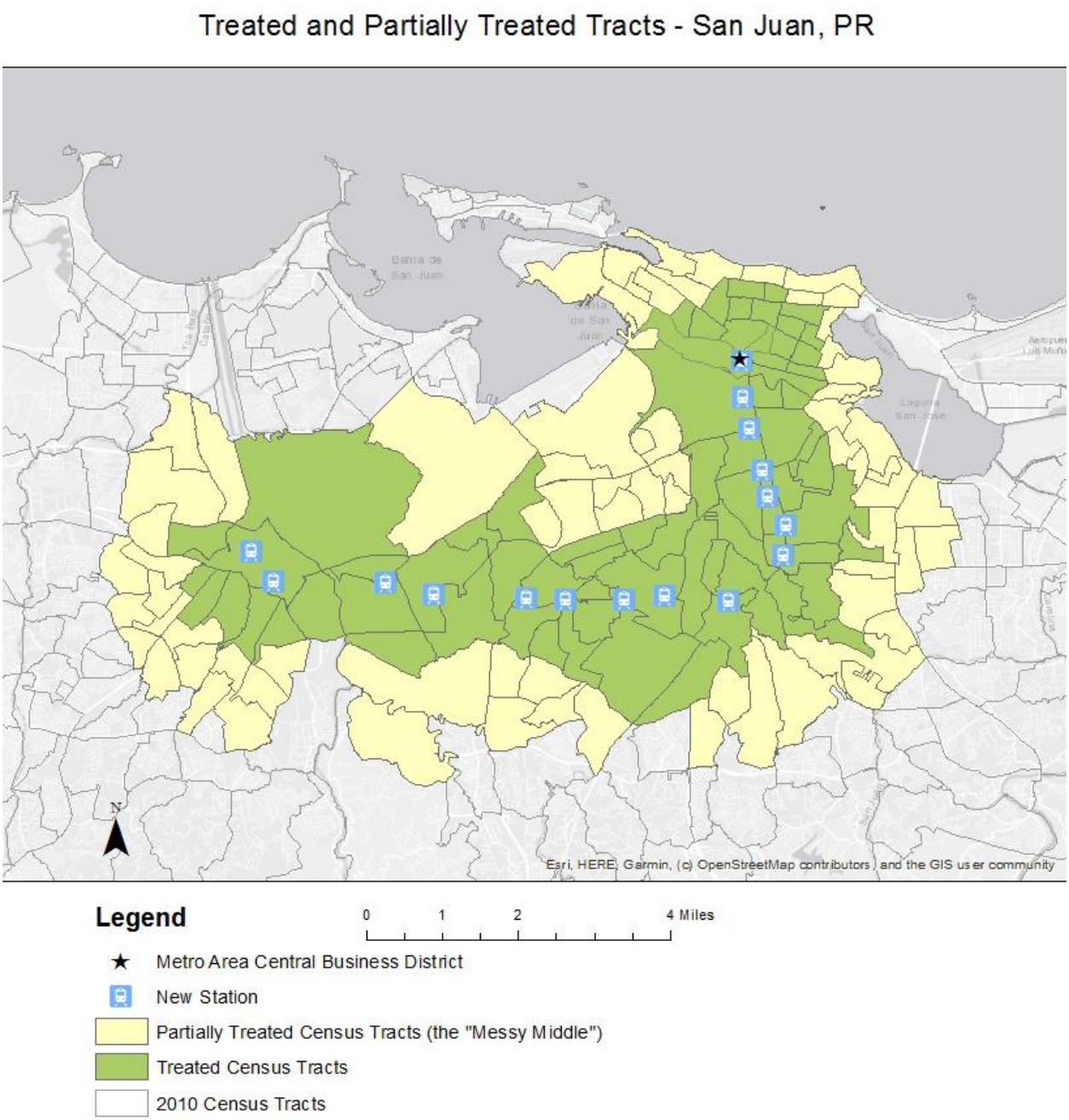

Figure 21: Treated and Partially Treated Tracts - San Juan, PR

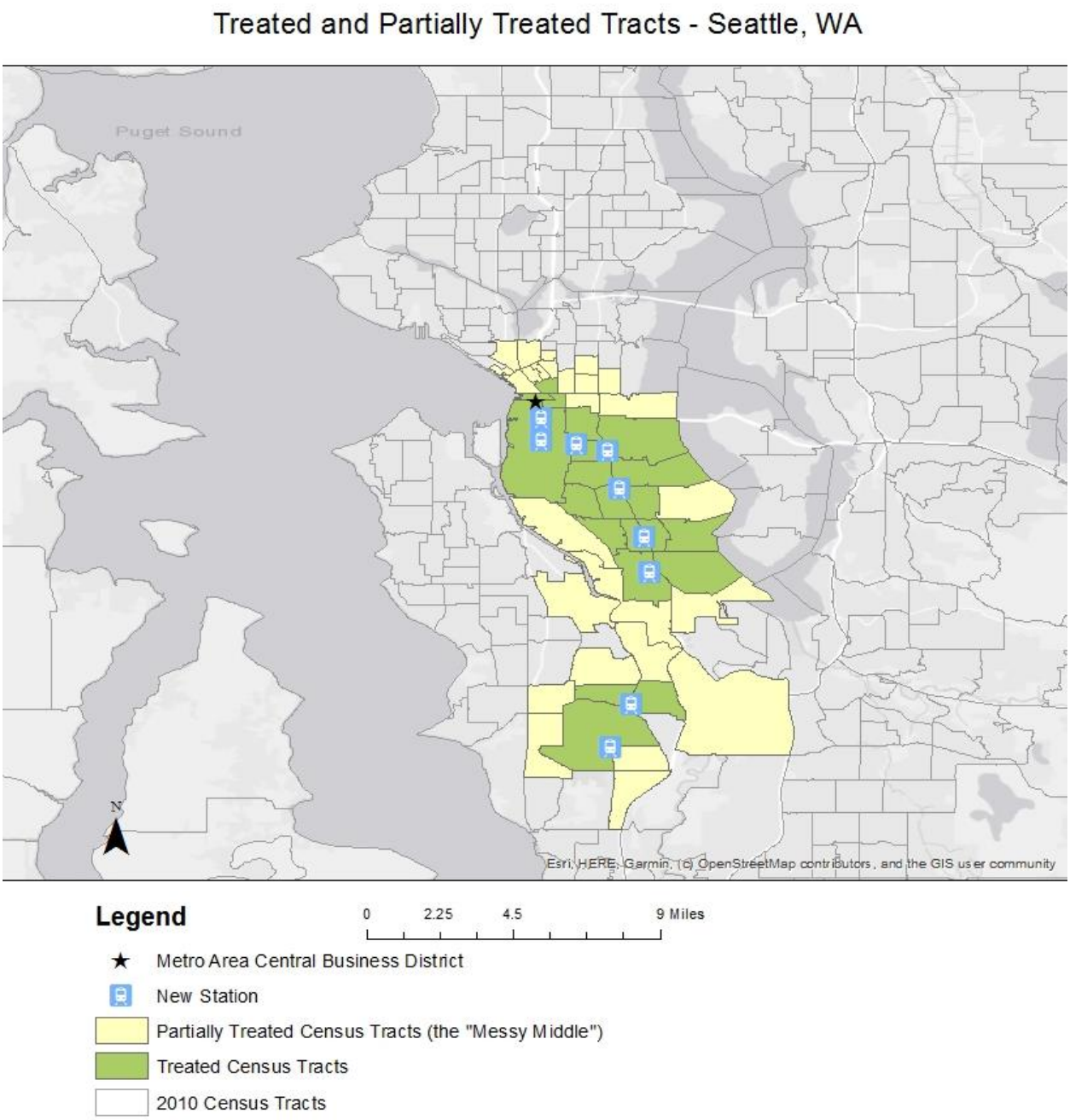

Figure 22: Treated and Partially Treated Tracts - Seattle, WA

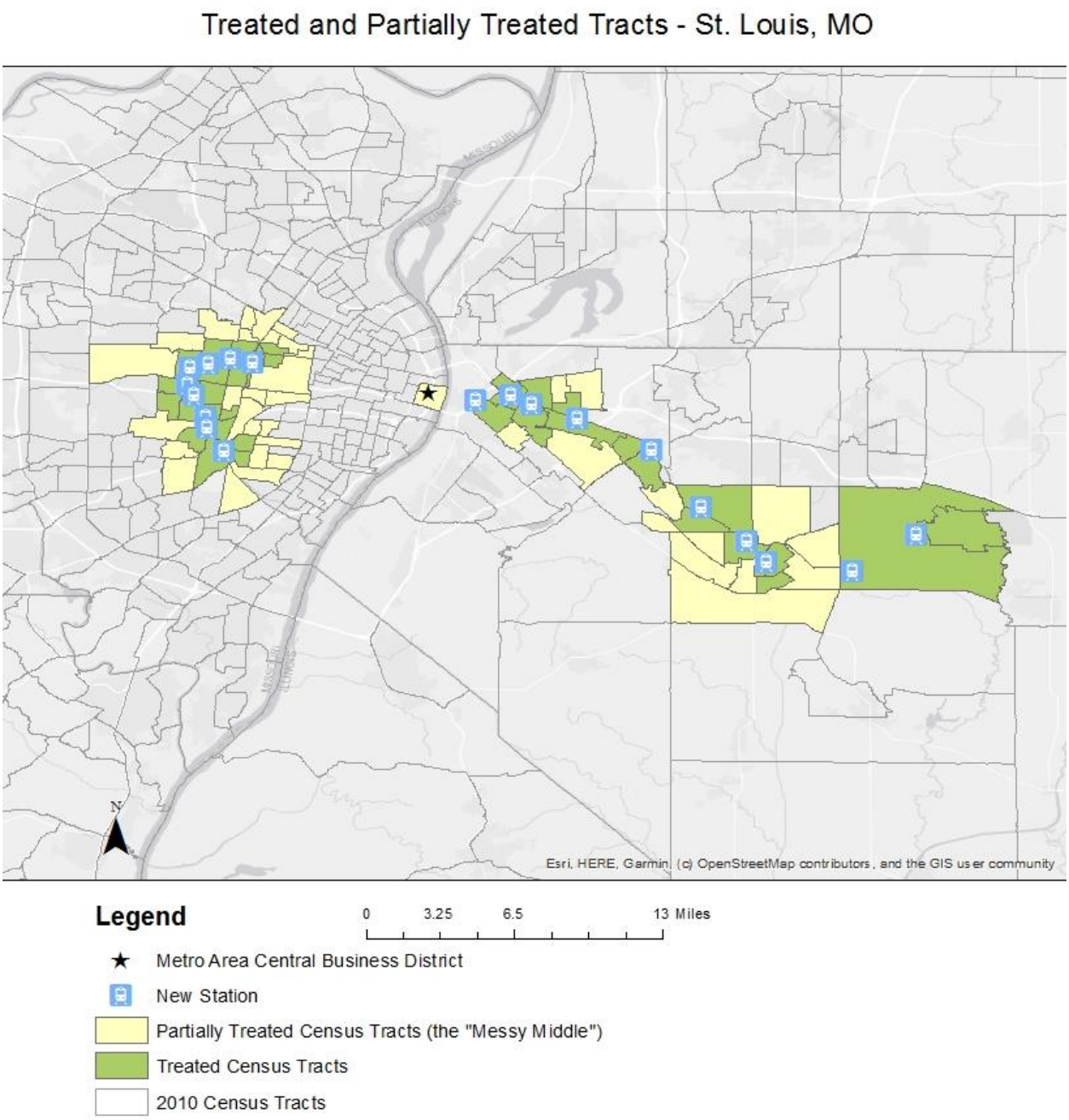

Figure 23: Treated and Partially Treated Tracts - St. Louis, MO

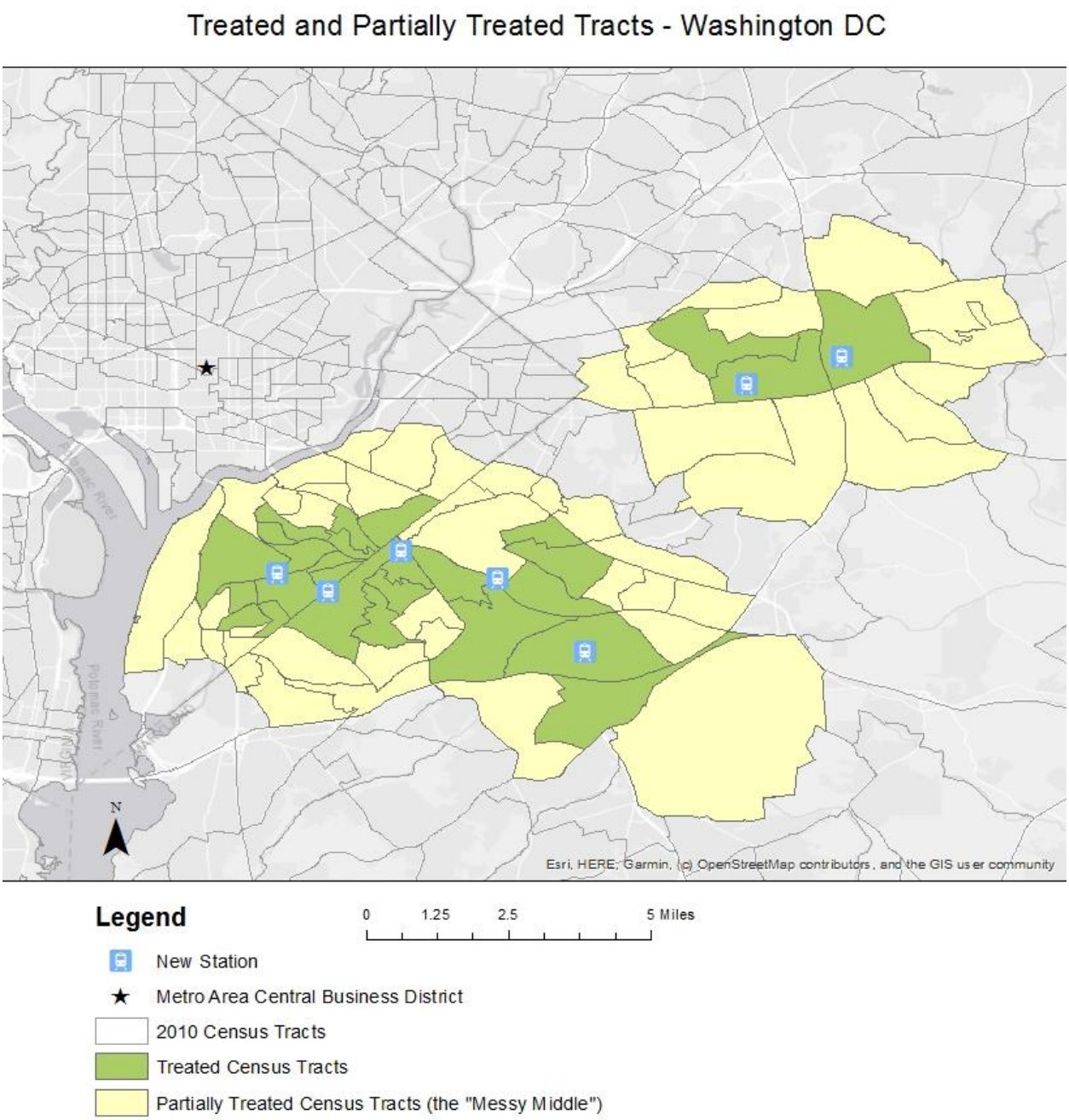

Figure 24: Treated and Partially Treated Tracts - Washington DC

APPENDIX B

Table 7: Establishing Common Support Regression

	(1) treated
prcntfemale	-0.503***
	(0.0223)
prcntnhlwhite	-0.0689***
	(0.00327)
prcnt65plus	0.138***
	(0.0122)
prcntbachelors	0.0373***
	(0.00563)
prcntpublictransit	-0.0393***
	(0.00552)
prcntrentals	0.0632***
	(0.00378)
logmeanrent	0.0117***
	(0.00250)
distance2cbd	-0.000868***
	(0.0000635)
_cons	0.212***
	(0.0197)
N	58287
adj. R^2	0.037

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

Table 8: Model #1 - Simple Regressions

	(1) logmeanrent	(2) logmeanrent	(3) prcntbachelors	(4) prcntbachelors	(5) prcntnhlwhite	(6) prcntnhlwhite	(7) prcnt65plus	(8) prcnt65plus
treated	0.0481***	0.0543***	0.0450***	0.0459***	0.0530***	0.0550***	-0.00863***	-0.00489*
	(0.010)	(0.011)	(0.005)	(0.006)	(0.006)	(0.006)	(0.002)	(0.002)
_lyear_2017	0.508***	0.503***	0.0572***	0.0566***	-0.0840***	-0.0850***	0.0179***	0.0142***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
_cons	6.409***	6.444***	0.253***	0.263***	0.594***	0.529***	0.118***	0.116***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
N	118513	76308	119168	76539	119169	76539	119169	76539
adj. R^2	0.881	0.885	0.912	0.910	0.936	0.926	0.787	0.795

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

Table 9: Second Model – Regression with the Introduction of Controls

	(1) logmeanrent	(2) logmeanrent	(3) prcntbachelors	(4) prcntbachelors	(5) prcntnhlwhite	(6) prcntnhlwhite	(7) prcnt65plus	(8) prcnt65plus
treated	-0.00307	0.00187	0.0232***	0.0222***	0.0313***	0.0288***	-0.0107***	-0.00597**
	(0.008)	(0.009)	(0.004)	(0.004)	(0.005)	(0.005)	(0.002)	(0.002)
prcntfemale	0.0187	0.0310	0.0622**	0.0791***	-0.0816***	-0.0961***	0.201***	0.196***
	(0.055)	(0.060)	(0.019)	(0.021)	(0.022)	(0.025)	(0.010)	(0.011)
prcntnhlwhite	0.290***	0.334***	0.231***	0.251***			0.0988***	0.0941***
	(0.016)	(0.017)	(0.007)	(0.008)			(0.003)	(0.004)
prcnt65plus	-0.404***	-0.473***	-0.223***	-0.267***	0.469***	0.488***		
	(0.033)	(0.038)	(0.013)	(0.016)	(0.016)	(0.019)		
prcntbachelors	0.683***	0.653***			0.430***	0.462***	-0.0877***	-0.0946***
	(0.024)	(0.026)			(0.013)	(0.014)	(0.005)	(0.006)
prcntpublictransit	-0.0202	-0.0106	0.0395*	0.0431**	-0.242***	-0.228***	0.00212	0.00793
	(0.033)	(0.033)	(0.016)	(0.016)	(0.019)	(0.019)	(0.008)	(0.008)
prcntrentals	-0.0684***	-0.0845***	-0.154***	-0.160***	-0.182***	-0.151***	-0.0550***	-0.0610***
	(0.018)	(0.017)	(0.007)	(0.008)	(0.010)	(0.011)	(0.004)	(0.005)
_lyear_2017	0.503***	0.503***	0.0464***	0.0407***	-0.141***	-0.156***	0.0425***	0.0413***
	(0.003)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)
logmeanrent			0.0770***	0.0911***	0.0611***	0.0856***	-0.0179***	-0.0234***
			(0.004)	(0.005)	(0.004)	(0.005)	(0.001)	(0.002)
_cons	6.137***	6.182***	-0.311***	-0.384***	0.190***	-0.0438	0.120***	0.174***
	(0.031)	(0.033)	(0.028)	(0.036)	(0.027)	(0.035)	(0.011)	(0.014)
N	118508	76308	118508	76308	118508	76308	118508	76308
adj. R^2	0.896	0.903	0.933	0.935	0.951	0.945	0.807	0.816

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

Table 10: Third Model – Regression with the Introduction of Controls and Distance 2 CBD

	(1) logmeanrent	(2) logmeanrent	(3) prcntbachelors	(4) prcntbachelors	(5) prcntnhlwhite	(6) prcntnhlwhite	(7) prcnt65plus	(8) prcnt65plus
treated	0.00535	0.00614	0.0222***	0.0218***	0.0261***	0.0228***	-0.00676***	-0.00392
	(0.008)	(0.009)	(0.004)	(0.004)	(0.005)	(0.005)	(0.002)	(0.002)
prcntfemale	0.0191	0.0207	0.0636***	0.0800***	-0.0825***	-0.0811***	0.196***	0.189***
	(0.053)	(0.059)	(0.019)	(0.021)	(0.022)	(0.025)	(0.010)	(0.011)
prcntnhlwhite	0.307***	0.351***	0.227***	0.250***			0.103***	0.101***
	(0.016)	(0.018)	(0.007)	(0.008)			(0.003)	(0.004)
prcnt65plus	-0.422***	-0.507***	-0.224***	-0.264***	0.502***	0.521***		
	(0.035)	(0.040)	(0.014)	(0.016)	(0.016)	(0.019)		
prcntbachelors	0.688***	0.653***			0.421***	0.448***	-0.0853***	-0.0920***
	(0.024)	(0.026)			(0.012)	(0.014)	(0.005)	(0.006)
prcntpublictransit	-0.0214	-0.00887	0.0403**	0.0429**	-0.237***	-0.225***	0.00512	0.00856
	(0.032)	(0.033)	(0.016)	(0.016)	(0.019)	(0.019)	(0.008)	(0.008)
prcntrentals	-0.0568***	-0.0855***	-0.154***	-0.160***	-0.176***	-0.146***	-0.0563***	-0.0607***
	(0.017)	(0.017)	(0.007)	(0.008)	(0.010)	(0.011)	(0.004)	(0.005)
year_distance2cbd	0.0000751***	0.0000816***	-0.00000381	-0.00000731	-0.0000568***	-0.000102***	0.0000364***	0.0000374***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
_lyear_2017	0.487***	0.492***	0.0433***	0.0415***	-0.135***	-0.140***	0.0357***	0.0361***
	(0.004)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)
logmeanrent			0.0849***	0.0914***	0.0701***	0.0884***	-0.0198***	-0.0248***
			(0.004)	(0.005)	(0.004)	(0.005)	(0.002)	(0.002)
_cons	4.246***	4.583***	-0.265***	-0.243*	1.554***	1.934***	-0.780***	-0.550***
	(0.311)	(0.347)	(0.067)	(0.096)	(0.087)	(0.125)	(0.042)	(0.054)
N	115939	76308	115939	76308	115939	76308	115939	76308
adj. R^2	0.901	0.903	0.933	0.935	0.951	0.946	0.812	0.819

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

Table 11: Fourth Model – Rent and Educational attainment DVs with Tests for Effect Heterogeneity

	(1) logmeanrent	(2) logmeanrent	(3) logmeanrent	(4) logmeanrent	(5) prcntbachelors	(6) prcntbachelors
treated	0.0165	0.0160	0.0222	0.0219	-0.0254***	-0.0256***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.006)	(0.006)
treatedbachelors	-0.0306	-0.0271				
	(0.035)	(0.035)				
prcntfemale	0.0190	0.0206	0.0194	0.0211	0.0622**	0.0783***
	(0.053)	(0.059)	(0.053)	(0.059)	(0.019)	(0.021)
prcntnhlwhite	0.306***	0.351***	0.306***	0.351***	0.227***	0.250***
	(0.016)	(0.018)	(0.016)	(0.018)	(0.007)	(0.008)
prcnt65plus	-0.422***	-0.507***	-0.423***	-0.508***	-0.221***	-0.259***
	(0.035)	(0.040)	(0.035)	(0.040)	(0.014)	(0.016)
prcntbachelors	0.690***	0.655***	0.691***	0.656***		
	(0.024)	(0.026)	(0.024)	(0.027)		
prcntpublictransit	-0.0213	-0.00886	-0.0218	-0.00933	0.0412**	0.0440**
	(0.032)	(0.033)	(0.032)	(0.033)	(0.015)	(0.016)
prcntrentals	-0.0570***	-0.0858***	-0.0570***	-0.0857***	-0.152***	-0.158***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.007)	(0.008)
year_distance2cbd	0.0000750***	0.0000815***	0.0000752***	0.0000817***	-0.00000405	-0.00000760
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
_lyear_2017	0.487***	0.492***	0.487***	0.492***	0.0434***	0.0416***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.003)
treatedwhite			-0.0473	-0.0446	0.133***	0.134***
			(0.031)	(0.032)	(0.013)	(0.013)
logmeanrent					0.0846***	0.0910***
					(0.004)	(0.005)
_cons	4.247***	4.585***	4.243***	4.581***	-0.257***	-0.236*
	(0.311)	(0.347)	(0.311)	(0.347)	(0.067)	(0.095)
N	115939	76308	115939	76308	115939	76308
adj. R^2	0.901	0.903	0.901	0.903	0.934	0.936

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

Table 12: Fourth Model – Race and Age DVs with Tests for Effect Heterogeneity

	(1) prcntnhlwhite	(2) prcntnhlwhite	(3) prcnt65plus	(4) prcnt65plus	(5) prcnt65plus	(6) prcnt65plus
treated	0.0637***	0.0648***	-0.00697*	-0.00503	-0.00239	0.0000100
	(0.010)	(0.010)	(0.003)	(0.003)	(0.003)	(0.003)
treatedbachelors	-0.104***	-0.116***	0.000577	0.00306		
	(0.024)	(0.024)	(0.009)	(0.009)		
prcntfemale	-0.0828***	-0.0816***	0.196***	0.189***	0.196***	0.189***
	(0.022)	(0.025)	(0.010)	(0.011)	(0.010)	(0.011)
prcnt65plus	0.502***	0.520***				
	(0.016)	(0.019)				
prcntbachelors	0.427***	0.456***	-0.0853***	-0.0922***	-0.0845***	-0.0912***
	(0.013)	(0.014)	(0.005)	(0.006)	(0.005)	(0.006)
prcntpublictransit	-0.237***	-0.224***	0.00512	0.00855	0.00501	0.00844
	(0.019)	(0.019)	(0.008)	(0.008)	(0.008)	(0.008)
prcntrentals	-0.177***	-0.146***	-0.0563***	-0.0606***	-0.0564***	-0.0607***
	(0.010)	(0.011)	(0.004)	(0.005)	(0.004)	(0.005)
logmeanrent	0.0699***	0.0880***	-0.0198***	-0.0248***	-0.0198***	-0.0248***
	(0.004)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)
year_distance2cbd	-0.0000569***	-0.000103***	0.0000364***	0.0000374***	0.0000364***	0.0000374***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
_lyear_2017	-0.135***	-0.140***	0.0357***	0.0361***	0.0357***	0.0361***
	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
prcntnhlwhite			0.103***	0.101***	0.103***	0.101***
			(0.003)	(0.004)	(0.003)	(0.004)
treatedwhite					-0.0123	-0.0111
					(0.008)	(0.008)
_cons	1.556***	1.940***	-0.780***	-0.551***	-0.780***	-0.551***
	(0.087)	(0.125)	(0.042)	(0.054)	(0.042)	(0.054)
N	115939	76308	115939	76308	115939	76308
adj. R^2	0.951	0.946	0.812	0.819	0.812	0.819

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.00$

APPENDIX C

Table 13: Included Metro Areas and Station Counts

Included Metro Areas	Included Stations
Atlanta, GA	2
Charlotte, NC	15
Dallas, TX	39
Denver, CO	20
Houston, TX	16
Jersey City/Newark, NJ	31
Los Angeles, CA	34
Miami, FL	3
Minneapolis, MN	19
Norfolk, VA	11
Oceanside, CA	15
Philadelphia, PA / Trenton, NJ	20
Phoenix, AZ	30
Pittsburg, PA	9
Portland/Beaverton, OR	32
Sacramento, CA	20
Salt Lake City, UT	22
San Diego, CA	5
San Francisco, CA	2
San Jose, CA	19
San Juan, PR	16
Seattle, WA	9
St. Louis, MO	19
Washington DC	7