EXAMINING INCOME-BASED RESIDENTIAL SEGREGATION AND AFFORDABLE HOUSING LOCATIONS, SAN FRANCISCO

A Thesis submitted to the faculty of
San Francisco State University
In partial fulfillment of
the requirements for
the Degree

Master of Science

In

Geographic Information Science

by

Sally Mae Shatford

San Francisco, California

May 2018

CERTIFICATION OF APPROVAL

I certify that I have read Examining income-based residential segregation and affordable housing locations, San Francisco by Sally Mae Shatford, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirement for the degree Masters of Science in Geographic Information Systems at San Francisco State University.

Xiaohang Liu, Ph.D.
Professor

Qian Guo, Ph.D.
Associate Professor

EXAMINING INCOME-BASED RESIDENTIAL SEGREGATION AND AFFORDABLE HOUSING LOCATIONS, SAN FRANCISCO

Sally Mae Shatford
San Francisco, California
2018

Income-based residential segregation is becoming worse in metropolitan cities across the United States due to an increase in income divide between low and high-income households. Low-income households choose where they live based on income and cost of housing. They cannot move out of poor neighborhoods without the help of subsidized housing. This makes affordable housing the most likely avenue for alleviating residential segregation. Using San Francisco as a case study, this research attempts to determine the relationship between affordable housing and income-based segregation and assess how residential segregation has changed from 2000 to 2015. There is a statistically significant relationship between affordable housing, income and segregated, low-income census tracts. The low-income population has shifted from the downtown area in 2000 to the downtown and southeastern area in 2015. The low-income population has less access to transit hubs, jobs and city amenities. Overall, residential segregation has become worse from 2000 to 2015 due to the movement of segregated areas.

I certify that the Abstract is a correct representation of the content of this thesis.

Chair, Thesis Committee Date

PREFACE AND/OR ACKNOWLEDGEMENTS

Much appreciation goes to my committee: Dr. Xiaohang Liu and Dr. Qian Guo. Many thanks to the San Francisco State University Department of Geography and the Environment, specifically Dr. Nancy Wilkinson and Dr. Jerry Davis for their support. Thank you to the Mayor's Office of Community Development and Housing of San Francisco for their data. Lastly, thank you to Dr. Ayse Pamuk and Peter Cohen for their valuable time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

Residential segregation is a growing problem in metropolitan areas across the United States (Massey and Denton 1988; Watson 2009; Wyly and DeFillipis 2010). Residential segregation refers to the limiting of housing areas to homogenous groups. Residential segregation can be based on a variety of demographics such as ethnicity or income. This paper will focus on income-based segregation.

Since 1970, the difference between the high and low-income classes has become more extreme in the United States, and the gap continues to grow (Taylor 2016; Zuk and Chapple 2016), which in turn has led to an increase in income-based residential segregation (Watson 2009; Bischoff and Reardon 2013). According to the Department of Housing and Urban Development, a household is considered low-income if it makes no more than 80% of the area's median income (AMI). It is important to understand residential segregation because it contributes to a worsening divide between class and race by disproportionately affecting lower income and minority populations who have less agency to choose where they live and how much to spend on housing.

Income-based segregation can manifest itself in large numbers of low-income families living in the same, more affordable area of a city because they cannot afford housing in a wealthier neighborhood. The contrasting example would be very wealthy households living in an area of expensive real estate because they don't want to live near a less wealthy population. The isolation of minority and/or low-income groups leads to unequal opportunity for education (Wyly and DeFillipis 2010; Quillian 2014) and unequal access to public services (Chen et al. 2015) for poorer areas in contrast to wealthier neighborhoods, where residents can afford to contribute to and maintain the quality, accessibility, and number of these goods and services (Tighe 2010; Bischoff and Reardon 2013). When poor families continue to live in the same area, segregation compounds (Watson 2009) and turns into extreme poverty, which can lead to decreased mobility and public health hazards (Anderson et al. 2003), sometimes even low birth weights (Grady 2006).

The literature on residential segregation focuses primarily on quantification methods (e.g. Reardon and O'Sullivan 2004) and negative impacts of this segregation on the communities (e.g. Quillian 2014). Moreover, residential segregation research focuses on ethnic minority populations (Massey and Denton 1988; Reardon and O'Sullivan 2004, etc.) instead of low-income populations. Other socioeceonomic determinants such as income (Owens 2015) and poverty (e.g. Wyly and DeFillipis 2010) are little examined. This research is one of the few that focuses on measuring low-income households as the segregated population. Despite the fact that residential segregation is a spatial measurement by definition (Newby 1981), some classical measurements of it, as overviewed by Philippe Apparacio and his co-authors (Apparacio et al. 2014), use data on population count only; the spatial distribution of the population is not considered (e.g. Duncan and Duncan 1955). This research will focus on measuring spatial variations of income-based segregation using spatial statistics.

Income affects where households decide to live. A household's income determines how much they are willing to pay for a housing location (Tiebout 1956). While wealthier households can move into poorer neighborhoods, it is unlikely that they are willing to do so because of the lack of public goods and services in poorer areas. On the other hand, low-income households are probably willing to move into a wealthier neighborhood, but they are not able to do so unless there is a subsidy. To reverse residential segregation, low-income and wealthier households need to be better integrated. Very poor households and wealthy households remain isolated. To this end, affordable housing is the most likely avenue to reverse residential segregation because it can provide housing subsidies to low-income households. The Department of Housing and Urban Development defines affordable housing in relative terms where rent does not comprise more than 30% of a household's income. By strategically placing affordable housing outside of the segregated neighborhoods, it is possible for low-income households to have housing opportunity in wealthier neighborhoods with access to public goods and services and upward mobility (Oakley 2008; Wyly and DeFillipis 2010; Chetty 2015).

There is little research on whether affordable housing has been successful in reversing residential segregation. Only a few articles mention the relationship between affordable housing and residential segregation (Anderson et al. 2003; Oakley 2008; Wyly and DeFillipis 2010; Chen et al. 2015; Owens 2015). These authors approach this problem from a variety of fields. Some authors measure the concentration of affordable housing using various spatial statistics in New York City (Wyly and DeFillipis 2010), four selected major metropolitan areas: Chicago, Atlanta, New York City and Los Angeles (Oakley 2008) and in Beijing, China (Chen 2015). Ann Owens uses traditional residential segregation indices to determine if assisted housing has reversed high-income and low-income based residential segregation across all 331 metropolitan statistical areas in the United States from 1980 to 2005 (Owens 2015). The federal government made efforts in the 1970s to deconcentrate subsidized housing. Contrary to what was expected, these efforts seemed to have caused high-income households to become more isolated (Owens 2015), and federally funded affordable housing developments to become more concentrated (Oakley 2008; Wyly and DeFillipis 2010). Over 40 years have elapsed since the deconcentration efforts, yet affordable housing in major cities is consistently spatially clustered in low-income neighborhoods (Oakley 2009; Dang et al. 2014; Chen et al. 2015). This research will determine if there is a relationship between areas of low income-based segregation and affordable housing locations in the more recent years 2000-2015 in San Francisco, CA.

San Francisco is a relatively small, coastal city/county in California, about 47 square miles. It is bounded by water on three sides. This eliminates any effect tracts have outside of the San Francisco County aside from the southern edge. Since 1999, San Francisco has undergone two tech booms: one in 1999 and one unfolding presently. These tech booms and San Francisco's proximity to Silicon Valley attract educated and generally wealthier workers, pushing out lower income residents and accelerating gentrification. The median income increased from $48,000 to $96,835 between 2000 and 2015. San Francisco has the highest cost of living 164% and highest housing costs 281% in the United States compared to the national average of 100% (Cost of Living Index 2010). While San Francisco's population increased on average by 2915.2 households per year from 2000-2015 (United States Census Bureau 2000; 2015), housing units increased on average by only 2672.2 units per year (San Francisco Planning Department 2015), leaving a deficiency of 243 households per year. As a result, there is an ongoing, worsening housing shortage reflected by the drastic increase of housing prices and fair market rent across the entire city. The fair market rent for a two bedroom apartment increased from $1,167 to $2,062 between 2000 and 2015 (Department of Housing and Urban Development 2015).

San Francisco is ethnically diverse and integrated, which makes income the driving force behind housing choice. For example, most landlords require proof of income three times the monthly rent. Based on the 2015 value for fair market rent, most low-income families cannot afford the fair market rent because they need a combined monthly income of at least $6,186. Additionally, rising costs impact different groups of individuals unequally, with low-income people being impacted most severely (Tighe 2010; Taylor 2015; Taylor 2016; Zuk and Chapple 2016). For such low-income households, the percentage of income spent on rent increases much more rapidly as housing costs rise than it does for households of higher income. This leaves them less funds for other necessities such as food and healthcare, leading to possible displacement, homelessness and poor health (Anderson et al. 2003; Wyly and DeFillipis 2010; Zuk and Chapple 2016). For such low-income households, affordable housing is critical in providing a housing option for those who cannot afford otherwise. Otherwise, these low-income households stay in poorer areas of cities where affordable housing is traditionally located (Anderson et al., 2003; Wyly and DeFillipis 2010). This pattern contributes to and amplifies the cycles of residential segregation (Dang 2014). Therefore, households are more likely to consider rent over the ethnic makeup of the neighborhood when making housing choices. This makes an income-based segregation approach versus an ethnicity-based segregation more relevant to the study of San Francisco.

San Francisco has witnessed a strong pace of economic growth in the last twenty years. Instead of this growth delivering an uptick in overall quality of life across San Francisco, it has instead created deepening of divides. These divides not only manifest themselves in income but also in access to affordable, quality housing. If residential segregation greatly handicaps upward mobility, and the location of affordable housing can reverse the isolation, the placement of affordable housing should be carefully considered with respect to income-based segregation. By adding a case study of San Francisco to the literature, the extent of areas of segregation and affordable housing units will be better revealed. Specifically, I will: (1) Analyze the current spatial patterns and locations of current affordable housing units in San Francisco; (2) Determine if there is income-based segregation in San Francisco in 2000 and 2015, and, if there is income-based segregation, describe the extent and location of segregated areas in 2000 and 2015; (3) Examine the relationship between income-based segregation and locations of affordable housing units, as well as if and how the relationship has changed between 2000 and 2015.

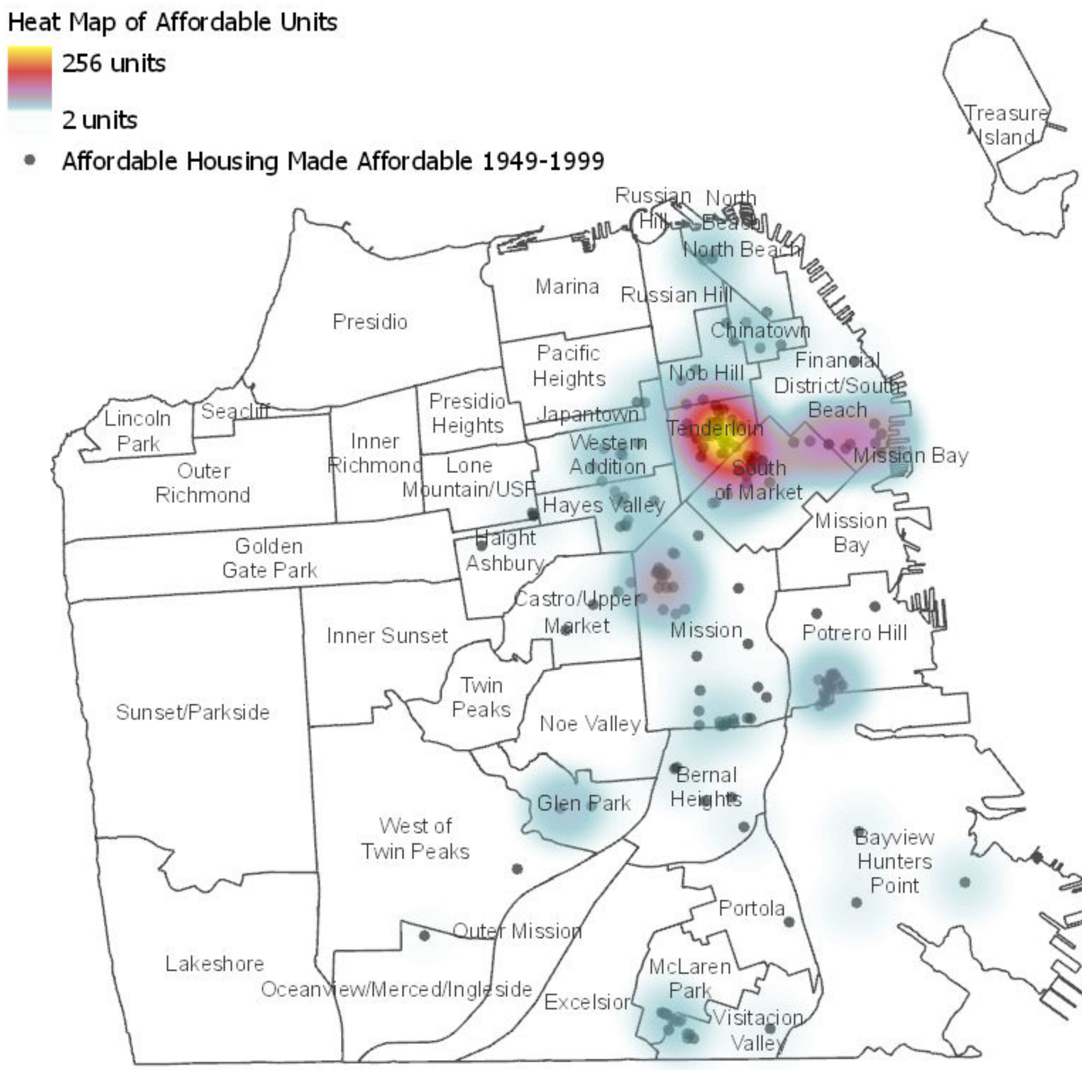
2. Data

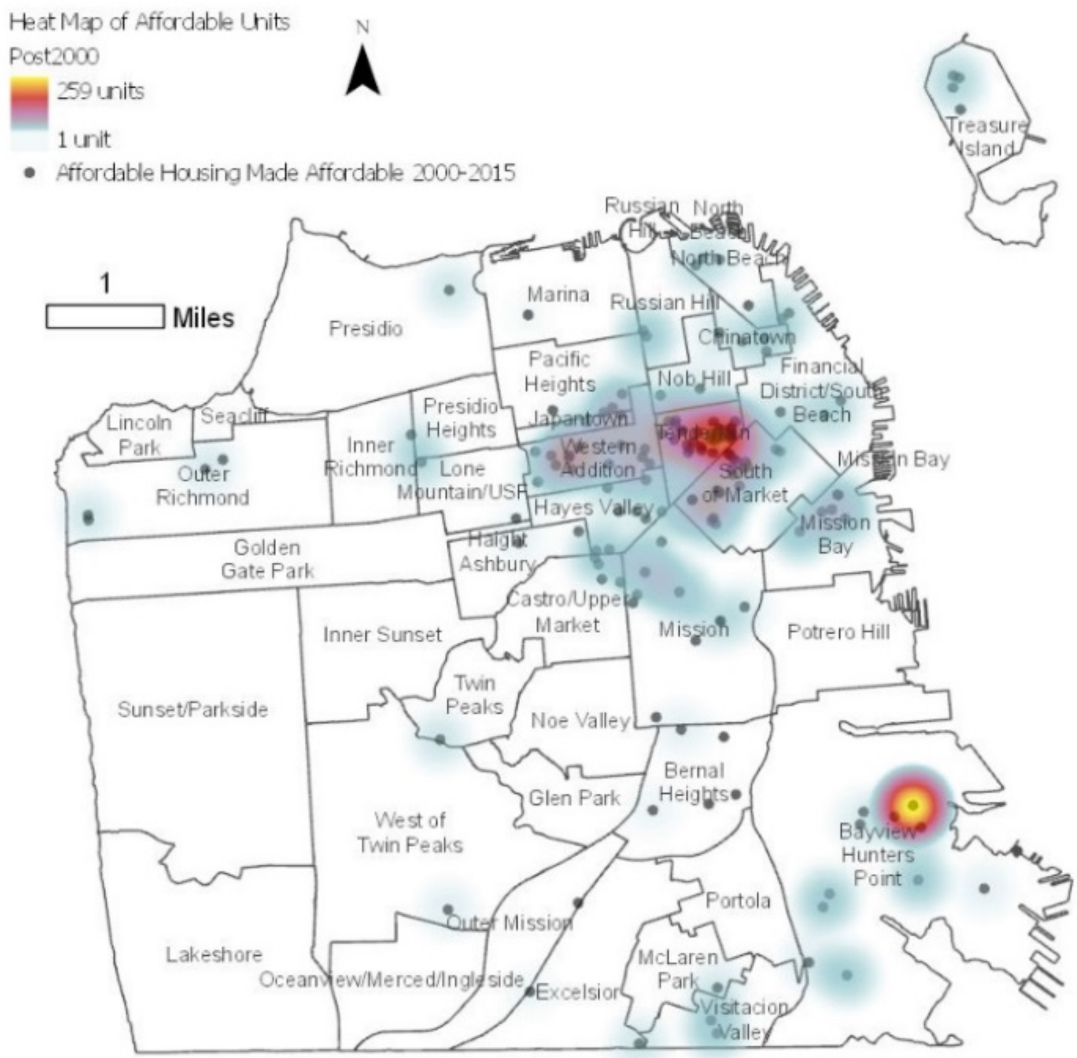
2.1 Affordable housing data

Affordable housing in San Francisco comes in three forms: buildings dedicated to low-income housing, housing vouchers, and inclusionary zoning. The first form is the traditional model for affordable housing where units and complexes are built for multifamily subsidized housing. This includes nonprofit developments, low income housing tax credit developments and public housing. An alternative to this traditional model is housing vouchers, which is also known as Section 8. In this model, the government gives a voucher to a qualifying low-income household or individual to subsidize their rent. The household or individual can take this voucher to a landlord and lives as a tenant in the open market. Another alternative is inclusionary zoning. In San Francisco, for-profit developers must allocate 12% of their onsite units to affordable housing or pay an Affordable Housing Fee. The inclusionary dataset was made available March 2018. This research will look only at traditional developments (i.e., nonprofit developments, low income tax credit developments and public housing) because inclusionary zoning (made available March 2018) and Section 8 data is not available.

The nonprofit developments affordable housing data for this research comes from the Mayor's Office of Housing and Community Development. The low-income housing tax credit development and public housing data was acquired from the Department of Housing and Urban Development. The year affordable for the three datasets have a combined range of 1940-2016. The year it was built and the year it was designated affordable were recorded for the nonprofit affordable housing developments. The year the building was made affordable is used for this research. Low-income housing tax credit developments were built exclusively for affordable housing, so the year built is the same as the year it became affordable. This research will look at the affordable housing made affordable from 1940 to 2015. The total number of units in each building and the number of affordable units in it were recorded. Each unit is a separate apartment that can have a different number of bedrooms. Affordable housing units dedicated to 0-80% area median income was used because that is the income range defined for low-income households. Other units were excluded from this study. Figure 1 shows the spatial distribution of affordable housing developments by 2015 as well as a heat map of the number of affordable units built at each location. The heat map shows particularly dense areas where more units are being built. Each development is categorized by whether it became affordable between 1940-1999 or 2000-2015.

Figure 1: Locations of affordable housing developments built or made affordable from 1940-1999 (a) on above and 2000-2015 (b) below in San Francisco. A heat map based on the number of affordable units designated for 0-80% area median income is shown.





.

2.2 Income data

Income data for the years 2000 and 2015 were acquired from the US Census Bureau. The Census Bureau provides income data by several categories such as households, families, married-couples, and non-family households. Since affordable housing is designated per household income, and household income in San Francisco is known to be positively skewed, median household income is used for this research. For the year 2000, there was a complete decennial survey at the household level but income data were released only at the census tract level. For year 2015, which is the nearest to year 2016 where affordable housing data end, income data come from the 2011-2015 American Community Survey (5 year). Considering that census tracts are the finest spatial unit where income data are available for both 1999 and 2015, census tracts are used for spatial analysis. As explained in the Introduction, a household is considered low income if it makes no more than 80% of the area's median income. Since the average household size in San Francisco was 2.3 in 2000 and 2.26 in 2015 (United States Department of Housing and urban Development, 2000, 2015), the median income for a two-person household was used to determine the area's median income. The value is $48,000 in 2000 and $98,835 in 2015. Using the 80% standard, a household is considered low-income in 2000 if its combined income in 2000 is less than $38,400. By the same token, a low-income household in 2015 makes less than $77,468 that year.

To count the number of low-income households in each census tract, the conversion in Table 1 is used. The Census does not record each individual's income directly. Instead, the distribution of median household income in a census tract is presented using the nine income ranges in Table 1. Each income bracket is categorized into low-income or not low-income based on whether it is less than 80% of the area median income. For example, the census income ranges $0-10,000, $10,000-24,900, and $25,000-34,900 in 2000 are categorized as low-income even the low-income threshold is $38,400 area median income for 2000. For 2000, households that fell into the census income range of $0-34,900 (~73% area median income) were categorized as low-income. For 2015, households that fell into the census income range of $0-74,899 (~76% area median income) were categorized as low income. The census income ranges were categorized to be as close as possible to 80% of the area median income for the appropriate year.

Table 1: Census income brackets and percent AMI designations for San Francisco in 2000 and 2015. Each income bracket recorded by the census is assigned percent median income values in order to determine the count of low-income populations per census tract.

	2000		2015	
Census Income Range	Income Category	80% Area Median Income	Income Category	80% Area Median Income
0-10,000	Low	less than 38,400	Low	less than 77,468
10,000-24,900	Low	less than 38,400	Low	less than 77,468
25,000-34,900	Low	less than 38,400	Low	less than 77,468
35,000-49,900	Not Low	more than 38,400	Low	less than 77,468
50,000-74,899	Not Low	more than 38,400	Low	less than 77,468
75,000-99,900	Not Low	more than 38,400	Not Low	more than 77,468
100,000-149,000	Not Low	more than 38,400	Not Low	more than 77,468
150,000-199,900	Not Low	more than 38,400	Not Low	more than 77,468
200,000+	Not Low	more than 38,400	Not Low	more than 77,468

3. Methods

3.1 Spatial distribution of affordable housing

There are two methods used to determine the spatial distribution of point data: average nearest neighbor (Chen et al. 2015) and Global Moran's I. Other literature considers the concentration of affordable housing across a census tract (e.g. Oakley 2008). This research will look at point data because it better reflects the actual location of affordable housing units. These two methods determine if point data is significantly clustered, significantly dispersed or random. Average nearest neighbor considers how the number of input points would be randomly distributed across the size of the study area. It expects the points to be a certain average distance from each other. If the distance between points in the dataset is larger or smaller than the randomly distributed data, the points are dispersed or clustered, respectively. If the distance between points in the dataset is comparable to that of a normal distribution, the points are considered to be randomly distributed. Global Moran's I works similarly to average nearest neighbor. Each point in the dataset is assigned a weight. Global Moran's I considers the average value of the weight and an average distance data points are from one another. Global Moran's I determines if the dataset is, on average, greater (dispersed), about the same (random) or less (clustered) than the expected values. The Moran's Index ranges from -1 to 1 and produces an expected value and distance for your dataset. Global Moran's I can provide a more detailed description because it measures spatialautocorrelation of the dataset based on a specified weight.

This research will consider the location of each development, number of affordable units and median income of the census tract. Three measurements are calculated for 2000 and 2015 respectively using point locations of affordable housing: (1) average nearest neighbor, (2) Global Moran's I using number of affordable housing units as a weight, and (3) Global Moran's I using household income as a weight.

3.2 Income-based residential segregation

3.2.1 Traditional indices on residential segregation

There are over 42 indices for measuring residential segregation (Apparacio et al. 2014) which are categorized into Massey and Denton's five dimensions of segregation: evenness, exposure, clustering, concentration and centralization (Massey and Denton 1988). Evenness refers to the pattern of dispersal of a subgroup across an area. A subgroup is the potentially segregated population. The more evenly dispersed the subgroup is, the less segregated the area. Exposure defines how much interaction subgroups have with other subgroups. The more the two groups are able to interact, the less segregated the area is. Clustering measures how continuous the area a subgroup occupies. The less continuous the subgroup population, the less segregated they are. Concentration describes the total area a subgroup takes up within the area. The denser the subgroup population is, the more segregated they are. Finally, centralization considers how close or far a subgroup is to the city center. By Massey and Denton's definition, the closer the subgroup is to the city center, the more segregated they are. The underlying assumption is that low-income minorities live in the city center, hence the closer the subgroup is to the city center the more segregated that group is. This dimension is becoming outdated, at least in San Francisco, as wealthier populations move out of the suburbs and closer to the downtown areas. Most of these indices produce a single summary value to describe one of the five dimensions of segregation (e.g. Wong 1999).

Of the five dimensions of segregation, concentration and clustering are most relevant to residential segregation because they define the density and area of the segregated population. This research will use the delta index to measure concentration and the absolute clustering index to measure the segregation of low-income households (0-80% are median income). Of indices that measure concentration, delta is the only spatial index that is used for measuring one group (low-income population). The absolute clustering index was chosen to measure the dimension of clustering because the other indices measure mean proximity between groups, which is similar to the value of average nearest neighbor and Global Moran's I.

The delta index quantifies the concentration dimension of residential segregation as defined in Equation 1 where X is the minority population count in the study area and x_i is the minority population in census tract i. Variable A is the sum of all the census tracts with area a_i. The delta index describes how concentrated a minority population is within the study area. The equation examines the concentration of the minority population in each census tract and determines how far it deviates from the mean concentration of the study area. If there is the same concentration of minorities in each census tract, the delta index would equal 0. If all the minorities are concentrated in one census tract, the delta index would equal 1. Therefore, 0 means not concentrated/not segregated and 1 means concentrated/segregated.

$$0.5\sum_{i=1}^{n}\left|\left(\frac{x_i}{X}\right)-\left(\frac{a_i}{A}\right)\right|$$

Equation 1: Delta index

The absolute clustering index quantifies the clustering dimension of residential segregation as defined in Equation 2 where X is the minority population count in the study area. Variable N is the number of census tracts, x_i is the minority population in census tract i, and x_j is the minority population in census tract j. Variable c_{ij} is the weighted distance (exp(-distance from i to j) from the centroid of tract i to tract j. Variable t_j is the total population of census tract j. The absolute clustering index describes how continuous the segregated census tracts are. It incorporates areas of influence defined by the inverse weighted distance. This area of influence calculates nearby proportion of low-income households to total population within that area of influence. If the proportion of low-income households remains the same across areas adjacent of influence, the low-income population is likely to be clustered. But, if the proportion differs across adjacent areas of influence, the low-income population is dispersed throughout the study area. The absolute clustering index ranges from 0 to 1 where 1 is completely clustered/segregated and 0 is not clustered/segregated at all.

$$\frac{\left\{\sum_{\mathrm{i=1}}^{\mathrm{n}}\left[\frac{x_i}{\mathrm{X}}\sum_{\mathrm{j=1}}^{\mathrm{n}}c_{ij}x_j\right]-\left[\frac{X}{n^2}\sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{j=1}^{n}c_{ij}\right]\right\}}{\left\{\sum_{\mathrm{i=1}}^{\mathrm{n}}\left[\frac{x_i}{\mathrm{X}}\sum_{\mathrm{j=1}}^{\mathrm{n}}c_{ij}t_j\right]-\left[\frac{X}{n^2}\sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{j=1}^{n}c_{ij}\right]\right\}}$$

Equation 2: Absolute clustering index

3.2.2 Local indicators of spatial association

While residential segregation is a spatial measurement by definition (Newby 1981), as it examines how two populations are spatially separated by their counts and locations, traditional methods don't always incorporate a spatial dimension because of the interdisciplinary nature of this topic. Early measures of residential segregation were aspatial, for example the dissimilarity index (e.g. Duncan and Duncan 1955). This aspatial index does not account for the arrangement of the area units with the study area and does not accurately reflect the level of segregation. In other words, this index does not address any of the five dimensions of segregation described earlier in this paper. Newer indices attempt to address the spatial nature of residential segregation by incorporating a spatial dimension (e.g. Morill 1991; Wong 1999). Despite attempts to incorporate a spatial dimension into segregation indices, local variation is lost due to the summary value produced by these indices. This summary value gives no indication as to the movement of areas of segregation and local variation. It is important to specify where the local areas of segregation are and consider the global trend.

To comprehend the local variation in segregation, Local Indicators of Spatial Association (LISA) in spatial statistics is used (Anselin 1995). LISA compares the observed spatial distribution and weight to a statistical random distribution. LISA creates an output where data is put into one of five categories: high-high, low-low, high-low, low-high, and insignificant. High-high and low-low clusters are adjacent census tracts with similar values. High-low and low-high census tracts are outliers. These census tracts differ significantly from their adjacent tracts. In the case of this study, a high-high cluster is formed by census tracts with a high proportion of low-income households that are spatially adjacent. In other words, this is where a statistically significant number of low-income households live. A low-low cluster is formed by census tracts of a very low proportion of low-income households. It is not necessarily a high-income area, but a cluster of very few low-income households. High-low areas refer to a census tract with a high proportion of low-income households surrounded by census tracts with a low proportion of low-income households – vice versa for low-high designations. Insignificant census tracts meet the values expected in a random distribution. They are not significantly similar or different than their adjacent tracts.

Previous methods have used LISA to examine the proportion of affordable housing units per census tract (Oakley 2009) and how that compares to poverty levels (Wyly and DeFillipis 2010). In this research, LISA will be applied for the first time to the study of income-based residential segregation to identify segregated areas within San Francisco. Census tracts from 2000 and 2015 will be used as the input bounding areas. The proportion of low-income households will be used as a weight as described in Section 2.2. The proportion of low-income households was calculated by adding up the number of low-income households (0-approximately 80% area median income) and dividing by the total number of households in each census tract. Furthermore, the local variations captured by LISA will be compared with delta index and absolute clustering index discussed in Section 3.2.1. Such a comparison will reveal the importance of comparing global trends with spatial statistics.

3.3 Poisson distribution

Income-based segregation and affordable housing units must be further examined to determine if a relationship exists between the two. The goal of this study is to determine if there is a relationship between income, segregated areas and affordable housing locations. A Poisson regression was used to determine if the previously mentioned variables have a significant relationship and if they are correlated. This analysis is used when the explanatory variable is continuous, and the response variable is a count. A Poisson regression always produces a value greater than 0 and is based on a Poisson distribution. This regression analysis determines if the relationship between the explanatory and response variable is statistically significant as well as the strength of the relationship.

Census tracts from 2000 and 2015 will be assigned the total number of affordable units, the median household income of that census tract, and a segregated or not segregated designation for 1940-1999 and 2000-2015, respectively. The segregated designation will be used as a categorical predictor variable. Outliers were removed from the data. Significance and correlation will be determined for income versus number of affordable units based on whether the tract is segregated or not segregated for 2000 and 2015. Additionally, number of units located in segregated and not segregated areas will be compared for 2000 and 2015.

4. Results and Discussion

4.1 Spatial distribution of affordable housing locations

Average nearest neighbor and Global Moran's I were run to determine whether the affordable housing developments shown in Figure 1 are randomly distributed, spatially clustered or spatially dispersed. There were 189 affordable housing buildings made affordable built from 1940-1999 and 179 from 2000 to 2015 with units designated for 0-80% area median income. The results are shown in Table 2. The average neighbor which considers only location, shows that affordable housing developments are spatially clustered before and after 2000. The average distance between affordable housing units is 161m while the expected distance apart is 312 m in 2000. Although the affordable housing locations are farther apart after 2000 than before 2000, they still remain spatially clustered. Global Moran's I suggest the same when location is weighted by median household income of the census tract and by total number of affordable housing units before and after 2000. This leads to the conclusion that the locations of affordable housing developments are spatially clustered in certain areas of median household income and areas with a high number of affordable housing units. Affordable housing locations are unequally distributed across San Francisco.

Table 2: Spatial clustering of affordable housing using average nearest neighbor and Global Moran's I method (income and affordable units as weights).

	1940-1999 189 locations			**2000-2015** 179 locations		
Spatial Test	**Spatial Distribution**	**Index Value**	**Expected Index Value**	**Spatial Distribution**	**Index Value**	**Expected Index Value**
Average Nearest Neighbor	Clustered	161m	312m	Clustered	217m	450m
Global Moran's I, Affordable Housing Unit Count	Clustered	0.67	-0.0053	Clustered	0.48	-0.0057
Global Moran's I, Household Income	Clustered	0.813	-0.0053	Clustered	0.49	-0.0057

4.2 Income-based residential segregation

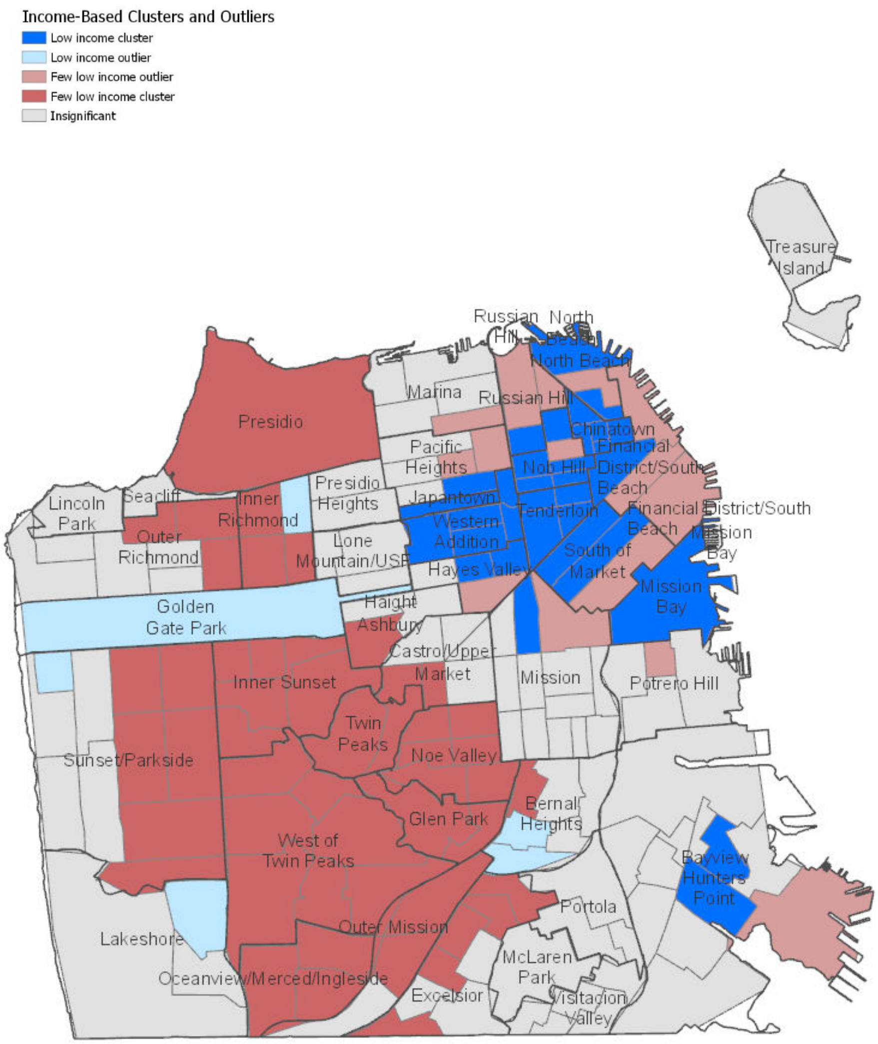
4.2.1 Delta index and absolute clustering index

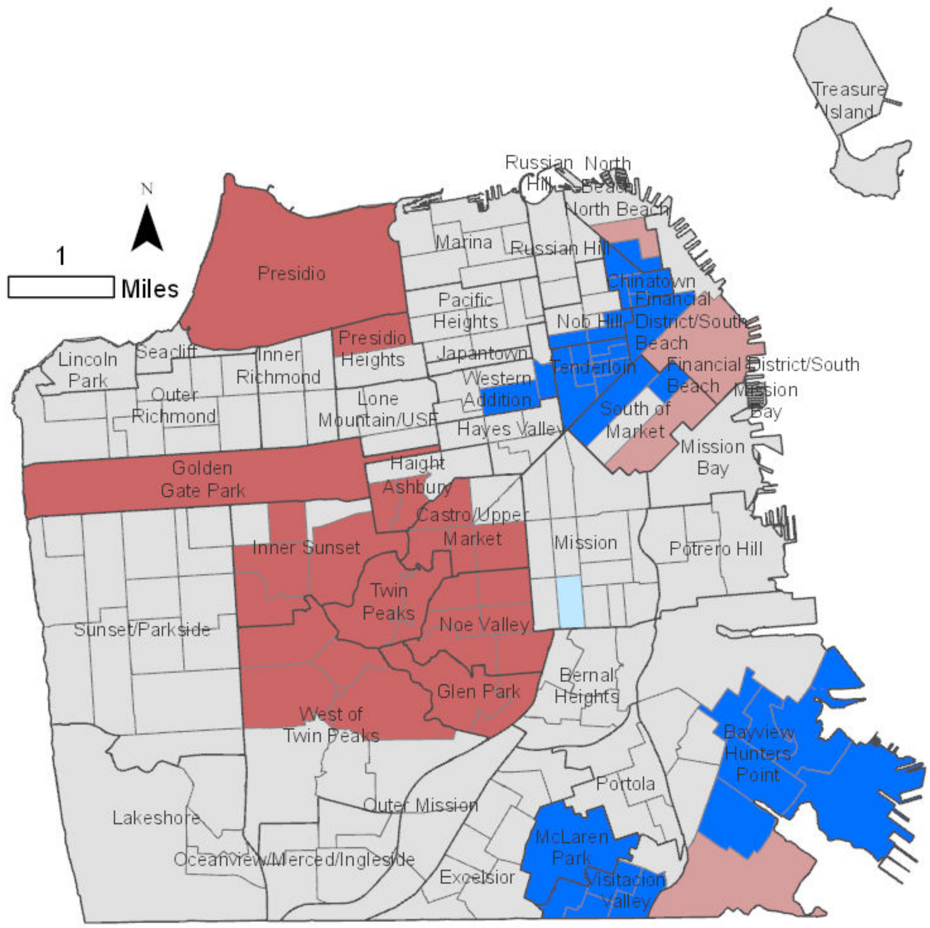
To assess the degree of income-based segregation in San Francisco, the delta index and the absolute clustering index were calculated. The delta index, which describes the dimension of concentration, yielded a value of 0.4134 in 2000 and 0.3730 in 2015. The absolute clustering index, which describes the dimension of clustering, generated a value of 0.2643 in 1999 and 0.1841in 2015. These values for San Francisco indicate there is a slight decrease in concentration and clustering and therefore, a slight overall decrease in residential income segregation between 2000 and 2015. The area occupied by low-income households may have become less continuous from 2000 to 2015. Additionally, the low-income families might have become more dispersed and less dense. These two indices produce global summary values for San Francisco. Note that these indices do not say anything about the specific locations of the low-income population, but measure only the degree to which the low-income population of San Francisco is segregated.

4.2.2 Local variation of income-based residential segregation

To understand how and where income-based residential segregation changed between 2000 and 2015, LISA was calculated based on the location of each census tract and its proportion of low-income households. Figure 2 shows the LISA output where each census tract is assigned one of the five categories mentioned in Section 3.2.2. The clusters with a high proportion of low-income households is of primary interest to this research. These are areas of low-income-based residential segregation.

Figure 2: Spatial clustering of census tracts based on their proportion of low-income households in (a) year 2000 shown above and (b) year 2015 shown below. A low-income cluster is formed if all census tracts in it have high proportion of low-income households. This is an area where income-based residential segregation occurs.





In 2000, there is one large cluster of low-income households in the Tenderloin/South of Market neighborhoods, a small cluster in North Beach, one small cluster in Mission Bay, and a cluster in Bayview. By 2015, there are just three relatively large clusters of low-income households, but the one in the Tenderloin shrank in size by losing South of Market. This is most likely due to rapid gentrification and increase in wealthier households moving in since 2000. This decreases the proportion of low-income households in that neighborhood. The clusters in Mission Bay and North Beach also disappeared. In the last 20 years, development in Mission Bay has taken place in the form of transit hubs and new medical centers. Mission Bay has become a neighborhood with more public services and amenities that wealthier households can afford to live in. Similarly, the neighborhood in North Beach has also disappeared. Wealthier households enjoy the luxury of water-front property and proximity to city centers. As wealthier households move in to lower-income neighborhoods, the overall income of that area increases. In comparison, the small Bayview cluster in 2000 has grown larger and a new low-income cluster emerged in the McLaren Park/Visitation Valley neighborhoods. In 2000, Bayview was still a post-industrial neighborhood with a small low-income cluster. The increase in size of this low-income cluster is most likely due to the lack of amenities provided in these neighborhoods. Bayview/McLaren Park/Visitation Valley don't have much in amenities in the form of transit, jobs etc. They are far away from the city center, so housing is less expensive. This attracts low-income households that want to remain in San Francisco but cannot afford to live near city amenities.

There are four low-income outliers in 2000 in Golden Gate Park, Lakeshore, Richmond, and Bernal Heights. Low-income outliers are census tracts that have a significantly higher proportion of low-income households than their neighboring tracts. The four 2000 outliers disappeared in 2015. Simultaneously, the cluster with few low-income households dropped census tracts in Twin Peaks (southern side) and added census tracts in Castro/Haight-Ashbury/Golden Gate Park neighborhoods (northern side). It seems likely that there are fewer low-income households in the areas closer to downtown areas and parks because wealthier households are attracted to these areas and can pay more for housing. The low-income outlier located in the Richmond neighborhood and the few-low income cluster is not present in 2015. More low-income households have moved into this neighborhood compared to its surrounding areas, which removes this few low-income household cluster and its corresponding outlier. The Richmond is farther from the city center and does not have metro line stops, making it less attractive for wealthier families. The outlier in Lakeshore disappears between 2000 and 2015 because it is no longer significantly different from its surrounding census tracts as the few-low-income cluster has shifted north and slightly east. The low-income outlier in 2015 appears in the

Mission. As the few-low-income cluster moved east, the low-income census tract became different from its adjacent tract in a much wealthier neighborhood, Noe Valley. The area with few-low-income shrank from 2000 to 2015. Additionally, census tracts farther from the city center lost low-income households from 2000 to 2015.

There has been a shift in the last 20 years where wealthier households want to live closer to city centers. They can be close to transit, cultural attractions, and other public amenities. This can lead to gentrification and displacement of low-income households because wealthier households can afford to pay higher rent in these attractive areas. This pattern is shown in the newly emerged low-income clusters that are mainly near San Francisco's southeastern boundary. The low-income clusters in the downtown areas in 2000 have shrunk while the low-income clusters on the city outskirts have increased in area. The areas with few low-income households remain have decreased in size but remained in the same area from 2000-2015. This is because it is much harder for a low-income household to move to a high-income neighborhood. They generally cannot afford the rent unless it is subsidized. There are many fewer options for low-income households. Conversely, it is much easier for a higher-income household to move to a less wealthy neighborhood. Depending on their income, they can decide to move almost anywhere in the city. The rent is cheaper, which leaves them with more disposable income.

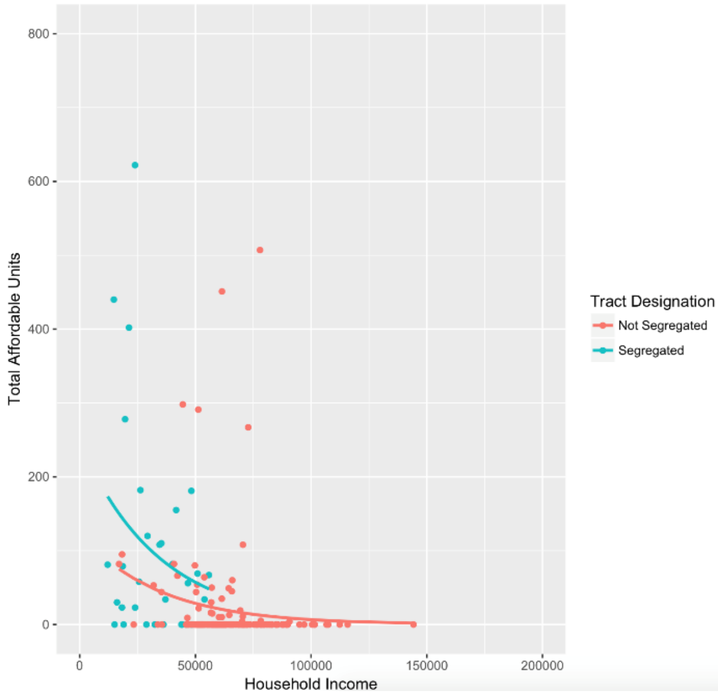
The areas of income-based segregation in Figure 2 show a change from one large, downtown cluster to two smaller clusters. This visualization remains consistent with the concentration and clustering index values calculated in Section 4.1.1. Based on both indices, there is a slight decrease in concentration and clustering in San Francisco from 2000-2015. The low-income clusters become more spread out. Yet, Figure 2 presents a detailed picture of the change in segregation. Although the overall concentration and clustering seems to decrease, the low-income cluster is moving away from the downtown area and towards the outskirts of San Francisco. This shift has negative impacts on the low-income residents who are farther away from public amenities, transit, etc. The local variation within the study area was not accounted for by the delta index and absolute clustering index. The indices show only that overall segregation was decreasing and not that another segregated area appeared.

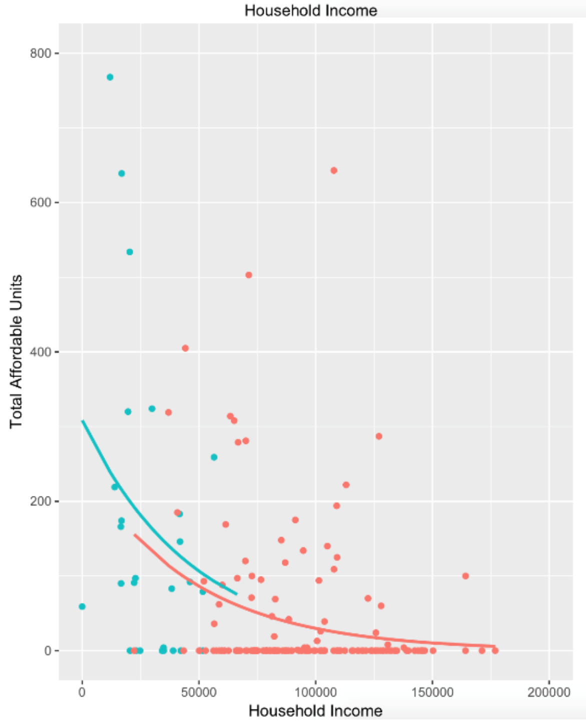
4.3 Relationship between affordable housing location and income-based segregation

This research uses a Poisson regression to determine what the relationship is between household income, number of affordable units and segregated designation. Now that low-income clusters have been located, how does the location of affordable housing relate to these low-income clusters? The location of affordable housing should be alleviating income-based residential segregation. Where affordable housing is being built impacts where low-income households live.

The Poisson regression was run using R. Each census tract has a household income, total number of affordable units, and the categorical variable of segregation. The model is shown in Figure 3. The p-value for the model was ~0 in for 1940-1999 and 2000-2015. The r^2 value was 0.136 for 1940-1999 and 0.189 for 2000-2015, respectively. These statistics mean the number of affordable units cannot be completely explained by household income and segregated designations because r^2 value is closer to 0 than to 1. The correlation between the two variables increased slightly after 2000. There is a negative relationship between affordable housing and income. For example, as affordable housing increases, the household income of that census tract will likely decrease and be in a segregated tract. Census tracts that are segregated are predicted to have more affordable units. Note the difference between segregated and not-segregated linear models in Figure 3. The census tracts with the most affordable units are built in low income tracts. The models shown in Figure 3 indicate a significant relationship between household income and the number of affordable units that are built in that census tract.

Figure 3: Linear models based on a Poison distribution to determine the relationship between household income and affordable housing based on a categorical variable of segregation designation for (a) 1940-1999 shown aboveand (b) 2000-2015 shown below.





How many of these affordable units are located in low-income clusters? There are a total of 7,725 affordable units made affordable or built as affordable housing between 1940-1999 and 14,087 between 2000 and 2015. There are 176 census tracts in 2000 and 195 in 2015. In 2000, 61.3% of affordable units are located in low-income segregated tracts. In 2015, 54.3% of affordable units made affordable after 1999 are located in low-income segregated tracts. There is a slight decrease in the percentage of affordable units being built in the low-income clusters.

In summary, this research examined the relationship between affordable housing locations and areas of income-based segregation in three ways. (1) Analyzed the spatial patterns of affordable housing before and after 2000. (2) Determined the level of income-based segregation using traditional segregation index values. (3) Performed a Poisson regression to determine if there is a correlation between affordable units, income and low-income clusters. Affordable housing remains spatially clustered before and after 2000 in San Francisco. The index values that describe the dimension of concentration and clustering indicated a slight decrease in the overall segregation. Local indicators of spatial association (LISA) shows the low-income clusters shifting from one downtown area to one smaller downtown area and a second area on the outskirts of the city. The low-income cluster is shifting away from the city center. The relationship between affordable housing, income and low-income clusters is statistically significant and becomes slightly more correlated in 2015. Additionally, the percent of affordable units being built in low-income segregated areas remains relatively the same from 2000 to 2015. The spatial patterns of affordable housing remain unchanged. Although the residential segregation indices indicate a decrease in overall segregation, there are still areas of income-based segregation, and affordable housing units continue to coincide with these segregated areas. The locations of low-income segregated areas have shifted, but affordable housing has not been able to reverse income-based residential segregation because most affordable units are still located in segregated areas. Additionally, these segregated areas shifted away from the downtown area where jobs, transportation and other amenities are readily available. This creates more challenges for low-income households pushed to the outskirts of the city because they have unequal access to these amenities.

5. Conclusion

The long-term health and vitality of a city depends on the equitable distribution of its resources, including: quality of housing, adequate public services, educational opportunities, employment opportunities, and accessible transportation. The trend of increased cost of living and correlated increase in segregated communities along wealth lines in San Francisco is a major barrier to its overall equality. Affordable housing provides an avenue to equalize housing opportunities across different household incomes. Therefore, the placement of affordable housing is a critical dimension of ensuring the success and means for mobility of low-income households and reverse income-based residential segregation.

This research shows that spatial statistics can play a role in assessing residential segregation. This study provides a method for locating areas of income-based segregation using local indicators of spatial association to capture local variation in the study area. It is important to identify these low-income clusters to avoid perpetuating residential segregation. Instead of looking at federal deconcentration efforts impact on residential segregation (Oakley 2008; Wyly and DeFillipis 2010; Owens 2015), this study examined the relationship between affordable housing, income and low-income clusters in a modern city (San Francisco) where rapid growth and accelerated gentrification is taking place, resulting in a deepening divide in class.

This San Francisco case study examined how residential segregation has changed from 2000 to 2015 and asked if there is a relationship between affordable housing, income and low-income clusters. Affordable housing was found to be clustered based on distance, household income, and number of units built using average nearest neighbor and Global Moran's I. The location of affordable housing and number of affordable housing units is spatially clustered in San Francisco and spread unequally across household incomes. Delta index and the absolute clustering index produce summary values for the dimensions of concentration and clustering, respectively. These values indicated a slight decrease in overall segregation, but it is unclear where the low-income population is moving within San Francisco. This local variation can be determined using local indicators of spatial association. This method provides a more detailed picture of where residential segregation occurs and how it has changed from 2000 to 2015. The low-income clusters, which are areas of income-based segregation, were concentrated in the downtown area in 2000. As of 2015, there are two areas of low-income clusters: one smaller one in the downtown area and another on the southeastern edge of San Francisco. Although the overall segregation of the city appears to be decreasing, the low-income households are moving farther away from the city center, transit and jobs. The residential segregation indices did not account for a second low-income cluster appearing. About half of the affordable housing units built or made affordable from 1940-1999 and 2000-2015 are located in the low-income clusters. This research revealed a negative relationship between affordable housing units, income and low-income clusters. If the income of the census tract increases, it is more likely to have fewer affordable housing units. This correlation has become slightly stronger from 2000 to 2015, which indicates a stronger relationship between the variables. In summary, although traditional indices indicate that income-based residential segregation has decreased, the segregated areas have simply shifted and created more challenges for low-income families. Additionally, half of affordable housing units are being built in segregated tracts. Income-based segregation in San Francisco has become worse because the segregated area is generating more inaccessibility and unequal housing opportunity for low-income households, and low-income households have little or no upward mobility because half of affordable housing was built in segregated tracts.

This study used spatial statistics (average nearest neighbor, Global Moran's I, local indicators of spatial association), traditional indices (delta index and absolute clustering index) and Poisson regression to determine if there is a relationship between affordable housing and income and to assess the change in residential segregation in San Francisco from 2000 to 2015. The spatial methods were successful in assessing distribution of affordable housing and locating where there are areas of income-based segregation in San Francisco. The traditional segregation indices were not adequate in determining how residential segregation had changed. It gave a summary value that did not describe the true change in the low-income population. Local indicators of spatial association provided a solution by identifying statistically significant areas of low-income clusters.

This research was successful in determining the change in residential segregation from 2000 to 2015 and examining the relationship between affordable housing, income, and low-income clusters. Yet, it was unsuccessful in determining if affordable housing created a segregated tract or vice versa: was the tract already segregated when affordable housing was built or did the continuous building of affordable housing create a segregated tract? Now that a correlation has been established, further research should explore what the relationship is. Is affordable housing perpetuating or alleviating income-based residential segregation?

Affordable housing is essential to subsidizing housing costs for the low-income population and alleviating residential segregation. Therefore, where affordable housing is being built matters in a material and long-term way for the city's entire population. Affordable housing can provide upward mobility (Chetty et al. 2015) to low-income families who would not be able to attain this mobility otherwise. A rapidly growing city like San Francisco needs to take measures to protect its low-income families from displacement and provide housing choice in the form of affordable housing. Without the intentional placement of subsidized housing in a manner that decreases residential segregation, income-based residential segregation will become worse.

References:

Anderson, L. M., St. Charles, J., Fullilove, M. T., Scrimshaw, S. C., Fielding, and J. E., Normand, J. 2003. Providing Affordable Family Housing and Reducing Residential Segregation by Income a Systematic Review. *American Journal of Preventative Medicine*. 24(3): 47-67.

Anselin, L. 1995. Local Indicators of Spatial Association – LISA. *Geographical Analysis.* 27: 93-115.

Apparicio, P., Martori, J. C., Pearson, A. L., Fournier, E., and Apparicio, D. 2014. An Open-Source Software for Calculating Indicies of Urban Residential Segregation. *Social Science and Computer Review.* 32 (1): 117-128.

Bischoff, K. and Reardon, S. 2013, Residential segregation by income 1970 – 2009. *Diversity and Disparities: America Enters a New Century*. 208-233

Chen, M., Zhang, W., and Dadao, L. 2015. Examining spatial pattern and location choice of affordable housing in Beijing, China: Developing a workable assessment framework. *Urban Studies* Col 52 (10): 1846-1863.

Chetty, R., Hendren, N., and Katz, L. 2016. The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving Opportunity Experiement. *American Economic Review* 106 (4): 855-902.

Cost of Living Index. 2010. United States Census Bureau. https://www2.census.gov/library/publications/2011/compendia/statab/.../12s0728.xls

Dang, Y., Liu, Z., and Zhang, W. 2014. Land-based interests and the spatial distribution of affordable housing development: The case of Beijing, China. *Habitat International.* 44: 137-145.

Duncan, O. D. and Duncan, B. 1955 A Methodological Analysis of Segregation indexes. *American Sociological Review*. 20 (2): 210-217.

Grady, S. C. 2006 Racial disparities in low birthweight and the contribution of residential segregation: A multilevel analysis. Social science and medicine. 63: 3013-3029.

Massey, D. and Denton, N. 1988. The Dimensions of Residential Segregation. *Social Forces*. 67(2): 281-315.

Newby, R. G. 1981 Segregation, desegregation, and racial balance: Status implications of these concepts. *The Urban Review*. 14: 17-24.

Oakley, D. 2008. Locational Patterns of Low-Income Housing Tax Credit Developments: A Sociospatial Analysis of Four Metropolitan Areas. *Urban Affairs Review*. 43 (5): 599-628.

Owens, A. 2015 Assisted Housing and Income Segregation among Neighborhoods in U.S. Metropolitan Areas. *The Annals of the American Academy of Political and Social Science*. 660 (1): 98-116.

Quillian, L. 2014 Does Segregation Create Winners and Losers? Residential Segregation and Inequality in Educational Attainment, *Social Problems*, 61(3): 402–426.

San Francisco Planning Department. 2015. 2015 San Francisco Housing Inventory. http://default.sfplanning.org/publications_reports/2015_Housing_Inventory_Final_Web.pdf

State of the Nation's Housing. 2017. Joint Center for Housing Studies of Harvard University. http://www.jchs.harvard.edu/research/publications/state-nations-housing-2017.

Taylor, M. 2015. Legislative Analyst's Office Report: *California's High Housing Costs Causes and Consequences.*

_______. 2016. Legislative Analyst's Office Report: *Perspectives on Helping Low-Income Californians Afford Housing.*

Tighe, R. J. 2010. Public Opinion and Affordable Housing: A Review of Literature. *Journal of Planning and Literature*. 25(1): 3-17.

Tiebout, C. M. 1956. A Pure Theory of Local Expenditure. *The Journal of Political Economy*. 64 (5): 416-424.

United States Census Bureau. 2000. SF-3DP-3 *Profile of Selected Economic Characteristics: 2000* [Data file]. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

_______. 2015. S1901 *Income in the past 12 months* [Data file]. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

United States Department of Housing and Urban Development. 2000. *2010 Adjusted Income Limits* [Data file]. Retried from https://www.hudexchange.info/resource/reportmanagement/published/HOME_IncomeLmts_State_CA_2000.pdf

_______. 2015. *2015 Adjusted Home Income Limits* [Data file]. Retried from https://www.hudexchange.info/resource/reportmanagement/published/HOME_IncomeLmts_State_CA_2015.pdf

_______.2017. *Low-income housing tax credits* [Data file]. Retrieved form https://lihtc.huduser.gov/

_______.2017. *Public Housing Buildings* [Data File]. Retried from https://egis-hud.opendata.arcgis.com/datasets/52a6a3a2ef1e4489837f97dcedaf8e27_0

Watson, T. 2009 Inequality and the measurement of residential segregation by income in American neighborhoods. *Review of Income and Wealth Series.* 55(3): 820-844.

Wong, D. 1999. Geostatistics as measures of spatial segregation. *Urban Geography.* 20 (7): 635-647.

Wyly, E. and DeFillipis, J. 2010. Mapping Public Housing: The Case of New York City. *City & Community*. 9: 61-86

Zuk, M. and Chapple, K. 2016. Housing Production, Filtering and Displacement: Untangling the Relationships. Institute of Governmental Studies Research Brief, University of California Berkeley