PREDICTIVE MODELING TO ASSESS GENTRIFICATION SUSCEPTIBILITY

Kevin Bogue

Background & Theoretical Foundation

For the last several years the Puget Sound real estate market – particularly the market in Seattle – has been growing at an astronomical rate and, while that is great news for homeowners, real estate agents, and real estate developers, it is causing serious issues for those closer to the low end of the socio-economic scale. Perhaps chief among those issues is gentrification.

However, defining gentrification can be problematic and nebulous. As Galster &Peacock suggest in *Urban Gentrification: Evaluating Alternative Indicators* (1986), the definition of gentrification – which is rather ambiguous – has an effect on the findings associated with studies examining it and its impact (p. 321-322). Keeping that in mind, it was necessary to choose an operational definition that was broad enough to be encompassing, yet specific enough to provide guidance.

For the purposes of this analysis, gentrification will be defined as the purchase and renovation of houses and stores in deteriorated urban neighborhoods by upper- or middle-income families or individuals ("Gentrification"). Frequently this leads to improved property values but often displaces low-income families and small businesses. It is perhaps too late for Seattle to be saved from widespread gentrification, but as the overflow from that market further spreads to surrounding cities, it may be possible to identify areas most susceptible to gentrification in the hope that civic leaders, armed with this information, are able to enact proactive measures to alleviate and mitigate the potential impacts.

With that in mind, this research was undertaken with the hypothesis that the process of gentrification is not random, but rather that areas that gentrify tend to follow a pattern which can be identified by the presence, or absence, of certain characteristics. As such, Tacoma, Washington was selected as a testing ground to identify areas within the city that are most susceptible to gentrification and, based on that initial analysis of Tacoma, to begin the development of an automated tool that will allow municipalities, urban planners, or other interested groups to identify areas susceptible to gentrification.

As a theoretical backbone, the work of Dr. Karen Chapple of the University of California, Berkeley and Interim Director of the Institute for Urban and Regional Development was heavily relied upon. Specifically, this analysis has been largely shaped by Chapple's *Mapping Susceptibility to Gentrification: The Early Warning Toolkit* (2009), which examined gentrification in the San Francisco Bay Area and identified 19 indicators with a positive or negative impact on the likelihood of an area to gentrify. While Dr. Chapple's work was aimed somewhat narrowly at the San Francisco Bay Area, the intent is to use her methods to develop a set of indicators that are broad enough to allow for the development of a predictive tool, but specific enough to produce an accurate analysis anywhere in the United States.

Of the 19 determinative indicators Chapple examined it is simply not feasible to conduct analysis on all of them within an automated tool – the knowledge and data needs would be prohibitive. However, by selecting indicators where data is widely available and somewhat standardized, it becomes possible to create the type of broad, yet specific analysis that this research seeks. Indicators fitting those criteria for this analysis include: median income, overburdened renters, ethnicity, single-family housing values, accessibility to parks and highways, and proximity to subsidized housing. Much of this data was accessible via the US Census Bureau which allowed for a high degree of standardization, accuracy, and availability while the remaining data could be analyzed by creating a network dataset to measure different types of distance.

Predicting areas where gentrification is likely might allow Tacoma to proactively enact policies to protect vulnerable populations and to mitigate the overall effects of gentrification even as the real estate market in the area continues to be amongst the hottest in the country (Rosenberg, 2017). Additionally, by using this analysis to write an automated tool, other municipalities and concerned actors would be able to assess susceptibility to gentrification anywhere in the United States.

Methods

Using the previously mentioned gentrification susceptibility indicators – median income, overburdened renters, ethnicity, single-family housing values, accessibility to parks and highways, and proximity to subsidized housing – data was obtained from a combination of the United States Census Bureau, Washington State Geospatial Data Archive, and Pierce County Open GeoSpatial Data Portal.

The most recent tabular data from 2015 was retrieved from the US Census Bureau –Median Household Income in the Past 12 Months (in 2015 Inflation-Adjusted Dollars), Median Gross Rent as a Percentage of Household Income in the Past 12 Months (Dollars), and Race – at the block group enumeration level to increase accuracy in the overall analysis.

Data used to determine single-family housing value – current as of 2008 – was obtained from the Washington State Geospatial Data Archive and was generalized to block groups where the average single-family housing value was calculated as the average of the value for all single-family dwelling parcels within each block group.

Analysis of accessibility to parks and highways, and proximity to subsidized housing was performed using a streets network dataset created specifically to carry out spatial analysis on these topics based on appropriate distance measurement – drive time to highways, walk time to parks, and miles from subsidized housing.

An index was developed where a range of data for each indicator was quantified into five classes, giving each class a numerical ‘score’ (1-5) where a ‘1’ indicates a lower susceptibility to gentrification and a ‘5’ indicates a higher susceptibility to gentrification based solely on the indicator being analyzed. Because no single indicator can directly predict susceptibility to gentrification, a calculation was performed using the quantified (1-5) ‘scores’ contained within each indicator, which produced a composite ‘score’ (7-35) where a ‘7’ indicates the lowest susceptibility to gentrification, and a ‘35’ indicates the highest susceptibility to gentrification based on the totality of the analyses performed.

Processes & Implementation

Household Median Income

When seeking to identify the susceptibility of an area to gentrification, “(i)ncome diversity is a very important indicator: if an area is more diverse, i.e., has relatively equal representation … among income groups, then it is more likely to attract this form of neighborhood change” (Chapple, p.6). Essentially, mixed-income neighborhoods have a greater likelihood to gentrify than those that are more homogenous in terms of household income. However, by running an analysis based on median income for Tacoma, it was discovered that with little variation, higher income block groups tended to cluster together as did lower income block groups. Due to this phenomena, Chapple’s assertion about income diversity as an indicator was slightly altered so that, rather than analyzing areas with great income diversity, areas were examined based on median income by block group, where areas with high median income were determined to be unlikely areas for future gentrification to occur (possibly because areas with high median income are more likely to have previously gentrified) and areas with low median income were determined to be more likely locations for future gentrification based on the definition of gentrification under which the entirety of this analysis was performed

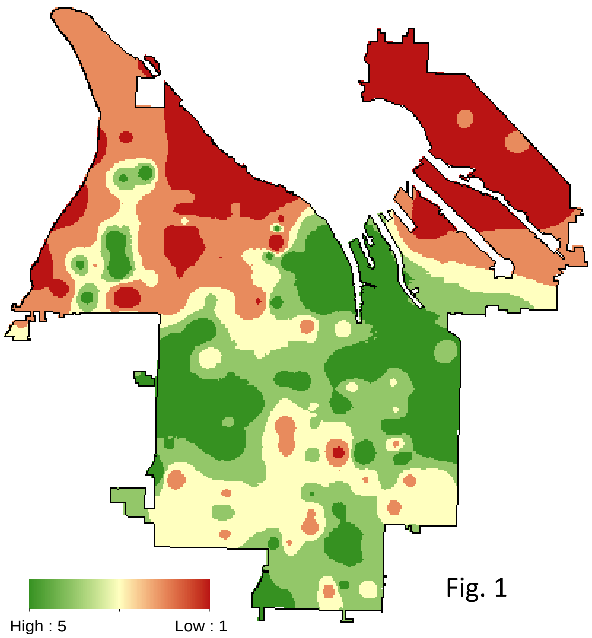

Fig. 1

This revised analysis of Tacoma's median household income by block group was performed by joining 2015 Census data for Median Household Income in the Past 12 Months (in 2015 Inflation-Adjusted Dollars) to block group polygons in Tacoma. Those polygons were converted to points, and a mean of $56,768.30 and standard deviation of $22,963.84 were calculated from which the z-score for each block group (the 'distance' median income for a block group deviates from the average, or mean, for all of the block groups in Tacoma) was determined. The Inverse Distance Weighted method of statistical interpolation was employed to model the z-score for household median income for the entirety of the city by 50-foot grid cells in raster format, where each cell contained a calculated value of median household income z-score.

Findings were then reclassified by quintile and given a numerical value (1-5) where a '1' represents the data's top 20-percent of household median income, and '5' represents the data's bottom 20-percent of household median income (Fig. 1). Findings are symbolized green-to-red where green indicates low median income and higher susceptibility to gentrification and red indicates higher median income and a lower susceptibility to gentrification. Based on the previously discussed definition of gentrification and the methodology employed in this analysis, the areas with higher household median income would be less likely locations for future gentrification than the lower income areas.

Overburdened Renters

As Chapple discovered in her analysis of gentrification in the San Francisco Bay Area, "(a)s reinvestment occurs and property values rise, the potential for different forms of indirect displacement rises. Able to command higher rents on the market, landlords will raise rents to the extent permitted by law, increasing tenant turnover. While these increases may impact any tenant not residing in permanently affordable housing, they are most likely to displace those already paying a disproportionate share of their income for rent, who are not able to squeeze their transportation or food budget any more to pay for housing" (Chapple, p.8). If "… (there are) a high share of renters who pay over 35% of their income for rent, then the neighborhood is more susceptible. It is easy to envision what occurs in this case: as an influx of newcomers increases area rents, these overburdened renters find themselves unable to pay an even higher share of their income for rent, so they depart, leaving more vacancies for new gentrifiers" (Chapple, p. 7).

Using Chapple's findings as a starting point it was determined that, based on the data, the average rent as a percent of household income was 33.99% with a standard deviation of 9.71%, which indicates that the difference between Chapple's definition of an overburdened renter and the average Tacoma renter is approximately 1%. Therefore, analysis for Tacoma deviates slightly from Chapple's, where Chapple was identifying areas where rent comprised 35% or more of household income, the Tacoma analysis will examine rent percentage of household income in relation to the mean for all Tacoma block groups.

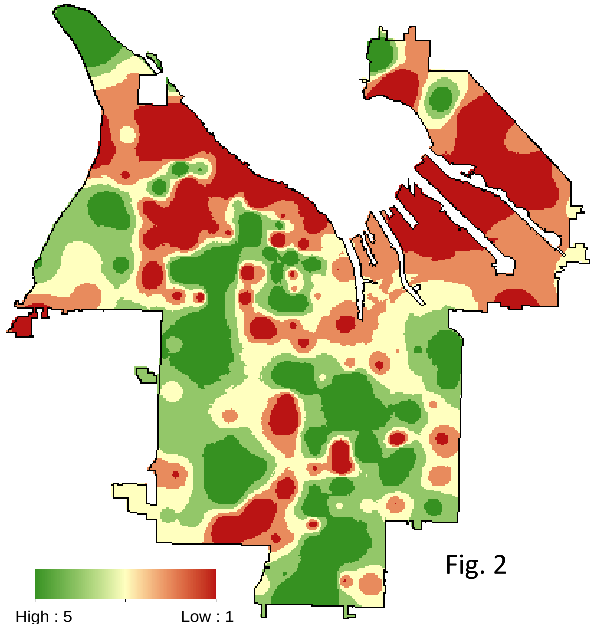

Fig. 2

To do this the tabular Census data for Median Gross Rent as a Percentage of Household Income in the Past 12 Months (Dollars) was joined to block group polygons within the City of Tacoma. In the same manner as determining household median income, the block group polygons with rent data joined were converted to points, the z-score for each block group was calculated, and the Inverse Distance Weighted method of statistical interpolation was used to model the z-score for rent as a percentage of household income for the entirety of Tacoma by 50-foot grid cells in raster form where each cell contained a calculated value for rent percentage z-score.

Findings were reclassified by quintile and given a numerical value (1-5) where a '1' represents the data's bottom 20-percent of rent as a percentage of household income, and '5' represents the data's top 20-percent of rent as a percentage of household income (Fig. 2). Findings are symbolized green-to-red where green indicates a higher rent percentage of household income and higher susceptibility to gentrification, while red indicates lower rent percentage of household income and a lower susceptibility to gentrification.

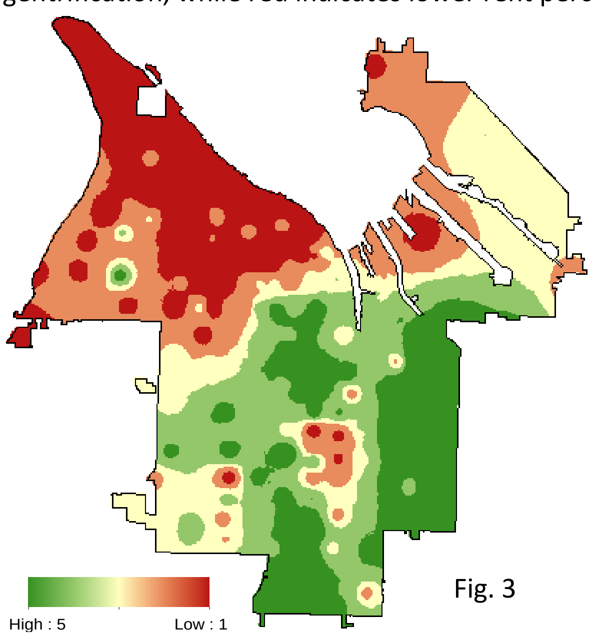

Fig. 3

Ethnicity

According to Chapple, "(t)he more non-Hispanic whites are in the area, the less likely it is to gentrify: the most susceptible areas are those where the majority is minorities" (Chapple, p.7). To spatially determine the ethnic composition of Tacoma, tabular demographic data from the US Census Bureau – specifically RACE 2015 ACS 5-year estimates– which contained data for total population as well as population per ethnic group for each block group in Pierce County was joined to the block groups within Tacoma keeping only matching records.

Using the provided data, the percentage of Non-Hispanic White population per block group was calculated, and an average of 67.6% was determined for all block groups in Tacoma with a standard deviation of 19.8%. Using these derived values, a z-score was calculated for each block group, and the block group polygons were converted to points, then statistically interpolated using the Inverse Distance Weighted method to model all of Tacoma by percentage of Non-Hispanic White population z-score in raster form by 50-foot grid cells, where each cell contained a calculated value for Non-Hispanic White population z-score.

Findings were reclassified by quintile and given a numerical value (1-5) where a '1' represents the data's bottom 20-percent of ethnic diversity – the highest percentage of Non-Hispanic White population – and '5' represents the top 20-percent of ethnic diversity – the lowest percentage of Non-Hispanic White population – (Fig. 3). Findings are symbolized green-to-red where green indicates a lower percentage of Non-Hispanic White population and higher susceptibility to gentrification, while red indicates a higher percentage of Non-Hispanic White population and a lower susceptibility to gentrification.

Single-Family Housing Value

Based on this analysis' definition of gentrification – the purchase and renovation of houses and stores in deteriorated urban neighborhoods by upper- or middle-income families or individuals ("Gentrification") – areas with high single-family property values are unlikely to gentrify compared to areas with lower single-family property values.

Single-family property values were derived from tax parcel polygon data obtained from the Washington State Geospatial Data Archive, from which single-family dwelling data was extracted, and, based upon the taxable value field within the data, a mean of $237,789.54, standard deviation of $133,582.58, and z-score were calculated. The polygon parcel data was converted to points, joined to block groups. Based on z-score the points were interpolated using the Inverse Distance Weighted method to model all of Tacoma by single-family housing taxable value z-score in raster form by 50-foot grid cells where each cell contained a value for the taxable value z-score.

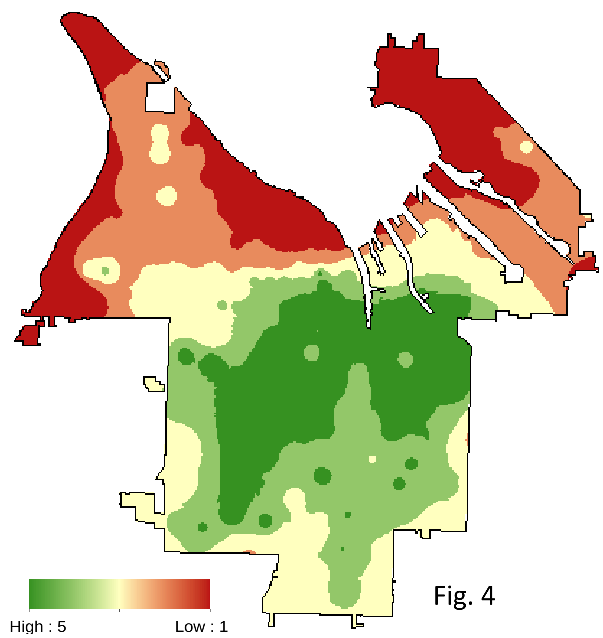

Fig. 4

Findings were reclassified by quintile and given a numerical value (1-5) where a '1' represents the data's top 20-percent of single-family housing by taxable value – the block groups with the highest single-family housing taxable value – and '5' represents the bottom 20-percent of single-family housing by taxable value – the block groups with the lowest single-family housing by taxable value (Fig. 4). Findings are symbolized green-to-red where green indicates a lower single-family housing taxable value and higher susceptibility to gentrification, while red indicates a higher single-family housing taxable value and a lower susceptibility to gentrification.

Access to Parks and Highways

Key factors "in attracting change and investment to (an) area (is) the proximity of amenities such as … public space (and to a lesser extent, small parks), as well as the convenient location of transit (as evidenced by a high share of transit commuters)" (Chapple, p. 6). Tacoma, and many municipalities around the country, have many parks within their boundaries, but while access to transit would undeniably be a major indicator of future gentrification, outside only a handful of major cities, very few municipalities offer transit infrastructure that would provide the same type of gentrification indicator Chapple was examining. However, in a similar vein, access to highways – which allow persons to travel great distances relatively quickly – is similar to access to transit and much more broadly applicable.

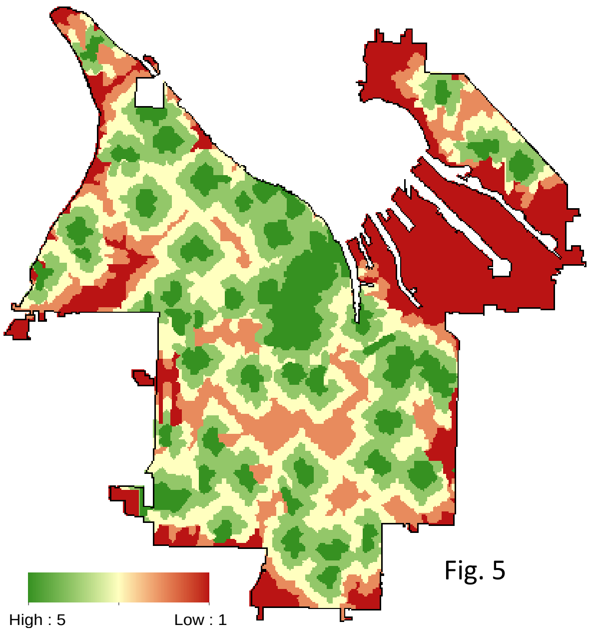

Fig. 5

In order to determine access to highways and parks a street network dataset was created within which drive time (in minutes), walk time (in minutes), and distance (in miles) were calculated for each line segment comprising streets within and around Tacoma (streets within the municipalities of University Place, Fircrest, Lakewood, Puyallup, Fife, and Federal Way were included) in order to accurately model travel since not all travel between locations occurs within a municipal boundary.

To determine parks access, a service area was created based on park location point data with break values at 5, 10, 15, and 20 in terms of walk time minutes that were calculated during the creation of the streets network dataset. The created service areas were merged by break value, converted to a raster, then reclassified into five quantified classes (1-5) and symbolized green-to-red based on walk time (Fig.5). A '1' represents areas more than 20 minutes' walk time from a park—symbolized red and indicating a lower susceptibility to future gentrification – and '5' represents areas within five minutes' walk time of a park – symbolized green and indicating a greater susceptibility to future gentrification.

To determine access to highways a similar analysis to that used to determine access to parks was employed. However, unlike park locations, data for highway on/off ramps was not readily available. From within the attribute tables of the feature classes used to create the streets network dataset, highway on/off ramps were identified and exported

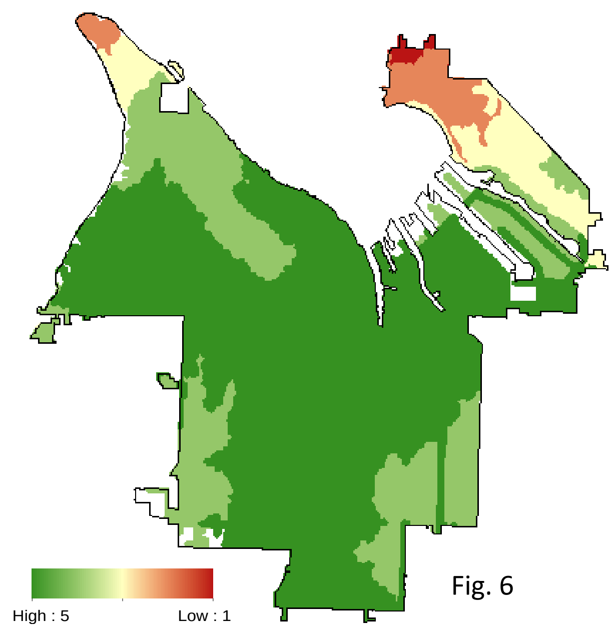

Fig. 6

to create a point feature class with which the analysis could be performed.

Using the highway on/off ramp feature class, a service area was created with 3, 6, 9, 12, and 15 minute breaks in terms of drive time minutes that were calculated during the creation of the streets network dataset. The service areas were merged by break values, converted to a raster, then reclassified into five quantified classes (1-5) and symbolized green-to-red based on drive time in minutes (Fig.6). A ‘1’ represents areas between 12 and 15 minutes’ drive time from a highway on/off ramp—symbolized red and indicating a lower susceptibility to future gentrification – and ‘5’ represents areas between 0 and 3 minutes of a highway on/off ramp – symbolized green and indicating a greater susceptibility to future gentrification. No areas in Tacoma that are accessible by streets were more than 15 minutes’ drive time from a highway on/off ramp. Areas within the city boundary that are not accessible by streets are symbolized white because there was no data for those areas.

Proximity to Subsidized Housing

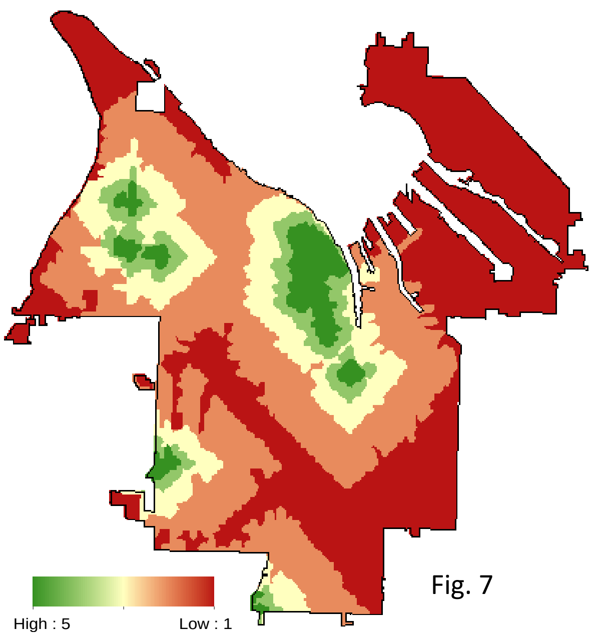

Fig. 7

Chapple states that, “the higher the number of public housing units, the more likely the area is to gentrify, perhaps because there is often a lot of mobility in neighborhoods adjacent to public housing” (Chapple, p. 7). Therefore, since it is unlikely that subsidized housing will be evenly spread throughout a city, it follows that the areas closer to subsidized housing would be more likely to gentrify due to Chapple’s mobility postulation than areas further away from subsidized housing.

To determine proximity to subsidized housing tax parcel data from Washington State Geospatial Data Archive was obtained, and, from within that data, subsidized housing parcels were isolated and converted to point data. Then, using the streets network dataset created to determine access to parks and highways, a service area was generated with breaks of 0.25, 0.5, 1, and 2 miles in terms of the distance values calculated within the network dataset. The service areas were merged by break values, converted to a raster, then reclassified into five quantified classes (1-5) and symbolized green-to-red based on distance (Fig. 7). A ‘1’ represents areas further than two miles from subsidized housing – symbolized red and indicating a lower susceptibility to gentrification – and ‘5’ represents areas 0.25 miles or less from subsidized housing – symbolized green and indicating a higher susceptibility to gentrification.

Findings

Final quantified analyses for median income, overburdened renters, ethnicity, single-family housing value, accessibility to parks and highways, and proximity to subsidized housing were used as inputs calculate a composite ‘score,’ based on the sum of the calculated values (1-5) for each analysis. The lowest possible score was ‘7’ – achieving the lowest score (1) for each input – which represents the least susceptibility to gentrification. The highest possible score was '35' – achieving the highest possible score (5) for each input – which represents the highest susceptibility to gentrification.

Actual calculated scores ranged 8-34 and have been visualized beneath an overlay of the neighborhoods comprising Tacoma (Fig. 8). The lowest scores (symbolized red and indicating the lowest susceptibility to gentrification) were located largely in the West End, North End, and North East neighborhoods. The highest scores (symbolized green and indicating the highest susceptibility to gentrification) were located largely in the Central, New Tacoma, Eastside, South End, and South Tacoma neighborhoods. The absolute highest scores (31-34) were largely found, and highly concentrated in the Hilltop area, which is the dark green area on the border of the New Tacoma, Central, and North End neighborhoods investigating that result showed that there is major neighborhood change already underway in the Hilltop area. The findings were in line with the expected outcome of the analysis.

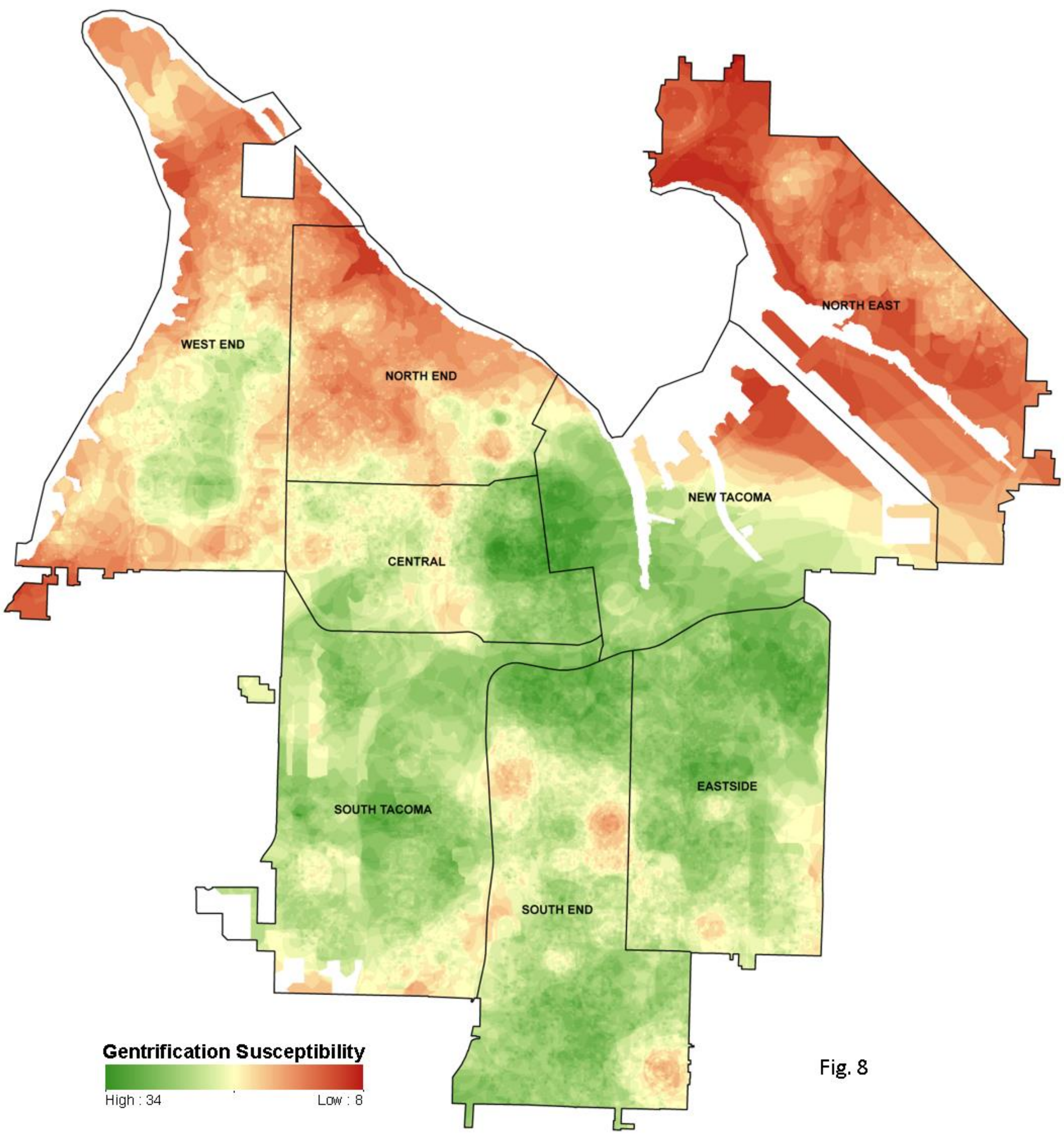

Fig. 8

Continued Research

Tacoma was used as a testing ground from which an automated tool is being developed. Upon completion, this tool will allow for the identification of areas susceptible to gentrification anywhere in the United States. However, before developing the tool, it was necessary to determine whether the seven predictive indicators produced a relatively accurate analysis. Future research will include testing the model in other locations and adjusting as necessary to ensure accuracy.

Currently a rudimentary iteration of the automated tool has been created using ArcGIS' ModelBuilder where four of the seven gentrification susceptibility indicators have been modeled: median income, overburdened renters, ethnicity, and single-family property value. However, the final three indicators – access to parks, access to highways, and proximity to subsidized housing – are currently unable to be modeled using ModelBuilder as they require the creation of a network dataset. Due to the incredible number of variables required to create a network dataset, it is beyond the capabilities of ModelBuilder and requires an in-depth working knowledge of python scripting.

Research will continue with the goal of creating a web-based tool that is capable of locating and pulling all required data autonomously to conduct the gentrification susceptibility analysis for interested parties anywhere in the United States.

Critical Analysis

This analysis of gentrification susceptibility, while comprehensive, is not above reproach. With a nearly infinite number of definitions for gentrification, simply defining the term is problematic due to its inherent ambiguousness. Furthermore, because of the semi-nebulous nature of the definition, there are exponentially more indicators and subsets of indicators that can help to predict the susceptibility gentrification than can realistically be modeled. So, while this analysis specifically sought to identify indicators that were broad enough to be applicable anywhere but specific enough to produce a meaningful analysis, the final map "is but one of an indefinitely large number of maps that might be produced for the same situation or from the same data" (Monmonier, p.2).

The data used in the analysis underwent a generalization and reclassification process which certainly made it easier to understand, but compromised precision to a degree. For instance, 50-foot grid cells for raster analysis may have been too large. Perhaps the analysis would have been better served by conducting examinations using 25-foot or smaller grid cells. Whether the generalization and reclassification processes have a negative impact on the overall analysis will be determined in future testing.

In order to preserve continuity of analyses and to ensure the use of the most recent data, perhaps single-family housing value should have been obtained from the US Census Bureau. Since the stated goal of this research was to produce an automated tool capable of running autonomously, pulling data that is both standardized and known to exist may have been preferable as it cannot be assumed that parcel data containing a taxable value field is widely available nationwide.

The findings being nearly exactly as expected is potentially problematic as it could be the product of bias. As Mark Monmonier states in his book *How to Lie with Maps* (1996), "(t)he map is as it is because the map author "knows" how it should look. This knowledge, of course, might be faulty, or the resulting geographic interpretation might differ significantly from that of another competent observer" (p. 42). As with the generalization and reclassification processes, the accuracy of the findings of this analysis will be examined as it is tested on areas unfamiliar to its author.

Sources

Chapple, K. (2009, August). Mapping Susceptibility to Gentrification: The Early Warning Toolkit. Retrieved February 3, 2017, from http://communityinnovation.berkeley.edu/reports/Gentrification-Report.pdf

Galster, G., & Peacock, S. (1986). Urban gentrification: Evaluating alternative indicators. *Social Indicators Research, 18*(3), 321-337. doi:10.1007/bf00286623

Gentrification. (n.d.). Dictionary.com Unabridged. Retrieved February 15, 2017 from Dictionary.com website http://www.dictionary.com/browse/gentrification

Monmonier, M. (1996). How to Lie with Maps (2nd ed.). Chicago, IL: The University of Chicago press.

Rosenberg, M. (2017, March 6). Seattle and Eastside home prices, after brief slowdown, surge to record highs. *The Seattle Times*. Retrieved April 24, 2017