

# APPENDIX 3

---

Methodology for Mapping Gentrifying  
Neighborhoods in Austin and  
Neighborhood Drilldowns

# Appendix 3

## Gentrification Mapping and Vulnerability Analysis Methodology

<b>Part I: Study Area Analysis</b> .....	<b>135</b>
Determining Study Area .....	135
Vulnerability .....	135
Demographic Change .....	138
Housing Market Change .....	142
Neighborhood Typology .....	145
<b>Part II: Neighborhood Drilldown</b> .....	<b>146</b>
American Community Survey .....	146
CHAS .....	147
Schools .....	147
HMDA .....	147
Affordable Housing .....	148
Building Permits and Residential Real Estate Listings .....	148

# PART I: STUDY AREA ANALYSIS

## Determining The Study Area

The study area includes all census tracts that have lie wholly or partially within the city limits. Even a small area within the city limits warrants inclusion in the study. However, these values will be compared to those for the five-county MSA (made up of Bastrop, Caldwell, Hays, Travis, and Williamson Counties), rather than limited to only the city of Austin. This is in order to take a regional approach to the study of gentrification and displacement.

### 1. VULNERABILITY:

#### Step 1: Locate or calculate the average value for the MSA in each of the categories.

For vulnerability, we considered five factors:

- **Renters:** Percentage of households within the tract who rent.
- **People of Color:** Percentage of people in the tract who identify as anything other than White Non-Hispanic alone.
- **Low-Income Households:** Percentage of households in the tract who earn less than 80% of AMFI (U.S. Department of Housing and Urban Development's Area Median Family Income).
- **Child Poverty:** Percentage of children (i.e., people under the age of 18) in the tract who are live in households below the poverty line.
- **Lack of College Education:** Percentage of adults ages 25 and up in the tract who lack a four-year college degree or higher level of education.

Note that for all five variables, a higher (more positive) number equates to a higher level of vulnerability in the tract. If the opposite is the case for the data that is downloaded, then the variable should be inverted (e.g., percentage of adults 25 and up with a four-year degree or greater should be converted to the same percentage without at least a four-year degree).

For all five vulnerability factors other than Low-Income Households, the MSA figure is available directly from the census. These can be downloaded easily from Social Explorer or American Fact Finder. Low-Income Households share is obtainable from CHAS (Comprehensive Housing Affordability Strategy) data published by HUD. For the CHAS data, the weighted average of the countywide values for the five counties within the MSA can be easily calculated using the CHAS query tool.<sup>1</sup> Collect the figure for each county, as well as the total population (to be used as a weight), then use the formula below to determine the weighted mean in Excel:

=SUMPRODUCT(range of values, range of weights)/SUM(range of weights)

#### Step 2: Download census tract data and calculate percentages

#### Step 3: Calculate Z-scores for each census tract

To help with the calculation of weighted standard deviation, create a new column to the right of the values, and name it  $(x-x_{\text{bar}})^2$ . This is a component to the weighted standard deviation formula and is based on how far the value is from the average. To calculate the  $(x-x_{\text{bar}})^2$  component, use the following formula. **NOTE: It helps to use \$ symbols on the MSA average to avoid having to retype the cell location. For example:**

"\$C\$354"=(value - MSA average)^2

Make sure to delete the resulting value for any census tracts that had a dubious 0—in other words, a 0 that was the result of missing data rather than a true value. Otherwise, this function will calculate the difference between “0” and the average, which will distort your result. Repeat this process with the other 4 sets of values.

Now you can calculate the weighted standard deviation for each set of values. The reason why you’ve included the total population, total housing units, etc. in this spreadsheet is because they will be used to weight each census tract in the calculation of the weighted standard deviation. Before calculating the weighted standard deviation, you will need one more component. This is based on the count of census tracts.

`=(count - 1) / count`

This will be used as you calculate the weighted standard deviation. Below is the formula.

`=SQRT(SUMPRODUCT(range of (x-x+bar)^2 component values, range of weights)/(((count-1)/count)*(SUM(range of weights))))`

It will look like this:

`=SQRT(SUMPRODUCT(B2:B351,D2:D351)/(C358*(SUM(B2:B351))))`

Now you are able to calculate the Z-score for each value, using the components you have calculated. To do this, the “Standardize” function is used.

`=STANDARDIZE(value, mean, standard deviation)`

Once you have a Z-score for every census tract (unless there was no information given for that tract), make a clean worksheet by pasting only the values from the calculation worksheet. You can delete everything except for the Z-scores for each of the 5 categories and the census tract ID for each tract. This will be enough to map in GIS.

#### **Step 4: Determine which Z-scores meet the threshold for consideration.**

If any individual Z-score is more than +0.5, it will be considered in terms of calculating vulnerability. In the worksheet created in the last step, add a column to the right of the Z-score (“vulnerability check”) and use an IF function to mark whether the Z-score meets the threshold:

`=IF(Z-score>.5,1,0)`

`=IF(B2>.5,1,0)`

Once copied to the entire column, you should have “0”s next to any Z-scores that are less than +.5 and “1”s next to any that are more. Repeat this for all five columns of Z-scores.

When this is complete, create a “sum” column that adds up the five vulnerability check columns. Make a new column with an IF function to mark columns with three or more vulnerability indicators over .5.

`=IF(sum column>2,“YES”,“NO”)`

Now any census tracts with “YES” listed are considered particularly vulnerable, while those with a “NO” are not. Also create a new column (“Z Average”) that calculates the average of the 5 Z-scores. (When these are mapped, the average Z-scores for these vulnerable tracts will be displayed, while the others will not.)

## Step 5: Map the Vulnerability Index calculation for each census tract in GIS

Copy into a new worksheet only the values for:

- The FIPS ID
- The five Z-scores (for the individual factors)
- The “YES” / “NO” determination
- The final average Z-score (the average of the five factors)

Make sure all the columns have headers, and that these headers contain no spaces or special characters. Also, check for any errors within the values, as these will result in problems. These values are now ready to be mapped in GIS.

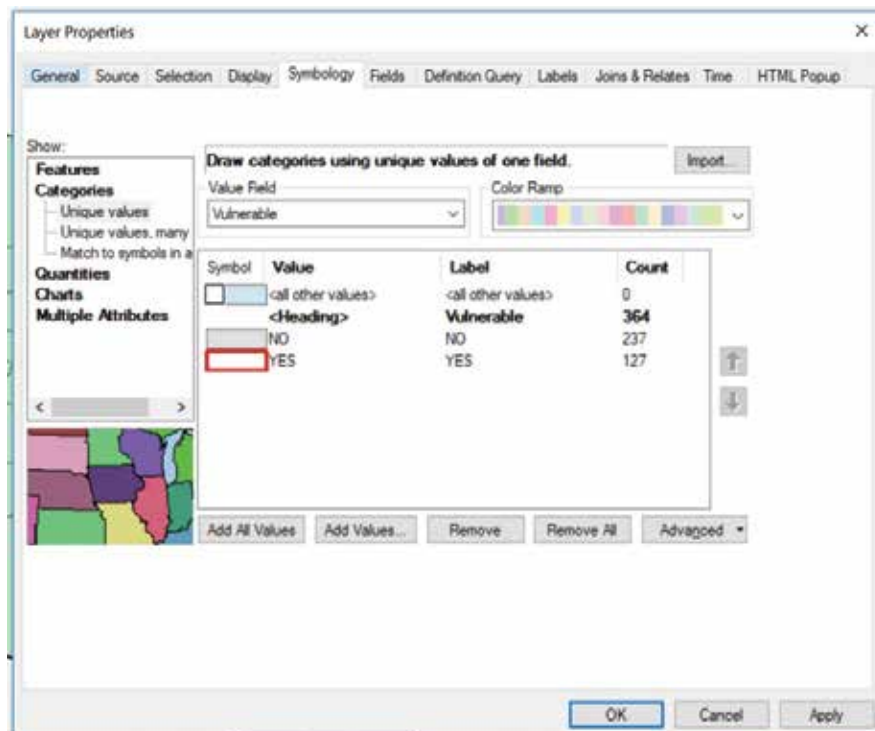
Open a new map, and add in the 2010 Census Tract Boundaries<sup>2</sup>, as well as the county and city limits, major roads and water bodies. (These last two are intended to provide familiar landmarks, since census tracts are unfamiliar geographic units to most people.)



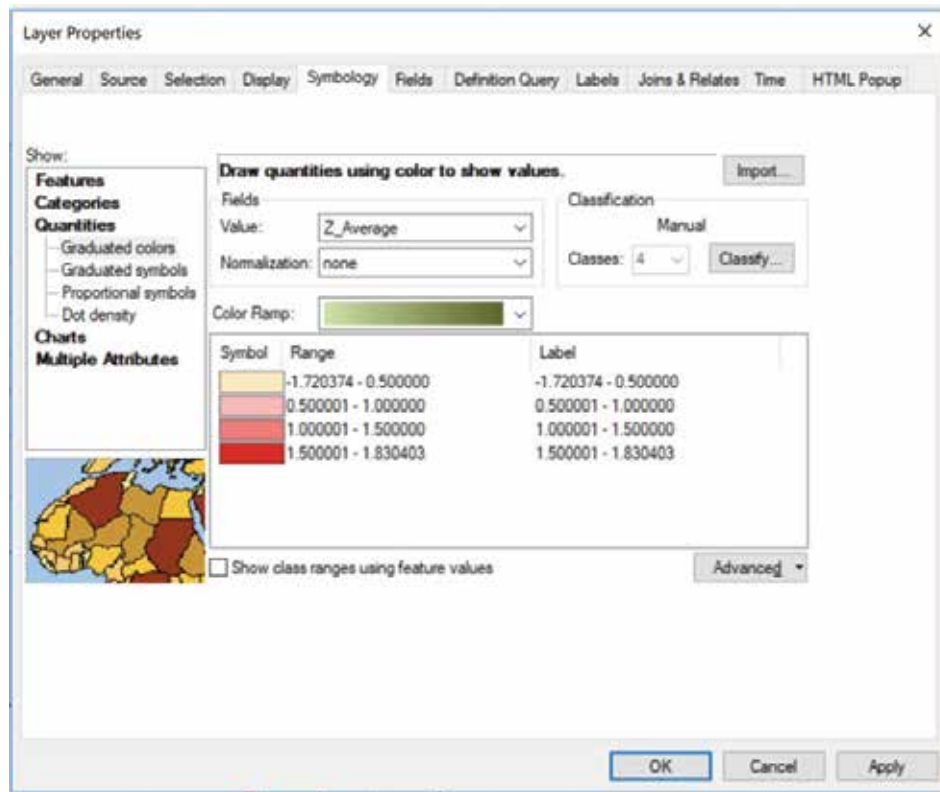
Right click the Census Tract Boundaries file and click Joins and Relates. Click Joins. Fill in the correct sheet to join from and the fields that will join. NOTE: In the TIGER shapefile, you will use the GEOID field to join to your census tract number field. It should look something like this, with the option for “Keep only matching records” selected.

Validate the join, and make sure there are no errors with invalid characters being used in the headings. Also check that the number of fields identified to join is the same as the number in your study area. If not, check your original Excel table and make sure all the census tracts are included. There should be a matching tract located in the TIGER file for every tract.

Once the join is performed, copy and paste this layer onto the data frame, so you have two copies of the same census tract file. Open the properties for one of these census tract layers. Go to the Symbology tab, and click “Categories, Unique Values.” Add all values (“YES” and “NO”) and unclick the option for “all other values”. Change the color for “NO” to a neutral color and “YES” to a red outline.



For the other census tract file, use the symbology tab to select Quantities, Graduated Colors. Create 4 breaks as shown below, and change the colors to reflect those standards:



## 2. DEMOGRAPHIC CHANGE:

In terms of change over time, four factors are considered:

- **Home Ownership:** Percentage Point Change in Owner Occupied Units as a Percent of Total Occupied Units
- **Race and Ethnicity:** Percentage Point Change in White Non-Hispanic Alone Population as a Percent of Total Population
- **Educational Attainment:** Percentage Point Change in the Percent of the Population over 25 years with a Bachelor's degree or higher
- **Income:** Percent change in Median Family Income

(Note the difference between "percentage point" change and "percent change." If the homeownership rate changes from 40% to 50%, there has been a percentage point increase of 10%, but a percent increase of 25%. If income changes from \$50,000 to \$60,000, there has been a 20% percent increase.)

Similar to the vulnerability analysis, instead of comparing the raw percent changes, we chose to look at Z-scores. This allowed us to measure how far above or below the mean change level each of the census tracts fell, giving us a better sense of where the true outliers were.

For this section of the analysis, we chose to look at demographic change over the period of 2000-2016 (based on five-year ACS estimates for 2012-2016). Because of the difference in census tract boundaries between these two years, crosswalk tables were used to "translate" between geographies. These were downloaded from the Longitudinal Tract Database website hosted by Brown University.<sup>3</sup> This process will be described in greater detail in the methodology below:

### Step 1: Calculate MSA averages for Demographic change

Download the four factors at the MSA level for both 2000 and 2012-2016. (Make sure the MSA hasn't changed between 2000 and 2016! If so, you may have to take county averages and combine them into one weighted average for the MSA.)

Calculate for each of these either the percent change or the percentage point change.

Percentage Point Increase/Decrease	Percent Change
Percent Non-Hispanic White Alone	MFI
Percent College Graduate	
Percent Homeowners	

For percent change, use this formula:  $=(2016-2000)/2000$

For percentage point change, use this formula:  $=2016-2000$

These figures will be used when calculating the Z-scores for demographic change in a later step.

### Step 2: Download and prepare ACS Data for 2012-2016

Using the Social Explorer or American Fact Finder websites, all data for 2012-2016 was downloaded and opened in Excel. For each of the points, a percentage of the total was calculated (see above list for denominators). For the percentage of the population without a Bachelor's degree, you will have to add up the first three columns (No High School, High School, Some College) to get the total of those.

### Step 3: Download Census and ACS Data for 2000

Again, using Social Explorer or American Fact Finder, all data can be obtained for 2000.

### Step 4: Download the crosswalk table to translate between geographies

The crosswalk table includes a separate row for every shape that is created when the 2000 and 2010 census tract boundaries are overlaid. The table identifies each of these areas by both their 2000 and 2010 FIPS ID. (Most have not changed, but in cities that experienced a lot of population change during that period, there are often boundary adjustments, splits or joins of 2000 census tracts to form new boundaries.) In the crosswalk table, each of the overlay boundaries is weighted according to how much of the 2000 population can be attributed to each new 2010 geography.

### Step 5: Use the Crosswalk Table to determine a weighted 2010 figure for each of the four factors in 2000.

The use of the crosswalk table is not complicated, but it can be a little bit counterintuitive.

It's easiest to begin is by using the =VLOOKUP function to insert all values from the 2000 data table from Social Explorer into a copy of the crosswalk data for your region. (Don't try to do this the opposite way, as there are usually significantly more census tracts in the later year. If you try to add the crosswalk data to the figures, you will miss some of the census tracts.) **NOTE: You should not delete the "Total Population", "Total Occupied Housing Units", and all other denominator fields because you will be using them to weight later on.** Therefore, the fields you should have in the crosswalk table are:

- 2000 GEOID
- 2010 GEOID
- Weight
- Total Population\*
- Total Occupied Housing Units\*
- Total Owner-Occupied Housing Units\*

- Population Over 25
- White Non-Hispanic Alone Population
- Population over 25 with a Bachelor's degree or more
- Median Family Income (In 2016 dollars)

\* = used for weighting - do not take the weighted average (See instructions below)

To begin, hide the column with the 2010 geographies so it isn't confusing as you're adding the data from the Social Explorer table. Once the data that corresponds to that field is entered, you don't need to use the 2000 geographies anymore and can hide this one, un hiding the 2010 field. Using a pivot table, get a count of how many of each FIPS ID are in the 2010 column. (Alternately, you could sort them in ascending order and check manually, but this can be tedious for larger areas.) It helps to highlight the rows of any matching IDs so that you can come back to them easily when you need to weight them. Most matches will be pairs, but you may find some where there are three or four instances of the same ID. This is not a problem—it just means that the new 2010 tract has been composed of pieces of more than two 2000 tracts.

Add a column to the right of the original. For all multiples of the same ID number, use the function below to calculate the weighted average:

=SUMPRODUCT(range of values, range of weights)/SUM(range of weights)

NOTE: If one of the values seems dubious - for example, a "\$0" listed for MFI, you should not factor it into the weighting, as it will bring down the average artificially. Calculate the weighted average of all other matches, excluding these.

After you've finished calculating all weighted averages for IDs with multiple entries, you can simply copy over the value from any ID without matches. Do not copy over values from fields with a weight listed as "0"—these do not translate to 2010 geographies.

You should now have a value for every ID that is either a weighted average of multiple 2000 values or simply a copy of a single 2000 value. Double check to make sure there are no zeros that seem dubious.

Complete this process for each of the nine fields in the list above.

For the weighting values (population, number of units, etc.), which are based on nominal population figures rather than averages or percentages, use the following formula when there are multiple values for the same census tract instead of calculating the weighted average:

= (value for 1st census tract \* weight of that census tract) + (value for 2nd census tract \* weight of that census tracts) + ...

At this point, keep this working folder with the formulas so you can return to it, as you may run into questions or errors later on that need to be checked. If you don't have the formulas visible, you won't be able to determine what went wrong. Copy and paste all these values into a new worksheet.

To make the worksheet easier to use, you will delete unnecessary information:

- Delete the rows for any fields with a "0" weight.
- Also, delete all but the row that includes each weighted average.



### Step 6: Calculate difference in values from 2000 to 2016.

Calculate for each of these either the percent change or the percentage point change.

Percentage Point Increase/Decrease	Percent Change
Percent Non-Hispanic White Alone	MFI
Percent College Graduate	
Percent Homeowners	

For percent change, use this formula:  $=(2016-2000)/2000$

For percentage point change, use this formula:  $=2016-2000$

### Step 7: Calculate the z-score for each factor

To help with the calculation of weighted standard deviation, create a new column to the right of the values, and name it  $(x-x_{\text{bar}})^2$ . This is a component to the weighted standard deviation formula and is based on how far the value is from the average. To calculate the  $(x-x_{\text{bar}})^2$  component, use the following formula. **NOTE: It helps to use \$ symbols on the MSA average to avoid having to retype the cell location.** For example, "\$C\$354")

$=(\text{value} - \text{MSA average})^2$

Make sure to delete the resulting value for any census tracts that had a dubious 0—in other words, a 0 that was the result of lack of information rather than a true value. Otherwise, this function will calculate the difference between "0" and the average, which will confuse your result. Repeat this process with the other three sets of values.

Now you can calculate the weighted standard deviation for each set of values. The reason why you've included the total population, total housing units, etc. in this spreadsheet is because they will be used to weight each census tract in the calculation of the weighted standard deviation. Before calculating the weighted standard deviation, you will need one more component. This is based on the count of census tracts.

$=(\text{count} - 1) / \text{count}$

This will be used as you calculate the weighted standard deviation. Below is the formula.

$=\text{SQRT}(\text{SUMPRODUCT}(\text{range of } (x-x_{\text{bar}})^2 \text{ component values, range of weights})/(((\text{count}-1)/\text{count})*(\text{SUM}(\text{range of weights}))))$

It will look like this:

$=\text{SQRT}(\text{SUMPRODUCT}(B2:B351,D2:D351)/(C358*(\text{SUM}(B2:B351))))$

Now you are able to calculate the Z-score for each value, using the components you have calculated. To do this, the "Standardize" function is used.

$=\text{STANDARDIZE}(\text{value, mean, standard deviation})$

Once you have a Z-score for every census tract (unless there was no information given for that tract), make a clean worksheet by pasting only the values from the calculation worksheet. You can delete all but the Z-scores for each of the four categories, as well as the census tract ID. This will be enough to map in GIS.

### **Step 8: Determine if a tract experienced demographic change**

For a tract to trigger a “yes” to demographic change, it must have an average Z-score of  $>.5$  and at least two of the two factors must have a z-score above  $.5$ . Create four columns to calculate whether a z-score was above  $.5$ .

Use the formula: `=IF(ZScore=TRUE,"1","0")`

This should mean a z-score above  $.5$  results in a 1 and below  $.5$  results in a 0. Then create a column to sum the four z-score results of the formula for each tract. Results should range from 0 to 4. Finally, to calculate if both conditions are met to trigger demographic change, create a column and enter the following formula.

`=IF(AND(SumFactorsAbove.5 $\geq$ 2,AvgZscore $\geq$ 0.5),"Yes","No")`

## **3. HOUSING MARKET CHANGE:**

For the housing market, we are interested in the median home values in 1990, 2000 and 2012-2016, as well as the percent median home value change from 1990 to 2012-2016 and from 2000 to 2012-2016. Unlike Vulnerability and Demographic Change, the housing market values are not normalized by the MSA average, but rather are broken into quintiles to provide a base understanding of where each tract falls within the range of values for the study area.

Some tracts will be missing Median Home Value information in the surveys because there are not enough homeowner occupied units to produce a reliable Median Home Value. In these cases we use median rent data to determine where a census tract falls in the analysis.

### **Summary of Housing Market Typology:**

**Accelerating tracts** are those with a low or moderate 2012-2016 home value (defined as the bottom 3 quintiles of median home value in 2016) AND with high appreciation between 2000 and 2012-2016 (i.e., top two quintiles of appreciation of median home value).

**Appreciated tracts** are those with a low (bottom three quintiles) median home value in 1990 AND High (top two quintiles) median home value in 2012-2016 AND High 1990 to 2012-2016 appreciation (top two quintiles).

**Adjacent tracts** are those with a low or moderate 2000 value (bottom 3 quintiles) AND low or moderate (bottom 3 quintiles) appreciation of median home value AND that touch the boundary of at least one tract with a high 2016 value and/or high 2000-2016 appreciation (top two quintiles). The following steps require using data from the decennial 1990 and 2000 censuses and the 5-year 2012-2016 American Community Survey (ACS). Because of the difference in census tract boundaries between these two years, crosswalk tables were used to “translate” between geographies. These were downloaded from the Longitudinal Tract Database website hosted by Brown University.<sup>4</sup> This process will be described in greater detail in the methodology below:

### **Step 1: Download and prepare Census 2000 and ACS Data for 2012-2016**

Using the Social Explorer or American Fact Finder websites, all data for 1990, 2000, and 2012-2016 were downloaded and opened in Excel. Download the Median Home Value and Median Rent for each time period.

### **Step 2: Download the crosswalk table to translate between geographies**

The crosswalk table includes a separate row for every shape that is created when the 1990, 2000, and 2010 census tract boundaries are overlaid. The table identifies each of these areas by their 1990, 2000, or 2010 FIPS ID. In the crosswalk table, each of the overlay boundaries is weighted

according to how much of the 1990 or 2000 population can be attributed to each new 2010 geography.

### **Step 3: Use the Crosswalk Table to determine a weighted 2016 figure for the 2016 Median Home Value.**

The use of the crosswalk table is not complicated, but it can be a little bit counterintuitive. It's easiest to begin by using the =VLOOKUP function to insert all values from the 1990 or 2000 data table from Social Explorer into a copy of the crosswalk data for your region. (Don't try to do this the opposite way, as there are usually significantly more census tracts in the later year. If you try to add the crosswalk data to the figures, you will miss some of the census tracts.) **NOTE:** *You should not delete the "Total Population", "Total Occupied Housing Units", and all other denominator fields because you will be using them to weight later on.* Therefore, the fields you should have in the crosswalk table are:

- 1990 or 2000 GEOID
- 2010 GEOID
- Weight
- Total Population\*
- Total Occupied Housing Units\*
- Total Owner-Occupied Housing Units\*
- Median Home Value (In 2016 dollars)
- Median Rent (In 2016 dollars)

\* = used for weighting - do not take the weighted average (See instructions below)

To begin, hide the column with the 2010 geographies so it isn't confusing as you're adding the data from the Social Explorer table. Once the data that corresponds to that field is entered, you don't need to use the 1990 or 2000 geographies anymore and can hide this one, un hiding the 2010 field.

Using a pivot table, get a count of how many of each FIPS ID are in the 2010 column. (Alternately, you could sort them in ascending order and check manually, but this can be tedious for larger areas.) It helps to highlight the rows of any matching IDs so that you can come back to them easily when you need to weight them. Most matches will be pairs, but you may find some where there are three or four instances of the same ID. This is not a problem—it just means that the new 2010 tract has been composed of pieces of more than two 1990 or 2000 tracts.

Add a column to the right of the original. For all multiples of the same ID number, use the function below to calculate the weighted average:

=SUMPRODUCT(range of values, range of weights)/SUM(range of weights)

NOTE: If one of the values seems dubious—for example, a "\$0" listed, you should not factor it into the weighting, as it will bring down the average artificially. Calculate the weighted average of all other matches, excluding these.

After you've finished calculating all weighted averages for IDs with multiple entries, you can simply copy over the value from any ID without matches. Do not copy over values from fields with a weight listed as "0" —these do not translate to 2010 geographies.

You should now have a value for every ID that is either a weighted average of multiple 1990 or 2000 values or simply a copy of a single 1990 or 2000 value. Double check to make sure there are no zeros that seem dubious or incorrect functions.

Complete this process for each of the nine fields in the list above.

definitions directly from Social Explorer—this option may be preferred to what is described here and require less effort.

For the weighting values (population, number of units, etc.), which are based on nominal population figures rather than averages or percentages, use the following formula when there are multiple values for the same census tract instead of calculating the weighted average:

= (value for 1st census tract \* weight of that census tract) + (value for 2nd census tract \* weight of that census tracts) + ...

At this point, keep this working folder with the formulas so you can return to it, as you may run into questions or errors later on that need to be checked. If you don't have the formulas visible, you won't be able to determine what went wrong. Copy and paste all these values into a new worksheet.

To make the worksheet easier to use, you will delete unnecessary information:

- Delete the rows for any fields with a "0" weight.
- Also, delete all but the row that includes each weighted average.

### **Step 5: Calculate difference in values from 1990 or 2000 to 2012-2016.**

For the percentage change in value from 2000 to 2016, you will have to create two new columns labeled HomeValChg\_00\_16 and RentValChng\_00\_16.

For percent change, use this formula: =(2016-2000)/2000

### **Step 6: Rank tracts by quintiles.**

In order to determine the housing market change type for the tract we need to know whether the Median Home Value and change in value (or Rent in the tracts missing home value data) were High or Low.

To the side of the document create new columns titled: HomeValue00\_Break, HomeValue16\_Break, HomeValChg\_Break, Rent00\_Break, Rent16\_Break, RentValChg\_Break.

Calculate the quintile breakpoints for each of the six data values for all of the tracts with the following:

1st Quintile =PERCENTILE.INC(\$B\$2:\$B\$200, 0.2)

2nd Quintile =PERCENTILE.INC(\$B\$2:\$B\$200, 0.4)

3rd Quintile =PERCENTILE.INC(\$B\$2:\$B\$200, 0.6)

4th Quintile =PERCENTILE.INC(\$B\$2:\$B\$200, 0.8)

Next we need to determine if a value is High (in the top two quintiles) or Low (in the bottom three quintiles). Create a home value set of columns titled: 00\_Quintile, 16\_Quintile, Chg\_Quintile, Housing\_Type, Adjacent? And a rent value set of columns titled: 00\_Quintile, 16\_Quintile, Chg\_Quintile, Housing\_Type, Adjacent?

Use the following formula pulling from the appropriate value's 3rd quintile breakpoint, (which were calculated in the previous step) where B2 in the example below is an individual tract value and \$L\$4 is the 3rd quintile breakpoint.

=IF(B2<=\$L\$4,("Low"),("High"))

Use formula to calculate all the Quintile columns.

The next step is to determine the Housing Market Typology using the following formula, where the F column is 16\_Quintile values, G2 is Chg\_Quintile values, and E2 is 00\_Quintile values.

=IF(AND(F2="Low",G2="High"), "Accelerating", IF(AND(E2="Low", F2="High", G2="High"), "Appreciated"))

The above formula can determine two of the three housing market typologies, however the Adjacent typology will require a different formula and the use of ArcGIS, since part of the typology definition is based on spatial adjacency to other high value or high rate of change tracts.

To calculate potential Adjacent tracts, use the following formula:

=IF(AND(E2="Low",G2="Low"),"Adjacent")

Now that you have a list of potentially Adjacent tracts you will need to move to ArcGIS. To prepare your excel data for ArcGIS, copy and paste only the value of the Housing Market Calculations onto a new sheet titled HousingMarket\_GIS. Make sure that none of the titles of your columns include symbols or spaces.

Open ArcGIS and the study tract area shapefile with your census tract outlines and IDs.

Join the study area shapefile to the HousingMarket\_GIS spreadsheet based on the GEOID reference number of the census tracts which should be in both the shapefile and the spreadsheet. Validate the join to check that all data is being properly cross referenced and then complete the join.

Check that the join was successful by opening the study area shapefile's attribute table. You should see all the HousingMarket\_GIS information included at the end of the table.

Export the joined shapefile to create a permanent shapefile titled HousingMarket\_0016. Add the new shapefile to the map document.

Now Select by Attribute all potential adjacent tracts. Make a layer of only the selected features and title it Potential\_Adjacent. Clear your selection.

Use the Select by Location function to select all Potential Adjacent tracts that are touching the boundary of a tract with a "High" 2016 Home Value AND/OR a "High Home Value Change from 2000-2016.

The results will be the Adjacent tracts. You can then update your spreadsheet to reflect which tracts are actually adjacent. Create a final Housing Market Type Column to record the final results.

#### **4. NEIGHBORHOOD TYPOLOGY**

Now that we have calculated Vulnerability, Demographic Change and Housing Market Types we can determine Neighborhood Typology.

Below is a summary of table of how each typology is defined:

Table 1: Tract-level neighborhood typology representing different stages of gentrification <sup>5</sup>

Neighborhood Type	Vulnerable population?	Demographic change?	Housing market condition
Susceptible	Yes	No	Adjacent
Early: Type 1	Yes	No	Accelerating
Early: Type 2	Yes	Yes	Adjacent
Dynamic	Yes	Yes	Accelerating
Late	Yes	Yes	Appreciated
Continued Loss	No	Has % white and % with BA increasing	Appreciated

**Step 1:** Create a new spreadsheet/tab to compare all the necessary information to determine Neighborhood Typology.

**Step 2:** Carefully copy and paste the following information into the spreadsheet: Census Tract GEOID (number), Census Tract Name, Vulnerability Status, Demographic Change Status, Housing Market Type, % Non-Hispanic White Alone increasing and % BA increasing.

**Step 3:** Use the chart to determine if the tract qualifies for a Neighborhood Typology.

## PART II: NEIGHBORHOOD DRILLDOWN

In order to better understand context-specific triggers of displacement and gentrification, Bates (2013) recommends performing “neighborhood drilldowns” that take a deep data dive into a neighborhood. This information should help inform the best tools and policies to combat gentrification and displacement in the particular neighborhood. It is easier for the purposes of working with census data to look at a neighborhood based on census tract boundaries.

### 1. AMERICAN COMMUNITY SURVEY DATA

A baseline understanding of demographic change can be established using American Community Survey data, which can be accessed via the websites of American Factfinder or Social Explorer. It is important to check that the census tract you are interested in has not had a change in its boundaries over the time period being investigated; if it has then you will need to work with Crosswalk data to ensure accurate representation.

**Step 1:** From the American Factfinder home page select the Advanced Search link, then select the “SHOW ME ALL” button.

**Step 2:** Select the “Geographies” button from the left side of the screen. Use the window options to narrow down results to the census tracts you are interested in. Then close out of the window.

**Step 3:** In the top search bar type: “DEMOGRAPHIC AND HOUSING ESTIMATES” and look for the 2016 ACS 5-year estimates result. Select the table.

**Step 4:** Download data – select the “Use the data” button. This should provide you with population, age, and race/ethnicity data.

**Step 5:** Repeat Steps 3 and 4 to search for the following ACS 2012-2016 data: Languages Spoken at Home for Population 5 years and over, Median Household Income, Educational Attainment of population over 25, and Tenure. Note that you can tell where the data comes from in the Dataset column of the search results table.

**Step 6: Repeat process of steps 3-5 to download the same data from the 2000 and 2010 censuses.**

## 2. CHAS DATA

HUD's Comprehensive Housing Affordability Strategy (CHAS) Data is useful for gaining information about the burden of homeownership or renting on different groups in a neighborhood.

**Step 1.** Download data from the HUD CHAS data by selecting the year you are interested in (2006-2014) then below the selected years select the "Data" button. From there select the year "2010-2014 ACS 5-year Average" and the Geographic level "Census Tracts", download as a csv file. This will provide you with information for all census tracts in the country,

**Step 2.** Use the CHAS Data Dictionary to find the correct Table csv file for the information you need. You will need to cross-reference the Data Dictionary Table column description with the coded column labels in table.

Note that the data is broken into different levels of HDMFI (HUD Defined Median Family Income). Depending on how data needs to be displayed it might be necessary to sum numbers across different levels.

## 3. SCHOOLS

Changes in a neighborhood's school population can reveal greater trends, which is why school data is included in the drilldown. Since high schools tend to have very large catchment areas they are left out of the analysis.

**Step 1.** Identify all elementary and middle schools whose catchment areas include portions of the drilldown census tracts. Use the catchment area maps for the local school districts (for example, AISD, Del Valle). For charter schools in Texas, check the Texas Education agency map.

**Step 2.** Research pertinent school information by looking up Texas Education Agency (TEA) Texas Performance Reporting System (TPRS) reports on the TEA website. Select the year that you wish to download a report, then select the "Campus Reports" link on the left side of the page. You can then search for a school by name.

**Step 3:** In order to understand roughly how much of the student -aged population is attending local schools you will need to combine ACS age data with the TRPS reported data on school attendance.

Note that since catchment areas for each school will likely extend beyond the drilldown neighborhood you will likely need to identify the additional census block groups or census tracts that most closely match the catchment area of each school by using the census tract map.

Once you have the age data for all the census tracts/block groups in a school catchment area you can sum the number of 5-9 year olds and see what percent of that total is enrolled in the elementary school, or in the case of a middle school students the sum of the number of 10-14 year olds.

## 4. HMDA DATA

The Federal Home Mortgage Disclosure Act (HMDA) requires banks to report on the applications for home mortgages and included with that information is socioeconomic information about the applicant as well as whether or not the mortgage application was approved. Note that the data does not reflect ALL mortgage applications but does provide an informative cross-section of the mortgage market.

**Step 1.** Go to the FFIEC website page for HMDA to download an Aggregate Report. Search for data in the most recent year available. The site will allow you to narrow down information to the MSA level.

**Step 2.** After selecting your MSA you will need to select the tables you are interested in analyzing. These include: Applications by tract, Home Purchase Loan Applicants by Race/Ethnicity and Home Purchase Loan Denial Rates by Race/Ethnicity. Download data in Excel form.

**Step 3.** Create a new spreadsheet tab to summarize the data you are interested in by census tract number.

## 5. AFFORDABLE HOUSING

Knowledge of the existing subsidized affordable housing in a tract can help inform which tools are appropriate to use in that tract.

**Step 1.** Download/acquire Subsidized Affordable Housing Inventory data from the appropriate authority (typically the city, but sometimes a local affordable housing organization).

**Step 2.** Using ArcGIS, add the study area shape file and the affordable housing data.

**Step 3.** Create individual shapefiles for each “drilldown” tract.

**Step 4.** Use the Select by Location tool to determine which affordable housing properties are located within a drilldown tract. Export the resulting table to use for reference and further analysis.

## 6. BUILDING PERMITS AND RESIDENTIAL REAL ESTATE LISTINGS

Building permits and residential real estate listings are ways to assess the “hotness” of a given neighborhood in terms of building and sales activity. Both data sources are more timely than Census data (e.g., as of the time of writing, in mid 2018, the most recent census tract-level home value data is from 2012-2016). They are also, in some ways, more accurate.<sup>7</sup> Access to data on city-level permitting activity varies by jurisdiction. In Austin it can be obtained from the city’s open data portal.

Residential real estate listings data (Multiple Listing Service, or MLS data) is not publicly available. We obtained MLS data for Austin courtesy of the Austin Board of Realtors (ABOR).

Both permit data and MLS data must be geoprocessed, likely using GIS, to identify data points that lie within the particular drilldown census tract to be analyzed.

## Endnotes

<sup>1</sup> <https://www.huduser.gov/portal/datasets/cp.html>.

<sup>2</sup> <https://www.census.gov/cgi-bin/geo/shapefiles/index.php>.

<sup>3</sup> <https://s4.ad.brown.edu/projects/diversity/Researcher/LTBDDload/Default.aspx>. Note that it recently became possible to download 2000 or 1990 data using 2010 census tract definitions directly from Social Explorer—this option may be preferred to what is described here and require less effort.

<sup>4</sup> <https://s4.ad.brown.edu/projects/diversity/Researcher/LTBDDload/Default.aspx>. As mentioned earlier in note 3, it recently became possible to download 2000 or 1990 data using 2010 census tract

<sup>5</sup> Bates, Lisa, “Gentrification and Displacement Study: implementing an inclusive development strategy in the context of gentrification,” May 18, 2013, at 31, <https://www.portlandoregon.gov/bps/article/454027>.

<sup>6</sup> Since we completed our drilldown analysis, 2011-2015 CHAS data have been released.

<sup>7</sup> Census data on home values relies on homeowner-occupants accurately reporting the value of their own homes to Census enumerators. Clearly homeowners vary in the how accurately they understand the value their home would fetch on the open market were it to be sold.