Calculating a gentrification metric across UWIN sites

Mason Fidino

The metric

We are basing our gentrification metric based on Freeman (2005). The general 'rules' tied to this metric are that a census tract must:

1. Have a median income less than the 40th percentile of the metropolitan area at the beginning of the intercentennial.
2. Has a percentage of housing built over the past 20 years that is below the 40th per centile for the metropolitan area.
3. Has a percentage increase in educational attainment that is greater than that of the metropolitan area.
4. Housing prices increased during the decade.

The data

The objective of this analysis is to determine how historical patterns of gentrification are associated to patterns of urban biodiversity. As such, we needed to compile census data from multiple years. To do so, I used the `tidycensus` package in `R` to query census data from the year 2000, 2010, 2015, and 2019. The 2000 data came from the 10-year decennial census whereas the remaining data come from the 5-year American Community Survey (ACS). The 10 year gap between 2000 and 2010 was because the 2005 5-year ACS data was not available. The 5-year ACS data was used because the 1-year ACS data did not contain estimates for smaller towns.

Across all of these years I compiled data on race, income, number of housing units educational attainment (Table 1) for all census tracts that fell within counties that were sampled. For a given city that is sampled, we are making the assumption that the metropoliation area is "the counties that are sampled."

Step 1. Median income at start of the 20 year period.

To calculate this I took the median income at the 40th percentile for each census tract that fell within a county that was sampled in a given UWIN city. Following this, I determined which census tracts were less then that specific value. I then intersected the site coordinates on those census tracts to determine if that specific site fell within a census tract that had a median income less than the 40th percentile of the studies area in 2000. Unlike other metrics we calculated here, we don't need to compare census tracts across time. Instead, all I did was calculated the 40th income percentile in 2000 and used that as the cutoff for the 2019 census tract levels.

Here is a table that shows how many sites are above (FALSE) and below (TRUE) the 40th percentile of median income. Some cities look to have a pretty even split, which is nice, while other cities (e.g. Salt Lake City, **scut** has many sites above the median income value).

Table 1: The number of sites below the 40th percentile of income in 1990 for each city.

	FALSE	TRUE
ahga	17	11
autx	17	15
boma	20	7
chil	80	31
deco	14	25
inin	29	16
ioio	27	10
jams	31	15
lrar	19	10
mawi	8	14
naca	46	29
phaz	43	49
poor	19	4
rony	9	14
scut	133	13
sewa	21	12
slmo	28	10
tawa	23	19
uril	19	16
wide	20	9

Step 2. Percent of housing built over 20 years

We now need to find locations whose percentage of housing built over the past 20 years is below the 40th per centile for the metropolitan area. To do this we need to figure out the percent change for each census tract from 2000 to 2019, calculate the 40th percentile, and then identify the census tracts that are below the 40th percentile. However, census tracts change a little bit each decade, and we need to be able to know how many houses (roughly) were present in 2000 in the 2019 census tracts. To do this:

1. I converted the raw housing counts into housing density based on the size of the census tracts in 2000 and 2019.
2. I rasterized the housing data from 2000 into equally sized grid cells among cities (500m x 500m).
3. I extracted the rasterized data from 2000 with the 2019 census tracts, taking the mean from each intersecting raster layer.
4. I unscaled the housing density value back to 'number of houses built' (i.e., multiplied by the area of the 2019 census tract).

If a census tract did not change over time, then this method would be equivalent to spatially joining the two datasets. If a census tract did change, then this method is similar to conducting areal interpolation across the 2000 census tracts that a given 2019 census tract intersects with. I decided on a 500 m resolution because extracting the rasterized data from 2000 with the 2000 census tracts returned the same values (i.e., there really wasn't a reason to go to finer spatial scales).

Here is a table that shows how many sites are above (FALSE) and below (TRUE) the 40th percentile of houses built.

Table 2: The number of sites below the 40th percentile of the number of houses built between 2000 and 2019.

	FALSE	TRUE
ahga	17	11
autx	21	11
boma	11	16
chil	68	45
deco	23	16
inin	26	19
ioio	31	6
jams	27	19
lrar	16	13
mawi	13	9
naca	28	47
phaz	61	31
poor	15	8
rony	9	14
scut	114	32
sewa	20	13
slmo	26	12
tawa	23	19
uril	19	16
wide	19	10

Step 3. Has a percentage increase in educational attainment that is greater than that of the metropolitan area.

Again, we need to know what the percent increase is for each census tract, so I needed to rasterize the 2000 census data.

Here is a table that shows how many sites are above (TRUE) and below (FALSE) the average change in in educational attainment.

Table 3: The number of sites below the 50th percentile of educational attainment (i.e,. college degree) between 2000 and 2019.

	FALSE	TRUE
ahga	19	9
autx	11	21
boma	18	7
chil	49	64
deco	14	25
inin	16	29
ioio	12	25
jams	20	26
lrar	10	17
mawi	11	11
naca	34	40
phaz	31	56

Step 4: Housing prices increased during the decade.

calculating this is similar to comparing changes in the number of houses built within a given census tract. However, we also need to account for inflation in these calculations. I went to the [U.S. Bureau of Labor Statistics website] (https://www.bls.gov/data/inflation_calculator.htm) and used their inflation calculator to determine how much the price of $1 has changed between January 2000 and January 2019 (it is $1.50). Thus, I multiplied the dollar values of median housing prices in 2000 by `1.50` before comparing changes in housing prices.

Here is a table that shows how many sites reside in census tracts that have increased in price (TRUE) over time and those that have not (FALSE)

Table 4: The number of sites that reside in census tracts whose housing prices have increased between 2000 and 2019.

	FALSE	TRUE
ahga	9	19
autx	2	30
boma	4	23
chil	53	60
deco	4	35
inin	28	17
ioio	4	33
jams	13	33
lrar	10	19
mawi	8	14
naca	6	69
phaz	23	69
poor	0	23
rony	14	9
scut	5	141
sewa	3	30
slmo	5	33
tawa	2	40
uril	17	18
wide	1	28