Typical Home Price & Neighborhood Change in Atlanta

Intro

This study looks at net changes in demographic data at the neighborhood statistical area level in Atlanta from 2010 to 2018, focusing on whether or not there is correlation between the rate of change of the following social variables (x) commonly tied to gentrification and displacement:

- % of change median household Income
- % of residents who are Non-Hispanic, white only
- % of residents age 18-39
- % of housing units that are owner-occupied
- % of housing units that are vacant
- % of population 25 years and over with a bachelor's degree or beyond

and changes in the typical home value of a neighborhood (y) using multivariate regression. Typical home value of a neighborhood (y) and neighborhood change, broadly (x) combined, is meant to be a proxy for understanding patterns of gentrification and displacement through financial means.

The goal will be to run two different multivariate regression models, determine which is "stronger," and interpret findings from the resulting statistics.

Literature Review

Questions about what's causing outmigration of historically excluded communities in urban centers are causing a stir. Chicago for example is witnessing a decline in its Black population while its suburbs gain Black residents (Kapos et. al, 2021). Referencing Marcuse, Valli points to four key types of displacement within gentrification literature: "(1) direct last-resident displacement; (2) direct chain displacement, including not only the displacement of the last tenants, but also previous ones in the history of the building; (3) exclusionary displacement; that is, the exclusion of certain kinds of households from gentrified housing stock; (4) displacement pressure" (2015, p. 1193). For the purposes of this study, I attempt to build a model in reference to the third type, *exclusionary displacement in relation to housing costs*.

The criticality of understanding trends at the neighborhood level is underscored by Neil Smith's rent gap theory, whereby the land's "highest and best use" is extracted. Stehl summarizes this as, "[...] the difference between present revenues from a given parcel of land - 'capitalized rents' - and the potential rents that parcel could yield, given prevailing rents at the metropolitan level" (2019, p. 8). The focus on cities becomes more imperative with this theory at hand, and with the history of disinvestment in cities; neighborhoods closer to city centers that may have been hit hard by suburbanization or the 2008 financial crisis are conveniently targets through which the rent gap may operate (Slater, 2018, pp. 57-59). I focus primarily on a singular type of real estate in single family housing. Lauermann observed that luxury housing for example occurs primarily in already gentrified, white, wealthy areas (2021, pp.13-15). Thus, I am interested if I can determine - without a "luxury" designation from the data - whether or not trends with the explanatory variables I have chosen are more or less prevalent in neighborhoods of different price points or rates of increase.

While the social variables in the ensuing analysis should not be interpreted as a way to find those to blame for gentrification, modern discourse provides me some trends which I can test. Holleran for example writes of the white, middle class, young, educated professionals that tend to make up the YIMBY movement and how this can lead to their influx into the city as well as gentrification (2021, p. 849). Additionally, the home-valuation process in the U.S. is noted to be highly correlated with race, in which white neighborhoods are valued at higher levels than communities of color (Howell, Korver-Glenn, 2021, pp.1052-1053). Atlanta is no stranger to the suburbanization of poverty as in the case of Sandy Springs (Lanari, 2019, pp. 369-370) and with the influx of rapid development over the past few decades, many people have been asking questions about displacement for a long time.

Data Description / Methodology

Explanatory Variables - City of Atlanta Data Set

In 2020, the City of Atlanta released their Neighborhood Change report using Atlanta Regional Commission's data, which transforms American Community Survey data to the Neighborhood Statistical Area level. In it are a number of socio-economic and demographic variables at the neighborhood level for the years 2010 and 2018. The data and report aim to help understand and identify growth, displacement, low-income concentration and population decline.

From this data, I have derived by explanatory variables. For the purposes of this report, the data identified as useful according to the literature review are:

- Median household income (USD) = *(med_household_income_percentchange)*
- % of residents who are Non-Hispanic, white only = *(white_percentchange)*
- % of residents age 18-39 = *(percentchange_18_39)*
- % of housing units that are owner-occupied = *(owner_occupied_percentchange)*
- % of housing units that are vacant = *(vacant_percentchange)*
- % of population 25 years and over with a bachelor's degree or beyond = *(baplus_percentchange)*

The data set itself provides an estimated percentage and margin of error in 2010 and 2018 for each of these variables. However, the explanatory variables for the analysis are the *percent change* of these values, calculated as:

(X - Y) / Y, where X = the value in 2018 and Y = the value in 2010.

Neighborhood Identification - ARC Data Set

Since the City of Atlanta data set provides neighborhoods as a unit of analysis but only by their NSA signifier code, I mapped the Atlanta Regional Commission's neighborhood names from their open source repository on neighborhood statistical areas. While little of the model makes use of these names explicitly, it will help interpret the results.

ZHVI Data Set

In order to account for housing prices, I joined an Atlanta-specific cut of the Zillow Home Value Index (ZHVI) to the City of Atlanta Neighborhood Change Data Set. The specific data used from Zillow is the ZHVI Single-Family Homes Time Series ($) at the Neighborhood geographic level. The home value produced is not the median home value of the region, but is rather calculated as a weighted average of the middle third of homes in a given region - in this case, neighborhoods. These reflect "typical values" for homes in the 35th-65th percentile range. I will henceforth refer to this unit of measurement as the "typical" home value for a neighborhood.

To prepare it for analysis and joining with the other data, I filtered for entries where the "City" field read Atlanta. This was to exclude metro-Atlanta neighborhoods that might be outside the City of Atlanta's data set. The Zillow data set lists the typical home value for the region by month over multiple years. I removed all years that were not 2010 or 2018.

Taxonomy and Aggregation of Neighborhoods

In order to turn these monthly units of typical home values into an annual unit and thus be comparable with the annual units in the City of Atlanta data set, I took the average of each Zillow-identified neighborhood for 2010 and 2018 separately. However, the Zillow-identified neighborhoods in the "Region Name" field were often smaller than what was in the "NEIGHBORHOOD" field in the Atlanta Regional Commission (ARC) data set

In joining the two data sets, I found that many of the variables under "Region Name" in the Zillow data were exact matches to or combinations of the "NEIGHBORHOOD" field in the ARC data set that I used to map statistical area codes to neighborhood names. For example, NSA V03 - Peoplestown in the ARC neighborhood data - was a one-to-one match because Peoplestown was also the name of a neighborhood in the ZHVI data set. In these cases, I simply took the average of all twelve 2010 and all twelve 2018 values for the neighborhood to be that neighborhood's typical home value.

However the neighborhood statistical area (NSA) I01 - according to the ARC data - has the neighborhood name, "Beecher Hills, Florida Heights, Westwood Terrace." In the Zillow data, these were considered three separate "Region Names" or neighborhoods. In these cases, I took a nested average. In this example, that would mean taking the average of each of the three neighborhoods individually (twelve values for 2010 and for 2018 each) and then averaging those three units together for each year.

Response Variables

The methodology described thus far creates the two Y variables to be tested in separate models:

- Percent change of the average typical home value = *(ZHVI_avg_avg_percentchange)*
- Net change of the average typical home value = *(ZHVI_avg_avg_netcash)*

Missing Data Treatment and Excluded NSAs

In joining the ZHVI data set to the City of Atlanta data set, there were two key types of missing data that occurred, resulting in their exclusion from the analysis:

First, there were a number of neighborhoods in the ZHVI data set that did not have values for 2010. The following NSAs - the unit of analysis for the fully joined data set - had one or more ZHVI neighborhoods with no typical home value data in 2010:

- Arlington Estates, Ben Hill, Butner/Tell, Elmco Estates, Fairburn, Fairburn Tell, Fairway Acres, Huntington, Lake Estates, Wildwood Forest

- Audobon Forest, Audobon Forest West, Chalet Woods, Harland Terrace, Peyton Forest, Westhaven
- Kingswood, Mt. Paran/Northside, Mt. Paran Parkway, Randall Mill, West Paces Ferry/Northside, Whitewater Creek
- Baker Hills, Bakers Ferry, Boulder Park, Fairburn Road/Wisteria Lane, Ridgecrest Forest, Wildwood (NPU-H), Wilson Mill Meadows, Wisteria Gardens
- Ben Hill Forest, Ben Hill Pines, Brentwood, Deerwood, Mellwood, Rue Royal, Tampa Park
- Ben Hill Acres, Briar Glen, Cascade Green, Heritage Valley, Meadowbrook Forest, Mt. Gilead Woods
- Lindbergh/Morosgo
- Chosewood Park, Englewood Manor
- South Atlanta, The Villages at Carver
- Adair Park, Pittsburgh
- Capitol View, Capitol View Manor
- Center Hill, Harvel Homes Community
- Hunter Hills, Mozley Park
- Almond Park, Carey Park
- Fairburn Mays, Mays
- Lakewood, Leila Valley, Norwood Manor, Rebel Valley Forest
- Carver Hills, Rockdale, Scotts Crossing, West Highlands
- Blair Villa/Poole Creek, Glenrose Heights, Orchard Knob, Rosedale Heights
- Atlanta Industrial Park, Bolton Hills, Brookview Heights, Chattahoochee, English Park, Lincoln Homes, Monroe Heights
- Bankhead Courts, Bankhead/Bolton, Carroll Heights, Fairburn Heights, Old Gordon
- Amal Heights, Betmar LaVilla, High Point, Joyland
- Browns Mill Park, Polar Rock, Swallow Circle/Baywood
- Thomasville Heights
- Dixie Hills, Penelope Neighbors, West Lake
- Georgia Tech, Marietta Street Artery

The second type of missing data was when one of the neighborhood names in the NSA was not found in the ZHVI data set. For example, in the NSA “Buckhead Heights, Lenox, Ridgedale Park,” neither Lenox nor Ridgedale Park existed as a neighborhood in the ZHVI data and thus I did not want to include potentially skewed typical home values. The full list of NSAs excluded due to similar issues are:

- Ashley Courts, Greenbriar Village, Niskey Cove, Niskey Lake, Sandlewood Estates
- Buckhead Heights, Lenox, Ridgedale Park
- Collier Hills, Collier Hills North, Colonial Homes
- Grant Park, Oakland
- Midwest Cascade, Regency Trace
- East Lake, The Villages at East Lake
- Capitol Gateway, Summerhill
- Ben Hill Terrace, Kings Forest, Old Fairburn Village
- Campbellton Road, Fort Valley, Pomona Park
- Ashview Heights, Harris Chiles, Just Us
- Fort McPherson, Venetian Hills
- Benteen Park, Boulevard Heights, Custer/McDonough/Guice, State Facility, Woodland Hills

Finally, some NSAs were identified as not mappable due to differences in naming conventions between the ARC names imported and the ZHVI neighborhood names, creating duplicate mapping or otherwise potentially inaccurate. These excluded NSAs are:

- Atlanta University Center, The Villages at Castleberry Hill
- Bankhead, Washington Park
- Bolton, Riverside, Whittier Mill Village
- Channing Valley, Memorial Park, Springlake, Wildwood (NPU-C)
- Candler Park, Druid Hills
- **NSA F05** *did not have a name match within the ARC data set.*
- **NSA X05** *is the Airport and does not have any data in either ZHVI or the City of Atlanta data set.*

After excluding these 42 NSAs, only 59 of the 101 (not counting F05 and X05) NSAs - or 58% of the City of Atlanta's NSAs - remain for analysis.

Other Notes on Variables

Age cohorts in the City of Atlanta Neighborhood Change data set outside of the 18-39 subset include:

- Under 18 years
- 40-64 years
- Age 65 years and over

Racial categories in the City of Atlanta Neighborhood Change data set outside of the not Hispanic, white alone subset include:

- Not Hispanic, Black or African American alone
- Not Hispanic, Asian alone
- Not Hispanic, other
- Not Hispanic or Latino (of any race)

In the City of Atlanta Neighborhood Change data set, there is a total housing units variable which is divided into two proportions: *vacant and occupied*. Likewise, the occupied housing units variable is divided into two proportions: *owner occupied and renter occupied.*

In the City of Atlanta Neighborhood Change data set, this is the only variable that looks at educational attainment, but it looks at it in conjunction with age and does not count individuals younger than 25 with bachelor's degrees.

Analysis & Results

Reminder, the below findings - unless otherwise specified - refer to changes between the years 2010 and 2018, using 59 of 103 - or 57% - of the city of Atlanta's neighborhood statistical areas.

When Y = % Change in Typical Home Value

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.448[a]	.201	.109	58.357592%

a. Predictors: (Constant), baplus_percentchange, vacant_percentchange, percentchange_18_39, white_percentchange, owner_occupied_percentchange, med_household_income_percentchange

With an **R value of .448, nearly .5,** there is a *moderate correlation* between the response variables and all of the explanatory variables. However, with an **R Square value of .201**, only 20.1% of the variance in the response variable is explained by the combined explanatory variables. In other words, it only reduces my prediction error by 20.1%.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	90.608	13.432		6.745	<.001	63.654	117.562		
	med_household_income_percentchange	-.375	.398	-.131	-.941	.351	-1.174	.424	.790	1.266
	white_percentchange	.067	.059	.146	1.124	.266	-.053	.186	.914	1.094
	percentchange_18_39	1.223	.545	.285	2.245	.029	.130	2.317	.951	1.051
	owner_occupied_percentchange	-.199	.481	-.057	-.413	.681	-1.163	.766	.821	1.218
	vacant_percentchange	-.197	.202	-.126	-.976	.333	-.602	.208	.925	1.081
	baplus_percentchange	.269	.281	.129	.957	.343	-.294	.831	.841	1.189

a. Dependent Variable: ZHVI_avg_avg_percentchange

For every **percent increase in the change in the proportion of individuals aged 18-39**, the percent change of the average typical home value will ***increase*** by **1.223%** holding all other variables constant. ***This is statistically significant*** at the 95% confidence level as the p-value is **.029**, meaning that if there was no correlation between the **percent increase in the change in the proportion of individuals aged 18-39** and the percent change in average typical home value, there would be a **2.9%** probability of observing the observed level of correlation between these variables in a random sample of this size.

For every **percent increase in the change in proportion of not Hispanic, white alone individuals**, the percent change of the average typical home value will ***increase*** by .**067%** holding all other variables constant. ***This is not statistically significant*** at the 95% confidence level as the p-value is **.266**, meaning that if there was no correlation between the **percent increase in the change in proportion of not Hispanic, white alone individuals** and the percent change in average typical home value, there would be a **26.6%** probability of observing the observed level of correlation between these variables in a random sample of this size.

For every **percent increase in the change in the proportion of vacant housing units**, the percent change of the average typical home value will ***decrease*** by **.197%** holding all other variables constant. ***This is not statistically significant*** at the 95% confidence level as the p-value is **.333**, meaning that if there was no correlation between the **percent increase in the change in the proportion of vacant housing units** and the percent change in average typical home value, there would be a **33.3%** probability of observing the observed level of correlation between these variables in a random sample of this size.

For every **percent increase in the change in the proportion of individuals 25 years or older and with a bachelor's degree or more**, the percent change of the average typical home value will ***increase*** by **.269%** holding all other variables constant. ***This is not statistically significant*** at the 95% confidence level as the p-value is **.343**, meaning that if there was no correlation between the **percent**

increase in the change in the proportion of individuals 25 years or older and with a bachelor's degree or more and the percent change in average typical home value, there would be a **34.3%** probability of observing the observed level of correlation between these variables in a random sample of this size.

For every **percent increase in the change in median household income**, the percent change of the average typical home value will ***decrease*** by .**375%** holding all other variables constant. ***This is not statistically significant*** at the 95% confidence level as the p-value is **.351**, meaning that if there was no correlation between the **percent increase in the change in median household income** and the percent change in average typical home value, there would be a **35.1%** probability of observing the observed level of correlation between these variables in a random sample of this size. According to the literature, it seems unlikely that higher positive rates of change in median household income would decrease the percent change in average typical home value in a neighborhood, so this may be a result of the multivariate model itself or the limited sample of data after excluding 42 NSAs in the city of Atlanta.

For every **percent increase in the change in the proportion of owner-occupied housing units**, the percent change of the average typical home value will ***decrease*** by **.199%** holding all other variables constant. ***This is not statistically significant*** at the 95% confidence level as the p-value is **.681**, meaning that if there was no correlation between the **percent increase in the change in the proportion of owner-occupied housing units** and the percent change in average typical home value, there would be a **68.1%** probability of observing the observed level of correlation between these variables in a random sample of this size.

According to the literature, it seems unlikely that higher positive rates of change in median household income or proportion of owner-occupied housing units would decrease the percent change in average typical home value in a neighborhood, so this may be a result of the multivariate model itself or the limited sample of data after excluding 42 NSAs in the city of Atlanta. However, this also means that even trends that confirm my understanding of the literature are merely reflections of this incomplete data set and the distribution of each neighborhood's explanatory social variables therein.

		Y = % Change in Typical Home Value						
		ZHVI_avg_avg_percentchange	med_household_income_percentchange	white_percentchange	percentchange_18_39	owner_occupied_percentchange	vacant_percentchange	baplus_percentchange
Pearson Correlation	ZHVI_avg_avg_percentchange	1.000	-.143	.246	.345	-.123	-.205	.117
	med_household_income_percentchange	-.143	1.000	-.122	-.056	.345	.017	.336
	white_percentchange	.246	-.122	1.000	.148	-.152	-.156	.109
	percentchange_18_39	.345	-.056	.148	1.000	-.009	-.144	.091
	owner_occupied_percentchange	-.123	.345	-.152	-.009	1.000	.194	.218
	vacant_percentchange	-.205	.017	-.156	-.144	.194	1.000	-.020
	baplus_percentchange	.117	.336	.109	.091	.218	-.020	1.000
Sig. (1-tailed)	ZHVI_avg_avg_percentchange	.	.140	.030	.004	.177	.059	.188
	med_household_income_percentchange	.140	.	.178	.337	.004	.450	.005
	white_percentchange	.030	.178	.	.132	.126	.119	.205
	percentchange_18_39	.004	.337	.132	.	.474	.138	.247
	owner_occupied_percentchange	.177	.004	.126	.474	.	.070	.049
	vacant_percentchange	.059	.450	.119	.138	.070	.	.440
	baplus_percentchange	.188	.005	.205	.247	.049	.440	.

From running this multivariate regression, I do not have the regression function for each explanatory variable against the response variable individually, but I do know the strength of their individual correlation and their statistical significance:

All variables share a weak correlation (R value between 0 and +- 0.3) with the percent change in typical home value ***except for the percent change of residents aged 18 to 39***, which shares a positive moderately strong correlation with an *R value of .345*. This means the higher the proportion of residents aged 18 to 39, the more likely there will be a higher percent change in typical home value associated with it.

This age cohort, with a p-value of .004, ***and the percent of residents who are white*** are the only other statistically significant relationships, meaning I can be 95% confident that they are correlated beyond my sample of data.

When Y = $ Change in Typical Home Value

Upon this regression's first attempt, Ansley Park / Sherwood Forest was flagged as an outlier, having a net cash change in typical home value that was larger than 3 standard deviations from the rest of the data. It was removed for the following model:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.326[a]	.106	.090	85730.3643	.106	6.649	1	56	.013

a. Predictors: (Constant), med_household_income_percentchange

b. Dependent Variable: ZHVI_avg_avg_netcash

With an **R value of .326,** there is a *weaker, but still moderate correlation* between the response variables and all of the explanatory variables than the previous model. With an **R Square value of .106**, only 10.6% of the variance in the response variable is explained by the combined explanatory variables. In other words, it only reduces my prediction error by 10.6%.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	186086.591	20189.115		9.217	<.001	145574.167	226599.015		
	med_household_income_percentchange	1063.657	598.455	.235	1.777	.081	-137.231	2264.544	.790	1.266
	white_percentchange	-83.948	89.409	-.115	-.939	.352	-263.360	95.464	.914	1.094
	percentchange_18_39	-1362.364	819.007	-.200	-1.663	.102	-3005.822	281.094	.951	1.051
	owner_occupied_percentchange	1372.460	722.455	.246	1.900	.063	-77.253	2822.173	.821	1.218
	vacant_percentchange	81.707	303.549	.033	.269	.789	-527.409	690.824	.925	1.081
	baplus_percentchange	-1041.363	421.619	-.317	-2.470	.017	-1887.403	-195.323	.841	1.189

a. Dependent Variable: ZHVI_avg_avg_netcash

For every **percent increase in the change in the proportion of individuals 25 years or older and with a bachelor's degree or more**, the net change of the average typical home value will ***decrease*** by **$1041.36** holding all other variables constant. ***This is statistically significant*** at the 95% confidence level as the p-value is **.017**, meaning there is a **1.7%** probability of the values observed in my data set happening by chance.

For every **percent increase in the change in the proportion of owner-occupied housing units**, the net change of the average typical home value will ***increase*** by **$1,372.46** holding all other variables constant. ***This is not statistically significant*** at the 95% confidence level as the p-value is **.063**, meaning that if there was no correlation between the **percent increase in the change in the proportion of owner-occupied housing units** and the percent change in average typical home value, there would be a **6.3%** probability of observing the observed level of correlation between these variables in a random sample of this size.

For every **percent increase in the change in median household income**, the net change of the average typical home value will ***increase*** by **$1,063.66** holding all other variables constant. ***This is not statistically significant*** at the 95% confidence level as the p-value is **.081**, meaning that if there was no correlation between the **percent increase in the change in median household income** and the percent change in average typical home value, there would be a **8.1%** probability of observing the observed level of correlation between these variables in a random sample of this size.

For every **percent increase in the change in the proportion of individuals aged 18-39**, the net change of the average typical home value will ***decrease*** by $**1,362.36** holding all other variables constant. ***This is not statistically significant*** at the 95% confidence level as the p-value is **.102**, meaning that if there was no correlation between the **percent increase in the change in the proportion of individuals aged 18-39** and the percent change in average typical home value, there would be a **10.2%** probability of observing the observed level of correlation between these variables in a random sample of this size.

For every **percent increase in the proportion of not Hispanic, white alone individuals**, the net change of the average typical home value will ***decrease*** by **$83.95** holding all other variables constant. ***This is not statistically significant*** at the 95% confidence level as the p-value is **.352**, meaning that if there was no correlation between the **percent increase in the proportion of not Hispanic, white alone individuals** and the percent change in average typical home value, there would be a **35.2%** probability of observing the observed level of correlation between these variables in a random sample of this size.

For every **percent increase in the change in the proportion of vacant housing units**, the net change of the average typical home value will ***increase*** by **$81.71** holding all other variables constant. ***This is not statistically significant*** at the 95% confidence level as the p-value is **.789**, meaning that if there was no correlation between the **percent increase in the change in the proportion of vacant housing units** and the percent change in average typical home value, there would be a **78.9%** probability of observing the observed level of correlation between these variables in a random sample of this size.

Again, there are strange findings that either contradict the literature or contradict the other model. It seems unlikely that higher positive rates of change in proportion of vacant housing units would increase the net cash change in average typical home value in a neighborhood, even if it’s not statistically significant. Therefore, this requires an exploration into what neighborhoods see higher or lower net cash changes and percent changes in average typical home values.

Y = $ Change in Typical Home Value

		ZHVI_avg_avg_netcash	med_household_income_percentchange	white_percentchange	percentchange_18_39	owner_occupied_percentchange	vacant_percentchange	baplus_percentchange
Pearson Correlation	ZHVI_avg_avg_netcash	1.000	.240	-.251	-.266	.284	.138	-.215
	med_household_income_percentchange	.240	1.000	-.122	-.056	.345	.017	.336
	white_percentchange	-.251	-.122	1.000	.148	-.152	-.156	.109
	percentchange_18_39	-.266	-.056	.148	1.000	-.009	-.144	.091
	owner_occupied_percentchange	.284	.345	-.152	-.009	1.000	.194	.218
	vacant_percentchange	.138	.017	-.156	-.144	.194	1.000	-.020
	baplus_percentchange	-.215	.336	.109	.091	.218	-.020	1.000
Sig. (1-tailed)	ZHVI_avg_avg_netcash	.	.034	.028	.021	.015	.149	.051
	med_household_income_percentchange	.034	.	.178	.337	.004	.450	.005
	white_percentchange	.028	.178	.	.132	.126	.119	.205
	percentchange_18_39	.021	.337	.132	.	.474	.138	.247
	owner_occupied_percentchange	.015	.004	.126	.474	.	.070	.049
	vacant_percentchange	.149	.450	.119	.138	.070	.	.440
	baplus_percentchange	.051	.005	.205	.247	.049	.440	.

All variables share a weak correlation (R value between 0 and +- 0.3) with the percent change in typical home value. Multiple explanatory variables are statistically significant with a p-values of less than 0.05. Notably ***the percent change of residents 18 to 39 suddenly has a negative impact on the response variable, as does the percentage change of residents who are white.***

Discrepancies Between Models; A Look at Underlying Data

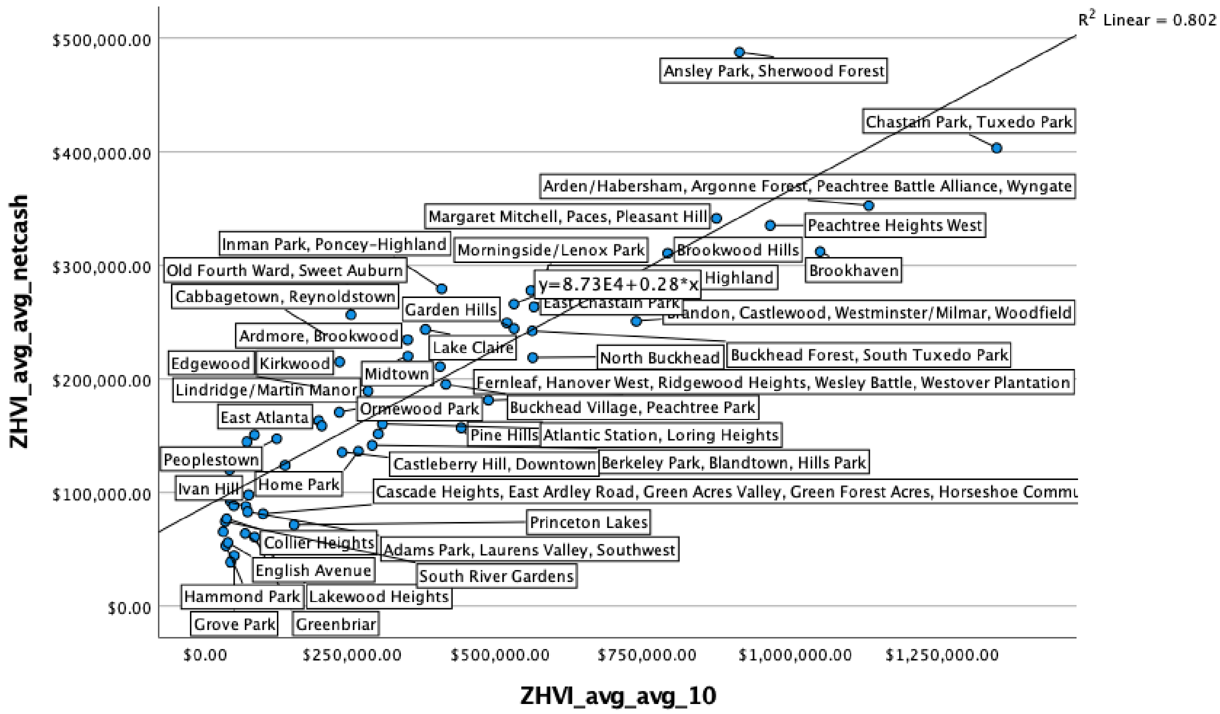
In the following bivariate regression model, the average typical home value of a neighborhood in 2010 was the explanatory variable (x) and the net cash change in average typical home value from 2010 to 2018 is the response variable (y).

Model Summary[b]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.896[a]	.802	.799	43981.6546	.802	230.853	1	57	<.001

a. Predictors: (Constant), ZHVI_avg_avg_10
b. Dependent Variable: ZHVI_avg_avg_netcash

Glancing at the data and which neighborhoods had higher net cash changes, I noticed one of the limitations of using net cash change in average typical home value as a response variable in the previous models is that it's highly correlated with neighborhoods that start out in 2010 as already expensive. **With an extremely strong correlation (R=.896) and 80.2% of the variance in net cash change of average typical home value by neighborhood can be explained by the initial average typical home value in 2010. In other words, it vastly reduces my prediction error by 80.2%.** Common sense does tell me that more expensive homes are more likely to have higher net differences after multiple years, and this data shows that to be true.



According to the linear regression model, for every **dollar increase in the average typical home value in a neighborhood in 2010**, the net cash change of the average typical home value is expected to ***increase*** by **$0.28**

In the following bivariate regression model, the average typical home value of a neighborhood in 2010 was the explanatory variable (x) and the percent change in average typical home value from 2010 to 2018 is the response variable (y).

Model Summary[b]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.667[a]	.445	.435	46.469559%	.445	45.633	1	57	<.001

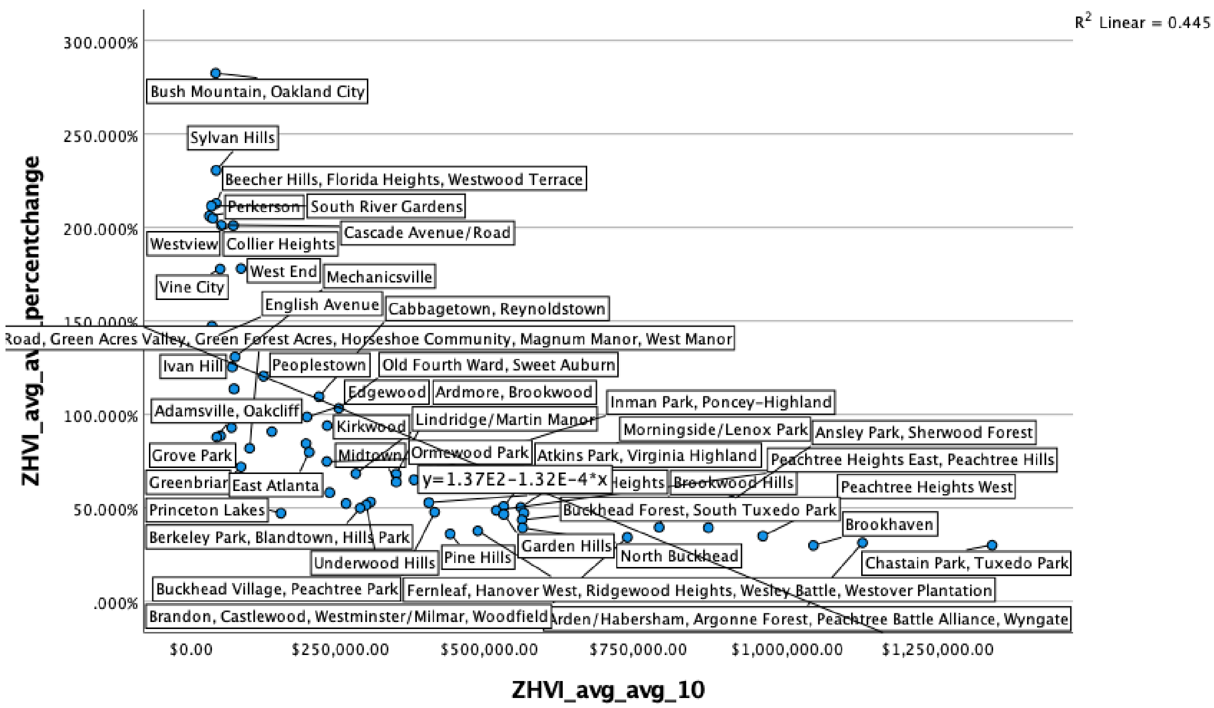
a. Predictors: (Constant), ZHVI_avg_avg_10

b. Dependent Variable: ZHVI_avg_avg_percentchange

Conversely, the typical home price in 2010 does have a strong negative correlation with the percent change in typical home value. While not quite as strong as the R value of .896 above, **there is a strong correlation (R=.667) between the typical home price in 2010 and the percent change in typical home value. 44.5% of the variance in percent change of average typical home value by neighborhood can be explained by the initial average typical home value in 2010. In other words, it**

somewhat reduces my prediction error by 44.5%. Meaning that less expensive homes experienced greater relative change in their home value than more expensive ones. It does nearly appear that the neighborhoods names in each graph have flipped:

Graph



According to the linear regression model, for every **dollar increase in the average typical home value in a neighborhood in 2010**, the percent change of the average typical home value is expected to ***decrease*** by **0.000132%**

Conclusion, Limitations, & Suggested Next Steps of Research

Many of the variables I expected to have a relationship with changes in typical home values in the city of Atlanta between 2010 and 2018 - whether percent change or net change - did not show any correlation. Proportion of owner-occupied housing units, vacant housing units, and median income did not seem to correlate in either model, nor as bivariate regressions against the response variable.

While the multivariate model using the percent change in typical home value of a neighborhood as the response variable was comparatively stronger, there are a few other things that I know: in this model, the age-cohort of 18 to 39 was the only statistically significant explanatory variable both within the model when controlling for other variables *and* when paired individually with the response variable.

This indicates moderate correlation between percentage increases in proportion of residents age 18-39 and percentage increases in a neighborhood's typical home value.

The second model - when contextualized with the discrepancies in the base data - also shows me interesting trends with respect to age, education, and race. When controlling for other variables, every percent increase in the change in the proportion of residents 25 years or older and with a bachelor's degree or more is expected to be correlated with the net change of the average typical home value in that neighborhood decreasing by $1041.36. This could alternatively be understood as over time, individuals who are 25 years or older with a bachelor's or more may not be moving to these neighborhoods in the first place - which I now know were often already more expensive to begin with. Outside of the multivariate regression model, both an increase in white residents and an increase in residents 18 to 39 were also shown to be associated with decreases in the net cash difference in typical home value between 2010 and 2018.

This could possibly have traces in the relationship between age cohort of 18 to 39 and percent increase and some of Holleran's YIMBY scholarship - young, educated, white millennials may be more likely to move to neighborhoods that have a higher percentage increase in typical home value between 2010 and 2018, which I now know had significantly lower base prices in 2010.

To test this theory, subsequent research can be done on the underlying data from this analysis. Knowing the geography of the city, many of these neighborhoods with percentage increases in typical home value are south and west and from glancing at the data, tend to have had typical home prices of roughly $40,000 - $100,000 in 2010 and closer to $200,000+ in 2018. Meanwhile the ones with higher net cash differences are north and east and often have typical home values around $500,000-600,000 and $800,000+ in 2010 and 2018 respectively. Analysis can be done to identify what types of neighborhoods are having a higher increase in young, educated, white residents. Do they start with a higher population of Black residents, lower median incomes, or higher vacancies? Perhaps the Zillow data did not help us learn anything new but rather, stand as a form to test my understanding.

Additional discrepancies between the two initial models may be a result of the largely white and affluent Ansley Park/Sherwood Forest neighborhood being included in the first, and excluded as an outlier from the second. Further analysis on this neighborhood's exact values for its social variables and its ZHVI values would help determine the neighborhood's inclusion/exclusion bearing on the model.

Lastly, the most glaring shortcoming of this analysis is the fact that it only utilizes barely over half of the neighborhood statistical areas due to complications joining it with Zillow's data set. With this in mind, ***the exclusion of 42 NSAs negatively impacts my ability to make inferences about the entire city of Atlanta from both this analysis and this data set.*** To understand the implications of these exclusions, analysis of the social components of both the included and excluded subsets can be done separately to see if the two samples were significantly different in key social variables like the X variables in this analysis. Additionally, they can be utilized and no longer excluded in any analysis that does not include the Zillow data set. Looking primarily at the variables already provided in the City of Atlanta Neighborhood Change data set would be the most logical next step to continue to understand how displacement is happening in Atlanta.

Bibliography

Atlanta Regional Commission (2014). City of Atlanta Neighborhood Statistical Areas. *Open Data Georgia Association of Regional Commissions*. Retrieved November 13, 2023, from https://opendata.atlantaregional.com/datasets/GARC::city-of-atlanta-neighborhood-statistical-areas/about

City of Atlanta (2021). Neighborhood Change Report. *City Planning Office of Housing and Community*

Development. Retrieved October 30, 2023, from https://www.atlantaga.gov/government/departments/city-planning/office-of-housing-community-development/neighborhood-change-report

Holleran, M. (2021). Millenial 'YIMBYs' and boomer 'NIMBYs': Generational views on housing affordability in the United States, Sociological Review, 69(4), 846-861.

Junia Howell, Elizabeth Korver-Glenn, The Increasing Effect of Neighborhood Racial Composition on Housing Values, 1980–2015, Social Problems, Volume 68, Issue 4, November 2021, Pages 1051–1071, https://doi.org/10.1093/socpro/spaa033

Kapos, S., J. P. JR., Moore, N. Y., & Bey, L. (2021, December 7). *Black people are leaving Chicago en masse. it's changing the city's power politics.* POLITICO. https://www.politico.com/news/magazine/2021/12/07/chicago-black-population-decline-523563

Lanari, E. (2019). Envisioning a new city center: Time, displacement, and Atlanta's suburban futures. City & Society, 31(3), 365-391.

Lauermann, J. (2022). Luxury housing and gentrification in New York City, 2010-2019, *Urban Geography*, 43:10, 1683-1701

Slater, T. (2021). 'Gentrification beyond false choice urbanism' in Shaking Up the City: Ignorance, Inequality, and the Urban Question (pp. 52-83) Berkeley: University of California Press.

Stehlin, J. (2019). Cyclescapes of the Unequal City: Bicycle Infrastructure and Uneven Development. *University of Minnesota Press.*

Valli, C. (2015). A sense of displacement: Long-time residents' feelings of displacement in gentrifying Bushwick, New York. International Journal of Urban and Regional Research, 39(6), 1191-1208.

Zillow. Zillow Home Value Index (ZHVI), Single-Family Homes Time Series ($), Neighborhood. *Zillow Housing Data*. Retrieved October 30, 2023, from https://www.zillow.com/research/data/