How the marketing of real estate properties explains mortgage applicants by race and income

Isabelle Nilsson[a]* and Elizabeth C. Delmelle[b]

*[a]Department of Geography & Earth Sciences, University of North Carolina at Charlotte, Charlotte, USA, inilsso1@charlotte.edu; [b]Department of City and Regional Planning, University of Pennsylvania, Philadelphia, USA; delmelle@design.upenn.edu; *corresponding author*

In this article, we study how the marketing of single-family homes explains the racial and income makeup of mortgage applicants in a neighborhood. We use a case study of the robust housing market of Charlotte, North Carolina, and annual, longitudinal real estate listing advertisements alongside mortgage lending data, to demonstrate how the share of properties advertised a certain way in a neighborhood in one year explains shares of mortgage applicants by race and income the following year. We classify property advertisement text using a semi-supervised learning algorithm into five categories following a housing investment, disinvestment to renewal continuum. We find stark racial disparities in mortgage applicants by housing type, even after controlling for income. We find that Black applicants nearly exclusively apply for mortgages in neighborhoods with a high share of properties advertised as disinvested with little profit-making promise. High-income White applicants rise as the share of advertised properties becomes more homogenous.

Keywords: discrimination, housing lifecycle, neighborhood change, mortgage lending, semi-supervised learning.

1. Introduction

The real estate industry – including the actions of agents, mortgage lending professionals, housing developers, and appraisers – has played a significant role in shaping the trajectories of neighbourhoods throughout the United States (Besbris, 2020a; Korver-Glenn, 2018, 2021; Mallach, 2024; Slater, 2021). Before The Fair Housing Act of 1968 prohibited discriminatory practices in the real estate market, real estate professionals used legal tactics to preclude Black homebuyers from purchasing homes in White neighborhoods (Korver-Glenn, 2021; Mallach, 2024). Despite the gradual dismantling of legal discrimination in the housing and mortgage markets and the increasing diversity of the United States population, racial and income segregation remains persistently high, especially for African Americans (Massey, 2020).

Previous research has demonstrated the more subtle ways that discrimination continues to percolate through the housing market including audits of realtors documenting the continued practice of steering prospective homebuyers of different races to different neighbourhoods (Christensen & Timmins, 2022; Galster & Godfrey, 2005), showing and recommending fewer homes to minorities than White home seekers (Turner et al., 2016), and persistent racial disparities in mortgage lending practices (Quillian et al., 2020). Important recent scholarship on this topic has used detailed qualitative methods to understand the intricacies behind the actions of real estate professionals that perpetually reproduce racially segregated neighborhoods (Besbris, 2020b; Korver-Glenn, 2021). Our contribution is to take a systematic, longitudinal approach that connects the marketing of real estate properties in neighborhoods with the race and income composition of mortgage applicants. We capitalize on an annual database of properties from the Multiple Listing Service (MLS) from 2001 to 2021 and annually updated mortgage application data from the Home Mortgage Disclosure Act (HMDA).

We begin by classifying property advertisement text into five categories following a housing investment-disinvestment continuum using a semi-supervised text classification algorithm. We then analyze the shares of these five classes of homes for sale in a census tract in one year, linking them to the shares of mortgage applicants by race, income, and the intersection of race and income. We examine to what extent the makeup of the types of homes for sale in a neighborhood is associated with the race and income of mortgage applicants. We hypothesize that the composition of housing types in a neighborhood provides a signal about the risk of the investment. Finally, we explore threshold effects or tipping points that might indicate a non-linear relationship between mortgage applicants and the composition of advertised homes in a neighborhood.

Our analysis reveals stark racial differences in homebuyers across our housing categories, even among similar income groups. We find that Black homebuyers overwhelmingly apply for mortgages in neighborhoods where homes are advertised within our 'disinvested' class—properties with no promise of profit-making and few desirable amenities listed. As the concentration of disinvested homes in a neighborhood increases, so does the share of Black applicants. Conversely, homes classified as expensive investments are positively associated with the share of White applicants. Notably, upper-income Black and White applicants display distinct patterns: Blacks avoid neighborhoods with a high concentration of expensive investments, preferring neighborhoods with a high share of disinvested housing and, to a lesser extent, New Suburban housing.

Our article proceeds with a review of the real estate industry's role in shaping neighborhood outcomes and how the marketing of properties is racially uneven, aiding in the reproduction of patterns of racial segregation. We then discuss our empirical study including an overview of our text classification procedure and estimation framework. Finally, we present and discuss our findings.

2. Literature review

2.1 The Real Estate's Role in Neighborhood Change

Housing developers, real estate agents, mortgage lenders, and home appraisers have all served as gatekeepers to shape the perpetually racially unequal housing markets of US Cities (Korver-Glenn, 2018). Many areas of concentrated poverty and non-Whites in today's cities are the enduring product of discriminatory practices in the housing market. From its onset, The National Association of Realtors operated under the belief that neighborhood racial homogeneity was crucial for preserving property values and housing conditions (Slater, 2021). They used racial covenants to restrict non-Whites from purchasing homes in certain neighborhoods and practiced racial steering when deciding which properties to show clients.

Decisions by mortgage bankers to not lend in certain neighborhoods because of their housing age, residential incomes, or racial compositions, commonly known as redlining, restricted credit flows to minority neighborhoods. During the post-war housing boom, entire neighborhoods were denied mortgage and home improvement loans backed by the Federal Housing Administration (FHA) because they were deemed too risky, as indicated by aging housing and racial makeup of a neighborhood. Eventually, FHA was reformed, and lending in inner cities was encouraged, resulting in FHA loans becoming synonymous with racial change (Bradford 1979). Realtors capitalized on this change and convinced Whites to quickly sell their homes at a discount, suggesting that racial changes were immanent. They, in turn, sold those properties to Blacks at inflated prices using FHA-insured loans - a practice commonly referred to as “blockbusting” (Bradford 1979).

More recently, many minority neighborhoods were impacted by predatory lending – a form of reverse redlining (Haupert, 2019). Subprime lending originated in the 1990s by lending to those with poor credit, but after 2000, these types of higher fee, higher interest, or adjustable interest rate mortgages were expanded to middle-class borrowers in rapidly growing parts of the country (Aalbers 2009). Predatory loans designed to exploit vulnerable, and disproportionately low-income and minority home borrowers ballooned during this time, ultimately leading to high rates of foreclosures and subsequent declining home values (Immergluck, 2009). Lower and middle-class Blacks and were more likely to be targeted for subprime loans and have concentrated rates of foreclosures (Rugh & Massey 2010). In turn, these foreclosed properties have increasingly been targeted by investors and corporate landlords who purchase large quantities and transform owner-occupied dwellings into rentals (Black et al. 2011).

Finally, the role of racism more broadly was a contributing factor in restricting the residential choices of Blacks to a few select neighborhoods. Physical violence against Black families and their homes and against the realtors who showed or sold homes to Blacks in White neighborhoods were common tactics used (Drake & Cayton, 1945; Pattillo, 2005).

On the flip side of neighborhood decline, gentrification, broadly defined as a change in the social class of residents, can also be viewed as a racialized process (Rucks-Ahidiana, 2021b). Gentrification occurs unevenly across cities depending on a neighborhood’s racial composition and history of disinvestment. Real estate professionals often perceive non-White neighborhoods as poor or risky investments (Rucks-Ahidiana, 2021b; Taylor, 2019). Consequently, they are less likely to undergo gentrification compared to White neighborhoods (Delmelle, 2016; Hwang & Ding, 2020; Rucks-Ahidiana, 2021a) and when they do, the resulting income changes are less than in the case of gentrification in White neighborhoods (Rucks-Ahidiana, 2021a). Thus, this process is arguably not strictly a calculus of profit-making potential and locational advantage, but one that is melded by neighborhood valuation, considering its racial makeup, history, and other contextual factors by homebuyers and real estate professionals. When real estate agents see potential value in a neighborhood, they can use levers of advertisement and marketing to accelerate the process. In very tight housing markets, the likelihood that a low-income Black neighborhood will be re-evaluated as a profit-making space increases (Rucks-Ahidiana, 2021b).

2.2 Real Estate Advertisements, Signals, and Sorting

Real estate agents are intermediaries between the evolving tastes and preferences of home seekers and available properties (Bridge, 2001; Perkins et al., 2008). They aid in the production of gentrifiable spaces by connecting working-class homes with middle-class buyers (Bridge, 2001). Early-stage gentrifiers differ from later-stage buyers, as they assume more risk for the potential of higher financial rewards, often purchasing at lower prices compared to homes in more advanced stages of gentrification. These buyers may value neighborhood diversity or the 'authentic' culture of the area differently (Rucks-Ahidiana 2021b). As initial homes are purchased and redeveloped, the risks of buying property in a transitioning neighborhood decrease. Eventually, the area becomes attractive to larger-scale developers interested in constructing high-end housing or condominiums (Bridge, 2001; Skaburskis, 2010). From a marketing perspective, realtors may highlight the potential opportunities in a location to prospective middle-class or white home buyers, hoping they will initiate the process (Perkins et al., 2008).

The number of properties in a neighborhood marketed as ‘opportunities’ for investment can signal the risk of purchasing a home there. In disadvantaged neighborhoods, a few such properties might indicate the earliest stages of gentrification and thus a higher risk. Similarly, if all homes are advertised as opportunities but none have undergone renovation or revitalization, this may also indicate a higher risk. These properties are unlikely to attract the wealthiest buyers (Bridges, 2001). Likewise, in new single-family developments or subdivisions, the earliest purchasers take on more risk than later homebuyers when the subdivision’s market has been established and the remaining home and neighborhood characteristics are solidified (Hollans et al., 2012).

2.3 Residential Search and Sorting

There has been some debate in the literature as to the extent to which persistent patterns of racial segregation are entirely the result of continued discriminatory practices in the residential selection process or a reflection of differences in racial preferences for neighborhoods. Early work generally concluded that Whites have a strong preference for living in predominantly White neighborhoods (Clark, 2008; Krysan et al., 2009; Krysan & Farley, 2002) and a negative preference as the share of Blacks in a neighborhood increases, even after controlling for correlated neighborhood factors like school quality, crime levels, and housing values (Emerson et al., 2001). Blacks, on the other hand, prefer racially mixed neighborhoods (Krysan et al., 2009; Krysan & Farley, 2002).

Recent scholarship has suggested that racial steering, rather than preference explains the disproportionate number of minorities living in poor neighborhoods. While Blacks are shown properties similar in number to Whites, their quality is lower in terms of school ratings, pollution, poverty, and education levels (Christensen & Timmins, 2022). Despite this, preference arguably still plays a role in segregation, as high-income Black households often choose low-income neighborhoods with some Black residents over high-income White neighborhoods (Aliprantis et al., 2024). This is compounded by the limited socioeconomic diversity among majority Black neighborhoods. This finding can be contextualized by the work of Pattillo (2013) who describes how middle and upper-class Black gentrifiers use their resources to improve disinvested neighborhoods, viewing their return to the Black ghetto as a "racial uplift project" (p. 301).

Housing choices vary by income group, with higher-income families making deliberate decisions based on preferences, while lower-income moves are often involuntary, addressing immediate needs with limited resources (Harvey et al., 2020). Consequently, lower-income and minority buyers often purchase lower-quality homes in less desirable or declining neighborhoods, frequently requiring expensive repairs (Van Zandt, 2007; Van Zandt & Rohe, 2011). These homes are typically older, cheaply built, inner-ring suburban properties lacking the advantages of city center locations or newer suburban areas (Hanlon, 2009; Lichter et al., 2023). In fast-growth cities, new developments on former industrial sites were marketed to minority and low-income households as an affordable path to homeownership during housing booms (Currie & Sorensen, 2019; Sorensen et al., 2014).

The concept of shopping externalities suggests that the characteristics of surrounding home listings signal current home seekers or developers. During an active buyer search, the number of surrounding listings affects home prices due to competition (Turnbull & Dombrow, 2006). Our analysis extends this idea, proposing that property marketing also signals future sorting behaviors, influencing neighborhood racial and income changes.

3. Data and methodological approach

To understand how property marketing explains neighborhood mortgage applicant characteristics, we classify real estate listings based on their descriptions using semi-supervised text classification. We create five housing categories along a continuum from investment to disinvestment and revitalization. We calculate the share of these housing types by census tract and link this with Home Mortgage Disclosure Act (HMDA) data on mortgage applicants. Through exploratory analysis, we examine patterns of housing types and mortgage applicant characteristics over time. We then estimate models using the share of specific mortgage applicant characteristics as the dependent variable and the share of homes in each housing category as independent variables, identifying thresholds where housing type share significantly explains mortgage applicant types.

Our analysis features a case study on the high-demand housing market of Charlotte, North Carolina, and its encompassing county, Mecklenburg. Charlotte was the fourth fastest-growing city among the 50 largest cities in the US between 2010 and 2020 (Frey, 2021). Like other fast-growing cities, Charlotte has confronted growing pains with increasing home prices, gentrification pressures in more urban neighborhoods, and the suburbanization of poverty (City of Charlotte, 2021; Delmelle et al., 2021a; Nilsson & Delmelle, 2023b; Smith & Graves, 2005).

The census tract, used as our neighborhood proxy, is the smallest geography for HMDA data. However, tracts have limitations: they can be too large or heterogeneous, masking significant differences (Sperling, 2012), or too small to represent a neighborhood, needing aggregation (Logan, 2018). In Charlotte, neighborhood profile areas derived from block groups are used for planning (City of Charlotte, 2024), meaning tracts may be too large and mask internal diversity.

3.1 Data

Data on advertisements (public remarks) comes from CoreLogic's MLS records, including 158,253 geocoded single-family home listings from 2001 to 2020 in Mecklenburg County, NC. The public remarks are cleaned of misspellings and abbreviations used by realtors (e.g., 'brm') are translated into their proper terms (e.g., bedroom) using Nowak et al.'s (2021) real estate dictionary. To avoid using specific spatial information contained in the public remarks that may influence the classification process, we set up a word list with local neighborhood and town names and we replaced these place-specific names with the generic placeholders of 'neighborhoodname' and 'townname'. Finally, to reduce the dimensionality of the data and focus on terms with high information value, we remove high-frequency stop words from the public remarks (e.g., 'this', 'and', 'the').

To understand homebuyers in different neighborhoods, we use Home Mortgage Disclosure Act (HMDA) data from 2001 to 2021. This annual data has been used to study and predict neighborhood change and includes information on the race and income of mortgage applicants, loan types, and census tracts (Delmelle & Nilsson, 2021; Delmelle et al., 2021b; Galster & Tatian, 2009; Nilsson & Delmelle, 2023b). Due to temporal inconsistencies in property and loan type coding, we focus on single-family home purchase loan originations, the majority of transactions in the county. Applications with missing information, validity, or quality edit failures are removed.

Mortgage lending applicants are classified by income as low, moderate, middle, or upper based on the Community Reinvestment Act (CRA) definitions (e-Cfr 2023). Low-income is less than 50% of the area median income, moderate is 50%-80%, middle is 80%-120%, and upper is 120% or higher. Applicant data by income and race is aggregated annually to the census tract level and interpolated to 2010 boundaries using crosswalks from the Longitudinal Tract Database (Logan et al., 2012). After removing missing observations, complete data is available for 212 of 233 tracts in Mecklenburg County.

3.2 Classification of real estate listings

Our housing classification procedure is developed and described in detail in Nilsson and Delmelle (2023a), but we recap it briefly here. We classify homes into five groups according to their stage in the housing lifecycle using the words of the advertisement contained in the public remarks (see Table 1). We use a semi-supervised text classification algorithm, Lbl2Vec (Schopf et al., 2021), which enables us to devise more theoretically grounded classes compared to unsupervised classification methods. The algorithm learns jointly embedded document and word vectors to classify documents (listings) based on how semantically similar they are to a set of user-defined classes described by provided keywords (Schopf et al., 2021).

The five housing classes—disinvestment, opportunity, renewed, new suburban, and expensive investment—reflect the response of housing supply to demand, encompassing renovation, new construction, or lack of reinvestment as homes age (Galster, 1996). For the semi-supervised algorithm, initial keywords were provided for each class, allowing the algorithm to learn similar words (see Table 1). Disinvestment is characterized by terms like 'foreclosure' and pertains to homes eligible for short-sale or backed by FHA or Fannie Mae's 'homepath' program. Disinvested homes may deteriorate further, present an opportunity for reinvestment, or be replaced. 'Fixer-upper', 'potential', and 'flip' typify homes marketed for reinvestment. 'Location' is often pivotal in distinguishing between continued disinvestment and opportunities. Renewed homes have undergone renovation or remodeling, often found in urban, walkable, and historic neighborhoods. The new suburban class represents typical US suburban housing with uniform, sprawling developments, often with HOAs and family-oriented amenities.

Depending on location attractiveness, new construction or reinvestment may involve expensive custom homes, designed with specific architecture and luxury brands. 'Neighborhoodname' serves as a placeholder for popular neighborhoods in both the renewed and expensive investment classes, reflecting their strong reinvestment predictors (Delmelle & Nilsson, 2021). For further insights into keyword selection and algorithm parameters, refer to Nilsson and Delmelle (2023a).

Table 1. Classes and keywords.

Class name	Keywords	Excerpt from sample remark
Disinvestment	addendum alloffer auction bank fha foreclosure hud homepath homesteps fnma fhmc mac preforeclosure shortsale subjectto	"FHA insured financing not available. Home sold as-is. HUD will not make repairs"
Opportunity	location imagination investment flip investor potential fixer builder	"Fixer upper! Great opportunity to enjoy this fantastic neighborhood at an unprecedented price! Instant equity for the ambitious homeowner or smart investor […]"
Renewed	renovated walkable neighborhoodname historic remodeled renovation location	"Historic [neighborhoodname] home completely renovated in 2012, home was taken down to the studs […]"
New suburban	school towninside amenities pool tennis playground elementary clubhouse	"[…] Amazing community amenities! Clubhouse, outdoor pool, tennis courts, pocket parks,

		playgrounds. Easy commute to Charlotte. Walk to [name] Elementary School”
Expensive investment	suite custom private exquisite neighborhoodname luxury luxurious architecture design pool	“Custom home, built by designer-builder for personal residence and impeccably maintained by its current owners. This incredible property includes countless upgrades and detail to architectural integrity […] Spacious owners suite features tray ceiling, bay window, sitting area and luxurious bath […] Welcome to [neighborhoodname]!”

3.3 Estimating Thresholds

We draw on Quercia and Galster’s (2000) review of thresholds and neighborhood change to develop our empirical approach. Following their review, a threshold effect can be operationalized by segmenting the independent variable into ranges and then creating categorical dummy variables indicating whether the observation has values within a range (1) or not (0). The coefficient for each dummy variable (range) is the average level of the dependent variable within that range. It appears graphically as a “step” function, an intuitive way of interpreting the results. A threshold is depicted by a large step or difference in coefficient values between adjacent categorical dummies.

To determine how the share of homes in a neighborhood marketed as being in various stages of the housing lifecycle is associated with the race and income of loan applicants in the neighborhood, we estimate the following model:

$$Y_{i,t} = \alpha + X_{i,t-1}\beta + Z_{i,t-1}\omega + u_{i,t} \tag{1}$$

where $Y_{i,t}$ is the share of mortgage applicants belonging to a racial (or income) group of interest in tract *i* in year *t*. $X_{i,t-1}$ is a vector of covariates that includes the shares of low-, moderate- and middle-income applicants (leaving out upper-income applicants to avoid multicollinearity) in year *t-1* when $Y_{i,t}$ is the change in the racial group, and the share of black applicants when $Y_{i,t}$ is the change in the income group. $u_{i,t}$ is an error term.

$Z_{i,t-1}$ is a vector of ordered categorical variables indicating the share of homes for sale classified as disinvestment, opportunity, new suburban, and renewed, respectively, in year *t-1*. To test whether there is a threshold effect, the shares of homes classified into the different classes were divided into five percentile levels: 31-50, 51-70, 71-90, 91-100. Neighborhoods with the lowest shares, below the 30th percentile, of each respective class, were used as reference categories. The choice of these percentiles was determined following an examination of the distribution of the different housing classes over time, where highly spatially concentrated classes such as expensive investment and disinvestment had very skewed distributions with many neighborhoods having zero classified listings, resulting in the 10th and 20th percentile being zero in many years. Table 2 provides descriptive statistics for the entire panel dataset which consists of 212 tracts over the period 2001-2020 (housing type shares) and 2001-2021 (mortgage lending applicant shares). Note that year-specific percentile levels for the regressions were calculated to reflect changes in the county-wide housing stock over time.

Table 2. Descriptive statistics of the variables in the panel dataset including the mean, standard deviation, and percentile values (aggregate for all years)

	Mean	**Std**	**10%**	**30%**	**50%**	**70%**	**90%**
Low-income (%)	11.32	13.94	0	2.44	6.25	13.51	29.44
Moderate-income (%)	24.27	15.29	5.00	13.79	23.53	33.33	43.37
Middle-income (%)	21.71	10.49	8.33	16.67	22.22	27.27	33.33
Upper-income (%)	41.41	24.94	10.74	23.81	37.78	56.86	76.92
White (%)	62.77	21.06	34.15	51.51	66.99	78.05	86.13
Black (%)	18.84	18.50	0	4.35	12.50	27.27	45.59
Disinvestment (%)	9.12	11.35	0	1.25	5.00	11.11	25.00
Opportunity (%)	35.26	21.53	10.00	20.41	32.00	45.45	66.67
Renewed (%)	8.08	7.14	0	3.85	7.05	10.34	17.39
New suburban (%)	32.81	20.15	5.00	20.00	33.33	44.87	59.75
Expensive investment (%)	13.35	14.79	0	2.50	8.70	17.39	35.29

Finally, given historical patterns of intra-urban residential sorting and segregation by socioeconomic and demographic characteristics (see Figures 1, S1-S2) there is reason to suspect that loan applicant characteristics at the neighborhood level are spatially dependent. Therefore, we perform panel versions of the Lagrange Multiplie tests obtained by pooling the cross-sectional versions (Anselin et al., 2008, 1996; Millo et al., 2023). These tests suggest a spatial lag process for the share of applicants of different races and a spatial error process for the share of applicants of different income groups. The model in Equation (1) is estimated as follows when $\boldsymbol{Y_{i,t}}$ is related to race:

$$Y_{i,t} = \alpha + \rho W Y_{i,t} + X_{i,t-1}\beta + Z_{i,t-1}\omega + u_{i,t} \tag{2}$$

and as a spatial error model when $Y_{i,t}$ is related to income group:

$$Y_{i,t} = \alpha + X_{i,t-1}\beta + Z_{i,t-1}\omega + u_{i,t} \tag{3}$$

$$u_{i,t} = \lambda W u_{i,t} + \varepsilon_{i,t}$$

where W is a row-standardized, contiguity-based spatial weights matrix, ρ and λ are spatial autoregressive parameters, and ε an independently distributed error term.

4. Results and discussion

We begin by mapping the shares of different housing classes by census tract and compare them to patterns of mortgage applicants by income and race for a subset of years (Figures 1, S1-S2). The purpose of this is twofold. First, it validates the classifications by comparing them to existing knowledge about segregation and neighborhoods in Charlotte. Second, it gives us a descriptive understanding of the association between our housing classes and applicant characteristics. Note that we did not include the renewed class in the maps as this class did not exhibit any strong patterns - likely due to widespread redevelopment across the city’s tight housing market due to rapid population growth.

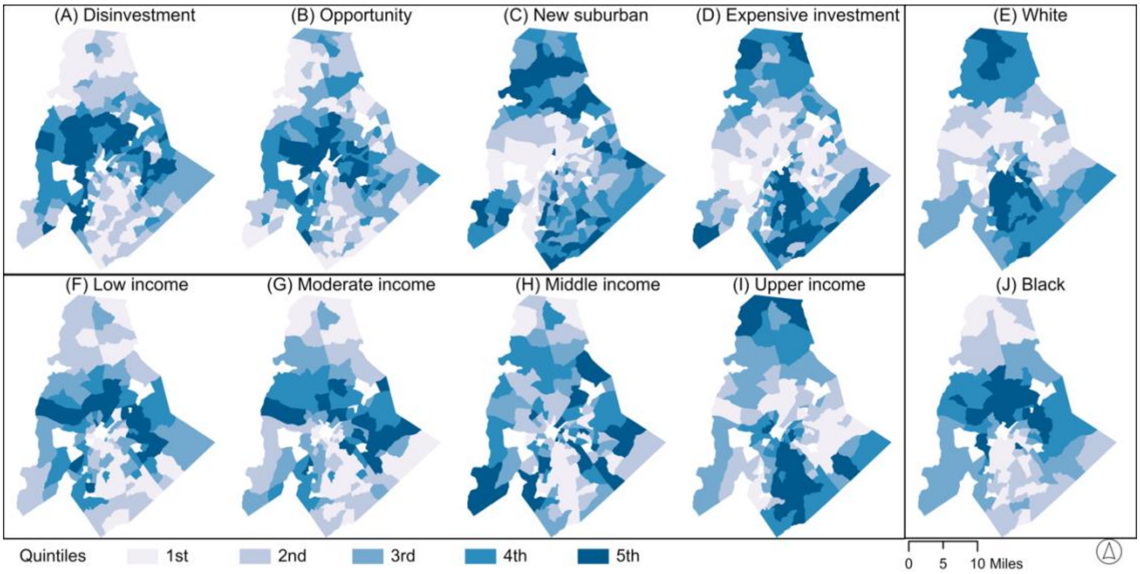

Fig 1. Shares of listings classified into different housing classes (A-D) mapped against the share of low through upper-income mortgage applicants (F-I) and the share of white (E) and black applicants (J) in 2016.

The maps in Figures 1, S1, and S2 demonstrate that the housing classification algorithm successfully demarcates well-known patterns of segregation and inequality in the city. Of note is the 'crescent' of neighborhoods with a high share of homes marketed as 'disinvestment' and 'opportunities'. These are spatially correlated with a high share of low-moderate income and Black mortgage applicants. Conversely, the southeastern wedge of 'wealth' featuring a high share of upper-income and White applicants corresponds to housing classified as 'expensive investments'. As expected, new suburban development is spread across the outer portions of the county and corresponds with patterns of upper-income and middle-income mortgage applicants.

We next explore the distribution of house prices among our five classes. The boxplots shown in S3 show that the mean listing price of each class fits with our theoretical foundation: highest for expensive investment, followed by new suburban, renewed, opportunity, and disinvestment. Importantly, these boxplots show overlap in the price distributions across different classes. This further underscores the necessity to segment homes based on their descriptions. Two properties might have similar prices, yet their valuations can vary significantly by the realtor crafting the advertisements. Consequently, one property might be promoted as an excellent investment opportunity, while the other could be marketed as a distressed property.

Next, we estimate the model in Equation 3 for low-income (%), moderate-income (%), middle-income (%), upper-income (%), respectively, and Equation 2 for White (%) and Black (%) applicants, respectively. The results are shown in Table 3 and the key findings are visualized in Figures 2, S4-S7.

Table 3. Regression results

	Dependent variable (%)					
	Low-income(t)	Moderate-income(t)	Middle-income(t)	Upper-income(t)	White(t)	Black(t)
Disinvestment percentile (t-1) (<30th reference group)						
30th-50th	0.462	2.231***	1.763***	-1.946***	1.071	0.837
50th-70th	1.491***	3.712***	2.660***	-3.940***	-1.538**	3.896***
70th-90th	4.126***	4.955***	1.633***	-4.013***	-5.132***	7.927***
>90th	9.847***	3.428***	0.006	-1.280	-2.377**	11.117***
Expensive investment percentile (t-1) (<30th reference group)						
30th-50th	-2.306***	-0.731	2.302***	4.808***	5.291***	-1.436**
50th-70th	-4.678***	-4.810***	2.377***	14.358***	10.637***	-3.827***
70th-90th	-5.714***	-8.735***	0.582	25.341***	16.435***	-5.693***
>90th	-6.156***	-10.838***	-2.922***	38.906***	24.693***	-7.301***
New suburban percentile (t-1) (<30th reference group)						
30th-50th	0.830	3.684***	1.333**	1.087	7.104***	-1.972**
50th-70th	-0.698	3.737***	3.765***	2.881***	9.012***	-2.839***

70^{th}-90^{th}	-1.483**	4.580***	5.625***	4.169***	11.598***	-3.304***
>90^{th}	-1.511**	5.218***	8.333***	6.054***	16.168***	-4.612***
Opportunity percentile (t-1) (<30^{th} reference group)						
30^{th}-50^{th}	1.174**	3.524***	1.935***	0.165	7.437***	-1.662**
50^{th}-70^{th}	2.154***	4.940***	3.983***	0.386	14.237***	-5.251***
70^{th}-90^{th}	5.204***	6.299***	3.842***	1.682	18.037***	-5.670***
>90^{th}	5.529***	5.405***	5.817***	5.643***	21.107***	-5.245***
Renewed percentile (t-1) (<30^{th} reference group)						
30^{th}-50^{th}	0.091	1.207**	1.537***	0.095	2.416***	-1.245*
50^{th}-70^{th}	0.019	1.919***	1.318***	0.911	3.138***	-0.701
70^{th}-90^{th}	0.885*	2.613***	1.951***	-0.261	4.521***	-1.318**
>90^{th}	0.716	4.436***	3.889***	-1.048	7.723***	-2.575***
Controls (%) (t-1)						
Low					-0.284***	0.330***
Moderate					-0.248***	0.330***
Middle					-0.101***	0.240***
Black	0.195***	0.241***	0.015**	-0.405***		
Constant	7.357***	13.601***	12.321***	34.646***	32.115***	2.372***
λ	0.329***	0.169***	0.143***	0.336***		
ρ					0.241***	0.245***
N	4,028	4,028	4,028	4,028	4,028	4,028
Pseudo R^2	0.38	0.38	0.15	0.55	0.38	0.38

Notes: Statistical significance at the *** 1%, ** 5%, and * 10% level.

The models generally account for about 40% of the variance in dependent variables, except for middle-income applicants, likely due to their less concentrated distribution compared to other income groups (Figures 1, S1-S2). This suggests middle-income applicants have more diverse housing options and less segregation compared to higher-income groups. They show a strong association with new suburban development (Table 3, Figure S4), with their share increasing steadily alongside the proportion of homes marketed as 'new suburban'. However, there's no apparent threshold effect here. Conversely, in neighborhoods with a significant proportion of expensive homes (>90th percentile), there's a notable decline in middle-income applicants, suggesting a potential threshold effect.

Moderate-income applicants tend to apply for mortgages in neighborhoods with a large share of new suburban development (Table 3, Figure S5). Given their weaker financial situation compared to middle-income earners, they are positively associated with neighborhoods containing a high share of disinvested homes and homes marketed as opportunities. As expected, these applicants are not likely to apply for homes in neighborhoods with significant shares of expensive investments.

The results clearly show the segregating force of expensive investments and disinvestment. The share of low-income applicants increases with the share of properties marketed as disinvested, with a notable difference in coefficient magnitude at the 70th and 90th percentiles (Table 3, Figure S6). Conversely, the larger the share of expensive and new suburban investment, the lower the share of low-income mortgage applicants. There is a strong relationship between the share of homes marketed as expensive investments and upper-income mortgage applicants (Figure S7).

Examining the results by race reveals that Black applicants predominantly purchase homes in neighborhoods with high shares of disinvested properties, with their numbers increasing as the share of such homes rises (Table 3, Figure 2). White applicants are nearly absent from these neighborhoods (Table 3, Figure 2). Conversely, White applicants rise with the share of expensive investments, a category negatively associated with Black applicants (Table 3, Figure 2). A notable threshold is observed: in neighborhoods with a high share (>50th percentile) of 'opportunity' homes, there is a significant increase in White applicants and a sharp decline in Black applicants.

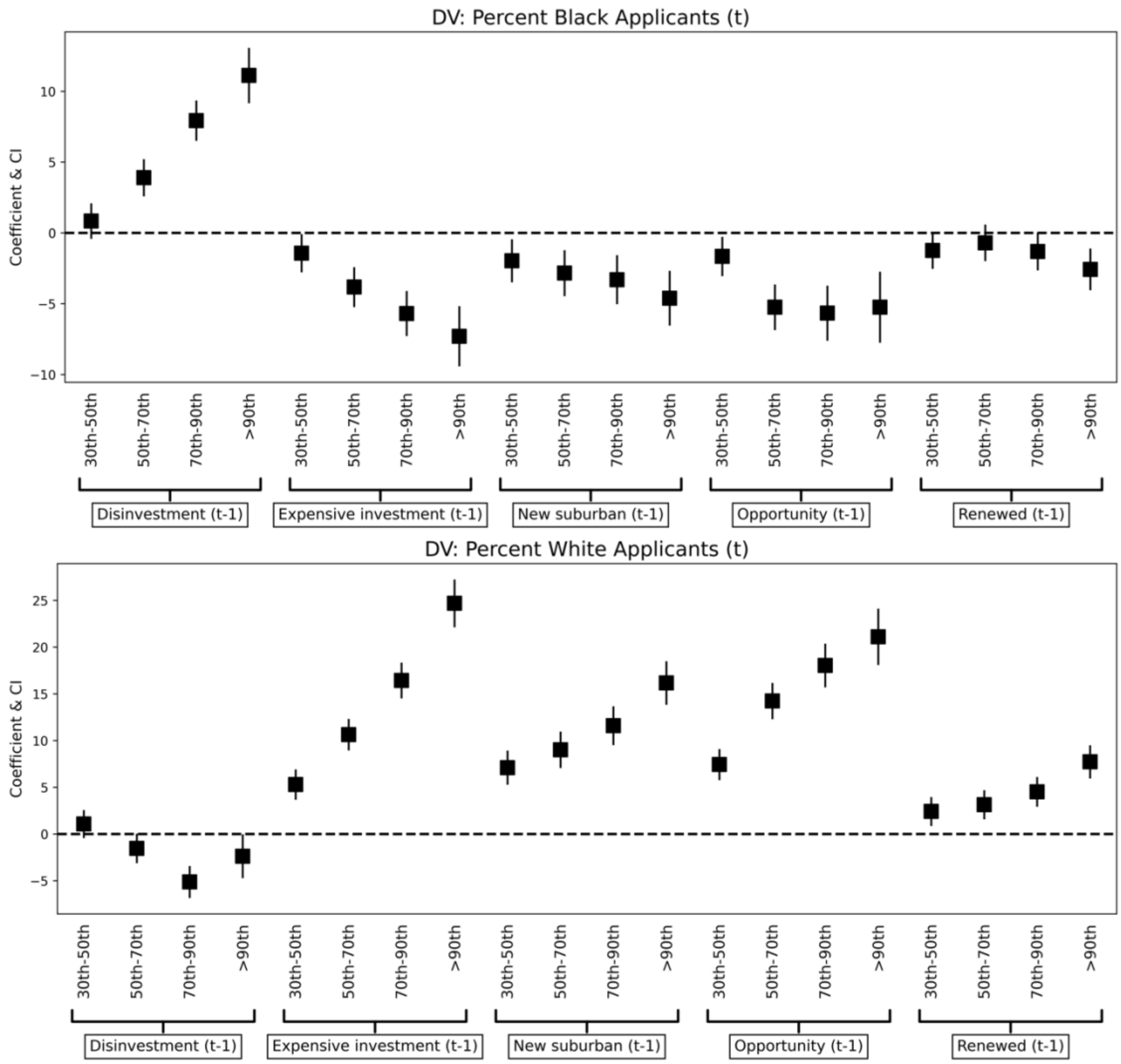

Fig 2. Estimated coefficients and 95% confidence intervals for housing class percentiles when the dependent variable is share of Black and White applicants, respectively.

To further explore the racial differences, we jointly estimate the effects of race and income in a set of models (Table S1, Figure S8). We find that while both Black and White low-income applicants show a possible threshold effect at the 90th percentile of disinvestment, this effect is significantly larger for Black applicants. Additionally, low-income White applicants are positively associated with neighborhoods featuring a moderate share of new suburban developments. In contrast, low-income Black applicants do not show a significant association with new suburban neighborhoods and have a negative relationship with areas having larger shares that housing class.

Both middle- and upper-income Black and White applicants tend to apply for mortgages in neighborhoods with a higher share of new suburban homes (Table S1). However, what differentiates them is their relationship to disinvestment and expensive investment. Middle- and upper-income Black applicants are positively associated with disinvestment, while their White counterparts show a negative association with disinvestment. Additionally, White middle- and upper-income applicants exhibit a strong positive relationship with expensive investments. In contrast, Black middle-income applicants have a significantly negative association with expensive investments. Upper-income Black applicants show only a weak positive association with moderate levels of expensive investment and a negative association with very high levels of expensive investment.

Finally, we investigated the mortgage application patterns of different income groups before and after the 2008 financial crisis (Table S2). The results indicate that post-2008, low-income applicants were more likely to apply for homes in the most disinvested neighborhoods. Before the crisis, they were more likely to apply for homes in neighborhoods with a moderate share of new suburban developments, but this trend disappeared after 2008. Similarly, while middle-income applicants had opportunities to apply for homes in neighborhoods with larger shares of expensive investments **before** the crisis, these opportunities vanished after 2008.

5. Discussion and Conclusions

In this paper, we examined how the marketing of single-family properties in a neighborhood explains shares of mortgage applicants by race and income in the subsequent year using a longitudinal case study on Mecklenburg County, North Carolina from 2000-2021. We applied a text classification algorithm to the property advertisements to derive a typology of five classes of homes marketed across the city: disinvestment, opportunities, renewed, new suburban, and expensive investment. We then estimated how the share of properties listed for sale in class in a neighborhood explained subsequent shares of mortgage applicants in that neighborhood.

Our analysis reveals stark patterns of segregation by income, race, and race and income jointly considered with the makeup of advertised homes in a neighborhood. We found that, overall, Black mortgage applicants nearly exclusively applied for mortgages in neighborhoods containing high shares of properties falling into our disinvested class – homes not advertised with a promise of a good investment or profit-making potential and with few advertised attractive amenities. This relationship held even when considering race and income together. Middle- and higher-income Blacks also largely purchased homes in neighborhoods with large shares of disinvested housing.

While our analysis cannot point to the causal mechanism explaining this finding, we can turn to the literature to posit a few hypotheses for future investigation. First, prior research has suggested that Black gentrifiers (middle- to upper-income homebuyers in previously disinvested Black neighborhoods), purposefully use their social and economic capital to uplift these neighborhoods (Pattillo, 2013). They value existing low-income Black communities and may choose these neighborhoods out of preference (Rucks-Ahidiana, 2021b). Second, current research continues to demonstrate that racial discrimination persists in the housing market search with Black home seekers consistently shown homes that are of lower quality than their White counterparts (Christensen & Timmins, 2022). Third, we know that in general, Black home seekers prefer racially mixed neighborhoods (Aliprantis et al., 2024), and given historical processes of discrimination and disinvestment in Charlotte, neighborhoods with moderate shares of Blacks also correspond to higher-poverty, lower-quality homes. Therefore middle-to upper-income Black homebuyers must consider a tradeoff between racial diversity and housing quality. In our case, like the finding by Aliprantis et al. (2024), we observe a preference for the former. However, in contrast to (Aliprantis et al., 2024), we do not control for wealth, only income. Given the growing racial wealth gap (Markley et al., 2020), differences in wealth could also contribute to these findings. Finally, previous research in Charlotte described how low-income Black residents were targeted to live in new, cheaply built suburban homes constructed on land close to environmentally toxic sites and far from any advantageous amenities – neighborhoods ‘built to fail’ that subsequently saw high rates of foreclosures (Currie & Sorensen, 2019; Sorensen et al., 2014).

Unlike low-income Black mortgage applicants, low-income White applicants were positively associated with moderate shares of new suburban homes and neighborhoods marketed as opportunities. Black low-income applicants were only positively associated with opportunities when the share of such properties was very high, suggesting riskier investments. White low- and middle-income applicants were significantly associated with opportunity neighborhoods at all housing shares. Thus, low-income White residents are more likely to take on the risk and potential financial reward of investing in housing on the frontiers of gentrification or in early new developments, while low-income Black residents purchase homes in areas of concentrated and continued disinvestment, with no advertised possibility of financial return. This racial disparity may contribute to the rising wealth gap stemming from homeownership (Markley et al., 2020). Our analysis linking realtor language to racial sorting outcomes supports the notion that the benefits of homeownership are not experienced uniformly by Black and White homebuyers.

High-income White mortgage applicants predominantly sought housing in New Suburban or Expensive Investment neighborhoods. As these neighborhoods became more homogenous, the share of wealthy White applicants sharply increased. The preference for racial and housing homogeneity among White home seekers is a longstanding trend, dating back to the formation of the National Realtor's Association and consistently supported by residential preference studies (Clark, 2002; Emerson et al., 2001).

We observe a strong correlation between new suburban housing developments and middle-income applicants. Initially, a small share of new suburban properties attracts significant low-income White applicants, but this response dissipates as the share increases. This may be due to the uncertainty in new developments, where investment risks lower prices, allowing low-income homeowners to enter the neighborhood (Hollans et al., 2012).

Finally, we observe changes before and after the Great Recession. After 2008, there is a stronger association between low-income applicants and neighborhoods with high disinvestment. Additionally, the pre-crisis association between middle-income applicants and neighborhoods with larger shares of expensive investment disappeared after 2008. This suggests fewer opportunities for upward mobility among multiple income groups after the crisis.

Taken together, our results paint a picture of sustained income and racial segregation in Charlotte, captured exclusively from the words stemming from real estate text. We do not suggest that the advertisements themselves are the causal mechanism behind these findings. But how realtors opt to craft the ads is reflective of their property valuation (Rucks-Ahidiana, 2021b) – a perspective that likely trickles down and influences further property marketing and showings. As Taylor (2019) contends, the language used to describe neighborhoods and housing continues to serve as a proxy for the location of Black housing (Taylor, 2019). Our analysis supports that idea.

Aside from the empirical findings of this study, our article also contributes to the growing body of literature that applies Natural Language Processing methods to real estate data, a previously under-utilized dataset for quantitative analyses of urban dynamics (Shahbazi et al., 2016; Hu et al., 2019). The analysis presented here can be extended to a more predictive framework to understand neighborhood dynamics at a fine temporal resolution and in near real-time.

Data availability statement

The historical MLS data from CoreLogic® is proprietary and not publicly available. However, public remarks for homes listed for sale can be obtained from websites like Zillow. The Python code for classifying real estate listings is available from Nilsson and Delmelle (2023a) at https://doi.org/10.6084/m9.figshare.20493012.v1, which also includes fabricated data similar to the dataset used in this study. HMDA-derived data and model estimation code can be accessed here: https://figshare.com/s/7491904cdf26acc03724.

References

Aliprantis, D., Carroll, D. R., & Young, E. R. (2024). What explains neighborhood sorting by income and race? Journal of Urban Economics, 141, 103508.

Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics, 26*(1), 77-104.

Anselin, L., Le Gallo, J., & Jayet, H. (2008). Spatial panel econometrics. In Matyas, L., & Sevestre, P. *The econometrics of panel data: Fundamentals and recent developments in theory and practice*. Springer-Verkag Berlin Heidelberg.

Besbris, M. (2020a). *Upsold: Real estate agents, prices, and neighborhood inequality*. University of Chicago Press.

Bridge, G. (2001). Estate agents as interpreters of economic and cultural capital: the gentrification premium in the Sydney housing market. *International Journal of Urban and Regional Research*, *25*(1), 87-101.

City of Charlotte (2024). Neighborhood resources: The quality of life explorer. https://www.charlottenc.gov/Streets-and-Neighborhoods/Get-Involved/Neighborhood-Resources#section-3 (last accessed 5/20/2024)

Christensen, P., & Timmins, C. (2022). Sorting or Steering: The Effects of Housing Discrimination on Neighborhood Choice. *Journal of Political Economy*, *130*(8), 2110–2163. https://doi.org/10.1086/720140/SUPPL_FILE/20191181DATA.ZIP

Clark, W. A. V. (2008). Changing Residential Preferences across Income, Education, and Age. *Http://Dx.Doi.Org/10.1177/1078087408321497*, *44*(3), 334–355.

Currie, M. A., & Sorensen, J. (2019). Repackaged “urban renewal”: Issues of spatial equity and environmental justice in new construction, suburban neighborhoods, and urban islands of infill. *Journal of Urban Affairs*, *41*(4), 464-485.

Delmelle, E. C., & Nilsson, I. (2021). The language of neighborhoods: A predictive-analytical framework based on property advertisement text and mortgage lending data. *Computers, Environment and Urban Systems*, *88*, 101658.

Delmelle, E. C., Nilsson, I., & Adu, P. (2021a). Poverty suburbanization, job accessibility, and employment outcomes. *Social Inclusion, 9*(2), 166-178.

Delmelle, E. C., Nilsson, I., & Schuch, J. C. (2021b). Who's moving in? A longitudinal analysis of home purchase borrowers in new transit neighborhoods. *Geographical Analysis, 53*(2), 237-258.

Emerson, M. O., Yancey, G., & Chai, K. J. (2001). Does Race Matter in Residential Segregation? Exploring the Preferences of White Americans. *Https://Doi.Org/10.1177/000312240106600607*, *66*(6), 922–935.

Frey, W. H. (2021). 2020 Census: Big cities grew and become more diverse, especially among their youth. Brookings Institution Report, October 28. https://www.brookings.edu/research/2020-census-big-cities-grew-and-became-more-diverse-especially-among-their-youth/ (last accessed 5/8/2023)

Galster, G. (1996). William Grigsby and the analysis of housing sub-markets and filtering. *Urban Studies*, *33*(10), 1797-1805.

Galster, G., & Godfrey, E. (2005). By words and deeds: Racial steering by real estate agents in the U.S. in 2000. *Journal of the American Planning Association*, *71*(3), 251–268.

Galster, G. C., Quercia, R. G., & Cortes, A. (2000). Identifying neighborhood thresholds: An empirical exploration. *Housing Policy Debate*, *11*(3), 701-732.

Galster, G., & Tatian, P. (2009). Modeling housing appreciation dynamics in disadvantaged neighborhoods. *Journal of Planning Education and Research, 29*(1), 7-22.

Hanlon, B. (2009). A typology of inner–ring suburbs: Class, race, and ethnicity in US suburbia. *City & Community*, *8*(3), 221-246.

Harvey, H., Fong, K., Edin, K., & DeLuca, S. (2020). Forever homes and temporary stops: Housing search logics and residential selection. *Social Forces*, *98*(4), 1498-1523.

Haupert, T. (2019). Racial Patterns in Mortgage Lending Outcomes During and After the Subprime Boom. *Housing Policy Debate*, *29*(6), 947–976.

Hollans, H., Martin, R. W., & Munneke, H. J. (2013). Measuring price behavior in new residential subdivisions. *The journal of real estate finance and economics*, *47*, 227-242.

Hu, Y., Mao, H., & McKenzie, G. (2019). A natural language processing and geospatial clustering framework for harvesting local place names from geotagged housing advertisements. International Journal of Geographical Information Science, 33(4), 714-738.

Korver-Glenn, E. (2018). Compounding Inequalities: How Racial Stereotypes and Discrimination Accumulate across the Stages of Housing Exchange. *American Sociological Review*, *83*(4), 627–656.

Korver-Glenn, E. (2021). *Race brokers: Housing markets and segregation in 21st century urban America*. Oxford University Press.

Krysan, M., Couper, M. P., Farley, R., & Forman, T. A. (2009). Does race matter in neighborhood preferences? Results from a video experiment. *American Journal of Sociology*, *115*(2), 527–559.

Krysan, M., & Farley, R. (2002). The Residential Preferences of Blacks: Do They Explain Persistent Segregation? *Social Forces*, *80*(3), 937–980.

Lichter, D. T., Thiede, B. C., & Brooks, M. M. (2023). Racial Diversity and Segregation: Comparing Principal Cities, Inner-Ring Suburbs, Outlying Suburbs, and the Suburban Fringe. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *9*(1), 26-51.

Logan, J. R. (2018). Relying on the census in urban social science. *City & Community, 17*(3), 540-553.

Logan, J. R., Xu, Z., & Stults, B. (2012). Interpolating US decennial census tract data from as early as 1970 to 2010: A longitudinal tract database. *The Professional Geographer, 66*(3), 412-420.

Mallach, A. (2024). Shifting the Redlining Paradigm: The Home Owners' Loan Corporation Maps and the Construction of Urban Racial Inequality. *Housing Policy Debate*.

Markley, S. N., Hafley, T. J., Allums, C. A., Holloway, S. R., & Chung, H. C. (2020). The Limits of Homeownership: Racial Capitalism, Black Wealth, and the Appreciation Gap in Atlanta. *International Journal of Urban and Regional Research*, *44*(2), 310–328.

Massey, D. S. (2020). Still the Linchpin: Segregation and Stratification in the USA. *Race and Social Problems*, *12*(1), 1–12.

Millo, G., Piras, G., & Bivand, R. (2023). 'splm': Econometric models for spatial panel data. https://cran.r-project.org/web/packages/splm/splm.pdf (last accessed 5/13/2024)

Nilsson, I., & Delmelle, E. C. (2023a). An embedding-based text classification approach for understanding micro-geographic housing dynamics. *International Journal of Geographic Information Science*

Nilsson, I., & Delmelle, E. C. (2023b). Smart growth as a luxury amenity? Exploring the relationship between the marketing of smart growth characteristics and neighborhood racial and income change. *Journal of Transport Geography*, *106*, 103522.

Nowak, A., & Smith, P. (2017). Textual analysis in real estate. *Journal of Applied Econometrics*, *32*(4), 896-918.

Nowak, A., Price, B. S., & Smith, P. S. (2021). Real Estate Dictionaries Across Space and Time. *Journal of Real Estate Finance and Economics, 62*(1), 139-163.

Pattillo, M. (2005). Black middle-class neighborhoods. *Annual Review of Sociology*, *31*(Volume 31, 2005), 305–329.

Pattillo, M. (2013). Black on the Block. *Black on the Block*. https://doi.org/10.7208/CHICAGO/9780226649337.001.0001

Perkins, H. C., Thorns, D. C., & Newton, B. M. (2008). Real estate advertising and intraurban place meaning: Real estate sales consultants at work. *Environment and Planning A*, *40*(9), 2061-2079.

Querica, R. G., & Galster, G. C. (2000). Threshold effects and neighborhood change. *Journal of Planning Education and Research*, 20(2), 146-162.

Quillian, L., Lee, J. J., & Honoré, B. (2020). Racial Discrimination in the U.S. Housing and Mortgage Lending Markets: A Quantitative Review of Trends, 1976–2016. *Race and Social Problems*, *12*(1), 13–28.

Rucks-Ahidiana, Z. (2021a). Racial composition and trajectories of gentrification in the United States. *Urban Studies*, *58*(13), 2721–2741.

Rucks-Ahidiana, Z. (2021b). Theorizing Gentrification as a Process of Racial Capitalism.

Schopf, T., Braun, D., & Matthes, F. (2021). Lbl2Vec: An embedding-based approach for unsupervised document retrieval on predefined topics. *Proceedings of the*

17th International Conference on Web Information Systems and Technologies - WEBIST, 124-132.

Shahbazi, M., Barr, J. R., Hristidis, V., & Srinivasan, N. N. (2016, February). Estimation of the investability of real estate properties through text analysis. In 2016 IEEE Tenth International Conference on Semantic Computing (ICSC) (pp. 301-306). IEEE.

Skaburskis, A. (2010). Gentrification in the context of 'risk society'. *Environment and Planning A*, *42*(4), 895-912.

Slater, G. (2021). Freedom to Discriminate: How Realtors Conspired to Segregate Housing and Divide America. Heyday.

Smith, H., & Graves, W. (2005). Gentrification as corporate growth strategy: The strange case of Charlotte, North Carolina and the Bank of America. *Journal of Urban Affairs*, *27*(4), 403-418.

Sorensen, J., Gamez, J., & Currie, M. (2014). Windy Ridge: A neighborhood built to fail. *Applied Geography*, *51*, 8–25.

Sperling, J. (2012). The tyranny of census geography: Small-area data and neighborhood statistics. *Cityscape, 14*(2), 219-223.

Taylor, K. Y. (2019). *Race for profit: How banks and the real estate industry undermined black homeownership*. UNC Press Books.

Turnbull, G. K., & Dombrow, J. (2006). Spatial competition and shopping externalities: Evidence from the housing market. *The Journal of Real Estate Finance and Economics*, *32*, 391-408

Turner, M. A., Santos, R., Levy, D. K., Wissoker, D. A., Aranda, C., & Pitingolo, R. (2016). *Housing discrimination against racial and ethnic minorities 2012: Executive summary*.

Van Zandt, S. (2007). Racial/ethnic differences in housing outcomes for first-time, low-income home buyers: Findings from a National homeownership education program. *Housing Policy Debate*, *18*(2), 431-474.

Van Zandt, S., & Rohe, W. M. (2011). The sustainability of low-income homeownership: The incidence of unexpected costs and needed repairs among low-income home buyers. *Housing Policy Debate*, *21*(2), 317-341.