The Unequal Availability of Rental Housing Information Across Neighborhoods

Abstract:
As more urban residents find their housing through online search tools, recent research has theorized the potential for online information to transform and equalize the housing search process. Yet we know very little about what rental housing information is available online. Using a corpus of millions of geocoded Craigslist advertisements for rental housing across the 50 largest metropolitan statistical areas in the U.S. merged with census-tract level data from the American Community Survey, we identify and describe the types of information commonly included in listings across different types of neighborhoods. We find that in the online housing market, renters are exposed to fundamentally different types of information depending on the ethnoracial and socioeconomic makeup of the neighborhoods in which they are searching.

Residential mobility decisions—which are predicated on information about available housing units and accessible neighborhoods—critically shape life chances (Bischoff and Owens 2019; Chetty and Hendren 2018; Sampson 2012; Sharkey and Faber 2014) as well as rates of residential segregation (Krysan 2002; Krysan and Crowder 2017; Massey and Denton 1993; South et al. 2011). Longstanding interest in variations in the availability of housing information across ethnoracial groups (e.g., Courant 1978) has culminated in recent insights regarding how homeseekers form their choice sets and make decisions about where to move (Bader and Krysan 2015; Bruch and Swait 2019; Havekes et al. 2016; Krysan and Crowder 2017). Broadly, this research is committed to the idea that differential housing outcomes by race/ethnicity—like those previously found in research on residential mobility (e.g., Bruch and Mare 2006; Bruch and Swait 2019; Crowder and South 2005; Logan and Alba 1993)—are both a product and cause of socio-spatial inequality (Krysan and Crowder 2017).

In this article, we gather data and employ methods rarely used by demographers to further research on residential mobility in the U.S. in two key, related ways. First, in light of the rapidly changing housing search process, we examine what kinds of information about rental housing are available online. The most recent American Housing Survey (2017) found that housing websites are now a primary source of information for all urban homeseekers. Yet despite the turn toward understanding the sources of information homeseekers rely on when making mobility decisions and the expanding use of online search tools, we lack a thorough description of the housing market information that is readily available online. We examine if this information is similar across neighborhoods, or, if like other sources of information, housing websites present segmented and segregated information that tracks with racial/ethnic and socioeconomic forms of socio-spatial inequality. Doing so helps adjudicate between perspectives that express some optimism about the potential for online search tools to reduce racial information inequalities in residential mobility decision making (e.g., Krysan and Crowder 2017; McLaughlin and Young 2018; Palm and Danis 2001) and others that argue that any new technology that fosters mobility will likely reproduce existing inequalities (e.g., Brannon 2017; Massey 2005; Stiel and Jordan 2018). Second, in analyzing an increasingly important source of information for homeseekers, we emphasize that mobility decisions are, in part, a product of the supply of information on available units.

To understand the supply of information, we collected 1.6 million geo-coded advertisements for rental housing from the 50 largest metropolitan areas in the U.S. posted on Craigslist, the dominant platform for today's metropolitan rental housing market (Boeing 2020; Boeing and Waddell 2017).[1] Using computational text analysis techniques, we first identify common types of information displayed in online housing advertisements. Next, we demonstrate that advertisements for rental housing largely reflect existing socio-spatial inequalities: the information about available housing units varies depending on the surrounding neighborhood's

[1] Craigslist operates as a classifieds website outside of the U.S. as well but is not the primary housing website in many other countries (see Rae 2015).

ethnoracial make-up and rate of households in poverty. Listings in poorer neighborhoods tend to focus on tenant qualifications like financial requirements and lack of eviction or criminal history, rather than describing the housing unit's amenities. But even among non-poor neighborhoods, listings in Black and Latino neighborhoods focus disproportionately on tenant (dis)qualifications compared to listings in otherwise similar White neighborhoods, underscoring the racialized nature of information in the online rental housing market. In contrast, listings in White and Asian neighborhoods—regardless of poverty level—are more likely to describe the aesthetic qualities of housing units. Finally, listings with higher asking rents in White and Asian neighborhoods, particularly those with higher poverty rates and thus more gentrification potential (Hwang 2015; 2016; Hwang and Sampson 2014; Timberlake and Johns-Wolfe 2017), are more likely to describe desirable neighborhood characteristics. These differences highlight the importance of studying the information environment itself (Bruch and Feinberg 2017). In the online housing market, renters are exposed to fundamentally different types of information depending on the ethnoracial and socioeconomic makeup of the neighborhoods in which they search. These information differences may attract or repel certain types of homeseekers and reify place reputations—key mechanisms of residential sorting which operate as both outcomes and causes of sociospatial inequality (Krysan and Crowder 2017).

INFORMATION AND RESIDENTIAL MOBILITY IN THE RENTAL MARKET

Since the Great Recession, a growing number of American households have become renters and the rental market continues to be where the majority of African-Americans, Latinos, and immigrants find their housing (Ellen and Karfunkel 2016; NMHC 2016; Schachter and Besbris 2017). Renters, compared to homeowners, have higher rates of residential mobility, different rates of racial/ethnic segregation, distinct choice constraints in their residential mobility decisions, and most metropolitan renters face a market where demand is higher relative to supply and costs are rising (DeLuca et al. 2013; Desmond and Shollenberger 2015; Friedman et al. 2013; JCHS 2019; Pilkauskas and Michelmore 2019). Landlords are therefore well positioned to exploit information asymmetries, potentially shaping how renters are sorted (Garboden and Rosen 2019; Rosen 2014).

As a growing number and proportion of Americans have become renters and rental housing affordability has declined, the housing market in general—and the market for rental housing in particular—increasingly operate online. Recent survey data show that, across racial/ethnic groups, internet sites like Craigslist are one of the top two most common ways homeseekers in urban areas find places to live along with word of mouth (American Housing Survey 2017).[2] In short, housing websites—in conjunction with the sharp decline in unequal access to the internet in urban areas (Anderson and Perrin 2016)—are transforming residential search and mobility processes (see Schachner and Sampson 2020:679).

Online rental housing advertisements are a point of connection for landlords, who control the supply of rental housing, with prospective renters, whose preferences shape demand. By serving as this point of connection, rental housing advertisements can influence the types of households who do and do not apply. In other words, advertisements for rental housing are a key source of information for renters making residential mobility decisions.

[2] The use of the internet as a rental housing search tool is differentiated across levels of education. Renters with Bachelors or advanced degrees report finding their housing via the internet at double the rate of renters with a high school degree or less. However, this trend reverses when looking at renters with families: renters with families with a high school education or less are nearly two times more likely to find housing through the internet than renters with families with bachelors or advanced degrees.

Online listings are also often the first signal that prospective renters receive about whether or not a particular unit matches their housing preferences, and the iterative and imbricated nature of the housing search process means that prospective tenants learn about potential places to live as they browse listings (Krysan and Crowder 2017). Moreover, online search tools allow for easier and faster comparison across units. On the one hand, this could help equalize searches since a wide range of listings can be accessed. On the other hand, it could also heighten particular signals as searchers quickly screen out a high volume of listings (Bruch, Feinberg, and Lee 2016).

When individuals embark on a search for housing, one of their initial tasks is to determine in which neighborhoods they would consider living (Bader and Krysan 2015). Of course, renters face structural constraints (e.g., price, geography), have pre-existing information about neighborhoods, and may have preferences for certain neighborhoods based on their social networks, commutes, etc. But this pre-existing information tends to be minimal (Krysan and Bader 2009; Lareau 2014).[3] Previous research has shown that certain types of information can affect homeseekers' understandings of different neighborhoods as more or less appropriate place to live. Descriptions of local amenities, for example, influence neighborhood selection, and homeseekers are sensitive to signals about crime and safety—though these may be proxies for ethnoracial makeup (Krysan and Crowder 2017; see also Quillian and Pager 2001). In fact, language that is not overtly racial can still provide cues about a given neighborhood's demographics (Besbris 2016; 2020; Besbris and Faber 2017; Howell and Emerson 2018;

[3] Additionally, according to the Current Population Survey, since 2005 3-5% of the U.S. population engage in cross-county residential mobility annually, meaning that, in a given year, tens of millions of Americans are gathering information on neighborhoods in cities where they do not currently live.

Kennedy et al. 2020; Korver-Glenn 2018)—and can subsequently affect residential mobility decisions and economic decisions more generally (Besbris et al. 2015; 2019; Krysan and Crowder 2017). Landlords may similarly be influenced by shared perceptions of neighborhoods when they compose their advertisements, and the information they provide about available housing likely both reflects existing patterns of residential sorting and perpetuates them.

After homeseekers select a neighborhood or set of neighborhoods in which to search, they must then compare available housing units (Krysan and Crowder 2017:53). Variation in descriptions about the units are a key component of the residential selection process (see Harvey et al. 2020; Rosenblatt and DeLuca 2012; Wood 2014). The information on sites like Craigslist is potentially influential for the selection of both neighborhoods and of individual housing units—it is far more robust than the types of information homeseekers tend to gather from their social networks (Carrillo et al. 2016; Lareau 2014) and is updated in real time as renters go through the search process.

Most online housing platforms like Craigslist require landlords to provide the location and price of a listed unit. However, landlords are free to choose what other types of information to include in their advertisements—e.g., describing the unit in text. Whether a landlord is motivated by profit, bias, or simply trying to provide relevant information to prospective tenants, the selective inclusion and exclusion of information in rental housing advertisements may influence homeseekers' residential mobility decisions. In other words, examining the content of online housing advertisements is essential since it reflects existing perceptions about the types of people that belong in particular neighborhoods, facilitates landlords' selection of particular kinds of renters, and enables renters to select neighborhoods and units that match their preferences.

DATA AND METHODS

Following growing recognition of the value of data collected online for understanding demographic processes (Cesare et al. 2018), we examine advertisements collected from Craigslist. Not all rental housing in the U.S. is advertised on Craigslist; indeed, Boeing (2020) found that, in 2014, advertisements for rental housing on Craigslist were overrepresented in neighborhoods with higher shares of White residents, demonstrating how offline inequalities are reproduced in the supply of information online. However, our goal is to understand what kinds of information are shared on Craigslist, which is the most comprehensive and timely source of housing market information in the U.S. (Boeing and Waddell 2017).

We designed a set of Python scripts to crawl Craigslist and gather information from rental ads, including listing date, rent (price), square footage and other unit characteristics, neighborhood name, geo-location, and the full text of the advertisement. We include all Craigslist sites that correspond to the 50 largest metropolitan statistical areas (MSAs) in the U.S.[4] Posters creating ads for rental housing are asked to supply the closest cross streets for the listing and this position is plotted on a Google maps image embedded within the advertisement. We identify each listing's location using the approximate longitude and latitude from the cross-street plot on Google maps. Across our metro areas, 12% of all listings are missing a geocode and are thus excluded from all analyses presented here. We use the geocodes to assign each advertisement to a Census tract using the ArcGIS geographic join tool, which returns 15

[4] For most metropolitan statistical areas, the corresponding Craigslist site closely matches Census MSA definitions; moreover, because we only use Census data at the tract level and follow Craigslist market definitions to determine metro area boundaries, any discrepancies do not impact our results (see Appendix).

character FIPS census tract code to indicate in which Census tract each geocode belongs.[5] The Python scripts revisit each MSA Craigslist site once per week, and check to see whether each currently posted listing is new, in which case all information is scraped, or if the listing is a repeat from the previous week, which is also noted in the database.[6]

From late May through the middle of February, 2018, we collected 3,950,558 listings across all 50 MSAs. We eliminate listings missing price information (about 1%), listings with prices higher than $10,000 per month (about 1%), as well as listings that are duplicates (about 50%), for a final dataset of 1,697,117 unique, geocoded listings.[7] We then merge our data with 2016 American Community Survey (ACS) 5-year pooled data on tract racial/ethnic composition, poverty status, and other neighborhood characteristics relevant to rental market dynamics. Because of missing data in our various covariates and listings that do not have enough text for topic modeling, our final sample size is 1,692,639.[8]

A large body of work documents that two dimensions of neighborhoods together structure the housing market overall and residential mobility in particular: ethnoracial and socioeconomic composition (Adelman 2005; Charles 2006; Clark 1992; 2009; Clark and Morrison 2012; Crowder and South 2008; Gabriel and Spring 2019; Krysan and Crowder 2017; Lee et al. 1994; Sampson and Sharkey 2008; Swaroop and Krysan 2011). Pervasive cross-neighborhood inequality off-line motivates us to test whether the entrenched socio-spatial

[5] Here, we measure information differences at the tract-level. In the appendix we run similar models at the zipcode-level in case there are any innacuracies in matching ads to tracts and find similar results.
[6] We exclude posts that do not post a date on which the housing is available because these tend to be spam posts. By scraping weekly, we miss listings that are posted and removed within one week.
[7] We present regression results with the sample that does not remove listings with price higher than $10,000 in the Appendix. The results are substantively similar.
[8] See the Appendix for tables on the distribution of ads across MSAs and tracts.

hierarchy in U.S. cities is reflected in the types of information contained in housing listings in different types of neighborhoods.[9] Similar to Wang et al. (2018), in most of our analyses we use 2016 5-year pooled ACS data to classify tracts into eight different neighborhood types: White non-poor, White poor, Black non-poor, Black poor, Latino non-poor, Latino poor, Asian non-poor, and Asian poor. Neighborhood racial composition is based on the plurality racial group, and we use a threshold of 30% of tract households living at or below the federal poverty line as our measure of neighborhood poverty. We have tested these cut offs and do not find substantive differences using alternative neighborhood classification schemas; these results are reported in Appendix tables A11 and A14 . By using a categorical measure of neighborhood type, we are better able to identify how neighborhood racial composition and socioeconomic status intersect. We find substantively similar results using continuous measures (see Appendix Table A15). In additional analyses we include posted unit rental price and a broader set of neighborhood measures (from 2016 ACS 5-year data) that are commonly correlated with neighborhood race/ethnicity and poverty rate, including the proportion of college educated residents, the proportion of foreign born population, the proportion of units renter occupied, the proportion of units built after 2010, and the neighborhood vacancy rate.[10] These variables, which measure either the quality of the units in a neighborhood or neighborhood demographics, were selected because they could plausibly affect the ways landlords list their units. For example, a more highly educated renter pool could prompt landlords to advertise certain types of amenities, while a higher vacancy rate might create more competition for potential renters and result in landlords

[9] The non-random selection of neighborhoods by economic conditions during the residential mobility process and the subsequent effect of these conditions on residents exists outside of the U.S. as well (McAvay 2018; van Ham et al. 2018).
[10] These additional measures are meant to further contextualize the main findings on differences in advertisements by neighborhood type (see Varian 2014).

adding more information to their listings (Boeing et al. 2020).

Method

Recent work suggests that information presented in online housing advertisements does indeed vary by neighborhood. For example, Craigslist advertisements from poorer neighborhoods and neighborhoods with more Black residents contain fewer words on average and are less likely to state an exact address in the ad (Boeing et al. 2020). In addition, the prevalence of Craiglist listings relative to underlying housing stock is greater in neighborhoods with whiter, higher-income, and/or more educated residents (Boeing 2020). Critically, however, there is little work that has examined the main textual description of listings, though Kennedy et al. (2020) is a recent major exception. While collecting and analyzing these data is more time and labor intensive, it allows for a deeper understanding of the information commonly included in online rental housing advertisements. As we demonstrate below, the text of advertisements tend to contain key details about housing and neighborhoods which likely influence prospective renters' residential mobility decisions. Moreover, unlike the 'check-boxes' analyzed in prior work which offer landlords limited options/flexibility to describe their units (Boeing et al. 2020), in the main listing description landlords have full discretion to include (or exclude) any type and amount of text-based information they consider necessary to attract their desired tenants. Exploring this discretionary information is critical given research in other settings which shows how unequal/discriminatory treatment can be more prevalent in contexts where gatekeepers have more discretion (Pager and Shepherd 2008).

We use two computational text analysis approaches to describe the kinds of information available in online housing advertisements and to test for variation across neighborhoods. First, we use Structural Topic Models (STM) to identify common topics or themes in advertisements. Topics are sets of words which frequently co-occur. For example, we might anticipate that rental housing listings are likely to include language about the number of bedrooms and bathrooms. If basic descriptions of the housing unit are a common type of information, topic model results will include a topic with words often found in these descriptions (like ‘bedroom’, ‘bathroom’, etc.). Topic models do not require researchers to know beforehand which themes will emerge; rather, they take a purely inductive approach by identifying commonly co-occurring sets of words. The researcher then examines the collections of words and identifies their substantive meaning. Thus, topic models allow us to characterize the different content areas that are commonly included in online housing listings without being influenced by any prior assumptions. In addition to providing a description of information commonly presented in online housing advertisements, STM can also be used to compare the prevalence of topics and specific word choices within a topic across different types of neighborhoods (DiMaggio, Nag, and Blei 2013; Roberts et al. 2014). In other words, we can use STM to estimate which types of topics are more likely to be used by landlords in advertisements for housing in different types of neighborhoods and to compare the types of words used across neighborhoods within the same topic.

To run our STM analysis, we first create a document term matrix. It contains information specifying how many times each term appears in each individual document, i.e. advertisement. In preprocessing the corpus, we convert capital letters to lowercase, remove numbers, stop words and punctuations, and conduct stemming to obtain more informative outcomes. Second, we remove low frequency words—words that appear in less than 1% of ads—which is standard and a crucial step to reduce noise in the outcome (Mosteller and Wallace 1963).

After preprocessing and creating a document-term matrix, we run a STM with 7 topics.[11] Our STM analysis occurs over three stages. First, we compute topic proportions by each document.[12] When we conduct STM, we include covariates that account for the document generating process.[13] However, including covariates in the STM estimation makes minimal difference in the topic model outcomes, and in the Appendix we show that STM without any covariates demonstrates identical results. Second, we run regression models to estimate the relationship between neighborhood type and topic proportions by including the same set of covariates. In both cases we include MSA fixed-effects. Finally, we compare word choices within the same topic between neighborhood types by running a new STM including a dummy variable that indicates whether the neighborhood is majority White or majority non-White (because this type of analysis can only be conducted across two groups rather than our 8-category neighborhood typology).[14]

As we describe in detail below, our STM analysis finds substantial variation in the information provided in advertisements across neighborhoods. To further explore these differences, we examine whether individual words are associated with neighborhood characteristics. Focusing only on topics might obscure specific words or phrases like 'Section 8' (a reference to subsidies in the form of vouchers provided to some poor- and moderate-income

[11] We follow Chang et al. (2009) in choosing a number of topics that are easily interpretable and convey cohesive meaning. In the Appendix, we present results that use different number of topics.
[12] STM also produces word-topic matrix, which represents the proportion of word use by each topic.
[13] Our preferred STM estimation includes the posted unit price, our 8-category neighborhood type classification, the proportion of college educated residents, the proportion of units renter occupied, the proportion of units built after 2014, the proportion of foreign born population, the vacancy rate, and MSA fixed effects as covariates.
[14] Because word comparisons within topics can only be calculated for binary neighborhood measures, here we classify all neighborhoods as majority White or majority non-White.

households for use in the private rental market) or ‘Whole Foods’ (a reference to an upscale grocery store) that vary systematically across different types of neighborhoods. In addition, to implement STM we have to make multiple decisions in the modeling and interpretation process which might inadvertently influence our findings. Using a second approach to understand patterns in out text data provides a key robustness test.

To identify individual words that tend to distinguish between neighborhoods, we use Multinomial Inverse Regression (MNIR), a powerful tool for measuring how words are associated with continuous measures of neighborhood characteristics.[15] MNIR incorporates high dimensional data—such as text data with large numbers of covariates—into statistical analysis to uncover the strength of correlation between each word and our covariate of choice. For example, MNIR estimates how each word in the corpus is correlated with the poverty rate of the neighborhood where the listing is posted.[16] If “section” (as in “section 8 housing”) has a strong correlation with listing tract poverty rate, then the MNIR coefficient for “section” will be high (note that because MNIR estimates the strength of association better with continuous measures, we do not use our neighborhood categorization measure).[17] MNIR achieves the same goal as OLS while dealing with statistical issues in high dimensional data such as text data. To run MNIR, we use the same preprocessing techniques as our STM analysis and prepare the

---

[15] We choose MNIR over other machine learning methods such as LASSO or ridge regression because MNIR is specifically developed for application to text data (Gentzkow et al. 2019).

[16] Past work has referenced unpublished research on STM analyses of Craigslist data (see Boeing et al. 2020) or used STM to analyze ads in one city (Kennedy et al. 2020) but to our knowledge no research has used MNIR to understand the characteristics of rental market information across neighborhoods.

[17] Since this is a question of association instead of prediction (see Grimmer 2012), prior work has shown log-odds and model based approaches like MNIR to be effective for identifying distinguishing words (Manning et al. 2008; Taddy 2013). However, we find that Mutual Information models (available upon request) and MNIR show very similar results.

document-term matrix. MNIR will produce a vector of coefficients that contains the strength of correlation between the selected neighborhood covariate and each word.[18]

Finally, we conduct additional, ancillary analyses presented in the Appendix to further quantify information differences across neighborhoods. In these analyses we use OLS to identify variation in numeric characteristics of online advertisements, including the overall number of words included in each advertisement and the number of pictures. These analyses identify clear differences in type of information included in advertisements depending on the neighborhood in which they are located, further supporting the text-based analyses presented below.

RESULTS

Identifying Information Types

We begin by identifying the types of information commonly included in online housing advertisements. Table 1 details our seven topics, listing their labels, the most common words within these topics, and each topic's prevalence average across the entire set of advertisements.[19]

[Table 1 here]

The *general information* topic focuses on the type of building (apartment building, duplex, single-family home, etc.) and on average accounts for about 25% of a listing's content. The *availability* topic includes information on how to contact the landlord and view the unit and accounts for 8.5% of listing content on average. *Unit description*, at 25% of average listing content, includes basic descriptions of the size of the unit (number of bedrooms, bathrooms,

[18] MNIR does this by assuming the document term matrix is a collection of draws from a multinomial distribution and inversing the regression framework by putting the high dimensional document term matrix on the left-hand side.
[19] Words that are displayed in Table 1 have simplified forms because they are all stemmed. Every word from STM will be displayed as a stemmed form.

etc.). The *Pet policy* topic (8.3% of content) covers whether and which types of pets are allowed in the unit. These topics cover basic, perfunctory information that is undoubtedly part of homeseekers' decisions. However, they are less subject to landlords' capriciousness in that they are generally objective characteristics about the unit. Additionally, much of this information, including number of bedrooms/bathrooms, the availability of a washer/dryer, and the pet policy, can all be signaled by the landlord in write-in options and checkboxes that are independent of the advertisement's main text and can be used by homeseekers to filter units when searching on Craigslist. Other work more thoroughly explored the usage of checkboxes in Craigslist advertisements (Boeing et al. 2020). As a result, our focus here is on describing the linguistic content of the text in ads.

[Figure 1 here]

We focus our analysis on three topics: *logistics*, *unit amenities,* and *neighborhood amenities*, because of their theoretical importance to the residential decision process (Desmond 2016; Harvey et al. 2020; Krysan and Crowder 2017; Rosen 2017; Rosenblatt and DeLuca 2012; Wood 2014). The first topic we focus on, *logistics*, captures language about the logistics of applying to rent a unit and desired and undesired renter characteristics, including language about housing vouchers, eviction history, and income and credit score requirements. The *logistics* topic accounts for about 10% of the content in an average listing. Landlords include this type of text in their advertisements to try and influence who will contact them for more information or apply to rent their unit (Rosen 2014). Figure 1, above, shows two examples of listings from our data that have a high proportion of the logistics topic.

The next topic we examine is *unit amenities.* Relative to the *unit descriptions* category, this topic captures more specific language about optional housing features (see Figure 1 for examples). Moreover, in additional analyses presented in Appendix Table A9 and Figure A1, we confirm that these topics are distinct by identifying different prevalence patterns across neighborhoods. On average the *unit amenities* topic accounts for about 12% of listing content. Our third and final topic of interest is *neighborhood amenities*. The *neighborhood amenities* topic contains language describing neighborhood characteristics. This topic, which accounts for about 10% of the content in an average listing, includes words about parks, shopping, restaurants and other information about neighborhood location or resources (see Figure 1).[20]

Testing for Variation in Available Information Across Neighborhoods

Next, to test whether each topic is disproportionately prominent in advertisements in certain types of neighborhoods, we treat topic prevalence, or the proportion of each advertisement's words that are dedicated to each topic, as a dependent variable. We use OLS models with MSA-level fixed effects to predict prevalence, and we cluster standard errors by MSA.[21] We log transform our dependent variables because the distribution of the variables are skewed to the right.[22] Table 2 reports coefficients predicting the logistics topic (model 1a), unit amenities topic (model 2a), and neighborhood amenities topic (model 3a) without additional control measures. We use these models to estimate pairwise differences across neighborhood types by changing the baseline category in our regression models and generating predicted

[20] Unit amenities and neighborhood amenities both potentially vary based on landlords' advertising strategies as well as underlying differences in housing and neighborhood characteristics. Our data cannot adjudicate between the two. However, given prior research showing that racial composition tends to predict residents' and homeseekers' *assumptions* about unit and neighborhood amenities (Krysan and Crowder 2017), it is unlikely that underlying quality/fixed characteristics of housing and neighborhoods can completely explain differences in ad text.

[21] We obtain similar results when we cluster by Census tract (see Appendix Table A20).

[22] See Appendix Figures A2 and A3 for histograms of dependent variables. Results are similar when we do not log transform our dependent variables (See Appendix Table A10 and Figure A4).

values. Figure 2 reports pairwise comparisons of neighborhood types for each of our topics. Because our dependent variables are log-transformed, we present percent change in topic proportion, rather than the regression coefficients, for ease of interpretation.

[Figure 2 here]

As shown in the first column of Figure 2, regardless of race, discussion of rental logistics—income requirements, background checks, renter disqualifications—is almost 50% more prevalent in poorer neighborhoods compared to their same-race, non-poor counterparts. However, the logistics topic is not just associated with poverty status; holding poverty constant, the logistics topic remains more prevelant in Black and Latino neighborhoods compared to White ones, ranging from 25 to 75% more prevelant in Black and Latino neighborhoods compared to White neighborhoods with similar poverty rates.[23] While discussion of rental logistics is clearly related to the socioeconomic status of neighborhoods, it is also racialized.

These results suggest that renters searching in lower-income and/or Black and Latino neighborhoods are more likely to encounter restrictive information in advertisements, which may encourage some searchers—those who can afford it—to select out of the prospective pool of renters before a landlord even begins to formally review applicants. This language may also lead some renters to believe that these neighborhoods have high rates of crime, eviction, and poverty. In other words, this language work in tandem with any pre-existing information to further stigmatize neighborhoods (Besbris et al. 2015; 2018; 2019; Sampson and Raudenbush 2004). However, some renters might appreciate having rental requirements presented upfront, and other words within this topic focus on subjects like lease terms and the rental application process.

[23] In these analyses, White, Black, and Latino poor neighborhoods have similar poverty rates (median 38%).

While our data do not allow us to test the effects of language on renter behavior we can examine whether there is any heterogeneity in the choice of words within the logistics topic by neighborhood racial composition. In this analysis we must use a binary neighborhood classification; we compare word choice in neighborhoods that are majority-White to those that are majority non-White. Figure 3 displays the results of this analysis. We find that within the rental logistics topic there is variation in the specific words being used in advertisements along racial lines. Specifically, words like "must", "credit", and "incom[e]", which imply a more exclusionary tone and focus on renter characteristics, predominate in advertisements in majority non-White neighborhoods. In contrast, in majority-White neighborhoods, we see words with a more neutral or welcoming tone and a focus on the general rental application process, including "will", "move", "applic[ation]", "leas[e]", and "free".

[Figure 3 about here]

Returning to the second column of Figure 2, we find that advertisements in poorer neighborhoods tend to have less language about unit amenities (for example, descriptions of building materials, appliances, etc.). Relative to same-race, non-poor neighborhoods, poor White, Black, Latino, and Asian neighborhoods all have about 20-30% less discussion of unit amenities (note the difference for Asian neighborhoods is not significant, perhaps due to small sample sizes). But like the logistics topic, we also find a clear racialized pattern of unit amenities language. Among non-poor neighborhoods, listings in Black and Latino neighborhoods are about 40% less likely to contain information about housing unit amenities compared to White neighborhoods. Similarly, listings in poor Black and Latino neighborhoods are over 40% less likely to contain information about housing unit amenities compared to poor White neighborhoods. In sum, compared to White neighborhoods, advertisements in Black and Latino neighborhoods over-emphasize renter qualifications and logistics, and under-emphasize unit amenities.

Discussion of neighborhood amenities (descriptions of nearby parks, restaurants, etc.) displays a more complex pattern. As shown in the final column of Figure 2, listings in poor neighborhoods tend to include *more* discussion of neighborhood amenities relative to their non-poor, same-race counterparts. The effect sizes vary from being about 25% more prevalent in poor Latino neighborhoods to being almost 90% more prevalent in poor White neighborhoods (and 100% more prevalent in poor Asian neighborhoods, though this estimate is very imprecise given our small sample size). Nevertheless, regardless of poverty status, listings in Black and Latino neighborhoods contain less neighborhood description compared to similar-poverty status White neighborhoods (both poor and non-poor). Models 1b, 2b, and 3b in Table 2 repeat our analysis but includes additional neighborhood covariates. While some effect sizes are attenuated, most differences by race and poverty status remain statistically significant.

[Table 2 here]

Why would advertisements for housing in poorer neighborhoods, and particularly in predominantly White and Asian neighborhoods with more college-educated residents, contain more language on neighborhood amenities? It may be due, in part, to landlords' desires to attract higher-SES renters. That is, in trying to attract higher-SES tenants who may be willing and able to pay more rent, landlords are incentivized to emphasize neighborhood amenities. While all lower-income neighborhoods have some gentrification potential, which could explain why we see more neighborhood amenities language in poor neighborhoods across racial composition, prior research has shown that poor non-Black neighborhoods are more likely than their Black counterparts to gentrify (Hwang 2015; 2016; Hwang and Sampson 2014; Timberlake and Johns-

Wolfe 2017) which could account for the variance in amenities language across neighborhoods of different racial compositions.

If this is the case, we would expect higher-priced rental units within poorer neighborhoods to be driving the observed greater prevalence of neighborhood amenities langauge. To test this, we first replicate Table 2 but classify listings based on neighborhood race and whether the listing is above (= high rent) or below (= low rent) the median asking rental price in their metro area (see Appendix Table A16). We find similar patterns for the logistics and unit amenities topics using listing rent as our SES measure compared to our neighborhood race by poverty typology. However, a distinct pattern emerges for neighborhood amenities: while we previously found that neighborhood amenities language is *more prevalent* in poorer neighborhoods, we find that it is *less* prevalent in lower-rent units. To better understand what is driving these different patterns, we next interact our binary measure of unit listing rent (high v. low) with our full, 8-cateogry neighborhood typology, creating 16 neighborhood categories in total (see Appendix Table A17 and Figure A6).

Even when we account for both neighborhood poverty and rent, large ethnoracial differences remain. With respect to neighborhood amenities, we find that listings in lower poverty neighborhoods still tend to have less such language and fewer differences by race or rent. Yet among higher poverty neighborhoods, we do find that higher-rent listings tend to have more neighborhood amenities language and that regardless of rent, listings in White and Asian neighborhoods tend to have more neighborhood amenities language. Altogether, these findings remain consistent with the racialized gentrification processes identified in prior research. Landlords listing higher rent units in potentially gentrifying neighborhoods tend to put more information about neighborhood amenities in their ads.

While our STM analysis identifies clear differences in information across neighborhoods depending on their racial/ethnic and socioeconomic composition as well as rent, each topic contains multiple words and it remains somewhat unclear exactly what words might systematically appear more/less frequently across neighborhoods. Thus, we next use MNIR to identify individual words that are correlated with neighborhood characteristics and to provide a robustness check on our STM findings. The MNIR results are detailed in Figure 4, which lists the words with the strongest correlations with three key neighborhood measures. We conduct MNIR with continuous neighborhood measures, examining the proportion of Black residents, the proportion of residents in poverty, and the proportion of college-educated residents. Additional analyses for other neighborhood race and SES measures show similar substantive findings and are reported in the Appendix. Note that, because of MNIR's limitations, these analyses consider just one neighborhood characteristic at a time.

For each measure we list the top 50 words with the highest association with each neighborhood characteristic. Figure 4 displays the words descending in order from strongest to weakest correlation (within this group of relatively highly correlated words). The correlation coefficients calculated in MNIR have no substantive meaning, and, in MNIR there are no standard or accepted cut-offs as to what constitutes a weak or strong correlation (unlike Pearsons' r). Thus, in MNIR correlation coefficients for specific words can only be interpreted relative to one another (see Appendix for MNIR coefficients).

[Figure 4 about here]

The MNIR results underscore how specific words clearly vary by neighborhood racial composition and socioeconomic status. Beginning with the language associated with a larger proportion of Black residents, we see words focused on renter characteristics and (dis)qualifications, such as: ‘evictions’, ‘section’ (short for Section 8), ‘criminal’ and ‘proof’ (of income). This pattern mirrors our findings for the logistics topic in our STM analysis. We also see words about affordability and finances, including: ‘discounts’, ‘affordable’, ‘money’, and ‘income’, though the correlation is not quite as strong; for example, ‘evictions’ is about 1.7 times more correlated with proportion Black than is ‘income’ (see Appendix). In other analyses not shown we confirm that these words are negatively correlated with the proportion of White residents.

We see a similar list when we examine words correlated with having a higher neighborhood poverty rate: ‘evictions’, ‘criminal’ (background), ‘section’ (8), and ‘proof’ (of income) are all prevalent and these words have some of the strongest associations with low-income neighborhoods. We also find similar words about affordability, including: ‘affordable’, and ‘income’, though again ‘evictions’ is about 1.6 times more correlated with proportion in poverty compared to ‘income’ (see Appendix). However, we also find that words related to college students, including ‘campus’, ‘students’, and ‘university’ are also highly correlated with neighborhood poverty. College students living in off campus housing often have little to no personal income, raising neighborhood poverty rates even when they receive familial/other non-income based financial support (Bishaw 2013). The correlation between neighborhood poverty and these words suggests that landlords in certain higher poverty neighborhoods might be targeting student renters and/or that in these neighborhoods poverty rates are high *because* they have a large number of student renters (Laidley 2014; Ehlenz 2019). More broadly, this variation across these two lists of words underscores the intersectional relationship between neighborhood race and poverty status.

Both lists starkly contrast with the words associated with a higher proportion of neighborhood residents with at least a college degree. Rather than mentioning renter qualifications or affordability, advertisements in neighborhoods with more college-educated residents tend to have words describing housing and neighborhood amenities. For example, words like ‘rooftop’, ‘concierge’, ‘marble’, ‘elevator’, and ‘backsplashes’ all describe housing unit/building amenities that are generally high-end. Other words seem to describe neighborhood or location amenities, such as ‘whole’ and ‘foods’ (as in Whole Foods), ‘museum’, ‘nightlife’, and ‘yoga’. Again, not only do these words appear to be more focused on neighborhood/location characteristics than do the words associated with large Black populations or more poor households, but they also imply a certain type of neighborhood associated with higher-SES lifestyles and amenities.

The MNIR results offer additional evidence that the content of advertisements depends on a neighborhood’s socioeconomic status and racial composition. In predominantly Black and Latino and/or poor neighborhoods, we find much less evidence of any discussion of the quality/amenities of the housing unit and neighborhood; instead, listings in these neighborhoods focus on affordability and renter qualifications. The emphasis on affordability and renter qualifications likely attracts some prospective renters and repels others; relatedly, the lack of emphasis on unit and neighborhood amenities may prevent some prospective renters from considering housing in these neighborhoods. In contrast, in neighborhoods with large proportions of highly-educated residents, advertisements do not tend to mention renter qualifications nor affordability. This does not mean that in these neighborhoods there are no expected or required renter qualifications; rather, it seems that landlords listing properties in these neighborhoods do not feel the need to mention them when soliciting renters.

It is also important to note the absence of certain words in both the MNIR and STM analyses: we find no evidence of explicit racial/ethnic words, perhaps because fair housing laws and Craigslist posting policies largely prohibit them. Previous studies of overt discrimination in online housing ads have found higher rates of discrimination against renters with children than any other protected category (Oliveri 2010) and various other forms of discrimination—racial steering, different response rates to inquiries from different raced homeseekers—remain prevalent (Besbris 2020; Hanson and Hawley 2011; Hogan and Berry 2011). Yet the absence of explicit racial/ethnic words in our sample underlines the potential additional influence of the more subtle language differences that we have identified.

DISCUSSION

Our results highlight the importance of understanding the information environment in which housing searches take place for demographic research on residential mobility. Mobility decisions are predicated on available information and, as shown above, the information sharing practices of landlords do not equalize information across neighborhoods. Multiple analyses reveal that advertisements for units in neighborhoods with more Black/Latino and/or poorer residents tend to have less language describing unit amenities and relatively more language devoted to tenant (dis)qualifications compared to ads from Whiter and/or lower-poverty neighborhoods. Even in low-poverty Black and Latino neighborhoods advertisements disproportionately focus on renter (dis)qualifications rather than unit amenities. In contrast, advertisements for housing in White and Asian neighborhoods are more likely to include positive descriptions of neighborhood characteristics; this is particularly true for higher-rent listings in poor White and Asian neighborhoods which may be undergoing—or poised to undergo—gentrification. Indeed, recent research has demonstrated how the gentrification potential of neighborhoods depend on their existing racial composition (Hwang 2015; 2016). While a key limitation of our data is that it is cross sectional, future research could scrape advertisements longitudinally to pinpoint the relationship between changing information and changing neighborhood demographics, as well as further unpack how other types of variation in neighborhoods (i.e., levels of diversity or segregation) are reflected in advertisements.

To illustrate our findings, Figure 5 (below), presents a pair of maps of central St. Louis, MO. The maps, which show rates of Black residents and the prevalance of the logistics topic, reveals the socio-spatial distribution of information on Criagslist. The maps also contain two dots indicating where the advertisements in Figure 6 originated from.[24] The first ad in Figure 6 is a listing for an apartment in a predominantly White neighborhood (about 50% of residents) with relatively high poverty levels that has been experiencing upscaling and demographic changes. The listing has multiple paragraphs which describe the housing amenities (like its 'open layout'). It also includes a description of the neighborhood, citing how it is close to multiple universities and near an area with 'hustle and bustle'.

[Figures 5 and 6 about here]

The second is a listing for an apartment in a predominantly Black neighborhood (about 90% of residents) with relatively high poverty levels. The text in this listing outlines multiple renter qualifications. As prospective renters look through their options on Craigslist, in the areas surrounding Figure 5 they will see additional, similar listings with text on both housing and neighborhood amenities. In the areas surrounding Figure 6 they also see other listings which include little if any information about the housing or neighborhood but a long list of renter

[24] These posts were captured after data collection for the results presented here had ended but are qualitatively similar to posts included in our dataset. Both listings were posted within two days of one another during the summer of 2018.

disqualifications. While St. Louis follows the overall trends identified in our sample, future work should test for differences across MSAs based on their levels of segregation and other characteristics to better understand whether and how the racialization of housing information depends on local conditions and histories (Kennedy et al. 2020).

If, as Krysan and Crowder (2017) convincingly argue, most prospective renters are likely to search for housing in neighborhoods they are familiar with through their lived experiences and social networks, then our results suggest that renters are receiving very different information about units and their surroundings depending on the neighborhoods in which they search. Black and Latino renters, who predominantly search for housing in Black or Latino neighborhoods, encounter strong messaging about their qualifications. Additionally, these differences—as well as the biased spatial availability of advertisements—reduce search costs for searchers in Whiter neighborhoods, who are more likely to be White, and expand their mobility options (Boeing 2020).

If at least some prospective (White, non-poor) renters are open to considering a more heterogeneous group of neighborhoods during the initial stages of their search, the different types of information searchers are exposed to likely influence their decisions. Language about tenant qualifications, which predominates in Black and Latino neighborhoods, could drive away potential renters who can afford to look elsewhere, particularly since these ads tend to lack text on amenities. More broadly, such differences in information contribute to the formation and maintenance of place reputations—acting in concert with homeseekers' existing information to reify perceptions of certain neighborhoods as more or less appropriate for different demographic groups (Krysan and Crowder 2017).

Crucially, future work should test the effects of language differences described here and explore the myriad potential consequences of our findings for patterns of integration/segregation. We have outlined various ways these differences may matter, but experiments could be used to measure to what extent homeseekers associate certain types of language with particular neighborhood demographics. While there is a growing interest in *how* homeseekers make their residential mobility decisions, more work is needed on the relative importance of various sources of information. In other words, how do homeseekers weigh information found online compared to other sources? What is clear from our findings is that online search tools do not serve to erase information differences about available housing across neighborhoods and, as a result, likely foster existing demographic differences in residential mobility.

REFERENCES

Adelman, R. M. (2005) The Roles of Race, Class, and Residential Preferences in the Neighborhood Racial Composition of Middle-Class Blacks and Whites. *Social Science Quarterly*, 86, 209-228.

Anderson, M. & Perrin, A. (2016) 13% of Americans don't use the internet. Who are they? *Pew Research* < http://www.pewresearch.org/fact-tank/2016/09/07/some-americans-dont-use-theinternet-who-are-they/>.

Bader, M. D.M. & Krysan, M. (2015) Community Attraction and Avoidance in Chicago: What's Race Got to Do with It? *ANNALS*, 660, 261-281.

Besbris, M. (2016) Romancing the Home: Emotions and the Interactional Creation of Demand in the Housing Market. *Socio-Economic Review*, 14, 461-482.

Besbris, M. (2020) *Upsold: Real Estate Agents, Prices, and Neighborhood Inequality*. Chicago: University of Chicago Press.

Besbris, M. & Faber, J. W. (2017) Investigating the Relationship between Real Estate Agents, Segregation, and House Prices: Steering and Upselling in New York State. *Sociological Forum*, 3, 850-873.

Besbris, M., Faber, J. W., & Sharkey, P. (2019) Disentangling the Effects of Race and Place in Economic Transactions: Findings from an Online Field Experiment. *City & Community*, 18, 529-555.

Besbris, M., Faber, J. W., Rich, R., & Sharkey, P. (2015) Effect of Neighborhood Stigma on Economic Transactions. *Proceedings of the National Academy of Sciences*, 112, 4994-4998.

Besbris, M., Faber, J. W., Rich, R., & Sharkey, P. (2018) The Geography of Stigma: Experimental Methods to Identify the Penalty of Place. In S. M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 159-177). New York: Springer.

Bischoff, K, & Owens, A. (2019) The Segregation of Opportunity: Social and Financial Resources in the Educational Contexts of Lower- and Higher-Income Children, 1990-2014. *Demography*, 56, 1635-1664.

Bishaw, A. (2013). Examining the effect of off-campus college students on poverty rates. Washington, DC: U.S. Census Bureau. http://www.census.gov/hhes/www/poverty/publications/bishaw.pdf

Boeing, G. (2020) Online Rental Housing Market Representation and the Digital Reproduction of Urban Inequality. *Environment and Planning A: Economy and Space*, 47, 590-608.

Boeing, G., Besbris, M., Schachter, A., & Kuk., J. (2020) Housing Search in the Age of Big Data: Smarter Cities or Same Old Blind Spots? *Housing Policy Debate*, doi: 10.1080/10511482.2019.1684336

Boeing, G., & Waddell, P. (2017) New Insights into Rental Housing Markets across the United States: Web Scraping and Analyzing Craigslist Rental Listings. *Journal of Planning Education and Research*, 37, 457-476.

Brannon, M. M. (2017) Datafied and Divided: Techno-Dimensions of Inequality in American Cities. *City & Community*, 16, 20-24.

Bruch, E., & Feinberg, F. (2017) Decision-Making Processes in Social Contexts. *Annual Review of Sociology* 43, 207-227.

Bruch, E., Feinberg, F., & Lee, K. Y. (2016) Extracting Multistage Screening Rules from Online Dating Activity Data. *Proceedings of the National Academy of Sciences*, 113, 10530-10535.

Bruch, E. E., & Mare, R. D. (2006) Neighborhood Choice and Neighborhood Change. *American Journal of Sociology*, 112, 667-709.

Bruch, E., & Swait, J. (2019) Choice Set Formation in Residential Mobility and Its Implications for Segregation Dynamics. *Demography*, 56, 1665-1692.

Carrillo, L., Pattillo, M., Hardy, E., & Acevedo-Garcia, D. (2016) Housing Decisions among Low-Income Hispanic Households in Chicago. *Cityscape*, 18, 109-150.

Cesare, N., Lee, H., McCormick, T., Spiro, E., & Zagheni, E. (2018) Promises and Pitfalls of Using Digital Traces for Demographic Research. *Demography*, 55, 1979-1999.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009) Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems*: 288-296.

Charles, C. Z. (2006) *Won't You Be My Neighbor? Race, Class, and Residence in Los Angeles*. New York: Russell Sage.

Chetty, R., & Hendren, N. (2018) The Impacts of Neighborhood on Intergenerational Mobility I: Childhood Exposure Effects. *Quarterly Journal of Economics*, 133, 1107-1162.

Clark, W. A. V. (1992) Residential Preferences and Residential Choices in a Multi-Ethnic Context. *Demography* 30, 451-466.

Clark, W A. V. (2009) Changing Residential Preferences Across Income, Education, and Age: Findings from a Multi-City Study of Urban Inequality. *Urban Affairs Review*, 44, 334-355.

Clark, W A. V. & Morrison, P. (2012) Socio-Spatial Mobility and Residential Sorting: Evidence from a Large Scale Survey. *Urban Studies*, 39, 3253-3270.

Crowder, K., & South, S. J. (2005) Race, Class, and Changing Patterns of Migration Between Poor and Nonpoor Neighborhoods. *American Journal of Sociology*, 110, 1715-1763.

Crowder, K., & South, S. J. (2008) Spatial Dynamics of White Flight: The Effect of Local and Extralocal Racial Conditions on Neighborhood Out-Migration. *American Sociological Review*, 73, 792-812.

Courant, P. N. (1978) Racial Prejudice in a Search Model of the Urban Housing Market. *Journal of Urban Economics*, 5, 329-345.

DeLuca, S., Garboden, P. M., & Rosenblatt, P. (2013) Segregating shelter: How housing policies shape the residential locations of low-income minority families. *The ANNALS of the American Academy of Political and Social Science*, 647, 268-299.

Desmond, M. (2016) *Evicted.* New York NY: Crown Publishers.

Desmond, M., & Shollenberger, T. (2015) Forced Displacement from Rental Housing: Prevalence and Neighborhood Consequences. *Demography*, 52, 1751-1772.

DiMaggio, P., Nag, M., & Blei, D. M. (2013) Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding. *Poetics*, 41, 570–606.

Ehlenz, M. M. (2019) The University and its Neighborhood: A Study of Place-Making and Change. *Journal of Urban Affairs*, 41, 776-794.

Ellen, I. G., & Karfunkel, B. (2016) Renting in America's Largest Metropolitan Areas. *NYU Furman Center/Capital One National Affordable Housing Landscape* http://furmancenter.org/files/CapOneNYUFurmanCenter__NationalRentalLandscape_MAY2015.pdf

Friedman, S., Tsao, H., & Chen, C. (2013) Housing Tenure and Residential Segregation in Metropolitan America. *Demography*, 50, 1477-1498.

Gabriel, R., & Spring, A. (2019) Neighborhood Diversity, Neighborhood Affluence: An Analysis of the Neighborhood Destination Choices of Mixed-Race Couples with Children. *Demography*, 56, 1051-1073.

Garboden, P. M. E., & Rosen, E. (2019) The Threat of Eviction: How Landlords Shape a Contingent Tenure. *City & Community*, 18, 638-661.

Gentzkow, M., & Shapiro, J. M.. (2006) Media bias and reputation. *Journal of Political Economy*, 114, 280-316.

Gentzkow, M., Kelly, B. T., & Taddy, M. (2019) Text as Data. *Journal of Economic Literature*, 57, 535-574.

Grimmer, J. (2012) Evaluating Model Performance in Fictitious Prediction Problems. Discussion of "Multinomial Inverse Regression for Text Analysis" by Matthew Taddy. Unpublished manuscript.

Hanson, A., & Hawley, Z. (2011) Do landlords discriminate in the rental market? *Journal of Urban Economics*, 70, 99-114.

Harvey, H., Fong, K., Edin, K., & DeLuca S. (2020) Forever Homes and Temporary Stops: Housing Search Logics and Residential Selection. *Social Forces*, 98, 1498-1523.

Havekes, E., Bader, M., & Krysan, M. (2016) Realizing Racial and Ethnic Neighborhood Preferences? Exploring the Mismatches Between What People Want, Where They Search, and Where They Live. *Population Research and Policy Review*, 35, 101-126.

Hogan, B., & Berry, B. (2011) Racial and Ethnic Biases in Rental Housing. *City & Community*, 10, 351-372.

Howell, J., & Emerson, M. O. (2018) Preserving Racial Hierarchy Amidst Changing Racial

Demographics: How Neighbourhood Racial Prferences are Changing While Maintaining Segregation. *Ethnic and Racial Studies*, 41, 2770-2789.

Hwang, J. (2015) Gentrification in Changing Cities: Immigration, New Diversity, and Racial Inequality in Neighborhood Renewal. *ANNALS,* 660, 319-340.

Hwang, J. (2016) Pioneers of Gentrification: Transformation in Global Neighborhoods in Urban America in the Late Twentieth Century. *Demography*, 53, 189-203.

Hwang, J., & Sampson, R. J. (2014) Divergent Pathways of Gentrification: Racial Inequality and the Social Order of Renewal in Chicago Neighborhoods. *American Sociological Review*, 79, 725-751.

JCHS. 2019. "The State of the Nations Housing 2019." http://www.jchs.harvard.edu/research/state_nations_housing

Kennedy, I., Hess, C., Paullada, A., & Chasins, S. 2020. Racialized Discourse in Seattle Rental Ad Text. *Social Forces* doi: 10.1093/sf/soaa075/5879475

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

Krysan, M. (2002) Whites Who Say They'd Flee: Who Are They, and Why Would They Leave. *Demography*, 39, 675-696.

Krysan, M., & Bader, M. (2009) Racial Blind Spots: Black-White-Latino Differences in Community Knowledge. *Social Problems*, 56, 677-701.

Krysan, M., & Crowder, K. (2017) *Cycle of Segregation*. New York: Russell Sage.

Laidley, T. M. (2014) The Privatization of College Housing: Poverty, Affordability, and the U.S. Public University. *Housing Policy Debate*, 24, 751-768.

Lareau, A. (2014) Schools, Housing, and the Reproduction of Inequality. In A. Lareau & K. Goyette (Eds.) *Choosing Homes, Choosing Schools* (pp. 169-206). New York: Russell Sage.

Lee, B. A., Oropesa, R. S., & Kanan, J. W. (1994) Neighborhood Context and Residential Mobility. *Demography*, 31, 249-270.

Logan, J. R., & Alba, R. (1993) Locational Returns to Human Capital: Minority Access to Suburban Community Resources. *Demography*, 30, 243-268.

Manning, C. D., Raghavan, P., & Schütze, H. (2008) *Introduction to information retrieval*. Cambridge: Cambridge University Press.

Massey, D. S. (2005) Racial Discrimination in Housing: A Moving Target. *Social Problems*, 52, 148-151.

Massey, D. S., & Denton, N. A. (1993) *American Apartheid: Segregation and the Making of the Underclass*. Cambridge: Harvard University Press.

McAvay, H. (2018) How Durable are Ethnoracial Segregation and Spatial Disadvantage? Intergenerational Contextual Mobility in France. *Demography*, 55, 1507-1545.

McLaughlin, R., & Young, C. (2018) Data Democratization and Spatial Heterogeneity in the Housing Market. In *A Shared Future: Fostering Communities of Inclusion in an Era of Inequality*. Cambridge: Harvard Joint Center for Housing Studies. http://www.jchs.harvard.edu/sites/jchs.harvard.edu/files/a_shared_future_data_democratization_spatial_heterogeneity.pdf

National Multifamily Housing Coalition. (2016) Quick Facts: Resident Demographics. < http://www.nmhc.org/Content.aspx?id=4708>

Oliveri, R. C. 2010. Discriminatory Housing Advertisements On-Line: Lessons from Craigslist. *Indiana Law Review*, 43, 1125-1183.

Palm, R., & Danis, M. A. (2001) Residential Mobility: The Impacts of Web-Based Information on the Search Process and Spatial Neighborhood Choice Patterns. *Urbam Geography*, 22, 641-655.

Pilkauskas, N., & Michelmore, K. (2019) The Effect of the Earned Income Tax Credit on Housing and Living Arrangments. *Demography*, 56, 1303-1326.

Quillian, L., & Pager, D. (2001) Black Neighbors, Higher Crime? The Role of Racial Stereotypes in Evaluations of Neighborhood Crime. *American Journal of Sociology* 107, 717-767.

Rae, A. (2015) Online housing search and the geography of submarkets. *Housing Studies*, 30, 453-472.

Roberts, M. E. et al. (2014) Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58, 1064–1082.

Rosen, E. (2014) Rigging the Rules of the Game: How Landlords Geographically Sort Low-Income Renters. *City & Community*, 13, 310-340.

Rosen, E. (2017) Horizontal Immobility: How Narratives of Neighborhood Violence Shape Housing Decisions. *American Sociological Review*, 82, 270-296.

Rosenblatt, P., & DeLuca, S. (2012) 'We Don't Live Outside, We Live in Here': Neighborhood and Residential Mobility Decisions Among Low-Income Families. *City & Community*, 11, 254-284.

Sampson, R. J. (2012) *Great American City.* Chicago: University of Chicago Press.

Sampson, R. J., & Raudenbush, S. W. (2004) Seeing disorder: Neighborhood stigma and the social construction of 'broken windows'. *Social Psychology Quarterly*, 67, 319-342.

Sampson, R. J., & Sharkey, P. (2008) Neighborhood Selection and the Social Reproduction of Concentrated Racial Inequality. *Demography*, 45, 1-29.

Schachner, J. N., & Sampson R. J. (2020) Skill-Based Contextual Sorting: How Parental Cognition and Residential Mobility Produce Unequal Environments for Children. *Demography* 57, 675-703.

Schachter, A., & Besbris, M. (2017) Immigration and Neighborhood Change: Methodological Possibilities for Future Research. *City & Community*, 16, 244-251.

Sharkey, P., & Faber, J. (2014) Where, When, Why, and For Whom Do Residential Contexts Matter? Moving Away from the Dichotomous Understanding of Neighborhood Effects. *Annual Review of Sociology*, 40, 559-579.

Stiel, J., & Jordan, R. (2018) Household Neighborhood Decision Making and Segregation. In *A Shared Future: Fostering Communities of Inclusion in an Era of Inequality*. Cambridge: Harvard Joint Center for Housing Studies. http://www.jchs.harvard.edu/sites/default/files/A_Shared_Future_Chapter_5_Household_Neighborhood_Decisionmaking.pdf

Swaroop, S., & Krysan, M. (2011) The Determinants of Neighborhood Satisfaction: Racial Proxy Revisted. *Demography*, 48, 1203-1229.

Taddy, M. (2013) Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, 108, 755–770.

The American Panel Survey. 2017. http://taps.wustl.edu/

The American Housing Survey. 2017. https://www.census.gov/programs-surveys/ahs.html

Timberlake, J. M., & Johns-Wolfe, E. (2017) Neighborhood Racial Composition and Gentrification in Chicago and New York 1980 to 2010. *Urban Affairs Review*, 53, 236-272.

van Ham, M., Boschman, S., & Vogel, M. (2018) Interpreting Neighborhood Choice in a Model of Neighborhood Effects on Income. *Demography*, 55, 1069-1090.

Varian, H. R. (2014) Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28, 3-28.

Wang, Q., Phillips, N. E., Small, M. L., & Sampson R. J. (2018) "Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences*, 115, 7735–7740.

Wood, H. (2014) When Only a House Makes a Home: How Home Selection Matters in the Residential Mobility Decisions of Lower-Income, Inner-City African American Families. *Social Service Review*, 88, 264-294.

Logistics Topic

This house will not last long. Looking for a motivated renter with NO evictions, GOOD rental history, and income must be $ per month. Application fees are waived for SERIOUS applicants. (91.9%)

You will love to call this house home! $ Deposit holds it and $ per month rent is what gets you in. Must have verifiable income and no past evictions within the last yrs. (90.9%)

Unit Amenities Topic

ad# ft. Ceilings Hardwoodood Style Flooring* Island Kitchens* Upgraded " Cabinets Deep Kitchen Sinks with Gooseneck Faucets Granite Counter Tops Stainless Steel Appliances Washer and Dryer Included Separate Walk−In Showers* Live−Work Units Available* Large Walk−In Closets (90.1%)

Granite counter tops washer and dryer ceiling fans wood floors, Stainless steel kitchen appliances Modern accents and LED Lighting (87.4%)

Neighborhood Amenities Topic

* Located close to public transportation with easy access to freeways /, Caltrain and corporate shuttles * Blocks from the AT&T stadium, UCSF campus, a farmer's market, Safeway, Whole Foods, eateries, and nightlife (90.0%)

Watch your step as you might trip from having too much fun walking to campus with your friends while staying at the Fine Arts building in the hippest place on earth, Berkeley! (89.3%)

Figure 1: Example Listings of Three Main Topics

Note: The numbers inside parentheses represent the proportion of the relevant topic in the listing. For example, the first listing in the figure on the left has 91.9% of logistics topic in the listing.

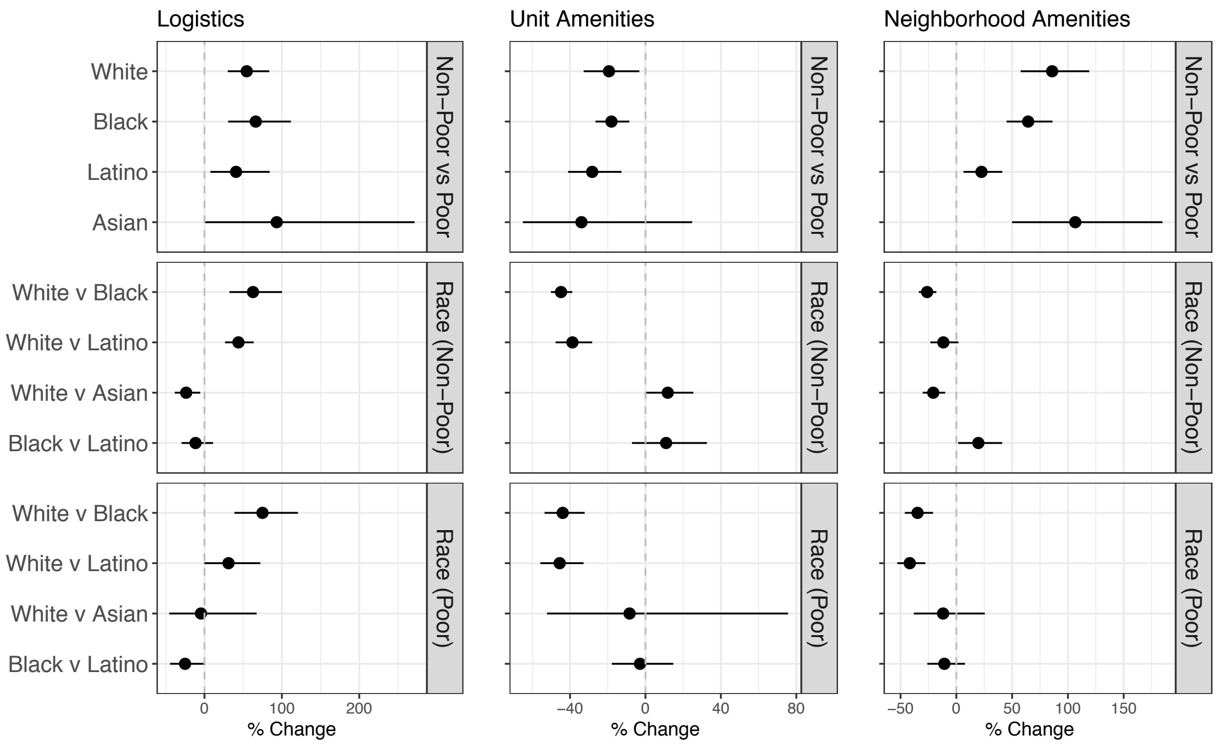

Figure 2: Pairwise Comparison of Topic Proportions Across Neighborhood Types

Note: The dependent variables are log transformed. The dots and bars indicate the percent change and confidence intervals derived from regression models. For the ease of interpretation, we present percent change instead of regression coefficients. The neighborhood written first (before "vs") is the base category. The first four rows (non-poor vs poor) display the regression coefficients for poor neighborhoods when the non-poor neighborhoods for the respective racial group is the base category. Negative value means poor neighborhoods have less information than non-poor neighborhoods. Positive value means non-poor neighborhoods have less information than poor neighborhoods. The following eight rows compare differences between different racial compositions. The first racial group is the base category. For example, the coefficient for "White v Latino" in the sixth row for the unit amenities topic indicates that Latino non-poor neighborhoods have 38.7 percent less topic proportions in unit amenities than White non-poor neighborhoods. Neighborhood racial composition and poverty rate are obtained from 2016 ACS 5-year pooled data. The regression models include MSA fixed effects. Standard errors are clustered at MSA level. Plots are based on results presented in Table 2.

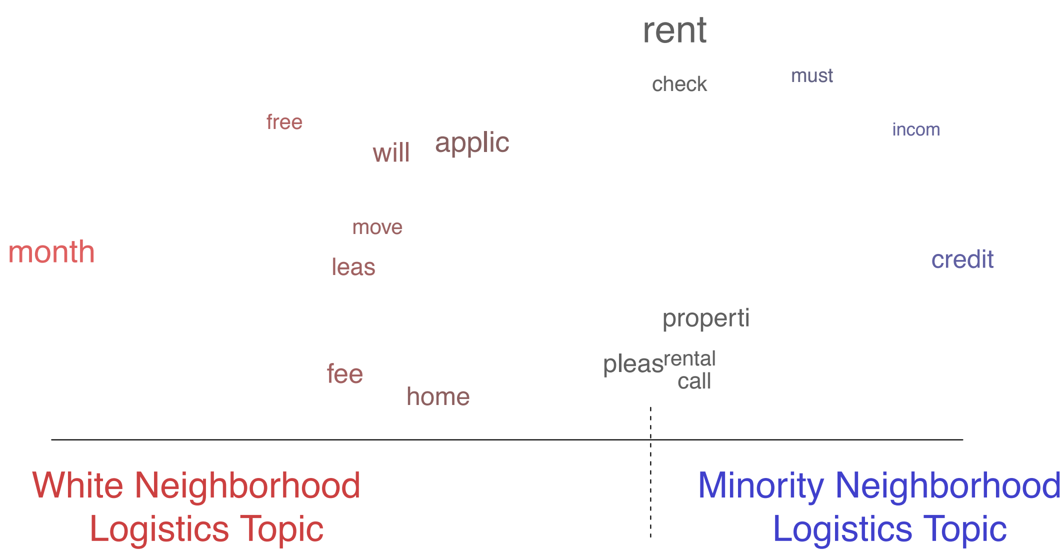

Figure 3: Choice of Words within the Logistics Topic by Neighborhood Racial Composition

Note: Words that appear in the right hand side of the figure are more likely to appear in majority non-White neighborhoods. The likelihood increases as the word is located farther right. Words that are located in the left of the dashed line have higher likelihood to be used in majority White neighborhoods. The size of the words represents how frequently the word will appear in the logistics topic.

Figure 4: Words with Top 50 Correlation with Neighborhood Covariates

Note: The size of the word is proportional to the coefficient from MNIR. If the size of the word is two times larger than the other one, it means the coefficient is two times higher.

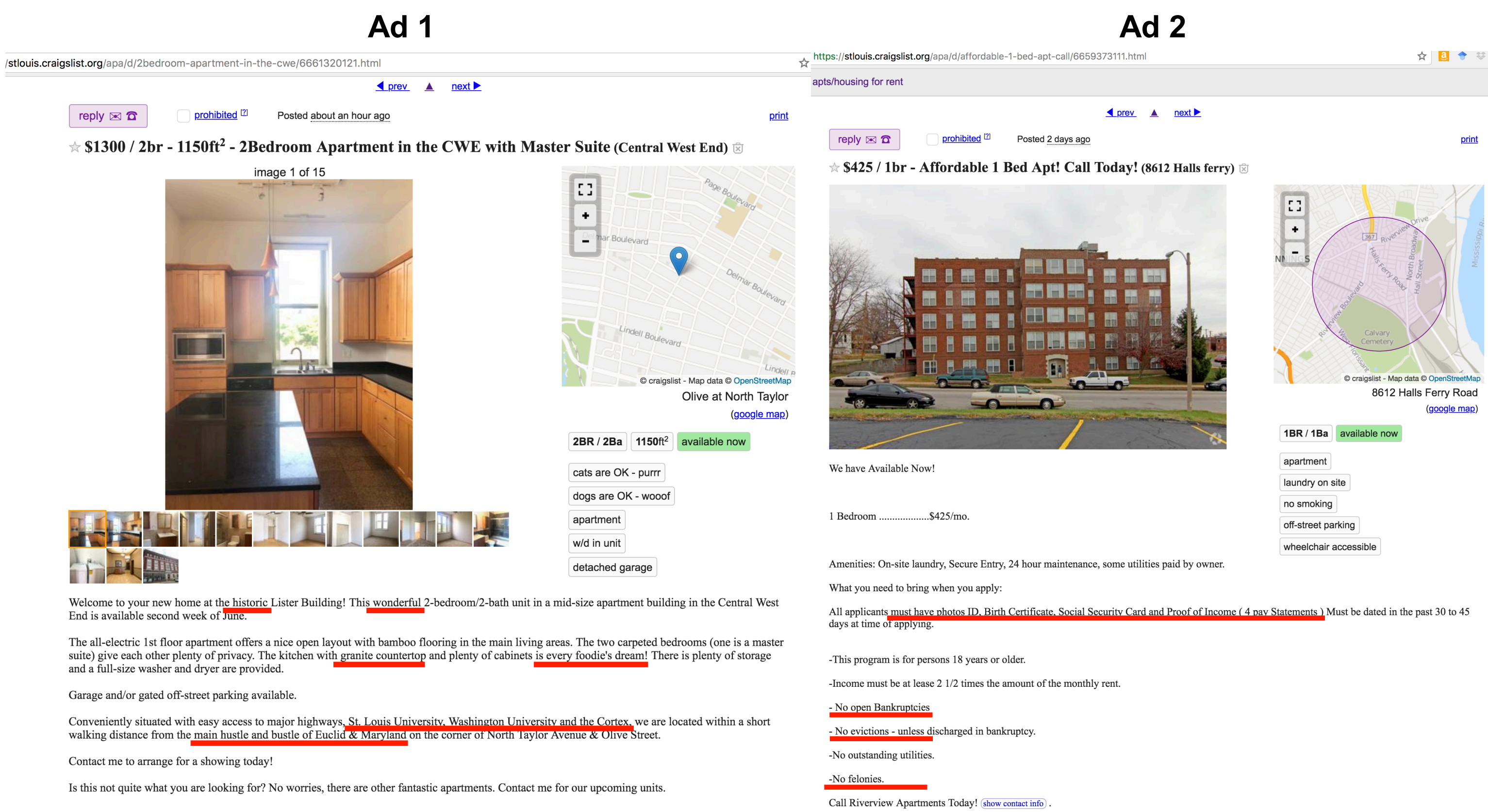

Figure 5: Real Housing Advertisement Posted from Predominantly White Neighborhood with Relatively High Poverty Levels in St. Louis, MO (Left) and from Predominantly Black Neighborhood with Relatively High Poverty Levels in St. Louis, MO (Right)

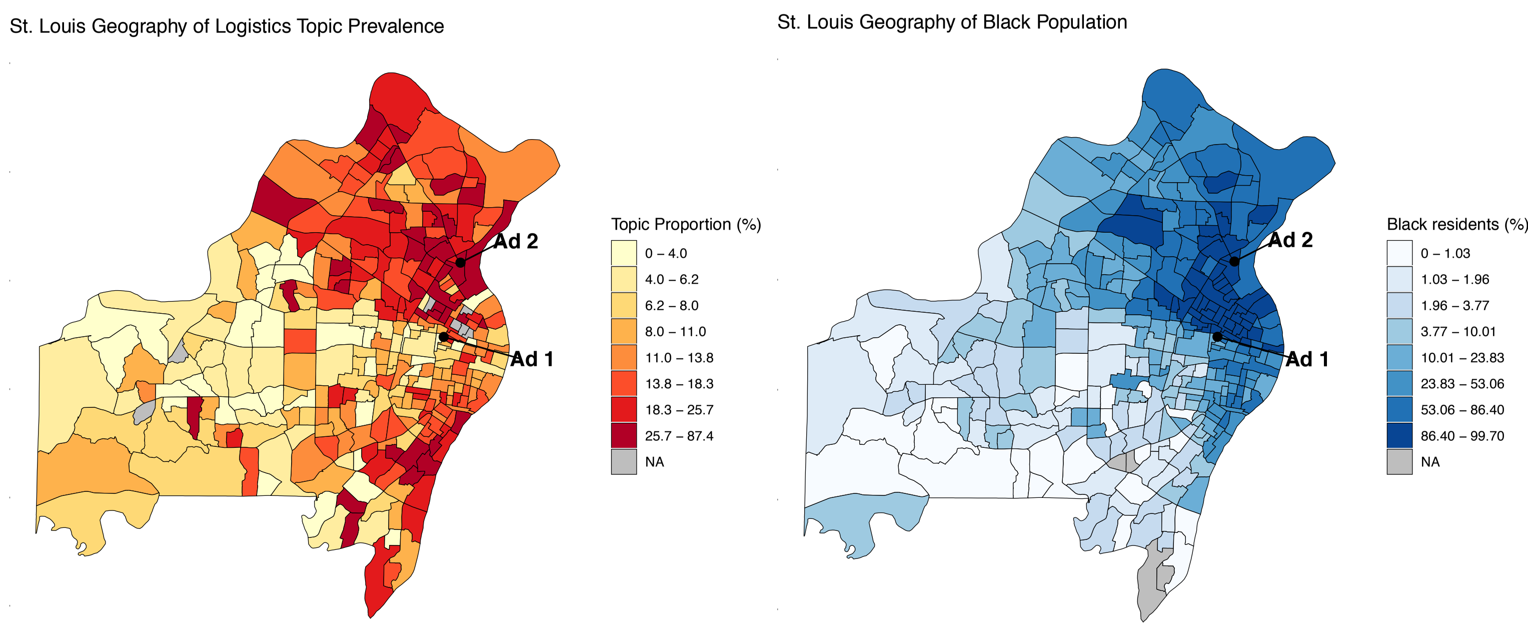

Figure 6: St. Louis, MO Geography of Logistics Topic Proportion (Left) and Black Population (Right)

Table 1: STM Topics from Craigslist Rental Listings

Label	Words	%
General	apart, home, communiti, center, pool, call, offer, locat, fit, bedroom	24.6
Logistics	rent, month, applic, will, fee, home, credit, pleas, properti, leas	10.6
Unit Amenity	kitchen, floor, center, applianc, room, communiti, loung, fit, stainless, area	12.7
Unit Description	bedroom, room, floor, kitchen, new, includ, larg, bath, bathroom, month	25.2
Pet Policy	pet, apart, home, communiti, polici, offic, restrict, now, hour, hous	8.3
Neighborhood Amenity	park, apart, locat, downtown, walk, shop, restaur, citi, minut, street	10.2
Availability	apart, avail, leas, today, price, unit, manag, chang, properti, call	8.5

Table 2: Regression Results Predicting Topic Proportions with Additional Neighborhood Covariates

	Dependent variable: Log Transformed Topic Proportion					
	Logistics Topic		Unit Amenities Topic		Neighborhood Amenities Topic	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Neighborhood Type						
White Poor	0.436**	0.520**	−0.215*	−0.234**	0.620**	0.382**
	(0.088)	(0.084)	(0.093)	(0.078)	(0.084)	(0.075)
Black Non-poor	0.487**	0.195+	−0.594**	−0.164**	−0.304**	−0.036
	(0.106)	(0.104)	(0.052)	(0.051)	(0.054)	(0.061)
Black Poor	0.995**	0.497**	−0.792**	−0.239**	0.193**	0.280**
	(0.112)	(0.122)	(0.063)	(0.066)	(0.064)	(0.064)
Latino Non-poor	0.364**	0.362**	−0.490**	−0.095	−0.124+	0.347**
	(0.065)	(0.084)	(0.080)	(0.072)	(0.072)	(0.055)
Latino Poor	0.706**	0.664**	−0.822**	−0.305**	0.080	0.478**
	(0.140)	(0.148)	(0.077)	(0.079)	(0.074)	(0.076)
Asian Non-poor	−0.270*	0.236*	0.112+	−0.081	−0.232**	−0.057
	(0.109)	(0.114)	(0.058)	(0.065)	(0.065)	(0.092)
Asian Poor	0.389	0.802**	−0.303	−0.393+	0.494**	0.439**
	(0.288)	(0.269)	(0.317)	(0.216)	(0.160)	(0.078)
Unit and Neighborhood Covariates						
Price ($1000)		−0.114**		0.336**		−0.005
		(0.026)		(0.037)		(0.016)
% College		−0.013**		0.017**		0.018**
		(0.002)		(0.001)		(0.001)
% Foreign Born		−0.017**		0.004*		−0.010**
		(0.004)		(0.002)		(0.002)
% Units Renter Occ		−0.009**		0.007**		0.010**
		(0.001)		(0.001)		(0.001)
% Unit Built after 2010		−0.013*		0.047**		−0.004
		(0.005)		(0.004)		(0.002)
% Vacancy		2.691**		−0.911**		1.592**
		(0.269)		(0.192)		(0.212)
MSA Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	1,692,639	1,692,639	1,692,639	1,692,639	1,692,639	1,692,639
Number of Census Tracts	37,319	37,319	37,319	37,319	37,319	37,319
R^2	0.179	0.230	0.231	0.349	0.241	0.346
Adjusted R^2	0.179	0.230	0.231	0.349	0.241	0.346
Residual Std. Error	1.664	1.611	1.342	1.235	1.104	1.025

$^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$

Note: Topic proportions for the three topics are computed by STM (Roberts et al. 2014). Dependent variables are log transformed. The base category for neighborhood type is White non-poor neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 30% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at census tract level.

Appendix Contents

A. Understanding CL Data

A-1 Craigslist Market vs Census MSA Definition

For most metros, the corresponding Craigslist site closely matches Census MSA definitions. There are a few discrepancies between Craigslist market and Census MSA definitions. Because we only use Census data at the tract level and follow Craigslist market definitions to determine metro area boundaries, any discrepancies do not impact our results. For example, the 'SFBay' Craigslist site covers both the San Francisco-Oakland-Hayward, CA MSA and the San Jose-Sunnyvale-Santa Clara, CA MSA. We refer to this site as the San Francisco Bay Area. Similarly, while the Census treats Miami, Fort Lauderdale, and West Palm Beach, FL as one MSA, in May of 2017 when we began data collection, each of these areas had its own Craigslist site. During data collection, Craigslist switched to using just one site, with the Fort Lauderdale and West Palm Beach sites now redirecting to the main Miami site. We combine all unique listings from these three sites and refer to them as the Miami metro area. The Los Angeles Craigslist covers Los Angeles County area rather than Los Angeles-Long Beach-Anaheim MSA, which includes Orange County. Craigslist has a separate site for Orange County, which we do not include as part of our Los Angeles metro area. While these examples demonstrate that Craigslist and Census data do not follow identical definitions of metro areas, in our paper we only use Census data at the tract level, and follow Craigslist market definitions to determine metro area boundaries for our MSA fixed effects, so these discrepancies do not impact our results.

A-2 Data Distribution across MSAs, Tracts, and Neighborhood Types

Table A1: Number of Listings by MSA

	MSA
Atlanta-Sandy Springs-Roswell	41974
Austin-Round Rock	29572
Baltimore-Columbia-Towson	45833
Birmingham-Hoover	21892
Boston-Cambridge-Newton	18746
Buffalo-Cheektowaga-Niagara Falls	22816
Charlotte-Concord-Gastonia	44128
Chicago-Naperville-Elgin	27501
Cincinnati	33301
Cleveland-Elyria	25635
Columbus	36046
Dallas-Fort Worth-Arlington	39943
Denver-Aurora-Lakewood	48877
Detroit-Warren-Dearborn	28637
Hartford-West Hartford-East Hartford	25035
Houston-The Woodlands-Sugar Land	35970
Indianapolis-Carmel-Anderson	34631
Jacksonville	34982
Kansas City	23029
Las Vegas-Henderson-Paradise	25894
Los Angeles-Long Beach-Anaheim	46669
Louisville-Jefferson County	29548
Memphis	25155
Miami-Fort Lauderdale-West Palm Beach	50896
Milwaukee-Waukesha-West Allis	30336
Minneapolis-St. Paul-Bloomington	39494
Nashville-Davidson-Murfreesboro-Franklin	35577
New Orleans-Metairie	29525
New York-Newark-Jersey City	42814
Oklahoma City	26889
Orlando-Kissimmee-Sanford	37223
Philadelphia-Camden-Wilmington	38382
Phoenix-Mesa-Scottsdale	37176
Pittsburgh	32074
Portland-Vancouver-Hillsboro	51815
Providence-Warwick	26694
Raleigh	46891
Richmond	38892
Riverside-San Bernardino-Ontario	42583
Sacramento-Roseville-Arden-Arcade	42214
Salt Lake City	46315
San Antonio-New Braunfels	25808
San Diego-Carlsbad	47671
San Francisco-Oakland-Hayward	54491
Seattle-Tacoma-Bellevue	52170
St. Louis	32525
Tampa-St. Petersburg-Clearwater	36588
Virginia Beach-Norfolk-Newport News	35220
Washington-Arlington-Alexandria	55669

Table A2: Number of listings in Craigslist and number of census tracts in the entire top 50 MSAs per neighborhood type: Plurality, 30% Poverty

Type	Number of Listings (%)	Number of Tracts in Top 50 MSAs (%)
White Nonpoor	1,196,496	24,643
	(69.73)	(64.10)
White Poor	79,115	976
	(4.61)	(2.54)
Black Nonpoor	137,838	3,081
	(8.03)	(8.01)
Black Poor	77,439	2,148
	(4.51)	(5.59)
Latino Nonpoor	142,051	4,755
	(8.28)	(12.37)
Latino Poor	50,162	1,734
	(2.92)	(4.51)
Asian Nonpoor	28,982	1,011
	(1.69)	(2.63)
Asian Poor	3,730	95
	(0.22)	(0.25)

B. STM Topic Model Robustness Checks

B-1 STM Topic Estimation without Covariates

STM is very similar to Latent Dirichlet Allocation (LDA), which is one of the most common forms of topic modeling. However, LDA assumes every document in a corpus is generated in the same way and therefore assumes the frequency of each topic and words likely to be used within each topic are the same across each document (Roberts et al. 2014). STM relaxes these assumptions and is better suited for our analyses since topics and word choices within topics will likely vary across advertisements not all housing units are the same and landlords may choose to address different themes in their ads. STM also allows us to include covariates when the model is estimating which words appears in each topic and how frequently each topic occurs in each document.

In this section, we demonstrate that STM estimation results and the regression on the topic proportions from the STM are robust to the selection of covariates included in the STM estimation process. We run an STM that does not include any covariates. Table A3 reports the STM topics from the model without covariates. When we compare Table A3 and Table 1, the top 10 words for each topics are identical. The only difference between Table A3 and Table 1 are the average topic proportions. However, the biggest difference in topic proportions is only 0.3 percentage point (for the neighborhood amenity topic).

Table A4 presents respective regression models to Table 2. We use the topic proportions computed by STM that does not include covariates and run regression models predicting these topic proportions. The results in Table A4 have minimal difference with results from Table 2. When we compare the coefficients from our neighborhood type variables, the biggest difference is 0.006.

Table A3: STM Topics from Craigslist Rental Listings: No Covariate STM

Label	Words	%
General	apart, home, communiti, center, pool, call, offer, locat, fit, bedroom	24.4
Logistics	rent, month, applic, will, fee, credit, home, pleas, leas, move	10.6
Unit Amenity	kitchen, floor, center, applianc, room, communiti, loung, fit, stainless, area	12.7
Unit Description	bedroom, room, floor, kitchen, new, includ, larg, bath, bathroom, month	25.1
Pet Policy	pet, polici, apart, restrict, offic, now, home, communiti, hous, hour	8.1
Neighborhood Amenity	park, apart, locat, downtown, walk, shop, restaur, citi, minut, just	10.5
Availability	avail, apart, leas, price, today, unit, manag, chang, properti, subject	8.7

Table A4: Regression on Topic Proportions Estimated from STM without Covariates

	Dependent variable:		
	Logistics Topic	Unit Amenities Topic	Neighborhood Amenities Topic
	(1)	(2)	(3)
White Poor	0.459**	−0.194**	0.305**
	(0.075)	(0.073)	(0.084)
Black Non-poor	0.104	−0.097*	−0.009
	(0.107)	(0.046)	(0.061)
Black Poor	0.436**	−0.175**	0.173**
	(0.122)	(0.065)	(0.062)
Latino Non-poor	0.289**	−0.115	0.206**
	(0.082)	(0.073)	(0.058)
Latino Poor	0.562**	−0.227**	0.323**
	(0.148)	(0.073)	(0.078)
Asian Non-poor	0.242*	−0.027	−0.051
	(0.101)	(0.064)	(0.092)
Asian Poor	0.726**	−0.249	0.384**
	(0.249)	(0.194)	(0.079)
Price ($1000)	−0.112**	0.341**	0.016
	(0.027)	(0.037)	(0.018)
% College	−0.009**	0.011**	0.011**
	(0.002)	(0.001)	(0.001)
% Foreign Born	−0.015**	0.004*	−0.008**
	(0.003)	(0.002)	(0.002)
% Units Renter Occ	−0.006**	0.006**	0.006**
	(0.001)	(0.001)	(0.001)
% Unit Built after 2010	−0.013*	0.038**	−0.002
	(0.005)	(0.004)	(0.002)
% Vacancy	2.377**	−0.655**	1.122**
	(0.268)	(0.182)	(0.225)
Observations	1,692,639	1,692,639	1,692,639
R^2	0.154	0.223	0.176
Adjusted R^2	0.154	0.223	0.176
Residual Std. Error	1.615	1.242	1.047

$^{+}$p<0.1; *p<0.05; **p<0.01

Note: Dependent variables are log transformed. The base category for neighborhood type is White non-poor neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 30% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

B-2 STM Topic Estimation with Different Number of Topics

In this section, we demonstrate the topic model results and regression results from STMs that use different numbers of topics. Results from topic models can vary when researchers choose the number of topics. We show that the results from STMs with 5 and 9 topics have similar results to the results from the main text (7 topics). In addition, the results in this section show that STM with 7 topic has more interpretable and cohesive results than 5 or 9 topics.

Table A5 presents the STM results with 5 topics (top 10 frequency words and proportions). The content of the topics is overall similar to the results from STM with 7 topics. However, there are a couple of topics that are combined together into a single topic. These combined topics are created because the number of topics is not large enough so topics that should be independent are compressed to a single topic. When we compare Table A8 to Table 1 in the main text, the first seven topics are identical to the ones from Table 1. The only differences are the last two topics. These last two topics labelled as apartment 1 and apartment 2 represent generic text from various apartment complexes. These two topics are less coherent than the first seven topics. The STM creates topics that are less interpretable as we increase the number of topics above the ideal number of topic (7).

The regression results reported in Table A7 and Table A8 demonstrate similar results to Table 2. The regression results from STM with 9 topics is more likely to be similar to the models with 7 topics than the results from STM with 5 topics. It is because STM with 9 topics share the same 7 topics as STM with 7 topics and because STM with 5 topics have a few merged topics. The results in Table A7 are similar to Table 2, especially for logistics and unit description topic and unit amenities topic. The regression coefficients for general and neighborhood amenities topic show similar direction as Table 2 but the strength of correlation is weaker than Table 2 because the topic contains a general topic which makes the topic less coherent. The results in Table A8 are very similar to Table 2. In fact, the results for unit amenities topic and neighborhood amenities topic have stronger correlation than Table 2. Altogether, these comparisons support our selection of a model with 7 topics as optimal.

Table A5: STM Topics from Craigslist Rental Listings: STM with 5 Topics

Label	Words	%
Availability and General	apart, home, today, call, avail, leas, price, locat, communiti, manag	16.0
Logistics and Unit Description	bedroom, rent, month, room, floor, includ, park, bath, pet, new	35.2
Unit Amenity	kitchen, park, stainless, applianc, room, steel, center, granit, countertop	24.3
Pet Policy	pet, communiti, apart, home, hour, hous, offic, restrict, now, polici	10.0
General and Neighborhood Amenity	apart, center, home, pool, communiti, call, bedroom, closet, fit, park	14.6

Table A6: STM Topics from Craigslist Rental Listings: STM with 9 Topics

Label	Words	%
General	apart, home, communiti, center, pool, call, offer, fit, locat, bedroom	24.4
Logistics	month, rent, applic, fee, home, will, credit, leas, deposit, move	9.6
Unit Amenity	kitchen, floor, applianc, center, room, stainless, featur, steel, communiti, countertop	11.9
Unit Description	bedroom, room, floor, kitchen, new, larg, includ, bath, bathroom, live	22.0
Pet Policy	pet, polici, restrict, apart, bath, dog, offic, breed, now, per	7.3
Neighborhood Amenity	apart, park, locat, walk, downtown, build, restaur, shop, citi, street	9.7
Availability	avail, apart, price, unit, leas, today, chang, properti, subject, special	6.9
Apartments	communiti, park, center, access, hour, pool, fit, amen, apart, creek	4.3
Apartments 2	beach, bedroom, artnt, pool, unit, view, bay, nth, downtown, rent	3.9

Table A7: Regression on Topic Proportions Estimated from STM with 5 Topics

	Dependent variable:		
	Logistics and Unit Description Topic	Unit Amenities Topic	Neighborhood Amenities Topic
	(1)	(2)	(3)
White Poor	0.670**	−0.672**	0.043
	(0.090)	(0.065)	(0.071)
Black Non-poor	0.042	−0.024	−0.117*
	(0.102)	(0.064)	(0.052)
Black Poor	0.466**	−0.544**	0.084
	(0.125)	(0.077)	(0.058)
Latino Non-poor	0.368**	−0.421**	0.140*
	(0.070)	(0.057)	(0.063)
Latino Poor	0.736**	−0.719**	0.101
	(0.138)	(0.082)	(0.068)
Asian Non-poor	0.216+	−0.211**	−0.050
	(0.116)	(0.077)	(0.067)
Asian Poor	0.861**	−0.702**	−0.108
	(0.252)	(0.232)	(0.159)
Price ($1000)	0.029	−0.082**	0.330**
	(0.031)	(0.026)	(0.036)
% College	−0.006**	−0.002	0.020**
	(0.002)	(0.001)	(0.001)
% Foreign Born	−0.023**	0.023**	−0.005*
	(0.004)	(0.003)	(0.002)
% Units Renter Occ	−0.009**	0.001	0.008**
	(0.002)	(0.001)	(0.001)
% Unit Built after 2010	−0.030**	−0.008*	0.039**
	(0.005)	(0.004)	(0.003)
% Vacancy	2.676**	−3.242**	0.432**
	(0.275)	(0.269)	(0.138)
Observations	1,692,639	1,692,639	1,692,639
R^2	0.284	0.310	0.404
Adjusted R^2	0.284	0.310	0.404
Residual Std. Error	1.456	1.149	1.039

$^{+}$p<0.1; *p<0.05; **p<0.01

Note: Dependent variables are log transformed. The base category for neighborhood type is White non-poor neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 30% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

Table A8: Regression on Topic Proportions Estimated from STM with 9 Topics

	Dependent variable:		
	Logistics Topic	Unit Amenities Topic	Neighborhood Amenities Topic
	(1)	(2)	(3)
White Poor	0.519^{**}	-0.285^{**}	0.441^{**}
	(0.081)	(0.076)	(0.074)
Black Non-poor	0.159	-0.172^{**}	-0.003
	(0.107)	(0.052)	(0.065)
Black Poor	0.431^{**}	-0.239^{**}	0.362^{**}
	(0.116)	(0.064)	(0.070)
Latino Non-poor	0.298^{**}	-0.111	0.398^{**}
	(0.079)	(0.070)	(0.058)
Latino Poor	0.637^{**}	-0.311^{**}	0.633^{**}
	(0.134)	(0.080)	(0.081)
Asian Non-poor	0.166	-0.066	-0.092
	(0.107)	(0.065)	(0.096)
Asian Poor	0.702^{**}	-0.447^{+}	0.608^{**}
	(0.263)	(0.234)	(0.075)
Price ($1000)	-0.143^{**}	0.326^{**}	-0.035^{*}
	(0.024)	(0.034)	(0.016)
% College	-0.014^{**}	0.016^{**}	0.019^{**}
	(0.002)	(0.001)	(0.001)
% Foreign Born	-0.016^{**}	0.005^{*}	-0.011^{**}
	(0.003)	(0.002)	(0.002)
% Units Renter Occ	-0.008^{**}	0.007^{**}	0.012^{**}
	(0.001)	(0.001)	(0.001)
% Unit Built after 2010	-0.014^{**}	0.046^{**}	-0.011^{**}
	(0.005)	(0.004)	(0.003)
% Vacancy	2.150^{**}	-1.001^{**}	1.910^{**}
	(0.286)	(0.186)	(0.204)
Observations	1,692,639	1,692,639	1,692,639
R^2	0.193	0.325	0.447
Adjusted R^2	0.193	0.325	0.447
Residual Std. Error	1.535	1.251	1.001

$^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$

Note: Dependent variables are log transformed. The base category for neighborhood type is White non-poor neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 30% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

B-3 Unit Description Topic

Table A9: Regression Model Predicting Unit Description Topic Proportion

	Dependent variable:
	Unit Description Topic
White Poor	0.520^{**}
	(0.084)
Black Non-poor	-0.007
	(0.089)
Black Poor	0.354^{**}
	(0.110)
Latino Non-poor	0.225^{**}
	(0.071)
Latino Poor	0.564^{**}
	(0.118)
Asian Non-poor	0.162
	(0.118)
Asian Poor	0.666^{**}
	(0.200)
Price ($1000)	0.067^{*}
	(0.032)
% College	-0.005^{**}
	(0.002)
% Foreign Born	-0.017^{**}
	(0.003)
% Units Renter Occ	-0.007^{**}
	(0.001)
% Unit Built after 2010	-0.021^{**}
	(0.004)
% Vacancy	2.066^{**}
	(0.270)
Observations	1,692,639
R^2	0.214
Adjusted R^2	0.214
Residual Std. Error	1.403

$^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$

Note: Dependent variables are log transformed. The base category for neighborhood type is White non-poor neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 30% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

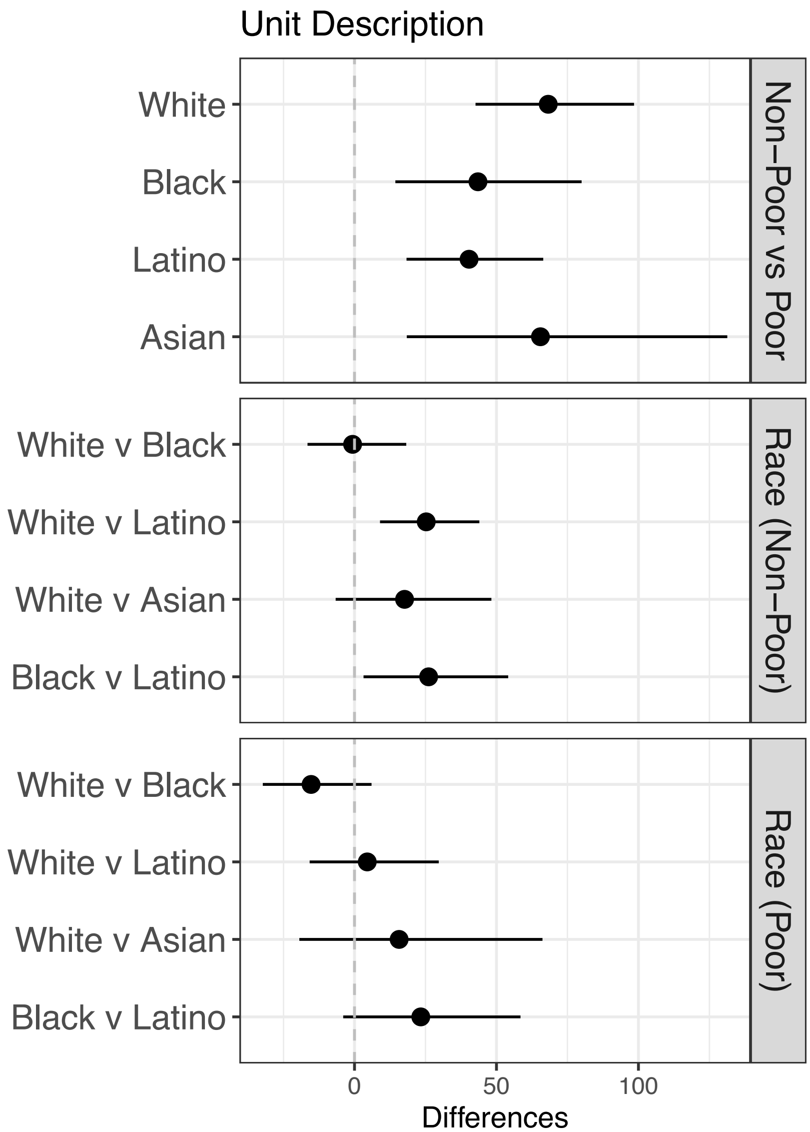

Figure A1: Pairwise Comparison of Unit Description Topic Proportion Across Neighborhood Types.
Note: The dependent variable is log transformed. The dots and bars indicate the percent change and confidence intervals from derived from regression models. For the ease of interpretation, we present percent change instead of regression coefficients. The neighborhood written first (before "vs") is the base category. The first four rows (non-poor vs poor) display the regression coefficients for poor neighborhoods when the non-poor neighborhoods for the respective racial group is the base category. Negative value means poor neighborhoods have less information than non-poor neighborhoods. Positive value means non-poor neighborhoods have less information than poor neighborhoods. The following eight rows compare differences between different racial compositions. The first racial group is the base category. For example, the coefficient for "White v Latino" in the sixth row for the unit amenities topic indicates that Latino non-poor neighborhoods have 25.3 percent more topic proportions in unit description than White non-poor neighborhoods. Neighborhood racial composition and poverty rate are obtained from 2016 ACS 5-year pooled data. The regression models include MSA fixed effects. Standard errors are clustered at MSA level. Plots are based on results presented in Table A9.

C. Modeling Topic Prevalence Robustness Checks

C-1 Log-transformation of the Dependent Variables

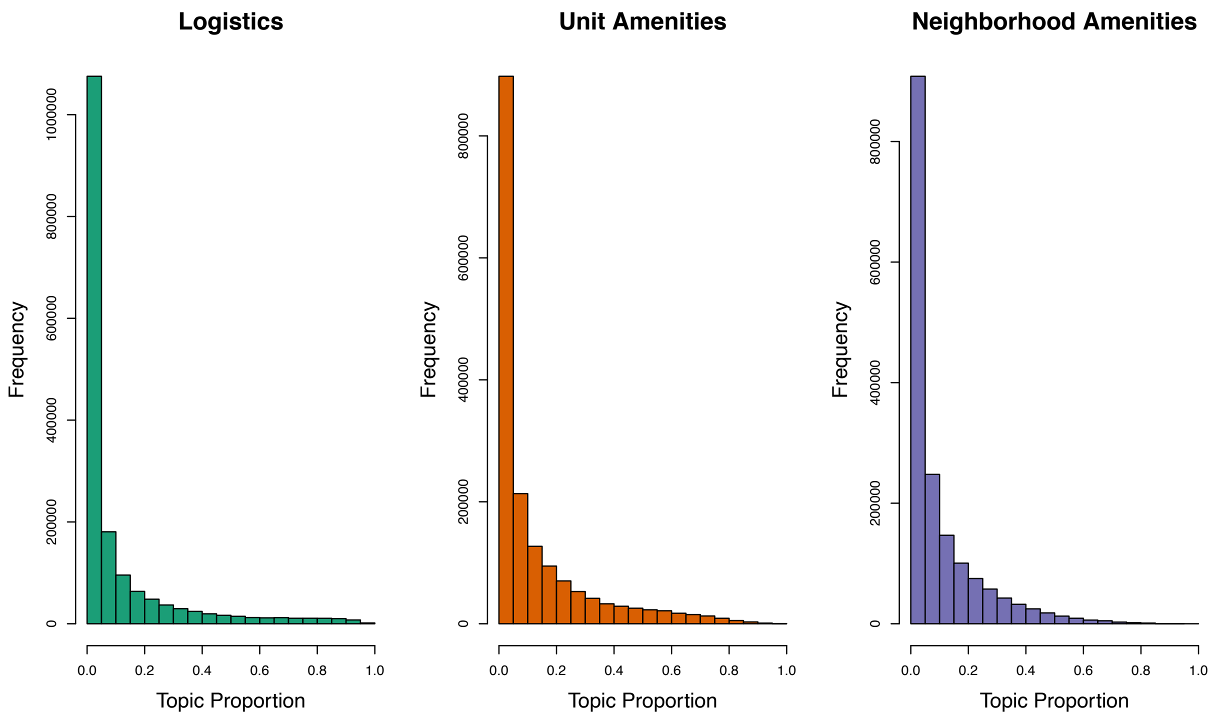

Figure A2: Histogram for Topic Proportions

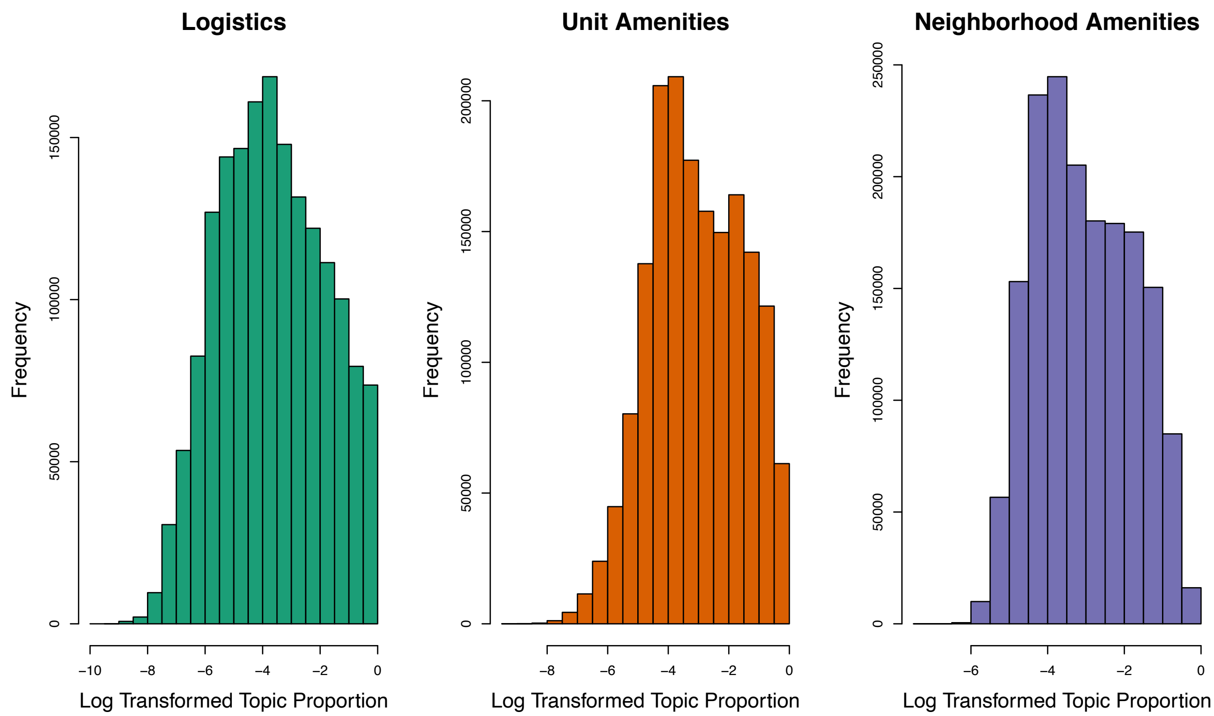

Figure A3: Histogram for Log-Transformed Topic Proportions

Table A10: Topic Proportions without Log Transformation

	Dependent variable:					
	Logistics Topic		Unit Amenities Topic		Neighborhood Amenities Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
White Poor	0.021**	0.028**	−0.013+	−0.016*	0.057**	0.036**
	(0.007)	(0.006)	(0.008)	(0.007)	(0.012)	(0.011)
Black Non-poor	0.038**	0.014+	−0.043**	−0.008	−0.024**	0.001
	(0.008)	(0.008)	(0.006)	(0.006)	(0.005)	(0.005)
Black Poor	0.077**	0.041**	−0.051**	−0.011	0.005	0.016**
	(0.010)	(0.010)	(0.007)	(0.007)	(0.005)	(0.006)
Latino Non-poor	0.037**	0.023*	−0.043**	−0.003	−0.014*	0.029**
	(0.009)	(0.010)	(0.009)	(0.008)	(0.006)	(0.006)
Latino Poor	0.063**	0.046**	−0.059**	−0.007	−0.004	0.033**
	(0.012)	(0.012)	(0.008)	(0.010)	(0.007)	(0.008)
Asian Non-poor	−0.008	0.009	0.016*	0.006	−0.024**	−0.008
	(0.006)	(0.006)	(0.007)	(0.009)	(0.008)	(0.009)
Asian Poor	0.022	0.041*	−0.027	−0.027	0.047*	0.043*
	(0.022)	(0.017)	(0.026)	(0.018)	(0.019)	(0.017)
Price ($1000)		−0.015**		0.043**		−0.002
		(0.003)		(0.005)		(0.003)
% College		−0.001**		0.001**		0.002**
		(0.0001)		(0.0002)		(0.0001)
% Foreign Born		−0.0005*		0.0001		−0.001**
		(0.0002)		(0.0003)		(0.0002)
% Units Renter Occ		−0.001**		0.0004**		0.001**
		(0.0001)		(0.0001)		(0.0001)
% Unit Built after 2010		−0.0002		0.007**		−0.0002
		(0.0003)		(0.001)		(0.0003)
% Vacancy		0.190**		0.006		0.102**
		(0.024)		(0.025)		(0.023)
Observations	1,692,639	1,692,639	1,692,639	1,692,639	1,692,639	1,692,639
R^2	0.119	0.149	0.133	0.233	0.152	0.230
Adjusted R^2	0.119	0.149	0.133	0.233	0.151	0.230
Residual Std. Error	0.174	0.171	0.165	0.155	0.122	0.116

$^{+}$p<0.1; *p<0.05; **p<0.01

Note: Dependent variables are topic proportions estimated by STM. The base category for neighborhood type is White non-poor neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 30% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

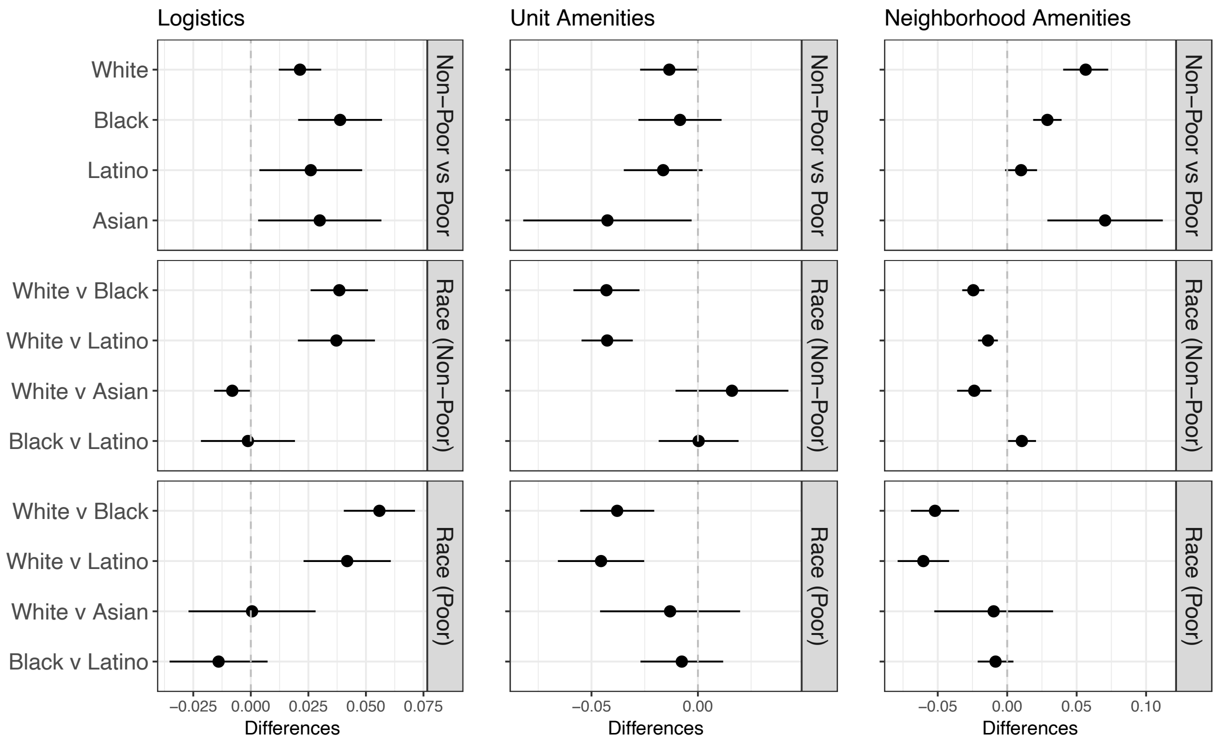

Figure A4: Dependent Variables without Log Transformation: Pairwise Comparison of Topic Proportions Across Neighborhood Types

Note: The dependent variables are topic proportions estimated by STM. The dots and bars indicate the percent change and confidence intervals from derived from regression models. For the ease of interpretation, we present percent change instead of regression coefficients. The neighborhood written first (before "vs") is the base category. The first four rows (non-poor vs poor) display the regression coefficients for poor neighborhoods when the non-poor neighborhoods for the respective racial group is the base category. Negative value means poor neighborhoods have less information than non-poor neighborhoods. Positive value means non-poor neighborhoods have less information than poor neighborhoods. The following eight rows compare differences between different racial compositions. The first racial group is the base category. For example, the coefficient for "White v Latino" in the sixth row for the unit amenities topic indicates that Latino non-poor neighborhoods have 0.043 (4.3 percentage point) less topic proportions in unit amenities than White non-poor neighborhoods. Neighborhood racial composition and poverty rate are obtained from 2016 ACS 5-year pooled data. The regression models include MSA fixed effects. Standard errors are clustered at MSA level. Based on Table A10.

C-2 Defining Neighborhoods

In this section, we test whether our regression results are robust to different neighborhood classifications.

C-2-a Majority v. Plurality

We change our classification for neighborhood racial composition. Instead of using plurality to determine the dominant group in each neighborhood, we use majority to classify the dominant group. Census tracts that do not have any majority group are classified as diverse neighborhoods. We use the same poverty rate threshold (16.6%) as Table A14. The overall results reported in Table A11 are similar to Table A14. Black and Latino non-poor neighborhoods show weaker correlations than Table 3. However, poor Black and Latino neighborhoods have stronger correlations than those reported in Table 3 (except for Latino poor neighborhoods for unit amenities topic). Regression coefficients for the neighborhood amenities topic are similar across the three different classification schemes. The new neighborhood types in this regression, diverse neighborhoods, demonstrate a mixed pattern. Diverse poor neighborhoods show the same direction but somewhat smaller magnitudes in terms of regression coefficients compared to Black and Latino poor neighborhoods. They have more logistics and neighborhood amenities topic, but less unit amenities topic than White non-poor neighborhoods. Diverse non-poor neighborhoods have less logistics topic compared White non-poor neighborhoods, although the results for the logistics topic is marginally significant at the 90% confidence level.

Again, overall these comparisons demonstrate robust support for our key conclusion: both neighborhood race and SES structure the types of information available to prospective renters. However, some specific findings are sensitive to our neighborhood racial and poverty composition cut-offs we use.

Table A11: Topic Proportion Regressions – Majority Racial Group and Average Poverty Rate (16.6%)

	Dependent variable:		
	Logistics Topic	Unit Amenities Topic	Neighborhood Amenities Topic
	(1)	(2)	(3)
White Poor	0.423**	−0.203**	0.194**
	(0.046)	(0.044)	(0.047)
Black Non-poor	0.017	−0.0001	−0.097
	(0.214)	(0.177)	(0.079)
Black Poor	0.654**	−0.355**	0.203**
	(0.109)	(0.059)	(0.051)
Latino Non-poor	0.196	−0.102	0.362**
	(0.127)	(0.080)	(0.084)
Latino Poor	0.594**	−0.358**	0.462**
	(0.134)	(0.082)	(0.082)
Asian Non-poor	0.230	−0.346**	−0.101
	(0.157)	(0.098)	(0.120)
Asian Poor	0.553*	0.113	0.495**
	(0.264)	(0.161)	(0.137)
Diverse Non-poor	−0.146+	−0.070	−0.057
	(0.083)	(0.064)	(0.062)
Diverse Poor	0.376**	−0.218**	0.191**
	(0.083)	(0.053)	(0.060)
Price ($1000)	−0.116**	0.334**	−0.006
	(0.025)	(0.037)	(0.017)
% College	−0.012**	0.016**	0.018**
	(0.002)	(0.001)	(0.001)
% Foreign Born	−0.015**	0.005*	−0.010**
	(0.004)	(0.002)	(0.002)
% Units Renter Occ	−0.011**	0.007**	0.010**
	(0.001)	(0.001)	(0.001)
% Unit Built after 2010	−0.011*	0.047**	−0.003
	(0.005)	(0.004)	(0.002)
% Vacancy	2.459**	−0.816**	1.596**
	(0.239)	(0.181)	(0.205)
Observations	1,692,639	1,692,639	1,692,639
R^2	0.234	0.351	0.344
Adjusted R^2	0.234	0.351	0.344
Residual Std. Error	1.606	1.233	1.026

$^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$

Note: Dependent variables are log transformed. The base category for neighborhood type is White non-poor neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the majority racial group. We use a threshold of 16.6% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

C-2-b Diversity

Next, we test an alternative conceptualization of neighborhood diversity developed by Ellis, Holloway and Wright (2015). We use 2010 data from their site, `mixedmetro.us`, to test alternative classifications of neighborhood racial/ethnic composition and diveristy. First, in Table A12 we categorize neighborhoods based on racial plurality and by diversity (with White and low-diversity neighborhoods as the reference categories), following their definition of diversity. As shown in Table A12, diversity is associated with variation in our key topics. Furthermore, Table A13 demonstrates that when we interact racial pluarity with diversity, we can find differences between low and moderately diverse same-race neighborhoods.

However, it is difficult to tell how these results compare to our main models which interact neighborhood race with poverty status. The key question is whether our findings for neighborhood race by poverty status are distinct for diverse neighborhoods. Thus, we next estimate a three-way interaction between neighborhood racial plurality, diversity level, and poverty status. To facilitate interpretation we plot the predicted change between low and moderate diversity neighborhoods by race and poverty status in Figure A5. Our results remain consistent with previous analyses. While neighborhoods with moderate v. low diversity are not identical within racial categories, the results are largely similar; Black and Latino neighborhoods have more language about rental logistics and less about unit amenities in contrast to White neighborhoods; poor non-White neighborhoods remain particularly disadvantaged.

Table A12: Plurality and Diversity of Neighborhoods

	Dependent variable:		
	Logistics Topic	Unit Amenities Topic	Neighborhood Amenities Topic
	(1)	(2)	(3)
Black	0.174	-0.176^{**}	-0.036
	(0.113)	(0.053)	(0.042)
Latino	0.290^{**}	-0.128	0.277^{**}
	(0.100)	(0.084)	(0.049)
Asian	0.302^{*}	-0.030	0.021
	(0.145)	(0.079)	(0.071)
Moderate Diversity	-0.111	0.032	-0.100^{**}
	(0.070)	(0.035)	(0.029)
High Diversity	-0.264^{*}	0.193^{*}	-0.054
	(0.113)	(0.082)	(0.064)
Price ($1000)	-0.146^{**}	0.336^{**}	-0.015
	(0.024)	(0.039)	(0.012)
% College	-0.013^{**}	0.016^{**}	0.017^{**}
	(0.002)	(0.001)	(0.001)
% Foreign Born	-0.015^{**}	0.004^{*}	-0.009^{**}
	(0.004)	(0.002)	(0.002)
% Units Renter Occ	-0.004^{**}	0.005^{**}	0.013^{**}
	(0.001)	(0.001)	(0.001)
% Unit Built after 2010	-0.014^{*}	0.052^{**}	0.002
	(0.007)	(0.005)	(0.003)
% Vacancy	2.843^{**}	-1.103^{**}	1.770^{**}
	(0.385)	(0.255)	(0.192)
Observations	1,168,994	1,168,994	1,168,994
R^2	0.216	0.359	0.332
Adjusted R^2	0.216	0.359	0.331
Residual Std. Error	1.581	1.229	1.028

$^{+}$p<0.1; *p<0.05; **p<0.01

Note: Dependent variables are log transformed.The base category for neighborhood type is White neighborhoods for racial composition and low diversity neighborhoods for diversity. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. Diversity classification follows Ellis, Holloway, and Wright (2012). Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

Table A13: Race × Diversity

	Dependent variable:					
	Logistics Topic		Unit Amenities Topic		Neighborhood Amenities Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
White Moderate Diversity	-0.149^{*}	-0.118	-0.044	-0.0004	-0.001	-0.091^{*}
	(0.071)	(0.078)	(0.049)	(0.045)	(0.058)	(0.041)
Black Low Diversity	0.969^{**}	0.198	-1.063^{**}	-0.253^{**}	-0.115	0.038
	(0.145)	(0.156)	(0.066)	(0.084)	(0.081)	(0.069)
Black Moderate Diversity	0.491^{**}	0.188^{+}	-0.453^{**}	-0.060	-0.044	-0.0004
	(0.109)	(0.109)	(0.086)	(0.079)	(0.072)	(0.053)
Latino Low Diversity	0.363	0.354	-0.730^{**}	-0.113	-0.246^{+}	0.444^{**}
	(0.239)	(0.229)	(0.224)	(0.163)	(0.136)	(0.137)
Latino Moderate Diversity	0.298^{**}	0.171	-0.536^{**}	-0.057	-0.069	0.246^{**}
	(0.099)	(0.129)	(0.069)	(0.086)	(0.084)	(0.075)
Asian Low Diversity	0.201	0.554^{**}	-0.549^{**}	-0.412^{**}	-0.200^{**}	0.321^{**}
	(0.125)	(0.176)	(0.080)	(0.087)	(0.062)	(0.107)
Asian Moderate Diversity	-0.241^{*}	0.139	0.035	-0.065	-0.063	-0.011
	(0.099)	(0.169)	(0.075)	(0.093)	(0.099)	(0.109)
High Diversity	-0.064	-0.122	-0.235^{**}	0.117	-0.070	0.077
	(0.111)	(0.135)	(0.084)	(0.094)	(0.085)	(0.077)
Price ($1000)		-0.149^{**}		0.335^{**}		-0.017
		(0.025)		(0.040)		(0.018)
% College		-0.013^{**}		0.017^{**}		0.018^{**}
		(0.002)		(0.001)		(0.001)
% Foreign Born		-0.014^{**}		0.004^{+}		-0.009^{**}
		(0.004)		(0.002)		(0.003)
% Units Renter Occ		-0.004^{**}		0.005^{**}		0.013^{**}
		(0.001)		(0.001)		(0.001)
% Unit Built after 2010		-0.014^{*}		0.052^{**}		0.001
		(0.006)		(0.005)		(0.003)
% Vacancy		2.747^{**}		-1.021^{**}		1.687^{**}
		(0.361)		(0.264)		(0.251)
Observations	1,168,994	1,168,994	1,168,994	1,168,994	1,168,994	1,168,994
R^2	0.179	0.217	0.246	0.358	0.203	0.332
Adjusted R^2	0.179	0.217	0.246	0.358	0.203	0.332
Residual Std. Error	1.618	1.581	1.333	1.229	1.122	1.028

$^{+}$p<0.1; *p<0.05; **p<0.01

Note: Dependent variables are log transformed. The base category for neighborhood type is White low diversity neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. Diversity classification follows Ellis, Holloway, and Wright (2012). Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

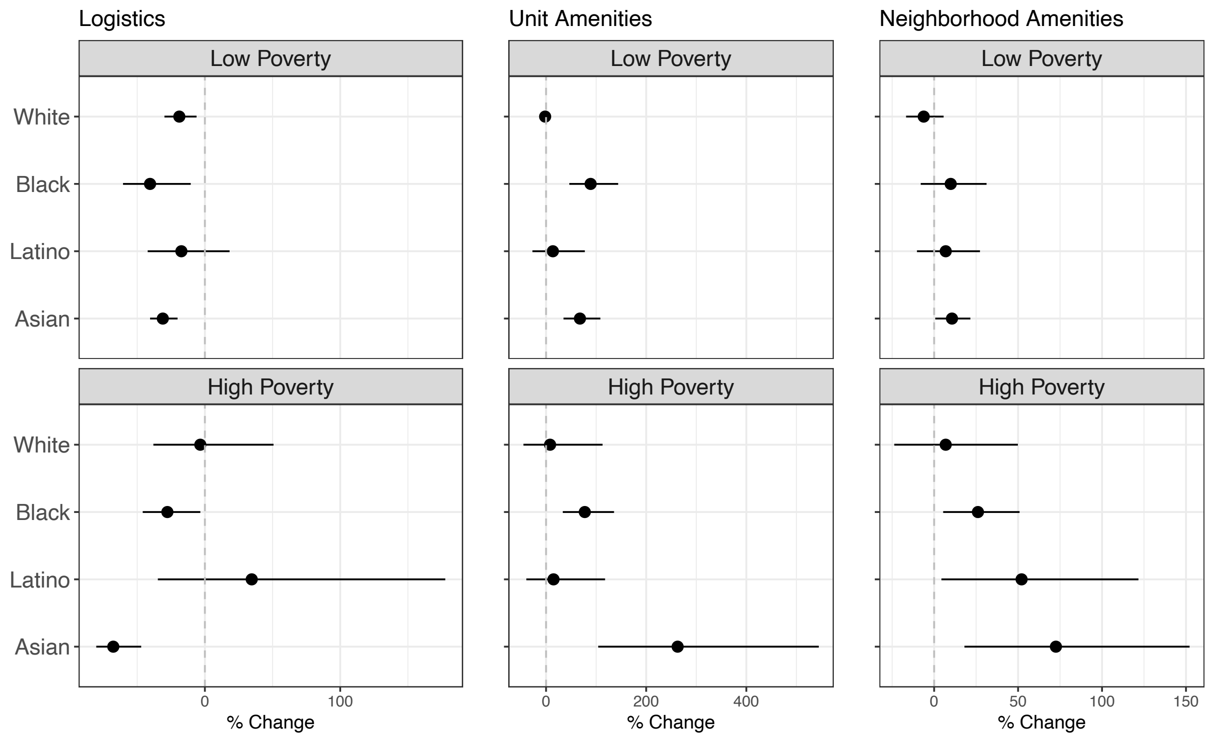

Figure A5: The Change in Topic Proportions When Low Diverse Neighborhoods Become Moderate Diverse Neighborhoods

Note: Dependent variables are log transformed. The dots and bars indicate the percent change and confidence intervals derived from regression models. For the ease of interpretation, we present percent change instead of regression coefficients. The first four rows (low poverty) display the percent change in topic proportion for low poverty neighborhoods when low diverse neighborhoods become moderate diverse neighborhoods. Negative value means moderate diversity neighborhoods have less information than low diversity neighborhoods. Positive value means moderate diversity neighborhoods have more information than low diversity neighborhoods. The following four rows compare differences between moderate and low diversity neighborhoods across high poverty neighborhoods. For example, the coefficient for "Black" in the sixth row for the unit amenities topic indicates that Black and high poverty neighborhoods have 77.36 percent more topic proportions in unit amenities when the low diversity neighborhoods become moderate diversity neighborhoods. Neighborhood racial composition and poverty rate are obtained from 2016 ACS 5-year pooled data. The regression models include MSA fixed effects. Standard errors are clustered at MSA level.

C-2-c Poverty Threshold

We test whether changing the poverty rate threshold (set at 30% in the main text) changes the results from the topic proportion regressions in Table A14. We set the average poverty rate, 16.6% as a threshold for classifying poor and non-poor neighborhoods. The direction and the magnitude of the coefficients are substantively similar to Table 3 except for Black non-poor and Latino non-poor neighborhoods.The differences between Table A14 and Table 3 are only pronounced for the logistic topic and unit amenities topic. The regression results for the neighborhood amenities topic in Table A14 is very similar to those from Table 3. The strength of correlations for Black poor neighborhood is stronger than the results reported in Table 3. This comparison demonstrates that our overall findings are robust to the poverty rate threshold we use to classify neighborhoods but that some results are sensitive to the poverty rate threshold.

Table A14: Topic Proportion Regressions – Poverty Rate Threshold: Average Poverty Rate (16.6%)

	Dependent variable:		
	Logistics Topic	Unit Amenities Topic	Neighborhood Amenities Topic
	(1)	(2)	(3)
White Poor	0.437**	−0.190**	0.201**
	(0.049)	(0.044)	(0.045)
Black Non-poor	−0.005	−0.072	−0.104
	(0.200)	(0.086)	(0.107)
Black Poor	0.620**	−0.323**	0.201**
	(0.098)	(0.050)	(0.053)
Latino Non-poor	0.350**	−0.039	0.355**
	(0.127)	(0.094)	(0.063)
Latino Poor	0.697**	−0.286**	0.468**
	(0.093)	(0.072)	(0.062)
Asian Non-poor	0.254+	−0.161+	−0.066
	(0.130)	(0.082)	(0.112)
Asian Poor	0.704**	−0.162	0.264*
	(0.221)	(0.163)	(0.106)
Price ($1000)	−0.114**	0.336**	−0.005
	(0.025)	(0.037)	(0.017)
% College	−0.012**	0.016**	0.018**
	(0.002)	(0.001)	(0.001)
% Foreign Born	−0.018**	0.005*	−0.011**
	(0.004)	(0.002)	(0.002)
% Units Renter Occ	−0.011**	0.007**	0.010**
	(0.001)	(0.001)	(0.001)
% Unit Built after 2010	−0.012*	0.047**	−0.003
	(0.005)	(0.004)	(0.002)
% Vacancy	2.536**	−0.837**	1.617**
	(0.244)	(0.178)	(0.208)
Observations	1,692,639	1,692,639	1,692,639
R^2	0.234	0.351	0.345
Adjusted R^2	0.234	0.350	0.345
Residual Std. Error	1.607	1.234	1.026

$^{+}$p<0.1; *p<0.05; **p<0.01

Note: Dependent variables are log transformed. The base category for neighborhood type is White non-poor neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 16.6% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

C-2-d Topic Regressions with Continuous Race and Poverty Variables

In this section, we report the results from regressions that include continuous race and poverty variables instead of our neighborhood type variable. Racial composition measures include % Black, % Latino, and % Asian. We do not include % White because of multicollinearity concerns. We add the poverty rate for each census tract to measure the extent of poverty existing in the neighborhood.

We present two models for each outcome variable. First, we report the results from a model that includes the race variables and other covariates, but does not include the poverty rate variable. We show the results from this model because the % Black and % Latino variables are highly correlated with poverty rate. Next, we include every covariate including the poverty rate. The first models show the proportion of logistics topic increases as % Black and % Latino increase. These correlations become weaker when we include the poverty rate variable. The regression coefficient for the poverty rate show a strong correlation between logistic topic proportion and poverty rate. The second models demonstrate the proportion of unit amenities topic decreases as % Black and % Latino increase. Similar to the logistics topic models, including the poverty rate measure weakens the correlations between unit amenities topic proportion and racial composition variables. Contrary to the first and second models, % Black and % Latino show different results in the third model. Specifically, the neighborhood amenities topic proportion increases as % Latino increases and % Asian decreases. The relationship between the proportion of neighborhood amenities topic and % Black is less consistent. There is no statistically significant relationship in the model without the poverty rate variable but the correlation becomes negative when we include the poverty rate variable.

The overall results from Table A15 are similar to Table 2, demonstrating the robustness of our key conclusions. Neighborhoods with higher % Black or % Latino have more logistics topic and less unit amenities topic. There is more neighborhood amenities topic in neighborhoods with a higher % Latino. Higher poverty rate is positively correlated with logistics topic proportion and neighborhood amenities topic proportion, but negatively correlated with unit amenities topic proportion. These results are largely consistent with the results of Table 2.

Table A15: Topic Regressions with Continuous Race and Poverty Variables

	Dependent variable:					
	Logistics Topic		Unit Amenities Topic		Neighborhood Amenities Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
% Black	0.003*	−0.0005	−0.003**	−0.001	0.0001	−0.002+
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
% Latino	0.006**	0.001	−0.004*	−0.002	0.007**	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
% Asian	−0.008*	−0.012**	−0.002	−0.0005	−0.007*	−0.009**
	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
Poverty Rate		0.028**		−0.010**		0.015**
		(0.003)		(0.002)		(0.002)
Price ($1000)	−0.114**	−0.122**	0.334**	0.337**	−0.006	−0.010
	(0.025)	(0.025)	(0.037)	(0.037)	(0.016)	(0.017)
% College	−0.013**	−0.009**	0.017**	0.015**	0.018**	0.020**
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
% Foreign Born	−0.012**	−0.009*	0.006**	0.004*	−0.008**	−0.006*
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
% Units Renter Occ	−0.008**	−0.014**	0.006**	0.009**	0.011**	0.008**
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
% Unit Built after 2010	−0.012*	−0.010+	0.047**	0.047**	−0.003	−0.002
	(0.005)	(0.005)	(0.004)	(0.004)	(0.002)	(0.002)
% Vacancy	0.030**	0.022**	−0.010**	−0.007**	0.019**	0.014**
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	1,692,643	1,692,643	1,692,643	1,692,643	1,692,643	1,692,643
R^2	0.226	0.240	0.348	0.351	0.342	0.350
Adjusted R^2	0.226	0.240	0.348	0.351	0.342	0.350
Residual Std. Error	1.616	1.601	1.236	1.233	1.028	1.021

+p<0.1; *p<0.05; **p<0.01

Note: Dependent variables are topic proportions estimated by STM. Listings that are more expensive than $10,000 are removed. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

C-2-e High v Low Rent Units

Next we test whether higher and lower priced (asking rent) units display distinct information patterns, even after accounting for neighborhood characteristics. To do so we categorize listings with an asking rent that is higher than the metro median as high rent, all other listings are classified as low rent. Next, we estimate a model interacting neighborhood race with rental price (see Table A16). Interestingly, we find that classifying listings by neighborhood racial composition and high v. low rent produces similar patterns for the logistics and unit amenities topics compared to our neighborhood race by poverty typology. However, a distinct pattern emerges for neighborhood amenities: while we find that neighborhood amenities language is more prevalent in poorer neighborhoods, we find that it is less prevalent among lower-rent listings.

To further explore this finding, we next created a 3-way interaction across neighborhood racial composition, neighborhood poverty status, and listing asking rent (creating 16 distinct neighborhood categories). These results are reported in Figure A6. We find that neighborhood racial composition, poverty status, and rent all contribute to the prevalence of logistics language in advertisements. Rent appears to be strongly tied to the level of logistics language. However, even after accounting for both neighborhood poverty and rent, racial differences remain: among low-poverty neighborhoods, listings in Black and Latino neighborhoods have more logistics language. This is particularly true for listings with lower rents. Additionally, listings in higher poverty neighborhoods tend to have more logistics language regardless of race or rent. Yet the differences between low and high poverty neighborhoods are greatest for Black and Latino neighborhoods. Altogether these patterns highlight how logistics language remains highly racialized.

Somewhat similarly, unit amenities language is more prevalent in higher-rent listings overall, and appears to have a weaker relationship with neighborhood poverty status compared to logistics language. However, once again racial gaps remain even after accounting for rent, such that listings in both poor and non-poor predominantly Black and Latino neighborhoods have less discussion of unit amenities compared to those in White neighborhoods with similar poverty rates.

Finally, in general, advertisements in higher poverty neighborhoods tend to include more discussion of neighborhood amenities compared to their same-race, lower poverty counterparts. Among high poverty neighborhoods, we see gaps both by race (listings in Black and Latino neighborhoods have less neighborhood amenities language compared to White neighborhoods) and by asking rent, with higher-rent units containing more neighborhood amenities language. Among low-poverty neighborhoods, there are only small (and sometimes non-significant) differences in the prevalence of neighborhood amenities language among higher and lower-rent units; racial differences among low-poverty units are also small, though Black neighborhoods have slightly lower levels than all others. We focus on this last set of findings in the main text of the paper.

Table A16: Race × High and Low Rent Units

	Dependent variable:					
	Logistics Topic		Unit Amenities Topic		Neighborhood Amenities Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
White Low Rent	0.435**	0.279**	−0.736**	−0.500**	−0.133**	0.022
	(0.057)	(0.050)	(0.035)	(0.032)	(0.026)	(0.023)
Black High Rent	0.410**	0.104	−0.377**	−0.041	−0.133	0.030
	(0.126)	(0.122)	(0.090)	(0.080)	(0.082)	(0.072)
Black Low Rent	1.039**	0.534**	−1.261**	−0.685**	−0.265**	0.015
	(0.108)	(0.103)	(0.051)	(0.052)	(0.046)	(0.052)
Latino High Rent	0.433**	0.428**	−0.402**	−0.082	−0.022	0.417**
	(0.071)	(0.091)	(0.079)	(0.071)	(0.071)	(0.063)
Latino Low Rent	0.732**	0.624**	−1.151**	−0.642**	−0.222**	0.303**
	(0.100)	(0.104)	(0.067)	(0.074)	(0.070)	(0.054)
Asian High Rent	−0.136	0.314*	0.021	−0.132	−0.215**	−0.035
	(0.104)	(0.125)	(0.069)	(0.085)	(0.072)	(0.081)
Asian Low Rent	0.142	0.523**	−0.595**	−0.608**	−0.238**	0.036
	(0.167)	(0.148)	(0.147)	(0.120)	(0.086)	(0.099)
% College		−0.013**		0.017**		0.017**
		(0.002)		(0.001)		(0.001)
% Foreign Born		−0.017**		0.004*		−0.011**
		(0.004)		(0.002)		(0.002)
% Units Renter Occ		−0.007**		0.006**		0.012**
		(0.001)		(0.001)		(0.001)
% Unit Built after 2010		−0.011*		0.045**		−0.004+
		(0.005)		(0.004)		(0.002)
% Vacancy		2.876**		−0.840**		1.818**
		(0.280)		(0.193)		(0.210)
Observations	1,692,639	1,692,639	1,692,639	1,692,639	1,692,639	1,692,639
R^2	0.188	0.230	0.287	0.358	0.228	0.341
Adjusted R^2	0.188	0.230	0.287	0.358	0.228	0.341
Residual Std. Error	1.655	1.611	1.292	1.227	1.113	1.029

$^{+}$p<0.1; *p<0.05; **p<0.01

Note: Topic proportions for the three topics are computed by STM (Roberts et al. 2014). The base category for neighborhood type is White high rent neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. Rental price higher than metro median is classified as high rent units. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

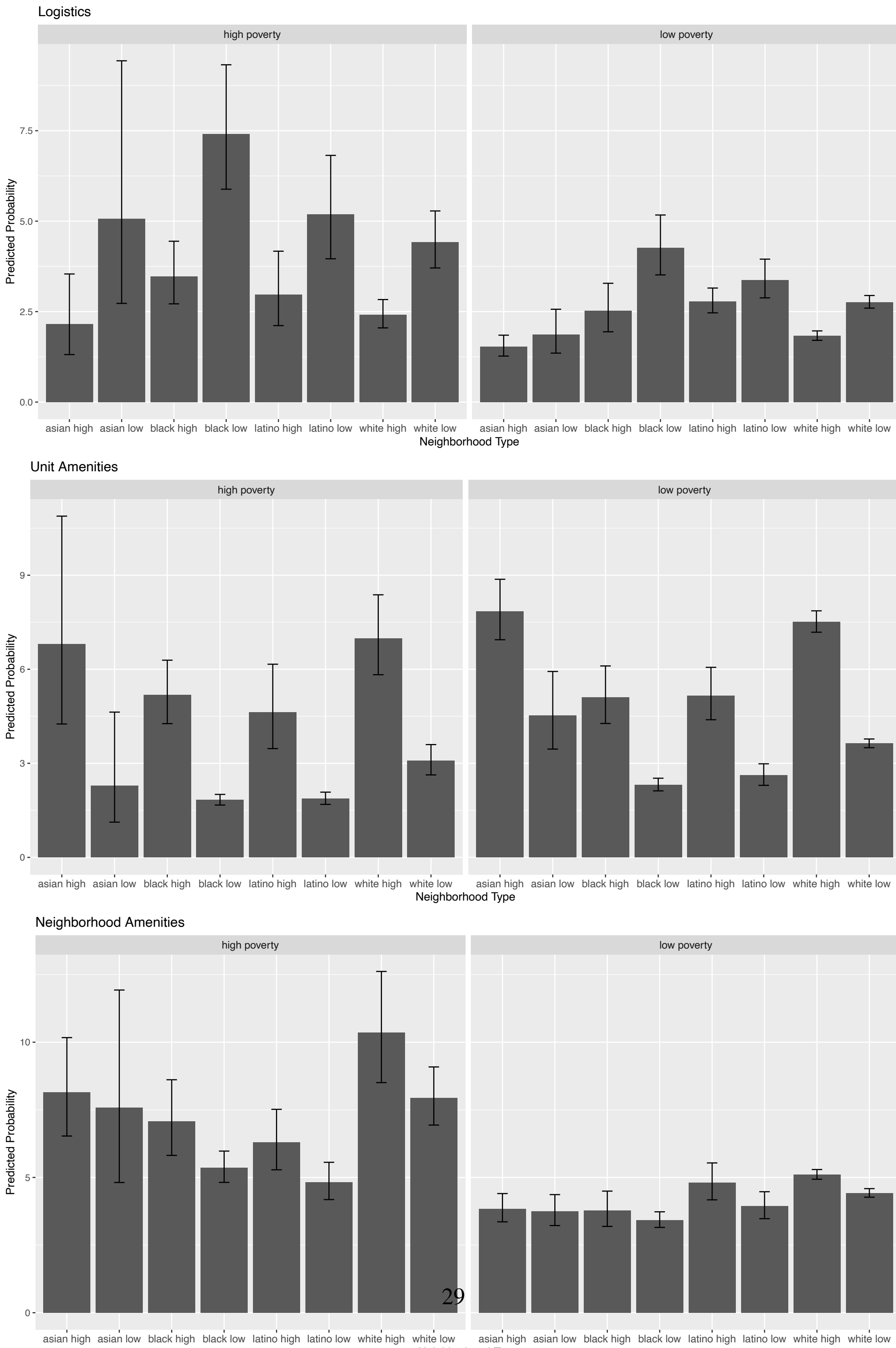

Figure A6: Predicted Probabilities of Topic Proportions across Neighborhood Types (Race × Poverty × High/Low Rents)

C-2-f Census Tracts v. Zip Codes

In this section we test if our results hold at the zip code level rather than using tracts. Zip codes are much larger than tracts and are not ideal for representing neighborhoods; however, because of how Craigslist collects geocoded information from posters, zip codes are less likely to be sensitive to any potential user errors. Because the distribution of poverty across zip codes is distinct from that of tracts, we use a poverty threshold of 15% in these models. While there are some differences when we use zip codes, our key results remain: listing information is highly racialized and also corresponds to zip code poverty rates.

Table A17: Zip Code

	Dependent variable:					
	Logistics Topic		Unit Amenities Topic		Neighborhood Amenities Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
White Poor	0.501^{**}	0.440^{**}	-0.223^{**}	-0.156^{**}	0.413^{**}	0.219^{**}
	(0.069)	(0.074)	(0.056)	(0.043)	(0.042)	(0.041)
Black Non-poor	0.245	0.005	-0.483^{**}	-0.144^{+}	-0.397^{**}	-0.184^{*}
	(0.280)	(0.232)	(0.140)	(0.083)	(0.066)	(0.081)
Black Poor	0.834^{**}	0.363^{**}	-0.667^{**}	-0.170^{*}	0.156^{*}	0.131^{*}
	(0.122)	(0.115)	(0.078)	(0.069)	(0.062)	(0.060)
Latino Non-poor	0.172	0.170	-0.317^{*}	0.013	−0.159	0.295^{**}
	(0.105)	(0.132)	(0.128)	(0.121)	(0.125)	(0.093)
Latino Poor	0.624^{**}	0.547^{**}	-0.623^{**}	-0.147^{*}	0.143^{*}	0.448^{**}
	(0.067)	(0.091)	(0.077)	(0.067)	(0.066)	(0.075)
Asian Non-poor	-0.220^{+}	0.367^{*}	0.091	-0.169^{+}	-0.194^{**}	0.114
	(0.113)	(0.147)	(0.111)	(0.089)	(0.067)	(0.077)
Asian Poor	0.797^{**}	1.081^{**}	-0.446^{+}	−0.352	0.109	0.349^{**}
	(0.165)	(0.173)	(0.253)	(0.215)	(0.130)	(0.079)
Price ($1000)		-0.117^{**}		0.345^{**}		0.015
		(0.025)		(0.036)		(0.018)
% College		-0.027^{**}		0.031^{**}		0.032^{**}
		(0.004)		(0.003)		(0.002)
% Foreign Born		-0.025^{**}		0.008^{**}		-0.012^{**}
		(0.005)		(0.002)		(0.003)
% Units Renter Occ		-0.005^{*}		0.007^{**}		0.014^{**}
		(0.002)		(0.001)		(0.001)
% Unit Built after 2010		-0.015^{**}		0.040^{**}		-0.005^{*}
		(0.005)		(0.003)		(0.002)
% Vacancy		0.026^{**}		-0.009^{**}		0.017^{**}
		(0.005)		(0.003)		(0.003)
Observations	1,690,982	1,690,982	1,690,982	1,690,982	1,690,982	1,690,982
R^2	0.184	0.225	0.226	0.342	0.243	0.331
Adjusted R^2	0.184	0.225	0.226	0.342	0.243	0.331
Residual Std. Error	1.659	1.616	1.347	1.241	1.103	1.036

$^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$

Note: Dependent variables are log transformed. The base category for neighborhood type is White non-poor neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 15% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

C-3 Modeling Decisions

C-3-a Regression Models Including Price Outliers

In this section, we show that our results are robust to the decision to remove listings with a posted price higher than $10,000. When we include these listings we find minimal differences in the results from the regression models reported in the main text.

Table A18: Including Listings Priced Higher than $10,000

	Dependent variable:		
	Logistics Topic	Unit Amenities Topic	Neighborhood Amenities Topic
	(1)	(2)	(3)
White Poor	0.519**	−0.232**	0.382**
	(0.084)	(0.080)	(0.075)
Black Non-poor	0.193+	−0.164**	−0.037
	(0.104)	(0.052)	(0.061)
Black Poor	0.491**	−0.223**	0.281**
	(0.123)	(0.066)	(0.064)
Latino Non-poor	0.360**	−0.093	0.346**
	(0.085)	(0.076)	(0.055)
Latino Poor	0.664**	−0.304**	0.477**
	(0.150)	(0.081)	(0.076)
Asian Non-poor	0.250*	−0.127+	−0.055
	(0.115)	(0.066)	(0.092)
Asian Poor	0.800**	−0.380	0.440**
	(0.276)	(0.234)	(0.078)
Price ($1000)	0.00000	0.00000**	0.00000*
	(0.00000)	(0.00000)	(0.00000)
% College	−0.015**	0.020**	0.018**
	(0.002)	(0.001)	(0.001)
% Foreign Born	−0.017**	0.004*	−0.010**
	(0.004)	(0.002)	(0.002)
% Units Renter Occ	−0.009**	0.006**	0.010**
	(0.001)	(0.001)	(0.001)
% Unit Built after 2010	−0.014**	0.050**	−0.004
	(0.005)	(0.004)	(0.002)
% Vacancy	2.653**	−0.788**	1.589**
	(0.273)	(0.209)	(0.212)
Observations	1,695,948	1,695,948	1,695,948
R^2	0.229	0.333	0.346
Adjusted R^2	0.229	0.333	0.346
Residual Std. Error	1.613	1.250	1.025

$^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$

Note: Dependent variables are log transformed. The base category for neighborhood type is White non-poor neighborhoods. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 30% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at MSA level.

C-3-b Clustering by Census Tract v MSA

Because our models include MSA fixed effects, we cluster standard errors by MSA. However, because the correlation among advertisements within tracts is greater than the correlation within MSAs, we also estimate models clustering standard errors by tract. The results are substantively unchanged. Some standard errors are slightly larger and some are slightly smaller.

Table A19: Clustering by Census Tract

	Dependent variable: Log Transformed Topic Proportion					
	Logistics Topic		Unit Amenities Topic		Neighborhood Amenities Topic	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
Neighborhood Type						
White Poor	0.436**	0.520**	−0.215**	−0.234**	0.620**	0.382**
	(0.075)	(0.076)	(0.075)	(0.069)	(0.058)	(0.052)
Black Non-poor	0.487**	0.195*	−0.594**	−0.164**	−0.304**	−0.036
	(0.083)	(0.082)	(0.060)	(0.055)	(0.048)	(0.045)
Black Poor	0.995**	0.497**	−0.792**	−0.239**	0.193**	0.280**
	(0.085)	(0.090)	(0.062)	(0.055)	(0.044)	(0.044)
Latino Non-poor	0.364**	0.362**	−0.490**	−0.095+	−0.124**	0.347**
	(0.061)	(0.077)	(0.048)	(0.052)	(0.038)	(0.042)
Latino Poor	0.706**	0.664**	−0.822**	−0.305**	0.080	0.478**
	(0.099)	(0.109)	(0.071)	(0.071)	(0.056)	(0.062)
Asian Non-poor	−0.270**	0.236*	0.112	−0.081	−0.232**	−0.057
	(0.089)	(0.099)	(0.083)	(0.077)	(0.056)	(0.072)
Asian Poor	0.389+	0.802**	−0.303+	−0.393**	0.494**	0.439**
	(0.216)	(0.205)	(0.179)	(0.125)	(0.144)	(0.081)
Unit and Neighborhood Covariates						
Price ($1000)		−0.114**		0.336**		−0.005
		(0.014)		(0.011)		(0.009)
% College		−0.013**		0.017**		0.018**
		(0.001)		(0.001)		(0.001)
% Foreign Born		−0.017**		0.004**		−0.010**
		(0.002)		(0.001)		(0.002)
% Units Renter Occ		−0.009**		0.007**		0.010**
		(0.001)		(0.001)		(0.001)
% Unit Built after 2010		−0.013**		0.047**		−0.004+
		(0.003)		(0.002)		(0.002)
% Vacancy		2.691**		−0.911**		1.592**
		(0.233)		(0.166)		(0.139)
MSA Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	1,692,639	1,692,639	1,692,639	1,692,639	1,692,639	1,692,639
Number of Census Tracts	37,319	37,319	37,319	37,319	37,319	37,319
R^2	0.179	0.230	0.231	0.349	0.241	0.346
Adjusted R^2	0.179	0.230	0.231	0.349	0.241	0.346
Residual Std. Error	1.664	1.611	1.342	1.235	1.104	1.025

$^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$

Note: Topic proportions for the three topics are computed by STM (Roberts et al. 2014). Dependent variables are log transformed. The base category for neighborhood type is White non-poor neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 30% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at census tract level.

C-3-c Including Month Fixed Effects

Given the potential for seasonality effects in the rental market, we also estimate models with month fixed effects. Results are unchanged.

Table A20: Month Fixed Effects

	Dependent variable:					
	Logistics Topic		Unit Amenities Topic		Neighborhood Amenities Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
White Poor	0.433**	0.517**	−0.213*	−0.231**	0.620**	0.382**
	(0.088)	(0.084)	(0.093)	(0.078)	(0.084)	(0.075)
Black Non-poor	0.488**	0.195+	−0.595**	−0.165**	−0.304**	−0.036
	(0.105)	(0.104)	(0.052)	(0.051)	(0.054)	(0.061)
Black Poor	0.992**	0.493**	−0.789**	−0.235**	0.193**	0.280**
	(0.112)	(0.122)	(0.063)	(0.066)	(0.064)	(0.063)
Latino Non-poor	0.364**	0.360**	−0.490**	−0.093	−0.124+	0.347**
	(0.065)	(0.084)	(0.081)	(0.072)	(0.072)	(0.055)
Latino Poor	0.704**	0.661**	−0.820**	−0.302**	0.080	0.477**
	(0.140)	(0.148)	(0.078)	(0.079)	(0.074)	(0.076)
Asian Non-poor	−0.266*	0.239*	0.108+	−0.083	−0.232**	−0.057
	(0.109)	(0.114)	(0.058)	(0.065)	(0.065)	(0.092)
Asian Poor	0.379	0.792**	−0.293	−0.381+	0.493**	0.438**
	(0.286)	(0.267)	(0.315)	(0.215)	(0.159)	(0.077)
Price ($1000)		−0.114**		0.337**		−0.005
		(0.026)		(0.037)		(0.016)
% College		−0.013**		0.017**		0.018**
		(0.002)		(0.001)		(0.001)
% Foreign Born		−0.017**		0.004*		−0.010**
		(0.004)		(0.002)		(0.002)
% Units Renter Occ		−0.009**		0.007**		0.010**
		(0.001)		(0.001)		(0.001)
% Unit Built after 2010		−0.013*		0.047**		−0.004
		(0.005)		(0.004)		(0.002)
% Vacancy		2.684**		−0.904**		1.592**
		(0.269)		(0.191)		(0.212)
MSA Fixed Effects	Y	Y	Y	Y	Y	Y
Month Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	1,692,635	1,692,635	1,692,635	1,692,635	1,692,635	1,692,635
R^2	0.180	0.231	0.233	0.351	0.241	0.346
Adjusted R^2	0.180	0.231	0.233	0.351	0.241	0.346
Residual Std. Error	1.662	1.610	1.341	1.233	1.104	1.025

+p<0.1; *p<0.05; **p<0.01

Note: Dependent variables are log transformed. The base category for neighborhood type is White non-poor neighborhoods. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 30% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Models include MSA and month fixed effects. Standard errors are clustered at MSA level.

D. MNIR Robustness Checks

D-1 MNIR Coefficients for Figure 4

Table A21: MNIR Coefficients for Figure 4

	% Black		% Poverty		% Education	
	Words	Coefficients	Words	Coefficients	Words	Coefficients
1	evictions	0.0196	campus	0.0543	foods	0.0450
2	section	0.0195	students	0.0432	rooftop	0.0382
3	polis	0.0151	exposed	0.0369	uptown	0.0326
4	applicants	0.0147	evictions	0.0358	lobby	0.0317
5	eat	0.0146	university	0.0351	concierge	0.0316
6	brick	0.0135	lofts	0.0303	boutique	0.0301
7	ups	0.0134	museum	0.0286	rise	0.0300
8	train	0.0133	section	0.0280	bicycle	0.0297
9	exposed	0.0133	historic	0.0274	midtown	0.0292
10	hook	0.0132	august	0.0272	union	0.0278
11	needed	0.0131	studios	0.0270	whole	0.0268
12	hookup	0.0131	midtown	0.0270	marble	0.0265
13	clothes	0.0129	proof	0.0259	red	0.0247
14	discounts	0.0128	duplex	0.0255	nw	0.0246
15	affordable	0.0127	brick	0.0250	subway	0.0244
16	wall	0.0127	secured	0.0245	museum	0.0239
17	perfectly	0.0127	ave	0.0241	nightlife	0.0239
18	criminal	0.0125	needed	0.0235	elevator	0.0238
19	proof	0.0122	study	0.0234	backsplashes	0.0237
20	de	0.0120	original	0.0230	hill	0.0235
21	money	0.0119	criminal	0.0227	bike	0.0233
22	military	0.0118	arts	0.0224	streets	0.0233
23	income	0.0116	applicants	0.0221	broker	0.0231
24	hospital	0.0115	recently	0.0218	underground	0.0231
25	porch	0.0115	sky	0.0218	skyline	0.0227
26	app	0.0115	intercom	0.0218	lined	0.0226
27	years	0.0114	income	0.0217	desk	0.0226
28	rates	0.0114	landlord	0.0213	wine	0.0225
29	metro	0.0111	block	0.0210	blocks	0.0223
30	exciting	0.0111	bus	0.0209	classes	0.0219
31	entrances	0.0111	sewer	0.0207	yoga	0.0216
32	choice	0.0111	field	0.0207	clubroom	0.0215
33	br	0.0109	medical	0.0206	urban	0.0214
34	basement	0.0109	street	0.0206	neighborhoods	0.0212
35	alarm	0.0109	porch	0.0205	showers	0.0212
36	must	0.0109	stadium	0.0204	charm	0.0209
37	townhomes	0.0107	building	0.0203	quartz	0.0209
38	connections	0.0107	de	0.0202	charging	0.0204
39	extraordinary	0.0107	painted	0.0199	racks	0.0204
40	rear	0.0106	skyline	0.0199	lines	0.0199
41	university	0.0106	roof	0.0196	building	0.0193
42	background	0.0104	br	0.0195	steps	0.0191
43	mini	0.0102	affordable	0.0193	sky	0.0190
44	historic	0.0101	college	0.0193	dry	0.0188
45	care	0.0099	pay	0.0191	conference	0.0188
46	anytime	0.0099	st	0.0189	shops	0.0187
47	pointe	0.0099	electricity	0.0188	starbucks	0.0185
48	largest	0.0097	pays	0.0186	deep	0.0185
49	portal	0.0097	line	0.0185	glass	0.0182
50	duplex	0.0096	lines	0.0185	stadium	0.0181

D-2 MNIR with Different Preprocessing

In this section, we demonstrate that our MNIR results are robust to our preprocessing procedures. We use 1% as a threshold for removing low frequency words in the MNIR model in the main text. Here, we report the MNIR result from a different threshold: 0.7%. We remove words that appear less than 0.7% of the documents. Given we have 1,696,499 documents, the word needs to appear at least in 11,875 documents to not be removed.

Higher % Black	Higher % Poverty	Higher % College
wall–to–wall	campus	floor–to–ceiling
stubs	student	foods
budget	students	trader
mills	exposed	capitol
priced	university	rooftop
evictions	evictions	triangle
section	lofts	concierge
eat–in	museum	uptown
renoted	section	lobby
scious	historic	facing
applition	august	panoramic
secuty	stubs	bicycle
renovations	midtown	midtown
polis	de	boutique
hook–ups	studios	union
applicants	proof	cafes
order	duplex	whole
suburban	brick	actual
qualify	hospitals	marble
recent	secured	queen
de	hollywood	broker
discounts	ave	tower
exposed	shuttle	subway
bedroo	off–street	nw
posit	sky	joe
hookup	recent	trendy
brick	study	elevator
affordable	original	museum
perfectly	criminal	nightlife
train	renoted	backsplashes
broad	hook	smoke–free
handicap	showings	red
criminal	sewage	hill
proof	arts	streets
ups	intercom	bike
hospital	recently	capital
military	needed	highlands
artnts	applicants	underground
app	block	desk
metro	music	blocks
hook	income	repair
livg	skyline	shuttle
confirm	landlord	wine
years	stadium	skyline
rates	eat	classes
exciting	medical	urban
money	bus	broad
interstates	secuty	walkable
porch	sewer	yoga
choice	roommate	santa

Figure A7: Words with Top 50 Correlation with Neighborhood Covariates

D-3 MNIR Results for Different SES Covariates

In this section, we present MNIR results with different covariates. We run MNIR with census tract median household income and % White. The first column lists the top 50 words that are strongly associated with higher median household income. The top 50 words show that advertisements in neighborhoods with higher median household income tend to have words that emphasize housing and neighborhood amenities. For example, words such as 'whole,' 'foods,' 'metro,' 'yoga,' 'starbucks' describe neighborhood amenities. Words that describe higher-end housing unit amenities 'marble,' 'concierge,' 'whirlpool,' 'high-end,' 'sinks,' 'quartz,' 'showers,' are more likely to appear as the neighborhood median household income increases.

The next column displays the top 50 words that are likely to appear when listings are in neighborhoods with lower median household income. Similar to the results from % Black (presented in the main text), words that emphasize renter qualifications such as 'evictions,' 'section' (8), 'criminal,' 'background,' 'screened,' 'income,' 'application,' 'money,' 'must' are more likely to appear as the neighborhood median household income decreases. There are a few words that describe the unit and the neighborhood. For example, 'historic' and 'hospital' are likely to describe the neighborhood. But the number of words describing neighborhood amenities are more limited compare to the first column. Words such as 'hookup,' 'lofts,' 'porch,' 'painted,' 'intercom' describe housing amenities that are not high-end features.

The results for % White (presented in the third column of Figure A8) show a less obvious pattern than other covariates. They includes word that describe neighborhood amenities ('whole,' 'foods,' 'theatre') and unit amenities ('whirlpool,' 'lawn,' 'carports'). However, these words appear less frequently. They are also more likely to represent high-end neighborhood and unit amenities.

Higher Median HH Income	Lower Median HH Income	Higher % White
ranch	evictions	snow
foods	campus	non–smoking
san	students	removal
wine	polis	salon
backsplashes	section	cherry
cabanas	criminal	woods
marble	hook–ups	commons
classes	hook	salt
hiking	off–street	acre
elementary	proof	underground
concierge	university	foods
exquisite	posit	country
metro	sewer	biking
union	affordable	nw
acre	duplex	collection
biking	background	pond
charging	hookup	tanning
whirlpool	screened	lawn
whole	hospital	basement
high–end	closed	hill
ge	ups	bike
broker	hookups	august
smoke–free	lofts	streets
pendant	applicants	union
sinks	clud	wine
ball	income	level
level	clothes	river
quartz	artnts	lower
elegant	historic	round
yoga	porch	overlooking
corporate	entrances	smoking
showers	pay	pendant
reflect	studios	charm
starbucks	painted	newer
pre–wired	intercom	whirlpool
gaming	vertical	main
guests	application	saltwater
winning	eat	acres
commitment	app	finished
hill	original	carports
nw	aable	july
party	money	whole
trails	secured	theatre
seating	specl	golf
attached	refreshing	bonus
cyber	must	utility
conference	patrol	indoor
prestigious	ve	polis
molding	facilities	appointed
soaking	facility	november

Figure A8: Words with Top 50 Correlation with Neighborhood Covariates

E. Supplemental Analysis

E-1 Variation in Amount and Types of Information

In supplemental analyses we examine non-text forms of information inequality. We created three simple measures based on our Craigslist data: (1) general information is a count of the number of distinct (optional) information fields that have been filled out in each post, including the number of bedrooms, bathrooms, and square footage, as well as contact information and the exact listing address; (2) number of pictures is a count of the number of pictures posted with the advertisement; and, (3) number of words is a count of the total number of words in the main text body of each advertisement. While these measures do not tell us anything about the content of advertisements, they offer initial, simple indicators of information differences (see Boeing et al. 2020 for similar analyses). Additionally, these measures capture important dimensions of advertisements that can impact the housing search process. For example, Craigslist allows prospective renters to filter which posts are shown to them based on these information categories; e.g., one can select to only be shown postings that contain pictures, or that have two or more bedrooms. We use a similar modeling approach here as with our topic proportions in the main text.

RESULTS

Table A22 reports descriptive statistics for our numerical (non-text) data. On average, advertisements contain 3.7 distinct fields of information, but range as low as 0 and as high as 5. We see wider ranges in the number of pictures (ranging from 0 to 24, with an average of 9) and overall word count (ranging from 6 – 3,782 words, with an average of about 183 words). Table A22 also reports descriptive statistics for all of our independent variables, which we draw from the ACS.[1]

[1] The sample size for information regressions and topic model regressions differ because there are plenty of listings that do not have enough text for topic modeling. There are 1,457 listings that have less than 5 English words in their text; in addition, because our text preprocessing procedures remove very low frequency and very high frequency words, we drop additional listings that do not have enough text after preprocessing.

Table A22: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max	Missing
Dependent Variables[a]								
General Information	1,693,890	3.74	0.99	0	3	4	5	0
Number of Pictures	1,693,890	9.02	5.96	0	5	12	24	4
Number of Words	1,693,890	182.57	115.04	6	99	244	3,782	0
Neighborhood Type[b]								
White Non-Poor	1,179,452	69.2						
White Poor	78,610	4.6						
Black Non-Poor	136,747	8.1						
Black Poor	75,942	4.4						
Latino Non-Poor	141,015	8.3						
Latino Poor	49,535	2.9						
Asian Non-Poor	28,892	1.7						
Asian Poor	3,697	0.2						
Unit Covariate[a]								
Price ($1,000)	1,693,890	1.42	0.80	0.001	0.91	1.70	10.00	18,604
Tract Covariates[b]								
% College	1,693,890	41.14	21.42	0.00	23.01	58.06	100.00	2,500
% Foreign Born	1,693,890	16.42	12.72	0.00	6.83	22.80	100.00	2,499
% Units Renter Occupied	1,693,890	52.82	23.72	0.00	34.72	71.06	100.00	3,250
% Units Built Post 2010	1,693,890	3.15	5.18	0	0	4.0	88.00	3,250
% Vacancy	1,693,890	10.104	7.62	0.00	5.09	13.03	95.77	3,202
Variables for MNIR[b]								
% Black	1,693,890	17.134	21.263	0.000	3.200	21.500	100.000	2,499
% College	1,693,890	41.136	21.424	0.000	23.010	58.060	100.000	2,500
Poverty Rate	1,693,890	16.161	11.915	0.000	7.370	21.910	100.000	3,224

[a] Source: Craigslist
[b] Source: 2016 American Community Survey 5-year pooled data

Clearly, there is wide variation in the amount of information included in Craigslist rental housing advertisements. To explore this variation, we estimate an OLS model regressing each outcome on neighborhood type, testing whether there is systematic variation across different types of neighborhoods. We then estimate pairwise difference across neighborhood types by changing the baseline category in our regression models and generating predicted values. Figure A9 reports pairwise comparisons of neighborhood types for each of our information outcomes.

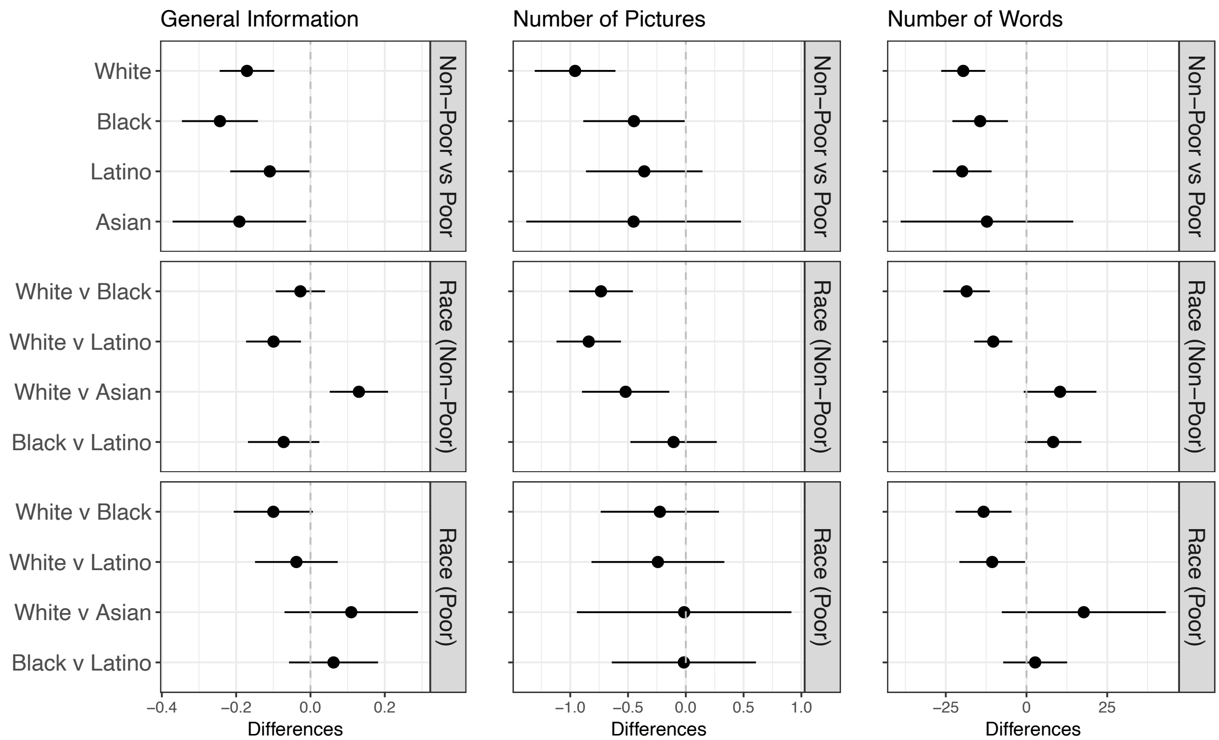

Figure A9: Pairwise Comparison of Information Outcomes Across Neighborhood Types

Note: The dots and bars indicate the regression coefficients and confidence intervals from regression models. The neighborhood written first (before "vs") is the base category. The first four rows (non-poor vs poor) display the regression coefficients for poor neighborhoods when the non-poor neighborhoods for the respective racial group is the base category. Negative value means poor neighborhoods have less information than non-poor neighborhoods. Positive value means non-poor neighborhoods have less information than poor neighborhoods. The following eight rows compare differences between different racial compositions. The first racial group is the base category. For example, the coefficient for "White v Latino" in the sixth row for the number of pictures indicates that Latino non-poor neighborhoods have 0.84 less pictures than White non-poor neighborhoods. Neighborhood racial composition and poverty rate are obtained from 2016 ACS 5-year pooled data. The regression models include MSA fixed effects. Standard errors are clustered at census tract level.

Figure A9 focuses on theoretically important comparisons across neighborhood types by plotting the differences in each outcome for given pairs of neighborhoods. Starting with the top panel for each outcome measure, we can see a clear pattern across outcomes: listings in poor neighborhoods tend to contain significantly less information than their same-race, non-poor counterparts. The magnitude and significance of these information gaps varies somewhat across our information measures, but it is clear that holding racial composition constant, advertisements in poorer neighborhoods provide far less information to searchers.

However, poverty status does not account for all the variation in information levels. The middle panel for each outcome compares information differences among non-poor neighborhoods by race. Across measures we consistently find that listings in non-poor Black and Latino neighborhoods contain less information than those in non-poor White neighborhoods. In other words, even when we just compare neighborhoods with similarly low poverty levels, Black and Latino neighborhoods face an information disadvantage compared to White neighborhoods. Advertisements for housing in non-poor Asian neighborhoods contain more overall information and number of words compared to those in non-poor White neighborhoods, but fewer pictures on average, suggesting that listings in Asian neighborhoods do not experience the same racial penalty as Black and Latino neighborhoods. Nevertheless, we find clear evidence of both a socioeconomic and racial hierarchy in terms of basic information: advertisements in neighborhoods with more Black and/or Latino residents and/or more poor households contain significantly less information—measured as the number of distinct information fields, the number of pictures, and the number of overall words provided within advertisements—than do advertisements in neighborhoods with more Asian, White, and/or non-poor residents.

Finally, the bottom panel in Figure A9 measures racial differences in listing information levels among poor neighborhoods. Even among poor neighborhoods we find some evidence of racial inequality: listings in poor Black and Latino neighborhoods have significantly fewer words than those in poor White neighborhoods. While racial differences among poor neighborhoods tend to be smaller in terms of magnitude and are not statistically significant across all outcome measures, they nevertheless underscore the importance of accounting for neighborhood race and poverty status simultaneously to fully understand differences in access to information.

Table A23 repeats our analysis but includes additional neighborhood covariates. We find that higher priced listings tend to contain more information on average, as do listings in neighborhoods with greater proportions of college-educated residents or immigrants. Interestingly, we also find that measures of stronger rental market competition (rental occupancy rate and an indicator of recent construction activity) are also associated with greater information, while higher vacancy rates—an indicator of a weaker rental market—are associated with less information. Searchers who are limited to looking in less desirable neighborhoods have access to far less information about their potential homes.

Overall, our supplementary analyses find clear evidence of neighborhood racial and socioeconomic inequalities in the amount of information presented in rental housing advertisements. These results offer futher support for the conclusions we draw using the text data about racial and socioeconomic inequality in the information provided in housing advertisements.

Table A23: Regression Results Predicting Information Measures with Additional Neighborhood Covariates

	Dependent variable:		
	General Information	Num Pics	Num Words
Neighborhood Type			
White Poor	-0.194^{**}	-1.008^{**}	-24.444^{**}
	(0.037)	(0.176)	(3.644)
Black Non-poor	0.016	-0.179	-10.079^{**}
	(0.035)	(0.135)	(3.737)
Black Poor	-0.171^{**}	-0.503^{*}	-20.563^{**}
	(0.045)	(0.208)	(3.536)
Latino Non-poor	-0.151^{**}	-0.427^{**}	-8.637^{*}
	(0.042)	(0.156)	(3.554)
Latino Poor	-0.255^{**}	-0.668^{**}	-27.832^{**}
	(0.047)	(0.237)	(4.648)
Asian Non-poor	0.006	-0.635^{**}	-1.596
	(0.048)	(0.215)	(6.202)
Asian Poor	-0.190^{*}	-1.283^{**}	-14.516
	(0.079)	(0.420)	(11.739)
Unit and Neighborhood Covariates			
Price ($1000)	0.182^{**}	1.956^{**}	16.301^{**}
	(0.009)	(0.053)	(0.931)
% College	0.0002	0.007^{**}	0.269^{**}
	(0.0005)	(0.002)	(0.052)
% Foreign Born	0.005^{**}	0.010^{*}	0.398^{**}
	(0.001)	(0.004)	(0.109)
% Units Renter Occ	0.003^{**}	0.012^{**}	0.357^{**}
	(0.0004)	(0.002)	(0.044)
% Unit Built after 2010	0.004^{**}	0.014^{*}	1.218^{**}
	(0.001)	(0.007)	(0.177)
% Vacancy	-0.011^{**}	-0.026^{**}	-0.791^{**}
	(0.001)	(0.005)	(0.117)
MSA Fixed Effects	Y	Y	Y
Observations	1,693,890	1,693,890	1,693,890
Number of Census Tracts	37,392	37,392	37,392
R^2	0.107	0.088	0.110
Adjusted R^2	0.107	0.088	0.110
Residual Std. Error	0.932	5.693	108.556

$^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$

Note: The base category for neighborhood type is White non-poor neighborhoods. Listings that are more expensive than $10,000 are removed. For neighborhood classification, racial composition is based on the plurality racial group. We use a threshold of 30% of tract poverty rate as our measure of neighborhood poverty. Neighborhood covariates are obtained from 2016 ACS 5-year pooled data. Standard errors are clustered at census tract level.