Seattle Rental Ad Texts and Processes of Segregation

Ian Kennedy

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Arts

University of Washington

2019

Committee

Kyle Crowder

Jerald Herting

John Wilkerson

Program Authorized to Offer Degree:

Sociology

University of Washington

Abstract

Seattle Rental Ad Texts and Processes of Segregation

Ian Kennedy

Chair of the Supervisory Committee:
Kyle Crowder
Department of Sociology

Recent work on racial residential segregation shows how individuals' perceptions of neighborhoods influence their housing choices and therefore contribute to the reproduction of segregation. Rental housing advertisements are a form of public discourse about neighborhoods which both influence and reflect those perceptions. I use n-gram regression and Structural Topic Models (STM) to investigate whether and how rental listings from the Seattle metropolitan area *Craigslist* page differ in association with neighborhood racial proportion. Neighborhoods with higher White proportion are associated with inviting words like 'restaurant' and 'charming,' while less-White neighborhoods are connected to security terms like 'gated' or transportation terms like 'light rail'. STM and qualitative analysis shows that listings from White neighborhoods emphasize connections to neighborhood history and culture, while listings from non-white neighborhoods offer more incentives and focus on development features, sundering these areas from their surroundings. Finally, analysis of security discourse reveals that not only is language about security more common in less White neighborhoods, but administrative data show that actual security systems are less common. Without mentioning race, these listings reveal racialized neighborhood perceptions which likely impact individual's neighborhood choice in ways that contribute to housing segregation.

In contrast to narratives of racial progress, housing segregation has persisted as a key contributor to disadvantage based on race. Though the Fair Housing Act passed in 1968, and despite the creation of national laws mandating fair treatment and a cabinet-level position to oversee its enforcement, racial housing segregation remains entrenched, changing least for people of all races near the bottom of the income distribution (Intrator, Tannen, and Massey 2016; Massey and Tannen 2015). Growing up in a segregated area has lasting influence on education (Massey and Denton 1993; Jencks and Mayer 1990), health (Williams and Collins 2001; Gibbons and Yang 2014), and other outcomes.

Fully understanding the reproduction of segregation requires attention on the processes that perpetuate it. State sponsored racist policies institutionalized segregation for decades before they were prohibited. The most common explanations for why unequal housing patterns continue—structural factors like racial differences in human capital, discrimination, and homophily—have been insufficient (Crowder and Krysan 2016). Instead, developments in theories of race and racism as well as new research in residential practices suggest that knowledge about and perceptions of neighborhoods shape the processes behind residential attainment—how people find their way to the area and home where they live, or how landlords market units and select tenants (Bader and Krysan 2015; Krysan and Crowder 2017). By viewing segregation as the outcome of aggregate individual action, this approach allows for the investigation of what contributes to neighborhood knowledge and perception while still accounting for structural influences on that action. While personal perceptions cannot be directly examined, discursive production about neighborhoods—like online rental listings—can be. The content and composition of that discourse is severely under researched.

The purpose of this study is to investigate the themes and narratives that are common in rental advertisements from neighborhoods with different racial compositions. I use rental housing listings from *Craigslist* in the area around Seattle, WA as a case study to understand how those texts vary systematically with neighborhood racial proportion. Identifying such patterns in a metro-wide analysis reveals not just individual associations or perceptions about particular neighborhoods, but widespread ideas Seattleites hold about the kinds of places where different groups of people live.

I begin by investigating the texts at the word level, testing the existence of significant discursive differences by neighborhood racial proportions. Then, I use unsupervised topic modeling, a method that recognizes groups of words that often appear together and uses these groups to identify topics. I use regression to identify how much each topic is associated with neighborhood racial proportion. I describe that association and the topics that represent it using qualitative analysis and deep reading. There are differences in rental listings across neighborhoods that are both statistically and practically significant and which may influence housing search patterns by shaping the way home-seekers think about the places they might choose to live. Those differences align with psychological and geographical work on the racialization of place (Bonam et al. 2016; Bonam et al. 2017; Bonam et al. 2018; Inwood and Yarbrough 2010) and the way new White residents often fail to integrate into existing community structures when moving to less White neighborhoods (Walton 2018). My analysis reveals Seattle's landscape of racialized neighborhood perception and explores how that landscape could be implicated in the reproduction of residential segregation.

Background

Like in many cities in the United States, residential segregation in Seattle was state sanctioned and enforced by redlining and restrictive covenants for decades (Silva 2009; McGee 2007; Rothstein 2017). However, segregation persisted even as those barriers fell. The city's racial geography, produced through law and state policy, seems to be maintained by more subtle forces acting through everyday practices like finding a home or choosing a tenant. Sociology has been successful in documenting the problem but has not identified a robust causal account of residential segregation. This paper uses text analysis methods to investigate a social process driven by public discourse that could contribute to that persistence by showing the different ways community, transportation, and safety appear in listings from neighborhoods with different racial compositions.

Historical discrimination and racist policies shaped contemporary housing dynamics and racial residential segregation. But state action and other explicit sources are not solely responsible for the reproduction of segregation. This paper uses a process-based approach to housing dynamics paired with a critical race theory perspective on racism to investigate possible ways racialized discourse could contribute to the systemic reproduction of segregation, in the absence of explicit discrimination or racist intention. This is possible by focusing on texts as a discursive site where individual and public perceptions—individual and aggregate action—meet. In those contexts, words and phrases associated with racial residential patterns reveal the racialized nature of discourse and residential attainment.

Racialized Neighborhood Perceptions

I define racialized perception as the correlation between, on one hand, the association of an idea and an object and, on the other, a race and that object. This definition borrows from psychological research into the racialization of space (Bonam et al. 2018). If people associate Whiter neighborhoods in Seattle with walkability in general, then I describe that association as racialized perception, even if (actually, regardless of whether) those neighborhoods are more walkable. Put another way, in public perception, Whiteness and walkability are associated because they both describe the same places. A carefully produced study of spatial variation in racialized neighborhood perceptions would give insight into a hitherto unseen factor in housing searches.

Moreover, racialized neighborhood perceptions might inform the way people act in neighborhoods after they move there. Walton (2018), through her ethnographic work in stably diverse neighborhoods in Boston, suggests some forms these racialized perceptions can take. She identifies two of what she calls ‘habits of Whiteness’: (1) anxiety—worry about the security of a neighborhood or if it is a sufficient place to live; and (2) ambivalence—uncertainty about the value of existing neighborhood culture, amenities, and social networks. My analysis shows that both anxiety and ambivalence are present in Seattle’s racialized neighborhood perceptions.

The concept of racialized perceptions is superficially similar to both the psychological concept of implicit bias and political dog whistle racism. All three are notable because they include racialized meaning without explicit racial content. Haney-Lopez (2015) and Bonilla-Silva (2014) both argue that apparently colorblind language is an essential contributor to the perpetuation of America’s racial order. However, unlike implicit bias or dog whistle racism, racialized perceptions do not rely on or assume conscious or unconscious racist intent. Dog whistle racism is coded language, where specific terms signal specific racial meanings. That is not true of racialized neighborhood perceptions where racialization is an association between language and neighborhood populations, and a simple mapping of terms is not sufficient. Implicit bias is more similar, especially in its focus on how unnoticed associations with race might cause us to act differently. However, neighborhood racial perceptions are public perceptions present in public discourse, a strong contrast with implicit bias, which exists within individual minds.

Systemic Race and the Reproduction of Residential Segregation

Racialized neighborhood perceptions align with contemporary sociological theories of racism. While traditional accounts of racial discrimination focus on racially biased treatment of one individual by another, alternatives suggest that discriminatory outcomes can occur as part of a process without explicit connection to race on the part of the actor (Bracey 2015; Golash-Boza 2016; Reskin 2012; Pager and Shepherd 2008). These new approaches focus on colorblind racism (Bonilla-Silva 2014) and racial habitus (Emirbayer and Desmond 2015), where racism is enacted and perpetuated in part through discourse and action that is not obviously racial. This view naturally incorporates the way that race and other social identifications are intertwined in the way they attach to people and the way they influence differential outcomes.

The racialization of neighborhood perceptions takes place in a social world with various axes of difference. In the context of this paper, understanding how race, income, and wealth are intertwined is especially important. Historical differences in economic resources by race, caused by systemic racism and discrimination, mean that less White neighborhoods tend to be poorer and have older housing stock. By the same token, non-White, especially Black, home-seekers also tend to have lower incomes and less wealth than White home-seekers. These differences are linked in the social imaginary, and in neighborhood perceptions, through stereotypes about black neighborhoods as poor and dangerous (Bonam, Bergsieker, and Eberhardt 2016; Quillian and Pager 2001; Bresbis, Faber, Rich, and Sharkey 2015). This intersection between race and class is not additive, in the sense that it can be accounted for by a race effect plus income, education, or wealth effects. Nor is it an interaction that can be explained by racially different slopes for economic or other variables (Zuberi and Bonilla-Silva 2008). Instead, in the lives of everyday people, the meaning of race and class are entwined so that each composes the other (Bonilla-Silva 2018), which requires complementary qualitative analysis to describe. In part, this view implies that discrimination in housing, school, labor, and credit are not independent, but form an integrated system. Small differences in each area compound to produce larger differences overall. Reskin (2012) refers to this larger difference as 'über discrimination,' a way to think about the complex interactions between areas of discrimination. Within that system, Reskin (2012) identifies housing segregation as both particularly intractable, and as a leverage point from which the whole system might be challenged.

Taking account of both the implicit discursive nature of contemporary discrimination and its systemic nature at the same time is difficult. Pager and Shepherd (2008) recognize that large scale studies are usually stuck with examining unequal outcomes, leaving investigating the role of preferences to smaller, qualitative scenarios like interviews and experiments. New data sources like *Craigslist* come with new challenges, but they also offer a chance to take hold of both the discursive and systemic aspects of contemporary racism. The crux is seeing *Craigslist* listing texts as discursive productions—and therefore rough reflections—of public neighborhood perception. Listing texts are not an inert record of neighborhood perception, but, by influencing the perceptions of the people who read them, they contribute to the reproduction and reinforcement of neighborhood perceptions through time. This means that the individual action of posting or reading an advertisement and the large-scale outcome of housing patterns are linked by the listing texts.

Using native data like *Craigslist* rental listings is new in research on neighborhood attainment. Most work on neighborhood perception and segregation has been done in experimental or interview settings (Bader and Krysan 2015, Pager and Shepherd 2008; Bonam et al. 2016), and it is unclear if those findings are externally valid, reflecting real residential attainment patterns. I use text analysis techniques, like topic modeling (Dimaggio et al. 2013; Egami et al. 2017) to analyze large corpuses of publicly available documents. These text analysis methods have been used to examine public perceptions of pollution (Tvinnereim, Liu and Jamelske 2017) and of popular music (Light, and Odden 2017). By training these methods on a sample of listings collected from Seattle's largest online housing market, I explore how apartment listings reflect neighborhood perceptions, which are central to the housing search processes that shape segregation. Moreover, I show how descriptions of less-White neighborhoods as dangerous and uninteresting conform to expectations based on empirical work on White habits in stably-diverse neighborhoods (Walton 2018).

A Process-based Approach to the Reproduction of Residential Segregation

The persistence of racial housing segregation has been instrumental to maintaining racial inequality over the last century (Bader and Krysan 2015; Crowell and Fossett 2017; Massey and Denton 1993). Rothstein (2017) argues that historical discrimination drives the current state of racial inequality; the active but subtle mechanisms in place today uphold a segregation that is not new. Logan (2017) shows that segregation has been central to urban life in the United States for over 100 years. Housing policy from the early and mid 20th century—during a period of booming suburbanization—drove not only the segregated population distribution in many metropolitan areas, but also the changes in the urban economy and built environment. Redlining by the Home Owners' Loan Corporation prohibited home loans in neighborhoods with high minority representation and prevented loans for rehabilitation and renovation, leading to dilapidated housing stock (Rothstein 2017; Massey and Rugh 2017). This created the baseline conditions for the segregation we see today. The association between neighborhood circumstances like housing stock age and quality, which are not explicitly racial, and the same neighborhood's racial makeup is the result of a complex history. Drawing from theories of systemic racism outlined above emphasizes that differences in rental price, in quality and newness of housing stock, in median income, in education, in quality of neighborhood schools, cannot be considered separately from the racial differences with which they are associated. It seems more reasonable to assume that there has been a long-term reciprocal causal relationship between all of those factors and racial proportion.

Contemporary housing segregation remains a significant problem (Massey and Tannen 2015) and is most calcified for poor people (Intrator, Tannen, and Massey 2016). However, the forces responsible for the reproduction of contemporary segregation are less explicit and less entwined with the state than they were prior to the Fair Housing Act of 1968. Some landlords have deployed a bevy of tactics of discrimination and tenant manipulation to maximize occupancy by preferred tenants (Greif 2018; Korver-Glenn 2018; Rosen 2014). It's important to note that landlords need not have racist intentions in order to reproduce racialized neighborhood segregation. Instead, even in the absence of conscious or unconscious racial intention, there may be associations between the way property owners and managers (POMs) habitually write about neighborhoods with racialized populations. This paper explores a possible way to make those hidden associations visible.

Understanding how rental listing texts align with residential patterns requires incorporating existing knowledge about those patterns. Recent literature on large scale residential outcomes, or location attainment, has generally followed two themes: spatial attainment and place stratification (Crowell and Fossett 2017; Pais, South and Crowder 2012). Spatial attainment models explain racial differences in housing outcomes by focusing on group differences in human capital (Massey and Denton 1985). This view is also roughly consistent with discrete choice models which assume that individuals search all available homes to find the rental that maximizes the amenities and features they want given their economic constraints (Quillian 2015; Galiani, Murphy and Pantano 2015). In contrast, place stratification focuses on how racial discrimination influences where people live, finding that a home-seeker's race strongly influences the neighborhoods where they can easily find a home. Crowell and Fossett (2017) use novel decomposition methods to show that, at least in their study of non-Hispanic White and Lantix home-seekers, both views have empirical support. I advance this literature by examining the raw material that goes into housing outcomes: listing text discourse and the neighborhood perceptions it may engender. By turning towards how people think and write about neighborhoods, I continue a theoretical reorientation away from individual resources and racial identification as predictors of residential attainment. My focus is on the social links between individual action—choosing a neighborhood, a unit, or a tenant—and aggregate outcomes like segregation. I illuminate this connection by exploiting the texts of rental listings as measures of neighborhood perception.

I use the foregoing historical and theoretical background to develop three hypotheses about the way listing text could vary with neighborhood racial proportion. First, theories of segregation more aligned with economic discrete choice models assume that individual home-seekers can objectively assess their options regardless of neighborhood variations. Under this model, we would reasonably expect no significant association in listing text with on neighborhood race. Differences in listing texts should be limited to the information home-seekers require to maximize their utility based on their resources. That leads to hypothesis 1: Craigslist *rental listings do not contain lexical and topic differences that are significantly associated with neighborhood racial composition*. Since neighborhood racial composition is correlated with many other neighborhood features, including economic status, access to amenities, and distance from central areas, a finding in line with hypothesis 1 would also suggest that listing text writers do not change how they write depending on those neighborhood characteristics either.

Considering spatial assimilation's supposition that racial differences in housing are due to group differences in resources implies hypothesis 2: Craigslist *rental listings contain lexical or topic differences that are associated with neighborhood race, but those associations are due to superficially non-racial neighborhood characteristics*. Findings that accord with hypothesis 2 would suggest that text writers may change the way they write about neighborhoods based on neighborhood characteristics like housing stock, median income, or proportion of commuters, which are correlated with racial proportion.

Finally, incorporating theories of racialization and contemporary discrimination along with place stratification theories of segregation suggests hypothesis 3: *Craigslist rental listings contain lexical or topic differences that are associated with neighborhood race, and those differences persist even when including covariates for superficially non-racial neighborhood characteristics.* Findings supporting this hypothesis would be consistent with the idea that text writers write differently about neighborhoods with different racial composition, even when those neighborhoods are similar in terms of other observed neighborhood characteristics.

Methods

Data Collection

I use a large set of rental listing texts from the Seattle, WA *Craigslist* apartments page to conduct a mixed methods analysis of discursive differences associated with neighborhood racial proportion. I identify those differences at the word level using n-gram logistic regression and at the document level using Structural Topic Models (STMs). Those analyses test the existence of discursive differences that could reflect racialized neighborhood perceptions. To investigate the content of those perceptions, I select representative texts using STM and qualitative coding and subject those texts to qualitative analysis and deep reading.

The corpus contains 278,005 rental advertisements posted between March 2017 and September 2018 and obtained using the Helena web-crawler (Chasins and Bodik 2017). The advertisements are geo-located using scraped addresses and then matched to the United States American Community Survey (ACS) data 2012-2016 five-year estimate at the tract level. Seattle is a particularly apt setting for this case study, given its low levels of racial segregation on traditional measures and high levels of neighborhood change (Thomas 2017). Seattle is also a large *Craigslist* market relative to its population, with an average of more than 2000 listings per day.

Duplicate documents weaken the explanatory capability of topic models (Schofield, Thompson, and Mimno 2017) so I reduce the number of near-duplicate texts in the corpus. I leverage a useful quality of topic models to remove very similar texts: duplicates tend to coalesce into a single topic. By iteratively fitting STMs, I can remove all texts that would appear as duplicates to a human coder.[1] In deciding which duplicates to discard, I keep the most recent listing of the ones that are ranked highly similar. This reduces the corpus to 45,358 listings from 848 census tracts in and around Seattle.

[1] I use the 10-gram Jaccard similarity to compare texts, marking as similar texts greater than .3 on that measure. Jaccard distance is fast to compute, straightforward to interpret, and automatically ranges from 0-1. Essentially, that means that to compare two texts, I first make one set for each text of all ordered sequences of 10 characters. I then divide the number of sequences in the intersection of the sets (the number of shared 10-grams) and divide by the number of sequences in the union of the sets. Texts with no matching 10-grams receive a similarity score of 0 and an exact match would be a 1. I found that texts in the sample with similarity scores of greater than .6 tended to be from the same listings but at different times. Scores between .3 and .6 were often from the same property or property manager, but from different units. Score from .2 to .3 sometimes included listings from the same property but differed significantly in the ways they described the unit or the neighborhood. Some listings between .2 and .3 were unrelated and were similar because they included common phrases like ‘granite countertops’ or ‘close to restaurants, shops and more.’ Based on this, I set the threshold for duplicates at .3.

I strip URLs and digits from the texts in the corpus. Additionally, I remove mentions of neighborhoods in the texts, as they are collinear with neighborhood racial makeup and thus not informative features for examining differential framing. Finally, I implement a neighborhood typology based on that used by Crowder, Pais and South (2012) and Hall, Crowder and Spring (2015). I identify seven neighborhood types based on tract-level racial proportions estimated in the 2012-2016 ACS (given in order of prevalence): "Predominantly White," "White Asian," "White Latinx," "White Mixed," "Mixed", "White Black" and "Majority Non-White". There are 246 "Predominantly White" tracts, including 9,991 listings, and 14 "Majority non-White" tracts, including 941 listings, in the sample. Details about the neighborhood typology can be found in Appendix III.

Lexical Differences

In order to establish that there are meaningful differences in the texts that make up housing advertisements in neighborhoods above median White proportion (high-White neighborhoods) and above median Black proportion (high-Black neighborhoods), I conduct two binary classification tasks using n-gram features, where n = 4, in two logistic regression models with L2 regularization, which reduces many coefficients to zero and improves the parsimony of the model.[2] I first train a model to predict whether an ad is from a high-White neighborhood or not, and then one to predict whether an ad is from a high-Black neighborhood or not.[3]

Structural Topic Modeling

[2] Formal model specifications for all 84 models included in the analysis are in Appendix I
[3] Note that these categories are not mutually exclusive: The dataset contains 2196 listings from areas that, at the Census tract level, have above median White and Black populations. However, dropping these from analysis did not change classifier performance, and the resulting feature coefficients were largely the same.

I use STM to quantitatively describe common discursive features of the listings and examine associations between those features and neighborhood racial composition. While the results of the n-gram regression are highly suggestive, they have some significant limitations. Most especially, the terms produced are limited to four words, and so do not take document context into account. There is also a danger that these sets of terms were produced by the analysis process and do not reflect trends in the documents themselves. The n-gram regression processes used will produce sets of terms which most distinguish any subsets of texts. This type of analysis is hard to extend meaningfully into a discussion of the reproduction of segregation. It is possible that these texts simply reflect the conditions in those neighborhoods, not necessarily differences in neighborhood perceptions. Finally, it is difficult to meaningfully include neighborhood and listing controls using n-gram regression. These high and low White neighborhood terms may be mostly about economic and not racial difference.

Topic modeling techniques address these limitations and have additional benefits. First, topic modeling is an unsupervised technique, and so it is possible to produce topic distributions without meaningful co-variation with neighborhood racial proportion. That is, unlike logistic regression on n-grams, topic modeling will not automatically produce variation across any co-variate. Topic modeling produces estimates of topic proportion at the document level, so I can use regression to assess the association between a particular topic and neighborhood racial composition. That shows the association between certain patterns of language (word co-occurrence in STM) and race in Seattle. Moreover, by including additional covariates in this analysis, Topic Models are conducive to analyzing which portions of that association are unexplained by neighborhood characteristics.

Topic modeling is subject to significant analyst discretion, so I confirm the robustness of these findings three ways. First, I check the dependence of the model output on the number of topics selected using the “robust LDA” method adjusted for STM to confirm that important topics appear with various choices of the number of topics, which they do (Casas, Bi, and Wilkerson 2018). Second, I perform a simulation test randomizing the census tract of each document. The observed associations are much larger than we would expect based on the randomization. Third, I report results only from a held-out test set of documents. Results in the training and test sets are substantively consistent.[4] Topic modeling takes into account word co-occurrence, which includes document context and also aligns with relational accounts of meaning (Dimaggio et al. 2013). Similarly, topic models produce output at the document level instead of the word level, making it easy to examine if certain topics are more common in certain neighborhoods.

STM is faster and more reliable than other methods, including Latent Dirichlet Allocation (LDA) topic models (Egami et al. 2017; Roberts, Stewart, and Tingley 2014). STM uses a more consistent initialization process than LDA and can simultaneously estimate topic proportions and associations with covariates.[5] This estimation method, combined with the train-test split and other robustness checks, helps ensure that the associations I observe between discursive forms and racial proportions are present in the corpus and are not induced by nature of the analysis.

I add covariates at both the listing and tract level. At the listing level, I include the log of the rent and the square footage of the listing. At the tract level, I include the neighborhood racial typology, poverty proportion, log of median income, population in thousands, proportion college educated, proportion commuting, proportion of units owner-occupied, proportion of units rented

---

[4] More details and the results of these checks are in Appendix IV.

[5] In the training set, this allows the vector of topic proportions for each document to be estimated as a latent variable, as in a structural equation model, by leveraging the covariance matrix between the topics and covariates. This process produces more reliable standard errors in the analysis of the training set. For the test set, I use STM’s ‘average’ setting to estimate topic proportions. This uses the average proportions from the training set as the priors in the test set estimation and is an appropriate choice when the original estimation included the covariates of interest. See Egami et al. (2017) for more details.

in buildings with more than 20 units, and the proportion of units rented in buildings built after 2010. I fit an STM of 40 topics[6] to a duplicate-cleaned training set of 22,679 documents. I use the results of the topic model, a vector of 40 topic proportions for each document, to estimate the relationship of each neighborhood type and the listing text. Topic proportions range between 0 and 1, are non-zero, and the 40 vectors associated with a particular document sum to 1. I estimate log-level OLS regressions of each topic on the neighborhood typology alone (40 models) and on the neighborhood typology and other covariates (40 models). I then validate the models on a test set of the same size and report the results from the test set. These results are shown in Figures 2 and 3.

In order to understand the content of the topics found to be associated with neighborhood racial composition, I perform qualitative analysis and coding on a subset of representative documents. For each of the 40 topics, I select the ten documents which have the highest proportional match for that topic, resulting a set of 400 representative documents. Using Atlas.ti, I code each document for neighborhood and unit features. I pay special attention to features that appear salient across more and less desirable neighborhoods: distance and modes of transportation, neighborhood descriptions, locations described as being nearby the unit, features of the property or development, especially security features, and features of the unit.

Qualitative analysis of 400 documents identified key terms which emphasized the security of units in listings. These terms most often indicated the presence of a security system or security patrol. I analyzed the prevalence of listings which included such security discourse and used logistic regression to estimate the association between neighborhood type and security

[6] Selecting the proper number of topics, K. is an essential part of using Topic Models, including STM, for text analysis. I used STM's measures of model likelihood and semantic coherence to find a K that balanced each and produced topics that made substantive sense. I experimented with K as large as 60 and as small as 12, but found that at smaller Ks, interesting topics were subsumed under more general topics.

discourse. Since it was possible that any relationship in that model might be due to the increased existence of security systems in those neighborhoods, I also used tax-assessor data, which indicates if a building has a security system installed, to estimate the relationship between installed security systems and neighborhood type. These results are shown in Figure 4.

Results

Quantitative Results: Systemic Textual Differences

Without considering neighborhood differences, *Craigslist* rental listings tend to have a consistent format and to present similar types of information. Since the purpose of these texts is to attract renters, the texts need to include information that home seekers might use to select a place to live. Accordingly, almost all advertisements include details about the unit for rent, like the size, number of bedrooms, amenities, and monthly costs. Information about the surrounding area is common, but not universal. If a listing includes explicit information about the neighborhood, that information is generally focused on nearby places to shop, eat, or visit, and local transportation options. I consider it unlikely for a listing to mention an amenity, nearby attraction, or unit feature that is not present, as that would be misleading. However, the opposite case is more imaginable. Even if a particular unit or neighborhood feature, say a café or a security system, is present in reality, it may not show up in a text. Results from each step of my analysis focus on how these discursive patterns are different on average for advertisements from neighborhoods with differently racialized populations.

Logistic regression shows that there are systematic lexical differences by neighborhood race at the word level. Table 1 shows the words that were associated with the largest increases in the log-odds of their containing document being from a listing in a neighborhood with above-median White and Black proportion.

Table 1: Terms associated with high-White and high-Black neighborhoods

Terms Associated with high-White neighborhoods	Terms Associated with high-Black neighborhoods
‘whole foods’ ‘shops’ ‘laundry’ ‘beach’ ‘classic’ ‘restaurants’ ‘charming’ ‘deck’ ‘basement’ ‘bike’	‘diverse’ ‘light rail’ ‘station’ ‘airport’ ‘community college’ ‘concierge’ ‘gated’

Terms associated with high-White neighborhoods imply a welcoming, accessible neighborhood with nice places to shop, eat, and visit. 7.9% of listings in high-White neighborhoods used the word ‘charming’ while only 4.6% of other listings did, and high-White listings used the word ‘classic’ more than twice as often (3.5% to 1.6%). These listings emphasized proximity to expensive grocery stores like Whole Foods (1.6% to 1.2%) and nearby attractions like restaurants (26.4% to 20.7%) and beaches (7.0% to 3.4%).

The terms associated with high-Black neighborhoods, on the other hand, focus on transportation, presumably away from the neighborhood. Mentions of ‘light rail’ occurred in 6.3% of listings from high-Black neighborhoods, but only 2.0% of other listings. However, the light rail in Seattle passes through more traditionally Black neighborhoods which could account for some of that difference. To measure that impact, I examined the prevalence of the term ‘light rail’ only for listings geocoded to within one mile of the train’s route, leaving 5,691 listings from above-median Black tracts and 1,774 listings from other tracts. In that subset, 21.5% of listings from high-Black neighborhoods included ‘light rail’, while only 13.0% of listings from other neighborhoods did.

Security terms were also notably associated with more Black neighborhoods. Take the term ‘concierge,’ which listings use to describe a doorman or security guard, usually in a development or apartment building. It showed up in 539 texts—only 1.1% of all listings—but 456 of those mentions were in high-Black neighborhoods. Other security terms, like 'gated' (3.7% to 2.4%) and 'control' (6.6% to 5.4%) were also more common in high-Black neighborhoods. These themes were common in the topic modeling results as well and motivated further analysis below.

The STM produced 40 topics, defined by a high probability of containing certain groups of words, and a vector of topic proportions for each document. I label each topic by examining the words most associated with it and reading example texts, and refer to it by that label and, in parentheses, the number it was assigned in the STM. Some topics, like the one I called "Cozy and Comfortable" (Topic 16), focused on unit features. Others, like "High Class Surroundings" (Topic 8) were more explicitly concerned with the surrounding area. I regress the log of the topic proportions for each document on the neighborhood typology and other covariates to assess that topic's association with neighborhood racial composition.

Figure 1: Log-Level Coefficients of Topic Distribution regressed on Neighborhood type

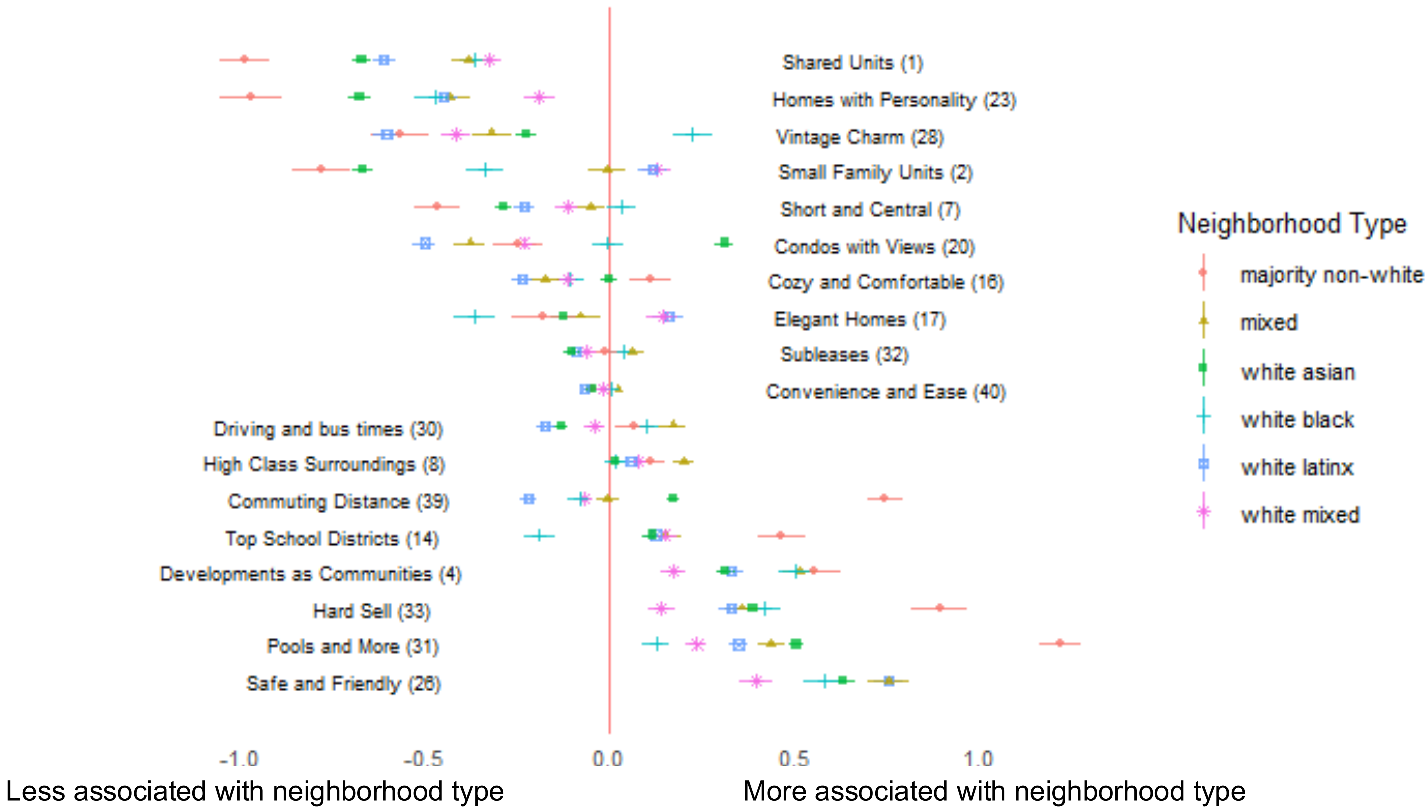
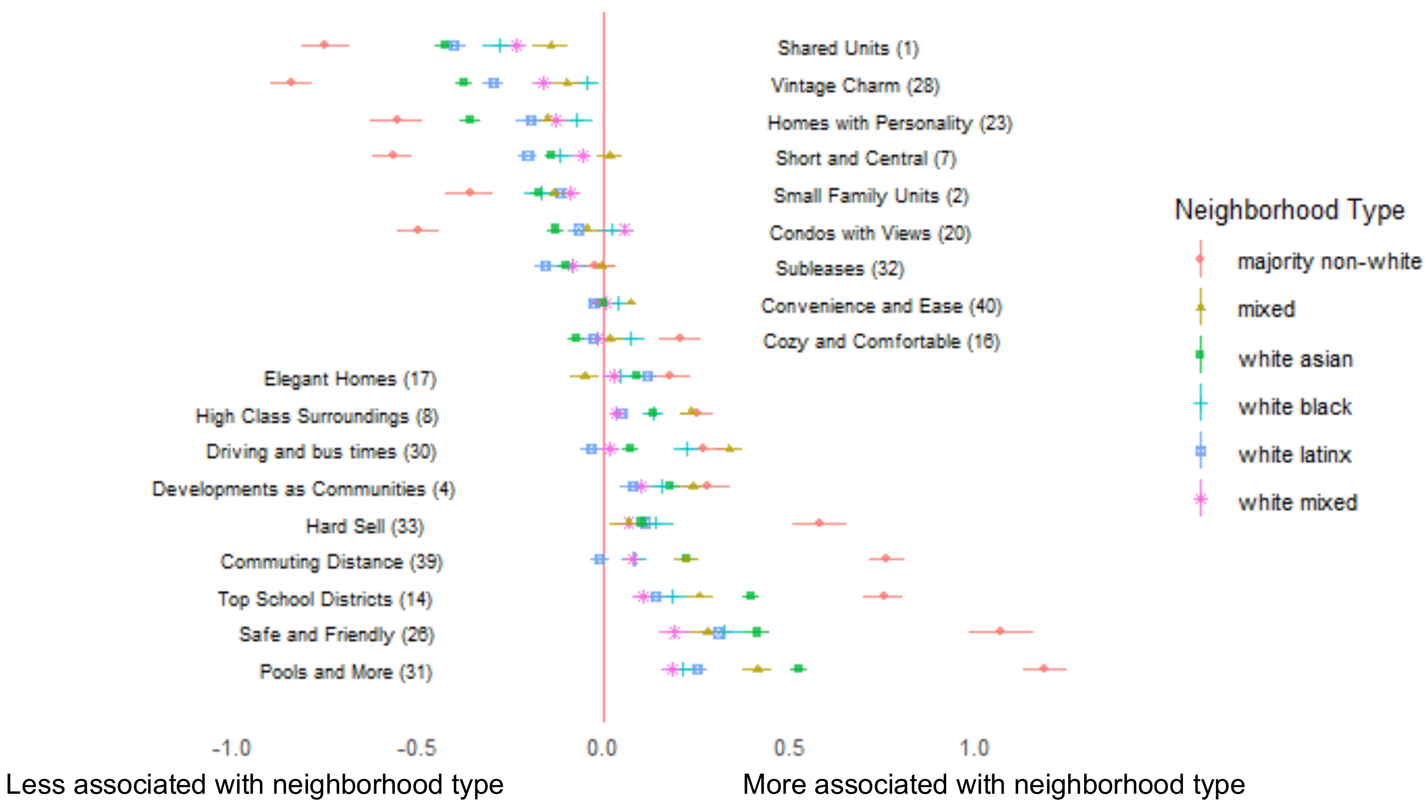


Figure 2: Log-Level Coefficients of Topic Distribution regressed on Neighborhood type and covariates



I report the results of log-level OLS regression of estimated topic proportions on a held-out test set of documents for two model types (see Cryer 2018 for another example of logging topic model output). Log-level coefficients of β on a neighborhood type dummy, let's say majority non-White, suggest that a switch from a listing in a predominantly White neighborhood to a majority non-White neighborhood is associated with an increase in the topic proportion of $\beta e^{\beta} \times 100\%$ (which can be approximated by $\beta \times 100\%$ for small values of β because in those cases e^{β} is approximately 1). For example, in the multi-variable regression of "Hard Sell" (Topic 33), a topic which centered on discounts, special deals, and other incentives for tenants to move in quickly, the coefficient for majority non-White neighborhoods is 0.58. That means that, on average, I expect that the average listing for a unit in a majority non-White neighborhood to include this topic $0.58e^{0.58} \times 100\%$ or 103% more than the average in a predominantly White neighborhood. In other words, the model suggests that the average proportion of "Hard Sell" (Topic 33) discourse is roughly twice as high in majority non-White neighborhoods than in predominantly White neighborhoods, ceteris paribus. I use this method to calculate the percentage increases reported below.

Figure 1 reports results from bivariate models which regressed each topic proportion on only the neighborhood racial typology, using 'Predominantly White' as the reference category. Figure 2 reports results from models which include covariates for the log of the rent and the square footage of the unit, and tract-level indicators for the neighborhood typology, poverty proportion, log of median income, population in thousands, proportion college educated, proportion commuting, proportion of owner-occupied units, proportion of units rented in buildings with more than 20 units, and the proportion of units rented in buildings built after

2010. In both figures, only a subset of topics of interest are shown.[7] In figures 1 and 2, appearing on the left side of the plot indicates that the relevant neighborhood type is associated with significantly less of that topic, appearing on the right indicates an association with significantly more of that topic. For instance, we saw above that "Hard Sell" (Topic 33) was more than twice as common in majority non-White neighborhoods than in predominantly White neighborhoods in the multi-covariate case. In contrast, "Convenience and Ease" (Topic 40), a topic which centered on easy to access storage and convenient parking and commuting, was not strongly associated with any neighborhood type in either bivariate or multi-variable cases.

The quantitative results identify the topics which vary with neighborhood racial proportion. Topics which have to do with trust (Topics 1 and 32), with personality and other less-quantifiable positive qualities (Topics 23, 28, and 16), and centrality (Topics 7 and 20) are associated more with predominantly White neighborhoods. Topics associated with travel (Topics 30 and 39), safety (Topic 26), and property, as opposed to unit, amenities (Topics 8, 4, and 31) are more associated with less-white neighborhoods.

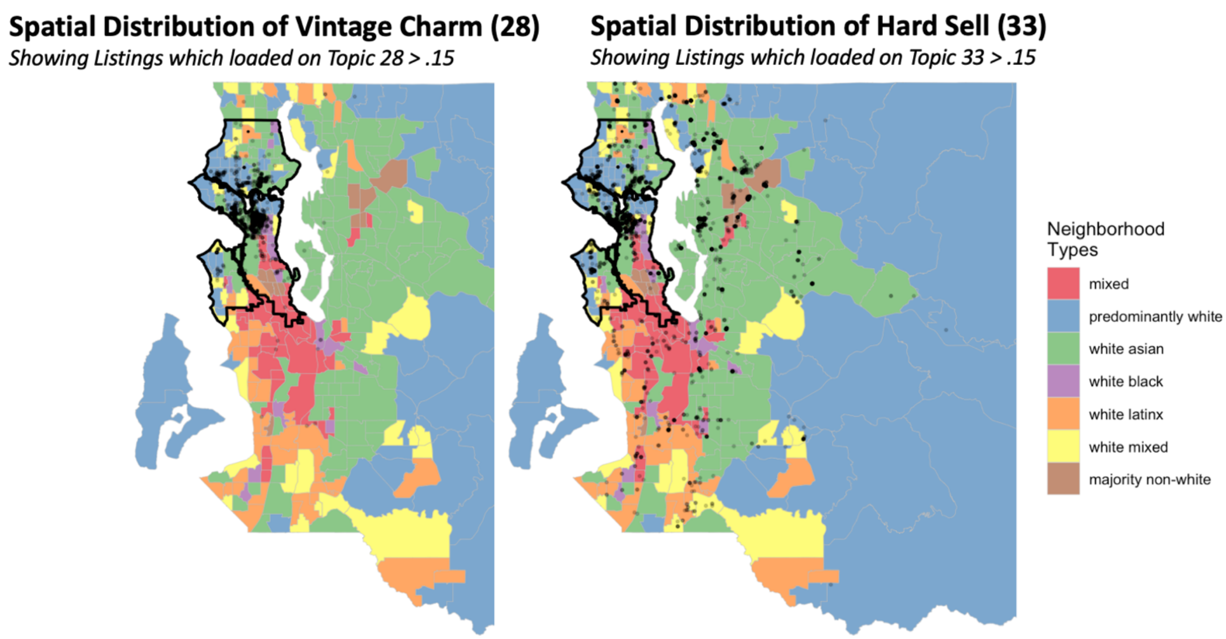
For example, I calculate the percent change in the two topics concerning trust. "Shared Units" (Topic 1) advertise units that are attached to the landlord's home or property, often called accessory units, and "Subleases" (Topic 32) are requests for new tenants to assume a lease or sublet for a short period of time. Both of these arrangements require high levels of trust between the two parties. The multi-variable model estimates that, compared to a listing from a predominantly White neighborhood, a listing from a less White tract contains 12.5% (for Mixed neighborhoods) to 35.4% (for majority non-White neighborhoods) less of "Shared Units" (Topic

[7] Full Regression and STM output and tables are available in Appendices I and II respectively. The subset of topics displayed was based on three criteria. First, limit the total number of topics to make the visualization relatively easy to read. Second, include all of the topics mentioned in the paper. And third, include at least some topics without strong associations with neighborhood racial proportion.

1) and 2.5% (for majority non-White neighborhoods) to 13.5% less (for White Latinx neighborhoods) of "Subleases" (Topic 32). This suggests that these high-trust arrangements may be most common in the Whitest neighborhoods. We can arrive at this same inference by examining figure 2 and noticing that Topics 1 and 32 are arranged on the left side of the plot, indicating their association with predominantly White neighborhoods.

The topics visualized here also have notable variation in their spatial distribution, with an example shown in Figure 3.

Figure 3:Spatial Distribution of Topics



We can see that "Vintage Charm" (Topic 28), a topic associated with predominantly White neighborhoods, is clustered in central Seattle (city border indicated by thick black line). Contrastingly, "Hard Sell" (Topic 33), includes a number of central listings, but has many more peripheral listings than "Vintage Charm" (Topic 28). This pattern—that topics associated with more White neighborhoods are also more central—occurs for the other topics as well. In part, this reflects the long-standing spatial and racial demographic order in Seattle. The oldest, most central, and most established neighborhoods have been Whiter because of explicitly racially motivated redlining and racial covenants enforced by real estate agents who could be expelled from the Real Estate Board for non-compliance (Rothstein 2017). These historical patterns have been exacerbated by changes in Seattle over the past decades, as non-White and poorer populations have been pushed out of more desirable central areas by rising rents and evictions (Thomas 2017). In other words, neighborhoods' racial composition and their peripheral status are intertwined. Also, note that all of the neighborhood types in this analysis occur both within Seattle and outside of it.

To fully understand the implications of this quantitative text analysis, it is essential to read example texts closely to understand both the nature of racialized discourse and to explore ways it could influence residential patterns.

Qualitative Results: Neighborhood Discourse and Habits of Whiteness

Quantitative analysis shows that differences in discourse are significant and spread across a variety of topics, reflecting racialized neighborhood descriptions which may be suggestive of racialized perceptions. In the analysis that follows, I use close readings of listing texts to investigate the content of that discourse.

Since the model controls for neighborhood income and poverty as well as unit price and square footage these results are not about the substantial economic differences between White and non-White neighborhoods. In fact, the topics "Elegant Homes" (Topic 17) and "High Class Surroundings" (Topic 8) were more associated with less White neighborhoods when I included economic controls. White Black neighborhoods were associated with an 25% decrease in prevalence of "Elegant Homes" (Topic 17) in the bivariate model but an *increase,* though not significant, of 4% with controls. Association of those neighborhoods with "High Class

Surroundings" (Topic 8) increased from an insignificant 2% increase to a significant 15% increase. That is, the quantitative analysis suggests that the important difference is not simply the amenities available in more and less White neighborhoods, but how those amenities and other property and neighborhood features are framed by text writers. Understanding that relationship requires qualitative analysis.

We begin with listings associated with "Vintage Charm" (Topic 28), mapped above. This example, from the high-White and high-profile neighborhood Capitol Hill, represents how listings from that topic describe the connection to their neighborhoods:

> The St. Florence is a vintage building with extremely modern conveniences! Located at the intersection of several main streets of Capitol Hill (East Denny, East Olive & Summit) there are endless options for entertainment.

This development emphasizes the care taken with the 'vintage' quality of the building, indicating a history worthy of preservation. People who live there then have the benefit of connecting to that neighborhood and building its community. It is clear that the 'endless options for entertainment' are also features of the neighborhood. By focusing on restoring this building, the development is marking Capitol Hill in both material and discursive terms as a place worth inhabiting and maintaining.

In contrast, consider this "Pools and More" (Topic 31) excerpt from another ad, this one in Bitter Lake, in an area with a relatively high Black population:

> Yet Another Veridian Cove Special! Spacious Kitchen and Bathroom. Includes all the amenities we are famous for. Life the way it should be! Relaxing after work in Club Veridian: 2 hot tubs, heated pool, Yoga/Pilates studio and Full Gym! Private Lake Access!

Or from this example from "Developments as Communities" (Topic 4): "Express yourself in a community of unique apartment homes nestled in between private courtyards and lush Northwest landscaping." The focus in these listings, and many others which matched highly in either topic, changes from the neighborhood to the property. There is still a community here, but it is on site. The pool and yoga studio take care of you while you are at home. You can relax in the private courtyards. The 'community' is identified with the unit type, apartment homes, rather than with actual people. Listings for developments in White Latinx and White Black neighborhoods include language about amenities and community, which could reflect conscious or unconscious attempts for POMs to assuage worries they expect prospective renters to have—again consciously or unconsciously—about neighborhood demographics.

We can also see this kind of discourse in another topic associated with less-White neighborhoods. "Hard Sell" (Topic 33) listings included special offers and discounts to attract tenants. Consider this listing:

> Sherwood Apartment Homes has a beautiful & cozy top floor one bedroom available now! Not only that, we are offering $300 off your move in costs! This home offers a open style kitchen, spacious floor plan and so much more! Don't miss out on this awesome deal!
>
> *pricing and availability subject to change*

Like listings that focused on the community inside a development, rather than around it, this approach attempts to coax recalcitrant renters to neighborhoods outside of their original search area. In a more direct way than emphasizing community, "Hard Sell" (Topic 33) listings imply a lower value for the surrounding neighborhood. Specifically, regardless of neighborhood, "Hard Sell" (Topic 33) listings show that a POM is willing to take less money than they were previously, which is a clear sign that the POM considers the unit to have lower value. The fact that such discourse was associated with less-White neighborhoods suggests that those neighborhoods are also associated with less value.

Taken together, this qualitative analysis shows that, while texts reflect the material reality of neighborhoods and POM's attempts to sell units effectively, these differences are also about the framing of unit surroundings. That is, any developer who has a pool to sell will sell it, but in more White neighborhoods, they might sell the pool and the neighborhood. In contrast, in less White neighborhoods, discourse about surrounding features declines and more focus is put on development amenities. That finding aligns with the quantitative association between topics like "Pools and More" (Topic 31) and "Developments as Communities" (Topic 4) as shown in figures 2 and 3. Not only is neighborhood context mentioned less in these areas, but the units are more often discounted, further eroding the perceived value of those places.

Security Discourse

Racial discourse in the United States, whether explicit or implicit, has often focused on security and danger. The n-gram regression found that security terms were more common in high-Black neighborhoods, and over 1,200 listings in the full dataset mentioned a 'Courtesy Patrol.' Many more mentioned a 'Night Patrol,' or a 'concierge.' In the STM, 'patrol' was a high-loading word for "Safe and Friendly" (Topic 26), which was associated with less White neighborhoods. Compared to predominantly White neighborhoods, the model expects between 35-45% more of "Safe and Friendly" (Topic 26) in Mixed, White Black, and White Latinx neighborhoods, 60% more in White Asian neighborhoods, and four times as much in majority non-White tracts. The website for one development explains, "[a] Courtesy patrol is on duty every night to ensure that you have a good night's rest." That is, this security language seems to directly address the White habit of anxiety.

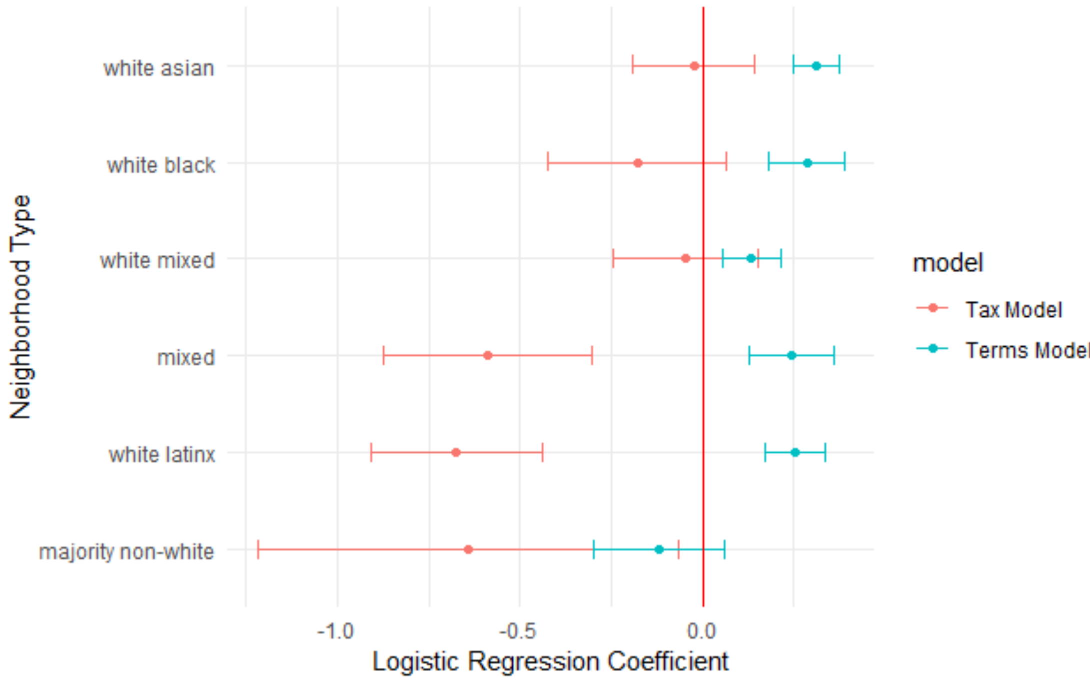
Like amenities, security was a common theme in topics associated with both high and low non-White proportions. Similarly, words like safety, security, secure, and controlled were present in listings from all neighborhood types. It was not only the incidence of security talk that distinguished Whiter listings, but the manner in which that talk was presented. Another text from "Vintage Charm" (Topic 28) demonstrates this:

> Quiet street in a safe, very walkable neighborhood, with many cafes, restaurants, and food-shopping options to choose from […] Very centrally located -- wherever you choose to go — Alki beach in West Seattle, hiking in Issaquah, the mall in Bellevue, etc, you're no more than a 20-min car ride away. The apartment is a rare find - an affordable, light-filled, spacious place on a quiet street in a great, safe, walkable and fun neighborhood. Come check it out!

Here safety is not only a feature of the neighborhood but is combined with other aspects of the neighborhood: walkability and fun. This is a marked contrast to the idea of a courtesy patrol in a development otherwise depicted as isolated from its surroundings.

These associations reflect a general trend in the listings independent of the prevalence of installed security systems. Nelson's (2017) computational grounded theory pushes scholars, after completing initial analysis and deep reading, to verify those findings with further quantitative analysis, often by investigating term frequency. In the case of safety, I examine how terms revealed in the qualitative analysis to be common indicators of security systems are associated with tract-level racial proportions. I also see how the same factors are associated with installed security systems reported in King County Tax assessor data. I estimate the association using logistic regression of reported security systems and incidence of security terms on neighborhood type and the tract-level covariates included in the topic model analysis. Figure 4 shows this analysis.

Figure 4: Results of Logistic Regression on Security Terms[8] *and Tax Data*
Error Bars show 95% Confidence Intervals



There is a significant association between the White-Black neighborhood type and higher incidence of safety terms, compared to the reference category of predominantly White neighborhoods. However, comparing this finding with the incidence of security systems tells a different story. Security discourse in the sample is slightly, but significantly, more common in less-White neighborhood types, except for majority non-White neighborhoods. Actual security systems are significantly *less* likely to occur in Mixed, White Latinx and majority non-White neighborhoods, and their odds of occurring are not statistically significantly different between White Asian, White Black, and White Mixed neighborhoods and the reference category of majority non-White.[9]

[8] Security terms are: secure, secured, security, control, controlled, patrol, gate, gated, protect, protection, protected, intercom, alarm

[9] Full regression output is available in Appendix I.

The difference between incidence of security discourse and actual security systems suggests that there is a racialized neighborhood perception in less-White neighborhoods that perspective renters will be more likely to want to read about security features. The fact that security systems tend to be less common there simply means that, in predominantly White neighborhoods, listing writers do not feel compelled to mention these security features even when they are present. That seems to reflect a racialized perception that the safety of White neighborhoods goes without saying.

Discussion

My mixed methods approach reveals significant and systemic differences in *Craigslist* listings by neighborhood type when examined at both the word and document level. This fails to support hypothesis 1, that variation in discourse is not associated with neighborhood race. Moreover, significant associations between differences in listing text discourse and neighborhood racial composition remain significant after addition of covariates. I therefore find no support for hypothesis 2 because variation in racialized neighborhood discourse are robust to economic factors, housing stock, and unit-level price and square footage. These findings, therefore, support hypothesis 3: *Craigslist* rental listings include lexical and topical differences associated with neighborhood race, and those differences persist even when including covariates for non-racial neighborhood characteristics.

The differences in discourse revealed by my analysis accord with observations in other changing urban areas. Trends in Seattle *Craigslist* listings align with Walton's (2018) finding that new White residents respond to less-White neighborhoods by adopting the White habits of anxiety and ambivalence. While language associated with predominantly White neighborhoods focuses on connections to neighborhood past and present, language associated with majority non-White neighborhoods seeks to assuage anxiety by emphasizing safety and to counter ambivalence by showing off property features that mirror the neighborhood features of White spaces. This discourse treats White neighborhoods as sufficient places for living, while framing non-White neighborhoods as places that need to be adorned and secured to be suitable for habitation.

While the cross-sectional nature of this study makes it difficult to assess the causes of these linguistic differences, considering them in the context of contemporary theoretical and empirical work on the search for housing and the reproduction of segregation is suggestive in at least four ways. First, for housing searchers with little local information and ample resources who initially include many areas in their search, these differences might lead them to exclude certain neighborhoods which seem to offer fewer amenities or connections to community. Second, these texts could contribute to building shared knowledge about neighborhoods, shared knowledge which could itself influence neighborhood selection in a housing search. Third, instead of (or in addition to) creating such shared knowledge of neighborhoods, the texts could represent or reflect existing shared knowledge. In that case, *Craigslist* listings act as a conduit for assimilating new or less knowledgeable rental housing seekers into the common knowledge of King County neighborhoods. Finally, these differences might reflect material differences in the neighborhoods, including the legacy of past and continued racial segregation. Far from mutually exclusive, these four ways the observed differences in rental texts could influence housing searches are synergistic. Taken together, they suggest further support for hypothesis 3. In addition to existing explicit discrimination and exclusionary tactics by landlords and real estate agents (Greif 2018; Korver-Glenn 2018), home seekers face a landscape of unequal neighborhood perceptions.

These texts indicate particular difficulties for vulnerable populations. One topic, "Tenant Restrictions" (Topic 19), was focused on restrictions on applications including credit checks, income requirements, background checks, required appointments, and details about housing rental laws. This topic was weakly, and often not significantly, associated with more White neighborhoods. It was, however, very significantly associated with unit price (the full regression table is included in Appendix II). This means that these restrictions fall more heavily on poorer people, regardless of the geography of their search. Since the model controls for the size of the apartments, this would make things particularly difficult for families, especially single-parent households with limited income. Adding interaction terms for racial proportions and poverty (not shown) reveals that "Tenant Restrictions" (Topic 19) is strongly and significantly associated with the interaction of White-Black neighborhood type and tract poverty.

It is harder to speak to gendered effects in the text. However, given that incomes tend to be lower for women than men, for mothers than non-mothers, and for Black and Latinx women than White women, I can imagine that the focus on tenant restrictions in "Tenant Restrictions" (Topic 19) might fall hardest on Black and Brown mothers.

Conclusion

Craigslist texts vary by neighborhood racial composition in a way that plausibly influences housing dynamics and the reproduction of segregation. This occurs through racialized public neighborhood perceptions which do not necessarily have racial content. These effects may extend beyond housing search. Moreover, these racialized perceptions do not seem limited to certain neighborhood types. Instead, White neighborhoods, majority non-White neighborhoods, and mixed-race neighborhoods are all subject to racialized perceptions. However, while perceptions of White neighborhoods tend to be varyingly positive, perceptions of less White neighborhoods eschew engagement with communities there, possibly weakening neighborhood collective efficacy. Therefore, housing policies which are not explicitly integrative will likely reproduce segregation even if they address traditional causes.

This work has implications for research into the structures of racism by supporting those views which hold that racialized perceptions are not limited to individuals' minds, or to boundary work in a localized sphere. Moreover, this is importantly different from implicit racism: The perceptions examined in this study may be conscious (or unconscious) but are not easily recognizable as racialized at the individual level. Instead, the racialized aspects of neighborhood discourse are most visible as aggregate associations, but they may still be impactful. Racialized neighborhood discourse combines distorted perceptions of neighborhoods, like the emphasis on security, with accurate perceptions of neighborhood difference caused by historical discrimination, like the peripheral location of less-White neighborhoods. When those combined perceptions are reproduced in discourse like *Craigslist* listing text, they show up as natural and objective. The discourse then perpetuates the association as normative: making it seem that not only are less-White neighborhoods more dangerous and farther from the city center, but that such a discrepancy is natural, not worth understanding. By tracing the association between discourse and neighborhood race, I do not yet pull back the curtain to reveal the causes of that association, but I do show that there is such a curtain. Further work with this perspective will begin to show what processes do the social work of making racialized neighborhood discourse and perception.

There are significant limitations to this study in its current form. While the validity of these conclusions within the greater Seattle area during the study period are strong, the conclusions are unlikely to apply without significant modification beyond that spatial and temporal scope. Moreover, these conclusions are based on data from *Craigslist*, which could introduce a number of issues. Most immediately, the connection between listing text and public neighborhood perception, while always present, is not always direct. It is possible, moreover, that something about the *Craigslist* platform biases text writers or home searchers in a way that undermines the strength of these results. I have selected covariates with an eye towards engaging with spatial assimilation theories of the reproduction of segregation that focus on group differences in aggregate human capital to explain differences in residential attainment. I therefore include covariates for class, income, and education in order to account for variations across neighborhoods that might be conflated with racial composition. However, I do not include spatial smoothing or other neighborhood level covariates, notably crime rates. I justify this choice by arguing that associations between neighborhood type and crime rates are both a product of historical discrimination.

There are some larger questions that go unaddressed in this work. Most pressing, from my perspective, concerns the racial and ethnic makeup of people who use *Craigslist* in Seattle. I would like to gather more information on the demographics of both users and text writers. That data would allow me to study the possibility that racialized neighborhood discourse and perception operates differently depending on individual racialization and class position. Additionally, the results are currently limited to Seattle. On one hand, this makes it easier to leverage local knowledge about neighborhoods and therefore deepens the qualitative analysis. At the same time, focusing on a single city severely limits the generalizability of the findings. However, even with these drawbacks, I find these initial results promising (though not heartening). Continued research with these methods can improve our understanding about how these public texts shape perceptions of neighborhoods.

The current analysis demonstrates that, though devoid of explicit racial phrases or words, *Craigslist* rental texts are produced and perceived in racial contexts that influence their content.

These methods can easily be applied to a sample including multiple metropolitan areas. I am part of a team that is already collecting similar data from the 100 largest metropolitan areas in the United States by population. Such an expansion would improve both external validity and generalizability, though care would be necessary to adequately account for the specificity of each locale. As the sample expands not only through space, but through time, we will also be able to test possible time-ordered associations that could show what kinds of neighborhood discourse precede, if not cause, neighborhood changes, as well as reflect those neighborhood changes.

Apart from more data, additional methods might improve our understanding of the relationship between neighborhood perception and residential attainment. Factorial design studies are excellent at supporting clear causal claims, but their generalizability is often weak due to unrealistic or far-fetched vignettes. A factorial design in this case might be stronger because the text examples could be based on real advertisements and their difference carefully measured using the original topic model. This study would be able to investigate whether the subtle changes in discourse outlined in this paper would cause people to change the kind of neighborhood they wanted to move into.

However, even the modest results reported here are not well accounted for in the existing residential attainment literature. Instead, new approaches towards segregation and a cultural sociology lens offer more insights. These results also align with Zuberi and Bonilla-Silva's (2008) call for new methodology for the sociological study of race. I join a burgeoning community of sociologists (Bonilla-Silva 2018, Golash-Boza 2016; Ray et al. 2017; Zuberi 2011) arguing to incorporate a strengthening thread of theoretical work from feminist Black Studies exemplified by the canonization of works by Audre Lorde (1984), Hortense Spillers (1987), Sylvia Wynter (2003), and Saidiya Hartman (1997). From this perspective, studying language used in *Craigslist* listings provides a window into the racialization process that translates into segregative mobility patterns.

Lay accounts of racial discrimination equate it to prejudice in the minds of individuals. Reskin (2012) complicates this paradigm by writing, "because über discrimination operates partly through distorting our thought processes (Greenwald and Banaji 1995), a third strategy for limiting discrimination involves implementing decision-making practices that minimize these distortions" (p. 31). *Craigslist* rental listings seem to be one mechanism that causes such distortion, and therefore, with further study, could provide insight into how to limit this operation of über discrimination.

In this account, differential outcomes by race are partially the result of apparently innocuous action by people who, regardless of their intentions, have their perceptions distorted by über discrimination. This view has remarkable overlap with contemporary Black feminist theory that pushes for a focus on processes of racialization which can drive apparently non-racial action and discourse (Wynter 2003; Hartman 1997; Spillers 1987; Weheliye 2014). Sociologists studying race and space should take up that focus and use it to improve our understanding of the pervasiveness and impact of those processes of racialization. If we can trace the processes that reproduce contemporary segregation, we may be able to learn how to slow down, halt, or reverse them.

References

Bader, Michael D. M. and Maria Krysan. 2015. "Community Attraction and Avoidance in Chicago: What's Race Got to Do with It?" *Annals of the American Academy of Political and Social Science.* 660(1):261–81.

Besbris, Max. 2016. "Romancing the home: emotions and the interactional creation of demand in the housing market." *Socio-Economic Review*, 14(3), 461-482.

Besbris, Max, Jacob William Faber, Peter Rich, and Patrick Sharkey. 2015. "Effect of neighborhood stigma on economic transactions." *Proceedings of the National Academy of Sciences*, 112.16: 4994-4998.

Bonam, Courtney M., Hilary B. Bergsieker, and Jennifer L. Eberhardt. 2016. "Polluting Black Space." *Journal of Experimental Psychology: General.* 145(11):1561–82.

Bonam, Courtney M., Caitlyn Yantis, and Valerie J. Taylor. 2018. "Invisible middle-class Black space: Asymmetrical person and space stereotyping at the race–class nexus." *Group Processes & Intergroup Relations*. 1-24.

Bonam, Courtney M., Valerie J. Taylor, and Caitlyn Yantis. 2017. "Racialized physical space as cultural product." *Social and Personality Psychology Compass.* 11.9.

Bonilla-Silva, Eduardo. 2014. *Racism without racists: Color-blind racism and the persistence of racial inequality in America*. Rowman & Littlefield.

Bonilla-Silva, Eduardo. 2018. "Feeling Race: Theorizing the Racial Economy of Emotions." *American Sociological Review*. Vol. 84(1) 1–25

Bracey, Glenn E. 2015. "Toward a Critical Race Theory of State." *Critical Sociology.* 41(3):553–72.

Casas, Andreu, Tianyi Bi, and John Wilkerson. 2018. "A Robust Latent Dirichlet Allocation Approach for the Study of Political Text." presented at *Conference of the American Political Science Association*. Boston.

Chasins, Sarah and Rastislav Bodik. 2017. "Skip Blocks: Reusing Execution History to Accelerate Web Scripts." *Proceedings of the ACM on Programming Languages.* 1(OOPSLA):1–28.

Crowder, Kyle and Maria Krysan. 2016. "Moving Beyond the Big Three: A Call for New Approaches to Studying Racial Residential Segregation." *City and Community.* 15(1):18–22.

Crowder, Kyle, Jeremy Pais, and Scott J. South. 2012. "Neighborhood diversity, metropolitan constraints, and household migration." *American Sociological Review*. 77(3), 325-353.

Crowell, Amber R., and Mark Fossett. 2017. "White and Latino Locational Attainments: Assessing the Role of Race and Resources in US Metropolitan Residential Segregation." Sociology of Race and Ethnicity, 2332649217748426.

Cryer, Jennifer. 2018. "Navigating Identity in Campaign Messaging: The Influence of Race & Gender on Strategy in U.S. Congressional Elections." Available at *SSRN*: 2863215.

Dimaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture : Application to Newspaper Coverage of U . S . Government Arts Funding." *Poetics.* 41(6):570–606.

Egami, Naoki, Christian J. Fong, Justin Grimmer, Margaret E. Roberts, and Brandon M. Stewart. 2017. "How to Make Causal Inferences Using Texts." arXiv:1802.02163.

Emirbayer, Mustafa, and Matthew Desmond. 2015. *The racial order*. Chicago: University of Chicago Press.

Galiani, Sebastian, Alvin Murphy, and Juan Pantano. 2015. “Estimating Neighborhood Choice Models: Lessons from a Housing Assistance Experiment.” *American Economic Review* 105(11):3385–3415.

Gibbons, Joseph, and Tse-Chuan Yang. 2014. "Self-rated health and residential segregation: how does race/ethnicity matter?" *Journal of Urban Health* .91.4: 648-660.

Golash-Boza, Tanya. 2016. “A Critical and Comprehensive Sociological Theory of Race and Racism.” *Sociology of Race and Ethnicity.* 2(2):129–41.

Greenwald, Anthony G., and Mahzarin R. Banaji. 1995. “Implicit social cognition: attitudes, self-esteem, and stereotypes.” *Psychological review*. 102(1), 4.

Greif, Meredith. 2018. “Regulating landlords: Unintended consequences for poor tenants.” *City & Community*, 17(3), 658-674.

Haney López, Ian. *Dog whistle politics: How coded racial appeals have reinvented racism and wrecked the middle class.* Oxford University Press, 2015.

Hartman, Saidiya V. 1997. *Scenes of subjection: Terror, slavery, and self-making in nineteenth-century America.* Oxford University Press.

Hall, Matthew, Kyle Crowder, and Amy Spring. 2015. "Neighborhood foreclosures, racial/ethnic transitions, and residential segregation." *American Sociological Review*. 80.3: 526-549.

Intrator, Jake, Jonathan Tannen, and Douglas S. Massey. 2016. “Segregation by Race and Income in the United States 1970–2010.” *Social Science Research.* 60:45–60.

Inwood, Joshua F., and Robert Yarbrough. 2010. “Racialized places, racialized bodies: The impact of racialization on individual and place identities.” *GeoJournal*. 75(3), 299-301.

Jencks, Christopher, and Susan E. Mayer. 1990. "The social consequences of growing up in a poor neighborhood." in *Inner-city poverty in the United States.* Robert Halpern eds. 111-186. Washington, DC: National Academies Press.

Krysan, Maria and Kyle Crowder. 2017. *Cycle of Segregation.* New York: Russell Sage Foundation.

Korver-Glenn, Elizabeth. 2018. “Compounding inequalities: How racial stereotypes and discrimination accumulate across the stages of housing exchange.” *American Sociological Review*, 83(4), 627-656.

Logan, John R. 2017. “Racial segregation in postbellum Southern cities: The case of Washington, DC.” *Demographic research*, 36, 1759.

Lorde, Audre. 1984. *Sister Outsider*. New York: Random House.

Light R, and Odden C. 2017. “Managing the Boundaries of Taste: Culture, Valuation, and Computational Social Science.” *Social Forces*. 96(2):877–908

Massey, Douglas S., and Nancy A. Denton. 1985. “Spatial assimilation as a socioeconomic outcome.” *American sociological review*. 94-106.

Massey, Douglas S., and Nancy A. Denton. 1993. *American apartheid: Segregation and the making of the underclass*. Harvard University Press.

Massey, Douglas S. and Jonathan Tannen. 2015. “A Research Note on Trends in Black Hypersegregation.” *Demography*. 52(3):1025–34.

Massey, Douglas S., and Jacob S. Rugh. 2017. "Zoning, Affordable Housing, and Segregation in US Metropolitan Areas." in *The Fight for Fair Housing: Causes, Consequences, and Future Implications of the 1968 Federal Fair Housing Act*. Gregory D. Squires ed. London: Routledge.

Nelson, L. K. 2017. “Computational Grounded Theory: A Methodological Framework.” *Sociological Methods & Research.*

Pais, Jeremy, Scott J. South, and Kyle Crowder. 2012. “Metropolitan heterogeneity and minority neighborhood attainment: Spatial assimilation or place stratification?” *Social problems.* 59(2), 258-281.

Pager, Devah, and Hana Shepherd. 2008. "The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets." *Annual Review of Sociology*. 34: 181-209.

Quillian, Lincoln. 2015. “A Comparison of Traditional and Discrete-Choice Approaches to the Analysis of Residential Mobility and Locational Attainment.” *Annals of the American Academy of Political and Social Science*. 660(1):240–60.

Quillian, Lincoln and Devah Pager. 2001. “Black Neighbors , Higher Crime ? The Role of Racial Stereotypes in Evaluations of Neighborhood Crime.” *American Journal of Sociology*. 107(3):717–67.

Ray, Victor Erik, Antonia Randolph, Megan Underhill, and David Luke. 2017. “Critical Race Theory, Afro-Pessimism, and Racial Progress Narratives.” *Sociology of Race and Ethnicity*. 3(2):147–58.

Reskin, Barbara. 2012. “The Race Discrimination System.” *Annual Review of Sociology*. 38(1):17–35.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2014. "stm: R package for structural topic models." *R package* 1: 12.

Rosen, Eva. 2014. "Rigging the Rules of the Game: How Landlords Geographically Sort Low-Income Renters." *City & Community*. 13.4: 310-340.

Rothstein, Richard. 2017. *The color of law: A forgotten history of how our government segregated America.* Liveright Publishing.

Schofield, Alexandra, Laure Thompson, and David Mimno. 2017. “Quantifying the Effects of Text Duplication on Semantic Models.” *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing.* 2727–37.

Spillers, Hortense J. 1987. “Mama’s Baby, Papa’s Maybe: An American Grammar Book.” *Diacritics.* 17(2):64–81.

Tvinnereim E, Liu X, and Jamelske, EM. 2017. “Public perceptions of air pollution and climate change: different manifestations, similar causes, and concerns.” *Climatic Change*. 140(3–4):399–412.

Thomas, Timothy Andrew. 2017. *Forced Out: Race, Market, and Neighborhood Dynamics of Evictions.* Diss.

Walton, Emily. 2018. “Habits of Whiteness: How Racial Domination Persists in Multiethnic Neighborhoods.” *Sociology of Race and Ethnicity*: 1-15.

Weheliye, Alexander G. 2014. *Habeas viscus: Racializing assemblages, biopolitics, and black feminist theories of the human.* Duke University Press.

Williams, David R., and Chiquita Collins. 2001. "Racial residential segregation: a fundamental cause of racial disparities in health." *Public health reports.* 116.5: 404.

Wynter, Sylvia. 2003. “Unsettling the coloniality of being/power/truth/freedom: Towards the human, after man, its overrepresentation—An argument.” *CR: The new centennial review,* 3(3), 257-337.

Zuberi, Tukufu. 2011. “Critical Race Theory of Society.” *CONNECTICUT LAW REVIEW* 43(5):1573–91.

Zuberi, Tukufu and Eduardo Bonilla-Silva, eds. 2008. *White Logic, White Methods : Racism and Methodology*. Rowman & Littlefield Publishers.

Seattle Rental Ad Texts and Processes of Segregation Appendices

Appendix I: Regression Specification and Output

This paper includes the results from 2+80+2 or 84 regression models. In this appendix I provide the model specifications for those models in general terms and the regression output for 82 of the models. The two L2 regularized logistic regressions on n-grams include covariates for each word in the corpus and so are not easily summarized.

N-gram L2 Regression

$$logit(HighWhite_i) = \beta X_i$$

For i = 1, …, n

Where $HighWhite_i$ is the probability of the neighborhood containing the listing *i* has a proportion White above the median for all tracts, β is a vector of coefficients estimated using L2 regularization, and X_i is a vector of dummies indicating if listing *i* contains each word in the corpus, and N is the number of listings.

Similarly for high-Black neighborhoods:

$$logit(HighBlack_i) = \beta X_i$$

For i = 1, …, N

Where $HighBlack_i$ is the probability of the neighborhood containing the listing *i* has a proportion Black above the median for all tracts, β is a vector of coefficients estimated using L2 regularization, and X_i is a vector of dummies indicating if listing *i* contains each word in the corpus, and N is the number of listings (45,358).

Topic Log-Linear OLS Regression: Bivariate

$$\log(Topic_{ik}) = \beta_0 + \gamma NT_i + \epsilon_i$$

$$\epsilon_i \sim N\left(0, \sigma^2_{logk}\right)$$

For i = 1, …, N

For k = 1, …, K

Where $Topic_{ik}$ is the proportion of document *i* for topic *k* (reflecting K separate regressions), β_0 is the linear intercept, γ is a vector of neighborhood type coefficients, and NT_i is a vector of dummies for each neighborhood type. σ^2_{logk} is the varience of the logged vector of topic proportions for topic k. N is the number of documents (22,679) and K is the number of topics (40).

Topic Log-Linear OLS Regression: Multi-variable

$$\log(Topic_{ik}) = \beta_0 + \gamma NT_i + \beta X_i + \epsilon_i$$

$$\epsilon_i \sim N\left(0, \sigma^2_{logk}\right)$$

For i = 1, …, N

For k = 1, …, K

Where $Topic_{ik}$ is the proportion of document *i* for topic *k* (reflecting K separate regressions), β_0 is the linear intercept, γ is a vector of neighborhood type coefficients, NT_i is a vector of dummies for each neighborhood type, β is a vector of coefficients for other covariates, and X_i is a vector of covariates. σ^2_{logk} is the varience of the logged vector of topic proportions for topic k. N is the number of documents (22,679) and K is the number of topics (40).

Security Terms Logistic Regression

$$logit(SecTerm_i) = \gamma NT_i + \ \beta X_i$$

For i = 1, …, N

Where $SecTerm_i$ is the probability that text *i* includes a security term, γ is a vector of neighborhood type coefficients, NT_i is a vector of dummies for each neighborhood type, β is a vector of coefficients for other covariates, and X_i is a vector of covariates. N is the number of listings (45,358).

Installed Security Systems Logistic Regression

$$logit(SecSystem_i) = \gamma NT_i + \ \beta X_i$$

For i = 1, …, n

Where $SecSystem_i$ is the probability that entry *i* in King County tax assessor data indicates an installed security system, γ is a vector of neighborhood type coefficients, NT_i is a vector of dummies for each neighborhood type, β is a vector of coefficients for other covariates, and X_i is a vector of covariates. N is the number of entries (7,609).

Model Output:

```
 [1] "Topic1"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.5002  -0.8959  -0.0576   0.8132   4.2778

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                4.551670   0.570029   7.985 1.47e-15 ***
N_type_mixed              -0.144238   0.045767  -3.152  0.00163 **
N_type_white asian        -0.429918   0.025473 -16.877  < 2e-16 ***
N_type_white black        -0.282785   0.042797  -6.608 3.99e-11 ***
N_type_white latinx       -0.404873   0.033441 -12.107  < 2e-16 ***
N_type_white mixed        -0.239677   0.032233  -7.436 1.08e-13 ***
N_type_majority non-white -0.751680   0.065476 -11.480  < 2e-16 ***
pov_proportion             2.615928   0.199406  13.119  < 2e-16 ***
log_income                -0.420977   0.052952  -7.950 1.95e-15 ***
pop_thousands             -0.066309   0.005465 -12.133  < 2e-16 ***
share_college              5.164207   0.168752  30.602  < 2e-16 ***
share_commuters           -1.544356   0.172739  -8.940  < 2e-16 ***
share_oo                   0.737214   0.095064   7.755 9.21e-15 ***
share_rental_over_20      -1.288494   0.092028 -14.001  < 2e-16 ***
share_built_after_10      -0.407556   0.259815  -1.569  0.11675
log_price                 -0.739736   0.040459 -18.284  < 2e-16 ***
log_sqft                   0.062503   0.029548   2.115  0.03441 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.687751)

    Null deviance: 44034  on 22678  degrees of freedom
Residual deviance: 38248  on 22662  degrees of freedom
AIC: 76249

Number of Fisher Scoring iterations: 2
```

```
[1] "Topic2"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.1156  -0.8473  -0.0071   0.8849   5.5938

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               1.502209   0.563520   2.666 0.007687 **
N_type_mixed             -0.135980   0.045244  -3.005 0.002655 **
N_type_white asian       -0.174517   0.025182  -6.930 4.31e-12 ***
N_type_white black       -0.170750   0.042308  -4.036 5.46e-05 ***
N_type_white latinx      -0.118078   0.033059  -3.572 0.000355 ***
N_type_white mixed       -0.093999   0.031865  -2.950 0.003182 **
N_type_majority non-white -0.361997   0.064729  -5.593 2.26e-08 ***
pov_proportion            0.206526   0.197129   1.048 0.294803
log_income               -0.769225   0.052347 -14.695  < 2e-16 ***
pop_thousands            -0.031633   0.005403  -5.855 4.84e-09 ***
share_college             0.215006   0.166825   1.289 0.197476
share_commuters          -0.215869   0.170766  -1.264 0.206201
share_oo                  1.182680   0.093978  12.585  < 2e-16 ***
share_rental_over_20     -2.024913   0.090977 -22.257  < 2e-16 ***
share_built_after_10      0.569283   0.256848   2.216 0.026673 *
log_price                -1.068683   0.039997 -26.719  < 2e-16 ***
log_sqft                  1.563232   0.029210  53.516  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.649427)

    Null deviance: 62960  on 22678  degrees of freedom
Residual deviance: 37379  on 22662  degrees of freedom
AIC: 75728

Number of Fisher Scoring iterations: 2

[1] "Topic3"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
```

```
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3779  -0.6037  -0.2100   0.3156   5.0117

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -3.3491025  0.4350778  -7.698 1.44e-14 ***
N_type_mixed             0.0195253  0.0349319   0.559  0.57620
N_type_white asian      -0.0449462  0.0194423  -2.312  0.02080 *
N_type_white black       0.0263101  0.0326650   0.805  0.42057
N_type_white latinx      0.0255662  0.0255242   1.002  0.31653
N_type_white mixed       0.0604200  0.0246021   2.456  0.01406 *
N_type_majority non-white -0.3026247  0.0499752  -6.056 1.42e-09 ***
pov_proportion           0.1367836  0.1521978   0.899  0.36881
log_income              -0.0940495  0.0404156  -2.327  0.01997 *
pop_thousands            0.0007141  0.0041714   0.171  0.86408
share_college            0.3560063  0.1288005   2.764  0.00571 **
share_commuters         -0.1324851  0.1318440  -1.005  0.31497
share_oo                -0.1022519  0.0725578  -1.409  0.15878
share_rental_over_20     0.0029166  0.0702406   0.042  0.96688
share_built_after_10     0.1478246  0.1983050   0.745  0.45601
log_price                0.2069877  0.0308802   6.703 2.09e-11 ***
log_sqft                -0.3384045  0.0225524 -15.005  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.9832144)

    Null deviance: 23065  on 22678  degrees of freedom
Residual deviance: 22282  on 22662  degrees of freedom
AIC: 63995

Number of Fisher Scoring iterations: 2

[1] "Topic4"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.9319  -0.8024  -0.1322   0.7139   4.5829

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               3.781398   0.517654   7.305 2.87e-13 ***
N_type_mixed              0.237884   0.041562   5.724 1.06e-08 ***
N_type_white asian        0.179651   0.023132   7.766 8.43e-15 ***
N_type_white black        0.157357   0.038865   4.049 5.16e-05 ***
N_type_white latinx       0.075864   0.030369   2.498 0.012493 *
N_type_white mixed        0.097132   0.029271   3.318 0.000907 ***
N_type_majority non-white  0.278326   0.059460   4.681 2.87e-06 ***
pov_proportion           -0.102867   0.181084  -0.568 0.569998
log_income                0.197048   0.048086   4.098 4.19e-05 ***
pop_thousands             0.030230   0.004963   6.091 1.14e-09 ***
share_college            -0.572324   0.153246  -3.735 0.000188 ***
share_commuters           0.947764   0.156867   6.042 1.55e-09 ***
share_oo                 -1.099284   0.086329 -12.734  < 2e-16 ***
share_rental_over_20      0.461084   0.083572   5.517 3.48e-08 ***
share_built_after_10     -0.394749   0.235943  -1.673 0.094327 .
log_price                -0.419894   0.036741 -11.428  < 2e-16 ***
log_sqft                 -1.109020   0.026833 -41.331  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.391853)

    Null deviance: 46662  on 22678  degrees of freedom
Residual deviance: 31542  on 22662  degrees of freedom
AIC: 71878

Number of Fisher Scoring iterations: 2

[1] "Topic5"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-6.5417  -1.0173  -0.1152   0.8980   5.7755
```

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)              -12.897232   0.659586 -19.554  < 2e-16 ***
N_type_mixed               0.094216   0.052957   1.779  0.07524 .
N_type_white asian         0.052814   0.029475   1.792  0.07317 .
N_type_white black         0.272005   0.049521   5.493 4.00e-08 ***
N_type_white latinx        0.076124   0.038695   1.967  0.04916 *
N_type_white mixed         0.117974   0.037297   3.163  0.00156 **
N_type_majority non-white  0.206281   0.075763   2.723  0.00648 **
pov_proportion            -0.490987   0.230735  -2.128  0.03335 *
log_income                 0.562458   0.061271   9.180  < 2e-16 ***
pop_thousands              0.036796   0.006324   5.819 6.01e-09 ***
share_college             -1.461388   0.195264  -7.484 7.46e-14 ***
share_commuters            0.432358   0.199878   2.163  0.03054 *
share_oo                  -0.610833   0.109999  -5.553 2.84e-08 ***
share_rental_over_20       2.073902   0.106486  19.476  < 2e-16 ***
share_built_after_10       0.966222   0.300634   3.214  0.00131 **
log_price                  1.694749   0.046815  36.201  < 2e-16 ***
log_sqft                  -1.758340   0.034190 -51.429  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 2.259731)

    Null deviance: 74445  on 22678  degrees of freedom
Residual deviance: 51210  on 22662  degrees of freedom
AIC: 82868

Number of Fisher Scoring iterations: 2

[1] "Topic6"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.64271  -0.44558  -0.01226   0.44250   2.21852

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)               -5.025680   0.272480 -18.444  < 2e-16 ***
N_type_mixed               0.001166   0.021877   0.053  0.95748
```

```
N_type_white asian        -0.018773   0.012176  -1.542  0.12314
N_type_white black        -0.014572   0.020457  -0.712  0.47627
N_type_white latinx       -0.024352   0.015985  -1.523  0.12767
N_type_white mixed        -0.019082   0.015408  -1.238  0.21556
N_type_majority non-white -0.034101   0.031298  -1.090  0.27593
pov_proportion             0.278206   0.095318   2.919  0.00352 **
log_income                -0.011204   0.025311  -0.443  0.65802
pop_thousands             -0.007015   0.002612  -2.685  0.00725 **
share_college              0.721591   0.080665   8.946  < 2e-16 ***
share_commuters            0.062129   0.082571   0.752  0.45180
share_oo                   0.051092   0.045441   1.124  0.26088
share_rental_over_20      -0.070194   0.043990  -1.596  0.11058
share_built_after_10      -0.242791   0.124194  -1.955  0.05060 .
log_price                 -0.113235   0.019340  -5.855 4.83e-09 ***
log_sqft                   0.425382   0.014124  30.117  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.3856419)

    Null deviance: 9635.5  on 22678  degrees of freedom
Residual deviance: 8739.4  on 22662  degrees of freedom
AIC: 42770

Number of Fisher Scoring iterations: 2

[1] "Topic7"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0251  -0.7181  -0.0478   0.7134   3.6901

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               10.362399   0.448546  23.102  < 2e-16 ***
N_type_mixed               0.014735   0.036013   0.409 0.682432
N_type_white asian        -0.141439   0.020044  -7.056 1.76e-12 ***
N_type_white black        -0.121797   0.033676  -3.617 0.000299 ***
N_type_white latinx       -0.206708   0.026314  -7.855 4.16e-15 ***
N_type_white mixed        -0.059490   0.025364  -2.345 0.019011 *
```

```
N_type_majority non-white -0.571129   0.051522 -11.085  < 2e-16 ***
pov_proportion            1.332070   0.156909   8.489  < 2e-16 ***
log_income               -0.279964   0.041667  -6.719 1.87e-11 ***
pop_thousands            -0.034578   0.004301  -8.040 9.39e-16 ***
share_college             4.740558   0.132788  35.700  < 2e-16 ***
share_commuters          -0.495816   0.135925  -3.648 0.000265 ***
share_oo                 -0.926892   0.074804 -12.391  < 2e-16 ***
share_rental_over_20     -1.055638   0.072415 -14.578  < 2e-16 ***
share_built_after_10     -1.154831   0.204444  -5.649 1.64e-08 ***
log_price                -1.024560   0.031836 -32.182  < 2e-16 ***
log_sqft                 -0.476552   0.023251 -20.496  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.045027)

    Null deviance: 35880  on 22678  degrees of freedom
Residual deviance: 23682  on 22662  degrees of freedom
AIC: 65378

Number of Fisher Scoring iterations: 2

[1] "Topic8"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2966  -0.5157  -0.1472   0.2496   5.5202

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              -6.440726   0.377195 -17.075  < 2e-16 ***
N_type_mixed              0.234536   0.030285   7.744 1.00e-14 ***
N_type_white asian        0.133396   0.016856   7.914 2.61e-15 ***
N_type_white black        0.133525   0.028319   4.715 2.43e-06 ***
N_type_white latinx       0.045350   0.022128   2.049   0.0404 *
N_type_white mixed        0.034292   0.021329   1.608   0.1079
N_type_majority non-white  0.250257   0.043326   5.776 7.75e-09 ***
pov_proportion           -0.490883   0.131949  -3.720   0.0002 ***
log_income               -0.137455   0.035039  -3.923 8.77e-05 ***
pop_thousands            -0.004540   0.003616  -1.255   0.2094
```

```
share_college            -0.611485   0.111665  -5.476 4.39e-08 ***
share_commuters           0.586739   0.114303   5.133 2.87e-07 ***
share_oo                  0.403819   0.062905   6.420 1.39e-10 ***
share_rental_over_20     -0.300991   0.060896  -4.943 7.76e-07 ***
share_built_after_10      0.147794   0.171922   0.860   0.3900
log_price                 0.024977   0.026772   0.933   0.3509
log_sqft                  0.227432   0.019552  11.632  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.7390025)

    Null deviance: 18385  on 22678  degrees of freedom
Residual deviance: 16747  on 22662  degrees of freedom
AIC: 57520

Number of Fisher Scoring iterations: 2

[1] "Topic9"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4320  -0.6756  -0.1847   0.4822   3.7490

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              -2.6533308  0.4211983  -6.299 3.04e-10 ***
N_type_mixed             -0.0674516  0.0338176  -1.995   0.0461 *
N_type_white asian       -0.0358838  0.0188221  -1.906   0.0566 .
N_type_white black       -0.0663491  0.0316230  -2.098   0.0359 *
N_type_white latinx      -0.0091098  0.0247099  -0.369   0.7124
N_type_white mixed       -0.0125963  0.0238173  -0.529   0.5969
N_type_majority non-white -0.3768273  0.0483809  -7.789 7.06e-15 ***
pov_proportion           -0.0862251  0.1473426  -0.585   0.5584
log_income               -0.1623741  0.0391263  -4.150 3.34e-05 ***
pop_thousands            -0.0008511  0.0040383  -0.211   0.8331
share_college             0.6196881  0.1246917   4.970 6.75e-07 ***
share_commuters          -0.2187224  0.1276380  -1.714   0.0866 .
share_oo                 -0.0861754  0.0702431  -1.227   0.2199
share_rental_over_20     -0.2845430  0.0679999  -4.184 2.87e-05 ***
```

```
share_built_after_10       0.0205942  0.1919789   0.107   0.9146
log_price                 -0.2075855  0.0298951  -6.944 3.92e-12 ***
log_sqft                   0.3076216  0.0218330  14.090  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.9214837)

    Null deviance: 21300  on 22678  degrees of freedom
Residual deviance: 20883  on 22662  degrees of freedom
AIC: 62525

Number of Fisher Scoring iterations: 2

[1] "Topic10"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.0082  -0.7029  -0.2367   0.4554   4.0275

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             -3.378817   0.469052  -7.203 6.05e-13 ***
N_type_mixed            -0.042400   0.037660  -1.126 0.260233
N_type_white asian       0.024405   0.020961   1.164 0.244308
N_type_white black      -0.056827   0.035216  -1.614 0.106612
N_type_white latinx     -0.018672   0.027517  -0.679 0.497424
N_type_white mixed       0.016355   0.026523   0.617 0.537488
N_type_majority non-white -0.166006   0.053878  -3.081 0.002064 **
pov_proportion          -0.199284   0.164083  -1.215 0.224557
log_income              -0.178349   0.043572  -4.093 4.27e-05 ***
pop_thousands            0.008281   0.004497   1.841 0.065578 .
share_college            0.239471   0.138858   1.725 0.084618 .
share_commuters          0.568678   0.142139   4.001 6.33e-05 ***
share_oo                -0.250311   0.078224  -3.200 0.001376 **
share_rental_over_20    -0.430292   0.075726  -5.682 1.35e-08 ***
share_built_after_10    -0.741489   0.213790  -3.468 0.000525 ***
log_price               -0.034180   0.033292  -1.027 0.304573
log_sqft                 0.182444   0.024314   7.504 6.43e-14 ***
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.142765)

    Null deviance: 26416  on 22678  degrees of freedom
Residual deviance: 25897  on 22662  degrees of freedom
AIC: 67406

Number of Fisher Scoring iterations: 2

[1] "Topic11"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.1473  -0.7695  -0.0661   0.6809   5.2304

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               -7.22440    0.50168 -14.401  < 2e-16 ***
N_type_mixed              -0.09098    0.04028  -2.259   0.0239 *
N_type_white asian        -0.27801    0.02242 -12.401  < 2e-16 ***
N_type_white black         0.07518    0.03766   1.996   0.0459 *
N_type_white latinx       -0.20863    0.02943  -7.089 1.39e-12 ***
N_type_white mixed        -0.04116    0.02837  -1.451   0.1468
N_type_majority non-white -0.70225    0.05763 -12.187  < 2e-16 ***
pov_proportion             0.34647    0.17550   1.974   0.0484 *
log_income                 0.24607    0.04660   5.280 1.30e-07 ***
pop_thousands             -0.02450    0.00481  -5.094 3.54e-07 ***
share_college              3.66333    0.14852  24.666  < 2e-16 ***
share_commuters           -1.96923    0.15203 -12.953  < 2e-16 ***
share_oo                  -0.80992    0.08366  -9.681  < 2e-16 ***
share_rental_over_20       1.43496    0.08099  17.717  < 2e-16 ***
share_built_after_10      -0.90882    0.22866  -3.975 7.07e-05 ***
log_price                  1.10349    0.03561  30.991  < 2e-16 ***
log_sqft                  -1.36485    0.02601 -52.485  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.307258)
```

```
    Null deviance: 57864  on 22678  degrees of freedom
Residual deviance: 29625  on 22662  degrees of freedom
AIC: 70456

Number of Fisher Scoring iterations: 2

[1] "Topic12"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income + 
    pop_thousands + share_college + share_commuters + share_oo + 
    share_rental_over_20 + share_built_after_10 + log_price + 
    log_sqft, data = new_fit_topics)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-3.3133  -0.6427  -0.1087   0.5444   4.5675  

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)             -11.237589   0.405021 -27.746  < 2e-16 ***
N_type_mixed              0.081435   0.032519   2.504  0.01228 *  
N_type_white asian        0.146378   0.018099   8.088 6.39e-16 ***
N_type_white black        0.122952   0.030408   4.043 5.29e-05 ***
N_type_white latinx       0.145817   0.023761   6.137 8.56e-10 ***
N_type_white mixed        0.065246   0.022902   2.849  0.00439 ** 
N_type_majority non-white  0.054026   0.046523   1.161  0.24554    
pov_proportion           -0.223043   0.141683  -1.574  0.11545    
log_income               -0.149131   0.037623  -3.964 7.40e-05 ***
pop_thousands             0.005398   0.003883   1.390  0.16448    
share_college            -0.143885   0.119902  -1.200  0.23014    
share_commuters           1.178677   0.122736   9.603  < 2e-16 ***
share_oo                  0.087534   0.067545   1.296  0.19501    
share_rental_over_20     -0.394652   0.065388  -6.036 1.61e-09 ***
share_built_after_10     -0.240935   0.184605  -1.305  0.19186    
log_price                 0.393688   0.028747  13.695  < 2e-16 ***
log_sqft                  0.767550   0.020994  36.560  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.8520569)

    Null deviance: 26993  on 22678  degrees of freedom
Residual deviance: 19309  on 22662  degrees of freedom
AIC: 60748
```

```
Number of Fisher Scoring iterations: 2

[1] "Topic13"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7406  -0.6658  -0.1534   0.6139   3.0775

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              -1.062663   0.382929  -2.775 0.005523 **
N_type_mixed              0.174924   0.030745   5.690 1.29e-08 ***
N_type_white asian        0.014538   0.017112   0.850 0.395574
N_type_white black        0.146896   0.028750   5.109 3.26e-07 ***
N_type_white latinx      -0.010841   0.022465  -0.483 0.629399
N_type_white mixed        0.052269   0.021653   2.414 0.015790 *
N_type_majority non-white  0.040542   0.043985   0.922 0.356683
pov_proportion           -0.165231   0.133955  -1.233 0.217411
log_income               -0.222096   0.035571  -6.244 4.35e-10 ***
pop_thousands            -0.013701   0.003671  -3.732 0.000191 ***
share_college             0.334957   0.113362   2.955 0.003133 **
share_commuters           0.118026   0.116041   1.017 0.309114
share_oo                  0.043084   0.063861   0.675 0.499903
share_rental_over_20     -0.575398   0.061822  -9.307  < 2e-16 ***
share_built_after_10     -0.440690   0.174536  -2.525 0.011579 *
log_price                 0.063254   0.027179   2.327 0.019957 *
log_sqft                 -0.100594   0.019849  -5.068 4.05e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.7616422)

    Null deviance: 17714  on 22678  degrees of freedom
Residual deviance: 17260  on 22662  degrees of freedom
AIC: 58204

Number of Fisher Scoring iterations: 2

[1] "Topic14"
```

```
Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.7324  -0.7152  -0.1295   0.6057   4.9229

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -14.484230   0.464817 -31.161  < 2e-16 ***
N_type_mixed              0.258557   0.037320   6.928 4.38e-12 ***
N_type_white asian        0.395585   0.020771  19.045  < 2e-16 ***
N_type_white black        0.184599   0.034898   5.290 1.24e-07 ***
N_type_white latinx       0.137824   0.027269   5.054 4.35e-07 ***
N_type_white mixed        0.102888   0.026284   3.914 9.09e-05 ***
N_type_majority non-white  0.754488   0.053391  14.131  < 2e-16 ***
pov_proportion            0.335149   0.162601   2.061 0.039298 *
log_income                0.100201   0.043178   2.321 0.020316 *
pop_thousands             0.016310   0.004457   3.660 0.000253 ***
share_college            -1.384379   0.137605 -10.061  < 2e-16 ***
share_commuters           1.338698   0.140856   9.504  < 2e-16 ***
share_oo                  0.852421   0.077517  10.997  < 2e-16 ***
share_rental_over_20     -0.776009   0.075042 -10.341  < 2e-16 ***
share_built_after_10     -0.661197   0.211860  -3.121 0.001805 **
log_price                -0.022160   0.032991  -0.672 0.501790
log_sqft                  1.083052   0.024094  44.951  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.122222)

    Null deviance: 42596  on 22678  degrees of freedom
Residual deviance: 25432  on 22662  degrees of freedom
AIC: 66994

Number of Fisher Scoring iterations: 2

[1] "Topic15"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
```

```
    log_sqft, data = new_fit_topics)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.7860  -0.6708  -0.1668   0.5521   3.5733  

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)               -4.7106734  0.4086646 -11.527  < 2e-16 ***
N_type_mixed              -0.0348737  0.0328112  -1.063 0.287857    
N_type_white asian        -0.0910284  0.0182620  -4.985 6.26e-07 ***
N_type_white black         0.0877082  0.0306820   2.859 0.004259 ** 
N_type_white latinx       -0.0133993  0.0239746  -0.559 0.576240    
N_type_white mixed        -0.0001119  0.0231085  -0.005 0.996135    
N_type_majority non-white -0.2235027  0.0469412  -4.761 1.94e-06 ***
pov_proportion            -0.2480704  0.1429580  -1.735 0.082707 .  
log_income                -0.1694537  0.0379620  -4.464 8.09e-06 ***
pop_thousands             -0.0058591  0.0039181  -1.495 0.134829    
share_college              0.1341766  0.1209812   1.109 0.267412    
share_commuters           -0.1320925  0.1238398  -1.067 0.286146    
share_oo                   0.2351702  0.0681529   3.451 0.000560 ***
share_rental_over_20      -0.2521294  0.0659764  -3.822 0.000133 ***
share_built_after_10      -0.0711173  0.1862661  -0.382 0.702610    
log_price                  0.2181386  0.0290055   7.521 5.66e-14 ***
log_sqft                   0.0929103  0.0211833   4.386 1.16e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.867458)

    Null deviance: 20286  on 22678  degrees of freedom
Residual deviance: 19658  on 22662  degrees of freedom
AIC: 61155

Number of Fisher Scoring iterations: 2

[1] "Topic16"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income + 
    pop_thousands + share_college + share_commuters + share_oo + 
    share_rental_over_20 + share_built_after_10 + log_price + 
    log_sqft, data = new_fit_topics)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
```

```
-3.3035  -0.7882  -0.1616   0.6153   4.5808

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -1.261e+01  5.025e-01 -25.102  < 2e-16 ***
N_type_mixed             1.368e-02  4.034e-02   0.339 0.734586
N_type_white asian      -7.610e-02  2.245e-02  -3.389 0.000702 ***
N_type_white black       7.173e-02  3.773e-02   1.901 0.057286 .
N_type_white latinx     -2.978e-02  2.948e-02  -1.010 0.312453
N_type_white mixed      -1.769e-02  2.841e-02  -0.622 0.533651
N_type_majority non-white  2.064e-01  5.772e-02   3.577 0.000349 ***
pov_proportion          -1.408e-01  1.758e-01  -0.801 0.423073
log_income               8.488e-02  4.668e-02   1.818 0.069003 .
pop_thousands            7.384e-05  4.818e-03   0.015 0.987772
share_college           -5.315e-01  1.488e-01  -3.573 0.000354 ***
share_commuters          3.862e-01  1.523e-01   2.536 0.011213 *
share_oo                 6.058e-01  8.380e-02   7.229 5.01e-13 ***
share_rental_over_20     5.578e-01  8.112e-02   6.876 6.33e-12 ***
share_built_after_10     9.074e-03  2.290e-01   0.040 0.968398
log_price                8.809e-01  3.567e-02  24.700  < 2e-16 ***
log_sqft                -9.747e-02  2.605e-02  -3.742 0.000183 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.311528)

    Null deviance: 32255  on 22678  degrees of freedom
Residual deviance: 29722  on 22662  degrees of freedom
AIC: 70530

Number of Fisher Scoring iterations: 2

[1] "Topic17"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income + 
    pop_thousands + share_college + share_commuters + share_oo + 
    share_rental_over_20 + share_built_after_10 + log_price + 
    log_sqft, data = new_fit_topics)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-4.4832  -0.7574   0.0071   0.7732   6.7617  

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)              -19.587043   0.496353 -39.462  < 2e-16 ***
N_type_mixed              -0.051080   0.039852  -1.282  0.19994
N_type_white asian         0.087324   0.022181   3.937 8.28e-05 ***
N_type_white black         0.040649   0.037265   1.091  0.27537
N_type_white latinx        0.114176   0.029119   3.921 8.84e-05 ***
N_type_white mixed         0.025513   0.028067   0.909  0.36335
N_type_majority non-white  0.178322   0.057013   3.128  0.00176 **
pov_proportion            -0.735281   0.173633  -4.235 2.30e-05 ***
log_income                -0.144000   0.046108  -3.123  0.00179 **
pop_thousands              0.019003   0.004759   3.993 6.54e-05 ***
share_college             -2.989657   0.146940 -20.346  < 2e-16 ***
share_commuters            1.007249   0.150412   6.697 2.18e-11 ***
share_oo                   1.411510   0.082777  17.052  < 2e-16 ***
share_rental_over_20      -0.328027   0.080133  -4.094 4.26e-05 ***
share_built_after_10       0.745823   0.226234   3.297  0.00098 ***
log_price                  0.247457   0.035229   7.024 2.21e-12 ***
log_sqft                   2.018273   0.025729  78.445  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.279662)

    Null deviance: 73950  on 22678  degrees of freedom
Residual deviance: 29000  on 22662  degrees of freedom
AIC: 69972

Number of Fisher Scoring iterations: 2

[1] "Topic18"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income + 
    pop_thousands + share_college + share_commuters + share_oo + 
    share_rental_over_20 + share_built_after_10 + log_price + 
    log_sqft, data = new_fit_topics)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.4305  -0.6384  -0.1828   0.4012   5.7507  

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)               -5.585694   0.426375 -13.100  < 2e-16 ***
N_type_mixed              -0.050531   0.034233  -1.476  0.13994    
N_type_white asian        -0.062039   0.019053  -3.256  0.00113 ** 
N_type_white black        -0.075643   0.032012  -2.363  0.01814 *  
```

```
N_type_white latinx       -0.038356   0.025014  -1.533  0.12519
N_type_white mixed        -0.027970   0.024110  -1.160  0.24602
N_type_majority non-white  0.043805   0.048976   0.894  0.37110
pov_proportion            -0.184963   0.149154  -1.240  0.21496
log_income                -0.074644   0.039607  -1.885  0.05950 .
pop_thousands              0.005951   0.004088   1.456  0.14550
share_college             -1.981637   0.126224 -15.699  < 2e-16 ***
share_commuters            0.071578   0.129207   0.554  0.57960
share_oo                   0.628544   0.071106   8.839  < 2e-16 ***
share_rental_over_20      -0.171789   0.068836  -2.496  0.01258 *
share_built_after_10       1.006926   0.194338   5.181 2.22e-07 ***
log_price                 -0.397396   0.030263 -13.132  < 2e-16 ***
log_sqft                   0.569022   0.022101  25.746  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.9442748)

    Null deviance: 25618  on 22678  degrees of freedom
Residual deviance: 21399  on 22662  degrees of freedom
AIC: 63079

Number of Fisher Scoring iterations: 2

[1] "Topic19"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7507  -0.8208  -0.2278   0.5820   4.7143

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                0.360277   0.509231   0.707  0.47927
N_type_mixed              -0.032150   0.040886  -0.786  0.43168
N_type_white asian        -0.135408   0.022756  -5.950 2.71e-09 ***
N_type_white black        -0.123326   0.038232  -3.226  0.00126 **
N_type_white latinx       -0.069297   0.029874  -2.320  0.02037 *
N_type_white mixed        -0.047007   0.028795  -1.632  0.10260
N_type_majority non-white -0.324123   0.058493  -5.541 3.04e-08 ***
pov_proportion             0.226994   0.178138   1.274  0.20258
```

```
log_income              -0.201266   0.047304  -4.255 2.10e-05 ***
pop_thousands            -0.009056   0.004882  -1.855  0.06363 .
share_college           -0.114772   0.150753  -0.761  0.44647
share_commuters          -0.632818   0.154315  -4.101 4.13e-05 ***
share_oo               0.204346   0.084924   2.406  0.01613 *
share_rental_over_20      -0.268959   0.082212  -3.272  0.00107 **
share_built_after_10       0.750703   0.232104   3.234  0.00122 **
log_price               -0.583178   0.036143 -16.135  < 2e-16 ***
log_sqft                0.237714   0.026396   9.006  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.346927)

    Null deviance: 31927  on 22678  degrees of freedom
Residual deviance: 30524  on 22662  degrees of freedom
AIC: 71134

Number of Fisher Scoring iterations: 2

[1] "Topic20"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.1046  -0.7802  -0.0976   0.6870   4.7624

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              -15.395104   0.501858 -30.676  < 2e-16 ***
N_type_mixed              -0.048034   0.040294  -1.192   0.2332
N_type_white asian         -0.131560   0.022427  -5.866 4.52e-09 ***
N_type_white black          0.018797   0.037679   0.499   0.6179
N_type_white latinx        -0.070707   0.029442  -2.402   0.0163 *
N_type_white mixed          0.054234   0.028378   1.911   0.0560 .
N_type_majority non-white  -0.501323   0.057646  -8.697  < 2e-16 ***
pov_proportion             0.005675   0.175559   0.032   0.9742
log_income                0.431744   0.046619   9.261  < 2e-16 ***
pop_thousands             -0.005592   0.004812  -1.162   0.2452
share_college             1.256118   0.148570   8.455  < 2e-16 ***
share_commuters            -0.094698   0.152081  -0.623   0.5335
```

```
share_oo                  -0.470024   0.083695  -5.616 1.98e-08 ***
share_rental_over_20       1.952982   0.081022  24.104  < 2e-16 ***
share_built_after_10      -0.697349   0.228743  -3.049   0.0023 **
log_price                  1.426024   0.035620  40.034  < 2e-16 ***
log_sqft                  -0.801337   0.026014 -30.804  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.308204)

    Null deviance: 44831  on 22678  degrees of freedom
Residual deviance: 29647  on 22662  degrees of freedom
AIC: 70472

Number of Fisher Scoring iterations: 2

[1] "Topic21"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3106  -0.6377  -0.2055   0.4688   3.8108

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              -5.660664   0.407038 -13.907  < 2e-16 ***
N_type_mixed              0.272414   0.032681   8.336  < 2e-16 ***
N_type_white asian       -0.085608   0.018189  -4.707 2.53e-06 ***
N_type_white black        0.049191   0.030560   1.610 0.107487
N_type_white latinx      -0.146784   0.023879  -6.147 8.03e-10 ***
N_type_white mixed        0.070898   0.023017   3.080 0.002070 **
N_type_majority non-white -0.064380   0.046754  -1.377 0.168529
pov_proportion            0.612549   0.142389   4.302 1.70e-05 ***
log_income               -0.100969   0.037811  -2.670 0.007582 **
pop_thousands            -0.019113   0.003903  -4.898 9.77e-07 ***
share_college             0.329741   0.120500   2.736 0.006216 **
share_commuters           0.787607   0.123347   6.385 1.74e-10 ***
share_oo                  1.048261   0.067882  15.442  < 2e-16 ***
share_rental_over_20      0.110327   0.065714   1.679 0.093186 .
share_built_after_10     -0.265213   0.185525  -1.430 0.152867
log_price                 0.092135   0.028890   3.189 0.001429 **
```

```
log_sqft                  0.075568   0.021099   3.582 0.000342 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.8605674)

    Null deviance: 21048  on 22678  degrees of freedom
Residual deviance: 19502  on 22662  degrees of freedom
AIC: 60974

Number of Fisher Scoring iterations: 2

[1] "Topic22"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income + 
    pop_thousands + share_college + share_commuters + share_oo + 
    share_rental_over_20 + share_built_after_10 + log_price + 
    log_sqft, data = new_fit_topics)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.7405  -0.6400  -0.1277   0.4097   7.2597  

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)              -3.204136   0.437903  -7.317 2.62e-13 ***
N_type_mixed             -0.025344   0.035159  -0.721 0.471005    
N_type_white asian       -0.188909   0.019569  -9.654  < 2e-16 ***
N_type_white black        0.019329   0.032877   0.588 0.556595    
N_type_white latinx      -0.198228   0.025690  -7.716 1.25e-14 ***
N_type_white mixed       -0.089437   0.024762  -3.612 0.000305 ***
N_type_majority non-white -0.317281   0.050300  -6.308 2.88e-10 ***
pov_proportion            1.010130   0.153186   6.594 4.37e-11 ***
log_income               -0.219665   0.040678  -5.400 6.73e-08 ***
pop_thousands            -0.020559   0.004198  -4.897 9.81e-07 ***
share_college             3.834237   0.129637  29.577  < 2e-16 ***
share_commuters          -0.501775   0.132700  -3.781 0.000156 ***
share_oo                 -0.301389   0.073029  -4.127 3.69e-05 ***
share_rental_over_20     -0.719929   0.070697 -10.183  < 2e-16 ***
share_built_after_10     -0.280680   0.199593  -1.406 0.159660    
log_price                 0.225790   0.031081   7.265 3.86e-13 ***
log_sqft                 -0.425229   0.022699 -18.733  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
(Dispersion parameter for gaussian family taken to be 0.9960264)

    Null deviance: 26341  on 22678  degrees of freedom
Residual deviance: 22572  on 22662  degrees of freedom
AIC: 64289

Number of Fisher Scoring iterations: 2

[1] "Topic23"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.8675  -0.9275  -0.0692   0.8987   6.3399

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -10.772748   0.602772 -17.872  < 2e-16 ***
N_type_mixed             -0.154304   0.048396  -3.188 0.001433 **
N_type_white asian       -0.359445   0.026936 -13.344  < 2e-16 ***
N_type_white black       -0.076763   0.045255  -1.696 0.089856 .
N_type_white latinx      -0.200165   0.035362  -5.660 1.53e-08 ***
N_type_white mixed       -0.132246   0.034085  -3.880 0.000105 ***
N_type_majority non-white -0.559720   0.069237  -8.084 6.57e-16 ***
pov_proportion            1.251286   0.210860   5.934 3.00e-09 ***
log_income               -0.574015   0.055993 -10.252  < 2e-16 ***
pop_thousands            -0.054190   0.005779  -9.377  < 2e-16 ***
share_college             2.250121   0.178445  12.610  < 2e-16 ***
share_commuters          -0.985620   0.182661  -5.396 6.89e-08 ***
share_oo                  1.457458   0.100524  14.499  < 2e-16 ***
share_rental_over_20     -1.481090   0.097314 -15.220  < 2e-16 ***
share_built_after_10      0.093711   0.274739   0.341 0.733037
log_price                 0.198469   0.042783   4.639 3.52e-06 ***
log_sqft                  1.514144   0.031245  48.460  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.887211)

    Null deviance: 68852  on 22678  degrees of freedom
Residual deviance: 42768  on 22662  degrees of freedom
```

```
AIC: 78783

Number of Fisher Scoring iterations: 2

[1] "Topic24"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2558  -0.7516  -0.0319   0.7422   3.2593

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                4.415036   0.435373  10.141  < 2e-16 ***
N_type_mixed               0.071568   0.034956   2.047   0.0406 *
N_type_white asian        -0.044863   0.019456  -2.306   0.0211 *
N_type_white black        -0.082482   0.032687  -2.523   0.0116 *
N_type_white latinx       -0.061439   0.025541  -2.405   0.0162 *
N_type_white mixed        -0.010584   0.024619  -0.430   0.6673
N_type_majority non-white -0.360934   0.050009  -7.217 5.47e-13 ***
pov_proportion             0.726958   0.152301   4.773 1.82e-06 ***
log_income                -0.312736   0.040443  -7.733 1.10e-14 ***
pop_thousands             -0.009886   0.004174  -2.368   0.0179 *
share_college              1.855341   0.128888  14.395  < 2e-16 ***
share_commuters           -0.010597   0.131933  -0.080   0.9360
share_oo                   0.003537   0.072607   0.049   0.9611
share_rental_over_20      -0.710023   0.070288 -10.102  < 2e-16 ***
share_built_after_10      -0.466480   0.198440  -2.351   0.0187 *
log_price                 -0.735551   0.030901 -23.803  < 2e-16 ***
log_sqft                   0.093826   0.022568   4.158 3.23e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.9845497)

    Null deviance: 24485  on 22678  degrees of freedom
Residual deviance: 22312  on 22662  degrees of freedom
AIC: 64026

Number of Fisher Scoring iterations: 2
```

```
[1] "Topic25"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4558  -0.8140  -0.1329   0.7709   3.5836

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                 0.971968   0.491532   1.977 0.048006 *
N_type_mixed                0.161280   0.039465   4.087 4.39e-05 ***
N_type_white asian          0.127587   0.021965   5.809 6.38e-09 ***
N_type_white black         -0.048340   0.036904  -1.310 0.190241
N_type_white latinx        -0.038926   0.028836  -1.350 0.177055
N_type_white mixed          0.064306   0.027794   2.314 0.020698 *
N_type_majority non-white -0.110371   0.056460  -1.955 0.050612 .
pov_proportion              0.296406   0.171947   1.724 0.084753 .
log_income                 -0.107334   0.045660  -2.351 0.018745 *
pop_thousands              -0.012363   0.004713  -2.623 0.008715 **
share_college               2.335160   0.145513  16.048  < 2e-16 ***
share_commuters             0.563342   0.148952   3.782 0.000156 ***
share_oo                   -0.369419   0.081973  -4.507 6.62e-06 ***
share_rental_over_20       -0.068277   0.079355  -0.860 0.389577
share_built_after_10       -1.315815   0.224037  -5.873 4.33e-09 ***
log_price                  -0.145749   0.034887  -4.178 2.96e-05 ***
log_sqft                   -0.469058   0.025479 -18.410  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.254927)

    Null deviance: 31711  on 22678  degrees of freedom
Residual deviance: 28439  on 22662  degrees of freedom
AIC: 69529

Number of Fisher Scoring iterations: 2

[1] "Topic26"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
```

```
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-6.0323  -1.1760  -0.2001   0.9562   5.8690

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              -9.155452   0.758433 -12.072  < 2e-16 ***
N_type_mixed              0.278046   0.060894   4.566 5.00e-06 ***
N_type_white asian        0.410698   0.033892  12.118  < 2e-16 ***
N_type_white black        0.321036   0.056942   5.638 1.74e-08 ***
N_type_white latinx       0.307178   0.044494   6.904 5.20e-12 ***
N_type_white mixed        0.190026   0.042887   4.431 9.43e-06 ***
N_type_majority non-white  1.070328   0.087117  12.286  < 2e-16 ***
pov_proportion           -1.896033   0.265313  -7.146 9.18e-13 ***
log_income                0.660821   0.070453   9.380  < 2e-16 ***
pop_thousands             0.082668   0.007272  11.369  < 2e-16 ***
share_college            -6.514337   0.224527 -29.014  < 2e-16 ***
share_commuters           2.452375   0.229832  10.670  < 2e-16 ***
share_oo                 -0.863246   0.126484  -6.825 9.02e-12 ***
share_rental_over_20      1.134696   0.122444   9.267  < 2e-16 ***
share_built_after_10      0.890835   0.345688   2.577  0.00997 **
log_price                 0.287612   0.053831   5.343 9.24e-08 ***
log_sqft                 -0.860551   0.039314 -21.889  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 2.987785)

    Null deviance: 76239  on 22678  degrees of freedom
Residual deviance: 67709  on 22662  degrees of freedom
AIC: 89202

Number of Fisher Scoring iterations: 2

[1] "Topic27"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)
```

```
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.38170  -0.58575  -0.09556   0.51253   2.96432

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              -6.900e+00  3.561e-01 -19.379  < 2e-16 ***
N_type_mixed              6.420e-02  2.859e-02   2.246 0.024730 *
N_type_white asian       -2.939e-02  1.591e-02  -1.847 0.064748 .
N_type_white black        3.035e-02  2.673e-02   1.135 0.256324
N_type_white latinx       3.281e-03  2.089e-02   0.157 0.875199
N_type_white mixed       -6.961e-05  2.013e-02  -0.003 0.997242
N_type_majority non-white  1.458e-01  4.090e-02   3.566 0.000364 ***
pov_proportion           -8.530e-02  1.246e-01  -0.685 0.493440
log_income                7.888e-02  3.308e-02   2.385 0.017095 *
pop_thousands             6.734e-03  3.414e-03   1.973 0.048550 *
share_college            -6.102e-01  1.054e-01  -5.789 7.20e-09 ***
share_commuters           2.833e-01  1.079e-01   2.626 0.008652 **
share_oo                  3.933e-01  5.938e-02   6.623 3.59e-11 ***
share_rental_over_20      1.668e-01  5.748e-02   2.902 0.003712 **
share_built_after_10     -1.796e-01  1.623e-01  -1.107 0.268466
log_price                 2.683e-01  2.527e-02  10.615  < 2e-16 ***
log_sqft                 -6.150e-02  1.846e-02  -3.332 0.000863 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.6585291)

    Null deviance: 15331  on 22678  degrees of freedom
Residual deviance: 14924  on 22662  degrees of freedom
AIC: 54905

Number of Fisher Scoring iterations: 2

[1] "Topic28"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.5311  -0.7392  -0.0484   0.6915   4.9011
```

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                9.020234   0.488419  18.468  < 2e-16 ***
N_type_mixed              -0.100322   0.039215  -2.558   0.0105 *
N_type_white asian        -0.379441   0.021826 -17.385  < 2e-16 ***
N_type_white black        -0.048476   0.036670  -1.322   0.1862
N_type_white latinx       -0.299471   0.028653 -10.451  < 2e-16 ***
N_type_white mixed        -0.163474   0.027618  -5.919 3.28e-09 ***
N_type_majority non-white -0.844800   0.056102 -15.058  < 2e-16 ***
pov_proportion             1.468695   0.170858   8.596  < 2e-16 ***
log_income                -0.184674   0.045371  -4.070 4.71e-05 ***
pop_thousands             -0.042556   0.004683  -9.088  < 2e-16 ***
share_college              6.287163   0.144592  43.482  < 2e-16 ***
share_commuters           -2.889177   0.148008 -19.520  < 2e-16 ***
share_oo                  -1.030631   0.081454 -12.653  < 2e-16 ***
share_rental_over_20      -0.003984   0.078852  -0.051   0.9597
share_built_after_10      -1.667037   0.222618  -7.488 7.23e-14 ***
log_price                 -0.290643   0.034666  -8.384  < 2e-16 ***
log_sqft                  -1.383850   0.025317 -54.660  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.239082)

    Null deviance: 63827  on 22678  degrees of freedom
Residual deviance: 28080  on 22662  degrees of freedom
AIC: 69241

Number of Fisher Scoring iterations: 2

[1] "Topic29"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1850  -0.6335  -0.1401   0.4127   5.1378

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               -8.231077   0.456609 -18.027  < 2e-16 ***
N_type_mixed               0.124655   0.036661   3.400 0.000674 ***
```

```
N_type_white asian        -0.012569   0.020404  -0.616 0.537904
N_type_white black         0.046021   0.034282   1.342 0.179463
N_type_white latinx        0.072681   0.026787   2.713 0.006667 **
N_type_white mixed         0.019823   0.025820   0.768 0.442641
N_type_majority non-white -0.259772   0.052448  -4.953 7.36e-07 ***
pov_proportion             0.117224   0.159730   0.734 0.463025
log_income                -0.307315   0.042416  -7.245 4.45e-13 ***
pop_thousands             -0.014011   0.004378  -3.201 0.001374 **
share_college              0.261945   0.135175   1.938 0.052657 .
share_commuters            0.093360   0.138369   0.675 0.499861
share_oo                   0.939260   0.076148  12.335  < 2e-16 ***
share_rental_over_20      -0.352895   0.073717  -4.787 1.70e-06 ***
share_built_after_10      -0.029738   0.208119  -0.143 0.886379
log_price                  0.064435   0.032408   1.988 0.046800 *
log_sqft                   0.755783   0.023668  31.932  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.082935)

    Null deviance: 31232  on 22678  degrees of freedom
Residual deviance: 24541  on 22662  degrees of freedom
AIC: 66186

Number of Fisher Scoring iterations: 2

[1] "Topic30"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income + 
    pop_thousands + share_college + share_commuters + share_oo + 
    share_rental_over_20 + share_built_after_10 + log_price + 
    log_sqft, data = new_fit_topics)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-3.3559  -0.7797  -0.0984   0.7108   3.9208  

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)                0.044505   0.477783   0.093 0.925785    
N_type_mixed               0.337104   0.038361   8.788  < 2e-16 ***
N_type_white asian         0.069205   0.021351   3.241 0.001191 ** 
N_type_white black         0.222449   0.035871   6.201 5.70e-10 ***
N_type_white latinx       -0.037059   0.028029  -1.322 0.186137    
N_type_white mixed         0.016624   0.027017   0.615 0.538342    
```

```
N_type_majority non-white  0.269887   0.054880   4.918 8.82e-07 ***
pov_proportion            1.186807   0.167137   7.101 1.28e-12 ***
log_income               -0.447733   0.044383 -10.088  < 2e-16 ***
pop_thousands            -0.028564   0.004581  -6.236 4.58e-10 ***
share_college             2.890311   0.141443  20.434  < 2e-16 ***
share_commuters          -0.137338   0.144785  -0.949 0.342853
share_oo                  0.572330   0.079680   7.183 7.04e-13 ***
share_rental_over_20     -1.005189   0.077135 -13.032  < 2e-16 ***
share_built_after_10     -0.126171   0.217770  -0.579 0.562341
log_price                -0.125056   0.033911  -3.688 0.000227 ***
log_sqft                  0.071829   0.024766   2.900 0.003732 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.185701)

    Null deviance: 28310  on 22678  degrees of freedom
Residual deviance: 26870  on 22662  degrees of freedom
AIC: 68242

Number of Fisher Scoring iterations: 2

[1] "Topic31"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5110  -0.7692  -0.1919   0.5821   4.0563

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               -7.909924   0.490415 -16.129  < 2e-16 ***
N_type_mixed               0.414031   0.039375  10.515  < 2e-16 ***
N_type_white asian         0.525572   0.021915  23.982  < 2e-16 ***
N_type_white black         0.208603   0.036820   5.666 1.48e-08 ***
N_type_white latinx        0.249997   0.028771   8.689  < 2e-16 ***
N_type_white mixed         0.181474   0.027731   6.544 6.12e-11 ***
N_type_majority non-white  1.189532   0.056331  21.117  < 2e-16 ***
pov_proportion            -1.399623   0.171556  -8.158 3.57e-16 ***
log_income                 0.383992   0.045556   8.429  < 2e-16 ***
pop_thousands              0.042332   0.004702   9.003  < 2e-16 ***
```

```
share_college            -2.680256   0.145183 -18.461  < 2e-16 ***
share_commuters           2.211281   0.148613  14.879  < 2e-16 ***
share_oo                 -0.348446   0.081786  -4.260 2.05e-05 ***
share_rental_over_20      0.490710   0.079175   6.198 5.82e-10 ***
share_built_after_10     -0.817664   0.223527  -3.658 0.000255 ***
log_price                 0.094301   0.034808   2.709 0.006750 **
log_sqft                 -0.326620   0.025421 -12.848  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.249229)

    Null deviance: 31330  on 22678  degrees of freedom
Residual deviance: 28310  on 22662  degrees of freedom
AIC: 69426

Number of Fisher Scoring iterations: 2

[1] "Topic32"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4560  -0.8044  -0.1786   0.7183   3.9971

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               1.144328   0.499041   2.293  0.02185 *
N_type_mixed             -0.006226   0.040067  -0.155  0.87651
N_type_white asian       -0.102752   0.022301  -4.608 4.10e-06 ***
N_type_white black       -0.085555   0.037467  -2.283  0.02241 *
N_type_white latinx      -0.157500   0.029277  -5.380 7.53e-08 ***
N_type_white mixed       -0.087468   0.028219  -3.100  0.00194 **
N_type_majority non-white -0.025504   0.057322  -0.445  0.65638
pov_proportion            1.001863   0.174573   5.739 9.65e-09 ***
log_income               -0.005749   0.046357  -0.124  0.90130
pop_thousands            -0.005099   0.004785  -1.066  0.28661
share_college             1.114338   0.147736   7.543 4.77e-14 ***
share_commuters          -0.487179   0.151227  -3.222  0.00128 **
share_oo                 -0.001446   0.083225  -0.017  0.98614
share_rental_over_20     -0.154379   0.080567  -1.916  0.05536 .
```

```
share_built_after_10       0.571033   0.227459   2.510  0.01206 *
log_price                 -0.544660   0.035420 -15.377  < 2e-16 ***
log_sqft                  -0.181992   0.025868  -7.035 2.04e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.293559)

    Null deviance: 31531  on 22678  degrees of freedom
Residual deviance: 29315  on 22662  degrees of freedom
AIC: 70217

Number of Fisher Scoring iterations: 2

[1] "Topic33"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.7498  -1.0383  -0.1915   0.9599   4.3360

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               -6.383385   0.635484 -10.045  < 2e-16 ***
N_type_mixed               0.063598   0.051022   1.246 0.212604
N_type_white asian         0.102114   0.028398   3.596 0.000324 ***
N_type_white black         0.138504   0.047711   2.903 0.003700 **
N_type_white latinx        0.107752   0.037281   2.890 0.003853 **
N_type_white mixed         0.067435   0.035934   1.877 0.060584 .
N_type_majority non-white  0.581619   0.072995   7.968 1.69e-15 ***
pov_proportion            -0.861487   0.222304  -3.875 0.000107 ***
log_income                 0.567839   0.059032   9.619  < 2e-16 ***
pop_thousands              0.039221   0.006093   6.437 1.24e-10 ***
share_college             -3.457907   0.188129 -18.381  < 2e-16 ***
share_commuters            0.746671   0.192574   3.877 0.000106 ***
share_oo                  -0.941842   0.105980  -8.887  < 2e-16 ***
share_rental_over_20       1.105286   0.102595  10.773  < 2e-16 ***
share_built_after_10       0.523435   0.289649   1.807 0.070754 .
log_price                  0.295716   0.045104   6.556 5.64e-11 ***
log_sqft                  -0.861833   0.032941 -26.163  < 2e-16 ***
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 2.097607)

    Null deviance: 54391  on 22678  degrees of freedom
Residual deviance: 47536  on 22662  degrees of freedom
AIC: 81180

Number of Fisher Scoring iterations: 2

[1] "Topic34"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income + 
    pop_thousands + share_college + share_commuters + share_oo + 
    share_rental_over_20 + share_built_after_10 + log_price + 
    log_sqft, data = new_fit_topics)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.7789  -0.6809  -0.1817   0.4311   5.0159  

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)               -6.731960   0.464022 -14.508  < 2e-16 ***
N_type_mixed              -0.081557   0.037256  -2.189 0.028598 *  
N_type_white asian        -0.183215   0.020736  -8.836  < 2e-16 ***
N_type_white black        -0.138450   0.034838  -3.974 7.09e-05 ***
N_type_white latinx       -0.246075   0.027222  -9.040  < 2e-16 ***
N_type_white mixed        -0.102723   0.026239  -3.915 9.07e-05 ***
N_type_majority non-white -0.079933   0.053300  -1.500 0.133711    
pov_proportion             1.428178   0.162323   8.798  < 2e-16 ***
log_income                 0.122135   0.043104   2.833 0.004608 ** 
pop_thousands             -0.012149   0.004449  -2.731 0.006322 ** 
share_college              1.239549   0.137369   9.023  < 2e-16 ***
share_commuters           -0.465918   0.140615  -3.313 0.000923 ***
share_oo                   0.286542   0.077385   3.703 0.000214 ***
share_rental_over_20       0.292206   0.074913   3.901 9.62e-05 ***
share_built_after_10      -0.051648   0.211497  -0.244 0.807078    
log_price                  0.529653   0.032935  16.082  < 2e-16 ***
log_sqft                  -0.629801   0.024053 -26.184  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.118384)
```

```
    Null deviance: 28335  on 22678  degrees of freedom
Residual deviance: 25345  on 22662  degrees of freedom
AIC: 66917

Number of Fisher Scoring iterations: 2

[1] "Topic35"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8314  -0.6484  -0.0517   0.6210   3.2295

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               -1.346439   0.382426  -3.521 0.000431 ***
N_type_mixed               0.104597   0.030705   3.407 0.000659 ***
N_type_white asian        -0.009250   0.017089  -0.541 0.588330
N_type_white black         0.093910   0.028712   3.271 0.001074 **
N_type_white latinx       -0.103765   0.022435  -4.625 3.77e-06 ***
N_type_white mixed         0.005105   0.021625   0.236 0.813365
N_type_majority non-white  0.056042   0.043927   1.276 0.202046
pov_proportion             0.818136   0.133779   6.116 9.78e-10 ***
log_income                -0.120119   0.035525  -3.381 0.000723 ***
pop_thousands              0.001112   0.003667   0.303 0.761681
share_college              2.949274   0.113214  26.051  < 2e-16 ***
share_commuters           -0.008746   0.115889  -0.075 0.939840
share_oo                  -0.208443   0.063777  -3.268 0.001084 **
share_rental_over_20      -0.277890   0.061740  -4.501 6.80e-06 ***
share_built_after_10      -0.797570   0.174307  -4.576 4.77e-06 ***
log_price                  0.097936   0.027143   3.608 0.000309 ***
log_sqft                  -0.370767   0.019823 -18.704  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.7596428)

    Null deviance: 19703  on 22678  degrees of freedom
Residual deviance: 17215  on 22662  degrees of freedom
AIC: 58145
```

```
Number of Fisher Scoring iterations: 2

[1] "Topic36"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7494  -0.9151  -0.0611   0.8848   3.1861

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                -1.622380   0.465979  -3.482 0.000499 ***
N_type_mixed               -0.025979   0.037413  -0.694 0.487448
N_type_white asian          0.039213   0.020823   1.883 0.059694 .
N_type_white black         -0.049703   0.034985  -1.421 0.155421
N_type_white latinx         0.007710   0.027337   0.282 0.777928
N_type_white mixed         -0.018349   0.026349  -0.696 0.486208
N_type_majority non-white -0.083847   0.053525  -1.567 0.117244
pov_proportion             -0.238987   0.163007  -1.466 0.142632
log_income                  0.000763   0.043286   0.018 0.985936
pop_thousands               0.008721   0.004468   1.952 0.050941 .
share_college              -0.644866   0.137948  -4.675 2.96e-06 ***
share_commuters             0.192458   0.141208   1.363 0.172916
share_oo                   -0.281513   0.077711  -3.623 0.000292 ***
share_rental_over_20       -0.081397   0.075229  -1.082 0.279270
share_built_after_10        0.124464   0.212389   0.586 0.557869
log_price                  -0.390133   0.033073 -11.796  < 2e-16 ***
log_sqft                    0.156167   0.024154   6.465 1.03e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.127837)

    Null deviance: 26053  on 22678  degrees of freedom
Residual deviance: 25559  on 22662  degrees of freedom
AIC: 67108

Number of Fisher Scoring iterations: 2

[1] "Topic37"
```

```
Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0272  -0.9087  -0.0256   0.9342   3.7068

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               -0.41508    0.54860  -0.757 0.449294
N_type_mixed               0.04435    0.04405   1.007 0.313950
N_type_white asian        -0.08748    0.02451  -3.569 0.000360 ***
N_type_white black        -0.11044    0.04119  -2.681 0.007338 **
N_type_white latinx       -0.15194    0.03218  -4.721 2.36e-06 ***
N_type_white mixed        -0.10742    0.03102  -3.463 0.000536 ***
N_type_majority non-white -0.23278    0.06302  -3.694 0.000221 ***
pov_proportion             1.38322    0.19191   7.208 5.87e-13 ***
log_income                -0.44873    0.05096  -8.805  < 2e-16 ***
pop_thousands             -0.02801    0.00526  -5.325 1.02e-07 ***
share_college              2.05771    0.16241  12.670  < 2e-16 ***
share_commuters           -0.30374    0.16625  -1.827 0.067701 .
share_oo                   0.80907    0.09149   8.843  < 2e-16 ***
share_rental_over_20      -1.32404    0.08857 -14.949  < 2e-16 ***
share_built_after_10       0.14816    0.25005   0.593 0.553502
log_price                 -0.60168    0.03894 -15.453  < 2e-16 ***
log_sqft                   0.79749    0.02844  28.044  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.563244)

    Null deviance: 41678  on 22678  degrees of freedom
Residual deviance: 35426  on 22662  degrees of freedom
AIC: 74511

Number of Fisher Scoring iterations: 2

[1] "Topic38"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
```

```
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.7163  -0.7623  -0.1143   0.7508   3.4295

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              -12.992004   0.447322 -29.044  < 2e-16 ***
N_type_mixed               0.100029   0.035915   2.785  0.00535 **
N_type_white asian         0.094356   0.019989   4.720 2.37e-06 ***
N_type_white black         0.282515   0.033584   8.412  < 2e-16 ***
N_type_white latinx        0.108081   0.026243   4.119 3.83e-05 ***
N_type_white mixed         0.081256   0.025294   3.212  0.00132 **
N_type_majority non-white  0.059809   0.051382   1.164  0.24443
pov_proportion            -0.413708   0.156481  -2.644  0.00820 **
log_income                 0.085900   0.041553   2.067  0.03872 *
pop_thousands              0.011687   0.004289   2.725  0.00643 **
share_college             -0.357437   0.132425  -2.699  0.00696 **
share_commuters            0.841239   0.135554   6.206 5.53e-10 ***
share_oo                   0.146723   0.074600   1.967  0.04922 *
share_rental_over_20       0.587311   0.072217   8.133 4.42e-16 ***
share_built_after_10      -0.204014   0.203886  -1.001  0.31702
log_price                  1.121673   0.031749  35.329  < 2e-16 ***
log_sqft                  -0.150532   0.023187  -6.492 8.64e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.039335)

    Null deviance: 27384  on 22678  degrees of freedom
Residual deviance: 23553  on 22662  degrees of freedom
AIC: 65254

Number of Fisher Scoring iterations: 2

[1] "Topic39"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
```

```
-3.0228  -0.6985  -0.0875   0.6263   3.5385

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -8.240790   0.419276 -19.655  < 2e-16 ***
N_type_mixed             0.219597   0.033663   6.523 7.02e-11 ***
N_type_white asian       0.223667   0.018736  11.938  < 2e-16 ***
N_type_white black       0.081598   0.031479   2.592  0.00954 **
N_type_white latinx     -0.011639   0.024597  -0.473  0.63609
N_type_white mixed       0.075607   0.023709   3.189  0.00143 **
N_type_majority non-white  0.762536   0.048160  15.833  < 2e-16 ***
pov_proportion           1.486217   0.146670  10.133  < 2e-16 ***
log_income               0.240698   0.038948   6.180 6.52e-10 ***
pop_thousands            0.007637   0.004020   1.900  0.05747 .
share_college            1.767512   0.124123  14.240  < 2e-16 ***
share_commuters          0.678586   0.127056   5.341 9.34e-08 ***
share_oo                -0.083443   0.069923  -1.193  0.23274
share_rental_over_20    -0.314027   0.067690  -4.639 3.52e-06 ***
share_built_after_10    -1.029348   0.191103  -5.386 7.26e-08 ***
log_price                0.247963   0.029759   8.332  < 2e-16 ***
log_sqft                -0.251822   0.021733 -11.587  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.9130938)

    Null deviance: 22245  on 22678  degrees of freedom
Residual deviance: 20693  on 22662  degrees of freedom
AIC: 62317

Number of Fisher Scoring iterations: 2

[1] "Topic40"

Call:
glm(formula = log(get(i)) ~ N_type_ + pov_proportion + log_income +
    pop_thousands + share_college + share_commuters + share_oo +
    share_rental_over_20 + share_built_after_10 + log_price +
    log_sqft, data = new_fit_topics)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.77796  -0.19285   0.02759   0.21991   1.21688

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)              -4.474909   0.136561 -32.769  < 2e-16 ***
N_type_mixed              0.072053   0.010964   6.572 5.09e-11 ***
N_type_white asian       -0.003618   0.006102  -0.593 0.553283
N_type_white black        0.037941   0.010253   3.701 0.000216 ***
N_type_white latinx      -0.030354   0.008011  -3.789 0.000152 ***
N_type_white mixed        0.001073   0.007722   0.139 0.889469
N_type_majority non-white -0.019055   0.015686  -1.215 0.224475
pov_proportion            0.249398   0.047771   5.221 1.80e-07 ***
log_income               -0.075329   0.012686  -5.938 2.92e-09 ***
pop_thousands            -0.005261   0.001309  -4.018 5.89e-05 ***
share_college             0.713378   0.040427  17.646  < 2e-16 ***
share_commuters           0.124044   0.041383   2.997 0.002725 **
share_oo                  0.030154   0.022774   1.324 0.185498
share_rental_over_20     -0.215811   0.022047  -9.789  < 2e-16 ***
share_built_after_10     -0.269083   0.062243  -4.323 1.55e-05 ***
log_price                -0.011875   0.009693  -1.225 0.220531
log_sqft                  0.016374   0.007079   2.313 0.020725 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.09686475)

    Null deviance: 2277.7  on 22678  degrees of freedom
Residual deviance: 2195.1  on 22662  degrees of freedom
AIC: 11436

Number of Fisher Scoring iterations: 2
```

Output from Security Terms regressions

	Data Source	
	Safety Terms in *Craigslist* Discourse	Security System Registered in Tax Records
Neighborhood Type: Mixed	0.245***	-0.588***
	(0.059)	(0.145)
Neighborhood Type: White Asian	0.312***	-0.023
	(0.032)	(0.085)
Neighborhood Type: White Black	0.285***	-0.179
	(0.053)	(0.124)
Neighborhood Type: White Latinx	0.255***	-0.674***
	(0.042)	(0.120)
Neighborhood Type: White Mixed	0.134***	-0.047
	(0.041)	(0.102)
Neighborhood Type: Majority non-White	-0.120	-0.642*

	(0.091)	(0.294)
Poverty Proportion	0.945***	-1.051
	(0.247)	(0.689)
Log Income	-0.487***	-0.341
	(0.064)	(0.180)
Population in Thousands	-0.017*	-0.013
	(0.007)	(0.020)
Proportion with College Degree	3.603***	4.822***
	(0.199)	(0.543)
Proportion Commuters	0.605**	-2.420***
	(0.213)	(0.620)
Proportion of residences Owner Occupied	1.117***	1.254***
	(0.117)	(0.326)
Proportion of residences rented in Buildings with 20+ units	0.418***	2.530***
	(0.113)	(0.319)
Proportion of residences built after 2010	0.009	-0.182
	(0.322)	(0.819)
Intercept	2.526***	1.541
	(0.694)	(2.024)
Observations	45,358	7,609
Log Likelihood	-25,969.550	-3,895.204
Akaike Inf. Crit.	51,969.100	7,820.408
Note:	* p<0.05; ** p<0.01; *** p<0.001	

Appendix II: Topic Output

Topic 1 Top Words:

Highest Prob: util, share, privat, includ, entranc, electr, quiet, one, small, separ, street, gas, water, basement, tenant

FREX: mother, cottag, mil, share, entranc, law, separ, respect, cabin, daylight, adu, power, occupi, basement, vape

Lift: airstream, burgundian, deborah, entranceway, gun, jct, jctn, kichan, lattic, neonicotinoid, nobel, pesticid, prism, rol, roseanna

Score: share, mother, cottag, basement, mil, law, entranc, util, electr, internet, quiet, separ, cabl, apart, privat

Topic 2 Top Words:

Highest Prob: yard, bedroom, bath, fenc, garag, washer, dryer, duplex, back, car, nice, includ, pet, smoke, larg

FREX: duplex, rambler, hook, hookup, yard, fenc, shed, nice, cute, back, stove, section, bdrm, negoti, newer

Lift: mitt, southhil, subsequ, tanwax, waller, jovita, joan, stoner, tahuya, pitt, fain, artondal, scra, karl, sylvia

Score: duplex, yard, fenc, rambler, shed, car, garag, back, nice, bath, cul, hookup, washer, sac, dryer

Topic 3 Top Words:

Highest Prob: tour, schedul, leas, properti, note, month, pet, onlin, appoint, park, book, avail, heat, cost, util

FREX: book, note, anytim, tour, pdt, jmw, cost, onlin, schedul, date, hyperlink, cool, browser, wire, none

Lift: agil, apmpro, apport, basso, cayc, christyric, delorey, docdro, exur, fridayï, heatheridg, hnn, linh, mondayï, offcampustour

Score: book, tour, note, onlin, jmw, anytim, pdt, schedul, properti, descript, type, cost, today, cool, windermer

Topic 4 Top Words:

Highest Prob: apart, access, communiti, park, site, manag, free, conveni, locat, avail, easi, mainten, close, balconi, control

FREX: afford, emerg, site, onsit, mainten, paperless, apart, profession, public, control, transport, manag, payment, free, villa

Lift: paperless, bbï, buena, canap, gian, issac, jlmb, mirada, miranda, pursuit, rogen, septem, solara, stoneg, stonemeadow

Score: apart, site, communiti, control, manag, emerg, access, onlin, mainten, elev, balconi, free, afford, facil, transport

Topic 5 Top Words:

Highest Prob: loung, communiti, fit, rooftop, resid, amen, design, studio, center, grill, area, live, bike, pet, quartz

FREX: loung, leed, inspir, insignia, newest, rooftop, packag, sustain, wash, confer, pit, deliveri, usb, cardio, certifi

Lift: arenanet, barista, bis, bowman, breatheasi, brewstat, cornhol, curat, cyren, danzeâ, directvâ, freedland, geographi, ginza, imac

Score: loung, rooftop, studio, quartz, fit, leed, bike, packag, communiti, hour, concierg, energi, insignia, wifi, cardio

Topic 6 Top Words:

Highest Prob: room, bedroom, larg, live, kitchen, space, closet, area, bathroom, full, two, dine, storag, spacious, size

FREX: larg, room, closet, live, dine, full, size, huge, two, area, space, bathroom, bedroom, kitchen, extra

Lift: replubican, mock, nock, sung, sheva, duckl, applaus, norwood, din, galley, imaginari, larg, roomi, allergi, hallway

Score: room, larg, bedroom, live, closet, dine, kitchen, space, full, bathroom, two, area, size, spacious, storag

Topic 7 Top Words:

Highest Prob: unit, apart, avail, build, park, bedroom, floor, one, street, dryer, washer, bus, locat, laundri, line

FREX: unit, plex, apt, triplex, build, coin, wsg, oper, locker, complex, bldg, similar, buslin, line, fourplex

Lift: alderlan, aol, banburi, buildingon, cornellandassoci, daren, flor, javista, kimi, prearrang, shunt, careek, jani, westcoastarm, galahad

Score: unit, apart, build, apt, coin, plex, bus, garbag, site, triplex, sewer, avail, park, laundri, street

Topic 8 Top Words:

Highest Prob: golf, point, park, cours, home, height, includ, nearbi, well, provid, great, bay, close, entertain, mani

FREX: cours, golf, jefferson, height, point, sand, jackson, defianc, meadowbrook, harbour, matthew, brown, arbor, chamber, hale

Lift: fiddler, glover, marilynn, norwegian, quetzal, refrigeratoâ, woodshop, cours, dragu, tipp, victori, jefferson, pierson, copperlin, london

Score: golf, cours, point, height, sand, jefferson, harbour, jackson, meadowbrook, defianc, home, matthew, arbor, chamber, club

Topic 9 Top Words:

Highest Prob: avail, properti, allow, squar, feet, bed, bath, offer, detail, dog, address, rent, cat, now, rental

FREX: feet, squar, address, allow, llc, mcgill, avenueoneresidenti, barb, bender, cat, marilyn, detail, properti, gregori, eclips

Lift: avenueoneresidenti, reserev, mcgill, amazonslu, avenueon, barb, bender, christensen, gladiat, gloria, gorder, gwenev, hankin, hillrent, kaci

Score: descript, properti, allow, feet, address, llc, squar, polici, cat, rental, applic, fee, detail, dog, avenu

Topic 10 Top Words:

Highest Prob: featur, park, leas, refriger, year, pet, rang, dishwash, oven, properti, sqft, microwav, bathroom, type, laundri

FREX: footag, durat, sqft, refriger, oven, type, rang, key, dispos, microwav, featur, year, forc, polici, storm

Lift: aleah, batiz, birchwood, ecobe, edt, flatiron, fpl, fulcrum, jiang, labri, leong, leteishia, lui, mensen, demartini

Score: durat, footag, type, oven, refriger, descript, featur, polici, key, rang, forc, dispos, sqft, storm, year

Topic 11 Top Words:

Highest Prob: build, amazon, loft, unit, space, score, needl, modern, center, citi, ceil, histor, walk, block, window

FREX: museum, foundat, slu, melinda, arena, landmark, loft, boutiqu, amazon, histor, needl, expos, industri, virginia, trendi

Lift: abuzz, agn, aia, aum, biom, bonnevill, detour, disney, fireman, flexi, foremost, herringbon, johnston, kress, kundig

Score: amazon, needl, build, loft, slu, score, histor, museum, foundat, unit, rooftop, facebook, melinda, walkscor, urban

Topic 12 Top Words:

Highest Prob: garag, bath, floor, car, bedroom, level, master, townhom, townhous, attach, privat, main, fireplac, open, gas

FREX: townhous, townhom, attach, level, garag, car, main, half, master, tandem, upper, town, second, powder, third

Lift: husrt, nothgat, facilitiesmicrowavebalconi, townhous, fairlak, shaun, bridgewat, lavizzo, dunhil, mae, holden, fay, between, geozon, sedan

Score: townhous, townhom, garag, master, attach, car, level, bath, main, floor, fireplac, gas, upstair, suit, upper

Topic 13 Top Words:

Highest Prob: new, remodel, floor, paint, brand, newli, carpet, updat, applianc, renov, kitchen, throughout, light, fresh, complet

FREX: new, newli, remodel, brand, paint, fresh, renov, fixtur, recent, updat, quartz, carpet, complet, throughout, upgrad

Lift: rot, handprint, allot, hyper, smartcar, suki, grout, quatz, jennieu, johnlscott, eatur, lon, refig, stencil, trimwork

Score: new, remodel, brand, paint, newli, renov, quartz, updat, carpet, fixtur, fresh, applianc, cabinet, throughout, recent

Topic 14 Top Words:

Highest Prob: school, high, district, elementari, middl, locat, hill, great, hous, famili, close, home, communiti, lake, win

FREX: newport, elementari, school, northshor, finn, tyee, chinook, inglewood, factoria, somerset, woodridg, middl, tahoma, district, win

Lift: alcott, apna, camwest, cavalero, christa, colina, elemetari, eton, goddard, hardword, jnr, kicthen, laurelcrest, lochmoor, schola

Score: school, elementari, district, newport, high, middl, northshor, award, win, factoria, cul, somerset, sac, tyee, lake

Topic 15 Top Words:

Highest Prob: window, door, heat, wall, doubl, effici, ceil, built, cover, sink, electr, floor, kitchen, microwav, energi

FREX: blind, wall, effici, insul, shelv, mini, energi, sink, doubl, ceram, mirror, glass, door, heater, panel

Lift: charmingâ, gaylord, gibb, pewter, refr, sidebar, cozyâ, edgmon, saka, indirect, stonewar, fenwick, shoulder, compressor, ceo

Score: energi, wall, effici, blind, window, doubl, sink, heat, pane, door, insul, glass, tile, mini, electr

Topic 16 Top Words:

Highest Prob: home, light, warm, even, also, cozi, perfect, summer, love, day, winter, keep, window, natur, morn

FREX: winter, warm, morn, nearest, keep, climat, cold, futur, even, poplarstreet, usa, dinner, pond, warmth, weather

Lift: blum, bromin, crate, dappl, gpf, label, longest, niagara, stcr, shortest, aair, althroughout, canning, chelan, fhs

Score: climat, warm, nearest, winter, poplarstreet, keep, morn, usa, pond, shortest, mile, even, longest, home, althroughout

Topic 17 Top Words:

Highest Prob: home, room, master, famili, bath, fireplac, beauti, larg, bonus, gas, upstair, dine, suit, bedroom, car

FREX: formal, bonus, piec, upstair, master, famili, pantri, island, soak, downstair, suit, room, butler, jet, den

Lift: bonusroom, handscrap, schilter, bedroo, canterwood, miniatur, saki, summerlin, tongu, pruitt, roadsid, bdba, hdwds, mas, butler

Score: master, home, room, formal, upstair, famili, bonus, fenc, suit, piec, yard, fireplac, gas, island, downstair

Topic 18 Top Words:

Highest Prob: home, right, left, list, onto, turn, exit, inform, ave, agent, rent, estat, direct, take, real

FREX: left, onto, exit, turn, broker, ago, deem, andrea, hors, sale, agent, discriminatori, buy, reliabl, program

Lift: ayub, cochran, crisi, deceiv, dol, edgemont, everroad, gasca, heartland, idaho, lockbox, locust, mapquest, mildew, nurtur

Score: left, turn, onto, exit, agent, broker, list, estat, program, buy, purchas, right, licens, discriminatori, deem

Topic 19 Top Words:

Highest Prob: applic, must, screen, incom, rental, credit, tenant, accept, histori, requir, report, time, landlord, reusabl, verifi

FREX: report, incom, reusabl, evict, histori, comprehens, verifi, gross, stub, must, record, prior, crimin, convict, criteria

Lift: comprehes, icx, liafw, longev, madeson, melia, misstat, mitig, moorman, nxnw, schiess, smartphon, spousal, teriann, procedur

Score: incom, histori, report, reusabl, applic, accept, evict, screen, must, credit, verifi, crimin, comprehens, rental, tenant

Topic 20 Top Words:

Highest Prob: view, deck, floor, citi, balconi, bay, top, amaz, sound, den, mountain, amen, gas, space, roof

FREX: tower, elliott, skylin, stun, concierg, penthous, spectacular, ashley, sweep, roof, incred, elliot, panoram, deckwash, bay

Lift: linehigh, availableconcierg, availablemicrowavebalconi, availableroof, brenden, builtsmart, conditioninghigh, deckclub, deckwash, disposalhardwood, disposalviewhardwood, enso, floorsair, floorscarpetfireplacestoragewalk, floorscarpetfireplacestorageyardmodern

Score: view, concierg, rooftop, needl, mountain, elliott, tower, sound, penthous, puget, balconi, skylin, olymp, bay, roof

Topic 21 Top Words:

Highest Prob: view, lake, min, beach, washington, sound, waterfront, deck, water, mountain, ferri, minut, puget, boat, rainier

FREX: min, ferri, beach, shipyard, boo, launch, boat, dock, kayak, peek, psns, marina, manett, lake, sea

Lift: charleston, indianola, multimillion, seagul, tug, burlesqu, elb, mel, pettus, rainner, soldier, solon, uss, kingfish, pepsi

Score: lake, min, beach, view, ferri, waterfront, washington, sound, mountain, puget, boat, olymp, rainier, sunset, dock

Topic 22 Top Words:

Highest Prob: coffe, nearbi, woodland, cafe, zoo, bakeri, pizza, includ, bar, breweri, bike, market, food, park, pub

FREX: glen, pizza, blake, burger, roanok, bever, mexican, boston, marketim, para, breweri, closest, lighthous, deli, zoo

Lift: almacenamiento, alquil, balcã, carrot, centro, cheshiahud, depã, durant, espacio, habitacion, jalisco, llamar, lugar, moroccan, nuevo

Score: woodland, zoo, glen, pizza, bakeri, coffe, burger, para, breweri, cafe, por, nearbi, pub, vita, una

Topic 23 Top Words:

Highest Prob: hous, home, yard, floor, hardwood, room, basement, backyard, fenc, charm, famili, lot, back, deck, street

FREX: craftsman, basement, driveway, backyard, unfinish, furnac, bungalow, detach, workshop, charm, yard, porch, attic, daylight, matur

Lift: venita, availab, fplce, hrdwood, varnish, vienna, waal, bunglow, houis, recroom, winterwood, inlay, pcms, schafer, bellewood

Score: basement, yard, fenc, hous, backyard, craftsman, charm, home, driveway, famili, hardwood, rambler, unfinish, furnac, bungalow

Topic 24 Top Words:

Highest Prob: fee, month, deposit, pet, per, refund, applic, rent, non, secur, person, move, addit, adult, util

FREX: refund, per, non, fee, hold, person, approv, deposit, adult, nonrefund, max, applic, month, pet, standard

Lift: nrf, wheeler, asses, clifthous, bayvista, asgard, astro, aujane, oriana, machell, mulkiteo, rochell, elmer, guyer, dolor

Score: refund, fee, per, deposit, applic, non, month, pet, adult, person, rent, secur, approv, hold, move

Topic 25 Top Words:

Highest Prob: includ, condo, water, park, rent, secur, sewer, garbag, unit, month, electr, leas, washer, dryer, spot

FREX: condo, sewer, garbag, spot, water, trash, hoa, secur, assign, electr, sewag, includ, condominium, pay, owner

Lift: abella, groovi, kirkton, makeup, nema, incess, quickcheck, applainc, coa, downton, palazzo, salesforc, miscellan, housebroken, equipt

Score: condo, sewer, garbag, water, unit, secur, includ, electr, spot, condominium, rent, hoa, park, assign, month

Topic 26 Top Words:

Highest Prob: home, pet, center, apart, fit, communiti, hour, amen, restrict, call, hous, offic, pool, featur, washer

FREX: breed, restrict, patrol, weight, select, friday, monday, hour, tan, clubhous, sparkl, bull, spa, courtesi, chow

Lift: admn, ambercrest, crossbre, ctl, greensview, heightsbearcreek, miramont, onlyâ, pig, piranha, rentplus, ridgellc, rivercroft, sonora, tarantula

Score: pool, apart, clubhous, hour, restrict, breed, fit, center, select, weight, spa, opportun, polici, home, communiti

Topic 27 Top Words:

Highest Prob: enjoy, set, like, privat, quiet, feel, home, beauti, just, well, make, surround, love, place, green

FREX: set, peac, feel, yet, like, seren, green, forest, nestl, seclud, surround, tuck, privaci, oasi, enjoy

Lift: pup, yael, chirp, confluenc, northridg, eden, splurg, celtic, yet, peac, northup, renal, seren, tuck, cherish

Score: feel, set, peac, enjoy, like, quiet, tree, seren, yet, green, surround, forest, privaci, privat, tranquil

Topic 28 Top Words:

Highest Prob: studio, build, apart, light, floor, walk, locat, block, charm, hardwood, just, rail, laundri, vintag, great

FREX: studio, vintag, classic, summit, brick, rail, charm, volunt, world, build, origin, old, howel, harvard, sccc

Lift: alanna, annaron, ansonia, antoinett, avlb, camara, clarwood, corona, coryel, delicaci, delmont, dino, franconia, harborvi, irv

Score: studio, apart, build, vintag, charm, block, rail, site, classic, light, brick, hardwood, walk, origin, laundri

Topic 29 Top Words:

Highest Prob: case, deposit, basi, month, consid, fee, pet, applic, home, leas, term, secur, addit, rent, appli

FREX: case, basi, lakeshor, consid, wpm, stage, burden, sec, app, word, spoken, dawnett, eas, potenti, ref

Lift: bimonth, samara, showmojo, burden, roundabout, augment, bannist, capor, caraway, contemp, franki, harman, lakeshor, limerick, perrinvill

Score: case, basi, consid, wpm, fee, deposit, lakeshor, applic, dawnett, spoken, appli, adult, burden, month, dir

Topic 30 Top Words:

Highest Prob: mile, block, minut, bus, hous, mall, hospit, colleg, away, garden, safeway, etc, stop, close, less

FREX: hospit, safeway, westwood, meyer, colleg, express, mile, children, trader, veget, walmart, joe, mall, walgreen, fred

Lift: belvider, goug, huos, kellog, medgar, thomson, twelfth, bigdaddi, cinnebarr, kremama, plume, ipic, smoothtop, marko, fil

Score: mile, block, hospit, safeway, mall, bus, trader, colleg, joe, meyer, fred, minut, hous, children, qfc

Topic 31 Top Words:

Highest Prob: pool, communiti, court, fireplac, wood, park, burn, outdoor, cover, unit, hot, tub, bath, swim, center

FREX: court, tenni, pool, swim, burn, sauna, sport, cabana, hot, basketbal, indoor, complex, clubhous, ground, exercis

Lift: caller, closetmaid, demeanor, larkspur, latasha, madera, pao, pratik, sixtyon, sunn, veridian, woodlak, carlisl, racquet, eastbridg

Score: pool, court, swim, condo, sauna, clubhous, tenni, burn, communiti, basketbal, cabana, fireplac, tub, hot, complex

Topic 32 Top Words:

Highest Prob: will, leas, can, look, move, pleas, rent, take, current, month, need, time, apart, interest, get

FREX: someon, subleas, leav, sinc, renew, know, send, roommat, messag, abl, asap, transfer, current, interest, answer

Lift: brittain, girlfriend, nsf, through, trap, acknowledg, born, chemistri, complic, florida, god, polit, raineer, sociabl, tenement

Score: will, someon, apart, leas, look, can, current, take, move, takeov, get, interest, know, subleas, thank

Topic 33 Top Words:

Highest Prob: home, today, apart, call, price, come, bedroom, move, tour, offer, avail, one, will, chang, communiti

FREX: chang, price, today, subject, special, wait, miss, thursday, hurri, tuesday, wednesday, don, monday, friday, come

Lift: altia, envious, hayworth, ibarra, infolimestonecourt, knack, knick, obstacl, ondin, sizzl, solana, andant, apartmetn, bushi, cesarston

Score: today, apart, home, chang, tour, price, call, subject, monday, thursday, opportun, special, come, wednesday, move

Topic 34 Top Words:

Highest Prob: furnish, fulli, bed, includ, furnitur, term, tabl, month, avail, short, need, size, queen, internet, chair

FREX: furnish, queen, mattress, furnitur, sofa, chair, couch, utensil, dresser, towel, tabl, twin, inclus, apodmentâ, stock

Lift: aft, agon, apodmentâ, armchair, barsala, berth, bevmo, chromecast, crockeri, custodi, dinghi, dungeo, firestick, flatwar, gel

Score: furnish, furnitur, queen, chair, tabl, couch, fulli, sofa, wifi, mattress, bed, towel, utensil, internet, dresser

Topic 35 Top Words:

Highest Prob: shop, restaur, wng, locat, distanc, park, bus, within, store, groceri, line, great, street, market, conveni

FREX: distanc, wng, within, restaur, store, groceri, shop, librari, farmer, junction, theater, movi, market, metro, line

Lift: flexibil, twnhse, cline, lowman, junaita, ldb, bagel, coff, hillwood, jersey, locatio, burbank, harborwood, refurnish, distanc

Score: distanc, wng, restaur, groceri, store, within, shop, bus, locat, market, line, park, coffe, farmer, librari

Topic 36 Top Words:

Highest Prob: contact, show, info, call, click, pleas, bedroom, bath, text, inform, manag, schedul, see, leas, email

FREX: click, show, info, contact, macpherson, cell, inc, call, cynthia, waâ, hardin, bob, tag, text, eldridg

Lift: dewar, erni, twyla, frimer, devonyâ, maesta, mangement, burney, dwtn, ruben, northfieldproperti, dortico, crowther, moniqu, townsquar

Score: info, contact, show, click, call, pleas, macpherson, manag, text, inform, schedul, cell, websit, inc, see

Topic 37 Top Words:

Highest Prob: month, rent, hous, first, last, requir, deposit, leas, check, tenant, credit, pleas, avail, background, year

FREX: last, background, check, requir, first, credit, minimum, damag, due, respons, sign, disturb, renter, hous, tenant

Lift: lakepoint, prescreen, tinney, youâ, automobil, halpin, monika, summeri, yummi, rotweil, backroad, ywca, gasolin, lanlord, zeke

Score: credit, background, hous, requir, last, check, month, tenant, rent, damag, deposit, first, due, minimum, leas

Topic 38 Top Words:

Highest Prob: stainless, applianc, granit, steel, floor, counter, top, tile, countertop, kitchen, hardwood, featur, beauti, walk, custom

FREX: granit, stainless, steel, counter, slab, custom, countertop, tile, applianc, cherri, marbl, gourmet, top, luxuri, gorgeous

Lift: countetop, rta, opendoor, buildium, feephotosmoreview, propertypow, winslow, garbur, pga, bartop, bravia, brava, variat, brazilian, travertin

Score: stainless, granit, steel, counter, applianc, countertop, tile, custom, top, slab, luxuri, floor, hardwood, cabinet, cherri

Topic 39 Top Words:

Highest Prob: minut, park, walk, trail, microsoft, commut, easi, bike, locat, campus, center, ride, just, access, drive

FREX: burk, trail, gilman, microsoft, campus, googl, ride, hike, minut, marymoor, commut, shuttl, jog, gillman, connector

Lift: guenther, ooba, queit, sensordri, tooba, truesteam, valor, walkscoredotcom, trb, lionsgat, burk, samena, silicon, goldsmith, gilman

Score: microsoft, trail, campus, minut, googl, bike, burk, ride, amazon, gilman, walk, commut, marymoor, park, drive

Topic 40 Top Words:

Highest Prob: locat, park, great, bedroom, avail, includ, conveni, beauti, easi, access, floor, dryer, close, washer, open

FREX: locat, great, conveni, beauti, easi, open, close, dryer, well, access, avail, bedroom, washer, park, appoint

Lift: locat, conveni, bbq, great, easi, well, excel, ampl, beauti, awesom, quick, appoint, close, outsid, just

Score: locat, great, park, conveni, bedroom, easi, beauti, avail, dryer, close, washer, access, includ, open, privat

Appendix III: Neighborhood Typology

This paper uses a novel neighborhood typology based on racial proportions, which is developed building on Crowder, Pais and South's (2012) work. This typology labels each neighborhood based on the proportions of the top three racial groups present. While previous typologies used only one or two racial groups, our typology incorporates proportions of ethnic Latinx people and non-hispanic White, Black, Asian, and Native Americans and includes 28 neighborhood types. To form these types, we begin by noting the three racial or ethnic groups with the highest proportion in a tract. If the largest group is non-hispanic White and the White proportion is over .8, then the neighborhood is labeled 'predominantly white'. If the largest group is not White and makes up more than 50 percent of the tract population, then the tract is labeled as predominantly that group. If no group makes up more than 40 percent of the tract proportion, the tract is marked as 'mixed'. If the first group is greater than 40 percent and the second group is greater than 10 percent then the tract is labeled with the name of the first and second groups in that order. If the first group is greater than 40 percent and no other group is greater than 10 percent, the label is the first group and then the word 'mixed'. Empty tracts are so labeled. By far the most common label is 'predominantly white', 37,207 2010 tracts with 1980 counts received this label. There are more than 6,000 tracts labeled 'white black' and 'white hispanic' each, and around 2,000 tracts labeled 'black white' and 'predominantly black' each, with around 1,600 more labeled 'latinx white'. No other labels have more than 1000 tracts. A full table of the counts of neighborhood types can be found in the appendix.

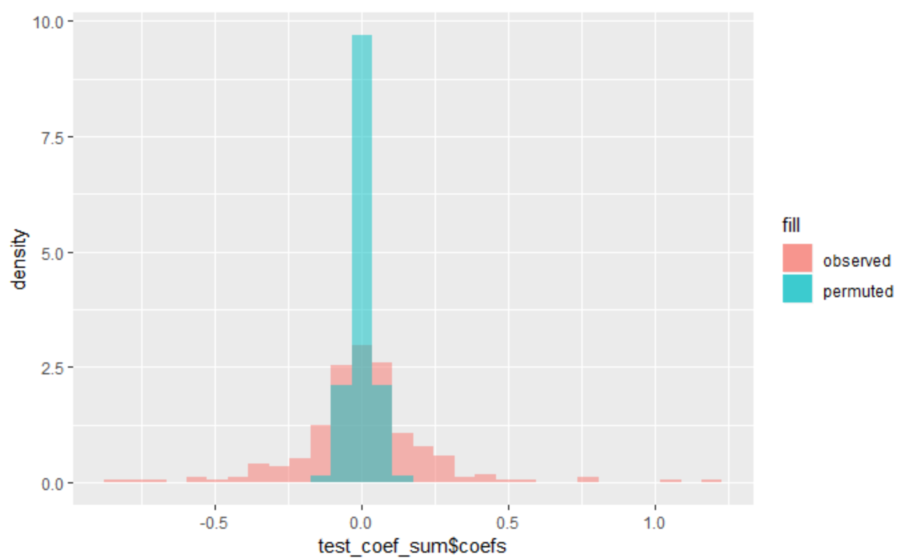
This method maintains the advantage of Crowder, Pais, and South's (2012) work in that it produces a typology that does not depend on metropolitan or country context yet conveys information about the neighborhood. As described below, this neighborhood typology is particularly useful for producing meaningful counterfactual scenarios. However, we include it as a dependent variable for instructive purposes.

Appendix IV: Robustness

Permutation Test:

Results comparing the observed effects with the distribution of 500 permutations randomizing the topics

The plot below shows a comparison of coefficient size in the observed relationships (reported above) and those produced by a permutation test. To produce the permuted coefficients, we ran simulated regressions where the vectors of topic proportions (produced by the STM routine in our analysis) were assigned randomly to census tracts, instead of being associated through the address of the listing. The visualization below shows the results of 10,000 simulations and the observed coefficients.



We can see that coefficients larger than 0.35 would be very unlikely if the associations were random (in fact, no coefficient with an absolute value that larger than 0.39 appeared in the permutation test). Most of the associations discussed in the paper are of that magnitude or greater. This suggests that the observed associations are not a result of random noise in the text or census data.

ldaRobust Topic Robustness Test: