Racialized Ideology and Structure in Online Rental Advertisements: Using Natural Language Processing to Uncover Racialized Discourse

Ian Kennedy

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Kyle Crowder, Chair

Alexes Harris

Nathalie Williams

Program Authorized to Offer Degree:

Sociology

University of Washington

Abstract

Racialized Ideology and Structure in Online Rental Advertisements:

Using Natural Language Processing to Uncover Racialized Discourse

Ian Kennedy

Chair of the Supervisory Committee:

Kyle Crowder

Department of Sociology

This dissertation analyzes how the discursive representation of racist ideology in rental ads link up with the racialized social system in the US, a system I define, following Bonilla Silva (1997) as a "societ[y] in which economic, political, social, and ideological levels are partially structured by the placement of actors in racial categories or race." Throughout this dissertation, I develop a connection between discourse and housing outcomes as likely implicated in processes that reproduce segregation and gentrification from three viewpoints. First, in Chapter 1 I examine a particular aspect of the advertisements: how they describe neighborhood quality. I use text analysis and regression techniques to see how those descriptions vary in neighborhoods with different racial compositions and different socioeconomic status. I find that advertisements in White and Asian neighborhoods describe those places more positively than advertisements in

Black or Latinx neighborhoods. I also investigate the ways that text writers in some Black neighborhoods might resist the trend of devaluing Black neighborhoods. In Chapter 2 I look at a second, somewhat broader area of discourse: exclusionary language. I investigate how it varies across race and socioeconomic class, and add consideration of gentrifying areas. I find that some kinds of discursive exclusion, like background and credit checks, are more common in Black neighborhoods, while others, like strong tenant expectation, are more common in White neighborhoods. Finally, in Chapter 3, I return to the discursive themes from the first two chapters, but consider their variation across the sixteen metropolitan areas in my study. I find that the racialization of discourse about neighborhood quality is consistent across the whole study area. I also find that, in general, some cities have more exclusionary discourse than others, and that that variation can be partially explained by metropolitan-level variables. For instance, cities with lower levels of segregation tend to have more language about background and credit checks.

Acknowledgements

This dissertation would not have happened without the advice, support, and generous feedback from so many people. Kyle Crowder, in addition to serving as Dissertation Committee Chair, has mentored me, and this project, from the day I walked into his office to ask if he thought it might be possible. Alexes Harris supported and guided me through graduate school in many ways, including as mentor, committee member, co-author, and incredible example of how people can make change from academic positions. As a committee member, Nathalie Williams kept me on track with excellent feedback and support. John Wilkerson, a gracious Graduate Student Representative, supported my growth as a computational text analyst and often pushed me towards new ways of thinking about the project.

I also had the support of other superlative faculty members beyond my committee. Jaklyn Hwang corresponded with me very generously to help me refine my operationalization of gentrification. Maria Krysan met with me more than once to discuss design issues when I doubted myself. Mahesh Somashekhar helped me think more about these results in the context of each city, and reminded me of the power keyword analysis can have. This dissertation uses data from the UW's National Rent project. Chris Hess literally made this project possible by helping to start the project and organizing the Craigslist Scraping process that we still use now. Sarah Chasins developed the scraping tool, Helena, that the project uses and has been extremely generous in supporting our, perhaps niche, feature requests. Alex Ramillier has made sure that our scraping processes keep running, and has kept up his work even after moving on from UW.

The project would be unable to scrape anything without the computer infrastructure supported by the Center for Studies in Demography and Ecology at UW. What that really means

is that I am indebted to Matthew Wetherford for keeping our servers running, updated, and helping us when we ran out of space (again).

The Bayesian models I use were only possible because of a statistics study group with Breon Haskett and Mark Igra. Without them, modeling the test-set data would have been much harder, and not as transparent and robust. I also got excellent feedback on my models from Connor Gilroy and Joe Bak-Coleman, each of them willing to have long Slack chains about setting priors or using non-centered parameterizations. Four people shared their time and brainpower to read near-final versions of these chapters. Patricia Taylor, my mother, read the first chapter. My partner, Quynh Tran, read the second chapter. My friends and colleagues, Mark Igra and Bron Haskett read the third chapter and the introduction, respectively. I am in their debt, a debt I can likely only really repay to Mark and Breon. Angelina Eimannsberger repeatedly assured me that I could do this. Turns out she was right as usual.

I was able to benefit from presenting early (and late) versions of this work. The UW Sociology Seminar and the CSDE seminar both graciously invited me to present chapters. The Context Working Group, led by Kyle Crowder, saw very raw versions of these analyses and helped me see what bits were important. Finally, this dissertation would have taken perhaps years more time to write without the support of the Eunice Kennedy Shriver National Institute of Child Health and Human Development training grant T32 HD101442-01.

Obviously there are scores of others who supported me and helped me keep going through the inevitable difficulties of dissertating. The earth herself literally supported my every step, not least by allowing me to work and live in the shadow and protection of Tahoma, the mother of waters. Everything that is good about my work derives from all of that help and support. The errors, typos, misunderstandings, lost chances, and other mistakes are my own

Introduction

Threads and Fabric

Sociology is usually more focused on thinking about describing the world with theories than with metaphors. Theories, though, are just less-artful metaphors, and sometimes good metaphors make good theories. For instance, a central motivation of this dissertation comes from Cathy Cohen (1997: 437), who writes

> Many of us continue to search for a new political direction and agenda, one that does not focus on integration into dominant structures but instead seeks to transform the basic fabric and hierarchies that allow systems of oppression to persist and operate efficiently.

This fabric is a telling metaphor: cloth is both complex, made of varied threads, and unitary, a single object. Systems of oppression can be similar: showing up as powerful unified objects which cause harm and reproduce oppression, but they are constructed through threads of norms, action, discourse, and ideology. One of the additions Cohen makes here is to connect fabric to hierarchical systems. That juxtaposition is particularly generative for me: hierarchies, including race and class hierarchies, rely on the particular way that social life is woven together–not just on dominance in seats of power and access to capital. Maintaining power means both controlling threads and controlling how they're woven.

In this dissertation, I apply sustained academic attention to a particular intersection in that basic fabric: looking at how the ideologies present in the text of rental advertisements on the Craigslist platform relate to the racialized and classed structure of housing outcomes. In other words, I analyze how the discursive representation of racist ideology in rental ads link up with the racialized social system in the US, a system I define, following Bonilla Silva (1997) as a "societ[y] in which economic, political, social, and ideological levels are partially structured by

the placement of actors in racial categories or race." Throughout this dissertation, I develop a connection between discourse and housing outcomes as likely implicated in processes that reproduce segregation and gentrification. Due to the descriptive nature of this research, I abstain from making claims about the breadth and depth of this connection. However, I believe that the circumstantial evidence is quite strong: rental ad discourse appears to be at the right place at the right time for contributing to the reproduction of racialized housing outcomes.

My grander hopes for this project, though, are based on the idea that the shape of this intersection of ideology and structure will help me and others understand the shape of other such intersections. That type of intersection–discursive ideology with structure–could be essential to the warp and weft of the fabric that supports systems of oppression, and so understanding those crossings of discourse and action, of ideology and structure, could lead to new ways of resisting those systems, rebuilding, and maybe eventually reweaving them.

The rest of this introduction touches on some of the theoretical ground which undergirds the dissertation. I explore various views of ideology, connected Marxist, Neomarxist, and Black Feminist accounts. Then I review how others have theorized the link, or the articulation, between ideology and structure in racial terms. Here I link cultural studies to contemporary sociological work. Once I have some of that groundwork laid, I examine my choice of Craigslist as a platform, and 16 metropolitan areas as a study area. Finally, I briefly introduce each chapter of the dissertation.

Complementary Views of Ideology

Ideology, as used within this dissertation (usually paired with the word 'racist') can be interpreted by its common definition as a set of beliefs–in this case the set of beliefs that falsely justify racist views–without missing out on any of the empirical findings. However, I take this

opportunity to explore how my use aligns not only with that common definition, but also usefully with Marxist, neo-Marxist, and contemporary critical-race conceptions of ideology. I consider these different views not as challenges to each other, but perhaps complementary.

Marx uses the word 'ideology' to refer to the power the ruling class holds over the ideas and beliefs of people in all classes (Marx and Engles 1965 [1845]). In that account, ideology particularly refers to those ideas that are not rooted in material reality: ideology presents a twisted version of reality. While people are equal and the working class has the ability to gain control of the means of production in truth, they are alienated from that fact, in part by ideology. However, ideology is only a superstructural mirage, not part of the base truth of material reality. Althusser (2006) breaks with this conception, arguing that rather than simply being a distorted form of ideas that justifies power, that ideology is essential to the reproduction of power relations. Ideology, he argues, sets the rules of the game and defines what appears as 'self-evident.' This kind of ideology defines what is possible, and instead of being something that can be thrust aside in favor of focus on material reality, must be a site of struggle because it is the site of the creation of social reality.

Both kinds of ideology are relevant when we consider contemporary racial ideology. In one sense, racist and white supremacist ideology is clearly of the traditional Marxist type: it falsely asserts a racial hierarchy as material truth. At the same time, though, it is an ideology in Althusser's sense: racist ideology structures how US Americans see the world in ways that they do not realize, consciously, as being racial. Racial conceptions of space change how people perceive disorder (Sampson and Raudenbush 2004), whether they want to live in a place (Krysan et al. 2009), or whether they think polluting the area is appropriate (Bonam et al. 2016).

Of course the ideology that I study is not freely available to me here, but mediated through the language of rental advertisements. I am interested in the aspects of that discourse that are colored by racist ideology, aspects I call racialized discourse, defined as writing or talk that has racial meaning, regardless of whether that meaning is explicit (Kennedy et al. 2021). Broadly, racialized discourse includes stereotypes and 'controlling images' (Collins 2002), and also the subtle ways that (especially) White people discuss race without naming it, what Bonilla-Silva (2003) calls 'colorblind racial ideology'. Seamster (2015) explicitly calls for more attention to the role of colorblind racial ideology in urban regimes.

That process mirrors a key aspect of Bonilla-Silva's (2003) analysis—what he calls the 'naturalization frame' of colorblind racism—which "allows whites to explain away racial phenomena by suggesting they are natural occurrences. For example—whites can claim 'segregation' is natural because people from all backgrounds gravitate toward likeness." (Bonilla-Silva 2003: 28). Bonilla-Silva highlights that this naturalization is easier in a context where institutional powers perpetuate the racial order: if policies are not explicitly racial but they produce differential outcomes, White people can believe that those outcomes are evidence of natural group differences, not the outcomes of racial structure. In other words, racialized discourse in housing advertisements could reflect a process of naturalization in which racial structure is encoded in discourse, but its racialized aspects are obfuscated.

While naturalized racialized discourse might not be explicit, it is also not unrecoverable. Bonilla-Silva exposes it in his interviews by attending to the ways his interlocutors discuss, or avoid discussing, race. He uses four frames to discuss their strategies: abstract liberalism—attributing racial differences to individual choices within an equal system; naturalization—asserting that racial differences are natural, and perhaps biological; cultural

racism—the idea that racial groups have different cultural values which explain differences in outcome; and the minimization of racism—framing racism as a past event, no longer relevant to people's life chances. In the context of this dissertation, the naturalization frame is most relevant. Insofar as advertisements present a racialized picture of neighborhood quality without explicit mention of race, they also naturalize that racialization. To understand the consequences of such an ideology, I examine theories that link structure and ideology.

Racial Structure and Ideology

Theories of articulation in cultural studies (Hall 1996[1980]), the sociology of race (Bonilla-Silva 1997), and Black Feminist thought (Weheliye 2014) emphasize that 'structure'—here broadly used to refer to the ordering of society along hierarchical racial lines—supposes a myriad of working processes linked together and moving in concert. In a restatement of Cohen's (1997) fabric metaphor, society is structured hierarchically based on race and this structure both supports and is supported by racial ideology (Collins 2002, Bonilla-Silva 2003, Golash-Boza 2016). The goal of this project is to investigate a particular moment where racism as a social structure is caught in the act of articulation with racial ideology. While there are many such moments, most of them, "do not escape concealment under the brush of discourse" (Spillers 1987: 67).

Racial ideology in rental ads, like much of contemporary color-blind racial ideology in the United States, is concealed under the brush of discourse. However, by combining natural language processing and demographic methods, I can remove part of that covering, showing what aspects of that discourse is racialized. Because rental advertisements exist in large numbers, and can be geographically linked to segregation as a piece of the racialized social system in the United States, they offer a particularly rich ground for understanding this articulation.

This project targets the articulation between racist ideology and racial structure in the hopes of understanding ways that articulation can be interrupted and eventually broken. The current housing situation in the United States was created by structural-legal systems which, by preventing Black families from living outside of certain neighborhoods and preventing loans for homes within Black neighborhoods, sapped Black communities of wealth (Oliver and Shapiro 1997) and allowed new White neighborhoods to flourish. These legally sanctioned housing policies were linked to the war on drugs, mass incarceration, and over-policing, and together contributed to creating a racialized hierarchy of housing, including certain neighborhoods that are both overwhelmingly Black and very poor, and others that are overwhelmingly rich and White. These are central to the structural and unequal harms and boons of our racialized system of housing, and demographers have been central to mapping their impact.

Historically, the politicians, bankers and bureaucrats who designed and implemented the structures of segregation justified their action with explicitly racist motivations, powered by an explicitly racist ideology (Rothstein 2017). Lawyers and activists resisted these policies, and their efforts yielded political and ideological successes. Politically, the Fair Housing Act offered legal protections for homeseekers, and culturally explicit racism became less common. However, racial structures and ideologies did not fall, but changed their form. Ideologically, racist language became less obvious, but remained pervasive, as colorblind racial ideology. Politicians implementing the war on drugs only sometimes relied on explicit mentions of race, turning instead to poverty and joblessness in Black neighborhoods—caused in part by the structure—as reasons for their implementation. They depicted—and their contemporary counterparts still depict—those neighborhoods as blighted, dangerous, and in need of revitalization, or worthy of simple destruction. This ideology, though now often ‘colorblind’ is but a new iteration of old

racism, and it has permeated housing dynamics. Korver-Glen (2018) shows how real estate professionals and loan officers readily brandish racialized understandings of housing and neighborhoods. Sociological studies of housing are only now taking account of the import of this ideology, largely in the study of neighborhood perceptions. This history of housing structure and ideology reflects Golash-Boza's (2016) framework for the mutual reinforcing of racial ideology and structure.

Housing Dynamics as Racialized Structure

Explicitly racist policies like redlining and housing covenants linked race and neighborhood quality through much of the 20th century (Rothstein 2017). Redlining marked less White and especially Black neighborhoods as 'hazardous' and essentially prevented people from using federally insured mortgages to buy homes there. Housing covenants, clauses included in property deeds that restricted possession to White people only, were often required in new, larger developments, essentially restricting non-White people to living in places with aging, crowded housing stock. Both processes were based on the racist assumption that neighborhoods where many Black people lived were of low quality, but also actively devalued those places, reproducing results that confirm the assumption. In that way, legal processes created a material reality of racially structured space.

Critical race theorist Richard Ford theorized how race and space were interwoven, a process he called 'racially identified political space' (Ford 1994). He argued that racial meanings clung to spaces, and that those meanings could be self-reinforcing: policies based on the stereotype of White neighborhoods as clean and safe might act to ensure those, and perhaps only those, neighborhoods remained clean and safe, disregarding non-White and especially Black neighborhoods. The result of such policies would not only be the perpetuation of material

differences in neighborhood quality depending on the race of the people who live there, but a naturalization of that difference as not being about race at all.

Empirical evidence about housing dynamics supports Ford's theoretical work, and affirms that people apply racist ideas to neighborhoods based on who they think lives there. Bonam (2018) shows that participants seeing a black family in a home cause them to devalue the neighborhood, or make them more likely to support allowing pollution in that neighborhood (Bonam 2016). Krysan et al. (2009) show that seeing Black people in videos of real Detroit and Chicago neighborhoods caused White participants to consider those neighborhoods less desirable. A similar effect seems to influence home appraisals, where racialized neighborhood perception influences racial differences in home appraisal value (Howell and Korver-Glenn 2018). In these cases racial ideas about space exist beyond material differences, effectively contributing, rather than simply reflecting or perpetuating, existing material differences.

Legal attempts to define racialized space, like redlining, and racialized perceptions of space, as revealed by social-psychological experiments, are both part of what McKittrick (2006) calls an "ongoing spatial project" (McKittrick 2006: 121) of both race and gender domination. Wynter (1990; 2001) explicitly connects the process of racialization of people and of spaces, parallel otherings which work to exclude some spaces and bodies from social life. Non-White spaces generally, but especially Black Female space have their meaning shaped in an ideological struggle, where dominant forces leverage powerfull tools like the census, imminent domain, and scholarly knowledge-production to control the ownership, status, and perception of space.

Craigslist as a Site

Craigslist is a particularly good research site for examining rental advertisement discourse. It is well known, and broadly used–the ranking company SimilarWeb ranks Craigslist

as the 35th most visited website in the US as of April 2022 (SimilarWeb 2022). The platform's affordances are also conducive to discourse analysis. While other rental platforms limit how landlords or property managers frame their rental, Craigslist allows people to write their ads however they like. The title, text, neighborhood, and address field are all open text fields. Text pulled from the advertisements, therefore is created by the posters themselves, rather than edited or written by the platform. That makes it native data about how those people think about their units, their marketing, and the surrounding neighborhood.

Craigslist is active in all fifty states and in hundreds of cities in the United States. However, I selected sixteen metropolitan areas as the study area for this dissertation. I chose sixteen cities from various regions, of various sizes, and with variation on metro-level Segregation. Sixteen cities were enough to get that variation, but few enough that I could build on-the-ground partnerships in as many areas as possible before starting work. I reached out to partners in all sixteen metro areas in 2021, and asked what kinds of questions they were struggling with, and if they were interested in my project on racialized discourse. I was able to share data and analysis based on their requests, like running an analysis of gentrification for a partner in Rhode Island, or helping the city of New Orleans create a census of recently rented units to help with tenants rights work. I introduce the dataset in more detail in the chapters below.

In the three chapters that follow, I examine the question of the articulation between racial ideology and housing structure from three viewpoints.

First, in Chapter 1 I examine a particular aspect of the advertisements: how they describe neighborhood quality. I use text analysis and regression techniques to see how those descriptions vary in neighborhoods with different racial compositions and different socioeconomic status. I

find that advertisements in White and Asian neighborhoods describe those places more positively than advertisements in Black or Latinx neighborhoods. I also investigate the ways that text writers in some Black neighborhoods might resist the trend of devaluing Black neighborhoods.

In Chapter 2 I look at a second, somewhat broader area of discourse: exclusionary language. I investigate how it varies across race and socioeconomic class, and add consideration of gentrifying areas. I find that some kinds of discursive exclusion, like background and credit checks, are more common in Black neighborhoods, while others, like strong tenant expectation, are more common in White neighborhoods.

Finally, in Chapter 3, I return to the discursive themes from the first two chapters, but consider their variation across the sixteen metropolitan areas in my study. I find that the racialization of discourse about neighborhood quality is consistent across the whole study area. I also find that, in general, some cities have more exclusionary discourse than others, and that that variation can be partially explained by metropolitan-level variables. For instance, cities with lower levels of segregation tend to have more language about background and credit checks.

References

Althusser, Louis. "Ideology and ideological state apparatuses (notes towards an investigation)." The anthropology of the state: A reader 9, no. 1 (2006): 86-98.

Anderson, Carol. 2016. *White rage: The unspoken truth of our racial divide*. Bloomsbury Publishing USA.

Bonam, Courtney, Caitlyn Yantis, and Valerie Jones Taylor. 2018. “Invisible Middle-Class Black Space: Asymmetrical Person and Space Stereotyping at the Race–Class Nexus.” Group Processes & Intergroup Relations 136843021878418. doi: 10.1177/1368430218784189.

Bonam, Courtney M., Hilary B. Bergsieker, and Jennifer L. Eberhardt. 2016. "Polluting Black Space." Journal of Experimental Psychology: General 145, no. 11.

Bonilla-Silva, Eduardo. 2003. Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in the United States. Rowman & Littlefield Publishers.

Bonilla-Silva, Eduardo. 1997. "Rethinking racism: Toward a structural interpretation." American sociological review: 465-480.

Cohen, Cathy J. 1997. Punks, Bulldaggers, and Welfare Queens: The Radical Potential of Queer Politics? GLQ 1 May; 3 (4): 437–465. doi: https://doi.org/10.1215/10642684-3-4-437

Collins, Patricia Hill. 2002. Black feminist thought: Knowledge, consciousness, and the politics of empowerment. Routledge.

Ford, Richard T. 1994. "The boundaries of race: political geography in legal analysis." The legal geographies reader: law, power, and space, Oxford: Blackwell 87104

Golash-Boza, Tanya. 2016. "A critical and comprehensive sociological theory of race and racism." Sociology of race and ethnicity 2, no. 2: 129-141.

Hall, Stuart. 1996. “Race, Articulation, and Societies Structured in Dominance.” in Black British Cultural Studies: A Reader, edited by Houston Baker, Manthia Diawara, and Ruth Lindeborg. Chicago: University of Chicago Press.

Howell, Junia, and Elizabeth Korver-Glenn. 2018. "Neighborhoods, race, and the twenty-first-century housing appraisal industry." Sociology of Race and Ethnicity 4, no. 4: 473-490.’

Kennedy, Ian, Chris Hess, Amandalynne Paullada, and Sarah Chasins,. 2021. “Racialized Discourse in Seattle Rental Ad Texts.” Social Forces 99(4) doi: 10.1093/sf/soaa075.

Korver-Glenn, Elizabeth. 2018. “Compounding Inequalities: How Racial Stereotypes and Discrimination Accumulate across the Stages of Housing Exchange.” American Sociological Review 83(4):627–56. doi: 10.1177/0003122418781774.

Krysan, Maria, Mick P. Couper, Reynolds Farley, and Tyrone A. Forman. 2009. “Does Race Matter in Neighborhood Preferences? Results from a Video Experiment.” American Journal of Sociology 115(2):527–59. doi: 10.1086/599248.

Marx, Karl, and Friedrich Engels. 1965 [1845]. "The German Ideology." Google Books.

McKittrick, Katherine. 2006. Demonic grounds: Black women and the cartographies of struggle. U of Minnesota Press.

Oliver, Melvin L., and Thomas M. Shapiro. 2006. Black Wealth, White Wealth: A New Perspective on Racial Inequality. Taylor & Francis.

Rothstein, Richard. 2017. The Color of Law: A Forgotten History of How Our Government Segregated America. New York: Liveright Publishing.

Sampson, Robert J., and Stephen W. Raudenbush. 2004. "Seeing disorder: Neighborhood stigma and the social construction of “broken windows”." Social psychology quarterly 67, no. 4: 319-342.

Seamster, Louise. 2015 "The white city: Race and urban politics." Sociology Compass 9, no. 12: 1049-1065.

Similarweb. 2022. Top Website Ranking. https://www.similarweb.com/top-websites/united-states/

Spillers, Hortense J. 1987. "Mama's baby, papa's maybe: An American grammar book." diacritics 17, no. 2: 65-81.

Weheliye, Alexander G. 2014. Habeas Viscus: Racializing Assemblages, Biopolitics, and Black Feminist Theories of the Human. Durham, NC: Duke University Press.

Wynter, Sylvia. "Beyond Miranda's Meanings: Un/silencingthe'Demonic Ground'of Caliban's' Woman,'." Out of the KumbIa: Woman and Caribbean Literature (1990): 355-72.

Wynter, Sylvia. 2001. "Towards the Sociogenic Principle: Fanon, Identity, the Puzzle of Conscious Experience, and What It Is Like to Be ‘Black.’." National identities and sociopolitical changes in Latin America: 30-66.

Chapter 1: Racialized Representations of Neighborhood Quality Across 16 US Metro Areas

Abstract

Housing dynamics in the United States have historically been racialized through explicitly discriminatory laws like redlining, linking neighborhood quality to racial composition. The persistence of segregation and gentrification suggests that neighborhood race is still entwined with perceptions of neighborhood quality. This study investigates how the connection between neighborhood race and perceived quality persists in present day rental advertisements. Using 500k listings from the online rental platform Craigslist spanning 16 U.S. metropolitan areas I apply computational text analysis to show that contemporary descriptions of neighborhoods online reflect the legacy of more traditional forms of neighborhood racialization. For a homeseeker, looking at advertisements on Craigslist does not give them an objective, neutral, or even only slightly skewed picture of the neighborhoods they might search, but instead presents a view that distorts neighborhood quality based on the race of neighborhood residents. Advertisements tend to describe White neighborhoods more positively than other neighborhoods: Black tracts with a median household income of $150,000 are described about as favorably as Asian and White tracts with a fraction of the median income. These findings have implications for understanding the perceptual nature of the reproduction of racialized housing dynamics.

Housing segregation by race has been a well documented source of disparities in life chances by race in the United States (Massey and Denton 1993, Charles 2003). Foundational to contemporary segregation are legal policies, like redlining, which explicitly devalued non-White neighborhoods generally and Black neighborhoods specifically. Federally subsidized loans were unavailable in those neighborhoods, based on the claim that they were bad investments (Rothstien 2017). In other words, legal policy implemented segregation in part by constructing a racialized hierarchy of neighborhood quality. Yet there is strong consensus that the forces that perpetuate segregation have become increasingly subtle—shifting from explicit forms of exclusionary discrimination, like redlining, to practices that are highly racialized but cloaked in colorblind racial ideology (Krysan and Crowder 2017, Bonilla-Silva 2003).

One of these subtle forces is the durability of racialized assumptions about neighborhood quality. By focusing on discursive variations in the description of neighborhoods in differentially racialized neighborhoods, I use Craigslist advertisements to examine the extent to which online rental advertisements present a racialized hierarchy of neighborhood quality. Those discursive variations are important because they at least reflect and likely also contribute to and reify the racialization of space. Following Ford (1994) and McKittrick (2006), spaces are racialized because past policy and contemporary practice has influenced the way people think about certain neighborhoods. As landlords and other text writers describe the places their units are in, they could include those racialized perceptions in their advertisements. They might do that because they think that such a description would help them attract the best possible tenant, or simply as honest descriptions of how they perceive those places, with no thought to race or economic motive.

For a homeseeker, looking at advertisements on Craigslist does not seem to give them an objective, neutral, or even only slightly skewed picture of the neighborhoods they might search, but instead presents a view that distorts neighborhood quality based on the way the neighborhood has been racialized. I find that Black neighborhoods are described less positively than other neighborhoods, and that Asian and White neighborhoods are described most positively, even accounting for economic and housing stock factors. Specifically, relatively high-income Black tracts are described no more favorably than majority White tracts with a fraction of the median income. These findings have implications for understanding the perceptual nature of the reproduction of racialized housing dynamics specifically, and for the social processes that drive inequality generally.

Background

To understand the role of Craigslist advertisements in perpetuating racialized ideas about neighborhood quality, I draw on theories of racialized space, colorblind racism, and emerging work on online rental markets.

Racialized Space

In the United States, space is racialized through material differences between places where people of different races live, in the perceptions that people hold about those differently-raced spaces, and through the reciprocal relationship between that material structure and perception. One aspect of racialization is how people think about and experience neighborhoods in terms of how nice of a place they are to live.

Explicitly racist policies like redlining and housing covenants linked race and neighborhood quality through much of the 20th century (Rothstein 2017). Redlining marked less

White and especially Black neighborhoods as ‘hazardous’ and essentially prevented people from using federally insured mortgages to buy homes there. Housing covenants, clauses included in property deeds that restricted possession to White people only, were often required in new, larger developments, essentially restricting non-White people to living in places with aging, crowded housing stock (Massey and Denton 1993). Both processes were based on the racist assumption that neighborhoods where many Black people lived were of low quality, but also actively devalued those places, producing results that confirm the assumption. In that way, legal processes created a material reality of racially structured space.

Critical race theorist Richard Ford theorized how race and space were interwoven, a process he called ‘racially identified political space’ (Ford 1994). He argued that racial meanings clung to spaces, and that those meanings could be self-reinforcing: policies based on the stereotype of White neighborhoods as clean and safe might act to ensure those, and perhaps only those, neighborhoods remained clean and safe, disregarding non-White and especially Black neighborhoods. As a result, the policies would also perpetuate material differences in neighborhood quality depending on the race of the people who live there, but a naturalization of that difference would make it seem as though those differences are not about race at all.

Empirical evidence about housing dynamics supports Ford’s theoretical work, and affirms that people apply racist ideas to neighborhoods based on who they think lives there. Bonam (2018) shows that participants seeing a Black family in a home cause them to devalue the neighborhood, or make them more likely to support allowing pollution in that neighborhood (Bonam 2016). Krysan et al. (2009) showed that seeing Black people in videos of real Detroit and Chicago neighborhoods caused White participants to consider those neighborhoods less desirable. A similar effect seems to influence home appraisals, where racialized neighborhood

perception influences racial differences in home appraisal value (Howell and Korver-Glenn 2018). In these cases racial ideas about space exist beyond material differences, effectively contributing to, rather than simply reflecting or perpetuating, existing material differences.

Legal attempts to define racialized space, like redlining, and racialized perceptions of space, as revealed by social-psychological experiments, are both part of what McKittrick (2006) calls an "ongoing spatial project" (McKittrick 2006: 121) of both race and gender domination. Wynter (1990; 2001) explicitly connects the process of racialization of people and of spaces, parallel otherings which work to exclude some spaces and bodies from social life. Non-White spaces generally, but especially Black Female space have their meaning shaped in an ideological struggle, where dominant forces leverage powerful tools like the census, eminent domain, and scholarly knowledge-production to control the ownership, status, and perception of space. More general discourse, as in rental advertisements, would also be subject to the same ideological struggle.

Racialized Discourse

The racialized meanings of space can be exposed, recorded, or shared in discourse. I define racialized discourse, broadly, as writing or talk that has racial meaning, regardless of whether that meaning is explicit (Kennedy et al. 2021). Broadly, racialized discourse includes stereotypes and 'controlling images' (Collins 2002), and also the subtle ways that (especially) White people discuss race without naming it. Bonilla-Silva (2003) calls it 'colorblind racial ideology', that is "a loosely organized set of ideas, phrases, and stores that help whites justify contemporary white supremacy" (Bonilla-Silva 2003: 262)–based on the falsehood that race no longer matters in the United States. In the context of rental advertisements, this means focusing on language in the advertisements which is not explicitly racial, but is statistically associated

with neighborhoods with different racial compositions. The development–and detection–of racialized language in public contexts like rental advertisements is subtle and more about variations in the descriptions of everyday things, like neighborhoods and nearby amenities, than in explicit racial exclusion.

That process mirrors a key aspect of Bonilla-Silva's (2003) analysis—what he calls the 'naturalization frame' of colorblind racism—which "allows whites to explain away racial phenomena by suggesting they are natural occurrences for example—whites can claim 'segregation' is natural because people from all backgrounds gravitate toward likeness." (Bonilla-Silva 2003: 28). Like Ford, Bonilla-Silva highlights that this naturalization is easier in a context where institutional powers perpetuate the racial order: if policies are not explicitly racial but they produce differential outcomes, White people can believe that those outcomes are evidence of group differences, not the outcomes of racial structure. In other words, racialized discourse in housing advertisements could reflect a process of naturalization in which racial structure is encoded in discourse, but its racialized aspects are obfuscated.

Resistance

As Patricia Hill Collins argues, "Western social theories place far more emphasis on explaining social order than they do in explaining political resistance, let alone aspiring to generate it." (Collins 2019: 117). Collins highlights the fact that people oppressed by the social order develop strategies and methods for living under, persisting, and resisting that oppression. McKittrick (2006) and Wynter (2001) draw attention to the way that marginalized spaces can grow into alternative sites of power and resistance, with special emphasis on the way Black women and subaltern subjects contest dominant spatial structures. One of the key realms Wynter

(2006) identifies where this struggle takes place is the discursive-ideological: where what matters is the way people think, talk, and write about space and race.

In the context of Craigslist advertisements and racialized discussions of neighborhood quality, studying only the hegemonic ideology would leave out important aspects of how racialized discourse shows up to text readers, and could lead to reifying that discourse as simply a reflection of neighborhood material differences. Highlighting existing practices of resistance also challenges the often implicit deficit frame of Black neighborhoods as only lacking (Hunter and Robinson 2016). Instead, this paper carefully examines different ways to describe neighborhoods, with special focus on how those ways might be racialized differently.

Broader segregative processes

Perceptions of neighborhood quality have clear repercussions for housing dynamics, especially racialized processes like gentrification and segregation. In a world where many housing searches take place online, considering how those online rental markets represent neighborhoods is essential. Kyrsan and Crowder's (2017) Social Structural Sorting Perspective (SSSP) on the reproduction of segregation emphasizes that homeseekers focus on housing units within a set of neighborhoods they consider appropriate. Homeseekers exclude neighborhoods that seem unwelcoming, or appear to have low quality. Perceptions of neighborhood quality are racially-driven (Bonam 2018, Sampson and Raudenbush 2004), reflecting and reifying the racialization of those spaces. White homeseekers may be excluding less-White neighborhoods based on a false assumption that the neighborhood is low-quality, in combination with or separate from racist attitudes. Non-White homeseekers may avoid an area because they have experienced or heard about discrimination or racism occuring there (Krysan and Crowder 2016). In the case of gentrification, spurious associations between neighborhood quality and race may

make gentrifiers define their neighborhoods differently from established boundaries to try and set apart their new high-quality area from older, perceptually more dangerous, and often less-White areas (Hwang 2016). It could also cause them to build social networks that exclude long-time residents (Walton 2021).

These changes come in the context of 21st century cities where the meanings of neighborhoods, including who should and should not live, work, and eat in them, is the subject of popular and political conflict. In the first decades of this century, while the share and number of White urban residents declined slightly, the wealth of those White residents and their influence in the affairs of their cities has not (Seamster 2015). Understanding those dynamics means making explicit comparisons between neighborhoods with different racial and economic characteristics, not focusing on non-White places and an imagined, implicitly White, referent (Howell 2019).

Online real estate platforms include rich data about real-world rental markets, including the asking rent, the address, details about the units for rent, and an often detailed textual description of the living space and its surrounding neighborhood. Since each listing can be located within a census tract, we can examine the distribution of information about rentals, estimate small area asking rents, or analyze how the language used in the advertisements varies across space. Existing work using that kind of data has shown that information about neighborhoods and housing units is unequally distributed: poorer and less-White areas tend to have both fewer advertisements (Boeing 2020, Hess et al. 2021), and less information in each advertisement, especially about the neighborhood (Bresibs et al. 2021). Kennedy et al. (2021) focused on the racialized content of advertisements, but their analysis was limited to listings from Seattle and had no clear findings on neighborhood quality.

Online markets have outsized potential to reproduce racially disparate outcomes. While Craigslist has anti-discrimination policies, discrimination persists (Reosti 2018, Hogan and Berry 2011). Online markets are built by mostly White men, who bring significant bias to their work (Noble 2019; Benjamin 2019). Discrimination gets baked into digital tools-even or especially when people's motives are 'primarily economic,' or even when they are motivated to help (Eubanks 2018). Platforms have a responsibility not only to push against explicitly discriminatory actions by individuals, but also to prevent 'digital redlining' (Cottom 2016; Gillard and Culik 2016) and to ensure that the way places are described on their platforms does not reproduce racial discrimination.

This paper looks explicitly at the racialization of descriptions of neighborhood quality by investigating three research questions. First, how do descriptions of neighborhood quality in online rental advertisements vary depending on how the neighborhood is racialized? Second, how do those descriptions change when we consider neighborhood racialization and socio-economic status together? And finally, if those descriptions tend to reflect the racial ideology of anti-Blackness that's prevalent in the United States, how do people resist that ideology? I address the first two questions by regressing a measure of positive neighborhood descriptions on tract-level data about neighborhood ethno-racial composition from the ACS 2019 5-year estimates, and covariates. To explore resistance to hegemonic ideology I use a combination of exploratory analysis on alternative ways to describe neighborhood quality and residual analysis to identify places where discourse is most different from what the models expect.

Data and Methods

This paper identifies how neighborhood quality is described differently in neighborhoods which are racialized differently. In collaboration with the University of Washington's National Rent Project, I collected a dataset of 2.5 million Craigslist advertisements from 16 metropolitan areas. I use computational text analysis to systematically describe these texts in terms of neighborhood quality and other factors, and regression analysis to examine how those descriptions vary with the racialization of the neighborhood where the ad is posted.

Scraped Data

The dataset I use contains 2.5 million rental advertisements scraped from the online platform Craigslist in 16 large metropolitan areas in the United States, collected between February 2017 and March 2021, using the web scraper Helena (Chasins et. al 2017) by UW's National Rent Project. I included all the advertisements that included addresses, rents, titles, unit descriptions, and lists of amenities. Because of a scraping issue, many listings are missing from August 2019 to October 2019. Of the original 2.5 million advertisements; about 2 million were dropped as being formal or near duplicates, missing a long enough listing text, missing rent, or not being geocodable to the tract level; this process is described in more detail in Appendix I. Almost all of the texts dropped were dropped during the deduplication process. Table 1 shows the number of listings retained after deduplication and pre-processing in each CBSA in the study area.

Table 1:

Metropolitan Area (CBSA)	Total Listings	Train Set Listings	Test Set Listings
New York-Newark-Jersey City, NY-NJ-PA	111,253	55,589	55,664
Seattle-Tacoma-Bellevue, WA	79,056	39,543	39,513
Los Angeles-Long Beach-Anaheim, CA	58,792	29,598	29,194
Boston-Cambridge-Newton, MA-NH	41,808	20,986	20,822
Chicago-Naperville-Elgin, IL-IN-WI	34,842	17,401	17,441
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	32,396	16,192	16,204
Urban Honolulu, HI	22,886	11,466	11,420
San Francisco-Oakland-Hayward, CA	22,792	11,355	11,437
New Orleans-Metairie, LA	18,121	9,105	9,016
Providence-Warwick, RI-MA	16,479	8,126	8,353
Riverside-San Bernardino-Ontario, CA	15,561	7,701	7,860
Dallas-Fort Worth-Arlington, TX	15,722	7,894	7,828
Atlanta-Sandy Springs-Alpharetta, GA	13,398	6,740	6,658
Charlotte-Concord-Gastonia, NC-SC	11,083	5,413	5,670
Worcester, MA-CT	7,263	3,612	3,651
Colorado Springs, CO	6,045	3,032	3,013
Total	507,497	253,753	253,744

Conceptualization and Racialization

Each advertisement is geolocated to the tract and joined with tract-level estimates from the 2019 ACS 5-year, including the racial composition, housing stock, economic characteristics, and commuting distance and time. In order to examine racialization in text, I measure the association between certain language and neighborhood racialization. I operationalize neighborhood racialization in a 5 level mutually exclusive typology. I label any tract that is majority White as White. For Latinx, Black, and Asian tracts, I use that label if that group is the largest ethnoracial group and if at least 40% of the tract population shares that ethno-racial identification according to the census. Tracts that I cannot classify as White, Latinx, Black, or Asian using those specifications I label as Other.[1]

[1] I have repeated the analysis with a more fine-grained typology and the results are broadly similar to those shown below. I intend to examine those results in future work.

Table 2: Summary Statistics (N = 507,506)

Covariate	mean	sd
Neighborhood Covariates		
Poverty Proportion	0.13	0.10
Log of tract Median Income	11.23	0.47
Share of Adults with College Degrees	0.20	0.10
Share Commuters by Car	0.33	0.14
Share With a Commute over 20 mins	0.35	0.10
Proportion of Owner-occupied Housing Units	0.46	0.25
Proportion of Rentals in Properties with >20 Units	0.19	0.22
Proportion of Units Built After 2010	0.03	0.07
Proportion of Units Single Family Detached Homes	0.38	0.39
Population Density	0.01	0.01
Listing Covariates		
Log of the Listing Rent	7.47	0.53
Number of Bedrooms	2.01	1.12
Neighborhood Type		
Black	0.09	0.31
Latinx	0.14	0.37
White	0.70	0.54
Other	0.08	0.29
Asian	0.06	0.25

STM and Tract-Level Associations

I use a Structural Topic Model (STM) (Roberts et al. 2014) to model topics associated with the neighborhood characteristics. STM is a variant of Latent Dirichlet Allocation (Blei 2003). Both methods assume that each document in the corpus is composed of K topics and that each topic is defined by a vector of probabilities, one for each word in the corpus's vocabulary.

Both methods iteratively attempt to fit Document-Topic proportions and Topic-Word probabilities to best fit the data. STM also estimates the covariance matrix between the topics and a set of covariates chosen by the analyst. By doing so, it can include the error of the topic modeling procedure in its regression estimates (Roberts et al. 2014).

After fitting the STM, but before estimating any regression models, I read the top 20 texts from each of the 50 topics and created a label for each topic.[2] I read each text closely with special attention to language about neighborhoods. I also took general notes about each topic after reading all of the texts from that topic. Using that qualitative analysis and the top terms created by STM's 'Frex' and 'score' methods, I produced a topic title for each of the 50 topics. Following (Egami et al. 2018), I aimed for titles that would minimize a reader's surprise when they compared a text with the title of the topic that most closely matched that text. When I refer to topics in the text below I use the format 'Topic Title (Topic Number)', so the 18th topic is 'Fees and Deposits (18)'.

To address the first two research questions, focused on discursive associations with neighborhood race, I use linear regression models fit using Baysian inference using STAN (Stan Development Team 2022) through the `rethinking` (McElreath 2020) and `cmdstan` (Stan Development Team 2022) packages in R. I present results from two model types below. The 'bivariate' model is a particular topic regressed only on the neighborhood ethno-racial typology. The 'multivariable' models regress a topic on that same neighborhood typology as well as a number of tract- and listing-level covariates. The tract-level covariates describe the socio-economic and housing stock characteristics of the neighborhood, while the listing-level

[2] All of the topic titles as well as some other info about the topics can be found in Appendix IV

covariates are the rent and number of bedrooms. Table 2 shows all of the covariates and their descriptive statistics.

The plots below show the expected value of a particular topic in an unseen advertisement from each of the neighborhood types, either holding all the other variables at their means and putting the expected value on the x-axis, or putting another covariate, like tract median income on the x-axis, expected topic proportion on the y-axis, and using color to plot the various neighborhood types. Put briefly, we can consider a topic to reflect racialized discourse if the contrast between the point estimates for differently racialized neighborhoods is large and significant.

Multi-Modal Validation

While STM results are a powerful tool for observing associations in text, they require careful validation through close reading and, if possible, quantitative assessment using keywords (Nelson, 2017). I validate my topic modeling results by using a train-test split, confirming that my topics of interest are not dependant on my choice of K=50, checking my coefficients using permutation tests, through close reading of texts, and by using keywords to triangulate my understanding of the topics.

There is a risk when using topic models that the associations observed are a result of overfitting the data. To avoid this, I use a train-test split method of out-of-sample testing. I divide my sample of texts into two equal subsets, a training set and a testing set. I fit the STM using only the training set, use the training set to explore the topics, and then use the test set for the quantitative analysis presented below.

Topic modeling results can be sensitive to the analyst's choice of the number of topics, or K. After testing models with different numbers of topics, but before conducting regression

analysis, I settled on a model with 50 topics. I based this choice on STM's metrics and qualitative assessments of models with as few as 12 topics and as many as 75. This assessment showed that 50 topics produced a model which covered all of the important themes I saw in models of any size, but also retained key themes in single topics. I further ensured that the topics I focus on below were not created only because of this choice of K. To do so, I followed Wilkerson and Casas's method (2017) and compared the word-topic vectors for my topics of interest to models fit with other choices of K. That showed that topics similar to topic 36 and 50 occurred with other choices of K. More details are in the Appendix II.

Finally, the topic labels and their interpretations are a product of my experience reading and analyzing Craigslist texts. I present one of the topics specifically as a measure of positive language surrounding neighborhoods, and that interpretation is well supported both by the high-matching texts in that topic and by the words most associated with that topic by the STM. Following Nelson (2017), I use keywords to confirm both that my interpretation of the topic is reasonable and that the analysis is robust to an alternate specification of discursive neighborhood quality. Details of that process are available in Appendix III.

Results

Descriptions of neighborhood quality vary significantly by neighborhood racial composition, such that White and Asian neighborhoods are more positively described than Black, Latinx, and other neighborhoods. However, there are aspects of the advertisements that contest that broad trend. Black neighborhoods are described as being as historic as White neighborhoods, though Asian neighborhoods are described as the least historic. Also, some

Black neighborhoods are described very positively, I examine some advertisements from those neighborhoods in detail and explore what might make those places different.

Neighborhood Discourse

Of the 50 topics estimated by the STM, ten of them focused on the neighborhood or the surrounding area rather than the features of units or the stipulations of the lease. All nine of those topics are shown in Table 3 along with the titles I gave them based on qualitative coding, and the top words the STM model. Six of the nine neighborhood topics concerned access to features like parks, beaches, schools, hospitals, highways, walking distances, and public transportation. The other three were ‘Historic Charm (36)’, ‘On-Site Community (41)’, and ‘Positive Adjectives: Location (50)’. While all three of these topics have interesting racialized patterns, ‘Positive Adjectives: Location (50)’ is the most explicit discussion of neighborhood quality. A similar table with details for all topics is shown in Appendix IV

Table 3: Neighborhood Topics

Topic Number	Topic Title	Top Words
Topic 4	Walkability	distance, stores, walking, banks, shops, grocery, restaurants, away, farmers, store, joe, markets, within, bakeries, pubs
Topic 6	Ocean Access	ocean, mahalo, uri, beachfront, kcc, ewa, barracks, ohana, surfboard, hpu, corps, canoe, oceanview, kunia, pupukea
Topic 13	Free Standing Homes	house, yard, home, family, back, single, fenced, backyard, neighborhood, front, quiet, nice, driveway, porch, bedrooms, rent, lot, large, hook, beautiful
Topic 23	Universities and Hospitals	drexel, delaware, ardmore, plymouth, wales, goshen, malvern, forge, abington, darby, villanova, brandywine, hup, radnor, hospital
Topic 27	Driving Distance	canyon, calabasas, saddleback, doheny, pedro, glassell, zaya, palmdale, rowland, danville, ortega, winnetka, springs, talega, freeway
Topic 36	Historic Charm	lined, original, ocf, tree, magazine, claw, restored, charm, character, victorian, clawfoot, antique, rowhouse, carriage, classic
Topic 41	On-Site Community	offer, comfort, furry, variety, convenience, deserve, residents, lifestyle, proud, choose, dip, invite, plans, experience, discover
Topic 44	Public Transportation Access	apartment, transportation, near, close, yasmine, located, keerah, conveniently, carollo, gesler, public, maintained, beautiful, bedroom, kept
Topic 47	Schools	elementary, clyde, proctor, finn, chinook, cougar, woodridge, tahoma, kingsgate, tyee, elem, rambler, wedgwood, cumming, school
Topic50	Positive Adjectives: Location	location, great, excellent, area, easy, terrific, convenient, open, access, value, prime, ideal, including, areas, friendly

Positive Adjectives

‘Positive Adjectives: Location (50)’ was almost completely focused on words like ‘great’ and ‘beautiful’ and ‘location’, and those words were in regard to the location, neighborhood, or

area around the unit. Another topic also focused on positive adjectives, which I called “Positive Adjectives: Unit (12)”. These two topics are compared in Table 4 which compares the top terms and some examples from both topics.

Table 4: Positive Adjectives Topic Comparison (emphasis added)

Positive Adjectives: Unit (12) won, see, right, summer, coming, perfect, fall, come, soon, awesome, make, gem, wow, miss, beat	Positive Adjectives: Location (50) location, great, excellent, area, easy, terrific, convenient, open, access, value, prime, ideal, including, areas, friendly
gorgeous 4 bedroom	**Great Location**
luxury kitchen is **perfect**	**prime area** of Downey,
beautifully remodeled	**Location, location location**

While 'Positive Adjectives: Unit (12)' includes words like 'perfect', 'awesome' and 'gem', it doesn't have any location-specific words among its top terms. On the other hand, 'Positive Adjectives: Location (50)' includes both very positive words like 'great', 'excellent', terrific', 'prime' and 'ideal', and location-specific terms like 'location' and 'area'. This is born out in examples where the adjectives in advertisements with high proportions of 'Positive Adjectives: Unit (12)' refer to the unit, as in 'beautifully remodeled', or to an aspect of the unit as in 'luxury kitchen is perfect.' In contrast, advertisements with a high proportion of 'Positive Adjectives: Location (50)' often connect positive adjectives with spatial terms like 'great location', 'prime area', or 'location, location, location.'

Regression Results

I regress 'Positive Adjectives: Location (50)' on the neighborhood typology based on the majority ethno-racial group in the tract where the unit was advertised.

Figure I.1

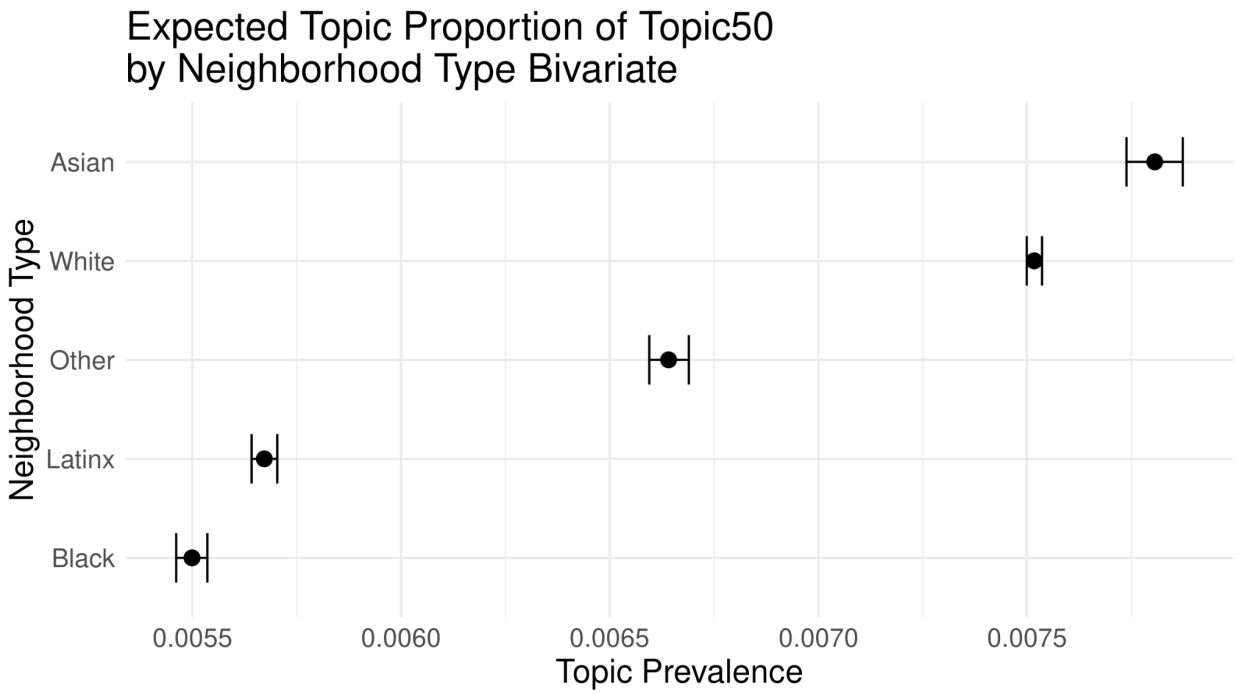


Figure I.1 Caption: Predicted prevalence of 'Positive Adjectives: Location (50)' on the X axis, neighborhood typology based on majority racial/ethnic group on the Y axis. This model only includes a regression of 'Positive Adjectives: Location (50)' on the neighborhood typology based on the ethno-racial majority group

In Figure I.1, values far to the right mean that the model predicts a lot of 'Positive Adjectives: Location (50)', or that it expects the neighborhood to be described positively. Values to the left mean the model expects less of that topic, that the neighborhood won't be described positively, or won't be described at all. These results are based on a bivariate association: just the 'Positive Adjectives: Location' topic regressed on the neighborhood typology with no other covariates. Asian and White neighborhoods have much higher expected values for this topic, while Black and Latinx neighborhoods have much lower expected values. In other words, Black and Latinx neighborhoods are systematically described with fewer positive adjectives than the

other three neighborhood types, and Asian neighborhoods are consistently described with more positive adjectives than the rest of the neighborhood types.

Figure I.1 is simply a descriptive plot, showing differences in descriptions of neighborhood quality depending on the racial composition of that neighborhood. The history of racialized housing means that race, economics, and housing stock hang together as part of a racialized social system (Bonilla-Silva 1997). This bivariate model does not try to disentangle those different aspects of structural racism in the United States, but rather measures their aggregate effects.

It is valuable, however, to understand how descriptions of neighborhood quality vary systematically with both neighborhood race and neighborhood socioeconomic characteristics. Doing so does not 'isolate the effect of race', but rather shows how racialized discourse might be distributed unevenly across neighborhoods that are similar in other ways. Practically, I investigate this by including socioeconomic and housing stock covariates in the regression model, which offers a different view of the processes that drive racialized differences in description of neighborhood quality. I also use an interaction between the neighborhood typology and the neighborhood income to see how race and socioeconomic status work together in this domain. These results are shown in Figure I.2.

Figure I.2

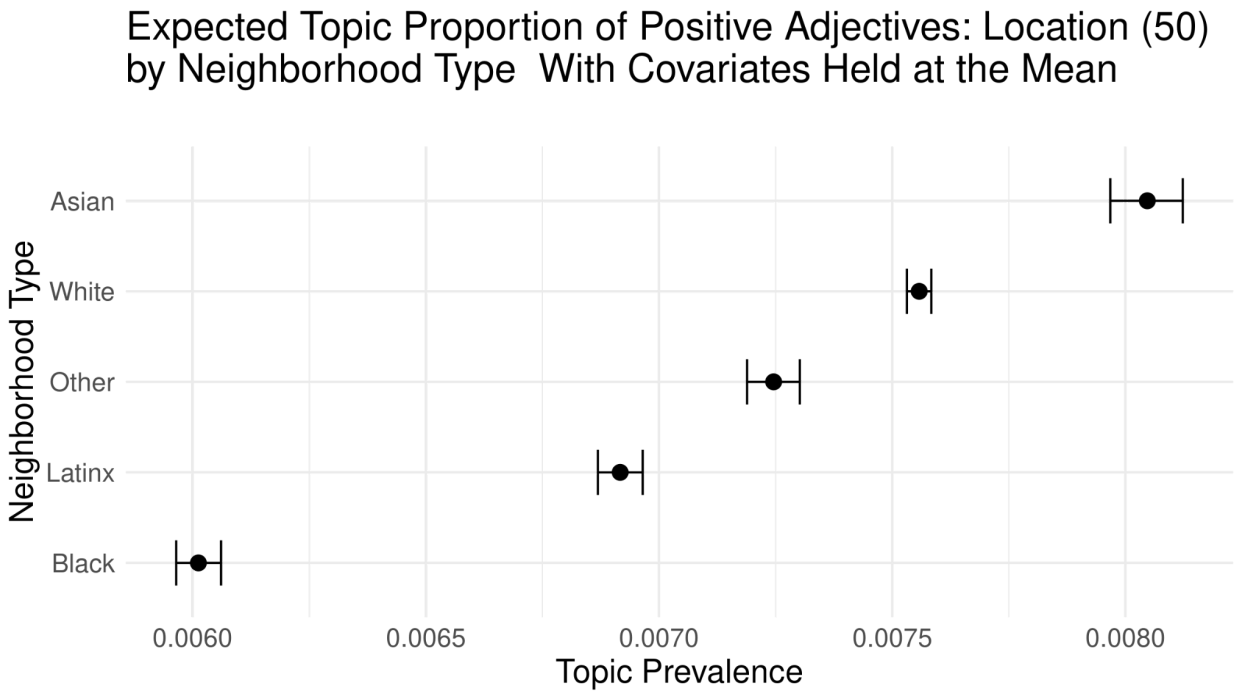


Figure I.2 Caption: Predicted prevalence of 'Positive Adjectives: Location (50)' on the X axis, neighborhood typology based on majority racial/ethnic group on the Y axis. This model includes a regression of 'Positive Adjectives: Location (50)' on the neighborhood typology based on the ethno-racial majority group, and all of the covariates shown in Table 2

With those other covariates and an interaction added, the results are broadly similar, though Latinx neighborhoods are now more similar to Other, White, and Asian neighborhoods in their expected value. That supports research on the complexity of how ethno-racialization and class connect in Latinx neighborhoods. Asian neighborhoods still have the highest expected proportion. And Black neighborhoods still have by far the lowest.

With or without controls, it seems that Craigslist rental advertisements systematically describe Black neighborhoods less positively than other neighborhoods. This aligns strongly with the critical race expectation that racism, and specifically anti-Blackness, is deeply embedded in the United States. This also connects to experimental work (Bonam et al. 2016;2018; Krysan and

Farley 2009) where participants devalued neighborhoods where they believed Black people lived. Finally, this reflects the history of redlining and legal segregation: those processes were devaluations of Black space encoded in law, but that devaluation has persisted in the way people describe Black space on Craigslist. This finding suggests that the racist judgements about neighborhood quality that drove redlining are still present in the way that landlords and other text writers describe places in Craigslist advertisements.

Interaction with Median Income

These results may suggest that racialized ideas about neighborhood quality are embedded in the neighborhood descriptions found in Craigslist home advertisements. Those analyses show that landlords and other text writers describe places as a ‘great location’ or a ‘prime area’ in part in response to the racial composition of the neighborhood. While my models control for socioeconomic factors, is it likely that part of the difference in prevalence of ‘Positive Adjectives: Location (50)’ still also reflects the way that the racialized social system in the United States ties race, place, and money together. Places that have residents with high incomes are also likely to be described positively, but that change may not happen the same way in neighborhoods with different racial composition. To investigate that possibility I run additional models which include an interaction between tract median income (logged in the model, but level in the y axis of the visualization) and neighborhood racialization.

Figure I.3

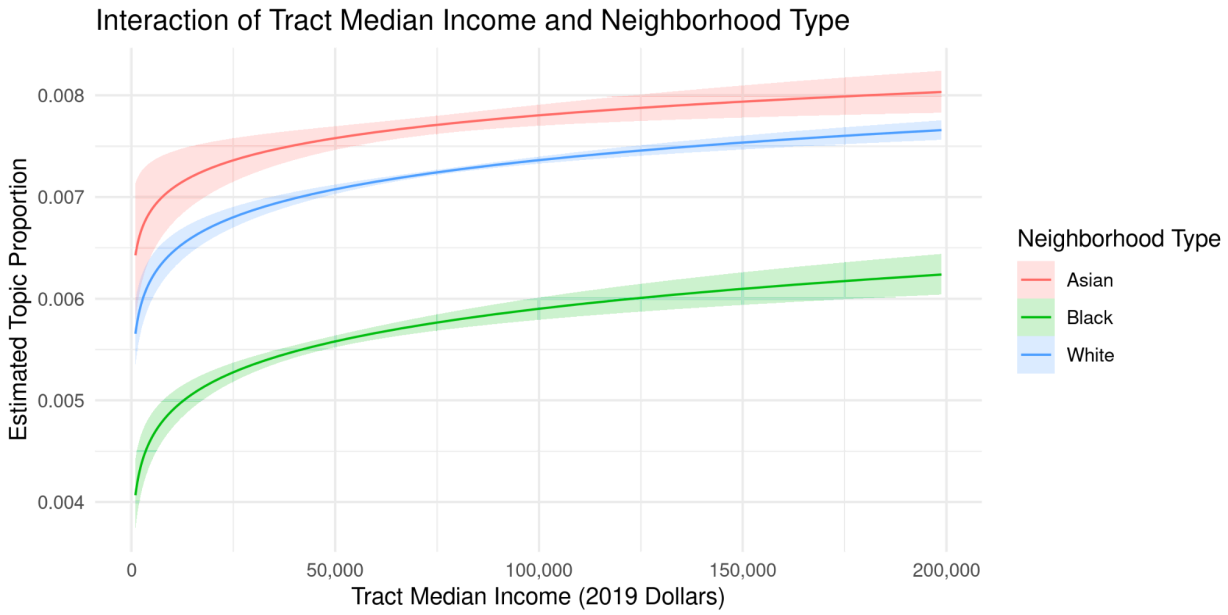


Figure I.3 Caption: This figure plots the race-income interaction, with tract median household income on the X axis, and the estimated proportion of 'Positive Adjective: Location' on the Y axis. This model includes a regression of 'Positive Adjectives: Location (50)' on the neighborhood typology based on the ethno-racial majority group,all of the covariates shown in Table 2, and an interaction between the neighborhood typology and the tract median income. The red, blue, and green lines represent expectations for Asian, White, and Black neighborhoods respectively. The shading represents 95% confidence intervals.

The expected topic prevalence for 'Positive Adjectives: Location (50)' rises in all three neighborhood types as tract median income increases. That makes sense following from the idea that, on average, richer tracts would be more likely to be described positively than poorer tracts. That trend is slightly less pronounced for White tracts and slightly more for Asian tracts. At the same time, the expectation for Black neighborhoods is well below those for Asian and White neighborhoods at all values. In fact, the expectation for Black tracts is so much lower that the expected prevalence of 'Positive Adjectives: Location (50)' is lower in Black tracts with a median income of $150,000--the richest Black tracts from which we consistently observe listings--than in White or Asian tracts at any income level. Put another way, this model suggests

that relatively high-income Black tracts are described no more favorably than majority White tracts with very low median income.

The results so far suggest that, in answer to research questions 1 and 2, Black neighborhoods are described less positively than other neighborhoods, and that that relationship is maintained when including socioeconomic and housing characteristics of neighborhoods. However, following critical race theory's focus on not just describing dominant structures, but also investigating how they might be resisted, I turn now to an investigation of discursive trends in the advertisements that diverge from the anti-Blackness discussed so far.

Measuring Resistance

In order to address the third research question, focused on the ways that people who write listings might resist or diverge from hegemonic racial ideology in the advertisements, I implement two approaches. First, I take a broader look at neighborhood descriptions, focusing specifically on how the advertisements describe the historic characteristics of neighborhoods. I use similar regression methods for this part of the analysis.

Secondly, I look directly for resistance to the tendency for advertisements to describe Black neighborhoods less positively than other neighborhoods. To do so, I use residual analysis to identify those Black tracts that have much higher incidence of 'Positive Adjectives: Location (50)' than a model would expect. I fit a new OLS model which regresses the average prevalence of 'Positive Adjectives: Location (50)' in a tract on the same tract-level socio-economic and housing stock covariates used in the other models (shown in Table 2) for Black tracts only. I then identify the ten tracts with the highest positive residuals. I read all of the listings from those tracts and present both quantitative and qualitative results below.

While we have so far been interested primarily in the ‘Positive Adjectives: Location (50)’ topic, Table 3 shows ten topics related to the neighborhood. While none of the others are directly about neighborhood quality, one stands out as an alternative to thinking about recognizing the value of a place: ‘Historic Charm (36).’

I titled topic 36 ‘Historic Charm’ because the advertisements I read emphasized units that were in historic or vintage buildings and neighborhoods. ‘Historic Charm (36)’ is a topic that combines unit descriptions, using words like floors, hardwood, brick, and victorian, with neighborhood descriptions, indicated by words like street, block, blocks, and lined. These come together in the advertisements, as in this excerpt: “a unique brownstone living experience on a beautiful/ quiet tree lined block next to the Famous historic Striver's Row of Harlem.” This advertisement connects the tree-lined street to the broader context of a historic area of Harlem.

I used the same regression method (including covariates) to estimate the racialization of this topic, shown in Figure I.4.

Figure I.4

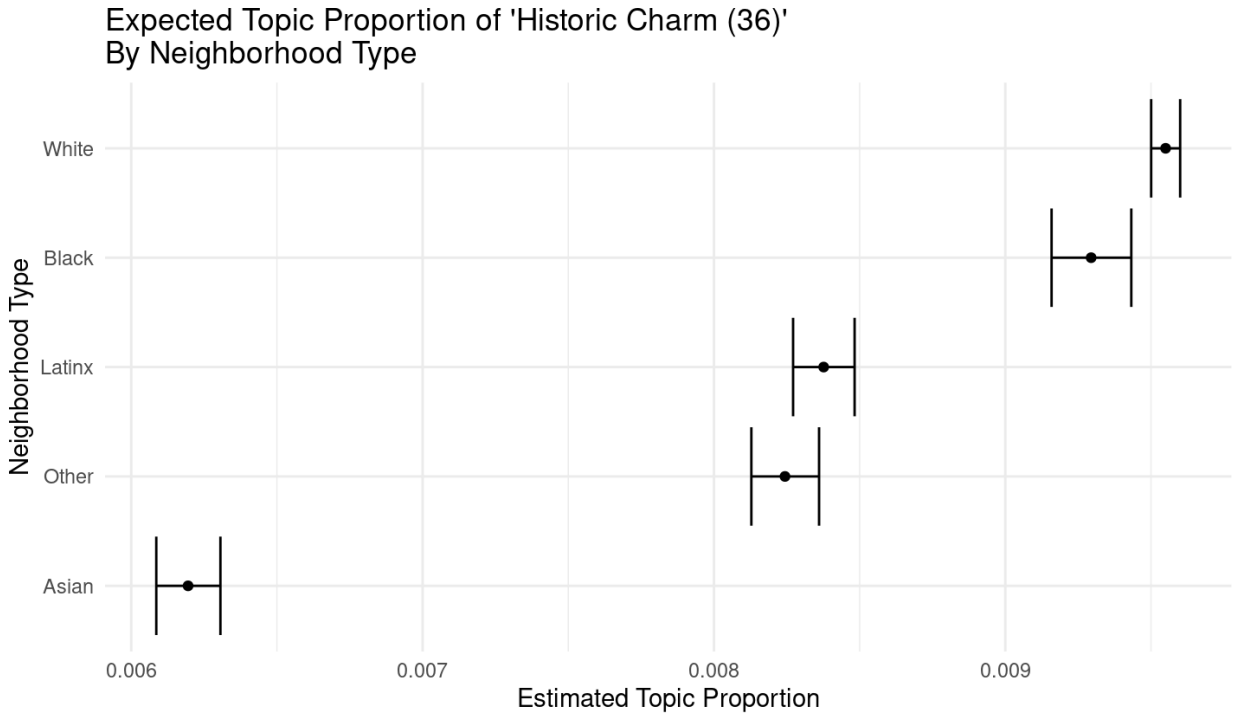


Figure I.4 Caption: Predicted prevalence of ‘Historic Charm (36)’ on the X axis, neighborhood typology based on majority racial/ethnic group on the Y axis. This model includes a regression of ‘Historic Charm (36)’ on the neighborhood typology based on the ethno-racial majority group, and all of the covariates shown in Table 2

In contrast to ‘Positive Adjectives: Location (50)’, the point estimate for the prevalence of ‘Historic Charm (36)’ is nearly as high in Black neighborhoods as in White neighborhoods.

Many Black neighborhoods in cities like New York, Atlanta, New Orleans, and Philadelphia have existed as long as those cities themselves, and are therefore historic. The Black communities in those cities are strong, and this ‘Historic Charm’ discourse could be a reflection of that strength. It is also true that segregation policies, like redlining, forced Black communities into areas with the oldest and worst housing stock. Now pre-war and other ‘historic’ architecture is desirable. So it is possible that these results reflect gentrification.

It's also notable that Asian neighborhoods have the lowest amount of this topic. This seems to align with the stereotype of Asian Americans as 'perpetual foreigners' (Lee 2008), but there's also a huge amount of diversity in the people labeled as Asian in the census.

Figure I.5

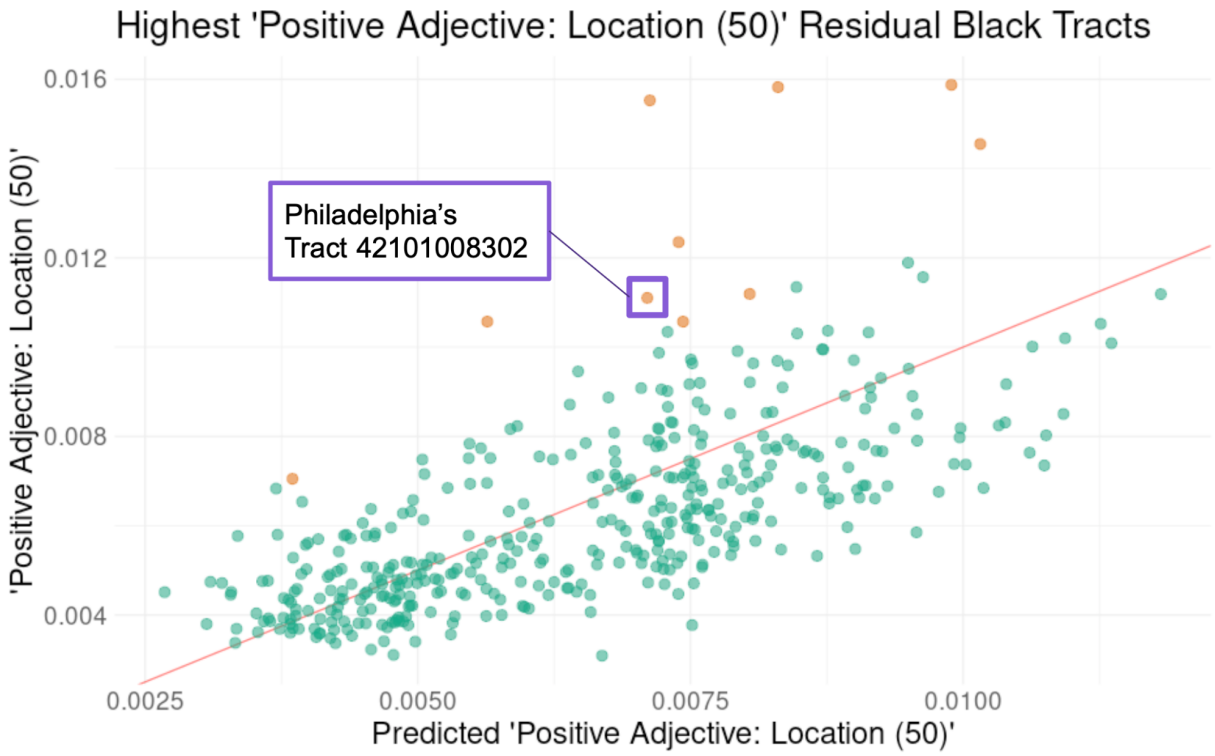


Figure I.5 Caption: Figure I.5 shows the predicted average topic proportions on the X axis and the observed average topic proportions on the Y axis for each Black tract in the sample. The orange tracts are the ten tracts with the largest positive residuals.

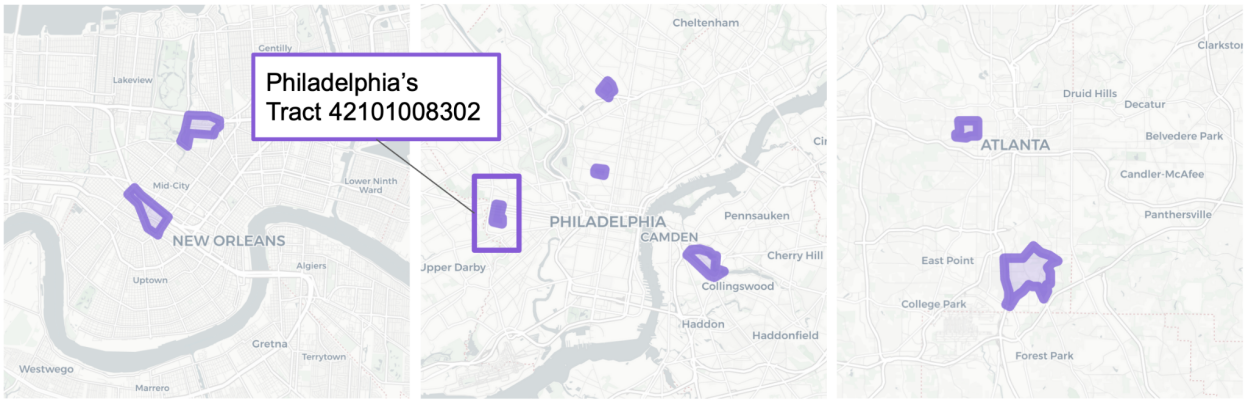
A second approach I use to investigate possible resistance to the lack of positive language in Black neighborhoods is by looking at residuals. At the mean, Black neighborhoods are described less positively than other neighborhood types, but the variation across neighborhoods is broad, and many Black neighborhoods are described very positively. I examine the Black tracts with more of the positive neighborhood adjectives than expected. My intuition is that there

may be different text writing strategies in those neighborhoods, strategies that resist the hegemonic ideology of anti-Blackness, and which perhaps could be replicated or encouraged elsewhere.

I took the average proportion of the 'Positive Adjectives: Location (50)' topic observed in every majority-Black tract with at least 10 advertisements. Then I regressed that tract-average topic proportion on the same tract-level covariates as in my main analysis.

Figure I.5 shows all of the Black tracks with more than 10 advertisements, with the predicted amount of 'Positive adjectives: location' on the X axis and the observed average on the Y axis. The orange tracts are the ten tracts with the highest positive residuals: where the observed value was the largest relative to the modeled value. I read all of the advertisements we collected from those ten tracts.

Figure I.6



New Orleans **Philadelphia** **Atlanta**

Figure I.6 Caption: Figure I.6 shows the locations of eight of the ten Black tracts with the largest positive residuals.

Here are eight of those ten (the other two were in Charlotte NC and Newark, NJ). These tracts are all in traditionally Black neighborhoods in cities with large Black populations. Each of these tracts has its own story, but I focus on one tract that was particularly interesting. This tract

is in West Philadelphia, and has relatively high poverty and low median income. One advertisement from that tract reads:

> Just renovated apartment in the West Philadelphia/Cobbs Creek area. This unit includes a modern renovated kitchen and bath.[…]
> This unit is located in a **vibrant Philadelphia neighborhood** and located within a two unit building that **sits near Cobbs Creek park**. The **neighborhood is adjacent to University City, Clark Park, Cedar Park, Malcolm X Park and close to Philadelphia International Airport.**
> For more info visit mahariyared.com [emphasis added]

This ad was posted by Mahari Yared, a local development company founded by Mahari Bailey, who grew up in West Philly. Mahari Yared's website is explicit about their mission to both invest in Philadelphia and offer modern, affordable housing. Bailey is also active in the art scene and commissions local artists to paint murals in his apartments. It seems that perhaps because of Bailey's connection to the neighborhood and Mahari Yared's mission, they might see more in the neighborhood to be positive about than an absentee landlord. Rucks-Ahidiana (2021) makes a similar point in her recent paper writing about gentrification where she argues that "Black and Latinx [folks might] preserve the racialized culture and racial composition of the neighborhood" (11).

Discussion

Craigslist advertisements do contain racialized descriptions of neighborhood quality which seems to reflect the anti-Blackness of racial ideology in the United States. Those descriptions are further connected to the racialized social structure in the United States, and, in the case of Black neighborhoods, persist when accounting for neighborhood socioeconomic status and housing stock. These findings have implications for how White, Black, Asian, and Latinx neighborhoods are described, and those findings are not uniform across those groups.

These models describe the central tendency of discourse in rental advertisements, which I consider hegemonically anti-Black. However, there are counter-hegemonic signals as well: historic Black neighborhoods are described accordingly, and text writers with connections to neighborhoods seem to do a better job of resisting hegemonic descriptions.

Discourse in online rental advertisements is linked to racial structure in at least three ways. First, the consistently less-positive descriptions of Black neighborhoods reflect the continued and pervasive ideology of anti-Blackness in the United States. Second, that these ads discursively devalue Black, Latinx, and Other neighborhoods is part of the legacy of redlining and other systemic devaluation of Black space. Finally, while this paper focuses on the discursive nature of the ads, they could be part of a reciprocal process: homeseekers could act in reaction to racialized discourse in a way that reproduces that discourse. Such a reciprocal process would be consistent with critical race theories (Ford 1994; Golash-Boza 2016).

Socioeconomic status and race are closely entwined in the United States, such that, on average, majority White or Asian neighborhoods have higher median incomes and higher shares of college graduates than majority Black or Latinx neighborhoods, or neighborhoods with no majority. That association—part of the United States's racialized social system (Bonilla-Silva 1997)—is reflected in the advertisements. When only considering neighborhood race, as in Figure I.1, Black, Latinx, and Other neighborhoods are described with less of the 'Positive Adjectives: Location (50)' topic than White or Asian neighborhoods. The model from Figure I.1 is sufficient as a description of racialized neighborhood discourse, not in spite of its lack of socioeconomic covariates, but because of that lack. It reflects the material differences in space from centuries of racist perception and explicitly racist policies that forced non-White people to

live in already poor neighborhoods and made it difficult for those neighborhoods to improve. Statistically, it uses a tract's racial composition in 2019 to stand in for that history.

While that model is sufficient, it cannot tell us how neighborhoods with similar socioeconomic and built environments, but different racialization, compare in terms of discourse. The model visualized in Figure I.2 is not just about racialized descriptions of neighborhood quality, but attempts to specifically address how neighborhood quality reflects not only the material differences wrought by the history of racism, but racialized perceptions of similar neighborhoods in the present. Of course the model neither measures those perceptions directly, nor accounts for all aspects of the racialized social system. It does, however, suggest that the lack of positive neighborhood descriptions in Black neighborhoods, compared especially to Asian and White neighborhoods, is not only due to observable neighborhood characteristics. Instead, it seems that landlords and text writers tend to describe Black neighborhoods with relatively few positive adjectives even when they are affluent. I argue that such a difference reflects a difference in the value those text writers ascribe to places based on the way that place has been racialized. Landlords and property managers seem to think that even in areas where the median income and rate of college education is high, that if neighborhood is mostly White or Asian then it could be a place to emphasize the 'Location, Location, Location', while Black neighborhoods, it seems, are best left unmentioned.

The way the advertisements describe race is not simply about White neighborhoods being described more positively than Black neighborhoods. Instead, majority Asian neighborhoods are the places which are most commonly described positively, with or without covariates in the model. Though Latinx neighborhoods are described less positively than White and Asian neighborhoods on average, those differences are smaller when I include covariates in the model.

In contrast, the stark results concerning Black neighborhoods seem to align with both a history and a present of material and discursive anti-Blackness.

I consider that anti-Blackness in regard to the description of places in the United States to be hegemonic: it is pervasive and persistent but not without challenge. Instead, I document how other aspects of the neighborhood description, like discussion of neighborhood history, do not distinguish between Black, White, and other neighborhoods. While text writers may not describe Black neighborhoods positively, those writers do sometimes recognize and emphasize the historic character of those neighborhoods. Also, anti-Black descriptions are not uniform: close readings of advertisements from Black neighborhoods with more positive language than the model expects imply that resistance to hegemonic racialized neighborhood descriptions does happen. Property development companies like Mahari Yared, which have both close ties to the neighborhoods they advertise in and a stated mission of racial justice, seem well placed to describe their neighborhoods positively and accurately. Further work should examine this aspect of the advertisements in more detail.

These findings have theoretical implications for the study of housing dynamics, racial ideology, and critical race theory. Theories of housing dynamics emphasize the importance of neighborhood perception in early search stages (Krysan and Crowder 2017). Insofar as readers adopt the neighborhood presentations in the advertisements, this work suggests that those perceptions could be racially structured in part through the ads themselves. Further work should directly address this causal link, but if demonstrated, this adds to our understanding of how segregation could be reproduced.

Racial ideology often considers ideas about race to be explicit and maintained in the minds and actions of individuals. In those cases, anti-racist ideology could be effective as long as

it focused on explicit discourse and education of individuals. Research on less-explicit forms of ideology tend to emphasize coded political speech or evasive talk mostly by White participants (Bonilla-Silva 2003). Critical race theories argue that anti-Black ideology is deeply embedded in the United States. This work supports that argument, but expands the range of locations that ideology might be present. Moreover, it adds a way that racialized perceptions can be stored over time, and where they might spread.

In the specific case of racialized descriptions of neighborhood quality, racist perceptions seem to be stored in and perpetuated by racialized discourse in a way that could have consequences for general perceptions of those places, for housing dynamics, and for policy decisions.

First of all, the texts are a seemingly neutral way of transmitting or perpetuating racialized perceptions of neighborhoods. Racialized ideas of neighborhoods perpetuated through any of those ways could influence people's housing choices. As discussed in Kennedy et al. (2021), there are at least three ways for racialized neighborhood discourse to influence homeseekers choices. First, people with little knowledge of the area may simply adopt the ideas of the neighborhoods described in the ads. That could reproduce segregation if White renters, for instance, exclude all of the units in a particular Black neighborhood because the neighborhood is not described as positively as its socioeconomic status would suggest. Second, while many home seekers might have their own ideas of neighborhoods before reading advertisements, those ideas may not be conscious, and the discourse in the advertisements is subtle, and might escape notice. In that way, even people with established ideas about neighborhoods might have their perceptions slowly and slightly altered. Finally, the advertisements may also naturalize racist ideas (conscious or not) about neighborhoods. Repeatedly reading many positive adjectives

connected to some neighborhoods and few connected to others might make it seem like those distinctions are simply objective descriptions. This seems especially probable given that many people develop strong ideas about neighborhoods without visiting them in person (Krysan and Crowder 2017). Future experimental work should test each of these possibilities more directly.

Regardless of how those neighborhood perceptions are transmitted from advertisements to homeseekers, they could have further effects on the social atmosphere of the neighborhood, on policing policy, on housing policy, and on resource allocation. Walton (2017) documents how new White residents in stably diverse neighborhoods in Boston form homogeneous social networks, challenge neighborhood traditions, institute changes to policies about public land use, and often call the police. Those behaviors are consistent with a negative neighborhood perception, or a neighborhood that these new residents feel they are improving. Alternatively, NIMBYism surrounding, say, development of multi-family homes in a White traditionally single-family zoned area may be motivated by racialized neighborhood perception that the neighborhood is a 'great location.'

While this discussion is focused on the case of rental advertisements, the contemporary digital world is full of potential loci of powerful racialized discourse. This is also a contribution to theories that expect strong connections between racial ideology and racial structure (Golash-Boza 2016). Because Craigslist rental advertisements are on a platform where ideology and structure meet, they can reveal these dynamics. In addition to these theoretical additions, this is a methodological one: scholars of ideology generally and racial ideology specifically should seek out places where we can directly observe the covariation between ideology and structure.

Though this descriptive study has rich theoretical implications, it also has some limitations. First, the sample of 16 metro areas was selected carefully and balanced across region

and metro characteristics. However, results from those metros may not generalize to other cities. Results from rural areas, where Craigslist listings are relatively rare, could also be different. Relatedly, the sample covers a large time span: 2017-2021, and that time span is different in different metro areas. More details about the implications of this sample are discussed in Appendix V. Second, this study is intentionally descriptive: it maps the linguistic contours of racialized neighborhood descriptions. I use the term racialized to reflect discursive associations with spaces where people who are racialized differently live. However, the models presented above cannot speak to what effects those differences in discourse across space might have. Like other descriptive studies, however, I describe the topology of racialized descriptions of neighborhood value with the aim of developing further, specific hypotheses that could be investigated using experimental or other methods to explore causal relationships. Finally, many of the critical race theories that I draw on emphasize the importance of understanding race and gender together, among other axes of difference. This paper focuses on race and socioeconomic status, and leaves a fuller discussion of gender and sexuality for future work.

Conclusion

More than a half century after racial discrimination in housing was outlawed, we still face cities where housing dynamics are shaped by race. This research helps us understand more about why those dynamics have been so persistent by showing how they could be connected to perceptions recorded and reproduced in online rental markets. The paper also contributes to debates about housing policies, especially concerning whether we should focus on developing neighborhoods, often called ‘place-based’ solutions, or encouraging people, usually with vouchers, to move to ‘higher opportunity’ places, or ‘person-based interventions’ (see, for

example, Ellen and Steil 2019). Both of those interventions, however, seem to assert the low quality of the starting neighborhoods, and dismiss the existing community organizations there. These findings, on the other hand, support the need to revitalize not only the material conditions of neighborhoods, but how people think and write about certain neighborhoods. Doing that likely means supporting and strengthening existing community organizations, perhaps alongside place-based and person-based approaches.

There is also ample space for intervention in the real estate industry, as called for by Bresbis and Faber (2020), with special focus on the role of online rental platforms, like Craigslist. Noble (2018), focusing on Google, has shown that digital platforms reproduce racism and racist representations of people. Benjamin (2019) points specifically to the danger of online platforms reproducing segregation, not because the mostly White developers necessarily intend to, but because algorithms that do not consider and resist structural racism will likely support it. So far, Craigslist and other online rental markets have received little scrutiny for their role in offering a platform where others write texts that reproduce racist ideology. However, if those platforms were proactive, they could become a model for how digital platforms can serve as sites of resistance to racism.

References

Benjamin, Ruha. 2019. Race After Technology: Abolitionist Tools for the New Jim Code. John Wiley & Sons.

Besbris, Max, Ariela Schachter, and John Kuk. 2021. "The unequal availability of rental housing information across neighborhoods." Demography 58, no. 4: 1197-1221.

Besbris, Max and Jacob W. Faber. 2020. Furthering Fair Housing in the Residential Real Estate Industry. Poverty & Race, 29(2): 7-8.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent dirichlet allocation." The Journal of machine Learning research 3: 993-1022.

Boeing, Geoff. 2020. "Online rental housing market representation and the digital reproduction of urban inequality." Environment and Planning A: Economy and Space 52, no. 2: 449-468.

Bonam, Courtney, Caitlyn Yantis, and Valerie Jones Taylor. 2018. “Invisible Middle-Class Black Space: Asymmetrical Person and Space Stereotyping at the Race–Class Nexus.” Group Processes & Intergroup Relations 136843021878418. doi: 10.1177/1368430218784189.

Bonam, Courtney M., Hilary B. Bergsieker, and Jennifer L. Eberhardt. 2016. "Polluting Black Space." Journal of Experimental Psychology: General 145, no. 11.

Bonilla-Silva, Eduardo. 2003. Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in the United States. Rowman & Littlefield Publishers.

Bonilla-Silva, Eduardo. 1997. "Rethinking racism: Toward a structural interpretation." American sociological review: 465-480.

Charles, Camille Zubrinsky. 2003. "The dynamics of racial residential segregation." Annual review of sociology 29, no. 1: 167-207.

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. doi: 10.1145/3133875.

Collins, Patricia Hill. 2002. Black feminist thought: Knowledge, consciousness, and the politics of empowerment. Routledge.

Collins, Patricia Hill. 2019. "Intersectionality as critical social theory." In Intersectionality as Critical Social Theory. Duke University Press.

Cottom, Tressie M. 2016. Black Cyberfeminism: Intersectionality, Institutions, and Digital Sociology. Bristol, UK: Policy Press.

Egami, Naoki, Christian J. Fong, Justin Grimmer, Margaret E. Roberts, and Brandon M. Stewart. 2018. "How to make causal inferences using texts." arXiv preprint arXiv:1802.02163.

Ellen, Ingrid, and Justin Steil, eds. 2019. The dream revisited: Contemporary debates about housing, segregation, and opportunity. Columbia University Press.

Eubanks, Virginia. 2018. Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Ford, Richard T. 1994. "The boundaries of race: political geography in legal analysis." The legal geographies reader: law, power, and space, Oxford: Blackwell 87104

Gilliard, Chris. and Culik, Hugh. 2016. “Digital redlining, access, and privacy: filtering content is often done with good intent, but filtering can also create equity and privacy issues.” Available at: https://www.commonsense.org/education/articles/digital-redlining-access-and-privacy.

Golash-Boza, Tanya. 2016. "A critical and comprehensive sociological theory of race and racism." Sociology of race and ethnicity 2, no. 2: 129-141.

Hess, Chris, Arthur Acolin, Rebecca Walter, Ian Kennedy, Sarah Chasins, and Kyle Crowder. 2021. "Searching for housing in the digital age: Neighborhood representation on internet rental housing platforms across space, platform, and metropolitan segregation." Environment and Planning A: Economy and Space: 0308518X211034177.

Hogan, Bernie, and Brent Berry. 2011. "Racial and ethnic biases in rental housing: An audit study of online apartment listings." City & community 10, no. 4: 351-372.

Howell, Junia. 2019. "The unstudied reference neighborhood: Towards a critical theory of empirical neighborhood studies." Sociology Compass 13, no. 1: e12649.

Howell, Junia, and Elizabeth Korver-Glenn. 2018. "Neighborhoods, race, and the twenty-first-century housing appraisal industry." Sociology of Race and Ethnicity 4, no. 4: 473-490.’

Hunter, Marcus Anthony, and Zandria F. Robinson. 2016. "The sociology of urban Black America." Annual Review of Sociology 42: 385-405.

Hwang, Jackelyn. 2016. "The social construction of a gentrifying neighborhood: Reifying and redefining identity and boundaries in inequality." Urban Affairs Review 52, no. 1: 98-128.

Kennedy, Ian, Chris Hess, Amandalynne Paullada, and Sarah Chasins,. 2021. “Racialized Discourse in Seattle Rental Ad Texts.” Social Forces 99(4) doi: 10.1093/sf/soaa075.

Krysan, Maria, Mick P. Couper, Reynolds Farley, and Tyrone A. Forman. 2009. “Does Race Matter in Neighborhood Preferences? Results from a Video Experiment.” American Journal of Sociology 115(2):527–59. doi: 10.1086/599248.

Krysan, Maria, and Kyle Crowder. 2017. Cycle of Segregation: Social Processes and Residential Stratification. New York: Russell Sage Foundation.

Lee, Stacey J. 2008. "The Model Minority and the Perpetual Foreigner: Stereotypes of Asian Americans." In Asian American Psychology, pp. 119-134. Psychology Press.

Massey, Douglas S., and Nancy A. Denton. 1994. American apartheid: Segregation and the making of the underclass. Cambridge, MA: Harvard University Press.

McElreath, Richard. 2020. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC.

McKittrick, Katherine. 2006. Demonic grounds: Black women and the cartographies of struggle. U of Minnesota Press.

Nelson, Laura K. 2017. “Computational Grounded Theory: A Methodological Framework.” Sociological Methods & Research 004912411772970. doi: 10.1177/0049124117729703.

Noble, Safiya Umoja. 2018. Algorithms of Oppression: How Search Engines Reinforce Racism. nyu Press.

Reosti, Anna Christine. 2018. "Tenant Screening and Fair Housing in the Information Age." PhD dissertation.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. “Structural Topic Models for Open-Ended Survey Responses: STRUCTURAL TOPIC MODELS FOR SURVEY RESPONSES.” American Journal of Political Science 58(4):1064–82. doi: 10.1111/ajps.12103.

Rothstein, Richard. 2017. The Color of Law: A Forgotten History of How Our Government Segregated America. New York: Liveright Publishing.

Rucks-Ahidiana, Zawadi. 2021. "Theorizing Gentrification as a Process of Racial Capitalism." City & Community: 15356841211054790.

Sampson, Robert J., and Stephen W. Raudenbush. 2004. "Seeing disorder: Neighborhood stigma and the social construction of “broken windows”." Social psychology quarterly 67, no. 4: 319-342.

Seamster, Louise. 2015 "The white city: Race and urban politics." Sociology Compass 9, no. 12: 1049-1065.

Stan Development Team. 2022. Stan Modeling Language Users Guide and Reference Manual, 2.29. https://mc-stan.org

Walton, Emily. 2021. "Habits of whiteness: How racial domination persists in multiethnic neighborhoods." Sociology of Race and Ethnicity 7, no. 1: 71-85.

Wilkerson, John, and Andreu Casas. "Large-scale computerized text analysis in political science: Opportunities and challenges." Annual Review of Political Science 20 (2017): 529-544.

Wynter, Sylvia. "Beyond Miranda's Meanings: Un/silencingthe'Demonic Ground'of Caliban's' Woman,'." Out of the KumbIa: Woman and Caribbean Literature (1990): 355-72.

Wynter, Sylvia. 2006. "On how we mistook the map for the territory, and reimprisoned ourselves in our unbearable wrongness of being, of desêtre: Black studies toward the human project." A Companion to African-American Studies: 107-118.

Wynter, Sylvia. 2001. "Towards the Sociogenic Principle: Fanon, Identity, the Puzzle of Conscious Experience, and What It Is Like to Be ‘Black.’." National identities and sociopolitical changes in Latin America: 30-66.

Chapter 2: Gentrification and Racialized Exclusion in Online Rental Markets

Ian Kennedy

Abstract

Segregation and neighborhood change continue to be a central problem in inequality across race and socioeconomic status. This paper looks at how home rental advertisements on the Craigslist platform include information about sorting processes through language that could signal exclusion. I start from the framework that housing advertisements are one way landlords can attempt to sort homeseekers to find an 'ideal tenant' to live in their property, and that this sorting takes place within a context of broader processes of racialization and neighborhood change. I analyze a corpus of 500,000 rental advertisements from the Craigslist platform collected from 16 metropolitan areas from 2017 to 2020. To attend to neighborhoods' racial composition, socioeconomic status, recent exposure to gentrification and racial turnover, I combine topic modeling, spatial regression, and qualitative close reading. I find that discursive practices of exclusion and enticement have racialized and classed dimensions, and are tied to gentrification processes in two ways. First, neighborhood-level signaling could exclude people from particular places based on described standards of home maintenance, cleanliness, and the use of Housing Choice Vouchers. Second, the advertisements seem to contain efforts to sort tenants within neighborhoods by using high application fees and implementing background, credit and income checks. Both of these trends are linked to broad ethnoracial and class-based stereotypes about who is clean, who is credit worthy, who is likely to have been incarcerated, or who might have a housing voucher. These patterns suggest that housing advertisements could contribute to the sorting of people into neighborhoods based on their class and ethnoracial identification.

Background

Segregation and neighborhood change continue to be a central problem in inequality by race and socioeconomic status. Sociologists looking at the persistence of the harms associated with those changes have often looked at the real estate sales market, showing how real estate agents have racially homogeneous networks (Bresbis 2020) and how discrimination amplifies across the home-buying process (Korver-Glen 2018).

Discriminatory sorting is well documented in the home buying real estate market. As homeseekers move through the home buying process, actors like loan officers and real estate agents can discriminate in ways that perpetuate racialized housing dynamics (Korver-Glen 2018, Bresbis 2020). In the rental market, on the other hand, those sorting processes are limited to fewer interactions, almost all with a single landlord or property manager. Rosen et al. (2021) found that both large and small landlords used discretion in ways that could lead to discrimination, but that large landlords especially relied on automated tools like background and credit checks. Krysan and Crowder (2017) showed that segregation could persist partially because of the way homeseekers conducted their search. People do not want to move to places that are uninviting to them and exclude those places from their housing search early on, never considering units there. In addition to the interpersonal discrimination that occurs when landlords or property managers meet prospective tenants, this study examines if and to what extent racialized housing dynamics could be perpetuated through exclusionary language that homeseekers are exposed to during their housing search. For instance by reading advertisements on Craigslist, especially when ads reflect landlords' assumptions about who is an appropriate tenant.

Conceptions of who makes an ideal tenant are likely based on the racial and class dynamics of the neighborhood, especially in terms of racialized and classed ideas about cleanliness, being able to pay rent on time, and having negative credentials like low credit scores. Therefore, ads have potentially segregative effects at the neighborhood level by signaling to potential renters the level of welcome they are likely to experience in the neighborhood: which I call 'neighborhood-level sorting.' At the same time, ads with exclusionary language could prevent negatively-credentialed homeseekers from applying for higher-quality units *within* a particular neighborhood due to credit-score limits, background checks, and by refusing people with Housing Choice Vouchers (HCVs). I call these processes 'unit-level exclusion.' Both neighborhood-level sorting and unit-level exclusion could restrict the housing searches of people racialized as non-white, of people with restricted incomes, and of people with negative credentials like contact with the criminal-legal system, low credit scores, or who hold HCVs.

The current paper looks at how home rental advertisements on the Craigslist platform include information about sorting processes–especially language that could exclude or dissuade homeseekers. I read those advertisements in their neighborhood context, paying particular attention to the neighborhoods' racial composition, its socioeconomic status, and its recent exposure to neighborhood change, especially to gentrification.

Restrictions in Real Estate

Existing work on the real estate market, both in home purchases and rentals, shows how landlords, property managers, real estate agents, and others discriminate, steer, exclude, attract, or otherwise guide home searches. This paper builds on qualitative work that investigates how home searches are not only about tenants finding the right homes, but also about landlords looking for the right person to live in their property (Rosen 2014). Korver-Glenn (2018) shows

how real estate agents use their perceptions of neighborhoods, and what kinds of people match those neighborhoods, to stereotypically match clients to homes.Central to that idea is the fact that there are multiple opportunities for gatekeeper intervention in the real estate market, allowing real estate agents, loan officers, and assessors (Howell and Korver-Glenn 2018) to all make discriminatory impacts. The rental market, by comparison, only has a few discretionary opportunities: the housing advertisement, the screening process, and the decision of whom to rent to. Moreover, all three of those interactions are usually with one entity: the landlord or property manager.

Landlords are careful to maximize their ability to control all three of those processes to try to rent to the tenants they want. Rosen (2014) emphasizes the work that landlords do to attract and maintain tenants with HCVs, commonly referred to as ‘Section 8’ vouchers, once again matching the tenants they want to the units they have. Voucher holders are particularly vulnerable to exclusion: Cunningham et al. (2018) demonstrate that source of income (SOI) discrimination is rampant, up to 67% of voucher holders could face discrimination in cities without SOI ordinances. Bresbis et al. (2022) show that advertisements with exclusionary language are fairly common on Craigslist and that advertisements welcoming HCV holders are concentrated in poorer, less White neighborhoods.

Even when they do not engage in discrimination, landlords and property managers seek to maintain discretion in whom they rent to. Grief (2018) shows that when landlords face external pressures, like the possibility of increased utility costs, they respond in part by increasing the screening process, hoping to regain control. Fields (2019) also focuses on the screening process, especially in the case of very large property management and owner companies. In those cases, the screening criteria are digitized and analyzed in a process she calls

the ‘automated landlord’. Fields (2019) argues that stringent screening could result in substantive changes to rental practices, like offering fewer amenities for cheaper units. Rosen (2021) compares how large and small landlords handle screening, showing that not only do large landlords rely on metrics like credit and background checks, but that those measures are dubious in terms of how well they select stable tenants. Landlord screening techniques, however, occur only after prospective tenants have completed applications. Before that, homeseekers seem to have most of the agency: they can look for homes wherever they like. Landlords and property managers, on the other hand, may try to sculpt the information landscape homeseekers face through rental advertisements. I argue that the shape of that information landscape could constrain tenant agency by leading tenants to self-screening. Neighborhood-level sorting could prune some neighborhoods from homeseekers’ consideration, and unit-level exclusion could stop people with negative credentials from applying to rent certain units.

The Craigslist Rental Market

This paper exposes a new vein of possible ways homeseekers are sorted into neighborhoods and units by revealing exclusionary, racialized language in Craigslist rental advertisements. Craigslist has become an exciting new venue for social science research because its advertisements are native data created in a simple format with a few text fields and images, and with lots of freedom for the poster: all of the fields are essentially open to whatever the poster wants to type. A large majority of advertisements have some kind of geolocatable information, whether an address or a map insert. That geolocation allows me to set each rental unit within its neighborhood context, specifically how that place fits in the racialized and classed social system in the United States, based on demographic data at the tract level. Finally, online

rental advertisements are directly connected to housing searches and outcomes. As a result, research on housing markets using Craigslist data has blossomed.

Initial work on Craigslist's rental markets focused on the potential to use scraped rental advertisements to estimate rents with high spatial and temporal granularity (Boeing and Wadell 2017). Later work focused on potential problems with the data, especially its representativeness across space, race, and socioeconomic status (Boeing et al. 2021; Bresbis et al. 2022; Hess et al. 2021). A third theme has been using the distribution of the advertisements through the racially-structured space of the United States to expose and describe racialized differences in the content of the advertisements (Bresbis et al. 2021; Kennedy et al. 2021; Somashekhar et al. 2021).

Focusing on racialized discourse Kennedy (2021) reveals how systematic differences in text reflect themes about gentrification and the implicit valuation of racialized space. Rucks-Ahidiana (2021) emphasizes that gentrification is embedded in processes of racial capitalism. Somashekhar et al (2021) show that the description of securitized amenities, like gates and controlled entry, are more commonly mentioned in areas with high income inequality. Bresbis et al (2021) focus on the metadata, like the number of images, and an overview of the language showing that advertisements in White neighborhoods do more to describe the neighborhood and have more images.

I argue that rental advertisements serve as a key selection step in the rental housing market. While the home buying market features many opportunities for gatekeeper discrimination, the rental market has fewer. Landlords in the rental market, then, use what means they can to find the tenant who fits their unit best (Rosen 2014, Grief 2018). While previous work has focused on the actual screening process landlords conduct, Krysan and Crowder's

(2017) model of housing search expects that a fair portion of neighborhood sorting will already have occurred by the time the landlord has the chance to deny a tenant. Instead, the housing advertisement itself, insofar as it contains language that could influence housing search, is where landlords and other text writers can influence, whether intentionally or not, a homeseeker's eventual destination. Those advertisements are the product of a complicated, though unobserved, process. Text writers can have many roles, including landlords, property managers, or–for large management companies–advertisement copy writers. They may use a template that is common across all of the company or landlord's holdings. However, all advertisements are designed not just to market housing units, but to market them effectively to a tenant that the landlord or property manager considers appropriate for the unit.

I focus on describing how language that could be exclusionary is unequally distributed through space, especially across differently racialized neighborhoods, neighborhoods with different socioeconomic status, and neighborhoods at various stages of gentrification. In interpreting those results, I explore possible explanations for and effects of that unequal distribution. However, the empirical evidence presented below is not sufficient to test those theories. Instead, in the discussion, I point towards ways that future research could examine the causes and consequences of the unequal distribution of exclusionary language.

I am particularly interested in two ways that an advertisement might contribute to tenant sorting processes: 'within-neighborhood selection' of tenants and 'between-neighborhood exclusion'. Within-neighborhood selection describes the way that advertisements might guide, draw, or repulse potential tenants from particular units within a neighborhood. This kind of selection reflects the fact that landlords and property managers are attempting to select the 'ideal tenant', but must make that selection constrained by who applies to their advertisement. In

neighborhoods where text writers expect many non-ideal applicants, exclusionary language might be more common as a way of pre-sorting the tenant applicant pool by excluding some negatively-credentialed applicants.

While within-neighborhood selection is a process whereby prospective tenants refrain from applying to a particular unit, there may also be cases where homeseekers stop considering entire neighborhoods because of what they read in the texts. I use the term between-neighborhood exclusion to refer to language that could convince a homeseeker that a whole neighborhood is or is not a good match for them. Some non-White homeseekers exclude neighborhoods from their housing search if they think they are likely to face discrimination there, often based on accounts from friends or relatives (Krysan and Crowder 2017). Reading advertisements with exclusionary language could have a similar effect, causing people to give up looking for units in certain neighborhoods.

Both within-neighborhood selection and between-neighborhood exclusion occur in a landscape of housing options shaped by the racialized social system and by housing processes like gentrification. I follow Hwang's (2020) definition of gentrification as a process through which a neighborhood that had lower income than the surrounding area experiences increases in income or college education greater than that surrounding area does. That process of neighborhood change is entwined with race. The U.S. racialized social system likely influences what kinds of people landlords consider to be good tenants, through stereotypes and racialized tropes about income, criminal legal contact, credit worthiness, and receiving federal benefits. In other words, the fact that perceptions of neighborhoods are racialized might influence a landlord or text-writer's expectations of their applicants, and therefore shape how they write texts. Gentrification could similarly change who landlords expect to apply for their unit. For instance,

in a gentrifying area, landlords or property managers might expect that more people from outside the neighborhood might apply compared to places that are stably rich or stably poor.

This leads me to three descriptive research questions. First, what kinds of potentially exclusionary language appear in the advertisements? Second, how are these discursive trends distributed across differently racialized and classed neighborhoods? Third and finally, how does that racialization intersect with neighborhood socioeconomic status and with processes of gentrification? I identify four kinds of potentially exclusionary discourse: tenant responsibilities like restrictions and stipulations, backgrounds and credit checks, move-in costs like fees and deposits, and the explicit exclusion of HCV holders. I find that some kinds of exclusionary language, like that discussing background and credit checks, or explicit exclusion of HCV holders, is more common in Black and Latinx neighborhoods. Other language, like that around move-in fees and tenant responsibilities, are more common in White neighborhoods.

Data and Methods

Scraped Rental Advertisements

This paper uses a sample of 2.5 million rental advertisements scraped and geocoded by the National Rent Project (NRP) at the University of Washington from the Craigslist platform between February 2017 and March 2021. The NRP uses the Helena web crawler (Chasins et. al 2017) to access advertisements from 16 different Core-Based Statistical Areas (CBSAs) during that period and collect information including the asking rent, the size of the unit for rent, like the square footage and number of bedrooms, and the listing text of the advertisement. Too many duplicates makes efficient topic modeling difficult (Schofield et al. 2017), so I deduplicated the

advertisements to exclude even very similar texts, since even separate units marketed by the same landlord or property manager may be almost identical. I kept the earliest observed similar advertisement when there were duplicates. The deduplicated dataset had 507,506 individual listings. More information about the deduplication process is available in Appendix I. The distribution of listings across CBSAs is shown in Table 1.

Table 1:

Metropolitan Area (CBSA)	Total Listings	Train Set Listings	Test Set Listings
New York-Newark-Jersey City, NY-NJ-PA	111,253	55,589	55,664
Seattle-Tacoma-Bellevue, WA	79,056	39,543	39,513
Los Angeles-Long Beach-Anaheim, CA	58,792	29,598	29,194
Boston-Cambridge-Newton, MA-NH	41,808	20,986	20,822
Chicago-Naperville-Elgin, IL-IN-WI	34,842	17,401	17,441
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	32,396	16,192	16,204
Urban Honolulu, HI	22,886	11,466	11,420
San Francisco-Oakland-Hayward, CA	22,792	11,355	11,437
New Orleans-Metairie, LA	18,121	9,105	9,016
Providence-Warwick, RI-MA	16,479	8,126	8,353
Riverside-San Bernardino-Ontario, CA	15,561	7,701	7,860
Dallas-Fort Worth-Arlington, TX	15,722	7,894	7,828
Atlanta-Sandy Springs-Alpharetta, GA	13,398	6,740	6,658
Charlotte-Concord-Gastonia, NC-SC	11,083	5,413	5,670
Worcester, MA-CT	7,263	3,612	3,651
Colorado Springs, CO	6,045	3,032	3,013
Total	**507,497**	**253,753**	**253,744**

Neighborhood Characteristics

The NRP geocodes each advertisement within a census tract, using an address included with the advertisement when possible, and the map included with the listing when not. This method allows the NRP to successfully geocode 94% of advertisements with either an address or a map. I use data from the census bureau to identify neighborhood characteristics, operationalizing neighborhoods as census tracts. Tracts are small enough to provide information about the immediate social context for each advertisement, but large enough so that American Community Survey (ACS) 2019 5-year estimates are available and reliable. I use the ACS 5-year estimates for information about neighborhood racial composition, socioeconomic status, and housing stock. To identify gentrifying tracts I also use data from the 2000 decennial census (Manson et al. 2021).

I use a neighborhood typology, rather than proportions of people with particular racial or ethnic identification, to operationalize how each tract fits into the racialized social system in the United States (Bonilla-Silva 1997). I use racial composition as a proxy for that social location, in part based on recent evidence that the way contemporary racialized populations are distributed in space is a good signal of systemic racism (Graetz and Esposito 2022). At the same time, racial composition is a reasonable proxy for how people–both within and beyond the tract–place that area in their racialized understandings of space. To create the typology, I label any tract that is majority White as White. For Latinx, Black, and Asian tracts, I use that label if that group is the largest ethnoracial group and if at least 40% of the tract population shares that ethno-racial identification according to the census. Tracts that I cannot classify as White, Latinx, Black, or Asian using those specifications I label as Other. Summary statistics for the neighborhood characteristics are in Table 2.

Table 2: Summary Statistics (N = 507,506)

Covariate	Mean	SD
Neighborhood Covariates		
Poverty Proportion	0.13	0.10
Log Of Tract Median Income	11.23	0.47
Share Of Adults With College Degrees	0.20	0.10
Share Commuters By Car	0.33	0.14
Share With A Commute Over 20 Mins	0.35	0.10
Proportion Of Owner-Occupied Housing Units	0.46	0.25
Proportion Of Rentals In Properties With >20 Units	0.19	0.22
Proportion Of Units Built After 2010	0.03	0.07
Proportion Of Units Single Family Detached Homes	0.38	0.39
Population Density	0.01	0.01
Gentrifying	0.32	0.47
Gentrifiable But Not Gentrifying	0.23	0.42
Not Gentrifiable	0.46	0.50
Listing Covariates		
Log Of The Listing Rent	7.47	0.53
Number Of Bedrooms	2.01	1.12
Neighborhood Type		
Black	0.09	0.31
Latinx	0.14	0.37
White	0.70	0.54
Other	0.08	0.29
Asian	0.06	0.25

The focus on racialized language of exclusion in the advertisement connects clearly to ideas about neighborhood change. In this paper, I consider two particular types of neighborhood

change: gentrification and racial turnover. For gentrification I borrow from Hwang's (2020) methodology which focuses on how tracts change between two time periods, relative to a larger spatial context. Specifically, Hwang (2020) considers three types of tracts: not gentrifiable tracts that had a median household income higher than their context's median at the first time period, gentrifying tracts that had a lower median income than their context in the first time period, but had higher than median increases in either rent or median income, and finally gentrifiable but not gentrifying tracts which had lower median income than their context at the first time period, but which did not have higher than median rent or income increases.

I use the 2000 decennial census as the first time point, since that predates the mortgage crisis and economic downturn of the 2000s, and the 2019 5-year ACS as my second time point. I examine gentrification in the context of the surrounding area. For tracts that overlap the borders of a Census-defined place, usually a city, with a population of more than 100,000 people, I use that city as the tract's spatial context. I group together tracts that fall outside of such places, but within the same CBSA to serve as a suburban or exurban context. More information about this process is in Appendix VI

Topic Modeling

The key conceptual outcome for this paper is the extent to which advertisements include language that could exclude someone from a particular neighborhood. I use Structural Topic Modeling, or STM (Roberts et al 2014; Egami et al 2019), to quantitatively describe my corpus of Craigslist advertisements. STM is based on Latent Dirichlet allocation, or LDA, (Blei 2003) an earlier topic modeling method. Like LDA, STM produces estimates of the topical composition for each document in a corpus, and estimates the topics based on word co-occurrence. However, STM builds on LDA by considering the topics as a latent variable in a

Structural Equation Model (Roberts et al. 2019), in which the documents relate, through the topics, to other document-level covariates. This allows for better estimates of the association between particular topics and covariates and, just as importantly, better estimates of the error on those associations.

I use a topic model with 50 topics, and include all of the covariates in Table 2 as covariates when I fit the model. The choice of 50 topics is based both on STM's built in fit metrics, and by carefully inspecting the topics and example texts from models with as few as 12 and as many as 75 topics, following the method in Grimmer et al. (2022; see also Nelson 2017; Egami et al. 2018). The model with 50 topics seemed to balance the goals of including important topics, avoiding many duplicate or very similar topics, and having large, indistinct topics. The topics include all of the themes that are common in online rental advertisements: details about the unit, about the property, about the lease, and about the surrounding area.

My first research question concerns the kinds of discourse that could lead to exclusion. To identify possibly exclusive topics, I read the top 20 matching texts for all 50 topics to identify possibly exclusive topics, which produced a short list of seven topics. I then read a further 30 texts for each of those seven topics, qualitatively coding for language that could be exclusionary. In this step I drop four topics where exclusionary language was present but not common in texts linked to the topic. The remaining three topics, shown in Table 3, concern move-in costs which I label 'Fees and Deposits (18)', tenant expectations in a topic I call 'Restrictions and Stipulations (38)', and barriers and requirements for application, labeled 'Background and Credit Checks (46)'. I provide detailed descriptions of each topic, including examples from advertisements, in the results section. A table analogous to Table 3 but including all 50 topics is available in Appendix IV.

One of the topics I excluded in the close-reading step connected specifically to language around HCVs. That topic combined welcoming and exclusionary language, making it a poor quantitative operationalization of exclusionary language. Since HCV discrimination is important to understand, I replaced the topic with a search-string-based measure of Section 8 exclusion. I iteratively tuned a set of phrases, or search-strings, to positively identify texts that explicitly excluded HCV holders. This list includes phrases like ‘not approved for voucher’, ‘not accepting section 8’, as well as common misspellings like ‘not excepting section 8.’ I updated and improved the list by testing it on sets of 100 random advertisements containing the words ‘section 8’ or ‘voucher.’ I stopped when my list of phrases correctly identified 200 advertisements with no errors. In total, the list includes 25 phrases, which are available in Appendix VII.

Table 3

Topic Title	Topic Number	Top Matching Terms
Restrictions and Stipulations	38	ago, consider, taken, rules, owners, noise, tenants, repairs, cleaned, install, portion, disturb, increase, lives, rented
Background and Credit Checks	46	verifiable, statements, felonies, stub, salary, copies, proof, returns, felony, paycheck, judgments, paystubs, requirement, credit, applicants
Fees and Deposits	18	refundable, nonrefundable, adult, insurance, non, damage, pdt, case, application, lbs, holding, processing, brink, charged, liability

Analytic Strategy

For each topic in Table 3, and for the HCV exclusion indication, I conduct three rounds of regression analysis aimed at my second and third research questions. The second question

focuses on the distribution of exclusionary language across differently racialized neighborhoods, so I regress each discursive marker on the neighborhood typology described above. The third question concerns the interaction between neighborhood socioeconomic status and gentrification with the racialization of exclusion, so I run regressions including the covariates shown in Table 2 and an interaction between the neighborhood type and the tract median income (logged). This stage two analysis allows me to investigate how much of the association with neighborhood type comes through other neighborhood characteristics. Similarly, I include an interaction between the neighborhood type and the neighborhood's history of gentrification, since gentrification has been linked to processes of exclusion and attraction.

The models are linear regression models fit using Baysian inference using STAN (Stan Development Team 2022) through the `rethinking` (McElreath 2020) and `cmdstan` (Stan Development Team 2022) packages in R.

Triangulation and Validity

To reduce overfitting risk, the results presented below come from a 50% test sample of advertisements completely distinct from the 50% training set used to train the STM model. To ensure that the results are not just a strange feature of the topic model itself, I validate each included topic by comparing the results to a search-string based analysis. Details on this validation can be found in Appendix III. Finally, I ensure that the topics I analyze do not overly depend on my choice of 50 topics using the robust LDA technique (Wilkerson and Casas 2014), available in Appendix II.

Results

I present mixed-methods results for four types of exclusionary discourse: language that sets tenant expectations, language about background and credit checks, descriptions of move-in fees, and the explicit exclusion of people with housing choice vouchers. For each discursive theme, I first introduce the kind of language texts tend to include in that theme with examples and interpretation. I then weave together descriptions and interpretations of the quantitative results. The results are statistical descriptions of the unequal distribution of each theme across differently racialized neighborhoods, neighborhoods with different socioeconomic status, and neighborhoods at different stages of gentrification. I interpret those results in terms of what implications they might have for landlords and prospective tenants.

Restrictions and Stipulations (38)

The restrictions and stipulations in this topic generally refer to how the landlord expects the tenant to interact with the property and the unit. The highest loading advertisements had language about tenants' responsibilities around taking care of the lawn, of the walls or floors, and of the furnishings. The advertisement with the closest match to this topic came from a White neighborhood in Philadelphia and began with a list of tenant responsibilities:

> 1 No smoking
> 2 Renter's Insurance Required
> 3 Pets permitted with owner approval and pet deposit
> 4 Tenant pays electric, gas, water, oil, cable, internet
> 5 Tenant responsible for lawn care to include grass cutting, weeding, leaf removal and snow removal
> 6 No alterations to home or grounds without written permission from landlord
> 7 use of area rugs or floor protectors required under all furniture resting on hardwood flooring
> 8 Tenant responsible for replacement of any batteries, bulbs or filters in the property during tenancy

The language in this text seems to reflect an expectation on the landlord's part that some tenants might not take proper care of the unit or property. This was explicit in the language of some advertisements where landlords referenced past bad experiences. One advertisement from Seattle which matched highly on this topic included the disclaimer: "Sorry for all the rules, but we have been BURNT too many times!!" This seems to reflect the intention on the part of the landlord that their new tenants not be like previous tenants who did not take proper care of the property.

At the unit level, this corresponds to a landlord or property manager's concern for the property, and a desire to rent to someone who will share that concern. The converse of that desire is to keep away potential tenants whom the landlord believes will not share that concern, a decision that landlords may make based on personal 'gut' discretion (Rosen 2021). This language frames tenants as falling on a spectrum between two possibilities: 'ideal tenants' who pay rent on time and take care of their units, 'professional tenants' (Grief 2018) who are trying to take advantage of, or 'burn' landlords, with a bevy of tenants between those two extremes. In places where this language is common, it seems to send a message: do not bother applying if you are the second type of tenant, in fact, perhaps you should consider a different neighborhood. If that caused tenants to exclude such a neighborhood from their search, it would result in what I call between-neighborhood exclusion: language that signals, and perhaps encourages, exclusivity in particular neighborhoods. The quantitative results show that that message has both a classed and a racial tone.

In both the bivariate model (Figure II.1.1) and the model with socioeconomic controls (Figure II.1.2), language around tenant expectations, as operationalized by the 'Restrictions and Stipulations (38)' topic, is less associated with Black neighborhoods than other neighborhood

types. Contrastingly, this language is most common in White and Asian neighborhoods. From a socio-economic standpoint, the statistical interaction between neighborhood race and median income, shown in Figure II.1.3, shows that this topic is most common in areas with lower median income, though the differences between neighborhoods racialized as Black and other racialized neighborhoods are consistent at all income levels.

That this language about responsibility seems connected to White neighborhoods, and perhaps particularly relatively poor White neighborhoods, seems to reflect controlling images and false stereotypes about what kind of person keeps an area clean and tidy (Collins 2002). This could also perhaps offer gatekeepers a purportedly non-racial explanation for excluding Black prospective tenants. Rosen (2021) documents how some landlords use their personal judgment about homeseekers' existing living space and children to make choices about to whom to rent. If that is the case, this could be reflective of between-neighborhood exclusion: this language might signal to some prospective tenants that this neighborhood is not for them. Alternatively, it could assuage the doubts of relatively wealthy tenants that renting a home in a poorer neighborhood does not mean that the expectations for cleanliness are lax. It could even provide an air of exclusivity to those neighborhoods.

In terms of gentrification, it is interesting that this topic is particularly uncommon in Black gentrifying neighborhoods, shown in Figure II.1.4. That might reflect a desire for landlords in those neighborhoods to be more open and less restrictive about who moves into those neighborhoods. It is interestingly inconsistent with the expectation that gentrifying neighborhoods might have higher levels of exclusionary language. However, there could be a spatial dynamic at play here: this topic may be more common in less-poor areas than in poorer areas. That hypothesis requires further investigation.

Figure II.1.1

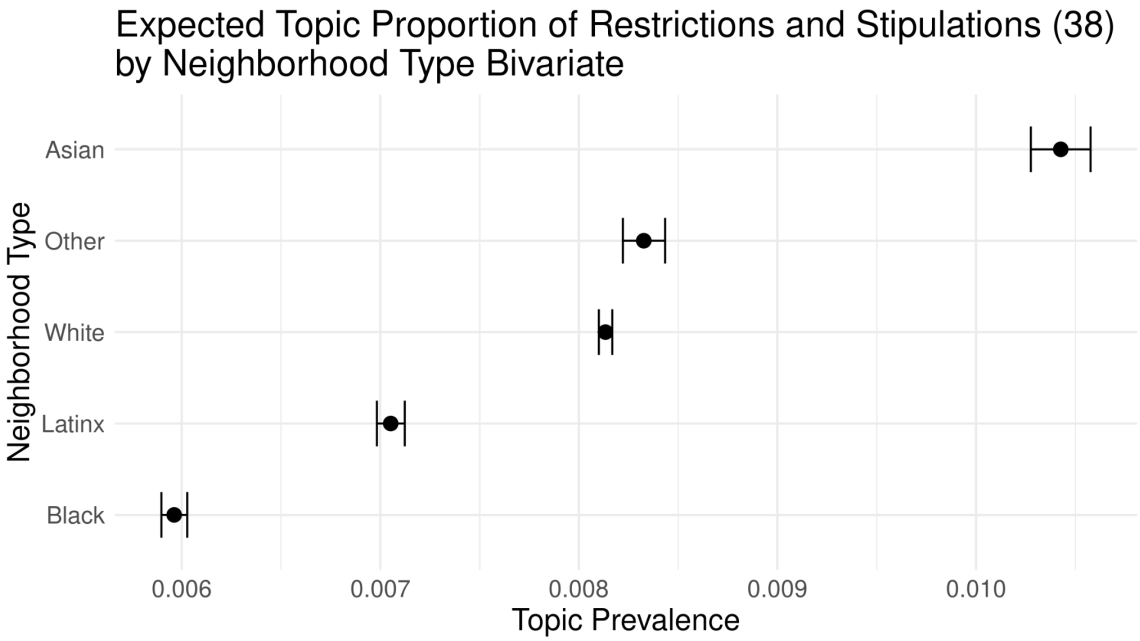


Figure II.1.1 shows the expected topic proportions of 'Restrictions and Stipulations (38)' for each neighborhood type from a model with no other covariates.

Figure II.1.2

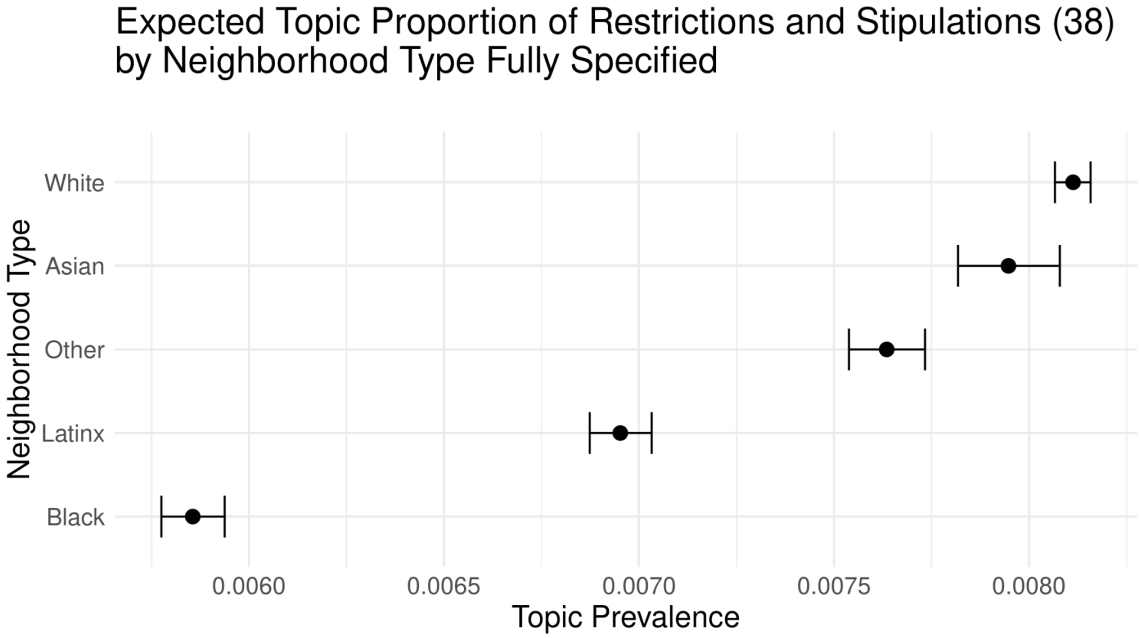


Figure II.1.2 shows the expected topic proportions of ‘Restrictions and Stipulations (38)’ for each neighborhood type from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Figure II.1.3

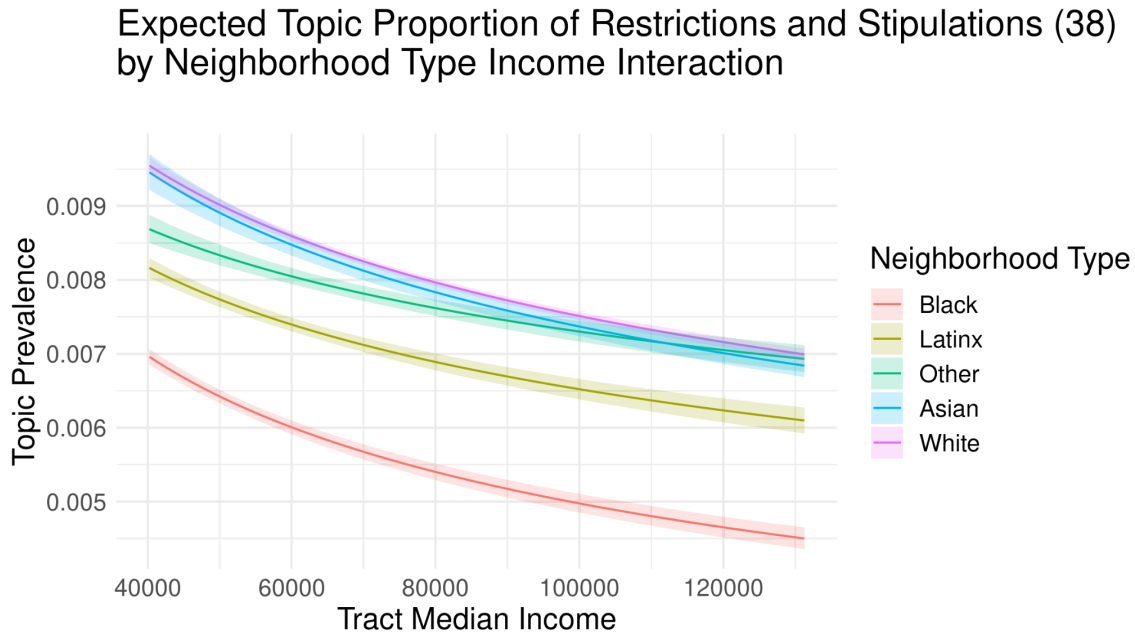


Figure II.1.3 shows the expected topic proportions of ‘Restrictions and Stipulations (38)’ at various hypothetical levels of tract median income from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Figure II.1.4

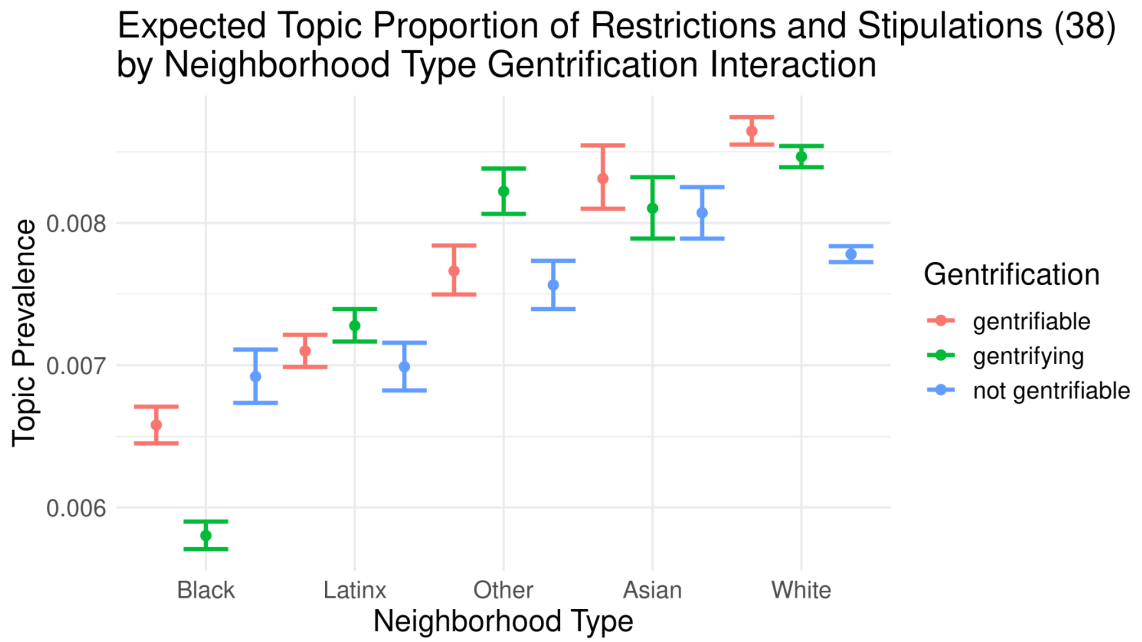


Figure II.1.4 shows the expected topic proportions of 'Restrictions and Stipulations (38)' for each neighborhood type hypothetically gentrifying, gentrifiable, or not gentrifiable from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Together, the results for tenant expectations show that landlords and property managers include more language about 'Restrictions and Stipulations (36)' in non-Black neighborhoods, especially White, Asian, and Other neighborhoods. In all neighborhoods, this language is more likely in neighborhoods with lower median income.

Background and Credit Checks (46)

While tenant expectations are a relatively subtle signal of exclusion, background and credit checks are one of the most straightforward ways the advertisements reflect exclusionary language. Text writers can specify clear boundaries, like a credit rating of 700 or no prior convictions, that explicitly exclude people who do not meet the criteria. One topic, 'Background and Credit Checks (46)' covered this kind of language as well as things like high application fees, specific types and outcomes of background and credit checks, and strict income

requirements. The highest matching text came from a Latinx neighborhood in Riverside, CA, and read in its entirety:

> REQUIREMENTS
> Complete application.
> Credit Check $35.00 per applicant money order
> Copy of valid Driver's License or ID
> Copy of Social Security Card
> ITIN 6 Months Verifiable Employment
> Proof of current income recent check stubs !!!
> Must Make 2.5X Monthly Rent !!!
> If self employed must provide 6 months of bank statements or most recent tax returns.

This advertisement is notable because it includes almost no information about the unit, but its detail about the application process is not out of the ordinary for texts highly matched on this topic. Instead, many advertisements outlined detailed multi-hoop processes for prospective tenants to leap through. One advertisement from a Latinx neighborhood in Los Angeles included 13 steps:

> 1. A completed and signed application from all proposed occupants over the age of 18 must be submitted. ALL LINES MUST BE FILLED IN. Incorrect or misinformation will disqualify you as a prospective renter. We will not contact you for missing information it must be supplied by completely filling out the application
> 2. A valid E Mail address is required in order to complete your credit check [...]
> 13. We will accept the first qualified applicant.

This advertisement, then, outlined a thirteen step process that all tenants were expected to pass through before even being eligible to rent the unit. While only one of the steps was a background check, all of the steps emphasize a relationship between tenant and landlord where the landlord gets to select the person who meets their criteria, and the tenant is judged based on their credentials. Those credentials include both economic measures like credit scores and proof of income, but also ‘social credit’ issues like criminal legal contact or references from past

landlords. Neither of these criteria are obviously good proxies for how good of a tenant someone will be, and both are highly correlated with a homeseeker's race and socioeconomic status (Rosen, 2021).

These types of advertisements could dissuade people from even applying to a rental if they think they will not meet the requirements, or even if the requirements themselves are simply too difficult to carry out. If prospective renters see many ads like this in a neighborhood, they might also believe that the neighborhood might not be for them. When I was looking for an apartment and had a low credit score, I avoided both particular ads and whole neighborhoods for that reason, as I discuss in detail in the discussion. In other words, I both found myself selected into particular units within neighborhoods, or 'within-neighborhood selection and excluding myself from other neighborhoods, or 'between-neighborhood exclusion' based on my expectations about how landlords and property managers might react to my credit score.

Quantitative analysis supports these possibilities. Discussion of background and credit checks is most common in Latinx and Black neighborhoods in the bivariate analysis (Figure II.2.1). Estimates for those neighborhood types remain high when including socioeconomic and housing stock controls, but Asian neighborhoods are estimated as even higher in that model (Figure II.2.2). That suggests that, like 'Restrictions and Stipulations (38)', the racialization of 'Background and Credit Checks (46)' is closely tied to both racialized and classed language. Looking at the interaction between median household income and neighborhood type shown in Figure II.2.3 explains this in the regression analysis: this topic is much more common at low income levels in all neighborhood types.

Landlords may use this language in poorer neighborhoods generally to try and select tenants who can pay, however the higher prevalence in poor Black and Latinx neighborhoods aligns with racist and classist stereotypes about wealth and about criminal legal contact.

Considering gentrification, shown in Figure II.2.4, complements the income analysis. The model expects the largest proportion of 'Background and Credit Checks (46)' in Black neighborhoods that were not gentrifying between 2000-2019. This language seems to serve to sort people within those neighborhoods. This kind of language is also relatively common in Asian neighborhoods that were gentrifying or which had relatively high income in 2000. That might reflect the one aspect of what Zhang and Logan (2016; see also Logan and Zhang 2010) call 'global neighborhoods': neighborhoods with a high proportion of Asian people which might exist between predominantly Black and predominantly White neighborhoods. In those cases, racialized motivations–property managers worried about prospective tenants from those nearby areas–might spur landlords and property managers to conduct more rigorous tenant screenings and to write about those screening procedures in the advertisements.

Some prospective tenants may feel excluded by that language, while others may feel safer thinking that their neighbors have been vetted, despite the fact that background and credit checks are not correlated with good tenancy (Rosen 2021). The racialized motivation in this case need not be explicit, and need not reflect a text writer's own feelings about the neighborhood in question. Instead, if the property owner or manager simply expects their prospective tenants to be worried about what sorts of people might move in, they may be motivated by 'market forces' to include this kind of exclusionary language. Put another way, rigorous requirements for income, credit, and background checks might help attract wealthier homeseekers by assuring them that their neighbors will be of similar socio-economic, and perhaps ethno-racial, status.

Figure II.2.1

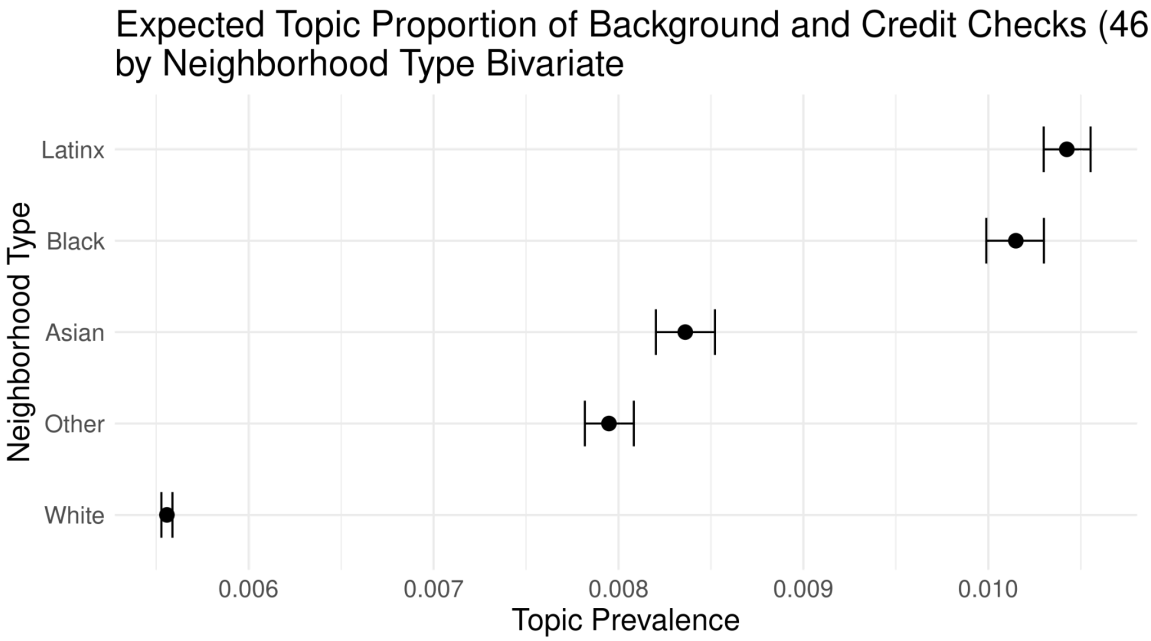


Figure II.2.1 shows the expected topic proportions of 'Background and Credit Checks (46)' for each neighborhood type from a model with no other covariates.

Figure II.2.2

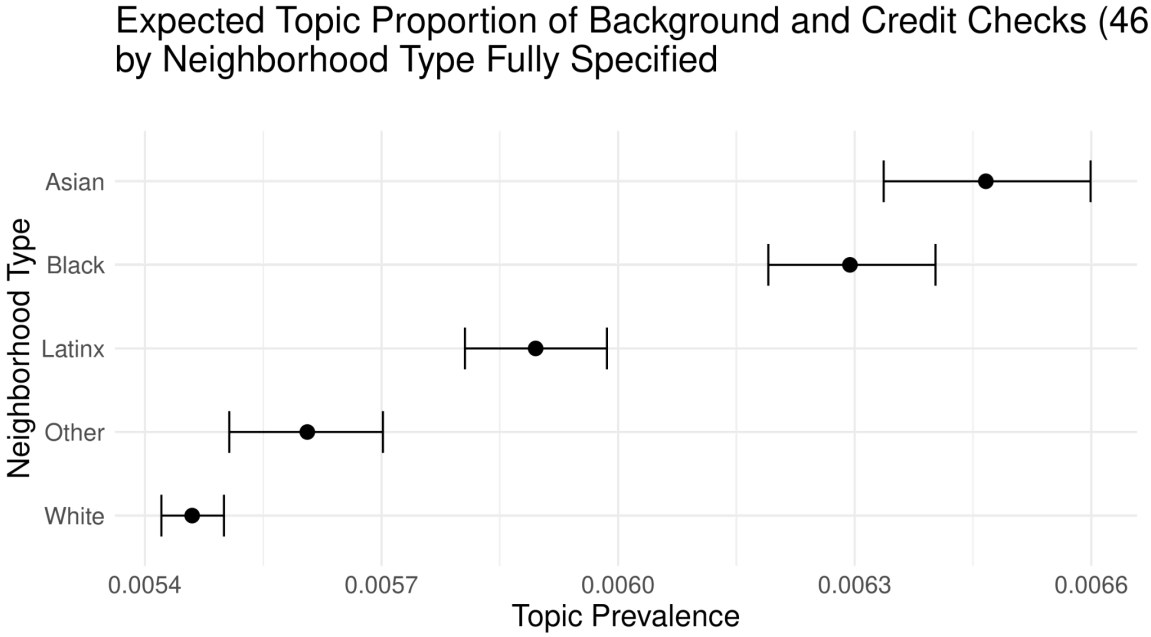


Figure II.2.2 shows the expected topic proportions of ‘Background and Credit Checks (46)’ for each neighborhood type from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Figure II.2.3

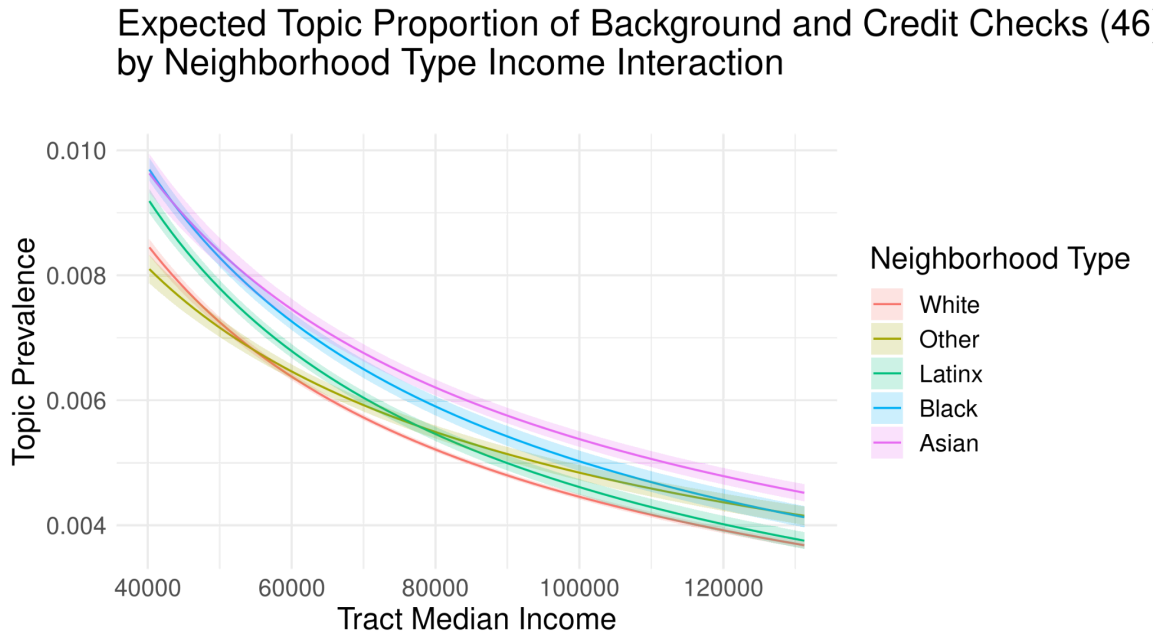


Figure II.2.3 shows the expected topic proportions of 'Background and Credit Checks (46)' at various hypothetical levels of tract median income from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Figure II.2.4

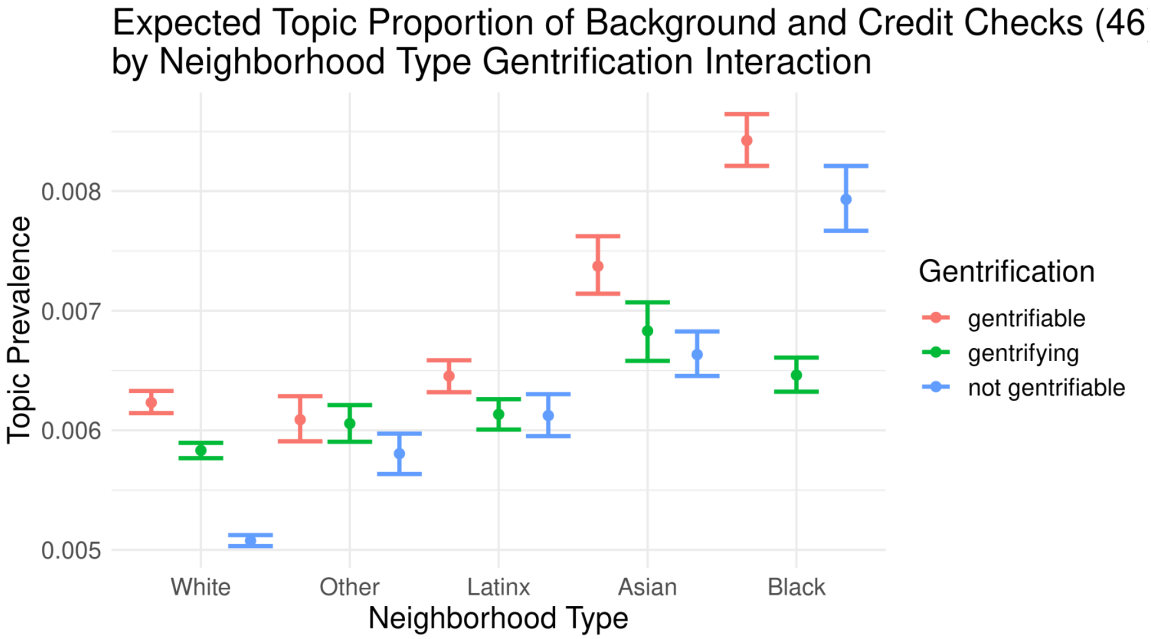


Figure II.2.4 shows the expected topic proportions of 'Background and Credit Checks (46)' for each neighborhood type hypothetically gentrifying, gentrifiable, or not gentrifiable from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Fees and Deposits (18)

While background and credit checks sort people based on negative credentials, text writers also include details about application and move-in costs which raise barriers for homeseekers. Not everyone has easy access to two or three months rent, or can pay multiple 20-50$ application fees, so a straightforward way to sort tenants is through raising the cost of applying to or renting the unit. While it might seem easy to increase that cost by raising the rent, that strategy is difficult: homeseekers are sensitive to monthly costs and may look elsewhere if the rent is too high. On Craigslist, rents are listed with the listing title on the search page, and homeseekers can easily filter their search by rent amount. Raising move-in costs, on the other

hand, may be a less visible option that still has the effect of ensuring that prospective tenants are able to come up with a large payment when necessary. That kind of language is captured in a topic I labeled ‘Fees and Deposits (18)’. As an example, this text was from an advertisement in a White neighborhood south of Seattle, WA:

> Pets Negotiable
> $250 Non Refundable Pet Fee plus additional fees if pet is approved
> $250 Non Refundable pest treatment fee
> $350 Carpet Cleaning Fee
> $100 Utility Deposit
> $150 Non Refundable Admin Fee
> TOTAL MOVE IN $3025

This list is typical of high-matching texts. Many texts had even longer lists of fees, often included extensive details about pet requirements, and were sometimes combined with application fees and requirements. All of these advertisements have, as is emphasized in the example above, high move-in costs relative to their rents. Most advertisements do not list the total move-in cost–over $3000 in this case–but that amount is not atypical.

Looking just at their association with neighborhood racialization, these strategies are most common in Asian and Other neighborhoods in the bivariate case, shown in Figure II.4.1. Those differences are somewhat attenuated when the model includes housing stock and socioeconomic controls, shown in Figure II.4.2, though in that case White neighborhoods have the highest point estimate. In other words, when comparing neighborhoods with similar median income, the model expects White neighborhoods to have more language about move-in costs than other neighborhood types. This kind of language is least common in Black neighborhoods in both modeling strategies. Interaction analysis shows that this kind of restrictive language is more common in all neighborhood types when the tract median income is low, however it drops off faster in Black and Latinx neighborhoods than in White and Other neighborhoods. Like

language about restrictions and background checks, it is possible that this could reflect both within-neighborhood sorting and between-neighborhood exclusion. In poor neighborhoods, landlords want to ensure that prospective tenants can pay the rent, and forcing them to pay fees and deposits upfront is a straightforward way to check if tenants have the capacity to pay: it excludes people who might be able to pay the monthly rent, but have difficulty getting two or three times that amount for a deposit and moving fees. In richer neighborhoods, however, this language may have a more racialized framing. These results are consistent with the idea that landlords in rich white neighborhoods may include this kind of language as a race and class based exclusionary signal. If that were true, White homeseekers could see these extra fees as part of a reasonable premium to live in an exclusive neighborhood.

These results are also consistent with other situations where demand for housing is high in a particular area, like surrounding a university. In those cases landlords might still have difficulty raising rents because most students may not be able to pay high amounts. However, landlords could use high application fees and move-in costs to try and capitalize on the desirability of those neighborhoods: students might more easily be able to borrow one-time fees from parents or on credit cards than handle increased monthly expenses associated with higher rent. But these economic explanations and the possibility of this language being exclusionary are not at odds: high move-in fees could serve both roles by excluding people who cannot pay those fees and by earning some extra money for landlords.

The results for gentrification are also interesting: White neighborhoods not eligible for gentrification because of high income in 2000 and Black neighborhoods that were gentrifying between 2000 and 2019 have similarly low point estimates, as shown in Figure II.4.4. I want to emphasize again that these data are descriptive and cannot identify causes. However, it is

possible that these two kinds of neighborhoods have low levels of language around move-in fees for different reasons. In Black gentrifying neighborhoods landlords may be interested in attracting people to the neighborhood and so limit exclusionary language. On the other hand, in White, stably rich neighborhoods, an emphasis on exclusionary language may not be necessary because the high income and racial concentration of the neighborhood are already exclusionary enough.

Figure II.3.1

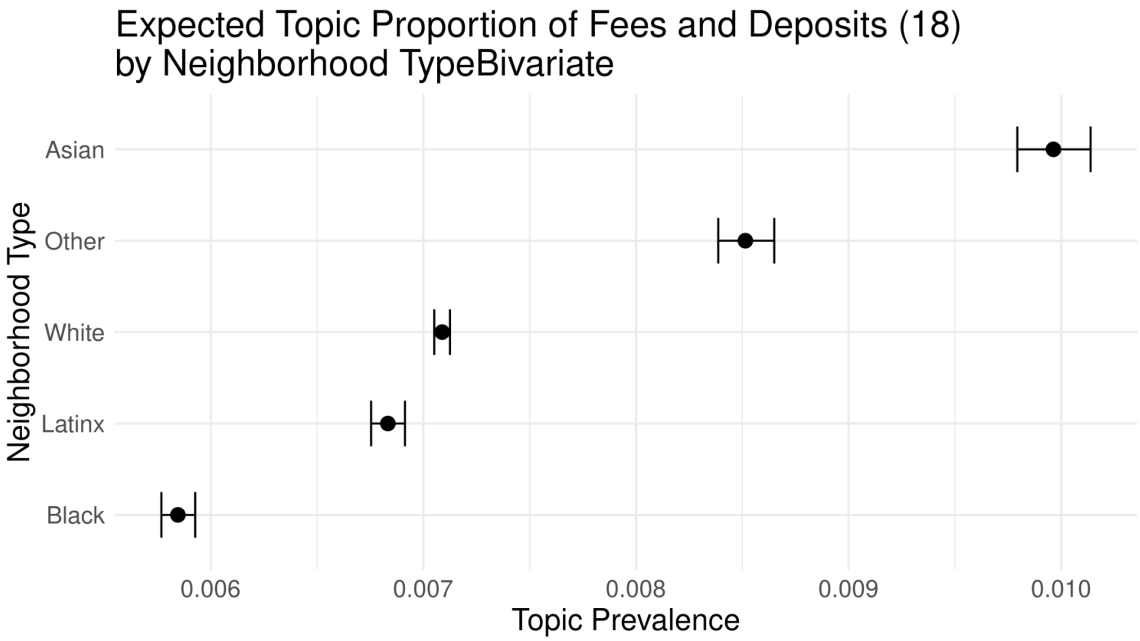


Figure II.3.1 shows the expected topic proportions of 'Fees and Deposits (18)' for each neighborhood type from a model with no other covariates.

Figure II.3.2

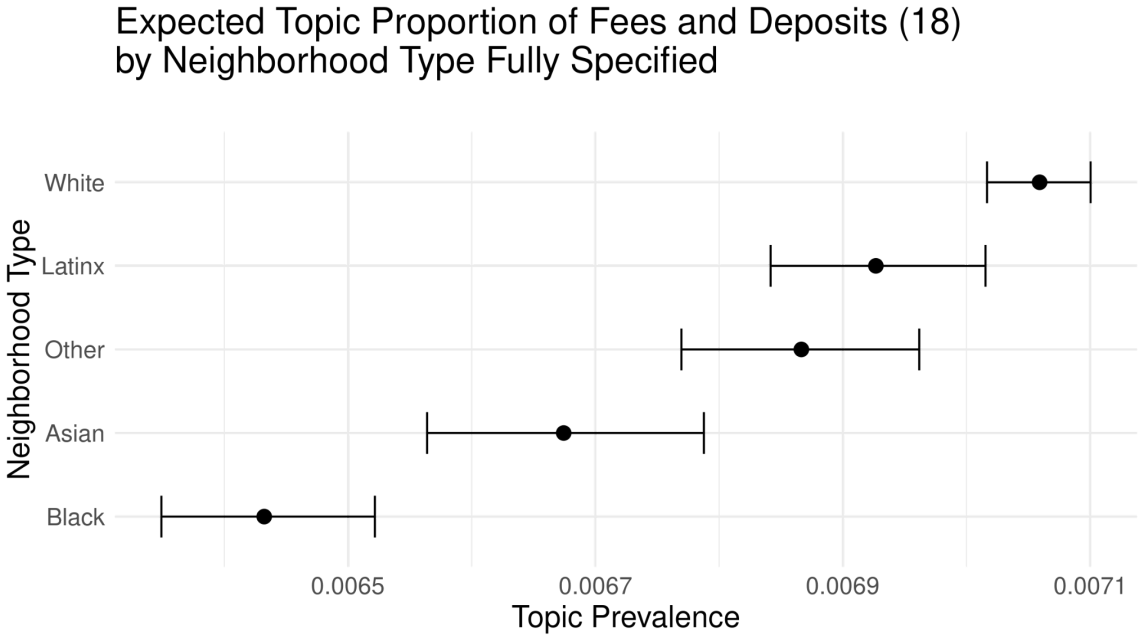


Figure II.3.2 shows the expected topic proportions of 'Fees and Deposits (18)' for each neighborhood type from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Figure II.3.3

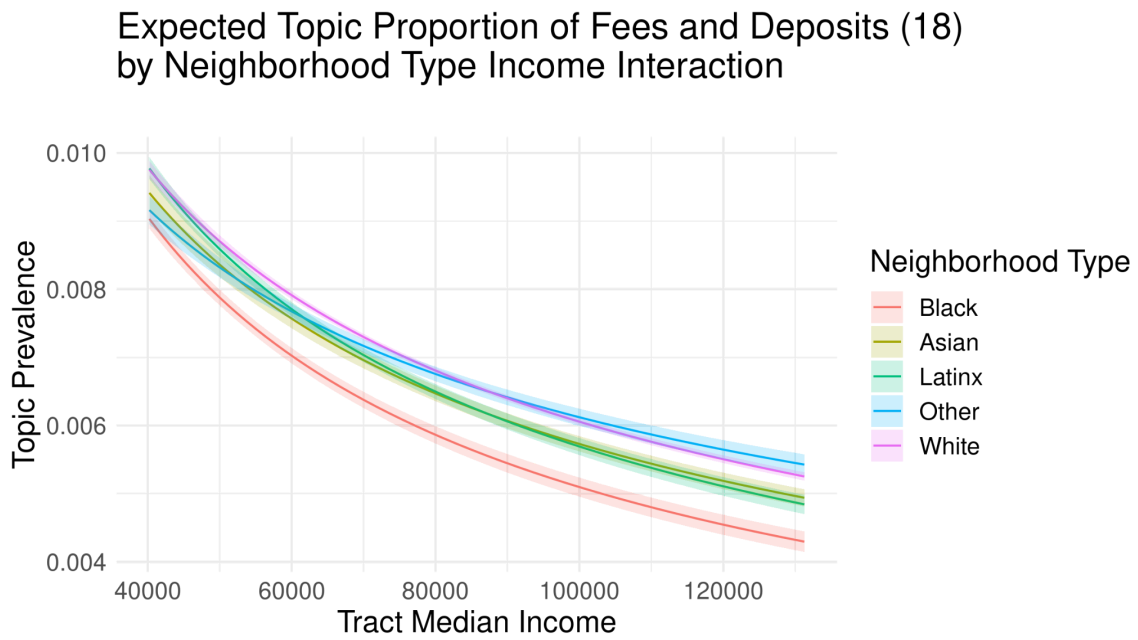


Figure II.3.3 shows the expected topic proportions of 'Fees and Deposits (18)' at various hypothetical levels of tract median income from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Figure II.3.4

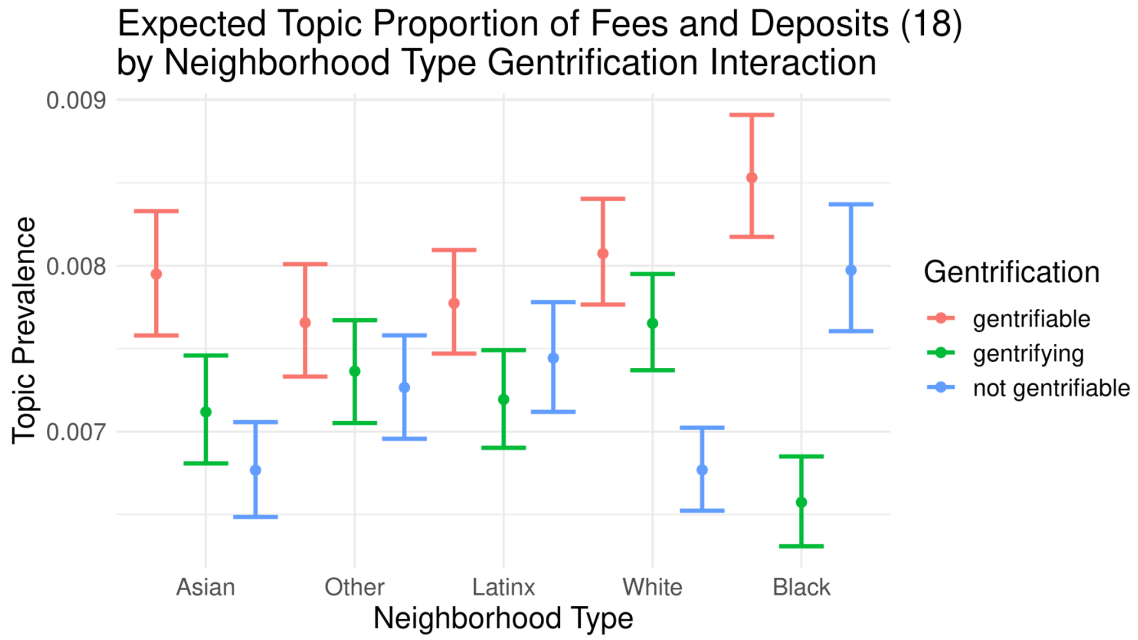


Figure II.3.4 shows the expected topic proportions of ‘Fees and Deposits (18)’ for each neighborhood type hypothetically gentrifying, gentrifiable, or not gentrifiable from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Housing Choice Voucher Exclusion

The most explicit exclusionary language that is common in the advertisements concerns HCV holders. Evidence shows that source-of-income discrimination, where landlords refuse to rent to people with Housing Choice Vouchers (HCVs), is rampant in the rental housing market. Cunningham et al (2018) show that in cities without SOI discrimination laws, up to 67% of HVC holders face discrimination. On the other hand, research shows that some landlords and property owners actively recruit HCV holders, especially in poor neighborhoods when the landlord considers the reliable voucher payment to be worth more than a potentially unreliably non-voucher tenant (Rosen 2014; 2021).

In the rental advertisements, people most commonly refer to HCVs as simply vouchers or with a reference to ‘section 8.’ However, such mentions are rare: only 2.8% of advertisements mention ‘section 8’ or ‘voucher.’ Of those, 36%, or 1% of the total sample, are exclusionary, including phrases like ‘no section 8’, or ‘not set up for section 8 at this time.’ While the other kinds of exclusionary language documented above were well operationalized using a topic from an STM, the topic most closely connected to HCV exclusion mixed together exclusionary and welcoming language. As a result, I operationalize section 8 exclusion here with a set of search strings, the development and validation of which are discussed in more detail in the methods section, and in Appendix VII.

The quantitative analysis shows how this language is unequally distributed between White neighborhoods and neighborhoods with other racialized statuses. The bivariate analysis,

shown in Figure II.4.1 shows that exclusionary advertisements are most common in Asian neighborhoods and least common in White neighborhoods. This is at least partially driven by the very high incidence of exclusionary advertisements in Honolulu, the city with the most Asian neighborhoods in this sample. Honolulu had the highest proportion of advertisements mentioning HCVs, 12.3%, and the highest proportion of exclusionary advertisements, 7.8%, meaning that 63% of HCV mentions in Honolulu were exclusionary, also the largest in the sample. The fully specified model, however, uses fixed effects to account for those kinds of differences between cities. That model expects Black neighborhoods to have the highest proportion of exclusionary advertisements, though that estimate is not statistically significantly different from the estimates for Asian, Latinx, or Other tracts. Each of those, however, is significantly different from White tracts.

Examining the interaction between neighborhood racialization and tract median income, depicted in Figure II.4.3, shows that this difference between White and non-White tracts seems especially pronounced at higher income levels. The gentrification interaction is shown simply for completeness: gentrification status does not seem to be associated with differences in prevalence of this kind of exclusionary language.

While I am confident that I have operationalized this kind of exclusion effectively, I worry if this analysis adequately describes the ways this kind of language is classed and racialized. First of all, since this language is very rare, measuring the overall prevalence may not be the best analysis strategy. Instead, it might be preferable to limit the analysis to tracts with at least one mention of HCVs and instead examine the proportion of exclusionary advertisements. Furthermore, discussions with community tenants-rights organizations suggest that landlords may use HCV welcoming language to fill hard to rent units in poorer or less-White

neighborhoods. My current analysis cannot capture those trends. Finally, exploratory spatial analysis points to the possibility that this kind of exclusion happens in non-poor tracts that border poor tracts. I look forward to addressing these possibilities in future work.

Figure II.4.1

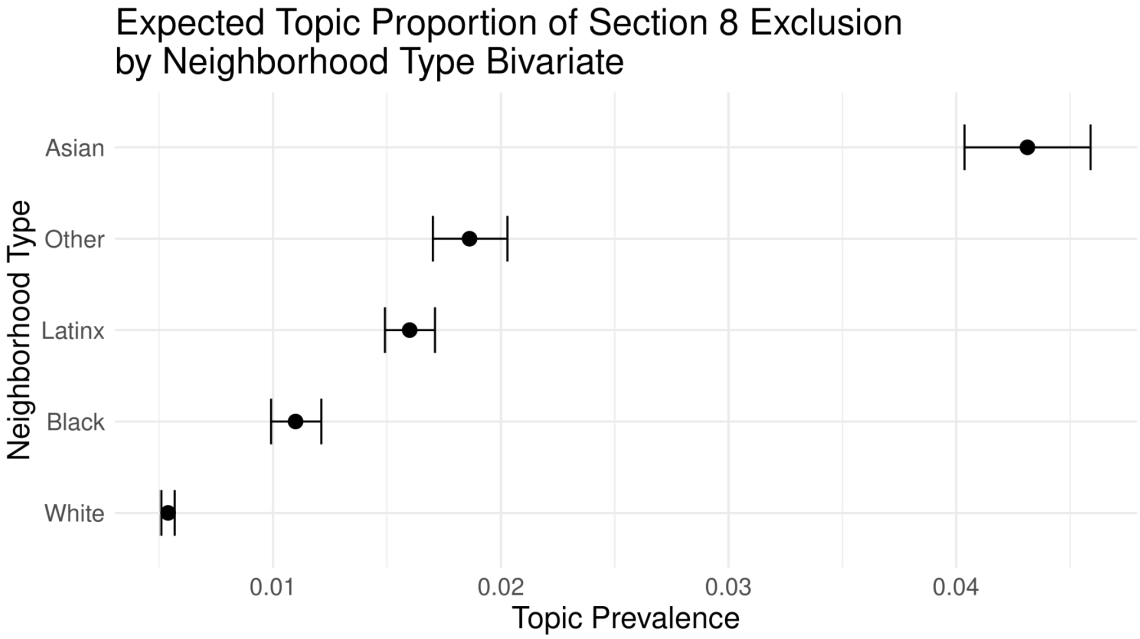


Figure II.4.1 shows the expected probability that an advertisement will include language that excludes HCV holders for each neighborhood type from a model with no other covariates.

Figure II.4.2

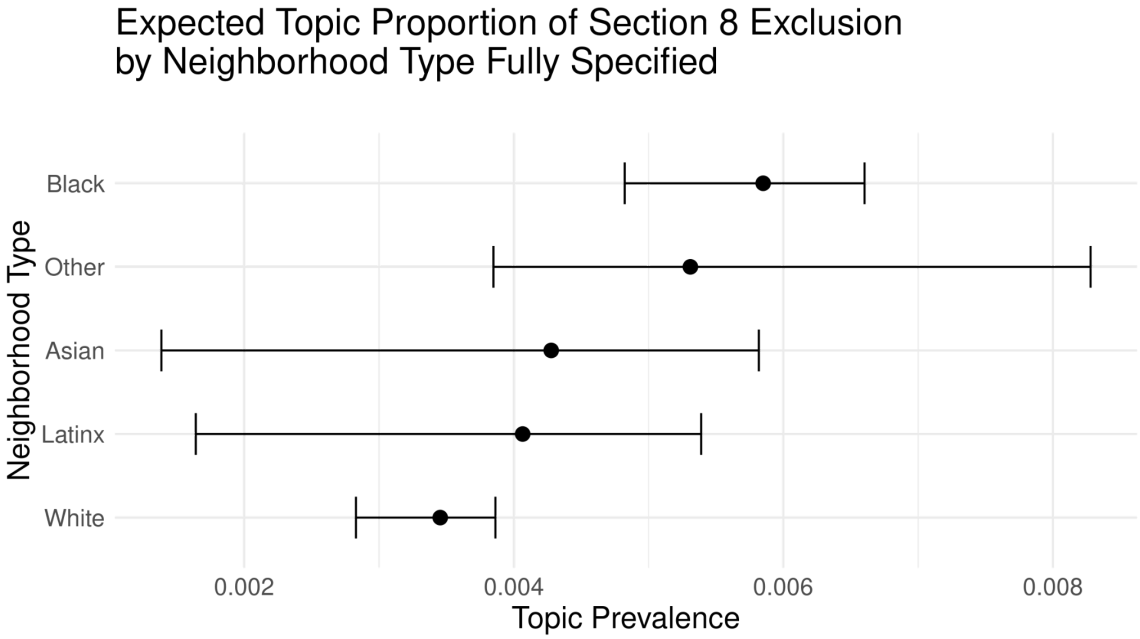


Figure II.4.2 shows the expected probability that an advertisement will include language that excludes HCV holders for each neighborhood type from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Figure II.4.3

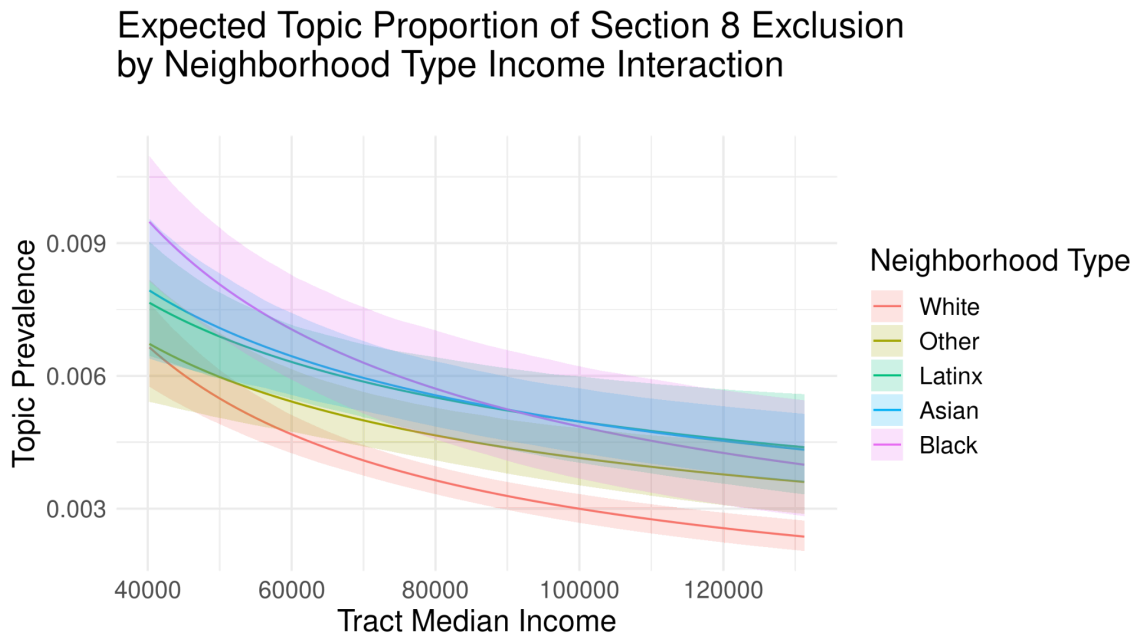


Figure II.4.3 shows the expected probability that an advertisement will include language that excludes HCV holders at various hypothetical levels of tract median income from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Figure II.4.4

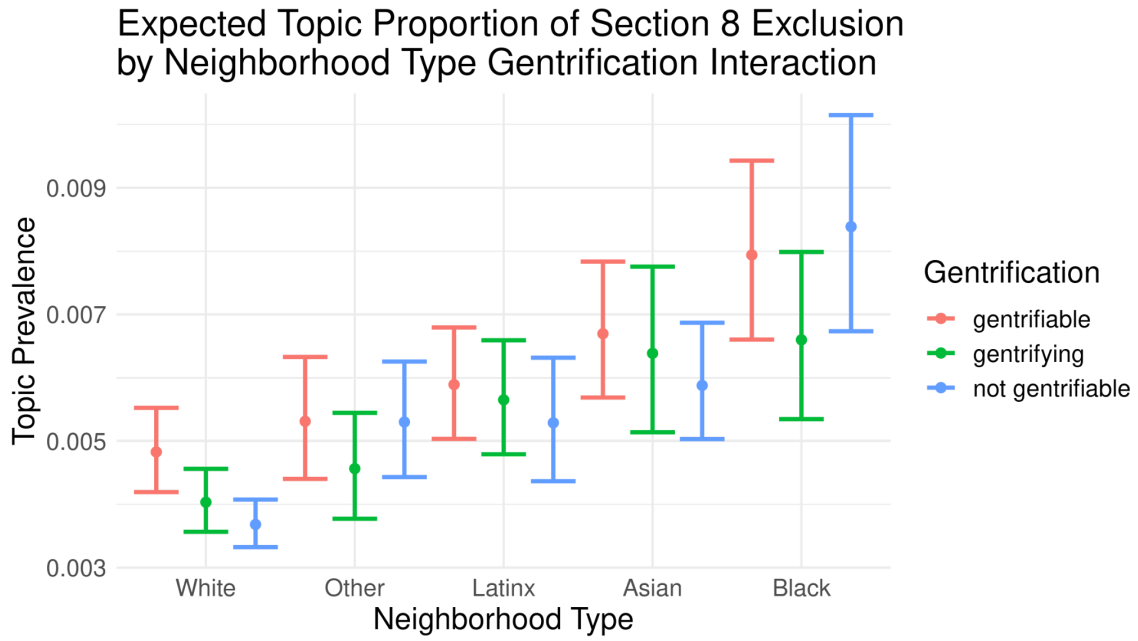


Figure II.4.4 shows the expected probability that an advertisement will include language that excludes HCV holders for each neighborhood type hypothetically gentrifying, gentrifiable, or not gentrifiable from a model with housing stock and socioeconomic covariates, where those covariates are all held at their means.

Discussion

I find that exclusionary language in four aspects of Craigslist advertisements follows race and class differences in neighborhoods, and that it is also associated in some cases with processes of gentrification. While results across the four themes are not identical, they are complementary. Language around 'Restrictions and Stipulations (38)' seems to have raced and classed implications for the kinds of tenants landlords can trust to meet those expectations, perhaps suggesting both within-neighborhood and between neighborhood sorting. In poor neighborhoods restrictive language is universally common, but it drops off more sharply in Black neighborhoods than in other neighborhood types, suggesting that one role of the restrictions in wealthier areas could be to exclude prospective Black residents. 'Background and Credit Checks (46)' also serves both purposes: selecting for the best available tenants in poor neighborhoods, especially non-gentrifying Black neighborhoods, and, perhaps signaling exclusivity in richer Asian neighborhoods. 'Fees and Deposits (18)' seem to be used for between neighborhood exclusion, though the results are less conclusive than for background and credit checks. Finally, language about HCVs seemed generally to be more exclusionary in non-White neighborhoods. These findings align with existing research about how real estate market actors generally (Korver-Glen 2018; 2021; Bresbis et al. 2021), and the rental market actors specifically (Rosen 2014, Rosen 2021) use discretion to drive housing outcomes. This adds to the literature by showing that housing advertisements may allow landlords and property managers to use

exclusionary language to dissuade some tenants from even applying, or perhaps even searching in a particular neighborhood.

The results have three main implications for the sociology of housing. First, while rental transactions are generally unregulated and unrecorded, the advertisements show how much discretion landlords have, and how that discretion might be unequally distributed among populations.

Second, this work broadens the range of possibilities for how exclusion works, the sources where people might get ideas that this place is not for them. For instance, Krysan and Crowder (2017) relate stories of people excluding neighborhoods from their housing search if they thought they might face discrimination in those neighborhoods. Those homeseekers based their neighborhood perception on stories from friends or family. However, rental advertisements might offer another path through which people could exclude neighborhoods, even if they initially consider the neighborhood appropriate for them and even if they can afford a unit there. Even for someone who can afford the rent, high credit checks, SOI discrimination, or the threat of a background check could discourage people with negative credentials from applying, or even from continuing to search in that neighborhood. Even those without negative credentials may stop their search in such areas because they get the impression that the neighborhood is exclusive. This could have real consequences for how we observe discrimination in the rental housing market. If people likely to face discrimination in a particular unit or neighborhood instead never apply, that could make it seem as though discrimination was rare, when in fact those people's housing searches had been significantly shaped by exclusionary language.

Finally, this study allows the consideration of exclusionary language directly in relation to existing patterns of ethno-racial and socioeconomic disparities across space, beyond

considering how landlords and property managers consciously think or act in terms of race or class. These methods can also capture and describe the way that text writers may write ads differently depending on their, possibly unconscious, perception of different places and who lives in those places. Home advertisements, after all, are designed to attract tenants, the 'ideal' tenant if possible. But the implicit conception of that ideal tenant changes significantly across neighborhoods, and these results describe that change in conception to a degree. Specifically, by using the ethnoracial composition and socioeconomic status of the neighborhood as proxies, the results presented above seem to suggest that expectations about potential tenants' class and race influence how landlords and property managers write advertisements. Note that this does not require the investigation to be about different landlords from different neighborhoods, but holds even if it were the same property managers writing about many units and changing language depending on neighborhood characteristics.

As written, this paper has some significant limitations. Craigslist advertisements are only a piece, albeit a large piece, of the online rental market. While many units are posted to Craigslist as well as other platforms, those units may be marketed differently depending on the platform. However, even with that caveat, these results still reflect discursive techniques used by real landlords and property managers, and my analysis accurately describes the distribution of those techniques through space and their association with different neighborhood characteristics.

This work solely considers differences between neighborhoods, using Metro-level fixed effects to account for metropolitan context, and careful preprocessing to avoid over-counting listings from large properties. However, considering policy options without understanding the metropolitan context is dangerous. Existing local policies likely influence apartment listing practices in particular ways, and adding new policies will likely have unforeseen effects.

Therefore further work should directly examine how these results change across metropolitan areas and if the history or policy context of those metropolitan areas impacts the results. Separately, there may be interesting variations in the language depending on the characteristics of landlords themselves. While surveys suggest that landlords of all sizes post on Craigslist (Crowder et al. 2018), other work suggests that large and small landlords deal differently with their discretion when making choices about to whom to rent (Rosen, 2021). Those differences could be interesting and could be investigated in further work that can match listings more directly to property size or ownership.

More broadly, while I speculate about what effect reading these advertisements has on homeseekers, I do not present direct evidence to support those claims. I know from personal experience that homeseekers sometimes adjust their behavior in response to screening tactics. When I was turned away from a rental property because of my credit rating, I changed my behavior and no longer looked for homes in that neighborhood or at advertisements with credit requirements. However, a more systematic investigation is surely needed. Surveys and interviews of homeseekers and renters could assess how much people think about exclusionary language while they search, and experimental methods could assess how people might react in the moment to reading such language.

Conclusion

Throughout this paper flows the spectre of intention: to what degree do landlords, property managers, and other text writers purposely engage in the clearly racist and classist stereotypes that echo in the advertisements presented above? But I prefer asking this question differently: to what extent do people who engage in racist speech use racist ideology and to what

extent does that ideology use them? This question is important both because its answer has ramifications for how we theorize racism, but also because our practical responses to discrimination should be different depending on the answer. If individual discrimination were the key problem, than our existing methods of identifying and prosecuting violators should be enough. However, in a racialized social system, as in the United States, racist ideology serves to buttress, naturalize, and hide the racial hierarchy (Bonilla-Silva 1997; 2003; Golash-Boza 2017). The dominant racial ideology makes it seem natural–and not racist or classist–for landlords to write advertisements with exclusionary effects. In order to resist the added harms of racialized language in advertisements, we will likely need to directly challenge that racial ideology, not just police the language of the advertisements.

Even just resisting discrimination at the level of individual landlords is difficult. One of the big missing pieces here is that these advertisements often concern the details of rental applications and of leases. Compared to the home buying market, those documents are generally unregulated. Whether through legal or academic means, we need more attention on those documents.

We know that SOI discrimination is rampant and harmful: it subverts one of the best tools we have to help housing insecure people. But we also know that many HCV holders never find a unit where they can use their voucher. This paper adds a further way that HCV holders may be restricted in their access not just to high-opportunity neighborhoods, but to any housing at all. When landlords who do not mention vouchers tend to discriminate, and when many of the advertisements that mention vouchers are exclusionary, voucher holders may feel like they have nowhere to go. In that situation, landlords can effectively create an applicant pool with no voucher holders without explicitly discriminating.

This connects to other ways that landlords might shape their applicant pool using exclusionary language, based on income, credit, and other 'social credit' indicators. As Rosen (2021) emphasizes, we know that those outcomes come from a generally unequal society and do not effectively indicate who will make a good tenant. Credit and background checks as they currently stand perpetuate the inequality they are created by, and should not be used to decide who gets housed.

This is about the fundamental injustice of having a housing search process where the stakes for one party are economic while the stakes for the other—having a home at all—are existential. Whether on or offline this deep injustice is hidden through the conceit of a housing 'market', and the attendant implication that market actors are somehow on equal footing. Until all homeseekers can compare apartments and houses with the security that no matter what, they will have a safe place to live, housing markets cannot help but produce incentives that cause landlords to discriminate.

References

Benjamin, Ruha. 2019. Race After Technology: Abolitionist Tools for the New Jim Code. John Wiley & Sons.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent dirichlet allocation." The Journal of machine Learning research 3: 993-1022.

Boeing, Geoff. 2020. "Online rental housing market representation and the digital reproduction of urban inequality." Environment and Planning A: Economy and Space 52, no. 2: 449-468.

Boeing, Geoff, and Paul Waddell. 2017. "New insights into rental housing markets across the United States: Web scraping and analyzing craigslist rental listings." Journal of Planning Education and Research 37, no. 4: 457-476.

Boeing, Geoff, Max Besbris, Ariela Schachter, and John Kuk. 2021. "Housing search in the age of big data: smarter cities or the same old blind spots?." Housing Policy Debate 31, no. 1: 112-126.

Bonilla-Silva, Eduardo. 2003. Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in the United States. Rowman & Littlefield Publishers.

Bonilla-Silva, Eduardo. 1997. "Rethinking racism: Toward a structural interpretation." American sociological review: 465-480.

Besbris, Max, John Kuk, Ann Owens, and Ariela Schachter. 2022. "Predatory Inclusion in the Market for Rental Housing: A Multicity Empirical Test." Socius 8: 23780231221079001.

Besbris, Max, Ariela Schachter, and John Kuk. 2021. "The unequal availability of rental housing information across neighborhoods." Demography 58, no. 4: 1197-1221.

Besbris, Max and Jacob W. Faber. 2020. Furthering Fair Housing in the Residential Real Estate Industry. Poverty & Race, 29(2): 7-8.

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. doi: 10.1145/3133875.

Collins, Patricia Hill. 2002. Black feminist thought: Knowledge, consciousness, and the politics of empowerment. Routledge.

Cunningham, Mary, Martha Galvez, Claudia L. Aranda, Robert Santos, Douglas A. Wissoker, Alyse D. Oneto, Rob Pitingolo, and James Crawford. 2018. A pilot study of landlord acceptance of housing choice vouchers. Washington, DC: US Department of Housing and Urban Development, Office of Policy Development and Research.

Crowder, Kyle, Erin Carll, Jerald Herting, and Chris Hess. 2018. Seattle Rental Housing Study: Final Report. City of Seattle Office of City Auditor.

Egami, Naoki, Christian J. Fong, Justin Grimmer, Margaret E. Roberts, and Brandon M. Stewart. 2018. "How to make causal inferences using texts." arXiv preprint arXiv:1802.02163.

Fields, Desiree. 2022. "Automated landlord: Digital technologies and post-crisis financial accumulation." Environment and Planning A: Economy and Space 54, no. 1: 160-181.

Golash-Boza, Tanya. 2016. "A critical and comprehensive sociological theory of race and racism." Sociology of race and ethnicity 2, no. 2: 129-141.

Graetz, Nick, and Michael Esposito. 2022. "Historical redlining and contemporary racial disparities in neighborhood life expectancy." Socarxiv. 10.31235/osf.io/q9gbx.

Greif, Meredith. 2018. "Regulating landlords: Unintended consequences for poor tenants." City & Community 17, no. 3: 658-674.

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. Text as data: A new framework for machine learning and the social sciences. Princeton University Press. Princeton, NJ.

Hess, Chris, Arthur Acolin, Rebecca Walter, Ian Kennedy, Sarah Chasins, and Kyle Crowder. 2021. "Searching for housing in the digital age: Neighborhood representation on internet rental housing platforms across space, platform, and metropolitan segregation." Environment and Planning A: Economy and Space: 0308518X211034177.

Howell, Junia, and Elizabeth Korver-Glenn. 2018. "Neighborhoods, race, and the twenty-first-century housing appraisal industry." Sociology of Race and Ethnicity 4, no. 4: 473-490.'

Hwang, Jackelyn. 2016. "The social construction of a gentrifying neighborhood: Reifying and redefining identity and boundaries in inequality." Urban Affairs Review 52, no. 1: 98-128.

Hwang, Jackelyn. 2020 "Gentrification without segregation? Race, immigration, and renewal in a diversifying city." City & Community 19, no. 3: 538-572.

Kennedy, Ian, Chris Hess, Amandalynne Paullada, and Sarah Chasins,. 2021. "Racialized Discourse in Seattle Rental Ad Texts." Social Forces 99(4) doi: 10.1093/sf/soaa075.

Korver-Glenn, Elizabeth. 2018. "Compounding Inequalities: How Racial Stereotypes and Discrimination Accumulate across the Stages of Housing Exchange." American Sociological Review 83(4):627–56. doi: 10.1177/0003122418781774.

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

Krysan, Maria, and Kyle Crowder. 2017. Cycle of Segregation: Social Processes and Residential Stratification. New York: Russell Sage Foundation.

Logan, John R., and Charles Zhang. 2010. "Global neighborhoods: New pathways to diversity and separation." American journal of sociology 115, no. 4: 1069-1109.

Nelson, Laura K. 2017. “Computational Grounded Theory: A Methodological Framework.” Sociological Methods & Research 004912411772970. doi: 10.1177/0049124117729703.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2021. IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V16.0

McElreath, Richard. 2020. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. "Stm: An R package for structural topic models." Journal of Statistical Software 91: 1-40.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. “Structural Topic Models for Open-Ended Survey Responses: STRUCTURAL TOPIC MODELS FOR SURVEY RESPONSES.” American Journal of Political Science 58(4):1064–82. doi: 10.1111/ajps.12103.

Rosen, Eva. 2014. "Rigging the Rules of the Game: How Landlords Geographically Sort Low–Income Renters." City & Community 13, no. 4: 310-340.

Rosen, Eva, Philip ME Garboden, and Jennifer E. Cossyleon. 2021. "Racial discrimination in housing: how landlords use algorithms and home visits to screen tenants." American Sociological Review 86, no. 5: 787-822.

Rucks-Ahidiana, Zawadi. 2021. "Theorizing Gentrification as a Process of Racial Capitalism." City & Community: 15356841211054790.

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, pp. 2737–47.

Somashekhar, Mahesh, Chris Hess, Ian Kennedy, and Kyle Crowder. 2021. “Home Security, Privatized Amenities, and the Promotion of Micro-Segregation in Craigslist Rental Advertisements”. American Sociology Association Annual Meeting 2021.

Stan Development Team. 2022. Stan Modeling Language Users Guide and Reference Manual, 2.29. https://mc-stan.org

Wilkerson, John, and Andreu Casas. "Large-scale computerized text analysis in political science: Opportunities and challenges." Annual Review of Political Science 20 (2017): 529-544.

Zhang, Wenquan, and John R. Logan. 2016. "Global neighborhoods: Beyond the multiethnic metropolis." Demography 53, no. 6: 1933-195

Chapter 3 Metropolitan Level Variation in Racialized Discourse: Neighborhood Quality, Exclusion, and Segregation

Abstract

Housing in the United States is structured by racism. This racialized structure is based in a history of legally enforced segregation (Rothstein 2017), White mob violence (Anderson 2016), and racist ideology. Contemporary housing dynamics—continued segregation, gentrification, and racial displacement—perpetuate this history. A key aspect of those processes were past policies like redlining which explicitly linked neighborhood quality and racial categories and essentially prevented non-White Americans from getting mortgage financing. Chapter 1 of this dissertation has shown that racialized conceptions of neighborhood quality persist in contemporary housing advertisements. Another contributor to the perpetuation of racialized housing dynamics are real estate actors. Recent research has shown that real estate actors and gatekeepers may contribute to this perpetuation by using explicit or implicit racial ideas about who should live where (Korver-Glen 2018; Bresbis 2022; Rosen 2014; 2021). Chapter 2 of this dissertation showed that rental advertisements that included exclusionary language were unequally distributed across differently racialized neighborhoods. However, that work has not focused on how those processes may differ across metropolitan areas, thereby preventing the possibility of observing large-scale contextual associations with policy or metropolitan level segregation. This paper investigates whether the racialized discourse exposed in the first two chapters is consistent across metropolitan areas, or more focused in some areas than others, and how evenly the prevalence of exclusionary language is across metropolitan areas. Building on that analysis, I investigate whether particular metro-level variables, especially segregation and certain rental policies, can account for some of that variation across metropolitan areas. I find that racialized discourse about neighborhood quality consistently devalues Black neighborhoods relative to White neighborhoods across metropolitan areas. Discourses of exclusion, on the other hand, vary across metropolitan areas. Language about background and credit checks was more common in less-Segregated metropolitan areas and in those which have rent control laws.

The aim of this chapter is to contribute to sociological understanding of how rental advertisements might vary across Metropolitan areas, and how metropolitan-level variables might explain part of that variation, especially in terms of advertisements that concern the value of the neighborhood, and advertisements that are exclusionary.

The United States is subject to a racialized social system based in histories of slavery, settler colonialism, exclusion, and internment. That system is nationwide, covering all 50 states and the U.S.'s remaining colonial territories. However, the way that system enacts racialized violence and other harms varies across regions and cities and neighborhoods. This study uses one measure of racial ideology–drawn from Craigslist advertisements–to see how much of the variation of that ideology is associated with history and policy at the city level: segregative rent control policies and laws preventing source of income (SOI) discrimination, compared to how much of that variation is associated with neighborhood characteristics, including the way the neighborhood is situated in the broader racialized social system and its socioeconomic status. Specifically, I examine how language about the value of the neighborhood and exclusionary language vary at those two levels of analysis. This paper investigates whether the racialized discourse exposed in the first two chapters is consistent across metropolitan areas, or more focused in some areas than others, and how evenly the prevalence of exclusionary language is across metropolitan areas. Building on that analysis, I investigate whether particular metro-level variables, especially segregation and certain rental policies, can account for some of that variation across metropolitan areas. I find that racialized discourse is indeed consistent across metropolitan areas, but that source of income laws, rent control, and metropolitan-level segregation explain only a small portion of variance in racialized and exclusionary discourse.

Background

I first review the literature on how racial structure and racial ideology intertwine in the United States. Then I consider existing research on the ways that real estate actors use race as a sorting mechanism, both in terms of how they describe the relative value of neighborhoods and decide what kinds of people deserve to live in a place. I then theorize how similar processes might influence how landlords, property managers and other text writers might compose rental advertisements, before turning towards a comparison of how the hyperlocal context of a census tract and the broader local context of a metropolitan area might have features that influence the way that landlords write texts. Finally, I briefly review the research on online rental platforms.

Racist Structure and Ideology

Studying racialized discourse necessarily means engaging with racial ideology: the discourse is an expression of that ideology, a reflection of that ideology, and a signal about the shape of that ideology. Given that these are rental advertisements that reflect the result of housing dynamics that are racially structured, the advertisements have the special feature of laying at a site where structure and ideology articulate, where they might move each other in a reciprocal process.

Critical race theories in sociology have focused on how racial disparities in health, wealth, and life chances generally can be traced back to engagement with the racialized social system in the United States (Bonilla-Silva 1997). A central way that system reproduces itself is through racial ideology that naturalizes and validates the existing system (Bonilla-Silva 2003; Golash-Boza 2016). Traditional forms of racial ideology in the U.S. focused on devaluing non-white spaces and bodies, often doing so through gendered and sexualized tropes (Collins 2002; Spillers 1987). In the decades following the civil rights movement–as affirmative action,

school integration, and other anti-racist polices had generally positive effects–explicitly racist language became less common in the public sphere. Instead, 'new racism' (Bobo and Fox 2003) or a 'colorblind racial ideology' (Bonilla-Silva 2003) emerged as a way to continue to justify racial discrepancies in life chances via seemingly race-neutral or unbiased processes. Contemporary racial discourse reflects both explicit and colorblind aspects, sometimes joined together.

Both explicit and colorblind ideologies serve a key purpose in the racialized social system: to uphold and justify it and thus encourage its reproduction. However, it is worth noting that those two types of ideology reflect two ways of thinking about ideology, one the more traditionally Marxist Antonio Gramsci (Bates 1975), and one from neo-Marxist Louis Althusser (2006 [1971]) thought. Gramsci (Bates 1975) argued that the ruling class used cultural hegemony, a warped account of the social world, to justify and preserve their status. White supremacist and explicitly racist discourse aligning with Gramsci's accounting of cultural hegemony: they explicitly argue for the perpetuation of a racist social system based on differences between racial groups. Althusser (2006) offers a different possibility for ideology, suggesting that perhaps instead of offering a twisted version of a fundamentally equal world, ideology could be an accurate representation of a twisted world. In other words, ruling class cultural hegemony sprouted from their powerful positions. While Gramsci (Bates 1975) argued that working classes could resist cultural hegemony with their own ideology, Althusser is more pessimistic.

Another thread of research that connects large scale ideology to individual and structural action is Fraz Fanon's work on colonial discourses (Fanon 1952). Sylvia Wynter (2001), building on Fanon's work, creates a theoretical connection between the enlightenment conception of Man

as white and male to justifications of colonization, and finally the production of racial categories which separate Man as fully human from non-white subjects as not-quite-human. Wynter (2001) uses Fanon's term sociogeny to describe this as not simply a single historical moment, but a constantly regenerating process of racialization whereby the racial world is re-written as the world 'objectively' discovered–but truly sociogenically created–by scientific observation. Katherine McKittrick (2006) develops Wynter's ideas by framing racial ideology as a specifically spatial project that helps show not only the meaning of places but also the meaning of ourselves: not just 'this isn't the place for me', but at the same time 'I'm not a person who belongs here'.

All of these theories rely on a connection between the macro and micro social worlds. Structural forces exist at the macro level, as do large-scale racial ideologies. However, the harms of racist structure and the utterances of racial ideologies exist at the level of individual homes lost and advertisements written. This paper attempts to explicitly bridge those levels by looking at the variation of micro-level housing advertisement discourse across the macro level of metropolitan areas. Ray (2019) and others (Golash-Boza 2016; Rucks-Ahidiana 2021) have pushed recent theory on race in the US into an interest in the mezzo level that joins the micro and macro and helps complete the reciprocal circle. This means that properly understanding the working of this structure-ideology connection means attuning to the ways it might vary or stay consistent in different contexts. Put another way, this ideology, though it is racialized and hegemonic, may not be total: there may be cites where it is weaker. Alternatively, racial ideology in the United States may be so pervasive and embedded that it seems relatively stable across place. In the context of the analysis below, is this language consistently associated with certain racial groups across metropolitan areas, or are there some places where racialized language is

consistently rare? If there are such places, what features do they have as municipalities that might explain the reduced prevalence of racialized language, and could those be adopted as techniques for reducing that kind of language elsewhere?

Neighborhood Valuation and Exclusion

This paper investigates two threads of racialized ideology in the U.S. in the context of online rental markets, using home rental advertisements as samples of native discourse. These texts are subject to both the broad racist ideology of the United States, but also the specific racialized contexts of both city and neighborhood. First, I focus on language relating to neighborhood quality or value. Second, I examine language related to both explicit and implicit exclusion.

Perceptions of neighborhood quality and race are intertwined in the United States in part because of the practice of redlining, in which the U.S. government labeled space by quality using explicitly racial justification (Rothstein 2017). Not only does that racist (de)valuation have impacts on contemporary life (Graetz and Espositio 2022), but racialized neighborhood valuation continues today (Kennedy, Chapter 1 this dissertation). However, neighborhood valuation happens within a local context where neighborhoods are compared to those around them, to those within the same city. This paper examines the extent to which observed racialized neighborhood valuations persist or vary across cities.

Neighborhood valuation is important because it seems to influence how people perceive the affordability of their rent and how they perceive disorder (Sampson 2004). If contemporary racialized neighborhood valuation, as reflected in housing advertisements, represents a version of the process that produced redlining, it is conceivable that it could have similar repercussions for racial disparities in life chances (Graetz and Esposito 2022). This kind of discourse is also clearly

connected to Gramcian ideology: a twisted representation of the world in order to justify continued inequality. Similarly, it reflects most directly a racialization of space that could reflect the pervasive anti-Blackness of American racial ideology.

However, it is possible that the implementation or manifestation of that ideology is unequal across space, and that such manifestation could be associated with measurable policy, historical, or compositional features of cities. If that is the case, identifying places which seem to be less embedded in anti-Black ideology could help develop methods for resisting that ideology more generally at the mezzo level.

A second pervasive and important kind of discourse is language of exclusion. While neighborhood valuation could sort home-seekers by shaping their perceptions of places, exclusionary language can directly exclude people with Housing Choice Vouchers (HCVs), with low incomes, with past criminal legal system involvement, or other negative credentials. Discrimination against people in those groups is still common (Turner et al. 2013; Turner et al. 2002). Even when explicit discrimination does not occur, language around these topics could shape people's perceptions of those neighborhoods. Krysan and Crowder (2017) document people adjusting their housing search based on their perceptions of neighborhoods, sorting themselves by race and socioeconomic status, perpetuating structural inequality.

Like with neighborhood valuation, it seems likely that exclusionary discourse may not be equally common in different urban contexts. Specifically, certain states and municipalities have explicit policies agianst source of income discrimination, against discrimination agianst people with past criminal legal system involvement, and designed to support tenants by controlling or limitiing rent increases. These factors could influence the prevalence of exclusionary language, and studying that variation could potentially lead to improved policy choices.

The Role of Text Writers

The people who write rental advertisements, often landlords and property managers, are embedded in the racial structure and ideology that are part of the U.S.'s racialized social system. As they compose rental advertisements, they do so with some of the same motivations as real estate actors who are processing applications and selecting tenants (Korver-Glen 2017, Besbris 2020). That is, text writers want to create an advertisement that will be as attractive as possible to 'good' tenants who will pay rent on time, and dissuade 'bad' tenants, people who cannot or will not pay rent on time, who might cause damage to the unit, who might disturb other tenants, or who might be problematic in other ways. But these categories are not neutral in terms of race. First, text writers have a racialized and classed understanding of what kinds of people already live near the unit, and racial and class assumptions about what kinds of people might want to live near the unit. They may also have racist or classist ideas about what kinds of people make 'good' or 'bad' tenants. Together, these two kinds of ideas may cause text writers to produce advertisements with racialized language–racialized because of its spatial association with racialized neighborhoods (Kennedy et al. 2021)–even while the text writers use no explicitly racial or classed language, and perhaps even without consciously thinking race. In other words, it could produce an unintentional but deeply racialized production of discourse that seems almost subconscious.

Contextual factors like segregation and rental policy could reasonably impact that process of writing advertisements. In highly segregated metropolitan areas, landlords might believe that tenants would be unlikely to seek a home in a neighborhood with a large majority of racial

others, and that the racial composition of potential tenants would closely match the racial composition of the immediate neighborhood surrounding the unit.

The policies I focus on in this paper, SOI laws and rent control, are explicitly designed to protect tenants. SOI laws generally forbid discrimination on the basis of a prospective tenant's source of income, whether that be from a job, a pension, or from a housing choice voucher. Rent control laws generally limit the amount that rent can increase each year and often include projections for tenants against eviction. These kinds of policies could therefore impact landlord's texts in at least two ways. First, landlords could respond to those laws 'as intended', by refraining from using exclusionary language. On the other hand, tenant protections might make landlords more intentional about their choice of the ideal tenant. In rent-controlled areas landlords realize that tenants may stay in units for very long time periods and so landlords might use all of the discretion they have access to, including in rental advertisements, to guide the matching process between tenants and units.

Local and Hyperlocal Contexts

At base, this paper is about comparing the variation associated with the hyperlocal context of each advertisement, operationalized as a tract, with the variation associated with the slightly larger municipal context, operationalized as a CBSA. The importance of metropolitan context is well established in research on location attainment and segregation (Crowder et al. 2012; South et al. 2011). Metropolitan areas shape how people experience neighborhoods, not least because they provide a reference set of places for people to compare. For text writers, CBSAs are also the broadest definition of rental markets–and they often, though not always, overlap with Craigslist's market definitions. Finally, the metropolitan context is often relevant in terms of policy.

These large contexts are essential if we want to understand how we can effectively limit the impact of racial ideology in the United States. Each of those areas has its local history of that place's relationship to segregation, its history of redlining, and its contemporary policy tools aimed, perhaps, to address those harms.

I consider five metro-level variables: the presence or absence of rent control laws, the presence or absence of a law banning source of income discrimination, metropolitan segregation, the region, and the average walk score of the city. Rent control laws could influence the way landlords or other text writers compose advertisements because they might make the expected tenure of a tenant longer and the expected return over that period lower. Source of income laws should ideally discourage landlords from explicitly discriminating against HCV holders. Segregation might change text writer behavior by changing the population that text writers might expect will apply for their units. In highly segregated metropolitan areas landlords and property managers may expect only people who match the composition of a unit's neighborhood might apply. In less segregated areas text writers might have the opposite expectation, that even in a predominantly Black neighborhood Asian, White, or Latinx homeseekers might be more likely to apply. The region, Northeast, Midwest, West, or South, may be a proxy for the metropolitan area's past and present relationship to racism, insofar as that is different across regions. Finally, the walk score may be a proxy for the city's overall neighborhood quality with respect to material differences and easy access to amenities.

Craigslist Rental Advertisements as Racial Ideology

Craigslist rental advertisements offer a particularly good opportunity to investigate how the racial ideology of the U.S. changes across different cities depending on the racial history and housing policy in those places. First, Craigslist advertisements have a format that allows text

writers to write long descriptions of neighborhoods and lease details without mediation by more specific platform affordances. That means that the text I analyze was written by the same person who posted the advertisement, not produced by a third party or as a result of a series of checkboxes or selections. Second, Craigslist advertisements are discursive records located within active rental markets, a central part of the process through which renters wind up living in a particular unit or neighborhood. As such, they lie at the articulation of both structure and racial ideology. Finally, Craigslist advertisements are linked to rental units through addresses that can be located within space, including within neighborhoods and cities. The advertisements are discursive records of people enmeshed in the process of housing sorting, and linked to the spatial structure of that sorting.

As a result, analyzing the advertisements allows me to examine how language varies across the local context of municipalities and the hyperlocal context of neighborhoods, examining how that variation is explained by covariates at both levels. The aim of that investigation is to uncover those places that might effectively resist the most harmful aspects of anti-Blackness in the United States, and perhaps which local policies might be effective at reducing discriminatory language.

Research Questions

I consider two ways that discourse could vary across metropolitan areas and possible explanations for that variation.

I first examine metropolitan-level variation in exclusionary language. Chapter 2 showed that Cragislist advertisements included language that was either implicitly or explicitly exclusionary. I identified four kinds of language that could exclude people: exclusion against

people with housing choice vouchers (HCVs), tenant expectations that could exclude some people, background and credit checks that could exclude people with past criminal legal contact or people with low credit scores, and finally exclusion through high move-in costs. In this chapter, I return to those kinds of exclusion and look at how their prevalence–not their racialization–varies across metropolitan areas. Put another way, are these strategies of exclusion more common in some CBSAs than in others? Building on that analysis, I examine whether any of that variation in prevalence across those metros can be attributed to metro-level variables, considering segregation, region, population, walkability, and rental policy.

I take this analysis one step further looking at racialized neighborhood discourse. Chapter 1 of this dissertation showed that on average, advertisements described Black neighborhoods less positively than White and Asian neighborhoods. This paper investigates whether that racist ideology is consistently present across metropolitan areas, or if there are some places where the ideology is weaker or different. I then examine if the observed variation in racialized discourse can be attributed to metro-level variables: segregation, region, population, walkability, and rental policy.

Data and Methods

This paper uses native data scraped from the Craigslist rental platform combined with ACS 5-year data ending in 2019, and with metro-level information about segregation, region, population, walkability, and rental policy. Building on the analyses presented in chapters 1 and 2 of this dissertation, which used metro fixed effects to account for metropolitan level differences, this chapter takes those differences as its main focus. I look at how topics of interest in the texts vary across metropolitan areas, how the description of neighborhood quality is racialized differently across metropolitan areas, and finally how that variation might be explained by differences in metropolitan policy, walkability, or segregation. To do so I use three-level hierarchical mixed-effects models with the outcome at the listing level, listings nested within census tracts, and tracts nested within CBSAs.

Data

I break down the data into three categories depending on their level of analysis. Listing level variables include those that were collected when the rental listing was scraped. Neighborhood level variables are ACS 5-year estimates of the housing stock, racialization, and socioeconomic status of the census tracts which contain the units from each listing. Finally, the metropolitan level variables refer to the policies, city-wide average walkability, and segregation of the metro as a whole.

Listing Level Variables

All of the listing-level variables were collected by the National Rent Project at the University of Washington in daily scrapes from the “apts/housing” section of the Craigslist rental platform from 2017 to 2021, from 16 different metropolitan areas. In total the project collected

2.5 million advertisements from those metropolitan areas in that time period. As part of the pre-processing for the topic modeling analysis, I use a greedy deduplication process which removes both identical and very similar advertisements from the sample. That process is designed to remove identical or similar units posted repeatedly during the study period if they are described similarly each time, while retaining instances where the text of the listing varies significantly. This reduces the corpus to 505,570 listings across the 16 metropolitan areas, half of which are reserved as a training set for the Structural Topic Modeling or STM (Roberts et al. 2019), and 252,785 of which are the test set used for the analysis below, as shown in Table 1.

Table 1

CBSA	Source of Income Law	Rent Control Law	Metro Population	Metro Dissimilarity	Average Walk Score	n Listings
Atlanta-Sandy Springs-Alpharetta, GA	Present	Absent	5,862,424	0.59	48	6,658
Boston-Cambridge-Newton, MA-NH	Present	Absent	4,832,346	0.65	83	20,822
Charlotte-Concord-Gastonia, NC-SC	Absent	Absent	2,545,560	0.52	26	5,670
Chicago-Naperville-Elgin, IL-IN-WI	Present	Absent	9,508,605	0.75	77	17,441
Colorado Springs, CO	Present	Absent	723,498	0.40	36	3,013
Dallas-Fort Worth-Arlington, TX	Absent	Absent	7,320,663	0.56	46	7,819
Los Angeles-Long Beach-Anaheim, CA	Present	Present	13,249,614	0.66	67	29,194
New Orleans-Metairie, LA	Absent	Absent	1,267,777	0.63	58	9,016
New York-Newark-Jersey City, NY-NJ-PA	Present	Present	19,294,236	0.76	88	54,714
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	Present	Absent	6,079,130	0.67	75	16,204
Providence-Warwick, RI-MA	Absent	Absent	1,618,268	0.58	76	8,353
Riverside-San Bernardino-Ontario, CA	Present	Present	4,560,470	0.46	42	7,860
San Francisco-Oakland-Hayward, CA	Present	Present	4,701,332	0.60	89	11,437
Seattle-Tacoma-Bellevue, WA	Present	Absent	3,871,323	0.52	74	39,513
Urban Honolulu, HI	Present	Absent	984,821	0.43	66	11,420
Worcester, MA-CT	Present	Absent	941,338	0.54	56	3,651
						252,785

Each advertisement was geocoded to the surrounding census tract using either an included address or the coordinates from an embedded map widget. I do not include advertisements which could not be accurately geocoded to the tract level (roughly 6% of all advertisements), or advertisements that are missing the rent or number of bedrooms. The advertisements also include a title, listing text, asking rent, and number of bedrooms. All of these are included in the analysis and descriptive statistics are available in Table 2.

Neighborhood Level Variables

Each advertisement is linked to neighborhood level variables taken from the 5-year ACS estimates with a 2019 end year, gathered from National Historical Geographic Information System (Manson et al. 2021). I include a neighborhood racialization typology which seeks to reflect the extent to which the neighborhood is represented within the racialized social system in the United States (Bonilla-Silva 1997). I mark majority White neighborhoods as 'White.' Neighborhoods with a plurality > 40% of people who marked Black, Asian, or Latinx on the census, and with no other group > 10% were marked with the plurality group. I labeled neighborhoods which did not fit into the White, Black, Asian, or Latinx designation as Other.

In addition to the neighborhood typology, I collect information about the population, socioeconomic status and housing stock of each neighborhood. These variables were selected to reflect features about the neighborhood which might be salient to landlords. The variables are shown with descriptive statistics in Table 2.

Metro Level Variables

The goal of this paper is to first describe metropolitan level variation in discourse about neighborhood quality and exclusion in Craigslist rental advertisements, to see if

metropolitan-level housing policy and segregation might explain some of that variation, and finally to examine whether discourse around neighborhood quality is racialized differently across metropolitan areas. To that end I gather appropriate metropolitan level variables. I use details on housing policy collected by the National Multifamily Housing Council (National Multifamily Housing Council 2022a; 2022b) to measure the existence of rent control laws and laws that ban discrimination on the basis of source of income in each metropolitan area. Finally, I use the 2019 ACS 5-year estimates of tract-level non-Hispanic White and Black populations to calculate the dissimilarity index (Bell 1954, Lieberson 1981, Massey and Denton 1988) for each metropolitan area as a measure of segregation. Descriptive statistics for metro level variables are shown in Table 1.

Table 2

Covariate	mean	sd
Neighborhood Covariates		
Poverty Proportion	0.13	0.10
Log of tract Median Income	11.23	0.47
Share of Adults with College Degrees	0.20	0.10
Share Commuters by Car	0.33	0.14
Share With a Commute over 20 mins	0.35	0.10
Proportion of Owner-occupied Housing Units	0.46	0.25
Proportion of Rentals in Properties with >20 Units	0.19	0.22
Proportion of Units Built After 2010	0.03	0.07
Proportion of Units Single Family Detached Homes	0.38	0.39
Population Density	0.01	0.01
Gentrifying	0.32	0.47
Gentrifiable but not Gentrifying	0.23	0.42

Not Gentrifiable	0.46	0.50
Listing Covariates		
Log of the Listing Rent	7.47	0.53
Number of Bedrooms	2.01	1.12
Neighborhood Type		
Black	0.09	0.31
Latinx	0.14	0.37
White	0.70	0.54
Other	0.08	0.29
Asian	0.06	0.25

Analysis

I use two methods to extract discursive features, first STM (Egami et al. 2018, Roberts et al. 2019), and second iteratively validated keywords. I then use each of those operationalizations as the outcome in mixed-effects hierarchical models to investigate associations.

Text Features

This paper is interested in two broad types of discourse in housing advertisements. First, a measure of how the advertisements describe the quality of the neighborhood, and second how the advertisements might reflect exclusionary language. The overall prevalence of that language in this corpus and its unequal distribution across differently racialized neighborhoods and across neighborhoods with different socioeconomic status was the focus of chapter 1 (neighborhood quality) and chapter 2 (exclusionary language) of this dissertation. This chapter returns to those topics using the same operationalizations.

I operationalize neighborhood quality and some measures of exclusionary language using topic modeling implemented in the stm package in R (Egami et al. 2018, Roberts et al. 2019). I

use a topic labeled ‘Positive Adjectives: Location (50)’ to operationalize descriptions of neighborhood quality. The topics that focus on exclusionary language include topics about ‘Restrictions and Stipulations (38)’, ‘Background and Credit Checks (46)’ and ‘Fees and Deposits (18)’.

An important aspect of exclusion and discrimination in rental markets concerns landlords’ willingness to rent to people with Housing Choice Vouchers (HCVs), often referred to as ‘section 8 vouchers’ or ‘housing vouchers’ by landlords and property managers in my corpus. While there was a topic which picked up on this kind of exclusion, the topic combined language that was exclusionary like ‘no section 8’ with language that was welcoming like ‘section 8 welcome’, or advertisements exclusively for HCV holders, ‘2bdrm voucher required’. This language fits into one topic because topic models cannot recover meaning that derives from word order, so cannot differentiate between “We welcome pets, no section 8” and “No pets, we welcome section 8”. To address this limitation of the topic modeling approach, I developed, in collaboration with housing advocacy and legal aid groups, a set of keywords which reliably identified exclusionary advertisements. More information on this process is available in Appendix VII.

Mixed Effects Analysis

I use three level mixed-effects models to examine the variation in discourse about neighborhood quality and exclusion across metropolitan areas. For discourse about exclusion, I run three models for each discursive theme, two to assess variation across metropolitan areas and one to assess association with metro-level variables. For discourse around neighborhood quality, I fit five models: all three used for metro-level variance and two more models that estimate

random slopes for neighborhood racialization. For more information on these choices and comparisons with frequentist results (which are substantially similar) see Appendix VIII.

As an example of the staged modeling strategy, consider my examination of neighborhood quality. First I examine just the variation in the neighborhood quality topic across metropolitan areas using a model with random intercepts for each metro (not shown). Then I create a similar model that maintains the metro-level intercepts and adds random intercepts for each tract. For models operationalizing discursive features with topic models, the model takes the form shown in (1).

(1)

$$Log(TopicPrevalence_i) \sim Normal(\mu\ ,\sigma\ )$$
$$\mu = \alpha_{metro} + \alpha_{tract}$$
$$\alpha_{metro} \sim Normal(\mu_{metro}, \sigma_{metro})$$
$$\alpha_{tract} \sim Normal(\mu_{tract}, \sigma_{tract})$$
$$\mu_{metro} \sim Normal(0,\ 1)$$
$$\mu_{tract} \sim Normal(0,\ 1)$$
$$\sigma_{metro} \sim Exponential(1)$$
$$\sigma_{tract} \sim Exponential(1)$$

Where α_{metro} is a metro-level random intercept, and α_{tract} is a tract-level intercept. Both have priors on their means of $Normal(0,\ 1)$ since the topic prevalence is normalized and demeaned.

Next I model the proportion of the total variance, and of metro-level variance that is accounted for when adding metro-level variables. This model is shown in (2)

(2)

$$Log(TopicPrevalence_i) \sim StudentT(\mu\ ,\sigma\ )$$
$$\mu = \alpha_{metro} + \alpha_{tract} + \mathbf{B}\ \mathbf{X}_i$$
$$\alpha_{metro} \sim Normal(\mu_{metro}, \sigma_{metro})$$
$$\mu_{metro} \sim Normal(0,\ 1)$$
$$\mu_{tract} \sim Normal(0,\ 1)$$
$$\sigma_{metro} \sim Exponential(1)$$
$$\sigma_{tract} \sim Exponential(1)$$

Where α_{metro}, μ_{metro}, and σ_{metro} are as in (1), $\mathbf{B}\ \mathbf{X}_i$ are a vector of coefficients (with Normal(0,1) priors) and a matrix of metro-level variables. These models so far are used for all of the discursive themes. The two models described below are used only to examine whether the racialization of discourse surrounding neighborhood quality is consistent across metropolitan areas.

I fit two models, both of which add random slopes, nested within metropolitan areas, for the neighborhood typology. One of them includes no other covariates and is not shown. The other is the model shown in (3) and also includes covariates at all three levels.

(3)

$$Log(TopicPrevalence_i) \sim Normal(\mu\ ,\sigma\ )$$
$$\mu = \alpha_{metro} + \beta_{ntype,\,metro} + \mathbf{B}_1\mathbf{X}_i$$
$$\alpha_{metro} \sim Normal(\mu_{metro}, \sigma_{metro})$$
$$\beta_{ntype,\,metro} \sim MultiNormal(0,\ \rho\ ,\ \sigma_{ntype})$$
$$\rho\ \sim LkjCorr(2)$$
$$\mu_{metro} \sim Normal(-\ 4,\ .7)$$
$$\sigma_{metro} \sim Exponential(1)$$
$$\sigma_{ntype} \sim Exponential(1)$$

Where all is as in (2) and $\beta_{ntype,\,metro}$ is the random slope for each neighborhood type and each metro area, with a $MultiNormal(0,\ \rho\ ,\ \sigma_{ntype})$ prior, where ρ is a covariance matrix with a LKJ Correlation prior with $\eta = 2$.

The analysis of HCV exclusionary language uses keywords to label texts as exclusionary or not, rather than using a topic. The modeling is broadly similar, but the outcome is a dichotomous variable indicating whether the advertisement is explicitly exclusionary or not. Therefore, I model that variable using a binomial distribution with a logit link. Details and model definitions are available in Appendix VIII.

Results

My analysis shows that discursive features related to neighborhood quality and exclusion do vary by metropolitan area. However, the variation at the metropolitan level is generally much less than the variation at the tract level. In some cases, exclusionary discourse is also associated with metropolitan-level characteristics, including the region, the metro population, the level of segregation, and the presence of rent control or SOI laws. Metro-average walkability was not significantly associated with any of the exclusionary or racialized discursive themes discussed.

In the case of racialized discourse about neighborhood quality, there is some variation in how that language is racialized across metropolitan areas. However, that variation is lessened with the inclusion of tract-level variables. In the fully-specified model all metropolitan areas shared a significant contrast between positive language in Black and White neighborhoods.

I first provide an overview of the goodness of fit statistics for models with various specifications. Then I examine the particular metro-level associations for different kinds of

exclusionary discourse. Finally, I examine racialized discourse about neighborhood quality and how that racialization changes–or stays consistent–across metropolitan areas.

Goodness of Fit

Table 3 shows various goodness of fit metrics for each discourse type across three model specifications. I consider four types of discursive exclusion: 'Fees and Deposits (18),' a topic dominated by language about move-in-fees; 'Restrictions and Stipulations (36),' a topic focused on tenant responsibilities; 'Background and Credit Checks (48),' a topic related to tenant screening; and 'HCV exclusion,' a series of search strings which identify advertisements that explicitly exclude HCV holders. See Chapter 2 for more details on those four discursive themes and their operationalization. I also examine 'Positive Adjectives (50)' a topic focused on positive descriptions of neighborhoods, which is racialized such that it is generally more common in White and Asian neighborhoods and less common in Black neighborhoods. See Chapter 1 for more information about that topic and its racialized dynamics.

For each of those 5 discursive themes, I consider three model specifications. First, a model with random intercepts at the metropolitan level labeled 'CBSA Random Intercepts.' This model aims to examine how much variation in the topic is present at the metropolitan level. Next, a model with both metropolitan and tract-level random intercepts labeled 'CBSA RIs + Tract RIs,' which compares the amount of variation explained at tract level to the amount explained at the metropolitan level. The third model 'RIs + Metro Vars' adds five metropolitan level variables: region, metropolitan population, metropolitan segregation, the presence of a SOI law, the presence of rent control, and the average walk score for the metropolitan area.

Table 3

Fees and Deposits (18)	**σ2**	**τ00 Tract**	**τ00 CBSA**	**ICC**	**Conditional R2**	**BIC**
Topic 18 ~ CBSA Random Intercepts	0.73[3]		0.17	0.19	0.191	637859.1
Topic 18 ~ CBSA RIs + Tract RIs	0.61	0.11	0.15	0.30	0.299	609702.4
Topic 18 ~ RIs + Metro Vars	0.61	0.11	0.08	0.24	0.382	609797.7
Restrictions and Stipulations (38)	**σ2**	**τ00 Tract**	**τ00 CBSA**	**ICC**	**Conditional R2**	**BIC**
Topic 38 ~ CBSA Random Intercepts	0.93		0.09	0.08	0.084	697892.4
Topic 38 ~ CBSA RIs + Tract RIs	0.81	0.11	0.06	0.18	0.175	678748.7
Topic 38 ~ RIs + Metro Vars	0.81	0.11	0.03	0.15	0.199	678853.7
Background and Credit Checks (46)	**σ2**	**τ00 Tract**	**τ00 CBSA**	**ICC**	**Conditional R2**	**BIC**
Topic 46 ~ CBSA Random Intercepts	0.95		0.06	0.06	0.062	704061.9
Topic 46 ~ CBSA RIs + Tract RIs	0.76	0.19	0.05	0.24	0.239	666963.2
Topic 46 ~ RIs + Metro Vars	0.76	0.19	0.04	0.23	0.25	667072
Positive Adjectives: Location (50)	**σ2**	**τ00 Tract**	**τ00 CBSA**	**ICC**	**Conditional R2**	**BIC**
Topic 50 ~ CBSA Random Intercepts	0.95		0.05	0.05	0.048	704683.5
Topic 50 ~ CBSA RIs + Tract RIs	0.83	0.13	0.04	0.17	0.174	684578.7
Topic 50 ~ RIs + Metro Vars	0.83	0.13	0.03	0.16	0.196	684685.2
HCV Exclusion[4]	**σ2**	**τ00 Tract**	**τ00 CBSA**	**ICC**	**Conditional R2**	**BIC**
HCV Exclusion ~ CBSA Random Intercepts	3.29		1.3	0.28	0.284	25528.8
HCV Exclusion ~ CBSA RIs + Tract RIs	3.29	1.42	1.24	0.45	0.447	24920.8
HCV Exclusion ~ RIs + Metro Vars	3.29	1.43	0.55	0.38	0.452	25008.2

Turning to the table itself, there are some broad similarities across discursive themes. In general a small but not insignificant portion of the total variation is accounted for at the CBSA level, ranging from 6% for 'Background and Credit Checks (46)' to 28% of the variation in HCV exclusion. Invariably, however, more variation is accounted for at the tract level than at the CBSA level, and in many cases tract-level variation accounted for some portion of CBSA-level

[3] Sharp-eyed readers will notice that the variance components sum to one. This is because the topic prevalence variable, which tends to be approximately log-normal (though are actually part of a multinomial Derichlet), has been logged and then scaled to have a mean of zero and a standard deviation of one. Predictions and expected values are inversely transformed in the text and in figures to be on the original scale.

[4] HCV exclusion is operationalized using a set of search strings which mark each advertisement as exclusionary or not. As such, while the other models used Linear Mixed Effects models, the HCV models are Generalized Linear Mixed Effects using a binomial likelihood and logit link to account for the binary outcome.

variation: τ00 CBSA is lower for models including random intercepts at both levels when compared to models with only CBSA intercepts. Finally, while Conditional R2–an estimate of the total variance explained by all random and fixed effects combined–improved for all discursive themes with the addition of metropolitan level variance, other measures of model fit like ICC and BIC, suggested a worse model fit for all models. That is because those metropolitan-level variables do not add information to the model that was not already present in the CBSA and tract-level random intercepts. However, for the purposes of my research question: do metropolitan-level variables explain part of the metropolitan-level variance, this result means we can generally answer in the affirmative: adding metropolitan-level variables reduced τ00 CBSA for all discursive themes.

Below I address each discursive theme in turn and examine which metropolitan-level variables had significant associations.

Fees and Deposits (18)

A key aspect of the advertisements is the inclusion of details about move-in costs–upfront fees and deposits that a tenant must pay as they assume possession of the unit. These themes are captured by a topic I label 'Fees and Deposits (18).' Chapter 2 provides details about the racialized tendencies of this language and some examples.

Figures III.1.1 and III.1.2 show the expected topic prevalence for advertisements from each metro, either with no covariates (18.CSBA Null) or with both tract random intercepts and metro-level variables (18.CBSA Full). While there is significant metro-level variation in the null model, no CBSA intercept is significantly different from the mean in the fully specified model.

Figure III.1.1

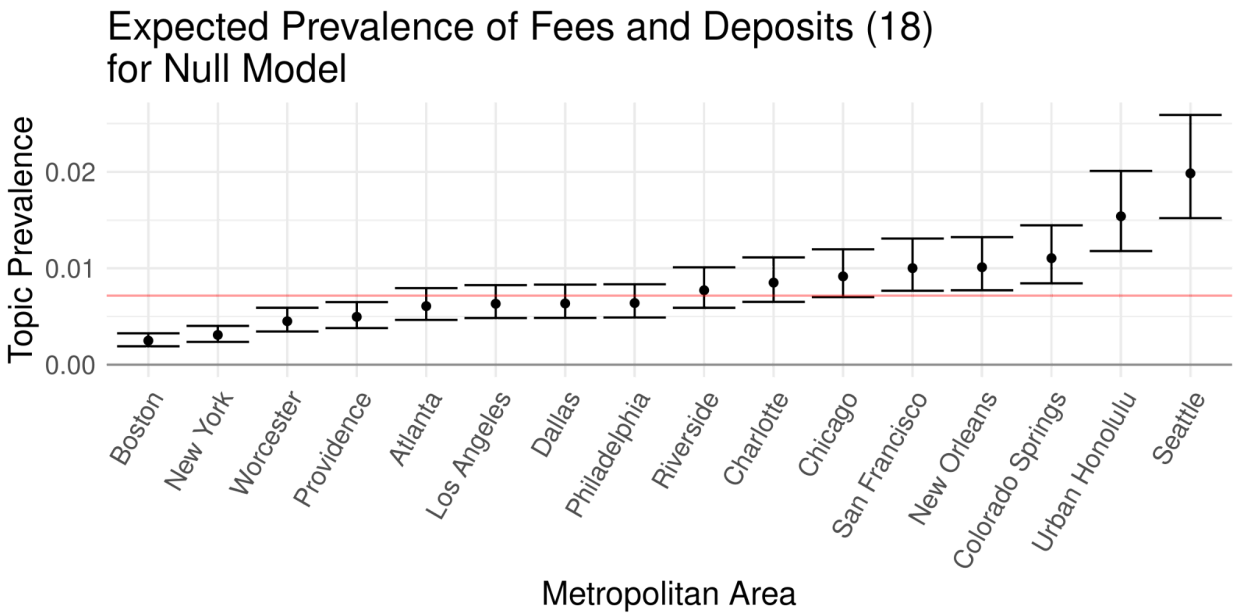


Figure III.1.1 Caption: Figure III.1.1 shows the expected topic prevalence on the y-axis for each metropolitan area on the x-axis in a model with metropolitan-level random intercepts alone.

Figure III.1.2

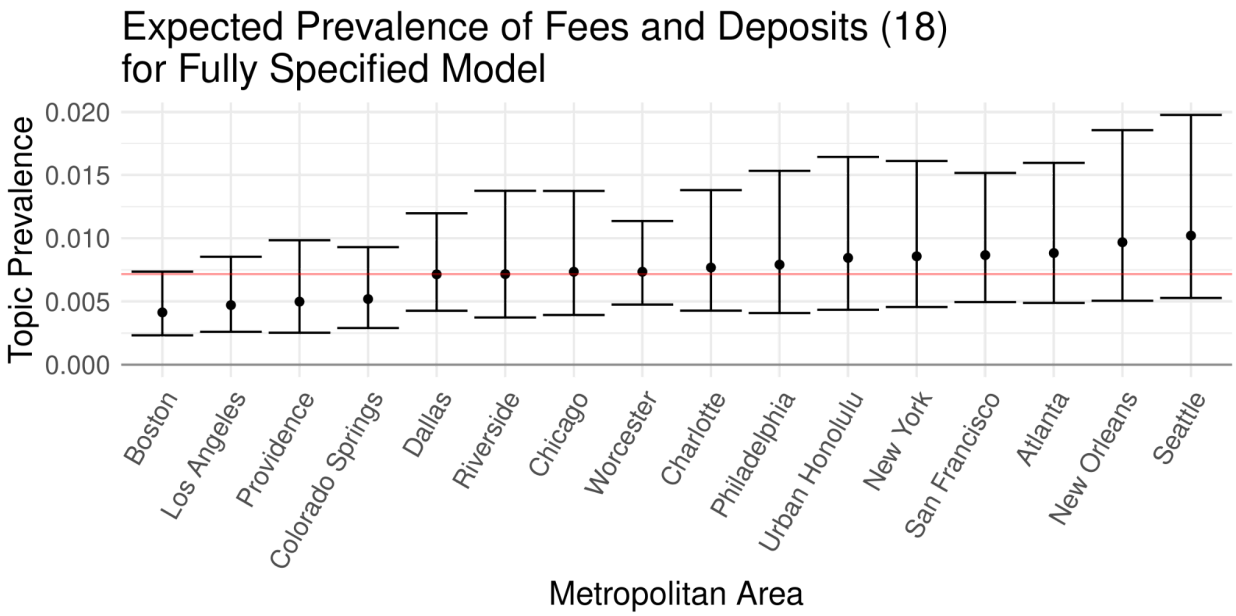


Figure III.1.2 Caption: Figure III.1.2 shows the expected topic prevalence on the y-axis for each metropolitan area on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and metropolitan-level covariates.

Part of that reason is that the metropolitan-level variables explain 46% [(0.15-0.08)/0.15] of the variation at the metropolitan level remaining after accounting for tract-level variation. However, while those variables account for metropolitan variation, they do not, in general, provide clear associations that are obviously statistically or practically significant. Of the metropolitan level variables, only region is significant, and expected topic prevalence for 'Fees and Deposits (18)' is shown in Figure III.1.3.

Figure III.1.3

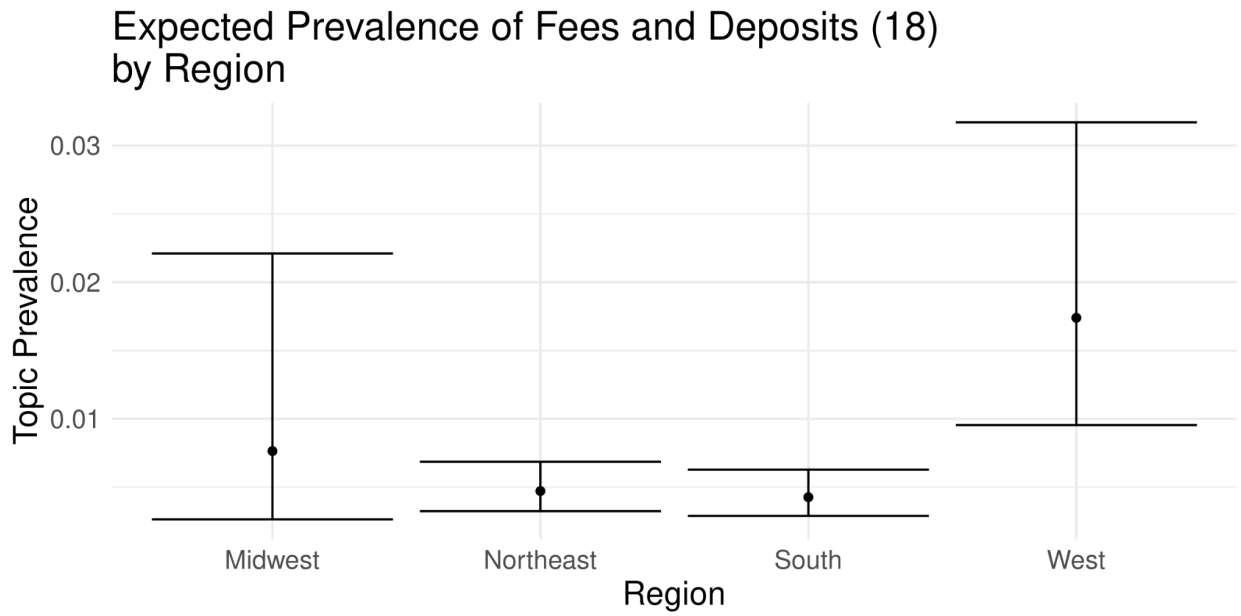


Figure III.1.3 Caption: Figure III.1.3 shows the expected topic prevalence on the y-axis for each region on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and metropolitan-level covariates.

The estimate for Western cities is significantly higher than the estimate for cities from other regions. This difference remains significant even when including other tract and advertisement level covariates (not shown). However, it may reflect the fact that the markets included from the west, from the LA area, the Bay Area, and Seattle, are very tight housing markets where landlords can demand higher move-in costs in general.

Restrictions and Stipulations (38)

As with the discursive features discussed above, I first look at how 'Restrictions and Stipulations (38)'--language related to tenant expectations–varies across the metropolitan areas with a model that just has random intercepts for each metro. Figure III.2.1 shows that variation. While this topic varies significantly by metropolitan area, that variation only accounts for a small

portion, around 8%, of the total variation in the topic prevalence. Adding tract random intercepts and metro-level variables account for two thirds of that variation, reflected in Figure III.2.2.

Figure III.2.1

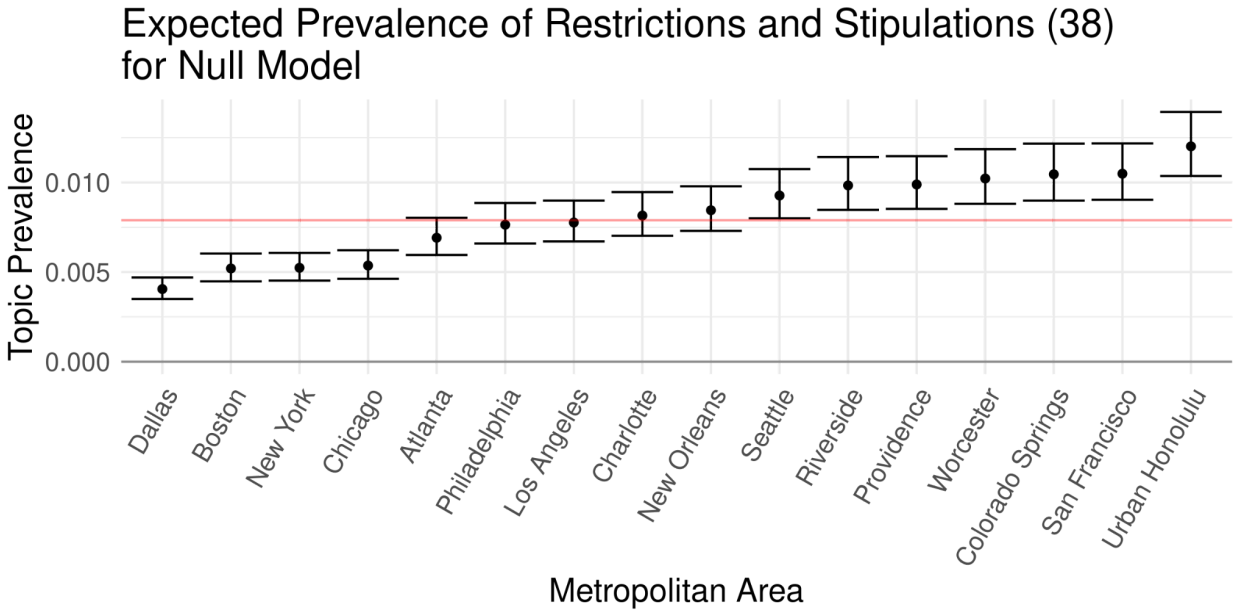


Figure III.2.1 Caption: Figure III.2.1 shows the expected topic prevalence on the y-axis for each metropolitan area on the x-axis in a model with metropolitan-level random intercepts alone.

Figure III.2.2

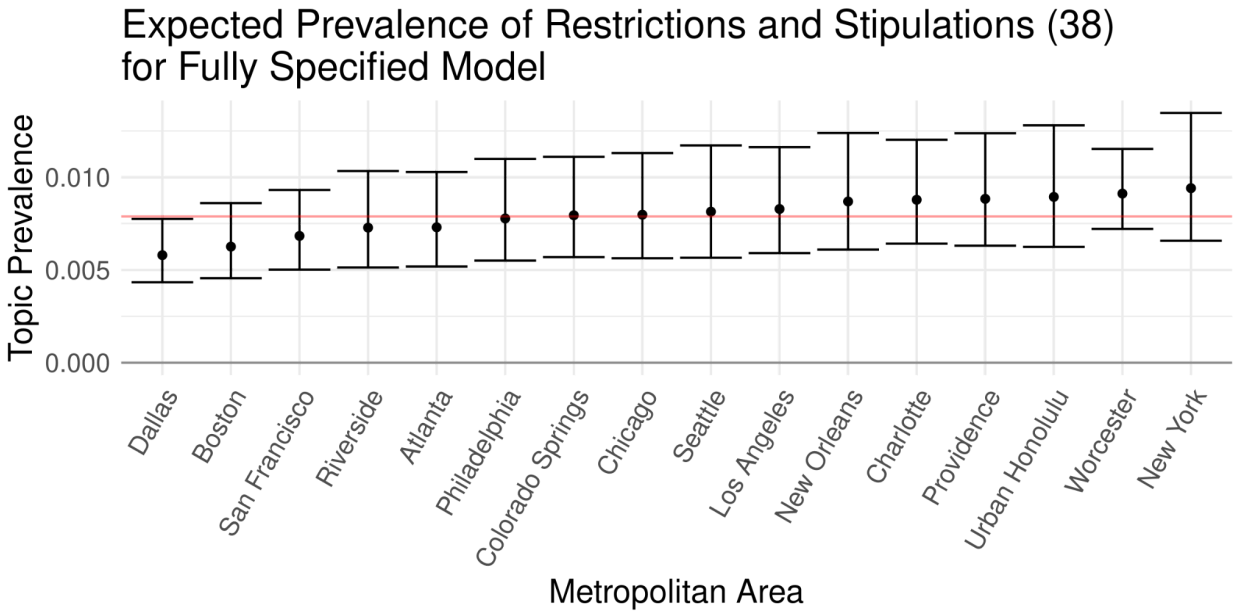


Figure III.2.2 Caption: Figure III.2.2 shows the expected topic prevalence on the y-axis for each metropolitan area on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and metropolitan-level covariates.

The only feature of metropolitan areas significantly associated with this discourse around tenant responsibilities was the metro population, which remains significant even with added explanatory variables at the advertisement and neighborhood levels (not shown). The expected prevalence of this topic in a very large metro area with 20 million people is little more than a third the size of the expected prevalence in a small metro area with 500,000 people. While in Chapter 2 I speculate that this topic could be a way for landlords to signal that the neighborhood is the kind of place where people take care of their homes and yards, this result suggests that it could also in part be about metropolitan-level norms. Landlords could include more of this language in smaller cities if they expect that new residents from larger cities may not understand the expectations for tenants.

Figure III.2.3

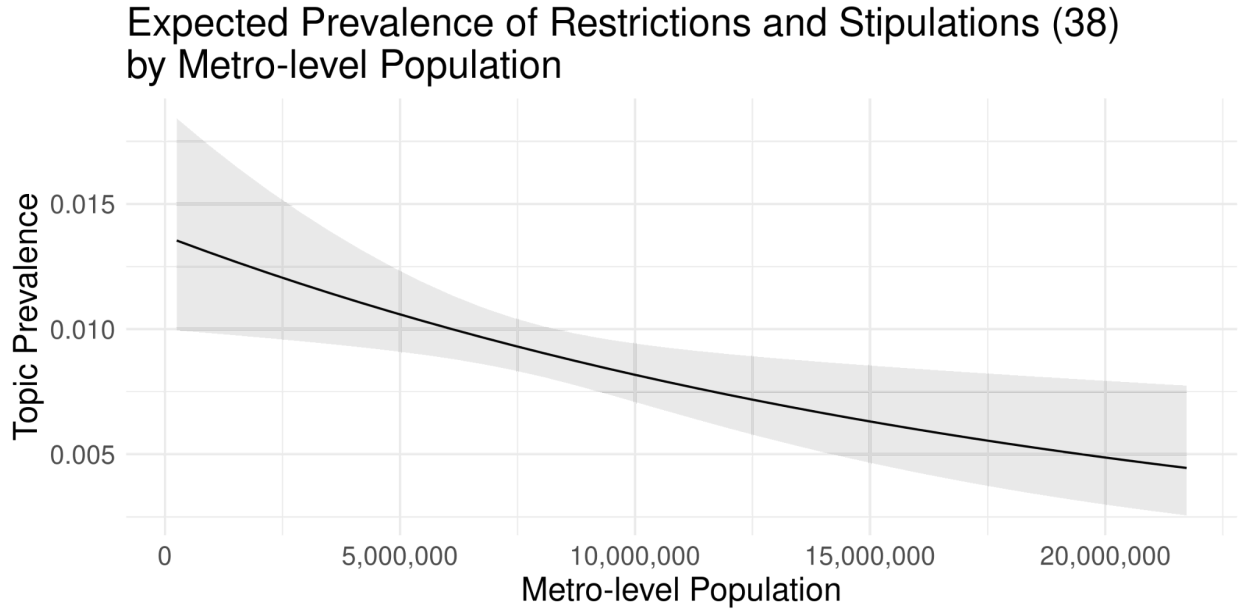


Figure III.2.3 Caption: Figure III.2.3 shows the expected topic prevalence on the y-axis by metro-level population on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and metropolitan-level covariates.

Topic 46 Background and Credit Checks

Credit and background checks are common ways that landlords and property managers screen tenants. One topic, 'Background and Credit Checks (46)' picked up on the language text writers use in rental advertisements surrounding these practices. Ads matching this topic often have detailed descriptions of the application process and include specific requirements for the background and credit check outcomes like 'no felonies' or 'credit over 650'.

As with the other discursive features, I look first at the distribution of this topic across metropolitan areas, and then at how that distribution changes as I complicate the model.

Metro-level intercepts only account for around 6% of all variation in the prevalence of this topic, falling to 4% when tract random intercepts and metro-level covariates are included.

Figure III.3.1

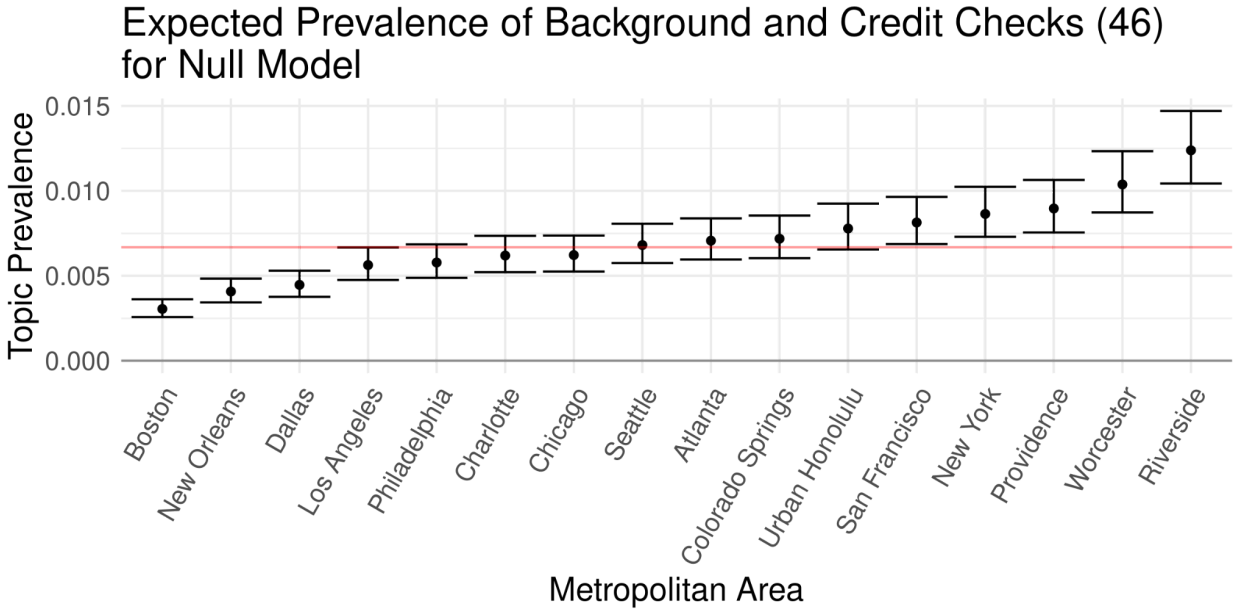


Figure III.3.1 Caption: Figure III.3.1 shows the expected topic prevalence on the y-axis for each metropolitan area on the x-axis in a model with metropolitan-level random intercepts alone.

Figure III.3.2

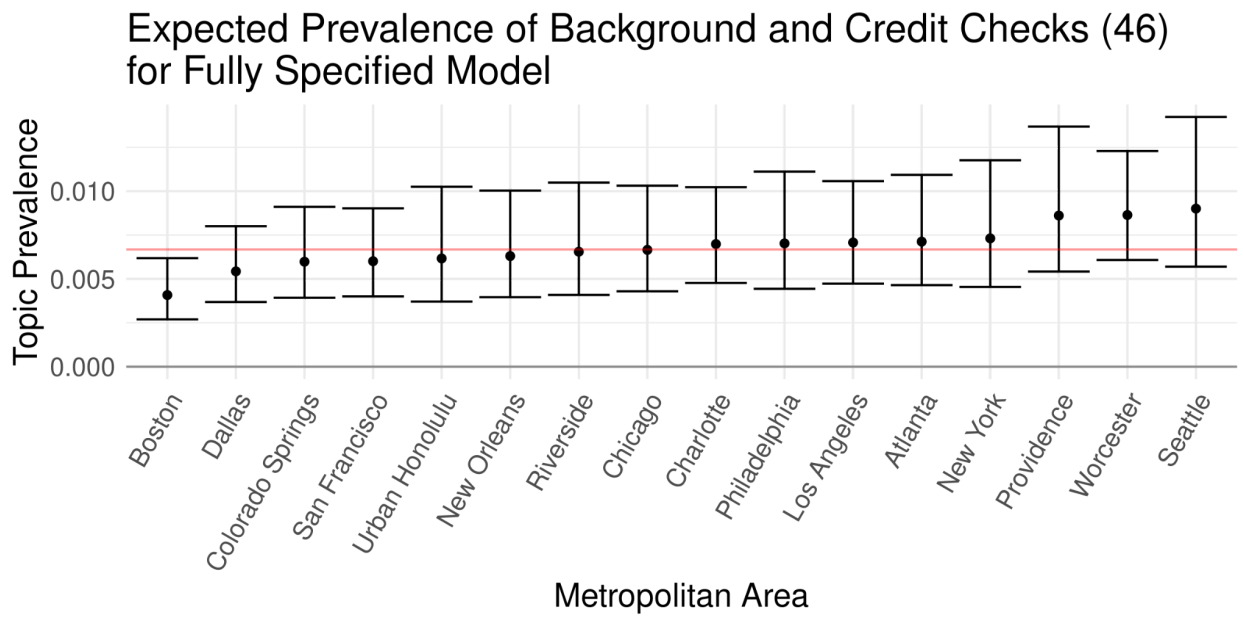


Figure III.3.2 Caption: Figure III.3.2 shows the expected topic prevalence on the y-axis for each metropolitan area on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and metropolitan-level covariates.

However, region, metro-level segregation, and rent control policy all have significant associations with the prevalence of this kind of exclusionary discourse. Landlords and property managers seem to include these kinds of screening techniques the most in the Midwest and the least in the West.

Figure III.3.3

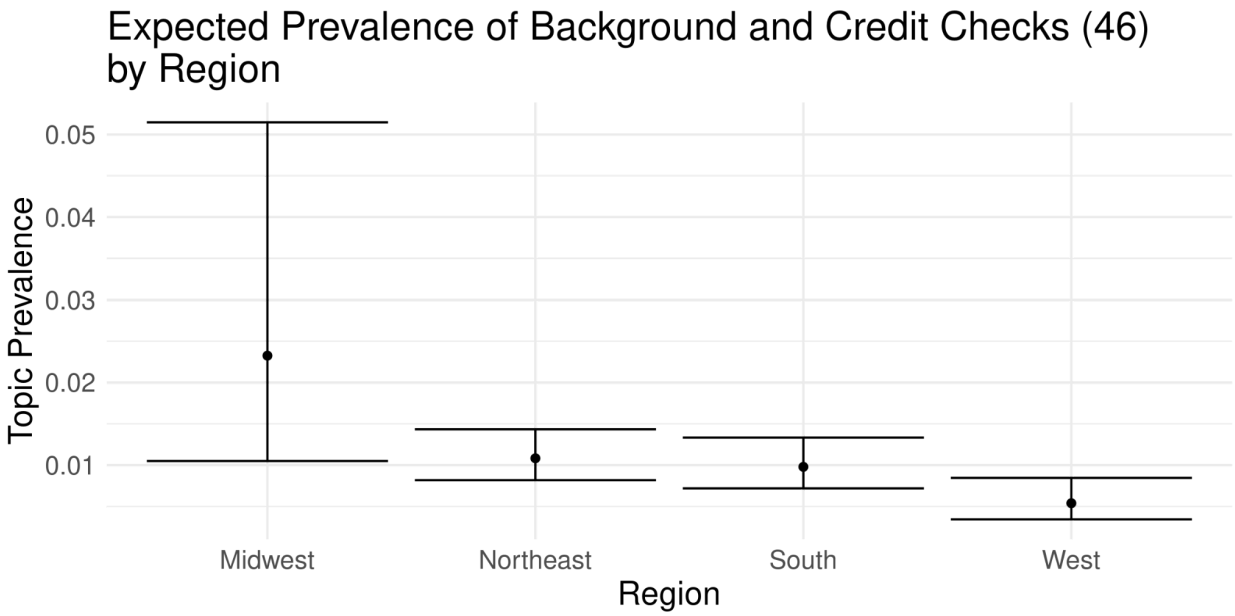


Figure III.3.3 Caption: Figure III.3.3 shows the expected topic prevalence on the y-axis for each region on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and metropolitan-level covariates.

The segregation association, shown in Figure III.3.4 is even more interesting. It shows that the prevalence of this topic declines significantly with the metro-level dissimilarity index. Indeed, the model expects ads from highly segregated metro areas to have less than a quarter of the prevalence of this kind of exclusion as ads from the least segregated metropolitan areas. This aligns strongly with expectation that landlords may use less exclusionary language in more segregated metropolitan area because they expect that, in general, the race and class demographics of potential tenants might more closely match the race and class demographics of the neighborhood in more segregated places.

Figure III.3.4

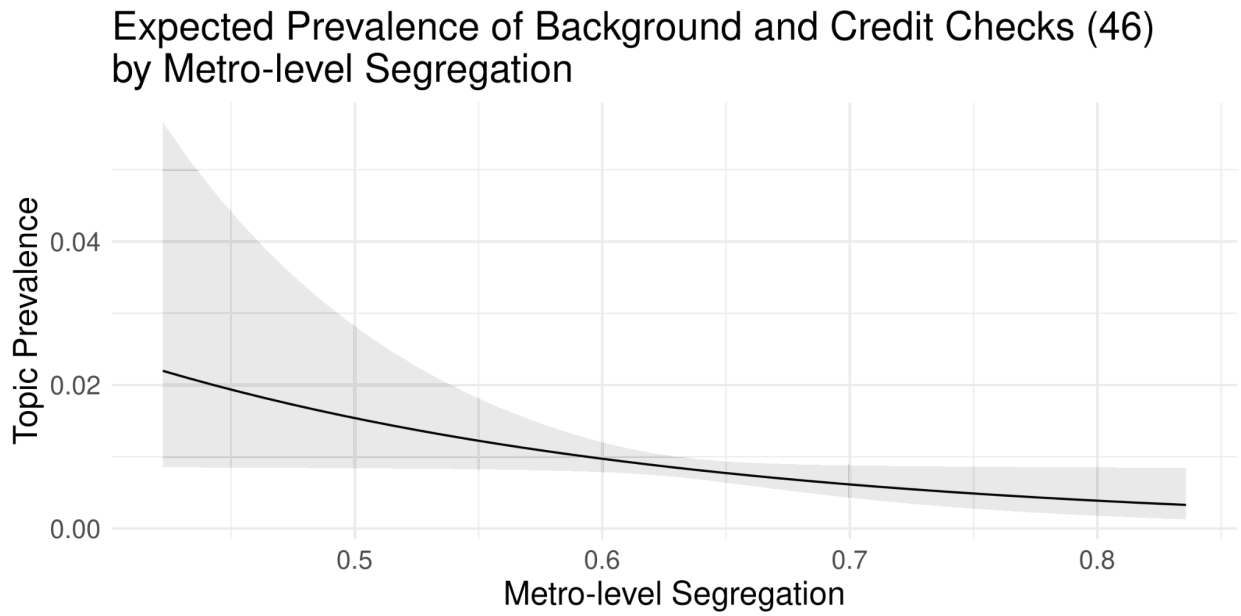


Figure III.3.4 Caption: Figure III.3.4 shows the expected topic prevalence on the y-axis by metro-level segregation on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and metropolitan-level covariates.

Finally, 'Background and Credit Checks (46)' is much more prevalent in cities with rent control than in cities without. This also aligns with the expectation that landlords in areas with rent control might use more exclusionary language as a way of maximizing their discretion during the home search process. That could be because rent control laws limit landlord discretion later on in terms of the rent they charge and often in terms of how easily they can evict a tenant.

Figure III.3.5

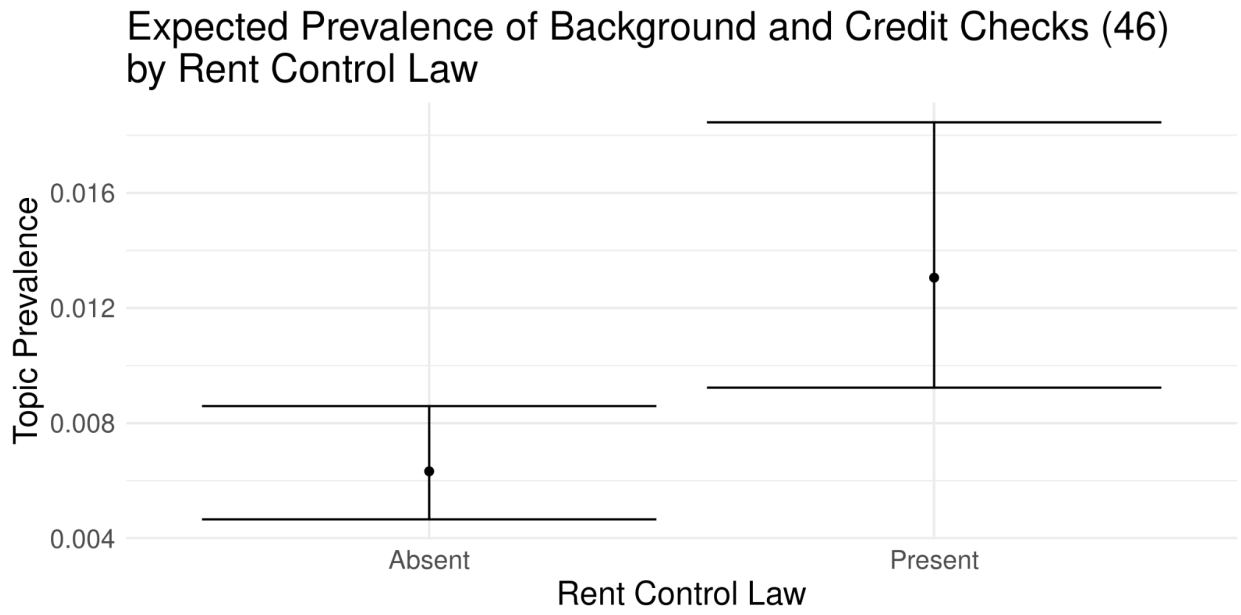


Figure III.3.5 Caption: Figure III.3.5 shows the expected topic prevalence on the y-axis by the presence of a metro-level rent control law on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and metropolitan-level covariates.

Housing Choice Voucher Exclusion

In addition to the three implicitly exclusionary discursive themes presented above, many advertisements exhibited language that explicitly excluded people with HCVs, often referred to as Section 8 vouchers in the advertisements. While there was a topic which focused on language about vouchers, the topic tended to combine exclusionary language with welcoming language. To establish a tighter focus on exclusionary language, I developed a set of search strings which identified exclusionary advertisements with very high accuracy. More information on this process is available in the methods section of Chapter 2 and Appendix VII.

There is strong reason to expect that this language will vary by metropolitan area: while the HCV program is federally-funded it is administered by local housing authorities. Moreover, discrimination based on source of income (SOI discrimination), including HCVs, is prohibited

by state or local law in 12 out of the 16 metros in my study area. I expected that this language would be more common in metropolitan areas without protections for HCV holders.

As with the other discursive features, I subjected my measure of section 8 exclusion to a three-stage modeling strategy. First, I regress the discursive feature on metro-level random intercepts alone.

Figure III.4.1 shows significant variation in the expected prevalence of exclusionary advertisements between different metropolitan areas. Indeed, the metro-level intercepts alone account for around 28% of the total variation in HCV exclusion. Honolulu has by far the highest expected prevalence with nearly 8% of advertisements expected to be exclusionary, followed by Riverside and New Orleans. Boston, New York and Chicago have a very small expected proportion of exclusionary advertisements. Once accounting for tract-level intercepts, however, only Boston and Colorado Springs still have expected values significantly lower than the population mean.

Figure III.4.1

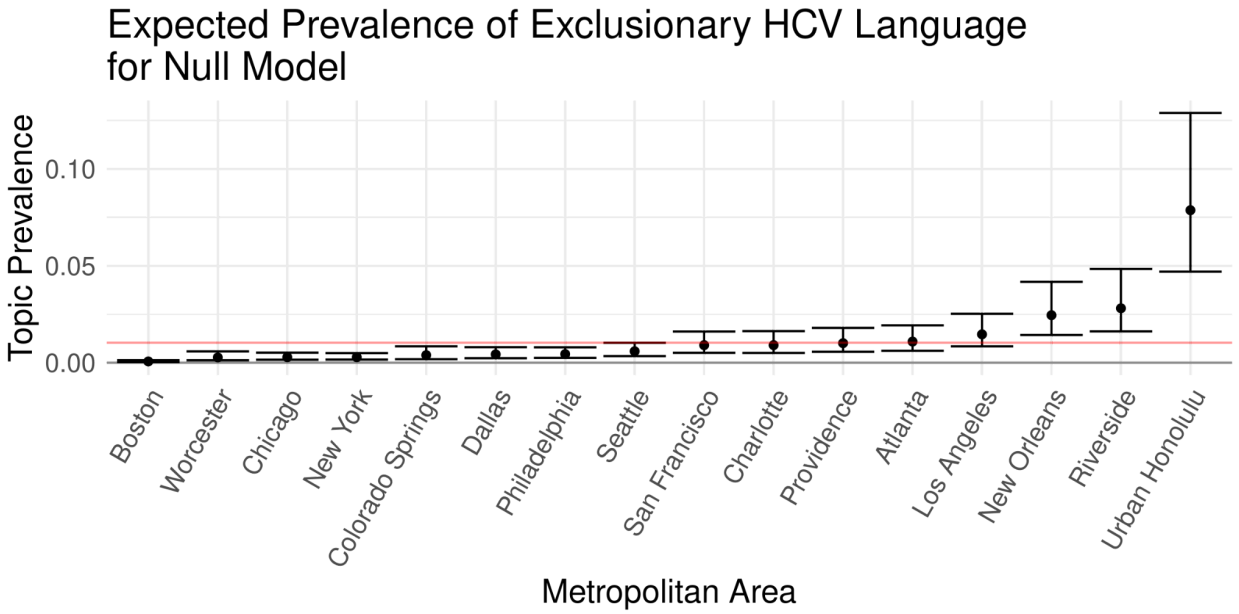


Figure III.4.1 Caption: Figure III.4.1 shows the expected probability of HCV exclusion on the y-axis for each metropolitan area on the x-axis in a model with metropolitan-level random intercepts alone.

Figure III.4.2

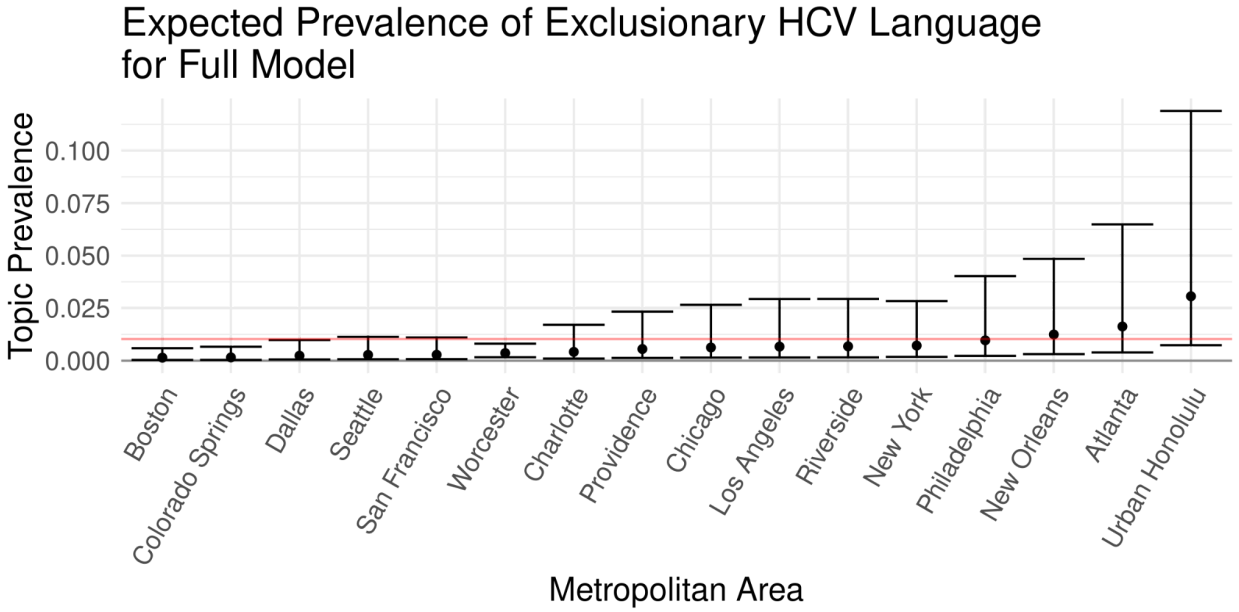


Figure III.4.2 Caption: Figure III.4.2 shows the expected probability of HCV exclusion on the y-axis for each metropolitan area on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and metropolitan-level covariates.

In the model with tract- and metro-level intercepts and metro-level covariates, HCV exclusionary discourse is significantly less prevalent in cities with an SOI law than in cities without one. None of the other metro-level variables–population, segregation, rent control, or walkability–had a significant association. This aligns with the expectation that SOI laws reduce this kind of discrimination. However, this association is not significant when ad and neighborhood level variables are included in the model, so this association might just reflect compositional differences in metropolitan areas with and without SOI laws.

Figure III.4.3

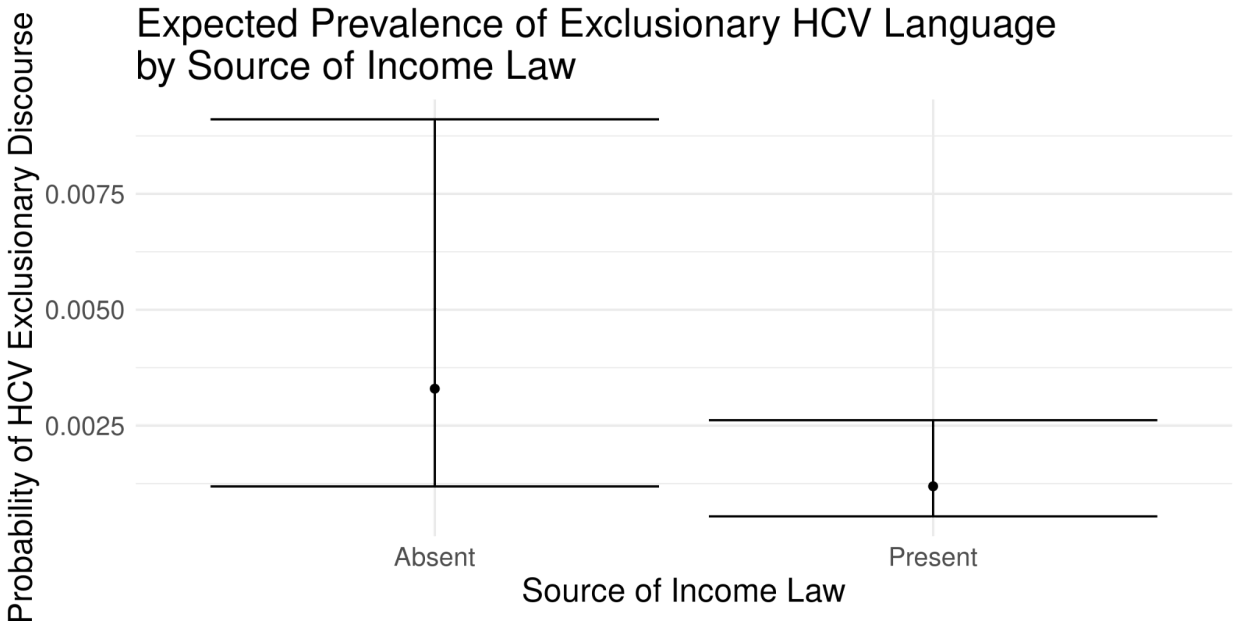


Figure III.4.3 Caption: Figure III.4.3 shows the expected probability of HCV exclusion on the y-axis by the presence of a metro-level SOI law on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and metropolitan-level covariates.

Topic 50 Positive Adjectives: Location

While my research questions about exclusionary language focused on differences in prevalence across metropolitan areas, my interest in racialized discourse about neighborhood quality focuses on differences in effect. Specifically, is this kind of language racialized differently in different places, or does it reflect a nation-wide ideology of anti-blackness?

Racialized housing dynamics in the US, including segregation and gentrification, are and have historically been tied to racialized notions about housing quality. Redlining was a legal practice whereby neighborhoods were marked nominally by quality, but that conception of quality was strongly driven by racist interpretations of who lived in the neighborhoods (Rothstein 2017). Gentrification also carries implicit notions about what kinds of people make a quality or authentic neighborhood (Rucks-Ahidiana 2021; Hwang 2016). A fuller discussion of this background is available in Chapter 1.

I operationalize descriptions of neighborhood quality by using a topic I label 'Positive Adjectives: Location (50).' Advertisements with a high proportion of this topic used phrases like 'location, location, location' and 'great neighborhood' to emphasize the quality of the neighborhood as a feature of the unit for rent.

I begin by examining how that topic varies in expected prevalence across the 16 metros in my study area. A model with only metro random intercepts, shown in Figure III.5.1 expects 'Positive Adjectives: Location (50)' to be most common in New Orleans, San Francisco, and Charlotte, and least common in Riverside and New York. There are no clear patterns based on region and city size. The metro-level random intercepts only account for around 5% of the variation in this topic, lower than for any other discursive theme I investigate.

Figure III.5.1

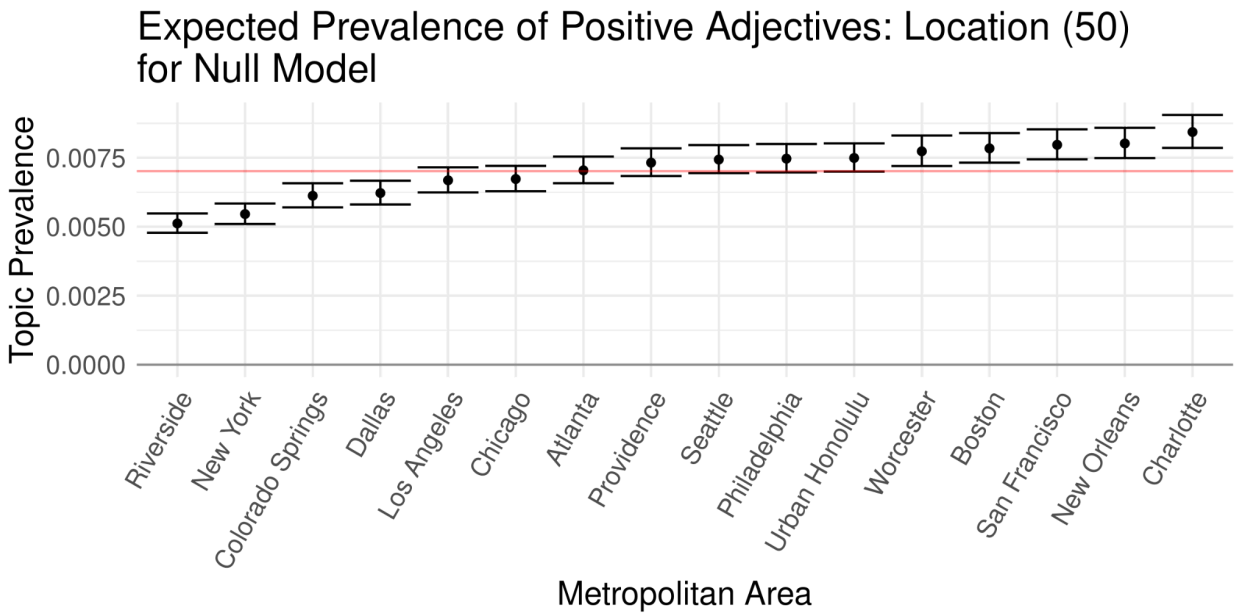


Figure III.5.1 Caption: Figure III.5.1 shows the expected topic prevalence on the y-axis for each metropolitan area on the x-axis in a model with metropolitan-level random intercepts alone.

Looking at Figure III.5.1, we can see that though the expected values for some metropolitan areas are significantly different from the population mean, those differences are small.

Figure III.5.2

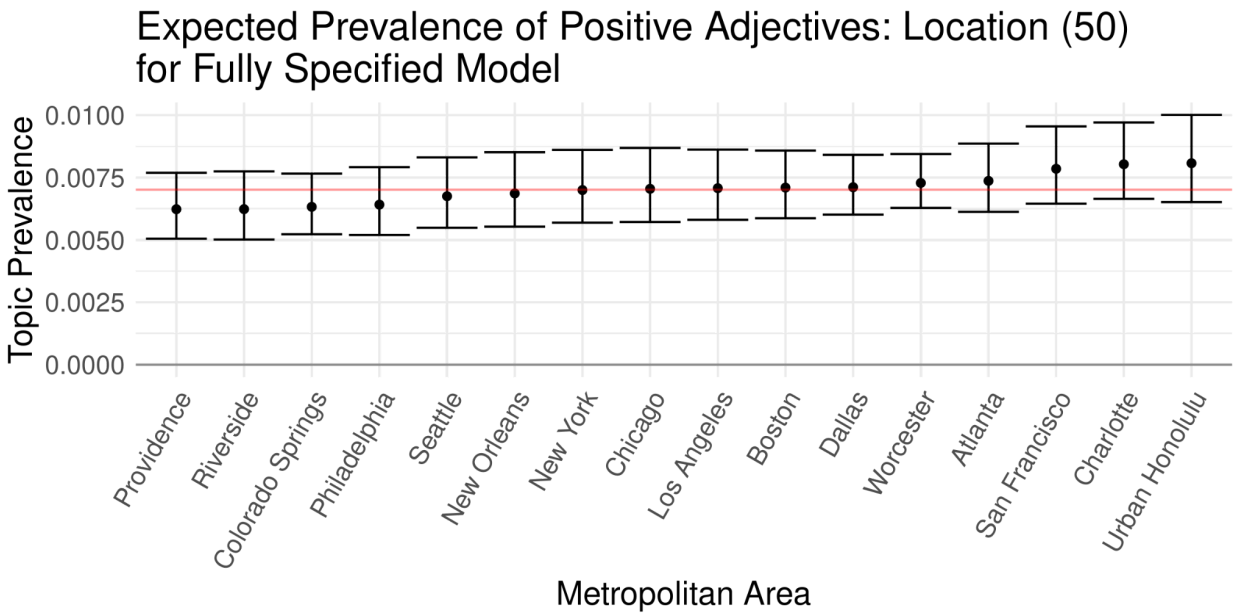


Figure III.5.2 Caption: Figure III.5.2 shows the expected topic prevalence on the y-axis for each metropolitan area on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and metropolitan-level covariates.

Indeed, adding tract-intercepts and metro-level variables shrinks those differences further, so that none are significantly different from the population mean.

To look at differences in racialization across metropolitan areas, I fit new models with random slopes for each of five neighborhood types: Black, White, Asian, Latinx, and Other.

Figure III.5.3

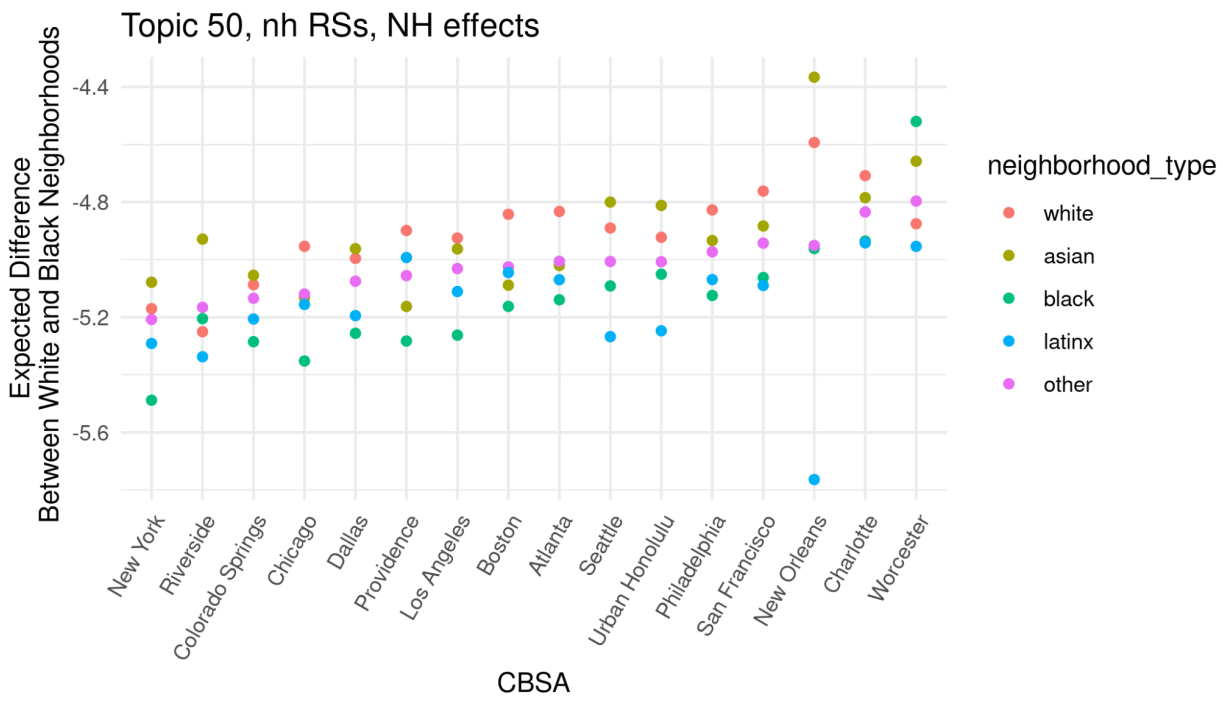


Figure III.5.3 Caption: Figure III.5.3 shows the expected topic prevalence on the y-axis for each metropolitan area on the x-axis, colored by neighborhood type in a model with tract-level random intercepts, metropolitan-level random intercepts, and neighborhood type random slopes.

Looking at a model with random slopes for neighborhood types within each metropolitan area, Figure III.5.3 shows that the variation between neighborhood types is comparable to the variation between metropolitan areas. The estimate for the neighborhood with the highest prevalence in Riverside, the metro with the lowest average prevalence, is higher than the estimate for the neighborhood with the lowest prevalence in Worcester, the metro with the highest average prevalence.

Moreover, there are clear racialized trends here, inline with the findings presented in Chapter 1. White and Asian neighborhoods have consistently the highest point estimates in most metro areas while Black and Latinx neighborhoods have consistently the lowest, with the only exception being Worcester. This suggests that the anti-blackness that was inherent in redlining

seems to be embedded in contemporary racialized discourse across this 16 metro sample. Put another way, this consistent racialization is reflective of a more general dynamic in the contemporary United States in which the composition of the neighborhood is erroneously accepted as an indication of its "quality."

Figure III.5.4

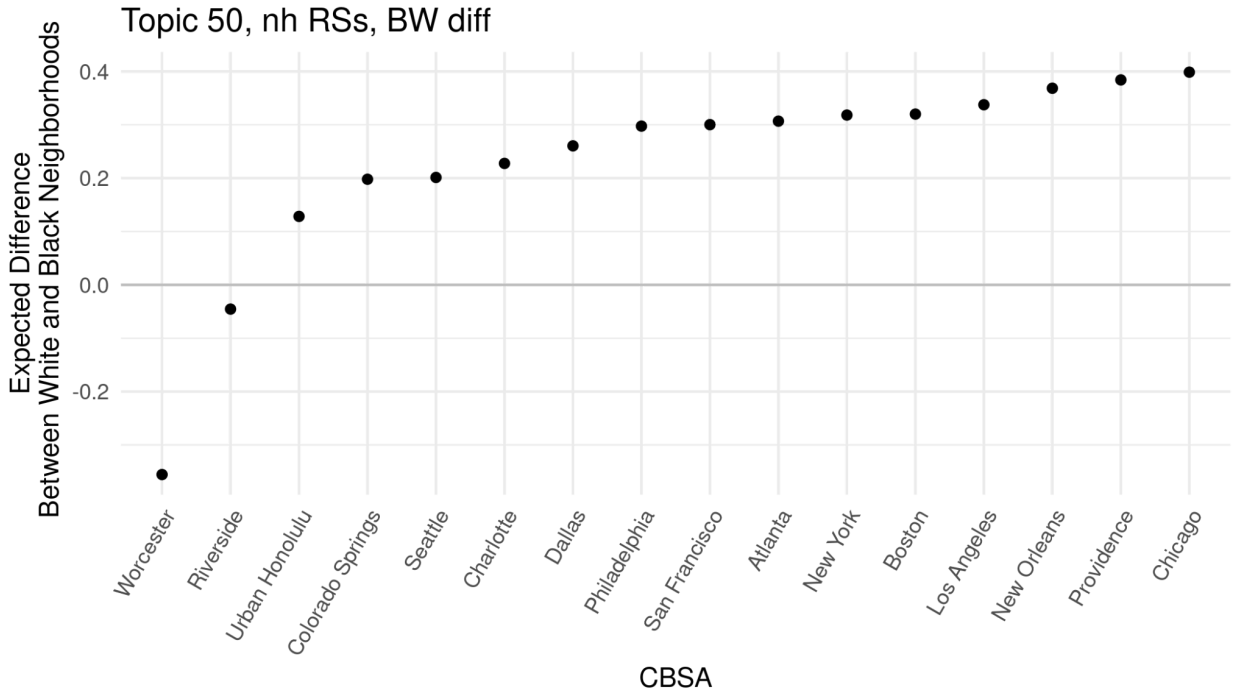


Figure III.5.4 Caption: Figure III.5.4 shows the expected contrast in topic prevalence between a White and a Black neighborhood on the y-axis in each metropolitan area on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, and neighborhood type random slopes.

That theme of anti-Blackness is perhaps even more clear looking at the contrasts. All but two metros reflect higher values for White neighborhoods than for Black neighborhoods in terms of ‘Positive Adjectives: Location (50).’ A core question, however, is whether there is something about Worcester and Riverside that helps resist this trend of anti-Blackness, or if these results

instead might reflect specificities about those metros that we could account for in our modeling strategy.

Figure III.5.5

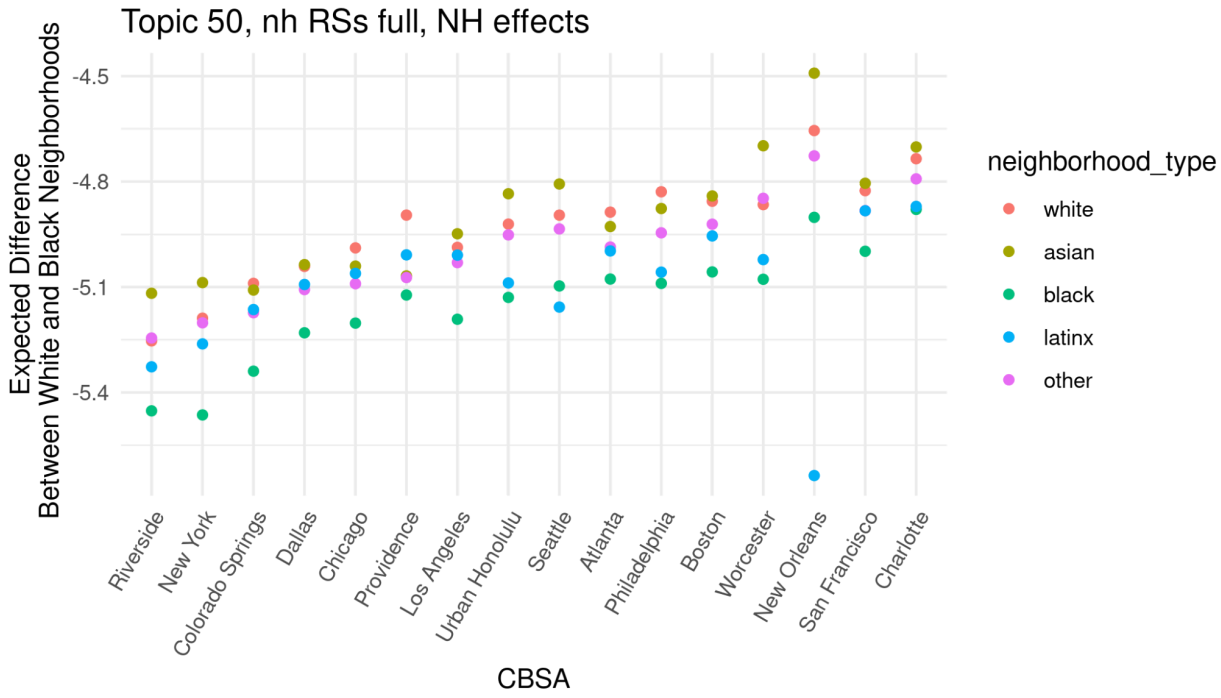


Figure III.5.5 Caption: Figure III.5.5 shows the expected topic prevalence on the y-axis for each metropolitan area on the x-axis, colored by neighborhood type in a model with tract-level random intercepts, metropolitan-level random intercepts, neighborhood type random slopes, and listing-, tract-, and metro-level covariates.

Figure III.5.5 shows a model where I add socio-economic and housing stock controls at the neighborhood level, and housing policy and segregation at the metro-level. With those variables included in the model as fixed effects, the racialization of 'Positive Adjectives: Location' is not only upheld, but in the case of anti-Blackness, may even be more stark. The pattern of Black and Latinx neighborhoods having the lowest expected values, while White or Asian neighborhoods have the highest is also strengthened. However, while the point estimates for Black neighborhoods are significantly lower than for other neighborhoods, the Latinx

neighborhood estimates are now not statistically significantly different from other neighborhoods in most metropolitan areas, with the exception of New Orleans, Seattle, and Honolulu.

Figure III.5.6

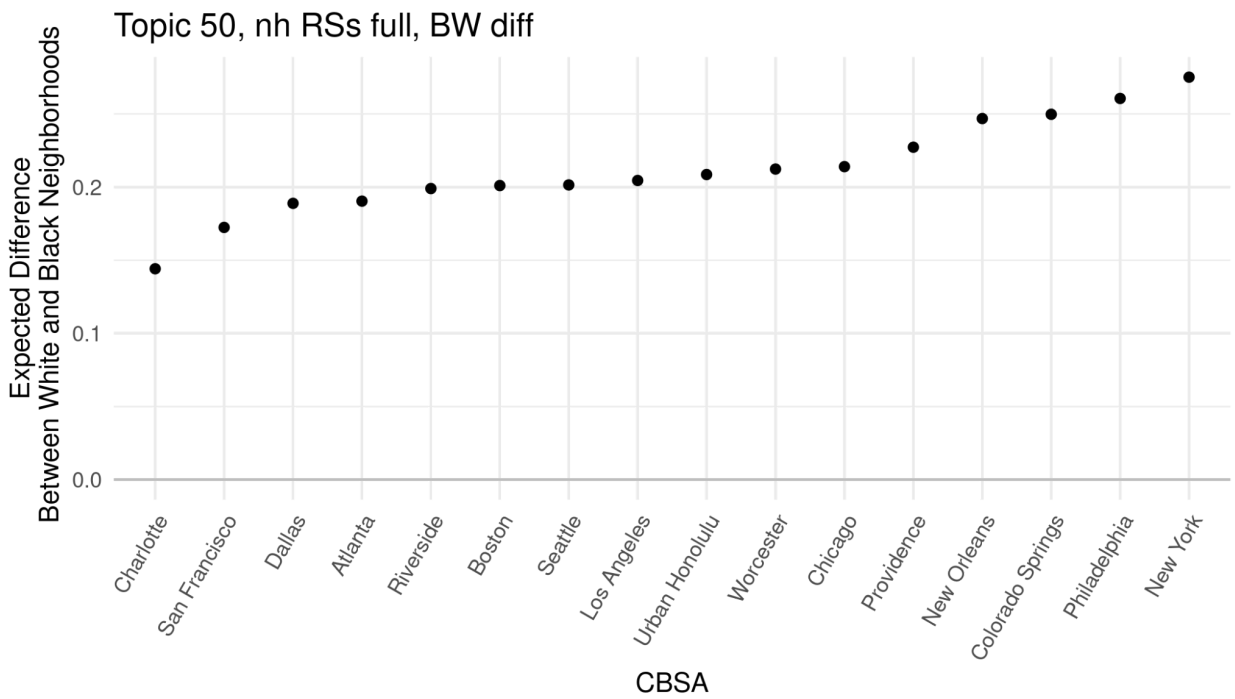


Figure III.5.6 Caption: Figure III.5.6 shows the expected contrast in topic prevalence between a White and a Black neighborhood on the y-axis in each metropolitan area on the x-axis in a model with tract-level random intercepts, metropolitan-level random intercepts, neighborhood type random slopes, and listing-, tract-, and metro-level covariates.

The size of the difference in the estimates of positive adjectives connected to the neighborhood for White and Black is even clearer when we look at the contrasts: while Worcester and Riverside had higher estimates for Black neighborhoods in the model without covariates, in the full model those metros have estimates near the median. That is likely because the other information about the neighborhoods helped allow for a better estimate of the prevalence and racialization of 'Positive Adjectives: Location (50).' Specifically, these tract-level covariates account for compositional and housing stock differences in Worcester and Riverside that likely hid this relationship. Similarly, when accounting for socioeconomic, housing, metro policy, and segregation, the magnitude of these contrasts are smaller, but their variation across metro areas is also smaller.

Discussion

This paper lays out two sets of research questions: first, does exclusionary language vary with the metropolitan area, and is that variation explained by metro-level variables? Secondly, is language about neighborhood discourse consistently racialized across metropolitan areas, or are there some places where racialization is stronger or weaker?

I find that exclusionary language is varied across metropolitan areas. Both move-in costs and tenant screening mechanisms are unequally distributed across regions. Language around tenant expectations is associated with smaller metropolitan areas by population. Tenant screening mechanisms are more common in less-segregated metropolitan areas, and more common in cities with rent control. Explicit exclusion of people with HCVs is less common in cities with SOI ordinances.

While exclusionary language is varied, the anti-Black ideology in discourse connected to neighborhood quality is consistent. White neighborhoods are described more positively in all 16 metropolitan areas when accounting for ad, tract, and metropolitan level covariates and allowing slopes for neighborhood types to vary across metropolitan areas.

Implications for Scholarship

The clearest implication of this paper is that the anti-Black ideology examined in detail in Chapter 1 is not simply a product of certain metropolitan areas, but rather a feature shared across all 16 metropolitan areas in the sample. Hypotheses that anti-Blackness would be more prevalent in metro areas with a strong legacy of segregation, or in metros with stronger connection the legacy of slavery are sensible, but are not supported by this data. In contrast to the other discursive features I focus on, the racialization of neighborhood quality seems to be the most stable across metropolitan areas. This highlights the fact that systemic racism, and the racial

ideology it is reciprocally linked with, is pervasive in the United States, and not limited to rural areas or to particular cities. Instead, the history of racial violence through slavery, Jim Crow, redlining, and school segregation persists throughout this country in the form of continued residential and school segregation, the war on drugs and its aftermath, ideological resistance to critical race theory, mass incarceration, police violence, and white identity politics. I argue that those explicit outcomes of our racialized social system could be undergirded by everyday forms of racialized thoughts, including discourse about neighborhood quality. One area of scholarship that is beginning to investigate this possibility is the measurement of structural or systemic racism literature (Hardeman et al. 2022, Brown and Homan 2022). Future measurement schemas could include the discursive racialization as one of their variables.

I further argue that the harms stemming from this ideology are not properly considered in the realm of symbolic violence (Bourdieu 1979), but rather as part of a racial project of naturalizing and perpetuating racial structures (Omi and Winant 1986; Golash-Boza 2017; Bonilla-Silva 1997; 2006; 2019). Resisting racialized discourses of neighborhood quality therefore should account for that reciprocality and address both investment in Black neighborhoods and intentional ideological adulation for the existing strengths of those neighborhoods.

I also find that exclusionary language is less common in metro areas without rent control and those that are more segregated. This aligns with an understanding of the production of advertisements that considers how landlords and property managers attempt to maximize their discretion (Rosen 2014; 2021; Grief 2018). In more segregated metro areas it seems like those text writers may rely on long standing racialized housing patterns to do some of the exclusion for them. In areas with rent control, it seems plausible that text writers respond to that restriction on

their control of their housing units by maximizing the discretion they have through tenant sorting techniques. Future research in rental housing dynamics should attend carefully to these contextual impacts on landlord and property manager's actions.

While less obviously impactful, I note that this is the first instance I am aware of a systematic multi-level analysis of large scale texts attentive to metro-level variation, and certainly so in the area of rental housing dynamics. However, it is just a start. These texts are a rich source of information about how landlords, property managers, and other people who contribute to this discourse think about space and place, at least in terms of the rental market. Future work could easily build on this beginning by expanding the study area, adding other rental platforms, and expanding the dimensionality by including temporal variation.

Implications for Policy

Municipal housing policy is local by definition. However, a comparative framework like the one presented here has the potential to inform that policy by placing it in the context of differences across metro areas. This paper does not present evidence that racialized descriptions of neighborhood quality are associated with SOI laws, rent control laws, or segregation. Perhaps future policies that more directly address racial ideology will be associated with a decrease in that kind of discourse. On the contrary, this paper adds to evidence presented in Chapter 2 that suggests landlords use advertisements as a space to influence what kinds of people apply for their units, and that landlords are more likely to attempt this strategy when tenant protections are in place.

On the other hand, I do show an association between certain types of exclusionary language and metropolitan context. Policy makers interested in equal access to housing should consider those associations moving forward. Landlords and other text writers might use more

exclusionary language in places with rent control because they worry that they will have less discretion once they sign a lease: tenant selection might feel more important because tenants may stay in the unit a long time and may–if their unit is rent controlled–represent a diminishing source of revenue compared to market-rate units over time. However, I want to explicitly emphasize that this analysis does not suggest that rent control laws are ineffective or should be abandoned. If anything, I argue that such laws should be strengthened and partnered with other tenant protections. At the same time, landlords' and property managers' urge for discretion and care in whom they rent to should be listened to, addressed, and accounted for in policy, but only insofar as they can be part of equitable rental markets. One possible solution would be to make lease agreements, and perhaps even tenant applications, public documents which would allow scholars and municipalities to directly investigate rates of discrimination and to target landlords who engage in it.

The case of HCVs and exclusionary language is especially interesting. Laws prohibiting SOI discrimination seem, in this sample, to have a limited influence on the prevalence of exclusionary language at the metropolitan level. While they remain uninvestigated in this paper, it seems likely that other features of the metropolitan area, like the availability and spatial distribution of affordable housing units, might play the most central role.

On the other hand, the existence of metropolitan areas with strong SOI laws but also high rates of exclusionary language, like in Honolulu, Riverside, Los Angeles, and Atlanta, may suggest an opportunity for increased enforcement of those existing laws. It is clear that having SOI laws on the books is not enough on its own to prevent exclusionary language, and presumably exclusionary action. Municipalities should at the very least ask landlords with exclusionary language in their advertisements to revise their texts, but can go even further by

auditing tenant screening practices to ensure their laws are being followed. One approach could be to encourage platforms not to host illegal advertisements. Craigslist and other platforms have the ability to remove or at least label advertisements that have illegally discriminatory language, and should do so immediately.

Limitations

In its current state, this analysis is limited. While the corpus of texts is large, trying to estimate mixed-effects models with only 16 groups is difficult: it is not clear how these findings would generalize to a larger sample. This is especially true when estimating random slopes for the neighborhood types. While I am confident that these results are internally valid, and there is strong evidence for external validity when results are consistent across metro areas (as for the evidence for anti-blackness in discourse about neighborhood quality, or about background and credit checks and restrictions and stipulations), it is hard to interpret, let alone make larger inferences, when there is variability across metro areas that is not explained by the limited metro-level variables included.

It is likely that those apparent inconsistencies actually reflect interesting and complex processes that produce different outcomes in different places, processes that my current modeling strategy simply cannot recover. For instance, it is possible that large property management companies with holdings in different cities may write similar advertisements in similar neighborhoods in the places where they are active, but obviously not in cities where they have no clients. Even excluding such effects, diffusion of discursive practices across text writers could induce regional confounding. And in some ways those worries are just the tip of the iceberg: there is no reason to expect these dynamics to be simple.

Another key limitation of this analysis is its data. While there are large numbers of listings from each metropolitan area, the distribution of those listings across time is not the same in all of the areas, due to variation in posting, but also due to data collection: we started collecting data in Seattle in 2017 but in 2019 for the other metropolitan areas. While results are similar when I restrict the data to only a 2019 subset, a topic model fit on only that data is substantively different (though the triangulation checks for the topics analyzed here are robust). More information about that can be found in Appendix V.

Future Work

Together, these implications and limitations point to rich possibilities for future work. First of all, data access means that I can now use a very robust set of data with a shared time frame from nearly 100 metropolitan areas. Such a dataset would probably be able to make better estimates of the distribution of random intercepts and slopes. Secondly, this analysis has focused on the qualities of the neighborhood, operationalized as a tract. However, the theories of how landlords think about sorting probably relate not just to potential tenants who live in the same tract as the unit, but also how landlords think about potential tenants from nearby tracts. Ongoing research from Stewart et. al. (2022) suggests that the features of adjacent tracts also influence the way landlords write advertisements. Including such analysis would strengthen the connection to theory here as well. Finally, this paper only attends to five metro-level variables: region, walkability, SOI policy, rent control policy, and segregation. However, there are many other policies and aspects of municipalities that could influence how landlords write texts differently. Understanding more about the metropolitan context would require investigating more of those possible covariates.

We know that housing processes, and the markets in which those processes unfold, are complex and shaped by both hyper-local and metropolitan context. Analyses such as this one can add to our understanding of that context and its impact, but they are also relatively gross instruments for measuring subtle interactions. Still, imperfect measures and understandings are better than none, as long as we understand how to limit our interpretations and to use these experiences to build better models in the future.

Conclusion

Contemporary rental housing markets in the United States are based increasingly on digital platforms, like Craigslist. In the 16 metros in my study area, advertisements on Craigslist consistently included racialized discourse both in their discussions of neighborhood quality, and in their use of exclusionary language. While some aspects of that discourse was racialized in similar ways across the 16 areas, other aspects were not, showing significant variation in different municipalities. Using metro-level variables I expose some of the factors that might explain that variation, but as with any process as complex as multi-city housing markets, there is ample space for further investigation into the impact of metropolitan context on both rental advertisements and landlord behavior generally.

These findings have significant implications for how we think about racial ideology and its relationship to racial structure, for understanding how ethnoraciality is co-constructed with socio-economic status, and for developing strong housing policies that protect renters while understanding landlords' motivations. They also have implications for platforms. Craigslist's hands-off approach currently allows landlords to flaunt laws that protect HCV holders, and that should stop. Finally, the roots of this problem are far deeper than in the strategies that landlords and property managers use to shape their advertisements, resting instead in the legacy of how property ownership and racialized space are intertwined in the United States, connecting to redlining, to the Chinese Exclusion Act, and to global phenomena like Settler Colonialism. Dismantling this kind of racist ideology also means rethinking what it means to own land, rethinking who has the right to safe housing, and what the values of our society truly are.

References

Althusser, Louis. "Ideology and ideological state apparatuses (notes towards an investigation)." The anthropology of the state: A reader 9, no. 1 (2006): 86-98.

Anderson, Carol. 2016. *White rage: The unspoken truth of our racial divide*. Bloomsbury Publishing USA.

Bates, Thomas R. 1975 "Gramsci and the Theory of Hegemony." Journal of the History of Ideas 36, no. 2: 351-366.

Bell, Wendell. 1954. "A Probability Model for the Measurement of Ecological Segregation." Social Forces 32:357-64.

Besbris, Max. "Upsold." In Upsold. University of Chicago Press, 2020.

Besbris, Max, John Kuk, Ann Owens, and Ariela Schachter. 2022. "Predatory Inclusion in the Market for Rental Housing: A Multicity Empirical Test." Socius 8: 23780231221079001.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent dirichlet allocation." The Journal of machine Learning research 3: 993-1022.

Bobo, Lawrence D., and Cybelle Fox. 2003. "Race, racism, and discrimination: Bridging problems, methods, and theory in social psychological research." Social psychology quarterly 66, no. 4: 319-332.

Bonilla-Silva, Eduardo. 2003. Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in the United States. Rowman & Littlefield Publishers.

Bonilla-Silva, Eduardo. 1997. "Rethinking racism: Toward a structural interpretation." American sociological review: 465-480.

Bonilla-Silva, Eduardo. 2019. "Feeling race: Theorizing the racial economy of emotions." American Sociological Review 84, no. 1: 1-25.

Bourdieu, Pierre. 1979. "Symbolic power." Critique of anthropology 4, no. 13-14: 77-85.

Brown, Tyson, and Patricia Homan. 2022. "Structural racism and health stratification in the US: connecting theory to measurement." Socarxiv. 10.31235/osf.io/3eacp.

Collins, Patricia Hill. 2002. Black feminist thought: Knowledge, consciousness, and the politics of empowerment. Routledge.

Crowder, Kyle, Jeremy Pais, and Scott J. South. 2012. "Neighborhood diversity, metropolitan constraints, and household migration." American sociological review 77.3: 325-353.

Egami, Naoki, Christian J. Fong, Justin Grimmer, Margaret E. Roberts, and Brandon M. Stewart. 2018. "How to make causal inferences using texts." arXiv preprint arXiv:1802.02163.

Fanon, Franz. 1952. Black Skin, White Masks. New York: Grove Press.

Gilliard, Chris. and Culik, Hugh. 2016. “Digital redlining, access, and privacy: filtering content is often done with good intent, but filtering can also create equity and privacy issues.” Available at: https://www.commonsense.org/education/articles/digital-redlining-access-and-privacy.

Golash-Boza, Tanya. 2016. "A critical and comprehensive sociological theory of race and racism." Sociology of race and ethnicity 2, no. 2: 129-141.

Graetz, Nick, and Michael Esposito. 2022. "Historical redlining and contemporary racial disparities in neighborhood life expectancy." Socarxiv. 10.31235/osf.io/q9gbx.

Greif, Meredith. 2018. "Regulating landlords: Unintended consequences for poor tenants." City & Community 17, no. 3: 658-674.

Hardeman, Rachel R., Patricia A. Homan, Tongtan Chantarat, Brigette A. Davis, and Tyson H. Brown. 2022. "Improving The Measurement Of Structural Racism To Achieve Antiracist Health Policy: Study examines measurement of structural racism to achieve antiracist health policy." Health Affairs 41, no. 2: 179-186.

Hwang, Jackelyn. 2016. "The social construction of a gentrifying neighborhood: Reifying and redefining identity and boundaries in inequality." Urban Affairs Review 52, no. 1: 98-128.

Kennedy, Ian, Chris Hess, Amandalynne Paullada, and Sarah Chasins,. 2021. “Racialized Discourse in Seattle Rental Ad Texts.” Social Forces 99(4) doi: 10.1093/sf/soaa075.

Korver-Glenn, Elizabeth. 2018. “Compounding Inequalities: How Racial Stereotypes and Discrimination Accumulate across the Stages of Housing Exchange.” American Sociological Review 83(4):627–56. doi: 10.1177/0003122418781774.

Krysan, Maria, and Kyle Crowder. 2017. Cycle of Segregation: Social Processes and Residential Stratification. New York: Russell Sage Foundation.

Liberson, Stanley. 1981. 'An Asymmetrical Approach to Segregation." Pp. 61-82 in Ethnic Segregation in Cities, edited by Ceri Peach, Vaughn Robinson, and Susan Smith. Croom Helm.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2021. IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V16.0

McElreath, Richard. 2020. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC.

Monnahan, Cole C., James T. Thorson, and Trevor A. Branch. 2017. "Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo." Methods in Ecology and Evolution 8, no. 3: 339-348.

Massey, Douglas S., and Nancy A. Denton. 1994. American apartheid: Segregation and the making of the underclass. Cambridge, MA: Harvard University Press.

McKittrick, Katherine. 2006. Demonic grounds: Black women and the cartographies of struggle. U of Minnesota Press.

National Multifamily Housing Council. 2022a. Rent Control By State Law. National Multifamily Housing Council: https://www.nmhc.org/globalassets/advocacy/rent-control/rent-control-by-state-law-chart_2022.pdf

—-. 2022b. Source of Income Laws by State, County and City. National Multifamily Housing Council: https://www.nmhc.org/research-insight/analysis-and-guidance/source-of-income-laws-by-state-county-and-city/

Omi, Michael, and Howard Winant. 1986. Racial formation in the United States. Abingdon, Oxfordshire: Routledge.

Ray, Victor. 2019. "A theory of racialized organizations." American Sociological Review 84, no. 1: 26-53.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. "Stm: An R package for structural topic models." Journal of Statistical Software 91: 1-40.

Rosen, Eva. 2014. "Rigging the Rules of the Game: How Landlords Geographically Sort Low–Income Renters." City & Community 13, no. 4: 310-340.

Rosen, Eva, Philip ME Garboden, and Jennifer E. Cossyleon. 2021. "Racial discrimination in housing: how landlords use algorithms and home visits to screen tenants." American Sociological Review 86, no. 5: 787-822.

Rothstein, Richard. 2017. The Color of Law: A Forgotten History of How Our Government Segregated America. New York: Liveright Publishing.

Rucks-Ahidiana, Zawadi. 2021. "Theorizing Gentrification as a Process of Racial Capitalism." City & Community: 15356841211054790.

Sampson, Robert J., and Stephen W. Raudenbush. 2004. "Seeing disorder: Neighborhood stigma and the social construction of “broken windows”." Social psychology quarterly 67, no. 4: 319-342.

South, Scott J., Kyle Crowder, and Jeremy Pais. 2011. "Metropolitan structure and neighborhood attainment: Exploring intermetropolitan variation in racial residential segregation." Demography 48.4: 1263-1292.

Spillers, Hortense J. 1987. "Mama's baby, papa's maybe: An American grammar book." diacritics 17, no. 2: 65-81.

Stan Development Team. 2022. Stan Modeling Language Users Guide and Reference Manual, 2.29. https://mc-stan.org

Stewart, Remy, Christian Hess, Ian Kennedy, Kyle Crowder. 2022. “The Use of Security Deposits as a Residential Sorting Mechanism: A computational Text Analysis of Craigslist Rental Listings.” Poster Presentation. PAA 2022 Annual Meeting. Atlanta.

Turner, Margery A., Stephen L. Ross, George C. Galster, John Yinger, Erin B. Godfrey, Beata A. Bednarz, Carla Herbig, Seon Joo Lee, AKM. Rezaul Hossain, Bo Zhao. 2002. Discrimination in Metropolitan Housing Markets: National Results from Phase I HDS 2000. The Urban Institute.

Turner, Margery A., Diane K. Levy, Dough Wissoker, Claudia L. Aranda, Rob Pitingolo, Rob Santos. 2013. Housing Discrimination Against Racial And Ethnic Minorities 2012. The Urban Institute.

Wynter, Sylvia. 2001. "Towards the Sociogenic Principle: Fanon, Identity, the Puzzle of Conscious Experience, and What It Is Like to Be ‘Black.’." National identities and sociopolitical changes in Latin America: 30-66.

Conclusion

I sought out a way to look at the overlapping intersection of two threads in the 'fabric' of hierarchical oppression: one structural and one ideological. I found evidence of how racialized housing structures aligned with and were possibly replicated in part by racist ideology embedded in rental ad discourse. Based on that investigation, this dissertation makes three major claims. First, that the language of Craigslist rental ad texts is indeed racialized, and that racialization aligns with what we know about anti-Black ideology in the United States. Second, that exclusionary language is common and diverse in the advertisements. That exclusionary language is also racialized and seems to align with the processes which perpetuate racialized housing patterns, like gentrification. Finally, examining variation in language across metropolitan areas, I find that anti-Black neighborhood discourse is consistent across metro areas, while language excluding negatively credentialed prospective tenants is more common in less segregated cities and in cities with rent control laws.

Racialized Discourse

When advertisements include positive neighborhood descriptions, like "great neighborhood" or "location, location, location," the analysis in Chapter 1 shows that such language is unevenly distributed across space. It is significantly more common in White and Asian neighborhoods, and significantly less common in Black neighborhoods. This difference persists even in models accounting for neighborhood socioeconomic status: rich Black neighborhoods with median incomes of $150,000 have the same expected prevalence of this language as poor White and Asian neighborhoods with a small fraction of that median income.

That unequal distribution of positive language aligns with the history of racial ideology and practice in the United States that has consistently devalued Black spaces. It also connects to arguments that racial ideology in the US is dominated by anti-Blackness. However, while that language is hegemonic, it does not stand without challenge. Instead, there are many Black neighborhoods that are described positively. Some of them are close to cultural resources, like sports stadiums, but others gain their positive descriptions from the concerted efforts of property management companies owned by local people. That's the case for Mahari Yared in West Philadelphia, owned and run by Mahari Bailey who grew up there. Mahari Yared's website is explicit that part of its mission is to lift up its community, and its advertisements are positive about the community resources in West Philadelphia.

Discursive Exclusion

While the most obvious purpose of a rental advertisement is to market a housing unit (and therefore also its surrounding neighborhood), another purpose is for landlords to select the prospective tenant whom they feel is the best fit for the unit, and the neighborhood. In Chapter 2 I find that landlords use a variety of exclusionary discourse: explicitly excluding housing choice voucher (HCV) holders, writing about credit and background checks, setting high move-in and application costs, and describing tenant responsibilities. I theorize that sometimes this language can be used to sort tenants within a neighborhood, and sometimes to signal homeseekers about what kind of a neighborhood the unit is in.

I find that exclusionary language fits with both racialized and classed stereotypes about what kinds of tenants are likely to be negatively credentialed. For instance, language about background and credit checks is more common in non-White neighborhoods and, within those

neighborhoods, in non-gentrifying neighborhoods. Language about tenant responsibilities, on the other hand, is most common in richer, Whiter neighborhoods, including gentrifying places. This could reflect landlord and text writer expectations about who needs to have their background and credit checked, compared to who can live up to detailed expectations.

Metro-Level Variation

Finally, I examine how the racialized and classed language from the first two chapters varies (or not) across metropolitan areas. In Chapter 3 I show that the anti-Black ideology that structures language about neighborhood quality is depressingly consistent across the 16 core-based statistical areas (CBSAs) in my study area. In the fully specified model, every CBSA had a significant difference in the expected proportion of positive neighborhood discourse between White and Black neighborhoods, and the variation in that difference was small. This aligns with the theory that anti-Blackness is a widespread phenomenon in US ideology. It is not limited to rural areas or to cities in the south or cities with high levels of segregation.

Exclusionary language, especially language around background and credit checks, on the other hand, does vary by CBSA. Metro areas with lower levels of segregation saw more of that topic on average, as did metro areas with rent control laws. That finding could align with landlords and property managers trying to maximize their discretion in places where they may have a diverse applicant pool or where it may be difficult to raise the rent on their tenants.

What Can We Do?

The easiest point of intervention in how housing advertisements are written is at the platform level. Platforms like Craigslist have a responsibility to uphold the public interest in free and fair access to housing. It is within their capabilities to ban explicitly exclusionary language against HCV holders and even to prompt text writers to include positive attributes about

less-White neighborhoods. However, while I believe that rental advertisement discourse can contribute to the perpetuation of racialized housing outcomes, changing that discourse will not be enough to dismantle those outcomes.

We also need to support local ownership and management of rental properties, as with Mahari Yared in West Philadelphia. People who already have strong connections to their neighborhoods are going to do the best job writing about those neighborhoods, and by keeping ownership local, they will also be investing in those communities.

That is an adjustment to our current market-based rental process, but we also need to build alternatives to that market. This paper is one of many identifying the problems that HCV holders have in facing housing in those markets, and traditional public housing models are also fraught. A possible alternative is social housing, whereby municipalities build mixed-income properties that are owned and managed by a public social-housing organization. Rents in these properties are generally capped at 30% of people's incomes, so no residents are rent burdened. Market-rate tenets help support subsidized tenants, and excess rents go towards the construction of further housing. It is possible that such a system could sidestep some of the ideological issues raised above, disturbing the weave of structure and ideology, at least in the domain of rental housing.

Future Directions

I am excited about continuing my work on this topic. I have already planned both ways to use survey experiment methods, and spatial analyses of segregation using the same data source I apply in this dissertation. The survey methods would allow me to supplement the descriptive analysis I offer here with some limited causal claims. I am particularly interested in whether the racialized discourse I uncover changes the way home seekers view neighborhoods. I will also be

working on an audit study examining landlord reactions to negatively credentialed tenants. That could uncover more about landlords' intentions during the process of filling their units. I also intend to use this data source to investigate more about how online rental ads might align with racialized housing dynamics through their role in defining named neighborhoods. I plan to use the named neighborhoods defined by the location of Craigslist ads to estimate segregation metrics, and then compare those metrics to those calculated using census tract boundaries.

Finally, I imagine that this is just a particular case study for the study of how structural and ideological threads meet in the fabric that supports systems of oppression. I plan to seek out other such intersections to see if my observations from Craigslist might apply there also.

Appendix

Appendix I: Deduplication

As I worked with this data, it became very clear that duplicate or near-duplicate listings were very common in our scraped Craigslist data. Sometimes the listings were for the same unit posted at various points in time, often with the same rent. Just as often, though, the duplicates were postings for different units in the same building, or in different buildings but managed by the same property manager. In some cases the duplicates reflect spam or fraud, where the poster copies the text and pictures from a legitimate advertisement. For many use cases, like for estimating rents for small areal units, these kinds of duplicate listings are not a large problem. Even for my analysis using STM, the model recovers very similar topics when including or not including these duplicates. However, I believe that including the duplicate listings raises two issues for my analysis as a whole, one quantitative and one qualitative. First, quantitatively I believe it is likely that high levels of duplicates are likely to be unequally spread through space: they could be more common in areas with higher rents, since landlords there may repost more often, and in areas with more large multi-family properties, since those areas likely have more nearly-identical postings. Since those features are likely associated with neighborhood socioeconomic status and racialization including the duplicates could bias my estimates.

Qualitatively, I use high-matching texts to interpret the STM output. However, when there are many duplicate or near-duplicate texts, the high-matching texts for each topic tend to be composed of those duplicates. That makes interpreting the topics difficult: I want to identify topic titles and understandings of the topics that can reasonably be generalized to other texts with a high proportion of that topic, but if there are only duplicate texts, I cannot be sure if what I’m

reading is about the topic, or just about this particular multi-family building, or a particular landlord or property manager.

For those two reasons, I use a very-greedy deduplication algorithm that excludes almost four fifths of my original 2.5million text sample. Directly comparing all 2.5 million texts to each other is not computationally feasible. Instead, I use a two-step approach which takes advantage of the data structure. First, I deduplicate within addresses: my experience with the data shows that most of the duplicates come from repeated postings of the same unit or similar posts about units in the same building. Comparing within addresses takes care of those kinds of duplicates. I always keep the earliest observed instance of a text. After address deduplication, I use STM itself to ferret out duplicates. When there are duplicates they tend to cluster as the highest-matching texts in a particular topic, I iteratively fit STM models and deduplicate within the top texts for each topic.

The deduplication itself uses the R stringdistmatrix package's Jaccard distance. I set a threshold of .7 which I arrived at after comparing my own qualitative coding of duplicates to deduplication using various measures of distance at various thresholds. For this use case I found a Jaccard distance of .7 often removed all of the texts I considered to be practically duplicates, while leaving almost all of the texts I considered distinct. I expect that with other corpora, even other corpora of Craigslist listings, the right distance measure and threshold could be quite different.

Appendix II: STM Robust

Topic models can be sensitive to the analyst's choice of the number of topics. Wilkerson and Casas (2017) outline a method to check whether similar topics are recovered when the

number of topics is adjusted. They develop that method for LDA topic models, and I have modified that method to use with STM topic models. For the analysis I use a topic model with 50 topics, for the robustness check I estimate new models with 47-53 topics. I then compare the topic-vocabulary vectors in my 50 topic STM to the topic-word vectors of the six alternative models to see if similar topics were recovered with those choices of K. I show the results below for the five topics I focus on in this paper:

Appendix Table 1:

Topic Name	47	48	49	**50**	51	52	53
Fees and Deposits (18)	0.99	0.99	1.00	-	1.00	0.97	0.99
Historic Charm (36)	0.99	0.99	0.98	-	0.99	0.99	0.99
Restrictions and Stipulations (38)	0.99	0.99	0.99	-	0.99	0.97	0.96
Background and Credit Checks (46)	0.99	0.99	1.00	-	1.00	0.99	0.99
Positive Adjectives: Location (50)	0.98	0.98	0.98	-	0.98	0.98	0.98

Appendix Table 1 Caption: This table shows how closely the topics I focus on match topics in different STM models fit with different numbers of topics, ranging from 47 to 53 topics. Green cells are those which have at least one topic with a cosine similarity of >.95 with a topic in the other model. The number is the cosine similarity of the closest match. Some cells show 1.00, but they're actually just very close. In other words, this analysis suggests that these topics are very stable by this measure across models with other choices of K.

Appendix III: Triangulation

To further validate the results presented in the main text, I compare the results using the topic model to results based on a set of keywords or search strings that reflect my understanding of each topic presented in chapters 1 and 2. To be clear, the expectation here is not that results be identical: the topic model uses every word in the vocabulary to assign its proportions, and my search strings will only use at most 12 words. What I am looking for is that the pattern of the keywords generated by my understanding of the topic roughly align with the results from the topic models. For each topic below I first reproduce the plot with covariates from either Chapter 1 or 2. Then I present 1 or more plots that use search strings or keywords I generated. I offer brief comments about the similarities and differences.

Positive Adjectives: Location (50):

Figure A.1: Topic Based Plot

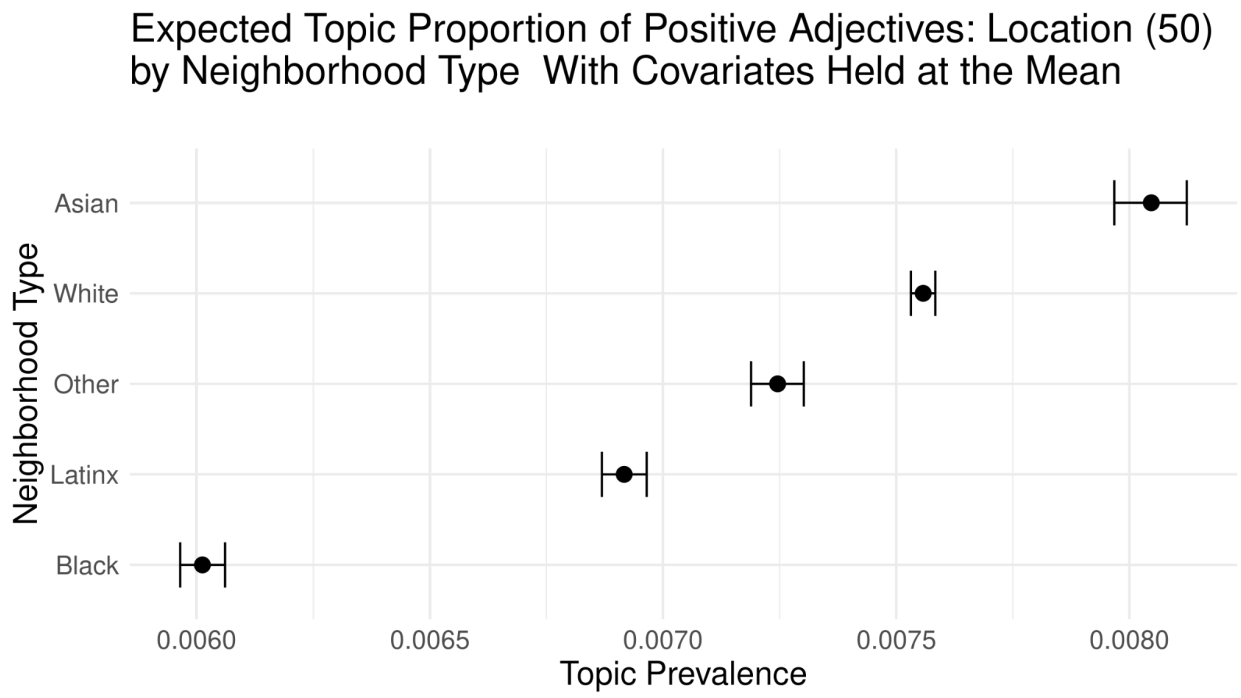
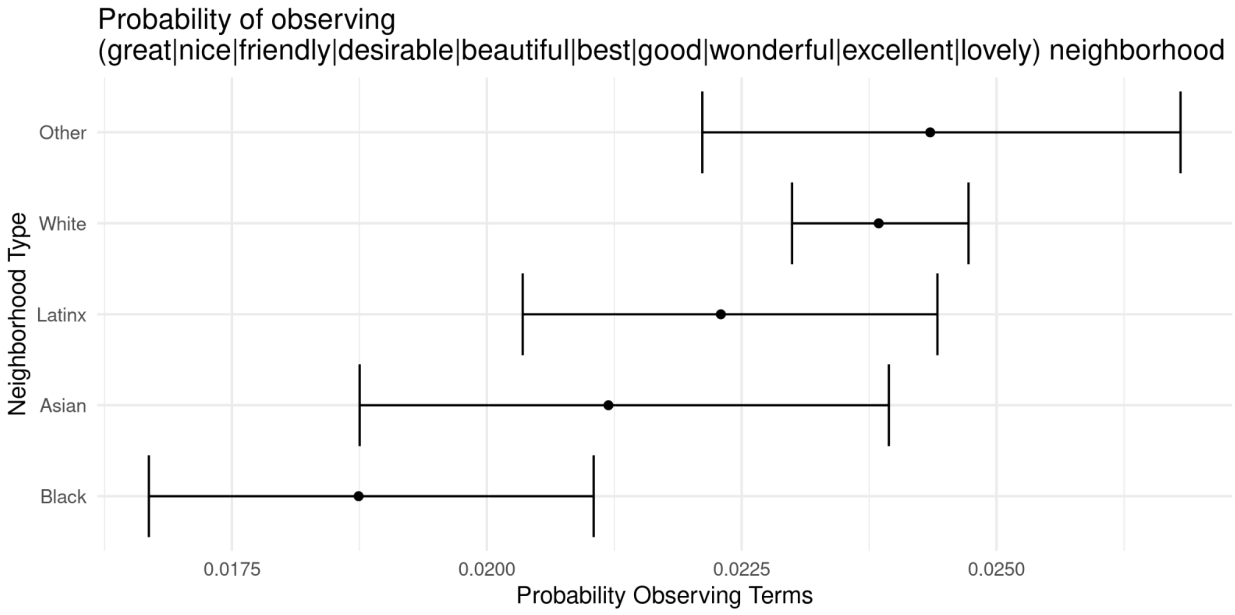
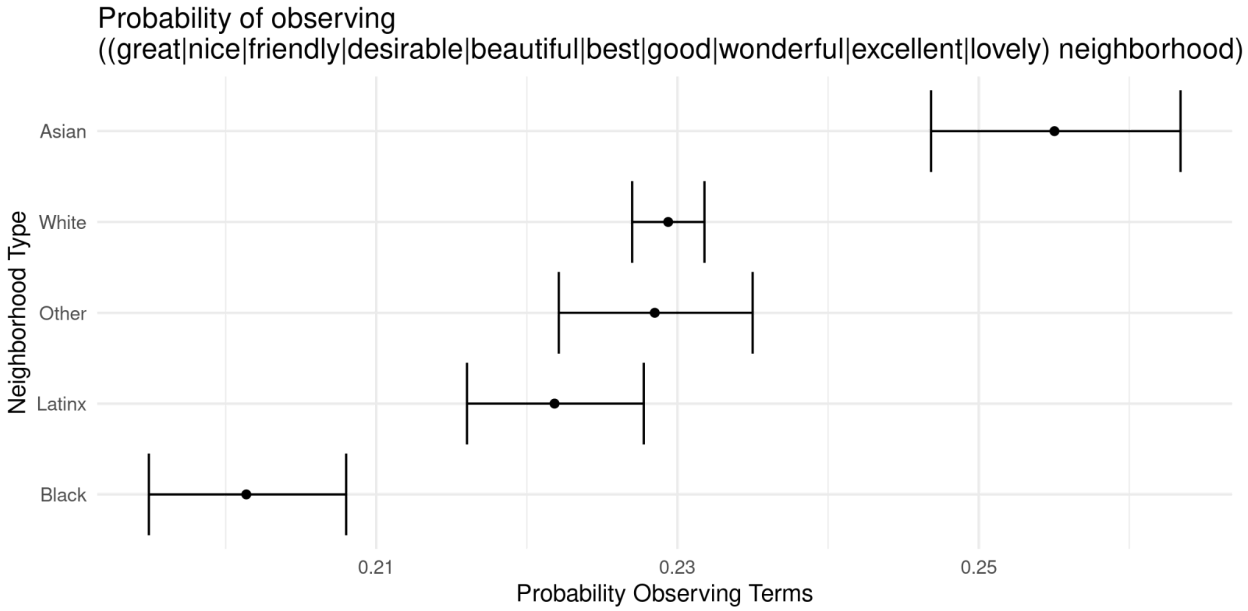


Figure A.1.2: Positive Neighborhood Descriptors



The first search string I used for Topic50 looks for a combination of any word with the word ‘neighborhood’. This has similar results to the main finding in that Black neighborhoods have significantly the lowest expected probability of observing that search string. However, Asian neighborhoods, which have a large portion of Topic50, seem to have this search string much less often. I realized that this search string only uses the term ‘neighborhood’ to operationalize the location portion of the topic. So I did an additional test adding the term ‘location.’

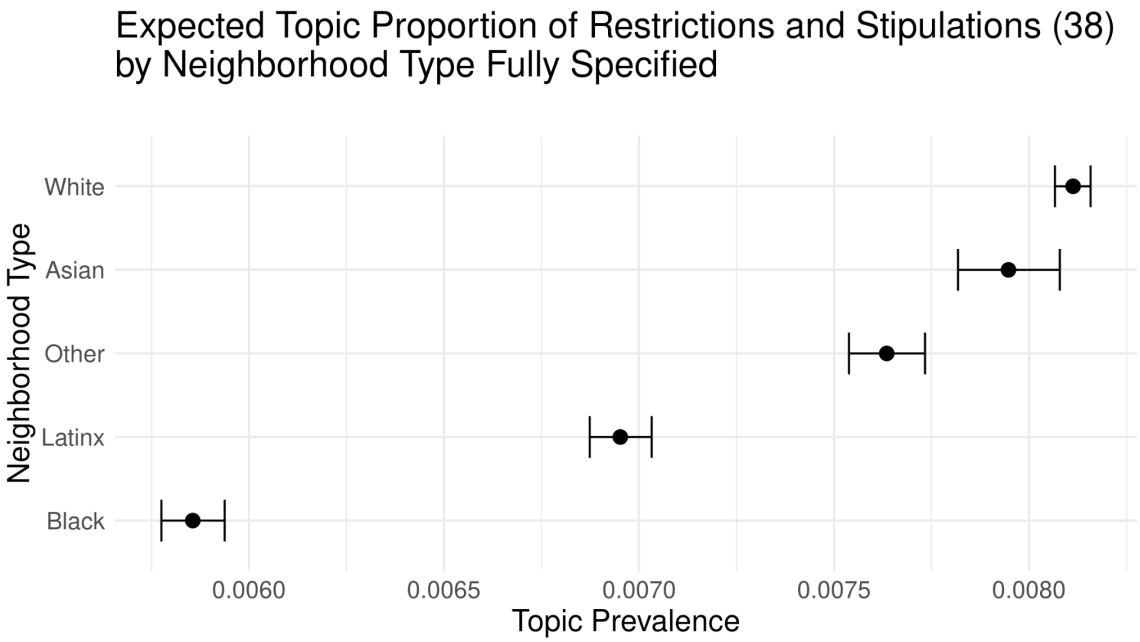
Figure A.2: Positive Neighborhood Descriptors + ‘Location’



Adding the term ‘location’ changes the plot so that it looks very much like the topic-generated plot. Interestingly it seems like the high prevalence of the term ‘location’ is what might partially account for the results for Asain neighborhoods.

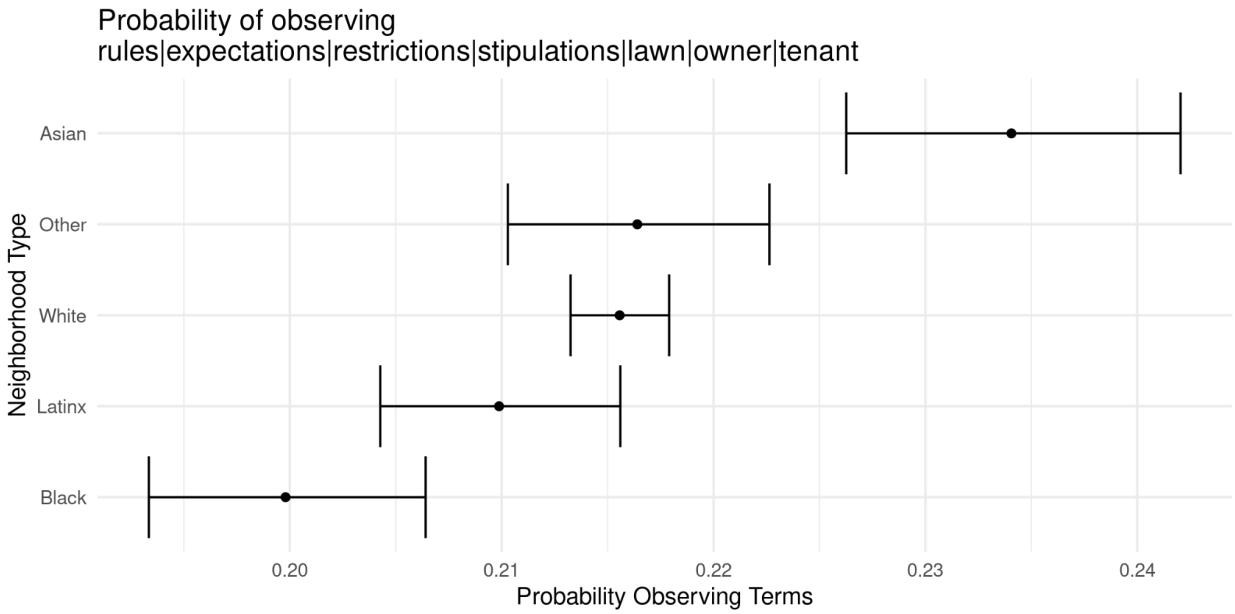
Restrictions and Stipulations (38):

Figure A.3: Topic Based Plot



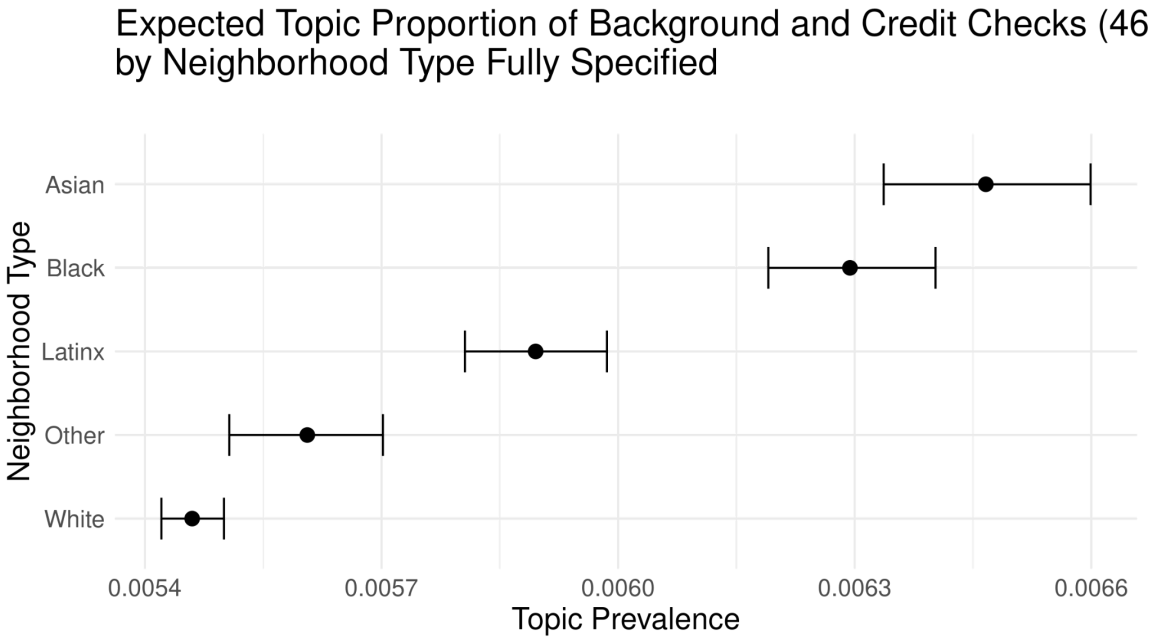
I used a set of keywords here that align with my interpretation of Topic38 as describing tenant responsibilities and expectations. It aligns reasonably well with the results from the topic model.

Figure A.4: Keyword Plot



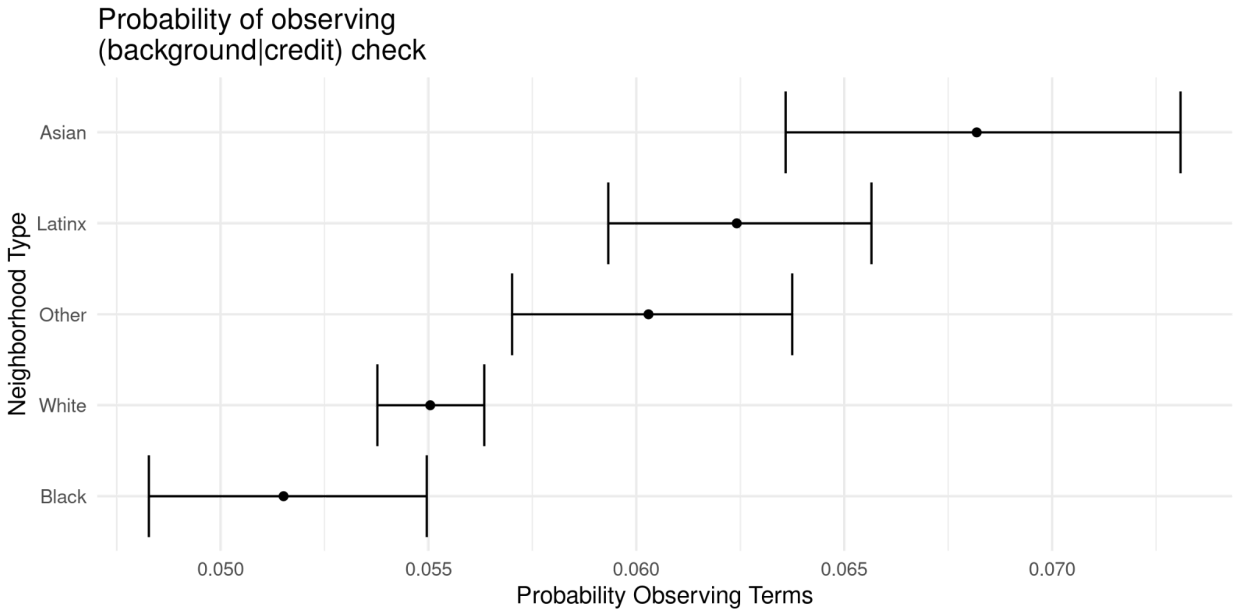
Background and Credit Checks 46

Figure A.1.6: Topic Based Plot



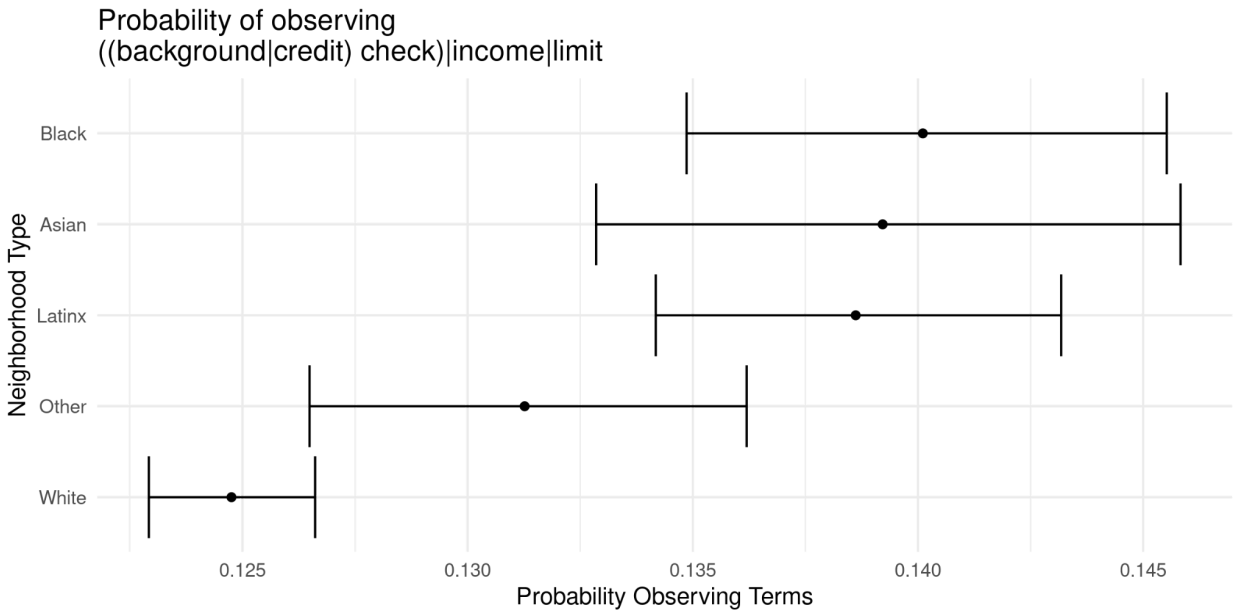
Here I started with a very simple search string that matches either ‘background check’ or ‘credit check.’ These results seem somewhat different from the fully specified model using the topic: the location of Black neighborhoods moves from second highest to lowest.

Figure A.5: Background and Credit Checks



I hypothesized that I may not have included all of the aspects of the topic. Going back to the example texts, I realized that many texts included details about income, or used the term limit. Adding those terms resulted in a plot much more similar to the topic-based analysis.

Figure A.1.8: Background and Credit + income and limit



Fees and Deposits (18)

Figure A.6: Topic Based Plot

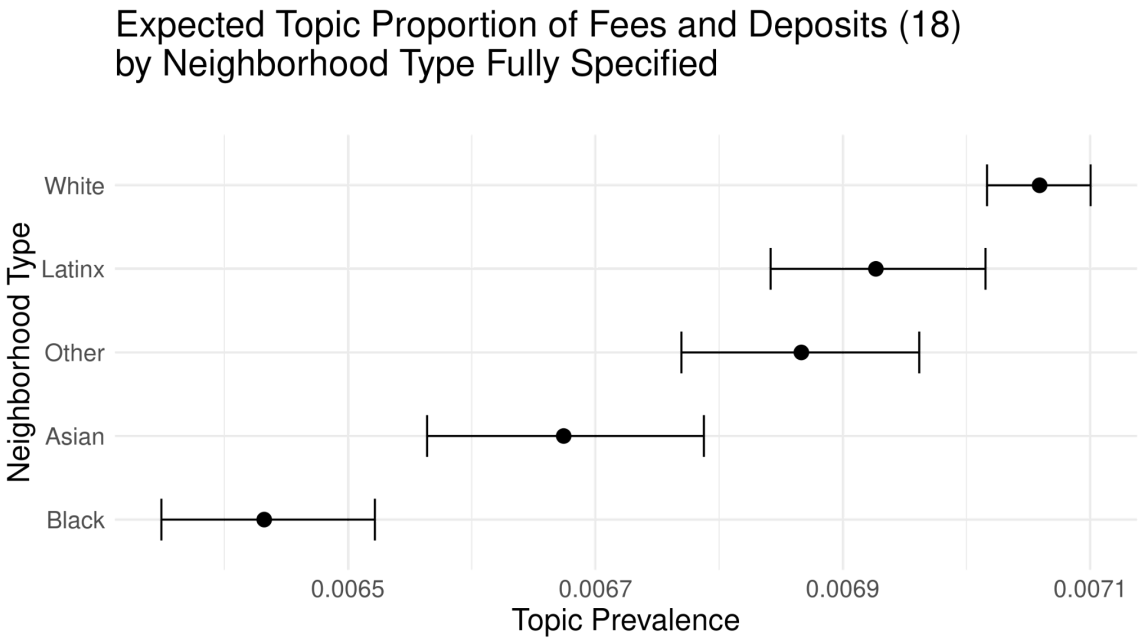
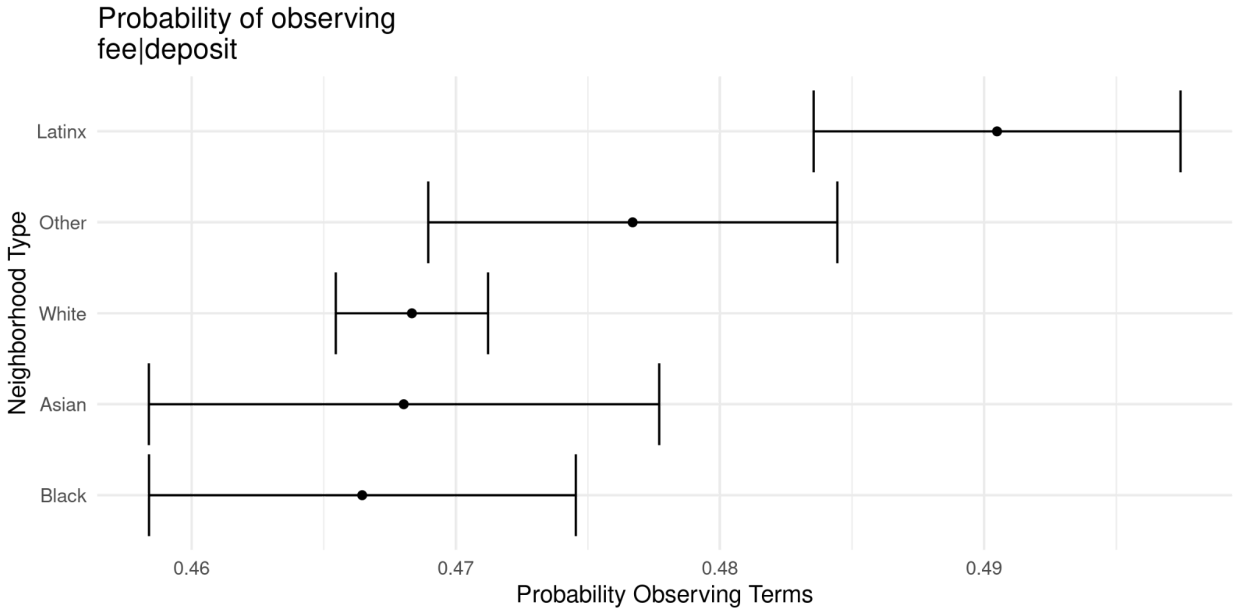


Figure A.1.11: Fee|Deposit



Looking just at the terms ‘fee’ and ‘deposit’ had somewhat different results than the topic-based analysis, though also strong similarities: The bottom two neighborhood types and the top three are the same. This suggests that going forward I may want to look more closely into the best way to operationalize this idea.

Appendix IV: Topic Descriptions

Appendix Table 2

Topic	topic title	Top Words
Topic1	Property Amenities	concierge, conference, valet, lounge, simulator, clubroom, cabanas, grooming, yoga, amenity, billiards, technogym, leed, cardio, trx
Topic2	Appointments and Hours	monday, tuesday, thursday, wednesday, friday, emergency, select, hours, townhomes, patios, saturday, garages, connections, bark, microwaves
Topic3	NYC Area Commute	journal, brunswick, rutherford, jsq, caldwell, fordham, metuchen, seton, njit, palisade, raritan, rahway, cranford, metropark, brkr
Topic4	Walkability	distance, stores, walking, banks, shops, grocery, restaurants, away, farmers, store, joe, markets, within, bakeries, pubs
Topic5	Scenic Views	needle, slu, elliott, elliot, olympics, seward, tableau, melinda, admiral, yesler, pointe, cascades, territorial, century, magnolia
Topic6	Ocean Access	ocean, mahalo, uri, beachfront, kcc, ewa, barracks, ohana, surfboard, hpu, corps, canoe, oceanview, kunia, pupukea
Topic7	Lease Details	signing, month, rent, security, months, due, first, last, year, upfront, total, requires, deposit, upon, lease
Topic8	Parking and Unit Location	condo, unit, spot, secured, condominium, assigned, secure, balcony, complex, reserved, parking, designated, underground, spots, locker
Topic9	On-Site Sports	tennis, courts, court, swimming, pool, basketball, pools, racquetball, playground, volleyball, sauna, spa, clubhouse, spas, resort
Topic10	Section 8 Availability	cargarage, appt, bdrm, duplex, bdr, section, ere, appointment, sec, call, gary, bth, showing, loretta, text
Topic11	Utilities	alll, plowing, snow, removal, numc, floria, eatin, coccia, annadale, bentahar, oil, tottenville, rwu, verdeschi, livingroom
Topic12	Positive Adjectives: Unit	won, see, right, summer, coming, perfect, fall, come, soon, awesome, make, gem, wow, miss, beat
Topic13	Free Standing Homes	house, fenced, yard, shed, single, shotgun, ups, driveway, family, bungalow, hook, back, backyard, unfinished, fence
Topic14	Pet Details	purrr, wooof, turbo, sja, macpherson, redside, sandco, martisa, shinsato, wpm, verraterra, description, terms, eberly, duration
Topic15	Appliances	fans, conditioning, wall, air, ceiling, blinds, refrigerator, fan, microwave, molding, stove, conditioner, ceramic, range, heating
Topic16	Furniture	sofa, linens, mattress, towels, stocked, pots, toaster, utensils, pans, channels, bedding, cookware, sheets, plates, sleeper
Topic17	Least Timing	sublease, takeover, july, sublet, march, min, august, april, dec, transfer, october, earlier, starting, november, february
Topic18	Fees and Deposits	refundable, nonrefundable, adult, insurance, non, damage, pdt, case, application, lbs, holding, processing, brink, charged, liability
Topic19	Lofts and Studios	unhide, studio, urth, posting, studios, deco, loft, lofts, bachelor, vermont, korea, melrose, factory, alcove, ktown

Topic20	Renovations	new, paint, newly, brand, fresh, completely, recently, remodeled, redone, remodel, totally, painted, renovated, freshly, plumbing
Topic21	Architectural Details	doors, sliding, vanity, faucet, enclosure, glass, sink, double, led, shower, french, shelving, porcelain, dispenser, panel
Topic22	Boston Commute	bus, minute, rail, lowell, commuter, stops, miles, merrimack, kenmore, artery, lines, andover, reservoir, hingham, massart
Topic23	Universities and Hospitals	drexel, delaware, ardmore, plymouth, wales, goshen, malvern, forge, abington, darby, villanova, brandywine, hup, radnor, hospital
Topic24	Disclaimers	reusable, comprehensive, race, origin, religion, familial, discrimination, rcw, discriminate, pursuant, prohibits, jmw, consumer, reports, sex
Topic25	Non-Rental: Scams, Offers, Requests	deemed, hyperlink, specialize, ads, obligation, purchase, lender, webpage, consultation, browser, homeownership, copyright, consult, scam, financing
Topic26	Utilities and Parking Included	electricity, pays, sewer, trash, electric, water, sewage, included, plex, paid, utilities, garbage, tenant, except, smoking
Topic27	Driving Distance	canyon, calabasas, saddleback, doheny, pedro, glassell, zaya, palmdale, rowland, danville, ortega, winnetka, springs, talega, freeway
Topic28	High-end Room Size	formal, curb, entertaining, bonus, cape, piece, opens, secondary, berdoom, cod, sliders, colonial, ranch, paver, finished
Topic29	Unit Features	deleaded, medford, applianced, waverley, salem, cushing, watermarked, rentsource, Blackstone, jwu, avl, marescia, sotheby, shilalis, lambergs
Topic30	Contact Information	inquiries, respond, reply, messages, thank, inquires, serious, name, please, message, number, responding, calls, questions, mail
Topic31	Room Count and Layout	washer, dryer, two, one, bedroom, private, three, full, dish, stackable, four, size, bathroom, dyer, den
Topic32	Special Deals	special, ready, hurry, immediate, save, reduced, receive, waived, today, savings, fast, ask, drop, priced, upgrades
Topic33	Roomate/home search	housemate, bit, respectful, stuff, really, hello, roommate, tell, trying, folks, lived, anything, looking, tidy, wanted
Topic34	Modern' Decor	tops, granite, stainless, appliances, steel, counter, countertops, floors, counters, quartz, hardwood, modern, gorgeous, throughout, cherry
Topic35	Boston Brokers	broker, cotter, peterborough, upright, hot, hhw, signature, gtn, valencius, comm, beds, mastercard, bradsfield, beacon, alex
Topic36	Historic Charm	lined, original, ocf, tree, magazine, claw, restored, charm, character, victorian, clawfoot, antique, rowhouse, carriage, classic
Topic37	Chicago Commute	bjb, hermitage, kwgc, rehabbed, welles, rehab, schiller, belden, hoyne, illinois, ukranian, mariano, fultongrace, healy, division
Topic38	Restrictions and Stipulations	ago, consider, taken, rules, owners, noise, tenants, repairs, cleaned, install, portion, disturb, increase, lives, rented
Topic39	Mid-Range Room Size	large, room, closet, living, lots, space, plenty, dining, separate, extra, huge, spacious, storage, ample, closets
Topic40	Garage and Bathroom Counts	townhouse, upstairs, attached, car, garage, level, downstairs, story, townhome, master, opener, upper, fireplace, lower, tri

Topic41	On-Site Community	offer, comfort, furry, variety, convenience, deserve, residents, lifestyle, proud, choose, dip, invite, plans, experience, discover
Topic42	Subject to change without notice	change, notes, vary, staffordshire, represent, chows, prices, availability, subject, presa, terriers, pricing, shepherds, actual, hybrids
Topic43	High-End Unit Descriptions	design, idwm, contemporary, elements, timeless, sophisticated, striking, incorporates, culture, chic, sleek, eclectic, incorporate, refined, elegance
Topic44	Public Transportation Access	apartment, transportation, near, close, yasmine, located, keerah, conveniently, carollo, gesler, public, maintained, beautiful, bedroom, kept
Topic45	High-End Apartment Buildings	uws, skips, flex, stabilized, doorman, streeteasy, trains, windowed, prewar, bochen, crosstown, hell, antagoniste, postwar, net
Topic46	Background and Credit Checks	verifiable, statements, felonies, stub, salary, copies, proof, returns, felony, paycheck, judgments, paystubs, requirement, credit, applicants
Topic47	Schools	elementary, clyde, proctor, finn, chinook, cougar, woodridge, tahoma, kingsgate, tyee, elem, rambler, wedgwood, cumming, school
Topic48	Virtual Tours	alamo, waiver, edwardian, void, contrary, biker, walker, webpass, tours, protocol, walkthrough, golden, cole, japantown, guerrero
Topic49	Anuncios en Español	por, cocina, mes, cerca, sala, grande, dos, piso, disponible, dormitorios, mas, bano, recamaras, esta, tiene
Topic50	Positive Adjectives: Location	location, great, excellent, area, easy, terrific, convenient, open, access, value, prime, ideal, including, areas, friendly

Appendix V: Sample

I use Craigslist advertisements from 16 metro areas, spanning from 2017 to 2021. The listings were scraped by the National Rent Project at the University of Washington, and the scraping process changed several times during that period, and was not always the same for all of the metropolitan areas. There are two notable changes over time in the scraping process. First, only listings from Seattle were scraped from late 2017 to early 2019. During that period the project was testing scraping infrastructure. Scraping expanded to 100 metropolitan areas in early 2019, and has continued since then. However, due to a change in the Craigslist platform, we were unable to effectively scrape rents from August 2019 through February 2020. During that period we still successfully scraped listing text and other data, but the rent information is mostly

incomplete. Since this paper uses the rent of the listing as a covariate in the regression analysis, I do not include any advertisements from that period which are missing rent information. Finally, the COVID-19 pandemic started in March 2020 and significantly altered rental markets. However, a large portion of the data included in this analysis comes from that period from March 2020 to April 2021, which was COVID-impacted. However, I run a number of robustness checks to confirm that the results do not depend on these aspects of the data.

The results of the regression analysis are consistent if we use the topics presented in the paper but limit the listings included. This is true if I exclude the Seattle advertisements from the period when only Seattle was scraped (2017-early 2019), and if I exclude the advertisements from the COVID-impacted period from March 2020 on. For instance, consider the interaction plot in the paper.

Figure A.7

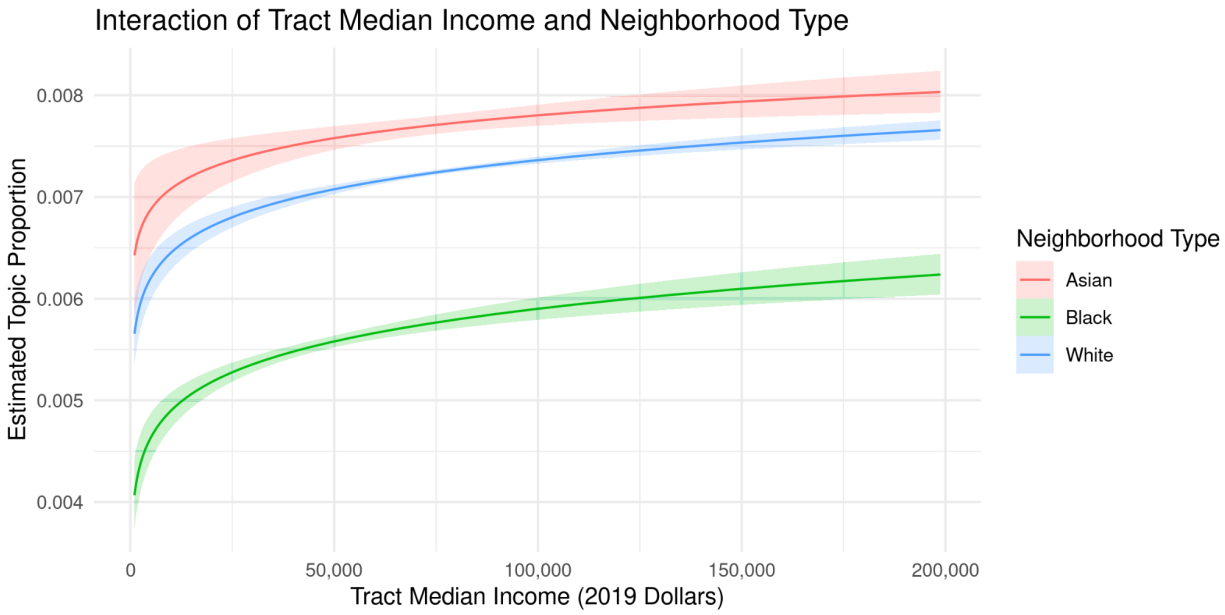
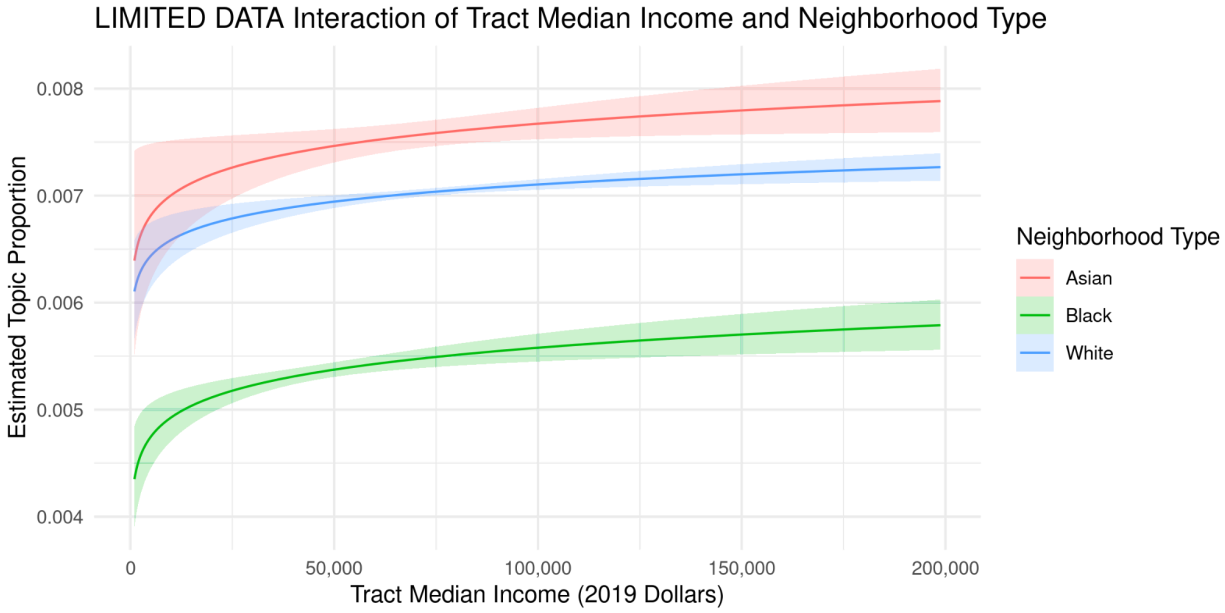


Figure A.8



The two plots are nearly identical, thought Figure A.8, which used the limited data sample has slightly wider errors. I include the whole regression output for these two models

below. Note that there are no changes in sign or significance across the models. The coefficient for the interaction between Black neighborhoods and median income is itself not significant in the limited model, but the interaction as a whole still adds to model fit in that case, and the plot is very similar.

Table 2

	Displayed Model			*Model With 2019 February-August Only*		
Predictors	Estimates	CI	p	Estimates	CI	p
(Intercept)	-3.78	-3.92 – -3.64	<0.001	-3.52	-3.72 – -3.33	<0.001
ntype [asian]	0.23	-0.03 – 0.49	0.081	0	-0.38 – 0.38	1
ntype [black]	-0.49	-0.71 – -0.27	<0.001	-0.48	-0.76 – -0.20	0.001
ntype [latinx]	-1.44	-1.67 – -1.22	<0.001	-1.44	-1.73 – -1.15	<0.001
ntype [other]	-0.61	-0.88 – -0.34	<0.001	-0.5	-0.85 – -0.15	0.005
log income	0.06	0.04 – 0.07	<0.001	0.03	0.02 – 0.05	<0.001
cbsa [Urban Honolulu, HI]	0.01	-0.01 – 0.03	0.277	0.02	-0.00 – 0.04	0.091
cbsa [Worcester, MA-CT]	-0.03	-0.06 – -0.00	0.021	-0.05	-0.09 – -0.02	0.001
pov proportion	0.05	0.00 – 0.10	0.032	-0.02	-0.09 – 0.04	0.461
share oo	-0.12	-0.14 – -0.09	<0.001	-0.22	-0.25 – -0.19	<0.001
share sf detached	-0.27	-0.28 – -0.26	<0.001	-0.16	-0.17 – -0.14	<0.001
share car commuters	-0.36	-0.39 – -0.33	<0.001	-0.35	-0.39 – -0.31	<0.001
share commute over 20	-0.31	-0.34 – -0.28	<0.001	-0.28	-0.32 – -0.23	<0.001
share rental	-0.31	-0.33 – -0.29	<0.001	-0.27	-0.30 – -0.25	<0.001

over 20						
share built before 40	0.08	0.06 – 0.09	<0.001	0.11	0.09 – 0.13	<0.001
share built after 10	-0.6	-0.65 – -0.56	<0.001	-0.61	-0.67 – -0.55	<0.001
share college	1.43	1.38 – 1.48	<0.001	1.4	1.34 – 1.46	<0.001
log price	-0.22	-0.22 – -0.21	<0.001	-0.22	-0.23 – -0.21	<0.001
beds	0.05	0.05 – 0.05	<0.001	0.05	0.05 – 0.06	<0.001
ntype [asian] * log income	-0.02	-0.04 – 0.01	0.2	0.01	-0.03 – 0.04	0.698
ntype [black] * log income	0.02	0.00 – 0.04	0.025	0.02	-0.00 – 0.05	0.111
ntype [latinx] * log income	0.12	0.10 – 0.14	<0.001	0.12	0.10 – 0.15	<0.001
ntype [other] * log income	0.05	0.03 – 0.08	<0.001	0.04	0.01 – 0.07	0.011
Observations	253744			141316		
R2 / R2 adjusted	0.135 / 0.135			0.129 / 0.128		

Appendix VI: Gentrification

I base my measure of gentrification on Hwang's (2020) operationalization where she compares Seattle neighborhoods from 1990 until 2013. I adapt that measure to account for the gentrification history of the neighborhoods in my sample leading up to the observed advertisements in the 2017-2021 period I analyze. Hwang's method requires two points in time, the first point acts as a baseline and identifies tracts as gentrified or not, the second time sorts the gentrified tracts into those which did gentrify and those that did not. Both time points create measures that are relational to a spatial context. In the first time point, tracts are gentrifiable if their median income is lower than the contextual median income. In the second time point, previously gentrifiable tracts are labeled as gentrifying (or gentrified) if indicators of both housing costs and socioeconomic status increased faster than the contextual measures. Housing cost measures include home value and median rent, and socioeconomic measures are median income and share college educated. If at least one measure from each category increased faster than the same measure for the contextual area, than the tract is marked as gentrifying. This process results in a three-level typology: gentrifying, gentrifiable but not gentrifying, and not gentrifiable.

This method requires the analyst to make two key decisions: what time points to use and what contextual areal unit to use. I corresponded with Hwang about the right choices for this particular case and followed her suggestion to use the 2000 decennial census as the starting time point and the 2019 5-year ACS as the second time point. Hwang recommended using city municipal boundaries as the areal context. However, since my data crosses CBSAs much of it is within very small cities or in unincorporated areas. Therefore, I used the city or municipal context, operationalized as a Census place, only if that place had a population greater than 100K.

For unincorporated tracts or tracts in smaller municipalities, I used all of the tracts that were similarly unincorporated or in smaller municipalities but within the same CBSA as the spatial context. However, results are substantially similar if I use the whole CBSA as the spatial context.

The choice of time periods may be more influential. When I repeated the analysis with a smaller time scale, the associations for gentrification were generally smaller. In my correspondence with Hwang, she suggests that might be because neighborhood change during that shorter period was generally less pronounced than in the longer period, possibly resulting in a noisier and less informative gentrification measure.

Appendix VII: HCV Search Strings

As explained in the main text, while other kinds of exclusionary language are operationalized with topics from the STM, HCV exclusion is operationalized using a set of search strings that labels each text as exclusionary or not. I developed the list iteratively with feedback from partner community organizations that work with HCV holders. I stopped when my list of phrases correctly identified 200 advertisements with no errors. In total, the list includes 25 phrases, each of which was only present in advertisements that I hand-coded as exclusionary. That is, on its own, each of these search strings will correctly mark some texts as exclusionary, will miss many exclusionary texts, but will never generate false positives. Note that the list includes many variations in spelling and form, includes different ways text writers refer to HCVs, and has variety in the kinds of explanations landlords give for exclusion.

1. no smoking, pets or section 8
2. cannot handle section 8
3. not accept section 8
4. don't accept section 8
5. no section 8

6. no pets or section 8
7. not section 8 certified
8. not participate in section 8
9. not accepting section 8
10. not excepting section 8
11. not accept section 8
12. not except section 8
13. no housing voucher
14. no voucher
15. not excepting any voucher
16. voucher not accepted
17. voucher not excepted
18. no section 8
19. do not accept hud voucher
20. do not accpet voucher
21. will not deal with vouchers
22. not accpeting sec.8
23. no smoking or vouchers
24. not approved for section 8
25. not approved for voucher

While not mentioned in the paper, I use this same set of search strings when working with community partners to identify potentially illegal advertisements in places that have SOI laws, like in Seattle and Chicago.

Appendix VIII: Modeling

The statistical modeling in Chapter 3 is complex. I run five different model specifications for five different outcome variables, including 4 continuous outcomes and one binary outcome. I ran all of the modeling using both frequentist and bayesian methods, and in general the estimates were very similar. The main text has details on the model specification, and mentions that the random slopes models reported were fit using frequentist methods. This appendix explains more about that choice, some of the difficulties in modeling these outcomes, and some ways this analysis could be improved. It takes the form of a narrative memo of the analysis for this chapter.

I began the analysis for this chapter in January 2022 when I decided on the metro-level variables I would use in the analysis. I had no intention of using bayesian methods at that time, though I understood that they might be able to include the error from the STM models even when fit on the test data. I had experience using the lme4 package for hierarchical modeling and used that for my exploratory analysis. Most of the models fit well, but models with random slopes included a warning that the fit was singular. This caused problems for estimating confidence intervals on estimated effects and expected values, problems I have not been able to resolve. However, at the time I read up on what might be causing the warnings. The documentation recommended scaling the data, which I did to no effect, or using a Baysian method that could include weakly informative priors or a non-centered parameterization. I had also learned to fit hierarchical models using BUGS, but I had not done so for years, and was interested in Baysian methods more generally, so I used this error as a motivation to refresh my Baysian knowledge with the hopes of fixing these models.

Those hopes ultimately proved fruitless, though I did learn a lot about Basyian inference, and indeed many of the models reported in this dissertation were fit using Baysian methods. No matter what I tried, however, I have not been able to satisfactorily fit models to my topic-derived outcomes using random slopes. Fitting the models usually takes ~40 hours, and results in warnings that the posterior has not been adequately explored. I have tried using more informative priors and non-centered parameterizations, and those changes improved the model, but not enough for me to be confident in its results. Because of these issues, in this version of the paper I report the lme4 estimates of the quantities of interest for Topic 50 without confidence intervals.

I also want to make a note about the difference in modeling HCV exclusion, which is a binary outcome. I modify the model specification in the main text to account for the binary outcome as shown in (A1). The resulting model uses a binomial likelihood and a logit link. Apart from that part of the model structure, the rest of the modeling, including the covariates, is the same for the HCV analysis as for the continuous outcomes.

(A1)

$$HCV_i \sim Binomial(p)$$
$$logit(p) = \alpha_{metro} + \alpha_{tract}$$
$$\alpha_{metro} \sim Normal(\mu_{metro}, \sigma_{metro})$$
$$\alpha_{tract} \sim Normal(\mu_{tract}, \sigma_{tract})$$
$$\mu_{metro} \sim Normal(0,\ 1)$$
$$\mu_{tract} \sim Normal(0,\ 1)$$
$$\sigma_{metro} \sim Exponential(1)$$
$$\sigma_{tract} \sim Exponential(1)$$

References

Hwang, Jackelyn. 2020 "Gentrification without segregation? Race, immigration, and renewal in a diversifying city." City & Community 19, no. 3: 538-572.

Wilkerson, John, and Andreu Casas. "Large-scale computerized text analysis in political science: Opportunities and challenges." Annual Review of Political Science 20 (2017): 529-544.