Digital Traces of Racial Stereotypes in the Marketing of Neighborhoods on Airbnb

Abstract

This study investigates how place-based entrepreneurs market their neighborhoods to prospective visitors in the face of deeply entrenched stereotypes about neighborhoods of color and the urban core. I leverage place-based geospatial sentiment data from 19,967 Airbnb listings located in Los Angeles, California located in 2,156 census block groups of which 1,316 (61%) were not majority white and apply NLP topic modeling. Hosts' descriptions of predominantly white neighborhoods focus on the physical and landscape aesthetics of the neighborhood, whereas minority-majority neighborhoods include more details about residents as well as the safety of the community. Language that deflects negative stereotypes about minority neighborhoods is most explicit in descriptions of urban core neighborhoods where Black residents are the majority or plurality. Overall, the findings suggest that the neighborhood marketing strategies in Airbnb's online platform transmit existing urban inequalities to prospective tourists by implicitly or explicitly engaging with stereotypes about race and urban life.

Introduction

The Internet and digital peer-to-peer platforms have increased how much and how quickly we can learn about what a city has to offer. In the tourism and leisure marketplace, Airbnb has facilitated information sharing about neighborhoods by providing home or apartment owners with a platform to be place-entrepreneurs who market their property and neighborhood to prospective guests in need of short or long-term home accommodations for their travels (Freytag & Bauder 2018). However, research on Airbnb focuses mostly on its impacts on the structural conditions of the city, such as how Airbnb has transformed the geography of urban tourism away from city-planned tourist bubbles and hotel corridors and into various residential neighborhoods (Fuller & Michel 2014), increased rent prices and reduced the stock of affordable housing (van Holm 2020), and heightened gentrification (Gant 2016). This body of research does not unpack how the Airbnb marketplace, by commodifying urban spaces, contains digital imprints of the existing cultural and structural inequalities of cities and conveys them to prospective tourists. Thus, it is critical to investigate how the neighborhood information provided by Airbnb hosts reflects and refracts existing racial and social inequalities across the urban landscape.

Race is an important dimension to study in the marketing of neighborhoods on Airbnb because research on neighborhood reputations has found strong associations between neighborhood racial composition and neighborhood perceptions. Studies consistently show that the desirability of neighborhoods follows a racial hierarchy in which predominantly black neighborhoods are perceived as low in desirability, especially among white people, because of deeply engrained stereotypes of these neighborhoods as spaces that are prone to social disorder and lacking in high quality amenities such as schools, recreational spaces, and food/dining options (Besbris, Faber, Rich & Sharkey 2018; Logan & Collver 1983; Sampson 2012). Neighborhoods with largely white residents, by contrast, have developed reputations as highly desirable because people perceive higher quality amenities and social stability to be more prevalent. The existence of this hierarchy of neighborhood desirability influences individuals' residential decision-making and contributes to racial segregation within cities, which implies that the demographic, natural, and built environmental features of neighborhoods are often racialized and ultimately shape how neighborhoods are perceived and whether people want to spend substantial amounts of their time in a neighborhood (Lewis 2003; Charles 2006; Krysan et al 2009; Emerson, Chai, & Yancey 2001; Hwang & Sampson 2014). Race plays such a key role in organizing the neighborhood hierarchy in the residential decision-making context that it is logical to inquire whether this pattern extends to the context of new urban tourism ushered in by Airbnb, wherein residential neighborhoods are marketed to outsiders for leisure and recreational consumption. Does the language that Airbnb hosts use to describe their neighborhoods reflect racial stereotypes or racialized perceptions of neighborhoods?

Airbnb brands itself as a travel resource for guests to "see the city like a local", so it is sensible to hypothesize that the relationship between neighborhood racial composition and perceptions of desirability observed in the residential context informs the discursive strategies that Airbnb hosts use to market their neighborhoods to tourists. The findings in the neighborhood perceptions research suggest that hosts in predominantly white neighborhoods will emphasize the features of their neighborhoods that are typically viewed as assets, such as higher quality leisure and dining amenities in their neighborhood descriptions, whereas hosts in neighborhoods with greater shares of racial minorities may adapt their descriptions to counter the negative stereotypes of disorder and disinvestment that are associated with minority neighborhoods. At the same time, hosts may be mindful that the average Airbnb guest is a member of the global middle class and described as cosmopolitan with culturally omnivorous tastes for natural, authentic, and gritty environments that include the presence of racial minorities, artists, and small and independent businesses (Sieler & Wasmuth 2015, Brown-Saracino 2009). Instead of being concerned with countering negative racial stereotypes, hosts in minority-majority and racially mixed neighborhoods may be able to capitalize on the population characteristics of their neighborhoods to appeal to the typical Airbnb guest.

In this study, I test these hypotheses about how the discursive strategies of marketing neighborhoods for leisure may reflect the racial stereotypes about neighborhoods and neighborhood desirability. I analyze neighborhood descriptions from 19,967 publicly available Airbnb rental listings located in Los Angeles, California, which is the largest Airbnb market and among the most diverse in terms of the neighborhood racial composition where properties are located. To analyze the language in listings and identify linguistic differences between neighborhoods with varying racial compositions, I use topic modeling and dictionary correlations. Then, I apply spatial statistics to the topics to identify the topics that display strong patterns of spatial concentration. The language and spatial analyses collectively uncover neighborhood descriptors that have a clear racial and spatial character.

The findings from analyses of the rich corpus of Airbnb listings reveal significant topical differences in the language used to describe majority white neighborhoods compared to non-white neighborhoods. Majority white neighborhoods tend to focus on details about the physical and landscape characteristics of the neighborhood, emphasizing nature and the availability of outdoor recreational and leisure activities. In contrast, the neighborhood descriptions of minority-majority neighborhoods emphasize details about the residents, homes, and urbanicity of neighborhoods. Descriptions of majority Black and Black-White mixed neighborhoods in the urban core demonstrated exceptionally strong evidence of using frames about population diversity and safety to counter negative stereotypes about minority communities and appeal to the prototypical Airbnb guest. The observed linguistic differences between neighborhoods align with the racialization of place geographically (Bonam et al. 2017) and indicate that neighborhood discourse is racialized. Overall, the study illuminates how the discursive neighborhood marketing approaches used in digital Airbnb marketplace can reflect and acknowledge racial inequalities in urban neighborhoods.

Data and Methods

Data. I collected a total of 19,967 unique Airbnb rental listings that were listed online between January and December 2022 from the platform InsideAirbnb.com. Listings come from a total of 11,294 unique Airbnb hosts. Geographic coordinates for each property are also included with the listing text. The map on the left-side of Figure 1 shows the locations of listings within Los Angeles. A spatial join of the Airbnb locations to census block group polygons, which I use as neighborhood boundaries, yielded 2,156 distinct census block groups (71% of total census block groups in the city) with Airbnb properties that serve as the analytic sample for this study. Neighborhoods contained an average of 10 Airbnb listings (SD = 20, min = 1, max = 558) (see Supplementary Figure S1).

The neighborhood racial and socioeconomic demographic data come from the U.S. Census Bureau's 2021 American Community Survey five-year estimates, which is the most recent release of neighborhood data available. Using the census block group identifiers, I joined the demographic variable to the listings and applied a commonly used racial neighborhood typology devised by Crowder and colleagues (2012) that is based on racial proportions for Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian, and Hispanic groups. The typology labels each neighborhood based on the proportions of the top three race groups present. To form these neighborhood race types, I first identified the three racial groups with the highest proportion in the block group. If the largest group is non-Hispanic white and the proportion is over 0.8, then the neighborhood is labeled predominantly white. If the largest group is not white and it makes up more than 50% of the population, then the block group is labeled as predominantly that group. If no group is more than 40% of the block group population, then the neighborhood is labeled mixed. If the first group is greater than 40% and the second group is greater than 10% then the block group is labeled with the name of the first and second groups in that order. The neighborhood race types formed under this typology for Los Angeles include Predominantly White, White Asian Mix, White Black Mix, White Hispanic Mix, Majority Black, Black White Mix, Majority Asian, Asian Mix, Majority Hispanic, Hispanic White Mix, Hispanic Black Mix, and Mixed-No Majority. The map on the right-side of Figure 1 shows the typology applied to block groups in Los Angeles.

Sample Description. Sample characteristics are shown in Table 1. Among the 2,156 neighborhoods, the most common racial types were Majority Hispanic (35%) and White-Hispanic Mixed (20%). The average median household income of neighborhoods was $94,367 (SD = 46,394). Neighborhoods where white residents were the most common group accounted for 39% of the sample, and they tended to be more affluent (average median household income = $110,064 vs. $72,518 for all other neighborhood types).

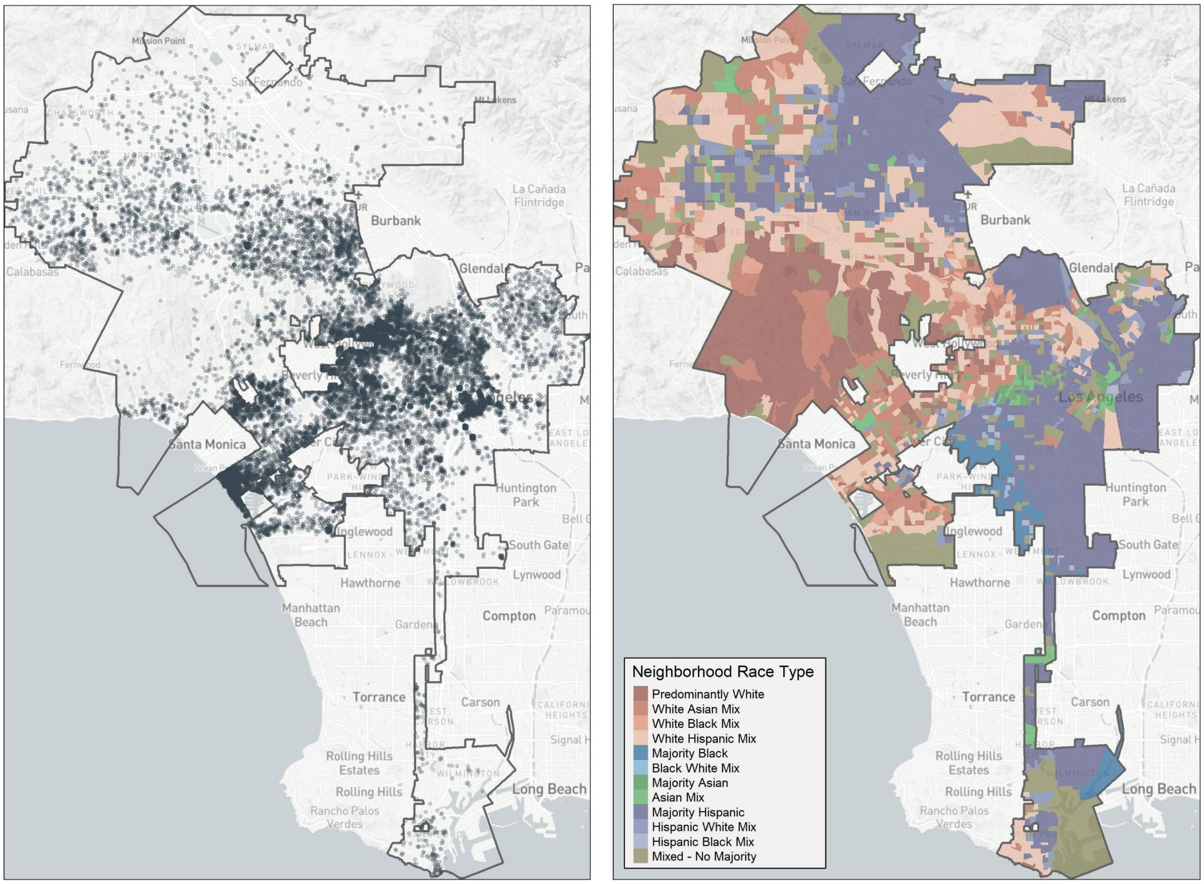

Figure 1. Maps of Airbnb Property Locations and the Neighborhood Racial Typology.

Methods. The analytic approach in this study operates from the perspective that without explicitly mentioning race, neighborhood descriptions can contain racialized language. I define racialized discourse as variations in language use that are associated with differences in neighborhood racial composition. By using NLP approaches, specifically topic modelling and dictionary correlations on Airbnb listing text, the observed association between discursive forms—themes, phrases, ideas, descriptors—and neighborhood racial composition can be detected.

Word Extraction. I determined the relative frequency with which listings and neighborhoods used words (unigrams) by using an open-source Python-based language analysis infrastructure (dlatk.wwbp.org; Schwartz et al. 2017). Individual listings averaged 87 words (SD = 64). On average, neighborhood descriptions had 2,265 (SD = 2,984) words (tokens) for a total of 4,278,221 words (tokens) (see Supplementary Figure S2).

Table 1

Means, standard deviations, and correlations with confidence intervals.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9
1. Predominantly White (>80%)	0.06	0.28									
2. White-Asian Mixed	0.11	0.33	-.12** [-.13, -.10]								
3. White-Hispanic Mixed	0.20	0.47	-.22** [-.23, -.20]	-.26** [-.27, -.25]							
4. Majority Asian	0.02	0.14	-.05** [-.06, -.03]	-.05** [-.07, -.04]	-.10** [-.11, -.09]						
5. Majority Hispanic	0.35	0.38	-.14** [-.16, -.13]	-.17** [-.18, -.16]	-.32** [-.33, -.30]	-.07** [-.08, -.05]					
6. Hispanic-White Mixed	0.05	0.19	-.06** [-.07, -.05]	-.07** [-.09, -.06]	-.14** [-.15, -.12]	-.03** [-.04, -.02]	-.09** [-.10, -.08]				
7. Hispanic-Black Mixed	0.01	0.06	-.02** [-.03, -.01]	-.02** [-.04, -.01]	-.04** [-.06, -.03]	-.01 [-.02, .00]	-.03** [-.04, -.01]	-.01 [-.03, .00]			
8. Majority Black	0.01	0.12	-.04** [-.05, -.02]	-.05** [-.06, -.03]	-.09** [-.10, -.07]	-.02* [-.03, -.00]	-.06** [-.07, -.04]	-.02** [-.04, -.01]	-.01 [-.02, .01]		
9. Black-White Mixed	0.01	0.04	-.01 [-.03, .00]	-.01* [-.03, -.00]	-.03** [-.04, -.01]	-.01 [-.02, .01]	-.02* [-.03, -.00]	-.01 [-.02, .01]	-.00 [-.02, .01]	-.00 [-.02, .01]	
10. Median Household Income (2021$)	$94,366.90	46,394.49	.33** [.32, .34]	.19** [.17, .20]	.09** [.08, .10]	-.11** [-.12, -.09]	-.37** [-.38, -.36]	-.09** [-.10, -.07]	-.03** [-.04, -.02]	-.06** [-.08, -.05]	-.00 [-.01, .01]

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

Topic Modeling. To summarize the content of the listings, I used Latent Dirichlet Allocation topic modeling as provided by the MALLET package (McCallum, 2002) to cluster the content of listings into 100 topics with an alpha of three. Akin to factor analysis, LDA produces clusters of words that occur in similar semantic contexts, yielding semantically coherent topics of related words (Blei, Ng & Jordan, 2003). Duplicate texts reduce the explanatory power of topic models, so duplicate listings from hosts with multiple properties located in the same neighborhood were removed from the sample before modeling topics. Additionally, words that do not contribute to linguistic variability like neighborhood names and the most frequent 55 words were treated as stop words and removed from the listing text prior to modelling. After modeling the topics, I extracted the relative frequency that the neighborhoods in the sample were described using these 100 topics (yielding 100 values per neighborhood), which summarized the language applied to neighborhoods as a distribution over topics.

Language Associations. To determine if a topic was associated with neighborhood composition, I tested the language features in an in-sample linear regression model controlling for median household income. In the models, the neighborhood types are dummy-coded, so that each neighborhood type is compared to all other neighborhood types. I report its standardized regression coefficient (β) with the associated significance, corrected for multiple comparisons. Given that the language analyses are exploratory, to guard against the possibility of spurious findings I corrected the customary $p = .05$ significance threshold with the Benjamini-Hochberg procedure (1995) for multiple comparisons for the number of language features explored.

Correlating Topics and LIWC Dictionaries. Then, I categorize the topics with statistically significant coefficients into established higher order groupings to organize the language patterns around themes. I correlated the topics with Linguistic Inquiry and Word Count (LIWC) dictionary categories (Pennebaker, Boyd, Jordan, & Blackburn 2015). I report the correlation coefficients, with p-values corrected for multiple comparisons, for topics that were positively correlated with LIWC dictionary categories.

Calculating Spatial Autocorrelation of Topics. The topic frequencies were extracted at the listing-level and normalized to capture the prevalence of the topic for a given listing. For each topic, I calculated a global Moran's I test for autocorrelation to determine whether topic usage in listings demonstrated spatial clustering or dispersion. Greater concentration represents a stronger, spatial character of a topic and implies that the topic reflects a common understanding about an aspect of a neighborhood's identity. Then, I ran local-level spatial clustering using the Getis-Ord-Gi* statistic to locate the neighborhoods where topic clustering emerged within the city. I report the topics that demonstrated significant global-level spatial clustering patterns ($p < 0.001$) and provide an accompanying series of maps that shows the hot and cold spots, where a topic was highly prevalent in listings' neighborhood descriptions and absent or sparingly present in listings, respectively.

Results

Topic Associations. Of the 100 topics modeled on the Airbnb listings dataset, I found 46 topics were statistically significant predictors of at least one of the nine types in the neighborhood race types controlling median household income. The caterpillar plots in Supplementary Figures 3 to 5 show the magnitude and direction of coefficients for the topics for each race type. To identify higher level patterns in the topics, I organized the topics into broader themes by identifying topics (n = 26) that demonstrated significant positive correlation coefficients with the LIWC Leisure or Family/Friends/Homes dictionaries. In this section on the topic associations, I address the broader topical themes in the language used in neighborhood descriptions within each race type that emerged from the dictionary correlations.

Leisure Activities and Amenities. The language used to describe leisure activities and amenities differed across racial neighborhood types (see Figure 3). As a group, the whitest neighborhood types had the greatest number of significant topics in this leisure activity/amenities LIWC dictionary category compared to all other race types. The leisure-related topics that surfaced in predominantly white neighborhood descriptions –upscale dining and shopping experiences (*fine dining, Grove, Rodeo Drive*) – can be construed as classed-amenities, which distinguishes the topics from the majority white mixed neighborhood types, which brought up shops and retail establishments (*shops, art, cafes, boutiques*) that did not signal class as clearly. Majority white Asian mixed and majority white Hispanic mixed types showed significant correlations with outdoor and physical activities (*park-going, golfing, tennis, beach-going*) and cultural institutions and arts activities (*theaters, art museum, galleries*). Albeit white residents are not the majority in Hispanic white mixed neighborhood, this type also contains topics associated with outdoor activities (*Echo Lake Park*, *hiking, aquarium, shoreline*) this suggests that as the share of white residents increases, neighborhood descriptions are more likely to surface outdoor activities or points of interest.

On the other hand, Majority Asian, Hispanic Black mixed, and Black white Mixed neighborhoods commonly referenced the many amenities of the downtown district (*nightlife, events, arts, businesses*) for guests to explore. Majority Black neighborhoods focused on proximity to major entertainment points of interest including professional sporting and gambling venues (*stadium, casino, coliseum*). Across the neighborhood types where minorities were the majority, the leisure amenities are part of the built environment associated with urban life.

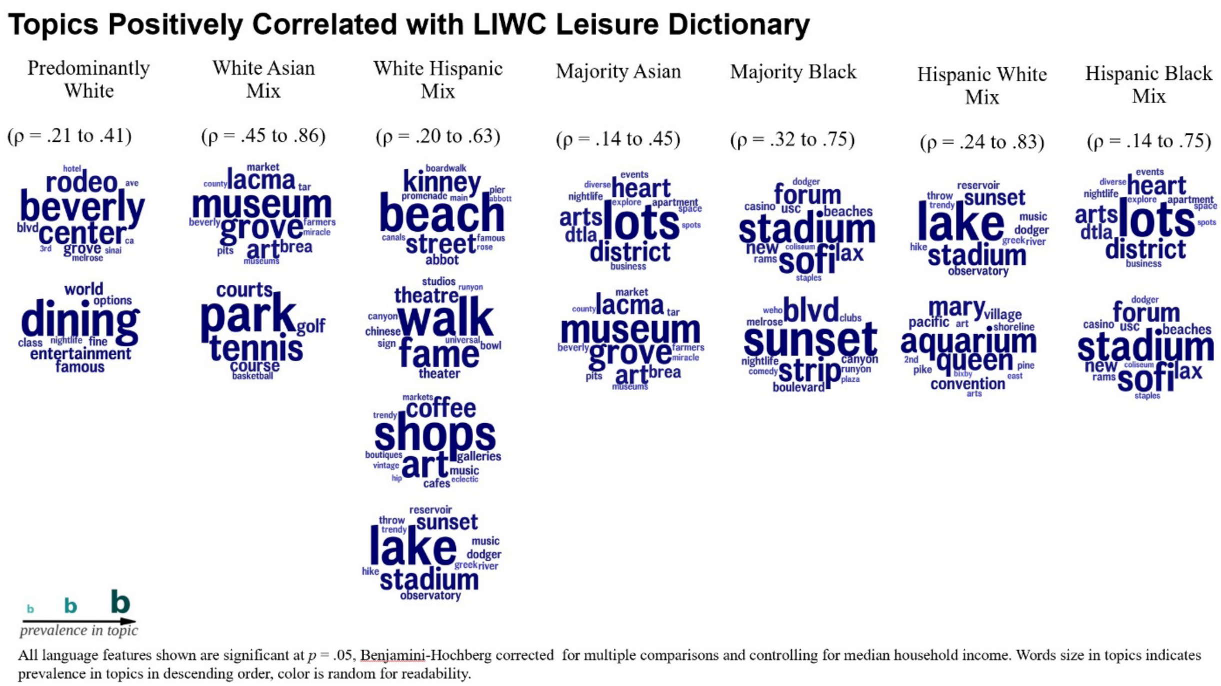

Figure 3. Topics correlated with the LIWC leisure dictionary that were significantly associated with racial types when controlling for median household income (from the 100 topics modeled on the Los Angeles county Airbnb listings corpus). The size of the word in the topic signifies prevalence within the topic, color shading is random for readability.

Family, Friends, and Homes. Topics that were correlated with LIWC's family, friends, and home dictionaries also demonstrated racialized patterns (see Figure 4). Nature and images associated with lower levels of urbanicity appear in the topics positively associated with the LIWC Friends/Family/Home dictionaries for the whitest neighborhood types and the most Hispanic neighborhoods. Predominantly white neighborhoods described the aesthetics of the neighborhood having beautiful views of nature (*views of mountains, hills, tree-lined streets, ocean*) and white Hispanic mixed neighborhoods emphasized their neighborhoods' location in hills (*up, high, canyon*). Descriptions of white Asian mixed neighborhoods focused on the physical layout of the neighborhood in the style of an urban village (*village, nestled, campus, historic*). By emphasizing characteristics of the neighborhood and apartments layout that signal medium-density housing and mixed-use zoning the hosts paint a picture of a less urbanized neighborhood. Similarly, in neighborhoods where Hispanics are the majority or plurality the descriptions contain language that signals more suburbanized style of housing by referencing zoning and home values (*single-family, million dollar*), quiet homes and streets (*private, gated, cul-de-sacs, no traffic, hidden gem*), and trendy college-town retail (*hip, spots*). The similarity across white Asian mixed, majority Hispanic, and Hispanic white mixed neighborhoods are the prevalence of descriptors that diverge from prototypical images of urbanicity. However, the sharpest divergence from high urbanicity appeared in the descriptors in predominantly white neighborhoods, which leaned into natural scenery present in exurban environments.

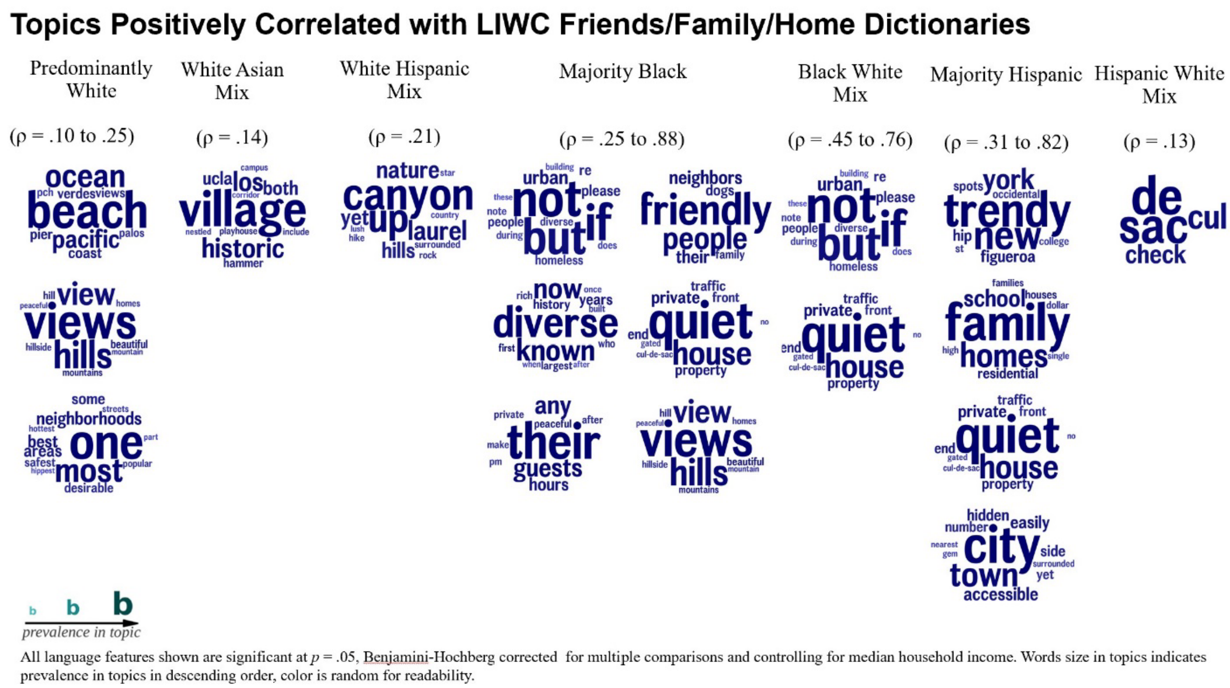

Figure 4. Topics correlated with LIWC Home, Friends, and Family dictionaries that were significantly associated with racial types when controlling for median household income (from the 100 topics modeled on the Los Angeles county Airbnb listings corpus). The size of the word in the topic signifies prevalence within the topic, color shading is random for readability.

In contrast to the majority White neighborhoods, the descriptions of majority Hispanic, Majority Black, and Black-White mixed neighborhoods attended to the demographic characteristics of the people in the neighborhood and physical assurances of safety offered by the housing and built environment. Majority Black and majority Hispanic neighborhood descriptions had the greatest number of topics correlated with this dictionary and the strongest magnitudes of correlations. One common descriptor across majority Black, majority Hispanic, and Black white mix neighborhoods is the image of quiet homes/streets. Noise-level also comes up for majority Black neighborhoods in the form of language that provides rules for guests about curfews and policies (*pm, guests, after hours*) to maintain the peacefulness and quiet of the neighborhood. In Hispanic neighborhoods, the emphasis was on a family-friendly environment (*family-oriented, schools, residential*), whereas descriptors of majority Black neighborhoods emphasized diversity (*culture, ethnicity, histories*) and friendliness (*dogs, neighborly, pleasant*) of the residents. Safety was discussed across both Majority Black and Hispanic neighborhoods, but largely in the context of the physical design of the homes (*cul-de-sacs, no traffic, private property, gated, peaceful, quiet hours, peacefulness, security*). Although safety appeared in the descriptions of predominantly white neighborhoods, it came up in general terms with superlative language that described the neighborhood's desirability (*safest, most desirable, best areas*) but did not feature significant discussion of physical infrastructures or means of offering greater safety. Lastly, "urban" was a common descriptor across Majority Black and Black-White Mixed neighborhoods. However, it is unique from the other topics because it contains negating language in relation to the homeless population (*urban, diverse, but, note, homeless, people)*. Below are two listings that demonstrate how hosts frame their neighborhood as desirable despite the presence of a homeless population:

> *"The area is ridiculously hip. Feel free to search "Arts District Los Angeles" and "Little Tokyo" to read up on the hood. Keep in mind that most parts of Los Angeles these days has (sic) a homeless problem. Downtown Los Angeles and its surrounding areas are not exempt. Luckily for us, we have a well lit building and a great security system."*

> *"LA Times just called us (Jefferson Park) the hottest up and coming neighborhood [sic]...however, We've lived here before it was "up and coming" and still loved it! If you need a Whole Foods, BMWs, and fancy lawns, you might want to stay in Beverly Hills. If you like dollar ice cream trucks, street tamale vendors and neighbors who look out for each other - we'd love to have you! Like most expensive cities, we have homeless people, some blight, and you're likely to hear several languages on my block (I'm Filipino-American and people on my block are of every persuasion). Most who come with open minds find it to be a friendly, diverse, working and middle-class community."*

Altogether, the topics that were most strongly associated with humanizing details provide portraits of the social, natural, and built environments of the neighborhoods and how they relate to neighborhood racial composition. Whereas the leisure-related topics identified neighborhood-level racial differences in the points of interest and activity spaces that are marketed in listings.

Spatial Clustering of Topics. Among the topics significantly correlated with the LIWC Leisure and Family/Friends/Home dictionaries, nine demonstrated statistically significant evidence of spatial clustering in the city's neighborhoods via the Global Moran's I scores. Higher global Moran's scores indicate that use of a topic in listings deviated more from a random distribution, whereas scores with a lower magnitude point to more spatial dispersion of the topic. The Getis-Ord-Gi* statistics reveal the locations of listings that contained language associated with a topic at high rates clustered and that lacked language associated with a topic, which are symbolized as hot and cold spots, respectively. The rate of clustering within compared to outside the urban core is a key differentiating point across these topics that had a spatial character.

Leisure Activities and Amenities. Among the topics correlated with the LIWC leisure dictionary, Upscale Shopping (I=0.390), Ocean/Beach (I=0.235), Downtown (I=0.479), Theaters (I =0.372), and Stadium/Casino (I=0.170) demonstrated statistically significant spatial clustering. Figure 5 presents the listings that had high and low prevalence of each of these topics via hot and cold spots determined with Getis-Ord-Gi* analysis. The Downtown, Stadium/Casino, and Theaters showed some presence of clustering downtown. Of these three topics, the Downtown and Theaters topics are most highly concentrated in downtown neighborhoods. The Stadium/Casino topic, which had the lowest global clustering score of all the significant topics, had the greatest dispersion of hot spots in both downtown neighborhoods and neighborhoods south of downtown around Inglewood. The region where the Stadiums topics are located is home to a large population of Black Angelenos, which implies that these topics have a greater racial character than a highly concentrated spatial character. Hosts invoked amenities that were not within the neighborhood where their property was located but may be seen as points of interest to members of the community and thereby essential to market to prospective guests seeking to explore the city as a local.

In neighborhoods outside of the urban core, the Upscale Shopping and Ocean/Beach topics demonstrated strong patterns of spatial clustering. Upscale Shopping was prevalent in the descriptions of Los Angeles' neighborhoods surrounding Beverly Hills, like La Brea, Cheviot Hills, Pico Robertson, and Fairfax. Cold spots appeared in downtown neighborhoods, where neighborhood descriptions had no mentions or infrequent mentions of the amenities and venues associated with upscale shopping. The Ocean/Beach topic had hot spots in neighborhoods clustered along the Pacific ocean and cold spots in downtown neighborhoods. Unsurprisingly, these topics possess a highly apparent spatial character that follows the distribution of the amenities in the physical and built environment of the city.

Family/Friends/Homes. Figure 6 presents the hot and cold spot maps for the four topics in the Family/Friends/Homes dictionary that have significant spatial concentrations: Canyon/High Up (I=0.251), Hills/Views (I=0.200), Trendy (I=0.436), and Urban, but (I=0.172). The qualified urban topic – Urban, but – was most highly concentrated in downtown Los Angeles and driven largely by Black-White mixed neighborhoods, but hot spots are more dispersed than other topics correlated with the Family/Friends/Homes dictionaries. Other small pockets of hot spots occur in listings around Hollywood, Brentwood, and Venice Beach. Even though one of the example quotes described homelessness as a problem throughout the city, the map does not reveal hot spots dispersed widely across listings in the city. The absence of widespread dispersion of listings that reference homelessness and the negating language of the Urban but topic indicates that Airbnb hosts in neighborhoods outside of downtown neighborhoods are not making it a point to address the homelessness in their neighborhoods or the city's crisis more broadly. Presumably, hosts in neighborhoods outside the urban core can avoid touching on the homeless population in their neighborhood by directing attention to other attributes of their neighborhood. For hosts in urban core neighborhoods, the denser and smaller spatial scale of the built environment potentially makes the homelessness crisis appear to be a bigger part of urban life. Therefore, hosts in these neighborhoods may feel pressure to grapple with this negative stereotype more directly than other negative racial stereotypes.

On the other hand, the Trendy, Canyons/Up High, and Hills/Views topics were not as prevalent in the urban core. The Trendy topic, which had the strongest city-wide autocorrelation, was highly concentrated in listings along Los Angeles' urban periphery, just to the east of Pasadena and south of Glendale. The geographic location of the hot spot for the Trendy topic matches the language of low density/suburban-ness that also appeared in some majority Hispanic neighborhoods. The language and spatial analyses indicate that the trendy descriptor for neighborhoods has a counterintuitive racial and spatial character. The locations of Canyons/Up High and Hills/Views topics, which described the aesthetics of predominantly white and white mixed neighborhoods, are not surprising because they overlap with neighborhoods where nature is most prevalent. Spatial clusters for these topics are located within the mountainous and hilly region to the north-west of downtown near Franklin Canyon Park, Laurel Canyon, and Beverly Glenn as well as near the coast around the Pacific Palisades.

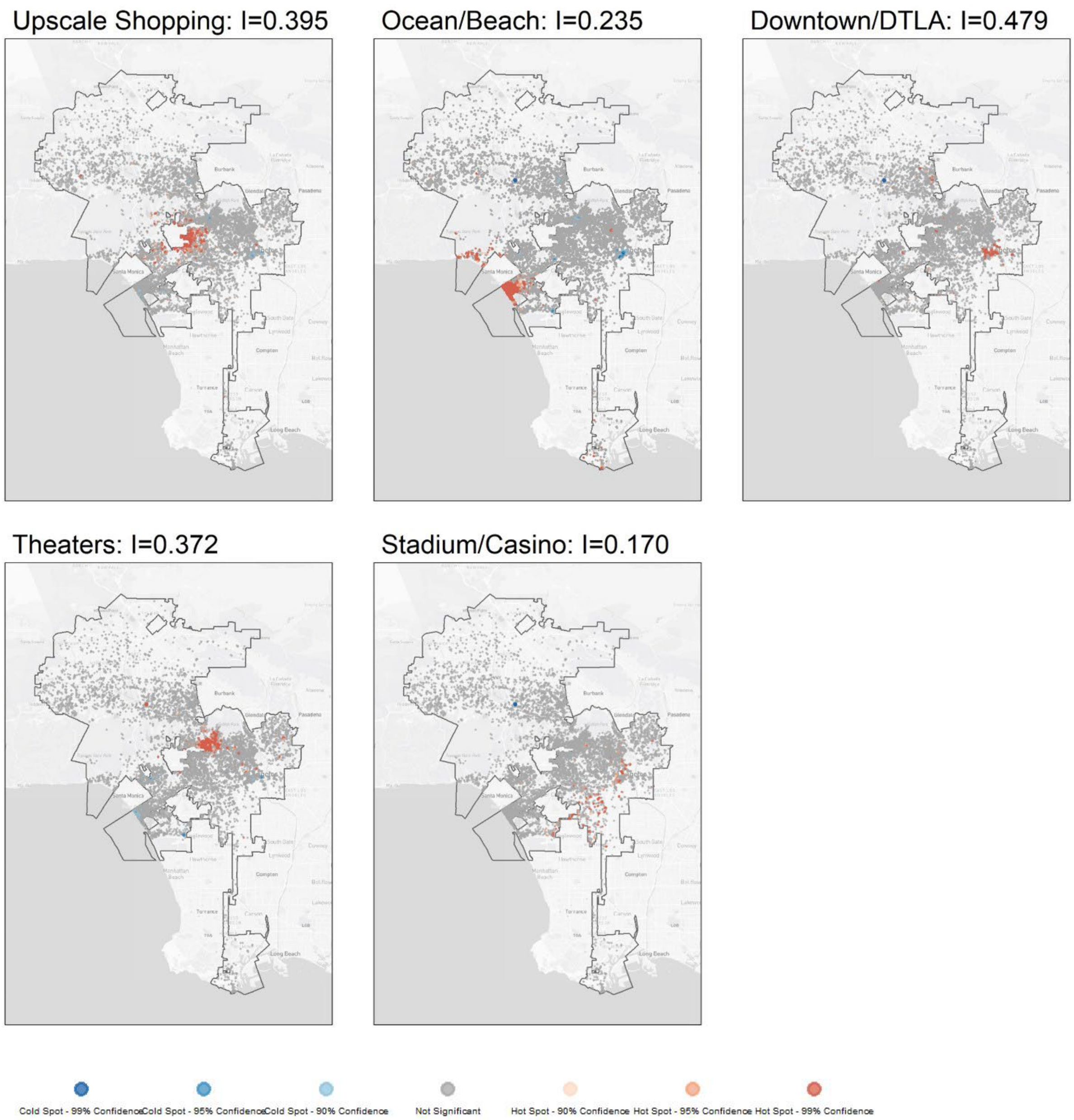

Figure 5. Leisure topics with significant Moran's *I* scores ($p<0.001$). Colors reflect local hot and cold spots, where topics were highly present and absent, respectively.

The spatial autocorrelation and hot-spot analyses confirm that neighborhoods reference leisure amenities located in or around their neighborhood rather than selling amenities in the city at large. On the other hand, the humanizing language used in topics like Trendy and Urban but have an important spatial dimension that the spatial statistics identify.

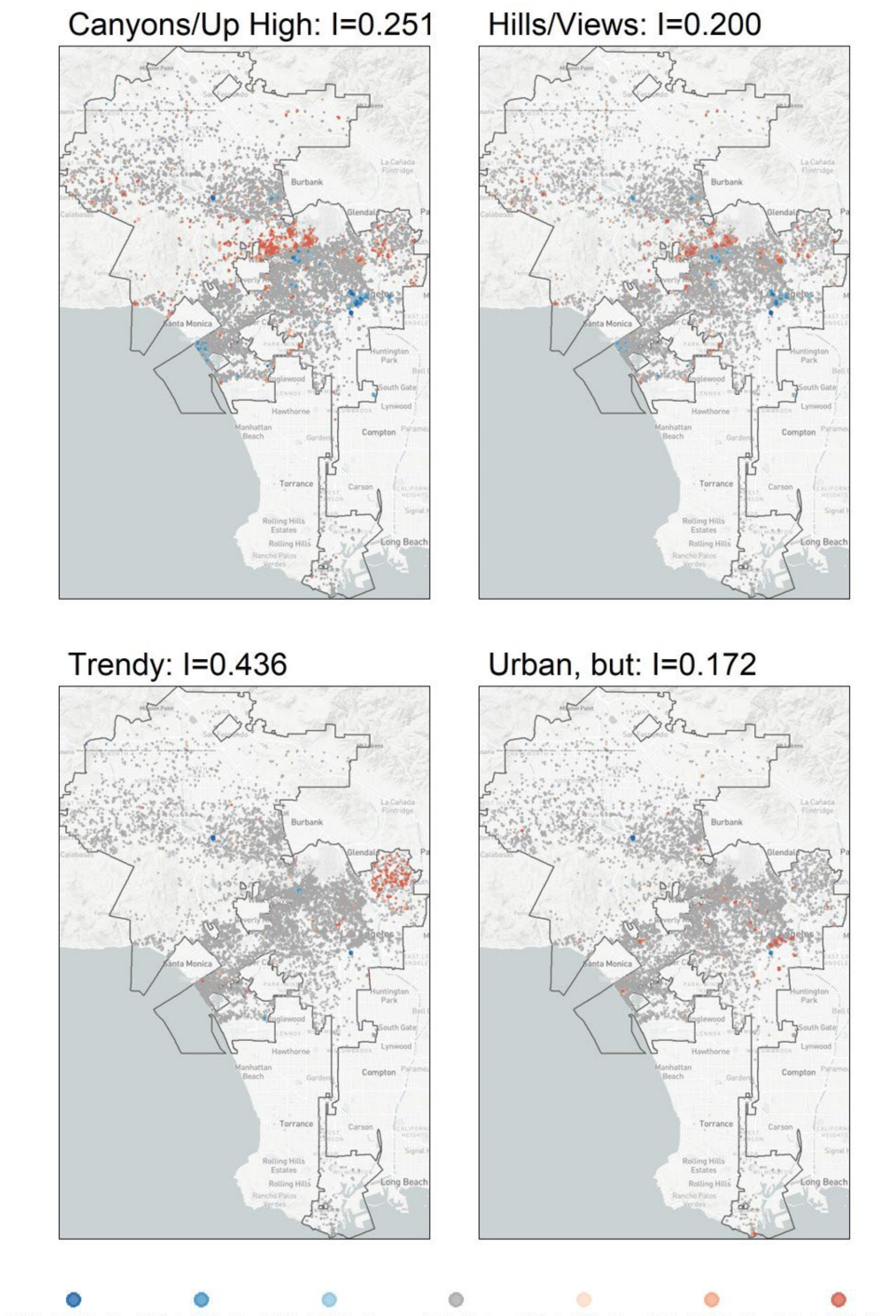

Figure 6. Topics correlated with Home/Family/Friends dictionaries that have significant Moran's *I* scores. Colors reflect local hot and cold spots, where topics were highly present and absent, respectively.

Discussion

The study has three major findings. Firstly, frames relating to nature and natural aesthetics were common in the neighborhood descriptions of leisure and other humanizing details in the whitest neighborhoods. Secondly, the frames in majority-minority neighborhoods focus on points of interest in the urban core and countering stereotypes about disorderly or unsafe communities with images of family friendly and diverse residential neighborhoods. Thirdly, urbanity, especially regarding homelessness and density, is featured most explicitly in descriptions of neighborhoods with a Black majority or Black plurality within and adjacent to the urban core, but implicitly in other neighborhood types.

The fixation on nature spaces and natural aesthetics in the listings located in predominantly white neighborhoods dovetails with research on the recreational and aesthetic tastes of the white urban gentry. Case studies of neighborhoods in the early stages of gentrification show that affluent white residents who are new to a neighborhood have typically had the most say in customizing the neighborhood's public spaces to their preferences, which often resulted in the creation of new public green spaces or upgrading of existing parks (Tissot 2011; Mele 2000; Zukin et al. 2009). The white gentry's preference for "natural" things has also been seen in the application of natural elements like wood, plants, open-air concepts within their homes and business establishments (Zukin 1987; Brown-Saracino 2012). Several durable associations have shaped the race-city-nature relationship: the formulation of nature as 1) an ideal type in Europe and North America concurrent with rapid urbanization ("the gray threat") after 1750 and 2) a culturally and economically valuable resource for improving the ills of urban life and home values, defined, and controlled by white individuals and white institutions (Loughran 2017). These historical associations have influenced the racialized nature-based practices like hiking or sailing, which are linked to inequalities in cultural capital (Bourdieu 1984; Finney 2014). These structural and cultural factors play into the framing with nature that appears in the whitest neighborhoods as well as the frames featuring the hard-physical infrastructure of public spaces and points of interest in the urban core that were referenced in majority-minority neighborhoods.

In minority-majority neighborhoods, the focus on the characteristics of residents, homes, and zoning patterns suggests that hosts are trying to counter negative stereotypes about minorities and minority communities. A substantial body of research has found that perceptions of minorities as lacking warmth and their communities being prone to social and physical disorder serve as the basis for prospective home buyers' and renters' negative evaluations of minority neighborhoods in the residential decision-making process (Hwang & Sampson 2014; Emerson, Chai, & Yancey 2001). Therefore, the language emphasizing family-friendliness, quietness, and safety afforded by physical infrastructures like security systems, gates, lighting, set-back streets, implicitly acknowledges the deeply entrenched racialized stereotypes people of color and minority spaces. Although the marketing of majority Black, and Black white mixed neighborhoods to a lesser extent, as diverse would appear to contradict the racialized hierarchy of neighborhood preferences in residential decision-making, this tactic of highlighting diversity may reflect Airbnb hosts' awareness of the tastes and preferences of the modal Airbnb guest. Prior studies have profiled Airbnb guests and labeled them cosmopolitans with omnivorous tastes for natural, authentic, and "gritty" environments that include the presence of racial minorities, artists, and small and independent businesses (Sieler & Wasmuth 2015, Brown-Saracino 2009). Thus, the framing of these neighborhoods with a majority Black population as diverse and historically noteworthy would make this neighborhood more desirable to the average prospective tourist.

Lastly, the tendency for neighborhood descriptions to include frames that situate neighborhoods along the urban-suburban spectrum also aligns with how urban space has been racialized and strongly associated with disorder and incivility (Wittmer & Parizeau 2016). The frames associated with suburban or peri-urban qualities like “single family homes”, “cul-de-sac”, and “urban village” mentioned in the white Asian and majority Hispanic neighborhoods stand in contrast to high density built and social environments in cities. The contrast becomes sharper when considering that other topics in these neighborhoods surface quiet-surroundings, privacy, and security - all elements that deflect from images of the physical and social disorder that tends to come along with density. On the other hand, descriptions of Black majority and Black plurality neighborhoods grapple with urbanity more directly. The term “urban” is often used as a synonym for Black or Blackness, so the strength of the association between the qualified urban (urban-but) topic and these racial types is powerful evidence of the stickiness of Blackness with urbanism. It is noteworthy that the framing of urbanity by hosts in these neighborhoods softens the “urbanity” of downtown neighborhoods by acknowledging homelessness as an undesirable element of the neighborhood and cushioning it with other frames that resonate with greater desirability such as 1) the physical infrastructure of security afforded by the building, property, and city and 2) the level of diversity in the community in terms of economic, ethnic, cultural, and lifestyle. This framing to soften the neighborhood’s urban image runs counter to the preference for “grittiness” among young cosmopolitans (Brown-Saracino 2009), but at the same time tries to re-package homelessness as part of the neighborhood’s “diversity”, which young cosmopolitans typically view positively. In sum, the frames employed in marketing neighborhoods that include more non-white residents imbue racial stereotypes with frames that appeal to the cosmopolitan tastes of the modal Airbnb guest.

Limitations. Although the validity of these conclusions of the city of Los Angeles is robust, the findings are limited in their generalizability to other cities. However, Los Angeles has the largest and most diverse Airbnb market by neighborhood composition of all markets listed on InsideAirbnb.com, which suggests that the findings could be applied to other highly diverse urban contexts. Otherwise, the current study is subject to limitations associated with data from online peer-to-peer platforms – geographic sparsity and non-random samples of places. Further, findings reported here are correlational, and thus cannot be used to infer causality.

Future Directions. Future work should further examine how elements of urbanity are negotiated across neighborhoods with different levels of density and economic disadvantage by expanding the sample to include cities with different urbanization and demographic profiles. Additionally, linking the frames found in this analysis to the perceptions of Airbnb guests via analysis of the written reviews from guests will shed light on how they evaluate the urban and non-urban qualities of the neighborhood that hosts described.

Conclusion. The current study demonstrates that the language Airbnb hosts use to market their neighborhoods to prospective tourists varies systematically according to the racial composition of the neighborhood, net of socioeconomic status. There are racial differences in the details provided about a neighborhood’s leisure amenities and opportunities as well as the homes, neighborhood residents, and community atmosphere. Hosts use racialized stereotypes that would likely appeal to the perceived normative leisure and cultural tastes of prospective guests while also deploying frames that run counter to or soften stereotypes to get around normative perceptions of disorder and disarray associated with minority, urban neighborhoods. Computational language analysis is a promising approach for learning about how efforts to market and brand neighborhoods for leisure and recreational purposes include racialized notions of space.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1), 289-300.

Besbris, M., Faber, J. W., Rich, P., & Sharkey, P. (2018). The geography of stigma: Experimental methods to identify the penalty of place. *Audit studies: Behind the scenes with theory, method, and nuance*, 159-177.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of Machine Learning Research, 3(Jan), 993-1022.

Bonam, C. M., Taylor, V. J., & Yantis, C. (2017). Racialized physical space as cultural product. *Social and Personality Psychology Compass*, *11*(9), e12340.

Brown-Saracino, J., & Ghaziani, A. (2009). The constraints of culture: Evidence from the Chicago Dyke March. *Cultural Sociology*, *3*(1), 51-75.

Brown-Saracino, J. (2012). Power at the Roots: Gentrification, Community Gardens, and the Puerto Ricans of the Lower East Side.

Brown-Saracino, J., & Parker, J. N. (2017). 'What is up with my sisters? Where are you?' The origins and consequences of lesbian-friendly place reputations for LBQ migrants. *Sexualities*, *20*(7), 835-874.

Charles, C. Z. (2006). *Won't you be my neighbor: Race, class, and residence in Los Angeles*. Russell Sage Foundation.

Crowder, Kyle, Jeremy Pais, and Scott J. South. 2012. Neighborhood diversity, metropolitan constraints, and household migration. *American Sociological Review,* 77(3):325–53.

Dillman, K. N., Horn, K. M., & Verrilli, A. (2017). The what, where, and when of place-based housing policy's neighborhood effects. *Housing Policy Debate*, *27*(2), 282-305.

Duke, J. (2010). Exploring homeowner opposition to public housing developments. *J. Soc. & Soc. Welfare*, *37*, 49.

Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., ... & Ungar, L. H. (2021). Closed-and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods*, 26(4), 398.

Emerson, M. O., Chai, K. J., & Yancey, G. (2001). Does race matter in residential segregation? Exploring the preferences of white Americans. American sociological review, 922-935.

Finney, C. (2014). Black faces, white spaces: Reimagining the relationship of African Americans to the great outdoors. UNC Press Books.

Freytag, T., & Bauder, M. (2018). Bottom-up touristification and urban transformations in Paris. *Tourism Geographies*, *20*(3), 443-460.

Füller, H., & Michel, B. (2014). 'Stop Being a Tourist!'New Dynamics of Urban Tourism in Berlin-Kreuzberg. *International Journal of Urban and Regional Research*, *38*(4), 1304-1318.

Gant, A. C. (2016). Holiday rentals: The new gentrification battlefront. *Sociological Research Online*, *21*(3), 112-120.

Gotham, K. F. (2014). Race, real estate, and uneven development: The Kansas City experience, 1900-2010 (2nd ed.). Albany, NY: State University of New York Press.

Hwang, J., & Sampson, R. J. (2014). Divergent pathways of gentrification: Racial inequality and the social order of renewal in Chicago neighborhoods. American sociological review, 79(4), 726-751.

Jackson, C., & Jones, N. (2012). Remember the Fillmore: The lingering history of urban renewal in Black San Francisco.

Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016). Gaining Insights From Social Media Language: Methodologies and Challenges. *Psychological Methods.*

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

Lareau, A., & Goyette, K. (Eds.). (2014). *Choosing homes, choosing schools*. Russell Sage Foundation.

Lewis, A. E. (2003). *Race in the schoolyard: Negotiating the color line in classrooms and communities*. Rutgers University Press.

Logan, J. R., & Collver, O. A. (1983). Residents' perceptions of suburban community differences. *American Sociological Review*, 428-433.

Logan, J. R., & Molotch, H. (1987). *Urban fortunes: The political economy of place*. Univ of California Press.

Logan, J. R., Whaley, R. B., & Crowder, K. (1997). The character and consequences of growth regimes: An assessment of 20 years of research. *Urban affairs review*, *32*(5), 603-630.

Maitland, R., & Newman, P. (2009). Conclusions for world tourism cities. Routledge.

McCallum, A. K. (2002). Mallet: A machine learning for language toolkit.

Parker, J. N. (2018). Negotiating the space between avant–garde and “hip enough”: Businesses and commercial gentrification in Wicker Park. *City & Community*, *17*(2), 438-460.

Pearce, J. (2012). The ‘blemish of place’: stigma, geography and health inequalities. A commentary on Tabuchi, Fukuhara & Iso. *Social Science and Medicine*, *75*(11), 1921.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC 2015.

Rosen, E. (2017). Horizontal immobility: How narratives of neighborhood violence shape housing decisions. *American Sociological Review*, *82*(2), 270-296.

Sampson, R. J. (2012). *Great American city: Chicago and the enduring neighborhood effect*. University of Chicago Press.

Sampson, R. J., & Raudenbush, S. W. (2004). Seeing disorder: Neighborhood stigma and the social construction of “broken windows”. *Social psychology quarterly*, *67*(4), 319-342.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinki, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, 8, e73791.

Schwartz, H. A., Giorgi, S., Sap, M., Crutchley, P., Ungar, L., & Eichstaedt, J. C. (2017). DLATK: Differential language analysis ToolKit. In Proceedings of the 2017 Conference on *Empirical Methods in Natural Language Processing*: System Demonstrations (pp. 55-60).

Semyonov, M., & Kraus, V. (1982). The social hierarchies of communities and neighborhoods. *Social science quarterly*, *63*(4), 780.

Small, M. L. (2002). Culture, cohorts, and social organization theory: Understanding local participation in a Latino housing project. *American Journal of Sociology*, *108*(1), 1-54.

Tach, L. M. (2009). More than bricks and mortar: Neighborhood frames, social processes, and the mixed–income redevelopment of a public housing project. *City & Community*, *8*(3), 269-299.
Trounstine, J. (2018). *Segregation by design: Local politics and inequality in American cities*. Cambridge University Press.
van Holm, E. J. (2020). Evaluating the impact of short-term rental regulations on Airbnb in New Orleans. *Cities*, *104*, 102803.
Wittmer, J., & Parizeau, K. (2016). Informal recyclers' geographies of surviving neoliberal urbanism in Vancouver, BC. Applied Geography, 66, 92-99.
Zukin, S. (1987). Gentrification: culture and capital in the urban core. *Annual review of sociology*, *13*(1), 129-147.
Zukin, S., Trujillo, V., Frase, P., Jackson, D., Recuber, T., & Walker, A. (2009). New retail capital and neighborhood change: Boutiques and gentrification in New York City. *City & Community*, *8*(1), 47-64.

Supplementary Figures

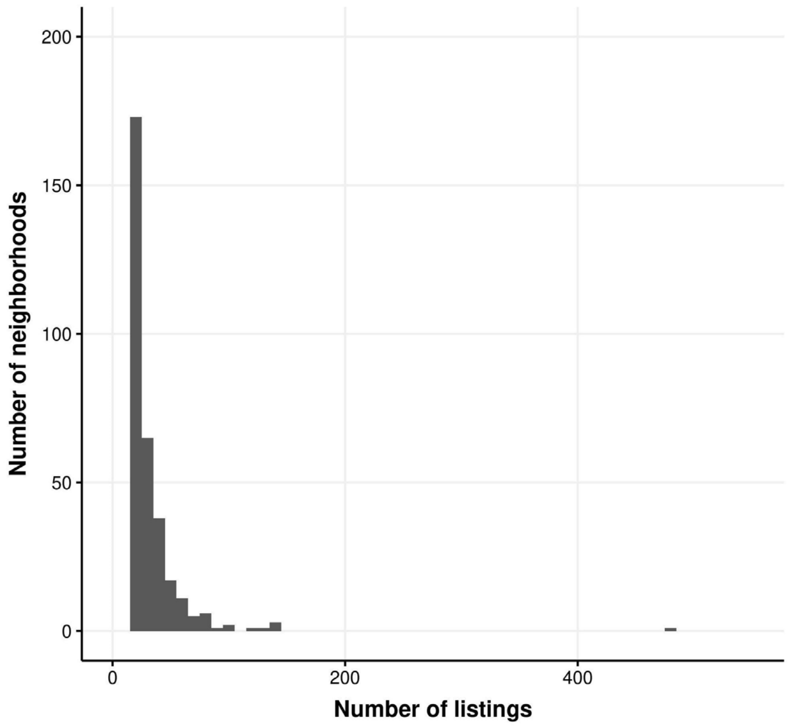

Figure S1. Histogram summarizing listing counts across neighborhoods.

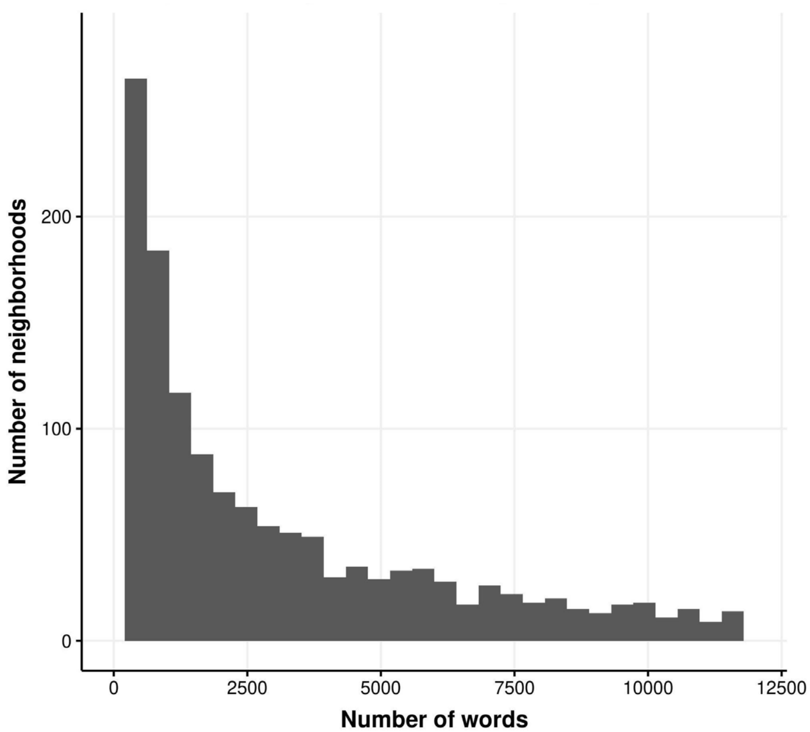

Figure S2. Histogram of words per neighborhood

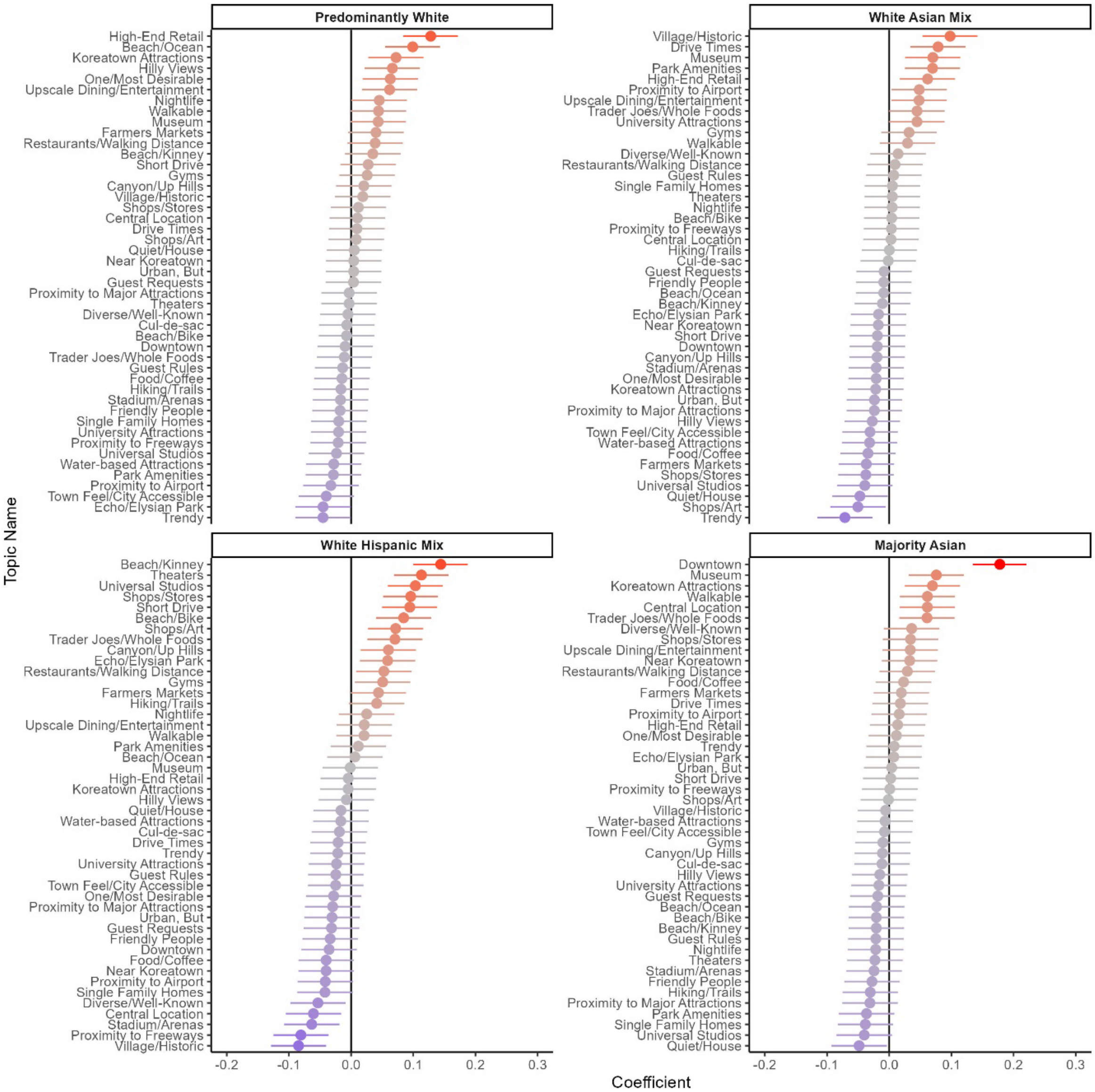

Figure S3. Topic coefficients for Predominantly White; White Asian Mix; White Hispanic Mix; and Majority Asian neighborhoods. *Note: White Black Mix and Asian White Mix neighborhoods are not featured because no topics had significant associations with these neighborhood types.*

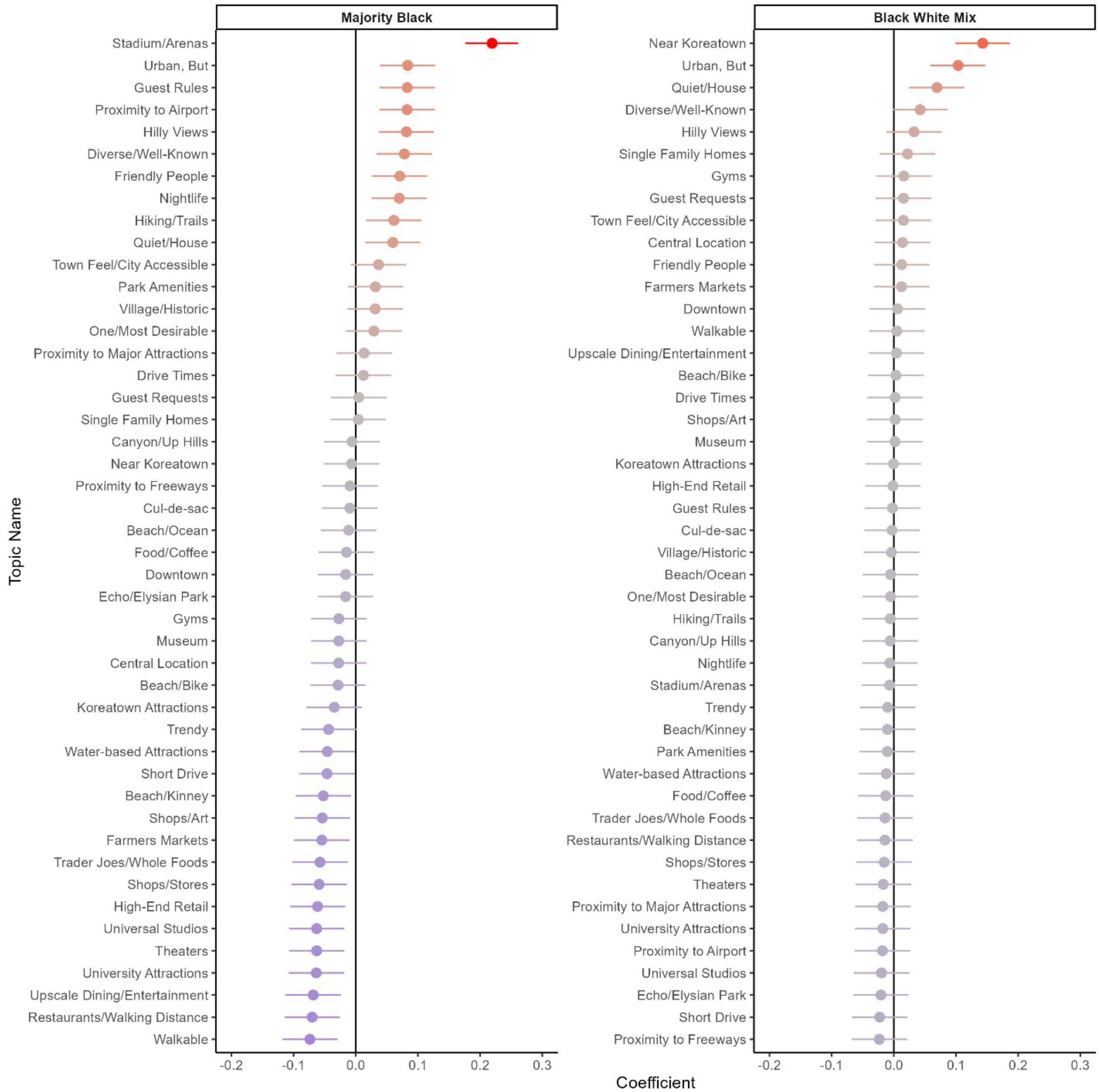

Figure S4. Topic coefficients for neighborhood race types with the greatest shares of Black residents. *Note: Black Hispanic Mix and Black Asian Mix neighborhoods are not featured because no topics had significant associations with these neighborhood types.*

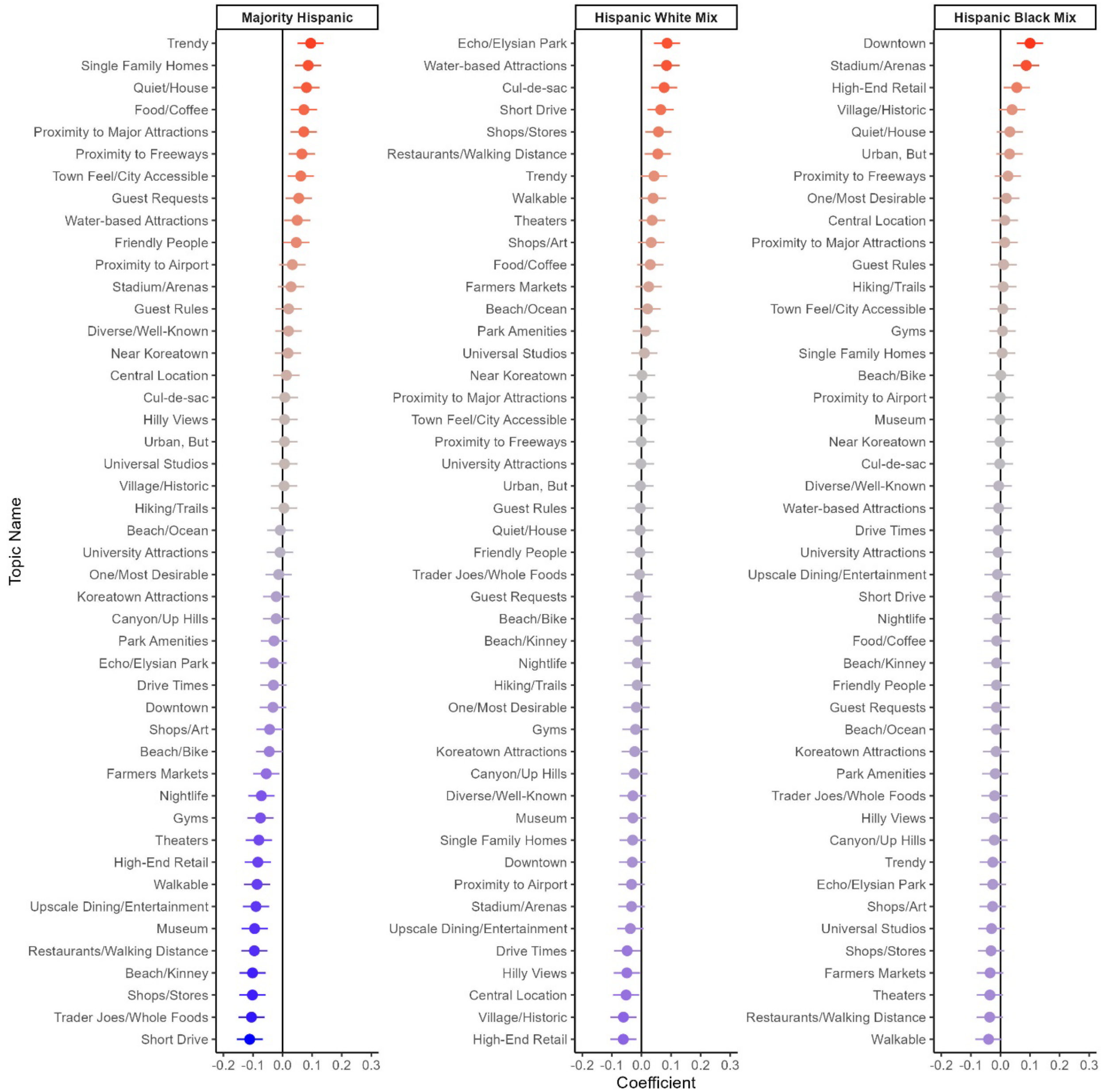

Figure S5. Topic coefficients for the neighborhood race types with the greatest shares of Hispanic residents. *Note: Hispanic Asian Mix neighborhoods are not featured because no topics had significant associations with these neighborhood types.*