

Environment, Racialization, and Territorial Stigma: Evidence from HOLC Texts

*Jonathan Tollefson
Department of Sociology
Mansueto Institute for Urban Innovation
University of Chicago

Abstract This paper uses network-based text analysis approaches to map the conceptual structure of HOLC assessment documents and to analyze how assessment discourses shaped HOLC grading decisions. I find that HOLC assessments reflected and mobilized a particular sociospatial imaginary: one that juxtaposed a close analysis of urban context with a generalized, legible assessment of urban space and value. I further argue that the HOLC assessment process articulated race, space, and value through a fundamentally socio-environmental framework: That urban built and natural environments formed a crucial terrain linking ideologies of racial “threat” to urban structure and logics of valuation. Supporting this argument, I find that natural and built environments were a key factor contributing to the grades assigned by HOLC cartographers: Neighborhoods that were assessed to be dirtier, more polluted, and in closer proximity to environmentally hazardous industry were more likely to be assigned low HOLC grades - and thus more likely to suffer the long-term impacts associated with being shut out from mortgage resources. Together, these findings provide a basis for a structured understanding of the HOLC assessment process, and provide evidence for the importance of urban environments in the early formation of segregated and unequal urban geographies.

Intro

Though urban environmental inequality is well-understood in the post-EPA period (see Mohai, Pellow, and Roberts 2009), rapid, iterative transformation in the urban built environment and poor environmental record-keeping on the part of private industry and state and local regulatory bodies mean that pre-EPA environmental site data are largely unavailable. As a result, sociologists have little systematic understanding of how urban environments shaped the early formation of neighborhood-level inequality (Frickel and Elliott 2018; Tollefson 2026).

This paper addresses this problem by sidestepping limitations in environmental site data entirely. Instead, I turn to the “Residential Security Maps” published by the Home Owners’ Loan Corporation (HOLC) in the 1930s. Throughout the latter half of the 1930s, HOLC deployed an army of researchers to document neighborhood built environments and sociodemographic conditions in order to guide new federally-backed mortgage policies. Neighborhoods were assigned one of four mortgage security grades: A (green, “best”); B (blue, “still desirable”); C (yellow, “declining”); or D (red, “hazardous”). Following the digitization of the HOLC “redlining” maps (Nelson et al. 2023), scholars, policymakers, and mass media have devoted significant attention to how redlining may have shaped neighborhoods over the decades following the HOLC mapping process (see Candipan and Tollefson 2024). Much of this work has argued that the mortgage security grades, in particular, significantly impacted neighborhood trajectories over the course of the 20th century, with consequences for long-run patterns of neighborhood inequality.

Along with the A-B-C-D grades, however, HOLC assessors also produced narrative “area descriptions” for each neighborhood, which consisted of paragraph-length texts meant to explain and contextualize the alphabetic grades they assigned. Area descriptions provide a rich account of how HOLC assessors perceived the conditions in specific neighborhoods and how their perceptions of neighborhood conditions informed the assignment of mortgage security ratings. HOLC area descriptions thus hold clues as to the role of urban environments in the HOLC grading process, allowing us to bypass more direct measures of neighborhood environmental conditions that, for various reasons, are extremely difficult to come by for the historical period in question. Doing so can address several gaps in sociologists’ understanding of the role of urban environments in the early formation of neighborhood inequality and segregation.

To extract information on urban environments from HOLC description systematically across the full range of US cities visited by HOLC assessors, I treat HOLC area descriptions as a text corpus, and use an emerging set of network-based computational text analysis tools (Hannud Abdo et al. 2022) to model the thematic structure of the grading decisions made by HOLC cartographers. I then link the thematic structure of the HOLC area descriptions to the grades assigned by HOLC cartographers. This approach makes it possible to assess how

various neighborhood characteristics shaped HOLC grading decisions in addition to - and beyond - the well-studied question of neighborhood demographic and racial and ethnic segregation.

I use this topic landscape alongside population estimates interpolated to HOLC boundaries to assess the relational interface between racial “threat” and (perceptions of) environmental risk in the articulation of a form of territorial stigmatization (cf. Wacquant, Slater, and Pereira 2014) specific to the early 20th century, by way of a critical perspective on the production of official knowledge and a Du Boisian understanding of the role of social-environmental relations in racialization (see Bhardwaj 2023). First, I use a random forests approach to select the set of discursive topics that best predict HOLC grading decisions. I then investigate whether a discursive analysis of HOLC documents adds predictive power compared to neighborhood demographics alone. Finally, I evaluate the coproduction of race and environment in the formation of territorial stigmatization through a set of models interacting neighborhood demographic information with topic content.

This paper finds that HOLC assessment were organized according to a particular sociospatial logic, providing the first structured analysis of HOLC assessment discourses across the full range of neighborhoods visited by HOLC cartographers. I further show that natural and built environments were a key factor contributing to the grades assigned by HOLC cartographers: Neighborhoods that were assessed to be dirtier, more polluted, and in closer proximity to environmentally hazardous industry were more likely to be assigned low HOLC grades - and thus more likely to suffer the long-term impacts associated with being shut out from mortgage resources.

Understanding how urban environments factored into the HOLC grading process illuminates the role of the environment in the early formation of neighborhood inequality, for two reasons: First, it can tell us how these central policy actors in this key moment in the (re)production of urban inequality imagined the relationship between urban environments and neighborhood security, risk, and desirability. More broadly, because the grades assigned by HOLC assessors themselves are associated with long-term impacts on neighborhood inequality and racial segregation, understanding the role of urban environments in the HOLC grading schema can reveal, directly, some of the forces that shaped an important mechanism of long-term urban inequality.

Data, variables, methods

Data

HOLC data

HOLC data are drawn from the *Mapping Inequality* project (Nelson et al. 2023), an effort by the University of Richmond Digital Scholarship Lab to digitize and

georeference HOLC maps across 196 US cities. The full *Mapping Inequality* data set contains information on 8,878 neighborhoods, 7,588 of which contain non-empty area description text data.

HOLC recorded area descriptions on standardized forms, with nine different form types used across the range of cities included in the *Mapping Inequality* dataset. Most area description forms included a range of numeric entries listing neighborhood demographics, housing characteristics, and mortgage and tax information, alongside three narrative sections: “favorable influences,” “unfavorable influences,” and “clarifying remarks.” To treat each neighborhood as a single unit in the corpus, I concatenate these three narrative sections into a single block of text for analysis.

While most of the AD forms for cities in the *Mapping Inequality* data set follow this general format, a small number of cities relied on AD forms that gathered all the numeric and descriptive data into a single paragraph, or, in some cases, omitted narrative description entirely. In order to maintain comparability across neighborhoods, I focus on cities with distinct and extensive narrative assessments, a data set totaling of 6,749 neighborhoods across 127 cities (full list is available in Appendix). This final subset represents 89% of the non-empty area descriptions in the *Mapping Inequality* data set.

Of the 6,749 narrative assessment documents analyzed, 713 (11%) were assigned an “A” grade; 1710 (25%) “B”; 2720 (40%) C; and 1606 (24%) a “D” grade. The distribution of HOLC grades varied greatly by city. Several cities - Lexington, MA, Hartford, CT, and the greater Stamford, CT area - consisted of 30% or more “A”-graded neighborhoods. At the other end of the spectrum, 20 cities, notably including Brooklyn, Queens, and Boston, were assigned a single “A” area; meanwhile, two small industrial cities - Springfield, IL and Everett, MA - consisted of more than three-quarters “D”-graded areas, with the remaining neighborhoods assigned “C” (or “Definitely Declining”) grades by HOLC assessors.

Text data are cleaned and prepared for topic analysis by removing punctuation, proper nouns, directional markers and frequent place-specific terms (e.g., “borough,” “boulevard,” etc).

Constructing a linked HOLC-census data set

My central concern is with the interaction between assessment discourses and neighborhood demographics, aligning with the interests of the current paper in the relational processes tying policy discourses on environment, housing, and neighborhood context to race and space. Accordingly, I begin by intersecting HOLC neighborhood assessments with enumeration district-level demographic data from the 1930 and 1940 US censuses. To do so, I use areal apportionment methods to reassign demographic data from census districts to HOLC boundaries, drawing on census maps from the Urban Transition HGIS project at Brown University (Logan and Zhang 2018).

HOLC assessments often extended into close-in suburbs that are absent from extant spatial census data from the period. With this in mind, I limit intersected data according to two criteria. First, I include only those HOLC areas where at least 50% of the region overlaps with available census boundaries. I then limit my analysis to cities where at least 50% of HOLC records are retained after applying the first selection criterion. Together, these selection criteria (1) ensure adequate population coverage while allowing for variation in how HOLC records and census data deal with uninhabited sectors (e.g., parks, forested areas, industrial zones, waterways, etc.); and (2) maintain intra-city variation despite imperfect agreement between my two data sources. After applying both selection criteria, the final data set consists of 1,031 HOLC records across 17 cities. Cases provide strong regional (three Western, four Northeastern, and five each of Midwestern and Southern cities) and sociodemographic variation.¹

Analytic methods

Topic analysis

This paper begins by analyzing the thematic structure of HOLC area descriptions at scale. To do so, I deploy a network-based topic modeling approach using the *Sashimi* toolset (Breucker et al. 2016-10-28, 2016-10; Hannud Abdo et al. 2022), an emerging methodology that uses Bayesian Stochastic Block Modeling (SBM) (Peixoto 2014) to construct a nested, hierarchical analysis of discursive topics across large-scale text corpora.

Similar to standard topic modeling approaches (cf. Blei 2012), this approach begins by developing a mutually exclusive set of “topics,” or groups of words that tend to occur together in the same document. Topics provide an initial landscape of the latent themes that appear throughout a corpus. After defining an initial topic universe, then next step is to construct a nested hierarchy of topic clusters based on the co-occurrence of topics across documents. Topic clusters are constructed using an SBM approach.

The result is that topics can be distinguished individually by their constitutive terms, and can also be evaluated by their proximity to other topics across the hierarchy of topic clusters. This nested structure allows for a multi-scalar analysis of the thematic content of a corpus as well as the relations between individual topics at multiple levels of granularity and abstraction. Topic analysis thus provides a basis of understanding for the primary thematic elements that characterize the HOLC assessment process, while higher-order topic cluster open a pathway to understand how the concepts captured by individual topics converge to produce broader discourses.

¹Using to Logan et al.’s (2024) typology of early 20th century segregation trajectories, six of the selected had remained primarily white from 1880 to the 1930s; five reflected distinctly Southern trajectories; and three were “destination cities” for Black migrants. Logan et al. consider Akron, OH a “mixed northern” city. Spokane, WA is absent from the Logan et al. study.

I follow standard topic modeling practices and code topics according to the frequency of topic terms, while returning to the original texts to place topic content in context. I code topic clusters according to the distribution of constitutive topics alongside a contextual reading of text documents. Iterating between topic terms and a close analysis of the original texts contributes to building a substantive understanding of topics and topic clusters.

I focus on the full universe of available assessment documents that contain adequate text data ($N=6,749$) to construct topic and topic clusters, in accordance with the generalizing logic of the HOLC assessment program, which sought to evaluate neighborhoods in cities across the US according to a universalized set of assessment criteria.

Because SBM relies on Markov Chain Monte Carlo simulations to construct block clusters, I conducted five separate iterations and selected the model with the lowest entropy score. Entropy is calculated using a negative log likelihood approach. SMB hyperparameters were defined according to Table 1.

Table 1: SBM hyperparameters.

Parameter	Specification	Notes
N -grams	2	N -grams permitted, up to $N=2$
Stop words	Topic	Stop words retained in input text. SBM approach clusters stop words in a single topic that is then removed from analysis.
Minimum blocks	Free	N -blocks allowed to vary freely and inferred from input data (Hannud Abdo et al. 2022)
Maximum blocks	Free	N -blocks allowed to vary freely and inferred from input data (Hannud Abdo et al. 2022)
Block levels	Free	N -levels allowed to vary freely. Hierarchical structure inferred from input data.

Selecting definitive topics

Topics capture individual concepts or terms across the full corpus of assessment documents. Because AD documents provide a broad and rich corpus for analysis, individual topics represent relatively narrow conceptual elements. HOLC grading decisions proceeded by synthesizing the broad array of neighborhood characteristics, demographic information, and the judgement of individual assessors to assign one of four ordinal grades. Building a structured understanding of the relationship between HOLC discourses and mortgage risk assessments thus requires sifting through the broad landscape of topics to select the key conceptual elements that best predict grading outcomes.

To do so, I treat each topic as a document-level variable. For each document, I measure the presence of every topic as a binary indicator: A topic is coded as

“present” if it accounts for at least 10% of the terms within a given document. I use a binary rather than continuous measure to account for variation in document length across the corpus.

Using the full universe of narrative assessments (N=6,749), I employ a random forests approach through the *VSURF* R package (Genuer, Poggi, and Tuleau-Malot 2015) to select topics that are strongly associated with neighborhood grading decisions. Document-level topic indicators are used as a matrix of input variables to predict ordinal A-B-C-D grading outcomes. This approach outputs a vector of topics that, together, best predict variation in the grades assigned by HOLC assessors. *VSURF* hyperparameters are outlined in Table 2.

Table 2: VSURF hyperparameters.

Parameter	Specification	Notes
Thresholding		
<i>N</i> -forests	20	
<i>N</i> -trees	500	
Interpolation		
<i>N</i> -forests	10	
<i>N</i> -trees	100	
Prediction		
<i>N</i> -forests	10	
<i>N</i> -trees	100	

Relational analysis of topic content and neighborhood demographics

I use the topic variables identified through a random forests approach to analyze the relational interface between assessment discourses and racial geography in HOLC grading decisions. Focusing on intersected HOLC-census data (N=1,031), I estimate a series of models predicting grading decisions according to selected topics and constructed sociodemographic variables. Models are specified to (1) evaluate the relative contribution of narrative assessment content to HOLC grading decisions, alongside and in addition to neighborhood demographics; (2) estimate the effect of environmental discourses on the distribution of grading outcomes; and (3) understand complex interactions in racial geography and environmental discourses in the production of neighborhood assessment decisions.

Results

Lexical structure of HOLC documents

I begin by presenting results from the topic analysis of the HOLC AD corpus. The 6,578 neighborhoods with AD data provide substantial data for analysis, allowing for a rich description of the primary themes that animated HOLC

assessments. The domain-topic analysis returned 129 distinct Level-1 topics², nested within 15 Level-2 and 3 Level-3 topic clusters.

I begin by presenting results from the topic analysis of the HOLC AD corpus. The 6,578 neighborhoods with AD data provide substantial data for analysis, allowing for a rich description of the primary themes that animated HOLC assessments. The domain-topic analysis returned 129 distinct Level-1 topics³, nested within 15 Level-2 and 3 Level-3 topic clusters.

At the lowest level, Level-1 topics provide a highly detailed map of the lexical structure of HOLC area descriptions. Topics include references to specific features of the built environment (industry, railroads, businesses); social structure (race, ethnicity, immigration, hetero- or homogeneity); housing characteristics (apartments, houses, rentals, house size); assessments of physical condition (well maintained, blighted, occupied, vacant); financial terms (house value, sales price, mortgages, tax); neighborhood structure (lot size, convenience, terrain); city services (water, sewer); and sociospatial context (proximity to industry, settlement patterns, “encroachment” of undesired groups). Topics thus represent the primary conceptual elements that make up the specialized, technical, and targeted language of HOLC neighborhood assessments. The frequency of different topics, in turn, measures the relative importance of the various conceptual elements that shaped the HOLC assessment process.

Higher-order topic clusters are constructed from the co-occurrence of base Level-1 topics across the corpus. The content of higher-order clusters reveals the relations among topics across the corpus, while also linking the granular focus of individual topics to a comprehensive map of the lexical structure of HOLC assessments. Neither highly granular nor obtusely abstract, the 15 Level-2 topic clusters provide a useful measure of the primary concerns that animated HOLC assessors as they explained and justified neighborhood grading decisions. The 3 Level-3 clusters, in contract, reflect a broad categorical framework for how HOLC assessors organized the observations, descriptions, appraisals, and ideological judgements that constituted each neighborhood description.

Figure 1 displays the relative frequency of each Level-1 topic, with topics grouped into Level-2 and Level-3 clusters. At the most abstract level, I find that HOLC assessments are organized according to a three-part framework, with each element of that framework corresponding to a distinct spatial-analytic scale. The most frequent topic group is also the most local in scale: The quality of housing and its local spatial context (*"Housing Qualities and Context"*) represents the most significant area of assessment, totaling about 45% of the corpus as a whole. Neighborhood-level qualities and the structure of neighborhood built environments (*"Neighborhood Qualities"*), meanwhile, account for about 30% of the overall topic structure. The final 25% is devoted to neighborhoods'

²In addition to these 129 substantive topics, the model additionally generated 18 topics consisting of “stopwords” and other terms without substantive interest.

³In addition to these 129 substantive topics, the model additionally generated 18 topics consisting of “stopwords” and other terms without substantive interest.

broader sociospatial, financial, and physical context within the city as a whole (*"Neighborhood Context"*). This nested framework further suggests a fundamental spatial and scalar awareness animating the HOLC assessment processes, a point that I return to in the discussion.

Zooming in, the organization of Level-2 topic clusters reveals the primary elements composing this broader three-level structure. Assessments focused on housing issues are primarily concerned with three elements: The construction and location of housing (*"Construction and Location"*); the sociospatial context of housing units (*"Housing, Society, Space"*, listed as "Spatial" in Figure 1); and issues related to occupancy and ownership, deed restrictions and zoning, and access to stores and businesses (*"Occupancy and Business"*). Assessments centered on qualities of a neighborhood focused on issues related to industry, zoning, civil society institutions, and social structure (*"Industry, Zoning and Social Structure"*); neighborhood character (*"Neighborhood Character"*); impacts associated with industrial land uses, restrictions, race, and labor (*"Industry, Race, and Labor"*); and social character and property values (*"Social Character and Value"*). Finally, neighborhood context consists of qualitative assessments and neighborhood trends (*"Assessment"*); the racial and sociospatial context of neighborhood districts (*"Districts"*); the makeup of a its housing stock (*"House Type"*); positive and negative influences on neighborhood valuation (*"Influences"*); and taxes, utilities, and rental rates (*Tax*). Level-1 topics, in turn, represent the granular linguistic elements that compose each of these broad focus areas (see Appendix for full Level-1 topic information). For instance, the *"location_character"* topic, one of the 12 topics that make up the (*"Neighborhood Character"*) cluster, is primarily defined by "location," "character," and "design," alongside other less frequent terms.

Space and place

Topic analysis suggests that the HOLC assessment process was rooted in a model of housing and neighborhood valuation as a function of the sociospatial structure of cities at multiple scales. As outlined in the previous section, this begins with the lexical structure of HOLC assessments at the most abstract level. At the highest level of abstract, the content of assessment documents is structured according to three discursive trunks representing three distinct scalar concerns: housing-level characteristics, neighborhood-level characteristics, and the relationship between neighborhoods and the city. These initial findings support my overall argument that neighborhood assessment was organized according to a vision of urban development that stemmed from a deep concern with the relationship between social and spatial processes. Moving to a more granular analysis of Level-1 topics and Level-2 topic clusters reveals the multiscale nature of this essential spatiality.

At the most basic level, this is evident in the diversity and breadth of granular Level-1 topics focused on the spatial structure of residential populations and urban built environments. Level-1 topics focused on spatial structure are

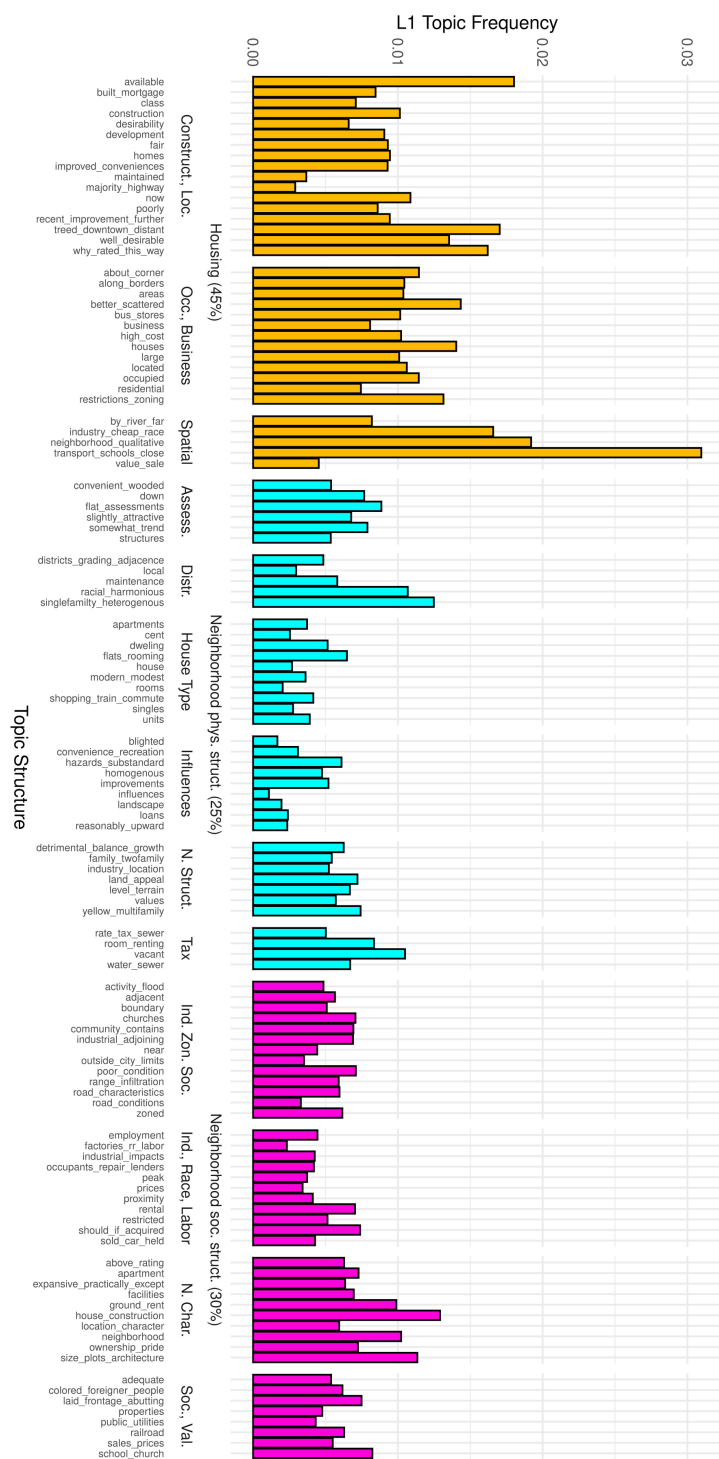


Figure 1: Three-level topic structure, with relative frequency of Level-1 topics displayed on the Y-axis.

distributed across all three primary thematic trunks, and form a key thematic element of many Level-2 topic clusters as well (see Table 3). This means that spatial structure and spatial relationships - location, distance, adjacency, density, boundaries, being within or outside, being surrounded by or separated from - form an essential connective tissue integrating the diverse range of elements that shaped HOLC assessment outcomes.

Table 3: Topics representing space and spatial relations.

Topic	Terms
Table in progress	

Turning our attention from the granular Level-1 topics to Level-2 topic clusters reveals how this essential spatial awareness operates within the thematic structure of the assessment corpus. While this pattern is evident across the topic structure, I focus on a Level-2 topic cluster focusing on the sociospatial context of urban housing (*"Housing, Society, Space"*) to develop this argument. *"Housing, Society, Space"* integrates topics on neighborhood spatial structure (*By_River_Far* and *Transport_Schools_Close*, listed in Table 3) with topics related to urban industry and race (*industry_cheap_race*), housing value (*value_sale*), and assessments of housing and neighborhood quality (*neighborhood_qualitative*). Here, spatial structure and spatial relations join assessments of housing value to the balance between spatial benefits - transportation and easy access to schools - and detriments, including environmental risks and what HOLC assessors termed “racial hazards.” Thus, not only is the content of assessment documents structured according this fundamental spatiality, but that sociospatial awareness is also an essential ingredient to how HOLC actors assemble diverse elements of concern into comprehensive assessment documents. Thus, a topic analysis suggest a sociospatial imaginary predicated on the granular measurement of local difference, as well as a deep concern with the relationship between local difference and interlocal context.

HOLC content and neighborhood valuation

Next, I investigate the relationship between the lexical structure of HOLC texts and the grading decisions of HOLC assessors.

Selecting definitive topics

Random forests analysis using the VSURF package in R identifies 24 out of 147 topics that together best predict HOLC grading decisions (see Table 4).

Table 4: Random Forests results for definitive topics.

Topic	Key Terms
value_sale	for (4714), value (376), sale (278), block (154), erected (86), dilapidated (65), accessibility (53), access (53), occasional (50), clearance (36), length (22)
industry_cheap_race	industry (650), cheap (576), negroes (289), plant (274), yards (260), company (235), workers (199), extreme (198), shacks (197), odors (144), undeveloped (137), railroads (137), oldest (136), manufacturing (131), bad (107)
neighborhood_qualitative	good (3723), none (661), spotty (170), clean (130), elevation (71), quiet (47)
construction	construction (2330), uniform (261), retarded (30)
development	development (1483), developed (763), developing (97)
homes	homes (2444)
areas	areas (732), lower (507), within (507), group (180), walking (150), tendency (115), portions (113), conditions (101), landscaped (100), individual (72), winding (47), beautifully (34), closein (18), adjustment (8)
restrictions_zoning	restrictions (908), zoning (612), while (469), upon (225), multiple (144), charm (112), extremely (97), basis (95), percent (92), subdivisions (78), ranges (74)
high_cost	high (2133), cost (199), make (183), highest (128)
better_scattered_mixed	better (1193), scattered (564), mixed (532), throughout (442), around (253), priced (249), work (148), bungalow (137), isolated (91), field (62), dead (30), flush (8)
somewhat_trend	somewhat (732), trend (592), slowly (195), downward (188), decline (163), static (96), reasonable (54), command (22)

Topic	Key Terms
convenient_wooded	convenient (1018), wooded (208), poorer (144), potential (22)
slightly_attractive	slightly (395), attractive (389), those (358), even (121), smaller (108), latter (95), stability (69), slopes (55), like (50), sufficient (47), anything (39), sharp (26)
homogenous	homogeneous (309), financing (208), against (169), provide (98), sites (90), enforced (53), permit (46), limit (45)
house_construction	brick (613), frame (586), buildings (524), bungalows (405), frames (104), stone (93), stucco (73), inferior (54), predominately (49),
road_conditions	paved (729), remainder (108), graveled (16)
community_contains	community (804), contains (317), club (222), golf (219), completed (76), opened (64), citizens (53), wealthiest (22), signal (9), mccallie (5)
industrial_adjoining	industrial (785), plants (334), adjoining (308), commercial (248), packing (46), lowcost (16), textile (13), equipped (13), lcoated (7), noyes (7), stevens (7)
poor_condition	poor (1249), condition (486), poorest (70), westward (30)
colored_foreigners_people	people (486), colored (371), foreigners (67), exists (65), paid (42), stock (41), width (38), whites (32), buffer (28)
restricted	restricted (728), accessible (142), fire (117), playgrounds (107), moderately (95), police (48), winter (46), jerrybuilt (19), tourists (10), deterring (9), poorness (5), obsoleacence (3)
employment	employment (504), rather (400), concentration (158), groups (88)
negro	negro (690)

Assessing predictive power of topics

I now turn to linked HOLC-census data to assess the predictive power of key topics alongside and in addition to neighborhood demographics. My intention in this section is to establish that the discursive content of HOLC documents

shaped assessment decisions in ways that were not fungible with the racial geography of early 20th century cities. To do so, I construct three nested models predicting poor HOLC decisions - measured as an ordinal A-B-C-D outcome variable - using (1) Black and foreign-born population interpolated to HOLC boundaries; (2) key topics derived from the random forests analysis outlined above; and (3) demographic data and topic information together. As Table 5 demonstrates, introducing topic content alongside sociodemographic information improves model fit to a significant degree.

Table 5: Poor HOLC grade by race, nativity, and HOLC topics.

Variable	M1	M2	M3
<i>Levels</i>			
A B	-2.89***	-3.27***	-4.79***
B C	-1.13**	-1.07***	-2.48***
C D	0.99**	1.69***	0.55
<i>Population</i>			
Black Pop.	0.57***		0.5***
Foreign-born Pop.	0.72***		0.84***
Total Pop.	-0.67***		-0.84***
<i>Topics</i>			
value_sale		0.36**	0.26.
industry_cheap_race		1.2***	1.23***
neighborhood_qualitative		-0.5***	-0.51***
why_rated_this_way		-0.16	-0.4**
construction		-0.31*	-0.19
development		-0.63***	-0.6***
homes		-0.41**	-0.42**
areas		-0.36**	-0.41**
restrictions_zoning		0.25	0.26
high_cost		-0.74***	-0.73***
better_scattered_mixed		0.31*	0.13
somewhat_trend		-0.17	-0.32*
convenient_wooded		-0.19	-0.11
slightly_attractive		-0.07	-0.14
homogenous		-0.6**	-0.76***
house_construction		0.15	-0.01
road_conditions		-0.21	-0.2
community_contains		-0.77***	-0.6***
industrial_adjoining		0.72***	0.74***
poor_condition		1.1***	1***
colored_foreigner_people		0.62***	0.59***
restricted		-1.08***	-1.04***
employment		0.21	0.16
negro		2.13***	1.65***

Variable	M1	M2	M3
AIC	2412.48	2074.05	1966.63

Relational interfaces

A final set of models focuses on the relational interface between neighborhood composition and key discourses of race, place, environment, and value, in order to understand the interaction between race and environment in the production and formalization of early 20th century territorial stigmatization. To do so, I interact topic variables with neighborhood-level Black population estimates, in order to understand how the relationship between the narrative text of HOLC documents and assessment decisions vary according to neighborhood demographics. Population data are logged.

Key findings:

Topics indicating hazardous industrial risks predict poor HOLC ratings for neighborhoods with at least some Black presence; this effect is absent for non-Black neighborhoods, suggesting that the effect of industrial risk on HOLC decisions was tied in part to its local intersection with ideologies of racial threat. The effect of this topic diminishes as Black population size increases (see Figure 2).

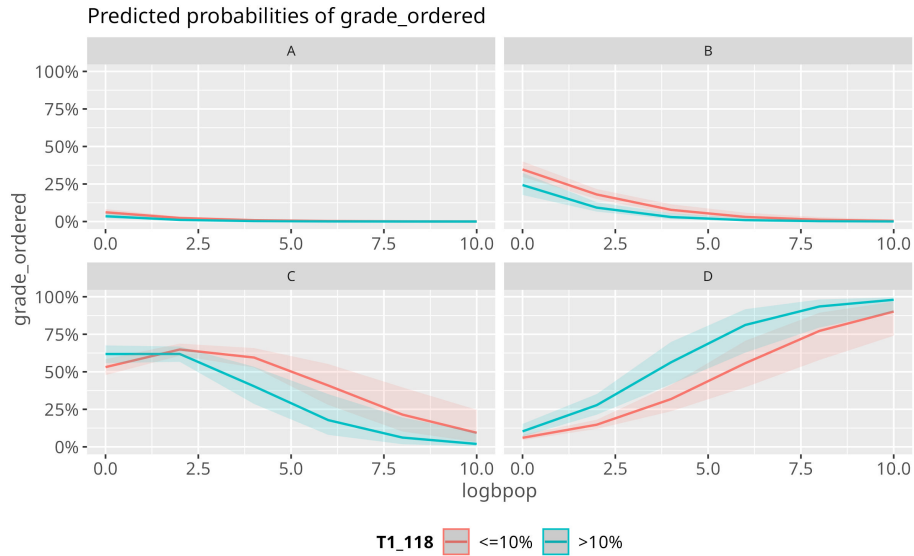


Figure 2: Effect of “industrial_adjoining” topic on HOLC assessments by Black populations size.

Topics linking industrial manufacturing with racialized discourses predict poor HOLC ratings for nearly all values of the Black population variable, until the far right tail of its distribution (see Figure 3).

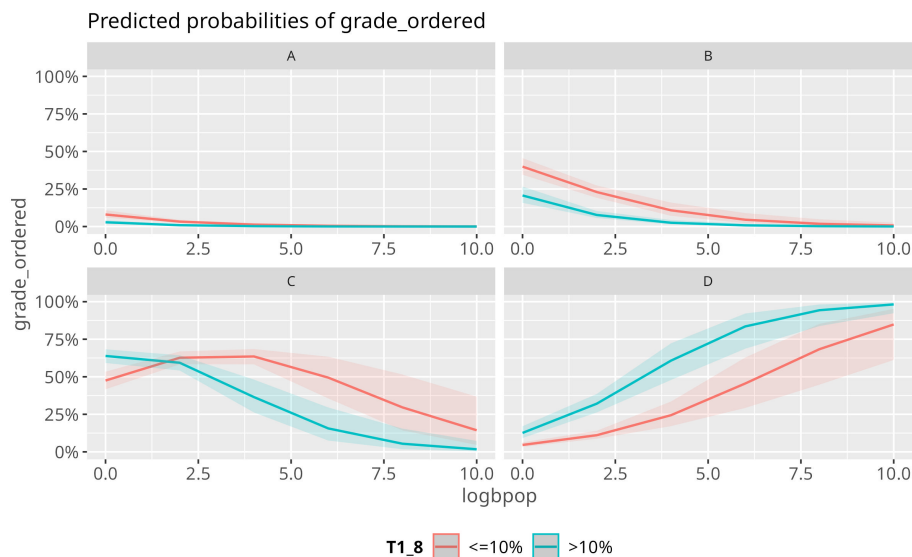


Figure 3: Effect of “industry_cheap_race” topic on HOLC assessments by Black populations size.

Topics indicating poor neighborhood conditions predict poor HOLC ratings for neighborhoods with at least some Black presence; this effect is absent for non-Black neighborhoods, suggesting that the effect of poor housing conditions on HOLC decisions was tied in part to its local intersection with ideologies of racial threat. This effect is maintained as Black population size increases, until well into the third quartile of the Black population variable (see Figure 4).

Topics indicating neighborhood restrictions predict high HOLC ratings; this effect diminishes as Black population size increases, until no significant difference is found for neighborhoods with more than about 40 Black residents (see Figure 5).

Discussion

To come.

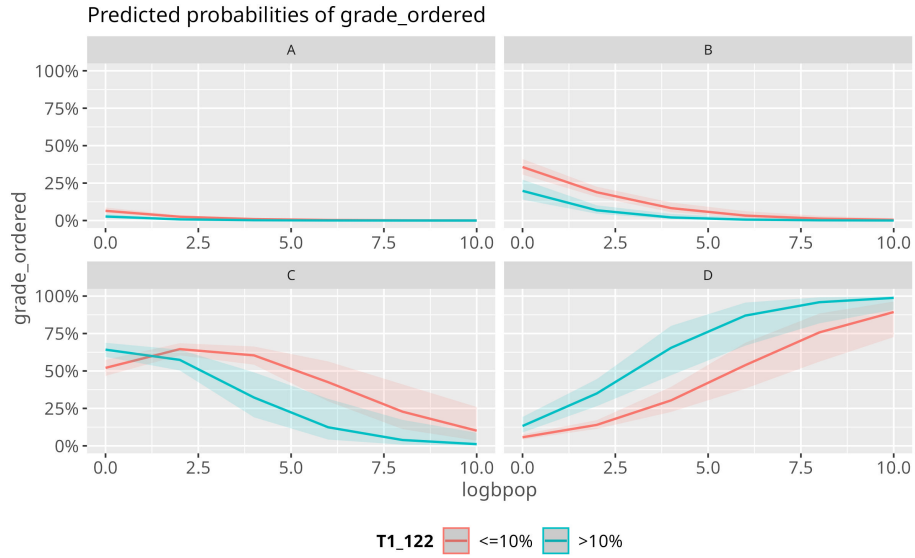


Figure 4: Effect of “poor_condition” topic on HOLC assessments by Black populations size.

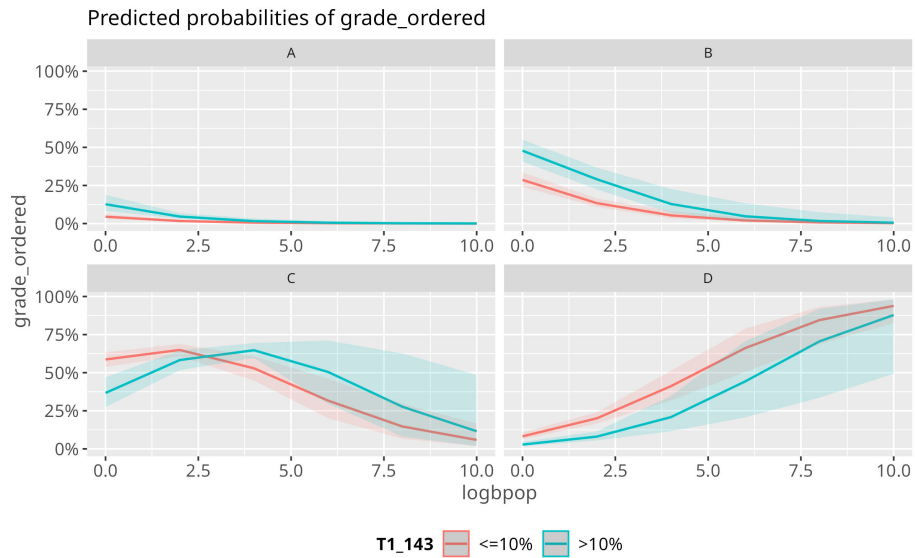


Figure 5: Effect of “restricted” topic on HOLC assessments by Black populations size.

References

- Bhardwaj, Ankit. 2023. “The Soils of Black Folk: W.E.B. Du Bois’s Theories of Environmental Racialization.” *Sociological Theory* 41(2):105–28.
- Blei, David M. 2012. “Probabilistic Topic Models.” *Communications of the ACM* 55(4):77–84.
- Breucker, Philippe, Jean-Philippe Cointet, Alexandre Hannud Abdo, Guillaume Orsal, Constance de Quatrebarbes, Tam-Kien Duong, Cristian Martinez, Juan Pablo Ospina Delgado, Luis Daniel Medina Zuluaga, Diego Fernando Gómez Peña, Tatiana Andrea Sánchez Castaño, Joenio Marques da Costa, Hajar Laglil, Lionel Villard, and Marc Barbier. 2016-10-28, 2016-10. “CorTexT Manager.”
- Candipan, Jennifer, and Jonathan Tollefson. 2024. “Machine Learning and Large-Scale Data for Understanding Urban Inequality.” in *The Oxford Handbook of the Sociology of Machine Learning*, edited by C. Borch and J. P. Pardo-Guerra. Oxford, UK: Oxford University Press.
- Frickel, Scott, and James R. Elliott. 2018. *Sites Unseen: Uncovering Hidden Hazards in American Cities*. New York: Russell Sage Foundation.
- Genuer, Robin, Jean-Michel Poggi, and Christine Tuleau-Malot. 2015. “VSURF: An R Package for Variable Selection Using Random Forests.” *The R Journal* 7(2):19–33.
- Hannud Abdo, Alexandre, Jean-Philippe Cointet, Pascale Bourret, and Alberto Cambrosio. 2022. “Domain-Topic Models with Chained Dimensions: Charting an Emergent Domain of a Major Oncology Conference.” *Journal of the Association for Information Science and Technology* 73(7):992–1011.
- Logan, John R., Elisabeta Minca, Benjamin Bellman, Amory Kisch, and H. Jacob Carlson. 2024. “From Side Street to Ghetto: Understanding the Rising Levels and Changing Spatial Pattern of Segregation, 1900–1940.” *City & Community* 23(2):155–79.
- Logan, John R., and Weiwei Zhang. 2018. “Developing GIS Maps for US Cities in 1930 and 1940.” in *The Routledge Companion to Spatial History*. Routledge.
- Mohai, Paul, David Pellow, and J. Timmons Roberts. 2009. “Environmental Justice.” *Annual Review of Environment and Resources* 34(1):405–30.
- Nelson, Robert K., LaDale Winling, Richard Marciano, and Nathan Connolly. 2023. “Mapping Inequality.” in *American Panorama*, edited by R. K. Nelson and E. L. Ayers.
- Peixoto, Tiago P. 2014. “Hierarchical Block Structures and High-Resolution Model Selection in Large Networks.” *Physical Review X* 4(1):011047.

- Tollefson, Jonathan. 2026. "Environment and the Racialization of Space in US Cities." *Social Forces* soag001.
- Wacquant, Loïc, Tom Slater, and Virgílio Borges Pereira. 2014. "Territorial Stigmatization in Action." *Environment and Planning A: Economy and Space* 46(6):1270–80.