CESifo WORKING PAPERS

11098
2024

April 2024

Mapping Out Institutional Discrimination: The Economic Effects of Federal “Redlining”

Disa M. Hynsjö, Luca Perdoni

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 11098*

Mapping Out Institutional Discrimination: The Economic Effects of Federal "Redlining"

Abstract

This paper proposes a novel empirical strategy to estimate the causal effects of federal "redlining" – the mapping and grading of US neighborhoods by the Home Owners' Loan Corporation (HOLC). In the late 1930s, a federal agency created color-coded maps to summarize the financial risk of granting mortgages in different neighborhoods, together with forms describing the presence of racial and ethnic minorities as "detrimental". Our analysis exploits an exogenous population cutoff: only cities above 40,000 residents were mapped. We employ a difference-in-differences design, comparing areas that received a particular grade with neighborhoods that would have received the same grade if their city had been mapped. The control neighborhoods are defined using a machine learning algorithm trained to draw HOLC-like maps using newly geocoded full-count census records. Our findings support the view that HOLC maps further concentrated economic disadvantage. For the year 1940, we find a substantial reduction in property values and a moderate increase in the share of African American residents in areas with the lowest grade. Such negative effects on property values persisted until the early 1980s. The magnitude of the results is higher in historically African American neighborhoods. The empirical results show that a government-supplied, data-driven information tool can coordinate exclusionary practices and amplify their consequences.

JEL-Codes: J150, R230, N920, N320.

Keywords: Redlining, neighborhoods, discrimination, machine-learning.

*Disa M. Hynsjö**
Department of Economics
Yale University, New Haven / CT / USA

Luca Perdoni
ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Munich / Germany
perdoni@ifo.de

* Disa passed away in Summer 2021. This paper is founded on her vision, brilliance and courage.

April 25, 2024
I am indebted to my Ph.D. committee members, Joseph Altonji, John Eric Humphries and Cormac O'Dea for their guidance, support and encouragement. Chelsea Ilarde and Alessandro Di Filippo provided excellent research assistance. For detailed comments and suggestions, I also thank Julian Aramburu, Jaime Arellano-Bover, Paula Calvo, Eric Chyn, Alberto Ciancio, Rebecca Diamond, Jose Antonio Espin-Sanchez, Paul Goldsmith-Pinkham, Sun Kyoung Lee, Trevon Logan, Costas Meghir, Helmut Rainer, Sarah Raviola, Jaehee Song, Ebonya Washington, and Seth Zimmerman. Conference participants at the NBER Summer Institutes on Urban Economics and the Development of the American Economy, the CESifo Area Conference on Labor Economics and the IEB Workshop on Urban Economics have provided precious feedback. The technical support of Barbara Esty and Miriam Olivares from Yale University Library, as well as Dan Feenberg and Mohan Ramanujan at NBER, has been fundamental. The previous version of the paper was circulated under the title "The Effects of Federal Redlining Maps: A New Empirical Strategy". This research has been supported by the Washington Center for Equitable Growth and the Cowles Foundation Program in Labor and Public Economics.

1. Introduction

For the past hundred years, the wealth gap between Black and white Americans has shown few signs of convergence and today remains the largest among racial economic differences (Derenoncourt et al., 2024). Since housing is a fundamental asset for most American families, mortgage lending is critical in determining this gap (Baradaran, 2017). Information tools have been employed since the first half of the 20th century by financial institutions and public administrations to evaluate credit applications. Did a novel data-driven tool developed by the federal government affect American neighborhoods, foreshadowing algorithmic decision-making in the housing market? Between 1935 and 1939, a federal agency – the Home Owners' Loan Corporation (HOLC) – drew *Residential Security Maps* for more than 200 US cities to summarize the financial risk of granting loans in different neighborhoods. Color-coded maps assigned each neighborhood one of four security grades, from A (green) to D (red).[1] Standardized forms attached to the maps (*Area Descriptions*) consistently described the presence of African Americans, Jews, and European immigrants as detrimental to a neighborhood's grade. In the late 1970s, urban historian K. T. Jackson rediscovered the maps at the National Archives and proposed them as an example of residential redlining, the systematic denial of mortgages to residents of a community (Jackson, 1980). Since then, the view that the HOLC maps were a source of systemic discrimination has steadily gained popularity (Coates, 2014; Rothstein, 2017). In this paper, we propose a new strategy to measure the causal effects of a government initiative proposed as a symbol of institutional discrimination by journalists, activists, scholars, and presidential candidates.[2]

This paper estimates the short and long-term causal effects of the Home Owners' Loan Corporation maps using a new empirical strategy. The outcomes of interest are homeownership rates, property values, rent prices, and shares of African American residents in 1940. We also analyze the evolution of these outcomes between 1960 and 2010. Our approach exploits an exogenous population threshold: only cities above 40,000 residents were mapped. We use a machine learning classification algorithm to draw residential security maps in control cities with a population below the 40,000-resident threshold. Using the grades predicted by the classification model, we apply a grouped difference-in-differences design to measure the effects of the HOLC maps. The causal effects are identified by the differences between neighborhoods in treated cities and areas that would have received the same security grade if their city had been mapped.

Our empirical strategy is feasible thanks to a spatial dataset we constructed using the 1910-1940 full-count census records (Ruggles et al., 2020), National Historical Geographic Information System (NHGIS) data (Manson et al., 2021), and CoreLogic property deeds. We clean household addresses for each census decade following best practices from the urban history literature (Logan and Zhang, 2018). Detailed geographic

1. As an example, a scan of the HOLC map for New Haven, CT, is available in Figure 1.

2. See Rothstein (2017) and Coates (2014). Historical government support of redlining practices has been proposed by President J. Biden and Senator E. Warren as a motivation for their housing plans.

HOLC Residential Security Map of New Haven, CT

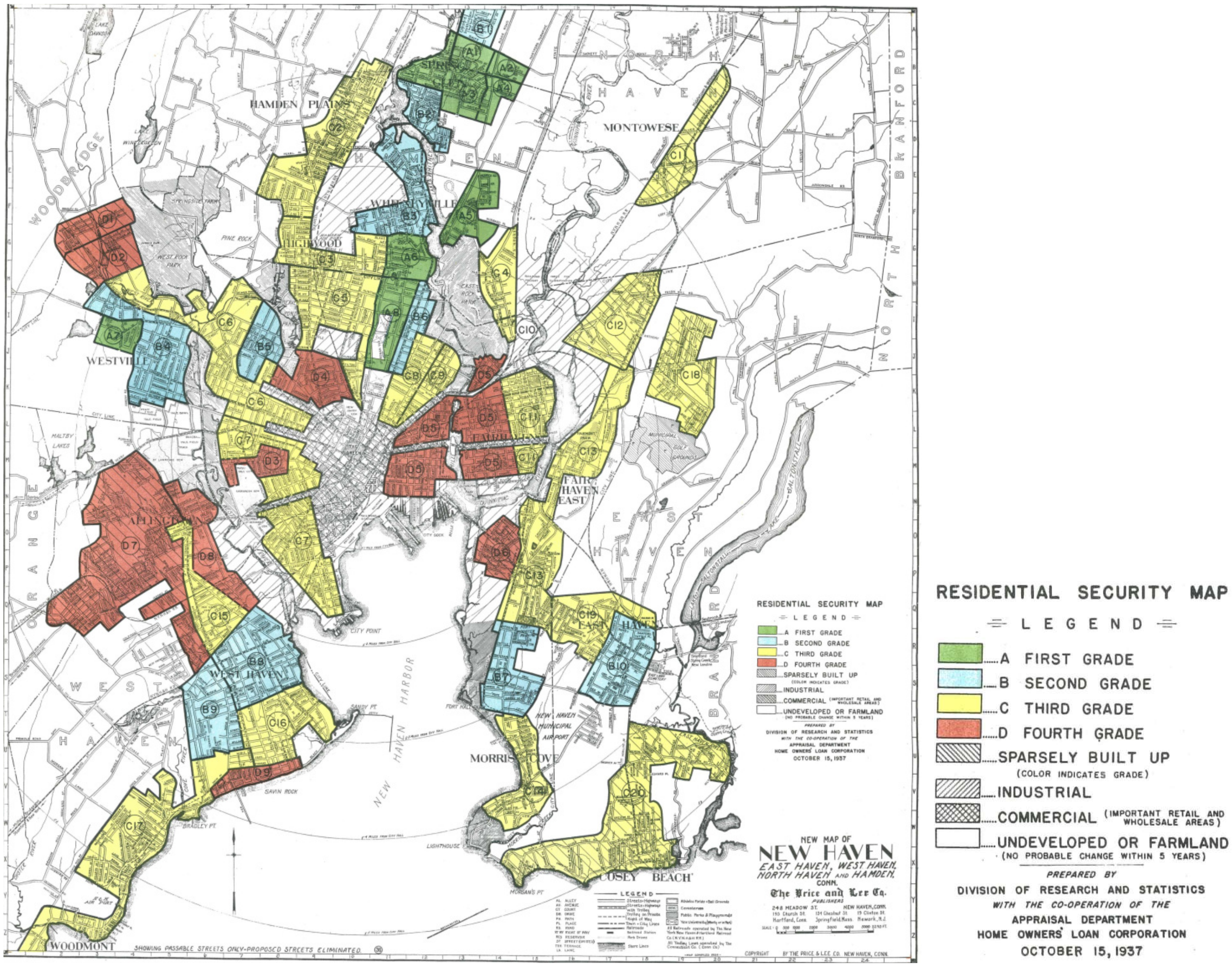


Figure 1 — The scan of the original Residential Security Map of New Haven, CT, has been provided by *Mapping Inequality* (Nelson et al., 2023).

coordinates are then assigned to census observations using a state-of-the-art locator. Georeferenced data allow us to match census records with digitized HOLC maps (Nelson et al., 2023) and alternative sources of information to expand our dataset to the years beyond 1940. We use sociodemographic information from the National Historical Geographic Information System (NHGIS) along with property transaction prices from the CoreLogic deed database to estimate the long-term causal effects of the maps. Our final dataset covers major US urban areas across more than 300 counties between 1910 and 2010.

The effects in red neighborhoods support the view that HOLC maps further concentrated economic disadvantage and led to a weaker housing market. In 1940, shortly after the introduction of the maps, we find a reduction in property prices and homeownership rates in D (red) areas, along with a moderate increase in the percentage of African Americans living in those neighborhoods. Property value reductions are also detected in C (yellow) areas.[3] The negative effects on property prices in yellow and red zones persisted until the early 1980s, shortly after legislative measures were introduced to improve access to credit.[4] We also construct an individual panel dataset thanks to the Census Linking Project crosswalks (Abramitzky et al., 2021), and we find positive effects of the D grade on residential moves for whites but none for Blacks, together with higher barriers to homeownership for African American movers. Taking advantage of digitized *Area Description* forms (Markley, 2023), we show that historically African American neighborhoods drive our results and that the maps reinforced pre-existing urban patterns for race and social class.

The credibility of our approach relies on the performance of the machine learning classification model. To assess its precision, we build a test dataset randomly excluding 25% of neighborhoods from the algorithm's training procedure. We then compare observed and predicted grades in the test dataset. Our trained random forest algorithm assigns the correct grade to more than 90% test neighborhoods, and its predicted maps are convincing replicas of those made by HOLC.

Our paper contributes to a growing literature in economics studying the consequences of HOLC maps.[5] This paper is most closely related to the work of Aaronson, Hartley and Mazumder (2021b), who use a border regression discontinuity design to measure the local effects of lower HOLC grades. Unlike Aaronson et al. (2021b), our estimation method compares similar neighborhoods in mapped and unmapped cities, avoiding endogeneity concerns due to differential pre-trends in socioeconomic variables on different sides of HOLC borders.[6] In particular, the counterfactual is made of similar neighborhoods not mapped by HOLC rather than nearby areas with a higher evaluation. Moreover, we are the first to propose a predictive model replicating the HOLC maps. Overall, the results obtained with our new empirical approach are a middle-ground between "*HOLC Culpablism*" and "*HOLC Scepticism*" (Markley, 2024), the two sides of the current debate about the legacy of the "redlining" maps. According to our results, the maps reinforced several dimensions of economic disadvantage, but the effects were concentrated only in specific neighborhoods and cannot be detected in present-day American cities.

Our empirical strategy will measure neither the consequences of maps' biases nor the impact of ex novo discrimination.[7] Instead, we will capture the effects of a tool speeding up the evaluations of mortgage

3. This result is consistent with a finding in Aaronson et al. (2021b) labelled by the authors as "yellow-lining".

4. The Equal Credit Opportunity Act (1974), the Home Mortgage Disclosure Act (1975), and the Community Reinvestment Act (1977) had the common goal of increasing access to mortgages in neighborhoods previously ignored by financial institutions.

5. All the recent papers on this topic either focus on different questions or employ different empirical approaches, datasets, and sample definitions. In addition to Aaronson et al. (2021b), see Fishback et al. (2023), Fishback et al. (2011), Faber (2020), Aaronson et al. (2021a), Fishback et al. (2022), and Anders (2023). There are also working papers on this topic using spatial regression discontinuity designs: see Appel and Nickerson (2016), and Krimmel (2020).

6. Aaronson et al. (2021b) and Fishback et al. (2023) document the differential pre-trends due to the non-random location of borders and non-random assignment of grades. Aaronson et al. (2021b) employ propensity scores and a subset of idiosyncratic borders to address endogeneity concerns.

7. Simplistic discussions of the maps often suggest that the HOLC single-handedly invented housing market discrimination, assigning arbitrarily low grades to minority neighborhoods. To back up such causal claims, we would need an unfeasible set of control cities with color-blind mortgage lending. Interpreting the effects of the HOLC maps as the direct impact of racial discrimination would be misleading and historically unsound.

applications. Such an instrument was created by a federal agency accepting and systematizing pre-existing discriminatory practices, leading to an unprecedented nationwide standardization of neighborhood assessment techniques. This is why sociologists often argue that the HOLC *institutionalized* discrimination (Small and Pager, 2020).[8] So far, economists have mainly focused on individuals who discriminate based on their taste (Becker, 1971) because of imperfect information[9] or implicit bias (Bertrand et al., 2005; Bertrand and Duflo, 2017). These different mechanisms originate from individual choices and cannot be readily applied to settings where something other than an individual discriminates. This paper provides an empirical analysis focused on institutional discrimination complementing existing research in economics on the role of discrimination in the US housing market.

The paper also contributes to the literature investigating the causes of segregation and urban inequality (Cutler et al., 1999; Glaeser and Vigdor, 2012; Logan and Parman, 2017). As outlined in Boustan (2012), residential segregation can result from individual choices by white households, often referred to as *White Flight* (Boustan, 2010, 2017), Black self-segregation,[10] or collective action.[11] Our findings give an example of the last of these causes since the federal agency's practices could strengthen residential exclusion. Moreover, if a government initiative reinforces neighborhood disadvantage, it can hamper intergenerational mobility because of childhood exposure effects and transform localized effects into long-term economic inequality (Chyn and Katz, 2021).[12]

Another relevant dimension of the HOLC initiative was its technological content. The agency undertook an unprecedented data collection effort, creating a data analytics tool at the forefront of real estate appraisal techniques of the time.[13] HOLC maps can then be interpreted as an innovation in statistical technology that led to increased automation in the processing of mortgage applications. Our results offer a cautionary tale of how institutional practices can coordinate individual choices and speak to today's concerns about algorithmic decision-making (Rambachan et al., 2020; Ludwig and Mullainathan, 2021; Fuster et al., 2022). In terms of methods, we contribute to a relatively recent body of literature using machine learning algorithms to build control groups in observational studies.[14]

8. See Small and Pager (2020) and Lang and Kahn-Lang Spitzer (2020) for a comparison of perspectives on discrimination between sociology and economics.

9. See Fang and Moro (2011) for a review of research on statistical discrimination originated by Phelps (1972) and Arrow (1973).

10. This possible source of segregation does not find strong empirical support. See Krysan and Farley (2002) and Ihlanfeldt and Scafidi (2002).

11. Collective action as a source of segregation can take many forms. Some examples are racial covenants (Jones-Correa, 2000; Sood et al., 2019), urban renewal programs (Collins and Shester, 2013), and public housing programs (Chyn, 2018; Almagro et al., 2023).

12. The importance of childhood exposure effects as a primary driver of neighborhood effects has been recently shown in several papers. See Chetty et al. (2016); Chetty and Hendren (2018), and Chyn (2018).

13. In particular, we provide an empirical analysis of the consequences of a collection of federal maps. See Nagaraj and Stern (2020) for a review of recent work on the Economics of Maps.

14. An example is Liberman et al. (2018). See Mullainathan and Spiess (2017) and Athey and Imbens (2019) for a review of machine learning algorithms within the econometric toolbox.

The paper proceeds as follows. Section 2 provides additional details about HOLC activities and the circulation of its maps, while Section 3 describes our novel dataset. We outline our empirical strategy in Section 4, along with results about the performance of our classification algorithm and an array of validity checks. The estimated effects of HOLC maps can be found in Section 5. Section 6 concludes.

2. Historical Background

In the aftermath of the Great Depression, the Roosevelt Administration developed several programs to tackle a mortgage crisis characterized by soaring default rates and falling property values. The Home Owners' Loan Corporation (HOLC) was created in 1933 to aid homeowners "*in hard straits largely through no fault of their own*" (Federal Home Loan Bank Board, 1937). Under the Federal Home Loan Bank Board (FHLBB) direction, the HOLC's first task was to refinance distressed mortgages with longer terms, lower interest rates, and higher loan-to-value ratios. In particular, the HOLC granted fully amortized loans with 15-year minimum terms at 5% interest rate, financing up to 80% of the property value.[15] The agency concluded its $3 billion lending effort in 1936 after refinancing over one million loans and holding approximately 10% of US non-farm mortgages (Jackson, 1980).

As a consequence of their lending program, HOLC gained considerable exposure to the housing market. Government officials believed that a healthier lending industry was necessary to safeguard the value of federal real estate investments (Hillier, 2005). In particular, they considered the standardization of appraisal techniques critical to achieving price stability (Woods, 2012). For this reason, the FHLBB directed HOLC to systematically evaluate US neighborhoods, following a growing interest in ecological models across the real estate industry.[16] In 1936, HOLC started the *City Surveys* program, producing maps (*Residential Security Maps*) and standardized forms (*Area Descriptions*) for 239 major U.S. cities. The initiative was completed by 1939.

Field agents drew HOLC maps based on published reports, public records, federal maps, and detailed surveys of local financial institutions (Michney, 2022). The availability of these sources varied between cities, and HOLC agents relied on their networks in the real estate community to supply any missing information. The result was meant to be "*a composite opinion of competent realtors engaged in residential brokerage, good mortgage lenders, and the HOLC appraisal staff.*"[17] HOLC agents traced boundaries to divide residential areas into homogeneous neighborhoods. They then assigned a grade on a four-level scale to summarize

15. These terms were much more convenient to homeowners than the 5-year interest-only loans, with interest up to 7%, that were prevalent in the market up to that time. See Fishback et al. (2011).

16. See Jackson (1980) and Light (2010) for a discussion about how ecological models, newly developed by the Chicago School of Sociology, became an influential theory for real estate appraisal.

17. Corwin A. Fergus to T. L. Williamson, October 2, 1936, Roll 431, Home Owners Loan Corporation, microfilm copies of General Administrative Correspondence 1933-36, National Archives II (College Park, MD). As cited in Michney (2022).

the financial security of real estate investment in each zone.[18] Areas colored in green (grade A) were the first-tier neighborhoods, while blue neighborhoods (grade B) were deemed still suitable. The color yellow (grade C) highlighted neighborhoods becoming obsolete or at risk of "*infiltration of a lower grade population*" (Hillier, 2005). Red neighborhoods (grade D) were considered "*hazardous*" (Hillier, 2003) for investment. The agency also produced each neighborhood's detailed *Area Descriptions*. In these forms, they described housing conditions, local amenities, and demographic composition.[19] The presence of African Americans, Jews, and certain European immigrants was always characterized as a "*detrimental influence*" that "*infiltrated*" the American social fabric with ominous effects on local housing markets. While the inclusion of racial and ethnic hierarchies in real estate appraisal was pervasive at that time, HOLC practices standardized these notions at an "*unprecedented scale*" (Jackson, 1980) with the coordinated effort of more than 20,000 employees distributed across more than 200 local offices and the stamp of "*federal endorsement*" (Hillier, 2003).

HOLC could not have used the results of the *City Surveys* in its lending decisions since the maps were created after the agency completed its refinancing effort. Therefore, the economic impact of the maps depends on how widely these documents circulated among other federal agencies and private financial institutions. The literature offers diverging views on this topic. Hillier (2003) reports that the FHLBB intended to restrict access to "*agencies within the FHLB*" and "*such government agencies having interests allied with those of the Board*" while no copies were granted to "*private interests*". However, the author concedes that the maps were in strong demand among the public and that local consultants employed by the HOLC had access to these documents. An opposite stance, first proposed by Jackson (1980) and more popular today, argues that HOLC's findings were widely distributed and quickly became a benchmark for real estate appraisal in government agencies and the private sector.

Even if we lack definitive evidence about the circulation of *City Surveys*, there is proof that another federal agency – the Federal Housing Administration (FHA) – received multiple copies of HOLC maps. The main task of FHA was to evaluate applications to its mortgage insurance program. To this end, the FHA employed a manual that openly described racial and ethnic segregation as a guarantee of housing market stability (Federal Housing Administration, 1936) and a collection of maps that categorized neighborhoods on a four-level scale according to their financial security (Aaronson et al., 2021b). At the same time, Fishback et al. (2023) argue that "*the FHA developed its own research using more precise block-level data from property inventories in hundreds of cities*", dismissing the use of HOLC maps by the FHA. Since we are not aware of historical evidence about different FHA practices or data sources following the 40,000-resident threshold, our identification strategy is meant to capture the effects of this specific HOLC information tool while controlling for other housing market institutions.[20] Notwithstanding many uncertainties about the maps' dissemination,

18. A scan of the HOLC map drawn in 1937 for New Haven, CT, can be found in Figure 1.

19. As an example, a scan of the HOLC *Area Description* for New Haven D-4 neighborhood is available in Appendix Figure A1.

20. The cities covered by FHA data tend to be larger and mostly do not overlap with the smaller ones included in our Diff-in-Diff sample. The historical evidence provided by Fishback et al. (2023) can then be interpreted as a caveat about the external

it is undeniable that HOLC *City Surveys* are valuable *"Windows into the governing racial–spatial ideology of real estate capital on the eve of the US's most dramatic expansion of homeownership to date"* (Markley, 2024).

The explicit inclusion of racial or ethnic criteria in real estate financing became illegal in 1968 with the introduction of the Fair Housing Act. A further series of federal laws enacted in the 1970s addressed concerns about the lasting effects of financial exclusion. In particular, the Equal Credit Opportunity Act (1974), the Home Mortgage Disclosure Act (1975), and the Community Reinvestment Act (1977) were meant to counteract redlining by reinforcing anti-discrimination legislation, introducing mortgage disclosure requirements, and supervising credit supply at the local level.

3. Data

We construct a new dataset combining three sources: digitized HOLC maps, census data, and CoreLogic deed records. Our classification algorithm is trained on 1930 census data merged with HOLC maps. The short-term effects of the HOLC maps are measured with 1930 and 1940 census data, while the long-term effects are estimated by combining full-count census data with CoreLogic and tract-level NHGIS data for the decades between 1960 and 2010. In the following sections, we provide additional details about each data source.

3.1 HOLC Residential Security Maps

We incorporate HOLC grades in our project using the digitized maps provided by the Digital Scholarship Lab at the University of Richmond (Nelson et al., 2023). The files contain maps for 216 cities in 39 states.[21] We convert neighborhood shapes originally traced by HOLC into a regular grid of hexagons. Hexagons will be our spatial unit of observation, and their use simplifies the construction of HOLC maps in control cities.[22] The area of one hexagon approximates the typical size of a block in US grid plan cities such as New York City or Chicago.[23] We assign a grade to a hexagon if one color occupies at least 75% of its surface.[24] This spatial transformation has a negligible impact on the overall distribution of the grades. Appendix Table A1 shows the proportions of each grade according to different spatial definitions, while Appendix Figure A2 compares the digitized version of the HOLC map of New Haven, CT, with its hexagon-level counterpart. While our hexagons have a fixed area, the HOLC neighborhoods do not, which explains the minor discrepancies in

validity of our results in bigger American cities.

21. We use the last updated version of the digitized maps released in December 2023. Compared to the previous versions used in the recent literature, this update adds twelve HOLC City Survey Maps (Nelson et al., 2023).

22. More details on why we choose to use a grid of hexagons can be found in Section 4.1.

23. The grid is made of regular hexagons with an area of 0.025 squared kilometers (7.3 acres) and a side of approximately 100 meters (328 feet).

24. The results are robust to modest variations in the 75% threshold. Given the small dimension of each spatial unit, the vast majority of hexagons (80.9%) contain only one grade. 7.6% of hexagons do not meet the 75% threshold and have a missing grade.

Table 1 — 1930 Population Distribution According to HOLC Grades

		Population Shares			
	N	A	B	C	D
General Population	38,350,330	3.5%	18.1%	41.8%	36.6%
By Race					
Whites	35,902,615	3.7%	19.1%	43.9%	33.3%
African Americans	2,391,444	0.8%	2.4%	11.3%	85.5%

Notes: The sample includes individuals in the 1930 census with a valid geocode in neighborhoods with a digitized HOLC map.

the shares for C and D grades. Moreover, the smaller average size of A and B neighborhoods explains the percentage reductions for these classes. We also incorporate data from *Area Description* forms (Markley, 2023) in our analysis to investigate heterogeneous effects across different types of D neighborhoods. More details about characterizing different red zones can be found in Section 5.4.

3.2 Full-Count Census Data

We rely on full-count census records for data between 1910 and 1940. We geocode the heads of household by taking advantage of the addresses available in the proprietary version supplied by *ancestry.com* and IPUMS (Ruggles et al., 2020). Census addresses are cleaned following best practices found in the spatial history literature (Logan and Zhang, 2018), and geographic coordinates are assigned by a state-of-the-art locator (ESRI StreetMap Premium 2019) that combines parcel centroids and street locations.[25] Detailed geographic coordinates allow us to construct neighborhood-level averages by combining individual observations with our hexagon grid. Population distributions according to the grades can be found in Table 1. Yellow and red areas include 78.4% of the general population in the 1930 census but contain 96.8% of African American individuals. Table 2 reports descriptive statistics of the 1930 census according to the HOLC grades. The agency's grades captured pre-existing socioeconomic urban patterns. African American percentages, homeownership rates, property values, and rent prices correlate with the HOLC scale. More details about the definitions of variables drawn from the full count census can be found in the Appendix Section A.2. In order to analyze short-term sorting patterns, we construct an individual panel dataset taking advantage of crosswalks provided by the Census Linking Project (Abramitzky et al., 2022). We provide more details about this extra dataset in Section 5.2.

25. The overall proportions of matched addresses for 1910, 1920, 1930, and 1940 are 60.5%, 65.4%, 76.1%, and 73.5%, respectively.

Table 2 — 1930 Descriptive Statistics According to HOLC Grades

	Grade			
	A	B	C	D
Black	0.01 (0.12)	0.01 (0.09)	0.02 (0.13)	0.15 (0.35)
Home Owner	0.67 (0.47)	0.56 (0.50)	0.46 (0.50)	0.33 (0.47)
House Value	21,080 (63,320)	10,600 (27,984)	8,193 (21,394)	8,170 (73,938)
Rent	110 (436)	82 (383)	63 (330)	49 (266)
Income Score	7.27 (0.35)	7.23 (0.32)	7.14 (0.32)	6.95 (0.37)
First Gen. Immigrant	0.17 (0.37)	0.20 (0.40)	0.27 (0.44)	0.33 (0.47)
Unemployed, Men	0.05 (0.22)	0.07 (0.25)	0.10 (0.30)	0.13 (0.34)
Owns a Radio	0.74 (0.44)	0.70 (0.46)	0.61 (0.49)	0.41 (0.49)

Notes: The sample includes individuals in the 1930 census with a valid geocode in neighborhoods with a digitized HOLC map. See Appendix Section A.2 for definitions of census variables in our dataset.

3.3 NHGIS Data

After 1940, we must rely on publicly available census data.[26] We obtain tract-level data for homeownership rates, property values, rent prices, and the shares of African Americans between 1960 and 2010 from the National Historical Geographic Information System (NHGIS) at IPUMS (Manson et al., 2021). We focus on census tracts since they are the smallest geographical units identifiable between 1960 and 2010.[27] Census tracts became available in smaller cities only in later decades. Hence, this source does not fully cover our sample of interest until 1980.[28] Despite these shortcomings, NHGIS is the best available nationwide source of geolocated data for household characteristics continuously available in the second half of the twentieth century.

26. IPUMS released a public version of the 1950 Full Count US Census in early 2024. However, the *ancestry.com* restricted version containing individual addresses has not been updated yet. Assigning observations to our hexagon grid with the public version of the census is not possible.

27. While all hexagons have the same area, census tract dimensions vary. A census tract is always bigger than the hexagons we defined in surface terms. The median census tract contains 34 hexagons in surface terms.

28. As mentioned in Section 4, we focus on cities with population between 30,000 and 60,000. Appendix Table A2 reports coverage rates of NHGIS data for our sample of interest. We do not use NHGIS data from 1950, given the very low coverage of the 1950 census tracts in our cities of interest.

3.4 CoreLogic Deed Records

We supplement NHGIS tract-level data with sale records obtained from CoreLogic, which contains transaction data collected from county assessors and deed registries, including information about sale prices, dates of sale, and the geographic coordinates of the buildings. In our dataset, transactions are binned into 5-year windows according to the sale year and month. As expected, the number of sales recorded in the dataset is much higher in recent years.[29] The nature of CoreLogic data is different from the data sources we have described so far. They are administrative records of realized sales, while census property values are the results of extensive surveys based on self-reports. The granular detail and administrative source of CoreLogic make it a promising addition to the datasets used to study the long-term effects of the HOLC maps.

4. A New Empirical Strategy

We propose a new strategy to measure the short and long-term effects of the HOLC maps. Our approach does not rely on border regression-discontinuity designs, which have been prevalent in the literature on the topic. Instead, we exploit an exogenous population threshold: HOLC staff surveyed only cities with at least 40,000 residents. To reduce the role of unobservable confounders, we focus on cities of a similar size. In particular, we define cities with a population between 40,000 and 60,000 as the treated cities, while the municipalities between 30,000 and 40,000 residents are included in the control group.[30] Figure 2 illustrates this threshold and highlights our definitions of treated cities in purple and control cities in orange. The causal effects are identified by the differences between neighborhoods in treated cities and areas that would have received the same security grade if their city had been mapped.

A machine learning classification algorithm assigns grades to neighborhoods in control cities, replicating HOLC evaluations. The algorithm is trained to associate 1930 census observables to observed grades from the HOLC maps. Using the predicted grades, we then apply a grouped difference-in-differences design to measure the causal effects of the maps' four different ratings. To provide an intuition, let us restrict the analysis to two cities: a treated city, Phoenix, AZ, and a control city, Raleigh, NC. To estimate the effect of the D grade, we compare the outcomes for observations geocoded in Phoenix D areas with observations in Raleigh neighborhoods that the HOLC would have rated D, according to our algorithm.

Building the control group with a machine learning model is a matching technique alternative to synthetic control methods (Abadie and Gardeazabal, 2003; Abadie et al., 2010). Instead of estimating a set of weights

29. Additional details about CoreLogic's coverage of our cities of interest can be found in Appendix Table A2.

30. The length of the population interval is doubled on the treated side to balance the number of cities in the control and treated groups. The number of control cities distant enough from a mapped place is 49. There are only 34 cities between 40,000 and 50,000 in the 1930 census, which are not suburbs of an American metropolis. Considering all not-suburb places until 60,000 people returns a treated group of 52 cities. The magnitudes of coefficients do not change substantially when restricting the treatment group to smaller cities, as shown in Section 5.5. See Appendix A.4 for a list of cities.

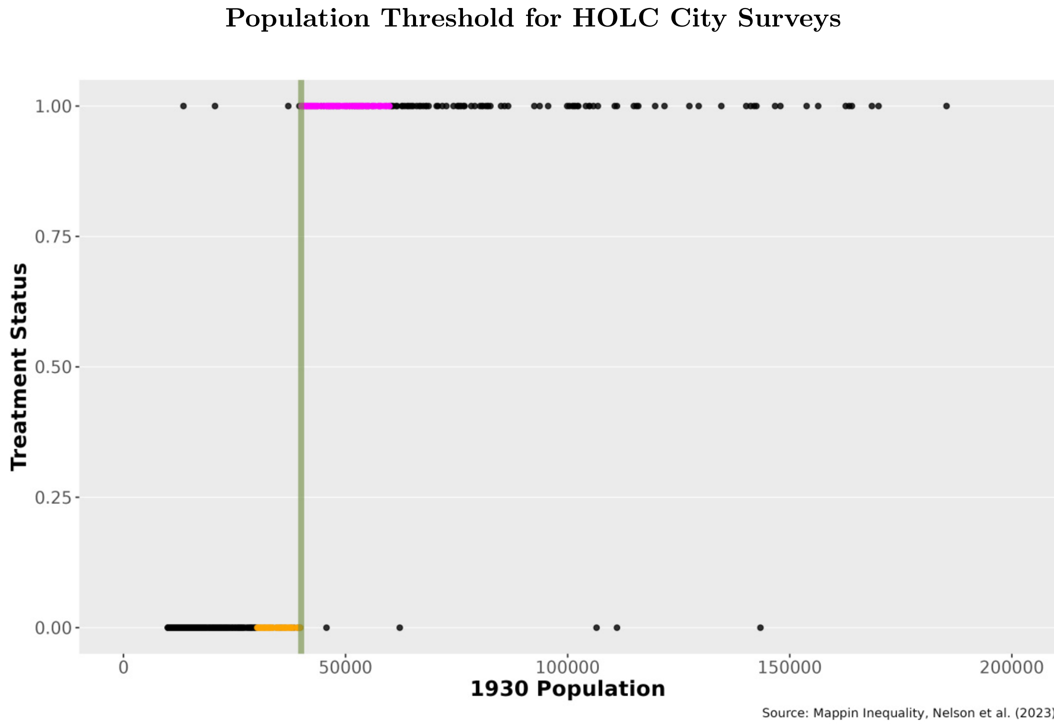


Figure 2 — The graph shows the treatment status of cities according to their 1930 population. The vertical green line highlights the 40,000 people threshold. Orange points identify cities in the control group, while purple ones highlight treated cities.

so that control areas could mirror treated areas, we classify neighborhoods by replicating HOLC grades. In the same spirit of synthetic controls, we do not use post-treatment data when designing the control group, and each observation's contribution to the counterfactual is explicit.[31] Unlike the synthetic control method, control group units are never used in the training procedure that determines the counterfactual composition in our procedure. Moreover, each class's "donor pool" can be easily visualized on a map. The resulting control group will have to meet validity checks not targeted during its design, such as the parallel trends assumption.[32] Hence, our approach reduces the possibility of manipulation in developing a synthetic counterfactual even more.

4.1 A Classification Algorithm

The success of our strategy relies on convincingly replicating HOLC evaluations using 1930 census data. In particular, we are interested in recovering a function that can credibly predict y, the HOLC grade, based on

31. Our approach can be thought of as a particular case of the synthetic control method where a classification model assigns weights. In particular, for treated observations with grade j, we are building a control group that assigns weights with only two values, 0 or 1. Let $\hat{g} \in \{A, B, C, D\}$ be the grade predicted by the classification model, we are assigning the 0 weight to all control observations such that $\hat{g} \neq j$ and a weight equal to 1 if $\hat{g} = j$. The resulting weighted mean will be rescaled by $n_{\hat{g}}$, the number of observations with a predicted class $\hat{g}$.

32. The machine learning training dataset does not contain information about pre-treatment trends. It only includes 1930 census information.

X, a set of neighborhood observables.

Since HOLC appraisers traced area boundaries and assigned grades simultaneously, our procedure should replicate both outcomes. We tackle these goals by classifying a regular grid of hexagons into four grades.[33] This approach imitates HOLC methods and tackles the complex task of drawing borders in control cities. While the hexagon grid helps imitate the original borders, 1930 census data are the best nationwide data source for replicating HOLC surveys. Our dataset provides very good detail about the sociodemographic composition of neighborhoods and housing prices. However, we lack information about mortgages, defaults, and interest rates that the agency regularly collected surveying local financial institutions.

We implement a random forest algorithm (Breiman, 2001) to classify hexagons into one of four HOLC grades. This machine learning method proves effective in dealing with the class imbalance of our classification problem[34] and outperforms other popular classification algorithms.[35] In short, the random forest is a nonparametric and nonlinear model based on decision trees. A tree is a hierarchical series of splitting rules for covariates X. The goal is to find the best binning structure of covariates X and the hierarchy of these splits to predict class y. Since random forests are widely employed in recent economics literature, we will highlight only a few relevant features of the algorithm.[36]

A decision tree provides flexible binning of multiple covariates to maximize the predictive power for the outcome class y. The definition of bins is entirely data-driven, and the process flexibly considers interactions between covariates. As a nonlinear multivariate function, the resulting bins define a link between covariates X and the predicted class $\hat{y}$. This approach usually returns a good in-sample fit but often suffers from poor out-of-sample predictions due to overfitting. Bootstrapped aggregation (bagging) techniques offer a remedy. The solution is to fit several trees on bootstrapped data samples, thus growing a forest. Moreover, each split is determined only by m randomly selected covariates.[37] These steps reduce the correlation between the predictions of each tree, characterizing the forest as "random" and providing reliable out-of-sample predictions. Once the algorithm is trained, the predicted class is the most voted by all the trees in the forest.

33. Hexagons, rather than triangles or squares, are well suited for our goals because they are the most circular-shaped polygon that can generate a regular grid. In particular, they reduce sampling bias, capture curved patterns more easily, reduce the projection distortion due to earth curvature, and provide a better definition of neighbors because of their centroid properties. For more details, see Birch et al. (2007).

34. The minority class share (Grade A) is 7.7% while the maximum one (Grade C) is 42.2%. More details can be found in Table A1.

35. More details about the performance of an ordered logit model in this setting can be found in Section 4.2.

36. See Fuster et al. (2022) for a more detailed explanation of the random forest algorithm.

37. The number of variables to use at each split (m) and the fraction of observations to sample for each tree are parameters that need tuning. More details on tuning our random forest can be found in the Appendix Section A.3.

4.2 Classification Model Results

We train the random forest algorithm with a hexagon-level dataset containing all cities mapped by HOLC.[38] The dataset includes 47 different variables from the 1930 census; most are included at different geographical levels, bringing the total number of training variables to 158.[39] The total number of neighborhoods in our dataset is 164,447.

We assess the performance of our classification model on a test set of hexagons randomly drawn from the original dataset.[40] These observations were excluded from the random forest training procedure and represent an out-of-sample validation of the model performance. Table 3 presents a matrix comparing the observed and predicted grades in the test set (*Confusion Matrix*). The probability of correctly classifying a neighborhood (*Accuracy*) is 91.31%, while the probabilities of correct predictions conditional on observed grades (Class-specific *Sensitivities*) are well above 90% for C and D classes. Comparing the predicted grade distribution (*Detection Prevalence*) with the observed class frequencies (*Prevalence*) shows that our model does not alter the overall distribution of the grades. Given that the identification strategy focuses on cities between 30,000 and 60,000 residents, we are interested in the performance of our model in smaller cities. Appendix Table A3 shows the results we obtain if we restrict our test set to cities with a population below 60,000. Accuracy is still high (89.52%), and performance metrics are similar.

It is worth comparing our machine learning procedure to classification models traditionally used in the economics literature. Accuracy levels above 90% are a substantial performance improvement compared to what we would obtain with an ordered logit. Appendix Table A4 shows the performance of an ordered logit estimated on the same dataset used by the random forest. Overall accuracy reaches only 63.35%, and the predicted grades severely overestimate the prevalence of C neighborhoods while underestimating the presence of D and A neighborhoods. It should be noted that a logit-type model is more direct than a random forest in characterizing the contribution of each variable to determine the probability of a grade. However, the improvement in predictive accuracy offered by the latter is so significant that it compensates for the loss in interpretability.

Prediction accuracy characterizes the model's precision, but it does not provide any insight into the spatial patterns of our predicted maps. In particular, the challenge is to obtain graded areas with a sufficient degree of compactness to mirror the HOLC maps. A comparison between the original HOLC map of Pittsburgh, PA, and our predicted map can be found in Figure 3. Overall, the predicted grades are coherent with the original

38. The results in this section are robust to variations in the population threshold determining which cities are included in the training procedure. The range of variation for this threshold is between 50,000 and 7,000,000 residents.

39. The complete list of variables is available in Appendix A.3.

40. The test dataset represents 25% of the original dataset, while 75% of the observations were used in training the model. The random sampling was stratified according to HOLC grade and city population. The results are robust to changes in the sampling procedure.

Table 3 — Random Forest Performance, Confusion Matrix

		Data: D	C	B	A
Prediction	D	12923	651	85	12
	C	967	18056	920	101
	B	86	581	8405	386
	A	8	26	123	2097
Accuracy		91.31%			
Class Sensitivity		92.41	93.49	88.17	80.77
Prevalence		30.78	42.52	20.99	5.71
Detection Prevalence		30.09	44.12	20.82	4.96

Notes: The matrix compares the observed and the predicted grades for a test set of observations excluded from the training procedure. The test set is a 25% random subsample of the original dataset selected with stratified sampling according to city population and HOLC grade. The level of observation is a neighborhood (hexagon). See Section 3.1 for details about the hexagon definition. The sample includes every hexagon in a mapped city containing at least 20 residents in 1930. A predicted grade is the class predicted by the trained random forest algorithm. See Section 4.1 and Appendix Section A.3 for details about the random forest training procedure. *Overall Accuracy* is the percentage of hexagons whose predicted grades correspond to observed ones. *Class sensitivity* for a grade j is the proportion of correctly predicted hexagons among the spatial units with grade j. *Prevalence* reports the share of each observed grade in the test set, while *detection prevalence* shows the distribution of predicted grades.

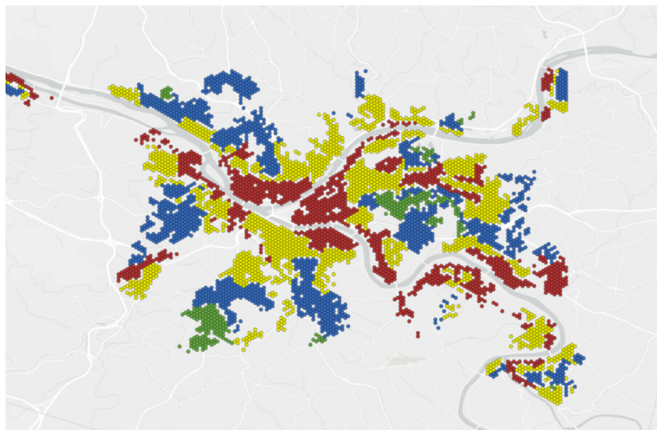
neighborhood shapes.[41] More HOLC maps drawn by the random forest algorithm for selected cities are shown in Figure 4. The final goal of replicating HOLC grades is to draw "redlining" maps in cities between 30,000 and 40,000 people. Figure 5 shows examples of predicted maps for control cities. The model identifies areas for all four grades, returning neighborhood shapes similar to the ones observed in the original maps in larger cities.

Our random forest algorithm has no spatial constraint that would guarantee an output visually similar to HOLC maps. The results rely on a training dataset that includes census observables at different levels of geographical definition. Appendix Figure A3 shows how the predicted map for New Haven, CT, changes when we train random forests with datasets including different levels of geographical aggregation. The top left map is the output we obtain when each hexagon only includes information about its area. Instead, the top right map adds city and county-level information to the training dataset. While generally accurate, these maps suffer from spatial noise, and the resulting neighborhoods cannot be easily encircled in a compact shape.[42] The compactness of the predicted neighborhoods increases when we include information about the area surrounding each hexagon. In particular, for each hexagon, we construct averages of surrounding census observables using 500-meter and 1,500-meter radii (0.31 miles and 0.93 miles, respectively). With the addition of these local averages, the classification algorithm returns predicted maps with plausible neighborhoods,

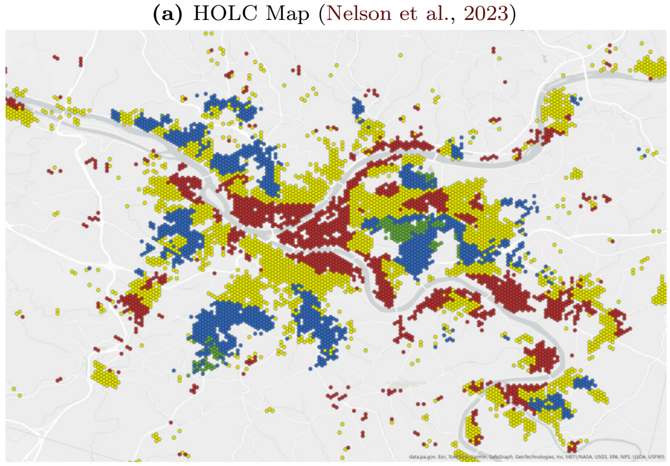
41. The surface covered by our hexagon-level maps is slightly smaller than the area covered initially by HOLC. This is because the federal agency mapped even scarcely populated areas, while our strategy focuses on hexagons with at least 20 residents.

42. The overall prediction accuracy of the random forest considering only neighborhood-level data is 68.49%, while it rises at 75.73% when including city and county level covariates.

Comparison of Observed and Predicted Maps for Pittsburgh, PA



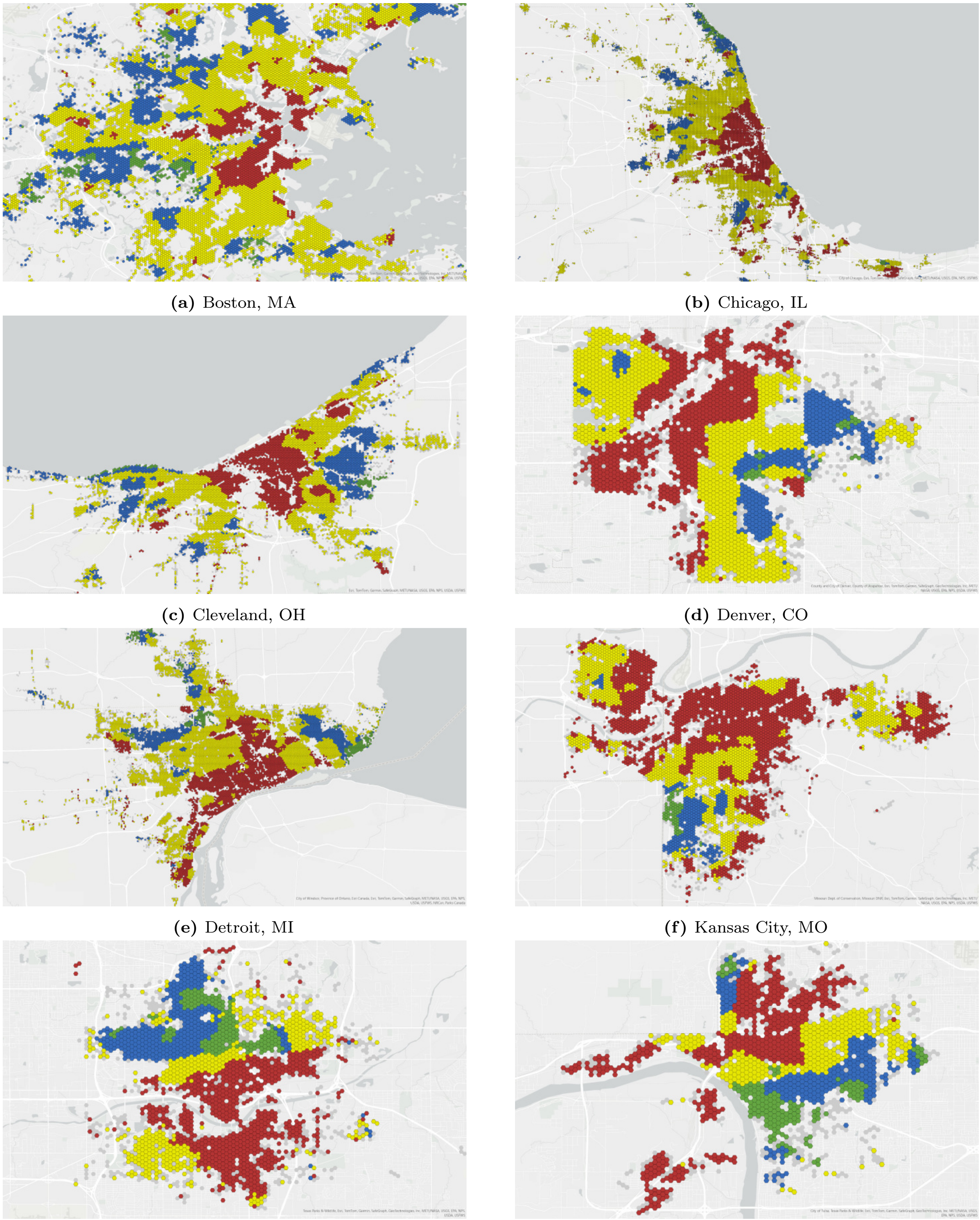
(a) HOLC Map (Nelson et al., 2023)



(b) Predicted Map, Random Forest Algorithm

Figure 3 — The figure compares the digitized version of the HOLC map for Pittsburgh, PA (Nelson et al., 2023) with the hexagon-level map we predict with the trained random forest algorithm. The correspondence between colors and grades is: Green=A, Blue= B, Yellow=C, Red=D. The complete collection of predicted maps can be accessed at this link: *Predicted Maps*. All the maps are north-oriented.

Predicted Maps of Selected Cities



(a) Boston, MA

(b) Chicago, IL

(c) Cleveland, OH

(d) Denver, CO

(e) Detroit, MI

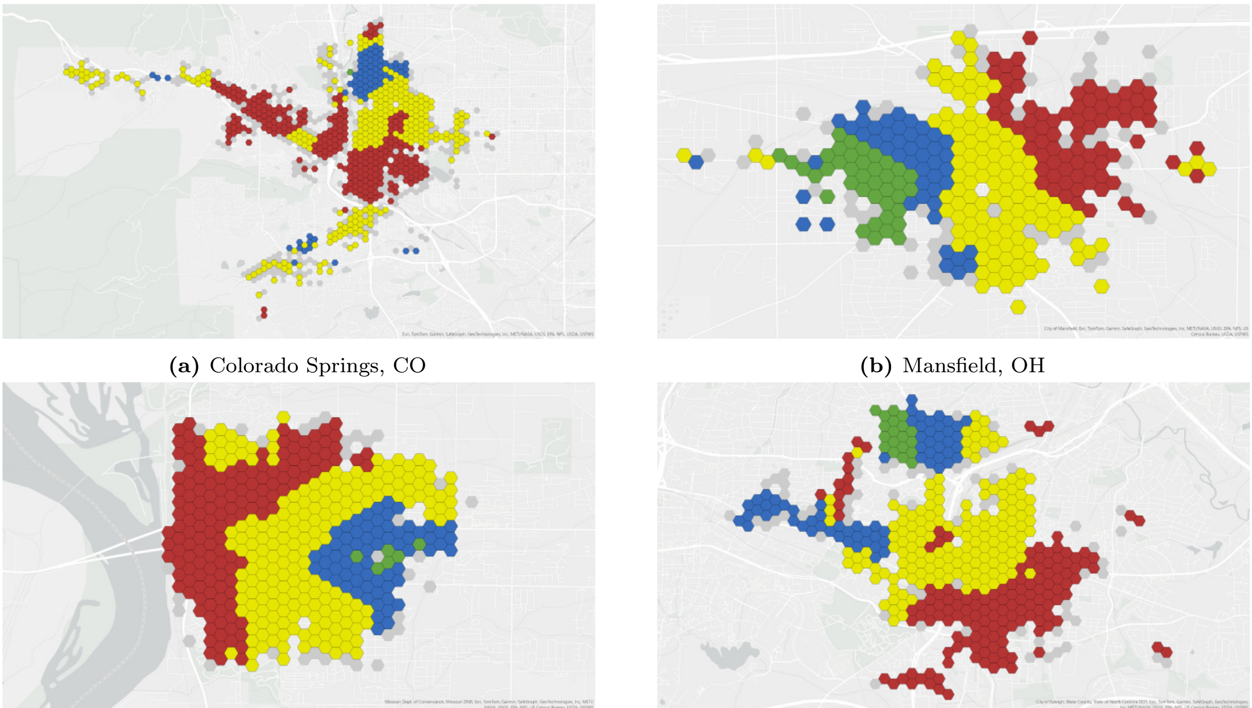
(f) Kansas City, MO

(g) Oklahoma City, OK

(h) Tulsa, OK

Figure 4 — The figure contains the hexagon-level maps predicted with the trained random forest algorithm for selected cities mapped by HOLC. Grey hexagons have fewer than 10 geocoded residents in 1930 and are excluded from the prediction exercise. All the maps are north-oriented. The complete collection of predicted maps can be accessed at this link: *Predicted Maps*.

Predicted Maps in Control Cities



(a) Colorado Springs, CO

(b) Mansfield, OH

(c) Quincy, IL

(d) Raleigh, NC

Figure 5 — The figure contains the hexagon-level maps predicted with the trained random forest algorithm for selected cities mapped by HOLC. The correspondence between colors and grades is Green=A, Blue=B, Yellow=C, Red=D. Grey hexagons have fewer than 10 geocoded residents in 1930 and are excluded from the prediction exercise. The complete collection of predicted maps can be accessed at this link: *Predicted Maps*. All the maps are north-oriented.

as shown in panels (c) and (d) of Appendix Figure A3.[43] Repeating the same procedure for Quincy, IL, a control city, returns the maps contained in Appendix Figure A4.

4.3 Measuring the Effects of HOLC Maps

Our goal is to estimate the grade-specific causal effects of introducing HOLC maps, an innovative information tool for neighborhood appraisal, in the real estate market. In the early stages of the 20th-century information revolution, the HOLC maps could act as a coordination device providing practical area evaluations to local financial institutions. To estimate these effects, we classify neighborhoods into four classes according to their predicted grades using our trained random forest. Then, we apply a difference-in-differences design comparing neighborhoods in treated cities, between 40,000 and 60,000 residents, with those in control cities, between 30,000 and 40,000 residents, separately for each grade. Our last pre-treatment period is 1930, and 1940 is our first post-treatment period. If the empirical design assumptions are deemed credible, the estimated coefficients will capture the global effects of each grade assigned by the federal "redlining" maps.

43. A random forest trained on a dataset including neighborhoods and information about their surroundings achieves an 89.84% accuracy. If we add city and county-level variables to the former dataset, accuracy increases to 91.31%.

In the short run, our specification is:

$$Y^{\hat{g}}_{i,c,t} = \alpha^{\hat{g}} D_c + \gamma^{\hat{g}} P_t + \beta^{\hat{g}} D_c P_t + \delta^{\hat{g}} X_{i,c} + \varepsilon^{\hat{g}}_{i,c,t} \tag{1}$$

In the equation, $Y^{\hat{g}}_{i,c,t}$ is the outcome for neighborhood i with grade $\hat{g}$, in city c, at time t. The term D_c is a treatment dummy, and P_t is a post-treatment indicator. $X_{i,c}$ includes neighborhood observables and information about the surrounding areas. Equation (1) will be estimated by group according to the grade $\hat{g} = G(Z_{i,c,1930})$ assigned by the random forest according to 1930 census observables $Z_{i,c,1930}$. The coefficients of interest are $\left\{\beta^A, \beta^B, \beta^C, \beta^D\right\}$. In Section 5.1, we provide results for different specifications of equation (1), replacing the term $\alpha^{\hat{g}} D_c$ with city fixed effects. The validity of this empirical framework relies on two main assumptions. First, the maps did not affect dependent variables in control cities. Second, outcomes would have evolved in parallel without the policy. We examine the validity of these assumptions in the following section.

We can extend equation (1) to measure medium and long-term effects. In particular, we estimate the following equation at the neighborhood level:

$$Y^{\hat{g}}_{i,c,t} = \alpha^{\hat{g}} D_c + \Gamma^{\hat{g}} \bar{P}_t + \sum_{t \in T} \beta^{\hat{g}}_t D_c P_t + \delta^{\hat{g}} X_{i,c} + \varepsilon^{\hat{g}}_{i,c,t} \tag{2}$$

where $T = \{1930, 1940, 1960, 1965, 1970, 1975, \cdots, 2010\}$. The only new terms compared to equation (1) are $\bar{P}_t = (P_{1940}, P_{1960}, \cdots P_{2010})$ a vector of year dummies for all elements of T (except 1930) and its corresponding vector of coefficients $\Gamma^{\hat{g}}$. In this case, $X_{i,c}$ includes time-invariant geographic controls. Section 5.3 contains the results for different specifications of (2) where we replace the treatment term $\alpha^{\hat{g}} D_c$ with city fixed effects.

Previous research on this topic has focused on estimating the *local* effects of a lower grade, such as D, compared to a higher one, e.g., C, with border regression discontinuity techniques. Such results characterize within-city impacts, while our strategy provides a different type of effect. We capture the *global* effects of the four HOLC grades. We are the first to estimate these effects, which are the most direct reduced-form characterization of the consequences of the federal "redlining" maps. A limitation of our analysis is related to external validity. Our effects are estimated for smaller cities and might not be appropriate to describe the effects of the maps in American metropolises. Another limitation is that our estimates rely on a first-stage classification model. The prediction errors of the random forest might attenuate or inflate the difference-in-differences estimates and affect their precision.[44] Ultimately, the high prediction accuracy of our trained machine learning algorithm and the soundness of the parallel trends assumption reassure us about the overall credibility of this strategy.

44. Our results are robust to substituting the predicted grades with the observed ones in treated cities, as detailed in Section 5.5. Moreover, we propose additional checks of biases introduced by our classification exercise in Section 4.4 to mitigate these concerns.

4.4 Validation of the Empirical Strategy

As noted in Section 4.3, our difference-in-differences framework relies on the no-treatment-spillover assumption and parallel pre-trends. In our context, the no-spillover assumption means that the HOLC intervention would not have affected control cities because of their geographic location. Appendix Figure A5 shows that control cities, in blue, are scattered throughout the country, and they are not suburbs of treated cities, in red. To strengthen the assumption, we include in our analysis only control cities with a distance of at least 20 km (12.4 mi) to the nearest treated municipality.[45] While it is safe to assume that HOLC mapping did not directly affect control cities, it is harder to argue that the assignment of grade g in a particular area does not affect surrounding neighborhoods.[46] If we are worried about the spillovers on surrounding areas in treated cities, the coefficient $\beta^{\hat{g}}$ will combine the effect of grade g and the correlations with other local grades. In Appendix Table A5, we provide the grade composition of an area's surroundings according to their evaluation and treatment status. Most surrounding neighborhoods belong to the same class for all grades. Since the distribution of surrounding evaluations is similar between the control and the treatment group, our empirical approach will partially address these concerns.

The soundness of the difference-in-differences framework hinges on the similar evolution of socioeconomic characteristics between treatment and control neighborhoods before the HOLC intervention. We graphically investigate the soundness of the parallel trends assumption in Figures 6 and A6 using data between 1910 and 1930. Overall, the trends for African American percentage and homeownership rate evolved parallelly in C (yellow) and D (red) areas. The same is true for grade B, while the assumption appears less valid for A (green) areas, and the results for this class, which represents approximately 3% of the sample, should be interpreted with caution. This validity check cannot be completed for property values and rents, two of our outcomes of interest, because the census started to record these variables only in 1930. As a partial remedy, we can investigate the trajectory of alternative socioeconomic variables. The remaining panels of Figures 6 and A6 compare trends for the share of immigrants, percentage of females, the share of married families and their size, together with IPUMS education score and an income measure we built based on 1940 census information.[47] Appendix Table A6 contains the results for an analytical check of the parallel trends assumption. There are no substantial differences between treated and control neighborhoods in the evolution of demographic and economic variables between 1920 and 1930 in B, C, and D areas. Once more, the validity of our assumptions appears weaker for grade A, the top HOLC class. The evolution of the Black percentage in D areas exhibited a 2.3 percentage point difference between 1920 and 1910. However, this discrepancy

45. The average distance between a control city and the closest mapped municipality is 83.9 km (52.1 mi). We provide a robustness check where we increase the minimum distance of control cities in section 5.5.

46. This challenging problem is similar to estimating the effects of an exogenous shock when dealing with non-random exposure, as described by Borusyak and Hull (2023). In our setting, even if it is credible to characterize treatment as random, we might think that neighborhood location will lead to non-random exposure to different grades from the surrounding areas.

47. The census did not record income before 1940. We impute an income score for wage-employed men aged 25-55 between 1910 and 1930 using income measures from 1940. More details are in the Appendix Section A.2.

vanished in the trends observed between 1930 and 1920, the most recent pre-treatment changes.

The classification algorithm we employ to replicate HOLC grades returns predicted maps that we can compare with the original ones in terms of socioeconomic characteristics. Appendix Table A7 compares averages according to observed and predicted grades, showing that the predicted maps do not alter the original socioeconomic composition of C and D neighborhoods. Moreover, we might be worried about the type of bias introduced in the analysis by spatial units receiving a "wrong" grade.[48] Appendix Table A8 shows that even when the model assigns a neighborhood to the wrong class, we are not introducing significant sources of bias, both in terms of 1930 observables or in terms of pre-trends.

Another assumption implicit in our empirical approach is that HOLC practices did not change across different cities. In particular, a predictive model trained with US metropolises might not accurately replicate HOLC grades in smaller cities, which is the ultimate goal of our prediction exercise. Appendix Figure A7 shows that the accuracy level of our random forest algorithm is robust to different training datasets according to the population of cities included. When we restrict our attention to predicting grades for neighborhoods in cities below 60,000 residents, overall accuracy is still close to 90%. We interpret these results as evidence of the high degree of standardization of HOLC grading procedures, making our predictive model a reliable source for HOLC evaluations in smaller cities.

5. The Effects of HOLC Maps

This section presents the difference-in-differences results for our sample of cities between 30,000 and 60,000 residents. Although treatment status is assigned at the city level according to the population threshold, we will measure the effects using neighborhood- or individual-level data. This is because a city-level analysis cannot recover grade-specific coefficients. Moreover, ignoring the heterogeneous impact of the four different grades would lead to empirical results that would mischaracterize the legacy of this federal intervention. Appendix Table A9 shows the results we obtain when we apply a difference-in-differences design to estimate a unique short-term effect of HOLC maps. We do not find any significant impact on our outcomes of interest, suggesting that cities on the two sides of the treatment threshold did not become overall different in the short term. However, to meaningfully capture the consequences of HOLC maps, we need to estimate the effects separately for each grade.

48. In our sample of treated cities, spatial units with different observed and predicted grades are approximately 12%.

1930-1910 Trends in Selected Variables for Neighborhood Grades C and D

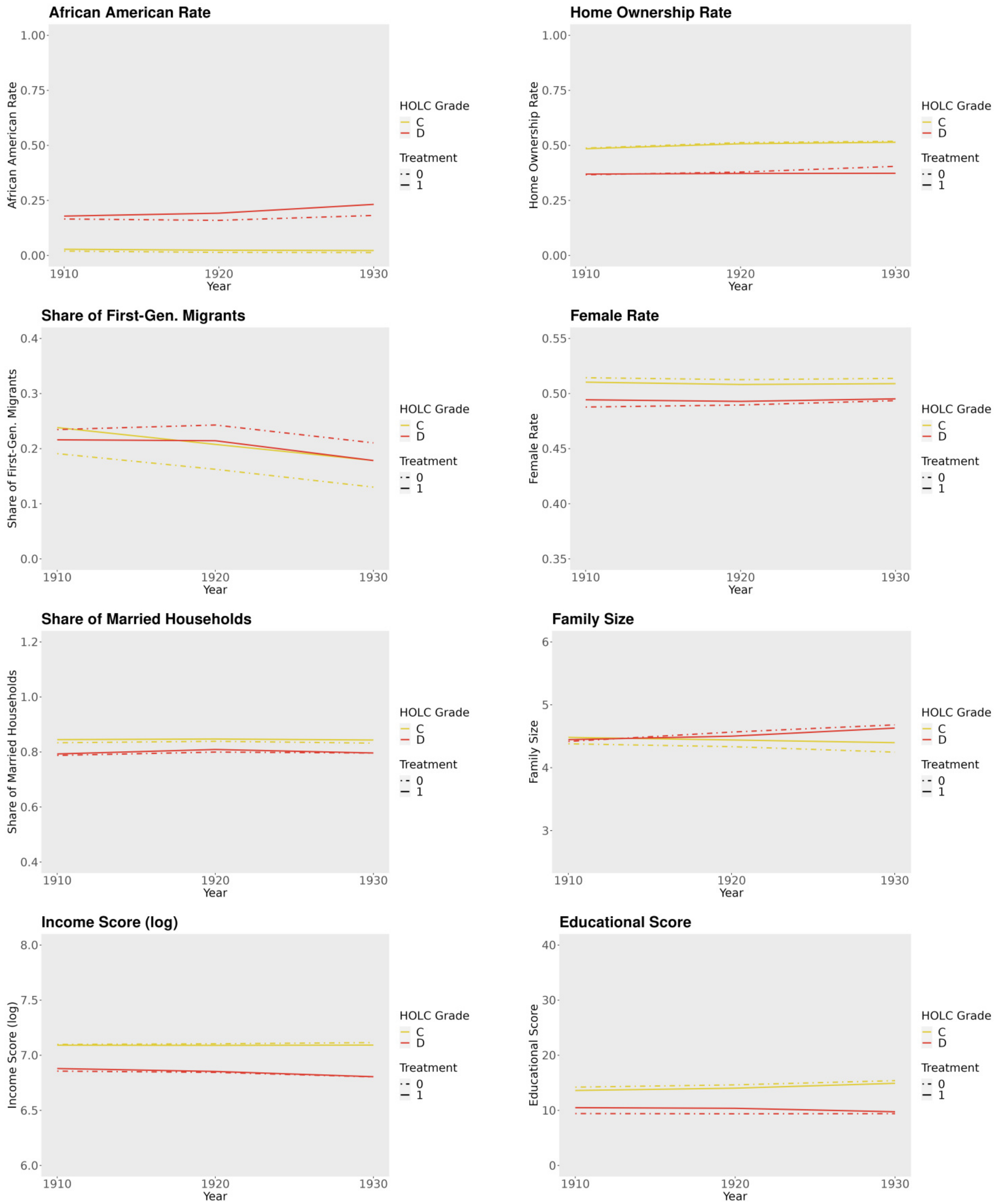


Figure 6 — The figure shows pre-trends for selected variables according to their predicted grade. The point estimates are averages of hexagon-level observations. Yellow lines indicate the C grade, and red ones correspond to the D evaluation. Solid lines report averages of treated observations, and dashed ones plot the control group means. The sample includes hexagons in cities with a 1930 population between 30,000 and 60,000, with at least 20 residents in 1930.

5.1 Short Term Results

We start by estimating equation (1) separately for each grade. The coefficients of interest reported in the tables of this section are $\{\beta^A, \beta^B, \beta^C, \beta^D\}$. The dataset includes neighborhood-level observations from 1930, the pre-intervention period, and 1940, the post-intervention one.

Table 4 — Short-Term Difference-in-Differences Results, by Grade

	Dependent Variables			
	African American Percentage	Home Ownership	Property Values (Logs)	Rent Prices
DiD_A	0.002	0.080***	0.101**	34.1
	(0.007)	(0.027)	(0.046)	(75.4)
DiD_B	0.003	0.010	0.000	0.4
	(0.002)	(0.009)	(0.020)	(17.6)
DiD_C	0.001	-0.010	-0.056**	6.9
	(0.001)	(0.008)	(0.026)	(8.9)
DiD_D	0.013***	-0.012*	-0.081**	-12.3**
	(0.004)	(0.007)	(0.040)	(5.8)

Notes: The table reports difference-in-differences coefficients we obtain estimating equation (1) by grade for four different outcomes with neighborhood-level observations. Each row contains the DiD coefficients for a given grade. The table shows the DiD coefficients resulting from a DiD framework with a city fixed-effect and geographic and demographic controls at the hexagon and local area level. The list of controls can be found in the notes of Tables 5. Standard errors, in parentheses, are clustered at the city-year level. The regression specification is analogous to the one in column (4) of Table 5. The sample includes neighborhoods in cities with a 1930 population between 30,000 and 60,000, with at least 20 residents. See Appendix Section A.4 for a list of cities. Significance: * 0.10, ** 0.05, *** 0.01.

Table 4 reports the four grade-specific coefficients for our outcomes of interest. Regarding African American percentage, we find a 1.3 percentage point increase in the lowest-rated areas (D, red), a 6.1% increase to the baseline. Given the near absence of Black Americans in A (green), B (blue), and C (yellow) neighborhoods, we do not find any effect in other areas.[49] The second column reports the results for homeownership rates. We find a 1.2 percentage point decrease in the percentage of homeowners in D zones in 1940, shortly after the introduction of the maps. This result is only significant at a 10% level when standard errors are clustered at the city-year level.[50] There is suggestive evidence of a weaker, negative effect in C (yellow) areas and a similar positive one in B neighborhoods. We find an 8.0 percentage point increase in the best-rated regions (A), but the caveats we mentioned in Section 4.4 apply in this case, so this result should be interpreted with skepticism.

49. The average percentages of Black residents in A, B, C, and D neighborhoods are respectively: 1.5%, 1.3%, 2.6%, and 21.2%.

50. In our research design, the treatment is assigned at the city level. Since 10-year gaps in our data separate different periods, we do not allow for within-city serial correlation of standard errors. This clustering choice does not address unobserved, within-city, serially correlated shocks over a ten-year time span. Another threat is that hexagon-specific error components could be serially correlated across decades. Note that our results are robust in including city-fixed effects, which will absorb constant city-level error components. We report results with standard errors clustered at the city level, our most conservative option, in Appendix Table A10. The median ratio between city-year clustered standard errors and city-clustered standard errors is 0.702; the average one is 0.782.

The estimated coefficients for property values, in logarithmic form, are also contained in Table 4. The assignment of C and D grades caused a relatively faster real estate depreciation in treated neighborhoods. The results for the log-outcome mirror the ones for the counterpart in levels as shown in Appendix Table A11: property values in 1940 are significantly lower in treated C and D hexagons relative to their 1930 baseline. While the reduction in property prices in red areas confirms the popular narrative about "redlining", the negative effect for C neighborhoods is more surprising, and it was first documented in Aaronson et al. (2021b). Our empirical design does not find any significant impact on B areas.[51] Regarding rent prices, we only find a significant decrease in D zones, providing additional evidence of a weakened housing market in neighborhoods colored red by HOLC agents.

Table 5 — Short-Term Difference-in-Differences Results. Property Values

	Dependent variable: Property Value (Logs)			
	(1)	(2)	(3)	(4)
DiD_A	0.057	0.085*	0.081*	0.101**
$\bar{Y}^A = 9.270$	(0.244)	(0.045)	(0.045)	(0.046)
DiD_B	0.003	0.003	-0.003	0.00
$\bar{Y}^B = 8.863$	(0.102)	(0.020)	(0.020)	(0.020)
DiD_C	-0.052	-0.052**	-0.057**	-0.056**
$\bar{Y}^C = 8.465$	(0.075)	(0.025)	(0.026)	(0.026)
DiD_D	-0.090	-0.092**	-0.089**	-0.081**
$\bar{Y}^D = 7.940$	(0.141)	(0.040)	(0.040)	(0.040)
City Fixed. Eff.		X	X	X
Neighborhood Controls			X	X
Local Area Controls				X

Notes: The table reports difference-in-differences coefficients obtained estimating equation (1) by grade for four different outcomes with neighborhood-level observations. Each row contains the DiD coefficients for a given grade. The table shows the DiD coefficients resulting from different DiD designs. The first column reports the results from equation (1) without controls and column two replaces the treatment indicator with a city fixed-effect. Columns three and four add controls at different geographic levels. The neighborhood-level controls include geographic coordinates, a scaled measure of distance from the city center, imputed income score, and average family size. The local-area controls' set includes averages of the following variables: African American percentage, homeownership rates, family size, radio ownership, property values, rent prices, income score, male unemployment rate, and family size. Standard errors, in parentheses, are clustered at the city-year level. We provide details about how local averages are defined in Appendix Section A.3. The sample includes neighborhoods in cities with a 1930 population between 30,000 and 60,000, with at least 20 residents. See Appendix Section A.4 for a list of cities. Significance: *0.10, **0.05, ***0.01.

Table 5 provides insights into our measurement of the short-term effects of the maps. The columns propose different versions of our DiD design for property values separately for the four grades. The results from a simple difference-in-differences design are reported in column one. Replacing the treatment indicator with a city-fixed effect does not substantially alter the coefficients but considerably increases the estimates' precision.

51. We are hesitant in interpreting the estimated coefficient β^A as the causal effect of grade A. The significantly smaller sample size, paired with weak evidence of parallel pre-trend for this class, invites caution when considering the results for this grade. Moreover, the result for the log-outcome regression is not confirmed by the linear-outcome regression results in Appendix Table A11.

Columns three and four combine a city-fixed effect with sociodemographic controls for each neighborhood and surrounding area. Our preferred specification is column four, which is reported in previous and subsequent tables. Analogous tables for the other outcomes of interest can be found in Appendix Tables A11, A12, A13, and A14.

5.2 Moves Across Neighborhoods

The previous section documented changes in neighborhood characteristics across treated and control cities. To what extent are these effects due to residential moves? Did the HOLC maps cause migration per se? To investigate these questions, we must switch from a neighborhood-level panel dataset to an individual-level one. We will link individuals between the 1930 and 1940 censuses using crosswalks provided by the Census Linking Project (CLP) (Abramitzky et al., 2022).[52] While our DiD sample contains 4,675,475 individuals in 1930 and 4,561,035 in 1940, we can only match 376,170 across the two decades using the CLP crosswalks, approximately 8% of our census observations. For these individuals, we can compare the locations in the two decades to identify who changed neighborhoods, moved into red areas, or stayed.

In Table 6, we measure differences in individual moves between treatment and control cities in HOLC red neighborhoods. For white Americans, we find a 3.3 percentage point difference in the probability of changing the neighborhood and a 3.9 percentage point increase in the likelihood of moving to an area with a better grade. We do not find similar results for the African American sample. This evidence suggests that white Americans drive a modest increase in move-out from red areas associated with the maps. The intuition is confirmed in Appendix Figure A8, where the two top panels show a larger shrinkage of D residents between 1930 and 1940 for whites in treated cities. Such difference cannot be observed in the lower panels for African Americans, who show similar grade-migration patterns between the treatment groups.

We continue to employ our panel data to replicate the DiD analysis for long-term residents of red areas (*Stayers*) and former dwellers that moved out of these zones by 1940 (*Movers*), separately by race. Table 7 reports the effect of the lowest HOLC grade for individual homeownership and property values, together with the Black American rate of their 1940 neighborhood. We do not find meaningful differences in these outcomes for white Americans who stayed in red neighborhoods. Black long-term residents lived in neighborhoods with a higher concentration of African Americans, but we do not find significant differences between homeownership and house values. Overall, the starkest differences can be found in the Black *Movers* sample, which shows a 5.3 percentage point lower homeownership rate and a higher local percentage of African Americans.[53] These

52. The crosswalks we use are obtained employing the Abramitzky, Boustan, and Eriksson (ABE) algorithm. Details about this linking algorithm and comparisons with some alternatives can be found in Abramitzky et al. (2021). In particular, we use the ABE algorithm with NYSIIS names. NYSIIS (New York State Identification and Intelligence System) provides a phonetic algorithm to standardize names *"based on their pronunciation so that names can be matched even if there are minor spelling differences"* Abramitzky et al. (2021). Race is also used as a matching variable.

53. The result for property values is based on a very small sample of Black homeowners and should be interpreted with skepticism.

Table 6 — Short-Term DiD Results, Individual Panel Dataset, Grade D

	Neighborhood Move		*Grade Move*	
	Whites	Blacks	Whites	Blacks
$Treated_c$	0.033**	0.010	0.039**	-0.003
	(0.016)	(0.021)	(0.019)	(0.013)
$\bar{Y}$	0.338	0.238	0.299	0.091
N	56,235	8,085	56,235	8,085

Notes: The table reports the coefficients obtained estimating differences in moving rates between control and treated cities. The first two columns show coefficients obtained from a regression where the dependent variable is an indicator for neighborhood move between 1930 and 1940. The last two columns come from a regression with an indicator for changing HOLC grade between 1930 and 1940 as the independent variable. In both cases, the independent variable of interest is a city-level treatment indicator. The regressions are estimated separately by race: columns one and three are based on the white sample, and columns two and four are estimated with the Black sample. The regressions include controls at the hexagon and local area level. The list of controls and the sample definition can be found in the notes of Tables 5. Standard errors, in parentheses, are clustered at the city level. See Appendix Section A.4 for a list of cities. Significance: * 0.10, ** 0.05, *** 0.01.

Table 7 — Short-Term DiD Results, Individual Panel Dataset, Grade D

	Whites			*African Americans*		
Dependent Var.	Black %	Own.	Prop. Val.	Black %	Own.	Prop. Val.
Pan. A: Stayers						
DiD_D	0.006**	-0.011*	0.054	0.033***	-0.007	0.097
	(0.003)	(0.007)	(0.046)	(0.008)	(0.018)	(0.063)
$\bar{Y}$	0.059	0.615	7.961	0.700	0.519	7.341
N	60,475	60,103	36,515	10,241	10,204	5,180
Pan. B: Movers						
DiD_D	0.002	-0.0002	-0.115**	0.027**	-0.053**	-0.292**
	(0.003)	(0.014)	(0.059)	(0.014)	(0.026)	(0.137)
$\bar{Y}$	0.058	0.370	7.805	0.630	0.271	7.113
N	37,838	37,349	13,293	3,799	3,765	957

Notes: The table reports difference-in-differences coefficients obtained estimating equation (1) with individual-level observations for three different outcomes. Panel A report estimates for individuals that did not move out of their residential D neighborhood (hexagon) between 1930 and 1940. Panel B provides the results for individuals who moved out of their residential D neighborhood (hexagon) between 1930 and 1940. The first three columns show coefficients obtained with a sample of white individuals, and the last three show the analogous estimates for the Black Americans sample. The table shows the coefficients resulting from a DiD framework with a city fixed-effect and geographic and demographic controls at the hexagon and local area level. The list of controls can be found in the notes of Tables 5. Standard errors, in parentheses, are clustered at the city-year level. The regression specification is analogous to the one in column (4) of Table 5. The sample includes individuals in cities with a 1930 population between 30,000 and 60,000. See Appendix Section A.4 for a list of cities. Significance: * 0.10, ** 0.05, *** 0.01.

results help us characterize the short-term effects of the D grade, suggesting that the maps reinforced the spatial concentration of Black Americans and increased barriers to homeownership for African American movers. At the same time, long-term residents in our individual panel were unaffected by property price drops.

5.3 Long Term Results

At the time of writing, full individual census data are unavailable starting in 1950. We must employ alternative data sources to investigate the effects of HOLC maps in the second half of the twentieth century. As mentioned in Section 3.3, we use tract-level NHGIS data between 1960 and 2010. Unfortunately, this data source did not provide sufficient coverage for our cities of interest in 1950, so we dropped this decade in the analysis. We estimate the model described in equation (2) separately for each grade with neighborhoods (hexagons) as the unit of observations. Since census tracts are always bigger than hexagon neighborhoods in our cities of interest, the geographical variation underlying these estimates is smaller than the one we exploited in previous estimates.

Figure 7 shows the DiD coefficients for African American shares, homeownership rates, and rent prices in C and D areas. The corresponding estimates are in Appendix Table A15. The increase in the local shares of Black Americans in D (red) areas we found in 1940 is confirmed in later years: we find moderate effects in D areas between the 1970s and the early 2000s. We do not find significant differences in homeownership rates and rent prices between treated and control neighborhoods in the long-term. This is one of the starkest differences in our results compared to the previous literature.[54] Our empirical strategy provides a more conservative picture of the long-term effects of the HOLC maps. Because of their high level of geographical aggregation, NHGIS data do not provide enough information to characterize the long-term effects of HOLC maps on property values. We include an additional data source in our long-term analysis to compensate for the lack of precision in these estimates.

We then turn to an alternative, more granular source of information: the CoreLogic deeds. This additional dataset allows us to assess the impact of HOLC maps on real estate transactions between 1965 and 2005.[55] We estimate equation (2) with neighborhood-level observations grouped in time bins with a 5-year frequency. Each transaction is assigned a neighborhood, hence a HOLC grade, using their geographic coordinates, which are available in CoreLogic. As in previous specifications, standard errors are clustered at the city-year level. Figure 8 shows the resulting coefficients and 95% confidence intervals. We find negative causal effects of HOLC maps on house values, in logarithmic form, between 1965 and 1980 in D (red) and C (yellow) areas. Statistically significant effects cannot be detected starting in the mid-1980s. Appendix Table A16 shows the corresponding results for the four different HOLC grades.[56] As we mentioned in Section 2, between 1974 and

54. It is not straightforward to directly compare our results with those in the existing literature. For example, Aaronson et al. (2021b) employs a different empirical method and sample of American cities. While we focus on cities with a population between 30,000 and 60,000 to safeguard the credibility of our DiD framework, Aaronson et al. (2021b) use 149 larger redlined cities for the border-discontinuity exercise.

55. The choice of this period is based on the coverage of CoreLogic deeds for our cities of interest. In particular, the dataset does not provide enough transactions to obtain reliable estimates before 1965. Additional details about CoreLogic's coverage of our cities of interest can be found in Appendix Table A2. We stopped the analysis in 2005 to avoid including the effects of the subprime mortgage crisis of the early 2000s.

56. The results are similar when considering the linear version of the outcome, as it is shown in Appendix Table A17.

Long-Term Difference-in-Differences Results, Census Data

Figure 7 — The figure shows the estimated coefficients for regression 2 and their 95% confidence intervals for African American percentage, homeownership rates, and rent prices. The figure includes the results for grades C and D between 1960 and 2010 obtained using NHGIS data described in Section 3.3. Yellow lines indicate the C grade, and red ones correspond to the D evaluation. The coefficients and standard errors are the ones reported in Table A15. In the reported specification, we replace the indicator for treatment, which is assigned at the city level, with a city fixed-effect. The list of controls includes geographic coordinates and their squares together with state-specific quadratic time trends. The sample consists of neighborhoods in cities with a 1930 population between 30,000 and 60,000, with at least 20 residents. See the Notes of Table A15 for more estimation details.

1977, three critical legislative measures were introduced with the primary goal of counteracting redlining in the mortgage market. Our results suggest that the combined effects of the Equal Credit Opportunity Act, the Home Mortgage Disclosure Act, and the Community Reinvestment Act might have been sufficient to offset the persistent effects of HOLC maps on property prices. Once more, our empirical approach provides a more moderate and nuanced description of the legacy of the HOLC maps. We find long-term increases in the percentage of African Americans in D neighborhoods and negative effects on property values in D and C neighborhoods up until the early 1980s. Unlike several studies that have associated maps with present-day outcomes, we do not find any meaningful effect of the maps on American neighborhoods in recent times.

Long-Term Difference-in-Differences Results, Property Values

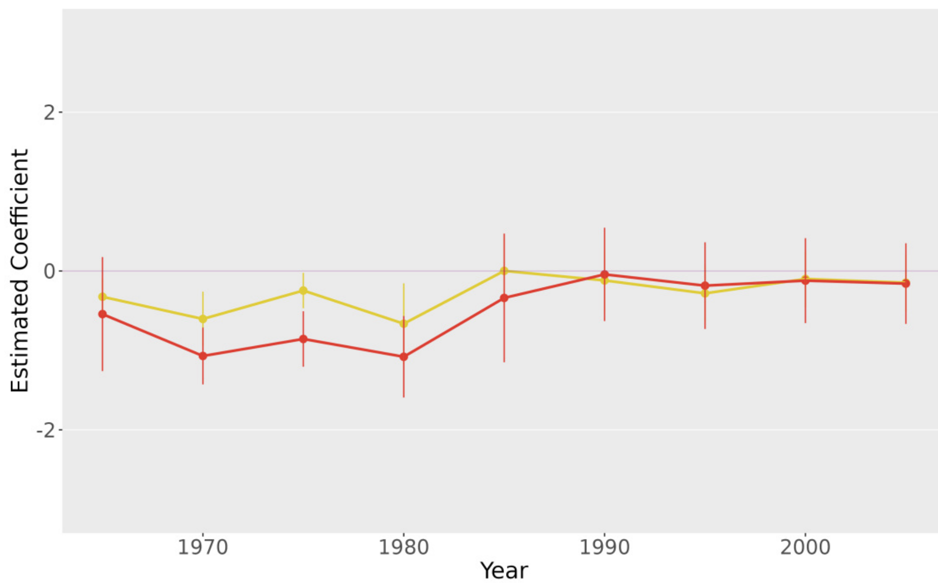


Figure 8 — The figure shows the estimated coefficients for regression 2 and their 95% confidence intervals for log-transformed property values. The figure includes the results for grades C and D between 1960 and 2005 obtained using CoreLogic data described in Section 3.4. Yellow lines indicate the C grade, and red ones correspond to the D evaluation. The coefficients and standard errors are the ones reported in Table A16. The sample consists of neighborhoods in cities with a 1930 population between 30,000 and 60,000, with at least 20 residents. See the notes of Appendix Table A16 for more estimation details.

5.4 Area Descriptions and Heterogenous Effects

Our analysis so far has focused only on the maps' grades, implicitly assuming that every neighborhood should be classified in one of four homogenous categories. While the simplicity and transparency of this approach are valuable, it should be noted that there is substantial heterogeneity in areas with the same evaluation (Fishback et al., 2023). To understand how the HOLC intervention affected different communities, we take advantage of the other component of HOLC *City Surveys*, the *Area Description* forms. Combining a digitized version of the forms provided by Markley (2023) with our geocoded census data, we can identify three neighborhood types to focus on: African American neighborhoods, European immigrant communities, and areas with lower-income residents.[57] Moreover, employing this additional data source addresses one of the major shortcomings of recent quantitative research on the maps, which usually ignores the *Area Descriptions* when interpreting the maps (Markley, 2024).

We repeat our short-term analysis separately for each neighborhood type.[58] Table 8 shows that effects on the percentage of Blacks can be found only in African American areas, clarifying that the red grade reinforced pre-existing patterns of racial segregation and did not affect racial composition in other disadvantaged neighborhoods. There are significant effects on homeownership in Black and low-income neighborhoods, approximately a 10% and a 5% drop compared to the baseline, respectively. At the same time, none of the results for the D grade from the general population applies when we focus on immigrant neighborhoods.[59] The negative effects on property values we found in Section 5.1 seem driven only by African American neighborhoods. The stark difference in results between immigrant areas and Black neighborhoods is confirmed in Appendix Table A18 where we further restrict the sample to low-income neighborhoods with a substantial presence of African Americans or European-born individuals. While we do not find any meaningful effect in disadvantaged immigrant neighborhoods, areas with "lower class" workers and a majority of Blacks show the starkest effects of the maps.

We can extend this heterogeneity analysis to the long-term effect of the maps estimating equation 2 for each sample of neighborhoods. The results for the long-term impact of the red grade on homeownership are reported in Table 9. Once more, there are clear differences in the effects of the HOLC intervention across neighborhoods with the same grade. We detect significant drops in homeownership rates in poorer neighborhoods starting in the 1960s, while no similar differences appear in immigrant neighborhoods or the general population. Appendix Table A19 reports the results for the long-term impact on African American percentages, confirming that the increases in Black Americans can only be found in historically Black neighborhoods. By acknowledging that HOLC grades are not homogeneous categories, we can clarify the impact of the "redlining" maps. Coloring a neighborhood in red did not have the same effect everywhere. In particular, the bulk of the negative impact of the maps is concentrated in African American areas. At the same time, the consequences of stronger barriers to homeownership can be found in low-income neighborhoods.

57. We cannot rely only on digitized *Area Description* data because they cover only treated cities. We need to find similar information in the 1930 census to define subsamples with control and treated units. In particular, we define a Black neighborhood as any hexagon with a percentage of at least 50% African Americans in one of the two sources. We label a neighborhood as an "immigrant" when the variable "Percentage Foreign Born" from the HOLC form is higher than 50%. Alternatively, we define an "immigrant" area if less than a third of residents are native-born individuals according to the 1930 census variable *nativity*. Low-income neighborhoods are identified in the *Area Description* data if the residents' occupations are classified as "low" in Markley (2023) or if the census-constructed occupational score is less than 2650$. This specific threshold was picked to mirror the distribution of the occupational score in the low-income neighborhood defined by the *Area Descriptions*.

58. Note that the three sets of neighborhoods are not exclusive. The same neighborhood could be included in multiple categories if it complies with more than one definition.

59. We find the same null results when defining immigrant neighborhoods using the ethnicities and the places of birth reported in the *Area Description* forms and the census.

Table 8 — Short-Term Difference-in-Differences Results, Grade D

	Neighborhood Samples		
	Afr. Am. Neighborhoods	Immigrant Neighborhoods	Low Income Neighborhoods
African American Perc.	0.044***	0.004	0.001
	(0.011)	(0.005)	(0.005)
Home Ownership	-0.037**	-0.008	-0.024***
	(0.015)	(0.009)	(0.008)
Property Values (Logs)	-0.129***	0.012	0.032
	(0.062)	(0.071)	(0.049)
$\bar{Y}^D$, Afr. Am.	0.812	0.030	0.171
$\bar{Y}^D$, Home Own.	0.357	0.538	0.444
$\bar{Y}^D$, Prop. Val.	7.53	7.89	7.86
N^D	3,948	3,106	6,014

Notes: The table reports difference-in-differences coefficients we obtain estimating equation (1) for three different outcomes with neighborhood-level D-grade observations. Different columns show the results employing different samples. An African American neighborhood is any hexagon with a percentage of at least 50% African Americans according to the *Area Description* forms or the 1930 census. We label a neighborhood as "immigrant" when the variable "Percentage Foreign Born" from the HOLC form is higher than 50% or if less than a third of residents are native-born individuals according to the census variable *nativity*. Low-income neighborhoods are identified in the *Area Description* data if the residents' occupations are classified as "low" in Markley (2023) or if the census-constructed occupational score is less than 2650$. Each row contains the DiD coefficients for a given outcome. The regression specification is analogous to the one in column (4) of Table 5. More details about the regressions estimates can be found in the notes of Table 5. Significance: * 0.10, ** 0.05, *** 0.01.

Table 9 — Long-Term DiD Results, Grade D, Homeownership Rates

	Neighborhood Samples			
	All Neighborhoods	Afr. Am. Neighborhoods	Immigrant Neighborhoods	Low Income Neighborhoods
DiD_{60}	-0.020	-0.059*	0.025	-0.061***
	(0.019)	(0.035)	(0.036)	(0.022)
DiD_{70}	-0.002	-0.053	0.032	-0.041*
	(0.019)	(0.037)	(0.037)	(0.022)
DiD_{80}	-0.011	-0.061*	0.047	-0.052**
	(0.018)	(0.035)	(0.021)	(0.021)
DiD_{90}	-0.004	-0.050	0.045	-0.056***
	(0.018)	(0.034)	(0.036)	(0.020)
DiD_{00}	-0.008	-0.039	0.031	-0.058***
	(0.017)	(0.032)	(0.036)	(0.019)
DiD_{10}	0.005	-0.016	0.033	-0.056***
	(0.018)	(0.034)	(0.035)	(0.019)
N	68,103	15,731	12,436	22,670
R^2	0.300	0.335	0.364	0.447

Notes: The table reports difference-in-differences coefficients we obtain estimating equation (2) for homeownership rates with neighborhood-level D-grade observations. Different columns show the results employing different samples. See the notes of Table 8 for details about the subsample definitions. Each row contains the DiD coefficients for a given census decade. The regression specification is analogous to the one in Appendix Table A15. More details about the regressions estimates can be found in the notes of Table A15. Significance: * 0.10, ** 0.05, *** 0.01.

5.5 Robustness

In the results we discussed so far, we grouped observations according to the grade the random forest algorithm predicted. In Appendix Table A20, we show that our short-term results are robust when we replace predicted grades with observed HOLC classes for treated observations. The same applies to the long-term results reported in Appendix Tables A21 and A22. Another robustness check is presented in Appendix Tables A23 and A24 where we change the definitions for the samples of cities in our treatment and control group. In the former table, we obtain similar results when employing treated observations from 45 cities between 40,000 and 50,000 residents (instead of 60,000). In the latter table, the results are similar when restricting the sample to 34 cities with 40,000 and 50,000 residents who are not a suburb of a major American city. In this last table, we also increase the distance threshold to include cities in the control group: they must be at least 30km (18.6 mi) away from a treated city.

Short-Term Diff-in-Diff Results, Placebo Treatment Thresholds

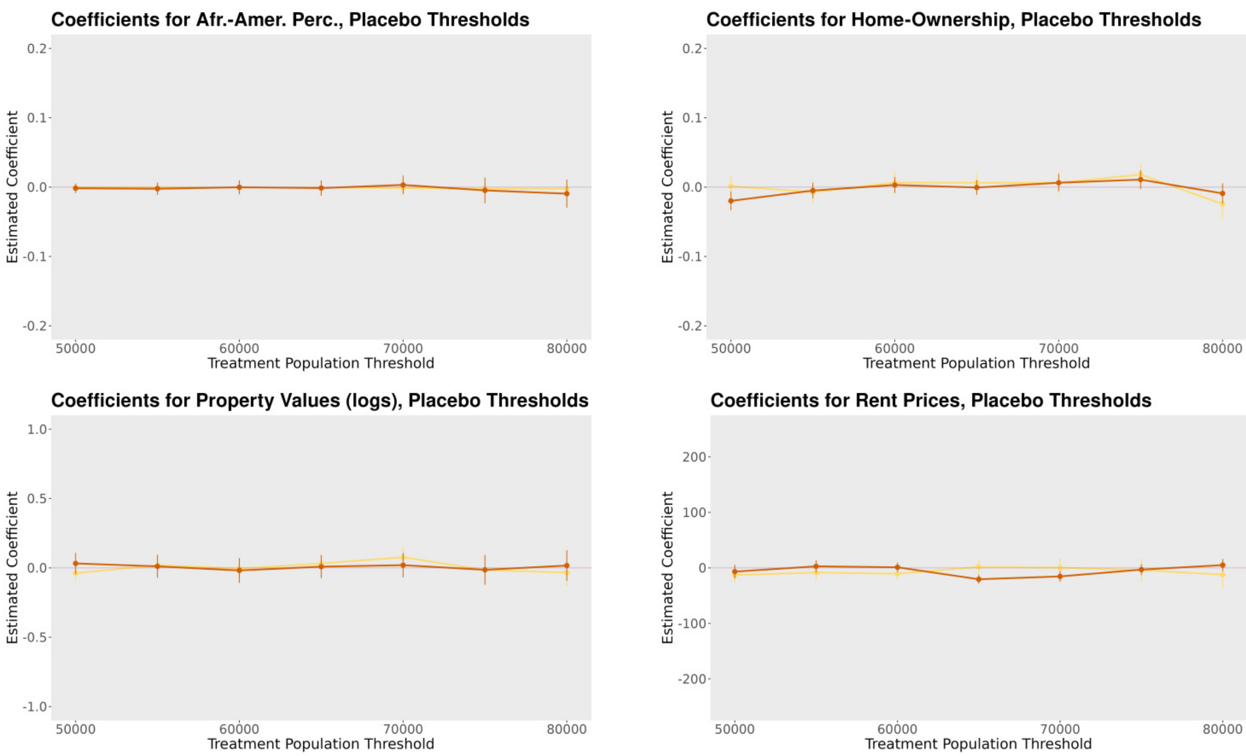


Figure 9 — The figure reports DiD coefficients we obtain estimating equation (1) for four different outcomes with neighborhood-level observations for C and D areas. The bars correspond to their 95% confidence intervals. Different panels show the results for different outcomes. The C grade results are reported in yellow, and the D grade results are plotted in red. The horizontal axis indicates the placebo treatment threshold used for that Diff-in-Diff regression. Each threshold defines a new treatment group and a new control group, imitating the construction of our main samples based on the original threshold of 40,000 residents. For each threshold c, the control group includes neighborhoods in cities with a population in the interval $[c - 10,000, c)$. The treatment group comprises cities in the population interval $[c, c + 20,000]$. The regression specification is analogous to the one in column (4) of Table 5. More details about the regressions estimates can be found in the notes of Table 5.

Our strategy might be capturing structural differences in the evolution of smaller versus bigger cities between 1930 and 1940. Appendix Table A25 shows that we do not find meaningful effects if we focus on alternative outcomes such as female percentage, marriage rate, age at first marriage, labor force inactivity, or male unemployment rate. As a final check, we replicate our empirical approach with seven placebo population thresholds between 50,000 and 80,000 residents, with 5,000-unit intervals. Each threshold defines a new treatment group and a new control group, imitating the construction of our main samples based on the original threshold of 40,000 residents. These results should not replicate our main results since HOLC practices did not differ according to these alternative thresholds. Figure 9 contains the results of this placebo exercise. Each panel plots the Diff-in-Diff coefficients for the C and D grades according to the population thresholds on the horizontal axis. We do not find any indication of a consistent difference between bigger and smaller cities according to the C and D grade classifications.

6. Conclusion

In the second half of the 1930s, a federal agency undertook an unprecedented survey of US neighborhoods in more than 200 cities. Its goal was to provide unified standards for real estate assessments and promote stability in a market that had just begun to recover from the Great Depression. The initiative was a data-driven effort based on the most advanced theories of urban development of the time, and its result was a data analytics tool sought after by real estate professionals. Less than a hundred years later, the Home Owners' Loan Corporation maps have become a symbol of structural racism in the popular press and the political debate. Today's negative judgments of HOLC practices are based on non-discrimination principles that have guided US public institutions since the civil rights movement. Such condemnations are backed by historical evidence and are valid independent of quantitative estimates of causal effects.

The main challenge in estimating the causal effects of different HOLC grades is that the agency's personnel precisely traced neighborhood borders and assigned evaluations. Different HOLC grades within a city closely mirror socioeconomic trends we can observe in the census data. Instead of relying on spatial discontinuity designs, we use an exogenous threshold to determine which cities the agency surveyed. Since we are interested in estimating the effects of different grades, we compare neighborhoods evaluated by HOLC with analogous neighborhoods in control cities. To classify neighborhoods in control cities, we train a random forest algorithm to replicate HOLC grades. Our spatial classification model has an out-of-sample accuracy of more than 90%, and returns predicted maps that are credible replicas of HOLC ones. We then estimate the HOLC grades' effects using the predicted grades with a grouped difference-in-differences design.

We find that this government intervention had adverse effects in areas that received the lowest grade. In the short-term, we find a sizable reduction in property prices and a moderate increase in the percentage of African American residents. We also find a decrease in property values in C neighborhoods. With the help of digitized *Area Description* forms, we can identify neighborhoods with different dimensions of disadvantage and measure the heterogeneous effects of the maps. We found starker effects for all our outcomes of interest in historically Black neighborhoods and a long-term negative impact on homeownership rates for low-income communities. For property prices, we exploit a more granular data source – CoreLogic deeds – to estimate the long-term evolution of the causal effects. We find significant negative effects on property prices in C and D neighborhoods until the early 1980s. This result differs from that of Aaronson et al. (2021b), who found significant effects on property prices until 2010, using a different set of cities and an alternative identification strategy. In our case, the effects of the maps can no longer be detected in the decades following the introduction of legislation targeting residential redlining. Compared to the existing quantitative literature on the maps, our results are more conservative and invite caution in blaming the HOLC for present-day urban inequality in American cities.

We have analyzed a set of maps drafted by the federal government to represent financial risk in the housing market. Since discriminatory practices were widespread at the time, it is not obvious whether a graphical representation of mortgage risk could have affected homeownership rates or property prices. Despite that, the "redlining" maps have become a symbol of racial exclusion, and their consequences are often inflated or misrepresented (Markley, 2024). We show that the D grade had some negative impacts on the general population but stronger adverse effects on neighborhoods and residents that were already disadvantaged. The maps did not break new ground in racial segregation or financial exclusion but reinforced pre-existing patterns. They represent yet another missed opportunity by the federal government to tackle residential segregation and the racial wealth gap. Our findings show that introducing a data-driven information tool can be an additional source of inequality, and they provide a cautionary tale about algorithmic decision-making from the last century.

References

Aaronson, Daniel, Jacob Faber, Daniel Hartley, Bhashkar Mazumder, and Patrick Sharkey (2021a) "The long-run effects of the 1930s HOLC "redlining" maps on place-based measures of economic opportunity and socioeconomic success," *Regional Science and Urban Economics*, 86, 103622, 10.1016/j.regsciurbeco.2020.103622.

Aaronson, Daniel, Daniel Hartley, and Bhashkar Mazumder (2021b) "The Effects of the 1930s HOLC "Redlining" Maps," *American Economic Journal: Economic Policy*, 13 (4), 355–392, 10.1257/pol.20190414.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010) "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105 (490), 493–505, 10.1198/jasa.2009.ap08746.

Abadie, Alberto and Javier Gardeazabal (2003) "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 93 (1), 113–132, 10.1257/000282803321455188.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez (2021) "Automated Linking of Historical Data," *Journal of Economic Literature*, 59 (3), 865–918, 10.1257/jel.20201599.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, Myera Rashid, and Santiago Pérez (2022) "Census Linking Project: 1850-1860 Crosswalk," 10.7910/DVN/KO5J44.

Akbar, Prottoy A., Sijie Li Hickly, Allison Shertzer, and Randall P. Walsh (2022) "Racial Segregation in Housing Markets and the Erosion of Black Wealth," *The Review of Economics and Statistics*, 1–45, 10.1162/rest_a_01276.

Almagro, Milena, Eric Chyn, and Bryan A. Stuart (2023) "Urban Renewal and Inequality: Evidence from Chicago's Public Housing Demolitions," 10.3386/w30838.

Anders, John (2023) "The long run effects of de jure discrimination in the credit market: How redlining increased crime," *Journal of Public Economics*, 222, 104857, 10.1016/j.jpubeco.2023.104857.

Appel, Ian and Jordan Nickerson (2016) "Pockets of Poverty: The Long-Term Effects of Redlining," *SSRN Scholarly Paper*, 10.2139/ssrn.2852856.

Arrow, Kenneth J. (1973) "The Theory of Discrimination," in Ashenfelter, Orley and Albert Rees eds. *Discrimination in Labor Markets*, 3–33: Princeton University Press, `http://www.jstor.org/stable/j.ctt13x10hs.5`.

Athey, Susan and Guido W. Imbens (2019) "Machine Learning Methods That Economists Should Know About," *Annual Review of Economics*, 11 (Volume 11, 2019), 685–725, 10.1146/annurev-economics-080217-053433.

Baradaran, Mehrsa (2017) *The Color of Money: Black Banks and the Racial Wealth Gap*: Harvard University Press, 10.2307/j.ctv24w649g.

Becker, Gary S. (1971) *The Economics of Discrimination*, Economics research studies of the Economics Research Center of the University of Chicago, Chicago, Ill.: University of Chicago Press, 2nd edition.

Bertrand, M. and E. Duflo (2017) "Field Experiments on Discrimination," in Banerjee, Abhijit Vinayak and Esther Duflo eds. *Handbook of Economic Field Experiments*, 1 of Handbook of Field Experiments, 309–393: North-Holland, 10.1016/bs.hefe.2016.08.004.

Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan (2005) "Implicit Discrimination," *American Economic Review*, 95 (2), 94–98, 10.1257/000282805774670365.

Birch, Colin P.D., Sander P. Oom, and Jonathan A. Beecham (2007) "Rectangular and hexagonal grids used for observation, experiment and simulation in ecology," *Ecological Modelling*, 206 (3), 347–359, 10.1016/j.ecolmodel.2007.03.041.

Borusyak, Kirill and Peter Hull (2023) "Nonrandom Exposure to Exogenous Shocks," *Econometrica*, 91 (6), 2155–2185, 10.3982/ECTA19367.

Boustan, Leah Platt (2010) "Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration," *The Quarterly Journal of Economics*, 125 (1), 417–443, `http://www.jstor.org/stable/40506285`.

——— (2012) "Racial Residential Segregation in American Cities," in *The Oxford Handbook of Urban Economics and Planning*: Oxford University Press, 10.1093/oxfordhb/9780195380620.013.0015.

——— (2017) *Competition in the Promised Land: black migrants in northern cities and labor markets*, NBER series on long-term factors in economic development, Princeton Oxford: Princeton University Press.

Breiman, Leo (2001) "Random Forests," *Machine Learning*, 45 (1), 5–32, 10.1023/A:1010933404324.

Chetty, Raj and Nathaniel Hendren (2018) "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects*," *The Quarterly Journal of Economics*, 133 (3), 1107–1162, 10.1093/qje/qjy007.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz (2016) "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *American Economic Review*, 106 (4), 855–902, 10.1257/aer.20150572.

Chyn, Eric (2018) "Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children," *American Economic Review*, 108 (10), 3028–56, 10.1257/aer.20161352.

Chyn, Eric and Lawrence F. Katz (2021) "Neighborhoods Matter: Assessing the Evidence for Place Effects," *Journal of Economic Perspectives*, 35 (4), 197–222, 10.1257/jep.35.4.197.

Coates, Ta-Nehisi (2014) "The Case for Reparations," *The Atlantic*.

Collins, William J. and Katharine L. Shester (2013) "Slum Clearance and Urban Renewal in the United States," *American Economic Journal: Applied Economics*, 5 (1), 239–73, 10.1257/app.5.1.239.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor (1999) "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 107 (3), 455–506, 10.1086/250069.

Derenoncourt, Ellora, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick (2024) "Wealth of Two Nations: The U.S. Racial Wealth Gap, 1860–2020*," *The Quarterly Journal of Economics*, 139 (2), 693–750, 10.1093/qje/qjad044.

Faber, Jacob W. (2020) "We Built This: Consequences of New Deal Era Intervention in America's Racial Geography," *American Sociological Review*, 85 (5), 739–775, 10.1177/0003122420948464.

Fang, Hanming and Andrea Moro (2011) "Theories of Statistical Discrimination and Affirmative Action: A Survey," in Benhabib, Jess, Alberto Bisin, and Matthew O. Jackson eds. *Handbook of Social Economics*, 1, 133–200: North-Holland, 10.1016/B978-0-444-53187-2.00005-X.

Federal Home Loan Bank Board (1937) "Fifth Annual Report of the Federal Home Loan Bank Board, For the Period July 1, 1936 through June 30, 1937."

Federal Housing Administration (1936) *Underwriting Manual: Underwriting Analysis Under Title II, Section 203 of the National Housing Act*: U.S. Government Printing Office.

Fishback, Price, Jonathan Rose, Kenneth A. Snowden, and Thomas Storrs (2022) "New Evidence on Redlining by Federal Housing Programs in the 1930s," *Journal of Urban Economics*, 103462, 10.1016/j.jue.2022.103462.

Fishback, Price V., Alfonso Flores-Lagunes, William C. Horrace, Shawn Kantor, and Jaret Treber (2011) "The Influence of the Home Owners' Loan Corporation on Housing Markets During the 1930s," *The Review of Financial Studies*, 24 (6), 1782–1813, 10.1093/rfs/hhq144.

Fishback, Price V., Jessica LaVoice, Allison Shertzer, and Randall P. Walsh (2023) "The HOLC Maps: How Race and Poverty Influenced Real Estate Professionals' Evaluation of Lending Risk in the 1930s," *The Journal of Economic History*, 83 (4), 1019–1056, 10.1017/S0022050723000475.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther (2022) "Predictably Unequal? The Effects of Machine Learning on Credit Markets," *The Journal of Finance*, 77 (1), 5–47, 10.1111/jofi.13090.

Glaeser, Edward and Jacob Vigdor (2012) "The End of the Segregated Century: Racial Separation in America's Neighborhoods, 1890–2010," *Manhattan Institute for Policy Research*.

Hillier, Amy E. (2003) "Redlining and the Home Owners' Loan Corporation," *Journal of Urban History*, 29 (4), 394–420, 10.1177/0096144203029004002.

——— (2005) "Residential Security Maps and Neighborhood Appraisals: The Home Owners' Loan Corporation and the Case of Philadelphia," *Social Science History*, 29 (2), 207–233, 10.1017/S014555320001292X.

Ihlanfeldt, Keith R. and Benjamin Scafidi (2002) "Black Self-Segregation as a Cause of Housing Segregation: Evidence from the Multi-City Study of Urban Inequality," *Journal of Urban Economics*, 51 (2), 366–390, 10.1006/juec.2001.2249.

Jackson, Kenneth T. (1980) "Race, Ethnicity, and Real Estate Appraisal: The Home Owners Loan Corporation and the Federal Housing Administration," *Journal of Urban History*, 6 (4), 419–452, 10.1177/009614428000600404.

Jones-Correa, Michael (2000) "The Origins and Diffusion of Racial Restrictive Covenants," *Political Science Quarterly*, 115 (4), 541–568, 10.2307/2657609.

Krimmel, Jacob (2020) "Persistence of Prejudice: Estimating the Long Term Effects of Redlining," *SocArXiv*, 10.31235/osf.io/jdmq9.

Krysan, Maria and Reynolds Farley (2002) "The Residential Preferences of Blacks: Do They Explain Persistent Segregation?*," *Social Forces*, 80 (3), 937–980, 10.1353/sof.2002.0011.

Lang, Kevin and Ariella Kahn-Lang Spitzer (2020) "Race Discrimination: An Economic Perspective," *Journal of Economic Perspectives*, 34 (2), 68–89, 10.1257/jep.34.2.68.

Liberman, Andres, Christopher Neilson, Luis Opazo, and Seth Zimmerman (2018) "The Equilibrium Effects of Information Deletion: Evidence from Consumer Credit Markets," *National Bureau of Economic Research Working Paper Series*, No. 25097, 10.3386/w25097.

Light, Jennifer S. (2010) "Nationality and Neighborhood Risk at the Origins of FHA Underwriting," *Journal of Urban History*, 36 (5), 634–671, 10.1177/0096144210365677.

Logan, John R. and Weiwei Zhang (2018) "Developing GIS maps for US cities in 1930 and 1940," in Gregory, Ian, Don DeBats, and Don Lafreniere eds. *The Routledge Companion to Spatial History*, 1st edition, 229–249, Milton Park, Abingdon, Oxon; New York, NY: Routledge, 2018.: Routledge, 10.4324/9781315099781-15.

Logan, Trevon D. and John M. Parman (2017) "The National Rise in Residential Segregation," *The Journal of Economic History*, 77 (1), 127–170, `http://www.jstor.org/stable/26348407`.

Ludwig, Jens and Sendhil Mullainathan (2021) "Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System," *Journal of Economic Perspectives*, 35 (4), 71–96, 10.1257/jep.35.4.71.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles (2021) "National Historical Geographic Information System: Version 16.0," 10.18128/D050.V16.0.

Markley, Scott (2023) "Tabulating Home Owners' Loan Corporation area description sheet data," *Environment and Planning B: Urban Analytics and City Science*, 50 (1), 268–280, 10.1177/23998083221133112.

——— (2024) "Federal 'redlining' maps: A critical reappraisal," *Urban Studies*, 61 (2), 195–213, 10.1177/00420980231182336.

Michney, Todd M. (2019) ""Undesign the Redline."," *Journal of American History*, 106 (3), 708–712, 10.1093/jahist/jaz510.

——— (2022) "How the City Survey's Redlining Maps Were Made: A Closer Look at HOLC's Mortgagee Rehabilitation Division," *Journal of Planning History*, 21 (4), 316–344, 10.1177/15385132211013361.

Mullainathan, Sendhil and Jann Spiess (2017) "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, 31 (2), 87–106, 10.1257/jep.31.2.87.

Nagaraj, Abhishek and Scott Stern (2020) "The Economics of Maps," *Journal of Economic Perspectives*, 34 (1), 196–221, 10.1257/jep.34.1.196.

Nelson, Robert K., LaDale Winling, Richard Marciano, and Nathan Connolly (2023) "Mapping Inequality: Redlining in New Deal America," American Panorama: An Atlas of United States History.

Phelps, Edmund S. (1972) "The Statistical Theory of Racism and Sexism," *The American Economic Review*, 62 (4), 659–661, `http://www.jstor.org/stable/1806107`.

Probst, Philipp, Marvin N. Wright, and Anne-Laure Boulesteix (2019) "Hyperparameters and tuning strategies for random forest," *WIREs Data Mining and Knowledge Discovery*, 9 (3), e1301, 10.1002/widm.1301.

Rambachan, Ashesh, Jon Kleinberg, Jens Ludwig, and Sendhil Mullainathan (2020) "An Economic Perspective on Algorithmic Fairness," *AEA Papers and Proceedings*, 110, 91–95, 10.1257/pandp.20201036.

Rothstein, Richard (2017) *The color of law: a forgotten history of how our government segregated America*, New York ; London: Liveright Publishing Corporation.

Ruggles, Steven, Catherine A. Fitch, Ronald Goeken et al. (2020) "IPUMS Restricted Complete Count Data: Version 2.0," 10.18128/D014.V2.0.

Small, Mario L. and Devah Pager (2020) "Sociological Perspectives on Racial Discrimination," *Journal of Economic Perspectives*, 34 (2), 49–67, 10.1257/jep.34.2.49.

Sood, Aradhya, William Speagle, and Kevin Ehrman-Solberg (2019) "Long Shadow of Racial Discrimination: Evidence from Housing Covenants of Minneapolis," *SSRN Electronic Journal*, 10.2139/ssrn.3468520.

Tach, Laura and Allison Dwyer Emory (2017) "Public Housing Redevelopment, Neighborhood Change, and the Restructuring of Urban Inequality," *American Journal of Sociology*, 123 (3), 686–739, https://www.jstor.org/stable/26546046.

Taylor, Keeanga-Yamahtta (2019) *Race for Profit: How Banks and the Real Estate Industry Undermined Black Homeownership*: University of North Carolina Press, https://www.jstor.org/stable/10.5149/9781469653686_taylor.

Woods, Louis Lee (2012) "The Federal Home Loan Bank Board, Redlining, and the National Proliferation of Racial Lending Discrimination, 1921–1950," *Journal of Urban History*, 38 (6), 1036–1059, 10.1177/0096144211435126.

A. Appendix

A.1 Geocoding Procedure

A.1.1 Address Cleaning

We cleaned addresses in decades between 1910 and 1940 following the procedure outlined in Logan and Zhang (2018). In particular:

- We cleaned street names. Names containing geographic indicators were removed if they were not street names, and dummy variables were created for group quarters (hotels, apartments, convents, hospitals, and group homes). House number information was extracted from street names.

- We cleaned house numbers. If the number found in the house number variable conflicted with the house number extracted from the street variable and the rented home, the house number variable was interpreted to represent an apartment number.

- We interpolated missing street names and house numbers conservatively. For observations on the same census page and within six house numbers from one another, missing streets were given the street name of the prior observation. For rented homes, missing house numbers were given the house number from the previous observation. For owned homes, if the street name was the same as the prior observation, missing house numbers were assigned a value equal to the house number of the preceding observation plus two.

A.1.2 Geocoding

We geocoded the head of each household using ESRI Streetmap Premium 2019. This new-generation locator combines street address routing coordinates and parcel centroids databases to improve the number and quality of the matches. The ESRI algorithm assigns each address-coordinate match a 0-100 score. We include only matches with a score of at least 85 in our analysis. This choice is conservative and reduces measurement errors due to incorrect locations in census households.

A.2 Census Variables Definitions

The following definitions are based on information provided by IPUMS documentation (Ruggles et al., 2020).

Census Variable Definitions	
Homeownership	Indicates whether the housing unit was owned, rather than rented, by its inhabitants.
African American	Based on census race variable. Before 1960, the census enumerator was responsible for categorizing persons and was not specifically instructed to ask the individual about their race.
Property Values	For 1930 and 1940, enumerators consulted with the owners to estimate the sale value of the housing unit.
Rent Prices	Amount of the household's monthly contract rent payment.
First-Generation-Migrant	Whether an individual was foreign-born. Based on the census variable *nativity*.
Unemployed, Men	Indicator defined accordind to census variable *empstat* for men between 18 and 65 years of age.
Radio Ownership	Whether any member of a family or housing unit owns a radio set.
Education Score	Census-built percentage of people in the respondent's occupational category who had completed one or more years of college.
Number of Children	The number of own children residing with each individual.
Population Density	For any neighborhood, the ratio between the area population and surface. The Hexagon surface is fixed at $0.025km^2$.

A.2.1 Income Score Imputation

We estimate a log-wage regression on 1940 census data focusing on men aged 25-55 living in urban areas employed for wages. We regress self-reported wage income on a second-degree polynomial in age, dummies indicating Black Americans, Hispanic Americans, and immigration status, 3-digit occupation, and state of residence indicators. Moreover, we include interactions between each of the race indicators and immigration ones with age polynomial and interactions between the demographic variables with 1-digit occupations and state of residence. The inputed income score in decades 1930, 1920, 1910 is the prediction based on the resulting estimates for men aged 25-55 who are employed for wages in those years.

A.3 Random Forest Training Procedure

We train the random forest algorithm with a hexagon-level dataset ($N = 164,447$) containing all cities mapped by the HOLC. The dataset includes 47 different 1930 census variables. The variables are included at various geographical levels, bringing the total number of training variables to 158. The geographical levels employed in the training procedure are hexagon, hexagon surroundings (500mt and 1500mt), city, and county. The 500-mt local averages are weighted by the number of individuals, while the 1,500-mt ones are weighted by the inverse distance between the hexagons. The random forest results are robust to using a 1km (0.62 mi) radius to define local averages. Results are similar when changing the definitions of the weights. In particular, the variables are:

Random Forest Training Variables

- Share of African Americans
- Share of Women
- Share of Home-Owners
- Share of Population Not Speaking English
- Share Married
- Share of families owning a Radio
- Family Size
- Number of Children
- Age at First Marriage
- House Values
- Rent Prices
- Imputed Income
- Earning Scores
- Educational Scores
- House Distance from City Center
- Neighborhood Population Density
- Labor Force Participation, by gender
- Unemployment Rates, by gender
- Self-Employed and waged employees
- Share of First Generation Immigrants
- Share of Second Generation Immigrants
- Domestic Migrants from the South
- Domestic Migrants from the Mid-West
- Detailed Job Categories shares
- Detailed Country of Birth shares

Before starting the training model, we follow a standard pre-processing machine learning procedure: we impute missing values with the corresponding median values and standardize all our predictors. The random forest is trained on a 75% random sample of the original dataset selected with stratified sampling according to HOLC grades and city population. We set the parameter $m = 53$, which determines the number of variables randomly chosen at each split, following the results of an automated tuning procedure employing model-based optimization (MBO) (Probst et al., 2019). The fraction of observations randomly sampled for each tree is set to 0.876, the minimum node size is 2, and the forest is made of 500 trees.

A.4 List of Cities

The difference-in-differences results are based on the control and treatment group definition outlined in Section 4. As a reminder, treated cities are municipalities surveyed by the HOLC with a population between 40,000 and 60,000. The control group includes cities between 30,000 and 40,000 residents with a distance of at least 30km (18.6 mi) from the nearest mapped city.

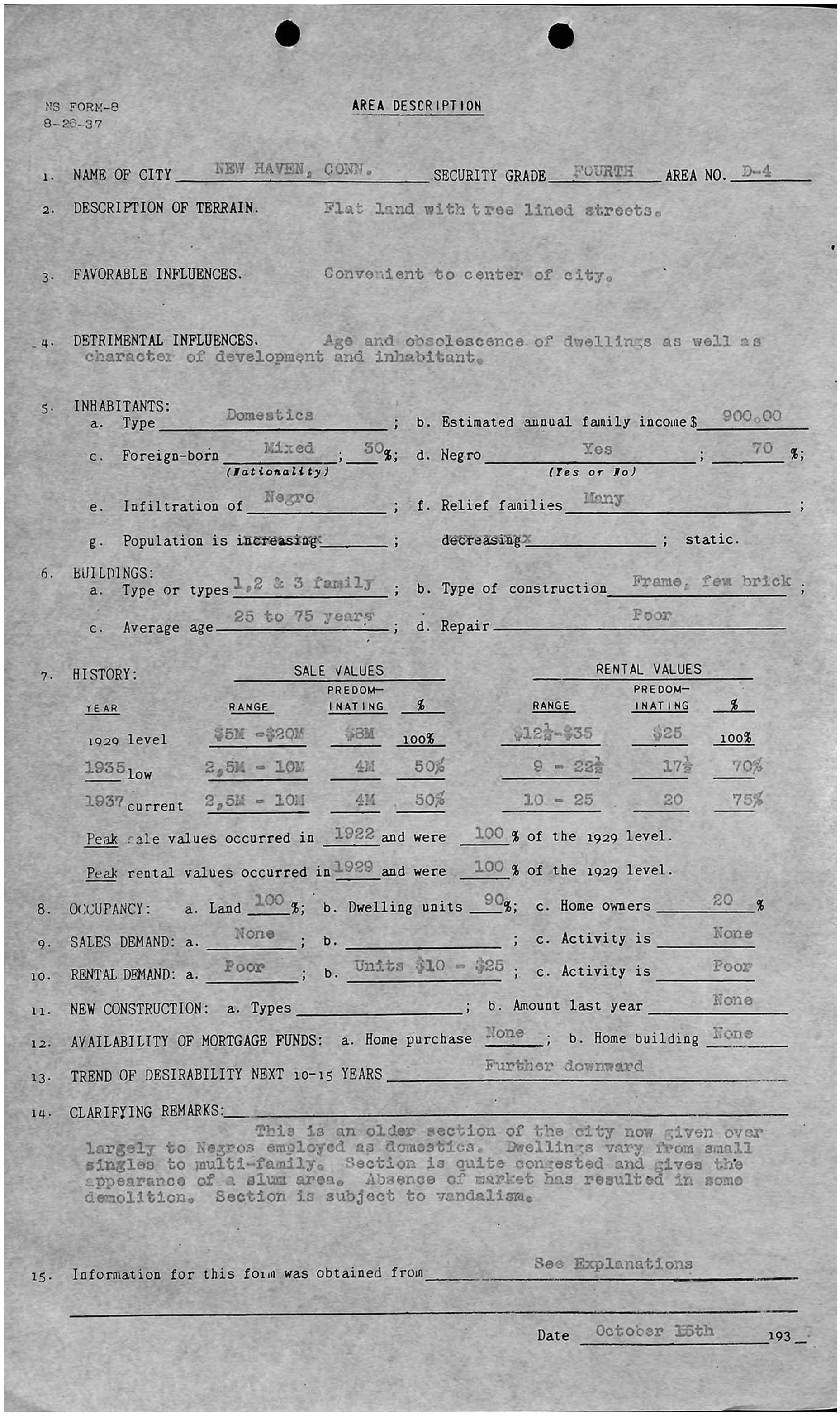
Treated Cities

- Amarillo, TX
- Asheville, NC
- Aurora, IL
- Austin, TX
- Battle Creek, MI
- Bay City, MI
- Beaumont, TX
- Cedar Rapids, IA
- Columbia, SC
- Columbus, GA
- Dubuque, IA
- Durham, NC
- Elmira, NY
- Fitchburg, MA
- Fresno, CA
- Greensboro, NC
- Hamilton, OH
- Haverhill, MA
- Jackson, MI
- Jackson, MS
- Jamestown, NY
- Joliet, IL
- Kalamazoo, MI
- Kenosha, WI
- Lancaster, PA
- Lexington, KY
- Lima, OH
- Lorain, OH
- Lynchburg, VA
- Macon, GA
- Madison, WI
- Muncie, IN
- Muskegon, MI
- New Castle, PA
- Ogden, UT
- Oshkosh, WI
- Perth Amboy, NJ
- Phoenix, AZ
- Pittsfield, MA
- Portsmouth, OH
- Poughkeepsie, NY
- Pueblo, CO
- Saint Petersburg, FL
- San Jose, CA
- Stamford, CT
- Stockton, CA
- Waco, TX
- Warren, OH
- Waterloo, IA
- Wichita Falls, TX
- Woonsocket, RI
- York, PA

Control Cities

- Alton, IL
- Amsterdam, NY
- Anderson, IN
- Auburn, NY
- Baton Rouge, LA
- Bellingham, WA
- Bloomington, IL
- Butte, MT
- Colorado Springs, CO
- Cumberland, MD
- Danville, IL
- Elgin, IL
- Elkhart, IN
- Everett, WA
- Fort Smith, AR
- Green Bay, WI
- Hazleton, PA
- High Point, NC
- Joplin, MO
- Kokomo, IN
- La Crosse, WI
- Laredo, TX
- Lewiston, ME
- Mansfield, OH
- Marion, OH
- Meridian, MS
- Muskogee, OK
- Newark, OH
- Newburgh, NY
- Norristown, PA
- Paducah, KY
- Pensacola, FL
- Port Huron, MI
- Quincy, IL
- Raleigh, NC
- Richmond, IN
- Rome, NY
- San Bernardino, CA
- Santa Ana, CA
- Santa Barbara, CA
- Sheboygan, WI
- Sioux Falls, SD
- Steubenville, OH
- Taunton, MA
- Tucson, AZ
- Watertown, NY
- Waukegan, IL
- Wilmington, NC
- Zanesville, OH

HOLC Area Description, Neighborhood D-4, New Haven, CT



NS FORM-8
8-28-37

AREA DESCRIPTION

1. NAME OF CITY NEW HAVEN, CONN. SECURITY GRADE FOURTH AREA NO. D-4
2. DESCRIPTION OF TERRAIN. Flat land with tree lined streets.
3. FAVORABLE INFLUENCES. Convenient to center of city.
4. DETRIMENTAL INFLUENCES. Age and obsolescence of dwellings as well as character of development and inhabitant.
5. INHABITANTS:
   a. Type Domestics ; b. Estimated annual family income $ 900.00
   c. Foreign-born Mixed (Nationality) ; 30 %; d. Negro Yes (Yes or No) ; 70 %;
   e. Infiltration of Negro ; f. Relief families Many ;
   g. Population is increasing ; decreasing x ; static.
6. BUILDINGS:
   a. Type or types 1, 2 & 3 family ; b. Type of construction Frame, few brick ;
   c. Average age 25 to 75 years ; d. Repair Poor
7. HISTORY:

YEAR	SALE VALUES RANGE	SALE VALUES PREDOM-INATING	SALE VALUES %	RENTAL VALUES RANGE	RENTAL VALUES PREDOM-INATING	RENTAL VALUES %
1929 level	$5M - $20M	$8M	100%	$12½-$35	$25	100%
1935 low	2.5M - 10M	4M	50%	9 - 22½	17½	70%
1937 current	2.5M - 10M	4M	50%	10 - 25	20	75%

Peak sale values occurred in 1922 and were 100 % of the 1929 level.
Peak rental values occurred in 1929 and were 100 % of the 1929 level.

8. OCCUPANCY: a. Land 100 %; b. Dwelling units 90 %; c. Home owners 20 %
9. SALES DEMAND: a. None ; b. ; c. Activity is None
10. RENTAL DEMAND: a. Poor ; b. Units $10 - $25 ; c. Activity is Poor
11. NEW CONSTRUCTION: a. Types ; b. Amount last year None
12. AVAILABILITY OF MORTGAGE FUNDS: a. Home purchase None ; b. Home building None
13. TREND OF DESIRABILITY NEXT 10-15 YEARS Further downward
14. CLARIFYING REMARKS: This is an older section of the city now given over largely to Negros employed as domestics. Dwellings vary from small singles to multi-family. Section is quite congested and gives the appearance of a slum area. Absence of market has resulted in some demolition. Section is subject to vandalism.
15. Information for this form was obtained from See Explanations

Date October 15th 193

Figure A1 — The scan of the *Area Description* for neighborhood D-4 of New Haven, CT, has been provided by *Mapping Inequality* (Nelson et al., 2023).

Comparison of HOLC Digitized Map and its Hexagon Version

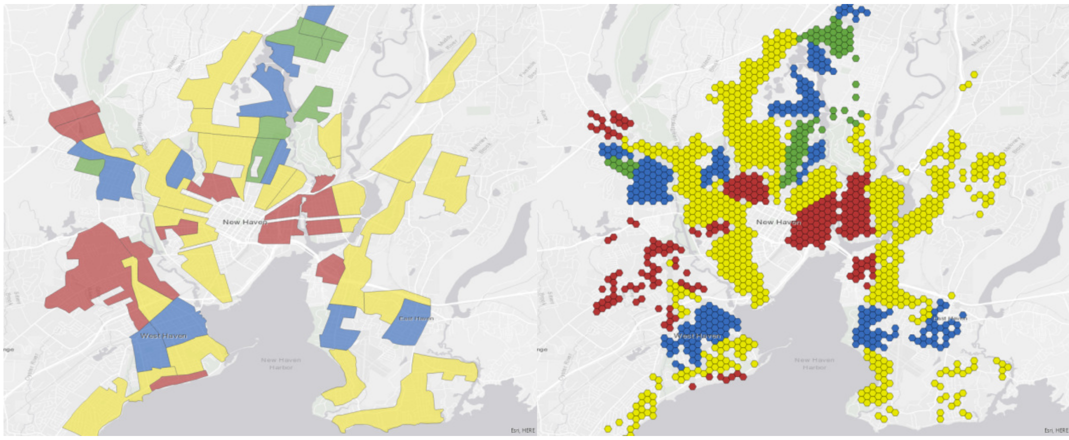
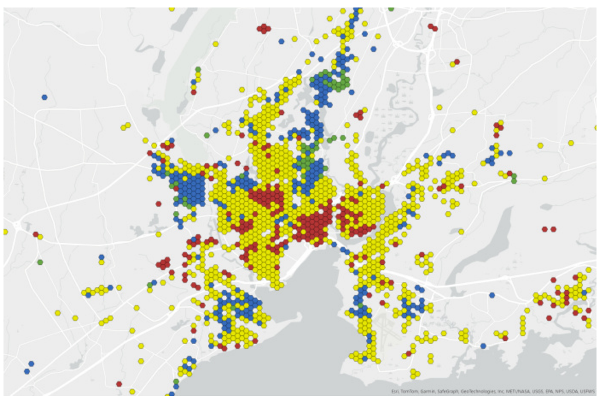
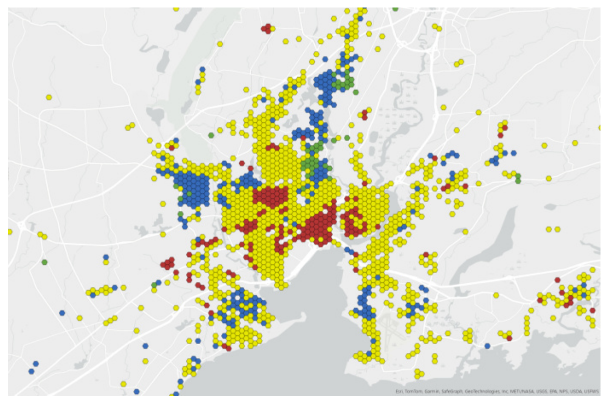


Figure A2 — The digitized version of the *Residential Security Map* of New Haven, CT, shown in the left panel, has been provided by *Mapping Inequality* (Nelson et al., 2023). Details about the definition of the hexagon grid can be found in Section 3.1. The right panel shows our hexagon-level replica of the original HOLC map. All the maps are north-oriented.

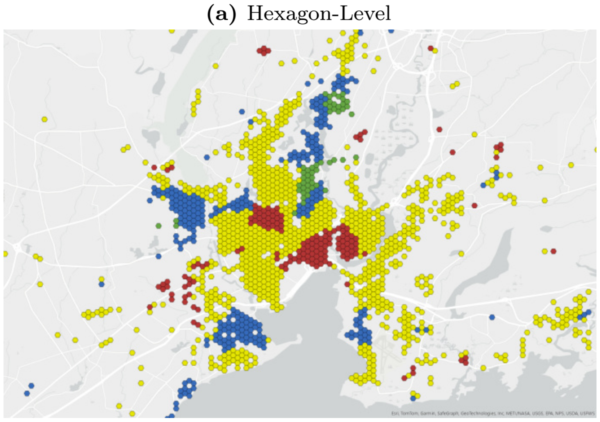
Comparison of Predicted Maps with Different Training Datasets. New Haven, CT



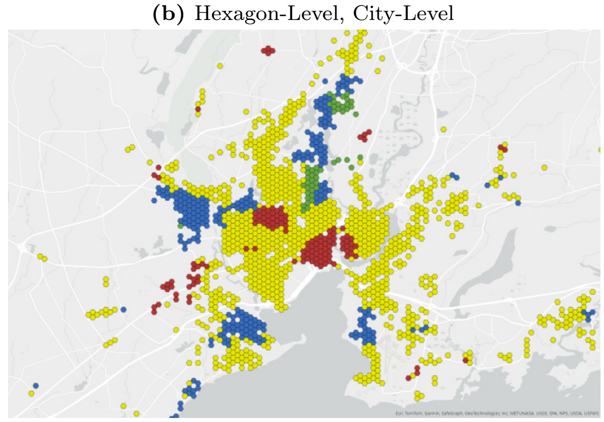
(a) Hexagon-Level



(b) Hexagon-Level, City-Level



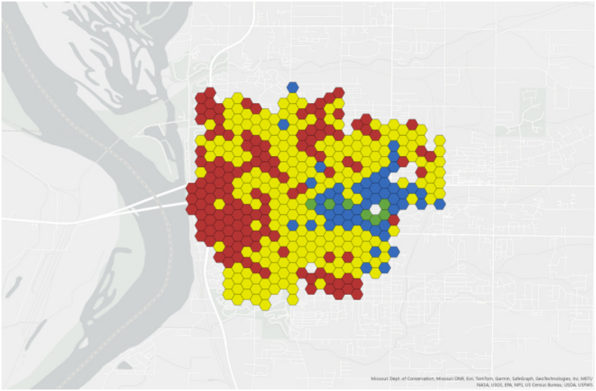
(c) Hexagon-Level and Local Level



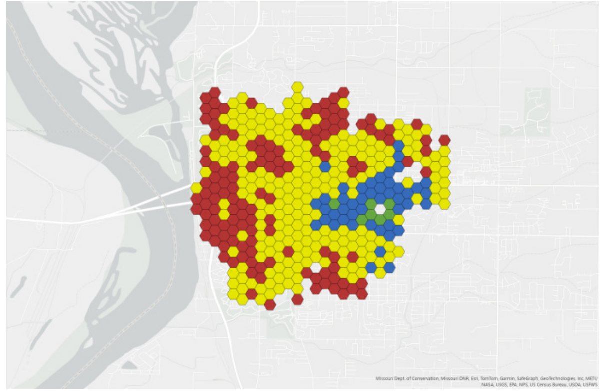
(d) Hexagon-Level, Local-Level, City-Level

Figure A3 — The maps show neighborhoods (hexagons) for New Haven, CT. The colors represent the grade predicted by the random forest algorithm. Different panels show predicted grades for random forests trained on four different sets of variables. The datasets differ in terms of their levels of geographical aggregation, but not because of the variables included. The top-left panel shows predicted grades when only hexagon-level variables are included. The top-right panel adds city-level variables. The bottom-left panel replaces city-level variables with local-level information about the surrounding area. The surrounding area consists of any hexagon whose centroids are within a 500mt (0.31mi) or 1,500mt (0.93mi) radius. The bottom-right panel shows the predicted grades when we include all the previously mentioned variables.

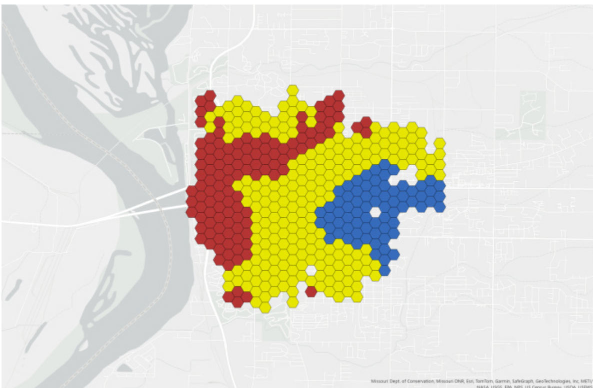
Comparison of Predicted Maps with Different Training Datasets. Quincy, IL



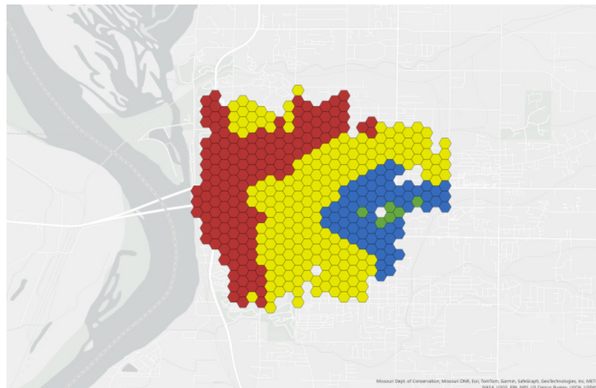
(a) Hexagon-Level



(b) Hexagon-Level, City-Level



(c) Hexagon-Level and Local Level



(d) Hexagon-Level, Local-Level, City-Level

Figure A4 — The maps show neighborhoods (hexagons) for Quincy, IL. The colors represent the grade predicted by the random forest algorithm. The correspondence between colors and grades is: Green=A, Blue= B, Yellow=C, Red=D. The details for the definitions of the four training datasets can be fund in the notes of Appendix Figure A3. All the maps are north-oriented.

Locations of Treatment and Control Cities

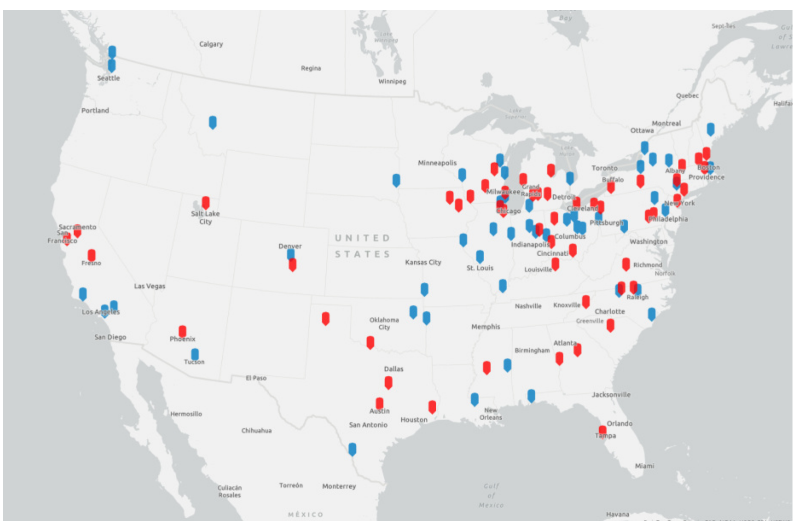


Figure A5 — The figure shows the cities' locations in the control and treatment groups. Control group cities are labeled in blue, while red pins are used for treatment group cities.

1930-1910 Trends in Selected Variables for Neighborhood Grades A and B

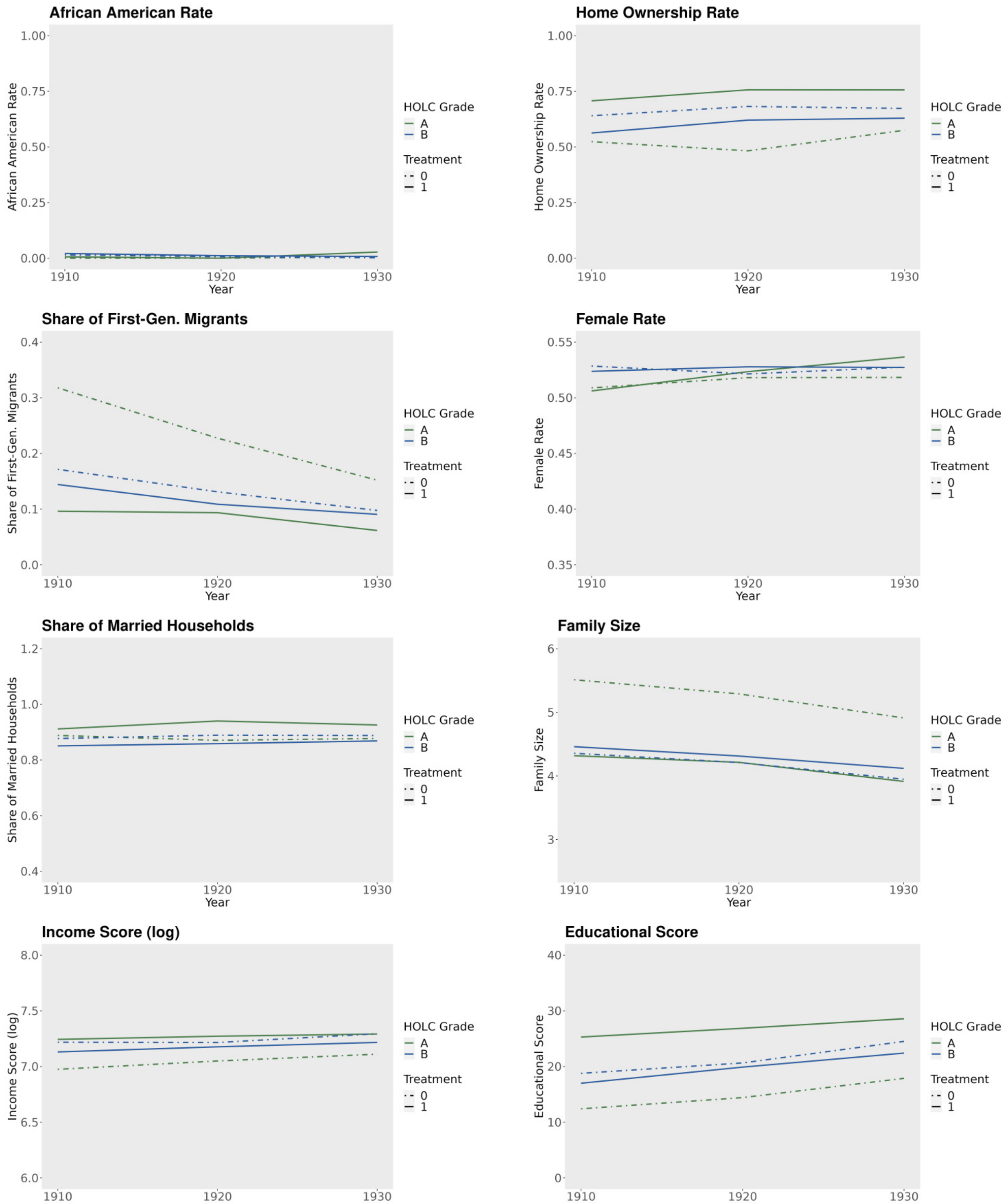


Figure A6 — The figure shows pre-trends for selected variables according to their predicted grade. The point estimates are averages of hexagon-level observations. Green lines indicate the A grade, and blue ones correspond to the B evaluation. Solid lines report averages of treated observations, and dashed ones plot the control group means. The sample includes hexagons in cities with a 1930 population between 30,000 and 60,000, with at least 20 residents in 1930.

Accuracy Levels According to Dataset Definition

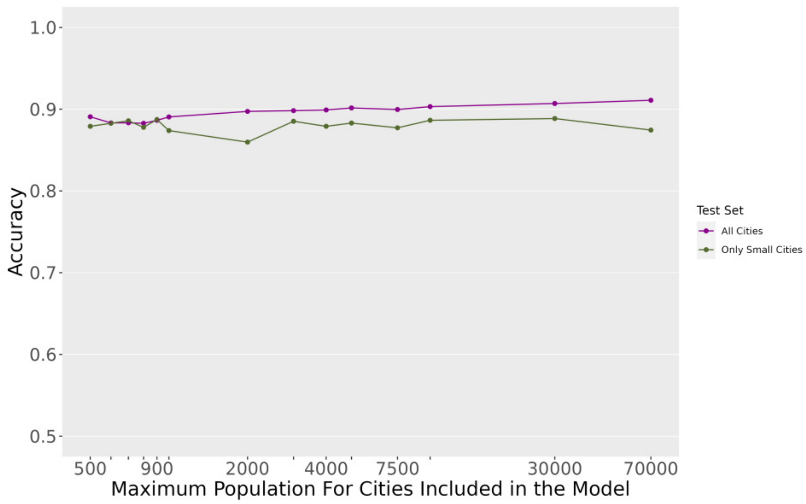
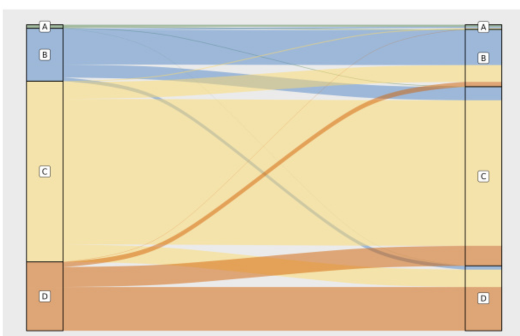
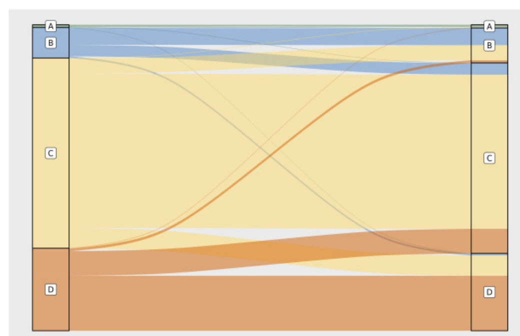


Figure A7 — The figure shows the accuracy level we obtain when we train the random forest classification algorithms with different datasets according to the size of cities we include. The purple line plots the accuracy obtained with a test set defined as a 25% random subsample of the original dataset, selected with stratified sampling according to city population and HOLC grade. The green line shows accuracy levels when we restrict the test set to cities with at most 60,000 residents.

Migration Flows Across Different HOLC Grade, by Treatment and Race

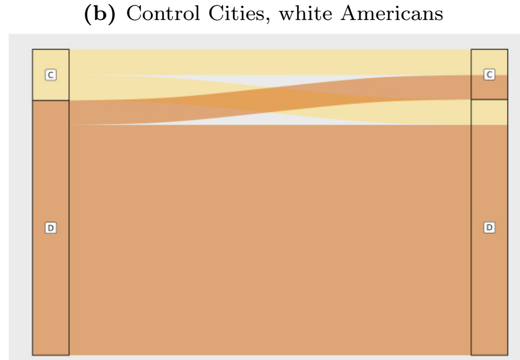


(a) Treated Cities, white Americans



(b) Control Cities, white Americans

(c) Treated Cities, African Americans



(d) Control Cities, African Americans

Figure A8 — The figure shows migration flows across HOLC grades separately for treatment status and race. In each panel, the left column indicates the 1930 individual grade distribution and the right indicates the 1940 corresponding distribution. The graphs are computed with individual panel datasets constructed with Census Linking Project crosswalks (Abramitzky et al., 2022). The streams from left to right indicate individual residential choices between 1930 and 1940. The color indicates the starting grade. The bottom panels report information only for C and D grades because of the minimal percentages of African Americans living in A and B areas.

Table A1 — Neighborhood Distribution according to HOLC Grades

		Proportions			
Spatial Unit	N	A	B	C	D
HOLC Neighborhoods	9,286	11.7%	26.6%	38.2%	23.5%
Hexagons	278,606	7.7%	22.2%	42.2%	27.8%

Notes: The sample of HOLC neighborhoods includes all the shapes digitized by Nelson et al. (2023) for 216 cities. We obtain the sample of hexagons overlaying a regular grid of hexagons with an area of 0.025 km^2 (6 acres) and a side of approximately 100 $mt.$ (330 $ft.$) over the digitized HOLC shapes. The grade of a hexagon is the one occupying the majority of its area. We keep only hexagons whose area is occupied by a single grade for at least 75% of its surface.

Table A2 — Census and CoreLogic coverage of Neighborhoods and Cities.

	Share of Coverage				
	Census Data			CoreLogic Deeds	
Decade	Neighborhood	City	Year	Neighborhood	City
1910	45.6%	98.8%			
1920	56.9%	100.0%			
1930	84.5%	100.0%			
1940	96.6%	100.0%			
1950	5.2%	14.8%			
1960	54.4%	51.9%			
			1965	3.4%	35.21%
1970	83.5%	80.2%	1970	5.6%	38.0%
			1975	8.6%	50.7%
1980	87.4%	93.8%	1980	11.4%	57.7%
			1985	21.1%	76.0%
1990	99.4%	100.0%	1990	35.8%	84.5%
			1995	52.7%	88.7%
2000	100.0%	100.0%	2000	64.4%	91.5%
			2005	76.0%	94.4%
2010	100.0%	100.0%			

Notes: The table reports the percentages of coverage for neighborhoods and cities of interest. The sample includes every hexagon in cities with a 1930 population between 30,000 and 60,000 and at least 20 residents. CoreLogic deeds are binned in 5-year time periods according to their sale year and month.

Table A3 — Random Forest Performance, Small Test, Confusion Matrix

		Data			
		D	C	B	A
Prediction	D	755	61	11	2
	C	92	1187	62	5
	B	9	39	562	24
	A	0	1	6	162
	Accuracy	89.52%			
	Class Sensitivity	88.20	92.16	87.68	83.93
	Prevalence	28.74	43.25	21.52	6.48
	Detection Prevalence	27.84	45.20	21.29	5.67

Notes: The matrix compares the observed and the predicted grades for a test set of observations excluded from the training procedure. The grades are predicted with the random forest algorithm whose results are detailed in Table 3. The test set is a 25% random subsample of the original dataset selected with stratified sampling according to city population and HOLC grade. The test set is restricted to observations from cities with at most 60,000 residents in the 1930 census. See the notes of Table 3 for additional details.

Table A4 — Logit Performance, Confusion Matrix

		Data			
		D	C	B	A
Prediction	D	9037	2627	187	19
	C	4801	14432	4373	326
	B	143	2222	4585	1529
	A	3	33	388	722
	Accuracy	63.35%			
	Class Sensitivity	64.62	74.72	48.10	27.81
	Prevalence	30.78	42.52	20.99	5.71
	Detection Prevalence	26.13	52.68	18.67	2.52

Notes: The matrix compares the observed and the predicted grades for a test set of observations excluded from the training procedure. The test set is a 25% random subsample of the original dataset selected with stratified sampling according to city population and HOLC grade. The level of observation is a neighborhood (hexagon). See Section 3.1 for details about the hexagon definition. The sample includes every hexagon in a mapped city containing at least 20 residents in 1930. A predicted grade is the class predicted by an ordered logit model. See the notes of Table 3 for additional details.

Table A5 — Shares of Local Grades, by Grade and Treatment Status

	Share A	Share B	Share C	Share D
Treatment Group				
A	65.2%	30.2%	4.5%	0.2%
B	4.9%	78.6%	14.1%	2.4%
C	0.3%	8.2%	82.2%	9.2%
D	0.05%	2.5%	17.5%	80.0%
Control Group				
A	78.6%	17.8%	2.0%	1.5%
B	1.5%	79.6%	18.7%	0.1%
C	0.07%	5.8%	88.9%	5.2%
D	0.2%	0.2%	20.4%	79.2%

Notes: The table reports average shares of surrounding grades according to neighborhoods grades and treatment status. Neighborhood surroundings are defined with a 500mt. radius (0.31 miles). The sample includes neighborhoods with at least 20 residents in 1930 in cities with a population between 30,000 and 60,000. See Appendix Section A.4 for a list of cities.

Table A6 — Testing Pre-Trends By Treatment Status

Dependent Variable	A	B	C	D
Panel A: 1930-1920 Trends				
Black	0.015	0.001	0.000	0.006
	(0.013)	(0.004)	(0.002)	(0.006)
Home Owner	-0.062	0.008	-0.011	-0.020
	(0.061)	(0.035)	(0.014)	(0.013)
First Gen. Immigrant	0.047**	0.012	-0.001	0.002
	(0.020)	(0.010)	(0.007)	(0.010)
Female	0.007	-0.007	0.001	-0.002
	(0.008)	(0.005)	(0.002)	(0.004)
Marriage rate	-0.021	-0.022**	-0.006	-0.015*
	(0.014)	(0.011)	(0.005)	(0.008)
Family size	0.042	0.064	0.025	0.015
	(0.130)	(0.057)	(0.037)	(0.050)
Panel B: 1920-1910 Trends				
Black	-0.006	0.002	0.002	0.023**
	(0.005)	(0.007)	(0.003)	(0.011)
Home Owner	0.092**	-0.010	-0.011	-0.008
	(0.039)	(0.016)	(0.010)	(0.015)
First Gen. Immigrant	0.055***	-0.001	0.002	-0.003
	(0.017)	(0.010)	(0.008)	(0.010)
Female	0.012	0.010	0.000	-0.001
	(0.017)	(0.007)	(0.002)	(0.004)
Marriage rate	0.010	0.023*	-0.009*	-0.004
	(0.040)	(0.012)	(0.005)	(0.008)
Family size	0.132	0.024	0.029	-0.068
	(0.256)	(0.079)	(0.036)	(0.078)

Notes: The table reports the coefficients from a set of regressions where the dependent variable is the ten-year change in the variable reported in the left column, and the independent variable is an indicator of treatment status. See Appendix Section A.2 for definitions of census variables in our dataset. The level of observation is a neighborhood (hexagon). The sample includes every hexagon in cities with a 1930 population between 30,000 and 60,000 and at least 20 residents. See Appendix Section A.4 for a list of cities. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01.

Table A7 — 1930 Descriptive Statistics According to HOLC and Predicted Grades

	Grade			
	C		D	
	HOLC	Predicted	HOLC	Predicted
Black	0.02 (0.11)	0.02 (0.08)	0.17 (0.32)	0.19 (0.33)
Home Owner	0.56 (0.27)	0.54 (0.24)	0.43 (0.28)	0.39 (0.24)
Property Value	7,772 (17,564)	8,066 (19,487)	6,220 (27,295)	6,474 (29,256)
Rent	54.62 (176.67)	55.39 (169.75)	41.99 (130.99)	40.98 (106.33)
First Gen Immigrant	0.22 (0.20)	0.23 (0.18)	0.23 (0.24)	0.24 (0.23)
Unemployed, Men	0.10 (0.13)	0.10 (0.10)	0.13 (0.14)	0.13 (0.11)
Owns a Radio	0.58 (0.27)	0.58 (0.23)	0.38 (0.27)	0.35 (0.23)
Neighborhood Population	131.82 (169.53)	149.33 (181.07)	184.92 (294.38)	221.97 (319.29)

Notes: The table reports averages of 1930 census variables according to different classifications. The first two columns compare means between hexagons classified as C by the HOLC with those classified as C by our random forest algorithm. The third and fourth columns do the same for grade D. The level of observation is a neighborhood (hexagon). The sample includes all the hexagons intersecting a HOLC neighborhood digitized by Nelson et al. (2023) in 216 maps. Standard deviations are reported in parentheses.

Table A8 — 1930 Descriptive Statistics According to Predicted Grades

	Predicted Grade			
	C		D	
	Correct	Wrong	Correct	Wrong
Panel A: 1930 Levels				
Black	0.02 (0.09)	0.01 (0.08)	0.20 (0.34)	0.11 (0.25)
Home Owner	0.54 (0.23)	0.56 (0.25)	0.39 (0.24)	0.44 (0.25)
Property Value	7,721 (15,569)	9,267 (29,256)	6,371 (30,300)	7.403 (17,243)
Rent	54.57 (160.82)	58.36 (198.49)	40.71 (107.72)	43.42 (92.61)
First Gen Immigrant	0.22 (0.18)	0.24 (0.19)	0.23 (0.23)	0.28 (0.24)
Unemployed, Men	0.10 (0.10)	0.10 (0.11)	0.13 (0.11)	0.13 (0.12)
Owns a Radio	0.58 (0.23)	0.60 (0.23)	0.34 (0.23)	0.40 (0.24)
Neighborhood Population	150.09 (167.16)	146.69 (222.75)	223.80 (323.20)	205.64 (281.37)
Panel B: 1930-1920 Trends				
Black	-0.005 (0.07)	-0.001 (0.07)	0.03 (0.16)	0.01 (0.19)
Home Owner	-0.01 (0.24)	0.03 (0.23)	-0.02 (0.25)	-0.03 (0.25)
Income Score	0.01 (0.18)	0.03 (0.19)	-0.02 (0.22)	-0.04 (0.25)
First Gen Immigrant	-0.02 (0.14)	-0.03 (0.15)	-0.05 (0.17)	-0.03 (0.14)

Notes: The table reports averages and trends of 1930 census variables according to different classifications. The first two columns compare means between hexagons classified as C by our classification model. The first column reports averages for hexagons whose observed grade is also C (correct prediction), while the second refers to neighborhoods with a HOLC grade other than C (wrong prediction). The third and fourth columns do the same for grade D. The level of observation is a neighborhood (hexagon). The sample includes all the hexagons intersecting a HOLC neighborhood digitized by Nelson et al. (2023) in 216 maps. Standard deviations are reported in parentheses.

Table A9 — Short-term Diff-in-Diff Results. No Grade Heterogeneity

	Dependent variables			
	African American Percentage	Homeownership Rates	Property Values (Logs)	Rent Prices
DiD	0.005 (0.029)	-0.004 (0.023)	-0.052 (0.089)	4.2 (11.4)
$\bar{Y}$	0.08	0.53	8.40	43.5
N	76,790	76,763	76,583	74,631

Notes: The table reports difference-in-differences coefficients we obtain estimating equation (1) for four different outcomes with neighborhood-level observations. Standard errors, in parentheses, are clustered at the city-year level. The sample includes neighborhoods in cities with a 1930 population between 30,000 and 60,000, with at least 20 residents. See Appendix Section A.4 for a list of cities. Significance: * 0.10, ** 0.05, *** 0.01.

Table A10 — Short-Term Difference-in-Differences Results, Clustered S.E.

	Dependent Variables			
	African American Percentage	Home Ownership	Property Values (Logs)	Rent Prices
DiD_A	0.002	0.080**	0.101	34.1
	(0.010)	(0.037)	(0.062)	(102.6)
DiD_B	0.003	0.010	0.000	0.4
	(0.003)	(0.013)	(0.028)	(24.8)
DiD_C	0.001	-0.010	-0.056	6.9
	(0.001)	(0.011)	(0.037)	(12.6)
DiD_D	0.013**	-0.012	-0.081	-12.3
	(0.005)	(0.011)	(0.056)	(8.28)

Notes: The table reports difference-in-differences coefficients we obtain estimating equation (1) by grade for four different outcomes with neighborhood-level observations. Each row contains the DiD coefficients for a given grade. Standard errors, in parentheses, are clustered at the city level. The regression specification is analogous to the one in Table 4. The sample includes neighborhoods in cities with a 1930 population between 30,000 and 60,000, with at least 20 residents. See Appendix Section A.4 for a list of cities. Significance: * 0.10, ** 0.05, *** 0.01.

Table A11 — Short-term Difference-in-Differences Results. Property Values

	Dependent variable: Property Value			
	(1)	(2)	(3)	(4)
DiD_A	175	-370	-454	-244
$\bar{Y}^A = 13577$	(3,264)	(1,005)	(1,016)	(965)
DiD_B	31	30	-15	13
$\bar{Y}^B = 8592$	(711)	(240)	(234)	(228)
DiD_C	-421	-433*	-463**	-459**
$\bar{Y}^C = 6203$	(439)	(228)	(229)	(231)
DiD_D	-707	-710***	-684***	-652**
$\bar{Y}^D = 4307$	(494)	(257)	(255)	(253)
City Fixed. Eff.		X	X	X
Spatial Unit Controls			X	X
Local Area Controls				X

Notes: The table reports difference-in-differences coefficients obtained estimating equation (1) by grade. Each row contains the DiD coefficients for a given grade. The regressions are estimated with neighborhood-level observations. See the Notes of Table 5 for additional details. Significance: * 0.10 ** 0.05 *** 0.01.

Table A12 — Short-term Difference-in-Differences Results. African American Percentage

	Dependent variable: African American Percentage			
	(1)	(2)	(3)	(4)
DiD_A	0.004	0.002	0.002	0.002
$\bar{Y}^A = 0.015$	(0.017)	(0.007)	(0.007)	(0.007)
DiD_B	0.002	0.003	0.003	0.003
$\bar{Y}^B = 0.013$	(0.005)	(0.002)	(0.002)	(0.002)
DiD_C	0.001	0.000	0.001	0.001
$\bar{Y}^C = 0.026$	(0.008)	(0.001)	(0.001)	(0.001)
DiD_D	0.014	0.012***	0.012***	0.013***
$\bar{Y}^D = 0.212$	(0.076)	(0.005)	(0.004)	(0.004)
City Fixed. Eff.		X	X	X
Spatial Unit Controls			X	X
Local Area Controls				X

Notes: The table reports difference-in-differences coefficients obtained estimating equation (1) by grade. Each row contains the DiD coefficients for a given grade. The regressions are estimated with neighborhood-level observations. See the Notes of Table 5 for additional details. Significance: * 0.10 ** 0.05 *** 0.01.

Table A13 — Short-term Difference-in-Differences Results. Home Ownership

	Dependent variable: Home Ownership			
	(1)	(2)	(3)	(4)
DiD_A	0.074	0.078***	0.077***	0.080***
$\bar{Y}^A = 0.695$	(0.117)	(0.025)	(0.027)	(0.027)
DiD_B	0.012	0.012	0.014	0.010
$\bar{Y}^B = 0.655$	(0.039)	(0.009)	(0.009)	(0.009)
DiD_C	-0.011	-0.012	-0.012	-0.010
$\bar{Y}^C = 0.547$	(0.025)	(0.008)	(0.008)	(0.008)
DiD_D	-0.008	-0.007	-0.007	-0.012*
$\bar{Y}^D = 0.424$	(0.037)	(0.008)	(0.008)	(0.007)
City Fixed. Eff.		X	X	X
Spatial Unit Controls			X	X
Local Area Controls				X

Notes: The table reports difference-in-differences coefficients obtained estimating equation (1) by grade. Each row contains the DiD coefficients for a given grade. The regressions are estimated with neighborhood-level observations. See the Notes of Table 5 for additional details. Significance: * 0.10 ** 0.05 *** 0.01.

Table A14 — Short-term Difference-in-Differences Results. Rent Prices

	Dependent variable: Rent Prices			
	(1)	(2)	(3)	(4)
DiD_A	2.7	17.6	30.6	34.1
$\bar{Y}^A = 88.3$	(76.7)	(76.8)	(76.0)	(75.4)
DiD_B	-2.5	-1.7	0.9	0.5
$\bar{Y}^B = 60.4$	(27.2)	(18.1)	(17.6)	(17.6)
DiD_C	7.4	7.2	6.9	6.9
$\bar{Y}^C = 42.9$	(13.2)	(8.9)	(9.0)	(9.0)
DiD_D	-12.4	-12.6**	-12.5**	-12.3**
$\bar{Y}^D = 32.6$	(9.1)	(5.8)	(5.9)	(5.9)
City Fixed. Eff.		X	X	X
Spatial Unit Controls			X	X
Local Area Controls				X

Notes: The table reports difference-in-differences coefficients obtained estimating equation (1) by grade. Each row contains the DiD coefficients for a given grade. The regressions are estimated with neighborhood-level observations. See the Notes of Table 5 for additional details. Significance: * 0.10 ** 0.05 *** 0.01.

Table A15 — Long-Term Difference-in-Differences Results, by Grade. Census Data

	Dependent variables			
	Homeownership Rates		African American Percentage	
	C	D	C	D
DiD_{60}	0.005	-0.020	-0.021	0.027
	(0.015)	(0.019)	(0.016)	(0.024)
DiD_{70}	-0.002	-0.002	0.005	0.037*
	(0.015)	(0.019)	(0.015)	(0.022)
DiD_{80}	0.000	-0.011	0.019	0.060***
	(0.013)	(0.018)	(0.015)	(0.023)
DiD_{90}	0.006	-0.004	0.029*	0.055**
	(0.012)	(0.018)	(0.015)	(0.023)
DiD_{00}	0.004	-0.008	0.032*	0.051**
	(0.012)	(0.017)	(0.017)	(0.022)
DiD_{10}	-0.002	0.005	0.020	0.022
	(0.012)	(0.018)	(0.019)	(0.024)
N	115,046	68,103	115,232	68,261
R^2	0.244	0.300	0.442	0.512

Notes: The table reports difference-in-differences coefficients obtained estimating equation (2) by grade. Each row contains the DiD coefficients for a given grade in the corresponding year. In the reported specification, we replace the indicator for treatment, which is assigned at the city level, with a city fixed-effect. The control list includes geographic coordinates, their squares, and state-specific quadratic time trends. The sample consists of neighborhoods in cities with a 1930 population between 30,000 and 60,000, with at least 20 residents. See Appendix Section A.4 for a list of cities. The data source for post-1940 outcomes is NHGIS; see Section 3.3 for details. Standard errors, in parentheses, are clustered at the city-year level. Significance:* 0.10 ** 0.05 *** 0.01.

Table A16 — Long-Term Difference-in-Differences Results, by Grade. CoreLogic

	Dependent variable: Property Values (Logs)			
	Grade			
	A	B	C	D
DiD_{65}	.	0.099	-0.325***	-0.543
	(.)	(0.089)	(0.119)	(0.366)
DiD_{70}	-0.051	-0.007	-0.604***	-1.070***
	(0.197)	(0.170)	(0.176)	(0.182)
DiD_{75}	0.181	0.034	-0.246**	-0.856***
	(0.169)	(0.131)	(0.114)	(0.178)
DiD_{80}	0.170	0.090	-0.664**	-1.080***
	(0.169)	(0.171)	(0.259)	(0.262)
DiD_{85}	0.231	0.701***	0.002	-0.339
	(0.542)	(0.252)	(0.227)	(0.413)
DiD_{90}	0.651**	0.415**	-0.119	-0.043
	(0.316)	(0.184)	(0.195)	(0.300)
DiD_{95}	0.465*	0.186	-0.282	-0.185
	(0.272)	(0.162)	(0.179)	(0.278)
DiD_{00}	0.705**	0.279	-0.100	-0.121
	(0.297)	(0.188)	(0.180)	(0.273)
DiD_{05}	0.480*	0.130	-0.147	-0.159
	(0.248)	(0.190)	(0.184)	(0.259)
DiD_{10}	0.806***	0.157	-0.210	-0.221
	(0.305)	(0.190)	(0.188)	(0.253)
N	2,201	24,024	91,730	41,434
R^2	0.193	0.067	0.063	0.139

Notes: The table reports difference-in-differences coefficients obtained estimating equation (2) by grade. The outcome, property value, is transformed in log form. Each row contains the DiD coefficients for a given grade in the corresponding year. The regressions are estimated with neighborhood-level observations. The sample includes neighborhoods with at least 20 residents in cities with a 1930 population between 30,000 and 60,000. See Appendix Section A.4 for a list of cities. The data source for post-1940 outcomes is CoreLogic, see Section 3.4 for details. Standard errors, in parentheses, are clustered at the city-year level. Significance: * 0.10 ** 0.05 *** 0.01.

Table A17 — Long-Term Difference-in-Differences Results, by Grade. CoreLogic

	Dependent variable: Property Values			
	Grade			
	A	B	C	D
DiD_{65}	.	1,847	-8,838**	-17,372***
	(.)	(3,903)	(3,708)	(5,652)
DiD_{70}	3,714	1,145	-12,658***	-15,846***
	(20,069)	(4,582)	(4,709)	(3,882)
DiD_{75}	14,570	-53	-9,996***	-12,632***
	(19,051)	(4,925)	(3,173)	(3,387)
DiD_{80}	16,115	-4,545	-16,781**	-17,804***
	(19,051)	(6,420)	(7,844)	(6,346)
DiD_{85}	11,370	18,945**	976	-5,189
	(25,567)	(8,874)	(9,150)	(10,674)
DiD_{90}	15,698	10,902	-8,937	-6,792
	(19,715)	(6,875)	(6,990)	(6,756)
DiD_{95}	16,850	586	-12,709**	-6,346
	(19,837)	(6,086)	(5,828)	(6,061)
DiD_{00}	38,087*	6,992	-7,633	-3,347
	(21,044)	(8,820)	(7,277)	(9,018)
DiD_{05}	36,969*	4,526	-4,669	3,105
	(21,272)	(11,269)	(9,455)	(9,691)
DiD_{10}	35,815*	1,525	-5,420	-1,593
	(20,995)	(7,742)	(7,016)	(7,947)
N	2,024	22,538	86,855	37,945
R^2	0.124	0.103	0.060	0.109

Notes: The table reports difference-in-differences coefficients obtained estimating equation (2) by grade. The table is analogous to Appendix Table A16, except that the outcome, property value, is in levels. See the notes of Appendix Table A16 for more details. Significance: * 0.10 ** 0.05 *** 0.01.

Table A18 — Short-Term Difference-in-Differences Results, Grade D

	DiD_D	
	Afr. Am. and Low Income Neighb.	Immigrant and Low Income Neighb.
Afr. Am. Perc.	0.056**	0.004
	(0.026)	(0.006)
Home Ownership	-0.118***	-0.014
	(0.043)	(0.012)
Property Values (Logs)	0.137	0.066
	(0.079)	(0.078)
$\bar{Y}^D$, Afr. Am.	0.726	0.083
$\bar{Y}^D$, Home Own.	0.344	0.488
$\bar{Y}^D$, Prop. Val.	7.75	8.16
N^D	1,078	1,424

Notes: The table reports difference-in-differences coefficients we obtain estimating equation (1) for three different outcomes with neighborhood-level D-grade observations. Different columns show the results employing different samples. More details about the sample definitions and the regressions estimates can be found in the notes of Table 8. Significance: * 0.10, ** 0.05, *** 0.01.

Table A19 — Long-Term Difference-in-Differences Results, Grade D

	DiD_D - African American Percentage			
	All Neighborhoods	Afr. Am. Neighborhoods	Immigrant Neighborhoods	Low Income Neighborhoods
DiD_{60}	0.027 (0.024)	0.141*** (0.043)	-0.023 (0.049)	-0.062* (0.033)
DiD_{70}	0.037* (0.022)	0.103*** (0.030)	-0.008 (0.049)	-0.051 (0.033)
DiD_{80}	0.060*** (0.023)	0.096*** (0.032)	0.020 (0.054)	-0.024 (0.033)
DiD_{90}	0.055** (0.023)	0.074*** (0.027)	0.026 (0.053)	-0.037 (0.032)
DiD_{00}	0.051** (0.022)	0.074*** (0.028)	0.021 (0.047)	-0.044 (0.032)
DiD_{10}	0.022 (0.024)	0.033 (0.034)	-0.010 (0.045)	-0.089*** (0.034)
N	68,261	15,756	12,395	22,736
R^2	0.512	0.294	0.623	0.631

Notes: The table reports difference-in-differences coefficients we obtain estimating equation (2) for African American percentages with neighborhood-level D-grade observations. Different columns show the results employing different samples. See the notes of Table 8 for details about the subsample definitions. Each row contains the DiD coefficients for a given census decade. The regression specification is analogous to the one in Appendix Table A15. More details about the regressions estimates can be found in the notes of Table A15. Significance: * 0.10, ** 0.05, *** 0.01.

Table A20 — Short-Term Difference-in-Differences Results, Actual Grades

	Dependent Variables			
	African American Percentage	Home Ownership	Property Values (Logs)	Rent Prices
DiD_A	0.005 (0.004)	0.064** (0.025)	0.148*** (0.039)	34.3 (72.5)
DiD_B	0.000 (0.002)	0.020** (0.009)	-0.011 (0.022)	-10.9 (16.9)
DiD_C	-0.001 (0.001)	-0.008 (0.008)	-0.037 (0.026)	8.9 (9.2)
DiD_D	0.014*** (0.004)	-0.021** (0.007)	-0.008 (0.035)	-7.8 (6.0)

Notes: The table reports difference-in-differences coefficients we obtain estimating equation (1) by grade for four different outcomes with neighborhood-level observations. Each row contains the DiD coefficients for a given grade. The grade for treated observations is assigned by HOLC instead of the ones predicted by the random forest. The table is analogous to Table 4; see its notes for additional details. Significance: * 0.10, ** 0.05, *** 0.01.

Table A21 — Long-Term Difference-in-Differences Results, Actual Grades. Census Data

	Dependent variable:			
	Homeownership Rates		African American Percentage	
	C	D	C	D
DiD_{60}	.	-0.016	.	0.000
	(.)	(0.015)	(.)	(0.023)
DiD_{70}	-0.020	-0.008	0.033**	0.019
	(0.013)	(0.013)	(0.014)	(0.018)
DiD_{80}	-0.020	-0.017	0.051***	0.039**
	(0.012)	(0.012)	(0.013)	(0.018)
DiD_{90}	-0.015	-0.015	0.059***	0.035**
	(0.012)	(0.011)	(0.014)	(0.017)
DiD_{00}	-0.019	-0.011	0.060***	0.028*
	(0.012)	(0.010)	(0.015)	(0.016)
DiD_{10}	-0.022*	.	0.044**	.
	(0.013)	(.)	(0.017)	(.)
N	85,733	53,996	85,880	54,128
R^2	0.284	0.360	0.480	0.596

Notes: The table reports difference-in-differences coefficients obtained estimating equation (2) by grade. Each row contains the DiD coefficients for a given grade in the corresponding year. The grade for treated observations is assigned by HOLC instead of the ones predicted by the random forest. The table is analogous to Appendix Table A15; see its notes for additional details. Significance: * 0.10, ** 0.05, *** 0.01.

Table A22 — Long-Term Difference-in-Differences Results, Actual Grades. CoreLogic

	Dependent variable: Property Values			
	Grade			
	A	B	C	D
DiD_{65}	.	6,024*	-10,631***	-15,959***
	(.)	(3,494)	(3,813)	(5,646)
DiD_{70}	15.375	5,270	-15,375***	-11,313*
	(17,591)	(4,401)	(4,110)	(6,412)
DiD_{75}	12,063	4,069	-10,944***	-11,533***
	(17,625)	(4,544)	(3,343)	(3,830)
DiD_{80}	11,414	-3,334	-18,043**	-14,110**
	(17,894)	(6,307)	(8,159)	(6,515)
DiD_{85}	25,559	19,173**	-589	675
	(19,528)	(7,603)	(9,374)	(9,227)
DiD_{90}	36,168**	12,489*	-9,005	-1,264
	(18,142)	(6,680)	(7,076)	(6,610)
DiD_{95}	27,502	1,882	-12,418**	-2,661
	(18,220)	(5,691)	(5,905)	(6,053)
DiD_{00}	54,502***	5,078	-7,581	-215
	(18,967)	(7,790)	(7,469)	(8,759)
DiD_{05}	54,716***	3,997	-3,954	5,165
	(20,419)	(10,361)	(9,590)	(9,293)
DiD_{10}	49,009**	2,423	-5,147	-127
	(19,595)	(7,278)	(7,115)	(7,659)
N	6,168	27,249	76,502	37,620
R^2	0.122	0.110	0.060	0.171

Notes: The table reports difference-in-differences coefficients obtained estimating equation (2) by grade. Each row contains the DiD coefficients for a given grade in the corresponding year. The grade for treated observations is assigned by HOLC instead of the ones predicted by the random forest. The table is analogous to Appendix Table A17; see its notes for additional details. Significance: * 0.10, ** 0.05, *** 0.01.

Table A23 — Short-Term Diff-in-Diff Results, by Grade. Alternative Treatment Group

	Dependent Variables			
	African American Percentage	Home Ownership	Property Values (Logs)	Rent Prices
DiD_A	-0.001	0.049*	0.004	33.7
	(0.006)	(0.028)	(0.060)	(77.6)
DiD_B	0.002	0.018*	-0.026	-8.7
	(0.002)	(0.009)	(0.024)	(17.4)
DiD_C	0.000	-0.012	-0.069***	4.5
	(0.001)	(0.008)	(0.026)	(9.5)
DiD_D	0.011***	0.001	-0.114***	-5.8
	(0.004)	(0.008)	(0.042)	(7.0)

Notes: The table reports difference-in-differences coefficients we obtain estimating equation (1) by grade for four different outcomes with neighborhood-level observations. Each row contains the DiD coefficients for a given grade. In this case, the treated group consists of observations from cities with a 1930 population between 40,000 and 50,000 (instead of 60,000). The table is analogous to Table 4; see its notes for additional details. Significance: * 0.10, ** 0.05, *** 0.01.

Table A24 — Short-Term Diff-in-Diff Results, by Grade. Restricted City Sample

	Dependent Variables			
	African American Percentage	Home Ownership	Property Values (Logs)	Rent Prices
DiD_A	0.005	0.068**	0.131***	-18.2
	(0.010)	(0.030)	(0.043)	(66.4)
DiD_B	0.005*	0.014	-0.011	11.7
	(0.003)	(0.011)	(0.024)	(17.3)
DiD_C	0.000	-0.013	-0.072**	-2.2
	(0.001)	(0.010)	(0.030)	(11.6)
DiD_D	0.013***	-0.006	-0.097**	-5.8
	(0.004)	(0.009)	(0.045)	(7.3)

Notes: The table reports difference-in-differences coefficients we obtain estimating equation (1) by grade for four different outcomes with neighborhood-level observations. Each row contains the DiD coefficients for a given grade. In this case, the treated group is restricted to observations from cities with a 1930 population between 40,000 and 50,000 (instead of 60,000) that are not suburbs of a bigger HOLC-mapped city. The control group is limited to cities whose distance from a treated city is at least 30km (18.6 mi). The table is analogous to Table 4; see its notes for additional details. Significance: * 0.10, ** 0.05, *** 0.01.

Table A25 — Short-Term Difference-in-Differences Results, by Grade. Alternative Outcomes

	Dependent Variables				
	Female	Marriage Rate	Age of Marriage	Out of Labor Force	Unemployed, Men
DiD_A	0.005	-0.008	-0.130	0.003	-0.024**
	(0.005)	(0.008)	(0.414)	(0.006)	(0.011)
DiD_B	0.000	0.018***	0.260	-0.002	0.001
	(0.002)	(0.004)	(0.165)	(0.003)	(0.004)
DiD_C	0.001	-0.003	0.075	-0.001	0.004
	(0.001)	(0.002)	(0.086)	(0.002)	(0.005)
DiD_D	0.002	0.002	0.069	-0.001	0.004
	(0.002)	(0.003)	(0.123)	(0.003)	(0.008)

Notes: The table reports difference-in-differences coefficients we obtain estimating equation (1) by grade for several different outcomes with neighborhood-level observations. Each row contains the DiD coefficients for a given grade. The table is analogous to Table 4; see its notes for additional details. See Appendix Section A.2 for definitions of census variables in our dataset. Significance: * 0.10, ** 0.05, *** 0.01.