Indexing Cards: Linking FHA and VA Borrowers to the Census

Katherine A. Thomas, Wenfei Xu, Thomas Storrs, and Jacob W. Faber

November 21, 2024

Abstract

The Federal Housing Administration (FHA) and Veterans' Administration (VA) mortgage programs played a key role in enabling the majority of Americans to become homeowners by the mid-20th century. While scholars have long argued that these programs' beneficiaries were disproportionately white, evidence of the identities of these early borrowers is both scant and disaggregated. We fill this void by linking records from the Reconstruction Finance Corporation Mortgage Company (RFCMC) identifying 120,000 FHA-insured and VA-guaranteed loans from 1935 to 1947 to the 1940 and 1950 full-count Censuses. The unbalanced nature of the RFCMC set allows us to concentrate on particular geographies, programs, and periods. We find, in line with expectations, that the FHA had few Black borrowers in South Carolina. Moreover, we argue that our numbers establish a likely upper bound on the proportion of Black borrowers given that RFCMC acted as a purchaser of last-resort for marginal lenders. Our results shed new light on the federal government's role in racially discriminatory mortgage practices.

Introduction

A wealth of scholarship across the social sciences has explored connections between New Deal Era housing policies and contemporary racial and spatial inequalities (Aaronson, Hartley, and Mazumder 2021; Faber 2020; Hynsjö and Perdoni 2024; Jackson 1980). While much of this work has focused on the Home Owners Loan Corporation (HOLC) because their infamous "Residential Security Maps" have been digitized, little is known about the Federal Housing Administration (FHA) and Veterans' Administration (VA) mortgage programs–both of which were substantially larger and operate up to the present. Together, these three programs were largely responsible for creating the American homeownership society by making mortgage credit affordable, widely available in a standardized form, and low risk for financial institutions to grant. Between 1930 and 1970, the US homeownership rate increased from 45% to 64%.[1]

Concurrent with the emergence of this new way of living, metropolitan areas experienced widespread suburbanization and a dramatic increase in residential segregation–especially between Black and white households (Connolly 2014; Freund 2007; Sugrue 2005). While scholars have long argued the New Deal policies were segregationist by design and largely responsible for shaping America's disparate racial geography, the recent digitization and publication of HOLC's maps by the Digital Scholarship Lab has resulted in research linking the program to twenty-first century inequalities in segregation, upward mobility, and health (Michney 2021). HOLC was primarily

[1] https://dqydj.com/historical-homeownership-rate-united-states/

intended to address the wave of foreclosures caused by the collapse of the economy during the Great Depression by refinancing mortgages of individuals at risk of default. From its start in 1933 to 1936, HOLC originated mortgages for about 1 million households (Fishback, Rose, and Snowden 2013; Harriss 1951). However, other scholars contest the historical significance of HOLC's maps, arguing that the maps were drafted after most of HOLC's lending was complete, HOLC lent to Black borrowers and in neighborhoods their own appraisers identified as "hazardous," and–most importantly–that any impact HOLC had was likely much smaller than that of the FHA, which began insuring mortgages in 1935, and VA, which began guaranteeing mortgages after the G.I. Bill was passed in 1944 (Hillier 2003; Mallach 2024; Markley 2024). While both the FHA and VA mortgage programs sought to encourage homeownership, their broader goals differed. The former also sought to foster a sound mortgage market, improve housing standards, and augment employment in construction. The latter aimed to compensate and reward veterans with easy credit (McMichael 1953; Weaver 1948).

In contrast to the rich data available on HOLC's mapping, we have very little data on either FHA or VA activities. FHA did draft maps of neighborhood-level lending risk similar to HOLC's, though only one of those maps is publicly available today (i.e., a map of Chicago, IL) (Xu 2022). The only published study of loan-level data for FHA analyzed 4,334 records in Baltimore, MD, Greensboro, NC, and Peoria, IL to show that African Americans received fewer than 1% of those mortgages (Fishback et al. 2022). No one has yet analyzed loan-level data for the VA.

Given the paucity of data on FHA and VA, scholars have drawn insights about their operations from public and private documentation and correspondence with government officials and financial institutions, as well as contemporaneous accounts of the agencies' activities provided by academics, civil rights activists, and reporters. The FHA's own underwriting manual warned against "inharmonious racial groups" and essentially mandated racially restrictive covenants in new developments (Federal Housing Administration 1938). In 1955, one commentator argued that "FHA adopted a racial policy that could well have been culled from the Nuremberg laws. From its inception FHA set itself up as the protector of the all-white neighborhood (Abrams 1955:229)." The 1950 Census of Housing shows 2.3% of FHA-insured and 2.4% of VA-guaranteed loans to nonwhite homeowners or fewer than 50,000 of more than 2 million such mortgages outstanding in on 1-family owner-occupied homes in 1950 (Bureau of the Census 1952).

In this paper, we digitize and analyze the first national dataset containing information on mortgages subsidized by FHA and VA. While local land records offices--such as registers of deeds--contain records of all FHA and VA loans for that locality, we are unaware of another set of national dataset of FHA and VA lending records from this era. Based on a pilot analysis linking FHA mortgages in South Carolina to the 1940 Census Full Count Data, we estimate that between 9–17% of FHA-insured borrowers were Black. These preliminary estimates indicate that the beneficiaries of these programs were predominantly white, offering valuable insights into how federal mortgage programs have contributed to the persistence of contemporary racial inequality.

Data

Records of the Reconstruction Finance Corporation Mortgage Company (RFCMC)

Our sample of FHA and VA loans is drawn from digitized records from the National Archives in College Park, MD, which holds "loan cards" describing borrower characteristics (See Figures 1 and 2) from mortgages purchased by the Reconstruction Finance Corporation Mortgage Company (RFCMC). Through the end of 1937, for example, RFCMC had purchased 9,683 FHA-insured mortgages. This amounted to 8.7% of the 111,474 of the new construction FHA mortgages during this time period along with 23.3% of the secondary market activity. The FHA acknowledged the RFCMC's importance in its early success, stating it "was of far greater significance in providing an adequate distribution of home-mortgage money on reasonable terms to finance new homes than the amount alone would indicate. The very fact that the discount facilities of this agency could be depended on served to increase confidence with which many institutions made insured mortgage loans[.]" (FHA 4th Annual Report, 13, 36-37, 57) After the Federal National Mortgage Association's (FNMA or "Fannie Mae") creation in 1938, RFCMC shifted to the purchase of FHA-insured mortgages on existing homes. In 1940, they began purchasing Title I, Class 3 FHA-insured mortgages and Title VI FHA-insured mortgages during World War II. RFCMC purchased VA-guaranteed mortgages for about 10 months between 1946 and 1947. Over these dozen years, RFCMC purchased selective credit mortgages to establish the FHA and VA as broad national programs by providing liquidity in capital poor areas of the country.

By the end of 1944, the FHA had insured about 3 million mortgages and we capture about 97,195 of these through RFCMC records. By the end of 1947, the VA had guaranteed about 1.3 million residential mortgages and we capture about 25,597 of these through RFCMC records. While we only see a few percent of the overall records for this era, RFCMC purchase patterns mean that we see approximately a tenth of records for specific programs in specific eras and a substantially higher proportion in specific geographies. For example all VA purchases fall in a ten month period and we see well over half of all FHA mortgages in South Carolina in some programs.

Despite the unbalanced nature of our sample, these results have implications for our understanding of how FHA and VA operated and the importance of these two massive interventions–not just in the housing market, but in American society more broadly. Homeownership is the most important source of wealth for most households–especially lower income households and Black households–and structures racially uneven patterns of residential mobility across a highly segregated geography of opportunity. Our evidence bolsters claims stretching from contemporary activists to present-day scholars of these programs' racially disparate and exclusionary impact with novel granularity.

More broadly, our sample illustrates that the RFCMC served as a hidden hand of government support for the FHA and VA mortgage programs that many Americans assumed were largely creatures of the free market. One historian argues that mid century white Americans' blindness to the government subsidies influencing and subsidizing their housing choices lead them to view the subsequent segregation of suburbia as a "natural" outcome of market forces (Freund 2007). Our work highlights the disproportionately white beneficiaries funded by RFCMC's purchases of FHA and VA

mortgages. While a private lender originated these loans, most were almost immediately purchased by RFCMC--a government owned corporation.

Digitizing Historical Mortgage Records

Figures 1 and 2 are representative examples of FHA and VA "loan cards," respectively. The top left corner of each card identifies the first name, last name and middle initial of the borrower, first name and middle initial of the borrower's spouse (if married), and the location of the home. The top right corner includes a unique ID number and the size of the loan; and information on the lender is in the middle of the card. FHA loan cards contain additional information on whether the mortgage is part of specific FHA programs, such as the smaller and more flexible Title I, Class 3 loans or Title VI loans for World War II housing, and whether the mortgage was eventually sold or transferred to FNMA.

To convert these records into a tabular dataset, we first use an Amazon Web Services (AWS) Textract OCR model to convert each image into a text file. Next, we fine-tune a custom entity recognition model (AWS Comprehend) to classify the OCR output into variables of interest, such as borrower name, borrower location, amount of mortgage, dates, etc. We then convert these entities into a tabular dataset containing features for each loan card and use regular expressions to clean strings. We match a string containing the borrower city (e.g., "East Orange" in Figure 1) to place names in the 1950 NHGIS Place Points file. We use an exact match between borrower city and place name for the majority of cards; for the remainder, we fuzzy match to place names, match to county names, or link manually. We identify the approximate date of the mortgage by using dates indicated on the cards (i.e. approval date) and, for cards without dates, through historical context of program effective dates. The final variables used from this constructed dataset are: borrower name(s), (first, last, and spouse); borrower location (county and state); approximate year of mortgage, and mortgage amount.

We start with 122,792 loan cards (97,195 FHA and 25,597 VA). In cleaning, we identify and remove atypical cards, such as FHA-insured loans to property developers (24,938 cards, 25.2%) and those that are references to, or indicate amendments to, other cards (23,731 cards, 19.2%). Our complete dataset includes 72,881 loans: 48,411 FHA-insured loans and 24,470 VA-guaranteed loans made to individuals or married couples between 1935 and 1947. Figure 3 shows the distribution of loan cards across states. Compared to all FHA and VA loans, the RFCMC-purchased FHA and VA loans are more likely to be in the Southern states, especially Texas, Oklahoma, Florida, Georgia, and South Carolina. Washington, Michigan, and Illinois also had large shares of RFCMC-purchased FHA and VA loans. Importantly, the Southern states all had larger Black populations, which may partially explain the higher share of Black borrowers that we observe compared to other studies.

Linking with Full Count Census Data

We use the IPUMS 1940 and 1950 Full Count Microdata, including names, for information on borrower race. Unlike most other studies that link individual data to Census microdata, the loan cards are more challenging because they lack a date of birth or street address—key features typically used for linking. In this extended abstract, we link a set of pre-1940 FHA loan cards in South Carolina

with the 1940 Census (n = 926). We select South Carolina for this pilot analysis because, relative to other states, RFCMC purchased the largest share of FHA-insured mortgages in South Carolina – nearly one in two. We are expanding our record linking approach to the full national set of RFCMC loan cards, including those in other states, those issued after 1940, and VA-guaranteed mortgages.

To generate a robust estimate of the racial composition of FHA and VA borrowers, we pilot two separate matching processes. In the first, we prioritize identifying exact matches (i.e. one-to-one matches) between borrowers in our loan cards and the 1940 Census. We do this with a multi-step record linkage approach between the loan cards' borrowers in South Carolina and the full population in South Carolina aged at least 15 years old in 1940. We seek matches by linking on the following variables, removing exact matches after each step: (1) first name and last name, (2) spouse first name and last name, (3) first name, last name, and spouse name, and (4) first name, last name, and spouse initial. Using this approach, we are able to find exact matches for 35.7% of pre-1940 FHA mortgages in South Carolina. We plan to expand this approach to include additional steps that incorporate fuzzy matching with Jaro-Winkler distance and that incorporate borrower county to narrow down one-to-many matches.

In the second approach, we calculate the probability that a given borrower is Black based on others' in their state with the same linking characteristics. We collapse the 1940 Census data into a series of summary tables that offer, for each set of linking characteristics used in the first approach, the number of matches and the proportion of each that are Black. Next, we link the entire set of loan card records to each of the summary tables. This approach uses information from exact matches alongside an imputed version of race for the loan cards for which we only have one-to-many matches. Because past scholarship suggests FHA and VA lent disproportionately to white households, this approach of imputing race of FHA and VA borrowers based on population characteristics will bias our estimate of the percent Black upwards. Finally, we plan to pilot probabilistic and machine learning-based approaches to record linkage (e.g., Feigenbaum 2016).

Preliminary Findings and Next Steps

When we prioritize exact matches in linking with the Census, we estimate that 9.7% of pre-1940 FHA loans in South Carolina went to Black borrowers. The racial composition of these FHA loan cards are presented in Figure 4 alongside demographics for the adult population of South Carolina in 1940. Nine percent of the cards that we matched using the first two steps step – which searches for exact matches using first name/last name and, separately, spouse first name/last name – were Black. For loan cards matched in the next step, for which we used first name and last name together with spouse name to identify exact links, we estimate that 10% of borrowers are Black. In the final step, using first name, last name, and spouse first initial to identify exact links, we find that 17% of matches are Black. One possible conclusion from this higher estimate in the final step of matching is that there may be false positives.

In our next step, relaxing the requirement to have an exact match and incorporating the probability that a given borrower is Black based on the race of all possible matches, we estimate that

approximately 17% of pre-1940 FHA borrowers in South Carolina were Black. Our first set of results, along with past scholarship on these programs, provides strong evidence that white households are overrepresented in the FHA and VA borrowers compared to the overall population. Thus, we consider this estimate to represent an upper bound of the proportion of Black borrowers.

We are pursuing several next steps to refine and extend this analysis in advance of the NBER Linking Historical Data Sources for Small Populations Conference. First, we will continue refining our approach for Census linkages, such as through incorporating within-county matching, fuzzy matching based on first name by Jaro-Winkler distance, and probabilistic linking approaches to reduce both the false positive and false negative rate. Second, we will expand our census linking to the entire national sample of RFCMC loan cards, which will include linking mortgages in and after 1940 to the 1950 Full Count Census. Third, we plan to incorporate additional Census variables in our analysis in order to describe borrowers nativity status, gender, and occupational status. And, lastly, we aim to provide an estimate of the total amount of FHA and VA mortgage dollars that went to Black versus white households.

Notably, our sample of FHA and VA loans only represents a small share of the millions of mortgages these two programs subsidized in the 1930s and 1940s. Importantly, while the loans we analyze are likely systematically different from the total population of FHA/VA loans, we have multiple reasons to believe sampling bias inflates our estimates of lending to Black households. One reason is that, because RFCMC was more active purchasing loans in more capital poor regions, our records are disproportionately concentrated in states with larger Black populations. Second, in concurrent strands of research analyzing RFCMC loan cards alongside local property records and city directories detailing the entire universe of FHA loans, we find emerging evidence that RFCMC disproportionately bought more distressed mortgages (i.e., those that were more likely to eventually be foreclosed on) and those deemed more "risky" (e.g., those in neighborhoods with more non-white residents). Future work could leverage publicly available, historical property records (i.e., similar to the research by Fishback et al.) to more comprehensively measure FHA/VA mortgage activity and accurately match those transactions to census records using street addresses in addition to names.

Although social scientists have long posited that FHA and VA discriminated against Black people, empirical analysis of these programs' activities has been limited by an almost complete lack of loan-level data. The only prior research to leverage data of this granularity studied approximately 4,000 mortgages in three cities (Fishback et al. 2021). The current manuscript is the first to use a national dataset to study federal mortgage subsidies during the New Deal Era–a period in which homeownership increased rapidly alongside suburbanization and racial segregation. Based on our pilot analysis of FHA mortgages in South Carolina, we find that while 40.6% of the South Carolina adult population was Black, between 9-17% of FHA-insured borrowers were Black. These estimates indicate a substantial racial disparities in access to federally subsidized mortgage credit. This work represents a substantial increase in our understanding of two of the largest housing market interventions in American history–both of which have been deeply implicated in contemporary racial inequalities in segregation and wealth.

References

Aaronson, Daniel, Daniel Hartley, and Bhashkar Mazumder. 2021. "The Effects of the 1930s HOLC 'Redlining' Maps." *American Economic Journal: Economic Policy* 13(4):355–92. doi: 10.1257/pol.20190414.

Abrams, Charles. 1955. *Forbidden Neighbors: A Study of Prejudice in Housing*. New York: Harper & Bros.

Bureau of the Census. 1952. *1950 Census of Housing Volume IV. Residential Financing: Mortgaged Nonfarm Properties*. Washington, D.C.: United States Department of Commerce.

Connolly, N. D. B. 2014. *A World More Concrete: Real Estate and the Remaking of Jim Crow South Florida*. Chicago ; London: The University of Chicago Press.

Faber, Jacob W. 2020. "We Built This: Consequences of New Deal Era Intervention in America's Racial Geography." *American Sociological Review* 85(5):739–75.

Federal Housing Administration. 1938. *Underwriting Manual: Underwriting and Valuation Procedure Under Title II of the National Housing Act*. US Government Printing Office.

Fishback, Price, Jonathan Rose, Ken Snowden, and Thomas Storrs. 2022. "New Evidence on Redlining by Federal Housing Programs in the 1930s." *Journal of Urban Economics* 103462.

Fishback, Price Van Meter, Jonathan Derek Rose, and Kenneth A. Snowden. 2013. *Well Worth Saving: How the New Deal Safeguarded Home Ownership*. Chicago ; London: The University of Chicago Press.

Freund, David M. P. 2007. *Colored Property: State Policy and White Racial Politics in Suburban America*. Chicago: University of Chicago Press.

Harriss, C. Lowell. 1951. "History and Policies of the Home Owners' Loan Corporation."

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

Hynsjö, Disa M., and Luca Perdoni. 2024. *The Effects of Federal "Redlining" Maps: A Novel Estimation Strategy*.

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):419–52. doi: 10.1177/009614428000600404.

Mallach, Alan. 2024. "Shifting the Redlining Paradigm: The Home Owners' Loan Corporation Maps and the Construction of Urban Racial Inequality." *Housing Policy Debate* 0(0):1–27. doi: 10.1080/10511482.2024.2321226.

Markley, Scott. 2024. "Federal 'Redlining' Maps: A Critical Reappraisal." *Urban Studies* 61(2):195–213. doi: 10.1177/00420980231182336.

McMichael, Stanley L. 1953. *How to Finance Real Estate*. New York, Prentice-Hall.

Michney, Todd M. 2021. "How the City Survey's Redlining Maps Were Made: A Closer Look at HOLC's Mortgagee Rehabilitation Division." *Journal of Planning History* 15385132211013361.

Sugrue, Thomas J. 2005. *The Origins of the Urban Crisis: Race and Inequality in Postwar Detroit: With a New Preface by the Author*. 1st Princeton Classic ed. Princeton: Princeton University Press.

Weaver, Robert C. 1948. *The Negro Ghetto*. New York: Harcourt, Brace.

Xu, Wenfei. 2022. "Legacies of Institutionalized Redlining: A Comparison Between Speculative and Implemented Mortgage Risk Maps in Chicago, Illinois." *Housing Policy Debate* 32(2):249–74.

Figures

Figure 1: Sample card for a FHA-insured mortgage

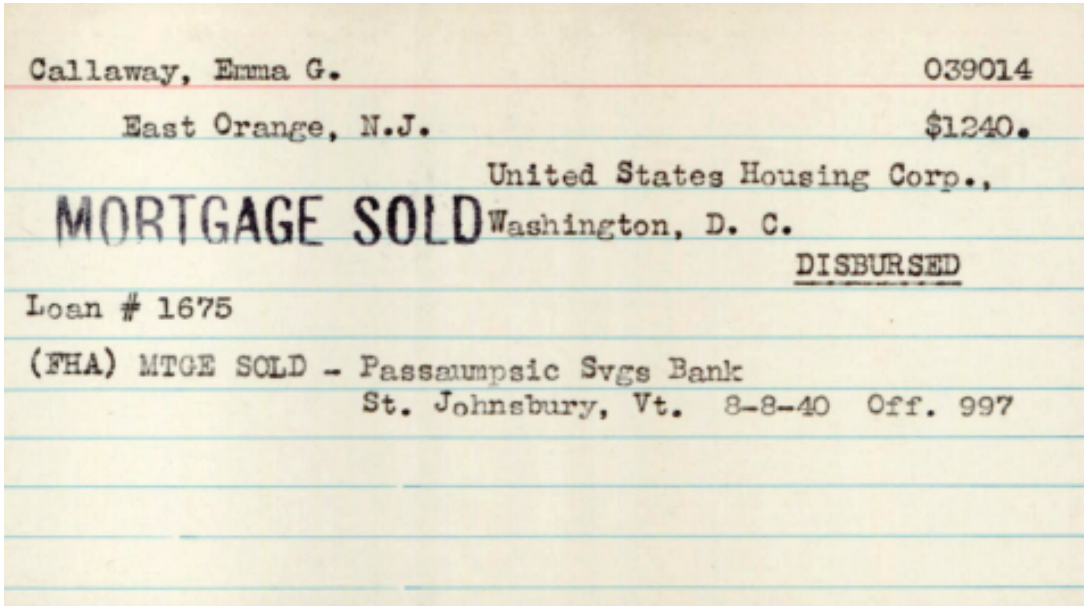


Figure 2: Sample card for VA-insured mortgage

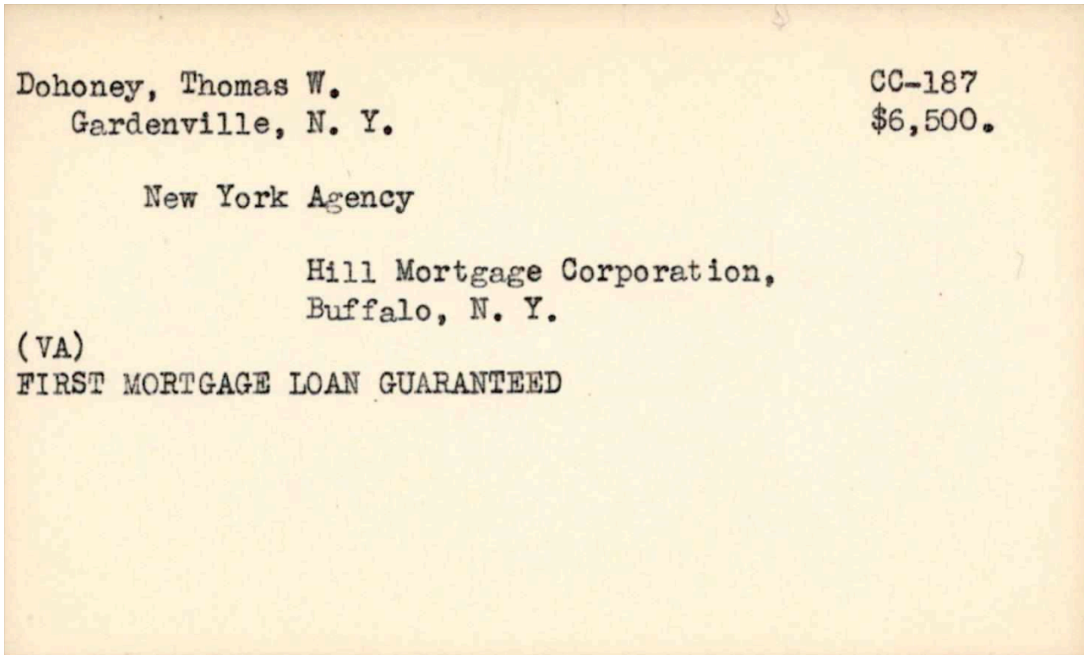


Figure 3: Spatial distribution of RFCMC FHA and VA loan cards across states

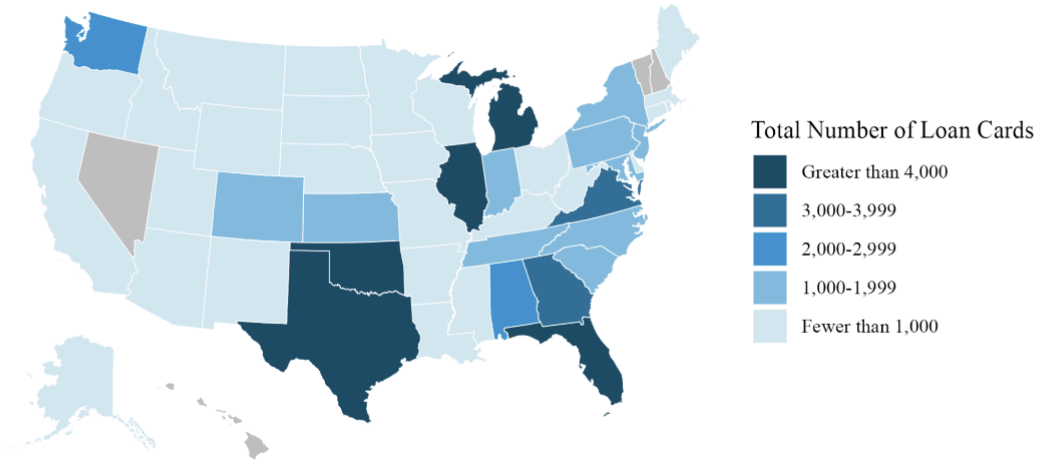


Figure 4: Percent Black among pre-1940 FHA loan cards in South Carolina compared to South Carolina adult population, using exact matches

