Automated Methods for Extracting Housing Data from Historic Land Records*

Extended abstract

Omer Ali Jacob Faber Kate Thomas Gabby Toborg

March 13, 2025

This paper proposes a method for extracting housing data from historic mortgage records. The approach described below has the potential to expand the period of time for which we have property-level housing data well before the 1990s and 2000s. In general, property- or address-level data are available either for the contemporary period or the period before 1940 when US Census data are de-anonymized and researchers can observe the name and address of respondents. In the contemporary period, studies make use of proprietary transactions data collected by real estate monitoring services. The earliest available transactions in these datasets are often in the 1990s, but comprehensive coverage usually begins around the year 2000. Examples of these studies include Kermani and Wong, 2021 who use data from ATTOM, Bayer, M. Casey, et al., 2017; Diamond and McQuade, 2019; Diamond, McQuade, and Qian, 2019 who use data from DataQuick, Bayer, M. D. Casey, et al., 2022 who use data from CoreLogic, and Deng, 2011 who uses data from Fidelity National Information Services Inc. The latter includes transactions beginning in 1985, but the data used were only for one county - Santa Clara, CA.

The historical literature, on the other hand, has leveraged address-level data on home-ownership and home values from the de-anonymized US Census. The names and addresses of respondents are available to view for all Censuses collected before 1950. However, neither home-ownership nor home values are available for 1950 as those responses were not preserved. They are both available for the Censuses of 1930 and 1940, while home-ownership is also available from 1900. Examples of recent studies that use these data to study housing markets include Hynsjö and Perdoni, 2024 and Akbar et al., 2022.

Few studies can examine the intervening period with the same granularity afforded by property- or address-level data. At the same time, this period saw important developments in housing markets, including the rise of the American home-ownership society, the development of the suburbs,

*Ali: University of Pittsburgh, omer.ali@pitt.edu. Faber: New York University, jacob.faber@nyu.edu. Thomas: New York University, katethomas@nyu.edu. Toborg: University of Pittsburgh, gat63@pitt.edu. Financial support from Russel Sage Foundation grant G-2310-45068, and a grant from NYU's Institute of Human Development and Social Change is gratefully acknowledged.

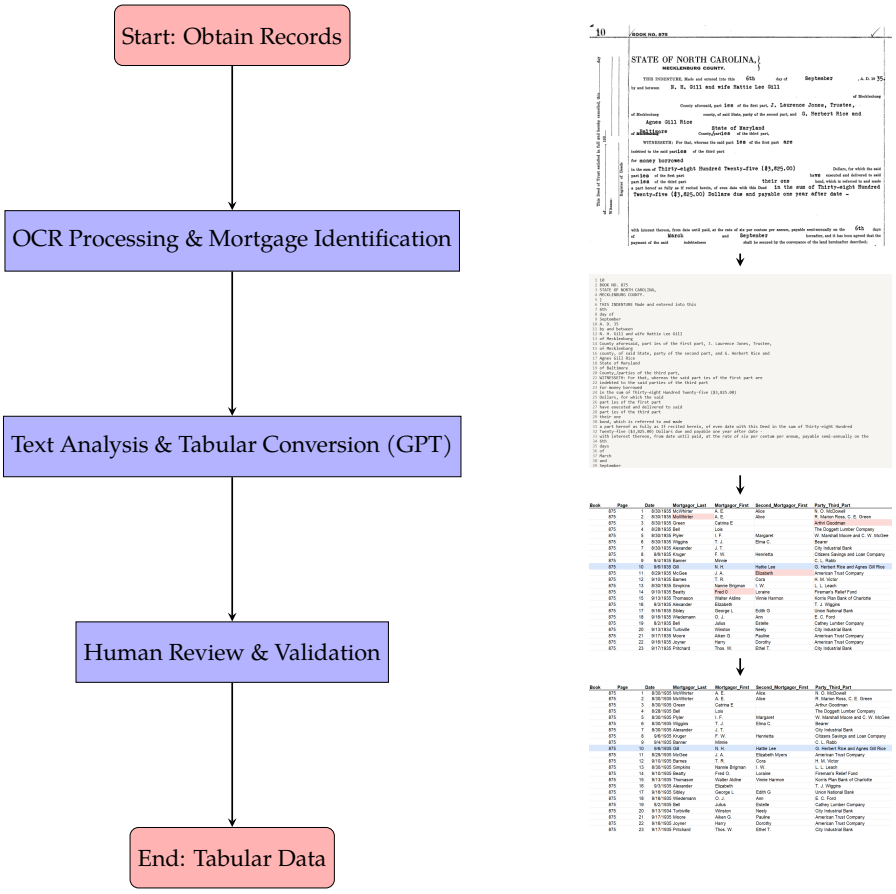
the exodus of white households from central cities, and a dramatic rise in residential segregation (Cutler, Glaeser, and Vigdor, 1999; Massey and Denton, 1993). Gordon and Bruch, 2019 study the trajectory of home values for a sample of roughly 100 properties in St Louis, MO between 1940 and 2016. The small sample size is due to the fact that following properties over this period required identifying relevant transactions in the archival records of the St Louis City Recorder of Deeds. Since property records were not digitized, this involved in-person visits to the office. Despite the small sample, the study uncovered important facts, including stark differences in the implications of home-ownership for wealth accumulation for Black and white residents. Whereas properties in white neighborhoods maintained their values as the city de-industrialized and lost a significant share of its population, those in Black neighborhoods experienced a precipitous decline over the same period. Another study that examines property-level data during this period is Rose and Hartley, 2023, which documents how blockbusting affected home values between 1940 and 1990. The authors use both property records and census tract data for their analysis. A detailed case study is conducted using property-level data for the neighborhood of Edmondson Village, Baltimore between 1954 and 1976. Those data were obtained from mortgages in Baltimore City, which were scanned and made available online, allowing the authors to compile a larger sample than Gordon and Bruch, 2019 and observe the terms of financing, among other details.[1] They find that realty companies engaged in blockbusting resold properties to Black families with a significant markup often through high-risk mortgages. A recent paper, Lyons et al., 2024, comes closest to bridging the housing data gap spanning the second half of the 20th century. Unlike the first set of papers that use real estate records maintained by county recorders, this study leverages real estate listings from newspapers to compile a dataset of 2.7 million properties offered for sale or rent between 1890 and 2006 across 30 cities. In addition to the rent and sale price, their data also include property characteristics, which are missing from mortgage records. On average, the dataset includes 90,000 listings per city and less than 800 per year for each city. For many cities, this constitutes a small share of transactions and may not capture the rich heterogeneity in housing markets across neighborhoods.

Although these recent, innovative, papers shed light on housing market dynamics during a period critical in the construction of America's racial geography, scholars are still limited in what is possible due to the lack of geographically and temporally detailed data on property transactions. Our paper seeks to fill this gap by introducing a method for generating data on property transactions from historic land records. Our method uses optical character recognition (OCR) and queries of large language models to compile information - such as buyer and seller names, address, mortgage lender, and home value - useful to housing scholars into a tabular dataset (see Figure 4).

Figure 1 displays our process for extracting housing data from historical property records, starting with obtaining said records from local counties (i.e., through county recorders/registrars, or

[1] Fishback et al., 2022 use a similar dataset to study the spatial distribution of FHA-insured mortgages in Baltimore and two other cities between 1935 and 1940.

Figure 1: Data extraction pipeline



state archives).[2] The first image on the right is an excerpt from one of these documents, in which N. H. Gill and Hattie Lee Gill secured a mortgage for the property at Lot 6, Square 10, in Crescent Heights, outside the City of Charlotte, NC on September 6, 1935. A larger version of this same record is reproduced in panel (b) of Figure 3.

The second image on the right shows the results of OCR processing for this property record, while the third displays the result of ChatGPT's extraction of pertinent information about the transaction. Notably, most of the record's text is not relevant to our current work, but standard legal language, which is ignored by ChatGPT as it follows the prompt received through its API. The prompt instructs the model to extract the book and page number, date of mortgage, first and last name of mortgagor(s), mortgagee (here called party of the third part), mortgage amount, interest rate, and legal description. Many of these fields are recorded exactly as they appear on the scanned record. We have added orange highlights to the cells that required correction in the subsequent

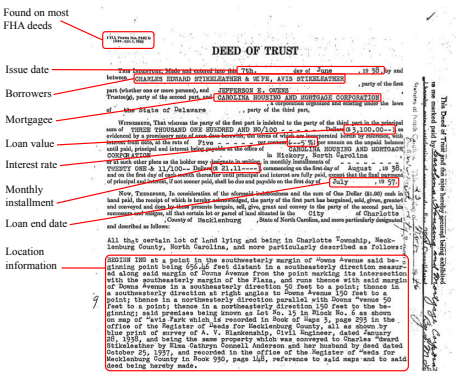
[2]We have acquired millions of scanned mortgages from more than twenty counties across the United States.

step. Finally, each record is reviewed and validated by a human research assistant. The fourth image on the right is a screenshot of our post-validation dataset, which includes minor corrections on some records. Overall, less than 10% of cells require manual correction. Most of these errors occur with missing or misspelled names. Other cells, including dates, amounts and other numbers, have error rates below 3%.

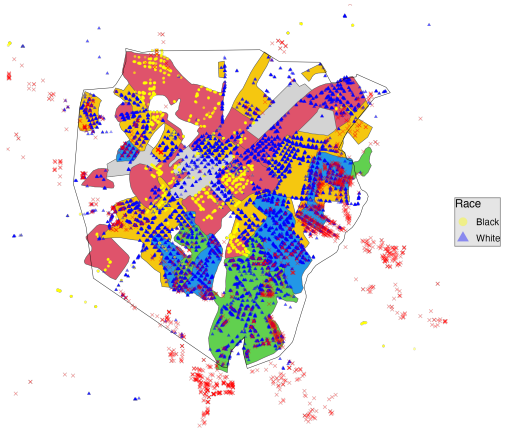
We can geolocate properties based on the legal description included in the mortgage. Some counties - Baltimore, for example - include the address within the property description. In most other counties, such as Tulsa, OK or Charlotte, NC, nearly all properties are uniquely identified by a subdivision name or corresponding plat map, as well as a lot and block number. These can often be linked directly with REGRID parcel data that includes the modern street address, as well as latitude and longitude. Some counties are not as straightforward. These may contain a mixture of addresses, subdivisions, and metes and bounds property descriptions or might not have available parcel data for geo-locating. Properties can still usually be geo-located manually by cross-referencing current and historical maps.

Figure 2

(a) FHA-insured mortgage



(b) HOLC map and FHA mortgages



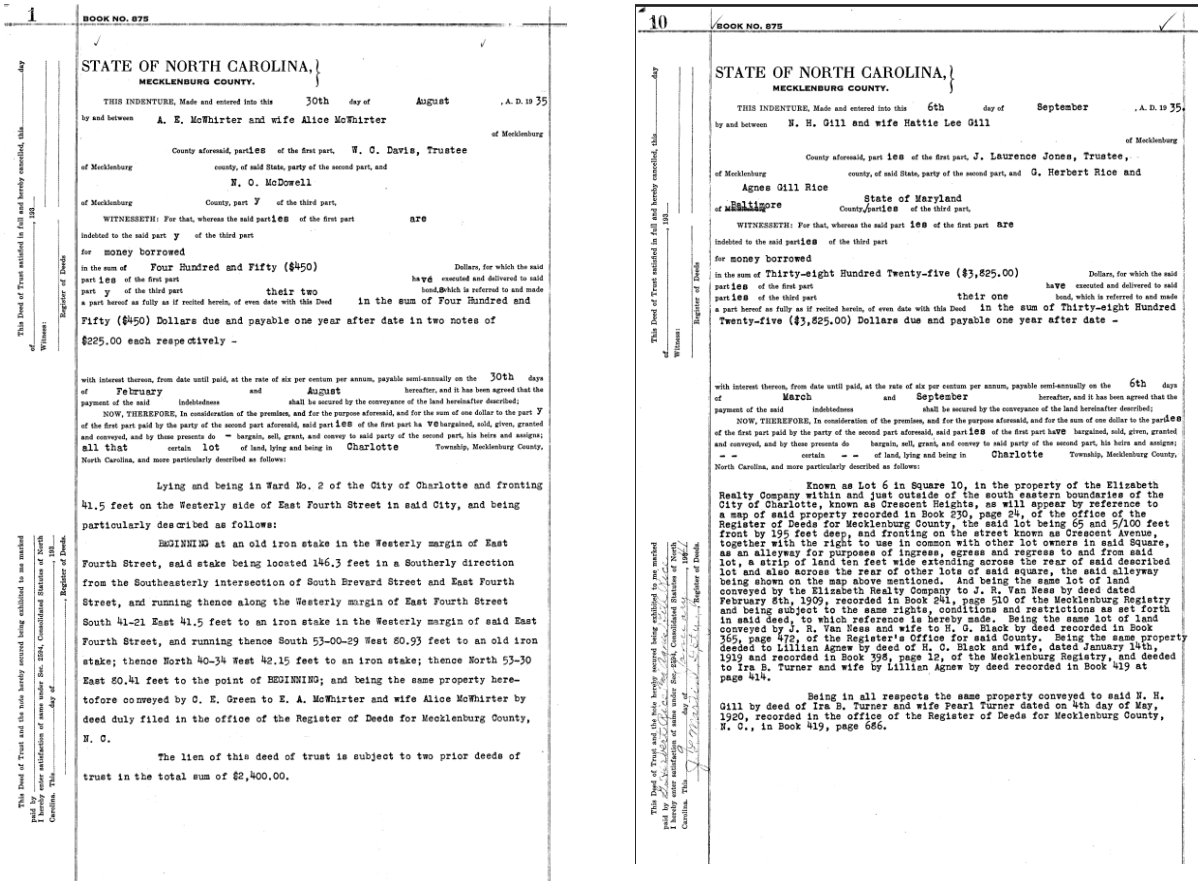
Note: Panel (a) shows a scanned image of an FHA-insured mortgage (called "deed of trust" in some states) issued in Mecklenburg County, annotated with the location of relevant information extracted in the automated data entry stage. Panel (b) shows the HOLC map of Charlotte, NC, city boundary (black line), location of households by race in 1940 (yellow circles and blue triangles), and location of FHA mortgages issued in 1935-40 (red crosses). The polygons correspond to neighborhoods rated in the HOLC map for Charlotte. Color-coded maps rated neighborhoods as best (green), still desirable (blue), definitely declining (yellow) and hazardous (red).

Figure 2 displays an excerpt from another property transaction with an FHA-insured mortgage (on the left) and provides an example of how these data can be used to explore racial and spatial inequalities in FHA mortgage provision (on the right). Specifically, we plotted FHA-insured mortgages issued in 1935-40 as red crosses, Black and White households (yellow circles and blue triangles, respectively), and neighborhood lending risk grades from the Home Owner's Loan Cor-

poration (HOLC) map for Charlotte, NC.

Figure 3: Conventional mortgages (deeds of trust) from Mecklenburg, NC, in 1935.

(a) (b)



1

BOOK NO. 875

STATE OF NORTH CAROLINA,
MECKLENBURG COUNTY.

THIS INDENTURE, Made and entered into this 30th day of August , A. D. 19 35
by and between A. E. McWhirter and wife Alice McWhirter
of Mecklenburg
County aforesaid, parties of the first part, W. O. Davis, Trustee
of Mecklenburg county, of said State, party of the second part, and
N. O. McDowell
of Mecklenburg County, part y of the third part,

WITNESSETH: For that, whereas the said parties of the first part are
indebted to the said part y of the third part
for money borrowed
in the sum of Four Hundred and Fifty ($450) Dollars, for which the said
parties of the first part have executed and delivered to said
part y of the third part their two bond, which is referred to and made
a part hereof as fully as if recited herein, of even date with this Deed in the sum of Four Hundred and
Fifty ($450) Dollars due and payable one year after date in two notes of
$225.00 each respectively -

with interest thereon, from date until paid, at the rate of six per centum per annum, payable semi-annually on the 30th days
of February and August hereafter, and it has been agreed that the
payment of the said indebtedness shall be secured by the conveyance of the land hereinafter described;

NOW, THEREFORE, In consideration of the premises, and for the purpose aforesaid, and for the sum of one dollar to the part y
of the first part paid by the party of the second part aforesaid, said parties of the first part ha ve bargained, sold, given, granted
and conveyed, and by these presents do - bargain, sell, grant, and convey to said party of the second part, his heirs and assigns;
all that certain lot of land, lying and being in Charlotte Township, Mecklenburg County,
North Carolina, and more particularly described as follows:

Lying and being in Ward No. 2 of the City of Charlotte and fronting
41.5 feet on the Westerly side of East Fourth Street in said City, and being
particularly described as follows:

BEGINNING at an old iron stake in the Westerly margin of East
Fourth Street, said stake being located 146.3 feet in a Southerly direction
from the Southeasterly intersection of South Brevard Street and East Fourth
Street, and running thence along the Westerly margin of East Fourth Street
South 41-21 East 41.5 feet to an iron stake in the Westerly margin of said East
Fourth Street, and running thence South 53-00-29 West 80.93 feet to an old iron
stake; thence North 40-34 West 42.15 feet to an iron stake; thence North 53-30
East 80.41 feet to the point of BEGINNING; and being the same property here-
tofore conveyed by C. E. Green to E. A. McWhirter and wife Alice McWhirter by
deed duly filed in the office of the Register of Deeds for Mecklenburg County,
N. C.

The lien of this deed of trust is subject to two prior deeds of
trust in the total sum of $2,400.00.

This Deed of Trust satisfied in full and hereby cancelled, this ____ day of ____, 19__
Witness: ____ Register of Deeds

This Deed of Trust and the note herein secured being exhibited to me marked paid by ____ I hereby enter satisfaction of same under Sec. 2594, Consolidated Statutes of North Carolina. This ____ day of ____, 19__ ____ Register of Deeds.

10

BOOK NO. 875

STATE OF NORTH CAROLINA,
MECKLENBURG COUNTY.

THIS INDENTURE, Made and entered into this 6th day of September , A. D. 19 35.
by and between N. H. Gill and wife Hattie Lee Gill
of Mecklenburg
County aforesaid, part ies of the first part, J. Laurence Jones, Trustee,
of Mecklenburg county, of said State, party of the second part, and G. Herbert Rice and
Agnes Gill Rice
of Baltimore State of Maryland County, parties of the third part,

WITNESSETH: For that, whereas the said part ies of the first part are
indebted to the said parties of the third part
for money borrowed
in the sum of Thirty-eight Hundred Twenty-five ($3,825.00) Dollars, for which the said
parties of the first part have executed and delivered to said
parties of the third part their one bond, which is referred to and made
a part hereof as fully as if recited herein, of even date with this Deed in the sum of Thirty-eight Hundred
Twenty-five ($3,825.00) Dollars due and payable one year after date -

with interest thereon, from date until paid, at the rate of six per centum per annum, payable semi-annually on the 6th days
of March and September hereafter, and it has been agreed that the
payment of the said indebtedness shall be secured by the conveyance of the land hereinafter described;

NOW, THEREFORE, In consideration of the premises, and for the purpose aforesaid, and for the sum of one dollar to the parties
of the first part paid by the party of the second part aforesaid, said parties of the first part have bargained, sold, given, granted
and conveyed, and by these presents do bargain, sell, grant, and convey to said party of the second part, his heirs and assigns;
- - certain - - of land, lying and being in Charlotte Township, Mecklenburg County,
North Carolina, and more particularly described as follows:

Known as Lot 6 in Square 10, in the property of the Elizabeth
Realty Company within and just outside of the south eastern boundaries of the
City of Charlotte, known as Crescent Heights, as will appear by reference to
a map of said property recorded in Book 230, page 24, of the office of the
Register of Deeds for Mecklenburg County, the said lot being 65 and 5/100 feet
front by 195 feet deep, and fronting on the street known as Crescent Avenue,
together with the right to use in common with other lot owners in said Square,
as an alleyway for purposes of ingress, egress and regress to and from said
lot, a strip of land ten feet wide extending across the rear of said described
lot and also across the rear of other lots of said square, the said alleyway
being shown on the map above mentioned. And being the same lot of land
conveyed by the Elizabeth Realty Company to J. R. Van Ness by deed dated
February 8th, 1909, recorded in Book 241, page 510 of the Mecklenburg Registry
and being subject to the same rights, conditions and restrictions as set forth
in said deed, to which reference is hereby made. Being the same lot of land
conveyed by J. R. Van Ness and wife to H. G. Black by deed recorded in Book
365, page 472, of the Register's Office for said County. Being the same property
deeded to Lillian Agnew by deed of H. G. Black and wife, dated January 14th,
1919 and recorded in Book 398, page 12, of the Mecklenburg Registry, and deeded
to Ira B. Turner and wife by Lillian Agnew by deed recorded in Book 419 at
page 414.

Being in all respects the same property conveyed to said N. H.
Gill by deed of Ira B. Turner and wife Pearl Turner dated on 4th day of May,
1920, recorded in the office of the Register of Deeds for Mecklenburg County,
N. C., in Book 419, page 686.

This Deed of Trust satisfied in full and hereby cancelled, this ____ day of ____, 19__
Witness: ____ Register of Deeds

This Deed of Trust and the note herein secured being exhibited to me marked paid by ____ I hereby enter satisfaction of same under Sec. 2594, Consolidated Statutes of North Carolina. This ____ day of ____, 19__ ____ Register of Deeds.

Note: These deeds represent one of the standard forms used in Mecklenburg, NC, in 1935. In the tabular output in Figure 4, the information for these particular records is highlighted.

Figure 4: Sample of tabular output for conventional mortgages in Mecklenburg, NC

Book	Page	Date	Mortgagor_Last	Mortgagor_First	Second_Mortgagor_First	Party_Third_Part	Street_Names	Amount	Interest_Rate
875	1	8/30/1935	McWhirter	A. E.	Alice	N. O. McDowell	East Fourth Street	450	6%
875	2	8/30/1935	MoWhirter	A. E.	Alice	R. Marion Ross, C. E. Green	East Fourth Street	672.42	6%
875	3	8/30/1935	Green	Catrina E		Arthri Goodman		2200	6%
875	4	8/28/1935	Bell	Lois		The Doggett Lumber Company	Steele Creek Road	800	6%
875	5	8/30/1935	Plyler	I. F.	Margaret	W. Marshall Moore and C. W. McGee	York Road	450	6%
875	6	8/30/1935	Wiggins	T. J.	Elma C.	Bearer	Hillcrest Avenue	3500	6%
875	7	8/30/1935	Alexander	J. T.		City Industrial Bank	Beatty's Ford Road	300	6%
875	8	9/6/1935	Kruger	F. W.	Henrietta	Citizens Savings and Loan Company	Lucena Street	800	6%
875	9	9/4/1935	Banner	Minnie		C. L. Rabb	East First Street	$2,189.58	6%
875	10	9/6/1935	Gill	N. H.	Hattie Lee	G. Herbert Rice and Agnes Gill Rice	Crescent Avenue	3825	6%
875	11	8/29/1935	McGee	J. A.	Elizabeth	American Trust Company	Wilson Avenue, Townes Road	3500	6%
875	12	9/10/1935	Barnes	T. R.	Cora	H. M. Victor	Oakland Avenue	250	6%
875	13	8/30/1935	Simpkins	Nannie Brigman	I. W.	L. L. Leach		3500	6%
875	14	9/10/1935	Beatty	Fred 0	Loraine	Fireman's Relief Fund		4500	6%
875	15	9/13/1935	Thomason	Walter Aldine	Vinnie Harmon	Korris Plan Bank of Charlotte	La Fayette Avenue	1000	6%
875	16	9/3/1935	Alexander	Elizabeth		T. J. Wiggins			
875	17	9/16/1935	Sibley	George L	Edith G	Union National Bank		3000	6%
875	18	9/16/1935	Wiedemann	O. J.	Ann	E. C. Ford	Beverly Drive	6000	6%
875	19	8/2/1935	Bell	Julius	Estelle	Cathey Lumber Company	Baldwin Avenue, Morgan Street,	201.7	6%
875	20	9/13/1934	Turbiville	Winston	Neely	City Industrial Bank	Third and Morrow	2500	6%
875	21	9/17/1935	Moore	Aiken G.	Pauline	American Trust Company	Park Avenue,Lyndhurst Avenue	3400	6%
875	22	9/16/1935	Joyner	Harry	Dorothy	American Trust Company		3900	6%
875	23	9/17/1935	Pritchard	Thos. W.	Ethel T.	City Industrial Bank		500	6%
875	24	9/20/1935	Chambless	Henry H	Lota	American Trust Company		2500	6%
875	25	10/1/1935	Chambless	Henry H.	Lota	James L. Gordon	Charlotte	200	6%
875	26	9/19/1935	Jarrell	C. S.	Annie Belle H.	The Mutual Building and Loan Associati	Park Avenue	4000	6%
875	27	9/17/1935	Stapleton	J.D.	Pearl	J.L. Gordon			6%
875	28	9/11/1935	Perry	M. D.	Ruth Maybell	Chas. S. Bostic	Eugene Street	650	6%
875	29	9/21/1935	Blount	W. B.	Emma	Katherine Goodman	Union Street	55	6%
875	30	8/27/1935	Starnes	L. J.	Mae	G. B. Pendleton, Ada C. Pendleton	Club Road	436	5%

Note: This figure presents a sample of the tabular data returned by ChatGPT from the OCR'd text of a sample of conventional mortgages from Mecklenburg, NC. The highlighted lines correspond to the example mortgages in Figure 3. These data represents a work in progress and may contain errors.

References

Akbar, Prottoy A., Sijie Li Hickly, Allison Shertzer, and Randall P. Walsh (Nov. 2022). "Racial Segregation in Housing Markets and the Erosion of Black Wealth". In: *The Review of Economics and Statistics*, pp. 1–45 (cit. on p. 1).

Bayer, Patrick, Marcus Casey, Fernando Ferreira, and Robert McMillan (Nov. 2017). "Racial and ethnic price differentials in the housing market". In: *Journal of Urban Economics* 102, pp. 91–105 (cit. on p. 1).

Bayer, Patrick, Marcus D. Casey, W. Ben McCartney, John Orellana-Li, and Calvin S. Zhang (Sept. 2022). *Distinguishing Causes of Neighborhood Racial Change: A Nearest Neighbor Design*. Working Paper (cit. on p. 1).

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor (1999). "The Rise and Decline of the American Ghetto". In: *Journal of Political Economy* 107.3, pp. 455–506 (cit. on p. 2).

Deng, Lan (2011). "The External Neighborhood Effects of Low-Income Housing Tax Credit Projects Built by Three Sectors". In: *Journal of Urban Affairs* 33.2, pp. 143–166 (cit. on p. 1).

Diamond, Rebecca and Tim McQuade (June 2019). "Who Wants Affordable Housing in Their Backyard? An Equilibrium Analysis of Low-Income Property Development". In: *Journal of Political Economy* 127.3, pp. 1063–1117 (cit. on p. 1).

Diamond, Rebecca, Tim McQuade, and Franklin Qian (Sept. 2019). "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco". In: *American Economic Review* 109.9, pp. 3365–3394 (cit. on p. 1).

Fishback, Price V., Jonathan Rose, Ken Snowden, and Thomas Storrs (May 2022). "New Evidence on Redlining by Federal Housing Programs in the 1930s". In: *Journal of Urban Economics* (cit. on p. 2).

Gordon, Colin and Sarah K. Bruch (Sept. 2019). "Home inequity: race, wealth, and housing in St. Louis since 1940". In: *Housing Studies* 0.0, pp. 1–24 (cit. on p. 2).

Hynsjö, Disa M. and Luca Perdoni (2024). *The Effects of Federal "Redlining" Maps: A Novel Estimation Strategy*. Tech. rep., p. 76 (cit. on p. 1).

Kermani, Amir and Francis Wong (Sept. 2021). *Racial Disparities in Housing Returns*. Working Paper 29306. National Bureau of Economic Research (cit. on p. 1).

Lyons, Ronan C., Allison Shertzer, Rowena Gray, and David N. Agorastos (June 2024). *The Price of Housing in the United States, 1890-2006*. Working Paper (cit. on p. 2).

Massey, Douglas S. and Nancy A. Denton (1993). *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press (cit. on p. 2).

Rose, Jonathan and Daniel Hartley (2023). *Blockbusting and the Challenges Faced by Black Families in Building Wealth through Housing in the Postwar United States*. Tech. rep. WP 2023-02. Federal Reserve Bank of Chicago (cit. on p. 2).