Where the Other Half Lives: Evidence on the Origin and Persistence of Poor Neighborhoods from New York City 1830-2011

Carlos Villarreal[1]
Draft: April 11, 2013

Abstract

I document the influence of early environmental conditions on the long-run distribution of housing prices and income on Manhattan Island by reconstructing and linking historical data in a spatial framework. An early topographic survey provides the location of historical marshes, areas endowed with poor natural drainage, which generated negative sanitation externalities at the time of settlement. I measure the influence of distance from those marshes on the rental price of housing from 1830 through 1940. An early aversion to historical marshes persists over time despite the introduction of sanitation infrastructure and the discovery of communicable disease pathways. The persistence of the aversion to the initial environmental disamenity arises from the entrenchment of low-income individuals on historical marsh sites while high-income individuals settle locations at ever increasing distances from the poor, which helps maintain the pattern of aversion to initial natural disamenities after the characteristics were no longer relevant.

Where are the tenements of today? … Crowding all the lower wards, wherever business leaves a foot of ground unclaimed; strung along both rivers, like a ball and chain tied to the foot of every street, and filling up Harlem with their restless, pent-up magnitudes, they hold within their clutch the wealth and business of New York

- Jacob Riis, *How the Other Half Lives*, p. 19, 1890

Introduction

Census data indicate that many low-income districts have been entrenched within US cities since at least 1880 despite the reshaping of metropolitan areas in the twentieth century. Prominent examples include Pilsen in Chicago, the Lower East Side and Harlem in New York City, and Mid City in New Orleans. These lower-income neighborhoods share a common natural endowment: they

[1] Department of Economics, University of Illinois at Chicago and the Center for Population Economics, University of Chicago. E-mail: carlos@uchicago.edu

encompass districts with poor natural drainage[2]. Early settled areas with good natural drainage, such as Park Slope and Brooklyn Heights in Brooklyn, Beacon Hill in Boston, Nob Hill in San Francisco, and Georgetown in Washington DC grew into high-income neighborhoods with high housing prices. That the early drainage environment might influence the long-run organization of people within cities is important because it links features of contemporary urban form to exogenous initial conditions at the time of settlement. The oldest US cities grew during a period when natural drainage influenced local sanitary conditions and consequently contributed to variation in the initial desirability of locations. New York City on Manhattan Island provides an ideal setting to disentangle the influence of poor drainage from that of elevation since its gentle topographic variation leaves limited opportunities for scenic views but determines the quality of local drainage. Early maps indicate that the location of historical marshes, the areas with the poorest natural drainage, were located throughout Manhattan Island including interior districts.

This project uses recent advances in geographic information system (GIS) technology to reconstruct a broader set of more spatially refined local amenities than was possible at the time of earlier work on real estate values (Edel and Sclar, 1975; McMillen, 1996; Atack and Margo, 1998) and the distribution of income (Gin and Sonstelie, 1992) within historical cities. A resurgence in research focusing on historical cities uses related GIS methods to evaluate the influence of urban conflagrations on the capital intensity of land use, residential density (James Siodla, ongoing), building and land values (Richard Hornbeck and Daniel Keniston, ongoing) as well as the influence of street grids on land values (Trevor O'Grady, ongoing), and the influence of bedrock depth on the distribution of skyscrapers (Barr, 2010; Barr, Tassier and Trendafilov, 2011). The long-run consequences of initial environmental conditions on the distribution of housing prices and people *within* cities remain underexplored.

[2] Median household income increased one percent with every ten meters of additional elevation within US cities in 2000.

Vestiges of the early housing market and people remain visible on Manhattan Island, as I will illustrate with both maps and regression results. I begin by showing that housing prices capitalize an aversion to historical marsh locations from 1830 through 1940, just before the initiation of rent controls. This is followed by an evaluation of possible persistence mechanisms that yields one plausible pathway: low-income households become entrenched on or near historical marshes while high-income households show a growing preference for residential locations away from the poor. Even Central Park, widely considered a valuable local amenity, cannot shield nearby blocks from the negative influence of historical marshes on housing prices and household income. Large-scale redevelopment also does not appear to provide neighborhoods an opportunity to break from their historical legacy. Areas with the greatest redevelopment, including Harlem and the Lower East Side, show the slowest growth in median household income and remain among the lowest income districts on Manhattan Island. Districts at the greatest distance from historical marshes, such as the Upper East and Upper West Sides, show among the largest increases in median income despite a lack of redevelopment. This suggests that features of the local environment that influence the desirability of locations at the time of settlement, even those that lost relevance long ago, may shape the long-run distribution of housing prices and income within cities.

The Distribution of Housing Prices and Income Within Cities

Traditional urban models follow Alonso (1964), Muth (1969), and Mills (1967, 1972) in emphasizing the importance of commuting transportation costs on the distribution of real estate value, population density, and income. Concisely, housing prices capitalize the value of access to the central business district (CBD) and housing amenities. The value of locations decreases as commuting costs rise with distance from the CBD, leading to higher prices in areas with lower transportation costs. Recent work, however, suggests that the influence of proximity to public transportation on housing prices can vary even within a city, leading to lower housing prices near public transportation in some neighborhoods while increasing housing prices elsewhere (McMillen and Redfearn, 2010). These models also do not predict the ascent of valuable areas outside the initial central business district such as the Upper West, Upper East, and Midtown districts on Manhattan Island, which this study finds were valuable districts before the construction of Grand Central Terminal or the opening of Central Park. Explaining these developments requires secondary sources of variation in the desirability of locations possibly unrelated to commuting transportation costs.

I make a simple modification to this model by adding environmental characteristics that were influential at the time of settlement into the set of amenities capitalized into housing prices. This is already partly implemented among studies that include proximity to contemporaneous natural amenities, such as distance to the nearest shore, in the hedonic analysis of housing prices. Those analyses overlook the possible long-run influence of historically prominent but now irrelevant or unobservable environmental characteristics on the distribution of housing prices. A better understanding the initial environment and distribution of housing prices is important because housing prices can influence the distribution of income within cities, with initial and potentially long-run consequences.

Evidence suggests that housing is a normal good in that the demand for housing amenities, capitalized into the price housing, rises with income as illustrated in Figure 1. However the influence of transportation costs on the spatial distribution of household income is theoretically ambiguous. The marginal cost of commuting time rises with income, which should attract higher-income households to residential locations near the central business district. This draw to the central business district is juxtaposed against the rising demand for larger quantities of housing and lower housing prices at greater distances from a central business district.

A clear prediction is that higher-income households will seek residential locations with lower commuting transportation costs, such as locations with rapid transit access. Instead, income falls in areas surrounding new access to public transportation (Glaeser, Kahn, and Rappaport, 2008) as lower-income households seek proximity to public transportation. This suggests that characteristics other than transportation costs influence the distribution of income. One proposition is that sufficient amenities in the central business district can entice the wealthy to reside near the central business district (Paris), while insufficient amenities in the central business district may not be able to attract the wealthy, eventually leading to their exodus to the suburbs (Detroit) (Brueckner, Thisse and Zenou, 1999).

I expand on this theme along several dimensions. I propose that exogenous natural amenities at the time of settlement, including those that are no longer observable, may influence both the initial and long-run distribution of housing prices, and income within cities[3]. These natural amenities can be dispersed throughout the settled area and need not directly relate to the location of the central business district[4]. I also identify a specific disamenity, historical marshes, as a source of variation in the desirability of locations at the time of settlement for New York City. The set of influential environmental amenities at the time of settlement surely varies within and among other cities, suburbs, and exurbs since each location faces unique conditions when the unsettled area is converted into to urban land use. A clearer picture of the forces underlying the initial desirability of locations and settlement patterns may improve our understanding of the growth and structure of contemporary cities.

[3] The quality of man-made amenities, including restaurants and shopping, may emerge from the same initial distribution of desirability determined by environmental characteristics at the time of settlement.

[4] Brueckner, Thisse , and Zenou (1999) discuss that natural amenities, historical amenities, and modern amenities can be found throughout cities. The study then focuses on a simplified model where amenities in the central business district determine whether the wealthy reside near the central business district or move to suburban locations.

Reconstructing and Measuring the Historical Record of New York City from 1830-2011

The strength of this research is in the reconstruction and linkage of historical records in a GIS environment. A data appendix contains a detailed description of the methodology I developed to accomplish this task. I use a variety of historical maps and aerial surveys to reconstruct the historical street centerlines for all of Manhattan Island[5], which form the foundation of the spatial reconstruction of all subsequent layers of historical data. Those data serve as the reference frame from which I overlay and extract the spatial extent of historical marshes from the *Topographic Atlas of the City of New York* (Viele, 1874). I then manually geocode the address-level information for every non-residential building listed in the first two building surveys of Manhattan Island (Dripps, 1851, 1852; Perris, 1852) in order to reconstruct the richest picture of location amenities commonly included in contemporary housing market analyses. This includes but is not limited to the location of police stations, schools, markets, parks, hospitals, churches and manufacturers. The geocoded point data fall within the historical extent of the property lots and achieve the same ten foot accuracy as my historical street centerlines.

I reconstruct the early housing market by geocoding the address-level rental housing price data from Robert Margo's sample of newspaper rental advertisements from 1830-1860 and build proximity measures from the environmental records described above. The housing data consists of 782 advertisements that include detailed descriptions of apartment locations and interior amenities. There is no way to verify the inclusion or exclusion of amenities within the housing unit from the advertisements. The quantity of housing characteristics might thus be measured with error. However, the number of listed amenities does not significantly vary within rental price quartiles across time, which suggests that estimates should not be biased but that the standard errors may be large. These data also represent the

[5] The street lines fall within a ten foot tolerance of contemporary street centerlines produced by the New York City Department of Information Technology and Telecommunications in areas with unchanged historical street paths. These data have been incorporated into the Historical Urban Ecological (HUE) data set of the Center for Population Economics at the University of Chicago, which are available for download at http://hue.uadata.org .

asking price of apartment space by property owners, which introduces uncertainty about the final contract price. Some advertisements indicate negotiable prices based on renter characteristics. This suggests that most of the advertisements may have been less open to negotiation. Additionally, several observations only report the nearest street intersection rather than the precise house number, and I therefore geocoded those properties to the specified intersection. I worry that this spatial generalization might adversely impact the analysis of location amenities on housing prices. However, the average historical street block segment length is only 376 feet. Half of the properties on a block report the same nearest intersection. Thus data geocoded to an intersection fall within 188 feet, or less than 1/25th of a mile, from the precise location of the property, which does not influence the results. I use the number of rooms as the measure of the quantity of housing space since the advertisements do not report square footage.

I then gather a set of block-level housing prices and characteristics from the 1940 Federal Housing Census. These are the last housing data from New York City before the implementation of rent controls and the New York City Housing Authority began work on the majority of their redevelopment projects[6]. I again use the historical street centerlines to reconstruct the block boundaries and generate location amenity proximity measures. The census data detail the characteristics of dwelling units at the block-level, including the total number of units occupied, units vacant, building age, number of private baths, quality of maintenance, density of occupation, and the average rental price. Table 2 presents the summary statistics for these data. No equivalent data exist for housing sales or mortgages. However, only 1.7 percent of dwelling units were owner-occupied on Manhattan Island in 1940. The rental price data thus capture this housing market.

[6] Only 5.63 percent of the public housing dwelling units owned by the New York City Housing Authority in May of 2012 were constructed before 1940.

Although these data include details about the quality of apartments, they do not report information about the size of the apartments. In order to generate trustworthy results I must generate a measure of the average size of dwelling units on a block in a manner that does not underestimate the average size of apartments at the block-level in 1940, which could inflate the estimated influence of floor space on apartment rental prices. Contemporary data from the New York City Department of Planning (2012) provide lot-level information on the age of structures, the number of residential units, and the total floor space of residential units. As of May 2012, 81.5 percent of lots representing 46.48 percent of the total lot space on Manhattan Island contained a structure built on or before 1940. I use those data to calculate the average lot-level apartment square footage for all lots with residential structures build on or before 1940 then employed a kriging estimation strategy to interpolate the average apartment square footage at a grid of 20 million points across Manhattan Island. I then reassembled the grid of interpolated square-footage values into the 1940 blocks to generate a measure of the block-level average apartment square footage. The exclusion of blocks using exclusively interpolated apartment square footage data does not influence the results.

A survivorship bias is built into the population of surviving historical structures. Better built and better maintained buildings are more likely to survive. This implicitly assumes that lower grade and poorly maintained buildings, including those in the slums with small apartment square footage, were less likely to survive. Estimates of the average apartment square footage in 1940 derived from data on surviving structures should therefore overestimate the average size of 1940 apartments and thus underestimate the influence of apartment square-footage on rental prices. Regression results will therefore provide lower bound estimates of the influence of apartment size on apartment rental prices.

The construction of the income data from 1880 through 2011 was more straightforward. I assemble the early income and population data from the 1880 IPUMS 100% Population Census data

(Ruggles et al, 2010). The enumeration district-level median household income statistics represent the income within households excluding boarding house residents and institutionalized individuals. I again use my historical street centerlines to reconstruct the 1880 enumeration districts. The tabular and spatial population and income data from the 1940, 1950, and 2000 censuses, as well as the 2007-2011 American Community Survey were drawn from the NHGIS (Minnesota Population Center, 2011). I also utilize contemporary lot-level data from the New York City Department of Planning discussed above to examine the influence of redevelopment efforts from 1950 through 2000 on the change in median household income during that period. This exercise requires the simple reassembly of the lot-level data into census tract-level statistics.

I estimate the influence of amenities on housing prices using a hedonic estimation strategy where the price of housing is a function of the value of the location and the quantity of housing in the dwelling unit. Characteristics that influence the housing price include dwelling unit size, the quality of workmanship, and interior fixtures. Amenities influencing the value of a location include distance from the central business district, distance from the nearest historical marsh, distance from the nearest park, and distance from the shoreline. Distance from Central Park is excluded from the 1830-1860 analysis because the park opened after the study period. The full battery of amenity controls is detailed below each set of regression results. I begin by estimating the value of amenities from the rental price of housing for the address-level data from 1830 through 1860 using the following strategy:

$$\ln(r_{it}) = \alpha + \beta D_{it} + \delta \ln(S_{it}) + \gamma M_{it} + \theta X_{it} + \mu_t + \varepsilon_{it}$$

Where the rental price of housing (r) of property (i) in time (t) is a function of its distance from the central business district (D), the size of the apartment (S), distance from the nearest historical marsh (M), a vector of housing and location characteristics (X), and year dummies (μ) to capture movement in the price level from 1830 through 1860, which allows me to pool the data into a single regression. The

1940 housing price data are observed in a single year, thus eliminating the need for the time component of the analysis. The 1940 analysis also includes distance from Central Park as a location amenity.

The analysis of the influence of historical marshes on the distribution of income within cities in 1880, 1950 and 2000 uses a similar estimation strategy but excludes housing amenities. This adjustment yields the following estimation strategy:

$$\ln(y_i) = \alpha + \beta D_i + \gamma M_i + \theta X_i + \varepsilon_i$$

Where the log median household income (y) in census district (i) is a function of distance from the central business district (D), distance from the nearest historical marsh (M), and a vector of location amenities (X) including distance from Central Park, distance from nearest park (not Central Park), distance from the shoreline, and distance from nearest rapid transportation station.

The Long-Run Aversion to Historical Marshes and the Origin and Persistence of Poor Neighborhoods

The rental price of housing from 1830 through 1860 shows aversion to historical marsh locations, with housing prices rising 30.47 percent with each additional mile of distance from a historical marsh in Model 1 of Table 3. The magnitude of the influence of historical marshes on housing prices remains large even after the introduction of location amenities in Model 2 and apartment amenities in Model 3. The controls for location and apartment amenities capture approximately 1/3 of the influence of the historical mashes on housing prices, which suggests that the distribution of amenities broadly follows from the initial pattern of aversion to historical marshes. A measure of proximity to Central Park, eventually a dominant feature of Manhattan Island, is excluded from these models since the park did not exist during this study period[7]. I include a measure of proximity to Central Park in the analysis of the distribution of housing prices and income in all subsequent years. The large coefficient on distance from

[7] The spatial extent of the park was finalized in 1861 but remained closed to the public for landscape redevelopment until 1873.

historical marshes in Model 3 shows a broad aversion to historical marshes even after the introduction of the location and apartment controls although the standard error indicates a mix of housing prices.

The aversion to historical marsh locations in the housing market persists through 1940. Apartment rental prices increase 81.25 percent with each additional mile of distance from a historical marsh In Model 1 of Table 4. Location amenities and block-level apartment characteristics respectfully account for 40.5 percent and 34.7 percent of the distribution of value arising from aversion to historical marshes in Models 2 and 3. The Model 3 coefficient on distance from historical marshes does not measurably differ from the results of the broadest model using the rental price of housing data from 1830 through 1860. Distance from historical marshes continues to explain variation in housing prices even after controlling for location amenities and apartment amenities. This suggests that the remaining distribution of value, including the location of unobserved amenities such as high-end restaurants and shopping, follow a distribution consistent with aversion to historical marsh locations. The median rental price of housing on historical marshes is \$22.87 per month, which falls into the 20th percentile of rental housing prices on Manhattan Island in 1940. The median rental price of housing in the old Five Points district, the earliest settled interior marsh represented by point number 1 in Figure 2, is \$18.77 (9th percentile among Manhattan blocks) despite its proximity to the central business district. The influence of historical marshes on the rental price of housing is also evident among the blocks abutting Central Park despite its value as a local amenity. Housing prices decrease 10.66 percent with each additional mile of distance from Central Park. The median rental price of housing on blocks with direct access to Central Park is \$94.21 per month but the median rental price of housing price among the subset of those blocks on the East Harlem marsh, with its epicenter identified as number 2 in Figure 3, is only \$33.61 (12th percentile among blocks abutting Central Park). The lowest monthly rental price of housing among blocks abutting Central Park (\$22.62) is located on a historical marsh location.

I next examine income sorting as a potential entrenchment mechanism for the aversion to historical marshes observed in the housing market. An analysis of the distribution of income on Manhattan Island form 1880 through 2011 in Table 5 reveals that high-income individuals increasingly choose residential locations away from historical marsh locations. In 1880 the enumeration-district level median household income rises 16.14 percent with each mile from a historical marsh in Model 1. The inclusion of location amenities in Model 2 captures approximately 2/3 of the marsh influence with a standard error that indicates income mixing. This result is partly due to the location of high-income households on the blocks surrounding 5th avenue from Washington Square Park through 85th Street, a narrow corridor away from historical marshes that also aligns with distance from the shoreline, a location amenity in the model with a coefficient of 0.1756 ($\hat{\sigma}$ = 0.0608). The four enumeration districts with the highest median household income include the Midtown blocks that now contain the Empire State Building, Rockefeller Center, and blocks abutting Central Park on the Upper East Side that fall into the highest 2% of median household income in 2011. Income varied across the rest of the island, with a small enclave of upper-middle income households located equidistant from two historical marshes in Harlem. The large standard errors on distance from historical marshes and Central Park reflect the partial settlement of the island, with 86 percent of the population and 91 percent of households in the upper quartile of income living south of Central Park near the initial central business district.

By 1950 settlement expanded to cover the entire island, with a concurrent growth in aversion to historical marsh locations reaching a 32.29 percent increase in income with each additional mile of distance from a historical marsh. As higher-income earners sorted away from historical marshes throughout the island, the highest-income enclaves emerged away from Central Park, as evident by its positive coefficient. Four of the five highest median income tracts are located over two miles from Central Park. The tracts of the Upper East and Upper West Sides, the neighborhoods flanking the Eastern and Western border of Central Park, remain high income, falling into the 70th percentile of income on

Manhattan tracts. However, 40 percent of tracts in the upper 50^{t} income percentile can be found abutting the shoreline, away from Central Park. As in the housing market, proximity to Central Park does not help raise the median income of districts on historical marsh locations. The lowest median income tract abutting Central Park lies on top of a historical marsh and falls into the 3rd income percentile on Manhattan Island.

Income sorting away from historical marshes grows until 2000 and then remains stable at a 70 percent increase in income with each mile from a historical marsh through 2011. Higher income households reveal a preference for locations near Central Park by the end of the period, with median household income decreasing 6.96 percent with each additional mile of distance away from the park. The highest income districts surrounding Central Park remain those away from historical marsh locations. Lower income districts show entrenchment on and near historical marsh locations. Blocks near the marshes of the Lower East Side and Harlem remain among the lowest median household block groups on Manhattan Island. In Washington Heights, the neighborhood north of Harlem at the greatest distance from a historical marsh remains a high income enclave. Higher income individuals also reveal a preference for locations near the initial central business district, as evident by the negative coefficient on distance from City Hall. The re-emergence of a preference for locations near the Financial District among higher income households yields the notable exception to the aversion to historical marshes observed throughout the rest of Manhattan Island. The block groups along the Hudson River from the Financial District through Greenwich Village emerge as the prominent high-income enclave located on a historical marsh.

These results suggest that higher-income households increasingly prefer residential locations away from historical marsh locations upon which lower income households reside since at least 1880. The precise externalities that spur the aversion of higher-income households to the poor remain unclear. Businesses in poor neighborhoods may provide the goods and services demanded by higher-income individuals, the area may have lower quality public services, such as lower quality public schools or poorly maintained parks, the districts may have grown into havens of petty crime and vice, or the districts may simply continue to encounter drainage problems and a higher risk of flooding - a historical legacy reaffirmed in the aftermath of Hurricane Sandy in 2012. Poverty generates myriad externalities for which historical data are scarce. What is certain, and measurable, is that the sorting of high income earners away from the poor helps entrench an aversion to historical marsh locations. High-income earners can choose to live wherever they prefer, and they reveal a preference for residential locations away from the poor.

The long-run persistence observed above begs the question: can large-scale interventions help a neighborhood break from a historical path of entrenched poverty? I focus my attention on the influence of redevelopment projects that raze and rebuild anew on the change in local median income, a measure of gentrification. In the period between 1950 and 2000, 96 percent of tracts on Manhattan Island experienced some redevelopment, with the median tract experiencing 23.5 percent of its lot space rebuilt and the maximum tract containing 94 percent redevelopment. The results in Table 6 indicate that redevelopment projects from 1950 through 2000 do not lead to a rise in median income. An increase in the percentage of redeveloped lot square footage within a 1950 tract does not raise median income in Model 1. One worry is that the lot area redevelopment measure may not fully capture the magnitude of redevelopment projects since modern construction techniques provide an opportunity for new buildings to dominate the available interior building space within a neighborhood with small lot footprints. I address this in Model 2 by employing the percent of redeveloped interior building area as the measure of redevelopment and find that this also does not influence the change in median income. Maybe the influence of redevelopment on gentrification differs by investor type. I disentangle the percentage of public residential, private residential, and exclusively commercial private redevelopment in Model 3 and find that private redevelopment does not measurably influence gentrification. Public residential redevelopment appears to concentrate poverty, with a one percent increase in public residential redevelopment leading to a 3 percent decrease in median income from 1950 through 2000. Historical marshes, however, show a strong influence on the distribution of median income growth in all specifications. One mile of additional distance from a historical marsh yields a 105 percent increase in median income during the study window. The distribution of redevelopment and median income growth displayed in Figure 3 show that neighborhoods including the Lower East Side and Harlem experienced among the highest percentage of redevelopment but achieved among the lowest growth in median real income (negative in several tracts). In contrast the Upper West and Upper East sides experienced some of the lowest percentages of reinvestment but achieved among greatest increases in median income, with larger income gains on the tracts at the greater distance from historical marshes[8]. This curious finding suggests that even large scale razing of districts may not be able change the long-run character of neighborhoods, almost as though their historical legacy is inescapable.

I explore two more potential entrenchment mechanisms: 1) that nativity or race may generate a barrier to neighborhood exit and entrench the local characteristics, and 2) that polluting manufacturers may help entrench the aversion historical marshes. If nativity or race prove a barrier to neighborhood exit then the early spatial distribution of those population characteristics should predict their later distribution. Instead I find that population characteristics became more spatially diluted as groups moved within and away from Manhattan Island from 1880 through 1940, the initial entrenchment period of the poor on historical marshes. Figure 4 illustrates that the location of the black population on Manhattan Island in 1880 does not predict their location in 1940. The spatial mobility is evident among a

[8] A similar analysis using the nominal change in median income using 2000 dollars, which adds tracts with missing income data in 1950, does not change the finding. With the exception of the negative influence of public projects, redevelopment does not influence the change in nominal median income. Distance from historical remains the best predictor of the long-run growth in median income.

variety of nativity groups in Table 7, where a one percent increase in 1880 percentage share of the population share only leading to an increase in nativity concentration (coefficient > 1.0) among Russians. There remains a residual Italian, Irish, and Polish population from 1880, but in decreased concentration. Higher income in 1880 negatively influences the entrenchment of the groups. This suggests that higher income individuals moved away and simply left the poor behind in the old neighborhoods, a finding consistent with the long-run income sorting observed earlier.

Polluting manufacturers could lead to the observed aversion to historical marshes if they primarily located on historical marshes and if their negative externalities show long-run capitalization into the housing market. Table 8 indicates that manufacturers in 1851 preferred locations within 0.25 miles of the shore, particularly industries with heavy inputs such brick makers and refineries. Among sites away from the shore firms and institutions show a strong aversion to locating on historical marshes, which indicates that manufacturers are not a potential mechanism for the entrenchment of the historical mash aversion. The persistence of polluting manufacturers might influence the long-run distribution of housing prices even though they did not locate on historical marsh sites. I estimate the influence of historical manufacturers on housing prices in 1830-1860, and 1940 in Table 9 and find that only the location of early foundries shows a negative influence on housing prices in 1940 while distance from historical marshes continues to show a large influence on housing prices even after the introduction of the historical manufacturing controls. These results suggest that the pollution from early manufacturers neither lead to the long-run aversion to historical marsh locations observed in the earlier analysis nor left a measurable historical legacy in the housing market.

Conclusion

Environmental conditions at the time of settlement shaped the growth of New York City on Manhattan Island. The distribution of housing prices show an aversion to historical marsh locations from

1830 through 1940, which appears to persist as high-income earners increasingly prefer locations away from the poor neighborhoods located on historical marshes. It is surprising that the location of historical marshes can predict any of the distribution of apartment rental prices through 1940 or the distribution of income through 2011 in New York City, the largest US city and among the most developed urban environments. This suggests that the initial environment can leave a lasting impression on urban form. Although these results only directly speak to conditions on Manhattan Island, similar forces may be at work in other cities. Future research should expand this study to evaluate the external validity of this finding.

Bibliography

William Alonso. *Location and Land Use*. Cambridge: Harvard University Press. 1964.

Jeremy Atack and Robert Margo. Location, Location Location! The Price Gradient for Vacant urban Land: New York, 1935 to 1900. *Journal of Real Estate Finance and Economics*. 16(2): 151-172, 1998.

Jason Barr, Troy Tassier, and Rossen Trendafilov. Depth to Bedrock and the Formation of the Manhattan Skyline, 1890-1915. *Journal of Economic History*. 71(4): 1060-1077, 2011.

Jason Barr. Skyscraper Height, *Journal of Real Estate Finance and Economics*. 45(3): 723-753, 2010.

Jan Brueckner, Jacques-François Thisse, and Yves Zenou, Why is Central Paris Rich and Downtown Detroit Poor? An Amenity-Based Theory. *European Economic Review*. 43(1): 91-107, 1999.

Matthew Dripps. *Map of That Portion of the City and County of New York North of 50th Street*. Matthew Dripps Publication, 1851 .

Matthew Dripps. *Map of the city of New York Extending Northward to 50th Street*. Matthew Dripps Publication, 1852.

Matthew Edel and Elliott Sclar. The Distribution of Real Estate Value Change: Metropolitan Boston, 1870-1970. *Journal of Urban Economics*. 2(4): 366-387, 1975

Alan Gin and Jon Sonstelie, The Streetcar and Residential Location in Nineteenth Century Philadelphia. *Journal of Urban Economics*. 32(1): 92-107, 1992.

Edward Glaeser, Matthew Kahn, and Jordan Rappaport. Why do the Poor Live in Cities? The Role of Public Transportation. *Journal of Urban Economics.* 63(1): 1-24, 2008.

Richard Hornbeck and Daniel Keniston. Creative Destruction From the Great Boston Fire of 1872: Barriers to Urban Growth Illuminated. (draft), 2012.

John Logan, et. al. Mapping America in 1880: The Urban Transition Historical GIS Project. *Historical Methods.* 44(1):49-60, 2011.

Robert Margo. The Rental Price of Housing in New York City, 1830-1860. *Journal of Economic History*. 56(3): 605-625, 1996.

Daniel McMillen. One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach. *Journal of Urban Economics*. 40(1): 100-124, 1996.

Daniel McMillen and Christian Redfearn. Estimation and Hypothesis Testing for Nonparametric Hedonic House Price Functions. *Journal of Regional Science*. 56(3): 712-733, 2010.

Edwin Mills. An Aggregative Model of Resource Allocation in a Metropolitan Area. *American Economic Review*. 57(2): 197-210, 1967.

Edwin Mills. *Studies in the Structure of the Urban Economy*. Baltimore: Johns Hopkins University Press, 1972.

Edwin Mills. *Urban Economics*. Glenview, Illinois: Scott Foresman, 1972.

Minnesota Population Center. National Historical Geographic Information System: Version 2.0. Minneapolis, MN: University of Minnesota 2011.

Richard Muth. *Cities and Housing*. Chicago: Univeristy of Chicago Press, 1969.

New York (City). Department of City Planning 2012. MapPLUTO (Edition 12v1) New York City: NYC Department of City Planning, 2012.

New York (City). Department of Information Technology and Telecommunications. Street Centerline Shapefile. New York City, 2012.

William Perris. *Maps of the City of New York*. New York City: Korff Brothers Lithography. 1852.

Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

James Siodla. Razing San Francisco: The 1906 Disaster as a Natural Experiment in Understanding Urban Land Use. (draft), 2012.

United States Bureau of the Census. *Sixteenth Census of the United States: 1940. Housing. Supplement to the First Series: Block Statistics for Cities*. Washington DC: US Government Printing Office, 1942.

Egbert Viele. *Topographical Atlas of the City of New York*. New York City: E L Viele Publication, 1874.

Tables

Table 1 - Summary Statistics - Rental Price of Housing 1830 - 1860

Variable	Obs	Mean	Std. Dev.	Min	Max
Asking Yearly Rental Price	782	316.89	216.68	20.00	1400.00
Number of Rooms	782	5.91	4.18	1.00	24.00
Miles from City Hall	782	1.70	1.27	0.10	7.66
Miles from Historical Marsh	782	0.28	0.25	0.00	1.36
Miles from Market	782	0.63	0.59	0.00	2.91
Miles from School	782	0.19	0.14	0.00	0.90
Miles from Police Station	782	0.39	0.39	0.00	2.54
Miles from Fire Department	782	0.25	0.18	0.00	1.40
Miles from Train Depot	782	0.60	0.34	0.00	2.09
Miles from Park	782	0.35	0.24	0.00	1.66
Miles from Merchant's Exchange	782	2.12	1.36	0.13	8.17
Miles from Bank	782	0.78	0.97	0.00	6.28
Miles from Dry Dock	782	1.58	0.76	0.07	6.20
Miles from Brewery	782	0.58	0.73	0.00	5.44
Miles from Oil Manufacturer	782	1.25	0.88	0.07	6.45
Miles from Refinery	782	1.13	1.05	0.00	6.81
Miles from Soap Manufacturer	782	0.65	0.50	0.00	4.46
Miles from Coal Yard	782	0.27	0.24	0.00	1.72
Miles from Foundry	782	0.41	0.32	0.00	2.07
Miles from Hospital	782	0.72	0.70	0.00	5.36
Dummy (Brick)	187				
Dummy (Modern)	286				
Dummy (Elegant)	109				
Dummy (Furnished)	119				
Dummy (Heated)	43				
Dummy (Partial Board)	18				
Dummy (Full Board)	40				
Dummy (Basement)	230				
Dummy (Attic)	92				
Dummy (Kitchen)	92				
Dummy (Private Bathroom)	121				
Dummy (Storage)	84				
Dummy (Prefer Native Tennant)	12				
Dummy (Yard)	103				
Dummy (Near Public Transportation)	83				
Dummy (Shared Room)	18				
Daily Contract	31				
Weekly Contract	130				
Monthly Contract	83				
Yearly Contract	538				

Table 2 - Summary Statistics - Block-Level Apartment Rental Data from the 1940 Federal Housing Census					
Variable	Obs	Mean	Std. Dev.	Min	Max
Block-Level Average Dwelling Unit Rental Price	2414	52.56	54.18	1.96	816
Block-Level Average Dwelling Unit Square Footage	2414	1140.27	699.62	406.33	9296.37
Miles from City Hall	2414	4.70	3.04	0.19	11.92
Miles from Historical Marsh	2414	0.36	0.32	0	1.49
Miles from Central Park	2414	1.73	1.36	0.032	5.38
Miles from a Park (Excluding Central Park)	2414	0.50	0.41	0	2.09
Miles from a Subway Station	2414	0.18	0.12	0	0.69
Elevation in Meters	2414	15.99	14.05	0.28	71.44
Percent of Dwelling Units Owner Occupied	2414	1.78	5.91	0	100
Percent of Dwelling Units Vacant and On the Market	2414	13.10	16.09	0	100
Percent of Dwelling Units Built Before 1900	2414	42.19	38.86	0	100
Percent of Dwelling Units Built From 1900 to 1919	2414	34.72	35.36	0	100
Percent of Dwelling Units Built From 1920 to 1929	2414	12.97	25.68	0	100
Percent of Dwelling Units Built From 1930 to 1940	2414	4.92	15.98	0	100
Percent of Dwelling Units With Over 1.5 Persons Per Room	2414	6.54	17.79	0	100
Percent of Dwelling Units in Need of Repair	2414	8.62	18.95	0	100
Percent of Dwelling Units With No Private Bath	2414	20.89	26.92	0	100

Table 3 - Regression Results - Dependent Variable : Log (Asking Apartment Rental Price) 1830 - 1860			
Variables	Model 1	Model 2	Model 3
Miles From City Hall	-0.0586**	-0.0244	-0.0486
	(0.0238)	(0.0381)	(0.0338)
Miles From Historical Marsh	0.3047**	0.2168	0.2099
	(0.1255)	(0.1537)	(0.1382)
Apartment Size	Yes	Yes	Yes
Location Amenities	No	Yes	Yes
Apartment Amenities	No	No	Yes
Observations	782	782	782
R^2	0.3520	0.3619	0.4212

Conley's corrected standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Notes: Location amenities include distance from public parks, police stations, markets, schools, and depots. Apartment amenities include the number of rooms, quality of fixtures, brick construction, furnished, heated, basement included, attic included, storage space included, kitchen included, yard included, only renting to native born Americans, rental contract length dummies, shared room, partial or full board included, and access to public transportation.

Table 4 - Regression Results - Dependent Variable : Log (1940 Block-Level Mean Apartment Rental Price)			
Variables	Model 1	Model 2	Model 3
Miles From City Hall	0.0493***	0.0304***	-0.0093
	(0.0054)	(0.0085)	(0.0069)
Miles From Historical Marsh	0.8125***	0.4831***	0.3153***
	(0.0685)	(0.0724)	(0.0555)
Miles From Central Park		-0.0939***	-0.1066***
		(0.0080)	(0.0139)
Apartment Size	Yes	Yes	Yes
Location Amenities	No	Yes	Yes
Apartment Amenities	No	No	Yes
Observations	2,414	2,414	2,414
R^2	0.5403	0.5721	0.6853

Conley's corrected standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Notes: Location amenities include distance from a subway station, distance from a park (not Central Park), distance from shore, and the mean elevation, which I include as a control for scenic views. Apartment amenities include the average dwelling-unit square footage, percent of dwelling-units owner occupied, percent of dwelling-units vacant and on the market, percent of dwelling-units built before 1900, percent of dwelling-units built between 1900 and 1919, percent of dwelling-units built between 1920 and 1930, percent of dwelling-units in need of repair, percent of dwelling-units with no private bathroom.

Table 5 - Regression Results - Dependend Variable: Log (Median Income)					
	1880		1950	2000	2007-2011
Variables	Model 1	Model 2	Model 2	Model 2	Model 2
Miles From City Hall	-0.0116**	-0.0075	-0.0057	-0.1182***	-0.0991***
	(0.0054)	(0.0092)	(0.0101)	(0.0143)	(0.0137)
Miles From Historical Marsh	0.1614***	0.0676	0.3229***	0.7062***	0.6994***
	(0.0502)	(0.0617)	(0.0659)	(0.0885)	(0.0825)
Miles From Central Park		-0.0011	0.0709***	-0.0393	-0.0696**
		(0.0161)	(0.0196)	(0.0293)	(0.0274)
Location Amenities	N	Y	Y	Y	Y
Observations	651	651	241	835	1,065
R^2	0.0376	0.0744	0.2110	0.4431	0.3559

Conley's corrected standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Notes: I assembled 1880 housholds from the IPUMS 100 percent individual-level population census data excluding residents of group quarters and institutionalized individuals.The 1950 data represent median family income, as these are the only data currently available at the tract-level. The 2000 data represent median household income at the block-group level from the population census. I compiled the 2007-2011 median household income records from the American Community Survey at the 2010 block-group level. The National Historical GIS (NHGIS) project provided the 1950 census, 2000 census, and the 2007-2011 American Community Survey tabular and spatial data

Table 6 - Analysis of the Influence of Redevelopment on Gentrification - Regression Results - Dependent Variable: Percentage Change in Tract-Level Real Median Income From 1950 - 2000			
Variables	Model1	Model 2	Model 3
Miles From City Hall	-27.75***	-27.33***	-27.43***
	(4.47)	(4.41)	(4.69)
Miles From Historical Marsh	144.50***	148.69***	105.55**
	(40.67)	(39.35)	(44.60)
Miles From Central Park	-17.99	-16.78	-20.22**
	(11.50)	(11.93)	(10.14)
Percent of Lot Area Redeveloped	-0.32		
	(0.62)		
Percent of Interior Building Area Redeveloped		0.12	
		(0.55)	
Percent of Lot Square Footage Redeveloped by Public Investment			-3.05***
			(0.70)
Percent of Lot Square Footage Redeveloped by Private Investment			0.35
			(0.92)
Percent of Lot Square Footage Redeveloped by Commercial Investment			1.15
			(1.09)
Constant	309.69***	290.55***	311.93***
	(49.83)	(50.10)	(53.69)
Observations	240	240	240
R^2	0.2298	0.2289	0.2687

Conley's corrected standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 7 - Regression Results - Tract-Level Analysis of Population Entrenchment - Dependent Variable: Percentage Share of Population Group Within 1940 Census Tracts in New York City

Variables	Population Groups													
	Italian		Irish		Polish		Russian		German		Black		Native White	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Percentage Share of Group in 1880	0.8202***	0.8833***	0.0412***	0.0335**	0.2134**	0.2334***	1.5822***	1.5603***	0.0028	0.0022	0.4105	0.2504	-0.8206***	-0.8096***
	(0.0940)	(0.0817)	(0.0155)	(0.0148)	(0.0855)	(0.0712)	(0.2440)	(0.2031)	(0.0155)	(0.0156)	(0.5736)	(0.5830)	(0.1674)	(0.1679)
Log (Mean Population Income in 1880)		-0.4200		-1.7836**		-2.2351**		-2.6506***		-1.0254		7.1678		13.2168
		(1.9617)		(0.8859)		(0.8668)		(0.9948)		(0.9153)		(18.0673)		(12.6563)
Constant	2.1559***	3.1931	-1.5075	6.5601**	0.7326***	7.9481***	0.8520***	9.6765***	3.9765*	4.6281	10.9361***	-11.3874	62.1710*	68.5865*
	(0.4708)	(6.3580)	(3.2801)	(2.9183)	(0.2269)	(2.8093)	(0.2206)	(3.2235)	(2.3314)	(2.9880)	(2.9527)	(58.5810)	(33.9620)	(40.5378)
R^2	0.2981	0.3112	0.0266	0.0399	0.0352	0.0637	0.2275	0.2016	0.0054	0.0051	0.0022	0.0013	0.0862	0.0825

These data include 262 tract observations. Conoly's corrected tandard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 8 - Location Summary Statistics of Non-Residential Sites in 1851			
Types of Sites	Obs	Percent Observed	
		At Shore	Not At Shore But On Marsh
Asylums	17	23.5	0
Banks	36	44.4	5
Breweries	15	20	8.3
Brick Makers	2	100	-
Church	229	7	8.5
Coal Yards	67	52.2	0
College	12	8.3	0
Corporate Headquarters	5	60	0
Courthouses	3	0	0
Train Depots	8	37.5	20
Distilleries	9	33.3	33.3
Entertainment	14	0	14.3
Factories	93	38.7	14
Fire Station Houses	53	18.9	2.3
Foundries	40	47.5	33.3
Hospitals	4	25	0
Hotels	111	38.7	5.9
Industrial Sites	49	83.7	12.5
Insurance Brokers	18	88.9	50
Markets	17	58.8	0
Newspaper Headquarters	6	16.7	0
Oil and Wax Works	5	80	0
Police Stations	53	34	17.1
Prisons	3	33.3	50
Refineries	8	87.5	0
Schools	84	27.4	13.1
Soap Manufacturers	10	40	0
Stables	20	15	5.9
Warehouses	6	100	-

Table 9 - Regression Results - Evaluating Influence of 1851 Manufacturers on the Rental Price of Housing
Dependent Variable: Log (Annual Rent)

Variables	1830 - 1860		1940	
	Model 3	Model 4	Model 3	Model 4
Miles From City Hall	-0.0486	0.1816*	0.0046	0.1581*
	(0.0338)	(0.1039)	(0.0056)	(0.0960)
Miles From Historical Marsh	0.2099	0.0342	0.4769***	0.2342***
	(0.1382)	(0.1772)	(0.0571)	(0.0535)
Miles From Central Park	-	-		-0.0279
				(0.0319)
Miles From Brewery		0.0690		-0.0392
		(0.1317)		(0.0831)
Miles From Coal Yard		0.1985		-0.0314
		(0.1994)		(0.0580)
Miles From Dry Dock		0.3107*		-0.1796
		(0.1794)		(0.1337)
Miles From Foundry		0.1430		0.1481***
		(0.1377)		(0.0397)
Miles From Hospital		-0.2515**		-0.1083**
		(0.1003)		(0.0476)
Miles from Oil Manufacturer		-0.4954**		0.1853
		(0.2436)		(0.1928)
Miles From Refinery		-0.0131		-0.0176
		(0.1656)		(0.1142)
Miles From Soap Manufacturer		0.0451		-0.0938
		(0.1135)		(0.0709)
Apartment Size	Y	Y	Y	Y
Spatial Controls	Y	Y	Y	Y
Housing Controls	Y	Y	Y	Y
Observations	782	782	2,414	2,414
R^2	0.4212	0.4483	0.6853	0.7281

Conley's corrected standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Figures

Figure 1 – 2000 Block Group-Level Median Contract Rent vs. Median Household Income on Manhattan Island

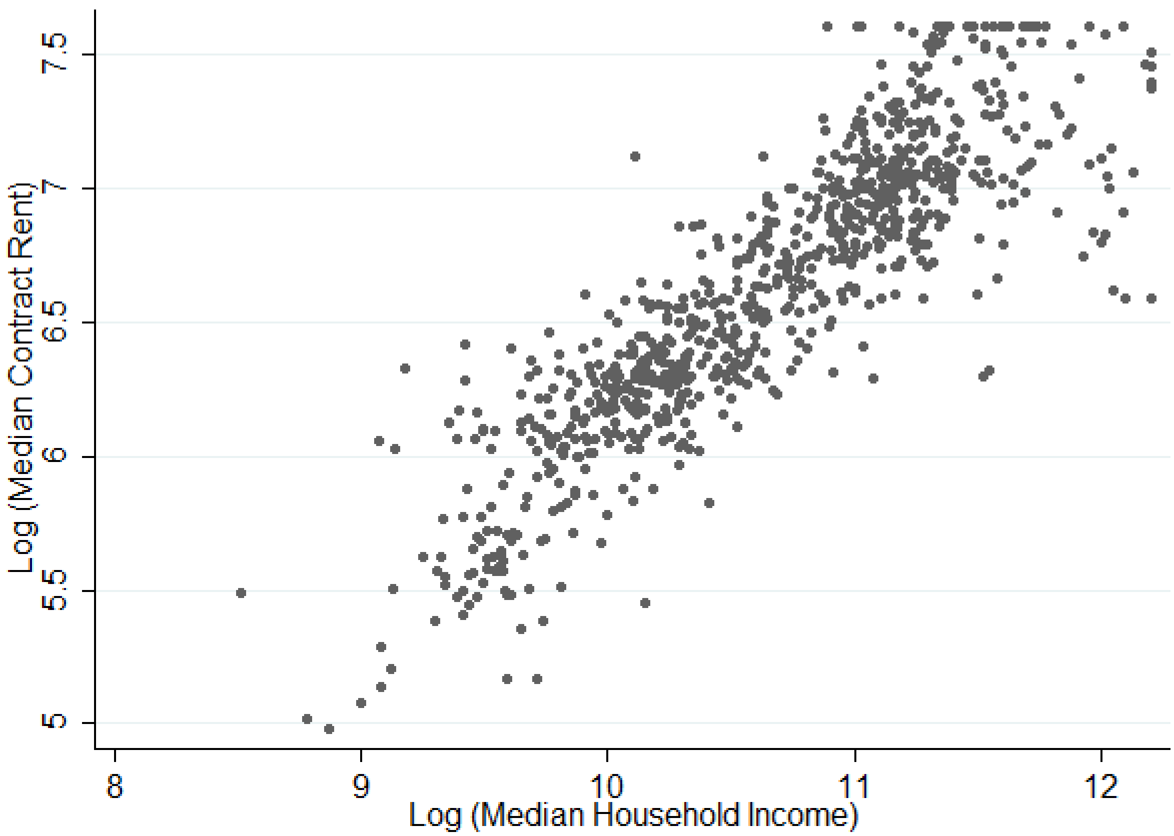


Figure 2 – The Location of Historical Marshes and 1940 Apartment Rental Prices

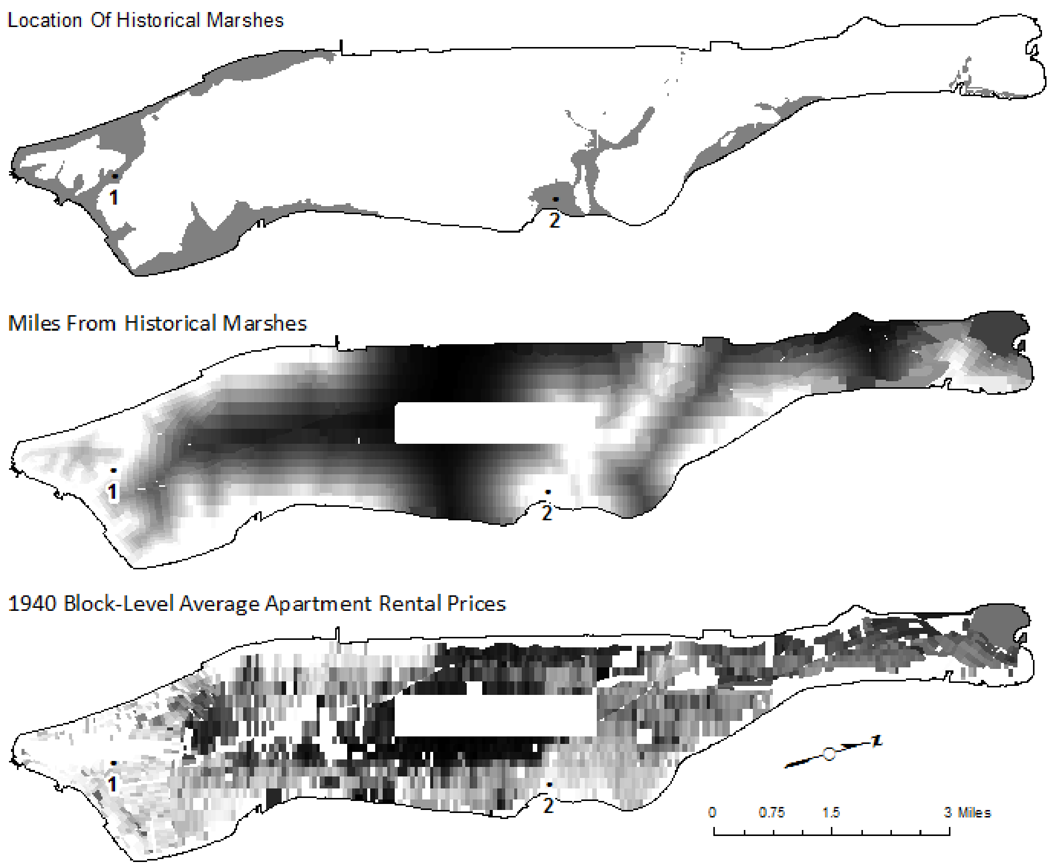
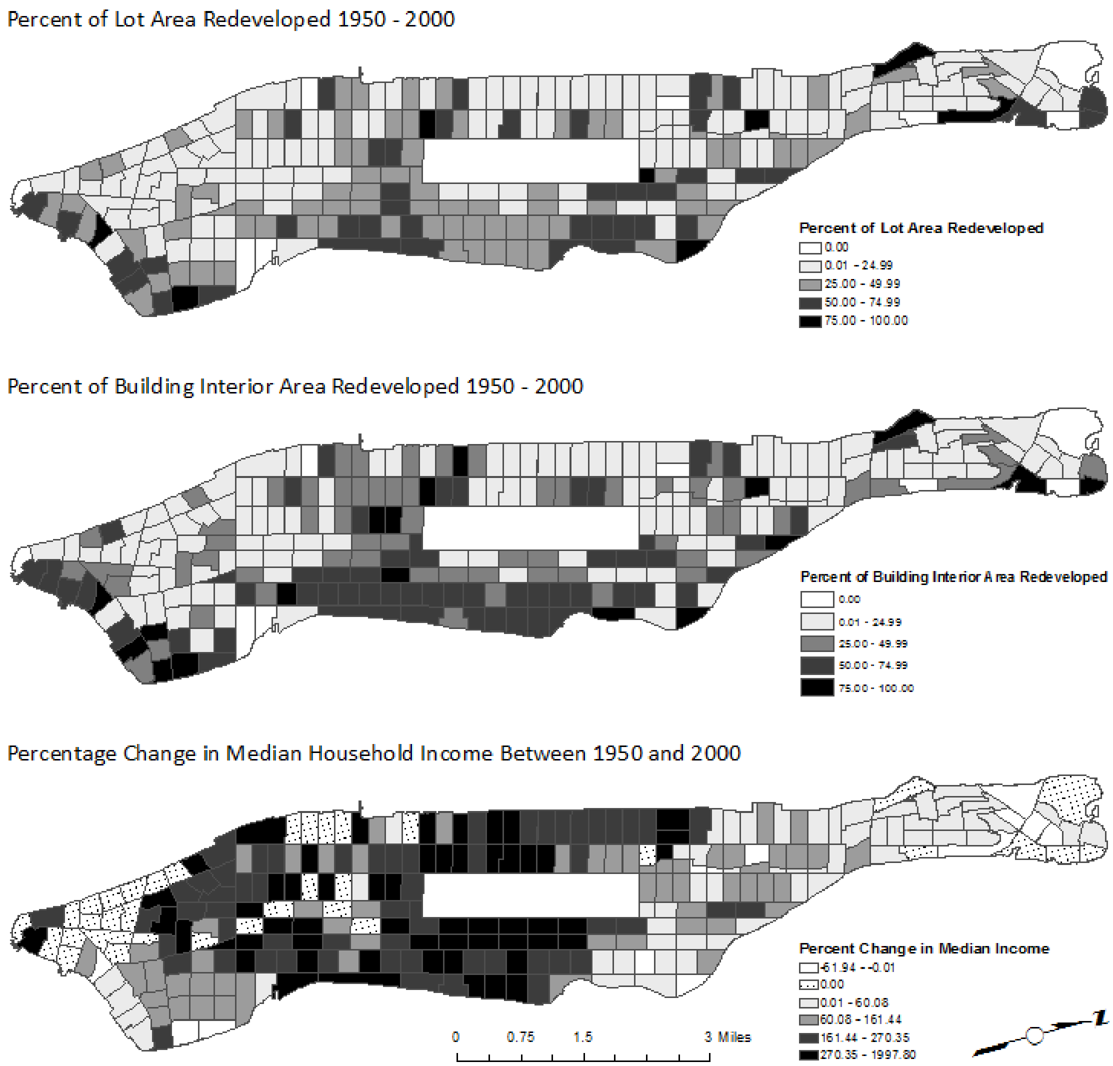
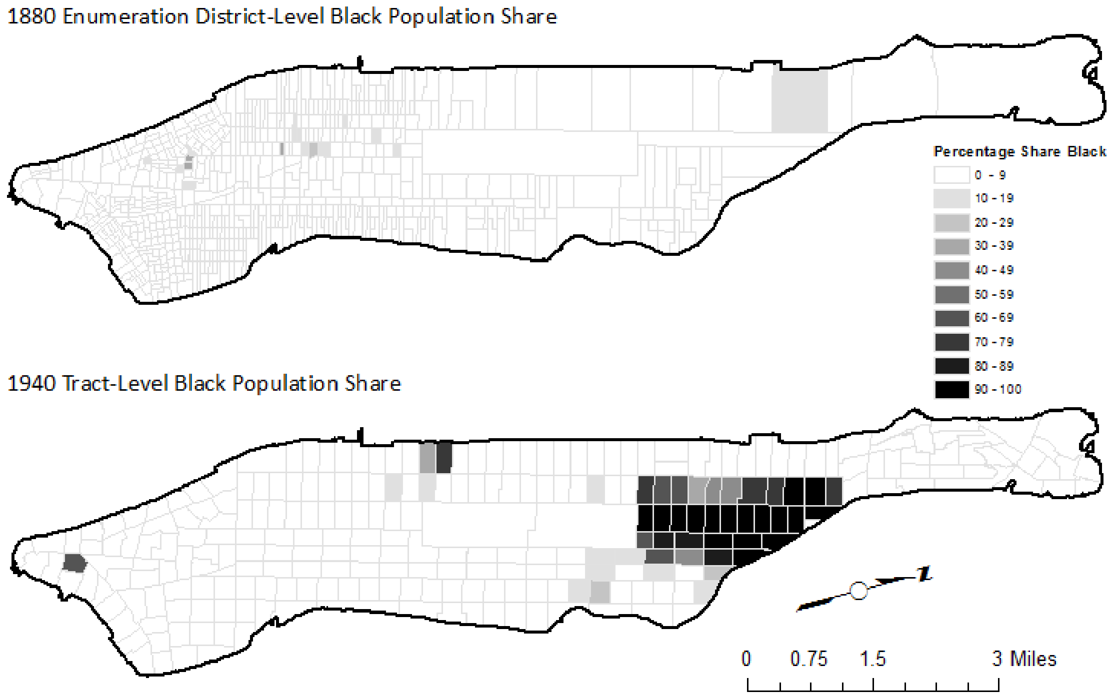


Figure 3 – The Influence of Redevelopment on the Change in Median Income 1950-2000



* dotted tracts do not contain 1950 income data

Figure 4 – Location of the Black Population in 1880 and 1940



Data Appendix

I use geographic information system (GIS) tools to generate representations of historical data in space and interact those data layers over time. I begin by drafting a set of historical streets centerlines that form the foundation from which reconstruct all subsequent data layers. These data are important because streets almost exclusively delimit the boundaries of administrative wards, assembly districts, census enumeration districts, and blocks. Accurate contemporary GIS data were not available at the start of the project, but would have been of limited use since contemporary streets no long follow historical street paths in many regions of the city. I utilized high resolution orthoimagery from the US Geological Survey to draft street centerlines for the streets with unmodified historical paths then georeferenced historical maps by the Bromley, Sanborn, and Perris fire insurance companies to draft historical street paths in razed districts. A comparison of my street reconstruction to newly released street centerline data created by the New York City Department of Technology and Telecommunications indicate that my streets achieve an accuracy tolerance of 10 feet in unmodified districts. Images from a 1924 aerial survey, which I found after the street reconstruction project, indicate that the historical streets file achieved similar accuracy in the razed districts.

The historical streets file allowed me to spatially link and extract a record of the initial location areas with poor natural drainage from the 1874 topographic survey of Manhattan Island by Egbert Viele, which details the initial water courses and historical marshes throughout Manhattan Island. The 1874 map incorporates revisions to of his 1865 survey and represents the best surviving record of the initial location of historical marshes. I georeference the topographic survey block-by-block across the entire island to correct for inaccuracies in the historical surveying tools and then build a polygon shapefile delimiting the space enclosed by historical marshes.

I then geocode the address-level rental data from 1830 through 1860. This process must be conducted manually because contemporary automated geocoders cannot accurately locate historical addresses because the building numbering system changed since the end of the study period. In order to locate the properties I used my historical streets to georeference maps of the first two building surveys of Manhattan Island by William Perris (1852) and Matthew Dripps (1851, and 1852) and further overlaid those data upon the contemporary building footprint and lot boundary shapefiles for Manhattan Island[9] . The combination of those data allowed me to locate the historical addresses from the 1830 through 1860 sample of rental price data to within 10 feet of the historical property. Summary statistics of those data are available in Table 1. I used the same address-level geocoding methodology to extract the locations of every non-residential building in the Perris and Dripps surveys. I verified the completeness of those data by checking a sample of listings from contemporaneous city directories, which revealed that the sample listings were already included in the geocoded data.

The remaining spatial data were more straightforward to reconstruct. I used the historical street centerlines to trace the 1880 enumeration districts using the boundary narratives distributed by the census bureau, which I double checked using addresses listed in the manuscript census. My historical street centerlines also guided the reconstruction of the block polygons representing the 1940 federal housing census data.

[9] Data generated by the New York City Department of Information Technology and Telecommunications