LUISS Guido Carli & EIEF

Economics and Finance - Rome Masters in Economics

Chair of Econometric Theory

The long shadow of racial segregation: Does it affect the infrastructure development of a city?

Professor Franco Peracchi
Supervisor

Professor Marco Lippi
Co-Supervisor

Pablo Mones
Candidate

Academic Year 2022/2023

The long shadow of racial segregation: Does it affect the infrastructure development of a city?

Pablo Mones

Abstract: This paper aims to disentangle the mechanisms underlying the well-documented association between race and the location of waste management facilities (WMFs). To do so, I first study whether WMFs were more likely to be located in neighborhoods with a higher share of black population after controlling for relevant covariates. Secondly, I analyze the socioeconomic consequences following the introduction of WMFs. Using historical data from the 1920-1940 U.S. Censuses and GIS data from the Urban Transition Project, I focus on the cities of Baltimore, Boston, Chicago, Cincinnati, Houston, and Nashville in the United States.

The results suggest that, contrary to expectations, there was no targeting of predominantly black neighborhoods for WMF placement after accounting for other geographic and socioeconomic factors. Instead, a more flexible model reveals an inverted U-shaped relationship, indicating that neighborhoods with mixed racial compositions were most affected by the presence of WMFs. These findings are consistent with the potential existence of blocking coalitions based on race, implying that homogeneous racial neighborhoods could resist the establishment of WMFs.

Furthermore, the study explores the consequences of WMF introduction on the affected regions. I find a decrease in urban development, measured by a decline in population density, following the construction of WMFs. Additionally, there is evidence of an increase in the share of black population in the areas where WMFs were introduced. These findings are consistent with a white flight from the affected neighborhoods.

Contents

1 Introduction

Literature has demonstrated that exposure to environmental risks is unequally distributed on the basis of race and socioeconomic status. Studies have established an association between nearby residents' racial and economic characteristics and the placement of hazardous environmental infrastructure (Bullard, 2008). These findings contribute to the ongoing discourse on environmental justice, revealing that marginalized social groups bear the consequences of the adverse effects of environmental pollution, exacerbating pre-existing structural inequalities within society.

Landfills and waste incinerators are essential for the proper functioning of every city. Even though they contribute to increase the inhabitants' overall social welfare, people living near these facilities often experience negative consequences. Various factors influence decisions regarding the location of waste management facilities (WMFs). Technical considerations play a role, and they depend on the geographical characteristics of the specific city under consideration. Additionally, political and socioeconomic factors come into play, as the allocation of WMFs entails winners and losers.

The potential negative externalities associated with WMFs have been widely studied in the literature. Waste incineration, for instance, leads to increased exposure to pollutants such as particles, acidic gases, aerosols, metals, and organic compounds. This exposure has been linked to elevated mortality rates and an increase in emergency hospital admissions (Dockery et al. (1994); Katsouyanni et al. (1997)). Furthermore, studies by Elliott et al. (1996) and Elliott et al. (2000) have identified a correlation between the distance from waste incinerators and the probability of developing certain types of cancer.

Exposure to dumps or landfills has also been proven to pose health risks to the surrounding inhabitants. Vrijheid (2000) conducted an in-depth literature review about the health effects of residing near waste landfills. The review revealed a link between exposure to hazardous waste sites and excess deaths from various types of cancers, as demonstrated by Griffith et al. (1989). Additionally, Berry and Bove (1997) discovered that infants born from parents who reside near landfills had a statistically significant lower average birth weight compared to the rest of the population.

The relationship between the location of WMFs and the racial composition of the surrounding areas has been extensively studied in the literature. The U.S. General Accounting Office (GAO, 1983) published one of the first technical reports on this topic in response to protests by environmental justice movements. The report found a positive association between the location of four hazardous waste landfills in the southern United States and the proportion of African-Americans living there. Subsequent studies have confirmed this trend, with a summary of the results available in Mohai et al. (2009).

However, establishing an association between race and the location of landfills and waste incinerators does not indicate a discriminatory process against minority populations. This scenario presents a classic 'Chicken and the Egg' situation, where it is unclear whether the racial composition caused the infrastructure's placement or vice versa. On the one hand, it is possible that authorities and private companies deliberately chose to locate WMFs in neighborhoods predominantly inhabited by racial minorities, resulting in a positive association in the data. On the other hand, it could be the case that the infrastructure was placed irrespective of race, but subsequent socioeconomic forces led to an increase in the share of black population. Alternatively, there may be other relevant socioeconomic and geographical factors simultaneously influencing both the location of WMFs and the settlement decisions of black inhabitants, thereby leading to the positive association observed in the data.

This study aims to disentangle these mechanisms and address two key research questions. Firstly, I examine whether WMFs were situated in neighborhoods with a higher historical proportion of black residents while controlling for a relevant set of covariates that capture other factors influencing the location of WMFs. Secondly, I explore the changes in socioeconomic characteristics within regions following the installation of a WMF. To address these questions, I focus on a group of cities in the United States, utilizing data from the U.S. Censuses conducted in 1920, 1930, and 1940, as well as GIS data from the Urban Transition Project (UTP) to map the census data onto geographical space. Additionally, I collect historical data on the locations of dumps, landfills, and waste incinerators operating before 2000 in these cities.

Regarding the first question, Louis (2004) argues that most waste management decisions were made at the municipal level during the 20^{th} century. According to the author, waste management had become an 'institutionally organized, technology-focused, municipally operated service' by the 1930s. By the 1960s, sanitary landfills had become the predominant method of municipal refuse disposal. Therefore, it is reasonable to assume that regulated WMFs were established under the supervision of municipal authorities. In contrast, unregulated landfills may have emerged out of necessity, negligence, lack of regulation, or ineffective political action. In many cases, these unregulated landfills were subsequently acquired by cities and formalized as official dumping sites.

My main hypothesis posits that during the study period, when minorities were politically underrepresented (Karnig, 1976), authorities may have chosen to impose the negative externalities associated with waste management infrastructure on these minorities (Mohai et al., 2009), as the political costs were lower than those associated with placing these facilities in predominantly white neighborhoods. If this hypothesis holds, it would be expectable to find a higher likelihood of observing a WMF in a neighborhood predominantly inhabited by black residents compared to a neighborhood with similar characteristics primarily inhabited by the white residents.

The magnitude of these effects may vary depending on city-specific characteristics. For instance, one might hypothesize that the effect could be less pronounced in cities with a higher percentage of black population, as they may have possessed greater representation power. Moreover, the political alignment of the municipal government could have influenced this decision. Wolfinger (1965) shows that a person's ethnicity is related to their likelihood of supporting a particular political party, even after controlling for other characteristics. Therefore, politicians may have chosen to place WMFs closer to neighborhoods associated with ethnicities that were unfavorable toward them to maintain their primary voter base content. Lastly, it is plausible that this effect was stronger in cities where the black population was not economically or culturally integrated.

This paper does not find robust evidence supporting the aforementioned hypothesis after controlling for relevant covariates that explain both the location of WMFs and the racial composition of neighborhoods. Houston, is the only city where there is a statistically significant and positive effect from the proportion of black population to the location of WMFs. In other cities, the effect is relatively weaker. The general pattern indicates an inverted U-shaped relationship between the variables, suggesting that neighborhoods with a mixed-race composition were the most affected.

Regarding the second question, this paper provides evidence indicating that introducing a WMF leads to an increase in the share of black population. Hypothetically, this could be attributed to two potentially complementary mechanisms: higher-income individuals may have chosen to relocate, but at the same time, those with lower economic resources may have moved in to take advantage of the decrease in land costs resulting from the WMF introduction (Hite et al., 2001). Historically, the black population in the United States tended to have lower income levels than the white population. Therefore, after establishing a WMF, white residents may have decided to leave the area, while black residents may have opted to settle there. I refer to the first case as the 'white flight' hypothesis and the second as the 'black settlement' hypothesis.

The estimation results indicate an increase in the proportion of black residents following the introduction of a WMF. Additionally, there is a significant decrease in population density, and individuals with worse economic outcomes (measured by unemployment) appear to have settled in the affected areas. These findings support the white flight hypothesis, suggesting that people prefer to leave the area after a WMF is introduced. However, as white individuals were, on average, wealthier than black individuals, they were more capable of doing so, resulting in a lower population density and an increase in the proportion of black residents.

The paper is organized as follows: Section 2 presents the data used for analysis and the methodology employed to ensure data comparability across years. Section 3 examines whether WMFs were located in neighborhoods with a higher proportion of black residents after accounting for various characteristics. Section 4 presents the empirical strategy and results regarding the effect of introducing a WMF, while Section 5 concludes the paper.

2 Data

To investigate this phenomenon, I utilize the IPUMS data set (Ruggles et al., 2021) derived from the censuses of 1920, 1930, and 1940. Additionally, I employ GIS data from the Urban Transition Project (Logan et al., 2011) to obtain historical Enumeration Districts (ED) maps. It is important to note that the selection of cities for study is contingent upon the availability of maps within the Urban Transition Project. Consequently, this study focuses on four Northern cities—Baltimore (MD), Boston (MA), Chicago (IL), Cincinnati (OH)—as well as two Southern cities—Houston (TX) and Nashville (TN). Therefore, it is imperative to recognize that the findings of this study may not generalize to the broader universe of US cities.

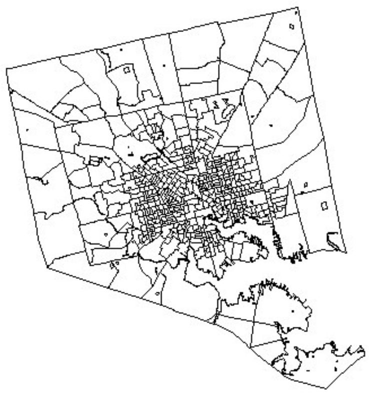
To merge the two databases, I employ the Enumeration District (ED) variable. This variable represents geographic areas designed to allow a census taker (enumerator) to visit every house in the district within two-weeks (in rural areas, the time allowed was one month). It is unique within the Counties of each State, and it enables localizing each person from the census in their respective ED, thereby providing a granular measure of location within the given time frame. The dataset derived from this approach yields detailed information at both the individual and household levels, encompassing multiple characteristics. Consequently, it permits a comprehensive understanding of the socioeconomic attributes of the EDs within each city.

Throughout this project, I focus on infrastructure related to waste management. The compilation of data on the locations of waste management facilities is accomplished by extracting information from various historical records, state and municipal technical reports, newspapers, and government bodies such as the Environmental Protection Agency (EPA). It is important to acknowledge that each information source may possess incomplete details regarding individual establishments, necessitating data from multiple sources to gather all the required information. Furthermore, in the case of irregular landfills—those developed without the supervision of municipal authorities—the precise commencement and cessation dates of operations remain unclear due to the absence of official records. Therefore, the sources explored in this study provide approximate dates based on historical reports. For further details on the infrastructure data, please refer to Appendix A, while the complete database can be made available upon request.

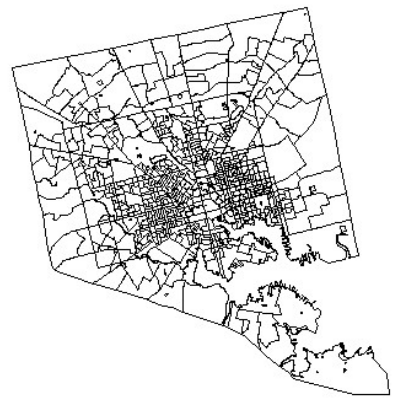
2.1 Synthetic neighborhoods

An essential feature of our data is that EDs are not constant throughout the censuses, so they are not comparable across years. In order to solve this problem, I follow the methodology proposed by Shertzer and Walsh (2019) to work with stable units of analysis. The procedure consists of creating a hexagonal grid of Synthetic Neighborhoods (SN) and then mapping information from Enumeration Districts to Synthetic Neighborhoods.

Figures 1 and 2 provide an example of why this is needed. As Baltimore's EDs are not comparable between 1920 and 1930, I create a hexagonal grid that is constant across years, thus providing stable units of analysis.



(a) 1920



(b) 1930

Figure 1: Baltimore's ED in 1920 and 1930

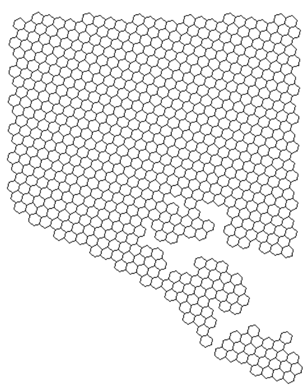


Figure 2: Baltimore's Hexagonal grid

The number and size of SNs varies across cities, and SNs are chosen to have approximately the same area as a city's average ED in 1930. After this grid is created, I follow the procedure of Shertzer and Walsh (2019) to map information from the EDs to the SNs.

Let i denote the SNs and j the EDs. First, I compute the spatial share of ED j included in SN i as the ratio between the intersection of the regions composed by ED j and SN i, and the area of ED j

$$a_{i,j} = \frac{area_{i \cap j}}{area_j}$$

Note that this fraction is always smaller or equal than 1, with the latter case occuring when ED j is fully included in SN i. Then, I assign the fraction $a_{i,j}$ of the total population of ED j, denoted as P_j, to SN i. After considering all the EDs that make up SN i, the total population of SN i is given by:

$$P_i = \sum_j a_{i,j} P_j$$

Note that this measure of population represents the expected population size of hexagon i under the assumption that population within each ED is equally distributed across the geographical space. Finally, the rest of the socioeconomic variables X are computed as a weighted average (on the basis of the area included and the population size) between the outcomes of all the EDs that make up SN i:

$$X_i = \sum_j X_j w_{i,j}$$

where $w_{i,j} = \frac{a_{i,j} P_j}{\sum_j a_{i,j} P_j}$ is the weight of ED j on the outcomes of SN i.

A graphical intuition behind the explained procedure is shown in Figure 3. After creating this hexagonal grid, I can study a city's racial division and socioeconomic characteristics at a micro level with stable units of analysis. This enables me to use and compare the information from the three censuses, 1920, 1930, and 1940.

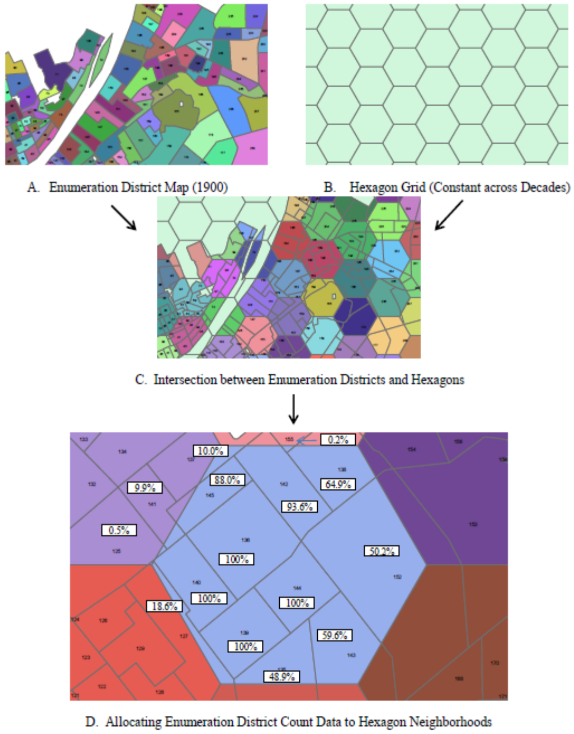


Figure 3: Procedure for creating the SNs. Source: Shertzer and Walsh (2019)

2.2 Area fixed effects

It is crucial to incorporate area-fixed effects into the analysis to account for the potential area-specific characteristics of the city in which SNs are located. This requires dividing the cities into distinct zones, where ideally, EDs within each zone exhibit similar characteristics.

I employ the ward variable available in the census databases and in the Urban Transition Project to address this concern. This variable identifies the political ward in which households were enumerated and represents a higher level of geographic aggregation than EDs, grouping geographically proximate EDs. Consequently, it allows for a division of the city into a disjoint number of areas. It is important to note that the ward variable is available for all cities under study except Houston. Given its variation across censuses, I adopt the wards observed in 1930 as the basis for delineating distinct areas within the city. As an example, Figure 4 displays the EDs of Baltimore in 1930 colored by their respective wards.

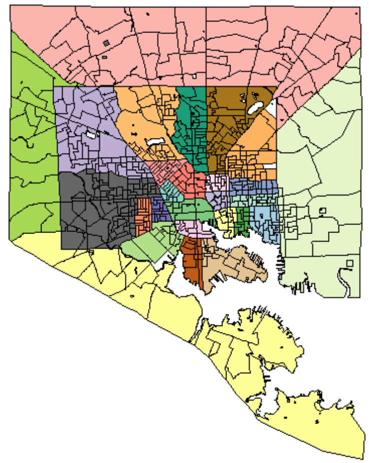


Figure 4: Baltimore's Wards in 1930

However, SNs do not inherently possess a ward, as they encompass multiple EDs. To overcome this, if all the EDs within a particular SN belong to the same ward, I assign that ward to the SN. Conversely, if the EDs constituting an SN span multiple wards, I assign the SN to the ward where the highest percentage of its constituent EDs is located.

3 Are WMFs installed in zones with a higher share of black population?

3.1 Empirical strategy

3.1.1 Framework

I estimate a linear probability model (LPM) for each city to examine the influence of the racial composition within Synthetic Neighborhoods on their likelihood of being affected by a waste management facility. Due to variations in institutional and geographical characteristics, regression analyses need to be city-specific to account for differences in allocation patterns. Ideally, it is desirable to study each SN's racial composition before introducing each WMF. However, since WMFs were implemented at different periods, incorporating corresponding dates into a single LPM becomes unclear. Therefore, I adopt a common specification where all variables are evaluated within the same period to ensure comparability across cities.

To address this issue, I evaluate the treatment and control variables in 1930 while considering the construction of landfills and waste incinerators built from 1930 until 2000. This approach allows for examining WMF introductions spanning different time periods within a single regression, which is both parsimonious and analytically convenient. However, it assumes that the racial composition in 1930 serves as a good predictor of WMF locations in $1930 + k$. While some readers may question the extent to which the racial composition in 1930 influences the placement of a WMF in, for instance, 1970, I provide a comprehensive discussion in subsection 3.1.3 to justify this reasoning and clarify the mechanism through which I believe the desired effect can be captured. Moreover, this strategy helps mitigate potential issues of reverse causality, as it would be hard to claim that the location of a landfill in $1930 + k$ causes the share of black population or any other socioeconomic characteristic in 1930.

To capture the relationship between WMFs allocation and ethnic composition, I adopt a selection on observables strategy. I control for the socioeconomic characteristics of each SN, their geographic characteristics, the presence of previous WMFs in the area, and attempt to account for unobservable SN characteristics using area-fixed effects determined by the ward variable, as described in subsection 2.2. Appendix B provides a list containing the variables included in the analysis. The data obtained from IPUMS is at the individual level, and I follow the approach outlined in section 2.1 to map individual-level data onto SNs.*

My identification strategy relies on the fact that after controlling for area-fixed effects and SN individual characteristics, the racial allocation among SNs of the same ward is uncorrelated with the error term. It is vital to note that the controls include several features of SNs that are crucial for upholding this assumption. Additionally, the geographical characteristics of SNs are incorporated to enhance the understanding of whether a particular SN is naturally more suitable for hosting a waste incinerator or a landfill.

A general formulation of the LPM that will be employed is presented as follows in Equation 1:

$$Y_{i,w,c} = \alpha_{w,c} + h_c(B_{i,w,c,30}) + g_c(\mathbf{z_{i,w,c,30}}) + \varepsilon_{i,w,c} \tag{1}$$

*In the case of the variable 'value of the house,' I trim the distribution to mitigate potential issues arising from outliers associated with data coding errors.

where i denotes the SN index, w the ward, and c the city. $Y_{i,w,c}$ is an indicator variable taking value 1 if there was a WMF in a radius of 1 kilometer of SN i in ward w of city c installed after 1930 and 0 otherwise. $B_{i,w,c,30}$ is the share of black population in SN i in ward w of city c in 1930, $\alpha_{w,c}$ represent area (ward) fixed effects, and $\mathbf{z_{i,w,c,30}}$ are the controls evaluated at 1930. $h_c(B_{i,w,c,30})$ and $g_c(\mathbf{z_{i,w,c,30}})$ are city-specific unknown and potentially complicated functions that express the relationship between the location of WMFs and the share of black population, and the controls, respectively. Note that the units of analysis are the SNs so, there are as many observations as SNs for each city.

An essential consideration in my empirical strategy is how to incorporate the share of black population into the model, as the results and their interpretation depend on this decision. Imposing a restriction on the functional form of $h_c(B_{i,w,c,30})$ allows for easier interpretation of the results but limits flexibility in the analysis. One possible approach to address this issue is to introduce the share of the black population as a linear term in the regression. This approach is straightforward and facilitates the interpretation of the results, albeit with the restriction that the marginal treatment effect is constant across all possible values of the share of the black population. Following this approach, I express the model specified in Equation 2 for each city as follows:

$$Y_{i,w,c} = \beta_c B_{i,w,c,30} + \alpha_{w,c} + g_c(\mathbf{z_{i,w,c,30}}) + \varepsilon_{i,w,c} \tag{2}$$

The primary parameter of interest in this regression analysis is denoted as β_c. This parameter is specific to each city and can vary based on their characteristics. It quantifies whether the racial composition of an SN influences the likelihood of that SN hosting WMFs. If β_c is positive, it suggests an increase in the probability of observing a WMF in a given SN when the share of the black population increases, all else being equal.

As previously mentioned, introducing the treatment variable as a linear regressor assumes a constant marginal treatment effect across all possible values within its range. To account for potential non-linearities in the relationship between $Y_{i,w,c}$ and $B_{i,w,c,30}$, while maintaining linearity in parameters, a more flexible approach is adopted by partitioning the treatment variable into group dummies.

To implement this approach, the observations are divided into four groups using the same cutoff points for all cities to ensure comparability in parameter interpretation. This enables the definition of the following indicator variables:

- $B^{G1}_{i,w,c,30} = \mathbb{1}_{0 \leq B_{i,w,c,30} < 5\%}$
- $B^{G2}_{i,w,c,30} = \mathbb{1}_{5\% < B_{i,w,c,30} \leq 15\%}$
- $B^{G3}_{i,w,c,30} = \mathbb{1}_{15\% < B_{i,w,c,30} \leq 30\%}$
- $B^{G4}_{i,w,c,30} = \mathbb{1}_{B_{i,w,c,30} > 30\%}$

The partitioning cutoffs were chosen to ensure a sufficient number of observations in each group for every city. Moreover, the results presented in the subsequent subsections are robust to variations in the cutoff values. By including these indicator variables as regressors, the model can accommodate a more flexible alternative specification, as shown in Equation 3:

$$Y_{i,w,c} = \beta_{c,2} B^{G2}_{i,w,c,30} + \beta_{c,3} B^{G3}_{i,w,c,30} + \beta_{c,4} B^{G4}_{i,w,c,30} + \alpha_{w,c} + g_c(\mathbf{z_{i,w,c,30}}) + \varepsilon_{i,w,c} \tag{3}$$

In this model I omit the variable $B^{G1}_{i,w,c,30}$. Hence, the rest of the parameters represent the differential change in the probability of being affected by a WMF with respect to the set of neighborhoods with low share of black population. In case the hypothesized relationship between the variables is true, $\beta_{c,p}$ should be greater than 0.

3.1.2 Control variables

Until now, I have not made any assumptions regarding the functional form through which the control variables may affect the outcome variable. To address this concern, I adopt the methodology proposed by Belloni et al. (2014), which provides a framework for performing valid inference on the linear-in-parameters treatment variable $B_{i,w,c,30}$ affecting the outcome variable $Y_{i,w,c}$, given a set of controls $\mathbf{z_{i,w,c,30}}$ that may have an unknown and complex relationship with the outcome through the function g_c.

The approach suggested in the paper involves approximating $g_c(\mathbf{z_{i,w,c,30}})$ by a linear expansion of the original controls $\mathbf{x_{i,w,c,30}} = P(\mathbf{z_{i,w,c,30}})$, that includes the constant, quadratic terms, interactions, and dummies. This approximation allows maintaining linearity in the parameters while capturing the potential nonlinear relationship between the controls and the outcome:

$$g_c(\mathbf{z_{i,w,c,30}}) \approx \mathbf{x}_{\mathbf{i,w,c,30}}^{\top}\delta_c$$

where $\mathbf{x_{i,w,c,30}}$ is a p-dimensional vector containing information about the expanded controls $P(\mathbf{z_{i,w,c,30}})$, δ_c is a city-specific p-dimensional vector with the coefficients associated with each term of the expansion. The approximation of $g_c(\mathbf{z_{i,w,c,30}})$ by this linear expansion has associated an approximation error $r_{i,w,c,30}$.

This approach is convenient when dealing with potential non-linear relationships between the control and outcome variables. However, it comes at the cost of increasing the model's dimensionality, as p can be quite large. A common but naive solution to address this challenge is to apply traditional machine learning techniques, such as LASSO, in order to reduce the dimensionality of the model specified in Equations 2 or 3 and then re-estimate the equations anew by Ordinary Least Squares (OLS) using the selected coefficients obtained through the LASSO procedure. This technique is known as 'post-LASSO.'

Nevertheless, following this approach might not be suitable when the primary objective is to infer the treatment's effect on the outcome variable. This concern has been extensively discussed and demonstrated by Belloni et al. (2014). The authors argue that a simple post-LASSO estimator can lead to biased parameter estimates, primarily due to the potential omitted variable bias that may arise from excluding a relevant regressor highly correlated with the treatment. Moreover, a basic post-LASSO procedure fails to accurately estimate standard errors since it does not account for the uncertainty resulting from the model selection step.

To address these limitations, Belloni et al. (2014) propose a post-double-selection estimator that selects the included covariates through two separate regressions. In the first step, the LASSO procedure is applied to identify the relevant controls that explain the treatment variable. Subsequently, another LASSO procedure is employed to select the relevant controls for explaining the outcome variable in the absence of treatment. Finally, an OLS regression where the outcome variable is regressed on the treatment variable and the selected controls from both previous regressions, is conducted. Additionally, the researcher can include another set of controls based on *ad-hoc* knowledge about the problem. This comprehensive procedure allows for valid inference regarding the parameter associated with the treatment. It also considers the uncertainty arising from the model selection step when estimating the standard errors. Under regularity conditions, the estimator is consistent and asymptotically normal.

For further details on the methodology proposed by Belloni et al. (2014) and the underlying assumptions, I refer interested readers to Appendix C.

After applying this procedure, the models for each of the two proposed specifications for $B_{i,w,c,30}$ are represented by the following equations:

$$Y_{i,w,c} = \beta_c B_{i,w,c,30} + \alpha_{w,c} + \mathbf{x}_{\mathbf{i,w,c,30}}^{\top} \delta_c^1 + \varepsilon_{i,w,c} \tag{4}$$

$$Y_{i,w,c} = \beta_{c,2} B_{i,w,c,30}^{G2} + \beta_{c,3} B_{i,w,c,30}^{G3} + \beta_{c,4} B_{i,w,c,30}^{G4} + \alpha_{w,c} + \mathbf{x}_{\mathbf{i,w,c,30}}^{\top} \delta_c^2 + \varepsilon_{i,w,c} \tag{5}$$

where δ_c^1 and δ_c^2 are city-specific p-dimensional vectors of coefficients associated with the expanded controls for both proposed specifications. The coefficients in $\delta_{c,j}^k$ will be non-zero if the methodology proposed by Belloni et al. (2014) suggests that control j should be included in specification k. Additionally, I include the area-fixed effects in the model.

It is important to acknowledge that employing a Linear Probability Model (LPM) may present certain drawbacks. For instance, an LPM may struggle to capture potential non-linear relationships between variables, particularly when values approach the distribution extremes. Furthermore, an LPM cannot ensure that the predicted probabilities $\hat{Y}_{i,w,c}$ are bounded within the range of [0,1]. While these limitations exist, the LPM generally performs well for values near the mean and, under regularity conditions, guarantees consistent estimation of the parameters of interest. Thus, considering that the primary objective of this paper is to infer the value of the parameters associated with race rather than doing prediction, I believe that the advantages of the LPM in terms of inference outweigh its potential limitations.

Moreover, it is crucial to note that this empirical strategy aims to capture a measure of direct discrimination, which differs from discrimination itself. Following the insights of Bohren et al. (2022), total discrimination can be divided into two categories: direct and systemic discrimination. Direct discrimination refers to instances where individuals (in this case, neighborhoods) with identical characteristics but different races experience disparate rewards or punishments (in this case, the presence of landfills or waste incinerators). On the other hand, systemic discrimination encompasses disparities that arise indirectly through race-based differences in the distribution of characteristics that are not specific to a particular group but still influence outcomes among equally qualified individuals.

By including in the regression covariates potentially highly correlated with being black (such as labor performance, unemployment, the share of house owners, and house values), I face a limitation in capturing the effect of total discrimination. These covariates, while informative, are not only associated with being black but are also consequences of past discrimination. Let's consider an example to illustrate this point: suppose that land value is the sole factor determining the placement of WMFs, rendering all other information irrelevant once I control for this variable. In a scenario of economic inequality between white and black individuals, given the historical discrimination they have faced, the house value of a white individual would be higher than that of a black individual. Even if the government were to locate WMFs considering solely land value, it would still perpetuate systemic discrimination. Therefore, finding statistically significant results would indicate the presence of direct discrimination in the siting of WMFs. Conversely, the absence of statistically significant results would suggest the absence of direct discrimination. However, the proposed methodology fails to capture the effect of systemic discrimination and, thus, cannot conclude the presence or absence of total discrimination.

3.1.3 Segregation stability

To establish an economic interpretation of the findings in this paper, it is crucial to understand how the racial composition in 1930 could impact the placement of WMFs in subsequent decades, as it is reasonable to believe that the municipal authority would look at racial composition at $1930+k$ to place a WMF in $1930+k$. This explanation relies on the assumption that settlement patterns in the cities under study remained relatively stable throughout the study period. If this assumption holds, the racial composition of a city in 1930 will influence its racial composition in the following decades due to path dependence. Consequently, the current racial composition may affect the decision-making process regarding the siting of WMFs. This theoretical framework provides a plausible channel through which the effect of historical racial composition on infrastructure location can be claimed.

The stability of black settlements in US cities is widely acknowledged in the academic literature (see Taylor (1979); Sharkey (2013)). To assess the reasonableness of this assumption in this specific case, I examine whether the racial settlement patterns of the black population remained stable in each city during 1920, 1930, and 1940. An initial evaluation of this issue can be conducted by studying maps of each city that illustrate the dynamics of black population settlement patterns. These maps, presented in Figures A.1 to A.6 in Appendix D, reveal the evolution of ethnic settlements throughout the study period and help identify areas predominantly inhabited by black residents. These maps suggest that the assumption of settlement stability holds true. While some changes in racial configuration are observed, they tend to occur gradually and exhibit a high degree of persistence in the concentration of the black population in specific areas of the cities.

Additionally, population data extracted from the censuses for each SN in each city, covering 1920, 1930, and 1940 for Northern cities and 1930 and 1940 for Southern cities, are available. This data allows me to construct a panel dataset encompassing these years. It is possible to investigate whether the past racial composition of an SN serves as a good predictor of its present racial composition for the cities under study. Consequently, I estimate the following regression for each city:

$$B_{i,c,t} = \rho_{0,c} + \rho_{1,c} B_{i,c,t-10} + \delta_{c,t} + \varepsilon_{i,c,t} \tag{6}$$

where $B_{i,c,t}$ represents the share of black individuals in SN i of city c at time t, and $\delta_{c,t}$ is a time-fixed effect (only for northern cities). Values of $\rho_{1,c}$ closer to 1 will provide evidence in favor of the aforementioned assumption, as they will reflect persistence in the racial composition of the SNs.

The results of regression 6 for the different cities are presented in Table 1. It is evident from the table that the settlement patterns of the black population appeared to be stable throughout our study period, as indicated by the ρ coefficient being close to 1. This result remains robust when weighting each observation by the population size of the SN or not (see Table A.2 in Appendix D).

	Baltimore	Boston	Chicago	Cincinnati	Houston	Nashville
$\hat{\rho}_{0,c}$	-0.001	0.001	0.019^{***}	0.006^{**}	-0.013	0.000
	(0.003)	(0.001)	(0.002)	(0.003)	(0.011)	(0.006)
$\hat{\rho}_{1,c}$	1.035^{***}	1.101^{***}	1.095^{***}	1.165^{***}	0.926^{***}	0.992^{***}
	(0.017)	(0.012)	(0.009)	(0.019)	(0.040)	(0.015)
$\hat{\delta}_{1940}$	0.019^{***}	0.000	-0.020^{***}	-0.007^{**}	-	-
	(0.004)	(0.001)	(0.002)	(0.004)	-	-
R^2	0.73	0.87	0.75	0.83	0.78	0.95
Num. obs.	1310	1222	4776	762	154	213

Note: Unweighted regression for Equation 6 specification for the percentage of Black population in each SN for each city under study. Standard errors are reported in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 1: Results of unweighted stability specification

3.2 Preliminary analysis and descriptive statistics

As mentioned before, I only use data from the 1930 census for this part of the analysis. Given the geographical nature of the problem, it is helpful to understand how the socioeconomic characteristics of each city are distributed across the geographic space during that time. By combining this information with the location of waste management facilities that started their operations between 1930 and 2000, it is possible to gain initial insights into the possible answer to the research question.

Presumably, certain variables may be determinants of the location of WMFs, to which I pay particular attention to in this analysis. Firstly, it is reasonable to expect that WMFs are placed in city areas that are not densely populated. This preference may stem both from the local government's desire to minimize the number of citizens affected by these infrastructures, and the physical space required, particularly for landfills. It is also worth noting that, as all hexagons have the same area, population density is proportional to population size, so I can freely interchange the terms during the analysis.

Regarding physical space, it would be ideal to have an indicator measuring land availability across different city areas. However, this type of variable was not available in 1930. Nevertheless, the census provides information that enables identifying farm households. To build this variable in the 1930 census 'enumerators simply asked respondents whether the house in which they lived was located on a farm.' I employ this variable as a proxy for land availability and to identify peripheral neighborhoods within the cities. It can be reasonably expected that landfills were located in those areas of the cities with large availability of physical space, thus anticipating a positive relationship between the share of farm households and the location of WMFs.

Literature in environmental justice has emphasized the importance of considering the income dimension in the unequal distribution of pollution. It would be interesting to investigate whether this assessment seemed true in my initial data description. However, income information for households was not available in the 1930 census. One potential approach to address this issue would be considering educational attainment as a proxy for income. However, this variable is also unavailable for 1930. Instead, I employ the Duncan Socioeconomic Index (SEI) (Duncan, 1961) as a proxy for economic performance. This index 'is a measure of occupational status based upon the income level and educational attainment associated with each occupation in 1950.'

Finally, the main variable of interest is the share of black population. It is important to understand how this variable is distributed across each city and its relationship with the other previously mentioned potential determinants.

Figures 5 to 10 show the maps of each city under study, as well as the geographical distribution of the variables mentioned before, all evaluated in 1930. The blue crosses indicate the locations of landfills and waste incinerators constructed between 1930 and 2000. The figures show that black communities tended to be located in areas with high population density, typically near the city center. This pattern was strong and persistent across all the cities under study. Moreover, there appears to be a strong relationship between population density and the location of infrastructure: most landfills and waste incinerators were situated in areas with low population density in 1930, while cases in which such facilities were installed in neighborhoods with high population density were relatively rare. Therefore, any reasonable model seeking to capture the effect of the share of black population on the location of WMFs should incorporate population density as an explanatory variable. Failing to do so may lead to downward biased estimates due to the positive correlation between population density and the share of black inhabitants.

Another noteworthy observation from the maps is the clear economic inequality between black and white individuals during the 1930 census. The neighborhoods with higher concentration of black population also exhibited the lowest levels of SEI. Given my interest in disentangling the effect of a higher share of black population from other factors, it is crucial to include SEI as a covariate. Failure to account for this variable may lead to upwardly biased estimations, as landfills and waste incinerators were more likely to be located in economically disadvantaged neighborhoods.

Regarding farming households, as anticipated, they tended to be primarily situated near the periphery of the cities. In cities such as Baltimore, Chicago, Cincinnati, and Nashville, it appears that some WMFs were located in these peripheral areas.

Lastly, it is essential to acknowledge that in certain cities, some of the landfills and waste incinerators constructed were outside the city limits, as defined by the Urban Transition Project maps. This issue was more prominent in the southern cities. For instance, in Houston (refer to Figure 9), there were at least four cases where I could not include the corresponding information in this model due to the lack of data for those areas of the city. Consequently, we must be cautious when interpreting the results obtained for the southern cities, as they provide only a partial view of the situation.

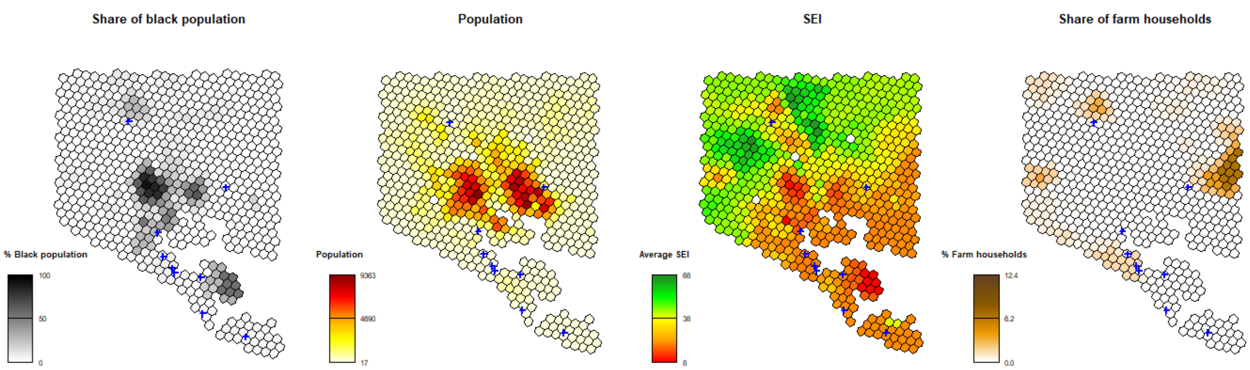


Figure 5: Socioeconomic characterization of Baltimore in 1930

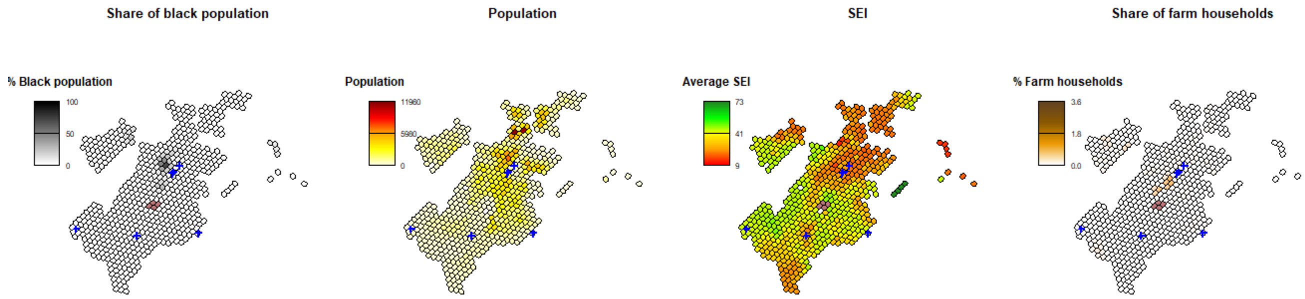


Figure 6: Socioeconomic characterization of Boston in 1930

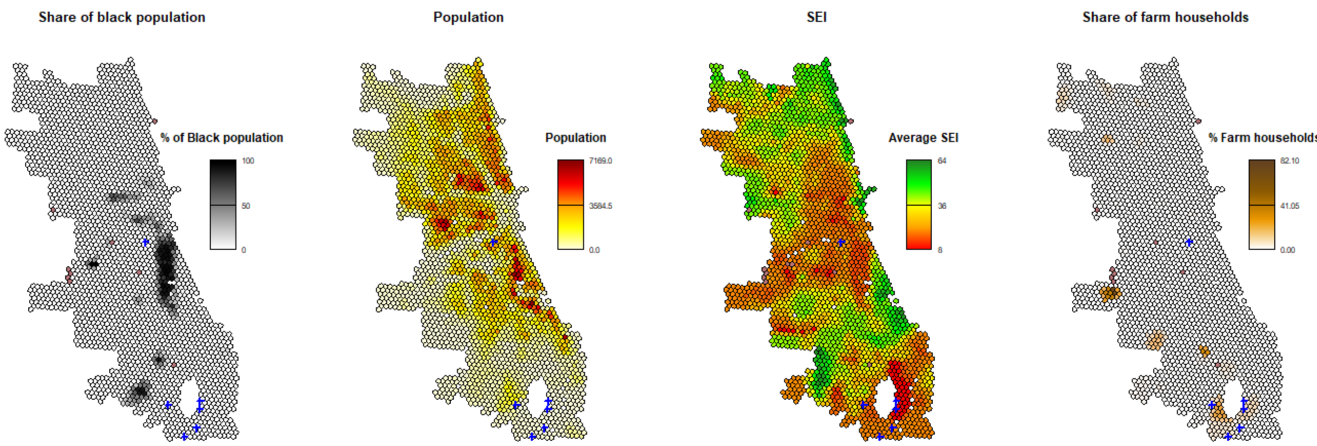


Figure 7: Socioeconomic characterization of Chicago in 1930

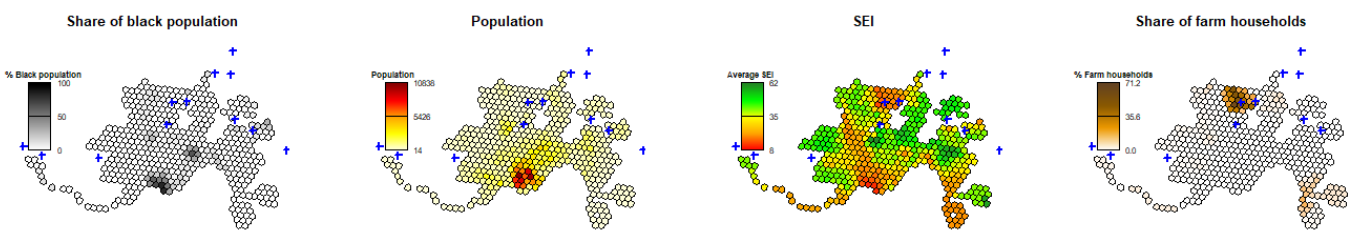


Figure 8: Socioeconomic characterization of Cincinnati in 1930

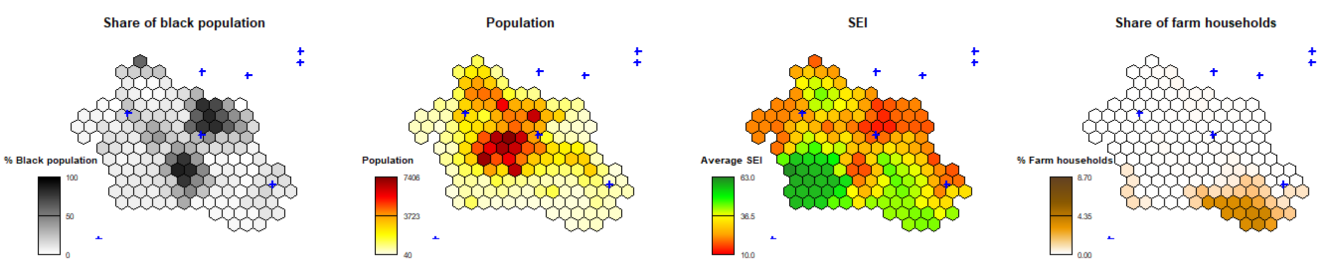


Figure 9: Socioeconomic characterization of Houston in 1930

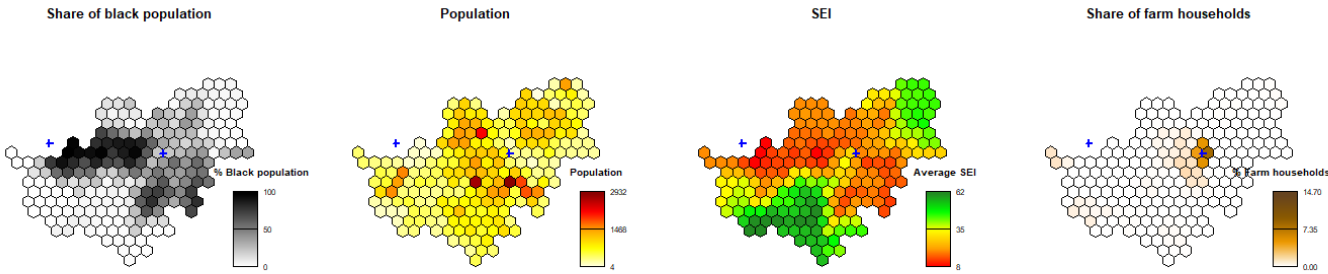


Figure 10: Socioeconomic characterization of Nashville in 1930

To further assess the relationship between explanatory variables and location of WMFs, I construct a conditional means table. This description provides an initial assessment of whether the direction of the relationships aligns with initial expectations.

For the variables share of black population, SEI, and population, I divide the individuals of each city into four different quartiles based on each variable. Specifically, for a given variable Q, let Q_p denote the p-th quartile, where Q_1 denotes the group of observations within the lower quartile and Q_4 the group of observations within the higher quartile of variable Q. I compute the share of SNs affected by a WMF (within a radius of 1 km) for each quartile of each variable, denoted as $\mathbb{E}[Y|Q_p]$.

Due to the high number of zero observations, quartiles are not appropriate for the share of farm households. Instead, I split the sample into two categories: neighborhoods with more than 1% of farm households ($F_{>1\%}$) and neighborhoods with less than 1% of farm households ($F_{<1\%}$).

Tables 2 to 5 present the results of this analysis. In general, the results align with our expectations. Table 2 reveals interesting patterns: with the exception of Chicago, there is a higher proportion of SNs affected by landfills or waste incinerators when comparing the highest quartile of the share of black population to the lowest quartile across all cities. Notably, these differences are substantial in cities like Baltimore, Houston, and Nashville.

Furthermore, the table suggests a nonlinear relationship between the share of black population and the likelihood of SNs being affected. The conditional means do not monotonically increase or decrease with an increasing share of black population. Instead, I observe discrete jumps in the probability for different quartiles of the distribution. This pattern is particularly notable in cities like Baltimore, Houston, and Nashville. In these cities, the relationship between the share of black population and the share of affected SNs does not appear to be strictly increasing. Instead, white neighborhoods seem relatively 'safe,' while neighborhoods with a medium or high share of black population are significantly more exposed. Therefore, marginal increases in the share of black population when it is already 'high enough' do not appear to have a substantial impact on the occurrence of landfills and waste incinerators.

	Baltimore	Boston	Chicago	Cincinnati	Houston	Nashville
$\mathbb{E}[Y\|B_1]$	0.05	0.11	0.05	0.13	0.02	0.00
$\mathbb{E}[Y\|B_2]$	0.14	0.17	0.02	0.11	0.12	0.00
$\mathbb{E}[Y\|B_3]$	0.20	0.14	0.01	0.09	0.28	0.28
$\mathbb{E}[Y\|B_4]$	0.17	0.16	0.02	0.11	0.17	0.19

Table 2: Share of affected SNs by quartiles of the share of black population

	Baltimore	Boston	Chicago	Cincinnati	Houston	Nashville
$\mathbb{E}[Y\|P_1]$	0.28	0.12	0.06	0.12	0.10	0.19
$\mathbb{E}[Y\|P_2]$	0.04	0.22	0.02	0.16	0.15	0.06
$\mathbb{E}[Y\|P_3]$	0.14	0.08	0.02	0.14	0.12	0.13
$\mathbb{E}[Y\|P_4]$	0.11	0.16	0.02	0.03	0.22	0.09

Table 3: Share of affected SNs by quartiles of population density

	Baltimore	Boston	Chicago	Cincinnati	Houston	Nashville
$\mathbb{E}[Y\|S_1]$	0.27	0.20	0.08	0.10	0.26	0.22
$\mathbb{E}[Y\|S_2]$	0.23	0.17	0.03	0.12	0.23	0.19
$\mathbb{E}[Y\|S_3]$	0.05	0.09	0.01	0.11	0.09	0.06
$\mathbb{E}[Y\|S_4]$	0.00	0.12	0.00	0.11	0.00	0.00

Table 4: Share of affected SNs by quartiles of SEI

	Baltimore	Boston	Chicago	Cincinnati	Houston	Nashville
$\mathbb{E}[Y\|F_{\leq 1\%}]$	0.14	0.14	0.02	0.11	0.17	0.08
$\mathbb{E}[Y\|F_{>1\%}]$	0.17	0.29	0.14	0.12	0.06	0.59

Table 5: Share of affected SNs by share of farm households

The results for the population density align with our expectations. SNs in the lowest quartiles of population density tended to have a higher proportion of affected neighborhoods. This pattern is consistent across all cities, except for Houston. However, it is important to note that our results for Houston may be biased due to the exclusion of four structures located on the city's periphery, for which I lack information.

The strongest results emerge when examining the SEI. In over half of the cities analyzed, no affected neighborhoods were in the highest quartile of income performance in 1930. This striking result highlights the presence of socioeconomic inequality in terms of exposure to pollutants. Instead, a substantial probability of being affected is observed among neighborhoods in the lowest income quartiles. Even if it might be argued that this is due to lower land prices in low-income neighborhoods, it does not negate the fact that individuals starting with initial disadvantages compared to the wealthier segments of the population are the ones who will bear the brunt of the negative health consequences of environmental pollution, thus exacerbating initial inequalities.

Regarding the relationship between the share of farm households and the location of WMFs, the results mostly align with our expectations, except for Houston. In cities like Chicago or Nashville, there is a substantial increase in the probability of being affected as the share of farm households increases.

3.3 Results

The results of estimating Equation 4 using the methodology proposed by Belloni et al. (2014) are presented in Table 6. The estimation was performed in R using the 'hdm' package developed by one of the authors of the paper (Chernozhukov et al., 2016). The number of observations corresponds to the number of SNs available in each city. Area-fixed effects were included whenever the variable was available. The table also indicates the number of controls selected by the procedure.

In order to comply with the requirements of the procedure, both the treatment variable and the controls were standardized. Therefore, the β_c coefficient can be interpreted as the change in the probability of an SN being affected by a landfill or waste incinerator following a one-standard-deviation increase in the share of black population.

	Baltimore	Boston	Chicago	Cincinnati	Houston	Nashville
$\hat{\beta}_c$	-0.030	0.011	-0.009^{**}	0.001	0.064	-0.001
	(0.034)	(0.021)	(0.003)	(0.021)	(0.053)	(0.064)
Area fixed effects	YES	YES	YES	YES	NO	YES
Controls selected	19	6	20	14	14	12
Num. obs.	652	614	2384	459	161	213

Note: Results of linear treatment specification in Equation 4. Standard errors computed following Belloni et al. (2014) are reported in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 6: Results from the linear specification

The estimated coefficients suggest that, in the cities examined, there is no strong evidence of direct discrimination in the allocation of landfills and waste incinerators when assuming a linear relationship between the variables. The point estimates for Boston, Cincinnati, and Houston were positive but not statistically significant at the 10% level.

In the case of Chicago, however, the estimated coefficient is negative and statistically significant. This finding is largely driven by the introduction of waste management facilities in the Lake Calumet Cluster, in the Southern part of the city, during the 1960s and 1970s. It is important to note that this area had a predominantly white population in 1930, which changed in the 1950s and 1960s, just before the introduction of the facilities. This particular case may highlight a limitation of the proposed methodology. When abrupt changes occur in the share of black population, as in this case, the methodology may fail to accurately capture the desired effect, as the racial composition used for the regression does not fully represent the reality of the city at the time when the infrastructures were constructed. However, it is worth noting that such abrupt changes are uncommon in the database.

	Baltimore	Boston	Chicago	Cincinnati	Houston	Nashville
$\hat{\beta}_2$	-0.048	-0.103	-0.016^{**}	0.130^{**}	0.225^{***}	-0.022
	(0.033)	(0.064)	(0.007)	(0.045)	(0.074)	(0.032)
$\hat{\beta}_3$	0.078	0.015	-0.038	0.026	0.300^{***}	0.078
	(0.070)	(0.044)	(0.027)	(0.049)	(0.104)	(0.049)
$\hat{\beta}_4$	-0.151	0.134	-0.024^{*}	-0.049	0.122	-0.111
	(0.100)	(0.339)	(0.013)	(0.033)	(0.127)	(0.083)
Area fixed effects	YES	YES	YES	YES	NO	YES
Controls selected	18	9	17	11	13	13
Num. obs.	652	614	2384	459	161	213

Note: Results from the categorical dummies treatment specification in Equation 5. Standard errors computed following Belloni et al. (2014) are reported in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 7: Results from the categorical dummies specification

Table 7 presents the results of the regression specification employing group dummies for the share of black population (Equation 5). This approach uncovers interesting insights not captured by the linear treatment assumption of the previous model. To address potential multicollinearity issues and enhance interpretability, the coefficient associated to $B^{G1}_{i,w,c}$ was excluded from the regression, implying that each coefficient representing the differential change in the probability of being affected by a WMF compared to the group of SNs with the lowest share of black population.

The findings from this model reveal a more nuanced pattern than those obtained from the linear specification. In the case of Houston, all coefficients are positive, significant, and substantial in magnitude, indicating a statistically significant increase in the probability of being affected by waste management facilities for neighborhoods with higher shares of black inhabitants. This effect, which was not fully captured by the previous model, appears to be more pronounced for neighborhoods with intermediate levels of black population and diminishes for neighborhoods with the highest shares of black population.

Furthermore, in Cincinnati, there is a positive effect for the second group, suggesting that areas with intermediate shares of black population also experienced a significantly higher probability of being affected by WMFs. This finding contrasts with the nearly negligible relationship observed in the previous linear model.

In Boston, the coefficients show an increasing trend, albeit with an initial negative coefficient that very likely influenced the low point estimation obtained in the initial specification.

Additionally, the estimation of coefficients in Table 7 reveals an inverted U-shaped relationship in many of the cities analyzed. This indicates that direct discrimination is stronger in neighborhoods with mixed-race compositions (G2 and G3). In contrast, predominantly white or black neighborhoods (G1 and G4) do not exhibit such a strong effect. This finding may be explained by the presence of blocking coalitions based on race, whereby neighborhoods with a more homogeneous racial composition were better able to mobilize and resist the introduction of waste management facilities, while mixed-race neighborhoods faced greater challenges in organizing against these facilities.

4 What happens after a WMF starts operating?

4.1 Empirical strategy

In this section, I study the effects of introducing a WMF on different socioeconomic variables. I study each city in which I have data availability separately, trying to provide some institutional insight into each case.

Shertzer and Walsh (2019) procedure enables having panel data at the SN level for 1920, 1930, and 1940 for the northern cities, and for 1930, and 1940 for the southern ones. This permits to study the effects of introducing a waste management facility on several socio-economic aspects of the city. In particular, I analyze the effect of their introductions on the share of black population, total population, and economic performance of the inhabitants of the SN. However, it is important to note that, due to the lack of information on income and education for 1920 and 1930, identifying the economic performance variable can be challenging. Nevertheless, I have data on the employment status of the censused individuals from 1930 and 1940, which allows me to compute the unemployment rate per SN as a proxy for economic performance.

It is important to acknowledge that this analysis is limited to waste management facility introductions between 1920 and 1940 in northern cities and between 1930 and 1940 in southern cities. However, such cases are relatively rare in my database, with only a few instances available for analysis. Given the data availability constraint, I can only explain the short-term impact of these introductions. However, it is still useful to obtain a first understanding of this phenomenon and gaining valuable insights into population dynamics.

I employ different methodological approaches based on data availability to explore the relationship between WMFs introductions and socio-economic variables. Specifically, for cases where information is only available for the 1930 and 1940 censuses, I adopt a simple Difference-in-Differences (DiD) approach.

$$Y_{i,t} = \beta_0 + \beta_1 D_i + \beta_2 T_{1940} + \beta_3 (D_i * T_{1940}) + \varepsilon_{i,t} \tag{7}$$

where $Y_{i,t}$ is an outcome of interest in SN i in year t, D_i is an indicator variable taking the value 1 if SN i was among those affected by the WMF (i.e. if SN i was among a 1 km radius from the installed WMF) and 0 otherwise, and T_{1940} is an indicator variable taking the value 1 if $t = 1940$ (post-treatment) and 0 if $t = 1930$ (pre-treatment). $D_i * T_{1940}$ is the interaction term. The parameter of interest is β_3, which captures the differential change in the outcome after introducing a WMF in that city area.

When there is information on all 3 years, I introduce a more flexible approach, allowing units to get treated in different periods:

$$Y_{i,t} = \alpha_i + \gamma_t + \beta_{T1} I_{K_{it}=1} + \beta_{T2} I_{K_{it}=2} + \varepsilon_{i,t} \tag{8}$$

where the census years are represented by $t = \{1920, 1930, 1940\}$, α_i are individual fixed effects, γ_t are time-fixed effects, and K_{it} indicates the time difference between t and the treatment. $K_{it} = 1$ when the WMF was introduced in between census $t - 10$ and t, and $K_{i,t} = 2$ represents the case when the WMF was introduced in between census $t - 20$ and $t - 10$ (in this case, this is only possible when $t = 1940$ and the WMF was introduced between 1920-1930). β_{T1} captures the effect of the introduction after one period, whereas β_{T2} captures the effect of the introduction after 2 periods.

4.2 Results

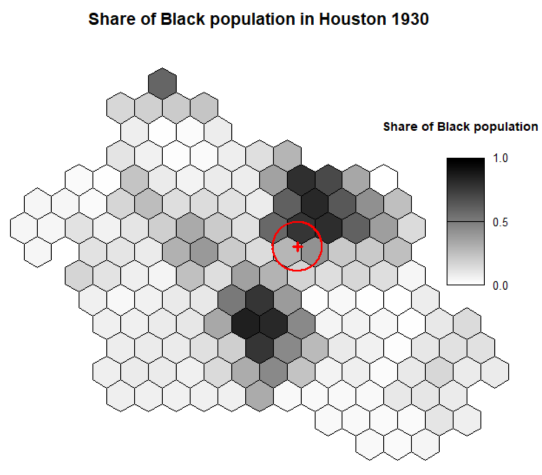
In this subsection, I present the results of the effects of introducing waste management facilities (WMFs) in the cities of Baltimore and Houston separately. †

The results go in the expected direction. The introduction of WMFs produces an increase in the share of black population and a negative effect in terms of the total population. Regarding the economic status of the affected inhabitants, no statistically significant effect was found, although the point estimates suggest that individuals with higher unemployment settled in those SNs. This seems to support the white outflow hypothesis: people left the neighborhood after the introduction of the landfill or waste incinerator. However, given that white individuals were, on average, wealthier than black individuals, they could more easily afford to relocate, increasing the share of black inhabitants.

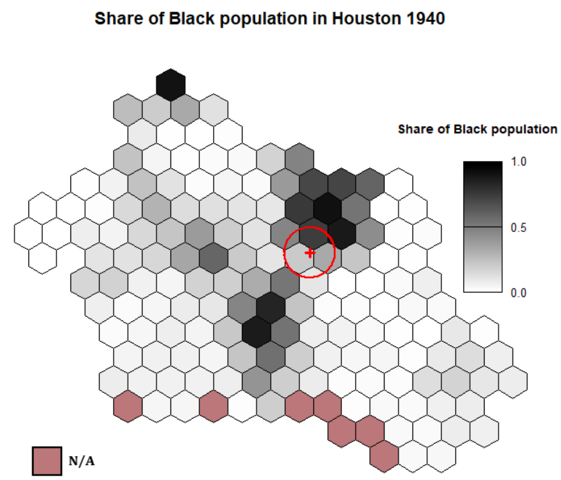
4.2.1 Houston

In the case of Houston, I study the effect of the introduction of the Velasco Incinerator. The Velasco Incinerator operated from the 1930s to the 1950s and was situated in the East downtown area of Houston. Historical sources, such as Bullard (1983), indicate that it was located in a predominantly black neighborhood, as depicted in Figure 11. Additionally, Figure 12 illustrates that the area surrounding the incinerator had a moderately high population density. The red cross on the figures represents the exact location of the incinerator, while the red circle represents the 1 km radius defining the affected area.

† While WMFs were also introduced in the 1930s in Cincinnati, the Urban Transition Project map for the city in 1940 does not provide information about peripheral neighborhoods (see Figure A.4 in Appendix D) where one of the WMFs was introduced. Therefore, Cincinnati is excluded from the analysis to ensure the completeness and accuracy of the results.

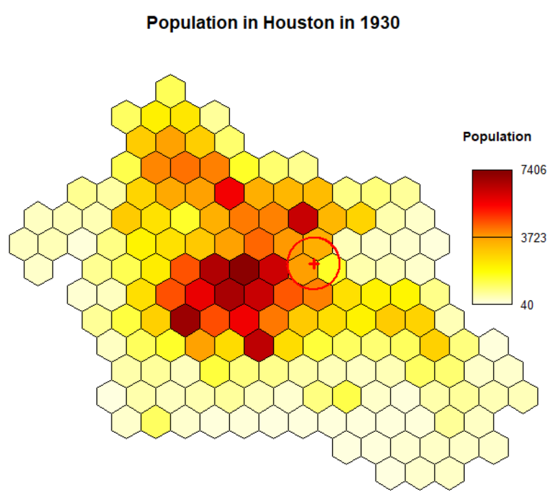


(a) 1930

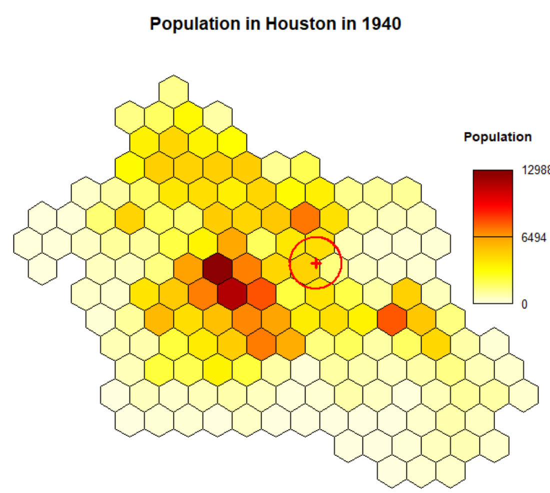


(b) 1940

Figure 11: Percentage of Black Population in Synthetic Neighborhoods in Houston 1930-1940



(a) 1930



(b) 1940

Figure 12: Population in Synthetic Neighborhoods in Houston 1930-1940

Figure 11 does not show major changes in the share of black population in the incinerator-surrounding SNs from 1930 to 1940. Instead, Figure 12 shows a decrease in population density after introducing the incinerator.

To accurately estimate this effect, I employ a DiD approach, as described in Equation 7. Table 8 presents the DiD estimation results for the various outcomes of interest. The estimated parameter signs align with expectations, but most of them lack statistical significance, except for population, which is significant at the 10% level. Regarding the share of black inhabitants, the effect is non-statistically significant, and the point estimation suggests that the introduction of the Velasco incinerator led to an increase of less than 2% in the share of black population in the neighborhoods located 1 km away, which is statistically indistinguishable from 0.

Furthermore, Table 9 displays the same results after standardizing the variables, allowing for a direct comparison of the effect of this introduction across the three different outcomes considered. The most substantial impact is observed in terms of population, with the introduction of the incinerator resulting in a decrease of approximately one-third of a standard deviation in that variable. Conversely, the effects on the share of unemployed inhabitants and the share of black population are less sizable.

	Share of Black	Population	Unemployment
$\hat{\beta}_3$	0.016	-763.47^{*}	0.002
	(0.041)	(392.27)	(0.006)
Num. obs.	308	308	308

Note: Difference-in-Differences estimation for the effect of introducing the Velasco Incinerator in Houston. Variables are in levels. Robust standard errors are reported in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 8: DiD estimations in levels for the effects of introducing the Velasco Incinerator, Houston

	Share of Black	Population	Unemployment
$\hat{\beta}_3$	0.074	-0.348^{*}	0.101
	(0.176)	(0.179)	(0.329)
Num. obs.	308	308	308

Note: Difference-in-Differences estimation for the effect of introducing the Velasco Incinerator in Houston. Variables are standardized. Robust standard errors are reported in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 9: Standardized DiD estimations for the effects of introducing the Velasco Incinerator, Houston

4.2.2 Baltimore

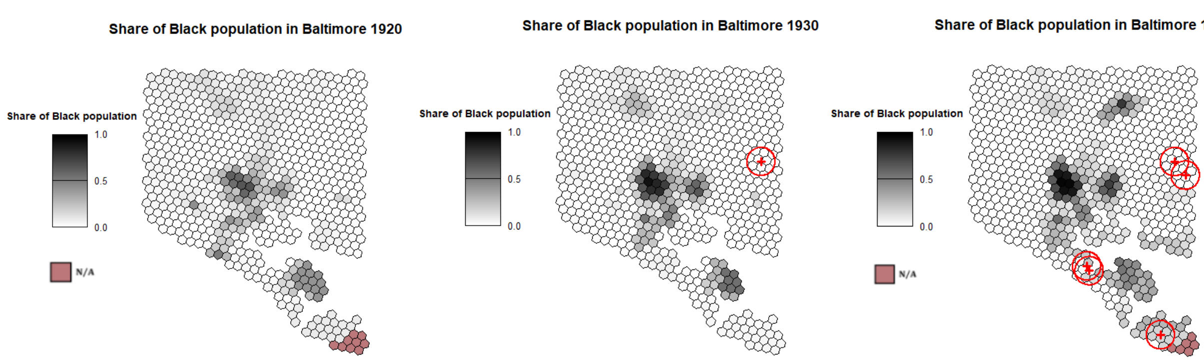
Baltimore is an interesting city to study this phenomenon, as four WMFs were introduced between 1920 and 1940. Regarding the projects, in 1927, Baltimore acquired the Bowley's

Lane Landfill site and used it as an unpermitted landfill until 1984. In the 1930s, the Reedbird Avenue Landfill and Incinerator started operations in South Baltimore. Moreover, the Pulaski Incinerator, located in East Baltimore, started its operations in 1931 after being constructed by the municipal government in 1927. Additionally, the Quarantine Road Landfill, located in the South of the city, started being used as an open dump during the 1930s, and it was later institutionalized as a sanitary landfill by the city of Baltimore.

These events provide an ideal framework for implementing a staggered DiD approach, allowing us to understand both the short and medium-run effects of the WMFs.

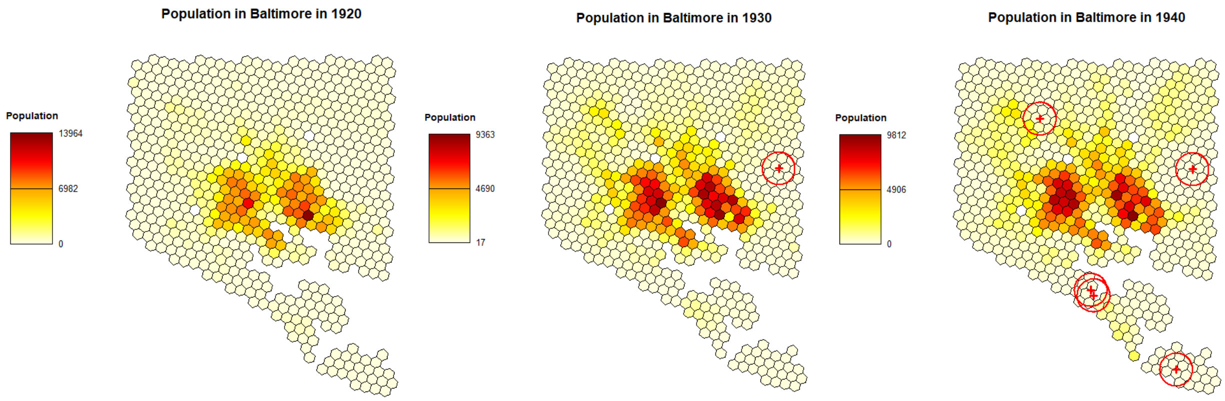
Data on population size and the share of black population is available for 1920, 1930, and 1940, while information on the unemployment rate is available only for 1930 and 1940. Therefore, for the first two variables, I use the specification described in Equation 8, while for unemployment, I employ the same Difference-in-Differences approach used for Houston (Equation 7).

Figures 13 and 14 show the evolution of the share of black population and the population size in the city of Baltimore for the 1920-1940 period. The figures also indicate the locations of newly introduced landfills and waste incinerators during this period (red crosses) and the 1 km radius around them (red circle), providing insight into the geographical areas of the city where these facilities were situated. It appears that population density played a significant role in determining the placement of these WMFs in Baltimore, as all the introduced WMFs were located in low-population-density neighborhoods.



(a) 1920 (b) 1930 (c) 1940

Figure 13: Percentage of Black Population in Synthetic Neighborhoods in Baltimore 1920-1940



(a) 1920 (b) 1930 (c) 1940

Figure 14: Population in Synthetic Neighborhoods in Baltimore 1920-1940

The results of the estimation derived from Equation 8 are visualized in Figure 15, while Table 10 presents the estimated coefficients for the share of black inhabitants and the population. The introduction of WMFs appears to have an impact on the population dynamics of the affected neighborhoods. In the short term (T1), there is a statistically significant increase in the share of black population, accompanied by a decrease in the overall population. This finding aligns with the white flight hypothesis, suggesting that as WMFs are introduced, residents tend to leave the neighborhood, as indicated by the negative coefficient estimated for population. However, the racial composition of those departing residents is not balanced; primarily, white individuals leave, resulting in an increase in the share of black population. These dynamics reinforce initial racial inequalities, as white residents can avoid the negative consequences associated with WMFs by leaving the neighborhood. However, this option is only available due to pre-existing disparities across households. This mechanism exacerbates racial inequalities, considering that residing near waste management facilities has known adverse health effects.

In terms of medium-term effects, I can only analyze the Bowley's Lane landfill case, as it was the only facility constructed between 1920-1930. Hence, I can identify the effect two decades post-introduction from a single landfill study, so we must be cautious when extrapolating these results. The estimation reveals no significant effect on the share of black population, but a decrease in the overall population. This finding also supports the white flight hypothesis, suggesting that there was no significant movement of new black residents into these neighborhoods, as evidenced by the population decline and minimal changes in the racial composition. As for the unemployment rate of inhabitants, the point estimate for the 1930s landfills and waste incinerators is positive but not statistically significant (see Table A.3 in Appendix E).

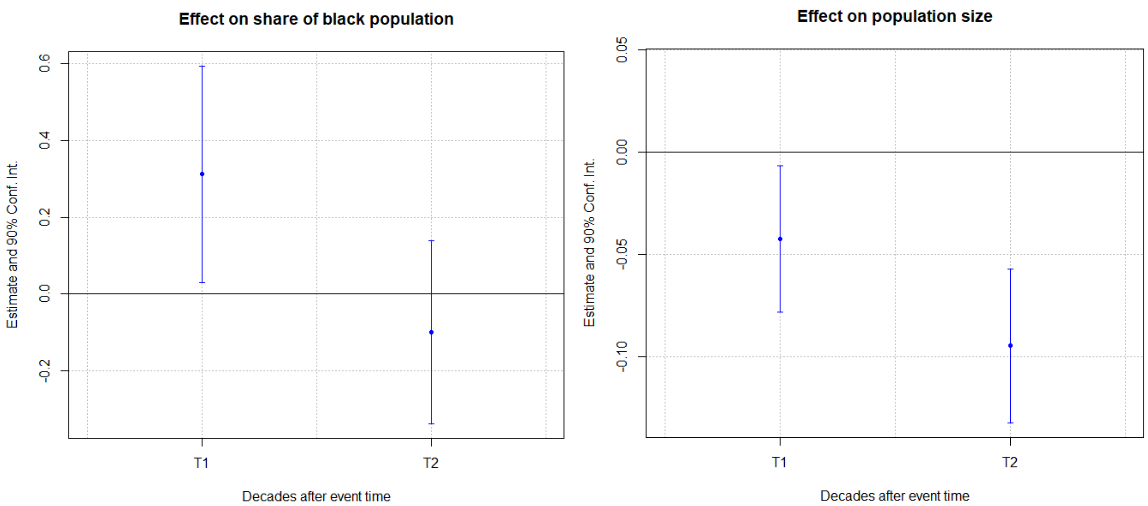


Figure 15: Staggered DiD point estimation and confidence intervals at the 90% level for the introduction of landfills and waste incinerators in Baltimore during 1920-1940

	Share of Black	Population
$\hat{\beta}_{T1}$	0.312^{*}	-0.042^{*}
	(0.171)	(0.022)
$\hat{\beta}_{T2}$	-0.099	-0.095^{***}
	(0.145)	(0.023)
Num. obs.	1965	1965

Note: Staggered Difference-in-Differences estimation for the effect of the introduction of landfills and waste incinerators in Baltimore. Variables are standardized. Clustered standard errors at the Ward-year level are reported in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 10: Staggered DiD estimations for standardized variables for Baltimore

5 Conclusions

Throughout this paper, I disentangle the possible confounding channels that could potentially explain the positive association between the location of landfills and waste incinerators and the share of black population. I employ historical data from the 1920, 1930, and 1940 censuses combined with maps from the Urban Transition Project to obtain a geographical description of the cities under study. This data was mapped at the Synthetic Neighborhood level to obtain stable units of analysis and perform meaningful comparisons across different censuses. Finally, I apply modern econometric techniques to pursue the analysis and to try to reduce the bias derived from model selection and functional form as much as possible.

Initially, I investigate whether landfills and waste incinerators were placed in predominantly black neighborhoods after controlling for other geographical and socioeconomic cofounders. The proposed methodology tries to capture the presence of direct discrimination in the allocations, which occurs when, after controlling for geographical and socioeconomic characteristics of the zones, there is still a statically significant effect of the share of black population. However, my results show that there did not seem to be targeting in terms of the share of black population regarding the location of waste management facilities, as measured by the linear specification: a marginal increase in the share of black population did not increase the probability of being affected by a WMF. Instead, when considering a more flexible model, I found an inverted U-shaped relationship between the variables. This result is consistent with the potential presence of blocking coalitions based on race, as neighborhoods with a homogeneous racial composition prevented the introduction of WMFs. In contrast, mixed-race ones were the most affected. Further exploration of this hypothesis is left as part of future research.

Additionally, I examine the consequences of WMF introduction on the affected regions. The most notable finding is a decrease in urban development, as measured by a decline in population density, following the introduction of these infrastructures. Furthermore, my analysis reveals an increase in the share of black population after the implementation of WMFs in Baltimore and Houston. The effect is more pronounced in Baltimore, while in Houston, the effect is positive but small in magnitude and statistically insignificant. These findings suggest that white flight from these neighborhoods was the factor that primarily drove the channel linking WMF location to race.

References

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

Berry, M. and Bove, F. (1997). Birth weight reduction associated with residence near a hazardous waste landfill. *Environmental health perspectives*, 105(8):856–861.

Bohren, J. A., Hull, P., and Imas, A. (2022). Systemic discrimination: Theory and measurement. Technical report, National Bureau of Economic Research.

Bullard, R. D. (1983). Solid waste sites and the black houston community. *Sociological inquiry*, 53(2-3):273–288.

Bullard, R. D. (2008). *Dumping in Dixie: Race, class, and environmental quality.* Avalon Publishing-(Westview Press).

Chernozhukov, V., Hansen, C., and Spindler, M. (2016). hdm: High-dimensional metrics. *arXiv preprint arXiv:1608.00354*.

Dockery, D. W., Pope, C. A., et al. (1994). Acute respiratory effects of particulate air pollution. *Annual review of public health*, 15(1):107–132.

Duncan, O. D. (1961). A socioeconomic index for all occupations. *Occupations and social status*.

Elliott, P., Eaton, N., Shaddick, G., and Carter, R. (2000). Cancer incidence near municipal solid waste incinerators in great britain. part 2: histopathological and case-note review of primary liver cancer cases. *British Journal of Cancer*, 82(5):1103–1106.

Elliott, P., Shaddick, G., Kleinschmidt, I., Jolley, D., Walls, P., Beresford, J., and Grundy, C. (1996). Cancer incidence near municipal solid waste incinerators in great britain. *British Journal of Cancer*, 73(5):702–710.

GAO, U. S. G. A. O. (1983). *Siting of Hazardous Waste Landfills and Their Correlation with Racial and Economic Status of Surrounding Communities: Report.*

Griffith, J., Duncan, R. C., Riggan, W. B., and Pellom, A. C. (1989). Cancer mortality in us counties with hazardous waste sites and ground water pollution. *Archives of Environmental Health: An International Journal*, 44(2):69–74.

Hite, D., Chern, W., Hitzhusen, F., and Randall, A. (2001). Property-value impacts of an environmental disamenity: the case of landfills. *The Journal of Real Estate Finance and Economics*, 22(2):185–202.

Karnig, A. K. (1976). Black representation on city councils: The impact of district elections and socioeconomic factors. *Urban Affairs Quarterly*, 12(2):223–242.

Katsouyanni, K., Touloumi, G., Spix, C., Schwartz, J., Balducci, F., Medina, S., Rossi, G., Wojtyniak, B., Sunyer, J., Bacharova, L., et al. (1997). Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 european cities: results from time series data from the APHEA project. *Bmj*, 314(7095):1658.

Logan, J. R., Jindrich, J., Shin, H., and Zhang, W. (2011). Mapping America in 1880: The urban transition historical GIS project. *Historical Methods*, 44(1):49–60.

Louis, G. E. (2004). A historical context of municipal solid waste management in the united states. *Waste management & research*, 22(4):306–322.

Mohai, P., Pellow, D., and Roberts, J. T. (2009). Environmental justice. *Annual review of environment and resources*, 34:405–430.

Ruggles, S., Fitch, C., Goeken, R., Hacker, J., Nelson, M., Roberts, E., Schouweiler, M., and Sobek, M. (2021). IPUMS ancestry full count data: Version 3.0 [dataset]. *Minneapolis, MN: IPUMS.*

Sharkey, P. (2013). *Stuck in place: Urban neighborhoods and the end of progress toward racial equality.* University of Chicago Press.

Shertzer, A. and Walsh, R. P. (2019). Racial sorting and the emergence of segregation in American cities. *Review of Economics and Statistics*, 101(3):415–427.

Taylor, R. L. (1979). Black ethnicity and the persistence of ethnogenesis. *American Journal of Sociology*, 84(6):1401–1423.

Vrijheid, M. (2000). Health effects of residence near hazardous waste landfill sites: a review of epidemiologic literature. *Environmental health perspectives*, 108(suppl 1):101–112.

Wolfinger, R. E. (1965). The development and persistence of ethnic voting. *American Political Science Review*, 59(4):896–908.

Appendix

A Landfills and Waste Incinerators

Name	City	Position	Open	Close	Status
Bowley's Lane LF	Baltimore	39.31,-76.54	1927	1985	Closed
Pennington Avenue LF	Baltimore	39.22,-76.59	1973	1981	Closed
Norris Farms LF	Baltimore	39.29,-76.48	1968	1985	Closed
Buck's Auto Dump	Baltimore	39.25,-76.62	1966	1975	Closed
Cylburn Park Dump	Baltimore	39.35,-76.65	1900s	2015	Closed
Fairfield SC Dump	Baltimore	39.24,-76.59	1955	Unknown	Closed
Coldspring Lane LF	Baltimore	39.34,-76.65	1930s		Active
Pulaski WI	Baltimore	39.31,-76.54	1927	2004	Closed
Reedbird Avenue WI	Baltimore	39.25,-76.62	1930s	1977	Closed
Reedbird Avenue LF	Baltimore	39.24,-76.62	1930s	1977	Closed
Wheelabrator WI	Baltimore	39.27,-76.63	1985		Active
Quarantine Road LF	Baltimore	39.2,-76.56	1930s		Active
Monument Street LF	Baltimore	39.3,-76.57	1973	1980	Closed
Barry Quarry	Boston	42.28,-71.11	1984	2014	Closed
Gardner Street LF	Boston	42.28,-71.18	1954	1999	Closed
Hallet Street LF	Boston	42.28,-71.04	1948	1966	Inactive
Reenergy Roxbury LLC	Boston	42.33,-71.07	1992		Active
Howard Transfer Station	Boston	42.33,-71.07	1999		Active
Spectacle Island LF	Boston	42.32,-70.99	1918	2006	Closed
South Bay WI	Boston	42.33,-71.07	1959	1975	Closed
CID Recycling and Disposal	Chicago	41.66,-87.58	1966	2007	Closed
Land & Lakes	Chicago	41.67,-87.57	1977	1998	Closed
Paxton LF	Chicago	41.68,-87.57	1976	1992	Closed
Stearn's Quarry	Chicago	41.84,-87.65	1969	1990	Closed
River Bend Prairie LF	Chicago	41.65,-87.59	1974	2014	Closed
Alburn WI	Chicago	41.68,-87.62	1967	1982	Closed

Promontory Point	Chicago	41.8,-87.58	1920s	1937	Closed
ELDA RDF	Cincinnati	39.19,-84.51	1973	1998	Closed
Rumpke SLF, Inc.	Cincinnati	39.27,-84.6	1945	2065	Active
Amberley Village LF	Cincinnati	39.22,-84.43	1950s	1980s	Closed
Anderson Township LF	Cincinnati	39.14,-84.32	1963	1986	Closed
Arlington Heights	Cincinnati	39.22,-84.46	1940s	1970s	Closed
Bass Island	Cincinnati	39.14,-84.36	1960s	1960s	Closed
Cheviot Closed LF	Cincinnati	39.13,-84.69	1954	1975	Closed
Cincinnati Milicron	Cincinnati	39.17,-84.43	1956	1970s	Closed
Village of Elmwood LF	Cincinnati	39.19,-84.49	1930s	1960s	Closed
Evendale LF	Cincinnati	39.24,-84.43	1960s		Closed
Glenway Crossing LF	Cincinnati	39.13,-84.61	1950s	1973	Closed
Dunbar WI	Cincinnati	39.16,-84.41	1931	1972	Closed
Gest street dump	Cincinnati	39.11,-84.54	Before the 30s		
Bellfort Boulevard LF	Houston	29.67,-95.37	1954	1970	Closed
Blue Bonnet LF	Houston	29.82,-95.24	1979	1998	Closed
McCarty Road LF	Houston	29.83,-95.24	1972	2037	Open
Whispering Pines LF	Houston	29.88,-95.27	1978	2047	Open
Velasco (and navigation) WI	Houston	29.76,-95.34	1930s	1950s	Closed
Fourth Ward/Gillette Street WI	Houston	29.76,-95.38	1920s	1940s	Closed
Holmes Road	Houston	29.67,-95.41	1930s		Closed
Kirkpatrick LF	Houston	29.81,-95.29	1971	1972	Closed
Sunnyside LF (Reed Road)	Houston	29.67,-95.43	1964	1970s	Closed
Acres Homes dump	Houston	29.84,-95.43	1960s	1970s	Closed
Northwest WI	Houston	29.86,-95.54	1972	Unknown	Closed
Kelley street WI	Houston	29.81,-95.34	1972	Unknown	
Westpark WI	Houston	29.73,-95.48	1972		Unknown
Bordeaux LF	Nashville	36.17,-86.84	1973	1994	Closed
Thermal Transfer Corporation	Nashville	36.17,-86.78	1974	2004	Closed

Table A.1: List of some of the considered Landfills and Waste Incinerators

B List of variables

B.1 Geographical variables:

Variable	Description	Availability	Source
ENUMDIST	Enumeration District	1920-1940	IPUMS - UTC
STATEICP	State	1920-1940	IPUMS
COUNTYICP	County	1920-1940	IPUMS
CITY	City	1920-1940	IPUMS
WARD	Ward of the ED	1920-1940	IPUMS
DCEN	Distance from SN to the city center	1920-1940	UTP
LAKE	1 if SN is next to a lake	1920-1940	MyGeodata
RIVER	1 if SN is next to a river	1920-1940	National Weather Service
ELEVATION	Average elevation of the SN	1920-1940	worldclim.org
PREVIOUS	Indicator taking the value 1 if there was a LF or WI in a radius of 1km	1920-1940	Self-constructed

B.2 Socioeconomic variables:

Variable	Description	Availability	Source
RACE	Race of the individual	1920-1940	IPUMS
POPULATION	Population Size of SN	1920-1940	IPUMS
FARM	Whether the household was a farm	1920-1940	IPUMS
IND	Industry of occupation	1930-1940	IPUMS
PRIM	Share of workers in primary sector	1930-1940	IPUMS
MANU	Share of workers in manufacture sector	1930-1940	IPUMS
SERVICE	Share of workers in service sector	1930-1940	IPUMS
DOMESTIC	Share of workers in domestic sector	1930-1940	IPUMS
SEI	Duncan Socioeconomic Index: Attach to each occupation a ranking	1920-1940	IPUMS
VALUEH	Value of the House	1930-1940	IPUMS
OWNERSHP	Whether the house is owned or not	1920-1940	IPUMS
AGE	Age of the individual	1920-1940	IPUMS
EMPSTAT	Employment status (employed, unemployed, etc)	1930-1940	IPUMS
IMM	Share of immigrants in SN	1920-1940	IPUMS

C Selecting controls

This section of the Appendix provides a resume of the selection on observables procedure proposed by Belloni et al. (2014). Throughout this chapter, I maintain the notation used by the authors: the outcome variable is denoted by y, the treatment by d and the vector of controls by $\mathbf{z}$.

C.1 Framework

Belloni et al. (2014) start from a partially linear model in which both the outcome and treatment variable depend on the controls:

$$y_i = d_i\alpha_0 + g(\mathbf{z_i}) + \zeta_i, \quad \mathbb{E}[\zeta_i|\mathbf{z_i}, d_i] = 0 \tag{A.1}$$

$$d_i = m(\mathbf{z_i}) + \upsilon_i, \quad \mathbb{E}[\upsilon_i|\mathbf{z_i}] = 0 \tag{A.2}$$

where $g(\mathbf{z_i})$ and $m(\mathbf{z_i})$ are unknown potentially complicated functions of the controls; ζ_i and υ_i are the error terms and α_0 is the parameter in which we are interested in making inference about. Although the functional relationship between the controls and the outcome can be potentially non-linear, the authors assume a linear relationship between the treatment and the outcome variable.

Thereafter, they proceed employing linear combinations of the control terms $\mathbf{x_i} = P(\mathbf{z_i})$ a p-dimensional vector in which there are included transformations of the original controls (polynomials, interactions, splines, dummies) in order to approximate both $g(\mathbf{z_i})$ and $m(\mathbf{z_i})$. The approximations are linear in the parameters and have approximation errors r_{g_i} and r_{m_i} respectively:

$$y_i = d_i\alpha_0 + \mathbf{x_i}^\top\beta_{g_0} + r_{g_i} + \zeta_i \tag{A.3}$$

$$d_i = \mathbf{x_i}^\top\beta_{m_0} + r_{m_i} + \upsilon_i \tag{A.4}$$

In order for this approximation to be valid and provide a framework where it is possible to carry out inference we require the original model to be sparse. Sparsity can be defined in terms of two conditions: Firstly, we require that there exist approximations $\mathbf{x_i}^\top \beta_{g_0}$ and $\mathbf{x_i}^\top \beta_{m_0}$ to $g(\mathbf{z_i})$ and $m(\mathbf{z_i})$ that require only a small number of non-zero coefficients to make the approximation errors r_{g_i} and r_{m_i} small relative to the estimation error. Formally, we require the existence of β_{g_0} and β_{m_0} such that at most $s = s_n \ll n$ elements of them are non zero and the resulting approximation errors should be small compared to the size of the estimation error:

$$\left[\bar{\mathbb{E}}\left(r_{g_i}^2\right)\right]^{\frac{1}{2}} \lesssim \sqrt{s/n} \quad \text{and} \quad \left[\bar{\mathbb{E}}\left(r_{m_i}^2\right)\right]^{\frac{1}{2}} \lesssim \sqrt{s/n}$$

C.2 Least squares after double selection

This subsection describes the procedure that needs to be followed in order to obtain the **post-double selection estimator**. Start by writing the reduced form of the model:

$$y_i = \mathbf{x_i}^\top \bar{\beta}_0 + \bar{r}_i + \bar{\zeta}_i \tag{A.5}$$

$$d_i = \mathbf{x_i}^\top \beta_{m_0} + r_{m_i} + \upsilon_i \tag{A.6}$$

Where the second equation is the same as before and the first one was obtained by simply substituting equation A.2 into A.1, where $\bar{\beta}_0 = \alpha_0 \beta_{m_0}$, $\bar{r}_i = \alpha_0 r_{m_i}$, and $\bar{\zeta}_i = \alpha_0 \upsilon_i + \zeta_i$.

The post-double selection estimator uses both reduced form equations A.5 and A.6 to select the set of controls. In particular, it will incorporate the controls that are relevant to explain either the outcome in absence of treatment (Equation A.5) and the treatment by itself (Equation A.6). The steps required to obtain the post-double selection estimator are described as follows:

Step 1: Apply a variable selection procedure (the authors use LASSO) to select the controls relevant to explain the outcome in absence of the treatment (Equation A.5). Denote by $\hat{\mathbf{I}}_1 = \{j \in 1, 2, \ldots p \text{ s.t } \hat{\bar{\beta}}_{0,j} \neq 0\}$ the controls whose coefficients were non-zero in the model selection procedure

Step 2: Apply a variable selection procedure (the authors use LASSO) to select the controls that are relevant to explain the treatment (Equation A.6). Denote by $\hat{\mathbf{I}}_2 = \{j \in 1, 2, \ldots p \text{ s.t } \hat{\beta}_{m_0,j} \neq 0\}$ the controls whose coefficients were non-zero in the model selection procedure

Step 3: Based on their knowledge of the problem, the researcher has the option of selecting another set of controls on an *ad hoc* basis that she believes are relevant to avoid bias in the estimation These controls are denoted by $\hat{\mathbf{I}}_3$. In our case, we imposed the area fixed effects to be included into the model.

Step 4: Define by $\hat{\mathbf{I}} = \hat{\mathbf{I}}_1 \cup \hat{\mathbf{I}}_2 \cup \hat{\mathbf{I}}_3$ the union of the controls selected in the previous three steps and by $\hat{s} = \|\hat{\mathbf{I}}\|_0$ the estimated dimension of the model. Now, regress by OLS y_i into d_i and the group of controls included in $\hat{\mathbf{I}}$ and define the post-double selection estimator as:

$$(\check{\alpha}, \check{\beta}) = \arg \min_{\alpha \in \mathbb{R}, \beta \in \mathbb{R}^p} \left[\mathbb{E}_n[(y_i - d_i\alpha - \mathbf{x_i}^\top \beta)^2] \ : \ \beta_j = 0, \forall j \notin \hat{\mathbf{I}}\right] \tag{A.7}$$

The main results established by the authors is that under sparsity conditions and a rich set of data-generating processes our post-double selection estimator is:

- Consistent: $\check{\alpha} \xrightarrow{p} \alpha_0$
- Asymptotically normal $\left(\left[\bar{\mathbb{E}}(v_i^2)\right]^{-1} \bar{\mathbb{E}}(v_i^2\zeta_i^2) \left[\bar{\mathbb{E}}(v_i^2)\right]^{-1}\right)^{-\frac{1}{2}} \sqrt{n}(\check{\alpha} - \alpha_0) \xrightarrow{d} N(0, 1)$

C.3 Selection of controls via feasible LASSO

Although the results presented in the previous section are valid for a wide group of model selectors, the authors use the LASSO selection procedure for most of the results of the paper. In this subsection, I will present some aspects of the LASSO procedure employed by the authors and the calibration of the parameters used in their R package hdm Chernozhukov et al. (2016).

For this part, the authors employ a version of the LASSO geared for heteroscedastic, non-Gaussian cases (Belloni et al., 2012) that solves:

$$\min_{\beta \in \mathbb{R}^p} \mathbb{E}_n \left[(y_i - \mathbf{x_i}^\top \beta)^2\right] + \frac{\lambda}{n}\|\hat{\Psi}\beta\|_1 \tag{A.8}$$

where $\|.\|_1$ is the $l-1$ norm, λ is the LASSO penalty coefficient, $\hat{\Psi} = diag(\hat{l}_1, \ldots, \hat{l}_p)$ is a diagonal matrix of penalty loadings, and $\|\hat{\Psi}\beta\|_1 = \sum_{j=1}^{p} |\hat{l}_j \beta_j|$. Following Belloni et al. (2012) we set the penalty level λ and loadings $\hat{l}_j$ to:

$$\lambda = 2c\sqrt{n}\Phi^{-1}\left(1 - \frac{\gamma}{2p}\right) \quad \text{and} \quad \hat{l}_j = l_j + op(1),\ l_j = \sqrt{\mathbb{E}_n\left[\bar{x}_{ij}^2 \epsilon_i^2\right]}$$

where $c > 1$ and $1 - \gamma$ is a confidence level. Following the package designed by one of the authors (Chernozhukov et al., 2016), we set the value of these parameters as $c = 1.1$ and $\gamma = 0.05$ (the default value of the package). Once these parameters are set, we obtain $\hat{l}_j$ by iteration following Belloni et al. (2012) .

D Segregation stability

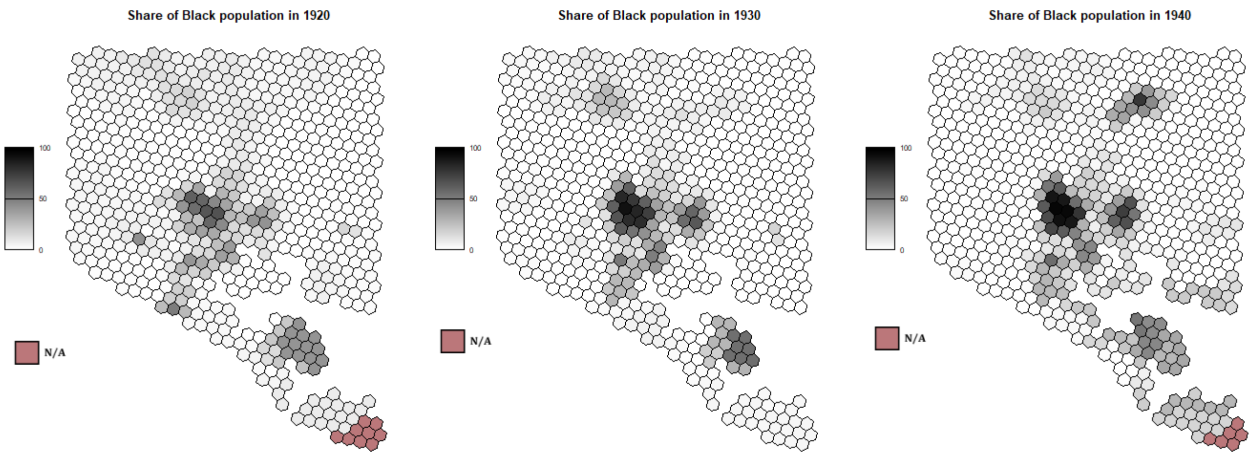


Figure A.1: Share of Black population in Baltimore 1920-1940

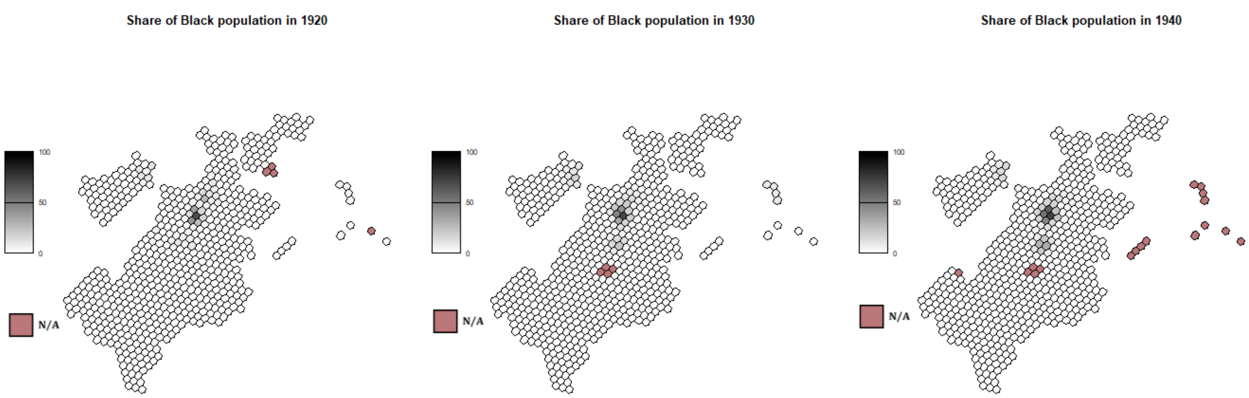


Figure A.2: Share of Black population in Boston 1920-1940

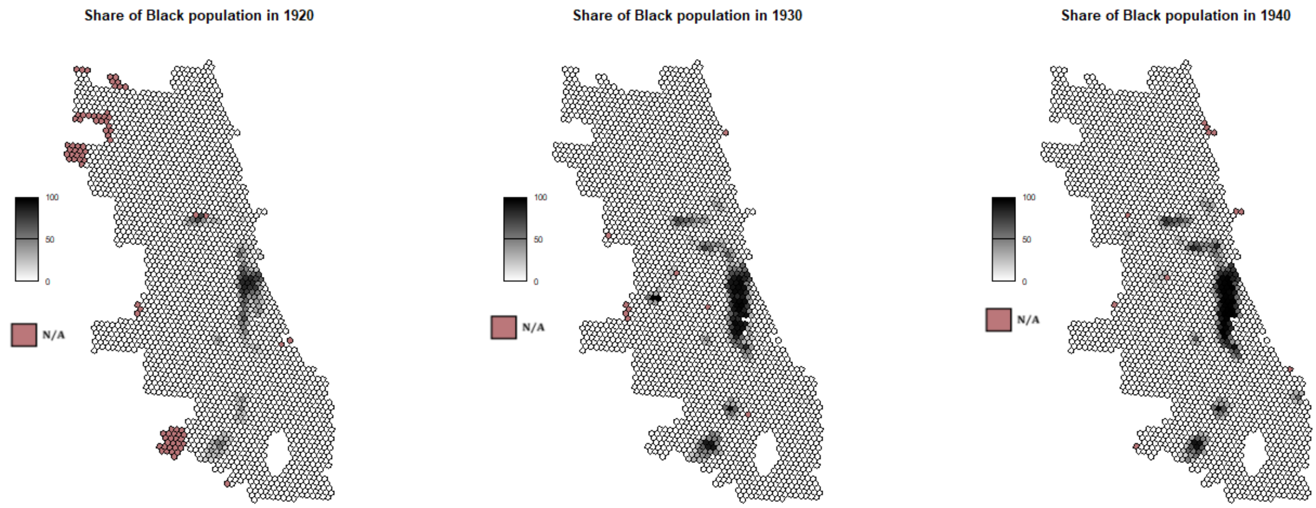


Figure A.3: Share of Black population in Chicago 1920-1940

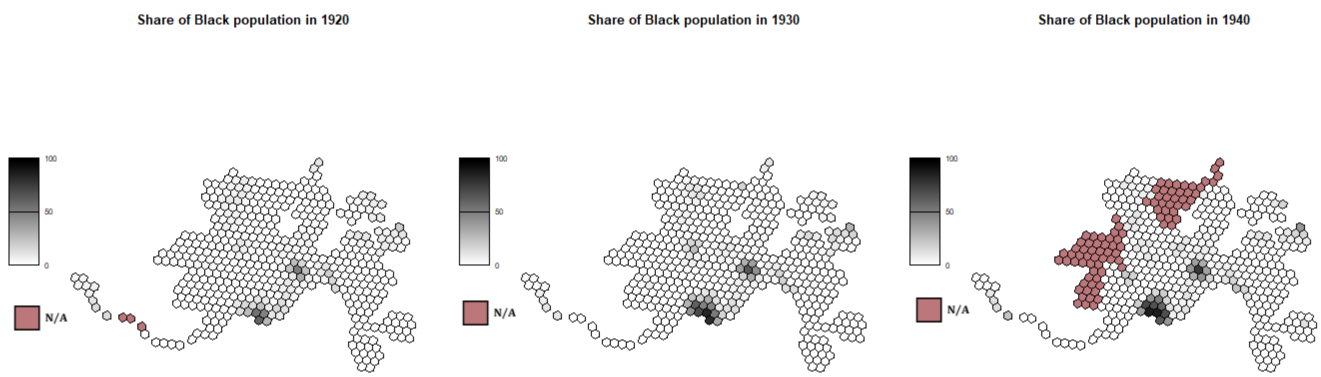


Figure A.4: Share of Black population in Cincinnati 1920-1940

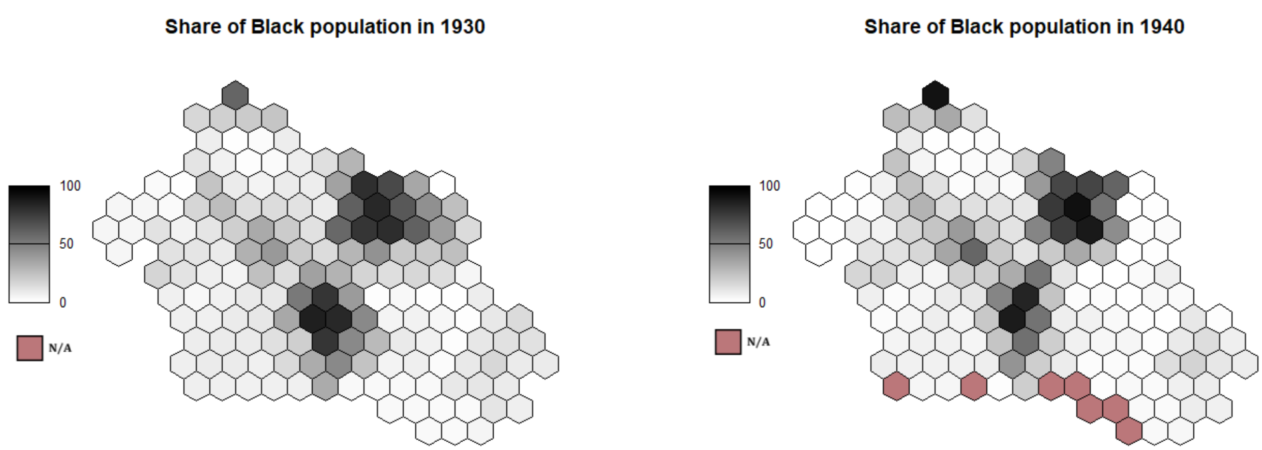


Figure A.5: Share of Black population in Houston 1930-1940

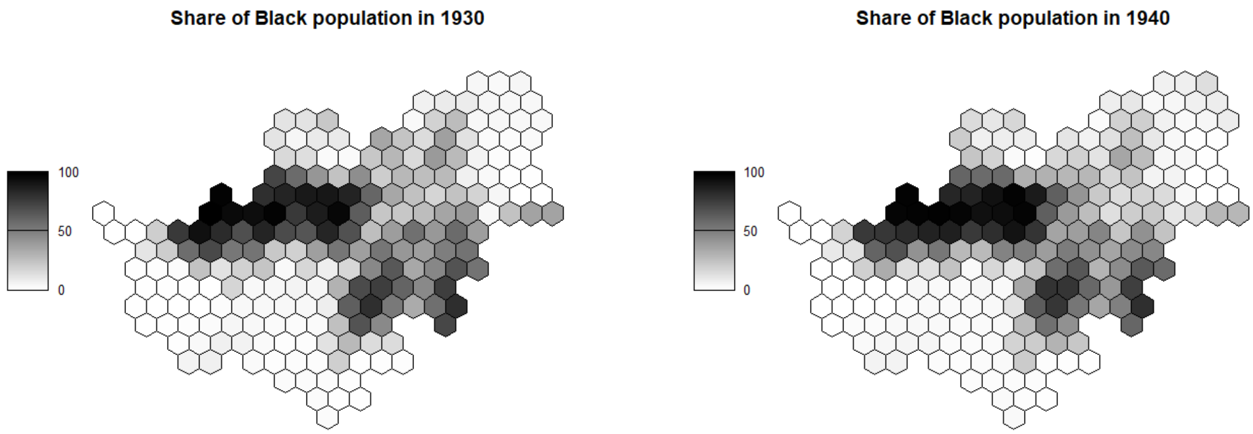


Figure A.6: Share of Black population in Nashville 1930-1940

	Baltimore	Boston	Chicago	Cincinnati	Houston	Nashville
Constant	0.034***	0.001	0.039***	0.033***	−0.001	0.003
	(0.004)	(0.001)	(0.002)	(0.004)	(0.011)	(0.006)
$B_{i,c,t-10}$	1.177***	1.107***	1.142***	1.242***	1.000***	1.011***
	(0.011)	(0.012)	(0.008)	(0.017)	(0.032)	(0.015)
$I_{t=1940}$	−0.044***	−0.001	−0.041***	−0.035***	-	-
	(0.005)	(0.001)	(0.003)	(0.006)	-	-
R^2	0.90	0.87	0.81	0.88	0.87	0.96
Num. obs.	1310	1222	4776	762	154	213

Note: Weighted regression for Equation 6 specification for the percentage of Black population for each city under study. Population weights correspond to the average population size of the Synthetic Neighborhoods in 1930-1940. Standard errors are reported in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A.2: Results of unweighted stability specification

E Other Tables and Figures

	Unemployment Rate
β	0.133
	(0.292)
Num. obs.	1310

Note: Difference-in-Differences estimation for the effect of the introduction of landfills and waste incinerators on the unemployment rate of the affected inhabitants in Baltimore in between 1930 and 1940. Variables are in standardized. Clustered standard errors at the Ward-year level are reported in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table A.3: Difference-in-Differences estimations for unemployment for Baltimore