4-2023

# Infrastructures of Race? Colonial Indigenous Segregation and Contemporaneous Urban Sorting

Luis Baldomero-Quintana
*William & Mary*, labaldomeroqui@wm.edu

Guillermo Woo-Mora

Enrique De la Rosa-Ramos

# Infrastructures of Race?
# Colonial Indigenous Segregation and Contemporary Urban Sorting*

Luis Baldomero-Quintana ⓡ †, L. Guillermo Woo-Mora ⓡ ‡, Enrique De la Rosa-Ramos§

April 4, 2023

**Abstract**

We study the impacts of a colonial segregation policy on modern-day spatial population patterns and residential sorting by human capital in Mexican cities. After the Conquest, the Spanish aimed to segregate Indigenous individuals into settlements called *pueblos de indios*. While the segregation policy lasted until the end of the colonial era, we use present-day census data at the block level on population, schooling and access to medical services to understand the persistent effects of *pueblos* on within-city structure. First, we document a spatial non-monotonic correlation between the location of the *pueblos* and population deagglomerations. Second, we study the causal impact of the *pueblos* on sorting by human capital by exploiting quasi-experimental variation created by *pueblos*' catchment area defined by colonial laws. Using a Regression Discontinuity Design, we find a slight increase in the blocks' share of households lacking access to medical services near the boundary of all *pueblos*. We further exploit the degree of success of the policy: it effectively isolated the Indigenous population only in a fraction of the settlements. We show that blocks near to the *pueblos* where the segregation policy was successful experience discontinuities on the blocks' measures of schooling of minus 0.38 years and the blocks' share of individuals with no access to medical services of 2.2 percentage points. Worse urban amenities and lower land prices explain our results. Our findings show that colonial segregation institutions have persistent impacts for centuries on urban sorting, even when the originally targeted group becomes a very small share of the modern-day population.

**JEL Codes:** N96, O18, J71, N36

**Keywords:** urban segregation, sorting, spatial persistence, cities and development

*We received financial support from the Economic History Association. All authors contributed equally to this manuscript. We randomized authors' names with the AEA Randomization Tool (code: uq6rhN4k90I_). We are thankful for the detailed comments of Camilo Acosta, Leah Brooks, John Lopresti, and John Parman, and the program Adopt a Paper by Elira Kulka and Danila Serra. We are grateful to Jennifer Alix-Garcia, Danae Hernandez, Eduardo Ortiz-Juarez, Felipe Castro Gutiérrez, and Dorothy Tanck de Estrada for data guidance. We benefited from feedback of Stefania Albanesi, Nathaniel Baum-Snow, Hoyt Bleakley, Kerem Cosar, Don Davis, Sebastian Ellingsen, Carlos Hurtado, Chinhui Juhn, Kalle Kappner, Nicolas Morales, Priya Mukherjee, Alexander Persaud, Jorge de la Roca, Yanos Zylberberg, Tom Zylkin, and participants at the MEA, AEA-ASSA, North American UEA, and European UEA annual meetings, FED Richmond, Smith College, University of Richmond, University of South Carolina, and the Scottish Economic Society Annual Conference. Annika Hoenack-Herrejon and Justin Marquez-Talavera provided excellent research assistance.

†Department of Economics, William & Mary, labaldomeroqui@wm.edu

‡Department of Economics, Paris School of Economics, guillermo.woo-mora@psemail.eu

§Department of International Development, King's College London, enriquedelarosa@kcl.ac.uk

# 1 Introduction

How do colonial segregation policies define the spatial sorting of contemporary cities in developing nations? Colonialism required the spatial division and organization of colonized territories to manage Indigenous populations effectively. The colonial spatial organization was sometimes based on new racial and ethnic hierarchies (e.g., U.S., Mexico, South Africa). In other cases, colonialism created spatial arrangements based on previous racial or ethnic divisions (e.g., India, Rwanda). Existing work allows us to understand the impacts of colonial institutions on economic growth and present-day institutions (Nunn, 2014, 2020; Koyama & Rubin, 2022). Recent literature also underlies the role of historical events on between and within-city structure (Hanlon & Heblich, 2022; Lin & Rauch, 2022). Although existing work has shown that segregation policies from the 19th and 20th centuries have persistent impacts on residential sorting (Shertzer et al., 2022; Yamagishi & Sato, 2022), whether colonial segregation policies have the same effect remains an empirical question. An answer to this question allows us to fully understand how history shapes present-day cities, and thus improve the design of placed-based policies.

We study the long-lasting effects of a colonial segregation policy that emerged five hundred years ago on residential sorting by human capital in Mexican cities. The Spanish Crown aimed to segregate the Indigenous population in small settlements called *pueblos de indios*. By the end of the colonial period, the segregation policy had an ambiguous success: as Estrada et al. (2005) explains, some *pueblos* remained inhabited exclusively by the Indigenous population (hereinafter referred to as *Indigenous Pueblos*), while others had also poor Spaniards, Indigenous, and other castes of individuals (hereinafter called *Mixed Pueblos*). The intervention was temporal since it ended *de facto* in 1821 with Mexico's Independence from Spain. Because these settlements were relatively small compared to the size of modern-day Mexican cities, we can use them to study within-city sorting patterns.

We estimate the impacts of *pueblos de indios* on two sets of outcomes at the block level: populations agglomerations and human capital sorting. We consider two measures of human capital: years of education and share of households with lack of access to medical services. Based on the 2020 census, we use the universe of city blocks within Mexican Metropolitan Statistical Areas (MSA) that contain at least one colonial *pueblo*. Our impacts exclusively consider how a *pueblo de indios* affects the closest city blocks. Thus, we obtain the distance of each modern-day city block to the nearest historical *pueblo*.

In the first part of the paper, we obtain spatial correlations to document the relationship between the *pueblos*' location and granular patterns of present-day population density. We find the empirical regularity that *pueblos de indios*' location is associated with non-monotonic population deagglomerations. Parametrically, a 10% rise in a block's distance to the nearest *pueblo de indios* is correlated with a reduction in the population density of the block by 1.07%, conditional on geography and market access. The correlations are robust to unobserved spatial heterogeneity.

In the second part of the paper, we exploit quasi-random variation in the segregation policy land use to estimate its long-lasting effects on human capital sorting. Colonial laws established that *pueblos de indios* must have a ratio of 600 *varas*, equivalent to 500 meters, to protect the settlements from land takeovers by the Spanish (Estrada et al., 2005; Castro Gutiérrez, 2010b). These boundaries allow us to use a spatial Regression Discontinuity Design (RDD) to obtain the causal estimates of *pueblos* on sorting patterns. To ensure balanced characteristics between control and treatment groups, we restrict our sample to blocks at most 1 km away from the location of the *pueblos*. Drawing on the *pueblos*' boundaries established by colonial laws, we define the *pueblos*' catchment area as a 0.5 km radius and define as running variable the distance of a city block's centroid to the boundary of the nearest *pueblo* catchment area. We then compare the human capital outcomes of city blocks within the *pueblos*' catchment area with those blocks outside of it (0.5 to 1 km away from the closest *pueblo*).

We use Calonico et al. (2014) for estimation and inference, and we implement a donut spatial RDD, as in Laudares & Caicedo (2023) and Barreca et al. (2016), to account for fuzzy borders and the potential sorting of city blocks into the catchment areas of the *pueblos* caused by urban design. Our spatial RDD estimates suggest there is sorting by human capital. Blocks within a *pueblos*' catchment area have a discontinuity in the share of households without medical services by 0.12 to 0.21 standard deviations, equivalent to a rise of 1.8 to 3.2 percentage points. However, we find no discontinuity in the blocks' average years of schooling.

Given the relevant heterogeneity in the segregation policy, we explore how the degree of effectiveness of the historical segregation policy shapes our results (Gelman et al., 2023). We split our sample of blocks in two according to the type of the nearest *pueblo*, *Indigenous* or *Mixed*, and obtain spatial RDD estimates for each sub-sample. We find that *Indigenous Pueblos* negatively impact blocks' contemporary human capital outcomes. The discontinuity in blocks' average years of education is minus 0.16 standard deviations in our most conservative estimate, equivalent to minus 0.38 years of education. The discontinuity in blocks' share of households with no access to medical services represents an increase in 0.14 standard deviations, equivalent to 2.2 percentage points. The results for the impacts of *Mixed Pueblos* on blocks' measures of human capital outcomes are inconclusive. The RDD estimates of the treatment of *Mixed Pueblos* on blocks' years of education are not statistically significant. Moreover, the estimates of the impacts of these *pueblos* on blocks' share of lack of access to medical services provide inconclusive results due to the lack of robustness. Our results lead us to conclude that city blocks' located in the catchment area of the *pueblos* where segregation was successful have worse measures of human capital relative to the control group, even two centuries after the end of the colonial policy.

Our heterogeneity results are robust to different bandwidths and cross-validation exercises, allowing us to rule out that a few *pueblos* drive the results. Moreover, we implement two falsification or placebo tests. First, we create mock-up boundaries holding constant *pueblos*' location and show that there is no discontinuity farther away from the actual cutoffs. Second, taking critiques of the persistence literature seriously (Abad & Maurer, 2021; Kelly, 2020), we develop a stringent falsification exercise where we randomly assign *pueblos* to fake locations and implement again our RDD. Our estimate of the effect of *Indigenous pueblos* on sorting by years of schooling is the only one that passed the exacting falsification test.

Finally, we examine the mechanisms driving our results. Using data from Mexico City MSA, the capital and biggest metropolitan area, we study the impact of the *pueblos de indios* on contemporary urban amenities and public goods. We observe that blocks in the catchment area of *Indigenous Pueblos* experience a reduction in the quality of the street relative to the control blocks, measured by characteristics such as pavement quality, lights, street signs, street sewage, sidewalks, etc. We also observe a reduction in land prices by 40% at the boundary of the *Indigenous Pueblos*. Thus, although *pueblos* do not directly affect human capital, they likely have effects on amenities and public goods that explain why individuals with lower human capital sort into these areas.

Our paper has three contributions to the economics literature. Our first and main contribution is that we provide evidence that colonial segregation policies can shape the spatial configuration of modern-day cities, even after such policies ended. Moreover, our results show that their effects depend on the degree of success of segregation in the past. Second, we document the historical causes of residential segregation in a developing nation context. Third, we provide evidence that historical shocks that occurred centuries ago can influence contemporary residential sorting in urban areas.

The main contribution of this paper concerns historical economic shocks that influence the spatial patterns of cities (Hanlon & Heblich, 2022; Lin & Rauch, 2022). There is a growing body of work regarding the persistent effects of these shocks on cities in Western high-income nations (Ahlfeldt et al., 2015; Ambrus et al., 2020; Bleakley & Lin, 2012; Brooks & Lutz, 2019; Heblich et al., 2021; Nitsch & Wolf, 2013; Shertzer et al., 2022, 2016; Redding et al., 2011). Recent studies have looked at the formation and path dependence of Latin American cities (Alix-Garcia & Sellars, 2020; Ellingsen, 2021). These empirical papers align with the theoretical framework of Allen & Donaldson (2020), who provide theoretical conditions under which a historical shock determines spatial equilibrium outcomes across time. We complement these strands of literature in two ways. First, we provide evidence that historical shocks can persist through many centuries, not just a few decades. Second, we document how colonial segregation policies shape the present-day spatial configuration within cities in a developing nation context.

Our paper also speaks to existing work on the historical factors of the 19th and 20th-century that drive contemporary urban segregation and inequality in American cities (Ananat, 2011; Boustan & Margo, 2013; Boustan, 2010; Cook et al., 2018; Roca et al., 2018; De la Roca et al., 2014; Logan & Parman, 2017; Rothstein, 2017; Baum-Snow & Lutz, 2011; Brinkman & Lin, 2022; Collins & Margo, 2007; Brooks et al., 2020; Shertzer et al., 2016). In addition, Shertzer et al. (2022) provides a detailed summary of how historical zoning regulations in the U.S. affect past and modern-day racial segregation. We add to this strand of work by presenting empirical evidence that past segregation impacts today's residential sorting by human capital in a context where the originally targeted group represents a very small fraction of the population in the modern-day: Indigenous population in Mexican cities represents only 4%.[1] Such an environment is different than the case of the U.S., where the ethnic and racial groups segregated in the past still represent a significant share of the population today.

The idea that urban segregation persists even in a historical environment with the presence of demographic shifts is studied by Yamagishi & Sato (2022) and Woo-Mora (2023). Firstly, Yamagishi & Sato (2022) studies the persistent impacts of an ethnic segregation policy in Japanese cities. In their environment, the targeted group is indistinguishable from the rest of the population; hence the historical persistence operates through the stigma of where the targeted individuals lived. They find large land price discounts across time in the initially segregated areas. We complement their work by studying a segregation policy that was eliminated two centuries ago with consequences on residential sorting by human capital. Moreover, their land price penalties are of a similar size to ours. Secondly, Woo-Mora (2023) documents the long-term negative impacts of a single *pueblo de indios* on individual outcomes of the inhabitants of Guadalajara city, Mexico. In his context, even though amenities do not differ significantly across the *pueblo*'s boundary, the housing market experiences a major land price discount over time. After he accounts for such a discount, he finds human capital gaps through multiple periods suggesting the persistence of placed-based stigma. We complement this work by expanding our analysis to all Mexican cities with at least one *pueblo*, 73% of all metropolitan areas, thus increasing the external validity of the persistent impacts of the *pueblos de indios*. Furthermore, we highlight the importance of the colonial segregation policy's success in determining the current urban sorting by human capital of Mexican cities.

We also contribute to the literature on the present-day determinants of residential sorting and urban segregation by income in cities of developing nations. (Pérez Pérez et al., 2022; Sabatini, 2003; Tsivanidis, 2022; Velazquez-Cabrera, 2021; Warnes, 2021; Selod & Zenou, 2003) focus on how public transportation and school choice impact urban sorting. Christopher (2001); Bharathi et al. (2019) describe modern-day segregation in India and South Africa, two nations with salient ethnic, racial, or caste identities. We complement this literature by providing evidence that historical shocks during colonial rule can influence current residential segregation outcomes in cities from developing nations.

This paper also adds to the literature on the persistent impact of colonial institutions in development outcomes in Latin American nations (Artiles, 2022; Dell, 2010; Garfias & Sellars, 2022; Laudares & Caicedo, 2023; Valencia Caicedo, 2019; Waldinger, 2017; Chiovelli et al., 2022). Another strand of literature highlights the importance role of pre-colonial and Indigenous institutions in shaping regional development outcomes with a focus on Mexico (Arias & Girod, 2014; Arteaga, 2018; Bobonis et al., 2021; Elizalde, 2020; Diaz-Cayeros et al., 2022; Diaz-Cayeros, 2011). We do not study institutions created by Indigenous individuals per se, but rather, our focus is on colonial institutions imposed on them through the urban space. Recent criticisms to the literature on historical persistence highlight the need to study with detail the spatial dimension of outcomes when economists study the causal impact of historical shocks

[1] The dynamics of the Indigenous population decline in urban areas respond to multiple factors such as epidemics, multiracial/multi-ethnic marriages or *mestizaje*, government policies, and identity shifts (Tenorio-Trillo, 2015)

on contemporary outcomes Abad & Maurer (2021); Kelly (2020). We consider the latter considerations in our study and complement the literature by showing that besides political economy, institutional and development outcomes, colonial institutions can also determine the urban space and the spatial distribution of the population. We provide evidence of the latter by showing the persistent effects of colonial segregation policies on residential sorting.

Lastly, our results speak to a large literature that studies the impacts of racial and ethnic diversity on economic growth and the provision of public goods, summarized by Alesina & Ferrara (2005). More recent work by Artiles (2022) finds that Indigenous ethnic diversity influences positively the provision of public goods in Peru. We add to this literature by contrasting the outcomes of areas characterized by successful Indigenous segregation vs. areas where the process of *mestizaje* (multiracial and multiethnic marriages) created mixed ancestry *pueblos*.[2] Our findings are compatible with the results of Artiles (2022): *pueblos* with successful indigenous segregation lead to worse present-day outcomes across city blocks in Mexican cities through the presence of worse urban amenities. However, *pueblos* with mixed-ancestry at the end of colonial times do not significantly impact blocks' measures of human capital, even though land prices affected by these *pueblos* have a premium.

The rest of the paper proceeds as follows. In Section 2 we document the historical context of the *pueblos de indios*. Section 3 describes our data, and we present descriptive statistics of the blocks in our sample. Section 4 presents our empirical model. We describe our main empirical findings in Section 5 and document the mechanisms behind our results in 6. In Section 7, we conclude.

## 2 Historical Context

Mexico's civilization origins have been traced back to 7200 B.C. (MacNeish, 1964).[3] Its spatial and political structures were diverse, ranging from stateless people, city-states, and a congeries of empires (Knight, 2002). However, most settlements in this period were small political units called *calpullis* where a group of families lived.[4] Moreover, archaeological evidence points out that pre-colonial cities consisted of communal centers with a market, a temple, and the residences of nobility surrounded by subordinates (Calnek, 1976; Alcántara, 2004; Gerhard, 1993). These settlements often had a nuclear pattern where density decreased from the center towards the periphery of the state (Sanders, 1956).

In contrast, sometimes, bigger areas with many *calpullis* formed states which, in some cases, dominated

[2] Historians define *mestizaje* to the phenomenon that defines the contemporary ethnic and racial configurations of Mexican society: a process of centuries of multiracial and multi-ethnic marriages between Spanish and Indigenous individuals. The process also included multiracial marriages with former African enslaved individuals, albeit to a lesser extent. The *mestizaje* diminished the presence of the Indigenous culture in Mexico. Moreover, Mexican governments implemented policies to accelerate *mestizaje* and reduced the weight of Indigenous traditions and culture in the country, Altamirano & Trejo (2016).

[3] Mexico's pre-colonial civilizations were located in the region named Mesoamerica. This region covered most of Mexico and Central America.

[4] A *calpulli* was an alliance of families bound by cultural, communal, political, agricultural, and educational ties (Colín, 2014)

other larger areas or regions that were subjected to taxation, known as *tributo* (Gerhard, 1993).[5] These largest states, in some cases, became empires. However, research indicates that empires were rare Gerhard (1993). Tenochtitlan was the most important empire at the time, often called the Mexica or Aztec Empire. The empire was located in what today is Mexico City, the most populated city in Mexico.

The above political, social, and economic organization was drastically altered with the Spaniards' conquest of Mesoamerica from 1518 to 1521. The fall of the Mexica or Aztec empire and the emergence of the Viceroyalty of New Spain marked the start of the colonial era. With the aid of different Aztecs' political and ethnic group rivals, the Spaniards cause a swift subjugation of the empire. This rapid conquest was a considerable shock to the political organization of pre-colonial settlements (Knight, 2002). However, the real political accomplishment of the Spanish conquerors was not the conquest itself but rather their capacity to politically control the Indigenous settlements by "recycling" some political contracts the Indigenous empires had used (Arteaga, 2018).

The geographic distribution of people and economic activity experienced an important change at the start of the colonial times. In 1538, the Spanish Crown enacted a law creating the institution of *Repúblicas de Indios* (Boix, 1841). *De jure*, such institution gave Indigenous communities rights over land and permitted some extent of self-government. Later on, spatial units with such Indigenous governing bodies were known as *pueblos de indios*. *De facto*, the communities were complex institutions: they were both a hostile places where Indigenous people had to comply with specific duties and taxes as well as a space where they could participate in their communities, maintain common goods, and celebrate their festivities (Castro Gutiérrez, 2013). Moreover, they had autonomy regarding the land tenure (Van Young, 1984). The *pueblos de indios* were useful for the Spanish because they also served for public administration and local governance purposes, including taxation, provision of public goods, and administrative records (Castro Gutiérrez, 2010a; Granados, 2013; De Estrada, 1999). Nevertheless, beyond its formal administrative purposes, the *pueblos de indios* served as a segregation policy (Nemser, 2017).

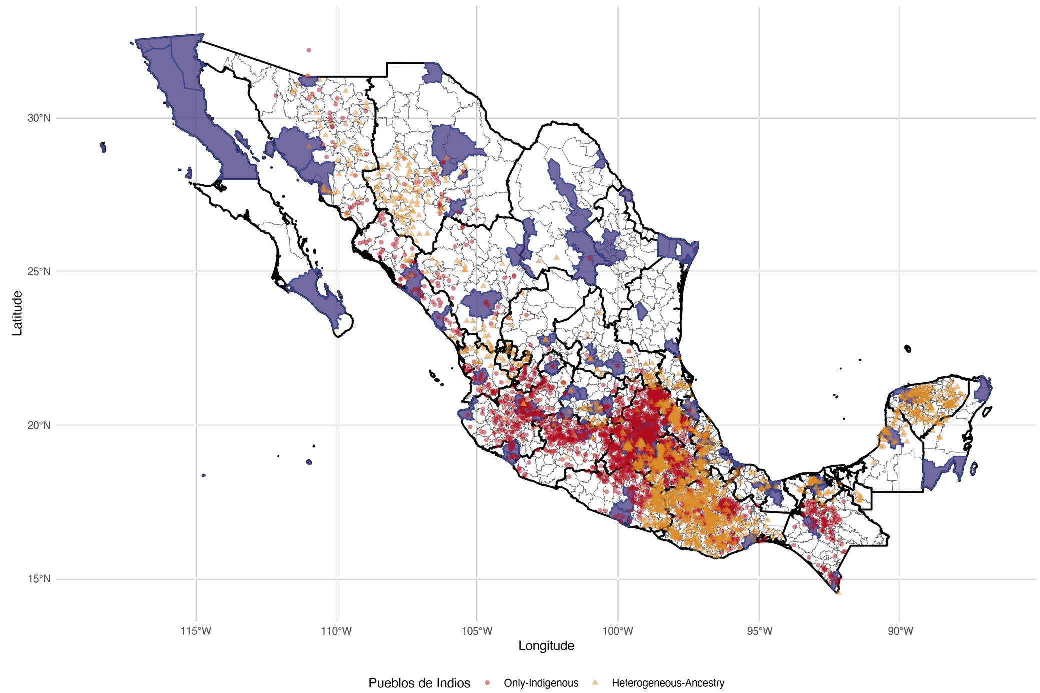
The newly created settlements were relatively small. Mexico's historiography has studied the size of these pueblos, and it coincides that a typical pueblo was about *"600 varas"* (Estrada et al., 2005; De Estrada, 1999; Castro Gutiérrez, 2010a). Moreover, Estrada et al. (2005) argues that by the eighteen century, a *pueblo de indios* had a population of about 360 individuals. The configuration of these *pueblos* consisted of a church, political authorities, and cropland. Estrada et al. (2005) also provides an alternative measure for extending a pueblo used informally at that time. This measure was the number of people that can be reached by the sound of a church bell.

There are two main historical origins explaining *pueblos*' origins. First, some emerged from pre-existing Indigenous settlements (Arteaga, 2018). A second type of *pueblos* were segregated towns artificially created by the Spaniards or the migrant Indigenous population during colonial times (Woo-Mora, 2023;

[5] This *tributo* was a payment for protection or imposed after a conflict between the Aztecs and the dominated societies.

Loreto López, 2010).[6] While some *pueblos* disappeared by the late colonial period (Diaz-Cayeros et al., 2022), their structure persisted throughout most of the colonial period because it permitted taxing isolated settlements. By the late 18th century, authorities in New Spain made substantial administrative reforms to centralize power and improve tax administration –also known as *Bourbon Reforms*. De Estrada (1999) argues that as part of these reforms, authorities documented more precisely the existence and location of the *pueblos de indios*. Figure 1 shows the spatial distribution of the *pueblos* across the territory encompassing nowadays Mexico. The *pueblos* were concentrated in central and southern Mexico.

Figure 1: *Pueblos de Indios*' location




**Notes**: The location of the *pueblos de indios* is based on Estrada et al. (2005). The map shows the borders of municipalities (light black), the delimitations of Mexican states (dark black), the Mexican MSAs (purple), and the *pueblos de indios* (red dots and yellow triangles).

The segregation policy had a heterogeneous success. There are historical documents outlining that given social and economic dynamics, there were two different kinds of *pueblos*. In one case, the *pueblos de indios* were relatively isolated from interactions with *Spanish Peninsulares* (Spaniards born in the Iberian Peninsula) or *Criollos* (the descendants of Spanish born in the Americas). As Castro Gutiérrez (2010a) documents, sometimes the isolated *Pueblos* saw a Spanish person only once a year for the collection of taxes for the Spanish Kingdom. The other case refers to those *pueblos* that became part of the colonial cities, as Lara-Cisneros (2013), Castro Gutiérrez (2010a), and De Estrada (1999) document. The inhabitants in the latter settlements constantly interacted with *Mestizos* (descendants of Spanish and Indigenous individuals), *Criollos*, and *Spanish Peninsulares*. Such interactions were so frequent that in some cases, the inhabitants of some *pueblos* sold their land to Spanish *Peninsulares*, *Criollos*, or *Mestizos*

[6]The historical literature suggests that most of the institutional patterns of the *Pueblos* were highly influenced by colonial policies that successfully copied or recycled former Indigenous institutions, such as the *calpulli*, to encourage local governments and maintain political stability.

looking for affordable housing.

The degree of interactions between *pueblo*'s inhabitants and other social actors depended on multiple geographic, economic, and social factors. The latter could occur due to agricultural, language, religion, trade, and labor relationships; location, the fact that natural barriers between the main colonial city and the *pueblos* were not large; or because the colonial urban sprawl started to reach the territories of *pueblos de indios* (Estrada et al., 2005). However, in some cases, *pueblos* only had an Indigenous population even if they were part of the colonial urban sprawl due to specific geographic characteristics, as Loreto López (2010) shows for the case of the city of Puebla. It is important to highlight that the *Pueblos* closer to the colonial cities likely had more frequent interactions with other castes or ethnic groups (Anderson, 1988).[7]

Socioeconomic interactions were not the only determinants of how the inhabitants of the *pueblos* interacted with other groups. Language was one major obstacle that could impede social interaction between indigenous inhabitants and Spaniards. When *pueblos de indios* emerged under colonial rule, Spaniard authorities debated if they should be taught religion in Spanish or in its native language. This situation lasted until the end of the seventieth century. From this period onwards, the Spaniards opened schools to spread the Spanish language across the native population (De Estrada, 1999).[8]

By the end of the colonial period, the Cadiz Constitution in 1812 eliminated the *pueblos de indios de jure*. Hence, the *pueblos* became *Ayuntamientos*. Moreover, the new legal order gave everyone the same rights regardless of their race. For example, the legal differences between Indians and Spaniards disappeared: Indigenous populations were given the same rights as Spaniards, such as citizenship, property rights, and commerce. As Lira (1995) noted, this was not an automatic process but the starting point that would define Mexico´s independence era in the forthcoming decades. The segregation policy in the form of *pueblos de indios* terminated with Mexico's independence from Spain, and the legal institution ended *de facto* in 1821.

In the aftermath of the colonial period, the geographies encompassing *pueblos* persisted but faced major changes. For instance, land tenure reverted drastically during the mid-decades of the eighteen century. Nineteenth-century liberal governments promulgated the *Ley Lerdo* (1856), one of the most dramatic changes in the land tenure regime Mexico has experienced, expropriating rural land. The new law prohibited ecclesiastical and corporate institutions' property ownership that was not being used for the purposes of these institutions (Bazant, 1966). These lands were auctioned or sold. Indigenous populations owning corporate lands held were forced to sell their properties (Fraser, 1972). Later on, by the turn of

[7] Spaniards implemented a social stratification policy. They designated themselves as *gente de razón* (rational people) as they considered themselves the only ones capable to make rational decisions. In contrast, they judged the Indigenous population as inferior and called them *república de indios* (republic of Indians). This social system is often referred as a *"dual republic"* (Vinson III, 2017)

[8] Other reforms also altered the relationship between the *pueblos* and Spaniards. In 1772, the Catholic authorities endeavored a major reform aiming to relocate the distribution of churches in Mexico. Before 1772 the churches´ location was determined merely by racial segregation (García Redondo, 2019), that is, indigenous populations and Spaniards attended different churches. The new reform allowed everyone to attend the church of their preference, regardless of their residency.

the twentieth century, *pueblos* were part of the new industrialized and growing size cities (Tenorio-Trillo, 2015; Lira, 1995). By 1929, only 19% of Mexico City's population identified as *indio* or indigenous, but inhabitants of former *pueblos* overall were considered as "*el pueblo*" or had low social status (Tenorio-Trillo, 2015). Moreover, while gentrifying Mexico City's downtown, elite urban planners and speculators often saw indigenous populations living in the city as individuals that the *new ideal city* would educate and civilize (Tenorio-Trillo, 2015).

# 3 Data and descriptive statistics

## 3.1 Data sources

We use three data sources: geospatial data on *pueblos de indios*' historical locations, Census data with information on present-day economic and geographic outcomes at the block level, and spatial datasets with information on public goods and urban amenities by block.

**Historical data**. We use georeferenced location data of the *pueblos de indios*. (Estrada et al., 2005) compiled information on all *pueblos*. They relied on the colonial censuses and administrative records produced by the Spanish Crown during the implementation of administrative and governance reforms to the *pueblos* during in the late 1700s (Estrada et al., 2005). The authors document the existence of 4,468 *pueblos*. The dataset indicates with a dummy variable whether a *pueblo* had only Indigenous inhabitants or if it had inhabitants of multiple ancestries (Indigenous, African, and Spanish).[9] Figure 1 displays the location of *pueblos* across Mexico.

**Contemporary outcomes**. We rely on Mexico's 2020 Census for the city block data. We construct a georeferenced dataset for the universe of urban blocks in municipalities that are part of MSAs that contain at least one *pueblo* (53 out of 73 metropolitan areas). This represents around 600,000 blocks. The sample includes both urbanized and non-urbanized blocks. In Mexico's Census, blocks consist of closed groups of housing units divided by streets and avenues. Table A.1 in the Appendix contains descriptive statistics of these blocks.

Mexico's Census also records variables on socioeconomic outcomes. We use four main outcomes of interest in our regressions at the block level: population, population density, average years of schooling, and share of households with lack of access to medical services. For balance tests and controls in some specifications, we use block-level data on the demographic composition of the block (age, sex, Indigenous descent, African descent, religion, sex of the head of household), migration, housing, and unemployment. Additionally, we use block-level data related to the geographic characteristics of each block. These characteristics are: altitude, pollution, the maximum and minimum temperature recorded in a year, average yearly

[9]The dataset of Estrada et al. (2005) contains other variables such as the *pueblos de indios*' names, coordinates, modern-day town name, contemporary municipality, present-day state, and the Catholic saint that the *pueblos de indios*' dwellers worshipped.

precipitation, distance administrative limits of the MSA, distance to the central business district (CBD),[10] and estimates of the historical population and location of colonial buildings.[11]

**Data for mechanisms.** We include six georeferenced datasets to understand the mechanisms behind our results. First, we use the teacher-student ratio for primary and secondary schools in Mexico City MSA from the last Education Census created in 2013 by Mexico's official statistical agency, INEGI. Second, we use records of every police report in 2019 from the Mexico City Prosecutor Office. These records include information on the type of crime, location, and date. The Prosecutor Office records are quite detailed and range from nonviolent minor robbery to major crimes like rape, homicide, or kidnapping.[12]

Third, we use INEGI's National Housing Inventory 2020. This dataset contains information on the quality of the street characteristics of every block.[13] Fourth, we use data on the location of retail trade establishments from INEGI's Census of Economic Units 2019. Lastly, we use information from the Cadastre Office's records in Mexico City on land prices. The city's Department of Finance uses these records to collect property taxes every year.

### 3.2 Descriptive statistics

We provide descriptive statistics for all city blocks in the metropolitan areas that have at least one *pueblo de indios*. We use this sample to obtain spatial correlations between distance to the *pueblos* and population density. We present sample statistics for demographic, housing, and geographic variables in Table A.1 in the Appendix. We highlight three characteristics of the blocks here. First, there are high levels of heterogeneity regarding housing units and population. The differences between quartiles for these variables are sizeable. Second, most blocks have similar demographics. Noticeably, female, Indigenous, and Afro-origin population shares remain relatively similar across quartiles. Third, the blocks have heterogeneous economic geography, as the last rows in Table A.1 show. Noticeably, the mean of the share of Indigenous inhabitants in a block is 4%. Hence, the originally segregated group is less likely to live in the city blocks that we consider in our sample.

[10]We define the location of the central business district as the mean point in latitude and longitude of all finance, insurance, real estate, and banking establishments reported by the National Statistical Directory of Economic Units (DENUE), a census of all establishments in Mexico created by the National Institute of Statistics and Geography (INEGI).

[11]INEGI produces the geospatial data for altitude, latitude, and longitude. The geospatial data on weather come from the National Aeronautics and Space Administration (NASA)'s Daily Surface Weather and Climatological Summaries (Daymet) project. Our geospatial data on pollution is for PM2.5 particles in 2018, and the Atmospheric Composition Analysis Group created the shapefiles at the Washington University in St Louis. Estimates of the historical population come from the HYDE project by Klein Goldewijk et al. (2017). Stangl (2019) produces the spatial data from colonial buildings.

[12]In Mexico, the local police and the state prosecutors' offices must coordinate during the investigation when a crime occurs. They are the first point of contact for any criminal investigation. It is possible that an investigation of a federal crime, such as a murder connected to mafia activities, will be transferred to the Federal Prosecutor Office. Nevertheless, the crime is recorded first by the state prosecutor's office.

[13]The National House Inventory dataset includes characteristics of the street where a blocks is located such as pavement quality, lights, crosswalk, transit signs, street name signs, trees, street sewer, street covering, sidewalks, wheelchair ramps, public transit stop, visual and audible stop lights, trees, bike infrastructure, and lack of informal vendors.

# 4 Empirical strategy

We study the effects of colonial segregation on shaping contemporary urban sorting in Mexican cities. To do this, we use two different empirical strategies. In the first, we document empirical regularities regarding the relationship between *pueblos de indios* and modern-day spatial population patterns. In the second empirical framework, we study *pueblos de indios*' effects on within-city sorting patterns.

**Spatial population patterns and *pueblos de indios*.**

We use the following specification to understand how a block's distance to its nearest *pueblo de indios* correlates with its population density.

$$PopDensity_b = \alpha_{MSA} + \beta \cdot (Distance\ to\ nearest\ Pueblo)_b + X_b'\Omega + \epsilon_b \tag{1}$$

where $b$ denotes a block, our unit of observation. Each block $b$ is within an MSA and we match it to the nearest *pueblo*; $PopDensity_b$ represents the population density in block $b$; $(Distance\ to\ nearest\ Pueblo)_b$ is the geodesic distance between block $b$ and its nearest *pueblo de indios*; and $X_b$ is a vector of geographic controls for block $b$ such as temperature (maximum and minimum), average yearly precipitation, altitude, slope, pollution (measured by PM 2.5 particles), longitude, latitude, and whether the block is urbanized or not. Our sample consists of all blocks (urbanized or not) within municipalities that are part of any MSA that contains at least one *pueblo*. A total of 53 out of 73 metropolitan statistical areas have *pueblos de indios*, representing about 38% of the country's population.

The parameter of interest is $\beta$. It represents the spatial correlation between the distance to *pueblos de indios* and the population density of a block. If $\beta < 0$, then we observe agglomeration patterns around the *pueblos* since blocks closer to the colonial settlements are denser; if $\beta > 0$, we have deagglomeration patterns near the *pueblos de indios*; and lastly, if $\beta = 0$, then there are neither agglomeration nor deagglomeration patterns around the *pueblos de indios* within Mexican cities.

**The causal impact of *pueblo de indios* on human capital sorting.**

We exploit the quasi-random variation created by colonial laws regarding the size of the *pueblos de indios*. Such laws indicated that the *pueblos* must have a minimum land size of 500-meter radius to protect the settlements from land theft by the Spaniards. This boundary allows us to implement a spatial regression discontinuity design (RDD) to estimate the *causal impact* of *pueblos de indios* on within-city sorting patterns.

We restrict our analysis to blocks within 1 km from a *pueblo de indios* to guarantee that our treatment and control units are relatively similar. Blocks within 0.5 km of a *pueblo de indios* are the treatment group, while blocks in a radius located between 0.5 and 1 km from a *pueblo de indios* act as our control group.

We use the following local-linear econometric specification:

$$Y_b = \alpha_{MSA} + \beta \cdot x_b + \delta \cdot D_b + \eta \cdot (x_b \cdot D_b) + \theta \cdot \bar{y}_{PI} + \Gamma' X_b + \varepsilon_b \tag{2}$$

where $b$ represents the block, our unit of analysis; $MSA$ is the metropolitan statistical area where block $b$ is located; we matched each block $b$ to the closest *pueblo*, $PI$. In the regression, $Y_b$ represent the economic outcome of interest (e.g., block's average years of education). Our treatment is represented by $D_b$ an indicator variable that takes the value of one when a block is within a 500-meter buffer of pueblo $PI$, and zero otherwise. The relative distance to the 500-meter buffer, or running variable, is represented by $x_b$ and takes negative values if block $b$ is within the *pueblo*'s 500-meter buffer or catchment area, and positive otherwise if block $b$ is between 0.5 and 1 km away from the nearest *pueblo* $PI$.

Our parameter of interest is $\delta$, which represents the causal impact of *pueblos de indios* on the modern-day economic outcomes of a block. If $\beta \neq 0$, this would confirm that the Spanish colonial interventions in the urban space impact current residential outcomes. This would match the predictions of the model by Allen & Donaldson (2022), which established theoretical conditions under which historical shocks shape present-day outcomes. If $\beta = 0$, then the *pueblos de indios* do not affect contemporary economic outcomes. This might be because of the dissipation of the effects of the colonial policy, the fact that geography is the sole determinant of within-city outcomes in Mexican cities, or the presence of heterogeneous treatment effects.

We include MSA fixed effects $\alpha_{msa}$ to account for differences in economic development across Mexican metropolitan statistical areas. Nonetheless, given that we are exploiting multiple boundaries, it is possible that our estimate simply reflects sorting within MSAs or differences between *pueblos de indios*. Instead of *pueblos de indios*' fixed effects, we include the average of the dependent variable of all units matched to a given *pueblo* as a control, $\bar{y}_{PI}$, in the spirit of Mundlak (1978). Lastly, we also control for a block's characteristics vector, $X_b$, which includes population density and geographic controls such as temperature, precipitation, altitude, slope, pollution, longitude, and latitude.

For estimation and inference, we follow Calonico et al. (2014). Additionally, we only consider 1st and 2nd-order polynomials in our estimation as Gelman & Imbens (2019) suggests. To account for multiple hypothesis testing, we consider a Bonferroni correction at a 10% significance level and use robust standard errors clustering by the nearest matched *pueblo*. In addition, for estimation we use a donut RD approach as in Barreca et al. (2016) and Laudares & Caicedo (2023) to account for the fuzzy boundaries of the *pueblos de indios*' catchment areas and the potential sorting of blocks into the buffer thresholds due to historical urban design factors. This means that in practice, we drop blocks located within 25 meters of the buffers. Panel (a) of Figure B.1 in the Appendix displays blocks' density around the threshold of 0.5 km around the *pueblos de indios* using (Cattaneo et al., 2018). We observe a rise in the blocks' density near the vicinity of the threshold. Panel (b) of Figure B.1 shows that after dropping the blocks in the vicinity of the boundary there is no discontinuity in the blocks' density.

Figure 2 displays the identification strategy using a subgroup of Mexico City blocks. All the blocks in the red or blue circles are in the treatment group, since they are located within the 0.5 km catchment area of a *pueblo de indios*. The blocks in the orange areas act as a control group, since they are located between 0.5 and 1 km from the nearest *pueblos de indios*. Any block outside these areas is dropped from the sample. Note that Figure 2 also shows that within Mexico City, the catchment areas of *pueblos* do not intersect. This suggests that the land use policy defined by colonial laws was likely enforced *de facto*.

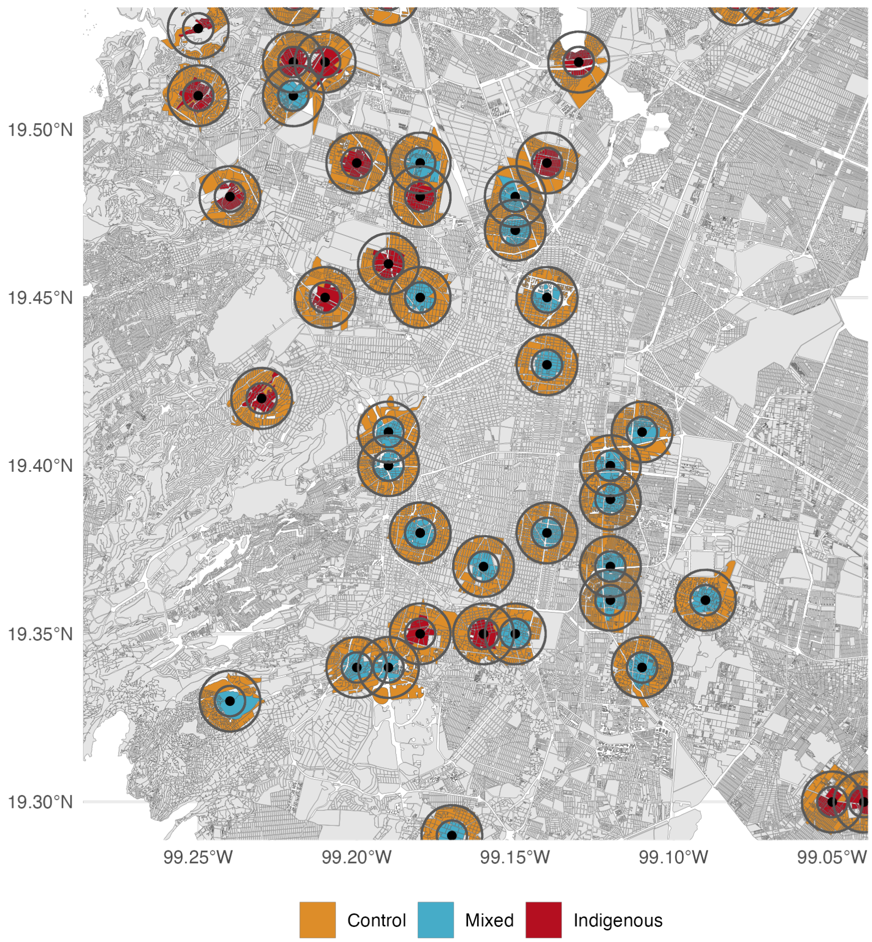



Figure 2: Identification strategy illustration: Spatial RDD
*Notes:* This figure displays a section of the Mexico City MSA showing a representation of the treatment group (red and blue) and control group (orange).

First, we estimate the local average treatment effect of all *pueblos de indios* on the blocks' economic outcomes. For our heterogeneity analysis, we estimate the model in equation (2) separately for two different sub-samples depending on the degree of success of the colonial segregation policy, as reported by Estrada et al. (2005). The first sample are those blocks near *pueblos* that only had Indigenous inhabitants by the end of the colonial period (red-orange circles in Figure 2). The second sample includes the blocks close to the *pueblos* that had heterogeneous ancestry (blue-orange circles in Figure 2).

We find balance on locational fundamentals and socioeconomic variables for control and treatment groups in both types of *pueblos*. Table B.2 shows the balance tests. The t-tests suggest balance in geographic characteristics in both cases, with a couple of exceptions (temperature and altitude in the case of *Mixed Pueblos*). The t-tests show there is balance in almost all socio-demographic characteristics. There is one exception for both types of *pueblos*: share of young population. For *Indigenous Pueblos*, there is no balance for number of children, children born alive, and share of Indigenous inhabitants. We control for these characteristics in our estimation.

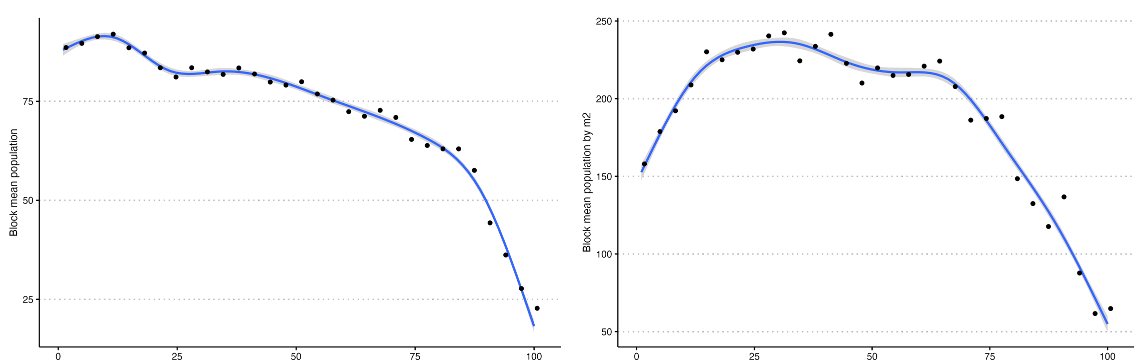
# 5 Results

We present two set of results. First, we present our OLS estimates showing the relationship between the location of *pueblos de indios* and spatial population patterns within metropolitan areas. Second, we provide RD estimates that measure the causal impact of *pueblos* on residential sorting by human capital. Initially, we show the estimates of the treatment effects of <u>all</u> *pueblos de indios* on blocks' human capital sorting. Afterward, we show how the estimates vary across different types of *pueblos* depending on the success of the segregation policy.

## 5.1 Pueblos and population density

First, we display descriptive data of blocks' population and population density around *pueblos* in Figure 3. The descriptive Figure shows two patterns. Blocks closer to a *pueblo de indios* have higher population (left panel in Figure 3), but it seems that city blocks near the *pueblos* are also larger in terms of area. Thus, we observe a quadratic relationship between distance to the *pueblos* and population density (right panel in Figure 3). Nevertheless, these descriptive patterns do not consider relevant factors such as market access, MSA characteristics, or geography. Since these elements can influence how the location of the *pueblos de indios* correlate with the spatial distribution of population, we use the model in equation (1) to control for these factors.

Figure 3: *Population distribution around pueblos de indios*




**Notes**: The graphs display the distribution of the population (left graph) and population density (right graph) of blocks around *pueblos de indios*. Each block is matched to the closest *pueblo*. Then we classify blocks in percentiles according to their distance to the nearest *pueblo*.

We observe the presence of non-linear deagglomerations around the *pueblos de indios*. To demonstrate this, we split our sample of blocks in vingtiles, according to their distance to the nearest *pueblo*. Then we estimate the model in equation (1) non-parametrically for each vingtile. We present our results in Figure 4. Moving from the first to the second vingtile is associated with an increase in population density of 6.6%. The seventh vingtile has 26% more population density than the closest vigntile to the *pueblo*. Interestingly, the deagglomeration forces are stronger among the blocks in the 10th to 19th vingtiles, relative to the blocks in the 1st to 7th vingtiles. The deagglomeration forces lose strength for the 18th to 20th vingtile.

Table C.3 in the Appendix displays the parametric estimates of the correlation between blocks' population density and the distance to the nearest *pueblos de indios* as we show in equation (1). Our spatial correlations show that an increase of one percent in the block's distance to the nearest *pueblo de indios* is associated with a reduction in block's population density by 0.107%, conditional on city characteristics, market access, and geography. These estimates are statistically significant. Our results are robust to eliminating spatially unobserved heterogeneity using Spatial First Differences (Druckenmiller & Hsiang, 2018), as shown in the last three columns of Table C.3 in the Appendix. Hence, once we account for market access and geography, the *pueblos de indios* generate spatial deagglomerations. Nonetheless, for blocks that are within a municipality encompassing the MSA but are rural, *pueblos* generate second-order spatial agglomerations.

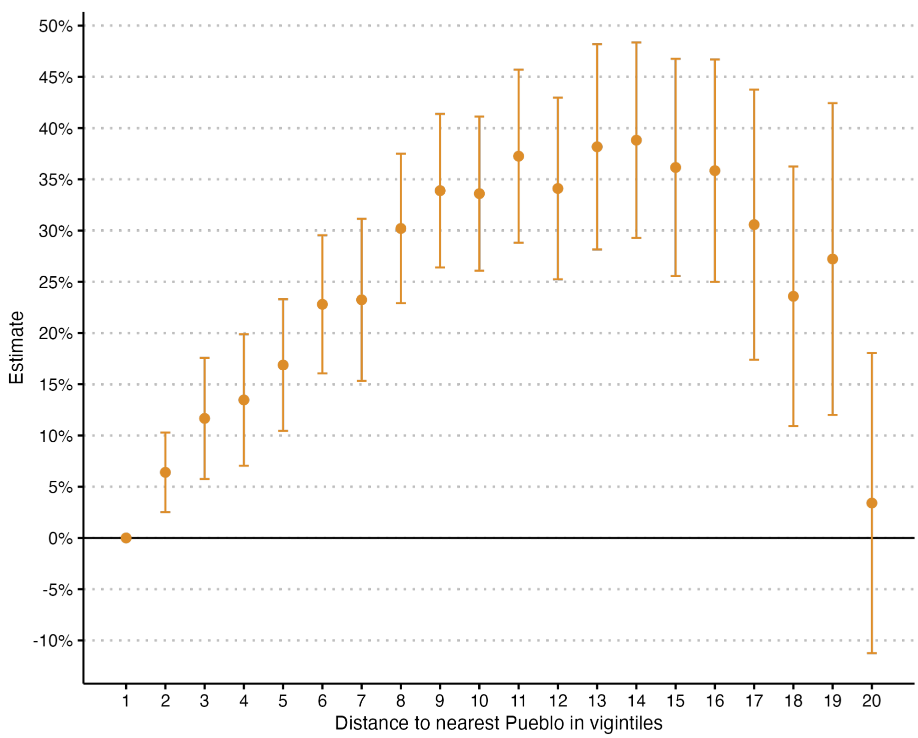


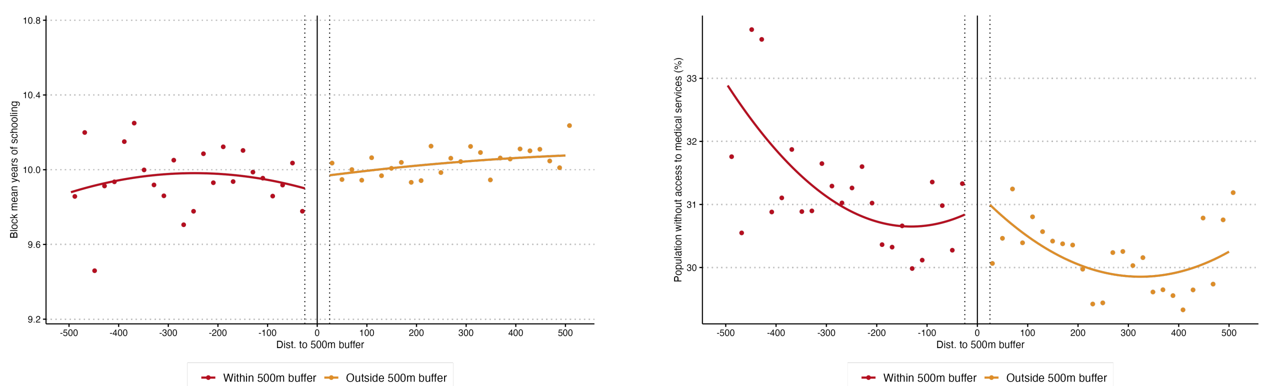

Figure 4: Pueblos' non-parametric agglomeration patterns

The estimates are expressed in semi-elasticities. Our unit of analysis is the block. Clustered standard errors in parentheses. We use clustering at the nearest *pueblo*. Controls include MSA fixed effects, urbanized block dummy, and distance to central business district. Geographic controls include max and min temperature, precipitation, altitude, slope, pollution (PM2,5), longitude, and latitude by block.

## 5.2 Average treatment effects of *pueblos* on residential sorting by human capital

In this section, we present the results of the impacts of *pueblos de indios* on the sorting patterns by human capital in Mexican cities. We consider two human capital outcomes: the blocks' average years of education and the blocks' share of households without access to medical services.

Figure 5 shows the spatial RDD plots for years of schooling and the population share without access to health services. The graphical evidence suggests a lack of discontinuity for the outcomes of interest. Point estimates are in Table C.4. Consistent with the graphical evidence, the RDD estimates of the causal impact of <u>all</u> *pueblos de indios* on the block's average years of education are not statistically significant and are not robust to different specifications. Nonetheless, point estimates show that the historical presence of *pueblos de indios* increases the share of households without access to medical services, as Table C.4 shows. Specifically, the *pueblos de indios* lead to a rise in this share between 0.12 and 0.21 standard deviations, which is equivalent to 1.8 to 3.2 percentage points. The estimates remain positive, but their significance varies across specifications.




(a) Years of schooling

(b) Population without access to health services

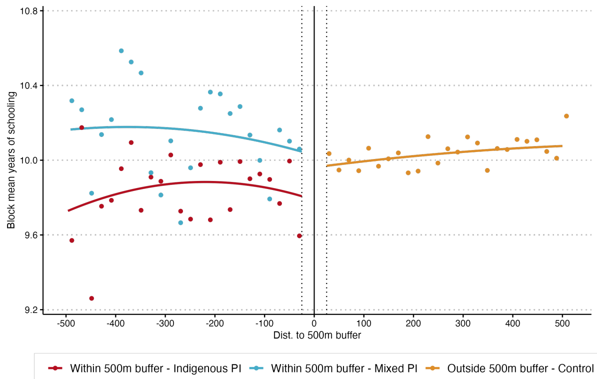
Figure 5: Spatial RDD

*Notes:* Figure (a) displays the discontinuity in blocks' average years of education. Figure (b) displays the discontinuity in blocks' population share without access to health services. The sample includes blocks near <u>all</u> *pueblos de indios*.

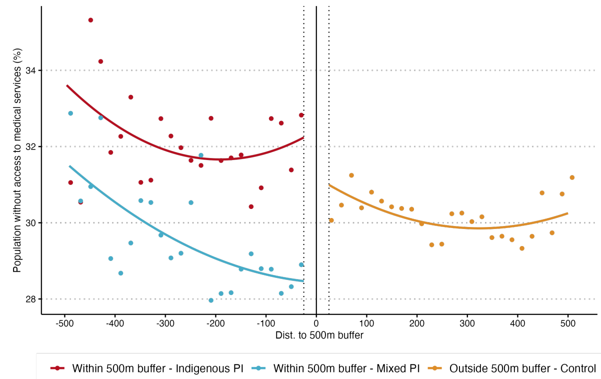
The effects are not robust when we consider the average treatment effect of <u>all</u> *pueblos de indios* on the blocks' human capital outcomes. Instead, these estimates might mask the heterogeneity of the segregation policy's success; that is, the effects of *pueblos* might vary between *Indigenous* and *Mixed Pueblos*. We proceed to do a heterogeneity analysis in the next sub-section.

## 5.3 Heterogeneous treatment effects of *pueblos de indios* by type (*Indigenous* vs *Mixed Pueblos*)

As we explain in the historical context, the spatial segregation policy of *pueblos de indios* intended to isolate Indigenous individuals during the colonial period, but the policy was not always successful. In some cases, the *pueblos* remained completely segregated until the end of the colonial era (*Indigenous Pueblos*), while in other cases, Spanish, Mestizos, Indios, and other caste individuals lived together in the *pueblos*. This led to multiethnic and multiracial marriages, thus creating *pueblos* with mixed ancestry (*Mixed Pueblos*). We use the classification created by Estrada et al. (2005) that labels each *pueblo* as Indigenous or Mixed to explore whether heterogeneous treatment effects explain the results in the previous section. As we discuss in detail in section 4, we split the sample according to the type of the nearest *pueblo*, and re-estimate the model in equation (2).




(a) Treatment vs. Control




(b) Treatment heterogeneity vs. Control

Figure 6: Spatial RDD: Heterogeneous effects by segregation success

*Notes:* Figure (a) displays a graphical discontinuity analysis of the blocks' share of households with no medical services access for the sample that includes blocks near <u>all</u> *pueblos de indios*. Figure (b) displays two discontinuities when we separate the sample into two parts, depending on the type of *pueblo*. We display the discontinuity of blocks' share of households with no medical service access for blocks near *Mixed Pueblos*, and the discontinuity corresponding to blocks adjacent to *Indigenous Pueblos*. We display the linear specification without geographic controls in both figures.

Figure 6 shows how treatment heterogeneity drives the aforementioned inconclusive results. The *Indigenous Pueblos* reduce the blocks' average years of schooling, whereas the *Mixed Pueblos* have a small positive impact. Consistently, *Indigenous Pueblos* lead to a lower population share with access to health services at the block level relative to the control, whereas blocks in the catchment area of *Mixed Pueblos* have more access to medical services relative to the control group. Therefore, by differentiating according to the type of *pueblo*, we demonstrate the colonial policy's lasting effects on human capital sorting.

Our estimates for the effects of different types of *pueblos* on the blocks' average education level are in Table 1. Crossing into the catchment area of *Indigenous Pueblos* reduces the blocks' average years of schooling between 0.16 and 0.19 standard deviations. The estimates are robust to different polynomials and the inclusion of geographical and socio-demographic covariates. The effect represents an increase of 0.38 and 0.44 years of schooling. On the other hand, while there is a positive estimate, there is no statistically significant effect of *Mixed Pueblos* on blocks' years of schooling.

We display the estimates from our heterogeneity analysis for the share of lack of access to medical services in Table 2. We find that crossing into the catchment area of *Indigenous Pueblos* increases the share of households with no access to medical services across blocks by 0.143 to 0.276 standard deviations, equivalent to 2.2 to 4.2 percentage points. The results are robust to different specifications, and the estimates in all models are statistically significant. In contrast, the RD estimates for *Mixed Pueblos*

Table 1: RDD: Years of schooling (z-score)

| Estimate | Lower CI | Upper CI | p-value | Covariates | Poly. | Bw. (m) | No. Obs. |
|---|---|---|---|---|---|---|---|
| Panel A. *Indigenous Pueblo* vs. Control | | | | | | | |
| -0.171 | -0.333 | -0.009 | 0.038 | No | 1 | 303.910 | 31266 |
| -0.194 | -0.389 | 0.000 | 0.050 | No | 2 | 378.375 | 39035 |
| -0.159 | -0.315 | -0.004 | 0.045 | Yes | 1 | 267.272 | 27297 |
| -0.187 | -0.366 | -0.009 | 0.040 | Yes | 2 | 366.379 | 37812 |
| Panel B. *Mixed Pueblo* vs. Control | | | | | | | |
| 0.167 | -0.035 | 0.370 | 0.105 | No | 1 | 313.278 | 28182 |
| 0.228 | -0.037 | 0.492 | 0.092 | No | 2 | 358.821 | 32575 |
| 0.117 | -0.041 | 0.275 | 0.147 | Yes | 1 | 323.757 | 29179 |
| 0.114 | -0.102 | 0.329 | 0.300 | Yes | 2 | 383.561 | 34955 |

*Notes:* We display estimates of the donut RD for two samples. The first sample includes the blocks matched to *Indigenous Pueblos*, and the second part includes the blocks matched to *Mixed Pueblos*. Then we obtain the donut RD estimates for each sample. For estimation and inference we follow Calonico et al. (2014). Following the recommendation of Gelman & Imbens (2019), we only consider 1st and 2nd order polynomials in our estimation. The confidence intervals consider multiple hypothesis testing using Bonferroni correction at a 10% significance level and use robust VCE clustering by *pueblo*. .The mean of the blocks' average years of schooling is 10.8 years, and the standard deviation is 2.2 years. The mean of the blocks' share of households with no access to medical care is 28.1% and the standard deviation is 15.1%. Poly. indicates the order of the polynomials, Bw. indicates bandwidth measured in meters.

provide inconclusive results. The RD estimates without controls are negative but switch signs once controlling for geography and socio-demographics. Lastly, the estimates are not statistically significant in any specification.

**Heterogeneity by pre-colonial population**

There is a possibility that the size of the pre-colonial population shapes our heterogeneity results. To explore this possibility, we show donut spatial RDD discontinuities for both types of *pueblos* according to the pre-colonial population in Figure C.2 in the Appendix. We find that for tertiles 1 and 3 of pre-colonial population, the direction of the estimates for each type of *pueblo* match the previous results i.e., *Indigenous Pueblos* negatively impact education and *Mixed Pueblos* have a positive effect on schooling. However, this is not true for tertile 2 of the *Mixed Pueblos*: the discontinuity indicates that this type of *pueblo* reduces schooling, contrary to the main results. Interestingly, we observe the presence of stronger discontinuities occur in the areas with higher pre-colonial population estimates.

For the second outcome, access to medical care, the discontinuities follow the same patterns as the main results across all tertiles of pre-colonial population. The *pueblos* with successful segregation increase the blocks' share of households with no access to medical services, while *pueblos* with mixed ancestry decrease this share. Interestingly, the stronger discontinuities occur in areas with lower pre-colonial population, in tertiles 1 and 2. In conclusion, pre-colonial population does not seem to affect substantially the direction

Table 2: RDD: Share without access to medical services (z-score)

| Estimate | Lower CI | Upper CI | p-value | Covariates | Poly. | Bw. (m) | No. Obs. |
|---|---|---|---|---|---|---|---|
| Panel A. *Indigenous Pueblo* vs. Control | | | | | | | |
| 0.212 | 0.040 | 0.383 | 0.015 | No | 1 | 295.968 | 32264 |
| 0.276 | 0.074 | 0.478 | 0.007 | No | 2 | 384.541 | 42068 |
| 0.143 | -0.006 | 0.293 | 0.060 | Yes | 1 | 292.312 | 31835 |
| 0.197 | 0.013 | 0.381 | 0.036 | Yes | 2 | 378.650 | 41464 |
| Panel B. *Mixed Pueblo* vs. Control | | | | | | | |
| -0.086 | -0.263 | 0.091 | 0.341 | No | 1 | 312.002 | 29946 |
| -0.051 | -0.265 | 0.162 | 0.636 | No | 2 | 395.616 | 38506 |
| 0.136 | -0.028 | 0.301 | 0.104 | Yes | 1 | 287.063 | 27309 |
| 0.127 | -0.064 | 0.318 | 0.193 | Yes | 2 | 384.770 | 37427 |

*Notes:* We display estimates of the donut RD for two samples. The first sample includes the blocks matched to *Indigenous Pueblos*, and the second part includes the blocks matched to *Mixed Pueblos*. Then we obtain the donut RD estimates for each sample. We follow Calonico et al. (2014) for estimation and inference. Following the recommendation of Gelman & Imbens (2019), we only consider 1st and 2nd order polynomials in our estimation. The confidence intervals consider multiple hypothesis testing using Bonferroni correction at a 10% significance level and use robust VCE clustering by *pueblo*. Poly. indicates the order of the polynomials, Bw. indicates bandwidth measured in meters.

of the discontinuities by type of *pueblo*. At most, they affect the size of the discontinuities.

### 5.4 Robustness and falsification checks

We implement two robustness and two falsification checks for the heterogeneous treatment effects. For our first robustness check, we obtain RD estimates with narrower bandwidths. Next, we implement cross-validation to verify that our results are not driven by a few *pueblos de indios*. In our placebo tests, we re-estimate our donut RD using a mock-up boundary, and we re-estimate the donut RD multiple times using random fake locations for the *pueblos de indios*.

**Robustness check 1. Narrower bandwidth**. We re-estimate our model using a Local Randomization RDD (Cattaneo et al., 2016) with a bandwidth of 50 meters around the cutoff for each one of the samples: blocks near *Mixed Pueblos*, and blocks near *Indigenous Pueblos*. We control for geographic and socio-demographic controls and MSA fixed effects. We display our results in Table C.5. We find similar patterns as in the previous subsection. Blocks within *Indigenous Pueblos*' catchment area have lower years of schooling and a higher population share of those without medical service access than those in the control group. In contrast, using the narrower bandwidth, blocks within *Mixed Pueblos* catchment area have higher average years of education and a lower share of households with no access to medical services relative to the control group. All these estimates are statistically significant using randomization inference.

**Robustness check 2. Are the results driven by a few *pueblos*?**. One potential concern is that our results might be driven by a small number of *pueblos de indios*. This could occur if, due to randomness or unobserved historical factors, a few *pueblos* experience strong discontinuities in education and lack of access to medical services, while there are no discontinuities in the modern-day blocks around the rest of the colonial settlements. To allay this concern, we implement a cross-validation exercise. We randomly drop 15% of the sample and re-estimate the model (2) 500 times. We report the means of the 500 estimates and confidence intervals with a 5% significance level in Table C.6. We compare them to the estimates in our preferred specification, the linear model with geographic controls (third row of every panel of tables 1 and 2). The magnitude and sign of the average estimate in this robustness check are similar to our original estimates. Furthermore, the estimates have the same statistical significance as our main results. Thus, our results are not driven by a few specific *pueblos*, but rather reflect the average treatment effect of the former on within-city sorting patterns.

**Falsification check 1. Mock-up cutoffs.** Is the assigned cutoff the relevant one? Would we find effects if we use bigger radius? In this falsification test, we modify the cutoffs, create mock-up boundaries of the *pueblos de indios* and re-estimate our donut RD for both types of pueblos. Without altering the *pueblos de indios*' location, we consider blocks between 1 and 2 km away from the nearest *pueblo de indios*. Second, we define a mock-up cutoff located 1.5 km away from the settlements. Thus, our treatment group contains blocks situated between 1 and 1.5 km away from the *pueblos*, while our control group consists of those blocks located between 1.5 km and 2 km away from the *pueblos*. Our results in Table C.7 show that the RD estimates using a mock-up boundary are not statistically significant. These results suggest that while we find effects using the cutoff of the 500-meter radius, as aligned with the historical policy determining *pueblos* land use, there are no effects with a larger mock-up radius.

**Falsification check 2. Mock-up random locations for *pueblos de indios***. One last concern is whether our results are driven by spatial noise. In this exercise, we randomly re-assign the location of the *pueblos de indios* to mock-up locations. Then, for each block within a MSA, we find the nearest mock-up *pueblo* and its respective distance. We create a new sample by exclusively keeping blocks between 0 and 1 km away from the randomly positioned mock-up *pueblos*. Then, we re-estimate our donut RD considering the linear specification with geographic controls. We repeat this process 1000 times and obtain the distribution of estimates. We show the results of the falsification test for the different types of *pueblos* and human capital outcomes in Figure C.3. Noticeably, we can only reject the null hypothesis that the impact of *Indigenous Pueblos* on blocks' years of education is due to random spatial noise, since the p-value of observing the same or a bigger estimate in absolute terms with *pueblos*'s random locations is 0.06. At conventional significance levels, we cannot reject that the other three estimates could be driven by random spatial noise, as their p-values indicate.

## 6 Evidence on mechanisms

In this section, we explore how colonial Indigenous segregation created long-term consequences for within-city sorting patterns. We consider two types of contemporary mechanisms: public good provision and urban amenities. Given the lack of comprehensive and harmonized data across Mexican MSAs, we focus on Mexico City metropolitan statistical area.

We explore two public good provision outcomes: teacher-student ratio as a measure of school quality and crime. We also look at two outcomes related to urban amenities: an urban amenities index, which is the first component of a principal component analysis using different measures on street quality, and the number of (formal) retail stores. Lastly, we analyze a proxy for hedonic pricing using land values data. We analyze these outcomes for both types of *pueblos*, Indigenous and Mixed.

We summarize our findings in Figure 7.[14] Overall, we find that *Indigenous Pueblos* have worse street quality. The index of urban amenities decreases by 0.25 standard deviations. Land prices within *Indigenous Pueblos*' catchment areas are 39.3% lower than land parcels outside these areas. This land value penalty is consistent with worse urban amenities and the evidence on land value penalties on previously segregated geographies (Yamagishi & Sato, 2022; Woo-Mora, 2023). The causal impacts of *Indigenous Pueblos* on the other three outcomes (crime, school quality, and business amenities) are not statistically significant. Hence, the successfully segregated *pueblos* affect current residential sorting via street amenities. Land prices capture these effects.

The results for mechanisms for the *pueblos* with mixed ancestry are puzzling. First, we find that land prices within the catchment area of *Mixed Pueblos* are 12.6% higher, but we do not find statistically significant discontinuities in school quality, the number of retail stores, or the index of street quality. However, the catchment areas of these *pueblos* do experience a discontinuous increase in crime, which may be explained by burglars targeting wealthier city blocks.

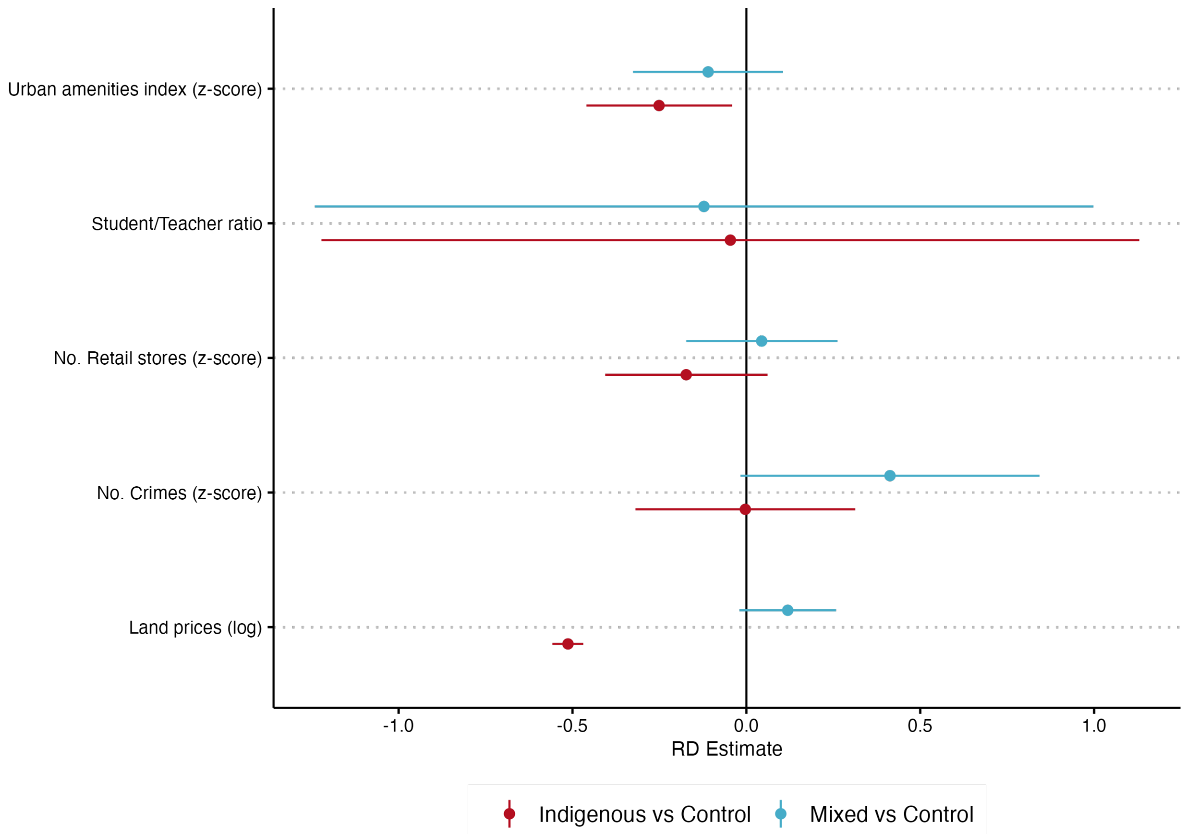
These results suggest that the long-lasting effects of *pueblos* on within-city sorting patterns might be primarily explained due to sorting into cheaper areas with worse urban amenities.

## 7 Conclusion

In this paper, we exploit the discontinuities of the boundaries of the *pueblos de indios*, settlements created during colonial rule to segregate the Indigenous population, to estimate the causal impact of colonial segregation policies in contemporary sorting by human capital in cities of a developing nation. Previous work has analyzed the effects of 19th and 20th-century segregation policies in American cities on residential sorting to show that historical governmental segregation policies have persistent effects in urban areas. To the best of our knowledge, existing work does not provide an empirical analysis of how

[14]We display point estimates in Table D.8.

Figure 7: *Mechanisms: Spatial RDD estimates*




**Notes**: The graphs display the donut spatial RDD estimates for five different outcomes related to urban amenities and public goods using data exclusively from Mexico City MSA. We split the sample in two: those schools/crime reports/land parcels/census blocks whose nearest pueblo is Indigenous, and those units that are closer to *Mixed Pueblo*. Then we proceed to obtain our donut spatial RDD estimates for each outcome and each type of pueblo as in equation (2). We report a linear specification that considers municipality fixed effects and the average level of the dependent variable for each *pueblo* in the spirit of Mundlak (1978). We report confidence intervals with a 5% significance level. We build the urban amenities index using characteristics of the front street of the block, as we explain in section 3.

colonial policies shape modern-day residential sorting in developing nations.

Using a Regression Discontinuity Design, we estimate the impacts of the *pueblos de indios* on residential sorting by human capital. We provide evidence that *pueblos de indios* do not impact blocks' average years of education, but they do affect blocks' share of households with lack of access to medical services. We find that the presence of heterogeneous treatment effects influences these results. To unmask the heterogeneity, we separate the sample in two according to the success of the segregation policy. We find that *pueblos* that were fully segregated reduce the blocks' contemporary measures of human capital. For *pueblos* where the colonial segregation policy failed, we do not have sufficient evidence that they had a persistent impact on present-day residential sorting by human capital. Rather than directly affecting human capital, our results suggest that the mechanisms are sorting due to amenities and land prices: there are worse urban amenities in the catchment areas of successfully segregated *pueblos* and lower land prices.

Our work faces two important limitations. First, we do not investigate how the colonial segregation policy shaped present-day outcomes between the end of the colonial era and the 20th century. While we are fully aware of the *compression of history* (Abad & Maurer, 2021), there exists a trade-off between historical specificity and external validity when we study the persistence of historical shocks across vast territories. Second, we cannot account for the political economy or collective action mechanisms that could also explain why successfully segregated *pueblos* have worse amenities and lower land values. Future work should explicitly include the historical specificity of different segregation patterns, exploit more historical data sources, and dig into the potential local political economy mechanisms explaining contemporary urban sorting.

Lastly, our work has implications for urban policies in developing nations. The design of place-based policies that improve distressed segregated neighborhoods requires us to understand the persistent impacts of colonial segregation policies and their mechanisms. In this sense, understanding that colonialism shapes modern-day residential sorting by human capital can help policymakers better target programs that aim to raise human capital in urban areas, and improve the design of taxes and subsidies that lead to the economic recovery of neighborhoods.

## References


Abad, L. A. & Maurer, N. (2021). History Never Really Says Goodbye: A Critical Review of the Persistence Literature. *Journal of Historical Political Economy*, 1(1), 31–68.

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2015). The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, 83(6), 2127–2189.

Alcántara, A. (2004). Los barrios de tenochtitlan. topografía, organización interna y tipología de sus predios. *Historia de la vida cotidiana en México*, 1, 167–198.

Alesina, A. & Ferrara, E. L. (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature*, 43(3), 762–800.

Alix-Garcia, J. & Sellars, E. A. (2020). Locational fundamentals, trade, and the changing urban landscape of Mexico. *Journal of Urban Economics*, 116.

Allen, T. & Donaldson, D. (2020). Persistence and Path Dependence in the Spatial Economy. *National Bureau of Economic Research Working Paper Series*, No. 28059.

Allen, T. & Donaldson, D. (2022). Persistence and path dependence: A primer. *Regional Science and Urban Economics*, 94, 103724.

Altamirano, M. & Trejo, G. (2016). The mexican color hierarchy: How race and skin tone still define chances 200 years after independence. In J. Hooker & J. T.A. (Eds.), *The Double Blind: The Politics of Racial and Class Inequalities in the Americas* (pp. 1–14). APSA.

Ambrus, A., Field, E., & Gonzalez, R. (2020). Loss in the time of cholera: Long-run impact of a disease epidemic on the urban landscape. *American Economic Review*, 110(2), 475–525.

Ananat, E. O. (2011). The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality. *American Economic Journal: Applied Economics*, 3(2), 34–66.

Anderson, R. D. (1988). Race and Social Stratification: A Comparison of Working-Class Spaniards, Indians, and Castas in Guadalajara, Mexico in 1821. *The Hispanic American Historical Review*, 68(2), 209–243.

Arias, L. M. & Girod, D. M. (2014). Indigenous Origins of Colonial Institutions. *Quarterly Journal of Political Science*, 9(3), 371–406.

Arteaga, F. (2018). The Historical Legacy of (Pre?)Colonial Indigenous Settlements in Mexico. *SSRN Electronic Journal*.

Artiles, M. (2022). Within-group heterogeneity in a multi-ethnic society. *MPRA Paper*.

Barreca, A. I., Guldi, M., Lindo, J. M., & Waddell, G. R. (2011). Saving babies? Revisiting the effect of very low birth weight classification. *Quarterly Journal of Economics*, 126(4), 2117–2123.

Barreca, A. I., Lindo, J. M., & Waddell, G. R. (2016). Heaping-induced bias in regression-discontinuity designs. *Economic Inquiry*, 54(1), 268–293.

Baum-Snow, N. & Lutz, B. F. (2011). School Desegregation, School Choice, and Changes in Residential Location Patterns by Race. *American Economic Review*, 101(7), 3019–46.

Bazant, J. (1966). La desamortización de los bienes corporativos en 1856. *Historia Mexicana*, 16(2), 193–212.

Bharathi, N., Malghan, D. V., & Rahman, A. (2019). Village in the City: Residential Segregation in Urbanizing India. *SSRN Electronic Journal*.

Bleakley, H. & Lin, J. (2012). Portage and Path Dependence. *The Quarterly Journal of Economics*, 127(2), 587–644.

Bobonis, G. J., Chaparro, J. C., Gonzalez-Navarro, M., & Rubio-Codina, M. (2021). Appropriate Institutions? Traditional Governance and Public Goods Provision in Oaxaca, Mexico. *AEA Papers and Proceedings*, 111, 243–247.

Boix, I. (1841). *Recopilación de Leyes de los Reynos de las Indias, mandadas imprimir y publicar por... Carlos II...: Va dividida en cuatro tomos, con el indice general, y al principio de cada tomo el indice especial de los títulos que contiene*, volume 3. Boix.

Boustan, L. P. (2010). Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration. *The Quarterly Journal of Economics*, 125(1), 417–443.

Boustan, L. P. & Margo, R. A. (2013). A silver lining to white flight? White suburbanization and African–American homeownership, 1940–1980. *Journal of Urban Economics*, 78, 71–80.

Brinkman, J. & Lin, J. (2022). Freeway Revolts! The Quality of Life Effects of Highways. *The Review of Economics and Statistics*, (pp. 1–45).

Brooks, L. & Lutz, B. (2019). Vestiges of Transit: Urban Persistence at a Microscale. *The Review of Economics and Statistics*, 101(3), 385–399.

Brooks, L., Rose, J., Shoag, D., & Veuger, S. (2020). The Long-Run Impact of the 1968 Washington, DC Civil Disturbance. *SSRN Electronic Journal*.

Calnek, E. E. (1976). The internal structure of Tenochtitlan BT - The valley of mexico : studies in pre-hispanic ecology and society. Advanced seminar series (pp. 287–325). Albuquerque: University of New Mexico Press.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6), 2295–2326.

Castro Gutiérrez, F. (2010a). Los indios y la ciudad. Panorama y perspectivas de investigación. In F. Castro-Gutiérrez (Ed.), *Los indios y las ciudades de Nueva España* (pp. 9–34). Ciudad de México: Universidad Nacional Autónoma de México.

Castro Gutiérrez, F. (2010b). *Los indios y las ciudades de Nueva España*. Universidad Nacional Autónoma de México, Instituto de Investigaciones Históricas.

Castro Gutiérrez, F. (2013). El origen y conformación de los barrios de indios. *Los Indios y las Ciudades de Nueva España*, (pp. 105–122).

Cattaneo, M. D., Arbor, A., Jansson, M., & Ma, X. (2018). *Manipulation testing based on density discontinuity*. Technical Report 1.

Cattaneo, M. D., Arbor, A., Titiunik, R., & Vazquez-Bare, G. (2016). Inference in regression discontinuity designs under local randomization. *The Stata Journal*, 16(2), 331–367.

Chiovelli, G., Mart, L. R., Torres, J. D., & Caicedo, F. V. (2022). Bourbon Reforms and State Capacity in the Spanish Empire. *Working Paper*.

Christopher, A. J. (2001). Urban Segregation in Post-apartheid South Africa. *Urban Studies*, 38(3), 449–466.

Colín, E. (2014). *Indigenous education through dance and ceremony: A Mexica palimpsest*. Springer.

Collins, W. J. & Margo, R. A. (2007). The Economic Aftermath of the 1960s Riots in American Cities: Evidence from Property Values. *The Journal of Economic History*, 67(4), 849–883.

Cook, L. D., Logan, T. D., & Parman, J. M. (2018). Rural Segregation and Racial Violence: Historical Effects of Spatial Racism. *American Journal of Economics and Sociology*, 77(3-4), 821–847.

De Estrada, D. T. (1999). *Pueblos de indios y educación en el México colonial, 1750 - 1821*. El Colegio de Mexico, 1 edition.

De la Roca, J., Ellen, I. G., & O'Regan, K. M. (2014). Race and neighborhoods in the 21st century: What does segregation mean today? *Regional Science and Urban Economics*, 47(1), 138–151.

Dell, M. (2010). The Persistent Effects of Peru's Mining Mita. *Econometrica*, 78(6), 1863–1903.

Diaz-Cayeros, A. (2011). Indian Identity, Poverty and Colonial Development in Mexico. In *Impact of Colonial and Post-Colonial Independence Institutions on Economic Development in Latin America Conference* Columbia University.

Diaz-Cayeros, A., Espinosa-Balbuena, J., & Jha, S. (2022). Pandemic Spikes and Broken Spears: Indigenous Resilience after the Conquest of Mexico. *Journal of Historical Political Economy*, 2(1), 89–133.

Druckenmiller, H. & Hsiang, S. (2018). Accounting for Unobservable Heterogeneity in Cross Section Using Spatial First Differences. *NBER Working Paper Series.*

Elizalde, A. (2020). On the economic effects of Indigenous institutions: Evidence from Mexico. *Journal of Development Economics*, 147, 102530.

Ellingsen, S. (2021). Long-distance trade and long-term persistence. *University of Pompeu Fabra Working Paper.*

Estrada, D. T., Alvarez Lobato, J. A., & Miranda, J. L. (2005). Atlas ilustrado de pueblos de indios de la Nueva Espana, 1800. *Journal of Latin American Geography*, 4(2), 97–109.

Fraser, D. J. (1972). La política de desamortización en las comunidades indígenas, 1856-1872. *Historia Mexicana*, 21(4), 615–652.

García Redondo, J. M. (2019). El proceso cartográfico de la reforma parroquial del arzobispo Lorenzana: Propuestas planimétricas de José Antonio de Alzate y recreación ilustrada de la Ciudad de México (1767-1772). *Historia Mexicana*, 68(3), 1001–1073.

Garfias, F. & Sellars, E. A. (2022). When State Building Backfires: Elite Coordination and Popular Grievance in Rebellion. *American Journal of Political Science*, 66(4), 977–992.

Gelman, A., Hullman, J., & Kennedy, L. (2023). Causal quartets: Different ways to attain the same average treatment effect. *arXiv*, (pp. 1–12).

Gelman, A. & Imbens, G. (2019). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, 37(3), 447–456.

Gerhard, P. (1993). *A Guide to the Historical Geography of New Spain.* University of Oklahoma Press.

Granados, L. (2013). Pasaportes neoclásicos. Identidad y cobro de tributo indígena en la Ciudad de México Borbónica. In F. Castro-Gutiérrez (Ed.), *Los indios y las ciudades de Nueva España.* Ciudad de México: Universidad Nacional Autónoma de México.

Hanlon, W. W. & Heblich, S. (2022). History and urban economics. *Regional Science and Urban Economics*, 94, 103751.

Heblich, S., Trew, A., & Zylberberg, Y. (2021). East-side story: Historical pollution and persistent neighborhood sorting. *Journal of Political Economy*, 129(5), 1508–1552.

Kelly, M. (2020). Understanding Persistence. *CEPR Press Discussion Paper No. 15246.*

Klein Goldewijk, K., Beusen, A., Doelman, J., & Stehfest, E. (2017). Anthropogenic land use estimates for the Holocene–HYDE 3.2. *Earth System Science Data*, 9(2), 927–953.

Knight, A. (2002). *Mexico: The Colonial Era.* Alan Knight. Cambridge University Press.

Koyama, M. & Rubin, J. (2022). *How the world became rich: The historical origins of economic growth*. John Wiley & Sons.

Lara-Cisneros, G. (2013). Religiosidad indigena en contextos urbanos. In F. Castro-Gutiérrez (Ed.), *Los indios y las ciudades de Nueva España*. Ciudad de México: Universidad Nacional Autónoma de México.

Laudares, H. & Caicedo, F. V. (2023). Tordesillas, Slavery and the Origins of Brazilian Inequality. *Working Paper*.

Lin, J. & Rauch, F. (2022). What future for history dependence in spatial economics? *Regional Science and Urban Economics*, 94, 103628.

Lira, A. (1995). *Comunidades indígenas frente a la ciudad de México: Tenochtitlan, Tlatelolco, sus pueblos y barrios, 1812-1919*. El Colegio de México.

Logan, T. D. & Parman, J. M. (2017). Segregation and Homeownership in the Early Twentieth Century. *American Economic Review*, 107(5), 410–414.

Loreto López, R. (2010). Los artífices de una ciudad. Los indios y sus territorialidades. Puebla de los Ángeles, 1777. In F. Castro-Gutiérrez (Ed.), *Los indios y las ciudades de Nueva España* (pp. 255–277). Ciudad de México: Universidad Autónoma de México Mexico City.

MacNeish, R. S. (1964). Ancient Mesoamerican Civilization. *Science*, 143(3606), 531–537.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1), 69–85.

Nemser, D. (2017). *Infrastructures of Race: Concentration and Biopolitics in Colonial Mexico*. Border Hispanisms. University of Texas Press.

Nitsch, V. & Wolf, N. (2013). Tear down this wall: on the persistence of borders in trade. *Canadian Journal of Economics/Revue canadienne d'économique*, 46(1), 154–179.

Nunn, N. (2014). Historical development. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth*, volume 2 (pp. 347–402). North-Holland.

Nunn, N. (2020). The Historical Roots of Economic Development. *Science*, 367(6485), eaaz9986.

Pérez Pérez, J., Vial Lecaros, F., & Zárate, R. D. (2022). Urban Transit Infrastructure: Spatial Mismatch and Labor Market Power. *Working Paper*.

Redding, S. J., Sturm, D. M., & Wolf, N. (2011). History and Industry Location: Evidence from German Airports. *The Review of Economics and Statistics*, 93(3), 814–831.

Roca, J. D. l., Ellen, I. G., & Steil, J. (2018). Does segregation matter for Latinos? *Journal of Housing Economics*, 40, 129–141.

Rothstein, R. (2017). *The Color of Law: A Forgotten History of How Our Government Segregated America*. Liveright Publishing.

Sabatini, F. (2003). La segregación social del espacio en las ciudades de América Latina.

Sanders, W. T. (1956). The central mexican symbiotic region: a study in prehistoric settlemetn patterns. *Viking Fund Publication*.

Selod, H. & Zenou, Y. (2003). Private versus public schools in post-Apartheid South African cities: theory and policy implications. *Journal of Development Economics*, 71(2), 351–394.

Shertzer, A., Twinam, T., & Walsh, R. P. (2016). Race, Ethnicity, and Discriminatory Zoning. *American Economic Journal: Applied Economics*, 8(3), 217–246.

Shertzer, A., Twinam, T., & Walsh, R. P. (2022). Zoning and segregation in urban economic history. *Regional Science and Urban Economics*, 94, 103652.

Stangl, W. (2019). Data: Territorial gazetteer for Spanish America, 1701-1808 (V3). *Harvard Dataverse*.

Tenorio-Trillo, M. (2015). *I Speak of the City: Mexico City at the turn of the twentieth century*. The University of Chicago Press.

Tsivanidis, N. (2022). Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio. *UC Berkeley (mimeo)*.

Valencia Caicedo, F. (2019). The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America*. *The Quarterly Journal of Economics*, 134(1), 507–556.

Van Young, E. (1984). Conflict and Solidarity in Indian Village Life: The Guadalajara Region in the Late Colonial Period. *Hispanic American Historical Review*, 64(1), 55–79.

Velazquez-Cabrera, D. (2021). Urban Transit Infrastructure: Gender Gap and Commuting Times. *University of Michigan Working Paper*.

Vinson III, B. (2017). *Before Mestizaje: The Frontiers of Race and Caste in Colonial Mexico*. Cambridge: Cambridge University Press.

Waldinger, M. (2017). The long-run effects of missionary orders in Mexico. *Journal of Development Economics*, 127, 355–378.

Warnes, P. E. (2021). Transport infrastructure improvements and spatial sorting: Evidence from buenos aires. *Working Paper*.

Woo-Mora, G. L. (2023). On the other side of the creek: The evolution and persistence of colonial zoning stigma. *Working Paper*.

Yamagishi, A. & Sato, Y. (2022). Measuring Discrimination in Spatial Equilibrium: 100 Years of Japan's Invisible Race. *Crepe Discussion Paper no. 116*.

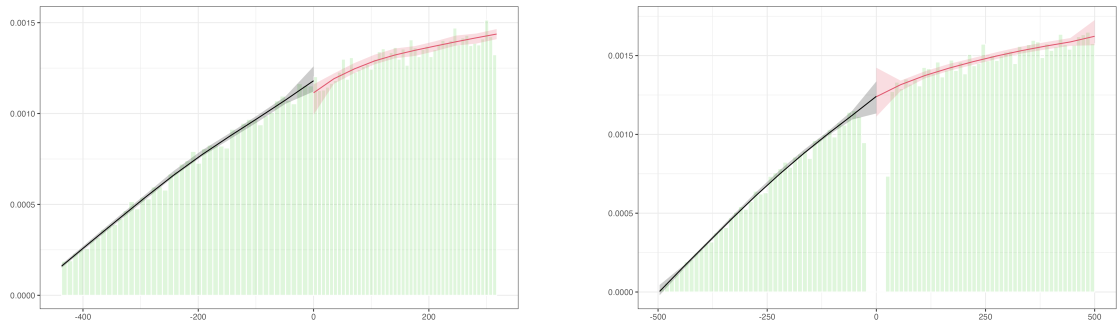
# A Data

Table A.1: 2020 Census Blocks: Summary statistics

| | SD | Minimum | Percentile 25 | Median | Mean | Percentile 75 | Maximum |
|---|---|---|---|---|---|---|---|
| Area ($m^2$) | 4.636 | 0.001 | 0.25 | 0.469 | 1.213 | 0.89 | 701.096 |
| Perimeter (m) | 0.593 | 0.012 | 0.226 | 0.319 | 0.456 | 0.463 | 33.87 |
| Population | 112.734 | 1 | 16 | 52 | 78.307 | 102 | 10484 |
| Female share | 0.062 | 0 | 0.485 | 0.519 | 0.519 | 0.552 | 1 |
| Indigenous share | 0.105 | 0 | 0 | 0 | 0.04 | 0.032 | 1 |
| Afro-origin share | 0.083 | 0 | 0 | 0 | 0.024 | 0 | 1 |
| Born in other state share | 0.165 | 0 | 0.086 | 0.157 | 0.202 | 0.273 | 1 |
| Living in other state at 2015 share | 0.057 | 0 | 0 | 0 | 0.029 | 0.04 | 1 |
| Housing units | 38.812 | 0 | 12 | 22 | 30.782 | 37 | 3284 |
| Inhabited housing units share | 0.146 | 0.005 | 0.789 | 0.882 | 0.849 | 0.952 | 1 |
| Years of schooling | 2.243 | 0 | 9.19 | 10.5 | 10.78 | 12.3 | 19.6 |
| People per room | 0.311 | 0 | 0.72 | 0.9 | 0.935 | 1.1 | 12 |
| Hh. without access to medical services share | 0.151 | 0.000 | 0.176 | 0.263 | 0.281 | 0.366 | 1 |
| Dist. to CBD (m) | 6797.813 | 12.612 | 4426.701 | 7870.069 | 9777.522 | 14040.3 | 32232.798 |
| Dist. to cathedral (m) | 6978.765 | 3.12 | 4412.093 | 8124.675 | 10020.135 | 14514.237 | 28552.409 |
| Dist. to nearest *pueblo de indios* (m) | 3483.928 | 3.502 | 1619.966 | 3107.917 | 4135.005 | 5555.92 | 18519.597 |

**Notes**: Based on 2020 Mexican Census. We include all blocks that belong to Metropolitan Statistical Areas with at least one *pueblo de indios* (73% of all MSAs). CBD is an acronym for the central business district. The administrative boundary corresponds to the boundary of the MSA created by its municipalities. For the case of distance to cathedral, we only consider those temples built during colonial times.

# B Empirical strategy




(a) Spatial RDD

(b) 'Donut' Spatial RDD

Figure B.1: RDD: Density test

**Notes**: We display a density test using Cattaneo et al. (2018). We show the density of blocks around the threshold in Figure (a). The density plot indicates an increase in density to the left of the threshold, and a decline in the density to the right of the threshold. Therefore, our main specification uses the sample as in Figure (b), by dropping the observations at the vicinity of the threshold.

Table B.2: Balance on observables

| Variable | *Indigenous Pueblo* vs. Control | | *Mixed Pueblo* vs. Control | |
|---|---|---|---|---|
| | Estimate | p-value | Estimate | p-value |
| *Geography* | | | | |
| Temperature max (C°) | 0.014 | 0.565 | 0.057 | 0.041 |
| Temperature min (C°) | 0.023 | 0.549 | 0.079 | 0.064 |
| Precipitation (mm/day) | -1.619 | 0.249 | -1.225 | 0.327 |
| Altitude (m) | -2.181 | 0.742 | -14.767 | 0.052 |
| Slope | -0.076 | 0.714 | -0.110 | 0.501 |
| Pollution (PM2.5) | 0.014 | 0.855 | 0.059 | 0.326 |
| *Socio-demographics* | | | | |
| Population by m2 (ln) | -0.028 | 0.813 | 0.026 | 0.852 |
| Ratio men/women (%) | 5.296 | 0.160 | 3.046 | 0.319 |
| 0-14 years old (%) | 2.160 | 0.067 | -1.874 | 0.065 |
| 15-64 years old (%) | -1.745 | 0.115 | 0.631 | 0.456 |
| 65 years old or more (%) | -1.092 | 0.207 | 1.276 | 0.223 |
| Children born alive (%) | 0.171 | 0.084 | -0.024 | 0.755 |
| Born in other state (%) | -0.317 | 0.852 | -0.457 | 0.825 |
| Living in other state by 2015 (%) | 0.530 | 0.285 | -0.187 | 0.794 |
| Indigenous (%) | -2.301 | 0.052 | -0.721 | 0.767 |
| Afro-origin (%) | -0.780 | 0.338 | -0.336 | 0.809 |
| Unemployment (%) | 0.000 | 0.963 | -0.001 | 0.929 |
| Catholic (%) | -0.345 | 0.838 | 0.393 | 0.824 |
| Households with female head (%) | 1.060 | 0.531 | 3.298 | 0.106 |

**Notes**. This table displays the balance test for blocks in the treatment group vs. control for each type of *pueblos*. Blocks in the control group are within a 0.5 km radius of the location of the *pueblo*, while blocks in the treatment are between 0.5 km to 1 km away from the *pueblos*. We only consider blocks that are part of MSAs that contain at least one *pueblo* (73% of all MSAs).

# C Results

Table C.3: Agglomerations

| | Population by m2 (ln) | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | | | SFD | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Dist. nearest PI (ln) | -0.356 | -0.168 | 0.080 | 0.107 | 0.082 | -0.137 | 0.087 |
| | (0.043) | (0.034) | (0.020) | (0.018) | (0.014) | (0.068) | (0.014) |
| Dist. CBD (ln) | | -0.875 | -0.245 | -0.197 | -0.062 | -0.349 | -0.057 |
| | | (0.054) | (0.028) | (0.029) | (0.026) | (0.103) | (0.026) |
| Observations | 613,096 | 613,096 | 613,096 | 514,406 | 433,170 | 2,650 | 430,520 |
| $R^2$ | 0.211 | 0.296 | 0.577 | 0.211 | 0.039 | 0.137 | 0.040 |
| Within $R^2$ | 0.023 | 0.128 | 0.477 | 0.058 | 0.039 | 0.032 | 0.039 |
| Sample | Full | Full | Full | Full | Full | Rural | Urban |
| MSA fixed effects | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Urban block FE | | | ✓ | ✓ | ✓ | | |
| Geographic controls | | | | ✓ | ✓ | ✓ | ✓ |

*Notes:* Unit of analysis is the block. Clustered standard errors in parentheses, clustering at the nearest *pueblo*. Geographic controls include maximum and minimum temperature, precipitation, altitude, slope, pollution (PM2.5), longitude, and latitude by block. We only include blocks that are part of municipalities in Mexican metropolitan statistical areas. Some municipalities contain both urban and rural blocks within their boundaries.

Table C.4: RDD: Average treatment of all *pueblos* on education and medical services access

| RD Estimate | Lower CI | Upper CI | p-value | Covariates | Poly. | Bw. (m) | No. Obs. |
|---|---|---|---|---|---|---|---|
| Panel A. *Dependent variable:* Years of schooling (z-score) | | | | | | | |
| -0.048 | -0.193 | 0.096 | 0.513 | No | 1 | 314.301 | 37399 |
| -0.023 | -0.218 | 0.171 | 0.816 | No | 2 | 329.974 | 39193 |
| -0.043 | -0.164 | 0.078 | 0.489 | Yes | 1 | 300.638 | 35796 |
| -0.074 | -0.232 | 0.085 | 0.363 | Yes | 2 | 369.206 | 43603 |
| Panel B. *Dependent variable:* Share of households without access to medical services (z-score) | | | | | | | |
| 0.117 | -0.035 | 0.269 | 0.131 | No | 1 | 287.005 | 36368 |
| 0.149 | -0.026 | 0.323 | 0.094 | Yes | 2 | 381.052 | 47779 |
| 0.164 | 0.018 | 0.311 | 0.028 | No | 1 | 260.541 | 33006 |
| 0.205 | 0.040 | 0.371 | 0.015 | Yes | 2 | 362.789 | 45674 |

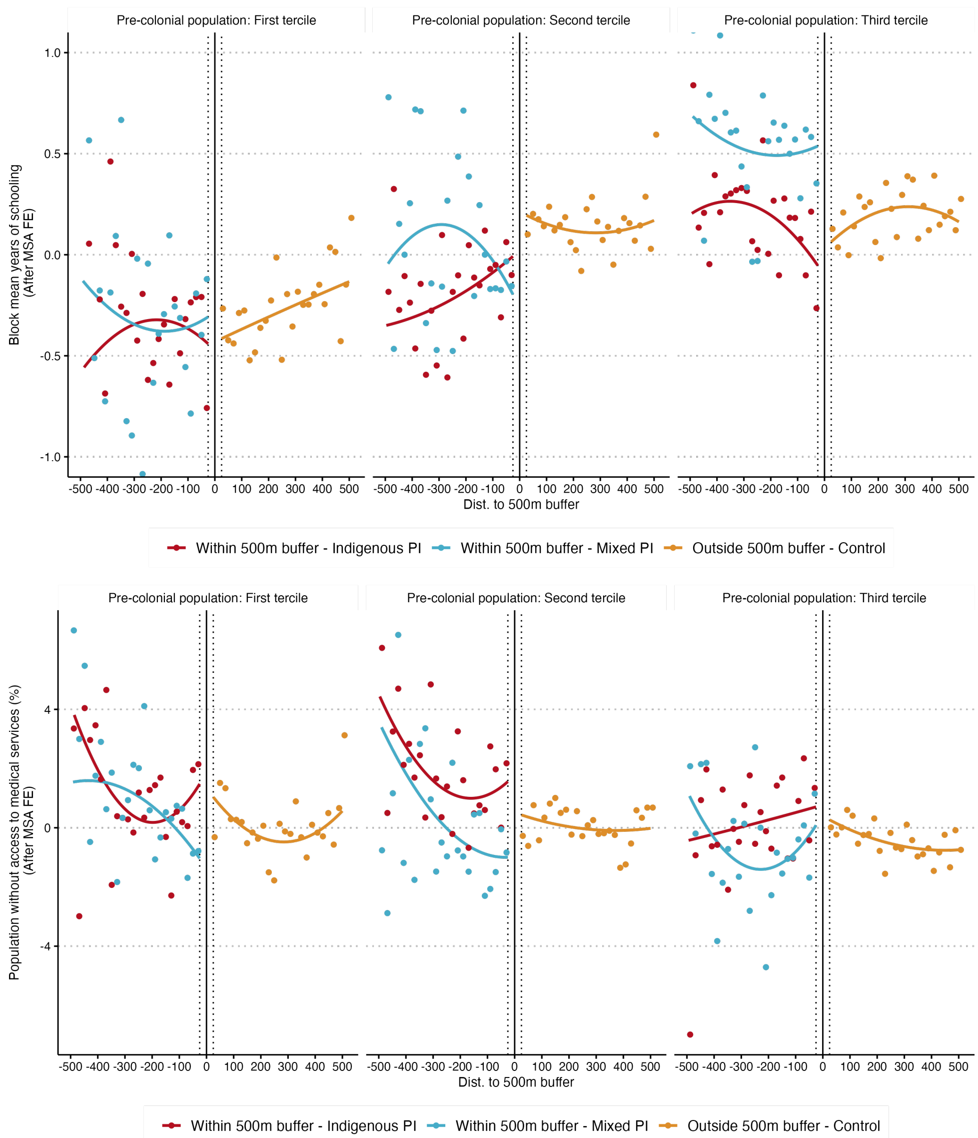
*Notes:* We display estimates of the donut RD. We follow Calonico et al. (2014) for estimation and inference. Following the recommendation of Gelman & Imbens (2019), we only consider 1st and 2nd order in our specifications. The confidence intervals consider multiple hypothesis testing using Bonferroni correction at a 10% significance level and use robust VCE clustering by *pueblo*. The mean of the blocks' average years of schooling is 10.8 years, and the standard deviation is 2.2 years. The mean of the blocks' share of households with no access to medical care is 28.1% and the standard deviation is 0.151.

Table C.5: Robustess: Local Randomization RDD - 50 meter window

| Variable | LR-RD Estimate | Lower CI | Upper CI | Asymptotic p-value |
|---|---|---|---|---|
| Panel A. *Indigenous Pueblo* vs. Control | | | | |
| Years of schooling (z-score) | -0.151 | -0.225 | -0.070 | 0.000 |
| No access to medical services (z-score) | 0.152 | 0.078 | 0.224 | 0.000 |
| Panel B. *Mixed Pueblo* vs. Control | | | | |
| Years of schooling (z-score) | 0.205 | 0.113 | 0.300 | 0.000 |
| No access to medical services (z-score) | -0.104 | -0.196 | -0.019 | 0.020 |

**Notes**: Specifications are local linear spatial RDD, without covariates. Blocks within a window of [-50,50] meters to the cutoff. We use the package *rdlocrand* for inference, which estimates robust standard errors accounting for SUTVA violation. The confidence intervals consider multiple hypothesis testing using Bonferroni correction at a 10% significance level.

Figure C.2: *Heterogeneity results by pre-colonial population estimates*




**Notes**: We show discontinuities using a donut spatial RDD. We use a linear specification with no covariates. We classify the block observations around *pueblos de indios* according to the population during pre-colonial times estimated by Klein Goldewijk et al. (2017). The first tertile contains the blocks around *pueblos* located in areas that had the lowest level of population during pre-colonial times, while the third tertile refers to the areas with the highest level of pre-colonial population.

Table C.6: Robustness test: Leave 15% of the sample out

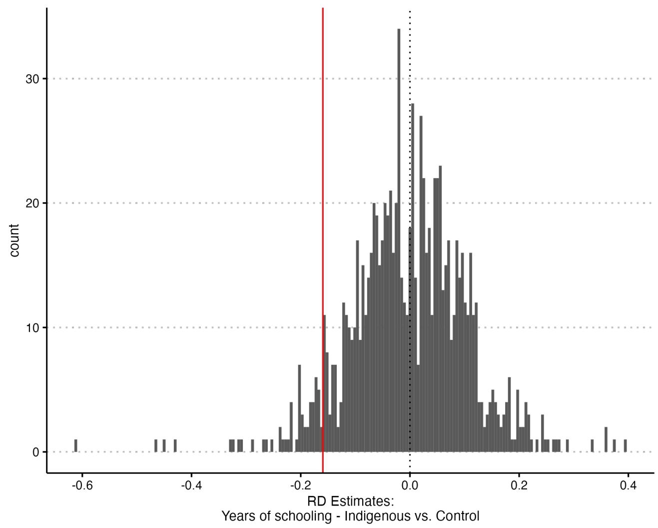
| Variable | Original | Drop 15% of the sample | | |
|---|---|---|---|---|
| | Estimate | Estimate | Lower CI | Upper CI |
| Panel A. *Indigenous Pueblo* vs. Control | | | | |
| Years of schooling (z-score) | -0.159 | -0.110 | -0.171 | -0.049 |
| No access to medical services (z-score) | 0.143 | 0.203 | 0.104 | 0.301 |
| Panel B. *Mixed Pueblo* vs. Control | | | | |
| Years of schooling (z-score) | 0.117 | 0.190 | 0.122 | 0.256 |
| No access to medical services (z-score) | 0.136 | 0.079 | -0.021 | 0.181 |

**Notes**: We report donut RD estimates for linear specifications with geographic covariates. We include the estimates from tables 1 and 2. For the estimates and confidence intervals that drop 15% of the sample, we take the mean and standard deviation of the estimates and confidence intervals after running our algorithm 500 times.

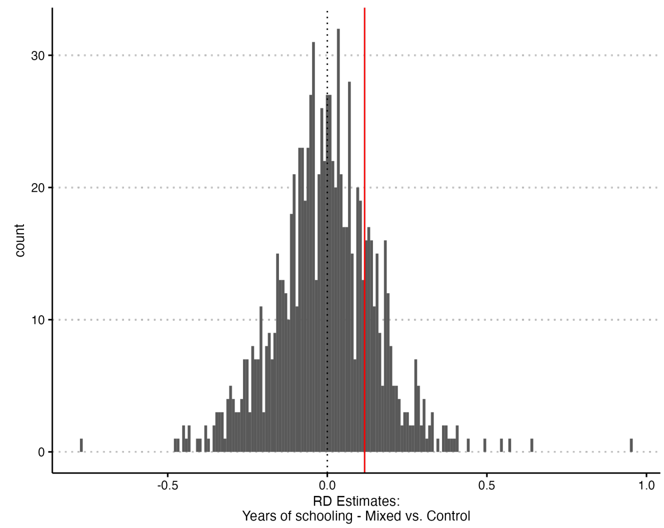
Table C.7: Falsification test: spatial RDD between 1 and 2 km from the nearest *pueblo*

| Variable | Estimate | Lower CI | Upper CI | p-value | Bw. (m) | No. Obs. |
|---|---|---|---|---|---|---|
| Panel A. *Indigenous Pueblo* vs. Control | | | | | | |
| Years of schooling (z-score) | -0.048 | -0.152 | 0.057 | 0.371 | 325.223 | 50003 |
| No access to medical services (z-score) | 0.035 | -0.058 | 0.129 | 0.457 | 319.309 | 49059 |
| Panel B. *Mixed Pueblo* vs. Control | | | | | | |
| Years of schooling (z-score) | -0.033 | -0.164 | 0.098 | 0.620 | 311.298 | 38179 |
| No access to medical services (z-score) | 0.000 | -0.098 | 0.097 | 0.993 | 346.746 | 42752 |

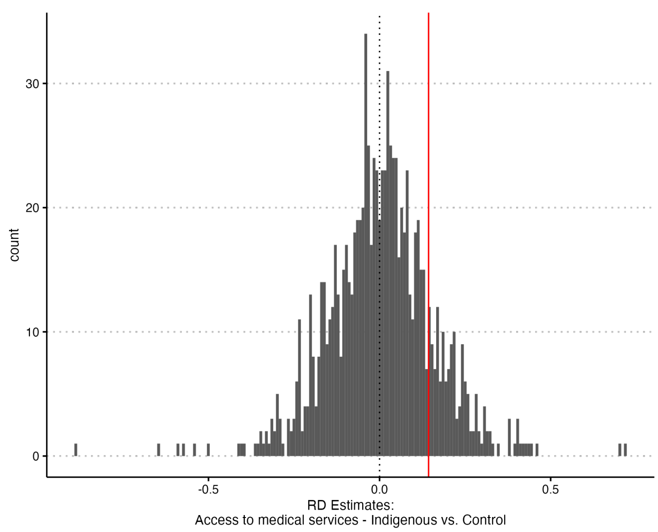
**Notes**: In this falsification test, our sample includes all blocks that are situated between 1 and 2 km away from the nearest *pueblo de indios*. The cutoff to treatment is 1.5 km. Thus, all blocks that are between 1 and 1.5 km away from the nearest *pueblos de indios* are classified as treated. All blocks located between 1.5 and 2 km away from the *pueblos* are part of the control group. Specifications are local linear spatial RDD, including covariates. We report donut RD estimates for linear specifications without covariates. We follow Calonico et al. (2014) for estimation and inference. In all confidence intervals, we use Bonferroni correction for multiple hypothesis testing at a 10% significance level and using robust VCE clustering by *pueblo*.



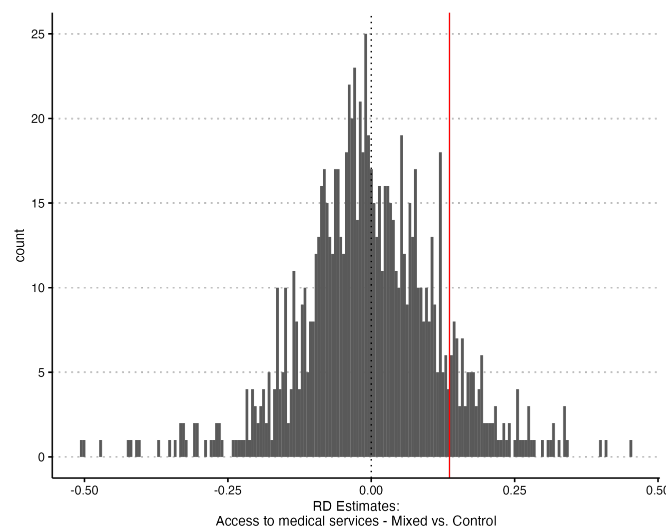

(a) *Indigenous Pueblos*, schooling




(b) *Mixed Pueblos*, schooling




(c) *Indigenous Pueblos*, access to medical services




(d) *Mixed Pueblos*, access to medical services

Figure C.3: Robustness check: Random mock-up locations of *pueblos de indios*

*Notes:* In this stringent falsification test, we randomize the location of the *pueblos de indios* and re-estimate our donut RDD with the same specification as our main results. We repeat this process 500 times and plot the donut RDD estimates. The red line in all figures displays the estimates of our linear specification with covariates from our heterogeneity results.

# D Mechanisms

Table D.8: Mechanisms

| Variable | Estimate | Lower CI | Upper CI | p-value | Bw. (m) | No. Obs. |
|---|---|---|---|---|---|---|
| Panel A. *Indigenous Pueblo* vs. Control | | | | | | |
| Land prices (log) | -0.513 | -0.566 | -0.460 | 0.000 | 198.653 | 65741 |
| No. Crimes (z-score) | -0.003 | -0.378 | 0.372 | 0.985 | 211.601 | 3846 |
| No. Retail stores (z-score) | -0.173 | -0.450 | 0.105 | 0.147 | 197.268 | 3559 |
| Student/Teacher ratio | -0.046 | -1.442 | 1.350 | 0.939 | 263.393 | 712 |
| Urban amenities index (z-score) | -0.251 | -0.500 | -0.002 | 0.019 | 152.806 | 15601 |
| Panel B. *Mixed Pueblo* vs. Control | | | | | | |
| Land prices (log) | 0.119 | -0.046 | 0.283 | 0.093 | 206.031 | 67452 |
| No. Crimes (z-score) | 0.413 | -0.098 | 0.923 | 0.060 | 241.304 | 4449 |
| No. Retail storess (z-score) | 0.044 | -0.214 | 0.303 | 0.690 | 271.299 | 5028 |
| Student/Teacher ratio | -0.122 | -1.450 | 1.207 | 0.831 | 246.006 | 699 |
| Urban amenities index (z-score) | -0.110 | -0.367 | 0.146 | 0.316 | 187.069 | 17904 |

**Notes**: Specifications are local linear donut spatial RDD as in Barreca et al. (2011). We include as controls municipality fixed effects and, instead of pueblos' fixed effects, we use the average value of the dependent variable in the spirit of Mundlak (1978). For estimation and inference, we follow Calonico et al. (2014). We account for multiple hypothesis testing in all confidence intervals using a Bonferroni correction considering a 10% significance level. The urban amenities index considers the characteristics of the street where the block is located as described in section 3.