Residential Segregation and Unequal Access to Local Public Services in India: Evidence from 1.5m Neighborhoods*

Sam Asher† ⓡ
Kritarth Jha‡ ⓡ
Paul Novosad§ ⓡ
Anjali Adukia¶ ⓡ
Brandon Tan‖ ⓡ

September 2025

Abstract

We study residential segregation and access to public services across 1.5 million urban and rural neighborhoods in India. Muslim and Scheduled Caste segregation in India is high by global standards, and only slightly lower than Black-White segregation in the U.S. Within cities, public facilities and infrastructure are systematically less available in Muslim and Scheduled Caste neighborhoods. Nearly all regressive allocation is across neighborhoods within cities—at the most informal and least studied form of government. These inequalities are not visible in the aggregate data typically used for research and policy.

JEL Codes: H41, J15, O15

*We are grateful for stellar research assistance from Ryu Matsuura, Alison Campion, Aneesha Mishra and Xinyu Ni, and for helpful comments from Gustavo Bobonis, Filipe Campante, Raj Chetty, John Friedman, Ingrid Gould Ellen, Ed Glaeser, Rema Hanna, Nina Harari, Rick Hornbeck, Larry Katz, Gaurav Khanna, Karthik Muralidharan, Sean Reardon, Mushfiq Mobarak, Rebecca White, and Maisy Wong, among many others. This project was in part supported by the IGC (project code IND-20165) and the Becker Friedman Institute at the University of Chicago. The ⓡ between author names signals that the name order was randomized using the AEA author randomization tool (TaxM_T0V9IFZ). When citing this paper, please follow the convention, such as using "S. Asher ⓡ K. Jha ⓡ P. Novosad ⓡ A. Adukia ⓡ B. Tan" in the references and/or using "Asher ⓡ al" instead of "et al" which signals the different meaning. More details can be found at https://www.aeaweb.org/journals/random-author-order.

†Imperial College London, sam.asher@imperial.ac.uk
‡Development Data Lab, jha@devdatalab.org
§Dartmouth College, paul.novosad@dartmouth.edu
¶University of Chicago, adukia@uchicago.edu
‖International Monetary Fund, btan@g.harvard.edu

1 Introduction

The concentration of marginalized social groups into poor neighborhoods is a key driver of persistent cross-group inequality in many contexts (Cutler et al., 2008; Ananat and Washington, 2009; Alesina and Zhuravskaya, 2011; Boustan, 2012; Chyn et al., 2022). Residential segregation can have a range of negative consequences: members of segregated groups may face worse discrimination in terms of provision of public services, they may have worse access to employment networks and labor market opportunities, and stereotypes in the wider population may be more difficult to break, among others (Massey and Denton, 1993; Cutler and Glaeser, 1997). Because homophily is a strong force in residential choices and residential settlement patterns tend to be highly persistent, these disadvantages can be particularly difficult to address.

Most of the empirical literature on residential segregation and neighborhood effects comes from developed countries, in large part due to the paucity of cross-neighborhood data in less developed countries. But the role of neighborhoods is particularly important to study in poorer countries. Cleavages across social groups are just as important in developing countries as they are in developed countries, if not more so. Developing countries are rapidly urbanizing, and thus the scope for policy to affect urban settlement patterns (which may stay in place for decades) is much greater than in richer countries. Whether cross-group disparities will be entrenched by urban settlement patterns remains to some extent a policy choice in cities that are still quickly growing.

In this paper, we mobilize new administrative data to describe settlement and segregation patterns of marginalized groups across Indian cities and villages, and the relationship between these settlement patterns and access to public services. Because of the substantial data requirements, there is little prior research on residential segregation in developing countries, and even basic descriptive facts about the extent of segregation and its relationship to public service access are lacking.

We focus on the segregation of members of Scheduled Castes, often called Dalits or (previously) Untouchables, and of Muslims. India is an important context in which to study these questions for several reasons. First, it is huge: the marginalized groups that we study number over 300 million individuals. Second, disparities across these groups are rooted in historical inequalities that have persisted for generations, but the extent to which those inequalities are being changed by market liberalization and urbanization remains an open question. Third, the policy and planning process in India remains focused on disparities at aggregate levels, such as the district (an aggregation of approximately 1000 villages and 10 towns); a recognition of how these aggregate plans translate to neighborhood-level outcomes is essential to understanding whether these policies are achieving their objectives.

We have two primary aims. First, we document the extent of residential segregation in rural and urban areas. Second, we describe how a range of neighborhood public services—schools, medical clinics, water, sewerage, and electricity—are distributed across marginalized group (MG) and non-marginalized group (non-MG) neighborhoods. Studying the causal effect of segregation on individual outcomes is left for future work, as it is complicated by the two-way relationship between being poor and living in a neighborhood with other poor people.

Our "neighborhoods" are groups of approximately 125 contiguous households (∼500 people) that were assigned together for census enumeration. Three subsequent Indian surveys used the same enumeration blocks, allowing us to link neighborhood demographics with information on access to public services. The enumeration blocks were drawn for the 2011 Population Census, which describes village and town demographics and public services in detail. We obtain household-level demographics and several living standard measures, including piped water, electricity and drainage, from the 2012 Socioeconomic and Caste Census (SECC). Scheduled Caste status is directly recorded, and we infer Muslim identity indirectly from the distinctive patterns of Muslim names.[1] We get information on public facilities from the 2013 Economic Census, which collected data on the universe of public and private schools and health centers, making it the only large-sample data source (to our knowledge) that can identify these services at the neighborhood level. Combining these datasets, we document individual demographics,

[1]We classify names as Muslim or non-Muslim using a long-short-term-memory (LSTM) neural network based on a training set of two million takers of the Indian Railways Exam. The out-of-sample accuracy against a set of manually classified names is 97% (Ash et al., 2025).

socioeconomic outcomes, and neighborhood-level public services for a broad sample covering over 60% of India's population.[2]

SCs and Muslims make up similar population shares in the country (17% and 14% respectively in 2011), but have distinct group histories. Scheduled Castes were consigned to the lowest occupational rungs of society for over a thousand years, but since independence, improving access to services for SCs has been an explicit goal of many governments. Empirical studies have found positive effects for some of the resulting affirmative action programs (Cassan, 2019; Gulzar et al., 2020; Asher et al., 2024).

Different Muslim groups have historically occupied heterogeneous positions in Indian society over the generations; some Muslims are descendants of India's 15th to 18th century ruling classes, while others descend from lower-caste groups who converted to Islam to escape their status at the bottom of the social hierarchy. Groups from both heritages increasingly find themselves politically marginalized and threatened. A large literature has discussed the relative outcomes of Scheduled Castes, and there is a more moderate literature on Muslims.[3]

We present three key findings. First, Muslims and SCs have notably segregated residential patterns, slightly lower than Black Americans and non-White people in England and Wales, but higher than minority groups in almost all other comparison countries. We focus on the dissimilarity and isolation indices, for which comparative statistics from other contexts are most widely available. The ranking of Muslim vs. SC segregation varies depending on the measure used, because the isolation index practically puts more weight on the most segregated neighborhoods, where there are more Muslims. 26% of Muslims live in neighborhoods that are more than 80% Muslim, compared with 16% of Scheduled Castes.

Urban and rural segregation are highly correlated across regions for both Muslims and SCs, suggesting that Indian cities are replicating rural settlement patterns that have been in place for hundreds of years.[4] Compared with SCs, Muslims are relatively more segregated in cities

[2] Our analysis dataset describes 400,000 urban neighborhoods (in 3500 cities and towns) and 1.1 million rural neighborhoods (in about 400,000 villages).

[3] For example, see Basant et al. (2010) and Jaffrelot and Gayer (2012).

[4] While the data do not record when these settlement patterns emerged, the historical record suggests that

than in rural areas. Marshaling the limited time series data available, we show that the urban segregation of Scheduled Castes has marginally diminished from 2001–11.[5]

Larger, poorer, and older cities tend to have higher Muslim and Scheduled Caste segregation, as do cities that have experienced more religious violence.[6] Cities with more Muslims have more Muslim segregation (in parallel with segregation patterns for Black communities in the U.S.), but there is no relationship between SC share and SC segregation. We find a strong negative correlation between Muslim segregation and upward mobility (defined as an increase in relative educational position across generations), which is notable given that Muslims are the least upwardly mobile major social group in India (Asher et al., 2024).

Next, we show that access to public services is systematically worse in neighborhoods where marginalized groups live. This holds for both Muslims and SCs, and for almost every local service that we could measure, including primary and secondary schools, medical clinics, piped water, electricity, and covered sewerage. Private providers do not make up for the reduced service access of marginalized groups; in fact, private services are also less accessible in MG neighborhoods, in part because these neighborhoods are poorer.[7]

The magnitude of the disparities is large. Compared with a 0% Muslim neighborhood, a 100% Muslim neighborhood in the same city is 10% less likely to have piped water and only half as likely to have a secondary school. For schools and clinics, facilities provided entirely by government, the disadvantage in Muslim neighborhoods is double the disadvantage in SC neighborhoods, echoing a consistent finding across the qualitative literature that Muslims report difficulty in getting public facilities from their representatives (Jaffrelot and Gayer, 2012). For electricity, water, and drainage, goods which have both a private (hook-up) and public (infrastructure) component,

rural Indians have been highly endogamous and segregated, such that village settlement patterns observed today have been static for decades, if not centuries.

[5]We calculate segregation in 2001 using the Population Census District Handbooks, which record enumeration block-level populations of Scheduled Castes, but not for any other group.

[6]This result is consistent with qualitative work arguing that religious violence has motivated Muslims to self-segregate for safety.

[7]Controlling for neighborhood living standards shrinks, but does not erase, MG disadvantages. Since our interest is in how ostensibly universal government services are allocated, we view the uncontrolled estimates as more relevant for our study.

SCs (who are somewhat poorer on average) face worse neighborhood-level disadvantages.

Disparities look different at higher levels of aggregation. Districts and subdistricts with many SCs have more public facilities on average. However, the cross-neighborhood allocation of these services *within* subdistricts and towns means that nearly all of these advantages are eliminated at the neighborhood level. Muslim neighborhoods, in contrast, have no advantage or disadvantage at higher levels of aggregation; the neighborhood disparity (which is large) is the aggregate disparity.

In short, marginalized groups face the most systematic and substantive disadvantages at the most local and informal levels of government — within towns and village clusters. These are the levels of government which operate with the least scrutiny, and at the greatest distance from the district and subdistrict levels at which affirmative action policies are codified.[8]

A key prior work in this space is Banerjee and Somanathan (2007), who showed that *districts* with more SCs and Muslims had fewer public services in 1971, and that the gap closed almost entirely or even reversed for SCs by 1991, but was little changed for Muslims. Our cross-district results for SCs are consistent with these results, but we show that the cross-district SC advantage is almost completely undone by the cross-neighborhood disadvantage. It remains an open question whether the aggregate convergence process of the 1970s and 1980s affected cross-neighborhood inequalities at all.

Systematic analysis of neighborhood access to public services in developing countries has been elusive because of an absence of neighborhood-level census data. While several of India's major sample surveys contain neighborhood identifiers, they are not powered to measure neighborhood characteristics like social group shares, nor do they have enough sample to measure urban segregation. Prior work on segregation in India includes a number of ward-level studies that use spatial units of population 30,000–200,000, making them 60–400 times more coarse than the neighbor-

[8]We found that people growing up in marginalized group neighborhoods—regardless of their social group—have systematically worse educational outcomes. These results are left for future work, because we lack the data to distinguish whether these outcome disparities are caused by unequal service access, discrimination, or just sorting of marginalized people into poor and un-serviced neighborhoods.

hoods in our analysis.[9] A series of recent studies has used enumeration block data similar to ours to document average patterns of segregation in a subset of Indian cities.[10] Other than a concurrent study using similar data in Brazil (Harari, 2024), we are aware of no prior work studying public service provision or individual outcomes at the neighborhood level in India, or any other major developing country.[11] Even at the village level in India, economic work on Muslim villages is rare, because data on village Muslim shares has not been previously available. Finally, we are aware of no prior quantitative work systematically studying access disparities *within* villages.

Importantly, the neighborhood-level disparities that we study are in many cases not apparent in aggregate data. Federal and state policies in India largely allocate funding for public services at aggregate levels (state, district, or subdistrict), while the cross-neighborhood distribution of those services is determined through less formal local processes. Consequently, a policy maker observing school allocation only at the district level could arrive at incorrect conclusions regarding access disparities and the efficacy of equalization policies. Our work underscores the return to leveraging high-resolution administrative data — which is available but under-used in many developing countries — to better understand and evaluate the performance of public programs.[12]

[9] Vithayathil and Singh (2012) use ward data from 2001 to show that residential segregation by caste is more prominent than by socioeconomic status in seven major cities. Singh et al. (2019) examine changes in caste-based segregation from 2001 and 2011, again at the ward level, finding that residential segregation by caste has persisted or worsened in 60% of the cities in their sample. Neither of these studies examine religion, which is less-often disaggregated than Scheduled Caste or Scheduled Tribe status in Indian Census microdata.

[10] Bharathi et al. (2021b) use enumeration block-level data to describe SC segregation and its correlates in the 150 largest cities. They find that blocks are more homogeneous than wards, highlighting the importance of working at this granular level. Like us, they find a weak relationship between city size and segregation (among large cities). Bharathi et al. (2021a) use similar-scale data on caste (jati) and religion to characterize segregation in rural Karnataka; they report a high level of rural segregation. Susewind (2017) measures Muslim segregation using microgeographic polling booth data in eleven cities. He compares the segregation measures with qualitative perceptions of marginalization and ghettoization, finding only a limited correspondence.

[11] Kumar (2025) uses municipal complaints data from Mumbai to show that Muslims have less consistent water access. Bharathi et al. (2022) uses block-level data to find that ward-level access to water and sewerage is correlated with ward-level SC segregation, and inversely correlated with block-level SC segregation; they do not look at outcomes in SC neighborhoods. Harari (2024) suggests a segregation measure incorporating distances to every neighborhood (and its demographics) in the city; we cannot calculate this in our setting because we observe neighborhood identifiers but cannot map them to geocoded locations.

[12] A public version of the neighborhood dataset will be published with this paper.

2 Context and Background

2.1 Marginalized Groups in India: Scheduled Castes and Muslims

Indian society has historically been stratified by social class, most directly through caste and endogamy. India's Scheduled Caste communities (SCs) are historically endogamous groups that occupy the lowest tiers of the caste system. They have experienced occupational and social segregation for thousands of years. Social norms have historically placed them in low-status occupations—like scavenging, emptying toilets, or handling animal carcasses—with virtually no prospect of upward mobility. The practice of "untouchability," now banned but still practiced in some form by many households, can take the form of segregation in schools, temples, and markets, restrictions against entering the homes or even wearing sandals in the presence of higher caste groups, among others. These restrictions have been enforced with various social sanctions, including violence and murder (Girard, 2021). Since independence, the government of India has worked to mitigate the socioeconomic disadvantages of SCs through a range of programs and policies. SC status is often used as a marker of poverty for means-tested welfare programs, and there are reserved positions for SCs in higher education, in politics, and in government. SC communities still experience substantial socioeconomic disadvantages, but by many measures the gap between SCs and general castes has shrunk somewhat over recent decades (Hnatkovska et al., 2012; Emran and Shilpi, 2015; Cassan, 2019; Asher et al., 2024).

Muslims occupy a similar share of the population to Scheduled Castes (14% for Muslims vs. 17% for SCs), with a higher share of members living in urban areas than any other social group. Like SCs, they on average have lower socioeconomic status than non-Muslim non-SCs. However, they experience fewer legal protections and have not been targeted by affirmative action policies, a few exceptions notwithstanding. While SCs have been gaining ground on higher-caste groups in socioeconomic terms, Muslims have if anything been losing ground, particularly in educational attainment, and have experienced significant losses in upward mobility in recent decades (Asher et al., 2024). Post-independence India has been characterized by waves of anti-Muslim activism, sometimes resulting in riots, property destruction, and violence. Various social movements and political parties have mobilized around the idea of Hinduism as a key pillar of Indian identity, to the exclusion of Muslims (Jaffrelot, 2021). Our analysis uses data from 2011–13, and thus predates the rise of the current Modi regime (which has roots in these social movements), though the BJP (Modi's party) held power nationally in the early 2000s, and held power in many states before and during our sample period.

While SCs and Muslims represent the largest disadvantaged groups in India, there are several other social groups not separately considered by our analysis. "Other Backward Castes" (OBCs) occupy an intermediate place in the caste system between general castes and SCs, comprising 40% of the population; IHDS 2011 reports that about half of Muslims are OBCs, though this share varies substantially across years and surveys. OBCs are not coded as such in any of the datasets that we use and their names are less distinctive, making it difficult to identify them (or their prevalence in any neighborhood) in our data. We also exclude Scheduled Tribes (STs) from our analysis; they are among the poorest social groups in India, but are concentrated in rural areas and have small population shares in the majority of cities.[13] Given the focus of this paper, we use the terms "marginalized groups" or "MGs" to describe SCs and Muslims, even though other groups in India could also reasonably be classified as such.

2.2 Marginalized Group Settlement Patterns in Rural and Urban India

Pre-independence cities in India were often characterized by neighborhoods with homogeneous occupational groups, often with mixed religion. In the absence of an effective municipal state, these neighborhoods were self-governing with respect to many public services, sometimes including even self-defense. Many neighborhoods had only a small number of entries, which made it possible to restrict access; this structure persists in many urban neighborhoods today, resulting in distinct boundaries between neighborhoods (Gist, 1957; Gould, 1965; Lynch, 1967; Doshi, 1991; Sachdev and Tillotson, 2002).[14] The ethnographic literature suggests a secular

[13] Only 4% of urban Indians report Scheduled Tribe status, compared with 15% who are SCs and 17% Muslims.

[14] These closed neighborhoods are described by different terms throughout the country: pols in Ahmedabad, mohallas in much of North India, paras in West Bengal, etc., often with names that reflect the occupational

trend of increasing segregation by religion rather than by occupation, as Hindu-Muslim violence has reduced Muslim feelings of safety in mixed neighborhoods. The resulting Muslim neighborhoods can house individuals from many classes, often with income segregation existing within the neighborhoods at a smaller scale. Jaffrelot and Gayer (2012) describe this pattern of Muslim segregation in a series of monographs spanning many parts of the country. In many of the case studies, Muslims report difficulties getting attention from politicians or access to public services in their segregated neighborhoods.

The literature on villages also suggests spatial separation between different classes and religions; individuals from lower status social groups often live in hamlets that are separated by a moderate walking distance from the village's primary agglomeration, where schools and health centers typically are found (Beteille, 2012; Himanshu et al., 2018).

While these patterns can be observed in many parts of the country, they are primarily documented in a qualitative literature (with some cites above), due to a general absence of large-scale data with neighborhood identifiers or of sufficient scale to characterize neighborhoods individually. There is a quantitative literature on unequal access to public services by caste *across* districts (Banerjee and Somanathan, 2007) and villages (Bailwal and Paul, 2021). The latter is possible because the Population Censuses record the SC population share of every village, along with a series of public services. Nationwide data on village-level Muslim shares did not exist before this paper, nor were there data on either social group shares or public services at the neighborhood level *within* villages. To our knowledge, there has also been no large-sample study of public service variation *within* cities; a key innovation of this paper is assembling near-universal urban neighborhood-level data that simultaneously describes both public services and marginalized group shares.

2.3 Public Service Provision and Levels of Government in India

India has a federal system of government with major powers divided between central, state, and local governments. There are 36 states and union territories (35 at the time of our sample),

origins of the space. Muchipara, for instance, is "the neighborhood (para) of cobblers (muchi)."

which have substantial administrative and legislative power. Public services are financed and allocated by both central and state government programs.

Local governance bodies are called panchayats in villages and municipalities in towns and cities. These bodies have elected representatives who can have de jure power over the selection and allocation of public services within their administrative areas, but have little control over their overall budgets, most of which derive from grants from higher levels of government. The 73rd and 74th Constitutional Amendments devolved power to allocate and provide public services to these local bodies. In villages, panchayat leaders often exercise de facto decision-making power over service allocation. Devolution in cities and towns is incomplete, however, and state-controlled authorities are often the final decision-makers.

The process through which urban allocation decisions are made is highly political and depends on citizen organization. Residents often self-organize into neighborhood welfare committees and organize protests, petitions and media to lobby for service access. Local, informal, party-affiliated power brokers are often the key mediators to whom citizens turn for help in accessing resources. The de facto informal struggle for resources, mediated by local brokers, is remarkably similar for the different types of public services studied in this paper — electricity, water, sewerage, schools, and health clinics (Jensenius, 2017; Kruks-Wisner, 2018; Auerbach, 2019; Auerbach and Kruks-Wisner, 2020).

India's government has implemented many high-profile policies intended to close disparities between marginalized and non-marginalized groups; many of these are conceived and designed at the state and central levels, and have prescribed allocations of public services across regions intended to remedy cross-group inequality, especially for Scheduled Castes. For instance, the District Primary Education Program targeted funds for building schools to districts with below-median female literacy (Khanna, 2023). The placement of new public facilities *within* districts, towns, and villages is less often directly prescribed by these high-level policies; it is instead agreed upon through consultation with local elected leaders, bureaucrats, and brokers. The extent to which policies target certain groups can therefore be different at different levels of aggregation;

the less formal decision-making process of local bodies could either enhance or undermine the progressivity of policies designed at higher levels of government (Alatas et al., 2012).

3 Neighborhood-level Data on Social Groups and Public Services

This section briefly describes the data sources used for this work. Additional detail on sources and data construction can be found in Appendix B.1.

3.1 Identifying Neighborhoods

Measuring segregation requires data with neighborhood granularity and a high neighborhood count both within and across cities, which cannot be done in any of India's major sample surveys. We identified and combined several data sources that recorded the internal location identifiers that were created for the administration of India's 2011 Population Census, called "enumeration blocks;" we call these neighborhoods. They consist of 100–125 households each (approximately 500 people) and describe a compact cluster of residences that an enumerator could visit in a single work session. In cities, these are typically city blocks, while in rural areas they are grouped clusters of residences or entire villages.[15]

Rural and urban enumeration blocks have similar populations, but the geography of urban and rural access to public facilities is quite different. Urban areas are dense, such that people can travel across many enumeration blocks for work or access to public services. Rural areas are more dispersed: neighboring enumeration blocks can be adjoining but can also be separated by several kilometers in the case of single-block villages. We did not have access to geocoordinates or boundaries for enumeration blocks.[16]

[15] Appendix Figure A.1 shows the enumeration block population distribution in the sample. We exclude outlier blocks with fewer than 150 (typically small villages) or more than 1000 people, which are less than 1% of the sample. Note that enumeration blocks are different from census blocks (sometimes simply called blocks), which have populations of about 200,000 each, and are unrelated to any units used in this paper.

[16] Enumeration block maps are sold individually as hand-drawn maps by the Indian Census. Geocoding the universe of 1.5 million enumeration blocks would be valuable, but was prohibitively costly. At the time of writing, they were sold for about $2 per individual block, and would additionally need to be georeferenced, as done by Gechter and Tsivanidis (2023) for Mumbai.

3.2 Demographic Data

Data on individuals comes from the 2011–12 Socioeconomic and Caste Census (SECC), a national census which recorded demographic and asset information on every person in India to determine eligibility for social programs. Individual-level data from SECC was posted online from 2013–14 in PDF format, and included respondent names and Scheduled Caste status; we scraped and cleaned the data in a lengthy process described in detail in Asher and Novosad (2020). Our dataset covers 196 million urban and 571 million rural respondents, compared with 385 million and 834 million in the 2011 Census respectively.[17] We used respondent names and an LSTM neural network to classify individuals as Muslims or non-Muslims; out-of-sample accuracy was 97%, due to the distinctive nature of Muslim names.[18] For comparisons with the United States, we used data from the 2020 U.S. Census and the Diversities and Disparities project, which is based on the 2010 U.S. Census.

3.3 Public Facilities and Public Infrastructure

We identified public facilities at the neighborhood level using the 2013 Economic Census (EC13), which used the same enumeration blocks as the 2011 Population Census. EC13 is a complete enumeration of non-farm establishments in the country, which includes schools, clinics, and hospitals, which are separately coded as private or public.[19] Health centers include hospitals, inpatient and outpatient clinics, and traditional care providers.[20]

EC13 also records whether a firm owner is Muslim or SC. The employment share in SC or

[17] To the best of our knowledge, missing data was a function of the actions of IT administrators and was unrelated to the data content. Block-level PDFs were to be posted in 30-day rolling periods; at some times, the SECC site was completely inaccessible, and some locations were posted for shorter periods or not posted at all. We discuss the representativeness of these data in Section 5. See Asher and Novosad (2020) and Asher et al. (2021) for more details on the scraping process.

[18] For more details, see Ash et al. (2025) and Appendix B.1.

[19] Earlier rounds of the Economic Census (1990, 1998, 2005) record similar data, but with neighborhood identifiers (urban frame survey units) that do not match any census. It was thus not possible to study changes in neighborhood-level service access over time.

[20] The Economic Census is imperfect for studying access to public services, given its focus on firms. For instance, Das et al. (2022) report that it undercounts health service providers by a factor of two in Madhya Pradesh, likely because it misses birth attendants and untrained providers. But there is no other data source of which we are aware that records universal information on public facilities at the neighborhood level.

Muslim firms is highly correlated with the group share in each neighborhood, providing some validation for the data on demographic shares below.

Data on household access to clean water, electricity, and sewerage was recorded at the individual level in the SECC. These are semi-public services, as households sometimes need to pay a hookup fee. In practice, the vast majority of neighborhoods in our sample had access shares close to 0 or 100%, indicating that the public infrastructure component is the primary determinant of access. We coded neighborhoods with a binary access measure for each service if more than 50% of households reported access.[21]

Neighborhoods were thus coded with binary access measures for both public facilities and services. While these are imperfect access measures for schools and clinics, which can be accessed in other neighborhoods, there is evidence that distance to schools matters in both cities and villages, especially for girls' schooling (Chandra and Staiger, 2007; Muralidharan and Prakash, 2017; Borker, 2021). For water, sewerage, and electric light, their presence in nearby neighborhoods is clearly a poor substitute for residential access. The presence of neighborhood schools and clinics are also relevant political economy outcomes even if they do not fully define access to those facilities.

Additional data sources used in the segregation correlates analysis (Section 5.2) are described in Appendix B.

4 Methods

4.1 Measuring and Comparing Residential Segregation

Our first objective is to document the extent of residential segregation of Muslims and SCs and to put it into international context. We calculated the canonical dissimilarity and isolation indices, as these are the most widely used measures in other studies. Comparability across contexts is still not straightforward, because calculating each measure requires several decisions which are not made consistently in the prior literature, and are sometimes not even reported. We first describe the measures that we use and then the details of making them comparable across contexts.

[21] Because so many neighborhoods were close to 0 or 100%, different thresholds had no material effect on results.

We describe each measure for the case of Scheduled Castes and take an analogous approach for the case of Muslims. When measuring Muslim segregation, we treat SCs as majority group members, and vice versa when measuring SC segregation. We take this approach because Muslim and SC segregation may have different dynamics, causes, and consequences, making it less useful to aggregate them in analysis.

The Dissimilarity Index. The Dissimilarity Index (**?**) is the most widely used measure of segregation. It ranges from zero to one and answers the question: what share of the marginalized group would need to change neighborhoods for the group to be evenly distributed within a city? We calculate this index for Scheduled Castes in a given city as:

$$DISSIMILARITY = \frac{1}{2}\sum_{i}^{N}\left|\frac{SC_i}{SC_{total}} - \frac{NON\_SC_i}{NON\_SC_{total}}\right|,$$

where SC_i is the Scheduled Caste population of block i and SC_{total} is the Scheduled Caste population of the city, and similarly for non-SCs.

The Isolation Index. The Isolation Index (Bell, 1954) measures the extent to which a population group is exposed only to members of its own group. In a given city, we calculate isolation on a [0,1] scale as follows:

$$ISOLATION = \sum_{i \in I}\left[\frac{SC_i}{SC_{total}} \cdot \frac{SC_i}{N_{total,i}}\right]$$

where definitions are the same as above and $N_{total,i}$ is the total block population and N_{total} is the city population.[22]

This specification of the isolation index can be directly interpreted as the marginalized group share in the average neighborhood of a member of the given marginalized group. Whether dissimilarity or isolation is the better measure depends on the functional form of the relationship between neighborhood share and social cost, which is not known. The two are highly correlated,

[22]Note that the minimum value of the isolation index under this definition is not zero, but rather the city-level marginalized group share.

but isolation is more responsive to highly concentrated MG neighborhoods (Massey and Denton, 1988).[23,24]

For urban areas, we calculate dissimilarity and isolation for each city/town, defining enumeration blocks as neighborhoods. In rural areas, we calculate the indices for each subdistrict, again with enumeration blocks as neighborhoods.[25]

International comparisons of segregation are challenging, because several measurement factors can directly affect segregation estimates. For example, U.S. segregation is higher when weighted across cities by Black population (to reflect the minority experience), when a smaller subset of large cities is used (because large U.S. cities are more segregated), when using MSA boundaries vs. city boundaries (because MSAs include disproportionately white suburbs), and when neighborhood units are smaller. For the India-U.S. comparison, we select comparable parameters to the main estimates in the U.S. literature; for other countries, where few estimates are available, we note below the differences which could affect results. Appendix B.2 describes how we selected and aligned all these patterns in more detail.

All of our measures are aggregates of block-level population that ignore the locations of the blocks relative to each other. As such, they all suffer from the "checkerboard" problem — a city with a checkerboard arrangement of neighborhoods has exactly the same segregation measure as a city where all the minority blocks are clustered together. Harari (2024) suggests a measure to address the checkboard problem and applies it in Brazil; unfortunately it cannot be applied in our context, because our data does not describe the location of each enumeration block within each city.

[23] Both measures take the value 1 for a maximally segregated city. A fully integrated city has a zero dissimilarity index, and an isolation index equal to the marginalized group share of the city (which in this case would be the MG share of every neighborhood as well). Some authors rescale the isolation index from 0 to 1, at the cost of losing the intuitive interpretation given here. Our international comparison below scales the isolation index as we do here.

[24] Many other measures are available in the literature (exposure, evenness, entropy), but all are highly correlated with the measures we use, and are worse in terms of either comparison availability or interpretability.

[25] A subdistrict consists of about 110 villages; there are about 5500 subdistricts in India. The rural measure thus captures a combination of segregation across villages and within villages.

4.2 Marginalized Group Shares and Neighborhood Public Services

Our second objective is to describe differences in access to public services between marginalized groups and non-marginalized-groups, and at which geographic levels they arise. We present an additive decomposition, where the total group disparity is the sum of the cross-state, cross-district, cross-city, and cross-neighborhood disparity.

We begin by describing how a fixed supply of public services is allocated across MG and non-MG neighborhoods within cities with the following neighborhood-level regression:

$$SERVICE_{i,c} = \beta_c MG\_Share_{i,c} + \Omega_c + \boldsymbol{\nu} POPULATION_{i,c} + \epsilon_{i,c}. \tag{1}$$

$SERVICE_{i,c}$ is a measure of the supply or availability of a given public service in neighborhood i and city c, such as an indicator for the presence of a secondary school. $MG\_Share_{i,c}$ is the share of people in neighborhood i from a given marginalized group. We control for neighborhood population, since it could be mechanically related to the supply of public services (though there is not much variation in neighborhood population). The city fixed effect Ω_c controls for differences in the availability of services in cities with more or fewer members of a given marginalized group.[26]

The coefficient β_c describes how service availability changes as the marginalized group share increases. A negative value indicates that a given public service in a city is allocated *away* from neighborhoods where marginalized groups live. We use the subscript c, because this measure describes a characteristic of the political economy equilibrium in each *city*.

We next examine how this local inequality relates to inequalities at higher levels of aggregation. If we replace the city fixed effect (Ω_c) in Equation 1 with a *district* fixed effect Ω_d, the coefficient on $MG\_SHARE_{i,c}$ then describes the allocation of services *within* districts; we call this β_d. This measure describes a combination of (i) the allocation of services across towns within districts; and (ii) the allocation of services across neighborhoods within towns. It is therefore useful to define $\alpha_d = \beta_d - \beta_c$, which specifically identifies the component of service

[26] Standard errors are clustered at the city level in the urban analysis and at the subdistrict level in the rural analysis, to account for correlated outcomes within regions.

access that comes from variation *within* districts and across towns.[27] We call this α_d because it describes the political economy equilibrium of the district — the outcome of the process by which public services are allocated within the district.

We repeat this process at progressively higher scales. Equation 1 with state fixed effects gives us $\alpha_s = \beta_s - \beta_d$, which describes how services are allocated across districts within states. The same equation with no fixed effects gives us $\alpha_f = \beta_o - \beta_s$, where α_f is the allocation of services across states.[28] The total disparity experienced by the marginalized group is β_o, an additive combination of political economy processes at different scales of geography and government, such that $\beta_o = \alpha_f + \alpha_s + \alpha_d + \alpha_c$.

All of the α terms are independently interesting, as they describe the allocation process at different scales of geography and government, where different forces apply. For example, if a state explicitly allocates services to districts with higher Scheduled Caste shares, this would suggest a positive value of α_s; this allocation could then be amplified or undermined at higher or lower geographies.

The decomposition also has implications for progressive policy. For example, suppose that α_c is highly negative (i.e. marginalized group neighborhoods have worse services, conditional on city fixed effects). In this case, the disparity can be reduced through policies that increase α_s (e.g. through affirmative action programs operating across districts), but this district-level targeting will be less efficient at reducing disparities than neighborhood-level targeting (which would affect α_c directly).

Identifying the geographic scale of disparity is particularly relevant given the different institutions controlling public service allocation at different geographic levels. Most policy research in India operates at the district level, as do many programs which allocate public services. High level policy-makers and researchers may not have access to local data, causing them to

[27]One could estimate a similar parameter directly in a town-level regression with district fixed effects. The advantage of our approach is that our estimates are additive across geographic levels. The approach is similar to the Blinder-Oaxaca decomposition, but with a disparity measured as a regression coefficient, and where the covariates are hierarchical locations.

[28]We use subscript "f" because α_f describes the federal (i.e. cross-state) political economy equilibrium.

misunderstand the nature of inequality. Our decomposition clarifies what information is lost by studying differences at aggregate levels. If we studied only the relationship between marginalized group share and public service outcomes at the district level, we would be measuring $\alpha_f + \alpha_s$, which is a biased measure of β_o if local disparities are large.

Our estimates do not isolate a causal effect of marginal group share on outcomes. For example, if marginalized groups are poor, and municipal governments undersupply public facilities to poor neighborhoods, then we would find $\beta_c < 0$ even if service provision was orthogonal to MG status, conditional on neighborhood income. In this case, MGs would still have worse access to public services—the outcome that we aim to measure.[29] Our null hypothesis is that the government allocates public facilities across neighborhoods equally, irrespective of neighborhood economic or social group status, in which case we would find $\beta_c = 0$.

We can think of the α terms as allocation rules; they describe the de facto outcomes of the allocation process at different geographic levels. For example, α_d can be thought of as the district allocation rule, which describes how a district's resources are allocated across towns in that district. These "rules" are outcomes of a complex and obscure political economy process influenced by politicians, bureaucrats, firms, and citizens. These "rules" are jointly determined by the public service allocation process and the decision choices of individuals. A negative α could reflect government discrimination, or it could reflect historical inequalities that make marginalized groups poorer and more likely to select into neighborhoods with worse public services. The α's are particularly informative for a decision-maker who wishes to target services to under-served groups.

5 Results

5.1 Segregation in Indian Villages and Cities

Table 1 presents summary statistics of the neighborhood-level sample, separately for 400,000 urban and 1.1 million rural neighborhoods. The difference in sample size reflects India's low urbanization rate (31% in 2011), slightly magnified by our worse sample coverage of urban

[29] We do not necessarily get closer to causal identification by adding control variables for neighborhood average education or consumption, because these outcomes are plausibly impacted by a shortage of public services.

places. Scheduled Caste individuals are relatively more likely to live in rural areas, while Muslims are more likely to live in towns and cities.

Table 2 reports town- and rural subdistrict-level summary statistics, including measures of segregation. The ordering of urban segregation of Muslims and Scheduled Castes depends on the measure used; Muslims have a higher isolation index (0.49 vs. 0.43 for SCs) but lower dissimilarity (0.52 vs 0.59 for SCs). The isolation index effectively puts more weight on neighborhoods with very little exposure to other groups and thus reflects a higher number of very concentrated Muslim neighborhoods. In rural areas, by contrast, SCs are a little bit more segregated by both measures.[30] The standard deviations of the segregation indices are between 0.1 and 0.25, reflecting substantial variation in the extent of segregation across cities.

Figure 1 helps to unpack these differences by showing the distribution of MG shares across neighborhoods. The Muslim distribution is notably bimodal, and a greater share of Muslims than SC lives in highly concentrated neighborhoods in both urban and rural areas. 26% of urban Muslims live in neighborhoods that are >80% Muslim, while 17% of urban SCs live in neighborhoods that are >80% SC.[31]

Figure 2 compares these results with estimates of minority segregation from other contexts. The comparison is limited by available estimates in the literature; we found few studies of segregation from low- and middle-income countries, and few measures of isolation outside of the United States. Immigrant populations are also emphatically different from the social groups we study here, in that the vast majority of SCs and Muslims are long-resident citizens from minority social groups. As noted in Section 4, isolation and dissimilarity are scale-dependent, so we plot neighborhood size on the X axis, and aggregate our measures to different neighborhood

[30] Table 2 also compares the sample characteristics with the full set of towns and villages in the Population Census. The rural sample covers 81% of rural subdistricts and 68% of rural people, and is highly representative. The urban sample covers 50% of cities and towns and 51% of the urban population. The town sample is less representative of small and young cities, which have slightly more public services, but similar marginalized group shares to older/larger towns. Because our sample has better coverage of large than small cities, it covers 83% of the urban population, making it more representative of people than it is of places. Reweighting the sample to match the full distribution of cities does not substantially alter any of our conclusions; Appendix Table A.1 shows the segregation indices after entropy rebalancing town demographic characteristics.

[31] In cities, the median Muslim lives in a neighborhood that is 47% Muslim. In rural areas, this is 37%. For SCs, these numbers are 38% and 46%, almost exactly the reverse.

sizes for comparison.[32]

After adjusting neighborhood scales, we find that the urban segregation of Indian Muslims is similar but a little bit lower than the current segregation of Black people in U.S. cities and towns. This makes India considerably more segregated than Brazil, the only other low- or middle-income country for which estimates were available, and it occupies a midpoint on the spectrum of immigrant segregation in Europe — higher than Italy and the Nordic capitals, but lower than immigrant segregation in Spain and non-White segregation in England and Wales.[33]

The segregation of Scheduled Castes is similar to that of Muslims at the small block scale used in our analysis. Aggregating neighborhoods brings down SC segregation considerably, indicating that segregated SC neighborhoods are less clustered than segregated Muslim neighborhoods.[34]

Appendix Figure A.3 shows maps of SC and Muslim segregation across the country. While there are pockets of high and low segregation, they do not follow obvious geographic patterns; the north, which is poorer and where people are less disposed toward cross-caste marriage — see, for example, Sahgal et al. (2021) — is no less segregated than the south.

We next examine whether rural segregation patterns are being replicated in cities. Given India's rapid urbanization in the second half of the twentieth century, settlement patterns in cities reflect more recent decisions and norms around integration and separation of social groups. Figure 3 shows binscatters of average urban and rural segregation levels in each district; rural and urban segregation are highly correlated for both Muslims and SCs ($\rho \in 0.43-0.73$), suggesting that the regional dynamics that lead to the separation of social groups in rural areas

[32] Unlike the U.S. estimates, Indian segregation is less variable by city size; these estimates use the full sample of 3504 cities/towns. Sources and details of each sample in the graph are described in Appendix Table A.2.

[33] The comparisons with the U.S. are the most reliable, as there are enough U.S. estimates to understand the extent to which those differ by sample selection and city definition. The literature does not give us information on how precise or stable are the estimates from the European cities.

[34] The U.S./India comparison is imperfect, because segregation patterns scale differently in different contexts, including across U.S. cities (Reardon et al., 2008), and we did not have data to calculate how U.S. estimates would change with smaller units. Nevertheless, our scaling adjustment should be superior to directly comparing estimates at different scales. The relevant scale for segregation-related policy may also differ across countries. Relative to U.S. cities, transport costs (in opportunity cost of both time and money) are higher in Indian cities, suggesting the reduced segregation at higher aggregate levels may be less relevant in India. More examination of the scaling characteristics of segregation would obviously be useful, but would necessitate geocoded enumeration blocks to which we do not have access.

are also important in neighboring cities and towns.

Finally, we marshal the limited data available to examine changes in Scheduled Caste segregation over time, using block-level demographic tables from the 2001 Census District Handbooks (see Appendix B for more details). Table 3 shows that, as measured by both dissimilarity and isolation, SC segregation decreased marginally from 2001–11, falling by a statistically insignificant 2–4%. Residential patterns are often stable, but finding only small changes in segregation is not a foregone conclusion, particularly given the Indian government's efforts to improve SC integration. For comparison, between 1980–2020, the Black/White dissimilarity index fell from 0.579 to 0.425, an average decline of 6% per year (Logan and Stults, 2021).

5.2 Correlates of Scheduled Caste and Muslim Segregation

We next examine the correlates of segregation at the town and city level. We run a regression of each segregation measure (Muslim or SC vs. Dissimilarity or Isolation) on a joint set of 10 covariates that are based on correlates of segregation identified in the literature from other countries. Figure 4 shows estimates from these regressions, where the X variables are normalized to make them more comparable. A coefficient of 0.1 on this graph implies that a 1 standard deviation increase in the covariate is associated with a 0.1 change in segregation; for reference, the dissimilarity difference between immigrants in Scandinavia and U.S. Blacks is about 0.2. Table 5 shows estimates from the four non-normalized regressions.[35]

We also generated non-parametric binscatters of each bivariate relationship; these can be seen at devdatalab.org/seg-correlates. We show a sample of these in Appendix Figure A.4.[36] We did not include highly correlated covariates because of multicollinearity; e.g. we included city per capita consumption but excluded average years of education, which was highly correlated.

Larger, poorer, and older cities are systematically more segregated for both Muslims and members of Scheduled Castes; consumption is particularly negatively monotonic in Muslim

[35]The sample size is limited by missing data on upward mobility and urban inequality. The full sample regression without these covariates is displayed in Appendix Table A.3.

[36]Note that we exclude the group population shares from the isolation index coefficient plot, as these are mechanically correlated with each other and can be misleading, but they are included in the multivariate regression.

segregation (Appendix Figure A.4). All three social groups (SCs, Muslims, and non-SC non-Muslims) are poorer in more segregated cities, a pattern notably different from the U.S., where Black residents are poorer in more segregated cities but the income slope for White residents is zero. Both MGs in India are more segregated in cities that experienced more Hindu-Muslim violence since 1950. This is consistent with the narrative of Muslim segregation as a defense against violence, but our analysis makes no claim of causality.

The remaining correlates differ in direction and significance for the two groups. Cities where upward mobility is low have systematically more segregated Muslim populations; this is notable, since Muslims are the group with the lowest upward mobility in India (Asher et al., 2024). Group shares are also predictive of segregation; Muslims are more segregated in cities with higher Muslim populations, but the effect is fully explained by city size and consumption level.

5.3 Access to Public Services in Marginalized Group Neighborhoods

In this section, we examine how the supply of public services varies across neighborhoods with and without concentrated marginalized groups. We focus on availability of public services at the most granular geographic level—the neighborhood—because it is the most relevant for individual access to services, and is also the least studied in prior work.

Figure 5 shows a binned scatterplot of the neighborhood-level relationship between the supply of secondary schools (an indicator for the presence of a neighborhood school) and the neighborhood marginalized group share, in both urban and rural areas. The urban series is residualized on city fixed effects and thus describes how schools are distributed across neighborhoods, conditional on the total supply of schools in a city. Secondary school availability falls monotonically with the neighborhood Muslim share (Panel A); raising the Muslim share of a neighborhood by 50 percentage points is associated with a 22% lower likelihood of the neighborhood having a public secondary school (approximately a 0.5 percentage point decline on a mean of 2.4%). Neighborhoods with a greater than 50% Muslim share stand out for being particularly underprovisioned; while only about 10% of neighborhoods have such a high Muslim share, over half of India's urban Muslims live in them (Figure 1). Rural locations look broadly similar, with the most Muslim neighborhoods having substantially fewer schools (Panel C).

The relationship between Scheduled Caste share and secondary school access is non-monotonic in both urban and rural areas; at low levels of SC shares, it is flat or rising in the SC share, but above a 20% SC share, secondary school presence falls precipitously, such that 50% SC neighborhoods have similar school availability to 50% Muslim neighborhoods (Panels B and D).[37]

We summarize this nonparametric relationship between neighborhood MG share and public facility presence with the linear estimator from Equation 1, with city fixed effects. SC and Muslim shares are included simultaneously to ensure that the allocation of facilities to one group's neighborhoods does not drive our estimate for the other group. Panel A of Table 4 shows that, in urban areas, SC and Muslim neighborhoods are systematically allocated fewer public services; with the exception of urban primary schools in SC neighborhoods, the point estimates are all negative, substantial, and highly significant. In rural areas (Panel B), the estimates are negative and significant for all facilities, for both groups. In short, the local political economy equilibrium systematically results in marginalized groups living in neighborhoods that are less well-served by public facilities.[38]

Table 6 shows analogous tests with private schools and health facilities, which could substitute for the absence of public sector facilities. In fact, we find that private facilities are also disproportionately allocated away from marginalized group neighborhoods, possibly because people in those neighborhoods have limited ability to pay for services. There are some exceptions: out of 12 group x urban/rural x facility estimates, 10 show statistically and economically significant allocation away from MG neighborhoods. The exceptions are private primary schools and health facilities, which are more common in rural Muslim neighborhoods.

We find similar results for household infrastructure services (access to electricity, closed drainage, and clean water, Table 7). These services are only measured in urban areas. All three

[37]Rural school shares are higher on average because rural areas are characterized by a greater number of smaller schools, reflecting the greater distance between neighborhoods.

[38]Results are similar when we use a measure of the scale of the facilities (log employment, shown in the even-numbered table columns). Results are virtually unchanged (i) by the inclusion of a control for whether a neighborhood is classified as a slum; and (ii) by restricting the sample to non-slum villages (Appendix Table A.4).

services are systematically less available in both Muslim and Scheduled Caste neighborhoods. For these infrastructure goods, the coefficients on the SC share are more negative than those on the Muslim share, suggesting that SC neighborhoods are the most poorly served by public infrastructure.[39]

The infrastructure services are particularly relevant, because urban residents may be able to travel to nearby schools and clinics in other neighborhoods, but water, electricity, and sewerage in nearby neighborhoods are less useful substitutes for own-neighborhood access.

5.4 Access Disparities at Different Geographic Levels of Aggregation

So far, we have found that public services are systematically allocated away from marginalized group neighborhoods at the most local level. However, this disparity does not summarize the total access disparity faced by marginalized groups, because there could be favorable or unfavorable differences in the supply of services at higher geographic levels of aggregation. For instance, districts with more Scheduled Castes might have more schools or better sanitation infrastructure; indeed, the Indian government has used the district or subdistrict Scheduled Caste share as a targeting mechanism for many programs (see Section 2).

We measure allocation at each geographic level of aggregation by varying the fixed effects in Equation 1. We can thus additively decompose the total urban access disparity into a disparity across neighborhoods, towns, districts, and states.

Panel A of Figure 6 summarizes the results for Muslim access to urban primary schools. We explain these figures in detail as they describe a central result of this paper. The outcome variable is the number of primary schools per 100,000 people; the sample mean of this variable is 15. The rightmost (dark gray) box (positioned at −1.9) tells us that a 100% Muslim neighborhood is estimated to have 1.9 fewer primary schools per 100,000 people than a 0% Muslim

[39]Note that these infrastructure services are not strictly public. They typically require some kind of household investment in addition to a base level of public infrastructure, but none of them can be accessed if that public infrastructure is not in place. Our estimations are run at the neighborhood level, and thus do not identify off of within-neighborhood differences in whether members of different social groups choose whether or not to hook up to each infrastructure service. The distributions of neighborhood availability of these services are highly bimodal, suggesting that the public component of the infrastructure is the key determinant of individual access.

neighborhood.[40] This is the coefficient from a regression of the primary school measure on the neighborhood Muslim share, with no fixed effects (β_o from Section 4.2). This coefficient reflects the total access disparity in Muslim neighborhoods, combining effects at all geographic levels.

This gap can then be decomposed into different geographic levels. The leftmost estimate $\alpha_f = -0.4$ tells us that states with more Muslims have fewer schools, and that 0.4 out of the 1.9 school gap above can be accounted for by this variation across states. The second estimate from the left ($\alpha_s = +1.1$) implies that — conditional on the number of primary schools in a state — districts with more Muslims on average have *more* primary schools.[41] The next two bars respectively give us α_d, which tells us how schools are allocated across towns/cities within districts, and α_c, which tells us how schools are allocated across neighborhoods within towns.[42]

The sum of all the α coefficients gives us the final estimate of -1.9 schools per 100,000 people. The graph shows that the neighborhood disadvantage faced by Muslims is driven almost entirely by the allocation of primary schools across urban neighborhoods within towns. In fact, the allocation combining all aggregates *above* the town level is marginally favorable to Muslims; but this small advantage is swamped by the unfavorable allocation across neighborhoods.

The remaining five panels of Figure 6 show how the other public facilities (secondary schools and health centers) are allocated across Muslim and non-Muslim neighborhoods, towns, districts, and states. We highlight several features of the combined results. First, the cross-neighborhood allocation (labeled "x-block") is systematically unfavorable for Muslims — as we saw in Table 4. Second, in urban areas, the magnitude of the cross-neighborhood inequality swamps the magnitude of the inequality at every other level of aggregation. It is at the lowest and most informal level of governance where Muslim neighborhoods are the most left out. In rural areas, allocation is unfavorable at every level of aggregation for all three facility types, and the impact is more uniform across geographic scales. Third, without neighborhood-level data, we

[40] The sample means for the other variables are in the figure note.

[41] We denote this α_s because it is informative about allocation choices at the *state government* level — it describes how schools are allocated across districts within states.

[42] The calculation of α_c is identical to that used for the coefficients in Table 4, except that the outcome here is denoted in schools per 100,000 individuals for easier interpretability of the effects across different geographies.

would detect no disadvantage in access to public facilities for Muslims in cities, and we would substantially underestimate the disadvantage in rural areas. Since the Indian government does not release data on Muslim shares below the subdistrict level, about half of the rural inequality in service access is invisible in the data available prior to this paper, as is all of the urban inequality.

The total disparity faced by Muslim neighborhoods is economically substantial. A fully Muslim neighborhood—recall that over half of Muslims live in neighborhoods that are more than 50% Muslim—has 13% fewer primary schools, 46% fewer secondary schools, and 20% fewer health centers than a neighborhood with no Muslims.

Figure 7 shows the same results for SC neighborhoods. The patterns are distinct from those observed for Muslims, even though both groups face substantial disadvantages at the most local level. A clear pattern emerges for secondary schools and health centers, in both rural and urban areas (Panels C–F). The allocation of these services is progressive across states, districts, towns, and villages; at all of these levels, areas with more SCs have more secondary schools and clinics. But *within* towns and villages, the distribution of schools and clinics is highly regressive across neighborhoods, undoing almost all of the progressivity at higher levels of government. Ignoring the cross-neighborhood allocation of secondary schools and clinics (which no prior data source has made visible) would make it appear that public services are strongly favorably targeted to places where SCs live, but in fact the total allocation is approximately neutral.

The allocation of primary schools to SC neighborhoods does not follow this pattern. Urban primary schools have progressive allocations for SCs at all levels of aggregation, while the allocation of rural primary schools is unfavorable to SCs at all geographic levels, but with the neighborhood being relatively unimportant. This distinct result could arise from the government's efforts to make primary schools universal across India, though clearly Muslim neighborhoods have been left out. The neutral to positive neighborhood allocation of primary schools could result from an interaction of that universal goal with a preference for segregating upper class children from SC children, but this is left as a topic for future work.

Section 5.3 showed that the cross-neighborhood allocation of public facilities was more unfavorable to Muslims than to SCs. This section shows that this is even more true across larger geographic units. When the effects are combined, we see that Muslim neighborhoods are systematically lacking in public facilities, while SC neighborhoods in the end have similar service levels to non-SC neighborhoods — the latter result arises from favorable allocation across large geographic units (like districts) but unfavorable allocation across neighborhoods.

Patterns like these could arise if affirmative action policies for Scheduled Castes (policies which have been prominent in India since independence) primarily affect the distribution of public services across higher units of aggregation, like states and districts. If these policies bind only at high levels of aggregation, and the less formal political processes of neighborhoods and municipal governments remain biased, then the cross-neighborhood allocation of services can undo some of the progressive allocation at higher levels of government. Muslims face the same or worse disadvantages as Scheduled Castes at the cross-neighborhood level, but with no systematic policy of affirmative action, there is no force to mitigate those disadvantages and Muslims end up substantially less well-served.

Figure 8 shows the same analysis for the infrastructure services: electric lighting, piped water, and closed drainage. For SCs, the cross-neighborhood variation in access drives almost all of the substantial access disparity, and there is little association between the SC share and infrastructure availability at the state, district, or town level. For Muslims, at the state and district levels, we find that piped water access is more common in districts with many Muslims, while electric light and drainage are less common. As noted above, the allocation across neighborhoods is economically significant and adverse for all of these services, for both groups.[43]

For the infrastructure services, there is thus less evidence of affirmative action in favor of any marginalized group, but both groups fare worse at the neighborhood level. Indeed, we are aware of no national programs to improve urban infrastructure services like these or to equalize

[43]Appendix FigureA.7 shows that there is no "intersectionality" effect, in that the SC disadvantage does not appear to be worse in neighborhoods with many Muslims, and vice versa. Appendix Figure A.8 shows the results as a function of city segregation. Interestingly, there is no clear relationship between city-level segregation and the cost of living in a segregated neighborhood; that is, segregated neighborhoods are not systematically more disadvantaged in cities with more of them. However, the average disadvantage naturally affects more members of marginalized groups in cities where more of them live in segregated neighborhoods.

access to them from the time period up to our sample. It is also notable that the relative access of the two groups is reversed for the infrastructure services; at both the cross-neighborhood and the overall level, Scheduled Castes neighborhoods have disproportionately worse access to water, electricity, and sewerage infrastructure than Muslims.[44]

6 Discussion

Our descriptive results raise three questions for future work. First, is segregation driven by discrimination or homophily? Second, why do segregated neighborhoods have worse public services? And third, are the poor outcomes in these neighborhoods *caused* by living in these neighborhoods? Answering these questions is beyond the scope of this descriptive work; but in this section, we discuss possible mechanisms and some of the existing evidence supporting it.

Empirical research on segregation in other contexts suggests that both selection and treatment play a role, and there is evidence of the same in India (Blank et al., 2004). Consistent with social groups in many other contexts, a large fraction of Indians from all social groups expresses preferences for living around members of their own group (Sahgal et al., 2021), and social group networks provide valuable services (such as insurance), for which members are willing to give up economic opportunities (Munshi and Rosenzweig, 2006, 2016). These choices are then further reinforced by formal and informal discrimination, such as landlords and home sellers refusing to work with buyers and renters from marginalized groups, another practice that echoes the U.S. experience (Banerjee and Knight, 1985; Sachar Committee Report, 2006; Thorat and Attewell, 2007; Madheswaran and Attewell, 2007; Thorat et al., 2015).

Since members of marginalized groups are poorer on average, these forces cause poor people to cluster together in neighborhoods with fewer amenities where they can afford to live. Any

[44] Appendix Figures A.5 and A.6 shows similar estimates to Figures 6 and 7 for private facilities. We spend less attention on these since there are few political forces driving their allocation at higher levels of aggregation. As noted in the prior section, cross-neighborhood allocation of private services is strongly unfavorable for marginalized group neighborhoods. Appendix Table A.5 shows the neighborhood-level public service estimates with controls for neighborhood consumption. Neighborhood consumption explains some but not all of the disparity in MG neighborhoods; results vary by group and public service. But these estimates do not help distinguish the existence of discrimination, because low consumption in MG neighborhoods could be in part the result of poor service access or other disamenities of those neighborhoods.

discrimination in state service provision (whether against the poor or against marginalized groups) further exacerbates that disadvantage. We do not have data to test whether specific public services and facilities appeared before or after social group settlement patterns were established. But settlement patterns move slowly, and a large share of India's public services have been built in the modern era. The number of urban high schools increased by a factor of three between 1990 and 2013, a period during which the urban population rose by only 75%.[45] We thus view it as unlikely that all of our neighborhood results could be explained away by selection of poor people into neighborhoods with few services, but that is undoubtedly an economic factor that exists. In villages, demographic characteristics are highly stable over time and thus predate the expansion of modern public schooling and infrastructure.

We are doubtful that disparities are meaningfully caused by low preference for public services among members of marginalized groups. Qualitative evidence suggests that Muslims would prefer better access to public services, but are unable to extract them from local politicians (Jaffrelot and Gayer, 2012; Tariq, 2025). The aggregate service access improvements for lower caste groups happened exactly as these groups were mobilizing politically and creating political parties to represent their interests, something that Muslims have never done successfully (Banerjee and Somanathan, 2007; Aneja and Ritadhi, 2022). Finally, we note that arguments about homophily and low demand for public services were frequently and erroneously invoked as a justification for U.S. policies like red-lining, which entrenched or increased disparities across races (Blank et al., 2004). The U.S. historical record suggests caution about treating cross-group inequalities in public services as benign.

Studies in other contexts have tackled the question of whether segregation causes adverse outcomes by identifying instruments, such as train tracks, that facilitated greater segregation in some areas than in others (Cutler and Glaeser, 1997; Ananat, 2011; Aaronson et al., 2021); this would be a useful direction for future work in India.

[45]The high school statistic is from the 1990 and 2013 Economic Censuses. The DISE 2012 school survey reports that 55% of urban high schools were built in the last 30 years. Meanwhile, in IHDS 2012, 75% of urban households have been in the same location for the last 30 years or longer. A large fraction of urban residents have been in place before their nearest high school was built.

7 Conclusion

Our paper presents a national-scale analysis of segregation and access to public services in India's urban and rural neighborhoods. Analysis of this kind has previously been impossible on a large scale due to the absence of sufficient neighborhood-level data to characterize neighborhood demographics and service access.

India's growing cities are significantly segregated. They are only marginally less segregated than rural areas, where neighborhood structure is strongly conditioned by centuries of occupation- and status-based division via the caste system. The religious and caste identity of the people who live in a given urban neighborhood are strongly predictive of public services in those neighborhoods. India's rapidly growing cities, considered to be engines of upward mobility, to a large degree have replicated the caste and religious disparities of its villages.

The public data generated by this project is a starting point for the systematic study of state resources and economic opportunity across Indian cities. While data on past segregation is scarce, our project will make time series analysis possible going forward. We have also identified substantial variation across cities and rural subdistricts in both segregation and the relative public service disadvantages experienced by marginalized groups. Future work that addresses the causes and consequences of these disparities is essential.

Concentration of marginalized groups and unequal provision of public services are persistent characteristics of the political economy of many countries. Modern India has had few of the types of state policies that contributed to racial segregation in the United States — there are thus fewer overtly harmful policies to remove. However, housing discrimination in India's cities is widely documented and has even been explicitly tolerated by the judiciary, echoing patterns from a too recent era in the U.S.

The historic tolerance for residential segregation and unequal access to public services has prevented generations of individuals from accessing opportunity, and is a central fracture in a highly polarized political system. At an earlier stage of development and with cities still rapidly growing, India has the opportunity to make a different set of choices. By highlighting segregation in India and documenting the concomitant disparities in access to public services, we draw attention to the critical choices that lie ahead for India and other urbanizing lower- and middle-income countries around the world.

References

Aaronson, Daniel, Daniel Hartley, and Bhashkar Mazumder, "The effects of the 1930s HOLC" redlining" maps," *American Economic Journal: Economic Policy*, 2021, *13* (4), 355–92.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, and Julia Tobias, "Targeting the poor: evidence from a field experiment in Indonesia," *American Economic Review*, 2012, *102* (4), 1206–1240.

Alesina, Alberto and Ekaterina Zhuravskaya, "Segregation and the Quality of Goverment in a Cross Section of Countries," *American Economic Review*, 2011, *101* (5), 1872–1911.

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

_ and Ebonya Washington, "Segregation and Black political efficacy," *Journal of Public Economics*, 2009, *93* (5-6), 807–822.

Andersen, Hans Skifter, Roger Andersson, Thomas Wessel, and Katja Vilkama, "The Impact of Housing Policies and Housing Markets on Ethnic Spatial Segregation: Comparing the Capital Cities of Four Nordic Welfare States," *International Journal of Housing Policy*, 2015, *16* (1), 1–30.

Aneja, Abhay and S.K. Ritadhi, "Can political parties improve minority welfare? Evidence from India's "Silent Revolution"," *Journal of Development Economics*, 2022, p. 102931.

Ash, Elliott, Sam Asher, Aditi Bhowmick, Sandeep Bhupatiraju, Daniel Chen, Tanaya Devi, Christoph Goessmann, Paul Novosad, and Bilal Siddiqi, "In-group bias in the Indian judiciary: Evidence from 5 million criminal cases," *Review of Economics and Statistics*, 2025, pp. 1–45.

Asher, Sam and Paul Novosad, "Rural roads and local economic development," *American Economic Review*, 2020, *110* (3), 797–823.

_ , _ , and Charlie Rafkin, "Intergenerational Mobility in India: New Measures and Estimates Across Time and Social Groups," *American Economic Journal: Applied Economics*, 2024, *16* (2).

— , **Tobias Lunt, Ryu Matsuura, and Paul Novosad**, "Development research at High Geographic Resolution: An Analysis of Night-lights, Firms, and Poverty in India using the SHRUG open data platform," *The World Bank Economic Review*, 2021, *35* (4), 845–871.

Auerbach, Adam Michael, *Demanding Development: The Politics of Public Goods Provision in India's Urban Slums*, Cambridge University Press, 2019.

— **and Gabrielle Kruks-Wisner**, "The geography of citizenship practice: How the poor engage the state in rural and urban India," *Perspectives on Politics*, 2020, *18* (4).

Bailwal, Neha and Sourabh Paul, "Caste Discrimination in Provision of Public Schools in Rural India," *The Journal of Development Studies*, 2021, *57* (11), 1830–1851.

Banerjee, Abhijit and Rohini Somanathan, "The Political Economy of Public Goods: Some Evidence from India," *Journal of Development Economics*, 2007, *82* (2).

Banerjee, Biswajit and John B Knight, "Caste discrimination in the Indian urban labour market," *Journal of Development Economics*, 1985, *17* (3), 277–307.

Basant, Rakesh, Abusaleh Shariff et al., "Handbook of Muslims in India: Empirical and policy perspectives," *OUP Catalogue*, 2010.

Bell, Wendell, "A probability model for the measurement of ecological segregation," *Social Forces*, 1954, *32* (4), 357–364.

Benassi, Federico, Ricardo Iglesias-Pascual, and Luca Salvati, "Residential Segregation and Social Diversification: Exploring Spatial Settlement Patterns of Foreign Population in Southern European Cities," *Habitat International*, 2020, *101*, 102200.

Beteille, Andre, *Caste, class and power: Changing patterns of stratification in a Tanjore village*, Oxford University Press, 2012.

Bharathi, Naveen, Deepak Malghan, and Andaleeb Rahman, "A permanent cordon sanitaire: Intra-village spatial segregation and social distance in India," *Contemporary South Asia*, 2021, *29* (2), 212–219.

— , — , **Sumit Mishra, and Andaleeb Rahman**, "Fractal urbanism: City size and residential segregation in India," *World Development*, 2021, *141*, 105397.

— , — , — , **and** — , "Residential segregation and public services in urban India," *Urban Studies*, 2022, *59* (14), 2912–2932.

Blank, Rebecca M, Marilyn Dabady, and Constance F Citro, *Measuring Racial Discrimination*, National Academies Press Washington, DC, 2004.

Borker, Girija, "Safety First: Perceived Risk of Street Harassment and Educational Choices of Women," 2021. World Bank Policy Research Working Paper No. 9731.

Boustan, Leah Platt, "Racial Residential Segregation in American Cities," in "The Oxford Handbook of Urban Economics and Planning," Oxford University Press, 2012.

Cassan, Guilhem, "Affirmative Action, Education and Gender: Evidence from India," *Journal of Development Economics*, 2019, *136.*

Chandra, Amitabh and Douglas O Staiger, "Productivity spillovers in health care: Evidence from the treatment of heart attacks," *Journal of Political Economy*, 2007, *115* (1).

Chyn, Eric, Kareem Haggag, and Bryan A Stuart, "The effects of racial segregation on intergenerational mobility: Evidence from historical railroad placement," 2022. NBER Working Paper No. 30563.

Cutler, David, Edward Glaeser, and Jacob Vigdor, "When are ghettos bad? Lessons from immigrant segregation in the United States," *Journal of Urban Economics*, 2008, *63* (3), 759–774.

Cutler, David M. and Edward L. Glaeser, "Are Ghettos Good or Bad?," *The Quarterly Journal of Economics*, 1997, *112* (3), 827–872.

_ , _ , and Jacob L. Vigdor, "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 1999, *107* (3), 455–506.

Das, Jishnu, Benjamin Daniels, Monisha Ashok, Eun-Young Shim, and Karthik Muralidharan, "Two Indias: The structure of primary health care markets in rural Indian villages with implications for policy," *Social Science & Medicine*, 2022, *301.*

Doshi, Harish, "Traditional neighbourhood in modern Ahmedabad: the Pol," *A Reader in Urban Sociology. New Delhi: Orient Longman*, 1991, pp. 179–208.

Elbers, Chris, Jean Lanjouw, and Peter Lanjouw, "Micro-level Estimation of Poverty and Inequality," *Econometrica*, 2003, *71* (1), 355–364.

Emran, M Shahe and Forhad Shilpi, "Gender, geography, and generations: Intergenerational educational mobility in post-reform India," *World Development*, 2015, *72*, 362–380.

Gechter, Michael and Nick Tsivanidis, "Spatial spillovers from high-rise developments: Evidence from the Mumbai Mills," 2023. Working paper.

Girard, Victoire, "Stabbed in the back? Mandated political representation and murders," *Social Choice and Welfare*, 2021, *56* (4), 595–634.

Gist, Noel, "The Ecology of Bangalore, India: An East-West Comparison," *Social Forces*, 1957, *35* (4), 356–365.

Gould, Harold A, "Lucknow Rickshawallas: The Social Organization of an Occupational Category," *International Journal of Comparative Sociology*, 1965, *6* (1), 24–47.

Gulzar, Saad, Nicholas Haas, and Benjamin Pasquale, "Does Political Affirmative Action Work, and for Whom? Theory and Evidence on India's Scheduled Areas," *American Political Science Review*, 2020, *114* (4), 1230–1246.

Harari, Mariaflavia, "Residential Patterns and Local Public Goods in Urban Brazil," 2024. Working paper.

Himanshu, Peter Lanjouw, and Nicholas Stern, *How lives change: Palanpur, India, and development economics*, Oxford University Press, 2018.

Hnatkovska, Viktoria, Amartya Lahiri, and Sourabh Paul, "Castes and labor mobility," *American Economic Journal: Applied Economics*, 2012, *4* (2), 274–307.

Jaffrelot, Christophe, *Modi's India: Hindu Nationalism and the Rise of Ethnic Democracy*, Princeton University Press, 2021.

_ **and Laurent Gayer**, *Muslims in Indian cities: Trajectories of marginalisation*, Columbia University Press, 2012.

Jensenius, Francesca, *Social Justice through Inclusion*, Oxford University Press, 2017.

Khanna, Gaurav, "Large-scale education reform in general equilibrium: Regression discontinuity evidence from india," *Journal of Political Economy*, 2023, *131* (2), 549–591.

Kruks-Wisner, Gabrielle, *Claiming the State: Active Citizenship and Social Welfare in Rural India*, Cambridge University Press, 2018.

Kumar, Tanu, "How the content of digital complaints shapes bureaucratic responsiveness in Mumbai," *Governance*, 2025, *38* (2).

Logan, John R., "Diversity and Disparities Project: Residential Segregation Data," 2021.

Logan, John R and Brian J Stults, "The Persistence of Segregation in the Metropolis: New Findings from the 2020 Census," *Diversity and Disparities Project, Brown University*, 2021.

Logan, John R. and Brian J. Stults, "Metropolitan Segregation: No Breakthrough in Sight," Technical Report 22-14, Center for Economic Studies, U.S. Census Bureau 2022. Working Papers 22-14.

Lynch, O. M., *Rural cities in India: continuities and discontinuities*, India and Ceylon: Unity and Diversity, 1967.

Madheswaran, Subramaniam and Paul Attewell, "Caste discrimination in the Indian urban labour market: Evidence from the National Sample Survey," *Economic and political Weekly*, 2007, pp. 4146–4153.

Massey, Douglas S and Nancy A Denton, "The dimensions of residential segregation," *Social Forces*, 1988, *67* (2), 281–315.

_ **and** _ , *American apartheid: Segregation and the making of the underclass*, Harvard university press, 1993.

Munshi, Kaivan and M Rosenzweig, "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *American Economic Review*, 2016, *106* (1), 46–98.

_ **and Mark Rosenzweig**, "Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy," *American Economic Review*, 2006, *96* (4), 1225–1252.

Muralidharan, K and N Prakash, "Cycling to School: Increasing Secondary School Enrollment for Girls in India," *American Economic Journal: Applied Economics*, 2017, *9* (3). NBER Working Paper No.19305.

Reardon, Sean F, Stephen A Matthews, David O'Sullivan, Barrett A Lee, Glenn Firebaugh, Chad R Farrell, and Kendra Bischoff, "The geographic scale of metropolitan racial segregation," *Demography*, 2008, *45* (3).

Sachar Committee Report, "Social, Economic and Educational Status of the Muslim Community of India," Technical Report, Government of India 2006.

Sachdev, Vibhuti and Giles Henry Rupert Tillotson, *Building Jaipur: the making of an Indian city*, Reaktion Books, 2002.

Sahgal, Neha, Jonathan Evans, Ariana Monique Salazar, Kelsey Jo Starr, and Manolo Corichi, "Religion in India: Tolerance and Segregation," Pew Research Center June 2021.

Simpson, Ludi, "Ghettos of the Mind: The Empirical Behaviour of Indices of Segregation and Diversity," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, March 2007, *170* (2), 405–424.

Singh, Gayatri, Trina Vithayathil, and Kanhu Charan Pradhan, "Recasting inequality: residential segregation by caste over time in urban India," *Environment and urbanization*, 2019, *31* (2), 615–634.

Susewind, Raphael, "Muslims in Indian cities: Degrees of segregation and the elusive ghetto," *Environment and Planning A*, 2017, *49* (6), 1286–1307.

Tariq, Asad, "Constituencies of Change," 2025. Working paper.

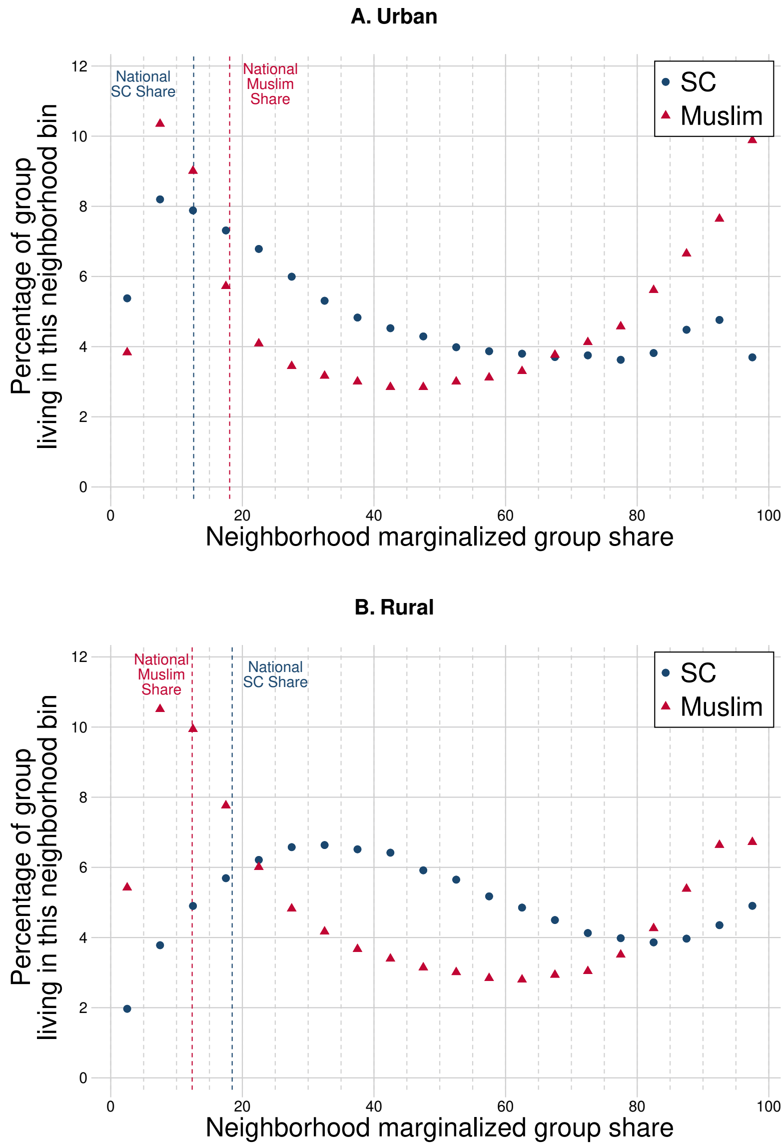
Thorat, Sukhadeo and Paul Attewell, "The legacy of social exclusion: A correspondence study of job discrimination in India," *Economic and political weekly*, 2007, pp. 4141–4145.

_ , Anuradha Banerjee, Vinod K Mishra, and Firdaus Rizvi, "Urban rental housing market: Caste and religion matters in access," *Economic and Political Weekly*, 2015, pp. 47–53.

Varshney, Ashutosh and Steven Wilkinson, "Varshney-Wilkinson Dataset on Hindu-Muslim Violence in India, 1950-1995, Version 2," 2006.

Vithayathil, Trina and Gayatri Singh, "Spaces of discrimination: Residential segregation in Indian cities," *Economic and Political Weekly*, 2012, pp. 60–66.

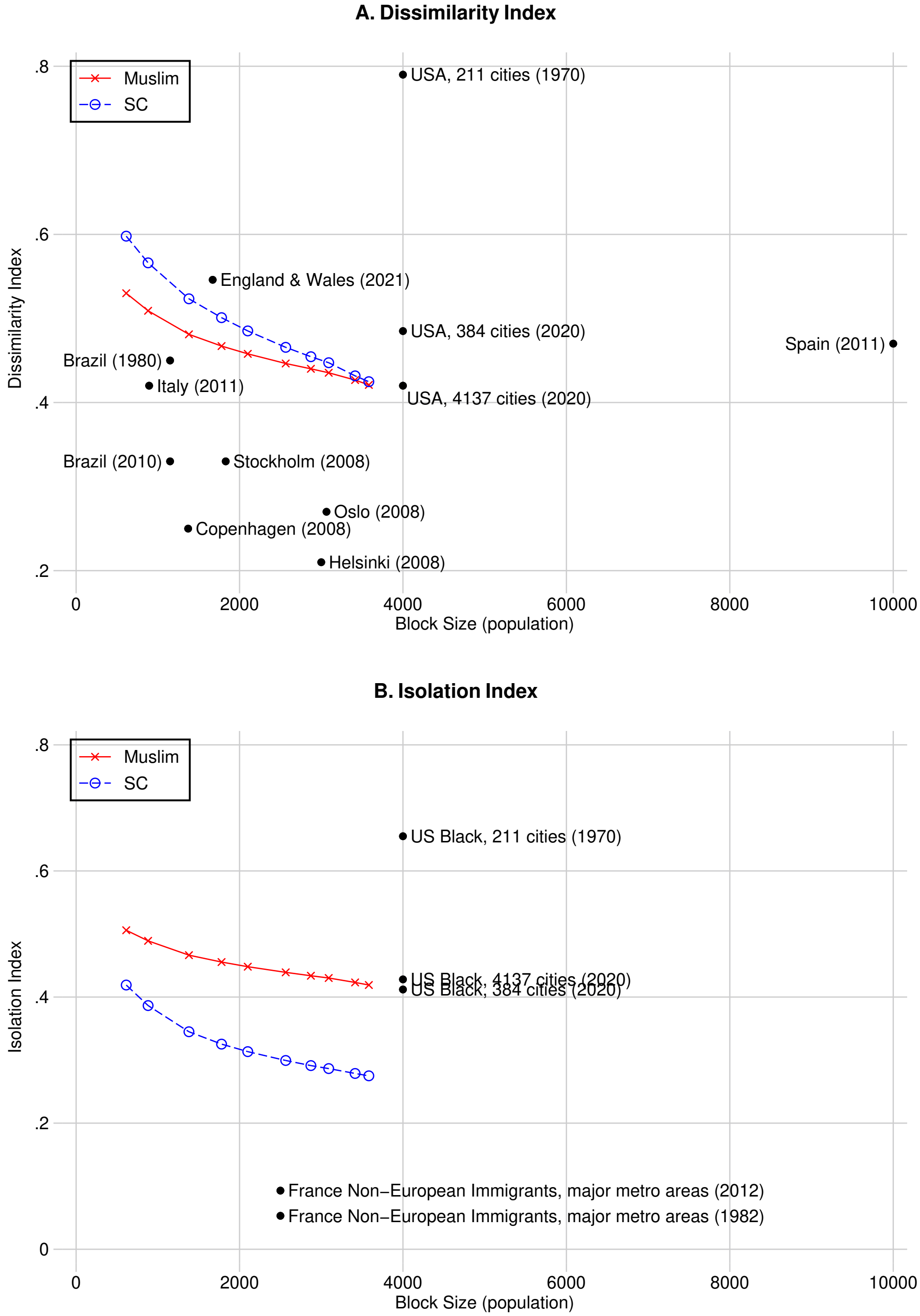
Figure 1
Population Distribution
as a Function of Marginalized Group Share



Notes: The figure shows the distribution of Scheduled Caste (SC) and Muslim population shares across their own neighborhood group share. For instance, the rightmost red triangle in Panel A shows that 6% (Y-axis) of Muslims live in neighborhoods where the Muslim share is between 95 and 100%.

Figure 2

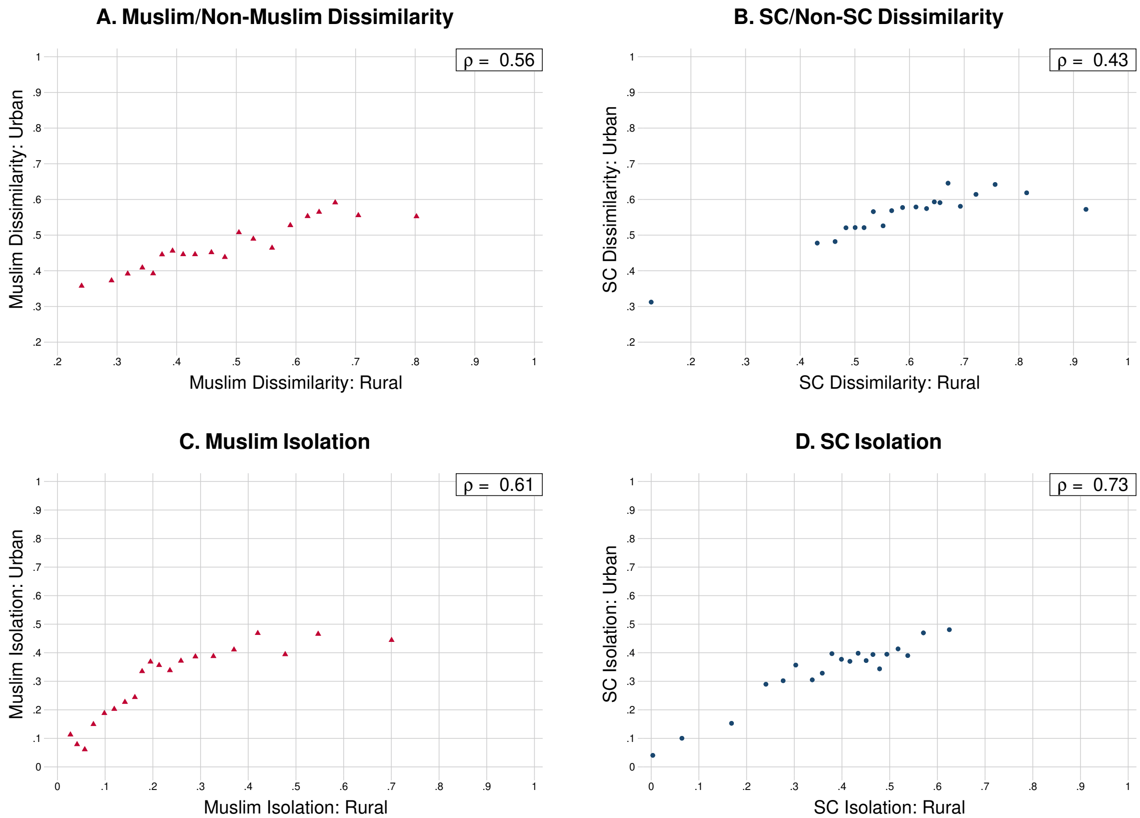
India's Urban Segregation in International Comparison



Notes: The figure shows indices measuring residential segregation in cities across different countries and minority groups. U.S. estimates are Black/White segregation; European and English estimates describe immigrant segregation. Appendix Table A.2 describes the data sources and additional details of each sample.

Figure 3

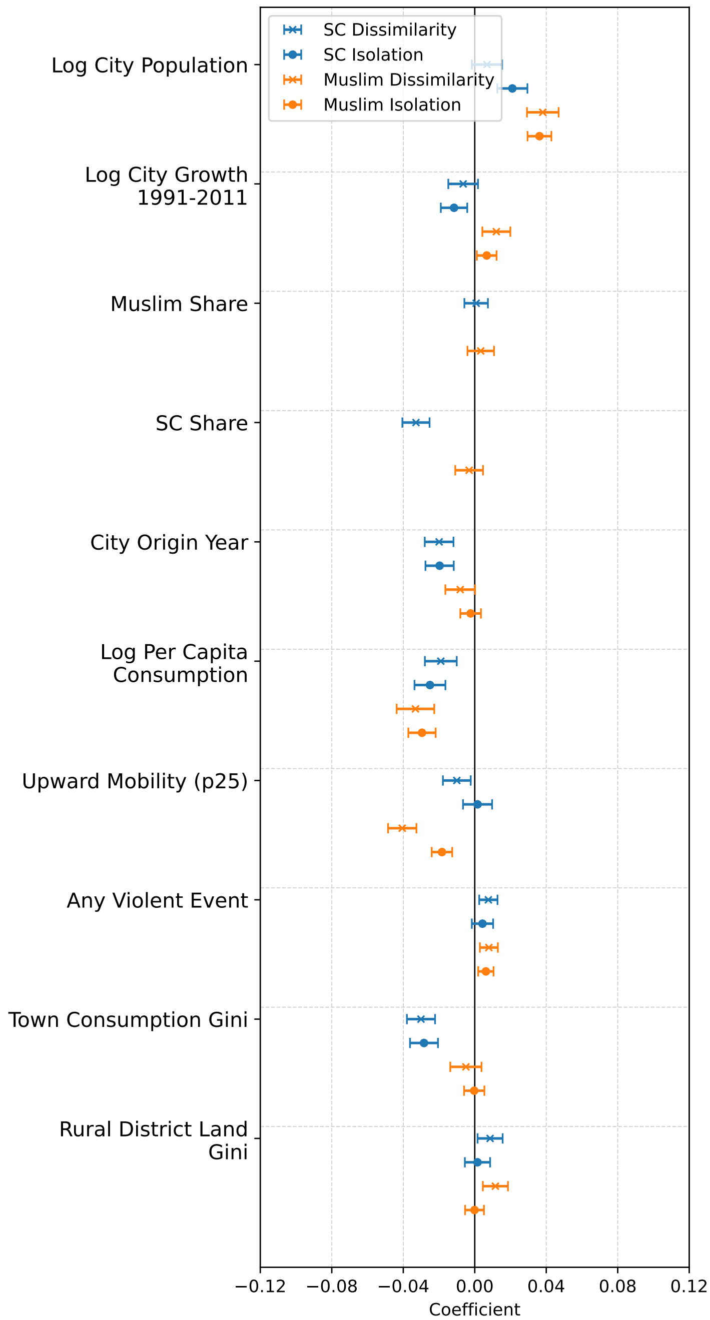
Urban vs Rural Segregation: District-level Comparisons



Notes: Each graph is a binscatter which compares urban and rural segregation measures in the same district, for Muslim and Scheduled Caste communities. Panels A–B use the dissimilarity index; Panels C–D use the isolation index. Each point is the mean urban value for roughly 20 subdistricts with mean rural value near the corresponding X-axis value. Source: SECC (2012).

Figure 4

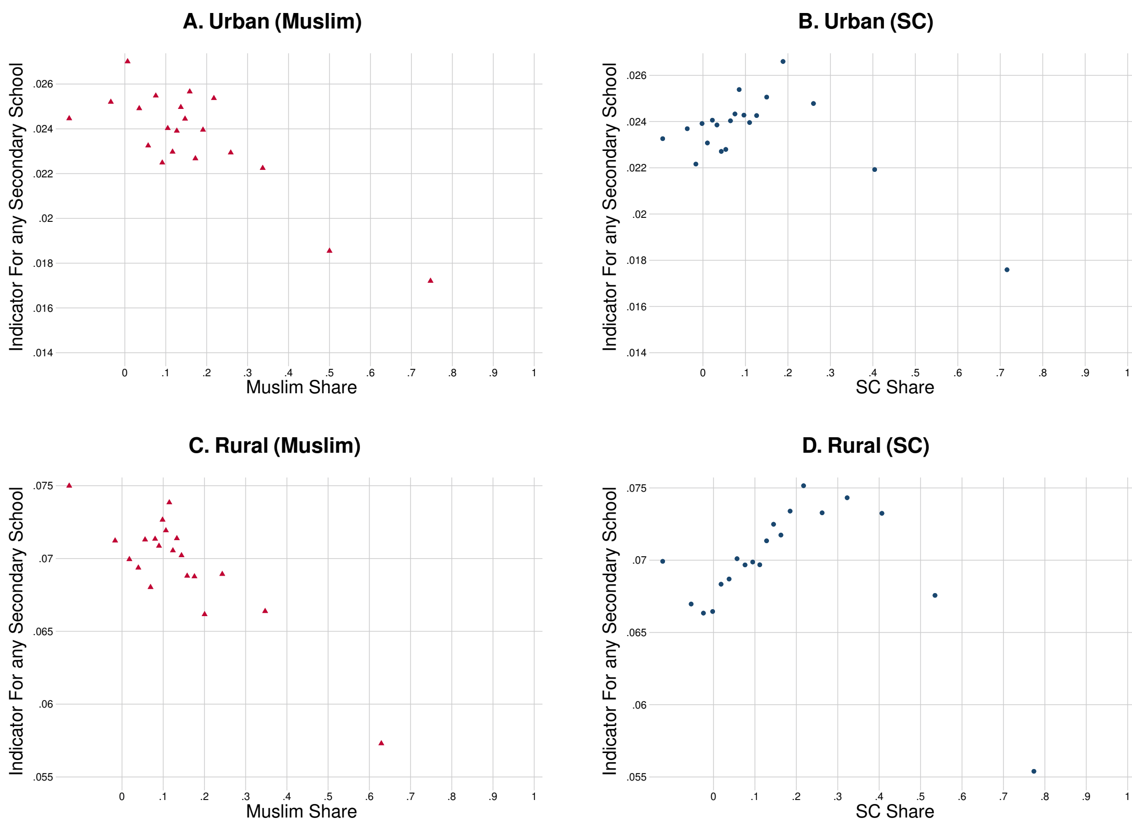
Multivariate Correlates of Muslim and Scheduled Caste Segregation



Notes: The figure shows estimates from four regressions on a set of normalized covariates, where the left-hand side variable was respectively SC dissimilarity, SC isolation, Muslim dissimilarity, and Muslim isolation. Each panel of the figure shows estimates from four regressions. For example, the estimates from the SC Dissimilarity regression are described by the first coefficient in every set of four. Estimates from a regression on raw (non-normalized) covariates are shown in Table 5. Muslim and SC shares were included in the isolation regression, but are excluded from the figure as their high positive correlation is mechanical and thus less informative.

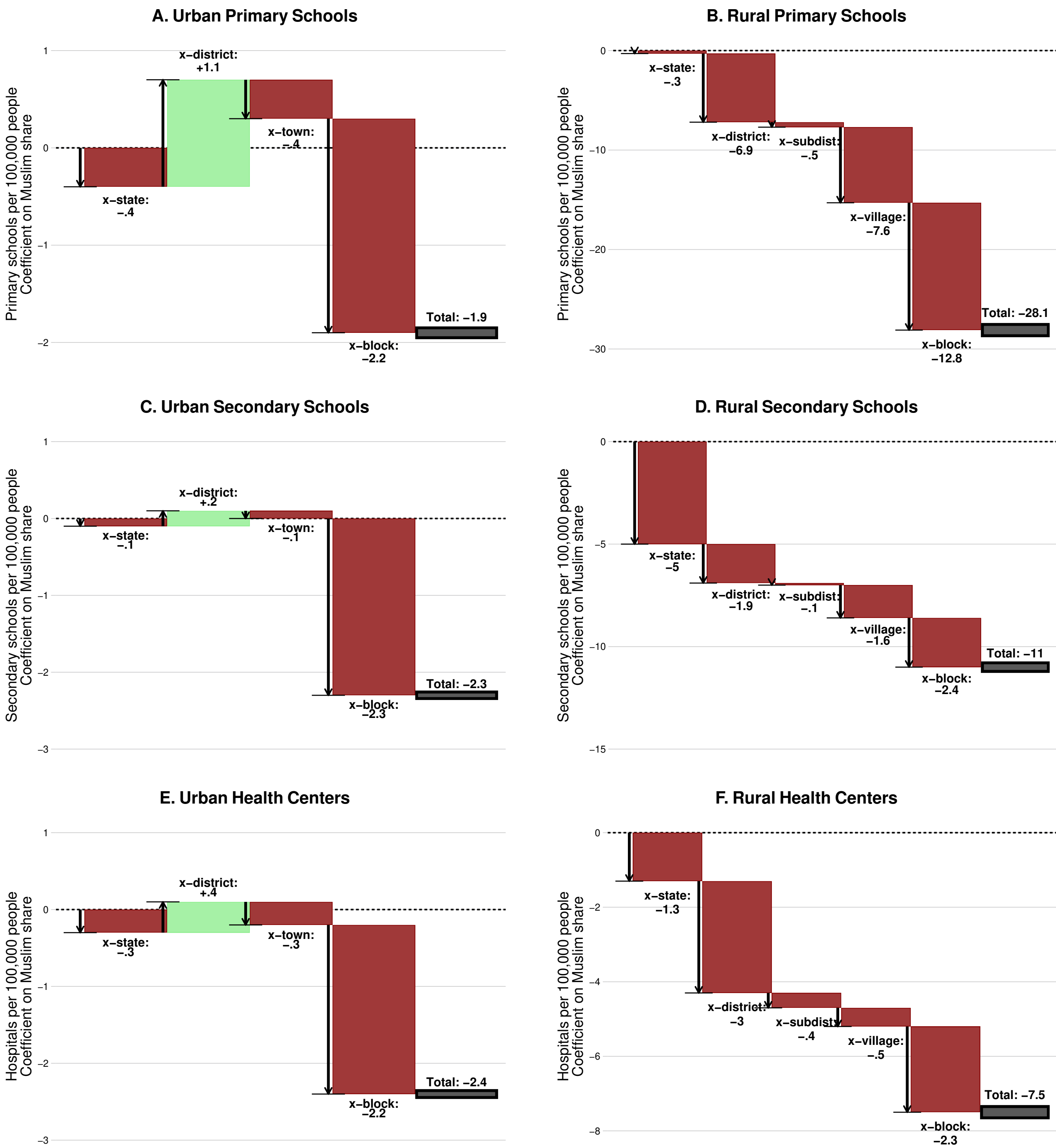
Figure 5

Access to Secondary Schools vs.
Neighborhood Marginalized Group Share



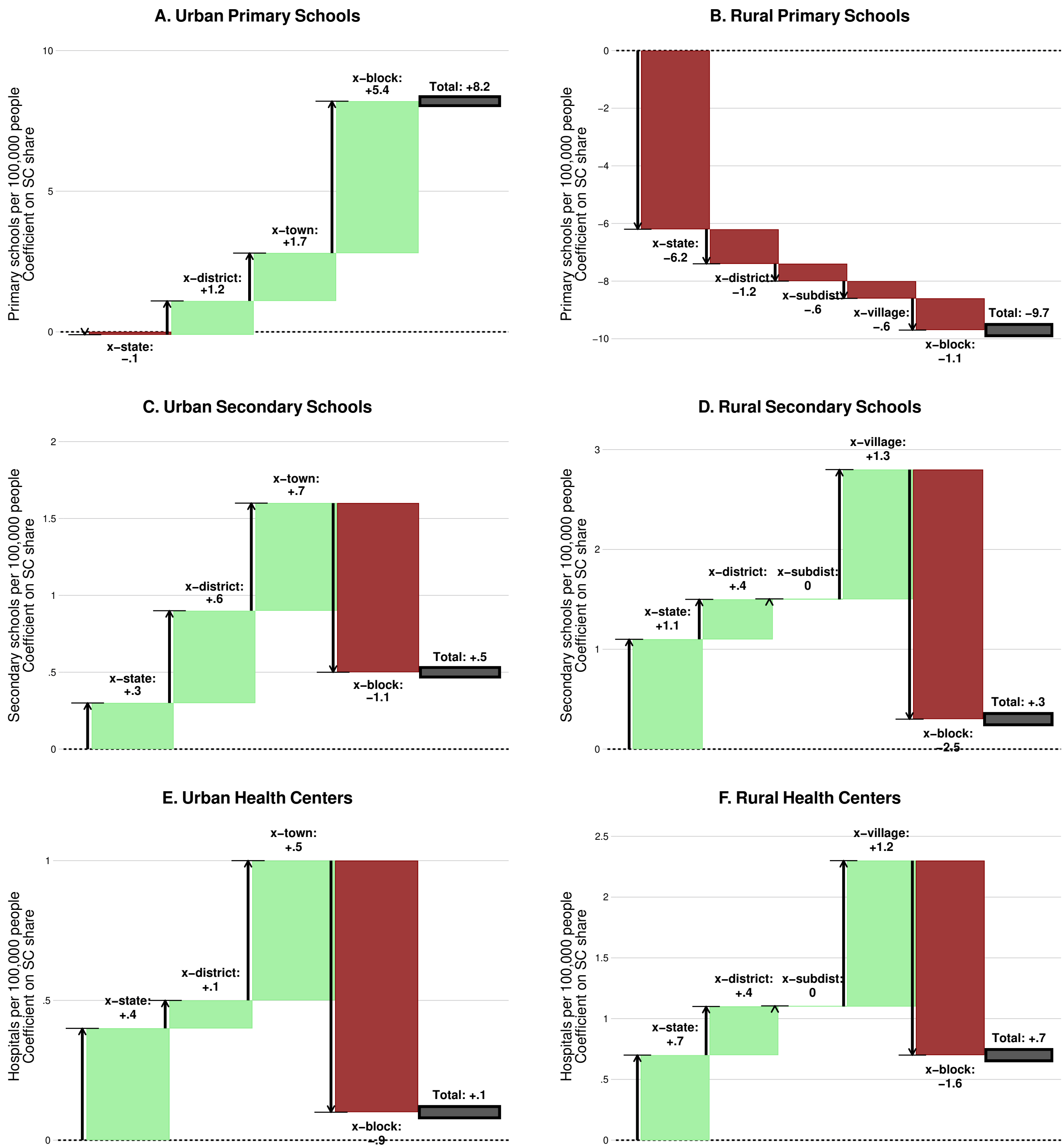
Notes: The figure shows binscatter plots of the percentage of neighborhoods that have a secondary school at a given level of SC/Muslim share. Each point represents the mean of 25,000 urban or 50,000 rural neighborhoods with a given marginalized group share. Sources: Economic Census 2013, SECC 2012.

Figure 6
Disparities in Public Facilities as a
Function of Neighborhood Muslim Share



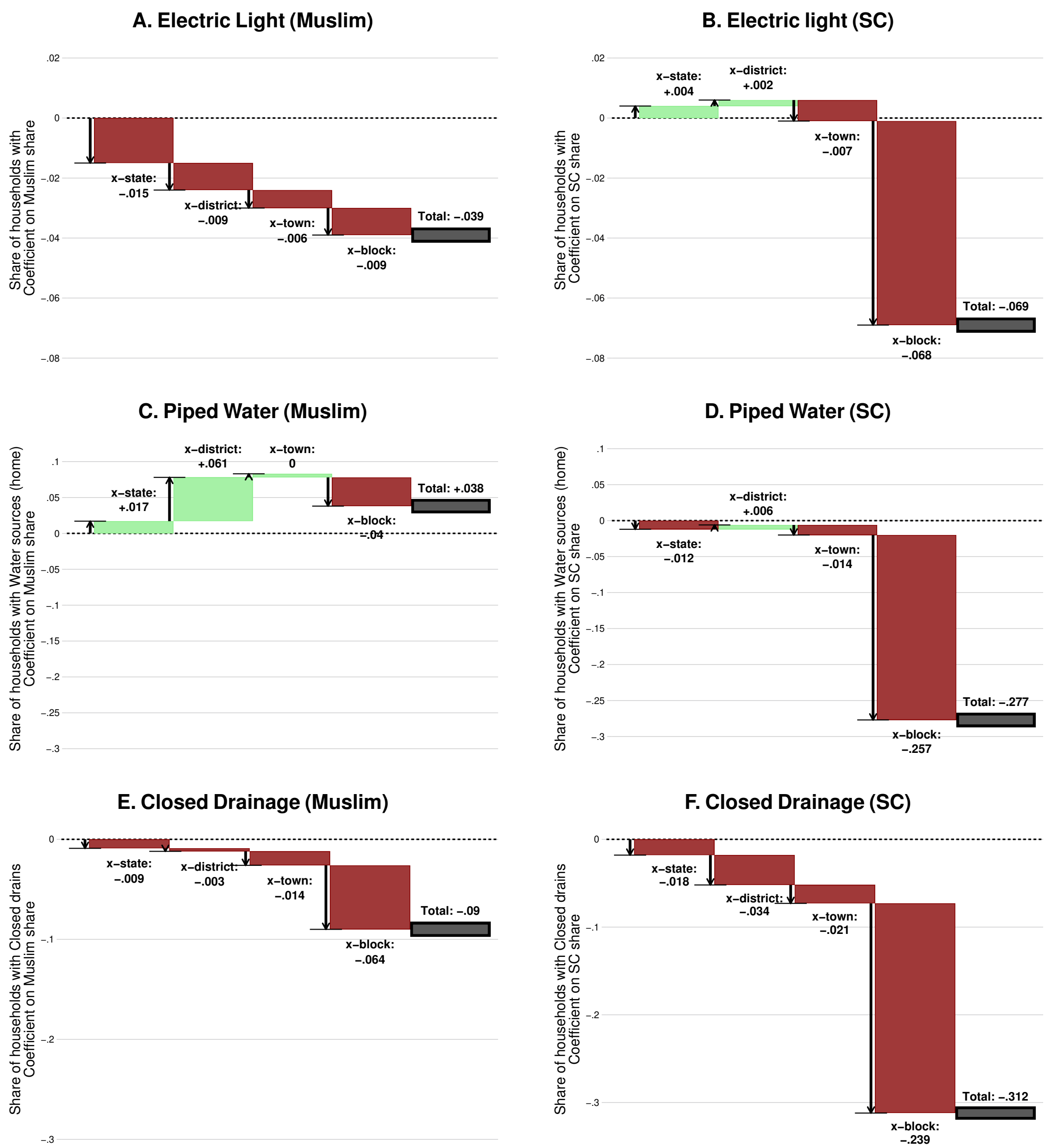
Notes: The figure describes the cross-neighborhood relationship between a neighborhood's Muslim share and a neighborhood's access to public facilities: primary and secondary schools, and health centers. The dark gray box shows the coefficient of a regression of a public facility indicator on the Muslim share in the full sample. A negative value implies that Muslim neighborhoods have fewer public facilities on average. The boxes to the left decompose that average effect into the effect arising at the cross-state, cross-district, cross-town/village, and cross-block levels. The outcome is the number of facilities per 100,000 people. The mean of this variable in rural areas is 74 for primary schools, 15 for secondary, and 12 for health centers. In urban areas, the means are respectively 15, 5, and 5. Sources: Economic Census 2013, SECC 2012.

Figure 7
Disparities in Public Facilities as a
Function of Neighborhood Scheduled Caste Share



Notes: The figure describes the cross-neighborhood relationship between a neighborhood's Scheduled Caste share and a neighborhood's access to public facilities: primary and secondary schools, and health centers. The dark gray box shows the coefficient of a regression of a public facility indicator on the Scheduled Caste share in the full sample. A negative value implies that Scheduled Caste neighborhoods have fewer public facilities on average. The boxes to the left decompose that average effect into the effect arising at the cross-state, cross-district, cross-town/village, and cross-block levels. The outcome is the number of facilities per 100,000 people. The mean of this variable in rural areas is 74 for primary schools, 15 for secondary, and 12 for health centers. In urban areas, the means are respectively 15, 5, and 5. Source: Economic Census 2013, SECC 2012.

Figure 8
Disparities in Urban Infrastructure Access
as a Function of Neighborhood Marginalized Group Share



Notes: The figure describes the cross-neighborhood relationship between a neighborhood's marginalized group share (SC or Muslim) and a neighborhood's access to public infrastructure. The sample is entirely urban. Each infrastructure measure is the share of people in a neighborhood who have access to that type of infrastructure. The dark gray box shows the coefficient of a regression of the infrastructure measure on the marginalized group share. This is the average disadvantage on this infrastructure service in marginalized group neighborhoods. The boxes to the left decompose that average effect into the effect arising at the cross-state, cross-district, cross-subdistrict, cross-town/village, and cross-block levels. The mean of the outcome variables are 0.95 for electric lighting, 0.73 for piped water and 0.56 for closed drainage. Source: SECC 2012.

Table 1
Neighborhood Summary Statistics

	Urban	Rural
Total Population	483 (165)	512 (170)
Scheduled Castes Population	56 (100)	86 (128)
Muslim Population	81 (124)	71 (117)
Scheduled Castes (Share)	0.11 (0.19)	0.17 (0.23)
Muslim (Share)	0.16 (0.23)	0.13 (0.20)
Has Public Primary School	0.07 (0.25)	0.33 (0.47)
Has Public Secondary School	0.02 (0.15)	0.07 (0.25)
Has Public Health Facility	0.02 (0.15)	0.06 (0.23)
Has Private Primary School	0.14 (0.34)	0.18 (0.38)
Has Private Secondary School	0.08 (0.27)	0.05 (0.22)
Has Private Health Facility	0.30 (0.46)	0.13 (0.33)
HH Has Closed Drains	0.56 (0.44)	NA
HH Has Electricity	0.95 (0.14)	NA
HH Has Water Source at Home	0.73 (0.34)	NA
Consumption Per Capita (SC)	30965 (17422)	16173 (8557)
Consumption Per Capita (Muslim)	27794 (14139)	15259 (7926)
Consumption Per Capita (Other)	31904 (12836)	17889 (6799)
Observations (Total)	400534	1108313

Notes: Standard deviations are in parentheses. The table shows average statistics at the enumeration block level for the analysis sample, separately for urban and rural areas. Semi-private goods (such as closed drains) are not measured in the SECC for rural areas. Consumption is measured in Indian Rupees for month. Sources: SECC (2012), Economic Census (2013).

Table 2

Sample Representativeness for Towns and Rural Subdistricts

	Towns		Subdistricts (Rural)	
	Our Sample	India (full)	Our Sample	India (full)
(Log) Population	10.31	9.87	11.51	11.39
	(1.08)	(1.03)	(0.98)	(1.20)
(Log) Area	2.33	2.00	10.34	10.25
	(1.09)	(1.10)	(0.92)	(1.08)
Scheduled Castes (Share)	0.14	0.15	0.16	0.16
	(0.09)	(0.11)	(0.10)	(0.11)
Muslim (Share)	0.18	0.19	0.09	0.09
	(0.19)	(0.22)	(0.16)	(0.16)
Town Origin Year	1947	1969		
	(42)	(43)		
Primary Schools per 100k	65.70	59.79	122.12	126.18
	(59.35)	(49.12)	(70.44)	(80.61)
Middle Schools per 100k	40.19	34.47	49.90	50.99
	(39.53)	(35.05)	(30.80)	(36.35)
Secondary Schools per 100k	22.83	20.67	19.48	19.44
	(21.66)	(21.36)	(14.91)	(15.15)
Hospitals per 100k	3.33	2.87	0.90	0.86
	(5.16)	(5.36)	(2.77)	(3.47)
Isolation Index (SC)	0.43		0.48	
	(0.13)		(0.11)	
Isolation Index (Muslim)	0.49		0.45	
	(0.20)		(0.23)	
Dissimilarity Index (SC)	0.59		0.58	
	(0.11)		(0.10)	
Dissimilarity Index (Muslim)	0.52		0.49	
	(0.14)		(0.15)	
Total Population	196,601,472	385,411,180	571,127,176	834,030,262
Observations	3504	7058	4759	5847

Notes: The table shows summary statistics at the town level (Columns 1-2) and subdistrict level (Columns 3-4) for key variables, comparing our sample (based on SECC 2012) and the all-India 2011 Population Census. The subdistrict data consists of the set of all villages in each subdistrict. Schools and health centers are measured per 100,000 people. Dissimilarity and isolation are weighted by the subdistrict/town marginalized group population. All other variables are unweighted. We do not show the SECC public infrastructure measures (e.g. electricity), because we do not observe these out of sample or in rural areas. Standard errors are in parentheses.

Table 3

Changes in Urban Scheduled Caste Segregation Over Time

A. Dissimilarity Index

Repr. Weights	Dissimilarity (2001)	Dissimilarity (2011)	Change	N
Yes	0.612	0.603	-0.016	1569
	(0.006)	(0.009)	(0.007)	
No	0.616	0.608	-0.016	1569
	(0.009)	(0.013)	(0.011)	

B. Isolation Index

Repr. Weights	Isolation (2001)	Isolation (2011)	Change	N
Yes	0.431	0.418	-0.017	1569
	(0.008)	(0.008)	(0.009)	
No	0.413	0.404	-0.015	1569
	(0.011)	(0.012)	(0.013)	

Notes: The table shows changes in Scheduled Caste segregation in a sample of 1569 towns. Segregation in 2001 is measured using the town enumeration block tables from the Census District Handbooks. Segregation in 2011 is measured using caste identifiers in the 2012 SECC. Enumeration block level data on Muslim populations is not included in the district handbooks. All samples are weighted by the Scheduled Caste town population in the given year; the changes are weighted by SC population 2001, and thus may not correspond to the exact difference between the 2001 and 2011 estimates displayed. The representation weights additionally weight towns by population, SC share, and literacy rate, to make the sample representative of the full set of cities and towns in India.

Table 4

Neighborhood-level Public Facilities
vs Marginalized Group Share

A. Urban Neighborhoods

	(1)	(2)	(3)	(4)	(5)	(6)
	Primary School		Secondary School		Health Facility	
	Indicator	Log Emp	Indicator	Log Emp	Indicator	Log Emp
SC Share	0.028***	0.033***	-0.005***	-0.020***	-0.004***	-0.005*
	(0.002)	(0.004)	(0.001)	(0.004)	(0.001)	(0.003)
Muslim Share	-0.004**	-0.010***	-0.010***	-0.031***	-0.009***	-0.021***
	(0.002)	(0.004)	(0.001)	(0.003)	(0.001)	(0.002)
Observations	357975	357975	357975	357975	357975	357975
Mean of Dependent Variable	0.07	0.12	0.02	0.06	0.02	0.04
Town FE	Yes	Yes	Yes	Yes	Yes	Yes

B. Rural Neighborhoods

	(1)	(2)	(3)	(4)	(5)	(6)
	Primary School		Secondary School		Health Facility	
	Indicator	Log Emp	Indicator	Log Emp	Indicator	Log Emp
SC Share	-0.006***	-0.017***	-0.007***	-0.016***	-0.002*	-0.003*
	(0.002)	(0.004)	(0.001)	(0.003)	(0.001)	(0.002)
Muslim Share	-0.085***	-0.142***	-0.021***	-0.040***	-0.014***	-0.016***
	(0.003)	(0.005)	(0.001)	(0.003)	(0.001)	(0.002)
Observations	978635	978635	978635	978635	978635	978635
Mean of Dependent Variable	0.33	0.54	0.07	0.15	0.06	0.08
Sub-district FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows results from neighborhood-level regressions of public facility presence on marginalized group share, for towns and rural subdistricts. Public facilities are measured either as an indicator for facility presence, or $log(employment+1)$ in the given type of facility. All regressions control for log neighborhood population and are clustered at the town (Panel A) or subdistrict (Panel B) levels.

Table 5
Correlates of Urban Segregation

	(1) Muslim Dissimilarity	(2) Muslim Isolation	(3) SC Dissimilarity	(4) SC Isolation
(Log) City Population	0.035***	0.032***	0.007*	0.019***
	(0.004)	(0.003)	(0.004)	(0.004)
City Growth Rate	1.005***	0.597***	-0.556	-0.959***
	(0.322)	(0.226)	(0.341)	(0.302)
Muslim (Share)	0.020	1.159***	-0.001	-0.053*
	(0.030)	(0.025)	(0.027)	(0.029)
Scheduled Castes (Share)	-0.046	-0.181***	-0.400***	1.157***
	(0.048)	(0.032)	(0.047)	(0.045)
City Origin Year ('00s)	-0.000*	-0.000	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
(Log) Per-capita Consumption	-0.123***	-0.107***	-0.063***	-0.084***
	(0.018)	(0.014)	(0.016)	(0.015)
City Upward Mobility	-0.005***	-0.002***	-0.001**	0.000
	(0.001)	(0.000)	(0.001)	(0.001)
Any Violent Event	0.029***	0.023***	0.029***	0.016
	(0.009)	(0.008)	(0.010)	(0.011)
City Consumption Gini	-0.092	-0.022	-0.431***	-0.408***
	(0.064)	(0.041)	(0.058)	(0.057)
Rural Land Gini	0.150***	0.006	0.107**	0.020
	(0.043)	(0.032)	(0.043)	(0.043)
Observations	1308	1308	1308	1308

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table shows estimates from city-level regressions of urban segregation measures on a set of city characteristics. The sample size is limited by the covariates for upward mobility and the urban gini. The larger-sample regression with these measures excluded is in Appendix Table A.3.

Table 6
Neighborhood-level Private Facilities
vs. Marginalized Group Share

A. Urban Neighborhoods

	(1)	(2)	(3)	(4)	(5)	(6)
	Primary School		Secondary School		Health Facility	
	Indicator	Log Emp	Indicator	Log Emp	Indicator	Log Emp
SC Share	-0.075***	-0.172***	-0.062***	-0.164***	-0.232***	-0.481***
	(0.003)	(0.006)	(0.002)	(0.006)	(0.004)	(0.007)
Muslim Share	-0.037***	-0.106***	-0.055***	-0.154***	-0.093***	-0.247***
	(0.003)	(0.006)	(0.002)	(0.006)	(0.004)	(0.007)
Observations	357975	357975	357975	357975	357975	357975
Mean of Dependent Variable	0.14	0.27	0.08	0.20	0.30	0.49
Town FE	Yes	Yes	Yes	Yes	Yes	Yes

B. Rural Neighborhoods

	(1)	(2)	(3)	(4)	(5)	(6)
	Primary School		Secondary School		Health Facility	
	Indicator	Log Emp	Indicator	Log Emp	Indicator	Log Emp
SC Share	-0.019***	-0.044***	-0.013***	-0.029***	-0.044***	-0.056***
	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)
Muslim Share	0.016***	0.014***	-0.004***	-0.011***	0.028***	0.039***
	(0.002)	(0.004)	(0.001)	(0.003)	(0.002)	(0.003)
Observations	978635	978635	978635	978635	978635	978635
Mean of Dependent Variable	0.18	0.29	0.05	0.10	0.13	0.14
Sub-district FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows results from neighborhood-level regressions of *private* facility presence on marginalized group share, for towns and rural subdistricts. Private facilities are measured either as an indicator for facility presence, or $log(employment+1)$ in the given type of facility. All regressions control for log neighborhood population and are clustered at the town (Panel A) and subdistrict (Panel B) levels.

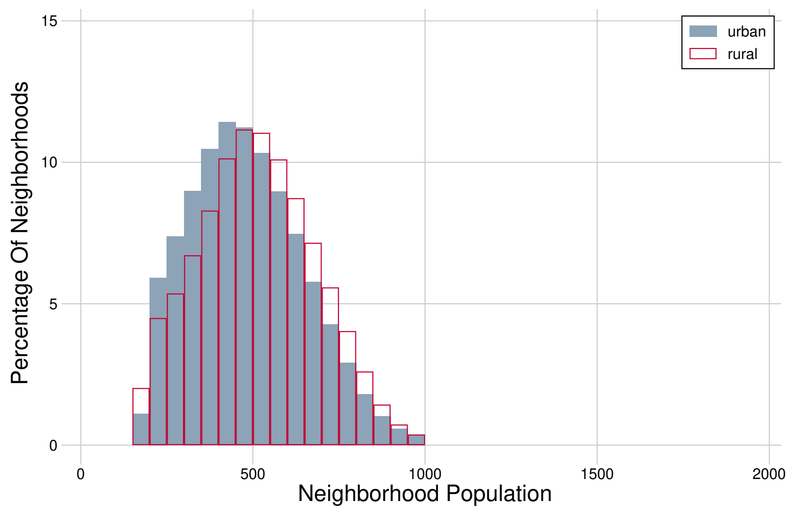
Table 7
Neighborhood-level Urban Infrastructure Services
vs Marginalized Group Share

	Closed Drains	Water Source (Home)	Light Source (Home)
SC Share	-0.258***	-0.285***	-0.069***
	(0.003)	(0.003)	(0.001)
Muslim Share	-0.099***	-0.082***	-0.019***
	(0.003)	(0.002)	(0.001)
Observations	388560	395244	389390
Mean of Dependent Variable	0.56	0.73	0.95
Town FE	Yes	Yes	Yes

Notes: The table shows results from neighborhood-level regressions of neighborhood-level infrastructure presence on marginalized group share. Results are only for cities; the given infrastructure is not measured in the rural data. Infrastructure is measured as the share of households in a neighborhood who have access to the service in question; in practice, this share is almost always very close to zero or one. All regressions control for log neighborhood population and are clustered at the town level.

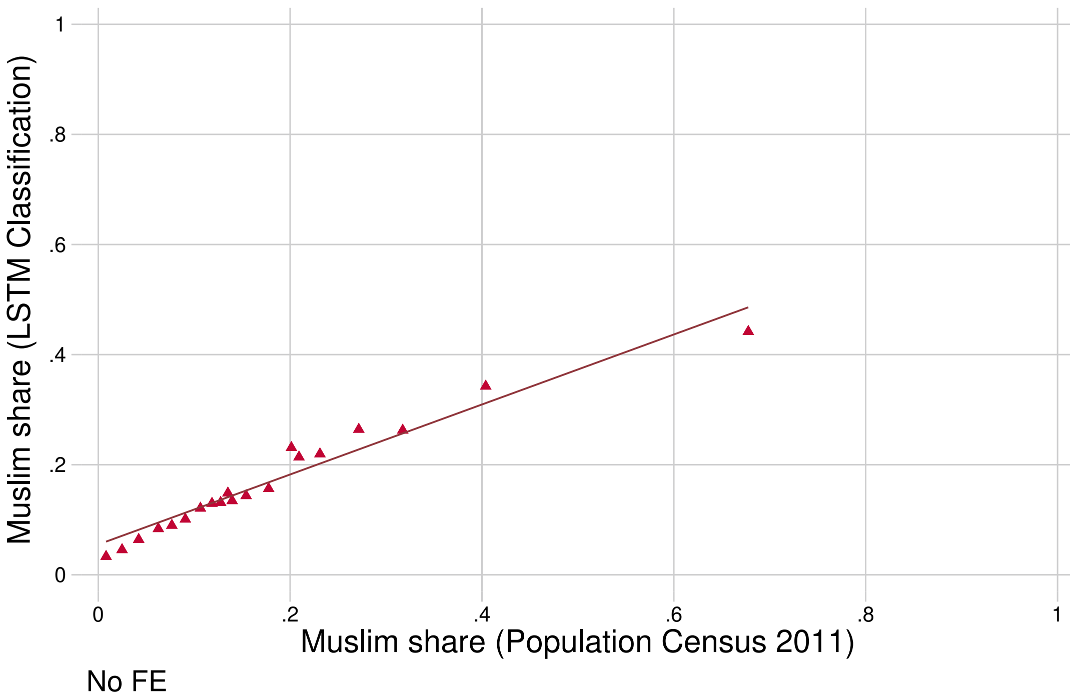
A Appendix: Additional Figures and Tables

Figure A.1
Neighborhood Population Distributions



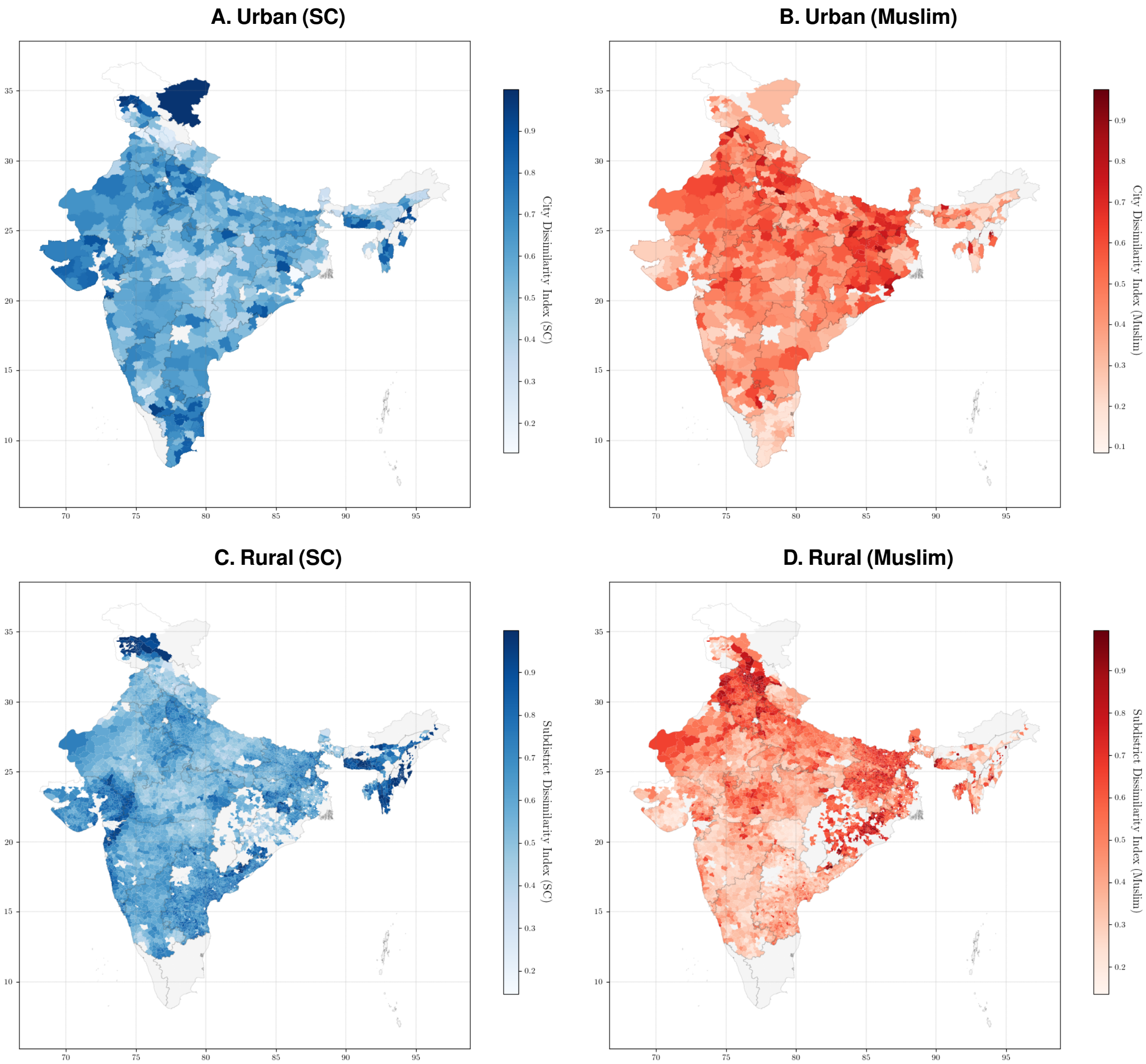
Notes: The figure shows the sample distribution of populations for neighborhoods in urban and rural areas used in our main results. Neighborhoods are excluded from the sample if they have fewer than 150 people or more than 1000.

Figure A.2
Validation of Muslim Name Classification:
Subdistrict Muslim Share in SECC vs Population Census



Notes: The figure shows a binned scatterplot of subdistrict-level Muslim shares using our classifier of SECC names, plotted against the subdistrict Muslim share recorded in the 2011 Population Census.

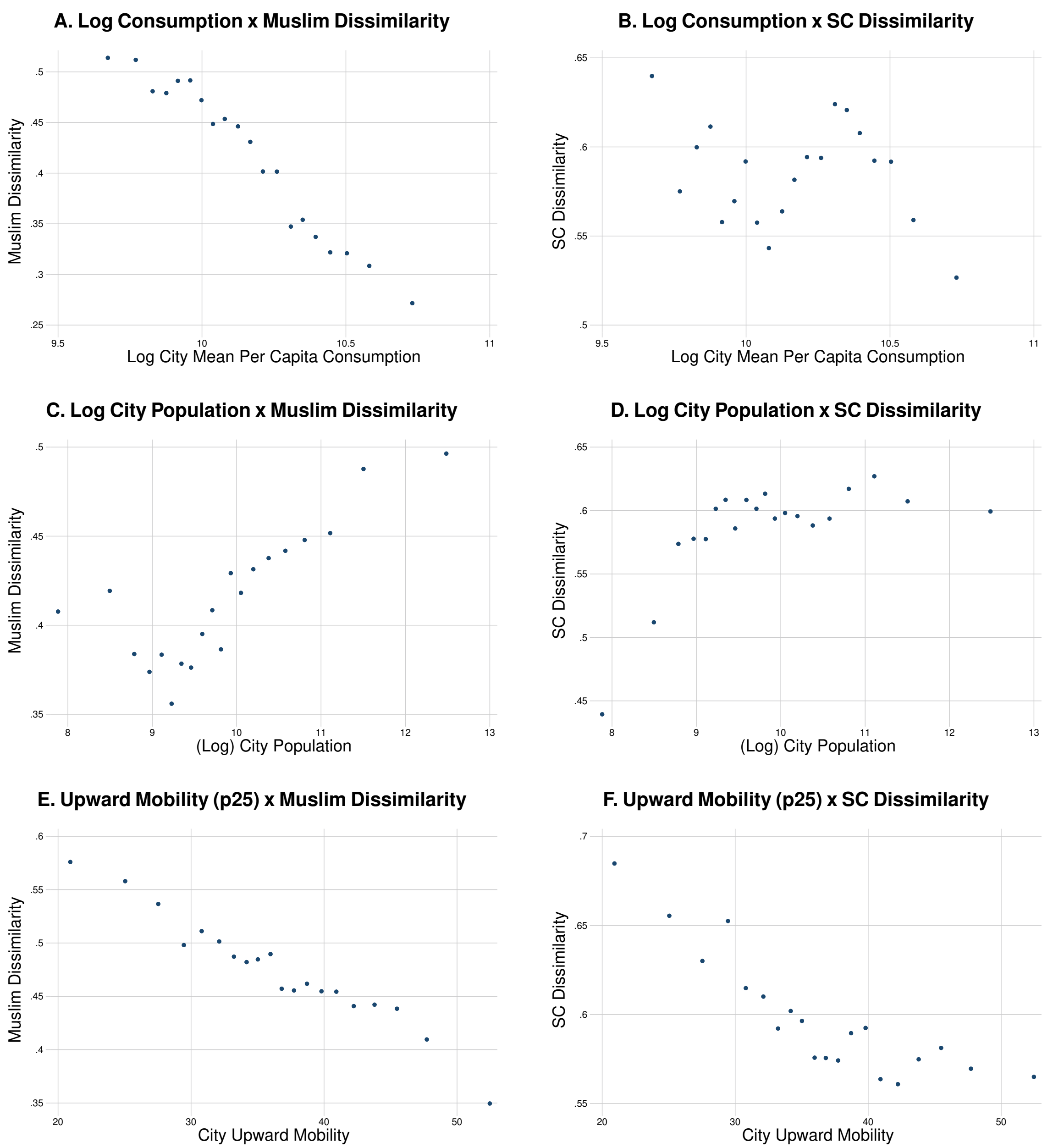
Figure A.3
Segregation Maps



Notes: The maps show the distribution of Scheduled Caste and Muslim segregation across rural and urban India. The town and subdistrict-level measures are aggregated to the district level for better visibility. For each district, the map shows the population-weighted mean of dissimilarity of locations in that district. Source: SECC 2012.

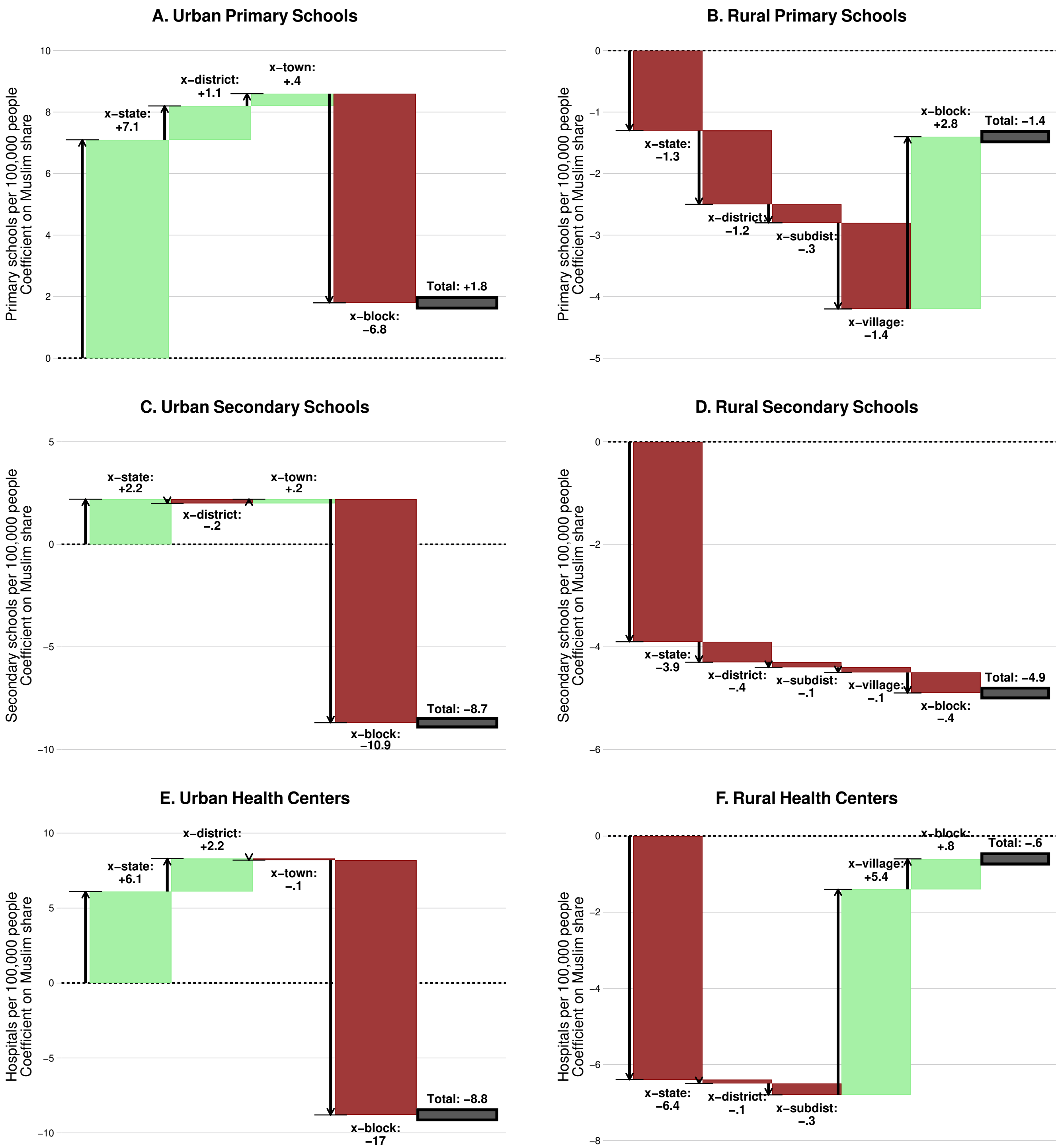
Figure A.4

City-Level Bivariate Correlates of Segregation



Notes: Each panel shows the bivariate relationship between a segregation measure and a selected socioeconomic indicator at the city level. Upward mobility is the education percentile of a child born in the bottom half of the parent education distribution (Asher et al., 2024). Additional covariates and can be viewed at devdatalab.org/segregation/correlates.html, along with these same graphs using the isolation index.

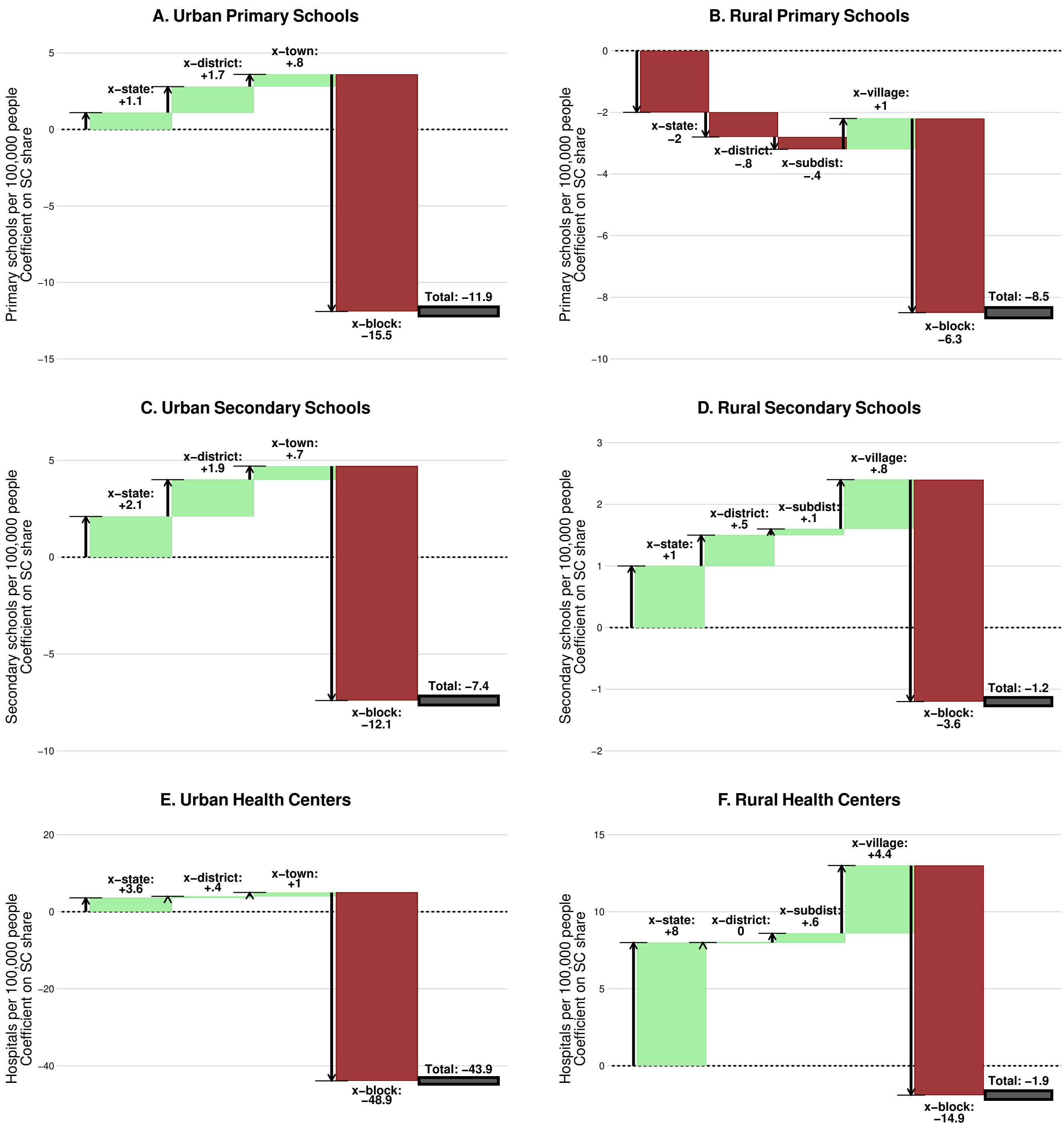
Figure A.5
Disparities in Private Facilities as a
Function of Muslim Share



Notes: The figure describes the cross-neighborhood relationship between a neighborhood's Muslim share and a neighborhood's access to private facilities: primary and secondary schools, and health centers. The dark gray box shows the coefficient of a regression of a private facility indicator on the Muslim share. This is the national advantage or disadvantage in access to the given facility in Muslim neighborhoods. The boxes to the left decompose that average effect into the effect arising at the cross-state, cross-district, cross-town/village, and cross-block levels. The outcome is the number of facilities per 100,000 people. The mean of this variable in rural areas is 38 for primary schools, 10 for secondary, and 26 for health centers. In urban areas, the means are respectively 31, 19, and 71. Sources: Economic Census 2013, SECC 2012.

Figure A.6

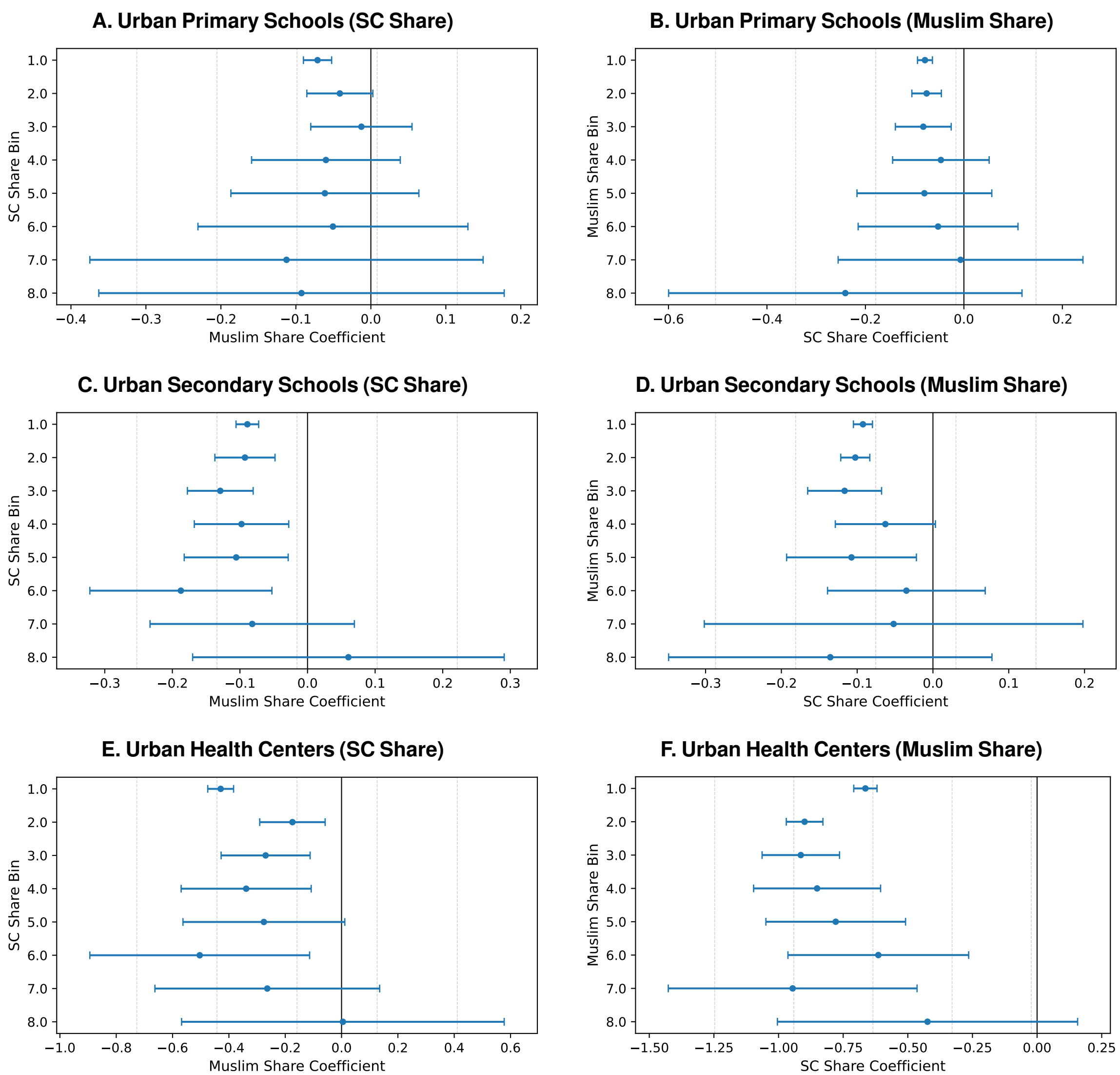
Disparities in Private Facilities as a
Function of Neighborhood Scheduled Caste Share



Notes: The figure describes the cross-neighborhood relationship between a neighborhood's Scheduled Caste share and a neighborhood's access to private facilities: primary and secondary schools, and health centers. The dark gray box shows the coefficient of a regression of a private facility indicator on the Scheduled Caste share. This is the national advantage or disadvantage in access to the given facility in Scheduled Caste neighborhoods. The boxes to the left decompose that average effect into the effect arising at the cross-state, cross-district, cross-town/village, and cross-block levels. The outcome is the number of facilities per 100,000 people. The mean of this variable in rural areas is 38 for primary schools, 10 for secondary, and 26 for health centers. In urban areas, the means are respectively 31, 19, and 71. Source: Economic Census 2013, SECC 2012.

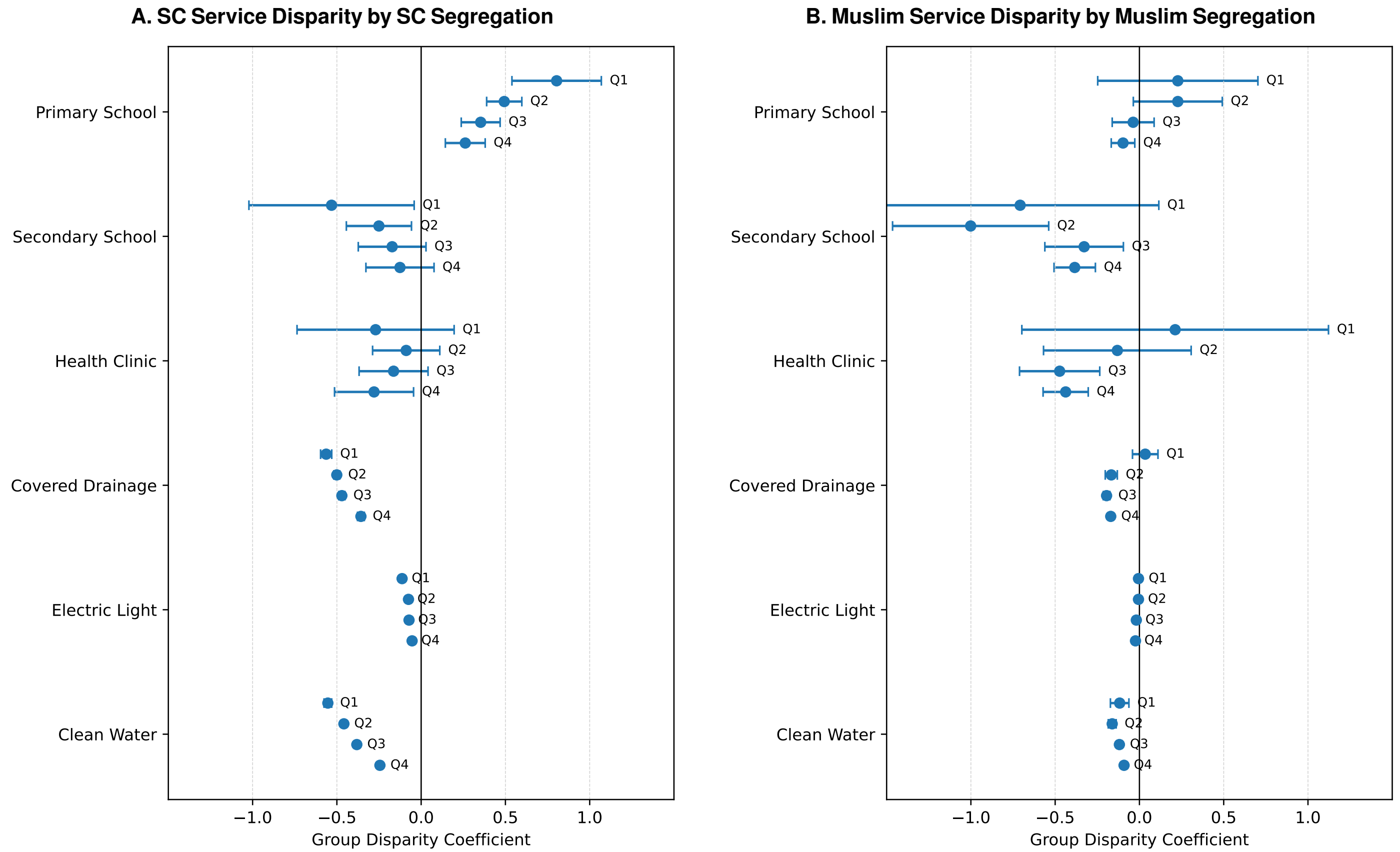
Figure A.7

Neighborhood Disparities in Public Facilities:
Social Group Interactions



Notes: The figure shows how estimates of one marginalized social group's neighborhood disadvantage change with the population share of the other marginalized social group. Each graph estimate shows the coefficient on the group share, split by the population share of the other group in 10 percentage point bins. We omit bins 9 and 10, since they leave too little variation in the first group share. For example, for Panel A, we estimate $\mathrm{SEC}_{nc} = \beta_0 \sum_{i=1}^{10} \delta_i \mathbf{1}\{\mathrm{MuslimBin}_{inc} = i\} + \sum_{i=1}^{10} \gamma_i \mathrm{SCShare}_{inc} \cdot \mathbf{1}\{\mathrm{MuslimBin}_{inc} = i\}$, where SEC is an indicator for secondary school presence in neighborhood n and city c, and MuslimBin indicates the Muslim neighborhood population share. The estimates tell us whether the relationship between SC share and secondary school access is different in neighborhoods with many and few Muslims.

Figure A.8
Neighborhood Disparities in Public Facilities:
By Quartile of Segregation



Notes: The figure shows estimates from a neighborhood-level regression of public service availability on the marginalized group share of the neighborhood, with town fixed effects. The regression is identical to those presented in Table 4, but split by city dissimilarity quartile. For SC disparities (Panel A), we split on SC dissimilarity, and for Muslim disparities (Panel B), we split on Muslim dissimilarity. The coefficients are scaled by the mean value of the public service in question.

Table A.1

Segregation Indices with Town Reweighting

	Unweighted Sample	Entropy-Weighted Sample	Reference
Log (population)	10.24	9.77	9.77
Log (area)	2.29	1.94	1.94
Scheduled Caste Share	0.14	0.15	0.15
Muslim Share	0.17	0.18	0.18
Town Origin Year	1948.10	1969.35	1969.40
Dissimilarity (SC)	0.59	0.60	
Dissimilarity (Muslim)	0.52	0.52	
Isolation Index (SC)	0.43	0.44	
Isolation Index (Muslim)	0.49	0.50	
Observations	3612	3612	7528

Notes: The table shows dissimilarity and isolation indices where towns have been reweighted to match the full sample of towns. Column 1 shows the primary sample of the paper. Column 3 shows summary statistics from the complete set of towns in the 2011 Population Census. Column 2 shows our sample means after entropy-weighting to rebalance our sample on these variables to match the Population Census. The dissimilarity and isolation indices are further weighted by the marginalized group's own population, as described in Section 4.

Table A.2

Sources for International Segregation Estimates

Source	# Cities	Country	Description	Year	Group	Weighted
Logan (2022)	404	US	Largest cities	2020	White vs. Black	Yes
	4137	US	All towns and cities	2020	White vs. Black	Yes
Cutler, Glaeser, Vigdor (1999)	211	US	Largest cities	1970	White vs. Black	Yes
Valente & Berry (2020)	35	Brazil	Brazil	1980	White vs. Black	No
	40	Brazil	Brazil	2010	White vs. Black	No
Simpson (2007)	62	England	England & Wales	1991	White vs. Non-White	No
	62	England	England & Wales	2001	White vs. Non-White	No
Andersen (2015)	1	Denmark	Copenhagen	2008	Immigrant vs. Native	No
	1	Finland	Helsinki	2008	Immigrant vs. Native	No
	1	Norway	Oslo	2008	Immigrant vs. Native	No
	1	Sweden	Stockholm	2008	Immigrant vs. Native	No
Benassi et al (2020)	8	Spain	Spain	2011	Immigrant vs. Native	No
	8	Italy	Italy	2011	Immigrant vs. Native	No
Verdugo & Toma (2018)	unknown	France	Major metro areas	1982	Non-European immigrants	Yes
	unknown	France	Major metro areas	2012	Non-European immigrants	Yes

Notes: The table shows the data sources and details for the international segregation estimates shown in Figure 2.

Table A.3
Multivariate Correlates of Urban Segregation

	(1) Muslim Dissimilarity	(2) Muslim Isolation	(3) SC Dissimilarity	(4) SC Isolation
(Log) City Population	0.019***	0.025***	0.014***	0.024***
	(0.003)	(0.002)	(0.003)	(0.003)
City Growth Rate	1.435***	0.490***	-1.227***	-1.373***
	(0.232)	(0.117)	(0.267)	(0.214)
Muslim (Share)	-0.039	1.213***	0.125***	0.069***
	(0.025)	(0.017)	(0.028)	(0.024)
Scheduled Castes (Share)	-0.070*	-0.119***	-0.104**	1.390***
	(0.038)	(0.023)	(0.050)	(0.035)
City Origin Year ('00s)	-0.042***	-0.012**	-0.040***	-0.025***
	(0.008)	(0.005)	(0.009)	(0.008)
(Log) Per-capita Consumption	-0.246***	-0.128***	-0.060***	-0.005
	(0.012)	(0.008)	(0.013)	(0.010)
Any Violent Event	0.048***	0.028***	0.003	-0.007
	(0.009)	(0.007)	(0.009)	(0.009)
Rural Land Gini	-0.152***	-0.027	0.378***	0.227***
	(0.037)	(0.022)	(0.042)	(0.034)
Observations	2792	2792	2792	2792

Standard errors in parentheses

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Notes: The table shows estimates from city-level regressions of urban segregation measures on a set of city characteristics. The regression is identical to that in Table 5, but upward mobility and the city Gini coefficient are excluded, to enable a larger sample.

Table A.4
Neighborhood-level Public Facilities vs.
Marginalized Group Share: Controlling/Excluding Slums

	(1)	(2)	(3)	(4)	(5)	(6)
	Slum Controls			No Slum		
	Primary School	Secondary School	Health Facility	Primary School	Secondary School	Health Facility
SC Share	0.028***	-0.005***	-0.004**	0.028***	-0.005***	-0.004**
	0.002	0.001	0.001	0.003	0.001	0.001
Muslim Share	-0.004	-0.010***	-0.009***	-0.004	-0.010***	-0.010***
	0.002	0.001	0.001	0.002	0.001	0.001
Observations	356271	356271	356271	308216	308216	308216
R^2	0.067	0.024	0.022	0.064	0.023	0.022
Town FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows results from a neighborhood-level regression of public facilities on the neighborhood marginalized group share, for towns, analogous to Table 4. Columns 1–3 show results with a control for whether or not the neighborhood is in a slum. Columns 4–6 show results for the set of urban neighborhoods that are not classified as slums.

Table A.5

Neighborhood-level Public Services and Marginalized Group Shares:
Estimates Controlling for Neighborhood Consumption

	(1)	(2)	(3)	(4)	(5)	(6)
	Primary		Secondary		Health	
SC Share	0.028***	0.019***	-0.005***	-0.002	-0.004***	0.002
	0.003	0.003	0.001	0.001	0.001	0.001
Muslim Share	-0.004*	-0.012***	-0.010***	-0.007***	-0.009***	-0.004***
	0.002	0.002	0.002	0.002	0.001	0.001
Consumption Control		X		X		X
Mean of Dependent Variable	0.07	0.07	0.02	0.02	0.02	0.02

	(7)	(8)	(9)	(10)	(11)	(12)
	Piped Water		Closed Drainage		Electric Light	
SC Share	-0.258***	-0.145***	-0.285***	-0.173***	-0.069***	-0.047***
	0.013	0.012	0.018	0.016	0.007	0.006
Muslim Share	-0.099***	0.008	-0.082***	0.017*	-0.019***	0.001
	0.013	0.010	0.014	0.009	0.003	0.002
Consumption Control		X		X		X
Mean of Dependent Variable	0.56	0.56	0.73	0.73	0.95	0.95

Notes: The table shows results from neighborhood-level regressions of urban public service availability on neighborhood marginalized group shares. Odd-numbered columns replicate results from Tables 4A and 7. Even-numbered columns show estimates from the same regression, with a control for average neighborhood per-capita consumption.

B Appendix: Data Sources

B.1 Data Sources

B.1.1 More Detail on SECC Collection and Use

The Socioeconomic and Caste Census (SECC) was a rapid poverty survey covering every household and individual in India. It collected a short list of assets, and demographic and education information. The data were made publicly available by the Government of India in fragmented form, spread across more than two million PDF files covering 825 million rural and 400 million urban individuals. The data were posted publicly to enable individuals to look up their entries and contest their content. The data were posted intermittently over a period of about two years, with each given file posted for several months at a time. We scraped the data as they were posted; our coverage is incomplete either because the server was down while some PDFs were posted, or some may not have been posted at all.

We developed a data extraction pipeline in Python to convert the PDF content into tables, and transliterated contents from regional languages and text into Latin characters and English words. We then matched each household record to 2011 Census urban and rural location codes using string cleaning and fuzzy name matching. The merge rate was high because the location names in the SECC were drawn directly from the Population Census. The enumeration block identifiers listed in each PDF has direct matches to enumeration block identifiers in the 2013 Economic Census, enabling us to match neighborhood demographics to neighborhood public facilities.

Consumption was not directly measured in the SECC, but we generated small area estimates of household per capita consumption using the SECC asset list and IHDS-II (2011–12), following Elbers et al. (2003). Rural and urban consumption distributions were broadly similar to direct survey measures from the same period (Asher et al., 2021).

B.1.2 Classifying Muslim Names

The SECC surveyed individual caste and religion, but religion was not released in the public data.[46] We therefore classify individuals as Muslims or non-Muslims using their first and last names, which were posted in the public data.

We developed a bidirectional LSTM neural network to classify individual names (combined first and last) as either Muslim or non-Muslim (Ash et al., 2025), using labeled training data obtained from the National Railway Exam (N = 1.4 million). The classifier outperformed a traditional fuzzy matching technique because it can effectively distinguish names with high string similarity, such as Khan (typically Muslim) vs. Khanna (typically non-Muslim).

SECC names originated in many different scripts (e.g. Kannada, Assamese, Hindu, Gujarati); we transliterated these to Latin, and normalized formatting before applying the classifier. We found that the same approach was not sufficiently accurate to classify individuals as SC/non-SC, ST/non-ST, or into jati categories.

We evaluated classifier performance with hold-out test sets, as well as on lists of names from SECC that were manually classified as characteristically Muslim or not by individuals with local knowledge. The Muslim classifier achieved 0.98 balanced accuracy and an F1 of 0.99, indicating extremely high reliability. This works well because Muslim names are very characteristic; caste names have more variation and cross categories more often. If the classifier predicted a Muslim probability between 0.35 and 0.65, we left the entry unlabeled; about

[46] Subcaste (also called jati) was also recorded but not released. The only caste identifier are broad indicators for Scheduled Caste or Scheduled Tribe status.

10% of individuals were unclassifiable in this way. In most case, failure to classify was not because of the name, but because the scanned text in the SECC was unreadable.

Our classification also closely predicts the subdistrict-level population share of Muslims (Appendix Figure A.2).[47] We pool Hindus with the 6% of Indians who are Jain, Christian, Sikh, or some other non-Hindu religion; we describe this group as "non-Muslims."[48]

B.1.3 Additional Data Sources

Data on Hindu-Muslim violence is from Varshney and Wilkinson (2006). It describes all Hindu-Muslim riots recorded between 1950–95 in the Times of India, India's newspaper of record at the time. We assign a riot to a town if it is described as occurred in the town or in the district in which the town is contained. We found similar results when we restricted to riots resulting in significant numbers of casualties.

U.S. segregation statistics were obtained from the Diversity and Disparities Project Logan (2021). The project harmonizes long-form census samples and ACS data into tract-level files, and provides measures of dissimilarity and isolation indices calculated in standard formats over time. For details of the specific indices used, see Section B.2.

Urban and rural inequality were both calculated from household-level SECC data. Rural land inequality was based on reported land (adding together irrigated, unirrigated, and other). In cities (where land ownership was not available), we used the small area consumption estimates described above. The urban inequality measures clearly dramatically understate urban inequality as the small area estimates are effectively top-coded for the people in the sample who have every listed SECC asset, but it captures inequality amongst the bottom 90%.

Upward mobility is the expected education level of a child born to a father in the bottom half of the education distribution. Details on its calculation are found in Asher et al. (2024).

City origin year is calculated from the Population Census, which lists city populations in every prior census going back to 1901. We define the origin year as the first year in which the city has a listed population over 5000 (the population threshold at which the Indian Census considers a location eligible for town status). Urban growth was calculated using population from the 1991 and 2011 Population Censuses.

B.1.4 Calculating Segregation in 2001 from Census District Handbooks

The Indian Census publishes District Handbooks for every district, large ($\sim$ 500–1000 page) volumes with detailed information on district demographics. Most of the district handbooks contain a 20–100 page listing of the population and Scheduled Caste population of every enumeration block in every town in the district. These are PDF tables, making data extraction somewhat more difficult. We used a heuristic search in the district handbooks to identify pages with urban enumeration block lists, and then used Google Gemini Pro 2.5 to transform the relevant PDF pages into tabular CSV files. The LLM had a much higher accuracy rate than human coders (98% accuracy against a double-entered series of pages), and (unlike human coders) was feasible to run on the complete set of handbook pages.

Our sample was contrained by the limited availability of 2001 District Handbooks in digital format. Out of 581 districts with at least one town with over 1000 people, we were able to obtain legible district handbooks for 398 districts, which had data on 2994 towns and cities. After merging these to the 2012 SECC urban sample (which has about 50% urban coverage), we had a sample of 1569 cities for which we could calculate SC segregation in both 2001 and 2011.

[47] Note that the Population Census reports the Muslim share only down to the subdistrict level.

[48] The non-Hindu, non-Muslim groups are small and we do not yet have an algorithm that can accurately classify them on the basis of names.

The District Handbooks do not contain information on block-level Muslim populations. The enumeration block boundaries are not necessarily consistent between 2001 and 2011, but this is unlikely to be a source of error, as they were constructed with exactly the same goal of providing enumerators with a list of residences that they could visit in sequence in a day or half-day of work. As such, in both periods they are likely to represent buildings, city blocks, or compact neighborhoods. The population size distribution of enumeration blocks is very similar in the two periods.

District Handbooks with similar data were produced for the 1991 and 2011 Censuses, but at this time we do not have a large sample of digital files.

B.2 Segregation Index Parameters

We describe each factor and the approach we take. We focus most on the U.S. comparison, as the U.S. is by far the most studied country in terms of segregation.

Neighborhood size: As the neighborhood grows larger, segregation measures fall mechanically.[49] Neighborhood size is determined by the census, and thus cannot be selected by the researcher. U.S. census tracts range from 1000–8000, with target population of 4000. Given the mechanical relationship between unit size, we show neighborhood size in every comparison. When we benchmark our segregation measures against the United States (and at no other place in the paper), we aggregate enumeration blocks based on their numeric identifiers to form neighborhoods of at least 4000 people.[50]

Weighting: To construct national estimates, we weight city-level segregation by the marginalized population in each city, so that measures reflect the experience of the marginalized group. We weight SC segregation by SC population, and Muslim segregation by Muslim population. The literature is mixed on weighting and sometimes does not specify whether weights are used; where possible, we use weights for international comparisons.

Town sample: In the U.S., large cities are more segregated. The sample city size threshold therefore mechanically affects national estimates. When comparing to the U.S., we use a similar city population threshold to achieve comparability. Otherwise we use the complete sample.

City boundaries: If suburbs and exurbs have group-correlated settlement patterns, segregation will be affected by the city boundaries used. In India, we use cities plus their outgrowths as defined by the Population Census; in the U.S., we use Metropolitan Statistical Areas which are the most comparable unit.

[49] To take an extreme example, if we defined a "neighborhood" as a single household, we would calculate a dissimilarity index close to 1, given the very high rates of caste and religious endogamy.

[50] In the handful of cities where we have enumeration block maps or neighborhood names, we confirmed that enumeration blocks with adjoining numbers are almost always geographically adjacent. Aggregating to 4000-person units based on block number inevitably adds noise to the neighborhood definition, which is why we use the disaggregated neighborhoods for everything except the U.S. comparison. Note that the U.S. Census defines neighborhoods according to existing informal boundaries, which are more likely to divide racial groups, and thus overstates segregation relative to an approach with arbitrary geographic units.