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‖ ⓡ

December 2023

Abstract

Urbanization in lower-income countries has the potential to cause substantial improvements in well-being, but the residential segregation of marginalized groups could reinforce inequality and limit access to new opportunities. We study residential segregation, access to public services, and economic outcomes across 1.5 million urban and rural neighborhoods for India's largest marginalized groups: Scheduled Castes (SCs) and Muslims. Levels of urban segregation in India are comparable to Black/White segregation in the United States. Within cities, public facilities and public infrastructure are systematically allocated away from neighborhoods where many Muslims and members of Scheduled Castes live. Nearly all of the regressive allocation is across neighborhoods within cities—at the most informal and least studied form of government. These inequalities are not visible in the more aggregated data typically used to study unequal service allocation. Children and young adults growing up in marginalized group neighborhoods have less schooling, even after controlling for parent education and household consumption. Unequal access to public services in India's highly segregated neighborhoods may be a significant contributor to disadvantages faced by marginalized groups.

JEL Codes: H41, J15, O15

*We are grateful for stellar research assistance from Ryu Matsuura and Alison Campion, and for helpful discussions in seminars at the Center for Policy Research, Centre de Sciences Humaines, IDinsight, Harvard Urban Development Workshop, IZA Labor Markets in South Asia Conference, NBER SI Development and Urban, Regional Disparities Seminar, Delhi Political Economy Workshop, SAIS, American University, University of Chicago, William T. Grant Foundation, and the AEA meetings. The order in which the authors names appear has been randomized using the AEA Author Randomization Tool (TaxM_T0V9IFZ).

†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, 2013; 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 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 describing settlement and segregation patterns of marginalized groups across Indian cities and villages, and we document the relationship between these settlement patterns and access to public services. The data requirements for such a task are significant, and we are unaware of any comparable comprehensive descriptive analysis of segregation and public service access in any developing country.

We focus on outcomes for Muslims and members of Scheduled Castes, often called dalits or (previously) untouchables. 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 like the district; 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 three 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. Third, we study the educational outcomes of young men and women who live in MG neighborhoods, providing suggestive evidence on the consequences of residential segregation.

To do this, we create a neighborhood-level dataset covering over 60% of India's population, the first such dataset to link neighborhood demographics with access to public services.[1] The challenge in constructing these data is that there is no systematic survey documenting public service availability across India. However, information about service availability can be inferred from India's firm and poverty censuses, which use the same small neighborhood coding scheme across the country. The Socioeconomic and Caste Census (SECC 2012) describes a short list of assets at the household level for *every household in India*, along with a household roster that records the education, occupation, and SC status for every household member. The asset list records whether urban households have piped water, electricity, and drainage; while household access may be privately purchased, neighborhood infrastructure is a precondition for these services and access within neighborhoods is largely homogeneous. Crucially, public listings of the SECC were released with respondent names; the distinctive naming patterns of Muslims allow us to predict the religious identity of household members (as Muslim or non-Muslim) with high accuracy.[2]

[1] Our analysis dataset describes 400,000 urban neighborhoods (in 3500 cities and towns) and 1.1 million rural neighborhoods (in about 400,000 villages). Subdistricts and towns that are unavailable in the administrative data are broadly similar on a wide range of variables to those in the study.

[2] 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., 2022).

We get information on public facilities from the Indian Economic Census, which records data on India's 65 million non-farm economic establishments. This census is conventionally used to study firms, but it also records data on public schools, health centers, and hospitals, making it the only large-sample data source (to our knowledge) that can identify these services at the neighborhood level. Combining these datasets, we can document individual demographics, socioeconomic outcomes, and neighborhood-level public services across every part of the country.

SCs and Muslims make up similar population shares in the country (17% and 14% respectively in 2011), but have distinct group histories. Scheduled Castes have been historically consigned to the lowest occupational rungs of society for over a thousand years, but have been targeted by decades of affirmative action policies since independence; empirical studies find positive effects for some of these programs (Gulzar et al., 2020; Asher et al., 2022).

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 of these 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] While disparities in access to public services have been documented for both groups, there is little systematic empirical work on disparities at the neighborhood level—the level at which public services are typically accessed.[4]

We present three key findings. First, Muslims and SCs have highly segregated residence patterns, comparable in magnitude to the contemporary urban segregation of Black people in the United States. Whether Muslims or SCs are more segregated depends on the measure of segregation used. SCs are more likely to experience moderate levels of segregation, while

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

[4]Because the Population Census records Scheduled Caste shares and the presence of a set of public services at the village level, the relationship between the village-level rural SC share and public services has been studied (Banerjee and Somanathan, 2007). However, village or neighborhood-level access to public services has not been studied on a national scale for Muslims, nor urban access to local services for either group.

the distribution of Muslim shares across neighborhoods is notably bimodal. As such, a greater share of Muslims live in neighborhoods that are almost entirely Muslim.[5]

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.[6] Compared with SCs, Muslims are relatively more segregated in cities than in rural areas. We marshal the limited time series data available to show that the extent of segregation is changing very slowly over time, if at all.

Second, we show that public services are systematically allocated away from 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 closed drainage. Private providers are *not* making up for the reduced service access of marginalized groups; in fact, private services also systematically locate away from MG neighborhoods, in part because these neighborhoods are poorer.[7]

The magnitude of the disparities is large. For example, compared with a 0% Muslim neighborhood, a 100% Muslim neighborhood in the same city is 10% less likely to have piped water infrastructure 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, 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

[5]For example, 26% of Muslims live in neighborhoods that are $>80\%$ Muslim, while 17% of SCs live in neighborhoods that are $>80\%$ SC.

[6]While the data do not record when these settlement patterns emerged, the historical record suggests that rural Indians have been highly endogamous, such that village settlement patterns observed today have been static for decades, if not centuries.

[7]The fact that marginalized groups live in poor neighborhoods does not explain away the public service results, since government service provision aims to be universal. Nor can they be explained by poor service provision in slums, as the results hold just as strongly in the sample of non-slum neighborhoods.

SCs have more public facilities on average, consistent with findings by Banerjee and Somanathan (2007). 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 are most systematically and substantively disadvantaged 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.

Finally, we examine the relationship between residential settlement patterns and outcomes for the next generation. We find that young people growing up in marginalized group neighborhoods have systematically worse educational outcomes. The disadvantages are substantially worse in Muslim neighborhoods than in SC neighborhoods and are economically large even after controlling for parent education and household consumption.[8] These disadvantages are experienced by members of all social groups—including non-marginalized groups—living in marginalized group neighborhoods.

These results are descriptive. Further research is needed to understand whether the disparities described here are causal effects of neighborhoods or driven by selection of marginalized groups into under-serviced neighborhoods. Equally, we do not prove that services are allocated away from marginalized group neighborhoods *because* of the people who live there, or because of some other characteristic of those neighborhoods. Our work serves as a necessary starting point for asking these questions, because these cross-neighborhood disparities have not previously been documented; even the extent of residential segregation has barely been measured, not only in India but in most developing countries. In Section 5, we discuss in detail the external evidence for and against causal interpretation of these results. But decisively disentangling the causal direction of these disparities is an important subject for future work.

[8] Rural SC neighborhoods are an exception; they have worse access to facilities but no worse educational outcomes.

Systematic analysis of access to public services at the neighborhood level 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, 30 times more coarse than the neighborhoods 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] but 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 have 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 value of leveraging high-resolution administrative data — which is available but under-used in many

[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 rarely available in Indian Census microdata.

[10] Bharathi et al. (2018) report enumeration block-level segregation based on SC status for five major cities. Bharathi et al. (2021) use similar-scale data on caste and religion to characterize segregation in urban Karnataka. Susewind (2017) measures Muslim segregation using microgeographic polling booth data in eleven cities.

[11] A concurrent study by Bharathi et al. (2022) examines the correlation between micro- and macro-segregation and access to water and sewerage, using block-level data. While we study outcomes for marginalized groups in neighborhoods where they are concentrated, Bharathi et al. (2022) study average outcomes in more and less segregated larger units (i.e. wards).

developing countries — to better understand and evaluate the performance of public programs.[12]

2 Context and Background

2.1 Scheduled Castes and Muslims in India

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 effectively compelled them to take on low-status occupations—like scavenging, emptying of 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, 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., 2022).

Muslims occupy a similar share of the population to Scheduled Castes (14% for Muslims vs. 17% for SCs). 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 general castes 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., 2022). Post-independence India has been characterized

[12] We are concurrently preparing a public version of the neighborhood dataset, which we will post with a revision of this paper.

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. Muslims have a higher share of members living in urban areas than any other major social group.

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 vast 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;

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

Doshi, 1991; Sachdev and Tillotson, 2002).[14] The ethnographic literature suggests a secular trend of increasing segregation by religion rather than by occupation, as Hindu-Muslim violence has reduced Muslim feelings of safety in mixed neighborhoods. These newly concentrated 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 difficulty getting attention from politicians or access to public services in their segregated neighborhoods.

The literature on villages also suggests a high degree of 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; Lanjouw et al., 2018).

While these patterns can be observed in many parts of the country, they are primarily documented in a qualitative literature (some of which is cited 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* villages (Banerjee and Somanathan, 2007; Bailwal and Paul, 2021), in part because the decennial Population Census records 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 simultaneously describing both public services and marginalized group shares.

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

2.3 Levels of government in India

India has a federal system of government with major powers divided between center, state, and local governments. The administrative apparatus is also decentralized, such that officials at different hierarchical levels have substantial autonomy.

There are 36 states and union territories (35 at the time of our sample), which have substantial administrative and legislative power. Public services are financed and allocated by both central and state government programs. Program implementation often lies in the hands of District Collectors, who are the top administrative officers of districts; there were 640 districts in our sample, though an additional 100 have been subdivided since then.

Local governance bodies are called panchayats in villages and municipalities in towns and cities. These bodies have elected representatives who can substantially influence 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 highest profile policies intended to close disparities between marginalized and non-marginalized groups are conceived and designed at the state and federal level, and often prescribe allocations of public services across regions. For instance, the District Primary Education Programme (Khanna, 2023) targeted funds for building schools to districts with below-median female literacy. The placement of new public facilities *within* districts, towns, and villages is rarely prescribed by these high-level policies; it is instead agreed upon through consultation with local elected leaders and bureaucrats or arbitrarily determined by on-the-ground implementers. 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

3.1 Identifying Neighborhoods

Studying neighborhood-level disparities requires data with granularity (to be able to identify neighborhoods) but also with scale (to be able to accurately measure neighborhood-level MG shares in many neighborhoods and many cities). Few of India's major sample surveys achieve this; they typically cover a small fraction of neighborhoods in any city, and too small a number of households in each sampling unit to measure MG shares, MG segregation, or disparities in MG outcomes.

To bridge this gap, we combine a set of census data sources which use the internal survey block identifiers (enumeration blocks) that were created for the administration of India's 2011 Population Census. These "neighborhoods" consist of 100–125 households each (or approximately 500 people) and describe a compact cluster of residences meant to be efficient for an enumerator to visit in a single session of work. In cities, these are typically city blocks, while in rural areas their boundaries are typically defined by grouped clusters of residences. When villages consist of fewer than 100–125 households—about half of villages—an enumeration block is a single village. Urban enumeration blocks are thus uniformly around 100–125 households, while rural blocks range from just a handful of households to the same upper limit around 125.[15] We exclude outlier neighborhoods which have fewer than 150 people (typically very small villages) or more than 1000 people.[16] Note that "enumeration blocks" are not the same units as "census blocks" (sometimes just called "blocks").[17]

While rural and urban enumeration blocks have similar populations, the geography of urban and rural access to public facilities are quite different. Urban areas are dense, such that indi-

[15] Appendix Figure A.1 shows the distribution of block population in the sample. Results are very similar if we exclude rural villages that are too small to have a 100 household enumeration block, an analysis which results in similar distribution of urban and rural blocks.

[16] These make up less than 1% of the population and our results are unchanged if they are returned to the sample. Very large enumeration blocks are excluded because they are anomalous (suggesting potential data collection issues), and because segregation measures are scale-dependent (see below). However, there are so few of these neighborhoods that the potential bias here would be small even if they were included.

[17] Census blocks have population of about 200,000 each, and are unrelated to any units used in this paper.

viduals can travel across many enumeration blocks for work or access to public services. Rural areas are more dispersed: neighboring enumeration blocks are separated by larger distances than neighboring blocks in cities. Because enumeration block boundaries are defined for the convenience of enumerators, multi-block villages with multiple hamlets will typically have enumeration block boundaries that keep hamlets self-contained within blocks. The distances between neighboring villages range from 0.5 to 5 km, depending on the region. The geocodes describing locations and polygons of enumeration blocks were not available to us — they are sold as hand-drawn maps at high cost by the Indian Census.

3.2 Public Facilities

The Population Census town and village directories report a wide range of public services, but these are only identified at the town/village level and map on to neighborhoods only for very small villages. To identify public facilities at the neighborhood level, we instead use the 2013 Economic Census (EC13). 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. EC13 records enumeration block identifiers, making it possible to identify public health centers, primary schools, and secondary schools at the neighborhood level.[18] Health centers include hospitals, inpatient and outpatient clinics, and traditional care providers. EC13 also records whether a firm owner is Muslim or SC. The employment share in SC or Muslim firms is highly correlated with the group share in each neighborhood. We measure public service availability with binary measures that indicate whether an enumeration block contains a given type of public facility.

3.3 Demographic Data

Data on individuals comes from the Socioeconomic and Caste Census (SECC), a national asset census which recorded information on every household and individual in India (mostly in 2012) to determine eligibility for social programs. The SECC describes age, gender, education,

[18] The 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 is thus not possible to study changes in neighborhood-level services over time.

occupation, and SC status for every household member, as well as a short list of assets used to rapidly assess socioeconomic status. The SECC was made publicly available online in a combination of formats; we scraped and processed the data, with the approach described in detail in Asher and Novosad (2020). The urban data were not posted in their entirety; our sample covers 196 million urban residents, compared with the census urban population of 385 million.[19]

Consumption is not directly measured in the SECC, but we generate small area estimates of household per capita consumption on the basis of all of the household assets on the SECC schedule, using the IHDS-II (2011–12) survey as our data source for consumption (Elbers et al., 2003). This process generates similar rural and urban consumption distributions to direct survey measures; see Asher and Novosad (2020) and Asher et al. (2021) for more details.

The SECC surveyed individual caste and religion, but religion was not released in the public data.[20] We therefore classify individuals as Muslims or non-Muslims using their first and last names, which were posted in the public data. Because of the distinctive naming patterns of Muslims, we can identify Muslim names with an out-of-sample accuracy of 97%. We do this with an long-short-term-memory (LSTM) neural network, which classifies names on the basis of repeated letter sequences, using a religion-labeled dataset of 2 million applicants to the Indian Railways as a training sample. This approach has much higher accuracy than a fuzzy merge; the latter creates classification errors when small letter substitutions change a name identity, such as *Khan* (a stereotypical Muslim name) vs. *Khanna* (a Hindu name). The neural network implementation is described in detail in Ash et al. (2022); a similar approach is taken by Chaturvedi and Chaturvedi (2023). We verified the classification accuracy on a withheld subset of the names in the railway data, as well as on a set of manually classified names in the SECC. Our classification also closely predicts the subdistrict-level population share of Muslims (Appendix Figure A.2). We pool Hindus with the 6% of Indians who are Jain, Christian, Sikh,

[19]To the best of our knowledge, missing data was a function of the actions of IT administrators and was unrelated to the data contents. Town and neighborhood data 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.

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

or some other non-Hindu religion; we describe this group as "non-Muslims."[21]

For comparison with the United States, we use data from the 2020 U.S. Census and the Diversities and Disparities project, which is based on the 2010 U.S. Census.

3.4 Neighborhood Time Series Data from the Population Census Handbooks

The Population Census publishes detailed District Handbooks, which have hundreds of pages of appendices with additional census tables. The 1991, 2001, and 2011 District Handbooks list each enumeration block in each city, along with its total, Scheduled Caste, and Scheduled Tribe population. The enumeration block identifiers are not persistent over time, but this resource makes it possible to calculate town-level segregation of Scheduled Castes (but not Muslims) in prior years. However, the data is embedded in PDF tables and scanned documents, often with formatting that varies across districts, creating a barrier to access.

We obtained copies of the Handbooks for 2001 and 2011; we could find 1991 handbooks for only a handful of districts. We developed a PDF parsing tool to extract tabular data from these handbooks and were able to parse enumeration block data for 1600 towns in our sample, representing about a third of India's urban population.[22] For the time series analysis, we use a set of 1400 towns for which we have data in both 2001 and 2001, where the total enumeration block population in the parsed data is within 50% of the total population recorded in the Census. A secondary sample is restricted to towns with a population mismatch of under 5%. Results are not materially affected under a range of alternate inclusion criteria.

4 Methods

4.1 Measuring and Comparing Residential Segregation

Our first objective is to document the extent of residential segregation of Muslims and SCs; we estimate the canonical dissimilarity and isolation indices.

[21]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.

[22]For the remaining towns, we were either unable to obtain the detailed District Handbook Appendices with enumeration block population counts, were unable to digitize the tables due to document quality, or did not have the data in either the SECC or equivalent 2011 District Handbooks. The observed towns cover all regions of India and are broadly representative of the full size distribution of towns.

The dissimilarity index ranges from zero to one and answers the question: what share of the marginalized group would need to change neighborhoods for it to be evenly distributed within a city? We calculate this index for marginalized group MG and majority group MAJ across the set of blocks B in city c as:

$$DISSIMILARITY_c = \frac{1}{2}\sum_{b\in B}\left|\frac{N_{\text{MG,b}}}{N_{\text{MG,c}}} - \frac{N_{\text{MAJ,b}}}{N_{\text{MAJ,c}}}\right|, \tag{1}$$

where $N_{\text{g,b}}$ is the number of members of group g in block b, and $N_{\text{g,c}}$ is the population of that group in the city.

The isolation index measures the extent to which a population group is exposed only to members of its own group. It can be summarized as the marginalized group share in the average neighborhood of a member of the marginalized group:

$$ISOLATION_c = \sum_{b\in B}\left|\frac{N_{\text{MG,b}}}{N_{\text{MG,c}}} \cdot \frac{N_{\text{MG,b}}}{N_{\text{total,b}}}\right|, \tag{2}$$

where $N_{\text{total,b}}$ is the population of all groups in the given block.

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 very different dynamics, causes, and consequences. Tripartite segregation measures exist, but they do not describe the dynamics that we aim to explore here, as we are specifically interested in differences between SC and Muslim segregation. We pool groups in this way only when calculating segregation; when looking at access to public service and social group outcomes below, we always separate Muslims, SCs, and non-Muslim non-SCs.

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.[23] Following convention in the U.S. segregation literature, we weight the indices across cities by each group's city or subdistrict population,

[23] 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.

providing an aggregate measure which reflects the experience of members of the marginalized group in question.

Measures of segregation can change depending on the level of aggregation used to define neighborhoods. 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. Our analysis defines neighborhoods at the enumeration block level (i.e. units of about 125 households or 500 people), as these are the most accurate contiguous residential units that we can identify. This scale also fits our intuitive understanding of the set of households with which individuals will most often interact.

The scale-variance of the segregation indices means that a comparison with the United States — where census tracts have populations ranging from 1000 to 8000 and average around 4000 — would be biased toward finding greater segregation in India. Therefore, 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.[24] While the level of the segregation index changes, except where noted, our comparative results are robust to aggregating neighborhoods to higher sizes.

4.2 Marginalized Group Shares and Neighborhood Public Services

Our second objective is to describe differences in access to public services between marginalized group and non-marginalized-group neighborhoods. We present the methods for the case of urban areas; the methods for rural areas are analogous.

Our main interest is in understanding how a fixed supply of public services is allocated across MG and non-MG neighborhoods within cities. We measure the allocation disparity with the following neighborhood-level regression:

[24] In the handful of cities where we have enumeration block maps or neighborhood names, we confirm that adjoining enumeration blocks are almost always adjoining in geography. 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, thus overstating segregation relative to an approach studying random geographic units. Replicating this approach in India is not possible given the data available. As a result, the U.S. segregation measures may be biased upward relative to those in India.

$$SERVICE_{n,c} = \beta_c MG\_Share_{n,c} + \Omega_c + \boldsymbol{\nu} POPULATION_{n,c} + \epsilon_{n,c}. \quad (3)$$

$SERVICE_{n,c}$ is a measure of the supply or availability of public services in neighborhood n and city c, such as an indicator for the presence of a secondary school. $MG\_Share_{n,c}$ is the marginalized group share. We include a control for neighborhood population, since it could be mechanically related to the supply of public services (though most neighborhoods are similarly populated). We include a city fixed effect Ω_c, which controls for differences in the availability of public services in cities with more or fewer members of marginalized groups.[25]

The coefficient β_c describes how service availability changes as the marginalized group share increases. A negative estimate indicates that a city's public services are 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*. For each service, we first show the non-parametric relationship between service availability and the MG share, and then summarize it with the linear estimate from Equation 3.

We then estimate disparities at higher levels of aggregation. If we estimate Equation 3, but we replace the city fixed effect with a *district* fixed effect (Ω_d), the coefficient on $MG\_SHARE_{n,c}$ will describe the allocation of services *within* districts; we call this β_d. This measure describes a combination of (1) the allocation of services across towns within districts; and (2) 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 access which comes from variation *within* districts and across towns.[26] 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 3 with state fixed effects gives

[25] 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.

[26] 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.

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.[27] 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 these district-level transfers will be less efficient at reducing disparities than neighborhood-level transfers (which would reduce α_c directly).

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

These 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

[27] We use subscript "f" because α_f describes the federal (i.e. cross-state) political economy equilibrium, and we use the subscript "o" to denote the estimate from Equation 3 with no fixed effects.

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.[28] 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 that is a function of decisions by politicians, bureaucrats, firms, and citizens. These "rules" are outcomes not only of the public service allocation process, but also of the decision choices of individuals. They are statistics that describe the entire political economy equilibrium. 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. It describes the equilibrium inequality in public service allocation at one level of governance, but does not attribute it to a specific policy or actor.

4.3 Marginalized Group Shares and Living Standards

Disparities in access to public services in marginalized group neighborhoods are most concerning if they result in unequal outcomes for people living in those neighborhoods, which would entrench inequality across groups. But if people in under-serviced neighborhoods can compensate by traveling to other neighborhoods for services or by consuming private services, then unequal allocation may be less harmful. The final part of our analysis therefore examines whether individuals experience worse outcomes in MG neighborhoods.

We use the following equation to examine the relationship between neighborhood MG share and the young generation's educational outcomes:

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

$$ED_{i,n,c} = \beta_1 MG\_Share_{n,c} + \beta_2 MG_i + \Omega_c + \boldsymbol{\nu} \boldsymbol{X}_{i,n,c} + \epsilon_{i,n,c}. \tag{4}$$

$ED_{i,n,c}$ is the number of years of education of individual i in neighborhood n and city c. $MG\_share_{n,c}$ is the marginalized group share of the neighborhood, and MG_i is an indicator for whether individual i is in the marginalized group. The inclusion of both lets us separate the effect of being a member of a marginalized group from the effect of living in a marginalized group neighborhood. We run the estimation with both marginalized group indicators and neighborhood measures (SC and Muslim) included simultaneously. The sample consists of men and women aged 17–18. We show unadjusted estimates, as well as estimates controlling for parent education and household consumption.

The analysis is descriptive, and is at best suggestive of the causal relationship between neighborhood and child outcomes for several reasons. First, we do not know individuals' place of birth. We focus on young people, and test robustness by looking at younger respondents (who are very likely to still live with parents), but we do not have the data to exclude recent migrants from the sample. More importantly, there may be unobserved characteristics of households which cause them to have worse outcomes and also choose to live in neighborhoods with other minorities.

Nevertheless, these results tell us the extent to which children growing up in MG neighborhoods are getting the human capital investments necessary for future socioeconomic success and shed light on the prospect for mean reversion across groups. We discuss the appropriate interpretation of the controlled and uncontrolled estimates of β_1 with the results in Section 5.4.

5 Results

5.1 Segregation in Indian Villages and Cities

Table 1 presents summary statistics of the neighborhood-level sample, separately for urban and rural neighborhoods. Both rural and urban neighborhoods have about 500 people each; there are about 1.1 million rural neighborhoods and 400,000 urban neighborhoods in the sample. The difference in sample size reflects India's low urbanization rate (31% in 2011), slightly magnified by our worse sample coverage of urban places. Scheduled Castes are relatively more likely to live in rural areas, while Muslims are more likely to live in towns and cities.

Table 2 describes the same data at a higher level of aggregation: the town or city for urban areas, and the subdistrict for rural. The table also compares our sample characteristics with the full set of towns and villages in the Population Census. In rural areas, our sample is highly representative, covering 81% of rural subdistricts and 68% of rural people. In urban areas, we have slightly higher sample coverage of larger cities (which are also older); smaller and more recent cities were less likely to have data posted by the SECC, making information about their residential composition unavailable.[29] Towns excluded from our sample have slightly fewer public services, but similar marginalized group shares. While not fully representative, our urban sample covers 50% of towns and 51% of India's urban population.

Differences in segregation across the two groups are not captured in a single dimension (Table 2). Measured by the dissimilarity index, Scheduled Castes are more segregated than Muslims in both rural and urban areas. According to the isolation index, however, Muslims are more segregated in urban areas. By both measures, Muslims are relatively more segregated in cities. Figure 1 shows the distribution of these measures across cities and rural subdistricts.

Appendix Figure A.3 helps to unpack these differences by showing the distribution of MG shares across neighborhoods. The Muslim distribution is notably bimodal, in both urban and rural areas. SCs are more segregated on average, but a greater share of Muslims live in the most segregated neighborhoods. 26% of urban Muslims live in neighborhoods that are >80% Muslim, while 17% of urban SCs live in neighborhoods that are >80% SC.[30]

It is useful to benchmark the segregation measures against those in MSAs in the United States, where segregation has been most studied. To match the definitions used by the U.S. Census, we aggregate neighborhoods to populations of at least 4000 people (as described in Section 4), and we limit our sample to cities with more than 100,000 people. Appendix Figure A.4 shows the density

[29]The table does not show the public infrastructure measured in the SECC, because we don't observe these out of sample or in rural areas.

[30]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.

functions of dissimilarity and isolation across cities. Using the 500 people per neighborhood definition in India, Muslims are systematically more segregated than U.S. Blacks, while for SCs it depends on the measure of segregation. Using the 4000 people per neighborhood definition, the distribution of U.S. Black segregation looks very close to that of Muslims. The first measure biases Indian segregation upward relative to U.S. segregation (because of the smaller neighborhood size), while the second biases Indian segregation downward (because of measurement error in neighborhood pooling). But the measures are consistent in showing that segregation in India, particularly of Muslims, is comparable in magnitude to that of Black people in the contemporary U.S..[31]

Figure 2 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,[32] 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. Panel A of Figure 3 plots the average urban Muslim dissimilarity index in each district, as a function of the rural Muslim dissimilarity index in the same district. The urban and rural dissimilarity indices are very highly correlated ($\rho = 0.56$), suggesting that the regional dynamics that lead to the separation of social groups in rural areas are also important in cities and towns. Panel B of the same figure shows that segregation patterns of Scheduled Castes are also highly correlated across urban and rural spaces, but less so than for Muslims ($\rho = 0.43$).

There are few data sources available in which we can observe changes in segregation over time. Using the District Handbook data described in Section 3.4, Table 3 shows changes over time in SC segregation under a range of measures and sample definitions. Scheduled Caste dissimilarity fell marginally between 2001–11 (between 0.002 and 0.014 on a base of about

[31]Note that this is a lower level of segregation than peak U.S. segregation in the 1960s and 1970s, when the weighted dissimilarity index across U.S. MSAs was close to 80, compared with 60 in the 2020 U.S. Census (Massey and Denton, 1993).

[32]See, for example, Pew Research Center (2021).

0.56), while the isolation index rose about 0.01 (on a base of 0.4). These changes are small; in comparison, the U.S. dissimilarity index has fallen by an average of 0.045 per decade between 1970 and 2020. Figure 4 shows the pattern of changes as a function of city size, showing that the changes in both dissimilarity and isolation are consistent across a broad range of city sizes.

The time series analysis is limited by the absence of neighborhoo-level data from before 2001, and the absence of neighborhood Muslim population before 2011. An alternate way to shed light on changes over time is by comparing recently urbanized places with more established settlements. Table 4 shows results from town-level regressions of Muslim and SC dissimilarity on the decade that a town or city first appeared as a town in the Population Census.[33] To account for the fact that our sample underrepresents the smallest and youngest towns, we also present results with controls for population and group shares.

In the long synthetic panel, we find that a town that is 10 years older has a dissimilarity index which is 0.009 points lower for Muslims and 0.007 points lower for SCs, with similar results for the isolation index. The city age effect is robust to the inclusion of additional town-level covariates.[34] The synthetic panel lets us look at a longer time series than the 2001–11 District Handbook data, but its interpretation is less clear. Segregation could indeed be falling over time, in the sense that newly settled cities develop less segregated neighborhood patterns than cities settled in the past. But the city age effect could also arise from cities themselves becoming more segregated over time, with marginalized group neighborhoods emerging and absorbing more group members over time. Under either interpretation, both approaches consistently imply that segregation in Indian cities is not changing very rapidly.

5.2 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

[33] The Census classifies a settlement as a town once it has more than 5000 people, an agricultural labor share (among men) below 25%, and a population density of at least 400 person per square kilometer.

[34] Appendix Figure A.5 shows that the dissimilarity difference between young and old cities is largely stable across settlements with different population sizes; it is not a mechanical function of size.

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 >50% Muslim share stand out for being particularly underprovisioned; there are not so many of these neighborhoods in India, but as shown in Appendix Figure A.3, a large share of Muslims live in them. 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).[35]

We summarize this nonparametric relationship between neighborhood MG share and public facility presence with the linear estimator from Equation 3, 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 5 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

[35]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. The relationship looks similar with a continuous measure of school size (total employment in the school) as an outcome (Appendix Figure A.6).

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.[36]

Table 6 shows analogous tests with private schools and clinics, 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 * urban/rural * 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.[37]

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 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.[38]

For the schools and clinics in urban areas, people can walk to a facility in a nearby neighborhood, mitigating the cost of not having a facility in one's own neighborhood. However in rural areas, the nearest facility outside the "neighborhood" can be quite far away. For the infrastructure goods, substitutes in nearby neighborhoods (e.g. for clean water) clearly imply substantial welfare costs.[39]

[36] 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.1).

[37] Note that the public facility results are not adequately explained by marginalized groups being poorer — since the role of government is ostensibly to provide equal access to public services whether people are rich or poor.

[38] 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.

[39] Because we only have GIS locations for neighborhoods in a handful of places, it is not possible to run

5.3 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 3. 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 take some time to explain these figures 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 neighborhood.[40] This is simply the coefficient from a regression of the primary school indicator 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 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

a test with the distance to the nearest neighborhood with a facility.

[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.

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 — the latter being exactly the estimates from the previous subsection (5.2).

The sum of all the α coefficients gives us the final estimate of -1.9. 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 — again, these bars are just the graphical representations of the estimates in Table 5. 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 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.

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.

The previous section 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; the effects combine to make Muslim neighborhoods severely 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 (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.

For the infrastructure services, there is thus less systematic evidence of affirmative action in favor of any marginalized group, but both groups systematically fare worse at the neighborhood level. Indeed, we are aware of no national programs to improve urban infrastructure services like these or to equalize 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.[42]

5.4 Outcomes for Marginalized Groups in Segregated Neighborhoods

This section examines the relationship between the educational outcomes of young people and the marginalized group share in the neighborhoods where they live. We use Equation 4, which describes a regression of individual years of education on neighborhood marginalized group shares, for individuals aged 17–18 years.

The evidence here is descriptive: raw differences in outcomes across neighborhoods reflect some combination of discrimination and preferences of those who live there. Controlling for individual and neighborhood characteristics is useful descriptively, but it does not necessarily improve the identification of discrimination, because those control variables may themselves be

[42] Appendix Figures A.7 and A.8 shows similar estimates to Figures 6 and 7 for private facilities. We spend less attention on these since there are no 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.

the result of discrimination. The empirical challenge is analogous to that of measuring gender discrimination in wages, where it is useful to know both the unadjusted gender wage gap and the wage gap controlling for job characteristics, but neither of these measures is a sufficient statistic for discrimination.

Table 8 shows the results for urban places; the outcome variable is years of education. We control for town/city fixed effects; the results are strictly across neighborhoods within cities. We also control for whether the individual is a Muslim or member of a Scheduled Caste. The ideal sample would be a set of people who grew up in the neighborhood and had completed their education. The best we can do is to focus on individuals aged 17–18 years old, though clearly some of them fail to meet both of these criteria, as we discuss below.

Column 1 shows the results for urban men aged 17–18, with only the household SC and Muslim indicators and town fixed effects; this shows the average difference between SC, Muslim, and non-SC non-Muslim outcomes, conditioning on town of residence. SCs have 1.1 fewer years of education than non-SC non-Muslims, and Muslims have 1.2 years fewer. Column 2 adds the neighborhood shares, which are the variable of interest and are indicated in bold in the table. Including the neighborhood share drives down the coefficient on the SC indicator by 45%, and on the Muslim indicator by 55% — about half of the disadvantage faced by marginalized groups is explained just by the marginalized group share of their neighborhood.

Still looking at Column 2, controlling only for the town fixed effects and the household SC and Muslim indicators, we see that Muslim and SC neighborhoods have significantly worse educational outcomes (the row in bold). 17–18-year-olds living in a 100% Muslim neighborhoods have 2.1 fewer years of education than those in 0% Muslim neighborhoods; the coefficient for Scheduled Caste neighborhoods is -1.6.

In Column 3, we add controls for both parents' years of education and household consumption. Unsurprisingly, all these controls are positively correlated with individual outcomes in the expected direction, and the inclusion of these controls brings down the magnitude of the coefficients on both of the neighborhood marginalized group shares.[43] The coefficient on the SC share is driven close to zero in the specification with controls, while the coefficient on the neighborhood Muslim share falls to less than half of its value in the unadjusted Column 2. The effect sizes have similar orders of magnitudes for members of marginalized and non-marginalized groups (Appendix Table A.2). Living in a marginalized group neighborhood is thus associated with much worse outcomes regardless of an individual's identity.

We interpret these results as follows. Young people in SC neighborhoods have systematically worse outcomes than those in non-SC neighborhoods — but the difference is mostly explained by the economic status of their families. This does not rule out a negative causal effect of growing up in an SC neighborhood on child outcomes, because those parent outcomes could themselves be caused by living in a bad neighborhood. For example, parents might invest less in their house (lowering their measured consumption) if they lack security of tenure.[44]

In Muslim neighborhoods, outcomes for young people are equally poor, but can be only partially explained by parent consumption and education. Children in these neighborhoods grow up in families with fewer resources, and yet have even worse outcomes (about 1 year lesser in a 100% Muslim neighborhood) than similarly poor children in non-Muslim neighborhoods. Again, this is a function of the neighborhood, not of the social group of the individual, as it holds for members of all social groups.

Columns 4–6 show the same results for rural areas, with results separated by social group in Appendix Table A.3; the results for Muslims are broadly similar. As in urban areas, young rural people in neighborhoods with high Muslim shares end up with substantially less education than those living in non-Muslim, non-SC neighborhoods. Rural SC neighborhoods do not show the same disadvantages; the coefficient on the SC share is close to zero, and even marginally positive for young women.[45] We graph the coefficients on the neighborhood group shares in

[43] The additional inclusion of mean neighborhood income has little effect beyond the individual consumption measures.

[44] The result is analogous to finding that a gender wage gap goes to zero if occupation, job description and job rank are controlled for — a result that does not disprove discrimination, since discrimination could result in different occupations and job ranks and descriptions.

[45] Young women in 100% SC neighborhoods have on average 0.07 additional years of education; the effect

Appendix Figure A.9 for easy comparison.

A limitation of these findings is that we do not observe individuals' places of birth, so the results here in part could be driven by less-educated 17- and 18-year-olds moving into neighborhoods with high marginalized group shares. We can show, however, that these results hold for children at all ages, including those arguably too young to be responsible for their own migration choices (see Appendix Figures A.10 and A.11).

These results suggest that marginalized group neighborhoods and the disadvantages associated with living in them reduce access to opportunity for people who grow up there. Our data does not allow us to calculate neighborhood exposure effects, as in Chetty and Hendren (2018) and Alesina et al. (2021), which would be even more dispositive; this is an important area for future research.

6 Conclusion

This paper presents a national-scale analysis of socio-economic outcomes 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 highly 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 both access to public services and of living standards in those neighborhoods. Within cities, Muslims and members of Scheduled Castes have much lower access to public services. India's rapidly growing cities, famous as engines of upward mobility, to a large degree have replicated the caste and religious structure of its villages.

Our research so far does not identify the causes of these disparities. However, discriminatory provision of public facilities to MG neighborhoods has been a persistent characteristic of the

is statistically significant, but economically small. Though in comparison with the other results in this paper, even a non-negative estimate here is notable.

political economy in many countries, including in India.

A limitation of our work is that it is largely based on cross-sectional data collected in 2012–13. The historical literature suggests that Scheduled Castes have been isolated at the neighborhood level for generations, but Muslim isolation has been exacerbated by Hindu-Muslim violence in the post-colonial era. Data from historical censuses could potentially shed light on changes in residential segregation over time.

That living standards are so much lower in SC and Muslim enclaves suggests that, as elsewhere, spatial concentration of marginalized groups may limit their economic opportunities. Modern India has never had the government regulations, such as redlining, 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 had disastrous consequences for the United States; it has prevented generations of individuals from access to 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 hope to draw attention to the critical choices that lie ahead for India and other urbanizing lower-income countries around the world.

References

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 Government in a Cross Section of Countries," *American Economic Review*, 2011, *101* (5), 1872–1911.

_ , Sebastian Hohmann, Elias Papaioannou, and Stelios Michalopoulos, "Intergenerational Mobility in Africa," *Econometrica*, 2021, *89* (1), 1–35.

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

Ash, Elliott, Sam Asher, Aditi Bhowmick, Sandeep Bhupatiraju, Daniel L Chen, Tanaya Devi, Christoph Goessmann, Paul Novosad, and Bilal Siddiqi, "In-group bias in the Indian judiciary: Evidence from 5 million criminal cases," 2022. Working paper.

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, Space, and Communities," 2022.

_ , 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.

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), 287–314.

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

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

_ **, Deepak V Malghan, and Andaleeb Rahman**, "Isolated by caste: Neighbourhood-scale residential segregation in Indian metros," *IIM Bangalore Research Paper*, 2018, *572.*

Boustan, Leah Platt, "Racial Residential Segregation in American Cities," *NBER Working Paper*, 2013, (No. w19045).

Cassan, Guilhem, "Affirmative action, education and gender: Evidence from India," *Journal of Development Economics*, 2019, *136*, 51–70.

Chaturvedi, Rochana and Sugat Chaturvedi, "Itâs All in the Name: A Character-Based Approach to Infer Religion," *Political Analysis*, 2023, pp. 1–16.

Chetty, Raj and Nathaniel Hendren, "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects," *The Quarterly Journal of Economics*, 2018, *133* (3), 1107–1162.

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.

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.

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.

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.

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

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

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

Massey, Douglas S and Nancy A Denton, *American apartheid: Segregation and the making of the underclass*, Harvard university press, 1993.

Pew Research Center, *Religion in India: Tolerance and Segregation* 2021.

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

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.

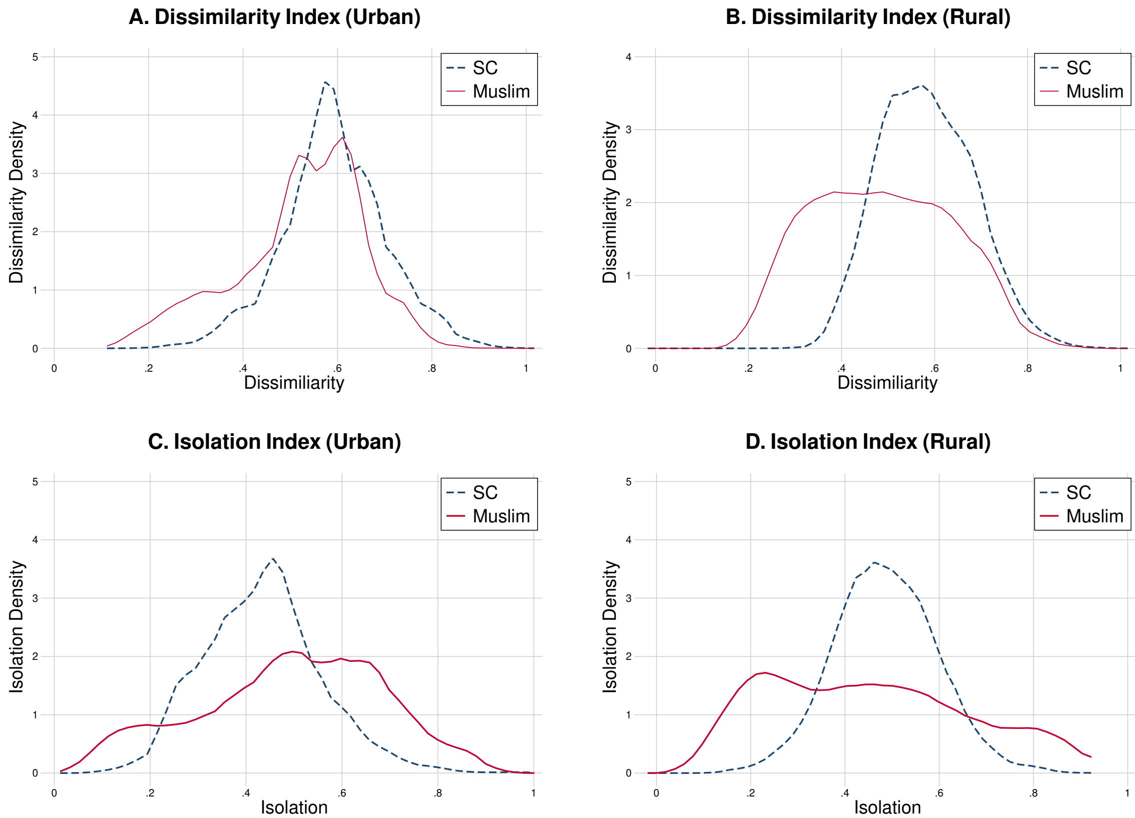
Susewind, Raphael, "Muslims in Indian cities: Degrees of segregation and the elusive ghetto," *Environment and Planning A*, 2017, *49* (6), 1286–1307.

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

Figures

Figure 1

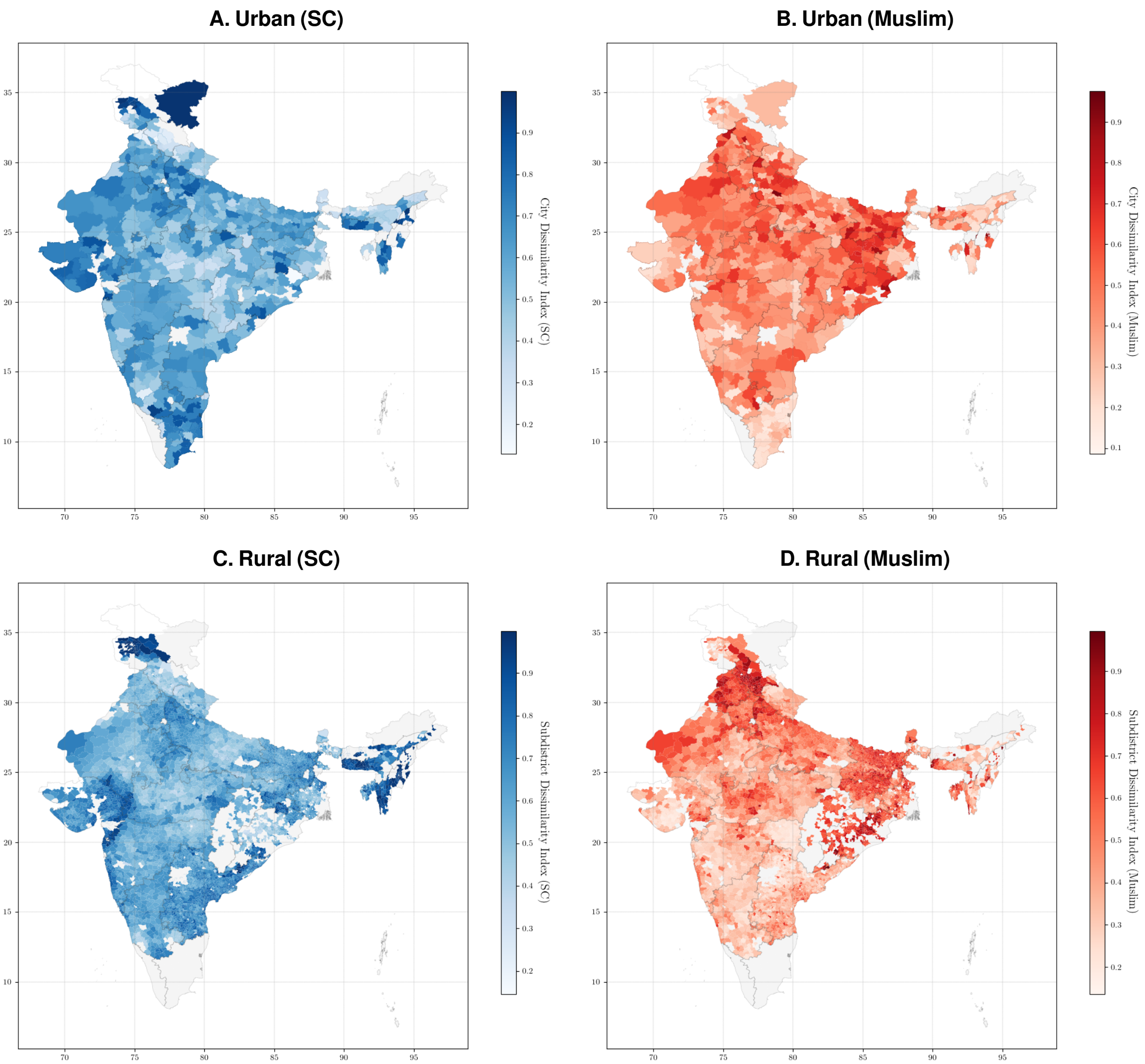
Segregation of Muslims and Scheduled Castes



Notes: The figure shows the distribution of segregation across cities/towns (Panels A and C) and rural subdistricts (Panels B and D), according to the dissimilarity and the isolation indices. The SC density functions are weighted by each town/subdistrict's SC population, and the same for Muslims, such that each curve represents the experience of members of the marginalized group. The urban sample has one observation per town, and the rural sample has one observation per subdistrict. Source: SECC 2012.

Figure 2

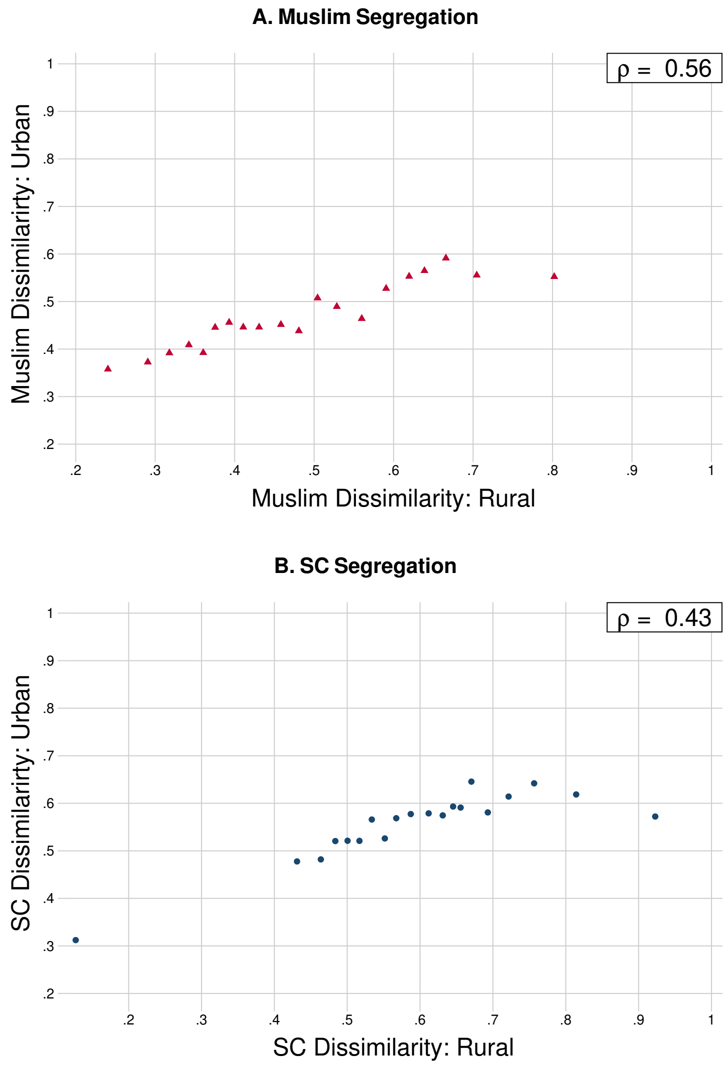
Segregation Maps



Notes: The maps show the distribution of segregation across 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.

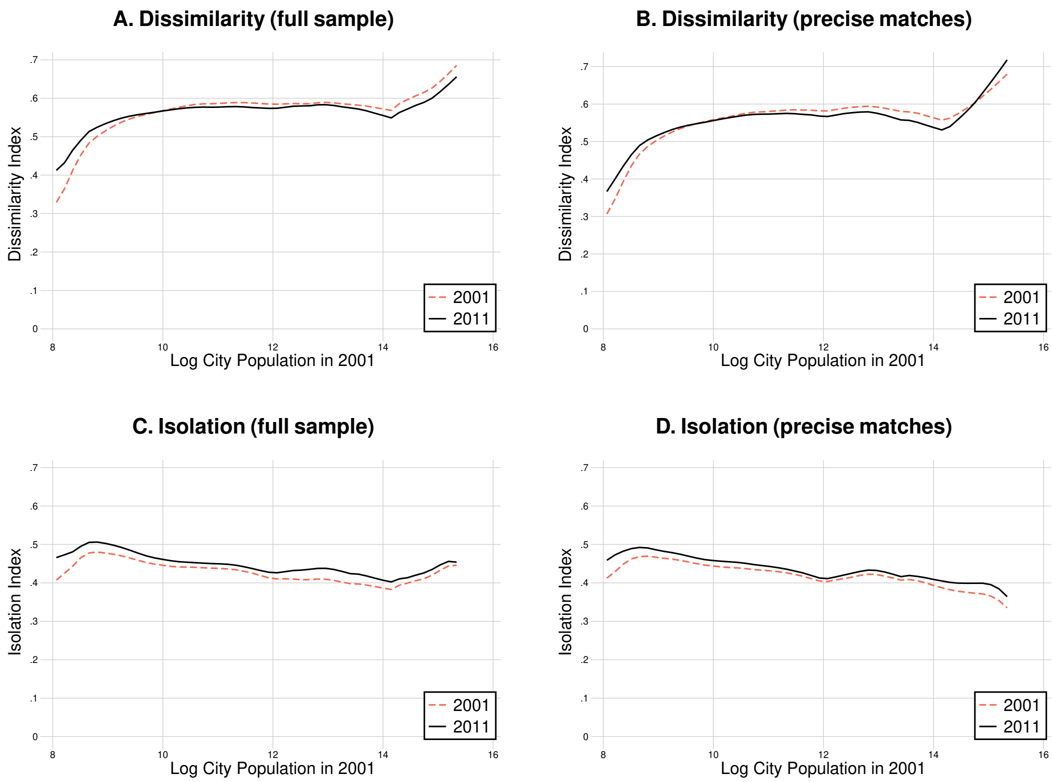
Figure 3

Urban vs Rural Segregation: District-level Comparisons



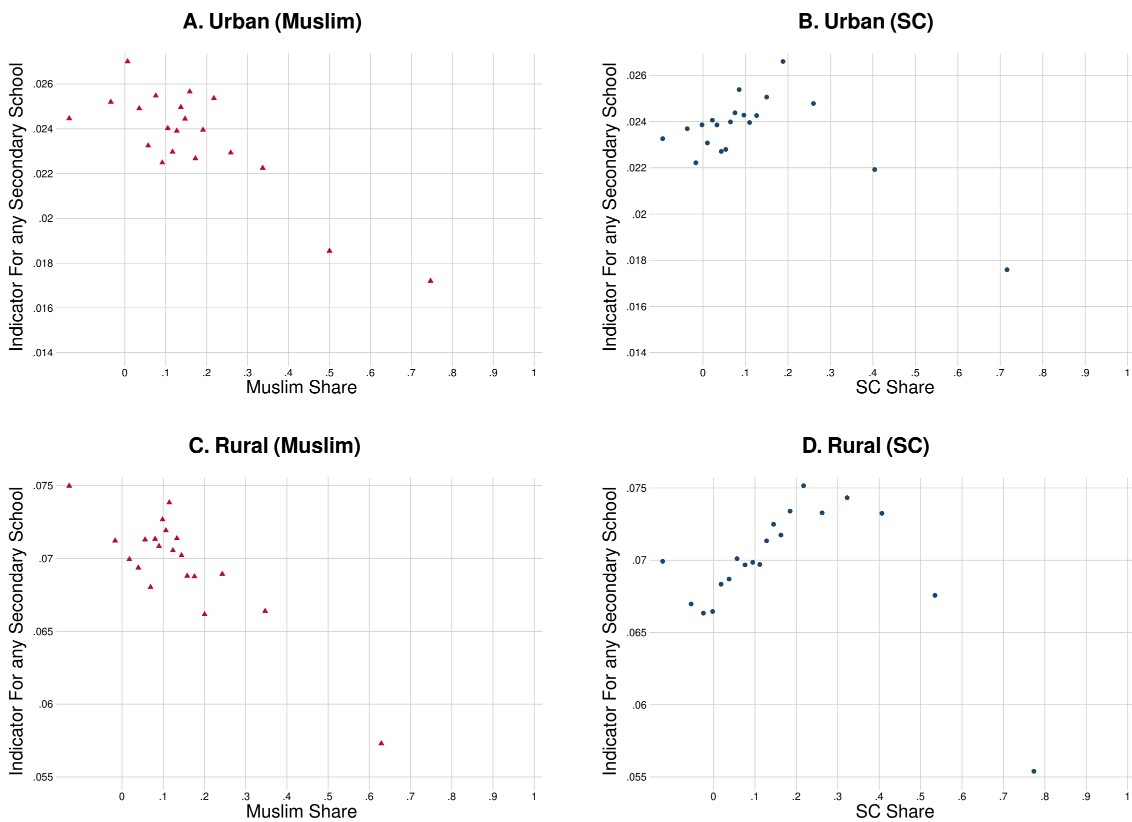
Notes: Each graph shows a binscatter representing the relationship between urban and rural segregation in the same district. Each point shows the mean urban dissimilarity across about 20 subdistricts with mean rural dissimilarity in the vicinity of the X axis value. Source: SECC 2012.

Figure 4
Changes in Urban Scheduled Caste Segregation (2001–2011)



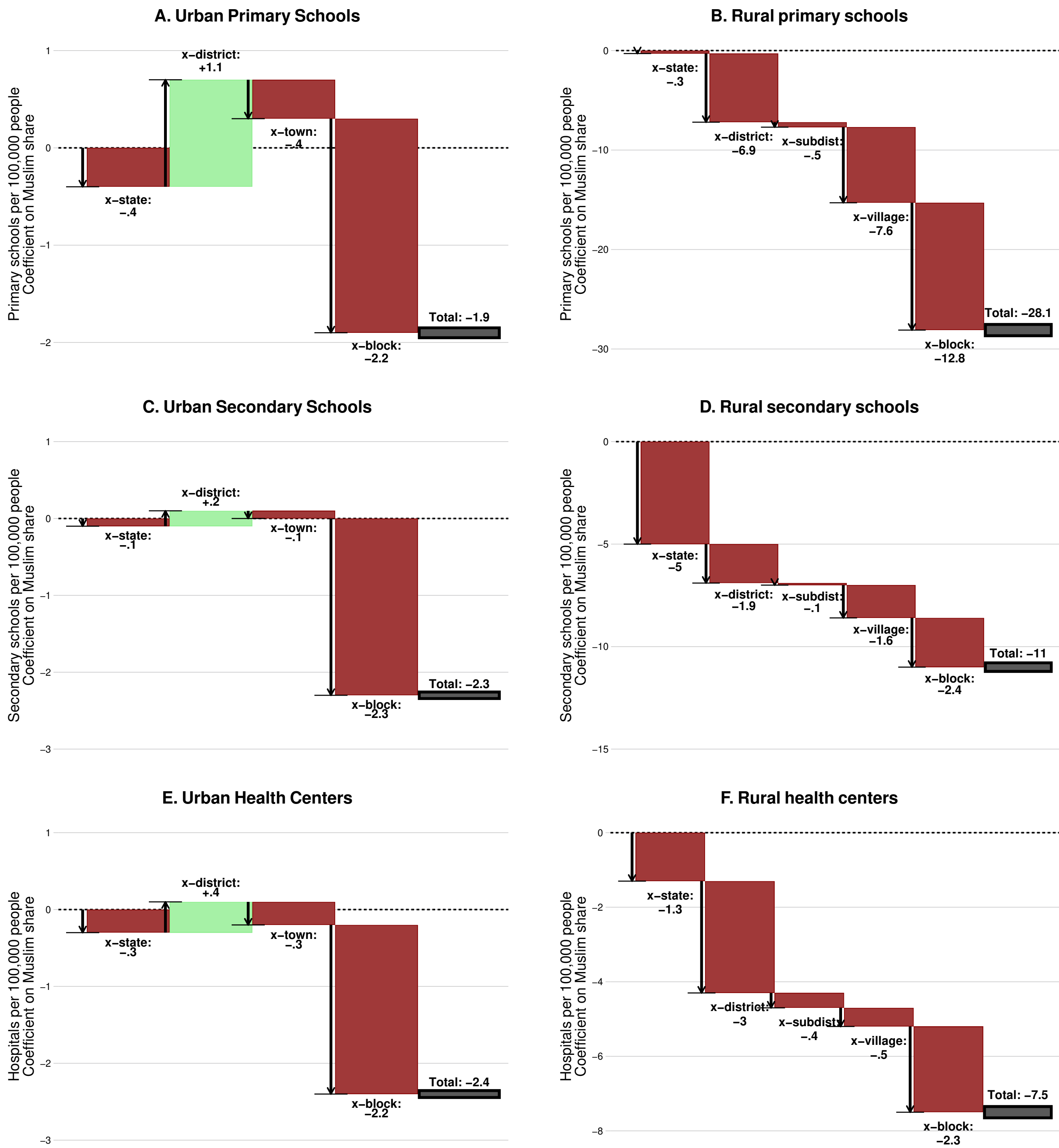
Notes: Each graph shows a local linear non-parametric regression of Scheduled Caste segregation (measured by dissimilarity or isolation) as a function of city log population, for 2001 and 2011. Data source: Census District Handbooks, 2001 and 2011.

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. Source: Economic Census 2013, SECC 2012.

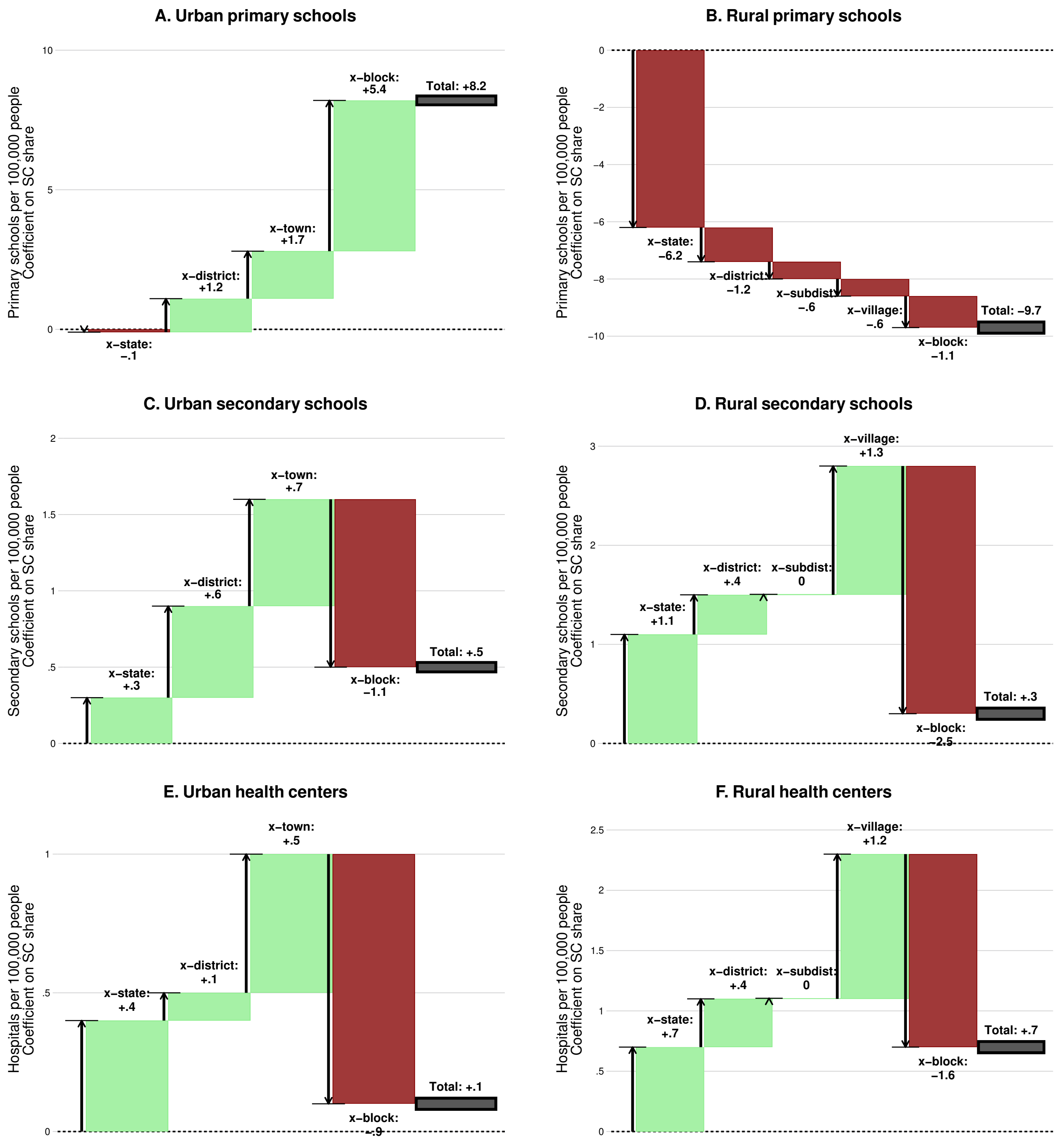
Figure 6
Disparities in Public Facilities as a
Function of Neighborhood Muslim Share



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. Source: Economic Census 2013, SECC 2012.

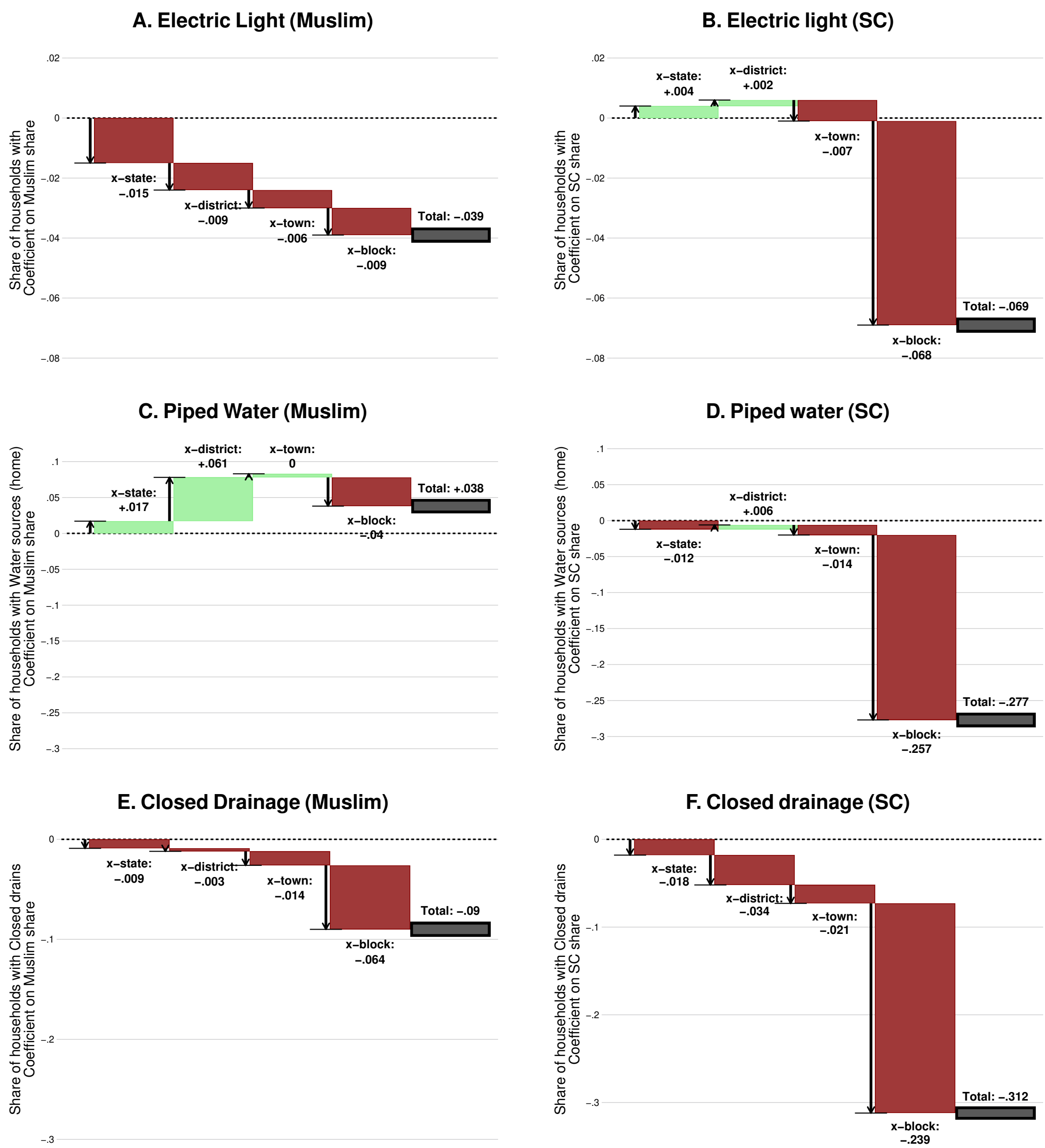
Figure 7

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.

Tables

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	
	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)
Dissimilarity Index (SC)	0.59		0.58	
	(0.11)		(0.10)	
Dissimilarity Index (Muslim)	0.52		0.49	
	(0.14)		(0.15)	
Isolation Index (SC)	0.43		0.48	
	(0.13)		(0.11)	
Isolation Index (Muslim)	0.49		0.45	
	(0.20)		(0.23)	
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. Standard errors are in parentheses.

Table 3

Changes in Urban Scheduled Caste Segregation (2001–2011)

A. Dissimilarity Index

Sample Definition	Weighted	Dissimilarity (2001)	Dissimilarity (2011)	Change	Sample Size
Full Sample	Yes	0.567	0.569	-0.002	1404
	No	0.573	0.587	-0.014	1404
Precise Population Match	Yes	0.556	0.559	-0.003	826
	No	0.563	0.576	-0.013	826

B. Isolation Index

Sample Definition	Weighted	Isolation (2001)	Isolation (2011)	Change	Sample Size
Full Sample	Yes	0.408	0.390	0.018	1404
	No	0.440	0.429	0.011	1404
Precise Population Match	Yes	0.401	0.385	0.016	826
	No	0.431	0.420	0.011	826

Notes: The table shows estimates of urban segregation of members of Scheduled Castes in 2001 and 2011, based on data from the Indian Population Census District Handbooks, 2001 and 2011. Each observation is a town or city. The full sample is the set of all towns parsed from these two data sources, for which the total District Handbook population was within 50% of the official population in the Population Census Abstract (PCA) in each year. The "Precise Population Match" sample requires population to be within 5% of the PCA in each year. In rows marked "weighted", town observations are weighted by the Scheduled Caste population. The district handbooks do not report Muslim population.

Table 4

Correlates of Urban Segregation

A. Dissimilarity Index

	SC Dissimilarity		Muslim Dissimilarity	
City Origin Decade	-0.007***	-0.004***	-0.009***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
SC share (town)		-0.012		-0.035
		(0.032)		(0.030)
Muslim share (town)		0.127***		0.193***
		(0.022)		(0.021)
City Population		0.012***		0.016***
		(0.003)		(0.003)
Observations	3519	3519	3519	3519
R2	0.03	0.04	0.05	0.09
Mean of Dependent Variable	0.59	0.59	0.42	0.42

B. Isolation Index

	SC Isolation		Muslim Isolation	
City Origin Decade	-0.004***	-0.004***	-0.016***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.000)
SC share (town)		1.344***		-0.074***
		(0.027)		(0.017)
Muslim share (town)		0.031		1.273***
		(0.018)		(0.012)
City Population		0.014***		0.024***
		(0.002)		(0.002)
Observations	3519	3519	3519	3519
R2	0.01	0.43	0.11	0.81
Mean of Dependent Variable	0.38	0.38	0.27	0.27

Notes: The table shows estimates from regressions of segregation (the town-level dissimilarity or isolation index) on a set of town characteristics. Sources: SECC (2012) and Population Census (2011).

Table 5
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) and subdistrict (Panel B) levels.

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***	-0.480***
	(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: Urban

	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.

Table 8
Education Attainment of Young People
in Scheduled Caste and Muslim Neighborhoods

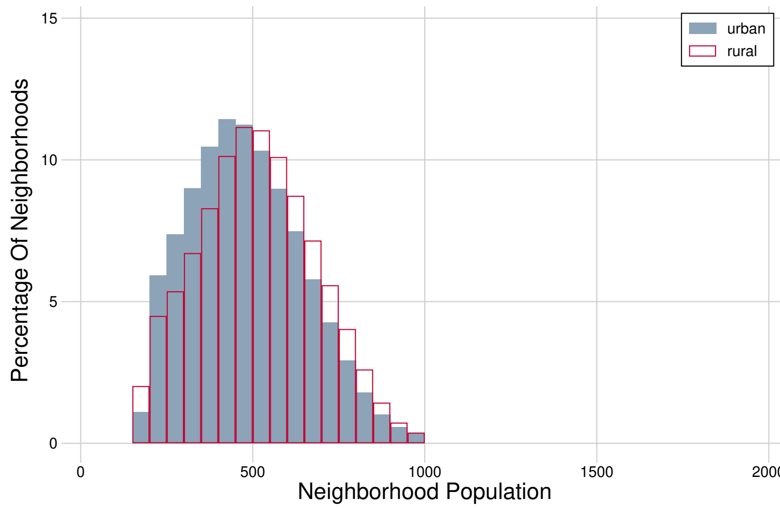
	(1)	(2)	(3)	(4)	(5)	(6)
	Urban			Rural		
SC (Individual)	-1.139***	-0.628***	-0.131***	-1.162***	-1.180***	-0.314***
	(0.009)	(0.008)	(0.007)	(0.007)	(0.006)	(0.006)
Muslim (Individual)	-1.249***	-0.572***	-0.267***	-1.163***	-0.760***	-0.370***
	(0.009)	(0.006)	(0.006)	(0.008)	(0.006)	(0.006)
Neighborhood SC Share		-1.632***	-0.149***		0.023	0.045***
		(0.024)	(0.021)		(0.018)	(0.016)
Neighborhood Muslim Share		-2.071***	-0.845***		-1.485***	-0.922***
		(0.024)	(0.021)		(0.025)	(0.021)
Father Ed			0.192***			0.282***
			(0.001)			(0.001)
Mother Ed			0.087***			0.091***
			(0.001)			(0.001)
Observations	5650186	5650186	3681721	8964412	8964412	6180712
R^2	0.15	0.16	0.33	0.16	0.17	0.32
Controls	None	None	Parent Ed + Cons	None	None	Parent Ed + Cons
Town/Subdistrict FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows results from individual-level regressions of years of education on individual and neighborhood characteristics. The variables of interest are the neighborhood SC and Muslim shares, which are accented in bold. Standard errors are clustered at the city (urban) or subdistrict (rural) level. Data source: a 10% random sample of individuals from the SECC.

A Appendix A. Additional Figures and Tables

Figure A.1

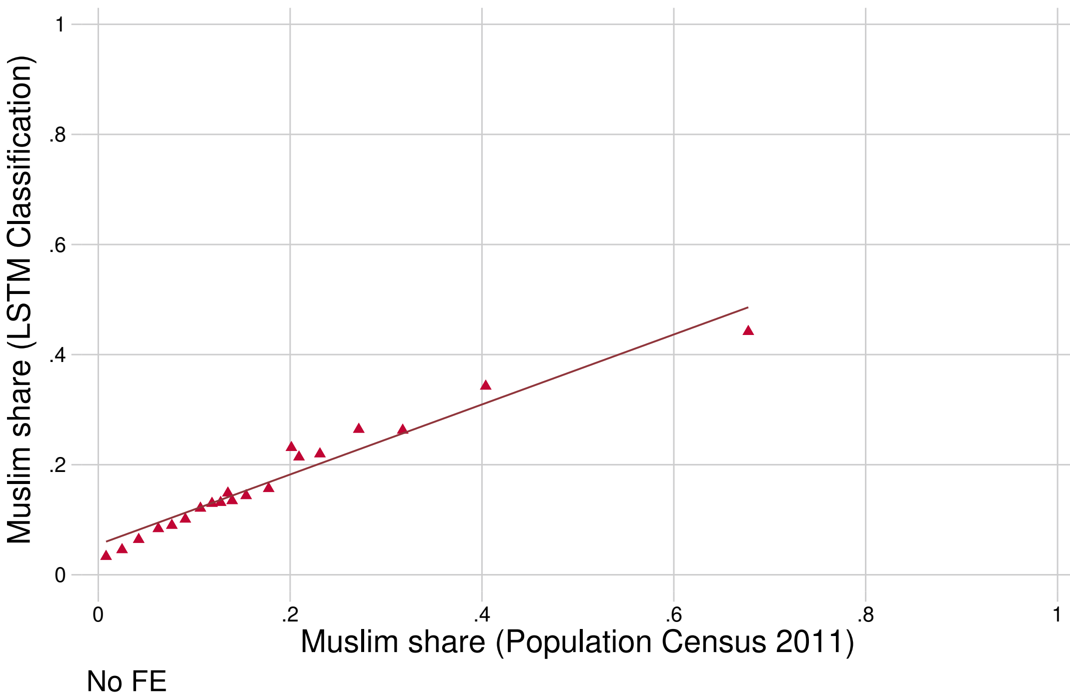
Neighborhood Population Distributions



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 less than 150 people or more than 1000.

Figure A.2

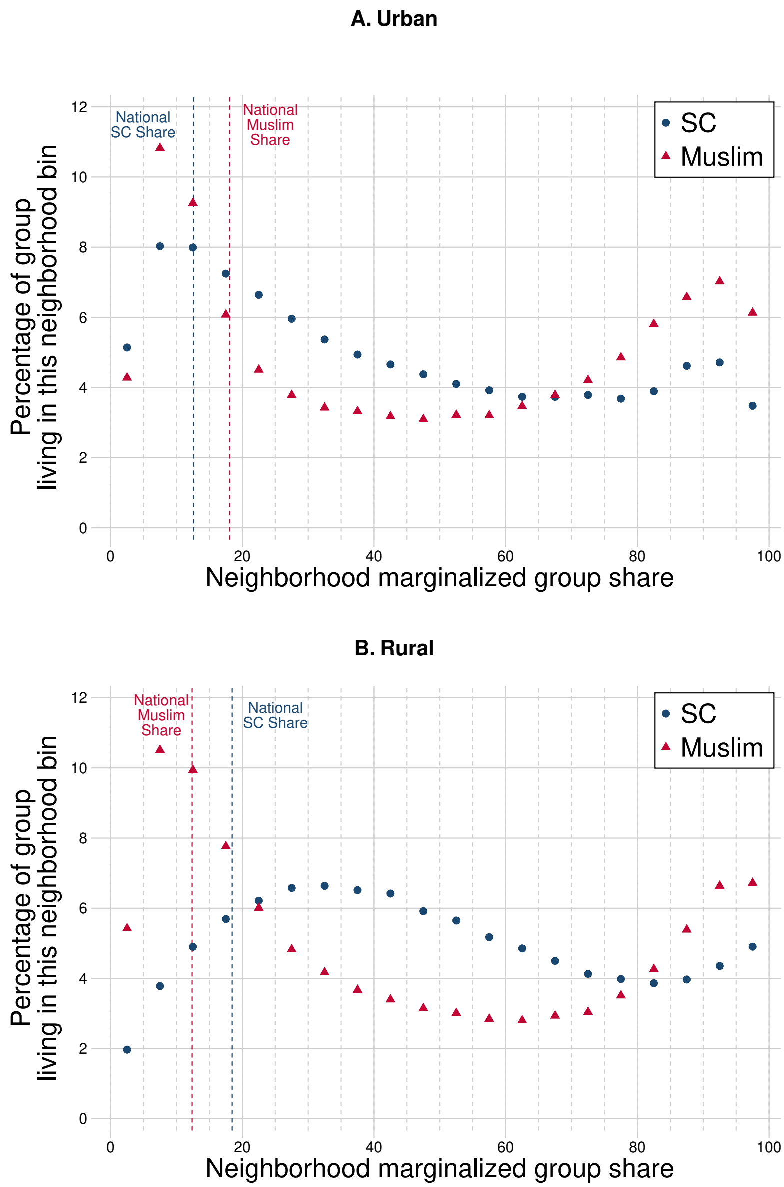
Validation of Muslim Name Classification:
Subdistrict Muslim Share in SECC vs Population Census



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.

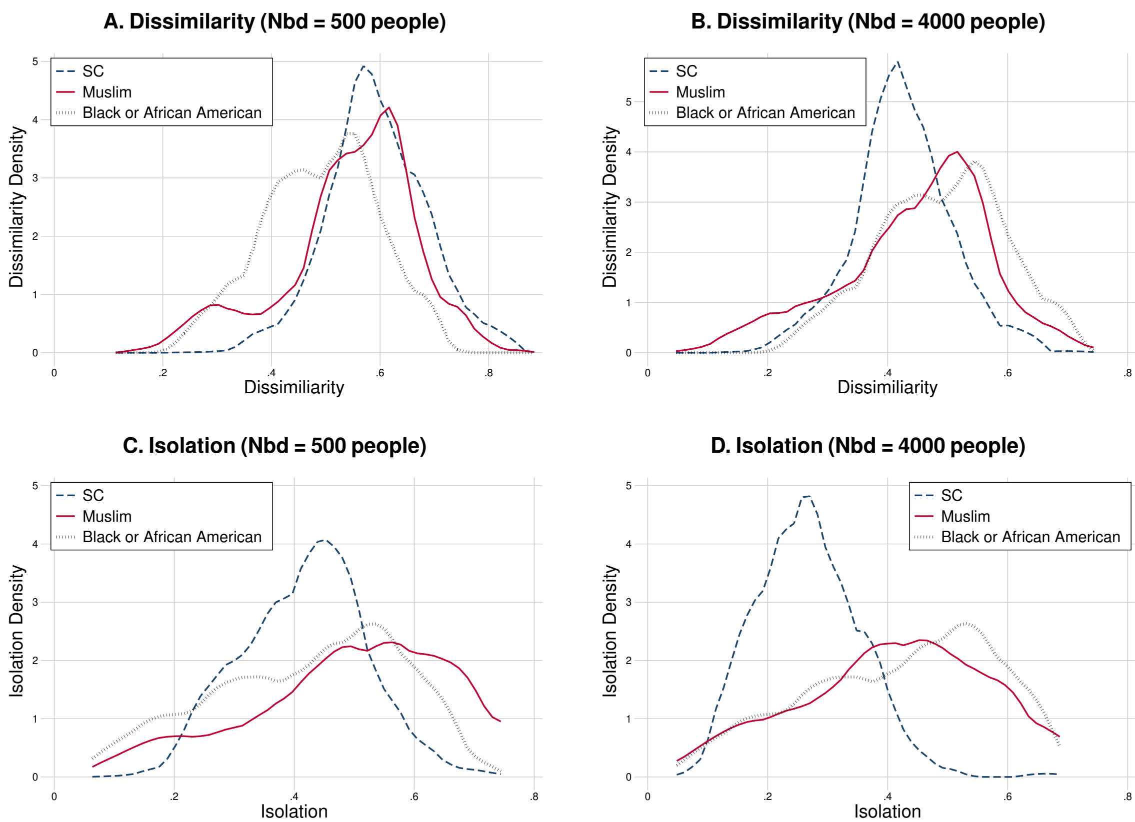
Figure A.3

Population Distribution
as a Function of Marginalized Group Share



Notes: The figure shows the distribution of Scheduled Caste 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%.

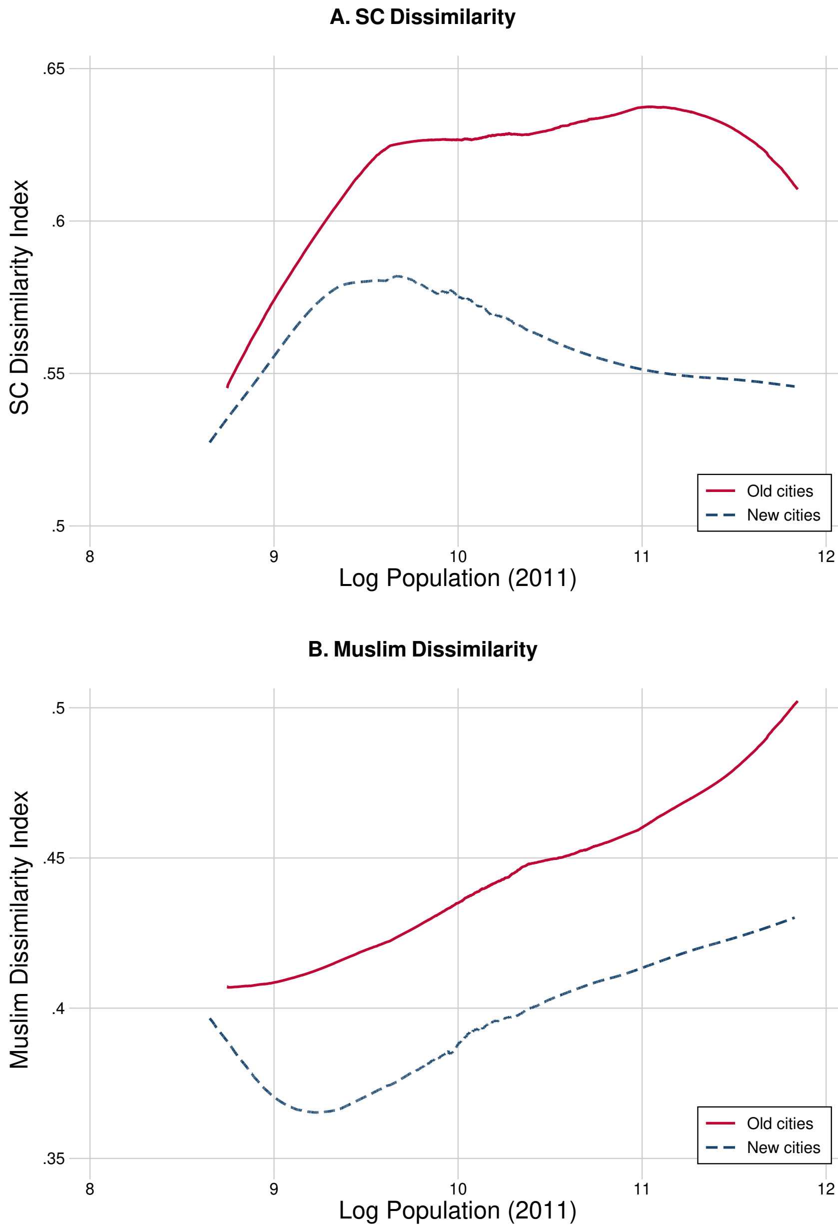
Figure A.4
Comparison of Urban Muslim and Scheduled Caste Segregation in India with Urban Black Segregation in the United States



Notes: The figure shows the distribution of the dissimilarity and isolation indices of Muslim and SCs, and compares them with similar Black/White measures in the U.S. Panels A and C define neighborhoods as enumeration blocks, which is the main definition used in the paper. Panels B and D aggregate enumeration blocks to have up to 4000 people in a neighborhood, for better comparability with the U.S. measures. All estimates are weighted by their respective marginalized group populations so that they reflect the experience of marginalized groups. All plots are calculated for the subset of Indian towns and American metropolitan statistical areas that have more than 100,000 people, to maximize comparability. Source: SECC 2012.

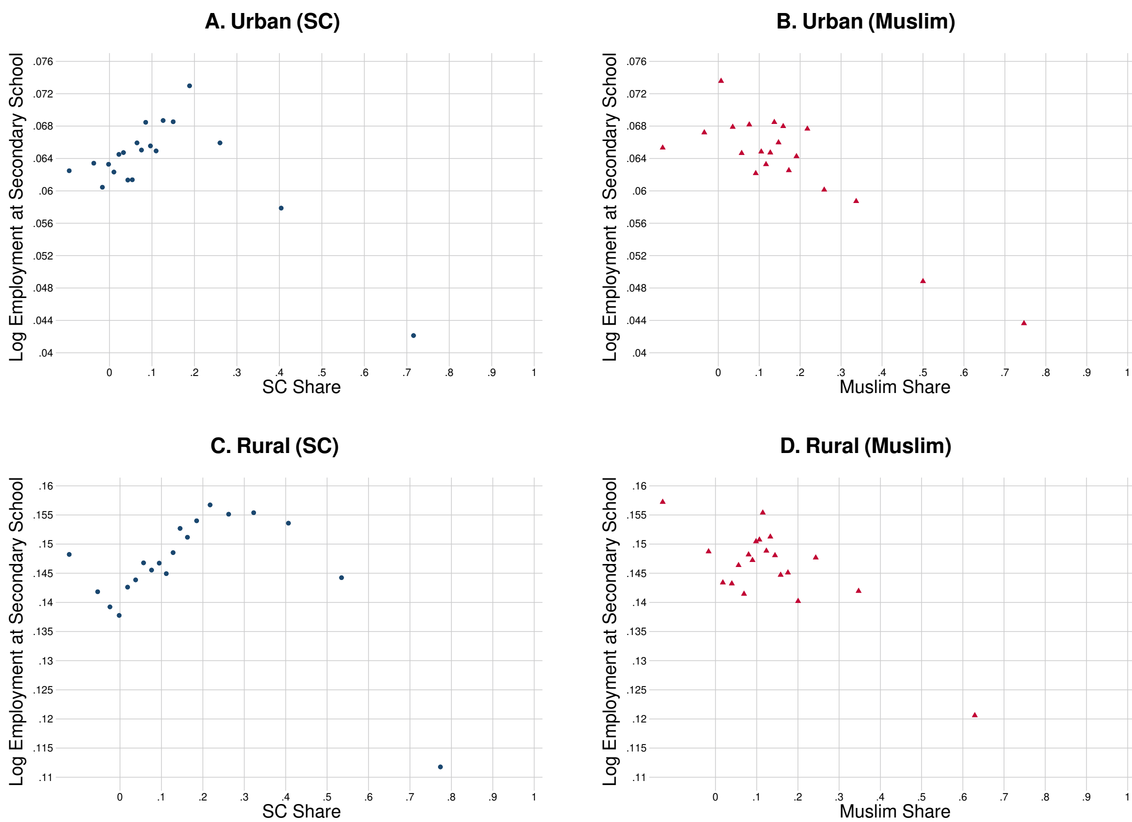
Figure A.5

Dissimilarity Indices by City Age



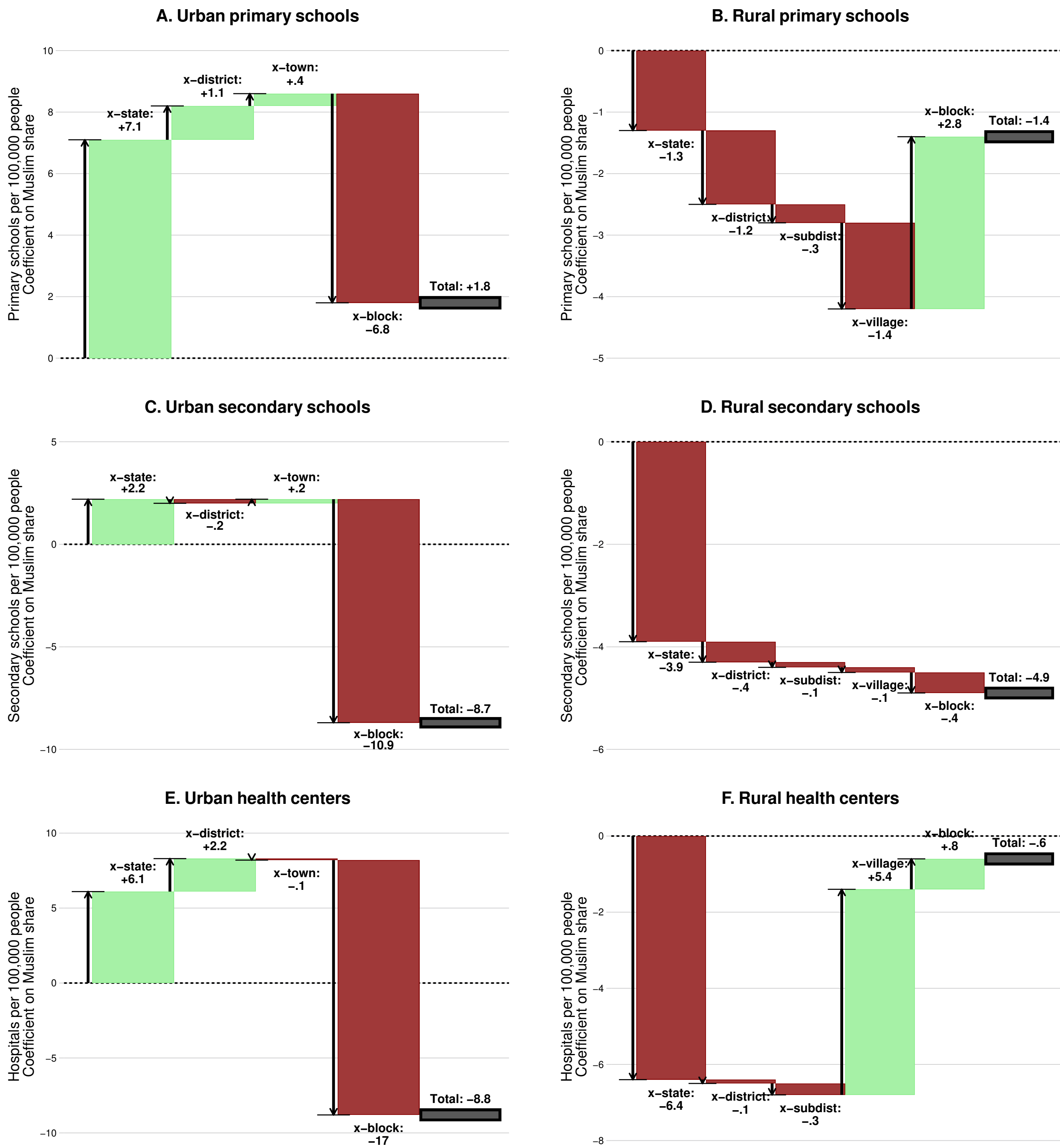
Notes: The figure shows lowess plots of dissimilarity measures against log city population, for SC and Muslim dissimilarity. Cities that are recorded in the decennial population census for the first time in 1922 or earlier are categorised as old cities. Those that are recorded in the decennial population census after 1922, are designated as new cities. Source: Population Census 2011, SECC 2012.

Figure A.6
Access to Public Secondary Schools vs.
Neighborhood Marginalized Group Shares:
Intensive Margin Estimates



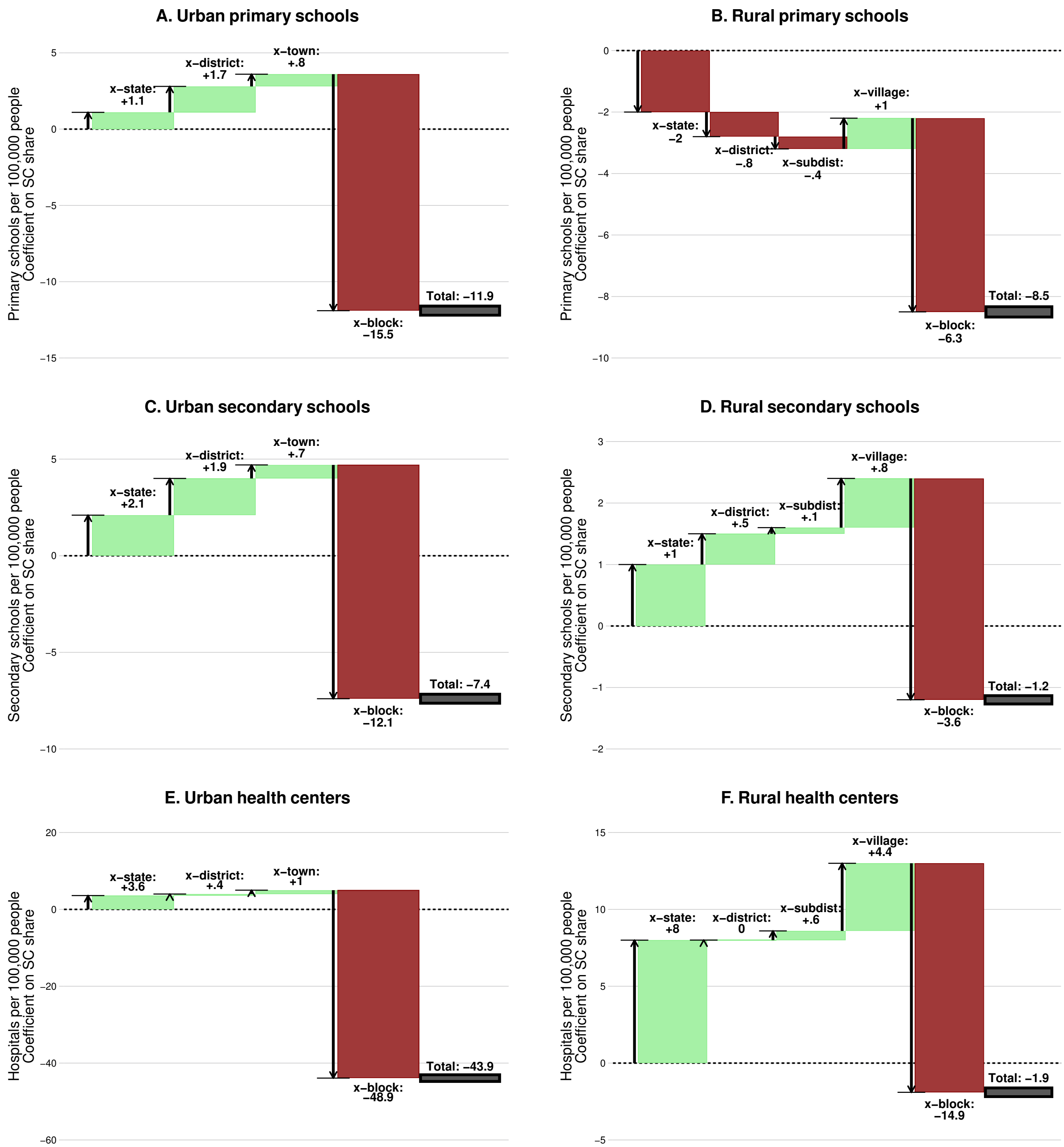
Notes: The figure shows binscatter plots of log employment (plus one) in neighborhood secondary schools as a function of the neighborhood SC or Muslim share. Source: Economic Census 2013, SECC 2012.

Figure A.7
Disparities in Private Facilities as a
Function of Muslim Share



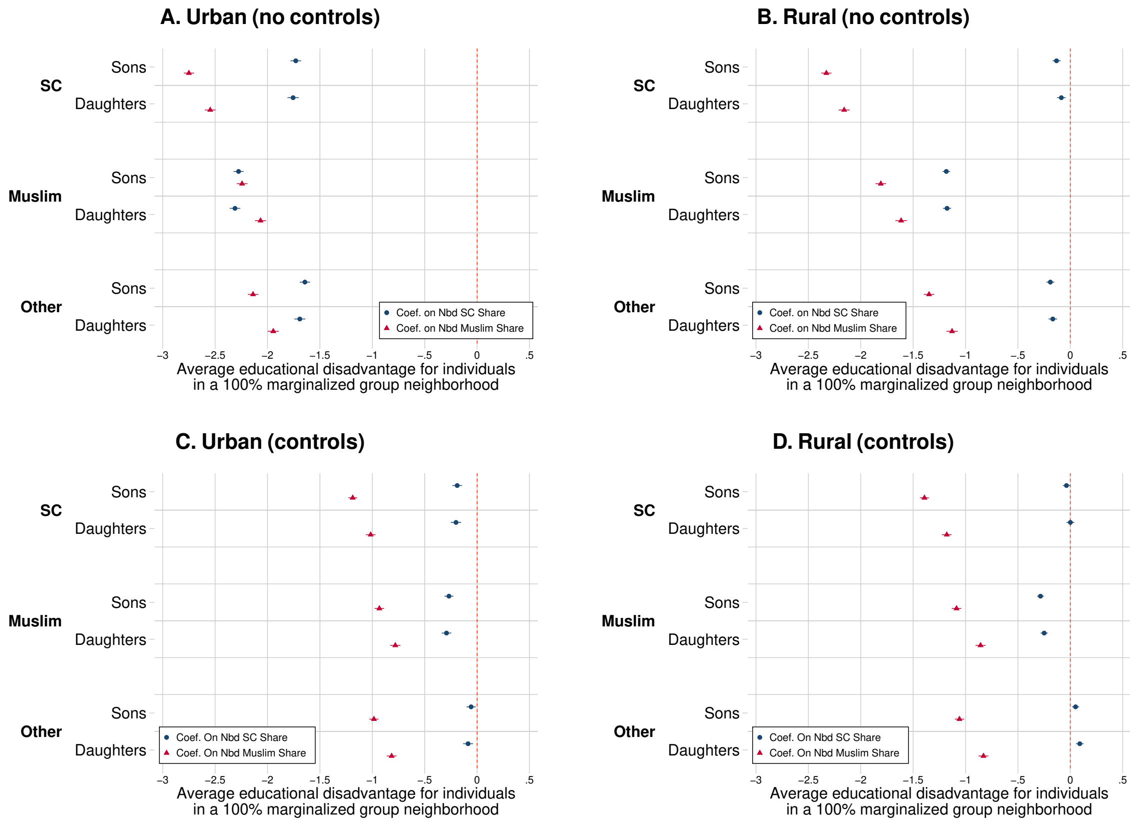
Notes: The figure describes the cross-neighborhood relationship between a neighborhood's Muslim share and a neighborhood's access to private facilities: primary and secondary schools, and health centers. The dark gray box shows the coefficient of a regression of a private facility indicator on the Muslim 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. Source: Economic Census 2013, SECC 2012.

Figure A.8
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.

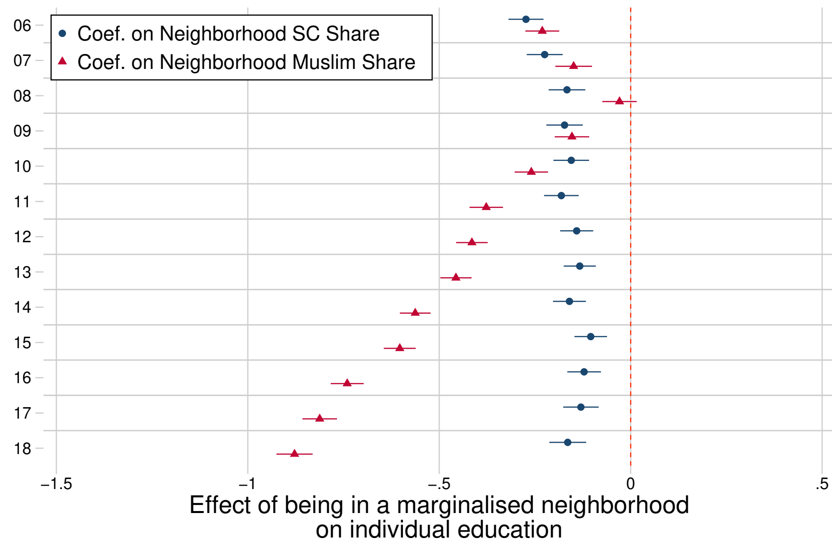
Figure A.9
Educational Attainment in Marginalized Group Neighborhoods



Notes: The figure shows coefficient estimates from Equation 4: a regression of individual years of education on the neighborhood marginalized group share. The data is a 10% sample of 17–18-year-olds from the SECC 2012. Panels A and B have no controls other than the neighborhood population. Panels C and D control for the individual's household consumption and parental education. The estimates are identical to those in Tables A.2 and A.3. All regressions include town (urban) or subdistrict (rural) fixed effects. Data source: SECC 2012.

Figure A.10

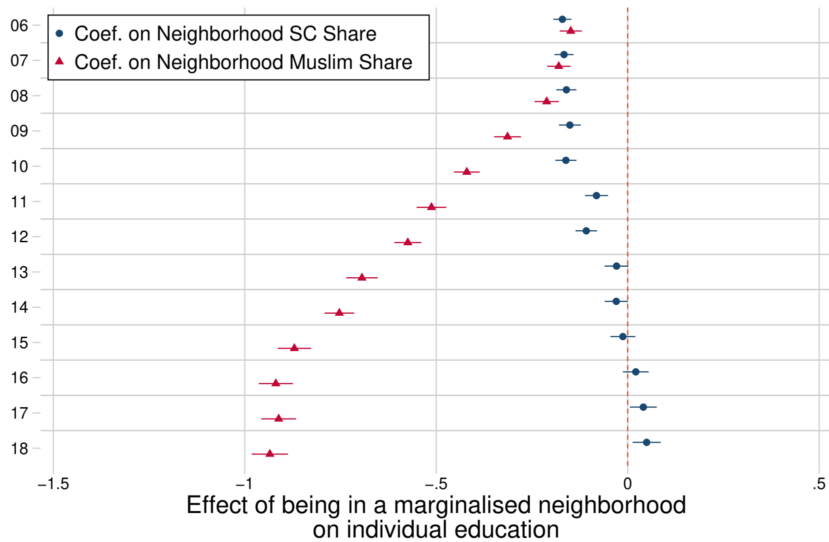
Educational Attainment in Marginalized Group Neighborhoods:
Urban Estimates by Age



Notes: The coefficient plot shows estimates from a regression of individual education on the neighborhood marginalized group share. These are identical to estimates from Table 8, but calculated for children at different ages (as indicated on the Y axis). Source: SECC 2012.

Figure A.11

Educational Attainment in Marginalized Group Neighborhoods: Rural Estimates by Age



Notes: The coefficient plot shows estimates from individual-level regressions of individual education on the neighborhood marginalized group share. These are identical to estimates from Table 8, but calculated for children at different ages (as indicated on the Y axis). Source: SECC 2012.

Table A.1

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 5. 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.2

Educational Attainment in Marginalized Group Neighborhoods, by Social Group: Urban

A. Young Men 17–18

	(1) All	(2) SC	(3) Muslim	(4) Other	(5) All	(6) SC	(7) Muslim	(8) Other
Neighborhood SC Share	-1.621***	-1.996***	-1.967***	-1.401***	-0.143***	-0.590***	0.116	0.164***
	0.026	0.039	0.059	0.035	0.023	0.037	0.060	0.031
Neighborhood Muslim Share	-2.156***	-1.973***	-2.051***	-2.224***	-0.917***	-0.711***	-0.686***	-0.952***
	0.026	0.084	0.035	0.034	0.023	0.082	0.033	0.031
Father's Education					0.194***	0.203***	0.214***	0.178***
					0.001	0.002	0.002	0.001
Mother's Education					0.080***	0.063***	0.118***	0.076***
					0.001	0.002	0.002	0.001
Log of per capital hh consumption					-0.171***	-0.328**	0.170*	-0.115**
					0.035	0.104	0.078	0.043
Observations	3023922	380627	554382	2088913	1965454	249868	366217	1349369
R2	0.16	0.17	0.20	0.12	0.33	0.33	0.38	0.28
Town FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

B. Young Women 17–18

	(1) All	(2) SC	(3) Muslim	(4) Other	(5) All	(6) SC	(7) Muslim	(8) Other
Neighborhood SC Share	-1.646***	-2.070***	-1.950***	-1.417***	-0.155***	-0.657***	0.112	0.147***
	0.028	0.040	0.062	0.038	0.025	0.040	0.063	0.034
Neighborhood Muslim Share	-1.978***	-1.752***	-1.861***	-2.041***	-0.764***	-0.588***	-0.535***	-0.815***
	0.027	0.085	0.037	0.036	0.025	0.086	0.035	0.033
Father's Education					0.190***	0.199***	0.202***	0.175***
					0.001	0.002	0.002	0.001
Mother's Education					0.096***	0.074***	0.134***	0.091***
					0.001	0.002	0.002	0.001
Log of per capital hh consumption					-0.593***	-0.729***	-0.178*	-0.540***
					0.038	0.117	0.083	0.047
Observations	2626264	333251	513803	1779210	1716267	217380	342672	1156215
R2	0.17	0.19	0.22	0.13	0.34	0.34	0.38	0.29
Town FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows estimates from individual-level regressions of education on the urban neighborhood marginalized group share. The estimates are analogous to those in Table 8, but estimated separately for members of each social group. Source: SECC 2012.

Table A.3

Educational Attainment in Marginalized Group Neighborhoods, by Social Group: Rural

A. Young Men 17–18

	(1) All	(2) SC	(3) Muslim	(4) Other	(5) All	(6) SC	(7) Muslim	(8) Other
Neighborhood SC Share	-0.001	-0.411***	-0.553***	0.507***	0.028	-0.331***	-0.017	0.344***
	0.020	0.030	0.044	0.026	0.018	0.029	0.044	0.023
Neighborhood Muslim Share	-1.573***	-0.806***	-1.539***	-1.488***	-1.029***	-0.627***	-1.033***	-0.815***
	0.026	0.065	0.036	0.037	0.023	0.064	0.033	0.033
Father's Education					0.275***	0.262***	0.299***	0.267***
					0.001	0.001	0.002	0.001
Mother's Education					0.071***	0.059***	0.113***	0.068***
					0.001	0.002	0.002	0.001
Log of per capital hh consumption					0.627***	0.570***	0.671***	0.601***
					0.020	0.050	0.052	0.024
Observations	4889326	852415	677549	3359362	3391599	611930	465140	2314529
R2	0.16	0.16	0.21	0.15	0.31	0.27	0.34	0.30
Subdistrict FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

B. Young Women 17–18

	(1) All	(2) SC	(3) Muslim	(4) Other	(5) All	(6) SC	(7) Muslim	(8) Other
Neighborhood SC Share	0.050*	-0.466***	-0.482***	0.646***	0.068***	-0.364***	0.060	0.419***
	0.021	0.032	0.048	0.029	0.019	0.031	0.047	0.025
Neighborhood Muslim Share	-1.375***	-0.555***	-1.319***	-1.326***	-0.800***	-0.451***	-0.750***	-0.693***
	0.028	0.071	0.039	0.040	0.025	0.070	0.036	0.036
Father's Education					0.294***	0.271***	0.300***	0.290***
					0.001	0.002	0.002	0.001
Mother's Education					0.116***	0.095***	0.164***	0.111***
					0.001	0.002	0.003	0.001
Log of per capital hh consumption					0.518***	0.506***	0.423***	0.528***
					0.022	0.056	0.056	0.026
Observations	4075086	685423	600899	2788764	2789113	484699	407228	1897186
R2	0.18	0.20	0.22	0.17	0.35	0.32	0.36	0.34
Subdistrict FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows estimates from individual-level regressions of education on the rural neighborhood marginalized group share. The estimates are analogous to those in Table 8, but estimated separately for members of each social group. Source: SECC 2012.