IIMB-WP N0. 712/2024

WORKING PAPER NO: 712

On Measuring Muslim Segregation in Urban India

Arpit Shah
Indian Institute of Management Bangalore
Harvard University
arpit.shah@iimb.ac.in

Anish Sugathan
Indian Institute of Management Ahmedabad
anishs@iima.ac.in

Naveen Bharathi
Indian Institute Technology Bombay
University of Pennsylvania
naveenbharathi@iitb.ac.in

Andaleeb Rahman
Cornell University
ar687@cornell.edu

Amit Garg
Indian Institute of Management Ahmedabad
amitgarg@iima.ac.in

Deepak Malghan
Indian Institute of Management Bangalore
Stockholm Environment Institute Asia Centre
dmalghan@iimb.ac.in

Year of Publication – September 2024

On Measuring Muslim Segregation in Urban India

Arpit Shah*[1,2], Anish Sugathan†[3], Naveen Bharathi‡[4,5], Andaleeb Rahman§[6], Amit Garg¶[3], and Deepak Malghan‖[1,7]

[1]Indian Institute of Management Bangalore
[2]Harvard University
[3]Indian Institute of Management Ahmedabad
[4]Indian Institute of Technology Bombay
[5]University of Pennsylvania
[6]Cornell University
[7]Stockholm Environment Institute Asia Centre

Abstract

The spatial segregation of Muslims in urban India is central to their social, economic, and political marginalization. However, the quantitative characterization of Muslim segregation has suffered from the lack of readily available demographic data at high spatial and temporal resolutions. This paper demonstrates the feasibility of accurately quantifying Muslim segregation in urban India using the latest electoral rolls data from Bengaluru (a megapolis of over 13 million residents) and an improved open-source algorithm to identify Muslim names. Our approach provides significant improvements over past efforts in this regard. We introduce two new metrics (diversity and local divergence) to account for substantial intra-city variation in the spatial segregation of Muslims. Our analysis suggests that the threefold ghetto-enclave-mixed taxonomy that the extant literature has quantified for entire towns can be found within large urban agglomerations such as Bengaluru. Our quantitative framework for Muslim segregation helps uncover the complex relationship between segregation and the ghettoization of Muslims in urban India. Our measurement framework uses publicly available data and can be applied to study segregation patterns across urban India.

Keywords: Residential Segregation, Racialization of Muslims, Electoral Data, India

*arpit.shah@iimb.ac.in
†anishs@iima.ac.in
‡naveenbharathi@iitb.ac.in
§ar687@cornell.edu
¶amitgarg@iima.ac.in
‖dmalghan@iimb.ac.in (corresponding author)

Acknowledgments

This paper would not have been possible without the sustained generosity of Raphael Susewind. Over several months, he unfailingly answered many questions that we had about data and methods in Susewind(2017), the starting point for our analysis. Susewind also shared GIS files corresponding to the Voronoi interpolation outputs published in Susewind(2017), thus allowing replication. We would also like to thank Rochana Chaturvedi and Sugat Chaturvedi for sharing their algorithm that we used to classify names into Muslim and non-Muslim categories. They answered multiple questions that we had regarding their published algorithm (Chaturvedi and Chaturvedi, 2023), and were kind enough to share their code and a pre-trained algorithm that we were able to use on the Bangalore voter rolls data. Arunima Malghan helped with the scraping of electoral rolls data. The usual disclaimers apply. Research reported here was supported by a grant from the EGAP network (https://tinyurl.com/egapAward).

On Measuring Muslim Segregation in Urban India

Abstract

The spatial segregation of Muslims in urban India is central to their social, economic, and political marginalization. However, the quantitative characterization of Muslim segregation has suffered from the lack of readily available demographic data at high spatial and temporal resolutions. This paper demonstrates the feasibility of accurately quantifying Muslim segregation in urban India using the latest electoral rolls data from Bengaluru (a megapolis of over 13 million residents) and an improved open-source algorithm to identify Muslim names. Our approach provides significant improvements over past efforts in this regard. We introduce two new metrics (diversity and local divergence) to account for substantial intra-city variation in the spatial segregation of Muslims. Our analysis suggests that the threefold ghetto-enclave-mixed taxonomy that the extant literature has quantified for entire towns can be found *within* large urban agglomerations such as Bengaluru. Our quantitative framework for Muslim segregation helps uncover the complex relationship between segregation and the ghettoization of Muslims in urban India. Our measurement framework uses publicly available data and can be applied to study segregation patterns across urban India.

Keywords: Residential Segregation, Electoral Data, Religion, Bengaluru, India

Introduction

The spatial segregation of Muslims in urban India is a central constitutive element of their ongoing social, political, and economic marginalization (Jaffrelot and Gayer, 2012). Muslims in India continue to remain a stigmatized minority with poor economic and educational attainments (Mishra Committee, 2007). The Muslim "ghetto" in urban India has been recognized by an official government committee that studied the economic and educational attainment gaps among Indian Muslims (Sachar Committee, 2006). Spatial isolation is central to how Muslims in India are increasingly denied full and equal citizenship rights (Chatterjee, 2017; Rahman and Kumar, 2023). The spatial segregation of Muslims is also intimately linked to episodes of sectarian religious violence and riots—referred to as "communal violence" in South Asia—that disproportionately impact Muslims (Dhattiwala, 2019). It has also been noted that religious violence in India is predominantly an urban phenomenon, making it essential to understand urban residential segregation based on religion (Galonnier, 2015).

The study of Muslim segregation in India has been dominated by detailed ethnographic accounts (Jaffrelot and Gayer, 2012; Jamil, 2017). The Indian national census does not release intra-city information on religious identity. While attempts have been made to use large-sample surveys to shed light on patterns of Muslim segregation (Chakravorty and Sircar, 2021), they cannot match the insights possible from census data. This paper uses the latest (2022-24) revision of electoral roll data to provide a detailed quantitative analysis of Muslim segregation in Bengaluru, a megapolis of 13 million residents. Our work improves upon Susewind's (2017) pioneering effort by leveraging recently released electoral geography maps to illustrate how improved methods and metrics can provide detailed insights into Muslim segregation in urban India.

We make both methodological and empirical contributions. Methodologically, we show how information in India's publicly released electoral rolls can be combined with electoral geography maps to provide a more accurate picture of spatial demography at the local level for urban India. Prior work on Muslim segregation has relied on the locations of electoral polling stations to determine the spatial religious demography of the cities being studied (Susewind, 2017). However, the location of the polling station is not the same as the region that the polling station covers. All polling stations in urban India cover voters from different areas in the town. These regions comprise the entirety of the city and are mutually exclusive in their geographical coverage. We leverage recently released data from the Karnataka State Election Commission that provides both the polling station location and the boundary that the polling station covers. The transition from point locations to actual regions provides a richer and more accurate picture of religious spatial demography in urban India. For instance, to understand the population density or access to public services that different religious groups might have, one would need information on religious demography *and* the boundaries/regions in which groups reside.

In addition, we leverage a recently released open-source algorithm to identify Muslim names from India's electoral rolls to develop a granular understanding of local demography (Chaturvedi and Chaturvedi, 2023). This algorithm offers higher classification speeds, comparable accuracy levels (in terms of specificity and sensitivity) to older algorithms, and greater coverage in terms of names classified. Our approach, which we demonstrate in the case of Bengaluru, can be applied to understand segregation using electoral data for the rest of the country. A detailed picture of spatial demography can also facilitate a richer understanding of how public services are impacted by residential segregation. Fundamentally, public services such as water or sewerage are provided at the neighbourhood level, and having accurate demographic information at the local level is crucial to understanding and addressing some of the most pernicious impacts of residential segregation (Trounstine, 2018; Bharathi et al., 2022).

Empirically, we provide a detailed characterization of intra-city Muslim segregation in Bengaluru at a neighbourhood-scale. In doing so, we introduce two new metrics to understand segregation quantitatively. Jaffrelot and Gayer (2012) and Susewind (2015) have argued that Muslim residential choices in urban India follow a threefold "ghetto-enclave-mixed" taxonomy. Gayer and Jaffrelot (2012) suggest that Muslim ghettos in Indian cities are characterized by political and social constraints over residential choices, neglect of the segregated area by authorities, regrouping of individuals based on ethnic/religious identity, estrangement from the city, and a sense of closure for the residents. Areas like Juhapura in Ahmedabad are considered typical of Muslim ghettos in India (Thomas, 2012). Enclaves have characteristics like ghettos but are characterized by the voluntary choice of residents to live in segregated areas. Mixed areas present the conceptual contrast to ghettos and enclaves. Galonnier (2015) studies Muslim segregation in the north Indian town of Aligarh and found that the "enclave, the citadel, and the ghetto" can be coterminous. While Susewind (2017) quantified entire towns using the threefold taxonomy, our results show that ghettos, enclaves, and mixed areas can be found *within* large urban agglomerations such as Bengaluru. Thus, our improvements to Susewind (2017) allow us to develop a nuanced portrait of intra-city variations in patterns of Muslim segregation.

The choice of Bengaluru as our empirical site is not only motivated by the availability of the best spatially explicit electoral data of any significant urban center in India but also by the spatial histories of Muslims in the city that help illustrate the utility of the methods we develop here.[1] An evolving economic landscape in a society rooted in strict occupational segregation and a long history of religious violence have driven neighborhood-scale religious demography in Bengaluru --- trajectories that are hardly sui generis (Chatterjee, 2009; Bhagat, 2018). Our central focus here is to bridge the gap between granular neighborhood-scale ethnographic accounts of Muslim segregation in India on the one hand and macro city-level summaries offered by quantitative studies on the other. The segregation measurement framework that we develop here explicitly recognizes that Muslim segregation in urban India is multi-scalar. While detailed ethnographic accounts can help describe how and why segregation occurs at micro-scales, our work complements such efforts by providing a granular portrait of the variation in segregation across the city at multiple geographical scales. Our multi-scalar approach also provides more nuanced insights than the macro portraits offered by large scale quantitative studies that analyse segregation at city-scale. While our multi-scalar quantitative framework can sketch the plausible phenomenological trajectories of spatial segregation, it is crucial to note that it is not a substitute for “boots on the ground” ethnography.

As an illustration, consider the historical spatial trajectories of neighbourhoods in the D.J. Halli ward of the city, which was the site of the last major violence targeted at Muslim residents in Bengaluru (Joshi, 2020). The high level of within-ward Muslim residential segregation in D.J. Halli is partly a result of previous communal riots in the ward, with Muslim residents fleeing mixed neighbourhoods to the perceived safety of segregated neighbouhoods.[2] This intra-ward segregation is also rooted in the historical employment patterns of Muslims and Dalits in the tannery industry. Over the decades, the primary demographic change in D.J. Halli has been the displacement of Dalits by incoming Muslim residents. These demographic processes can only be fully understood by measuring segregation at multiple scales, including neighbourhood scales.

While the state of Karnataka and its capital city Bengaluru have historically witnessed lower communal violence than the North of India immediately after independence (Varshney, 2003), there has been growing religious polarization in Bengaluru in recent decades. Attempts to drive the marginalization of Muslims have been accompanied by efforts leading to the 'othering' of Muslims. For instance, a recent decision by the Karnataka State Government banned Muslim women from wearing hijabs (head scarves) in public schools (Biswas, 2022). The usually insulated capitalist class in Bengaluru has expressed worry about the impacts of polarization on their businesses, with a rare but widely publicized comment from a billionaire urging the state government to 'resolve the growing religious divide' (Biswas, 2022). The othering of Muslims is linked to their spatial segregation. For instances, calls for the economic boycott of Muslim businesses (Biswas, 2022), and variations in the provisioning of public services like water and sanitation share a dialectical relationship with the spatial isolation of Muslims (Bharathi et al, 2022).

Another important motivation for our segregation measurement comes from ethical concerns surrounding the use of spatial maps representing Muslim demography in urban

India (discussed in detail below). When ethnographic accounts use spatial maps, they are typically limited to, at most, a handful of neighborhoods. However, the large-scale quantitative data we use here can potentially be used to develop spatial maps of Muslim demography for entire cities. We demonstrate that our measurement framework yields crucial sociological insights without the need for such spatial representation that is fraught in an environment like contemporary India, where Muslims are systematic targets of organized violence.

Data and Methods

Election Geography in India

As our segregation measurement relies on spatially explicit voter data, we summarize India's complex electoral geography. India is the world's largest electoral democracy, with over 900 million voters. Operationalizing universal adult franchise in a large and diverse country has led India to pioneer new institutions and methods for election ground operations (Shani, 2017). A key cog in India's complex electoral machinery is the well-honed election geography that divides India's large landmass into street-level election precincts (ECI, 2019, 2020). Our primary unit of segregation measurement is the polling area (PA) — a uniquely defined areal unit containing 1000– 1500 voters in urban India. Every polling area is assigned a unique polling station where voters physically vote on election day (ECI, 2020, §2.3). We provide a detailed glossary of terms used in India's electoral geography in Table A1 (Online Appendix).

Polling areas are particularly attractive elementary spatial units for studying Muslim segregation in urban India. The Election Commission of India implicitly accounts for neighbourhoods numerically dominated by "minorities" — a term used in India to refer to citizens not belonging to the majority Hindu religion (ECI, 2020, §2.6.4, §7.2.1). Muslims are by far the most significant minority demographic group. The Election Commission requires that neighbourhoods "predominantly inhabited" by minority groups be enumerated in a polling station located within these neighbourhoods to ensure that minority groups are not intimidated and that their franchise is protected (ECI, 2020, §7.2.1). The Chief Electoral Officer (CEO) in every state is required to delineate polling areas and polling station locations to ensure that minorities and "weaker sections" (a term used to refer to groups at the bottom of India's caste totem) are not disenfranchised. Among other measures, the CEO must work with these groups' civic leaders to delineate polling areas (ECI, 2020, §7.2.1). Thus, polling areas (especially in urban areas) capture minority residents' "mental maps" of their neighbourhood (Gould and White, 2012).

Polling Areas

We used two sources of electoral data in this study. First, we followed Susewind (2017) and extracted information from publicly available 2022-23 electoral rolls for Bengaluru from the website of the Chief Election Commissioner of the state of Karnataka (henceforth, CEC-K Rolls). We used the Tesseract OCR Engine with custom extensions (implemented in Python and R) to scrape voter names and polling station location information from the CEC-K Rolls. The CEC-K Rolls also provide the street address or neighbourhood descriptions (typically a collection of streets) constituting unique polling areas. Polling areas in urban India are wholly contained within wards (the smallest political boundary). We extracted ward information (that then uniquely maps to the assembly and parliamentary constituency). Second, we used geo-registered polling area boundaries recently made publicly available by the Karnataka State Election Commission.[3] This publicly available map also contains details about the polling station, including the address of the polling station building housing the polling station.

Merging the geo-registered polling boundaries with the CEC-K Rolls data was challenging and required significant amounts of iterative hand-coding. While the CEC-K Rolls data contain the polling station address, they do not provide the unique identifier used by the Election Commission to identify a polling station. Additionally, the street addresses across the two datasets do not match perfectly. For example, a "government exercise school" in one data set might be referred to as a "Vyayamshala" (the equivalent term in the local language) in the other dataset. A polling station building address could be called "Room no 1, Government Primary School, Hosur Main Road" in one dataset and "Hosur Road, Government School, Room 1" in the other.

Further, one or both datasets could record polling station building addresses as simply a generic "Government Primary School" or "Anganwadi Kendra" (public childcare center). While these incomplete address fields do not pose any ambiguity for the actual conduct of elections, they pose significant challenges for merging data across the city — the government operates hundreds of primary schools and childcare centers in the city. These idiosyncrasies made it impossible to algorithmically merge our two datasets (using fuzzy string-matching procedures, for example).

Without a clear algorithmic pathway for merging our two datasets, we developed an efficient procedure for hand-merging. First, we inserted unique ward and assembly constituency identifiers (by hand-matching names across the two datasets). We then used fuzzy string-matching algorithms within each ward to support our hand-matching and iteratively distill a narrower possibility set. If the match was still ambiguous, we used neighbourhood and street extent data on the rolls (ref Fig A1 in the Online Appendix). For example, the Anganwadi Kendra in the Rayapuram ward of the Chamrajapet assembly constituency in Central Bangalore covers the 13th Main Road and Blocks B and C near the Venkataswamy Garden neighbourhood. We used Open Street and Google Maps to search for these specific streets and neighbourhoods. For Bengaluru, we matched data for 7,526 out of 7,529 polling areas (or coverage of over 99.96 %) using these procedures.

We have described the data-merge process in detail to underscore that it is possible to geocode polling areas using only information in the electoral rolls' "parts" data. While admittedly laborious and time-consuming, we want to emphasize that the procedure described here is not Bengaluru-specific and can be applied to any town in India. It is also important to note that in using actual polling area boundaries, our approach improves upon Susewind (2017), who used point locations of polling stations to represent polling areas. Without polling boundary information, Susewind (2017) relies on spatial interpolation to estimate each polling station's polling area. Moving from point locations to polling areas provides a more accurate and complete picture of spatial demography at the neighbourhood-level. The gains in accuracy and spatial resolution by shifting to polling boundary data can be mathematically demonstrated (for brevity, we have not reported the accuracy gains in this work).

Muslim Name Classification

Next, we used a recently released algorithm (Chaturvedi and Chaturvedi, 2023) to classify individual voters into Muslim and non-Muslim categories (referred to as the C2 algorithm in the rest of the manuscript). Susewind (2015) uses a string-matching algorithm called name2community based on a reference list to classify names into religious categories. Being a dictionary-based approach, name2community cannot classify unseen names and has documented issues dealing with orthographic variations. When implemented on a new dataset from Uttar Pradesh, Chaturvedi and Chaturvedi (2023) report that name2community is able to classify 74% of names, while C2 achieves 100% coverage. The C2 algorithm, rooted in onomastics, relies on the variations in linguistic roots between Muslim and non-Muslim names in South Asia. While Muslim names are derived from Classical Arabic, most non-Muslim Indian names have Indo-Aryan, Dravidian, and Tibeto-Burnam roots. In addition to improved coverage and comparable accuracy, C2 is many orders of magnitude faster than name2community (50,000 names/second vs. 0.4 names/second, respectively). For large datasets such as Bengaluru (~9.5 million registered voters), the computational gains are significant.

Susewind (2017) has clearly shown why the (registered) electorate's religious composition reflects the general population's religious composition, including children and adults not registered to vote. We do not repeat those arguments here. If anything, in the intervening years, voter registration proportions have seen a modest increase. Further, with our precise polling area identification, we are confident that our data describes the true extent of Muslim spatial segregation in Bengaluru.

Religious Demography and Cartography in Contemporary India: Ethical Concerns

High spatial-resolution religious identity data is sensitive information in contemporary India, where calls for the social and economic boycott of Muslims are commonplace. It is an ethical obligation to account for how the data and methods introduced in this paper could potentially be used in a communally polarized environment. Researchers have argued that events such as the 2002 pogrom in the Western Indian state of Gujarat were aided and abetted by voter roll data (Jaffrelot, 2007; Blom Hansen, 2008). Our data and methods could potentially also abet Muslim voter suppression (Das, 2023). We are, therefore, not publishing our map of Muslim spatial demography in Bengaluru. Our polling area map with Muslim demography was made available for peer review. It was also reviewed by civic experts in Bengaluru who have decades of local "boots on the ground" experience. The multi-scalar segregation analysis presented here characterizes the extent of local segregation without directly making high-resolution demographic composition data public. Another essential part of the data protocol we followed was not to save any individual voter demographic information on any kind of record. Our workflow involved using a published algorithm to process voter rolls where the program only returns a single number for each polling area --- Muslim population share.

Citywide Spatial Demography

The CEC-K Rolls data for Bengaluru that we have used here contained ≈ 9.5 million registered voters spread across 29 assembly constituencies (the geographical unit at which elections for State legislatures in India are conducted) and 243 wards (the lowest political and administrative unit for urban governance in India). Across the city, 12.7% of registered voters are Muslims. This proportion is comparable to the national Census of 2011 (14%) and about 4% higher than what is reported by Susewind (2017) using the 2014 electoral rolls (as classified by the name2community algorithm). Figures A2 and A3 in the Online Appendix show the polling areas for Bengaluru.

We estimate the overall Muslim dissimilarity index (*D*, Duncan and Duncan, 1955) for Bengaluru to be ≈ 0.59, which is comparable to the value of 0.57 estimated by Susewind (2017). However, this citywide measure of segregation masks important intra-city variations in how Muslims experience spatial segregation (*cf*. Online Appendix for the conceptual and computational foundations of the dissimilarity index). Detailed ethnographic portraits over the last two decades have shown how Muslim segregation in urban India is multi-scalar with significant intra-city variation (Dupont, 2004; Jaffrelot and Gayer, 2012; Jamil, 2017). Quantifying segregation at multiple scales is important in understanding the true import of inter-group contact theories (Allport et al., 1954). Contact theories have shown that intergroup contact can potentially ameliorate out-group animosity only when such contacts are sufficiently deep and sustained over time (Nathan and Sands, 2023). Such deep and sustained contact is more likely at the neighbourhood or even intra-neighbourhood scales. Intergroup contact is of particular relevance in contemporary India, which has witnessed the growing social, political, and economic marginalization of Muslims (Jamil, 2017).

To uncover patterns of intra-city variation in Muslim segregation, we use the *D*-index at sub-city scales (assembly constituencies and wards) and compute diversity and demographic divergence at the neighbourhood (polling area or PA) scale. Panel-A of Figure 1 shows the distribution of diversity (computed as entropy) across areal units in Bengaluru at three levels of spatial aggregations — the polling area (PA), the ward, and the assembly constituency (AC). Not surprisingly, the largest spatial aggregation, the ACs are the most diverse. This panel also clearly shows the bimodal nature of Muslim segregation in the city. Panel-B of Figure 1 shows how ward population shares of Muslims are only partially correlated ($r \approx 0.2$) with segregation within the ward (as measured by the *D*-index). In Panel-C of Figure 1, we unpack this partial correlation using the Kullback-Leibler relative entropy divergence (D_{KL} of Kullback and Leibler, 1951) between the demographic distributions of the polling areas and the wards containing them. The KL Divergence metric (that we normalize for straightforward interpretation) measures local PA-level segregation. It measures how a polling area's (PA) demographic composition diverges from the larger ward's. The non-monotonic relationship between ward Muslim demography and the KL Divergence metric seen in Panel-C once again shows the significant variation in how Muslim residents of Bangalore experience spatial marginalization.

[Insert Figure 1 here]

Figure 2 uses religious identity information from 9.5 million voters to show how Muslims in Bengaluru "experience" segregation differently from others. The top two panels show the relationship between diversity and local segregation. In Panel-A, we plot the normalized Kullback Leibler divergence between the polling area (PA) and assembly constituency AC) as a function of AC diversity, as experienced by ≈ 9.5 million residents in our data. In Panel-B, we replicate this at the ward level (this time using PA-Ward normalized Kullback Leibler divergence and ward diversity). At both the AC and ward levels, there is a sharp increase in the segregation of non-Muslims in the most diverse ACs and wards. This phenomenon is also seen in the bottom two panels of Figure 3 (Panels C, D), where we plot local segregation (normalized Kullback-Leibler divergence) as a function of the dissimilarity index (that measures unevenness at the AC or Ward levels). Panel-D shows that as wards get diverse, Muslims and non-Muslims grow further apart.

[Insert Figure 2 here]

Spatial Results

Diversity & Divergence

Figure 3 presents the spatial extent of multi-scalar and multi-dimensional Muslim segregation in Bengaluru. Panel A shows the spatial distribution of polling area (PA) level diversity measured as $\log_2$ entropy. Panel B depicts the PA-Ward normalized Kullback Liebler Divergence for all the PAs in our data. Both panels show respective quartiles (measured as percentages) for straightforward spatial interpretation. Representing diversity and divergence as quartiles also underscores the non-monotonic relationship these metrics share with Muslim demography (*cf.* discussion around Figure 2 above). The most diverse polling areas (quartile-4 with entropy measurements between 50% and 100 %) encompass a wide range of PA-level Muslim demographic shares (10% – 90%). We have used the information-theoretic entropy (Shannon, 1948) to measure diversity rather than the somewhat easier-to-interpret fractionalization index because the entropy measure is perfectly additively decomposable and is thus best suited for characterizing multi-scalar segregation. However, our entropy measure and fractionalization are nearly perfectly correlated (r = 0.995 for our dataset). With two subgroups (Muslims and non-Muslims in our case), a maximum entropy of 100% is the equivalent of a fractionalization of 50%. It is achieved when both groups have equal population shares.

[Insert Figure 3 here]

The true import of the Panel A in Figure 3 lies in the fact that it depicts the limited potential for intergroup contact within the neighbourhood — the kind of sustained contact central to building trust across group boundaries. Most of the neighbourhoods are homogeneous, with little scope for intergroup contact. The spatial distribution of diversity also makes clear how ghettos and enclaves are two sides of the same segregation coin. The non-monotonic relationship between Muslim demographic share and diversity means that a (non-Muslim) enclave lacks intergroup diversity as much as a (Muslim) ghetto.

Panel B in Figure 3, which shows the demographic divergence of a neighbourhood (polling area) from the larger ward, is central to the intra-city characterization of ghettoization. The Kullback Liebler Divergence metric (normalized and expressed in percentage for straightforward interpretation) is a relative entropy measure.[4] A neighbourhood (polling area) demographically dominated by Muslims in an overwhelmingly non-Muslim ward will have a high divergence. However, a similar neighbourhood in a ward with a considerable Muslim presence will have a lower divergence. Thus, demographic divergence is a direct measure of the alienation or "estrangement" (Jaffrelot and Gayer, 2012) experienced by a neighbourhood. Neighbourhood-scale intra-ward divergence that we represent in Figure 3 is especially salient. As the elementary political and administrative unit in urban India, the ward is also central to how public services such as water and sanitation are provided, and demographic divergence is implicated in patterns of discrimination and favoritism (Bharathi et al., 2022).

The role played by demographic divergence in engendering state discrimination or favoritism is one of the central conditions for the ghettoization of Muslims in urban India. As both Jaffrelot and Gayer (2012) and Susewind (2017) note, it is not merely spatial segregation but the "neglect" of segregated neighbourhoods by local authorities that creates and sustains ghettos. In Figure 3, the relationship between demographic shares and divergence is non-monotonic, with the top quartile of divergence associated with a broad range of PA-level Muslim population shares (0.28 – 0.96). Contiguous neighbourhoods with high divergence do not necessarily imply a spatial concentration of ghettos (or enclaves). For example, many neighbourhoods in the top quartile of divergence measure in Figure 3 appear connected. This spatial contiguity only suggests that all these neighbourhoods have a significant Muslim presence (a minimum of 28%

demographic share). These neighbourhoods span the three-fold “ghetto-enclave-mixed” taxonomy studied by Jaffrelot and Gayer (2012) and Susewind (2017).

Our spatial results provide large-scale evidence for the nexus between spatial segregation and the “racialization” of Muslim identity (Hochman, 2019; Sikka, 2022). We can uncover these patterns because we have quantified neighborhood-scale demographic divergence. The spatial patterns of Muslim segregation in Bengaluru are not sui generis. While accounting for regional variations is essential (especially intra-Muslim spatial segregation by class and caste), ethnographic studies in other major urban centers in India have uncovered the process by which Muslim spaces are racialized. For example, Chatterjee (2018) shows that a "[M]usholman para [neighborhood] in Kolkata does not merely mean a locality where a majority of the population is Muslim. A culturally loaded term, it rather signifies a space of difference" (emphasis not in original). The "communally defined [Muslim] neighborhoods … carry an entrenched negative characterisation, a stigma which revolved around their categorisation as communally defined spaces of difference" (Chatterjee, 2017: 6; emphasis in the original). The patterns of local divergence that we have uncovered in Bangalore (Figure 3, Panel-B) help identify such "spaces of difference." As discussed above, the local divergence method to such identification is also an ethical obligation as it would be wholly inappropriate to release actual Muslim demographic proportions on a spatial map publicly when Muslims in contemporary India are relegated to the "margins of citizenship" (Chatterjee, 2017).

Dissimilarity

The dissimilarity index (*D* of Duncan and Duncan, 1955) is, by a wide margin, the most widely used measure of segregation. While the *D* measures only one of the several "dimensions" (Massey and Denton, 1988) of segregation, its utility lies in its ease of interpretation. As a measure of "evenness" (Massey and Denton, 1988), the *D*-index simply represents the proportion of one of the two groups that must be relocated to achieve complete evenness (zero segregation). While the *D* is usually computed as a single citywide summary measure, the metric can also be used to study patterns of intra-city unevenness. The choice of the two spatial scales used is arbitrary and is a normative choice made by the researcher. For example, Susewind (2017) reports Muslim segregation dissimilarity indices at the level of municipal boundaries as well as larger urban agglomerations (in each case, the smaller area units are the relevant sets of polling areas).

In urban India, it is sociologically, politically, and economically meaningful to characterize intra-city segregation at the ward or assembly constituency level (AC). In Figure 4, we juxtapose the intra-city distribution of *D*-indices computed at assembly constituency and ward levels with the citywide *D*-indices computed by Susewind (2017). Not surprisingly, the citywide dissimilarity is greater in magnitude than intra-city measurements in Bengaluru (respective means are shown in Figure 4). On average, the demographic difference between entire cities and neighbourhoods (polling areas) is greater than such differences relative to assembly constituencies or wards that are much smaller areal units. However, this difference in means between the three distributions is not the most critical information in Figure 4. Instead, it is the fact that there is a substantial overlap between the three distributions that underscore the salience of intra-city unevenness.

[Insert Figure 4 here]

The utility of measuring intra-city dissimilarity is even more evident in Figure 5. Here, we show the spatial distribution of *D*-indices for Bengaluru computed at two scales — assembly constituency (AC), and ward. While Susewind (2017) calculates dissimilarity at the city-level to classify cities into the ghetto-mixed-enclave taxonomy, Figure 5 shows intra-city variations in dissimilarity at the ward and assembly-constituency scales. First, Panel-A shows that it is never the case that every Muslim neighbourhood in a city is a ghetto (enclave). All available evidence shows that Muslim segregation cuts across class lines (Dupont, 2004; Jamil, 2017). Second, the comprehensive five-fold criteria developed by Jaffrelot and Gayer (2012) is most meaningful when segregation is measured at the most local political units (for example, as the demographic between neighbourhoods and wards).

Our discussion around Figures 3-5 shows how not accounting for multi-scalar segregation can lead to biases similar to those induced by the familiar modifiable areal unit problem or the MAUP (Openshaw, 1984). In our case, the MAUP-type ecological inference problem concerns the appropriate comparison unit for measuring dissimilarity. Should the neighbourhood unit (polling area) be compared to the ward, the assembly constituency, the municipal boundary of the city, or the urban agglomeration? Moving from an aspatial *D*-index to some spatial variant of the *D*-index does not solve this problem. We are not dealing with an analytic question but one about "preanalytic vision" (Schumpeter, 1954). Before she begins an analysis of segregation, the researcher must make a normative choice on *why* she is measuring segregation. In the present case, if the primary focus of segregation analysis is to understand how segregation is associated with Muslim marginalization in urban India, we have shown that we need to focus on finer intra-city spatial scales. The spatial structure of segregation in Figure 5 also shows how any study of Muslim segregation in urban India must pay attention to the “fractal” (Bharathi et al., 2021) geometry of Indian urban space. With fractal segregation, patterns of segregation repeat at every spatial scale. In our case, assembly constituencies within the city are segregated; wards within assembly constituencies are segregated, and neighbourhoods (polling areas) within wards are segregated.

The maps in Figures 3 and 5 provide evidence of the power of segregation portraits that use neighbourhood polygons. Using actual neighbourhood (polling area) boundaries is what allowed us to connect aspatial and spatial portraits of Muslim segregation. Describing multidimensional segregation using diversity and local divergence is most meaningful when they can also be mapped spatially. Our results suggest that using actual neighbourhood boundaries in conjunction with multi-scalar segregation measurement is a necessary preliminary to investigate the complex relationship between segregation and residents’ subjective perceptions of their residential space. Beyond phenomenological fidelity, the spatial mapping of neighborhood-scale segregation can help account for spatial inequality in the provisioning of public services such as water and sanitation. An ethnographic study of public services in Bengaluru found that segregation at two different scales modulates such inequality (Bharathi et al., 2022). While local neighborhood-scale "micro" segregation can help neighborhoods overcome intergroup collective action problems and mount effective *demand* for public services, "macro" segregation at higher levels (wards, for example) can engender state discrimination or favoritism in the *supply* of these services (Bharathi et al., 2022).

Discussion and Conclusion

Using updated data and methods, we provide a detailed characterization of Muslim residential segregation in Bengaluru, building on the seminal contributions to studying Muslim segregation in urban India (Jaffrelot and Gayer, 2012; Susewind, 2017). One of Susewind's (2017) central findings was the possible disjunction between citywide quantitative segregation measures and the ghettoization of Muslims as recorded by detailed ethnographic accounts (Jaffrelot and Gayer, 2012). We developed a quantitative framework allowing a more direct test of this disjunction hypothesis. Susewind (2017) (rightly) emphasized the centrality of residents' "mental maps" (Gould and White, 2012) in studying the true import of spatial segregation. We have placed these mental maps at the heart of our quantitative framework in four ways. First, our segregation analysis demonstrated a novel methodology to combine electoral roll data with electoral geography maps applicable across India. Our approach provides improved accuracy and more granular insights into segregation. Specifically, we can capture the mental maps of everyday lived reality — especially for the most spatially marginalized groups. Any spatial study of the "urban outcast" (Wacquant, 2008) must be rooted in places that the outcast occupies.

Second, our analysis demonstrates the centrality of intra-city variations in segregation patterns in interrogating the mental maps of residents. The intra-city variation is trivially seen even with the range of values taken by Bangalore's 29 assembly constituencies (AC). The least segregated AC has a *D*-index of 0.164, and the most segregated one has a *D*-index of 0.709 — a range comparable to the 0.29–0.73 citywide dissimilarity range reported by Susewind across the 11 cities studied by him (Susewind, 2017, Table-1). Not surprisingly, our ward-level dissimilarity indices display an even greater range across 243 wards of Bengaluru — [0.07,0.7]. A similar order of magnitude range is found for diversity measures across both assembly constituencies and wards. This intra-city variation shows how the threefold ghetto-enclave-mixed taxonomy that Susewind (2017) quantified for entire towns can be found *within* large urban agglomerations such as Bengaluru.

Third, we introduce the demographic divergence of individual neighbourhoods (here, proxied by polling areas) to capture the segregation experienced by individual neighbourhoods. A quantitative measure of demographic divergence like the KL-divergence metric used here is crucial to understanding patterns of state discrimination or favoritism — one of the critical components of spatial marginalization, including ghettoization. The divergence of an individual neighbourhood (from either the ward or the assembly constituency) is politically salient and vital to any quantitative description of the relationship between segregation and ghettoization (for example, *cf.* Figure 3). All five components of ghettoization listed by Jaffrelot and Gayer (2012) intersect with neighbourhood-scale divergence. For example, the "estrangement" that a neighbourhood might experience is directly related to demographic divergence — a Muslim-dominated neighbourhood in a Muslim-dominated ward has a lower divergence and likely lower estrangement.

Fourth, we provided incontrovertible evidence for how Muslim segregation in Bengaluru is multi-scalar. Any adequate characterization of the extent of Muslim spatial marginalization must account for how segregation operates across different spatial scales. In our example, there is segregation between assembly constituencies in the city, wards in an assembly constituency, and polling areas within wards. The entropy class metrics we have used for diversity and segregation are perfectly additively decomposable across spatial scales and can shed light on the scale at which segregation is concentrated.

While our analysis has used Bengaluru as an illustrative example, the framework we have developed here can be used to study Muslim segregation anywhere in urban India. Our methodology relies only on publicly available electoral roll data and an open-source algorithm to classify names on these rolls. We used Bengaluru as a convenient illustration, as the local electoral authorities provide a geo-registered polling area map. However, the polling area map can be geo-registered for any town in India using data from the electoral rolls. It is also essential to recognize that a complete portrait of urban segregation in India must necessarily combine caste and religion information. The data and methods introduced here can easily be combined with the national census data now publicly available at the neighborhood scale (Census Enumeration Blocks) for urban India (Bharathi et al., 2021). Census enumeration blocks are smaller than polling areas (for example, Bengaluru contains approximately 17,000 enumeration blocks) and thus easily combined with religious demography derived from electoral rolls.

The data and methods introduced here also contribute to the emerging literature on segregation in the Global South (Garrido, 2021). Unlike racial segregation in North American cities that dominates the scholarly literature, urban centers in the Global South are not reducible to a simple inner-city versus suburb morphology. Segregation in the Global South is appropriately characterized as a "patchwork" of locally segregated neighborhoods (Garrido, 2019). The patchwork quilt is characterized by everyday economic interactions between the "ghetto" and the "enclave" and best describes Muslim segregation in large metropolitan centers of India (for the national capital, Delhi, see Jamil, 2017; Chakravorty and Sircar, 2021). This stands in direct contrast to the complete isolation that characterizes the classical ghetto.

The spatial marginalization of Muslims in contemporary India is best understood as a "boundary maintenance" (Barth, 1998) strategy adopted by the state as well as non-state actors. Our empirical framework for Muslim segregation allows for the large-scale characterization of how the maintenance of spatial boundaries is a constitutive feature of the subordination and othering of Muslims in contemporary India. Muslim segregation in urban India can take on *de jure* and *de facto* forms. Our multi-scalar measurement framework can help spatialize both state-led segregation efforts, such as the Disturbed Areas Act in Gujarat (Tejani, 2023), as well as segregation driven by the quotidian acts of non-state actors such as housing societies leading hate campaigns against Muslims (Ashraf, 2024).

The detailed multi-scalar segregation portrait we have developed here can seed future research in cognate areas of urbanization and planning. For example, the portrait of segregation can be used to answer questions about the political economy of public goods provisioning (Trounstine, 2018) or explore questions of environmental justice. The question of "who gets what and why" is often more usefully cast as "who gets what and where" (Lobao et al., 2007). The latter question is at the heart of the just cities transition in the Global South.

Notes

[1] For a comprehensive history spanning the pre-colonial, colonial, and the post-colonial, *cf.* Nair (2005).

[2] The D.J. Halli Ward of Bangalore has been visited by major and minor communal violence. Joshi (2020) documents nearly a dozen incidents in the ward in the last three decades that have shaped local spatial demography.

[3] The map for Bengaluru is available from https://kgis.ksrsac.in/pollinginfo/ (accessed, 10 March, 2023).

[4] With our normalized percentage measure, 0% implies that the PA demography is identical to ward demography, and 100% corresponds to maximum demographic divergence.

References

Allport GW, Clark K and Pettigrew T (1954) *The Nature of Prejudice*. New York: Addison-Wesley.

Ashraf, A (2024) Inside WhatsApp Groups of Housing Societies Where Muslim Hate is Real. *Decode,* 20 February 2024.

Barth, F. (1998). Ethnic groups and boundaries: The social organization of culture difference. Waveland Press.

Bhagat, R. B. (2018). Hindu—Muslim Riots in India: A Demographic Perspective. In Lives of Muslims in India (pp. 163-186). New Delhi: Routledge.

Bharathi N, Malghan D, Mishra S and Rahman A (2021) Fractal Urbanism: City Size and Residential Segregation in Urban India. *World Development* 141: 105397.

Bharathi N, Malghan D, Mishra S and Rahman A (2022) Residential segregation and public services in urban India. *Urban Studies* 59(14): 2912–2932.

Biswas, S (2022) Bangalore: How polarisation is dividing India's Silicon Valley. *BBC International, April 7 2022.*

Blom-Hansen, T. (2008). The political theology of violence in contemporary India. *South Asia Multidisciplinary Academic Journal*, (2).

Chakravorty S and Sircar N (2021) *Colossus: The Anatomy of Delhi*. New Delhi: Cambridge University Press.

Chatterjee, A (2017). *Margins of Citizenship: Muslim Experiences in Urban India*. New Yor: Routledge.

Chatterjee, A (2018). Muslims: An address in a ghetto. *Hindu Business Line*, January 11, 2018.

Chatterjee, I. (2009). Social conflict and the neoliberal city: a case of Hindu–Muslim violence in India. *Trans. Inst. Br. Geogr*. 143-160.

Chaturvedi R and Chaturvedi S (2023) It's All in the Name: A Character-Based Approach to Infer Religion. *Political Analysis* : 1–16.

Das, S. (2023). Democratic backsliding in the world's largest democracy. Available at SSRN 4512936.

Dhattiwala R (2019) *Keeping the Peace: Spatial Differences in Hindu-Muslim Violence in Gujarat in 2002*. New Delhi: Cambridge University Press.

Duncan O and Duncan B (1955) A Methodological Analysis of Segregation Indexes. *American Sociological Review* 20(2): 210–217.

Dupont V (2004) Socio-Spatial Differentiation and Residential Segregation in Delhi: A Question of Scale? *Geoforum* 35(2): 157–175.

ECI (2019) *Elections FAQ*. New Delhi: Election Commission of India. URL https://eci.gov.in/faqs/elections/.

ECI (2020) *Manual on Polling Stations, Second Edition*. New Delhi: Election Commission of India.

Galonnier J (2015) The enclave, the citadel and the ghetto: The threefold segregation of upper-class Muslims in India. *International Journal of Urban and Regional Research*, 39(1), 92-111.

Garrido, M. (2019). *The patchwork city: Class, space, and politics in Metro Manila*. Chicago: University of Chicago Press.

Garrido, M. (2021). Reconceptualizing segregation from the global south. *City & Community,* 20(1), 24-37.

Gayer L and Jaffrelot C (2012) Muslims of the Indian city. From centrality to marginality.

Gould P and White R (2012) *Mental Maps*. New York: Routledge.

Hochman, A., 2019. Racialization: A defense of the concept. *Ethnic and Racial Studies*, 42(8), pp.1245-1262.

Jaffrelot, C. (2007). The 2002 pogrom in Gujarat, in Hinnelss, J and King, R (Ed.) *Religion and violence in South Asia: Theory and practice*, Routledge.

Jaffrelot C and Gayer L (2012) *Muslims in Indian Cities: Trajectories of Marginalisation*. New York: Columbia University Press.

Jamil G (2017) *Accumulation by Segregation: Muslim Localities in Delhi*. New Delhi: Oxford University Press.

Joshi S.K. (2020) *D.J. Halli Riots Factfinding Report*. Bengaluru: Bangalore Factfinding Collective.

Kullback S and Leibler R (1951) On Information and Sufficiency. *The Annals of Mathematical Statistics* 22(1): 79–86.

Lobao LM, Hooks G and Tickamyer AR (2007) *The sociology of spatial inequality*. Buffalo, NY: Suny Press.

Massey D and Denton N (1988) The Dimensions of Residential Segregation. *Social forces* 67(2): 281–315.

Mishra Committee (2007) *Report of the National Commission for Religious and Linguistic Minorities*. New Delhi: Government of India.

Nair J (2005) *The Promise of the Metropolis: Bangalore's Twentieth Century*. New Delhi: Oxford University Press.

Nathan NL and Sands ML (2023) Context and Contact: Unifying the Study of Environmental Effects on Politics. *Annual Review of Political Science* 26.

Openshaw S (1984) Ecological fallacies and the analysis of areal census data. *Environment and planning A* 16(1): 17–31.

Rahman A and Kumar V (2023) 20 Years After Gujarat Riots, Muslim Survivors From Naroda Patiya Live Near A Mountain Of Trash. *Article14* 3(1): 1–9.

Sachar Committee (2006) *Report on the Social, Economic, and Educational Status of the Muslim Community of India*. New Delhi: Government of India.

Schumpeter JA (1954) *History of economic analysis*. New York: Psychology Press.

Shani O (2017) *How India became democratic: Citizenship and the making of the universal franchise*. Cambridge and New Delhi: Cambridge University Press.

Shannon CE (1948) A mathematical theory of communication. *The Bell System Technical Journal* 27(3): 379–423.

Sikka, S., 2022. Indian Islamophobia as Racism. *The Political Quarterly*, 93(3), pp.469-477.

Susewind R (2015) What's in a name? probabilistic inference of religious community from south asian names. *Field Methods* 27(4): 319–332.

Susewind R (2017) Muslims in Indian Cities: Degrees of Segregation and the Elusive Ghetto. *Environment and Planning A* 49(6): 1286–1307.

Tejani, S. (2023). Saffron geographies of exclusion: The Disturbed Areas Act of Gujarat. *Urban Studies*, 60(4), 597-619.

Thomas C (2012) Facing ghettoisation in 'riot-city'. old Ahmedabad and Juhapura between victimisation and self-help.

Trounstine J (2018) *Segregation by Design: Local Politics and Inequality in American cities*. Cambridge: Cambridge University Press.

Varshney, A., 2003. *Ethnic conflict and civic life: Hindus and Muslims in India*. Yale University Press.

Wacquant L (2008) *Urban Outcasts: A Comparative Sociology of Advanced Marginality*. New York: Polity.

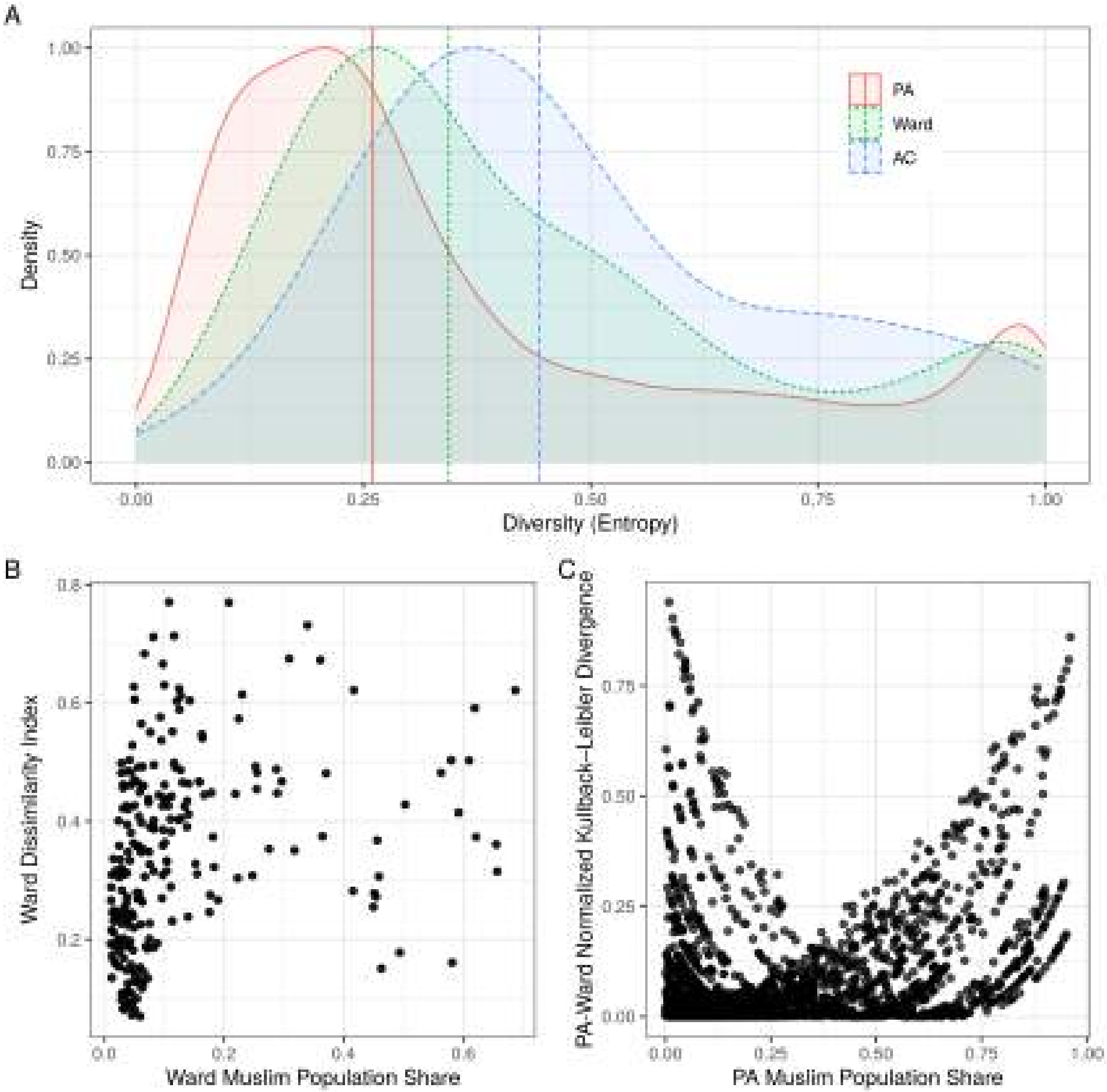

Figure 1: ***Muslim Segregation: Macro Portrait***. *Data from 7,526 PAs located in 243 wards and 29 assembly constituencies (AC). See main text for details.*

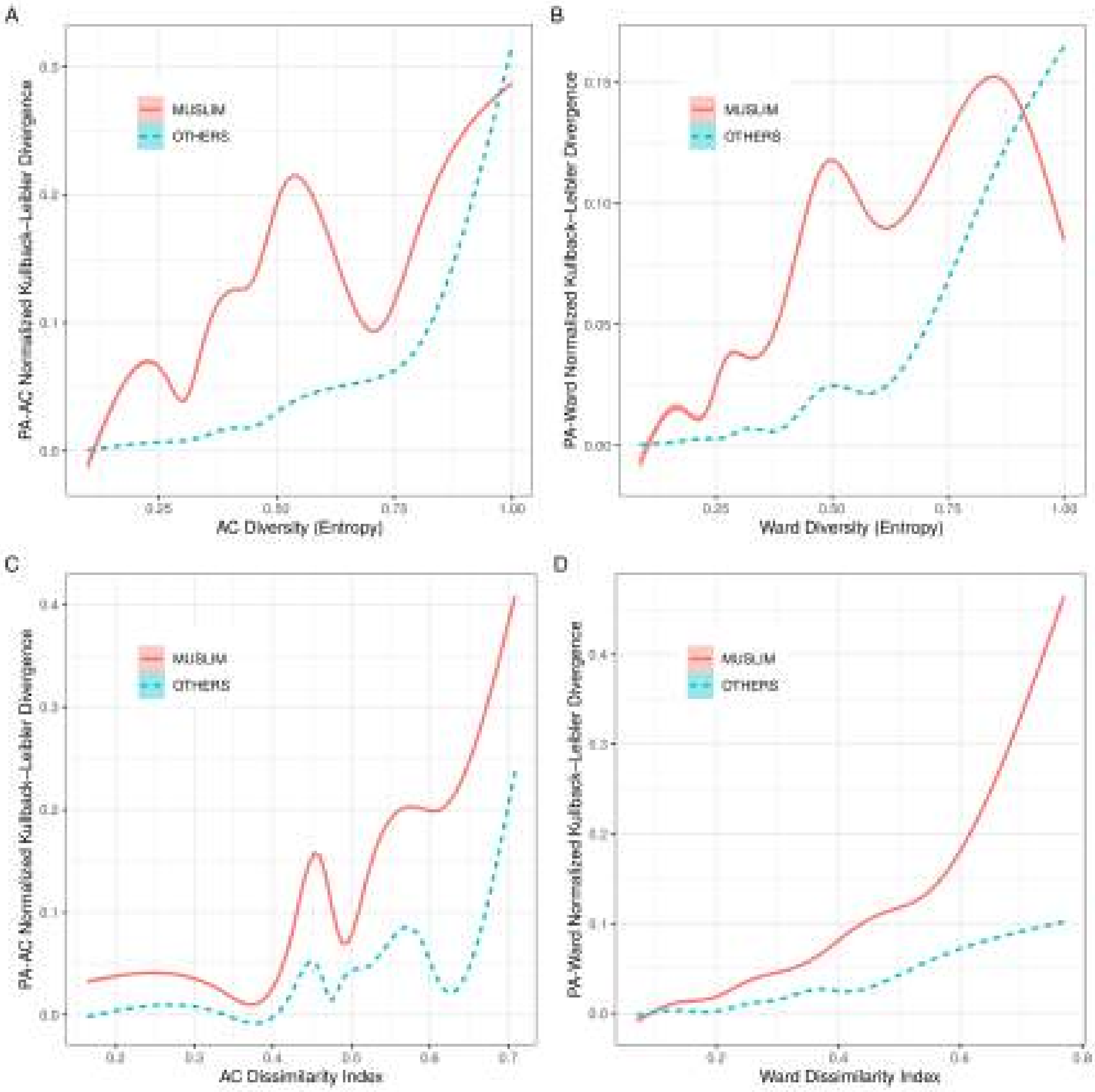

Figure 2: ***Muslim Segregation: Micro Portrait***. *Data from ≈ 9.5 million voters in Bengaluru. LOESS bands (95%) are shown in all panels. See main text for details.*

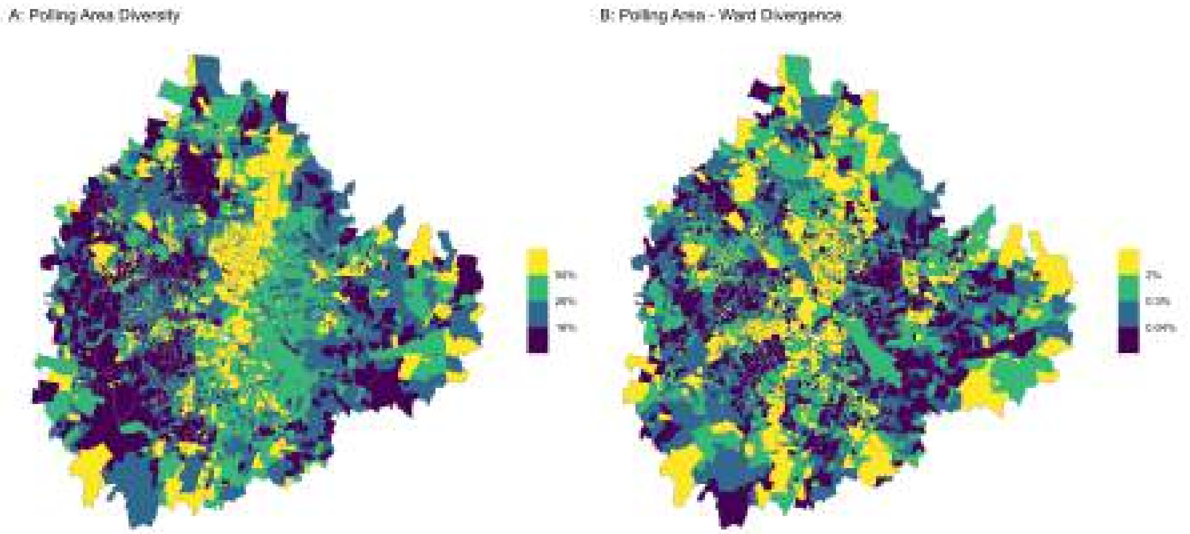

Figure 3: ***Diversity and Divergence Quartiles***. ***(A)*** *Polling Area (PA) Diversity Quartiles (Normalized Entropy, percent).* ***(B)*** *PA-Ward Normalized Kullback Liebler Divergence Quartiles (percent). Data from 7,526 polling areas (PAs). See main text for more details.*

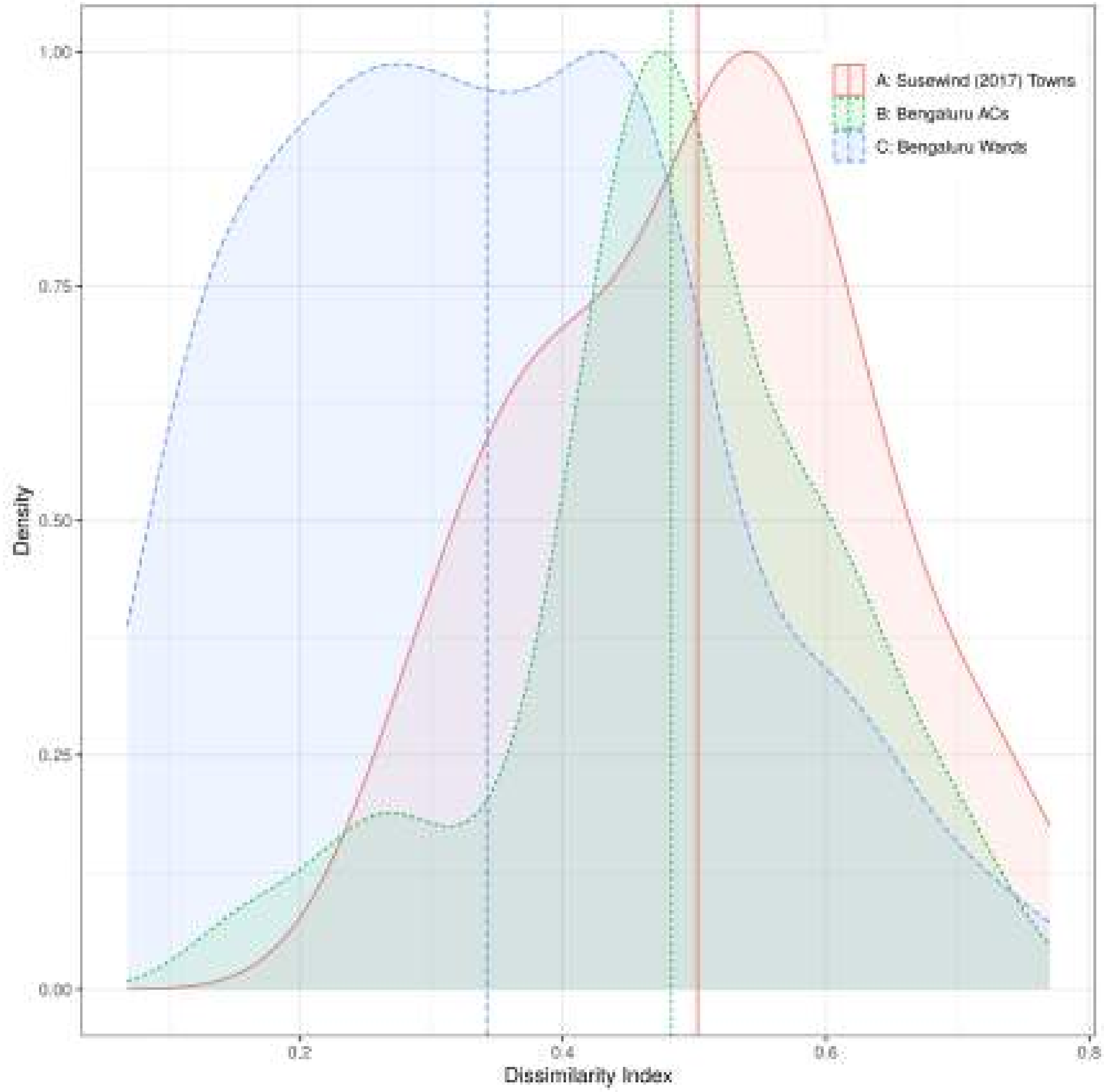

Figure 4: ***Dissimilarity Indices*** *(D). Data from 20 municipal towns and urban agglomerations in (**?**, Table-1); 29 Assembly Constituencies in Bengaluru (ACs), and 243 Bengaluru wards. Mean dissimilarity (D) is shown for each distribution. See main text for more details.*

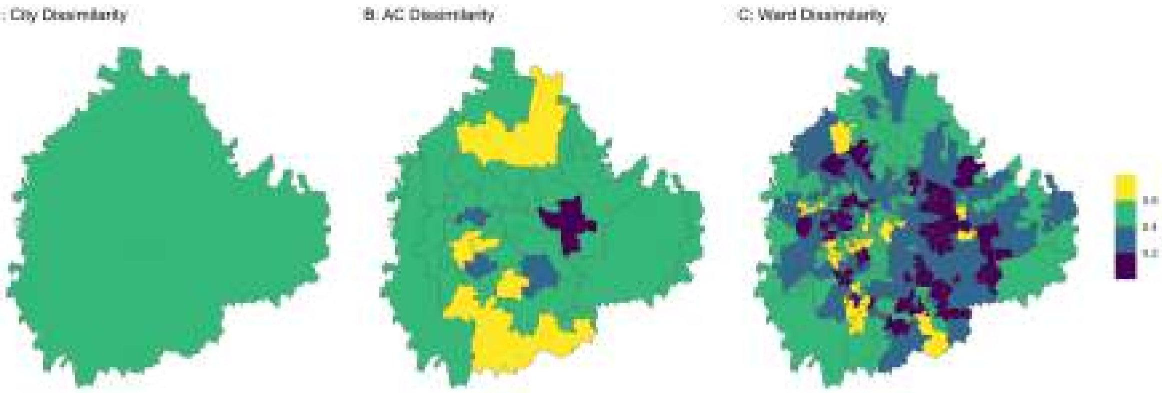

Figure 5: ***City, Ward and AC Dissimilarity***. ***(A)*** *Dissimilarity Index for the city of Bengaluru* ***(B)*** *Dissimilarity Index for 29 Assembly Constituencies* ***(C)*** *Dissimilarity Index for 243 wards. See main text for more details.*

On Measuring Muslim Segregation in Urban India

Online Appendix

Contents

List of Tables

List of Figures

Segregation Metrics

In this short appendix section, we describe the segregation metrics used to measure multi-scalar Muslim segregation in our paper. Besides providing details of how the segregation metrics were computed, we also offer an intuition for their interpretation in the context of Muslim segregation in urban India. For each metric, we also provide an extensive bibliography detailing how these metrics have been used in cognate fields.

Notation

Throughout this appendix, π_m and π_o represent demographic proportions of Muslims and non-Muslims (all others), respectively. Any segregation metric compares demographic distributions between two areal units — for example, between polling areas (neighborhoods) and the ward that contains them. We use $\vec{\pi}_P$ for the demographic distribution of the smaller areal unit, P, and $\vec{\pi}_Q$ to represent the distribution in the larger unit, Q that contains P. As we measure segregation using two groups (Muslims and non-Muslims), these distributions are simply vectors with two elements — $\vec{\pi}_P = (\pi_{Pm}, \pi_{Po})$, and $\vec{\pi}_Q = \big(\pi_{Qm}, \pi_{Qo}\big)$. m_i and o_i are, respectively, the absolute numbers of Muslim and non-Muslim residents in a smaller areal, i, at the level of P. Similarly, M and O are the total number of Muslim and non-Muslim residents at the level of Q. For example, m_i could be the total count of Muslim residents in a single polling area i, and M is the total count of Muslims across all polling areas of interest.

Dissimilarity Index

By a wide margin, the most commonly used metric for measuring residential segregation (including Muslim segregation in India) is the dissimilarity index, D (Duncan and Duncan, 1955; Taeuber and Taeuber, 1976). The D-index is a useful summary measure of segregation between two groups. It is easily interpreted as the proportion of the population-wide minority group that must be moved to eliminate all spatial clustering and achieve perfect "evenness" (Massey and Denton, 1988; Massey, White and Phua, 1996).

We used Eq. A1 to compute dissimilarity indices reported in the paper. In Eq. A1, m_i and o_i are, respectively, the number of Muslim and non-Muslim residents in a polling area i located in a ward, constituency, or city with a total population of M Muslims and O non-Muslims.

$$D = \frac{1}{2}\sum_i \left|\frac{m_i}{M} - \frac{o_i}{O}\right| \tag{A1}$$

Diversity

The extent of spatial unevenness captured by the dissimilarity index represents only one "dimension" of segregation (Massey and Denton, 1988). Residents in a neighborhood ultimately experience segregation in isolation. Among the four dimensions of the segregation taxonomy introduced by Massey and Denton (1988), "isolation" is the most phenomenologically relevant dimension. Among the many metrics available to measure isolation, diversity is the most suitable proxy for characterizing multi-scalar segregation (Bharathi et al., 2022). A more diverse neighborhood provides opportunities for greater outgroup contact.

Following contemporary segregation literature, we use the entropy metric rooted in information theory to measure diversity (Shannon, 1948; Theil, 1967; Theil and Finizza, 1971, 1992). Empirically, diversity measured using the information theory entropy metric is perfectly correlated with the ethno-linguistic fractionalization index, which is widely used in social sciences outside of segregation measurement. This empirical correlation is helpful in intuitively understanding what the information theory entropy index measures. We recollect that the fractionalization index (Hirschman, 1964) represents the probability that two randomly chosen individuals (from a given neighborhood) will belong to different sub-groups. For two groups (in our case, Muslims and non-Muslims), the fractionalization index takes on the maximum value of 0.5 when both groups have identical demographic shares. The entropy index measures the information content of a given demographic distribution. Consider three demographic distributions — ($\pi_m = 0, \pi_o = 1$), ($\pi_m = 0.5, \pi_o = 0.5$), ($\pi_m = 0.8, \pi_o = 0.2$). The first distribution, which is evidently the least diverse, has no information content, and the second distribution (with equal demographic proportions) has the greatest information content. This becomes clear in Eq. A2 that we use to compute entropic diversity.

$$H = -\left(\pi_m \log_e(\pi_m) + \pi_o \log_e(\pi_o)\right) \tag{A2}$$

In Eq. A2, diversity is computed as the sum of weighted inverse probabilities, with each term representing a subgroup's contribution to overall information content. Conversely, the individual inverse probabilities represent the corresponding subgroup's contribution to "uncertainty reduction." This direct correspondence between information content and uncertainty reduction makes the entropy metric an attractive measure of diversity to understand the isolation of segregated groups.

Using diversity as a measure of isolation stemming from segregation also sets the stage for measuring multi-scalar segregation. As discussed above, segregation measurements typically compare demographic proportions in individual neighborhoods with citywide demography. In contrast, the diversity measure uses demographic distributions from only the neighborhood. Thus, instead of a single summary measure of segregation for the entire city, we can show how each neighborhood experiences segregation. We build on this intuition (and the entropy metric) to present evenness measures at multiple scales.

Nieghborhood Divergence

To develop neighborhood-scale segregation measures, we ask the following question: how does the demographic distribution of the neighborhood diverge from the larger ward, electoral constituency, or the city containing it? This question is implicit even in the computation of the dissimilarity index. In Eq. A1, $\left|m_i/M - o_i/O\right|$ represents the contribution of neighborhood i to overall unevenness in the city. However, these individual contributions are not easily interpreted or compared across neighborhoods (Elbers, 2021). We use the information theory concept of "cross entropy" (Good, 1956, explained below) to characterize neighborhood-scale segregation.

Following the notation introduced above, consider a neighborhood P (in our case, an individual polling area) located in a larger areal unit (say, ward or electoral constituency), Q. The diversity of the neighborhood P ($H(P)$, measured using Eq. A2) is not independent of the diversity of the larger unit Q. For example, neighborhoods in a homogeneous larger unit are not expected to be diverse. Cross entropy represents the estimated diversity in Q given the demographic distribution in the larger unit P containing Q. Cross entropy shares a one-to-one correspondence with the more familiar likelihood function — maximizing log likelihood corresponds to minimizing cross entropy (so that $\vec{\pi}_Q$, the demographic distributoin in the larger unit, Q,is a good proxy-estimate of $\vec{\pi}_P$, the actual distribution in P).

If cross entropy is the estimate of neighborhood diversity based on the larger unit's diversity, the divergence between this estimate and the actual diversity is a meaningful and intuitive measure of segregation at the neighborhood scale. This difference between cross-entropy and entropy is the Kullback-Leibler (KL) divergence (relative entropy) that forms the basis of our neighborhood-scale segregation measurement (Kullback and Leibler, 1951; Kullback, 1959). Given this straightforward interpretation, KL Divergence has been adopted widely in the social sciences to measure group divergences in disparate fields (prominent examples include Walsh and O'Kelly, 1979; Mora and Ruiz-Castillo, 2003; Mori, Nishikimi and Smith, 2005; Bloome, 2014; Sasson, 2016). The KL Divergence has also increasingly found favor as the metric of choice among scholars of residential segregation (Ceccato and Karlström, 2000; Mora and Ruiz-Castillo, 2009, 2010; Bruun and Bearden, 2014; Roberto, 2016; Olteanu, Randon-Furling and Clark, 2019; Olteanu et al., 2020). We build on this robust tradition to measure the multi-scalar segregation of Muslims in urban India.

The KL divergence computation is described in Eq. A3. $\boxed{\mathbf{D}_{\mathrm{KL}}\left(\vec{\pi}_P \parallel \vec{\pi}_Q\right)}$ is the KL Divergence "from Q to P" (the divergence of the demographic distribution from the larger unit, Q to the smaller neighborhood unit, P). All other terms in Eq. A3 follow the notation we have defined above.

$$\mathbf{D}_{\mathrm{KL}}\left(\vec{\pi}_P \parallel \vec{\pi}_Q\right) = \pi_{Pm}\log_{\mathrm{e}}\left(\frac{\pi_{Pm}}{\pi_{Qm}}\right) + \pi_{Po}\log_{\mathrm{e}}\left(\frac{\pi_{Po}}{\pi_{Qo}}\right) \tag{A3}$$

To make KL Divergence values of the neighborhoods more intuitive (especially in comparisons across space and time), we normalize $\mathbf{D}_{\mathrm{KL}}\left(\vec{\pi}_P \parallel \vec{\pi}_Q\right)$ to the familiar $[0, 1]$ scale. We accomplish this normalization in Eq A4 by dividing KL-divergence by the log of the inverse demographic proportion of the smaller of the two groups.

$$\left.\begin{aligned} \mathbf{L}_{\mathrm{KL}}\left(\vec{\pi}_P \parallel \vec{\pi}_Q\right) &= \left(\frac{\mathbf{D}_{\mathrm{KL}}\left(\vec{\pi}_P \parallel \vec{\pi}_Q\right)}{-\log_{\mathrm{e}}\left(\pi_{Qg^*}\right)}\right); \\ \pi_{Qg^*} &= \min\left(\pi_{Qm}, \pi_{Qo}\right) \end{aligned}\right\} \tag{A4}$$

In Eq. A4, $\boxed{\mathbf{L}_{\mathrm{KL}}\left(\vec{\pi}_P \parallel \vec{\pi}_Q\right)}$ now represents the *normalized* KL Divergence from Q to P. The demographic proportion of the smallest group is used in this normalization because this is when cross-entropy is maximized. We have used $\mathbf{L}_{\mathrm{KL}}$ in our analysis of Muslim segregation in Bengaluru.

Election Geography in India

Our framework for measuring Muslim segregation in urban India is centered on the election geography of India. Given that there is considerable confusion in the literature on the various terms used to describe India's election geography, we present a glossary in Table A1. Figure A1 shows a typical polling station, polling station building, and the associated polling area (ECI, 2020).

Table A1: ***Indian Election Geography: A Glossary***

Terminology	Details
Electoral Rolls	The list of all registered voters, organized by assembly (state legislature) constituency.
Part	The elementary unit of electoral rolls that maps uniquely to a "polling area" and a "polling station."
Polling Area (PA)	Uniquely defined geographic boundary corresponding to a "polling station" and a single "part" of the electoral rolls. We use the PA as our primary areal unit of Muslim segregation measurement.
Polling Station	Polling station, in the current official usage, is the physical location with a one-to-one correspondence with a particular "part" of the electoral rolls. With some exceptions (where women only polling stations are setup), the polling stations also share a one-to-one correspondence with a "polling area."
Polling Station Building	The building where the actual polling happens (often, a public school). The single polling station building can contain up to four "polling stations" (this number is often exceeded in dense urban areas). Susewind (2017) uses interpolated polygons at this level in his measurement of Muslim segregation.
Precinct	Officially accepted synonym for "polling area."
Voter List	The colloquial term for "electoral rolls," and widely used in media reporting of Indian elections.
Polling Booth (Historical)	Historical (1951–80) term for the current "polling station." This term continues to be used colloquially as a synonym for "polling station." Also referred to as "electoral booth" in media reporting as well as scholarly work.
Polling Station (Historical)	Before the current terminology came into effect in 1980, the "polling station" was used to refer to the building where the voting took place. A collection of historical "polling booths."

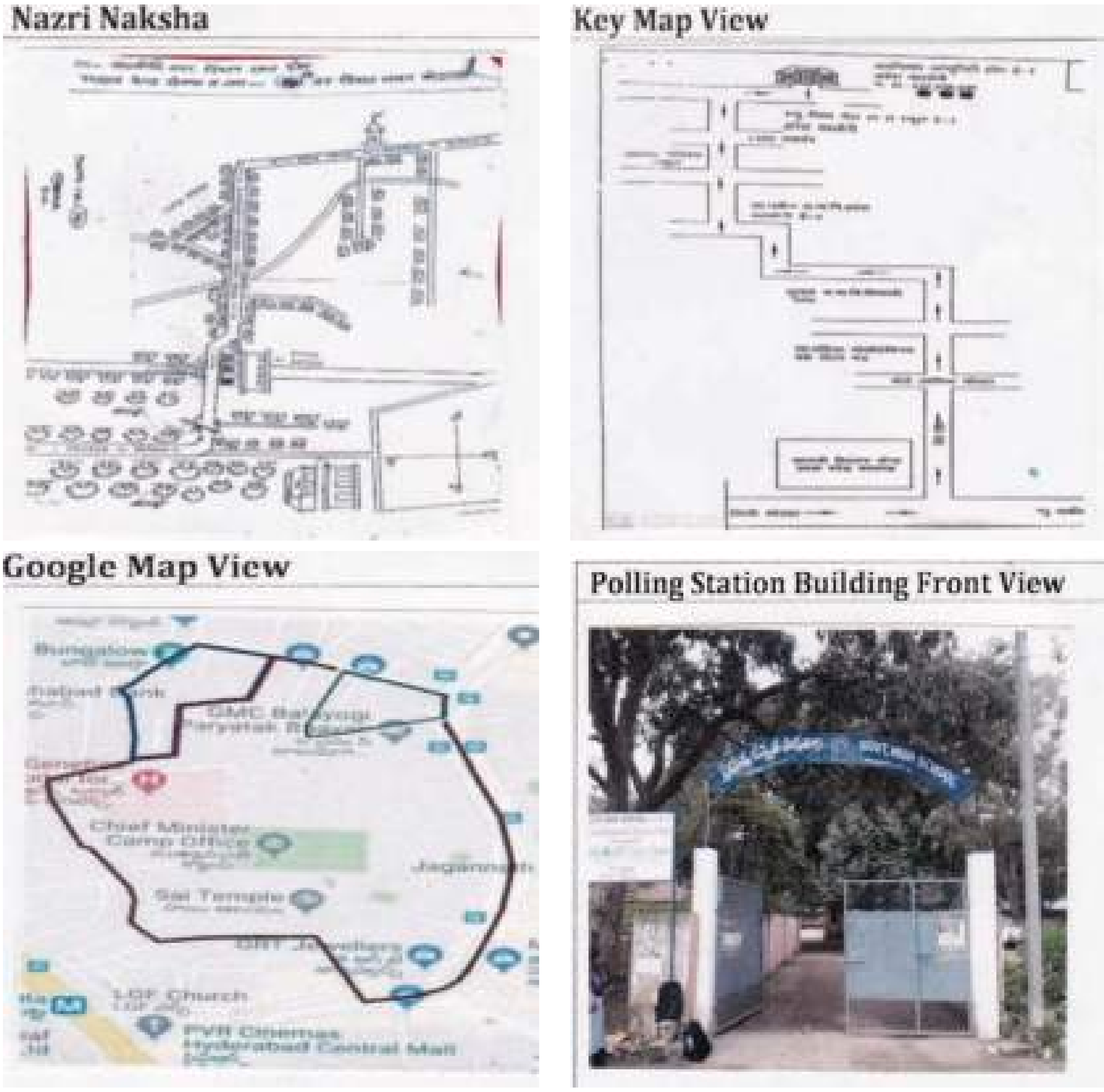

Figure A1: ***Polling Station, Polling Area, and Polling Station Building***. *Reproduced from (ECI, 2020, pp. 18–20)*

Voronoi Interpolation Bias

In building our framework for Muslim segregation measurement, we have used actual polling area polygons rather than pseudo-polygons generated by performing a Voronoi interpolation over polling station point coordinates. Figures A2 – A3 document how our method corrects the Voronoi interpolation bias.

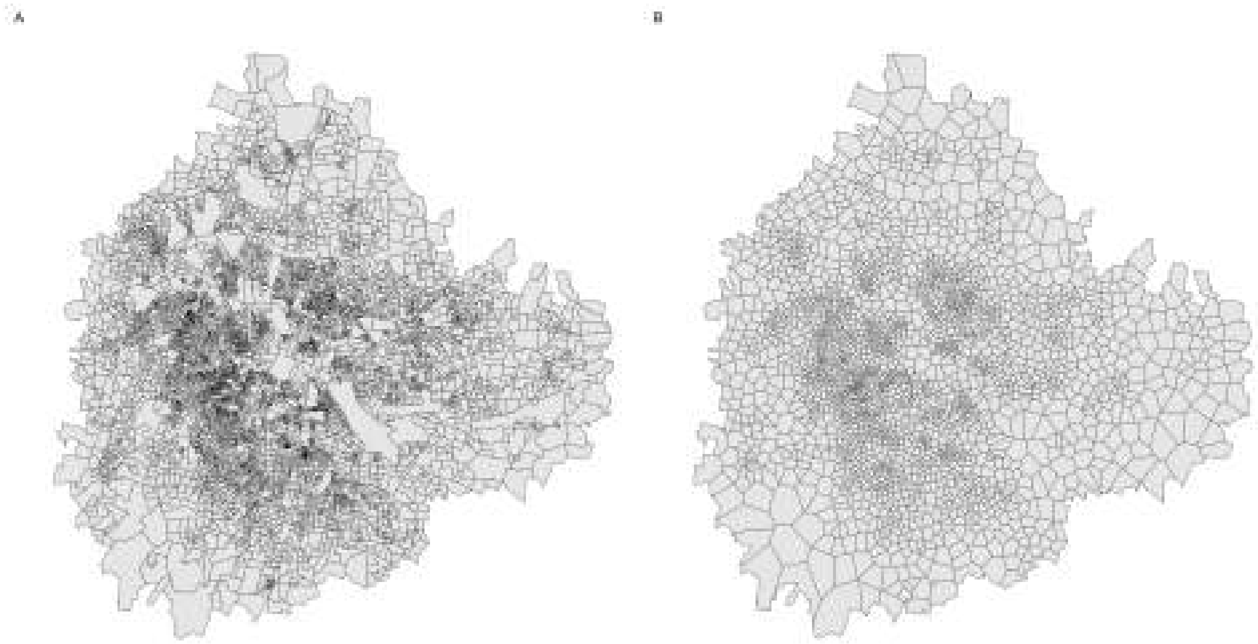

Figure A2: ***Loss in Spatial Resolution***. *Panel-A shows actual extent of 7,526 polling areas in Bengaluru. Panel-B shows 2,935 meta-aggregates obtained using Voronoi interpolation. These meta-aggregates correspond to the number of polling station buildings.*

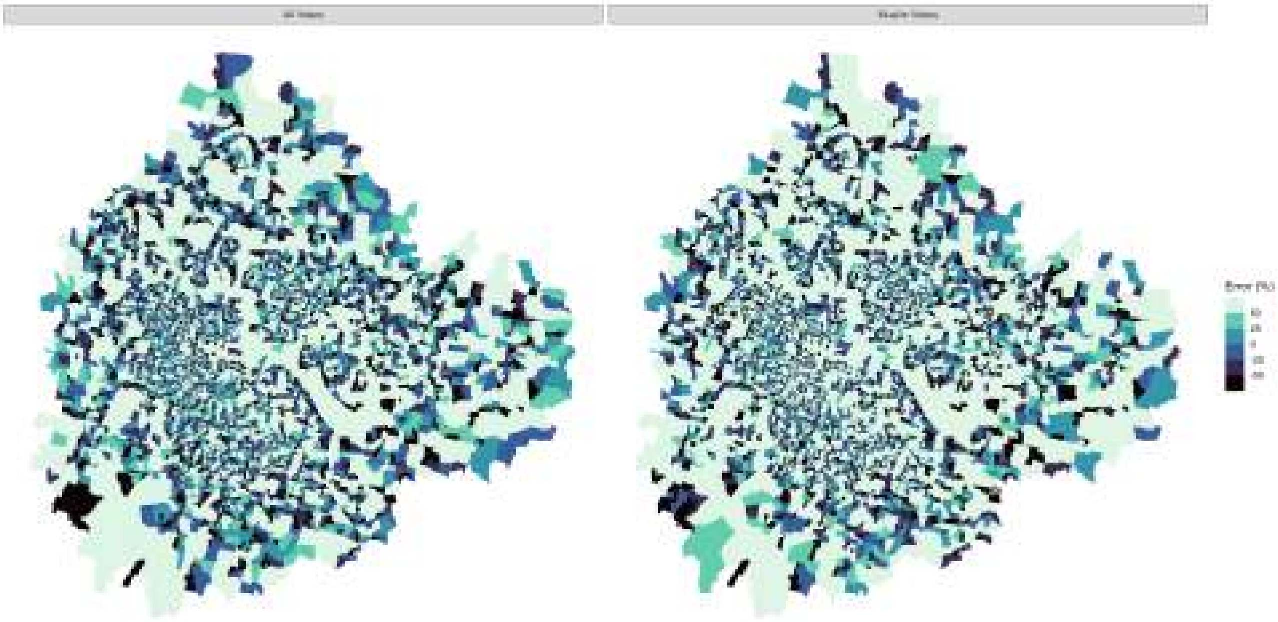

Figure A3: ***Spatial Distribution of Voronoi Interpolation Errors***. *Errors from Voronoi interpolation (shown as a percentage error relative to actual electoral polling area demography) in overall population estimates and Muslim population estimates are depicted for each of the 7,526 polling areas in Bengaluru. See main text for details.*

Online Appendix References

Bharathi, Naveen, Deepak Malghan, Sumit Mishra, and Andaleeb Rahman. 2022. "Residential segregation and public services in urban India." *Urban Studies*, 59(14): 2912–2932.

Bloome, Deirdre. 2014. "Racial Inequality Trends and the Intergenerational Persistence of Income and Family Structure." *American Sociological Review*, 79(6): 1196–1225.

Bruun, Jesper, and Ian G Bearden. 2014. "Time Development in the Early History of Social Networks: Link Stabilization, Group Dynamics, and Segregation." *PLoS One*, 9(11): e112775.

Ceccato, Vania, and Anders Karlström. 2000. "A New Information Theoretical Measure of Global and Local Spatial Association." *European Regional Science Association*, 29.

Duncan, Otis, and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indexes." *American Sociological Review*, 20(2): 210–217.

ECI. 2020. *Manual on Polling Stations, Second Edition.* Election Commission of India.

Elbers, Benjamin. 2021. "A Method for Studying Differences in Segregation Across Time and Space." *Sociological Methods & Research*, 0049124121986204.

Good, IJ. 1956. "Some terminology and notation in information theory." *Proceedings of the IEE-Part C: Monographs*, 103(3): 200–204.

Hirschman, Albert O. 1964. "The Paternity of an Index." *The American Economic Review*, 54(5): 761–762.

Kullback, Solomon. 1959. *Information Theory and Statistics.* John Wiley.

Kullback, Solomon, and Richard Leibler. 1951. "On Information and Sufficiency." *The Annals of Mathematical Statistics*, 22(1): 79–86.

Massey, Douglas, and Nancy Denton. 1988. "The Dimensions of Residential Segregation." *Social forces*, 67(2): 281–315.

Massey, Douglas, Michael White, and Voon-Chin Phua. 1996. "The Dimensions of Segregation Revisited." *Sociological Methods & research*, 25(2): 172–206.

Mora, Ricardo, and Javier Ruiz-Castillo. 2003. "Additively Decomposable Segregation Indexes. The case of Gender Segregation by Occupations and Human Capital levels in Spain." *The Journal of Economic Inequality*, 1(2): 147–179.

Mora, Ricardo, and Javier Ruiz-Castillo. 2009. "The Statistical Properties of the Mutual Information Index of Multigroup Segregation." *Documentos de trabajo. Economic series (Universidad Carlos III., Department of Economics)*, 1.

Mora, Ricardo, and Javier Ruiz-Castillo. 2010. "A Kullback-Leibler Measure of Conditional Segregation." *Documentos de trabajo. Economic series (Universidad Carlos III., Department of Economics)*, 2.

Mori, Tomoya, Koji Nishikimi, and Tony E Smith. 2005. "A Divergence Statistic for Industrial Localization." *Review of Economics and Statistics*, 87(4): 635–651.

Olteanu, Madalina, Julien Randon-Furling, and William Clark. 2019. "Spatial Analysis in High Resolution Geo-Data." *ESANN Proceedings*, 2019.

Olteanu, M, C de Bezenac, W Clark, and J Randon-Furling. 2020. "Revealing Multiscale Segregation Effects from Fine-Scale Data: A Case Study of Two Communities in Paris." *Spatial Demography*, In press.

Roberto, Elizabeth. 2016. "The Divergence Index: A Decomposable Measure of Segregation and Inequality." *arXiv Preprint*, arXiv:1508.01167v2: 1–34.

Sasson, Isaac. 2016. "Trends in Life Expectancy and Lifespan Variation by Educational Attainment: United States, 1990–2010." *Demography*, 53(2): 269–293.

Shannon, Claude Elwood. 1948. "A mathematical theory of communication." *The Bell system technical journal*, 27(3): 379–423.

Taeuber, Karl, and Alma Taeuber. 1976. "A Practitioner's Perspective on the Index of Dissimilarity." *American Sociological Review*, 41(5): 884–889.

Theil, Henri. 1967. *Economics and Information Theory.* North Holland.

Theil, Henri, and Anthony J Finizza. 1971. "A Note on The Measurement of Racial Integration of Schools by Means of Informational Concepts." *The Journal of Mahematical Sociology*, , (1): 187 – 193.

Theil, Henri, and Anthony J Finizza. 1992. "A Note on The Measurement of Racial Integration of Schools by Means of Informational Concepts." In *Henri Theil's Contributions to Economics and Econometrics.* 1033–1040. Springer.

Walsh, Jim, and Morton Edward O'Kelly. 1979. "An Information Theoretic Approach to Measurement of Spatial Inequality." *Economic and Social Review*, 10(4): 267–286.