The Causes and Consequences of Religious Segregation: Evidence from India *

Abhishek Arora † Aditya Chhabra ‡

Abstract

We devise a deep learning method to digitize cadastral maps from 20,000 villages in Karnataka, India to obtain coordinates of over 11 million land parcels and use the data to measure rural segregation at an unprecedented resolution. We provide rich descriptions of the levels of religious segregation and become the first to document it at the village level. Then, we study how the religious composition of neighborhoods affects public good provision. To tackle the endogenous selection of minority households into under-serviced neighborhoods and other unobservables, we use a novel instrumental variable, the presence of a *Dargah* - a *Sufi* shrine - to instrument for Muslim share within a neighborhood. For the first time, we provide causal evidence of the effect of neighborhood concentration of Muslim inhabitants on public good provision, specifically, that schools and health centers are more likely to be located in neighborhoods with a larger share of Muslim households, especially in those villages where the overall share of Muslim inhabitants is lower. However, the scale of concentration matters - neighborhoods in villages under the same local government with a larger share of Muslim inhabitants are less likely to contain public facilities. Finally, we isolate historical path dependence as one of the possible causes of observed differences in regional spatial concentration by utilizing a spatial Regression Discontinuity Design to show that the region that was ruled by the Nizam of Hyderabad more than 75 years ago is more segregated than the rest of the neighboring regions.

*Preliminary Draft - Please do not circulate

†Harvard University; Email: abhishekarora@fas.harvard.edu
‡World Bank; Email: achhabra@worldbank.org

1 Introduction

Development may not be inclusive and marginalized groups can be left behind even in the presence of rapid improvements in the overall standard of living. Spatial divides can lead to developmental divides emanating from potentially negative consequences of segregation like poorer access to public goods and employment networks for the marginalized group. (Cutler & Glaeser, 1997; Massey & Denton, 1993; Cutler *et al.*, 2008; Alesina & Zhuravskaya, 2011; Ananat & Washington, 2009; Boustan, 2013). We study religious segregation in rural India at an unprecedented scale. We zoom into the state of Karnataka to study how segregation affects public good provision both within and across villages while also isolating a historical channel that continues to drive present-day segregation of Muslim households.

Studying segregation in developing countries is difficult due to a lack of data at the neighborhood or individual level. Recent work by Adukia *et al.* (2023) documents segregation for most of India using Census data from 2012-13. They use neighborhoods defined at the level of the "enumeration block" in the census which generally contains 100-125 households (around 700 inhabitants). While these are potent in documenting segregation in Urban areas at the Town level, about 50% of the villages are contained within a single enumeration block. Without the coordinates of each inhabitant in the village, it then becomes impossible to compute standard measures of segregation at the village level. Thus, they end up measuring rural segregation at the level of the subdistrict which contains around 100 villages on average. Studying segregation at the level of the village - and within a village, would therefore require a much finer resolution of data than what is available from the Census.[1] Due to this inherent difficulty, most work has exclusively focused on Urban segregation and has generally used much larger neighborhoods - wards containing more than 20,000-30,000 inhabitants.

[1] Naveen Bharathi & Rahman (2021) do look at "micro-segregation" at the village level in Karnataka, India that is estimated by comparing consecutive "runs" of households to a hypothetical random distribution. It relies on the fact that nearby household IDs in the dataset represented spatial proximity allowing for the measurement of local concentration. However, without the locations of the households, their membership in appropriately defined and small-enough neighborhoods, or at least their proximity to key public facilities it is impossible to look at the effects of this concentration at the local level with their data.

(Susewind, 2017; Bharathi *et al.*, 2022; Naveen Bharathi & Rahman, 2021; Vithayathil & Singh, 2012; Singh *et al.*, 2019). 64% of the population of the country still lives in rural areas making it imperative to fully understand how residential patterns can affect access. We introduce a deep learning-based method to digitize cadastral maps from 20,000 villages in the state of Karnataka, India, and link it with administrative data to obtain geographic coordinates of over 16 million land owners - the finest georeferenced dataset available in a developing country context. This allows us to define arbitrarily large "neighborhoods" by tessellating a village into a grid and using them to measure segregation at the village level for the very first time. We then introduce an effective large-language model-based approach to train a binary classifier to classify property owner names into Muslim/non-Muslim and apply it to our dataset. Building on Adukia *et al.* (2023), we explore how the religious composition of a neighborhood affects minority access to public facilities who found descriptive evidence of public facilities being located away from neighborhoods (much larger than ours) and villages containing a larger share of Muslims. First, our data allows us to look at the effects on much smaller, within-village neighborhoods. Second, our data is from more than a decade after the last Census, and the enumeration of the next Census which was due in 2021 has been indefinitely postponed. Third, we use an instrumental variable to isolate the effects of segregation to deal with the possibility that minority households select into under-provisioned neighborhoods or if these neighborhoods are worse or better off for any reasons other than the religious composition. More specifically, we use the "presence of a *Dargah* - a *Sufi* shrine in the neighborhood to instrument for its Muslim share. Locations of these shrines, unlike those of mosques, are historically predetermined, and Muslim households tend to cluster around them. With appropriate controls, the IV can satisfy both relevance and (conditional) exogeneity conditions. Our data on public facilities comes from an equally rich source - GEOSADAK which is an open dataset that records the universe of public facilities in each village, prepared and maintained by the Ministry of Rural Development [2]. Knowing the

[2] The data can be found at https://geosadak-pmgsy.nic.in/

coordinates of each landowner and the neighborhoods they fall into allows us to spatially join other geospatial datasets. Specifically, we can ascertain whether a neighborhood contains one of four types of facilities in the dataset - Educational (Schools/Universities), Medical (Health centres, Hospitals etc.), Agricultural (Markets, Grain storage, processing etc.) and Administrative and Transportation facilities (Government offices, Bus Stands etc.). We find that neighborhoods with a larger share of Muslim households are more likely to contain educational and medical facilities. However, the benefit of this local spatial concentration is more prominent in villages where the Muslim share is lower overall. Moreover, the scale of concentration matters - neighborhoods in villages under the same local government that have a larger share of Muslim inhabitants are less likely to contain public goods. Larger neighborhoods in previous work thus hide the potential benefits of belonging to a local enclave, especially in those villages where the overall share of the minority is low. There is qualitative evidence on how Muslims report difficulty in getting public facilities from their representatives (Jaffrelot & Gayer, 2012), and perhaps in those villages where the overall share of Muslims is lower, locating around other households from the same community is beneficial. We aren't the first to find positive effects of segregation in the Indian context. Jaffrelot & Gayer (2012)also document qualitative evidence that Muslim inhabitants of neighborhoods segregated by communal violence in the state of Gujarat did better than their peers from less segregated neighborhoods despite having worse access to public services. Geruso & Spears (2018) find positive spillovers for households having more Muslim neighbors driven by lower open-defecation rates in the community. More recently, Kalra (2021) finds that communal violence is associated with an improvement in early education outcomes of children who began their schooling after violent events and provides descriptive evidence of that effect being linked to higher segregation levels in communities that faced violence. That there are possible benefits of spatial concentration is not limited to this context. Cutler *et al.* (2008) for instance find that in the U.S., residential segregation is beneficial for more educated groups. Besides noting the effects of segregation on access, we observe considerable regional heterogeneity in the segregation and overall Muslim shares at the village level and it is important to understand its potential causes while designing policies to deal with its consequences. We note that villages in the northeastern part of the state are more segregated. This area largely corresponds to the territory once ruled by the Nizam of Hyderabad, a Muslim ruler, over 75 years ago and is now a part of the "Kalyana-Karnataka" region [3]. The surrounding regions were ruled directly or indirectly by the British. We implement a spatial regression discontinuity design on the lines of Dell (2010) on the border separating the regions once ruled by the Nizam from the British presidencies of Bombay and Madras. We find that among villages around the Nizam border, those on the side of Nizam have a 30% higher segregation and Muslim share. That a boundary within the same state from 75% years ago affects present-day differences in religious segregation across villages, points towards the important role historical differences continue to play in the present day. These differences need to be taken into account when planning policies to reduce inter-community disparities. We plan to uncover the underlying mechanism driving this result in the future.

In Section 2, we give the present-day and historical context. Section 3 and Section 4 introduce the data and the digitization effort respectively. Section 5 documents village-level segregation measured with our data. Section 6 explores the effects of segregation on public good access, Section 7 explores a potential historical reason behind present-day segregation, and we conclude in Section 8.

2 Background

The state of Karnataka The state is divided into 31 districts. Each of them is subdivided into talukas (known as blocks in other Indian states) which are further divided into groups of villages called Gram Panchayats (GPs), with each GP typically consisting of 4-5 villages. Each GP is governed by a council consisting of elected representatives who are

[3] The Bellary district, also a part of the Kalyana Karnataka region wasn't under Nizam rule but under direct British rule.

elected for a term of 5 years. While it contains many urban centers like Bangalore, the state is predominantly rural (61%) and its villages are home to 37.5 million people (Census 2011). Muslims form approximately 13% of the population of the state, but the share of Muslims is higher in Urban areas (21%) relative to rural areas (7.5%).

The region has a rich history. It was only fully organized as a state within its present-day boundaries in 1956 but many empires and dynasties have ruled over the region in the past. Notably, it has seen alternating periods of Hindu and Muslim rule. More recently, after the defeat of Tipu Sultan in the Siege of Seringapatam (1799), the region was mainly divided between the Bombay and Madras presidency (direct British rule), the Kingdom of Mysore (as a British protectorate) and the Nizam of Hyderabad (also under British protection). The Nizams ruled over the present-day districts of the northeastern part of Karnataka - Bidar, Gulbarga, Yadgir, Kopal, and Raichur, now a part of the Kalyana-Karnataka region. The state's historical partitions are depicted in Figure A1. Later, the States Reorganisation Act of 1956 organized the regions formerly under the Hyderabad, Bombay, Madras, Coorg, and Mysore states into the state of Mysore - which was later renamed Karnataka.

The Hyderabad State The territory under the Hyderabad state constituted the present-day state of Telangana, parts of modern-day Karnataka, and parts of Maharashtra, India. The Nizams ruled directly from 1724 to 1857 and subsequently came under British Suzerainty as a Princely State but largely retained internal autonomy. After the end of British rule in India in 1947, the incumbent Nizam announced that the state did not wish to join India or Pakistan. During this period until the Indian government's Operation Polo in 1948, which led to the accession of the state to India, the region was rife with communal violence. Based on the recently declassified report[4] by the Pandit Sunderlal committee - a government-instituted panel, 27,000 - 40,000 Muslims died during and after the operation that led to the integration of the territory under the Nizam into newly independent India.

[4] Available on internet archive : https://archive.org/details/pandit-sundarlal-committee-report-on-the-massacres-in-hyderabad-1948/page/n13/mode/2up

3 Data

3.1 Property Records in Karnataka

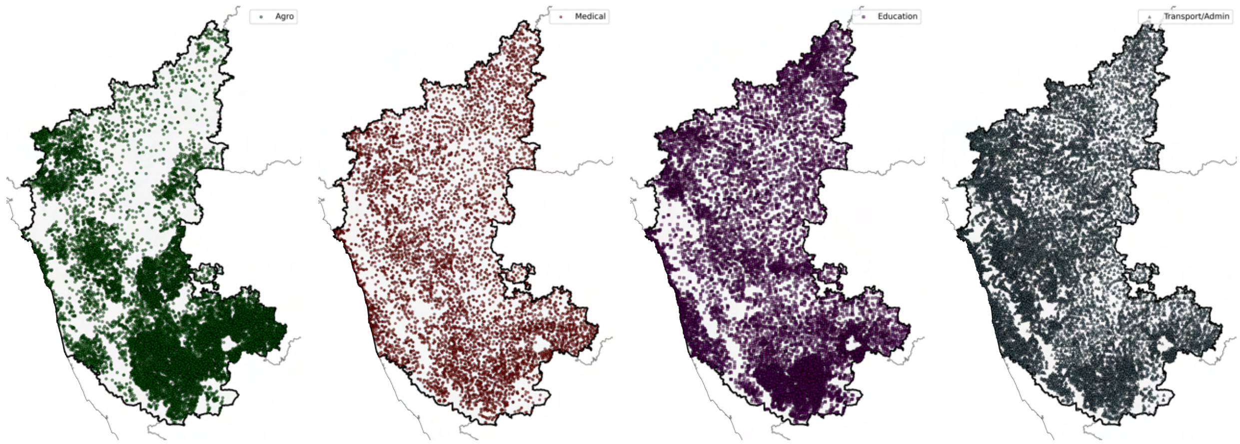
Property registration and the enforcement of property rights for the most part lie under the purview of the respective state government. Proper maintenance of the records of property rights can Over several decades now, the Central government has pushed for the digitization of land records by funding states to carry out a comprehensive digitization process under what is now known as the Digital India Land Record Modernization Programme (DILRMP). In the year 200, Karnataka's land record digitization effort was organized under the project "BHOOMI" aimed at digitizing paper records of property which in turn can lead to better protection of property rights, tackling bureaucratic corruption, improving access to credit, and assist property buyers in their due diligence. The project has since then won both national and international acclaim for making Karnataka one of the first Indian states to digitize properties at a massive scale. We use two datasets from Bhoomi - Property IDs and Owner names for over 18 million property owners and property-level village maps for 24,000 villages in the datasets. Both datasets are in the public domain by design. While a lot more detail is available for each property like ownership history, area, legal encumbrances, cultivation details, and soil type, we only limit our data collection effort to retrieving names of the property owners from the portal. Each property has 3 components that define a unique property in our dataset - "Survey Number", "Surnoc" and "Hissa Number". Each land parcel corresponding to a survey number can be subdivided into "Hissas" and each "Hissa" can have multiple owners. We observe all owners of a Hissa. We retrieve plot IDs and names of 18 million property owners, owning 11 million unique properties. Note that the administrative data gives us an identifier of the "main owner" of a plot, allowing us to keep only unique land parcels. We then use a deep learning-based approach to classify names of property owners into "Muslim" or "Non-muslim" by training a multilingual BERT model (Devlin *et al.*, 2018). We outline the process in Section 4.

As for the maps, the Bhoomi portal provides PDF files of village maps on their website at the Survey Number level - each land parcel has the survey number neatly annotated within it but the images are not georeferenced - they are just digital drawings of the parcels within the village. While the Central and state government has invested heavily in the digitization of these land maps (and going to a finer scale of the hissa), progress has been slow and expensive. Only recently in June 2023, the government launched the largest drone-based land survey to map land parcels in 10 of the 31 districts of Karnataka [5] while the plan has been in the works for several years now. Hence, the only available data source that can potentially provide the locations of each of the plots in our dataset is the collection of un-gereoferenced images available on the Bhoomi portal. The portal contains 24,000 villages and georeferencing them manually using standard tools would be a monumental task which would not be feasible on a research budget. We develop a novel deep-learning pipeline to georefernce these maps as outlined in Section 4.

3.2 Public Facilities

Our data is very granular - we know the coordinates of each plot (at least the level of survey number). To fully be able to use the richness of the information coming from property ownership data while studying household access to public goods, we use rich data from the Ministry of Rural Development's GEOSADAK open dataset containing coordinates of 61,042 state-provided facilities across the state, neatly categorized into 25,100 "Educational" (Schools/Colleges), 16049 "Market/Agro" (Markets/Agricultural Facilities), 12290 "Transport/Admin" (Government offices, Bus stands, Railway Stations) and 7603 "Healthcare" (Health care centers). We show their distribution in the state in Figure 1.

[5]Largest Drone-Based Land Parcel Mapping Contract in Karnataka Given to 2 Organisations." The Hindu, Bengaluru, 28 June 2023, 08:12 pm, Updated 29 June 2023, 09:45 am IST



(a) Agricultural/Markets (b) Medical (c) Educational (d) Administrative/Transport

Figure 1: The spread of the four types of facilities in the state of Karnataka, India

3.3 Locations of *Dargahs*

A *Dargah* is a shrine built over the grave of a revered religious figure, often a *Sufi* saint or *Dervish* who typically lived several centuries ago. They often include a mosque, meeting rooms, Islamic religious schools, and community buildings, but they can also be limited to a very small structure around the tomb. Historically, these shrines have been a common place of worship for Muslims - but also people of other religions. While the majority of visitors to a *Dargah* may be Muslim, they have been a symbol of co-existence in South Asia for centuries. *Dargahs* unlike other mosques would therefore have their locations predetermined by history - only the exteriors can be renovated or rebuilt. We will later show how this property of *Dargahs* can assist us in identification.

We used the Google Places API to get the locations of *Dargahs* in the state. We make the query using the keyword "dargah" in a radius of 10km around all of the state's villages' centroids. Note that while we specify a radius, the Google API may not limit itself to staying within it - it just uses the radius to rank top results. We pick the top 20 results (because we are interested in the nearest *Dargahs* to a village) and de-de-duplicate the results of the queries to get coordinates of 2525 *Dargahs* in and near the state's boundaries. While it may not contain the entirety of *Dargahs*, it has very good coverage. Figure 2 shows the spread across the state. 736 villages have a dargah within them - a village can have multiple

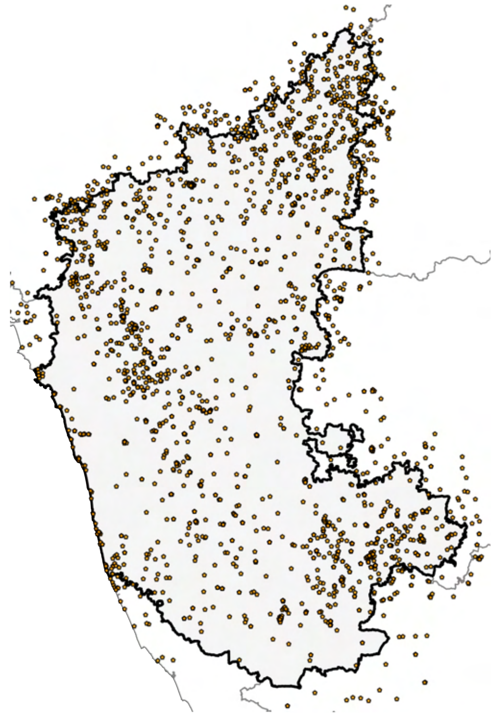


Figure 2: Locations of *Dargahs* (Sufi Shrines) in the state of Karnataka obtained using Google Places API

Dargahs. 1335 villages are GPs that have at least one *Dargah*.

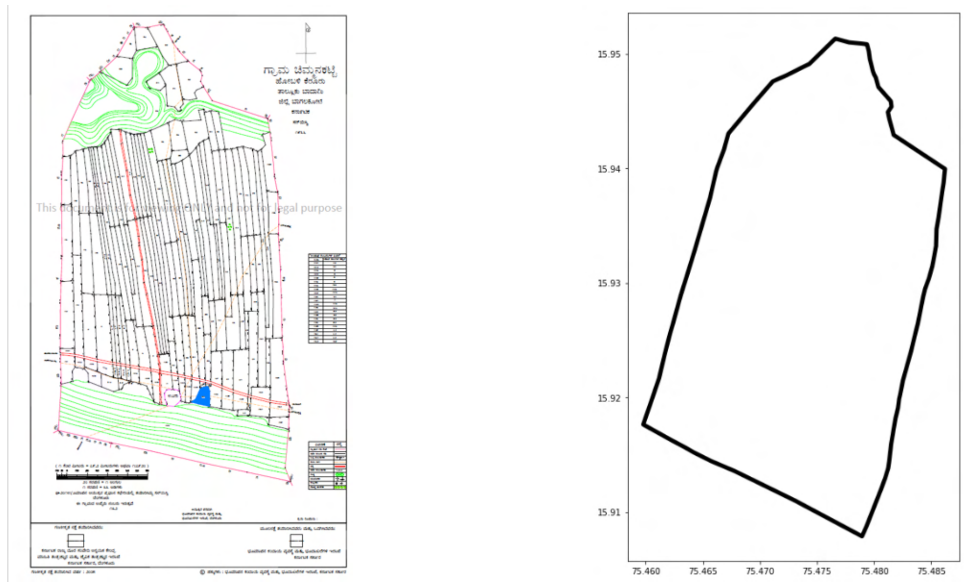
We may be under-counting the *Dargahs* because Google's coverage over some locations might be thin but, to the best of our knowledge, there is no other source that can provide the coordinates of the universe of *Dargahs* in the state (or in any other part of the country).

3.4 Historic Maps

We digitize the historic map of Karnataka's regions to obtain the boundaries of interest. We provide the source map and the georeferenced map in the Appendix.

4 Digitisation

Georeferencing Maps While there are massive map collections that are yet to be digitized, there are no tools available that can do it at scale accurately. Digitizing maps can be a laborious and expensive task.



(a) Raw map from the Bhoomi portal **(b)** Village polygon from a .shp file

Figure 3: An example village - its plot level map and digital polygon

We develop a novel pipeline to digitize property-level village maps at scale. Specifically, we break this problem down into georeferencing the map, followed by processing its contents. While the polygon of each plot would be important for getting the area of the plot, we do not tackle that problem - we only attempt to get the coordinates of each plot (note that the lowest unit on the map is a survey number, not the smallest unit of land in our ownership data - which is the hissa however, we will use the term plot for these for simplicity). The approach however can be extended to include plot level polygons as well using a semantic segmentation model - we do not attempt that because coordinates are sufficient to measure segregation and public good access - our main variables of interest. To put our approach succinctly, georeferencing a map in our case can be reduced to finding a transformation that can align a map to its corresponding digital polygon. While noisy, the boundaries of the village can be sufficient to do this task. Thus, the problem reduces to aligning the boundary of the map to the digital polygon.

First, we pre-process the raw image to remove all artifacts from the interior and exterior of the map such that only the boundary of the map remains in the image. We then frame the alignment problem as one of learning an affine transformation to map points in the source map to the target polygon. An affine transformation would only allow rotation, skewing, scaling and optionally, shearing of an image. An image can be represented as a matrix with dimensions ($pixels_x$, $pixels_y$, $channel$) and similarly, an affine transformation can be represented as a matrix of parameters that can be represented as in Equation 1

$$T_{affine} = \begin{pmatrix} sx \cdot \cos(a) + \text{shear\_x} \cdot \sin(a) & -sx \cdot \sin(a) + \text{shear\_x} \cdot \cos(a) & tx \\ sy \cdot \sin(a) + \text{shear\_y} \cdot \cos(a) & sy \cdot \cos(a) - \text{shear\_y} \cdot \sin(a) & ty \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

The alignment transformation can be represented as

$$M_{target} = T_{affine} M_{source} \tag{2}$$

Where M_{target} is the target polygon, M_{source} is the raw map's boundary mask and T_{affine} is the (learned) affine transformation matrix.

We chose to restrict ourselves to learning an affine transformation because the maps we are working with are digitally drawn and it is unlikely that they would require a non-linear transform like a thin-plate spline which is often useful for georeferencing scanned maps, but this method can easily be extended to those transforms - it's only a matter of learning more parameters.

We design a novel self-supervised algorithm to learn a transformation between a source and target image using synthetically generated data. Our algorithm is inspired by a Spatial Transformer Network which is designed as a drop-in module to make models perform tasks like digit classifications by allowing the model to spatially transform the input data while performing the task in a very efficient manner, without the need for extra training or special optimization processes (Jaderberg *et al.*, 2016). We build on the idea that neural networks can be used to learn image transformations. We design an architecture that would take in a

Source Image (M_{source}) and a target image (M_{target}) and return the T_{affine} matrix. We do this by first feeding in the images to two different encoding networks E_{source} and E_{target} to get their vector representations. In order to keep the computations lightweight and within a small budget, we utilize a MobilenetV3 encoder (Howard *et al.*, 2019) to encode our images into a vector (embedding) of 512 dimensions each. We then concatenate both of these embeddings into a vector of 1024 dimensions and feed it through a 3-layer Multi-layer-perceptron with *relu* activation (MLP), with the final layer having an output dimension of 6 - the unknowns in the affine transformation matrix. So, essentially, we need to learn the parameters of E_{source}, E_{target}, and MLP. We do this by utilizing a massive dump of shp files scraped from the internet and generating paired (source, target) images by rendering each polygon in the file as images and then randomly skewing one of them. We then learned to align these synthetic pairs of images by feeding them through the networks and learning to maximize the image similarity between the transformed source and target image. We measure image similarity by the Multi-Scale Structural Similarity (MSS) metric (Wang *et al.*, 2003). While any other measure of image similarity can also work (like Mean Sqaured Error), MSS outperformed other objective functions empirically.

Figure 4 outlines the training architecture described above. The objective function can be represented as

$$\min -\mathrm{MSS}\left(T_{\text{affine}}\left(M_{\text{source}}\right), \left(M_{\text{target}}\right)\right) \tag{3}$$

Here, T_{affine} represents the affine transformation parameters to be optimized, the source and target images are represented by M_{source} and M_{target} respectively, and MSS represents the Multi-Scale Structural Similarity function between the two images. Note that the f and g encoders don't appear in the objective directly - the actual loss calculation only utilizes the learned transformation matrix, to transform the source image and compare it with the original target image. However, note that those encoders' and also the MLP's parameters are learned during training.

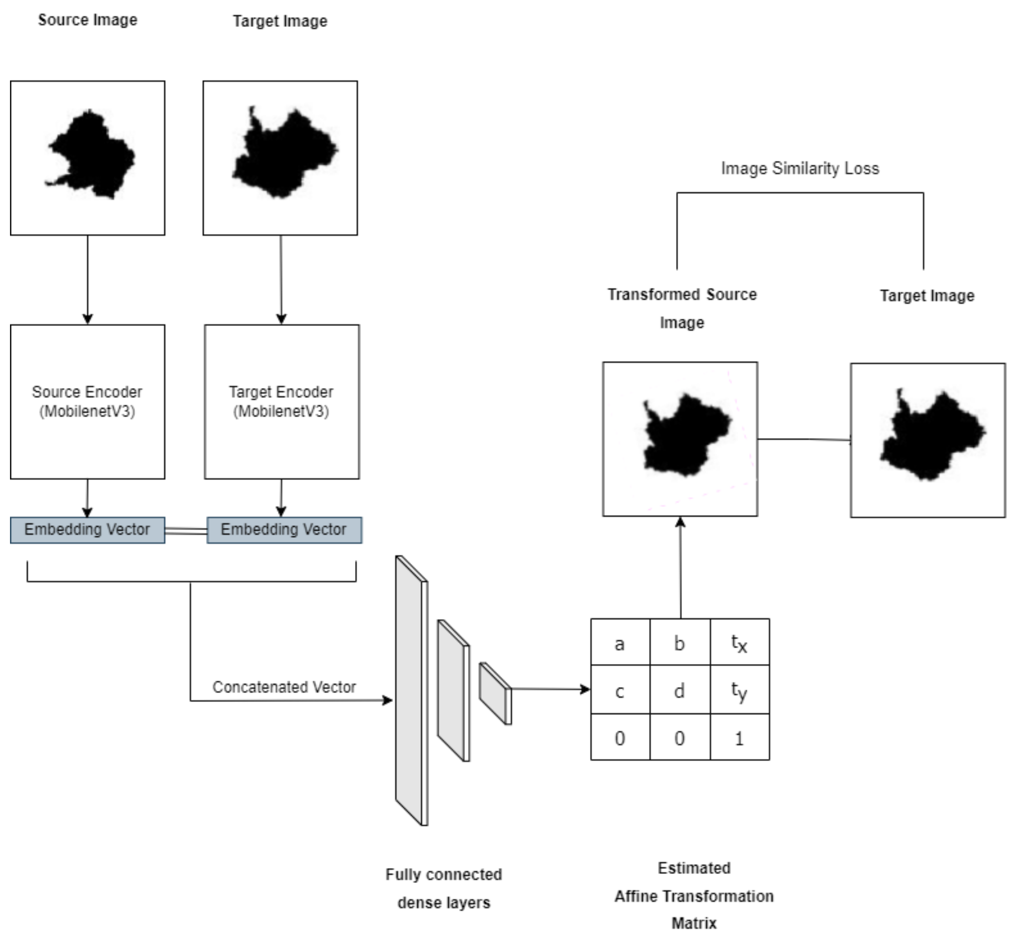


Figure 4: Model Architecture

Once trained, we use the model to take in paired map images (boundary masks) and polygon render of the corresponding shape file and obtain the affine transformation needed to align the map to the polygon.

Since the raw map is aligned to the polygon render whose pixels by definition can be mapped to the geographic plane, we now can map each pixel within the raw map's image to the geographic plane, effectively georeferencing it!

We designed this architecture to fit into our compute budget - all of the training was completed on a free-tier google colab instance within 120 minutes (with a GPU). Our pipeline is also efficient in inference - we can georeference a map within 2 seconds. We will release methods paper and a package to explain the architecture further.

(a)

(b)

(c)

Figure 5: Some examples of georefenced images

The alignment performance may not be perfect at times - either the map or the polygon may be inaccurate. We use an Intserection-over-Union (IoU) measure to evaluate alignment performance. It is defined as the ratio of the intersection of the area covered by the mask (outer boundary) of the map-specific area within the aligned raw image and the area of the village polygon to the union of the two when both images are superimposed on each other. An example where we demonstrate how the correct match has a high IoU and an incorrect match has a low-IoU is shown in Figure 5.

90% of our data has an IoU greater than 80% depicting a high proportion of well-matched maps. The algorithm fails to perform when a village is represented as fragments over multiple pages. Less than 10% of all villages had this issue. Other reasons for failure can include the non-existence of an accurate shape file for a particular village, fixing which is beyond the

(a) Rejected transformation (IoU < 50%)

(b) Accepted transformation (IoU > 90%)

Figure 6: Comparison of rejected and accepted transformations based on IoU

scope of our method.

We show the distribution of our georeferenced villages over a map of Karnataka in Figure 7. The grey regions can be either a missing village or an urban region.

Linking maps to Admin data Since locations are only available at the level of survey number, we merge maps at the survey number level. The maps on the Bhoomi portal were created in the year 2005 while our data on property owners is from September 2023. It is possible that some survey numbers were split or new survey numbers were created due to reasons such as government acquisition. Out of the 24,000 villages, we are only able to digitize 18,945 of them because either their files were corrupted or the village maps spanned multiple pages - where only a portion of each village was displayed on each page. Our pipeline can't handle these villages for now, and we drop them. Despite these two limitations, we can link 70% of the plots in our dataset to their geographic coordinates which gives us a massive set of 11 million properties and 16 million property owners.

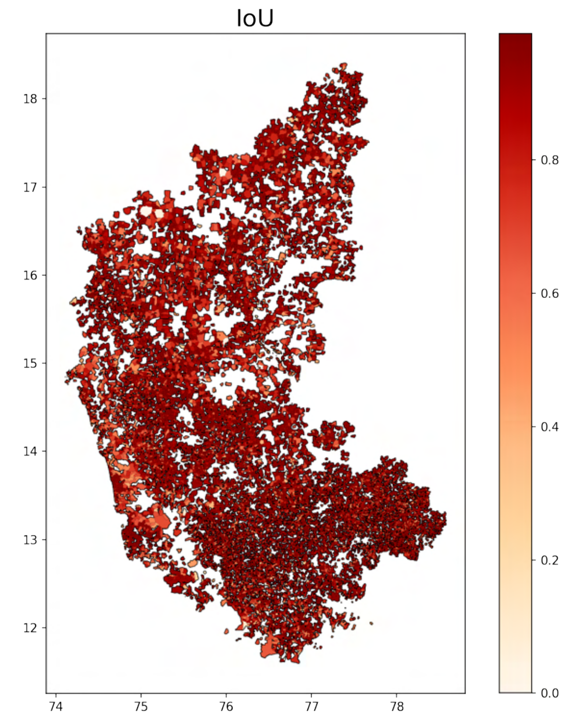


Figure 7: Polygons of georeferenced villages shaded by IoU. Note that gaps within the State polygon either represent areas where there are no villages (like Urban areas), or villages whose maps we couldn't digitize

Predicting religion Names in India can be very informative, especially in the case of classifying religion. Their use to classify religion and caste at scale is very well-documented (Ash *et al.*, 2021; Adukia *et al.*, 2023; Chaturvedi & Chaturvedi, 2020). We slightly digress from common approaches that use an LSTM neural network by using a transformer language model (BERT-based) instead (Devlin *et al.*, 2019). A transformer language model allows us to use massive pertaining of the network making the training of the classifier very efficient with minimal effort required to pre-process the text. We use the LinkTransformer library (Arora & Dell, 2023) to train the model with training hyperparameters given in the Appendix.

The names in our dataset are written in either the vernacular (Kannada) or in English. This was thus an appropriate setting to use a multilingual BERT model (bert-basemultilingual-cased) that supports both Kannada and English languages. BERT models are used to seeing longer texts (names are very short), so we make a template for all the names structured as "The name of the property owner is {name}". We treat this as a binary classification task - to classify names as Muslim and non-Muslim. We use a set of both secondary sources and hand-annotated labels to get a training dataset for training the classifier. Since we are working with rural property ownership data, more commonly used sources for training data such as national/urban lists of people (voter rolls/exam candidates etc.). First, we use a list of local government leaders that was shared with us by the state government. Out of 43,000 leaders in it, only 2,250 of them are Muslim. (5.2%). We keep 12,000 non-Muslim names and all of the Muslim names from this dataset. Names of elites may not represent the average name of a village property owner. We augmented this set from a list of people from Karnataka selected for a "Hajj Yatra" - a state-sponsored pilgrimage program - a list of 9,790 names. Additionally, we labeled 10,000 Muslim and 10,000 Hindu names from our dataset of property owners to get more in-domain data. Combining data from all of these sources gives us around 22,000 Muslim and non-Muslim names each. We then split the data into a 70-15-15 train-validation-test split to train and evaluate the model's performance. We then transliterated the Kannada (English) names to English (Kannada) and appended that to the dataset. This allowed us to train a multi-lingual model that achieved an F1 score of 97% (Accuracy of 96.7%, recall of 95.6%, and precision of 98.1%) on the test dataset - which was unseen in the training. [6] We then use this model to predict the religion on the names of property owners in our dataset of a massive 16.3 million properties. 4.7 percent of all names in our dataset are classified as Muslim - closely resembling the proportion of elite Muslim households in the village leaders dataset, but significantly lower than the state's average proportion of Muslims in rural areas (7.6% as per Census 2011). This is likely because Muslim households in India are among the least wealthy social groups in the country and it is entirely

[6] Note that the test dataset comprises a mix of manually annotated names from the property ownership dataset (where annotators predict the religion based on the name) and names from the village leaders and pilgrimage datasets, for which the religion is known with certainty. For the subsample of names whose religion is known with certainty, the F1 score (and accuracy) are slightly higher, at 98% and 97% respectively.

plausible that a smaller proportion of Muslim households would be in the land-owning class than their non-Muslim counterparts. We thus focus on segregation among property owners and not residents per se. However, inter-district differences in proportions are similar in terms of their ranking within the state (rural) between both our data and data from Census 2011. The correlation coefficient between the two series is 0.92. We demonstrate the same in Figure 8.

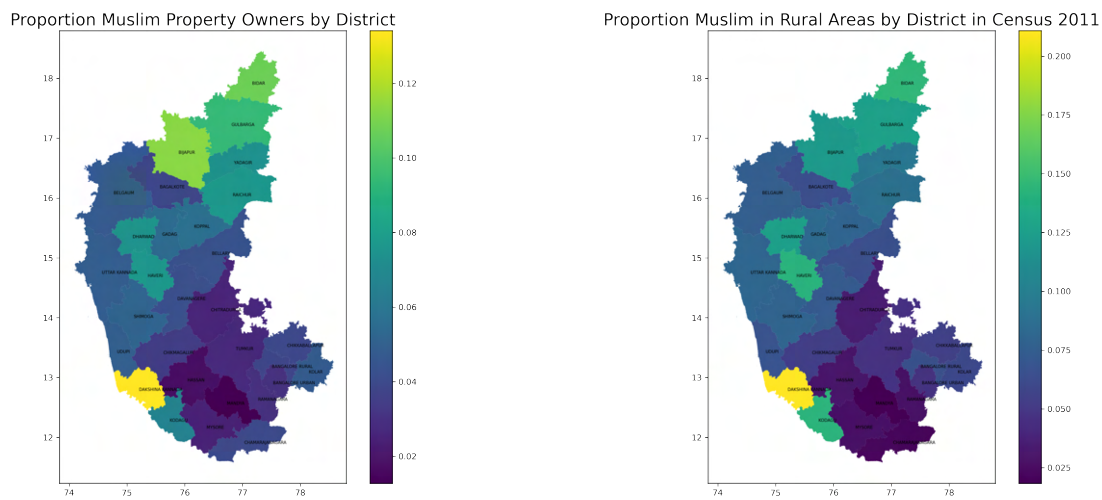


Figure 8: Benchmarking the district-level proportion of Muslim Property owners with estimates from Census 2011. The correlation between the two series is 92%.

We show the distribution of non-Muslim and Muslim households within a random village from the state in Figure A2.

Neighbourhood Construction We tessellate villages into grids of equal areas. We used several resolutions for documenting the levels of segregation in the state (Section 5). For downstream analyses, we use neighborhoods of size 500m X 500m to measure neighborhood Muslim shares.

Merging-in Public Facilities and Dargahs We have the geographic coordinates of all public facilities and Dargahs, so we spatially join them into the neighborhoods in our dataset. A neighborhood may contain multiple Dargahs or Facilities, but we only focus on the extensive margin - whether or not a neighborhood contains the establishment. By design, we allow for multiple neighborhoods (say, at the adjacent corners of 2 villages) to contain the same establishment as our main outcome (and instrument).

5 Documenting village-level segregation

To examine the extent of segregation in the state, we use several measures at different levels of aggregation. We use the widely used Dissimilarity and Isolation indices. These indices are defined as an aggregation over neighborhoods, usually census tracts (Massey & Denton, 1993; Cutler & Glaeser, 1997; Adukia *et al.*, 2023).

$$DISSIMILARITY_V = \frac{1}{2}\sum_{i=1}^{n}\left|\frac{m_i}{M} - \frac{a_i}{A}\right| \tag{4}$$

$$ISOLATION_V = \sum_{i=1}^{n}\frac{m_i}{M}\cdot\frac{m_i}{m_i + a_i} \tag{5}$$

where m_i is the number of Muslim landowners in neighborhood i, M is the total number of Muslim landowners in the village, a_i is the number of non-Muslims in the neighborhood and A is the total number of non-Muslim land-owners in the village. For the rest of the paper, we will use the terms landowner and households interchangeably.

$DISSIMILARITY_V$ represents the proportion of Muslim households who would need to move (in this case, landowners) to another neighborhood to bring the share of Muslims in each neighborhood equal to their share in the entire village.

$ISOLATION_V$ on the other hand represents the extent to which a Muslim landowner is only exposed to other Muslim landowners in the village. It can be understood as the share of Muslims in the neighborhood of an average Muslim inhabitant in the village.

Both indices range from 0 to 1, with higher values implying more segregation.

Our dataset does not contain predefined neighborhoods, so we need to define our own. Since we know the location of each land parcel within a village, we can define neighborhoods by splitting a village into grids with gridcells of a specific size. Segregation measures are not scale invariant, so different levels of aggregation would lead to different estimates. Specifically, the smaller a neighborhood gets, the larger the Dissimilarity and Isolation indices would be. While absolute levels differ, relative patterns continue to hold across different levels of aggregation. First, it is important to note that amongst all the villages in our sample, the average proportion of Muslim households is just around 4%. On the other hand, calculated on neighborhoods defined on 500m X 500m, dissimilarity is at 55% and Isolation is at 14.5%. The difference between the average Muslim share and Isolation reveals the extent of the levels of segregation - the proportion of Muslim neighbors in the neighborhood of an average Muslim inhabitant is over 3 times higher than the overall Muslim share in the village. Within a neighborhood of 250m X 250m, Isolation rises to 23%. One in twenty households in the village is Muslim, but a quarter of neighbors of an average Muslim household are Muslim.

In the Indian setting, recent work by Adukia *et al.* (2023) measures segregation on neighborhoods defined from census enumeration blocks - each typically containing 100-125 households (700 people). They point out that 50% of villages in India make up an entire neighborhood which makes traditional segregation measures powerless at the village level. Therefore, for rural areas, they end up calculating segregation indices at the level of the subdistrict that can contain about 110 villages. Our data is different from the Census used by Adukia *et al.* (2023) in two ways - one, it is from a decade after the census, and two, it is a record of all property owners - landless people are not covered. For simplicity, we would use the term household while talking about land-owners. Segregation in our dataset technically, would be exactly defined as segregation among property owners. In the near future, we will perform a benchmarking exercise with their estimates to understand the extent of the discrepancy -

if any[7]. For now, we measure the indices on neighborhoods with varying sizes, 1kmX1km, 500mX500m, and 250mX250m.

In a separate benchmarking exercise, we compare our segregation estimates with Black/White segregation which has been extensively studied in the context of U.S. cities. We take indices calculated for the year 2000 by Cutler *et al.* (2010). These estimates have been calculated for Metropolitan Statistical Areas (MSAs) which contain at least 1000 black residents with each neighbourhood within it containing approximately 4000 residents. While our focus is on rural segregation, it is helpful to compare magnitudes from a well-studied context. To make the populations more comparable, we measure Isolation and Dissimilarity over subdistricts instead of villages and drop those subdistricts that have less than 250 Muslim households. The population density in Karnataka is projected to be about 349 people per square km in 2021 according to the latest available population projections [8]. An appropriate neighborhood for comparison with the U.S. context would thus be larger than most villages. So instead of dividing up a village into neighborhoods, we tessellate a subdistrict into grids ranging from 3.5 X 3.5km (comparable to U.S. tract sizes of 4000 people) and 1.5km X 1.5km (comparable to the enumeration block size of 700 residents) [9]. Note that the population density is for the entire state, not just rural areas. It is likely therefore that our segregation levels are an overestimate. We thus want to provide two more neighborhood sizes - 5km X 5km and 2 km X 2km to give some more conservative estimates of segregation in the future. For now, we show densities with for an initial benchmarking exercise in Figure .

It is evident from the figure that religious segregation in rural Karnataka is not as high as Black/White segregation in terms of Isolation or Dissimilarity. Note that this is expected mechanically for the Isolation index because it is highly correlated with the proportion of Muslim households in the region - which is way lower in rural India than U.S. MSAs.

[7]The authors are in the process of preparing their data for us - they don't have data for a specific state publicly available.

[8]Population projections from the 2011 Census by National Health Mission, Government of India: https://nhm.gov.in/New_Updates_2018/Report_Population_Projection_2019.pdf

[9]A density of 349 per squared kilometers implies that a neighborhood of 4000 people requires an area of 12 squared kilometers and a neighborhood of 700 people requires an area of 2 squared kilometers.

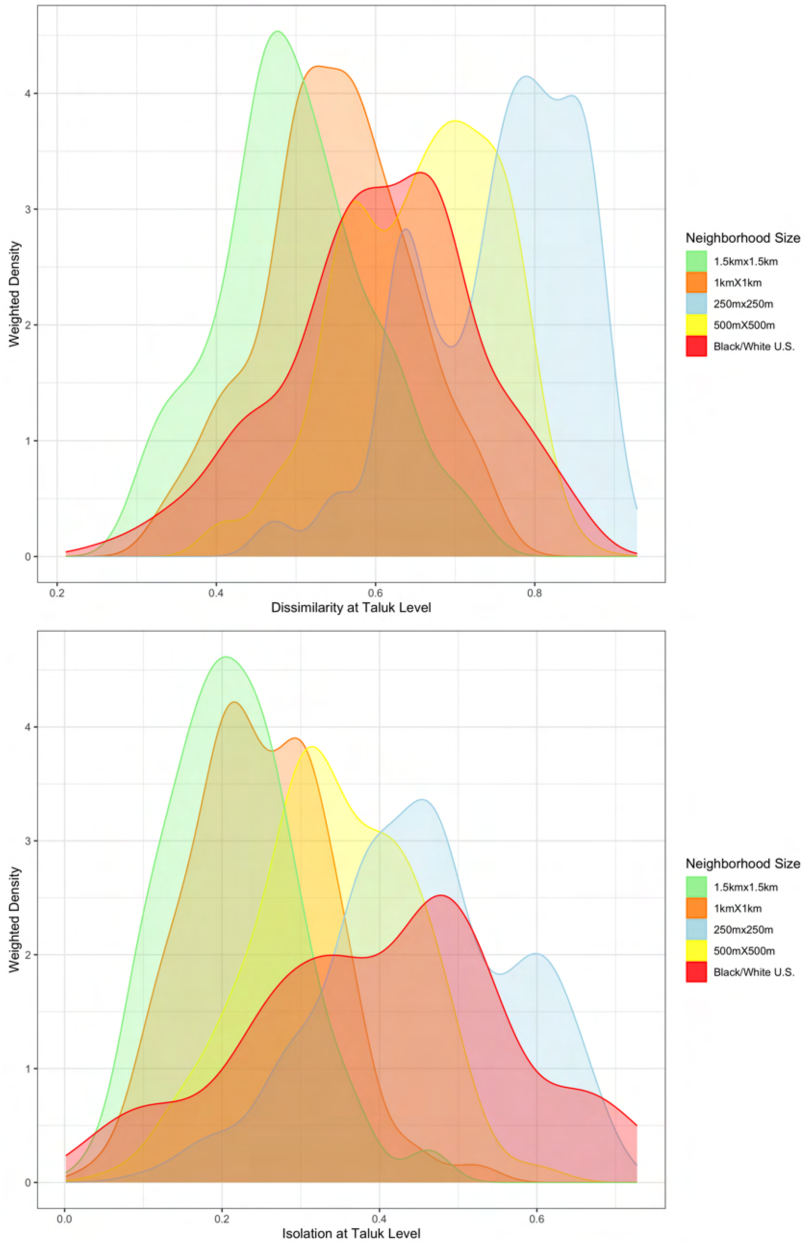


Figure 9: Distribution of Isolation and Dissimilarity Indices at the Taluk/Subdistrict level. Only for this figure, we weight the densities by the number of Muslims living in the subdistrict. U.S. estimates were obtained from (Cutler *et al.*, 2010).

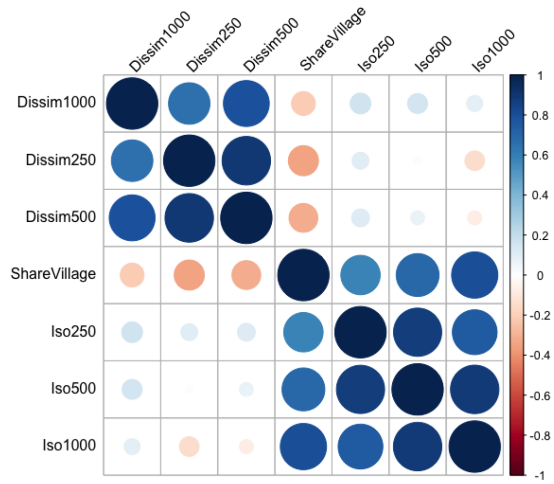


Figure 10: Correlation matrix between the various measures we use to document segregation. We only keep villages that have at least one Muslim household before computing correlations

Note that due to the resolution of our dataset, for the first time, we can capture village-level segregation. We report the Dissimilarity and Isolation indices at the **village** level at three different sizes of each neighborhood (250m x 250m, 500m X 500m and 1km X 1km) in Figure A3. There is interesting geographic heterogeneity, regions in the north-eastern part of the state (which was also a part of the Hyderabad state ruled by the Nizam until 1948) have higher values Isolation, a pattern that is also seen in the district-level proportions of Muslim residents in 8. Dissimilarity values on the other hand vary widely across the state but are still higher in the north relative to the south. Larger neighborhood sizes would mechanically attenuate the magnitudes of segregation measures but it is clear from the figure that relative patterns continue to largely hold.

Next, we plot the correlations between the village-level estimates of Dissimilarity, Isolation (at the three resolutions), and Village level Muslim shares in 10. For better comparability across the measures, we drop those villages that contain no Muslim households.

It is evident from the plot that both segregation measures do not have a high correlation with each other - both of them capture different dimensions of segregation. Isolation is closely linked to overall shares in the village and isolation measured with larger neighborhood sizes is closely linked to the overall village's Muslim share.

While DISSIMILARITY and ISOLATION estimate village-level segregation, they don't fully utilize the richness of our data. Since we observe the coordinates of each land parcel, we can define more granular and intuitive measures of clustering at the village level that take the geographic spread of minority households into account. We leave that out of the scope of this work in the interest of brevity, but it is something that we can use to provide richer descriptions of segregation which the standard indices don't provide.

One caveat while using our data to comment on segregation is that we don't exactly know who the resident of each household is. Since we are working with property ownership data, we are by design looking at wealthier households who own properties. If wealthier Muslims would choose to live in less segregated areas, what we have is likely a lower bound on the actual segregation levels. However, if wealthier households choose to co-locate with community members, the levels could go in the opposite direction.

6 Minority Access to Public Goods

6.1 Empirical Strategy

We now try to understand the link between a household's access to public goods and the religious composition of its neighborhood. Unlike previous work, the resolution of our dataset allows us to look at the effects of segregation both within and across villages.

First, we tessellate each village into grid cells of size 500m x 500m. Then, we count the total number of households Muslim and non-Muslim households in the neighborhood. For each neighborhood i, we calculate the share of Muslim households in the neighborhood ($SHARE_i$). We then examine public good access within a village.

$$D_{ig} = \beta_0 + \beta_1 SHARE_i + X_i + Z + \epsilon_i \tag{6}$$

where D_{ig} is a dummy that equals 1 when the neighborhood i contains any public good of type g (Categories include Health, Education, Transport/Admin or Agricultural/Market). $SHARE_i$ is the proportion of Muslim households in the neighborhood i. Z contains Village fixed-effects. We cluster the standard errors at the village level. Public goods within a village may be located in areas with a larger population and possibly, in those containing a higher share of households in the village (neighborhood density). We add controls for these in X_i. We remove those villages from the analysis which have no Muslim households.

A negative coefficient on β_1 would imply that on average, neighborhoods with a larger proportion of Muslim land-owners are less likely to contain a public good.

The level of aggregation can also matter when looking at the relationship between segregation and access to public goods. The placement of public goods may be planned by the local government (GP) which is usually in-charge of several villages (4-5 on average). The GP may favor one village over others while deciding the location of public goods. We estimate another specification to look for the effect of the Share of Muslims in a Village, within the GP. The instrument for this variable would simply be a dummy variable that takes the value 1 if the village contains a *Dargah* - villages that contain these medieval tombs would tend to have a higher share of Muslim property owners. Here instead of Village FE, Z would contain GP FE. Since we are looking at households from a coarser level, we can add Village-level controls from Census 2011 and SECC 2013 including Village Area, Distance to the Nearest town, proportion of scheduled caste households, prop. of scheduled tribe households, per capita consumption, the share of households employed in cultivation, village area, average elevation, vegetation, and terrain ruggedness. Borrowing the notation from above, the equation we estimate would take the form.

$$D_{ig} = \beta_0 + \beta_1 SHARE_i + \beta_2 SHAREVilllage + X_i + Z + \epsilon_i \tag{7}$$

Here, X_i also contains village-level controls described above (same across households in the village) apart from household-level controls and Z corresponds to GP fixed effects. As before β_1 gives the relationship between the neighborhood share of Muslims and the presence of public goods. β_2 on the other hand gives the relationship between village-level share of Muslims and the presence of public goods of type g - a negative coefficient would imply that neighborhoods in villages with more Muslim households relative to non-Muslim ones are less likely to contain public goods.

These relationships, however, are not causal. It may be the case that Muslim households are choosing to live in under-provisioned neighborhoods (or villages) which in turn may be under-provisioned due to other reasons. This would mean that the lack of access to public goods in neighborhoods or villages with a higher share of Muslims may just be a result of selection and not stemming from the level of minority share in the area. Moreover, we don't observe everything about the property owners. For instance, it may be the case that most of the agricultural land in the village is owned by non-Muslims, which tends to be away from village centers that contain more public goods - skewing the relationship in the positive direction. Crucially, we don't have a measure of the wealth of the property owner, the value of the property, or even the type of the parcel - agricultural, commercial, or residential.

To address these endogeneity concerns, we construct a novel Instrumental Variable to look at the effect of high Muslim shares in a neighborhood on public good provision. We use a dummy that takes the value 1 if the neighborhood contains a *Dargah* (Sufi Shrine described in Section 3) as an instrument for $SHARE_{id}$. More Muslims live near these shrines - a fact that is supported both anecdotally and seen in our data from a strong first stage in Table 2 where we confirm that the share of Muslim neighbors is strongly correlated (negatively) with the *Dargah* dummy (and presence of Darghah in a village is strongly positively related with the proportion of Muslim households in the village - instrument for village level Muslim shares for the next specification). Neighborhoods with a Dargah have a 5.6% higher share of Muslims relative to neighborhoods that don't have a Dargah - almost a 50% difference. Within a Gram Panchayat (GP), villages that have a Dargah have a 5% larger share of Muslim land-owners relative to other villages in that GP which is 55% more than the average village-level Muslim share. Moreover, these shrines were built between

13-19th Centuries - it is unlikely that public goods built centuries later had any relationship with the locations of these shrines. Even if one argues that these shrines may have been built around important village centers or places of high population density centuries ago and that these places continue to be important, controlling for present-day neighborhood population and density should be helpful in strengthening exogeneity. Another possibility is that these Dargahs were built in places that historically had more public or private facilities and more sprang up over time as the village grew. We further strengthen our IV by controlling for amount of built-up area in the neighborhood. There are satellite-imagery-based datasets out there for the same (like the one provided by Meta[10]), but for now, we use a proxy for them. We focus on only Education and Health facilities because they are more cleanly defined in our facilities dataset as "Admin/Transport" and "Agro" are very broad categories and can encapsulate a wide range of local infrastructure including Bus stands, railway stations, markers, food processing centers etc. We instead use these broad categories - specifically, the presence of any of the facilities of the types "Admin" or "Agro" in the neighborhood as proxies for the level built-up area. We will stick to these proxies for now, but the fact that the plots are all georeferenced would allow us to merge these satellite imagery-based datasets with relative ease. The presence of any Mosque (not a *Dargah*) on the other hand doesn't share the nice properties of this IV. New mosques can be built in areas where more Muslims live - violating the exogeneity assumption. Similarly, one can't use the presence of Hindu temples as a way to isolate these effects because there is no clean way for us to know the date of establishment of these temples. *Dargahs*, due to their historically predetermined location from centuries ago, combined with appropriate controls thus provide us with a valid instrument that can help us isolate the effect that we are interested in.

Using this instrument, we therefore can provide for the first time, a causal link between the religious composition of a neighborhood or a village and minority access to public goods.

There is a caveat, however. *Dargahs* are not present in every village. To ensure that our

[10]High-Resolution Population Density Maps: https://data.humdata.org/dataset/pakistan-india_all-files-high-resolution-population-density-maps

results are not due to spurious reasons, our IV estimations, we restrict our sample to villages that have a Dargah for the within-village regressions and to GPs that contain at least one village containing a Dargah for the within-GP regressions.

6.2 Early Results

Table 1 shows Ordinary Least Squares estimates from equation 6 in Panel A and equation 7 in Panel B. Neighbourhoods with a higher Muslim share tend to contain more public goods, of all types both within a village and within a GP. Specifically, compared to neighborhoods with no Muslim inhabitants, all-Muslim neighborhoods in the same village are 0.5 percentage points more likely to have an educational facility (A 31% increase compared to the mean). We will focus the discussion on educational facilities but the results are symmetric with health facilities. While this effect looks large, it needs to be unpacked further. A median neighborhood in our sample doesn't have a Muslim household and a neighborhood on the 75$^{\text{th}}$ percentile has a Muslim share of 2.5%. This implies that compared to the median neighborhood, a neighborhood on the third quartile in terms of Muslim share is only 0.0125 % more likely to have an educational facility - which is less than a less than 1% increase compared to the average likelihood of having an educational facility. As discussed in the previous section, endogeneity can lead to a severe bias. Nevertheless, it is important to note that the coefficient on the *Share* is positive and largely significant for all four types of facilities. For our IV specifications in 3, as discussed previously, we only consider only those villages (or GPs) that contain a *Dargah* and focus our attention on Educational and Medical facilities while using the presence of other facilities as a control for the built-up area.

First, we find that our coefficients of interest in the IV specifications are strongly significant and have a much larger magnitude than OLS estimates. Endogeneity had caused a significant underestimation of the effect size. In Panel A, within a village, we find that an increase of 10% in the Muslim share in a neighborhood increases the likelihood of the presence of a school by 5% relative to other neighborhoods of similar population levels and density. However, the scale of concentration matters. Within a GP (villages with the same local government), neighborhoods in villages with a 10% higher overall share of Muslim inhabitants would have a 4% lower likelihood of having a school. The compound effect can understood by considering a hypothetical scenario. Suppose that for any reason, 10% non-Muslim households move out of each neighborhood in a village and are replaced with Muslim inhabitants - increasing both the overall village share and neighborhood share by 10%. The benefits that neighborhoods get by the presence of 10% additional neighbors who are Muslims would reduce by 80% (from 5% to 1%).

This becomes clearer when we split the sample into villages with more and less than median village shares. The benefit of local clustering is amplified in villages that have a smaller share of Muslims at the village level - an 11% increase in the likelihood of having a school for every 10% increase in the neighborhood share compared to a 3.7% increase for neighborhoods in villages than have an above-median proportion of Muslim inhabitants.

One important point to note here is that we are trying to measure if neighborhoods can bargain for better placement of public goods in the village depending upon their demographic composition. For now, we don't know the date the facility was constructed. However, that doesn't necessarily invalidate the analysis. Muslim households clustering around Dargahs is not a recent phenomenon. Even if these facilities were constructed decades ago, as long as current neighborhoods' Muslim shares can proxy for the share further back in time, one can argue these effects are indeed driven by the concentration of minority households and not any other reasons. Nevertheless, we plan to enrich our analysis by using data (that we have already obtained) that records the universe of development works across all villages in Karnataka, including its entire life cycle. By using only works started in the past couple of years to measure access, we can test whether this effect of clustering is not driven just by previously constructed works. Moreover, if we can obtain richer property-level data, specifically, the date the landowner purchased the property (or inherited) it, we can look at the dynamics more holistically.

The fact that the IV coefficients are much larger than the OLS coefficients warrants further exploration of the channels that may be driving this result. We plan to explore them in future work.

We focus our attention on Public Good access as we don't have more information about the household beyond what's available to us from the property ownership and geospatial data. The relationship between access to educational and health facilities on related outcomes has been well studied. Enrollment in schools is higher, especially for girls if a school is nearby (Das & Das, 2023; Burde & Linden, 2012; Muralidharan & Prakash, 2017). Proximity to health facilities is positively linked to several health outcomes including institutional deliveries, increase in treatment-seeking, and maternal and child health (Okwaraji & Edmond, 2012; Santosh Kumar & Murray, 2014). We leave an exploration of the effects of segregation on educational and health outcomes for future work.

7 Historical roots of present-day segregation

7.1 Empirical Strategy

Besides isolating the effects of segregation on a household's access to public goods, we utilize Karnataka's historic partitions to look at historical channels of segregation in the state.

Specifically, we employ a geographic Regression Discontinuity Design based on Dell (2010) around the border of the territory once ruled by the Nizam of Hyderabad. As described in section 2, the Hyderabad state ruled by Nizam overlapped with 5 present-day districts of Karnataka spanning an area of 37,200 square km. The border between this region (referred to as "Nizam border" henceforth) and other regions of present-day Karnataka - the Bombay and Madras presidencies ruled by the British from 1799-1947 was approximately 900km long. Figure 11 depicts the villages in a 100km bandwidth around the boundary.

Our outcomes of interest are the two village-level segregation indices Dissimilarity and Isolation (constructed using 500m X 500m neighborhoods within each village) and the over-

Table 1: Access to Public Goods - Number of Public Facilties in Neighborhood

	(1)	(2)	(3)	(4)
	Educational	Medical	Agri/Markets	Admin/Transport
	Panel A : Within a Village			
Share	0.005***	0.002*	0.002	0.008***
	(0.001)	(0.001)	(0.001)	(0.002)
Observations	393585	393585	393585	393585
Mean Y	0.016	0.009	0.023	0.026
	Panel B : Within a GP			
Share	0.007***	0.003**	0.003**	0.011***
	(0.002)	(0.001)	(0.002)	(0.002)
Share Village	0.000	–0.004	–0.013**	–0.002
	(0.005)	(0.004)	(0.005)	(0.006)
Observations	315166	315166	315166	315166
Mean Y	0.016	0.009	0.023	0.027
Controls				
Village/GP FE	✓	✓	✓	✓

Notes: The dependent variable is the Number of Public Facilties in Neighborhood of the type Educational (School/Collage), Medical (Health Centre, Hospital/ Dispensary), Agro/Markets (Agricultural Faciltiies and Market Infrastructure) and Admin/Transport (Administrative office/Transport facility like a Bus Stand). Share of Muslim Neighbours is the proportion of muslim neighbors in the neighborhood. We control for local population density by adding the share of households in the village that are in the neighborhood. Standard Errors are clustered on the village level. Panel A has Village Fixed effects and Panel B has GP Fixed Effects. For Panel B, we additionnaly control for village-level covariates obtained by linking to Census 2011 and SECC 2013 - Distance to the nearest town, proportion of scheduled caste households, prop. of scheduled tribe households, per capita consumption, share of households employed in cultivation, village area, average elevation, vegetation and TRI. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: First Stage

	(1)	(2)	(3)
	Within Village	Within GP	
	Share	Share	Share Village
Dargah in NBD	0.056***	0.067***	0.004*
	(0.008)	(0.010)	(0.002)
Dargah in Village		0.048***	0.050***
		(0.007)	(0.007)
Observations	33590	42218	42218
Mean Y	0.109	0.085	0.089
Controls			
Village FE	✓		
GP FE		✓	✓

Notes: The dependent variable is the Number of Public Facilties in Neighborhood of the type Educational (School/Collage), Medical (Health Centre, Hospital/ Dispensary), Agro/Markets (Agricultural Faciltiies and Market Infrastructure) and Admin/Transport (Administrative office/Transport facility like a Bus Stand). Col 1 gives contains all villages containing a Dargah, Col 2 and 3 further subset by keeping those villages which have below and above median Muslim share (0.0587). Share of Muslim Neighbours is the proportion of muslim neighbors in the neighborhood. We control for local population density by adding the share of households in the village that are in the neighborhood. Standard Errors are clustered on the village level. Panel A has Village Fixed effects and Panel B has GP Fixed Effects. For Panel B, we additionnaly control for village-level covariates obtained by linking to Census 2011 and SECC 2013 - Distance to the nearest town, proportion of scheduled caste households, prop. of scheduled tribe households, per capita consumption, share of households employed in cultivation, village area, average elevation, vegetation and TRI. IV regressions in Panel A uses A dummy for Dargah in Neighborhod as an instrument and Panel B uses A dummy for Dargah in Neighborhod and dargah in village as instruments. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Access to Public Goods - Number of Public Facilties in Neighborhood

	(1)	(2)	(3)	(4)
	OLS		IV	
	Educational	Medical	Educational	Medical
	Panel A : Within a Village			
Share	0.003	–0.000	0.534***	0.609***
	(0.003)	(0.003)	(0.179)	(0.179)
Observations	33590	33590	33590	33590
Mean Y	0.018	0.011	0.018	0.011
FS F-Stat			47.939	47.939
	Panel B : Within a GP			
Share	0.011***	0.003	0.532***	0.674***
	(0.004)	(0.003)	(0.197)	(0.215)
Share Village	–0.007	–0.004	–0.433**	–0.649***
	(0.010)	(0.007)	(0.199)	(0.214)
Observations	42218	42218	42218	42218
Mean Y	0.015	0.010	0.015	0.010
FS F-Stat			18.721	18.721
Controls				
Village/GP FE	✓	✓	✓	✓

Notes: The dependent variable is the Number of Public Facilties in Neighborhood of the type Educational (School/Collage), Medical (Health Centre, Hospital/ Dispensary), Agro/Markets (Agricultural Faciltiies and Market Infrastructure) and Admin/Transport (Administrative office/Transport facility like a Bus Stand). Share of Muslim Neighbours is the proportion of muslim neighbors in the neighborhood. We control for local population density by adding the share of households in the village that are in the neighborhood. Standard Errors are clustered on the village level. Panel A has Village Fixed effects and Panel B has GP Fixed Effects. Panel A only contains those villages that have a Dargah and Panel B contains those Panchayats that have a Dargah - so naturally they will also have villages which don't have a Dargah. For Panel B, we additionnaly control for village-level covariates obtained by linking to Census 2011 and SECC 2013 - Distance to the nearest town, proportion of scheduled caste households, prop. of scheduled tribe households, per capita consumption, share of households employed in cultivation, village area, average elevation, vegetation and TRI. IV regressions in Panel A uses A dummy for Dargah in Neighborhod as an instrument and Panel B uses A dummy for Dargah in Neighborhod and dargah in village as instruments. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Access to Educational Facilities - Heterogeneity by Village Shares

	(1)	(2)	(3)
	Overall	$<= Median$	$> Median$
	Panel A : Educational Facility in Neighborhood		
Share	0.534***	1.109**	0.375**
	(0.179)	(0.531)	(0.174)
Observations	33590	15558	18032
Mean Y	0.018	0.019	0.017
FS F-Stat	47.939	14.302	33.019
	Panel B : Health Facility in Neighborhood		
Share	0.609***	0.878*	0.547***
	(0.179)	(0.470)	(0.190)
Observations	33590	15558	18032
Mean Y	0.011	0.011	0.010
FS F-Stat	47.939	14.302	33.019
Controls			
Village/GP FE	✓	✓	✓

Notes: The dependent variable is the Number of Public Facilties in Neighborhood of the type Educational (School/Collage), Medical (Health Centre, Hospital/ Dispensary), Agro/Markets (Agricultural Faciltiies and Market Infrastructure) and Admin/Transport (Administrative office/Transport facility like a Bus Stand). Share of Muslim Neighbours is the proportion of muslim neighbors in the neighborhood. We control for local population density by adding the share of households in the village that are in the neighborhood. Standard Errors are clustered on the village level. Panel A has Village Fixed effects and Panel B has GP Fixed Effects. Panel A only contains those villages that have a Dargah and Panel B contains those Panchayats that have a Dargah - so naturally they will also have villages which don't have a Dargah. For Panel B, we additionnaly control for village-level covariates obtained by linking to Census 2011 and SECC 2013 - Distance to the nearest town, proportion of scheduled caste households, prop. of scheduled tribe households, per capita consumption, share of households employed in cultivation, village area, average elevation, vegetation and TRI. IV regressions in Panel A uses A dummy for Dargah in Neighborhod as an instrument and Panel B uses A dummy for Dargah in Neighborhod and dargah in village as instruments. *** p<0.01, ** p<0.05, * p<0.1.

all proportion of Muslim landowners in the village. We calculate the distance of each village centroid from the Nizam border. We then use that as the continuous running variable in a local linear regression with covariates (Calonico *et al.*, 2019a) with the border forming the multidimensional cutoff point. We determine optimal bandwidths using a common Mean Squared Optimal bandwidth selector and use a triangular kernel on each observation's distance to the Nizam border. We estimate it with different bandwidths (50%, 100%, 150% of optimal), and with both linear and quadratic polynomials. The covariates include 30km border segment fixed effects, longitude, latitude, vegetation, ruggedness, elevation, village area, proportion of scheduled castes and scheduled tribe, distance to the nearest town (urban access), per capita consumption, share of households engaged in cultivation, manufacturing or services, the literacy rate and the total population of the village either obtained from survey/census data available on the SHRUG (cite) platform. We also use a specification that controls for Distance to the metros - Mumbai (Capital of the Bombay presidency), Bangalore (Present day capital of Karnataka), Hyderabad (Capital of the Nizam), Mysore (Capital of the Mysore kingdom), and Chennai (Capital of the Madras Presidency) to account for proximity to historic administrative centers. We conservatively cluster standard errors at the sub-district (taluka) level.

$$Y_{ivb} = \alpha + \gamma \cdot nizam_v + \beta X_{id} + f(Distance\ Nizam) + \phi_b + \epsilon_{ivb} \tag{8}$$

A positive coefficient on γ would imply that the outcome of interest is higher in a village on the Nizam side compared to a village on the non-Nizam side.

We perform standard falsification tests. First, we test for covariate balance around the Nizam boundary (using the optimal bandwidth retrieved from our main specification) in Table 6. Second, we test for continuity of other covariates around the boundary in Table 5.

The geographic covariates are largely balanced and do not vary discontinuously at the boundary. Other covariates are unbalanced - but they are not exogenous and could themselves be a result of the institutional differences at the Nizam boundary.

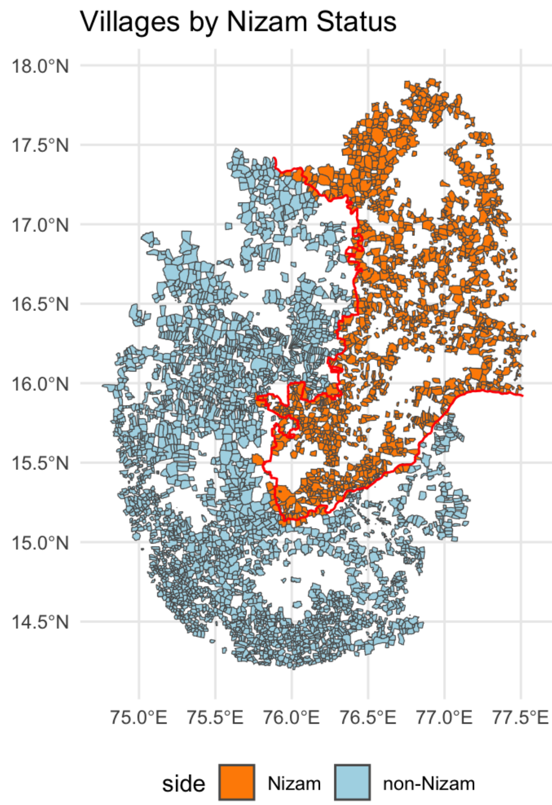


Figure 11: Villages around the Nizam boundary. The gaps between villages can be due to the presence of an Urban area, geographical obstructions or villages that could not be digitized.

7.2 Early Results

Table 7 contains the estimates of the *Nizam* effect on Isolation Index and Proportion of Muslim (landowners) households in the village [11]. We find no difference in Dissimilarity around the boundary while there is significantly higher isolation on the *Nizam* side across all permutations of bandwidths, polynomials, and covariates we have tried. This is plausible because, as noted in Section 5, Isolation and Dissimilarity in this context are not highly correlated. The isolation index can be high with low dissimilarity when while some Muslim households have been integrated into non-Muslim neighborhoods, an average Muslim household still lives in neighborhoods that are predominantly Muslim. Villages on the *Nizam* side have a 25-35% isolation index compared to other villages. Moreover, villages on the Nizam side have a higher share of Muslim property with a 28-38% higher Muslim share.

It is worth noting that like in Dell (2010), we are not estimating heterogenous treatment effects but account for them with border segment fixed effects. At this stage, we do not explore the mechanisms behind the effects we see. It is plausible that there is heterogeneity in the effects due to an underlying mechanism. Communal conflict could be a mechanism behind what we see in our results. At the time of the British withdrawal from India and right through the Indian "police action" termed Operation Polo which aimed to integrate the territory ruled by the Nizam into newly independent India, the region was rife with communal conflict. A "goodwill" mission was sent by the government which was led by Pandit Sunderlal. This report which has been declassified since then (in 2013), documents the loss of life and property in the region - and how it varied across districts. Among the five present-day districts in Karnataka that were a part of Nizam's kingdom, two of them - Gulbarga and Bidar were worst affected by the violence. If conflict leads to more segregation then it is plausible to expect heterogeneity in the "Nizam" effect. We plan to use the approach outlined by Keele & Titiunik (2015) that allows for the estimation of heterogeneous treatment effects at different points on the boundary in the future. We aim

[11] Omitted Dissimilarity from the table for concision.

to explore the mechanism behind these differences in the near future.

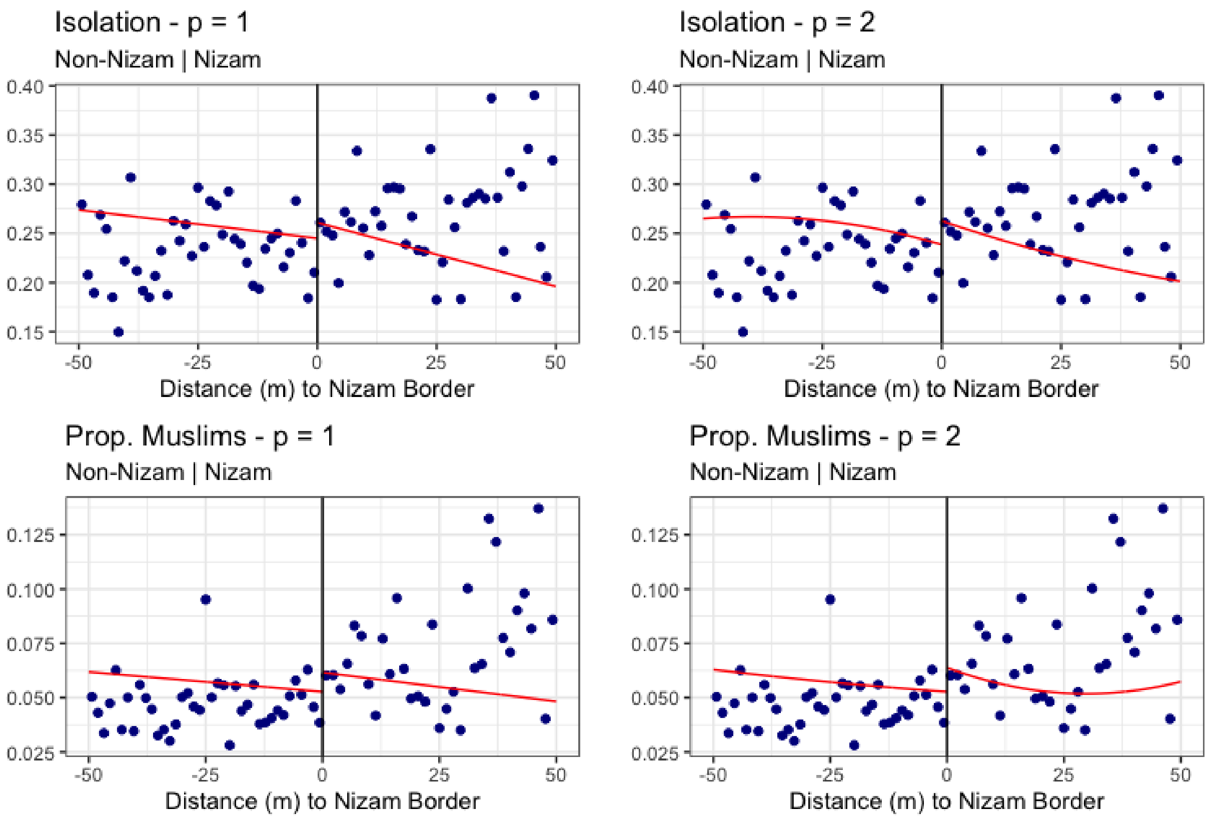


Figure 12: RD plots with different polynomial fits for Isolation and Muslim Share in a Village. All specifications control for SC/ST Share, lat, long, border segment fixed effects, and geographic controls (vegetation, elevation, and TRI. Optimal bandwidth was chosen using the *rdrobust* package (Calonico *et al.*, 2019b).

Table 5: Covariate Continuity around the Nizam Boundary

Covariate	Conventional	Bias-corrected	Robust	Observations
Veg. Cov. Mean	0.801	0.873	0.873	15953
	(0.876)	(0.876)	(0.958)	
Topo. Rug. Mean	-0.140	-0.0565	-0.0565	15953
	(0.389)	(0.389)	(0.450)	
Avg. Elev.	9.320	6.803	6.803	15953
	(17.81)	(17.81)	(18.30)	
Nearest Town Dist.	5.655***	4.881***	4.881**	15149
	(1.684)	(1.684)	(2.203)	
Total Pop.	470.6**	556.0***	556.0***	15159
	(195.4)	(195.4)	(186.1)	
Vil. Area (Sq m)	1943870.6	3319652.5**	3319652.5**	18945
	(1332892.5)	(1332892.5)	(1331430.3)	
SC Pop. Prop.	-0.0138	-0.0289*	-0.0289	15953
	(0.0169)	(0.0169)	(0.0214)	
ST Pop. Prop.	0.0429**	0.0443**	0.0443***	15953
	(0.0175)	(0.0175)	(0.0172)	
Rural Cons. Mean	-579.4	-426.4	-426.4	15109
	(1155.4)	(1155.4)	(1252.6)	
Household Cult. Income	-0.0700*	-0.0729*	-0.0729*	15159
	(0.0378)	(0.0378)	(0.0386)	
Manuf. Emp. 2013	-0.0348**	-0.0371**	-0.0371*	14499
	(0.0167)	(0.0167)	(0.0220)	
Service Emp. 2013	0.0462**	0.0531***	0.0531**	14499
	(0.0188)	(0.0188)	(0.0256)	
Literate Prop.	-0.00909	0.000239	0.000239	15159
	(0.00931)	(0.00931)	(0.0103)	

Standard Errors are clustered at the subdistrict/taluk level. $*p < 0.1, **p < 0.05, ***$
We use rdrobust (Calonico *et al.*, 2019b) to estimate the Spatial RD at the Nizam boundary with covariates
Controls include latitude, longitude, and the nearest 30km border segment.

Table 6: Covariate Balance around the Nizam Boundary

Variable	Coeff.	Std. Err.	Obs.
Veg. Cover	-0.730	(0.659)	2028
Topo. Rugged.	-0.153	(0.248)	2028
Avg. Elev.	-54.68**	(22.90)	2028
Nearest Town Dist.	2.770*	(1.406)	1888
Total Pop.	-116.5	(168.2)	1888
Village Area	-820632.5	(964928.4)	2227
SC Pop. Prop.	0.0162	(0.0124)	2028
ST Pop. Prop.	0.0680***	(0.0225)	2028
Rural Per Capita Cons.	649.2	(686.9)	1888
Households in Cultivation	-0.0519	(0.0333)	1888
Mfg Sector Employees	-0.0210**	(0.00967)	1868
Service Sector Employees	0.0274***	(0.00998)	1868
Literate Share	-0.0561***	(0.0154)	1888

Standard errors clustered at the subdistrict (taluk) level in parentheses
$*p < 0.1, **p < 0.05, ***p < 0.01$

Table 7: Muslim Segregation around the Nizam Border

	(1) No Cov	(2) +LatLon	(3) +Geographic	(4) +Demo. and Dist.
		Panel A : Isolation Index		
Conventional	0.026**	0.034***	0.026**	0.043***
	(0.012)	(0.013)	(0.011)	(0.013)
Bias-corrected	0.031***	0.037***	0.028**	0.045***
	(0.012)	(0.013)	(0.011)	(0.013)
Robust	0.031**	0.037***	0.028**	0.045***
	(0.013)	(0.014)	(0.013)	(0.014)
Mean-left (eff)	.13	.13	.13	.13
N-left (eff)	1375	1116	1245	906
N-right (eff)	1382	1241	1213	1032
Segment FE	No	Yes	Yes	Yes
		Panel B : Proportion of Muslim Property Owners		
Conventional	0.015***	0.014***	0.011**	0.014***
	(0.005)	(0.005)	(0.005)	(0.005)
Bias-corrected	0.015***	0.014***	0.012**	0.014***
	(0.005)	(0.005)	(0.005)	(0.005)
Robust	0.015***	0.014**	0.012**	0.014**
	(0.006)	(0.006)	(0.006)	(0.006)
Mean-left (eff)	.04	.04	.04	.04
N-left (eff)	1192	1116	1066	924
N-right (eff)	1278	1241	1114	1049
Segment FE	No	Yes	Yes	Yes

Robust Standard Errors are clustered at the subdistrict (taluka) level. We used the rdrobust package (Calonico *et al.*, 2019b) to compute the estimates and chose MSE optimal bandwidths. We use a linear polynomial for estimation. Geographic controls include vegetation cover, elevation, ruggedness and village area. In Distances, we include the distance to the nearest town distance to Bangalore, distance to Hyderabad, distance to Mysore, distance to Mumbai and distance to Chennai, cultivation, manufacturing and services shares in employment, average per capita consumption, literacy rate, total population, proportion of Scheduled Castes (SC) and proportion of Scheduled Tribes (ST). $*p < 0.1, **p < 0.05, ***p < 0.01$

8 Conclusion

We make several novel contributions. First, we introduce a method to digitize map collections at scale to obtain the coordinates of over 16 million landowners in Karnataka, India. The richness of our data allows us to document the religious segregation at the village level and we find it to be significantly high across the state. We then launch an exploration into the consequences of spatial concentration. We use a novel instrumental variable "Presence of a Dargah" to causally estimate its effect on access to public facilities, specifically, schools and health centers. Due to the unprecedented resolution of our data, we can look at the effect of clustering of Muslim households in neighborhoods both within and across villages. We find that neighborhoods with a larger proportion of Muslim households are more likely to contain public goods. However, the benefits of this spatial concentration are driven by those villages that have a lower share of Muslim households overall - residing in an enclave has benefits, especially when the overall share of the minority households is lower in the village. Moreover, the scale of segregation matters. Across villages in the same Grama Panchayat (GP), neighborhoods in villages with a higher share of Muslim households are less likely to have public goods - significantly muting the benefits of within-village local concentration. We see considerable regional heterogeneity in segregation. Given the consequences we have documented, it is important to understand the potential causes of these patterns. We employ a spatial Regression Discontinuity Design on the border separating the region once ruled by the Nizam 75 years ago from British-ruled areas to look at the historical underpinnings of segregation. We find that villages that were once under the Nizam exhibit 30% higher segregation compared to other villages around the boundary.

This work is very early stage and requires further exploration to solidify the mechanisms behind the effects we observe in our data. The fact that we have georeferenced it provides us with an easy way to merge a wide range of geospatial and administrative datasets.

References

Adukia, Anjali, Sam Asher, Kritarth Jha, Novosad, Paul, & Tan, Brandon Joel. 2023. Residential Segregation and Unequal Access to Local Public Services: Evidence from 1.5m Neighborhoods in India.

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

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

Arora, Abhishek, & Dell, Melissa. 2023. *LinkTransformer: A Unified Package for Record Linkage with Transformer Language Models.*

Ash, Elliott, Asher, Sam, Bhowmick, Aditi, Bhupatiraju, Sandeep, Chen, Daniel, Devi, Tanaya, Goessmann, Christoph, Novosad, Paul, & Siddiqi, Bilal. 2021. In-group bias in the Indian judiciary: Evidence from 5 million criminal cases. *Available at SSRN or the specific repository [Replace this text with the actual journal name if the paper was published in a journal]*, 8. [Online; accessed (insert current date)].

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

Boustan, Leah Platt. 2013 (May). *Racial Residential Segregation in American Cities.* NBER Working Papers 19045. National Bureau of Economic Research, Inc.

Burde, Dana, & Linden, Leigh L. 2012 (May). *The Effect of Village-Based Schools: Evidence from a Randomized Controlled Trial in Afghanistan.* Working Paper 18039. National Bureau of Economic Research.

Calonico, Sebastian, Cattaneo, Matias, Farrell, Max H., & Titiunik, Rocio. 2019a. Regression Discontinuity Designs Using Covariates. *The Review of Economics and Statistics*, **101**(3), 442–451.

Calonico, Sebastian, Cattaneo, Matias D., Farrell, Max H., & Titiunik, Rocío. 2019b. Regression Discontinuity Designs Using Covariates. *The Review of Economics and Statistics*, **101**(3), 442–451.

Chaturvedi, Rochana, & Chaturvedi, Sugat. 2020. *It's All in the Name: A Character Based Approach To Infer Religion.*

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

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

Cutler, David M., Glaeser, Edward L., & Vigdor, Jacob L. 2010. *Segregation Data, United States, 1890-2000*. Accessed: 2023-04-04.

Das, Bibhunandini, & Das, Amarendra. 2023. Is Distance to Secondary School a Barrier to Secondary and Higher Education in India? *Millennial Asia*, **14**(1), 102–126.

Dell, Melissa. 2010. THE PERSISTENT EFFECTS OF PERU'S MINING "MITA". *Econometrica*, **78**(6), 1863–1903.

Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton, & Toutanova, Kristina. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR*, **abs/1810.04805**.

Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton, & Toutanova, Kristina. 2019. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.

Geruso, Michael, & Spears, Dean. 2018. Neighborhood Sanitation and Infant Mortality. *American Economic Journal: Applied Economics*, **10**(2), 125–62.

Howard, Andrew, Sandler, Mark, Chu, Grace, Chen, Liang-Chieh, Chen, Bo, Tan, Mingxing, Wang, Weijun, Zhu, Yukun, Pang, Ruoming, Vasudevan, Vijay, Le, Quoc V., & Adam, Hartwig. 2019. *Searching for MobileNetV3*.

Jaderberg, Max, Simonyan, Karen, Zisserman, Andrew, & Kavukcuoglu, Koray. 2016. *Spatial Transformer Networks*.

Jaffrelot, Christophe, & Gayer, Laurent. 2012. *Muslims in Indian Cities: Trajectories of Marginalisation*. Columbia University Press.

Kalra, Aarushi. 2021 (Oct). *A 'Ghetto' of One's Own: Communal Violence, Residential Segregation and Group Education Outcomes in India*.

Keele, Luke J., & Titiunik, Rocío. 2015. Geographic Boundaries as Regression Discontinuities. *Political Analysis*, **23**(1), 127–155.

Massey, Douglas S, & Denton, Nancy A. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press.

Muralidharan, Karthik, & Prakash, Nishith. 2017. Cycling to School: Increasing Secondary School Enrollment for Girls in India. *American Economic Journal: Applied Economics*, **9**(3), 321–50.

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

Okwaraji, Yemisrach Behailu, & Edmond, Karen M. 2012. Proximity to health services and child survival in low- and middle-income countries: a systematic review and meta-analysis. *BMJ Open*, **2**(4).

Santosh Kumar, Emily A. Dansereau, & Murray, Christopher J. L. 2014. Does distance matter for institutional delivery in rural India? *Applied Economics*, **46**(33), 4091–4103.

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

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

Vithayathil, Trina, & Singh, Gayatri. 2012. Spaces of Discrimination: Residential Segregation in Indian Cities. *Economic and Political Weekly*, **47**(37), 60–66.

Wang, Z., Simoncelli, E.P., & Bovik, A.C. 2003. Multiscale structural similarity for image quality assessment. *Pages 1398–1402 Vol.2 of: The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2.

Data Appendix

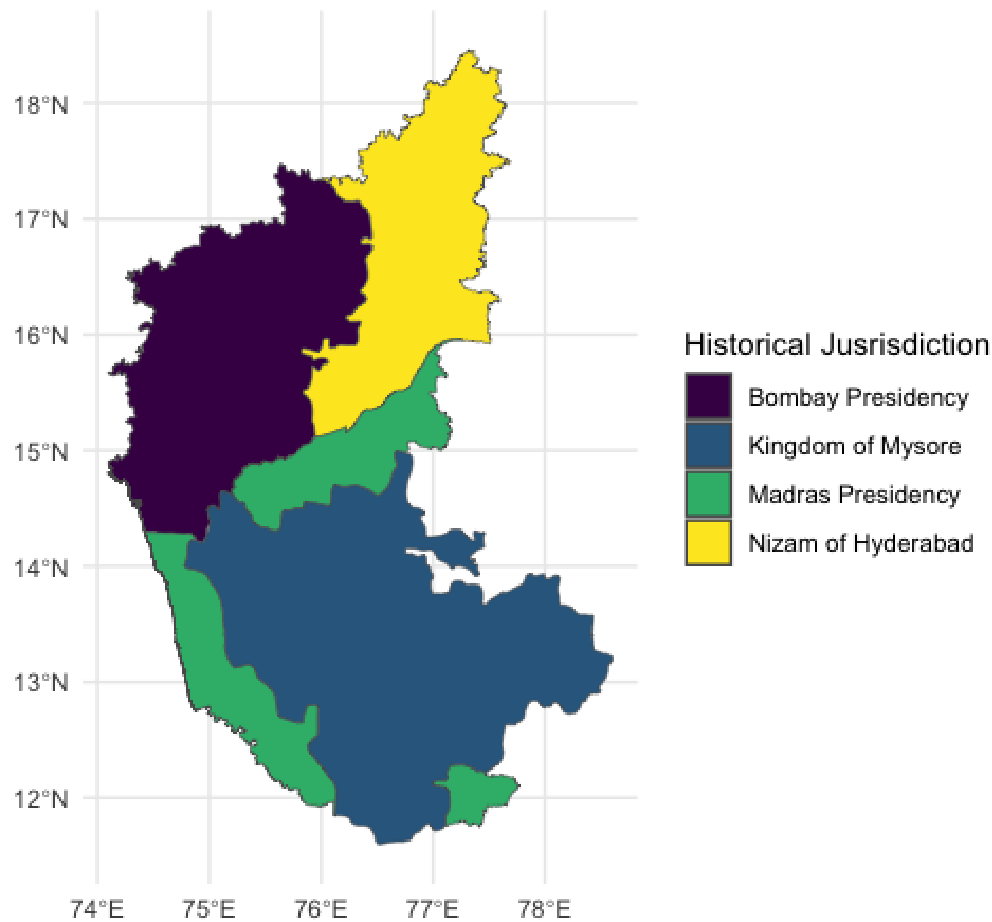


Figure A1: Historical Partitions of present-day Karnataka

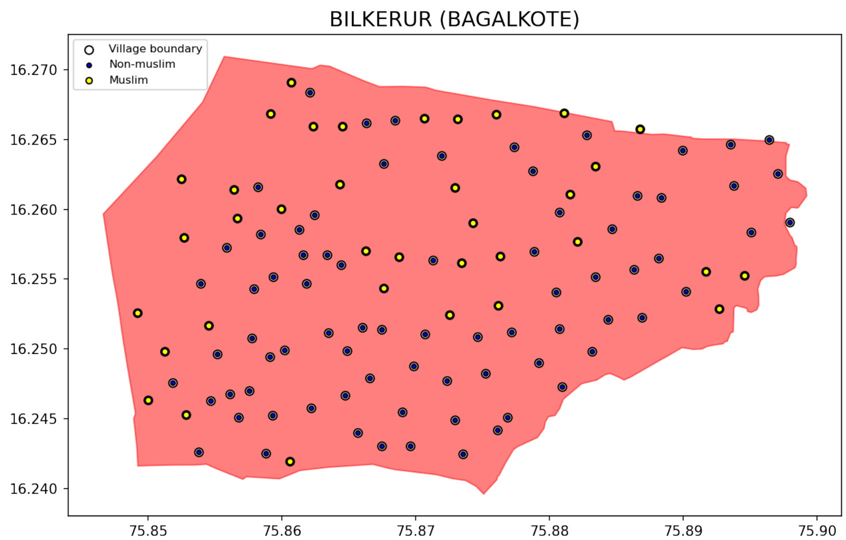


Figure A2: Distribution of Hindu and Muslim owned properties in georeferenced village

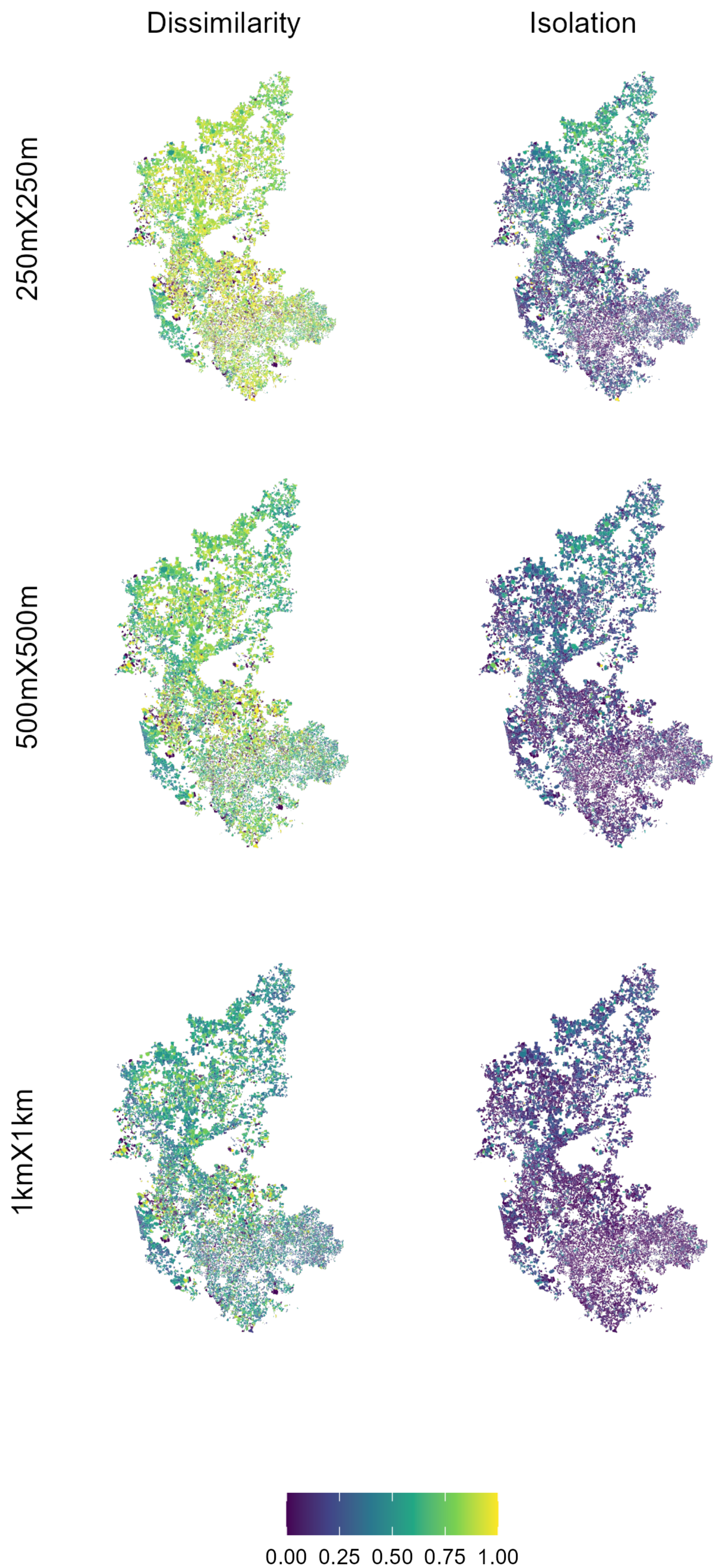


Figure A3: Segregation measures with different neighbourhood sizes

Table A1: Muslim Segregation around the Nizam Border (50% of automatic bandwidth)

	(1) No Cov	(2) +LatLon	(3) +Geographic	(4) +Distances
		Panel A : Isolation Index		
Conventional	0.034**	0.036***	0.031**	0.044***
	(0.013)	(0.013)	(0.013)	(0.013)
Bias-corrected	0.045***	0.044***	0.033***	0.043***
	(0.013)	(0.013)	(0.013)	(0.013)
Robust	0.045**	0.044**	0.033*	0.043**
	(0.018)	(0.018)	(0.018)	(0.018)
Mean-left (eff)	.13	.13	.13	.13
N-left (eff)	1016	1016	950	841
N-right (eff)	1161	1161	1033	988
Segment FE	No	Yes	Yes	Yes
		Panel B : Proportion of Muslim Property Owners		
Conventional	0.013**	0.012*	0.007	0.010
	(0.006)	(0.006)	(0.007)	(0.007)
Bias-corrected	0.016**	0.013**	0.007	0.009
	(0.006)	(0.006)	(0.007)	(0.007)
Robust	0.016*	0.013	0.007	0.009
	(0.008)	(0.008)	(0.009)	(0.009)
Mean-left (eff)	.04	.04	.04	.04
N-left (eff)	613	613	571	501
N-right (eff)	866	866	778	742
Segment FE	No	Yes	Yes	Yes

Robust Standard Errors are clustered at the subdistrict (taluka) level. We used the rdrobust package Calonico *et al.* (2019b) to compute the estimates and chose MSE optimal bandwidths. We use a linear polynomial for estimation. Geographic controls include vegetation cover, elevation, ruggedness and village area. In Distances, we include the distance to the nearest town distance to Bangalore, distance to Hyderabad, distance to Mysore, distance to Mumbai and distance to Chennai, cultivation, manufacturing and services shares in employment, average per capita consumption, literacy rate, total population, proportion of Scheduled Castes (SC) and proportion of Scheduled Tribes (ST). $*p < 0.1, **p < 0.05, ***p < 0.01$

Table A2: Muslim Segregation around the Nizam Border (150% of automatic bandwidth)

	(1) No Cov	(2) +LatLon	(3) +Geographic	(4) +Distances
		Panel A : Isolation Index		
Conventional	0.012	0.022**	0.017*	0.033***
	(0.009)	(0.009)	(0.009)	(0.009)
Bias-corrected	0.025***	0.034***	0.024***	0.043***
	(0.009)	(0.009)	(0.009)	(0.009)
Robust	0.025**	0.034***	0.024**	0.043***
	(0.012)	(0.012)	(0.012)	(0.012)
Mean-left (eff)	.13	.13	.13	.13
N-left (eff)	3697	3697	3282	3056
N-right (eff)	1857	1857	1665	1599
Segment FE	No	Yes	Yes	Yes
		Panel B : Proportion of Muslim Property Owners		
Conventional	0.014***	0.014***	0.010**	0.014***
	(0.004)	(0.004)	(0.004)	(0.004)
Bias-corrected	0.016***	0.014***	0.011**	0.015***
	(0.004)	(0.004)	(0.004)	(0.004)
Robust	0.016***	0.014**	0.011*	0.015**
	(0.006)	(0.006)	(0.006)	(0.006)
Mean-left (eff)	.04	.04	.04	.04
N-left (eff)	1587	1587	1451	1305
N-right (eff)	1455	1455	1298	1240
Segment FE	No	Yes	Yes	Yes

Robust Standard Errors are clustered at the subdistrict (taluka) level. We used the rdrobust package Calonico *et al.* (2019b) to compute the estimates and chose MSE optimal bandwidths. We use a linear polynomial for estimation. Geographic controls include vegetation cover, elevation, ruggedness and village area. In Distances, we include the distance to the nearest town distance to Bangalore, distance to Hyderabad, distance to Mysore, distance to Mumbai and distance to Chennai, cultivation, manufacturing and services shares in employment, average per capita consumption, literacy rate, total population, proportion of Scheduled Castes (SC) and proportion of Scheduled Tribes (ST). $*p < 0.1, **p < 0.05, ***p < 0.01$