Bayesian Spatial Modeling of School Disparities in Brunei Darussalam

Alvin Bong

Abstract

Understanding the spatial distribution of schools is essential for promoting equitable access to education. This study investigates spatial disparities in school availability across Brunei Darussalam, with the aim of identifying potentially underserved areas. An initial exploratory data analysis (EDA) was conducted to examine school counts across mukims. The core analysis applie Bayesian spatial Poisson model using Integrated Nested Laplace Approximation (INLA) to estimate the relative abundance of schools, adjusting for expected counts based on population, size of geographic area, and a socioeconomic indicator (median house price, partially simulated). Both spatially structured and unstructured random effects were included to account for latent spatial processes. Posterior estimates revealed several mukims with significantly lower school availability relative to the national baseline, offering valuable insights to support future policy planning and equitable school placement.

Introduction

Education is a foundational pillar of national development and its people, influencing social well-being, economic growth, and long-term sustainability. The global significance of education is recognized in Sustainable Development Goal 4, which promotes inclusive and equitable quality education for all (Assembly 2015). Nationally, Brunei Darussalam's national vision, Wawasan Brunei 2035, positions education as a cornerstone of the country's long-term development goals (Brunei Darussalam 2020). Ensuring equitable access to education through sufficient infrastructure and fair resource distribution is critical to delivering quality learning experiences.

While several studies have examined general aspects of education in Brunei, there have been limited studies based on quantitative methods (Ebil and Shahrill 2023; Abdul Latif, Matzin, and Escoto-Kemp 2021; Salbrina, Deterding, and Nur Raihan 2024; Mohamad et al. 2018). This project aims to addresses that gap by conducting spatial analysis on schools across the mukims in Brunei, specifically using Bayesian hierarchical models to identify adminstrative regions in Brunei where school availability falls significantly below the national baseline, supporting future policy planning and school placements.

Data

This study focuses exclusively on government primary and secondary schools in Brunei Darussalam, as these institutions serve as the main access points to education for most youth. Bayesian spatial modeling was conducted using Brunei's 39 mukims (administrative region) as the unit of analysis

(areal analysis) to achieve a balance between geographic detail and interpretability. Key data variables that were used include *school counts*, *administrative boundary data*, *population*, and *house prices*. These datasets were mostly sourced from Bruneiverse Github Page, particularly via the `bruneimap` R package, and were subsequently cleaned, wrangled, and merged primarily using `left_join()` and `rbind()` (Jamil 2025; Jamil et al. 2025).

The school dataset from 2018 was used as it is the most recent year for which disaggregated school-level data is available. Although more recent statistics exist, they are published only in summary form. Population data is drawn from the 2021 national census, the most recent census available in Brunei, despite the mismatch in years with the school dataset. Brunei conducts its national census every ten years, making the 2021 data the best option for population estimates.

To incorporate a socioeconomic indicator, we used median house prices derived from approximately 30,000 property listings spanning 1993–2025. These were calculated at the mukim level and included as a covariate in the Bayesian model. In cases where house price data were missing, values were imputed using predictions from an INLA-based Gaussian model. Manual imputations based on local knowledge were initially tested, but the INLA-predicted values were ultimately adopted, as both methods produced similar model outcomes. Given the nature of the data, house prices were treated as partially simulated estimates and may not fully reflect actual market values.

Methods

Choropleth map is used for exploratory data analysis to give an overview of school counts per mukim and overall geography of Brunei.

Spatial regression model

Let Y_i and E_i denote the observed and expected counts of schools, respectively, in mukim $i \in \{1, ..., n\}$. Let θ_i represent the *relative abundance* of schools in mukim i, analogous to a relative risk in disease mapping. The Bayesian hierarchical model (Banerjee, Carlin, and Gelfand 2003), incorporating spatial Besag-York-Mollié (BYM) model (Besag, York, and Mollié 1991) is specified as follow:

$$Y_i \mid \theta_i \sim \text{Poisson}(E_i \cdot \theta_i), \quad i = 1, ..., n$$

$$\log(\theta_i) = \beta_0 + \beta_1 \cdot \text{pop}_i + \beta_2 \cdot \text{area}_i + \beta_3 \cdot \text{hp}_i + u_i + v_i,$$

where:

- β_0 is the intercept,
- β_1, β_2, and β_3 are regression coefficients for the standardized covariates:
  - pop_i: population (in units of 10,000),
  - area_i: mukim size (in units of 10 km²),
  - hp_i: median house price (in BND \$1,000,000),
- u_i is a structured spatial effect, modelled using an intrinsic conditional autoregressive (CAR) prior $u_i \mid u_{-i} \sim \mathcal{N}(\bar{u}_{\delta_i}, \frac{1}{\tau_u n_{\delta_i}})$
- v_i is an unstructured random effect, $v_i \sim \text{Normal}(0, \frac{1}{\tau_v})$

The spatial random effect u_i requires a neighborhood (adjacency) matrix. Here, We define two mukims as neighbors if they share at least one boundary point (Queen contiguity). The neighborhood graph is constructed using the `poly2nb()` function from the `spdep` package.

Model fitting was performed in a Bayesian framework using the Integrated Nested Laplace Approximation (INLA) (Rue, Martino, and Chopin 2009). Model adequacy with respect to spatial structure is assessed by by examining the spatial autocorrelation ofstandardiszed Pearson residual defined as:

$$residual_i = \frac{Y_i - \mu_i}{\sqrt{\mu_i}} = \frac{Y_i - E_i \cdot \theta_i}{\sqrt{E_i \cdot \theta_i}},$$

where Y_i is the observed count, E_i is the expected count (offset), and θ_i is the relative risk. Significant residual spatial autocorrelation may indicate unaccounted spatial structure in the model.

Global Moran's I

To examine whether Pearson residuals of our model exhibit a clustered, dispersed, or random spatial pattern, we apply the Global Moran's I test using the `global_moran_test()` function from the `spdep` package. This test is computed for each mukim in the study area, indexed by $i, j = 1, 2, ..., N$. The Moran's I test statistic is defined as follows:

$$I = \frac{N}{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \in [-1, 1],$$

where:

- x_i is the Pearson residual in mukim i,
- $\bar{x}$ is the mean residual per mukim,
- w_{ij} is the spatial weight between mukims i and j.

For simplicity, the same Queen Contiguity neighbour is used. Moran's I values are standardized, with values close to $+1$ indicating positive spatial autocorrelation (i.e., clustering), where high or low values are near each other. Values close to -1 indicate negative spatial autocorrelation (i.e., dispersion), where neighboring values differ significantly. Values near 0 suggest randomness, indicating an absence of spatial pattern. Figure 1 shows the three configurations of areas.
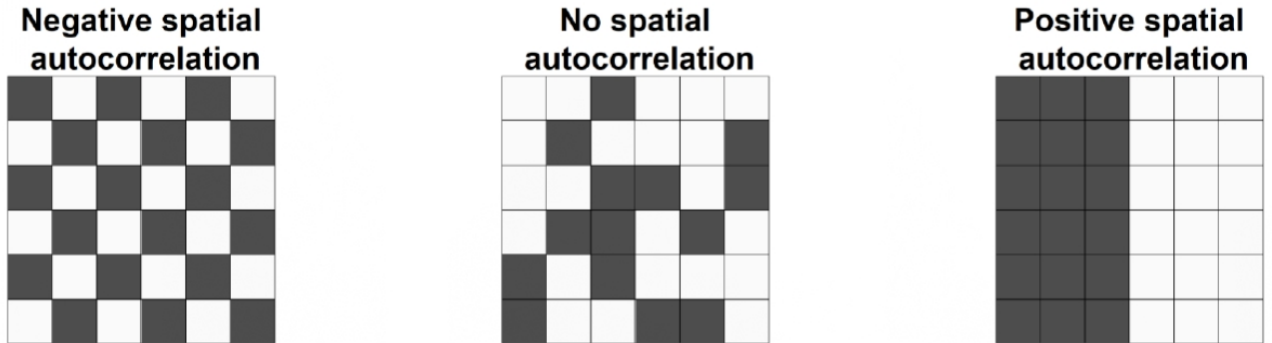

Figure 1: Examples of configurations of areas showing different types of spatial autocorrelation

To determine the significance of the Moran's I statistic, we employ the Central Limit Theorem to calculate p-values based on a Z-score, allowing us to test the following hypotheses:

- $H_0 : I = 0$ (no spatial autocorrelation),
- $H_1 : I \neq 0$ (presence of spatial autocorrelation).

Results

Exploratory Data Analysis

Figure 2 shows that mukims with higher schools count located in the coastal region near South China Sea. As expected, majority of mukims with high school count are in Brunei-Muara District (north-east of mainland), the most populated and urban district in Brunei.
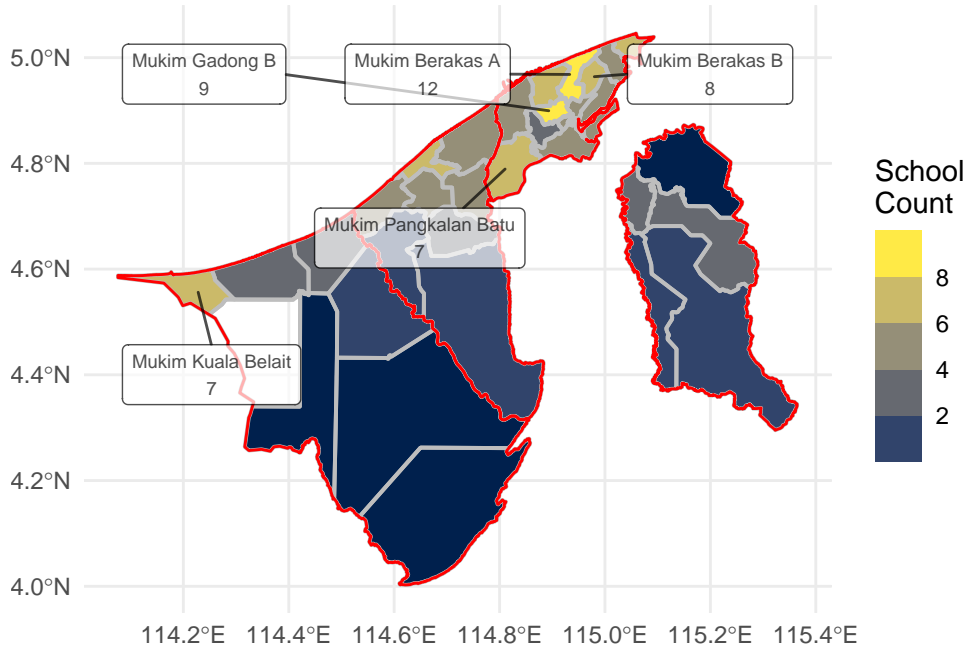

Figure 2: Choropleth Map of school count by mukim. Grey boundaries denote mukim. Red boundaries are the 4 districts in Brunei.

Model

The model results show that the intercept is estimated at $\hat{\beta}_0 = 1.077$, with a 95% credible interval of $(0.239, 1.894)$. The coefficient for population, $\beta_1 = -0.436$, with a 95% credible interval of $(-0.579, -0.295)$, indicates a statistically significant negative relationship between population size and relative school abundance. Specifically, for every 10,000 increase in population, the relative abundance of schools decreases by approximately 35%, since $\exp(-0.436) \approx 0.647$. In contrast, the covariates: mukim size and house price are not statistically significant. The coefficient for mukim size is $\hat{\beta}_2 = 0.015$ (95% CI: -0.002, 0.032), and for house price is $\hat{\beta}_3 = -0.759$ (95% CI: -3.174, 1.704).

The estimated relative abundance (RA) of schools across mukims (Figure 3)vindicates lower values in the northern coastal mukims, particularly in northern Brunei-Muara District, along the South China Sea. Conversely, higher RA values appear inland, especially in less densely populated regions.

To identify mukims with potentially inadequate school provision, non-exceedance probabilities were computed for a threshold of $RA < 0.7$. The analysis (Figure 4) indicates that it is highly likely that several mukims fall below this threshold, including **Mukim Sengkurong, Mukim Gadong A, Mukim Gadong B, Mukim Berakas B, and Mukim Mentiri**.

Lastly, the global Moran's I statistic of 0.213 ($p < 0.05$) provides statistically significant evidence to reject the null hypothesis of no spatial autocorrelation in the residuals.
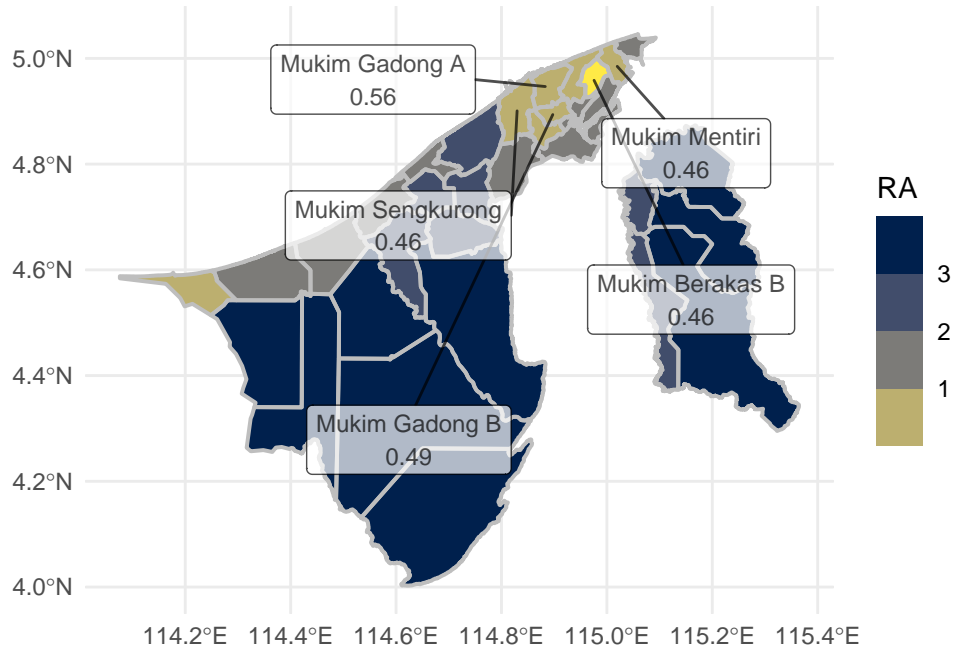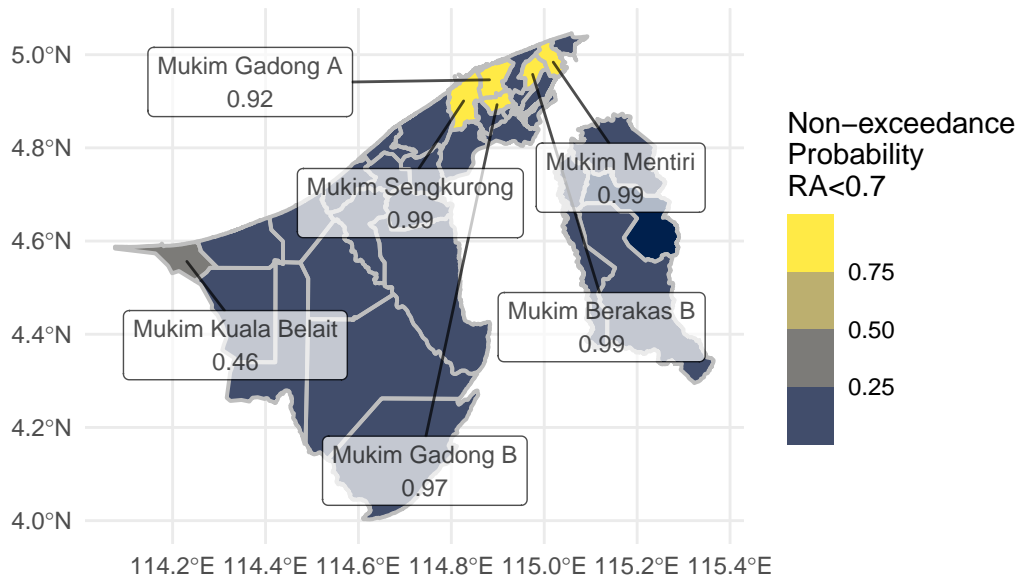

Figure 3: Relative Abudance of Schools in each mukim.

Figure 4: Mukims with non-exceedance probability of Relative Abundance < 0.7
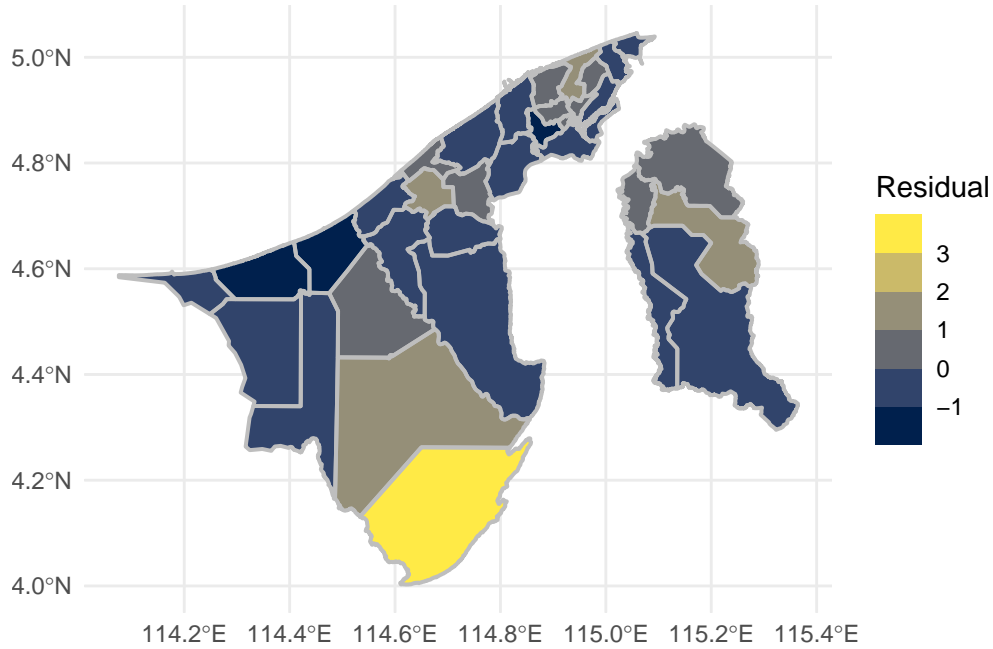
Figure 5: Spatial distribution of Pearson residuals from the Poisson Bayesian model of school counts per mukim

Discussion

The results in Section offer insights into both overall model performance and patterns of educational equity. Global Moran's I statistic (0.213) is not substantially high (<0.3), suggesting that the model captures most, but not all, of the residual spatial dependence. This is illustrated in Figure 5, where the variation in residual colors suggests a reasonably good fit, though further improvement may be possible with additional spatial covariates or refined random effects. The negative relationship between population and school counts, as well as higher RA values in inland (rural) mukims than coastal areas with larger population suggests that school availability in rural mukims seem to be adequate, potentially due to legacy planning policies or intentional efforts to ensure equitable access in remote areas.

In contrast, urban and peri-urban mukims in the Brunei-Muara District, the country's most densely populated and economically active region, show signs of disparities in school access. Notably, the high non-exceedance probability of $RA < 0.7$ in **Mukim Sengkurong, Gadong A & B, Berakas B, and Mentiri** indicates that they have fewer schools than expected relative to their population sizes. These disparities highlight areas that may be underinvested in educational infrastructure, warranting closer policy attention.

Importantly, these mukims are also encompass several new government housing developments, such as Perpindahan Lugu and Perpindahan Tanah Jambu. This suggests a potential planning gap, where population is increasing due to housing developments, but educational infrastructure has not yet caught up.

However, inspection of Figure 2 reveals that these areas have a similar number of schools compared to their neighboring mukims. This suggests that while the absolute number of schools may not be unusually low, the rapid increase in population within these new housing areas may have outpaced school capacity, leading to a situation where demand exceeds supply within these specific zones.

Nevertheless, this situation represents a strategic opportunity. Prioritizing school construction in these fast-growing neighborhoods could significantly improve access to education, reduce commute times, lower transportation costs, and enhance the overall quality of life for residents. Locating schools closer to homes supports national goals related to sustainable urban development, walkability, and equity in public services.

Limitations

This study is limited to public schools, excluding private and international institutions that may affect school availability in some areas. The data is from 2018, potentially missing recent changes, especially in fast-growing neighborhoods. Next, our model used the general population data and is not age-standardized, are important for demand estimation. Housing price was used as a proxy for socioeconomic status, but data relies on listing prices, which may not reflect actual market values due to negotiation factors. In some areas, price data was also simulated, introducing further uncertainty.

Conclusions

In summary, our model reveals a negative relationship between population size and relative school abundance, suggesting urban areas have fewer schools per capita than rural ones. Several mukims in the Brunei-Muara District, including Sengkurong, Gadong A & B, Berakas B, and Mentiri, appear to have fewer schools than expected, despite ongoing population growth driven by new housing developments. While total school counts may seem adequate, local demand in these areas may exceed capacity. Addressing this gap offers a strategic opportunity to improve access, reduce travel time, and support equitable urban planning.

References

Abdul Latif, Siti Norhedayah, Rohani Matzin, and Aurelia Escoto-Kemp. 2021. "The Development and Growth of Inclusive Education in Brunei Darussalam." *Globalisation, Education, and Reform in Brunei Darussalam*, 151–75. https://doi.org/10.1007/978-3-030-77119-5_8.

Assembly, United Nations General. 2015. "Transforming Our World: The 2030 Agenda for Sustainable Development." https://sustainabledevelopment.un.org/post2015/transformingourworld.

Banerjee, Sudipto, Bradley P Carlin, and Alan E Gelfand. 2003. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman; Hall/CRC.

Besag, Julian, Jeremy York, and Annie Mollié. 1991. "Bayesian Image Restoration, with Two Applications in Spatial Statistics." *Annals of the Institute of Statistical Mathematics* 43 (1): 1–20.

Brunei Darussalam, Government of. 2020. "Brunei Darussalam Long-Term Development Plan. Wawasan Brunei 2035."

Ebil, Syazana, and Masitah Shahrill. 2023. "Overview of Education in Brunei Darussalam." In *International Handbook on Education in South East Asia*, 1–21. Springer. https://doi.org/10.1007/978-981-16-8136-3_46-1.

Jamil, Haziq. 2025. *Bruneimap: Maps and Spatial Data of Brunei (R Package Version 0.3.1.9001)*. https://bruneiverse.github.io/bruneimap/.

Jamil, Haziq, Amira Barizah Noorosmawie, Hafeezul Waezz Rabu, and Lutfi Abdul Razak. 2025. "From Archives to AI: Residential Property Data Across Three Decades in Brunei Darussalam." *Data in Brief*, March. https://doi.org/10.1016/j.dib.2025.111505.

Mohamad, Hanapi, Rosyati M Yaakub, Emma Claire Pearson, and Jennifer Tan Poh Sim. 2018. "Towards Wawasan Brunei 2035: Early Childhood Education and Development in Brunei Darussalam." *International Handbook of Early Childhood Education*, 551–67. https://doi.org/10.1007/978-94-024-0927-7_25.

Rue, Håvard, Sara Martino, and Nicolas Chopin. 2009. "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 71 (2): 319–92.

Salbrina, Sharbawi, David Deterding, and Mohamad Nur Raihan. 2024. "Education in Brunei." In *Brunei English: A New Variety in a Multilingual Society*, 17–32. Springer. https://doi.org/10.1007/978-3-031-60303-7_2.