

# Implementing a BYM Model in Stan to Fit Boston Housing Price Data

(Proposal)

Marcel Gietzmann-Sanders

STAT641 - Bayesian Statistics  
University of Alaska Fairbanks

## 1 In Short

My Master's research is centered around understanding spatial movement of Chinook salmon in the Gulf of Alaska. As such I thought it would be interesting to look into a spatial Bayesian model for my project. Specifically I would like to follow in the footsteps of a paper that implemented the Besag York Mollie (BYM) model in Stan (Mitzi Morris (2019)) and then apply that to a housing dataset about Boston, MA where I live (Moraga (2023)). My project report would be organized according to the following outline:

1. Explain the BYM model
2. Showcase an implementation in Stan
3. Introduce the dataset and the question - housing price as a function of crime rates and number of bedrooms
4. Fit a BYM model and report on the results

## 2 Brief on the Besag York Mollie Model

The BYM model deals with modeling the distribution of a specific outcome across a series of distinct spatial units/areas given some set of covariates  $\mathbf{x}_i$  where  $i$  indexes the spatial unit in question.

If, as in the Boston housing prices example, we assume that our observations per unit  $Y_i$  are given by:

$$Y_i \sim N(\mu_i, \sigma^2)$$

Then the BYM model is:

$$\mu_i = \eta + \mathbf{x}_i\boldsymbol{\beta} + \phi_i + \theta_i$$

where  $\eta$  is our overall intercept,  $\boldsymbol{\beta}$  represents our regression coefficients,  $\phi_i$  is a spatial random variable and  $\theta_i$  is just a normal non-spatial random variable.

The priors for  $\eta$ ,  $\boldsymbol{\beta}$ , and  $\boldsymbol{\theta}$  are all assumed to be normal. Part of my project will be determining what I would like to use as a prior on  $\sigma^2$ . The spatial term  $\phi_i$ , however, gets a tad more complicated.

Each spatial interaction term  $\phi_i$  is modeled as conditional on the other terms as:

$$\phi_i|\phi_j \sim N\left(\sum_j w_{ij}\phi_j, \sigma^2\right), i \neq j$$

A key result that Besag proved (Julian (1974)) is that the joint distribution  $\boldsymbol{\phi}$  ends up being multivariate normal random variable centered at 0

$$\boldsymbol{\phi} \sim N(0, Q^{-1})$$

where  $Q = D(I - \alpha A)$  where  $D$  is a diagonal "neighbors" matrix (each element on the diagonal is the number of neighbors unit  $i$  has),  $A$  is an adjacency matrix where if  $i, j$  are neighbors then the  $i, j$  element is 1, and  $\alpha$  lets us control spatial dependence. In our case, as we will be following the stan implementation from the paper (Mitzi Morris (2019)) we will be setting  $\alpha = 1$ .

### 3 Brief on the Data

This data comes from the R package spData (Bivand (2022)) and contains information on housing in 506 Boston census tracts. What we'll be interested in for our study (which will more or less follow (Moraga (2023))) is the median housing price (Fig. 1), crime rate (Fig. 2), and average number of rooms per dwelling (Fig. 3) per tract.

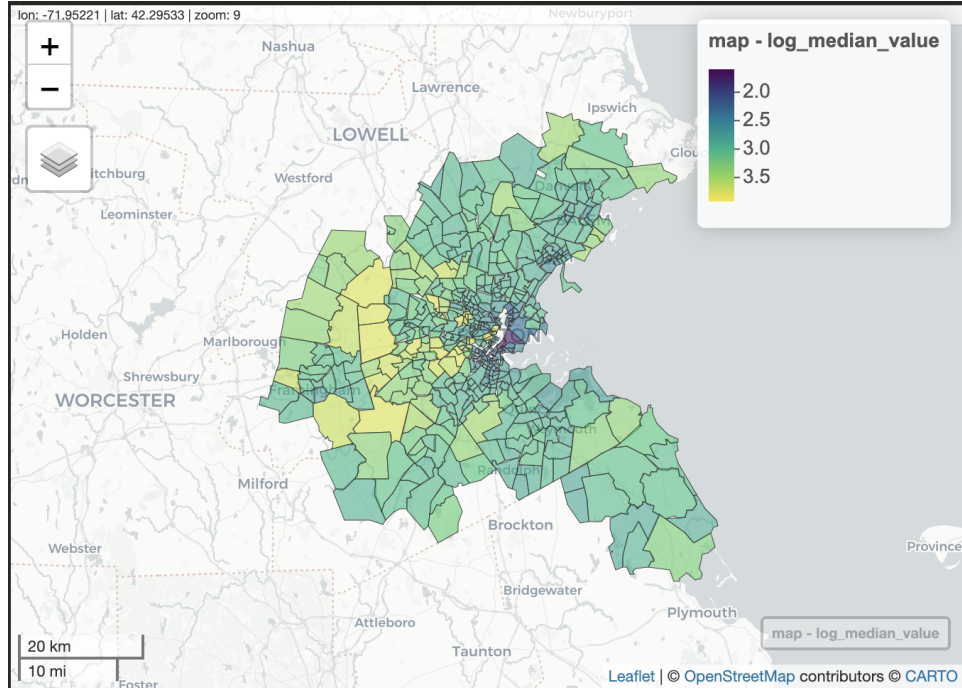


Figure 1: Log Median Housing Price

The logarithm of the median house value (in \$1000USD) by census tract in the Greater Boston area.

What we will be interested in understanding is the degree to which crime rates and number of rooms per dwelling affect housing prices in the Greater Boston area. I.e., we are specifically interested in the  $\beta$  component of our BYM model.

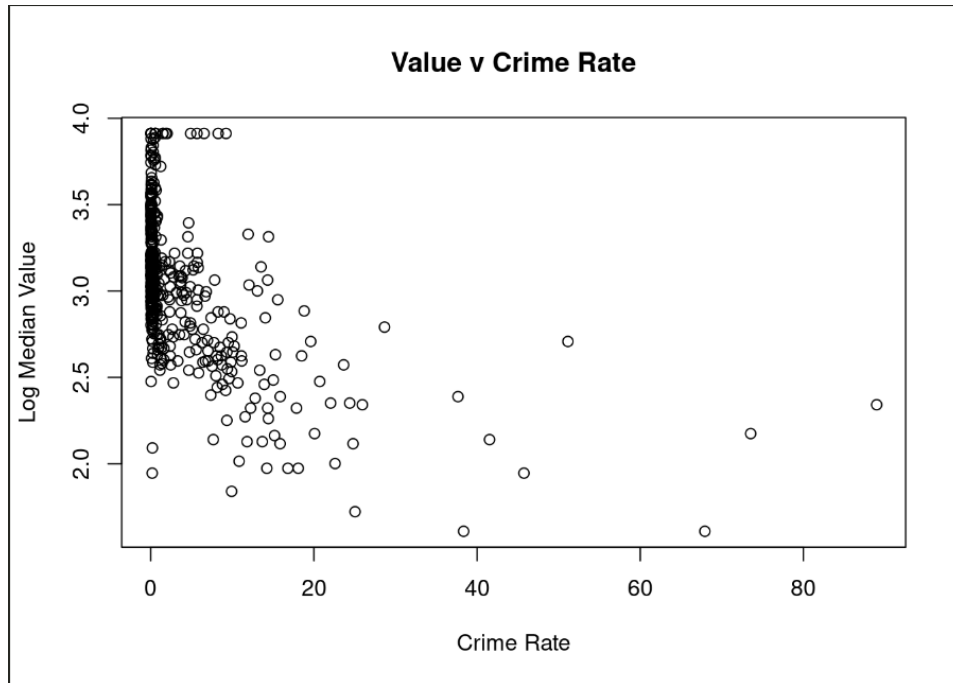


Figure 2: Value vs Crime Rate

The logarithm of the median house value (in \$1000USD) by census tract vs per capita crime rate.

## References

- Bivand, Roger, J. N.-R. L. (2022). *spdata: Datasets for spatial analysis*.
- Julian, B. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*.
- Mitzi Morris, Katherine Wheeler-Martin, D. S. S. J. M. A. G. C. D. (2019). Bayesian hierarchical spatial models: Implementing the besag york mollié model in stan. *Spat Spatiotemporal Epidemiol*.
- Moraga, P. (2023). *Spatial Statistics for Data Science: Theory and Practice with R*. American Fisheries Society.

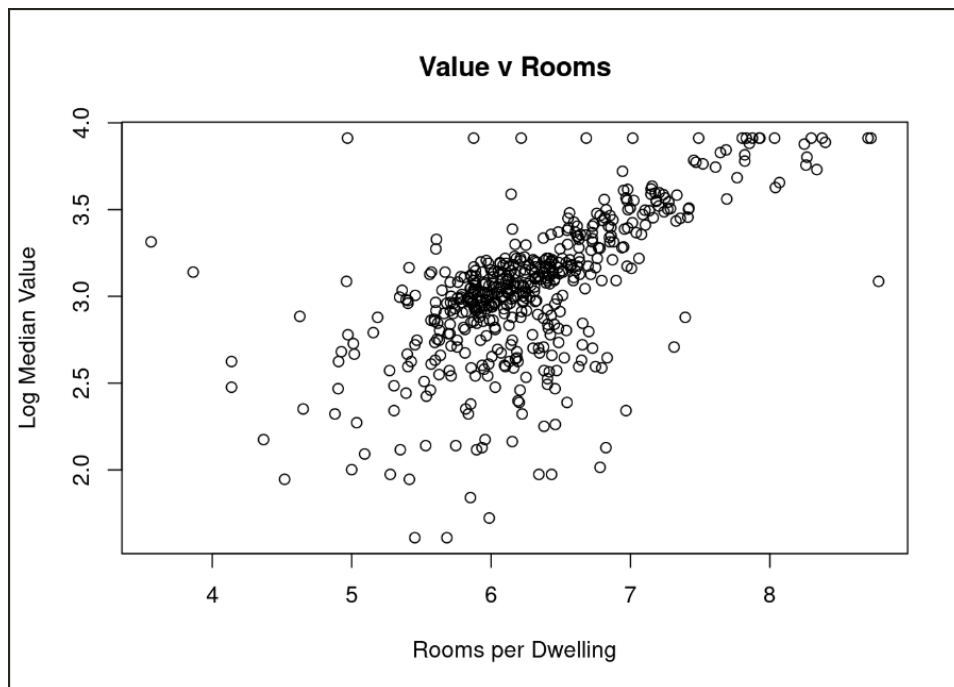


Figure 3: Value vs Rooms

The logarithm of the median house value (in \$1000USD) by census tract vs average rooms per dwelling.