Modelling U.S. Home Prices Using Bayesian Hierarchical Regression

Lorenzo Galli, Gabriele Socrate

Introduction

The dataset is about median housing price for neighbourhoods in North America in 2012. The data was downloaded from Kaggle and was collected and published by researchers on Zenodo: each row represents a neighbourhood, which are in total 20428, while the seven variables are the following:

- **housing median age** (in years): the median age of the housing in the neighbourhood;
- **total rooms**: total number of rooms in the neighbourhood;
- **total bedrooms**: total number of bedrooms in the neighbourhood;
- **population**: the population of the neighbourhood;
- **households**: the number of households in the neighbourhood;
- **median income** (in thousands of dollars): the median income of the neighbourhood;
- **ocean proximity** (categorical) with levels: Near Bay, <1 Hour from Ocean, Inland, Near Ocean;

The quantitative response is: Median house value (in dollars) of the neighbourhood.

Goal of the study

This analysis has two main objectives: first, to examine whether the median house value in different neighborhoods varies with respect to their distance from the ocean; and second, to investigate how the influence of the variables on the response changes depending on this distance.

Exploratory Data Analysis

Before fitting the model, we should have a general understanding of the variables through summary statistics and graphical representations.

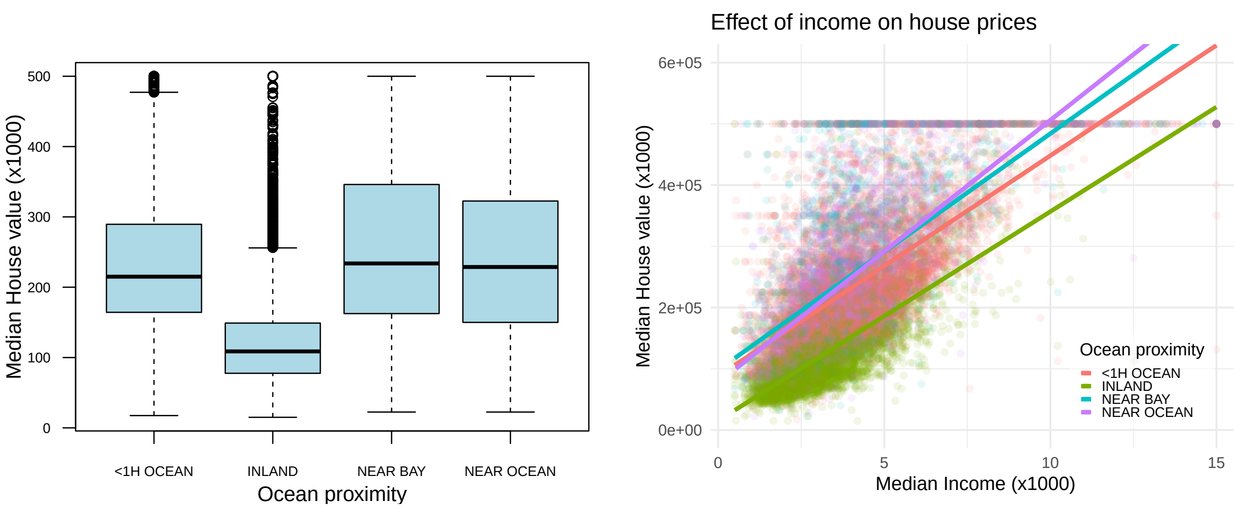
```
## NA: 0   Duplicate rows: 0
```

```
##  housing_median_age  total_rooms    total_bedrooms     population
##  Min.   : 1.00      Min.   :    2   Min.   :   1.0   Min.   :    3
##  1st Qu.:18.00      1st Qu.: 1450   1st Qu.: 296.0   1st Qu.:  788
##  Median :29.00      Median : 2127   Median : 435.0   Median : 1166
##  Mean   :28.63      Mean   : 2637   Mean   : 537.9   Mean   : 1425
##  3rd Qu.:37.00      3rd Qu.: 3143   3rd Qu.: 647.0   3rd Qu.: 1723
##  Max.   :52.00      Max.   :39320   Max.   :6445.0   Max.   :35682
```

```
##    households      median_income     median_house_value ocean_proximity
##  Min.   :    1.0   Min.   : 0.4999   Min.   : 14999     Length:20428
##  1st Qu.: 280.0    1st Qu.: 2.5634   1st Qu.:119475     Class :character
##  Median : 409.0    Median : 3.5375   Median :179700     Mode  :character
##  Mean   : 499.5    Mean   : 3.8714   Mean   :206822
##  3rd Qu.: 604.0    3rd Qu.: 4.7441   3rd Qu.:264700
##  Max.   :6082.0    Max.   :15.0001   Max.   :500001
```

To see if the median house value changes across the distance of the ocean, we can see from a boxplot; for simplicity, we will consider the Median House price in thousands of dollars.

```
## `geom_smooth()` using formula = 'y ~ x'
```



Before fitting any models, the boxplot reveals a clear difference in house prices across the groups. In particular, the "Inland" group shows significantly lower prices—approximately half those of the other groups—while the "<1H OCEAN" group displays less variability in price compared to the "Near Bay" and "Near Ocean" categories. Additionally, the relationship between *median income* and house value appears consistent across all groups, as indicated by the similar slopes of the fitted trend lines. This pattern holds for the other predictor variables as well: their effect on the response does not appear to vary substantially between groups.

Hierarchical Model

Since we have seen from the graphs the difference in the median house prices between different ocean distance, we want to fit a hierarchical model to better adress the diversity between groups: the four groups (in our case, the ocean distance's levels) are the first level, while the units within groups (neighbourhoods) are the second level. We assume that the observations are independent and identically distributed given a group specific parameter and a common variance:

$$(y_{1j}, \dots, y_{nj}) \mid \theta_j, \sigma^2 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta_j, \sigma^2)$$

We are interested in the group-specific means $(\theta_1, \dots, \theta_m)$, the within-group variability σ^2 that is assumed common for each group, and the mean and variance of the $\underline{\theta} : (\mu, \tau^2)$. We assume that:

$$(\theta_1, \dots, \theta_m) \mid \mu, \tau^2 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \tau^2)$$

while the independent priors are:

$$\frac{1}{\sigma^2} \sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2\right)$$

$$\frac{1}{\tau^2} \sim \text{Gamma}\left(\frac{\eta_0}{2}, \frac{\eta_0}{2}\tau_0^2\right)$$

$$\mu \sim \mathcal{N}\left(\mu_0, \gamma_0^2\right)$$

As prior hyperparameters, we have chosen values: $\mu_0 = 200$ is the median price of USA houses in 2012; γ_0^2, τ_0^2 and σ_0^2 are all fixed as large values in order to have weakly informative priors.

```
mu.0 = 200; gamma.0_2 = 2500
nu.0 = 2; sigma.0_2 = 1000
eta.0 = 2; tau.0_2 = 1000
```

Since we are in the semi-conjugate case, the joint posterior distribution is:

$$p(\mu, \tau^2, \sigma^2, \theta_1, \ldots, \theta_m \mid \underline{y}_1, \ldots, \underline{y}_m) \propto \prod_{j=1}^{m} \left\{ \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j, \sigma^2) \right\} \cdot \prod_{j=1}^{m} p(\theta_j \mid \mu, \tau^2) \cdot p(\mu) \cdot p(\tau^2) \cdot p(\sigma^2)$$

that can be approximated by using Gibbs Sampling from the full conditional distributions of μ, τ^2, θ_J and σ_2.
The full conditional of μ is obtained through:

$$p(\mu \mid \text{rest}) \propto p(\mu) \times \prod_{j=1}^{m} p(\theta_j \mid \mu, \tau^2)$$

$$\mu \mid \text{rest} \sim \mathcal{N}\left(\frac{\frac{\mu_0}{\gamma_0^2} + \frac{m\bar{\theta}}{\tau^2}}{\frac{1}{\gamma_0^2} + \frac{m}{\tau^2}}, \quad \left(\frac{1}{\gamma_0^2} + \frac{m}{\tau^2}\right)^{-1} \right) \qquad \text{where } \bar{\theta} = \frac{1}{m}\sum_{j=1}^{m} \theta_j$$

μ_n is a weighted average between the prior mean μ_0 and the sample mean $\bar{\theta}$, where the weights are the prior precision $1/\tau_0^2$ and the sample precision m/τ^2; γ_n^2 is the sum of the sample variance and prior variance.

The full conditional of τ^2 is obtained through:

$$p(\tau^2 \mid \text{rest}) \propto p(\tau^2) \times \prod_{j=1}^{m} p(\theta_j \mid \mu, \tau^2)$$

$$\frac{1}{\tau^2} \mid \text{rest} \sim \text{Gamma}\left(\frac{\eta_0 + m}{2}, \quad \frac{\eta_0 \tau_0^2 + \sum_{j=i}^{m} (\theta_j - \mu)^2}{2} \right)$$

Since τ^2 is the variance between groups, then m can be considered as the sample size of τ^2. So η_n is the sum between prior sample size η_0 and actual sample size m; τ_n^2, that coincides with the posterior mean of τ^2, is the sum of prior mean of τ^2 and the sum of the square deviation of each θ_j from their mean μ, which represents a measure of the variability of each θ_j.

The full conditional of θ_J is obtained through:

$$p(\theta_j \mid \text{rest}) \propto p(\theta_j \mid \mu, \tau^2) \times \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j, \sigma^2)$$

$$\theta_j \mid \text{rest} \sim \mathcal{N}\left(\frac{\frac{\mu}{\tau^2}+\frac{n_j\bar{y}_j}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n_j}{\sigma^2}},\quad \left(\frac{1}{\tau^2}+\frac{n_j}{\sigma^2}\right)^{-1}\right) \qquad \text{where } \bar{y}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij}$$

The posterior expectation of θ_j is the weighted average between the prior mean μ of $\underline{\theta}$ and the sample mean $\bar{y_j}$, where the weights are the prior precision $1/\tau^2$ and sample precision n_j/σ^2. The posterior variance of θ_j is the sum of the prior variance and sample variance.

The full conditional of σ^2 is obtained through:

$$p(\sigma^2 \mid \text{rest}) \propto p(\sigma^2) \times \prod_{j=i}^{m}\prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j, \sigma^2)$$

$$\frac{1}{\sigma^2} \mid \text{rest} \sim \text{Gamma}\left(\frac{\nu_0+n}{2},\quad \frac{\nu_0\sigma_0^2 + \sum_{j=1}^{m}\sum_{i=1}^{n_j}(y_{ij}-\theta_j)^2}{2}\right) \qquad \text{where } n = \sum_{j=1}^{m} n_j$$

ν_n is the sum between prior sample size ν_0 and actual sample size n, while σ_n^2 is the posterior expectation of σ^2, that is the sum of the prior mean σ_0^2 and the sum of squared deviation of each observation y_{ij} from the mean of their respective group θ_j.

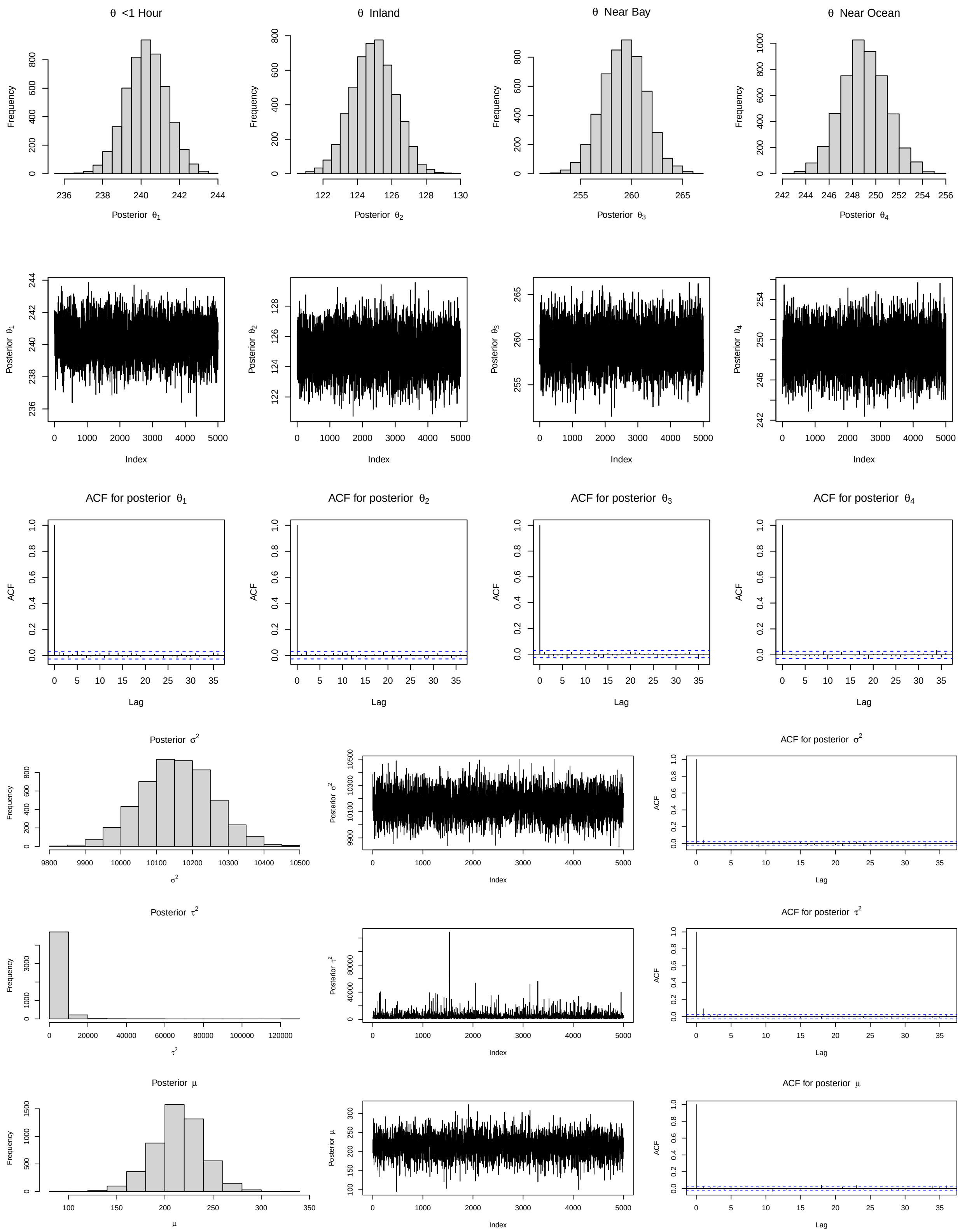
To approximate the posterior distribution, we can consider the following Gibbs Sampling scheme:

$$\begin{aligned}
&\text{Initialize } \theta_1^{(0)}, \dots, \theta_m^{(0)},\ (\tau^2)^{(0)},\ (\sigma^2)^{(0)} \\
&\textbf{for } s = 1, \dots, 5000: \\
&\quad \text{Update } (\mu,\ \tau^2): \\
&\qquad \text{Sample } \mu^{(s)} \sim p\left(\mu \mid \theta_1^{(s-1)}, \dots, \theta_m^{(s-1)}, (\tau^2)^{(s-1)}\right) \\
&\qquad \text{Sample } (\tau^2)^{(s)} \sim p\left(\tau^2 \mid \theta_1^{(s-1)}, \dots, \theta_m^{(s-1)}, \mu^{(s)}\right) \\
&\quad \text{Update } (\theta_1, \dots, \theta_m): \\
&\qquad \textbf{for } j = 1 \text{ to } 4: \\
&\qquad\quad \text{Sample } \theta_j^{(s)} \sim p\left(\theta_j \mid \underline{y}_j,\ \mu^{(s)},\ (\tau^2)^{(s)},\ (\sigma^2)^{(s-1)}\right) \\
&\quad \text{Update } \sigma^2: \\
&\qquad \text{Sample } (\sigma^2)^{(s)} \sim p\left(\sigma^2 \mid \underline{y},\ \theta_1^{(s)}, \dots, \theta_m^{(s)}\right)
\end{aligned}$$

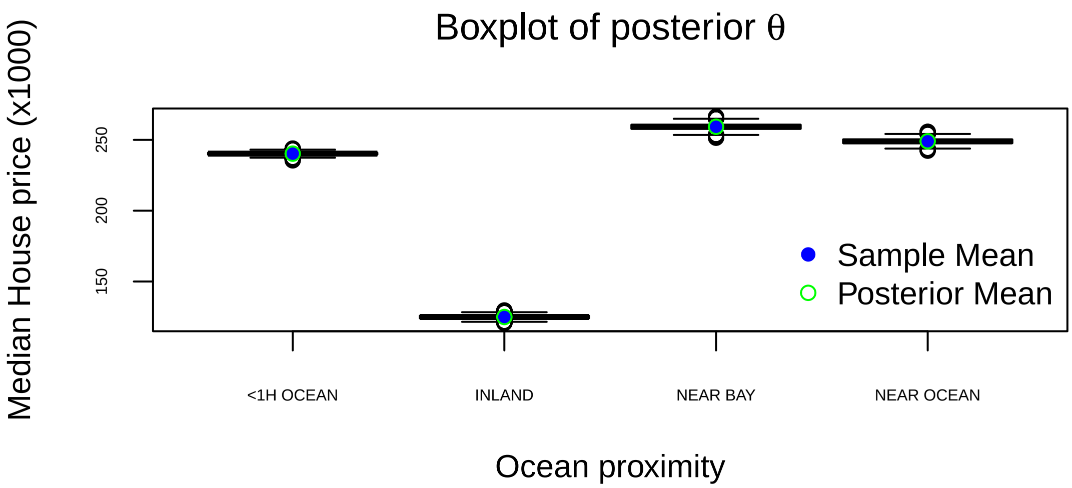
We initialize $\underline{\theta}$, σ^2 and τ^2 in the following way:

```
x = as.numeric(factor(Housing_data$ocean_proximity))
y = Housing_data$median_house_value/1000

theta = c()
  vars = c()

  for (j in unique(x)){
    theta[j] = mean(y[x == j])
    vars[j] = var(y[x == j])
  }

  tau_2 = var(theta)
  sigma_2 = mean(vars)
```

where each θ is the mean of the response in the j-th group, τ^2 is the variance between groups (variance of θ) and σ^2 is the common variance across groups.



Comparing the histograms of the parameters and the respective traceplots, the Markov chains are centered around the posterior mode. Moreover, we can see that there is no seasonality nor trends in the chains, meaning that there is very low dependence. To better adress the independence, the autocorrelation plots show that each parameter value doesn't depend on the past values: a rule-of-thumb is that we should we should thin the chain if the acf is greater than 0.2 at different lags, but this is not our case, so no thinning nor burn-in is needed.



The graph above shows the posterior θ boxplots: we can see that the sample mean (frequentist approach) coincide with the posterior mean of each θ, due to the fact that our prior is not informative and, as percevied before doing any posterior analysis, there is a difference between the median price of the houses depending on the ocean proximity.

Geweke test

The idea of the Geweke test is to take two "windows" of the chain and compare the means, that should be equal if the chained has converged to the posterior distribution. The Geweke statistic Z_n is distributed as a N(0,1) as the number of iteration goes to infinity:

$$Z_n = \frac{\bar{\theta}_I - \bar{\theta}_L}{\sqrt{\hat{s}_I^2 + \hat{s}_L^2}} \rightarrow \mathcal{N}(0,1) \quad \text{as} \quad n \rightarrow \infty$$

where $\bar{\theta_I}$ and $\bar{\theta_L}$ are the sample means of the first 30% iterations and last 30% iterations in the chain, $n = n_I + n_L$ and $\hat{s_I}^2$ and $\hat{s_L}^2$ are the sample variances of θ_I and θ_L.
The hypothesis are:

$$\begin{cases} H_0 : \bar{\theta}_I = \bar{\theta}_L \\ H_1 : \bar{\theta}_I \neq \bar{\theta}_L \end{cases}$$

```
##          theta[1] theta[2] theta[3] theta[4] sigma^2 tau^2    mu
## Geweke_Z    1.102    0.745    1.231   -1.142   0.497 0.682 0.611
```

If the absolute value of the statistic is less than 1.96 ($z_{1-a/2}$ with $\alpha = 0.05$), we cannot reject H_0: this is the case in our analysis.

Hierarchical Linear Regression Model

Once that we established that there is a different between the median house prices with respect to the groups, we would like to study how the covariates influence the response. As stated in the exploratory data analysis, the effect of the variables into the response do not change across groups, so in the linear regression model the β_s will be common, while the intercept will be group-specific.
The hierarchical linear regression model is:

$$Y_{ij} \mid \beta_{0j},\ \boldsymbol{\beta},\ \sigma^2 = \beta_{0j} + \boldsymbol{\beta}^\top x_{ij} + \varepsilon_{ij}, \quad \text{where } \varepsilon_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

The model assumes that the observations are independent given the group-specific intercept $\beta_0 j$, the coefficients $\underline{\beta}$ and the common variance σ^2, and they are distributed as a Normal:

$$y_{ij} \mid \beta_{0j},\ \boldsymbol{\beta},\ \sigma^2 \overset{\text{ind}}{\sim} \mathcal{N}\left(\beta_{0j} + \boldsymbol{\beta}^\top x_{ij},\ \sigma^2\right)$$

As a consequence, the likelihood is the following:

$$p(\underline{y} \mid \text{rest}) = \prod_{j=1}^{4} \prod_{i=1}^{n_j} p\left(y_{ij} \mid \beta_{0j},\ \boldsymbol{\beta},\ \sigma^2\right)$$

For this model, we will consider the following priors:

$$\boldsymbol{\beta} \sim \mathcal{N}_p\left(\boldsymbol{\beta}_0,\ V^{-1}\right)$$

$$\frac{1}{\sigma^2} \sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2\right)$$

$$\beta_{0j} \sim \mathcal{N}(\mu, \tau^2)$$

$$\mu \sim \mathcal{N}(\mu_0, \gamma_0^2)$$

$$\frac{1}{\tau^2} \sim \text{Gamma}\left(\frac{\eta_0}{2}, \frac{\eta_0}{2}\tau_0^2\right)$$

Since the posterior inference is highly sensitive from the choice of a fixed prior hyperparameter for the precision matrix $\text{V}^{-}1$ of the $\boldsymbol{\beta}$, we decided to consider an additive prior on V: by doing so, we allow the data to have more influence on the posterior distribution of the precision matrix. The prior is:

$$V \sim \text{Wishart}(a, U), \quad \text{where } U \text{ is a } p \times p \text{ matrix}$$

As prior hyperparameters, we have chosen the following value: for μ_0, γ_0^2, η_0, τ_0^2, ν_0, and σ_0^2 we have taken the same values of the hierarchical model; for $(beta_0)$ we have chosen a null vector of length 6 (number of predictors), so a priori we don't know if the effect is positive or negative; for U we have chosen a diagonal matrix (6x6) in order to have a weakly informative prior on V, and a = 6 because it must holds that a > p - 1.

```r
X = as.matrix(Housing_data[, -c(7, 8)]) # do not consider the response and the group variable
p = ncol(X)
Beta0  = rep(0, p)
U = diag(0.1, p)
a = p
```

Since we are in the semi-conjugate case, the joint posterior distribution is

$$p\left(\mu, \tau^2, \sigma^2, \beta_{01}, \ldots, \beta_{0m}, \underline{\beta}, V \mid \underline{y}_1, \ldots, \underline{y}_m\right) \propto \prod_{j=1}^{m} \left\{\prod_{i=1}^{n_j} p\left(y_{ij} \mid \beta_{0j}, \underline{\beta}, \sigma^2\right)\right\} \cdot \prod_{j=1}^{m} p\left(\beta_{0j} \mid \mu, \tau^2\right) \cdot p(\mu) \cdot p(\tau^2) \cdot p(\sigma^2) \cdot p(\underline{\beta} \mid V) \cdot p(V)$$

that can be approximated by using Gibbs Sampling from the full conditional distributions. The full conditional distribution of $\boldsymbol{\beta}$ is obtained through:

$$p(\boldsymbol{\beta} \mid \text{rest}) \propto \prod_{j=1}^{m} \prod_{i=1}^{n_j} p(y_{ij} \mid \beta_{0j}, \boldsymbol{\beta}, \sigma^2) \times p(\boldsymbol{\beta})$$

$$\boldsymbol{\beta} \mid \text{rest} \sim \mathcal{N}_p\left(A^{-1}\underline{\mathbf{b}},\ A^{-1}\right)$$

$$\text{where} \quad A = \left(\frac{\sum_{j=1}^{m} \sum_{i=1}^{n_j} x_{ij}^2}{\sigma^2}\right) + V, \quad \underline{\mathbf{b}} = \left(\frac{\sum_{j=1}^{m} \sum_{i=1}^{n_j} x_{ij}(y_{ij} - \beta_{0j})}{\sigma^2}\right) + V\boldsymbol{\beta}_0$$

The A matrix is the sum between the prior precision matrix V and the sum of the squared predictors' values x_{ij}^2 over σ^2, that is the expected Fisher information (that is also the sample precision); so the posterior variance of $\underline{\beta}$ is the sum between the sample covariance matrix and the prior covariance matrix. The vector $\underline{b}$ is the sum of the numerator of OLS estimates for $\underline{\beta}$ (OLS: $\hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^{m} \sum_{i=1}^{n_j} \mathbf{x}_{ij}\mathbf{x}_{ij}^\top\right)^{-1} \left(\sum_{j=1}^{m} \sum_{i=1}^{n_j} \mathbf{x}_{ij}(y_{ij} - \beta_{0j})\right)$) over σ^2 and of the prior precision matrix $V * \beta_0$ prior mean.

The full conditional distribution of $\tilde{\sigma}^2 = 1/\sigma^2$ is obtained through:

$$p(\tilde{\sigma}^2 \mid \text{rest}) \propto \prod_{j=1}^{m} \prod_{i=1}^{n_j} p(y_{ij} \mid \beta_{0j}, \boldsymbol{\beta}, \sigma^2) \times p(\tilde{\sigma}^2)$$

$$\tilde{\sigma}^2 \mid \text{rest} \sim \text{Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n}{2}\sigma_n^2\right) \qquad \text{where } \nu_n = \nu_0 + n, \quad \sigma_n^2 = \frac{1}{\nu_n}\left[\nu_0\sigma_0^2 + \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left((y_{ij} - \beta_{0j}) - \underline{\boldsymbol{\beta}}^\top \mathbf{x}_{ij}\right)^2\right]$$

ν_n is the sum between prior sample size ν_0 and actual sample size n, while σ_n^2 is the posterior mean of σ^2, that is the sum of prior mean σ_0^2 and the residual sum of squares RSS $\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left((y_{ij} - \beta_{0j}) - \underline{\boldsymbol{\beta}}^\top \mathbf{x}_{ij}\right)^2$.

The full conditional distribution of $\beta_0 j$ is obtained through:

$$p(\beta_{0j} \mid \text{rest}) \propto \prod_{i=1}^{n_j} p(y_{ij} \mid \beta_{0j}, \underline{\beta}, \sigma^2) \cdot p(\beta_{0j})$$

$$\beta_{0j} \mid \text{rest} \sim \mathcal{N}\left(\frac{b}{a}, \frac{1}{a}\right) \qquad \text{where } a = \frac{n_j}{\sigma^2} + \frac{1}{\tau^2} \ \text{ and } \ b = \sum_{i=1}^{n_j} \frac{y_{ij} - \underline{\beta}^\top x_{ij}}{\sigma^2} + \frac{\mu}{\tau^2}$$

b is the sum between the sample mean's numerator of $\beta_0 j$ ($\bar{\beta}_{0j} = \frac{1}{n_j}\sum_{i=1}^{n_j}\left(y_{ij} - \boldsymbol{\beta}^\top \mathbf{x}_{ij}\right)$) over σ^2 and the prior mean μ over prior variance τ^2. The posterior variance is the sum between prior variance τ^2 and the sample variance σ^2/n_j.

The full conditional distribution of μ is obtained through:

$$p(\mu \mid \text{rest}) \propto \prod_{j=1}^{m} p(\beta_{0j} \mid \mu, \tau^2) \cdot p(\mu)$$

$$\mu \mid \text{rest} \sim \mathcal{N}\left(\frac{b}{a}, \frac{1}{a}\right) \qquad \text{where } a = \frac{m}{\tau^2} + \frac{1}{\gamma_0^2} \;\text{ and }\; b = \frac{\mu_0}{\gamma_0^2} + \sum_{j=1}^{m} \frac{\beta_{0j}}{\tau^2}$$

The posterior variance of μ is the sum between prior variance γ_0^2 and the sample variance τ^2/m, while b is the sum between the prior mean μ_0 over prior variance γ_0^2 and the sample mean's numerator ($\sum_{j=1}^{m} \beta_{0j}$) over τ^2.

The full conditional distribution of $\tilde{\tau}^2 = 1/\tau^2$ is obtained through:

$$p(\tilde{\tau}^2 \mid \text{rest}) \propto \prod_{j=1}^{m} p(\beta_{0j} \mid \mu, \tau^2) \cdot p(\tilde{\tau}^2)$$

$$\tilde{\tau}^2 \mid \text{rest} \sim \text{Gamma}\left(\frac{\eta_n}{2}, \frac{\eta_n}{2}\tau_n^2\right) \qquad \text{where } \eta_n = \eta_0 + m \;\text{ and }\; \tau_n^2 = \frac{1}{\eta_n}\left(\sum_{j=1}^{m}(\beta_{0j} - \mu)^2 + \eta_0\tau_0^2\right)$$

η_n is the sum between prior sample size η_0 and actual sample size m, while τ_n^2 is the posterior mean of τ^2, that is the sum of the prior mean τ_0^2 and the sum of the square deviation of each $\beta_0 j$ from their mean μ, which represents a measure of the variability of the $\beta_0 js$.

The full conditional distribution of V is obtained through:

$$p(V \mid \text{rest}) \propto p(\underline{\beta} \mid V) \cdot p(V)$$

$$V \mid \text{rest} \sim \text{Wishart}(a_n, U_n) \qquad \text{where} \quad a_n = a + 1 \;\text{ and }\; U_n = U + (\underline{\beta} - \underline{\beta}_0)(\underline{\beta} - \underline{\beta}_0)^\top$$

The matrix U_n is the sum between the prior hyperparameter U and the outer product of the difference between the coefficients and their prior means $\underline{\beta}_0$.

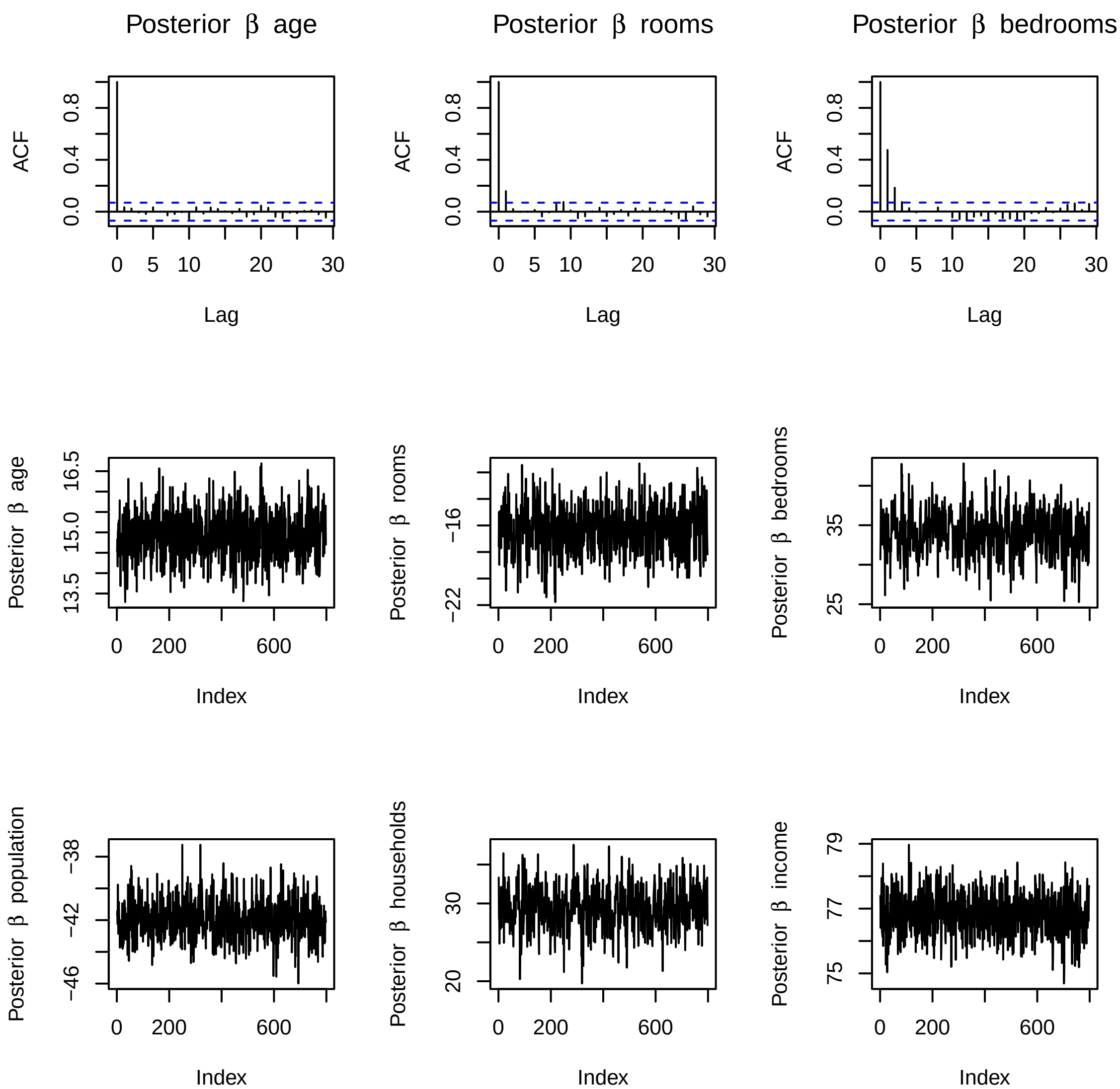
To approximate the full posterior distribution, we can consider the following Gibbs Sampling scheme:

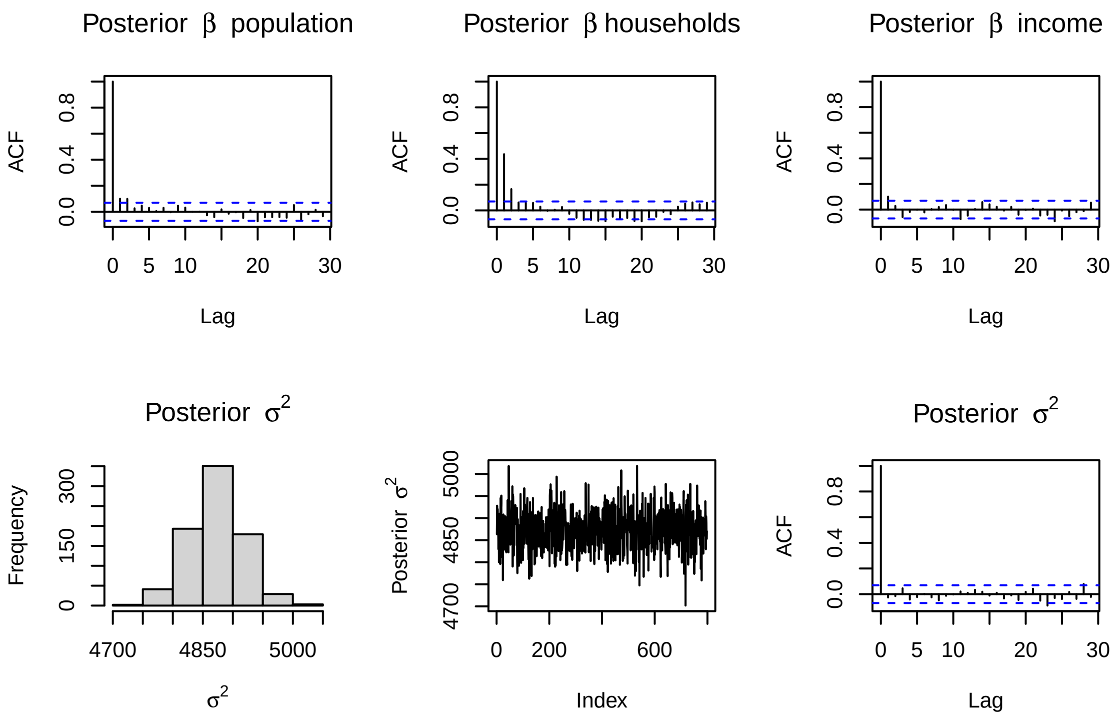
$$\begin{aligned}
&\text{Initialize } \underline{\beta}_0^{(0)},\ \tau^{2(0)},\ \underline{\beta}^{(0)},\ \sigma^{2(0)},\ V^{(0)}\\
&\text{For } s = 1, \ldots, 5000:\\
&\quad \text{Update } (\mu, \tau^2):\\
&\qquad \text{Sample } \mu^{(s)} \sim p\left(\mu \mid \underline{\beta}_0^{(s-1)},\ \tau^{2(s-1)}\right)\\
&\qquad \text{Sample } \tau^{2(s)} \sim p\left(\tau^2 \mid \underline{\beta}_0^{(s-1)},\ \mu^{(s)}\right)\\
&\quad \text{Update } \underline{\beta}_0:\\
&\qquad \text{For } j = 1, \ldots, 4:\\
&\qquad\quad \text{Sample } \beta_{0j}^{(s)} \sim p\left(\beta_{0j} \mid \mu^{(s)},\ \tau^{2(s)},\ \sigma^{2(s-1)},\ \underline{y}_j,\ \underline{\beta}^{(s-1)}\right)\\
&\quad \text{Update } \underline{\beta}:\\
&\qquad \text{Sample } \underline{\beta}^{(s)} \sim p\left(\underline{\beta} \mid V^{(s-1)},\ \sigma^{2(s-1)},\ \underline{\beta}_0^{(s)},\ \underline{y}\right)\\
&\quad \text{Update } V:\\
&\qquad \text{Sample } V^{(s)} \sim p\left(V \mid \underline{\beta}^{(s)}\right)\\
&\quad \text{Update } \sigma^2:\\
&\qquad \text{Sample } \sigma^{2(s)} \sim p\left(\sigma^2 \mid \underline{\beta}^{(s)},\ \underline{\beta}_0^{(s)},\ \underline{y}\right)
\end{aligned}$$

We initialized $\underline{\beta_0}^{(0)}$, $\underline{\beta}^{(0)}$, $\left(\tau^2\right)^{(0)}$, $\left(\sigma^2\right)^{(0)}$, $\mathrm{V}^{(0)}$ in the following way:

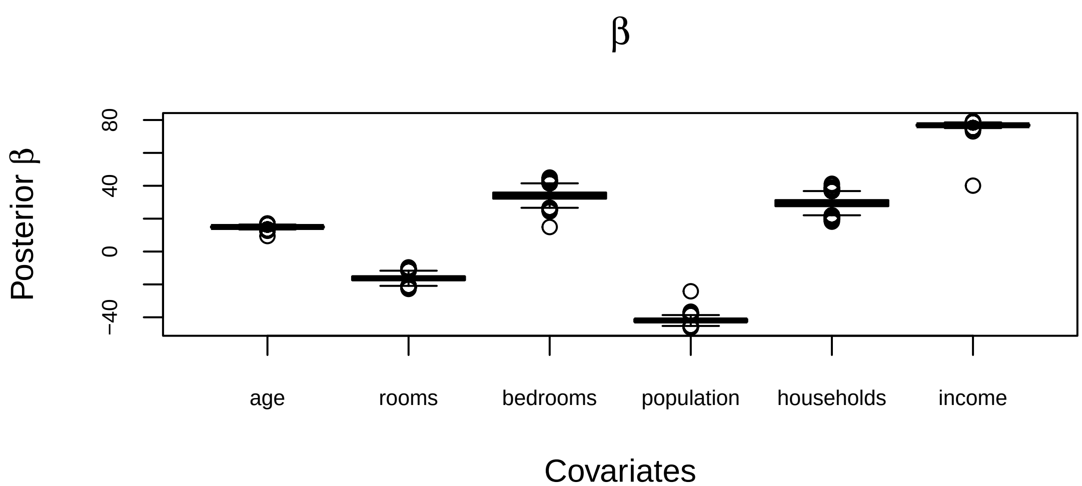
```
Beta_0_j = colMeans(output$theta.post)
tau_2 = var(Beta_0_j)
Beta   = c(solve(t(X)%*%X)%*%t(X)%*%y)
sigma2 = c(t(y - X%*%Beta)%*%(y - X%*%Beta)/(n - p - 1))
V0 = (1/sigma2)*t(X)%*%X
```

$\underline{\beta}^{(0)}$, $\left(\sigma^2\right)^{(0)}$, $\mathrm{V}^{(0)}$ are the frequentist estimates, while the $\underline{\beta_0}^{(0)}$ is the posterior expectation of the $\underline{\theta}$ obtained from the previous Gibbs Sampling of the hierarchical model, and $\left(\tau^2\right)^{(0)}$ is the variance of the $\beta_0 j$.





After thinning the chains every 25 iterations and burning the first 20 iterations, we can see from the ACF graphs and traceplots that the autocorrelation decreased noticeably: there is still some dependence in the initial lags, but as they increase the dependence becomes negligible. Since we have done thinning and burn-in for the β, to maintain consistency across iterations between every parameters, we have to do thinning and burn-in also for all the other parameters. By looking at the graphs related to σ^2, there is no dependence in the chain nor seasonality and trends. The graphs for μ, $\beta_0 j$ (θ_j in the hierarchical model) and τ^2 are very similar to those obtained in the hierarchical model, and even in this case they do not show dependency in the chains.



The graph above shows the boxplots for the posterior coefficients: they are all significantly different from zero, since none of the interquantiles contains the zero value. Moreover, as the median age of the houses in

the neighbourhood is higher, the median house price increases. The same holds for the number of bedrooms in the neighbourhood, the number of households and the median income of the neighbourhoods, since they all have a positive effect on the response.
On the other hand, the higher is the number of rooms and bigger is the population in the neighbourhood, the median house price decreases, since they have a negative effect on the response.

Geweke Test

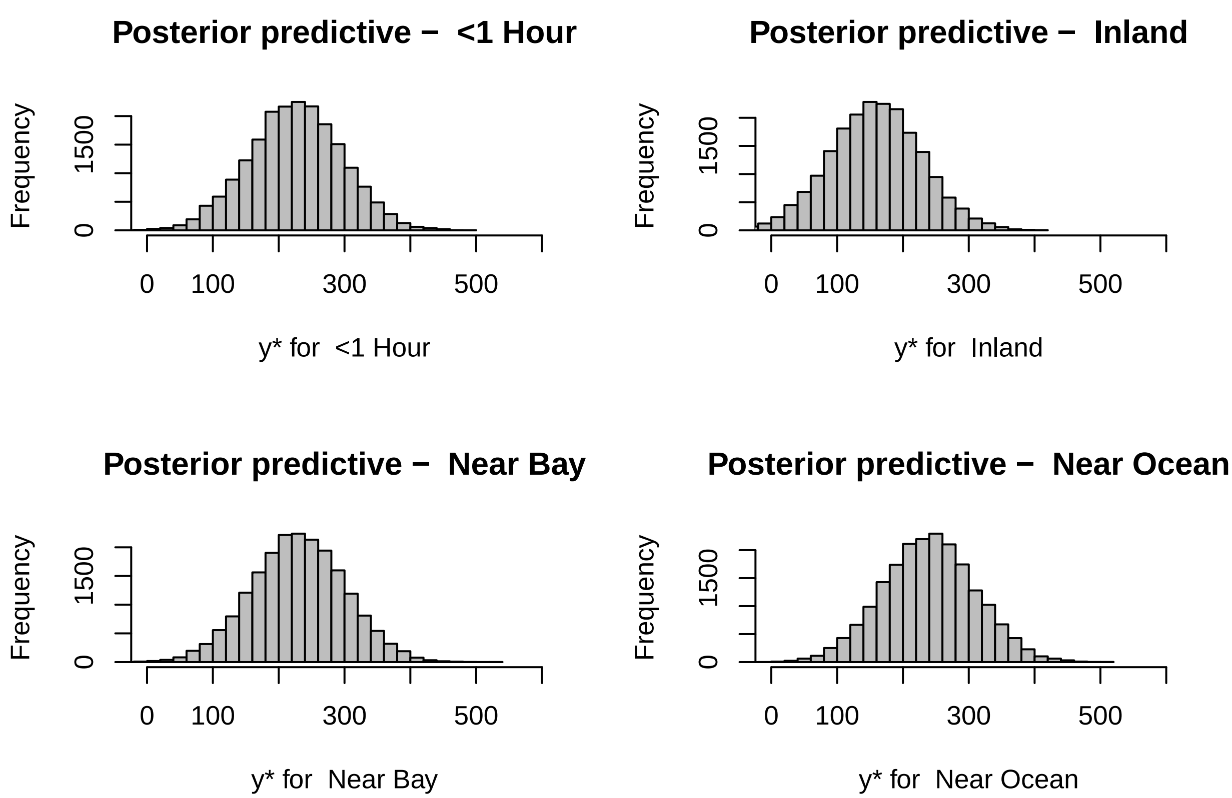
As we did for the hierarchical model, we perform the Geweke test on the posterior $\underline{\beta}$ and posterior σ^2.

```
##           beta[1] beta[2] beta[3] beta[4] beta[5] beta[6] sigma^2     mu  tau^2
## Geweke_Z  -0.774  -1.226   0.449   1.093  -0.293   1.287   0.622 -1.521 -1.503
```

If the absolute value of the statistic is less than 1.96 ($z_{1-a/2}$ with $\alpha = 0.05$), we cannot reject H_0: this is the case in our analysis.

Posterior predictive distributions

Let us consider a hypothetical new neighborhood, not included in the original dataset, whose predictor values correspond to the average covariate values observed in the neighborhoods already analyzed. Our objective is to estimate the median house price in this new neighborhood, depending on its location relative to the ocean So we have computed the posterior predictive distribution for each group




```
## [1] 225.6266 158.2940 229.8380 239.5534
```

The values above are the means of the posterior predictive distributions of each group: if the neighbourhood is "<1 Hour" far from ocean, the predicted median house price will approximately be 226000 dollars; if the neighbourhood is "Inland" the predicted median house price will be approximately 158000 dollars; if the neighbourhood is in "Near Bay", the predicted median house price will be approximately 229500 dollars; if the neighbourhood is "Near ocean", the predicted median house price will be approximately 239600 dollars.

Conclusion

Thanks to the hierarchical model, we can conclude that the data show a difference in median house prices across neighbourhoods based on the ocean proximity, with a particularly significant difference for inland neighbourhoods. After performing hierarchical linear regression, where we assumed that the covariates' effect on the response does not depends on the distance from the ocean, we found out that the housing median age, total number of bedrooms, number of households and median income within a neighbourhood have a positive effect on the response. Specifically, as these factors increase, the median house price tends to rise. On the other hand, the population and total number of rooms in the neighbourhood have a negative effect on the response, with increases in these variables associated with decreases in median house prices. The data was collected in 2012, so further research could be to check if data collected in recent years confirm our results.