Please describe all your work in clear terms, before implementing R code. Each question should include a description of your approach with clear indication of where I can download the associated source code. Your code should be attached to your assignment or uploaded online to a repository I can freely access.

Bayesian Model Selection.
Consider the data-set `BostonHousing`, available in R loading the `mlbench` library. The outcome of interest is `medv`, the median value of howner occupied homes. Several other covariates are measured by census tract. Let X be the matrix of covariate information, including the intercept term, and Y be the median home value.

Consider the following regression model

$$Y = X\beta + \epsilon$$

with $X : n \times p$, $\beta : p \times 1$ and $\epsilon \sim N(0, \sigma^2 I_n)$.

We are interested in the selection of the most likely model, including only a subset of the original p predictors. The number of possible models is 2^p.

Let $\gamma \in \{0, 1\}^p$ define a p-dimensional vector of predictor indicators, such that $\gamma_j = 1$ if X_j is included in the regression, otherwise $\gamma_j = 0$, $(j = 1, \ldots, p)$. Also, let X_γ and β_γ denote the subset of predictors and regression coefficients associated with components of γ that are equal to 1.

Conditional on γ the sampling model is

$$Y = X_\gamma \beta_\gamma + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n),$$

with improper prior:

$$p(\beta_\gamma, \sigma^2 \mid \gamma) \propto (\sigma^2)^{-\frac{q_\gamma}{2}} \exp\left[-\frac{g}{2\sigma^2}\beta'_\gamma (X'_\gamma X_\gamma)\beta_\gamma\right]$$

where $q_\gamma = \sum_{j=1}^{p} \gamma_j$ is the model size, given a realization of γ and $g = 0.001$. The prior on the model space is defined as: $p(\gamma) \propto 2^{-q_\gamma}$.

a. Describe and implement a Reversible Jumps MCMC algorithm exploring the model space $p(\gamma \mid Y)$. Report the estimated inclusion probabilities $p(\gamma_j = 1 \mid Y)$.

b. Compare your results in (a) with results obtained sampling directly from $p(\gamma \mid Y)$. You should be able to implement this without the need for RJ-MCMC.

c. For the MCMC implementations in (a) and (b), produce plots for the posterior distribution of the regression coefficients, when the median model is selected (select covariates with marginal inclusion probability above 0.5), and when we average over all explored models.

d. Split the data into a training $D_0 = (Y_0, X_0)$ set and a test set $D_1 = (Y_1, X_1)$ (including approximately one third of the sample). Train your model on D_0 and produce a Monte Carlo sample from the predictive distribution $p(Y_1 \mid X_1)$ associated with covariate information in the test set X_1. Describe how $p(Y_1 \mid X_1)$ when related to the test data Y_1, may be used to provide formal probabilistic understanding of predictive performance in terms of accuracy and uncertainty.