PROJECT

1. Consider the Parana dataset from the geoR library (data(parana), help(parana)) and **Identify the region of interest, design, response variable and covariates, if any**. Plot what the map would look like

2. Consider elevation *data(elevation*) as a simple linear regression problem, with elevation as the response and north-south coordinate as the explanatory variable. Fit the model and examine the residuals of the model.

   Do you consider that a more sophisticated model is necessary for the analysis of spatial variation?

   Do you consider that a more sophisticated model is necessary for the analysis of spatial variation, justify your answer.

3. Consider the following models for a data set, whose response Yi = 1, 2, ..., n associated with a sequence of xi positions along a spatial axis of one x-dimension.

   a) $Y_i = \alpha + \beta x_i + \epsilon_i$ where α and β are parameters and the ϵ_i are mutually independent with mean zero and variance σ^2_ϵ.

   b) $Y_i = A + Bx_i + \epsilon_i$ where the ϵ_i are as in the previous subsection, but A and B are random variables independent of each other and of the ϵ_i, each with zero mean and variances σ^2_α y σ^2_β, respectively

- For each of these models, find the expected value and variances of Y_i, the covariance between Y_i and Y_i, for $i \neq j$. Plot if possible

4. Look for and reproduce figure 3.2 from the book "Applied Spatial Data Analyses with R".

5. Consider the following

$$Y \sim Bern(0.5), X|Y = 0 \sim \mathcal{N}((0,0), \mathbb{A}) \text{ y } X|Y = 1 \sim \mathcal{N}((1,2), \mathbb{A})$$

$$\mathbb{A} = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$$

a) Calculate the optimal Bayesian classifier for a symmetric cost function.

b) Generate samples (X, Y) of size n = 50, 100, 500.

c) Find and discuss suitable decision trees for these data and compare the best models obtained with cross-validation with the optimal Bayesian classifier, **both visually and quantitatively**.