# Lab 1: Spatial exploration analysis

April 19, 2021


## Summary

In this lab session, you will learn the following:

- how to explore and visualise geostatistical data;
- how to explore the association between outcomes and covariates;
- how to construct a variogram;
- understand how the spatial correlation can (or cannot be reflected) in the empirical variogram

The session is divided into three parts; 1) preliminary exploration of the dataset in R; 2) using the app for spatial exploratory analysis; and 3) variogram analysis.


## 1 Preliminary exploration of the lead concentration dataset in R

The purpose of this section to get familiar with the variables in the data.

1. Open the R studio
2. Load the dataset named `GaliciaData.csv` into R

3. Check the columns of the data and ensure that you understand what each column means. Note: see the description of the dataset in the following link `https://mbgworkshop2021.netlify.app/materials.html`

4. Plot the histogram of the lead concentration, log of the lead concentration and the $PM_{10}$. You can also look at other summary such as the mean, median and variance.

# 2 Using the app for spatial exploratory analysis

In this section, you will use the app to perform further exploratory analysis of the `GaliciaData.csv` dataset.

1. Open the MBG app in R. Type the following into the R console: `shiny::runGitHub(repo="MBGapp", username= "olatunjijohnson", ref="main", subdir = "inst/MBGapp")`

2. Load the `GaliciaData.csv` into the app.

3. Display on a map the log of the lead concentration. What do you think about the sampling design?

4. Plot the scatter plot of the log of the lead concentration and $PM_{10}$. Is there any substantial association between the variables?

# 3 Variogram analysis

In this section, you will learn how to check for the evidence of spatial correlation using the variogram.

1. Using `GaliciaData.csv` dataset, construct a variogram for residual $\hat{Z}$ from

$$Y_i = \beta_0 + \beta_1 d(x_i) + Z_i,$$

where $Y_i$ is the log-lead concentration and $d(x)$ is the $PM_{10}$. Is there evidence of spatial correlation.

2. Remove the covariate and observe how the variogram changes. Is there any difference in the variogram and what does this imply?

3. Perform a confirmatory test for the existence of spatial correlation using the Monte Carlo strategy. Can you confirm an existence of spatial correlation? Is there any effect of the number of permutation on your conclusion?

# 4 Exercise

Using the dataset named `loaloaCameroon.csv`, Fit an appropriate model and examine the residual for the evidence of spatial correlation. This dataset is from Loaloa prevalence survey collected in Cameroon and some part of Nigeria. Hint: Repeat the task in section 1 to 3.

# 5 Additional exercise

In this exercise, you will search for a geostatistical data yourself and perform a spatial exploratory analysis on it. Hint: To download data on neglected tropical diseases, go to the EPSEN portal, `https://espen.afro.who.int/tools-resources/download-data`; select the disease of interest (e.g Loiasis); select "site level" data; choose a country of interest (e.g Nigeria); choose the range of the years (e.g start year 2010, end year 2015); and the download it.