Homework 2 (50 marks)

For this homework, use the **Columbus, OH crime dataset** introduced by Luc Anselin. This is one of the celebrated pioneering datasets, that justified the use of spatial regression models to the literature.

Submission: html output and R code

To load the dataset (and gal file) directly run:

# data(columbus, package="spData")

To load the shapefile (useful for maps) run:

columbus = st\_read(system.file("shapes/columbus.shp", package="spData"))

To load the gal file:

col.gal.nb = read.gal(system.file("weights/columbus.gal", package="spData"))

1. OLS1: Fit the variable CRIME on HOVAL, INC, OPEN, PLUMB, and DISCBD. Report the summary. (2 marks)
2. OLS2: (nested): Fit the variable CRIME only on HOVAL and INC. Report the summary. (2 marks)
3. Determine which model is a better fit for CRIME. Explain. (3 marks)
4. At 5% level of significance, test whether the nested model (OLS2) suffers from non-normality of errors and heteroskedasticity of errors. (4 marks)
5. Plot the CRIME, HOVAL and INC values on three different maps of Columbus. What spatial relationships do you observe? (6 marks)
6. Conduct Moran’s tests on the variables in the nested model (OLS2) to detect any spatial autocorrelation. Report your conclusions at 5% level of significance. (6 marks)
7. Are there any “residual” spill-over effects in the nested model? (2 marks)
8. Fit CRIME on the usual spatial regression models and report your results. (4 marks)
9. Use the LM test to determine whether the SLM model or the SEM model is a better fit for the data. (4 marks)
10. Use the likelihood ratio test to determine whether.
    1. SARAR(1,1) model is superior to SLM or SEM (4 marks)
    2. SLM model is superior to SAR. (2 marks)
11. Combine the results of (i) and (j) and any other diagnostics and identify the most suitable spatial model. (3 marks)
12. Compare the model selected in (k) to the nested linear model (OLS2) and determine whether the spatial specification is superior to the usual linear regression specification for this dataset. (3 marks)
13. Determine whether a higher lag order is more suitable for the model selected in (k). (5 marks)

**\* When conducting model selection, it is better to have more than one line of arguments supporting your decision (AIC, BIC, R-square, tests, etc.)**