Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
June 13th, 2023

Problem n.3

The shapefile `P_World_Energy_use_per_capita` contains world countries borders for which the energy use per capita in 2019 is available. The spatial latitude/longitude coordinates (source here) have already been *projected* and converted to EPSG:3395 (source here), a variant of the Mercator projection used for maritime navigation, represented in meters.

The world countries are *univocally* identified by an `ISO` (a two-letter acronym, to be used as ID variable). The information related to the primary energy consumption per capita (kWh/person) is included within the variable `En_Cons`.

By using **either** the `GeoDa` application **or** the R package `rgeoda`, answer the following questions. **Remark**: in the solution file, specify which application you are using. Moreover, if you use `GeoDa` application, either you submit a .pdf file with the screenshot of the main performed steps, or you explain those steps in words on the .pdf file of your solutions.

a) What is the advantage of having changed the projection? Why should you be careful before starting an analysis regarding whether your data is projected or not?

b) Define a Rook contiguity-based spatial weight with an order of contiguity equal to 1.

- Report the *minimum* and *maximum*, number of neighbors, as well as the *sparsity percentage* (i.e., the percentage of non-zero values).

- *How many* countries present the *minimum* number of neighbors you found?

- Comment on why rook contiguity-based spatial weight might not be appropriate for such a dataframe.

c) Define now a distance-band-based spatial weight with Geometric centroids and Euclidean distance.

- What is the critical distance cutoff such that each location has at least one neighbour

- How many neighbours does Egypt (`ISO` code is 'EG') have

d) Focus on `En_Cons`; based on the weight defined in c), report

- The LISA cluster Map. What socio-economic conclusions can you draw from the results displayed in the LISA cluster map?

- The slope of the line displayed in the Moran's scatterplot. Among the countries with a positive spatial autocorrelation, report the `ISO` code of the country with the highest $\sum_j w_{ij}(z_j - \overline{z})$

Remark: If you are working in `R`, set `legend(...., cex = 0.5)`

Upload your solution here