Assignment 5

Objective: Introduce students to spatial analysis using Python, focusing on spatial statistics, understanding spatial dependence, creating spatial weight matrices, and running a basic spatial regression.

Dataset: For this assignment, students are expected to find, download and merge required data, which are expected to be about US counties' socio-economic data, including spatial data (latitude, longitude, shapefiles) and socio-economic indicators.

Tasks:

1. Load and Visualize the Dataset:
   - Use geopandas to import the dataset.
   - Find a shapefile, merge,
   - Plot a map showing the spatial distribution of counties.
2. Spatial Weight Matrix:
   - Create a spatial weights matrix for the counties.
3. Spatial Autocorrelation:
   - Search and learn what is Moran's I.
   - Calculate Moran's I for a socio-economic indicator (e.g., income levels).
4. Spatial Regression:
   - Run a spatial regression model with a socio-economic variable as the dependent variable.
   - Interpret the coefficients briefly in comments.
5. Visualization:
   - Create a choropleth map showing the distribution of the selected socio-economic indicator.

Submission: A Jupyter Notebook containing the code for each task.

Data access: Students can access the US counties' socio-economic data from a few public data repositories.

US Census Bureau: The Census Bureau provides detailed socio-economic data at the county level. You can access this data through the American FactFinder tool or the new data platform, data.census.gov. This platform allows for searching and downloading various data sets, including demographic, economic, and geographic data.

The National Historical Geographic Information System (NHGIS): NHGIS offers historical census data and GIS-compatible boundary files for the United States between 1790 and the present. It's a great resource for socio-economic and demographic data and can be accessed at nhgis.org.

The United States Geological Survey (USGS): For more geographical or environmental data, the USGS is a valuable resource. Their website, usgs.gov, provides access to a wide range of data, including geographic information system (GIS) data.

The National Atlas of the United States: This resource provides a comprehensive, map-like view into the enormous wealth of data collected by various agencies of the federal government. It can be accessed at nationalmap.gov.

When using these datasets, it's crucial for you to understand how to navigate these platforms, select relevant data, and download it in a format that can be used for their analysis. Most of these platforms provide data in CSV format, which can be easily imported into Python using libraries like Pandas, and shapefile format for spatial data, which can be used with geopandas.