

Assignment 5: Geographic Data Analysis

Assignment Description

In this assignment, you will be using a database of geographic data provided for you in the PySal library to create two plots: a choropleth map and a proportional symbol map.

In addition to these two plots, you will compute the value of Moran's I for this data.

Hint: Use `"%matplotlib inline"` to display the graph on the Jupyter Notebook (To allow the grader to identify the PartID, add the cell magic after the Graded Cell and PartID comments.).

Directions

The data for this assignment includes the United States' lower 48 states. In addition to the state-by-state data, the dataset contains shape files for each state that you can use to create the choropleth and proportional symbol maps.

Using the data provided, perform the required analyses and create the requested maps.

Submission Directions for Assignment Deliverables

This assignment will be auto-graded. We recommend that you use Jupyter Notebook in your browser to complete and submit this assignment. In order for your answers to be correctly registered in the system, you must place the code for your answers in the cell indicated for each question. In addition, you should submit the assignment in the cell's display area. The display area should contain only your answer with no extraneous information, or else the answer may not be picked up correctly.

Each cell that is going to be graded has a set of comment lines at the beginning of the cell. These lines are extremely important and must not be modified or removed. (Graded Cell and PartID comments must be in the same line for proper execution of code.)

Please execute each cell in Jupyter Notebook before submitting.

NOTE: For each question, be sure to add your code in the exact cell that has the comment `"# Graded Cell, PartID: _____"`. If you add extra cells or split your code up into multiple cells, ensure you are adding your output `print()` statement and the chart in the cell containing the comment with regard to each question.

NOTE: If you lose the partId that is mentioned in the first line of each cell with regard to each question, use the syntax described below to add it.

- Question 1: # Graded Cell, PartID : CkcsR
- Question 2: # Graded Cell, PartID : FqNRm
- Question 3: # Graded Cell, PartID : CtQYv

Question 1:

Using the PySAL Data, create a choropleth map of the United States that depicts the per capita income of each US state in 2009.

You only need to draw plot for the main land states (except Alaska and Hawaii). You need to show clear illustration of latitude and longitude in the plot.

Note: The PySAL and GeoPandas libraries both contain utility functions that may make this task easier.

```
In [1]: # Graded Cell, PartID: CkcsR

import pysal
import geopandas as gpd
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline

# your code here
import os
os.environ['USE_PYGEOS'] = '0'
import pysal.lib as pysl
from pysal.explore import esda
from pysal.lib import weights, examples

us_income = pd.read_csv(pysal.lib.examples.get_path('usjoin.csv'))
us_income_shape = gpd.read_file(pysal.lib.examples.get_path('us48.shp'))
us_income_shape['per-capita-income-2009'] = np.array(us_income["2009"])
ax = us_income_shape.plot(column = "per-capita-income-2009" , legend = True)
ax.set_title("Per Capita Income in 2009")
plt.rcParams["figure.figsize"] = (40,8)
plt.show()
```

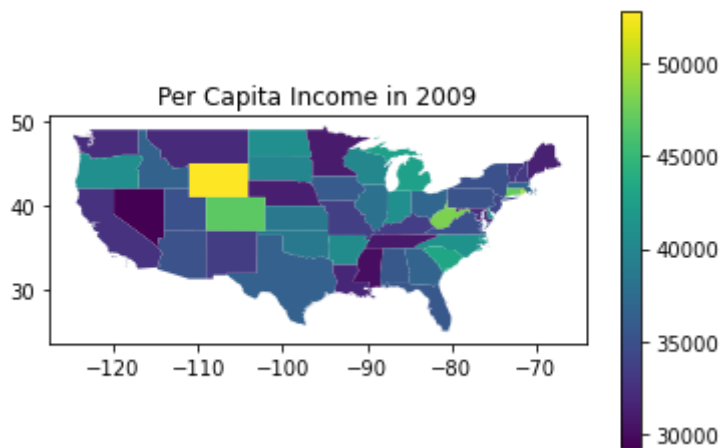
D:\Anaconda\lib\site-packages\geopandas_compat.py:123: UserWarning: The Shapely GEOS version (3.11.1-CAPI-1.17.1) is incompatible with the GEOS version PyGEOS was compiled with (3.10.4-CAPI-1.16.2). Conversions between both will be slow.

warnings.warn(
<ipython-input-1-759ab0e95f94>:4: UserWarning: Shapely 2.0 is installed, but because PyGEOS is also installed, GeoPandas will still use PyGEOS by default for now. To force to use and test Shapely 2.0, you have to set the environment variable USE_PYGEOS=0. You can do this before starting the Python process, or in your code before importing geopandas:

```
import os
os.environ['USE_PYGEOS'] = '0'
import geopandas
```

In a future release, GeoPandas will switch to using Shapely by default. If you are using PyGEOS directly (calling PyGEOS functions on geometries from GeoPandas), this will then stop working and you are encouraged to migrate from PyGEOS to Shapely 2.0 (https://shapely.readthedocs.io/en/latest/migration_pygeos.html).

```
import geopandas as gpd
D:\Anaconda\lib\site-packages\spaghetti\network.py:39: FutureWarning: The next major release of pysal/spaghetti (2.0.0) will drop support for all ``libpysal.cg`` geometries. This change is a first step in refactoring ``spaghetti`` that is expected to result in dramatically reduced runtimes for network instantiation and operations. Users currently requiring network and point pattern input as ``libpysal.cg`` geometries should prepare for this simply by converting to ``shapely`` geometries.
warnings.warn(f"{dep_msg}", FutureWarning)
```



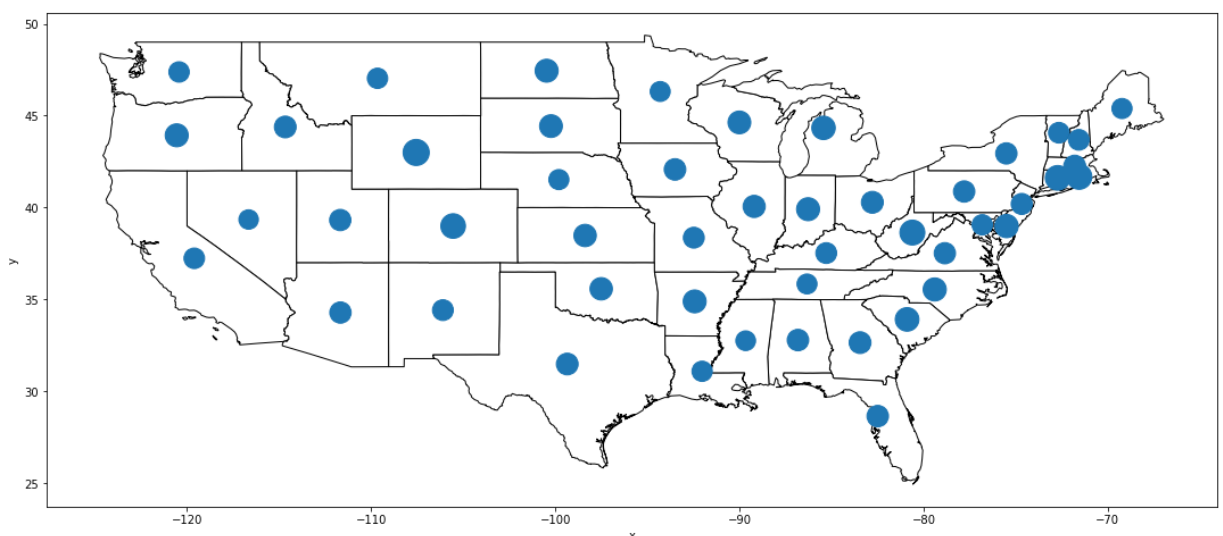
Question 2:

Again using the PySAL Data, create a proportional symbol map showing a dot at the centroid of each state that is scaled to the per capita income of each US state in 2009.

You only need to draw plot for the main land states (except Alaska and Hawaii). You need to show clear illustration of latitude and longitude in the plot.

Notes: The demonstration notebook for this unit contains code that performs a similar task and may be a useful reference for your assignment.

```
In [2]: # Graded Cell, PartID: FqNRm
# your code here
us_income_shape['centroid_column'] = us_income_shape.centroid
centroids = list(us_income_shape['centroid_column'])
final_df = pd.DataFrame({'y':[centroids[i].y for i in range(len(centroids))], \
                          'x':[centroids[i].x for i in range(len(centroids))], \
                          'data':list(us_income_shape['per-capita-income-2009'])})
base = us_income_shape.plot(color = 'white', edgecolor = 'black')
plt.rcParams["figure.figsize"] = (40,8)
final_df.plot( kind = 'scatter', x = 'x', y = 'y', s = final_df['data'] * 0.01, ax =
plt.show()
```



Question 3:

Using the same data, compute the value of Moran's I for the per capita income of each US state in 2009 using Rook Continuity. Report the value of I rounded to 4 decimal places (i.e. x.XXXX)

You only need to print a single value in this question.

Notes: Again, the PySal and GeoPandas libraries may contain useful utility functions.

```
In [3]: # Graded Cell, PartID: CtQYv
# your code here
y = us_income_shape['per-capita-income-2009']
w = weights.Rook.from_shapefile(examples.get_path("us48.shp"))
moran_stat = esda.Moran(y,w, two_tailed = False)
print (round(moran_stat.I, 4))

0.0407
```

```
In [ ]:
```