Assignment 3 - GGS366 Spatial Computing

Due February 25th 2024

Please find the assignment questions enclosed, totaling 100 available points.

Remember, if you just rely on genAI, you will not learn the fundamental basics, and thus fail the graded closed-book test/exam planned for later in the semester. So, the important thing here is to really make a good attempt at solving these problems without any assistive tools, based on the materials you have learnt in the first few weeks of the class.

To submit the work, you need to write your answers in a colab .ipynb notebook and then:

1. Print the .ipynb notebook to a .pdf and submit for review on blackboard.
2. Submit also the actual .ipynb notebook on blackboard allowing your code to be easily run.

Without submitting both of these files like this, you will receive a 20-point penalty to your overall grade. Submitted work may be checked for plagiarism, including for GenAI usage. The Mason honor code applies.

```
# Q1A: Load in the `regional_data.csv` file for Bangladesh, which
# represents sub-national regions with associated population, area
# and population density data. Explore/print the df metadata and provide a
# detailed explanation of what the different columns mean. Note: a
# 'detailed explanation' requires more than a 1 line description (10 points).

# Q1B: Subset the df on GID_1, GID_3, and the population column. Check
# how many rows are in the original df. Compare to your subset after
# removing incomplete rows. Print the answers (10 points).

# Q2: Present summary statistics for the population, area_km2, and
# population density at the GID_3 level (e.g., sum, mean, median,
# min, max etc.) (20 points).

# Q3: Now present the same summary statistics at the GID_1 level
#(e.g., sum, mean, median, min, max etc.). Provide a detailed description
# of what these statistical metrics represent. Good answers will include
# math formulae for the statistical metrics explanations (20 points).

# Q4: Convert the area to square meters, and then present the summary
# statistics at the GID_1 level. Report your metrics using scientific
# notation (20 points).

# Q5A: Add a new variable to your df which represents the population
# in millions (10 points).

# Q5B: Write out the three different datasets you generated to .csv
# and .xlsx. Do not include the index (10 points).
```