GDAA 1001 - Fundamentals of Spatial Data Analytics

Assignment #1 - Data Wrangling in R

Due: Friday, October 21 at 11:59pm | Value: 20%

Instructions

For this assignment, you will demonstrate your ability to carry out a range of data transformation tasks in R using functions from the `dplyr` package (and others).

The data required for this assignment are available on Brightspace. Download the zipped folder called "world95". Unzip its contents to your working directory. You should end up with five files in total, including the *world.shp* and *world_data.csv* files.

Once you have downloaded and unzipped the required files, proceed to complete the data transformation and mapping tasks outlined below.

Submit a PDF created from an RMarkdown file. The PDF should include all of your code and any text that you deem necessary to include. Also submit your .Rmd file.

Do not submit a script file (.R), nor an HTML output of your RMarkdown document. You must submit a PDF output and the corresponding .Rmd file.

Use the class notes from the week of Oct. 11 as a guide to completing the tasks.

Task #1 - Subsetting Rows and Columns

Using `dplyr` functions, subset the `world_data` dataset so that only only countries that meet the following criteria are included:

- They are located in Eastern Europe
- Male life expectancy is greater that 68 years
- Population growth is negative

In addition, subset the dataset so that only variables/columns relevant to the above three criteria are included, along with the country names.

Include your code and the resulting tibble output. Your code should make use of piping syntax.

Task #2 - Summary Statistics at the Regional Scale

Calculate the male to female life expectancy ratio and output a grouped summary of the average male to female life expectancy at the regional scale (i.e, using the `Region` variable).

This should be accomplished in a single code chunk using piping syntax, and make exclusive use of `dplyr` functions. Show your input code and the resulting `tibble` output.

Task #3 - Mapping Spatial Data in R

Merge the `world` and `world_data` objects into a new object called `world_join`. Remove any countries with male and female literacy rates and daily calorie intake values of zero. Calculate the population density for the remaining countries. Calculate the male to female literacy ratio for the remaining countries.

Show (i.e., "print") a `data.frame` or `tibble` including only the top five countries in terms of population density, in descending order. That is, the country with the highest overall population density should appear at the top. Include *only* the country and population density columns.

Show (i.e., "print") another `data.frame` or `tibble` including only the bottom five countries in terms of male to female literacy, in ascending order. That is, the country with the lowest male to female literacy rate should appear at the top. Include *only* the country and male to female ratio columns.

Produce two choropleth maps showing:

1. Population Density
2. Male to Female Literacy

Each map should have a different color scheme. Make sure you include all of your code.