SI 618 Winter 2017 HW6

In this homework, you'll load data into R to do some aggregation and visualization, which is a typical workflow for exploratory data analysis.

First, download the attached TSV file: this 'countrydata\_withregion.tsv' file is your input file. Your ultimate goal is to generate a PDF report using RMarkdown that looks like 'si618hw6\_sample\_report.pdf'.

**Step 1: Load data (20 points)**

First load the provided TSV data file using the \*\*read.table()\*\* function.  Then print the first 15 rows of the data frame.

**Step 2: Scatter plot of log transformed data (20 points)**

Compute the natural logarithm of the area and the population of each country to produce a scatter plot using the plot(), qplot() or ggplot() functions.

**Step 3: Data aggregation by region (30 points)**

Sum the areas and populations of all countries by region using the \*\*aggregate()\*\* function, respectively. Then create two pie charts using the \*\*pie()\*\* function showing the areas of regions and populations of regions, respectively.

**Step 4: Visualization of Population per sq km of Region (30 points)**

Create a new data frame to hold the population per sq km of each region using the \*\*data.frame()\*\* function. Then sort the data frame by population per sq km in decreasing order with the help of the \*\*order()\*\* function. Finally, create a bar plot using **qplot() or ggplot()** function with **geom=“bar”**.

**What to submit:**

A zip file named 'si618hw6\_youruniquename.zip' containing:

* The R Markdown file you wrote named 'si618hw6\_report\_youruniquename.Rmd
* a PDF file created from your Rmd file using Knit PDF named 'si618hw6\_report\_youruniquename.pdf'