Programming Assignment 1
STAT 311

Please complete the following problems and submit a file named `STAT311-HW1.R` to Gradescope. You should start from the provided `STAT311-HW1.R` file on Canvas.

Remember:

- Do not rename provided data files or edit them in any way.
- Do not use global paths in your script. Instead, use `setwd()` interactively in the console, but do not forget to remove or comment out this part of the code before you submit. The directory structure of your machine is not the same as the one on Gradescope's virtual machines.
- Do not destroy or overwrite any requested variables in your program. I check them only after I have run your entire program from start to finish.
- Check to make sure you do not have any syntax errors. Reset the working environment and rerun your entire assignment to ensure it runs without errors using the source command.
- Make sure to name your submission `STAT311-HW1.R`

Overview

We will be looking at the `county` dataset from the `usdata` package that looks at a number of variables for US counties. You will need to install the package using `install.packages("usdata")` exactly once. You can import the package with `library(usdata)` each time you start RStudio or clear the workspace.

Part 1

Treating the `county` dataset as the population of US counties, create the dataframe `my.SRS` that samples from `county` and represents a simple random sample of $n = 250$ individual counties from all counties in the US.

Part 2

Treating the `county` dataset as the population of US counties create the dataframe `my.Stratified` that represents a stratified sample of individual counties from

all counties in the US, statified along the level of education using the `median_edu` variable. Due to the different sizes of strata, you should sample:

- 1 county from `below_hs`
- 14 from `hs_diploma`
- 17 from `some_college`
- 4 from `bachelors`

Part 3

Treating the `county` variable as a population of US counties create the dataframe `my.Clustered` that represents a cluster sample of individual counties from all counties in the US, clustered by state using the `state` variable. You should randomly sample all counties from a total of 5 clusters.