



Problem 1 (Micromap)

The attached CSV file contains age-adjusted death rates (based upon 100,000 people) and average annual percent change (trend) over 5 years from 2011-2015 for uterus cancer in females ages 50 and over. 95% confidence intervals for the death rate and trend are also included. Data were generated from the CDC State Cancer Profile website with the following filters selected (<https://statecancerprofiles.cancer.gov/map/map.withimage.php?00&136&058&00&2&02&0&1&5&0#results>). Follow the instructions and answer ALL of the questions below:

- I. Read the dataset into RStudio using the read.csv() function. [uteruscancer.csv](#)
- II. Using the micromapST() function, create a linked micromap that contains the following features and formatting exactly as follows:
 - a. The first column should display the maps where states are highlighted from the extremes to the middle state.
 - b. The second column should display the names of the states and their corresponding colors (full names or abbreviations are acceptable).
 - c. The third column should display the confidence intervals for age-adjusted death rates.
 - d. The fourth column should display the confidence intervals for the recent 5-year trend.
 - e. The rows should be sorted according to the recent 5-year trend in decreasing order from top to bottom.
 - f. Appropriate titles and labeling should be included to completely and accurately describe the information in the micromap and in each column including the source of the information.
 - g. Produce a PDF file containing the micromap.
- I. Provide a few sentences of commentary about the micromap to answer ALL of the following questions:
 - a. Which geographical areas are experiencing the largest increase in death rates over the 5-year period? Why should we care about this?
 - b. Which geographical areas are experiencing little to no change in death rates over the 5-year period? Why should we care about this?
 - c. Which geographical areas have a low age-adjusted death rate, but are experiencing a large increase in death rates over the 5-year period? Why should we care about this?
 - d. Are certain data elements missing? Why should we care about this?

Problem 2 (Probability Distributions in R)

Set your seed (pick a number)

- a. Create a vector 'x' containing 10000 random samples from a chi-square distribution with 10 degrees of freedom (hint: ?rchisq)
- b. Find the value of the chi-square distribution with 10 degrees of freedom corresponding to the 99th percentile (hint: ?qchisq)
- c. What percentage of observations in your sample fall above this value? Is it exactly 1%? Should it be? Discuss.
- d. Using ggplot, display a density histogram of the vector 'x' (with 50 bins and your choice of colors for the boundaries and fill) and overlay the histogram with the density plot of the vector (size of the curve should be greater than or equal to 1.5 and adjust value should be 1.5).

Submission Instructions

Please submit 1) a script that can be run to produce the corresponding plots (don't forget your library statements), 2) a PDF document containing the micromap and histogram, and 3) a word or PDF document containing your commentary. Alternatively, you may submit an RMarkdown file and a knitted PDF or Word document. Please e-mail me if you have any questions about this assignment or need clarification on the problem statement.