Practice Problem Set 2

MPH 5962 - R Skills Lab Spring 2017

January 26, 2017

Instructions

To complete the problems, create an R script and write code according to the instructions in each. Construct the syntax so that the answer to the item prints in the console. For example, if you are to create a vector of the numbers 1-3, the corresponding code should read:

x <- c(1,2,3)

x

so that the answer prints in the console after calling x. If the answer calls for a written response that cannot be printed through syntax, type a note in the script using the hashtag (*e.g.*, #here is my answer). Save the R script and name it with your last name underscore classEx*X*, where *X* is the exercise number. For example if your last name were Smith, name the first set of exercises *Smith_classEx1*

These practice problems use the County Health Rankings and Roadmap data. On Blackboard, find and download the abbreviated and cleaned dataset called **"CountyHealth.csv"**. You will also need to download the list of variable labels **"CountyHealthVariableLabels.csv"**. For more information on the data, visit Social Explorer, click Tables, choose Health Data, and go to the 2016 Data Dictionary. Social Explorer is accessible through WUSTL Libraries http://library.wustl.edu/. For even more information, visit the official site http://www.countyhealthrankings.org/.

Loading & exploring data from a CSV

1. Download and save the data in a place to which you know the file path. Load the data. You can also load the variables labels CSV as a different data frame for reference, or just open it in Excel. Choose whether to load string variables as factors or characters. Which is the default?

2. How many variables does the CountyHealth.csv have? How many of them are health indicators? How many observations are in the data frame?

3. How many variables have missing values? Which variable has the most missing values?

4. How many different classes of variables (*i.e.*, variable types) are in the dataset? What are they?

Investigating health measures & outcomes in US counties

5. Which county in the US has the highest primary care physicians per 1,000,000 population? What is the rate?

6. Which county in the US has the highest adult smoking rate? What is the rate? Which county has the lowest, and what is the rate?

7. What is the average rate of teen births in US Counties?

8. To get a specific range of values of a variable, you can use the & in an index of a data frame, i.e., r[r$x > 8$&r$x<11,] would call the rows of r where x is less than 11 but greater than 8. Find two counties that are around the average for teen birth rates in the US.

9. Are different measures of mortality correlated within US counties? Create a correlation matrix of infant, child, and premature age-adjusted mortality to find out. Which pair of the two are most correlated? Which pair are least correlated?

10. Plot the rates of low birth weight against infant mortality in US counties. Do the two seem correlated?

11. Compare California Counties to the overall average health indicators for US counties. On average, are California counties healthier than others?