Fall 2018
ISYE 670

Module 5 Homework
30 points

Download the data on Blackboard. Use the dataset to make following changes. It contains income generated in the years 2002 to 2015 for all the states.

1) (2 points) Import the data with the headers. How is the data organized? (i.e. what are the column names? What does each row of the data represent?)

Use only dplyr functions for the following questions.

2) Lookup the sample_n and sample_frac functions.
   a) (2 points) Use the appropriate function to randomly choose 15 rows.
   b) (2 points) Use the appropriate function to randomly choose 40% of the rows.

3) Output certain columns.
   a) (2 points) Output only the values for the year 2005 and States
   b) (2 points) Output only the values for all the years. (Hint, use the -)
   c) (2 points) Output only the values for columns that begin with a Y

4) Output only certain observations.
   a) (2 points) Output only the observations with an index values of A and C and N.
   b) (2 points) Output only the observations related to Illinois and California.

The next questions require using multiple functions.

5) (2 points) Output data that shows only the index, state, and 2010 income values greater than $1,500,000.
   a) (Extra Credit: 2 points) The same code can be re-written concisely using the pipe operator %>%. What is that command?

6) (3 points) Output only the state and 2006 values, and arrange the values in descending order of 2006 values. What are the 3 highest income generating states?

7) (3 points) Using the mutate function, create a new column "ratio" which divides the income of 2015 by 2014. Use the appropriate combination of commands so only the state, 2014, 2015, and the new column is saved or output.

8) (3 points) Using group_by and summarise, create a table grouped by Index and then summarizes each group by taking the mean of Y2007.

9) (3 points) Group the observations by index. Output data that shows the average ratio of 2015 divided by 2014 for each group.