Jeffrey Chao

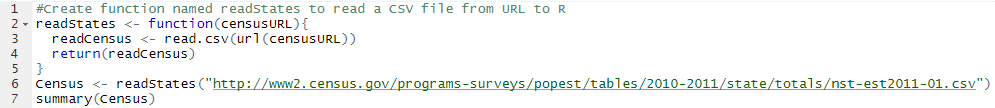
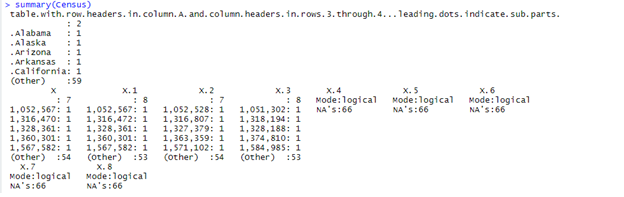
IST 687 – Homework 3

Date Due: 02/04/2020

Date Submitted: 02/02/2020

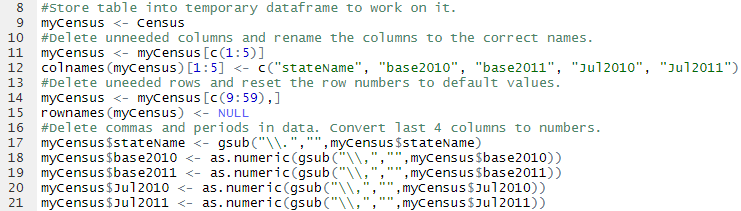
**Step 1: Create a function (named readStates) to read a CSV file into R**

1. *Note that you are to read a URL, not a file local to your computer.*
2. *The files is a dataset on state populations (within the United States).*

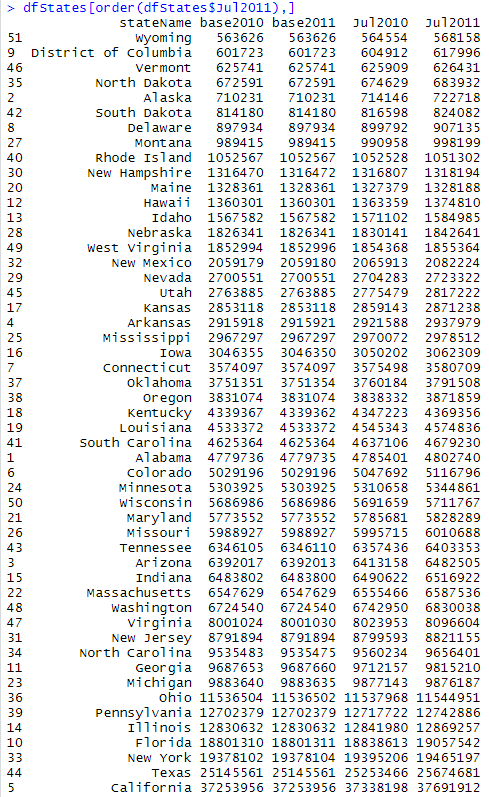
  


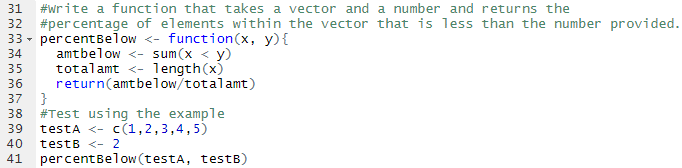
**Step 2: Clean the dataframe**

1. *Note the issues that need to be fixed.*
2. *Within your function, make sure there are 51 rows and 5 columns with the names:  
   (stateName, base2010, Base2011, Jul2010, Jul2011)*
3. *Make sure the last four columns are numbers (i.e. not strings*

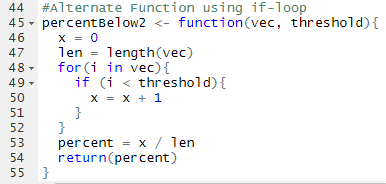


**Step 3: Store and Explore the dataset**

1. *Store the dataset into a dataframe called dfStates*
2. *Test your dataframe by doing mean(dfStates$Jul2011), the answer should be 6,109,645*  
     
   **Step 4: Find the state with the Highest Population**
3. *Based on the July2011 data, what is the highest population in a state?   
   What is the name of the state?*
4. *Sort the data, in increasing order, based on July2011 data.***Step 5: Explore the distribution of the states**
5. *Write a function that takes two parameters, a vector and a number.*
6. *The function returns the percentage of elements in the vector less than the number*
7. *For example, with inputs (1,2,3,4,5) and 2, the function would return 0.2*


1. *Test the function with the vector ‘dfStates$Jul2011’ and the mean of ‘dfStates$Jul2011’*
2. *There are many ways to write functions so please try to write multiple versions of this function – which do you think is best?*





Although it gives the same answers, I prefer the function without the if-loop since it seemed to be more concise and used less code.