Programming for BIG Data

Practical Assessment 4

In this practical assessment you will have to:

1) Create an R program that computes the overall sum of a set of numbers. The numbers are stored in three files called *data1.txt, data2.txt* and *data3.txt* . Your program should create a new file called *sum.txt*. The overall sum of the numbers read from the three input files should be written to *sum.txt*.

Your R program should contain:
i) A function that receives as input a filename, reads the numbers from the file and returns a vector of the numbers read.
Note: the numbers in data1.txt, data2.txt and data3.txt are placed one per line.
ii) A function that receives as input a vector of numbers and returns the sum of all the numbers in the input list of numbers.
ii) A function for writing a number to a file.
iv) Code to call the functions as appropriate.
v) Comments should be used to explain how each function processes its input.

2) Carry out an analysis of the LifeCycleSavings dateset using R.

Please provide the following information:
i) A description of the dataset and of the dataset format.
ii) R code and description of at least 1 analysis on the LifeCycleSavings data (e.g., average, standard deviation, frequency, etc.).
iii) R code and description of at least 1 plot on LifeCycleSaving data (e.g. histogram, barplot, pie chart, etc.).
iv) R code to save LifeCycleSavings dataset to a .csv file.
v) The code should be commented.