Parallel Programming

Practicing R

The data file for this assignment can be found at Sakai -> Resources ->TitanicPassengers.csv.

Read this data file into R – might have to research reading in a .csv file.

Test your input with a quick print of the top 100 rows of code. Also check the data type of the input – is it a data frame. Comment these two lines out to begin.

Complete the following (include the code but comment it out if this code gets too long as you continue execution):

1. Print out the number of rows and the number of columns in our new data frame.
2. What data structure is created from the above operation?
3. Print out only the column headers of our data frame.
4. Change the first column from “X” to “UniqID”.
5. Obtain (print) all of the categories of the categorical variable passengerClass.
6. Using the summary function – print out a quick summary of your data frame.
7. Using the str ( ) function find out the type of objects contained within the data frame.
8. Print out how many survivors were female / male.
9. Print out how many of those who died were female / male.
10. Print out how many under the age of 18 and female died.
11. Print out how many under the age of 18 and male died.
12. Print out how many over 25 survived in first class, second class and third class.
13. Print out how many under 25 survived in first class, second class and third class.
14. Using column names, retrieve a data frame slice of the two columns survived and age.
15. Find the mean age of all survivors.
16. Find the mean age of all non-survivors.
17. Create a histogram of the age of all the passengers on the titanic – add color.
18. Create a plot of survived versus sex of the survivor – add color and a legend.
19. Create a scatter plot of age versus survivor – add color.
20. Create a scatter plot of age versus not a survivor – add color.

Add comments organizing your code with all the above numbers and statements so I can see the exact operations for each step.