Assignment 1

For this assignment, you will use the Cereal datasets. Work all the questions in an R script. Create a document with the question, your answer and the output, which you shall submit on the platform.

1. Read this CSV file into R and use the *head* command to look at the first few rows.
2. Determine the number of records in this dataset.
3. Use an appropriate command to display the data types of the variables in the data frame.
4. Augment this dataset by adding a new column *'totalcarbo'* which is the sum of 'carbo' and 'sugars'
5. Determine the number of cereals in the dataframe which are 'cold' cereals.
   Hint: Based on this criteria, make an appropriate *subset* of the data and then count the number of rows.
6. Determine the number of distinct Suppliers in the dataset. Hint: Use *length* and *unique*.
7. Get a subset of the dataframe of all cereals having *'carbo'* between 5 and 10 units inclusive.
8. Obtain a subset of the dataframe containing cereals with zero fat and keep only the variables '*Cereal.name*', *'carbo'* and *'sugars'*. Use the head command to view the top rows of this dataframe.
9. Rename the column *'Supplier'* to *'Distributor'*
10. Produce a scatter plot with 'calories' on the x axis and 'carbo' on the y-axis.
11. Produce a bar chart (barplot) with the distinct distributors against the number of cereals that they distribute.
12. List the names of all the objects found in the workspace.
13. Create a vector *'fat'* with the data from the column *'fat'*. Then display its minimum, mean and maximum.
14. Write a function named 'retrieve' that takes three arguments: *'startCalorie'*, *'endCalorie'*, and *'range'*. Given that range is a vector of numerical ids, the function will loop across the records found that these positions and will print the row if the calorie of the cereal falls in the range *'startCalorie'* to *'endCalorie'*. You need to use a for loop and if statements.
15. Test the above function by making at least 5 different function calls.