CS 5301 – Programming Foundations for Data Analytics
Assignment 1

Due: 09/21/2021, 23:59

Submission: You should write all your solutions into a single **.r** file (R Script). Submit your **.r** file through Blackboard. Include the question number as a comment. Your **.r** file should be in the following format:

```
#Q1.a
...

#Q1.b
...

#Q1.c
...

#Q2.a
...
 ⋮
```

Q1 (15 pts). In your R script, compute (and print) the following sums using vector arithmetic and sum function:

a. $1 \times 1000 + 3 \times 998 + 5 \times 996 + \cdots + 997 \times 4 + 999 \times 2$

b. $\frac{1}{2} - \frac{3}{4} + \frac{5}{6} - \frac{7}{8} + \cdots - \frac{99}{100}$

c. $1 + \frac{2}{3} + \frac{4}{9} + \cdots + \frac{2^{25}}{3^{25}}$

Q2 (20 pts). `sample` function in R can be used generate random numbers. The first argument needs to be a vector of one or more elements from which to choose and the second argument is the number of random values to be generated. The third argument replace is FALSE by default and hence the generated numbers are unique. Changing it to TRUE allows to include a number more than once. In your R script, generate 100 different random numbers from 1 to 1000. Store the generated numbers in a vector and print the following values in the vector:

a. Second smallest number by getting the second element after sorting the values in increasing order.
b. All odd numbers.
c. Mean and standard deviation of all numbers.
d. Total number of values greater than 400 and less than 600.

Q3 (30 pts). In your R script, store the monthly average precipitation (inches) of Houston and Miami as given in the following table.

	Houston	Miami
January	3.68	1.88
February	2.98	2.07
March	3.36	2.56
April	3.60	3.36
May	5.15	5.52
June	5.35	8.54
July	3.18	5.79
August	3.83	8.63
September	4.33	8.38
October	4.50	6.19
November	4.19	3.43
December	3.69	2.18

a. Create two different vectors for Houston and Miami. Set the names attributes as the names of the months. Print the vectors for Houston and Miami.
b. Print the number of months in which the precipitation in Houston is greater than Miami. Note that you can use sum function to count the TRUE values in a logical vector.
c. Print the months in which the precipitation difference between Houston and Miami is less than 1 inch.
d. In Houston, it is expected that the precipitation in 2022 will be 2% greater than the averages. Compute the expected precipitation of 2022 and print monthly precipitation after rounding to 2 decimal places using **round** function.
e. Compute and print the monthly percentage difference in precipitation. Percentage difference is equal to the difference between two values divided by the average of the two values. For instance, percentage difference of 15 and 25 is computed as $\frac{|25-15|}{\left(\frac{25+15}{2}\right)} \times 100$, which is equal to 50%. Therefore, your output should be:

```
  January  February     March     April       May      June      July
64.748201 36.039604 27.027027  6.896552  6.935333 45.932325 58.193980
   August September   October  November  December
77.046549 63.729347 31.618335 19.947507 51.448041
```

f. Convert the precipitation to cm and print the statistics of Houston and Miami using summary function. The formula for conversion is cm <- inch*2.54.

Q4 (35 pts). In your R script:

a. (10 pts) Store following age category values in a factor named age:

"21-40", "41-60", "41-60", "21-40", "0-20", "21-40", "0-20", "41-60", "21-40", "0-20"

Set the order of the levels as "0-20" < "21-40" < "41-60" and print the factor.

Add one more value ("60+") to your factor as the last element. Since it is not a defined level in the factor, you need to add the level first. Print the factor after adding the new value.

b. (10 pts) Create a list named myList with the following four elements (with their names):
   1. **numbers**: The even numbers from 2 to 100.
   2. **strings**: Five character strings “x”,”y”,”z”,”a”,”b”.
   3. **matrix**: One 2x3 matrix $\begin{matrix} 9 & 3 & 7 \\ 8 & 2 & 6 \end{matrix}$.
   4. **logical**: Two logical values TRUE and FALSE.

Print the sum of the **numbers**. Change the value of the last element of the **strings** as “k”. Change the **matrix** with its transpose. Remove **logical** from the list. Print the list.

c. (15 pts)

	Monday	Tuesday	Wednesday	Thursday	Friday
Week1	325	262	229	337	465
Week2	199	223	148	389	513
Week3	244	412	403	399	336

- Store the daily sales of a product (given above) for three weeks in a matrix object.
- Assign the row names and column names as given in the table above.
- Compute and print the average sales of each week (row) and average sales of each day (column).
- Insert the sales on the weekend to the matrix using **`cbind`** function.

	Saturday	Sunday
Week1	872	719
Week2	666	702
Week3	598	504

- Print the sales of Week1 and Week3 on Tuesday, Thursday, and Saturday as a submatrix.