R PROGRAMMING ASSIGNMENMT-6

1.Write a programme that uses the column name to retrieve a specific column from a Data Frame?

ANS:

R Programming Code :

```
exam_data = data.frame(

name = c('Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael',
'Matthew', 'Laura', 'Kevin', 'Jonas'),

score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),

attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1),

qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes')

)

print("Original dataframe:")

print(exam_data)

print("Extract Specific columns:")

result <- data.frame(exam_data$name,exam_data$score)

print(result)
```

Output:

```
[1] "Original dataframe:"
       name score attempts qualify
```

```
1  Anastasia  12.5        1     yes
2       Dima   9.0        3      no
3  Katherine  16.5        2     yes
4      James  12.0        3      no
5      Emily   9.0        2      no
6    Michael  20.0        3     yes
7    Matthew  14.5        1     yes
8      Laura  13.5        1      no
9      Kevin   8.0        2      no
10     Jonas  19.0        1     yes
[1] "Extract Specific columns:"
   exam_data.name exam_data.score
1       Anastasia            12.5
2            Dima             9.0
3       Katherine            16.5
4           James            12.0
5           Emily             9.0
6         Michael            20.0
7         Matthew            14.5
8           Laura            13.5
9           Kevin             8.0
10          Jonas            19.0
```

2.Can you write a programme to extract the 2nd and 5th rows from a Data Frame with the 1st and 4th columns?

ANS:

R Programming Code :

```
exam_data = data.frame(
name = c('Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'),
score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),
attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1),
qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes')
)
print("Original dataframe:")
print(exam_data)
print("Extract 2rd and 5th rows with 1st and 4rd columns :")
result =  exam_data[c(2,5),c(1,4)]
print(result)
```

Output:

```
[1] "Original dataframe:"
        name score attempts qualify
1  Anastasia  12.5        1     yes
2       Dima   9.0        3      no
3  Katherine  16.5        2     yes
4      James  12.0        3      no
```

5 Emily 9.0 2 no

6 Michael 20.0 3 yes

7 Matthew 14.5 1 yes

8 Laura 13.5 1 no

9 Kevin 8.0 2 no

10 Jonas 19.0 1 yes

[1] "Extract 2rd and 5th rows with 1st and 4th columns :"

name attempts

3 Dima 2

5 Michael 2

3.What is the function sample() in R?

ANS:

The **sample()** function in R allows you to take a random sample of elements from a dataset or a vector, either with or without replacement.

The basic syntax for the sample() function is as follows:

sample(x, size, replace = FALSE, prob = NULL)

WHERE;

x: a dataset or vector from which to choose the sample
size: size of the sample
replace: should sampling be with replacement? (this is FALSE by

default)

prob: a vector of probability weights for obtaining the elements of the vector being sampled

4.In R, how can you make a table without using an external file?

ANS:

MyTable= data.frame ()

edit (MyTable)

The above code will open an Excel Spreadsheet for entering data into MyTable

Method 2:

The following code shows how to create a table with 4 columns a 2 rows from scratch:

```
#create matrix with 4 columns
tab <- matrix(rep(2, times=8), ncol=4, byrow=TRUE)

#define column names and row names of matrix
colnames(tab) <- c('A', 'B', 'C', 'D')
rownames(tab) <- c('F', 'G')

#convert matrix to table
tab <- as.table(tab)

#view table
```

```
tab


  A B C D
F 2 2 2 2
G 2 2 2 2
```

5.How would you write an R programme to add a new column to a DataFrame?

ANS:

R Programming Code:

```
exam_data = data.frame(

name = c('Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael',
'Matthew', 'Laura', 'Kevin', 'Jonas'),

score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),

attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1),

qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes')

)

print("Original dataframe:")

print(exam_data)

print("New data frame after adding the 'country' column:")

exam_data$country =
c("USA","USA","USA","USA","USA","USA","USA","USA","USA","USA")
```

```
print(exam_data)
```

Output:

```
[1] "Original dataframe:"
        name score attempts qualify
1  Anastasia  12.5        1     yes
2       Dima   9.0        3      no
3  Katherine  16.5        2     yes
4      James  12.0        3      no
5      Emily   9.0        2      no
6    Michael  20.0        3     yes
7    Matthew  14.5        1     yes
8      Laura  13.5        1      no
9      Kevin   8.0        2      no
10     Jonas  19.0        1     yes
[1] "New data frame after adding the 'country' column:"
        name score attempts qualify country
1  Anastasia  12.5        1     yes     USA
2       Dima   9.0        3      no     USA
3  Katherine  16.5        2     yes     USA
4      James  12.0        3      no     USA
5      Emily   9.0        2      no     USA
```

6	Michael	20.0	3	yes	USA
7	Matthew	14.5	1	yes	USA
8	Laura	13.5	1	no	USA
9	Kevin	8.0	2	no	USA
10	Jonas	19.0	1	yes	USA

6.How can I get RStudio and install it in Anaconda?

ANS:

To Install RStudio in Anaconda for Windows

Here are the steps to install RStudio in Anaconda for Windows:

Step 1) Open the downloaded exe and click Next

Step 2) Accept the License Agreement

Step 3) Select Just Me and click Next

Step 4) Select Destination Folder and Click Next

Step 5) Click Install in next Screen

Step 6) Installation will begin

Once done, Anaconda will be installed

7.Can you write a programme to remove a Data Frame column?

ANS:

Data Frame:

```
set.seed(456)
mydata <- data.frame(a=letters[1:5], x=runif(5,10,50), y=sample(5), z=rnorm(5))
```

Output:

	x	y	z
a	13.58206	2	0.3240611
b	18.42049	1	0.6906430
c	39.31821	4	0.2505479
d	44.08534	3	1.0073523
e	41.53592	5	0.5732347

Removing a column from the data frame:

```
df = subset(mydata, select = -c(x,z) )
```

Output:

```
a y
1 a 2
2 b 1
3 c 4
4 d 3
5 e 5
```