

DATA SCIENCE FOR ENGINEERS

Assignment-1

1. Choose the variable name that is invalid in R.

Solution: c

Variable name which starts with a number is invalid.

2. The command to access help in R Studio is—

Solution: d

There is no function called help.stand in R whereas others are valid.

3. In the R code given below, the value of "i" at which the loop breaks is ---

```
n=100
sum=0
for(i in seq(1,n,3)){
  sum=sum+i
  print(c(i,sum))
  if(sum>15)
    break
}
```

Solution: c

```

> n=100
> sum=0
> for(i in seq(1,n,3)){
+     sum=sum+i
+     print(c(i,sum))
+     if(sum>15)
+         break
+ }
[1] 1 1
[1] 4 5
[1] 7 12
[1] 10 22

```

4. The library that supports right `_join ()` function in R is---

Solution: a

`dplyr` supports right `_join ()` function in R

5. The value of sum and month when `i = 9` is -----

```

n=5
sum=0
for(i in 1:11){
  sum=sum+(i+2)
  print(c(month.abb[i+1],sum))
}

```

Solution: c

```

> n=5
> sum=0
> for(i in 1:11){
+     sum=sum+(i+2)
+     print(c(month.abb[i+1],sum))
+ }
[1] "Feb" "3"
[1] "Mar" "7"
[1] "Apr" "12"
[1] "May" "18"
[1] "Jun" "25"
[1] "Jul" "33"
[1] "Aug" "42"
[1] "Sep" "52"
[1] "Oct" "63"
[1] "Nov" "75"
[1] "Dec" "88"

```

6. Table 1 provides the scores of the students in three subjects. Create a data frame called `students_scores` out of Table 1. Which of the following options gives Table 2 as an output?

Table 1

Name	Mathematics	English	Science
Ram	85	80	79
Prabhu	70	79	96
Sita	90	73	95
Santosh	95	90	80
Lohith	80	93	87

Table 2

Name	Variable	Value
Ram	Mathematics	85
Prabhu	Mathematics	70
Sita	Mathematics	90
Santosh	Mathematics	95
Lohith	Mathematics	80
Ram	English	80
Prabhu	English	79
Sita	English	73
Santosh	English	90
Lohith	English	93
Ram	Science	79
Prabhu	Science	96
Sita	Science	95
Santosh	Science	80
Lohith	Science	87

Solution: a

```
> students_scores=data.frame("Name"=c("Ram","Prabhu","Sita","Santosh","Lohith"),
+                             "Mathematics"=c(85,70,90,95,80),
+                             "English" = c(80,79,73,90,93),
+                             "Science" = c(79,96,95,80,87))
> library(reshape2)
> Df=melt(students_scores, id.vars = c("Name") , measure.vars =c("Mathematics", "English", "Science"))
> print(Df)
```

	Name	variable	value
1	Ram	Mathematics	85
2	Prabhu	Mathematics	70
3	Sita	Mathematics	90
4	Santosh	Mathematics	95
5	Lohith	Mathematics	80
6	Ram	English	80
7	Prabhu	English	79
8	Sita	English	73
9	Santosh	English	90
10	Lohith	English	93
11	Ram	Science	79
12	Prabhu	Science	96
13	Sita	Science	95
14	Santosh	Science	80
15	Lohith	Science	87

Create a data frame with given vectors below.

rank = c(1,2,3)

competitor = c("Usain","Tyson","Yohan")

mark = c(9.58, 9.69, 9.65)

Store the data frame in the variable named – **“athletics”**.

7. The command to add a new row to the data frame “athletics” with the following values passed to each vector?

rank= 4, **competitor**="Asafa", **mark**=9.72

Solution: c

```
> rank = c(1,2,3)
> competitor = c("Usain","Tyson","Yohan")
> mark = c(9.58, 9.69, 9.65)
> athletics=data.frame(rank,competitor,mark)
> print(athletics)
  rank competitor mark
1    1      Usain 9.58
2    2       Tyson 9.69
3    3       Yohan 9.65
> athletics=rbind(athletics,data.frame(rank=4,competitor="Asafa",mark=9.72))
> print(athletics)
  rank competitor mark
1    1      Usain 9.58
2    2       Tyson 9.69
3    3       Yohan 9.65
4    4       Asafa 9.72
```

Answer question 8 based on the data frame created at the end of Q7.

8. The command to add a new column to the data frame **“athletics”** with vector **“nationality”** taking values **“JAM”,”USA”,”JAM”,”JAM”** is:-

Solution: b

```
> rank = c(1,2,3)
> competitor = c("Usain","Tyson","Yohan")
> mark = c(9.58, 9.69, 9.65)
> athletics=data.frame(rank,competitor,mark)
> print(athletics)
  rank competitor mark
1    1      Usain 9.58
2    2       Tyson 9.69
3    3       Yohan 9.65
> athletics=rbind(athletics,data.frame(rank=4,competitor="Asafa",mark=9.72))
> athletics=cbind(athletics,data.frame(nationality = c("JAM","USA","JAM","JAM")))
> print(athletics)
  rank competitor mark nationality
1    1      Usain 9.58         JAM
2    2       Tyson 9.69         USA
3    3       Yohan 9.65         JAM
4    4       Asafa 9.72         JAM
```

Answer question 9 based on the data frame created at the end of question 8.

9. The correct way to extract all elements for which "**mark**" is less than 9.69 using the "**subset**" command is ---

Solution: a

```
> rank = c(1,2,3)
> competitor = c("Usain","Tyson","Yohan")
> mark = c(9.58, 9.69, 9.65)
> athletics=data.frame(rank,competitor,mark)
> print(athletics)
  rank competitor mark
1    1      Usain 9.58
2    2       Tyson 9.69
3    3       Yohan 9.65
> athletics=rbind(athletics,data.frame(rank=4,competitor="Asafa",mark=9.72))
> athletics=cbind(athletics,data.frame(nationality = c("JAM","USA","JAM","JAM")))
> print(athletics)
  rank competitor mark nationality
1    1      Usain 9.58          JAM
2    2       Tyson 9.69          USA
3    3       Yohan 9.65          JAM
4    4       Asafa 9.72          JAM
> subset(athletics, athletics$mark <9.69)
  rank competitor mark nationality
1    1      Usain 9.58          JAM
3    3       Yohan 9.65          JAM
```

10. Which of the following defined functions will return the output as **9.37**?

Solution: c

```
> func_multi= function(a,b,c)
+ {
+   result=(a*b)**0.5+(a*c)**0.5
+   return(result)
+ }
>
> func_multi(4,5,6)
[1] 9.371115
```