Class06

Safiya Sayd(PID:A18027139)

Today we will get more exposure to function in R. We call functions to do all our work and today we will learn how to write our own.

First function

Arguments 2 and 3 have a default values (because we set y=0 and z=1)

```
add <- function(x, y=0, z=1) {
  x+ y+ z }
```

Can I use this?

```
add(1,1)
```

```
[1] 3
```

```
add(1, c(10, 100))
```

```
[1]  12 102
```

```
add(100)
```

```
[1] 101
```

```
add(100, 1, 1)
```

```
[1] 102
```

Second function

lets write a function that generates random nucleotide sequences.

We can make use of the in-built **sample()** function in R to help us here

```
sample(x=1:10, size=9)
```

```
[1]  7  6  2  9  4  3  8  1 10
```

```
sample(x=1:10, size=11, replace= TRUE)
```

```
 [1]  5  9  3  2  3 10  1  9  1  6  6
```

> Q. Can you use **sample()** to generate a random nucleotide sequence of length 5?

```
sample(x= c("A", "T", "G","C"), size= 5, replace= TRUE)
```

```
[1] "C" "C" "T" "C" "T"
```

> Q. Write a function **generate_DNA()** that makes a nucleotide sequence of a user specified length.

Every function in R has at least 3 things:

- A **name** (in our case "generate_DNA")
- One or more **input arguments** (the length of sequence we want)
- A **body** (R code that does the work)

```
Generate_DNA <- function(length){
  sample(x= c("A", "T", "G","C"), replace= TRUE)
  size=length
}
```

```
Generate_DNA(10)
```

> Can you write a **generate_protein()** function that returns amino acid seqeunce of a user requested length?

2

```
generate_protein <-function(length){
aa <- bio3d::aa.table$aa1[1:20]
s <- sample(aa, size=length, replace=TRUE)
  paste(s, collapse="")
}
```

I want my output of this function not to be a vector with once amino acid per element but rather a single string.

```
generate_protein(20)
```

```
[1] "RFEADICQSCKSNHEWHNVP"
```

Q. Generate protein sequence from length 6 to 12

We can use utility function `sapply()` to help us "apply" our function over all the values 6 to 12.

```
ans<- sapply(X=6:12, generate_protein)
```

```
cat( paste(">ID", 6:12, sep="", "\n", ans, "\n", collapse = ) )
```

```
>ID6
KMNPTT
 >ID7
SNTKDDS
 >ID8
FYDNKRGI
 >ID9
CMPNCNMGG
 >ID10
NGFSNCPAME
 >ID11
KFTVEFDTTRC
 >ID12
NMQQMCGLYPHK
```

Are any of these sequences unique in nature - i.e. never found in nature. We can search "refseq-protein" and look for 100% Ide and 100% coverage

None of the sequences generated are unique in nature. All of the 7 sequences provided had match sequences from blast that had 100% coverage and 100% Id values.