Class 6: R functions

Berne Chu (A18608434)

All functions in R have at least 3 things:

- A **name**, we pick this and use it to call the function.
- Input **arguments**, there can be multiple comma sperated inputs to the function.
- The **body**, lines of R code that do the work

Our first wee function:

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

let's test our function

```
add(c(1,2,3), 120)
```

```
[1] 121 122 123
```

```
add(10)
```

```
[1] 11
```

```
add(10,100)
```

```
[1] 110
```

A second function

Let's try somthing more interesting. Make a sequence generation tool.

The `sample()` function could be useful here

```
sample(1:10, size=3)
```

```
[1] 3 2 9
```

change this to work with the nucleotides A C G and T and return 3 of them

```
n <- c("A", "C", "G", "T")
sample(n, size=5, replace = T)
```

```
[1] "G" "G" "A" "C" "T"
```

Turn this snipet into a function that returns a user specifed length dna sequence. Let's call it `generate_dna()` ...

```
generate_dna <- function(len=10,fasta=F){
  n <- c("A", "C", "G", "T")
  v <- sample(n, size=len, replace = T)

  # Make a single element vector
  s <- paste(v,collapse = "")

  cat("Well done you!\n")
  if(fasta){
    return(s)
  }
  else{
    return(v)
  }
}
```

```
generate_dna(5)
```

```
Well done you!
```

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

```
s <- generate_dna(10,fasta=F)
```

```
Well done you!
```

```
s
```

```
 [1] "C" "C" "G" "G" "C" "C" "G" "T" "G" "A"
```

I want to the option to return a single element character vector with my sequence all together like this: "GGAGTAC"

```
s
```

```
 [1] "C" "C" "G" "G" "C" "C" "G" "T" "G" "A"
```

```
paste(s,collapse = "")
```

```
[1] "CCGGCCGTGA"
```

A more advance example

Make a third function the generates protein sequence of user specified length and format

```
generate_protein <- function(len=15,fasta=T){
  aa <- c(
  "A", "R", "N", "D", "C", "E", "Q", "G",
  "H", "I", "L", "K", "M", "F", "P", "S",
  "T", "W", "Y", "V")
  seq <- sample(aa, size=len, replace=T)
  if(fasta){
    return(paste(seq, collapse=""))
  }
  else{
    return(seq)
  }
}
```

Try this out...

```
generate_protein(10)
```

```
[1] "PMTCCICHML"
```

> Q. Please generate random protein sequences between lengths 5 and 12 amino acids.

One approach is to do this by brute force callin gour function for each length 5 to 12.

Another approach is to write a `for()` loop to itterate over the input valued 5 to 12.

A very useful third R specific approach is to use the `sapply()` function

```
seq_lengths <- 5:12
for(i in seq_lengths){
  cat(">", i,"\n")
  cat(generate_protein(i))
  cat("\n")
}
```

```
> 5
RKTKS
> 6
THDLWG
> 7
WINDGGN
> 8
GLIFQIMG
> 9
WNHWDTYAR
> 10
WMYRGQSCTR
> 11
EDRQHALKLMS
> 12
HGDTQMKWTERV
```

```
sapply(5:12, generate_protein)
```

```
[1] "HNWDP"        "DSEMAI"       "NKSHQWN"      "CPTPECGQ"     "MYLSVQHES"
[6] "SFRTCVLNML"   "NPYCWRFPDTI"  "IVLCEFWSLNMK"
```

Key-Point: Writing functions in R is doble but not the eaisest thing. Starting with a working snippet of code and then using LLM tools to improve and generalize your function code is a productive approach.