# HW LE 6

AUTHOR
Nundini Varshney (PID: A16867985)

## Section 1: Improving analysis code by writing functions

```
# (A. Can you improve this analysis code?
 df <- data.frame(a=1:10, b=seq(200,400,length=10),c=11:20,d=NA)
df
```

```
   a        b  c  d
1  1 200.0000 11 NA
2  2 222.2222 12 NA
3  3 244.4444 13 NA
4  4 266.6667 14 NA
5  5 288.8889 15 NA
6  6 311.1111 16 NA
7  7 333.3333 17 NA
8  8 355.5556 18 NA
9  9 377.7778 19 NA
10 10 400.0000 20 NA
```

```
df$a <- (df$a - min(df$a)) / (max(df$a) - min(df$a))
 df$b <- (df$b - min(df$a)) / (max(df$b) - min(df$b))
 df$c <- (df$c - min(df$c)) / (max(df$c) - min(df$c))
 df$d <- (df$d - min(df$d)) / (max(df$d) - min(df$d))
 df
```

```
          a        b         c  d
1  0.0000000 1.000000 0.0000000 NA
2  0.1111111 1.111111 0.1111111 NA
3  0.2222222 1.222222 0.2222222 NA
4  0.3333333 1.333333 0.3333333 NA
5  0.4444444 1.444444 0.4444444 NA
6  0.5555556 1.555556 0.5555556 NA
7  0.6666667 1.666667 0.6666667 NA
8  0.7777778 1.777778 0.7777778 NA
9  0.8888889 1.888889 0.8888889 NA
10 1.0000000 2.000000 1.0000000 NA
```

```
library("bio3d")
 # Can you improve this analysis code?
 s1 <- read.pdb("4AKE")  # kinase with drug
```

```
  Note: Accessing on-line PDB file
```

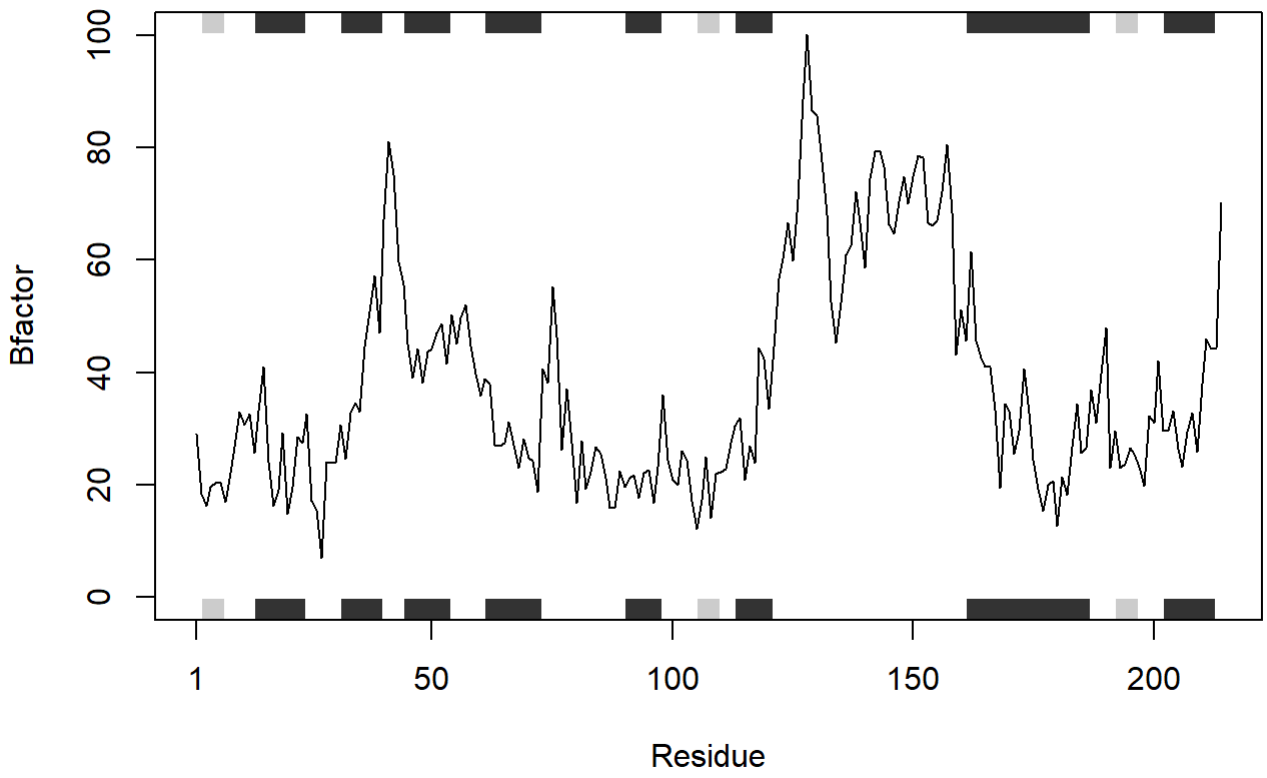
```
s2 <- read.pdb("1AKE")  # kinase no drug
```

```
Note: Accessing on-line PDB file
 PDB has ALT records, taking A only, rm.alt=TRUE
```

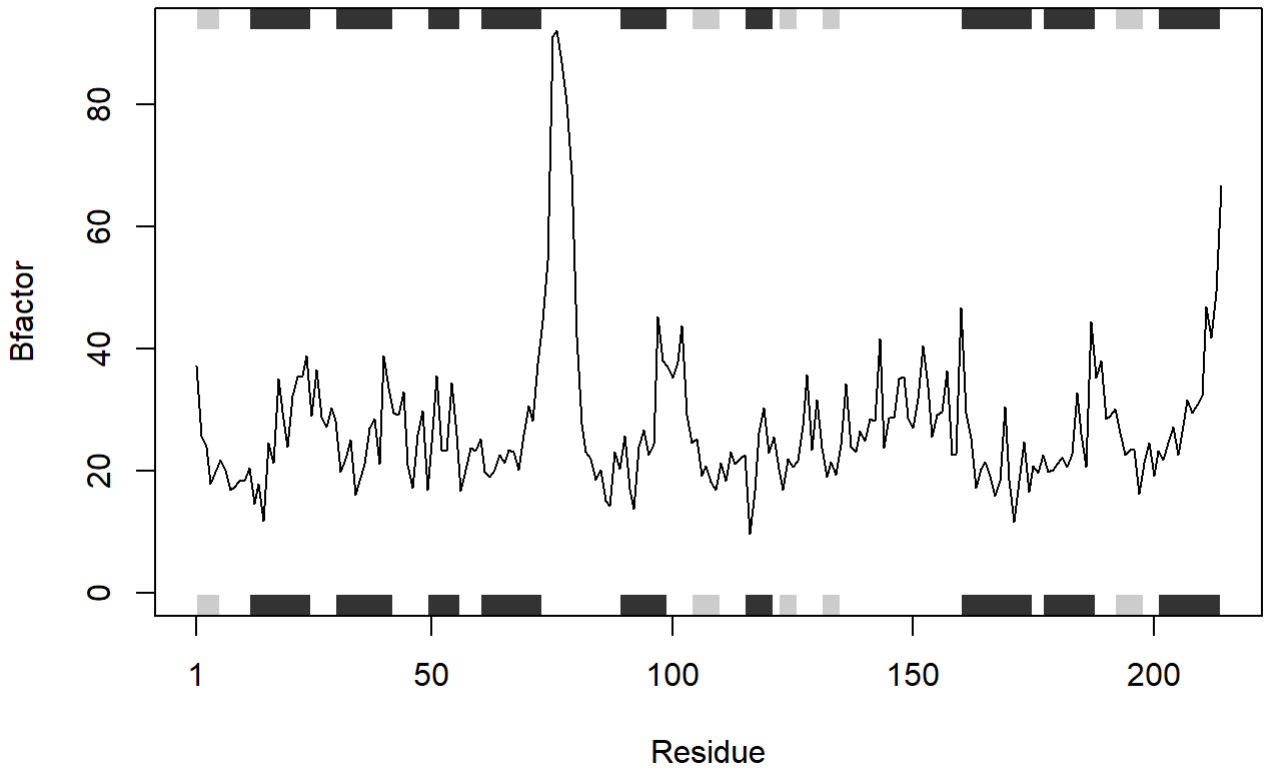
```
s3 <- read.pdb("1E4Y")  # kinase with drug
```

```
Note: Accessing on-line PDB file
```

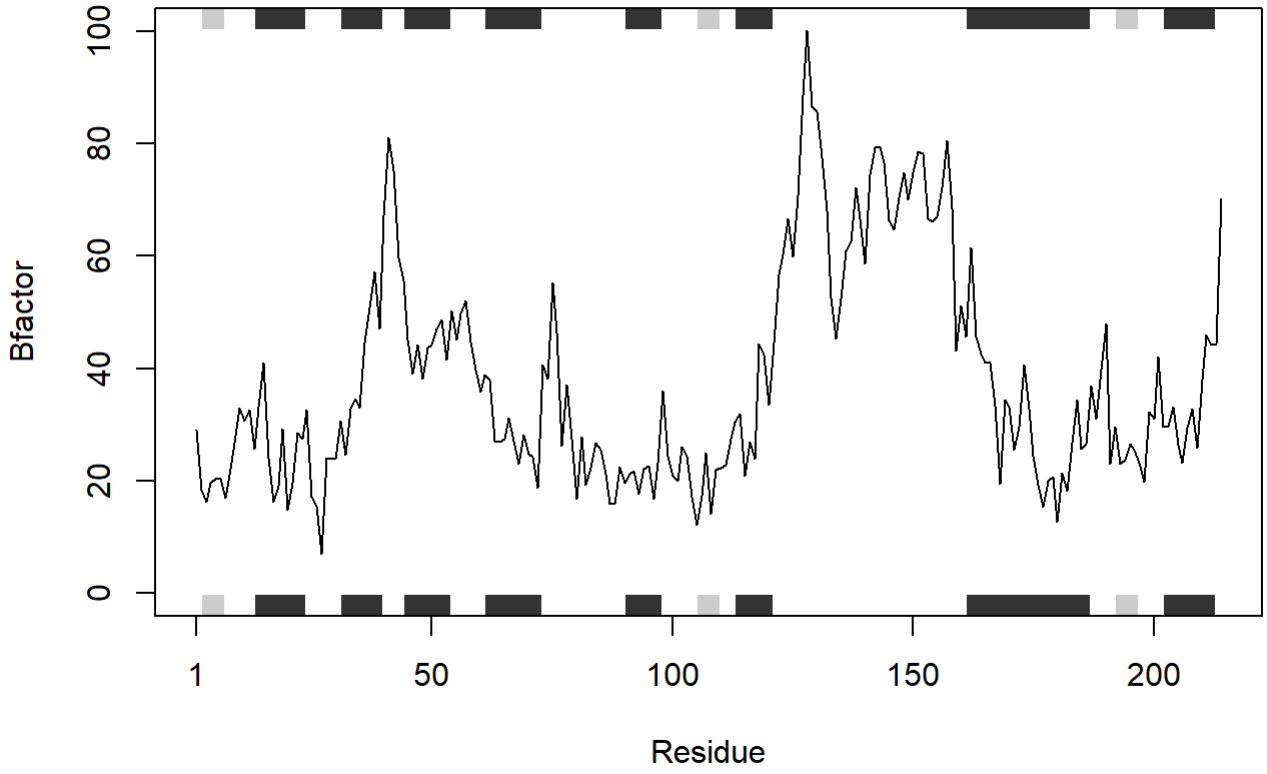
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s1, chain="A", elety="CA")
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```




```
plotb3(s2.b, sse=s2.chainA, typ="l", ylab="Bfactor")
```




```
plotb3(s3.b, sse=s3.chainA, typ="l", ylab="Bfactor")
```




> Q. 1

This function reads the protein data.

> Q. 2

Shortens it to the proteins needed for the analysis.

> Q. 3

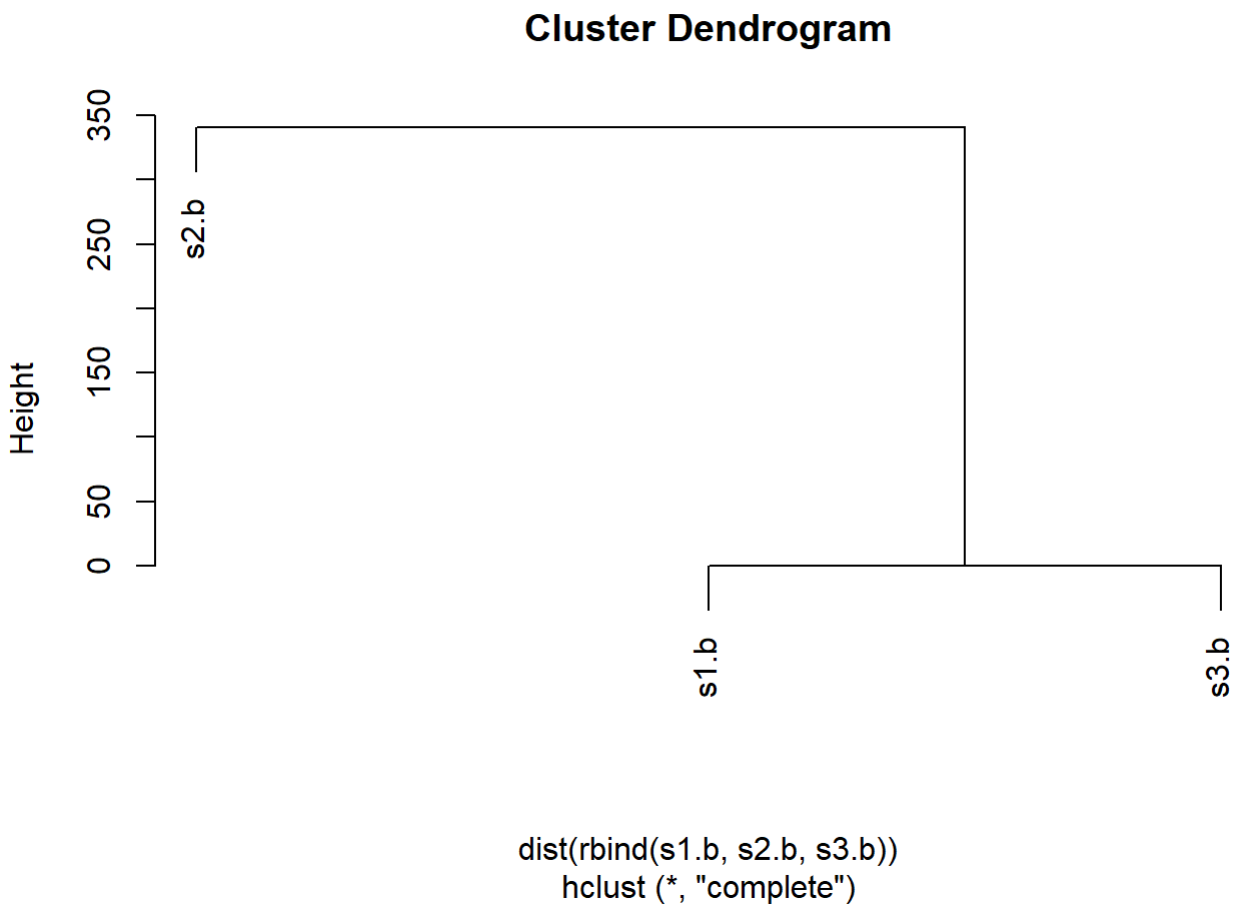
sse feature removes the black and grey rectangles, which represents the secondary structures

> Q. 4

They all look very similar to me functionally but the first plot looks more clear with the peaks.

> Q. 5

```
hc <- hclust( dist( rbind(s1.b, s2.b, s3.b) ) )
plot(hc)
```




s1 and s3 are closest together

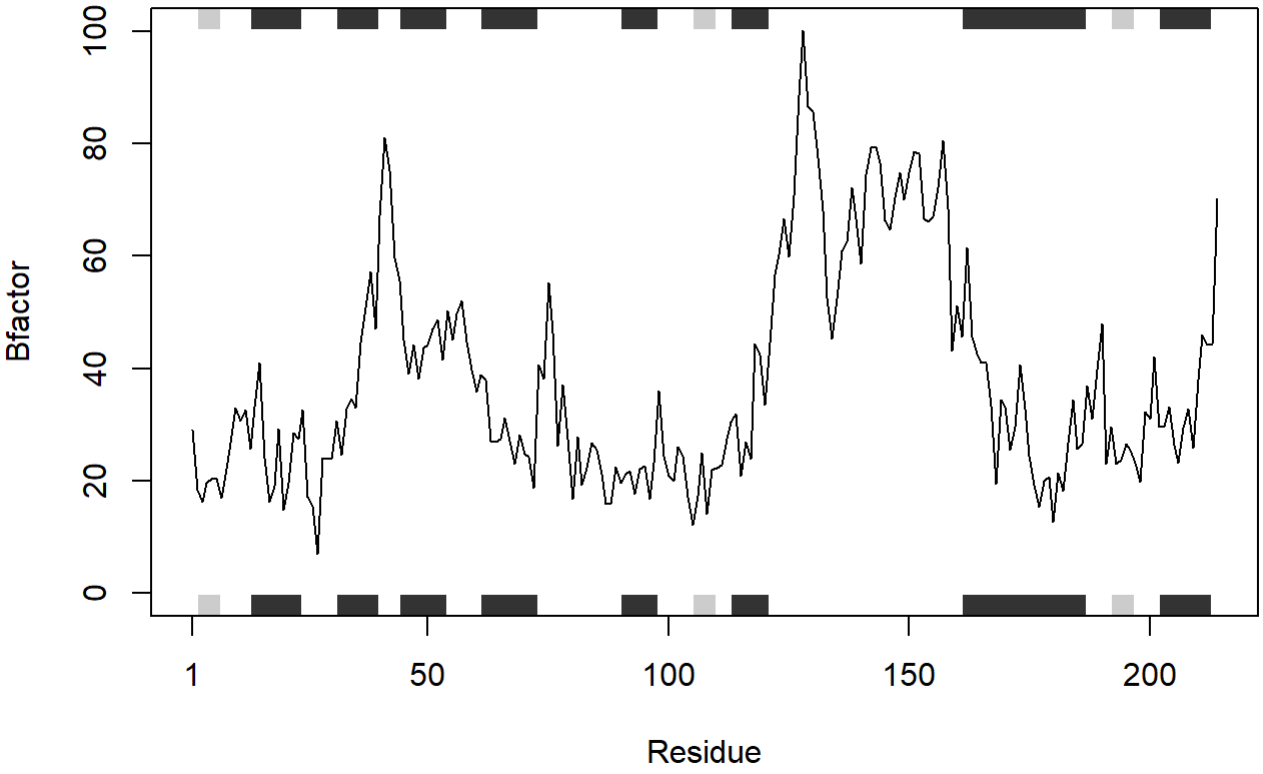
> Q6. How would you generalize the original code above to work with any set of input protein structures?

```
pr <- function(x) {
  y <- read.pdb(x)
  y.chainA <-trim.pdb(y, chain="A", elety="CA")
  y.b <- y.chainA$atom$b
  plotb3(y.b, sse=y.chainA, typ="l", ylab="Bfactor")
}

pr("4AKE")
```

```
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\Juhi\AppData\Local\Temp\Rtmp2x6jl9/4AKE.pdb exists. Skipping download
```


Bfactor
0
20
40
60
80
100
1
50
100
150
200
Residue