homework06

Xueran Zou

4/23/23

A. Improve the analysis code below.

```
df <- data.frame(a=1:10, b=seq(200,400,length=10),c=11:20,d=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$a) - min(df$d))
```

Write the function.

```
function_1 = function(x){
  (x - min(x)) / (max(x) - min(x))
}
```

Apply the function:

```
function_1(df$a)
```

```
 [1] 0.0000000 0.1111111 0.2222222 0.3333333 0.4444444 0.5555556 0.6666667
 [8] 0.7777778 0.8888889 1.0000000
```

B. Improve the below example code for the analysis of protein drug interactions.

```
library(bio3d)
```

```
s1 <- read.pdb("4AKE") # kinase with drug
```

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

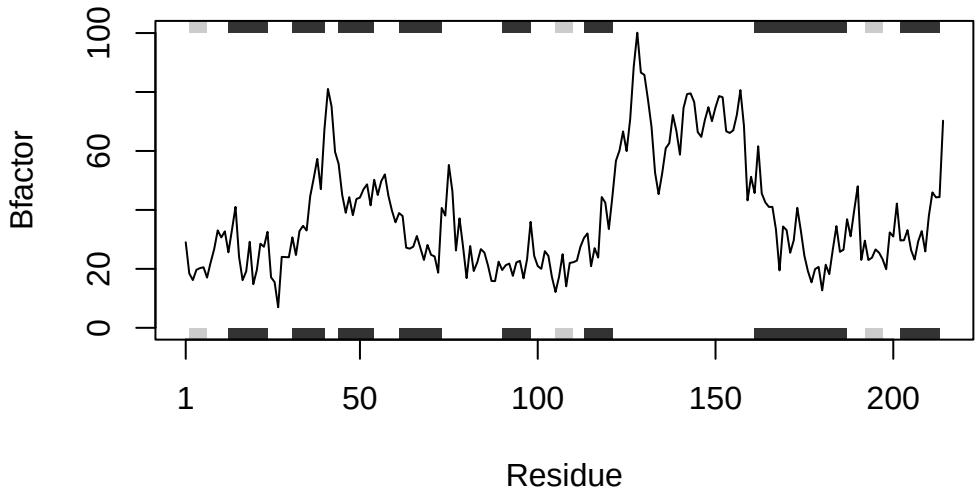
```
s2 <- read.pdb("1AKE") # kinase no drug
```

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

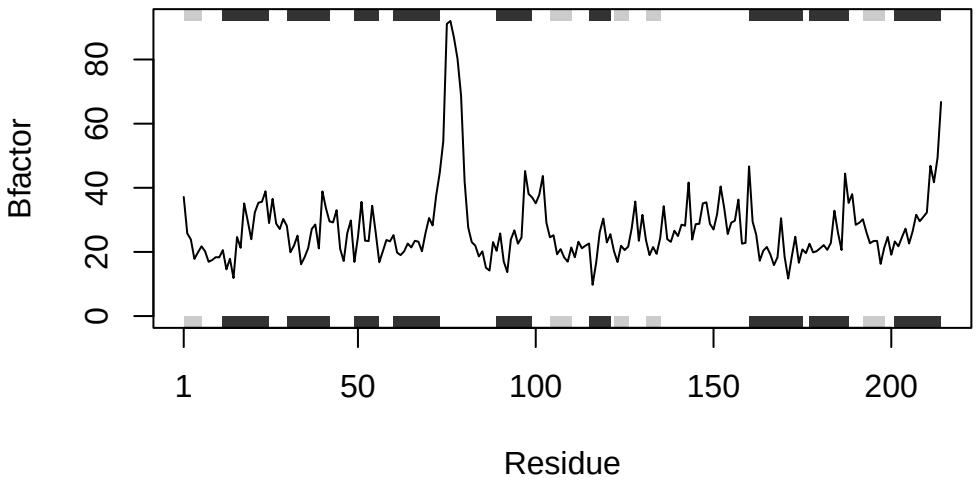
```
s3 <- read.pdb("1E4Y") # kinase with drug
```

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

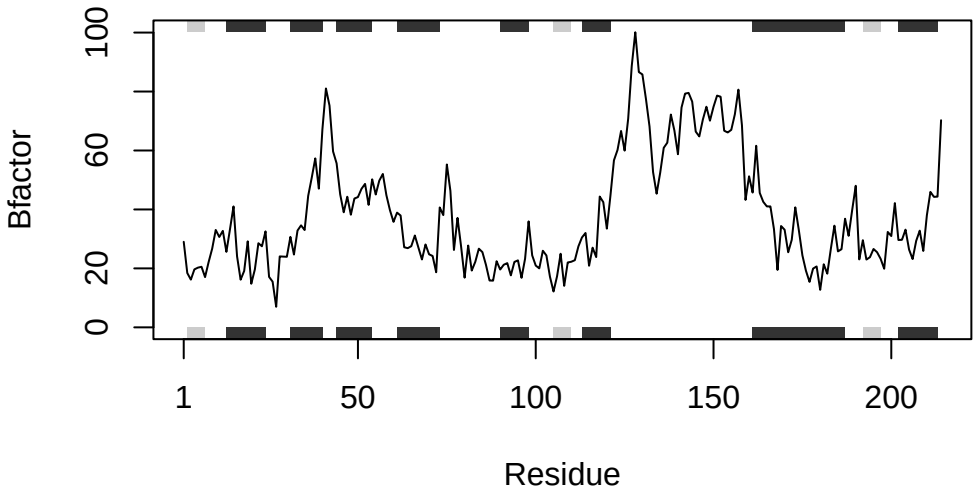
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s1, chain="A", elety="CA")
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```



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




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



Write the function.

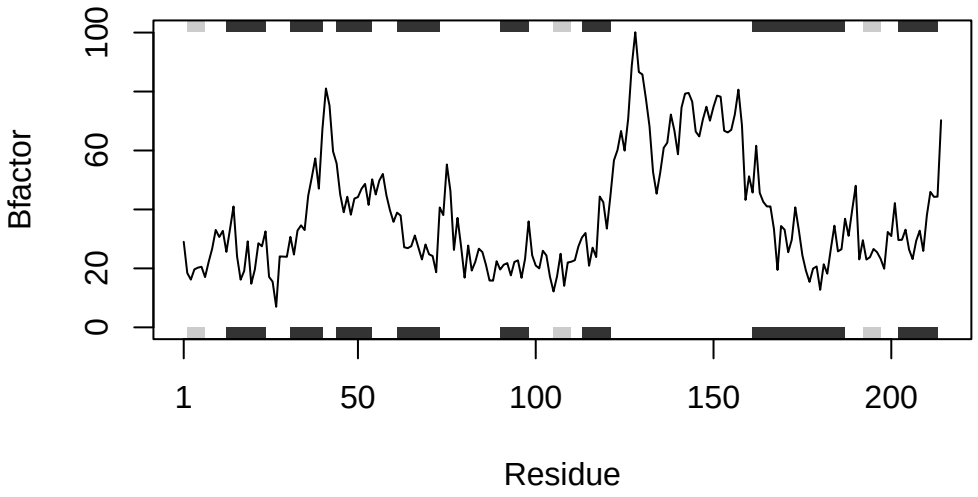
```
function_2 <- function(x){
  #input: PDB
  s <- read.pdb(x)
  s.chainA <- trim.pdb(s, chain="A", elety="CA")
  s.b <- s.chainA$atom$b
  plotb3(s.b, sse=s.chainA, typ="l", ylab="Bfactor")
  #output: a standard scatter plot with optional secondary structure in the marginal regio
}
```

Apply function:

```
function_2("4AKE")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\Lenovo\AppData\Local\Temp\RtmpYnJmt3/4AKE.pdb exists. Skipping
download
```



Q1. What type of object is returned from the read.pdb() function?

```
class(s1)
```

```
[1] "pdb" "sse"
```

The type of object returned from the read.pdb() function is large pdb.

Q2. What does the trim.pdb() function do?

The function is used to produce a new smaller PDB object, containing a subset of atoms, from a given larger PDB object.

Q3. What input parameter would turn off the marginal black and grey rectangles in the plots and what do they represent in this case?

```
s1 <- read.pdb("4AKE")
```

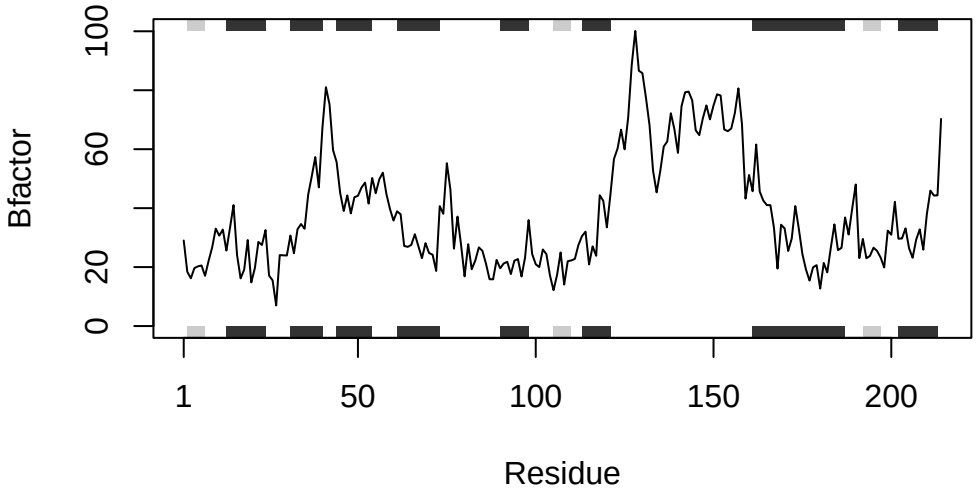
```
  Note: Accessing on-line PDB file
```

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\Lenovo\AppData\Local\Temp\RtmpYnJmt3/4AKE.pdb exists. Skipping
download
```

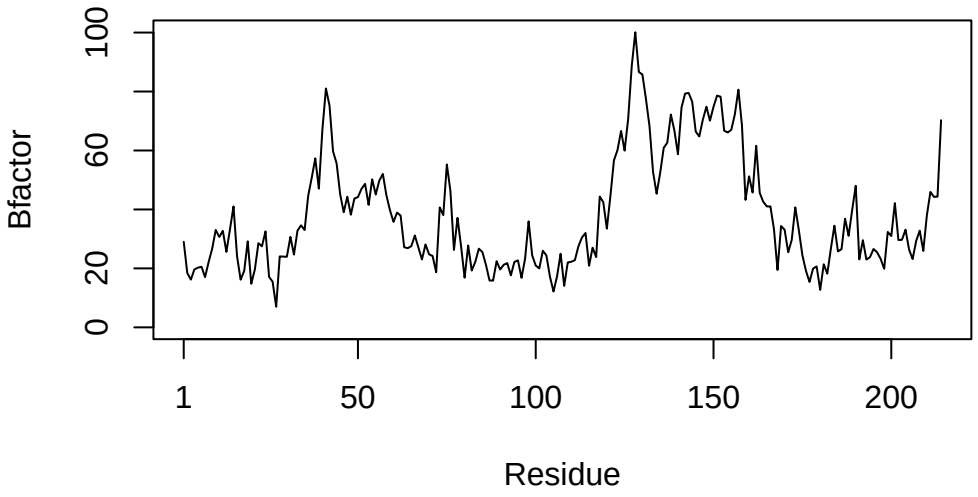
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
function_2("4AKE")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\Lenovo\AppData\Local\Temp\RtmpYnJmt3/4AKE.pdb exists. Skipping
download
```




```
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor", top=FALSE, bot=FALSE)
```

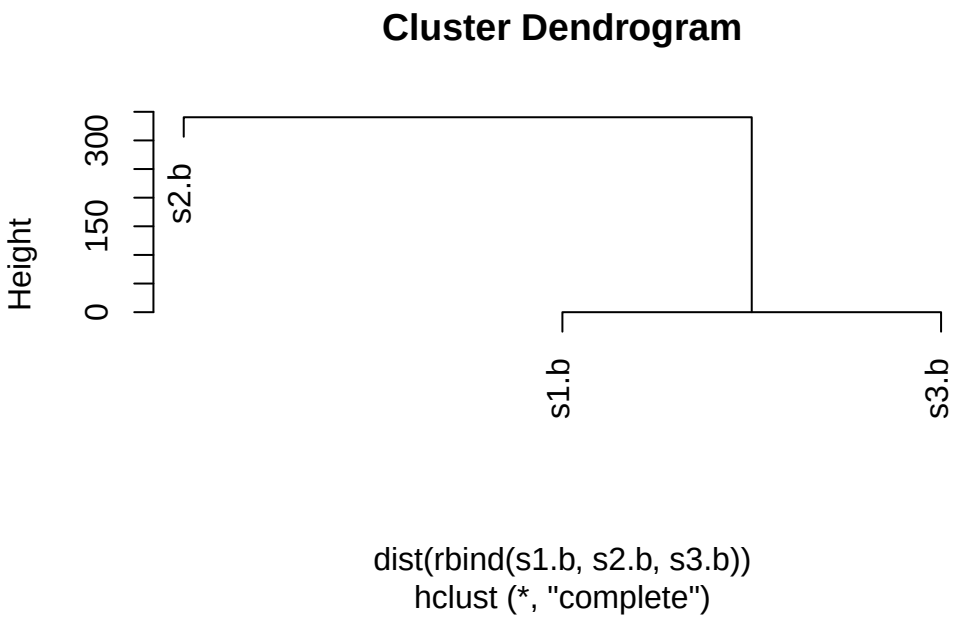


The input parameter "top=FALSE" and "bot=FALSE" would turn off the marginal black and grep rectangles in the plot.

In this case, the black ones represent alpha helices, and the grey ones represent beta strands.

Q4. What would be a better plot to compare across the different proteins?

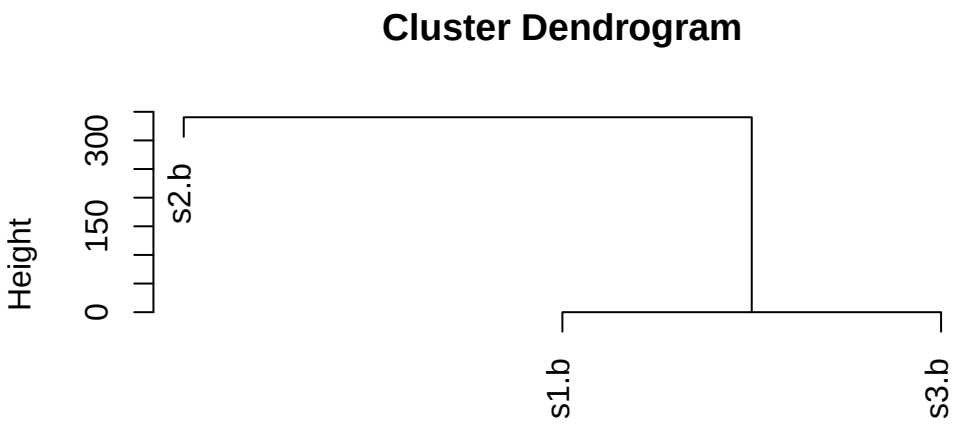
```
hc <- hclust(dist(rbind(s1.b, s2.b, s3.b)))
plot(hc)
```



The difference distance matrix can be a better plot to compare across the different proteins.

Q5. Which proteins are more similar to each other in their B-factor trends. How could you quantify this?

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



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

```
function_2 <- function(x){
  s <- read.pdb(x)
  s.chainA <- trim.pdb(s, chain="A", elety="CA")
  s.b <- s.chainA$atom$b
  plotb3(s.b, sse=s.chainA, typ="l", ylab="Bfactor")
}
```