HW6

Yufei Liu (A16222438)

Section 1: Improving analysis code by writing functions

A.

```
df <- data.frame(a=1:10, b=seq(200,400,length=10),c=11:20,d=NA)
Analyze <- function (x) {
  (x-min(x)) / (max(x)-min(x))
}

AnalyzeDf <- apply(df, 2, Analyze)
AnalyzeDf
```

```
              a         b         c  d
 [1,] 0.0000000 0.0000000 0.0000000 NA
 [2,] 0.1111111 0.1111111 0.1111111 NA
 [3,] 0.2222222 0.2222222 0.2222222 NA
 [4,] 0.3333333 0.3333333 0.3333333 NA
 [5,] 0.4444444 0.4444444 0.4444444 NA
 [6,] 0.5555556 0.5555556 0.5555556 NA
 [7,] 0.6666667 0.6666667 0.6666667 NA
 [8,] 0.7777778 0.7777778 0.7777778 NA
 [9,] 0.8888889 0.8888889 0.8888889 NA
[10,] 1.0000000 1.0000000 1.0000000 NA
```

B.

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

It is a pdb object that stores detailed structural informations of the protein.

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

It trims the pdb object to a smaller subset of atoms.

For instance, the trim.pdb(s, chain="A", elety="CA") command used above returns only atoms named "CA" in chain A

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

Can remove the sse argument to turn off the rectangle:

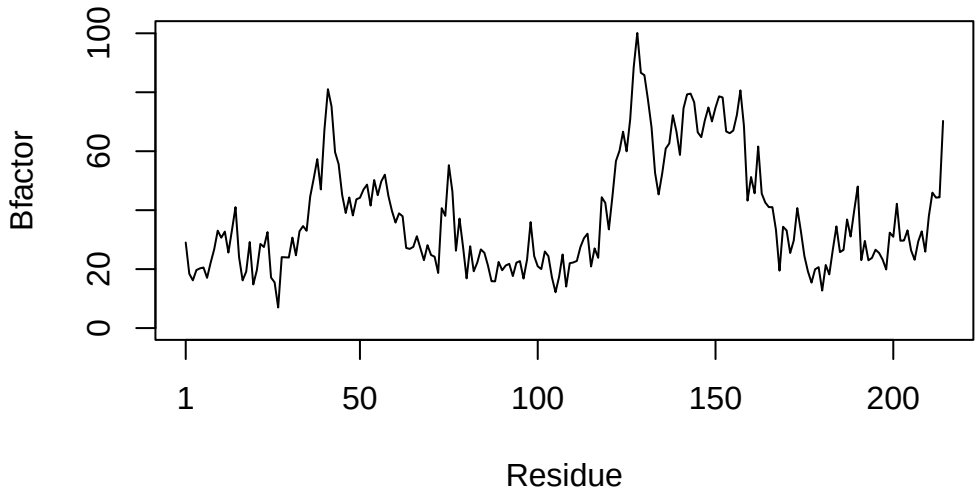
```
library(bio3d)
```

```
Warning: package 'bio3d' was built under R version 4.3.1
```

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

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

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



In this case, it highlights secondary structures from the trimmed chainA.

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

Combine them on the same graph:

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

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

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

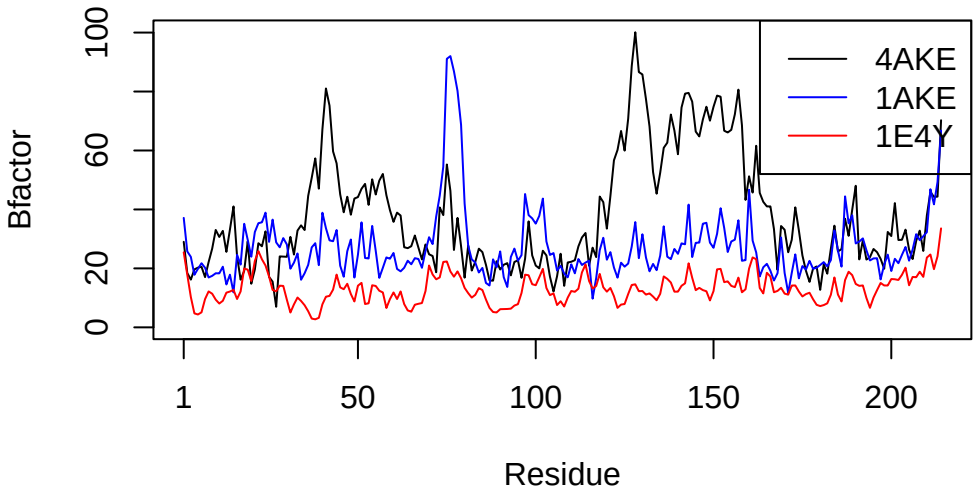
```
s2 <- read.pdb("1AKE") # kinase no drug
```

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

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

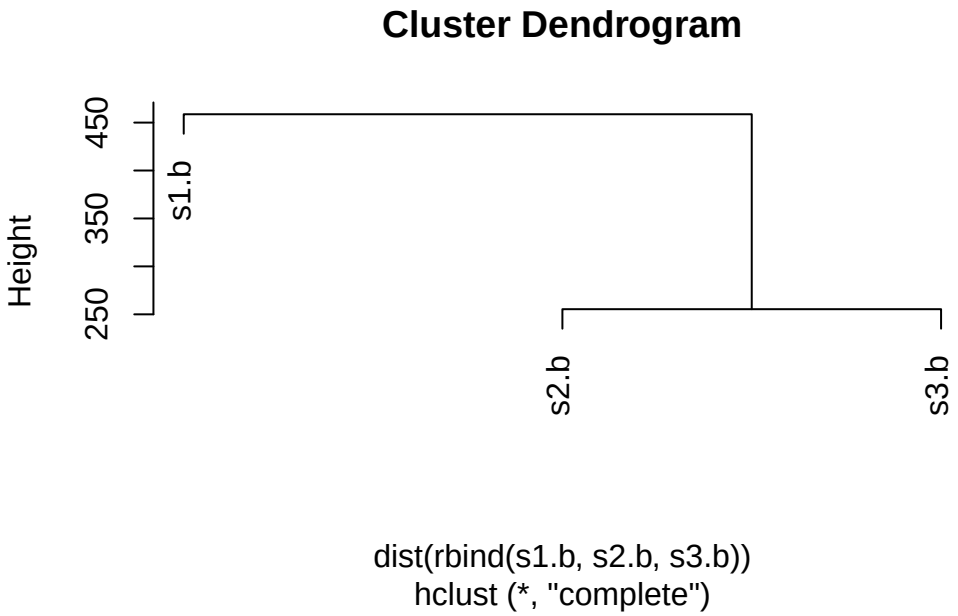
```
  Note: Accessing on-line PDB file
```

```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s3, 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, typ="l", ylab="Bfactor")
lines(s2.b, col="blue")
lines(s3.b, col="red")

legend("topright", legend=c("4AKE", "1AKE", "1E4Y"),
       col=c("black", "blue", "red"), lty=1)
```



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)
```



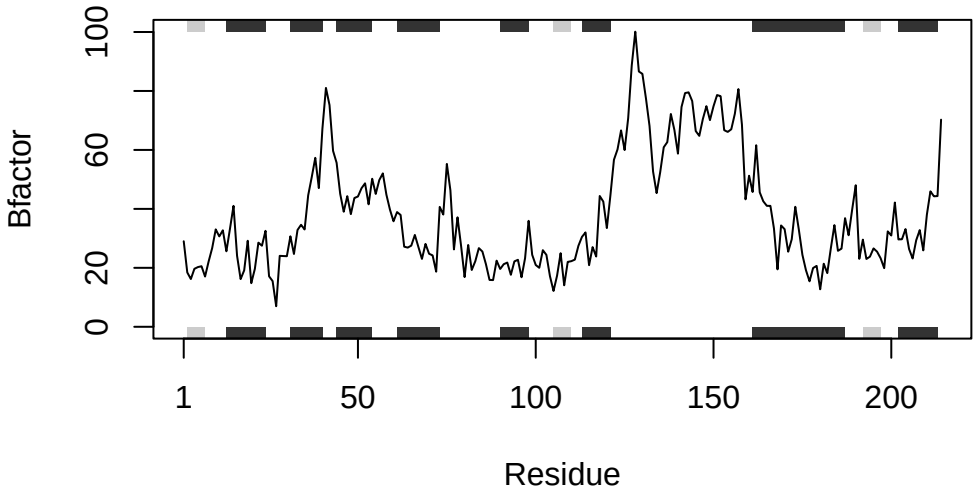
"1AKE" and "1E4Y" are more similar to each other.

Q6. Write your own function starting from the code above that analyzes protein drug interactions by reading in any protein PDB data and outputs a plot for the specified protein.

```
# library bio3d package to use its functions
library(bio3d)

# Function to analyze a protein with given PDB id, by plotting the Bfactor against residue
# input = protein PDB id
# output = plot of B factor against residue number.
# usage: analyzeP("PDB id of the target protein")
analyzeP <- function(id){
  s <- read.pdb(id)
  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")
}

# example
analyzeP("4AKE") # kinase with drug
```

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

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

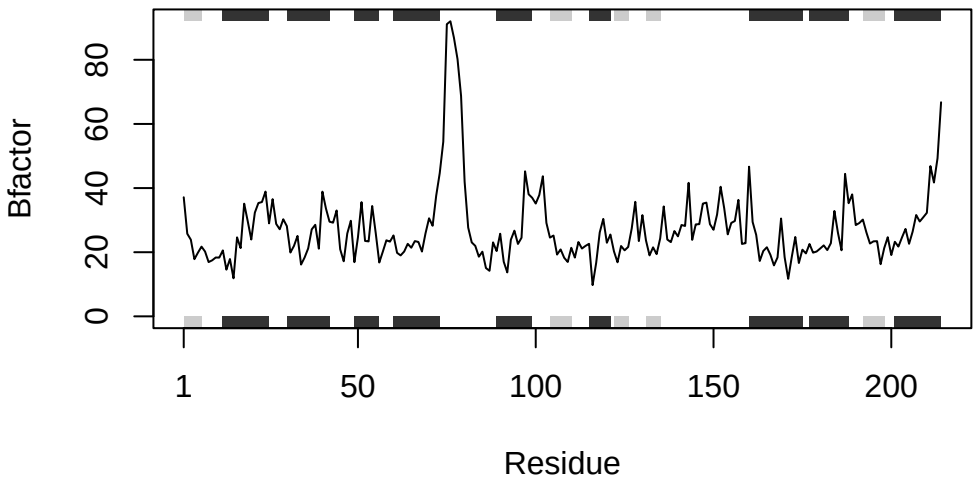



```
analyzeP("1AKE") # kinase no drug
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\liuyu\AppData\Local\Temp\RtmpmGVlTm/1AKE.pdb exists. Skipping download
```

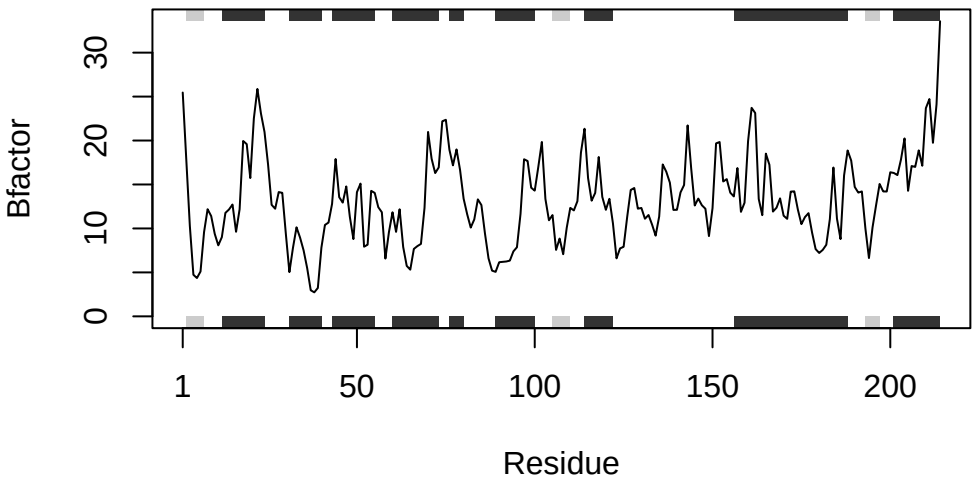
```
   PDB has ALT records, taking A only, rm.alt=TRUE
```




```
analyzeP("1E4Y") # kinase with drug
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\liuyu\AppData\Local\Temp\RtmpmGVlTm/1E4Y.pdb exists. Skipping download
```


Bfactor
0
10
20
30
1
50
100
150
200
Residue