R functions

Yinuo Song

4/23/23

Section 1: Improving analysis code by writing functions

A

```
# (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$a <- (df$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))
```

create function:

```
## input is the column of a data frame
nasty <- function(x){
     x <- (x - min(x)) / (max(x) - min(x))
}
##out put a new column
```

This function can be used on a vector and produce a ew vector.

```
nasty(df$a)
nasty(df$b)
nasty(df$c)
nasty(df$d)
```

```r
##testing
df$a
```

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

```r
df$b
```

```
[1] 1.000000 1.111111 1.222222 1.333333 1.444444 1.555556 1.666667 1.777778
[9] 1.888889 2.000000
```

```r
df$c
```

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

```r
df$d
```

```
[1] NA NA NA NA NA NA NA NA NA NA
```

B

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

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

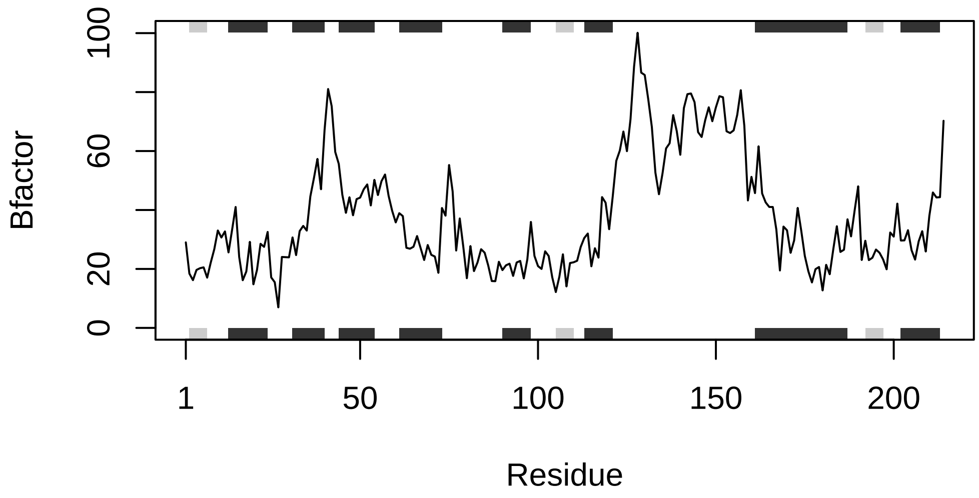
```r
s2 <- read.pdb("1AKE") # kinase no drug
```

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

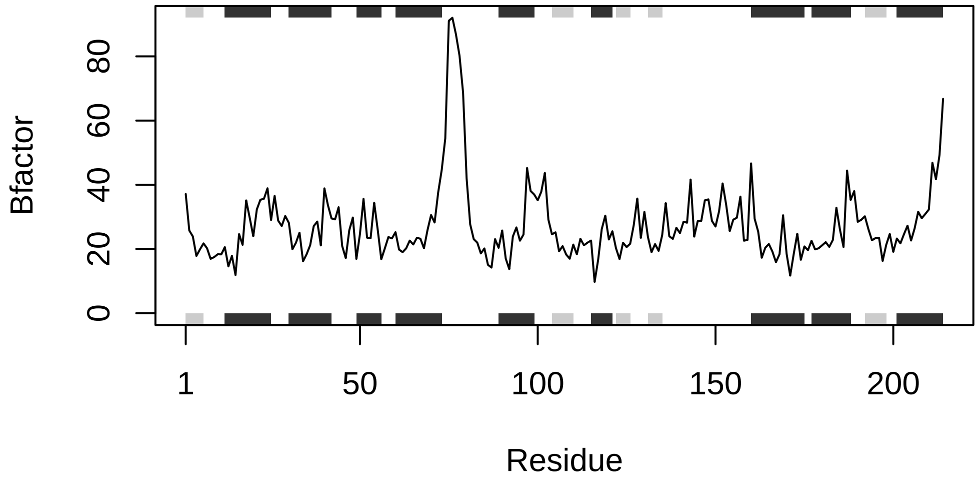
```
s3 <- read.pdb("1E4Y") # kinase with drug
```

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

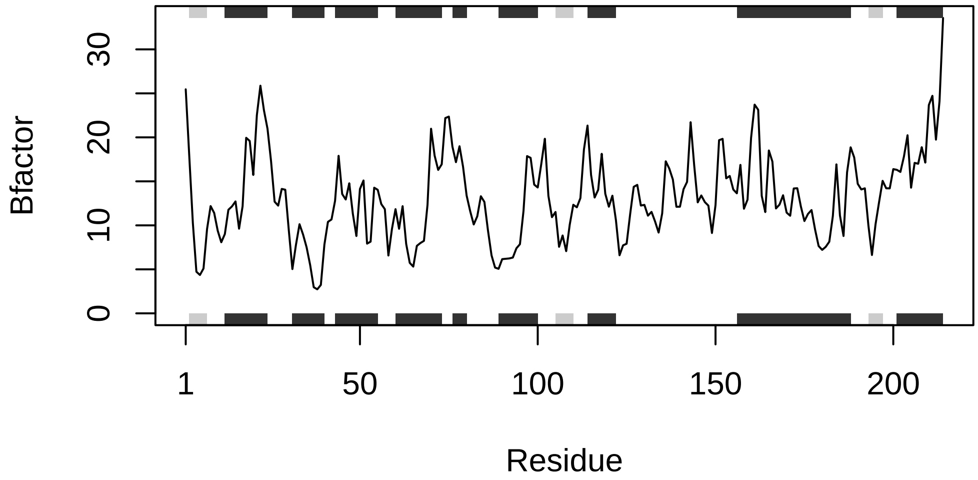
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s3, chain="A", elety="CA")
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```



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



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



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

```
typeof(s1)
```

```
[1] "list"
```

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

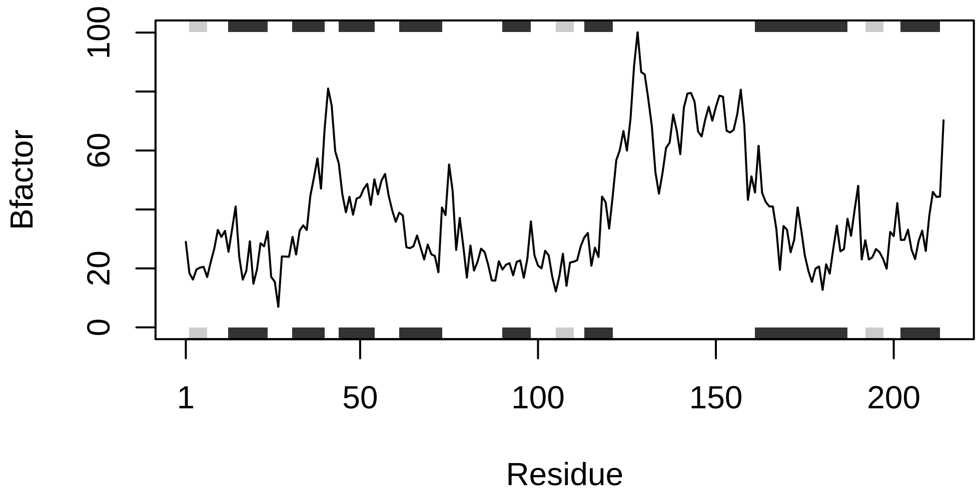
```
help(trim.pdb)
```

Produce a new smaller PDB object, containing a subset of atoms, from a given larger PDB object.

Create a new function:

```
## input is the list object of protein
pdbplot <- function(x){
     ## create chainA for the secondary structure
     x.chainA <- trim.pdb(s1, chain="A", elety="CA")
     ##create x.b for the x-y value
     x.b <- x.chainA$atom$b
     ## use plotb3 to plot
     plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
}
## ouput is a standard scatter plot with secondary structure in the marginal region.
```

```
pdbplot(s1)
```



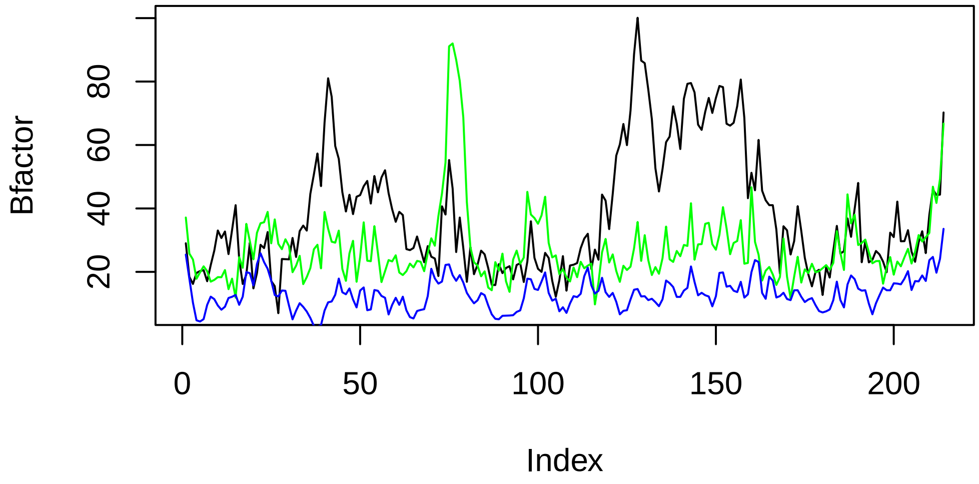
Q3. What input parameter would turn off the marginal black and grey rectangles in the plots and what do they represent in this case?

```
help(plotb3)
```

sse is the parameter that would turn off the marginal black and grey rectangles in the plots. and in this case they represent chainA data.

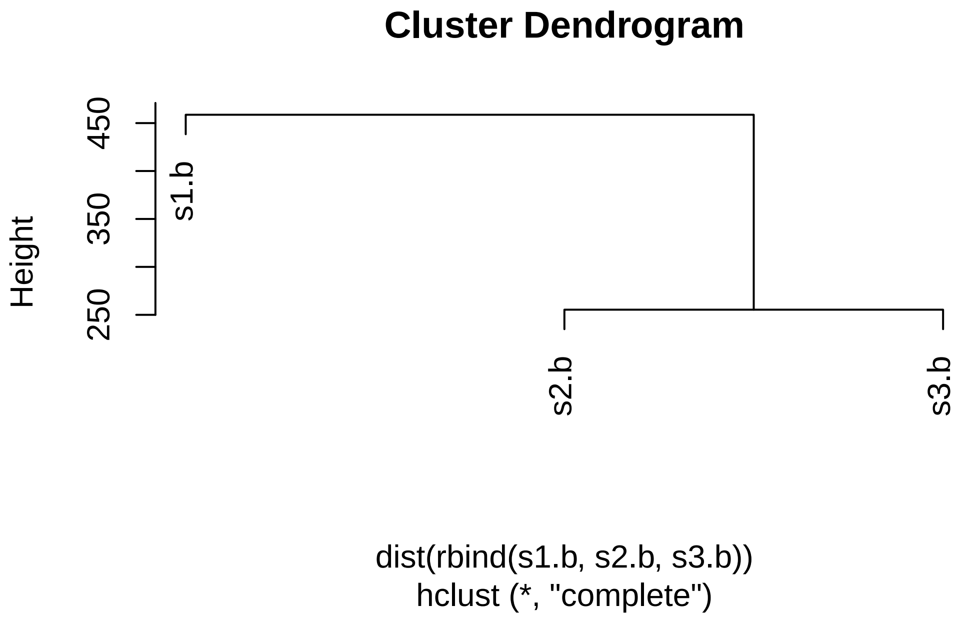
Q4. What would be a better plot to compare across the different proteins?

```
plot(s1.b, typ="l", ylab="Bfactor")
lines(s2.b,typ="l",col="green")
 lines(s3.b,,typ="l",col="blue")
```



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(s2) and s3(1E4Y) are more similar to each other.