HW6

Jae Kim

R Functions

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

```
NewFunction <- function(x){(x-min(x)) / (max(x)-min(x))}
NewFunction(df$a)
```

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

Improve the below example code for the analysis of protein drug interactions by abstracting the main activities in your own new function.

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

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

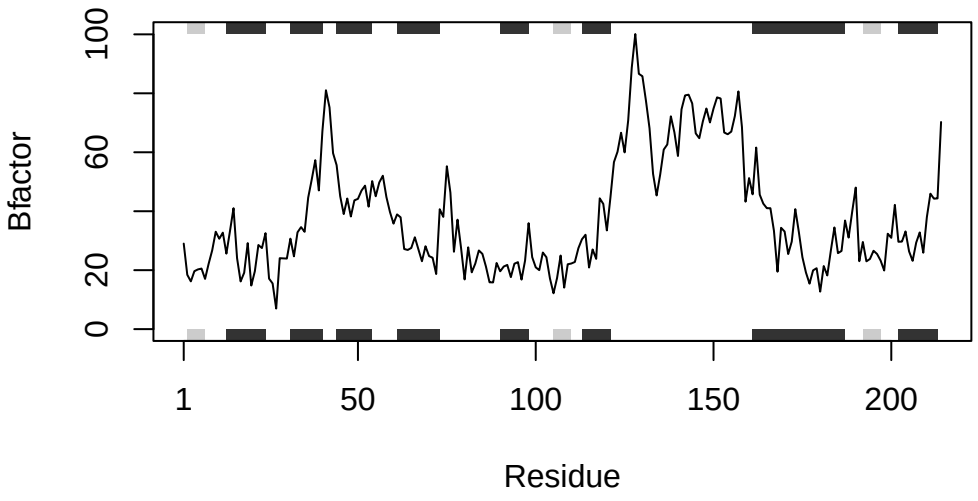
```
s2 <- read.pdb("1AKE")  # kinase no drug
```

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

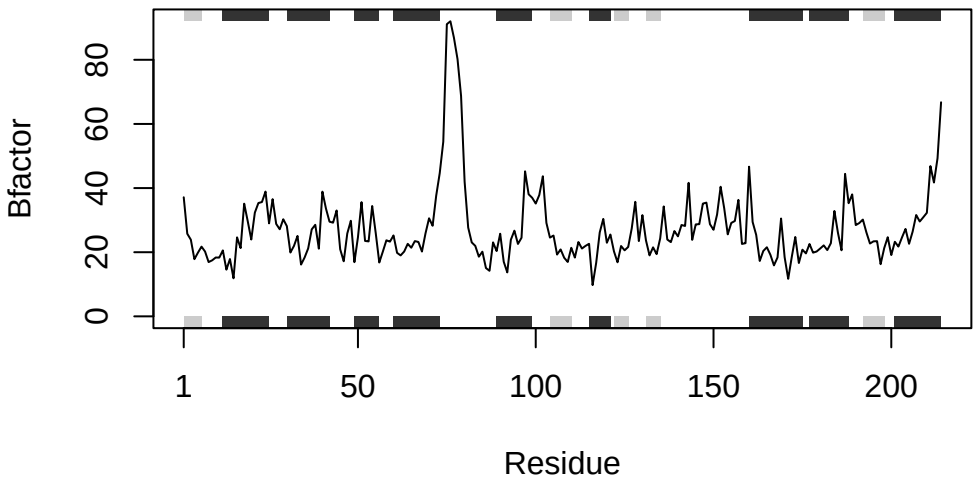
```
s3 <- read.pdb("1E4Y")  # kinase with drug
```

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

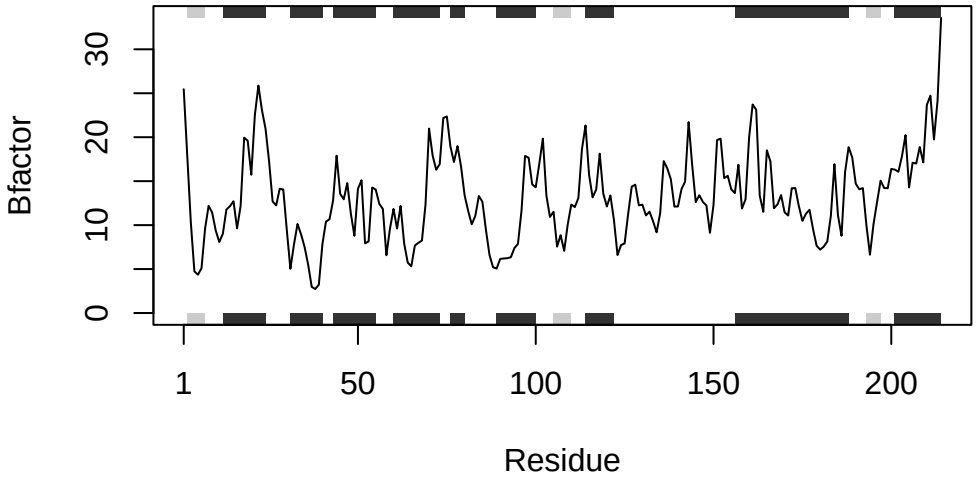
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s3, chain="A", elety="CA")

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")
```



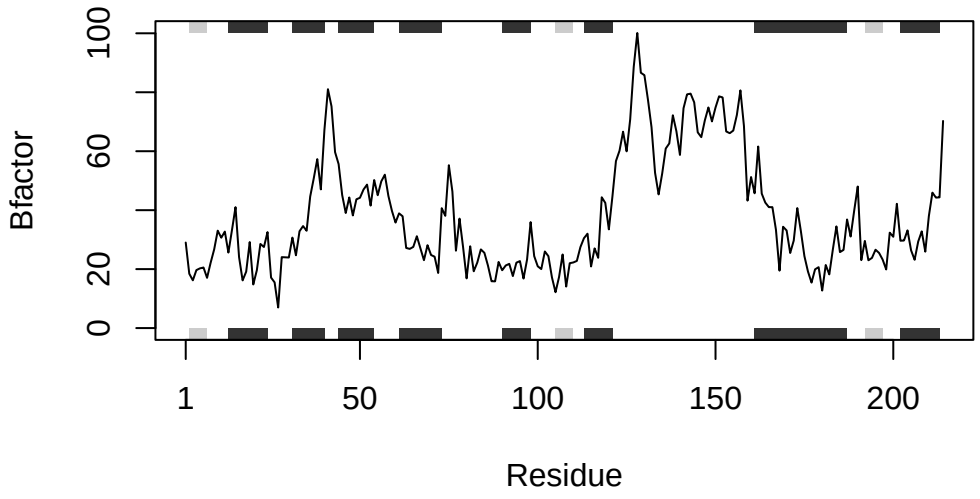
Q6. How would you generalize the original code above to work with any set of input protein structures?

```
# To read protein structure data in PDB format; used to load the protein as well.
a <- read.pdb("4AKE")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\kjj01\AppData\Local\Temp\RtmpKKM69q/4AKE.pdb exists. Skipping download
```

```
# It returns a list of atom coordinates, atom names, residue numbers, chain IDs, and etc.

# To produce a smaller PDB object, containing a subset of atoms, from a given larger PDB o
c <- trim.pdb(a, chain="A", elety="CA")

# To obtain B-factor for each residue of trimmed PDB object
b <- c$atom$b

# To obtain a plot of Bfactor vs residue
plotb3(b, sse=c, typ="l", ylab="Bfactor")
```



```
# function:
protein_plot <- function(x){
  a <- read.pdb(x)
  c <- trim.pdb(a, chain="A", elety="CA")
  b <- c$atom$b
  plotb3(b, sse=c, typ="l", ylab="Bfactor")
}
```

```
# Input any protein PDB data
# Outputs a plot for the specified protein
protein_plot("4AKE")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\kjj01\AppData\Local\Temp\RtmpKKM69q/4AKE.pdb exists. Skipping download
```