06_Class6_HW

Jacklyn Jung

2024-01-28

R 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$b - min(df$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))
df$a
```

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

```
df$b
```

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

```
df$c
```

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

```
df$d
```

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

```r
df <- data.frame(a=1:10, b=seq(200,400,length=10),c=11:20,d=NA)
normalize_column <- function(column) {
  return((column - min(column, na.rm = TRUE)) / (max(column, na.rm=TRUE) - min(column, na.
}
```

```r
normalize_column(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
normalize_column(df$b)
```

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

```r
normalize_column(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
normalize_column(df$d)
```

```
Warning in min(column, na.rm = TRUE): no non-missing arguments to min;
returning Inf

Warning in max(column, na.rm = TRUE): no non-missing arguments to max;
returning -Inf

Warning in min(column, na.rm = TRUE): no non-missing arguments to min;
returning Inf

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

```r
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
```

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

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/tl/t15g9j5d71b2pn77bpfjy59r0000gn/T//RtmpC3lswO/4AKE.pdb exists.
Skipping download
```

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

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/tl/t15g9j5d71b2pn77bpfjy59r0000gn/T//RtmpC3lswO/1AKE.pdb exists.
Skipping download
```

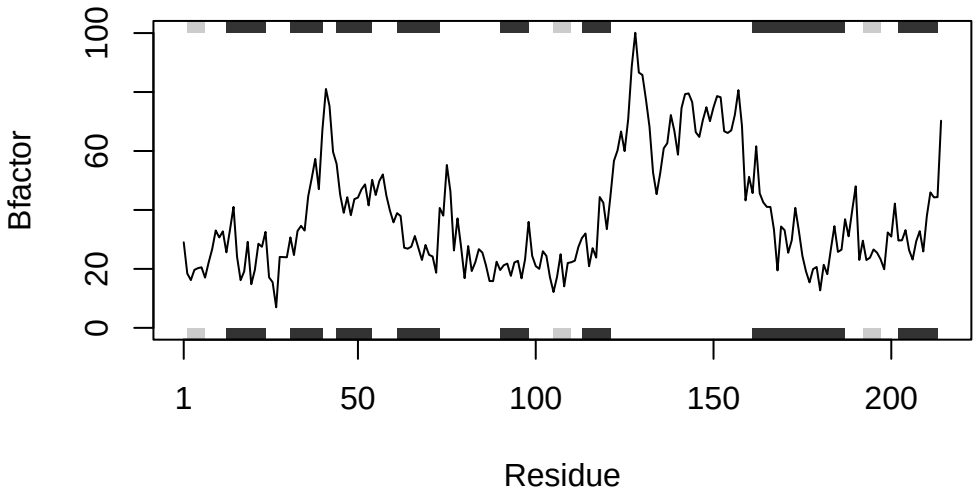
```
   PDB has ALT records, taking A only, rm.alt=TRUE
```

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

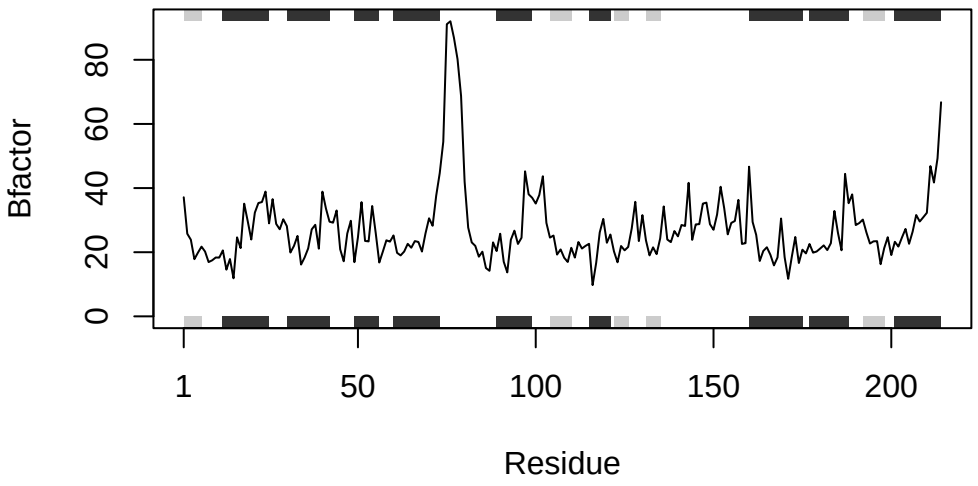
```
  Note: Accessing on-line PDB file
```

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/tl/t15g9j5d71b2pn77bpfjy59r0000gn/T//RtmpC3lswO/1E4Y.pdb exists.
Skipping download
```

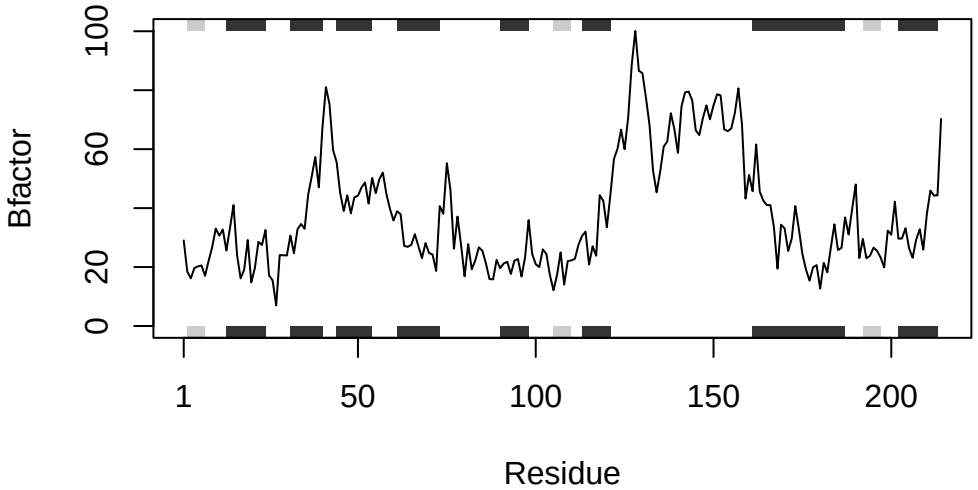
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s1, chain="A", elety="CA")
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```




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




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



“‘

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

- The read.pdb() function returns class of protein structure data in PDB format (three-dimensional structures)

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

- The trim.pdb() function helps to specify protein data structure that is returned.

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

- We would turn off the margin input parameter to remove the marginal black and grey rectangles in the plot. We can add the mar=FALSE argument to the function.

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

- ggplot2 might be a better plot to compare across different proteins with more aesthetically pleasing features.

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