HW6

Ziyuan_Han

10/16/2021

##A Can you improve this analysis code? ####Input: a dataframe, the minimum value and maxium value of the dataframe in each column or each row (depending on the apply function) ####Use the function to normalize the data by calculating through each column, the output will be a matrix with the normalized values

```
df <- data.frame(a=1:10, b=seq(200,400,length=10),c=11:20,d=NA)
functionA <-function(x){
  x <- (x-min(x)) / (max(x) - min(x))
}
apply(df,2,functionA)
```

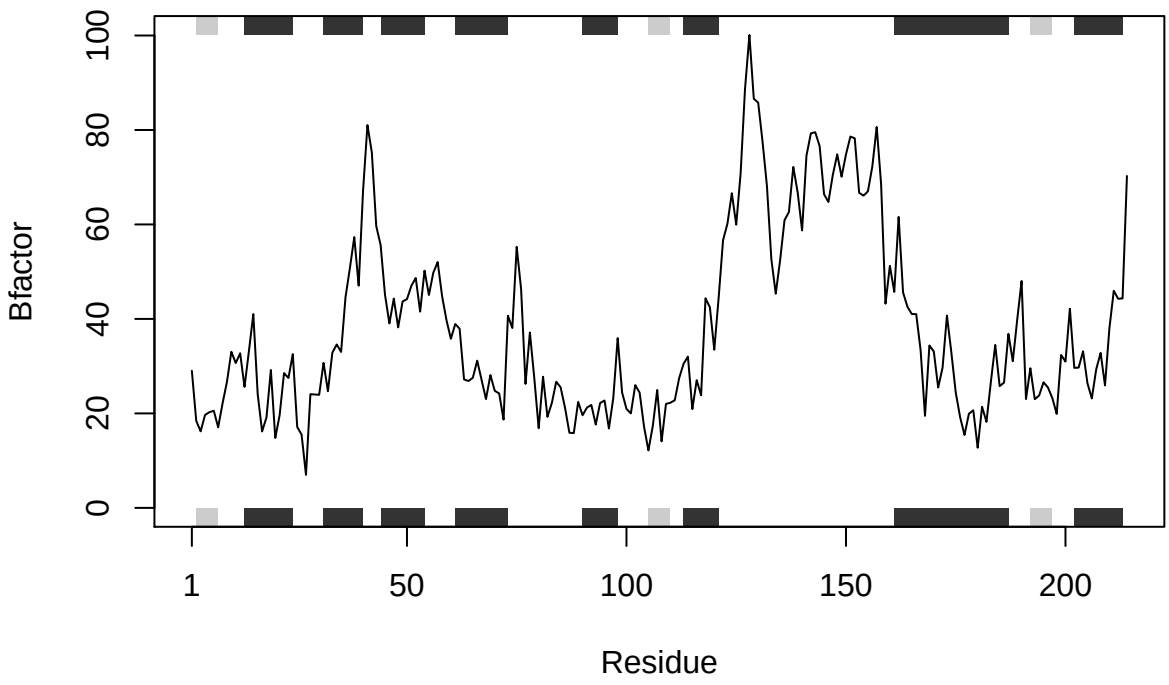
```
##               a         b         c  d
##  [1,] 0.0000000 0.0000000 0.0000000 NA
##  [2,] 0.1111111 0.1111111 0.1111111 NA
##  [3,] 0.2222222 0.2222222 0.2222222 NA
##  [4,] 0.3333333 0.3333333 0.3333333 NA
##  [5,] 0.4444444 0.4444444 0.4444444 NA
##  [6,] 0.5555556 0.5555556 0.5555556 NA
##  [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 Improve the code for protein analysis ###Input of the functions will be the PDB code for the objects

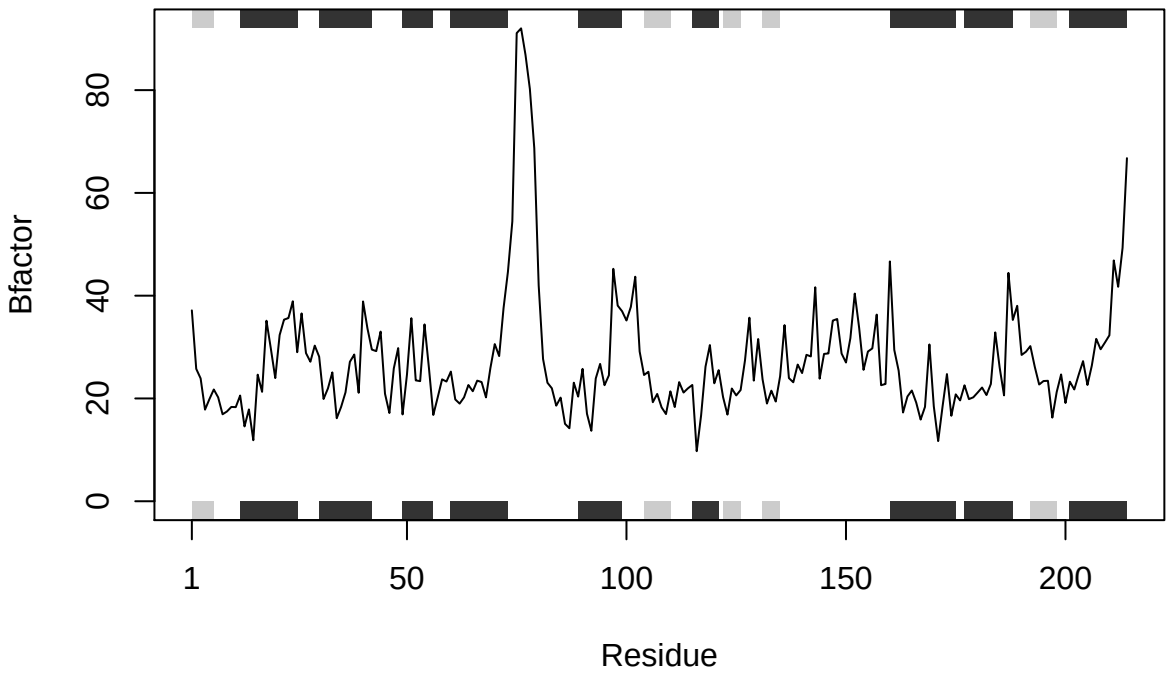
```
#install.packages("bio3d")
functionP <- function(noDrug = "NA", drug = "NA", mut = "NA") {
library(bio3d)
protein.code = c(noDrug, drug, mut)
for (id in protein.code) {
  x <- read.pdb(id) #Q1: this will return the pdb and sse objects
  schainA <- trim.pdb(x, chain="A", elety="CA")
  #this will produce a smaller PDB object including subset of atoms
  sb <- schainA$atom$b
  #this will subtract the b object in the atom data frame from the smaller PDB object
  plotb3(sb, sse=schainA, typ="l", ylab="Bfactor")
  #this is to plot a standard scatter plot with secondary structure in the marginal regions
  }
}

functionP("4AKE", "1AKE", "1E4Y")
```

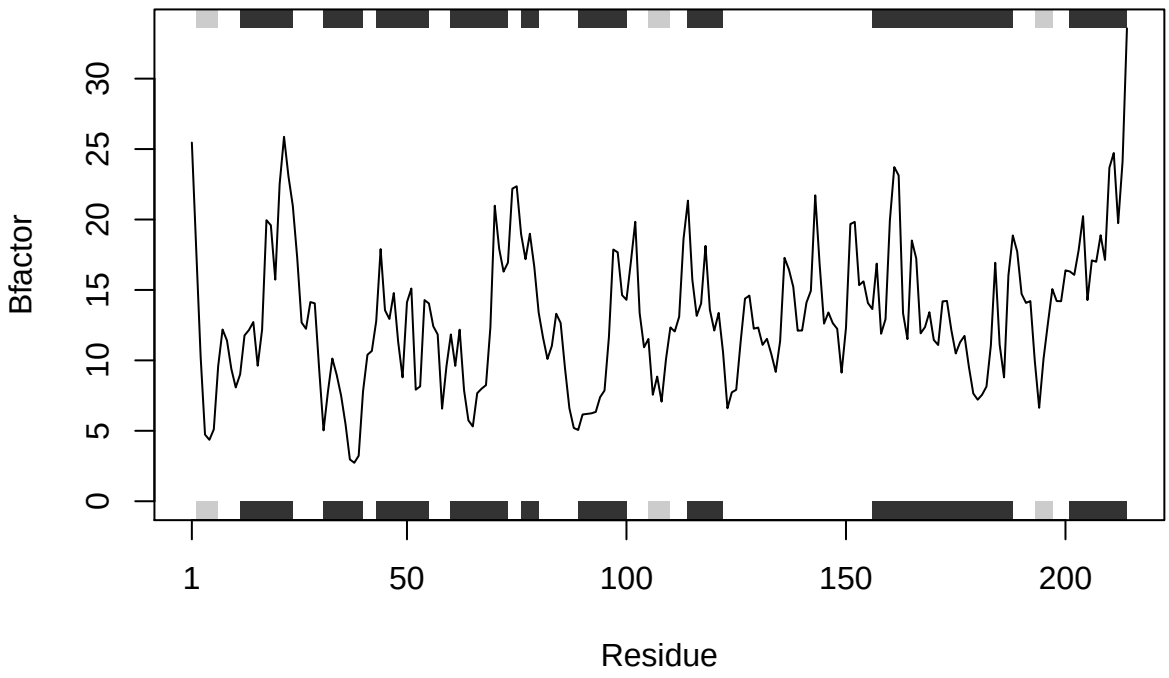
```
##   Note: Accessing on-line PDB file
```




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




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



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

Q2. What does the trim.pdb() function do? It produces 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? The top and bot parameter will determine the marginal black and grey rectangles in the plots. If they both equal to FALSE, the marginal rectangles will be disappeared. They represent the classical version of secondary structure annotation.

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

```
protein_drugplot <- function(file, chain, elmnt, fctr) {
  # allows different colors to be presented in the graph
  plot_colors <- c("cyan", "orange", "magenta")
  # to read through every value of the file vector
  for (i in 1:length(file)) {
  x <- read.pdb(file[i])
  xchain <- trim.pdb(x, chain = chain, elety = elmnt)
  atom_df <- xchain$atom
  # To subtract all the atom information and selects an entire column based on the factor input
  xfctr <- atom_df[, fctr]
  # To create the first plot
  if (i == 1) {
    plotb3(xfctr, sse = xchain, typ = "l", ylab = paste(toupper(fctr), "factor", sep = ""), col = plot_

    # adds other plots
  } else {
```

```
    lines(xfctr, col = plot_colors[i])
  }
  }

  # adds a legend for the graph
  legend("topright", title = "PDB File Name", file, fill = plot_colors, horiz=TRUE, cex = 0.3, inset = 
}
```