hw_class6

April 28, 2024

0.0.1 HW Class 6 (R Functions)

NAME: Edwin Ruiz

PID: A17136339

[1]:
```
install.packages("bio3d")
library(bio3d)
ls("package:bio3d")
```

```
The downloaded binary packages are in
/var/folders/vw/6c5wjngs433234dthdjypz800000gn/T//RtmptQF66A/downloaded_packages
```

1. 'aa.index' 2. 'aa.table' 3. 'aa123' 4. 'aa2index' 5. 'aa2mass' 6. 'aa321' 7. 'aanma' 8. 'aanma.pdb' 9. 'aanma.pdbs' 10. 'aln2html' 11. 'amsm.xyz' 12. 'angle.xyz' 13. 'as.fasta' 14. 'as.pdb' 15. 'as.pdb.default' 16. 'as.pdb.mol2' 17. 'as.pdb.prmtop' 18. 'as.select' 19. 'as.xyz' 20. 'atom.index' 21. 'atom.select' 22. 'atom.select.mol2' 23. 'atom.select.pdb' 24. 'atom.select.pdbs' 25. 'atom.select.prmtop' 26. 'atom2ele' 27. 'atom2ele.default' 28. 'atom2ele.pdb' 29. 'atom2mass' 30. 'atom2mass.default' 31. 'atom2mass.pdb' 32. 'atom2xyz' 33. 'basename.pdb' 34. 'bhattacharyya' 35. 'bhattacharyya.array' 36. 'bhattacharyya.enma' 37. 'bhattacharyya.matrix' 38. 'bhattacharyya.nma' 39. 'bhattacharyya.pca' 40. 'binding.site' 41. 'biounit' 42. 'blast.pdb' 43. 'bounds' 44. 'bounds.sse' 45. 'build.hessian' 46. 'bwr.colors' 47. 'cat.pdb' 48. 'chain.pdb' 49. 'check.utility' 50. 'clean.pdb' 51. 'cmap' 52. 'cmap.default' 53. 'cmap.pdb' 54. 'cmap.pdbs' 55. 'cmap.xyz' 56. 'cna' 57. 'cna.dccm' 58. 'cna.ensmb' 59. 'cnapath' 60. 'com' 61. 'com.pdb' 62. 'com.xyz' 63. 'combine.select' 64. 'community.aln' 65. 'community.tree' 66. 'consensus' 67. 'conserv' 68. 'convert.pdb' 69. 'core.cmap' 70. 'core.find' 71. 'core.find.default' 72. 'core.find.pdb' 73. 'core.find.pdbs' 74. 'cov.enma' 75. 'cov.nma' 76. 'covsoverlap' 77. 'covsoverlap.enma' 78. 'covsoverlap.nma' 79. 'dccm' 80. 'dccm.egnm' 81. 'dccm.enma' 82. 'dccm.gnm' 83. 'dccm.nma' 84. 'dccm.pca' 85. 'dccm.xyz' 86. 'deformation.nma' 87. 'diag.ind' 88. 'difference.vector' 89. 'dist.xyz' 90. 'dm' 91. 'dm.pdb' 92. 'dm.pdbs' 93. 'dm.xyz' 94. 'dssp' 95. 'dssp.pdb' 96. 'dssp.pdbs' 97. 'dssp.xyz' 98. 'elements' 99. 'entropy' 100. 'ff.aaenm' 101. 'ff.aaenm2' 102. 'ff.anm' 103. 'ff.calpha' 104. 'ff.pfanm' 105. 'ff.reach' 106. 'ff.sdenm' 107. 'filter.cmap' 108. 'filter.dccm' 109. 'filter.identity' 110. 'filter.rmsd' 111. 'fit.xyz' 112. 'fluct.nma' 113. 'formula2mass' 114. 'gap.inspect' 115. 'geostas' 116. 'geostas.default' 117. 'geostas.enma' 118. 'geostas.nma' 119. 'geostas.pdb' 120. 'geostas.pdbs' 121. 'geostas.xyz' 122. 'get.blast' 123. 'get.pdb' 124. 'get.seq' 125. 'gnm' 126. 'gnm.pdb' 127. 'gnm.pdbs' 128. 'hclustplot' 129. 'hivp' 130. 'hmmer' 131. 'identify.cna' 132. 'inner.prod' 133. 'inspect.connectivity' 134. 'is.gap' 135. 'is.mol2' 136. 'is.pdb' 137. 'is.pdbs' 138. 'is.select' 139. 'is.xyz' 140. 'kinesin' 141. 'layout.cna' 142. 'lbio3d' 143. 'load.enmff' 144. 'mask' 145. 'mask.dccm' 146. 'mktrj' 147. 'mktrj.enma' 148. 'mktrj.nma' 149. 'mktrj.pca' 150. 'mono.colors'

151. 'motif.find' 152. 'mustang' 153. 'network.amendment' 154. 'nma' 155. 'nma.pdb' 156. 'nma.pdbs' 157. 'normalize.vector' 158. 'orient.pdb' 159. 'overlap' 160. 'pairwise' 161. 'pca' 162. 'pca.array' 163. 'pca.pdbs' 164. 'pca.tor' 165. 'pca.xyz' 166. 'pdb.annotate' 167. 'pdb.pfam' 168. 'pdb2aln' 169. 'pdb2aln.ind' 170. 'pdb2sse' 171. 'pdbaln' 172. 'pdbfit' 173. 'pdbfit.pdb' 174. 'pdbfit.pdbs' 175. 'pdbs2pdb' 176. 'pdbs2sse' 177. 'pdbseq' 178. 'pdbsplit' 179. 'pfam' 180. 'plot.bio3d' 181. 'plot.blast' 182. 'plot.cmap' 183. 'plot.cna' 184. 'plot.cnapath' 185. 'plot.core' 186. 'plot.dccm' 187. 'plot.dmat' 188. 'plot.ecna' 189. 'plot.ecnapath' 190. 'plot.enma' 191. 'plot.fasta' 192. 'plot.fluct' 193. 'plot.geostas' 194. 'plot.hmmer' 195. 'plot.matrix.loadings' 196. 'plot.nma' 197. 'plot.pca' 198. 'plot.pca.loadings' 199. 'plot.pca.score' 200. 'plot.pca.scree' 201. 'plot.rmsip' 202. 'plotb3' 203. 'print.cna' 204. 'print.cnapath' 205. 'print.core' 206. 'print.enma' 207. 'print.fasta' 208. 'print.geostas' 209. 'print.mol2' 210. 'print.nma' 211. 'print.pca' 212. 'print.pdb' 213. 'print.prmtop' 214. 'print.rle2' 215. 'print.select' 216. 'print.sse' 217. 'print.xyz' 218. 'project.pca' 219. 'prune.cna' 220. 'pymol' 221. 'pymol.dccm' 222. 'pymol.modes' 223. 'pymol.nma' 224. 'pymol.pca' 225. 'pymol.pdbs' 226. 'read.all' 227. 'read.cif' 228. 'read.crd' 229. 'read.crd.amber' 230. 'read.crd.charmm' 231. 'read.dcd' 232. 'read.fasta' 233. 'read.fasta.pdb' 234. 'read.mol2' 235. 'read.ncdf' 236. 'read.pdb' 237. 'read.pdb2' 238. 'read.pdcBD' 239. 'read.pqr' 240. 'read.prmtop' 241. 'rgyr' 242. 'rle2' 243. 'rmsd' 244. 'rmsf' 245. 'rmsip' 246. 'rmsip.default' 247. 'rmsip.enma' 248. 'rot.lsq' 249. 'rtb' 250. 'sdENM' 251. 'seq2aln' 252. 'seqaln' 253. 'seqaln.pair' 254. 'seqbind' 255. 'seqidentity' 256. 'setup.ncore' 257. 'sip' 258. 'sip.default' 259. 'sip.enma' 260. 'sip.nma' 261. 'sse.bridges' 262. 'store.atom' 263. 'stride' 264. 'struct.aln' 265. 'summary.cna' 266. 'summary.cnapath' 267. 'summary.pdb' 268. 'torsion.pdb' 269. 'torsion.xyz' 270. 'transducin' 271. 'trim' 272. 'trim.mol2' 273. 'trim.pdb' 274. 'trim.pdbs' 275. 'trim.xyz' 276. 'unbound' 277. 'uniprot' 278. 'var.pdbs' 279. 'var.xyz' 280. 'vec2resno' 281. 'vmd' 282. 'vmd_colors' 283. 'vmd.cna' 284. 'vmd.cnapath' 285. 'vmd.ecna' 286. 'vmd.ecnapath' 287. 'wrap.tor' 288. 'write.crd' 289. 'write.fasta' 290. 'write.mol2' 291. 'write.ncdf' 292. 'write.pdb' 293. 'write.pir' 294. 'write.pqr' 295. 'xyz2atom' 296. 'xyz2z.pca' 297. 'z2xyz.pca'

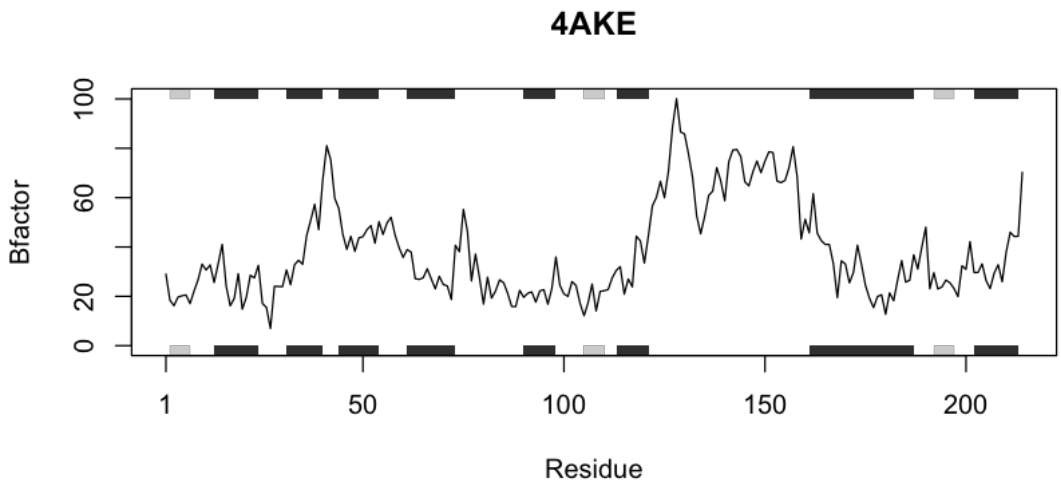
[2]:
```
"aa.index" %in% ls("package:bio3d")
```

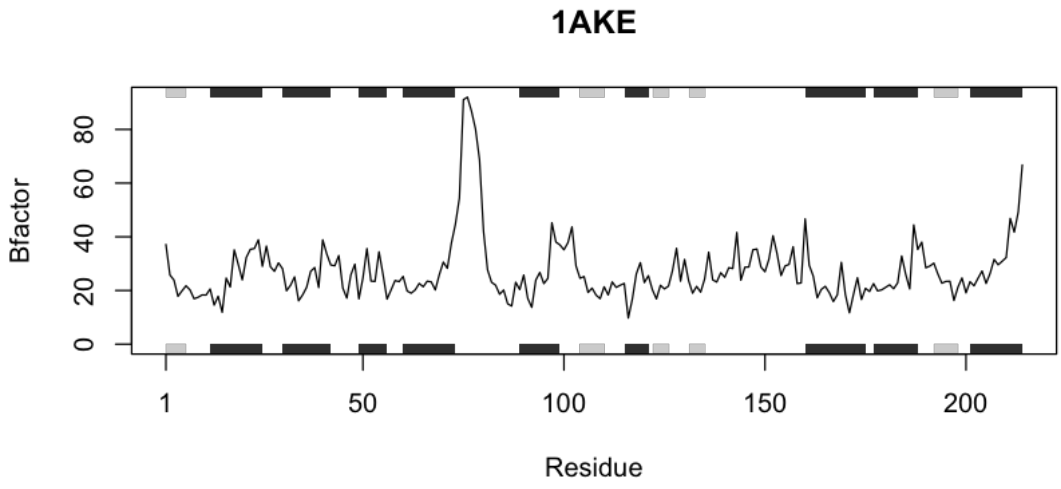
TRUE

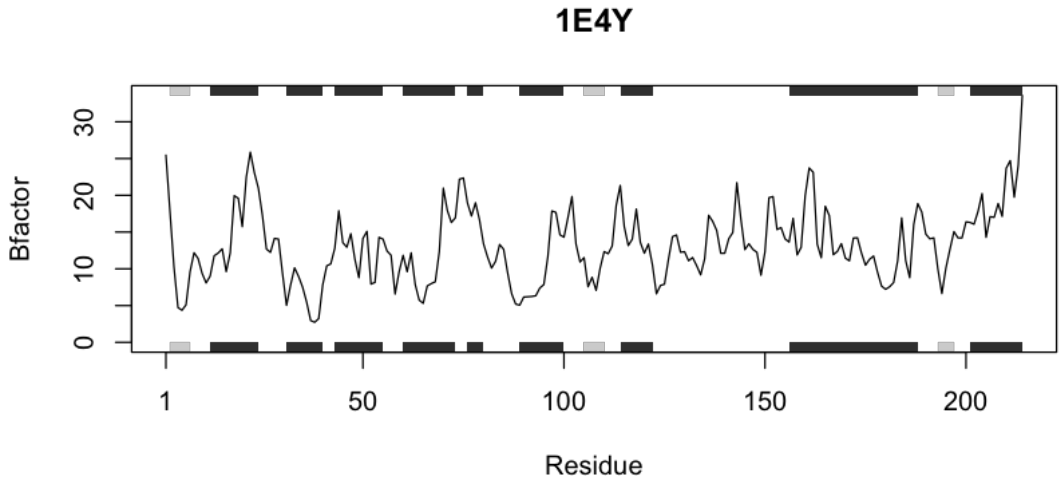
[3]:
```
#reading protein data
s1 <- read.pdb("4AKE") # kinase with drug
s2 <- read.pdb("1AKE") # kinase no drug
s3 <- read.pdb("1E4Y") # kinase with drug

#trimming
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")

#extracting b-factors
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b

#plotting b-factors
par(mfrow=c(2, 1))
```

```
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor", main="4AKE")
plotb3(s2.b, sse=s2.chainA, typ="l", ylab="Bfactor", main="1AKE")
plotb3(s3.b, sse=s3.chainA, typ="l", ylab="Bfactor", main="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
```



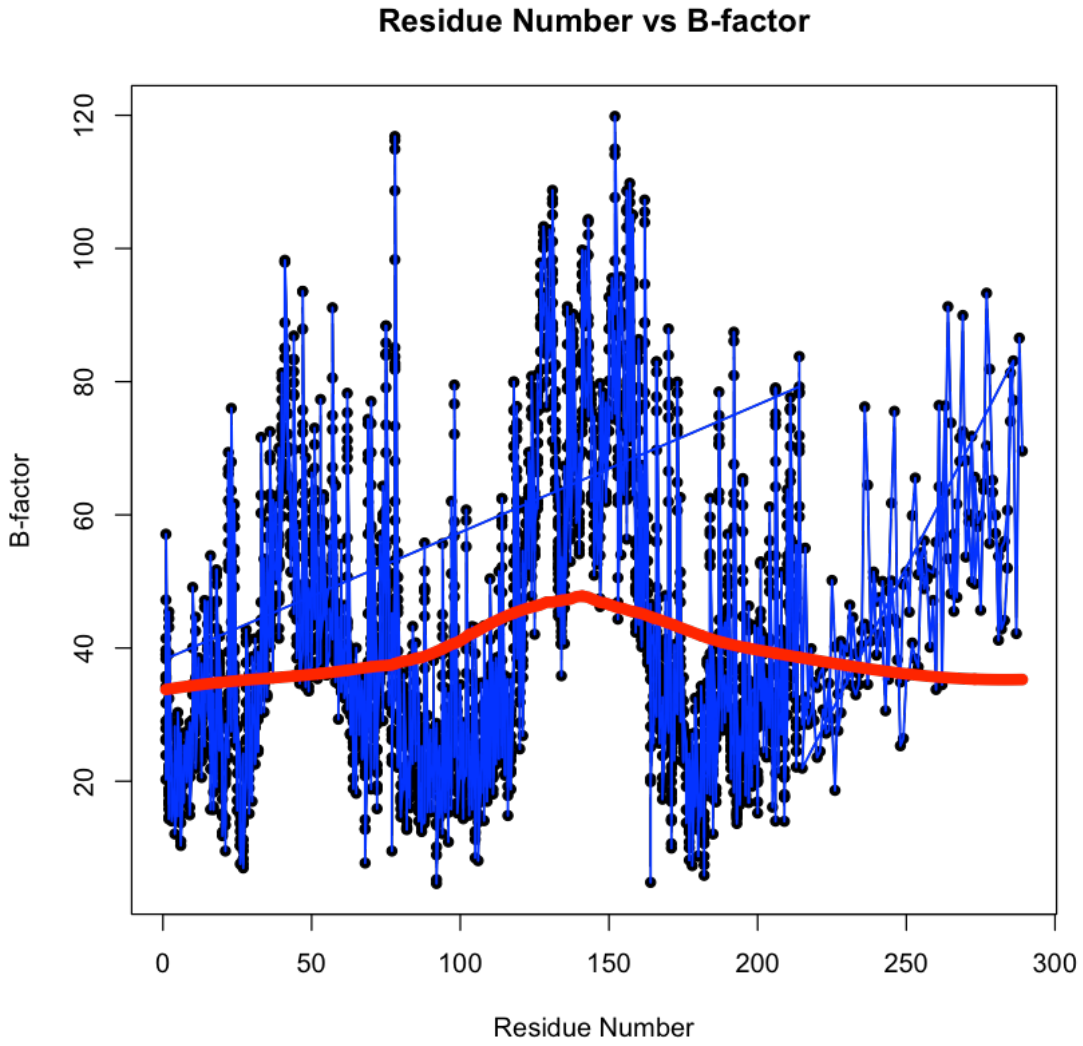




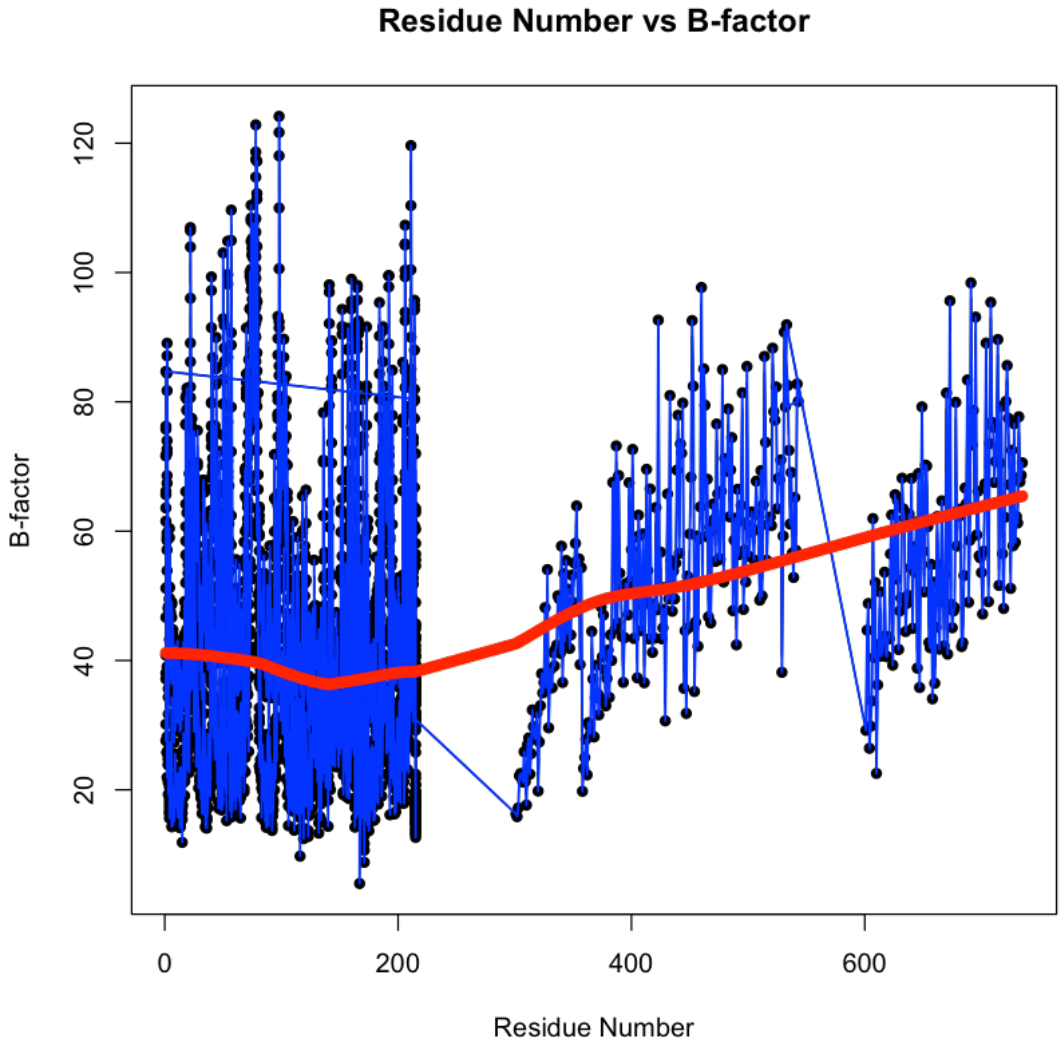
[4]:
```
#This function plots the residue number (amino acid position) vs b-factor (Atom mobility measure)
plot_residue_bfactor <- function(protein) { #protein object

    #validating input
    if (is.null(protein$atom$b) || is.null(protein$atom$resno))
    {stop("The B-factors or residue numbers are not valid")}

    #B-factors & residue numbers are extracted here
    protein.b <- protein$atom$b
    residue_numbers <- protein$atom$resno

    #checking array equality between residue numbers & B-factors
```

```
    if (length(protein.b) != length(residue_numbers))
    {stop("Mismatch between the number of B-factors and residue numbers.")}

    #creating the plot outputs
    plot(residue_numbers, protein.b, type = "b", col = "black", pch = 20, cex =␣
 ↪1.2,
         xlab = "Residue Number", ylab = "B-factor", main = "Residue Number vs␣
 ↪B-factor")

    #data point lines configurations
    lines(residue_numbers, protein.b, type = "l", col = "blue", lwd = 1.5)

    #trend line configurations
    lines(lowess(residue_numbers, protein.b), col = "red", lwd = 7)

    #output dimension debugging
    cat("Plot of", length(protein.b), "data points\n")
}
#Plotting s1, s2, s3 data
plot_residue_bfactor(s1)
plot_residue_bfactor(s2)
plot_residue_bfactor(s3)
```

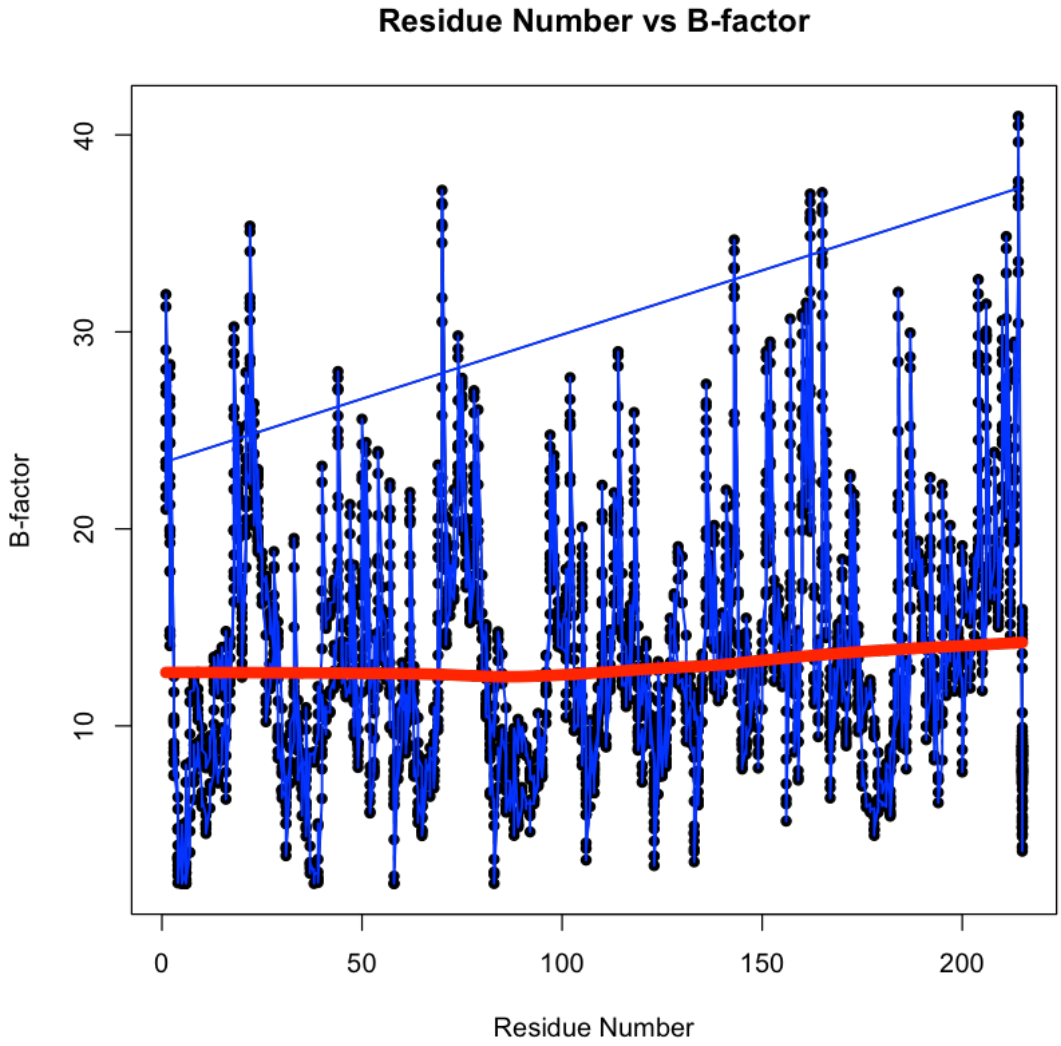
```
Plot of 3459 data points
```



Plot of 3804 data points



Plot of 3428 data points


Residue Number vs B-factor
B-factor
10
20
30
40
0
50
100
150
200
Residue Number