**Course Project**

1. Merge the training and the test sets to create one data set.
2. Use descriptive activity names to name the activities in the data set
3. Appropriately label the data set with descriptive variable names.
4. Extract only the measurements on the mean and standard deviation for each measurement.
5. From the clean data set, create a second, independent tidy data set with the average of each variable for each activity and each subject.

You will be required to submit: 1) a tidy data set, 2) a link to a Github repository with your run\_analysis.R script for performing the analysis, 3) a code book called CodeBook.md that describes the variables, the data, and any transformations or work that you performed to clean up the data, and 4) a README.md in the repo with your scripts.

1. Read in both huge files, then Merge files
2. Read in, then Cbind subject\_train and y\_train, correlate y\_train to activity\_labels
3. Read in, then Rbind features.txt
4. Many …
5. Write.table, row.name=FALSE

Test files have 561 columns.

Columns names are in the features.txt file. They can be added using rbind, or “names” command.

There are 2948 rows in all files. The columns can be combined using cbind.

Measurements are in 11 groups of 17. 3 measurements are taken for all. 11\*17\*3=561

We want the Mean (first measurement in each group) and Std (second measurement in each group). This means that we want the first 6 measurements for each group of 11, for a total of 66 columns.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 41 | 42 | 43 | 44 | 45 | 46 |
| 81 | 82 | 83 | 84 | 85 | 86 |
| 121 | 122 | 123 | 124 | 125 | 126 |
| 161 | 162 | 163 | 164 | 165 | 166 |
| 201 | 202 | 214 | 215 | 227 | 228 |
| 240 | 241 | 253 | 254 |  |  |
| 266 | 267 | 268 | 269 | 270 | 271 |
| 345 | 346 | 347 | 348 | 349 | 350 |
| 424 | 425 | 426 | 427 | 428 | 429 |
| 503 | 504 | 516 | 517 | 529 | 530 |

542, 543

rep(16,6),rep(-16,35), rep(16,6),rep(-16,35), rep(16,6),rep(-16,35), rep(16,6),rep(-16,35), rep(16,6),rep(-16,35),

rep(16,2),rep(-16,11), rep(16,2),rep(-16,11), rep(16,2),rep(-16,11), rep(16,2),rep(-16,11), rep(16,2),rep(-16,11),

rep(16,6),rep(-16,35), rep(16,6),rep(-16,73), rep(16,6),rep(-16,73), rep(16,2),rep(-16,11), rep(16,2),rep(-16,11), rep(16,2),rep(-16,11), rep(16,2),rep(-16,11)

c(features[1:6,2],features[41:46,2],features[81:86,2],features[121:126,2],features[161:166,2],features[201:202,2],features[214:215,2],features[227:228,2],features[240:241,2],features[253:254,2],features[266:271,2],features[345:350,2],features[424:429,2],features[503:504,2],features[516:517,2],features[529:530,2],features[542:543,2])

The 9 small data sets in the subdirectories have 128 columns in them. I’m not sure why we need those files.