Machine Learning Project

Contents

load data

```
training<-read.csv('./data/pml-training.csv',stringsAsFactors=TRUE,na.strings = "#DIV/0!")
dim(training)
```

```
## [1] 19622   160
```

```
str(training)
```

```
## 'data.frame':    19622 obs. of  160 variables:
##  $ X                       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ user_name               : Factor w/ 6 levels "adelmo","carlitos",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ raw_timestamp_part_1    : int  1323084231 1323084231 1323084231 1323084232 1323084232 1323084232 1
##  $ raw_timestamp_part_2    : int  788290 808298 820366 120339 196328 304277 368296 440390 484323 484
##  $ cvtd_timestamp          : Factor w/ 20 levels "02/12/2011 13:32",..: 9 9 9 9 9 9 9 9 9 9 ...
##  $ new_window              : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ num_window              : int  11 11 11 12 12 12 12 12 12 12 ...
##  $ roll_belt               : num  1.41 1.41 1.42 1.48 1.48 1.45 1.42 1.42 1.43 1.45 ...
##  $ pitch_belt              : num  8.07 8.07 8.07 8.05 8.07 8.06 8.09 8.13 8.16 8.17 ...
##  $ yaw_belt                : num  -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 ...
##  $ total_accel_belt        : int  3 3 3 3 3 3 3 3 3 3 ...
##  $ kurtosis_roll_belt      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ kurtosis_picth_belt     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ kurtosis_yaw_belt       : logi  NA NA NA NA NA NA ...
##  $ skewness_roll_belt      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ skewness_roll_belt.1    : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ skewness_yaw_belt       : logi  NA NA NA NA NA NA ...
##  $ max_roll_belt           : Factor w/ 196 levels "-0.2","-0.4",..: 196 196 196 196 196 196 196 196 1
##  $ max_picth_belt          : Factor w/ 23 levels "10","11","17",..: 23 23 23 23 23 23 23 23 23 23 ..
##  $ max_yaw_belt            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_roll_belt           : Factor w/ 185 levels "-0.2","0.3","-0.5",..: 185 185 185 185 185 185 185
##  $ min_pitch_belt          : Factor w/ 17 levels "0","1","15","16",..: 17 17 17 17 17 17 17 17 17 17
##  $ min_yaw_belt            : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_roll_belt     : Factor w/ 149 levels "0","0.1","0.13",..: 149 149 149 149 149 149 149 14
##  $ amplitude_pitch_belt    : Factor w/ 14 levels "0","1","10","11",..: 14 14 14 14 14 14 14 14 14 14
##  $ amplitude_yaw_belt      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ var_total_accel_belt    : Factor w/ 66 levels "0","0.0217","0.0278",..: 66 66 66 66 66 66 66 66 6
##  $ avg_roll_belt           : Factor w/ 192 levels "0","0.1","-0.2",..: 192 192 192 192 192 192 192 19
##  $ stddev_roll_belt        : Factor w/ 70 levels "0","0.091","0.0957",..: 70 70 70 70 70 70 70 70 70
##  $ var_roll_belt           : Factor w/ 97 levels "0","0.0083","0.0092",..: 97 97 97 97 97 97 97 97 97
##  $ avg_pitch_belt          : Factor w/ 215 levels "-0.2","-0.4",..: 215 215 215 215 215 215 215 215 1
##  $ stddev_pitch_belt       : Factor w/ 44 levels "0","0.0571","0.1",..: 44 44 44 44 44 44 44 44 44 44
##  $ var_pitch_belt          : Factor w/ 64 levels "0","0.0033","0.0393",..: 64 64 64 64 64 64 64 64 64
##  $ avg_yaw_belt            : Factor w/ 241 levels "-0.1","0.1","0.3",..: 241 241 241 241 241 241 241
##  $ stddev_yaw_belt         : Factor w/ 59 levels "0","0.0407","0.0522",..: 59 59 59 59 59 59 59 59 59
```

```
##  $ var_yaw_belt           : Factor w/ 146 levels "0","0.0017","0.0027",..: 146 146 146 146 146 146 1
##  $ gyros_belt_x           : num  0 0.02 0 0.02 0.02 0.02 0.02 0.02 0.02 0.03 ...
##  $ gyros_belt_y           : num  0 0 0 0 0.02 0 0 0 0 0 ...
##  $ gyros_belt_z           : num  -0.02 -0.02 -0.02 -0.03 -0.02 -0.02 -0.02 -0.02 -0.02 0 ...
##  $ accel_belt_x           : int  -21 -22 -20 -22 -21 -21 -22 -22 -20 -21 ...
##  $ accel_belt_y           : int  4 4 5 3 2 4 3 4 2 4 ...
##  $ accel_belt_z           : int  22 22 23 21 24 21 21 21 24 22 ...
##  $ magnet_belt_x          : int  -3 -7 -2 -6 -6 0 -4 -2 1 -3 ...
##  $ magnet_belt_y          : int  599 608 600 604 600 603 599 603 602 609 ...
##  $ magnet_belt_z          : int  -313 -311 -305 -310 -302 -312 -311 -313 -312 -308 ...
##  $ roll_arm               : num  -128 -128 -128 -128 -128 -128 -128 -128 -128 -128 ...
##  $ pitch_arm              : num  22.5 22.5 22.5 22.1 22.1 22 21.9 21.8 21.7 21.6 ...
##  $ yaw_arm                : num  -161 -161 -161 -161 -161 -161 -161 -161 -161 -161 ...
##  $ total_accel_arm        : int  34 34 34 34 34 34 34 34 34 34 ...
##  $ var_accel_arm          : Factor w/ 396 levels "0","0.0179","0.02",..: 396 396 396 396 396 396 396
##  $ avg_roll_arm           : Factor w/ 331 levels "0","-0.7853",..: 331 331 331 331 331 331 331 331 3
##  $ stddev_roll_arm        : Factor w/ 331 levels "0","0.05","0.1081",..: 331 331 331 331 331 331 331
##  $ var_roll_arm           : Factor w/ 331 levels "0","0.0025","0.0117",..: 331 331 331 331 331 331 3
##  $ avg_pitch_arm          : Factor w/ 331 levels "0","0.1573","0.3422",..: 331 331 331 331 331 331 3
##  $ stddev_pitch_arm       : Factor w/ 331 levels "0","0.0153","0.135",..: 331 331 331 331 331 331 33
##  $ var_pitch_arm          : Factor w/ 331 levels "0","0.0182","0.0275",..: 331 331 331 331 331 331 3
##  $ avg_yaw_arm            : Factor w/ 331 levels "0","-0.0188",..: 331 331 331 331 331 331 331 331 3
##  $ stddev_yaw_arm         : Factor w/ 328 levels "0","0.3471","0.3594",..: 328 328 328 328 328 328 3
##  $ var_yaw_arm            : Factor w/ 328 levels "0","0.1205","0.1292",..: 328 328 328 328 328 328 3
##  $ gyros_arm_x            : num  0 0.02 0.02 0.02 0 0.02 0 0.02 0.02 0.02 ...
##  $ gyros_arm_y            : num  0 -0.02 -0.02 -0.03 -0.03 -0.03 -0.03 -0.02 -0.03 -0.03 ...
##  $ gyros_arm_z            : num  -0.02 -0.02 -0.02 0.02 0 0 0 0 -0.02 -0.02 ...
##  $ accel_arm_x            : int  -288 -290 -289 -289 -289 -289 -289 -289 -288 -288 ...
##  $ accel_arm_y            : int  109 110 110 111 111 111 111 111 109 110 ...
##  $ accel_arm_z            : int  -123 -125 -126 -123 -123 -122 -125 -124 -122 -124 ...
##  $ magnet_arm_x           : int  -368 -369 -368 -372 -374 -369 -373 -372 -369 -376 ...
##  $ magnet_arm_y           : int  337 337 344 344 337 342 336 338 341 334 ...
##  $ magnet_arm_z           : int  516 513 513 512 506 513 509 510 518 516 ...
##  $ kurtosis_roll_arm      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ kurtosis_picth_arm     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ kurtosis_yaw_arm       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ skewness_roll_arm      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ skewness_pitch_arm     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ skewness_yaw_arm       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_roll_arm           : Factor w/ 291 levels "0","-0.1","0.1",..: 291 291 291 291 291 291 291 29
##  $ max_picth_arm          : Factor w/ 264 levels "0","-0.8","100",..: 264 264 264 264 264 264 264 26
##  $ max_yaw_arm            : Factor w/ 52 levels "13","14","15",..: 52 52 52 52 52 52 52 52 52 52 ..
##  $ min_roll_arm           : Factor w/ 279 levels "0","-0.2","0.2",..: 279 279 279 279 279 279 279 27
##  $ min_pitch_arm          : Factor w/ 291 levels "0","0.3","-1",..: 291 291 291 291 291 291 291 291
##  $ min_yaw_arm            : Factor w/ 39 levels "1","10","11",..: 39 39 39 39 39 39 39 39 39 39 ...
##  $ amplitude_roll_arm     : Factor w/ 307 levels "0","0.03","0.5",..: 307 307 307 307 307 307 307 30
##  $ amplitude_pitch_arm    : Factor w/ 295 levels "0","1","10","101.1",..: 295 295 295 295 295 295 29
##  $ amplitude_yaw_arm      : Factor w/ 52 levels "0","1","10","11",..: 52 52 52 52 52 52 52 52 52 52
##  $ roll_dumbbell          : num  13.1 13.1 12.9 13.4 13.4 ...
##  $ pitch_dumbbell         : num  -70.5 -70.6 -70.3 -70.4 -70.4 ...
##  $ yaw_dumbbell           : num  -84.9 -84.7 -85.1 -84.9 -84.9 ...
##  $ kurtosis_roll_dumbbell : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ kurtosis_picth_dumbbell: num  NA NA NA NA NA NA NA NA NA NA ...
##  $ kurtosis_yaw_dumbbell  : logi  NA NA NA NA NA NA ...
```

```
##  $ skewness_roll_dumbbell  : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ skewness_pitch_dumbbell : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ skewness_yaw_dumbbell   : logi  NA NA NA NA NA NA ...
##  $ max_roll_dumbbell       : Factor w/ 339 levels "0","10","101",..: 339 339 339 339 339 339 339 339
##  $ max_picth_dumbbell      : Factor w/ 340 levels "-100.1","100.2",..: 340 340 340 340 340 340 340 34
##  $ max_yaw_dumbbell        : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ min_roll_dumbbell       : Factor w/ 333 levels "0","0.9","1",..: 333 333 333 333 333 333 333 333 3
##  $ min_pitch_dumbbell      : Factor w/ 357 levels "0","-100.1","-100.2",..: 357 357 357 357 357 357 3
##  $ min_yaw_dumbbell        : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ amplitude_roll_dumbbell : Factor w/ 388 levels "0","0.96","1",..: 388 388 388 388 388 388 388 388
##   [list output truncated]
```

tidying

write a function to do this, as we will need to tidy the test data also cvtd timestamp -> to date kurtosis_roll_belt - div/0 ->NA kurtosis_picth_belt