Machine Learning Project

Contents

```
6
glm
load data
training<-read.csv('./data/pml-training.csv',stringsAsFactors=TRUE,na.strings = c("#DIV/0!","NA"))
dim(training)
## [1] 19622
              160
str(training,list.len=ncol(training))
                  19622 obs. of 160 variables:
## 'data.frame':
## $ X
                            : int 1 2 3 4 5 6 7 8 9 10 ...
                            : Factor w/ 6 levels "adelmo", "carlitos", ...: 2 2 2 2 2 2 2 2 2 ...
##
   $ user_name
## $ raw_timestamp_part_1
                                  1323084231 1323084231 1323084231 1323084232 1323084232 1323084232
## $ raw_timestamp_part_2
                                  788290 808298 820366 120339 196328 304277 368296 440390 484323 484
                            : Factor w/ 20 levels "02/12/2011 13:32",..: 9 9 9 9 9 9 9 9 9 9 ...
## $ cvtd_timestamp
## $ new_window
                            : Factor w/ 2 levels "no", "yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ num window
                                  11 11 11 12 12 12 12 12 12 12 ...
## $ roll belt
                                  1.41 1.41 1.42 1.48 1.48 1.45 1.42 1.42 1.43 1.45 ...
                            : num
## $ pitch belt
                                  8.07 8.07 8.07 8.05 8.07 8.06 8.09 8.13 8.16 8.17 ...
## $ yaw_belt
                                  -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 ...
                            : num
## $ total_accel_belt
                                  3 3 3 3 3 3 3 3 3 3 . . .
                            : int
                                  NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_roll_belt
                            : num
## $ kurtosis_picth_belt
                            : num
                                  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 NA NA NA NA NA NA NA NA ...
##
   $ skewness_roll_belt.1
                            : num
##
   $ skewness_yaw_belt
                            : logi NA NA NA NA NA NA ...
## $ max_roll_belt
                            : num NA NA NA NA NA NA NA NA NA ...
## $ max_picth_belt
                            : int NA NA NA NA NA NA NA NA NA ...
                                  NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_belt
                            : num
## $ min_roll_belt
                            : num NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_belt
                            : int NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_belt
                            : num NA NA NA NA NA NA NA NA NA ...
##
   $ amplitude roll belt
                            : num
                                  NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_belt
                            : int
                                  NA NA NA NA NA NA NA NA NA ...
## $ amplitude yaw belt
                            : num NA NA NA NA NA NA NA NA NA ...
## $ var_total_accel_belt
                            : num NA NA NA NA NA NA NA NA NA ...
## $ avg roll belt
                                  NA NA NA NA NA NA NA NA NA ...
                            : num
## $ stddev_roll_belt
                            : num NA NA NA NA NA NA NA NA NA ...
## $ var_roll_belt
                            : num
                                 NA NA NA NA NA NA NA NA NA ...
## $ avg_pitch_belt
                                  NA NA NA NA NA NA NA NA NA ...
                            : num
## $ stddev_pitch_belt
                            : num NA NA NA NA NA NA NA NA NA ...
```

```
## $ var_pitch_belt
                                 NA NA NA NA NA NA NA NA NA ...
                           : num
## $ avg_yaw_belt
                                 NA NA NA NA NA NA NA NA NA ...
                           : num
## $ stddev yaw belt
                           : num
                                 NA NA NA NA NA NA NA NA NA ...
## $ var_yaw_belt
                                 NA NA NA NA NA NA NA NA NA ...
                           : num
## $ gyros_belt_x
                           : num
                                 ## $ gyros_belt_y
                                 0 0 0 0 0.02 0 0 0 0 0 ...
                           : num
                                 -0.02 -0.02 -0.02 -0.03 -0.02 -0.02 -0.02 -0.02 -0.02 0 ...
## $ gyros_belt_z
                           : num
                                 -21 -22 -20 -22 -21 -21 -22 -22 -20 -21 ...
## $ accel belt x
                           : int
   $ accel_belt_y
##
                           : int
                                 4 4 5 3 2 4 3 4 2 4 ...
## $ accel_belt_z
                                 22 22 23 21 24 21 21 21 24 22 ...
                           : int
## $ magnet_belt_x
                           : int
                                 -3 -7 -2 -6 -6 0 -4 -2 1 -3 ...
##
                                 599 608 600 604 600 603 599 603 602 609 ...
   $ magnet_belt_y
                           : int
## $ magnet_belt_z
                           : int
                                 -313 -311 -305 -310 -302 -312 -311 -313 -312 -308 ...
## $ roll_arm
                                 : num
## $ pitch_arm
                                 22.5 22.5 22.5 22.1 22.1 22 21.9 21.8 21.7 21.6 ...
                           : num
##
   $ yaw_arm
                                 : num
## $ total_accel_arm
                                 34 34 34 34 34 34 34 34 34 ...
                           : int
## $ var accel arm
                                 NA NA NA NA NA NA NA NA NA ...
                           : num
                                 NA NA NA NA NA NA NA NA NA ...
## $ avg_roll_arm
                           : num
## $ stddev roll arm
                           : num
                                 NA NA NA NA NA NA NA NA NA ...
## $ var_roll_arm
                           : num
                                NA NA NA NA NA NA NA NA NA ...
## $ avg_pitch_arm
                                 NA NA NA NA NA NA NA NA NA ...
                           : num
## $ stddev_pitch_arm
                                 NA NA NA NA NA NA NA NA NA ...
                           : num
## $ var_pitch_arm
                                 NA NA NA NA NA NA NA NA NA ...
                           : num
## $ avg_yaw_arm
                           : num
                                 NA NA NA NA NA NA NA NA NA ...
## $ stddev_yaw_arm
                           : num
                                NA NA NA NA NA NA NA NA NA ...
## $ var_yaw_arm
                                 NA NA NA NA NA NA NA NA NA ...
                           : num
## $ gyros_arm_x
                           : num
                                 ## $ gyros_arm_y
                                 0 -0.02 -0.02 -0.03 -0.03 -0.03 -0.03 -0.02 -0.03 -0.03 ...
                           : num
## $ gyros_arm_z
                                 -0.02 -0.02 -0.02 0.02 0 0 0 0 -0.02 -0.02 ...
                           : num
## $ accel_arm_x
                           : int
                                 ## $ 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
                                 -368 -369 -368 -372 -374 -369 -373 -372 -369 -376 ...
                           : int
## $ magnet_arm_y
                                 337 337 344 344 337 342 336 338 341 334 ...
                           : int
## $ 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 ...
## $ kurtosis_picth_arm
                           : num
                                 NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_arm
                                 NA NA NA NA NA NA NA NA NA ...
                           : num
## $ skewness_roll_arm
                           : num
                                 NA NA NA NA NA NA NA NA NA ...
## $ skewness_pitch_arm
                                NA NA NA NA NA NA NA NA NA ...
                           : num
## $ skewness_yaw_arm
                                 NA NA NA NA NA NA NA NA NA ...
                           : num
                                 NA NA NA NA NA NA NA NA NA ...
## $ max roll arm
                           : num
## $ max_picth_arm
                                 NA NA NA NA NA NA NA NA NA ...
                           : num
## $ max_yaw_arm
                           : int
                                 NA NA NA NA NA NA NA NA NA ...
##
                                 NA NA NA NA NA NA NA NA NA ...
   $ min_roll_arm
                           : num
##
   $ min_pitch_arm
                           : num
                                 NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_arm
                           : int
                                 NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_arm
                           : num
                                NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_arm
                           : num
                                 NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_arm
                           : int
                                 NA NA NA NA NA NA NA NA NA ...
## $ 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 ...
##
   $ kurtosis_picth_dumbbell : num NA ...
   $ kurtosis yaw dumbbell
                             : logi
                                   NA NA NA NA NA ...
##
   $ skewness_roll_dumbbell
                                   NA NA NA NA NA NA NA NA NA ...
                            : num
   $ skewness_pitch_dumbbell : num
##
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ skewness yaw dumbbell
                             : logi
                                   NA NA NA NA NA ...
   $ max roll dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ max_picth_dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ max_yaw_dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ min_roll_dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
   $ min_pitch_dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
                                   NA NA NA NA NA NA NA NA NA ...
   $ min_yaw_dumbbell
                             : num
##
   $ amplitude_roll_dumbbell : num
                                   NA NA NA NA NA NA NA NA NA ...
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ amplitude_pitch_dumbbell: num
##
                                   NA NA NA NA NA NA NA NA NA ...
   $ amplitude_yaw_dumbbell
                            : num
##
   $ total_accel_dumbbell
                             : int
                                   37 37 37 37 37 37 37 37 37 ...
##
   $ var_accel_dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ avg roll dumbbell
                                   NA NA NA NA NA NA NA NA NA ...
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ stddev_roll_dumbbell
                             : num
##
   $ var roll dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
## $ avg_pitch_dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
                                   NA NA NA NA NA NA NA NA NA ...
   $ stddev_pitch_dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ var_pitch_dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ avg yaw dumbbell
                             : num
##
   $ stddev_yaw_dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
   $ var_yaw_dumbbell
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
                                   0 0 0 0 0 0 0 0 0 0 ...
   $ gyros_dumbbell_x
                             : num
                                   -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 ...
##
   $ gyros_dumbbell_y
                             : num
##
  $ gyros_dumbbell_z
                             : num
                                   0 0 0 -0.02 0 0 0 0 0 0 ...
##
   $ accel_dumbbell_x
                                   -234 -233 -232 -232 -233 -234 -232 -234 -232 -235 ...
                             : int
##
   $ accel_dumbbell_y
                             : int
                                   47 47 46 48 48 48 47 46 47 48 ...
##
   $ accel_dumbbell_z
                             : int
                                   -271 -269 -270 -269 -270 -269 -270 -272 -269 -270 ...
##
   $ magnet_dumbbell_x
                             : int
                                   -559 -555 -561 -552 -554 -558 -551 -555 -549 -558 ...
##
   $ magnet_dumbbell_y
                                   293 296 298 303 292 294 295 300 292 291 ...
                             : int
##
   $ magnet dumbbell z
                                   -65 -64 -63 -60 -68 -66 -70 -74 -65 -69 ...
                             : num
##
   $ roll forearm
                             : num
                                   28.4 28.3 28.3 28.1 28 27.9 27.9 27.8 27.7 27.7 ...
##
  $ pitch forearm
                             : num
                                   -63.9 -63.9 -63.9 -63.9 -63.9 -63.9 -63.8 -63.8 -63.8 ...
##
                             : num
                                   $ yaw_forearm
##
   $ kurtosis roll forearm
                                   NA NA NA NA NA NA NA NA NA ...
                             : num
##
   $ kurtosis_picth_forearm
                            : num
                                   NA NA NA NA NA NA NA NA NA ...
   $ kurtosis_yaw_forearm
                             : logi
                                   NA NA NA NA NA ...
##
   $ skewness roll forearm
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ skewness_pitch_forearm
                            : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ skewness_yaw_forearm
                             : logi NA NA NA NA NA ...
   $ max_roll_forearm
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
                                   NA NA NA NA NA NA NA NA NA ...
   $ max_picth_forearm
                             : num
##
   $ max_yaw_forearm
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ min_roll_forearm
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ min_pitch_forearm
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
##
   $ min_yaw_forearm
                                   NA NA NA NA NA NA NA NA NA ...
                             : num
## $ amplitude_roll_forearm : num
                                   NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_forearm : num
                                   NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_forearm
                             : num
                                   NA NA NA NA NA NA NA NA NA ...
   $ total accel forearm
                             : int 36 36 36 36 36 36 36 36 36 ...
```

```
## $ var accel forearm
                                  NA NA NA NA NA NA NA NA NA ...
                            : num
## $ avg_roll_forearm
                                  NA NA NA NA NA NA NA NA NA ...
                            : num
## $ stddev roll forearm
                            : num
                                  NA NA NA NA NA NA NA NA NA ...
                                  NA NA NA NA NA NA NA NA NA ...
## $ var_roll_forearm
                            : num
## $ avg_pitch_forearm
                            : num
                                  NA NA NA NA NA NA NA NA NA ...
                            : num NA NA NA NA NA NA NA NA NA ...
## $ stddev_pitch_forearm
## $ var_pitch_forearm
                                  NA NA NA NA NA NA NA NA NA . . .
                            : num
                                  NA NA NA NA NA NA NA NA NA ...
##
   $ avg_yaw_forearm
                            : num
##
   $ stddev_yaw_forearm
                            : num
                                  NA NA NA NA NA NA NA NA NA ...
## $ var_yaw_forearm
                            : num
                                  NA NA NA NA NA NA NA NA NA ...
## $ gyros_forearm_x
                            : num
                                  ##
   $ gyros_forearm_y
                                  0 0 -0.02 -0.02 0 -0.02 0 -0.02 0 0 ...
                            : num
##
   $ gyros_forearm_z
                                  -0.02 -0.02 0 0 -0.02 -0.03 -0.02 0 -0.02 -0.02 ...
                            : num
## $ accel_forearm_x
                            : int
                                  192 192 196 189 189 193 195 193 193 190 ...
## $ accel_forearm_y
                            : int
                                  203 203 204 206 206 203 205 205 204 205 ...
## $ accel_forearm_z
                            : int
                                  -215 -216 -213 -214 -214 -215 -215 -213 -214 -215 ...
## $ magnet_forearm_x
                            : int
                                  -17 -18 -18 -16 -17 -9 -18 -9 -16 -22 ...
## $ magnet_forearm_y
                            : num 654 661 658 658 655 660 659 660 653 656 ...
                            : num 476 473 469 469 473 478 470 474 476 473 ...
## $ magnet_forearm_z
## $ classe
                            : Factor w/ 5 levels "A", "B", "C", "D", ...: 1 1 1 1 1 1 1 1 1 1 ...
```

training\$cvtd_time<-strptime(training\$cvtd_timestamp,format = "%d/%m/%Y %H:%M",list=FALSE)

tidying

write a function to do this, as we will need to tidy the test data also cvtd timestamp -> to date \$ max_roll_belt max_picth_belt min_roll_belt min_pitch_belt amplitude_roll_belt

A bunch of others

set things to missing, impute the missing values

Then procede

```
na_list<-sapply(training,function(x) { sum(is.na(x))})
good<- na_list < 1900
slimmed<-training[good]
str(slimmed)</pre>
```

```
## 'data.frame':
                  19622 obs. of 60 variables:
##
   $ X
                        : int 1 2 3 4 5 6 7 8 9 10 ...
                        : Factor w/ 6 levels "adelmo", "carlitos", ...: 2 2 2 2 2 2 2 2 2 ...
## $ user_name
## $ raw_timestamp_part_1: int 1323084231 1323084231 1323084231 1323084232 1323084232 1323084232 13230
   $ raw_timestamp_part_2: int 788290 808298 820366 120339 196328 304277 368296 440390 484323 484434
##
   $ 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
                              11 11 11 12 12 12 12 12 12 12 ...
                        : int
## $ roll_belt
                               1.41 1.41 1.42 1.48 1.48 1.45 1.42 1.42 1.43 1.45 ...
                        : num
## $ pitch_belt
                              8.07 8.07 8.07 8.05 8.07 8.06 8.09 8.13 8.16 8.17 ...
                        : num
## $ yaw belt
                              -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 ...
                        : num
## $ total_accel_belt
                              3 3 3 3 3 3 3 3 3 ...
                        : int
## $ gyros belt x
                        : num
                              ## $ gyros_belt_y
                        : num 00000.020000 ...
```

```
## $ gyros belt z
                             -0.02 -0.02 -0.02 -0.03 -0.02 -0.02 -0.02 -0.02 -0.02 0 ...
                       : num
## $ accel_belt_x
                             -21 -22 -20 -22 -21 -21 -22 -22 -20 -21 ...
                       : int
                             4 4 5 3 2 4 3 4 2 4 ...
## $ accel belt y
                       : int
## $ accel_belt_z
                             22 22 23 21 24 21 21 21 24 22 ...
                       : int
## $ 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 ...
                             -313 -311 -305 -310 -302 -312 -311 -313 -312 -308 ...
## $ magnet_belt_z
                       : int
                             ## $ roll arm
                       : num
                       : num
##
   $ pitch arm
                             22.5 22.5 22.5 22.1 22.1 22 21.9 21.8 21.7 21.6 ...
## $ yaw_arm
                       : num
                             $ total_accel_arm
                       : int
                             34 34 34 34 34 34 34 34 34 ...
##
                             $ gyros_arm_x
                       : num
##
   $ 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
## $ 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
                             -368 -369 -368 -372 -374 -369 -373 -372 -369 -376 ...
                       : int
                             337 337 344 344 337 342 336 338 341 334 ...
## $ magnet_arm_y
                       : int
## $ magnet_arm_z
                       : int
                             516 513 513 512 506 513 509 510 518 516 ...
## $ roll_dumbbell
                       : num
                             13.1 13.1 12.9 13.4 13.4 ...
## $ pitch dumbbell
                             -70.5 -70.6 -70.3 -70.4 -70.4 ...
                       : num
## $ yaw_dumbbell
                             -84.9 -84.7 -85.1 -84.9 -84.9 ...
                       : num
## $ total_accel_dumbbell: int
                             37 37 37 37 37 37 37 37 37 ...
## $ gyros_dumbbell_x
                       : num
                             0 0 0 0 0 0 0 0 0 0 ...
## $ gyros_dumbbell_y
                       : num
                             -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 ...
## $ gyros_dumbbell_z
                             0 0 0 -0.02 0 0 0 0 0 0 ...
                       : num
## $ accel_dumbbell_x
                       : int
                             -234 -233 -232 -232 -233 -234 -232 -234 -232 -235 ...
## $ accel_dumbbell_y
                       : int
                             47 47 46 48 48 48 47 46 47 48 ...
## $ accel_dumbbell_z
                             -271 -269 -270 -269 -270 -269 -270 -272 -269 -270 ...
                       : int
##
   $ magnet_dumbbell_x
                       : int
                             -559 -555 -561 -552 -554 -558 -551 -555 -549 -558 ...
## $ magnet_dumbbell_y
                       : int
                             293 296 298 303 292 294 295 300 292 291 ...
## $ magnet_dumbbell_z
                             -65 -64 -63 -60 -68 -66 -70 -74 -65 -69 ...
                       : num
## $ roll_forearm
                             28.4 28.3 28.3 28.1 28 27.9 27.9 27.8 27.7 27.7 ...
                       : num
## $ pitch forearm
                             -63.9 -63.9 -63.9 -63.9 -63.9 -63.9 -63.8 -63.8 -63.8 ...
                       : num
## $ yaw_forearm
                             : num
## $ total accel forearm : int
                             36 36 36 36 36 36 36 36 36 ...
## $ gyros_forearm_x
                             : num
## $ gyros_forearm_y
                             0 0 -0.02 -0.02 0 -0.02 0 -0.02 0 0 ...
                       : num
## $ gyros_forearm_z
                             -0.02 -0.02 0 0 -0.02 -0.03 -0.02 0 -0.02 -0.02 ...
                       : num
## $ accel_forearm_x
                       : int 192 192 196 189 189 193 195 193 193 190 ...
## $ accel forearm y
                             203 203 204 206 206 203 205 205 204 205 ...
                       : int
                       : int
## $ accel forearm z
                             -215 -216 -213 -214 -214 -215 -215 -213 -214 -215 ...
## $ magnet_forearm_x
                             -17 -18 -18 -16 -17 -9 -18 -9 -16 -22 ...
                       : int
## $ magnet_forearm_y
                       : num
                             654 661 658 658 655 660 659 660 653 656 ...
## $ magnet_forearm_z
                             476 473 469 469 473 478 470 474 476 473 ...
                       : num
                       : Factor w/ 5 levels "A", "B", "C", "D", ...: 1 1 1 1 1 1 1 1 1 1 ...
   $ classe
library(ggplot2)
```

library(caret)

glm

We'll try a glm (binomial) model (logit). Maybe this does a bunch of models? make sure we do the preprocessing within the train statement, or else it messes up cross validation (We should split the data into folds and conduct PCA within each fold, rather than transforming via PCA and then splitting).