Practical Machine Learning Project

Alfredo Aranda Nunez

2 de julio de 2018

Table of Contents

The goal of this project is to predict the manner in which people did the exercise

We read the training and test variables using the "readr" library

```
# Reading the data
library(readr)

## Warning: package 'readr' was built under R version 3.4.3

pml_training <- read_csv("pml-training.csv")

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:
## cols(
##    .default = col_double(),
##    X1 = col_integer(),
##    user_name = col_character(),
##    raw_timestamp_part_1 = col_integer(),
##    raw_timestamp_part_2 = col_integer(),
##    cvtd_timestamp = col_character(),
##    new_window = col_character(),
##    num_window = col_integer(),
##    total_accel_belt = col_integer(),
##    kurtosis_roll_belt = col_character(),
##    kurtosis_picth_belt = col_character(),
##    kurtosis_yaw_belt = col_character(),
##    skewness_roll_belt = col_character(),
##    skewness_roll_belt.1 = col_character(),
##    skewness_yaw_belt = col_character(),
##    max_picth_belt = col_integer(),
##    max_yaw_belt = col_character(),
##    min_pitch_belt = col_integer(),
##    min_yaw_belt = col_character(),
##    amplitude_pitch_belt = col_integer(),
##    amplitude_yaw_belt = col_character()
##    # ... with 46 more columns
## )

## See spec(...) for full column specifications.

## Warning in rbind(names(probs), probs_f): number of columns of result i
s not
## a multiple of vector length (arg 1)

## Warning: 185 parsing failures.
## row # A tibble: 5 x 5 col      row                  col expected  actual
file expected    <int>             <chr>     <chr>   <chr>               <ch
r> actual 1  2231 kurtosis_roll_arm a double #DIV/0! 'pml-training.csv' f
ile 2  2231 skewness_roll_arm a double #DIV/0! 'pml-training.csv' row 3
2255 kurtosis_roll_arm a double #DIV/0! 'pml-training.csv' col 4  2255 sk
```

```
ewness_roll_arm a double #DIV/0! 'pml-training.csv' expected 5  2282 kurt
osis_roll_arm a double #DIV/0! 'pml-training.csv'
## ... ................ ... ........................................
................ ........... ......................................
.............. ......... ......................................
............ .... ......................................
..... .. ...................................................... .
.. .......................................................... .......
................................................................
## See problems(...) for more details.

pml_testing <- read_csv("pml-testing.csv")

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:
## cols(
##   .default = col_character(),
##   X1 = col_integer(),
##   raw_timestamp_part_1 = col_integer(),
##   raw_timestamp_part_2 = col_integer(),
##   num_window = col_integer(),
##   roll_belt = col_double(),
##   pitch_belt = col_double(),
##   yaw_belt = col_double(),
##   total_accel_belt = col_integer(),
##   gyros_belt_x = col_double(),
##   gyros_belt_y = col_double(),
##   gyros_belt_z = col_double(),
##   accel_belt_x = col_integer(),
##   accel_belt_y = col_integer(),
##   accel_belt_z = col_integer(),
##   magnet_belt_x = col_integer(),
##   magnet_belt_y = col_integer(),
##   magnet_belt_z = col_integer(),
##   roll_arm = col_double(),
##   pitch_arm = col_double(),
##   yaw_arm = col_double()
##   # ... with 37 more columns
## )
## See spec(...) for full column specifications.

# Knowing the data
nrow(pml_training)

## [1] 19622

nrow(pml_testing)

## [1] 20

summary(pml_training)
```

```
##        X1             user_name         raw_timestamp_part_1
##  Min.    :    1   Length:19622       Min.    :1.322e+09
##  1st Qu.: 4906   Class :character   1st Qu.:1.323e+09
##  Median : 9812   Mode  :character   Median :1.323e+09
##  Mean    : 9812                      Mean    :1.323e+09
##  3rd Qu.:14717                       3rd Qu.:1.323e+09
##  Max.    :19622                      Max.    :1.323e+09
##
##  raw_timestamp_part_2 cvtd_timestamp       new_window
##  Min.    :    294      Length:19622       Length:19622
##  1st Qu.:252912        Class :character   Class :character
##  Median :496380        Mode  :character   Mode  :character
##  Mean    :500656
##  3rd Qu.:751891
##  Max.    :998801
##
##    num_window      roll_belt        pitch_belt          yaw_belt
##  Min.    :  1.0   Min.    :-28.90   Min.    :-55.8000   Min.    :-180.00
##  1st Qu.:222.0   1st Qu.:  1.10   1st Qu.:  1.7600   1st Qu.: -88.30
##  Median :424.0   Median :113.00   Median :  5.2800   Median : -13.00
##  Mean    :430.6   Mean    : 64.41   Mean    :  0.3053   Mean    : -11.21
##  3rd Qu.:644.0   3rd Qu.:123.00   3rd Qu.: 14.9000   3rd Qu.:  12.90
##  Max.    :864.0   Max.    :162.00   Max.    : 60.3000   Max.    : 179.00
##
##  total_accel_belt kurtosis_roll_belt kurtosis_picth_belt
##  Min.    : 0.00   Length:19622       Length:19622
##  1st Qu.: 3.00   Class :character   Class :character
##  Median :17.00   Mode  :character   Mode  :character
##  Mean    :11.31
##  3rd Qu.:18.00
##  Max.    :29.00
##
##  kurtosis_yaw_belt  skewness_roll_belt skewness_roll_belt.1
##  Length:19622        Length:19622        Length:19622
##  Class :character    Class :character    Class :character
##  Mode  :character    Mode  :character    Mode  :character
##
##
##
##
##  skewness_yaw_belt  max_roll_belt      max_picth_belt   max_yaw_belt
##  Length:19622        Min.    :-94.300   Min.    : 3.00   Length:19622
##  Class :character    1st Qu.:-88.000   1st Qu.: 5.00   Class :character
##  Mode  :character    Median : -5.100   Median :18.00   Mode  :character
##                      Mean    : -6.667   Mean    :12.92
##                      3rd Qu.: 18.500   3rd Qu.:19.00
##                      Max.    :180.000   Max.    :30.00
##                      NA's    :19216   NA's    :19216
##  min_roll_belt     min_pitch_belt   min_yaw_belt       amplitude_roll_b
elt
```

```
##  Min.   :-180.00   Min.    : 0.00   Length:19622      Min.    :   0.000
##  1st Qu.: -88.40   1st Qu.: 3.00   Class :character   1st Qu.:   0.300
##  Median :  -7.85   Median :16.00   Mode  :character   Median :   1.000
##  Mean   : -10.44   Mean    :10.76                      Mean    :   3.769
##  3rd Qu.:   9.05   3rd Qu.:17.00                       3rd Qu.:   2.083
##  Max.   : 173.00   Max.    :23.00                      Max.    : 360.000
##  NA's   :19216     NA's    :19216                      NA's    :19216
##  amplitude_pitch_belt amplitude_yaw_belt var_total_accel_belt
##  Min.   : 0.000      Length:19622        Min.    : 0.000
##  1st Qu.: 1.000      Class :character    1st Qu.: 0.100
##  Median : 1.000      Mode  :character    Median : 0.200
##  Mean   : 2.167                          Mean    : 0.926
##  3rd Qu.: 2.000                          3rd Qu.: 0.300
##  Max.   :12.000                          Max.    :16.500
##  NA's   :19216                           NA's    :19216
##  avg_roll_belt     stddev_roll_belt var_roll_belt      avg_pitch_belt
##  Min.   :-27.40    Min.    : 0.000   Min.    :  0.000   Min.   :-51.400
##  1st Qu.:  1.10    1st Qu.: 0.200   1st Qu.:  0.000   1st Qu.:  2.025
##  Median :116.35    Median : 0.400   Median :  0.100   Median :  5.200
##  Mean   : 68.06    Mean    : 1.337   Mean    :  7.699   Mean    :  0.520
##  3rd Qu.:123.38    3rd Qu.: 0.700   3rd Qu.:  0.500   3rd Qu.: 15.775
##  Max.   :157.40    Max.    :14.200   Max.    :200.700   Max.    : 59.700
##  NA's   :19216     NA's    :19216   NA's    :19216   NA's    :19216
##  stddev_pitch_belt var_pitch_belt    avg_yaw_belt       stddev_yaw_belt
##  Min.   :0.000     Min.    : 0.000   Min.    :-138.300   Min.    :  0.000
##  1st Qu.:0.200     1st Qu.: 0.000   1st Qu.: -88.175   1st Qu.:  0.100
##  Median :0.400     Median : 0.100   Median :  -6.550   Median :  0.300
##  Mean   :0.603     Mean    : 0.766   Mean    :  -8.831   Mean    :  1.341
##  3rd Qu.:0.700     3rd Qu.: 0.500   3rd Qu.:  14.125   3rd Qu.:  0.700
##  Max.   :4.000     Max.    :16.200   Max.    : 173.500   Max.    :176.600
##  NA's   :19216     NA's    :19216   NA's    :19216   NA's    :19216
##   var_yaw_belt        gyros_belt_x        gyros_belt_y
##  Min.   :    0.000   Min.    :-1.040000   Min.    :-0.64000
##  1st Qu.:    0.010   1st Qu.:-0.030000   1st Qu.: 0.00000
##  Median :    0.090   Median : 0.030000   Median : 0.02000
##  Mean   :  107.487   Mean    :-0.005592   Mean    : 0.03959
##  3rd Qu.:    0.475   3rd Qu.: 0.110000   3rd Qu.: 0.11000
##  Max.   :31183.240   Max.    : 2.220000   Max.    : 0.64000
##  NA's   :19216
##   gyros_belt_z       accel_belt_x       accel_belt_y      accel_belt_z
##  Min.   :-1.4600   Min.    :-120.000   Min.    :-69.00   Min.    :-275.00
##  1st Qu.:-0.2000   1st Qu.: -21.000   1st Qu.:  3.00   1st Qu.:-162.00
##  Median :-0.1000   Median : -15.000   Median : 35.00   Median :-152.00
##  Mean   :-0.1305   Mean    :  -5.595   Mean    : 30.15   Mean    : -72.59
##  3rd Qu.:-0.0200   3rd Qu.:  -5.000   3rd Qu.: 61.00   3rd Qu.:  27.00
##  Max.   : 1.6200   Max.    :  85.000   Max.    :164.00   Max.    : 105.00
##
##  magnet_belt_x     magnet_belt_y    magnet_belt_z      roll_arm
##  Min.   :-52.0    Min.    :354.0    Min.    :-623.0    Min.    :-180.00
##  1st Qu.:  9.0    1st Qu.:581.0    1st Qu.:-375.0    1st Qu.: -31.77
```

```
##  Median : 35.0   Median :601.0   Median :-320.0   Median :   0.00
##  Mean   : 55.6   Mean   :593.7   Mean   :-345.5   Mean   :  17.83
##  3rd Qu.: 59.0   3rd Qu.:610.0   3rd Qu.:-306.0   3rd Qu.:  77.30
##  Max.   :485.0   Max.   :673.0   Max.   : 293.0   Max.   : 180.00
##
##    pitch_arm         yaw_arm          total_accel_arm var_accel_arm
##  Min.   :-88.800   Min.   :-180.0000   Min.   : 1.00   Min.   :  0.00
##  1st Qu.:-25.900   1st Qu.: -43.1000   1st Qu.:17.00   1st Qu.:  9.03
##  Median :  0.000   Median :   0.0000   Median :27.00   Median : 40.61
##  Mean   : -4.612   Mean   :  -0.6188   Mean   :25.51   Mean   : 53.23
##  3rd Qu.: 11.200   3rd Qu.:  45.8750   3rd Qu.:33.00   3rd Qu.: 75.62
##  Max.   : 88.500   Max.   : 180.0000   Max.   :66.00   Max.   :331.70
##                                                        NA's   :19216
##   avg_roll_arm     stddev_roll_arm    var_roll_arm        avg_pitch_arm
##  Min.   :-166.67   Min.   :  0.000   Min.   :    0.000   Min.   :-81.7
73
##  1st Qu.: -38.37   1st Qu.:  1.376   1st Qu.:    1.898   1st Qu.:-22.7
70
##  Median :   0.00   Median :  5.702   Median :   32.517   Median :  0.0
00
##  Mean   :  12.68   Mean   : 11.201   Mean   :  417.264   Mean   : -4.9
01
##  3rd Qu.:  76.33   3rd Qu.: 14.921   3rd Qu.:  222.647   3rd Qu.:  8.2
77
##  Max.   : 163.33   Max.   :161.964   Max.   :26232.208   Max.   : 75.6
59
##  NA's   :19216     NA's   :19216     NA's   :19216       NA's   :19216
## stddev_pitch_arm var_pitch_arm      avg_yaw_arm        stddev_yaw_arm
##  Min.   : 0.000   Min.   :   0.000   Min.   :-173.440   Min.   :  0.00
0
##  1st Qu.: 1.642   1st Qu.:   2.697   1st Qu.: -29.198   1st Qu.:  2.57
7
##  Median : 8.133   Median :  66.146   Median :   0.000   Median : 16.68
2
##  Mean   :10.383   Mean   : 195.864   Mean   :   2.359   Mean   : 22.27
0
##  3rd Qu.:16.327   3rd Qu.: 266.576   3rd Qu.:  38.185   3rd Qu.: 35.98
4
##  Max.   :43.412   Max.   :1884.565   Max.   : 152.000   Max.   :177.04
4
##  NA's   :19216    NA's   :19216      NA's   :19216       NA's   :19216
##   var_yaw_arm        gyros_arm_x        gyros_arm_y
##  Min.   :    0.000   Min.   :-6.37000   Min.   :-3.4400
##  1st Qu.:    6.642   1st Qu.:-1.33000   1st Qu.:-0.8000
##  Median :  278.309   Median : 0.08000   Median :-0.2400
##  Mean   : 1055.933   Mean   : 0.04277   Mean   :-0.2571
##  3rd Qu.: 1294.850   3rd Qu.: 1.57000   3rd Qu.: 0.1400
##  Max.   :31344.568   Max.   : 4.87000   Max.   : 2.8400
##  NA's   :19216
##   gyros_arm_z        accel_arm_x        accel_arm_y        accel_arm_z
```

```
##   Min.   :-2.3300   Min.   :-404.00   Min.   :-318.0   Min.   :-636.00
##   1st Qu.:-0.0700   1st Qu.:-242.00   1st Qu.: -54.0   1st Qu.:-143.00
##   Median : 0.2300   Median : -44.00   Median :  14.0   Median : -47.00
##   Mean   : 0.2695   Mean   : -60.24   Mean   :  32.6   Mean   : -71.25
##   3rd Qu.: 0.7200   3rd Qu.:  84.00   3rd Qu.: 139.0   3rd Qu.:  23.00
##   Max.   : 3.0200   Max.   : 437.00   Max.   : 308.0   Max.   : 292.00
##
##   magnet_arm_x     magnet_arm_y     magnet_arm_z    kurtosis_roll_arm
##   Min.   :-584.0   Min.   :-392.0   Min.   :-597.0   Min.   :-1.809
##   1st Qu.:-300.0   1st Qu.:  -9.0   1st Qu.: 131.2   1st Qu.:-1.345
##   Median : 289.0   Median : 202.0   Median : 444.0   Median :-0.894
##   Mean   : 191.7   Mean   : 156.6   Mean   : 306.5   Mean   :-0.366
##   3rd Qu.: 637.0   3rd Qu.: 323.0   3rd Qu.: 545.0   3rd Qu.:-0.038
##   Max.   : 782.0   Max.   : 583.0   Max.   : 694.0   Max.   :21.456
##                                                      NA's   :19294
##  kurtosis_picth_arm kurtosis_yaw_arm  skewness_roll_arm
##  Length:19622       Length:19622      Min.   :-2.541
##  Class :character   Class :character  1st Qu.:-0.561
##  Mode  :character   Mode  :character  Median : 0.040
##                                       Mean   : 0.068
##                                       3rd Qu.: 0.671
##                                       Max.   : 4.394
##                                       NA's   :19293
##  skewness_pitch_arm skewness_yaw_arm   max_roll_arm
##  Length:19622       Length:19622      Min.   :-73.100
##  Class :character   Class :character  1st Qu.: -0.175
##  Mode  :character   Mode  :character  Median :  4.950
##                                       Mean   : 11.236
##                                       3rd Qu.: 26.775
##                                       Max.   : 85.500
##                                       NA's   :19216
##  max_picth_arm      max_yaw_arm     min_roll_arm     min_pitch_arm
##  Min.   :-173.000   Min.   : 4.00   Min.   :-89.10   Min.   :-180.00
##  1st Qu.:  -1.975   1st Qu.:29.00   1st Qu.:-41.98   1st Qu.: -72.62
##  Median :  23.250   Median :34.00   Median :-22.45   Median : -33.85
##  Mean   :  35.751   Mean   :35.46   Mean   :-21.22   Mean   : -33.92
##  3rd Qu.:  95.975   3rd Qu.:41.00   3rd Qu.:  0.00   3rd Qu.:   0.00
##  Max.   : 180.000   Max.   :65.00   Max.   : 66.40   Max.   : 152.00
##  NA's   :19216      NA's   :19216   NA's   :19216    NA's   :19216
##   min_yaw_arm     amplitude_roll_arm amplitude_pitch_arm amplitude_yaw_
## arm
##  Min.   : 1.00   Min.   :  0.000    Min.   :  0.000    Min.   : 0.00
##  1st Qu.: 8.00   1st Qu.:  5.425    1st Qu.:  9.925    1st Qu.:13.00
##  Median :13.00   Median : 28.450    Median : 54.900    Median :22.00
##  Mean   :14.66   Mean   : 32.452    Mean   : 69.677    Mean   :20.79
##  3rd Qu.:19.00   3rd Qu.: 50.960    3rd Qu.:115.175    3rd Qu.:28.75
##  Max.   :38.00   Max.   :119.500    Max.   :360.000    Max.   :52.00
##  NA's   :19216   NA's   :19216      NA's   :19216      NA's   :19216
##  roll_dumbbell     pitch_dumbbell     yaw_dumbbell
##  Min.   :-153.71   Min.   :-149.59   Min.   :-150.871
```

```
##    1st Qu.: -18.49   1st Qu.: -40.89   1st Qu.: -77.644
##    Median :  48.17   Median : -20.96   Median :  -3.324
##    Mean   :  23.84   Mean   : -10.78   Mean   :   1.674
##    3rd Qu.:  67.61   3rd Qu.:  17.50   3rd Qu.:  79.643
##    Max.   : 153.55   Max.   : 149.40   Max.   : 154.952
##
##  kurtosis_roll_dumbbell kurtosis_picth_dumbbell kurtosis_yaw_dumbbell
##  Min.   :-2.174         Min.   :-2.200          Length:19622
##  1st Qu.:-0.682         1st Qu.:-0.721          Class :character
##  Median :-0.033         Median :-0.133          Mode  :character
##  Mean   : 0.452         Mean   : 0.286
##  3rd Qu.: 0.940         3rd Qu.: 0.584
##  Max.   :54.998         Max.   :55.628
##  NA's   :19221          NA's   :19218
##  skewness_roll_dumbbell skewness_pitch_dumbbell skewness_yaw_dumbbell
##  Min.   :-7.384         Min.   :-7.447          Length:19622
##  1st Qu.:-0.581         1st Qu.:-0.526          Class :character
##  Median :-0.076         Median :-0.091          Mode  :character
##  Mean   :-0.115         Mean   :-0.035
##  3rd Qu.: 0.400         3rd Qu.: 0.505
##  Max.   : 1.958         Max.   : 3.769
##  NA's   :19220          NA's   :19217
##  max_roll_dumbbell max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbe
## ll
##  Min.   :-70.10    Min.   :-112.90    Min.   :-2.20    Min.   :-149.60
##  1st Qu.:-27.15    1st Qu.: -66.70    1st Qu.:-0.70    1st Qu.: -59.67
##  Median : 14.85    Median :  40.05    Median : 0.00    Median : -43.55
##  Mean   : 13.76    Mean   :  32.75    Mean   : 0.45    Mean   : -41.24
##  3rd Qu.: 50.58    3rd Qu.: 133.22    3rd Qu.: 0.90    3rd Qu.: -25.20
##  Max.   :137.00    Max.   : 155.00    Max.   :55.00    Max.   :  73.20
##  NA's   :19216     NA's   :19216      NA's   :19221    NA's   :19216
##  min_pitch_dumbbell min_yaw_dumbbell amplitude_roll_dumbbell
##  Min.   :-147.00    Min.   :-2.20    Min.   :  0.00
##  1st Qu.: -91.80    1st Qu.:-0.70    1st Qu.: 14.97
##  Median : -66.15    Median : 0.00    Median : 35.05
##  Mean   : -33.18    Mean   : 0.45    Mean   : 55.00
##  3rd Qu.:  21.20    3rd Qu.: 0.90    3rd Qu.: 81.04
##  Max.   : 120.90    Max.   :55.00    Max.   :256.48
##  NA's   :19216      NA's   :19221    NA's   :19216
##  amplitude_pitch_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
##  Min.   :  0.00           Min.   :0              Min.   : 0.00
##  1st Qu.: 17.06           1st Qu.:0              1st Qu.: 4.00
##  Median : 41.73           Median :0              Median :10.00
##  Mean   : 65.93           Mean   :0              Mean   :13.72
##  3rd Qu.: 99.55           3rd Qu.:0              3rd Qu.:19.00
##  Max.   :273.59           Max.   :0              Max.   :58.00
##  NA's   :19216            NA's   :19221
##  var_accel_dumbbell avg_roll_dumbbell stddev_roll_dumbbell
##  Min.   :  0.000    Min.   :-128.96   Min.   :  0.000
##  1st Qu.:  0.378    1st Qu.: -12.33   1st Qu.:  4.639
```

```
##   Median :  1.000    Median :  48.23   Median : 12.204
##   Mean   :  4.388    Mean   :  23.86   Mean   : 20.761
##   3rd Qu.:  3.434    3rd Qu.:  64.37   3rd Qu.: 26.356
##   Max.   :230.428    Max.   : 125.99   Max.   :123.778
##   NA's   :19216      NA's   :19216     NA's   :19216
##   var_roll_dumbbell  avg_pitch_dumbbell stddev_pitch_dumbbell
##   Min.   :   0.00    Min.   :-70.73    Min.   : 0.000
##   1st Qu.:  21.52    1st Qu.:-42.00    1st Qu.: 3.482
##   Median :  148.95   Median :-19.91    Median : 8.089
##   Mean   : 1020.27   Mean   :-12.33    Mean   :13.147
##   3rd Qu.:  694.65   3rd Qu.: 13.21    3rd Qu.:19.238
##   Max.   :15321.01   Max.   : 94.28    Max.   :82.680
##   NA's   :19216      NA's   :19216     NA's   :19216
##   var_pitch_dumbbell avg_yaw_dumbbell   stddev_yaw_dumbbell
##   Min.   :   0.00    Min.   :-117.950  Min.   :  0.000
##   1st Qu.:  12.12    1st Qu.: -76.696  1st Qu.:  3.885
##   Median :  65.44    Median :  -4.505  Median : 10.264
##   Mean   : 350.31    Mean   :   0.202  Mean   : 16.647
##   3rd Qu.: 370.11    3rd Qu.:  71.234  3rd Qu.: 24.674
##   Max.   :6836.02    Max.   : 134.905  Max.   :107.088
##   NA's   :19216      NA's   :19216     NA's   :19216
##   var_yaw_dumbbell   gyros_dumbbell_x    gyros_dumbbell_y
##   Min.   :   0.00    Min.   :-204.0000  Min.   :-2.10000
##   1st Qu.:  15.09    1st Qu.:  -0.0300  1st Qu.:-0.14000
##   Median :  105.35   Median :   0.1300  Median : 0.03000
##   Mean   : 589.84    Mean   :   0.1611  Mean   : 0.04606
##   3rd Qu.: 608.79    3rd Qu.:   0.3500  3rd Qu.: 0.21000
##   Max.   :11467.91   Max.   :   2.2200  Max.   :52.00000
##   NA's   :19216
##   gyros_dumbbell_z  accel_dumbbell_x  accel_dumbbell_y  accel_dumbbell_
## z
##   Min.   : -2.380   Min.   :-419.00   Min.   :-189.00   Min.   :-334.00
##   1st Qu.: -0.310   1st Qu.: -50.00   1st Qu.:  -8.00   1st Qu.:-142.00
##   Median : -0.130   Median :  -8.00   Median :  41.50   Median :  -1.00
##   Mean   : -0.129   Mean   : -28.62   Mean   :  52.63   Mean   : -38.32
##   3rd Qu.:  0.030   3rd Qu.:  11.00   3rd Qu.: 111.00   3rd Qu.:  38.00
##   Max.   :317.000   Max.   : 235.00   Max.   : 315.00   Max.   : 318.00
##
##   magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z  roll_forearm
##   Min.   :-643.0    Min.   :-3600     Min.   :-262.00   Min.   :-180.00
## 00
##   1st Qu.:-535.0    1st Qu.:  231     1st Qu.: -45.00   1st Qu.:  -0.73
## 75
##   Median :-479.0    Median :  311     Median :  13.00   Median :  21.70
## 00
##   Mean   :-328.5    Mean   :  221     Mean   :  46.05   Mean   :  33.82
## 65
##   3rd Qu.:-304.0    3rd Qu.:  390     3rd Qu.:  95.00   3rd Qu.: 140.00
## 00
##   Max.   : 592.0    Max.   :  633     Max.   : 452.00   Max.   : 180.00
```

```
00
##                                         NA's   :1
##  pitch_forearm     yaw_forearm       kurtosis_roll_forearm
##  Min.   :-72.50   Min.   :-180.00   Length:19622
##  1st Qu.:  0.00   1st Qu.: -68.60   Class :character
##  Median :  9.24   Median :   0.00   Mode  :character
##  Mean   : 10.71   Mean   :  19.21
##  3rd Qu.: 28.40   3rd Qu.: 110.00
##  Max.   : 89.80   Max.   : 180.00
##
##  kurtosis_picth_forearm kurtosis_yaw_forearm skewness_roll_forearm
##  Length:19622           Length:19622         Length:19622
##  Class :character        Class :character     Class :character
##  Mode  :character        Mode  :character     Mode  :character
##
##
##
##
##  skewness_pitch_forearm skewness_yaw_forearm max_roll_forearm
##  Length:19622           Length:19622         Min.   :-66.60
##  Class :character        Class :character     1st Qu.:  0.00
##  Mode  :character        Mode  :character     Median : 26.80
##                                               Mean   : 24.49
##                                               3rd Qu.: 45.95
##                                               Max.   : 89.80
##                                               NA's   :19216
##  max_picth_forearm max_yaw_forearm    min_roll_forearm  min_pitch_fore
arm
##  Min.   :-151.00   Length:19622       Min.   :-72.500   Min.   :-180.0
0
##  1st Qu.:   0.00   Class :character   1st Qu.: -6.075   1st Qu.:-175.0
0
##  Median : 113.00   Mode  :character   Median :  0.000   Median : -61.0
0
##  Mean   :  81.49                      Mean   : -0.167   Mean   : -57.5
7
##  3rd Qu.: 174.75                      3rd Qu.: 12.075   3rd Qu.:   0.0
0
##  Max.   : 180.00                      Max.   : 62.100   Max.   : 167.0
0
##  NA's   :19216                        NA's   :19216     NA's   :19216
##  min_yaw_forearm    amplitude_roll_forearm amplitude_pitch_forearm
##  Length:19622      Min.   :  0.000        Min.   :  0.0
##  Class :character   1st Qu.:  1.125        1st Qu.:  2.0
##  Mode  :character   Median : 17.770        Median : 83.7
##                     Mean   : 24.653        Mean   :139.1
##                     3rd Qu.: 39.875        3rd Qu.:350.0
##                     Max.   :126.000        Max.   :360.0
##                     NA's   :19216          NA's   :19216
##  amplitude_yaw_forearm total_accel_forearm var_accel_forearm
```

```
##   Length:19622          Min.   :  0.00     Min.   :  0.000
##   Class :character      1st Qu.: 29.00     1st Qu.:  6.759
##   Mode  :character      Median : 36.00     Median : 21.165
##                         Mean   : 34.72     Mean   : 33.502
##                         3rd Qu.: 41.00     3rd Qu.: 51.240
##                         Max.   :108.00     Max.   :172.606
##                                            NA's   :19216
##   avg_roll_forearm    stddev_roll_forearm var_roll_forearm
##   Min.   :-177.234    Min.   :  0.000     Min.   :    0.00
##   1st Qu.:  -0.909    1st Qu.:  0.428     1st Qu.:    0.18
##   Median :  11.172    Median :  8.030     Median :   64.48
##   Mean   :  33.165    Mean   : 41.986     Mean   : 5274.10
##   3rd Qu.: 107.132    3rd Qu.: 85.373     3rd Qu.: 7289.08
##   Max.   : 177.256    Max.   :179.171     Max.   :32102.24
##   NA's   :19216       NA's   :19216       NA's   :19216
##   avg_pitch_forearm stddev_pitch_forearm var_pitch_forearm
##   Min.   :-68.17    Min.   : 0.000       Min.   :   0.000
##   1st Qu.:  0.00    1st Qu.: 0.336       1st Qu.:   0.113
##   Median : 12.02    Median : 5.516       Median :  30.425
##   Mean   : 11.79    Mean   : 7.977       Mean   : 139.593
##   3rd Qu.: 28.48    3rd Qu.:12.866       3rd Qu.: 165.532
##   Max.   : 72.09    Max.   :47.745       Max.   :2279.617
##   NA's   :19216     NA's   :19216        NA's   :19216
##   avg_yaw_forearm   stddev_yaw_forearm var_yaw_forearm    gyros_forearm
_x
##   Min.   :-155.06   Min.   :  0.000    Min.   :    0.00   Min.   :-22.0
00
##   1st Qu.: -26.26   1st Qu.:  0.524    1st Qu.:    0.27   1st Qu.: -0.2
20
##   Median :   0.00   Median : 24.743    Median :  612.21   Median :  0.0
50
##   Mean   :  18.00   Mean   : 44.854    Mean   : 4639.85   Mean   :  0.1
58
##   3rd Qu.:  85.79   3rd Qu.: 85.817    3rd Qu.: 7368.41   3rd Qu.:  0.5
60
##   Max.   : 169.24   Max.   :197.508    Max.   :39009.33   Max.   :  3.9
70
##   NA's   :19216     NA's   :19216      NA's   :19216
##   gyros_forearm_y    gyros_forearm_z    accel_forearm_x    accel_forear
m_y
##   Min.   : -7.02000  Min.   : -8.0900   Min.   :-498.00    Min.   :-632
.0
##   1st Qu.: -1.46000  1st Qu.: -0.1800   1st Qu.:-178.00    1st Qu.:  57
.0
##   Median :  0.03000  Median :  0.0800   Median : -57.00    Median : 201
.0
##   Mean   :  0.07517  Mean   :  0.1512   Mean   : -61.65    Mean   : 163
.7
##   3rd Qu.:  1.62000  3rd Qu.:  0.4900   3rd Qu.:  76.00    3rd Qu.: 312
.0
```

```
## Max.   :311.00000   Max.   :231.0000   Max.   : 477.00   Max.   : 923
.0
##
## accel_forearm_z   magnet_forearm_x   magnet_forearm_y magnet_forearm_z
## Min.   :-446.00   Min.   :-1280.0   Min.   :-896.0   Min.   :-973.0
## 1st Qu.:-182.00   1st Qu.: -616.0   1st Qu.:   2.0   1st Qu.: 191.0
## Median : -39.00   Median : -378.0   Median : 591.0   Median : 511.0
## Mean   : -55.29   Mean   : -312.6   Mean   : 380.1   Mean   : 393.6
## 3rd Qu.:  26.00   3rd Qu.:  -73.0   3rd Qu.: 737.0   3rd Qu.: 653.0
## Max.   : 291.00   Max.   :  672.0   Max.   :1480.0   Max.   :1090.0
##                                      NA's   :1        NA's   :1
## 　   classe
## Length:19622
## Class :character
## Mode  :character
##
##
##
##
```

```r
# What kind of classes do we have?
table(pml_training$classe)
```

```
##
##    A    B    C    D    E
## 5580 3797 3422 3216 3607
```

To know the behavior of the variables, a significance statistic test is implemented (ANOVA test)

```r
# To do an ANOVA statistiv test, we must change the "classe" variable wit
h number 1,2,3,4 and 5
                for(i in 1:nrow(pml_training)){
                if(pml_training$classe[i] == "A"){
                  pml_training$new_classe[i] <- 1
                          }
                if(pml_training$classe[i] == "B"){
                  pml_training$new_classe[i] <- 2
                          }
                if(pml_training$classe[i] == "C"){
                  pml_training$new_classe[i] <- 3
                          }
                if(pml_training$classe[i] == "D"){
                  pml_training$new_classe[i] <- 4
                          }
                if(pml_training$classe[i] == "E"){
                  pml_training$new_classe[i] <- 5
                          }
                          }

                      # Exploratory analysis
```

```r
table(pml_training$new_classe)
hist(pml_training$new_classe, xlab = "Classe Variable", main = "Histogram
                      of Classe variable")
```

Histogram of Classe variable
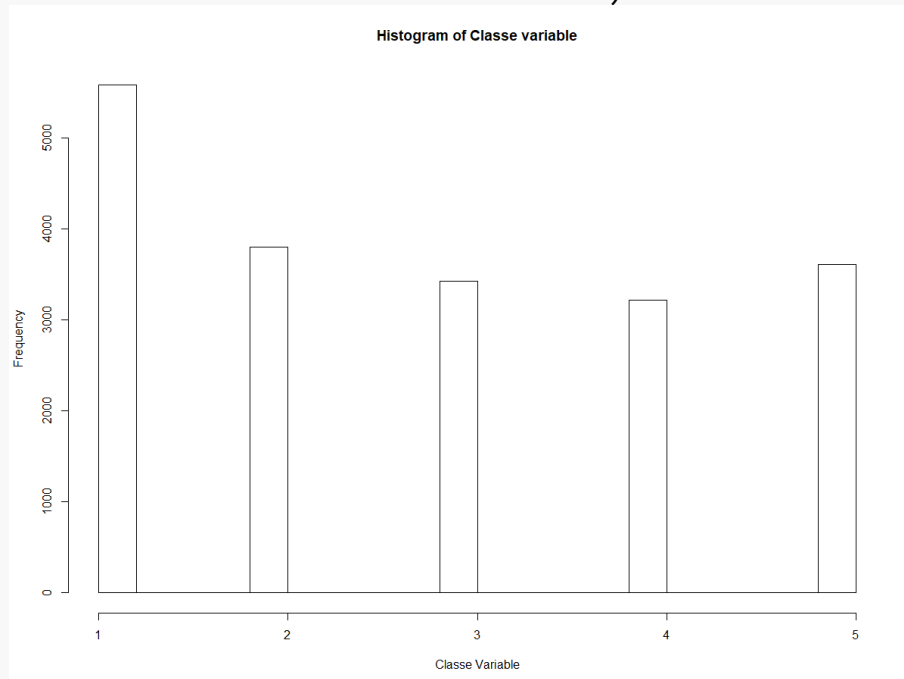

```r
# Selecting just the numeric columns
library("dplyr")
pml_training_numeric <- select_if(pml_training, is.numeric)


# Removing columns with NA
not_any_na <- function(x) all(!is.na(x))
pml_training_numeric_WNA <- pml_training_numeric %>% select_if(not_any_na
                              )
pml_training_numeric_WNA$new_classe_factor <- as.factor(pml_training_nume
                      ric_WNA$new_classe)
pml_training_numeric_WNA_filter <- pml_training_numeric_WNA[-1]
pml_training_numeric_WNA_filter <- pml_training_numeric_WNA_filter[-54]


# ANOVA test
aov <- aov(new_classe_factor ~ ., data = pml_training_numeric_WNA_filter)
summary(aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
raw_timestamp_part_1	1	24	23.9	19.013	1.30e-05	***
raw_timestamp_part_2	1	11	10.6	8.413	0.003729	**
num_window	1	16	16.0	12.780	0.000351	***
roll_belt	1	158	157.5	125.452	< 2e-16	***
pitch_belt	1	27	26.8	21.357	3.84e-06	***
yaw_belt	1	529	529.4	421.534	< 2e-16	***
total_accel_belt	1	324	324.3	258.262	< 2e-16	***
gyros_belt_x	1	92	91.8	73.109	< 2e-16	***

gyros_belt_y	1	107	107.3	85.414	< 2e-16	***
gyros_belt_z	1	158	158.3	126.058	< 2e-16	***
accel_belt_x	1	222	222.2	176.912	< 2e-16	***
accel_belt_y	1	2725	2724.7	2169.662	< 2e-16	***
accel_belt_z	1	687	687.4	547.374	< 2e-16	***
magnet_belt_x	1	48	48.2	38.375	5.96e-10	***
magnet_belt_y	1	1537	1536.6	1223.563	< 2e-16	***
magnet_belt_z	1	102	101.8	81.057	< 2e-16	***
roll_arm	1	399	399.2	317.842	< 2e-16	***
pitch_arm	1	1001	1000.6	796.725	< 2e-16	***
yaw_arm	1	4	4.2	3.341	0.067609	.
total_accel_arm	1	431	431.5	343.579	< 2e-16	***
gyros_arm_x	1	5	5.3	4.183	0.040856	*
gyros_arm_y	1	8	7.6	6.090	0.013605	*
gyros_arm_z	1	2	2.2	1.728	0.188706	
accel_arm_x	1	1537	1536.6	1223.592	< 2e-16	***
accel_arm_y	1	73	73.2	58.303	2.35e-14	***
accel_arm_z	1	939	938.9	747.619	< 2e-16	***
magnet_arm_x	1	392	391.8	311.966	< 2e-16	***
magnet_arm_y	1	371	371.4	295.772	< 2e-16	***
magnet_arm_z	1	520	520.3	414.332	< 2e-16	***
roll_dumbbell	1	178	178.4	142.019	< 2e-16	***
pitch_dumbbell	1	126	126.2	100.519	< 2e-16	***
yaw_dumbbell	1	961	960.7	765.022	< 2e-16	***
total_accel_dumbbell	1	179	178.9	142.426	< 2e-16	***
gyros_dumbbell_x	1	1	1.4	1.131	0.287585	
gyros_dumbbell_y	1	19	19.3	15.371	8.86e-05	***
gyros_dumbbell_z	1	1	1.3	1.072	0.300446	
accel_dumbbell_x	1	196	196.0	156.056	< 2e-16	***
accel_dumbbell_y	1	3	3.0	2.375	0.123315	
accel_dumbbell_z	1	8	7.7	6.147	0.013170	*
magnet_dumbbell_x	1	86	86.1	68.534	< 2e-16	***
magnet_dumbbell_y	1	17	17.1	13.619	0.000225	***
roll_forearm	1	104	104.0	82.839	< 2e-16	***
pitch_forearm	1	1944	1943.8	1547.793	< 2e-16	***
yaw_forearm	1	1	0.9	0.733	0.391820	
total_accel_forearm	1	1348	1348.3	1073.612	< 2e-16	***
gyros_forearm_x	1	9	9.1	7.247	0.007110	**
gyros_forearm_y	1	12	12.0	9.564	0.001988	**
gyros_forearm_z	1	1	0.9	0.687	0.407072	
accel_forearm_x	1	45	44.7	35.632	2.42e-09	***
accel_forearm_y	1	72	72.3	57.606	3.35e-14	***
accel_forearm_z	1	376	376.0	299.430	< 2e-16	***
magnet_forearm_x	1	3	3.2	2.564	0.109353	

```
Residuals                 19569   24575      1.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If we analyze the p value results, we can set our significance statistic variables (rejecting the null hypothesis, P_value <= 0.001) in the training model.

The model was built using the significant variables and a cross-validation test was used for assessing how the results of a statistical analysis will generalize to an independent data set. The goal of cross-validation is to estimate the expected level of fit of a model to a data set that is independent of the data that were used to train the model. It can be used to estimate any quantitative measure of fit that is appropriate for the data and model. For example, for binary classification problems, each case in the validation set is either predicted correctly or incorrectly. In this situation the misclassification error rate can be used to summarize the fit, although other measures like positive predictive value could also be used. When the value being predicted is continuously distributed, the mean squared error, root mean squared error or median absolute deviation could be used to summarize the errors.

```
# Models to predict the "Classe" variable
library(caret)
set.seed(30334)
trControl <- trainControl(method = "cv", number = 3)
pml_training_numeric_WNA_filter$new_classe_factor <- as.character(pml_tra
ining_numeric_WNA_filter$new_classe)
pml_training_numeric_WNA_filter <- pml_training_numeric_WNA_filter[-53]

# Random Forest
set.seed(30334)

rf <- train(new_classe_factor ~ ., data = pml_training_numeric_WNA_filter
,
            method = "rf", prox = TRUE, trControl = trControl)
printRandom Forest
```

```
19622 samples
   52 predictor
    5 classes: '1', '2', '3', '4', '5'


No pre-processing
Resampling: Cross-Validated (3 fold)
Summary of sample sizes: 13082, 13081, 13081
Resampling results across tuning parameters:
```

```
 mtry   Accuracy      Kappa
   2    0.9964835   0.9955517
  27    0.9989807   0.9987108
  52    0.9982163   0.9977439
```
Accuracy was used to select the optimal model using the largest value.
The final value used for the model was mtry = 27. (rf)
print(rf$results)

```
  mtry  Accuracy       Kappa     AccuracySD        KappaSD
1    2 0.9964835 0.9955517 0.0012137181 0.0015356398
2   27 0.9989807 0.9987108 0.0001764922 0.0002232666
3   52 0.9982163 0.9977439 0.0006892976 0.0008718879
```
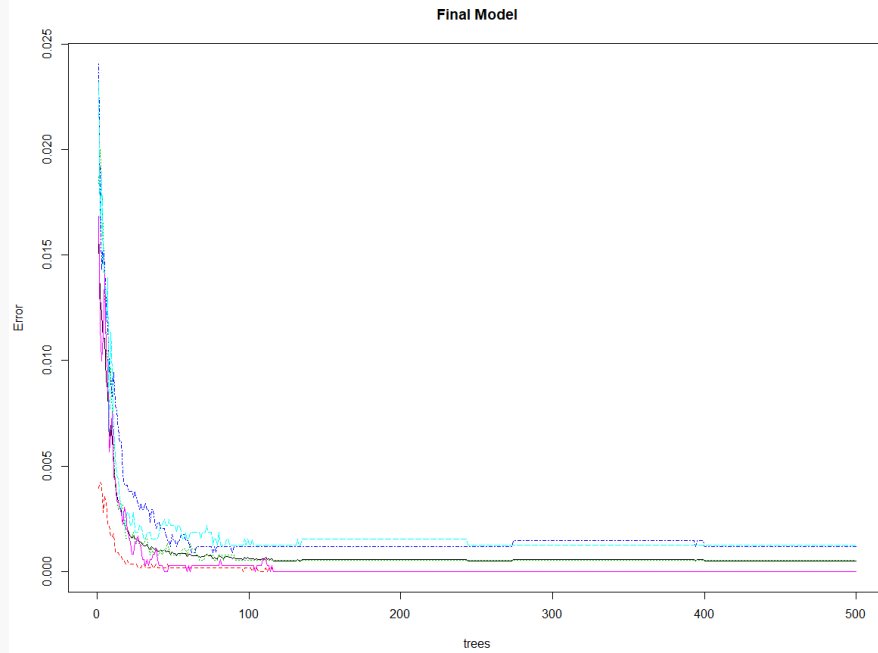
Accuracy cross validations respect to randomly selected predictors:
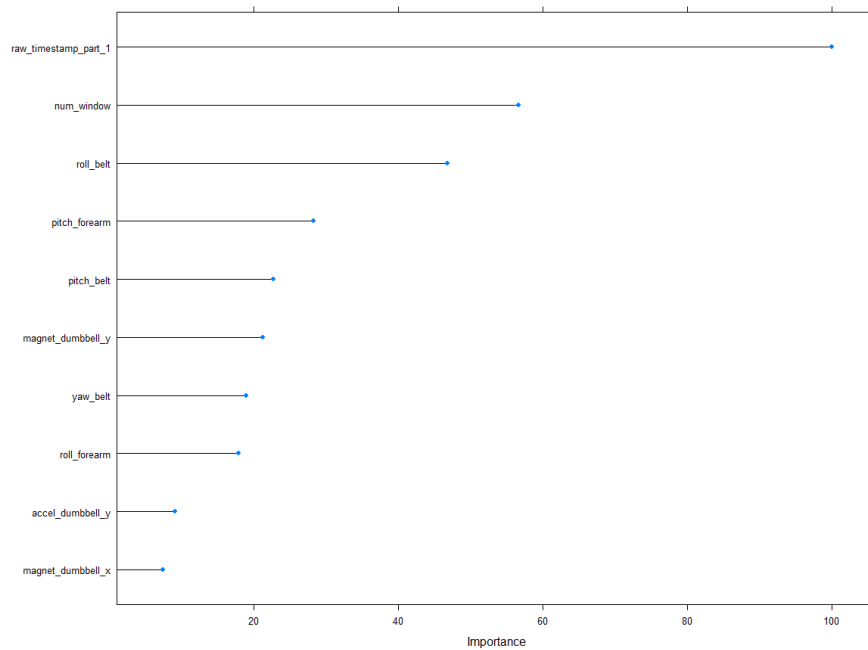plot(rf, pch=19,lty=6, lwd=2)
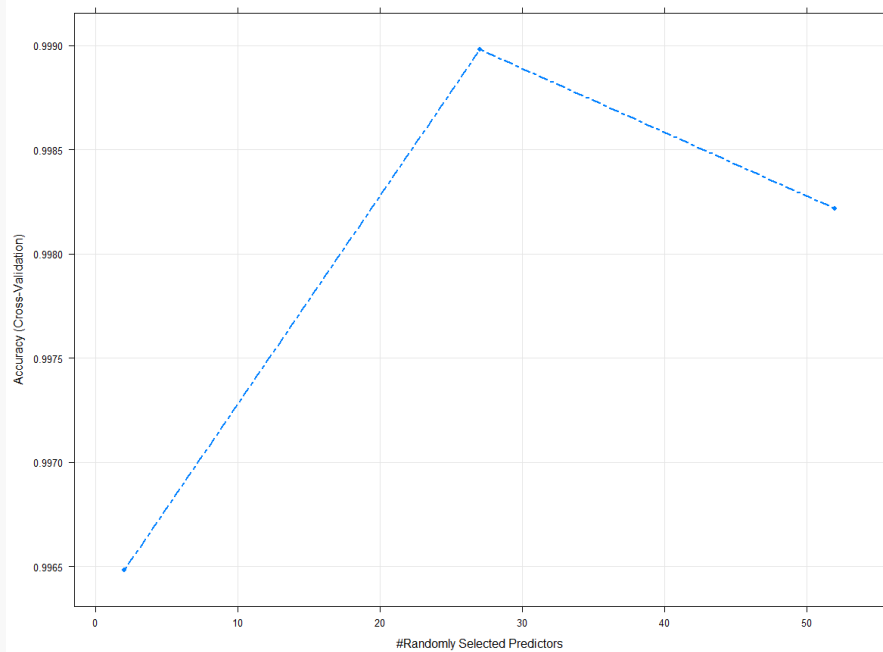
Final model errors:
plot(rf$finalModel, main = "Final Model")

Final Model
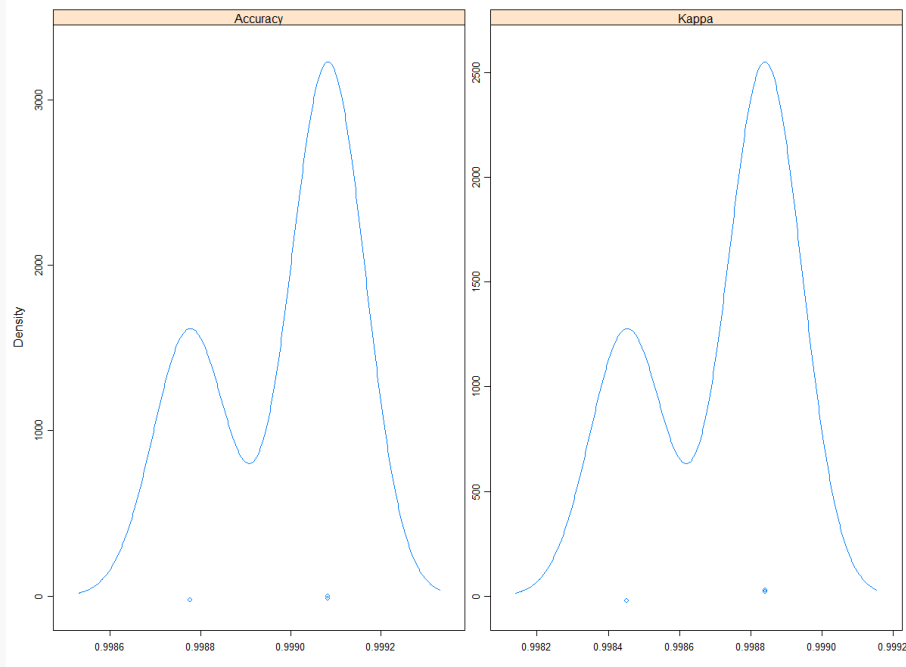
Importance of variables, top 10:

```
plot(varImp(rf), top = 10)
```


resample Histogram:

```
resampleHist((rf))
```

```
pred_train <- predict(rf, newdata = pml_training)
pml_training$new_classe_factor <- as.factor(pml_training$new_classe)
confusionMatrix(pred_train, pml_training$new_classe_factor)
```

Confusion Matrix and Statistics

```
          Reference
Prediction    1    2    3    4    5
         1 5580    0    0    0    0
         2    0 3797    0    0    0
         3    0    0 3422    0    0
         4    0    0    0 3216    0
         5    0    0    0    0 3607
```

Overall Statistics

```
              Accuracy : 1
                95% CI : (0.9998, 1)
   No Information Rate : 0.2844
   P-Value [Acc > NIR] : < 2.2e-16

                 Kappa : 1
 Mcnemar's Test P-Value : NA
```

Statistics by Class:

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	1.0000	1.0000	1.0000	1.0000	1.0000
Specificity	1.0000	1.0000	1.0000	1.0000	1.0000
Pos Pred Value	1.0000	1.0000	1.0000	1.0000	1.0000
Neg Pred Value	1.0000	1.0000	1.0000	1.0000	1.0000
Prevalence	0.2844	0.1935	0.1744	0.1639	0.1838
Detection Rate	0.2844	0.1935	0.1744	0.1639	0.1838
Detection Prevalence	0.2844	0.1935	0.1744	0.1639	0.1838
Balanced Accuracy	1.0000	1.0000	1.0000	1.0000	1.0000

We observe that it can possibly exist an overtfit in the training model, however, this one was trained with the statical significant variables, analyzed in the ANOVa test, and then, a cross validation test was used. So, the results of the prediction are shown as follow:

```
# Predictions
pred <- predict(rf, newdata = pml_testing)
# Results

[1] 2 1 2 1 1 5 4 2 1 1 2 3 2 1 5 5 1 2 2 2

Levels: 1 2 3 4 5
```