Project - Practical Machine Learning

David Berkowitz

March 3, 2015

```
library (caret, quietly=TRUE)
library (randomForest, quietly=TRUE)
set.seed (123)
```

Read in the data

```
## [1] "RDS exists, reading it"
## [1] "."
```

Feature selection (reduce the number of columns)

Many of the columns are statistical calculations (min_, max_, avg_, stddev_, var_, skewness_, kurtosis_) on the raw data measurements. I chose to build my classifier only on raw data columns: gyro_, accel_ and magnet_.

```
names = names (data.raw) # get the list of column names


gyros = grep ("^gyros", names)
accel = grep ("^accel", names)
magnet = grep ("^magnet", names)
class = grep ("^classe", names) # add the activity column
user = grep ("^user", names) # add the user name column


data = data.raw [, c(gyros, accel, magnet, class, user)] # only include wanted columns
data = na.omit (data) # omit NA values which are not appreciated by later functions
```

Partition the data

```
set.seed (12345)
dp = createDataPartition (y = data$classe, p = 0.6, list=FALSE)
myTraining = data [dp,] # training set has random 60%
myTesting = data [-dp, ] # testing set has the remaining 40%


a = nrow (data) ; b = nrow (myTraining) ; c = nrow (myTesting)
check = a - b - c
cbind (a, b, c, check) # check should equal 0
```

```
##             a      b      c check
## [1,] 19622 11776 7846      0
```

Create the classifier based on my training set

```
rf = randomForest (classe ~ ., data=myTraining); rf
```

```
##
## Call:
##  randomForest(formula = classe ~ ., data = myTraining)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 6
##
##          OOB estimate of  error rate: 1.55%
## Confusion matrix:
##      A    B    C    D    E class.error
## A 3341    2    1    3    1 0.002090800
## B   34 2223   21    0    1 0.024572181
## C    2   24 2026    2    0 0.013631938
## D    8    3   58 1858    3 0.037305699
## E    0    4    5   10 2146 0.008775982
```

Evaluate the classifier based on my testing set

```
predictions = predict (rf, myTesting)
confusionMatrix (predictions, myTesting$classe)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 2226   22    2    5    2
##          B    2 1484   18    0    0
##          C    0   12 1345   40    7
##          D    4    0    2 1237    6
##          E    0    0    1    4 1427
##
## Overall Statistics
##
##                Accuracy : 0.9838
##                  95% CI : (0.9808, 0.9865)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9795
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                    Class: A Class: B Class: C Class: D Class: E
## Sensitivity          0.9973   0.9776   0.9832   0.9619   0.9896
## Specificity          0.9945   0.9968   0.9909   0.9982   0.9992
## Pos Pred Value       0.9863   0.9867   0.9580   0.9904   0.9965
## Neg Pred Value       0.9989   0.9946   0.9964   0.9926   0.9977
## Prevalence           0.2845   0.1935   0.1744   0.1639   0.1838
## Detection Rate       0.2837   0.1891   0.1714   0.1577   0.1819
## Detection Prevalence 0.2877   0.1917   0.1789   0.1592   0.1825
## Balanced Accuracy    0.9959   0.9872   0.9870   0.9800   0.9944
```

This simple predictor does quite well with 98.38% accuracy, which is in line with the OOB estimated error rate of 1.55%.

Predict based on the real testing set

```
testingData = read.csv ("pml-testing.csv")
testingData = testingData [, c(gyros, accel, magnet, class, user)] # same dataset filtering a
s before
testingData = na.omit (testingData)

predictions = predict (rf, testingData); predictions
```

```
##    1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##    B  A  B  A  A  E  D  B  A  A  B  C  B  A  E  E  A  B  B  B
## Levels: A B C D E
```

This "simple" classifier was sufficient to correctly predict all 20 values for the course project submission.

Cross Validation

```
table (data$user_name, data$classe)
```

```
##
##              A     B     C     D     E
##    adelmo   1165   776   750   515   686
##    carlitos  834   690   493   486   609
##    charles   899   745   539   642   711
##    eurico    865   592   489   582   542
##    jeremy   1177   489   652   522   562
##    pedro     640   505   499   469   497
```

```
#ggplot (data, aes (x=classe)) + geom_bar() + facet_wrap (~ user_name)
```

There appear to be sufficient data points for each activity for each user. I will cross validate by user_name.

```
crossValidate = function (name)
  {
  testingRowNumbers = training = testing = NULL
  a = b = c = check = NULL
  rf = cm = NULL

  testingRowNumbers = grep (name, data$user_name)
  training = data [-testingRowNumbers,]
  testing = data [testingRowNumbers,]

  a = nrow (data) ; b = nrow (training) ; c = nrow (testing); check = a - b - c
  cbind (a, b, c, check) # check should equal 0

  rf = randomForest (classe ~ ., data=training)
  predictions = predict (rf, testing)
  cm = confusionMatrix (predictions, testing$classe)
  cat (name, "accuracy= ", cm$overall ["Accuracy"], "\n")
  return (rf);
  }
```

```
rf = crossValidate ("adelmo"); rf
```

```
## adelmo accuracy=  0.2834018
```

```
##
## Call:
##  randomForest(formula = classe ~ ., data = training)
##                Type of random forest: classification
##                     Number of trees: 500
## No. of variables tried at each split: 6
##
##         OOB estimate of  error rate: 0.83%
## Confusion matrix:
##      A    B    C    D    E class.error
## A 4409    1    0    4    1 0.001359003
## B   24 2977   20    0    0 0.014564714
## C    3   14 2655    0    0 0.006362275
## D    4    0   46 2649    2 0.019252129
## E    0    2    2    7 2910 0.003765834
```

```
rf = crossValidate ("carlitos"); rf
```

```
## carlitos accuracy=  0.5780848
```

```
##
## Call:
##  randomForest(formula = classe ~ ., data = training)
##                Type of random forest: classification
##                     Number of trees: 500
## No. of variables tried at each split: 6
##
##         OOB estimate of  error rate: 0.79%
## Confusion matrix:
##      A    B    C    D    E class.error
## A 4741    2    0    2    1 0.001053519
## B   14 3079   13    1    0 0.009011909
## C    2   25 2901    1    0 0.009559577
## D    5    0   54 2667    4 0.023076923
## E    0    1    0    5 2992 0.002001334
```

```
rf = crossValidate ("charles"); rf
```

```
## charles accuracy=  0.5616516
```

```
##
## Call:
##   randomForest(formula = classe ~ ., data = training)
##                  Type of random forest: classification
##                        Number of trees: 500
## No. of variables tried at each split: 6
##
##         OOB estimate of  error rate: 0.75%
## Confusion matrix:
##      A    B    C    D    E class.error
## A 4673    2    0    5    1 0.001709037
## B   18 3023   11    0    0 0.009501966
## C    2   20 2861    0    0 0.007630940
## D    5    0   46 2521    2 0.020590521
## E    0    1    2    6 2887 0.003107735
```

```
rf = crossValidate ("eurico"); rf
```

```
## eurico accuracy=  0.1765472
```

```
##
## Call:
##   randomForest(formula = classe ~ ., data = training)
##                  Type of random forest: classification
##                        Number of trees: 500
## No. of variables tried at each split: 6
##
##         OOB estimate of  error rate: 0.9%
## Confusion matrix:
##      A    B    C    D    E class.error
## A 4710    1    0    4    0 0.001060445
## B   21 3166   18    0    0 0.012168487
## C    3   25 2905    0    0 0.009546539
## D    3    0   61 2567    3 0.025436598
## E    0    0    3    7 3055 0.003262643
```

```
rf = crossValidate ("jeremy"); rf
```

```
## jeremy accuracy=  0.5684891
```

```
##
## Call:
##   randomForest(formula = classe ~ ., data = training)
##                 Type of random forest: classification
##                       Number of trees: 500
## No. of variables tried at each split: 6
##
##           OOB estimate of  error rate: 0.86%
## Confusion matrix:
##       A    B    C    D    E class.error
## A 4392    2    3    5    1 0.002498297
## B   24 3264   20    0    0 0.013301088
## C    2   20 2747    1    0 0.008303249
## D    4    0   50 2636    4 0.021529324
## E    0    1    0    3 3041 0.001313629
```

```
rf = crossValidate ("pedro"); rf
```

```
## pedro accuracy=  0.1938697
```

```
##
## Call:
##   randomForest(formula = classe ~ ., data = training)
##                 Type of random forest: classification
##                       Number of trees: 500
## No. of variables tried at each split: 6
##
##           OOB estimate of  error rate: 0.81%
## Confusion matrix:
##       A    B    C    D    E class.error
## A 4935    2    0    2    1 0.001012146
## B   18 3252   21    1    0 0.012150668
## C    0   19 2904    0    0 0.006500171
## D    1    1   59 2682    4 0.023662177
## E    0    1    2    6 3101 0.002893891
```

Results

```
cbind (names, oobErrorRates, accuracy)
```

```
##      names       oobErrorRates accuracy
## [1,] "adelmo"   "0.82"        "0.27"
## [2,] "carlitos" "0.84"        "0.58"
## [3,] "charles"  "0.82"        "0.58"
## [4,] "eurico"   "0.93"        "0.18"
## [5,] "jeremy"   "0.83"        "0.55"
## [6,] "pedro"    "0.81"        "0.22"
```

```
mean (oobErrorRates) # average expected error rate (percentage)
```

```
## [1] 0.8416667
```

```
mean (accuracy) # average measured accuracy
```

```
## [1] 0.3966667
```

Further Investigations

After the later video lectures I wanted to start over and apply some of the techniques discussed. Here are the results.

Read in the data set

```
setwd ("/Users/davidberkowitz/Projects/coursera/Practical Machine Learning")

data = read.csv ("pml-training.csv",
                 header=TRUE,
                 as.is = TRUE,
                 stringsAsFactors = FALSE,
                 #sep=',',
                 )
dim (data)
```

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

```
summary (data)
```

```
##        X              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_belt
##  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.   :-81.773
## 1st Qu.: -38.37   1st Qu.:  1.376   1st Qu.:    1.898   1st Qu.:-22.770
## Median :   0.00   Median :  5.702   Median :   32.517   Median :  0.000
## Mean   :  12.68   Mean   : 11.201   Mean   :  417.264   Mean   : -4.901
## 3rd Qu.:  76.33   3rd Qu.: 14.921   3rd Qu.:  222.647   3rd Qu.:  8.277
## Max.   : 163.33   Max.   :161.964   Max.   :26232.208   Max.   : 75.659
## 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.   :  0.000
## 1st Qu.: 1.642   1st Qu.:   2.697   1st Qu.: -29.198   1st Qu.:  2.577
## Median : 8.133   Median :  66.146   Median :   0.000   Median : 16.682
## Mean   :10.383   Mean   : 195.864   Mean   :   2.359   Mean   : 22.270
## 3rd Qu.:16.327   3rd Qu.: 266.576   3rd Qu.:  38.185   3rd Qu.: 35.984
## Max.   :43.412   Max.   :1884.565   Max.   : 152.000   Max.   :177.044
## 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    Length:19622
```

```
##   1st Qu.:-300.0   1st Qu.: -9.0   1st Qu.: 131.2   Class :character
##   Median : 289.0   Median : 202.0   Median : 444.0    Mode  :character
##   Mean   : 191.7   Mean   : 156.6   Mean   : 306.5
##   3rd Qu.: 637.0   3rd Qu.: 323.0   3rd Qu.: 545.0
##   Max.   : 782.0   Max.   : 583.0   Max.   : 694.0
##
##   kurtosis_picth_arm kurtosis_yaw_arm   skewness_roll_arm
##   Length:19622        Length:19622       Length:19622
##   Class :character    Class :character   Class :character
##   Mode  :character    Mode  :character   Mode  :character
##
##
##
##
##   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
```

```
##    Length:19622                Length:19622               Length:19622
##    Class :character            Class :character           Class :character
##    Mode  :character            Mode  :character           Mode  :character
## 
## 
## 
## 
##    skewness_roll_dumbbell skewness_pitch_dumbbell skewness_yaw_dumbbell
##    Length:19622                Length:19622               Length:19622
##    Class :character            Class :character           Class :character
##    Mode  :character            Mode  :character           Mode  :character
## 
## 
## 
## 
##    max_roll_dumbbell max_picth_dumbbell max_yaw_dumbbell    min_roll_dumbbell
##    Min.   :-70.10     Min.   :-112.90    Length:19622        Min.   :-149.60
##    1st Qu.:-27.15     1st Qu.: -66.70    Class :character    1st Qu.: -59.67
##    Median : 14.85     Median :  40.05    Mode  :character    Median : -43.55
##    Mean   : 13.76     Mean   :  32.75                        Mean   : -41.24
##    3rd Qu.: 50.58     3rd Qu.: 133.22                        3rd Qu.: -25.20
##    Max.   :137.00     Max.   : 155.00                        Max.   :  73.20
##    NA's   :19216      NA's   :19216                          NA's   :19216
##    min_pitch_dumbbell min_yaw_dumbbell    amplitude_roll_dumbbell
##    Min.   :-147.00     Length:19622        Min.   :  0.00
##    1st Qu.: -91.80     Class :character    1st Qu.: 14.97
##    Median : -66.15     Mode  :character    Median : 35.05
##    Mean   : -33.18                         Mean   : 55.00
##    3rd Qu.:  21.20                         3rd Qu.: 81.04
##    Max.   : 120.90                         Max.   :256.48
##    NA's   :19216                           NA's   :19216
##    amplitude_pitch_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
##    Min.   :  0.00           Length:19622            Min.   : 0.00
##    1st Qu.: 17.06           Class :character        1st Qu.: 4.00
##    Median : 41.73           Mode  :character        Median :10.00
##    Mean   : 65.93                                   Mean   :13.72
##    3rd Qu.: 99.55                                   3rd Qu.:19.00
##    Max.   :273.59                                   Max.   :58.00
##    NA's   :19216
##    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.   :-180.0000
##    1st Qu.:-535.0    1st Qu.:  231     1st Qu.: -45.00    1st Qu.:  -0.7375
##    Median :-479.0    Median :  311     Median :  13.00    Median :  21.7000
##    Mean   :-328.5    Mean   :  221     Mean   :  46.05    Mean   :  33.8265
##    3rd Qu.:-304.0    3rd Qu.:  390     3rd Qu.:  95.00    3rd Qu.: 140.0000
##    Max.   : 592.0    Max.   :  633     Max.   : 452.00    Max.   : 180.0000
##
##    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_forearm
##  Min.   :-151.00   Length:19622       Min.   :-72.500   Min.   :-180.00
##  1st Qu.:   0.00   Class :character   1st Qu.: -6.075   1st Qu.:-175.00
##  Median : 113.00   Mode  :character   Median :  0.000   Median : -61.00
##  Mean   :  81.49                      Mean   : -0.167   Mean   : -57.57
##  3rd Qu.: 174.75                      3rd Qu.: 12.075   3rd Qu.:   0.00
##  Max.   : 180.00                      Max.   : 62.100   Max.   : 167.00
##  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.   :-22.000
##  1st Qu.: -26.26   1st Qu.:  0.524    1st Qu.:    0.27  1st Qu.: -0.220
##  Median :   0.00   Median : 24.743    Median :  612.21  Median :  0.050
##  Mean   :  18.00   Mean   : 44.854    Mean   : 4639.85  Mean   :  0.158
##  3rd Qu.:  85.79   3rd Qu.: 85.817    3rd Qu.: 7368.41  3rd Qu.:  0.560
##  Max.   : 169.24   Max.   :197.508    Max.   :39009.33  Max.   :  3.970
##  NA's   :19216     NA's   :19216      NA's   :19216
##  gyros_forearm_y   gyros_forearm_z   accel_forearm_x   accel_forearm_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
##
##     classe
##  Length:19622
##  Class :character
##  Mode  :character
##
##
##
##
```

There are 160 columns in the raw data set.

Remove columns with NAs

```
NAindex = apply (data, 2, function(x) {sum(is.na(x))})
NAindex
```

```
##                            X              user_name     raw_timestamp_part_1
##                            0                      0                        0
##         raw_timestamp_part_2         cvtd_timestamp               new_window
##                            0                      0                        0
##                   num_window              roll_belt               pitch_belt
##                            0                      0                        0
##                     yaw_belt        total_accel_belt       kurtosis_roll_belt
##                            0                      0                        0
##          kurtosis_picth_belt       kurtosis_yaw_belt       skewness_roll_belt
##                            0                      0                        0
##         skewness_roll_belt.1        skewness_yaw_belt             max_roll_belt
##                            0                      0                    19216
##                max_picth_belt            max_yaw_belt            min_roll_belt
##                        19216                      0                    19216
##                min_pitch_belt             min_yaw_belt       amplitude_roll_belt
##                        19216                      0                    19216
##           amplitude_pitch_belt     amplitude_yaw_belt     var_total_accel_belt
##                        19216                      0                    19216
##                 avg_roll_belt         stddev_roll_belt            var_roll_belt
##                        19216                  19216                    19216
##                avg_pitch_belt        stddev_pitch_belt           var_pitch_belt
##                        19216                  19216                    19216
##                  avg_yaw_belt          stddev_yaw_belt             var_yaw_belt
##                        19216                  19216                    19216
##                  gyros_belt_x            gyros_belt_y             gyros_belt_z
##                            0                      0                        0
##                  accel_belt_x            accel_belt_y             accel_belt_z
##                            0                      0                        0
##                 magnet_belt_x           magnet_belt_y            magnet_belt_z
##                            0                      0                        0
##                     roll_arm               pitch_arm                 yaw_arm
##                            0                      0                        0
##               total_accel_arm           var_accel_arm             avg_roll_arm
##                            0                  19216                    19216
##               stddev_roll_arm             var_roll_arm            avg_pitch_arm
##                        19216                  19216                    19216
##              stddev_pitch_arm            var_pitch_arm              avg_yaw_arm
##                        19216                  19216                    19216
##                stddev_yaw_arm              var_yaw_arm              gyros_arm_x
##                        19216                  19216                        0
##                   gyros_arm_y              gyros_arm_z              accel_arm_x
##                            0                      0                        0
##                   accel_arm_y              accel_arm_z             magnet_arm_x
```

```
##                            0                         0                         0
##                 magnet_arm_y              magnet_arm_z          kurtosis_roll_arm
##                            0                         0                         0
##            kurtosis_picth_arm          kurtosis_yaw_arm          skewness_roll_arm
##                            0                         0                         0
##            skewness_pitch_arm          skewness_yaw_arm               max_roll_arm
##                            0                         0                     19216
##                 max_picth_arm               max_yaw_arm               min_roll_arm
##                        19216                     19216                     19216
##                 min_pitch_arm               min_yaw_arm          amplitude_roll_arm
##                        19216                     19216                     19216
##           amplitude_pitch_arm        amplitude_yaw_arm               roll_dumbbell
##                        19216                     19216                         0
##                pitch_dumbbell             yaw_dumbbell       kurtosis_roll_dumbbell
##                            0                         0                         0
##     kurtosis_picth_dumbbell     kurtosis_yaw_dumbbell     skewness_roll_dumbbell
##                            0                         0                         0
##     skewness_pitch_dumbbell     skewness_yaw_dumbbell           max_roll_dumbbell
##                            0                         0                     19216
##            max_picth_dumbbell        max_yaw_dumbbell           min_roll_dumbbell
##                        19216                         0                     19216
##            min_pitch_dumbbell        min_yaw_dumbbell     amplitude_roll_dumbbell
##                        19216                         0                     19216
## amplitude_pitch_dumbbell     amplitude_yaw_dumbbell        total_accel_dumbbell
##                        19216                         0                         0
##            var_accel_dumbbell        avg_roll_dumbbell        stddev_roll_dumbbell
##                        19216                     19216                     19216
##             var_roll_dumbbell       avg_pitch_dumbbell       stddev_pitch_dumbbell
##                        19216                     19216                     19216
##            var_pitch_dumbbell         avg_yaw_dumbbell        stddev_yaw_dumbbell
##                        19216                     19216                     19216
##             var_yaw_dumbbell          gyros_dumbbell_x          gyros_dumbbell_y
##                        19216                         0                         0
##             gyros_dumbbell_z          accel_dumbbell_x          accel_dumbbell_y
##                            0                         0                         0
##             accel_dumbbell_z         magnet_dumbbell_x         magnet_dumbbell_y
##                            0                         0                         0
##            magnet_dumbbell_z               roll_forearm              pitch_forearm
##                            0                         0                         0
##                 yaw_forearm     kurtosis_roll_forearm     kurtosis_picth_forearm
##                            0                         0                         0
##        kurtosis_yaw_forearm     skewness_roll_forearm     skewness_pitch_forearm
##                            0                         0                         0
##        skewness_yaw_forearm           max_roll_forearm          max_picth_forearm
##                            0                     19216                     19216
##             max_yaw_forearm           min_roll_forearm          min_pitch_forearm
##                            0                     19216                     19216
```

```
##          min_yaw_forearm   amplitude_roll_forearm amplitude_pitch_forearm
##                        0                    19216                   19216
##    amplitude_yaw_forearm       total_accel_forearm          var_accel_forearm
##                        0                        0                   19216
##          avg_roll_forearm         stddev_roll_forearm          var_roll_forearm
##                    19216                    19216                   19216
##         avg_pitch_forearm        stddev_pitch_forearm         var_pitch_forearm
##                    19216                    19216                   19216
##           avg_yaw_forearm          stddev_yaw_forearm           var_yaw_forearm
##                    19216                    19216                   19216
##           gyros_forearm_x             gyros_forearm_y           gyros_forearm_z
##                        0                        0                        0
##           accel_forearm_x             accel_forearm_y           accel_forearm_z
##                        0                        0                        0
##          magnet_forearm_x            magnet_forearm_y          magnet_forearm_z
##                        0                        0                        0
##                    classe
##                        0
```

```
data = data [,which (NAindex == 0)]


dim (data)
```

```
## [1] 19622     93
```

```
summary (data)
```

```
##        X            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_yaw_belt       min_yaw_belt
##  Length:19622       Length:19622       Length:19622
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##  amplitude_yaw_belt gyros_belt_x        gyros_belt_y
##  Length:19622       Min.   :-1.040000   Min.   :-0.64000
##  Class :character   1st Qu.:-0.030000   1st Qu.: 0.00000
##  Mode  :character   Median : 0.030000   Median : 0.02000
##                     Mean   :-0.005592   Mean   : 0.03959
##                     3rd Qu.: 0.110000   3rd Qu.: 0.11000
##                     Max.   : 2.220000   Max.   : 0.64000
##   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 gyros_arm_x
##  Min.   :-88.800   Min.   :-180.0000   Min.   : 1.00   Min.   :-6.37000
```

```
##   1st Qu.:-25.900    1st Qu.: -43.1000   1st Qu.:17.00   1st Qu.:-1.33000
##   Median :  0.000    Median :   0.0000   Median :27.00   Median : 0.08000
##   Mean   : -4.612    Mean   :  -0.6188   Mean   :25.51   Mean   : 0.04277
##   3rd Qu.: 11.200    3rd Qu.:  45.8750   3rd Qu.:33.00   3rd Qu.: 1.57000
##   Max.   : 88.500    Max.   : 180.0000   Max.   :66.00   Max.   : 4.87000
##   gyros_arm_y        gyros_arm_z         accel_arm_x        accel_arm_y
##   Min.   :-3.4400    Min.   :-2.3300     Min.   :-404.00    Min.   :-318.0
##   1st Qu.:-0.8000    1st Qu.:-0.0700     1st Qu.:-242.00    1st Qu.: -54.0
##   Median :-0.2400    Median : 0.2300     Median : -44.00    Median :  14.0
##   Mean   :-0.2571    Mean   : 0.2695     Mean   : -60.24    Mean   :  32.6
##   3rd Qu.: 0.1400    3rd Qu.: 0.7200     3rd Qu.:  84.00    3rd Qu.: 139.0
##   Max.   : 2.8400    Max.   : 3.0200     Max.   : 437.00    Max.   : 308.0
##   accel_arm_z        magnet_arm_x        magnet_arm_y       magnet_arm_z
##   Min.   :-636.00    Min.   :-584.0      Min.   :-392.0     Min.   :-597.0
##   1st Qu.:-143.00    1st Qu.:-300.0      1st Qu.:  -9.0     1st Qu.: 131.2
##   Median : -47.00    Median : 289.0      Median : 202.0     Median : 444.0
##   Mean   : -71.25    Mean   : 191.7      Mean   : 156.6     Mean   : 306.5
##   3rd Qu.:  23.00    3rd Qu.: 637.0      3rd Qu.: 323.0     3rd Qu.: 545.0
##   Max.   : 292.00    Max.   : 782.0      Max.   : 583.0     Max.   : 694.0
##   kurtosis_roll_arm  kurtosis_picth_arm  kurtosis_yaw_arm
##   Length:19622       Length:19622        Length:19622
##   Class :character   Class :character    Class :character
##   Mode  :character   Mode  :character    Mode  :character
##
##
##
##   skewness_roll_arm  skewness_pitch_arm  skewness_yaw_arm
##   Length:19622       Length:19622        Length:19622
##   Class :character   Class :character    Class :character
##   Mode  :character   Mode  :character    Mode  :character
##
##
##
##   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
##   Length:19622           Length:19622            Length:19622
##   Class :character       Class :character        Class :character
##   Mode  :character       Mode  :character        Mode  :character
##
##
##
```

```
##    skewness_roll_dumbbell skewness_pitch_dumbbell skewness_yaw_dumbbell
##    Length:19622           Length:19622            Length:19622
##    Class :character        Class :character         Class :character
##    Mode  :character        Mode  :character         Mode  :character
##
##
##
##    max_yaw_dumbbell    min_yaw_dumbbell    amplitude_yaw_dumbbell
##    Length:19622        Length:19622        Length:19622
##    Class :character    Class :character    Class :character
##    Mode  :character    Mode  :character    Mode  :character
##
##
##
##    total_accel_dumbbell gyros_dumbbell_x      gyros_dumbbell_y
##    Min.   : 0.00        Min.   :-204.0000    Min.   :-2.10000
##    1st Qu.: 4.00        1st Qu.:  -0.0300    1st Qu.:-0.14000
##    Median :10.00        Median :   0.1300    Median : 0.03000
##    Mean   :13.72        Mean   :   0.1611    Mean   : 0.04606
##    3rd Qu.:19.00        3rd Qu.:   0.3500    3rd Qu.: 0.21000
##    Max.   :58.00        Max.   :   2.2200    Max.   :52.00000
##    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.   :-180.0000
##    1st Qu.:-535.0    1st Qu.:  231     1st Qu.: -45.00   1st Qu.:  -0.7375
##    Median :-479.0    Median :  311     Median :  13.00   Median :  21.7000
##    Mean   :-328.5    Mean   :  221     Mean   :  46.05   Mean   :  33.8265
##    3rd Qu.:-304.0    3rd Qu.:  390     3rd Qu.:  95.00   3rd Qu.: 140.0000
##    Max.   : 592.0    Max.   :  633     Max.   : 452.00   Max.   : 180.0000
##    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_yaw_forearm
##  Length:19622            Length:19622           Length:19622
##  Class :character        Class :character       Class :character
##  Mode  :character        Mode  :character       Mode  :character
##
##
##
##  min_yaw_forearm     amplitude_yaw_forearm total_accel_forearm
##  Length:19622        Length:19622          Min.   :  0.00
##  Class :character    Class :character      1st Qu.: 29.00
##  Mode  :character    Mode  :character      Median : 36.00
##                                            Mean   : 34.72
##                                            3rd Qu.: 41.00
##                                            Max.   :108.00
##  gyros_forearm_x    gyros_forearm_y      gyros_forearm_z
##  Min.   :-22.000    Min.   : -7.02000    Min.   : -8.0900
##  1st Qu.: -0.220    1st Qu.: -1.46000    1st Qu.: -0.1800
##  Median :  0.050    Median :  0.03000    Median :  0.0800
##  Mean   :  0.158    Mean   :  0.07517    Mean   :  0.1512
##  3rd Qu.:  0.560    3rd Qu.:  1.62000    3rd Qu.:  0.4900
##  Max.   :  3.970    Max.   :311.00000    Max.   :231.0000
##  accel_forearm_x    accel_forearm_y     accel_forearm_z    magnet_forearm_x
##  Min.   :-498.00    Min.   :-632.0      Min.   :-446.00    Min.   :-1280.0
##  1st Qu.:-178.00    1st Qu.:  57.0      1st Qu.:-182.00    1st Qu.: -616.0
##  Median : -57.00    Median : 201.0      Median : -39.00    Median : -378.0
##  Mean   : -61.65    Mean   : 163.7      Mean   : -55.29    Mean   : -312.6
##  3rd Qu.:  76.00    3rd Qu.: 312.0      3rd Qu.:  26.00    3rd Qu.:  -73.0
##  Max.   : 477.00    Max.   : 923.0      Max.   : 291.00    Max.   :  672.0
##  magnet_forearm_y magnet_forearm_z    classe
##  Min.   :-896.0   Min.   :-973.0    Length:19622
##  1st Qu.:   2.0   1st Qu.: 191.0    Class :character
##  Median : 591.0   Median : 511.0    Mode  :character
##  Mean   : 380.1   Mean   : 393.6
##  3rd Qu.: 737.0   3rd Qu.: 653.0
##  Max.   :1480.0   Max.   :1090.0
```

Number of columns reduced from 160 to 93.

Preprocess the numeric columns. Preserve the outcome column.

```
c = data$classe # save

v = which (lapply (data, class) %in% "numeric") # numerica columns only
v
```

```
##  [1]   8   9  10  21  22  23  30  31  32  34  35  36  49  50  51  62  63  64  70  71  72  73  84
## [24] 85 86 91 92
```

```
ppo = preProcess (data [,v], method=c ('knnImpute', 'center', 'scale'))
data <- predict (ppo, data [,v])

data$classe = as.factor (c) # restore

dim (data)
```

```
## [1] 19622    28
```

```
summary (data)
```

```
##    roll_belt         pitch_belt          yaw_belt          gyros_belt_x
##  Min.   :-1.4870   Min.   :-2.51016   Min.   :-1.77317   Min.   :-4.9892
##  1st Qu.:-1.0089   1st Qu.: 0.06508   1st Qu.:-0.80987   1st Qu.:-0.1177
##  Median : 0.7744   Median : 0.22257   Median :-0.01886   Median : 0.1717
##  Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000
##  3rd Qu.: 0.9337   3rd Qu.: 0.65297   3rd Qu.: 0.25322   3rd Qu.: 0.5575
##  Max.   : 1.5553   Max.   : 2.68418   Max.   : 1.99808   Max.   :10.7346
##   gyros_belt_y      gyros_belt_z        roll_arm          pitch_arm
##  Min.   :-8.6864   Min.   :-5.5091   Min.   :-2.7195   Min.   :-2.7439
##  1st Qu.:-0.5060   1st Qu.:-0.2878   1st Qu.:-0.6819   1st Qu.:-0.6938
##  Median :-0.2504   Median : 0.1265   Median :-0.2451   Median : 0.1503
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.9000   3rd Qu.: 0.4581   3rd Qu.: 0.8175   3rd Qu.: 0.5153
##  Max.   : 7.6744   Max.   : 7.2540   Max.   : 2.2293   Max.   : 3.0347
##     yaw_arm          gyros_arm_x         gyros_arm_y
##  Min.   :-2.51363   Min.   :-3.21669   Min.   :-3.73849
##  1st Qu.:-0.59528   1st Qu.:-0.68859   1st Qu.:-0.63768
##  Median : 0.00867   Median : 0.01867   Median : 0.02006
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.00000
##  3rd Qu.: 0.65151   3rd Qu.: 0.76607   3rd Qu.: 0.46639
##  Max.   : 2.53097   Max.   : 2.42137   Max.   : 3.63767
##   gyros_arm_z       roll_dumbbell      pitch_dumbbell      yaw_dumbbell
##  Min.   :-4.69925   Min.   :-2.5390   Min.   :-3.7524   Min.   :-1.84865
##  1st Qu.:-0.61370   1st Qu.:-0.6054   1st Qu.:-0.8139   1st Qu.:-0.96124
##  Median :-0.07137   Median : 0.3479   Median :-0.2751   Median :-0.06057
```

```
##   Mean    : 0.00000    Mean    : 0.0000    Mean    : 0.0000    Mean    : 0.00000
##   3rd Qu.: 0.81444    3rd Qu.: 0.6259    3rd Qu.: 0.7644    3rd Qu.: 0.94488
##   Max.    : 4.97230    Max.    : 1.8547    Max.    : 4.3300    Max.    : 1.85753
##   gyros_dumbbell_x      gyros_dumbbell_y      gyros_dumbbell_z
##   Min.    :-135.33567    Min.    :-3.51818    Min.    : -0.98444
##   1st Qu.:  -0.12667    1st Qu.:-0.30502    1st Qu.: -0.07916
##   Median :  -0.02061    Median :-0.02632    Median : -0.00044
##   Mean    :  0.00000    Mean    : 0.00000    Mean    : 0.00000
##   3rd Qu.:   0.12523    3rd Qu.: 0.26876    3rd Qu.: 0.06953
##   Max.    :   1.36483    Max.    :85.17175    Max.    :138.69025
##   magnet_dumbbell_z   roll_forearm        pitch_forearm        yaw_forearm
##   Min.    :-2.2010    Min.    :-1.9792    Min.    :-2.95628    Min.    :-1.9299
##   1st Qu.:-0.6506    1st Qu.:-0.3199    1st Qu.:-0.38038    1st Qu.:-0.8507
##   Median :-0.2362    Median :-0.1122    Median :-0.05209    Median :-0.1861
##   Mean    : 0.0000    Mean    : 0.0000    Mean    : 0.00000    Mean    : 0.0000
##   3rd Qu.: 0.3497    3rd Qu.: 0.9828    3rd Qu.: 0.62866    3rd Qu.: 0.8796
##   Max.    : 2.9004    Max.    : 1.3530    Max.    : 2.81018    Max.    : 1.5578
##   gyros_forearm_x      gyros_forearm_y      gyros_forearm_z
##   Min.    :-34.1618    Min.    : -2.28823    Min.    : -4.69725
##   1st Qu.: -0.5827    1st Qu.: -0.49510    1st Qu.: -0.18880
##   Median : -0.1664    Median : -0.01457    Median : -0.04061
##   Mean    :  0.0000    Mean    :  0.00000    Mean    :  0.00000
##   3rd Qu.:  0.6199    3rd Qu.:  0.49821    3rd Qu.:  0.19308
##   Max.    :  5.8772    Max.    :100.27489    Max.    :131.57654
##   magnet_forearm_y   magnet_forearm_z   classe
##   Min.    :-2.5053    Min.    :-3.7009    A:5580
##   1st Qu.:-0.7423    1st Qu.:-0.5487    B:3797
##   Median : 0.4140    Median : 0.3179    C:3422
##   Mean    : 0.0000    Mean    : 0.0000    D:3216
##   3rd Qu.: 0.7006    3rd Qu.: 0.7024    E:3607
##   Max.    : 2.1593    Max.    : 1.8859
```

```
str (data)
```

```
## 'data.frame':    19622 obs. of  28 variables:
##  $ roll_belt         : num  -1 -1 -1 -1 -1 ...
##  $ pitch_belt        : num  0.347 0.347 0.347 0.347 0.347 ...
##  $ yaw_belt          : num  -0.874 -0.874 -0.874 -0.874 -0.874 ...
##  $ gyros_belt_x      : num  0.027 0.123 0.027 0.123 0.123 ...
##  $ gyros_belt_y      : num  -0.506 -0.506 -0.506 -0.506 -0.25 ...
##  $ gyros_belt_z      : num  0.458 0.458 0.458 0.417 0.458 ...
##  $ roll_arm          : num  -2 -2 -2 -2 -2 ...
##  $ pitch_arm         : num  0.884 0.884 0.884 0.871 0.871 ...
##  $ yaw_arm           : num  -2.25 -2.25 -2.25 -2.25 -2.25 ...
##  $ gyros_arm_x       : num  -0.0215 -0.0114 -0.0114 -0.0114 -0.0215 ...
##  $ gyros_arm_y       : num  0.302 0.278 0.278 0.267 0.267 ...
##  $ gyros_arm_z       : num  -0.523 -0.523 -0.523 -0.451 -0.487 ...
##  $ roll_dumbbell     : num  -0.154 -0.153 -0.157 -0.149 -0.15 ...
##  $ pitch_dumbbell    : num  -1.61 -1.62 -1.61 -1.61 -1.61 ...
##  $ yaw_dumbbell      : num  -1.05 -1.05 -1.05 -1.05 -1.05 ...
##  $ gyros_dumbbell_x  : num  -0.107 -0.107 -0.107 -0.107 -0.107 ...
##  $ gyros_dumbbell_y  : num  -0.108 -0.108 -0.108 -0.108 -0.108 ...
##  $ gyros_dumbbell_z  : num  0.0564 0.0564 0.0564 0.0477 0.0564 ...
##  $ magnet_dumbbell_z : num  -0.793 -0.786 -0.779 -0.758 -0.815 ...
##  $ roll_forearm      : num  -0.0502 -0.0512 -0.0512 -0.053 -0.0539 ...
##  $ pitch_forearm     : num  -2.65 -2.65 -2.65 -2.65 -2.65 ...
##  $ yaw_forearm       : num  -1.67 -1.67 -1.66 -1.66 -1.66 ...
##  $ gyros_forearm_x   : num  -0.197 -0.213 -0.197 -0.213 -0.213 ...
##  $ gyros_forearm_y   : num  -0.0242 -0.0242 -0.0307 -0.0307 -0.0242 ...
##  $ gyros_forearm_z   : num  -0.0976 -0.0976 -0.0862 -0.0862 -0.0976 ...
##  $ magnet_forearm_y  : num  0.538 0.551 0.546 0.546 0.54 ...
##  $ magnet_forearm_z  : num  0.223 0.215 0.204 0.204 0.215 ...
##  $ classe            : Factor w/ 5 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...
```

Remove columns with near zero variance

```
nzv <- nearZeroVar (data, saveMetrics=TRUE)
nzv
```

```
##                     freqRatio percentUnique zeroVar    nzv
## roll_belt            1.101904     6.7781062    FALSE  FALSE
## pitch_belt           1.036082     9.3772296    FALSE  FALSE
## yaw_belt             1.058480     9.9734991    FALSE  FALSE
## gyros_belt_x         1.058651     0.7134849    FALSE  FALSE
## gyros_belt_y         1.144000     0.3516461    FALSE  FALSE
## gyros_belt_z         1.066214     0.8612782    FALSE  FALSE
## roll_arm            52.338462    13.5256345    FALSE  FALSE
## pitch_arm           87.256410    15.7323412    FALSE  FALSE
## yaw_arm             33.029126    14.6570176    FALSE  FALSE
## gyros_arm_x          1.015504     3.2769341    FALSE  FALSE
## gyros_arm_y          1.454369     1.9162165    FALSE  FALSE
## gyros_arm_z          1.110687     1.2638875    FALSE  FALSE
## roll_dumbbell        1.022388    84.2065029    FALSE  FALSE
## pitch_dumbbell       2.277372    81.6685353    FALSE  FALSE
## yaw_dumbbell         1.132231    83.4828254    FALSE  FALSE
## gyros_dumbbell_x     1.003268     1.2282132    FALSE  FALSE
## gyros_dumbbell_y     1.264957     1.4167771    FALSE  FALSE
## gyros_dumbbell_z     1.060100     1.0498420    FALSE  FALSE
## magnet_dumbbell_z    1.020833     3.4451126    FALSE  FALSE
## roll_forearm        11.589286    11.0895933    FALSE  FALSE
## pitch_forearm       65.983051    14.8557741    FALSE  FALSE
## yaw_forearm         15.322835    10.1467740    FALSE  FALSE
## gyros_forearm_x      1.059273     1.5187035    FALSE  FALSE
## gyros_forearm_y      1.036554     3.7763735    FALSE  FALSE
## gyros_forearm_z      1.122917     1.5645704    FALSE  FALSE
## magnet_forearm_y     1.246914     9.5403119    FALSE  FALSE
## magnet_forearm_z     1.000000     8.5771073    FALSE  FALSE
## classe               1.469581     0.0254816    FALSE  FALSE
```

```
data = data [,nzv$nzv==FALSE]

dim (data)
```

```
## [1] 19622    28
```

```
summary (data)
```

```
##    roll_belt          pitch_belt          yaw_belt          gyros_belt_x
## Min.   :-1.4870   Min.   :-2.51016   Min.   :-1.77317   Min.   :-4.9892
## 1st Qu.:-1.0089   1st Qu.: 0.06508   1st Qu.:-0.80987   1st Qu.:-0.1177
## Median : 0.7744   Median : 0.22257   Median :-0.01886   Median : 0.1717
## Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000
## 3rd Qu.: 0.9337   3rd Qu.: 0.65297   3rd Qu.: 0.25322   3rd Qu.: 0.5575
## Max.   : 1.5553   Max.   : 2.68418   Max.   : 1.99808   Max.   :10.7346
```

```
##    gyros_belt_y       gyros_belt_z        roll_arm          pitch_arm
##  Min.   :-8.6864    Min.   :-5.5091    Min.   :-2.7195    Min.   :-2.7439
##  1st Qu.:-0.5060    1st Qu.:-0.2878    1st Qu.:-0.6819    1st Qu.:-0.6938
##  Median :-0.2504    Median : 0.1265    Median :-0.2451    Median : 0.1503
##  Mean   : 0.0000    Mean   : 0.0000    Mean   : 0.0000    Mean   : 0.0000
##  3rd Qu.: 0.9000    3rd Qu.: 0.4581    3rd Qu.: 0.8175    3rd Qu.: 0.5153
##  Max.   : 7.6744    Max.   : 7.2540    Max.   : 2.2293    Max.   : 3.0347
##     yaw_arm           gyros_arm_x         gyros_arm_y
##  Min.   :-2.51363   Min.   :-3.21669   Min.   :-3.73849
##  1st Qu.:-0.59528   1st Qu.:-0.68859   1st Qu.:-0.63768
##  Median : 0.00867   Median : 0.01867   Median : 0.02006
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.00000
##  3rd Qu.: 0.65151   3rd Qu.: 0.76607   3rd Qu.: 0.46639
##  Max.   : 2.53097   Max.   : 2.42137   Max.   : 3.63767
##   gyros_arm_z       roll_dumbbell      pitch_dumbbell     yaw_dumbbell
##  Min.   :-4.69925   Min.   :-2.5390    Min.   :-3.7524    Min.   :-1.84865
##  1st Qu.:-0.61370   1st Qu.:-0.6054    1st Qu.:-0.8139    1st Qu.:-0.96124
##  Median :-0.07137   Median : 0.3479    Median :-0.2751    Median :-0.06057
##  Mean   : 0.00000   Mean   : 0.0000    Mean   : 0.0000    Mean   : 0.00000
##  3rd Qu.: 0.81444   3rd Qu.: 0.6259    3rd Qu.: 0.7644    3rd Qu.: 0.94488
##  Max.   : 4.97230   Max.   : 1.8547    Max.   : 4.3300    Max.   : 1.85753
##  gyros_dumbbell_x    gyros_dumbbell_y    gyros_dumbbell_z
##  Min.   :-135.33567   Min.   :-3.51818   Min.   : -0.98444
##  1st Qu.:  -0.12667   1st Qu.:-0.30502   1st Qu.: -0.07916
##  Median :  -0.02061   Median :-0.02632   Median : -0.00044
##  Mean   :   0.00000   Mean   : 0.00000   Mean   :  0.00000
##  3rd Qu.:   0.12523   3rd Qu.: 0.26876   3rd Qu.:  0.06953
##  Max.   :   1.36483   Max.   :85.17175   Max.   :138.69025
##  magnet_dumbbell_z  roll_forearm       pitch_forearm      yaw_forearm
##  Min.   :-2.2010    Min.   :-1.9792    Min.   :-2.95628   Min.   :-1.9299
##  1st Qu.:-0.6506    1st Qu.:-0.3199    1st Qu.:-0.38038   1st Qu.:-0.8507
##  Median :-0.2362    Median :-0.1122    Median :-0.05209   Median :-0.1861
##  Mean   : 0.0000    Mean   : 0.0000    Mean   : 0.00000   Mean   : 0.0000
##  3rd Qu.: 0.3497    3rd Qu.: 0.9828    3rd Qu.: 0.62866   3rd Qu.: 0.8796
##  Max.   : 2.9004    Max.   : 1.3530    Max.   : 2.81018   Max.   : 1.5578
##  gyros_forearm_x     gyros_forearm_y     gyros_forearm_z
##  Min.   :-34.1618   Min.   : -2.28823   Min.   : -4.69725
##  1st Qu.: -0.5827   1st Qu.: -0.49510   1st Qu.: -0.18880
##  Median : -0.1664   Median : -0.01457   Median : -0.04061
##  Mean   :  0.0000   Mean   :  0.00000   Mean   :  0.00000
##  3rd Qu.:  0.6199   3rd Qu.:  0.49821   3rd Qu.:  0.19308
##  Max.   :  5.8772   Max.   :100.27489   Max.   :131.57654
##  magnet_forearm_y  magnet_forearm_z  classe
##  Min.   :-2.5053    Min.   :-3.7009    A:5580
##  1st Qu.:-0.7423    1st Qu.:-0.5487    B:3797
##  Median : 0.4140    Median : 0.3179    C:3422
##  Mean   : 0.0000    Mean   : 0.0000    D:3216
```

```
##   3rd Qu.: 0.7006    3rd Qu.: 0.7024    E:3607
##   Max.    : 2.1593   Max.    : 1.8859
```

```
str (data)
```

```
## 'data.frame':    19622 obs. of  28 variables:
##  $ roll_belt        : num   -1 -1 -1 -1 -1 ...
##  $ pitch_belt       : num   0.347 0.347 0.347 0.347 0.347 ...
##  $ yaw_belt         : num   -0.874 -0.874 -0.874 -0.874 -0.874 ...
##  $ gyros_belt_x     : num   0.027 0.123 0.027 0.123 0.123 ...
##  $ gyros_belt_y     : num   -0.506 -0.506 -0.506 -0.506 -0.25 ...
##  $ gyros_belt_z     : num   0.458 0.458 0.458 0.417 0.458 ...
##  $ roll_arm         : num   -2 -2 -2 -2 -2 ...
##  $ pitch_arm        : num   0.884 0.884 0.884 0.871 0.871 ...
##  $ yaw_arm          : num   -2.25 -2.25 -2.25 -2.25 -2.25 ...
##  $ gyros_arm_x      : num   -0.0215 -0.0114 -0.0114 -0.0114 -0.0215 ...
##  $ gyros_arm_y      : num   0.302 0.278 0.278 0.267 0.267 ...
##  $ gyros_arm_z      : num   -0.523 -0.523 -0.523 -0.451 -0.487 ...
##  $ roll_dumbbell    : num   -0.154 -0.153 -0.157 -0.149 -0.15 ...
##  $ pitch_dumbbell   : num   -1.61 -1.62 -1.61 -1.61 -1.61 ...
##  $ yaw_dumbbell     : num   -1.05 -1.05 -1.05 -1.05 -1.05 ...
##  $ gyros_dumbbell_x : num   -0.107 -0.107 -0.107 -0.107 -0.107 ...
##  $ gyros_dumbbell_y : num   -0.108 -0.108 -0.108 -0.108 -0.108 ...
##  $ gyros_dumbbell_z : num   0.0564 0.0564 0.0564 0.0477 0.0564 ...
##  $ magnet_dumbbell_z: num   -0.793 -0.786 -0.779 -0.758 -0.815 ...
##  $ roll_forearm     : num   -0.0502 -0.0512 -0.0512 -0.053 -0.0539 ...
##  $ pitch_forearm    : num   -2.65 -2.65 -2.65 -2.65 -2.65 ...
##  $ yaw_forearm      : num   -1.67 -1.67 -1.66 -1.66 -1.66 ...
##  $ gyros_forearm_x  : num   -0.197 -0.213 -0.197 -0.213 -0.213 ...
##  $ gyros_forearm_y  : num   -0.0242 -0.0242 -0.0307 -0.0307 -0.0242 ...
##  $ gyros_forearm_z  : num   -0.0976 -0.0976 -0.0862 -0.0862 -0.0976 ...
##  $ magnet_forearm_y : num   0.538 0.551 0.546 0.546 0.54 ...
##  $ magnet_forearm_z : num   0.223 0.215 0.204 0.204 0.215 ...
##  $ classe           : Factor w/ 5 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...
```

Number of columns reduced from 93 to 28. Pretty good compared to the original 160.

Partition the data into training and test sets.

Note that all preprocessing/cleaning of the data is performed **before** the partitioning.

```
set.seed (12345)


dp = createDataPartition (y = data$classe, p = 0.6, list=FALSE)
myTraining = data [dp,] # training set has random 60%
myTesting = data [-dp, ] # testing set has the remaining 40%


dim (myTraining)
```

```
## [1] 11776    28
```

```
dim (myTesting)
```
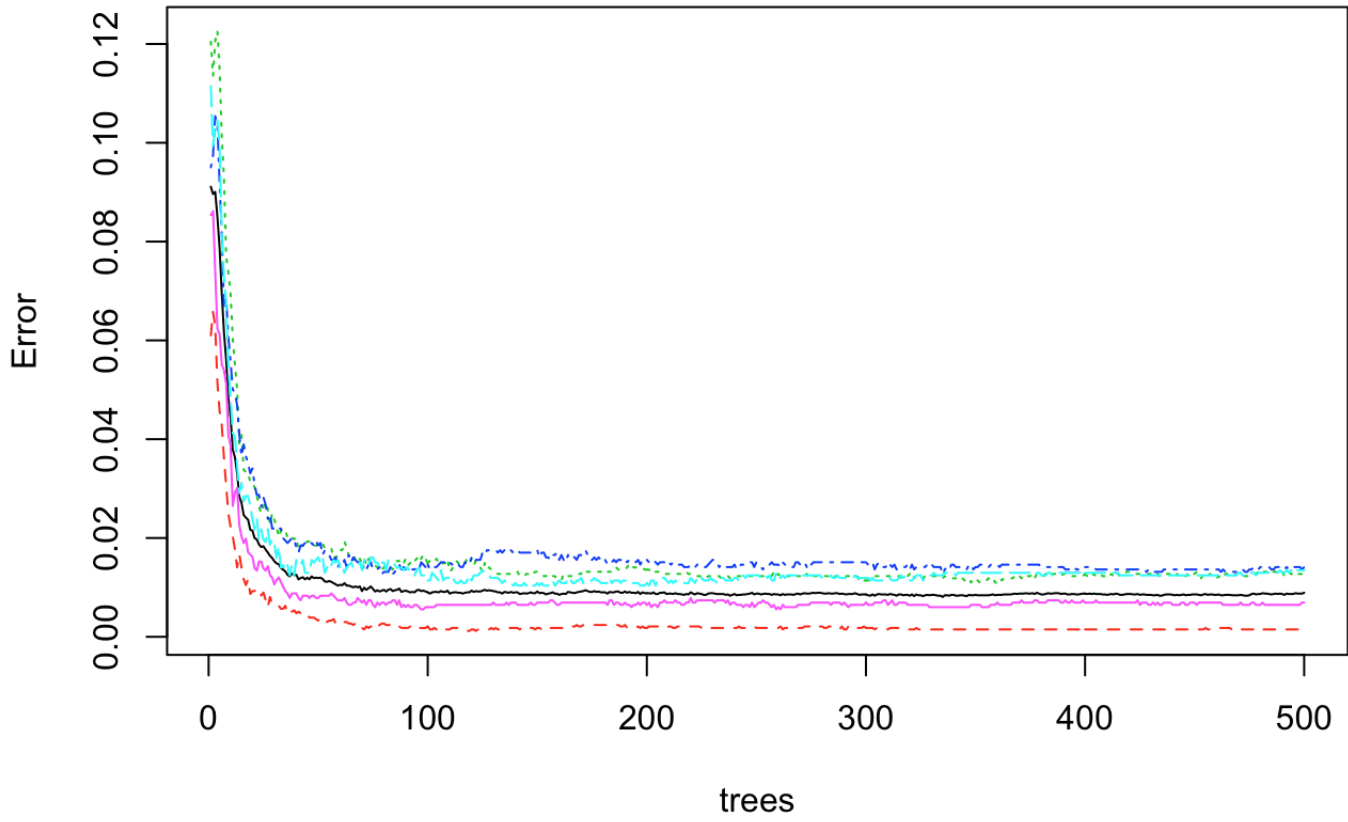
```
## [1] 7846    28
```

Train the model - fit

```
fit = randomForest (classe ~ ., data=myTraining)
fit
```

```
##
## Call:
##   randomForest(formula = classe ~ ., data = myTraining)
##                  Type of random forest: classification
##                         Number of trees: 500
## No. of variables tried at each split: 5
##
##          OOB estimate of  error rate: 0.89%
## Confusion matrix:
##       A    B    C    D    E class.error
## A 3343    4    0    0    1 0.001493429
## B   16 2249   11    1    2 0.013163668
## C    0   12 2025   16    1 0.014118793
## D    0    0   21 1904    5 0.013471503
## E    0    2    4    9 2150 0.006928406
```

```
plot (fit)
```

fit

```
imp = varImp (fit)
imp
```

```
##                        Overall
## roll_belt            1184.98424
## pitch_belt            664.87727
## yaw_belt              876.04683
## gyros_belt_x          126.47359
## gyros_belt_y          157.22140
## gyros_belt_z          348.56799
## roll_arm              328.88533
## pitch_arm             170.61394
## yaw_arm               320.83210
## gyros_arm_x           129.82550
## gyros_arm_y           145.29218
## gyros_arm_z            74.90312
## roll_dumbbell         492.98727
## pitch_dumbbell        264.46993
## yaw_dumbbell          330.94705
## gyros_dumbbell_x      142.90256
## gyros_dumbbell_y      322.83452
## gyros_dumbbell_z       93.18073
## magnet_dumbbell_z     700.79537
## roll_forearm          546.39488
## pitch_forearm         771.49432
## yaw_forearm           216.67365
## gyros_forearm_x        96.00044
## gyros_forearm_y       135.67157
## gyros_forearm_z        95.30065
## magnet_forearm_y      243.45499
## magnet_forearm_z      328.29754
```

Expected error rate of 0.8% sounds pretty good to me :-)

Train the model in a slightly different way - fit2

Note this method takes a **loooong** time.

```
fit2 = train (classe ~.,
              method="rf",
              data=myTraining,
              trControl=trainControl (method='cv'), number=5, allowParallel=TRUE )
fit2
```

```
## Random Forest
##
## 11776 samples
##     27 predictor
##      5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
##
## Summary of sample sizes: 10598, 10598, 10598, 10599, 10599, 10599, ...
##
## Resampling results across tuning parameters:
##
##   mtry  Accuracy   Kappa      Accuracy SD  Kappa SD
##    2    0.9898956  0.9872175  0.002921488  0.003696799
##   14    0.9895553  0.9867884  0.002155995  0.002728546
##   27    0.9862440  0.9826019  0.001818795  0.002302239
##
## Accuracy was used to select the optimal model using  the largest value.
## The final value used for the model was mtry = 2.
```

```
imp = varImp (fit2)
imp
```

```
## rf variable importance
##
##    only 20 most important variables shown (out of 27)
##
##                       Overall
## roll_belt             100.00
## yaw_belt               80.86
## magnet_dumbbell_z      64.30
## pitch_forearm          60.16
## pitch_belt             60.09
## roll_forearm           45.86
## roll_dumbbell          44.07
## roll_arm               32.96
## gyros_belt_z           30.69
## yaw_dumbbell           30.58
## gyros_dumbbell_y       29.29
## yaw_arm                28.61
## pitch_dumbbell         27.46
## magnet_forearm_z       27.13
## magnet_forearm_y       25.06
## yaw_forearm            19.57
## pitch_arm              14.54
## gyros_arm_x            13.14
## gyros_dumbbell_x       11.93
## gyros_arm_y            11.80
```

Note that this model was cross validated 10 fold.

Check against the test set - both fit and fit2

```
predictions = predict (fit, myTesting) #fit
confusionMatrix (predictions, myTesting$classe)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 2225    9    0    0    0
##          B    6 1501    6    1    1
##          C    0    8 1352   10    2
##          D    0    0   10 1271    3
##          E    1    0    0    4 1436
##
## Overall Statistics
##
##                Accuracy : 0.9922
##                  95% CI : (0.99, 0.994)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9902
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9969   0.9888   0.9883   0.9883   0.9958
## Specificity            0.9984   0.9978   0.9969   0.9980   0.9992
## Pos Pred Value         0.9960   0.9908   0.9854   0.9899   0.9965
## Neg Pred Value         0.9988   0.9973   0.9975   0.9977   0.9991
## Prevalence             0.2845   0.1935   0.1744   0.1639   0.1838
## Detection Rate         0.2836   0.1913   0.1723   0.1620   0.1830
## Detection Prevalence   0.2847   0.1931   0.1749   0.1637   0.1837
## Balanced Accuracy      0.9976   0.9933   0.9926   0.9932   0.9975
```

```
predictions = predict (fit2, myTesting) #fit2
confusionMatrix (predictions, myTesting$classe)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 2226   11    0    0    0
##          B    6 1499    4    0    0
##          C    0    8 1359   14    2
##          D    0    0    5 1268    3
##          E    0    0    0    4 1437
##
## Overall Statistics
##
##                Accuracy : 0.9927
##                  95% CI : (0.9906, 0.9945)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9908
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                     Class: A Class: B Class: C Class: D Class: E
## Sensitivity           0.9973   0.9875   0.9934   0.9860   0.9965
## Specificity           0.9980   0.9984   0.9963   0.9988   0.9994
## Pos Pred Value        0.9951   0.9934   0.9826   0.9937   0.9972
## Neg Pred Value        0.9989   0.9970   0.9986   0.9973   0.9992
## Prevalence            0.2845   0.1935   0.1744   0.1639   0.1838
## Detection Rate        0.2837   0.1911   0.1732   0.1616   0.1832
## Detection Prevalence  0.2851   0.1923   0.1763   0.1626   0.1837
## Balanced Accuracy     0.9977   0.9930   0.9949   0.9924   0.9980
```

Conclusion

The accuracy of the second model, which takes much longer to calculate, is **not** more accurate than the first model. In either case, I am very happy with a 99% accuracy.