# PracticalMachineLearningCourseProject

Katie Santo

6/21/2020

## Background

This is the course project for Practical Machine Learning within the Data Science: Statistics and Machine Learning Specialization. The project uses data from accelerometers on the belt, forearm, arm, and dumbbell of six participants. The participants performed barbell lifts five different ways (exactly according to specs, throwing elbows to the front, lifting the dumbbell halfway, lowering dumbbell half way, or throwing the hips to the front). The goal of this project is to predict the manner in which the participants performed the barbell lift.

## Data preparation

First, I will load in the data and do some basic exploratory analysis. In the training dataset there are 160 variables and 19,622 observations. From the summary, it looks like there are some variables with majority of the values missing. There are also some variables that do not seem pertinent to predicting the outcome (X, user_name, raw_timestamp_part_1, raw_timestamp_part_2, cvd_timestamp, new_window, num_window). My final training dataset will have 19,622 observations and 53 variables. The testing dataset has 160 variables and 20 observations.

```
# Loading in the training data and getting an overview of the data
train <- read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv", header=TRUE, n
dim(train)
```

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

```
summary(train)
```

```
##        X            user_name    raw_timestamp_part_1 raw_timestamp_part_2
##  Min.   :    1   adelmo  :3892   Min.   :1.322e+09    Min.   :   294
##  1st Qu.: 4906   carlitos:3112   1st Qu.:1.323e+09    1st Qu.:252912
##  Median : 9812   charles :3536   Median :1.323e+09    Median :496380
##  Mean   : 9812   eurico  :3070   Mean   :1.323e+09    Mean   :500656
##  3rd Qu.:14717   jeremy  :3402   3rd Qu.:1.323e+09    3rd Qu.:751891
##  Max.   :19622   pedro   :2610   Max.   :1.323e+09    Max.   :998801
##
##           cvtd_timestamp  new_window    num_window      roll_belt
##  28/11/2011 14:14: 1498   no :19216   Min.   :  1.0   Min.   :-28.90
##  05/12/2011 11:24: 1497   yes:  406   1st Qu.:222.0   1st Qu.:  1.10
##  30/11/2011 17:11: 1440               Median :424.0   Median :113.00
##  05/12/2011 11:25: 1425               Mean   :430.6   Mean   : 64.41
##  02/12/2011 14:57: 1380               3rd Qu.:644.0   3rd Qu.:123.00
##  02/12/2011 13:34: 1375               Max.   :864.0   Max.   :162.00
##  (Other)         :11007
##    pitch_belt          yaw_belt       total_accel_belt kurtosis_roll_belt
```

```
##  Min.   :-55.8000   Min.   :-180.00   Min.   : 0.00    #DIV/0!  :   10
##  1st Qu.:  1.7600   1st Qu.: -88.30   1st Qu.: 3.00    -1.908453:    2
##  Median :  5.2800   Median : -13.00   Median :17.00    -0.016850:    1
##  Mean   :  0.3053   Mean   : -11.21   Mean   :11.31    -0.021024:    1
##  3rd Qu.: 14.9000   3rd Qu.:  12.90   3rd Qu.:18.00    -0.025513:    1
##  Max.   : 60.3000   Max.   : 179.00   Max.   :29.00    (Other)  :  391
##                                                        NA's     :19216
##  kurtosis_picth_belt kurtosis_yaw_belt skewness_roll_belt skewness_roll_belt.1
##  #DIV/0!  :   32     #DIV/0!:  406     #DIV/0!  :    9    #DIV/0!  :   32
##  47.000000:    4     NA's   :19216     0.000000 :    4    0.000000 :    4
##  -0.150950:    3                       0.422463 :    2    -2.156553:    3
##  -0.684748:    3                       -0.003095:    1    -3.072669:    3
##  -1.750749:    3                       -0.010002:    1    -6.324555:    3
##  (Other)  :  361                       (Other)  :  389    (Other)  :  361
##  NA's     :19216                       NA's     :19216    NA's     :19216
##  skewness_yaw_belt max_roll_belt     max_picth_belt   max_yaw_belt
##  #DIV/0!:  406     Min.   :-94.300   Min.   : 3.00   -1.1   :   30
##  NA's   :19216     1st Qu.:-88.000   1st Qu.: 5.00   -1.4   :   29
##                    Median : -5.100   Median :18.00   -1.2   :   26
##                    Mean   : -6.667   Mean   :12.92   -0.9   :   24
##                    3rd Qu.: 18.500   3rd Qu.:19.00   -1.3   :   22
##                    Max.   :180.000   Max.   :30.00   (Other):  275
##                    NA's   :19216     NA's   :19216   NA's   :19216
##  min_roll_belt     min_pitch_belt   min_yaw_belt   amplitude_roll_belt
##  Min.   :-180.00   Min.   : 0.00   -1.1   :   30   Min.   :  0.000
##  1st Qu.: -88.40   1st Qu.: 3.00   -1.4   :   29   1st Qu.:  0.300
##  Median :  -7.85   Median :16.00   -1.2   :   26   Median :  1.000
##  Mean   : -10.44   Mean   :10.76   -0.9   :   24   Mean   :  3.769
##  3rd Qu.:   9.05   3rd Qu.:17.00   -1.3   :   22   3rd Qu.:  2.083
##  Max.   : 173.00   Max.   :23.00   (Other):  275   Max.   :360.000
##  NA's   :19216     NA's   :19216   NA's   :19216   NA's   :19216
##  amplitude_pitch_belt amplitude_yaw_belt var_total_accel_belt avg_roll_belt
##  Min.   : 0.000       #DIV/0!:   10      Min.   : 0.000       Min.   :-27.40
##  1st Qu.: 1.000       0.00   :   12      1st Qu.: 0.100       1st Qu.:  1.10
##  Median : 1.000       0.0000 :  384      Median : 0.200       Median :116.35
##  Mean   : 2.167       NA's   :19216      Mean   : 0.926       Mean   : 68.06
##  3rd Qu.: 2.000                          3rd Qu.: 0.300       3rd Qu.:123.38
##  Max.   :12.000                          Max.   :16.500       Max.   :157.40
##  NA's   :19216                           NA's   :19216        NA's   :19216
##  stddev_roll_belt var_roll_belt     avg_pitch_belt    stddev_pitch_belt
##  Min.   : 0.000   Min.   :  0.000   Min.   :-51.400   Min.   :0.000
##  1st Qu.: 0.200   1st Qu.:  0.000   1st Qu.:  2.025   1st Qu.:0.200
##  Median : 0.400   Median :  0.100   Median :  5.200   Median :0.400
##  Mean   : 1.337   Mean   :  7.699   Mean   :  0.520   Mean   :0.603
##  3rd Qu.: 0.700   3rd Qu.:  0.500   3rd Qu.: 15.775   3rd Qu.:0.700
##  Max.   :14.200   Max.   :200.700   Max.   : 59.700   Max.   :4.000
##  NA's   :19216    NA's   :19216     NA's   :19216     NA's   :19216
##  var_pitch_belt    avg_yaw_belt      stddev_yaw_belt    var_yaw_belt
##  Min.   : 0.000   Min.   :-138.300   Min.   :  0.000   Min.   :    0.000
##  1st Qu.: 0.000   1st Qu.: -88.175   1st Qu.:  0.100   1st Qu.:    0.010
##  Median : 0.100   Median :  -6.550   Median :  0.300   Median :    0.090
##  Mean   : 0.766   Mean   :  -8.831   Mean   :  1.341   Mean   :  107.487
##  3rd Qu.: 0.500   3rd Qu.:  14.125   3rd Qu.:  0.700   3rd Qu.:    0.475
##  Max.   :16.200   Max.   : 173.500   Max.   :176.600   Max.   :31183.240
```

```
##  NA's   :19216    NA's   :19216      NA's   :19216     NA's   :19216
##   gyros_belt_x        gyros_belt_y       gyros_belt_z      accel_belt_x
##  Min.   :-1.040000   Min.   :-0.64000   Min.   :-1.4600   Min.   :-120.000
##  1st Qu.:-0.030000   1st Qu.: 0.00000   1st Qu.:-0.2000   1st Qu.: -21.000
##  Median : 0.030000   Median : 0.02000   Median :-0.1000   Median : -15.000
##  Mean   :-0.005592   Mean   : 0.03959   Mean   :-0.1305   Mean   :  -5.595
##  3rd Qu.: 0.110000   3rd Qu.: 0.11000   3rd Qu.:-0.0200   3rd Qu.:  -5.000
##  Max.   : 2.220000   Max.   : 0.64000   Max.   : 1.6200   Max.   :  85.000
##
##   accel_belt_y     accel_belt_z     magnet_belt_x   magnet_belt_y
##  Min.   :-69.00   Min.   :-275.00   Min.   :-52.0   Min.   :354.0
##  1st Qu.:  3.00   1st Qu.:-162.00   1st Qu.:  9.0   1st Qu.:581.0
##  Median : 35.00   Median :-152.00   Median : 35.0   Median :601.0
##  Mean   : 30.15   Mean   : -72.59   Mean   : 55.6   Mean   :593.7
##  3rd Qu.: 61.00   3rd Qu.:  27.00   3rd Qu.: 59.0   3rd Qu.:610.0
##  Max.   :164.00   Max.   : 105.00   Max.   :485.0   Max.   :673.0
##
##  magnet_belt_z       roll_arm         pitch_arm          yaw_arm
##  Min.   :-623.0   Min.   :-180.00   Min.   :-88.800   Min.   :-180.0000
##  1st Qu.:-375.0   1st Qu.: -31.77   1st Qu.:-25.900   1st Qu.: -43.1000
##  Median :-320.0   Median :   0.00   Median :  0.000   Median :   0.0000
##  Mean   :-345.5   Mean   :  17.83   Mean   : -4.612   Mean   :  -0.6188
##  3rd Qu.:-306.0   3rd Qu.:  77.30   3rd Qu.: 11.200   3rd Qu.:  45.8750
##  Max.   : 293.0   Max.   : 180.00   Max.   : 88.500   Max.   : 180.0000
##
##  total_accel_arm var_accel_arm     avg_roll_arm     stddev_roll_arm
##  Min.   : 1.00   Min.   :  0.00   Min.   :-166.67   Min.   :  0.000
##  1st Qu.:17.00   1st Qu.:  9.03   1st Qu.: -38.37   1st Qu.:  1.376
##  Median :27.00   Median : 40.61   Median :   0.00   Median :  5.702
##  Mean   :25.51   Mean   : 53.23   Mean   :  12.68   Mean   : 11.201
##  3rd Qu.:33.00   3rd Qu.: 75.62   3rd Qu.:  76.33   3rd Qu.: 14.921
##  Max.   :66.00   Max.   :331.70   Max.   : 163.33   Max.   :161.964
##                  NA's   :19216    NA's   :19216     NA's   :19216
##   var_roll_arm       avg_pitch_arm     stddev_pitch_arm var_pitch_arm
##  Min.   :    0.000   Min.   :-81.773   Min.   : 0.000   Min.   :   0.000
##  1st Qu.:    1.898   1st Qu.:-22.770   1st Qu.: 1.642   1st Qu.:   2.697
##  Median :   32.517   Median :  0.000   Median : 8.133   Median :  66.146
##  Mean   :  417.264   Mean   : -4.901   Mean   :10.383   Mean   : 195.864
##  3rd Qu.:  222.647   3rd Qu.:  8.277   3rd Qu.:16.327   3rd Qu.: 266.576
##  Max.   :26232.208   Max.   : 75.659   Max.   :43.412   Max.   :1884.565
##  NA's   :19216       NA's   :19216     NA's   :19216    NA's   :19216
##   avg_yaw_arm       stddev_yaw_arm     var_yaw_arm         gyros_arm_x
##  Min.   :-173.440   Min.   :  0.000   Min.   :    0.000   Min.   :-6.37000
##  1st Qu.: -29.198   1st Qu.:  2.577   1st Qu.:    6.642   1st Qu.:-1.33000
##  Median :   0.000   Median : 16.682   Median :  278.309   Median : 0.08000
##  Mean   :   2.359   Mean   : 22.270   Mean   : 1055.933   Mean   : 0.04277
##  3rd Qu.:  38.185   3rd Qu.: 35.984   3rd Qu.: 1294.850   3rd Qu.: 1.57000
##  Max.   : 152.000   Max.   :177.044   Max.   :31344.568   Max.   : 4.87000
##  NA's   :19216      NA's   :19216     NA's   :19216
##   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 skewness_roll_arm
##  #DIV/0! :   78    #DIV/0! :   80     #DIV/0! :   11   #DIV/0! :   77
##  -0.02438:    1    -0.00484:    1     0.55844 :    2   -0.00051:    1
##  -0.04190:    1    -0.01311:    1     0.65132 :    2   -0.00696:    1
##  -0.05051:    1    -0.02967:    1     -0.01548:    1   -0.01884:    1
##  -0.05695:    1    -0.07394:    1     -0.01749:    1   -0.03359:    1
##  (Other) :  324    (Other) :  322     (Other) :  389   (Other) :  325
##  NA's    :19216    NA's    :19216     NA's    :19216   NA's    :19216
##  skewness_pitch_arm skewness_yaw_arm  max_roll_arm     max_picth_arm
##  #DIV/0! :   80     #DIV/0! :   11   Min.   :-73.100   Min.   :-173.000
##  -0.00184:    1     -1.62032:    2   1st Qu.: -0.175   1st Qu.:  -1.975
##  -0.01185:    1     0.55053 :    2   Median :  4.950   Median :  23.250
##  -0.01247:    1     -0.00311:    1   Mean   : 11.236   Mean   :  35.751
##  -0.02063:    1     -0.00562:    1   3rd Qu.: 26.775   3rd Qu.:  95.975
##  (Other) :  322     (Other) :  389   Max.   : 85.500   Max.   : 180.000
##  NA's    :19216     NA's    :19216   NA's   :19216     NA's   :19216
##   max_yaw_arm     min_roll_arm    min_pitch_arm      min_yaw_arm
##  Min.   : 4.00   Min.   :-89.10   Min.   :-180.00   Min.   : 1.00
##  1st Qu.:29.00   1st Qu.:-41.98   1st Qu.: -72.62   1st Qu.: 8.00
##  Median :34.00   Median :-22.45   Median : -33.85   Median :13.00
##  Mean   :35.46   Mean   :-21.22   Mean   : -33.92   Mean   :14.66
##  3rd Qu.:41.00   3rd Qu.:  0.00   3rd Qu.:   0.00   3rd Qu.:19.00
##  Max.   :65.00   Max.   : 66.40   Max.   : 152.00   Max.   :38.00
##  NA's   :19216   NA's   :19216    NA's   :19216     NA's   :19216
##  amplitude_roll_arm amplitude_pitch_arm amplitude_yaw_arm roll_dumbbell
##  Min.   :  0.000    Min.   :  0.000     Min.   : 0.00     Min.   :-153.71
##  1st Qu.:  5.425    1st Qu.:  9.925     1st Qu.:13.00     1st Qu.: -18.49
##  Median : 28.450    Median : 54.900     Median :22.00     Median :  48.17
##  Mean   : 32.452    Mean   : 69.677     Mean   :20.79     Mean   :  23.84
##  3rd Qu.: 50.960    3rd Qu.:115.175     3rd Qu.:28.75     3rd Qu.:  67.61
##  Max.   :119.500    Max.   :360.000     Max.   :52.00     Max.   : 153.55
##  NA's   :19216      NA's   :19216       NA's   :19216
##  pitch_dumbbell     yaw_dumbbell      kurtosis_roll_dumbbell
##  Min.   :-149.59   Min.   :-150.871   #DIV/0!:    5
##  1st Qu.: -40.89   1st Qu.: -77.644   -0.2583:    2
##  Median : -20.96   Median :  -3.324   -0.3705:    2
##  Mean   : -10.78   Mean   :   1.674   -0.5855:    2
##  3rd Qu.:  17.50   3rd Qu.:  79.643   -2.0851:    2
##  Max.   : 149.40   Max.   : 154.952   (Other):  393
##                                       NA's   :19216
##  kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_roll_dumbbell
##  -0.5464:    2           #DIV/0!:  406         #DIV/0!:    4
##  -0.9334:    2           NA's   :19216         -0.9324:    2
```

```
##  -2.0833:    2                                   0.1110 :    2
##  -2.0851:    2                                   1.0312 :    2
##  -2.0889:    2                                   -0.0082:    1
##  (Other):  396                                   (Other):  395
##  NA's   :19216                                   NA's   :19216
##  skewness_pitch_dumbbell skewness_yaw_dumbbell max_roll_dumbbell
##  -0.2328:    2           #DIV/0!:  406         Min.   :-70.10
##  -0.3521:    2           NA's   :19216         1st Qu.:-27.15
##  -0.7036:    2                                 Median : 14.85
##  0.1090 :    2                                 Mean   : 13.76
##  1.0326 :    2                                 3rd Qu.: 50.58
##  (Other):  396                                 Max.   :137.00
##  NA's   :19216                                 NA's   :19216
##  max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell min_pitch_dumbbell
##  Min.   :-112.90    -0.6   :   20    Min.   :-149.60   Min.   :-147.00
##  1st Qu.: -66.70    0.2    :   19    1st Qu.: -59.67   1st Qu.: -91.80
##  Median :  40.05    -0.8   :   18    Median : -43.55   Median : -66.15
##  Mean   :  32.75    -0.3   :   16    Mean   : -41.24   Mean   : -33.18
##  3rd Qu.: 133.22    -0.2   :   15    3rd Qu.: -25.20   3rd Qu.:  21.20
##  Max.   : 155.00    (Other):  318    Max.   :  73.20   Max.   : 120.90
##  NA's   :19216      NA's   :19216    NA's   :19216     NA's   :19216
##  min_yaw_dumbbell amplitude_roll_dumbbell amplitude_pitch_dumbbell
##  -0.6   :   20    Min.   :  0.00          Min.   :  0.00
##  0.2    :   19    1st Qu.: 14.97          1st Qu.: 17.06
##  -0.8   :   18    Median : 35.05          Median : 41.73
##  -0.3   :   16    Mean   : 55.00          Mean   : 65.93
##  -0.2   :   15    3rd Qu.: 81.04          3rd Qu.: 99.55
##  (Other):  318    Max.   :256.48          Max.   :273.59
##  NA's   :19216    NA's   :19216           NA's   :19216
##  amplitude_yaw_dumbbell total_accel_dumbbell var_accel_dumbbell
##  #DIV/0!:    5          Min.   : 0.00        Min.   :  0.000
##  0.00   :  401          1st Qu.: 4.00        1st Qu.:  0.378
##  NA's   :19216          Median :10.00        Median :  1.000
##                         Mean   :13.72        Mean   :  4.388
##                         3rd Qu.:19.00        3rd Qu.:  3.434
##                         Max.   :58.00        Max.   :230.428
##                                              NA's   :19216
##  avg_roll_dumbbell stddev_roll_dumbbell var_roll_dumbbell  avg_pitch_dumbbell
##  Min.   :-128.96   Min.   :  0.000      Min.   :    0.00   Min.   :-70.73
##  1st Qu.: -12.33   1st Qu.:  4.639      1st Qu.:   21.52   1st Qu.:-42.00
##  Median :  48.23   Median : 12.204      Median :  148.95   Median :-19.91
##  Mean   :  23.86   Mean   : 20.761      Mean   : 1020.27   Mean   :-12.33
##  3rd Qu.:  64.37   3rd Qu.: 26.356      3rd Qu.:  694.65   3rd Qu.: 13.21
##  Max.   : 125.99   Max.   :123.778      Max.   :15321.01   Max.   : 94.28
##  NA's   :19216     NA's   :19216        NA's   :19216      NA's   :19216
##  stddev_pitch_dumbbell var_pitch_dumbbell avg_yaw_dumbbell
##  Min.   : 0.000        Min.   :   0.00    Min.   :-117.950
##  1st Qu.: 3.482        1st Qu.:  12.12    1st Qu.: -76.696
##  Median : 8.089        Median :  65.44    Median :  -4.505
##  Mean   :13.147        Mean   : 350.31    Mean   :   0.202
##  3rd Qu.:19.238        3rd Qu.: 370.11    3rd Qu.:  71.234
##  Max.   :82.680        Max.   :6836.02    Max.   : 134.905
##  NA's   :19216         NA's   :19216      NA's   :19216
##  stddev_yaw_dumbbell var_yaw_dumbbell   gyros_dumbbell_x    gyros_dumbbell_y
```

```
##  Min.   :  0.000     Min.   :    0.00   Min.   :-204.0000   Min.   :-2.10000
##  1st Qu.:  3.885     1st Qu.:   15.09   1st Qu.:  -0.0300   1st Qu.:-0.14000
##  Median : 10.264     Median :  105.35   Median :   0.1300   Median : 0.03000
##  Mean   : 16.647     Mean   :  589.84   Mean   :   0.1611   Mean   : 0.04606
##  3rd Qu.: 24.674     3rd Qu.:  608.79   3rd Qu.:   0.3500   3rd Qu.: 0.21000
##  Max.   :107.088     Max.   :11467.91   Max.   :   2.2200   Max.   :52.00000
##  NA's   :19216       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   #DIV/0!:   84
##  1st Qu.:  0.00   1st Qu.: -68.60   -0.8079:    2
##  Median :  9.24   Median :   0.00   -0.9169:    2
##  Mean   : 10.71   Mean   :  19.21   -0.0227:    1
##  3rd Qu.: 28.40   3rd Qu.: 110.00   -0.0359:    1
##  Max.   : 89.80   Max.   : 180.00   (Other):  316
##                                     NA's   :19216
##  kurtosis_picth_forearm kurtosis_yaw_forearm skewness_roll_forearm
##  #DIV/0!:   85          #DIV/0!:  406        #DIV/0!:   83
##  -0.0073:    1          NA's   :19216        -0.1912:    2
##  -0.0442:    1                               -0.4126:    2
##  -0.0489:    1                               -0.0004:    1
##  -0.0523:    1                               -0.0013:    1
##  (Other):  317                               (Other):  317
##  NA's   :19216                               NA's   :19216
##  skewness_pitch_forearm skewness_yaw_forearm max_roll_forearm max_picth_forearm
##  #DIV/0!:   85          #DIV/0!:  406        Min.   :-66.60   Min.   :-151.00
##  0.0000 :    4          NA's   :19216        1st Qu.:  0.00   1st Qu.:   0.00
##  -0.6992:    2                               Median : 26.80   Median : 113.00
##  -0.0113:    1                               Mean   : 24.49   Mean   :  81.49
##  -0.0131:    1                               3rd Qu.: 45.95   3rd Qu.: 174.75
##  (Other):  313                               Max.   : 89.80   Max.   : 180.00
##  NA's   :19216                               NA's   :19216    NA's   :19216
##  max_yaw_forearm min_roll_forearm  min_pitch_forearm min_yaw_forearm
##  #DIV/0!:   84   Min.   :-72.500   Min.   :-180.00   #DIV/0!:   84
##  -1.2   :   32   1st Qu.: -6.075   1st Qu.:-175.00   -1.2   :   32
##  -1.3   :   31   Median :  0.000   Median : -61.00   -1.3   :   31
##  -1.4   :   24   Mean   : -0.167   Mean   : -57.57   -1.4   :   24
##  -1.5   :   24   3rd Qu.: 12.075   3rd Qu.:   0.00   -1.5   :   24
##  (Other):  211   Max.   : 62.100   Max.   : 167.00   (Other):  211
```

```
##  NA's   :19216   NA's   :19216     NA's   :19216     NA's   :19216
##  amplitude_roll_forearm amplitude_pitch_forearm amplitude_yaw_forearm
##  Min.   :  0.000        Min.   :  0.0           #DIV/0!:   84
##  1st Qu.:  1.125        1st Qu.:  2.0           0.00   :  322
##  Median : 17.770        Median : 83.7           NA's   :19216
##  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
##  total_accel_forearm var_accel_forearm avg_roll_forearm   stddev_roll_forearm
##  Min.   :  0.00      Min.   :  0.000   Min.   :-177.234   Min.   :  0.000
##  1st Qu.: 29.00      1st Qu.:  6.759   1st Qu.:  -0.909   1st Qu.:  0.428
##  Median : 36.00      Median : 21.165   Median :  11.172   Median :  8.030
##  Mean   : 34.72      Mean   : 33.502   Mean   :  33.165   Mean   : 41.986
##  3rd Qu.: 41.00      3rd Qu.: 51.240   3rd Qu.: 107.132   3rd Qu.: 85.373
##  Max.   :108.00      Max.   :172.606   Max.   : 177.256   Max.   :179.171
##                      NA's   :19216     NA's   :19216      NA's   :19216
##  var_roll_forearm   avg_pitch_forearm stddev_pitch_forearm var_pitch_forearm
##  Min.   :    0.00   Min.   :-68.17    Min.   : 0.000       Min.   :   0.000
##  1st Qu.:    0.18   1st Qu.:  0.00    1st Qu.: 0.336       1st Qu.:   0.113
##  Median :   64.48   Median : 12.02    Median : 5.516       Median :  30.425
##  Mean   : 5274.10   Mean   : 11.79    Mean   : 7.977       Mean   : 139.593
##  3rd Qu.: 7289.08   3rd Qu.: 28.48    3rd Qu.:12.866       3rd Qu.: 165.532
##  Max.   :32102.24   Max.   : 72.09    Max.   :47.745       Max.   :2279.617
##  NA's   :19216      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 classe
##  Min.   :-446.00   Min.   :-1280.0   Min.   :-896.0   Min.   :-973.0   A:5580
##  1st Qu.:-182.00   1st Qu.: -616.0   1st Qu.:   2.0   1st Qu.: 191.0   B:3797
##  Median : -39.00   Median : -378.0   Median : 591.0   Median : 511.0   C:3422
##  Mean   : -55.29   Mean   : -312.6   Mean   : 380.1   Mean   : 393.6   D:3216
##  3rd Qu.:  26.00   3rd Qu.:  -73.0   3rd Qu.: 737.0   3rd Qu.: 653.0   E:3607
##  Max.   : 291.00   Max.   :  672.0   Max.   :1480.0   Max.   :1090.0
##
```

```
View(head(train))

# Removing 100 variables with 'NA'values
train1 <- train[, apply(train, 2, function(x) !any(is.na(x)))]
```

```
# Removing first 7 variables as they are not pertinent features
train2 <- train1[, -c(1:7)]

# Loading in the testing data
test <- read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv", header=TRUE)
dim(test)
```

```
## [1]  20 160
```

I will now take the training dataset and split it into training and testing sets using the createDataParition function within the caret package.

```
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
split_train <- createDataPartition(y=train2$classe, p=0.7, list=FALSE)
training1 <- train2[split_train, ]
dim(training1)
```

```
## [1] 13737    53
```

```
testing1 <- train2[-split_train, ]
dim(testing1)
```

```
## [1] 5885   53
```

## Model Building

I will test three models (decision trees, random forests, and generalized boosted trees) and see which performs better. I will do a 3-fold cross-validation.

```
## Cross-validation
trcontrol <- trainControl(method="cv", number=3)

## Decision Trees
fit1 <-train(classe ~. ,data=training1, method="rpart", trControl=trcontrol)

## Random forests
rf1 <- train(classe ~. ,data=training1, method="rf", trControl=trcontrol)

## Generalized  boosted trees
gbm1 <- train(classe ~. ,data=training1, method="gbm", trControl=trcontrol)
```

```
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1339
##      2        1.5219             nan     0.1000    0.0890
##      3        1.4621             nan     0.1000    0.0691
##      4        1.4175             nan     0.1000    0.0537
##      5        1.3815             nan     0.1000    0.0505
##      6        1.3483             nan     0.1000    0.0463
##      7        1.3190             nan     0.1000    0.0349
##      8        1.2959             nan     0.1000    0.0319
##      9        1.2731             nan     0.1000    0.0260
##     10        1.2546             nan     0.1000    0.0298
```

```
##     20        1.0950             nan     0.1000    0.0181
##     40        0.9234             nan     0.1000    0.0101
##     60        0.8145             nan     0.1000    0.0051
##     80        0.7359             nan     0.1000    0.0042
##    100        0.6736             nan     0.1000    0.0041
##    120        0.6219             nan     0.1000    0.0029
##    140        0.5783             nan     0.1000    0.0013
##    150        0.5586             nan     0.1000    0.0025
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1869
##      2        1.4890             nan     0.1000    0.1348
##      3        1.4021             nan     0.1000    0.1033
##      4        1.3357             nan     0.1000    0.0870
##      5        1.2784             nan     0.1000    0.0729
##      6        1.2325             nan     0.1000    0.0690
##      7        1.1893             nan     0.1000    0.0661
##      8        1.1466             nan     0.1000    0.0505
##      9        1.1144             nan     0.1000    0.0477
##     10        1.0841             nan     0.1000    0.0393
##     20        0.8842             nan     0.1000    0.0183
##     40        0.6723             nan     0.1000    0.0117
##     60        0.5497             nan     0.1000    0.0058
##     80        0.4598             nan     0.1000    0.0042
##    100        0.3952             nan     0.1000    0.0053
##    120        0.3447             nan     0.1000    0.0033
##    140        0.3024             nan     0.1000    0.0029
##    150        0.2848             nan     0.1000    0.0021
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.2355
##      2        1.4593             nan     0.1000    0.1543
##      3        1.3592             nan     0.1000    0.1286
##      4        1.2762             nan     0.1000    0.1089
##      5        1.2081             nan     0.1000    0.0871
##      6        1.1522             nan     0.1000    0.0875
##      7        1.0976             nan     0.1000    0.0605
##      8        1.0586             nan     0.1000    0.0631
##      9        1.0193             nan     0.1000    0.0528
##     10        0.9856             nan     0.1000    0.0489
##     20        0.7494             nan     0.1000    0.0199
##     40        0.5278             nan     0.1000    0.0094
##     60        0.4011             nan     0.1000    0.0048
##     80        0.3199             nan     0.1000    0.0036
##    100        0.2641             nan     0.1000    0.0017
##    120        0.2191             nan     0.1000    0.0024
##    140        0.1853             nan     0.1000    0.0015
##    150        0.1716             nan     0.1000    0.0013
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1306
##      2        1.5221             nan     0.1000    0.0884
##      3        1.4630             nan     0.1000    0.0688
##      4        1.4169             nan     0.1000    0.0523
```

```
##      5        1.3810             nan     0.1000    0.0458
##      6        1.3510             nan     0.1000    0.0453
##      7        1.3222             nan     0.1000    0.0341
##      8        1.2985             nan     0.1000    0.0360
##      9        1.2755             nan     0.1000    0.0334
##     10        1.2523             nan     0.1000    0.0295
##     20        1.1003             nan     0.1000    0.0180
##     40        0.9279             nan     0.1000    0.0075
##     60        0.8212             nan     0.1000    0.0065
##     80        0.7421             nan     0.1000    0.0048
##    100        0.6777             nan     0.1000    0.0040
##    120        0.6250             nan     0.1000    0.0018
##    140        0.5811             nan     0.1000    0.0025
##    150        0.5617             nan     0.1000    0.0016
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1790
##      2        1.4892             nan     0.1000    0.1329
##      3        1.4036             nan     0.1000    0.1061
##      4        1.3349             nan     0.1000    0.0801
##      5        1.2833             nan     0.1000    0.0683
##      6        1.2388             nan     0.1000    0.0587
##      7        1.2005             nan     0.1000    0.0644
##      8        1.1608             nan     0.1000    0.0525
##      9        1.1261             nan     0.1000    0.0429
##     10        1.0985             nan     0.1000    0.0418
##     20        0.8945             nan     0.1000    0.0201
##     40        0.6784             nan     0.1000    0.0091
##     60        0.5530             nan     0.1000    0.0080
##     80        0.4650             nan     0.1000    0.0082
##    100        0.3945             nan     0.1000    0.0032
##    120        0.3440             nan     0.1000    0.0024
##    140        0.3062             nan     0.1000    0.0023
##    150        0.2886             nan     0.1000    0.0027
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.2260
##      2        1.4619             nan     0.1000    0.1668
##      3        1.3556             nan     0.1000    0.1229
##      4        1.2796             nan     0.1000    0.0941
##      5        1.2186             nan     0.1000    0.0796
##      6        1.1651             nan     0.1000    0.0786
##      7        1.1158             nan     0.1000    0.0687
##      8        1.0711             nan     0.1000    0.0656
##      9        1.0285             nan     0.1000    0.0599
##     10        0.9906             nan     0.1000    0.0604
##     20        0.7499             nan     0.1000    0.0268
##     40        0.5220             nan     0.1000    0.0105
##     60        0.3983             nan     0.1000    0.0051
##     80        0.3156             nan     0.1000    0.0034
##    100        0.2613             nan     0.1000    0.0035
##    120        0.2192             nan     0.1000    0.0018
##    140        0.1847             nan     0.1000    0.0017
##    150        0.1715             nan     0.1000    0.0015
```

```
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1340
##      2        1.5216             nan     0.1000    0.0881
##      3        1.4615             nan     0.1000    0.0676
##      4        1.4172             nan     0.1000    0.0554
##      5        1.3811             nan     0.1000    0.0458
##      6        1.3515             nan     0.1000    0.0439
##      7        1.3227             nan     0.1000    0.0407
##      8        1.2960             nan     0.1000    0.0308
##      9        1.2764             nan     0.1000    0.0309
##     10        1.2562             nan     0.1000    0.0347
##     20        1.0968             nan     0.1000    0.0153
##     40        0.9246             nan     0.1000    0.0083
##     60        0.8183             nan     0.1000    0.0079
##     80        0.7382             nan     0.1000    0.0040
##    100        0.6782             nan     0.1000    0.0027
##    120        0.6258             nan     0.1000    0.0034
##    140        0.5818             nan     0.1000    0.0024
##    150        0.5627             nan     0.1000    0.0022
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1840
##      2        1.4880             nan     0.1000    0.1296
##      3        1.4028             nan     0.1000    0.1023
##      4        1.3355             nan     0.1000    0.0843
##      5        1.2804             nan     0.1000    0.0751
##      6        1.2329             nan     0.1000    0.0606
##      7        1.1930             nan     0.1000    0.0628
##      8        1.1536             nan     0.1000    0.0595
##      9        1.1157             nan     0.1000    0.0402
##     10        1.0886             nan     0.1000    0.0433
##     20        0.8885             nan     0.1000    0.0198
##     40        0.6760             nan     0.1000    0.0108
##     60        0.5469             nan     0.1000    0.0062
##     80        0.4621             nan     0.1000    0.0052
##    100        0.3955             nan     0.1000    0.0031
##    120        0.3437             nan     0.1000    0.0030
##    140        0.3019             nan     0.1000    0.0022
##    150        0.2844             nan     0.1000    0.0016
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.2436
##      2        1.4560             nan     0.1000    0.1642
##      3        1.3519             nan     0.1000    0.1262
##      4        1.2731             nan     0.1000    0.1029
##      5        1.2071             nan     0.1000    0.0906
##      6        1.1506             nan     0.1000    0.0743
##      7        1.1015             nan     0.1000    0.0649
##      8        1.0593             nan     0.1000    0.0665
##      9        1.0176             nan     0.1000    0.0533
##     10        0.9831             nan     0.1000    0.0448
##     20        0.7510             nan     0.1000    0.0211
##     40        0.5238             nan     0.1000    0.0113
```

```
##     60        0.4036             nan     0.1000    0.0075
##     80        0.3252             nan     0.1000    0.0049
##    100        0.2646             nan     0.1000    0.0027
##    120        0.2221             nan     0.1000    0.0028
##    140        0.1855             nan     0.1000    0.0010
##    150        0.1728             nan     0.1000    0.0008
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.2338
##      2        1.4600             nan     0.1000    0.1578
##      3        1.3591             nan     0.1000    0.1250
##      4        1.2792             nan     0.1000    0.0987
##      5        1.2165             nan     0.1000    0.0954
##      6        1.1569             nan     0.1000    0.0771
##      7        1.1090             nan     0.1000    0.0697
##      8        1.0648             nan     0.1000    0.0601
##      9        1.0273             nan     0.1000    0.0520
##     10        0.9942             nan     0.1000    0.0575
##     20        0.7583             nan     0.1000    0.0259
##     40        0.5314             nan     0.1000    0.0112
##     60        0.4093             nan     0.1000    0.0058
##     80        0.3258             nan     0.1000    0.0043
##    100        0.2666             nan     0.1000    0.0048
##    120        0.2236             nan     0.1000    0.0026
##    140        0.1926             nan     0.1000    0.0007
##    150        0.1791             nan     0.1000    0.0012
```

## Prediction

From the confusion matrix output, the accuracy for decision trees is 0.4929 resulting in an out of sample error of .5071. This is not a very good model as it is no better than random guessing. Next, I'll look at the random forest model. The accuracy here is excellent with an out of sample error of .0093. Generalized boosted trees also result in a good accuracy. The out of sample error is .0399. Overall, the random forest model predicts the best.

```
# Decision Tree Prediction
fit1predict <- predict(fit1, newdata=testing1)
confusionMatrix(testing1$classe, fit1predict)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1536   26  107    0    5
##          B  472  389  278    0    0
##          C  500   24  502    0    0
##          D  427  168  369    0    0
##          E  153  158  303    0  468
##
## Overall Statistics
##
##                Accuracy : 0.4919
##                  95% CI : (0.4791, 0.5048)
##     No Information Rate : 0.5247
```

```
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.3355
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.4974   0.5085   0.3220       NA  0.98943
## Specificity            0.9507   0.8535   0.8789   0.8362  0.88655
## Pos Pred Value         0.9176   0.3415   0.4893       NA  0.43253
## Neg Pred Value         0.6314   0.9208   0.7825       NA  0.99896
## Prevalence             0.5247   0.1300   0.2649   0.0000  0.08037
## Detection Rate         0.2610   0.0661   0.0853   0.0000  0.07952
## Detection Prevalence   0.2845   0.1935   0.1743   0.1638  0.18386
## Balanced Accuracy      0.7240   0.6810   0.6004       NA  0.93799
```

```
#Random Forest Prediction
rf1predict <- predict(rf1, newdata=testing1)
confusionMatrix(testing1$classe, rf1predict)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1670    3    1    0    0
##          B    8 1129    2    0    0
##          C    0    2 1019    5    0
##          D    0    1   10  953    0
##          E    0    0    2    8 1072
##
## Overall Statistics
##
##                Accuracy : 0.9929
##                  95% CI : (0.9904, 0.9949)
##     No Information Rate : 0.2851
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.991
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9952   0.9947   0.9855   0.9865   1.0000
## Specificity            0.9990   0.9979   0.9986   0.9978   0.9979
## Pos Pred Value         0.9976   0.9912   0.9932   0.9886   0.9908
## Neg Pred Value         0.9981   0.9987   0.9969   0.9974   1.0000
## Prevalence             0.2851   0.1929   0.1757   0.1641   0.1822
## Detection Rate         0.2838   0.1918   0.1732   0.1619   0.1822
## Detection Prevalence   0.2845   0.1935   0.1743   0.1638   0.1839
## Balanced Accuracy      0.9971   0.9963   0.9920   0.9922   0.9990
```

```r
#Generalized Boosted Trees
gbm1predict <- predict(gbm1, newdata=testing1)
confusionMatrix(testing1$classe, gbm1predict)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1652   14    7    0    1
##          B   45 1066   25    1    2
##          C    0   29  982   14    1
##          D    0    4   29  921   10
##          E    1   21   11   17 1032
##
## Overall Statistics
##
##                Accuracy : 0.9606
##                  95% CI : (0.9553, 0.9654)
##     No Information Rate : 0.2885
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9501
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9729   0.9400   0.9317   0.9664   0.9866
## Specificity            0.9947   0.9846   0.9909   0.9913   0.9897
## Pos Pred Value         0.9869   0.9359   0.9571   0.9554   0.9538
## Neg Pred Value         0.9891   0.9857   0.9852   0.9935   0.9971
## Prevalence             0.2885   0.1927   0.1791   0.1619   0.1777
## Detection Rate         0.2807   0.1811   0.1669   0.1565   0.1754
## Detection Prevalence   0.2845   0.1935   0.1743   0.1638   0.1839
## Balanced Accuracy      0.9838   0.9623   0.9613   0.9789   0.9881
```

I will use the random forest model to predict the manner in which the 20 participants did the exercise using the test dataset.

```r
predict20 <- predict(rf1, newdata=test)
predict20
```

```
##  [1] 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
```