Balish Practical Machine Learning

Background per assignment:

Using devices such as Jawbone Up, Nike FuelBand, and Fitbit it is now possible to collect a large amount of data about personal activity relatively inexpensively. One thing that people regularly do is quantify how much of a particular activity they do, but they rarely quantify how well they do it. In this project, your goal will be to use data from accelerometers on the belt, forearm, arm, and dumbell of 6 participants. They were asked to perform barbell lifts correctly and incorrectly in 5 different ways.

Goal:

The goal of the project is to predict the manner in which they did the exercise. This is the “classe” variable in the training set. You may use any of the other variables to predict with. You should create a report describing how you built your model, how you used cross validation, what you think the expected out of sample error is, and why you made the choices you did.

Data provided per web link includes a training and a test set. I will read in the data (training) and the needed caret library and explore some features:

```
library(caret)

## Warning: package 'caret' was built under R version 3.1.2

## Loading required package: lattice
## Loading required package: ggplot2

pmltrain<-read.csv("pml-training.csv")
summary(pmltrain)

##        X            user_name    raw_timestamp_part_1 raw_timestamp_part_2
##  Min.   :    1   adelmo  :3892   Min.   :1.32e+09     Min.   :   294
##  1st Qu.: 4906   carlitos:3112   1st Qu.:1.32e+09     1st Qu.:252912
##  Median : 9812   charles :3536   Median :1.32e+09     Median :496380
##  Mean   : 9812   eurico  :3070   Mean   :1.32e+09     Mean   :500656
##  3rd Qu.:14717   jeremy  :3402   3rd Qu.:1.32e+09     3rd Qu.:751891
##  Max.   :19622   pedro   :2610   Max.   :1.32e+09     Max.   :998801
##
##           cvtd_timestamp  new_window    num_window    roll_belt
##  28/11/2011 14:14: 1498   no :19216   Min.   :  1   Min.   :-28.9
##  05/12/2011 11:24: 1497   yes:  406   1st Qu.:222   1st Qu.:  1.1
##  30/11/2011 17:11: 1440               Median :424   Median :113.0
##  05/12/2011 11:25: 1425               Mean   :431   Mean   : 64.4
##  02/12/2011 14:57: 1380               3rd Qu.:644   3rd Qu.:123.0
##  02/12/2011 13:34: 1375               Max.   :864   Max.   :162.0
##  (Other)         :11007
##    pitch_belt        yaw_belt      total_accel_belt kurtosis_roll_belt
```

```
##  Min.   :-55.80   Min.   :-180.0   Min.   : 0.0             :19216
##  1st Qu.:  1.76   1st Qu.: -88.3   1st Qu.: 3.0     #DIV/0!  :   10
##  Median :  5.28   Median : -13.0   Median :17.0     -1.908453:    2
##  Mean   :  0.31   Mean   : -11.2   Mean   :11.3     -0.016850:    1
##  3rd Qu.: 14.90   3rd Qu.:  12.9   3rd Qu.:18.0     -0.021024:    1
##  Max.   : 60.30   Max.   : 179.0   Max.   :29.0     -0.025513:    1
##                                                     (Other)  :  391
##  kurtosis_picth_belt kurtosis_yaw_belt skewness_roll_belt
##          :19216            :19216              :19216
##  #DIV/0!  :   32     #DIV/0!:  406     #DIV/0!  :    9
##  47.000000:    4                       0.000000 :    4
##  -0.150950:    3                       0.422463 :    2
##  -0.684748:    3                       -0.003095:    1
##  -1.750749:    3                       -0.010002:    1
##  (Other)  :  361                       (Other)  :  389
##  skewness_roll_belt.1 skewness_yaw_belt max_roll_belt   max_picth_belt
##          :19216             :19216     Min.   :-94     Min.   : 3
##  #DIV/0!  :   32      #DIV/0!:  406     1st Qu.:-88     1st Qu.: 5
##  0.000000 :    4                       Median : -5     Median :18
##  -2.156553:    3                       Mean   : -7     Mean   :13
##  -3.072669:    3                       3rd Qu.: 18     3rd Qu.:19
##  -6.324555:    3                       Max.   :180     Max.   :30
##  (Other)  :  361                       NA's   :19216   NA's   :19216
##   max_yaw_belt   min_roll_belt   min_pitch_belt   min_yaw_belt
##         :19216   Min.   :-180    Min.   : 0             :19216
##  -1.1   :   30   1st Qu.: -88    1st Qu.: 3      -1.1   :   30
##  -1.4   :   29   Median :  -8    Median :16      -1.4   :   29
##  -1.2   :   26   Mean   : -10    Mean   :11      -1.2   :   26
##  -0.9   :   24   3rd Qu.:   9    3rd Qu.:17      -0.9   :   24
##  -1.3   :   22   Max.   : 173    Max.   :23      -1.3   :   22
##  (Other):  275   NA's   :19216   NA's   :19216   (Other):  275
##  amplitude_roll_belt amplitude_pitch_belt amplitude_yaw_belt
##  Min.   :  0         Min.   : 0                  :19216
##  1st Qu.:  0         1st Qu.: 1           #DIV/0!:   10
##  Median :  1         Median : 1           0.00   :   12
##  Mean   :  4         Mean   : 2           0.0000 :  384
##  3rd Qu.:  2         3rd Qu.: 2
##  Max.   :360         Max.   :12
##  NA's   :19216       NA's   :19216
##  var_total_accel_belt avg_roll_belt   stddev_roll_belt var_roll_belt
##  Min.   : 0           Min.   :-27     Min.   : 0       Min.   :  0
```

```
##  1st Qu.: 0           1st Qu.:  1     1st Qu.: 0       1st Qu.:  0
##  Median : 0           Median :116     Median : 0       Median :  0
##  Mean   : 1           Mean   : 68     Mean   : 1       Mean   :  8
##  3rd Qu.: 0           3rd Qu.:123     3rd Qu.: 1       3rd Qu.:  0
##  Max.   :16           Max.   :157     Max.   :14       Max.   :201
##  NA's   :19216        NA's   :19216   NA's   :19216    NA's   :19216
##  avg_pitch_belt  stddev_pitch_belt var_pitch_belt   avg_yaw_belt
##  Min.   :-51     Min.   :0         Min.   : 0      Min.   :-138
##  1st Qu.:  2     1st Qu.:0         1st Qu.: 0      1st Qu.: -88
##  Median :  5     Median :0         Median : 0      Median :  -7
##  Mean   :  1     Mean   :1         Mean   : 1      Mean   :  -9
##  3rd Qu.: 16     3rd Qu.:1         3rd Qu.: 0      3rd Qu.:  14
##  Max.   : 60     Max.   :4         Max.   :16      Max.   : 174
##  NA's   :19216   NA's   :19216     NA's   :19216   NA's   :19216
##  stddev_yaw_belt  var_yaw_belt    gyros_belt_x      gyros_belt_y
##  Min.   :  0     Min.   :    0   Min.   :-1.0400   Min.   :-0.6400
##  1st Qu.:  0     1st Qu.:    0   1st Qu.:-0.0300   1st Qu.: 0.0000
##  Median :  0     Median :    0   Median : 0.0300   Median : 0.0200
##  Mean   :  1     Mean   :  107   Mean   :-0.0056   Mean   : 0.0396
##  3rd Qu.:  1     3rd Qu.:    0   3rd Qu.: 0.1100   3rd Qu.: 0.1100
##  Max.   :177     Max.   :31183   Max.   : 2.2200   Max.   : 0.6400
##  NA's   :19216   NA's   :19216
##   gyros_belt_z     accel_belt_x      accel_belt_y    accel_belt_z
##  Min.   :-1.460   Min.   :-120.00   Min.   :-69.0   Min.   :-275.0
##  1st Qu.:-0.200   1st Qu.: -21.00   1st Qu.:  3.0   1st Qu.:-162.0
##  Median :-0.100   Median : -15.00   Median : 35.0   Median :-152.0
##  Mean   :-0.130   Mean   :  -5.59   Mean   : 30.1   Mean   : -72.6
##  3rd Qu.:-0.020   3rd Qu.:  -5.00   3rd Qu.: 61.0   3rd Qu.:  27.0
##  Max.   : 1.620   Max.   :  85.00   Max.   :164.0   Max.   : 105.0
##
##  magnet_belt_x   magnet_belt_y magnet_belt_z     roll_arm
##  Min.   :-52.0   Min.   :354   Min.   :-623   Min.   :-180.0
##  1st Qu.:  9.0   1st Qu.:581   1st Qu.:-375   1st Qu.: -31.8
##  Median : 35.0   Median :601   Median :-320   Median :   0.0
##  Mean   : 55.6   Mean   :594   Mean   :-346   Mean   :  17.8
##  3rd Qu.: 59.0   3rd Qu.:610   3rd Qu.:-306   3rd Qu.:  77.3
##  Max.   :485.0   Max.   :673   Max.   : 293   Max.   : 180.0
##
##    pitch_arm         yaw_arm        total_accel_arm var_accel_arm
##  Min.   :-88.80   Min.   :-180.00   Min.   : 1.0    Min.   :  0
##  1st Qu.:-25.90   1st Qu.: -43.10   1st Qu.:17.0    1st Qu.:  9
```

```
##  Median :  0.00   Median :   0.00   Median :27.0    Median : 41
##  Mean   : -4.61   Mean   :  -0.62   Mean   :25.5    Mean   : 53
##  3rd Qu.: 11.20   3rd Qu.:  45.88   3rd Qu.:33.0    3rd Qu.: 76
##  Max.   : 88.50   Max.   : 180.00   Max.   :66.0    Max.   :332
##                                                     NA's   :19216
##   avg_roll_arm   stddev_roll_arm  var_roll_arm   avg_pitch_arm
##  Min.   :-167    Min.   :  0     Min.   :    0   Min.   :-82
##  1st Qu.: -38    1st Qu.:  1     1st Qu.:    2   1st Qu.:-23
##  Median :   0    Median :  6     Median :   33   Median :  0
##  Mean   :  13    Mean   : 11     Mean   :  417   Mean   : -5
##  3rd Qu.:  76    3rd Qu.: 15     3rd Qu.:  223   3rd Qu.:  8
##  Max.   : 163    Max.   :162     Max.   :26232   Max.   : 76
##  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       Min.   :   0    Min.   :-173    Min.   :  0
##  1st Qu.: 2       1st Qu.:   3    1st Qu.: -29    1st Qu.:  3
##  Median : 8       Median :  66    Median :   0    Median : 17
##  Mean   :10       Mean   : 196    Mean   :   2    Mean   : 22
##  3rd Qu.:16       3rd Qu.: 267    3rd Qu.:  38    3rd Qu.: 36
##  Max.   :43       Max.   :1885    Max.   : 152    Max.   :177
##  NA's   :19216    NA's   :19216   NA's   :19216   NA's   :19216
##   var_yaw_arm     gyros_arm_x      gyros_arm_y      gyros_arm_z
##  Min.   :    0   Min.   :-6.370   Min.   :-3.440   Min.   :-2.33
##  1st Qu.:    7   1st Qu.:-1.330   1st Qu.:-0.800   1st Qu.:-0.07
##  Median :  278   Median : 0.080   Median :-0.240   Median : 0.23
##  Mean   : 1056   Mean   : 0.043   Mean   :-0.257   Mean   : 0.27
##  3rd Qu.: 1295   3rd Qu.: 1.570   3rd Qu.: 0.140   3rd Qu.: 0.72
##  Max.   :31345   Max.   : 4.870   Max.   : 2.840   Max.   : 3.02
##  NA's   :19216
##   accel_arm_x      accel_arm_y      accel_arm_z      magnet_arm_x
##  Min.   :-404.0   Min.   :-318.0   Min.   :-636.0   Min.   :-584
##  1st Qu.:-242.0   1st Qu.: -54.0   1st Qu.:-143.0   1st Qu.:-300
##  Median : -44.0   Median :  14.0   Median : -47.0   Median : 289
##  Mean   : -60.2   Mean   :  32.6   Mean   : -71.2   Mean   : 192
##  3rd Qu.:  84.0   3rd Qu.: 139.0   3rd Qu.:  23.0   3rd Qu.: 637
##  Max.   : 437.0   Max.   : 308.0   Max.   : 292.0   Max.   : 782
##
##   magnet_arm_y   magnet_arm_z  kurtosis_roll_arm kurtosis_picth_arm
##  Min.   :-392   Min.   :-597           :19216            :19216
##  1st Qu.:  -9   1st Qu.: 131   #DIV/0! :   78    #DIV/0! :   80
##  Median : 202   Median : 444   -0.02438:    1    -0.00484:    1
```

```
##  Mean   : 157   Mean   : 306   -0.04190:    1    -0.01311:    1
##  3rd Qu.: 323   3rd Qu.: 545   -0.05051:    1    -0.02967:    1
##  Max.   : 583   Max.   : 694   -0.05695:    1    -0.07394:    1
##                                (Other) :  324    (Other) :  322
##  kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm skewness_yaw_arm
##         :19216           :19216            :19216             :19216
##  #DIV/0! :   11   #DIV/0! :   77    #DIV/0! :   80     #DIV/0! :   11
##  0.55844 :    2   -0.00051:    1    -0.00184:    1     -1.62032:    2
##  0.65132 :    2   -0.00696:    1    -0.01185:    1     0.55053 :    2
##  -0.01548:    1   -0.01884:    1    -0.01247:    1     -0.00311:    1
##  -0.01749:    1   -0.03359:    1    -0.02063:    1     -0.00562:    1
##  (Other) :  389   (Other) :  325    (Other) :  322     (Other) :  389
##   max_roll_arm   max_picth_arm    max_yaw_arm     min_roll_arm
##  Min.   :-73     Min.   :-173    Min.   : 4      Min.   :-89
##  1st Qu.:  0     1st Qu.:  -2    1st Qu.:29      1st Qu.:-42
##  Median :  5     Median :  23    Median :34      Median :-22
##  Mean   : 11     Mean   :  36    Mean   :35      Mean   :-21
##  3rd Qu.: 27     3rd Qu.:  96    3rd Qu.:41      3rd Qu.:  0
##  Max.   : 86     Max.   : 180    Max.   :65      Max.   : 66
##  NA's   :19216   NA's   :19216   NA's   :19216   NA's   :19216
##  min_pitch_arm    min_yaw_arm    amplitude_roll_arm amplitude_pitch_arm
##  Min.   :-180    Min.   : 1      Min.   :  0        Min.   :  0
##  1st Qu.: -73    1st Qu.: 8      1st Qu.:  5        1st Qu.: 10
##  Median : -34    Median :13      Median : 28        Median : 55
##  Mean   : -34    Mean   :15      Mean   : 32        Mean   : 70
##  3rd Qu.:   0    3rd Qu.:19      3rd Qu.: 51        3rd Qu.:115
##  Max.   : 152    Max.   :38      Max.   :120        Max.   :360
##  NA's   :19216   NA's   :19216   NA's   :19216      NA's   :19216
##  amplitude_yaw_arm roll_dumbbell    pitch_dumbbell    yaw_dumbbell
##  Min.   : 0        Min.   :-153.7   Min.   :-149.6   Min.   :-150.87
##  1st Qu.:13        1st Qu.: -18.5   1st Qu.: -40.9   1st Qu.: -77.64
##  Median :22        Median :  48.2   Median : -21.0   Median :  -3.32
##  Mean   :21        Mean   :  23.8   Mean   : -10.8   Mean   :   1.67
##  3rd Qu.:29        3rd Qu.:  67.6   3rd Qu.:  17.5   3rd Qu.:  79.64
##  Max.   :52        Max.   : 153.6   Max.   : 149.4   Max.   : 154.95
##  NA's   :19216
##  kurtosis_roll_dumbbell kurtosis_picth_dumbbell kurtosis_yaw_dumbbell
##         :19216                 :19216                  :19216
##  #DIV/0!:    5          -0.5464:    2           #DIV/0!:  406
##  -0.2583:    2          -0.9334:    2
##  -0.3705:    2          -2.0833:    2
```

```
##  -0.5855:    2          -2.0851:    2
##  -2.0851:    2          -2.0889:    2
##  (Other):  393          (Other):  396
##  skewness_roll_dumbbell skewness_pitch_dumbbell skewness_yaw_dumbbell
##         :19216                 :19216                 :19216
##  #DIV/0!:    4          -0.2328:    2          #DIV/0!:  406
##  -0.9324:    2          -0.3521:    2
##  0.1110 :    2          -0.7036:    2
##  1.0312 :    2          0.1090 :    2
##  -0.0082:    1          1.0326 :    2
##  (Other):  395          (Other):  396
##  max_roll_dumbbell max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell
##  Min.   :-70       Min.   :-113              :19216    Min.   :-150
##  1st Qu.:-27       1st Qu.: -67       -0.6   :   20    1st Qu.: -60
##  Median : 15       Median :  40       0.2    :   19    Median : -44
##  Mean   : 14       Mean   :  33       -0.8   :   18    Mean   : -41
##  3rd Qu.: 51       3rd Qu.: 133       -0.3   :   16    3rd Qu.: -25
##  Max.   :137       Max.   : 155       -0.2   :   15    Max.   :  73
##  NA's   :19216     NA's   :19216      (Other):  318    NA's   :19216
##  min_pitch_dumbbell min_yaw_dumbbell amplitude_roll_dumbbell
##  Min.   :-147              :19216    Min.   :  0
##  1st Qu.: -92       -0.6   :   20    1st Qu.: 15
##  Median : -66       0.2    :   19    Median : 35
##  Mean   : -33       -0.8   :   18    Mean   : 55
##  3rd Qu.:  21       -0.3   :   16    3rd Qu.: 81
##  Max.   : 121       -0.2   :   15    Max.   :256
##  NA's   :19216      (Other):  318    NA's   :19216
##  amplitude_pitch_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
##  Min.   :  0                     :19216          Min.   : 0.0
##  1st Qu.: 17              #DIV/0!:    5          1st Qu.: 4.0
##  Median : 42              0.00   :  401          Median :10.0
##  Mean   : 66                                     Mean   :13.7
##  3rd Qu.:100                                     3rd Qu.:19.0
##  Max.   :274                                     Max.   :58.0
##  NA's   :19216
##  var_accel_dumbbell avg_roll_dumbbell stddev_roll_dumbbell
##  Min.   :  0        Min.   :-129      Min.   :  0
##  1st Qu.:  0        1st Qu.: -12      1st Qu.:  5
##  Median :  1        Median :  48      Median : 12
##  Mean   :  4        Mean   :  24      Mean   : 21
##  3rd Qu.:  3        3rd Qu.:  64      3rd Qu.: 26
```

```
##  Max.   :230        Max.   : 126      Max.   :124
##  NA's   :19216      NA's   :19216     NA's   :19216
##  var_roll_dumbbell avg_pitch_dumbbell stddev_pitch_dumbbell
##  Min.   :    0     Min.   :-71        Min.   : 0
##  1st Qu.:   22     1st Qu.:-42        1st Qu.: 3
##  Median :  149     Median :-20        Median : 8
##  Mean   : 1020     Mean   :-12        Mean   :13
##  3rd Qu.:  695     3rd Qu.: 13        3rd Qu.:19
##  Max.   :15321     Max.   : 94        Max.   :83
##  NA's   :19216     NA's   :19216      NA's   :19216
##  var_pitch_dumbbell avg_yaw_dumbbell stddev_yaw_dumbbell var_yaw_dumbbell
##  Min.   :   0       Min.   :-118     Min.   :  0         Min.   :    0
##  1st Qu.:  12       1st Qu.: -77     1st Qu.:  4         1st Qu.:   15
##  Median :  65       Median :  -5     Median : 10         Median :  105
##  Mean   : 350       Mean   :   0     Mean   : 17         Mean   :  590
##  3rd Qu.: 370       3rd Qu.:  71     3rd Qu.: 25         3rd Qu.:  609
##  Max.   :6836       Max.   : 135     Max.   :107         Max.   :11468
##  NA's   :19216      NA's   :19216    NA's   :19216       NA's   :19216
##  gyros_dumbbell_x  gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x
##  Min.   :-204.00   Min.   :-2.10    Min.   : -2.4    Min.   :-419.0
##  1st Qu.:  -0.03   1st Qu.:-0.14    1st Qu.: -0.3    1st Qu.: -50.0
##  Median :   0.13   Median : 0.03    Median : -0.1    Median :  -8.0
##  Mean   :   0.16   Mean   : 0.05    Mean   : -0.1    Mean   : -28.6
##  3rd Qu.:   0.35   3rd Qu.: 0.21    3rd Qu.:  0.0    3rd Qu.:  11.0
##  Max.   :   2.22   Max.   :52.00    Max.   :317.0    Max.   : 235.0
##
##  accel_dumbbell_y accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y
##  Min.   :-189.0   Min.   :-334.0   Min.   :-643      Min.   :-3600
##  1st Qu.:  -8.0   1st Qu.:-142.0   1st Qu.:-535      1st Qu.:  231
##  Median :  41.5   Median :  -1.0   Median :-479      Median :  311
##  Mean   :  52.6   Mean   : -38.3   Mean   :-328      Mean   :  221
##  3rd Qu.: 111.0   3rd Qu.:  38.0   3rd Qu.:-304      3rd Qu.:  390
##  Max.   : 315.0   Max.   : 318.0   Max.   : 592      Max.   :  633
##
##  magnet_dumbbell_z  roll_forearm     pitch_forearm     yaw_forearm
##  Min.   :-262.0    Min.   :-180.00   Min.   :-72.50   Min.   :-180.0
##  1st Qu.: -45.0    1st Qu.:  -0.74   1st Qu.:  0.00   1st Qu.: -68.6
##  Median :  13.0    Median :  21.70   Median :  9.24   Median :   0.0
##  Mean   :  46.1    Mean   :  33.83   Mean   : 10.71   Mean   :  19.2
##  3rd Qu.:  95.0    3rd Qu.: 140.00   3rd Qu.: 28.40   3rd Qu.: 110.0
##  Max.   : 452.0    Max.   : 180.00   Max.   : 89.80   Max.   : 180.0
```

```
##
##  kurtosis_roll_forearm kurtosis_picth_forearm kurtosis_yaw_forearm
##         :19216                :19216                :19216
##  #DIV/0!:   84         #DIV/0!:   85          #DIV/0!:  406
##  -0.8079:    2         -0.0073:    1
##  -0.9169:    2         -0.0442:    1
##  -0.0227:    1         -0.0489:    1
##  -0.0359:    1         -0.0523:    1
##  (Other):  316         (Other):  317
##  skewness_roll_forearm skewness_pitch_forearm skewness_yaw_forearm
##         :19216                :19216                :19216
##  #DIV/0!:   83         #DIV/0!:   85          #DIV/0!:  406
##  -0.1912:    2         0.0000 :    4
##  -0.4126:    2         -0.6992:    2
##  -0.0004:    1         -0.0113:    1
##  -0.0013:    1         -0.0131:    1
##  (Other):  317         (Other):  313
##  max_roll_forearm max_picth_forearm max_yaw_forearm min_roll_forearm
##  Min.   :-67      Min.   :-151             :19216   Min.   :-72
##  1st Qu.:  0      1st Qu.:   0      #DIV/0!:   84   1st Qu.: -6
##  Median : 27      Median : 113      -1.2   :   32   Median :  0
##  Mean   : 24      Mean   :  81      -1.3   :   31   Mean   :  0
##  3rd Qu.: 46      3rd Qu.: 175      -1.4   :   24   3rd Qu.: 12
##  Max.   : 90      Max.   : 180      -1.5   :   24   Max.   : 62
##  NA's   :19216    NA's   :19216     (Other):  211   NA's   :19216
##  min_pitch_forearm min_yaw_forearm amplitude_roll_forearm
##  Min.   :-180             :19216   Min.   :  0
##  1st Qu.:-175      #DIV/0!:   84   1st Qu.:  1
##  Median : -61      -1.2   :   32   Median : 18
##  Mean   : -58      -1.3   :   31   Mean   : 25
##  3rd Qu.:   0      -1.4   :   24   3rd Qu.: 40
##  Max.   : 167      -1.5   :   24   Max.   :126
##  NA's   :19216     (Other):  211   NA's   :19216
##  amplitude_pitch_forearm amplitude_yaw_forearm total_accel_forearm
##  Min.   :  0                    :19216         Min.   :  0.0
##  1st Qu.:  2             #DIV/0!:   84         1st Qu.: 29.0
##  Median : 84             0.00   :  322         Median : 36.0
##  Mean   :139                                   Mean   : 34.7
##  3rd Qu.:350                                   3rd Qu.: 41.0
##  Max.   :360                                   Max.   :108.0
##  NA's   :19216
```

```
##  var_accel_forearm avg_roll_forearm stddev_roll_forearm var_roll_forearm
##  Min.   :  0       Min.   :-177     Min.   :  0         Min.   :    0
##  1st Qu.:  7       1st Qu.:  -1     1st Qu.:  0         1st Qu.:    0
##  Median : 21       Median :  11     Median :  8         Median :   64
##  Mean   : 34       Mean   :  33     Mean   : 42         Mean   : 5274
##  3rd Qu.: 51       3rd Qu.: 107     3rd Qu.: 85         3rd Qu.: 7289
##  Max.   :173       Max.   : 177     Max.   :179         Max.   :32102
##  NA's   :19216     NA's   :19216    NA's   :19216       NA's   :19216
##  avg_pitch_forearm stddev_pitch_forearm var_pitch_forearm avg_yaw_forearm
##  Min.   :-68       Min.   : 0           Min.   :   0      Min.   :-155
##  1st Qu.:  0       1st Qu.: 0           1st Qu.:   0      1st Qu.: -26
##  Median : 12       Median : 6           Median :  30      Median :   0
##  Mean   : 12       Mean   : 8           Mean   : 140      Mean   :  18
##  3rd Qu.: 28       3rd Qu.:13           3rd Qu.: 166      3rd Qu.:  86
##  Max.   : 72       Max.   :48           Max.   :2280      Max.   : 169
##  NA's   :19216     NA's   :19216        NA's   :19216     NA's   :19216
##  stddev_yaw_forearm var_yaw_forearm gyros_forearm_x   gyros_forearm_y
##  Min.   :  0        Min.   :    0   Min.   :-22.000   Min.   : -7.02
##  1st Qu.:  1        1st Qu.:    0   1st Qu.: -0.220   1st Qu.: -1.46
##  Median : 25        Median :  612   Median :  0.050   Median :  0.03
##  Mean   : 45        Mean   : 4640   Mean   :  0.158   Mean   :  0.08
##  3rd Qu.: 86        3rd Qu.: 7368   3rd Qu.:  0.560   3rd Qu.:  1.62
##  Max.   :198        Max.   :39009   Max.   :  3.970   Max.   :311.00
##  NA's   :19216      NA's   :19216
##  gyros_forearm_z  accel_forearm_x  accel_forearm_y accel_forearm_z
##  Min.   : -8.09   Min.   :-498.0   Min.   :-632    Min.   :-446.0
##  1st Qu.: -0.18   1st Qu.:-178.0   1st Qu.:  57    1st Qu.:-182.0
##  Median :  0.08   Median : -57.0   Median : 201    Median : -39.0
##  Mean   :  0.15   Mean   : -61.7   Mean   : 164    Mean   : -55.3
##  3rd Qu.:  0.49   3rd Qu.:  76.0   3rd Qu.: 312    3rd Qu.:  26.0
##  Max.   :231.00   Max.   : 477.0   Max.   : 923    Max.   : 291.0
##
##  magnet_forearm_x magnet_forearm_y magnet_forearm_z classe
##  Min.   :-1280    Min.   :-896     Min.   :-973     A:5580
##  1st Qu.: -616    1st Qu.:   2     1st Qu.: 191     B:3797
##  Median : -378    Median : 591     Median : 511     C:3422
##  Mean   : -313    Mean   : 380     Mean   : 394     D:3216
##  3rd Qu.:  -73    3rd Qu.: 737     3rd Qu.: 653     E:3607
##  Max.   :  672    Max.   :1480     Max.   :1090
##
```

Data Cleaning

There are many variables with mainly NA and for simplicity will eliminate them. Will check for variables with near zero variance and elimiate them as well.

```
# There are many variables with mainly NA and use the following code to eliminate

pmltrainok<-pmltrain[,!apply(is.na(pmltrain), 2, any)]

#There remain variables with near zero variance and some of which have undefined
#values. Will find these and then also remove them leaving the pmltrainred dataset

nsv <- nearZeroVar(pmltrainok,saveMetrics=TRUE)
good<-rownames(nsv)[nsv$nzv==FALSE]
pmltrainred<-pmltrainok[,names(pmltrainok)%in%good]
```

The first 6 variables are time and subject names and numbers and will also eliminate those with the following:

```
pmltrainokk<-pmltrainred[,-c(1:6)]
```

Create a training and validation (testing) set from the original training set

Will set seed and then create a training and testing set from the original training set to allow some validation.

```
set.seed(3455)
inTrain <- createDataPartition(y=pmltrainokk$classe,
                                    p=0.4, list=FALSE)
training <- pmltrainokk[inTrain,]
testing <- pmltrainokk[-inTrain,]
```

The classe variable is now the 53rd column and will look at the degree of correlations among other variables to get further insight into models to choose:

```
M <- abs(cor(training[,-53]))
diag(M) <- 0
which(M > 0.8,arr.ind=T)

##                   row col
## yaw_belt            3   1
## total_accel_belt    4   1
## accel_belt_y        9   1
## accel_belt_z       10   1
## accel_belt_x        8   2
```

```
## magnet_belt_x     11   2
## roll_belt          1   3
## roll_belt          1   4
## accel_belt_y       9   4
## accel_belt_z      10   4
## pitch_belt         2   8
## magnet_belt_x     11   8
## roll_belt          1   9
## total_accel_belt   4   9
## accel_belt_z      10   9
## roll_belt          1  10
## total_accel_belt   4  10
## accel_belt_y       9  10
## pitch_belt         2  11
## accel_belt_x       8  11
## gyros_arm_y       19  18
## gyros_arm_x       18  19
## magnet_arm_x      24  21
## accel_arm_x       21  24
## magnet_arm_z      26  25
## magnet_arm_y      25  26
## accel_dumbbell_x  34  28
## accel_dumbbell_z  36  29
## pitch_dumbbell    28  34
## yaw_dumbbell      29  36
```

Rationale for approach

There remain as seen above many correlated variables interestingly many correlated with the row number. One approach would be to use a simple model and use PCA to limit variables. Since random forests are good for nonlinear and complex problems I chose to use this without further preprocessing since random forrests are robust. I use cross validation in the call to rf - cv method and allowParallel to be true. I chose these parameters since the boot strap method took too long on my lowly PC.

```
modFit <- train(classe~.,data=training,method="rf",
                trControl=trainControl(method="cv",number=5),
                prox=TRUE,allowParallel=TRUE)

## Loading required package: randomForest
## randomForest 4.6-10
## Type rfNews() to see new features/changes/bug fixes.

#we now look at the fit
modFit
```

```
## Random Forest
##
## 7850 samples
##   52 predictor
##    5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
##
## Summary of sample sizes: 6279, 6280, 6281, 6280, 6280
##
## Resampling results across tuning parameters:
##
##   mtry  Accuracy  Kappa   Accuracy SD  Kappa SD
##    2    0.9786    0.9729  0.002869     0.003625
##   27    0.9803    0.9750  0.003180     0.004031
##   52    0.9771    0.9710  0.003944     0.004993
##
## Accuracy was used to select the optimal model using  the largest value.
## The final value used for the model was mtry = 27.
```

Cross validation

The fit looks excellent and we further cross validate with the testing subset of the original training set. We generate the confusion matrix and this suggests an excellent ability to extrapolate beyond our original training data an accuracy close to 90%.

```
confusionMatrix(testing$classe,predict(modFit,testing))

## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 3338    6    2    0    2
##          B   39 2224   15    0    0
##          C    0   14 2027   12    0
##          D    0    1   34 1892    2
##          E    0    1    7    6 2150
##
## Overall Statistics
##
##                Accuracy : 0.988
##                  95% CI : (0.986, 0.99)
```

```
##     No Information Rate : 0.287
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.985
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity             0.988    0.990    0.972    0.991    0.998
## Specificity             0.999    0.994    0.997    0.996    0.999
## Pos Pred Value          0.997    0.976    0.987    0.981    0.994
## Neg Pred Value          0.995    0.998    0.994    0.998    1.000
## Prevalence              0.287    0.191    0.177    0.162    0.183
## Detection Rate          0.284    0.189    0.172    0.161    0.183
## Detection Prevalence    0.284    0.194    0.174    0.164    0.184
## Balanced Accuracy       0.994    0.992    0.985    0.993    0.998
```

Application to real test data

Finally we read in the actual testing data and use our model to predict the classe of each of the 20.

```
pmltest<-read.csv("pml-testing.csv")
pred <- predict(modFit,pmltest)
pred

##  [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
```

I would love to have tried other models but the computation time on my lowly windows PC has limited me greatly.