Week 4 Project Write Up

- Author: “Emilie Worsham”
- Date: “12/28/2017”

Project Background

This project is the final project in the Practical Machine Learning Course.

Project Background and Objective from the Course Website: 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. These type of devices are part of the quantified self movement - a group of enthusiasts who take measurements about themselves regularly to improve their health, to find patterns in their behavior, or because they are tech geeks. 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. More information is available from the website here: http://web.archive.org/web/20161224072740/http:/groupware.les.inf.puc-rio.br/har (http://web.archive.org/web/20161224072740/http:/groupware.les.inf.puc-rio.br/har) (see the section on the Weight Lifting Exercise Dataset).

Data

The Following Data Sources were used for this project:

- Training Data: https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv (https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv)
- Test Data: https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv (https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv)
- All of the data came from this source, and we thank them for allowing us to use this data for our projects:

http://web.archive.org/web/20161224072740/http:/groupware.les.inf.puc-rio.br/har (http://web.archive.org/web/20161224072740/http:/groupware.les.inf.puc-rio.br/har).

Download and Clean the Training Data

```
## download the training dataset
download.file(url = "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv",
              destfile = "C:/Users/erobe/Desktop/Saved Items/Coursera/Practical_Machine_learning/Practical_Machine_Learning/
pml-training.csv")

## Load training dataset
training <- read.csv("C:/Users/erobe/Desktop/Saved Items/Coursera/Practical_Machine_learning/Practical_Machine_Learning/pml-
training.csv", na.strings=c("NA","#DIV/0!",""))

# Download the testing data
download.file(url = "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv",
              destfile = "C:/Users/erobe/Desktop/Saved Items/Coursera/Practical_Machine_learning/Practical_Machine_Learning/
pml-testing.csv")

# Load the testing dataset
testing <- read.csv("C:/Users/erobe/Desktop/Saved Items/Coursera/Practical_Machine_learning/Practical_Machine_Learning/pml-t
esting.csv", na.strings=c("NA","#DIV/0!",""))
```

Exploring the Data & Cleaning the Data

First I wanted to see which columns are in both data sets. To be able to use both sets in modeling I will want to make sure that both sets have the same columns.

```
head(testing)
```

X user_name raw_timestamp_part_1 raw_timestamp_part_2 cvtd_timestamp 1 1 pedro 1323095002 868349 05/12/2011 14:23 2 2 jeremy 1322673067 778725 30/11/2011 17:11 3 3 jeremy 1322673075 342967 30/11/2011 17:11 4 4 adelmo 1322832789 560311 02/12/2011 13:33 5 5 eurico 1322489635 814776 28/11/2011 14:13 6 6 jeremy 1322673149 510661 30/11/2011 17:12 new_window num_window roll_belt pitch_belt yaw_belt total_accel_belt 1 no 74 123.00 27.00 -4.75 20 2 no 431 1.02 4.87 -88.90 4 3 no 439 0.87 1.82 -88.50 5 4 no 194 125.00 -41.60 162.00 17 5 no 235 1.35 3.33 -88.60 3 6 no 504 -5.92 1.59 -87.70 4 kurtosis_roll_belt kurtosis_picth_belt kurtosis_yaw_belt 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA skewness_roll_belt skewness_roll_belt.1 skewness_yaw_belt max_roll_belt 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA max_picth_belt max_yaw_belt min_roll_belt min_pitch_belt min_yaw_belt 1 NA NA NA NA NA 2 NA NA NA NA NA 3 NA NA NA NA NA 4 NA NA NA NA NA 5 NA NA NA NA NA 6 NA NA NA NA NA amplitude_roll_belt amplitude_pitch_belt amplitude_yaw_belt 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA var_total_accel_belt avg_roll_belt stddev_roll_belt var_roll_belt 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA avg_pitch_belt stddev_pitch_belt var_pitch_belt avg_yaw_belt 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA stddev_yaw_belt var_yaw_belt gyros_belt_x gyros_belt_y gyros_belt_z 1 NA NA -0.50 -0.02 -0.46 2 NA NA

-0.06 -0.02 -0.07 3 NA NA 0.05 0.02 0.03 4 NA NA 0.11 0.11 -0.16 5 NA NA 0.03 0.02 0.00 6 NA NA 0.10 0.05 -0.13 accel_belt_x accel_belt_y accel_belt_z magnet_belt_x magnet_belt_y 1 -38 69 -179 -13 581 2 -13 11 39 43 636 3 1 -1 49 29 631 4 46 45 -156 169 608 5 -8 4 27 33 566 6 -11 -16 38 31 638 magnet_belt_z roll_arm pitch_arm yaw_arm total_accel_arm var_accel_arm 1 -382 40.7 -27.80 178 10 NA 2 -309 0.0 0.00 0 38 NA 3 -312 0.0 0.00 0 44 NA 4 -304 -109.0 55.00 -142 25 NA 5 -418 76.1 2.76 102 29 NA 6 -291 0.0 0.00 0 14 NA avg_roll_arm stddev_roll_arm var_roll_arm avg_pitch_arm stddev_pitch_arm 1 NA NA NA NA NA 2 NA NA NA NA NA 3 NA NA NA NA NA 4 NA NA NA NA NA 5 NA NA NA NA NA 6 NA NA NA NA NA var_pitch_arm avg_yaw_arm stddev_yaw_arm var_yaw_arm gyros_arm_x 1 NA NA NA NA -1.65 2 NA NA NA NA -1.17 3 NA NA NA NA 2.10 4 NA NA NA NA 0.22 5 NA NA NA NA -1.96 6 NA NA NA NA 0.02 gyros_arm_y gyros_arm_z accel_arm_x accel_arm_y accel_arm_z magnet_arm_x 1 0.48 -0.18 16 38 93 -326 2 0.85 -0.43 -290 215 -90 -325 3 -1.36 1.13 -341 245 -87 -264 4 -0.51 0.92 -238 -57 6 -173 5 0.79 -0.54 -197 200 -30 -170 6 0.05 -0.07 -26 130 -19 396 magnet_arm_y magnet_arm_z kurtosis_roll_arm kurtosis_picth_arm 1 385 481 NA NA 2 447 434 NA NA 3 474 413 NA NA 4 257 633 NA NA 5 275 617 NA NA 6 176 516 NA NA kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm skewness_yaw_arm 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA max_roll_arm max_picth_arm max_yaw_arm min_roll_arm min_pitch_arm 1 NA NA NA NA NA 2 NA NA NA NA NA 3 NA NA NA NA NA 4 NA NA NA NA NA 5 NA NA NA NA NA 6 NA NA NA NA NA min_yaw_arm amplitude_roll_arm amplitude_pitch_arm amplitude_yaw_arm 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA roll_dumbbell pitch_dumbbell yaw_dumbbell kurtosis_roll_dumbbell 1 -17.73748 24.96085 126.23596 NA 2 54.47761 -53.69758 -75.51480 NA 3 57.07031 -51.37303 -75.20287 NA 4 43.10927 -30.04885 -103.32003 NA 5 -101.38396 -53.43952 -14.19542 NA 6 62.18750 -50.55595 -71.12063 NA kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_roll_dumbbell 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA skewness_pitch_dumbbell skewness_yaw_dumbbell max_roll_dumbbell 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell min_pitch_dumbbell 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA min_yaw_dumbbell amplitude_roll_dumbbell amplitude_pitch_dumbbell 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA amplitude_yaw_dumbbell total_accel_dumbbell var_accel_dumbbell 1 NA 9 NA 2 NA 31 NA 3 NA 29 NA 4 NA 18 NA 5 NA 4 NA 6 NA 29 NA avg_roll_dumbbell stddev_roll_dumbbell var_roll_dumbbell 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA avg_pitch_dumbbell stddev_pitch_dumbbell var_pitch_dumbbell 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA avg_yaw_dumbbell stddev_yaw_dumbbell var_yaw_dumbbell gyros_dumbbell_x 1 NA NA NA 0.64 2 NA NA NA 0.34 3 NA NA NA 0.39 4 NA NA NA 0.10 5 NA NA NA 0.29 6 NA NA NA -0.59 gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y 1 0.06 -0.61 21 -15 2 0.05 -0.71 -153 155 3 0.14 -0.34 -141 155 4 -0.02 0.05 -51 72 5 -0.47 -0.46 -18 -30 6 0.80 1.10 -138 166 accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z 1 81 523 -528 -56 2 -205 -502 388 -36 3 -196 -506 349 41 4 -148 -576 238 53 5 -5 -424 252 312 6 -186 -543 262 96 roll_forearm pitch_forearm yaw_forearm kurtosis_roll_forearm 1 141 49.30 156.0 NA 2 109 -17.60 106.0 NA 3 131 -32.60 93.0 NA 4 0 0.00 0.0 NA 5 -176 -2.16 -47.9 NA 6 150 1.46 89.7 NA kurtosis_picth_forearm kurtosis_yaw_forearm skewness_roll_forearm 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA skewness_pitch_forearm skewness_yaw_forearm max_roll_forearm 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA max_picth_forearm max_yaw_forearm min_roll_forearm min_pitch_forearm 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA min_yaw_forearm amplitude_roll_forearm amplitude_pitch_forearm 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA amplitude_yaw_forearm total_accel_forearm var_accel_forearm 1 NA 33 NA 2 NA 39 NA 3 NA 34 NA 4 NA 43 NA 5 NA 24 NA 6 NA 43 NA avg_roll_forearm stddev_roll_forearm var_roll_forearm avg_pitch_forearm 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA stddev_pitch_forearm var_pitch_forearm avg_yaw_forearm 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA stddev_yaw_forearm var_yaw_forearm gyros_forearm_x gyros_forearm_y 1 NA NA 0.74 -3.34 2 NA NA 1.12 -2.78 3 NA

NA 0.18 -0.79 4 NA NA 1.38 0.69 5 NA NA -0.75 3.10 6 NA NA -0.88 4.26 gyros_forearm_z accel_forearm_x accel_forearm_y accel_forearm_z 1 -0.59 -110 267 -149 2 -0.18 212 297 -118 3 0.28 154 271 -129 4 1.80 -92 406 -39 5 0.80 131 -93 172 6 1.35 230 322 -144 magnet_forearm_x magnet_forearm_y magnet_forearm_z problem_id 1 -714 419 617 1 2 -237 791 873 2 3 -51 698 783 3 4 -233 783 521 4 5 375 -787 91 5 6 -300 800 884 6

```
head(training)
```

X user_name raw_timestamp_part_1 raw_timestamp_part_2 cvtd_timestamp 1 1 carlitos 1323084231 788290 05/12/2011 11:23 2 2 carlitos 1323084231 808298 05/12/2011 11:23 3 3 carlitos 1323084231 820366 05/12/2011 11:23 4 4 carlitos 1323084232 120339 05/12/2011 11:23 5 5 carlitos 1323084232 196328 05/12/2011 11:23 6 6 carlitos 1323084232 304277 05/12/2011 11:23 new_window num_window roll_belt pitch_belt yaw_belt total_accel_belt 1 no 11 1.41 8.07 -94.4 3 2 no 11 1.41 8.07 -94.4 3 3 no 11 1.42 8.07 -94.4 3 4 no 12 1.48 8.05 -94.4 3 5 no 12 1.48 8.07 -94.4 3 6 no 12 1.45 8.06 -94.4 3 kurtosis_roll_belt kurtosis_picth_belt kurtosis_yaw_belt 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA skewness_roll_belt skewness_roll_belt.1 skewness_yaw_belt max_roll_belt 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA max_picth_belt max_yaw_belt min_roll_belt min_pitch_belt min_yaw_belt 1 NA NA NA NA NA 2 NA NA NA NA NA 3 NA NA NA NA NA 4 NA NA NA NA NA 5 NA NA NA NA NA 6 NA NA NA NA NA amplitude_roll_belt amplitude_pitch_belt amplitude_yaw_belt 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA var_total_accel_belt avg_roll_belt stddev_roll_belt var_roll_belt 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA avg_pitch_belt stddev_pitch_belt var_pitch_belt avg_yaw_belt 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA stddev_yaw_belt var_yaw_belt gyros_belt_x gyros_belt_y gyros_belt_z 1 NA NA 0.00 0.00 -0.02 2 NA NA 0.02 0.00 -0.02 3 NA NA 0.00 0.00 -0.02 4 NA NA 0.02 0.00 -0.03 5 NA NA 0.02 0.02 -0.02 6 NA NA 0.02 0.00 -0.02 accel_belt_x accel_belt_y accel_belt_z magnet_belt_x magnet_belt_y 1 -21 4 22 -3 599 2 -22 4 22 -7 608 3 -20 5 23 -2 600 4 -22 3 21 -6 604 5 -21 2 24 -6 600 6 -21 4 21 0 603 magnet_belt_z roll_arm pitch_arm yaw_arm total_accel_arm var_accel_arm 1 -313 -128 22.5 -161 34 NA 2 -311 -128 22.5 -161 34 NA 3 -305 -128 22.5 -161 34 NA 4 -310 -128 22.1 -161 34 NA 5 -302 -128 22.1 -161 34 NA 6 -312 -128 22.0 -161 34 NA avg_roll_arm stddev_roll_arm var_roll_arm avg_pitch_arm stddev_pitch_arm 1 NA NA NA NA NA 2 NA NA NA NA NA 3 NA NA NA NA NA 4 NA NA NA NA NA 5 NA NA NA NA NA 6 NA NA NA NA NA var_pitch_arm avg_yaw_arm stddev_yaw_arm var_yaw_arm gyros_arm_x 1 NA NA NA NA 0.00 2 NA NA NA NA 0.02 3 NA NA NA NA 0.02 4 NA NA NA NA 0.02 5 NA NA NA NA 0.00 6 NA NA NA NA 0.02 gyros_arm_y gyros_arm_z accel_arm_x accel_arm_y accel_arm_z magnet_arm_x 1 0.00 -0.02 -288 109 -123 -368 2 -0.02 -0.02 -290 110 -125 -369 3 -0.02 -0.02 -289 110 -126 -368 4 -0.03 0.02 -289 111 -123 -372 5 -0.03 0.00 -289 111 -123 -374 6 -0.03 0.00 -289 111 -122 -369 magnet_arm_y magnet_arm_z kurtosis_roll_arm kurtosis_picth_arm 1 337 516 NA NA 2 337 513 NA NA 3 344 513 NA NA 4 344 512 NA NA 5 337 506 NA NA 6 342 513 NA NA kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm skewness_yaw_arm 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA max_roll_arm max_picth_arm max_yaw_arm min_roll_arm min_pitch_arm 1 NA NA NA NA NA 2 NA NA NA NA NA 3 NA NA NA NA NA 4 NA NA NA NA NA 5 NA NA NA NA NA 6 NA NA NA NA NA min_yaw_arm amplitude_roll_arm amplitude_pitch_arm amplitude_yaw_arm 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA roll_dumbbell pitch_dumbbell yaw_dumbbell kurtosis_roll_dumbbell 1 13.05217 -70.49400 -84.87394 NA 2 13.13074 -70.63751 -84.71065 NA 3 12.85075 -70.27812 -85.14078 NA 4 13.43120 -70.39379 -84.87363 NA 5 13.37872 -70.42856 -84.85306 NA 6 13.38246 -70.81759 -84.46500 NA kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_roll_dumbbell 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA skewness_pitch_dumbbell skewness_yaw_dumbbell max_roll_dumbbell 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA

6 NA NA NA max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell min_pitch_dumbbell 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA min_yaw_dumbbell amplitude_roll_dumbbell amplitude_pitch_dumbbell 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA amplitude_yaw_dumbbell total_accel_dumbbell var_accel_dumbbell 1 NA 37 NA 2 NA 37 NA 3 NA 37 NA 4 NA 37 NA 5 NA 37 NA 6 NA 37 NA avg_roll_dumbbell stddev_roll_dumbbell var_roll_dumbbell 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA avg_pitch_dumbbell stddev_pitch_dumbbell var_pitch_dumbbell 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA avg_yaw_dumbbell stddev_yaw_dumbbell var_yaw_dumbbell gyros_dumbbell_x 1 NA NA NA 0 2 NA NA NA 0 3 NA NA NA 0 4 NA NA NA 0 5 NA NA NA 0 6 NA NA NA 0 gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y 1 -0.02 0.00 -234 47 2 -0.02 0.00 -233 47 3 -0.02 0.00 -232 46 4 -0.02 -0.02 -232 48 5 -0.02 0.00 -233 48 6 -0.02 0.00 -234 48 accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z 1 -271 -559 293 -65 2 -269 -555 296 -64 3 -270 -561 298 -63 4 -269 -552 303 -60 5 -270 -554 292 -68 6 -269 -558 294 -66 roll_forearm pitch_forearm yaw_forearm kurtosis_roll_forearm 1 28.4 -63.9 -153 NA 2 28.3 -63.9 -153 NA 3 28.3 -63.9 -152 NA 4 28.1 -63.9 -152 NA 5 28.0 -63.9 -152 NA 6 27.9 -63.9 -152 NA kurtosis_picth_forearm kurtosis_yaw_forearm skewness_roll_forearm 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA skewness_pitch_forearm skewness_yaw_forearm max_roll_forearm 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA max_picth_forearm max_yaw_forearm min_roll_forearm min_pitch_forearm 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA min_yaw_forearm amplitude_roll_forearm amplitude_pitch_forearm 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA amplitude_yaw_forearm total_accel_forearm var_accel_forearm 1 NA 36 NA 2 NA 36 NA 3 NA 36 NA 4 NA 36 NA 5 NA 36 NA 6 NA 36 NA avg_roll_forearm stddev_roll_forearm var_roll_forearm avg_pitch_forearm 1 NA NA NA NA 2 NA NA NA NA 3 NA NA NA NA 4 NA NA NA NA 5 NA NA NA NA 6 NA NA NA NA stddev_pitch_forearm var_pitch_forearm avg_yaw_forearm 1 NA NA NA 2 NA NA NA 3 NA NA NA 4 NA NA NA 5 NA NA NA 6 NA NA NA stddev_yaw_forearm var_yaw_forearm gyros_forearm_x gyros_forearm_y 1 NA NA 0.03 0.00 2 NA NA 0.02 0.00 3 NA NA 0.03 -0.02 4 NA NA 0.02 -0.02 5 NA NA 0.02 0.00 6 NA NA 0.02 -0.02 gyros_forearm_z accel_forearm_x accel_forearm_y accel_forearm_z 1 -0.02 192 203 -215 2 -0.02 192 203 -216 3 0.00 196 204 -213 4 0.00 189 206 -214 5 -0.02 189 206 -214 6 -0.03 193 203 -215 magnet_forearm_x magnet_forearm_y magnet_forearm_z classe 1 -17 654 476 A 2 -18 661 473 A 3 -18 658 469 A 4 -16 658 469 A 5 -17 655 473 A 6 -9 660 478 A

As you can see most of the columns are in both sets. Now I would like to see how many null values are in each data set

```
## This function will find the null values in each column in testing
na_count_testing <-sapply(testing, function(y) sum(length(which(is.na(y)))))

## Show how many nulls are in each column using the new function
na_count_testing <- data.frame(na_count_testing)
na_count_testing
```

```
##                         na_count_testing
## X                                       0
## user_name                               0
## raw_timestamp_part_1                    0
## raw_timestamp_part_2                    0
## cvtd_timestamp                          0
## new_window                              0
## num_window                              0
## roll_belt                               0
## pitch_belt                              0
## yaw_belt                                0
## total_accel_belt                        0
## kurtosis_roll_belt                     20
## kurtosis_picth_belt                    20
## kurtosis_yaw_belt                      20
## skewness_roll_belt                     20
## skewness_roll_belt.1                   20
## skewness_yaw_belt                      20
## max_roll_belt                          20
## max_picth_belt                         20
## max_yaw_belt                           20
## min_roll_belt                          20
## min_pitch_belt                         20
## min_yaw_belt                           20
## amplitude_roll_belt                    20
## amplitude_pitch_belt                   20
## amplitude_yaw_belt                     20
## var_total_accel_belt                   20
## avg_roll_belt                          20
## stddev_roll_belt                       20
## var_roll_belt                          20
## avg_pitch_belt                         20
## stddev_pitch_belt                      20
## var_pitch_belt                         20
## avg_yaw_belt                           20
## stddev_yaw_belt                        20
## var_yaw_belt                           20
## gyros_belt_x                            0
## gyros_belt_y                            0
## gyros_belt_z                            0
```

```
## accel_belt_x                0
## accel_belt_y                0
## accel_belt_z                0
## magnet_belt_x               0
## magnet_belt_y               0
## magnet_belt_z               0
## roll_arm                    0
## pitch_arm                   0
## yaw_arm                     0
## total_accel_arm             0
## var_accel_arm              20
## avg_roll_arm               20
## stddev_roll_arm            20
## var_roll_arm               20
## avg_pitch_arm              20
## stddev_pitch_arm           20
## var_pitch_arm              20
## avg_yaw_arm                20
## stddev_yaw_arm             20
## var_yaw_arm                20
## gyros_arm_x                 0
## gyros_arm_y                 0
## gyros_arm_z                 0
## accel_arm_x                 0
## accel_arm_y                 0
## accel_arm_z                 0
## magnet_arm_x                0
## magnet_arm_y                0
## magnet_arm_z                0
## kurtosis_roll_arm          20
## kurtosis_picth_arm         20
## kurtosis_yaw_arm           20
## skewness_roll_arm          20
## skewness_pitch_arm         20
## skewness_yaw_arm           20
## max_roll_arm               20
## max_picth_arm              20
## max_yaw_arm                20
## min_roll_arm               20
## min_pitch_arm              20
## min_yaw_arm                20
```

```
## amplitude_roll_arm                20
## amplitude_pitch_arm               20
## amplitude_yaw_arm                 20
## roll_dumbbell                      0
## pitch_dumbbell                     0
## yaw_dumbbell                       0
## kurtosis_roll_dumbbell            20
## kurtosis_picth_dumbbell           20
## kurtosis_yaw_dumbbell             20
## skewness_roll_dumbbell            20
## skewness_pitch_dumbbell           20
## skewness_yaw_dumbbell             20
## max_roll_dumbbell                 20
## max_picth_dumbbell                20
## max_yaw_dumbbell                  20
## min_roll_dumbbell                 20
## min_pitch_dumbbell                20
## min_yaw_dumbbell                  20
## amplitude_roll_dumbbell           20
## amplitude_pitch_dumbbell          20
## amplitude_yaw_dumbbell            20
## total_accel_dumbbell               0
## var_accel_dumbbell                20
## avg_roll_dumbbell                 20
## stddev_roll_dumbbell              20
## var_roll_dumbbell                 20
## avg_pitch_dumbbell                20
## stddev_pitch_dumbbell             20
## var_pitch_dumbbell                20
## avg_yaw_dumbbell                  20
## stddev_yaw_dumbbell               20
## var_yaw_dumbbell                  20
## gyros_dumbbell_x                   0
## gyros_dumbbell_y                   0
## gyros_dumbbell_z                   0
## accel_dumbbell_x                   0
## accel_dumbbell_y                   0
## accel_dumbbell_z                   0
## magnet_dumbbell_x                  0
## magnet_dumbbell_y                  0
## magnet_dumbbell_z                  0
```

```
## roll_forearm                0
## pitch_forearm               0
## yaw_forearm                 0
## kurtosis_roll_forearm      20
## kurtosis_picth_forearm     20
## kurtosis_yaw_forearm       20
## skewness_roll_forearm      20
## skewness_pitch_forearm     20
## skewness_yaw_forearm       20
## max_roll_forearm           20
## max_picth_forearm          20
## max_yaw_forearm            20
## min_roll_forearm           20
## min_pitch_forearm          20
## min_yaw_forearm            20
## amplitude_roll_forearm     20
## amplitude_pitch_forearm    20
## amplitude_yaw_forearm      20
## total_accel_forearm         0
## var_accel_forearm          20
## avg_roll_forearm           20
## stddev_roll_forearm        20
## var_roll_forearm           20
## avg_pitch_forearm          20
## stddev_pitch_forearm       20
## var_pitch_forearm          20
## avg_yaw_forearm            20
## stddev_yaw_forearm         20
## var_yaw_forearm            20
## gyros_forearm_x             0
## gyros_forearm_y             0
## gyros_forearm_z             0
## accel_forearm_x             0
## accel_forearm_y             0
## accel_forearm_z             0
## magnet_forearm_x            0
## magnet_forearm_y            0
## magnet_forearm_z            0
## problem_id                  0
```

```
## This function will find the null values in each column in training
na_count_training <-sapply(training, function(y) sum(length(which(is.na(y)))))

## Show how many nulls are in each column using the new function
na_count_training <- data.frame(na_count_training)
na_count_training
```

```
##                         na_count_training
## X                                       0
## user_name                               0
## raw_timestamp_part_1                    0
## raw_timestamp_part_2                    0
## cvtd_timestamp                          0
## new_window                              0
## num_window                              0
## roll_belt                               0
## pitch_belt                              0
## yaw_belt                                0
## total_accel_belt                        0
## kurtosis_roll_belt                  19226
## kurtosis_picth_belt                 19248
## kurtosis_yaw_belt                   19622
## skewness_roll_belt                  19225
## skewness_roll_belt.1                19248
## skewness_yaw_belt                   19622
## max_roll_belt                       19216
## max_picth_belt                      19216
## max_yaw_belt                        19226
## min_roll_belt                       19216
## min_pitch_belt                      19216
## min_yaw_belt                        19226
## amplitude_roll_belt                 19216
## amplitude_pitch_belt                19216
## amplitude_yaw_belt                  19226
## var_total_accel_belt                19216
## avg_roll_belt                       19216
## stddev_roll_belt                    19216
## var_roll_belt                       19216
## avg_pitch_belt                      19216
## stddev_pitch_belt                   19216
## var_pitch_belt                      19216
## avg_yaw_belt                        19216
## stddev_yaw_belt                     19216
## var_yaw_belt                        19216
## gyros_belt_x                            0
## gyros_belt_y                            0
## gyros_belt_z                            0
```

```
## accel_belt_x                   0
## accel_belt_y                   0
## accel_belt_z                   0
## magnet_belt_x                  0
## magnet_belt_y                  0
## magnet_belt_z                  0
## roll_arm                       0
## pitch_arm                      0
## yaw_arm                        0
## total_accel_arm                0
## var_accel_arm              19216
## avg_roll_arm               19216
## stddev_roll_arm            19216
## var_roll_arm               19216
## avg_pitch_arm              19216
## stddev_pitch_arm           19216
## var_pitch_arm              19216
## avg_yaw_arm                19216
## stddev_yaw_arm             19216
## var_yaw_arm                19216
## gyros_arm_x                    0
## gyros_arm_y                    0
## gyros_arm_z                    0
## accel_arm_x                    0
## accel_arm_y                    0
## accel_arm_z                    0
## magnet_arm_x                   0
## magnet_arm_y                   0
## magnet_arm_z                   0
## kurtosis_roll_arm          19294
## kurtosis_picth_arm         19296
## kurtosis_yaw_arm           19227
## skewness_roll_arm          19293
## skewness_pitch_arm         19296
## skewness_yaw_arm           19227
## max_roll_arm               19216
## max_picth_arm              19216
## max_yaw_arm                19216
## min_roll_arm               19216
## min_pitch_arm              19216
## min_yaw_arm                19216
```

```
## amplitude_roll_arm               19216
## amplitude_pitch_arm              19216
## amplitude_yaw_arm                19216
## roll_dumbbell                        0
## pitch_dumbbell                       0
## yaw_dumbbell                         0
## kurtosis_roll_dumbbell           19221
## kurtosis_picth_dumbbell          19218
## kurtosis_yaw_dumbbell            19622
## skewness_roll_dumbbell           19220
## skewness_pitch_dumbbell          19217
## skewness_yaw_dumbbell            19622
## max_roll_dumbbell                19216
## max_picth_dumbbell               19216
## max_yaw_dumbbell                 19221
## min_roll_dumbbell                19216
## min_pitch_dumbbell               19216
## min_yaw_dumbbell                 19221
## amplitude_roll_dumbbell          19216
## amplitude_pitch_dumbbell         19216
## amplitude_yaw_dumbbell           19221
## total_accel_dumbbell                 0
## var_accel_dumbbell               19216
## avg_roll_dumbbell                19216
## stddev_roll_dumbbell             19216
## var_roll_dumbbell                19216
## avg_pitch_dumbbell               19216
## stddev_pitch_dumbbell            19216
## var_pitch_dumbbell               19216
## avg_yaw_dumbbell                 19216
## stddev_yaw_dumbbell              19216
## var_yaw_dumbbell                 19216
## gyros_dumbbell_x                     0
## gyros_dumbbell_y                     0
## gyros_dumbbell_z                     0
## accel_dumbbell_x                     0
## accel_dumbbell_y                     0
## accel_dumbbell_z                     0
## magnet_dumbbell_x                    0
## magnet_dumbbell_y                    0
## magnet_dumbbell_z                    0
```

```
## roll_forearm                    0
## pitch_forearm                   0
## yaw_forearm                     0
## kurtosis_roll_forearm       19300
## kurtosis_picth_forearm      19301
## kurtosis_yaw_forearm        19622
## skewness_roll_forearm       19299
## skewness_pitch_forearm      19301
## skewness_yaw_forearm        19622
## max_roll_forearm            19216
## max_picth_forearm           19216
## max_yaw_forearm             19300
## min_roll_forearm            19216
## min_pitch_forearm           19216
## min_yaw_forearm             19300
## amplitude_roll_forearm      19216
## amplitude_pitch_forearm     19216
## amplitude_yaw_forearm       19300
## total_accel_forearm             0
## var_accel_forearm           19216
## avg_roll_forearm            19216
## stddev_roll_forearm         19216
## var_roll_forearm            19216
## avg_pitch_forearm           19216
## stddev_pitch_forearm        19216
## var_pitch_forearm           19216
## avg_yaw_forearm             19216
## stddev_yaw_forearm          19216
## var_yaw_forearm             19216
## gyros_forearm_x                 0
## gyros_forearm_y                 0
## gyros_forearm_z                 0
## accel_forearm_x                 0
## accel_forearm_y                 0
## accel_forearm_z                 0
## magnet_forearm_x                0
## magnet_forearm_y                0
## magnet_forearm_z                0
## classe                          0
```

As you can see there are quite a few nulls in both datasets so I will want to remove those before creating any of the models.

```
nonull_training <- training[,colSums(is.na(training)) == 0]
nonull_testing  <- testing[,colSums(is.na(testing)) == 0]
head(nonull_training)
```

```
##   X user_name raw_timestamp_part_1 raw_timestamp_part_2   cvtd_timestamp
## 1 1  carlitos           1323084231               788290 05/12/2011 11:23
## 2 2  carlitos           1323084231               808298 05/12/2011 11:23
## 3 3  carlitos           1323084231               820366 05/12/2011 11:23
## 4 4  carlitos           1323084232               120339 05/12/2011 11:23
## 5 5  carlitos           1323084232               196328 05/12/2011 11:23
## 6 6  carlitos           1323084232               304277 05/12/2011 11:23
##   new_window num_window roll_belt pitch_belt yaw_belt total_accel_belt
## 1         no         11      1.41       8.07    -94.4                3
## 2         no         11      1.41       8.07    -94.4                3
## 3         no         11      1.42       8.07    -94.4                3
## 4         no         12      1.48       8.05    -94.4                3
## 5         no         12      1.48       8.07    -94.4                3
## 6         no         12      1.45       8.06    -94.4                3
##   gyros_belt_x gyros_belt_y gyros_belt_z accel_belt_x accel_belt_y
## 1         0.00         0.00        -0.02          -21            4
## 2         0.02         0.00        -0.02          -22            4
## 3         0.00         0.00        -0.02          -20            5
## 4         0.02         0.00        -0.03          -22            3
## 5         0.02         0.02        -0.02          -21            2
## 6         0.02         0.00        -0.02          -21            4
##   accel_belt_z magnet_belt_x magnet_belt_y magnet_belt_z roll_arm
## 1           22            -3           599          -313     -128
## 2           22            -7           608          -311     -128
## 3           23            -2           600          -305     -128
## 4           21            -6           604          -310     -128
## 5           24            -6           600          -302     -128
## 6           21             0           603          -312     -128
##   pitch_arm yaw_arm total_accel_arm gyros_arm_x gyros_arm_y gyros_arm_z
## 1      22.5    -161              34        0.00        0.00       -0.02
## 2      22.5    -161              34        0.02       -0.02       -0.02
## 3      22.5    -161              34        0.02       -0.02       -0.02
## 4      22.1    -161              34        0.02       -0.03        0.02
## 5      22.1    -161              34        0.00       -0.03        0.00
## 6      22.0    -161              34        0.02       -0.03        0.00
##   accel_arm_x accel_arm_y accel_arm_z magnet_arm_x magnet_arm_y
## 1        -288         109        -123         -368          337
## 2        -290         110        -125         -369          337
## 3        -289         110        -126         -368          344
## 4        -289         111        -123         -372          344
```

```
## 5        -289         111        -123         -374          337
## 6        -289         111        -122         -369          342
##   magnet_arm_z roll_dumbbell pitch_dumbbell yaw_dumbbell
## 1          516      13.05217      -70.49400    -84.87394
## 2          513      13.13074      -70.63751    -84.71065
## 3          513      12.85075      -70.27812    -85.14078
## 4          512      13.43120      -70.39379    -84.87363
## 5          506      13.37872      -70.42856    -84.85306
## 6          513      13.38246      -70.81759    -84.46500
##   total_accel_dumbbell gyros_dumbbell_x gyros_dumbbell_y gyros_dumbbell_z
## 1                   37                0            -0.02             0.00
## 2                   37                0            -0.02             0.00
## 3                   37                0            -0.02             0.00
## 4                   37                0            -0.02            -0.02
## 5                   37                0            -0.02             0.00
## 6                   37                0            -0.02             0.00
##   accel_dumbbell_x accel_dumbbell_y accel_dumbbell_z magnet_dumbbell_x
## 1             -234               47             -271              -559
## 2             -233               47             -269              -555
## 3             -232               46             -270              -561
## 4             -232               48             -269              -552
## 5             -233               48             -270              -554
## 6             -234               48             -269              -558
##   magnet_dumbbell_y magnet_dumbbell_z roll_forearm pitch_forearm
## 1               293               -65         28.4         -63.9
## 2               296               -64         28.3         -63.9
## 3               298               -63         28.3         -63.9
## 4               303               -60         28.1         -63.9
## 5               292               -68         28.0         -63.9
## 6               294               -66         27.9         -63.9
##   yaw_forearm total_accel_forearm gyros_forearm_x gyros_forearm_y
## 1        -153                  36            0.03            0.00
## 2        -153                  36            0.02            0.00
## 3        -152                  36            0.03           -0.02
## 4        -152                  36            0.02           -0.02
## 5        -152                  36            0.02            0.00
## 6        -152                  36            0.02           -0.02
##   gyros_forearm_z accel_forearm_x accel_forearm_y accel_forearm_z
## 1           -0.02             192              203            -215
## 2           -0.02             192              203            -216
## 3            0.00             196              204            -213
```

```
## 4            0.00             189             206            -214
## 5           -0.02             189             206            -214
## 6           -0.03             193             203            -215
##   magnet_forearm_x magnet_forearm_y magnet_forearm_z classe
## 1              -17              654              476      A
## 2              -18              661              473      A
## 3              -18              658              469      A
## 4              -16              658              469      A
## 5              -17              655              473      A
## 6               -9              660              478      A
```

```
head(nonull_testing)
```

```
##   X user_name raw_timestamp_part_1 raw_timestamp_part_2   cvtd_timestamp
## 1 1     pedro           1323095002               868349 05/12/2011 14:23
## 2 2    jeremy           1322673067               778725 30/11/2011 17:11
## 3 3    jeremy           1322673075               342967 30/11/2011 17:11
## 4 4    adelmo           1322832789               560311 02/12/2011 13:33
## 5 5    eurico           1322489635               814776 28/11/2011 14:13
## 6 6    jeremy           1322673149               510661 30/11/2011 17:12
##   new_window num_window roll_belt pitch_belt yaw_belt total_accel_belt
## 1         no         74    123.00      27.00    -4.75               20
## 2         no        431      1.02       4.87   -88.90                4
## 3         no        439      0.87       1.82   -88.50                5
## 4         no        194    125.00     -41.60   162.00               17
## 5         no        235      1.35       3.33   -88.60                3
## 6         no        504     -5.92       1.59   -87.70                4
##   gyros_belt_x gyros_belt_y gyros_belt_z accel_belt_x accel_belt_y
## 1        -0.50        -0.02        -0.46          -38           69
## 2        -0.06        -0.02        -0.07          -13           11
## 3         0.05         0.02         0.03            1           -1
## 4         0.11         0.11        -0.16           46           45
## 5         0.03         0.02         0.00           -8            4
## 6         0.10         0.05        -0.13          -11          -16
##   accel_belt_z magnet_belt_x magnet_belt_y magnet_belt_z roll_arm
## 1         -179           -13           581          -382     40.7
## 2           39            43           636          -309      0.0
## 3           49            29           631          -312      0.0
## 4         -156           169           608          -304   -109.0
## 5           27            33           566          -418     76.1
## 6           38            31           638          -291      0.0
##   pitch_arm yaw_arm total_accel_arm gyros_arm_x gyros_arm_y gyros_arm_z
## 1    -27.80     178              10       -1.65        0.48       -0.18
## 2      0.00       0              38       -1.17        0.85       -0.43
## 3      0.00       0              44        2.10       -1.36        1.13
## 4     55.00    -142              25        0.22       -0.51        0.92
## 5      2.76     102              29       -1.96        0.79       -0.54
## 6      0.00       0              14        0.02        0.05       -0.07
##   accel_arm_x accel_arm_y accel_arm_z magnet_arm_x magnet_arm_y
## 1          16          38          93         -326          385
## 2        -290         215         -90         -325          447
## 3        -341         245         -87         -264          474
## 4        -238         -57           6         -173          257
```

```
## 5        -197         200         -30         -170          275
## 6         -26         130         -19          396          176
##   magnet_arm_z roll_dumbbell pitch_dumbbell yaw_dumbbell
## 1          481     -17.73748       24.96085    126.23596
## 2          434      54.47761      -53.69758    -75.51480
## 3          413      57.07031      -51.37303    -75.20287
## 4          633      43.10927      -30.04885   -103.32003
## 5          617    -101.38396      -53.43952    -14.19542
## 6          516      62.18750      -50.55595    -71.12063
##   total_accel_dumbbell gyros_dumbbell_x gyros_dumbbell_y gyros_dumbbell_z
## 1                    9             0.64             0.06            -0.61
## 2                   31             0.34             0.05            -0.71
## 3                   29             0.39             0.14            -0.34
## 4                   18             0.10            -0.02             0.05
## 5                    4             0.29            -0.47            -0.46
## 6                   29            -0.59             0.80             1.10
##   accel_dumbbell_x accel_dumbbell_y accel_dumbbell_z magnet_dumbbell_x
## 1               21              -15               81               523
## 2             -153              155             -205              -502
## 3             -141              155             -196              -506
## 4              -51               72             -148              -576
## 5              -18              -30               -5              -424
## 6             -138              166             -186              -543
##   magnet_dumbbell_y magnet_dumbbell_z roll_forearm pitch_forearm
## 1              -528               -56          141         49.30
## 2               388               -36          109        -17.60
## 3               349                41          131        -32.60
## 4               238                53            0          0.00
## 5               252               312         -176         -2.16
## 6               262                96          150          1.46
##   yaw_forearm total_accel_forearm gyros_forearm_x gyros_forearm_y
## 1       156.0                  33            0.74           -3.34
## 2       106.0                  39            1.12           -2.78
## 3        93.0                  34            0.18           -0.79
## 4         0.0                  43            1.38            0.69
## 5       -47.9                  24           -0.75            3.10
## 6        89.7                  43           -0.88            4.26
##   gyros_forearm_z accel_forearm_x accel_forearm_y accel_forearm_z
## 1           -0.59            -110             267            -149
## 2           -0.18             212             297            -118
## 3            0.28             154             271            -129
```

```
## 4            1.80             -92            406             -39
## 5            0.80            131             -93            172
## 6            1.35            230            322            -144
##   magnet_forearm_x magnet_forearm_y magnet_forearm_z problem_id
## 1             -714              419              617          1
## 2             -237              791              873          2
## 3              -51              698              783          3
## 4             -233              783              521          4
## 5              375             -787               91          5
## 6             -300              800              884          6
```

```
colnames(nonull_training, prefix =  'col')
```

```
##  [1] "X"                    "user_name"            "raw_timestamp_part_1"
##  [4] "raw_timestamp_part_2" "cvtd_timestamp"       "new_window"
##  [7] "num_window"           "roll_belt"            "pitch_belt"
## [10] "yaw_belt"             "total_accel_belt"     "gyros_belt_x"
## [13] "gyros_belt_y"         "gyros_belt_z"         "accel_belt_x"
## [16] "accel_belt_y"         "accel_belt_z"         "magnet_belt_x"
## [19] "magnet_belt_y"        "magnet_belt_z"        "roll_arm"
## [22] "pitch_arm"            "yaw_arm"              "total_accel_arm"
## [25] "gyros_arm_x"          "gyros_arm_y"          "gyros_arm_z"
## [28] "accel_arm_x"          "accel_arm_y"          "accel_arm_z"
## [31] "magnet_arm_x"         "magnet_arm_y"         "magnet_arm_z"
## [34] "roll_dumbbell"        "pitch_dumbbell"       "yaw_dumbbell"
## [37] "total_accel_dumbbell" "gyros_dumbbell_x"     "gyros_dumbbell_y"
## [40] "gyros_dumbbell_z"     "accel_dumbbell_x"     "accel_dumbbell_y"
## [43] "accel_dumbbell_z"     "magnet_dumbbell_x"    "magnet_dumbbell_y"
## [46] "magnet_dumbbell_z"    "roll_forearm"         "pitch_forearm"
## [49] "yaw_forearm"          "total_accel_forearm"  "gyros_forearm_x"
## [52] "gyros_forearm_y"      "gyros_forearm_z"      "accel_forearm_x"
## [55] "accel_forearm_y"      "accel_forearm_z"      "magnet_forearm_x"
## [58] "magnet_forearm_y"     "magnet_forearm_z"     "classe"
```

```
colnames(nonull_testing, prefix =  'col')
```

```
##  [1] "X"                    "user_name"            "raw_timestamp_part_1"
##  [4] "raw_timestamp_part_2" "cvtd_timestamp"       "new_window"
##  [7] "num_window"           "roll_belt"            "pitch_belt"
## [10] "yaw_belt"             "total_accel_belt"     "gyros_belt_x"
## [13] "gyros_belt_y"         "gyros_belt_z"         "accel_belt_x"
## [16] "accel_belt_y"         "accel_belt_z"         "magnet_belt_x"
## [19] "magnet_belt_y"        "magnet_belt_z"        "roll_arm"
## [22] "pitch_arm"            "yaw_arm"              "total_accel_arm"
## [25] "gyros_arm_x"          "gyros_arm_y"          "gyros_arm_z"
## [28] "accel_arm_x"          "accel_arm_y"          "accel_arm_z"
## [31] "magnet_arm_x"         "magnet_arm_y"         "magnet_arm_z"
## [34] "roll_dumbbell"        "pitch_dumbbell"       "yaw_dumbbell"
## [37] "total_accel_dumbbell" "gyros_dumbbell_x"     "gyros_dumbbell_y"
## [40] "gyros_dumbbell_z"     "accel_dumbbell_x"     "accel_dumbbell_y"
## [43] "accel_dumbbell_z"     "magnet_dumbbell_x"    "magnet_dumbbell_y"
## [46] "magnet_dumbbell_z"    "roll_forearm"         "pitch_forearm"
## [49] "yaw_forearm"          "total_accel_forearm"  "gyros_forearm_x"
## [52] "gyros_forearm_y"      "gyros_forearm_z"      "accel_forearm_x"
## [55] "accel_forearm_y"      "accel_forearm_z"      "magnet_forearm_x"
## [58] "magnet_forearm_y"     "magnet_forearm_z"     "problem_id"
```

Now the datasets have the same modeling columns, since we are using the model to predict the “classe” variable the nonull_training dataset is the modeling dataset and the nonull_testing is now the crossvalidation dataset. This step is just to rename them for coding simplicity

```
testdf <- nonull_testing
traindf <- nonull_training
```

Choosing Models

For this project we were given the ablity to choose which modeling techiniques we wanted to use. Since my job includes some modeling I will chose a few that we use most.

- Decision Tree
- Random Forrest

Decision Tree Model

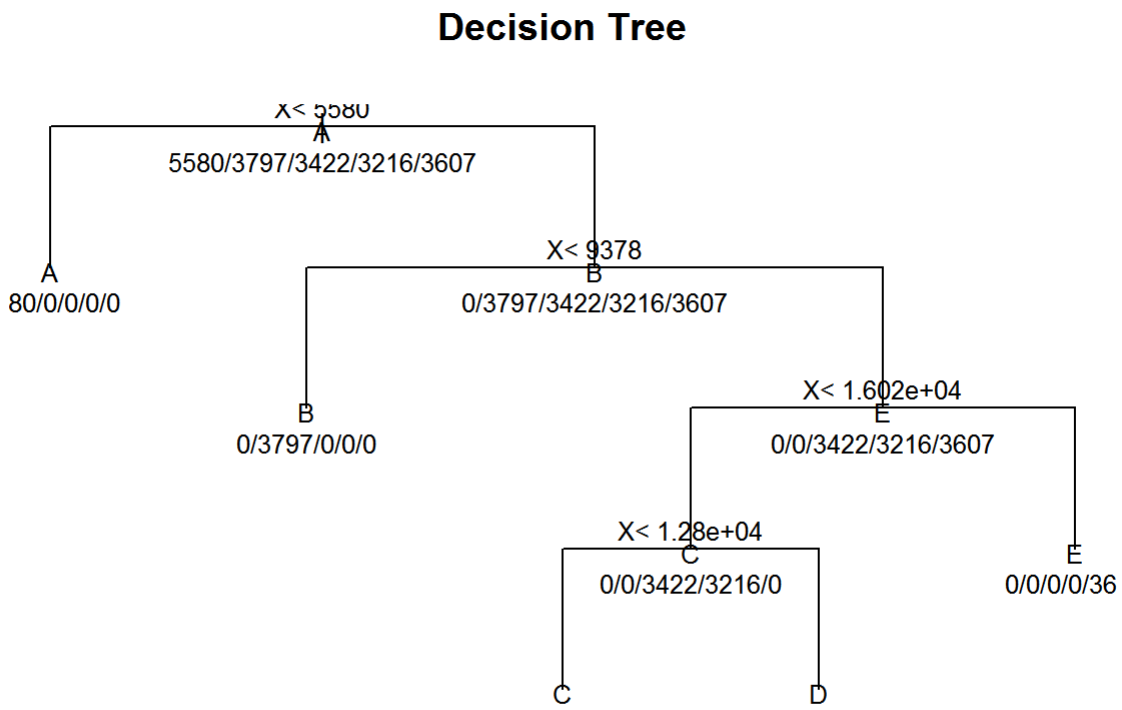
```
library(rpart)
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 3.3.3
```

```
model_tree <- rpart(
                  classe ~ .,
                  data=traindf,
                  method='class')

## view Model
printcp(model_tree)
```

```
##
## Classification tree:
## rpart(formula = classe ~ ., data = traindf, method = "class")
##
## Variables actually used in tree construction:
## [1] X
##
## Root node error: 14042/19622 = 0.71563
##
## n= 19622
##
##         CP nsplit rel error     xerror       xstd
## 1 0.27040      0   1.00000 1.00000000 0.00450019
## 2 0.25687      1   0.72960 0.72966814 0.00498294
## 3 0.24370      2   0.47272 0.47286711 0.00472013
## 4 0.22903      3   0.22903 0.22916963 0.00369375
## 5 0.01000      4   0.00000 0.00021364 0.00012334
```

```
## Print Tree
plot(model_tree, uniform=TRUE,
    main="Decision Tree")
    text(model_tree, use.n=TRUE, all=TRUE, cex=.8)
```




```
    ## View Summary
summary(model_tree)
```

```
## Call:
## rpart(formula = classe ~ ., data = traindf, method = "class")
##   n= 19622
##
##          CP nsplit rel error       xerror         xstd
## 1 0.2704031      0 1.0000000 1.0000000000 0.0045001913
## 2 0.2568722      1 0.7295969 0.7296681384 0.0049829367
## 3 0.2436975      2 0.4727247 0.4728671129 0.0047201330
## 4 0.2290272      3 0.2290272 0.2291696340 0.0036937483
## 5 0.0100000      4 0.0000000 0.0002136448 0.0001233384
##
## Variable importance
##                    X       cvtd_timestamp            roll_belt
##                   41                   20                    4
##        pitch_forearm       pitch_dumbbell raw_timestamp_part_1
##                    4                    4                    4
##        roll_dumbbell    magnet_dumbbell_y         accel_belt_z
##                    3                    3                    3
##     accel_dumbbell_x        magnet_belt_y        magnet_belt_z
##                    3                    3                    3
##              yaw_arm          accel_arm_x           pitch_belt
##                    2                    2                    1
##    magnet_dumbbell_z
##                    1
##
## Node number 1: 19622 observations,    complexity param=0.2704031
##   predicted class=A  expected loss=0.7156253  P(node) =1
##     class counts:  5580  3797  3422  3216  3607
##    probabilities: 0.284 0.194 0.174 0.164 0.184
##   left son=2 (5580 obs) right son=3 (14042 obs)
##   Primary splits:
##       X              < 5580.5     to the left,  improve=4995.2530, (0 missing)
##       cvtd_timestamp splits as  LLLRLLRRLLRLLRLLRLLR, improve=2977.5510, (0 missing)
##       roll_belt      < 130.5      to the left,  improve=1477.9800, (0 missing)
##       pitch_forearm  < -33.95     to the left,  improve=1079.6910, (0 missing)
##       accel_belt_z   < -187.5     to the right, improve= 903.7195, (0 missing)
##   Surrogate splits:
##       cvtd_timestamp       splits as  LLRRLRRRLRRLRRLRRLRR, agree=0.880, adj=0.578, (0 split)
##       pitch_forearm        < -26.65     to the left,  agree=0.797, adj=0.284, (0 split)
##       raw_timestamp_part_1 < 1322490000 to the left,  agree=0.759, adj=0.153, (0 split)
```

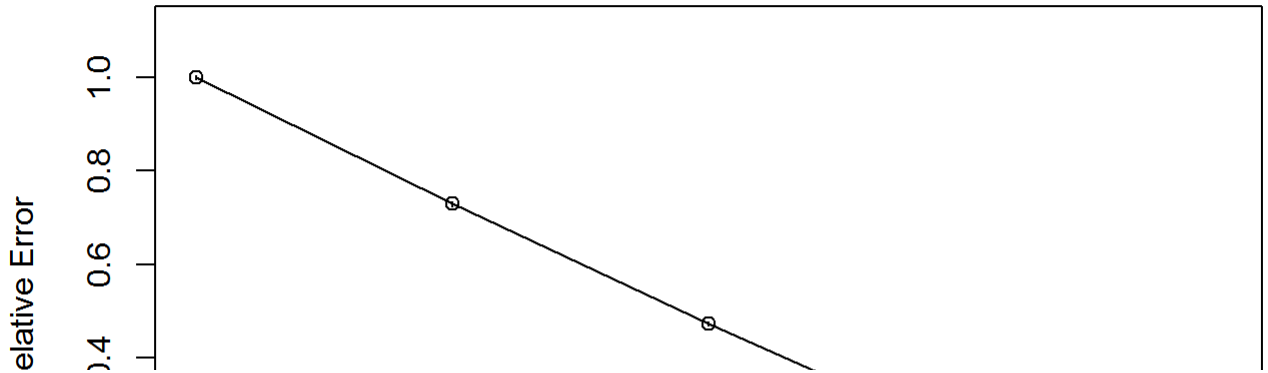
```
##       yaw_arm              < -115.5     to the left,  agree=0.756, adj=0.142, (0 split)
##       accel_arm_x          < -272.5     to the left,  agree=0.755, adj=0.140, (0 split)
##
## Node number 2: 5580 observations
##   predicted class=A  expected loss=0  P(node) =0.2843747
##     class counts:  5580     0     0     0     0
##    probabilities: 1.000 0.000 0.000 0.000 0.000
##
## Node number 3: 14042 observations,    complexity param=0.2568722
##   predicted class=B  expected loss=0.7295969  P(node) =0.7156253
##     class counts:     0  3797  3422  3216  3607
##    probabilities: 0.000 0.270 0.244 0.229 0.257
##   left son=6 (3797 obs) right son=7 (10245 obs)
##   Primary splits:
##       X                < 9377.5     to the left,  improve=3695.7260, (0 missing)
##       cvtd_timestamp   splits as  -LLR-LRR-LR-LR-LR-LR, improve=2428.2970, (0 missing)
##       roll_belt        < 128.5      to the left,  improve=1461.3500, (0 missing)
##       accel_belt_z     < -183.5     to the right, improve= 859.7751, (0 missing)
##       total_accel_belt < 20.5       to the left,  improve= 754.0623, (0 missing)
##   Surrogate splits:
##       cvtd_timestamp       splits as  -LRR-LRR-LR-LR-RR-RR, agree=0.843, adj=0.418, (0 split)
##       raw_timestamp_part_1 < 1322490000 to the left,  agree=0.772, adj=0.155, (0 split)
##       pitch_belt           < -42.85     to the left,  agree=0.769, adj=0.145, (0 split)
##       pitch_dumbbell       < 60.57863   to the right, agree=0.748, adj=0.067, (0 split)
##       magnet_dumbbell_z    < -114.5     to the left,  agree=0.747, adj=0.065, (0 split)
##
## Node number 6: 3797 observations
##   predicted class=B  expected loss=0  P(node) =0.1935073
##     class counts:     0  3797     0     0     0
##    probabilities: 0.000 1.000 0.000 0.000 0.000
##
## Node number 7: 10245 observations,    complexity param=0.2436975
##   predicted class=E  expected loss=0.6479258  P(node) =0.522118
##     class counts:     0     0  3422  3216  3607
##    probabilities: 0.000 0.000 0.334 0.314 0.352
##   left son=14 (6638 obs) right son=15 (3607 obs)
##   Primary splits:
##       X              < 16015.5    to the left,  improve=3506.7280, (0 missing)
##       cvtd_timestamp splits as  --LR--LR-LR-LR-LR-LR, improve=1811.6610, (0 missing)
##       roll_belt      < 128.5      to the left,  improve=1317.0020, (0 missing)
##       accel_belt_z   < -178.5     to the right, improve= 828.9546, (0 missing)
```

```
##       magnet_belt_y  < 578.5      to the right, improve= 687.6283, (0 missing)
##   Surrogate splits:
##       roll_belt      < 128.5      to the left,  agree=0.818, adj=0.482, (0 split)
##       cvtd_timestamp splits as  --LR--LR-LL-LR-LR-LR, agree=0.776, adj=0.362, (0 split)
##       accel_belt_z   < -178.5     to the right, agree=0.762, adj=0.325, (0 split)
##       magnet_belt_y  < 578.5      to the right, agree=0.755, adj=0.305, (0 split)
##       magnet_belt_z  < -379.5     to the right, agree=0.752, adj=0.296, (0 split)
##
## Node number 14: 6638 observations,    complexity param=0.2290272
##   predicted class=C  expected loss=0.4844833  P(node) =0.3382938
##     class counts:     0     0  3422  3216     0
##    probabilities: 0.000 0.000 0.516 0.484 0.000
##   left son=28 (3422 obs) right son=29 (3216 obs)
##   Primary splits:
##       X                 < 12799.5    to the left,  improve=3315.8040, (0 missing)
##       cvtd_timestamp    splits as  --LR--R--LR-LR-RR-LR, improve=1090.0790, (0 missing)
##       roll_dumbbell     < 59.05733   to the left,  improve= 590.0113, (0 missing)
##       magnet_dumbbell_y < 290.5      to the left,  improve= 480.0476, (0 missing)
##       pitch_dumbbell    < -1.223359  to the left,  improve= 431.1945, (0 missing)
##   Surrogate splits:
##       cvtd_timestamp    splits as  --LR--R--LR-LR-LR-LR, agree=0.781, adj=0.549, (0 split)
##       roll_dumbbell     < 57.73165   to the left,  agree=0.702, adj=0.385, (0 split)
##       magnet_dumbbell_y < 290.5      to the left,  agree=0.691, adj=0.362, (0 split)
##       pitch_dumbbell    < -1.223359  to the left,  agree=0.678, adj=0.336, (0 split)
##       accel_dumbbell_x  < -0.5       to the left,  agree=0.678, adj=0.335, (0 split)
##
## Node number 15: 3607 observations
##   predicted class=E  expected loss=0  P(node) =0.1838243
##     class counts:     0     0     0     0  3607
##    probabilities: 0.000 0.000 0.000 0.000 1.000
##
## Node number 28: 3422 observations
##   predicted class=C  expected loss=0  P(node) =0.1743961
##     class counts:     0     0  3422     0     0
##    probabilities: 0.000 0.000 1.000 0.000 0.000
##
## Node number 29: 3216 observations
##   predicted class=D  expected loss=0  P(node) =0.1638977
##     class counts:     0     0     0  3216     0
##    probabilities: 0.000 0.000 0.000 1.000 0.000
```

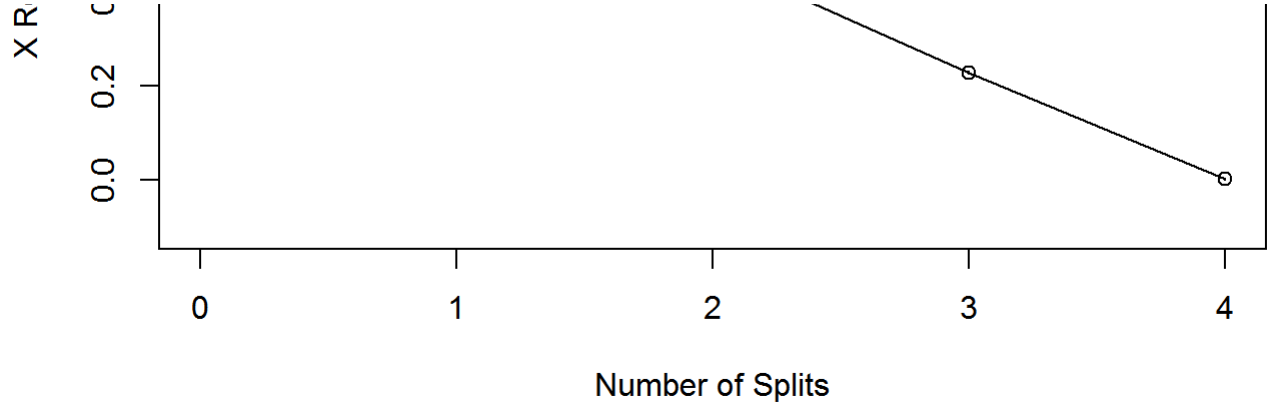
```
## View RSquared

rsq.rpart(model_tree)
```

```
##
## Classification tree:
## rpart(formula = classe ~ ., data = traindf, method = "class")
##
## Variables actually used in tree construction:
## [1] X
##
## Root node error: 14042/19622 = 0.71563
##
## n= 19622
##
##         CP nsplit rel error     xerror       xstd
## 1 0.27040      0   1.00000 1.00000000 0.00450019
## 2 0.25687      1   0.72960 0.72966814 0.00498294
## 3 0.24370      2   0.47272 0.47286711 0.00472013
## 4 0.22903      3   0.22903 0.22916963 0.00369375
## 5 0.01000      4   0.00000 0.00021364 0.00012334
```

```
## Warning in rsq.rpart(model_tree): may not be applicable for this method
```



I'm not really happy with the outcome



of the decision tree, so I'm hoping that the outcome of the Random Forest is better.

Using the Random Forest Model

Now I will set up the Random Forest Model, according to my coworkers, it is the model that they use 90% of the time so my guess before i get started is the one I will end up using to predict the outcome.

```
library(randomForest)
```

```
## randomForest 4.6-12
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
library(caret)
```

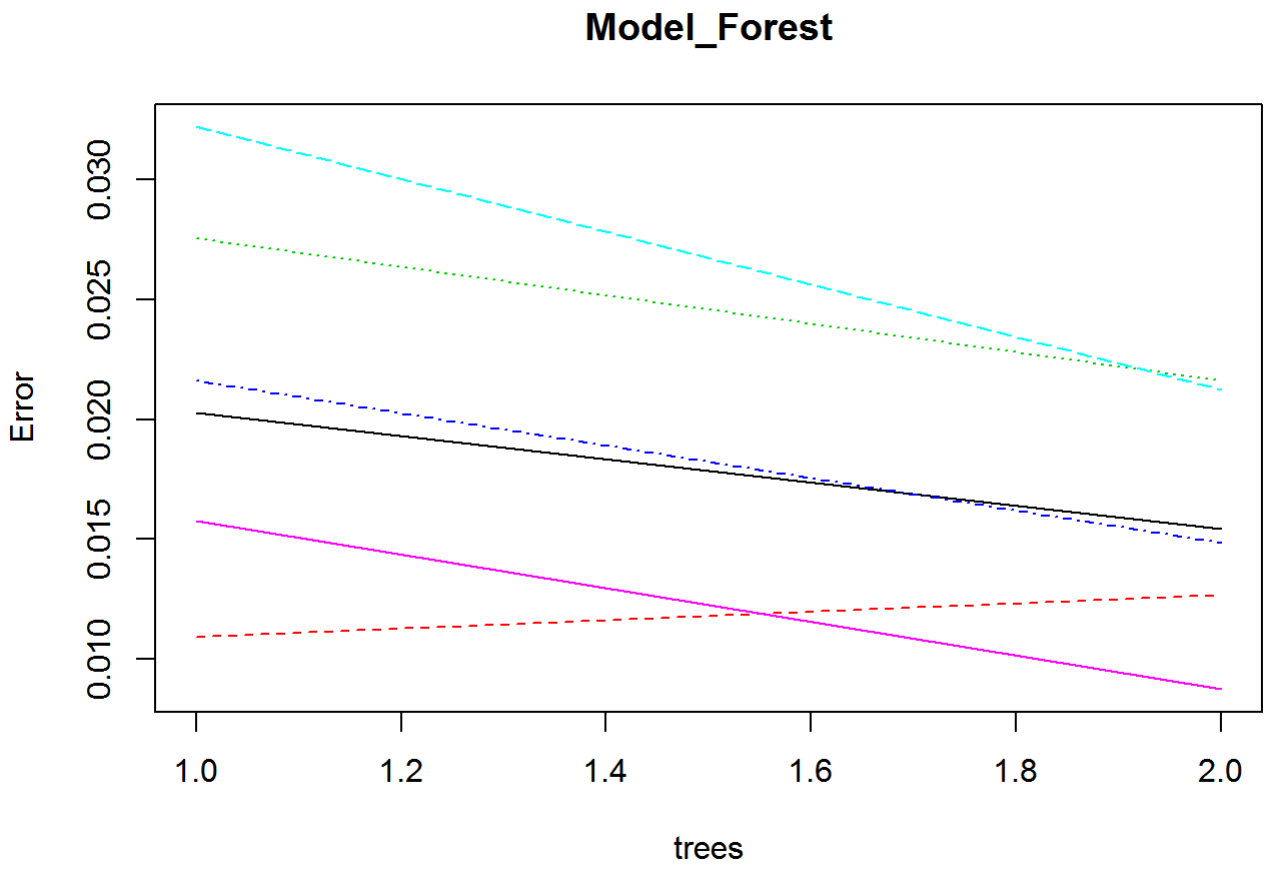
```
## Loading required package: lattice
```

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

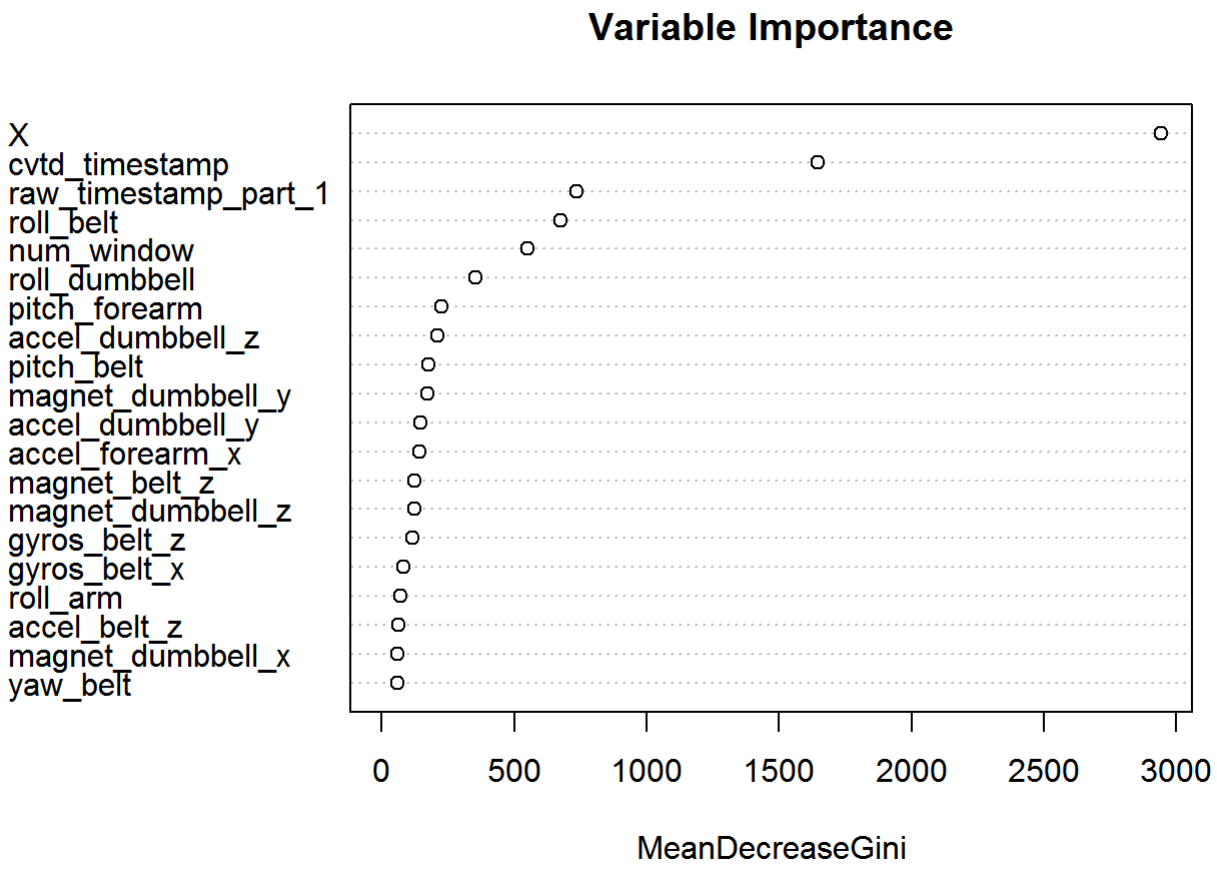
```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:randomForest':
##
##     margin
```

```
## Split traindf into two groups. Model Development and Model Validation.
sample.ind <- sample(2,
                     nrow(traindf),
                     replace = T,
                     prob = c(0.6,0.4))
traindf_dev <- traindf[sample.ind==1,]
traindf_val <- traindf[sample.ind==2,]


## using the dev training data to create the model
Model_Forest <- randomForest(classe ~ ., data=traindf_dev,  method='rf', ntree=2)

## Predicting using the model I just created using the training validation data set.

pred_train_val <- predict(Model_Forest, traindf_val)



plot(Model_Forest)
```




```
## Looking at the importance of each variable in the model.
varImpPlot(Model_Forest,
           sort = T,
           main="Variable Importance",
           n.var=20)
```




```
conf <- confusionMatrix(pred_train_val, traindf_val$classe)

conf
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 2210   17    0    3    2
##          B   13 1537   12    1    0
##          C    0   21 1331    5    0
##          D    0    1    8 1259    9
##          E    5    1    4   13 1451
##
## Overall Statistics
##
##                Accuracy : 0.9854
##                  95% CI : (0.9826, 0.988)
##     No Information Rate : 0.2819
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9816
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9919   0.9746   0.9823   0.9828   0.9925
## Specificity            0.9961   0.9959   0.9960   0.9973   0.9964
## Pos Pred Value         0.9901   0.9834   0.9808   0.9859   0.9844
## Neg Pred Value         0.9968   0.9937   0.9963   0.9967   0.9983
## Prevalence             0.2819   0.1995   0.1715   0.1621   0.1850
## Detection Rate         0.2796   0.1945   0.1684   0.1593   0.1836
## Detection Prevalence   0.2824   0.1978   0.1717   0.1616   0.1865
## Balanced Accuracy      0.9940   0.9853   0.9892   0.9901   0.9945
```

As you can see this model has an accuracy of 97%! That's unheard of in the "real world" so clearly Random Forrest is our choice.

Using the Model to Predict the Outcome

Now I will use the Random Forest model I just built to predict the outcome using the testing data.

```
common <- intersect(names(traindf), names(testdf))
for (p in common) {
  if (class(traindf[[p]]) == "factor") {
    levels(testdf[[p]]) <- levels(traindf[[p]])
  }
}

## Predictions for the course submission quiz.
pred_val <- predict(Model_Forest, testdf[,names(testdf)!="problem_id"])
pred_val
```

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

After putting these values into the quiz you I can see that the random forest model and it's predictions was the one to use.