Machine Learning Course Project

Robert W Allen

March 9, 2016

## Download and Clean Testing and Training Data for Project

Download the Testing and Training Data from the web.

There are fields that have the Excel division by year indicator as a value. These need to be converted to NA.

train\_url <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"  
download.file(train\_url, "pml-training.csv")  
raw\_train\_data <- read.csv("pml-training.csv", na.strings = c("NA", ""))  
  
test\_url <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"  
download.file(test\_url, "pml-testing.csv")  
raw\_test\_data <- read.csv("pml-testing.csv", na.strings = c("NA", ""))

## Descriptive Analysis

Before we can begin the anlaysis we need to understand the variables, especially the one one we hope to predict, classe.

head(raw\_train\_data)

## 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

summary(raw\_train\_data)

## 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  
## #DIV/0! : 32 #DIV/0!: 406 #DIV/0! : 9   
## 47.000000: 4 NA's :19216 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   
## NA's :19216 NA's :19216   
## skewness\_roll\_belt.1 skewness\_yaw\_belt max\_roll\_belt max\_picth\_belt   
## #DIV/0! : 32 #DIV/0!: 406 Min. :-94.300 Min. : 3.00   
## 0.000000 : 4 NA's :19216 1st Qu.:-88.000 1st Qu.: 5.00   
## -2.156553: 3 Median : -5.100 Median :18.00   
## -3.072669: 3 Mean : -6.667 Mean :12.92   
## -6.324555: 3 3rd Qu.: 18.500 3rd Qu.:19.00   
## (Other) : 361 Max. :180.000 Max. :30.00   
## NA's :19216 NA's :19216 NA's :19216   
## max\_yaw\_belt min\_roll\_belt min\_pitch\_belt min\_yaw\_belt   
## -1.1 : 30 Min. :-180.00 Min. : 0.00 -1.1 : 30   
## -1.4 : 29 1st Qu.: -88.40 1st Qu.: 3.00 -1.4 : 29   
## -1.2 : 26 Median : -7.85 Median :16.00 -1.2 : 26   
## -0.9 : 24 Mean : -10.44 Mean :10.76 -0.9 : 24   
## -1.3 : 22 3rd Qu.: 9.05 3rd Qu.:17.00 -1.3 : 22   
## (Other): 275 Max. : 173.00 Max. :23.00 (Other): 275   
## NA's :19216 NA's :19216 NA's :19216 NA's :19216   
## amplitude\_roll\_belt amplitude\_pitch\_belt amplitude\_yaw\_belt  
## Min. : 0.000 Min. : 0.000 #DIV/0!: 10   
## 1st Qu.: 0.300 1st Qu.: 1.000 0.00 : 12   
## Median : 1.000 Median : 1.000 0.0000 : 384   
## Mean : 3.769 Mean : 2.167 NA's :19216   
## 3rd Qu.: 2.083 3rd Qu.: 2.000   
## Max. :360.000 Max. :12.000   
## NA's :19216 NA's :19216   
## var\_total\_accel\_belt avg\_roll\_belt stddev\_roll\_belt var\_roll\_belt   
## Min. : 0.000 Min. :-27.40 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 0.100 1st Qu.: 1.10 1st Qu.: 0.200 1st Qu.: 0.000   
## Median : 0.200 Median :116.35 Median : 0.400 Median : 0.100   
## Mean : 0.926 Mean : 68.06 Mean : 1.337 Mean : 7.699   
## 3rd Qu.: 0.300 3rd Qu.:123.38 3rd Qu.: 0.700 3rd Qu.: 0.500   
## Max. :16.500 Max. :157.40 Max. :14.200 Max. :200.700   
## 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. :-51.400 Min. :0.000 Min. : 0.000 Min. :-138.300   
## 1st Qu.: 2.025 1st Qu.:0.200 1st Qu.: 0.000 1st Qu.: -88.175   
## Median : 5.200 Median :0.400 Median : 0.100 Median : -6.550   
## Mean : 0.520 Mean :0.603 Mean : 0.766 Mean : -8.831   
## 3rd Qu.: 15.775 3rd Qu.:0.700 3rd Qu.: 0.500 3rd Qu.: 14.125   
## Max. : 59.700 Max. :4.000 Max. :16.200 Max. : 173.500   
## NA's :19216 NA's :19216 NA's :19216 NA's :19216   
## stddev\_yaw\_belt var\_yaw\_belt gyros\_belt\_x   
## Min. : 0.000 Min. : 0.000 Min. :-1.040000   
## 1st Qu.: 0.100 1st Qu.: 0.010 1st Qu.:-0.030000   
## Median : 0.300 Median : 0.090 Median : 0.030000   
## Mean : 1.341 Mean : 107.487 Mean :-0.005592   
## 3rd Qu.: 0.700 3rd Qu.: 0.475 3rd Qu.: 0.110000   
## Max. :176.600 Max. :31183.240 Max. : 2.220000   
## NA's :19216 NA's :19216   
## gyros\_belt\_y gyros\_belt\_z accel\_belt\_x accel\_belt\_y   
## Min. :-0.64000 Min. :-1.4600 Min. :-120.000 Min. :-69.00   
## 1st Qu.: 0.00000 1st Qu.:-0.2000 1st Qu.: -21.000 1st Qu.: 3.00   
## Median : 0.02000 Median :-0.1000 Median : -15.000 Median : 35.00   
## Mean : 0.03959 Mean :-0.1305 Mean : -5.595 Mean : 30.15   
## 3rd Qu.: 0.11000 3rd Qu.:-0.0200 3rd Qu.: -5.000 3rd Qu.: 61.00   
## Max. : 0.64000 Max. : 1.6200 Max. : 85.000 Max. :164.00   
##   
## accel\_belt\_z magnet\_belt\_x magnet\_belt\_y magnet\_belt\_z   
## Min. :-275.00 Min. :-52.0 Min. :354.0 Min. :-623.0   
## 1st Qu.:-162.00 1st Qu.: 9.0 1st Qu.:581.0 1st Qu.:-375.0   
## Median :-152.00 Median : 35.0 Median :601.0 Median :-320.0   
## Mean : -72.59 Mean : 55.6 Mean :593.7 Mean :-345.5   
## 3rd Qu.: 27.00 3rd Qu.: 59.0 3rd Qu.:610.0 3rd Qu.:-306.0   
## Max. : 105.00 Max. :485.0 Max. :673.0 Max. : 293.0   
##   
## roll\_arm pitch\_arm yaw\_arm total\_accel\_arm  
## Min. :-180.00 Min. :-88.800 Min. :-180.0000 Min. : 1.00   
## 1st Qu.: -31.77 1st Qu.:-25.900 1st Qu.: -43.1000 1st Qu.:17.00   
## Median : 0.00 Median : 0.000 Median : 0.0000 Median :27.00   
## Mean : 17.83 Mean : -4.612 Mean : -0.6188 Mean :25.51   
## 3rd Qu.: 77.30 3rd Qu.: 11.200 3rd Qu.: 45.8750 3rd Qu.:33.00   
## Max. : 180.00 Max. : 88.500 Max. : 180.0000 Max. :66.00   
##   
## var\_accel\_arm avg\_roll\_arm stddev\_roll\_arm var\_roll\_arm   
## Min. : 0.00 Min. :-166.67 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 9.03 1st Qu.: -38.37 1st Qu.: 1.376 1st Qu.: 1.898   
## Median : 40.61 Median : 0.00 Median : 5.702 Median : 32.517   
## Mean : 53.23 Mean : 12.68 Mean : 11.201 Mean : 417.264   
## 3rd Qu.: 75.62 3rd Qu.: 76.33 3rd Qu.: 14.921 3rd Qu.: 222.647   
## Max. :331.70 Max. : 163.33 Max. :161.964 Max. :26232.208   
## NA's :19216 NA's :19216 NA's :19216 NA's :19216   
## avg\_pitch\_arm stddev\_pitch\_arm var\_pitch\_arm avg\_yaw\_arm   
## Min. :-81.773 Min. : 0.000 Min. : 0.000 Min. :-173.440   
## 1st Qu.:-22.770 1st Qu.: 1.642 1st Qu.: 2.697 1st Qu.: -29.198   
## Median : 0.000 Median : 8.133 Median : 66.146 Median : 0.000   
## Mean : -4.901 Mean :10.383 Mean : 195.864 Mean : 2.359   
## 3rd Qu.: 8.277 3rd Qu.:16.327 3rd Qu.: 266.576 3rd Qu.: 38.185   
## Max. : 75.659 Max. :43.412 Max. :1884.565 Max. : 152.000   
## NA's :19216 NA's :19216 NA's :19216 NA's :19216   
## stddev\_yaw\_arm var\_yaw\_arm gyros\_arm\_x   
## Min. : 0.000 Min. : 0.000 Min. :-6.37000   
## 1st Qu.: 2.577 1st Qu.: 6.642 1st Qu.:-1.33000   
## Median : 16.682 Median : 278.309 Median : 0.08000   
## Mean : 22.270 Mean : 1055.933 Mean : 0.04277   
## 3rd Qu.: 35.984 3rd Qu.: 1294.850 3rd Qu.: 1.57000   
## Max. :177.044 Max. :31344.568 Max. : 4.87000   
## 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  
## Min. : 0.000 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 5.425 1st Qu.: 9.925 1st Qu.:13.00   
## Median : 28.450 Median : 54.900 Median :22.00   
## Mean : 32.452 Mean : 69.677 Mean :20.79   
## 3rd Qu.: 50.960 3rd Qu.:115.175 3rd Qu.:28.75   
## Max. :119.500 Max. :360.000 Max. :52.00   
## NA's :19216 NA's :19216 NA's :19216   
## roll\_dumbbell pitch\_dumbbell yaw\_dumbbell   
## Min. :-153.71 Min. :-149.59 Min. :-150.871   
## 1st Qu.: -18.49 1st Qu.: -40.89 1st Qu.: -77.644   
## Median : 48.17 Median : -20.96 Median : -3.324   
## Mean : 23.84 Mean : -10.78 Mean : 1.674   
## 3rd Qu.: 67.61 3rd Qu.: 17.50 3rd Qu.: 79.643   
## Max. : 153.55 Max. : 149.40 Max. : 154.952   
##   
## kurtosis\_roll\_dumbbell kurtosis\_picth\_dumbbell kurtosis\_yaw\_dumbbell  
## #DIV/0!: 5 -0.5464: 2 #DIV/0!: 406   
## -0.2583: 2 -0.9334: 2 NA's :19216   
## -0.3705: 2 -2.0833: 2   
## -0.5855: 2 -2.0851: 2   
## -2.0851: 2 -2.0889: 2   
## (Other): 393 (Other): 396   
## NA's :19216 NA's :19216   
## skewness\_roll\_dumbbell skewness\_pitch\_dumbbell skewness\_yaw\_dumbbell  
## #DIV/0!: 4 -0.2328: 2 #DIV/0!: 406   
## -0.9324: 2 -0.3521: 2 NA's :19216   
## 0.1110 : 2 -0.7036: 2   
## 1.0312 : 2 0.1090 : 2   
## -0.0082: 1 1.0326 : 2   
## (Other): 395 (Other): 396   
## NA's :19216 NA's :19216   
## max\_roll\_dumbbell max\_picth\_dumbbell max\_yaw\_dumbbell min\_roll\_dumbbell  
## Min. :-70.10 Min. :-112.90 -0.6 : 20 Min. :-149.60   
## 1st Qu.:-27.15 1st Qu.: -66.70 0.2 : 19 1st Qu.: -59.67   
## Median : 14.85 Median : 40.05 -0.8 : 18 Median : -43.55   
## Mean : 13.76 Mean : 32.75 -0.3 : 16 Mean : -41.24   
## 3rd Qu.: 50.58 3rd Qu.: 133.22 -0.2 : 15 3rd Qu.: -25.20   
## Max. :137.00 Max. : 155.00 (Other): 318 Max. : 73.20   
## NA's :19216 NA's :19216 NA's :19216 NA's :19216   
## min\_pitch\_dumbbell min\_yaw\_dumbbell amplitude\_roll\_dumbbell  
## Min. :-147.00 -0.6 : 20 Min. : 0.00   
## 1st Qu.: -91.80 0.2 : 19 1st Qu.: 14.97   
## Median : -66.15 -0.8 : 18 Median : 35.05   
## Mean : -33.18 -0.3 : 16 Mean : 55.00   
## 3rd Qu.: 21.20 -0.2 : 15 3rd Qu.: 81.04   
## Max. : 120.90 (Other): 318 Max. :256.48   
## NA's :19216 NA's :19216 NA's :19216   
## amplitude\_pitch\_dumbbell amplitude\_yaw\_dumbbell total\_accel\_dumbbell  
## Min. : 0.00 #DIV/0!: 5 Min. : 0.00   
## 1st Qu.: 17.06 0.00 : 401 1st Qu.: 4.00   
## Median : 41.73 NA's :19216 Median :10.00   
## Mean : 65.93 Mean :13.72   
## 3rd Qu.: 99.55 3rd Qu.:19.00   
## Max. :273.59 Max. :58.00   
## NA's :19216   
## var\_accel\_dumbbell avg\_roll\_dumbbell stddev\_roll\_dumbbell  
## Min. : 0.000 Min. :-128.96 Min. : 0.000   
## 1st Qu.: 0.378 1st Qu.: -12.33 1st Qu.: 4.639   
## Median : 1.000 Median : 48.23 Median : 12.204   
## Mean : 4.388 Mean : 23.86 Mean : 20.761   
## 3rd Qu.: 3.434 3rd Qu.: 64.37 3rd Qu.: 26.356   
## Max. :230.428 Max. : 125.99 Max. :123.778   
## NA's :19216 NA's :19216 NA's :19216   
## var\_roll\_dumbbell avg\_pitch\_dumbbell stddev\_pitch\_dumbbell  
## Min. : 0.00 Min. :-70.73 Min. : 0.000   
## 1st Qu.: 21.52 1st Qu.:-42.00 1st Qu.: 3.482   
## Median : 148.95 Median :-19.91 Median : 8.089   
## Mean : 1020.27 Mean :-12.33 Mean :13.147   
## 3rd Qu.: 694.65 3rd Qu.: 13.21 3rd Qu.:19.238   
## Max. :15321.01 Max. : 94.28 Max. :82.680   
## NA's :19216 NA's :19216 NA's :19216   
## var\_pitch\_dumbbell avg\_yaw\_dumbbell stddev\_yaw\_dumbbell  
## Min. : 0.00 Min. :-117.950 Min. : 0.000   
## 1st Qu.: 12.12 1st Qu.: -76.696 1st Qu.: 3.885   
## Median : 65.44 Median : -4.505 Median : 10.264   
## Mean : 350.31 Mean : 0.202 Mean : 16.647   
## 3rd Qu.: 370.11 3rd Qu.: 71.234 3rd Qu.: 24.674   
## Max. :6836.02 Max. : 134.905 Max. :107.088   
## NA's :19216 NA's :19216 NA's :19216   
## var\_yaw\_dumbbell gyros\_dumbbell\_x gyros\_dumbbell\_y   
## Min. : 0.00 Min. :-204.0000 Min. :-2.10000   
## 1st Qu.: 15.09 1st Qu.: -0.0300 1st Qu.:-0.14000   
## Median : 105.35 Median : 0.1300 Median : 0.03000   
## Mean : 589.84 Mean : 0.1611 Mean : 0.04606   
## 3rd Qu.: 608.79 3rd Qu.: 0.3500 3rd Qu.: 0.21000   
## Max. :11467.91 Max. : 2.2200 Max. :52.00000   
## NA's :19216   
## gyros\_dumbbell\_z accel\_dumbbell\_x accel\_dumbbell\_y accel\_dumbbell\_z   
## Min. : -2.380 Min. :-419.00 Min. :-189.00 Min. :-334.00   
## 1st Qu.: -0.310 1st Qu.: -50.00 1st Qu.: -8.00 1st Qu.:-142.00   
## Median : -0.130 Median : -8.00 Median : 41.50 Median : -1.00   
## Mean : -0.129 Mean : -28.62 Mean : 52.63 Mean : -38.32   
## 3rd Qu.: 0.030 3rd Qu.: 11.00 3rd Qu.: 111.00 3rd Qu.: 38.00   
## Max. :317.000 Max. : 235.00 Max. : 315.00 Max. : 318.00   
##   
## magnet\_dumbbell\_x magnet\_dumbbell\_y magnet\_dumbbell\_z roll\_forearm   
## Min. :-643.0 Min. :-3600 Min. :-262.00 Min. :-180.0000   
## 1st Qu.:-535.0 1st Qu.: 231 1st Qu.: -45.00 1st Qu.: -0.7375   
## Median :-479.0 Median : 311 Median : 13.00 Median : 21.7000   
## Mean :-328.5 Mean : 221 Mean : 46.05 Mean : 33.8265   
## 3rd Qu.:-304.0 3rd Qu.: 390 3rd Qu.: 95.00 3rd Qu.: 140.0000   
## Max. : 592.0 Max. : 633 Max. : 452.00 Max. : 180.0000   
##   
## pitch\_forearm yaw\_forearm kurtosis\_roll\_forearm  
## Min. :-72.50 Min. :-180.00 #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  
## #DIV/0!: 85 #DIV/0!: 406 Min. :-66.60   
## 0.0000 : 4 NA's :19216 1st Qu.: 0.00   
## -0.6992: 2 Median : 26.80   
## -0.0113: 1 Mean : 24.49   
## -0.0131: 1 3rd Qu.: 45.95   
## (Other): 313 Max. : 89.80   
## NA's :19216 NA's :19216   
## max\_picth\_forearm max\_yaw\_forearm min\_roll\_forearm min\_pitch\_forearm  
## Min. :-151.00 #DIV/0!: 84 Min. :-72.500 Min. :-180.00   
## 1st Qu.: 0.00 -1.2 : 32 1st Qu.: -6.075 1st Qu.:-175.00   
## Median : 113.00 -1.3 : 31 Median : 0.000 Median : -61.00   
## Mean : 81.49 -1.4 : 24 Mean : -0.167 Mean : -57.57   
## 3rd Qu.: 174.75 -1.5 : 24 3rd Qu.: 12.075 3rd Qu.: 0.00   
## Max. : 180.00 (Other): 211 Max. : 62.100 Max. : 167.00   
## NA's :19216 NA's :19216 NA's :19216 NA's :19216   
## min\_yaw\_forearm amplitude\_roll\_forearm amplitude\_pitch\_forearm  
## #DIV/0!: 84 Min. : 0.000 Min. : 0.0   
## -1.2 : 32 1st Qu.: 1.125 1st Qu.: 2.0   
## -1.3 : 31 Median : 17.770 Median : 83.7   
## -1.4 : 24 Mean : 24.653 Mean :139.1   
## -1.5 : 24 3rd Qu.: 39.875 3rd Qu.:350.0   
## (Other): 211 Max. :126.000 Max. :360.0   
## NA's :19216 NA's :19216 NA's :19216   
## amplitude\_yaw\_forearm total\_accel\_forearm var\_accel\_forearm  
## #DIV/0!: 84 Min. : 0.00 Min. : 0.000   
## 0.00 : 322 1st Qu.: 29.00 1st Qu.: 6.759   
## NA's :19216 Median : 36.00 Median : 21.165   
## Mean : 34.72 Mean : 33.502   
## 3rd Qu.: 41.00 3rd Qu.: 51.240   
## Max. :108.00 Max. :172.606   
## NA's :19216   
## avg\_roll\_forearm stddev\_roll\_forearm var\_roll\_forearm   
## Min. :-177.234 Min. : 0.000 Min. : 0.00   
## 1st Qu.: -0.909 1st Qu.: 0.428 1st Qu.: 0.18   
## Median : 11.172 Median : 8.030 Median : 64.48   
## Mean : 33.165 Mean : 41.986 Mean : 5274.10   
## 3rd Qu.: 107.132 3rd Qu.: 85.373 3rd Qu.: 7289.08   
## Max. : 177.256 Max. :179.171 Max. :32102.24   
## NA's :19216 NA's :19216 NA's :19216   
## avg\_pitch\_forearm stddev\_pitch\_forearm var\_pitch\_forearm   
## Min. :-68.17 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.336 1st Qu.: 0.113   
## Median : 12.02 Median : 5.516 Median : 30.425   
## Mean : 11.79 Mean : 7.977 Mean : 139.593   
## 3rd Qu.: 28.48 3rd Qu.:12.866 3rd Qu.: 165.532   
## Max. : 72.09 Max. :47.745 Max. :2279.617   
## NA's :19216 NA's :19216 NA's :19216   
## avg\_yaw\_forearm stddev\_yaw\_forearm var\_yaw\_forearm gyros\_forearm\_x   
## Min. :-155.06 Min. : 0.000 Min. : 0.00 Min. :-22.000   
## 1st Qu.: -26.26 1st Qu.: 0.524 1st Qu.: 0.27 1st Qu.: -0.220   
## Median : 0.00 Median : 24.743 Median : 612.21 Median : 0.050   
## Mean : 18.00 Mean : 44.854 Mean : 4639.85 Mean : 0.158   
## 3rd Qu.: 85.79 3rd Qu.: 85.817 3rd Qu.: 7368.41 3rd Qu.: 0.560   
## Max. : 169.24 Max. :197.508 Max. :39009.33 Max. : 3.970   
## NA's :19216 NA's :19216 NA's :19216   
## gyros\_forearm\_y gyros\_forearm\_z accel\_forearm\_x accel\_forearm\_y   
## Min. : -7.02000 Min. : -8.0900 Min. :-498.00 Min. :-632.0   
## 1st Qu.: -1.46000 1st Qu.: -0.1800 1st Qu.:-178.00 1st Qu.: 57.0   
## Median : 0.03000 Median : 0.0800 Median : -57.00 Median : 201.0   
## Mean : 0.07517 Mean : 0.1512 Mean : -61.65 Mean : 163.7   
## 3rd Qu.: 1.62000 3rd Qu.: 0.4900 3rd Qu.: 76.00 3rd Qu.: 312.0   
## Max. :311.00000 Max. :231.0000 Max. : 477.00 Max. : 923.0   
##   
## accel\_forearm\_z magnet\_forearm\_x magnet\_forearm\_y magnet\_forearm\_z  
## Min. :-446.00 Min. :-1280.0 Min. :-896.0 Min. :-973.0   
## 1st Qu.:-182.00 1st Qu.: -616.0 1st Qu.: 2.0 1st Qu.: 191.0   
## Median : -39.00 Median : -378.0 Median : 591.0 Median : 511.0   
## Mean : -55.29 Mean : -312.6 Mean : 380.1 Mean : 393.6   
## 3rd Qu.: 26.00 3rd Qu.: -73.0 3rd Qu.: 737.0 3rd Qu.: 653.0   
## Max. : 291.00 Max. : 672.0 Max. :1480.0 Max. :1090.0   
##   
## classe   
## A:5580   
## B:3797   
## C:3422   
## D:3216   
## E:3607   
##   
##

summary(raw\_train\_data$classe)

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

There are a lot of columnss, and some look like they have a lot of missing (na) values. We will clean these up now.

dim(raw\_train\_data)

## [1] 19622 160

sum\_na <- sapply(raw\_train\_data, function(x) {sum(is.na(x))})  
rmv\_columns = names(sum\_na[sum\_na>=10000])  
clean\_train\_data = raw\_train\_data[, !names(raw\_train\_data) %in% rmv\_columns]

We will also remove the frist 7 columns that just contain a lot of metadata that we will not neeed.

clean\_train\_data = clean\_train\_data[,-c(1:7)]

Now we can take our training data and partition it into a training set and test for cross validation.

library(caret)

## Loading required package: lattice

## Loading required package: ggplot2

inTrain = createDataPartition(y=clean\_train\_data$classe, p=0.7, list=FALSE)  
train\_data = clean\_train\_data[inTrain,]  
test\_data = clean\_train\_data[-inTrain,]

## Modeling Building

### Model: Random Forest

First we will build some models predicting the classe variable.  
Model 1: rpart (Decision Tree)  
Model 2: GBM (Boosting with Trees)  
Model 3: Random Forest

library(randomForest)

## randomForest 4.6-12

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

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':  
##   
## margin

library(e1071)  
library(rpart)  
library(gbm)

## Loading required package: survival

##   
## Attaching package: 'survival'

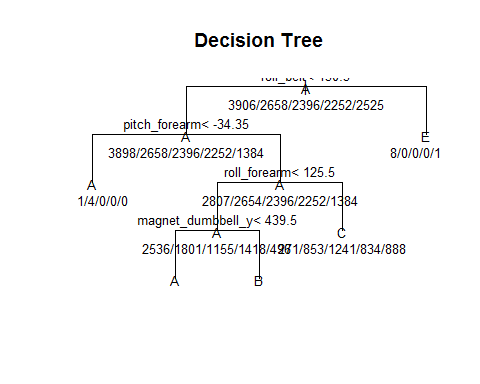
## The following object is masked from 'package:caret':  
##   
## cluster

## Loading required package: splines

## Loading required package: parallel

## Loaded gbm 2.1.1

set.seed(1234)  
  
mod\_rpart <- train(classe~., method = "rpart", data=train\_data)  
plot(mod\_rpart$finalModel, uniform=TRUE, main="Decision Tree")  
text(mod\_rpart$finalModel, use.n=TRUE, all=TRUE, cex=.8)



mod\_gbm <- train(classe~., method = "gbm", data=train\_data, verbose=FALSE)

## Loading required package: plyr

mod\_rf <- train(classe~., method = "rf", data=train\_data)

Then we will use this model to predict on the testing data set and check the results. We will calculate the mean of correct answers. The first mean will be for the rpart model, the second mean will be for the gbm model and the final mean for the Random Forest model.

mean\_rpart <- mean(predict(mod\_rpart, test\_data) == test\_data$classe) \* 100  
mean\_rpart

## [1] 47.79949

mean\_gbm <- mean(predict(mod\_gbm, test\_data) == test\_data$classe) \* 100  
mean\_gbm

## [1] 96.85641

mean\_rf <- mean(predict(mod\_rf, test\_data) == test\_data$classe) \* 100  
mean\_rf

## [1] 99.62617

## Conclusion

This test shows that the Random Forest model is more accurante than the GBM model and RPART model. The random forest model has a testing error (1 - Accuracy of Testing) of 0.3738318, while the GBM model has a testing error of 3.1435854 and the rpart model having an error of 52.2005098.