# Course Project

*ahs*

*December 14, 2014*

Project Goal The goal of this project is to determine how well people perform a particular exercise.

```r
# Load libraries
library(caret)
```

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

```r
library(rattle)
```

```
## Rattle: A free graphical interface for data mining with R.
## Version 3.3.0 Copyright (c) 2006-2014 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(randomForest)
```

```
## randomForest 4.6-10
## Type rfNews() to see new features/changes/bug fixes.
```

```r
# Load the data
train_data <- read.csv('pml-training.csv')
```

1. Load the training data.
2. Do some exploration.
    - PCA to figure out what variables to get rid of.
    - 
3. Decide on a model

- Regression/linear model
- Classification tree
- Random forest

# Exploration

How to deal with NA values

```
# Figure out if any column is all NAs
sapply(train_data, function(x)all(is.na(x)))
```

```
##                        X                user_name     raw_timestamp_part_1
##                    FALSE                    FALSE                    FALSE
##     raw_timestamp_part_2           cvtd_timestamp               new_window
##                    FALSE                    FALSE                    FALSE
##               num_window                roll_belt               pitch_belt
##                    FALSE                    FALSE                    FALSE
##                 yaw_belt         total_accel_belt       kurtosis_roll_belt
##                    FALSE                    FALSE                    FALSE
##      kurtosis_picth_belt        kurtosis_yaw_belt       skewness_roll_belt
##                    FALSE                    FALSE                    FALSE
##     skewness_roll_belt.1        skewness_yaw_belt            max_roll_belt
##                    FALSE                    FALSE                    FALSE
##           max_picth_belt             max_yaw_belt            min_roll_belt
##                    FALSE                    FALSE                    FALSE
##           min_pitch_belt             min_yaw_belt      amplitude_roll_belt
##                    FALSE                    FALSE                    FALSE
##     amplitude_pitch_belt       amplitude_yaw_belt     var_total_accel_belt
##                    FALSE                    FALSE                    FALSE
##            avg_roll_belt         stddev_roll_belt            var_roll_belt
##                    FALSE                    FALSE                    FALSE
##           avg_pitch_belt        stddev_pitch_belt           var_pitch_belt
##                    FALSE                    FALSE                    FALSE
##             avg_yaw_belt          stddev_yaw_belt             var_yaw_belt
##                    FALSE                    FALSE                    FALSE
##             gyros_belt_x             gyros_belt_y             gyros_belt_z
##                    FALSE                    FALSE                    FALSE
##             accel_belt_x             accel_belt_y             accel_belt_z
##                    FALSE                    FALSE                    FALSE
##            magnet_belt_x            magnet_belt_y            magnet_belt_z
##                    FALSE                    FALSE                    FALSE
##                 roll_arm                pitch_arm                  yaw_arm
##                    FALSE                    FALSE                    FALSE
##          total_accel_arm            var_accel_arm             avg_roll_arm
##                    FALSE                    FALSE                    FALSE
##          stddev_roll_arm             var_roll_arm            avg_pitch_arm
##                    FALSE                    FALSE                    FALSE
##         stddev_pitch_arm            var_pitch_arm              avg_yaw_arm
##                    FALSE                    FALSE                    FALSE
##           stddev_yaw_arm              var_yaw_arm              gyros_arm_x
##                    FALSE                    FALSE                    FALSE
##              gyros_arm_y              gyros_arm_z              accel_arm_x
##                    FALSE                    FALSE                    FALSE
##              accel_arm_y              accel_arm_z             magnet_arm_x
##                    FALSE                    FALSE                    FALSE
##             magnet_arm_y             magnet_arm_z        kurtosis_roll_arm
##                    FALSE                    FALSE                    FALSE
##       kurtosis_picth_arm         kurtosis_yaw_arm        skewness_roll_arm
##                    FALSE                    FALSE                    FALSE
##       skewness_pitch_arm         skewness_yaw_arm             max_roll_arm
##                    FALSE                    FALSE                    FALSE
##            max_picth_arm              max_yaw_arm             min_roll_arm
##                    FALSE                    FALSE                    FALSE
##            min_pitch_arm              min_yaw_arm       amplitude_roll_arm
##                    FALSE                    FALSE                    FALSE
##      amplitude_pitch_arm        amplitude_yaw_arm            roll_dumbbell
##                    FALSE                    FALSE                    FALSE
##           pitch_dumbbell             yaw_dumbbell   kurtosis_roll_dumbbell
##                    FALSE                    FALSE                    FALSE
##  kurtosis_picth_dumbbell    kurtosis_yaw_dumbbell   skewness_roll_dumbbell
##                    FALSE                    FALSE                    FALSE
##  skewness_pitch_dumbbell    skewness_yaw_dumbbell        max_roll_dumbbell
##                    FALSE                    FALSE                    FALSE
##       max_picth_dumbbell         max_yaw_dumbbell        min_roll_dumbbell
##                    FALSE                    FALSE                    FALSE
##       min_pitch_dumbbell         min_yaw_dumbbell  amplitude_roll_dumbbell
##                    FALSE                    FALSE                    FALSE
## amplitude_pitch_dumbbell   amplitude_yaw_dumbbell     total_accel_dumbbell
##                    FALSE                    FALSE                    FALSE
##       var_accel_dumbbell        avg_roll_dumbbell     stddev_roll_dumbbell
##                    FALSE                    FALSE                    FALSE
##        var_roll_dumbbell       avg_pitch_dumbbell    stddev_pitch_dumbbell
##                    FALSE                    FALSE                    FALSE
##       var_pitch_dumbbell         avg_yaw_dumbbell      stddev_yaw_dumbbell
##                    FALSE                    FALSE                    FALSE
##         var_yaw_dumbbell         gyros_dumbbell_x         gyros_dumbbell_y
##                    FALSE                    FALSE                    FALSE
##         gyros_dumbbell_z         accel_dumbbell_x         accel_dumbbell_y
##                    FALSE                    FALSE                    FALSE
##         accel_dumbbell_z        magnet_dumbbell_x        magnet_dumbbell_y
##                    FALSE                    FALSE                    FALSE
##        magnet_dumbbell_z             roll_forearm            pitch_forearm
##                    FALSE                    FALSE                    FALSE
##              yaw_forearm    kurtosis_roll_forearm   kurtosis_picth_forearm
##                    FALSE                    FALSE                    FALSE
##     kurtosis_yaw_forearm    skewness_roll_forearm   skewness_pitch_forearm
##                    FALSE                    FALSE                    FALSE
##     skewness_yaw_forearm         max_roll_forearm        max_picth_forearm
##                    FALSE                    FALSE                    FALSE
##          max_yaw_forearm         min_roll_forearm        min_pitch_forearm
##                    FALSE                    FALSE                    FALSE
##          min_yaw_forearm   amplitude_roll_forearm  amplitude_pitch_forearm
##                    FALSE                    FALSE                    FALSE
##    amplitude_yaw_forearm      total_accel_forearm        var_accel_forearm
##                    FALSE                    FALSE                    FALSE
##         avg_roll_forearm      stddev_roll_forearm         var_roll_forearm
##                    FALSE                    FALSE                    FALSE
##        avg_pitch_forearm     stddev_pitch_forearm        var_pitch_forearm
##                    FALSE                    FALSE                    FALSE
##          avg_yaw_forearm       stddev_yaw_forearm          var_yaw_forearm
##                    FALSE                    FALSE                    FALSE
##          gyros_forearm_x          gyros_forearm_y          gyros_forearm_z
##                    FALSE                    FALSE                    FALSE
##          accel_forearm_x          accel_forearm_y          accel_forearm_z
##                    FALSE                    FALSE                    FALSE
##         magnet_forearm_x         magnet_forearm_y         magnet_forearm_z
##                    FALSE                    FALSE                    FALSE
##                   classe
##                    FALSE
```

```
# It's not, so figure out how many columns are MOSTLY NAs.
sapply(train_data,
       function(x)NROW(na.omit(x))/NROW(x))
```

```
##                        X                user_name     raw_timestamp_part_1
##               1.00000000               1.00000000               1.00000000
##     raw_timestamp_part_2           cvtd_timestamp               new_window
##               1.00000000               1.00000000               1.00000000
##               num_window                roll_belt               pitch_belt
##               1.00000000               1.00000000               1.00000000
##                 yaw_belt         total_accel_belt       kurtosis_roll_belt
##               1.00000000               1.00000000               1.00000000
##      kurtosis_picth_belt        kurtosis_yaw_belt       skewness_roll_belt
##               1.00000000               1.00000000               1.00000000
##     skewness_roll_belt.1        skewness_yaw_belt            max_roll_belt
##               1.00000000               1.00000000               0.02069106
##           max_picth_belt             max_yaw_belt            min_roll_belt
##               0.02069106               1.00000000               0.02069106
##           min_pitch_belt             min_yaw_belt      amplitude_roll_belt
##               0.02069106               1.00000000               0.02069106
##     amplitude_pitch_belt       amplitude_yaw_belt     var_total_accel_belt
##               0.02069106               1.00000000               0.02069106
##            avg_roll_belt         stddev_roll_belt            var_roll_belt
##               0.02069106               0.02069106               0.02069106
##           avg_pitch_belt        stddev_pitch_belt           var_pitch_belt
##               0.02069106               0.02069106               0.02069106
##             avg_yaw_belt          stddev_yaw_belt             var_yaw_belt
##               0.02069106               0.02069106               0.02069106
##             gyros_belt_x             gyros_belt_y             gyros_belt_z
##               1.00000000               1.00000000               1.00000000
##             accel_belt_x             accel_belt_y             accel_belt_z
##               1.00000000               1.00000000               1.00000000
##            magnet_belt_x            magnet_belt_y            magnet_belt_z
##               1.00000000               1.00000000               1.00000000
##                 roll_arm                pitch_arm                  yaw_arm
##               1.00000000               1.00000000               1.00000000
##          total_accel_arm            var_accel_arm             avg_roll_arm
##               1.00000000               0.02069106               0.02069106
##          stddev_roll_arm             var_roll_arm            avg_pitch_arm
##               0.02069106               0.02069106               0.02069106
##         stddev_pitch_arm            var_pitch_arm              avg_yaw_arm
##               0.02069106               0.02069106               0.02069106
##           stddev_yaw_arm              var_yaw_arm              gyros_arm_x
##               0.02069106               0.02069106               1.00000000
##              gyros_arm_y              gyros_arm_z              accel_arm_x
##                1.00000000               1.00000000               1.00000000
##               accel_arm_y              accel_arm_z             magnet_arm_x
##                1.00000000               1.00000000               1.00000000
##              magnet_arm_y             magnet_arm_z        kurtosis_roll_arm
##                1.00000000               1.00000000               1.00000000
##        kurtosis_picth_arm         kurtosis_yaw_arm        skewness_roll_arm
##                1.00000000               1.00000000               1.00000000
##        skewness_pitch_arm         skewness_yaw_arm             max_roll_arm
##                1.00000000               1.00000000               0.02069106
##             max_picth_arm              max_yaw_arm             min_roll_arm
##                0.02069106               0.02069106               0.02069106
##             min_pitch_arm              min_yaw_arm       amplitude_roll_arm
##                0.02069106               0.02069106               0.02069106
##       amplitude_pitch_arm        amplitude_yaw_arm            roll_dumbbell
##                0.02069106               0.02069106               1.00000000
##            pitch_dumbbell             yaw_dumbbell   kurtosis_roll_dumbbell
##                1.00000000               1.00000000               1.00000000
##   kurtosis_picth_dumbbell    kurtosis_yaw_dumbbell   skewness_roll_dumbbell
##                1.00000000               1.00000000               1.00000000
##   skewness_pitch_dumbbell    skewness_yaw_dumbbell        max_roll_dumbbell
##                1.00000000               1.00000000               0.02069106
##        max_picth_dumbbell         max_yaw_dumbbell        min_roll_dumbbell
##                0.02069106               1.00000000               0.02069106
##        min_pitch_dumbbell         min_yaw_dumbbell  amplitude_roll_dumbbell
##                0.02069106               1.00000000               0.02069106
## amplitude_pitch_dumbbell   amplitude_yaw_dumbbell     total_accel_dumbbell
##                0.02069106               1.00000000               1.00000000
##        var_accel_dumbbell        avg_roll_dumbbell     stddev_roll_dumbbell
##                0.02069106               0.02069106               0.02069106
##         var_roll_dumbbell       avg_pitch_dumbbell    stddev_pitch_dumbbell
##                0.02069106               0.02069106               0.02069106
##        var_pitch_dumbbell         avg_yaw_dumbbell      stddev_yaw_dumbbell
##                0.02069106               0.02069106               0.02069106
##          var_yaw_dumbbell         gyros_dumbbell_x         gyros_dumbbell_y
##                0.02069106               1.00000000               1.00000000
##          gyros_dumbbell_z         accel_dumbbell_x         accel_dumbbell_y
##                1.00000000               1.00000000               1.00000000
##          accel_dumbbell_z        magnet_dumbbell_x        magnet_dumbbell_y
##                1.00000000               1.00000000               1.00000000
##         magnet_dumbbell_z             roll_forearm            pitch_forearm
##                1.00000000               1.00000000               1.00000000
##               yaw_forearm    kurtosis_roll_forearm   kurtosis_picth_forearm
##                1.00000000               1.00000000               1.00000000
##      kurtosis_yaw_forearm    skewness_roll_forearm   skewness_pitch_forearm
##                1.00000000               1.00000000               1.00000000
##      skewness_yaw_forearm         max_roll_forearm        max_picth_forearm
##                1.00000000               0.02069106               0.02069106
##           max_yaw_forearm         min_roll_forearm        min_pitch_forearm
##                1.00000000               0.02069106               0.02069106
##           min_yaw_forearm   amplitude_roll_forearm  amplitude_pitch_forearm
##                1.00000000               0.02069106               0.02069106
##     amplitude_yaw_forearm      total_accel_forearm        var_accel_forearm
##                1.00000000               1.00000000               0.02069106
##          avg_roll_forearm      stddev_roll_forearm         var_roll_forearm
##               0.02069106               0.02069106               0.02069106
##        avg_pitch_forearm     stddev_pitch_forearm        var_pitch_forearm
##               0.02069106               0.02069106               0.02069106
##          avg_yaw_forearm       stddev_yaw_forearm          var_yaw_forearm
##               0.02069106               0.02069106               0.02069106
##          gyros_forearm_x          gyros_forearm_y          gyros_forearm_z
##               1.00000000               1.00000000               1.00000000
##          accel_forearm_x          accel_forearm_y          accel_forearm_z
##               1.00000000               1.00000000               1.00000000
##         magnet_forearm_x         magnet_forearm_y         magnet_forearm_z
##               1.00000000               1.00000000               1.00000000
##                   classe
##               1.00000000
```

```
# Looking at the results, we see that all columns have either all measures
#   or the same number of non-measures. Let's get rid of the columns
#   that have 2% of values.
sapply(train_data,
       function(x)NROW(na.omit(x)))
```

```
##                        X                user_name     raw_timestamp_part_1
##                    19622                    19622                    19622
##     raw_timestamp_part_2           cvtd_timestamp               new_window
##                    19622                    19622                    19622
##               num_window                roll_belt               pitch_belt
##                    19622                    19622                    19622
##                 yaw_belt         total_accel_belt       kurtosis_roll_belt
##                    19622                    19622                    19622
##      kurtosis_picth_belt        kurtosis_yaw_belt       skewness_roll_belt
##                    19622                    19622                    19622
##     skewness_roll_belt.1        skewness_yaw_belt            max_roll_belt
##                    19622                    19622                      406
##           max_picth_belt             max_yaw_belt            min_roll_belt
##                      406                    19622                      406
##           min_pitch_belt             min_yaw_belt      amplitude_roll_belt
##                      406                    19622                      406
##     amplitude_pitch_belt       amplitude_yaw_belt     var_total_accel_belt
##                      406                    19622                      406
##            avg_roll_belt         stddev_roll_belt            var_roll_belt
##                      406                      406                      406
##           avg_pitch_belt        stddev_pitch_belt           var_pitch_belt
##                      406                      406                      406
##             avg_yaw_belt          stddev_yaw_belt             var_yaw_belt
##                      406                      406                      406
##             gyros_belt_x             gyros_belt_y             gyros_belt_z
##                    19622                    19622                    19622
##             accel_belt_x             accel_belt_y             accel_belt_z
##                    19622                    19622                    19622
##            magnet_belt_x            magnet_belt_y            magnet_belt_z
##                    19622                    19622                    19622
##                 roll_arm                pitch_arm                  yaw_arm
##                    19622                    19622                    19622
##          total_accel_arm            var_accel_arm             avg_roll_arm
##                    19622                      406                      406
##          stddev_roll_arm             var_roll_arm            avg_pitch_arm
##                      406                      406                      406
##         stddev_pitch_arm            var_pitch_arm              avg_yaw_arm
##                      406                      406                      406
##           stddev_yaw_arm              var_yaw_arm              gyros_arm_x
##                      406                      406                    19622
##              gyros_arm_y              gyros_arm_z              accel_arm_x
##                    19622                    19622                    19622
##              accel_arm_y              accel_arm_z             magnet_arm_x
##                    19622                    19622                    19622
##             magnet_arm_y             magnet_arm_z        kurtosis_roll_arm
##                    19622                    19622                    19622
##       kurtosis_picth_arm         kurtosis_yaw_arm        skewness_roll_arm
##                    19622                    19622                    19622
##       skewness_pitch_arm         skewness_yaw_arm             max_roll_arm
##                    19622                    19622                      406
##            max_picth_arm              max_yaw_arm             min_roll_arm
##                      406                      406                      406
##            min_pitch_arm              min_yaw_arm       amplitude_roll_arm
##                      406                      406                      406
##      amplitude_pitch_arm        amplitude_yaw_arm            roll_dumbbell
##                      406                      406                    19622
##           pitch_dumbbell             yaw_dumbbell   kurtosis_roll_dumbbell
##                    19622                    19622                    19622
##  kurtosis_picth_dumbbell    kurtosis_yaw_dumbbell   skewness_roll_dumbbell
##                    19622                    19622                    19622
##  skewness_pitch_dumbbell    skewness_yaw_dumbbell        max_roll_dumbbell
##                    19622                    19622                      406
##       max_picth_dumbbell         max_yaw_dumbbell        min_roll_dumbbell
##                      406                    19622                      406
##       min_pitch_dumbbell         min_yaw_dumbbell  amplitude_roll_dumbbell
##                      406                    19622                      406
## amplitude_pitch_dumbbell   amplitude_yaw_dumbbell     total_accel_dumbbell
##                      406                    19622                    19622
##       var_accel_dumbbell        avg_roll_dumbbell     stddev_roll_dumbbell
##                      406                      406                      406
##        var_roll_dumbbell       avg_pitch_dumbbell    stddev_pitch_dumbbell
##                      406                      406                      406
##       var_pitch_dumbbell         avg_yaw_dumbbell      stddev_yaw_dumbbell
##                      406                      406                      406
##         var_yaw_dumbbell         gyros_dumbbell_x         gyros_dumbbell_y
##                      406                    19622                    19622
##         gyros_dumbbell_z         accel_dumbbell_x         accel_dumbbell_y
##                    19622                    19622                    19622
##         accel_dumbbell_z        magnet_dumbbell_x        magnet_dumbbell_y
##                    19622                    19622                    19622
##        magnet_dumbbell_z             roll_forearm            pitch_forearm
##                    19622                    19622                    19622
##              yaw_forearm    kurtosis_roll_forearm   kurtosis_picth_forearm
##                    19622                    19622                    19622
##     kurtosis_yaw_forearm    skewness_roll_forearm   skewness_pitch_forearm
##                    19622                    19622                    19622
##     skewness_yaw_forearm         max_roll_forearm        max_picth_forearm
##                    19622                      406                      406
##          max_yaw_forearm         min_roll_forearm        min_pitch_forearm
##                    19622                      406                      406
##          min_yaw_forearm   amplitude_roll_forearm  amplitude_pitch_forearm
##                    19622                      406                      406
##    amplitude_yaw_forearm      total_accel_forearm        var_accel_forearm
##                    19622                    19622                      406
##         avg_roll_forearm      stddev_roll_forearm         var_roll_forearm
##                      406                      406                      406
##        avg_pitch_forearm     stddev_pitch_forearm        var_pitch_forearm
##                      406                      406                      406
##          avg_yaw_forearm       stddev_yaw_forearm          var_yaw_forearm
##                      406                      406                      406
##          gyros_forearm_x          gyros_forearm_y          gyros_forearm_z
##                    19622                    19622                    19622
##          accel_forearm_x          accel_forearm_y          accel_forearm_z
##                    19622                    19622                    19622
##         magnet_forearm_x         magnet_forearm_y         magnet_forearm_z
##                    19622                    19622                    19622
##                   classe
##                    19622
```

```
# Turn it into a logical expression so we can use it for filtering
keep_cols <- sapply(train_data,
       function(x)NROW(na.omit(x))!=406)

train_data <- train_data[keep_cols]

# We still have a bunch of columns with empty values, because
# it sees "" as a factor level. I think these also have "#DIV/0" as a factor
# level. Let's see if we can figure out which columns can be removed based
# on this.

# We have a couple columns we need to fix.
train_data$cvtd_timestamp <- as.character(train_data$cvtd_timestamp)
# I'm not quite sure if this is how i want to handle this, but it's what it is for now.
train_data$new_window <- as.character(train_data$new_window)

keep_cols_2 <- sapply(train_data, function(x)!any(levels(x)=='#DIV/0!'))
train_data <- train_data[keep_cols_2]
```

Now we have 60 columns of data.

```
train_data <- train_data[2:60]

# modFit <- train(classe ~ .,method="rpart", data=train_data[1:2000,])
# print(modFit$finalModel)
# plot(modFit$finalModel, uniform=TRUE, main="Classification Tree")
# text(modFit$finalModel, use.n=TRUE, all=TRUE, cex=.8)
# fancyRpartPlot(modFit$finalModel)
```

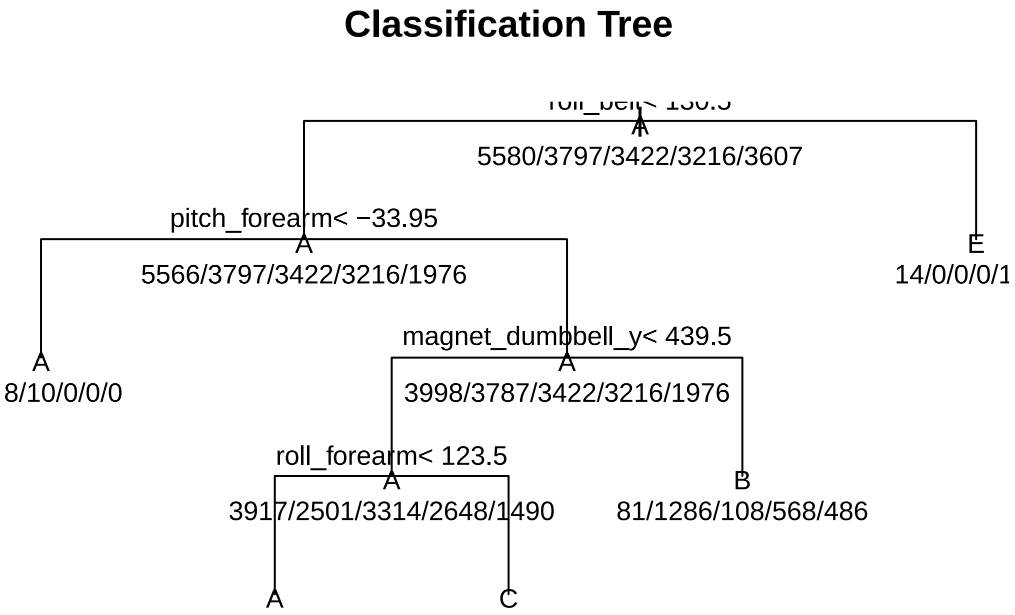
```
train_data$win_ind <- ave(train_data$new_window=="yes", FUN = function(x) {
            cumsum(x) })
```

```r
# foo <- data.frame(diff(train_data$raw_timestamp_part_2))
# foo<- rbind(0,foo)
# colnames(foo) <- c("t2_delta")
# foo[foo$some_diff < 0,] <- 0

# Another way to do this, but it creates a group, which I may or may not want.
train_data <- group_by(train_data, num_window) %>% mutate(t2_delta = c(0,diff(raw_timestamp_part_2)))
```

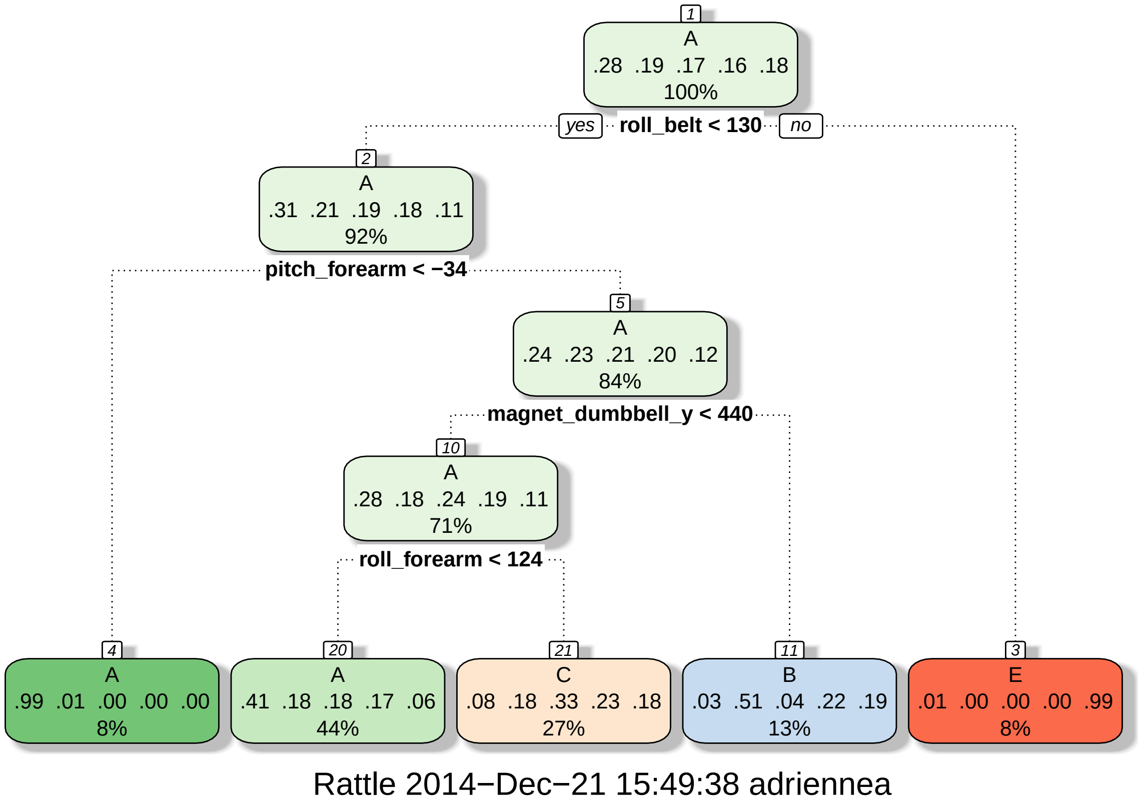
```r
# Okay, building a ctree again.
train_data <- train_data[6:61]

#modFit <- train(classe ~ .,method="rpart", data=train_data[,-55]) # excluding "win_ind"
#save(modFit, file="myClassTree.RData")

# Instead of training again, just load the saved one from file.
load("myClassTree.RData")

print(modFit$finalModel)
```

```
## n= 19622
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##  1) root 19622 14042 A (0.28 0.19 0.17 0.16 0.18)
##    2) roll_belt< 130.5 17977 12411 A (0.31 0.21 0.19 0.18 0.11)
##      4) pitch_forearm< -33.95 1578    10 A (0.99 0.0063 0 0 0) *
##      5) pitch_forearm>=-33.95 16399 12401 A (0.24 0.23 0.21 0.2 0.12)
##       10) magnet_dumbbell_y< 439.5 13870  9953 A (0.28 0.18 0.24 0.19 0.11)
##         20) roll_forearm< 123.5 8643  5131 A (0.41 0.18 0.18 0.17 0.061) *
##         21) roll_forearm>=123.5 5227  3500 C (0.077 0.18 0.33 0.23 0.18) *
##       11) magnet_dumbbell_y>=439.5 2529  1243 B (0.032 0.51 0.043 0.22 0.19) *
##    3) roll_belt>=130.5 1645    14 E (0.0085 0 0 0 0.99) *
```

```r
plot(modFit$finalModel, uniform=TRUE, main="Classification Tree")
text(modFit$finalModel, use.n=TRUE, all=TRUE, cex=.8)
```



```
fancyRpartPlot(modFit$finalModel)
```



Enough of this, let's generate a random forest.

```
in_small_train <- createDataPartition(y=train_data$classe,p=0.2, list=FALSE)
small_td <- train_data[in_small_train,]

#modFit_rf <- train(classe~.,data=small_td,method="rf",prox=TRUE)
```

```r
#save(modFit_rf, file="myRandomForest.RData")

load(file="myRandomForest.RData")
```

Evaluating the result on the data we trained with:

```r
getTree(modFit_rf$finalModel,k=2)
```

```
##    left daughter right daughter split var split point status prediction
## 1              2              3        54   108.00000      1          0
## 2              0              0         0     0.00000     -1          1
## 3              4              5        54   188.50000      1          0
## 4              6              7        42   -43.40000      1          0
## 5              8              9        54   326.50000      1          0
## 6              0              0         0     0.00000     -1          1
## 7             10             11         4   -93.40000      1          0
## 8             12             13        54   257.50000      1          0
## 9             14             15         1   848.00000      1          0
## 10            16             17         8    -0.01000      1          0
## 11            18             19        34    -1.11500      1          0
## 12             0              0         0     0.00000     -1          3
## 13            20             21        15   141.50000      1          0
## 14             0              0         0     0.00000     -1          5
## 15            22             23        29    33.54095      1          0
## 16             0              0         0     0.00000     -1          2
## 17             0              0         0     0.00000     -1          1
## 18            24             25         3    15.04000      1          0
## 19             0              0         0     0.00000     -1          2
## 20            26             27        42    -0.82000      1          0
## 21             0              0         0     0.00000     -1          3
## 22             0              0         0     0.00000     -1          5
## 23             0              0         0     0.00000     -1          4
## 24             0              0         0     0.00000     -1          2
## 25             0              0         0     0.00000     -1          3
## 26            28             29        49   273.00000      1          0
## 27            30             31        54   258.50000      1          0
## 28             0              0         0     0.00000     -1          3
## 29             0              0         0     0.00000     -1          4
## 30            32             33        39   275.50000      1          0
## 31             0              0         0     0.00000     -1          4
## 32             0              0         0     0.00000     -1          3
## 33             0              0         0     0.00000     -1          4
```

```r
pred_train <- predict(modFit_rf, small_td)
small_td$predRight <- pred_train==small_td$classe
table(pred_train, small_td$classe)
```

```
##
## pred_train    A    B    C    D    E
##          A 1115    0    0    0    0
##          B    1  760    0    0    0
##          C    0    0  685    0    0
##          D    0    0    0  644    0
##          E    0    0    0    0  722
```

```
pred_train2 <- predict(modFit_rf, train_data)
train_data$predRight <- pred_train2==train_data$classe
table(pred_train2, train_data$classe)
```

```
##
## pred_train2    A    B    C    D    E
##            A 5579    0    0    0    0
##            B    1 3797    0    0    0
##            C    0    0 3422    0    0
##            D    0    0    0 3216    0
##            E    0    0    0    0 3607
```

```
clean_test_data <- function(data){

  data <- data[keep_cols]
  data <- data[keep_cols_2]

  # Get rid of X column
  data <- data[2:60]


  # Add the time diff column (t2_delta)
  data <- group_by(data, num_window) %>% mutate(t2_delta =
                                                  c(0,diff(raw_timestamp_part_2)))

   data$win_ind <- ave(data$new_window=="yes", FUN = function(x) {
            cumsum(x) })

  # Get rid of other cols
  data <- data[6:61]


  # Remove win_ind col

}
```

# Testing

```
testing<- read.csv('pml-testing.csv')

# Repeat the transforms, which are captured in a function.

clean_test <- clean_test_data(testing)

pred_test <- predict(modFit_rf, clean_test)
pred_test
```

```
##  [1] A A A A A A A A A A A A A A A A A A A A
## Levels: A B C D E
```