Assesment

Keyur Nandu

5/12/2020

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 (see the section on the Weight Lifting Exercise Dataset).

```r
library(caret)
library(ggplot2)
library(rpart)
library(randomForest)
library(dplyr)
```

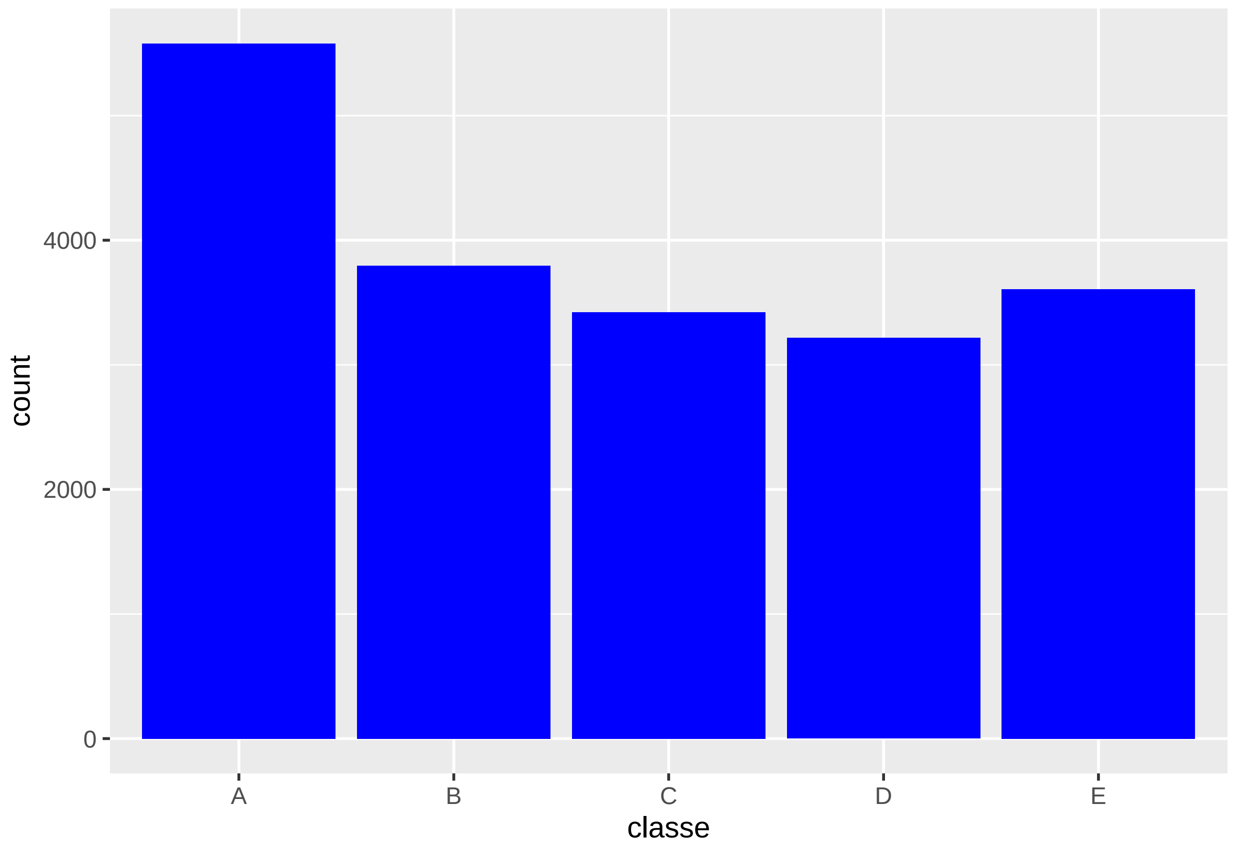
Loadind the data set and exploring it!

```r
#along with importing data we are adding NA to blank cells of csv files
raw_train <- read.csv("data/pml-training.csv", na.strings=c("","NA"))
raw_test <- read.csv("data/pml-testing.csv", na.strings=c("","NA"))

head(raw_train)
```

```
##   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 skewness_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>
##   skewness_roll_belt.1 skewness_yaw_belt max_roll_belt max_picth_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_yaw_belt min_roll_belt min_pitch_belt min_yaw_belt amplitude_roll_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_pitch_belt amplitude_yaw_belt var_total_accel_belt avg_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
##   stddev_roll_belt var_roll_belt avg_pitch_belt stddev_pitch_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
##   var_pitch_belt avg_yaw_belt stddev_yaw_belt var_yaw_belt gyros_belt_x
## 1             NA           NA              NA           NA         0.00
## 2             NA           NA              NA           NA         0.02
## 3             NA           NA              NA           NA         0.00
## 4             NA           NA              NA           NA         0.02
## 5             NA           NA              NA           NA         0.02
## 6             NA           NA              NA           NA         0.02
##   gyros_belt_y gyros_belt_z accel_belt_x accel_belt_y accel_belt_z
## 1         0.00        -0.02          -21            4           22
## 2         0.00        -0.02          -22            4           22
## 3         0.00        -0.02          -20            5           23
## 4         0.00        -0.03          -22            3           21
## 5         0.02        -0.02          -21            2           24
## 6         0.00        -0.02          -21            4           21
##   magnet_belt_x magnet_belt_y magnet_belt_z roll_arm pitch_arm yaw_arm
## 1            -3           599          -313     -128      22.5    -161
## 2            -7           608          -311     -128      22.5    -161
## 3            -2           600          -305     -128      22.5    -161
## 4            -6           604          -310     -128      22.1    -161
## 5            -6           600          -302     -128      22.1    -161
## 6             0           603          -312     -128      22.0    -161
##   total_accel_arm var_accel_arm avg_roll_arm stddev_roll_arm var_roll_arm
## 1              34            NA           NA              NA           NA
## 2              34            NA           NA              NA           NA
## 3              34            NA           NA              NA           NA
## 4              34            NA           NA              NA           NA
## 5              34            NA           NA              NA           NA
## 6              34            NA           NA              NA           NA
##   avg_pitch_arm stddev_pitch_arm var_pitch_arm avg_yaw_arm stddev_yaw_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_yaw_arm gyros_arm_x gyros_arm_y gyros_arm_z accel_arm_x accel_arm_y
## 1          NA        0.00        0.00       -0.02        -288         109
## 2          NA        0.02       -0.02       -0.02        -290         110
## 3          NA        0.02       -0.02       -0.02        -289         110
## 4          NA        0.02       -0.03        0.02        -289         111
## 5          NA        0.00       -0.03        0.00        -289         111
## 6          NA        0.02       -0.03        0.00        -289         111
##   accel_arm_z magnet_arm_x magnet_arm_y magnet_arm_z kurtosis_roll_arm
## 1        -123         -368          337          516              <NA>
## 2        -125         -369          337          513              <NA>
## 3        -126         -368          344          513              <NA>
## 4        -123         -372          344          512              <NA>
## 5        -123         -374          337          506              <NA>
## 6        -122         -369          342          513              <NA>
##   kurtosis_picth_arm kurtosis_yaw_arm skewness_roll_arm skewness_pitch_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>
##   skewness_yaw_arm max_roll_arm max_picth_arm max_yaw_arm min_roll_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_pitch_arm min_yaw_arm amplitude_roll_arm amplitude_pitch_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
##   amplitude_yaw_arm roll_dumbbell pitch_dumbbell yaw_dumbbell
## 1                NA      13.05217      -70.49400    -84.87394
## 2                NA      13.13074      -70.63751    -84.71065
## 3                NA      12.85075      -70.27812    -85.14078
## 4                NA      13.43120      -70.39379    -84.87363
## 5                NA      13.37872      -70.42856    -84.85306
## 6                NA      13.38246      -70.81759    -84.46500
##   kurtosis_roll_dumbbell kurtosis_picth_dumbbell kurtosis_yaw_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_roll_dumbbell skewness_pitch_dumbbell skewness_yaw_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_roll_dumbbell max_picth_dumbbell max_yaw_dumbbell min_roll_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_pitch_dumbbell min_yaw_dumbbell amplitude_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
##   amplitude_pitch_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
## 1                       NA                   <NA>                   37
## 2                       NA                   <NA>                   37
## 3                       NA                   <NA>                   37
## 4                       NA                   <NA>                   37
## 5                       NA                   <NA>                   37
## 6                       NA                   <NA>                   37
##   var_accel_dumbbell avg_roll_dumbbell stddev_roll_dumbbell var_roll_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
##   avg_pitch_dumbbell stddev_pitch_dumbbell var_pitch_dumbbell avg_yaw_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
##   stddev_yaw_dumbbell var_yaw_dumbbell gyros_dumbbell_x gyros_dumbbell_y
## 1                  NA               NA                0            -0.02
## 2                  NA               NA                0            -0.02
## 3                  NA               NA                0            -0.02
## 4                  NA               NA                0            -0.02
## 5                  NA               NA                0            -0.02
## 6                  NA               NA                0            -0.02
##   gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y accel_dumbbell_z
## 1             0.00             -234               47             -271
## 2             0.00             -233               47             -269
## 3             0.00             -232               46             -270
## 4            -0.02             -232               48             -269
## 5             0.00             -233               48             -270
## 6             0.00             -234               48             -269
##   magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z roll_forearm
## 1              -559               293               -65         28.4
## 2              -555               296               -64         28.3
## 3              -561               298               -63         28.3
## 4              -552               303               -60         28.1
## 5              -554               292               -68         28.0
## 6              -558               294               -66         27.9
##   pitch_forearm yaw_forearm kurtosis_roll_forearm kurtosis_picth_forearm
## 1         -63.9        -153                  <NA>                   <NA>
## 2         -63.9        -153                  <NA>                   <NA>
## 3         -63.9        -152                  <NA>                   <NA>
## 4         -63.9        -152                  <NA>                   <NA>
## 5         -63.9        -152                  <NA>                   <NA>
## 6         -63.9        -152                  <NA>                   <NA>
##   kurtosis_yaw_forearm skewness_roll_forearm skewness_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>
##   skewness_yaw_forearm max_roll_forearm max_picth_forearm max_yaw_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_roll_forearm min_pitch_forearm min_yaw_forearm amplitude_roll_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
##   amplitude_pitch_forearm amplitude_yaw_forearm total_accel_forearm
## 1                      NA                  <NA>                  36
## 2                      NA                  <NA>                  36
## 3                      NA                  <NA>                  36
## 4                      NA                  <NA>                  36
## 5                      NA                  <NA>                  36
## 6                      NA                  <NA>                  36
##   var_accel_forearm avg_roll_forearm stddev_roll_forearm var_roll_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
##   avg_pitch_forearm stddev_pitch_forearm var_pitch_forearm avg_yaw_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_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
```

```
raw_train %>% ggplot(aes(classe)) + geom_bar(stat = "count", fill = "blue")
```



As there are null values and many columns in the columns we need to do some data wrangling!

```
# removing columns with NA greater than 50%
training <- raw_train[, which(colMeans(!is.na(raw_train)) > 0.5)]
testing <- raw_test[,which(colMeans(!is.na(raw_train))>0.5)]

#First seven columns of the dataset are not so useful we can remove them
sam <- training[,-c(1:7)]
testing <- testing[,-c(1:7)]

#creating data partion of training sample for modeling
intrain <- createDataPartition(y= sam$classe, p=0.7, list = FALSE)

sam_train <- sam[intrain,]
sam_test <- sam[-intrain,]
```

#Decision Tree

```
model_dt <- rpart(classe ~ ., data = sam_train, method = "class")
pred_dt <- predict(model_dt,sam_test, type = "class")

confusionMatrix(table(pred_dt,sam_test$class))
```

```
## Confusion Matrix and Statistics
##
```

```
##
## pred_dt    A    B    C    D    E
##       A 1515  250   18  115   44
##       B   37  610   52   23   76
##       C   39  119  822  137  118
##       D   55   70   55  609   49
##       E   28   90   79   80  795
##
## Overall Statistics
##
##                Accuracy : 0.7393
##                  95% CI : (0.7279, 0.7505)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6686
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9050   0.5356   0.8012   0.6317   0.7348
## Specificity            0.8986   0.9604   0.9150   0.9535   0.9423
## Pos Pred Value         0.7801   0.7644   0.6656   0.7267   0.7416
## Neg Pred Value         0.9597   0.8960   0.9561   0.9297   0.9404
## Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Rate         0.2574   0.1037   0.1397   0.1035   0.1351
## Detection Prevalence   0.3300   0.1356   0.2099   0.1424   0.1822
## Balanced Accuracy      0.9018   0.7480   0.8581   0.7926   0.8385
```

#Random Forest

```r
library(parallel)
library(doParallel)
```

```
## Loading required package: foreach
```

```
## Loading required package: iterators
```

```r
cluster <- makeCluster(detectCores() - 1) # convention to leave 1 core for OS
registerDoParallel(cluster)

fitControl <- trainControl(method = "cv",
                           number = 5,
                           allowParallel = TRUE)



model_rf <- train(classe ~ ., method = "rf", data = sam_train, trControl =fitControl )

stopCluster(cluster)
registerDoSEQ()
```

```
pred_rf <- predict(model_rf,sam_test)

confusionMatrix(table(pred_rf,sam_test$class))
```

```
## Confusion Matrix and Statistics
##
##
## pred_rf    A    B    C    D    E
##       A 1672   11    0    0    0
##       B    0 1124    4    1    0
##       C    2    4 1018   10    1
##       D    0    0    4  952    5
##       E    0    0    0    1 1076
##
## Overall Statistics
##
##                Accuracy : 0.9927
##                  95% CI : (0.9902, 0.9947)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9908
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9988   0.9868   0.9922   0.9876   0.9945
## Specificity            0.9974   0.9989   0.9965   0.9982   0.9998
## Pos Pred Value         0.9935   0.9956   0.9836   0.9906   0.9991
## Neg Pred Value         0.9995   0.9968   0.9984   0.9976   0.9988
## Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Rate         0.2841   0.1910   0.1730   0.1618   0.1828
## Detection Prevalence   0.2860   0.1918   0.1759   0.1633   0.1830
## Balanced Accuracy      0.9981   0.9929   0.9944   0.9929   0.9971
```

Final prediction using testing data set on random forest model as we are getting 99.25 accuracy

```
# This will give us the prediction of the 20 variables in testing dataset
list(predict(model_rf, testing))
```

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