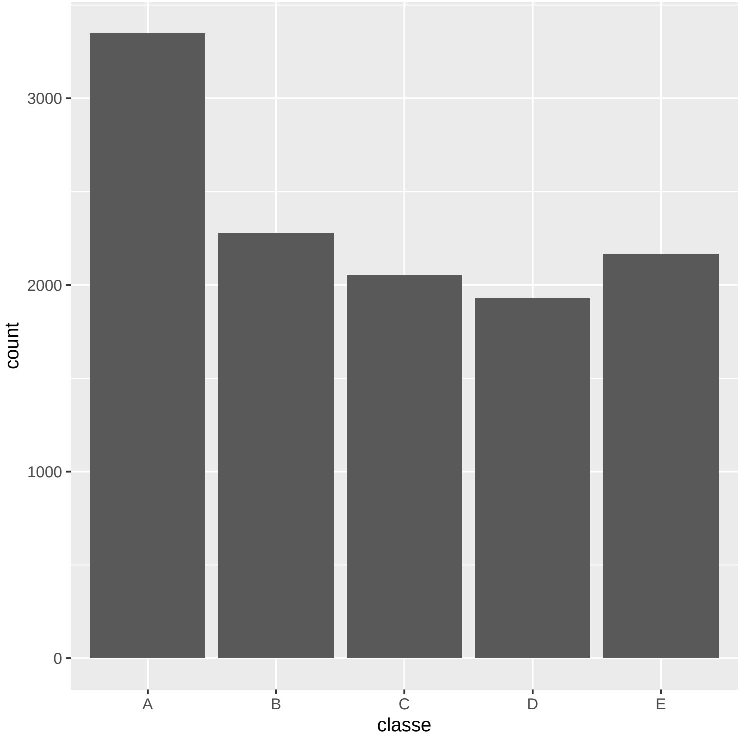
August 30, 2020

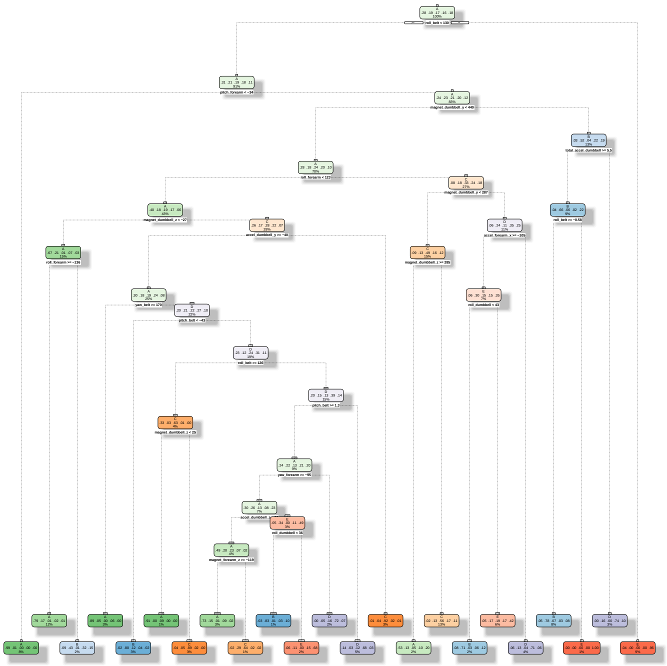
The results below are generated from an R script.

```
## The required packages

library(lattice)
library(caret)
library(ggplot2)
library(randomForest)
library(rpart)
library(rattle)


set.seed(1234)

## Downloading the files

trainingUrl <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"
testingUrl <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"

if (!file.exists(’./pml-testing.csv’) & !file.exists(’./pml-training.csv’)){
  download.file(testingUrl,’./pml-testing.csv’, mode = ’wb’)
  download.file(trainingUrl,’./pml-training.csv’, mode = ’wb’)
}

## Loading the files

train <- read.csv("pml-training.csv", na.strings = c("NA","#DIV/0!",""))
test <- read.csv("pml-testing.csv", na.strings = c("NA","#DIV/0!",""))

## Deleting columns with missing values only

train<-train[,colSums(is.na(train)) == 0]
test<-test[,colSums(is.na(test)) == 0]

## Removing non-relevant variables

train <- train[,-c(1:7)]
test <-test[,-c(1:7)]


## Partitioning the training set

inTrain <- createDataPartition(y=train$classe, p = 0.6, list = FALSE)
trainTrain <- train [inTrain, ]
trainTest <- train[-inTrain, ]
```

```
## Visualizing the training set
classe <- trainTrain$classe
ggplot(data.frame(classe), aes (x = classe)) + geom_bar()
```

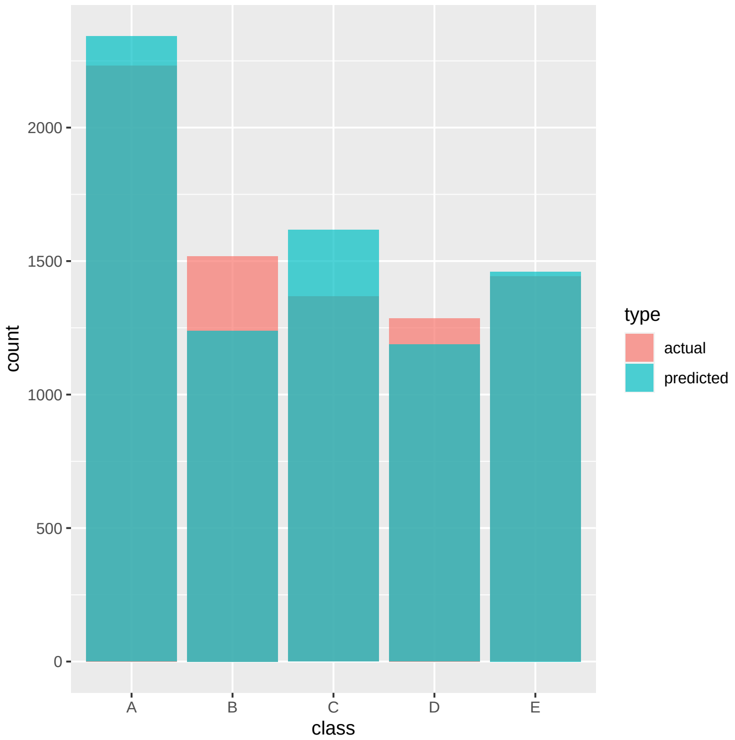


```
## Decision Tree model with its plot and prediction

modFit1 <- rpart(classe ~ ., data=trainTrain, method = "class")
fancyRpartPlot(modFit1)

## Warning:  labs do not fit even at cex 0.15, there may be some overplotting
```



Rattle 2020−8−30 22:25:55 TA

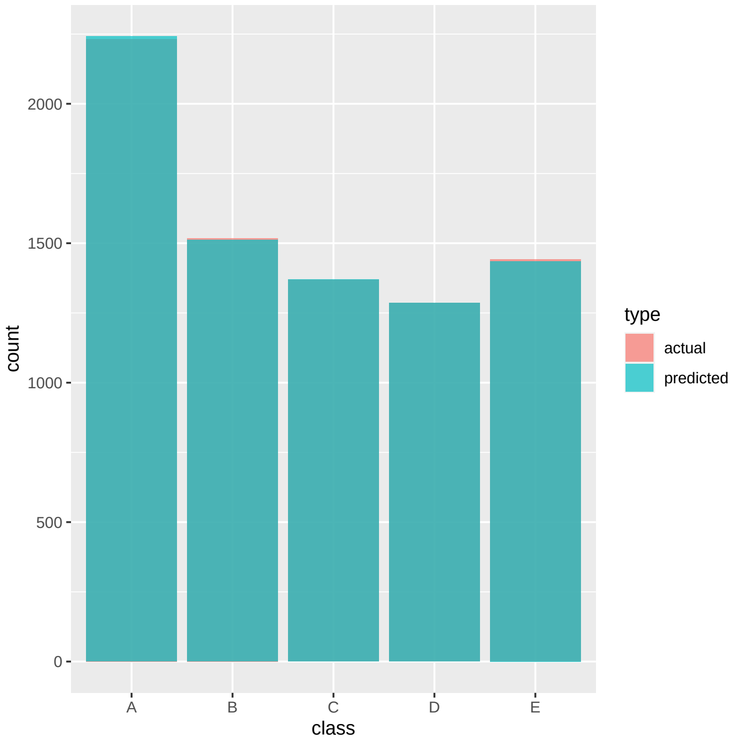
```
prediction1 <- predict(modFit1, trainTest, type = "class")

## Plot the prediction in comparison to the actual value

p1 <- data.frame(class = prediction1)
cla <- data.frame(class = trainTest$classe)
p1$type <- 'predicted'
cla$type <- 'actual'
v <- rbind(p1, cla)
ggplot(v, aes(class, fill = type))+ geom_histogram(alpha = 0.7,
                                                   stat = "count", position = 'identity')

## Warning:  Ignoring unknown parameters:  binwidth, bins, pad
```



```
## Checking if the predictions and classe variables are same level factors

str(prediction1)

##  Factor w/ 5 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...
##  - attr(*, "names")= chr [1:7846] "1" "2" "6" "8" ...

str(trainTest$classe)

##  chr [1:7846] "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" ...

## Turning the classe variable into appropriate factor
factoredClasse <- factor(trainTest$classe, levels = c ("A", "B", "C", "D", "E"))

## Testing the prediction
confusionMatrix(prediction1, factoredClasse)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1930  231   47   81   53
##          B   82  866   67  103  121
##          C   59  189 1062  173  134
##          D   94  118   85  821   70
##          E   67  114  107  108 1064
##
## Overall Statistics
##
##                Accuracy : 0.732
##                  95% CI : (0.722, 0.7417)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6605
##
##  Mcnemar’s Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.8647   0.5705   0.7763   0.6384   0.7379
## Specificity            0.9266   0.9411   0.9143   0.9441   0.9382
## Pos Pred Value         0.8241   0.6990   0.6568   0.6911   0.7288
## Neg Pred Value         0.9451   0.9013   0.9509   0.9302   0.9408
## Prevalence             0.2845   0.1935   0.1744   0.1639   0.1838
## Detection Rate         0.2460   0.1104   0.1354   0.1046   0.1356
## Detection Prevalence   0.2985   0.1579   0.2061   0.1514   0.1861
## Balanced Accuracy      0.8957   0.7558   0.8453   0.7912   0.8380
```

```
## Random forest model with its prediction

modFit2 <- randomForest(as.factor(classe) ~. , data = trainTrain)
prediction2 <- predict(modFit2, trainTest, type = "class")

## Plot the prediction in comparison to the actual value

p2 <- data.frame(class = prediction2)
p2$type <- ’predicted’
v2 <- rbind(p2, cla)
ggplot(v2, aes(class, fill = type))+ geom_histogram(alpha = 0.7,
                                                    stat = "count", position = ’identity’)
```

```
## Warning:  Ignoring unknown parameters:  binwidth, bins, pad
```



```
## Testing the result

confusionMatrix(prediction2, factoredClasse)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 2230   12    0    0    0
##          B    1 1506    5    0    0
##          C    0    0 1357   10    3
##          D    1    0    6 1274    5
##          E    0    0    0    2 1434
##
## Overall Statistics
##
##                Accuracy : 0.9943
##                  95% CI : (0.9923, 0.9958)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9927
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9991   0.9921   0.9920   0.9907   0.9945
## Specificity            0.9979   0.9991   0.9980   0.9982   0.9997
## Pos Pred Value         0.9946   0.9960   0.9905   0.9907   0.9986
## Neg Pred Value         0.9996   0.9981   0.9983   0.9982   0.9988
## Prevalence             0.2845   0.1935   0.1744   0.1639   0.1838
## Detection Rate         0.2842   0.1919   0.1730   0.1624   0.1828
## Detection Prevalence   0.2858   0.1927   0.1746   0.1639   0.1830
## Balanced Accuracy      0.9985   0.9956   0.9950   0.9944   0.9971
```

The R session information (including the OS info, R version and all packages used):

```
sessionInfo()

## R version 4.0.2 (2020-06-22)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19041)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=Korean_Korea.949  LC_CTYPE=Korean_Korea.949    LC_MONETARY=Korean_Korea.949
## [4] LC_NUMERIC=C                 LC_TIME=Korean_Korea.949
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] Hmisc_4.4-1         Formula_1.2-3       survival_3.2-3      RColorBrewer_1.1-2
##  [5] rpart.plot_3.0.8    rattle_5.4.0        bitops_1.0-6        tibble_3.0.1
##  [9] rpart_4.1-15        randomForest_4.6-14 caret_6.0-86        ggplot2_3.3.2
## [13] lattice_0.20-41
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.4.6         lubridate_1.7.9      png_0.1-7            class_7.3-17
##  [5] packrat_0.5.0        digest_0.6.25        ipred_0.9-9          foreach_1.5.0
##  [9] R6_2.4.1             plyr_1.8.6           backports_1.1.7      stats4_4.0.2
## [13] evaluate_0.14        e1071_1.7-3          highr_0.8            pillar_1.4.4
## [17] rlang_0.4.6          rstudioapi_0.11      data.table_1.13.0    Matrix_1.2-18
## [21] checkmate_2.0.0      rmarkdown_2.3        labeling_0.3         splines_4.0.2
## [25] gower_0.2.2          stringr_1.4.0        foreign_0.8-80       htmlwidgets_1.5.1
## [29] tinytex_0.24         munsell_0.5.0        xfun_0.15            compiler_4.0.2
## [33] pkgconfig_2.0.3      base64enc_0.1-3      htmltools_0.5.0      nnet_7.3-14
## [37] tidyselect_1.1.0     gridExtra_2.3        htmlTable_2.0.1      prodlim_2019.11.13
## [41] codetools_0.2-16     crayon_1.3.4         dplyr_1.0.0          withr_2.2.0
## [45] MASS_7.3-51.6        recipes_0.1.13       ModelMetrics_1.2.2.2 grid_4.0.2
## [49] nlme_3.1-148         gtable_0.3.0         lifecycle_0.2.0      magrittr_1.5
## [53] pROC_1.16.2          scales_1.1.1         stringi_1.4.6        farver_2.0.3
## [57] reshape2_1.4.4       latticeExtra_0.6-29  timeDate_3043.102    ellipsis_0.3.1
## [61] generics_0.0.2       vctrs_0.3.1          lava_1.6.7           iterators_1.0.12
## [65] tools_4.0.2          glue_1.4.1           purrr_0.3.4          jpeg_0.1-8.1
## [69] rsconnect_0.8.16     yaml_2.2.1           colorspace_1.4-1     cluster_2.1.0
## [73] knitr_1.29

Sys.time()

## [1] "2020-08-30 22:26:42 KST"
```