RF

```
library(mlbench)
data(Glass)
library(randomForest)
```

```
## randomForest 4.6-14
```

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

```
library(MASS)
library(caret)
```

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

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

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

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

```
Glass$Type <- as.factor(c(rep('Window', 163), rep('Non-Window', 51)))
set.seed(42)
train <- sample(1:nrow(Glass), nrow(Glass)/2)
```

```
# sample size
ceiling(.632*nrow(Glass[-train,]))
```

```
## [1] 68
```

```
# number of vars at each split
floor(sqrt(ncol(Glass)))
```

```
## [1] 3
```

```
set.seed(42)
rf_class <- randomForest(Type ~ ., data = Glass,
                         subset = train,
                         mtry = 3,
                         sampsize = 68,
                         importance = T)
rf_class
```

```
##
## Call:
##  randomForest(formula = Type ~ ., data = Glass, mtry = 3, sampsize = 68,      importance = T, subset = train)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 5.61%
## Confusion matrix:
##             Non-Window Window class.error
## Non-Window         18      4  0.18181818
## Window              2     83  0.02352941
```

```
# test
est_medv <- predict(rf_class, newdata = Glass[-train,])
mean(est_medv != Glass$Type[-train])
```

```
## [1] 0.1028037
```

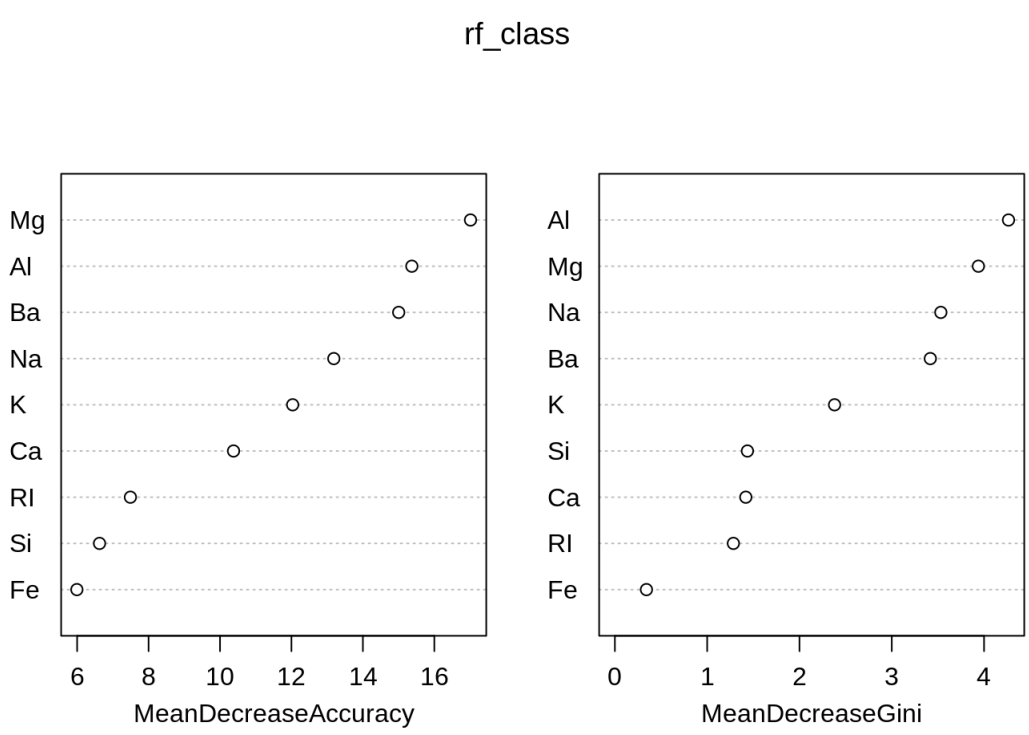
```
set.seed(42)
ntrees <- 500
rf_class <- randomForest(Type ~ ., data = Glass,
                         subset = train,
                         ntree = ntrees,
                         mtry = 3,
                         sampsize = 68,
                         importance = T,
                         do.trace = ntrees/10)
```

```
## ntree      OOB      1      2
##    50:   7.48% 22.73%  3.53%
##   100:   5.61% 13.64%  3.53%
##   150:   4.67% 13.64%  2.35%
##   200:   5.61% 18.18%  2.35%
##   250:   5.61% 18.18%  2.35%
##   300:   6.54% 22.73%  2.35%
##   350:   6.54% 22.73%  2.35%
##   400:   5.61% 18.18%  2.35%
##   450:   5.61% 18.18%  2.35%
##   500:   5.61% 18.18%  2.35%
```

```
rf_class
```

```
##
## Call:
##  randomForest(formula = Type ~ ., data = Glass, ntree = ntrees,      mtry = 3, sampsize = 68, importance = T, do.trace = ntrees/10,      subset = train)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 5.61%
## Confusion matrix:
##             Non-Window Window class.error
## Non-Window         18      4  0.18181818
## Window              2     83  0.02352941
```

```
varImpPlot(rf_class)
```



```
set.seed(42)
ntrees <- 150
rf_class <- randomForest(Type ~ ., data = Glass,
                         subset = train,
                         ntree = ntrees,
                         mtry = 3,
                         sampsize = 68,
                         importance = T,
                         do.trace = ntrees/25)
```

```
## ntree      OOB      1      2
##     6:   5,66%  9,09%  4,76%
##    12:   6,54% 13,64%  4,71%
##    18:   7,48% 22,73%  3,53%
##    24:   7,48% 22,73%  3,53%
##    30:   7,48% 27,27%  2,35%
##    36:   6,54% 22,73%  2,35%
##    42:   6,54% 22,73%  2,35%
##    48:   5,61% 22,73%  1,18%
##    54:   7,48% 22,73%  3,53%
##    60:   6,54% 18,18%  3,53%
##    66:   6,54% 18,18%  3,53%
##    72:   6,54% 22,73%  2,35%
##    78:   6,54% 22,73%  2,35%
##    84:   6,54% 22,73%  2,35%
##    90:   4,67% 13,64%  2,35%
##    96:   4,67% 13,64%  2,35%
##   102:   5,61% 13,64%  3,53%
##   108:   5,61% 13,64%  3,53%
##   114:   5,61% 13,64%  3,53%
##   120:   5,61% 13,64%  3,53%
##   126:   5,61% 13,64%  3,53%
##   132:   5,61% 13,64%  3,53%
##   138:   5,61% 13,64%  3,53%
##   144:   4,67% 13,64%  2,35%
##   150:   4,67% 13,64%  2,35%
```

```
rf_class
```

```
##
## Call:
##  randomForest(formula = Type ~ ., data = Glass, ntree = ntrees,      mtry = 3, sampsize = 68, importance = T, do.trace = ntrees/25,      subset = train)
##                Type of random forest: classification
##                      Number of trees: 150
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 4,67%
## Confusion matrix:
##             Non-Window Window class.error
## Non-Window         19      3  0,13636364
## Window              2     83  0,02352941
```

```
# test
est_medv <- predict(rf_class, newdata = Glass[-train,])
mean(est_medv != Glass$Type[-train])
```

```
## [1] 0,1028037
```

```
set.seed(42)
ntrees <- 90
rf_class <- randomForest(Type ~ ., data = Glass,
                         subset = train,
                         ntree = ntrees,
                         mtry = 3,
                         sampsize = 68,
                         importance = T,
                         do.trace = ntrees/30)
```

```
## ntree      OOB      1      2
##     3:   7,69% 17,65%  5,41%
##     6:   5,66%  9,09%  4,76%
##     9:   8,41% 13,64%  7,06%
##    12:   6,54% 13,64%  4,71%
##    15:   6,54% 13,64%  4,71%
##    18:   7,48% 22,73%  3,53%
##    21:   7,48% 22,73%  3,53%
##    24:   7,48% 22,73%  3,53%
##    27:   5,61% 18,18%  2,35%
##    30:   7,48% 27,27%  2,35%
##    33:   6,54% 22,73%  2,35%
##    36:   6,54% 22,73%  2,35%
##    39:   7,48% 22,73%  3,53%
##    42:   6,54% 22,73%  2,35%
##    45:   6,54% 22,73%  2,35%
##    48:   5,61% 22,73%  1,18%
##    51:   7,48% 22,73%  3,53%
##    54:   7,48% 22,73%  3,53%
##    57:   7,48% 22,73%  3,53%
##    60:   6,54% 18,18%  3,53%
##    63:   5,61% 18,18%  2,35%
##    66:   6,54% 18,18%  3,53%
##    69:   6,54% 18,18%  3,53%
##    72:   6,54% 22,73%  2,35%
##    75:   6,54% 22,73%  2,35%
##    78:   6,54% 22,73%  2,35%
##    81:   6,54% 22,73%  2,35%
##    84:   6,54% 22,73%  2,35%
##    87:   5,61% 13,64%  3,53%
##    90:   4,67% 13,64%  2,35%
```

```
rf_class
```

```
##
## Call:
##  randomForest(formula = Type ~ ., data = Glass, ntree = ntrees,      mtry = 3, sampsize = 68, importance = T, do.trace = ntrees/30,      subset = train)
##                Type of random forest: classification
##                      Number of trees: 90
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 4,67%
## Confusion matrix:
##             Non-Window Window class.error
## Non-Window         19      3  0,13636364
## Window              2     83  0,02352941
```

```
# test
est_medv <- predict(rf_class, newdata = Glass[-train,])
mean(est_medv != Glass$Type[-train])
```

```
## [1] 0,1028037
```

```
set.seed(42)
ntrees <- 96
rf_class_2 <- randomForest(Type ~ ., data = Glass,
                      subset = train,
                      ntree = ntrees,
                      mtry = 3,
                      sampsize = 68,
                      importance = T)
rf_class_2
```

```
##
## Call:
##  randomForest(formula = Type ~ ., data = Glass, ntree = ntrees,      mtry = 3, sampsize = 68, importance = T, subset = train)
##                Type of random forest: classification
##                      Number of trees: 96
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 4,67%
## Confusion matrix:
##             Non-Window Window class.error
## Non-Window          19      3  0,13636364
## Window               2     83  0,02352941
```

```
# test
est_medv_2 <- predict(rf_class_2, newdata = Glass[-train,])
mean(est_medv_2 != Glass$Type[-train])
```

```
## [1] 0,1028037
```