1 RandomForest with R

```
library(randomForest)

# download Titanic Survivors data
data <- read.table("http://math.ucdenver.edu/RTutorial/titanic.txt", h=T, sep="\t")
# make survived into a yes/no
data$Survived <- as.factor(ifelse(data$Survived==1, "yes", "no"))

# split into a training and test set
idx <- runif(nrow(data)) <= .75
data.train <- data[idx,]
data.test <- data[-idx,]
```

Train a random forest

```
rf <- randomForest(Survived ~ PClass + Age + Sex,
             data=data.train, importance=TRUE, na.action=na.omit)
```

How important is each variable in the model?

```
imp <- importance(rf)
o <- order(imp[,3], decreasing=T)
imp[o,]
#             no      yes MeanDecreaseAccuracy MeanDecreaseGini
```

```
#Sex    51.49855 53.30255            55.13458         63.46861
#PClass 25.48715 24.12522            28.43298         22.31789
#Age    20.08571 14.07954            24.64607         19.57423
```

Display the confusion matrix

```
# confusion matrix [[True Neg, False Pos], [False Neg, True Pos]]
table(data.test$Survived, predict(rf, data.test),
  dnn=list("actual", "predicted"))
#      predicted
#actual  no yes
#   no  427  16
#   yes 117 195
```

```
library(caret)
mod <- train(Species ~ ., data = iris,
         method = "cforest",
         controls = cforest_unbiased(ntree = 10))
varImp(mod)
```

returns:

```
cforest variable importance
Overall
Petal.Width  100.0000
Petal.Length  86.6279
Sepal.Length   0.5814
Sepal.Width    0.0000
```