Soil and climate model for predict admixture klass of arabidopsis

Random forest based on all variables

Match admixture group for every sample base on the next simple rule, if component number N in vector > 0.7, then this sample belong to the group N.

```r
clear.just.parametrs <- just.parametrs

lbound <- .7
groups <- apply(all_data[, c(1:adm.count)], 1, function(current.row) {
    group <- ((current.row > lbound) * c(1:adm.count))[current.row > lbound]
    if (length(group) == 0) {
        return(NA)
    }
    return(group)
})
str(groups)
```

```
##  int [1:1048] 6 6 6 NA NA NA 11 11 11 11 ...
```

```r
clear.just.parametrs <- just.parametrs
prepared.data <- data.frame(group = groups)
prepared.data <- na.omit(cbind(prepared.data, clear.just.parametrs))
prepared.data$group <- factor(prepared.data$group)
ncol(prepared.data)
```

```
## [1] 59
```

```r
set.seed(666)

train.indexes <- sample(1:nrow(prepared.data), 0.8 * nrow(prepared.data))
train <- prepared.data[train.indexes, ]
test <- prepared.data[-train.indexes, ]

res.forest <- randomForest(group ~ ., data = train)
t_pred.forest <- predict(res.forest, test)
confMat <- table(test$group, t_pred.forest)
sum(diag(confMat)) / sum(confMat)
```

```
## [1] 0.8591549
```

```r
important.dot.plot(res.forest)
```