satRday Budapest, September 3, 2016

Uncovering political connections of firms using machine learning methods

emarsys

» János Divényi @janosdivenyi «

CEU MICRODATA

» Jenő Pál @paljenczy «

CEU Microdata

Miklós Koren

Ádám Szeidl

CEU MICRODATA

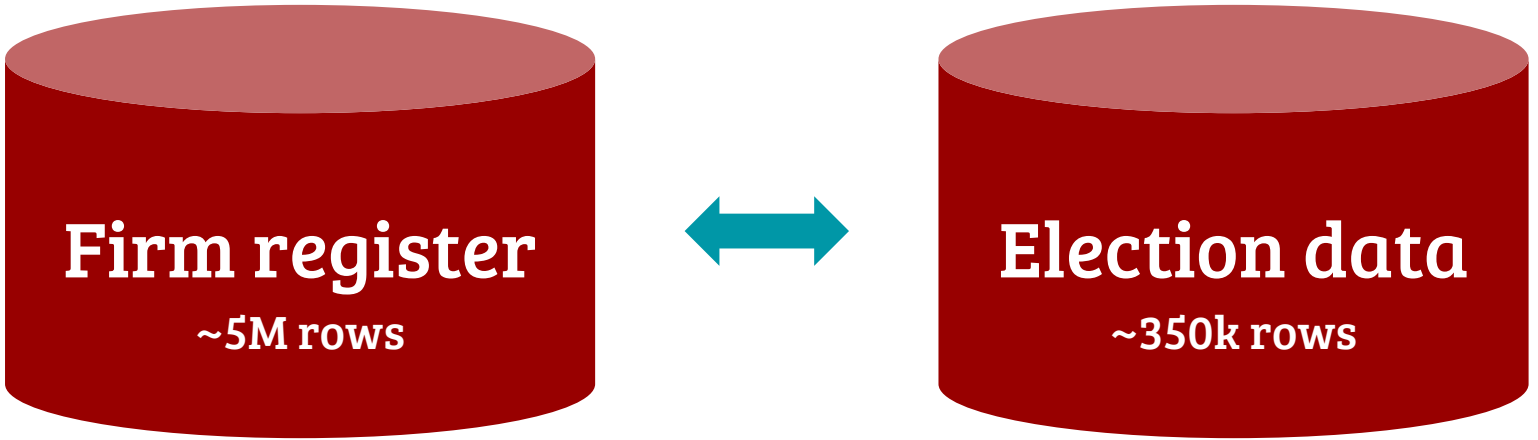
Political connections and favoritism in Hungary

Political connections matter

e'lios

Elios Innovatív Energetikai Zrt.

Information

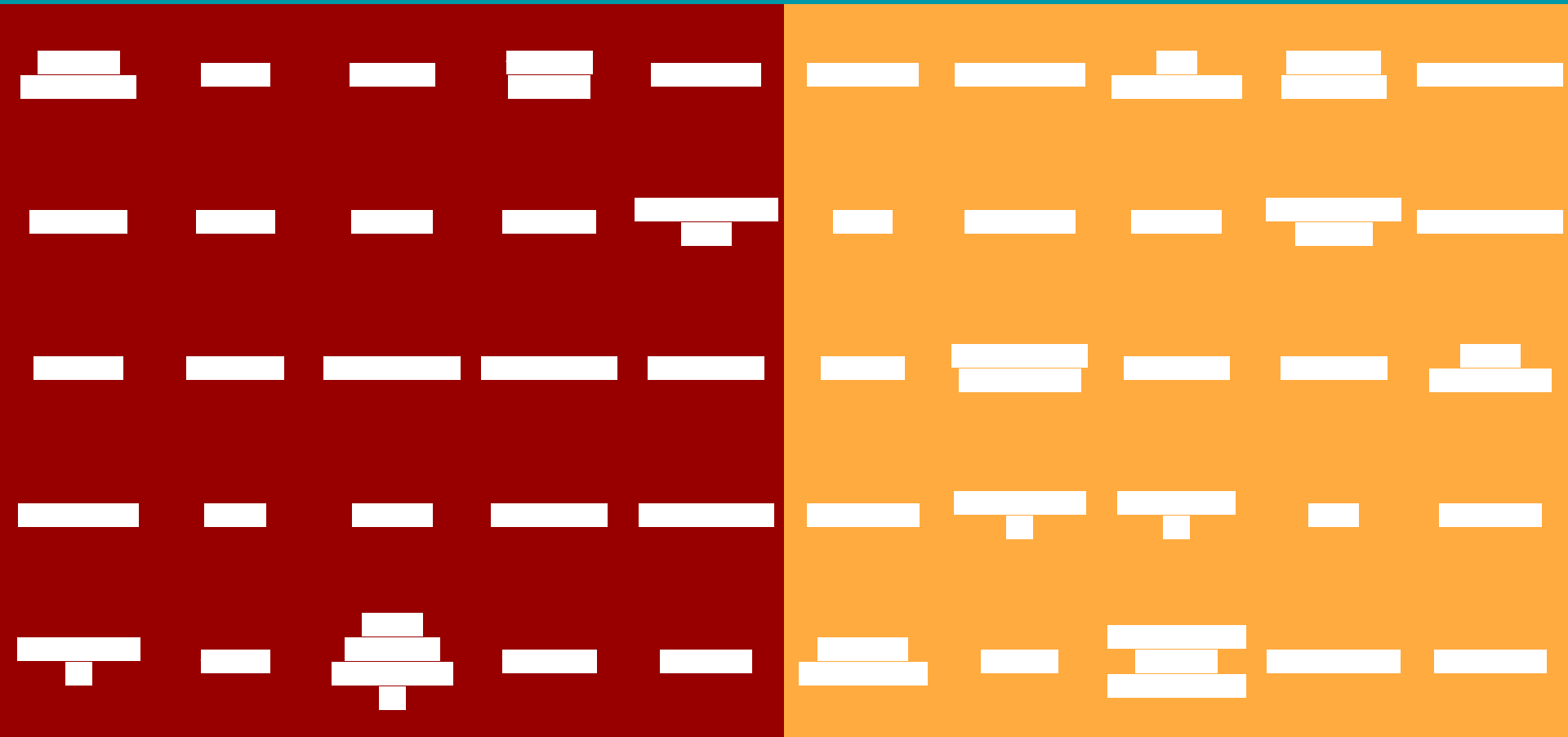


data.table

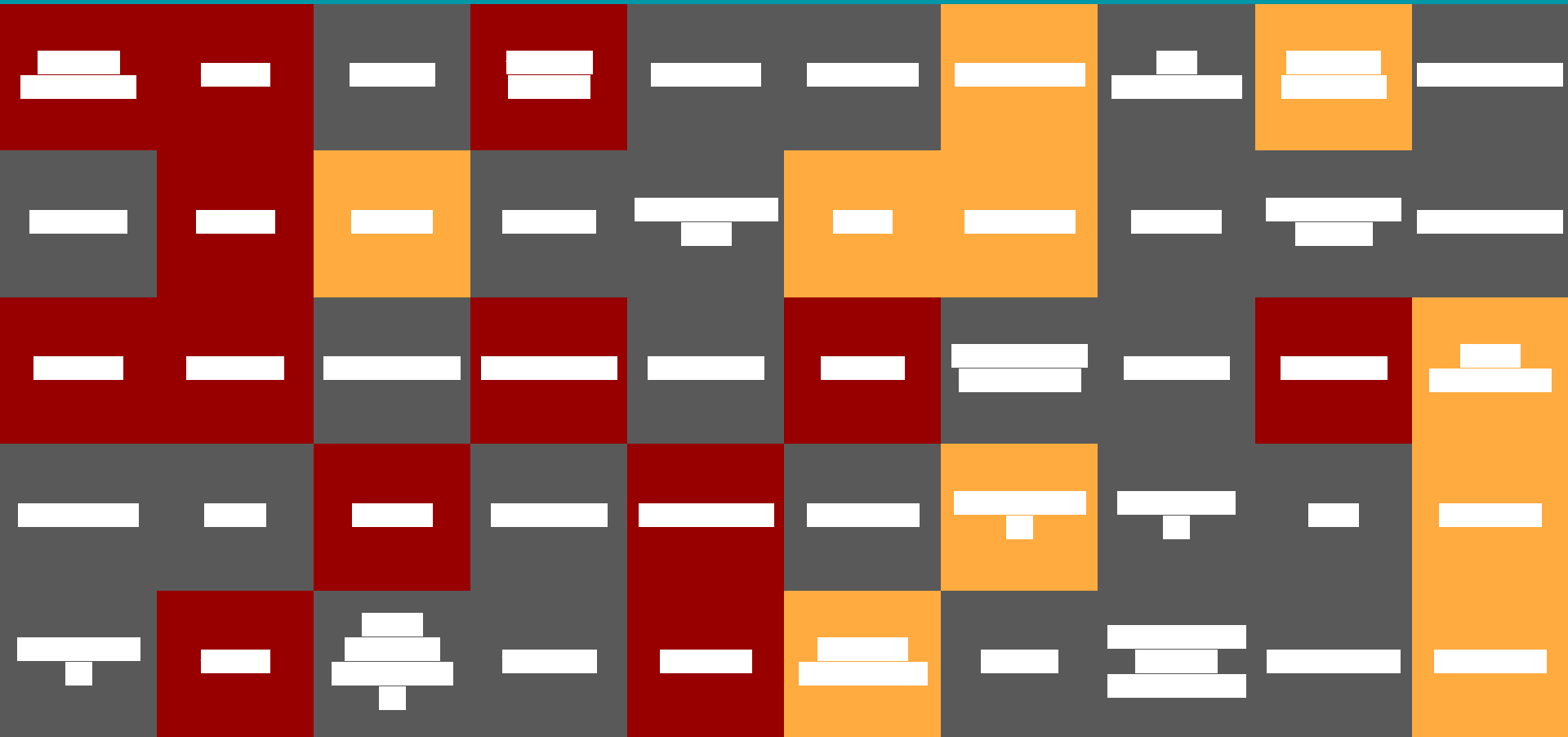
Decision rule

The firm is right
if there are
more right than left politicians
in the firm

Test firms: left or right

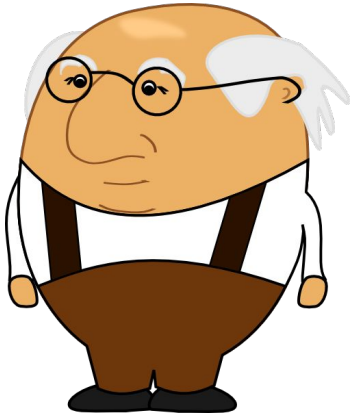


Test firms: left or right



Improve data

Ferenc Gyurcsány
PM of left coalition
2004-2009

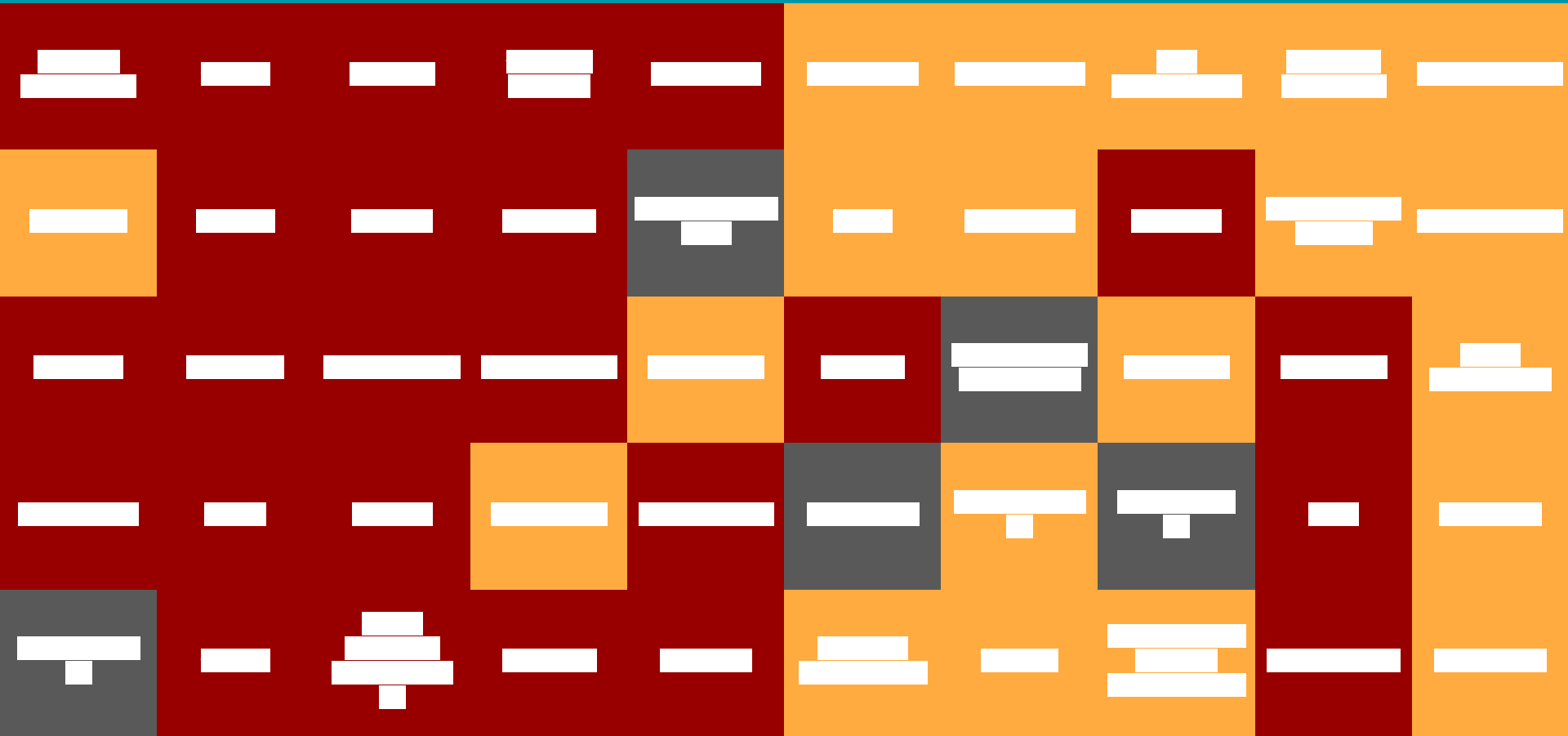


Ferenc Gyurcsán
local representative at
Nyíregyháza
1998

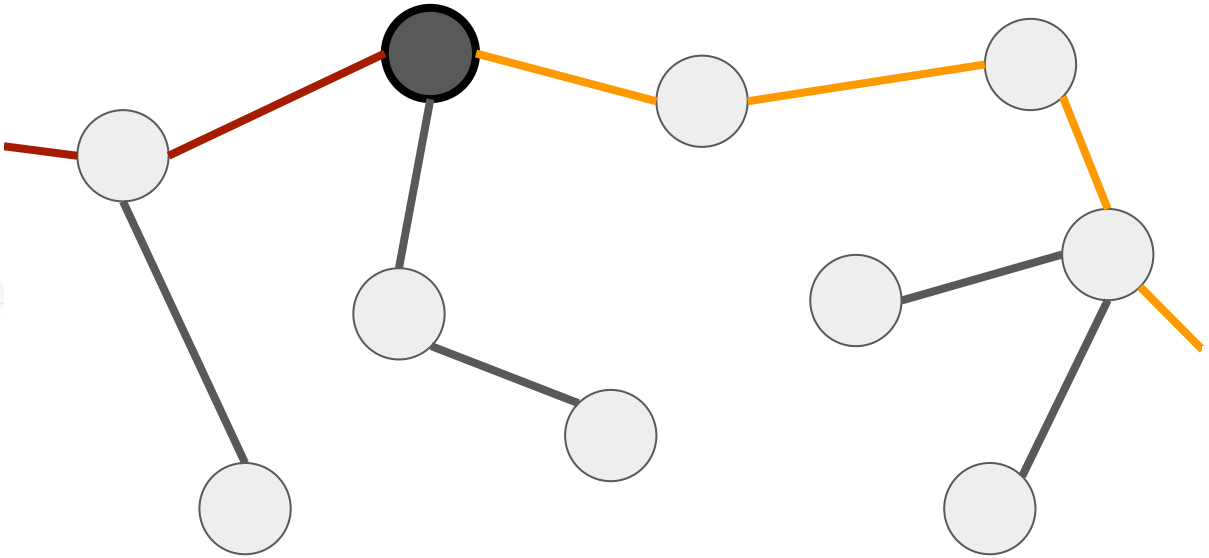
Probabilistic coloring

Prime minister:
69% left 31% other

Test firms: left or right



Improve information



igraph

Links: common ownership or location

Improve decision rule

caret

classification **a**nd
regression **t**raining

one interface to many algorithms

streamlines the process of machine learning

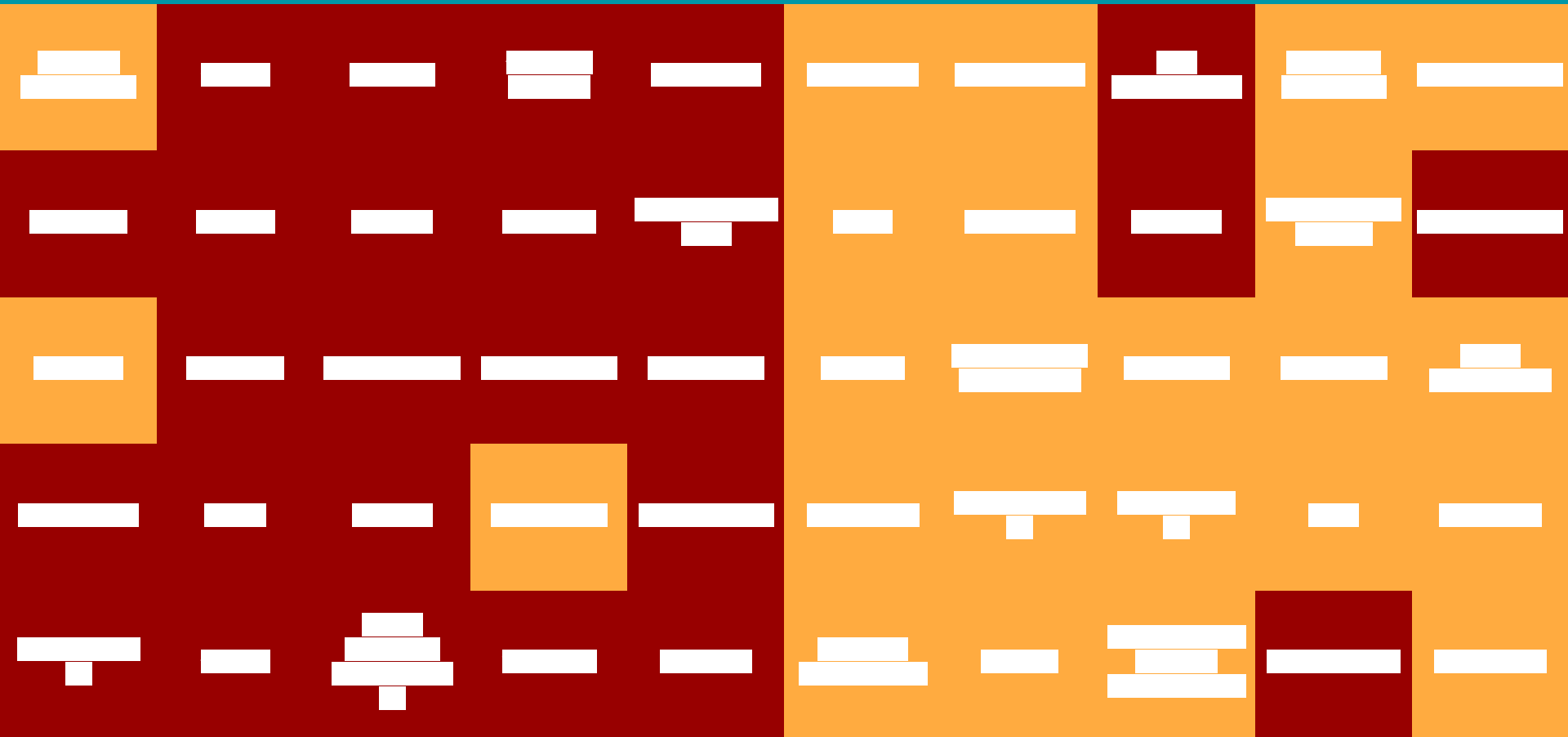
parallel computation with reproducibility

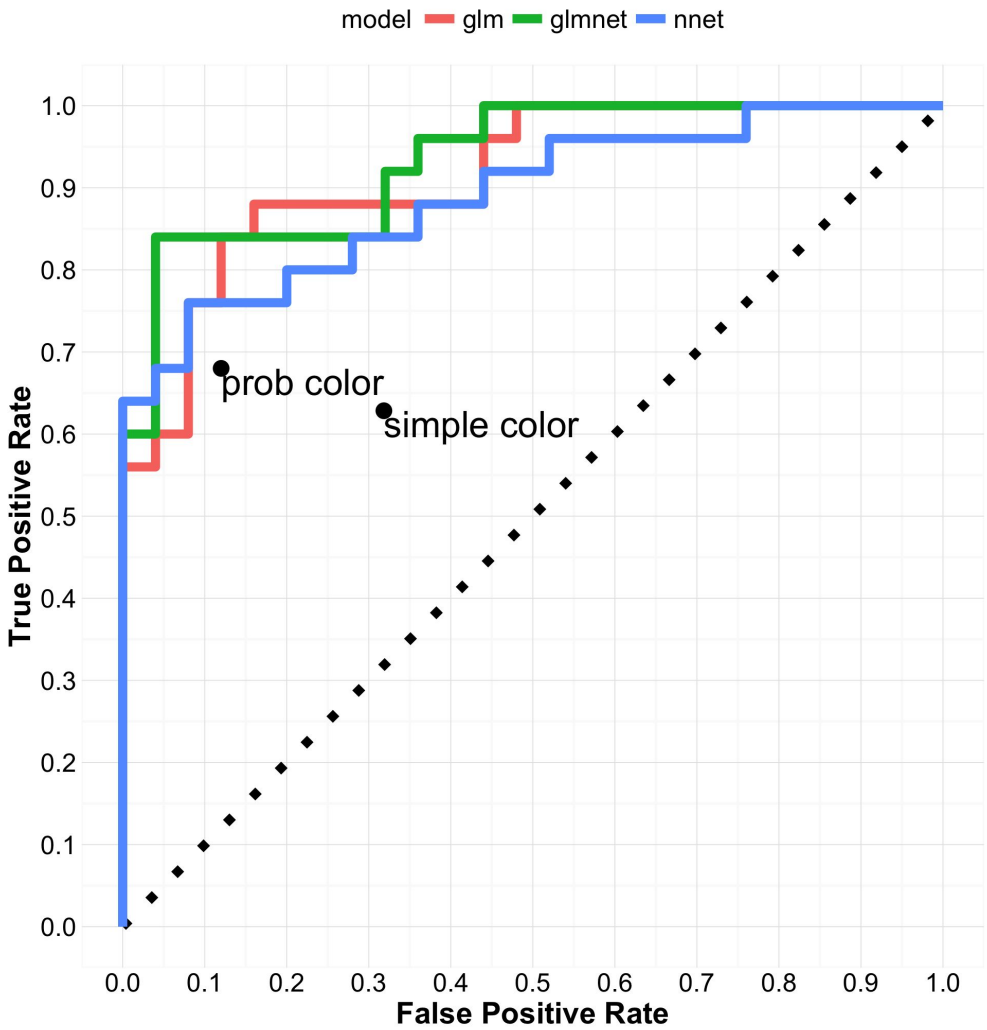
doParallel

The train function

```
fitControl <- trainControl(# 10-fold CV
                           method = "repeatedcv",
                           number = 10,
                           # repeated 5 times
                           repeats = 5)
set.seed(123)
model <- train(color ~ .,
               data = data.tr,
               method = "glmnet",
               tuneGrid = expand.grid(
                                  alpha = seq(0, 1, by = 0.1),
                                  lambda = 10^(-3:-1)),
               preProcess = c("center", "scale"),
               trControl = fitControl)
```

Test firms: left or right




model
glm
glmnet
nnet
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
True Positive Rate
prob color
simple color
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
False Positive Rate

Takeaways

iterative process

involving **manipulation**, **visualization**, **modelling**, etc