A graphical summary of your random forest

randomForestExplainer

April 10, 2019

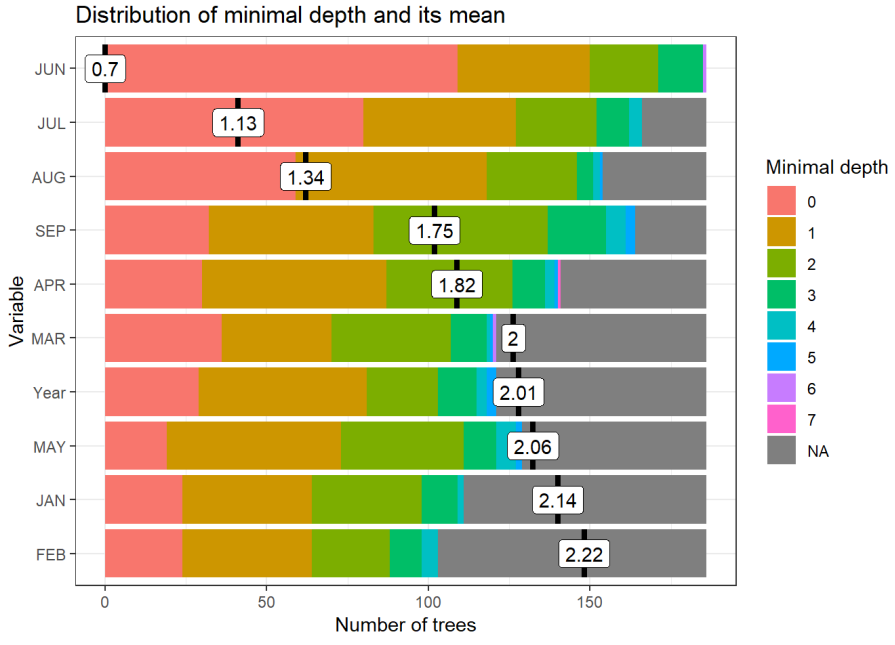
Details of your forest

```
##
## Call:
##  randomForest(x = x, y = y, mtry = res[which.min(res[, 2]), 1],      importance = TRUE, proximity = TRUE, stepfactor = 0.5)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 4
##
##           Mean of squared residuals: 4389479166
##                     % Var explained: -16.49
```

Distribution of minimal depth

The plot below shows the distribution of minimal depth among the trees of your forest. Note that:

- the mean of the distribution is marked by a vertical bar with a value label on it (the scale for it is different than for the rest of the plot),
- the scale of the X axis goes from zero to the maximum number of trees in which any variable was used for splitting.



Minimal depth for a variable in a tree equals to the depth of the node which splits on that variable and is the closest to the root of the tree. If it is low than a lot of observations are divided into groups on the basis of this variable

Importance measures

Below you can explore the measures of importance for all variables in the forest:

Show 10 **entries** **Search:**

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root	p_value
1	APR	1.8179	166	13,623,724.0238	2,082,983,656.6047	141	30	0.0452
2	AUG	1.3390	175	162,709,146.4992	5,537,407,493.8068	154	59	0.0075
3	DEC	2.4631	100	-17,162,865.5426	1,311,502,286.5867	90	17	1.0000
4	FEB	2.2208	111	-170,387,185.9676	2,092,384,419.3550	103	24	0.9992
5	JAN	2.1374	121	34,426,854.5753	1,542,904,310.8797	111	24	0.9866
6	JUL	1.1252	189	348,061,177.1050	6,507,435,423.8781	166	80	0.0002
7	JUN	0.7043	199	311,859,571.4473	9,283,742,198.8410	186	109	0.0000
8	MAR	1.9968	131	-22,639,425.0976	2,587,462,678.7790	121	36	0.9054
9	MAY	2.0586	142	-74,057,176.4393	1,850,476,518.6195	129	19	0.6347
10	NOV	2.3057	124	-134,205,850.5092	2,376,402,585.0013	110	27	0.9740

Showing 1 to 10 of 13 entries

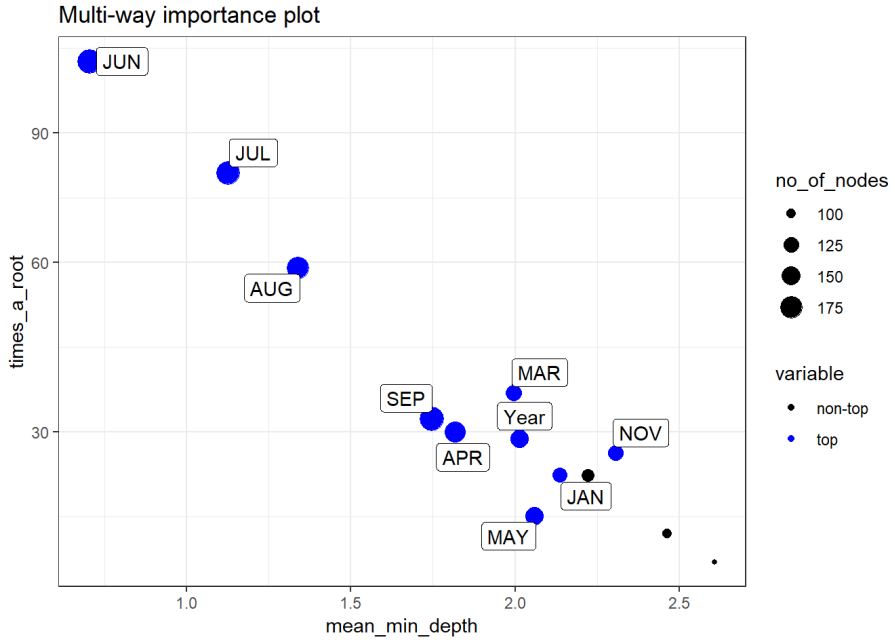
Previous 1 2 Next

Multi-way importance plot

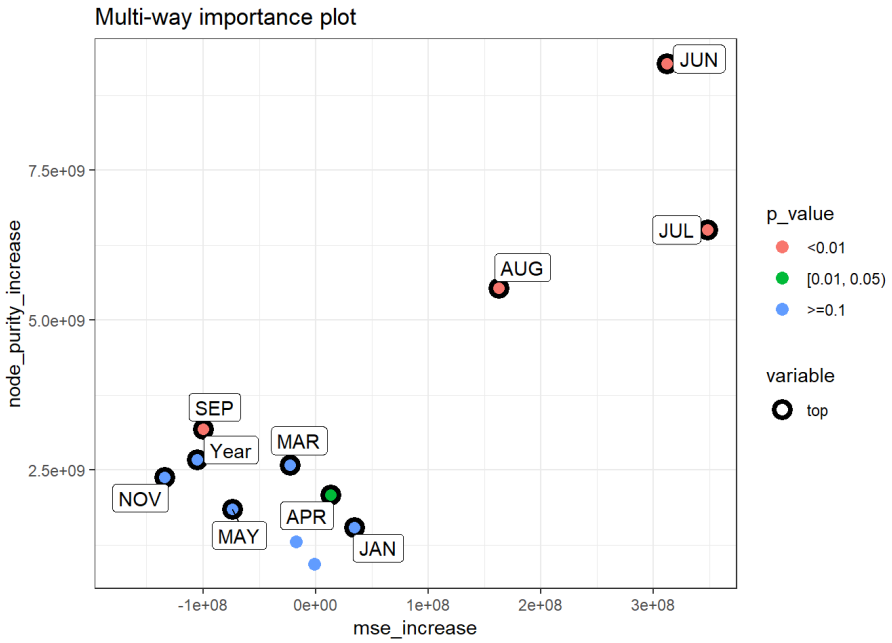
The multi-way importance plot shows the relation between three measures of importance and labels 10 variables which scored best when it comes to these three measures (i.e. for which the sum of the ranks for those measures is the lowest).

The first multi-way importance plot focuses on three importance measures that derive from the structure of trees in the forest:

- mean depth of first split on the variable,
- number of trees in which the root is split on the variable,
- the total number of nodes in the forest that split on that variable.



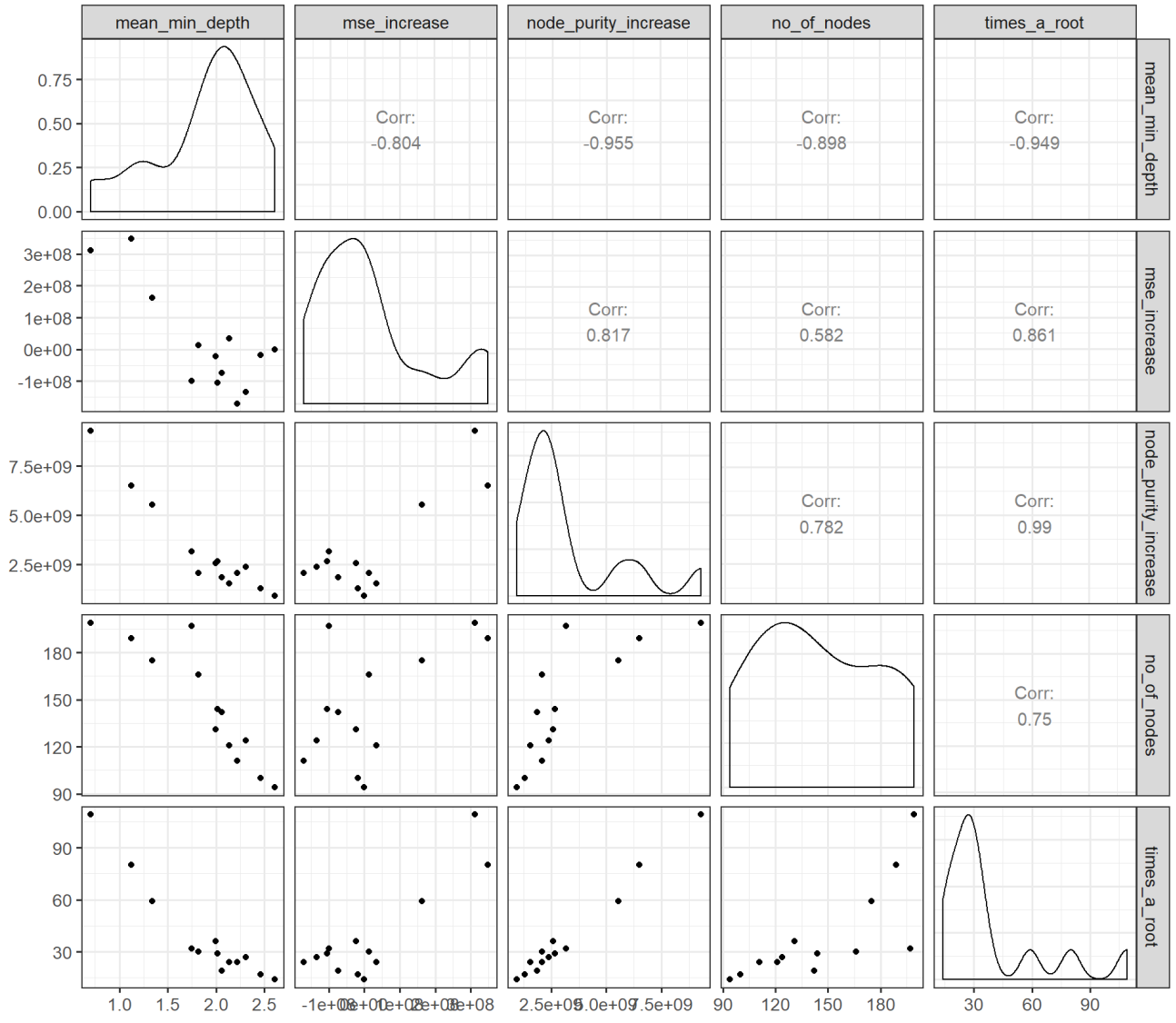
The second multi-way importance plot shows two importance measures that derive from the role a variable plays in prediction: with the additional information on the p-value based on a binomial distribution of the number of nodes split on the variable assuming that variables are randomly drawn to form splits (i.e. if a variable is significant it means that the variable is used for splitting more often than would be the case if the selection was random).



Compare importance measures

The plot below shows bilateral relations between the following importance measures: , if some variables are strongly related to each other it may be worth to consider focusing only on one of them.

Relations between measures of importance



Compare rankings of variables

The plot below shows bilateral relations between the rankings of variables according to chosen importance measures. This approach might be useful as rankings are more evenly spread than corresponding importance measures. This may also more clearly show where the different measures of importance disagree or agree.

Relations between rankings according to different measures