

permute_rf_strobl_mtry

Kelvyn Bladen

2024-05-17

```
s <- Sys.time()
library(ggplot2)
library(GGally)
library(ggeasy)
library(randomForest)
library(dplyr)
library(randomForestVIP)
library(tidyr)

rsq = vector(length = 12)

rf_oob_t <- mat.or.vec(8, 12)
rf_oob_f <- mat.or.vec(8, 12)
# rf_pdp <- mat.or.vec(8, 12)

perm_train <- mat.or.vec(8, 12)
drop_train <- mat.or.vec(8, 12)

perm_valid <- mat.or.vec(8, 12)
drop_valid <- mat.or.vec(8, 12)

mrep <- 20
n_size = 1000
set.seed(123)

for (j in seq_len(mrep)) {

  sig <- diag(1, 12, 12)

  for (ii in 1:4) {
    for (jj in 1:4) {
      sig[ii, jj] <- ifelse(ii == jj, 1, 0.95)
    }
  }

  strobl <- MASS::mvrnorm(n_size, mu = rep(0, 12), Sigma = sig)

  y <- 5 * strobl[, 1] + 5 * strobl[, 2] + 2 * strobl[, 3] +
    5 * strobl[, 5] + 5 * strobl[, 6] + 2 * strobl[, 7] +
    rnorm(n_size, mean = 0, sd = 1)
  strobl <- data.frame(cbind(strobl, y))
}
```

```

dfv <- MASS::mvrnorm(n_size, mu = rep(0, 12), Sigma = sig)
y <- 5 * dfv[, 1] + 5 * dfv[, 2] + 2 * dfv[, 3] +
    5 * dfv[, 5] + 5 * dfv[, 6] + 2 * dfv[, 7] +
    rnorm(n_size, mean = 0, sd = 1)
dfv <- data.frame(cbind(dfv, y))

for (k in seq_len(12)) {
  r <- randomForest(y ~ ., data = strobl, mtry = k,
                    importance = T)

  impt <- sqrt(as.data.frame(pmax(randomForest::importance(r, scale = T), 0)))
  impt <- impt$`%IncMSE`[1:8]

  impf <- sqrt(as.data.frame(pmax(randomForest::importance(r, scale = F), 0)))
  impf <- impf$`%IncMSE`[1:8]

  # vimp = pdp_compare(r, var_vec = 1:8, trellis = F)
  # impv = vimp$imp[c(1, 4)] %>% arrange(var) %>% pull(sd)

  # vimp = vip::vi_firm(r, train = strobl)
  # impv <- vimp$Importance[1:8]

  p <- predict(r, strobl)
  m = mean((p-strobl$y)^2)

  rq = r$rsq[500]

  vp <- predict(r, dfv)
  mv = mean((vp-dfv$y)^2)

  perm_impr <- vector(length = 8)
  perm_impv <- vector(length = 8)
  drop_impr <- vector(length = 8)
  drop_impv <- vector(length = 8)

  for (i in seq_len(8)) {
    df_new <- strobl
    df_new[i] <- df_new[sample(1:n_size), i]

    p <- predict(r, df_new)
    new_m = mean((p-strobl$y)^2)
    perm_impr[i] <- new_m - m

    #####

    v_new <- dfv
    v_new[i] <- v_new[sample(1:n_size), i]

    vp <- predict(r, v_new)
    new_vm = mean((vp-dfv$y)^2)
    perm_impv[i] <- new_vm - mv

    #####
  }
}

```

```

df_new <- strobl
df_new[, i] <- 0

p <- predict(r, df_new)
new_m = mean((p-strobl$y)^2)
drop_impr[i] <- new_m - m

#####

v_new <- dfv
v_new[, i] <- 0

vp <- predict(r, v_new)
new_vm = mean((vp-dfv$y)^2)
drop_impv[i] <- new_vm - mv
}

rf_oob_t[,k] <- rf_oob_t[,k] + impt / mrep
rf_oob_f[,k] <- rf_oob_f[,k] + impf / mrep

#rf_pdp[,k] <- rf_pdp[,k] + impp / mrep

rsq[k] <- rsq[k] + rq / mrep

simpr <- sqrt(pmax(perm_impr, 0))
perm_train[,k] <- perm_train[,k] + simpr / mrep

simpv <- sqrt(pmax(perm_impv, 0))
perm_valid[,k] <- perm_valid[,k] + simpv / mrep

dsimpr <- sqrt(pmax(drop_impr, 0))
drop_train[,k] <- drop_train[,k] + dsimpr / mrep

dsimpv <- sqrt(pmax(drop_impv, 0))
drop_valid[,k] <- drop_valid[,k] + dsimpv / mrep
}
}

```

```

for (i in seq_len(12)){
  sdf <- data.frame(coef = c(5, 5, 2, 0, 5, 5, 2, 0),
    rf_oob_t = rf_oob_t[,i],
    rf_oob_f = rf_oob_f[,i],
    #rf_pdp = rf_pdp[,i],
    perm_train = perm_train[,i],
    drop_train = drop_train[,i],
    perm_valid = perm_valid[,i],
    drop_valid = drop_valid[,i])

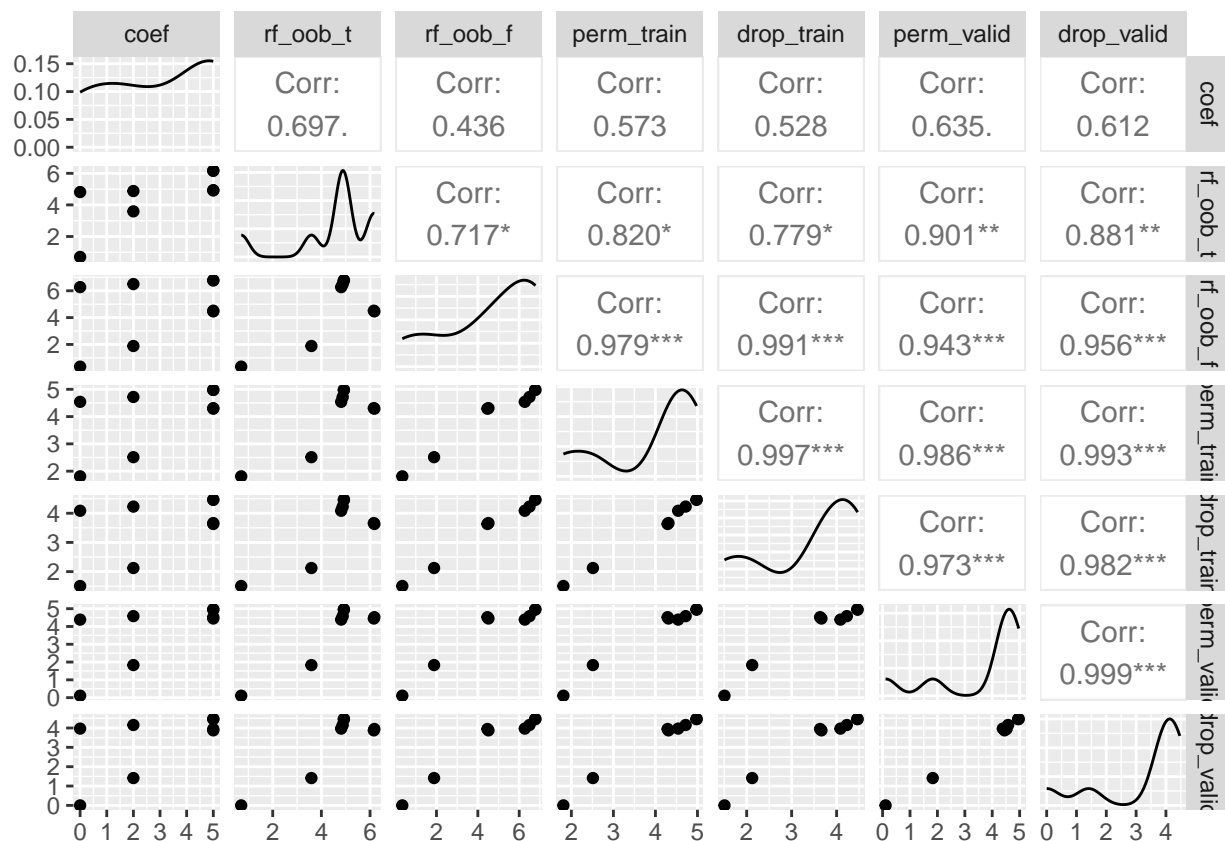
  print(sdf)
  print(ggpairs(sdf))

  sdf <- sdf %>% select(coef, rf_oob_f, #rf_pdp,
    perm_train, perm_valid)

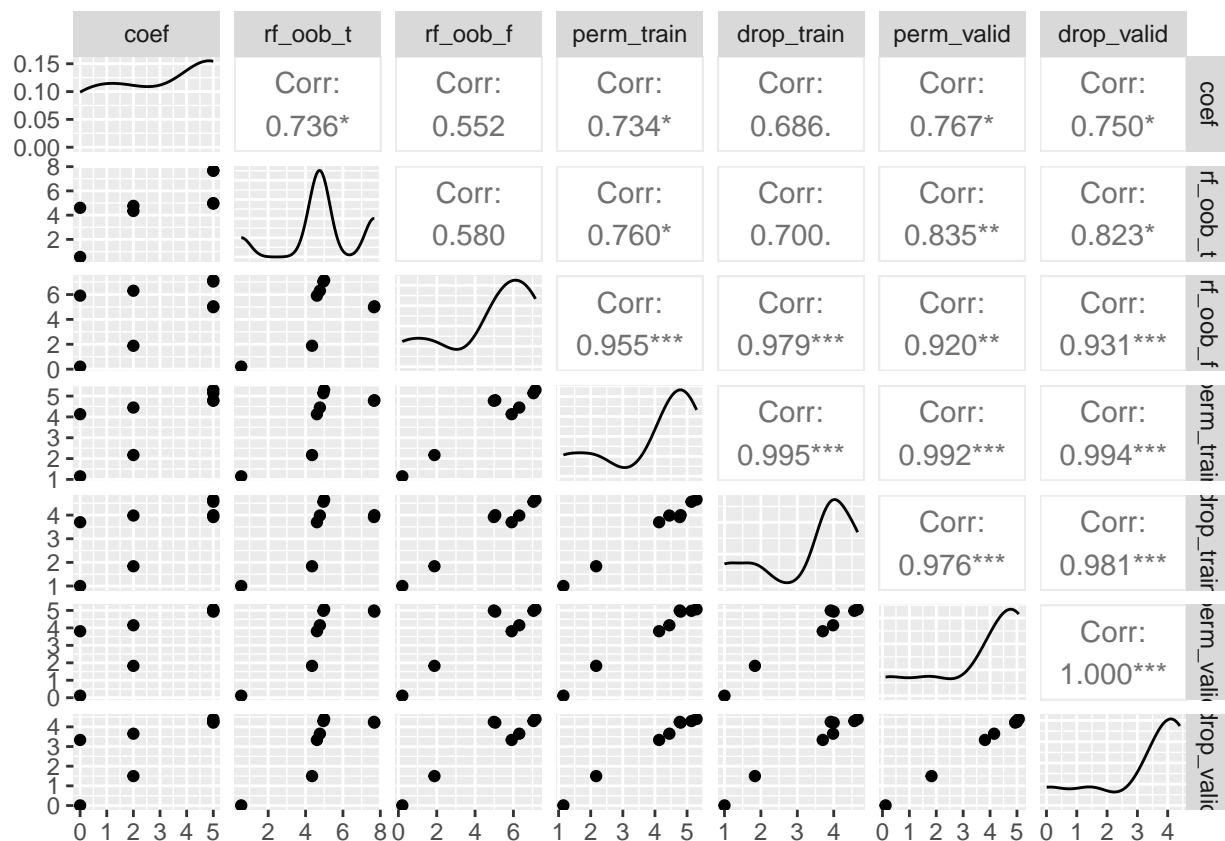
  print(sdf)
  print(ggpairs(sdf))
}

```

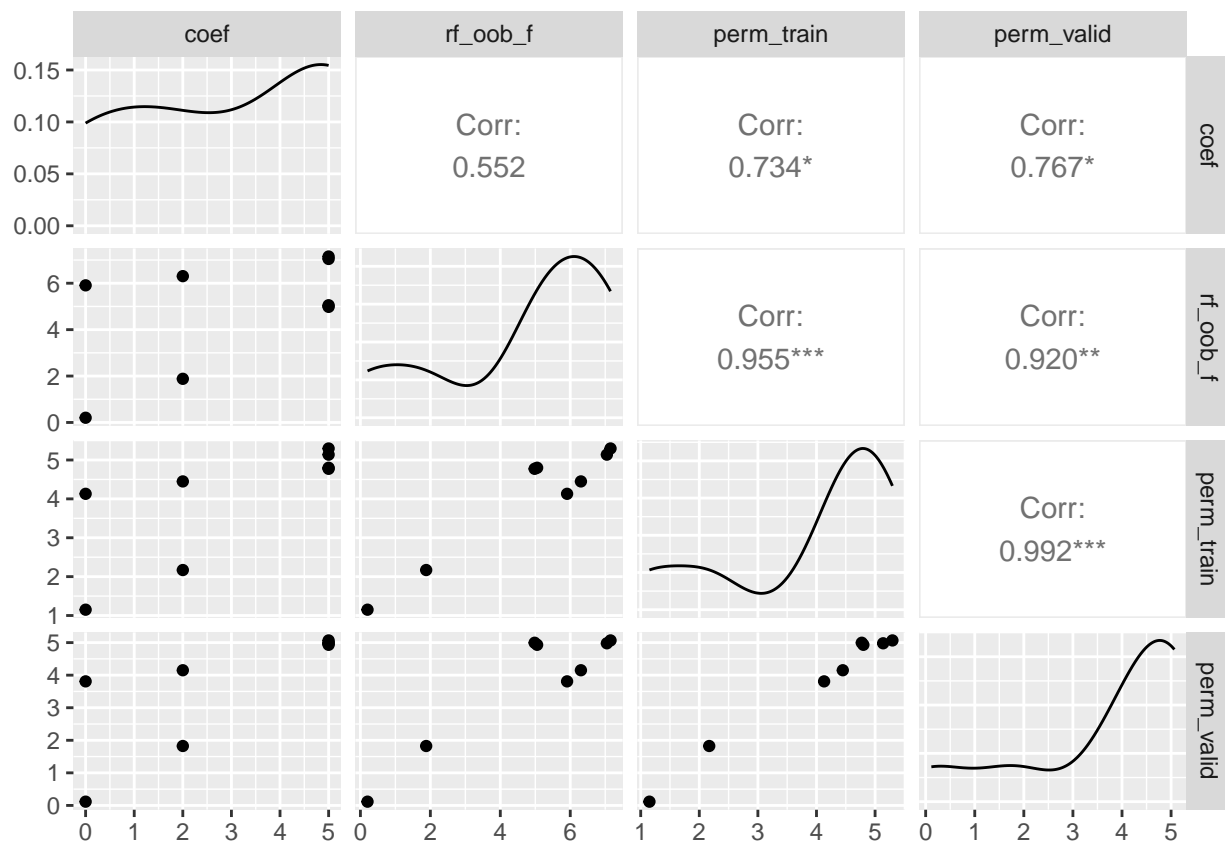
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	4.9232047	6.7867518	4.978387	4.458614	4.9673050	4.466013
## 2	5	4.9282341	6.7699025	4.978248	4.468112	4.9445419	4.458731
## 3	2	4.8775348	6.4991365	4.718588	4.227459	4.5804960	4.154014
## 4	0	4.8145681	6.2707637	4.540875	4.085461	4.3866911	3.968607
## 5	5	6.1561843	4.5130184	4.309551	3.663021	4.4489223	3.876903
## 6	5	6.1789173	4.4629638	4.290063	3.632390	4.5227268	3.948305
## 7	2	3.5906952	1.8829693	2.512124	2.122503	1.8299843	1.413030
## 8	0	0.7047689	0.3459405	1.810456	1.506748	0.1083828	0.000000



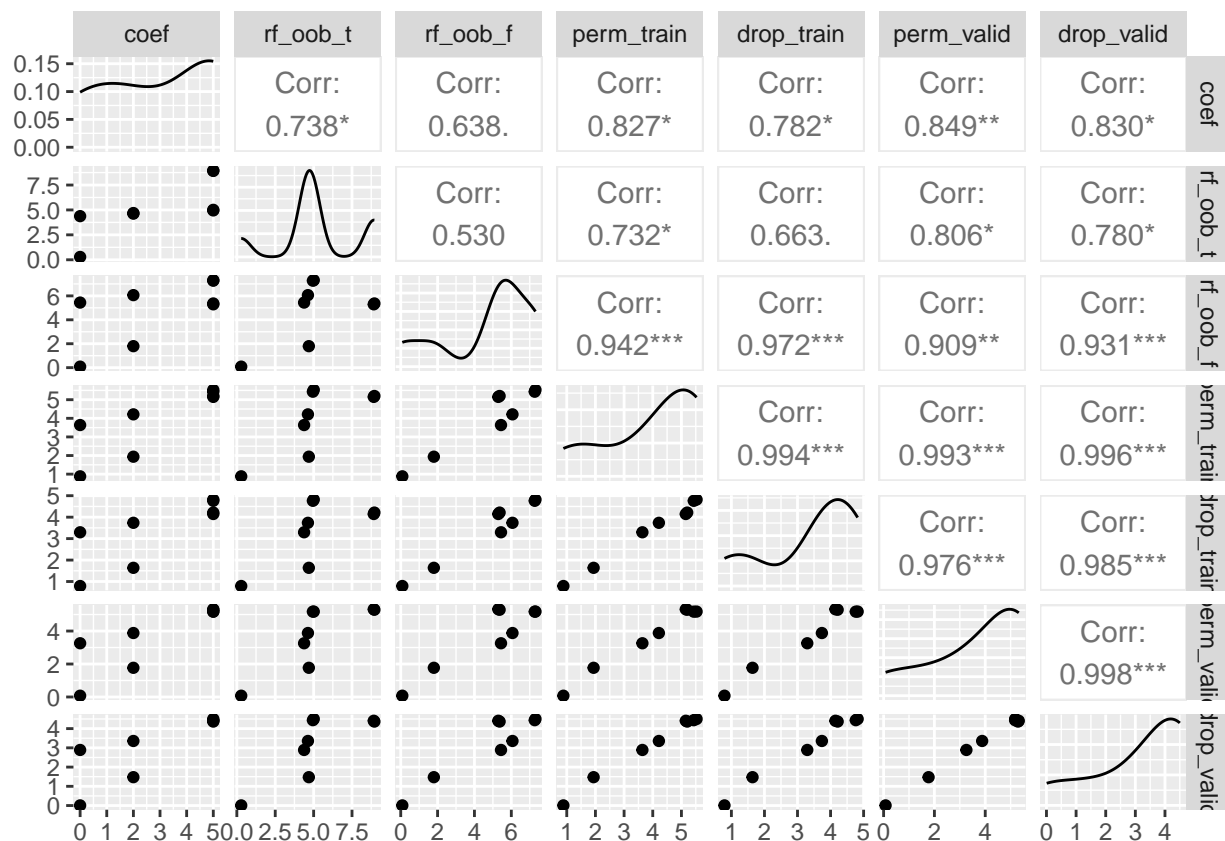
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  6.7867518   4.978387  4.9673050
## 2      5  6.7699025   4.978248  4.9445419
## 3      2  6.4991365   4.718588  4.5804960
## 4      0  6.2707637   4.540875  4.3866911
## 5      5  4.5130184   4.309551  4.4489223
## 6      5  4.4629638   4.290063  4.5227268
## 7      2  1.8829693   2.512124  1.8299843
## 8      0  0.3459405   1.810456  0.1083828
```

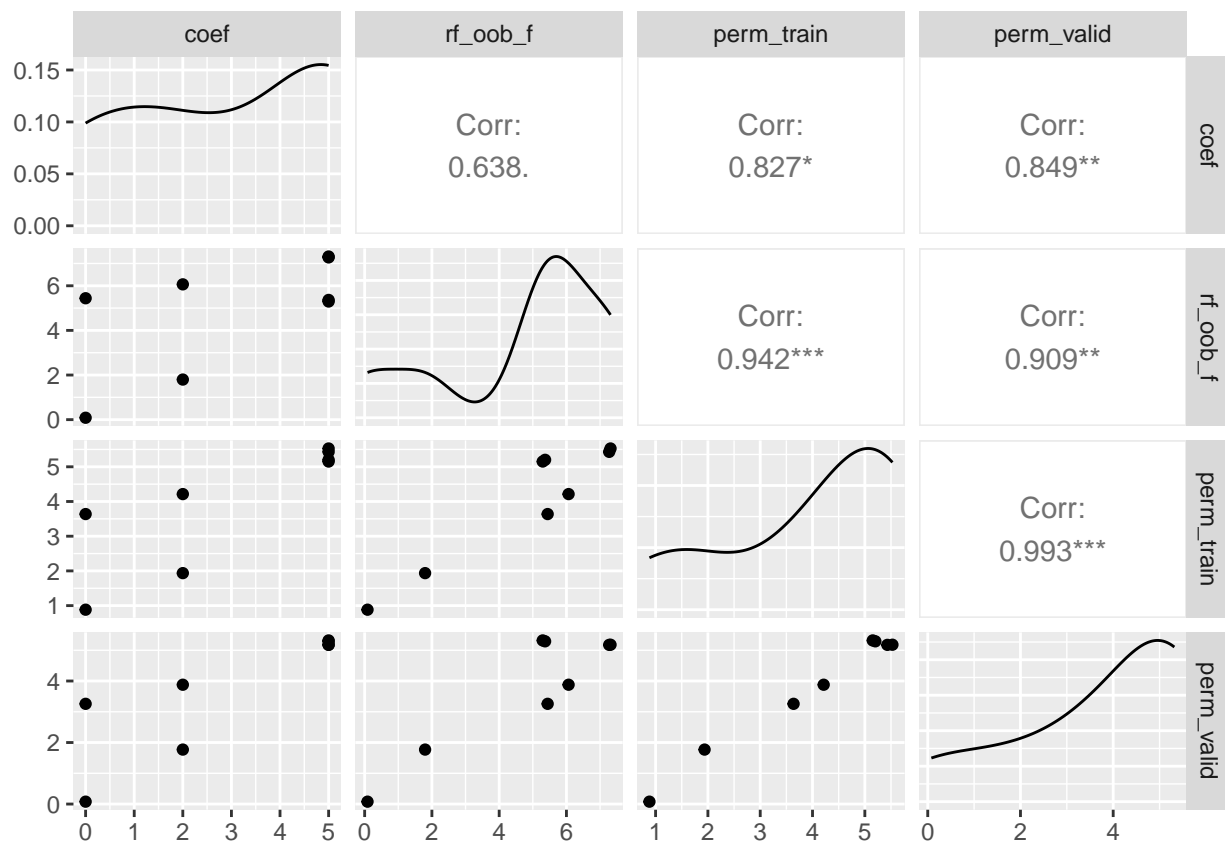
```
## coef rf_oob_f perm_train perm_valid
## 1 5 7.1513545 5.296275 5.0665306
## 2 5 7.0473135 5.137815 4.9787830
## 3 2 6.3051554 4.449097 4.1507461
## 4 0 5.9093782 4.130213 3.8098398
## 5 5 5.0524243 4.800096 4.9326233
## 6 5 4.9815931 4.773114 4.9950902
## 7 2 1.8808532 2.169620 1.8249657
## 8 0 0.2034577 1.149218 0.1156092
```



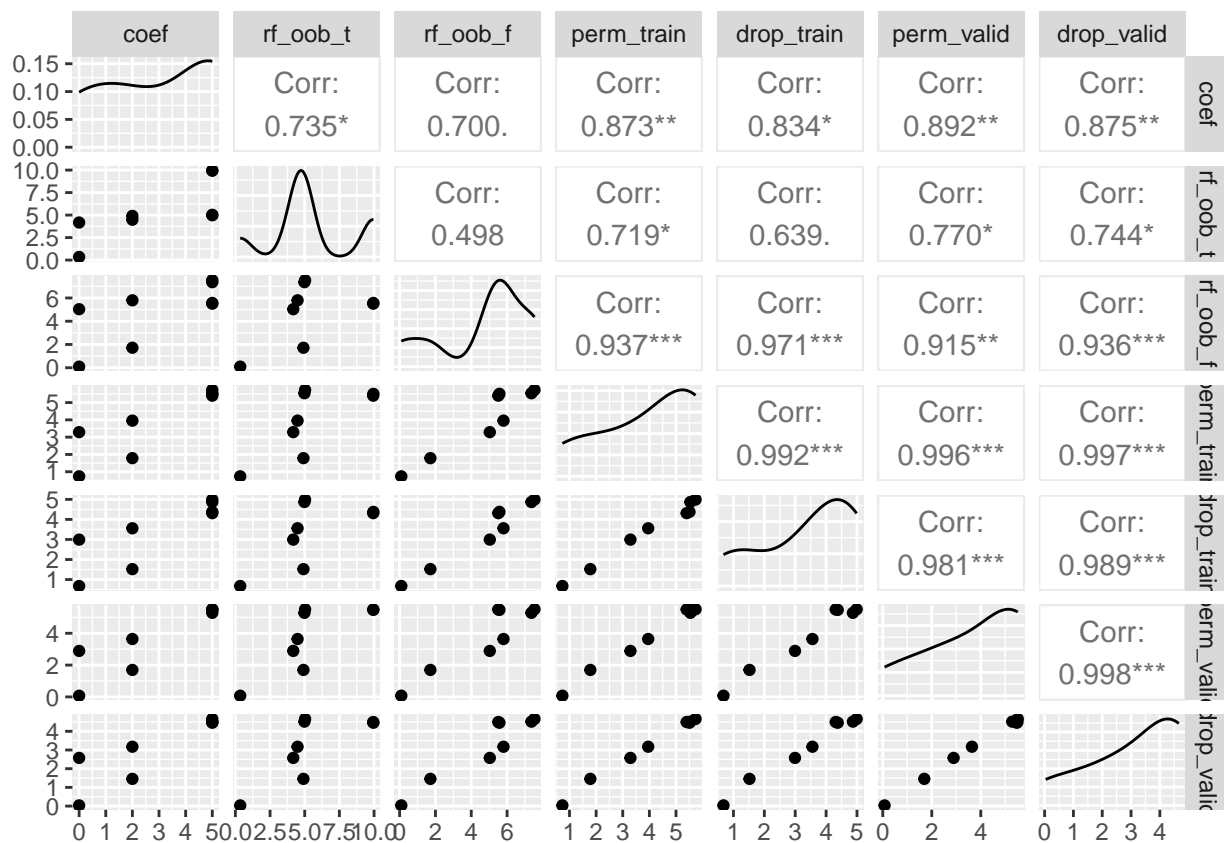
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.0088119	7.31104074	5.5276229	4.8194631	5.17986676	4.501597502
## 2	5	4.9528212	7.27169130	5.4306932	4.7626447	5.17481926	4.438836039
## 3	2	4.6293417	6.06361193	4.2131719	3.7374639	3.87960121	3.357840556
## 4	0	4.3793144	5.44118131	3.6378959	3.2952634	3.25844434	2.890165041
## 5	5	8.9565273	5.36503129	5.2010516	4.2142850	5.29026298	4.367322437
## 6	5	8.9156177	5.29500408	5.1512369	4.1466929	5.31915256	4.407176568
## 7	2	4.6893794	1.79739246	1.9364216	1.6388403	1.77082453	1.475175956
## 8	0	0.2841558	0.08550787	0.8823575	0.7906776	0.07791576	0.007453621



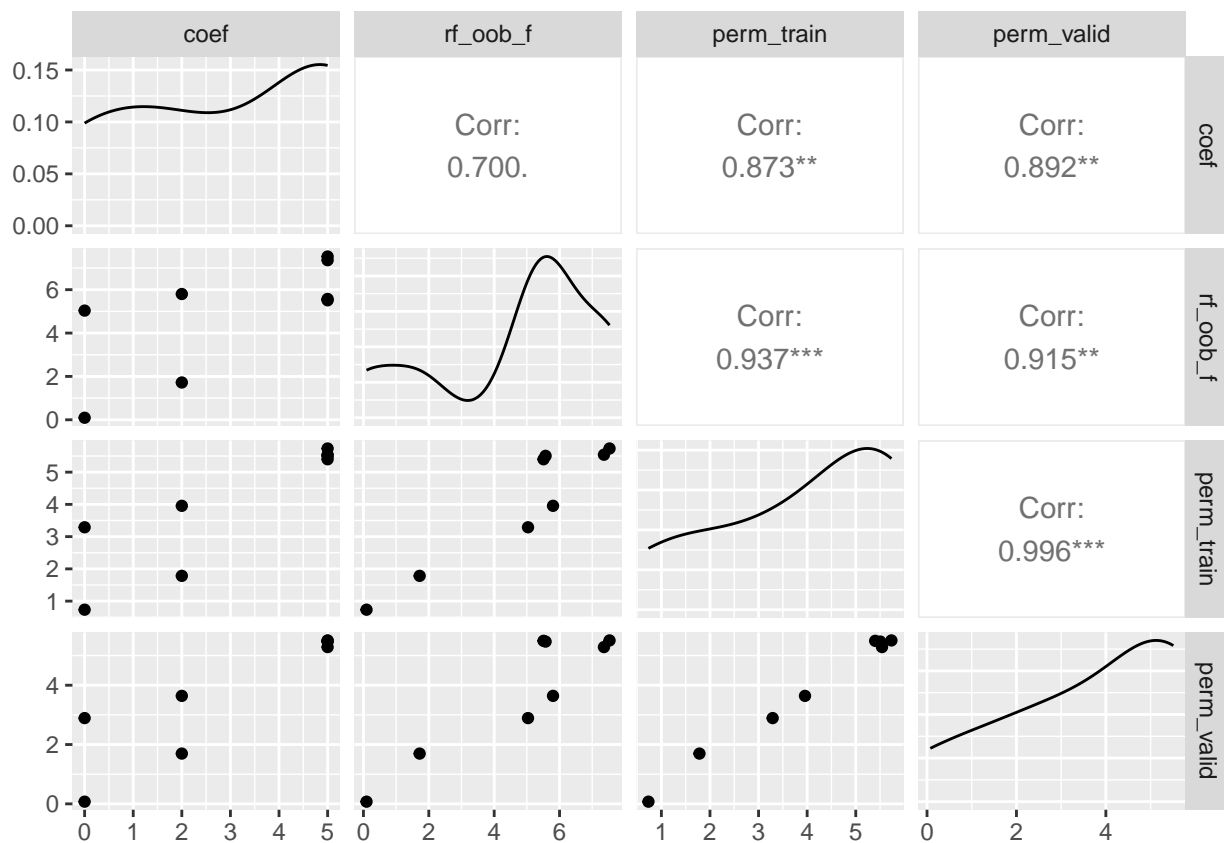
```
##      coef      rf_oob_f perm_train perm_valid
## 1      5 7.31104074  5.5276229 5.17986676
## 2      5 7.27169130  5.4306932 5.17481926
## 3      2 6.06361193  4.2131719 3.87960121
## 4      0 5.44118131  3.6378959 3.25844434
## 5      5 5.36503129  5.2010516 5.29026298
## 6      5 5.29500408  5.1512369 5.31915256
## 7      2 1.79739246  1.9364216 1.77082453
## 8      0 0.08550787  0.8823575 0.07791576
```



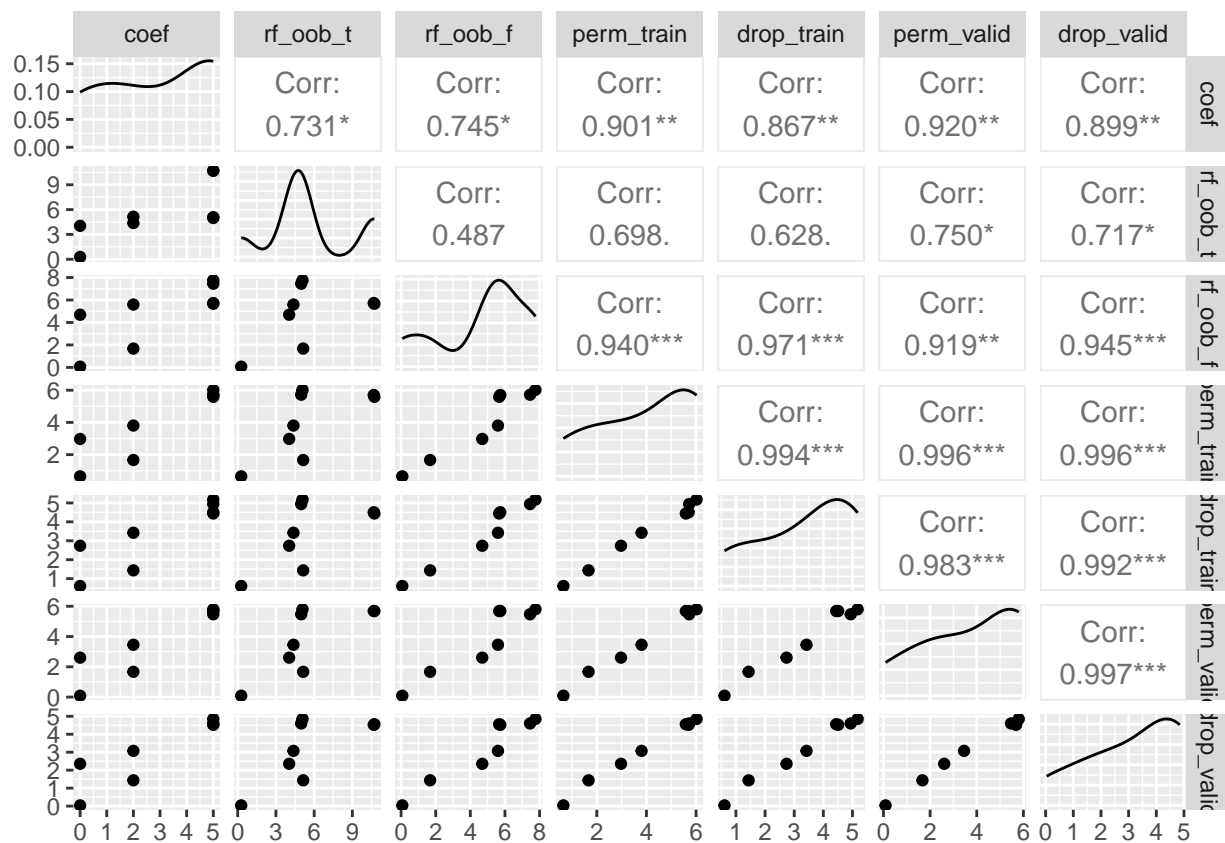
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.026444	7.52920570	5.7344680	4.9896816	5.50951395	4.6595373
## 2	5	4.980408	7.35920498	5.5405138	4.8686386	5.28551219	4.5173247
## 3	2	4.484315	5.80034654	3.9555255	3.5581697	3.64068490	3.1745422
## 4	0	4.174226	5.03531000	3.2899097	2.9929565	2.89139557	2.5765828
## 5	5	9.951880	5.57295915	5.5057166	4.3771017	5.47095042	4.4581379
## 6	5	9.938897	5.51043914	5.4000131	4.3121305	5.49637301	4.4973423
## 7	2	4.909441	1.71969019	1.7834158	1.5193668	1.69537217	1.4557844
## 8	0	0.348055	0.09458247	0.7335068	0.6772297	0.07430075	0.0367021



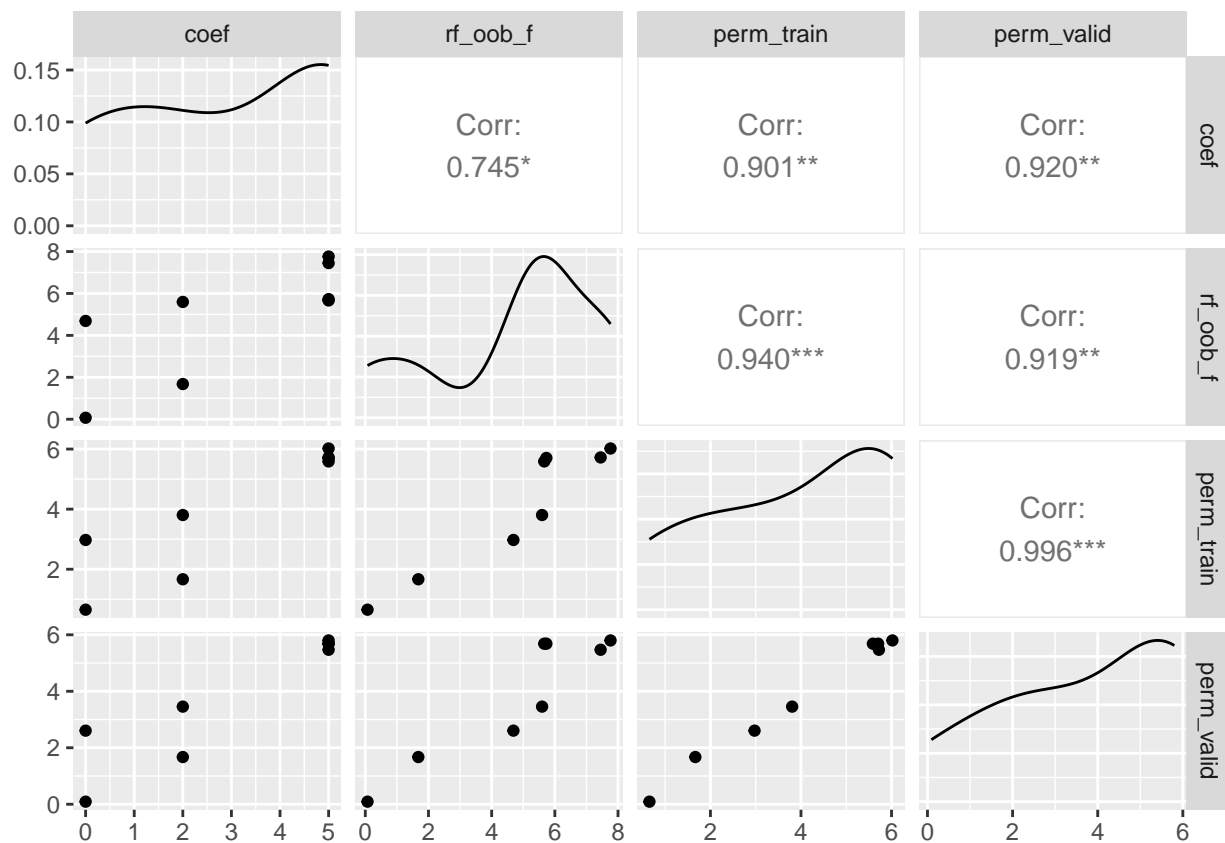
```
##      coef    rf_oob_f perm_train perm_valid
## 1      5 7.52920570  5.7344680 5.50951395
## 2      5 7.35920498  5.5405138 5.28551219
## 3      2 5.80034654  3.9555255 3.64068490
## 4      0 5.03531000  3.2899097 2.89139557
## 5      5 5.57295915  5.5057166 5.47095042
## 6      5 5.51043914  5.4000131 5.49637301
## 7      2 1.71969019  1.7834158 1.69537217
## 8      0 0.09458247  0.7335068 0.07430075
```



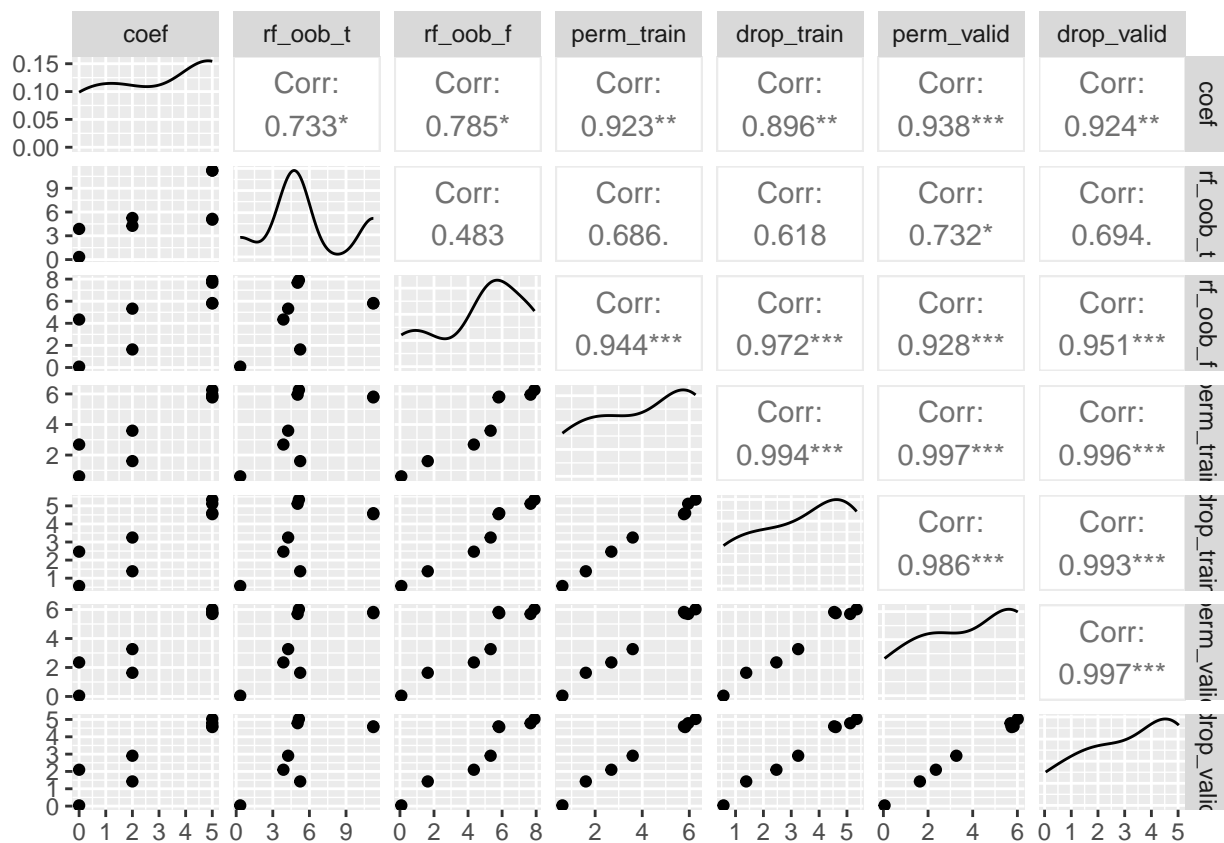
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.0873757	7.76293770	6.019923	5.1767055	5.80056173	4.85111492
## 2	5	4.9830830	7.45039988	5.721002	4.9391532	5.47036298	4.60177102
## 3	2	4.3764783	5.59611918	3.805101	3.4227932	3.45636594	3.07399372
## 4	0	4.0399458	4.68930164	2.974604	2.7373476	2.60575708	2.35382258
## 5	5	10.6781543	5.73101268	5.703247	4.5082562	5.68565086	4.52562349
## 6	5	10.7246838	5.66569073	5.588356	4.4412369	5.68433064	4.55746162
## 7	2	5.1402382	1.67837606	1.666437	1.4357574	1.67100958	1.43522835
## 8	0	0.2770213	0.07032433	0.652883	0.6077716	0.09079098	0.03757979



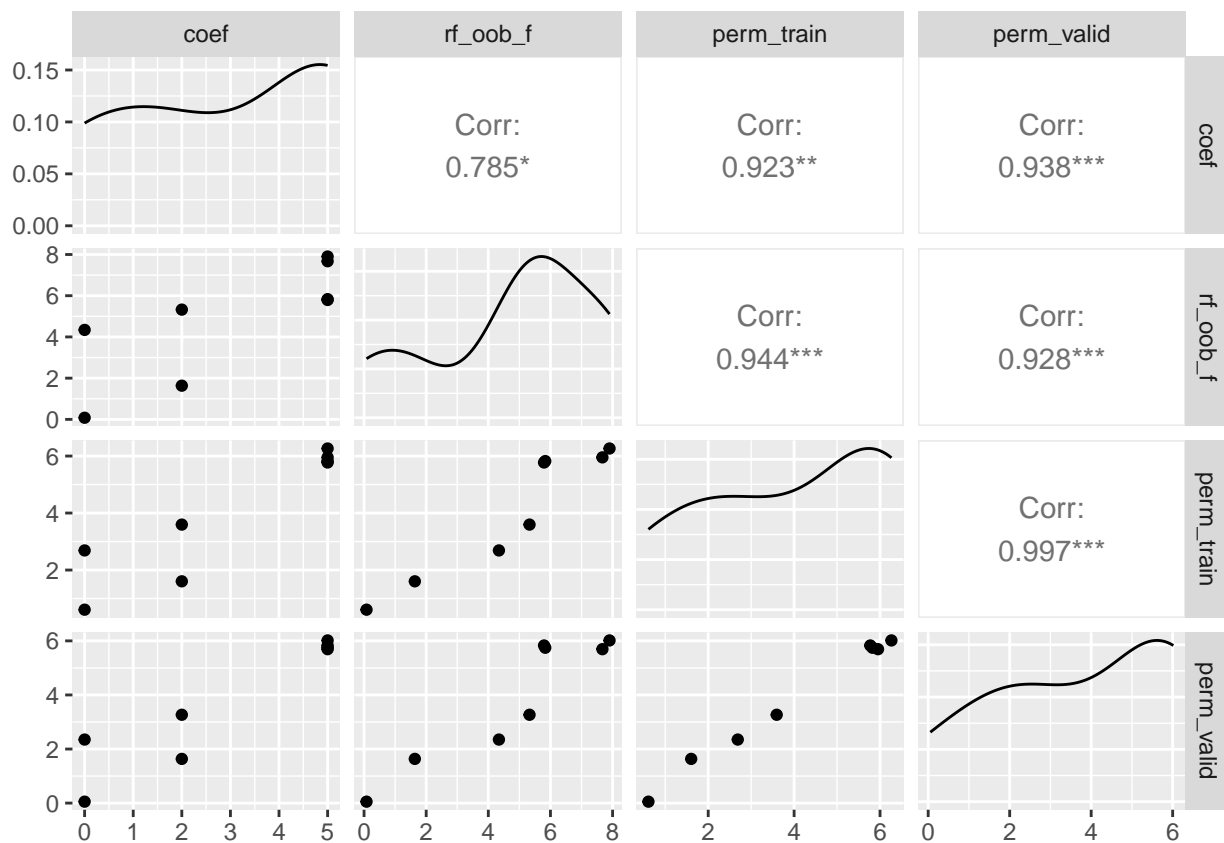
```
##      coef      rf_oob_f perm_train perm_valid
## 1      5 7.76293770    6.019923 5.80056173
## 2      5 7.45039988    5.721002 5.47036298
## 3      2 5.59611918    3.805101 3.45636594
## 4      0 4.68930164    2.974604 2.60575708
## 5      5 5.73101268    5.703247 5.68565086
## 6      5 5.66569073    5.588356 5.68433064
## 7      2 1.67837606    1.666437 1.67100958
## 8      0 0.07032433    0.652883 0.09079098
```



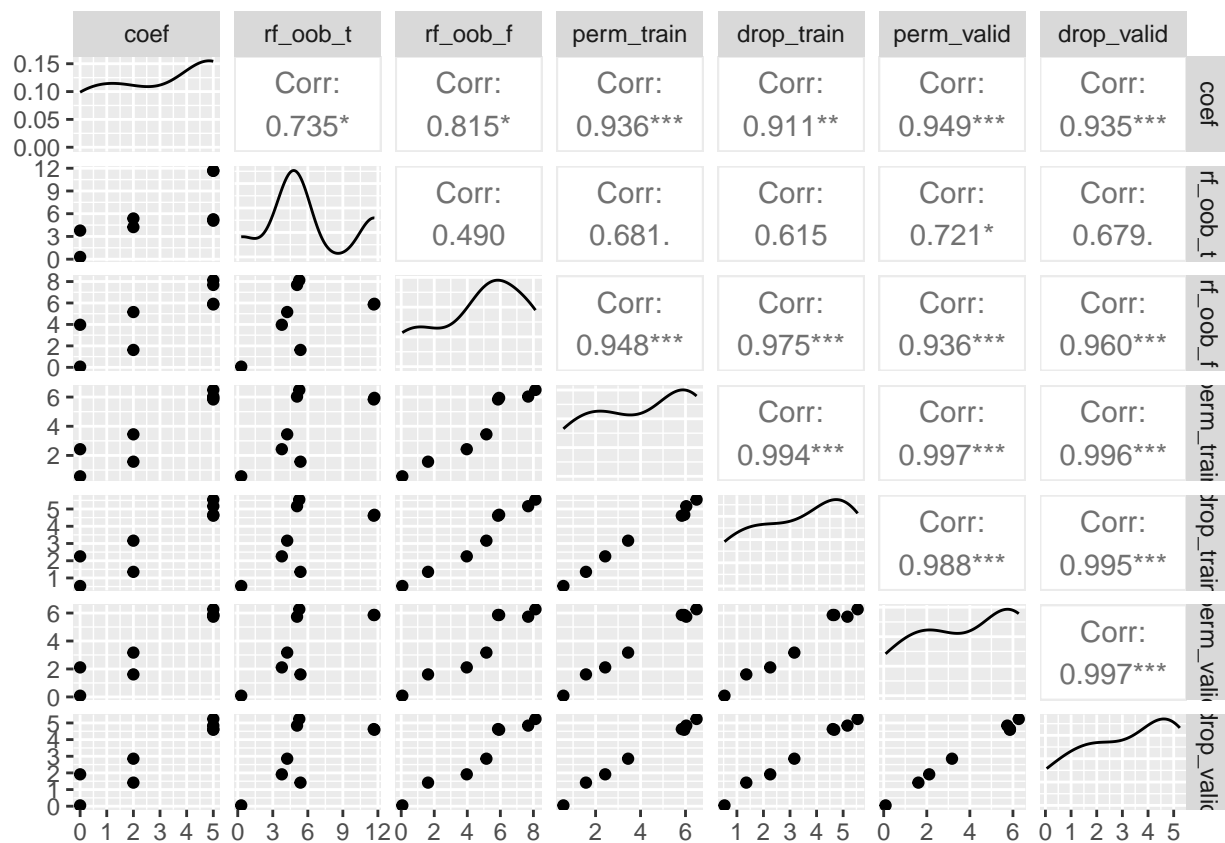
```
##      coef  rf_oob_t  rf_oob_f perm_train drop_train perm_valid drop_valid
## 1      5  5.139695  7.90161197  6.2656488  5.3477905  6.01647030  5.02849097
## 2      5  5.035742  7.67232044  5.9533098  5.1164581  5.69279133  4.78310235
## 3      2  4.257215  5.32639190  3.5948159  3.2488883  3.26736533  2.90578390
## 4      0  3.869313  4.34170737  2.6883788  2.4669876  2.35028501  2.09823500
## 5      5 11.241166  5.83346173  5.8233576  4.5956637  5.74630746  4.56175424
## 6      5 11.233567  5.79418930  5.7737948  4.5409567  5.82982226  4.60213987
## 7      2  5.243890  1.63449771  1.6039763  1.3837933  1.63482904  1.42420635
## 8      0  0.329755  0.07963356  0.6083674  0.5672437  0.05309593  0.04062444
```



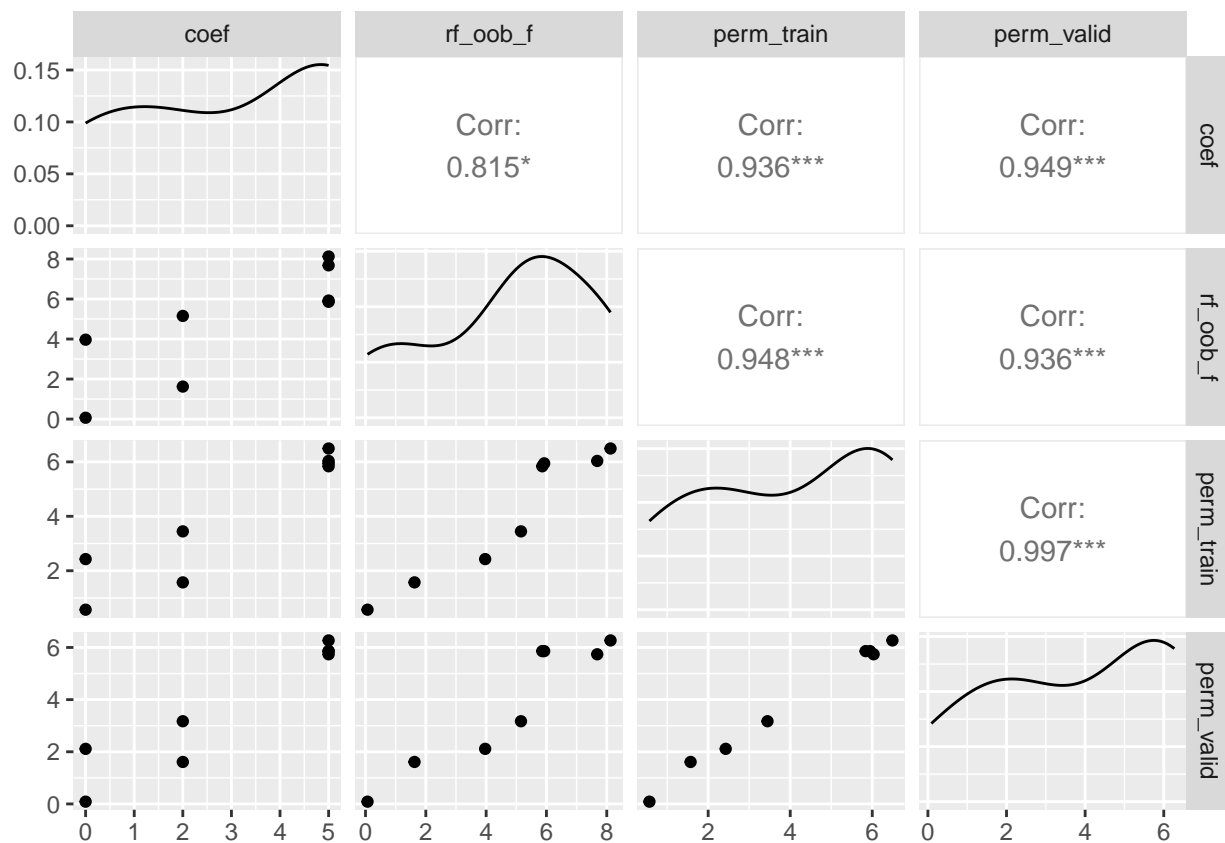
```
##      coef      rf_oob_f perm_train perm_valid
## 1      5 7.90161197  6.2656488 6.01647030
## 2      5 7.67232044  5.9533098 5.69279133
## 3      2 5.32639190  3.5948159 3.26736533
## 4      0 4.34170737  2.6883788 2.35028501
## 5      5 5.83346173  5.8233576 5.74630746
## 6      5 5.79418930  5.7737948 5.82982226
## 7      2 1.63449771  1.6039763 1.63482904
## 8      0 0.07963356  0.6083674 0.05309593
```



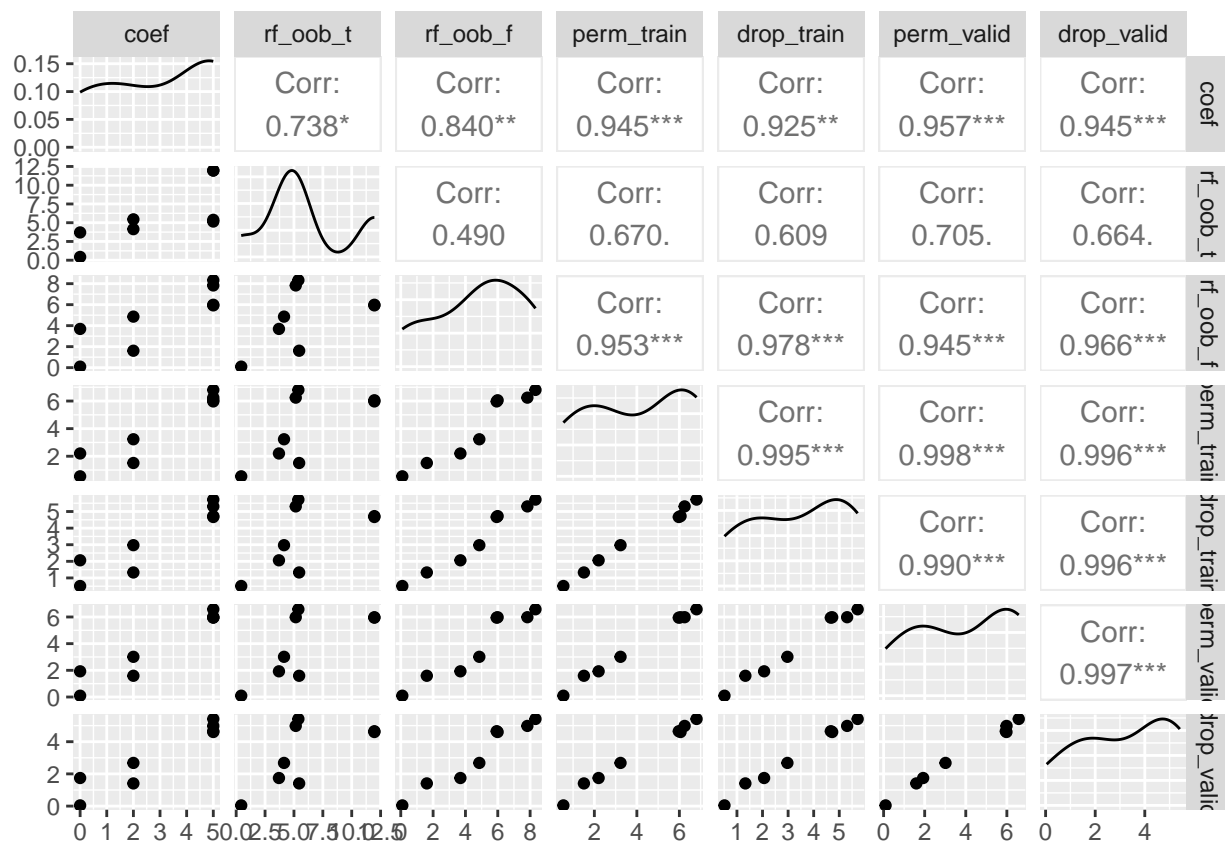
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.2678293	8.12512041	6.4920941	5.5472088	6.26903910	5.23973723
## 2	5	5.0764956	7.68376416	6.0336504	5.1633670	5.73856435	4.84078089
## 3	2	4.2330716	5.15614385	3.4462226	3.1633341	3.17075409	2.85267482
## 4	0	3.7752683	3.96769466	2.4275613	2.2542915	2.11244832	1.91469551
## 5	5	11.6992159	5.92497809	5.9454761	4.6647810	5.85986774	4.59039116
## 6	5	11.6488031	5.86210291	5.8378156	4.6117651	5.85938689	4.63174002
## 7	2	5.3708634	1.62500360	1.5717637	1.3549289	1.61052435	1.41583551
## 8	0	0.2961517	0.06663468	0.5680162	0.5331161	0.08780541	0.04853494



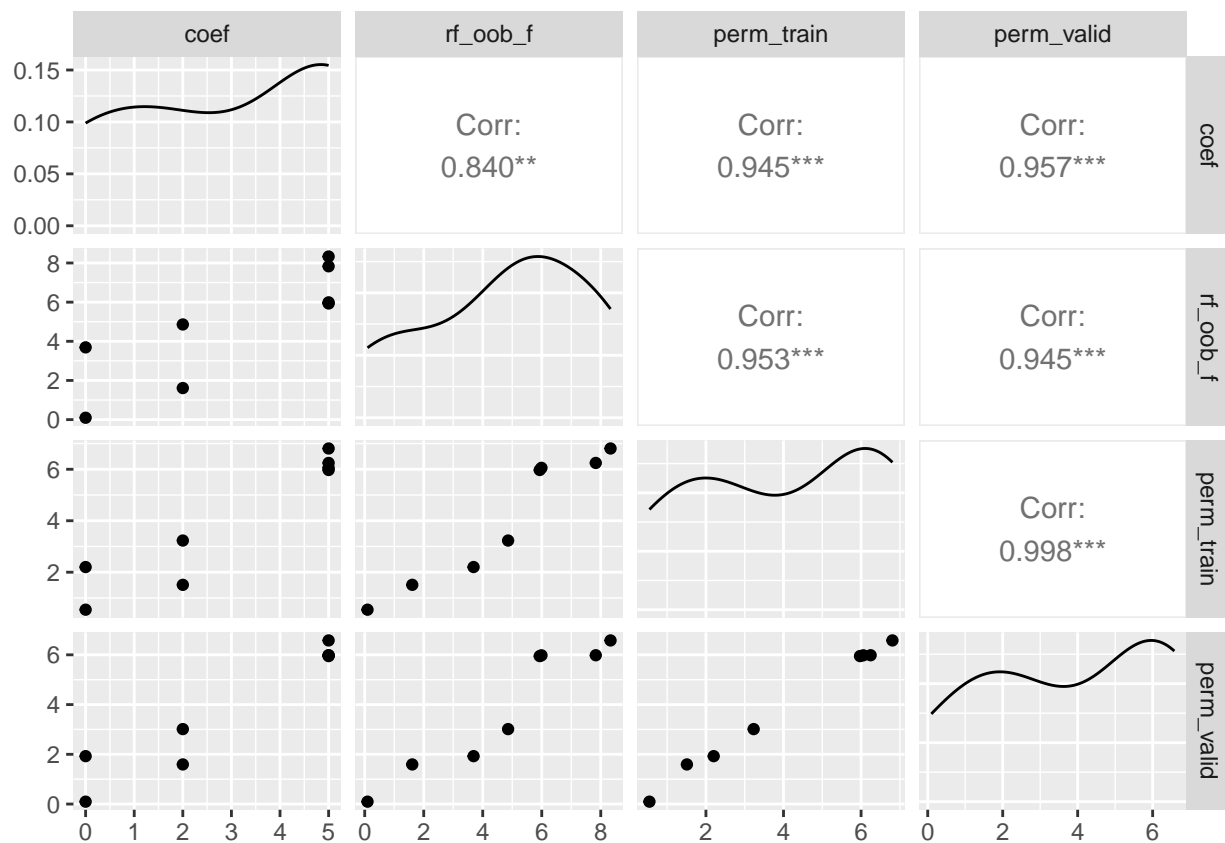
```
##      coef      rf_oob_f perm_train perm_valid
## 1      5 8.12512041  6.4920941 6.26903910
## 2      5 7.68376416  6.0336504 5.73856435
## 3      2 5.15614385  3.4462226 3.17075409
## 4      0 3.96769466  2.4275613 2.11244832
## 5      5 5.92497809  5.9454761 5.85986774
## 6      5 5.86210291  5.8378156 5.85938689
## 7      2 1.62500360  1.5717637 1.61052435
## 8      0 0.06663468  0.5680162 0.08780541
```



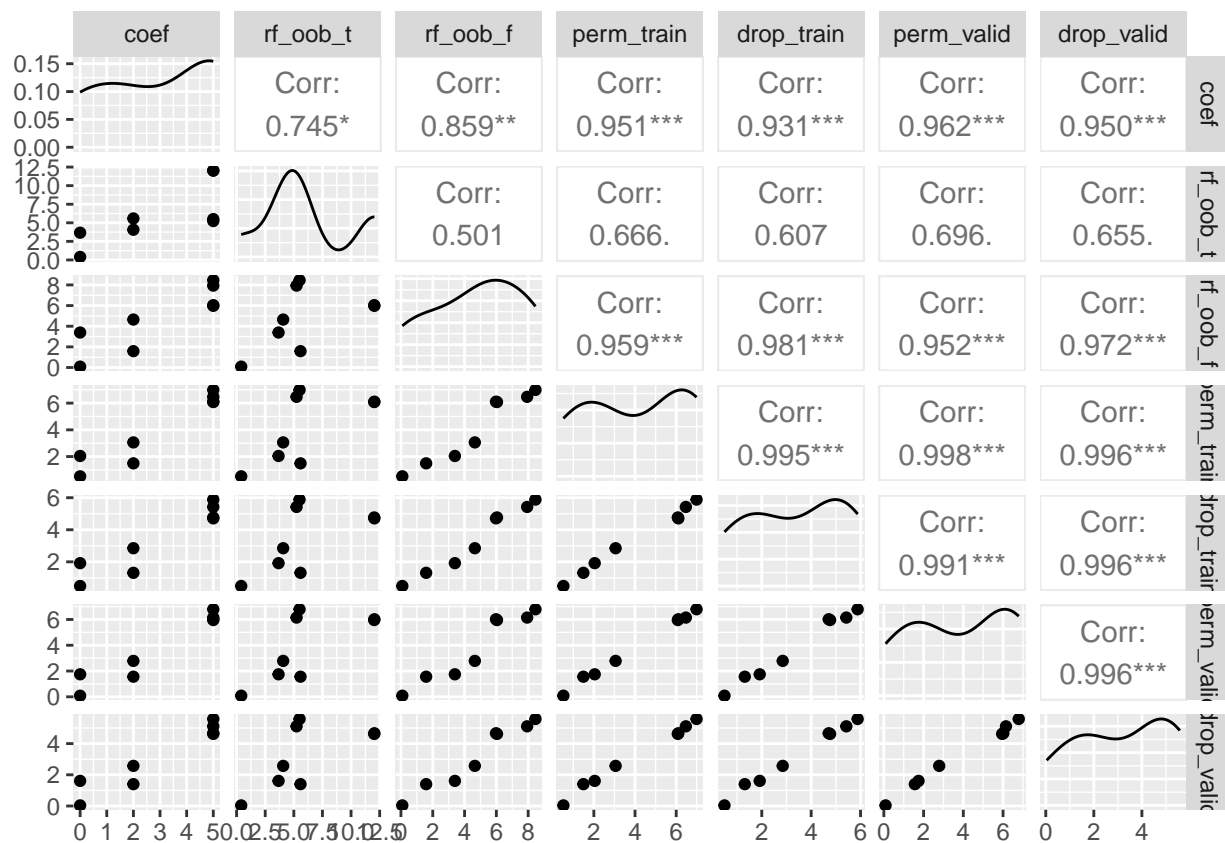
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.383096	8.3306383	6.808319	5.7185074	6.57892628	5.41340947
## 2	5	5.156924	7.8312700	6.246110	5.3073274	5.98383025	4.98152444
## 3	2	4.140259	4.8596433	3.230937	2.9730868	3.01424780	2.68205576
## 4	0	3.703159	3.6894599	2.200514	2.0620397	1.92639613	1.74490239
## 5	5	11.960561	5.9899890	6.055869	4.7224827	5.98024727	4.60345764
## 6	5	11.960913	5.9318768	5.972427	4.6701990	5.94962517	4.65239260
## 7	2	5.449595	1.6116273	1.509395	1.3302293	1.59286828	1.40788185
## 8	0	0.436030	0.0955173	0.543616	0.5131221	0.09653673	0.04298273



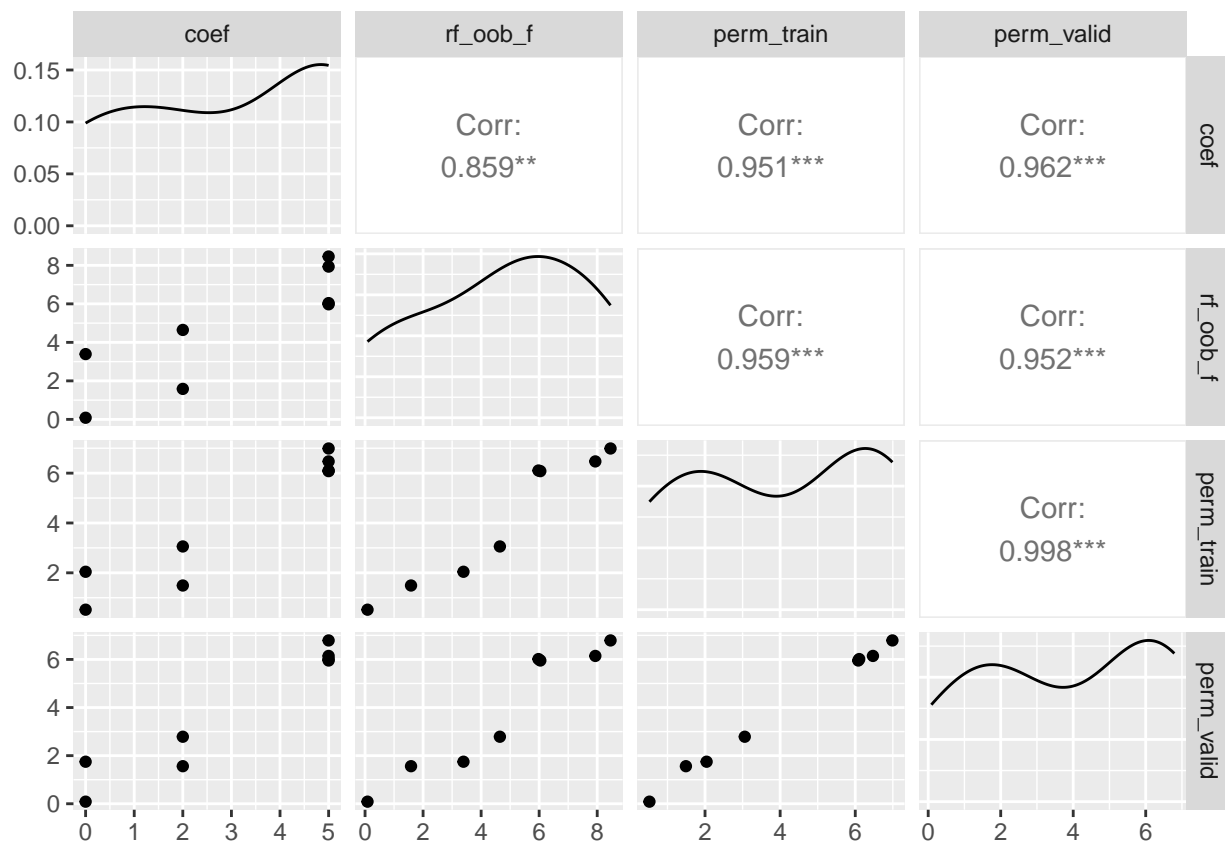
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  8.3306383   6.808319  6.57892628
## 2      5  7.8312700   6.246110  5.98383025
## 3      2  4.8596433   3.230937  3.01424780
## 4      0  3.6894599   2.200514  1.92639613
## 5      5  5.9899890   6.055869  5.98024727
## 6      5  5.9318768   5.972427  5.94962517
## 7      2  1.6116273   1.509395  1.59286828
## 8      0  0.0955173   0.543616  0.09653673
```



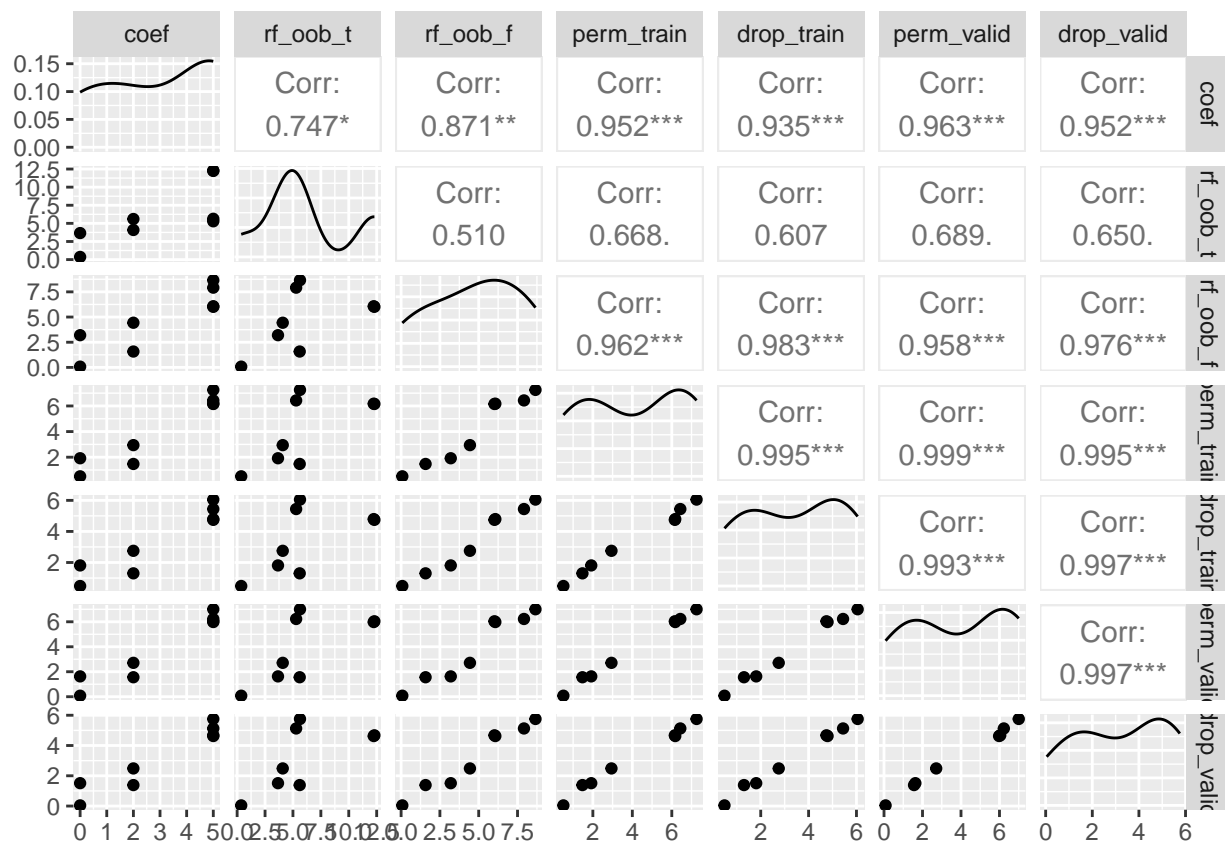
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.5002423	8.45876764	6.9898438	5.8828430	6.78978262	5.5782741
## 2	5	5.2378646	7.93522137	6.4705987	5.4239212	6.14612998	5.1065130
## 3	2	4.0684464	4.64659418	3.0579097	2.8486299	2.78676527	2.5714591
## 4	0	3.6656751	3.39189955	2.0418506	1.9176734	1.74741470	1.6084527
## 5	5	12.0232101	6.03991127	6.0806077	4.7676019	5.95579306	4.6206665
## 6	5	12.0382256	5.97643712	6.1052853	4.7112544	6.01442763	4.6618576
## 7	2	5.5829889	1.58460490	1.4910927	1.3133025	1.56202320	1.3978069
## 8	0	0.4026846	0.08556278	0.5197297	0.4906284	0.08712145	0.0376822



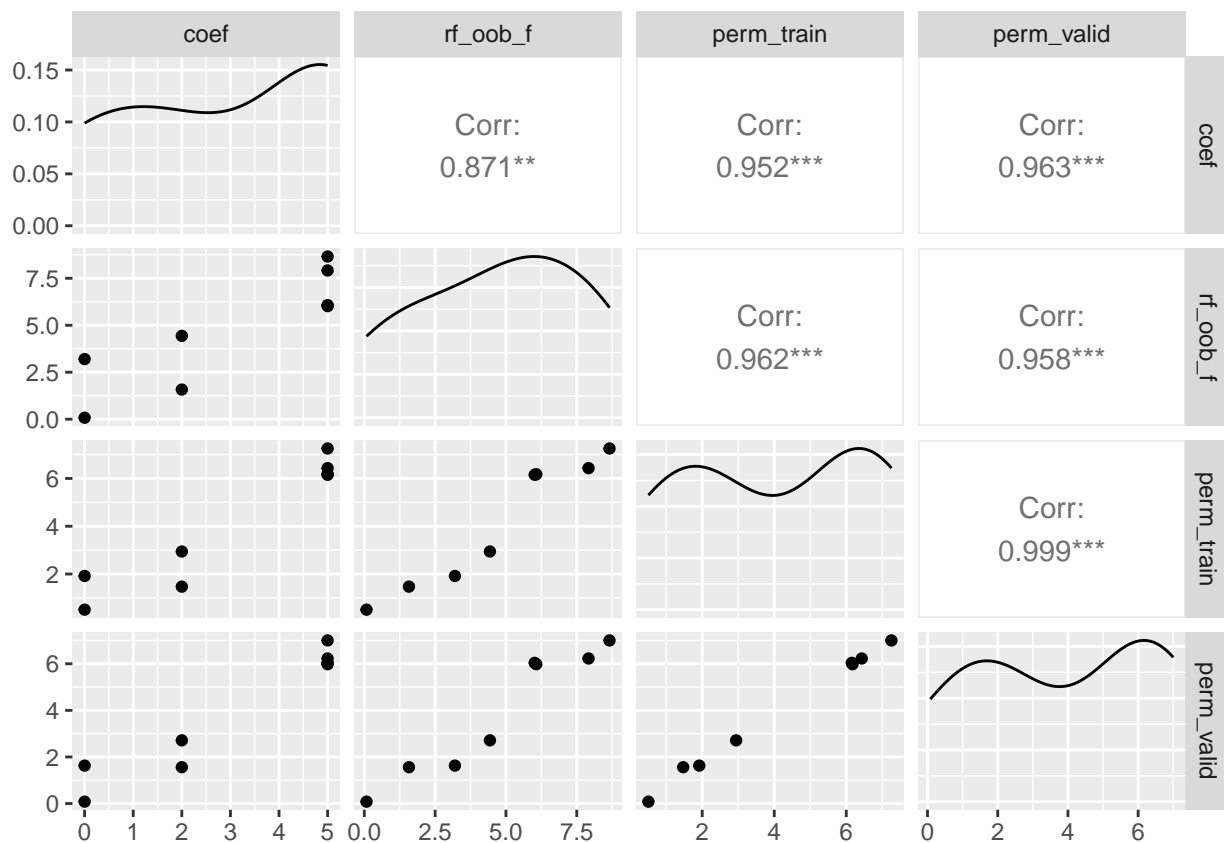
##	coef	rf_oob_f	perm_train	perm_valid
## 1	5	8.45876764	6.9898438	6.78978262
## 2	5	7.93522137	6.4705987	6.14612998
## 3	2	4.64659418	3.0579097	2.78676527
## 4	0	3.39189955	2.0418506	1.74741470
## 5	5	6.03991127	6.0806077	5.95579306
## 6	5	5.97643712	6.1052853	6.01442763
## 7	2	1.58460490	1.4910927	1.56202320
## 8	0	0.08556278	0.5197297	0.08712145



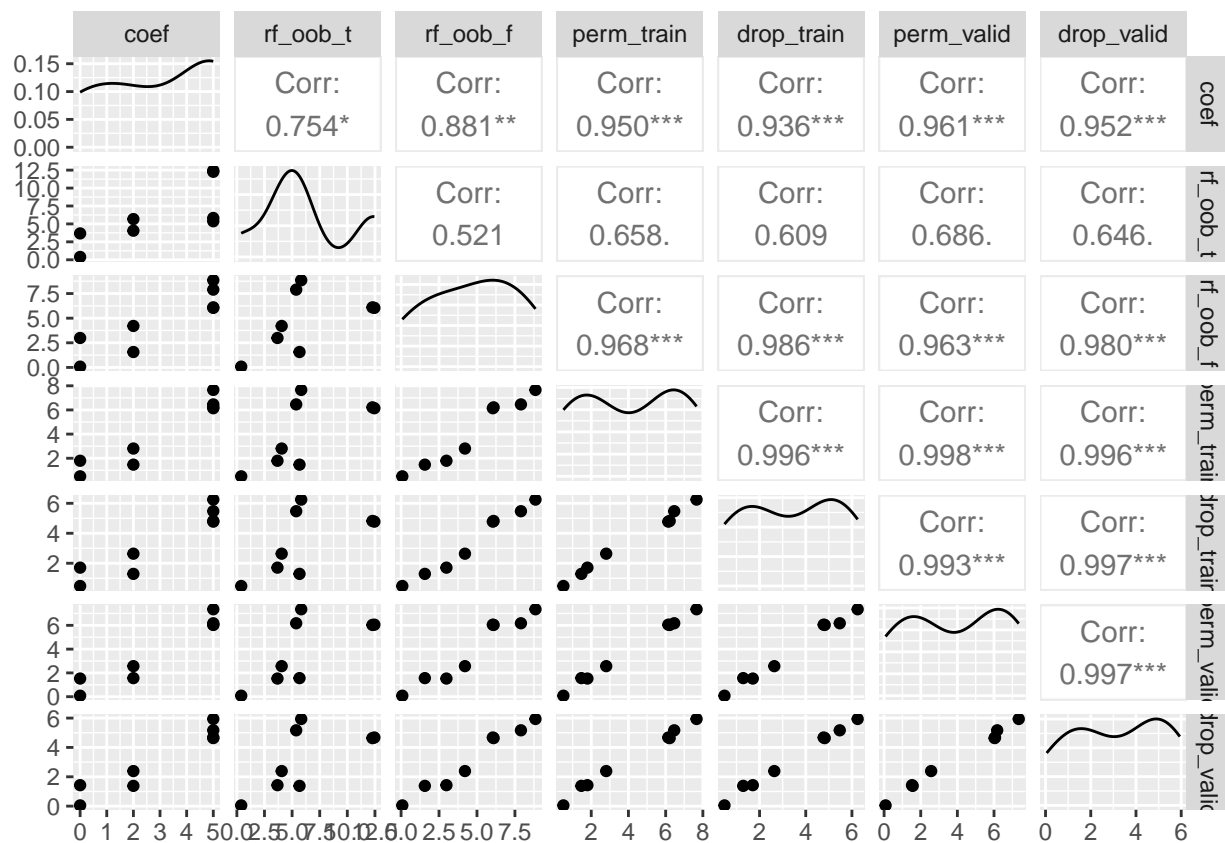
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.6329061	8.65640832	7.254421	6.0497662	7.00188486	5.75526487
## 2	5	5.2860683	7.91592974	6.430818	5.4398668	6.22650734	5.12735342
## 3	2	4.0860122	4.43537386	2.942242	2.7531667	2.71247132	2.48618329
## 4	0	3.6615137	3.19922724	1.918348	1.8114260	1.62891910	1.51080234
## 5	5	12.2376631	6.07148302	6.174279	4.7930632	5.98082293	4.62680220
## 6	5	12.2950496	6.01549839	6.154301	4.7457812	6.04002110	4.67477360
## 7	2	5.6051320	1.57614139	1.471565	1.3017190	1.56016321	1.38568551
## 8	0	0.3657013	0.07786646	0.506296	0.4862462	0.08097994	0.04305275



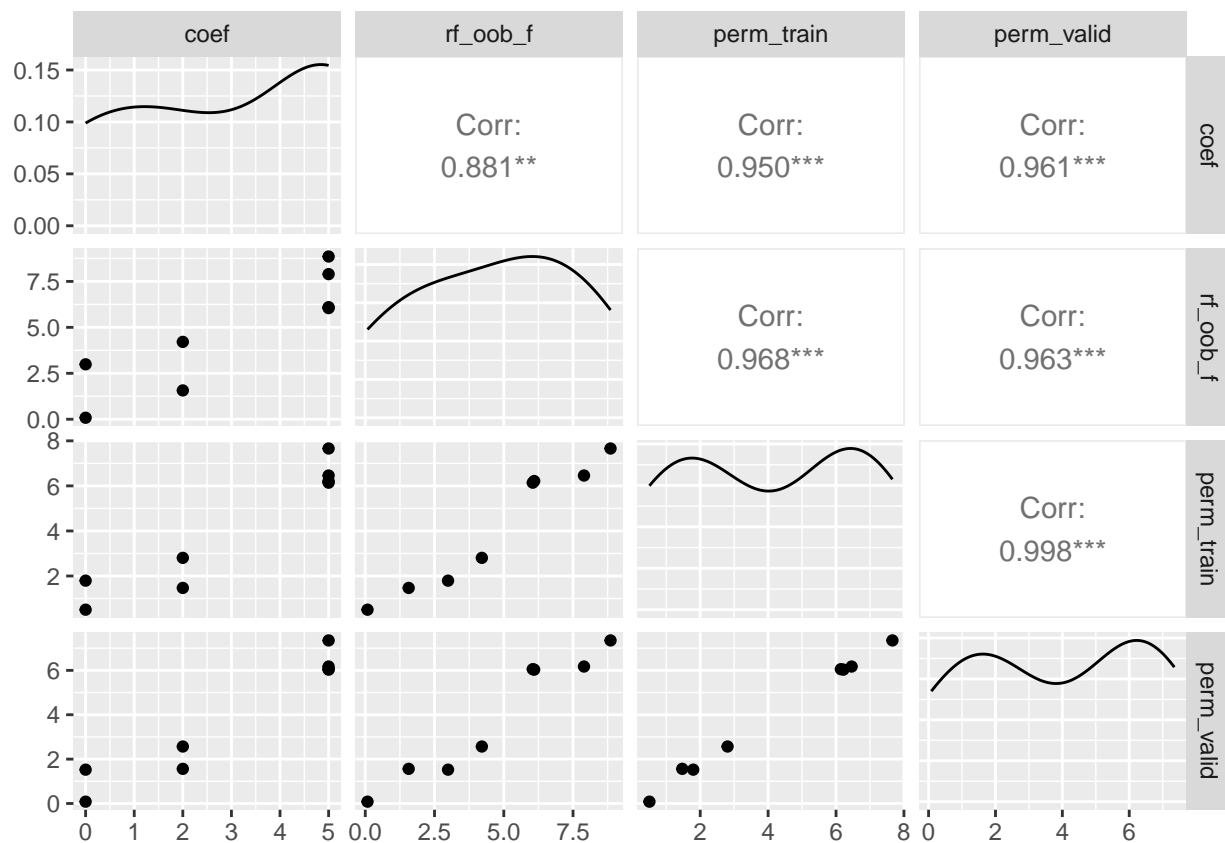
```
##      coef      rf_oob_f perm_train perm_valid
## 1      5 8.65640832    7.254421 7.00188486
## 2      5 7.91592974    6.430818 6.22650734
## 3      2 4.43537386    2.942242 2.71247132
## 4      0 3.19922724    1.918348 1.62891910
## 5      5 6.07148302    6.174279 5.98082293
## 6      5 6.01549839    6.154301 6.04002110
## 7      2 1.57614139    1.471565 1.56016321
## 8      0 0.07786646    0.506296 0.08097994
```



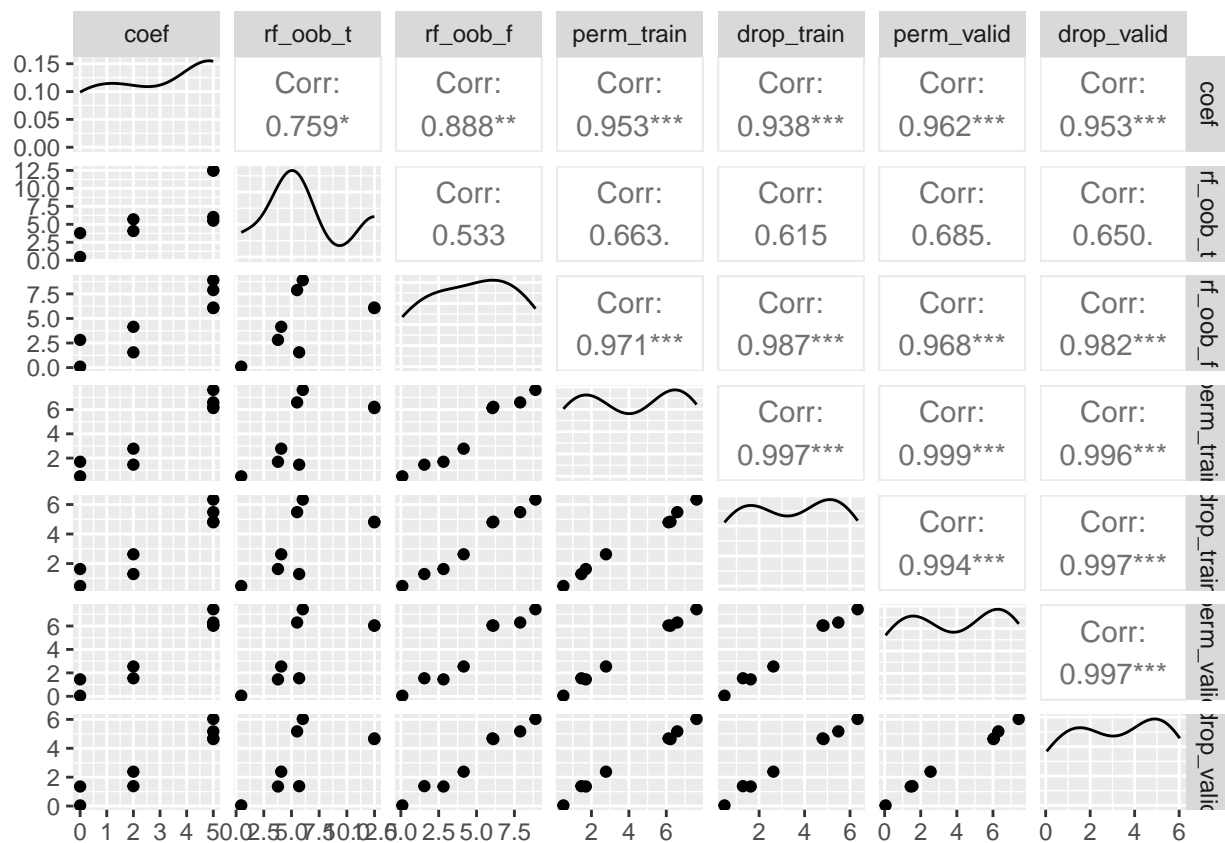
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.8265227	8.85377982	7.661362	6.2476671	7.34928037	5.93942507
## 2	5	5.3706267	7.89579428	6.460074	5.4721967	6.17266630	5.16738751
## 3	2	4.0582653	4.20881295	2.805603	2.6375272	2.56785404	2.38883621
## 4	0	3.6755401	2.98832197	1.794127	1.7072198	1.52446979	1.42699263
## 5	5	12.2735230	6.09779102	6.216068	4.8208116	6.03255609	4.63550670
## 6	5	12.4483873	6.04642915	6.142736	4.7755532	6.05380372	4.67593268
## 7	2	5.6725031	1.56704361	1.470628	1.2917256	1.56144818	1.38103770
## 8	0	0.3946355	0.08126967	0.503865	0.4808615	0.08193564	0.05024782



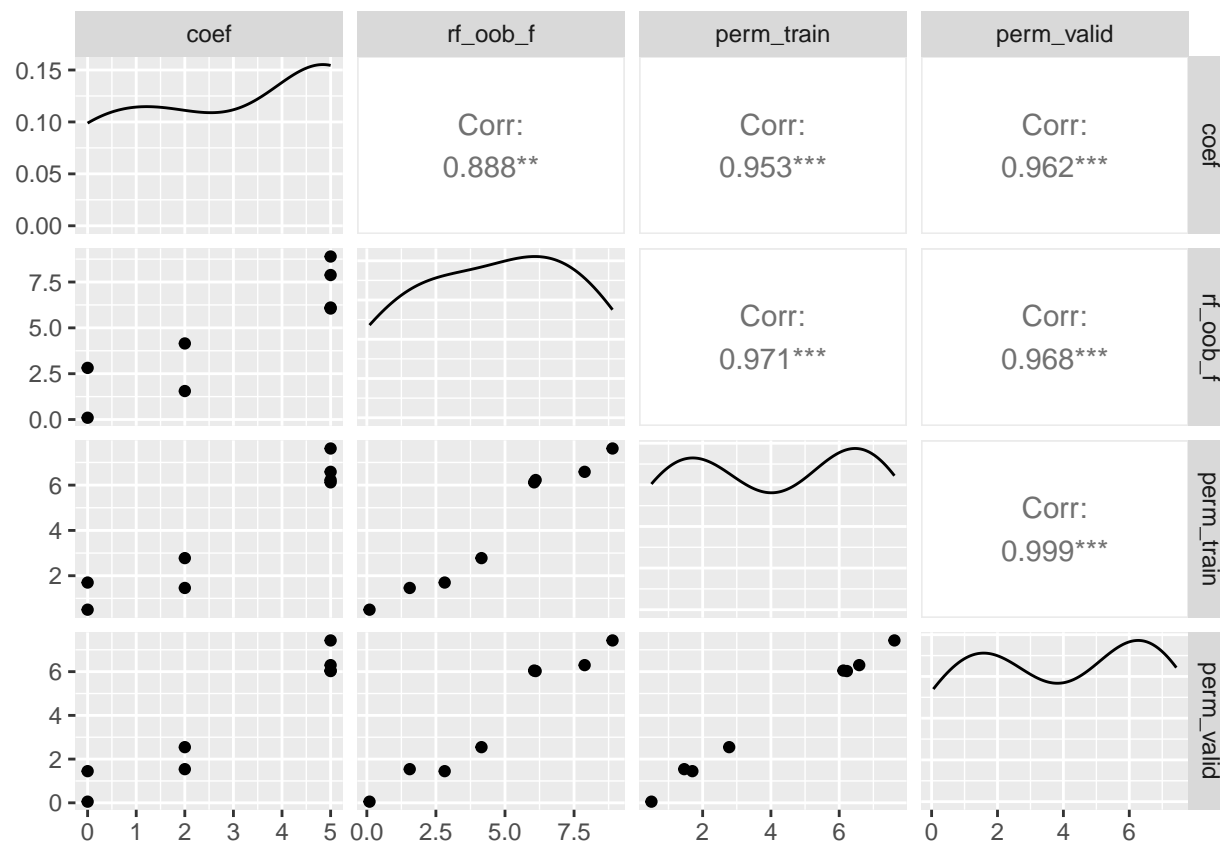
```
##      coef      rf_oob_f perm_train perm_valid
## 1      5 8.85377982    7.661362 7.34928037
## 2      5 7.89579428    6.460074 6.17266630
## 3      2 4.20881295    2.805603 2.56785404
## 4      0 2.98832197    1.794127 1.52446979
## 5      5 6.09779102    6.216068 6.03255609
## 6      5 6.04642915    6.142736 6.05380372
## 7      2 1.56704361    1.470628 1.56144818
## 8      0 0.08126967    0.503865 0.08193564
```



##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.0222742	8.89188976	7.6171784	6.3348151	7.42798613	6.03039936
## 2	5	5.5066072	7.88286859	6.5861615	5.4838336	6.29789755	5.17008432
## 3	2	4.0738069	4.14940134	2.7774567	2.6273046	2.54605911	2.38049724
## 4	0	3.7805343	2.81722921	1.6991040	1.6282520	1.44805658	1.35447366
## 5	5	12.4730002	6.10969742	6.2246851	4.8343230	6.02774253	4.63713875
## 6	5	12.4863316	6.05559396	6.1218387	4.7949016	6.04922435	4.68324427
## 7	2	5.6956883	1.55252283	1.4606751	1.2876046	1.53940435	1.37668195
## 8	0	0.4621747	0.09704756	0.5002986	0.4756553	0.05277177	0.04545093



```
##      coef      rf_oob_f perm_train perm_valid
## 1      5 8.89188976  7.6171784 7.42798613
## 2      5 7.88286859  6.5861615 6.29789755
## 3      2 4.14940134  2.7774567 2.54605911
## 4      0 2.81722921  1.6991040 1.44805658
## 5      5 6.10969742  6.2246851 6.02774253
## 6      5 6.05559396  6.1218387 6.04922435
## 7      2 1.55252283  1.4606751 1.53940435
## 8      0 0.09704756  0.5002986 0.05277177
```



rsq

```
## [1] 0.8651792 0.9080931 0.9233991 0.9318195 0.9373068 0.9405644 0.9424866
## [8] 0.9441860 0.9449884 0.9452774 0.9454542 0.9453583
```

```
Names = c("Cor1", "Cor2", "Cor3", "Cor4", "V5", "V6", "V7", "V8")
mag <- ifelse(Names %in% c("Cor1", "Cor2", "V5", "V6"), 5,
             ifelse(Names %in% c("Cor3", "V7"), 2, 0))

Names <- factor(Names,
               levels = c("Cor1", "Cor2", "Cor3", "Cor4",
                         "V5", "V6", "V7", "V8"), ordered = T)
Names <- factor(Names, ordered = F)

rf_oob_f1 = data.frame(rf_oob_f, Names, mag)
#rf_pdp1 = data.frame(rf_pdp, Names, mag)
perm_train1 = data.frame(perm_train, Names, mag)
drop_valid1 = data.frame(drop_valid, Names, mag)
perm_valid1 = data.frame(perm_valid, Names, mag)

colnames(rf_oob_f1)[1:12] <- 1:12
rf_oob_f1 <- rf_oob_f1 %>% pivot_longer(!c(Names, mag), names_to = "mtry",
                                       values_to = "Imp")
rf_oob_f1$mtry <- as.numeric(rf_oob_f1$mtry)
```

```

# colnames(rf_pdp1)[1:12] <- 1:12
# rf_pdp1 <- rf_pdp1 %>% pivot_longer(!c(Names,mag), names_to = "mtry",
#                                     values_to = "Imp")
# rf_pdp1$mtry <- as.numeric(rf_pdp1$mtry)

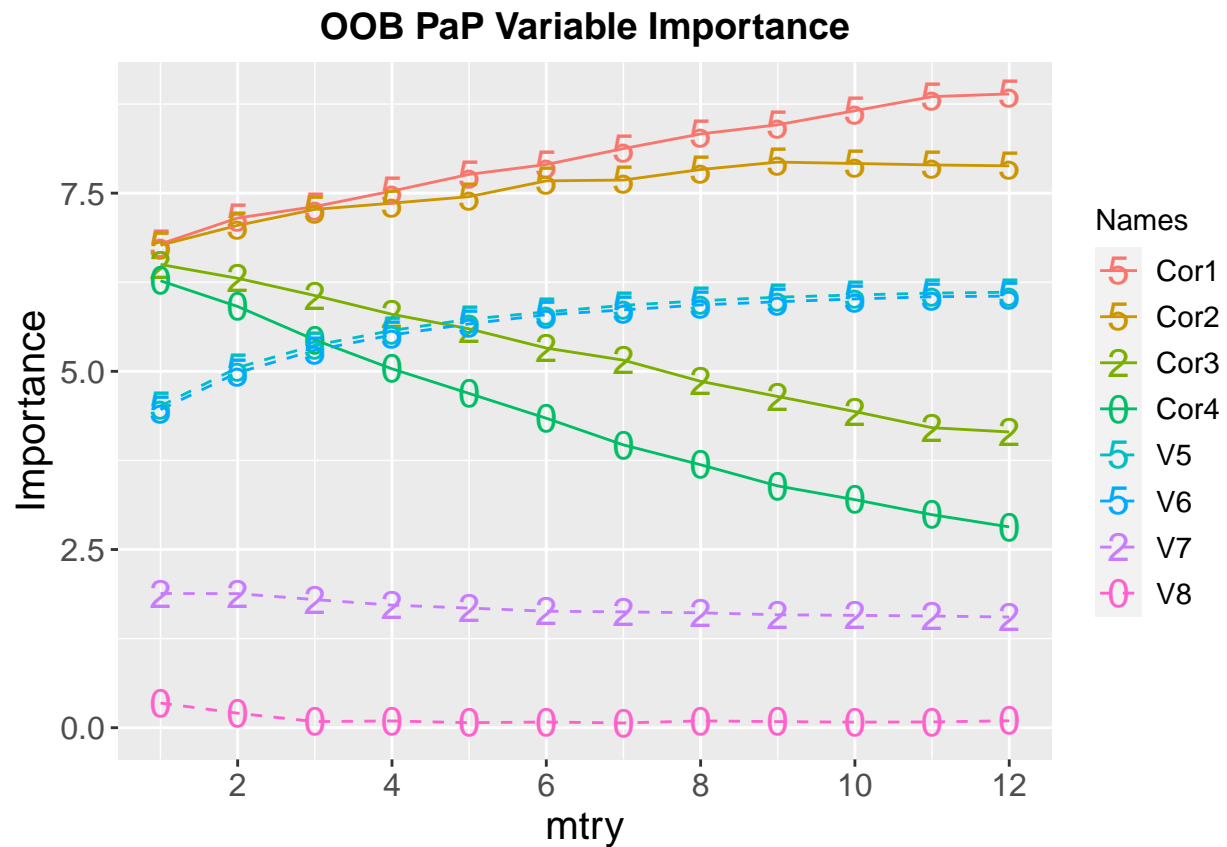
colnames(perm_train1)[1:12] <- 1:12
perm_train1 <- perm_train1 %>%
  pivot_longer(!c(Names,mag), names_to = "mtry", values_to = "Imp")
perm_train1$mtry <- as.numeric(perm_train1$mtry)

colnames(drop_valid1)[1:12] <- 1:12
drop_valid1 <- drop_valid1 %>%
  pivot_longer(!c(Names,mag), names_to = "mtry", values_to = "Imp")
drop_valid1$mtry <- as.numeric(drop_valid1$mtry)

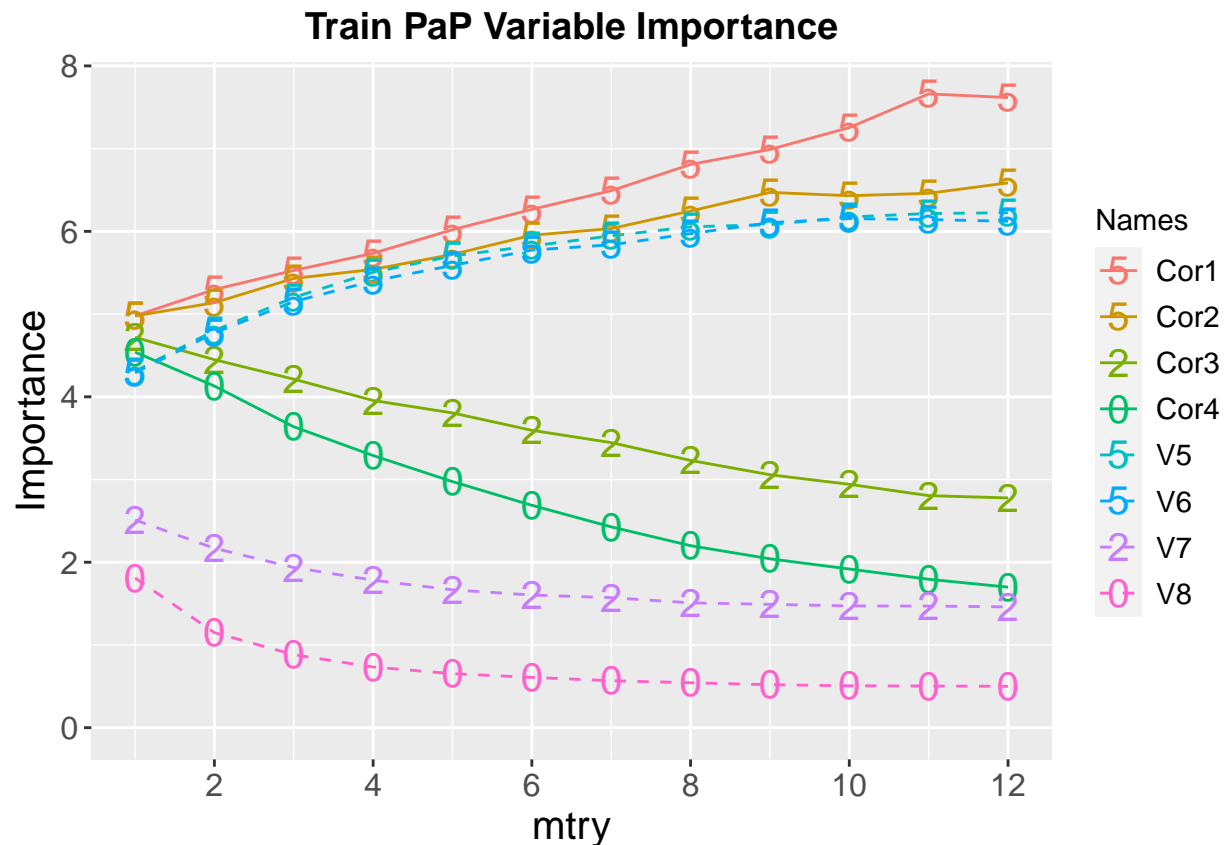
colnames(perm_valid1)[1:12] <- 1:12
perm_valid1 <- perm_valid1 %>%
  pivot_longer(!c(Names,mag), names_to = "mtry", values_to = "Imp")
perm_valid1$mtry <- as.numeric(perm_valid1$mtry)

gr <- rf_oob_f1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,max(rf_oob_f1$Imp))) +
  ggtitle("OOB PaP Variable Importance") +
  geom_point(size = 5) +
  scale_linetype_manual(values = rep(c(1, 2), each = 4)) +
  scale_shape_manual(values = c(53,53,50,48,53,53,50,48)) +
  scale_size(range = c(6,6)) +
  ylab("Importance") +
  guides(size = "none") +
  theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 15),
        plot.title = element_text(size = 14, face = "bold")) +
  easy_center_title() + easy_plot_legend_size(size = 11)
gr

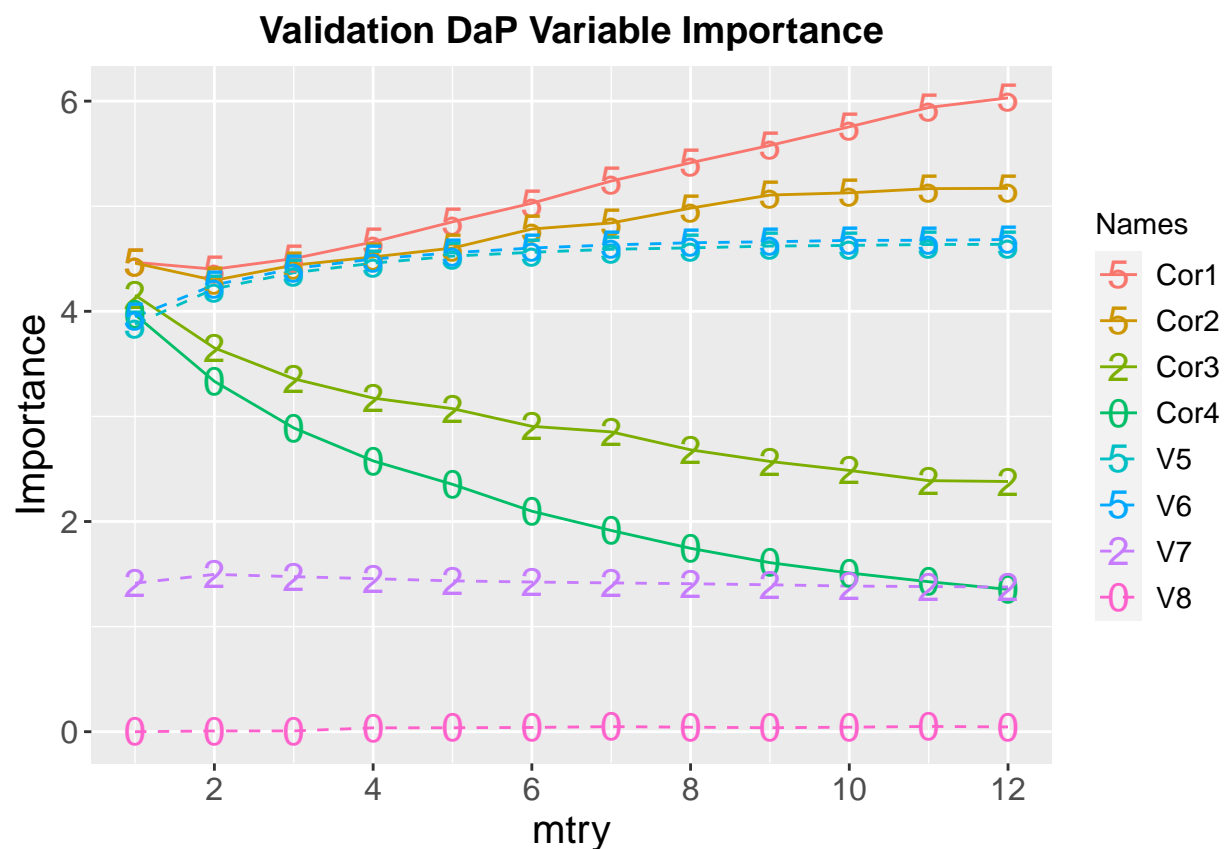
```



```
gp <- perm_train1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,max(perm_train1$Imp))) +
  ggtitle("Train PaP Variable Importance") +
  geom_point(size = 5) +
  scale_linetype_manual(values = rep(c(1, 2), each = 4)) +
  scale_shape_manual(values = c(53,53,50,48,53,53,50,48)) +
  scale_size(range = c(6,6)) +
  ylab("Importance") +
  guides(size = "none") +
  theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 15),
        plot.title = element_text(size = 14, face = "bold")) +
  easy_center_title() + easy_plot_legend_size(size = 11)
gp
```

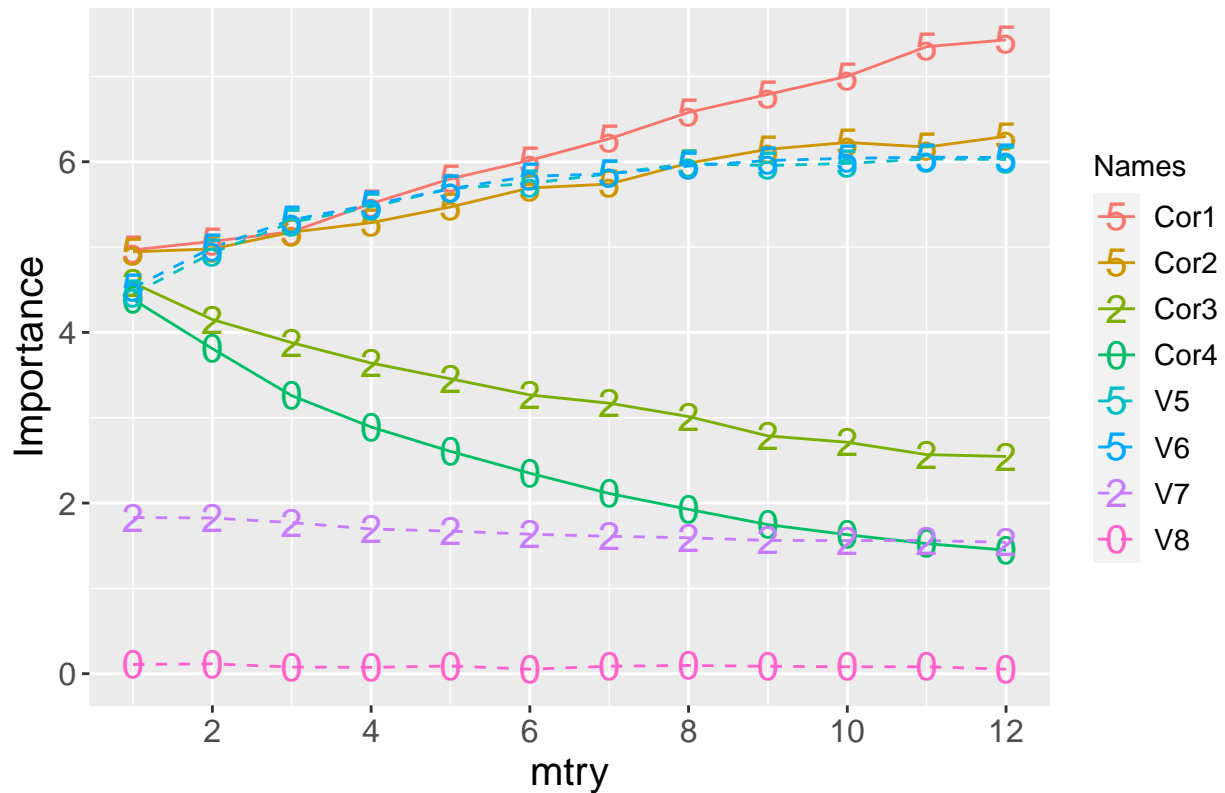


```
gd <- drop_valid1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,max(drop_valid1$Imp))) +
  ggtitle("Validation DaP Variable Importance") +
  geom_point(size = 5) +
  scale_linetype_manual(values = rep(c(1, 2), each = 4)) +
  scale_shape_manual(values = c(53,53,50,48,53,53,50,48)) +
  scale_size(range = c(6,6)) +
  ylab("Importance") +
  guides(size = "none") +
  theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 15),
        plot.title = element_text(size = 14, face = "bold")) +
  easy_center_title() + easy_plot_legend_size(size = 11)
gd
```



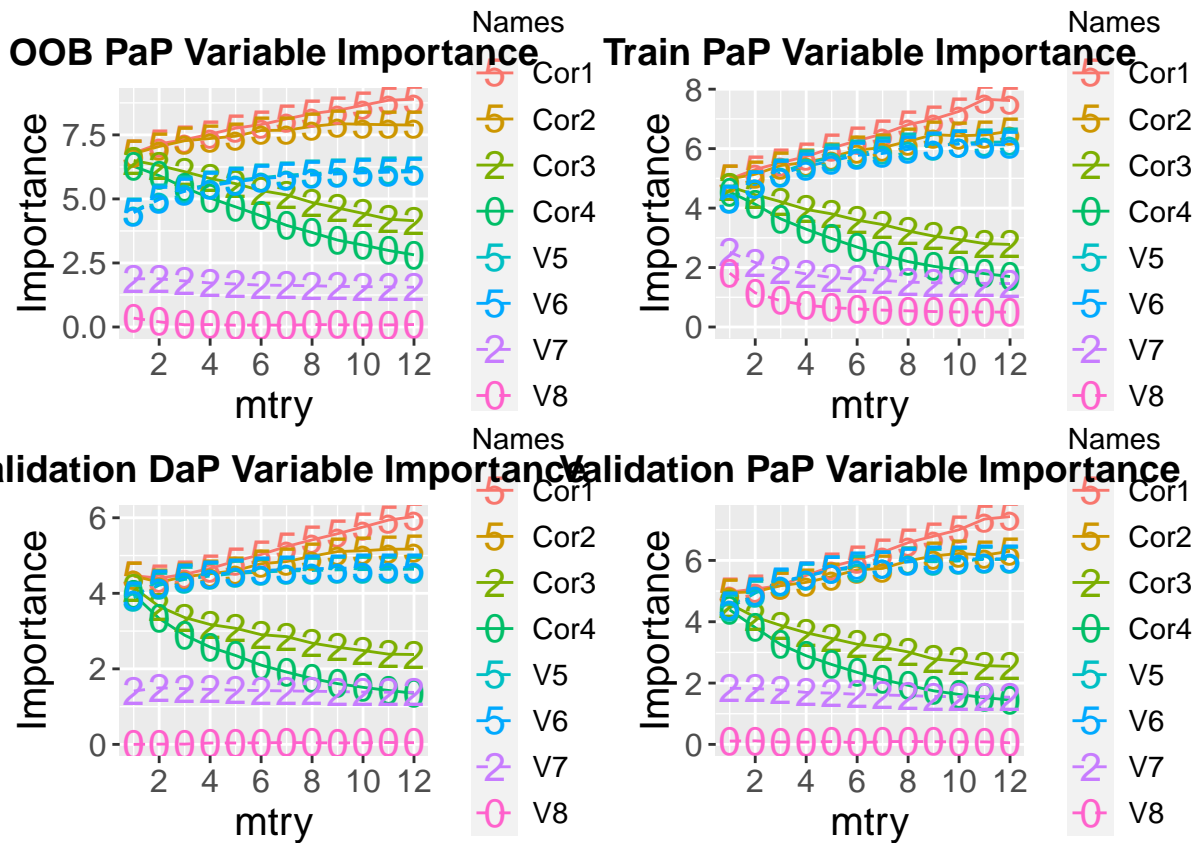
```
gv <- perm_valid1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,max(perm_valid1$Imp))) +
  ggtitle("Validation PaP Variable Importance") +
  geom_point(size = 5) +
  scale_linetype_manual(values = rep(c(1, 2), each = 4)) +
  scale_shape_manual(values = c(53,53,50,48,53,53,50,48)) +
  scale_size(range = c(6,6)) +
  ylab("Importance") +
  guides(size = "none") +
  theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 15),
        plot.title = element_text(size = 14, face = "bold")) +
  easy_center_title() + easy_plot_legend_size(size = 11)
gv
```


Validation PaP Variable Importance



```
# gpp <- rf_pdp1 %>%
#   ggplot(aes(x = mtry, y = Imp, color = Names,
#             group = Names, linetype = Names,
#             shape = Names)) +
#   geom_line() +
#   scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
#   scale_y_continuous(limits = c(0,max(rf_pdp1$Imp))) +
#   ggtitle("PDP Variable Importance") +
#   geom_point(size = 5) +
#   scale_linetype_manual(values = rep(c(1, 2), each = 4)) +
#   scale_shape_manual(values = c(53,53,50,48,53,53,50,48)) +
#   scale_size(range = c(6,6)) +
#   ylab("Importance") +
#   guides(size = "none") +
#   theme(axis.text = element_text(size = 12),
#         axis.title = element_text(size = 15),
#         plot.title = element_text(size = 14, face = "bold")) +
#   easy_center_title() + easy_plot_legend_size(size = 11)
# gpp

library(patchwork)
gr + gp + gd + gv
```

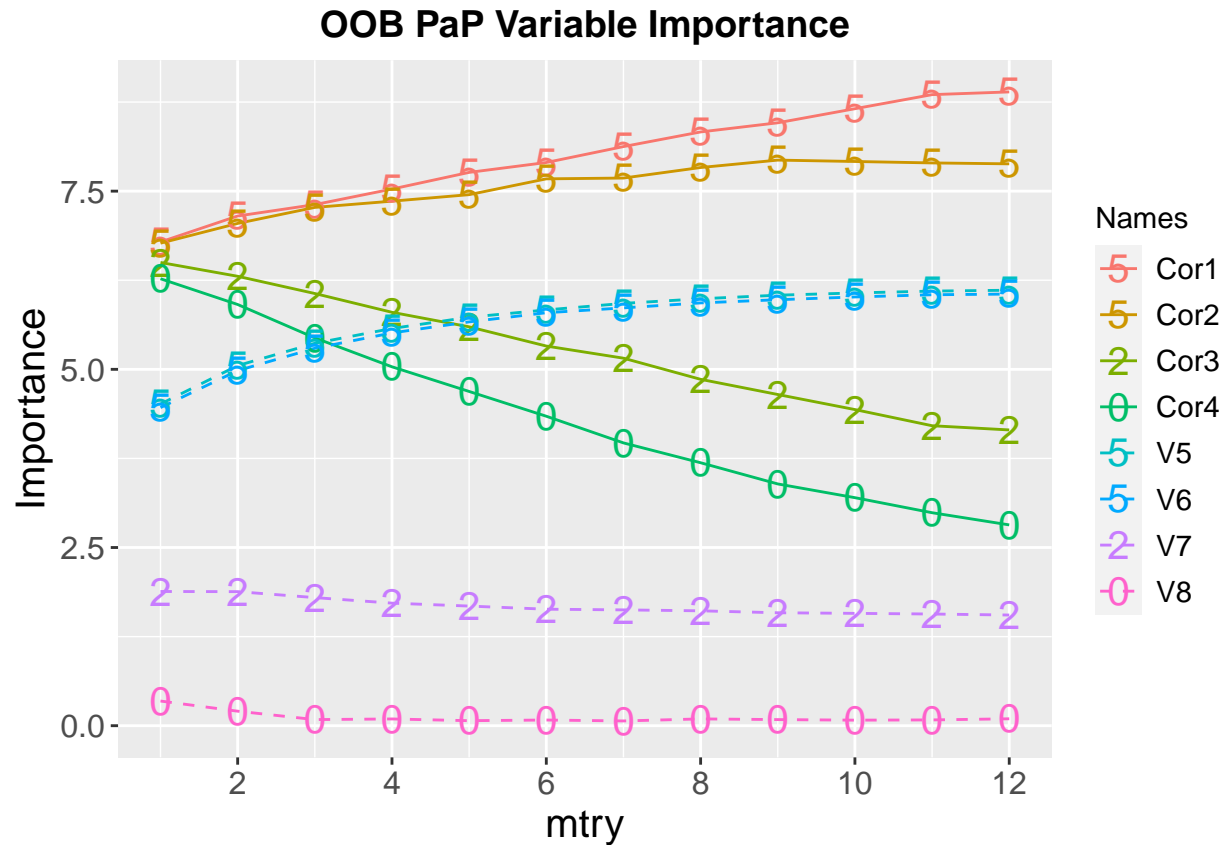


```
ggsave("x_all_zoom.pdf", plot = gr + gp + gd + gv, dpi = 2400,
       width = 9, height = 9)
```

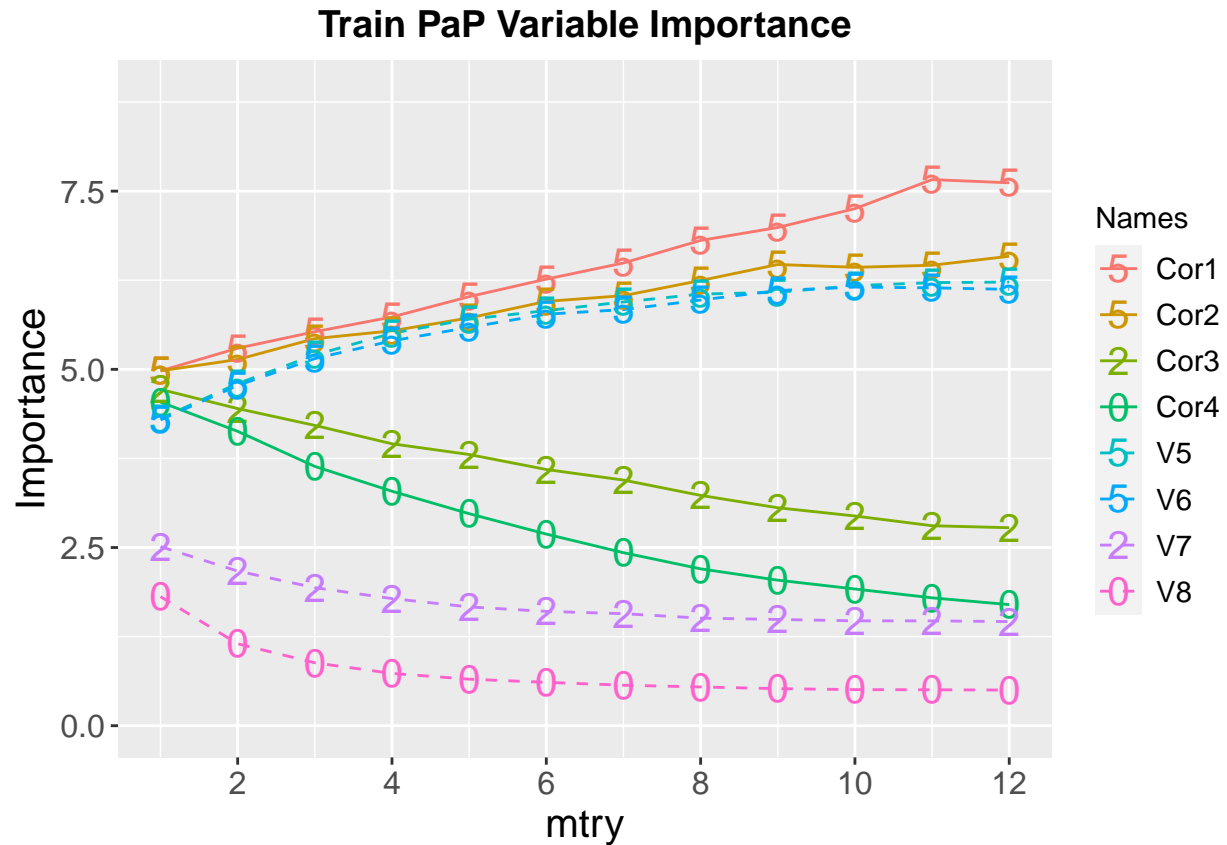
```
ma = max(rf_oob_f1$Imp, perm_train1$Imp, perm_valid1$Imp, drop_valid1$Imp)
# mp = max(rf_pdp1$Imp)

gr <- rf_oob_f1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,ma)) +
  ggtitle("OOB PaP Variable Importance") +
  geom_point(size = 5) +
  scale_linetype_manual(values = rep(c(1, 2), each = 4)) +
  scale_shape_manual(values = c(53,53,50,48,53,53,50,48)) +
  scale_size(range = c(6,6)) +
  ylab("Importance") +
  guides(size = "none") +
  theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 15),
        plot.title = element_text(size = 14, face = "bold")) +
  easy_center_title() + easy_plot_legend_size(size = 11)

gr
```



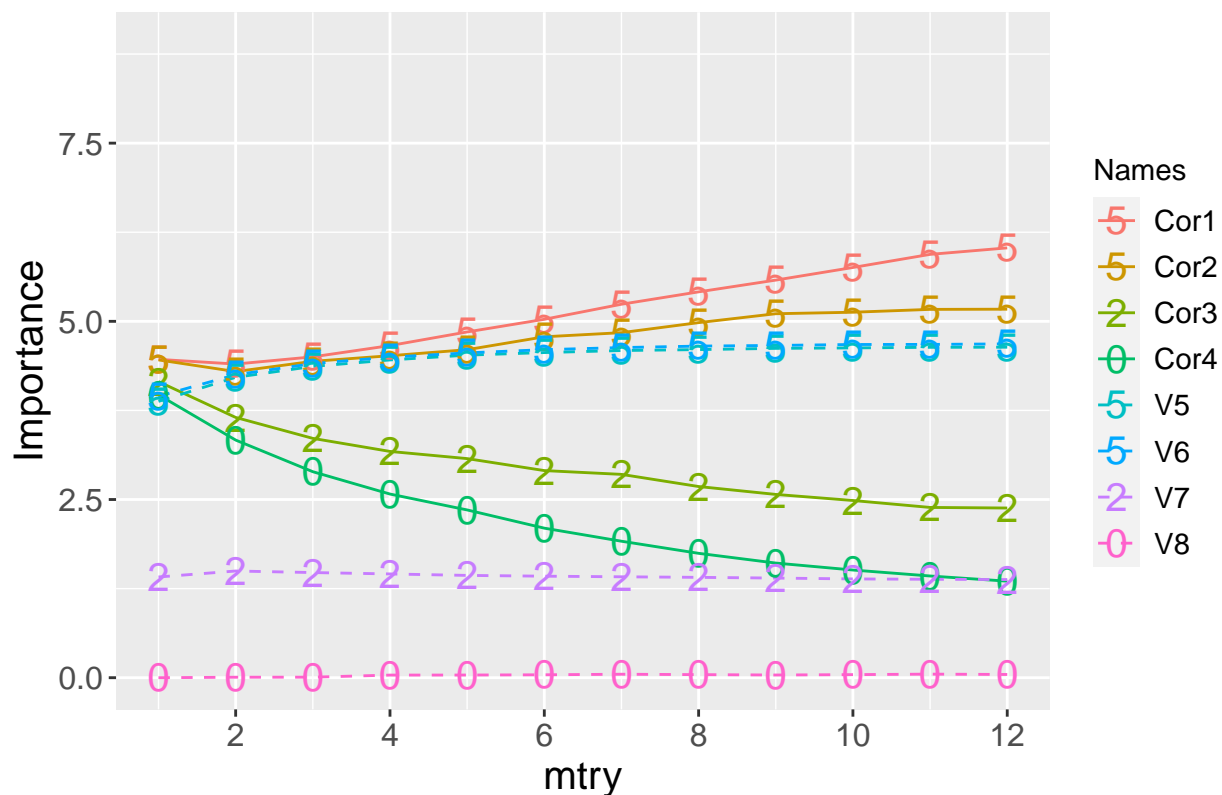
```
gp <- perm_train1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,ma)) +
  ggtitle("Train PaP Variable Importance") +
  geom_point(size = 5) +
  scale_linetype_manual(values = rep(c(1, 2), each = 4)) +
  scale_shape_manual(values = c(53,53,50,48,53,53,50,48)) +
  scale_size(range = c(6,6)) +
  ylab("Importance") +
  guides(size = "none") +
  theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 15),
        plot.title = element_text(size = 14, face = "bold")) +
  easy_center_title() + easy_plot_legend_size(size = 11)
gp
```



```
gd <- drop_valid1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,ma)) +
  ggtitle("Validation DaP Variable Importance") +
  geom_point(size = 5) +
  scale_linetype_manual(values = rep(c(1, 2), each = 4)) +
  scale_shape_manual(values = c(53,53,50,48,53,53,50,48)) +
  scale_size(range = c(6,6)) +
  ylab("Importance") +
  guides(size = "none") +
  theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 15),
        plot.title = element_text(size = 14, face = "bold")) +
  easy_center_title() + easy_plot_legend_size(size = 11)
```

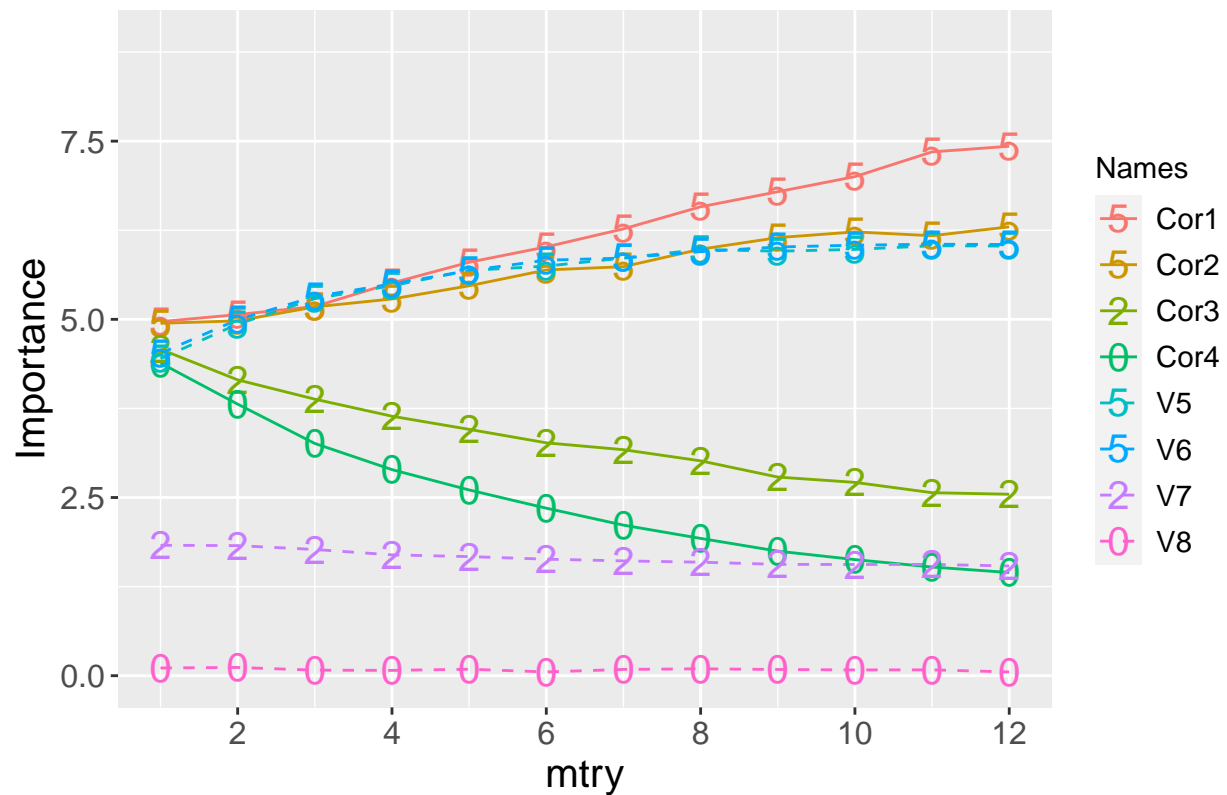
gd

Validation DaP Variable Importance



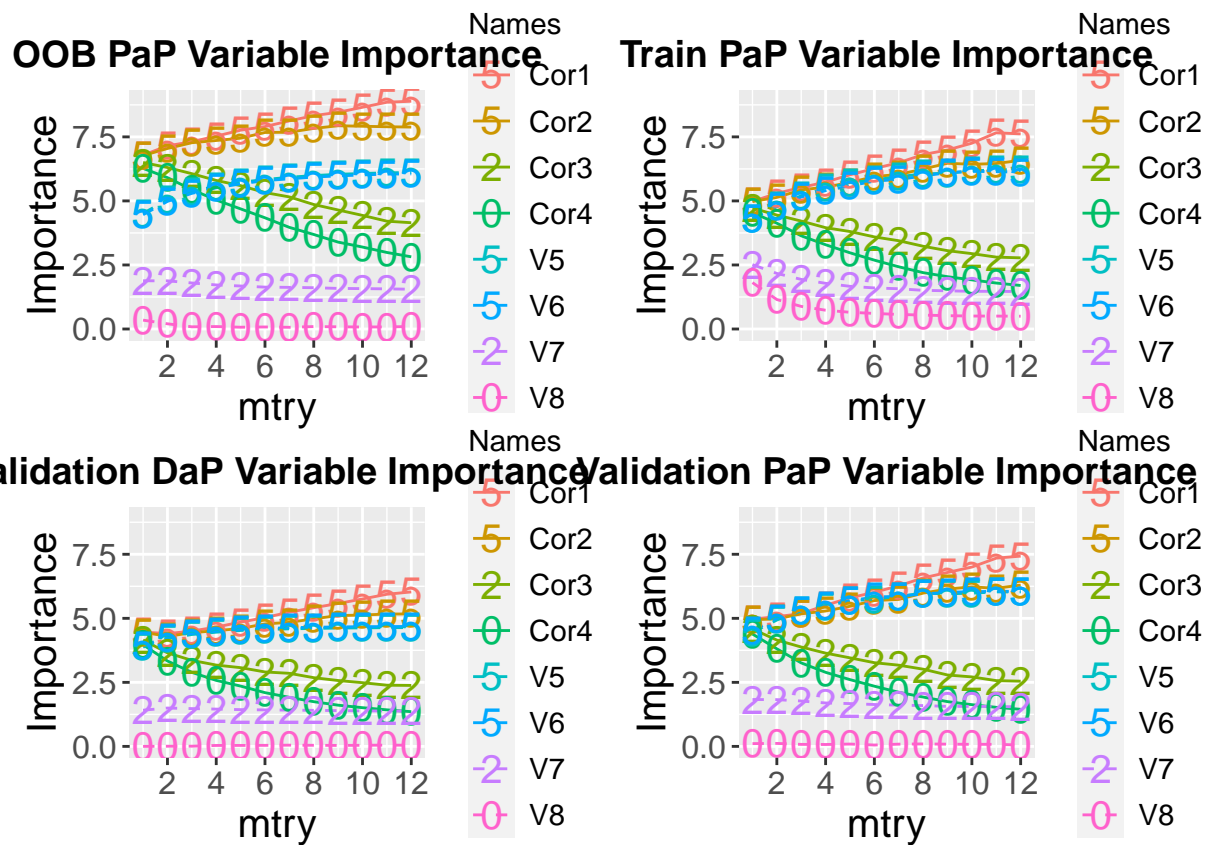
```
gv <- perm_valid1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,ma)) +
  ggtitle("Validation PaP Variable Importance") +
  geom_point(size = 5) +
  scale_linetype_manual(values = rep(c(1, 2), each = 4)) +
  scale_shape_manual(values = c(53,53,50,48,53,53,50,48)) +
  scale_size(range = c(6,6)) +
  ylab("Importance") +
  guides(size = "none") +
  theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 15),
        plot.title = element_text(size = 14, face = "bold")) +
  easy_center_title() + easy_plot_legend_size(size = 11)
gv
```

Validation PaP Variable Importance



```
# gpp <- rf_pdp1 %>%
#   ggplot(aes(x = mtry, y = Imp, color = Names,
#             group = Names, linetype = Names,
#             shape = Names)) +
#   geom_line() +
#   scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
#   scale_y_continuous(limits = c(0,mp)) +
#   ggtitle("PDP Variable Importance") +
#   geom_point(size = 5) +
#   scale_linetype_manual(values = rep(c(1, 2), each = 4)) +
#   scale_shape_manual(values = c(53,53,50,48,53,53,50,48)) +
#   scale_size(range = c(6,6)) +
#   ylab("Importance") +
#   guides(size = "none") +
#   theme(axis.text = element_text(size = 12),
#         axis.title = element_text(size = 15),
#         plot.title = element_text(size = 14, face = "bold")) +
#   easy_center_title() + easy_plot_legend_size(size = 11)
# gpp

library(patchwork)
gr + gp + gd + gv
```



```
ggsave("x_all.pdf", plot = gr + gp + gd + gv, dpi = 2400,
       width = 9, height = 9)
```

```
Sys.time() - s
```

```
## Time difference of 17.60173 mins
```