

permute_rf_strobl_x2_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)
  strobl = apply(strobl, 2, pnorm)
  strobl = (strobl - 0.5)*2

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

```

dfv <- MASS::mvrnorm(n_size, mu = rep(0, 12), Sigma = sig)
dfv = apply(dfv, 2, pnorm)
dfv = (dfv - 0.5)*2

y <- 5 * dfv[, 1]^2 + 5 * dfv[, 2]^2 + 2 * dfv[, 3]^2 +
  5 * dfv[, 5]^2 + 5 * dfv[, 6]^2 + 2 * dfv[, 7]^2 +
  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] + impdp / 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)

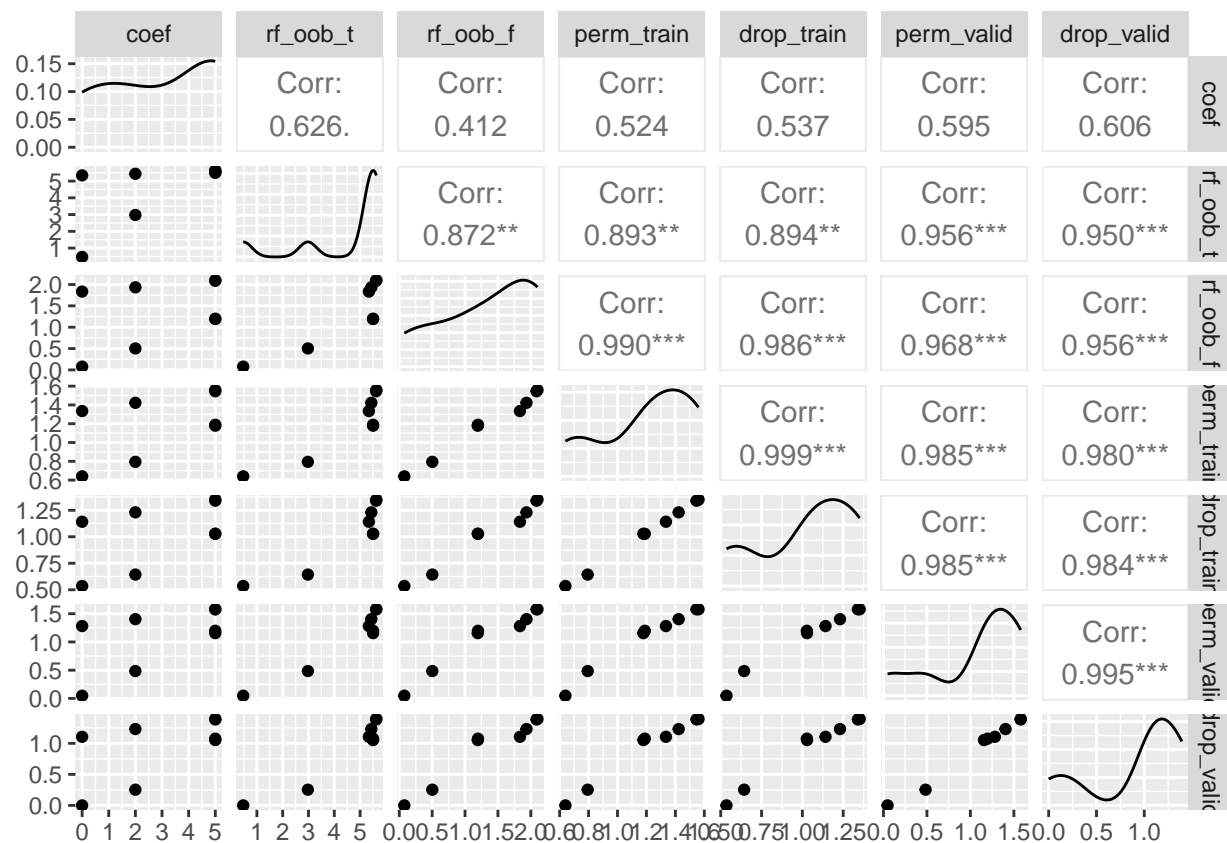
  print(sdf)
  print(ggpairs(sdf))
}

```

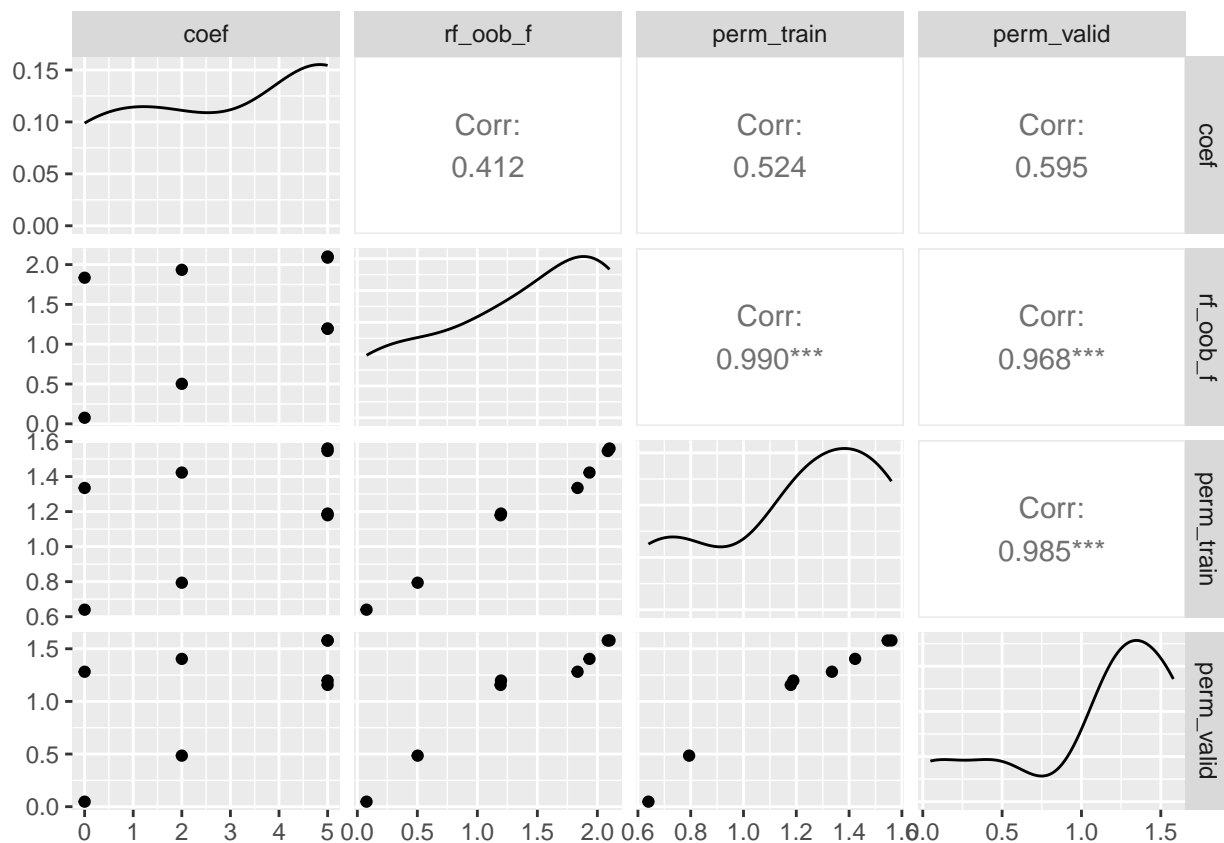
```

##      coef rf_oob_t  rf_oob_f perm_train drop_train perm_valid drop_valid
## 1      5 5.6315460 2.10126124 1.5603453 1.3507245 1.5790763 1.3929395
## 2      5 5.6196781 2.08715071 1.5461228 1.3400635 1.5782962 1.3857018
## 3      2 5.4368014 1.93434525 1.4227970 1.2304617 1.4039357 1.2304632
## 4      0 5.3418515 1.83491679 1.3349752 1.1416316 1.2817469 1.1054158
## 5      5 5.5035889 1.19366743 1.1790811 1.0284278 1.1566230 1.0529314
## 6      5 5.5023673 1.19721442 1.1887575 1.0274990 1.1981855 1.0751072
## 7      2 2.9768437 0.50266121 0.7940690 0.6433299 0.4848397 0.2530355
## 8      0 0.4704651 0.07664643 0.6398891 0.5352094 0.0477671 0.0000000

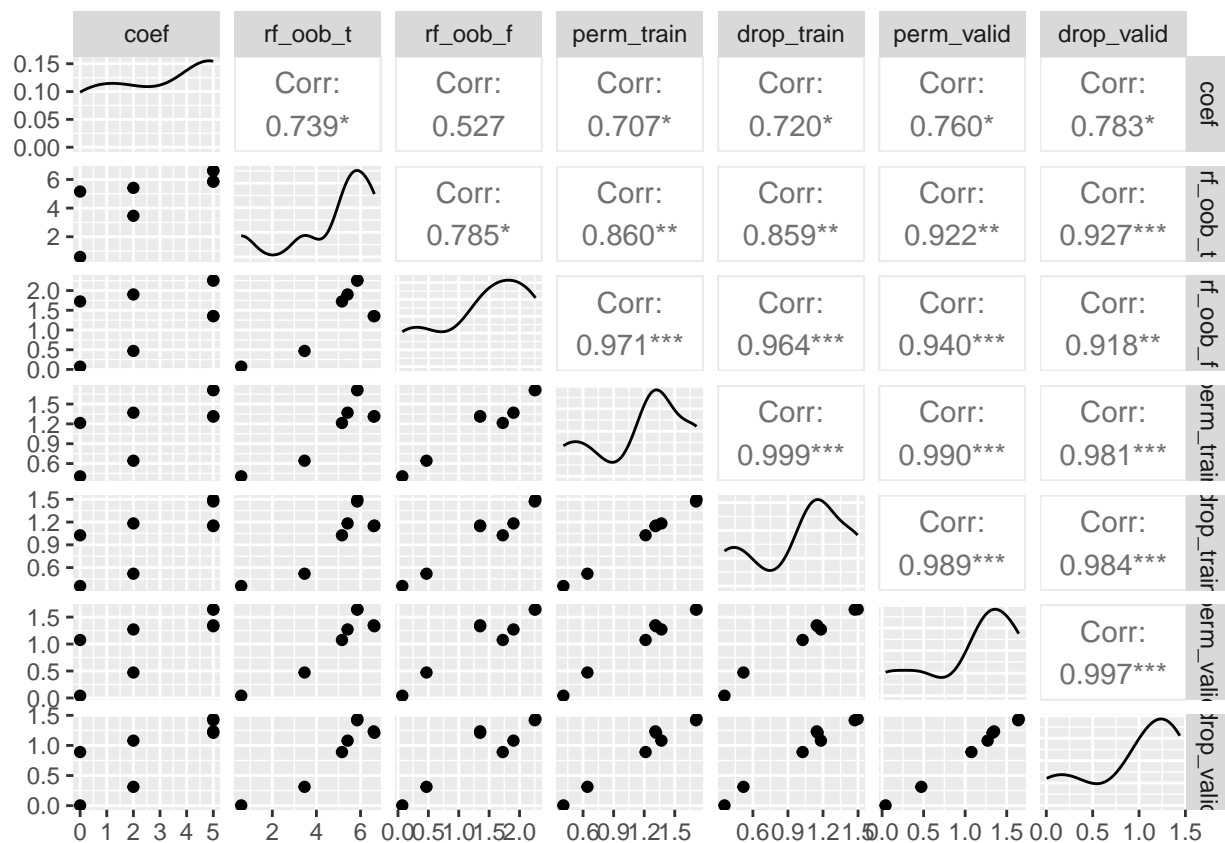
```



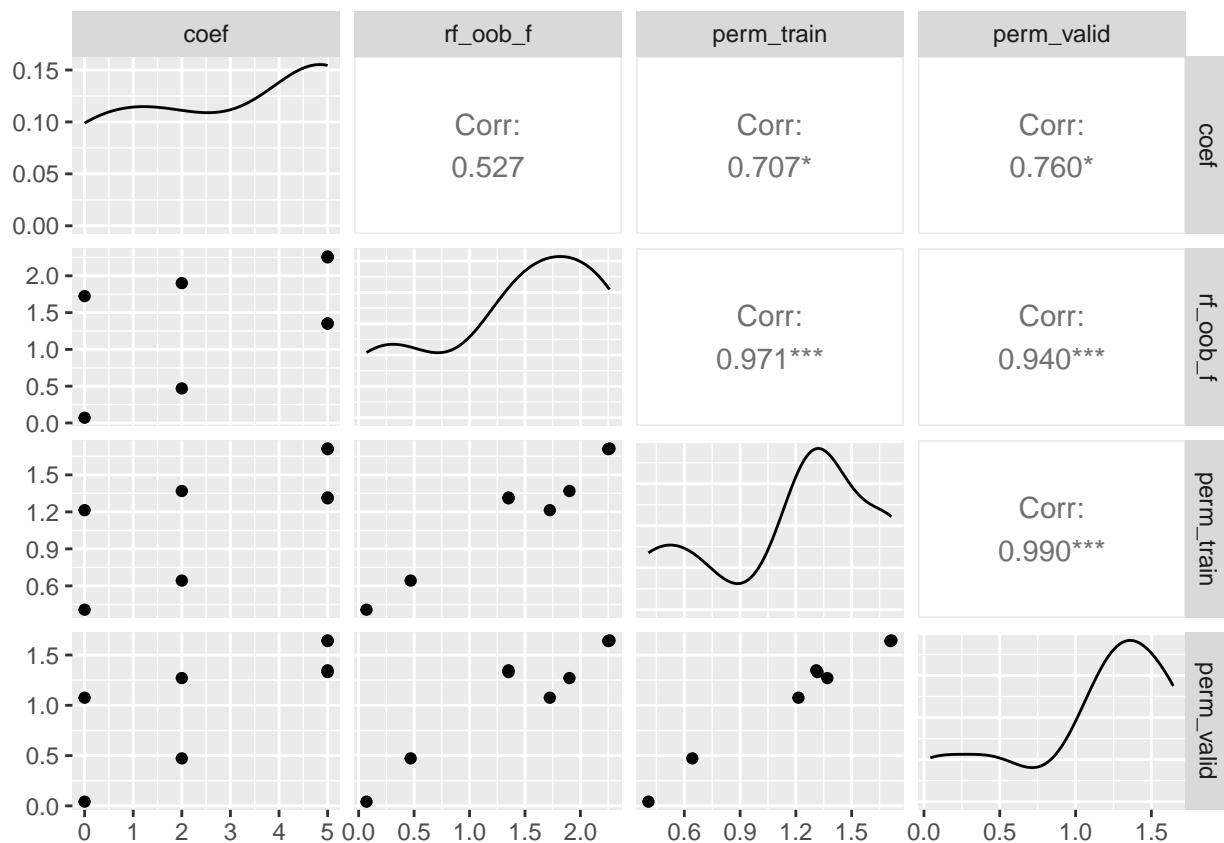
```
##      coef    rf_oob_f perm_train perm_valid
## 1      5 2.10126124  1.5603453  1.5790763
## 2      5 2.08715071  1.5461228  1.5782962
## 3      2 1.93434525  1.4227970  1.4039357
## 4      0 1.83491679  1.3349752  1.2817469
## 5      5 1.19366743  1.1790811  1.1566230
## 6      5 1.19721442  1.1887575  1.1981855
## 7      2 0.50266121  0.7940690  0.4848397
## 8      0 0.07664643  0.6398891  0.0477671
```



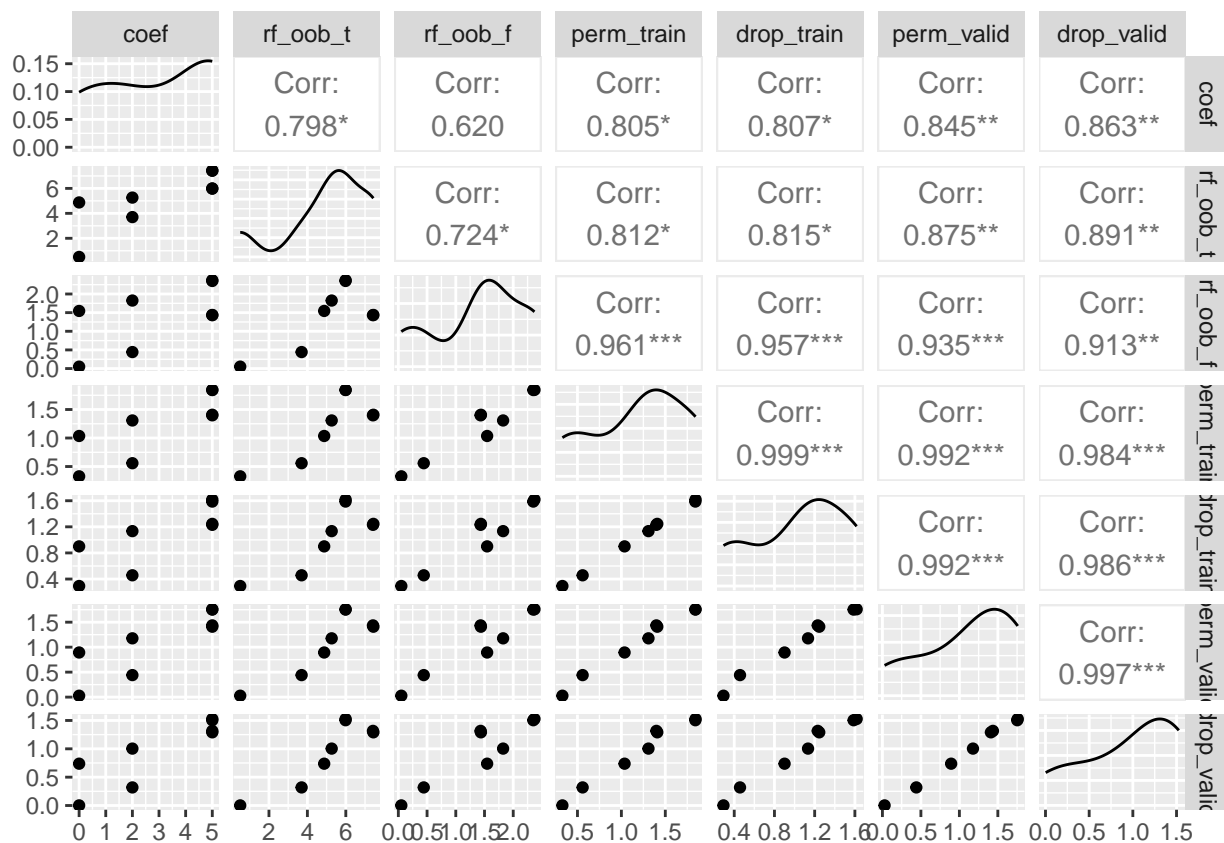
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.8595886	2.26231804	1.7140671	1.4964755	1.64497444	1.439733333
## 2	5	5.8471079	2.25082360	1.7085760	1.4714757	1.63887527	1.418271427
## 3	2	5.4129790	1.90086748	1.3693120	1.1821800	1.27076967	1.079535238
## 4	0	5.1579545	1.72414552	1.2139600	1.0257461	1.07547946	0.890323469
## 5	5	6.6270509	1.35187407	1.3168803	1.1546765	1.33015588	1.210252220
## 6	5	6.5983824	1.35125049	1.3097948	1.1467797	1.34790949	1.233525032
## 7	2	3.4613143	0.46925162	0.6428365	0.5192387	0.47246538	0.312302634
## 8	0	0.5885521	0.07109284	0.4069572	0.3570491	0.04211942	0.004074448



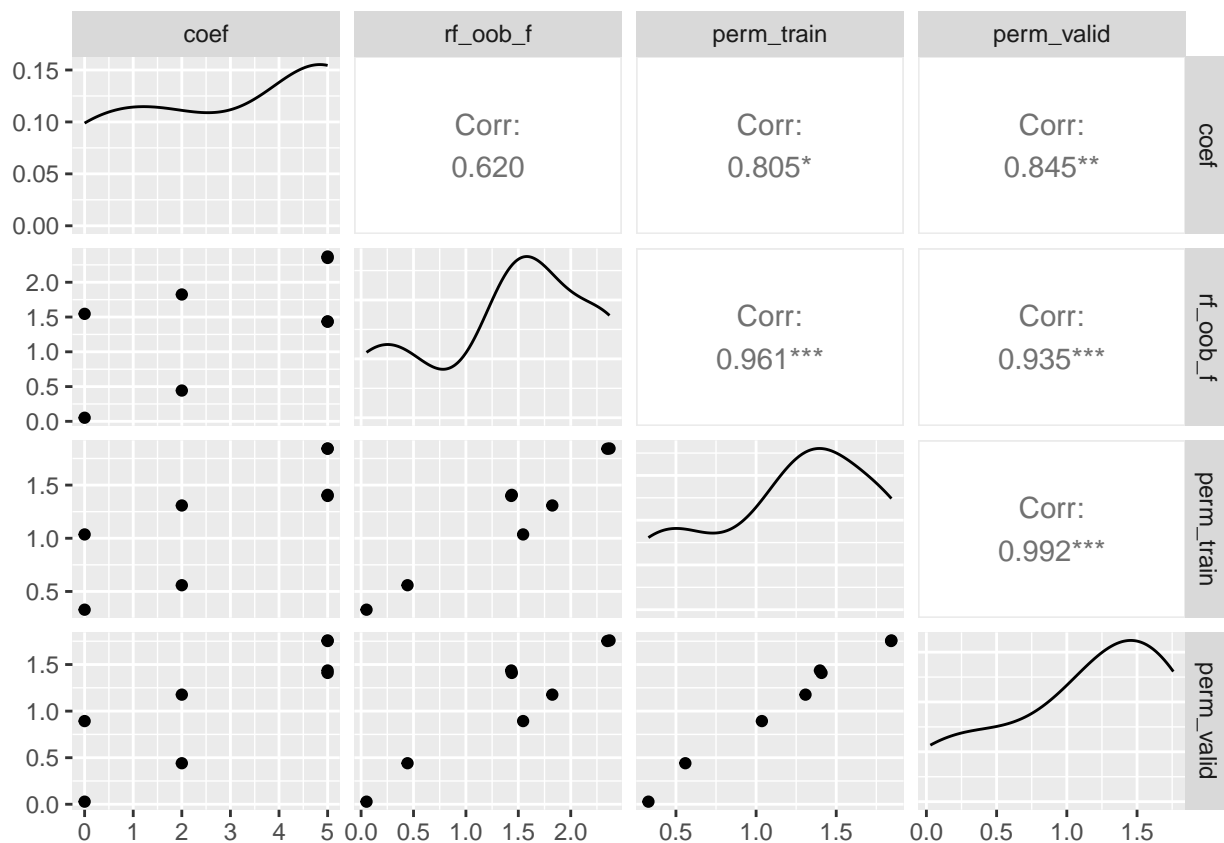
```
##      coef    rf_oob_f perm_train perm_valid
## 1      5 2.26231804  1.7140671 1.64497444
## 2      5 2.25082360  1.7085760 1.63887527
## 3      2 1.90086748  1.3693120 1.27076967
## 4      0 1.72414552  1.2139600 1.07547946
## 5      5 1.35187407  1.3168803 1.33015588
## 6      5 1.35125049  1.3097948 1.34790949
## 7      2 0.46925162  0.6428365 0.47246538
## 8      0 0.07109284  0.4069572 0.04211942
```



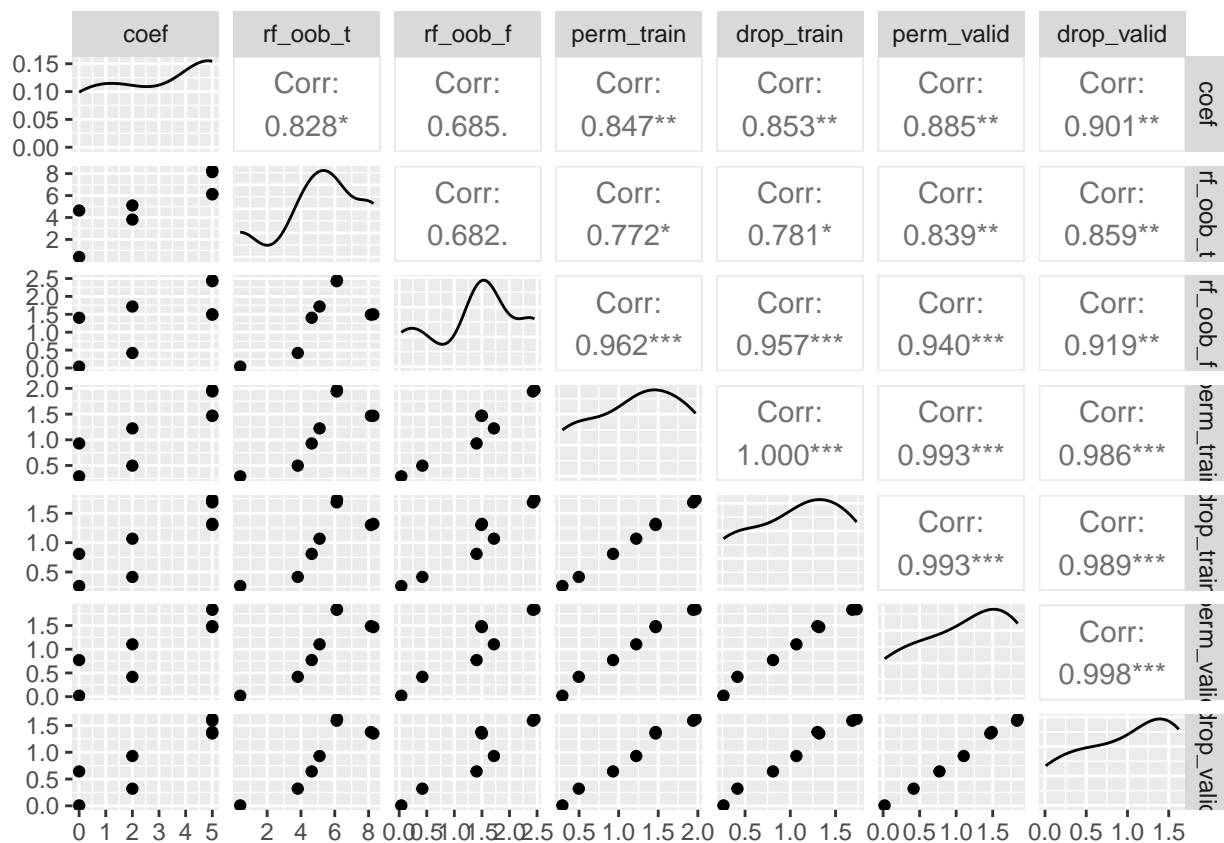
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.9910412	2.37009635	1.8443513	1.6176002	1.75853876	1.526464845
## 2	5	5.9873152	2.34964643	1.8420449	1.5873685	1.75243329	1.504125255
## 3	2	5.2652466	1.82420440	1.3083268	1.1339724	1.17676785	1.003648770
## 4	0	4.8748216	1.54573244	1.0360840	0.9005374	0.89310357	0.737661175
## 5	5	7.4347030	1.43768172	1.4082845	1.2454485	1.40982899	1.290864769
## 6	5	7.4141694	1.43260795	1.3982569	1.2306568	1.43585535	1.316368614
## 7	2	3.7002356	0.44294319	0.5586008	0.4573657	0.44061646	0.319186675
## 8	0	0.5048603	0.05175393	0.3283762	0.2929773	0.02804963	0.002030947



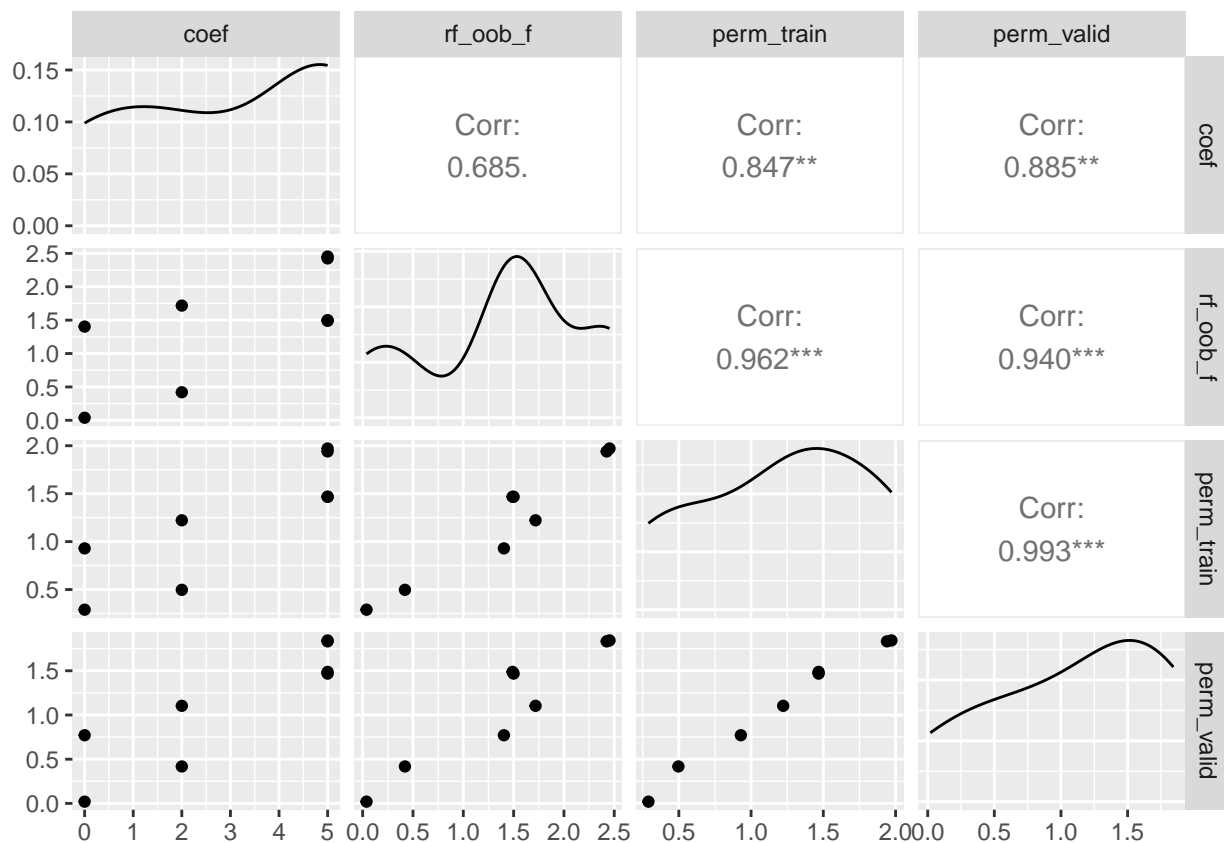
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  2.37009635  1.8443513  1.75853876
## 2      5  2.34964643  1.8420449  1.75243329
## 3      2  1.82420440  1.3083268  1.17676785
## 4      0  1.54573244  1.0360840  0.89310357
## 5      5  1.43768172  1.4082845  1.40982899
## 6      5  1.43260795  1.3982569  1.43585535
## 7      2  0.44294319  0.5586008  0.44061646
## 8      0  0.05175393  0.3283762  0.02804963
```



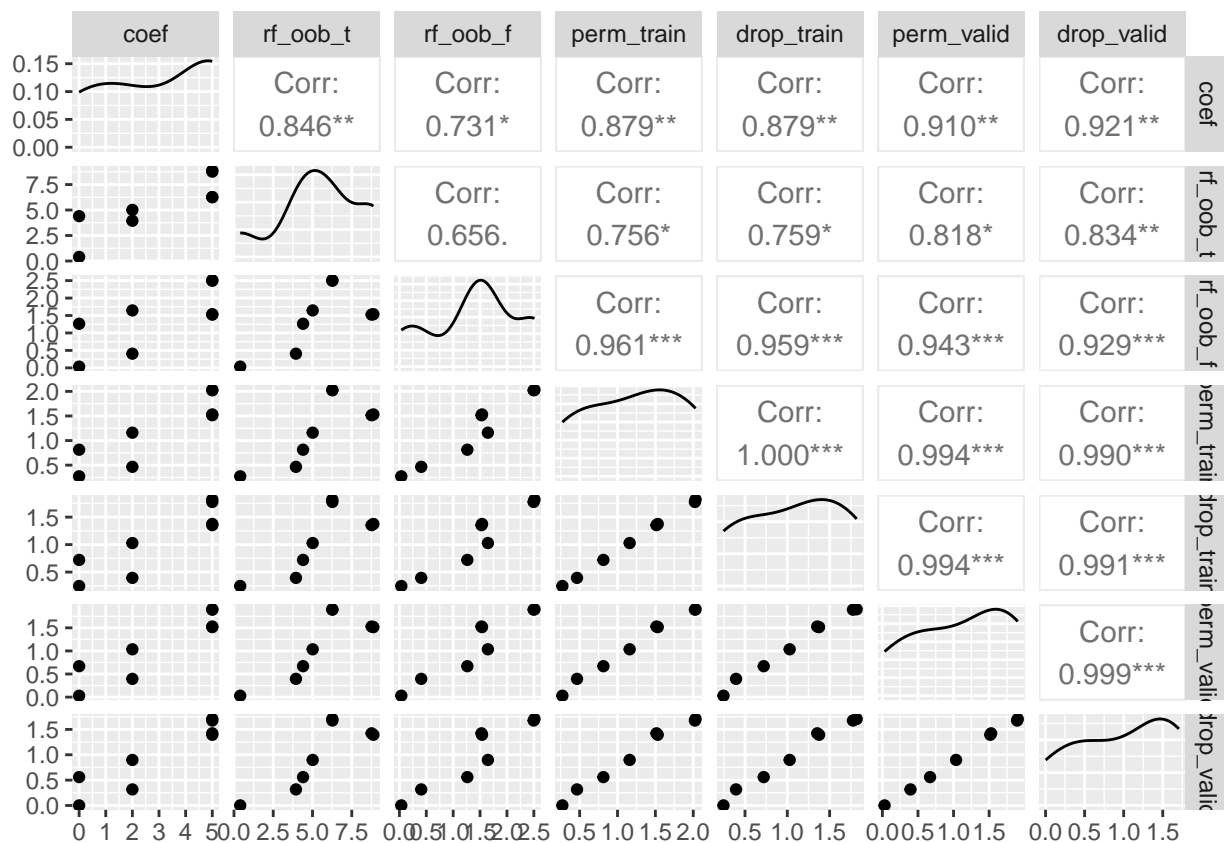
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.1311018	2.45304619	1.9714200	1.7337619	1.84315499	1.622210051
## 2	5	6.1158076	2.42463314	1.9405115	1.6855395	1.83279643	1.589292347
## 3	2	5.1020952	1.71805911	1.2223006	1.0675727	1.10359676	0.930523314
## 4	0	4.6375833	1.40349387	0.9293252	0.8086471	0.77118218	0.642955264
## 5	5	8.2836179	1.49993041	1.4675649	1.3204664	1.46788281	1.350556588
## 6	5	8.1385308	1.48963195	1.4674453	1.3008618	1.48725275	1.378845578
## 7	2	3.8126732	0.41997254	0.4971122	0.4159990	0.41810277	0.320803515
## 8	0	0.4063083	0.03829306	0.2897407	0.2613705	0.01991631	0.008390267



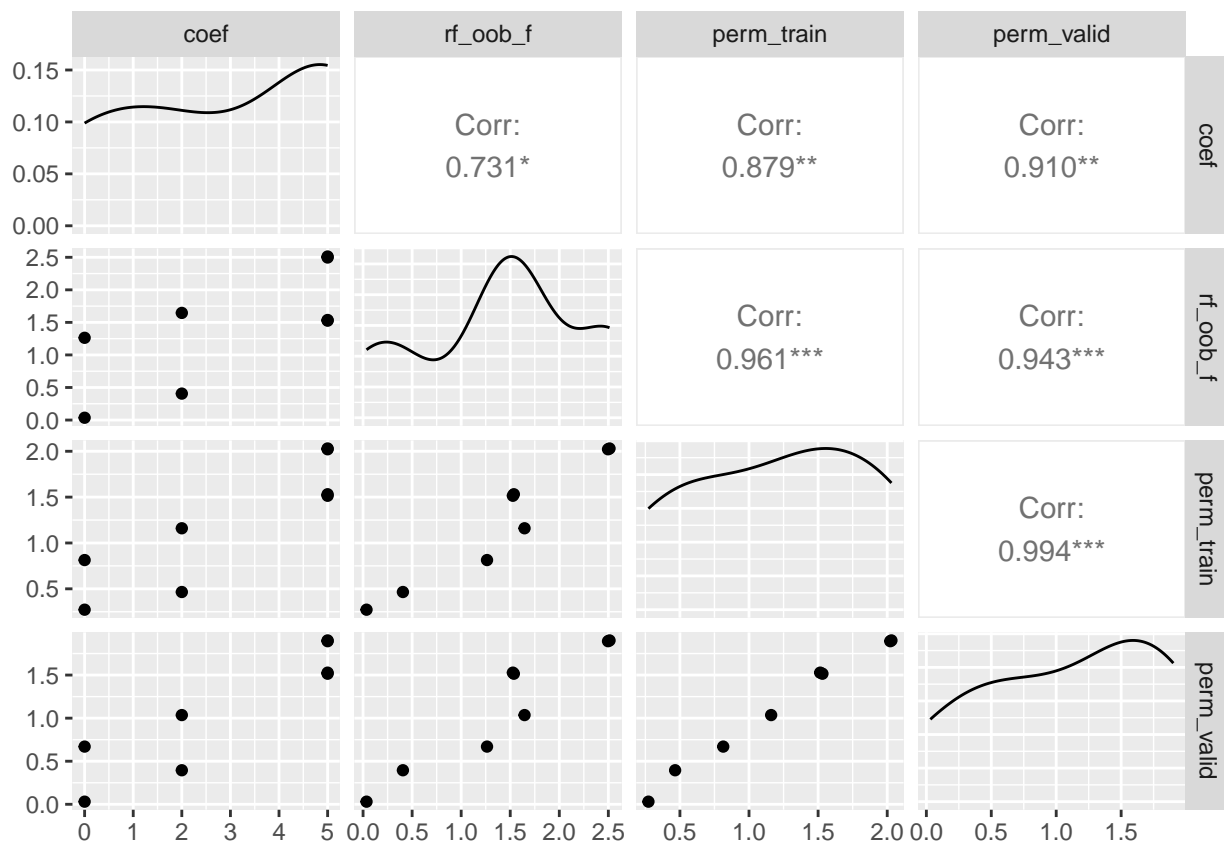
```
##      coef    rf_oob_f perm_train perm_valid
## 1      5  2.45304619  1.9714200  1.84315499
## 2      5  2.42463314  1.9405115  1.83279643
## 3      2  1.71805911  1.2223006  1.10359676
## 4      0  1.40349387  0.9293252  0.77118218
## 5      5  1.49993041  1.4675649  1.46788281
## 6      5  1.48963195  1.4674453  1.48725275
## 7      2  0.41997254  0.4971122  0.41810277
## 8      0  0.03829306  0.2897407  0.01991631
```



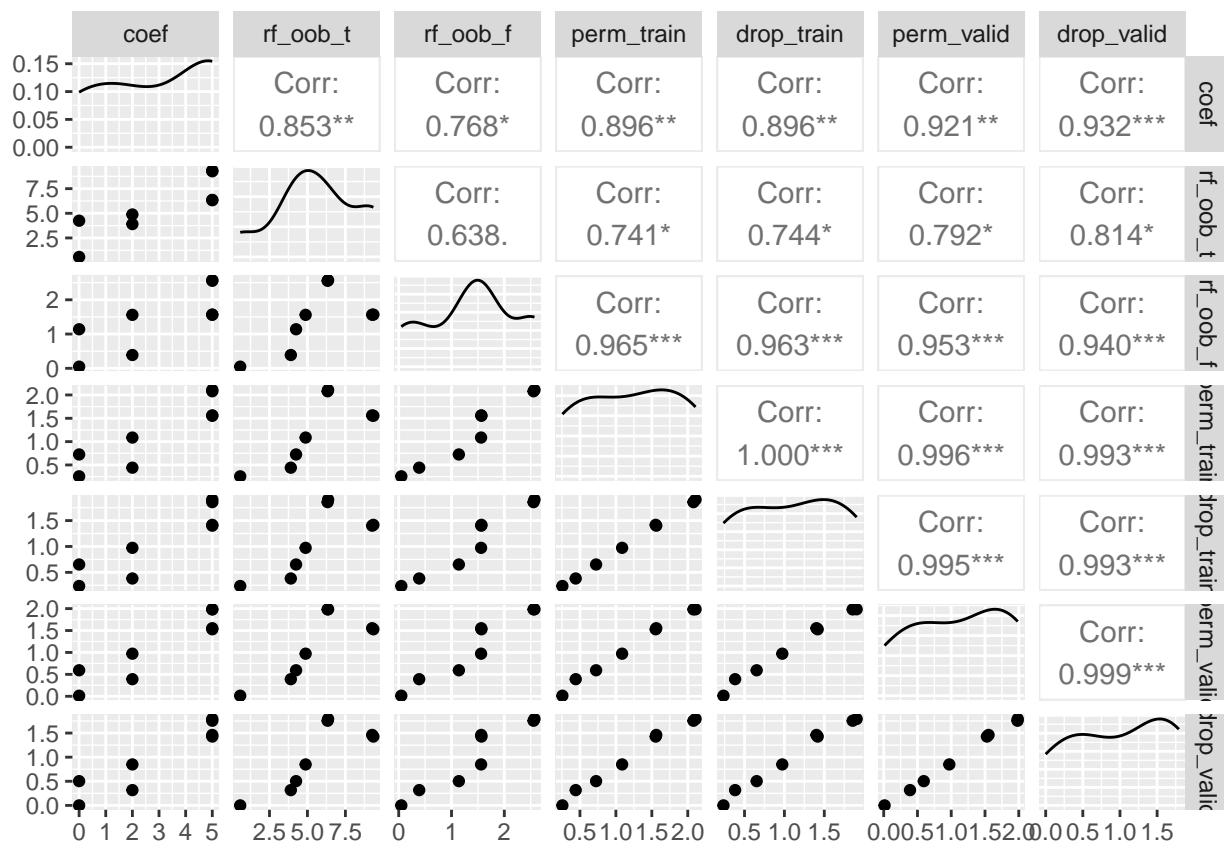
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.2617884	2.51191710	2.0305103	1.8192042	1.90206842	1.705957937
## 2	5	6.2673755	2.49667700	2.0220274	1.7758017	1.89354154	1.675538164
## 3	2	5.0124867	1.64570834	1.1603197	1.0279294	1.03575355	0.897970576
## 4	0	4.3991958	1.26346296	0.8136020	0.7213263	0.66972351	0.559556282
## 5	5	8.8768850	1.53670802	1.5312845	1.3753609	1.51590867	1.394172352
## 6	5	8.7642727	1.52676607	1.5148183	1.3538048	1.52806228	1.423038608
## 7	2	3.9471706	0.40615803	0.4653806	0.3940939	0.39520089	0.316429385
## 8	0	0.3977708	0.03526087	0.2723005	0.2459907	0.03065944	0.001572027



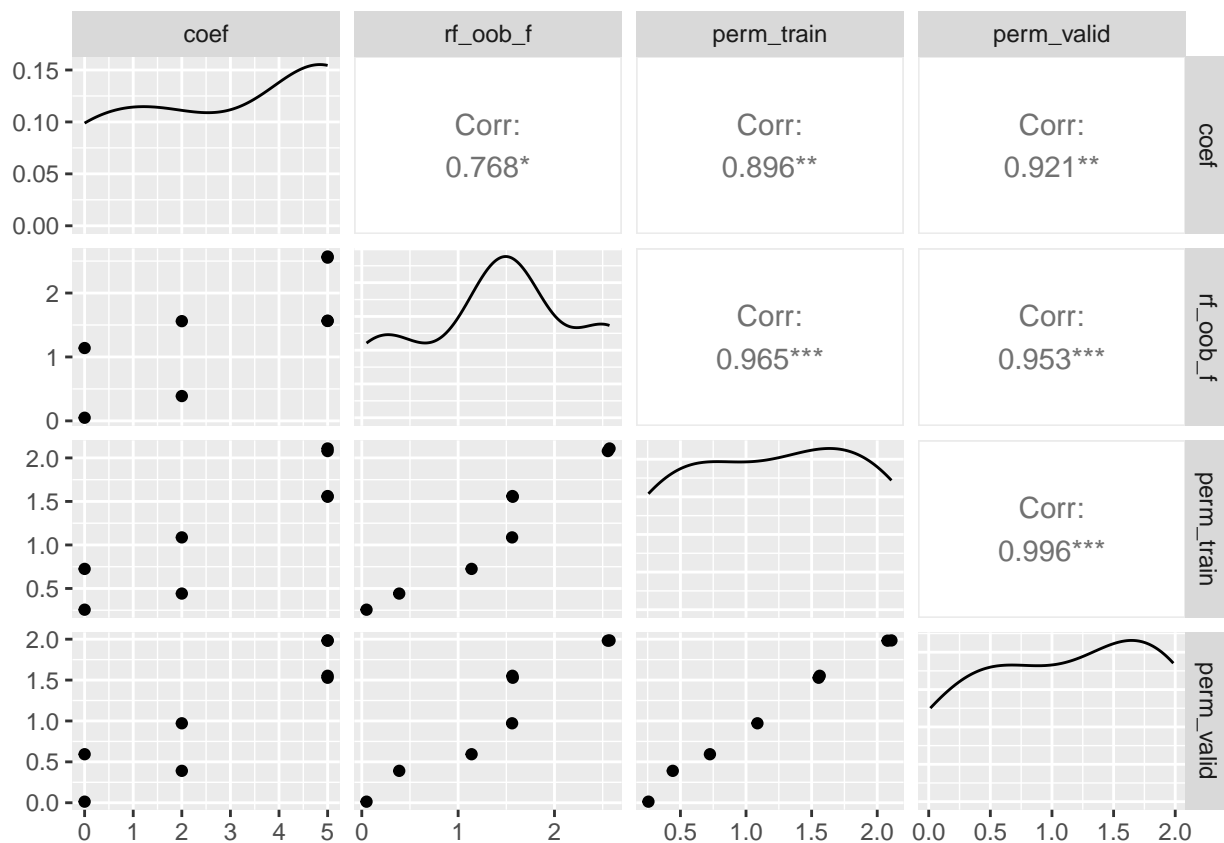
```
##      coef      rf_oob_f perm_train perm_valid
## 1      5 2.51191710  2.0305103  1.90206842
## 2      5 2.49667700  2.0220274  1.89354154
## 3      2 1.64570834  1.1603197  1.03575355
## 4      0 1.26346296  0.8136020  0.66972351
## 5      5 1.53670802  1.5312845  1.51590867
## 6      5 1.52676607  1.5148183  1.52806228
## 7      2 0.40615803  0.4653806  0.39520089
## 8      0 0.03526087  0.2723005  0.03065944
```



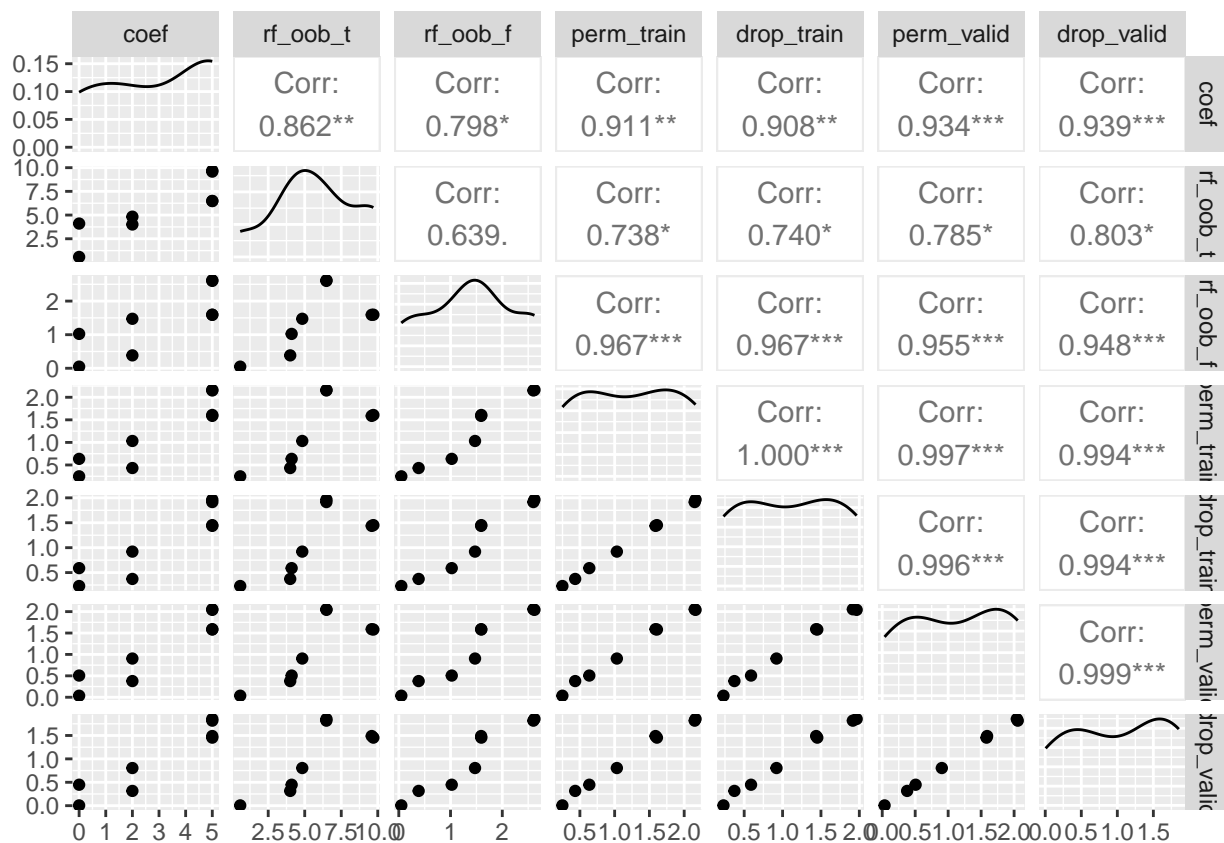
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.3606136	2.57127292	2.1083661	1.9056720	1.98479411	1.7922914
## 2	5	6.3303528	2.55406850	2.0788680	1.8559950	1.98199109	1.7556795
## 3	2	4.8804718	1.56009270	1.0869285	0.9737737	0.97153824	0.8491810
## 4	0	4.2505941	1.13945570	0.7248762	0.6523137	0.59327278	0.5029879
## 5	5	9.3427657	1.56709721	1.5536474	1.4172422	1.52861339	1.4278303
## 6	5	9.2630867	1.56317803	1.5617490	1.4018940	1.55245735	1.4607111
## 7	2	3.9060569	0.38838027	0.4414027	0.3818221	0.39008844	0.3169906
## 8	0	0.5722051	0.04906813	0.2560616	0.2363691	0.01315683	0.0000000



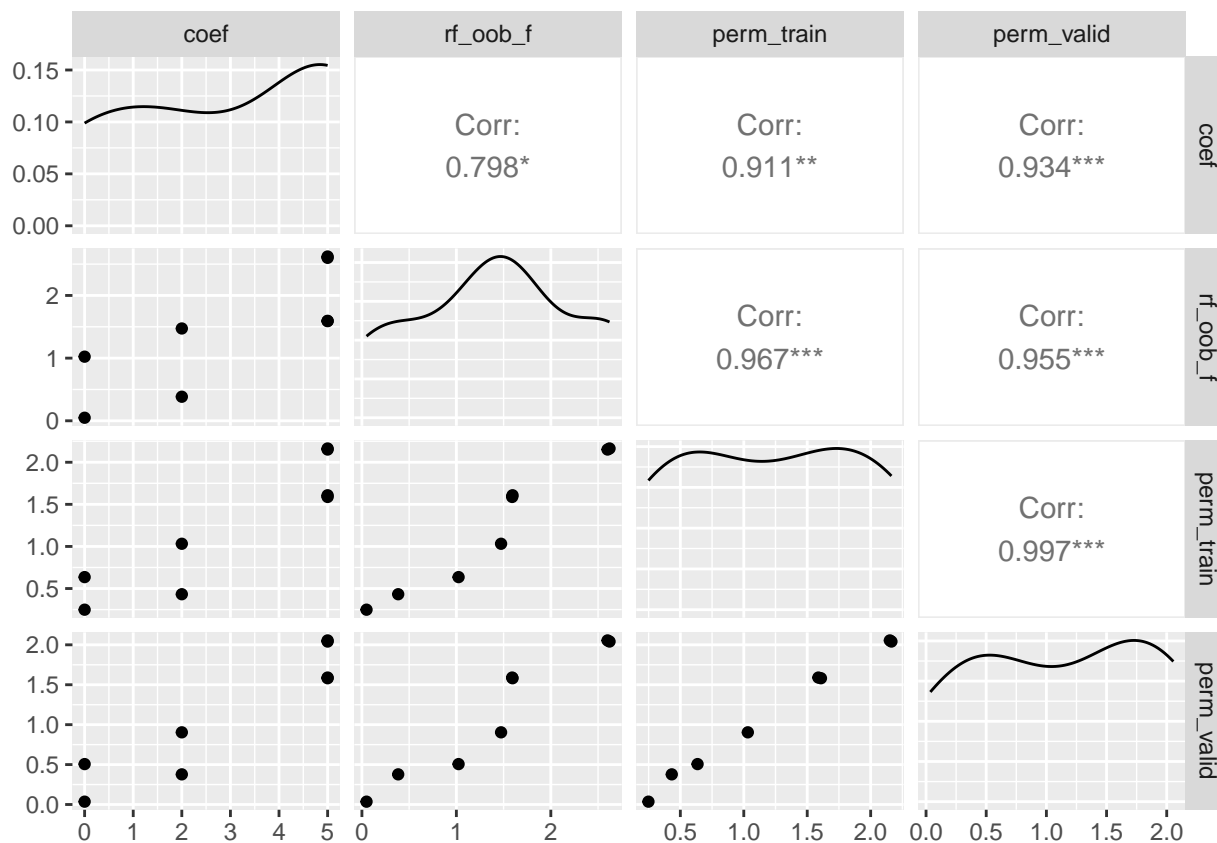
```
##      coef    rf_oob_f perm_train perm_valid
## 1      5 2.57127292  2.1083661 1.98479411
## 2      5 2.55406850  2.0788680 1.98199109
## 3      2 1.56009270  1.0869285 0.97153824
## 4      0 1.13945570  0.7248762 0.59327278
## 5      5 1.56709721  1.5536474 1.52861339
## 6      5 1.56317803  1.5617490 1.55245735
## 7      2 0.38838027  0.4414027 0.39008844
## 8      0 0.04906813  0.2560616 0.01315683
```



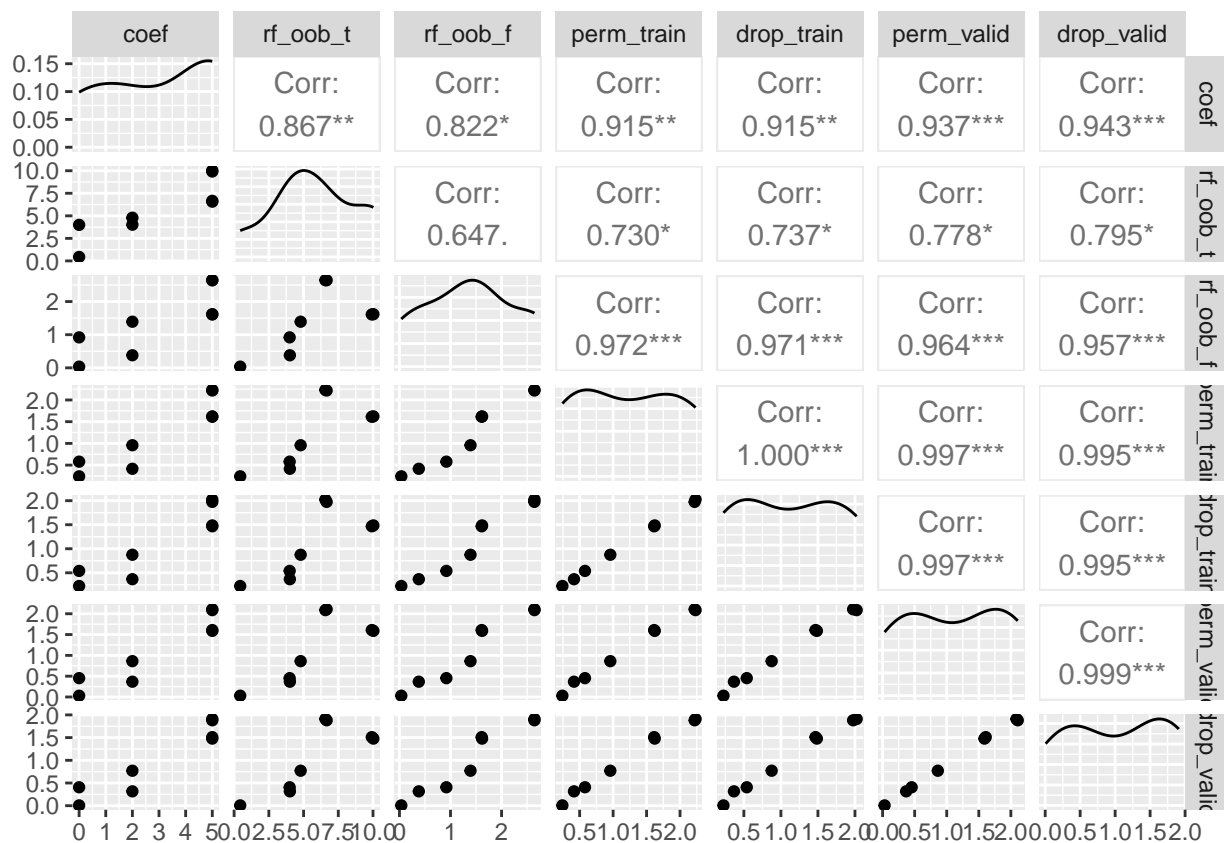
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.4880966	2.62160712	2.1637986	1.9650330	2.04136654	1.853377554
## 2	5	6.4828871	2.60085518	2.1514157	1.9152415	2.05503429	1.817446919
## 3	2	4.8320427	1.47328089	1.0318619	0.9209417	0.90392856	0.804436299
## 4	0	4.1047103	1.02326883	0.6355923	0.5892736	0.50639129	0.447546188
## 5	5	9.6998637	1.59482042	1.6081372	1.4521727	1.58160961	1.454231463
## 6	5	9.5853038	1.59079436	1.5880002	1.4373777	1.59021524	1.488282146
## 7	2	4.0069279	0.38405539	0.4323907	0.3724681	0.37761084	0.315509788
## 8	0	0.5836146	0.04872025	0.2484347	0.2303589	0.03572439	0.005784122



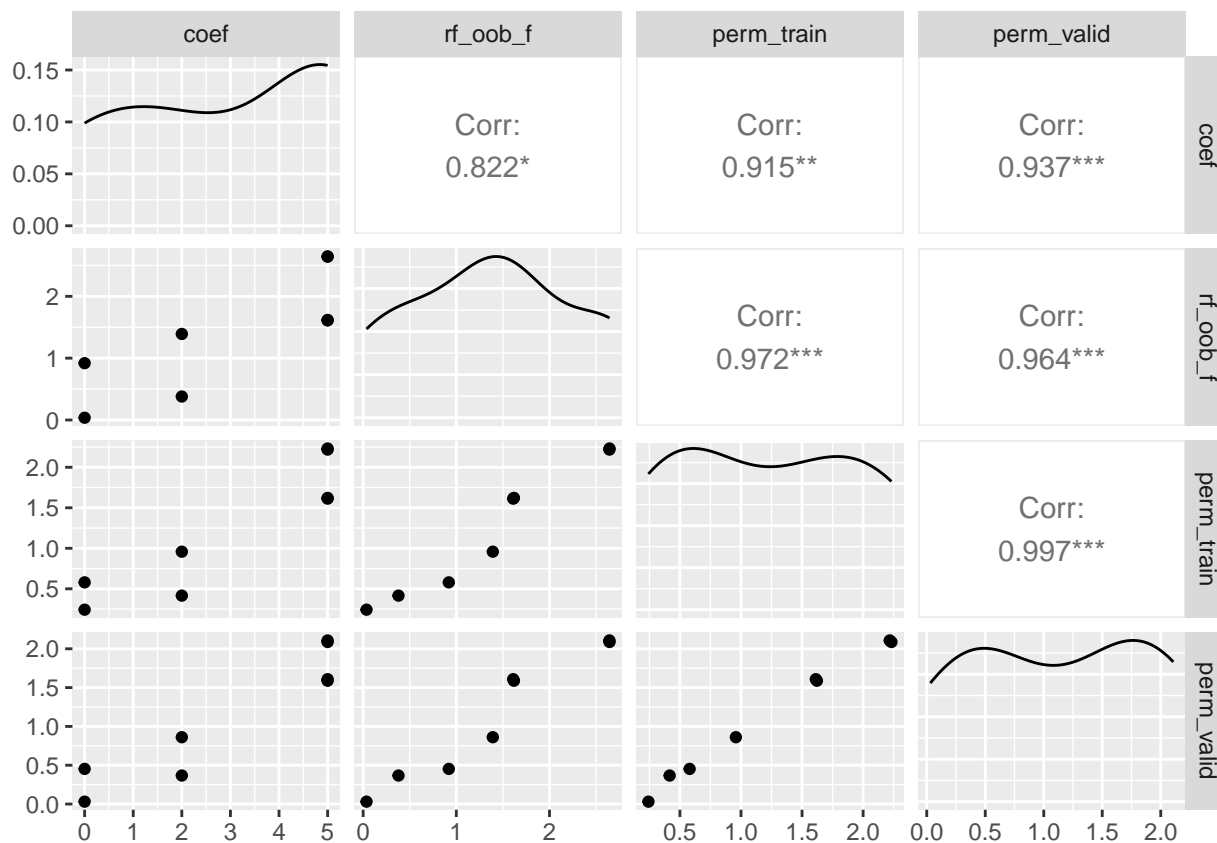
```
##      coef      rf_oob_f perm_train perm_valid
## 1      5 2.62160712  2.1637986 2.04136654
## 2      5 2.60085518  2.1514157 2.05503429
## 3      2 1.47328089  1.0318619 0.90392856
## 4      0 1.02326883  0.6355923 0.50639129
## 5      5 1.59482042  1.6081372 1.58160961
## 6      5 1.59079436  1.5880002 1.59021524
## 7      2 0.38405539  0.4323907 0.37761084
## 8      0 0.04872025  0.2484347 0.03572439
```



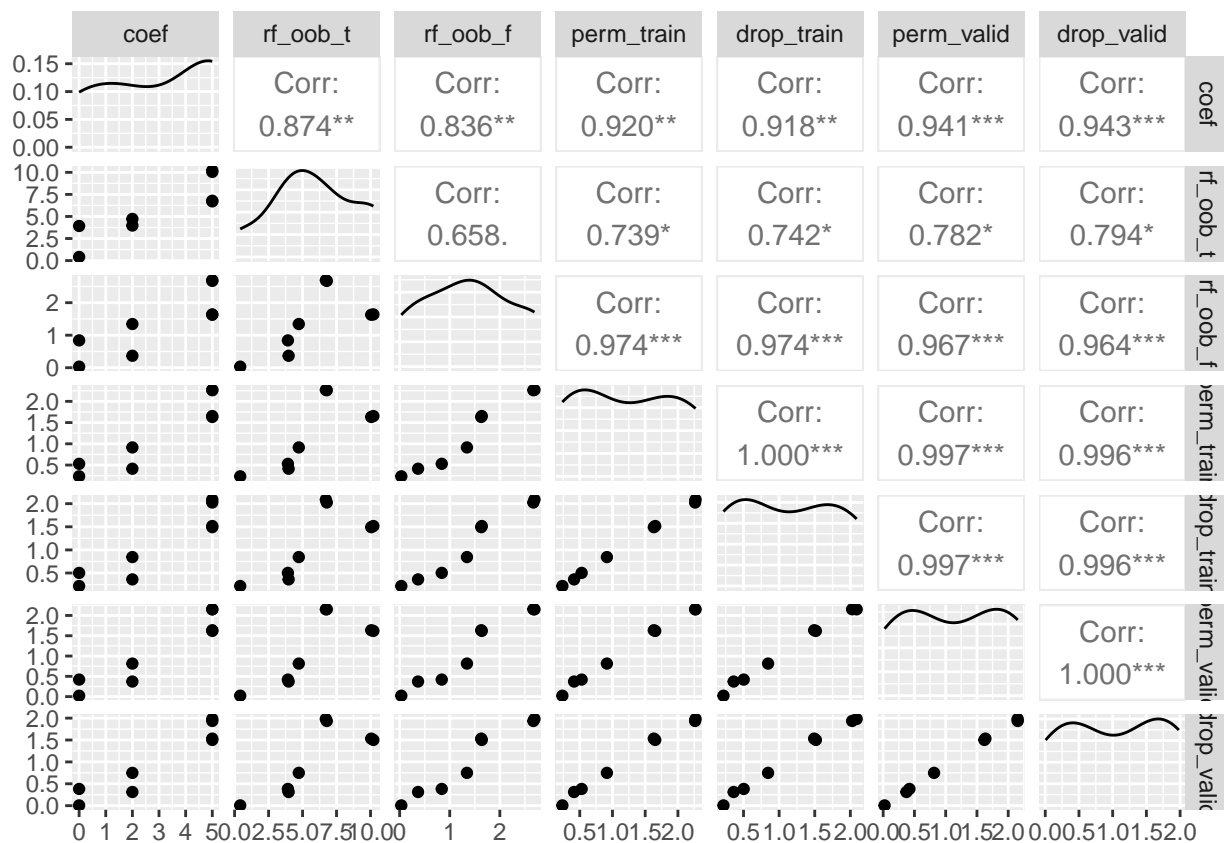
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.5708089	2.6441955	2.2319733	2.0230117	2.08756803	1.910563428
## 2	5	6.6595051	2.6432598	2.2181730	1.9753458	2.10764168	1.881175955
## 3	2	4.7913910	1.3917322	0.9582012	0.8751111	0.86114535	0.770540182
## 4	0	3.9981802	0.9193707	0.5795398	0.5390547	0.45301499	0.406283155
## 5	5	10.0255515	1.6182118	1.6218412	1.4856478	1.58864214	1.478107748
## 6	5	9.9144022	1.6106715	1.6129542	1.4674021	1.60844797	1.510298103
## 7	2	4.0152416	0.3793012	0.4157964	0.3662116	0.36733935	0.316349299
## 8	0	0.4390688	0.0363236	0.2421156	0.2246693	0.03074239	0.006873533



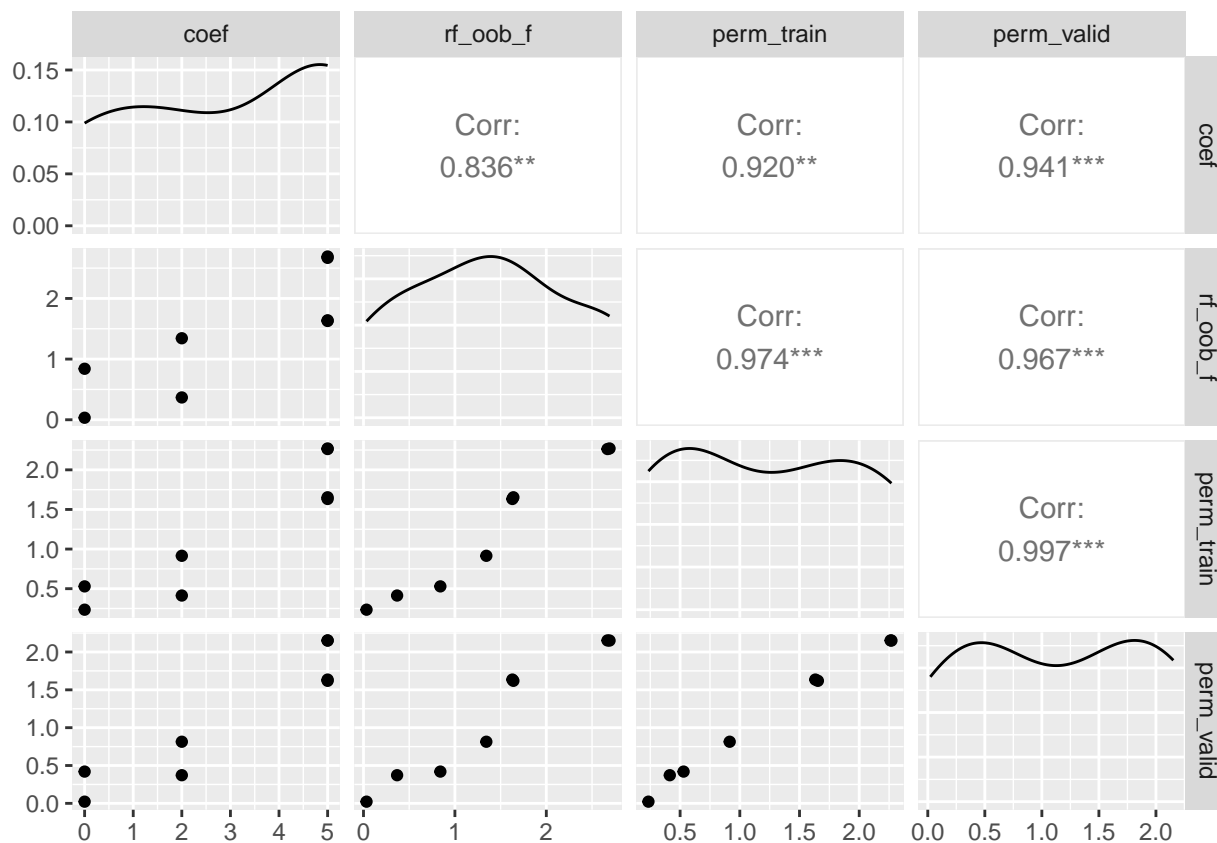
```
## coef rf_oob_f perm_train perm_valid
## 1 5 2.6441955 2.2319733 2.08756803
## 2 5 2.6432598 2.2181730 2.10764168
## 3 2 1.3917322 0.9582012 0.86114535
## 4 0 0.9193707 0.5795398 0.45301499
## 5 5 1.6182118 1.6218412 1.58864214
## 6 5 1.6106715 1.6129542 1.60844797
## 7 2 0.3793012 0.4157964 0.36733935
## 8 0 0.0363236 0.2421156 0.03074239
```



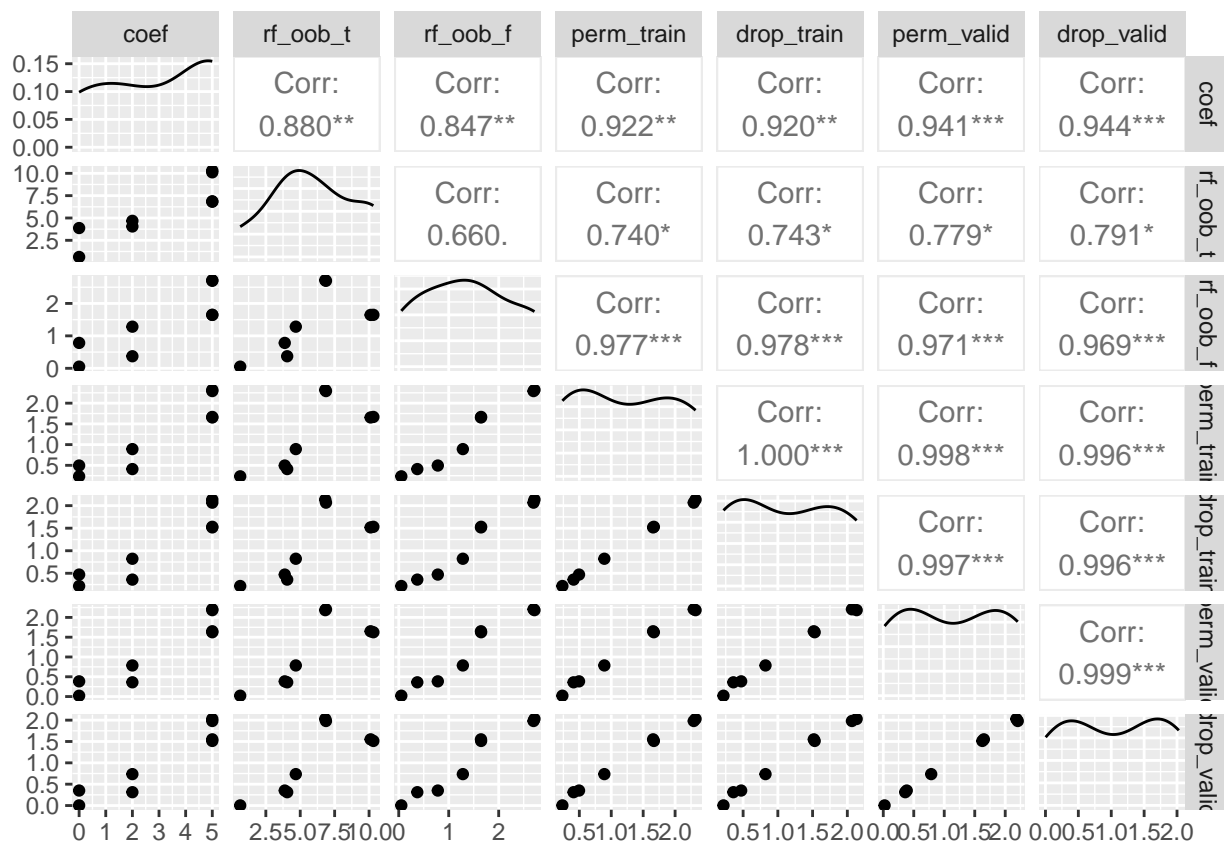
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.7137916	2.6912684	2.2696316	2.0883836	2.1511109	1.981082008
## 2	5	6.7823608	2.6692731	2.2625859	2.0247437	2.1525692	1.935331070
## 3	2	4.7242001	1.3436451	0.9151075	0.8446747	0.8139681	0.749559755
## 4	0	3.9220161	0.8408415	0.5291089	0.5027549	0.4198702	0.382527770
## 5	5	10.2098715	1.6395870	1.6535763	1.5159807	1.6191902	1.502278095
## 6	5	10.0524600	1.6286994	1.6322134	1.4920606	1.6354264	1.532103543
## 7	2	3.9745162	0.3676453	0.4135532	0.3618604	0.3720432	0.312272726
## 8	0	0.4132125	0.0331461	0.2347434	0.2188792	0.0225081	0.006418321



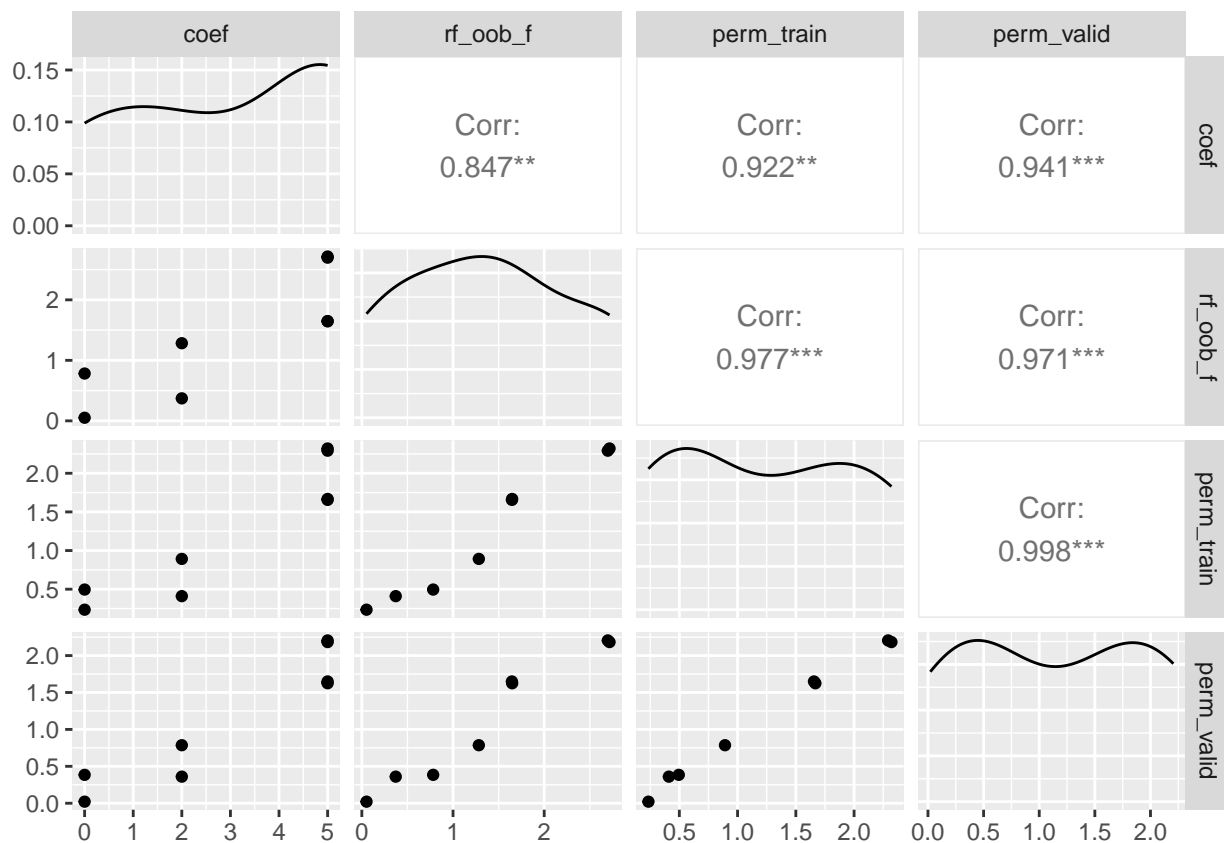
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  2.6912684  2.2696316  2.1511109
## 2      5  2.6692731  2.2625859  2.1525692
## 3      2  1.3436451  0.9151075  0.8139681
## 4      0  0.8408415  0.5291089  0.4198702
## 5      5  1.6395870  1.6535763  1.6191902
## 6      5  1.6286994  1.6322134  1.6354264
## 7      2  0.3676453  0.4135532  0.3720432
## 8      0  0.0331461  0.2347434  0.0225081
```



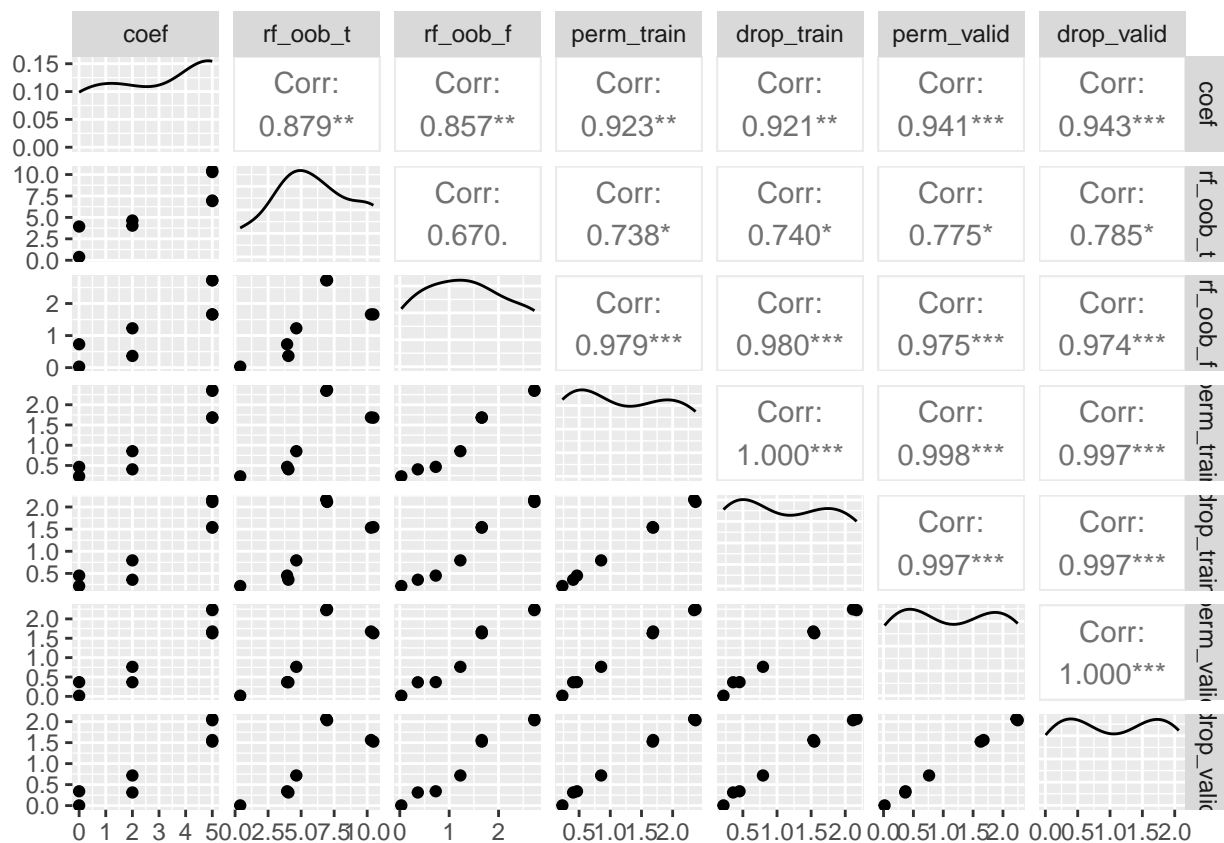
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.8174930	2.71612279	2.3207404	2.1342472	2.18323374	2.02778880
## 2	5	6.8693700	2.69740396	2.2922895	2.0664034	2.20544572	1.98021114
## 3	2	4.6802243	1.28314993	0.8919948	0.8222686	0.78598095	0.73575245
## 4	0	3.8747358	0.78231840	0.4952265	0.4717550	0.38459796	0.35096506
## 5	5	10.3141009	1.64851580	1.6677458	1.5302844	1.62511056	1.51304754
## 6	5	10.1119874	1.64568691	1.6558245	1.5174340	1.64950173	1.55018004
## 7	2	4.0513190	0.37187688	0.4100156	0.3599206	0.36027523	0.31259289
## 8	0	0.6390598	0.05141346	0.2347242	0.2175952	0.02136585	0.00308643



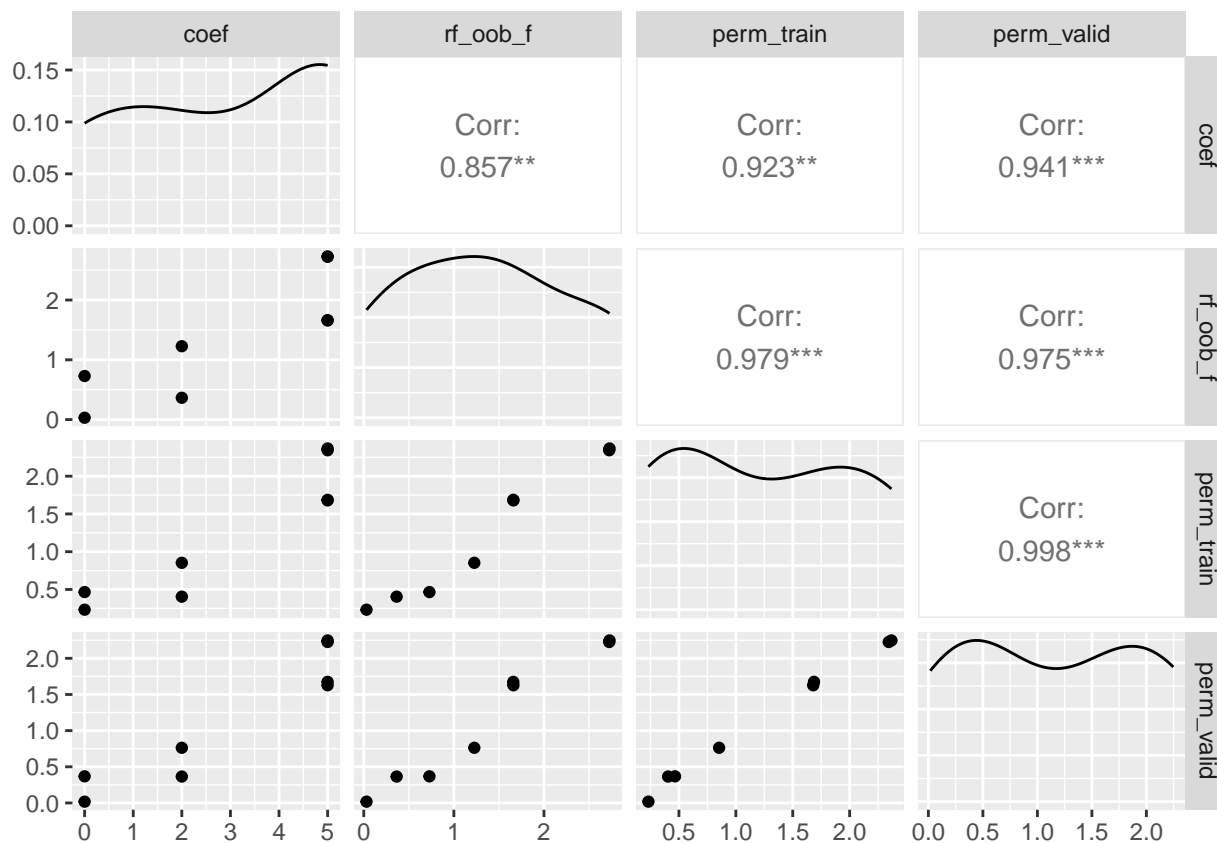
```
##      coef    rf_oob_f perm_train perm_valid
## 1      5 2.71612279  2.3207404 2.18323374
## 2      5 2.69740396  2.2922895 2.20544572
## 3      2 1.28314993  0.8919948 0.78598095
## 4      0 0.78231840  0.4952265 0.38459796
## 5      5 1.64851580  1.6677458 1.62511056
## 6      5 1.64568691  1.6558245 1.64950173
## 7      2 0.37187688  0.4100156 0.36027523
## 8      0 0.05141346  0.2347242 0.02136585
```



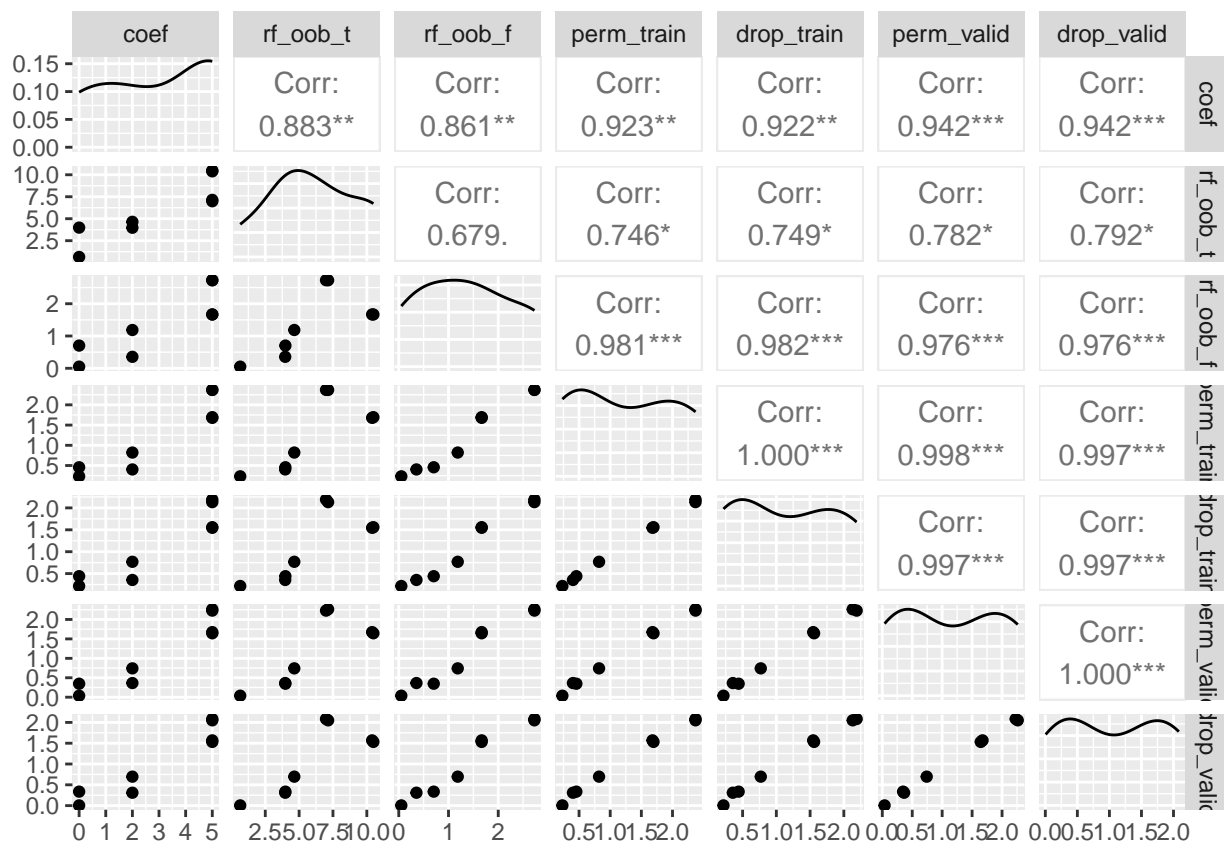
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.9058957	2.72389632	2.3433426	2.1669046	2.22601791	2.063774829
## 2	5	6.9672482	2.72748948	2.3669543	2.1128130	2.24743511	2.033363005
## 3	2	4.6384375	1.22696082	0.8531060	0.7943170	0.76295965	0.717846137
## 4	0	3.9371956	0.72881817	0.4658053	0.4509495	0.36811361	0.341270925
## 5	5	10.4700523	1.66088762	1.6794615	1.5460709	1.62769077	1.524241777
## 6	5	10.2881458	1.65948106	1.6867202	1.5325478	1.67468059	1.561598250
## 7	2	4.0362461	0.36498896	0.4047672	0.3574144	0.36508893	0.312745296
## 8	0	0.3873326	0.03067535	0.2321874	0.2159703	0.01782771	0.004845259



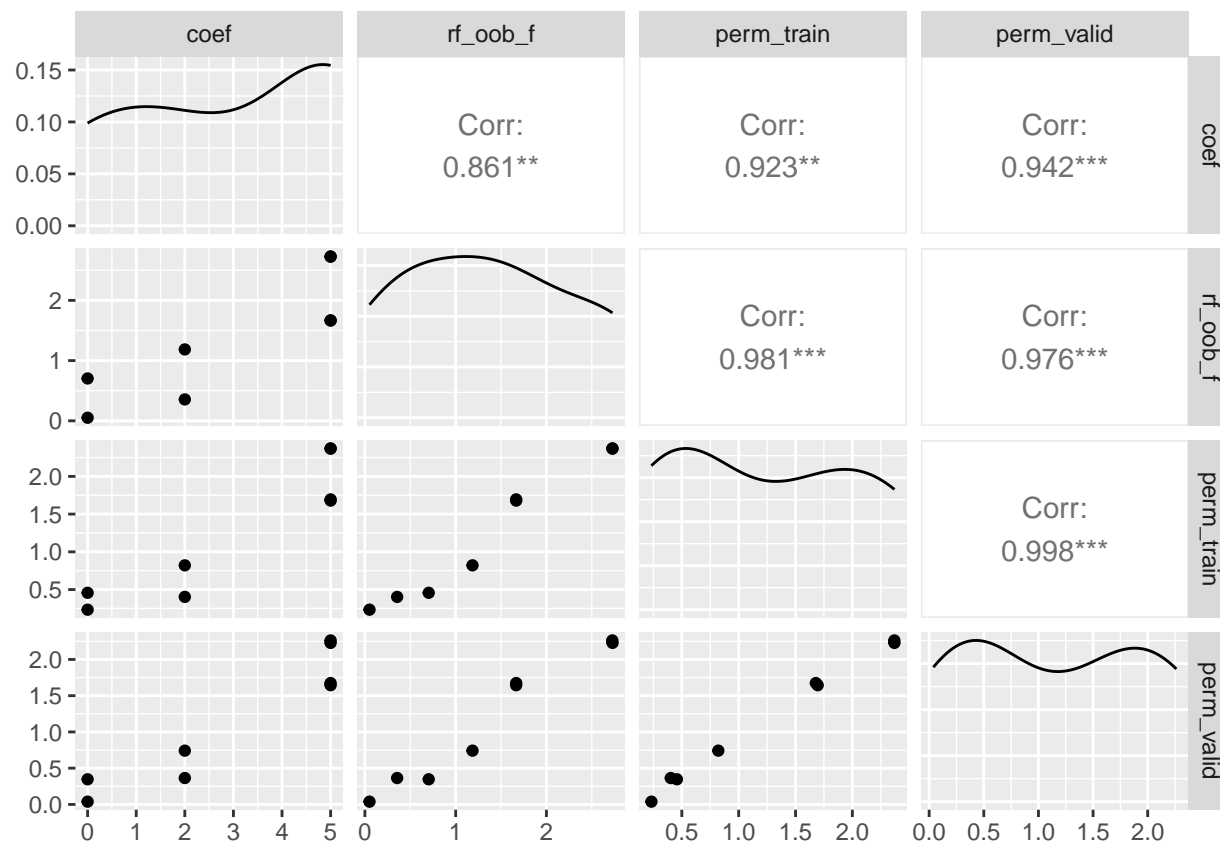
```
##      coef    rf_oob_f perm_train perm_valid
## 1      5 2.72389632  2.3433426 2.22601791
## 2      5 2.72748948  2.3669543 2.24743511
## 3      2 1.22696082  0.8531060 0.76295965
## 4      0 0.72881817  0.4658053 0.36811361
## 5      5 1.66088762  1.6794615 1.62769077
## 6      5 1.65948106  1.6867202 1.67468059
## 7      2 0.36498896  0.4047672 0.36508893
## 8      0 0.03067535  0.2321874 0.01782771
```



##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.9888291	2.72831584	2.3700187	2.1896874	2.22826367	2.083359283
## 2	5	7.1423652	2.72646639	2.3692144	2.1312496	2.26221131	2.050820354
## 3	2	4.6450102	1.18575131	0.8208566	0.7693898	0.74189573	0.696502954
## 4	0	3.9866866	0.70327887	0.4568263	0.4411822	0.34706820	0.336389319
## 5	5	10.4846720	1.66660597	1.6951411	1.5590400	1.64463793	1.531967474
## 6	5	10.3907685	1.66711242	1.6789708	1.5456231	1.67429284	1.568635633
## 7	2	3.9733702	0.35546257	0.4025704	0.3551615	0.36431457	0.309874802
## 8	0	0.6489715	0.05110907	0.2327085	0.2167247	0.03833951	0.005449022



```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  2.72831584  2.3700187  2.22826367
## 2      5  2.72646639  2.3692144  2.26221131
## 3      2  1.18575131  0.8208566  0.74189573
## 4      0  0.70327887  0.4568263  0.34706820
## 5      5  1.66660597  1.6951411  1.64463793
## 6      5  1.66711242  1.6789708  1.67429284
## 7      2  0.35546257  0.4025704  0.36431457
## 8      0  0.05110907  0.2327085  0.03833951
```



```
rsq
```

```
## [1] 0.8121291 0.8676367 0.8839176 0.8921492 0.8968934 0.8999322 0.9024276
## [8] 0.9046362 0.9065682 0.9071677 0.9078156 0.9076545
```

```
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)

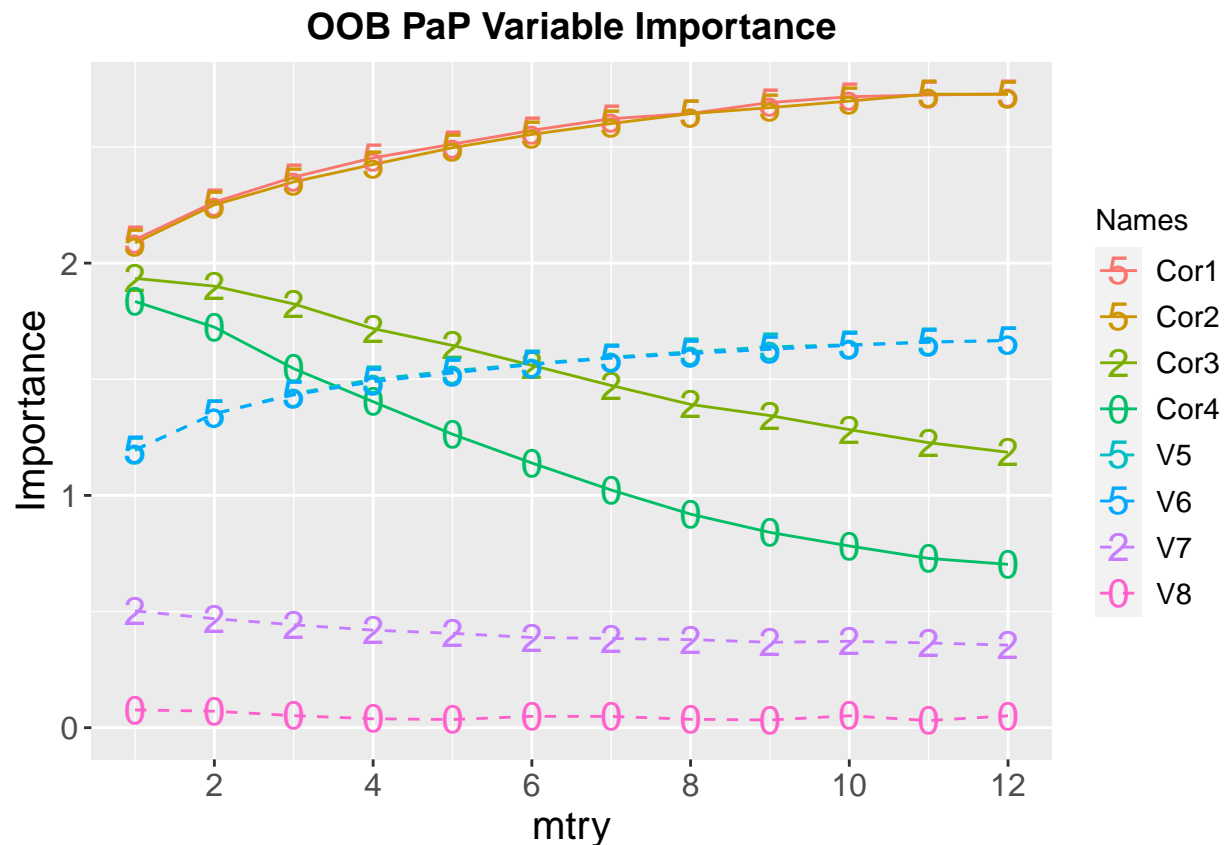
colnames(perm_train1)[1:12] <- 1:12
perm_train1 <- perm_train1 %>%
  pivot_longer(!c(Names,mag), names_to = "mtry", values_to = "Imp")
perm_train1$mtry <- as.numeric(perm_train1$mtry)

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)

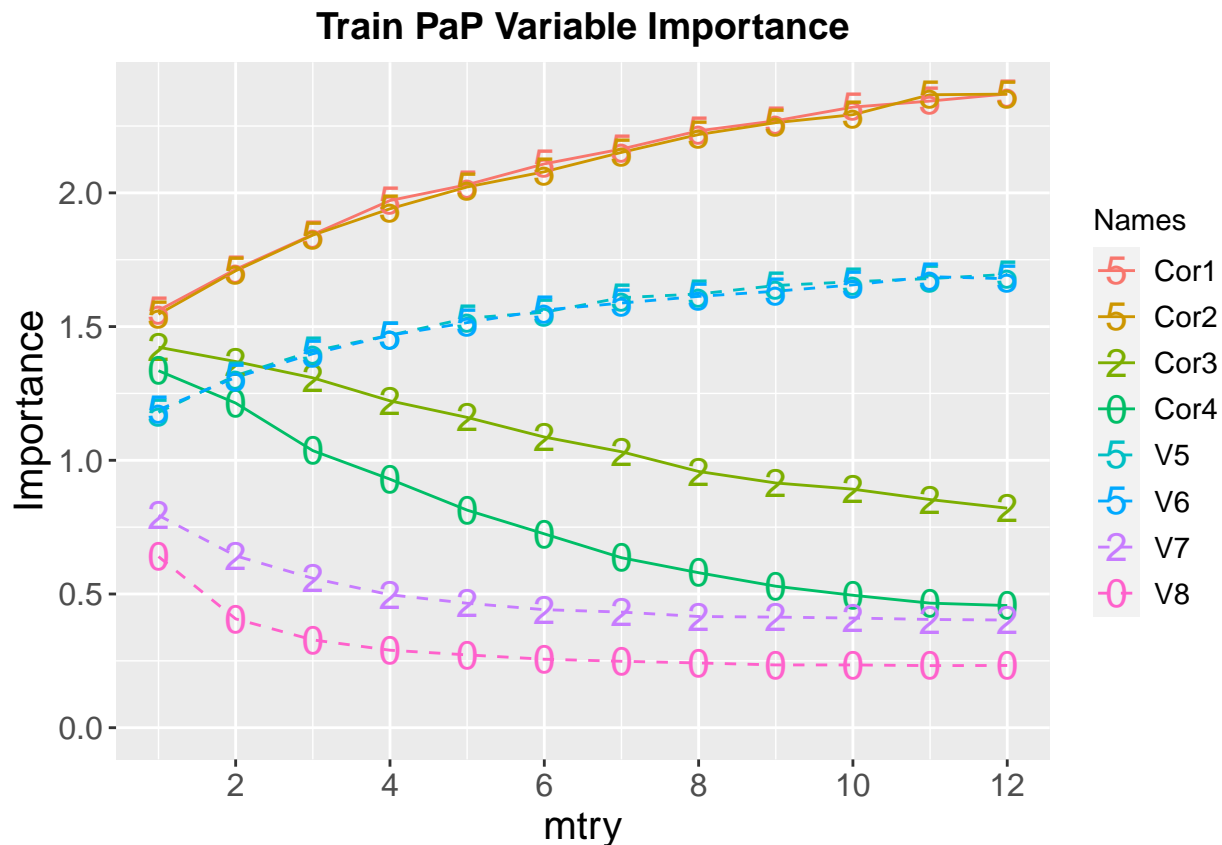
colnames(perm_valid1)[1:12] <- 1:12
perm_valid1 <- perm_valid1 %>%
  pivot_longer(!c(Names,mag), names_to = "mtry", values_to = "Imp")
perm_valid1$mtry <- as.numeric(perm_valid1$mtry)

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

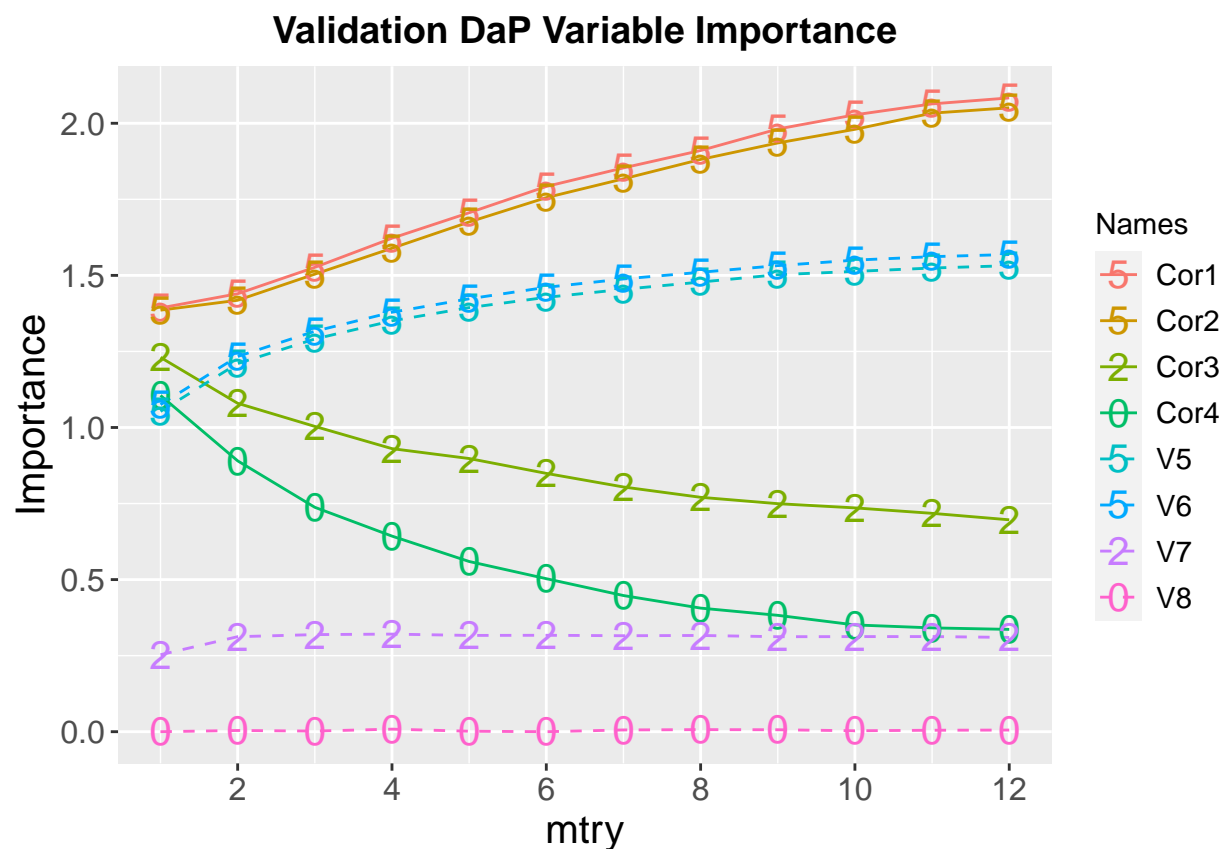
```



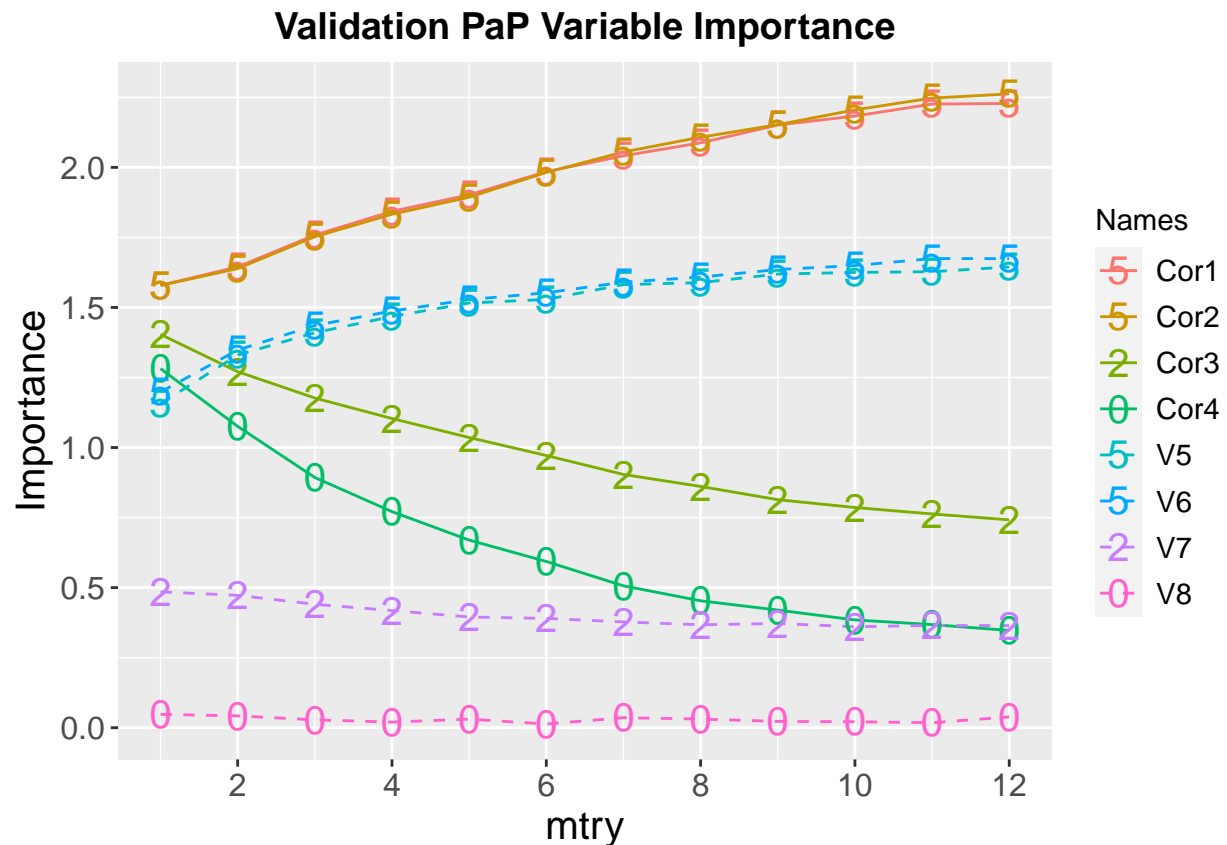
```
gp <- perm_train1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,max(perm_train1$Imp))) +
  ggtitle("Train 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
```

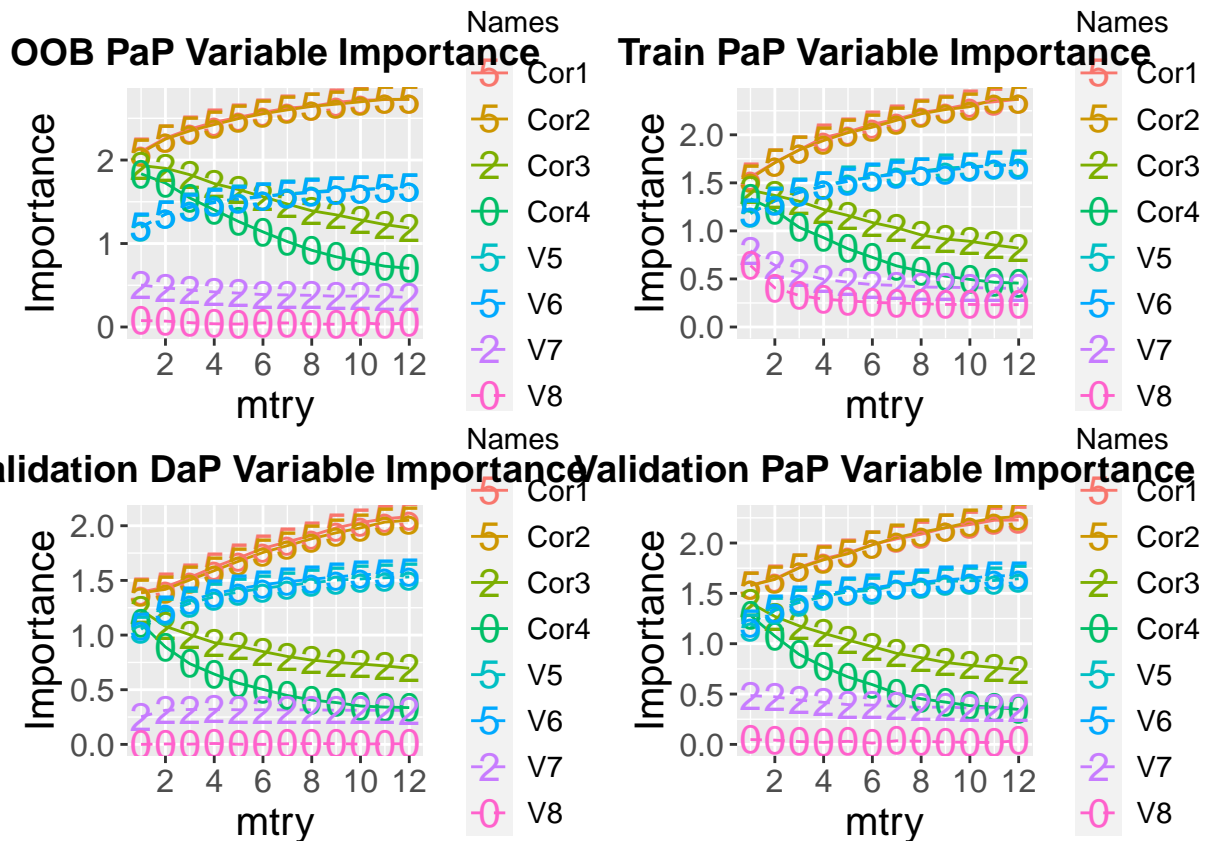


```
gv <- perm_valid1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,max(perm_valid1$Imp))) +
  ggtitle("Validation 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
```

```
# 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
```

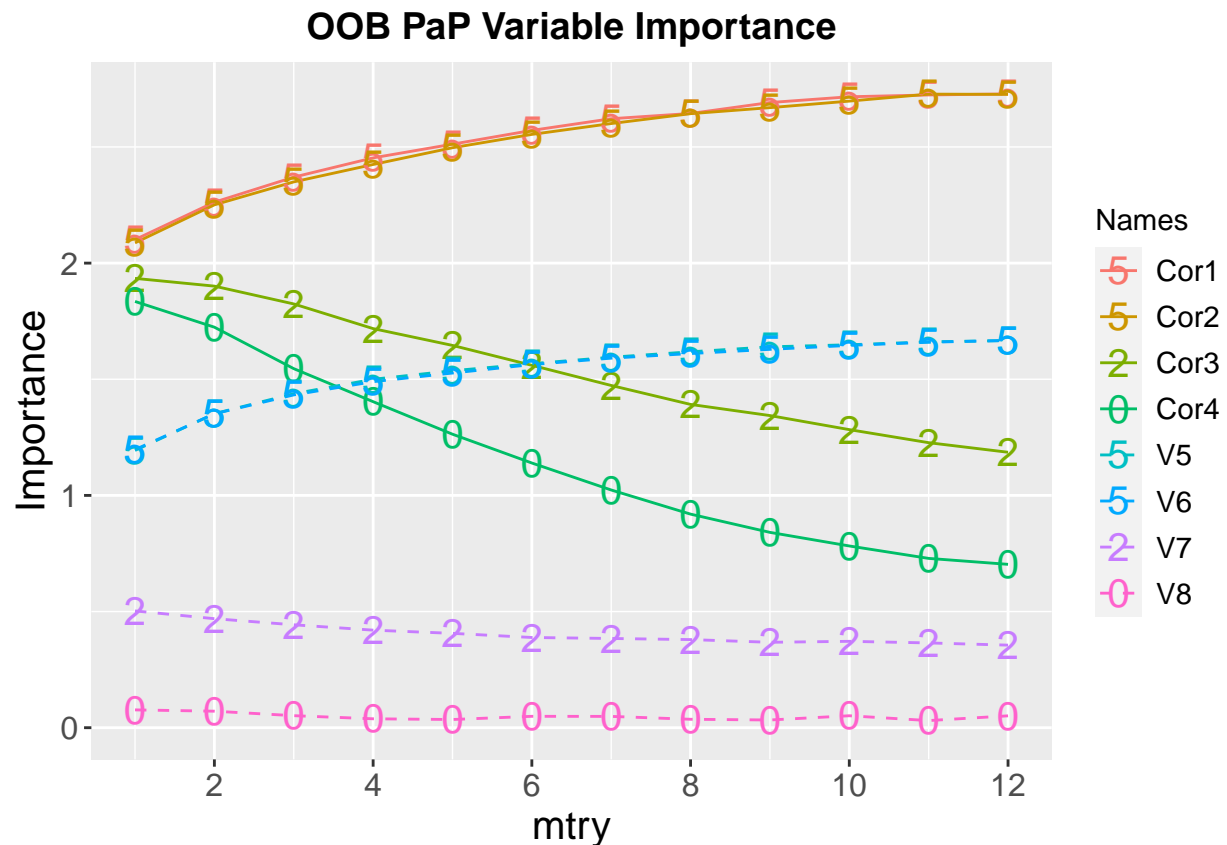


```
ggsave("x2_all_zoom.pdf", plot = gr + gp + gd + gv, dpi = 2400,
       width = 9, height = 9)
```

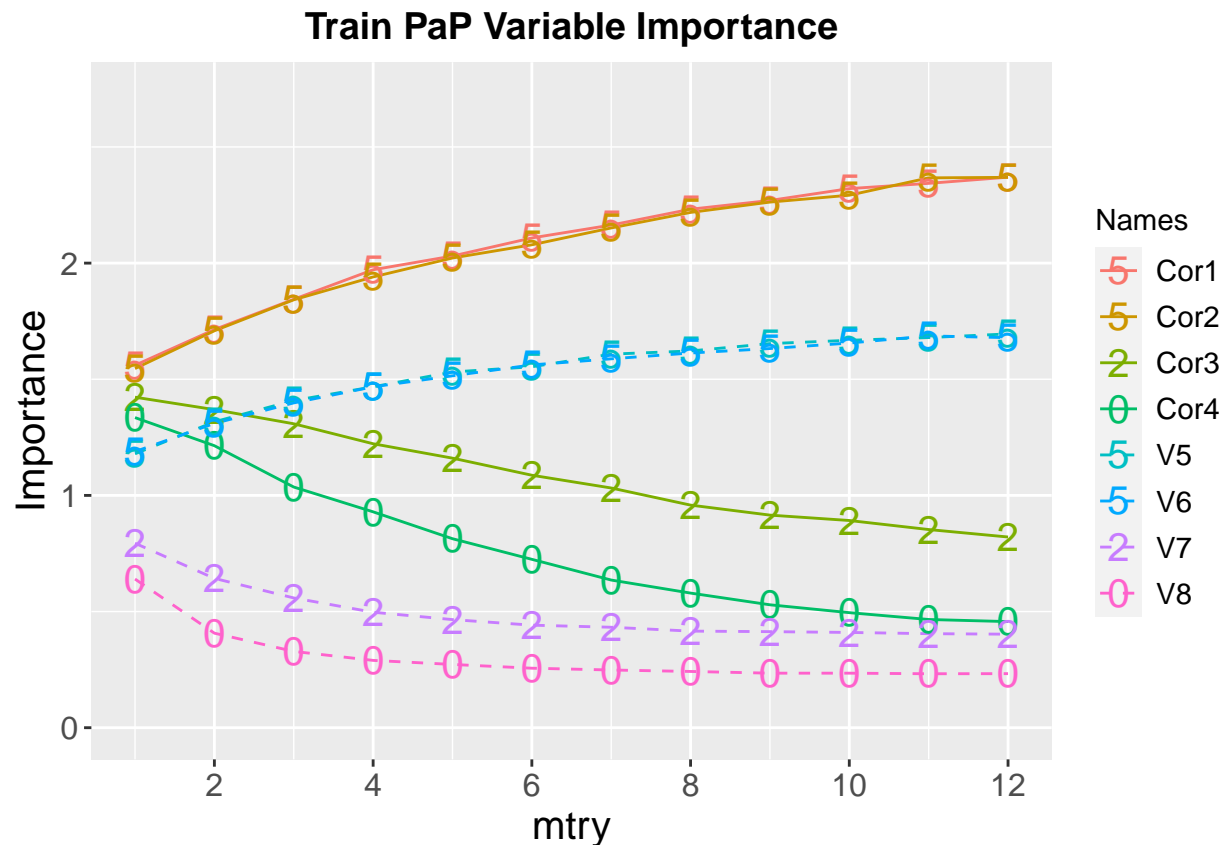
```
ma = max(rf_oob_f1$Imp, perm_train1$Imp, perm_valid1$Imp, drop_valid1$Imp)
# mp = max(rf_pdp1$Imp)

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
```

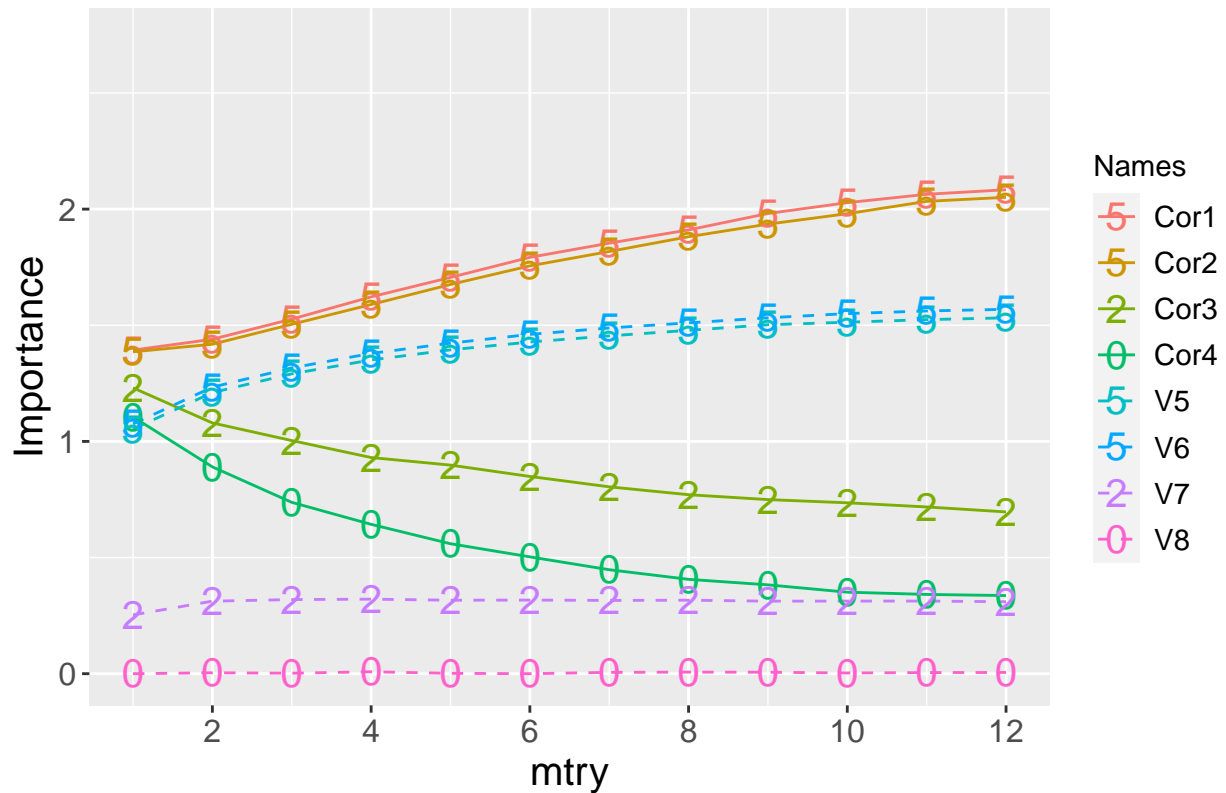


```
gp <- perm_train1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,ma)) +
  ggtitle("Train PaP Variable 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
```

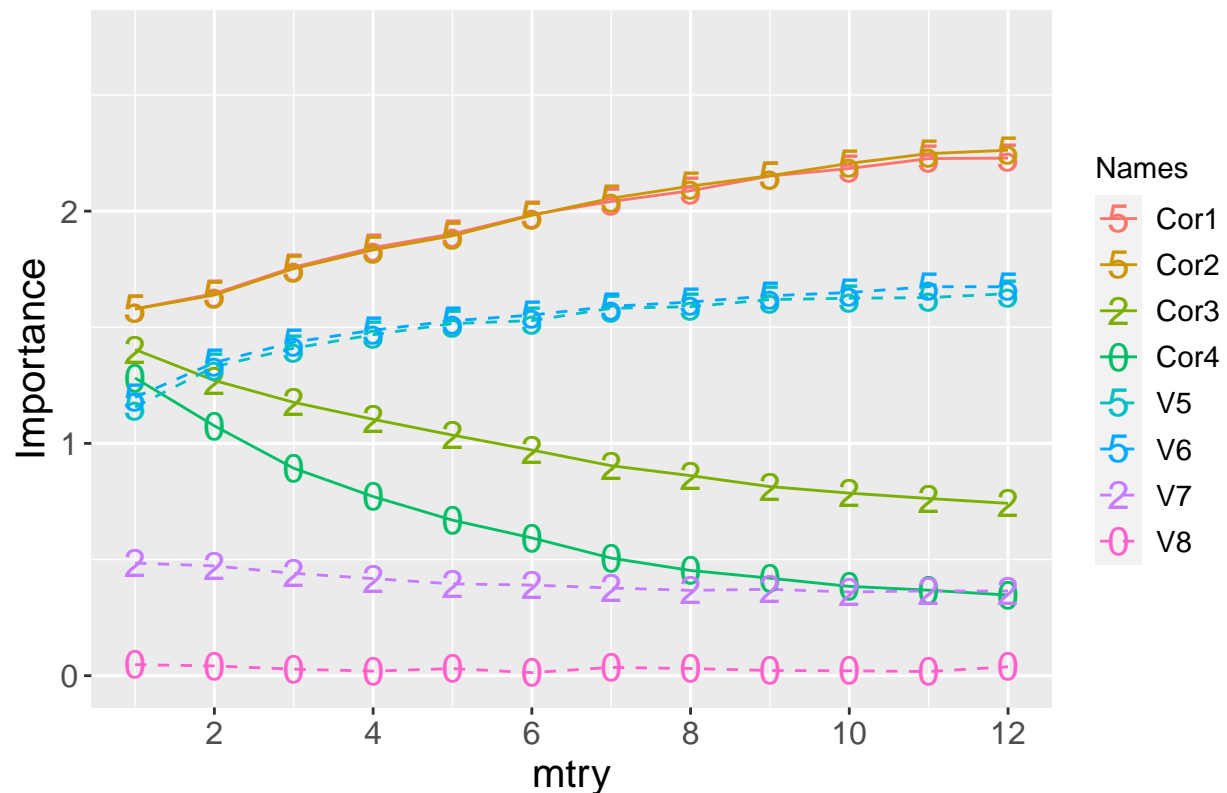
Validation DaP Variable Importance



```
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)
```

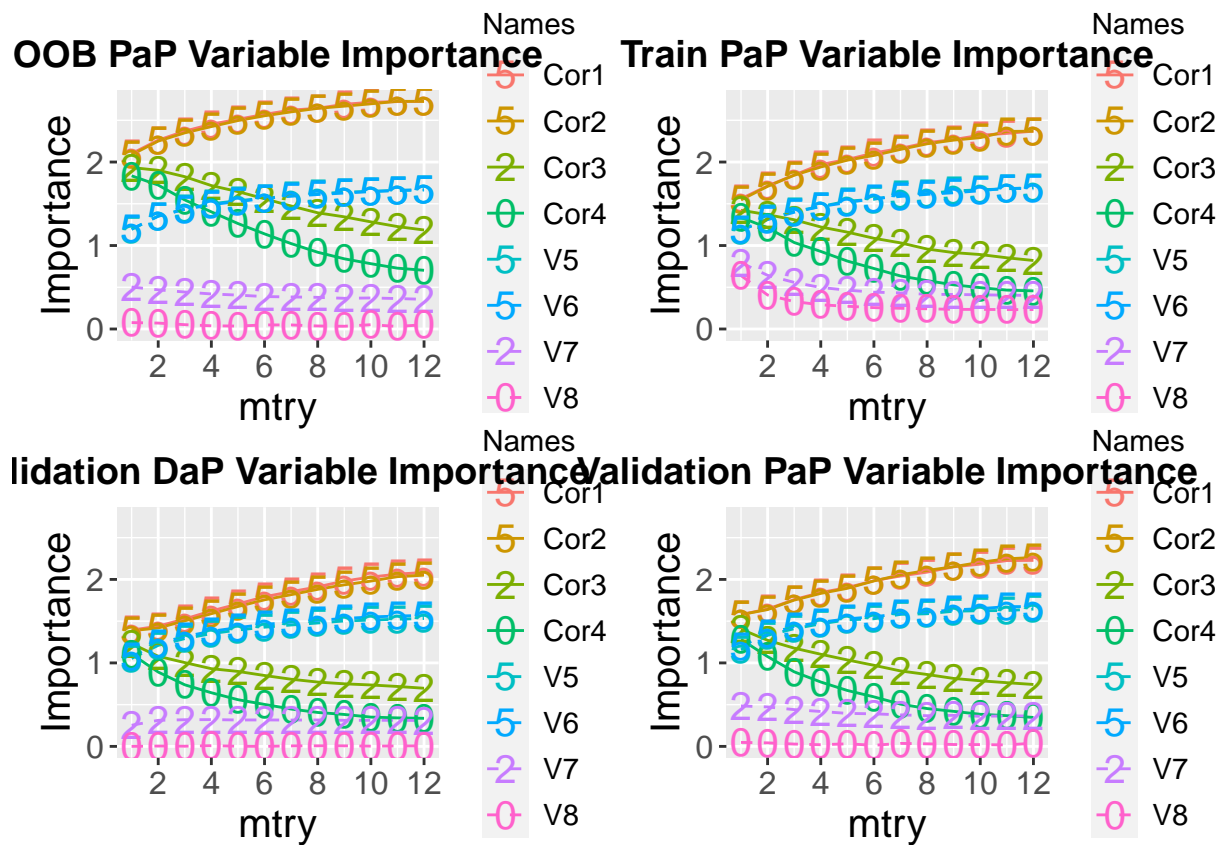
gv

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("x2_all.pdf", plot = gr + gp + gd + gv, dpi = 2400,
       width = 9, height = 9)
```

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

```
## Time difference of 20.56767 mins
```