

permute_rf_strobl_cos_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)*8*pi

  y <- 5 * cos(strobl[, 1]) + 5 * cos(strobl[, 2]) + 2 * cos(strobl[, 3]) +
    5 * cos(strobl[, 5]) + 5 * cos(strobl[, 6]) + 2 * cos(strobl[, 7])
  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)*8*pi
y <- 5 * cos(dfv[, 1]) + 5 * cos(dfv[, 2]) + 2 * cos(dfv[, 3]) +
  5 * cos(dfv[, 5]) + 5 * cos(dfv[, 6]) + 2 * cos(dfv[, 7])
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])

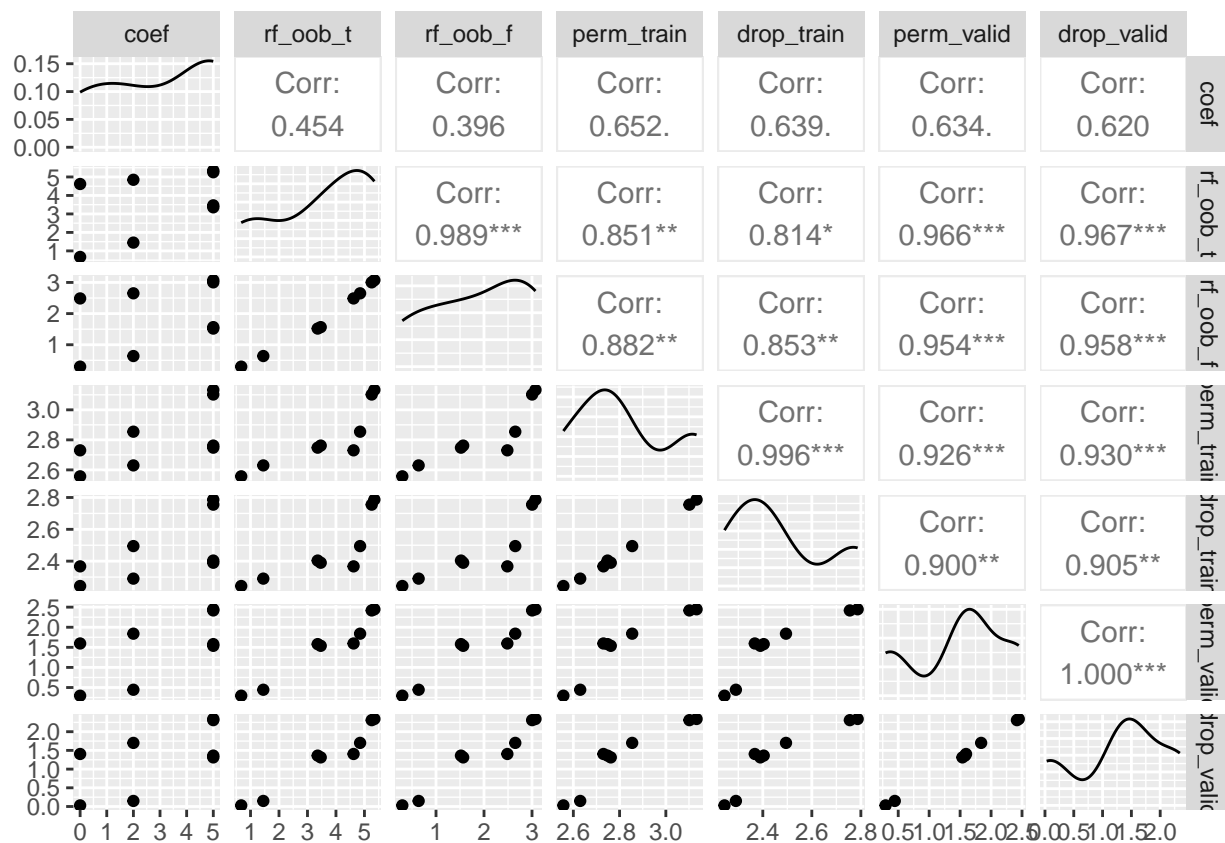
  print(sdf)
  print(ggpairs(sdf))

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

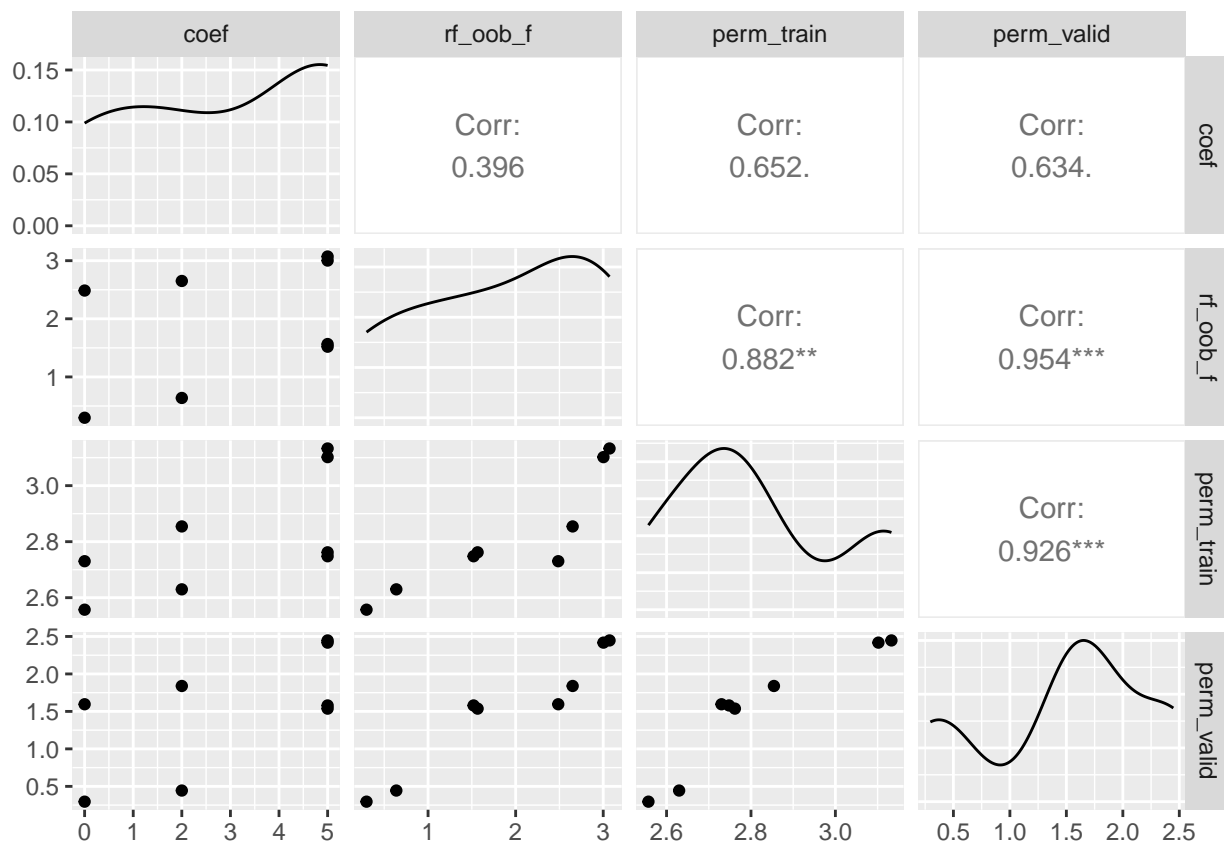
  print(sdf)
  print(ggpairs(sdf))
}

```

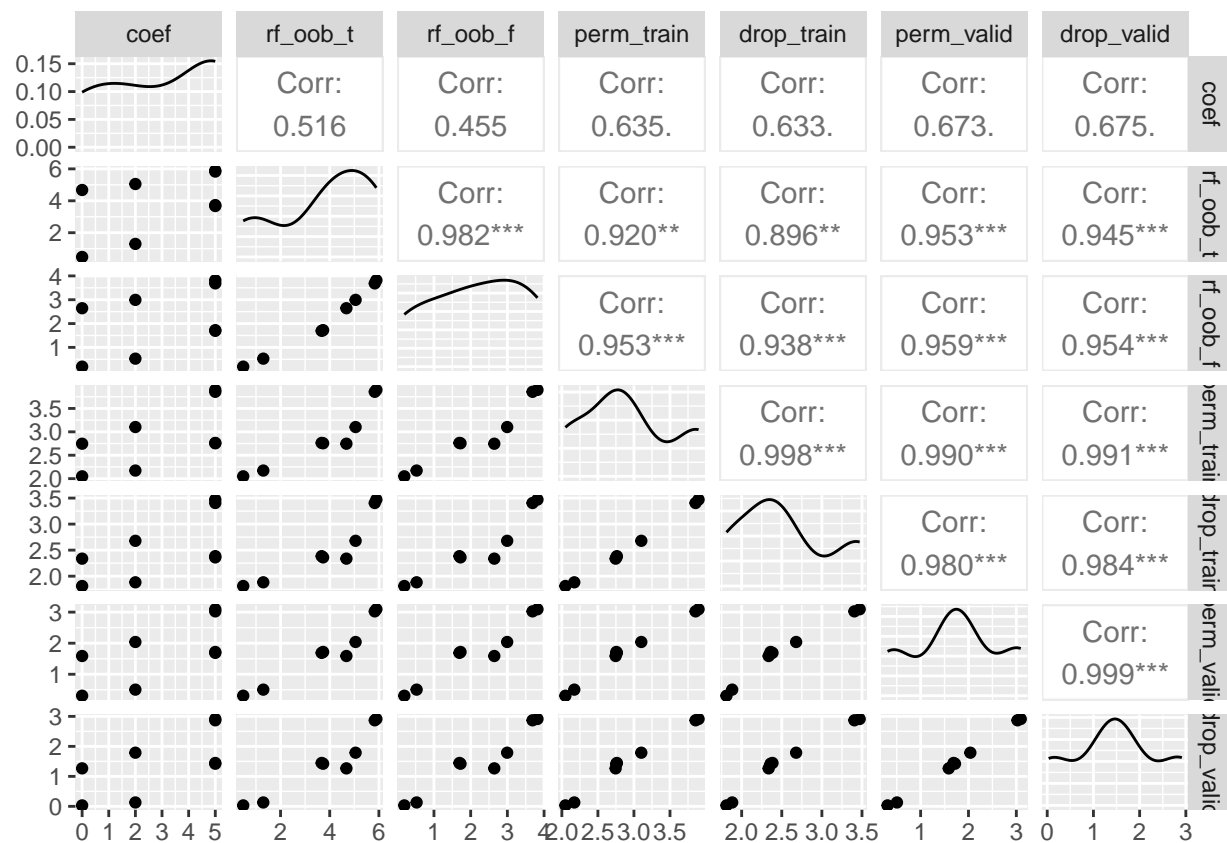
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.3492382	3.0718141	3.132773	2.786894	2.4471267	2.34072091
## 2	5	5.2555940	3.0023750	3.101917	2.755087	2.4181123	2.30920839
## 3	2	4.8478470	2.6509077	2.854390	2.494609	1.8395391	1.70081555
## 4	0	4.6188586	2.4864278	2.730186	2.367339	1.5955607	1.40447507
## 5	5	3.3583371	1.5198267	2.747933	2.403665	1.5797771	1.36131882
## 6	5	3.4646971	1.5657523	2.761906	2.390058	1.5372139	1.31305681
## 7	2	1.4492161	0.6379204	2.629774	2.290226	0.4443253	0.14877124
## 8	0	0.6734467	0.2976441	2.557050	2.243701	0.2958607	0.02972449



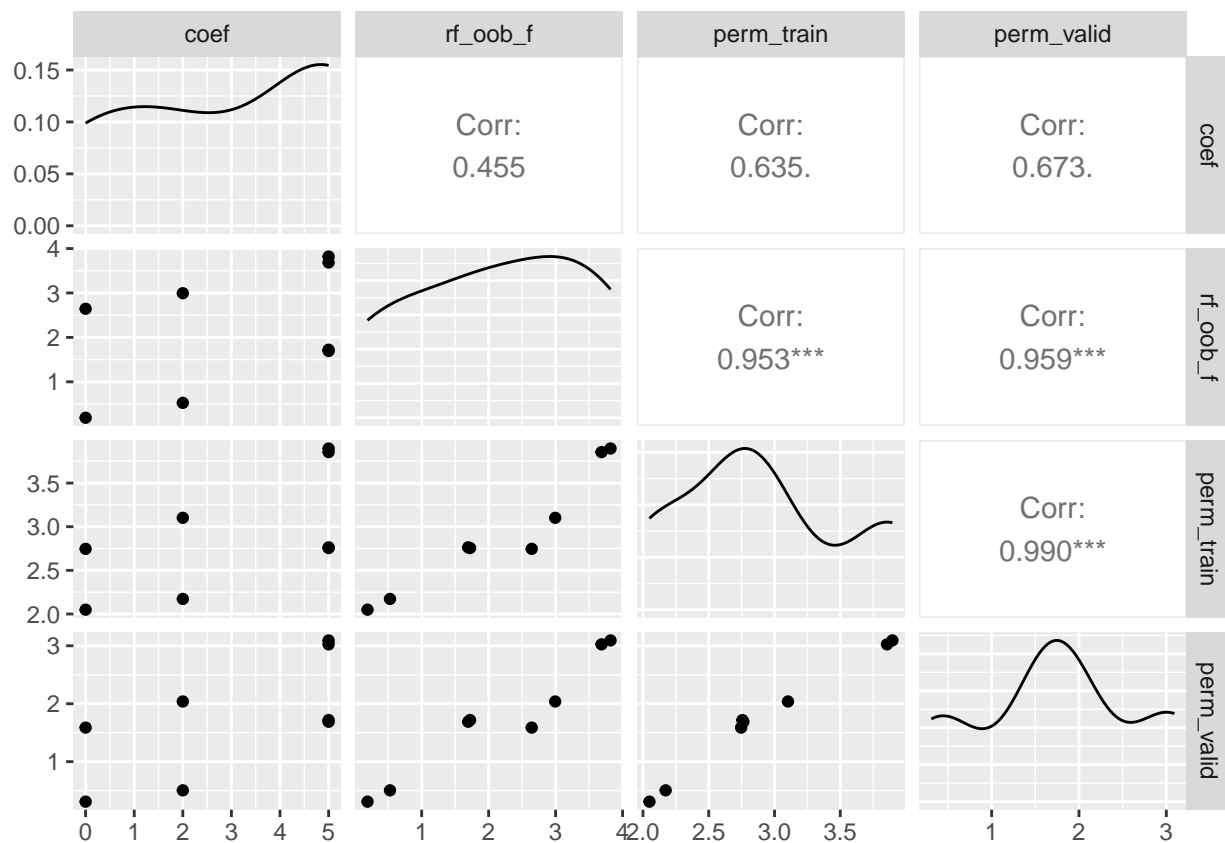
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5 3.0718141   3.132773  2.4471267
## 2      5 3.0023750   3.101917  2.4181123
## 3      2 2.6509077   2.854390  1.8395391
## 4      0 2.4864278   2.730186  1.5955607
## 5      5 1.5198267   2.747933  1.5797771
## 6      5 1.5657523   2.761906  1.5372139
## 7      2 0.6379204   2.629774  0.4443253
## 8      0 0.2976441   2.557050  0.2958607
```



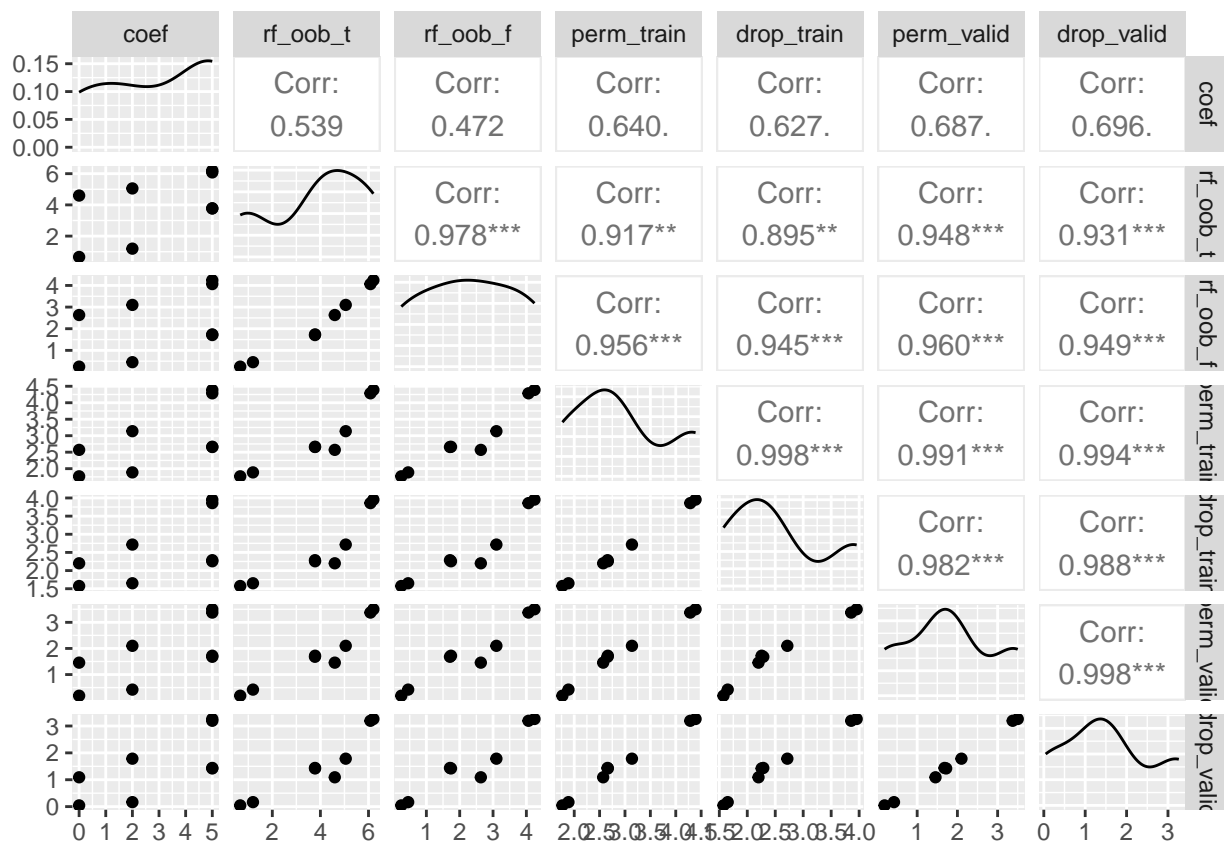
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	5.898014	3.8200511	3.896417	3.470851	3.0922245	2.91340601
## 2	5	5.832425	3.6858997	3.855050	3.405375	3.0243537	2.86900550
## 3	2	5.052200	2.9943599	3.102319	2.679148	2.0387708	1.78868062
## 4	0	4.678204	2.6436970	2.746196	2.336721	1.5877478	1.26647581
## 5	5	3.671519	1.6947179	2.763481	2.384269	1.6887480	1.44652100
## 6	5	3.731330	1.7227453	2.755953	2.360391	1.7178230	1.42106443
## 7	2	1.297920	0.5272373	2.172612	1.882038	0.5051816	0.12999569
## 8	0	0.480974	0.1917718	2.049221	1.810921	0.3094841	0.03185021



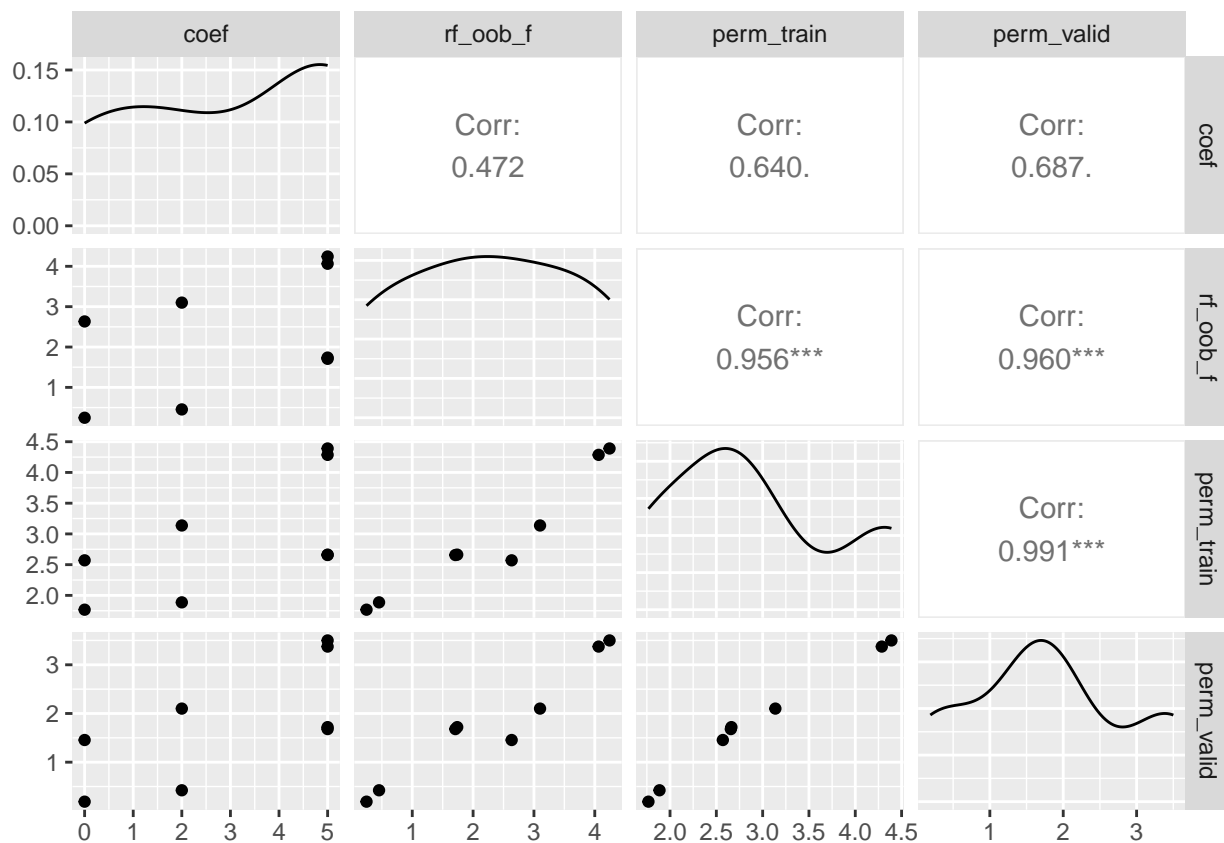
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  3.8200511   3.896417  3.0922245
## 2      5  3.6858997   3.855050  3.0243537
## 3      2  2.9943599   3.102319  2.0387708
## 4      0  2.6436970   2.746196  1.5877478
## 5      5  1.6947179   2.763481  1.6887480
## 6      5  1.7227453   2.755953  1.7178230
## 7      2  0.5272373   2.172612  0.5051816
## 8      0  0.1917718   2.049221  0.3094841
```



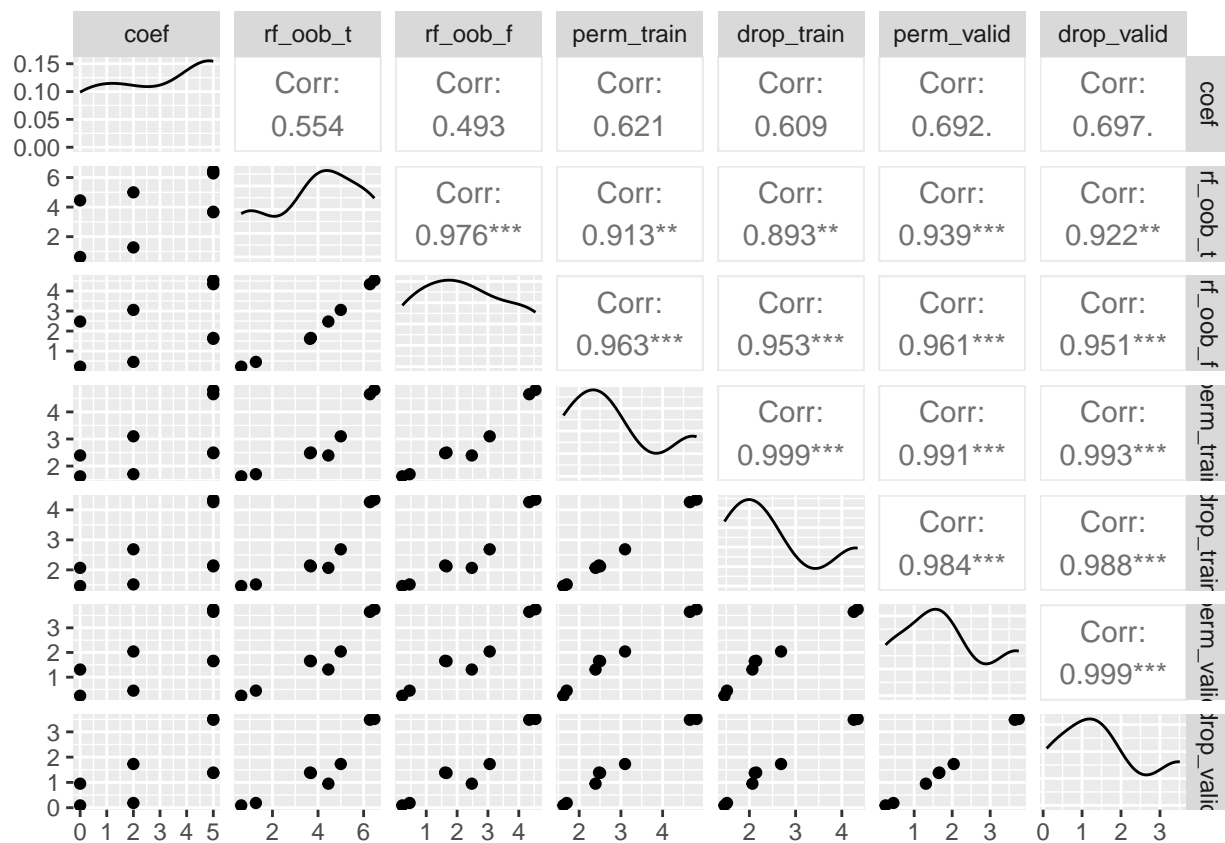
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.2030586	4.2427355	4.390200	3.955664	3.5003846	3.26220208
## 2	5	6.0808269	4.0634196	4.286134	3.858348	3.3745423	3.19403937
## 3	2	5.0538449	3.1001852	3.137226	2.716189	2.1007229	1.78428717
## 4	0	4.5983106	2.6338748	2.570211	2.200632	1.4553501	1.09121503
## 5	5	3.7740665	1.7082580	2.656811	2.286207	1.6801846	1.43825566
## 6	5	3.7790542	1.7385309	2.662559	2.256798	1.7202176	1.42179256
## 7	2	1.1960618	0.4537062	1.886498	1.650082	0.4233554	0.16570031
## 8	0	0.6652057	0.2478392	1.767018	1.574510	0.1894171	0.04488649



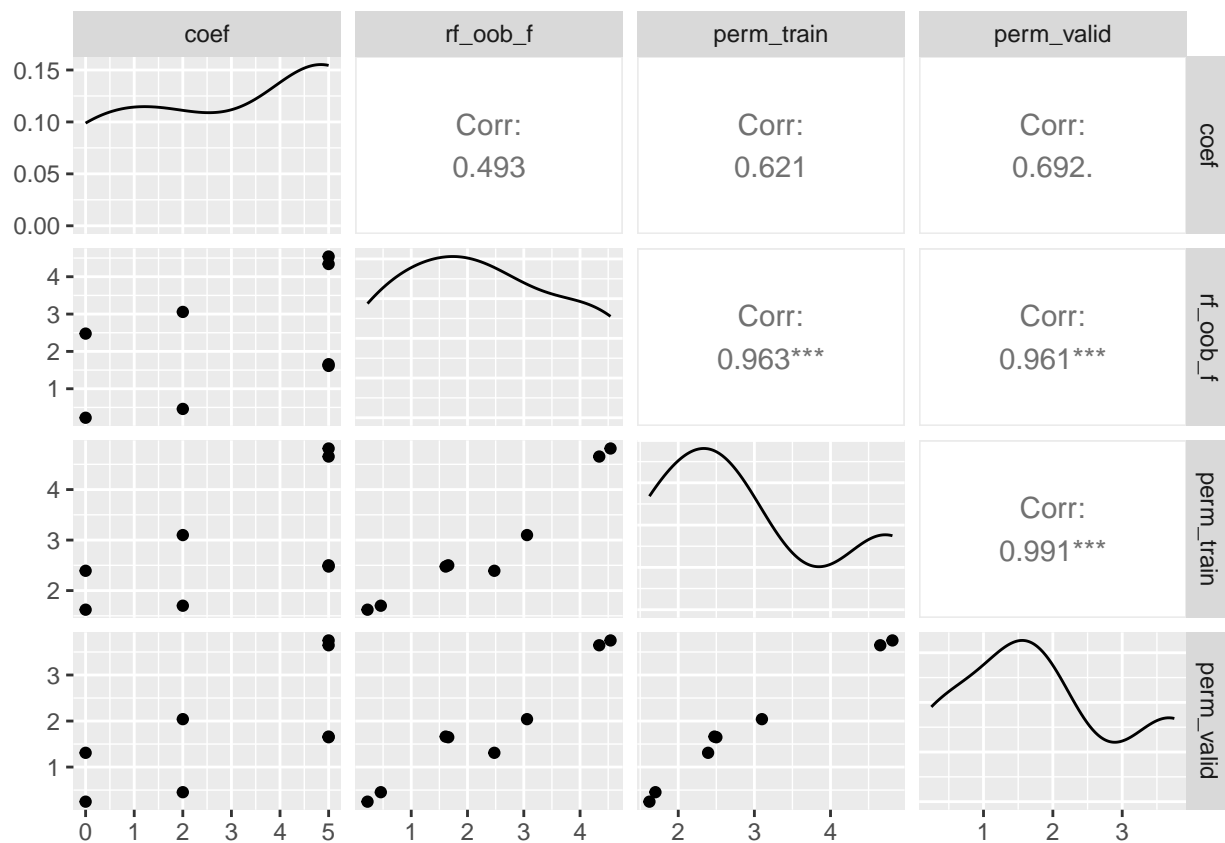
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  4.2427355  4.390200  3.5003846
## 2      5  4.0634196  4.286134  3.3745423
## 3      2  3.1001852  3.137226  2.1007229
## 4      0  2.6338748  2.570211  1.4553501
## 5      5  1.7082580  2.656811  1.6801846
## 6      5  1.7385309  2.662559  1.7202176
## 7      2  0.4537062  1.886498  0.4233554
## 8      0  0.2478392  1.767018  0.1894171
```



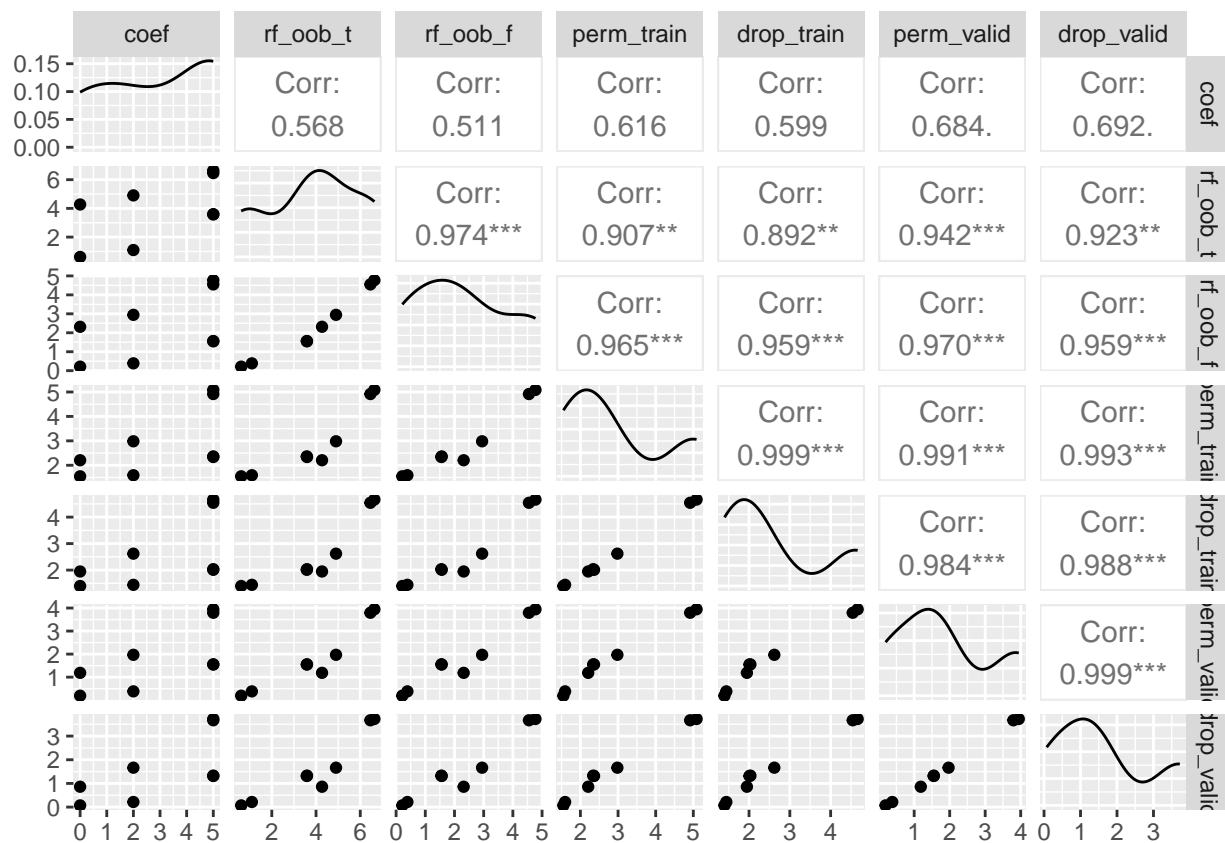
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.4728007	4.5413164	4.813505	4.342128	3.7513637	3.51074479
## 2	5	6.2773182	4.3408885	4.653758	4.257794	3.6467589	3.47921630
## 3	2	4.9979150	3.0574975	3.099135	2.685158	2.0406689	1.72894280
## 4	0	4.4538769	2.4766453	2.392174	2.066182	1.3095076	0.95015787
## 5	5	3.6488048	1.6117388	2.475786	2.147759	1.6619852	1.39009059
## 6	5	3.6852973	1.6574127	2.500824	2.117023	1.6478616	1.37476127
## 7	2	1.2721913	0.4593237	1.700158	1.515736	0.4536673	0.18684652
## 8	0	0.6284068	0.2231164	1.621445	1.462087	0.2480495	0.09124214



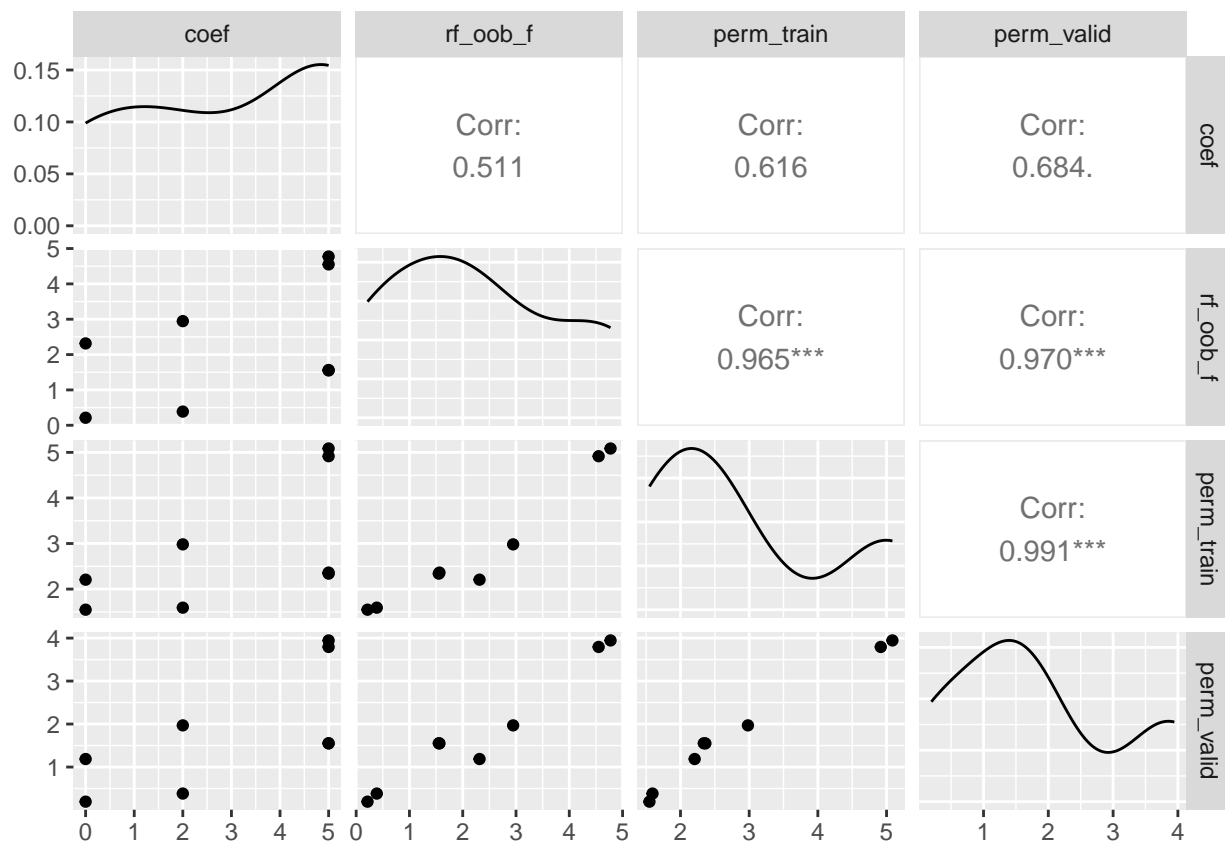
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  4.5413164  4.813505  3.7513637
## 2      5  4.3408885  4.653758  3.6467589
## 3      2  3.0574975  3.099135  2.0406689
## 4      0  2.4766453  2.392174  1.3095076
## 5      5  1.6117388  2.475786  1.6619852
## 6      5  1.6574127  2.500824  1.6478616
## 7      2  0.4593237  1.700158  0.4536673
## 8      0  0.2231164  1.621445  0.2480495
```



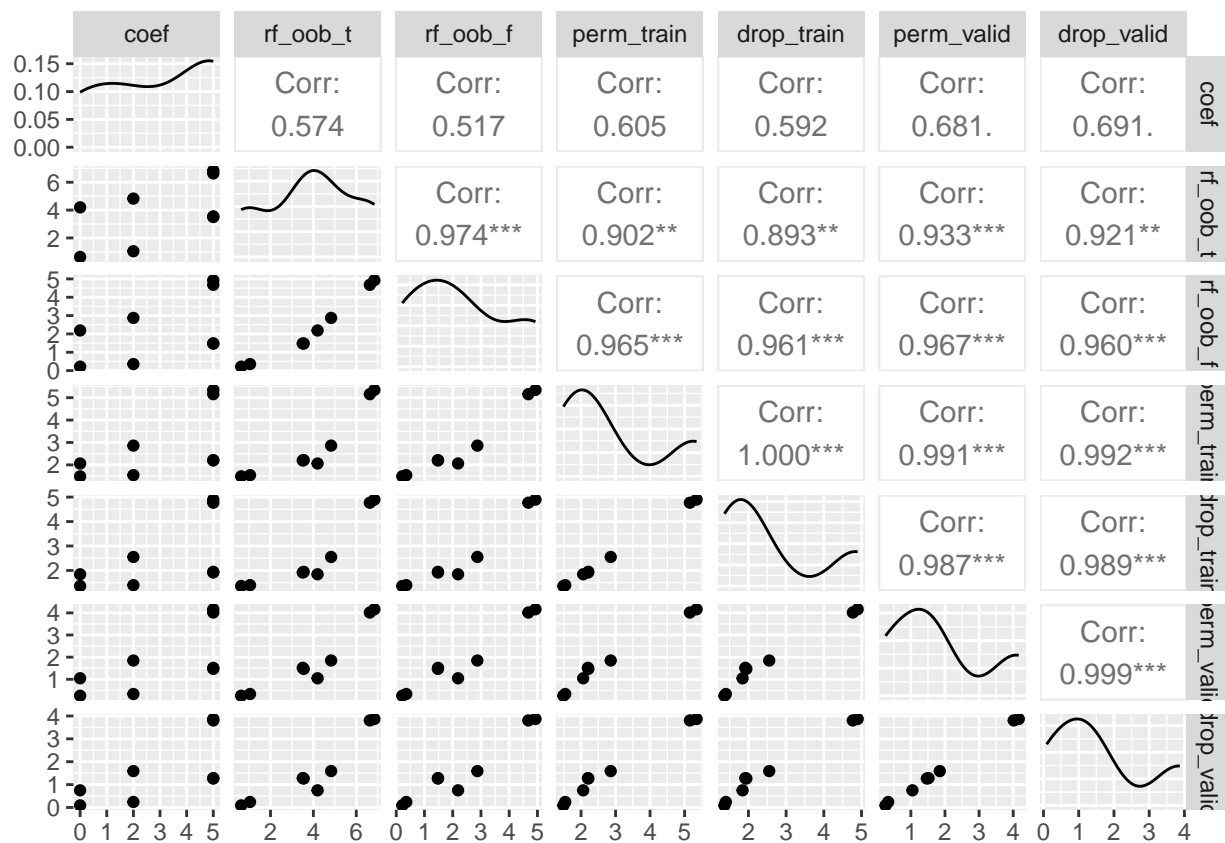
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.6380598	4.7740658	5.085831	4.662380	3.9455522	3.7223677
## 2	5	6.4587117	4.5508576	4.916161	4.541794	3.7938993	3.6674041
## 3	2	4.9040566	2.9459914	2.982040	2.619863	1.9695329	1.6643803
## 4	0	4.2716857	2.3154096	2.206229	1.948080	1.1865003	0.8601539
## 5	5	3.5930310	1.5509020	2.339543	2.033790	1.5528836	1.3279352
## 6	5	3.5816995	1.5629941	2.362347	2.010944	1.5499550	1.3150912
## 7	2	1.0993909	0.3867996	1.592634	1.443503	0.3827159	0.2156443
## 8	0	0.6177717	0.2128736	1.547906	1.398567	0.1944262	0.0718928



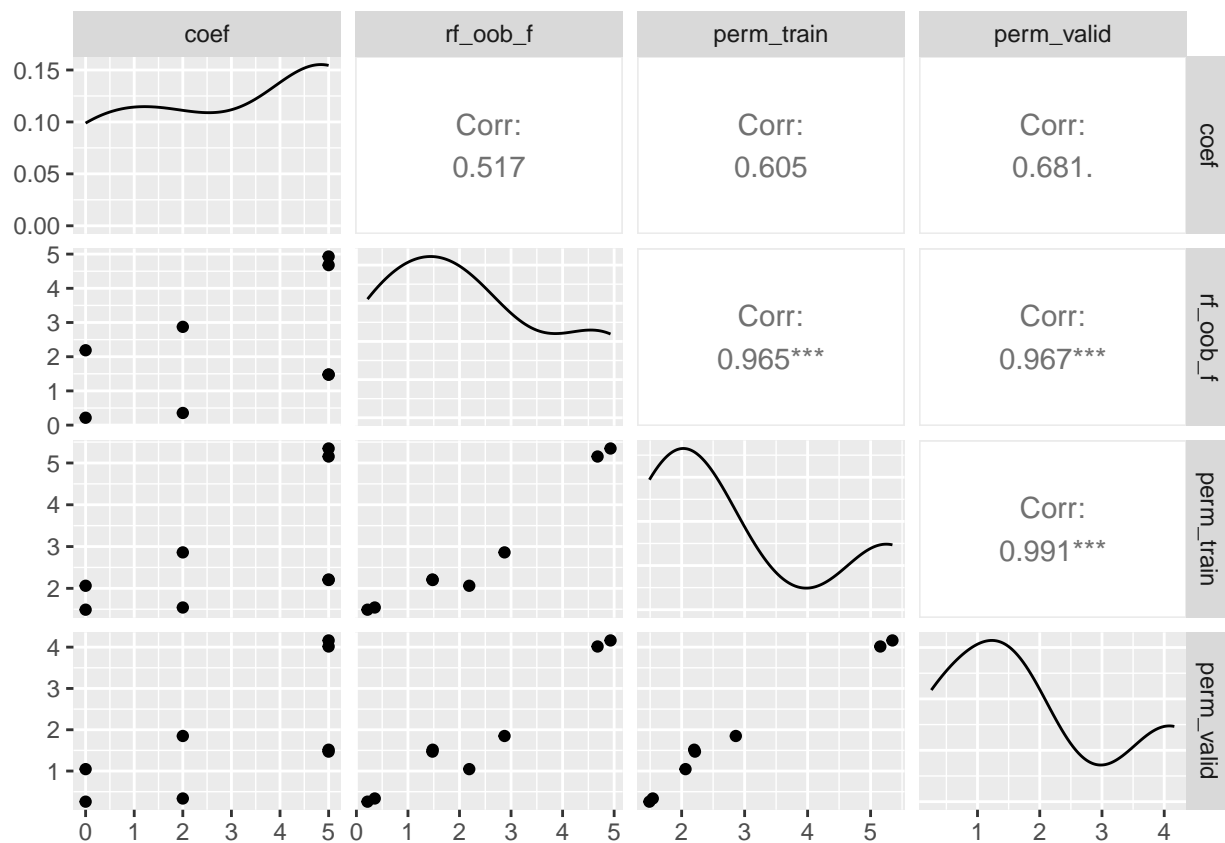
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5 4.7740658   5.085831  3.9455522
## 2      5 4.5508576   4.916161  3.7938993
## 3      2 2.9459914   2.982040  1.9695329
## 4      0 2.3154096   2.206229  1.1865003
## 5      5 1.5509020   2.339543  1.5528836
## 6      5 1.5629941   2.362347  1.5499550
## 7      2 0.3867996   1.592634  0.3827159
## 8      0 0.2128736   1.547906  0.1944262
```



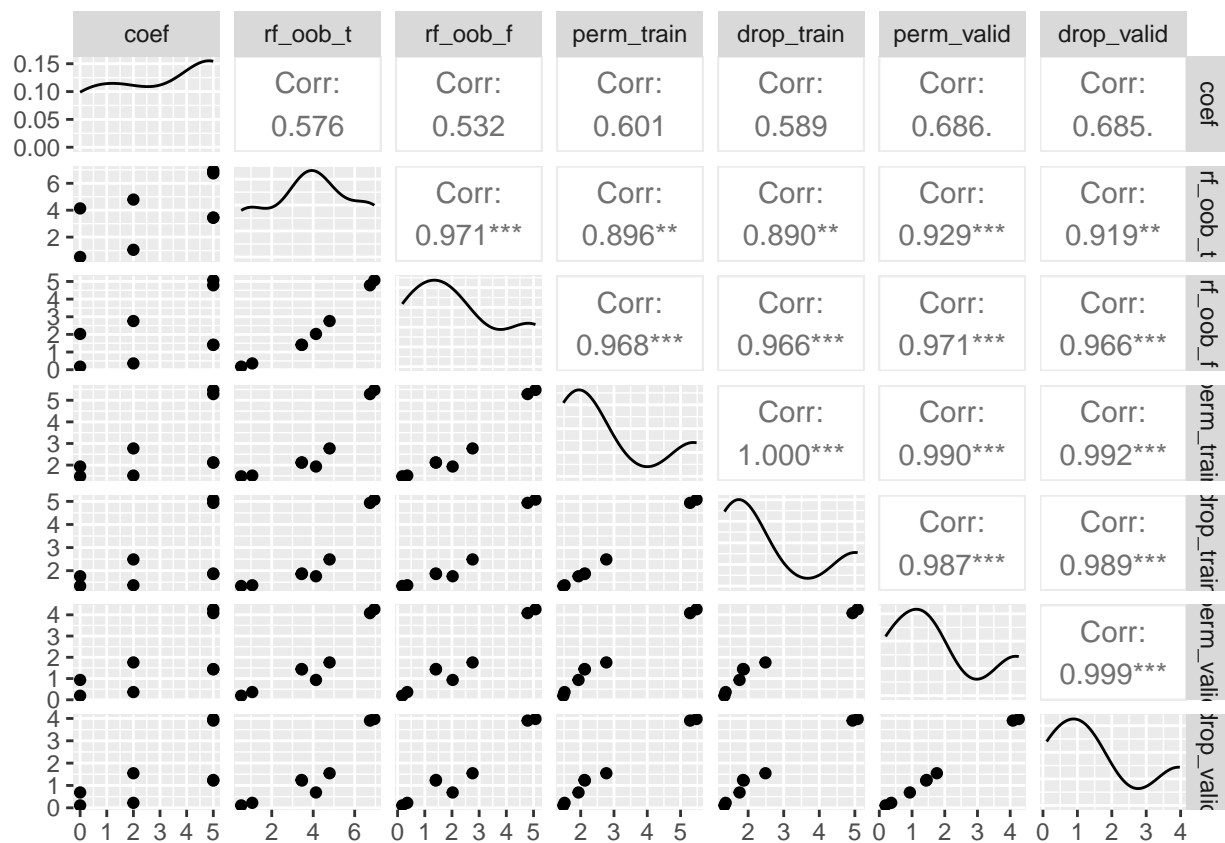
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.8440243	4.9295594	5.346901	4.898064	4.1625463	3.86975290
## 2	5	6.6339935	4.6782192	5.154607	4.770977	4.0157074	3.80595281
## 3	2	4.8267456	2.8733400	2.860120	2.551678	1.8487772	1.59099322
## 4	0	4.1959959	2.1887038	2.061728	1.844763	1.0465461	0.74992529
## 5	5	3.5495710	1.4747634	2.211585	1.944378	1.4704628	1.26638372
## 6	5	3.5056908	1.4795872	2.199144	1.917367	1.5167342	1.28420599
## 7	2	1.0459518	0.3553673	1.541535	1.399810	0.3351153	0.24019121
## 8	0	0.6354223	0.2154561	1.488717	1.364233	0.2586388	0.09089509



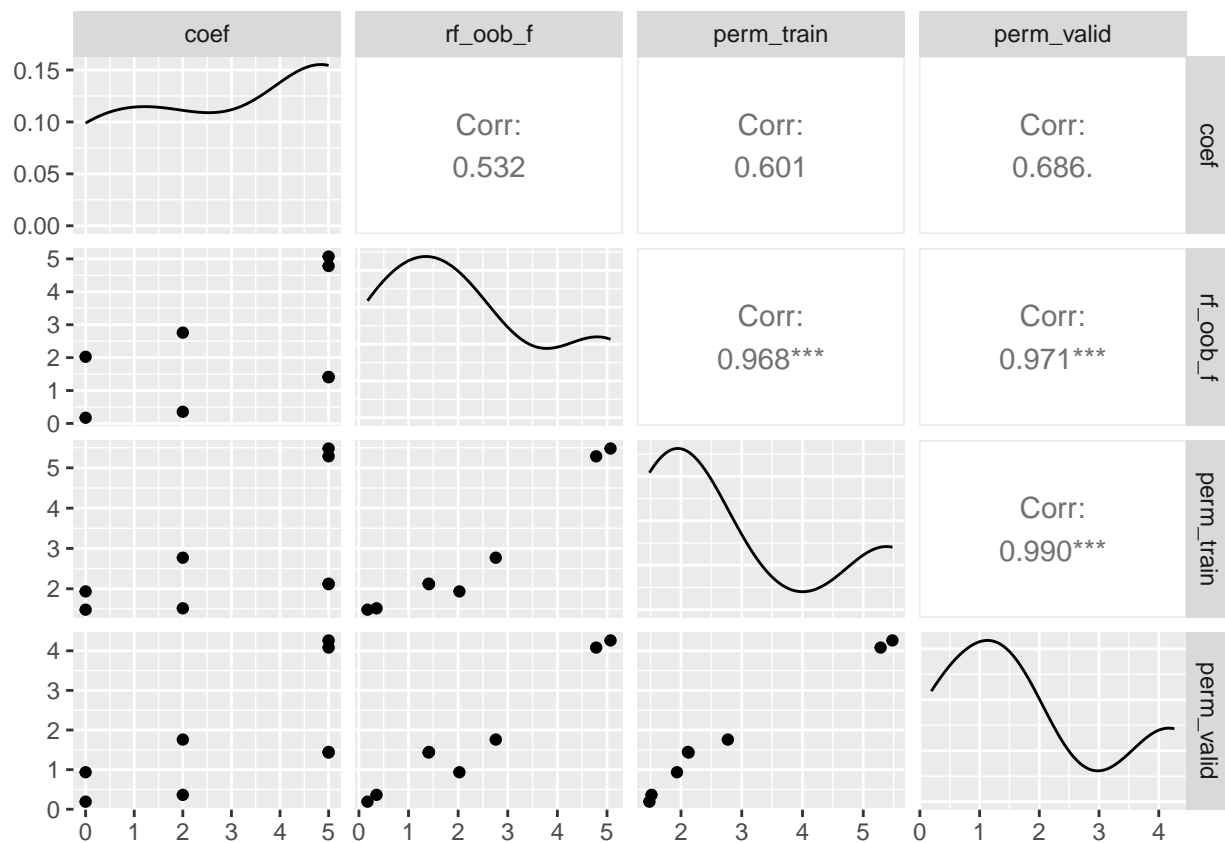
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  4.9295594   5.346901  4.1625463
## 2      5  4.6782192   5.154607  4.0157074
## 3      2  2.8733400   2.860120  1.8487772
## 4      0  2.1887038   2.061728  1.0465461
## 5      5  1.4747634   2.211585  1.4704628
## 6      5  1.4795872   2.199144  1.5167342
## 7      2  0.3553673   1.541535  0.3351153
## 8      0  0.2154561   1.488717  0.2586388
```



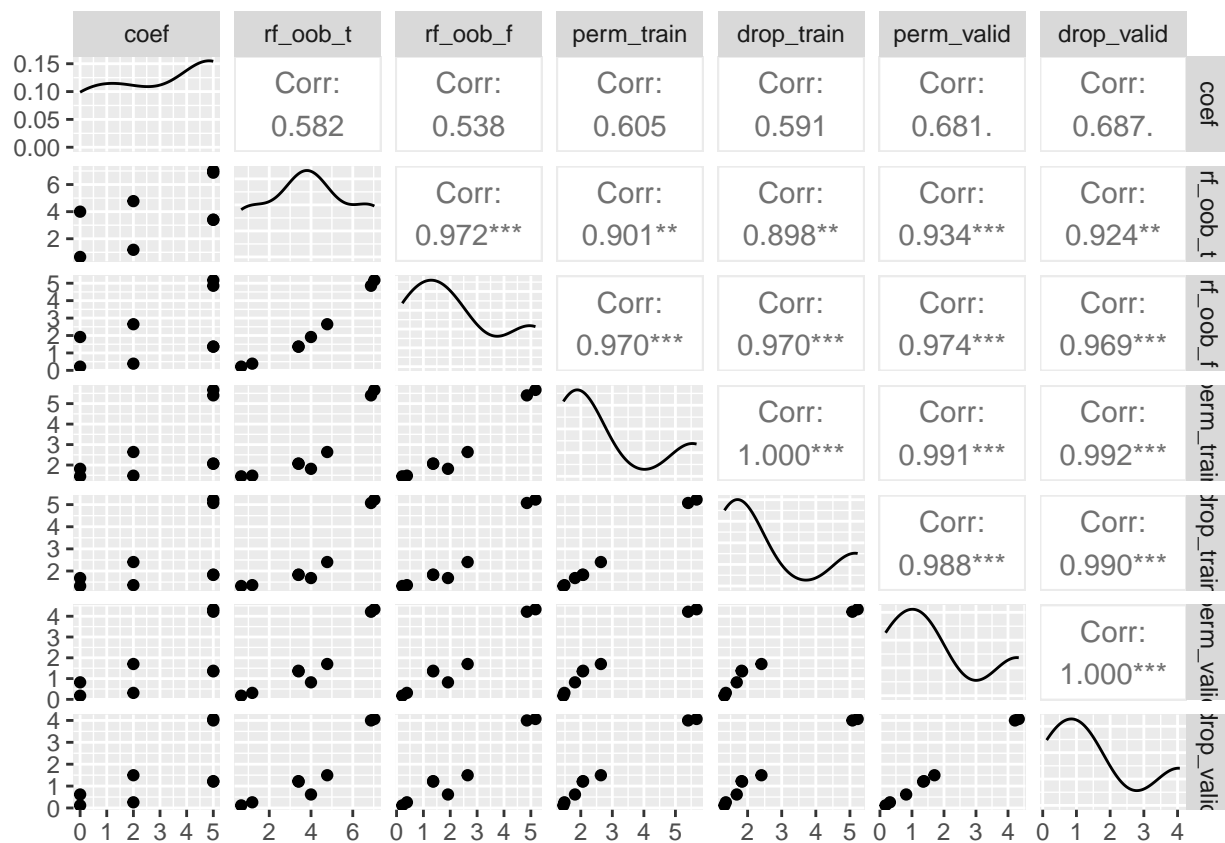
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	6.950919	5.0712617	5.480555	5.078058	4.2612138	3.9767618
## 2	5	6.739171	4.7822503	5.286416	4.935361	4.0806387	3.9049678
## 3	2	4.790585	2.7587235	2.770774	2.487057	1.7588730	1.5470886
## 4	0	4.133470	2.0256465	1.934032	1.758310	0.9341046	0.6871278
## 5	5	3.457785	1.4046777	2.125745	1.877026	1.4315530	1.2199783
## 6	5	3.435593	1.4130722	2.113455	1.856431	1.4464601	1.2448807
## 7	2	1.058166	0.3555464	1.515093	1.368380	0.3634862	0.2221296
## 8	0	0.526965	0.1736995	1.480252	1.335531	0.1918324	0.1076382



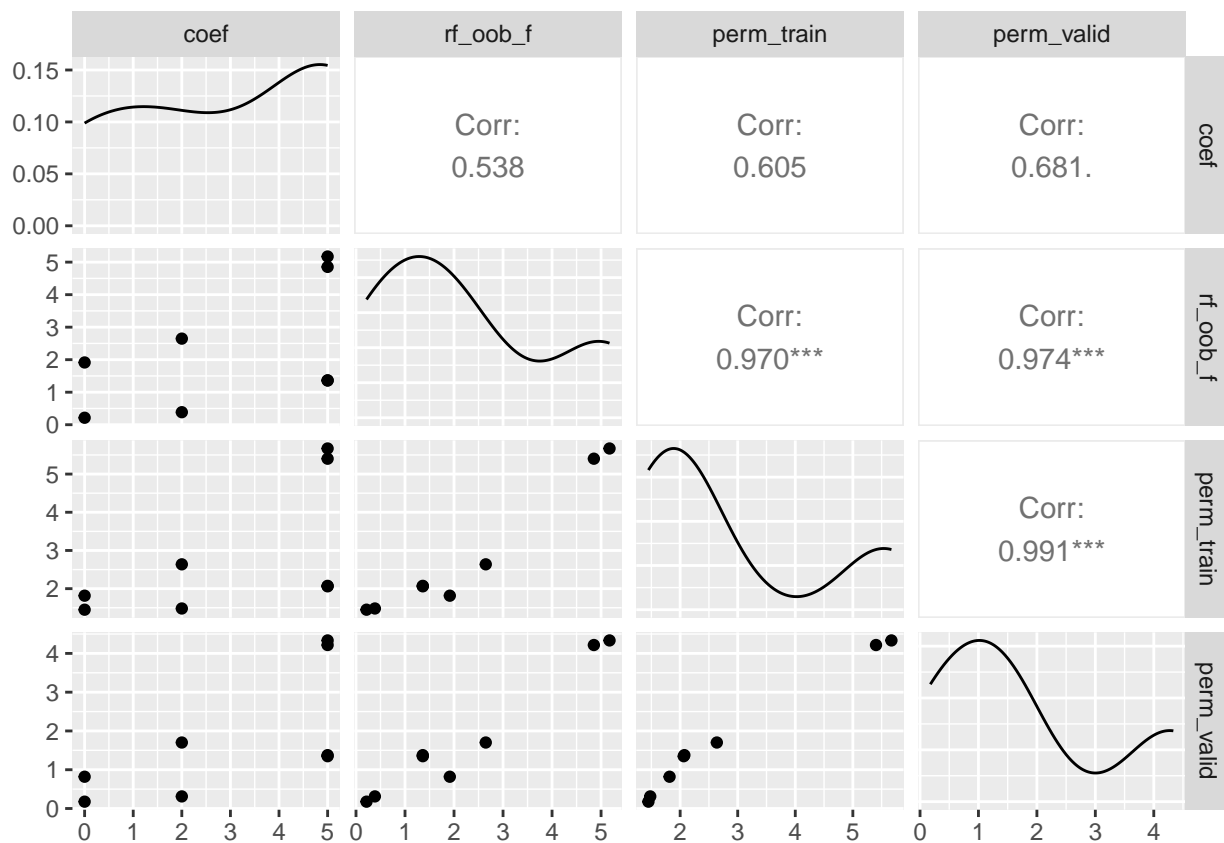
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  5.0712617   5.480555  4.2612138
## 2      5  4.7822503   5.286416  4.0806387
## 3      2  2.7587235   2.770774  1.7588730
## 4      0  2.0256465   1.934032  0.9341046
## 5      5  1.4046777   2.125745  1.4315530
## 6      5  1.4130722   2.113455  1.4464601
## 7      2  0.3555464   1.515093  0.3634862
## 8      0  0.1736995   1.480252  0.1918324
```



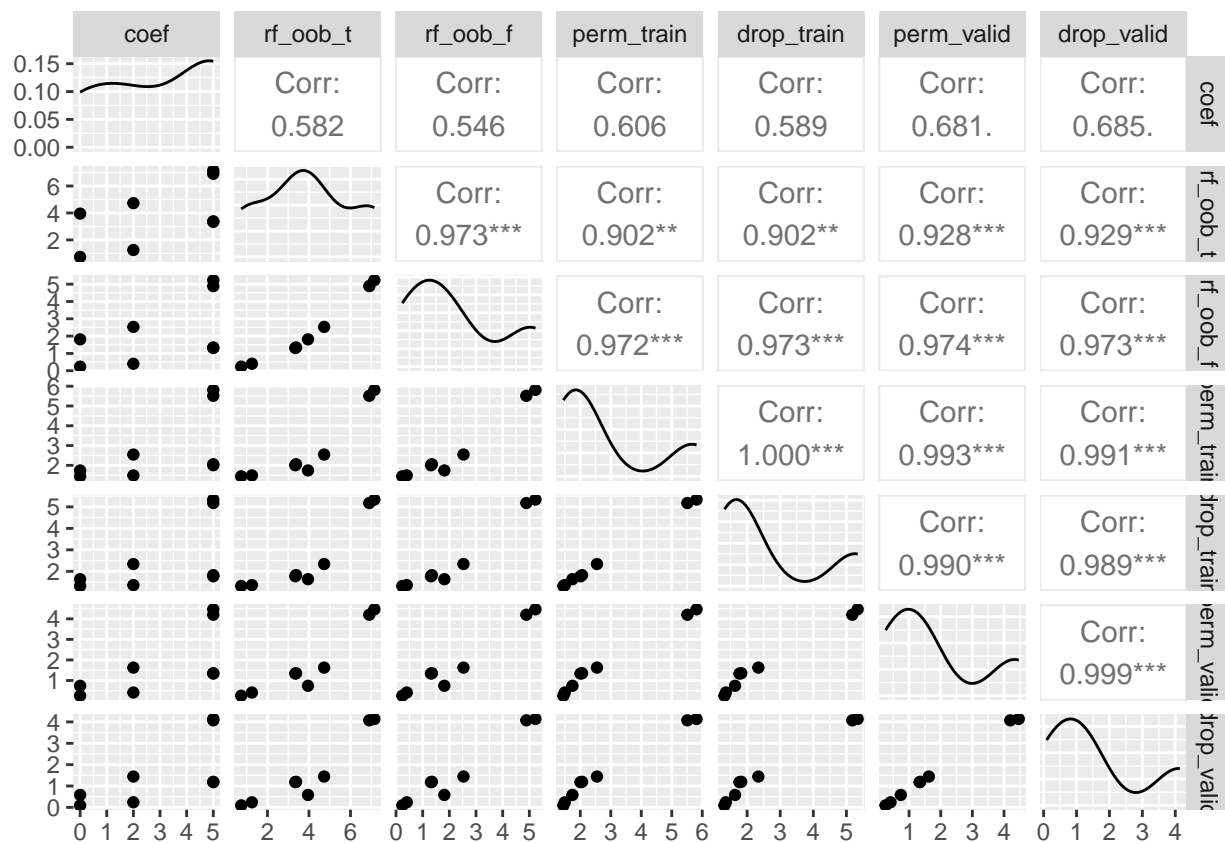
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	7.0346127	5.1726370	5.670379	5.226804	4.3348545	4.0674498
## 2	5	6.8739689	4.8536281	5.402523	5.072490	4.2166362	3.9997970
## 3	2	4.7760274	2.6478008	2.637392	2.405334	1.7021935	1.4957088
## 4	0	3.9988809	1.9144679	1.816278	1.680745	0.8198051	0.6156966
## 5	5	3.4070472	1.3629167	2.061897	1.838205	1.3489401	1.2037020
## 6	5	3.3948927	1.3622719	2.073262	1.823328	1.3770890	1.2267314
## 7	2	1.1852215	0.3855476	1.480847	1.361257	0.3123432	0.2586627
## 8	0	0.6566842	0.2146848	1.449130	1.321562	0.1766847	0.1171913



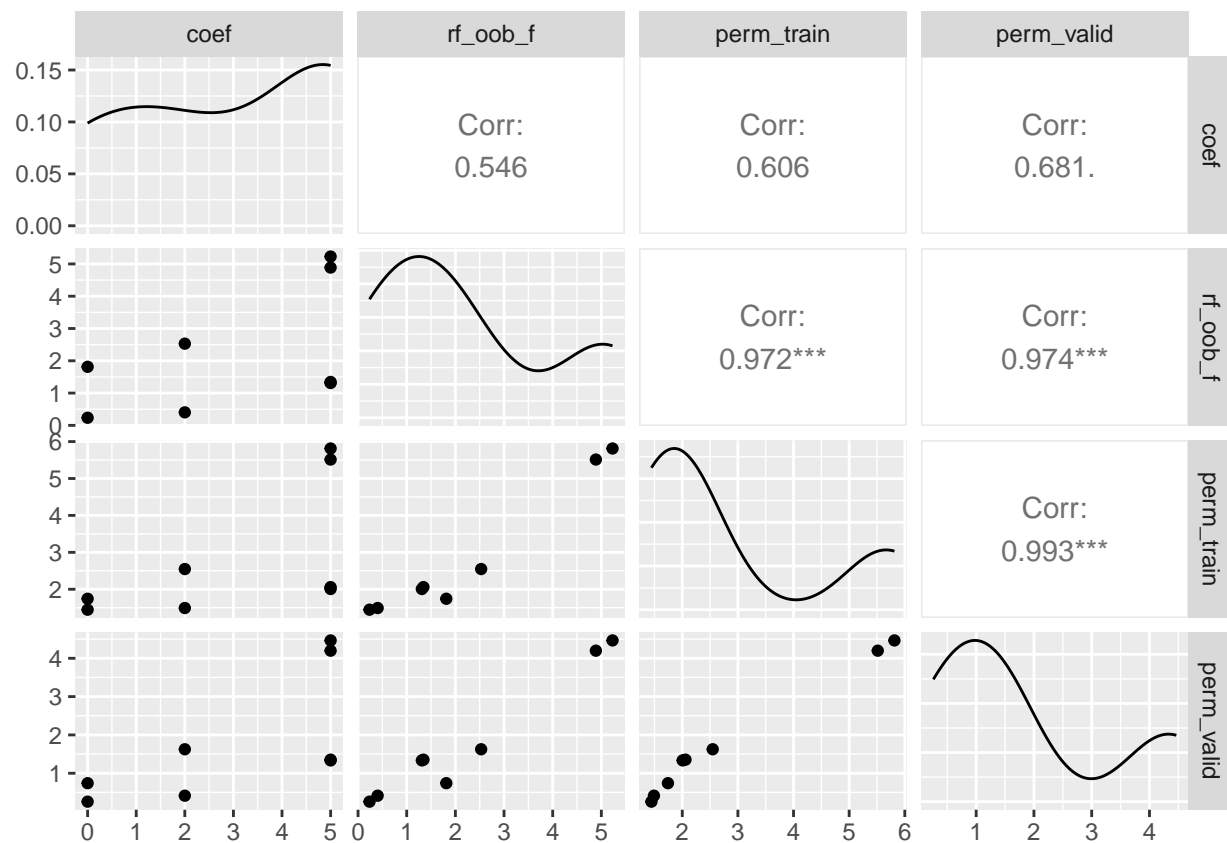
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  5.1726370  5.670379  4.3348545
## 2      5  4.8536281  5.402523  4.2166362
## 3      2  2.6478008  2.637392  1.7021935
## 4      0  1.9144679  1.816278  0.8198051
## 5      5  1.3629167  2.061897  1.3489401
## 6      5  1.3622719  2.073262  1.3770890
## 7      2  0.3855476  1.480847  0.3123432
## 8      0  0.2146848  1.449130  0.1766847
```



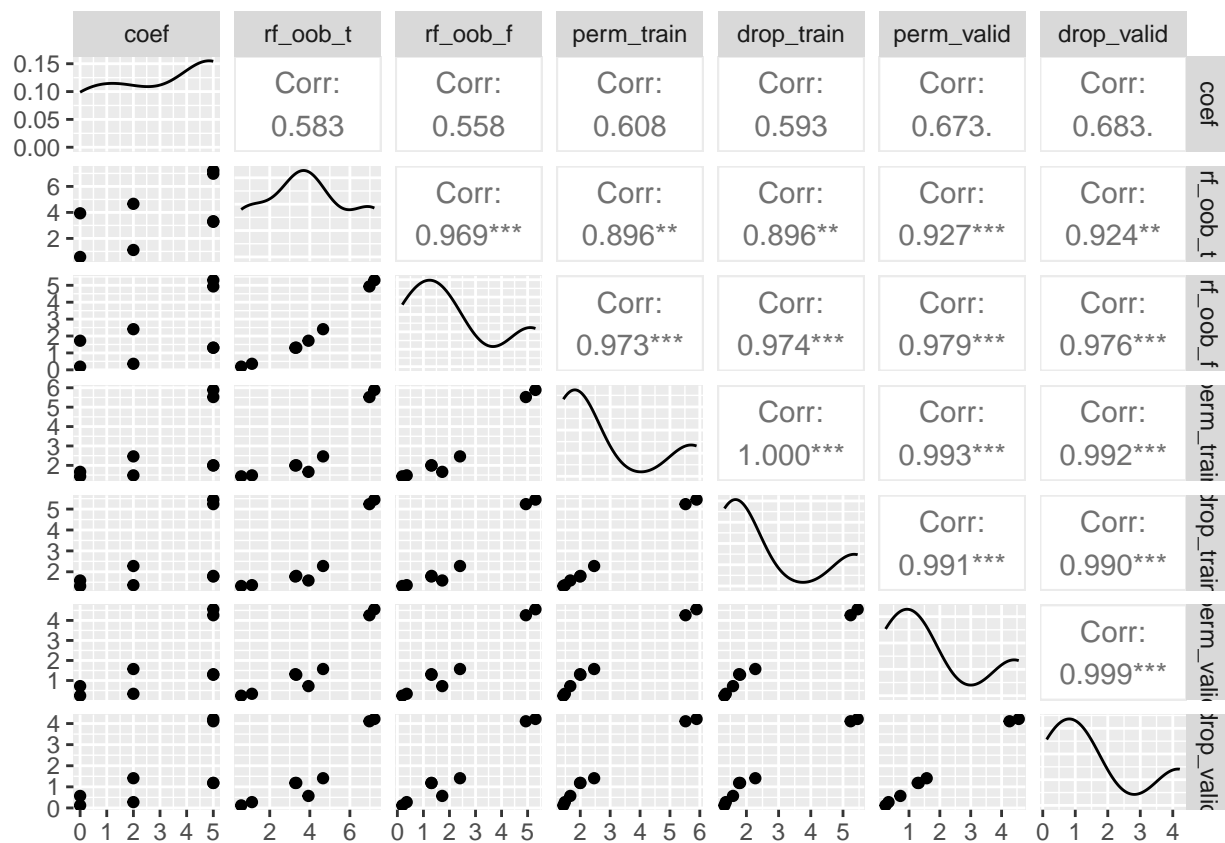
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	7.151048	5.2295701	5.813691	5.339819	4.4656121	4.13865176
## 2	5	6.906390	4.8878136	5.513883	5.178054	4.1982018	4.07417037
## 3	2	4.730771	2.5300479	2.546998	2.339617	1.6257358	1.44313632
## 4	0	3.952820	1.8134107	1.742294	1.634864	0.7404748	0.58258477
## 5	5	3.377943	1.3440095	2.057727	1.823713	1.3542001	1.19842386
## 6	5	3.340656	1.3116209	2.006598	1.769489	1.3352427	1.18125157
## 7	2	1.251532	0.4029976	1.490717	1.358221	0.4146162	0.23916946
## 8	0	0.736085	0.2358295	1.445613	1.316184	0.2585753	0.09303114



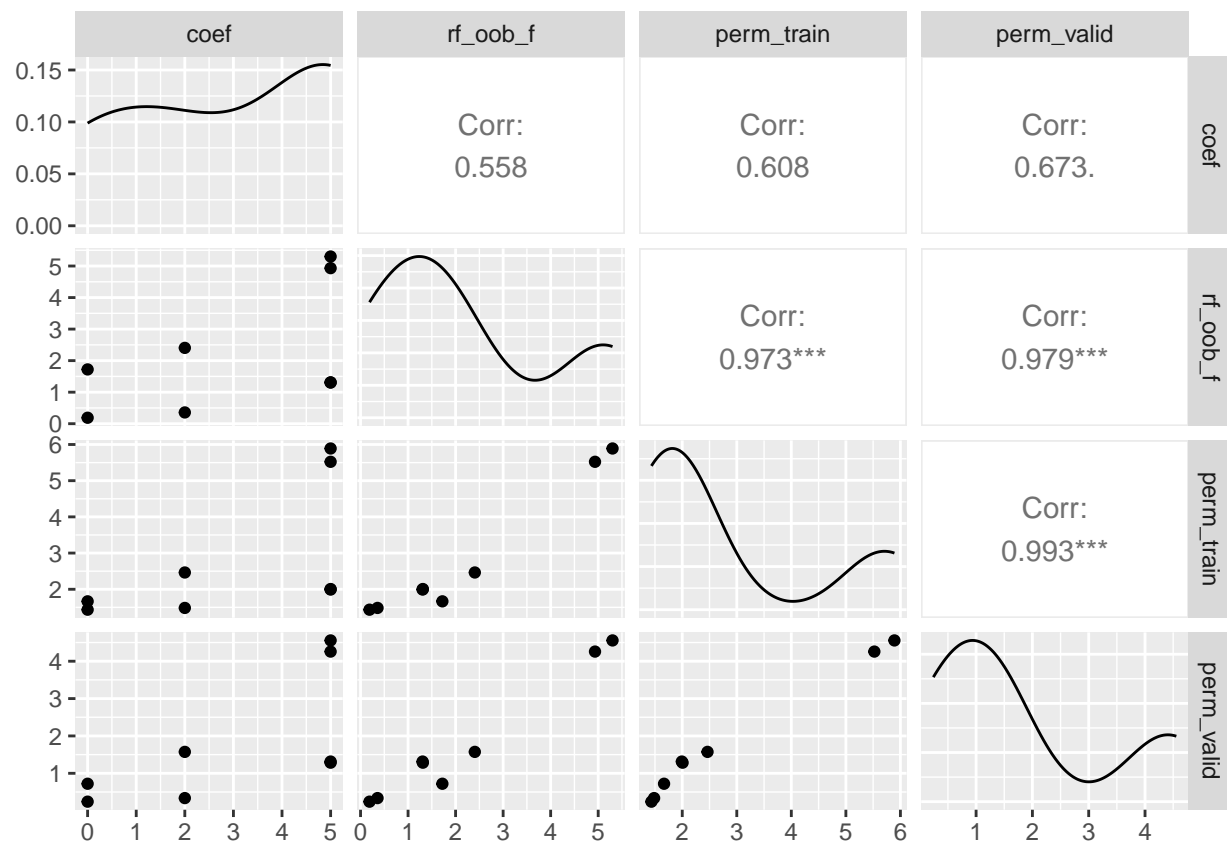
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  5.2295701   5.813691  4.4656121
## 2      5  4.8878136   5.513883  4.1982018
## 3      2  2.5300479   2.546998  1.6257358
## 4      0  1.8134107   1.742294  0.7404748
## 5      5  1.3440095   2.057727  1.3542001
## 6      5  1.3116209   2.006598  1.3352427
## 7      2  0.4029976   1.490717  0.4146162
## 8      0  0.2358295   1.445613  0.2585753
```



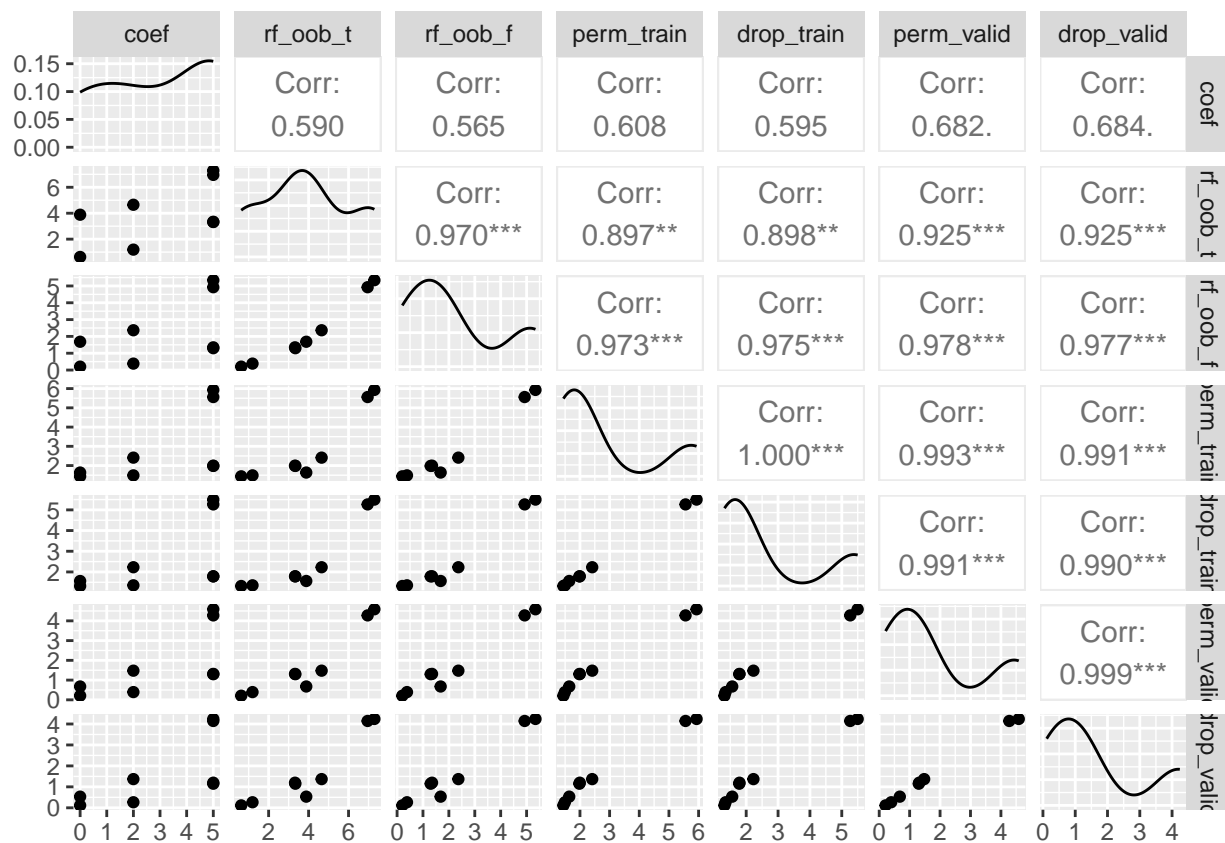
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	7.2227190	5.3016535	5.890436	5.455353	4.5562269	4.2171928
## 2	5	6.9756634	4.9334477	5.522062	5.237373	4.2581144	4.1088617
## 3	2	4.6617431	2.4045546	2.463593	2.277188	1.5733275	1.4081680
## 4	0	3.9334358	1.7217517	1.666298	1.583890	0.7205499	0.5686878
## 5	5	3.3236180	1.3113490	2.006207	1.803075	1.2840498	1.1845666
## 6	5	3.2776349	1.3054231	1.992384	1.773866	1.3145956	1.1834193
## 7	2	1.1106770	0.3570527	1.484123	1.360487	0.3365681	0.2756856
## 8	0	0.5756868	0.1890665	1.435231	1.319287	0.2409902	0.1203596



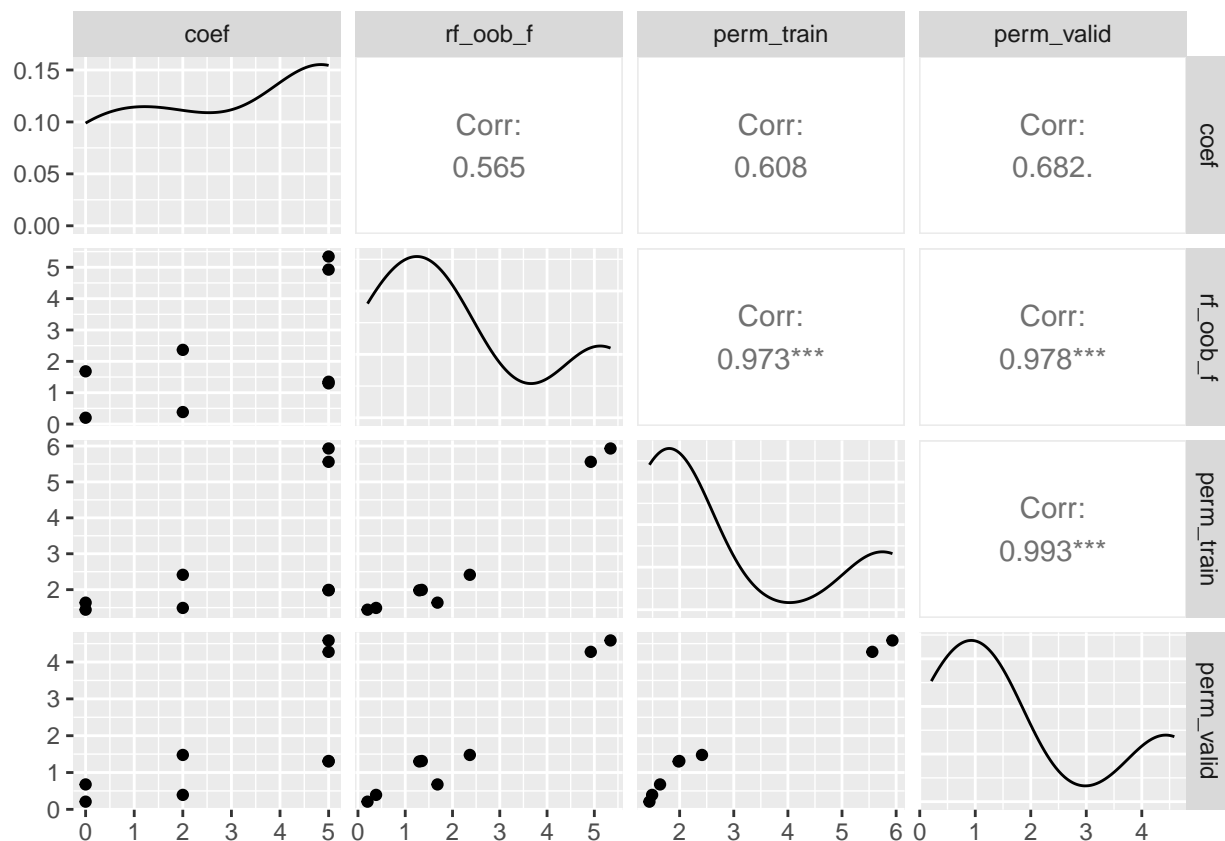
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  5.3016535   5.890436  4.5562269
## 2      5  4.9334477   5.522062  4.2581144
## 3      2  2.4045546   2.463593  1.5733275
## 4      0  1.7217517   1.666298  0.7205499
## 5      5  1.3113490   2.006207  1.2840498
## 6      5  1.3054231   1.992384  1.3145956
## 7      2  0.3570527   1.484123  0.3365681
## 8      0  0.1890665   1.435231  0.2409902
```



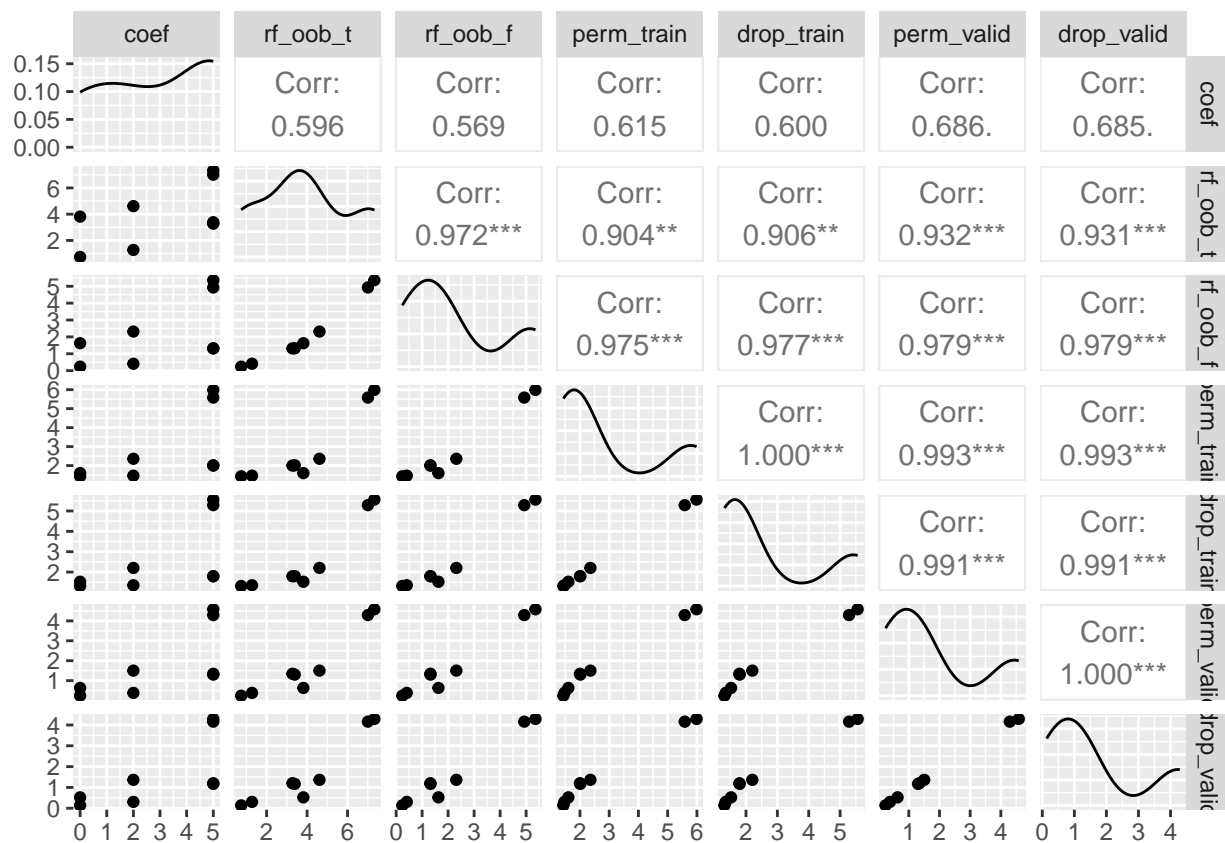
##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	7.2933589	5.3427789	5.931413	5.498007	4.5863925	4.2405328
## 2	5	6.9554370	4.9253883	5.559668	5.264931	4.2763126	4.1453611
## 3	2	4.6522106	2.3667451	2.413100	2.227210	1.4771412	1.3706961
## 4	0	3.8841850	1.6824494	1.637939	1.561600	0.6765382	0.5324861
## 5	5	3.3370855	1.2984157	1.981471	1.783216	1.3004438	1.1502654
## 6	5	3.3201387	1.3498802	1.994042	1.789568	1.3127114	1.1983372
## 7	2	1.1921051	0.3830356	1.491450	1.352978	0.3928741	0.2648641
## 8	0	0.6272377	0.2022429	1.441787	1.319019	0.2089783	0.1151039



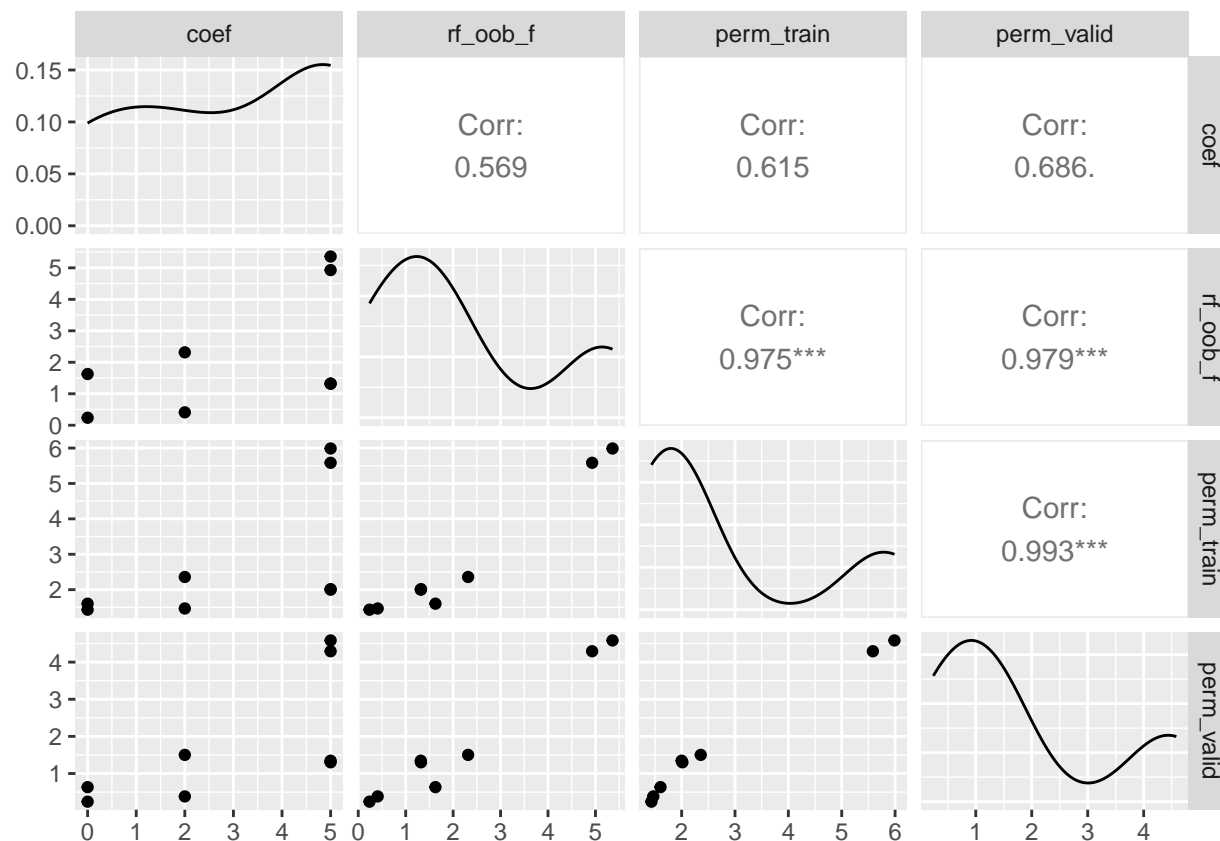
```
##      coef  rf_oob_f perm_train perm_valid
## 1      5  5.3427789   5.931413  4.5863925
## 2      5  4.9253883   5.559668  4.2763126
## 3      2  2.3667451   2.413100  1.4771412
## 4      0  1.6824494   1.637939  0.6765382
## 5      5  1.2984157   1.981471  1.3004438
## 6      5  1.3498802   1.994042  1.3127114
## 7      2  0.3830356   1.491450  0.3928741
## 8      0  0.2022429   1.441787  0.2089783
```



##	coef	rf_oob_t	rf_oob_f	perm_train	drop_train	perm_valid	drop_valid
## 1	5	7.3274727	5.3599532	5.987952	5.560224	4.5838607	4.2904863
## 2	5	7.0079207	4.9300345	5.583890	5.286510	4.2927042	4.1605779
## 3	2	4.6130651	2.3157313	2.358177	2.202229	1.5021702	1.3658789
## 4	0	3.8168274	1.6267072	1.603779	1.524820	0.6335992	0.5234447
## 5	5	3.3887226	1.3224451	2.014368	1.799189	1.3019839	1.1696339
## 6	5	3.2767930	1.3181803	2.000359	1.790938	1.3450224	1.2013226
## 7	2	1.2770504	0.4092293	1.468904	1.352660	0.3835341	0.3063159
## 8	0	0.7429277	0.2374625	1.434010	1.314291	0.2424372	0.1363539



```
## coef rf_oob_f perm_train perm_valid
## 1 5 5.3599532 5.987952 4.5838607
## 2 5 4.9300345 5.583890 4.2927042
## 3 2 2.3157313 2.358177 1.5021702
## 4 0 1.6267072 1.603779 0.6335992
## 5 5 1.3224451 2.014368 1.3019839
## 6 5 1.3181803 2.000359 1.3450224
## 7 2 0.4092293 1.468904 0.3835341
## 8 0 0.2374625 1.434010 0.2424372
```



```
rsq
```

```
## [1] 0.3059584 0.4002955 0.4463916 0.4702054 0.4807750 0.4863761 0.4892879
## [8] 0.4891247 0.4898518 0.4889229 0.4882676 0.4873865
```

```
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)

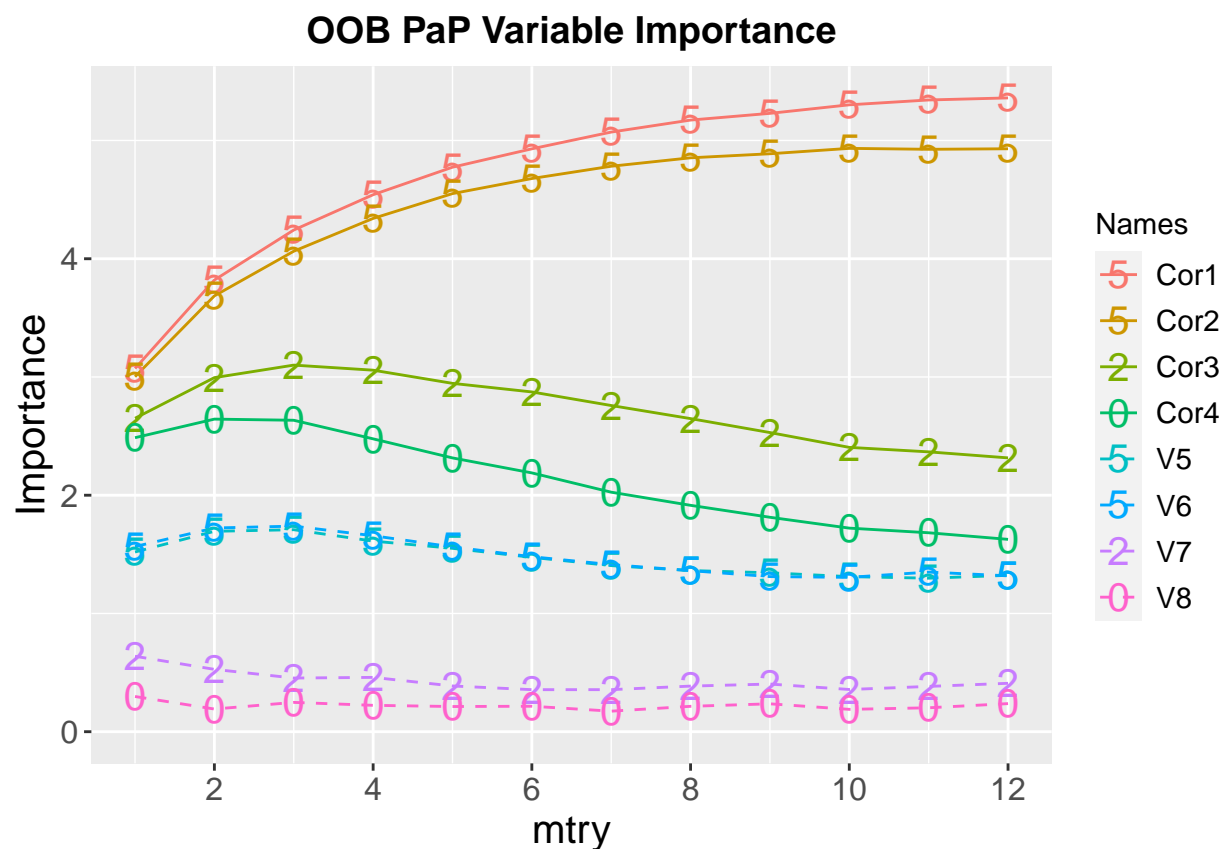
colnames(perm_train1)[1:12] <- 1:12
perm_train1 <- perm_train1 %>%
  pivot_longer(!c(Names,mag), names_to = "mtry", values_to = "Imp")
perm_train1$mtry <- as.numeric(perm_train1$mtry)

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)

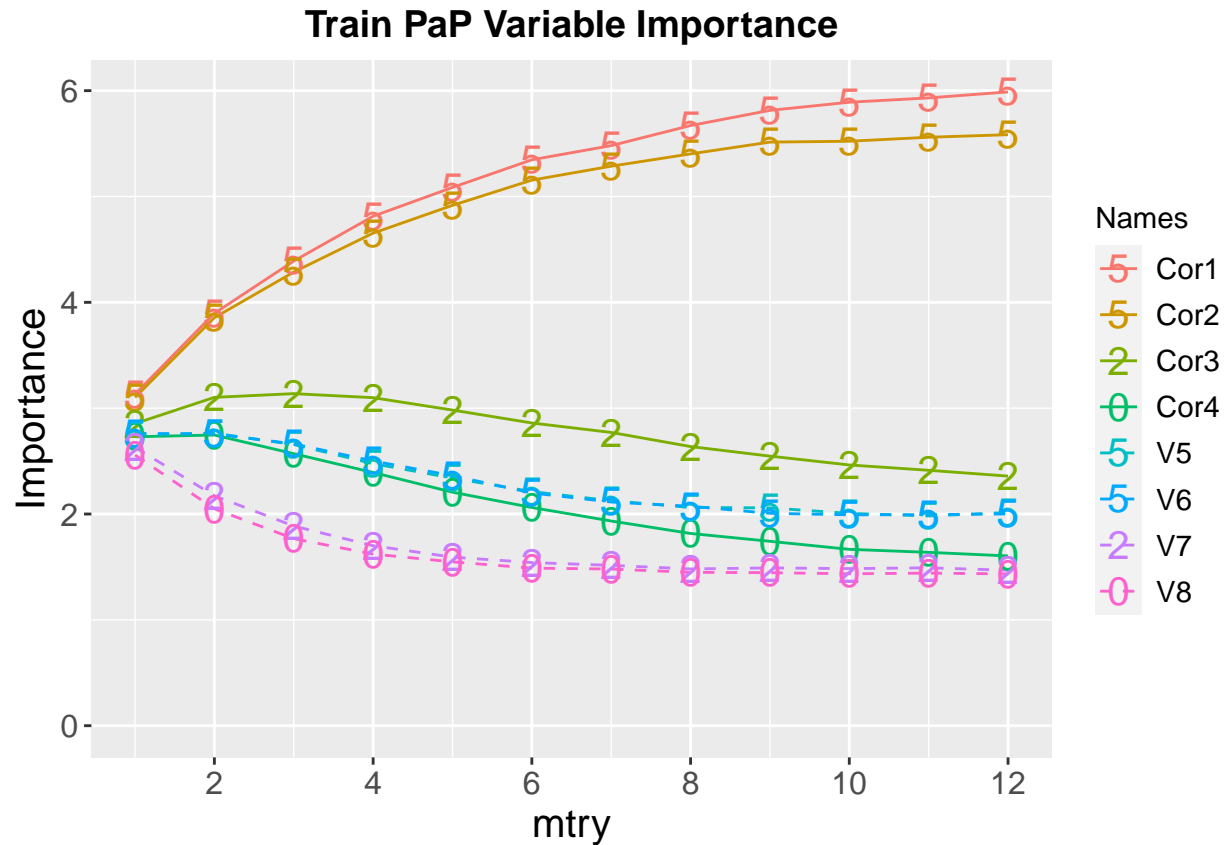
colnames(perm_valid1)[1:12] <- 1:12
perm_valid1 <- perm_valid1 %>%
  pivot_longer(!c(Names,mag), names_to = "mtry", values_to = "Imp")
perm_valid1$mtry <- as.numeric(perm_valid1$mtry)

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

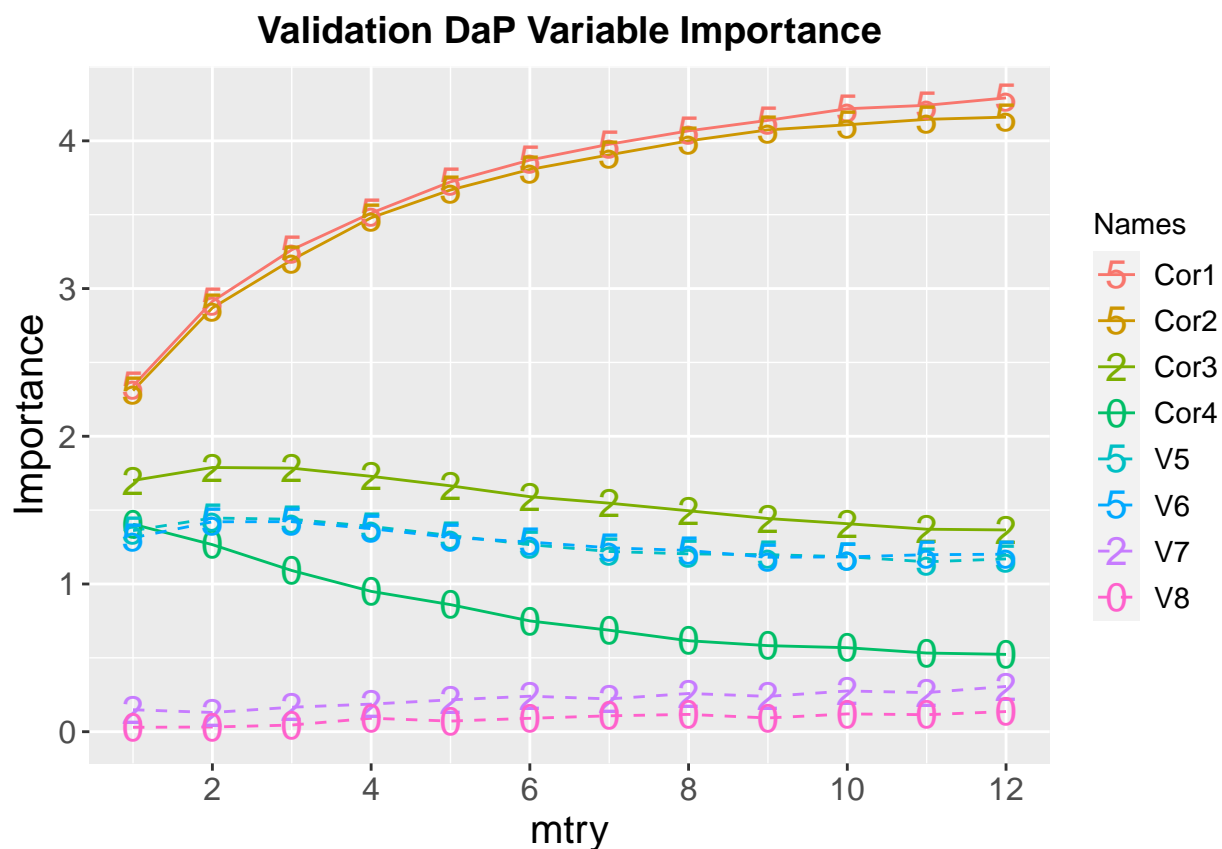
```



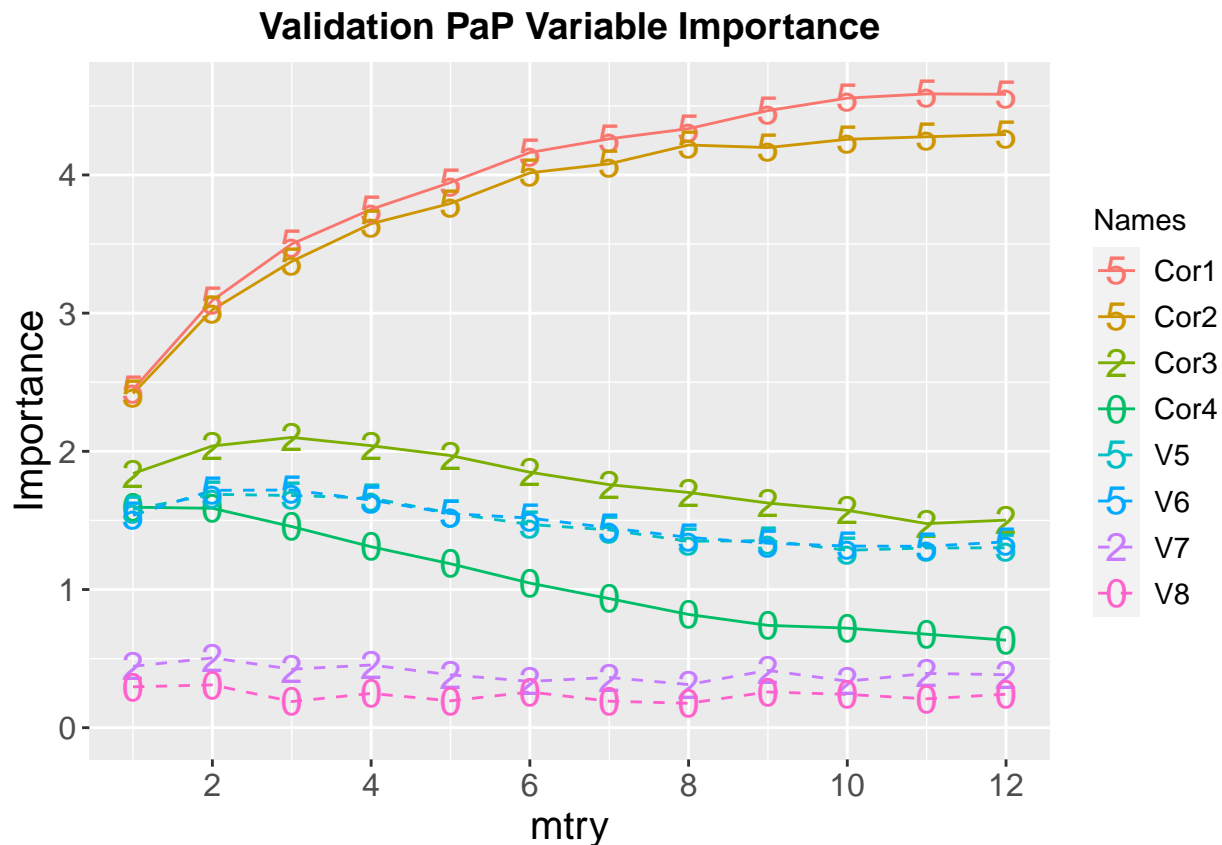
```
gp <- perm_train1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,max(perm_train1$Imp))) +
  ggtitle("Train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

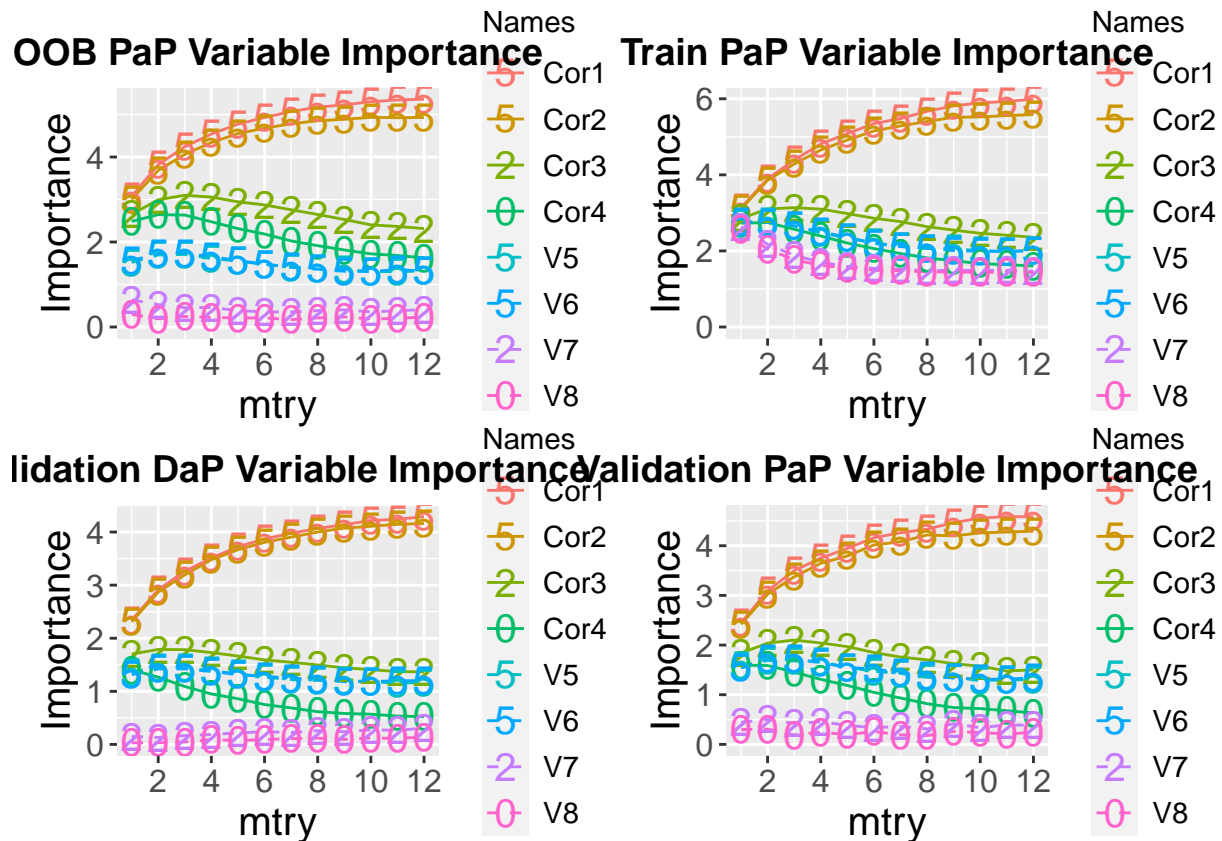


```
gv <- perm_valid1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,max(perm_valid1$Imp))) +
  ggtitle("Validation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

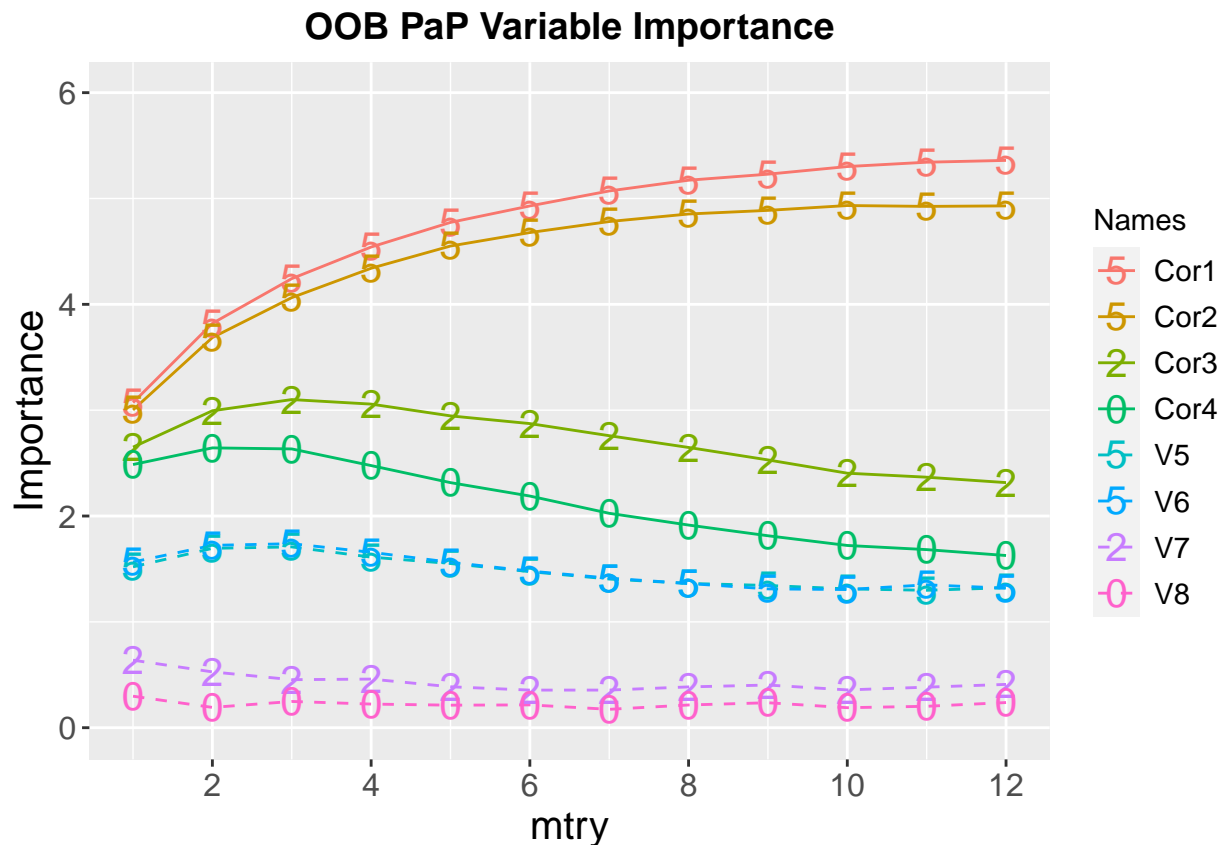


```
ggsave("xcos_all_zoom.pdf", plot = gr + gp + gd + gv, dpi = 2400,
       width = 9, height = 9)
```

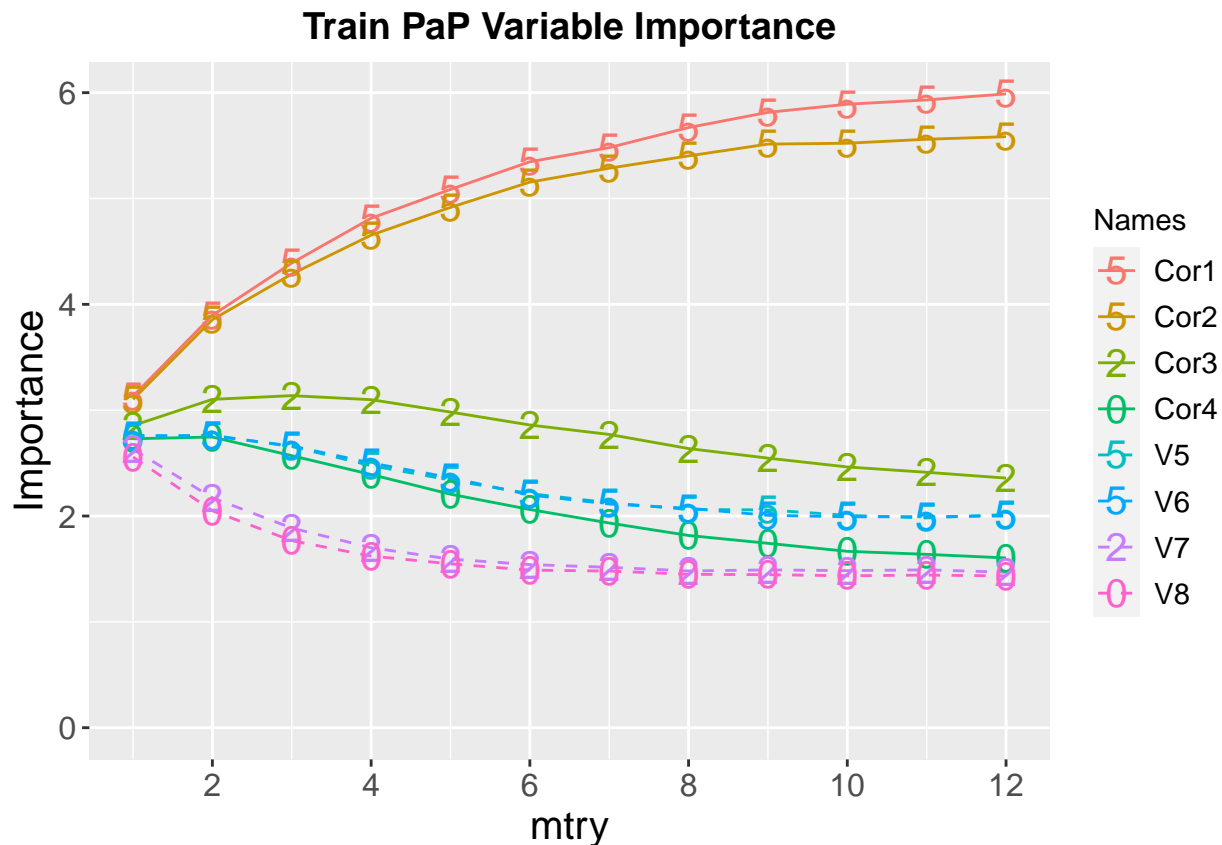
```
ma = max(rf_oob_f1$Imp, perm_train1$Imp, perm_valid1$Imp, drop_valid1$Imp)
# mp = max(rf_pdp1$Imp)
```

```
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)
```

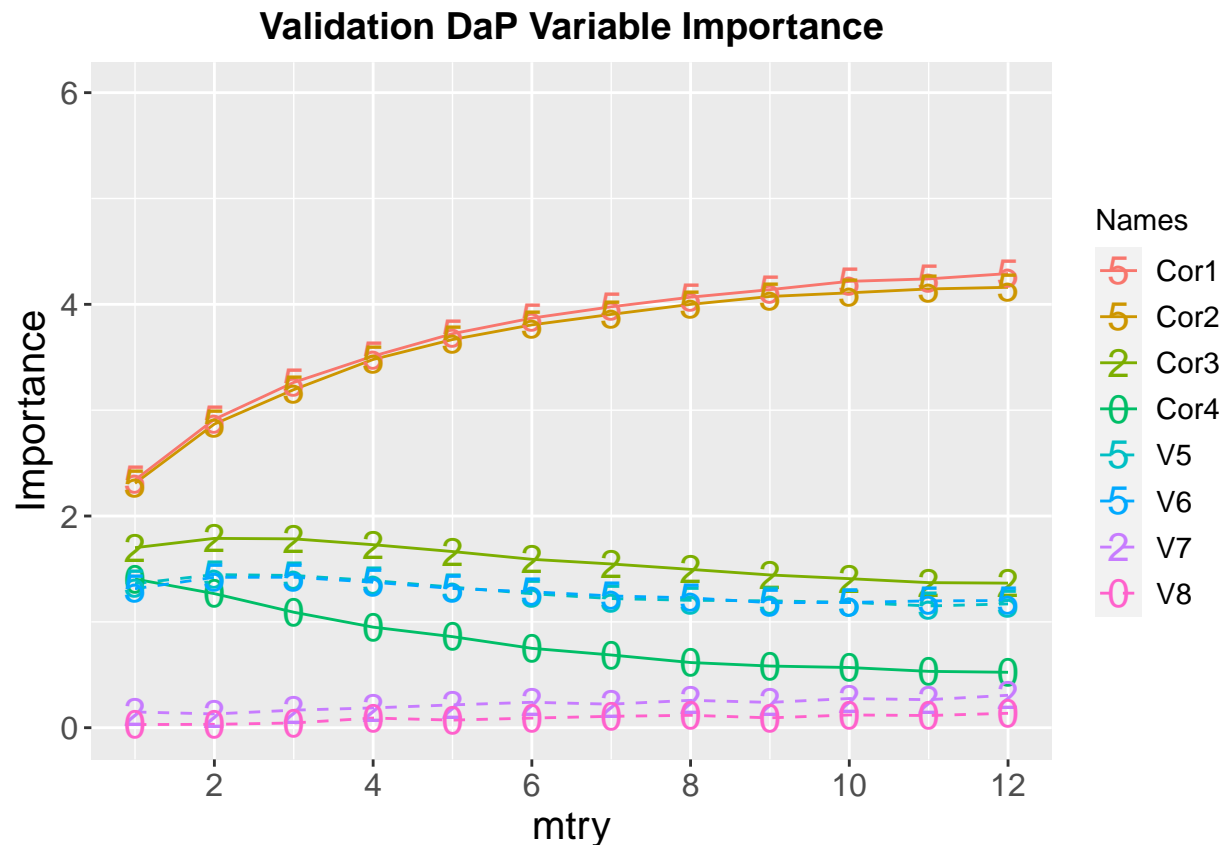
```
gr
```



```
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
```

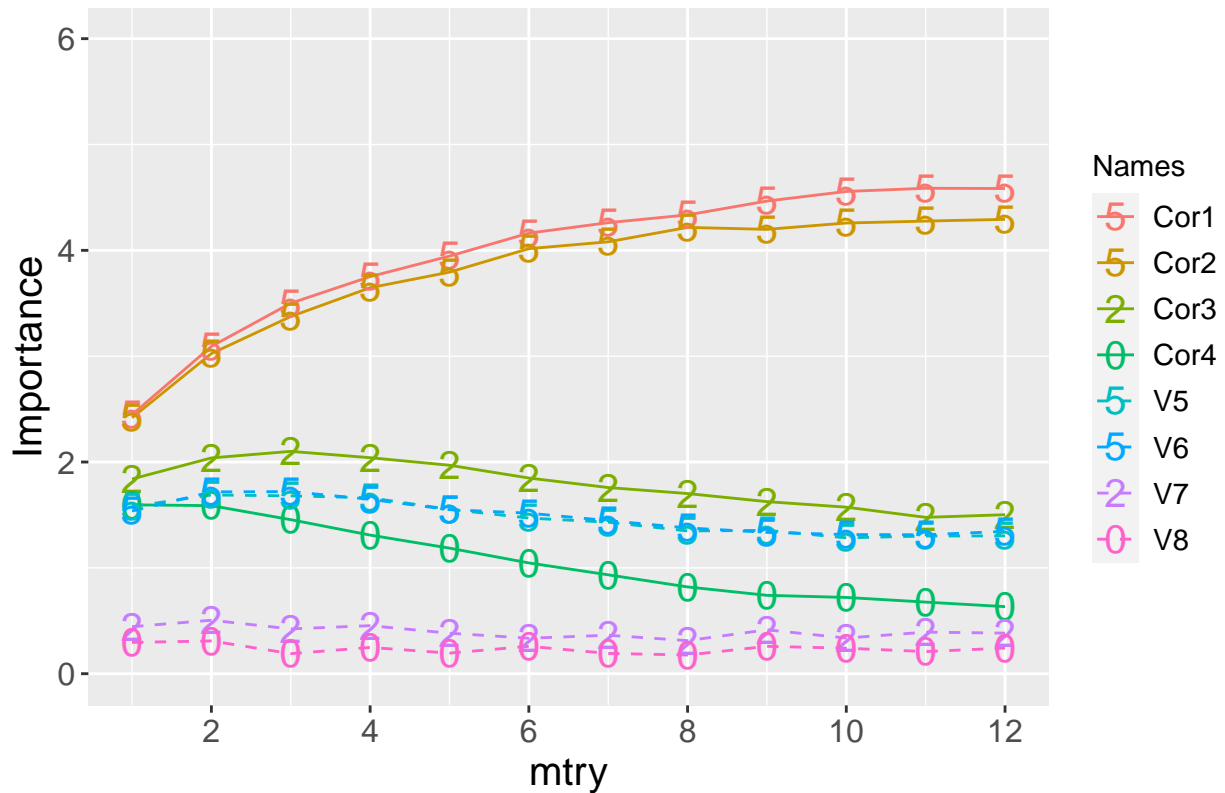


```
gd <- drop_valid1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,ma)) +
  ggtitle("Validation DaP Variable 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)
```



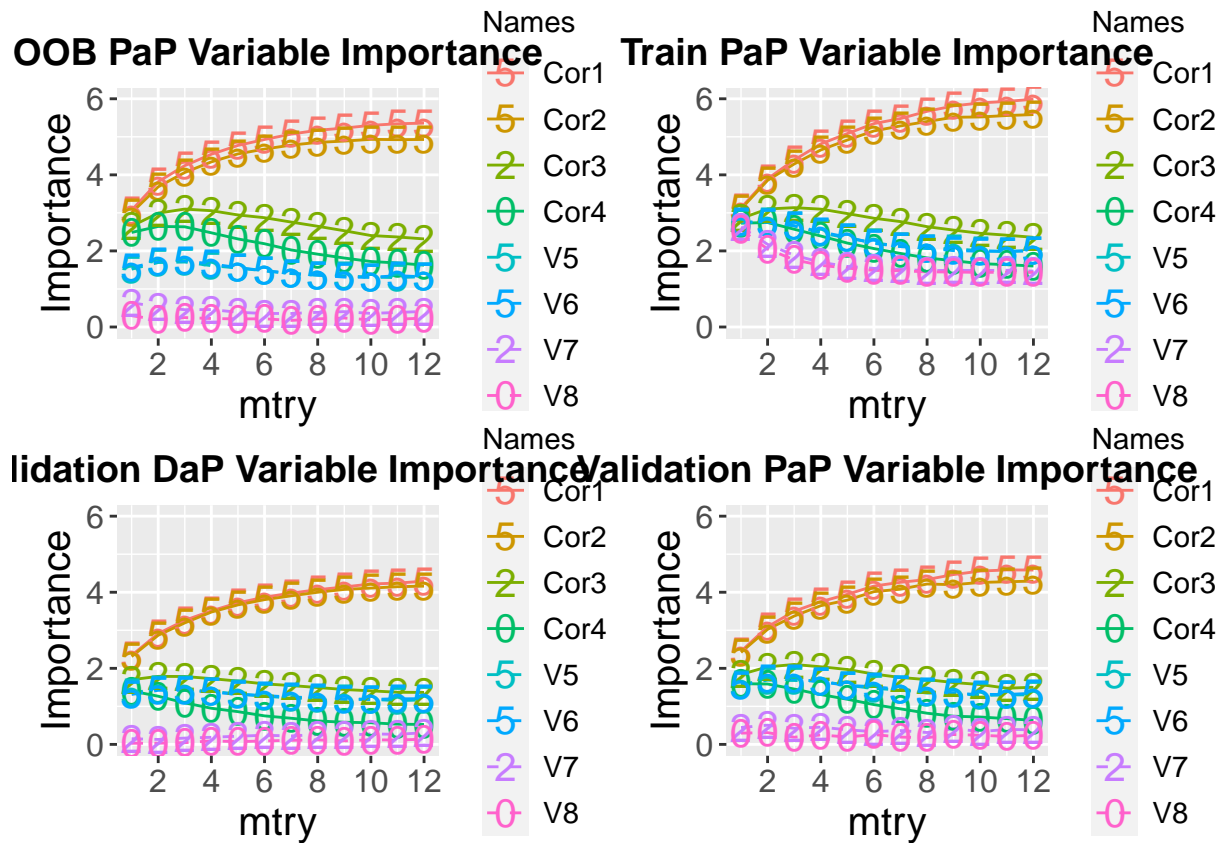
```
gv <- perm_valid1 %>%
  ggplot(aes(x = mtry, y = Imp, color = Names,
             group = Names, linetype = Names,
             shape = Names)) +
  geom_line() +
  scale_x_continuous(limits = c(1,12), breaks = seq(2,12,by=2)) +
  scale_y_continuous(limits = c(0,ma)) +
  ggtitle("Validation PaP Variable 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("xcos_all.pdf", plot = gr + gp + gd + gv, dpi = 2400,
       width = 9, height = 9)
```

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

```
## Time difference of 22.15713 mins
```