Rachael Blake

January 27, 2016

Random Forest Analysis Exploration for DMX Linkages

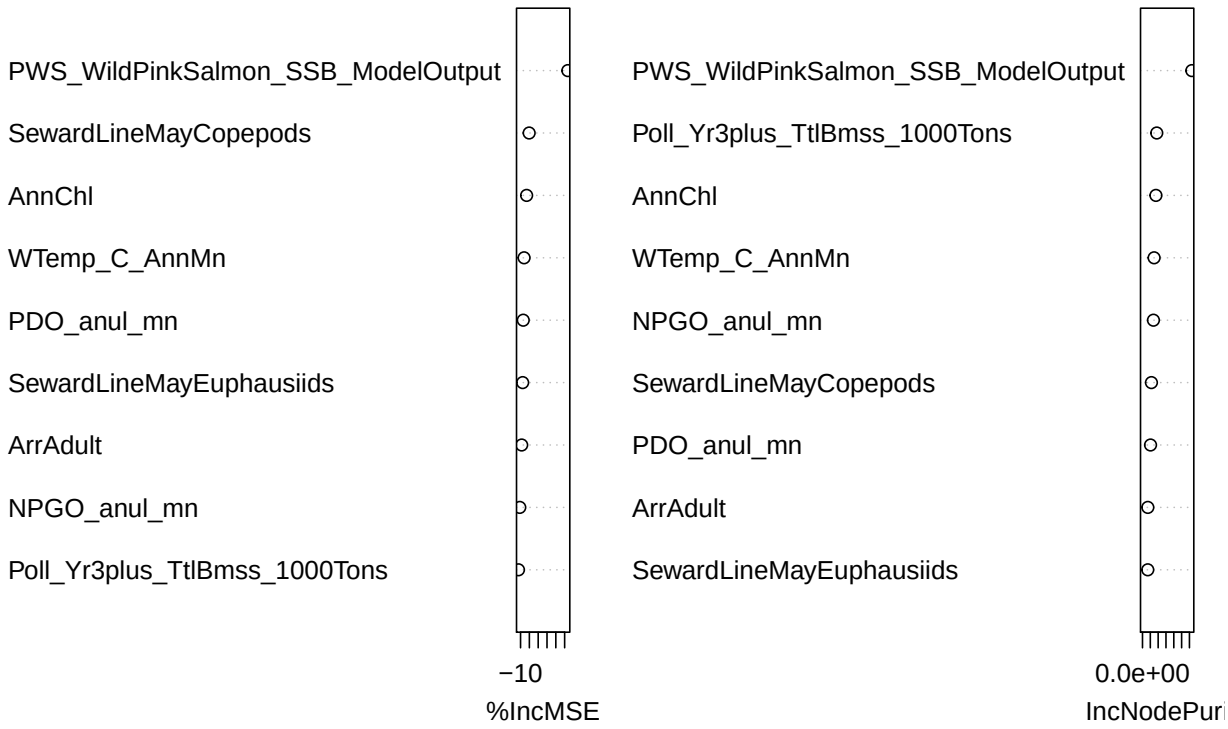
```
###
# Pink Salmon
CPD_Pink <- CPD %>%
            select(PDO_anul_mn,NPGO_anul_mn,#WndSp_m_s_AnnMn,WndDir_degT_AnnMn,
                   WTemp_C_AnnMn,AnnChl,SewardLineMayEuphausiids,SewardLineMayCopepods,
                   goaPinkCatchNum,PWS_WildPinkSalmon_SSB_ModelOutput,ArrAdult,
                   Poll_Yr3plus_TtlBmss_1000Tons) %>%
            filter(complete.cases(.))
goaPink <- randomForest(goaPinkCatchNum ~., data=CPD_Pink, importance=T, do.trace=1000, ntree=5000)
```

```
##      |      Out-of-bag   |
## Tree |      MSE  %Var(y) |
## 1000 | 6.754e+08    89.91 |
## 2000 | 6.647e+08    88.48 |
## 3000 | 6.621e+08    88.14 |
## 4000 | 6.706e+08    89.27 |
## 5000 | 6.615e+08    88.07 |
```

```
print(goaPink)
```

```
##
## Call:
##  randomForest(formula = goaPinkCatchNum ~ ., data = CPD_Pink,      importance = T, do.trace = 1000, n
##                Type of random forest: regression
##                      Number of trees: 5000
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 661539690
##                     % Var explained: 11.93
```

```
#plot(goaPink)
varImpPlot(goaPink)
```




```
goaPink$importance
```

```
##                                      %IncMSE IncNodePurity
## PDO_anul_mn                       -20090445.1     512657317
## NPGO_anul_mn                      -39270643.0     699152713
## WTemp_C_AnnMn                     -19089981.2     741417186
## AnnChl                             -9491189.1     858157379
## SewardLineMayEuphausiids          -20159957.2     336844264
## SewardLineMayCopepods               -609485.2     569584662
## PWS_WildPinkSalmon_SSB_ModelOutput 435618400.7    3151921292
## ArrAdult                          -22529767.7     341388813
## Poll_Yr3plus_TtlBmss_1000Tons     -52628342.0     911916150
```

```
# Arrowtooth
CPD_Arrow <- CPD %>%
             select(PDO_anul_mn,NPGO_anul_mn,#WndSp_m_s_AnnMn,WndDir_degT_AnnMn,
                    WTemp_C_AnnMn,AnnChl,SewardLineMayEuphausiids,Poll_Age1_recruits_millions,
                    ArrAdult,arth_tons,arth_real_rev,arth_vessels,arth_processors,arth_real_price) %>%
             filter(complete.cases(.))
arrow <- randomForest(ArrAdult ~., data=CPD_Arrow, importance=T, do.trace=1000, ntree=5000)
```

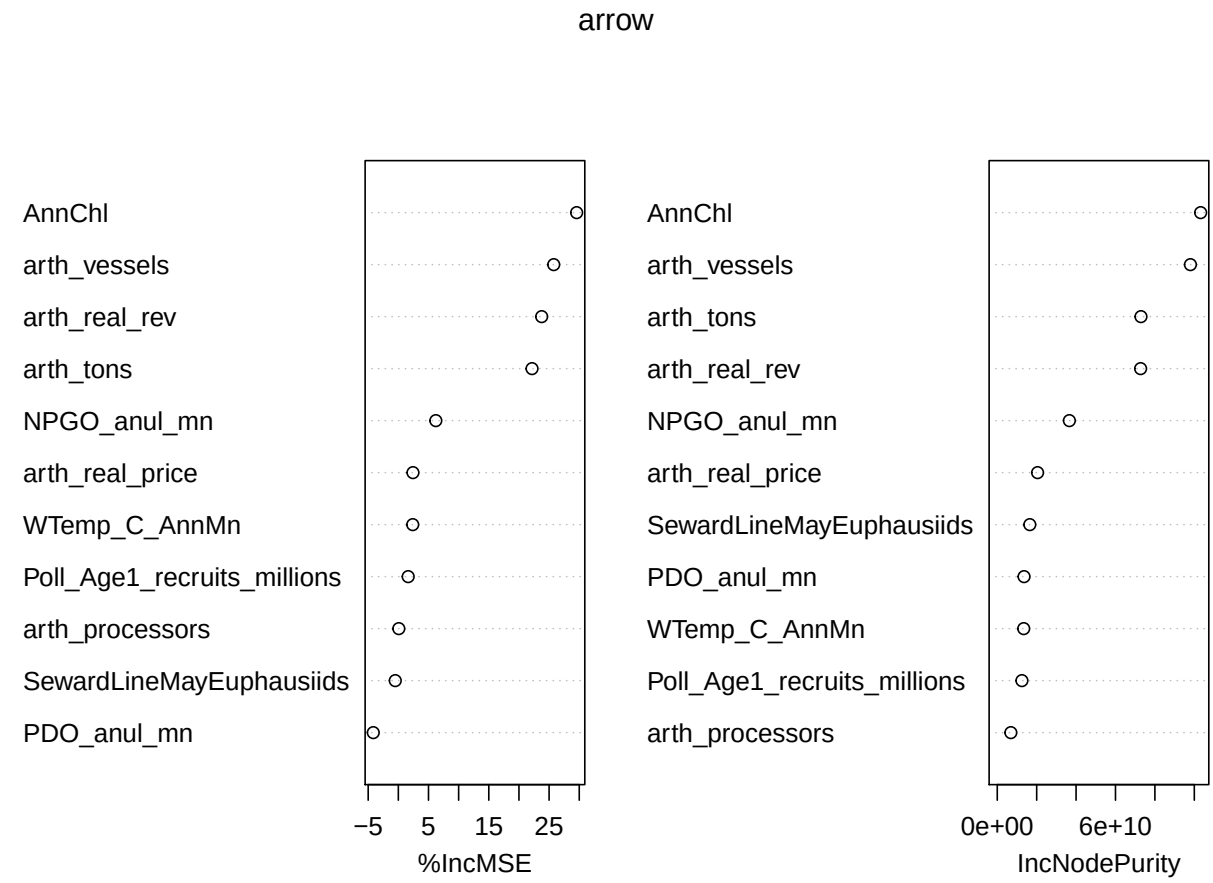
```
##      |      Out-of-bag   |
```

```
## Tree |      MSE  %Var(y) |
## 1000 | 1.365e+10    33.21 |
## 2000 | 1.342e+10    32.65 |
## 3000 | 1.376e+10    33.50 |
## 4000 | 1.351e+10    32.88 |
## 5000 | 1.335e+10    32.50 |
```

```
print(arrow)
```

```
##
## Call:
##  randomForest(formula = ArrAdult ~ ., data = CPD_Arrow, importance = T,      do.trace = 1000, ntree =
##                Type of random forest: regression
##                      Number of trees: 5000
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 13353600825
##                     % Var explained: 67.5
```

```
#plot(arrow)
varImpPlot(arrow)
```



```
arrow$importance
```

```
##                                %IncMSE IncNodePurity
## PDO_anul_mn                 -477160151   13594530483
## NPGO_anul_mn                1157942414   36595326288
## WTemp_C_AnnMn                201158865   13429955280
## AnnChl                     12602062035  103229453363
## SewardLineMayEuphausiids      -44158412   16515360674
## Poll_Age1_recruits_millions   141740340   12514181111
## arth_tons                    6453167900   72930974461
## arth_real_rev                7701157446   72749312442
## arth_vessels                 9309516115   98034094984
## arth_processors                 5286482    6907744158
## arth_real_price               250486278   20470993587
```

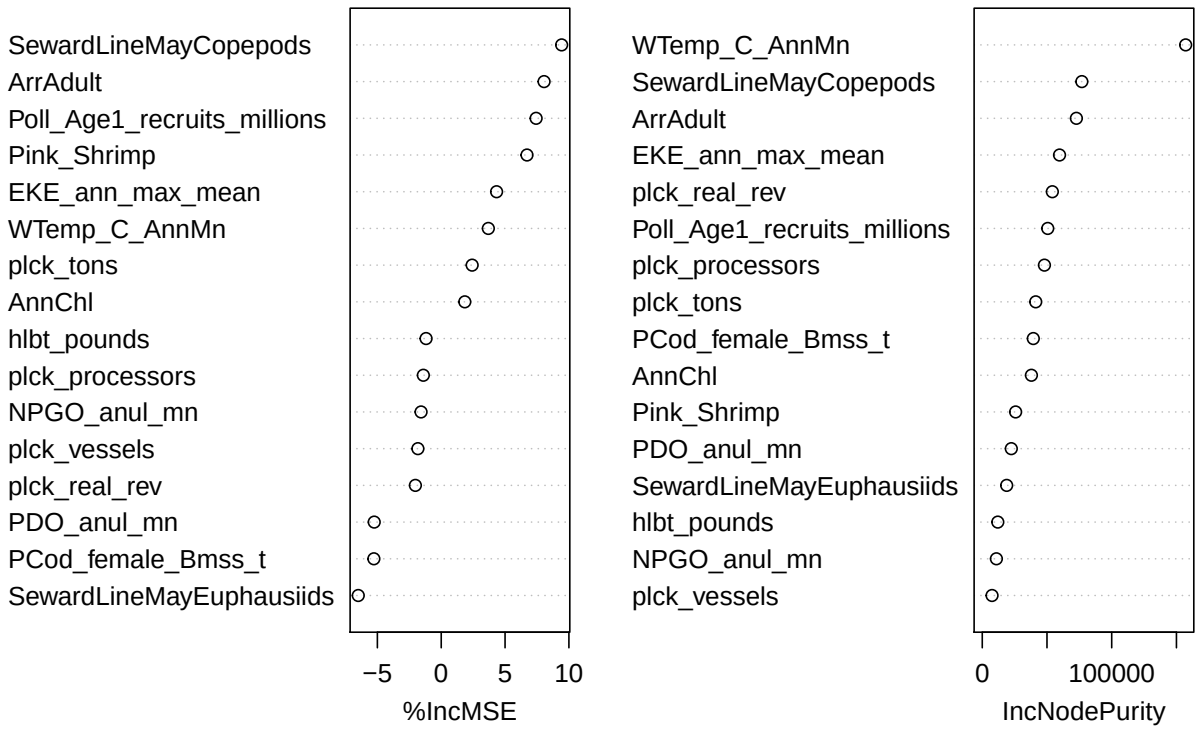
```
# Pollock Adults
CPD_Poll <- CPD %>%
            select(PDO_anul_mn,NPGO_anul_mn,#WndSp_m_s_AnnMn,WndDir_degT_AnnMn,
                   WTemp_C_AnnMn,AnnChl,EKE_ann_max_mean,SewardLineMayEuphausiids,
                   SewardLineMayCopepods,Pink_Shrimp,Poll_Age1_recruits_millions,
                   Poll_Yr3plus_TtlBmss_1000Tons,ArrAdult, PCod_female_Bmss_t,hlbt_pounds,
                   plck_tons,plck_real_rev,plck_vessels,plck_processors) %>%
            filter(complete.cases(.))
poll_a <- randomForest(Poll_Yr3plus_TtlBmss_1000Tons ~., data=CPD_Poll, importance=T, do.trace=1000, nt
```

```
##      |      Out-of-bag   |
## Tree |      MSE  %Var(y) |
## 1000 | 7.358e+04   109.83 |
## 2000 | 7.423e+04   110.80 |
## 3000 | 7.464e+04   111.41 |
## 4000 | 7.464e+04   111.41 |
## 5000 | 7.538e+04   112.52 |
```

```
print(poll_a)
```

```
##
## Call:
##  randomForest(formula = Poll_Yr3plus_TtlBmss_1000Tons ~ ., data = CPD_Poll,      importance = T, do.
##                Type of random forest: regression
##                      Number of trees: 5000
## No. of variables tried at each split: 5
##
##           Mean of squared residuals: 75381.6
##                     % Var explained: -12.52
```

```
#plot(poll_a)
varImpPlot(poll_a)
```




```
poll_a$importance
```

```
##                               %IncMSE IncNodePurity
## PDO_anul_mn                 -957.0998     22427.432
## NPGO_anul_mn                -212.1263     10913.432
## WTemp_C_AnnMn               1938.3586    157129.152
## AnnChl                       404.3940     37969.837
## EKE_ann_max_mean            1073.1554     59694.484
## SewardLineMayEuphausiids   -1299.7921     18862.370
## SewardLineMayCopepods       3428.9089     76934.817
## Pink_Shrimp                 1282.4778     25780.672
## Poll_Age1_recruits_millions 2318.9059     50590.419
## ArrAdult                    2981.4986     72709.491
## PCod_female_Bmss_t         -1295.4460     39429.603
## hlbt_pounds                 -141.7567     12002.622
## plck_tons                    546.0648     41309.136
## plck_real_rev               -326.1507     54135.794
## plck_vessels                -147.3356      7549.144
## plck_processors             -208.4154     48028.514
```

```
# Pollock Juvenile
CPD_JPoll <- CPD %>%
             select(PDO_anul_mn,NPGO_anul_mn,#WndSp_m_s_AnnMn,WndDir_degT_AnnMn,
```

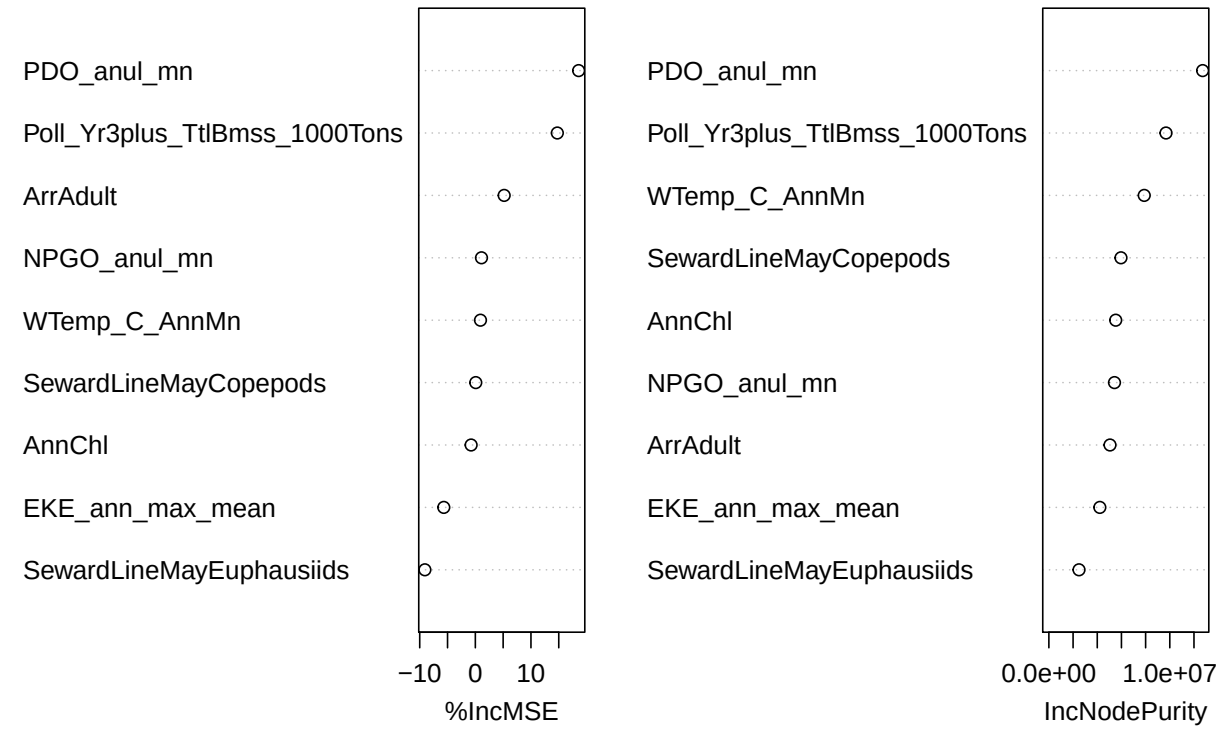
```
                    WTemp_C_AnnMn,AnnChl, EKE_ann_max_mean,SewardLineMayEuphausiids,
                    SewardLineMayCopepods,Poll_Age1_recruits_millions,
                    Poll_Yr3plus_TtlBmss_1000Tons,ArrAdult) %>%
             filter(complete.cases(.))
poll_j <- randomForest(Poll_Age1_recruits_millions ~., data=CPD_JPoll, importance=T, do.trace=1000, ntr
```

```
##      |      Out-of-bag   |
## Tree |      MSE  %Var(y) |
## 1000 | 5.598e+06   102.14 |
## 2000 | 5.513e+06   100.59 |
## 3000 | 5.531e+06   100.93 |
## 4000 | 5.534e+06   100.97 |
## 5000 | 5.515e+06   100.63 |
```

```
print(poll_j)
```

```
##
## Call:
##  randomForest(formula = Poll_Age1_recruits_millions ~ ., data = CPD_JPoll,      importance = T, do.t
##                Type of random forest: regression
##                      Number of trees: 5000
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 5514883
##                     % Var explained: -0.63
```

```
#plot(poll_j)
varImpPlot(poll_j)
```




```
poll_j$importance
```

```
##                                %IncMSE IncNodePurity
## PDO_anul_mn                 833385.356      12716167
## NPGO_anul_mn                 29712.021       5425393
## WTemp_C_AnnMn                27613.027       7884966
## AnnChl                      -17849.302       5527456
## EKE_ann_max_mean           -129918.422       4213952
## SewardLineMayEuphausiids   -182148.119       2491129
## SewardLineMayCopepods         2053.969       5961496
## Poll_Yr3plus_TtlBmss_1000Tons  534448.850       9685797
## ArrAdult                      131449.090       5054524
```

```
# Halibut
CPD_Hal <- CPD %>%
           select(PDO_anul_mn,NPGO_anul_mn,#WndSp_m_s_AnnMn,WndDir_degT_AnnMn,
                  WTemp_C_AnnMn,AnnChl,Poll_Yr3plus_TtlBmss_1000Tons, Poll_Age1_recruits_millions,
                  TotTCrab,Pink_Shrimp,hlbt_pounds,hlbt_real_rev,hlbt_vessels,hlbt_processors) %>%
           filter(complete.cases(.))
hlbt_lbs <- randomForest(hlbt_pounds ~., data=CPD_Hal, importance=T, do.trace=1000, ntree=5000)
```

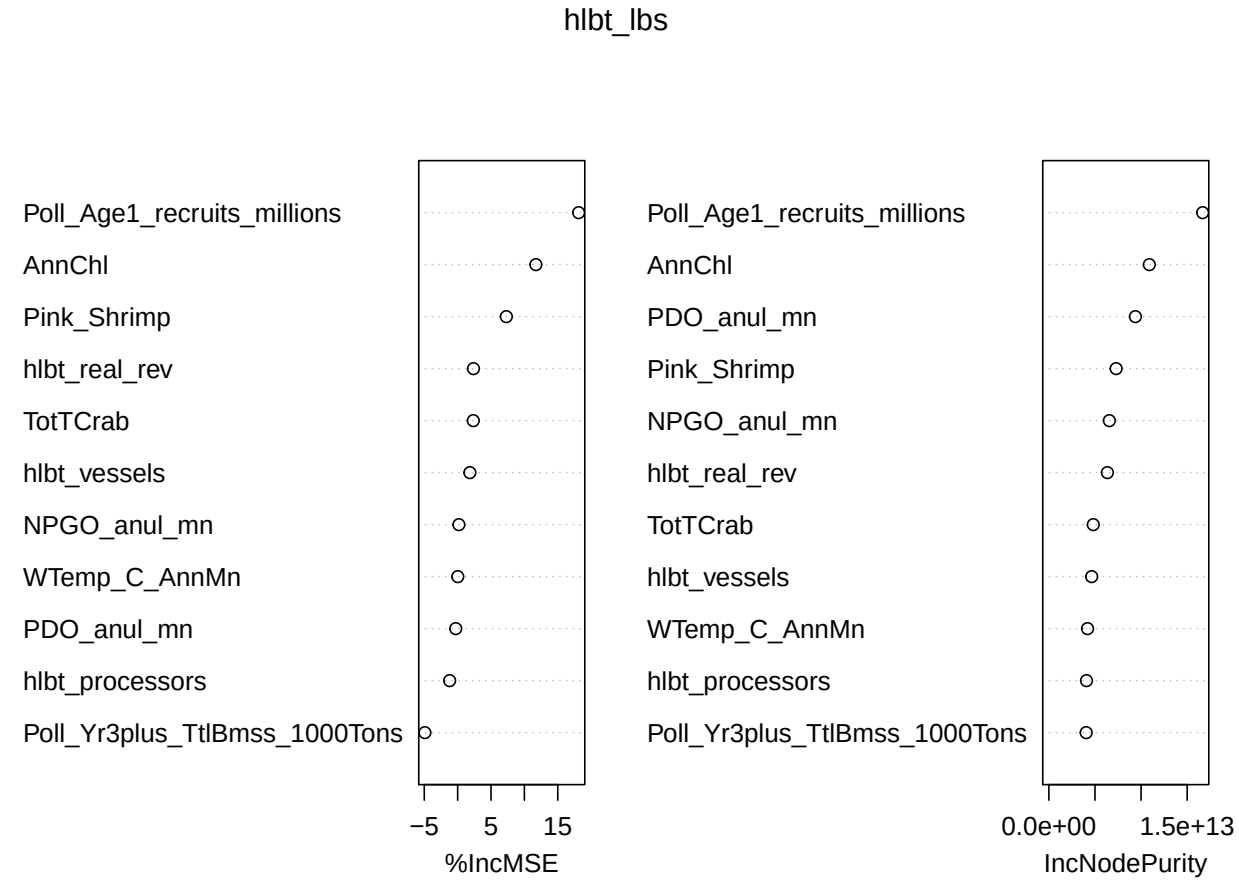
```
##      |      Out-of-bag   |
```

```
## Tree |      MSE  %Var(y) |
## 1000 | 8.159e+12   107.56 |
## 2000 | 8.152e+12   107.47 |
## 3000 | 8.067e+12   106.34 |
## 4000 | 8.059e+12   106.25 |
## 5000 | 8.051e+12   106.14 |
```

```
print(hlbt_lbs)
```

```
##
## Call:
##  randomForest(formula = hlbt_pounds ~ ., data = CPD_Hal, importance = T,      do.trace = 1000, ntree
##                Type of random forest: regression
##                      Number of trees: 5000
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 8.050789e+12
##                     % Var explained: -6.14
```

```
#plot(hlbt_lbs)
varImpPlot(hlbt_lbs)
```



```
hlbt_lbs$importance
```

```
##                                    %IncMSE IncNodePurity
## PDO_anul_mn                    -10739664795  9.383004e+12
## NPGO_anul_mn                     6423047180  6.559912e+12
## WTemp_C_AnnMn                     552018766  4.194333e+12
## AnnChl                          513692226449  1.089447e+13
## Poll_Yr3plus_TtlBmss_1000Tons -121932501831  4.046998e+12
## Poll_Age1_recruits_millions   1069456441911  1.667865e+13
## TotTCrab                        73342931198  4.814391e+12
## Pink_Shrimp                    227494056533  7.294152e+12
## hlbt_real_rev                   65798174468  6.344041e+12
## hlbt_vessels                    52743591728  4.642313e+12
## hlbt_processors                -32638972199  4.078487e+12
```