w04-2

wk04-2-REG-df2015na

- 실행시간 측정

```
time1 <- Sys.time()
time1
```

```
## [1] "2021-05-09 02:50:53 KST"
```

회귀모형

자료 설명

- Size Korea 2015년 인체계측자료 일부: n=300 (nc=268), d=13

패키지

```
suppressWarnings(suppressMessages(library(tidyverse)))
library(tidymodels)
```

```
## Warning: package 'tidymodels' was built under R version 4.0.5
```

```
## -- Attaching packages -------------------------------------- tidymodels 0.1.3 --
```

```
## v broom        0.7.6      v rsample      0.1.0
## v dials        0.0.9      v tune         0.1.5
## v infer        0.5.4      v workflows    0.2.2
## v modeldata    0.1.0      v workflowsets 0.0.2
## v parsnip      0.1.5      v yardstick    0.0.8
## v recipes      0.1.16
```

```
## Warning: package 'broom' was built under R version 4.0.5
```

```
## Warning: package 'dials' was built under R version 4.0.5
```

```
## Warning: package 'infer' was built under R version 4.0.5
```

```
## Warning: package 'modeldata' was built under R version 4.0.5
```

```
## Warning: package 'parsnip' was built under R version 4.0.5
```

```
## Warning: package 'recipes' was built under R version 4.0.5
```

```
## Warning: package 'tune' was built under R version 4.0.5
```

```
## Warning: package 'workflows' was built under R version 4.0.5
```

```
## Warning: package 'workflowsets' was built under R version 4.0.5
```

```
## Warning: package 'yardstick' was built under R version 4.0.5
```

```
## -- Conflicts --------------------------------------- tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
## * Use tidymodels_prefer() to resolve common conflicts.
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.0.5
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following objects are masked from 'package:yardstick':
##
##     precision, recall, sensitivity, specificity
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(skimr)
library(naniar)
```

```
##
## Attaching package: 'naniar'
```

```
## The following object is masked from 'package:skimr':
##
##     n_complete
```

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(scales)
```

```
DF <- as.data.frame(read_csv('D:/Github/Statics/DataMining/0509/df2015na.csv'))
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   gnd = col_character(),
##   age = col_double(),
##   ht = col_double(),
##   wt = col_double(),
##   wa = col_double(),
##   hdln = col_double(),
##   hdwd = col_double(),
##   ftln = col_double(),
##   ftwd = col_double(),
##   bld = col_character(),
##   lft = col_double(),
##   smk = col_double(),
##   alc = col_double()
## )
```

```
dim(DF)
```

```
## [1] 300  13
```

```
str(DF)
```

```
## 'data.frame':    300 obs. of  13 variables:
##  $ gnd : chr  "M" "M" "F" "F" ...
##  $ age : num  42 23 32 30 NA 26 20 24 18 58 ...
##  $ ht  : num  165 188 162 162 160 ...
##  $ wt  : num  79 77.8 59.6 52 58.9 76.3 61.4 63 65.2 58.3 ...
##  $ wa  : num  96.4 76.6 83 65.8 75 83.8 70.1 72.4 73.5 83.1 ...
##  $ hdln: num  17.9 20.4 17.5 16.4 17.4 19 19.8 18.5 20.1 15.6 ...
##  $ hdwd: num  8.6 8.3 7.7 6.6 7.7 8.6 7.8 7.9 8 7.6 ...
##  $ ftln: num  24.5 28.9 23.3 23.7 24.1 25.7 26.5 25.7 25.8 22.1 ...
##  $ ftwd: num  9.6 10.6 10.4 8.4 9.4 10.7 9.6 10.2 10.3 8.7 ...
##  $ bld : chr  "0" "0" "A" "B" ...
##  $ lft : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ smk : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ alc : num  1 0 0 0 0 0 0 0 0 0 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   gnd = col_character(),
##   ..   age = col_double(),
##   ..   ht = col_double(),
##   ..   wt = col_double(),
##   ..   wa = col_double(),
##   ..   hdln = col_double(),
##   ..   hdwd = col_double(),
##   ..   ftln = col_double(),
##   ..   ftwd = col_double(),
##   ..   bld = col_character(),
##   ..   lft = col_double(),
##   ..   smk = col_double(),
##   ..   alc = col_double()
##   .. )
```

```
head(DF)
```

```
##   gnd age    ht   wt   wa hdln hdwd ftln ftwd bld lft smk alc
## 1   M  42 165.0 79.0 96.4 17.9  8.6 24.5  9.6   0   0   0   1
## 2   M  23 187.5 77.8 76.6 20.4  8.3 28.9 10.6   0   0   1   0
## 3   F  32 161.7 59.6 83.0 17.5  7.7 23.3 10.4   A   0   0   0
## 4   F  30 162.0 52.0 65.8 16.4  6.6 23.7  8.4   B   0   0   0
## 5   F  NA 160.0 58.9 75.0 17.4  7.7 24.1  9.4   A   0   0   0
## 6   M  26 179.0 76.3 83.8 19.0  8.6 25.7 10.7   B   0   0   0
```

변수 조정

- 문자변수 (gnd, bld)를 factor화
- {0,1}로 코딩된 이산형 변수를 숫자로 처리하거나 factor해서 사용가능

```
DF <- mutate(DF,
             gnd=factor(gnd), bld=factor(bld),
             lft=factor(lft, labels=c('N','Y')),
             smk=factor(smk, labels=c('N', 'Y')),
             alc=factor(alc, labels=c('N', 'Y')))
str(DF)
```

```
## 'data.frame':    300 obs. of  13 variables:
##  $ gnd : Factor w/ 2 levels "F","M": 2 2 1 1 1 2 2 2 2 1 ...
##  $ age : num  42 23 32 30 NA 26 20 24 18 58 ...
##  $ ht  : num  165 188 162 162 160 ...
##  $ wt  : num  79 77.8 59.6 52 58.9 76.3 61.4 63 65.2 58.3 ...
##  $ wa  : num  96.4 76.6 83 65.8 75 83.8 70.1 72.4 73.5 83.1 ...
##  $ hdln: num  17.9 20.4 17.5 16.4 17.4 19 19.8 18.5 20.1 15.6 ...
##  $ hdwd: num  8.6 8.3 7.7 6.6 7.7 8.6 7.8 7.9 8 7.6 ...
##  $ ftln: num  24.5 28.9 23.3 23.7 24.1 25.7 26.5 25.7 25.8 22.1 ...
##  $ ftwd: num  9.6 10.6 10.4 8.4 9.4 10.7 9.6 10.2 10.3 8.7 ...
##  $ bld : Factor w/ 4 levels "A","AB","B","O": 4 4 1 3 1 3 2 1 4 3 ...
##  $ lft : Factor w/ 2 levels "N","Y": 1 1 1 1 1 1 1 1 1 1 ...
##  $ smk : Factor w/ 2 levels "N","Y": 1 2 1 1 1 1 1 1 1 1 ...
##  $ alc : Factor w/ 2 levels "N","Y": 2 1 1 1 1 1 1 1 1 1 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   gnd = col_character(),
##   ..   age = col_double(),
##   ..   ht = col_double(),
##   ..   wt = col_double(),
##   ..   wa = col_double(),
##   ..   hdln = col_double(),
##   ..   hdwd = col_double(),
##   ..   ftln = col_double(),
##   ..   ftwd = col_double(),
##   ..   bld = col_character(),
##   ..   lft = col_double(),
##   ..   smk = col_double(),
##   ..   alc = col_double()
##   .. )
```

결측

- skimr::skim(data, …): summary()에 결측정보를 추가. group_by와 연결
- 결측 현황: skim이나 naniar로 확인

```
DF %>% skim()
```

Data summary

Name	Piped data
Number of rows	300
Number of columns	13
_____	
Column type frequency:	
factor	5
numeric	8
_____	
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
gnd	0	1.00	FALSE	2	M: 155, F: 145
bld	0	1.00	FALSE	4	B: 110, A: 84, O: 68, AB: 38
lft	5	0.98	FALSE	2	N: 283, Y: 12
smk	2	0.99	FALSE	2	N: 237, Y: 61
alc	6	0.98	FALSE	2	N: 256, Y: 38

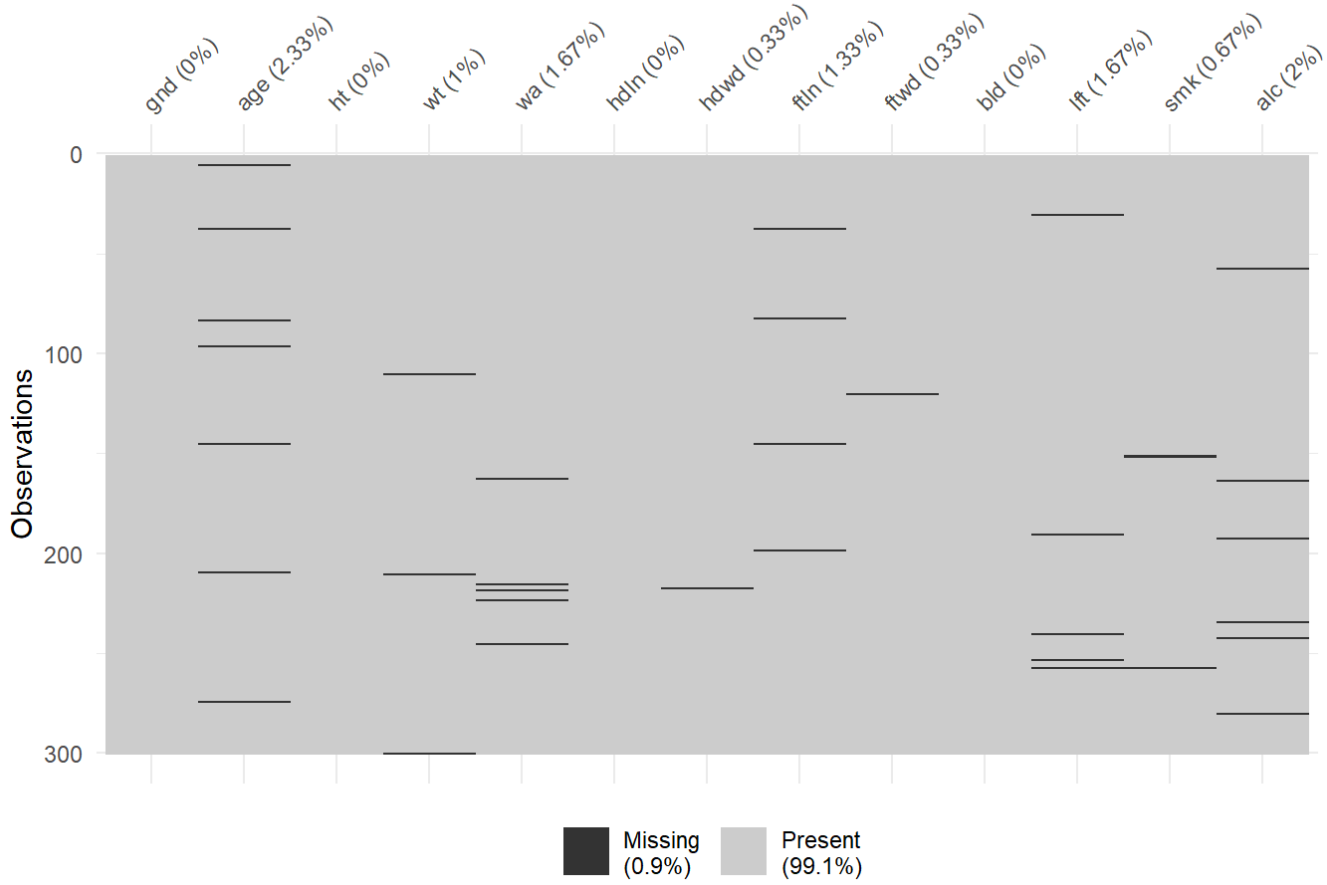
Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
age	7	0.98	29.91	13.33	15.0	19.0	26.00	37.00	68.0	
ht	0	1.00	165.64	9.03	144.3	158.0	165.85	172.83	190.5	
wt	3	0.99	63.77	12.38	39.9	55.6	61.80	71.20	106.1	
wa	5	0.98	78.23	9.72	59.0	71.0	77.20	84.15	111.6	
hdln	0	1.00	17.57	1.09	14.9	16.8	17.50	18.30	20.8	
hdwd	1	1.00	7.80	0.55	6.4	7.4	7.80	8.20	9.2	
ftln	4	0.99	24.37	1.51	19.9	23.3	24.30	25.50	28.9	
ftwd	1	1.00	9.68	0.70	8.0	9.2	9.70	10.15	12.2	

```
# 완전한 관측값 비율
sum(complete.cases(DF))/nrow(DF)*100
```

```
## [1] 89.66667
```

```
# 변수별 결측비율
naniar::vis_miss(DF)
```



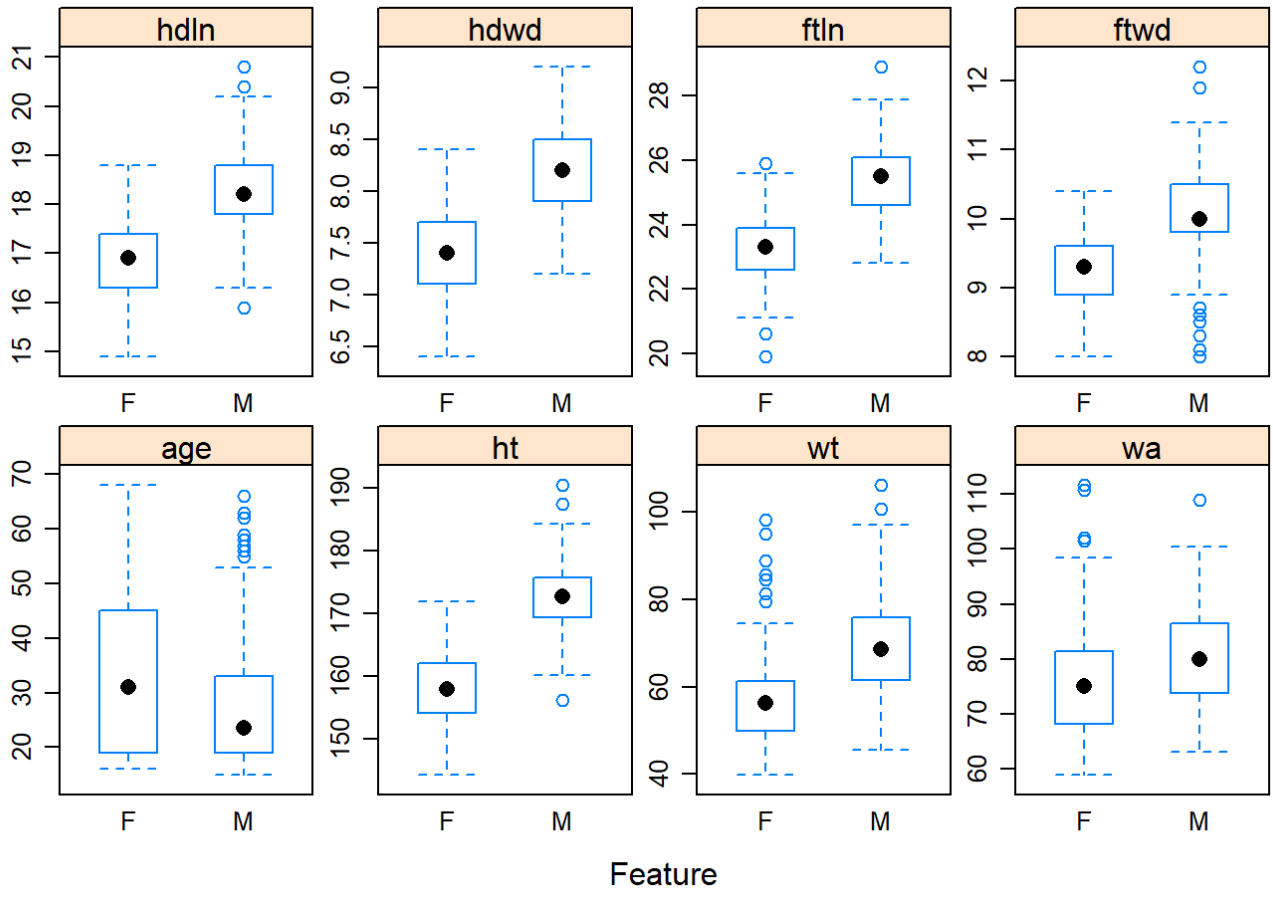

```
naniar::miss_var_summary(DF)
```

```
## # A tibble: 13 x 3
##    variable n_miss pct_miss
##    <chr>     <int>    <dbl>
##  1 age           7    2.33
##  2 alc           6    2
##  3 wa            5    1.67
##  4 lft           5    1.67
##  5 ftln          4    1.33
##  6 wt            3    1
##  7 smk           2    0.667
##  8 hdwd          1    0.333
##  9 ftwd          1    0.333
## 10 gnd           0    0
## 11 ht            0    0
## 12 hdln          0    0
## 13 bld           0    0
```

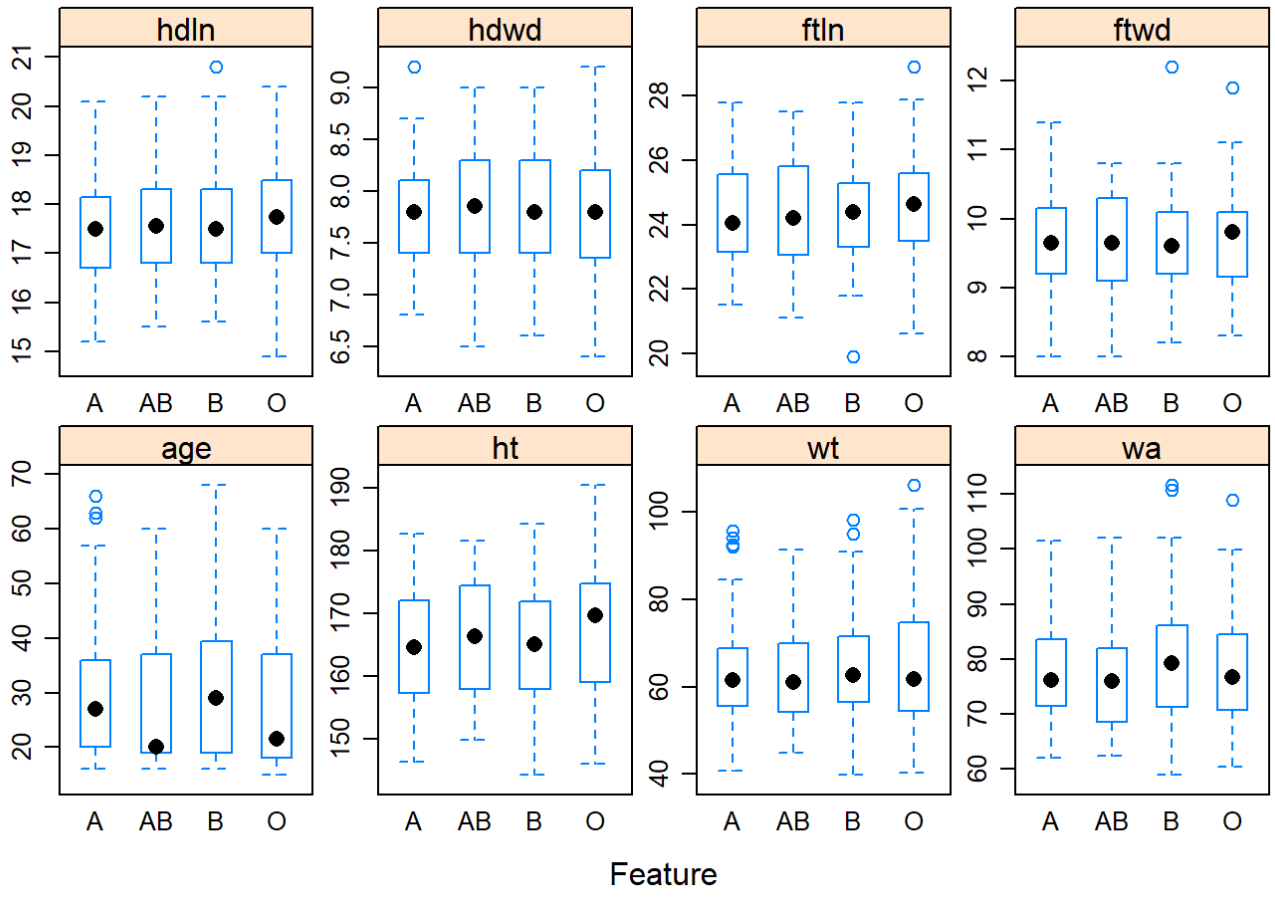
간단 탐색

- caret::featurePlot(x:연속형, y:요인, plot='strip,scatter,box,density')

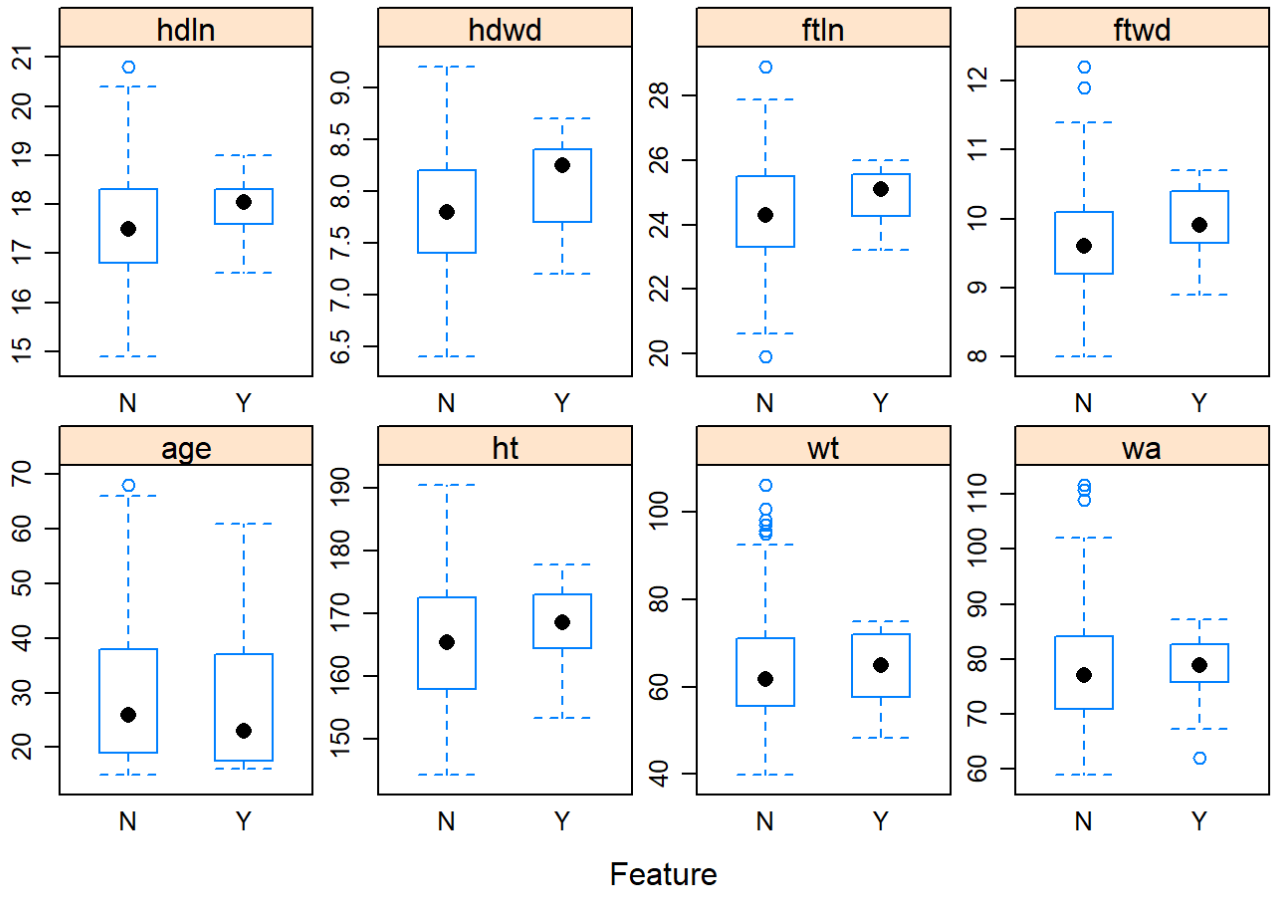
```
featurePlot(x=select_if(DF, is.numeric), y=DF$gnd,
            plot='box',
            scales=list(x=list(relation='free'), y=list(relation='free')))
```



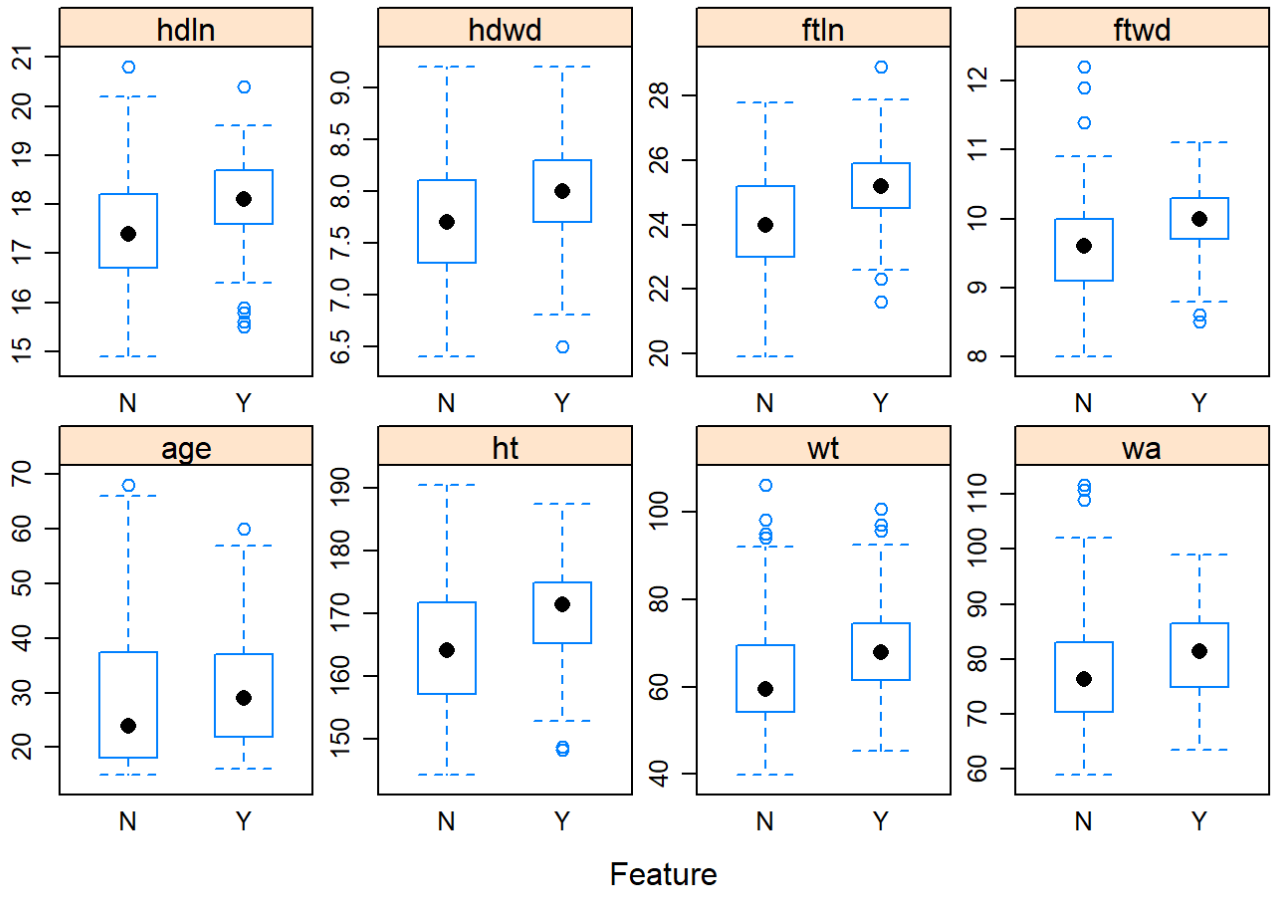

```
featurePlot(x=select_if(DF, is.numeric), y=DF$bld,
            plot='box',
            scales=list(x=list(relation='free'), y=list(relation='free')))
```



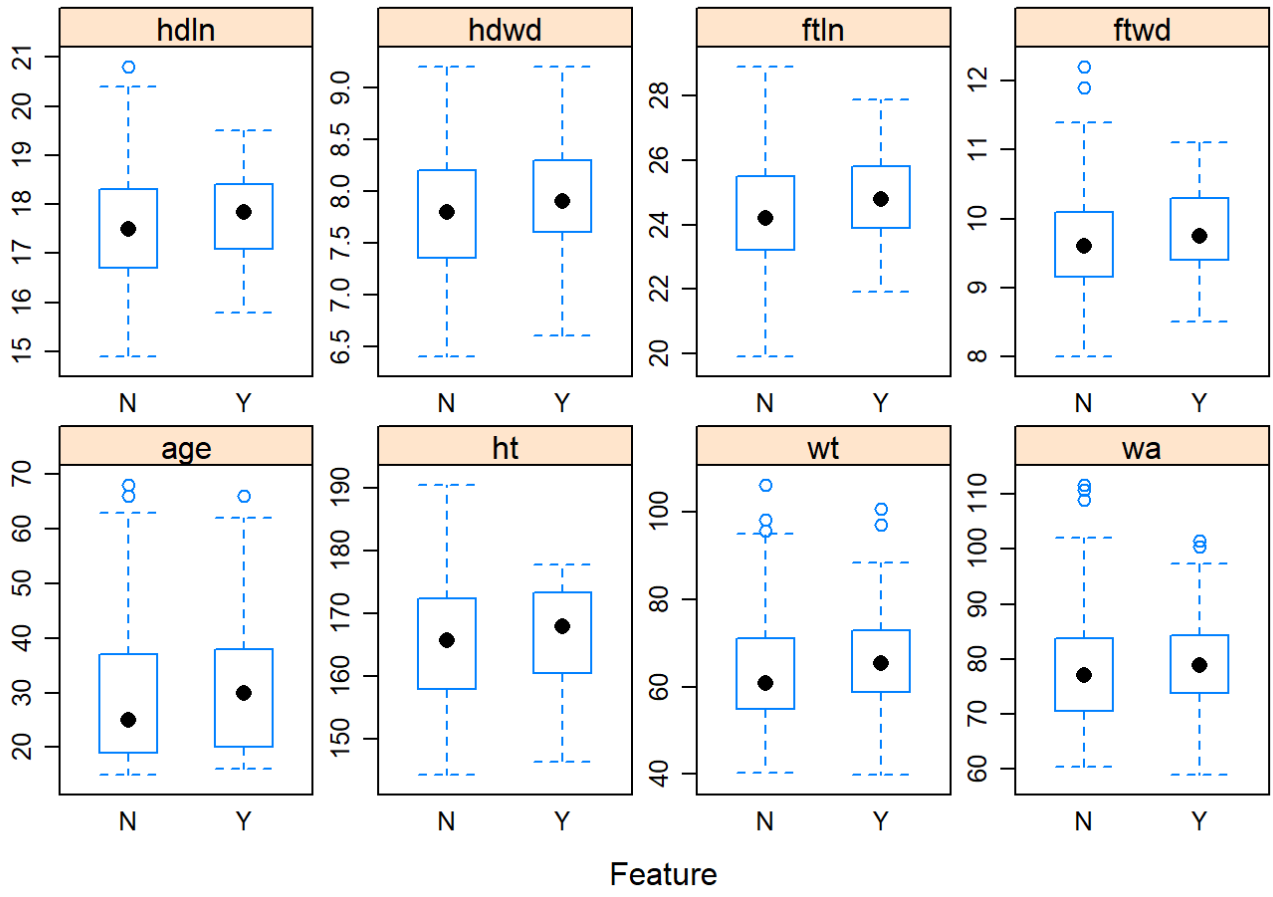

```
featurePlot(x=select_if(DF, is.numeric), y=DF$lft,
            plot='box',
            scales=list(x=list(relation='free'), y=list(relation='free')))
```




```
featurePlot(x=select_if(DF, is.numeric), y=DF$smk,
            plot='box',
            scales=list(x=list(relation='free'), y=list(relation='free')))
```




```
featurePlot(x=select_if(DF, is.numeric), y=DF$alc,
            plot='box',
            scales=list(x=list(relation='free'), y=list(relation='free')))
```



분할/예측값 저장소 준비

- TR:TS를 0.75:0.25로 분할

```
set.seed(20180968)

# rsplit 객체: strata에 NA가 있으면 안 됨
Ris <- initial_split(DF, prop=0.75)
TR <- training(Ris)
TS <- testing(Ris)

#예측값을 저장할 장소
TROUT <- TR %>% dplyr::select(ht)
TSOUT <- TS %>% dplyr::select(ht)
```

전처리

- recipe 객체 생성
  - 연속형 변수는 medianimpute
  - 이산형 변수(이진 가변수 포함)는 modeimpute
  - 이산형 변수(이진 가변수 제외)는 dummy화

```
RC <-
  recipe(ht~., data=TR) %>%
  step_medianimpute(all_numeric(), -all_outcomes()) %>%  #age, wt, wa, hdln, hdwd, ftln, ftwd
  step_modeimpute(all_nominal(), -all_outcomes()) %>%    #gnd, bld, lft, smk, alc
  step_dummy(all_nominal(), -all_outcomes())
```

```
## Warning: `step_modeimpute()` was deprecated in recipes 0.1.16.
## Please use `step_impute_mode()` instead.
```

```
## Warning: `step_medianimpute()` was deprecated in recipes 0.1.16.
## Please use `step_impute_median()` instead.
```

```
RC
```

```
## Data Recipe
##
## Inputs:
##
##       role #variables
##    outcome          1
##  predictor         12
##
## Operations:
##
## Median Imputation for all_numeric(), -all_outcomes()
## Mode Imputation for all_nominal(), -all_outcomes()
## Dummy variables from all_nominal(), -all_outcomes()
```

튜링계획 지정

```
trCntl <- trainControl(method='cv', number=10)
```

lm: 선형회귀모형

- 튜닝모수 없음. intercept는 튜닝 안함

```
modelLookup('lm')
```

```
##   model parameter     label forReg forClass probModel
## 1    lm intercept intercept   TRUE    FALSE     FALSE
```

- 적합

```
set.seed(20180968)
Mlm <- train(RC, data=TR,
             method='lm',
             trControl = trCntl)
             # metric='RMSE.Rsquared' 회귀 모형 선택 기준

Mlm
```

```
## Linear Regression
##
## 225 samples
##  12 predictor
##
## Recipe steps: impute_median, impute_mode, dummy
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 203, 201, 203, 204, 201, 203, ...
## Resampling results:
##
##   RMSE      Rsquared  MAE
##   3.240378  0.875855  2.607549
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

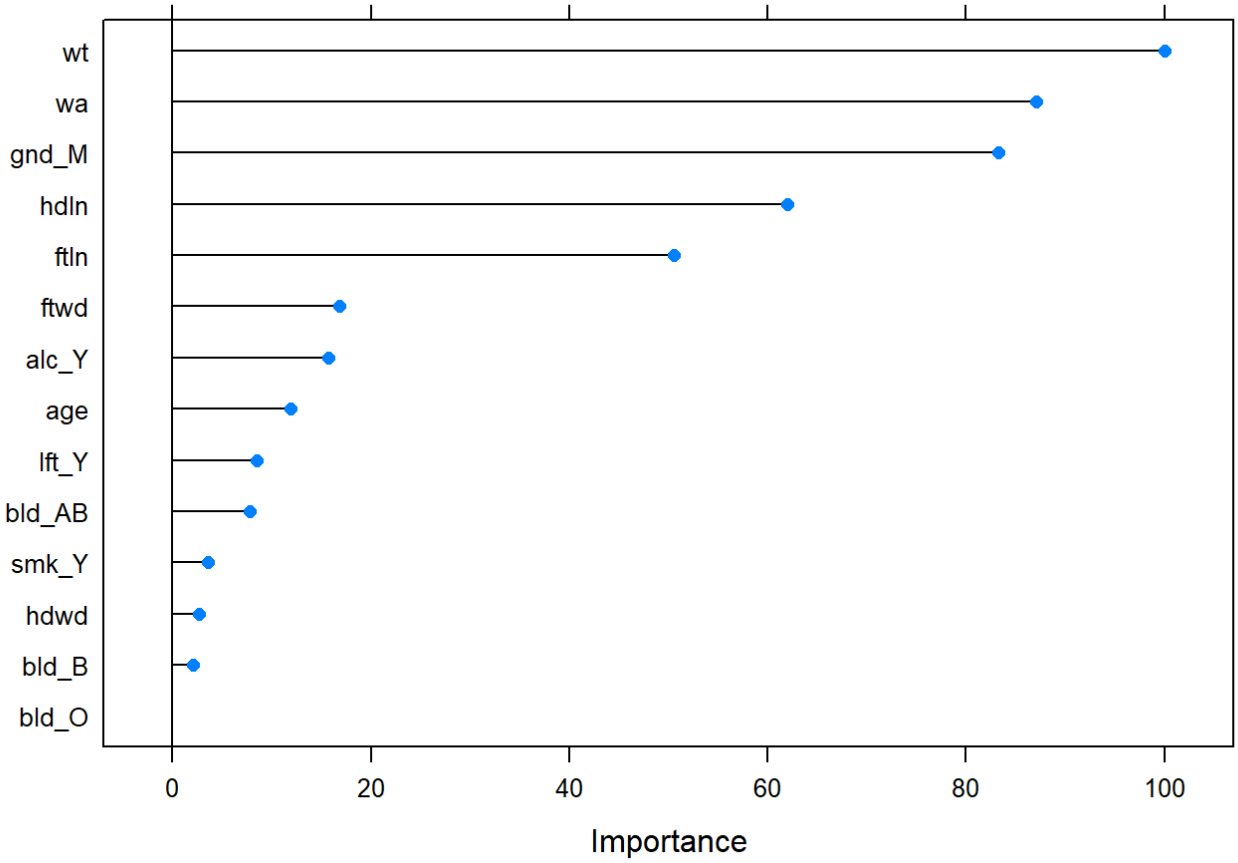
```
Mlm$results
```

```
##   intercept     RMSE Rsquared      MAE    RMSESD RsquaredSD     MAESD
## 1      TRUE 3.240378 0.875855 2.607549 0.4049714 0.02847711 0.3627034
```

```
# (X) plot(Mlm)
summary(Mlm)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2703 -2.2954  0.0097  1.8560  8.9259
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 106.73219    6.84843  15.585  < 2e-16 ***
## age          -0.02557    0.02305  -1.110    0.268
## wt            0.41691    0.04854   8.589 1.99e-15 ***
## wa           -0.41080    0.05480  -7.497 1.81e-12 ***
## hdln          2.10581    0.39249   5.365 2.13e-07 ***
## hdwd         -0.24662    0.74068  -0.333    0.739
## ftln          1.40383    0.31925   4.397 1.74e-05 ***
## ftwd         -0.71519    0.46609  -1.534    0.126
## gnd_M         5.57678    0.77733   7.174 1.23e-11 ***
## bld_AB        0.54383    0.71326   0.762    0.447
## bld_B         0.15639    0.56081   0.279    0.781
## bld_O        -0.06370    0.61290  -0.104    0.917
## lft_Y        -0.96495    1.17149  -0.824    0.411
## smk_Y         0.23101    0.57382   0.403    0.688
## alc_Y        -0.93218    0.64826  -1.438    0.152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.179 on 210 degrees of freedom
## Multiple R-squared:  0.8834, Adjusted R-squared:  0.8756
## F-statistic: 113.7 on 14 and 210 DF,  p-value: < 2.2e-16
```

```
plot(varImp(Mlm))
```




```
Mlm$bestTune
```

```
##   intercept
## 1      TRUE
```

```
Mlm$finalModel #lm객체(TR을 재적합한 모형)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Coefficients:
## (Intercept)          age           wt           wa         hdln         hdwd
##   106.73219     -0.02557      0.41691     -0.41080      2.10581     -0.24662
##        ftln         ftwd        gnd_M       bld_AB        bld_B        bld_O
##     1.40383     -0.71519      5.57678      0.54383      0.15639     -0.06370
##       lft_Y        smk_Y        alc_Y
##    -0.96495      0.23101     -0.93218
```

```
Mlm$resample
```

```
##        RMSE  Rsquared      MAE Resample
## 1  3.828806 0.8618895 3.252938   Fold01
## 2  3.370912 0.8753072 2.780300   Fold02
## 3  2.999581 0.8812158 2.423339   Fold03
## 4  3.053545 0.8437521 2.581288   Fold04
## 5  2.392532 0.9113453 1.898744   Fold05
## 6  3.328033 0.9052774 2.586671   Fold06
## 7  3.435689 0.8645778 2.621925   Fold07
## 8  3.676944 0.8251774 2.975894   Fold08
## 9  3.319015 0.8793539 2.600240   Fold09
## 10 2.998727 0.9106540 2.354147   Fold10
```

```
TROUT <- TR %>% dplyr::select(ht)
TSOUT <- TS %>% dplyr::select(ht)
TROUT <- TROUT %>% bind_cols(yhlm=predict(Mlm, newdata=TR))
TSOUT <- TSOUT %>% bind_cols(yhlm=predict(Mlm, newdata=TS))
head(TSOUT)
```

```
##      ht     yhlm
## 1 173.6 174.0110
## 2 150.3 151.3521
## 3 150.3 152.5093
## 4 149.4 155.8523
## 5 152.0 156.9321
## 6 180.3 176.9327
```

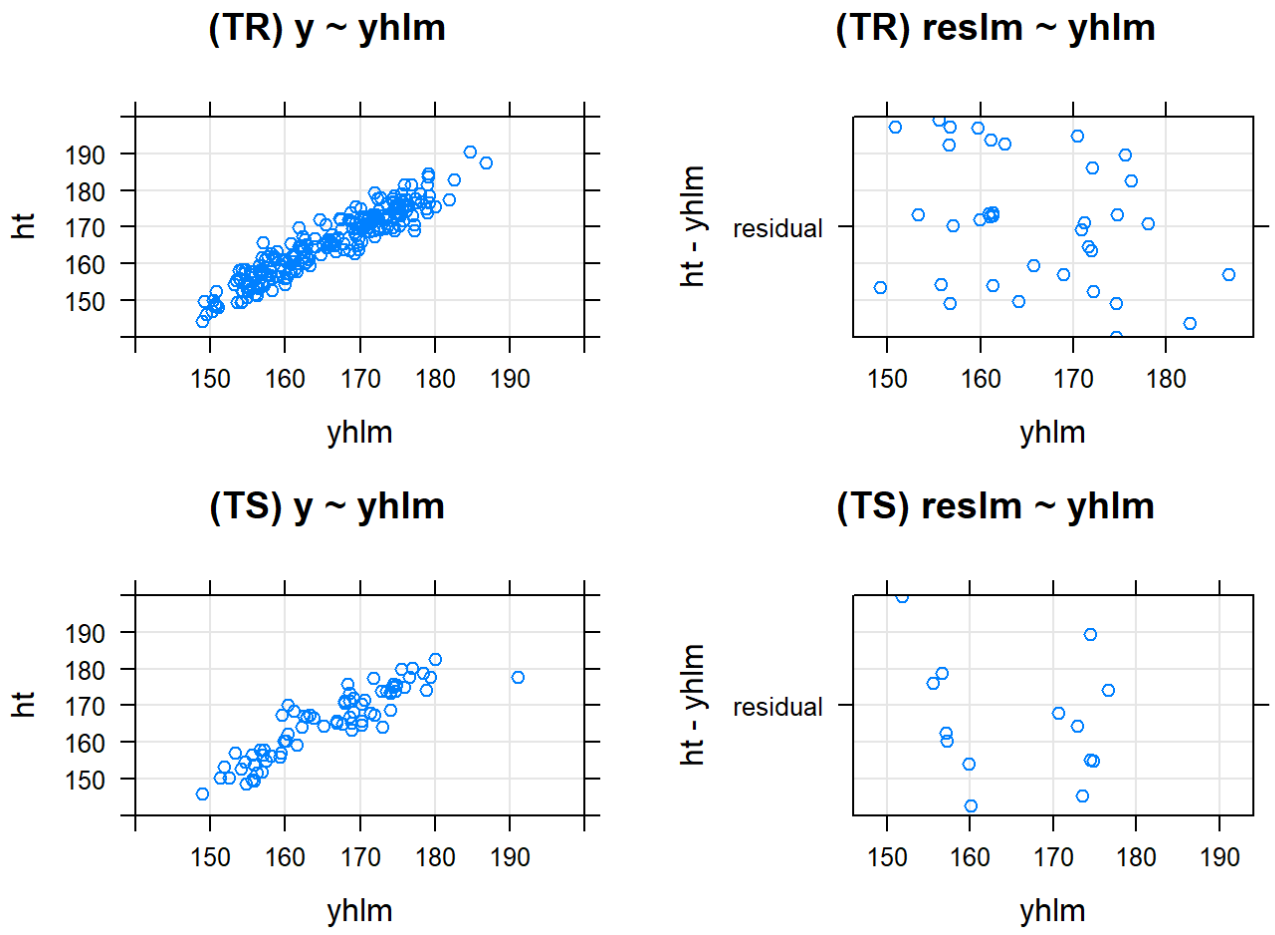
```
#For REG, yardstick::mae, rmse, rsq
foo <- function(y, yh) {
  c(rmse=rmse_vec(y, yh), mae=mae_vec(y, yh), rsq=rsq_vec(y, yh))
}
foo(TSOUT$ht, TSOUT$yhlm)
```

```
##      rmse       mae       rsq
## 3.9868174 3.0618789 0.8092686
```

```
METlm <-
  bind_cols(
    bind_rows(foo(TROUT$ht, TROUT$yhlm), foo(TSOUT$ht, TSOUT$yhlm)),
    data.frame(model='lm', TRTS=c('TR', 'TS')))
METlm
```

```
## # A tibble: 2 x 5
##    rmse   mae   rsq model TRTS
##   <dbl> <dbl> <dbl> <chr> <chr>
## 1  3.07  2.46 0.883 lm    TR
## 2  3.99  3.06 0.809 lm    TS
```

```
g1 <- xyplot(ht~yhlm, data=TROUT,
             type=c('p','g'), xlim=c(140, 200), ylim=c(140,200),
             main='(TR) y ~ yhlm')
g2 <- xyplot(ht-yhlm~yhlm, data=TROUT,
             type=c('p','g'),
             ylim='residual',
             main='(TR) reslm ~ yhlm')
g3 <- xyplot(ht~yhlm, data=TSOUT,
             type=c('p','g'), xlim=c(140, 200), ylim=c(140,200),
             main='(TS) y ~ yhlm')
g4 <- xyplot(ht-yhlm~yhlm, data=TSOUT,
             type=c('p','g'),
             ylim='residual',
             main='(TS) reslm ~ yhlm')
grid.arrange(g1, g2, g3, g4, ncol=2)
```




```
# 자유도로 확인가능
refitCntrl <- trainControl(method='none')
Flm <- train(RC, data=TR,
             method='lm',
             trControl=trainControl(method='none'),
             tuneGrid=Mlm$bestTune)
Flm
```

```
## Linear Regression
##
## 225 samples
##  12 predictor
##
## Recipe steps: impute_median, impute_mode, dummy
## Resampling: None
```

lmStepAIC: AIC 변수선택

- 튜닝모수 없음. intercept는 튜닝 안함
- parsnip에 없음

```
modelLookup('lmStepAIC')
```

```
##       model parameter     label forReg forClass probModel
## 1 lmStepAIC parameter parameter   TRUE    FALSE     FALSE
```

- 적합

```
set.seed(20180968)
Mstep <- train(RC, data=TR,
               method='lmStepAIC',
               trControl=trCntl)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
## Start:  AIC=477.03
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_0 + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - hdwd    1      0.03 1836.0 475.03
## - smk_Y   1      0.15 1836.1 475.05
## - bld_B   1      0.23 1836.2 475.06
## - bld_0   1      1.48 1837.5 475.19
## - ftwd    1      4.87 1840.8 475.57
## - alc_Y   1      5.21 1841.2 475.61
## - age     1      5.38 1841.4 475.63
## - bld_AB  1     12.48 1848.5 476.41
## <none>                1836.0 477.03
## - lft_Y   1     20.27 1856.2 477.26
## - hdln    1    194.39 2030.4 495.46
## - ftln    1    199.81 2035.8 496.00
## - gnd_M   1    415.55 2251.5 516.45
## - wa      1    513.80 2349.8 525.12
## - wt      1    696.52 2532.5 540.32
##
## Step:  AIC=475.03
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_B + bld_0 + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - smk_Y   1      0.14 1836.1 473.05
## - bld_B   1      0.22 1836.2 473.06
## - bld_0   1      1.46 1837.5 473.20
## - alc_Y   1      5.18 1841.2 473.61
## - ftwd    1      5.33 1841.3 473.62
## - age     1      6.85 1842.8 473.79
## - bld_AB  1     12.46 1848.5 474.41
## <none>                1836.0 475.03
## - lft_Y   1     20.61 1856.6 475.30
## - ftln    1    204.91 2040.9 494.51
## - hdln    1    206.02 2042.0 494.62
## - wa      1    513.78 2349.8 523.12
## - gnd_M   1    554.88 2390.9 526.64
## - wt      1    714.52 2550.5 539.76
##
## Step:  AIC=473.05
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_B + bld_0 + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_B   1      0.22 1836.4 471.07
## - bld_0   1      1.43 1837.6 471.21
## - alc_Y   1      5.04 1841.2 471.61
## - ftwd    1      5.47 1841.6 471.65
## - age     1      6.72 1842.9 471.79
## - bld_AB  1     12.68 1848.8 472.45
## <none>                1836.1 473.05
## - lft_Y   1     20.66 1856.8 473.32
## - hdln    1    206.18 2042.3 492.65
## - ftln    1    207.05 2043.2 492.74
## - wa      1    514.53 2350.7 521.20
```

```
## - gnd_M   1    581.99 2418.1 526.94
## - wt      1    717.31 2553.4 538.00
##
## Step:  AIC=471.07
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_0 + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_0   1      2.92 1839.3 469.40
## - alc_Y   1      5.09 1841.5 469.64
## - ftwd    1      5.54 1841.9 469.69
## - age     1      6.56 1842.9 469.80
## - bld_AB  1     14.45 1850.8 470.67
## <none>                1836.4 471.07
## - lft_Y   1     20.49 1856.8 471.33
## - hdln    1    206.13 2042.5 490.67
## - ftln    1    208.13 2044.5 490.87
## - wa      1    538.32 2374.7 521.26
## - gnd_M   1    583.90 2420.3 525.12
## - wt      1    747.29 2583.7 538.38
##
## Step:  AIC=469.4
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - ftwd    1      5.31 1844.6 467.98
## - age     1      5.33 1844.6 467.98
## - alc_Y   1      6.18 1845.5 468.08
## <none>                1839.3 469.40
## - bld_AB  1     18.91 1858.2 469.47
## - lft_Y   1     19.55 1858.8 469.54
## - hdln    1    204.99 2044.3 488.85
## - ftln    1    206.63 2045.9 489.01
## - wa      1    541.93 2381.2 519.82
## - gnd_M   1    582.29 2421.6 523.23
## - wt      1    751.75 2591.0 536.96
##
## Step:  AIC=467.98
## .outcome ~ age + wt + wa + hdln + ftln + gnd_M + bld_AB + lft_Y +
##     alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - age     1      4.57 1849.2 466.48
## - alc_Y   1      6.10 1850.7 466.65
## <none>                1844.6 467.98
## - lft_Y   1     19.47 1864.0 468.11
## - bld_AB  1     20.65 1865.2 468.24
## - hdln    1    200.12 2044.7 486.89
## - ftln    1    201.47 2046.0 487.02
## - wa      1    554.01 2398.6 519.29
## - gnd_M   1    578.48 2423.1 521.35
## - wt      1    747.66 2592.2 535.06
##
## Step:  AIC=466.48
## .outcome ~ wt + wa + hdln + ftln + gnd_M + bld_AB + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
```

```
## - alc_Y   1      6.97 1856.1 465.25
## - lft_Y   1     17.99 1867.1 466.45
## <none>                1849.2 466.48
## - bld_AB  1     20.54 1869.7 466.72
## - hdln    1    195.75 2044.9 484.91
## - ftln    1    200.34 2049.5 485.36
## - gnd_M   1    590.25 2439.4 520.72
## - wa      1    874.86 2724.0 543.12
## - wt      1    994.36 2843.5 551.84
##
## Step:  AIC=465.25
## .outcome ~ wt + wa + hdln + ftln + gnd_M + bld_AB + lft_Y
##
##          Df Sum of Sq    RSS    AIC
## <none>                1856.1 465.25
## - lft_Y   1     19.74 1875.9 465.39
## - bld_AB  1     21.33 1877.5 465.57
## - ftln    1    196.86 2053.0 483.71
## - hdln    1    200.70 2056.8 484.09
## - gnd_M   1    583.67 2439.8 518.75
## - wa      1    895.19 2751.3 543.14
## - wt      1   1013.85 2870.0 551.72
## Start:  AIC=478.1
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_0 + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - smk_Y   1      0.49 1868.5 476.15
## - bld_0   1      0.67 1868.7 476.17
## - hdwd    1      2.92 1871.0 476.41
## - bld_B   1      7.32 1875.4 476.88
## - ftwd    1      9.56 1877.6 477.12
## - lft_Y   1      9.58 1877.6 477.13
## - age     1     10.17 1878.2 477.19
## - alc_Y   1     13.24 1881.3 477.52
## - bld_AB  1     14.28 1882.3 477.63
## <none>                1868.0 478.10
## - ftln    1    132.08 2000.1 489.83
## - hdln    1    268.17 2136.2 503.06
## - wa      1    440.62 2308.7 518.66
## - gnd_M   1    527.26 2395.3 526.07
## - wt      1    578.46 2446.5 530.32
##
## Step:  AIC=476.15
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_0 + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_0   1      0.61 1869.1 474.22
## - hdwd    1      2.90 1871.4 474.46
## - bld_B   1      7.41 1875.9 474.95
## - lft_Y   1      9.68 1878.2 475.19
## - ftwd    1      9.72 1878.2 475.19
## - age     1      9.82 1878.3 475.20
## - alc_Y   1     12.80 1881.3 475.52
## - bld_AB  1     14.85 1883.4 475.74
## <none>                1868.5 476.15
## - ftln    1    133.16 2001.7 487.99
```

```
## - hdln    1    267.75 2136.3 501.07
## - wa      1    446.99 2315.5 517.26
## - gnd_M   1    544.14 2412.7 525.52
## - wt      1    588.14 2456.7 529.15
##
## Step:  AIC=474.22
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - hdwd    1      2.77 1871.9 472.51
## - lft_Y   1      9.45 1878.6 473.23
## - ftwd    1      9.62 1878.8 473.25
## - age     1      9.63 1878.8 473.25
## - bld_B   1     12.83 1882.0 473.59
## - alc_Y   1     13.18 1882.3 473.63
## <none>                1869.1 474.22
## - bld_AB  1     20.68 1889.8 474.43
## - ftln    1    133.07 2002.2 486.04
## - hdln    1    267.60 2136.7 499.11
## - wa      1    448.06 2317.2 515.41
## - gnd_M   1    543.67 2412.8 523.53
## - wt      1    590.62 2459.8 527.41
##
## Step:  AIC=472.51
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_B + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - lft_Y   1     10.42 1882.3 471.63
## - bld_B   1     11.33 1883.2 471.73
## - ftwd    1     12.37 1884.3 471.84
## - alc_Y   1     12.94 1884.8 471.90
## - age     1     17.07 1889.0 472.34
## - bld_AB  1     18.72 1890.6 472.51
## <none>                1871.9 472.51
## - ftln    1    142.77 2014.7 485.29
## - hdln    1    274.06 2146.0 497.98
## - wa      1    450.85 2322.8 513.89
## - wt      1    591.58 2463.5 525.71
## - gnd_M   1    662.03 2533.9 531.38
##
## Step:  AIC=471.63
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_B + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_B   1     10.07 1892.4 470.70
## - ftwd    1     12.53 1894.9 470.96
## - alc_Y   1     14.33 1896.7 471.15
## - age     1     14.50 1896.8 471.17
## - bld_AB  1     16.79 1899.1 471.41
## <none>                1882.3 471.63
## - ftln    1    143.41 2025.7 484.39
## - hdln    1    270.78 2153.1 496.64
## - wa      1    462.17 2344.5 513.76
## - wt      1    607.08 2489.4 525.82
## - gnd_M   1    652.23 2534.6 529.43
```

```
##
## Step:  AIC=470.7
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_AB  1     10.77 1903.2 469.84
## - age     1     10.98 1903.4 469.86
## - ftwd    1     14.07 1906.5 470.19
## - alc_Y   1     16.13 1908.5 470.41
## <none>                1892.4 470.70
## - ftln    1    148.88 2041.3 483.92
## - hdln    1    267.15 2159.6 495.24
## - wa      1    497.03 2389.4 515.58
## - gnd_M   1    643.22 2535.6 527.51
## - wt      1    650.81 2543.2 528.11
##
## Step:  AIC=469.84
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - age    1     11.40 1914.6 469.04
## - ftwd   1     15.41 1918.6 469.46
## - alc_Y  1     16.13 1919.3 469.54
## <none>               1903.2 469.84
## - ftln   1    150.65 2053.8 483.15
## - hdln   1    265.63 2168.8 494.10
## - wa     1    502.78 2406.0 514.96
## - gnd_M  1    645.50 2548.7 526.54
## - wt     1    653.05 2556.2 527.14
##
## Step:  AIC=469.04
## .outcome ~ wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - ftwd   1     16.42 1931.0 468.76
## - alc_Y  1     18.45 1933.0 468.97
## <none>               1914.6 469.04
## - ftln   1    149.51 2064.1 482.16
## - hdln   1    256.28 2170.8 492.29
## - gnd_M  1    664.84 2579.4 526.95
## - wa     1    936.77 2851.3 547.10
## - wt     1    940.17 2854.7 547.34
##
## Step:  AIC=468.76
## .outcome ~ wt + wa + hdln + ftln + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - alc_Y  1     17.95 1948.9 468.62
## <none>               1931.0 468.76
## - ftln   1    135.31 2066.3 480.37
## - hdln   1    243.45 2174.4 490.63
## - gnd_M  1    649.07 2580.1 525.00
## - wt     1    924.97 2856.0 545.43
## - wa     1    947.82 2878.8 547.03
##
## Step:  AIC=468.62
## .outcome ~ wt + wa + hdln + ftln + gnd_M
```

```
##
##         Df Sum of Sq    RSS    AIC
## <none>               1948.9 468.62
## - ftln   1    129.10 2078.0 479.51
## - hdln   1    254.77 2203.7 491.31
## - gnd_M  1    631.65 2580.6 523.05
## - wt     1    965.77 2914.7 547.52
## - wa     1   1000.97 2949.9 549.93
## Start:  AIC=487.99
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_O   1      0.23 1938.0 486.01
## - smk_Y   1      0.25 1938.0 486.01
## - bld_B   1      0.43 1938.2 486.03
## - hdwd    1      1.29 1939.1 486.12
## - bld_AB  1      2.89 1940.7 486.29
## - age     1      6.86 1944.7 486.71
## - lft_Y   1     15.31 1953.1 487.58
## <none>                1937.8 487.99
## - ftwd    1     22.04 1959.8 488.28
## - alc_Y   1     33.89 1971.7 489.51
## - ftln    1    191.68 2129.5 505.14
## - hdln    1    283.63 2221.4 513.72
## - gnd_M   1    449.53 2387.3 528.34
## - wa      1    532.60 2470.4 535.28
## - wt      1    690.33 2628.1 547.85
##
## Step:  AIC=486.01
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - smk_Y   1      0.22 1938.2 484.03
## - bld_B   1      1.10 1939.1 484.13
## - hdwd    1      1.22 1939.2 484.14
## - bld_AB  1      4.24 1942.3 484.46
## - age     1      6.77 1944.8 484.72
## - lft_Y   1     15.29 1953.3 485.61
## <none>                1938.0 486.01
## - ftwd    1     22.09 1960.1 486.31
## - alc_Y   1     34.09 1972.1 487.55
## - ftln    1    191.52 2129.5 503.14
## - hdln    1    283.41 2221.4 511.72
## - gnd_M   1    449.59 2387.6 526.36
## - wa      1    533.37 2471.4 533.36
## - wt      1    691.99 2630.0 545.99
##
## Step:  AIC=484.03
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_B   1      1.06 1939.3 482.14
## - hdwd    1      1.22 1939.5 482.16
## - bld_AB  1      4.34 1942.6 482.49
## - age     1      6.61 1944.8 482.72
```

```
## - lft_Y   1     15.45 1953.7 483.65
## <none>                1938.2 484.03
## - ftwd    1     22.13 1960.4 484.34
## - alc_Y   1     33.96 1972.2 485.56
## - ftln    1    192.64 2130.9 501.27
## - hdln    1    283.63 2221.9 509.76
## - gnd_M   1    468.63 2406.9 525.99
## - wa      1    537.39 2475.6 531.71
## - wt      1    698.39 2636.6 544.50
##
## Step:  AIC=482.14
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - hdwd    1      0.92 1940.2 480.24
## - bld_AB  1      3.46 1942.8 480.51
## - age     1      6.38 1945.7 480.81
## - lft_Y   1     14.90 1954.2 481.70
## <none>                1939.3 482.14
## - ftwd    1     22.53 1961.8 482.49
## - alc_Y   1     34.82 1974.1 483.76
## - ftln    1    195.05 2134.3 499.60
## - hdln    1    282.57 2221.9 507.76
## - gnd_M   1    471.43 2410.7 524.32
## - wa      1    553.81 2493.1 531.14
## - wt      1    711.59 2650.9 543.60
##
## Step:  AIC=480.24
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_AB  1      3.16 1943.4 478.57
## - age     1     11.14 1951.3 479.40
## - lft_Y   1     15.52 1955.7 479.86
## <none>                1940.2 480.24
## - ftwd    1     25.01 1965.2 480.84
## - alc_Y   1     34.41 1974.6 481.81
## - ftln    1    202.98 2143.2 498.44
## - hdln    1    293.79 2234.0 506.86
## - wa      1    553.82 2494.0 529.22
## - gnd_M   1    600.34 2540.6 532.97
## - wt      1    725.65 2665.9 542.74
##
## Step:  AIC=478.57
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + lft_Y +
##     alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - age    1     11.07 1954.4 477.72
## - lft_Y  1     15.21 1958.6 478.15
## <none>               1943.4 478.57
## - ftwd   1     26.10 1969.5 479.28
## - alc_Y  1     34.94 1978.3 480.19
## - ftln   1    203.91 2147.3 496.83
## - hdln   1    292.06 2235.4 504.99
## - wa     1    559.40 2502.8 527.93
```

```
## - gnd_M  1    602.55 2545.9 531.40
## - wt     1    731.63 2675.0 541.44
##
## Step:  AIC=477.72
## .outcome ~ wt + wa + hdln + ftln + ftwd + gnd_M + lft_Y + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - lft_Y  1     12.17 1966.6 476.98
## <none>               1954.4 477.72
## - ftwd   1     24.39 1978.8 478.24
## - alc_Y  1     39.06 1993.5 479.74
## - ftln   1    202.00 2156.4 495.69
## - hdln   1    281.68 2236.1 503.06
## - gnd_M  1    613.66 2568.1 531.16
## - wa     1    947.92 2902.4 556.00
## - wt     1   1007.46 2961.9 560.12
##
## Step:  AIC=476.98
## .outcome ~ wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## <none>               1966.6 476.98
## - ftwd   1     24.68 1991.3 477.52
## - alc_Y  1     39.32 2005.9 479.00
## - ftln   1    205.87 2172.5 495.19
## - hdln   1    275.49 2242.1 501.60
## - gnd_M  1    602.75 2569.4 529.26
## - wa     1    957.52 2924.1 555.51
## - wt     1   1028.11 2994.7 560.35
## Start:  AIC=488.08
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_O   1      0.00 1926.8 486.08
## - bld_B   1      0.24 1927.0 486.10
## - smk_Y   1      1.20 1928.0 486.20
## - hdwd    1      3.75 1930.5 486.47
## - lft_Y   1      5.13 1931.9 486.62
## - bld_AB  1      5.18 1932.0 486.63
## - age     1     13.56 1940.3 487.51
## - alc_Y   1     18.77 1945.5 488.05
## <none>               1926.8 488.08
## - ftwd    1     23.94 1950.7 488.60
## - ftln    1    168.63 2095.4 503.19
## - hdln    1    297.09 2223.8 515.33
## - gnd_M   1    431.83 2358.6 527.33
## - wa      1    552.87 2479.6 537.54
## - wt      1    760.75 2687.5 553.96
##
## Step:  AIC=486.08
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_B   1      0.33 1927.1 484.11
## - smk_Y   1      1.20 1928.0 484.20
## - hdwd    1      3.76 1930.5 484.48
```

```
## - lft_Y   1      5.15 1931.9 484.62
## - bld_AB  1      6.19 1933.0 484.73
## - age     1     13.61 1940.4 485.51
## <none>                1926.8 486.08
## - alc_Y   1     19.03 1945.8 486.08
## - ftwd    1     23.94 1950.7 486.60
## - ftln    1    168.66 2095.4 501.20
## - hdln    1    297.68 2224.4 513.39
## - gnd_M   1    431.84 2358.6 525.33
## - wa      1    554.33 2481.1 535.66
## - wt      1    763.30 2690.1 552.16
##
## Step:  AIC=484.11
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - smk_Y   1      1.12 1928.2 482.23
## - hdwd    1      3.48 1930.6 482.48
## - lft_Y   1      5.03 1932.1 482.64
## - bld_AB  1      5.93 1933.0 482.74
## - age     1     13.41 1940.5 483.53
## <none>                1927.1 484.11
## - alc_Y   1     19.88 1947.0 484.21
## - ftwd    1     24.33 1951.4 484.67
## - ftln    1    170.31 2097.4 499.39
## - hdln    1    297.64 2224.7 511.41
## - gnd_M   1    438.41 2365.5 523.93
## - wa      1    569.62 2496.7 534.94
## - wt      1    777.37 2704.5 551.25
##
## Step:  AIC=482.23
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - hdwd    1      3.27 1931.5 480.58
## - lft_Y   1      5.08 1933.3 480.77
## - bld_AB  1      6.10 1934.3 480.87
## - age     1     13.05 1941.3 481.61
## <none>                1928.2 482.23
## - alc_Y   1     19.15 1947.4 482.25
## - ftwd    1     24.98 1953.2 482.86
## - ftln    1    173.66 2101.9 497.82
## - hdln    1    296.52 2224.7 509.41
## - gnd_M   1    452.78 2381.0 523.26
## - wa      1    569.56 2497.8 533.03
## - wt      1    777.64 2705.8 549.35
##
## Step:  AIC=480.58
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_AB  1      5.50 1937.0 479.16
## - lft_Y   1      5.80 1937.3 479.19
## - alc_Y   1     18.74 1950.2 480.55
## <none>                1931.5 480.58
```

```
## - age     1     24.41 1955.9 481.14
## - ftwd    1     28.76 1960.2 481.59
## - ftln    1    187.74 2119.2 497.50
## - hdln    1    302.76 2234.2 508.28
## - wa      1    568.87 2500.4 531.24
## - gnd_M   1    575.89 2507.4 531.81
## - wt      1    791.58 2723.1 548.64
##
## Step:  AIC=479.16
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + lft_Y +
##     alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - lft_Y  1      5.08 1942.1 477.69
## <none>               1937.0 479.16
## - alc_Y  1     19.96 1956.9 479.25
## - age    1     23.65 1960.6 479.63
## - ftwd   1     30.53 1967.5 480.35
## - ftln   1    187.59 2124.6 496.01
## - hdln   1    306.00 2243.0 507.08
## - gnd_M  1    574.70 2511.7 530.16
## - wa     1    576.56 2513.5 530.31
## - wt     1    798.92 2735.9 547.60
##
## Step:  AIC=477.69
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## <none>               1942.1 477.69
## - alc_Y  1     21.89 1964.0 477.98
## - age    1     21.90 1964.0 477.98
## - ftwd   1     30.50 1972.6 478.87
## - ftln   1    188.68 2130.7 494.61
## - hdln   1    303.79 2245.8 505.34
## - gnd_M  1    569.65 2511.7 528.16
## - wa     1    582.09 2524.2 529.17
## - wt     1    810.02 2752.1 546.81
## Start:  AIC=490.32
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_B   1      0.24 1985.4 488.34
## - hdwd    1      0.55 1985.7 488.37
## - lft_Y   1      0.97 1986.1 488.42
## - smk_Y   1      1.32 1986.5 488.45
## - bld_O   1      1.46 1986.6 488.47
## - bld_AB  1      2.97 1988.1 488.62
## - alc_Y   1     15.22 2000.4 489.85
## - age     1     15.99 2001.2 489.93
## <none>               1985.2 490.32
## - ftwd    1     25.74 2010.9 490.91
## - ftln    1    184.50 2169.7 506.18
## - hdln    1    263.21 2248.4 513.35
## - wa      1    466.81 2452.0 530.77
## - gnd_M   1    471.39 2456.5 531.14
## - wt      1    630.86 2616.0 543.79
##
```

```
## Step:  AIC=488.34
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_0 + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - hdwd    1      0.49 1985.9 486.39
## - lft_Y   1      0.99 1986.4 486.44
## - smk_Y   1      1.32 1986.7 486.48
## - bld_AB  1      2.83 1988.2 486.63
## - bld_0   1      2.85 1988.2 486.63
## - alc_Y   1     15.40 2000.8 487.90
## - age     1     15.87 2001.3 487.94
## <none>                1985.4 488.34
## - ftwd    1     25.79 2011.2 488.94
## - ftln    1    184.86 2170.3 504.24
## - hdln    1    262.97 2248.4 511.35
## - gnd_M   1    477.98 2463.4 529.70
## - wa      1    488.89 2474.3 530.59
## - wt      1    656.88 2642.3 543.79
##
## Step:  AIC=486.39
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_0 + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - lft_Y   1      1.10 1987.0 484.50
## - smk_Y   1      1.35 1987.2 484.53
## - bld_0   1      2.62 1988.5 484.66
## - bld_AB  1      2.80 1988.7 484.68
## - alc_Y   1     15.26 2001.2 485.93
## <none>                1985.9 486.39
## - age     1     21.50 2007.4 486.56
## - ftwd    1     27.50 2013.4 487.16
## - ftln    1    189.04 2174.9 502.67
## - hdln    1    276.39 2262.3 510.58
## - wa      1    488.93 2474.8 528.63
## - gnd_M   1    587.54 2573.4 536.49
## - wt      1    674.81 2660.7 543.19
##
## Step:  AIC=484.5
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_0 + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - smk_Y   1      1.50 1988.5 482.66
## - bld_0   1      2.46 1989.4 482.75
## - bld_AB  1      2.58 1989.6 482.77
## - alc_Y   1     15.31 2002.3 484.05
## <none>                1987.0 484.50
## - age     1     21.00 2008.0 484.62
## - ftwd    1     27.52 2014.5 485.27
## - ftln    1    188.64 2175.6 500.74
## - hdln    1    276.86 2263.8 508.72
## - wa      1    488.99 2476.0 526.73
## - gnd_M   1    588.18 2575.2 534.62
## - wt      1    676.19 2663.2 541.38
##
## Step:  AIC=482.66
```

```
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_O + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_O   1      2.20 1990.7 480.88
## - bld_AB  1      2.86 1991.3 480.94
## - alc_Y   1     14.26 2002.7 482.09
## - age     1     19.80 2008.3 482.65
## <none>                1988.5 482.66
## - ftwd    1     27.72 2016.2 483.44
## - ftln    1    193.95 2182.4 499.36
## - hdln    1    275.38 2263.9 506.73
## - wa      1    494.61 2483.1 525.30
## - gnd_M   1    610.22 2598.7 534.45
## - wt      1    683.26 2671.7 540.02
##
## Step:  AIC=480.88
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_AB  1      4.40 1995.1 479.32
## - alc_Y   1     15.45 2006.1 480.43
## - age     1     18.18 2008.9 480.71
## <none>                1990.7 480.88
## - ftwd    1     27.79 2018.5 481.66
## - ftln    1    193.27 2183.9 497.50
## - hdln    1    273.82 2264.5 504.78
## - wa      1    497.99 2488.7 523.76
## - gnd_M   1    610.67 2601.4 532.66
## - wt      1    687.81 2678.5 538.53
##
## Step:  AIC=479.32
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - alc_Y  1     15.35 2010.4 478.86
## - age    1     18.10 2013.2 479.14
## <none>               1995.1 479.32
## - ftwd   1     30.19 2025.3 480.34
## - ftln   1    194.25 2189.3 496.00
## - hdln   1    276.51 2271.6 503.41
## - wa     1    501.98 2497.1 522.43
## - gnd_M  1    610.25 2605.3 530.96
## - wt     1    692.04 2687.1 537.18
##
## Step:  AIC=478.86
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M
##
##         Df Sum of Sq    RSS    AIC
## <none>               2010.4 478.86
## - age    1     20.84 2031.3 478.94
## - ftwd   1     29.13 2039.5 479.75
## - ftln   1    188.05 2198.5 494.84
## - hdln   1    290.07 2300.5 503.95
## - wa     1    505.86 2516.3 521.97
## - gnd_M  1    595.52 2605.9 529.01
## - wt     1    695.48 2705.9 536.58
```

```
## Start:  AIC=483.46
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - hdwd    1      0.00 1895.0 481.46
## - bld_B   1      0.00 1895.0 481.46
## - bld_O   1      0.87 1895.9 481.55
## - bld_AB  1      2.26 1897.3 481.70
## - lft_Y   1      8.56 1903.6 482.37
## - smk_Y   1      9.54 1904.6 482.48
## <none>                1895.0 483.46
## - alc_Y   1     20.45 1915.5 483.64
## - age     1     21.21 1916.2 483.72
## - ftwd    1     30.96 1926.0 484.75
## - ftln    1    160.51 2055.6 497.96
## - hdln    1    280.38 2175.4 509.47
## - gnd_M   1    478.55 2373.6 527.17
## - wa      1    487.39 2382.4 527.92
## - wt      1    650.21 2545.2 541.34
##
## Step:  AIC=481.46
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_B   1      0.00 1895.0 479.46
## - bld_O   1      0.87 1895.9 479.55
## - bld_AB  1      2.28 1897.3 479.70
## - lft_Y   1      8.57 1903.6 480.37
## - smk_Y   1      9.54 1904.6 480.48
## <none>                1895.0 481.46
## - alc_Y   1     20.46 1915.5 481.64
## - age     1     25.46 1920.5 482.17
## - ftwd    1     31.85 1926.9 482.84
## - ftln    1    162.29 2057.3 496.14
## - hdln    1    297.90 2192.9 509.10
## - wa      1    488.58 2383.6 526.02
## - gnd_M   1    637.25 2532.3 538.31
## - wt      1    659.83 2554.9 540.11
##
## Step:  AIC=479.46
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_O   1      1.11 1896.2 477.58
## - bld_AB  1      3.01 1898.0 477.78
## - lft_Y   1      8.57 1903.6 478.37
## - smk_Y   1      9.56 1904.6 478.48
## <none>                1895.0 479.46
## - alc_Y   1     20.47 1915.5 479.64
## - age     1     25.67 1920.7 480.19
## - ftwd    1     31.85 1926.9 480.84
## - ftln    1    162.48 2057.5 494.16
## - hdln    1    297.98 2193.0 507.11
## - wa      1    496.73 2391.8 524.72
## - gnd_M   1    642.22 2537.3 536.70
```

```
## - wt      1    671.27 2566.3 539.01
##
## Step:  AIC=477.58
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_AB  1      4.28 1900.4 476.04
## - lft_Y   1      8.23 1904.4 476.46
## - smk_Y   1      9.03 1905.2 476.54
## <none>                1896.2 477.58
## - alc_Y   1     20.99 1917.1 477.81
## - age     1     24.72 1920.9 478.21
## - ftwd    1     32.10 1928.2 478.99
## - ftln    1    162.46 2058.6 492.27
## - hdln    1    297.43 2193.6 505.16
## - wa      1    496.18 2392.3 522.76
## - gnd_M   1    641.33 2537.5 534.72
## - wt      1    670.51 2566.7 537.04
##
## Step:  AIC=476.04
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + lft_Y +
##     smk_Y + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - lft_Y  1      7.38 1907.8 474.82
## - smk_Y  1     10.07 1910.5 475.11
## <none>               1900.4 476.04
## - alc_Y  1     21.32 1921.8 476.30
## - age    1     25.05 1925.5 476.69
## - ftwd   1     32.94 1933.4 477.52
## - ftln   1    160.68 2061.1 490.51
## - hdln   1    299.30 2199.7 503.73
## - wa     1    504.76 2405.2 521.85
## - gnd_M  1    641.40 2541.8 533.07
## - wt     1    679.92 2580.3 536.12
##
## Step:  AIC=474.82
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + smk_Y +
##     alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - smk_Y  1     10.12 1917.9 473.90
## <none>               1907.8 474.82
## - age    1     22.91 1930.7 475.25
## - alc_Y  1     23.11 1930.9 475.27
## - ftwd   1     34.09 1941.9 476.42
## - ftln   1    159.41 2067.2 489.11
## - hdln   1    298.35 2206.2 502.32
## - wa     1    515.39 2423.2 521.37
## - gnd_M  1    634.59 2542.4 531.11
## - wt     1    701.02 2608.8 536.35
##
## Step:  AIC=473.9
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - alc_Y  1     18.66 1936.6 473.86
```

```
## <none>               1917.9 473.90
## - age    1     20.55 1938.5 474.06
## - ftwd   1     36.38 1954.3 475.71
## - ftln   1    167.00 2084.9 488.85
## - hdln   1    293.52 2211.4 500.80
## - wa     1    517.77 2435.7 520.41
## - gnd_M  1    692.53 2610.5 534.48
## - wt     1    706.54 2624.5 535.56
##
## Step:  AIC=473.86
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M
##
##         Df Sum of Sq    RSS    AIC
## <none>               1936.6 473.86
## - age    1     23.81 1960.4 474.34
## - ftwd   1     35.64 1972.2 475.56
## - ftln   1    160.84 2097.4 488.06
## - hdln   1    303.98 2240.6 501.46
## - wa     1    518.59 2455.2 520.03
## - gnd_M  1    674.62 2611.2 532.54
## - wt     1    709.44 2646.0 535.23
## Start:  AIC=481.77
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_B   1      0.04 1879.3 479.77
## - bld_O   1      0.34 1879.6 479.80
## - hdwd    1      3.21 1882.5 480.11
## - lft_Y   1      4.91 1884.2 480.30
## - smk_Y   1      5.30 1884.6 480.34
## - bld_AB  1      6.43 1885.7 480.46
## - age     1      7.81 1887.1 480.61
## <none>               1879.3 481.77
## - alc_Y   1     22.22 1901.5 482.15
## - ftwd    1     42.16 1921.5 484.27
## - ftln    1    216.51 2095.8 501.90
## - hdln    1    231.11 2110.4 503.31
## - gnd_M   1    469.22 2348.5 525.01
## - wa      1    548.75 2428.1 531.77
## - wt      1    731.23 2610.5 546.48
##
## Step:  AIC=479.77
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_O   1      0.32 1879.7 477.80
## - hdwd    1      3.27 1882.6 478.12
## - lft_Y   1      4.93 1884.3 478.30
## - smk_Y   1      5.26 1884.6 478.34
## - age     1      7.88 1887.2 478.62
## - bld_AB  1      9.11 1888.5 478.75
## <none>               1879.3 479.77
## - alc_Y   1     22.18 1901.5 480.15
## - ftwd    1     42.20 1921.5 482.28
## - ftln    1    216.89 2096.2 499.94
## - hdln    1    233.16 2112.5 501.51
```

```
## - gnd_M   1    479.96 2359.3 523.94
## - wa      1    565.95 2445.3 531.21
## - wt      1    752.41 2631.8 546.13
##
## Step:  AIC=477.8
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - hdwd    1      3.09 1882.7 476.14
## - lft_Y   1      4.78 1884.4 476.32
## - smk_Y   1      5.08 1884.7 476.35
## - age     1      7.68 1887.3 476.63
## - bld_AB  1     10.64 1890.3 476.95
## <none>                1879.7 477.80
## - alc_Y   1     22.70 1902.3 478.24
## - ftwd    1     42.39 1922.0 480.33
## - ftln    1    216.90 2096.6 497.97
## - hdln    1    232.91 2112.6 499.52
## - gnd_M   1    480.65 2360.3 522.03
## - wa      1    566.04 2445.7 529.24
## - wt      1    752.72 2632.4 544.17
##
## Step:  AIC=476.14
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - smk_Y   1      5.15 1887.9 474.69
## - lft_Y   1      5.49 1888.2 474.73
## - bld_AB  1      9.93 1892.7 475.21
## - age     1     15.43 1898.2 475.79
## <none>                1882.7 476.14
## - alc_Y   1     22.74 1905.5 476.58
## - ftwd    1     46.14 1928.9 479.05
## - ftln    1    224.44 2107.2 497.00
## - hdln    1    232.93 2115.7 497.82
## - wa      1    565.40 2448.1 527.45
## - gnd_M   1    613.10 2495.8 531.36
## - wt      1    768.61 2651.3 543.63
##
## Step:  AIC=474.69
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - lft_Y   1      5.65 1893.5 473.30
## - bld_AB  1     10.86 1898.8 473.86
## - age     1     14.94 1902.8 474.29
## <none>                1887.9 474.69
## - alc_Y   1     20.45 1908.3 474.88
## - ftwd    1     47.75 1935.6 477.76
## - ftln    1    228.82 2116.7 495.92
## - hdln    1    230.87 2118.8 496.11
## - wa      1    562.94 2450.8 525.67
## - gnd_M   1    663.23 2551.1 533.81
## - wt      1    770.49 2658.4 542.17
##
```

```
## Step:  AIC=473.3
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_AB  1     10.11 1903.7 472.38
## - age     1     13.36 1906.9 472.73
## <none>                1893.5 473.30
## - alc_Y   1     22.12 1915.7 473.66
## - ftwd    1     47.51 1941.0 476.33
## - ftln    1    229.11 2122.7 494.48
## - hdln    1    229.21 2122.8 494.49
## - wa      1    570.49 2464.0 524.76
## - gnd_M   1    657.66 2551.2 531.82
## - wt      1    782.82 2676.4 541.54
##
## Step:  AIC=472.38
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - age    1     13.76 1917.4 471.84
## <none>                1903.7 472.38
## - alc_Y  1     21.65 1925.3 472.68
## - ftwd   1     49.59 1953.2 475.60
## - ftln   1    228.46 2132.1 493.39
## - hdln   1    231.28 2134.9 493.66
## - wa     1    577.85 2481.5 524.19
## - gnd_M  1    658.97 2562.6 530.72
## - wt     1    785.88 2689.5 540.54
##
## Step:  AIC=471.84
## .outcome ~ wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## <none>                1917.4 471.84
## - alc_Y  1     24.57 1942.0 472.43
## - ftwd   1     47.17 1964.6 474.78
## - hdln   1    220.75 2138.2 491.96
## - ftln   1    226.85 2144.3 492.54
## - gnd_M  1    666.83 2584.2 530.43
## - wa     1   1051.91 2969.3 558.63
## - wt     1   1121.93 3039.3 563.36
## Start:  AIC=474.84
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - hdwd    1      0.78 1827.8 472.92
## - bld_O   1      1.29 1828.3 472.98
## - smk_Y   1      1.52 1828.5 473.00
## - lft_Y   1      1.55 1828.6 473.01
## - bld_AB  1      2.27 1829.3 473.09
## - bld_B   1      4.35 1831.3 473.32
## - age     1      7.87 1834.9 473.71
## - alc_Y   1     14.17 1841.2 474.40
## - ftwd    1     14.32 1841.3 474.41
## <none>                1827.0 474.84
## - ftln    1    165.53 1992.5 490.36
```

```
## - hdln    1    272.69 2099.7 500.94
## - gnd_M   1    471.88 2298.9 519.25
## - wa      1    543.41 2370.4 525.44
## - wt      1    618.66 2445.7 531.75
##
## Step:  AIC=472.92
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_O   1      1.46 1829.2 471.08
## - smk_Y   1      1.50 1829.3 471.09
## - lft_Y   1      1.69 1829.5 471.11
## - bld_AB  1      2.10 1829.9 471.16
## - bld_B   1      4.21 1832.0 471.39
## - age     1     12.08 1839.9 472.25
## - alc_Y   1     14.12 1841.9 472.48
## - ftwd    1     15.66 1843.4 472.65
## <none>                1827.8 472.92
## - ftln    1    172.52 2000.3 489.14
## - hdln    1    287.14 2114.9 500.40
## - wa      1    544.13 2371.9 523.56
## - gnd_M   1    612.10 2439.9 529.27
## - wt      1    626.96 2454.7 530.50
##
## Step:  AIC=471.08
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_B + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_AB  1      1.07 1830.3 469.20
## - smk_Y   1      1.73 1831.0 469.28
## - lft_Y   1      1.88 1831.1 469.29
## - bld_B   1      2.78 1832.0 469.39
## - age     1     12.97 1842.2 470.51
## - alc_Y   1     13.93 1843.2 470.62
## - ftwd    1     15.73 1845.0 470.81
## <none>                1829.2 471.08
## - ftln    1    175.12 2004.4 487.55
## - hdln    1    286.85 2116.1 498.51
## - wa      1    548.39 2377.6 522.05
## - gnd_M   1    611.35 2440.6 527.33
## - wt      1    627.82 2457.1 528.69
##
## Step:  AIC=469.2
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_B +
##     lft_Y + smk_Y + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - lft_Y  1      1.69 1832.0 467.39
## - smk_Y  1      1.82 1832.1 467.40
## - bld_B  1      2.01 1832.3 467.42
## - age    1     12.89 1843.2 468.62
## - alc_Y  1     14.07 1844.4 468.75
## - ftwd   1     15.98 1846.3 468.96
## <none>               1830.3 469.20
## - ftln   1    174.46 2004.8 485.59
## - hdln   1    289.44 2119.8 496.86
```

```
## - wa     1    553.94 2384.2 520.61
## - gnd_M  1    610.36 2440.7 525.34
## - wt     1    631.12 2461.4 527.05
##
## Step:  AIC=467.39
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_B +
##     smk_Y + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - smk_Y  1      1.84 1833.8 465.59
## - bld_B  1      2.00 1834.0 465.61
## - age    1     12.07 1844.1 466.72
## - alc_Y  1     14.98 1847.0 467.03
## - ftwd   1     15.50 1847.5 467.09
## <none>               1832.0 467.39
## - ftln   1    174.56 2006.6 483.77
## - hdln   1    287.85 2119.8 494.87
## - wa     1    560.33 2392.3 519.30
## - gnd_M  1    609.37 2441.4 523.39
## - wt     1    640.08 2472.1 525.92
##
## Step:  AIC=465.59
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_B +
##     alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - bld_B  1      1.91 1835.8 463.80
## - age    1     11.20 1845.0 464.82
## - alc_Y  1     13.75 1847.6 465.10
## - ftwd   1     16.07 1849.9 465.35
## <none>               1833.8 465.59
## - ftln   1    177.61 2011.5 482.27
## - hdln   1    286.21 2120.1 492.89
## - wa     1    563.73 2397.6 517.74
## - wt     1    647.60 2481.4 524.68
## - gnd_M  1    651.53 2485.4 525.00
##
## Step:  AIC=463.8
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - age    1     10.44 1846.2 462.95
## - alc_Y  1     14.76 1850.5 463.42
## - ftwd   1     16.29 1852.0 463.59
## <none>               1835.8 463.80
## - ftln   1    178.82 2014.6 480.58
## - hdln   1    286.07 2121.8 491.06
## - wa     1    577.85 2413.6 517.08
## - gnd_M  1    653.79 2489.5 523.34
## - wt     1    667.48 2503.2 524.45
##
## Step:  AIC=462.95
## .outcome ~ wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - ftwd   1     15.27 1861.5 462.61
## - alc_Y  1     16.71 1862.9 462.77
## <none>               1846.2 462.95
```

```
## - ftln   1    177.49 2023.7 479.49
## - hdln   1    276.43 2122.6 489.13
## - gnd_M  1    672.02 2518.2 523.65
## - wt     1    916.55 2762.7 542.37
## - wa     1    951.50 2797.7 544.91
##
## Step:  AIC=462.61
## .outcome ~ wt + wa + hdln + ftln + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - alc_Y  1     16.08 1877.5 462.35
## <none>               1861.5 462.61
## - ftln   1    162.51 2024.0 477.52
## - hdln   1    264.05 2125.5 487.41
## - gnd_M  1    658.32 2519.8 521.78
## - wt     1    906.92 2768.4 540.79
## - wa     1    975.70 2837.2 545.74
##
## Step:  AIC=462.35
## .outcome ~ wt + wa + hdln + ftln + gnd_M
##
##         Df Sum of Sq    RSS    AIC
## <none>               1877.5 462.35
## - ftln   1    157.74 2035.3 476.64
## - hdln   1    268.83 2146.4 487.38
## - gnd_M  1    642.82 2520.4 519.83
## - wt     1    947.74 2825.3 542.90
## - wa     1   1035.68 2913.2 549.09
## Start:  AIC=482.73
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_B   1      0.04 1888.3 480.73
## - smk_Y   1      0.29 1888.5 480.76
## - hdwd    1      0.84 1889.1 480.82
## - bld_AB  1      1.22 1889.4 480.86
## - bld_O   1      1.31 1889.5 480.87
## - lft_Y   1      3.75 1892.0 481.13
## - age     1     16.15 1904.4 482.46
## <none>               1888.2 482.73
## - alc_Y   1     21.75 1910.0 483.05
## - ftwd    1     30.09 1918.3 483.94
## - ftln    1    155.70 2043.9 496.81
## - hdln    1    310.47 2198.7 511.63
## - gnd_M   1    457.45 2345.7 524.77
## - wa      1    502.96 2391.2 528.67
## - wt      1    691.36 2579.6 544.06
##
## Step:  AIC=480.73
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - smk_Y   1      0.29 1888.5 478.76
## - hdwd    1      0.89 1889.2 478.83
## - bld_O   1      1.54 1889.8 478.90
## - bld_AB  1      1.87 1890.1 478.93
```

```
## - lft_Y   1      3.74 1892.0 479.13
## - age     1     16.19 1904.5 480.47
## <none>                1888.3 480.73
## - alc_Y   1     21.72 1910.0 481.05
## - ftwd    1     30.08 1918.3 481.94
## - ftln    1    156.13 2044.4 494.86
## - hdln    1    310.76 2199.0 509.66
## - gnd_M   1    465.82 2354.1 523.49
## - wa      1    514.18 2402.4 527.62
## - wt      1    703.54 2591.8 543.02
##
## Step:  AIC=478.76
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_0 + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - hdwd    1      0.89 1889.4 476.86
## - bld_0   1      1.49 1890.0 476.92
## - bld_AB  1      1.99 1890.5 476.98
## - lft_Y   1      3.74 1892.3 477.16
## - age     1     16.04 1904.6 478.48
## <none>                1888.5 478.76
## - alc_Y   1     21.43 1910.0 479.05
## - ftwd    1     30.25 1918.8 479.99
## - ftln    1    158.57 2047.1 493.13
## - hdln    1    310.84 2199.4 507.69
## - gnd_M   1    482.98 2371.5 522.99
## - wa      1    513.90 2402.4 525.62
## - wt      1    703.26 2591.8 541.02
##
## Step:  AIC=476.86
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_0 + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_0   1      1.19 1890.6 474.99
## - bld_AB  1      1.84 1891.3 475.06
## - lft_Y   1      4.14 1893.6 475.30
## <none>                1889.4 476.86
## - alc_Y   1     20.92 1910.4 477.09
## - age     1     22.89 1912.3 477.30
## - ftwd    1     32.98 1922.4 478.37
## - ftln    1    165.18 2054.6 491.87
## - hdln    1    324.11 2213.6 507.00
## - wa      1    514.69 2404.1 523.76
## - gnd_M   1    629.38 2518.8 533.22
## - wt      1    713.61 2603.1 539.90
##
## Step:  AIC=474.99
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_AB  1      2.77 1893.4 473.28
## - lft_Y   1      3.83 1894.5 473.40
## <none>                1890.6 474.99
## - age     1     21.82 1912.5 475.32
## - alc_Y   1     21.97 1912.6 475.33
```

```
## - ftwd    1     32.86 1923.5 476.48
## - ftln    1    164.75 2055.4 489.95
## - hdln    1    322.96 2213.6 505.00
## - wa      1    516.70 2407.3 522.03
## - gnd_M   1    629.15 2519.8 531.30
## - wt      1    717.51 2608.1 538.30
##
## Step:  AIC=473.28
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + lft_Y +
##     alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - lft_Y  1      3.86 1897.3 471.70
## <none>               1893.4 473.28
## - age    1     21.19 1914.6 473.54
## - alc_Y  1     21.50 1914.9 473.58
## - ftwd   1     34.67 1928.1 474.97
## - ftln   1    163.94 2057.3 488.14
## - hdln   1    325.55 2219.0 503.49
## - wa     1    547.62 2441.0 522.85
## - gnd_M  1    627.10 2520.5 529.36
## - wt     1    745.11 2638.5 538.65
##
## Step:  AIC=471.7
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## <none>               1897.3 471.70
## - age    1     20.30 1917.6 471.86
## - alc_Y  1     22.97 1920.2 472.14
## - ftwd   1     34.02 1931.3 473.31
## - ftln   1    164.60 2061.9 486.59
## - hdln   1    323.65 2220.9 501.67
## - wa     1    547.32 2444.6 521.15
## - gnd_M  1    623.27 2520.5 527.36
## - wt     1    745.49 2642.8 536.97
## Start:  AIC=485.55
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - hdwd    1      0.77 1927.3 483.63
## - smk_Y   1      0.97 1927.5 483.65
## - lft_Y   1      3.02 1929.6 483.87
## - bld_O   1      3.48 1930.0 483.92
## - bld_B   1      3.63 1930.2 483.93
## - bld_AB  1      9.43 1936.0 484.54
## - age     1      9.46 1936.0 484.54
## <none>               1926.5 485.55
## - ftwd    1     26.02 1952.6 486.26
## - alc_Y   1     26.86 1953.4 486.35
## - ftln    1    179.16 2105.7 501.51
## - hdln    1    206.38 2132.9 504.11
## - wa      1    479.29 2405.8 528.43
## - gnd_M   1    488.37 2414.9 529.19
## - wt      1    617.37 2543.9 539.70
##
## Step:  AIC=483.63
```

```
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - smk_Y   1      1.01 1928.3 481.74
## - lft_Y   1      3.25 1930.5 481.97
## - bld_B   1      3.31 1930.6 481.98
## - bld_O   1      3.66 1931.0 482.02
## - bld_AB  1      9.00 1936.3 482.57
## - age     1     14.40 1941.7 483.14
## <none>                1927.3 483.63
## - alc_Y   1     26.71 1954.0 484.41
## - ftwd    1     28.05 1955.4 484.55
## - ftln    1    185.51 2112.8 500.20
## - hdln    1    215.77 2143.1 503.07
## - wa      1    479.41 2406.7 526.51
## - gnd_M   1    616.15 2543.5 537.67
## - wt      1    631.01 2558.3 538.84
##
## Step:  AIC=481.74
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_B + bld_O + lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - lft_Y   1      3.29 1931.6 480.08
## - bld_B   1      3.41 1931.7 480.10
## - bld_O   1      3.90 1932.2 480.15
## - bld_AB  1      9.46 1937.8 480.73
## - age     1     13.99 1942.3 481.20
## <none>                1928.3 481.74
## - alc_Y   1     25.71 1954.0 482.41
## - ftwd    1     28.14 1956.5 482.67
## - ftln    1    187.05 2115.4 498.44
## - hdln    1    214.85 2143.2 501.08
## - wa      1    479.85 2408.2 524.63
## - wt      1    633.00 2561.3 537.08
## - gnd_M   1    646.03 2574.3 538.11
##
## Step:  AIC=480.08
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_B + bld_O + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_B   1      3.27 1934.9 478.43
## - bld_O   1      4.75 1936.4 478.58
## - bld_AB  1      8.99 1940.6 479.02
## - age     1     12.89 1944.5 479.43
## <none>                1931.6 480.08
## - alc_Y   1     27.28 1958.9 480.92
## - ftwd    1     27.99 1959.6 480.99
## - ftln    1    188.55 2120.2 496.90
## - hdln    1    212.92 2144.5 499.21
## - wa      1    483.59 2415.2 523.22
## - wt      1    639.67 2571.3 535.87
## - gnd_M   1    645.05 2576.7 536.29
##
## Step:  AIC=478.43
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
```

```
##     bld_0 + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_0   1      2.13 1937.0 476.65
## - bld_AB  1      5.96 1940.8 477.05
## - age     1     12.19 1947.1 477.69
## <none>                1934.9 478.43
## - alc_Y   1     27.39 1962.3 479.27
## - ftwd    1     27.98 1962.8 479.32
## - ftln    1    190.50 2125.4 495.39
## - hdln    1    211.34 2146.2 497.36
## - wa      1    508.21 2443.1 523.54
## - gnd_M   1    641.78 2576.7 534.29
## - wt      1    667.66 2602.5 536.31
##
## Step:  AIC=476.65
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_AB  1      4.63 1941.6 475.13
## - age     1     14.17 1951.2 476.12
## <none>                1937.0 476.65
## - alc_Y   1     26.33 1963.3 477.37
## - ftwd    1     27.67 1964.7 477.51
## - ftln    1    189.41 2126.4 493.49
## - hdln    1    214.02 2151.0 495.82
## - wa      1    506.87 2443.9 521.60
## - gnd_M   1    644.36 2581.4 532.66
## - wt      1    666.73 2603.7 534.40
##
## Step:  AIC=475.13
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - age    1     14.25 1955.9 474.61
## <none>               1941.6 475.13
## - alc_Y  1     26.29 1967.9 475.85
## - ftwd   1     28.54 1970.2 476.08
## - ftln   1    189.98 2131.6 491.99
## - hdln   1    213.04 2154.7 494.16
## - wa     1    512.78 2454.4 520.47
## - gnd_M  1    643.96 2585.6 530.99
## - wt     1    672.45 2614.1 533.20
##
## Step:  AIC=474.61
## .outcome ~ wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## <none>               1955.9 474.61
## - ftwd   1     26.43 1982.3 475.32
## - alc_Y  1     31.21 1987.1 475.80
## - ftln   1    183.78 2139.7 490.75
## - hdln   1    206.11 2162.0 492.84
## - gnd_M  1    674.60 2630.5 532.46
## - wa     1    902.34 2858.2 549.24
## - wt     1    932.13 2888.0 551.33
## Start:  AIC=534.91
```

```
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + bld_O + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_O   1      0.11 2122.1 532.92
## - bld_B   1      0.79 2122.8 532.99
## - hdwd    1      1.12 2123.1 533.02
## - smk_Y   1      1.64 2123.7 533.08
## - bld_AB  1      5.87 2127.9 533.53
## - lft_Y   1      6.86 2128.9 533.63
## - age     1     12.44 2134.5 534.22
## <none>                2122.0 534.91
## - alc_Y   1     20.89 2142.9 535.11
## - ftwd    1     23.79 2145.8 535.41
## - ftln    1    195.38 2317.4 552.72
## - hdln    1    290.88 2412.9 561.81
## - gnd_M   1    520.10 2642.1 582.23
## - wa      1    567.89 2689.9 586.26
## - wt      1    745.43 2867.5 600.64
##
## Step:  AIC=532.92
## .outcome ~ age + wt + wa + hdln + hdwd + ftln + ftwd + gnd_M +
##     bld_AB + bld_B + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - hdwd    1      1.08 2123.2 531.03
## - bld_B   1      1.52 2123.7 531.08
## - smk_Y   1      1.59 2123.7 531.09
## - lft_Y   1      6.79 2128.9 531.64
## - bld_AB  1      7.94 2130.1 531.76
## - age     1     12.35 2134.5 532.22
## <none>                2122.1 532.92
## - alc_Y   1     21.10 2143.2 533.14
## - ftwd    1     23.80 2145.9 533.43
## - ftln    1    195.28 2317.4 550.72
## - hdln    1    290.81 2412.9 559.81
## - gnd_M   1    520.07 2642.2 580.24
## - wa      1    569.32 2691.4 584.39
## - wt      1    747.72 2869.8 598.83
##
## Step:  AIC=531.03
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     bld_B + lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_B   1      1.21 2124.4 529.16
## - smk_Y   1      1.58 2124.8 529.20
## - lft_Y   1      7.24 2130.5 529.80
## - bld_AB  1      7.36 2130.6 529.81
## - age     1     18.86 2142.1 531.02
## <none>                2123.2 531.03
## - alc_Y   1     20.84 2144.1 531.23
## - ftwd    1     26.35 2149.6 531.81
## - ftln    1    203.29 2326.5 549.60
## - hdln    1    302.38 2425.6 558.99
## - wa      1    570.60 2693.8 582.59
## - gnd_M   1    665.55 2788.8 590.38
## - wt      1    758.41 2881.6 597.75
```

```
##
## Step:  AIC=529.16
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     lft_Y + smk_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - smk_Y   1      1.53 2125.9 527.32
## - bld_AB  1      6.23 2130.7 527.82
## - lft_Y   1      6.94 2131.4 527.89
## - age     1     17.88 2142.3 529.05
## <none>                2124.4 529.16
## - alc_Y   1     21.71 2146.1 529.45
## - ftwd    1     26.57 2151.0 529.96
## - ftln    1    204.56 2329.0 547.84
## - hdln    1    301.52 2425.9 557.02
## - wa      1    591.44 2715.9 582.42
## - gnd_M   1    665.53 2790.0 588.48
## - wt      1    786.34 2910.8 598.02
##
## Step:  AIC=527.32
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + bld_AB +
##     lft_Y + alc_Y
##
##          Df Sum of Sq    RSS    AIC
## - bld_AB  1      6.62 2132.6 526.02
## - lft_Y   1      7.08 2133.0 526.07
## - age     1     17.14 2143.1 527.13
## <none>                2125.9 527.32
## - alc_Y   1     20.49 2146.4 527.48
## - ftwd    1     27.06 2153.0 528.17
## - ftln    1    207.95 2333.9 546.32
## - hdln    1    300.07 2426.0 555.03
## - wa      1    593.35 2719.3 580.71
## - gnd_M   1    703.65 2829.6 589.65
## - wt      1    790.65 2916.6 596.47
##
## Step:  AIC=526.02
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + lft_Y +
##     alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - lft_Y  1      6.35 2138.9 524.69
## - age    1     17.06 2149.6 525.81
## <none>               2132.6 526.02
## - alc_Y  1     20.61 2153.2 526.18
## - ftwd   1     28.64 2161.2 527.02
## - ftln   1    207.47 2340.0 544.91
## - hdln   1    301.60 2434.2 553.78
## - wa     1    603.46 2736.0 580.09
## - gnd_M  1    703.94 2836.5 588.20
## - wt     1    799.00 2931.6 595.62
##
## Step:  AIC=524.69
## .outcome ~ age + wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## - age    1     15.41 2154.3 524.30
## <none>               2138.9 524.69
```

```
## - alc_Y  1     22.03 2160.9 525.00
## - ftwd   1     28.40 2167.3 525.66
## - ftln   1    207.97 2346.9 543.57
## - hdln   1    299.27 2438.2 552.15
## - wa     1    610.20 2749.1 579.16
## - gnd_M  1    697.60 2836.5 586.20
## - wt     1    810.33 2949.2 594.97
##
## Step:  AIC=524.3
## .outcome ~ wt + wa + hdln + ftln + ftwd + gnd_M + alc_Y
##
##         Df Sum of Sq    RSS    AIC
## <none>               2154.3 524.30
## - alc_Y  1     25.39 2179.7 524.94
## - ftwd   1     26.76 2181.1 525.08
## - ftln   1    205.29 2359.6 542.78
## - hdln   1    287.66 2442.0 550.50
## - gnd_M  1    716.45 2870.8 586.90
## - wa     1   1068.61 3222.9 612.94
## - wt     1   1135.52 3289.8 617.56
```

```
#metric='RMSE.Rsquared' 회귀 모형 선택 기준

Mstep
```

```
## Linear Regression with Stepwise Selection
##
## 225 samples
##  12 predictor
##
## Recipe steps: impute_median, impute_mode, dummy
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 203, 201, 203, 204, 201, 203, ...
## Resampling results:
##
##   RMSE      Rsquared   MAE
##   3.263257  0.8739242  2.629523
```

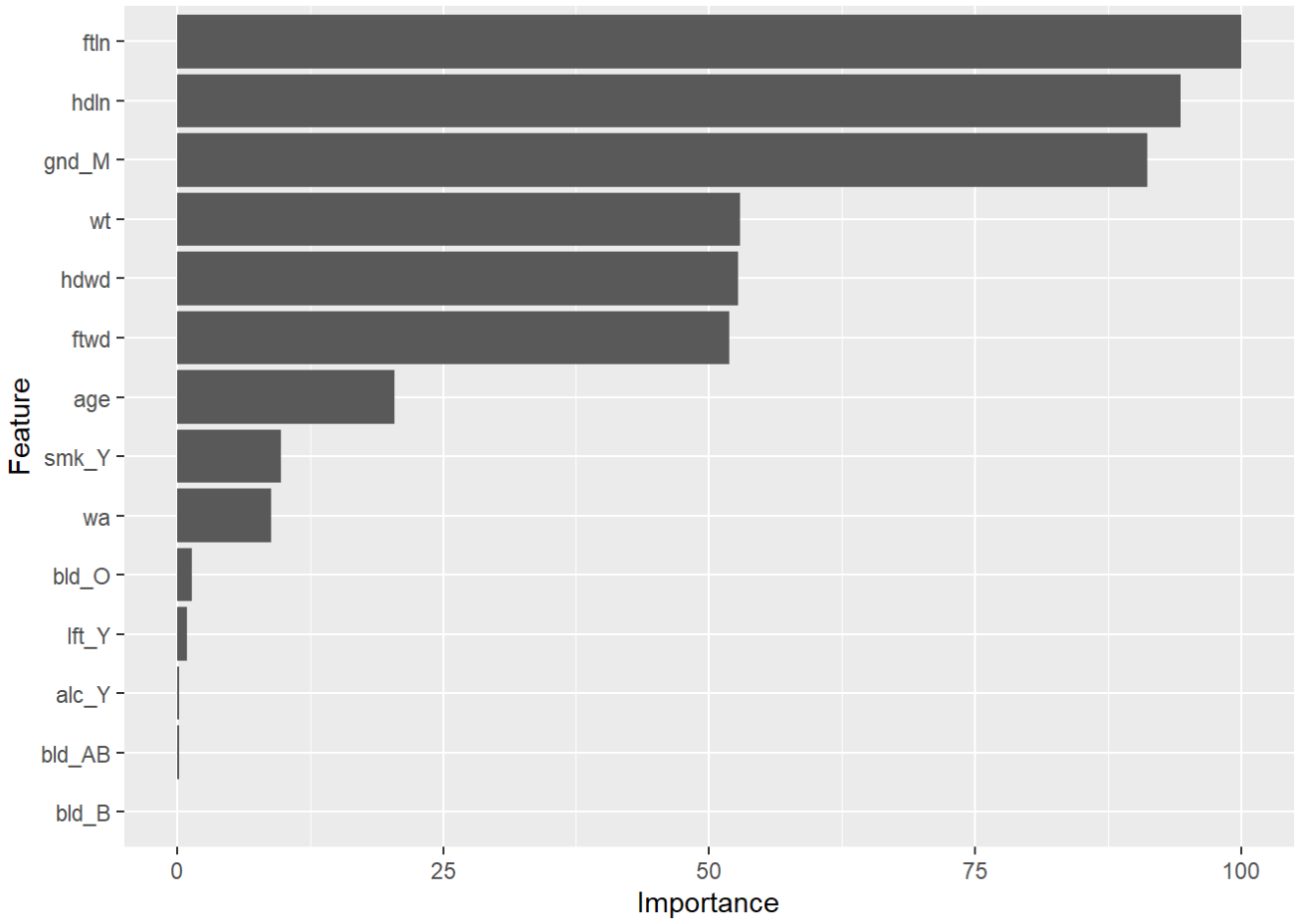
```
Mstep$results
```

```
##   parameter     RMSE  Rsquared      MAE    RMSESD RsquaredSD    MAESD
## 1      none 3.263257 0.8739242 2.629523 0.4413687  0.0323226 0.403908
```

```
# (X) plot(Mstep)
summary(Mstep)
```

```
##
## Call:
## lm(formula = .outcome ~ wt + wa + hdln + ftln + ftwd + gnd_M +
##     alc_Y, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3548 -2.3313 -0.0639  1.8351  8.6458
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 107.6314     5.8727  18.327  < 2e-16 ***
## wt            0.4470     0.0418  10.695  < 2e-16 ***
## wa           -0.4565     0.0440 -10.375  < 2e-16 ***
## hdln          2.0014     0.3718   5.383 1.90e-07 ***
## ftln          1.4200     0.3123   4.547 9.03e-06 ***
## ftwd         -0.7449     0.4537  -1.642    0.102
## gnd_M         5.4899     0.6462   8.495 3.16e-15 ***
## alc_Y        -1.0064     0.6293  -1.599    0.111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.151 on 217 degrees of freedom
## Multiple R-squared:  0.8816, Adjusted R-squared:  0.8778
## F-statistic: 230.9 on 7 and 217 DF,  p-value: < 2.2e-16
```

```
ggplot(varImp(Mstep))
```



```
Mstep$bestTune
```

```
##   parameter
## 1      none
```

```
Mstep$finalModel #lm 객체(TR을 재적합한 모형)
```

```
##
## Call:
## lm(formula = .outcome ~ wt + wa + hdln + ftln + ftwd + gnd_M +
##     alc_Y, data = dat)
##
## Coefficients:
## (Intercept)           wt           wa         hdln         ftln         ftwd
##    107.6314       0.4470      -0.4565       2.0014       1.4200      -0.7449
##       gnd_M        alc_Y
##      5.4899      -1.0064
```

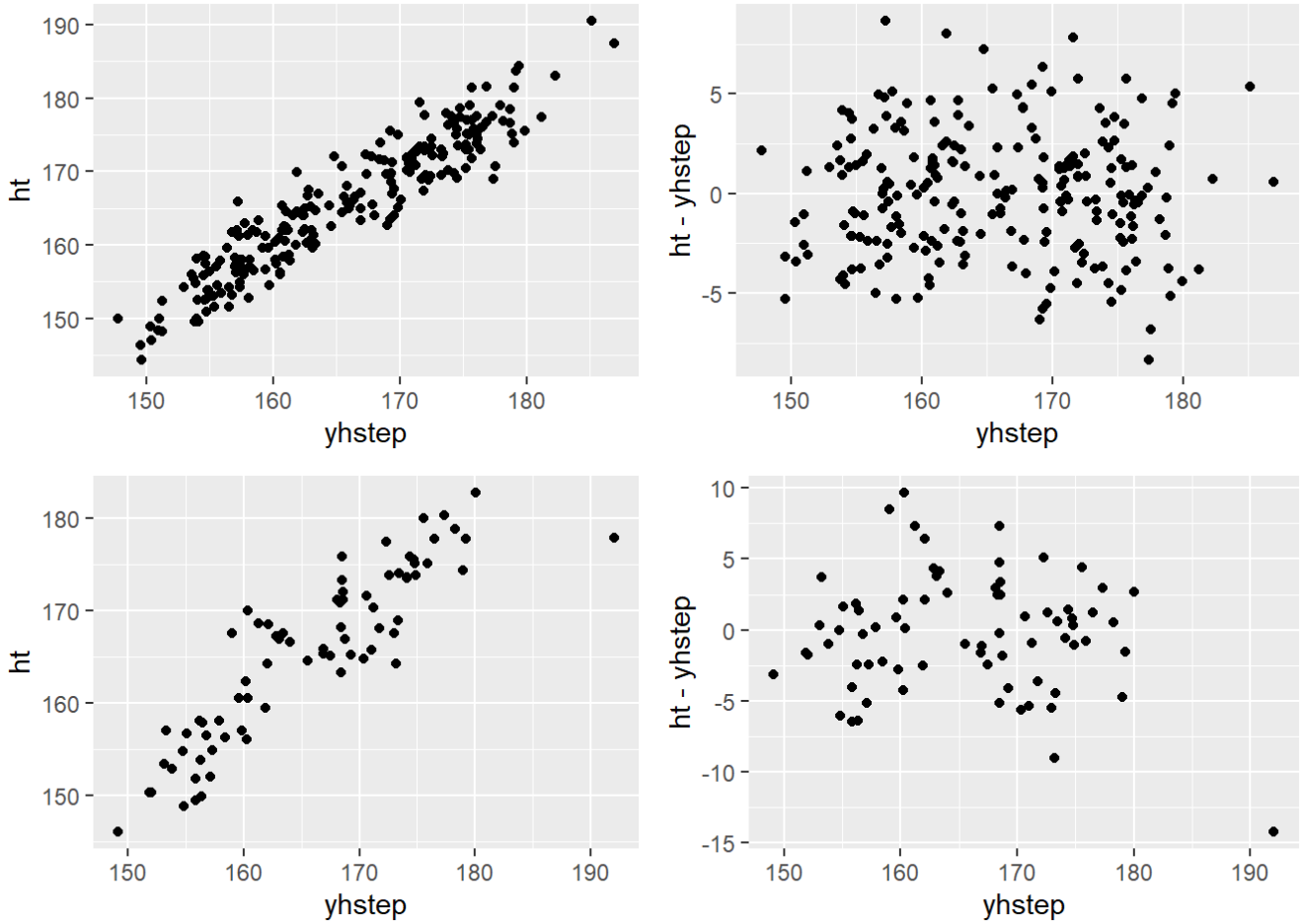
```
Mstep$resample
```

```
##        RMSE  Rsquared      MAE Resample
## 1  4.040736 0.8466410 3.428224   Fold01
## 2  3.317638 0.8785077 2.690494   Fold02
## 3  2.977722 0.8829252 2.315698   Fold03
## 4  3.055259 0.8443997 2.609918   Fold04
## 5  2.473252 0.9046306 1.946962   Fold05
## 6  3.276441 0.9099531 2.604198   Fold06
## 7  3.358299 0.8714640 2.619000   Fold07
## 8  3.805950 0.8111460 3.075412   Fold08
## 9  3.355144 0.8770957 2.625376   Fold09
## 10 2.972126 0.9124791 2.379946   Fold10
```

```
TROUT <- TROUT %>% mutate(yhstep=predict(Mstep, newdata=TR))
TSOUT <- TSOUT %>% mutate(yhstep=predict(Mstep, newdata=TS))
head(TSOUT)
```

```
##      ht     yhlm   yhstep
## 1 173.6 174.0110 174.1567
## 2 150.3 151.3521 151.8827
## 3 150.3 152.5093 152.0495
## 4 149.4 155.8523 155.8620
## 5 152.0 156.9321 157.1251
## 6 180.3 176.9327 177.3713
```

```
g1 <- ggplot(TROUT, aes(x=yhstep, y=ht)) + geom_point()
g2 <- ggplot(TROUT, aes(x=yhstep, y=ht-yhstep)) + geom_point()
g3 <- ggplot(TSOUT, aes(x=yhstep, y=ht)) + geom_point()
g4 <- ggplot(TSOUT, aes(x=yhstep, y=ht-yhstep)) + geom_point()
grid.arrange(g1, g2, g3, g4, ncol=2)
```




```
# For REG, yardstick::mae, rmse, rsq
METstep <-
  bind_cols(
    bind_rows(foo(TROUT$ht, TROUT$yhstep), foo(TSOUT$ht, TSOUT$yhstep)),
    data.frame(model='lmStepAIC', TRTS=c('TR', 'TS')))
METstep
```

```
## # A tibble: 2 x 5
##    rmse   mae   rsq model     TRTS
##   <dbl> <dbl> <dbl> <chr>     <chr>
## 1  3.09  2.49 0.882 lmStepAIC TR
## 2  4.08  3.13 0.802 lmStepAIC TS
```

glmnet, elasticnet, lasso, ridge

- enet은 분류분석에 사용 못함. glmnet 사용해야 함
- glmnet: nlambda=100개를 사전 탐색한 후 lambda를 정함

```
modelLookup('enet')
```

```
##   model parameter                   label forReg forClass probModel
## 1  enet  fraction Fraction of Full Solution   TRUE    FALSE     FALSE
## 2  enet    lambda             Weight Decay   TRUE    FALSE     FALSE
```

```
modelLookup('glmnet')
```

```
##    model parameter                    label forReg forClass probModel
## 1 glmnet     alpha        Mixing Percentage   TRUE     TRUE      TRUE
## 2 glmnet    lambda Regularization Parameter   TRUE     TRUE      TRUE
```

- 적합

```
set.seed(20180968)
glmnetGrid <- expand.grid(alpha=seq(0.1, by=0.24), lambda=seq(0.0, 0.1, by=0.01))
Mglmnet <- train(RC, data=TR,
                 method='glmnet',
                 trControl=trCntl,
                 tuneGrid = glmnetGrid)
```

```
## Loading required namespace: glmnet
```

```
## Warning: package 'glmnet' was built under R version 4.0.5
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
## Loaded glmnet 4.1-1
```

```
Mglmnet
```

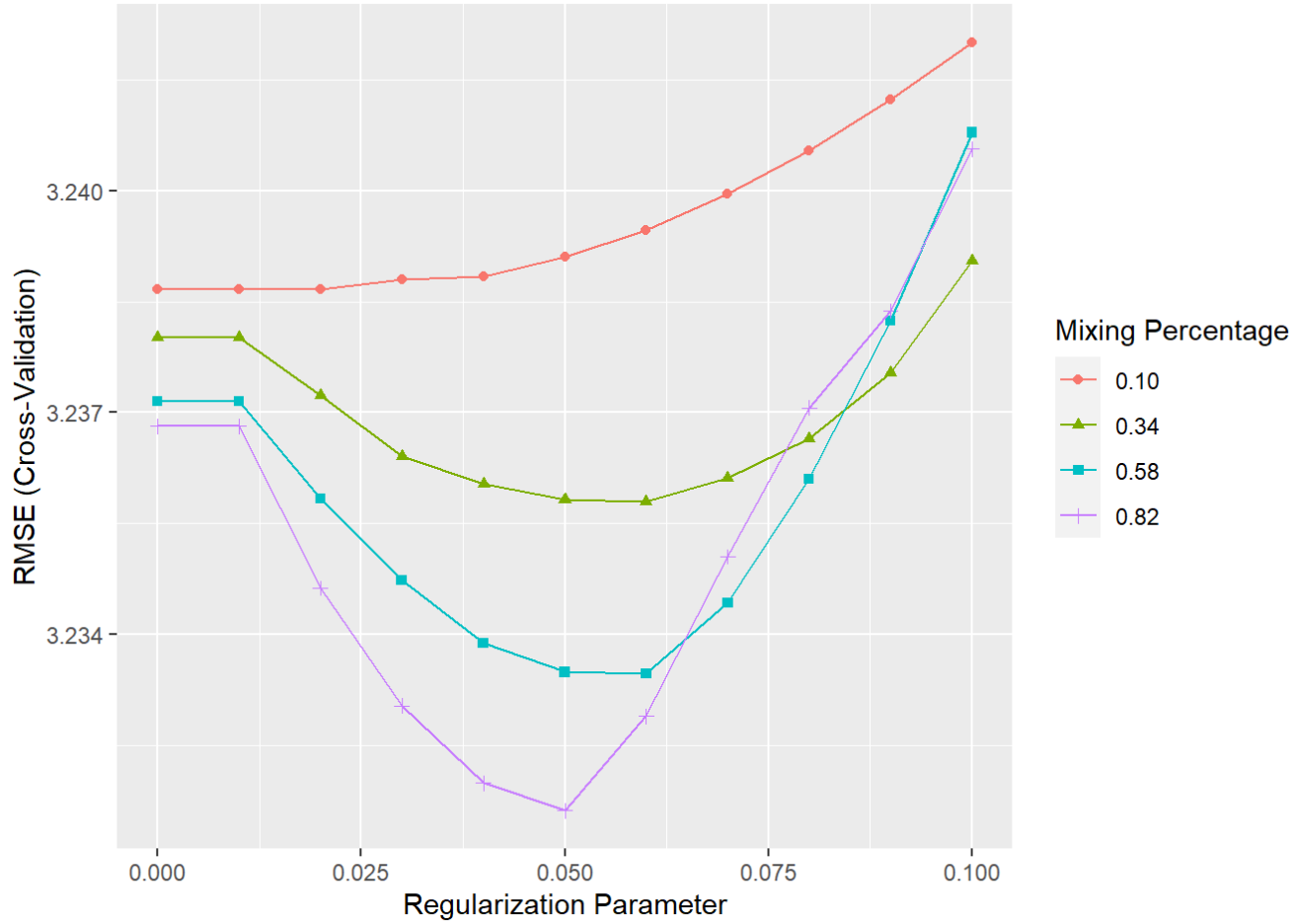
```
## glmnet
##
## 225 samples
##  12 predictor
##
## Recipe steps: impute_median, impute_mode, dummy
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 203, 201, 203, 204, 201, 203, ...
## Resampling results across tuning parameters:
##
##   alpha  lambda  RMSE      Rsquared   MAE
##   0.10   0.00    3.238669  0.8759964  2.602713
##   0.10   0.01    3.238669  0.8759964  2.602713
##   0.10   0.02    3.238669  0.8759964  2.602713
##   0.10   0.03    3.238803  0.8759938  2.602120
##   0.10   0.04    3.238843  0.8760033  2.601167
##   0.10   0.05    3.239104  0.8759953  2.600537
##   0.10   0.06    3.239470  0.8759813  2.600067
##   0.10   0.07    3.239948  0.8759602  2.599665
##   0.10   0.08    3.240541  0.8759342  2.599344
##   0.10   0.09    3.241231  0.8759015  2.599286
##   0.10   0.10    3.241998  0.8758614  2.599315
##   0.34   0.00    3.238018  0.8760387  2.603436
##   0.34   0.01    3.238018  0.8760387  2.603436
##   0.34   0.02    3.237235  0.8760845  2.602650
##   0.34   0.03    3.236408  0.8761461  2.601461
##   0.34   0.04    3.236032  0.8761800  2.600361
##   0.34   0.05    3.235823  0.8762094  2.599466
##   0.34   0.06    3.235795  0.8762227  2.598511
##   0.34   0.07    3.236116  0.8762084  2.598338
##   0.34   0.08    3.236644  0.8761821  2.598580
##   0.34   0.09    3.237538  0.8761326  2.598750
##   0.34   0.10    3.239051  0.8760326  2.599743
##   0.58   0.00    3.237153  0.8760821  2.603627
##   0.58   0.01    3.237153  0.8760821  2.603627
##   0.58   0.02    3.235839  0.8761678  2.602431
##   0.58   0.03    3.234736  0.8762497  2.601191
##   0.58   0.04    3.233887  0.8763168  2.599591
##   0.58   0.05    3.233495  0.8763468  2.598680
##   0.58   0.06    3.233474  0.8763520  2.598714
##   0.58   0.07    3.234428  0.8762864  2.599528
##   0.58   0.08    3.236103  0.8761658  2.601339
##   0.58   0.09    3.238241  0.8760062  2.603172
##   0.58   0.10    3.240782  0.8758199  2.605579
##   0.82   0.00    3.236813  0.8760982  2.603814
##   0.82   0.01    3.236813  0.8760982  2.603814
##   0.82   0.02    3.234628  0.8762382  2.602225
##   0.82   0.03    3.233037  0.8763536  2.600394
##   0.82   0.04    3.231996  0.8764216  2.599195
##   0.82   0.05    3.231622  0.8764462  2.598881
##   0.82   0.06    3.232900  0.8763419  2.600745
##   0.82   0.07    3.235052  0.8761705  2.603261
##   0.82   0.08    3.237065  0.8760085  2.605000
##   0.82   0.09    3.238369  0.8759142  2.606319
##   0.82   0.10    3.240561  0.8757655  2.608958
##
```

```
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were alpha = 0.82 and lambda = 0.05.
```

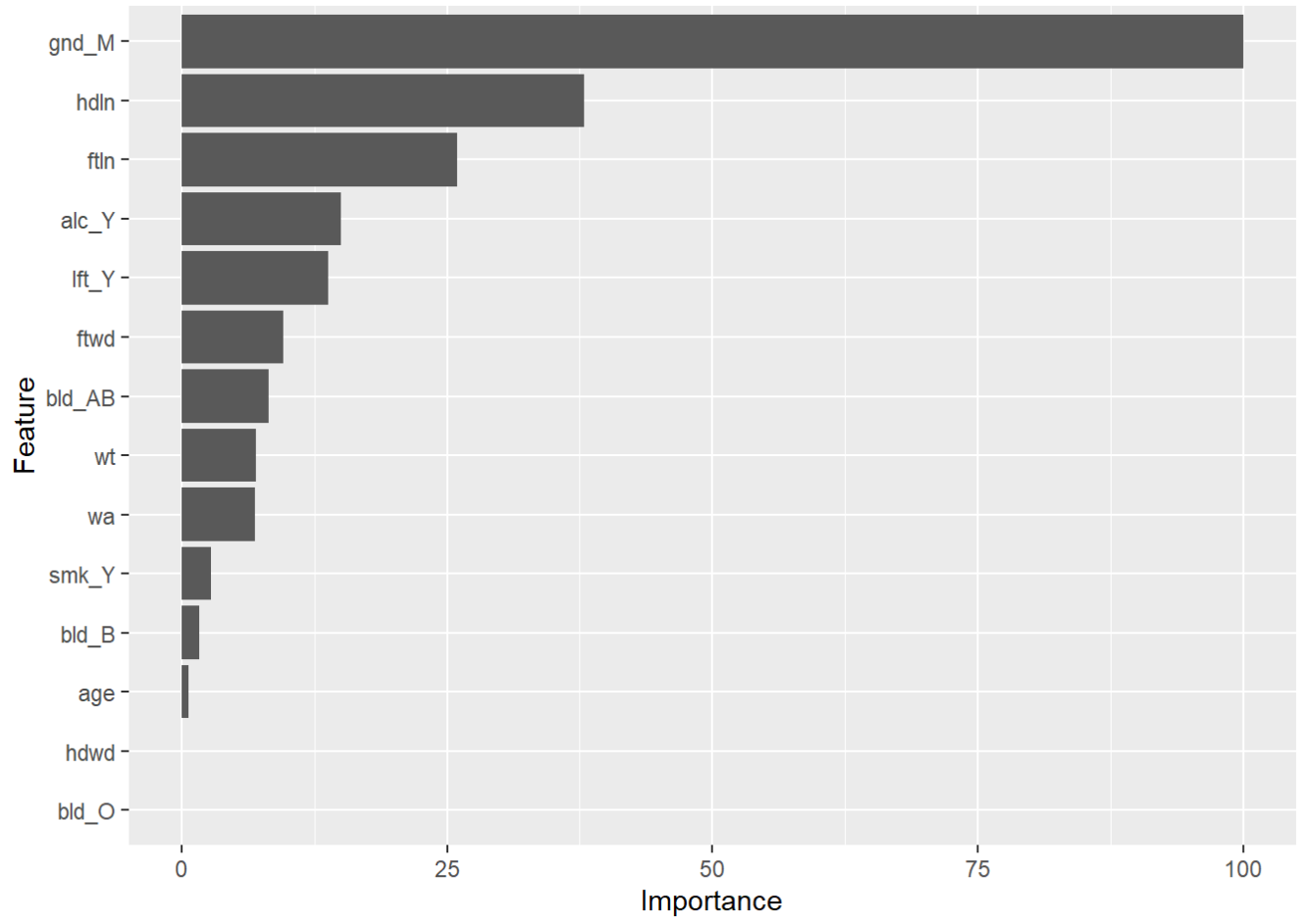
```
Mglmnet$results # 튜닝결과
```

```
##    alpha lambda     RMSE  Rsquared      MAE    RMSESD RsquaredSD     MAESD
## 1   0.10   0.00 3.238669 0.8759964 2.602713 0.4022375 0.02800729 0.3608910
## 2   0.10   0.01 3.238669 0.8759964 2.602713 0.4022375 0.02800729 0.3608910
## 3   0.10   0.02 3.238669 0.8759964 2.602713 0.4022375 0.02800729 0.3608910
## 4   0.10   0.03 3.238803 0.8759938 2.602120 0.4012571 0.02785710 0.3605752
## 5   0.10   0.04 3.238843 0.8760033 2.601167 0.4005820 0.02770127 0.3605766
## 6   0.10   0.05 3.239104 0.8759953 2.600537 0.3998738 0.02754754 0.3604456
## 7   0.10   0.06 3.239470 0.8759813 2.600067 0.3992940 0.02740470 0.3604087
## 8   0.10   0.07 3.239948 0.8759602 2.599665 0.3988164 0.02726978 0.3605304
## 9   0.10   0.08 3.240541 0.8759342 2.599344 0.3984344 0.02714932 0.3609684
## 10  0.10   0.09 3.241231 0.8759015 2.599286 0.3981340 0.02703585 0.3615686
## 11  0.10   0.10 3.241998 0.8758614 2.599315 0.3979540 0.02693186 0.3622089
## 12  0.34   0.00 3.238018 0.8760387 2.603436 0.4034927 0.02811666 0.3618156
## 13  0.34   0.01 3.238018 0.8760387 2.603436 0.4034927 0.02811666 0.3618156
## 14  0.34   0.02 3.237235 0.8760845 2.602650 0.4029405 0.02806889 0.3613539
## 15  0.34   0.03 3.236408 0.8761461 2.601461 0.4022138 0.02791069 0.3615389
## 16  0.34   0.04 3.236032 0.8761800 2.600361 0.4017329 0.02776580 0.3620840
## 17  0.34   0.05 3.235823 0.8762094 2.599466 0.4016562 0.02766184 0.3631831
## 18  0.34   0.06 3.235795 0.8762227 2.598511 0.4016739 0.02756314 0.3646064
## 19  0.34   0.07 3.236116 0.8762084 2.598338 0.4017239 0.02747004 0.3660596
## 20  0.34   0.08 3.236644 0.8761821 2.598580 0.4019662 0.02739215 0.3676707
## 21  0.34   0.09 3.237538 0.8761326 2.598750 0.4023297 0.02732310 0.3693572
## 22  0.34   0.10 3.239051 0.8760326 2.599743 0.4027831 0.02728084 0.3712590
## 23  0.58   0.00 3.237153 0.8760821 2.603627 0.4039608 0.02816728 0.3621186
## 24  0.58   0.01 3.237153 0.8760821 2.603627 0.4039608 0.02816728 0.3621186
## 25  0.58   0.02 3.235839 0.8761678 2.602431 0.4035612 0.02808709 0.3619292
## 26  0.58   0.03 3.234736 0.8762497 2.601191 0.4034347 0.02795389 0.3628893
## 27  0.58   0.04 3.233887 0.8763168 2.599591 0.4038501 0.02787407 0.3649023
## 28  0.58   0.05 3.233495 0.8763468 2.598680 0.4043704 0.02780232 0.3670124
## 29  0.58   0.06 3.233474 0.8763520 2.598714 0.4050228 0.02773365 0.3692124
## 30  0.58   0.07 3.234428 0.8762864 2.599528 0.4060353 0.02770051 0.3718702
## 31  0.58   0.08 3.236103 0.8761658 2.601339 0.4071329 0.02771245 0.3747487
## 32  0.58   0.09 3.238241 0.8760062 2.603172 0.4084574 0.02777533 0.3777942
## 33  0.58   0.10 3.240782 0.8758199 2.605579 0.4102326 0.02786780 0.3810825
## 34  0.82   0.00 3.236813 0.8760982 2.603814 0.4049175 0.02824250 0.3627914
## 35  0.82   0.01 3.236813 0.8760982 2.603814 0.4049175 0.02824250 0.3627914
## 36  0.82   0.02 3.234628 0.8762382 2.602225 0.4042263 0.02810668 0.3626411
## 37  0.82   0.03 3.233037 0.8763536 2.600394 0.4049950 0.02803947 0.3650044
## 38  0.82   0.04 3.231996 0.8764216 2.599195 0.4059884 0.02797860 0.3678403
## 39  0.82   0.05 3.231622 0.8764462 2.598881 0.4073427 0.02793348 0.3708866
## 40  0.82   0.06 3.232900 0.8763419 2.600745 0.4089465 0.02799484 0.3746125
## 41  0.82   0.07 3.235052 0.8761705 2.603261 0.4110752 0.02813034 0.3787522
## 42  0.82   0.08 3.237065 0.8760085 2.605000 0.4132912 0.02828278 0.3825758
## 43  0.82   0.09 3.238369 0.8759142 2.606319 0.4137737 0.02832759 0.3839515
## 44  0.82   0.10 3.240561 0.8757655 2.608958 0.4140395 0.02832761 0.3850228
```

```
ggplot(Mglmnet)
```




```
# (X) summary(Mglmnet)
ggplot(varImp(Mglmnet))
```

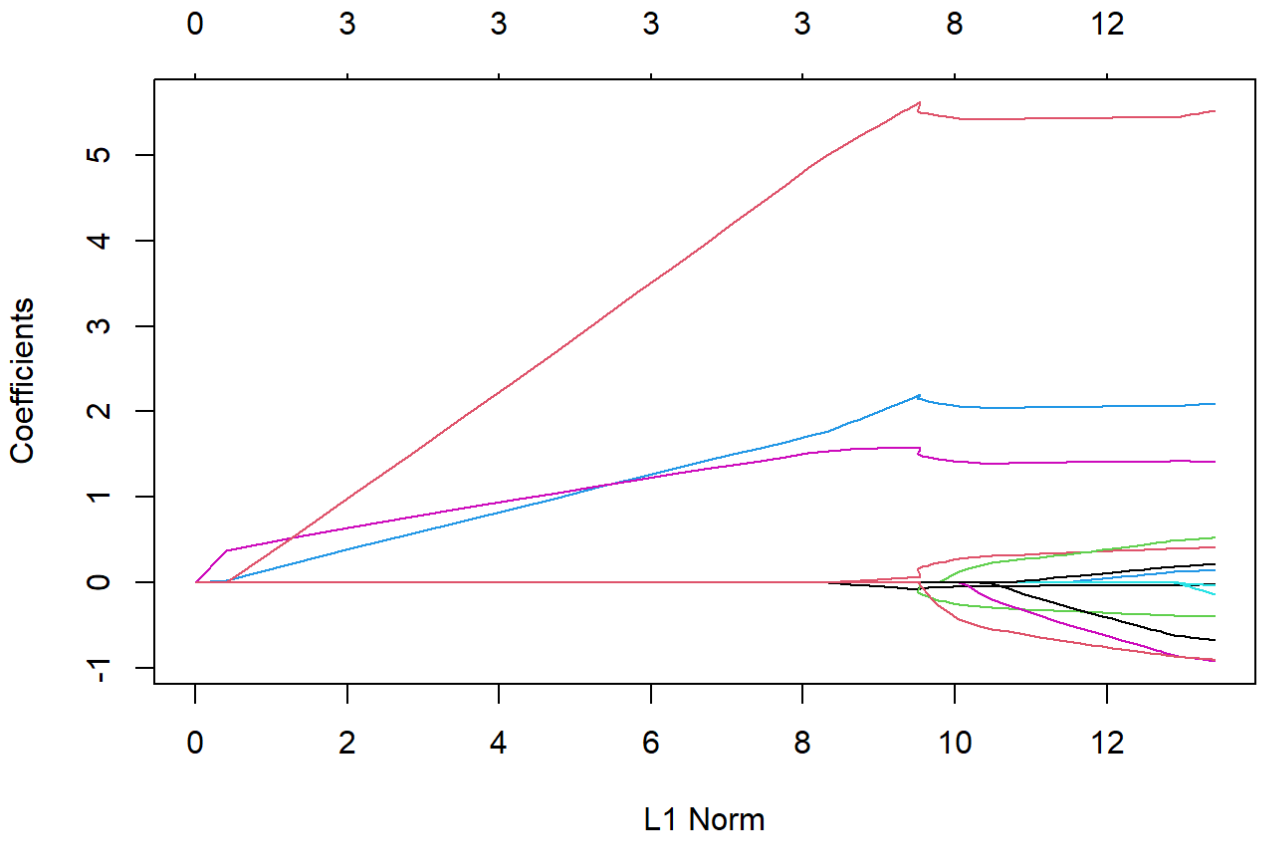


```
MglmnetSbestTune # alpha=1이므로 lasso가 됨됨
```

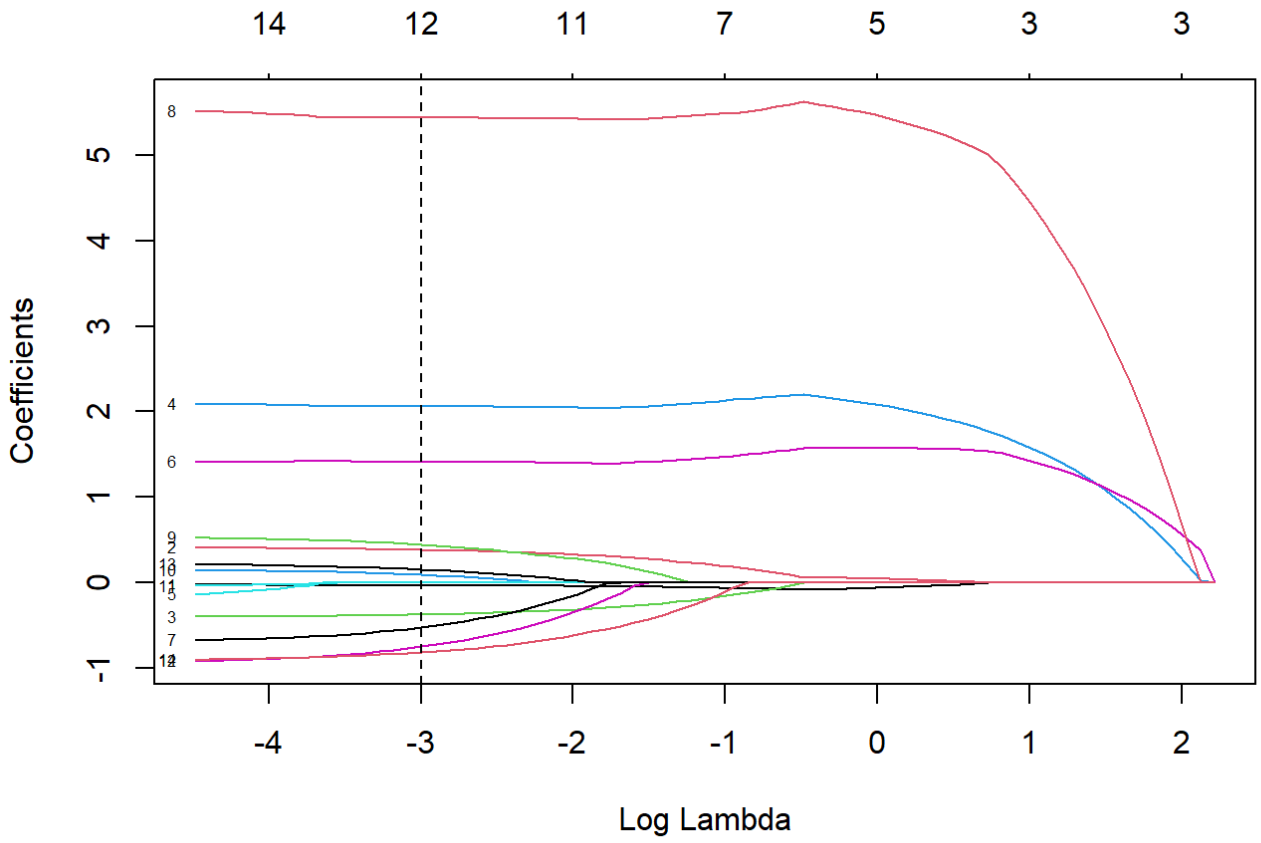
```
##    alpha lambda
## 39  0.82   0.05
```

```
# (X) MglmnetSfinalModel # glmnet 객체

# lasso plot: x: L1 Norm vs Ceofficients
plot(MglmnetSfinalModel)
```




```
# lasso plot: x: log(lamnda) vs Coefficients
plot(MglmnetSfinalModel, xvar='lambda', label=TRUE)
abline(v=log(MglmnetSbestTune$lambda), lty=2)
```




```
coef(Mglmnet$final, s=Mglmnet$bestTune$lambda)
```

```
## 15 x 1 sparse Matrix of class "dgCMatrix"
##                        1
## (Intercept) 103.27218166
## age          -0.03331879
## wt            0.38122537
## wa           -0.37508211
## hdln          2.06280719
## hdwd          .
## ftln          1.41344162
## ftwd         -0.52355250
## gnd_M         5.44200753
## bld_AB        0.44564697
## bld_B         0.09075481
## bld_O         .
## lft_Y        -0.75187603
## smk_Y         0.15234525
## alc_Y        -0.81647968
```

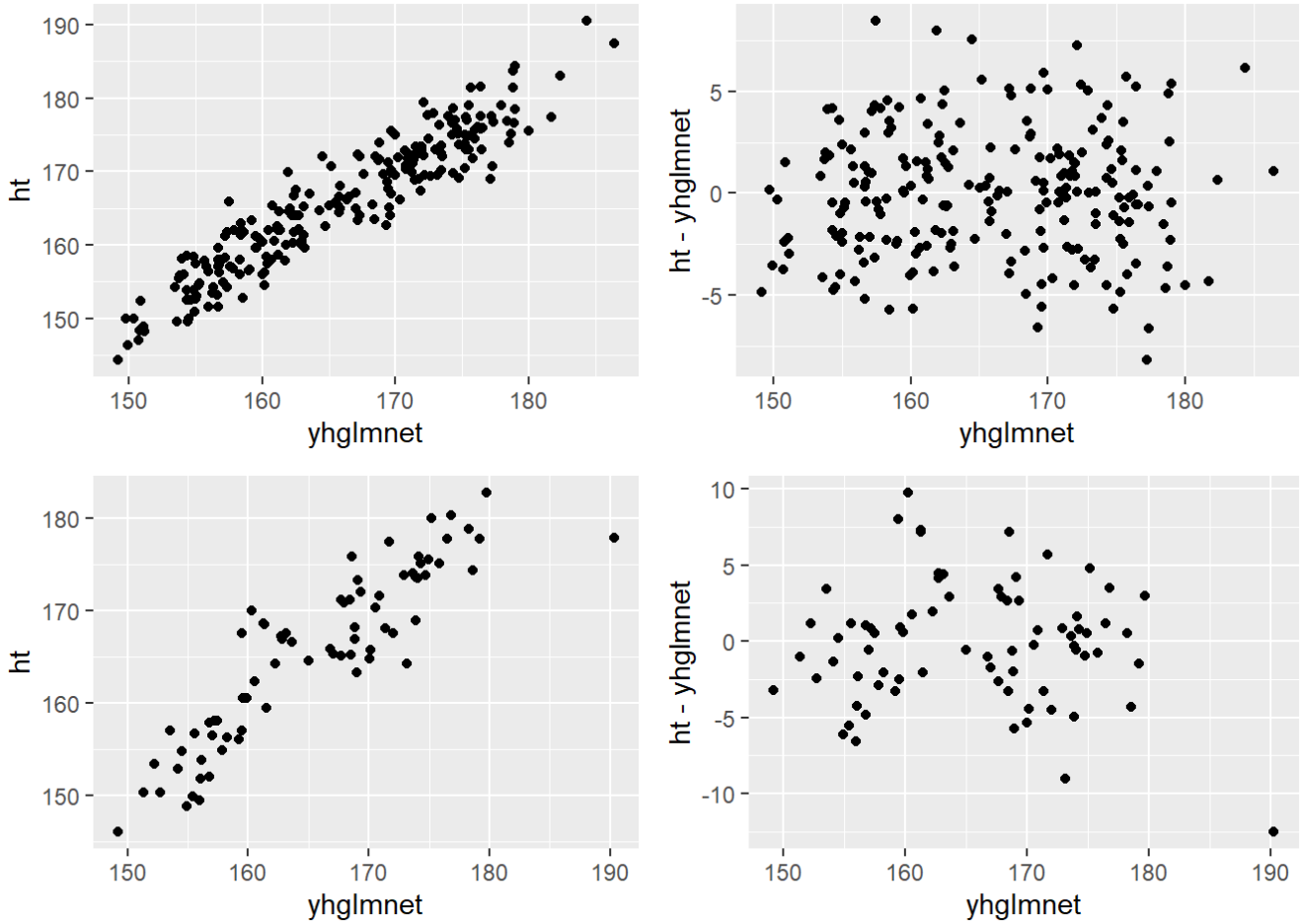
```
Mglmnet$resample # CV 폴더별 평가측도. densityplot(Mglmnet)
```

```
##        RMSE  Rsquared      MAE Resample
## 1  3.702864 0.8240406 2.996344   Fold08
## 2  2.468003 0.9046150 1.938548   Fold05
## 3  3.259792 0.8805444 2.522557   Fold09
## 4  3.326708 0.9062176 2.588750   Fold06
## 5  3.357782 0.8761392 2.695399   Fold02
## 6  2.990605 0.9117040 2.393667   Fold10
## 7  3.917740 0.8569581 3.337803   Fold01
## 8  3.325528 0.8720157 2.520638   Fold07
## 9  2.954271 0.8851321 2.413958   Fold03
## 10 3.012925 0.8470951 2.581145   Fold04
```

```
TROUT <- TROUT %>% mutate(yhglmnet=predict(Mglmnet, newdata=TR))
TSOUT <- TSOUT %>% mutate(yhglmnet=predict(Mglmnet, newdata=TS))
head(TSOUT)
```

```
##      ht     yhlm   yhstep yhglmnet
## 1 173.6 174.0110 174.1567 173.9229
## 2 150.3 151.3521 151.8827 151.3485
## 3 150.3 152.5093 152.0495 152.7120
## 4 149.4 155.8523 155.8620 155.9364
## 5 152.0 156.9321 157.1251 156.8017
## 6 180.3 176.9327 177.3713 176.8266
```

```
g1 <- ggplot(TROUT, aes(x=yhglmnet, y=ht)) + geom_point()
g2 <- ggplot(TROUT, aes(x=yhglmnet, y=ht-yhglmnet)) + geom_point()
g3 <- ggplot(TSOUT, aes(x=yhglmnet, y=ht)) + geom_point()
g4 <- ggplot(TSOUT, aes(x=yhglmnet, y=ht-yhglmnet)) + geom_point()
grid.arrange(g1, g2, g3, g4, ncol=2)
```




```
# For REG, yardstick::mae, rmse, rsq
METglmnet <-
  bind_cols(
    bind_rows(foo(TROUT$ht, TROUT$yhglmnet), foo(TSOUT$ht, TSOUT$yhglmnet)),
    data.frame(model='glmnet', TRTS=c('TR', 'TS')))
METglmnet
```

```
## # A tibble: 2 x 5
##    rmse   mae   rsq model  TRTS
##   <dbl> <dbl> <dbl> <chr>  <chr>
## 1  3.08  2.46 0.883 glmnet TR
## 2  3.96  3.06 0.811 glmnet TS
```

rpart

- rpart(회귀나무)

```
modelLookup('rpart') # 튜닝모수 cp:Complexity parameter
```

```
##   model parameter                label forReg forClass probModel
## 1 rpart        cp Complexity Parameter   TRUE     TRUE      TRUE
```

```
modelLookup('rpart2') # 튜닝모수 maxdepth
```

```
##    model parameter          label forReg forClass probModel
## 1 rpart2  maxdepth Max Tree Depth   TRUE     TRUE      TRUE
```

- 적합

```
set.seed(20180968)
rpartGrid <- expand.grid(cp=seq(0,0.2,length=10))

Mrpart <- train(RC, data=TR,
                method='rpart',  #Regression Tree
                trControl=trCntl,
                tuneGrid = rpartGrid)  #tuneLength=5, metric='RMSE,Rsquared'
```

```
##
## Attaching package: 'rpart'
```

```
## The following object is masked from 'package:dials':
##
##     prune
```

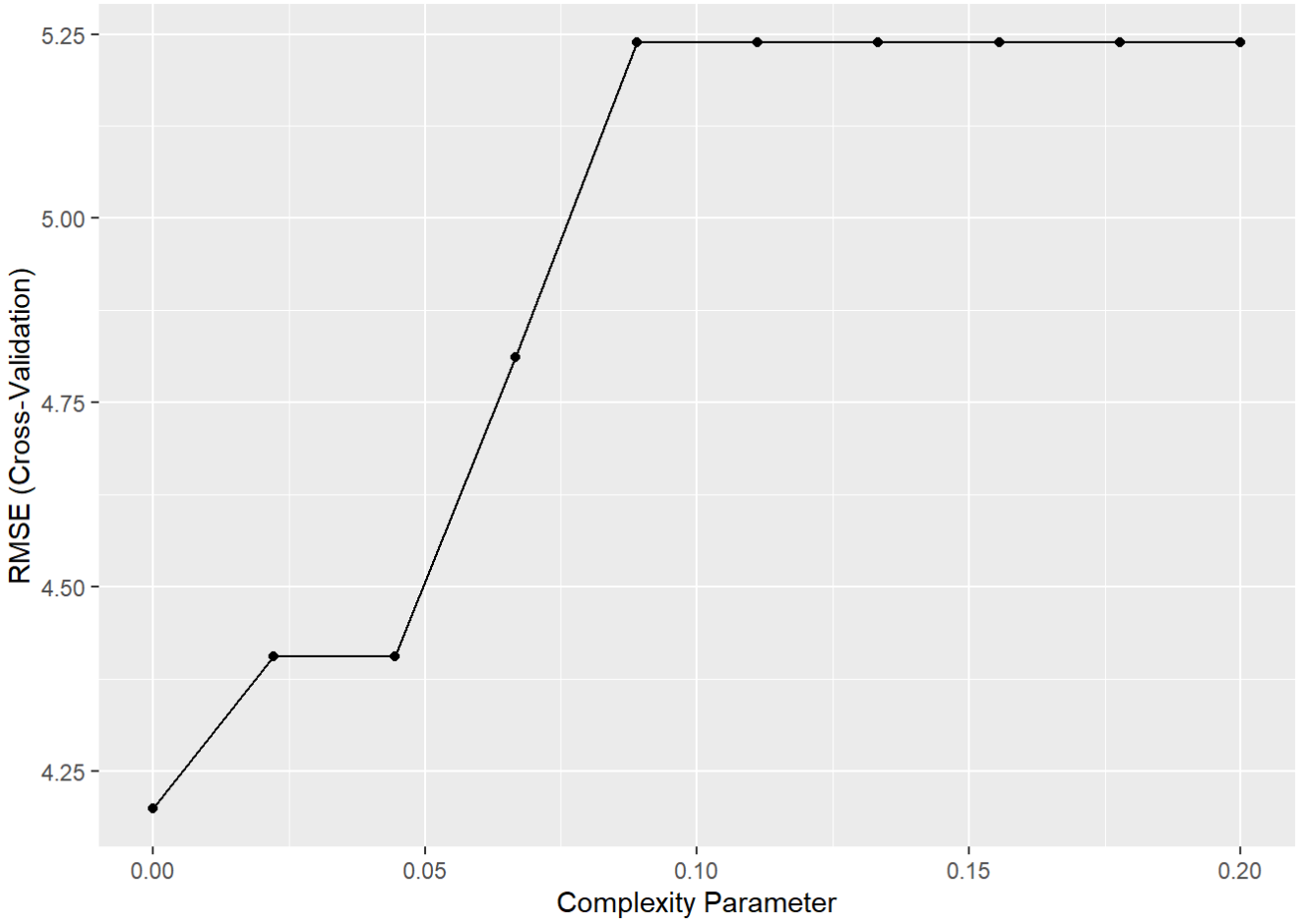
```
Mrpart
```

```
## CART
##
## 225 samples
##  12 predictor
##
## Recipe steps: impute_median, impute_mode, dummy
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 203, 201, 203, 204, 201, 203, ...
## Resampling results across tuning parameters:
##
##   cp          RMSE      Rsquared   MAE
##   0.00000000  4.198665  0.7955190  3.385144
##   0.02222222  4.405040  0.7716339  3.541678
##   0.04444444  4.405040  0.7716339  3.541678
##   0.06666667  4.811622  0.7249634  3.824715
##   0.08888889  5.238482  0.6719855  4.168392
##   0.11111111  5.238482  0.6719855  4.168392
##   0.13333333  5.238482  0.6719855  4.168392
##   0.15555556  5.238482  0.6719855  4.168392
##   0.17777778  5.238482  0.6719855  4.168392
##   0.20000000  5.238482  0.6719855  4.168392
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was cp = 0.
```

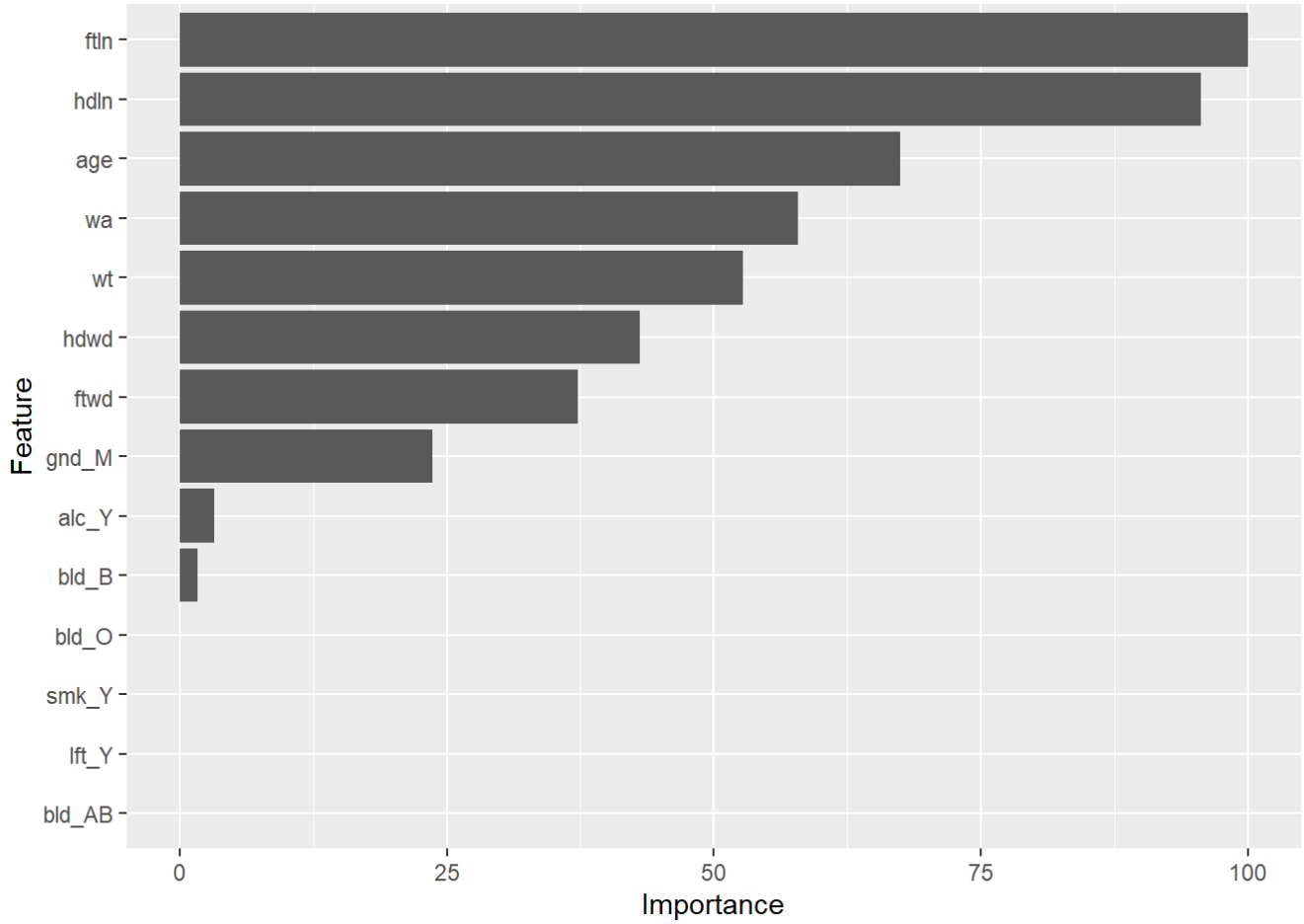
```
Mrpart$results # 튜닝 결과
```

```
##            cp     RMSE  Rsquared      MAE    RMSESD RsquaredSD     MAESD
## 1  0.00000000 4.198665 0.7955190 3.385144 0.3797437 0.03417153 0.3646081
## 2  0.02222222 4.405040 0.7716339 3.541678 0.6470732 0.04711557 0.4493578
## 3  0.04444444 4.405040 0.7716339 3.541678 0.6470732 0.04711557 0.4493578
## 4  0.06666667 4.811622 0.7249634 3.824715 1.0209247 0.07875000 0.7770192
## 5  0.08888889 5.238482 0.6719855 4.168392 0.9367782 0.06624606 0.7023085
## 6  0.11111111 5.238482 0.6719855 4.168392 0.9367782 0.06624606 0.7023085
## 7  0.13333333 5.238482 0.6719855 4.168392 0.9367782 0.06624606 0.7023085
## 8  0.15555556 5.238482 0.6719855 4.168392 0.9367782 0.06624606 0.7023085
## 9  0.17777778 5.238482 0.6719855 4.168392 0.9367782 0.06624606 0.7023085
## 10 0.20000000 5.238482 0.6719855 4.168392 0.9367782 0.06624606 0.7023085
```

```
ggplot(Mrpart) #M$results 시각화 size(#Hidden Units) vs RMSE
```




```
# (long) summary(Mrpart)
ggplot(varImp(Mrpart))
```



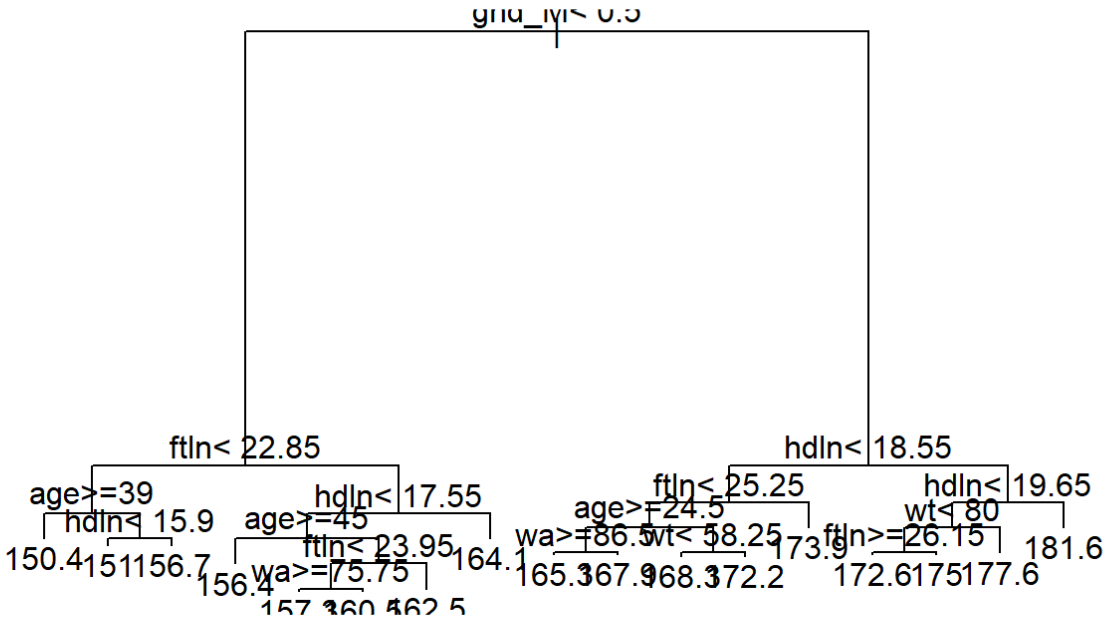

```
Mrpart$bestTune #alpha=1이므로 lasso가 됨
```

```
##   cp
## 1  0
```

```
Mrpart$finalModel #nnet 객체
```

```
## n= 225
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 225 18200.53000 165.7084
##    2) gnd_M< 0.5 107  3076.40300 158.0813
##      4) ftln< 22.85 35   582.24740 153.0514
##        8) age>=39 12   130.06920 150.3583 *
##        9) age< 39 23   319.73650 154.4565
##         18) hdln< 15.9 9    37.00000 150.9667 *
##         19) hdln>=15.9 14   102.66000 156.7000 *
##      5) ftln>=22.85 72  1178.22000 160.5264
##       10) hdln< 17.55 58   731.33880 159.6603
##         20) age>=45 12    24.14667 156.4333 *
##         21) age< 45 46   549.62980 160.5022
##           42) ftln< 23.95 29   303.40690 159.3103
##             84) wa>=75.75 11   106.64910 157.3091 *
##             85) wa< 75.75 18   125.78000 160.5333 *
##           43) ftln>=23.95 17   134.75880 162.5353 *
##       11) hdln>=17.55 14   223.15710 164.1143 *
##    3) gnd_M>=0.5 118  3255.31900 172.6246
##      6) hdln< 18.55 71  1246.32200 170.2352
##       12) ftln< 25.25 52   727.29690 168.8923
##         24) age>=24.5 24   296.27830 167.1417
##           48) wa>=86.5 7    96.59714 165.2571 *
##           49) wa< 86.5 17   164.58470 167.9176 *
##         25) age< 24.5 28   294.41860 170.3929
##           50) wt< 58.25 13    95.95077 168.3385 *
##           51) wt>=58.25 15    96.04933 172.1733 *
##       13) ftln>=25.25 19   168.59790 173.9105 *
##      7) hdln>=18.55 47   991.32550 176.2340
##       14) hdln< 19.65 38   461.34550 174.9658
##         28) wt< 80 29   306.51240 174.1517
##           56) ftln>=26.15 10    40.30900 172.5900 *
##           57) ftln< 26.15 19   228.97680 174.9737 *
##         29) wt>=80 9    73.68889 177.5889 *
##       15) hdln>=19.65 9   210.78890 181.5889 *
```

```
plot(Mrpart$finalModel)
text(Mrpart$finalModel)
```




```
library(partykit)
```

```
## Warning: package 'partykit' was built under R version 4.0.5
```

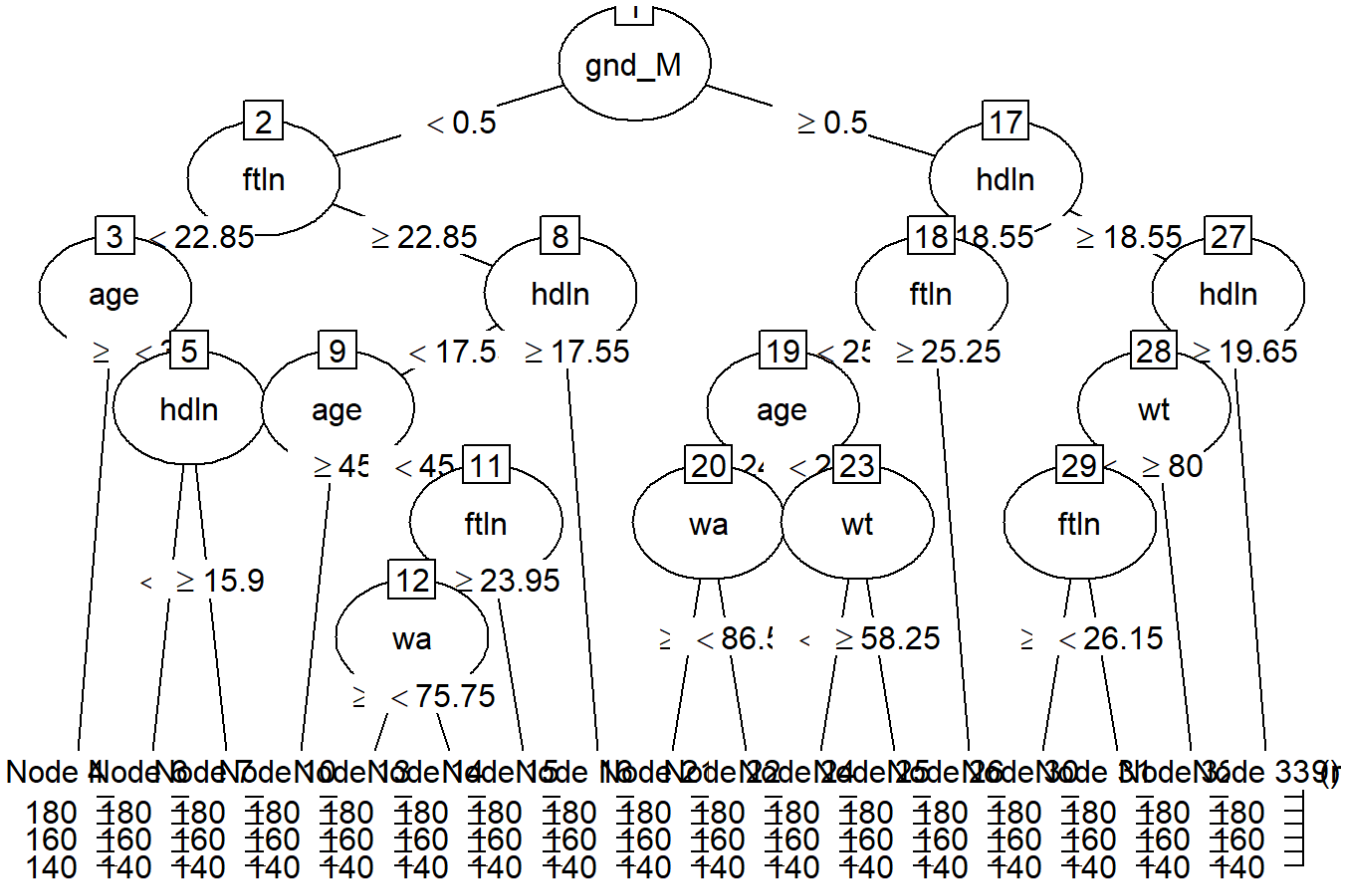
```
## Loading required package: grid
```

```
## Loading required package: libcoin
```

```
## Warning: package 'libcoin' was built under R version 4.0.5
```

```
## Loading required package: mvtnorm
```

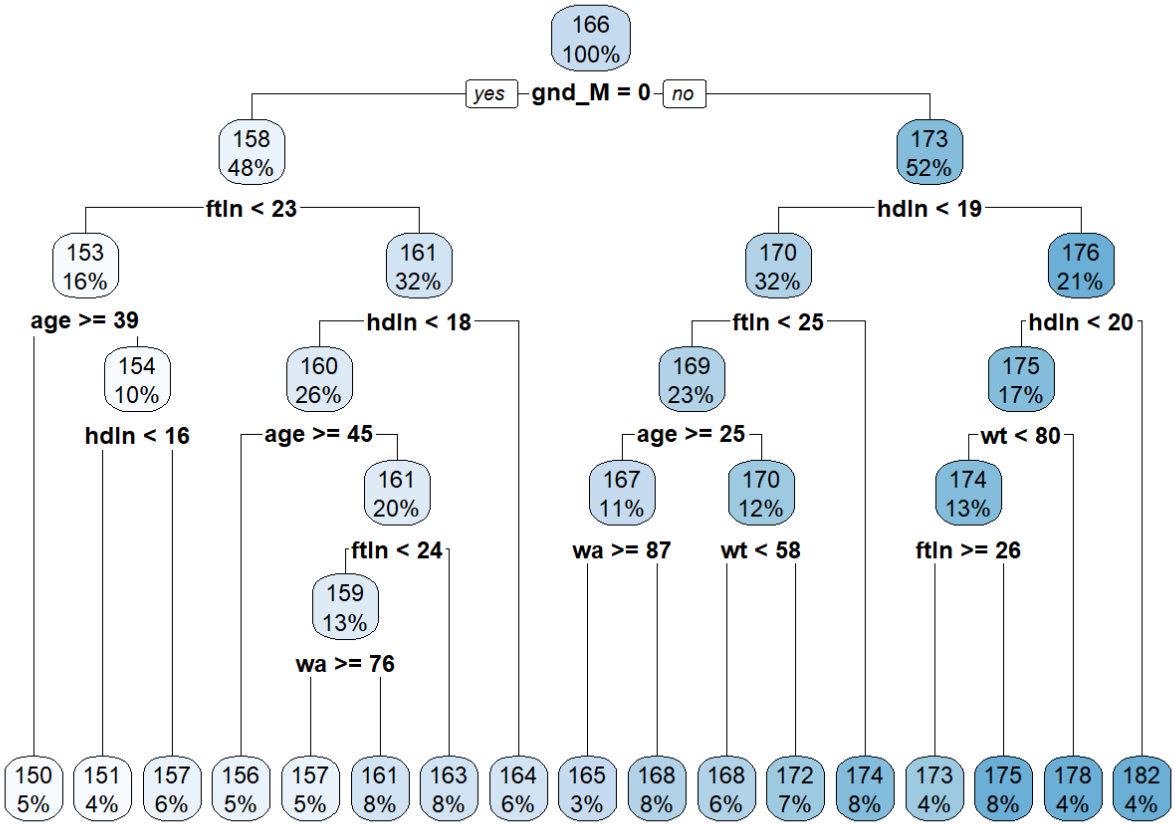
```
plot(as.party(Mrpart$finalModel))
```




```
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 4.0.5
```

```
rpart.plot::rpart.plot(Mrpart$finalModel)
```



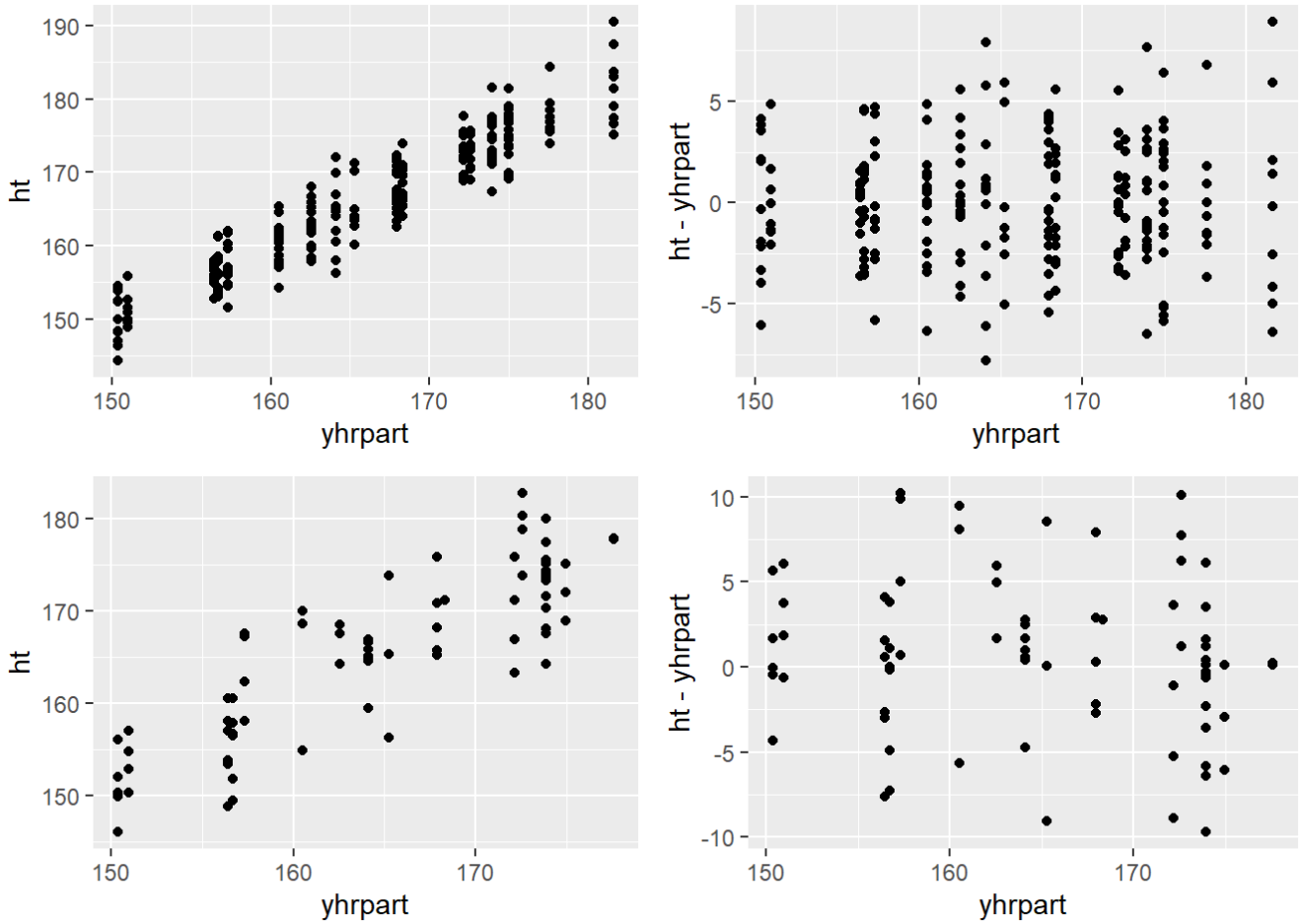

```
Mrpart$resample # CV 폴더별 평가측도 densityplot(Mnbet)
```

```
##        RMSE  Rsquared      MAE Resample
## 1  4.333332 0.7922302 3.424862   Fold02
## 2  3.878418 0.8529099 3.191465   Fold01
## 3  3.742828 0.7816095 2.722178   Fold05
## 4  4.433519 0.8242377 3.606185   Fold06
## 5  4.770831 0.7788670 3.649214   Fold10
## 6  4.142697 0.7616215 3.324100   Fold04
## 7  4.254620 0.7647782 3.660051   Fold08
## 8  4.444421 0.7590654 3.637495   Fold09
## 9  3.526314 0.8453712 2.845062   Fold03
## 10 4.459667 0.7944993 3.790824   Fold07
```

```
TROUT <- TROUT %>% mutate(yhrpart=predict(Mrpart, newdata=TR))
TSOUT <- TSOUT %>% mutate(yhrpart=predict(Mrpart, newdata=TS))
head(TSOUT)
```

```
##      ht     yhlm   yhstep yhglmnet  yhrpart
## 1 173.6 174.0110 174.1567 173.9229 173.9105
## 2 150.3 151.3521 151.8827 151.3485 150.3583
## 3 150.3 152.5093 152.0495 152.7120 150.9667
## 4 149.4 155.8523 155.8620 155.9364 156.7000
## 5 152.0 156.9321 157.1251 156.8017 150.3583
## 6 180.3 176.9327 177.3713 176.8266 172.5900
```

```
g1 <- ggplot(TROUT, aes(x=yhrpart, y=ht)) + geom_point()
g2 <- ggplot(TROUT, aes(x=yhrpart, y=ht-yhrpart)) + geom_point()
g3 <- ggplot(TSOUT, aes(x=yhrpart, y=ht)) + geom_point()
g4 <- ggplot(TSOUT, aes(x=yhrpart, y=ht-yhrpart)) + geom_point()
grid.arrange(g1, g2, g3, g4, ncol=2)
```




```
# For REG, yardstick::mae, rmse, rsq
METrpart <-
  bind_cols(
    bind_rows(foo(TROUT$ht, TROUT$yhrpart), foo(TSOUT$ht, TSOUT$yhrpart)),
    data.frame(model='rpart', TRTS=c('TR', 'TS')))
METrpart
```

```
## # A tibble: 2 x 5
##    rmse   mae   rsq model TRTS
##   <dbl> <dbl> <dbl> <chr> <chr>
## 1  3.03  2.40 0.887 rpart TR
## 2  4.71  3.58 0.741 rpart TS
```

ranger

- ranger:fast random forest

```
modelLookup('ranger') # 튜닝모수 mrtry, splitrule, min.nide.size
```

```
##    model     parameter                          label forReg forClass probModel
## 1 ranger          mtry #Randomly Selected Predictors   TRUE     TRUE      TRUE
## 2 ranger     splitrule                Splitting Rule   TRUE     TRUE      TRUE
## 3 ranger min.node.size             Minimal Node Size   TRUE     TRUE      TRUE
```

- 적합

```
set.seed(20180968)
rangerGrid <- expand.grid(
  mtry=seq(2, ncol(TR) -1, by=2), #mtry=seq(10, ncol(trn)-1, by=2) 조정
  min.node.size=1:3,
  splitrule = c('extratrees') # splitrule='gini, extratrees'. gini는 REG 적용 불가
)

Mranger <- train(RC, data=TR,
                 method='ranger', importance='impurity',
                 trControl=trCntl,
                 tuneGrid = rangerGrid)
```

```
## Loading required namespace: e1071
```

```
## Loading required namespace: ranger
```

```
## Warning: package 'e1071' was built under R version 4.0.5
```

```
##
## Attaching package: 'e1071'
```

```
## The following object is masked from 'package:tune':
##
##     tune
```

```
## The following object is masked from 'package:rsample':
##
##     permutations
```

```
## Warning: package 'ranger' was built under R version 4.0.5
```

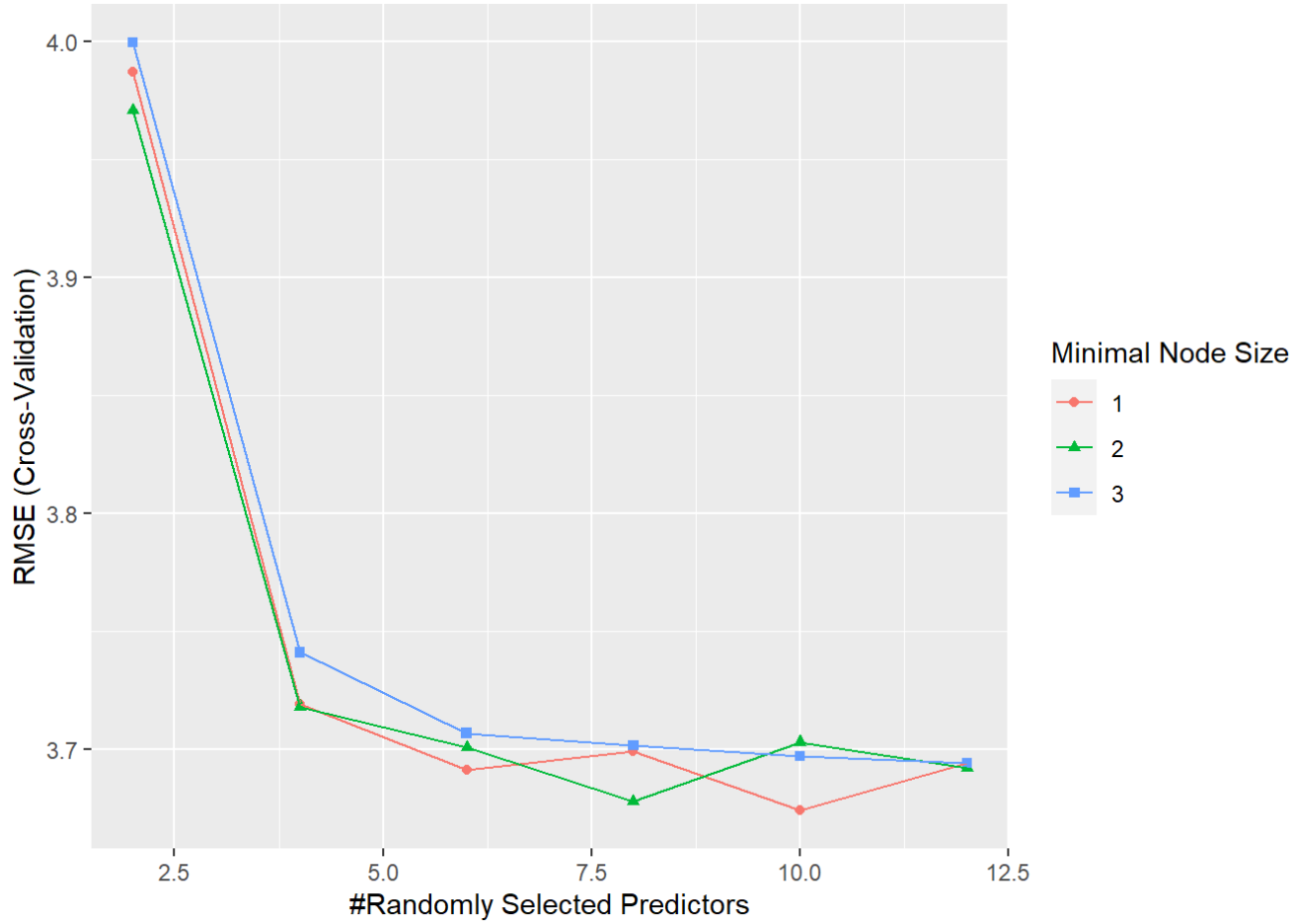
```
#tuneLength=5
#metric='RMSE.Rsquared' 회귀 모형선택기준
Mranger # 튜닝결과 (M$result)
```

```
## Random Forest
##
## 225 samples
##  12 predictor
##
## Recipe steps: impute_median, impute_mode, dummy
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 203, 201, 203, 204, 201, 203, ...
## Resampling results across tuning parameters:
##
##   mtry  min.node.size  RMSE      Rsquared   MAE
##    2    1              3.987041  0.8285565  3.207374
##    2    2              3.970594  0.8296593  3.203099
##    2    3              3.999417  0.8267738  3.224721
##    4    1              3.719415  0.8401963  2.985713
##    4    2              3.718169  0.8399347  2.987976
##    4    3              3.741117  0.8387950  3.008723
##    6    1              3.691443  0.8412773  2.981850
##    6    2              3.701017  0.8398350  2.986421
##    6    3              3.706896  0.8392526  2.988725
##    8    1              3.699122  0.8392267  2.977850
##    8    2              3.677906  0.8414831  2.968961
##    8    3              3.701763  0.8394250  2.987705
##   10    1              3.674095  0.8411646  2.974966
##   10    2              3.703184  0.8384910  2.978624
##   10    3              3.697164  0.8396012  2.976126
##   12    1              3.694088  0.8396032  2.967661
##   12    2              3.692049  0.8397208  2.976615
##   12    3              3.694218  0.8395932  2.977085
##
## Tuning parameter 'splitrule' was held constant at a value of extratrees
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were mtry = 10, splitrule = extratrees
##  and min.node.size = 1.
```

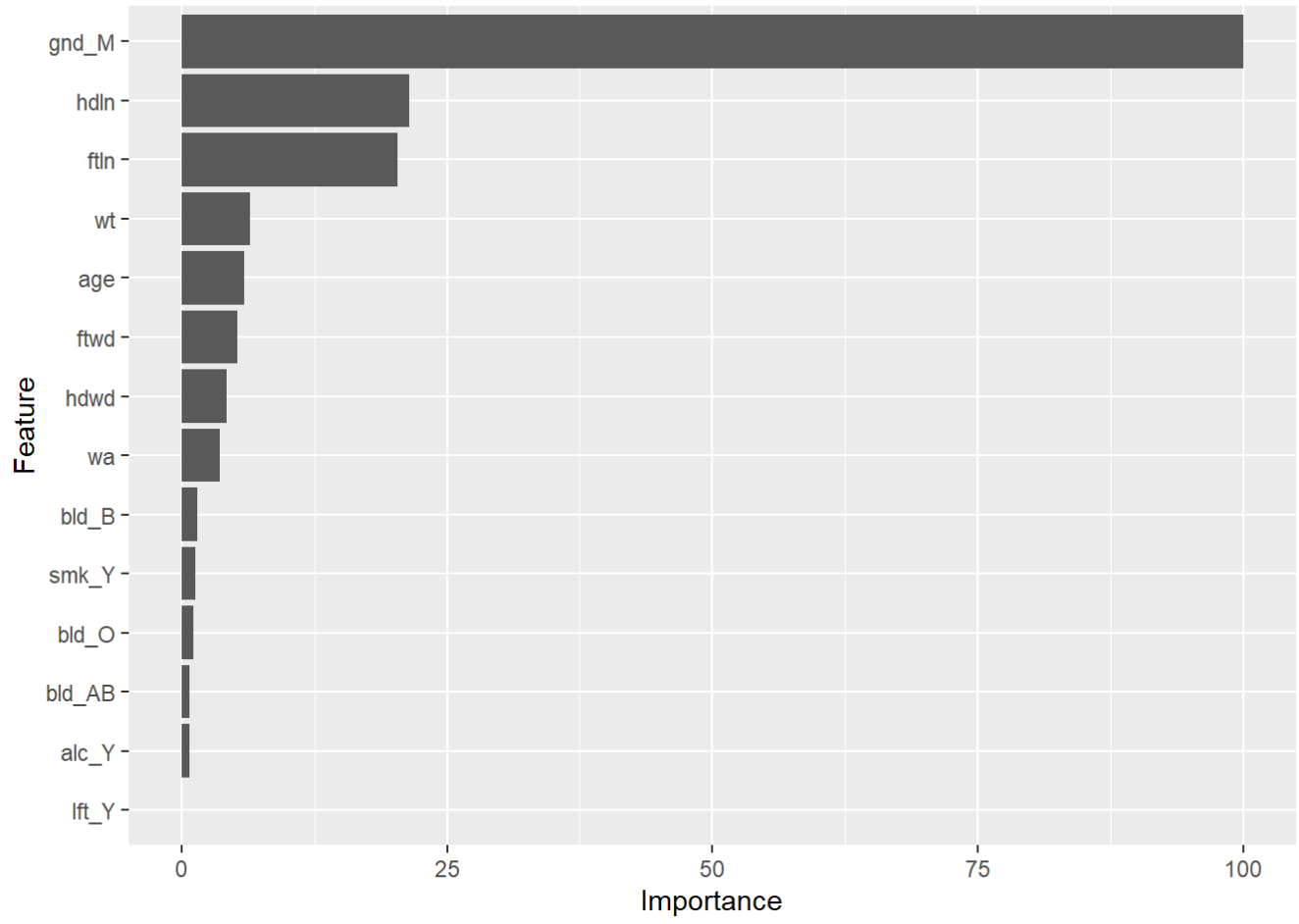
```
Mranger$results # 튜닝 결과
```

```
##    mtry min.node.size  splitrule     RMSE  Rsquared      MAE    RMSESD
## 1     2             1 extratrees 3.987041 0.8285565 3.207374 0.6564907
## 2     2             2 extratrees 3.970594 0.8296593 3.203099 0.6932792
## 3     2             3 extratrees 3.999417 0.8267738 3.224721 0.6888343
## 4     4             1 extratrees 3.719415 0.8401963 2.985713 0.5715269
## 5     4             2 extratrees 3.718169 0.8399347 2.987976 0.5618391
## 6     4             3 extratrees 3.741117 0.8387950 3.008723 0.5822348
## 7     6             1 extratrees 3.691443 0.8412773 2.981850 0.5099818
## 8     6             2 extratrees 3.701017 0.8398350 2.986421 0.5209589
## 9     6             3 extratrees 3.706896 0.8392526 2.988725 0.5650144
## 10    8             1 extratrees 3.699122 0.8392267 2.977850 0.4930408
## 11    8             2 extratrees 3.677906 0.8414831 2.968961 0.5198898
## 12    8             3 extratrees 3.701763 0.8394250 2.987705 0.5360149
## 13   10             1 extratrees 3.674095 0.8411646 2.974966 0.5036205
## 14   10             2 extratrees 3.703184 0.8384910 2.978624 0.4854389
## 15   10             3 extratrees 3.697164 0.8396012 2.976126 0.5116537
## 16   12             1 extratrees 3.694088 0.8396032 2.967661 0.4640200
## 17   12             2 extratrees 3.692049 0.8397208 2.976615 0.4603835
## 18   12             3 extratrees 3.694218 0.8395932 2.977085 0.5008541
##    RsquaredSD     MAESD
## 1  0.02738352 0.5628528
## 2  0.03018059 0.6068475
## 3  0.03267828 0.5828634
## 4  0.02763858 0.5110938
## 5  0.02730533 0.4867954
## 6  0.02746087 0.5141568
## 7  0.02700057 0.4434520
## 8  0.02989780 0.4490604
## 9  0.03007150 0.4750659
## 10 0.03027175 0.4169199
## 11 0.03120652 0.4433196
## 12 0.03067062 0.4500110
## 13 0.03234429 0.4229224
## 14 0.02962456 0.4135926
## 15 0.03113380 0.4331756
## 16 0.03012230 0.3815675
## 17 0.03129333 0.3842931
## 18 0.03254624 0.4228567
```

```
ggplot(Mranger) #M$results 시각화 size(#Hidden Units) vs RMSE
```




```
# (X) summary(Mranger)
ggplot(varImp(Mranger)) # train(importance='impurity') 지정해야 함
```



```
Mranger$bestTune #alpha=1이므로 lasso가 됨
```

```
##    mtry  splitrule min.node.size
## 13   10 extratrees             1
```

```
Mranger$finalModel #nnet 객체
```

```
## Ranger result
##
## Call:
##  ranger::ranger(dependent.variable.name = ".outcome", data = x,      mtry = min(param$mtry,
ncol(x)), min.node.size = param$min.node.size,      splitrule = as.character(param$splitrule),
write.forest = TRUE,      probability = classProbs, ...)
##
## Type:                             Regression
## Number of trees:                  500
## Sample size:                      225
## Number of independent variables:  14
## Mtry:                             10
## Target node size:                 1
## Variable importance mode:         impurity
## Splitrule:                        extratrees
## Number of random splits:          1
## OOB prediction error (MSE):       14.18224
## R squared (OOB):                  0.8254545
```

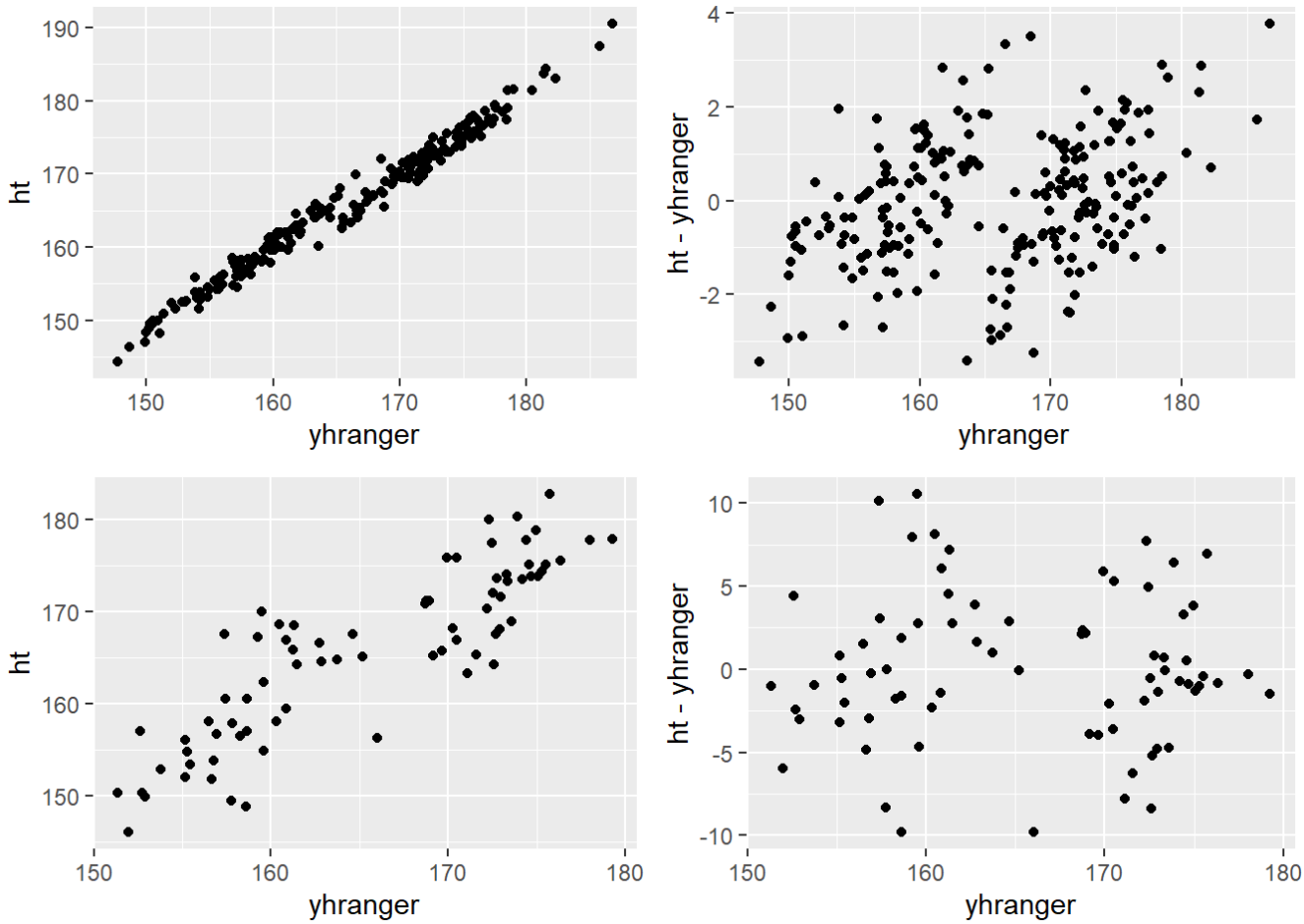
```
Mranger$resample # CV 폴더별 평가측도 densityplot(Mnnet)
```

```
##        RMSE  Rsquared      MAE Resample
## 1  3.079419 0.8864051 2.361155   Fold09
## 2  3.856201 0.8225490 3.041155   Fold07
## 3  3.360090 0.8179441 2.725517   Fold05
## 4  4.030379 0.7993189 3.363617   Fold08
## 5  3.046537 0.8871518 2.540436   Fold03
## 6  4.063592 0.8606841 3.431282   Fold06
## 7  3.828178 0.8362731 2.866708   Fold02
## 8  4.568847 0.8028576 3.700845   Fold01
## 9  3.134879 0.8317018 2.696667   Fold04
## 10 3.772829 0.8667604 3.022278   Fold10
```

```
TROUT <- TROUT %>% mutate(yhranger=predict(Mranger, newdata=TR))
TSOUT <- TSOUT %>% mutate(yhranger=predict(Mranger, newdata=TS))
head(TSOUT)
```

```
##      ht     yhlm   yhstep yhglmnet  yhrpart yhranger
## 1 173.6 174.0110 174.1567 173.9229 173.9105 172.7710
## 2 150.3 151.3521 151.8827 151.3485 150.3583 151.3288
## 3 150.3 152.5093 152.0495 152.7120 150.9667 152.7180
## 4 149.4 155.8523 155.8620 155.9364 156.7000 157.7586
## 5 152.0 156.9321 157.1251 156.8017 150.3583 155.1710
## 6 180.3 176.9327 177.3713 176.8266 172.5900 173.8952
```

```
g1 <- ggplot(TROUT, aes(x=yhranger, y=ht)) + geom_point()
g2 <- ggplot(TROUT, aes(x=yhranger, y=ht-yhranger)) + geom_point()
g3 <- ggplot(TSOUT, aes(x=yhranger, y=ht)) + geom_point()
g4 <- ggplot(TSOUT, aes(x=yhranger, y=ht-yhranger)) + geom_point()
grid.arrange(g1, g2, g3, g4, ncol=2)
```




```
# For REG, yardstick::mae, rmse, rsq
METranger <-
  bind_cols(
    bind_rows(foo(TROUT$ht, TROUT$yhranger), foo(TSOUT$ht, TSOUT$yhranger)),
    data.frame(model='ranger', TRTS=c('TR', 'TS')))
METranger
```

```
## # A tibble: 2 x 5
##    rmse   mae   rsq model  TRTS
##   <dbl> <dbl> <dbl> <chr>  <chr>
## 1  1.38  1.11 0.979 ranger TR
## 2  4.48  3.50 0.756 ranger TS
```

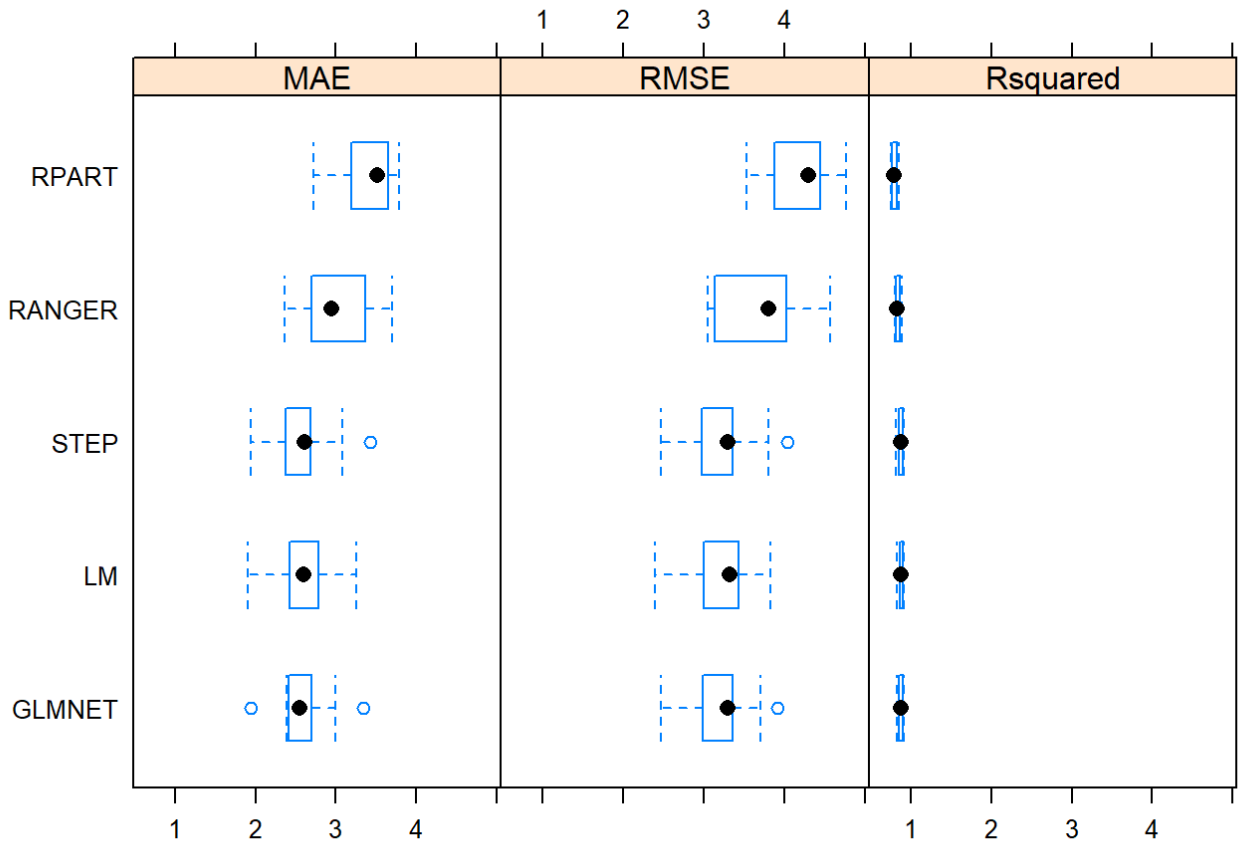
평가

CV 평가

```
RESAMP <- resamples(list(LM=Mlm,
                         STEP=Mstep,
                         GLMNET=Mglmnet,
                         RPART=Mrpart,
                         RANGER=Mranger))
summary(RESAMP)
```

```
##
## Call:
## summary.resamples(object = RESAMP)
##
## Models: LM, STEP, GLMNET, RPART, RANGER
## Number of resamples: 10
##
## MAE
##            Min.  1st Qu.   Median     Mean  3rd Qu.     Max. NA's
## LM     1.898744 2.462826 2.593456 2.607549 2.740707 3.252938    0
## STEP   1.946962 2.436009 2.614459 2.629523 2.674214 3.428224    0
## GLMNET 1.938548 2.440628 2.551851 2.598881 2.668737 3.337803    0
## RPART  2.722178 3.224623 3.515524 3.385144 3.646285 3.790824    0
## RANGER 2.361155 2.703879 2.944493 2.974966 3.283002 3.700845    0
##
## RMSE
##            Min.  1st Qu.   Median     Mean  3rd Qu.     Max. NA's
## LM     2.392532 3.013072 3.323524 3.240378 3.419495 3.828806    0
## STEP   2.473252 2.997106 3.297040 3.263257 3.357510 4.040736    0
## GLMNET 2.468003 2.996185 3.292660 3.231622 3.350014 3.917740    0
## RPART  3.526314 3.944488 4.293976 4.198665 4.441695 4.770831    0
## RANGER 3.046537 3.191182 3.800504 3.674095 3.986835 4.568847    0
##
## Rsquared
##              Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## LM     0.8251774 0.8625616 0.8773306 0.8758550 0.8992620 0.9113453    0
## STEP   0.8111460 0.8528468 0.8778017 0.8739242 0.8992042 0.9124791    0
## GLMNET 0.8240406 0.8607225 0.8783418 0.8764462 0.8997443 0.9117040    0
## RPART  0.7590654 0.7683004 0.7869198 0.7955190 0.8168031 0.8529099    0
## RANGER 0.7993189 0.8190953 0.8339875 0.8411646 0.8652413 0.8871518    0
```

```
bwplot(RESAMP)
```



TS 평가

```
MET <- bind_rows(METlm, METstep, METglmnet, METrpart, METranger)
arrange(MET, TRTS, rmse)
```

```
## # A tibble: 10 x 5
##     rmse   mae   rsq model     TRTS
##    <dbl> <dbl> <dbl> <chr>     <chr>
##  1  1.38  1.11 0.979 ranger    TR
##  2  3.03  2.40 0.887 rpart     TR
##  3  3.07  2.46 0.883 lm        TR
##  4  3.08  2.46 0.883 glmnet    TR
##  5  3.09  2.49 0.882 lmStepAIC TR
##  6  3.96  3.06 0.811 glmnet    TS
##  7  3.99  3.06 0.809 lm        TS
##  8  4.08  3.13 0.802 lmStepAIC TS
##  9  4.48  3.50 0.756 ranger    TS
## 10  4.71  3.58 0.741 rpart     TS
```

```
arrange(MET, TRTS, mae)
```

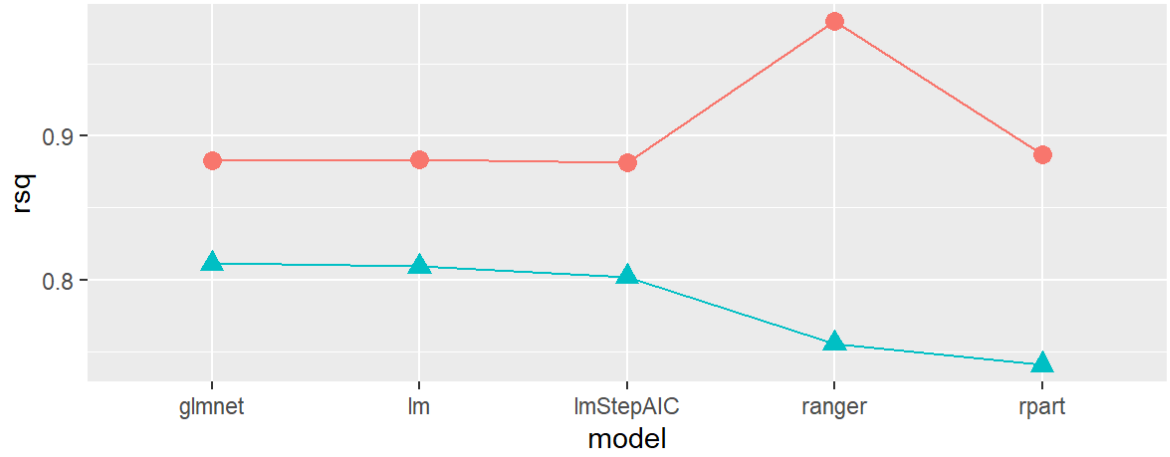
```
## # A tibble: 10 x 5
##     rmse   mae   rsq model     TRTS
##    <dbl> <dbl> <dbl> <chr>     <chr>
##  1  1.38  1.11 0.979 ranger    TR
##  2  3.03  2.40 0.887 rpart     TR
##  3  3.07  2.46 0.883 lm        TR
##  4  3.08  2.46 0.883 glmnet    TR
##  5  3.09  2.49 0.882 lmStepAIC TR
##  6  3.96  3.06 0.811 glmnet    TS
##  7  3.99  3.06 0.809 lm        TS
##  8  4.08  3.13 0.802 lmStepAIC TS
##  9  4.48  3.50 0.756 ranger    TS
## 10  4.71  3.58 0.741 rpart     TS
```

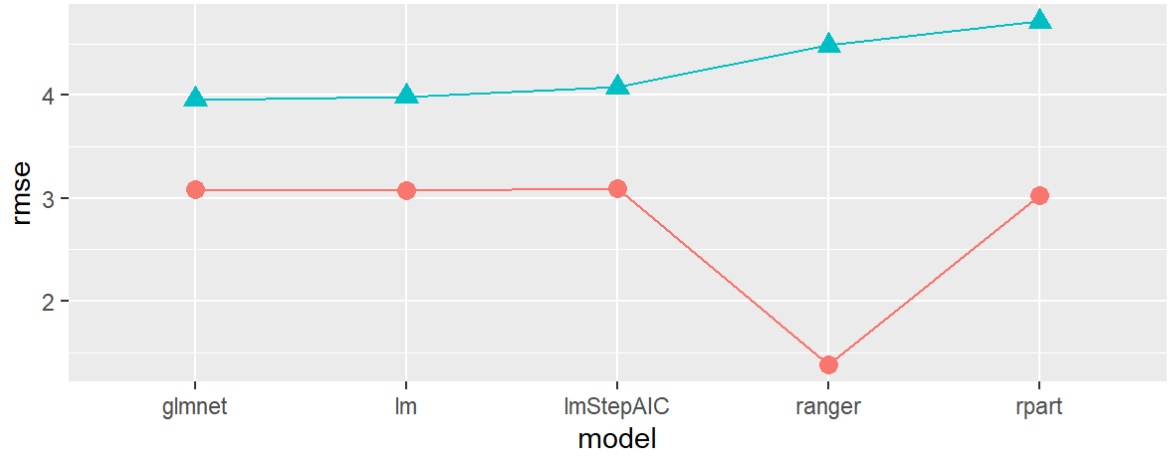
```
arrange(MET, TRTS, desc(rsq))
```

```
## # A tibble: 10 x 5
##     rmse   mae   rsq model     TRTS
##    <dbl> <dbl> <dbl> <chr>     <chr>
##  1  1.38  1.11 0.979 ranger    TR
##  2  3.03  2.40 0.887 rpart     TR
##  3  3.07  2.46 0.883 lm        TR
##  4  3.08  2.46 0.883 glmnet    TR
##  5  3.09  2.49 0.882 lmStepAIC TR
##  6  3.96  3.06 0.811 glmnet    TS
##  7  3.99  3.06 0.809 lm        TS
##  8  4.08  3.13 0.802 lmStepAIC TS
##  9  4.48  3.50 0.756 ranger    TS
## 10  4.71  3.58 0.741 rpart     TS
```

```
g1 <- ggplot(MET, aes(x=model, y=rsq, shape=TRTS, col=TRTS, group=TRTS)) +
  geom_line() +
  geom_point(size=3)
g2 <- ggplot(MET, aes(x=model, y=rmse, shape=TRTS, col=TRTS, group=TRTS)) +
  geom_line() +
  geom_point(size=3)

grid.arrange(g1, g2, nrow=2, ncol=1)
```





실행시간

```
time2 <- Sys.time()
time2
```

```
## [1] "2021-05-09 02:51:39 KST"
```

```
time2 - time1
```

```
## Time difference of 45.26508 secs
```