빨리하고 산책가야지!!

```
library(MASS)
library(caret)
```

```
## Warning: 패키지 'caret'는 R 버전 4.1.3에서 작성되었습니다
```

```
## 필요한 패키지를 로딩중입니다: ggplot2
```

```
## 필요한 패키지를 로딩중입니다: lattice
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
## v purrr   0.3.4
```

```
## Warning: 패키지 'tibble'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'tidyr'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'dplyr'는 R 버전 4.1.3에서 작성되었습니다
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x purrr::lift()   masks caret::lift()
## x dplyr::select() masks MASS::select()
```

```
library(tidymodels)
```

```
## Warning: 패키지 'tidymodels'는 R 버전 4.1.3에서 작성되었습니다
```

```
## -- Attaching packages --------------------------------------- tidymodels 0.2.0 --
```

```
## v broom        0.7.12     v rsample      0.1.1
## v dials        0.1.1      v tune         0.2.0
## v infer        1.0.0      v workflows    0.2.6
## v modeldata    0.1.1      v workflowsets 0.2.1
## v parsnip      0.2.1      v yardstick    0.0.9
## v recipes      0.2.0
```

```
## Warning: 패키지 'broom'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'dials'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'infer'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'modeldata'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'parsnip'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'recipes'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'rsample'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'tune'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'workflows'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'workflowsets'는 R 버전 4.1.3에서 작성되었습니다
```

```
## Warning: 패키지 'yardstick'는 R 버전 4.1.3에서 작성되었습니다
```

```
## -- 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 purrr::lift()            masks caret::lift()
## x yardstick::precision()   masks caret::precision()
## x yardstick::recall()      masks caret::recall()
## x dplyr::select()          masks MASS::select()
## x yardstick::sensitivity() masks caret::sensitivity()
## x yardstick::spec()        masks readr::spec()
## x yardstick::specificity() masks caret::specificity()
## x recipes::step()          masks stats::step()
## * Learn how to get started at https://www.tidymodels.org/start/
```

```
library(skimr)
```

```
## Warning: 패키지 'skimr'는 R 버전 4.1.3에서 작성되었습니다
```

```
library(gridExtra)
```

```
##
## 다음의 패키지를 부착합니다: 'gridExtra'
```

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

#1. 자료설명

```
gss <- as.data.frame(read.csv("C:WWUsersWWUserWWDesktopWWdataWWgss2018trump.csv",stringsAsFactors = T))
str(gss)
```

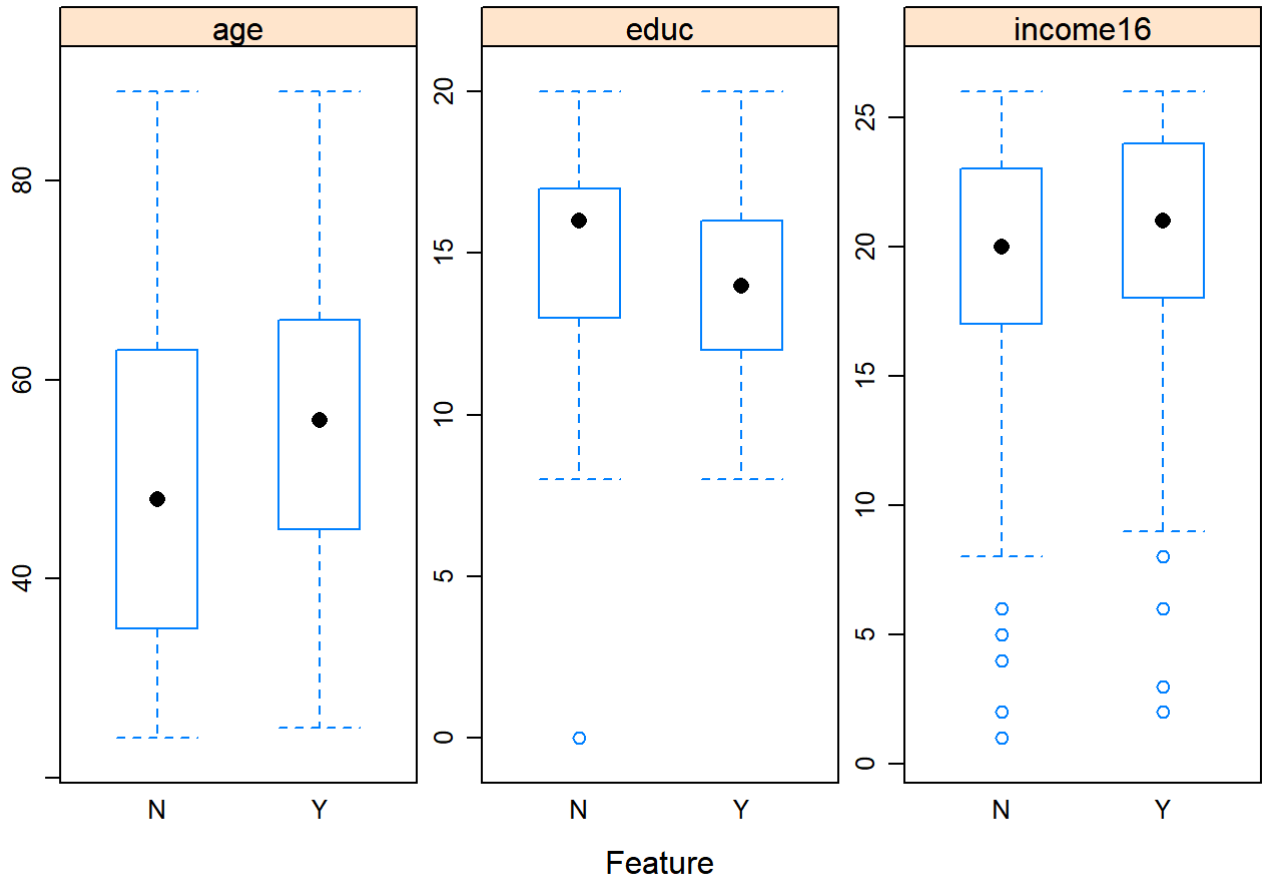
```
## 'data.frame':    646 obs. of  18 variables:
##  $ race    : Factor w/ 3 levels "1W","2B","3O": 1 1 2 1 1 2 1 1 1 1 ...
##  $ class   : Factor w/ 4 levels "1LOWER","2WORKING",..: 2 3 4 1 3 1 3 2 3 3 ...
##  $ age     : int  74 42 71 62 59 41 75 55 40 40 ...
##  $ sex     : Factor w/ 2 levels "1M","2F": 2 1 1 2 1 2 2 2 2 2 ...
##  $ degree  : int  1 3 4 0 4 1 1 3 4 2 ...
##  $ educ    : int  10 16 18 8 19 12 12 16 20 14 ...
##  $ marital : Factor w/ 5 levels "1MARRIED","2WIDOW",..: 4 1 3 2 3 5 2 1 3 1 ...
##  $ attend  : int  2 2 8 0 4 0 7 1 4 1 ...
##  $ income16: int  16 25 26 5 25 2 20 22 21 22 ...
##  $ sei10   : num  14.8 83.4 68.6 21.8 77.4 21.6 73.9 84.2 82.5 84.2 ...
##  $ madeg   : int  0 3 1 1 1 3 1 1 4 3 ...
##  $ padeg   : int  0 1 1 1 1 1 1 2 1 3 ...
##  $ maeduc  : int  8 16 12 12 12 16 12 13 19 16 ...
##  $ paeduc  : int  0 12 12 12 12 15 12 14 15 16 ...
##  $ masei10 : num  13.2 35.8 21.8 25.7 84.2 38.2 13.3 19.6 82.5 61.4 ...
##  $ pasei10 : num  24.6 77.4 58.4 46.6 52 32 41 65.1 39.7 59.1 ...
##  $ prestg10: int  22 61 53 35 72 28 53 64 64 64 ...
##  $ trump   : Factor w/ 2 levels "N","Y": 2 2 2 1 2 1 1 1 1 1 ...
```

```
gss <- gss %>% mutate(degree = factor(degree,labels=c('HS-','HS','COLLEGE','BA','GRAD')),
                      attend = factor(attend, labels= c('NVR','<1/YR', '1/YR', 'SVRL/YR', '1/MO', '2-3X/MO', 'NRLY EVRYWK',
'EVRYWK', 'MORE THN ONCE WK')), madeg = factor(madeg,labels=c('HS-','HS','COLLEGE','BA','GRAD')),padeg = factor(padeg,labels=c('HS-','HS','COLLEGE','BA','GRAD')))
gss$age <- as.numeric(gss$age)
gss$educ <- as.numeric(gss$educ)
gss$income16 <- as.numeric(gss$income16)
gss$maeduc <- as.numeric(gss$maeduc)
gss$paeduc <- as.numeric(gss$paeduc)
str(na.omit(gss))
```

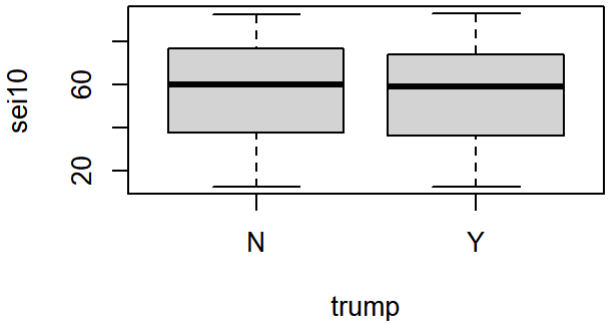
```
## 'data.frame':    646 obs. of  18 variables:
##  $ race    : Factor w/ 3 levels "1W","2B","3O": 1 1 2 1 1 2 1 1 1 1 ...
##  $ class   : Factor w/ 4 levels "1LOWER","2WORKING",..: 2 3 4 1 3 1 3 2 3 3 ...
##  $ age     : num  74 42 71 62 59 41 75 55 40 40 ...
##  $ sex     : Factor w/ 2 levels "1M","2F": 2 1 1 2 1 2 2 2 2 2 ...
##  $ degree  : Factor w/ 5 levels "HS-","HS","COLLEGE",..: 2 4 5 1 5 2 2 4 5 3 ...
##  $ educ    : num  10 16 18 8 19 12 12 16 20 14 ...
##  $ marital : Factor w/ 5 levels "1MARRIED","2WIDOW",..: 4 1 3 2 3 5 2 1 3 1 ...
##  $ attend  : Factor w/ 9 levels "NVR","<1/YR",..: 3 3 9 1 5 1 8 2 5 2 ...
##  $ income16: num  16 25 26 5 25 2 20 22 21 22 ...
##  $ sei10   : num  14.8 83.4 68.6 21.8 77.4 21.6 73.9 84.2 82.5 84.2 ...
##  $ madeg   : Factor w/ 5 levels "HS-","HS","COLLEGE",..: 1 4 2 2 2 4 2 2 5 4 ...
##  $ padeg   : Factor w/ 5 levels "HS-","HS","COLLEGE",..: 1 2 2 2 2 2 2 3 2 4 ...
##  $ maeduc  : num  8 16 12 12 12 16 12 13 19 16 ...
##  $ paeduc  : num  0 12 12 12 12 15 12 14 15 16 ...
##  $ masei10 : num  13.2 35.8 21.8 25.7 84.2 38.2 13.3 19.6 82.5 61.4 ...
##  $ pasei10 : num  24.6 77.4 58.4 46.6 52 32 41 65.1 39.7 59.1 ...
##  $ prestg10: int  22 61 53 35 72 28 53 64 64 64 ...
##  $ trump   : Factor w/ 2 levels "N","Y": 2 2 2 1 2 1 1 1 1 1 ...
```

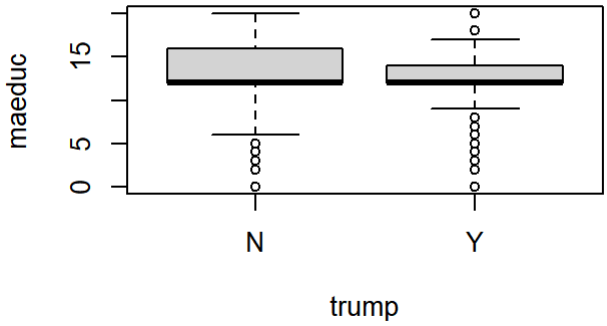
#2. 간단탐색

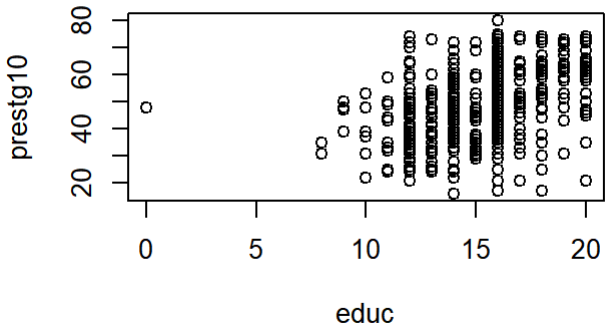
```
featurePlot(x=gss %>% dplyr::select(age,educ,income16), y= gss$trump, plot= 'box',
scales = list(x=list(relation='free'), y = list(relation='free')))
```

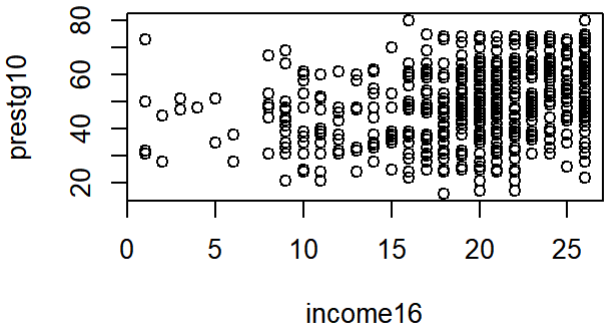


```
par(mfrow=c(2,2))
boxplot(sei10~trump, data=gss)
boxplot(maeduc~trump, data=gss)
plot(prestg10~educ, data=gss)
plot(prestg10~income16, data=gss)
```

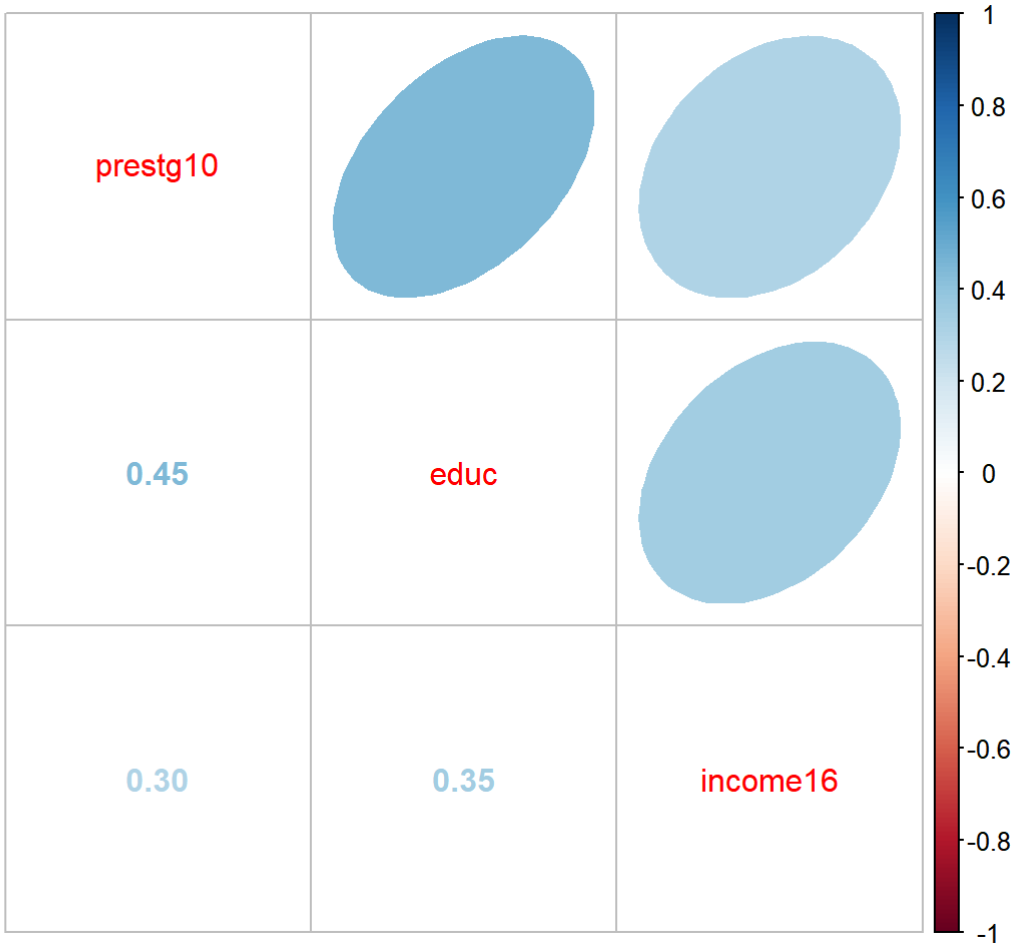









```
R <- cor(gss %>% dplyr::select(prestg10, educ, income16))
corrplot::corrplot.mixed(R, upper='ellipse')
```



#3. 분할

```
set.seed(20180178)
ISP <- initial_split(gss, prop=2/3)
TR <- training(ISP)
TS <- testing(ISP)
rbind(dim(TR), dim(TS))
```

```
##      [,1] [,2]
## [1,]  430   18
## [2,]  216   18
```

#4. 모형

```
Mglm <- glm(trump ~. , data=TR, family=binomial)
summary(Mglm)
```

```
##
## Call:
## glm(formula = trump ~ ., family = binomial, data = TR)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3893  -0.7822  -0.2381   0.8337   2.3455
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -0.512397   1.673824  -0.306 0.759511
## race2B                 -4.130665   1.117538  -3.696 0.000219 ***
## race30                 -0.886403   0.607455  -1.459 0.144508
## class2WORKING          -0.151934   0.582418  -0.261 0.794195
## class3MID              -0.651457   0.621388  -1.048 0.294459
## class4UP               -0.079528   0.887676  -0.090 0.928612
## age                     0.024836   0.010420   2.384 0.017146 *
## sex2F                  -0.992919   0.272040  -3.650 0.000262 ***
## degreeHS               -0.289056   0.907707  -0.318 0.750146
## degreeCOLLEGE          -0.283346   1.060137  -0.267 0.789259
## degreeBA               -0.631264   1.132998  -0.557 0.577416
## degreeGRAD             -0.604180   1.286529  -0.470 0.638626
## educ                   -0.087315   0.096132  -0.908 0.363731
## marital2WIDOW          -1.773013   0.646553  -2.742 0.006102 **
## marital3DIV            -0.266269   0.353375  -0.754 0.451148
## marital4SEP             0.168536   0.887807   0.190 0.849439
## marital5NEVER          -1.187438   0.452828  -2.622 0.008735 **
## attend<1/YR             0.257343   0.531656   0.484 0.628357
## attend1/YR              0.422758   0.428288   0.987 0.323599
## attendSVRL/YR           0.793885   0.444875   1.785 0.074341 .
## attend1/MO              0.918596   0.571140   1.608 0.107758
## attend2-3X/MO          -0.348135   0.636919  -0.547 0.584659
## attendNRLY EVRYWK       1.167892   0.658446   1.774 0.076111 .
## attendEVRYWK            1.354085   0.377556   3.586 0.000335 ***
## attendMORE THN ONCE WK  2.296570   0.827808   2.774 0.005532 **
## income16                0.008651   0.033838   0.256 0.798218
## sei10                   0.012224   0.011332   1.079 0.280726
## madegHS                 0.360921   0.566754   0.637 0.524240
## madegCOLLEGE            0.074260   0.847069   0.088 0.930141
## madegBA                -0.309994   0.915870  -0.338 0.735010
## madegGRAD              -1.824786   1.255601  -1.453 0.146136
## padegHS                -0.301658   0.525521  -0.574 0.565957
## padegCOLLEGE           -1.193386   0.850844  -1.403 0.160739
## padegBA                -1.723350   0.892934  -1.930 0.053609 .
## padegGRAD              -1.368743   1.019643  -1.342 0.179474
## maeduc                  0.075737   0.097145   0.780 0.435612
## paeduc                  0.067472   0.092801   0.727 0.467190
## masei10                 0.008393   0.006771   1.240 0.215159
## pasei10                 0.001939   0.007505   0.258 0.796117
## prestg10               -0.026418   0.018635  -1.418 0.156279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 568.69  on 429  degrees of freedom
## Residual deviance: 404.81  on 390  degrees of freedom
## AIC: 484.81
##
## Number of Fisher Scoring iterations: 7
```

##Mlgm 로지스틱 회귀식 ##yh <- -0.51-4.13race2B-0.88race3O-0.15class2WORKING-0.65class3MID-0.07class4UP+0.02age-0.99sex2F-0.28degreeHS-0.28degreeCOLLEGE-0.63degreeBA-0.6degreeGRAD-0.08educ-1.77marital2WIDOW-0.26marital3DIV+0.16marital4SEP-1.18marital5NEVER+0.25attend<1/YR+0.42attend1/YR+0.79attendSVRL/YR+0.91attend1/MO-0.34attend2-3X/MO+1.16attendNRLY EVRYWK+1.35attendEVRYWK+2.29attendMORE THN ONCE WK~~-0.02prestg10 ##사후확률 예측식 yh = -4.13race2B+0.02age-0.99sex2F-1.77marital2WIDOW-1.18marital5NEVER+1.35attendEVRYWK+2.29attendMORE THN ONCE WK

```
#4.2Mstep
Mstep <- MASS::stepAIC(Mglm, direction ="backward")
```

```
## Start:  AIC=484.81
## trump ~ race + class + age + sex + degree + educ + marital +
##     attend + income16 + sei10 + madeg + padeg + maeduc + paeduc +
##     masei10 + pasei10 + prestg10
##
##            Df Deviance    AIC
## - degree    4   405.42 477.42
## - class     3   408.08 482.08
## - income16  1   404.88 482.88
## - pasei10   1   404.88 482.88
## - paeduc    1   405.34 483.34
## - maeduc    1   405.43 483.43
## - educ      1   405.64 483.64
## - padeg     4   411.75 483.75
## - sei10     1   405.99 483.99
## - masei10   1   406.36 484.36
## <none>          404.81 484.81
## - prestg10  1   406.84 484.84
## - madeg     4   413.54 485.54
## - age       1   410.60 488.60
## - marital   4   419.54 491.54
## - attend    8   428.68 492.68
## - sex       1   418.64 496.64
## - race      2   443.78 519.78
##
## Step:  AIC=477.42
## trump ~ race + class + age + sex + educ + marital + attend +
##     income16 + sei10 + madeg + padeg + maeduc + paeduc + masei10 +
##     pasei10 + prestg10
##
##            Df Deviance    AIC
## - class     3   409.16 475.16
## - income16  1   405.45 475.45
## - pasei10   1   405.46 475.46
## - paeduc    1   406.04 476.04
## - maeduc    1   406.08 476.08
## - sei10     1   406.35 476.35
## - padeg     4   412.48 476.48
## - masei10   1   407.10 477.10
## - prestg10  1   407.32 477.32
## <none>          405.42 477.42
## - madeg     4   415.01 479.01
## - educ      1   410.66 480.66
## - age       1   411.44 481.44
## - marital   4   420.26 484.26
## - attend    8   429.01 485.01
## - sex       1   419.75 489.75
## - race      2   444.47 512.47
##
## Step:  AIC=475.16
## trump ~ race + age + sex + educ + marital + attend + income16 +
##     sei10 + madeg + padeg + maeduc + paeduc + masei10 + pasei10 +
##     prestg10
##
##            Df Deviance    AIC
```

```
## - income16  1   409.19 473.19
## - pasei10   1   409.19 473.19
## - paeduc    1   409.71 473.71
## - sei10     1   409.81 473.81
## - padeg     4   415.84 473.84
## - maeduc    1   410.06 474.06
## - masei10   1   410.34 474.34
## - prestg10  1   410.72 474.72
## <none>          409.16 475.16
## - age       1   413.72 477.72
## - madeg     4   419.78 477.78
## - educ      1   415.52 479.52
## - marital   4   422.85 480.85
## - attend    8   433.14 483.14
## - sex       1   422.32 486.32
## - race      2   446.70 508.70
##
## Step:  AIC=473.19
## trump ~ race + age + sex + educ + marital + attend + sei10 +
##     madeg + padeg + maeduc + paeduc + masei10 + pasei10 + prestg10
##
##            Df Deviance    AIC
## - pasei10   1   409.22 471.22
## - paeduc    1   409.73 471.73
## - sei10     1   409.81 471.81
## - padeg     4   415.84 471.84
## - maeduc    1   410.08 472.08
## - masei10   1   410.38 472.38
## - prestg10  1   410.72 472.72
## <none>          409.19 473.19
## - age       1   413.76 475.76
## - madeg     4   419.79 475.79
## - educ      1   415.92 477.92
## - marital   4   423.78 479.78
## - attend    8   433.16 481.16
## - sex       1   422.33 484.33
## - race      2   446.70 506.70
##
## Step:  AIC=471.22
## trump ~ race + age + sex + educ + marital + attend + sei10 +
##     madeg + padeg + maeduc + paeduc + masei10 + prestg10
##
##            Df Deviance    AIC
## - paeduc    1   409.81 469.81
## - sei10     1   409.83 469.83
## - padeg     4   415.89 469.89
## - maeduc    1   410.09 470.09
## - masei10   1   410.46 470.46
## - prestg10  1   410.73 470.73
## <none>          409.22 471.22
## - madeg     4   419.81 473.81
## - age       1   413.83 473.83
## - educ      1   415.94 475.94
## - marital   4   423.88 477.88
## - attend    8   433.38 479.38
## - sex       1   422.33 482.33
```

```
## - race      2   446.83 504.83
##
## Step:  AIC=469.81
## trump ~ race + age + sex + educ + marital + attend + sei10 +
##     madeg + padeg + maeduc + masei10 + prestg10
##
##            Df Deviance    AIC
## - sei10     1   410.39 468.39
## - masei10   1   411.03 469.03
## - prestg10  1   411.21 469.21
## - maeduc    1   411.66 469.66
## <none>          409.81 469.81
## - padeg     4   418.10 470.10
## - age       1   414.24 472.24
## - madeg     4   422.23 474.23
## - educ      1   416.28 474.28
## - marital   4   424.33 476.33
## - attend    8   433.77 477.77
## - sex       1   423.07 481.07
## - race      2   448.04 504.04
##
## Step:  AIC=468.39
## trump ~ race + age + sex + educ + marital + attend + madeg +
##     padeg + maeduc + masei10 + prestg10
##
##            Df Deviance    AIC
## - prestg10  1   411.28 467.28
## - masei10   1   411.65 467.65
## - maeduc    1   412.04 468.04
## <none>          410.39 468.39
## - padeg     4   418.57 468.57
## - age       1   415.09 471.09
## - educ      1   416.36 472.36
## - madeg     4   422.37 472.37
## - marital   4   425.07 475.07
## - attend    8   434.72 476.72
## - sex       1   423.42 479.42
## - race      2   448.40 502.40
##
## Step:  AIC=467.28
## trump ~ race + age + sex + educ + marital + attend + madeg +
##     padeg + maeduc + masei10
##
##           Df Deviance    AIC
## - masei10  1   412.29 466.29
## - maeduc   1   413.02 467.02
## <none>         411.28 467.28
## - padeg    4   419.68 467.68
## - age      1   415.66 469.66
## - madeg    4   423.29 471.29
## - marital  4   425.33 473.33
## - educ     1   419.80 473.80
## - attend   8   435.80 475.80
## - sex      1   423.77 477.77
## - race     2   449.47 501.47
##
```

```
## Step:  AIC=466.29
## trump ~ race + age + sex + educ + marital + attend + madeg +
##     padeg + maeduc
##
##           Df Deviance    AIC
## - padeg    4   420.04 466.04
## - maeduc   1   414.12 466.12
## <none>         412.29 466.29
## - age      1   416.52 468.52
## - madeg    4   423.32 469.32
## - marital  4   426.34 472.34
## - educ     1   420.40 472.40
## - attend   8   435.81 473.81
## - sex      1   424.63 476.63
## - race     2   450.63 500.63
##
## Step:  AIC=466.04
## trump ~ race + age + sex + educ + marital + attend + madeg +
##     maeduc
##
##           Df Deviance    AIC
## - maeduc   1   421.69 465.69
## <none>         420.04 466.04
## - age      1   426.18 470.18
## - marital  4   433.61 471.61
## - madeg    4   434.06 472.06
## - attend   8   442.10 472.10
## - educ     1   430.57 474.57
## - sex      1   431.99 475.99
## - race     2   456.46 498.46
##
## Step:  AIC=465.69
## trump ~ race + age + sex + educ + marital + attend + madeg
##
##           Df Deviance    AIC
## <none>         421.69 465.69
## - age      1   427.47 469.47
## - madeg    4   434.13 470.13
## - attend   8   443.02 471.02
## - marital  4   435.67 471.67
## - educ     1   431.17 473.17
## - sex      1   432.85 474.85
## - race     2   458.36 498.36
```

```
##제외된 변수  class,degree,income16,sei10,padeg,maeduc, paeduc ,masei10 , pasei10 ,prestg10
```

##제외된 변수 class,degree,income16,sei10,padeg,maeduc, paeduc ,masei10 , pasei10 ,prestg10 #5 TR,TS에서 평가

TR 평가

```
#5.1 TR 평가 Mglm
#TR Mglm
MTROUT <- TR %>% mutate(phMglm = predict(Mglm,newdata=TR,type='response'),yhglm=factor(ifelse(p
hMglm >0.5,'Y','N')))
confusionMatrix(MTROUT$trump, MTROUT$yhglm,mode='everything')
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   N   Y
##          N 227  42
##          Y  59 102
##
##                Accuracy : 0.7651
##                  95% CI : (0.7221, 0.8044)
##     No Information Rate : 0.6651
##     P-Value [Acc > NIR] : 3.946e-06
##
##                   Kappa : 0.4877
##
##  Mcnemar's Test P-Value : 0.1114
##
##             Sensitivity : 0.7937
##             Specificity : 0.7083
##          Pos Pred Value : 0.8439
##          Neg Pred Value : 0.6335
##               Precision : 0.8439
##                  Recall : 0.7937
##                      F1 : 0.8180
##              Prevalence : 0.6651
##          Detection Rate : 0.5279
##    Detection Prevalence : 0.6256
##       Balanced Accuracy : 0.7510
##
##        'Positive' Class : N
##
```

```
#AUC-ROC
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## 다음의 패키지를 부착합니다: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
MTRoc <- roc(MTROUT$trump, MTROUT$phMglm)
```

```
## Setting levels: control = N, case = Y
```

```
## Setting direction: controls < cases
```

```
auc(MTRoc)
```

```
## Area under the curve: 0.8396
```

```
roc_auc_vec(MTROUT$trump, MTROUT$phMglm,event_level='second')
```

```
## [1] 0.8395715
```

```
#AUC-PR
pr_auc_vec(MTROUT$trump, MTROUT$phMglm,event_level='second')
```

```
## [1] 0.7275009
```

```
#youden
coords(MTRoc,x='best',best.method ='youden')
```

```
##   threshold specificity sensitivity
## 1 0.3962478   0.7657993   0.8012422
```

```
#5.1 TR 평가 Mstep
TROUT <- TR %>% mutate(phMstep = predict(Mstep,newdata=TR,type='response'),yhglm=factor(ifelse
(phMstep >0.5,'Y','N')))
confusionMatrix(TROUT$trump, TROUT$yhglm,mode='everything')
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   N   Y
##          N 224  45
##          Y  62  99
##
##                Accuracy : 0.7512
##                  95% CI : (0.7075, 0.7913)
##     No Information Rate : 0.6651
##     P-Value [Acc > NIR] : 6.744e-05
##
##                   Kappa : 0.4573
##
##  Mcnemar's Test P-Value : 0.1219
##
##             Sensitivity : 0.7832
##             Specificity : 0.6875
##          Pos Pred Value : 0.8327
##          Neg Pred Value : 0.6149
##               Precision : 0.8327
##                  Recall : 0.7832
##                      F1 : 0.8072
##              Prevalence : 0.6651
##          Detection Rate : 0.5209
##    Detection Prevalence : 0.6256
##       Balanced Accuracy : 0.7354
##
##        'Positive' Class : N
##
```

```
#TR Mstep
#AUC-ROC
library(pROC)
TRoc <- roc(TROUT$trump, TROUT$phMstep)
```

```
## Setting levels: control = N, case = Y
```

```
## Setting direction: controls < cases
```

```
auc(TRoc)
```

```
## Area under the curve: 0.8213
```

```
roc_auc_vec(TROUT$trump, TROUT$phMstep,event_level='second')
```

```
## [1] 0.8213073
```

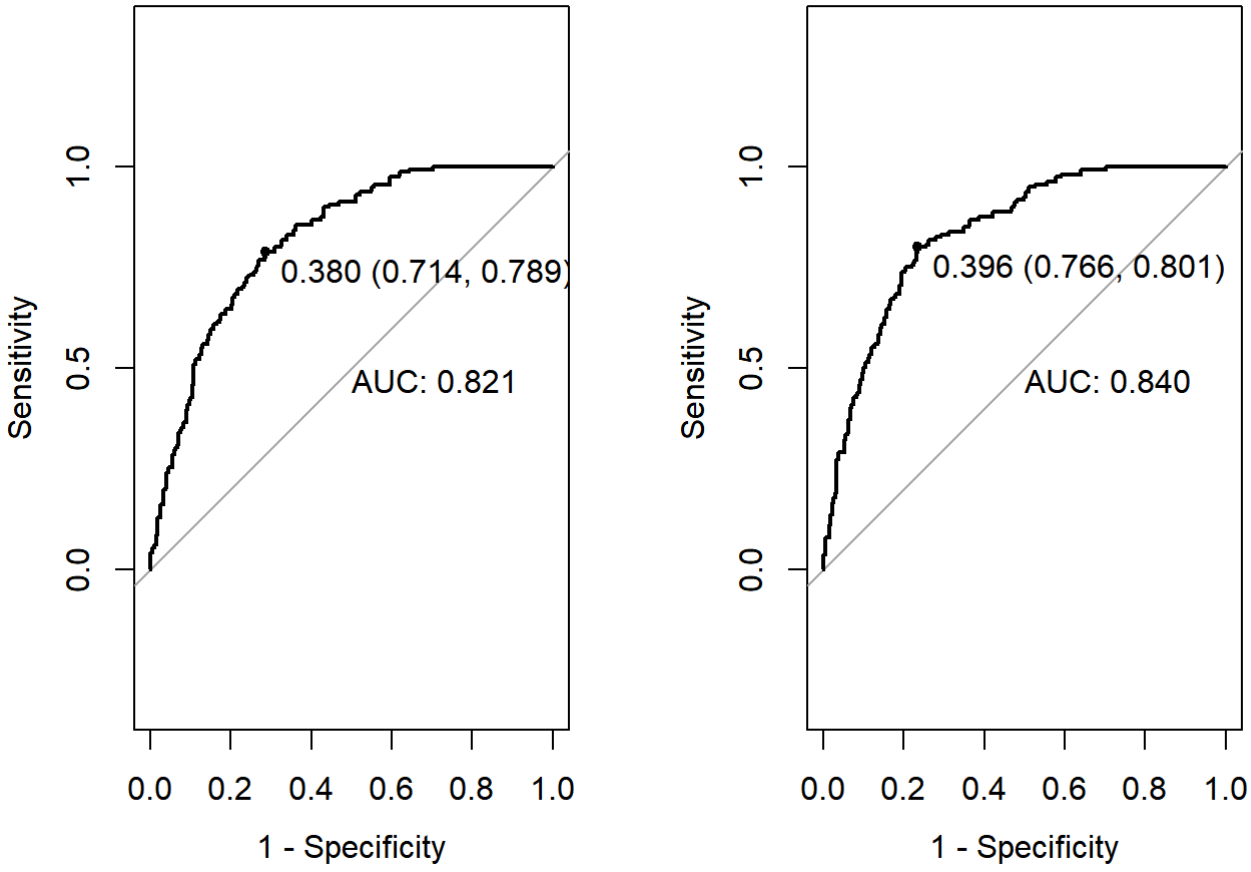
```
#AUC-PR
pr_auc_vec(TROUT$trump, TROUT$phMstep,event_level='second')
```

```
## [1] 0.6954614
```

```
#youden
coords(TRoc,x='best',best.method ='youden')
```

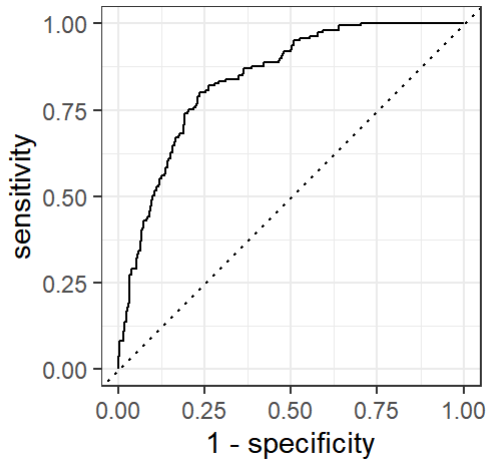
```
##   threshold specificity sensitivity
## 1 0.3804743   0.7137546   0.7888199
```

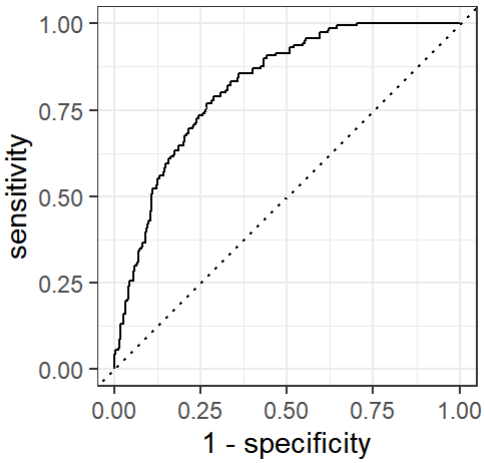
```
#Mglm, Mstep ROC 곡선
par(mfrow=c(1,2))
plot(TRoc,legacy=T,print.auc=T,print.thres='best',print.thres.best.method='youden')
plot(MTRoc,legacy=T,print.auc=T,print.thres='best',print.thres.best.method='youden')
```

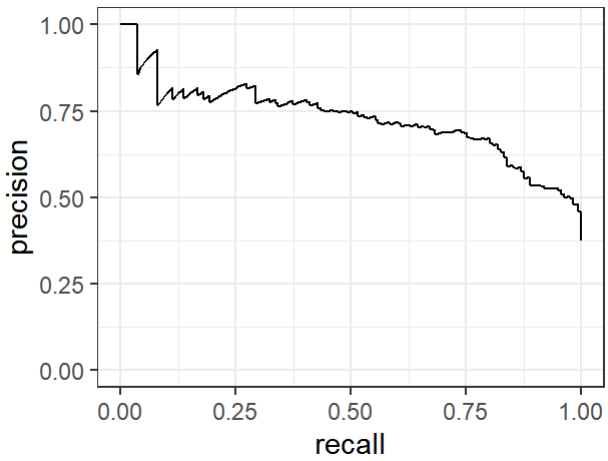


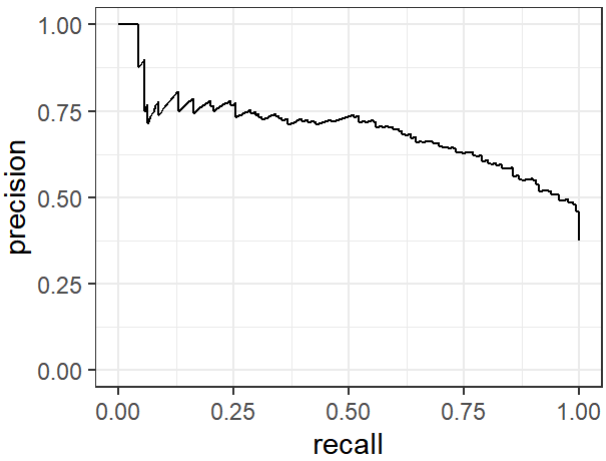

```
#Mglm, Mstep ROC, PR 곡선
g1 <- autoplot(roc_curve(MTROUT, 'trump', 'phMglm', event_level='second'))
g2 <- autoplot(roc_curve(TROUT, 'trump', 'phMstep', event_level='second'))
g3 <- autoplot( pr_curve(MTROUT, 'trump', 'phMglm', event_level='second'))
g4 <- autoplot( pr_curve(TROUT, 'trump', 'phMstep', event_level='second'))

grid.arrange(g1, g2, g3, g4, ncol=2)
```









TS 평가

```
#5.2 TS 평가 Mglm
#TS Mglm
MTSOUT <- TS %>% mutate(phMglm = predict(Mglm,newdata=TS,type='response'),yhglm=factor(ifelse(p
hMglm >0.5,'Y','N')))
confusionMatrix(MTSOUT$trump, MTSOUT$yhglm,mode='everything')
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   N   Y
##          N 105  28
##          Y  37  46
##
##                Accuracy : 0.6991
##                  95% CI : (0.6332, 0.7594)
##     No Information Rate : 0.6574
##     P-Value [Acc > NIR] : 0.1107
##
##                   Kappa : 0.3508
##
##  Mcnemar's Test P-Value : 0.3211
##
##             Sensitivity : 0.7394
##             Specificity : 0.6216
##          Pos Pred Value : 0.7895
##          Neg Pred Value : 0.5542
##               Precision : 0.7895
##                  Recall : 0.7394
##                      F1 : 0.7636
##              Prevalence : 0.6574
##          Detection Rate : 0.4861
##    Detection Prevalence : 0.6157
##       Balanced Accuracy : 0.6805
##
##        'Positive' Class : N
##
```

```
#AUC-ROC
library(pROC)
MTSroc <- roc(MTSOUT$trump, MTSOUT$phMglm)
```

```
## Setting levels: control = N, case = Y
```

```
## Setting direction: controls < cases
```

```
auc(MTSroc)
```

```
## Area under the curve: 0.761
```

```
roc_auc_vec(MTSOUT$trump, MTSOUT$phMglm,event_level='second')
```

```
## [1] 0.7610291
```

```
#AUC-PR
pr_auc_vec(MTSOUT$trump, MTSOUT$phMglm,event_level='second')
```

```
## [1] 0.6370204
```

```
#youden
coords(MTSroc,x='best',best.method ='youden')
```

```
##   threshold specificity sensitivity
## 1 0.2050371    0.556391   0.8795181
```

```
#5.2 TS 평가 Mstep
TSOUT <- TS %>% mutate(phMstep = predict(Mstep,newdata=TS,type='response'),yhglm=factor(ifelse
(phMstep >0.5,'Y','N')))
confusionMatrix(TSOUT$trump, TSOUT$yhglm,mode='everything')
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   N   Y
##          N 105  28
##          Y  40  43
##
##                Accuracy : 0.6852
##                  95% CI : (0.6187, 0.7465)
##     No Information Rate : 0.6713
##     P-Value [Acc > NIR] : 0.3614
##
##                   Kappa : 0.3161
##
##  Mcnemar's Test P-Value : 0.1822
##
##             Sensitivity : 0.7241
##             Specificity : 0.6056
##          Pos Pred Value : 0.7895
##          Neg Pred Value : 0.5181
##               Precision : 0.7895
##                  Recall : 0.7241
##                      F1 : 0.7554
##              Prevalence : 0.6713
##          Detection Rate : 0.4861
##    Detection Prevalence : 0.6157
##       Balanced Accuracy : 0.6649
##
##        'Positive' Class : N
##
```

```
#TS Mstep
#AUC-ROC
library(pROC)
TSroc <- roc(TSOUT$trump, TSOUT$phMstep)
```

```
## Setting levels: control = N, case = Y
```

```
## Setting direction: controls < cases
```

```
auc(TSroc)
```

```
## Area under the curve: 0.7701
```

```
roc_auc_vec(TSOUT$trump, TSOUT$phMstep,event_level='second')
```

```
## [1] 0.7700879
```

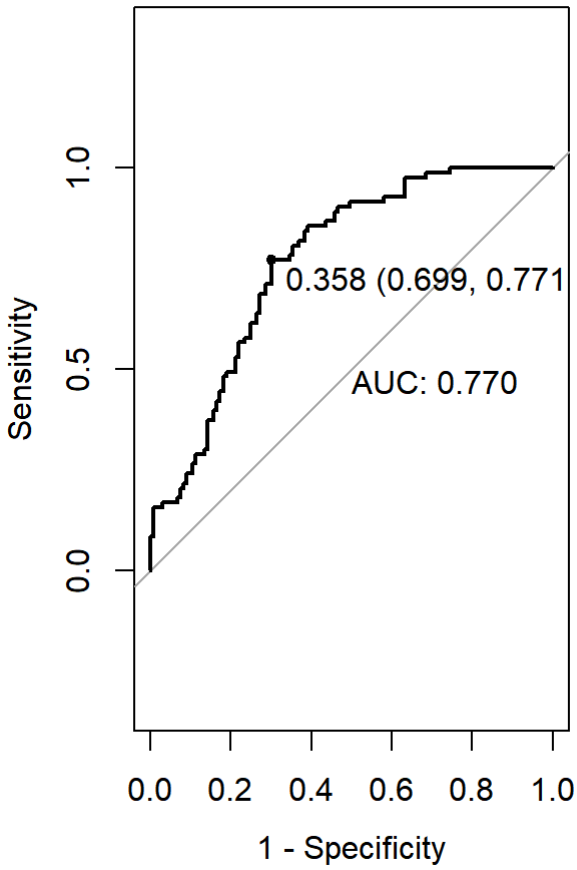
```
#AUC-PR
pr_auc_vec(TSOUT$trump, TSOUT$phMstep,event_level='second')
```

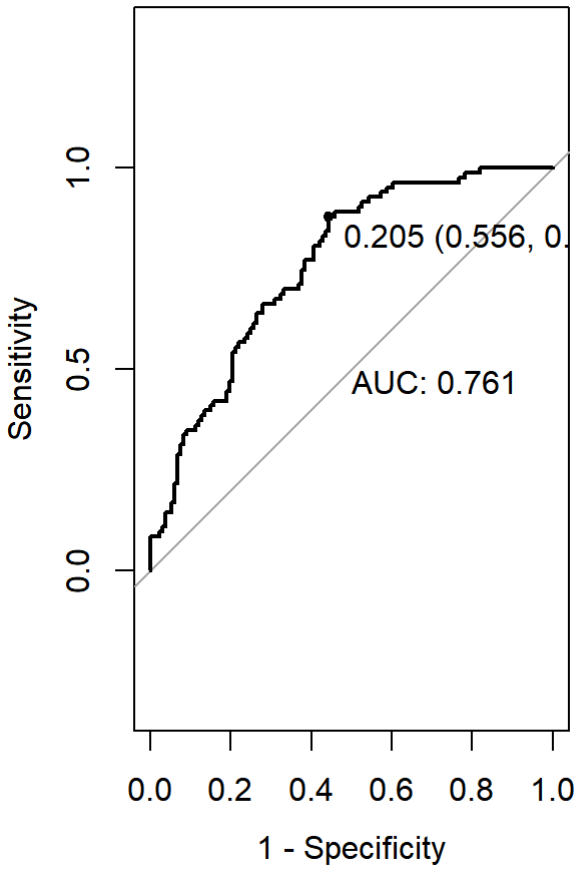
```
## [1] 0.6457725
```

```
#youden
coords(TSroc,x='best',best.method ='youden')
```

```
##   threshold specificity sensitivity
## 1 0.3578687   0.6992481   0.7710843
```

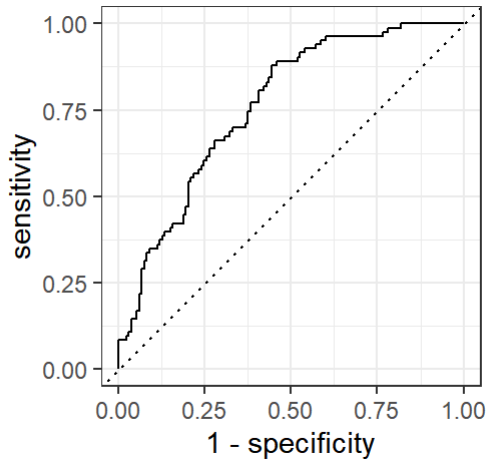
```
#Mglm, Mstep ROC 곡선
par(mfrow=c(1,2))
plot(TSroc,legacy=T,print.auc=T,print.thres='best',print.thres.best.method='youden')
plot(MTSroc,legacy=T,print.auc=T,print.thres='best',print.thres.best.method='youden')
```

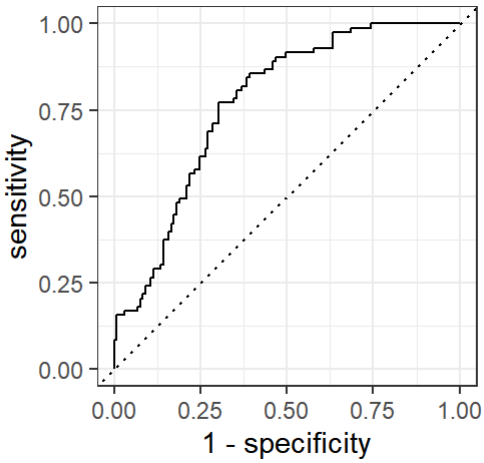


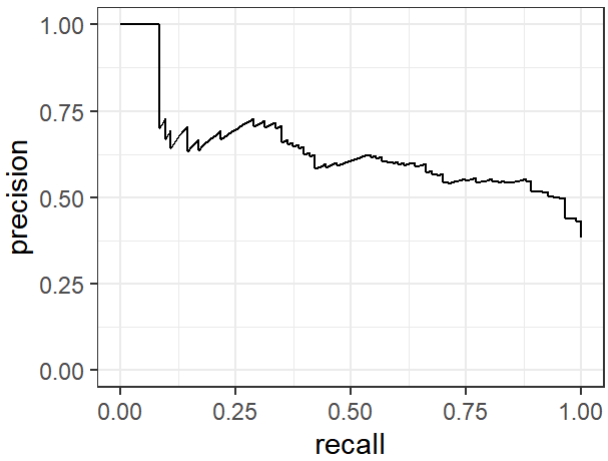


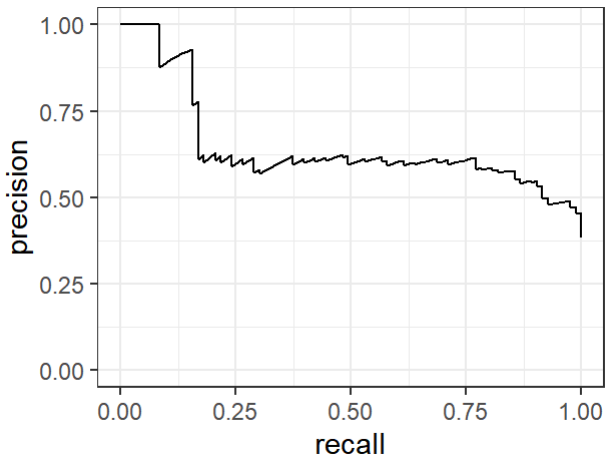

```
#Mglm, Mstep ROC, PR 곡선
g1 <- autoplot(roc_curve(MTSOUT, 'trump', 'phMglm', event_level='second'))
g2 <- autoplot(roc_curve(TSOUT, 'trump', 'phMstep', event_level='second'))
g3 <- autoplot( pr_curve(MTSOUT, 'trump', 'phMglm', event_level='second'))
g4 <- autoplot( pr_curve(TSOUT, 'trump', 'phMstep', event_level='second'))

grid.arrange(g1, g2, g3, g4, ncol=2)
```









데이터	모형	ACC	F1	AUC-ROC	AUC-PR
TR	Mglm	0.7651	0.8180	0.8395	0.7275
	Mstep	0.7512	0.8072	0.8213	0.6954
TS	Mglm	0.6991	0.7636	0.7610	0.6370
	Mstep	0.6852	0.7554	0.77	0.64577