Eric_Hirsch_621_Assignment_3

Eric Hirsch

4/7/2022

Contents

1. Data Exploration 1
2. Data Preparation 7

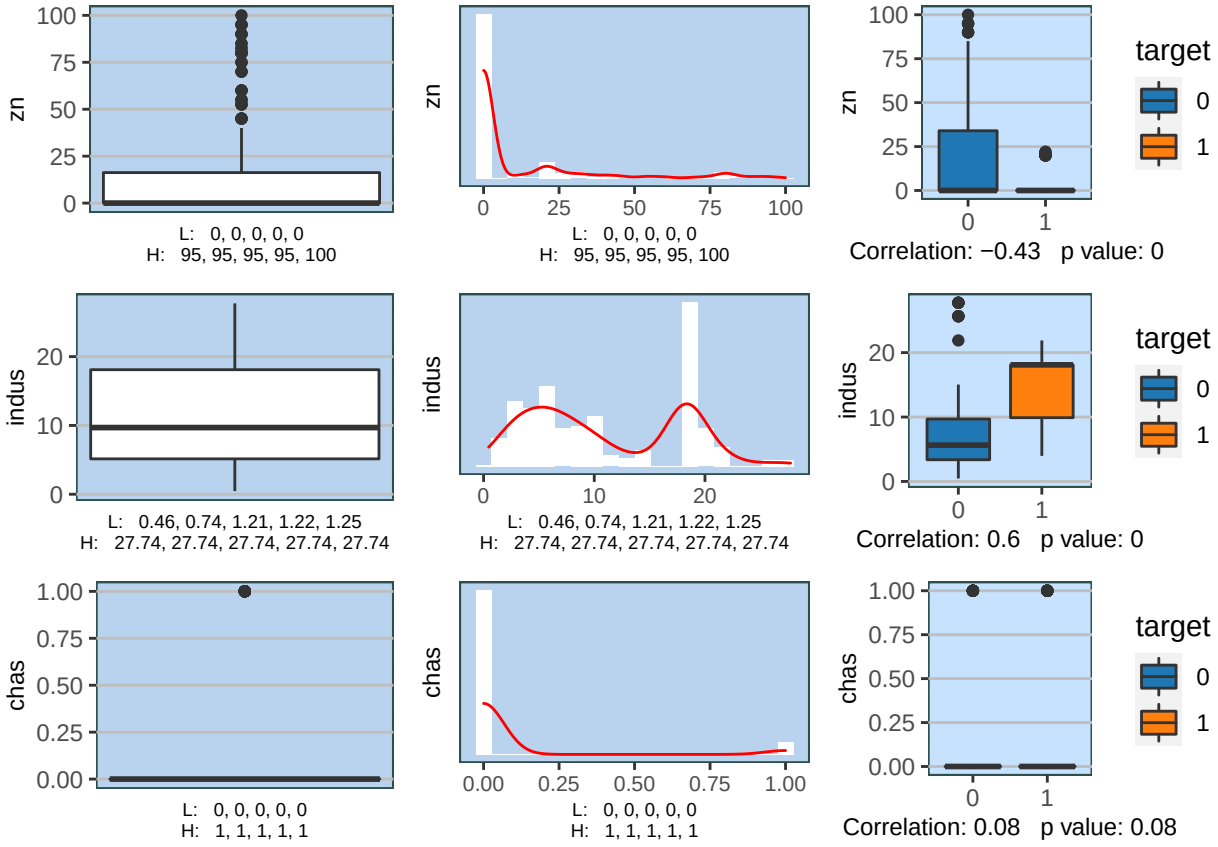
1. Data Exploration

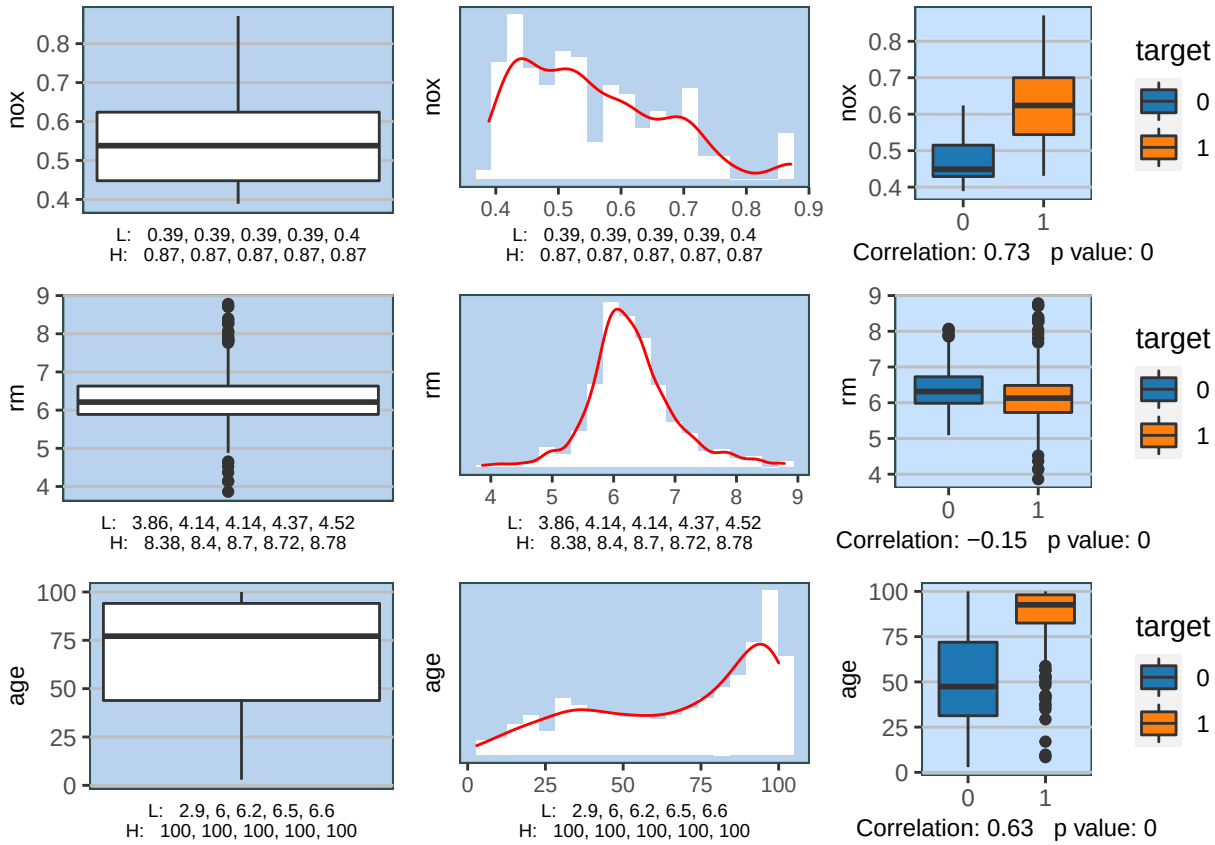
```
##        zn             indus             chas              nox
##  Min.   :  0.00   Min.   : 0.460   Min.   :0.00000   Min.   :0.3890
##  1st Qu.:  0.00   1st Qu.: 5.145   1st Qu.:0.00000   1st Qu.:0.4480
##  Median :  0.00   Median : 9.690   Median :0.00000   Median :0.5380
##  Mean   : 11.58   Mean   :11.105   Mean   :0.07082   Mean   :0.5543
##  3rd Qu.: 16.25   3rd Qu.:18.100   3rd Qu.:0.00000   3rd Qu.:0.6240
##  Max.   :100.00   Max.   :27.740   Max.   :1.00000   Max.   :0.8710
##        rm              age              dis              rad
##  Min.   :3.863   Min.   :  2.90   Min.   : 1.130   Min.   : 1.00
##  1st Qu.:5.887   1st Qu.: 43.88   1st Qu.: 2.101   1st Qu.: 4.00
##  Median :6.210   Median : 77.15   Median : 3.191   Median : 5.00
##  Mean   :6.291   Mean   : 68.37   Mean   : 3.796   Mean   : 9.53
##  3rd Qu.:6.630   3rd Qu.: 94.10   3rd Qu.: 5.215   3rd Qu.:24.00
##  Max.   :8.780   Max.   :100.00   Max.   :12.127   Max.   :24.00
##       tax           ptratio          lstat             medv
##  Min.   :187.0   Min.   :12.6   Min.   : 1.730   Min.   : 5.00
##  1st Qu.:281.0   1st Qu.:16.9   1st Qu.: 7.043   1st Qu.:17.02
##  Median :334.5   Median :18.9   Median :11.350   Median :21.20
##  Mean   :409.5   Mean   :18.4   Mean   :12.631   Mean   :22.59
##  3rd Qu.:666.0   3rd Qu.:20.2   3rd Qu.:16.930   3rd Qu.:25.00
##  Max.   :711.0   Max.   :22.0   Max.   :37.970   Max.   :50.00
##      target
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.4914
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

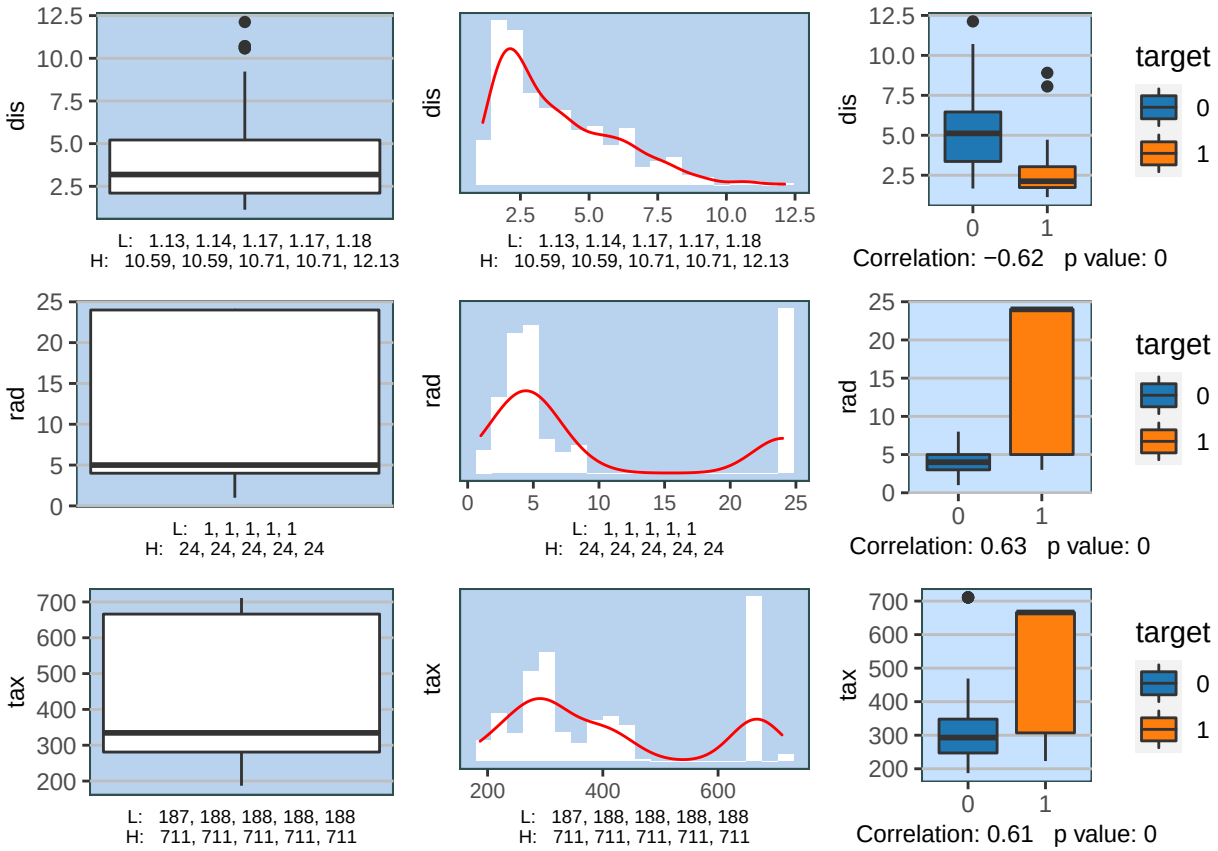
```
## 'data.frame':    466 obs. of  13 variables:
##  $ zn     : num  0 0 0 30 0 0 0 0 0 80 ...
##  $ indus  : num  19.58 19.58 18.1 4.93 2.46 ...
```

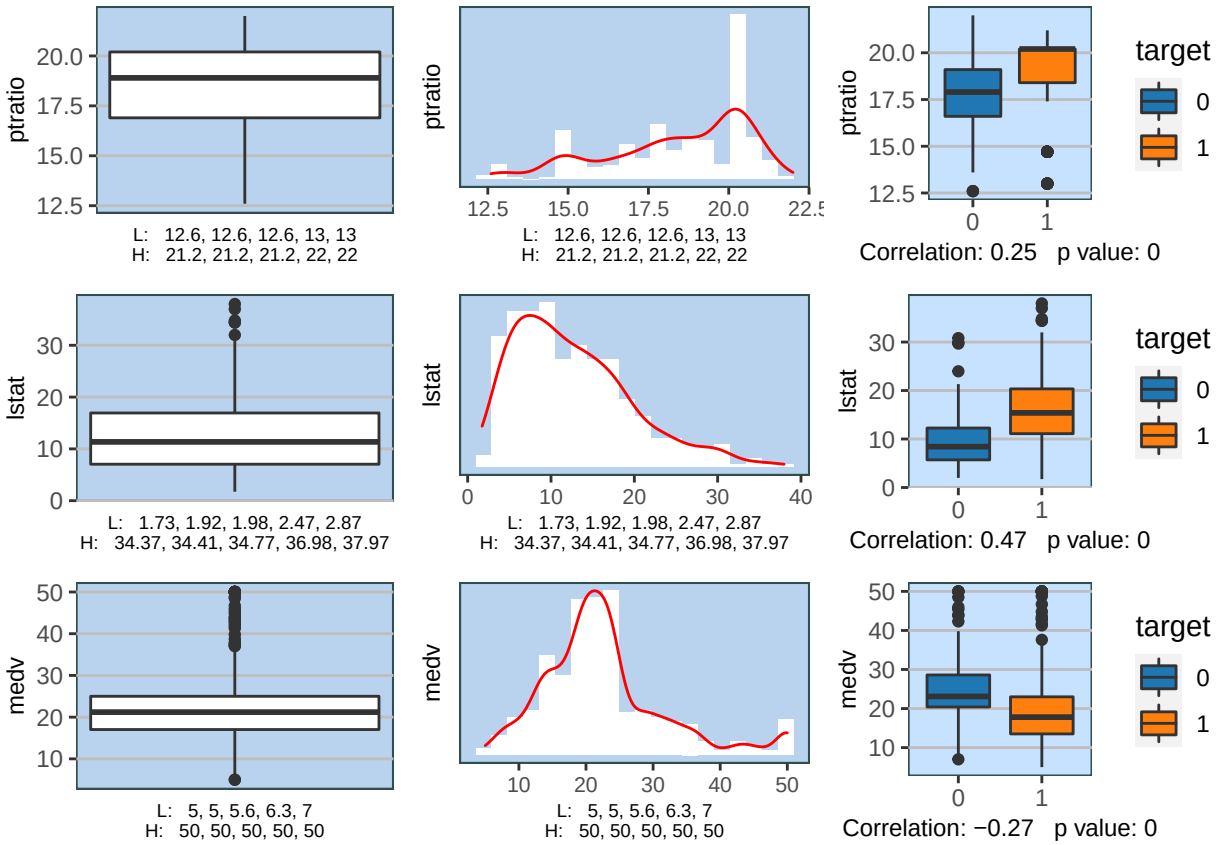
```
##  $ chas   : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ nox    : num  0.605 0.871 0.74 0.428 0.488 0.52 0.693 0.693 0.515 0.392 ...
##  $ rm     : num  7.93 5.4 6.49 6.39 7.16 ...
##  $ age    : num  96.2 100 100 7.8 92.2 71.3 100 100 38.1 19.1 ...
##  $ dis    : num  2.05 1.32 1.98 7.04 2.7 ...
##  $ rad    : int  5 5 24 6 3 5 24 24 5 1 ...
##  $ tax    : int  403 403 666 300 193 384 666 666 224 315 ...
##  $ ptratio: num  14.7 14.7 20.2 16.6 17.8 20.9 20.2 20.2 20.2 16.4 ...
##  $ lstat  : num  3.7 26.82 18.85 5.19 4.82 ...
##  $ medv   : num  50 13.4 15.4 23.7 37.9 26.5 5 7 22.2 20.9 ...
##  $ target : int  1 1 1 0 0 0 1 1 0 0 ...
```

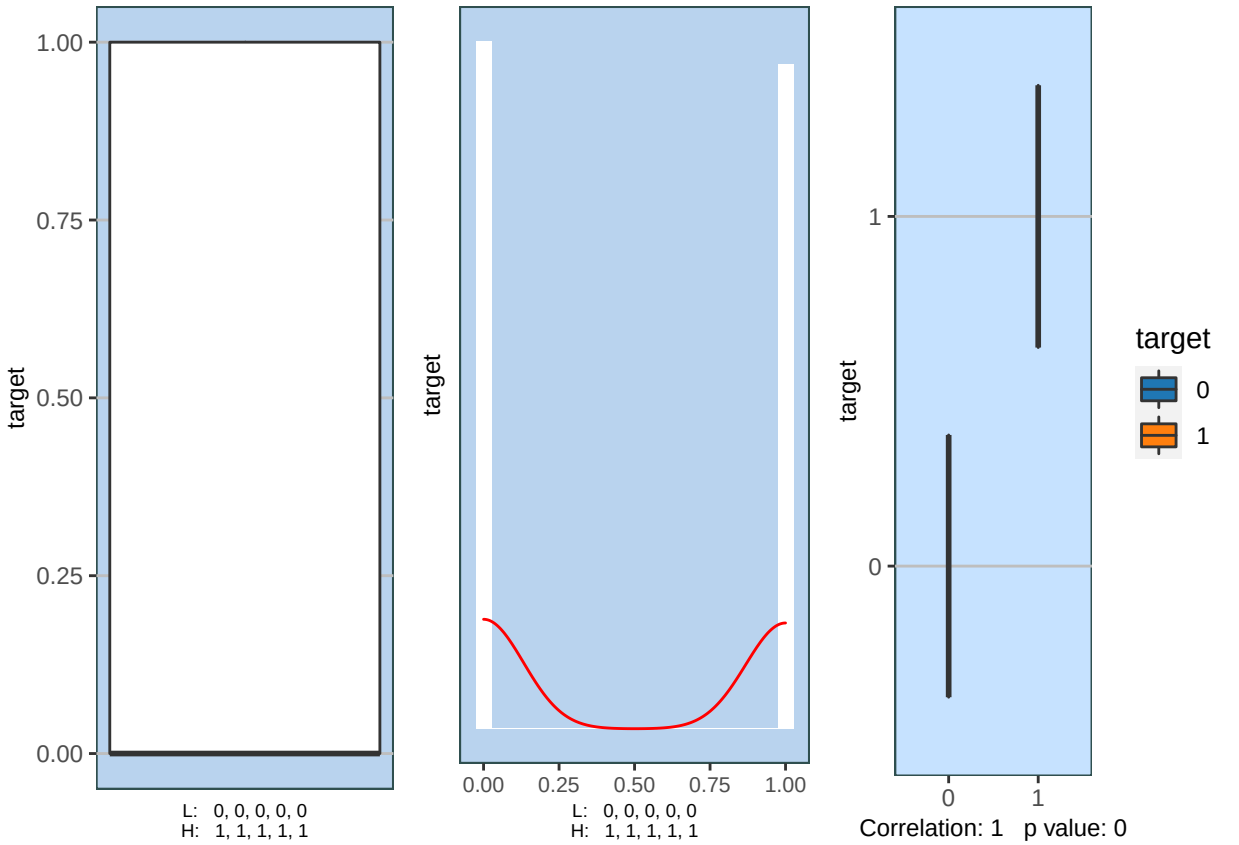
A. Summary Statistics The data consists of 466 observations and 13 variables, all numeric. Two are binary, including the target. There are no missing values. The target appears to be relatively balanced (which makes sense, as it is an indicator of being above or below the median.)




nox
0.8
0.7
0.6
0.5
0.4
L: 0.39, 0.39, 0.39, 0.39, 0.4
H: 0.87, 0.87, 0.87, 0.87, 0.87
nox
0.4 0.5 0.6 0.7 0.8 0.9
L: 0.39, 0.39, 0.39, 0.39, 0.4
H: 0.87, 0.87, 0.87, 0.87, 0.87
nox
0.8
0.7
0.6
0.5
0.4
0 1
target
0
1
Correlation: 0.73 p value: 0
rm
9
8
7
6
5
4
L: 3.86, 4.14, 4.14, 4.37, 4.52
H: 8.38, 8.4, 8.7, 8.72, 8.78
rm
4 5 6 7 8 9
L: 3.86, 4.14, 4.14, 4.37, 4.52
H: 8.38, 8.4, 8.7, 8.72, 8.78
rm
9
8
7
6
5
4
0 1
target
0
1
Correlation: −0.15 p value: 0
age
100
75
50
25
0
L: 2.9, 6, 6.2, 6.5, 6.6
H: 100, 100, 100, 100, 100
age
0 25 50 75 100
L: 2.9, 6, 6.2, 6.5, 6.6
H: 100, 100, 100, 100, 100
age
100
75
50
25
0
0 1
target
0
1
Correlation: 0.63 p value: 0

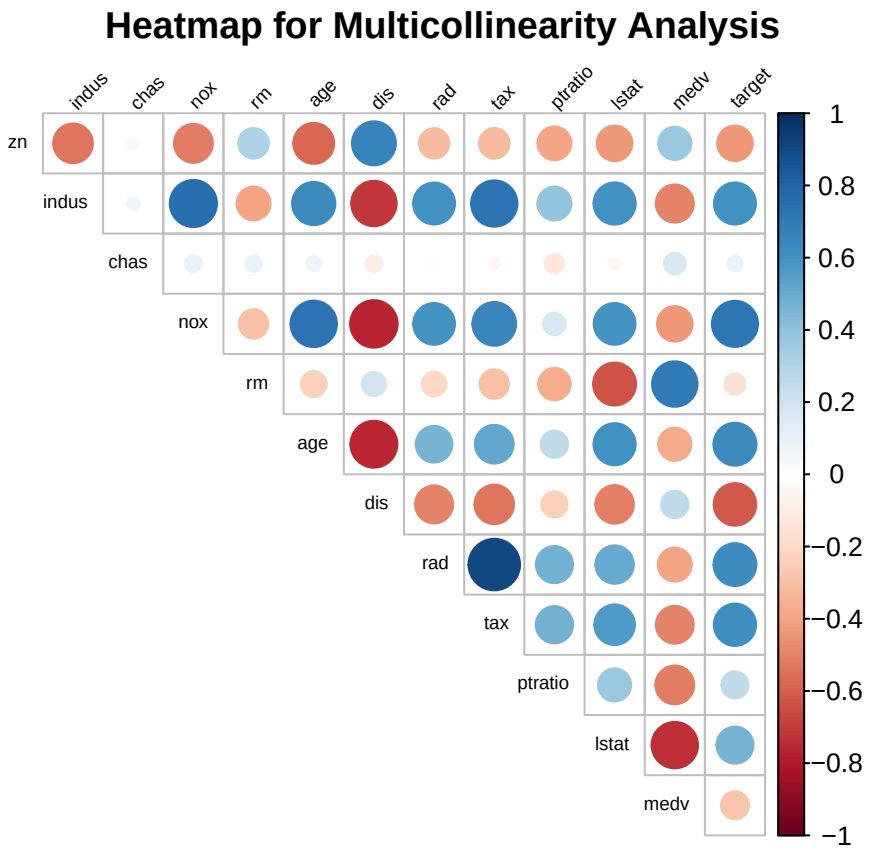

dis
L: 1.13, 1.14, 1.17, 1.17, 1.18
H: 10.59, 10.59, 10.71, 10.71, 12.13
dis
L: 1.13, 1.14, 1.17, 1.17, 1.18
H: 10.59, 10.59, 10.71, 10.71, 12.13
dis
target
0
1
Correlation: −0.62 p value: 0
rad
L: 1, 1, 1, 1, 1
H: 24, 24, 24, 24, 24
rad
L: 1, 1, 1, 1, 1
H: 24, 24, 24, 24, 24
rad
target
0
1
Correlation: 0.63 p value: 0
tax
L: 187, 188, 188, 188, 188
H: 711, 711, 711, 711, 711
tax
L: 187, 188, 188, 188, 188
H: 711, 711, 711, 711, 711
tax
target
0
1
Correlation: 0.61 p value: 0


ptratio
20.0
17.5
15.0
12.5
L: 12.6, 12.6, 12.6, 13, 13
H: 21.2, 21.2, 21.2, 22, 22
ptratio
12.5
15.0
17.5
20.0
L: 12.6, 12.6, 12.6, 13, 13
H: 21.2, 21.2, 21.2, 22, 22
ptratio
target
0
1
Correlation: 0.25 p value: 0
lstat
30
20
10
0
L: 1.73, 1.92, 1.98, 2.47, 2.87
H: 34.37, 34.41, 34.77, 36.98, 37.97
lstat
0
10
20
30
40
L: 1.73, 1.92, 1.98, 2.47, 2.87
H: 34.37, 34.41, 34.77, 36.98, 37.97
lstat
target
0
1
Correlation: 0.47 p value: 0
medv
50
40
30
20
10
L: 5, 5, 5.6, 6.3, 7
H: 50, 50, 50, 50, 50
medv
10
20
30
40
50
L: 5, 5, 5.6, 6.3, 7
H: 50, 50, 50, 50, 50
medv
target
0
1
Correlation: −0.27 p value: 0



Looking at boxplots, histograms, and boxplots against the target variable, we see some areas of interest. A number of distributions are broken (e.g. zn, indus, nox and rad), suggesting there may be hidden grouping within the variables. For example, zn = 0 may include areas that are different in their makeup from zn>0. In fact, there may be a common phenomenon among all opf them which identifies certain areas as highly industrial, as opposed to the mixed industrial and residential areas for the rest of the observations.

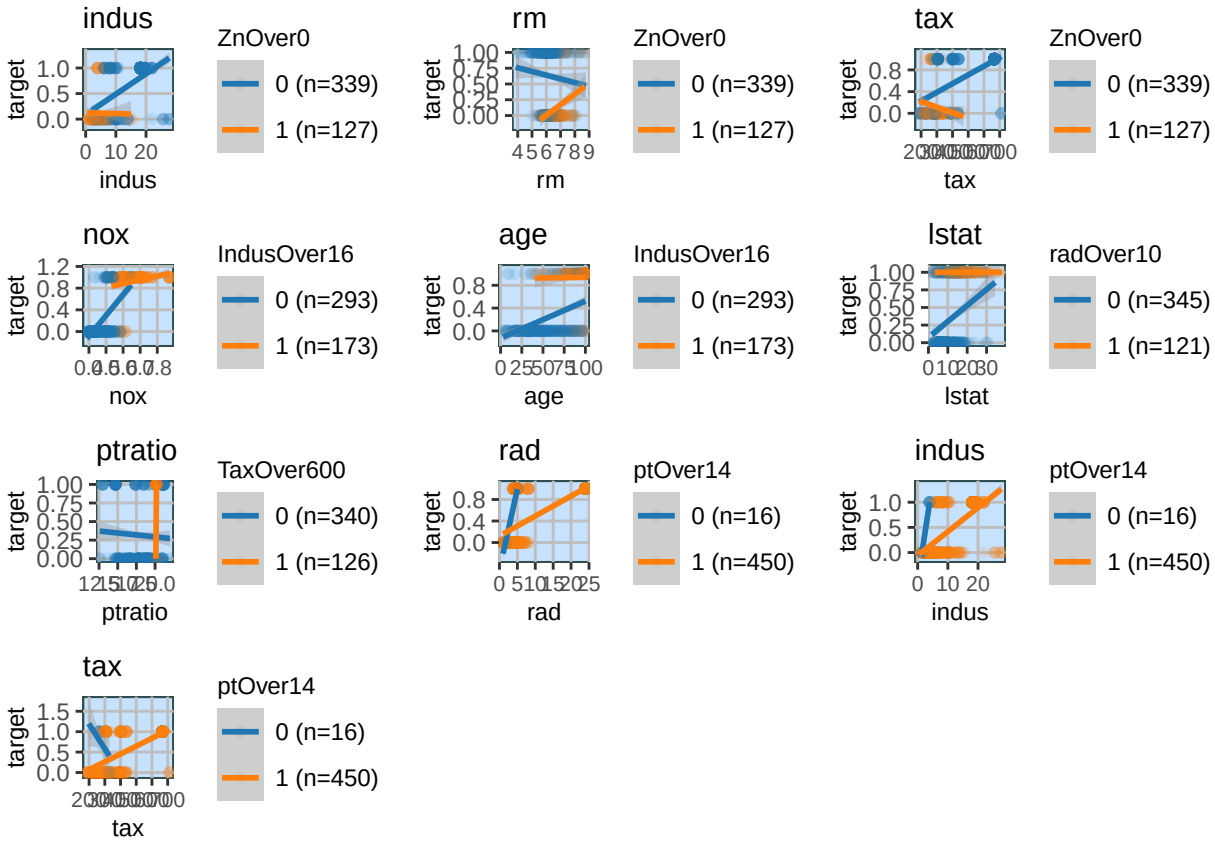
Most of the correlations are unsurprising, with the exception of tax rate, which increases with increase in crime.



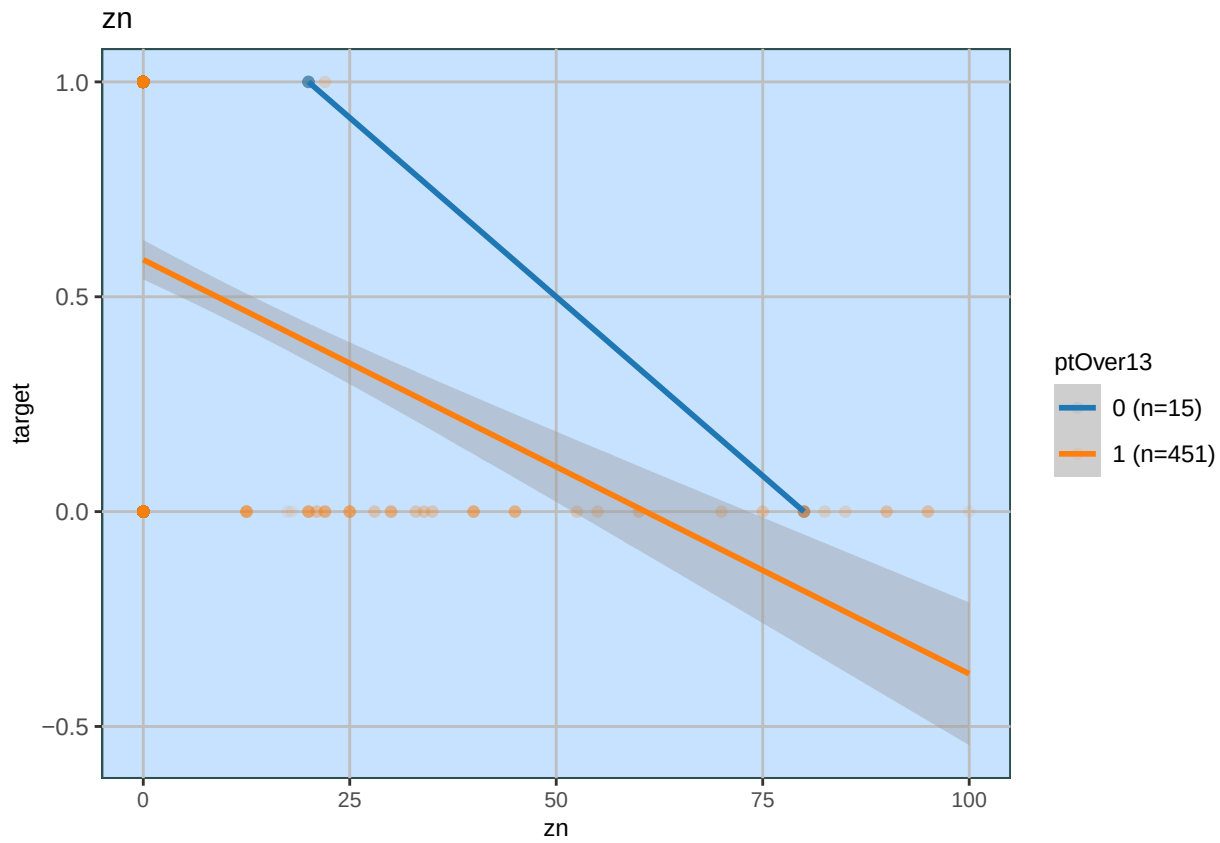
B. Multicollinearity
Mutlicollinearity is highly evident in the database - not surprisingly. The correlation between rad and tax is over 90%. Because rad is slightly more correlated, and has slightly less variation, than tax, tax will be dropped from the analysis.

2. Data Preparation

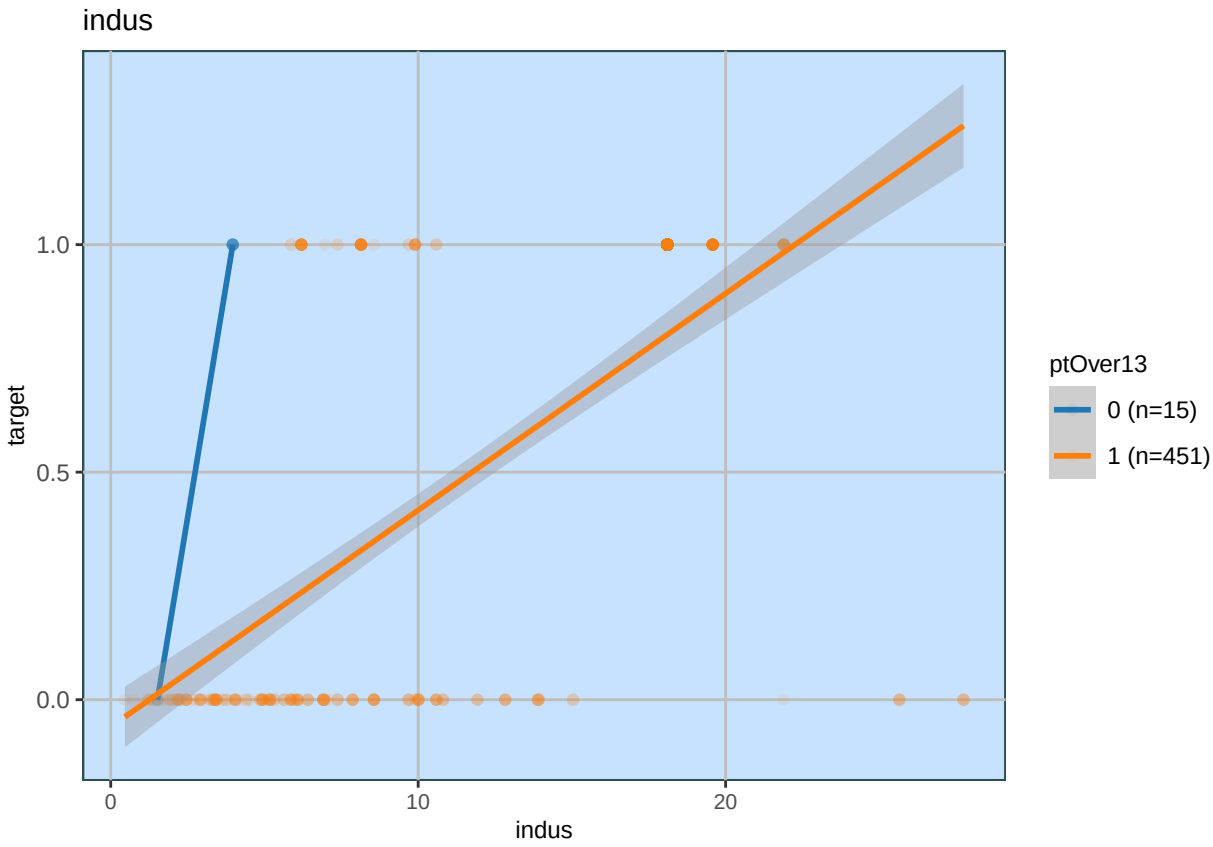
A. Interaction terms The dataset appears to hold the potential for many interaction terms, as many distributions suggest areas of very low indistrialization and very high industrialization, which may affect the slope of other variables. The following are just some of the possible interactions affecting the dataset:



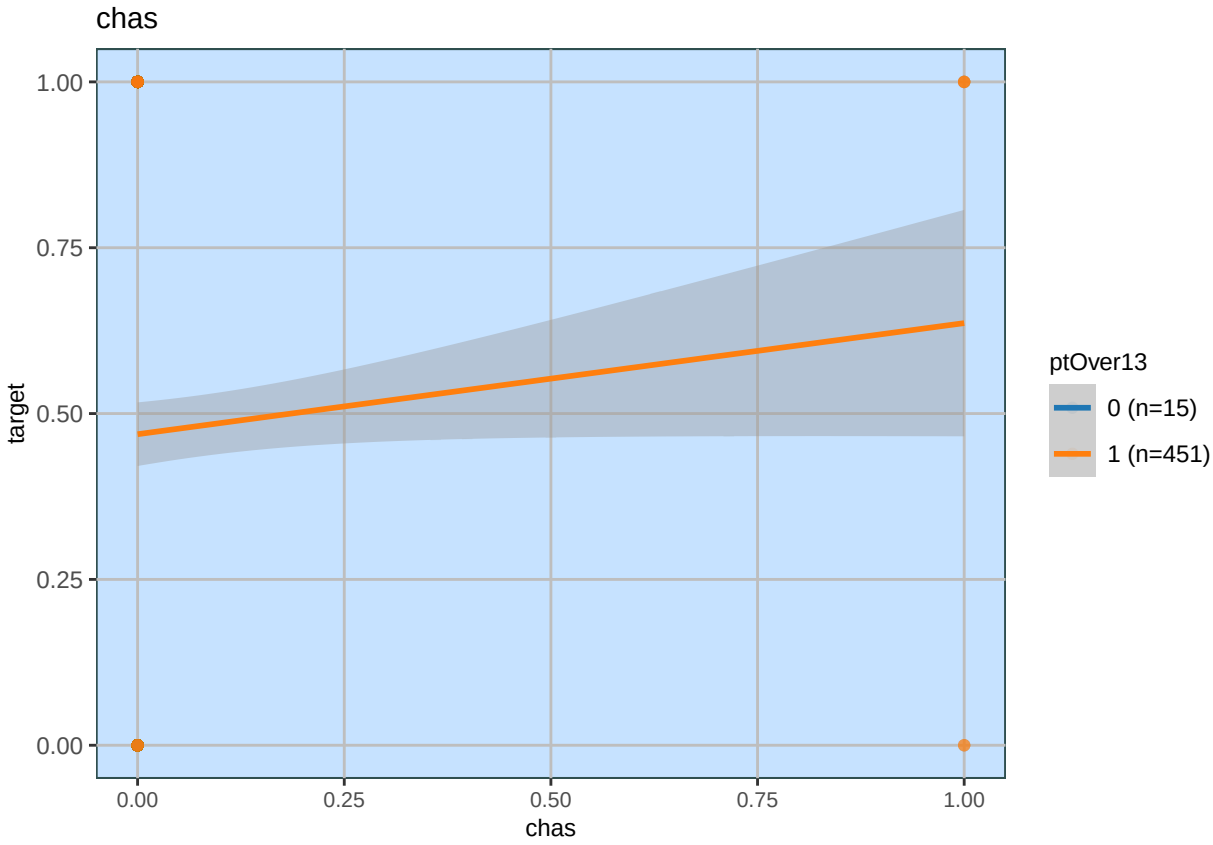

```
## 'data.frame':    466 obs. of  14 variables:
##  $ zn      : num  0 0 0 30 0 0 0 0 0 80 ...
##  $ indus   : num  19.58 19.58 18.1 4.93 2.46 ...
##  $ chas    : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ nox     : num  0.605 0.871 0.74 0.428 0.488 0.52 0.693 0.693 0.515 0.392 ...
##  $ rm      : num  7.93 5.4 6.49 6.39 7.16 ...
##  $ age     : num  96.2 100 100 7.8 92.2 71.3 100 100 38.1 19.1 ...
##  $ dis     : num  2.05 1.32 1.98 7.04 2.7 ...
##  $ rad     : int  5 5 24 6 3 5 24 24 5 1 ...
##  $ tax     : int  403 403 666 300 193 384 666 666 224 315 ...
##  $ ptratio : num  14.7 14.7 20.2 16.6 17.8 20.9 20.2 20.2 20.2 16.4 ...
##  $ lstat   : num  3.7 26.82 18.85 5.19 4.82 ...
##  $ medv    : num  50 13.4 15.4 23.7 37.9 26.5 5 7 22.2 20.9 ...
##  $ target  : int  1 1 1 0 0 0 1 1 0 0 ...
##  $ ptOver13: num  1 1 1 1 1 1 1 1 1 1 ...

## [[1]]
```



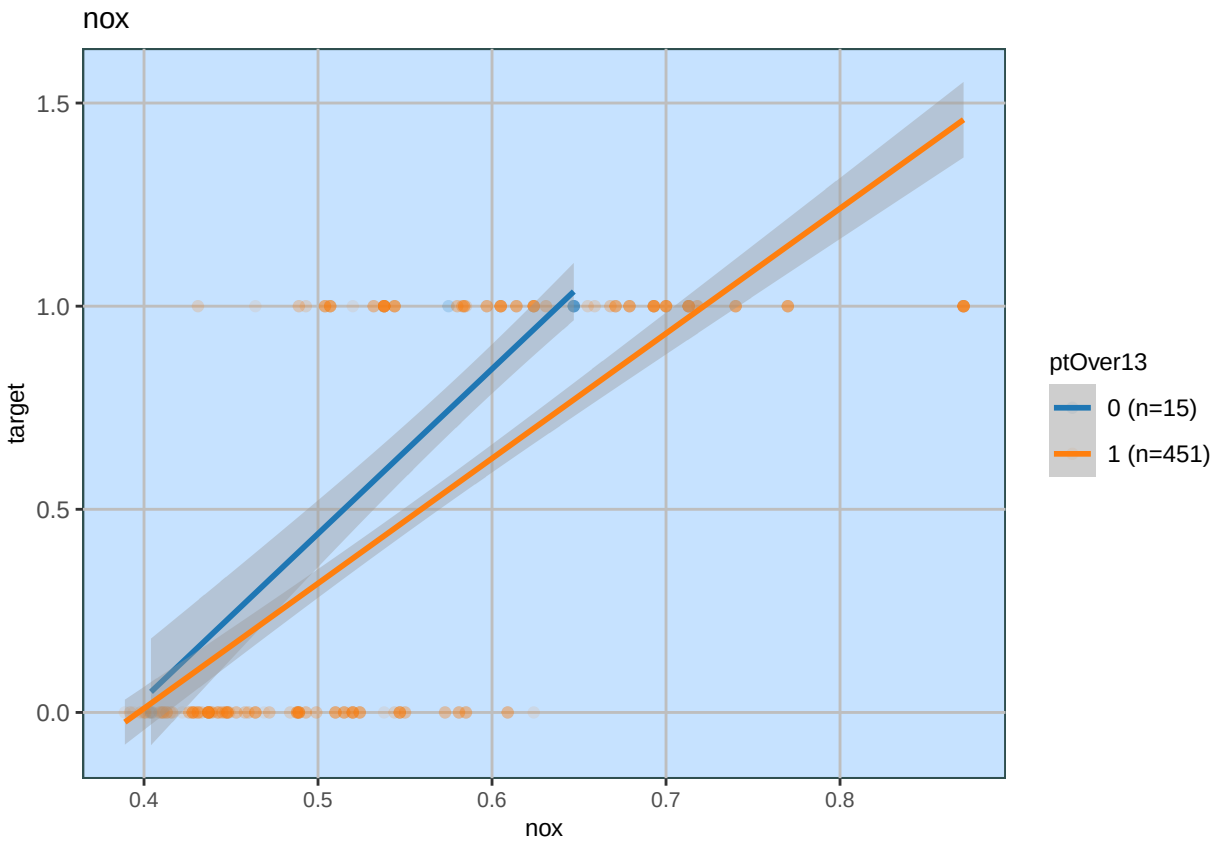

```
##
## [[2]]
```



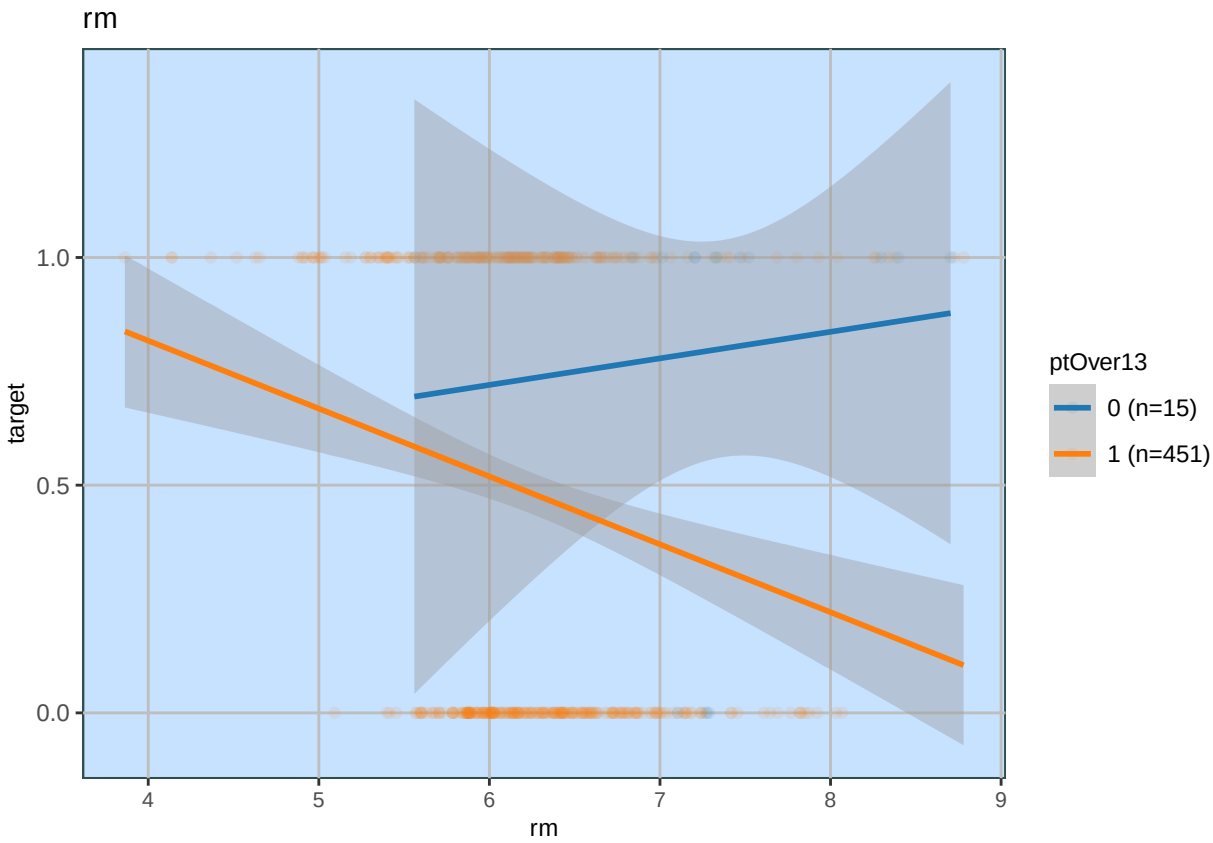

```
##
## [[3]]
```


chas
1.00
0.75
0.50
0.25
0.00
target
0.00
0.25
0.50
0.75
1.00
chas
ptOver13
0 (n=15)
1 (n=451)

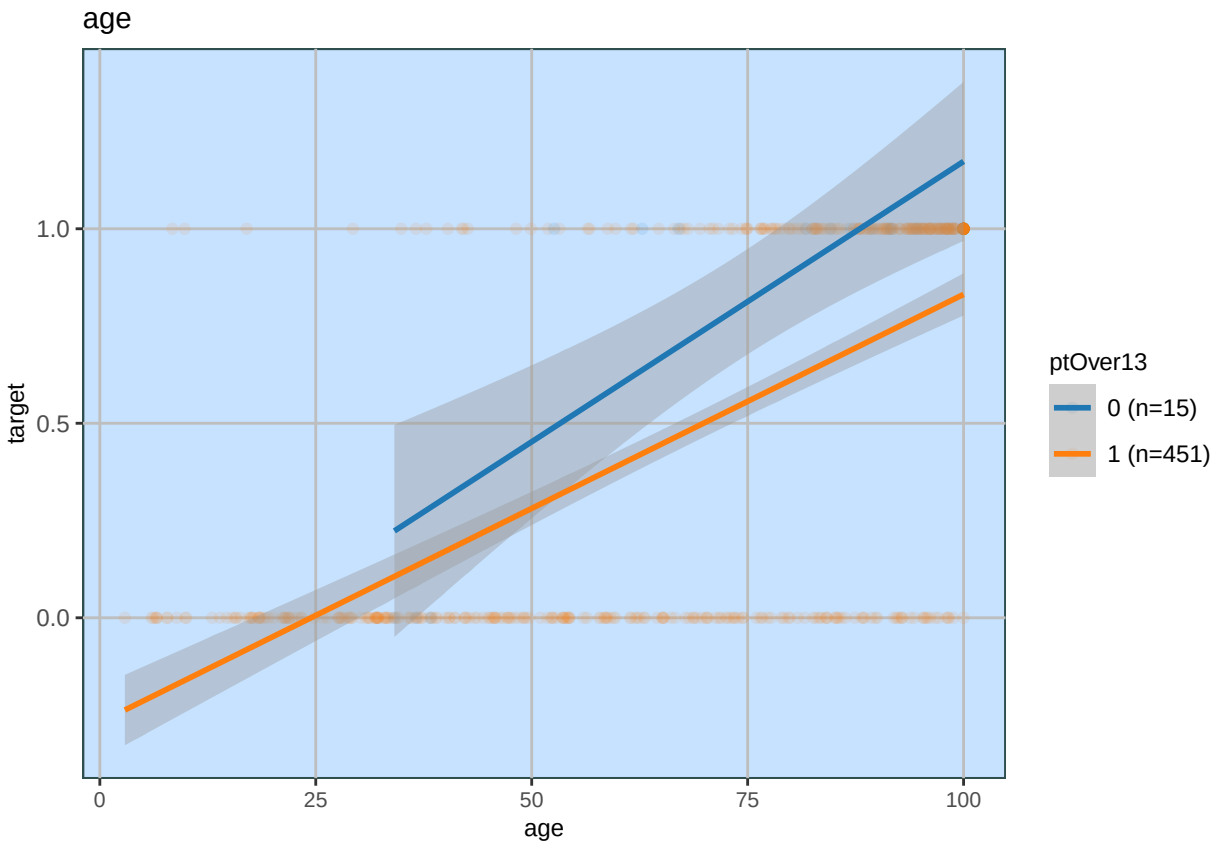
```
##
## [[4]]
```



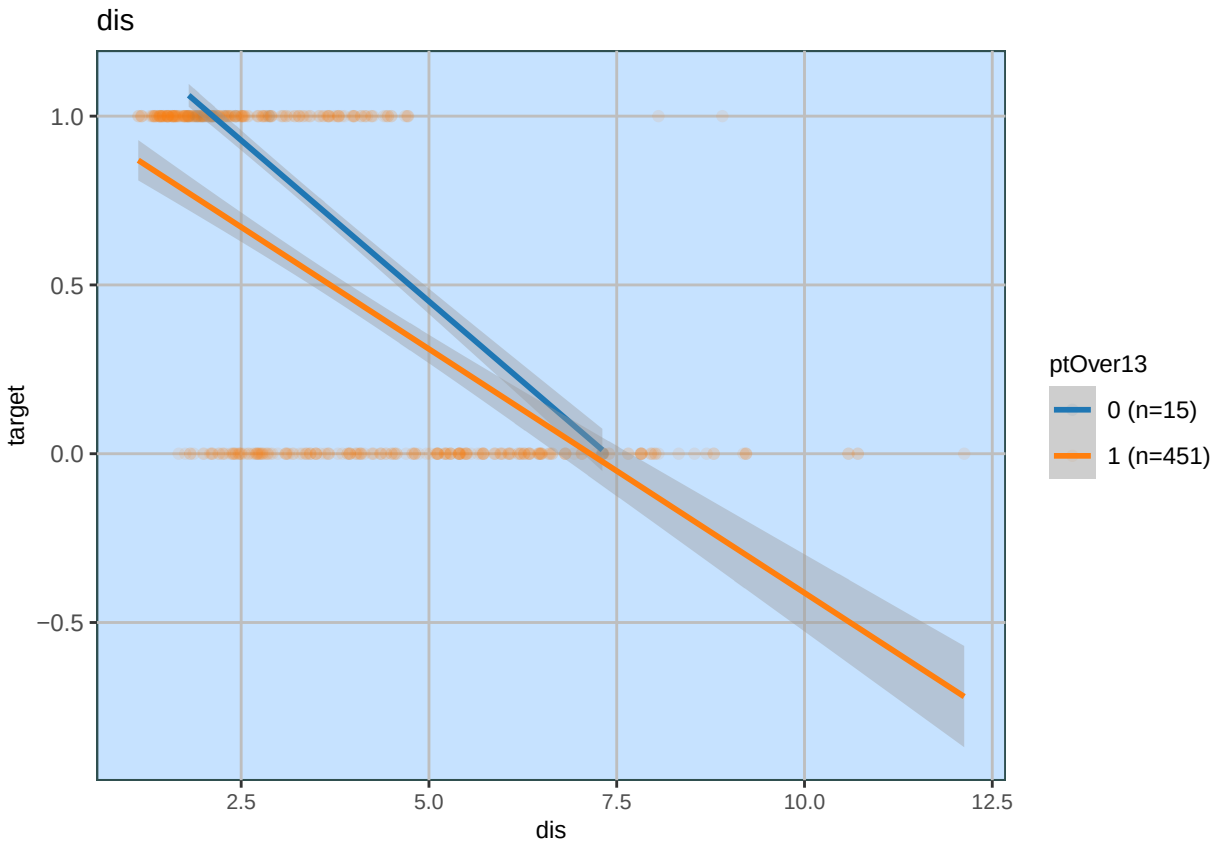

```
##
## [[5]]
```



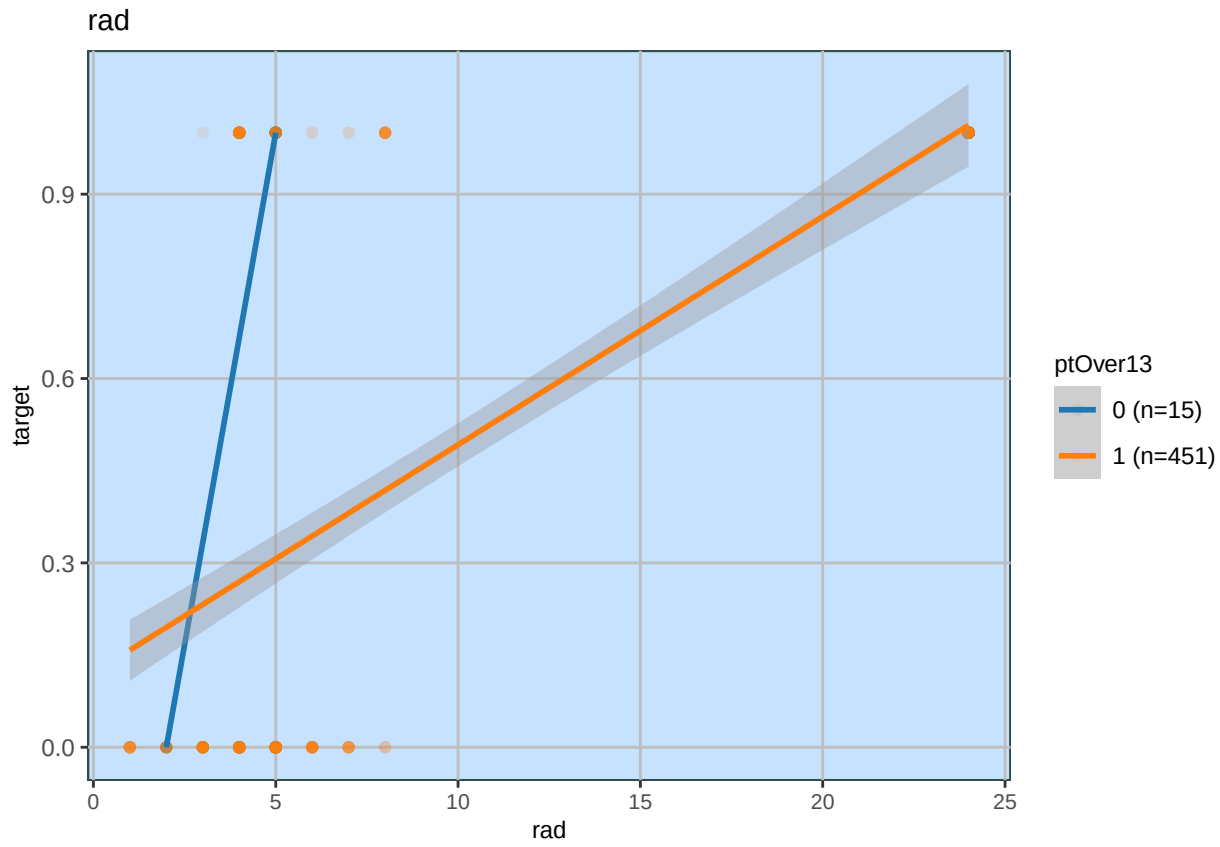

```
##
## [[6]]
```



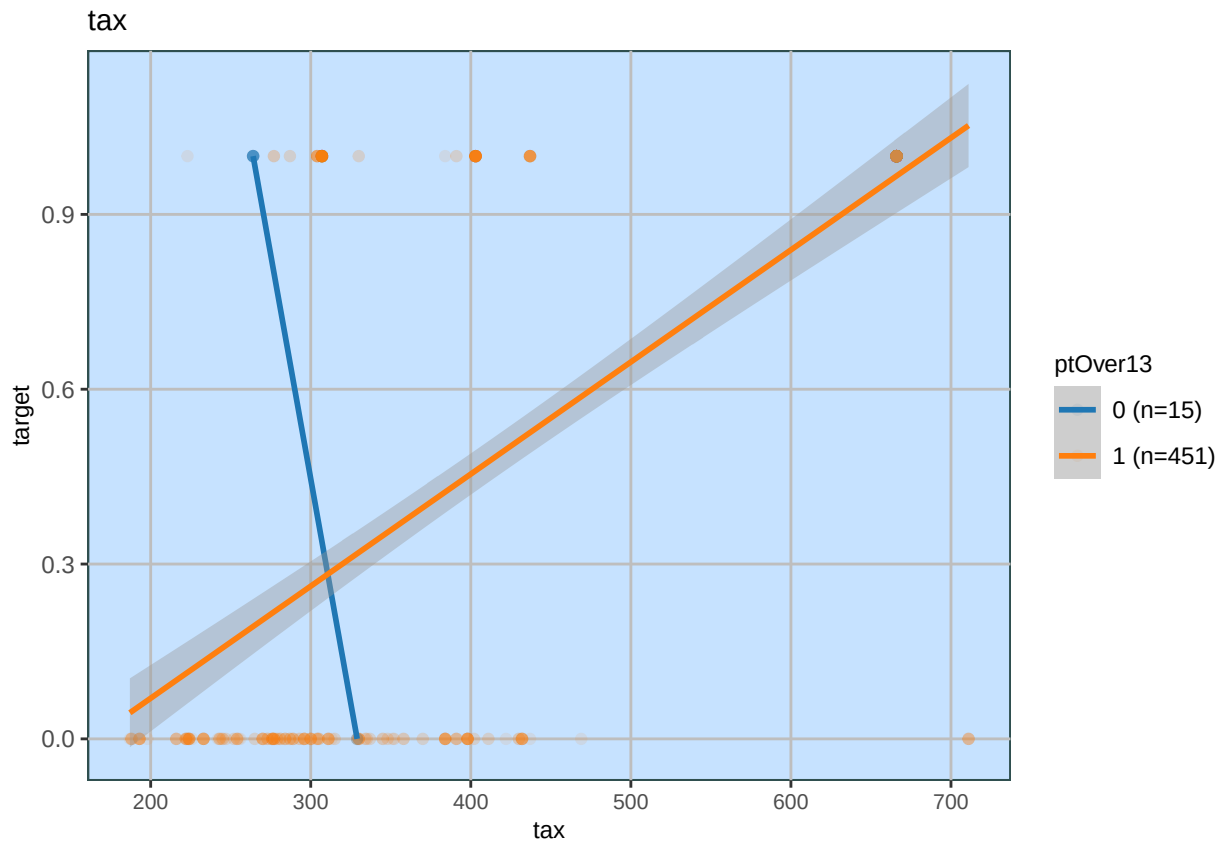

```
##
## [[7]]
```



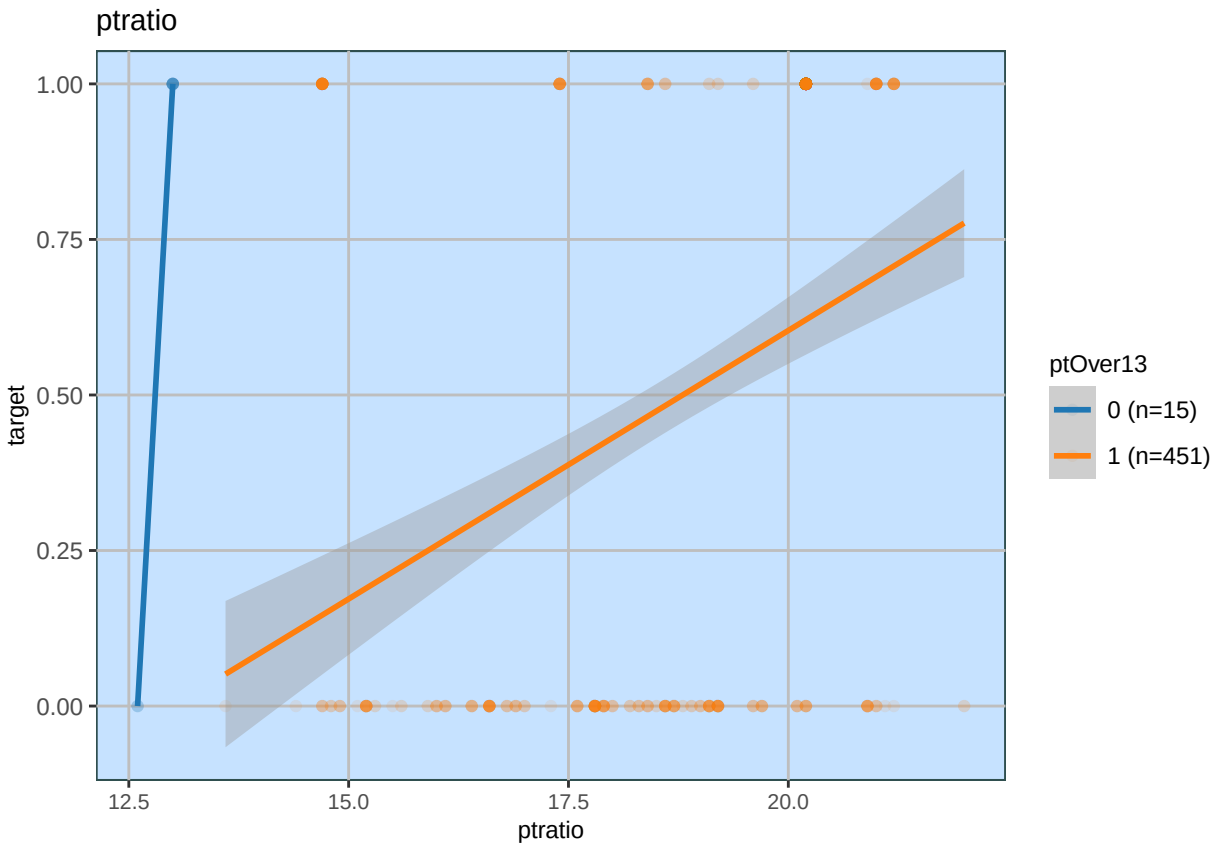

```
##
## [[8]]
```



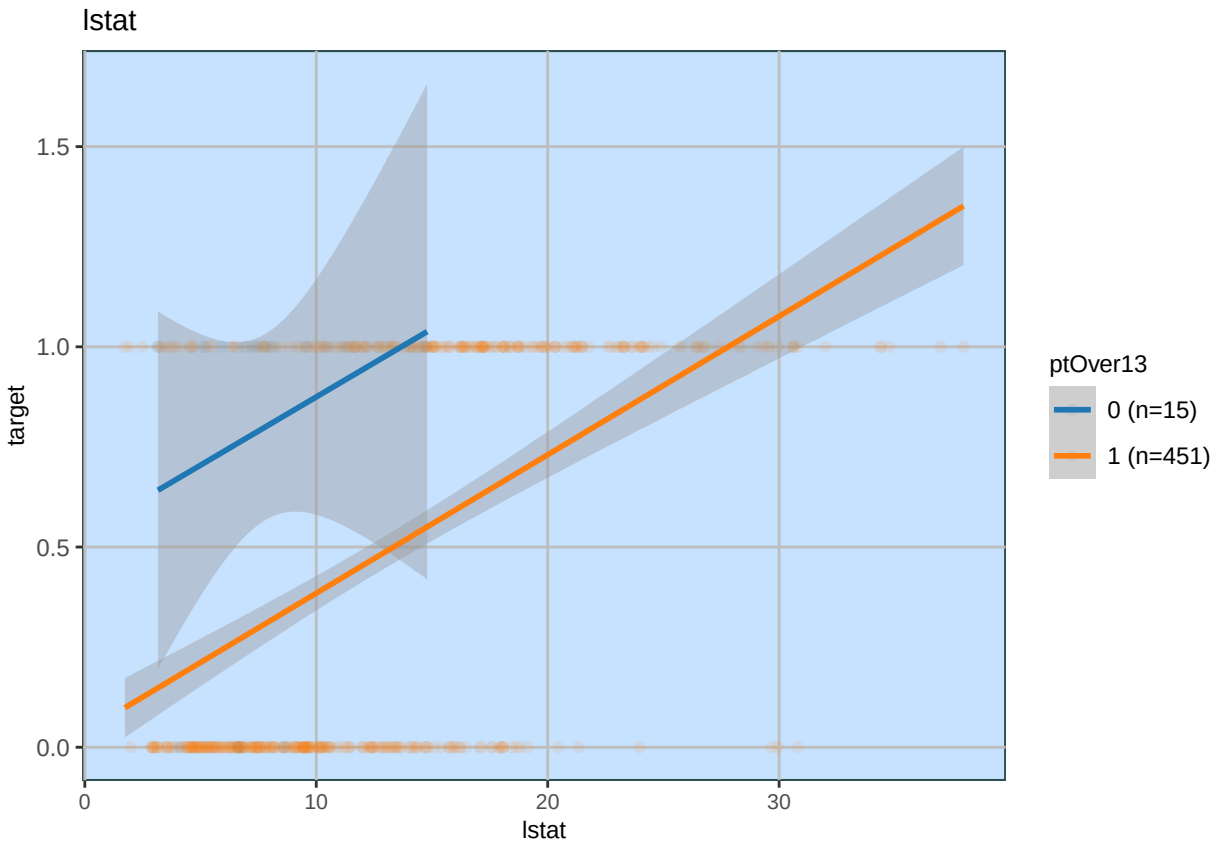

```
##
## [[9]]
```


tax
target
0.0
0.3
0.6
0.9
200
300
400
500
600
700
tax
ptOver13
0 (n=15)
1 (n=451)

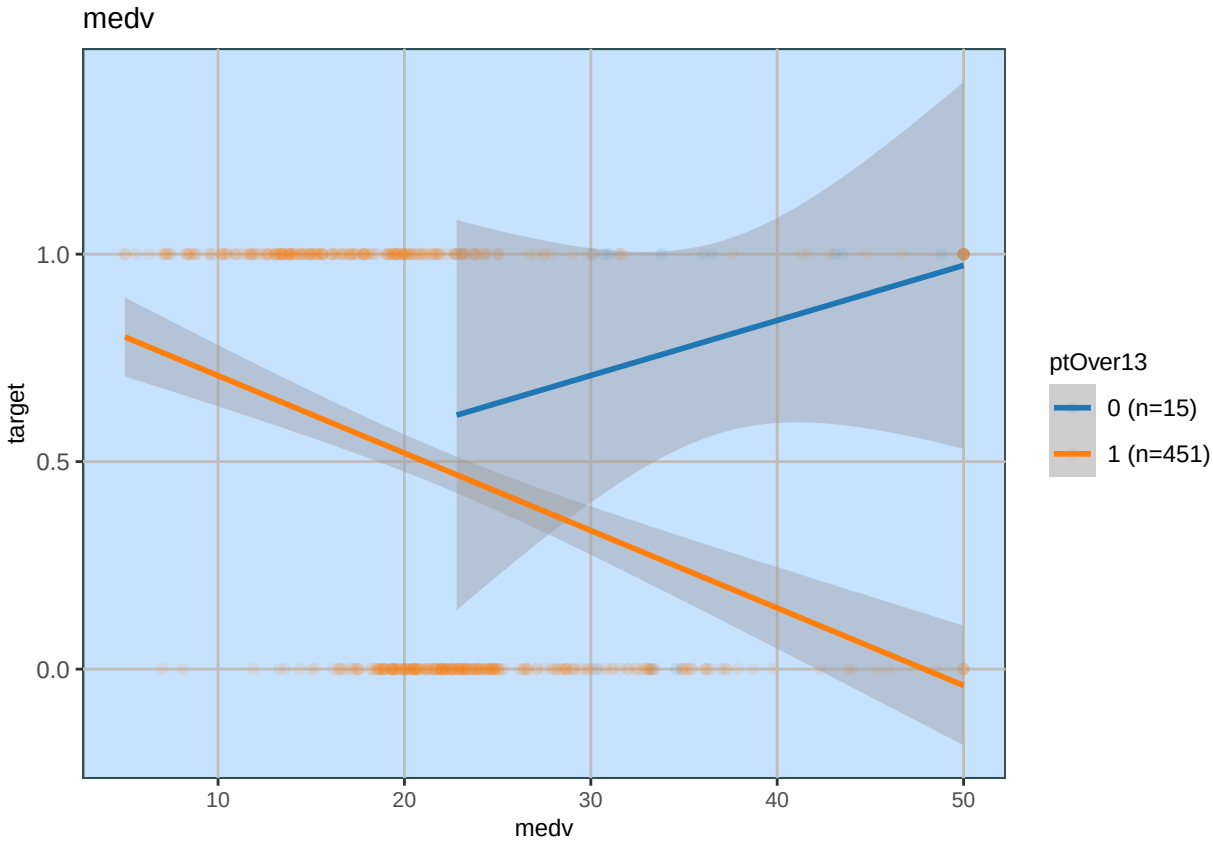
```
##
## [[10]]
```


ptratio
1.00
0.75
0.50
0.25
0.00
target
12.5
15.0
17.5
20.0
ptratio
ptOver13
0 (n=15)
1 (n=451)

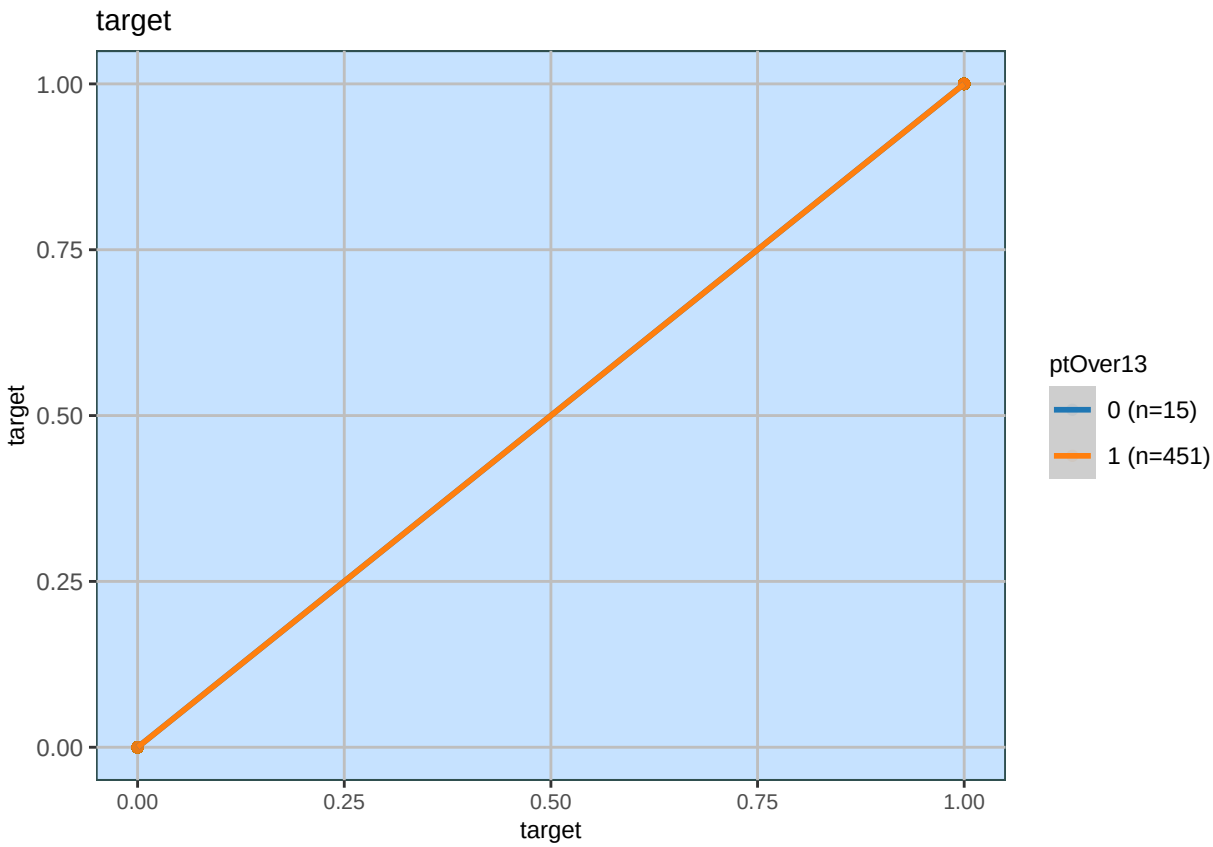
```
##
## [[11]]
```



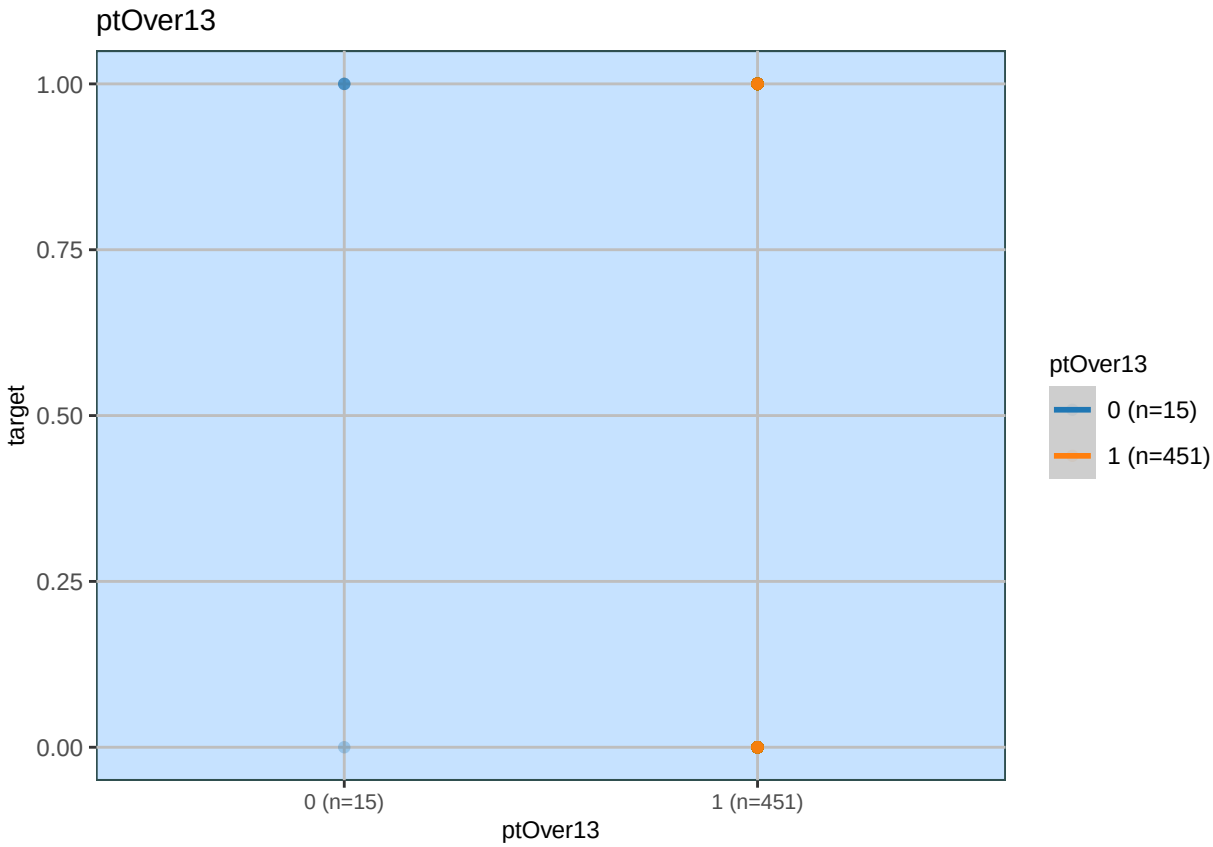

```
##
## [[12]]
```



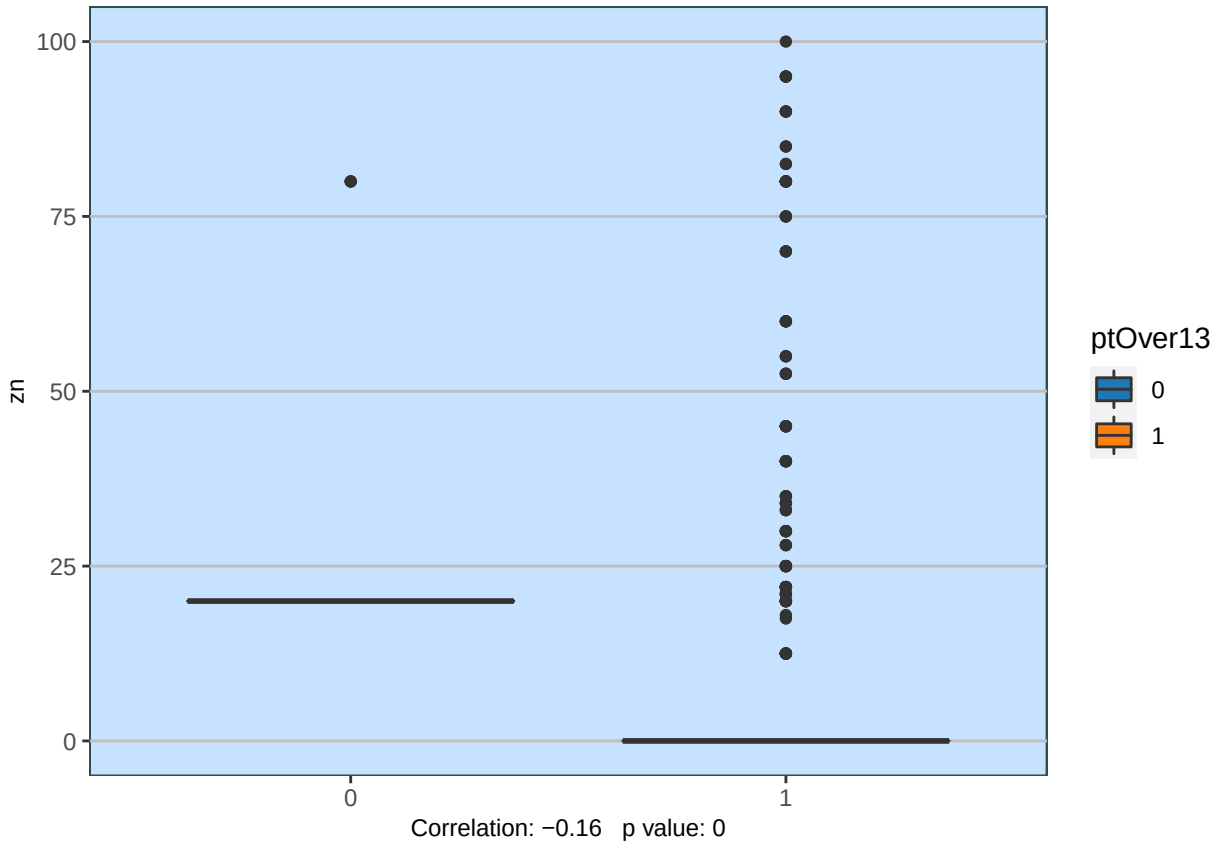

```
##
## [[13]]
```


target
1.00
0.75
0.50
0.25
0.00
target
0.00
0.25
0.50
0.75
1.00
target
ptOver13
0 (n=15)
1 (n=451)

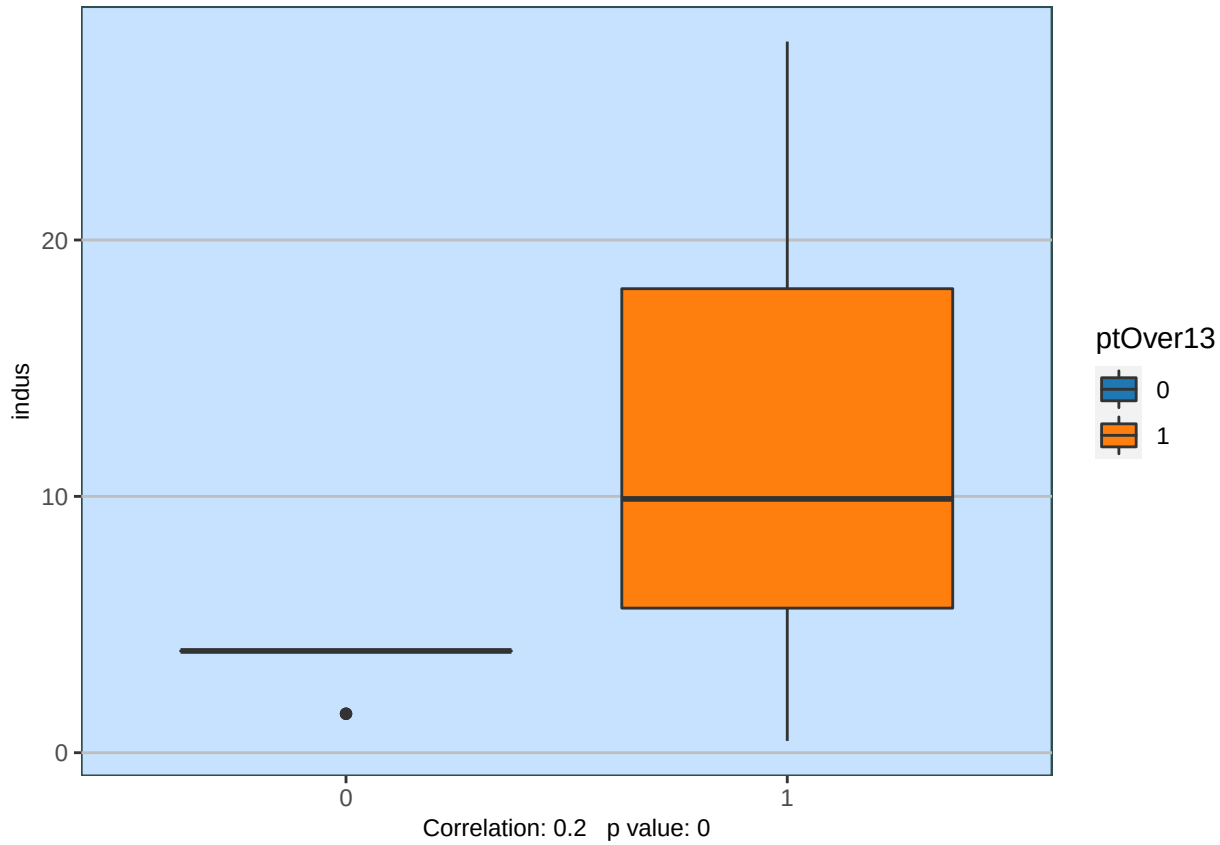
```
##
## [[14]]
```


ptOver13
1.00
0.75
0.50
0.25
0.00
target
0 (n=15)
1 (n=451)
ptOver13
ptOver13
0 (n=15)
1 (n=451)

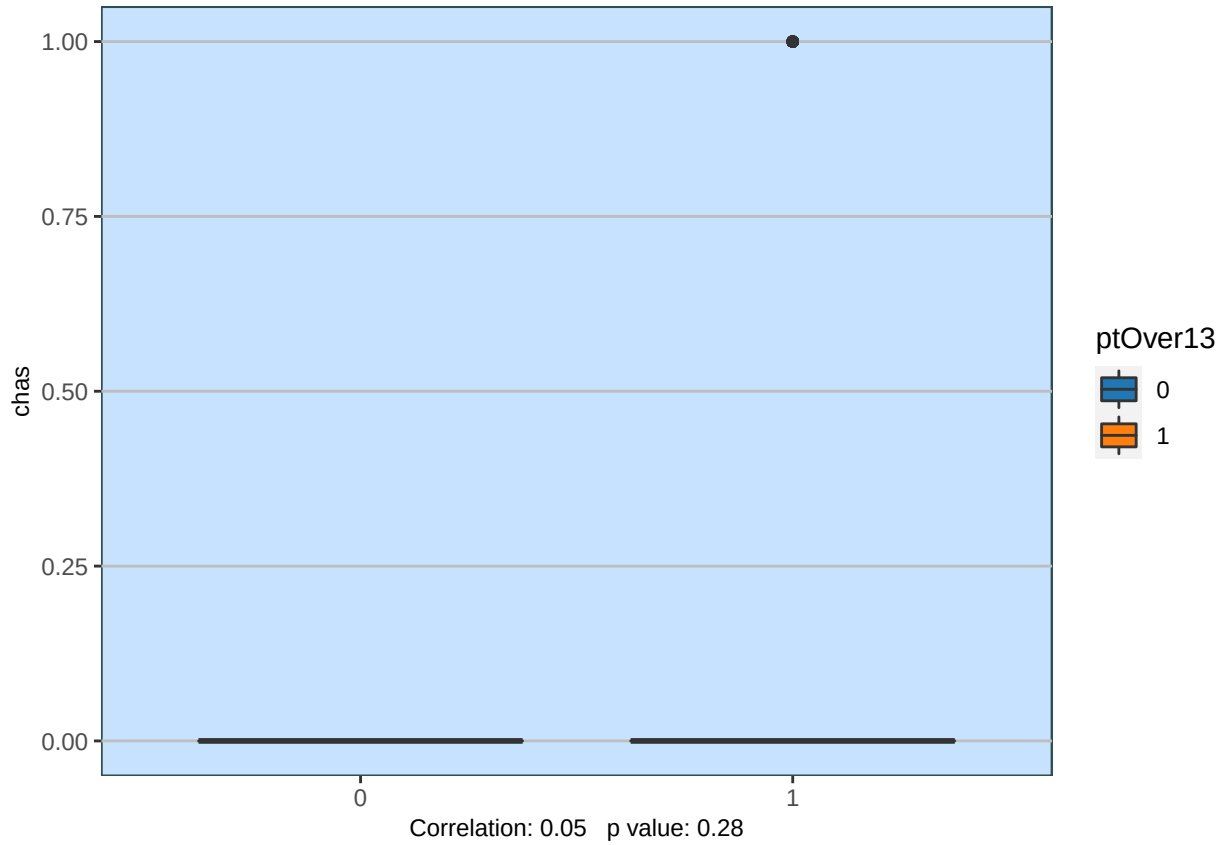
```
## [[1]]
```



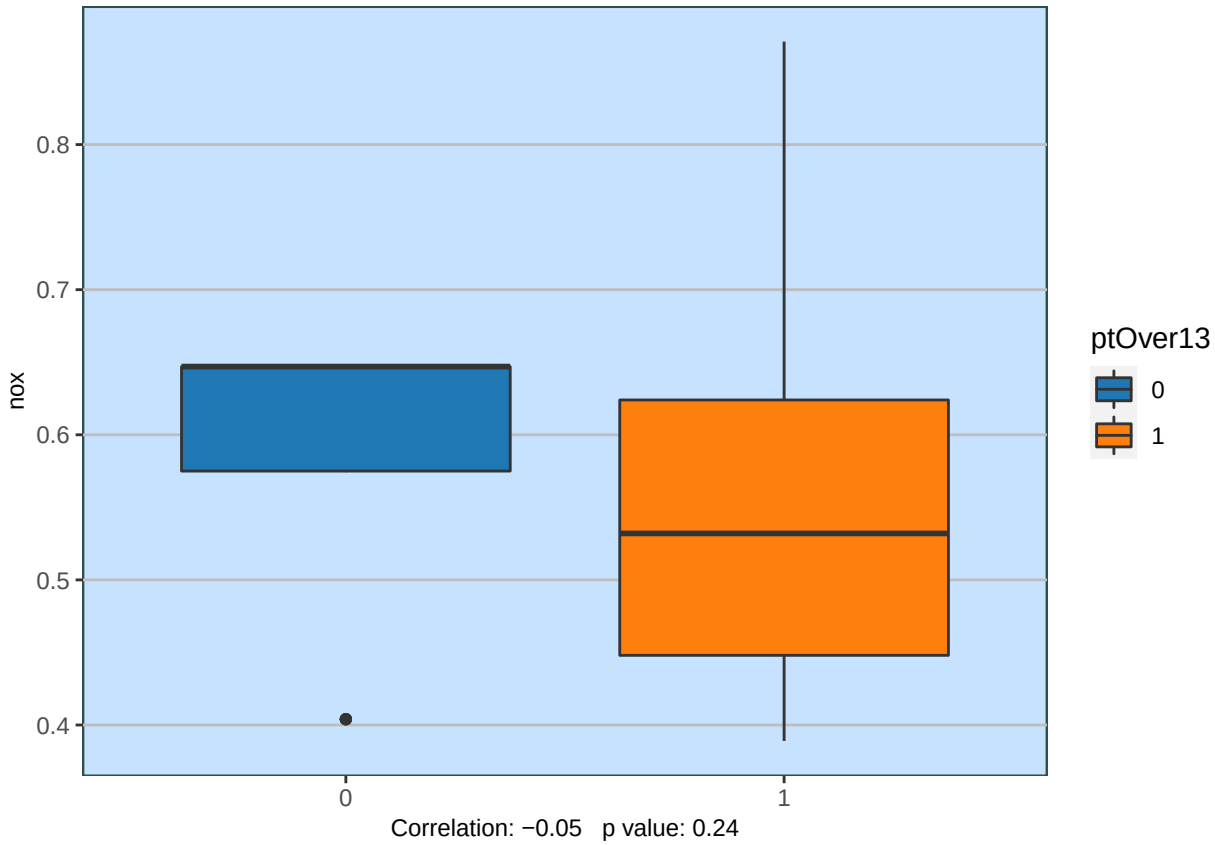

```
##
## [[2]]
```



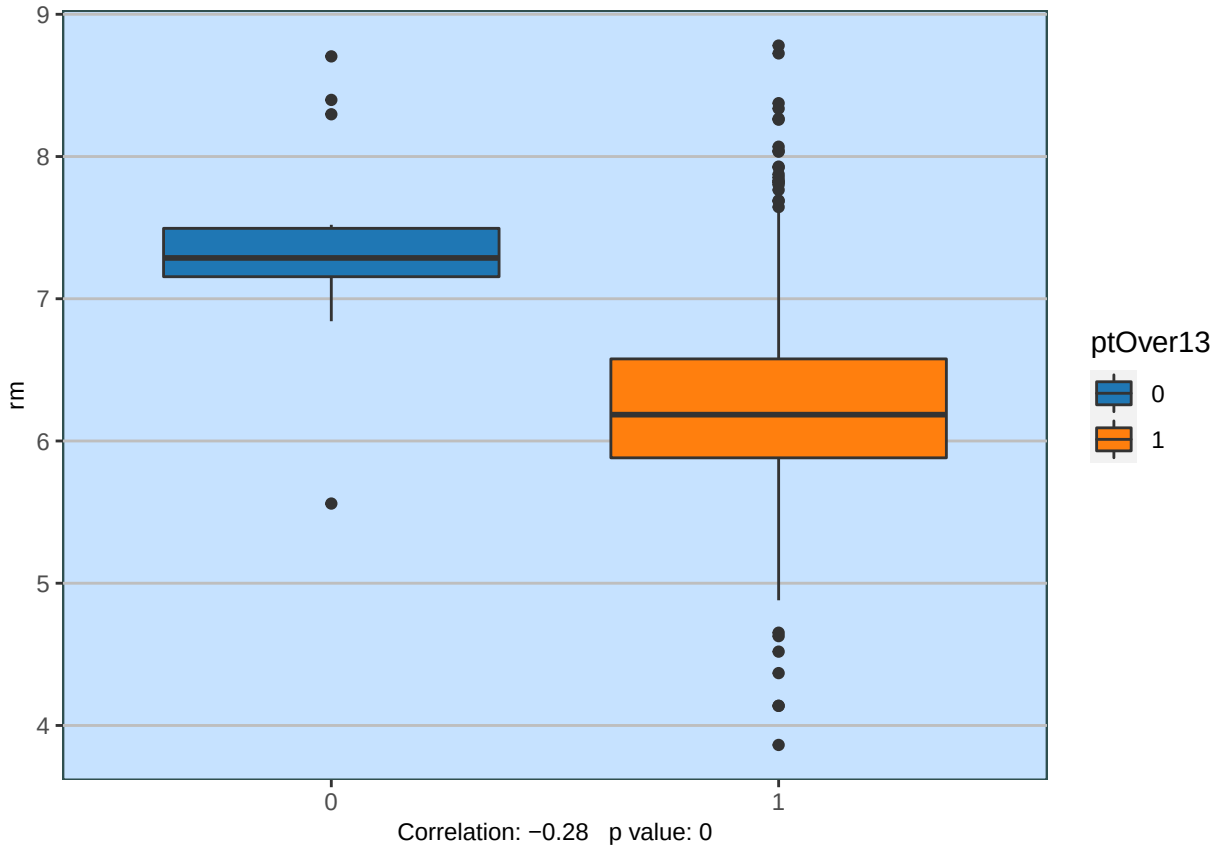

```
##
## [[3]]
```


1.00
0.75
0.50
0.25
0.00
chas
0
1
Correlation: 0.05 p value: 0.28
ptOver13
0
1

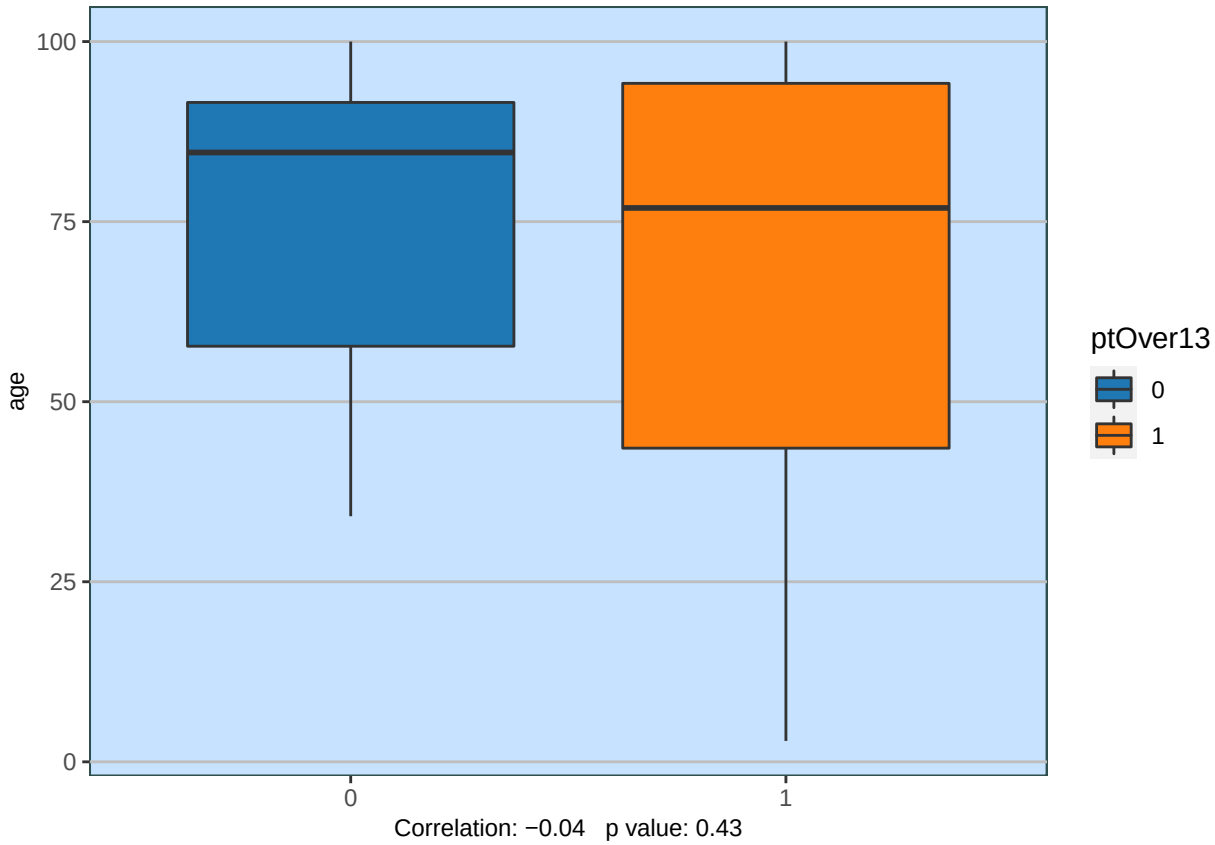
```
##
## [[4]]
```



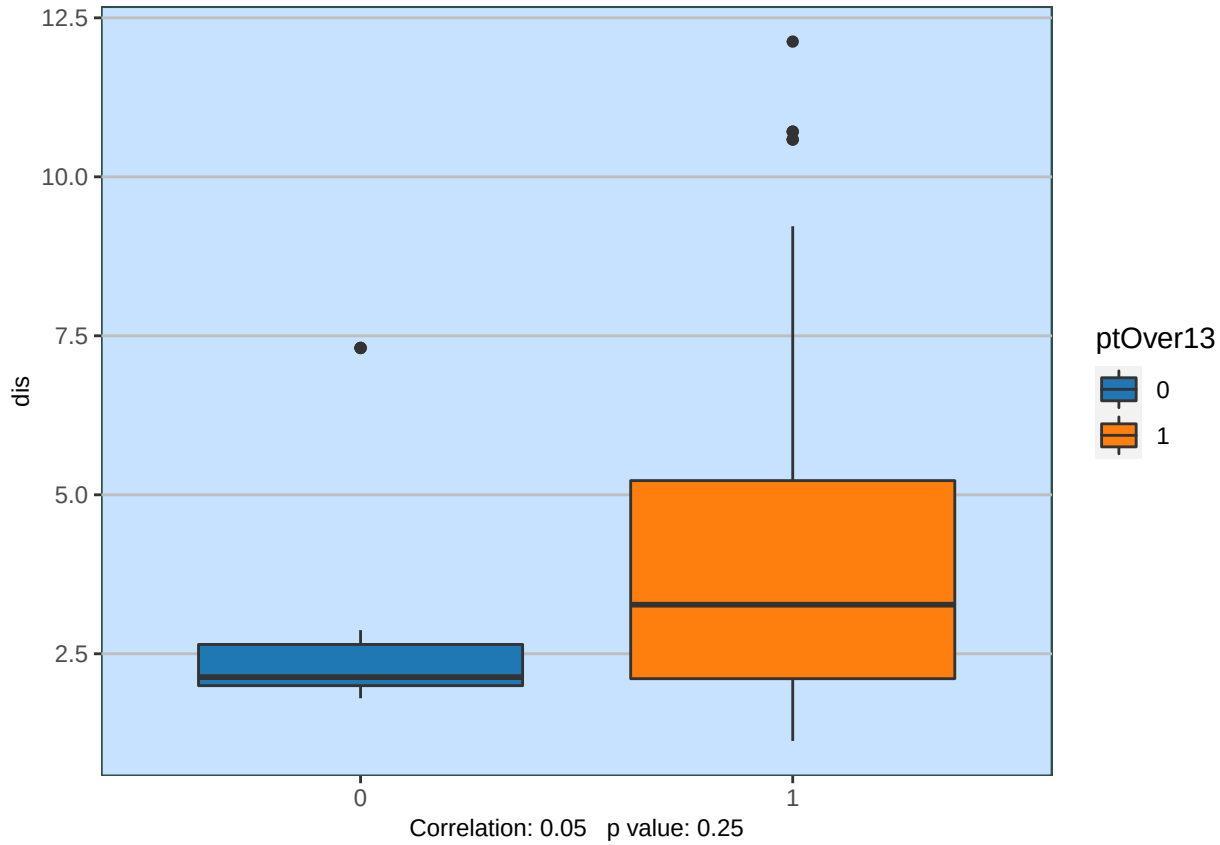

```
##
## [[5]]
```



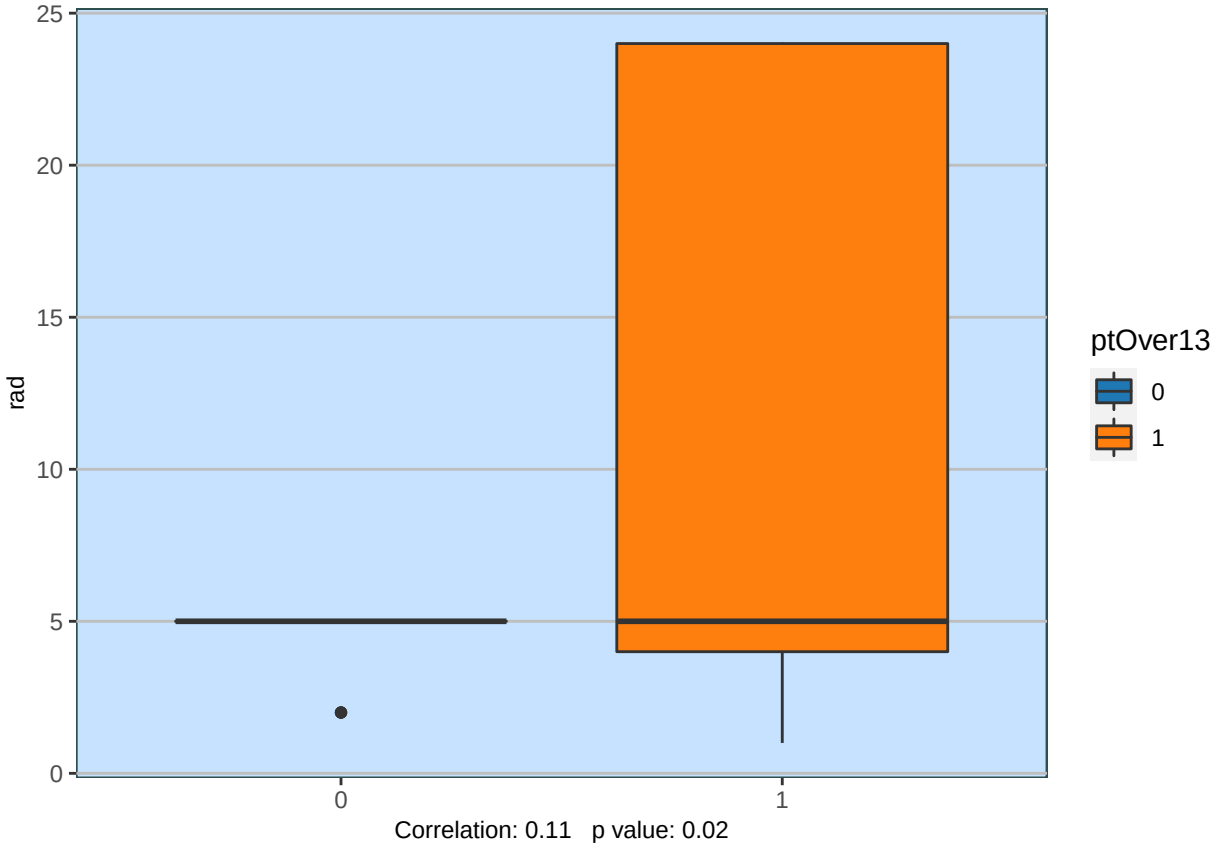

```
##
## [[6]]
```



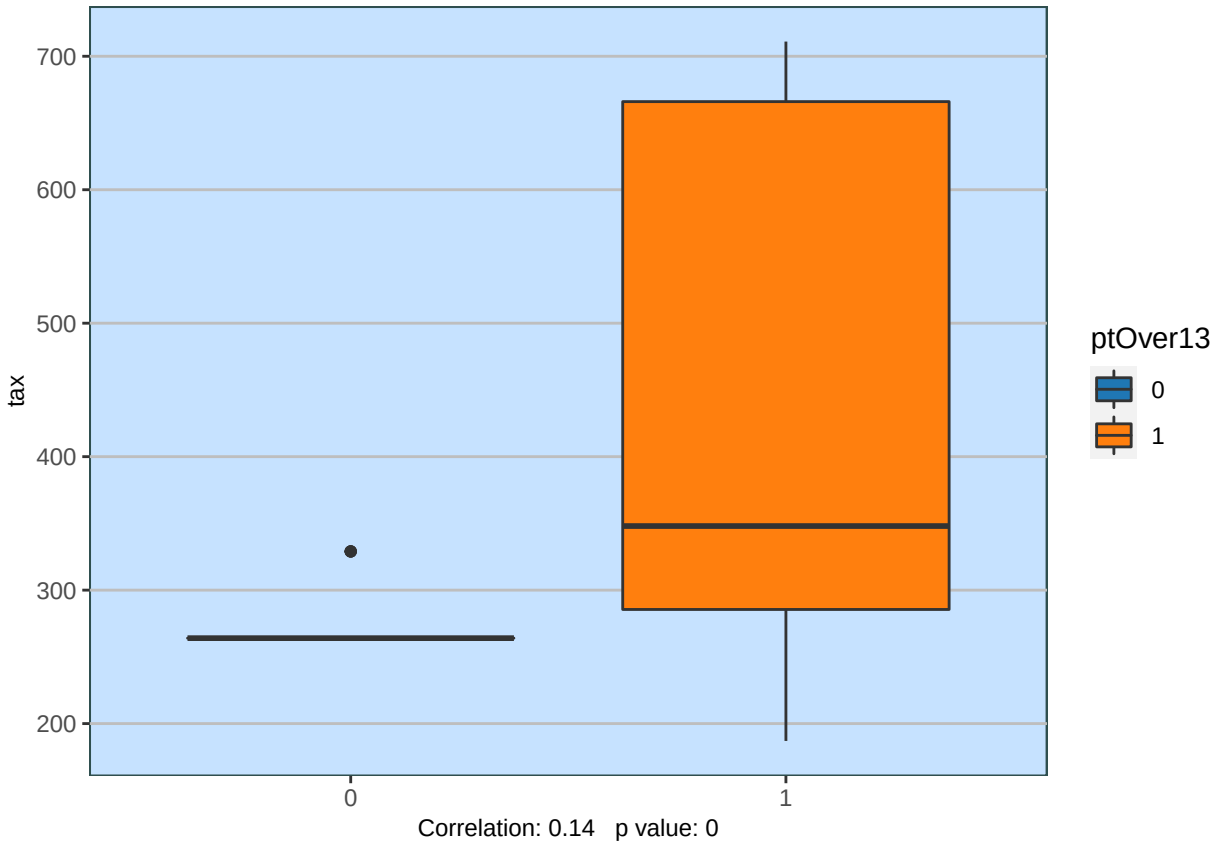

```
##
## [[7]]
```


12.5
10.0
7.5
5.0
2.5
dis
0
1
Correlation: 0.05 p value: 0.25
ptOver13
0
1

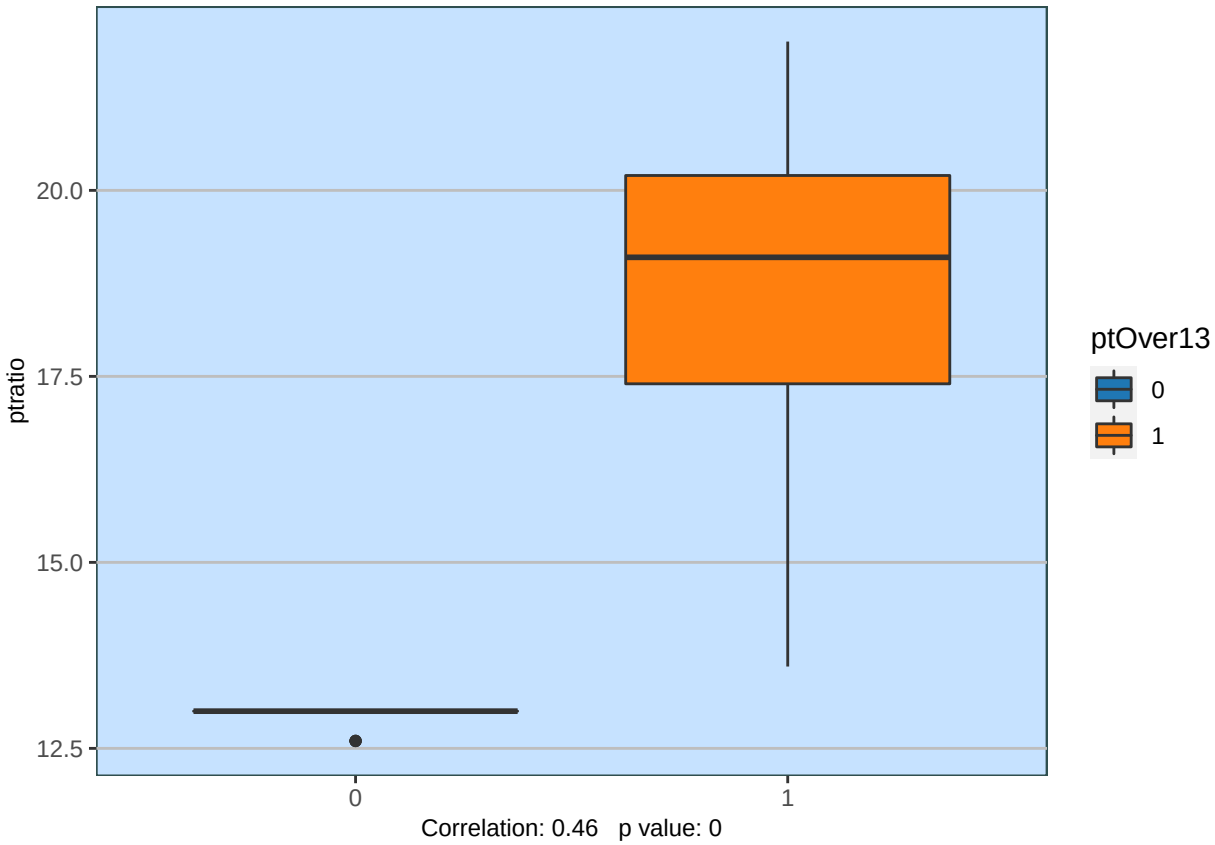
```
##
## [[8]]
```



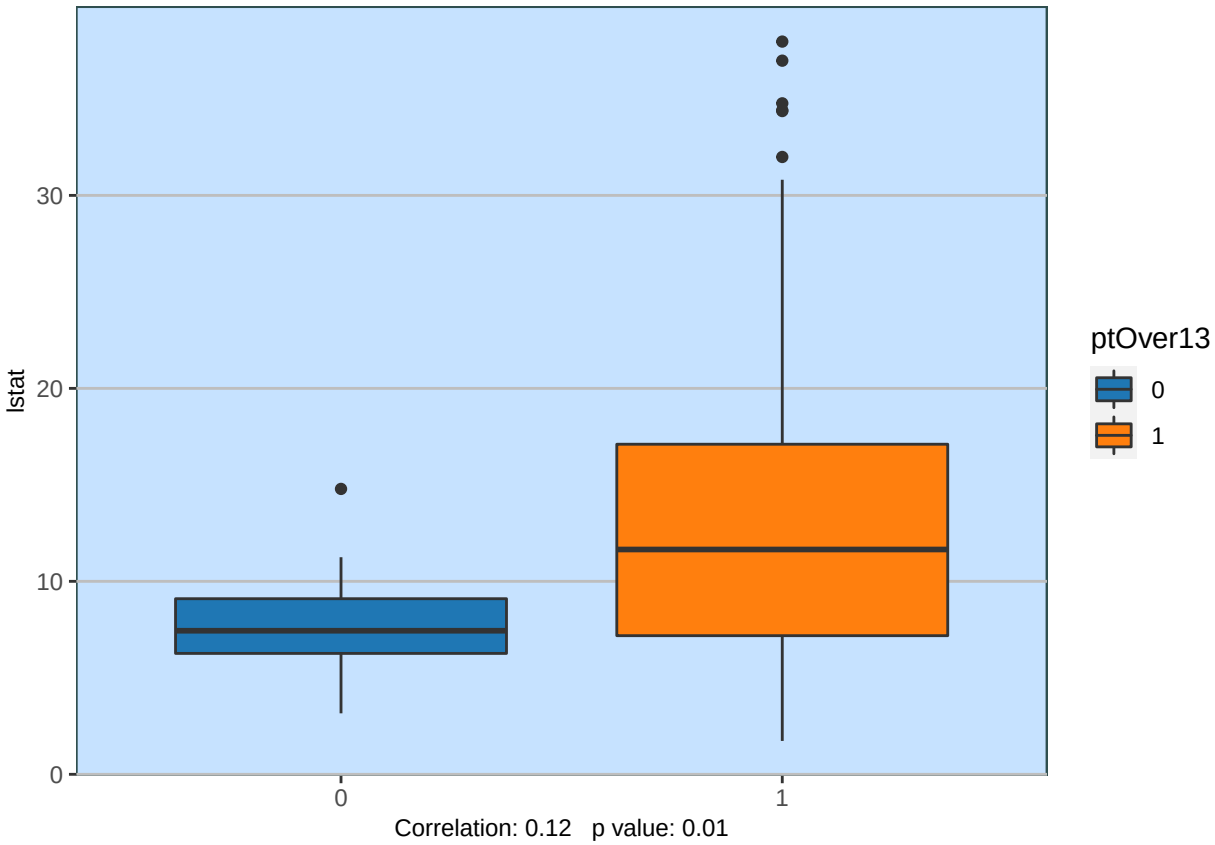

```
##
## [[9]]
```


700
600
500
400
300
200
tax
0
1
Correlation: 0.14 p value: 0
ptOver13
0
1

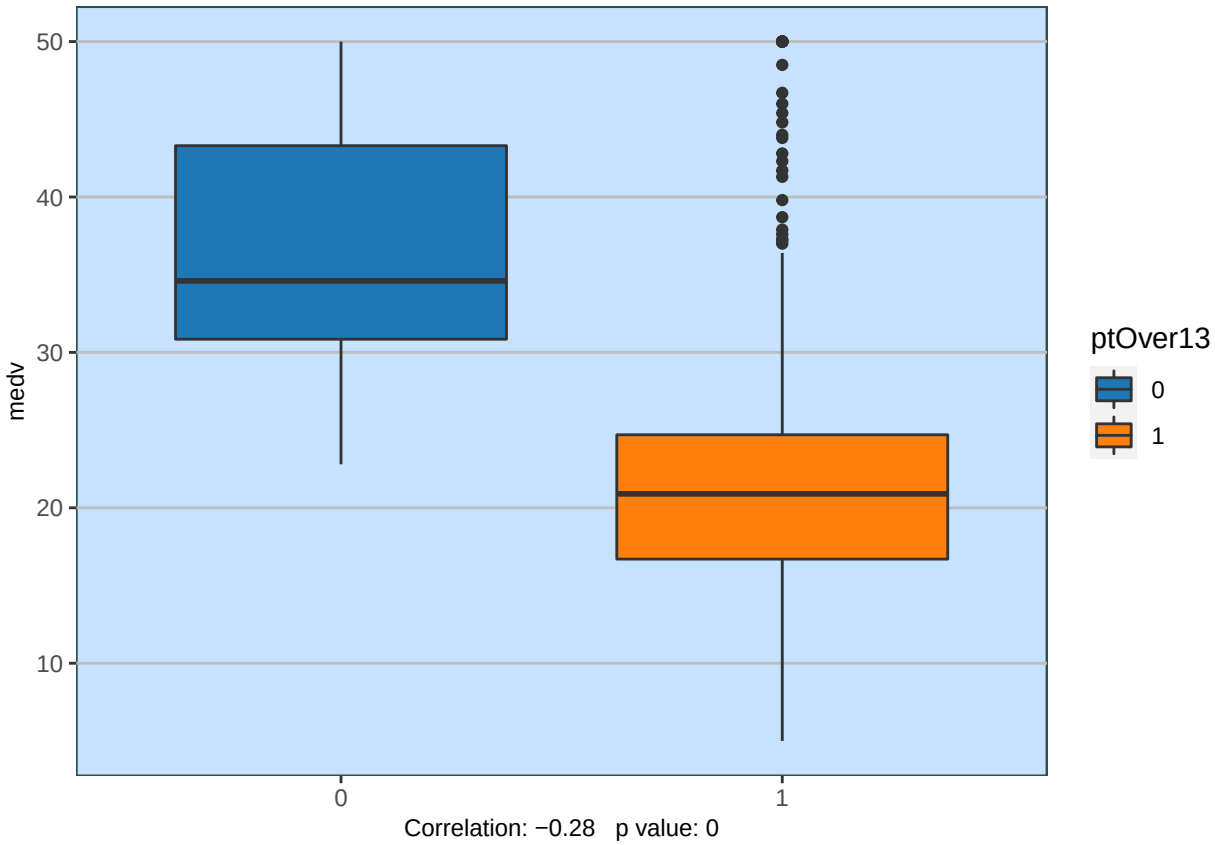
```
##
## [[10]]
```


20.0
17.5
15.0
12.5
ptratio
0
1
Correlation: 0.46 p value: 0
ptOver13
0
1

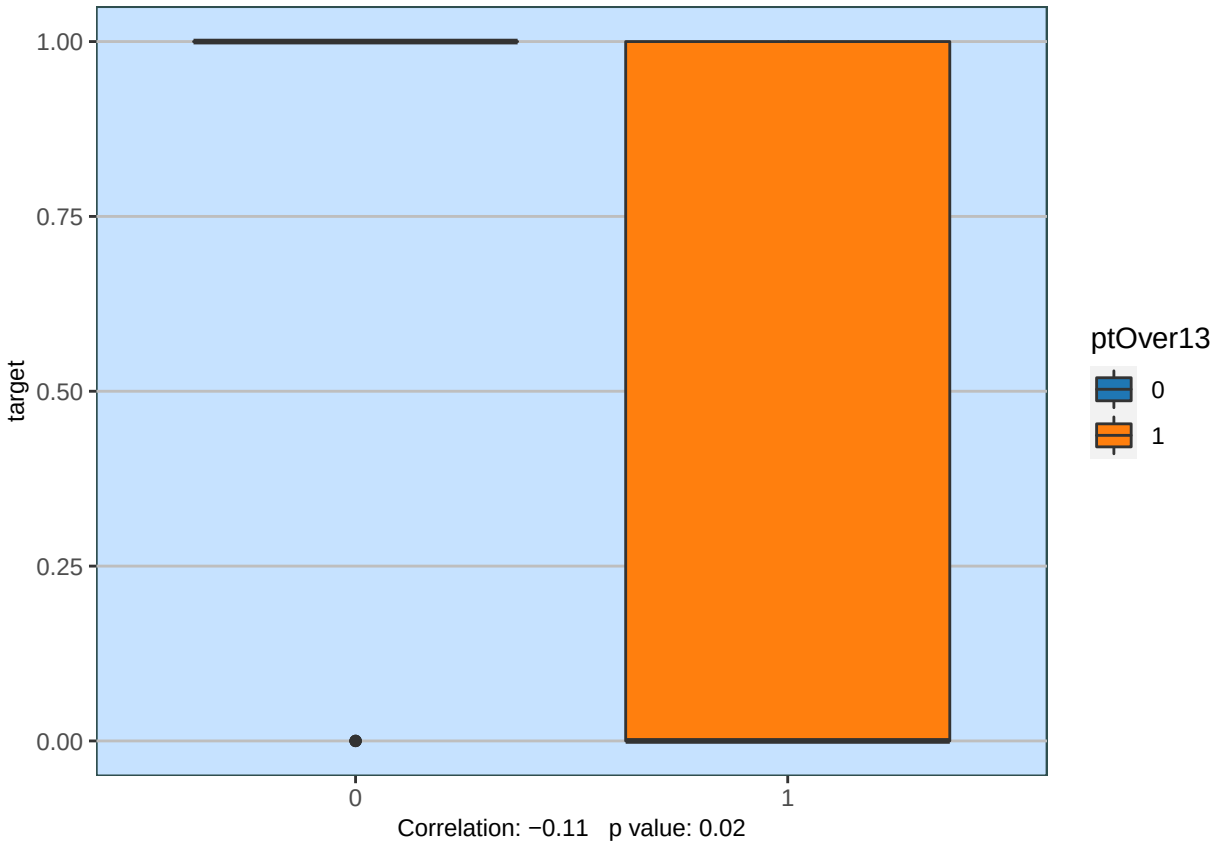
```
##
## [[11]]
```



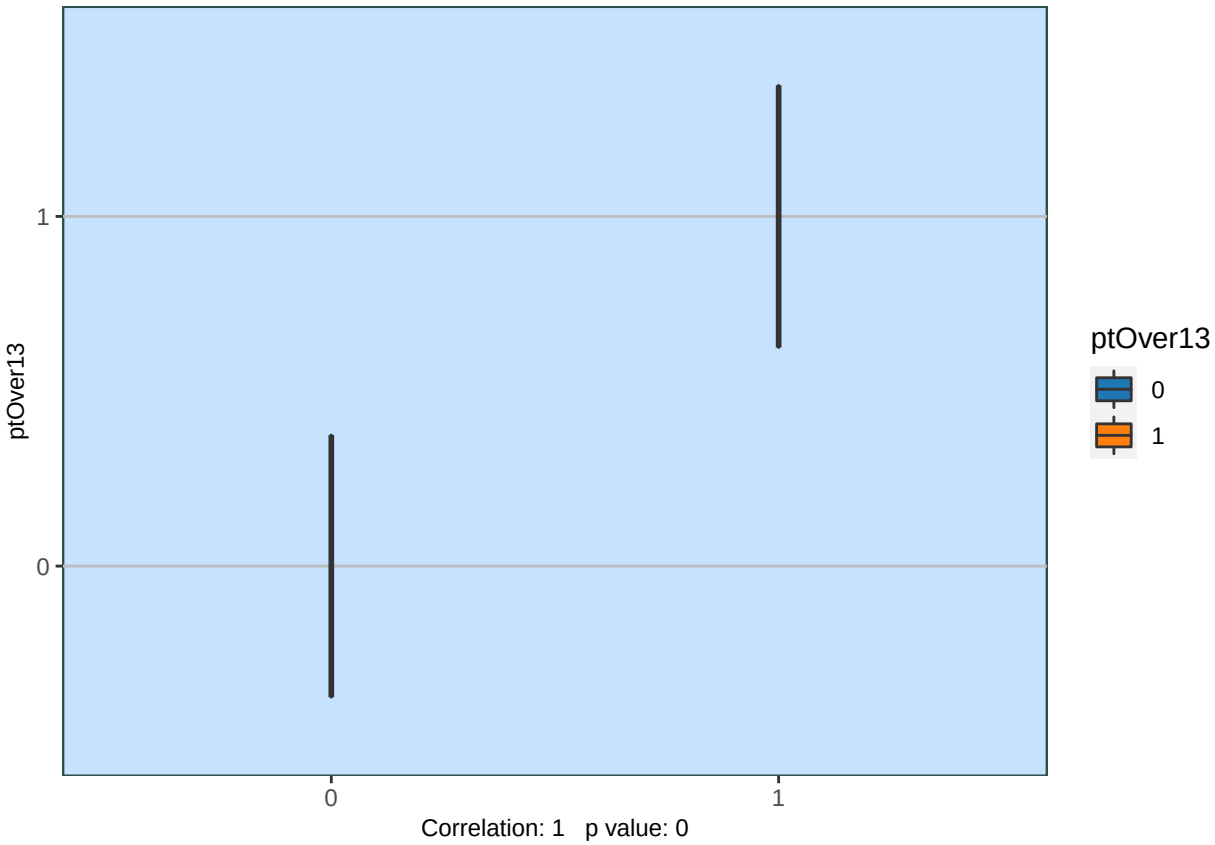

```
##
## [[12]]
```



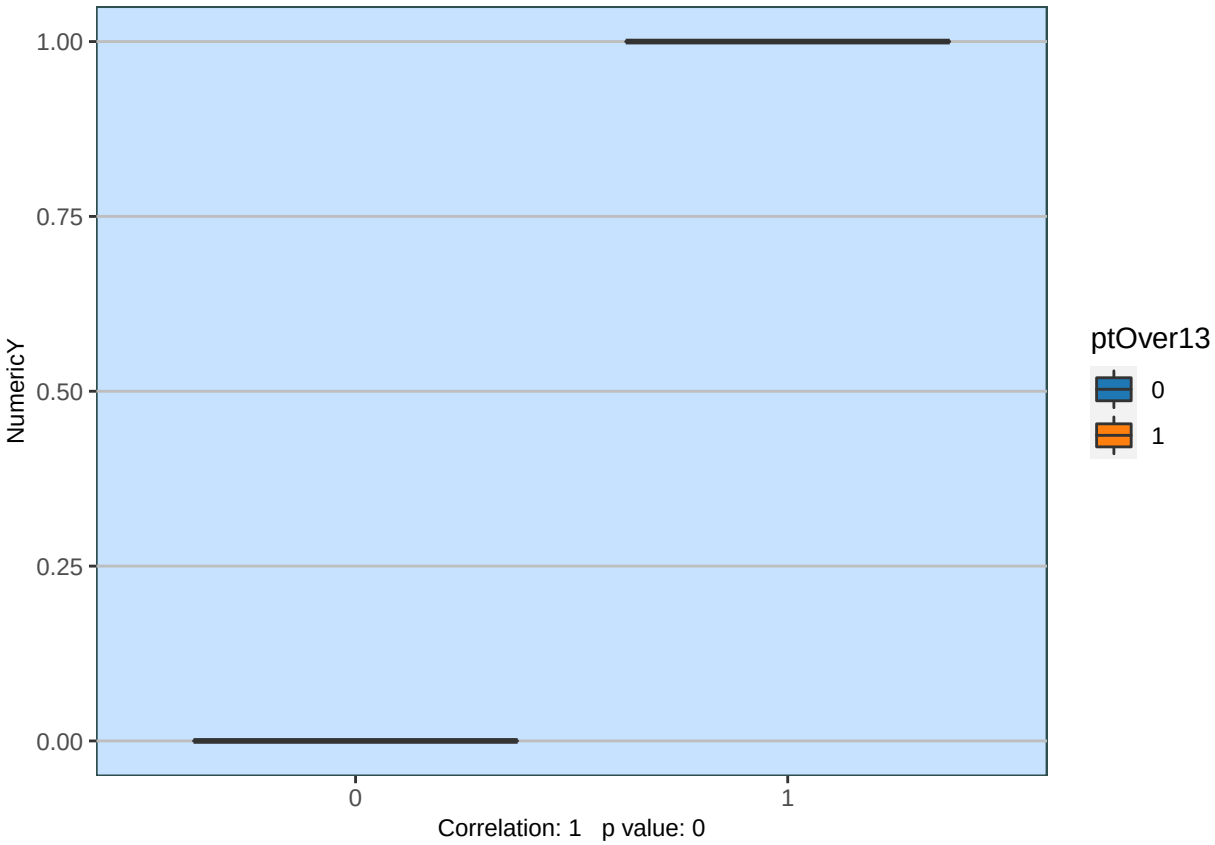

```
##
## [[13]]
```


1.00
0.75
0.50
0.25
0.00
target
0
1
Correlation: −0.11 p value: 0.02
ptOver13
0
1

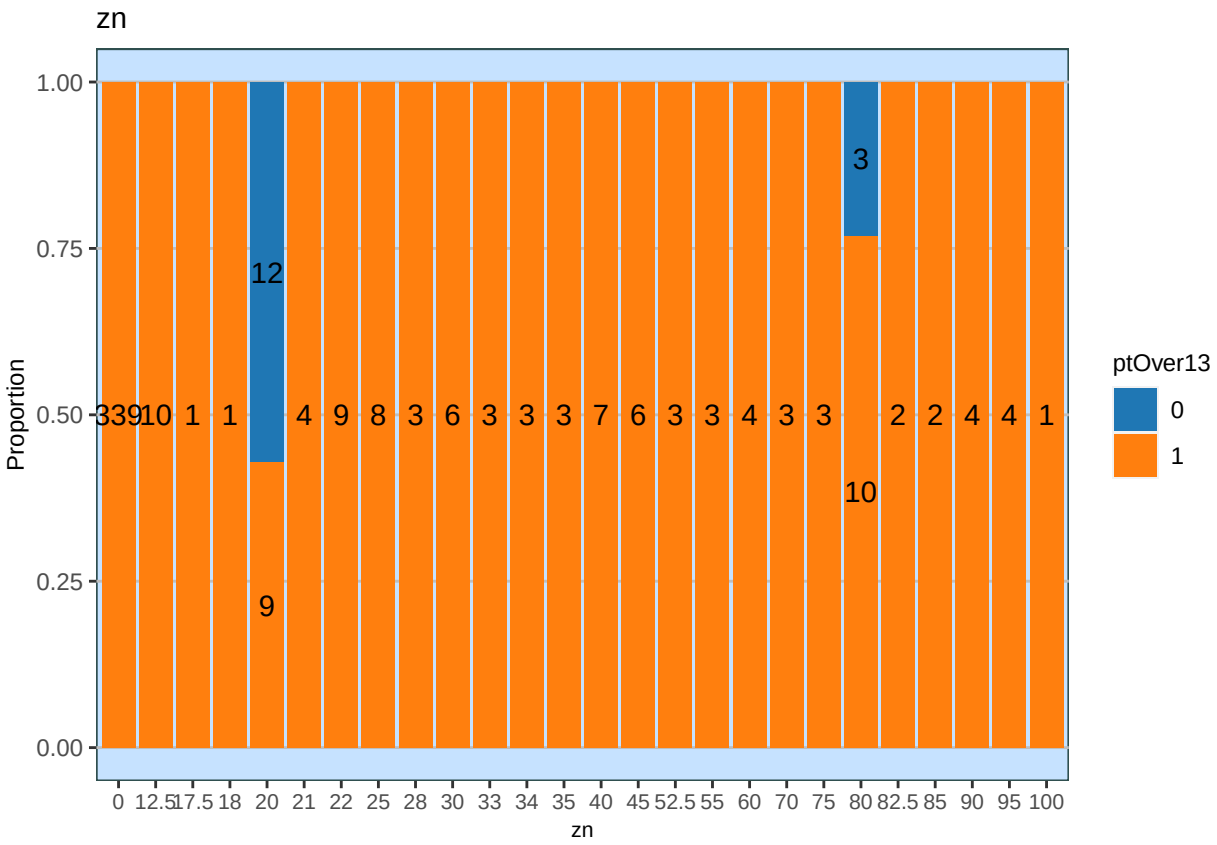
```
##
## [[14]]
```



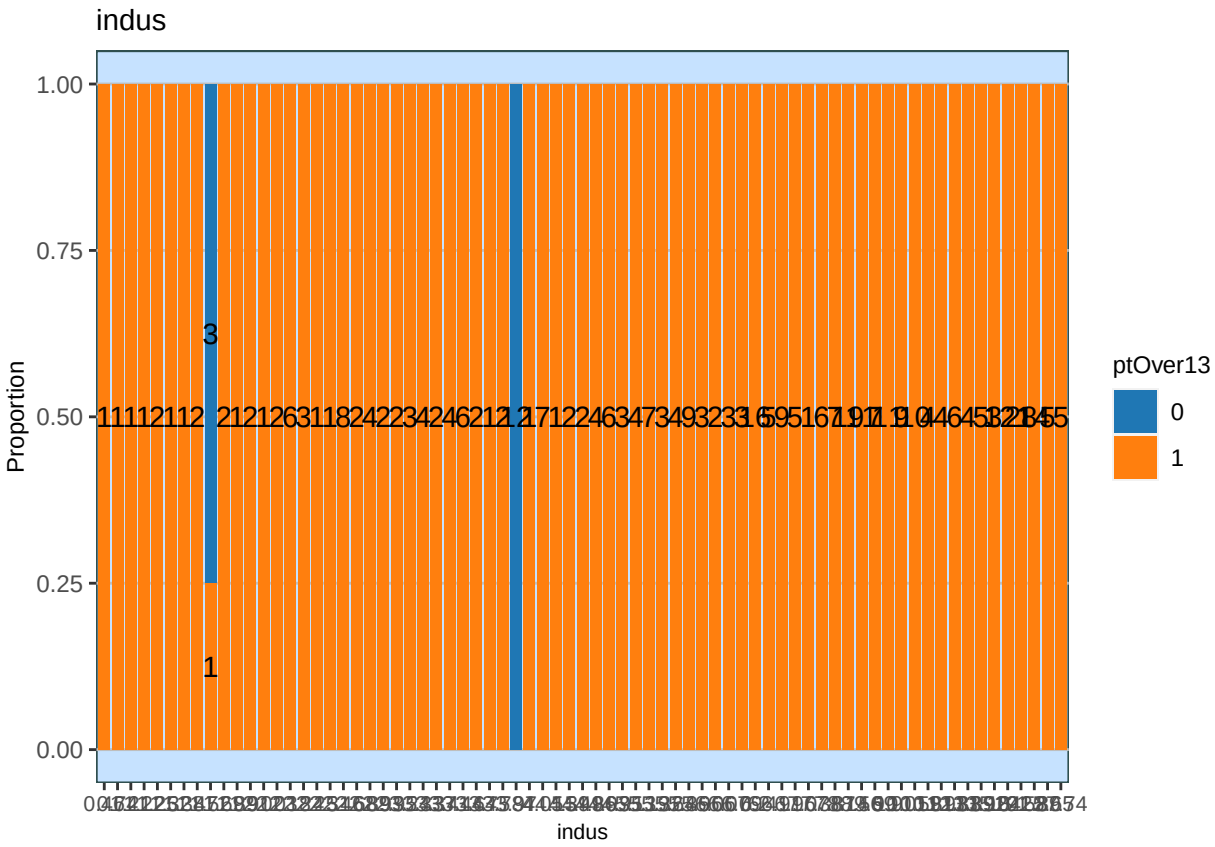

```
##
## [[15]]
```


1.00
0.75
0.50
0.25
0.00
NumericY
0
1
Correlation: 1 p value: 0
ptOver13
0
1

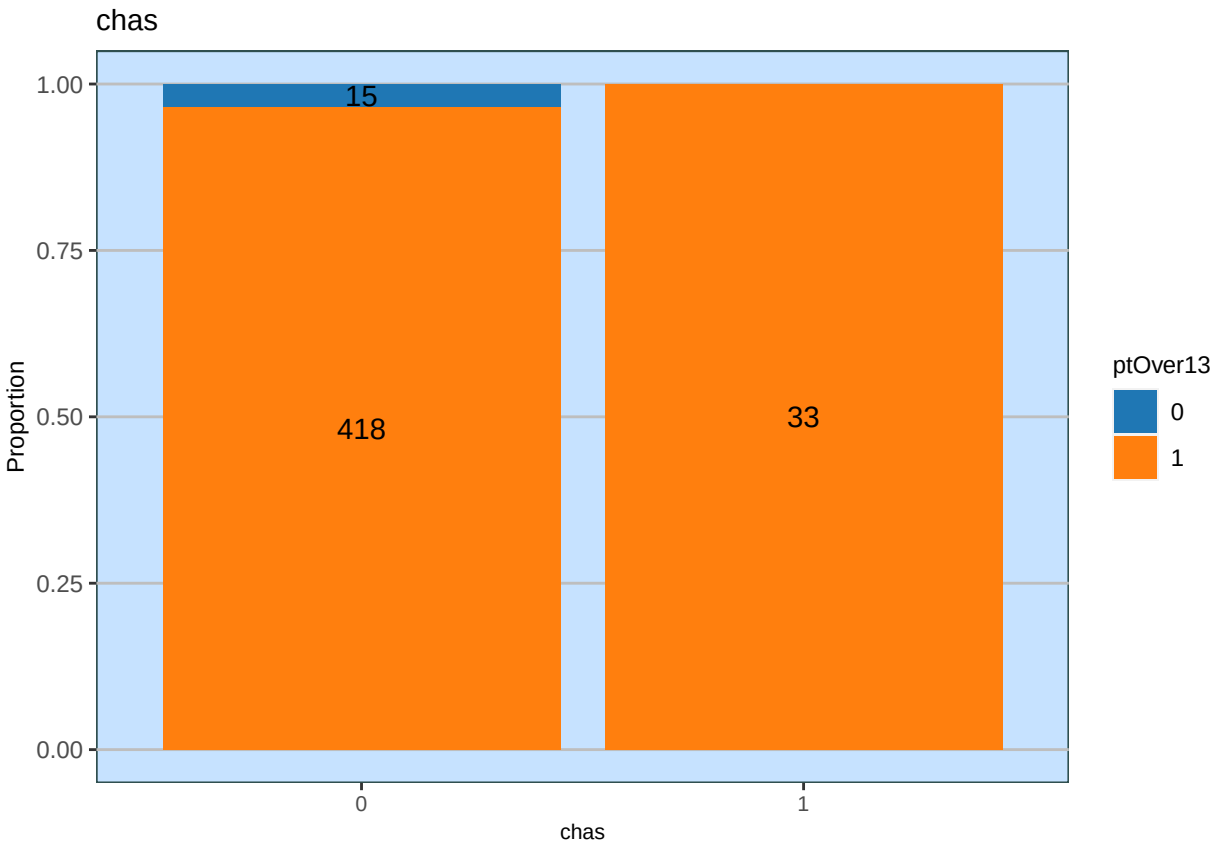
```
## [[1]]
```


zn
1.00
0.75
0.50
0.25
0.00
Proportion
3 3 9 10 1 1 12 9 4 9 8 3 6 3 3 3 7 6 3 3 4 3 3 3 10 2 2 4 4 1
0 12.5 17.5 18 20 21 22 25 28 30 33 34 35 40 45 52.5 55 60 70 75 80 82.5 85 90 95 100
zn
ptOver13
0
1

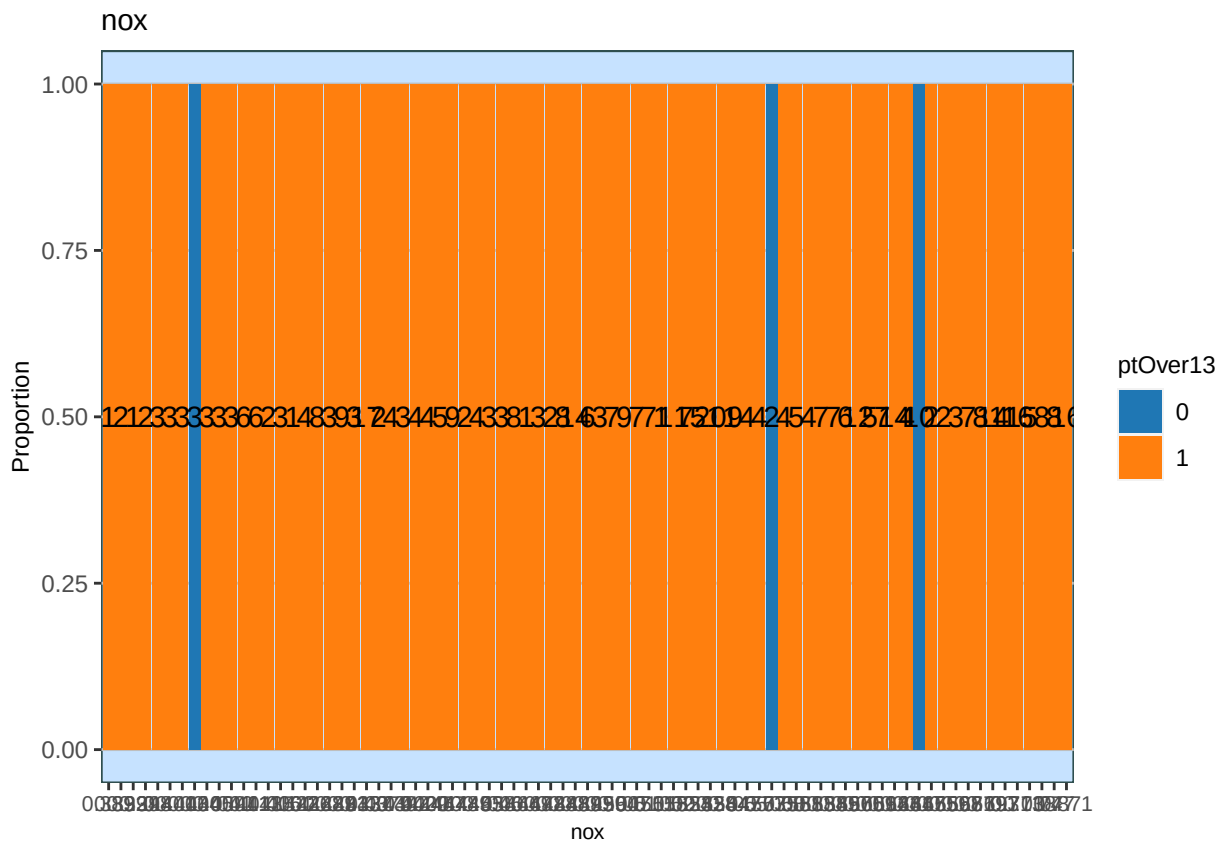
```
##
## [[2]]
```



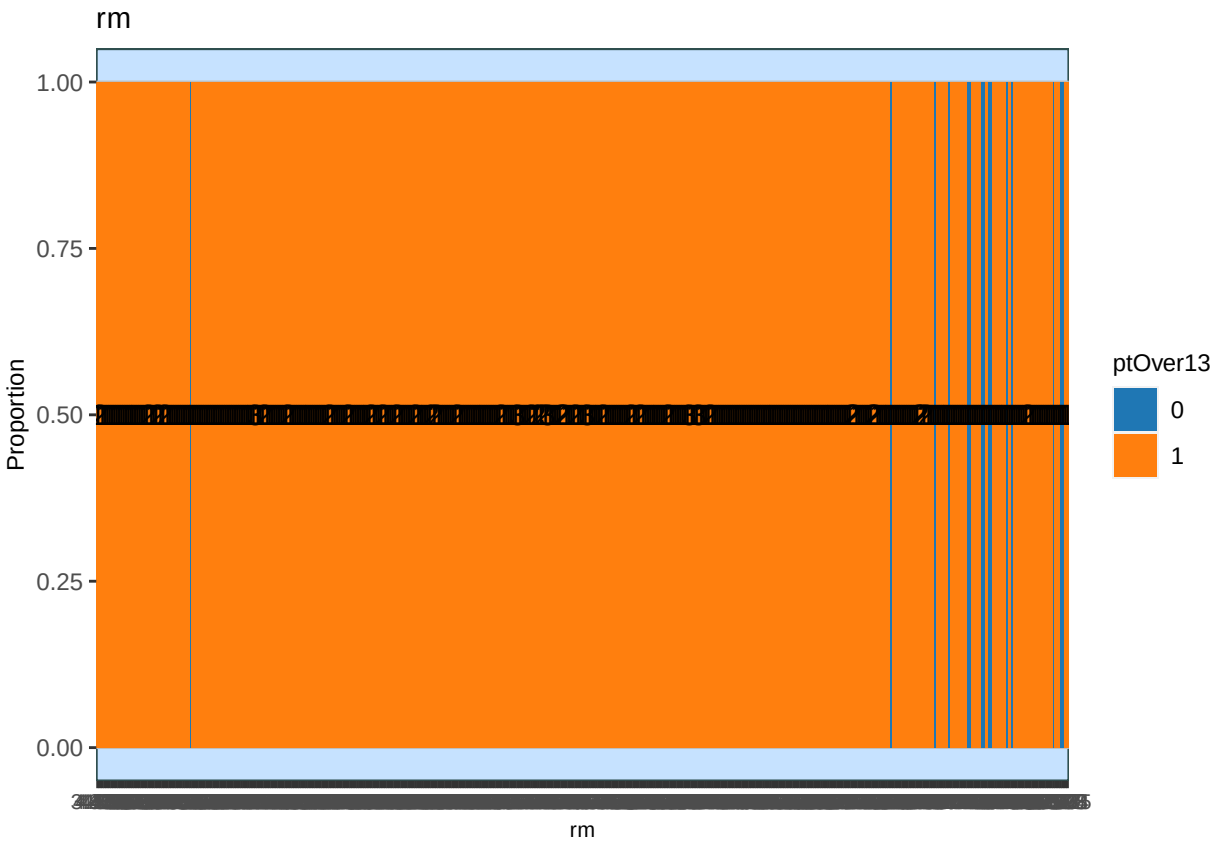

```
##
## [[3]]
```


chas
1.00
0.75
0.50
0.25
0.00
Proportion
15
418
33
0
1
chas
ptOver13
0
1

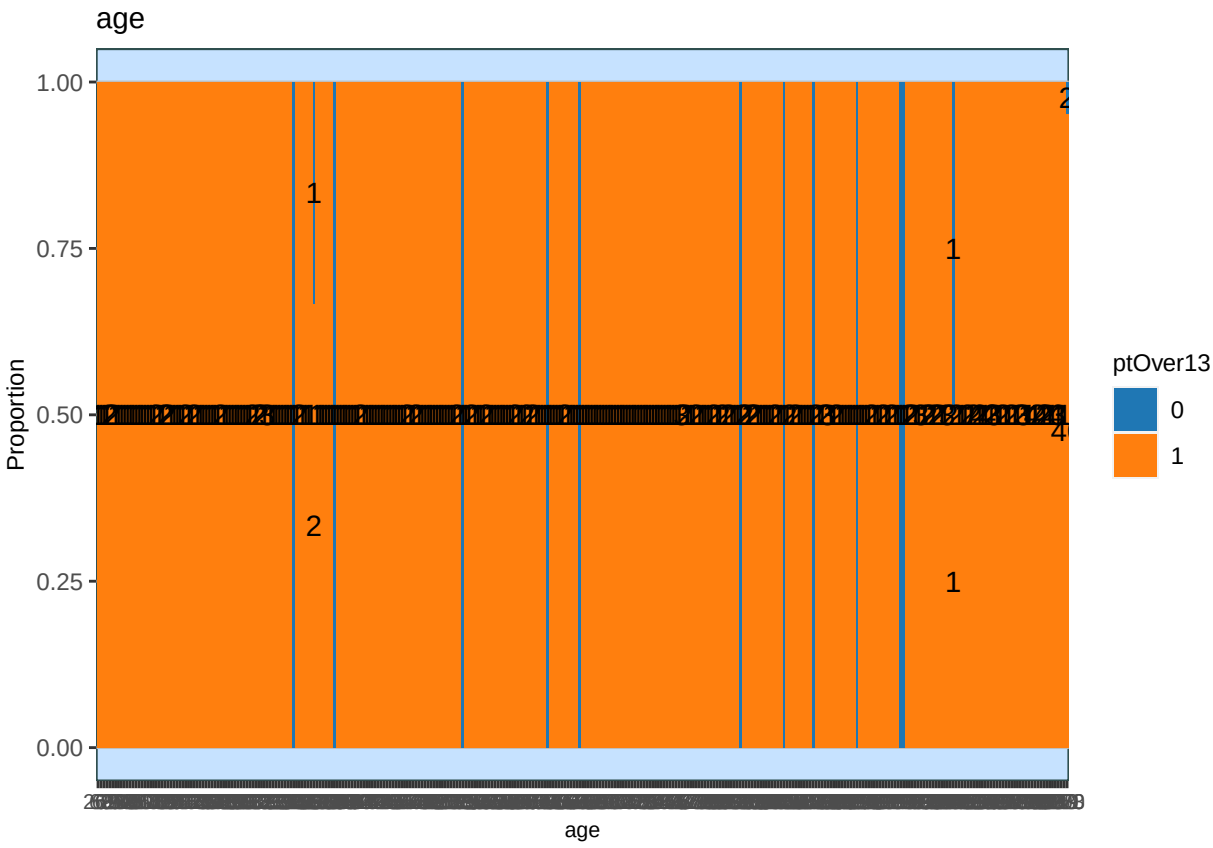
```
##
## [[4]]
```



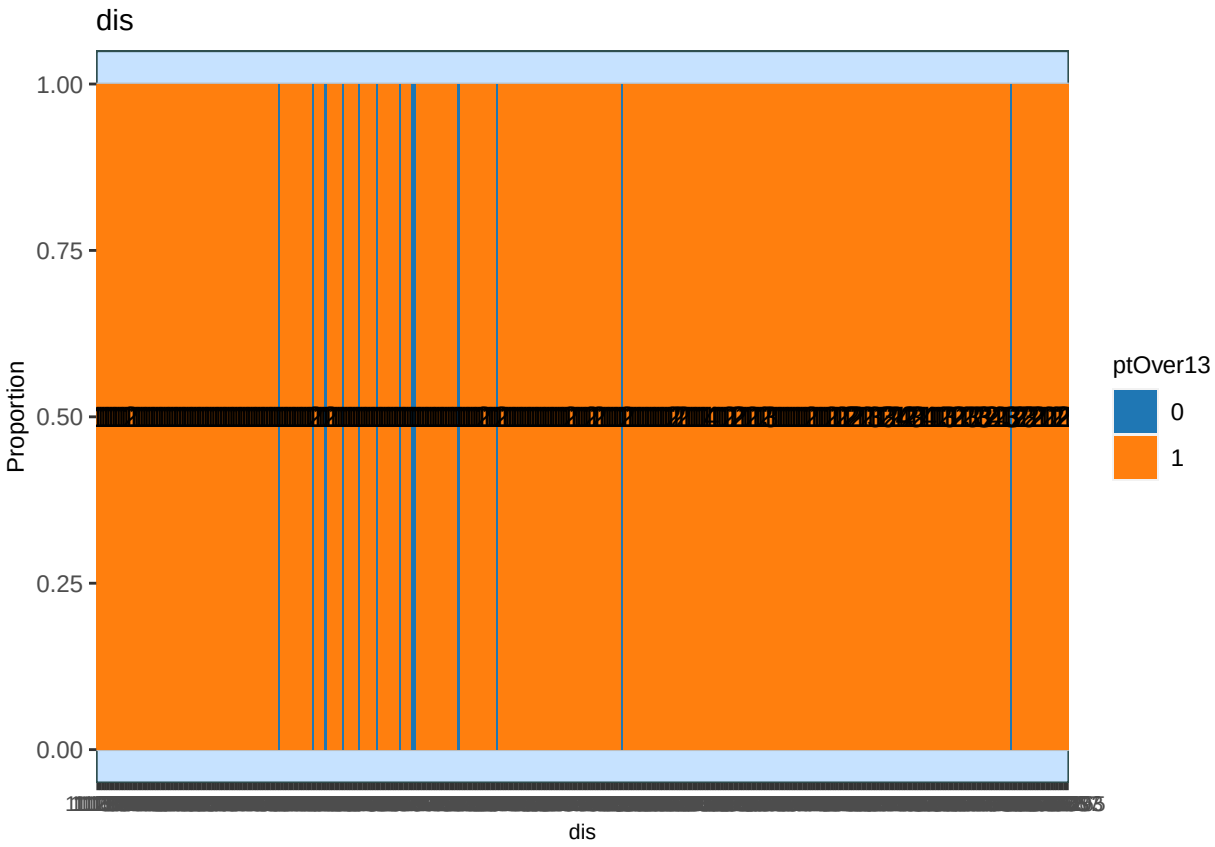

```
##
## [[5]]
```



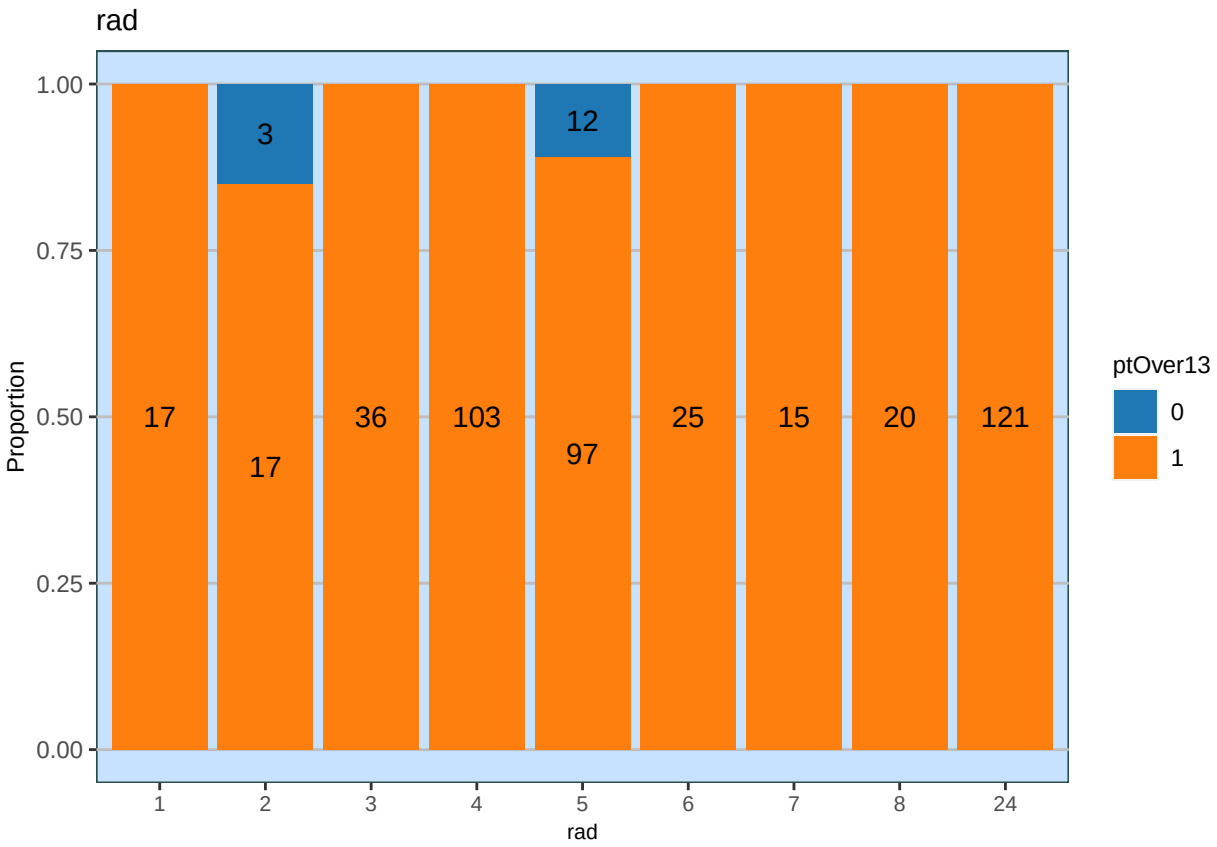

```
##
## [[6]]
```



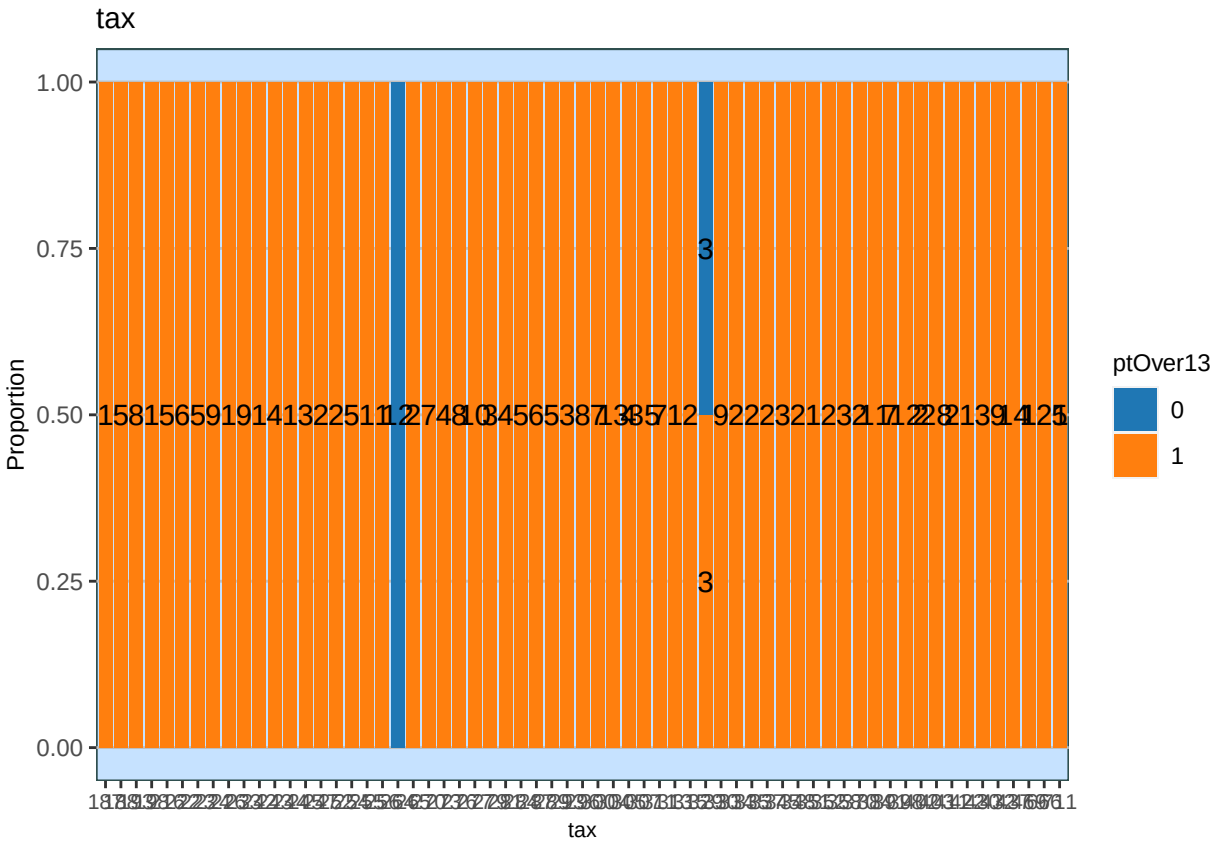

```
##
## [[7]]
```



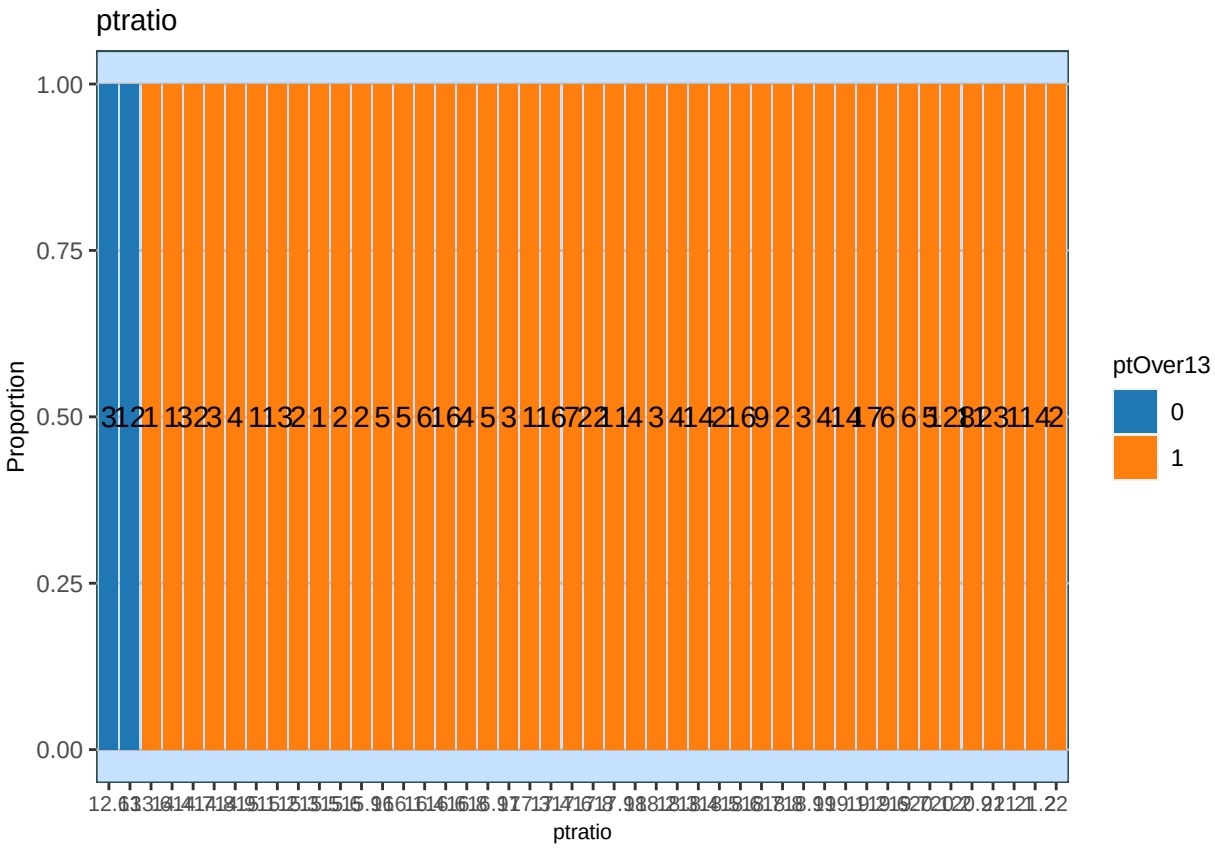

```
##
## [[8]]
```


rad
1.00
0.75
0.50
0.25
0.00
Proportion
17
3
17
36
103
12
97
25
15
20
121
1
2
3
4
5
6
7
8
24
rad
ptOver13
0
1

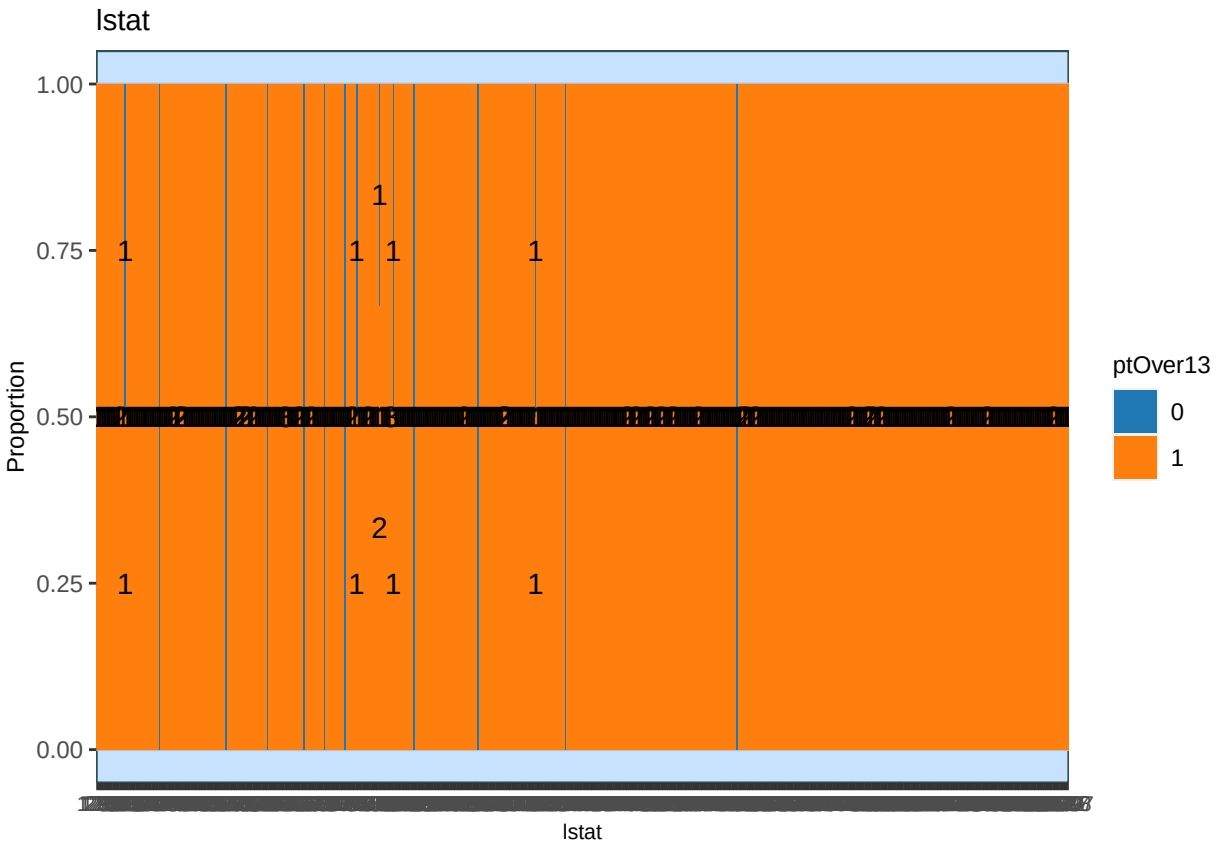
```
##
## [[9]]
```



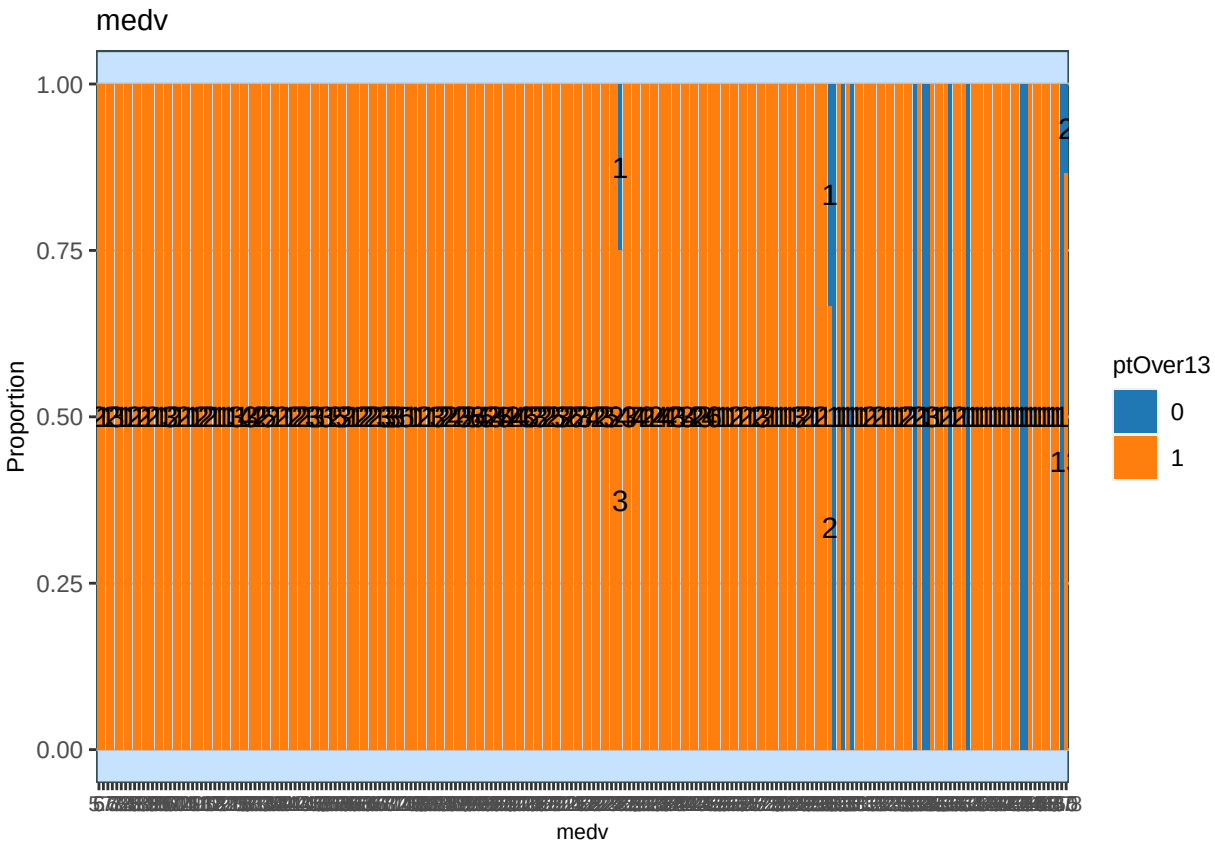

```
##
## [[10]]
```



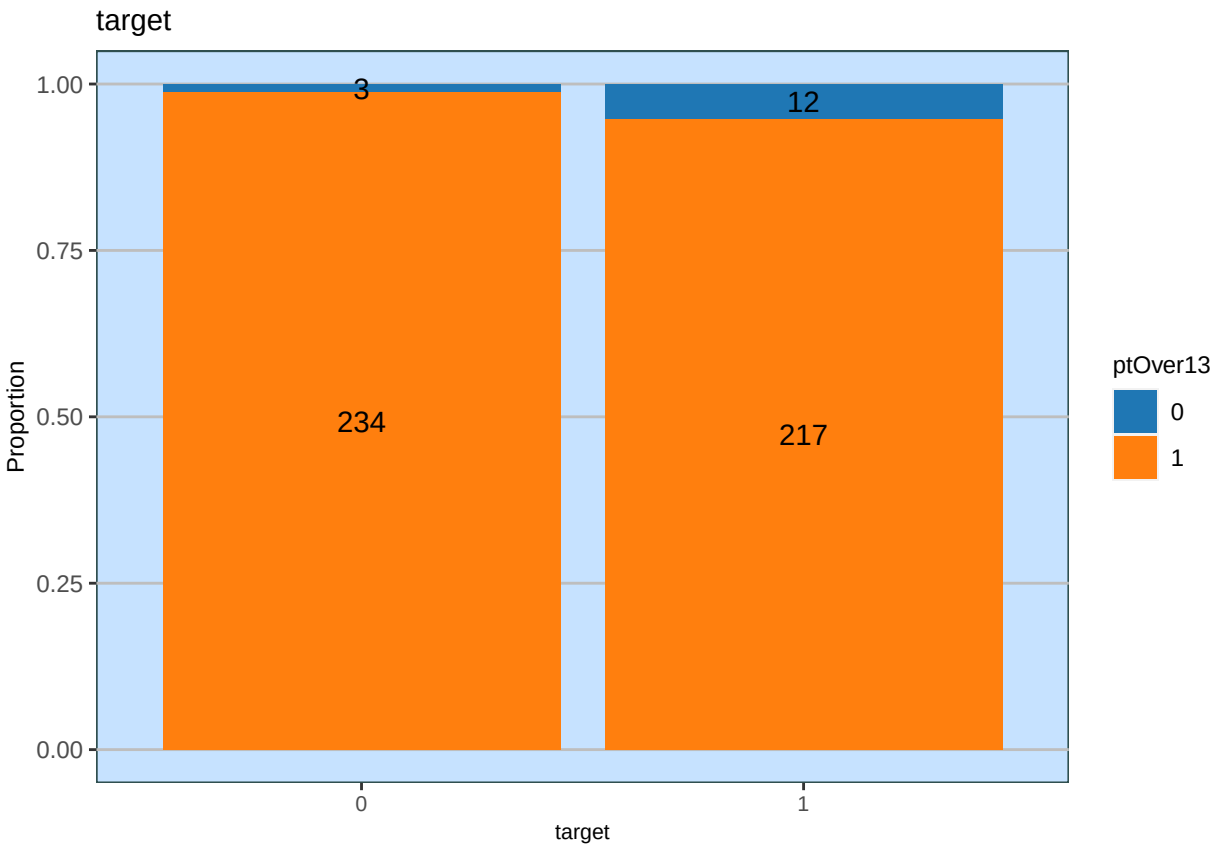

```
##
## [[11]]
```



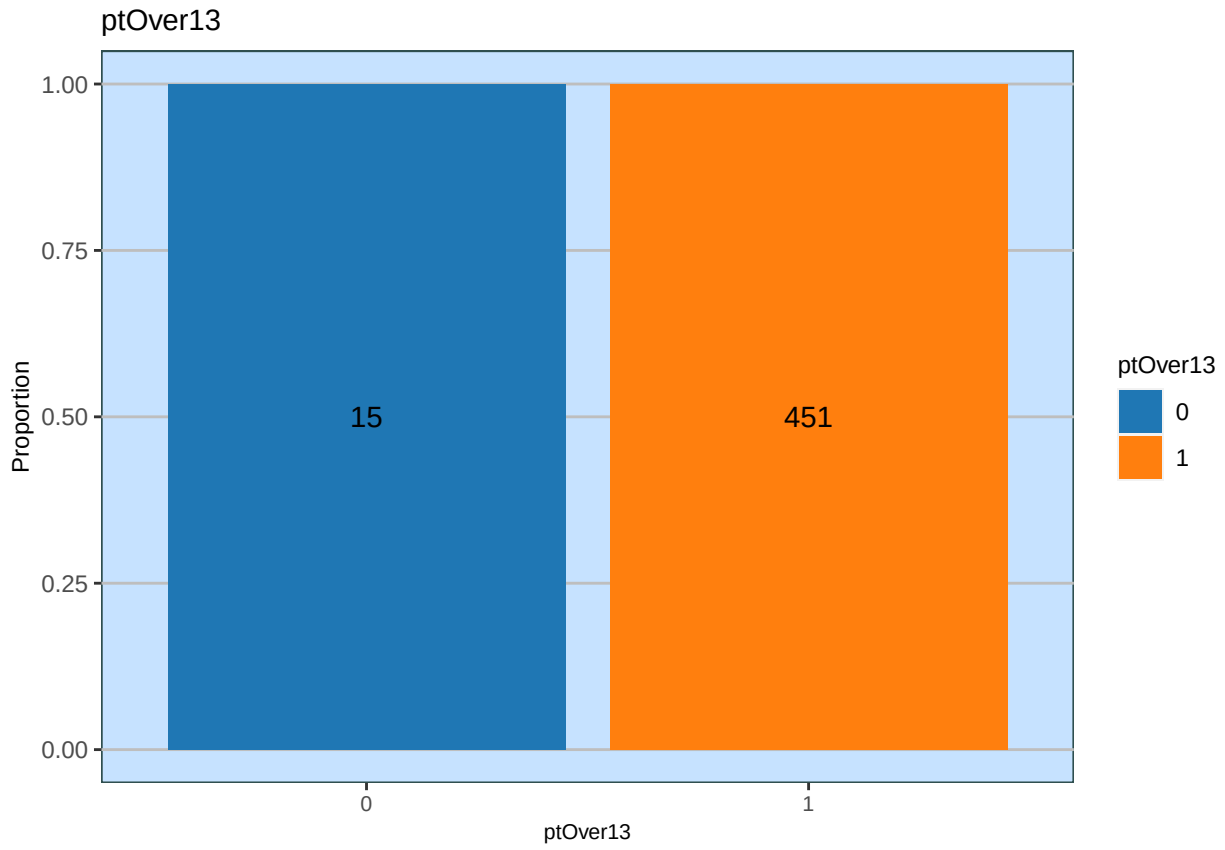

```
##
## [[12]]
```




```
##
## [[13]]
```


target
1.00
0.75
0.50
0.25
0.00
Proportion
3
12
234
217
0
1
target
ptOver13
0
1

```
##
## [[14]]
```




```
##       ptOver13
## target   0   1
##      0   3 234
##      1  12 217
```

```
## # A tibble: 466 x 4
## # Groups:   ptOver13 [2]
##    ptOver13 ave_lstat ave_medv ave_target
##       <dbl>     <dbl>    <dbl>      <dbl>
##  1        0      7.80     37.0        0.8
##  2        0      7.80     37.0        0.8
##  3        0      7.80     37.0        0.8
##  4        0      7.80     37.0        0.8
##  5        0      7.80     37.0        0.8
##  6        0      7.80     37.0        0.8
##  7        0      7.80     37.0        0.8
##  8        0      7.80     37.0        0.8
##  9        0      7.80     37.0        0.8
## 10        0      7.80     37.0        0.8
## # ... with 456 more rows
```

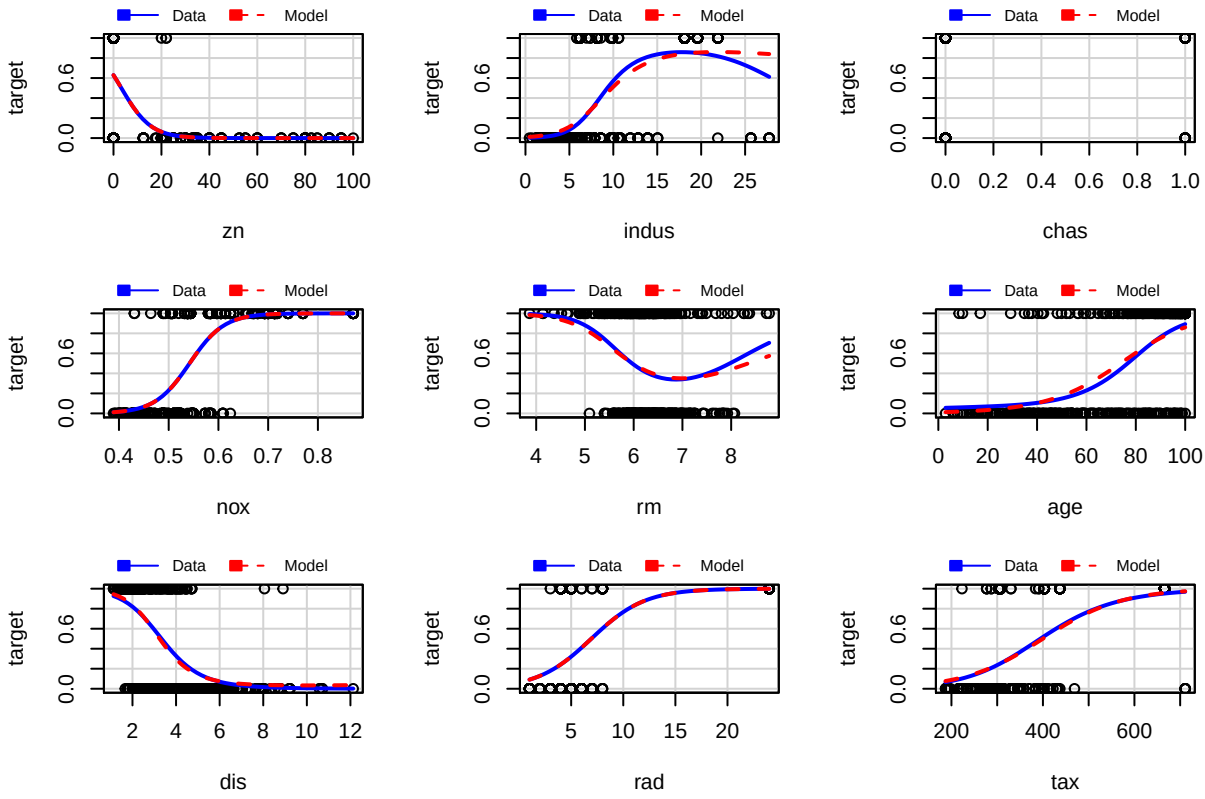
```
## 'data.frame':    466 obs. of  13 variables:
##  $ zn     : num  0 0 0 30 0 0 0 0 0 80 ...
##  $ indus  : num  19.58 19.58 18.1 4.93 2.46 ...
##  $ chas   : int  0 1 0 0 0 0 0 0 0 0 ...
```

```
##  $ nox    : num  0.605 0.871 0.74 0.428 0.488 0.52 0.693 0.693 0.515 0.392 ...
##  $ rm     : num  7.93 5.4 6.49 6.39 7.16 ...
##  $ age    : num  96.2 100 100 7.8 92.2 71.3 100 100 38.1 19.1 ...
##  $ dis    : num  2.05 1.32 1.98 7.04 2.7 ...
##  $ rad    : int  5 5 24 6 3 5 24 24 5 1 ...
##  $ tax    : int  403 403 666 300 193 384 666 666 224 315 ...
##  $ ptratio: num  14.7 14.7 20.2 16.6 17.8 20.9 20.2 20.2 20.2 16.4 ...
##  $ lstat  : num  3.7 26.82 18.85 5.19 4.82 ...
##  $ medv   : num  50 13.4 15.4 23.7 37.9 26.5 5 7 22.2 20.9 ...
##  $ target : int  1 1 1 0 0 0 1 1 0 0 ...
```

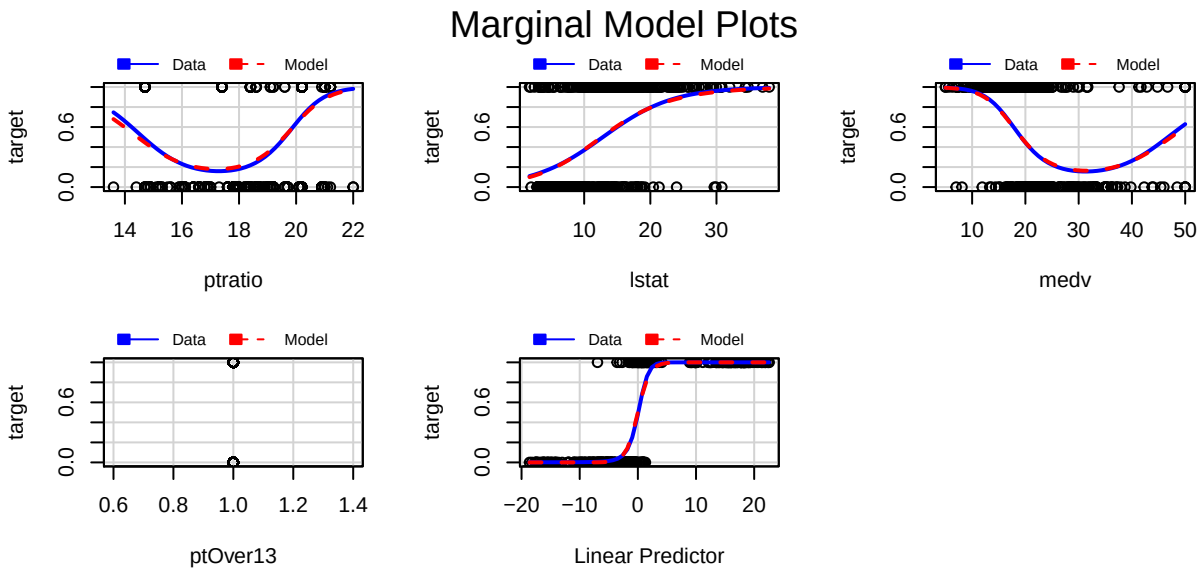
```
##
## Call:  glm(formula = target ~ ., family = "binomial", data = df)
##
## Coefficients:
## (Intercept)           zn        indus         chas          nox           rm
##  -40.822934    -0.065946    -0.064614     0.910765    49.122297    -0.587488
##         age          dis          rad          tax      ptratio        lstat
##    0.034189     0.738660     0.666366    -0.006171     0.402566     0.045869
##        medv
##    0.180824
##
## Degrees of Freedom: 465 Total (i.e. Null);  453 Residual
## Null Deviance:       645.9
## Residual Deviance: 192   AIC: 218
```

```
##
## Call:  glm(formula = target ~ ., family = "binomial", data = df23)
##
## Coefficients:
## (Intercept)           zn        indus         chas          nox           rm
##  -37.333521    -0.129787     0.004314     0.749600    38.875661    -0.716138
##         age          dis          rad          tax      ptratio        lstat
##    0.039393     0.878725     0.758820    -0.005851     0.450693     0.034661
##        medv     ptOver13
##    0.178087           NA
##
## Degrees of Freedom: 450 Total (i.e. Null);  438 Residual
## Null Deviance:       624.6
## Residual Deviance: 186.3     AIC: 212.3
```

```
## Error in smooth.construct.tp.smooth.spec(object, dk$data, dk$knots) :
##   A term has fewer unique covariate combinations than specified maximum degrees of freedom
## Error in smooth.construct.tp.smooth.spec(object, dk$data, dk$knots) :
##   A term has fewer unique covariate combinations than specified maximum degrees of freedom
```

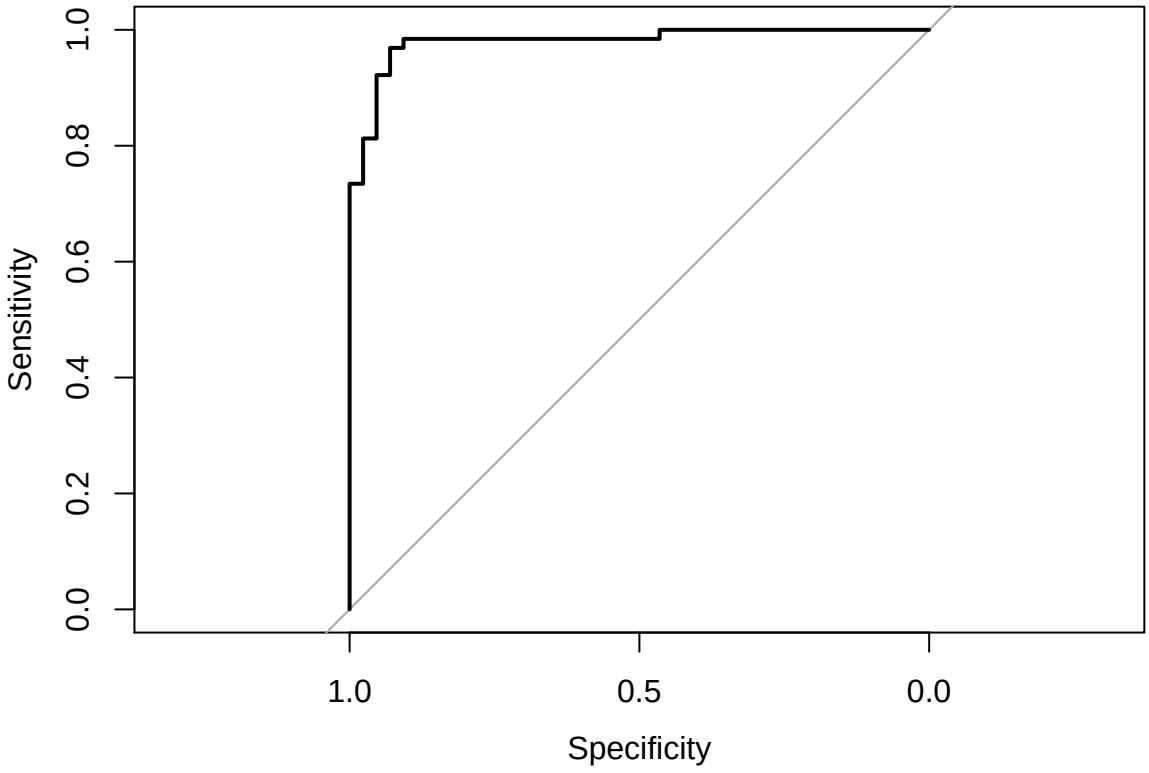



```
## Error in smooth.construct.tp.smooth.spec(object, dk$data, dk$knots) :
##   A term has fewer unique covariate combinations than specified maximum degrees of freedom
## Error in smooth.construct.tp.smooth.spec(object, dk$data, dk$knots) :
##   A term has fewer unique covariate combinations than specified maximum degrees of freedom
```




```
##
## Call:  glm(formula = fla, family = "binomial", data = df)
##
## Coefficients:
## (Intercept)           zn        indus         chas          nox           rm
##      2.3290      -1.5408      -0.4423       0.2339       5.7309      -0.4141
##         age          dis          rad          tax      ptratio        lstat
##      0.9683       1.5563       5.7880      -1.0362       0.8844       0.3258
##        medv
##      1.6708
##
## Degrees of Freedom: 465 Total (i.e. Null);  453 Residual
## Null Deviance:       645.9
## Residual Deviance: 192   AIC: 218
##
## Call:
## glm(formula = fla, family = "binomial", data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8464  -0.1445  -0.0017   0.0029   3.4665
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.3290     0.7195   3.237  0.00121 **
## zn           -1.5408     0.8097  -1.903  0.05706 .
```

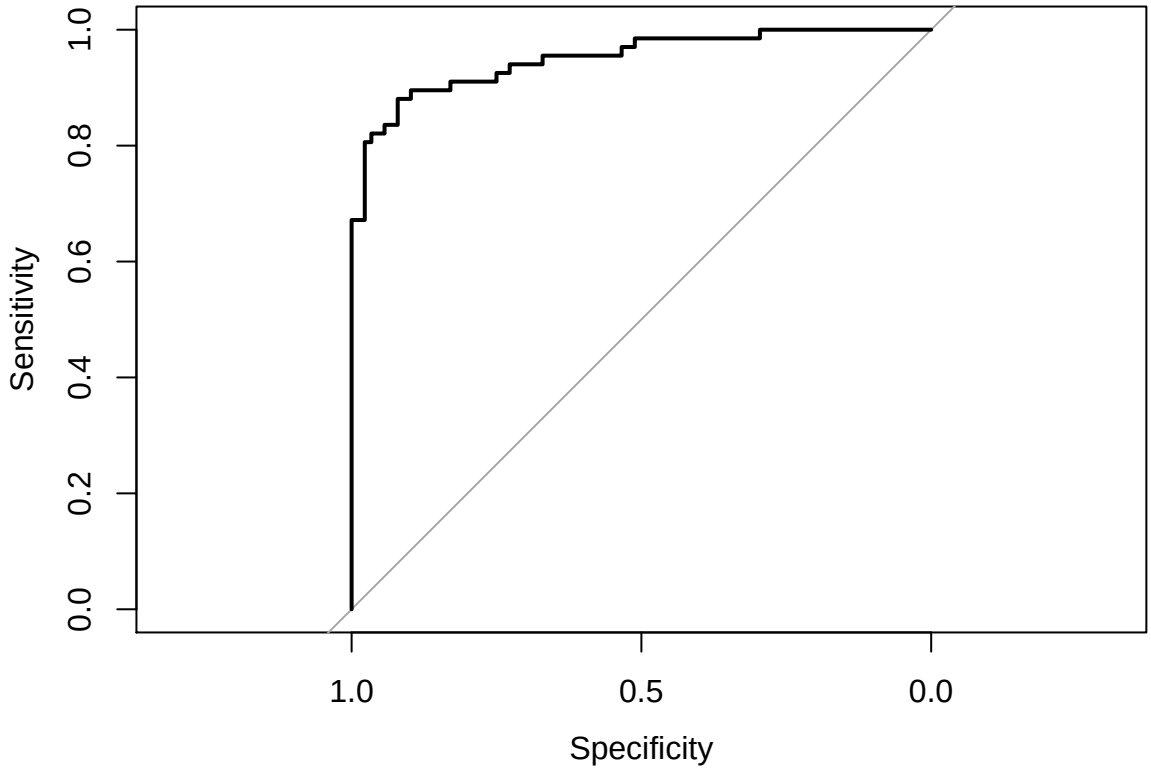
```
## indus        -0.4423     0.3260  -1.357  0.17485
## chas          0.2339     0.1940   1.205  0.22803
## nox           5.7309     0.9254   6.193 5.90e-10 ***
## rm           -0.4141     0.5095  -0.813  0.41637
## age           0.9683     0.3912   2.475  0.01333 *
## dis           1.5563     0.4852   3.208  0.00134 **
## rad           5.7880     1.4171   4.084 4.42e-05 ***
## tax          -1.0362     0.4961  -2.089  0.03674 *
## ptratio       0.8844     0.2782   3.179  0.00148 **
## lstat         0.3258     0.3838   0.849  0.39608
## medv          1.6708     0.6310   2.648  0.00810 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 645.88  on 465  degrees of freedom
## Residual deviance: 192.05  on 453  degrees of freedom
## AIC: 218.05
##
## Number of Fisher Scoring iterations: 9
##
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 40  5
##          1  3 59
##
##                Accuracy : 0.9252
##                  95% CI : (0.858, 0.9672)
##     No Information Rate : 0.5981
##     P-Value [Acc > NIR] : 2.008e-14
##
##                   Kappa : 0.8457
##
##  Mcnemar's Test P-Value : 0.7237
##
##             Sensitivity : 0.9302
##             Specificity : 0.9219
##          Pos Pred Value : 0.8889
##          Neg Pred Value : 0.9516
##              Prevalence : 0.4019
##          Detection Rate : 0.3738
##    Detection Prevalence : 0.4206
##       Balanced Accuracy : 0.9261
##
##        'Positive' Class : 0
##
```




```
## [1] "AUC:  0.980014534883721"
##
## Call:
## roc.default(response = dfPred_raw$class, predictor = dfPred_raw$predict_reg,     plot = TRUE)
##
## Data: dfPred_raw$predict_reg in 43 controls (dfPred_raw$class 0) < 64 cases (dfPred_raw$class 1).
## Area under the curve: 0.98

##
## Call:  glm(formula = fla, family = "binomial", data = df)
##
## Coefficients:
## (Intercept)           zn        indus         chas          nox           rm
##      2.3290      -1.5408      -0.4423       0.2339       5.7309      -0.4141
##         age          dis          rad          tax      ptratio        lstat
##      0.9683       1.5563       5.7880      -1.0362       0.8844       0.3258
##        medv
##      1.6708
##
## Degrees of Freedom: 465 Total (i.e. Null);  453 Residual
## Null Deviance:       645.9
## Residual Deviance: 192   AIC: 218

##
## Call:  glm(formula = fla, family = "binomial", data = df)
```

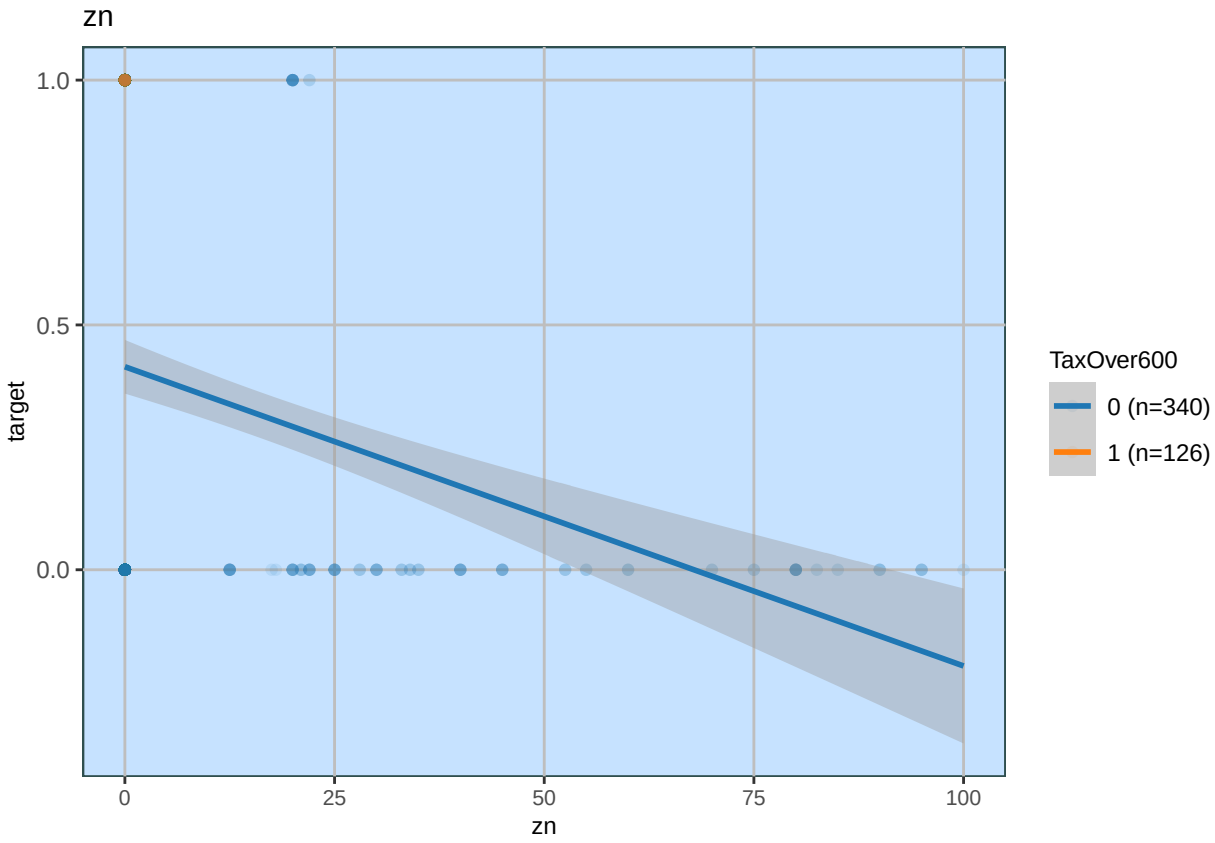
```
##
## Coefficients:
## (Intercept)          nox          dis          rad      ptratio         medv
##   -31.27121     37.37652      0.29535      0.51558      0.28586      0.08635
##
## Degrees of Freedom: 465 Total (i.e. Null);  460 Residual
## Null Deviance:       645.9
## Residual Deviance: 225.3     AIC: 237.3
##
## Call:
## glm(formula = fla, family = "binomial", data = df)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.06137  -0.31295  -0.04733   0.00705   2.81210
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -31.27121    4.82619  -6.479 9.20e-11 ***
## nox          37.37652    5.56582   6.715 1.88e-11 ***
## dis           0.29535    0.14902   1.982  0.04748 *
## rad           0.51558    0.11531   4.471 7.77e-06 ***
## ptratio       0.28586    0.09877   2.894  0.00380 **
## medv          0.08635    0.02832   3.050  0.00229 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 645.88  on 465  degrees of freedom
## Residual deviance: 225.32  on 460  degrees of freedom
## AIC: 237.32
##
## Number of Fisher Scoring iterations: 8
##
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 83 12
##          1  5 55
##
##                Accuracy : 0.8903
##                  95% CI : (0.8302, 0.9348)
##     No Information Rate : 0.5677
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.7737
##
##  Mcnemar's Test P-Value : 0.1456
##
##             Sensitivity : 0.9432
##             Specificity : 0.8209
##          Pos Pred Value : 0.8737
```

```
##          Neg Pred Value : 0.9167
##              Prevalence : 0.5677
##          Detection Rate : 0.5355
##    Detection Prevalence : 0.6129
##       Balanced Accuracy : 0.8820
##
##        'Positive' Class : 0
##
```

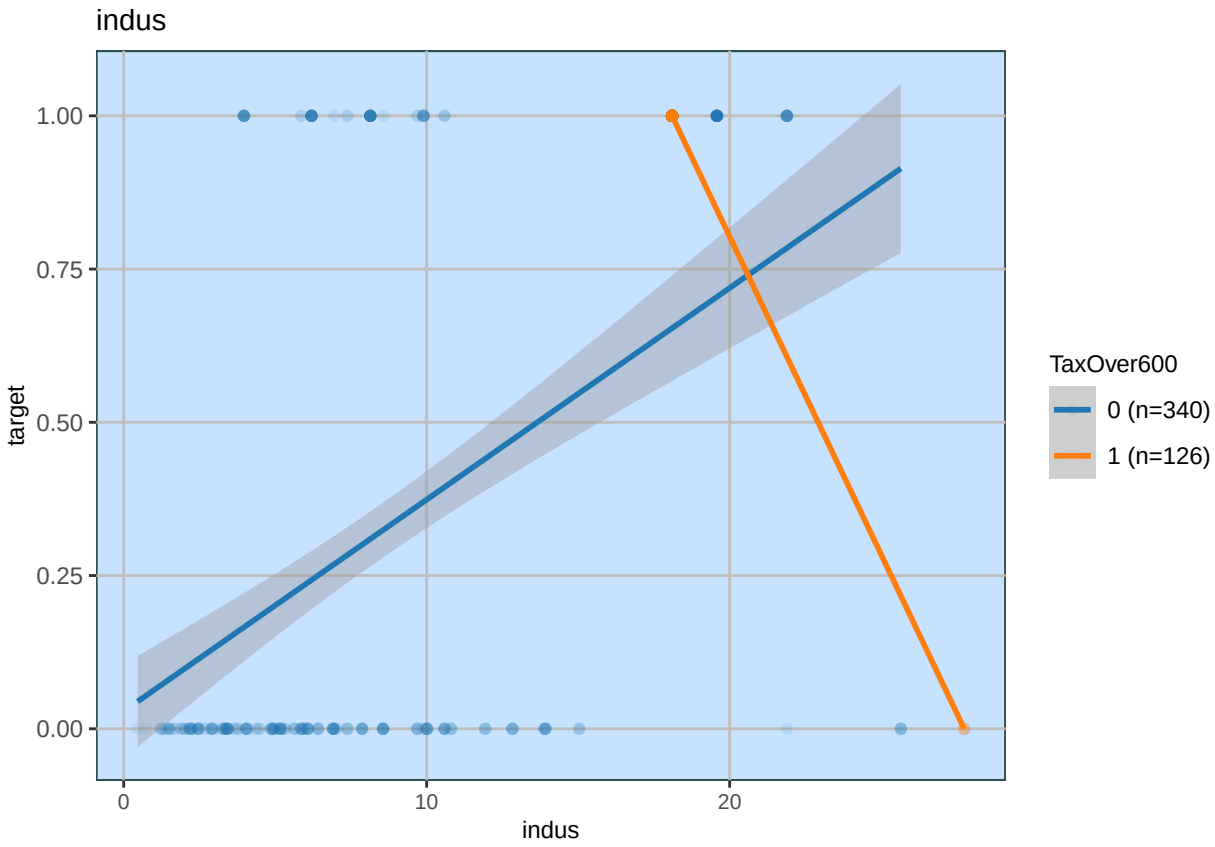



```
## [1] "AUC:  0.950474898236092"
##
## Call:
## roc.default(response = dfPred_raw$class, predictor = dfPred_raw$predict_reg,     plot = TRUE)
##
## Data: dfPred_raw$predict_reg in 88 controls (dfPred_raw$class 0) < 67 cases (dfPred_raw$class 1).
## Area under the curve: 0.9505

##
## Call:  glm(formula = fla, family = "binomial", data = df)
##
## Coefficients:
## (Intercept)          nox          dis          rad      ptratio         medv
##   -31.27121     37.37652      0.29535      0.51558      0.28586      0.08635
##
## Degrees of Freedom: 465 Total (i.e. Null);  460 Residual
```

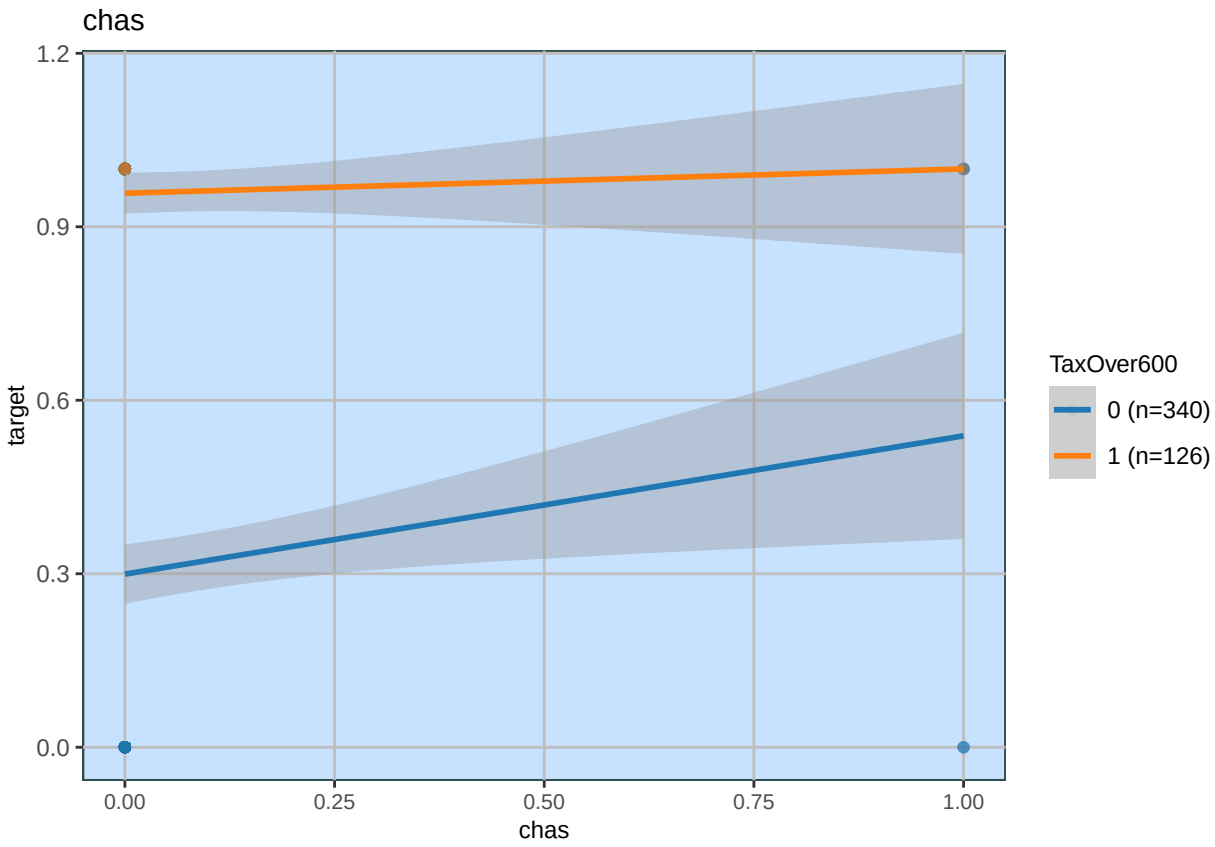
```
## Null Deviance:       645.9
## Residual Deviance: 225.3     AIC: 237.3

## [[1]]
```



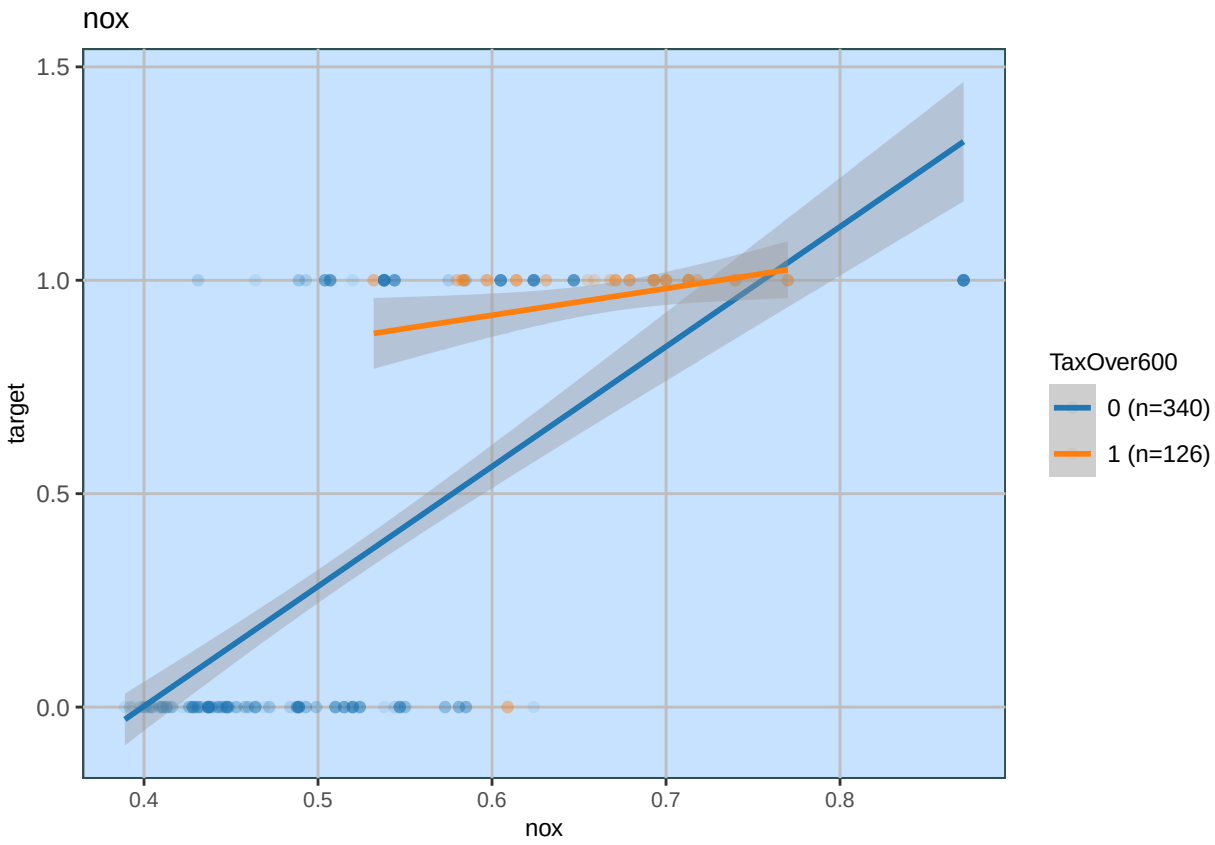

```
##
## [[2]]
```


indus
1.00
0.75
0.50
0.25
0.00
target
0
10
20
indus
TaxOver600
0 (n=340)
1 (n=126)

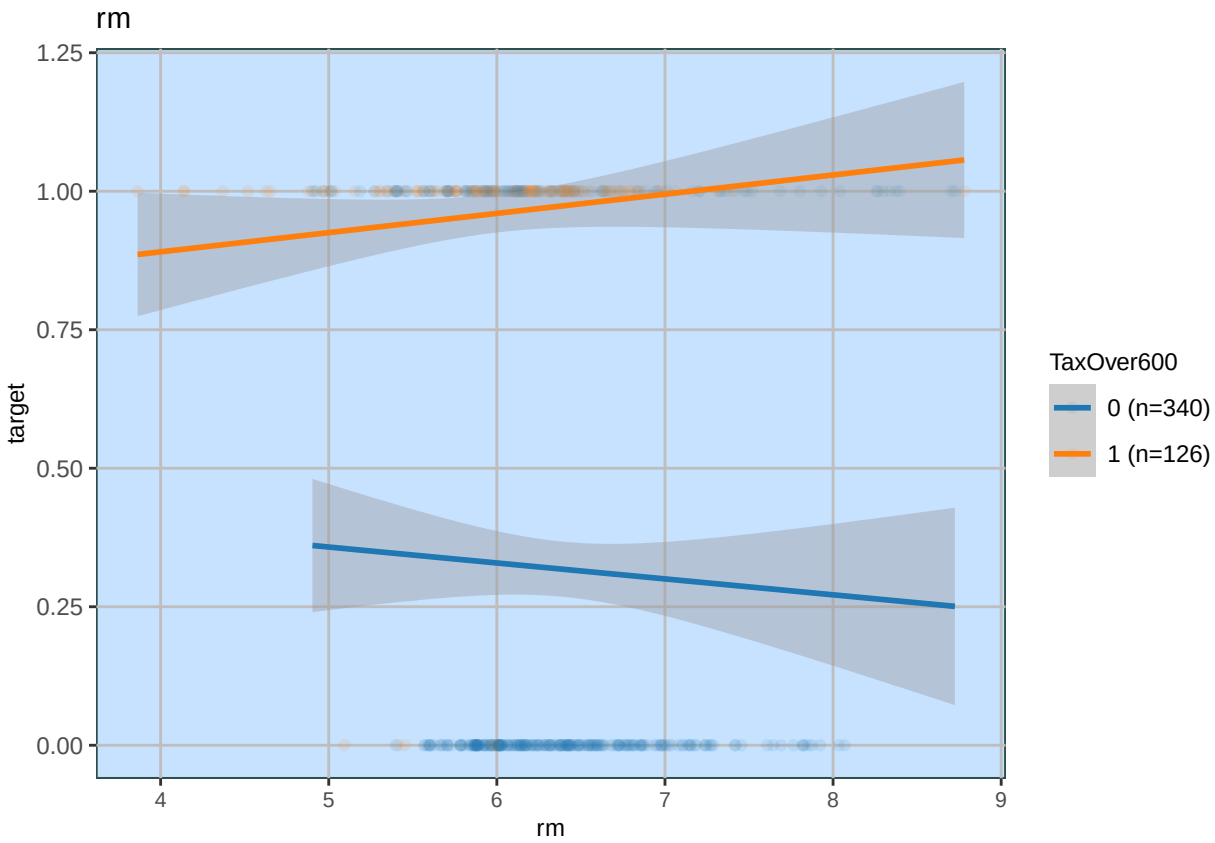
```
##
## [[3]]
```


chas
1.2
0.9
0.6
0.3
0.0
target
0.00
0.25
0.50
0.75
1.00
chas
TaxOver600
0 (n=340)
1 (n=126)

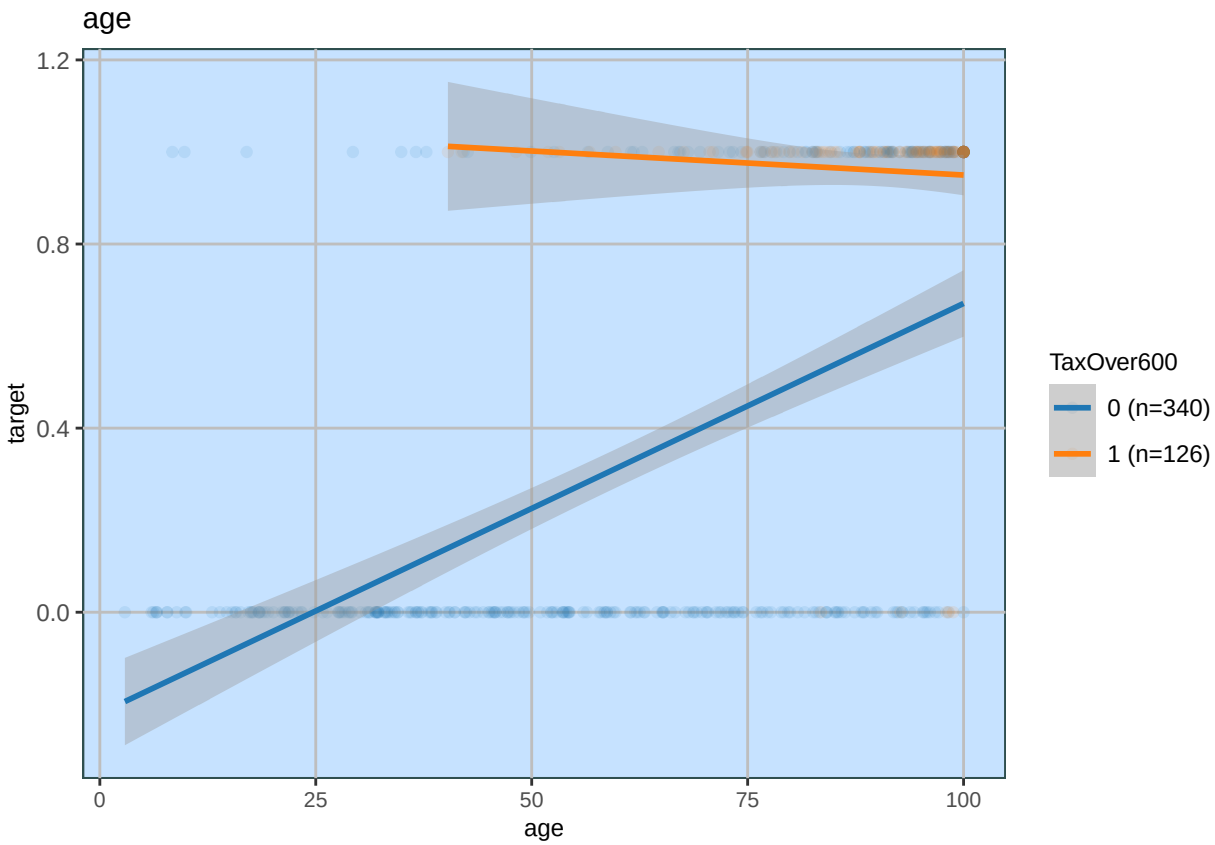
```
##
## [[4]]
```


nox
1.5
1.0
0.5
0.0
target
0.4
0.5
0.6
0.7
0.8
nox
TaxOver600
0 (n=340)
1 (n=126)

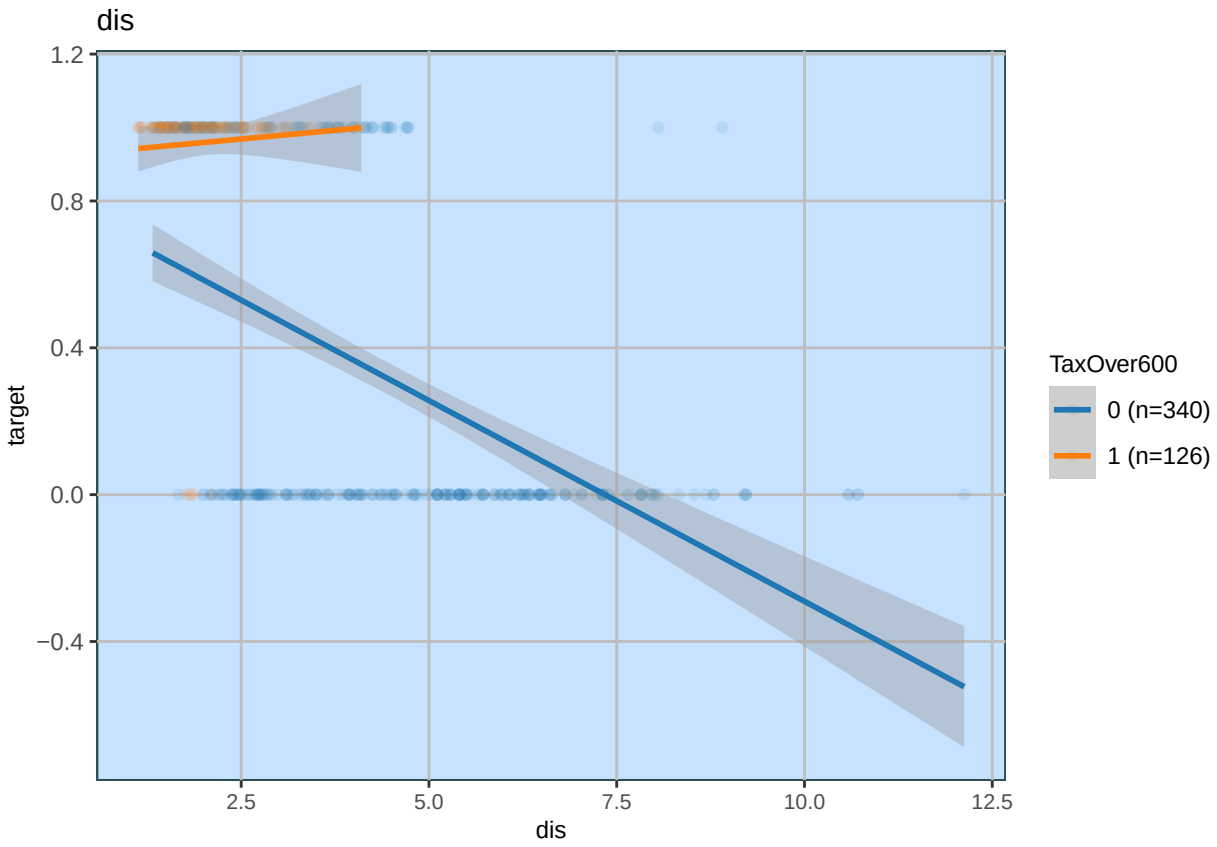
```
##
## [[5]]
```


rm
1.25
1.00
0.75
0.50
0.25
0.00
target
4
5
6
7
8
9
rm
TaxOver600
0 (n=340)
1 (n=126)

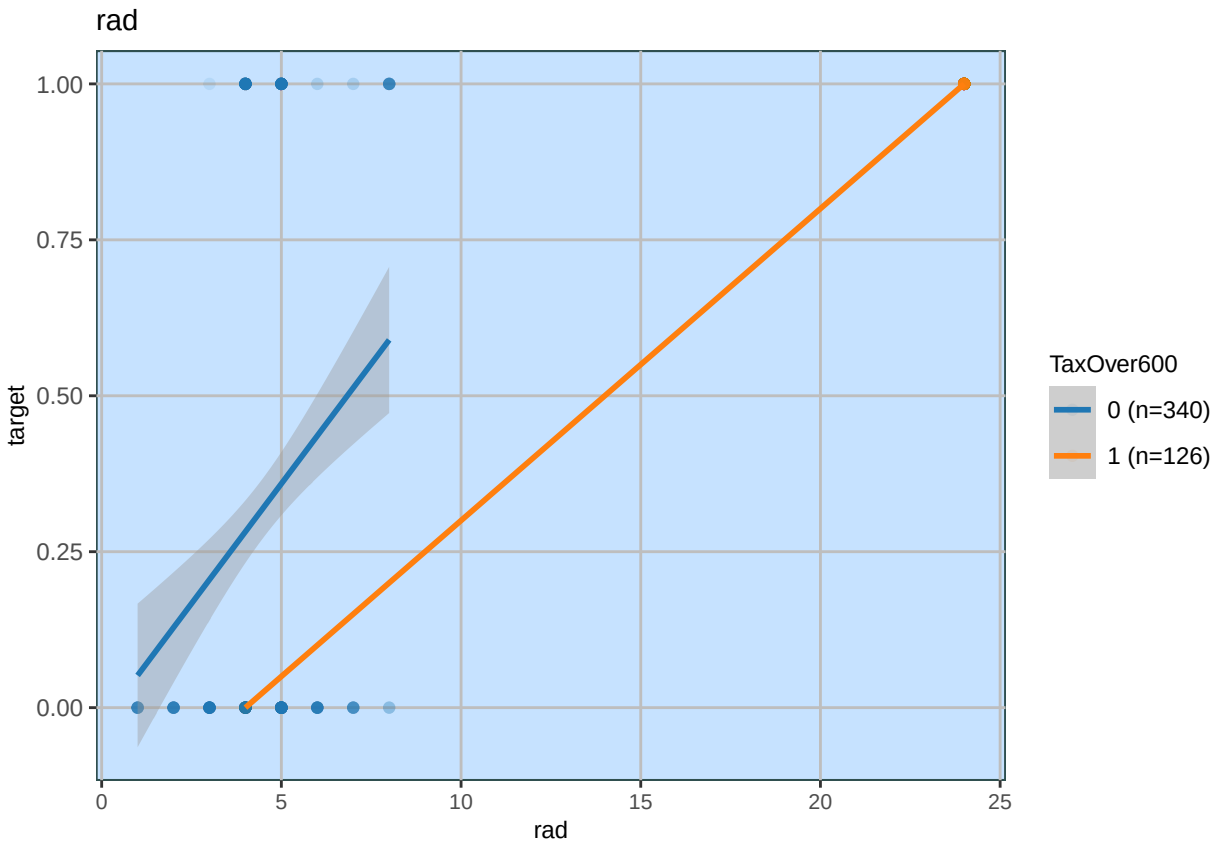
```
##
## [[6]]
```


age
1.2
0.8
0.4
0.0
target
0
25
50
75
100
age
TaxOver600
0 (n=340)
1 (n=126)

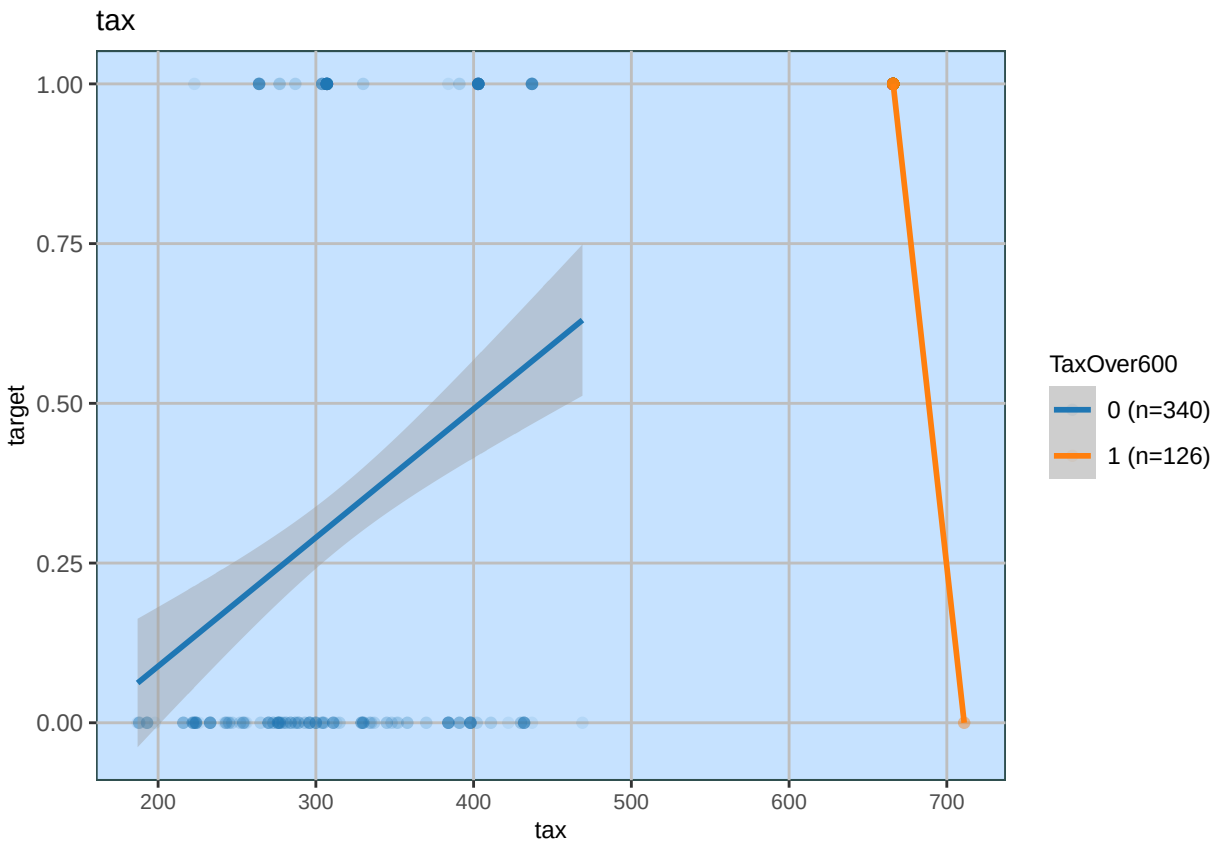
```
##
## [[7]]
```


dis
1.2
0.8
0.4
0.0
−0.4
target
2.5
5.0
7.5
10.0
12.5
dis
TaxOver600
0 (n=340)
1 (n=126)

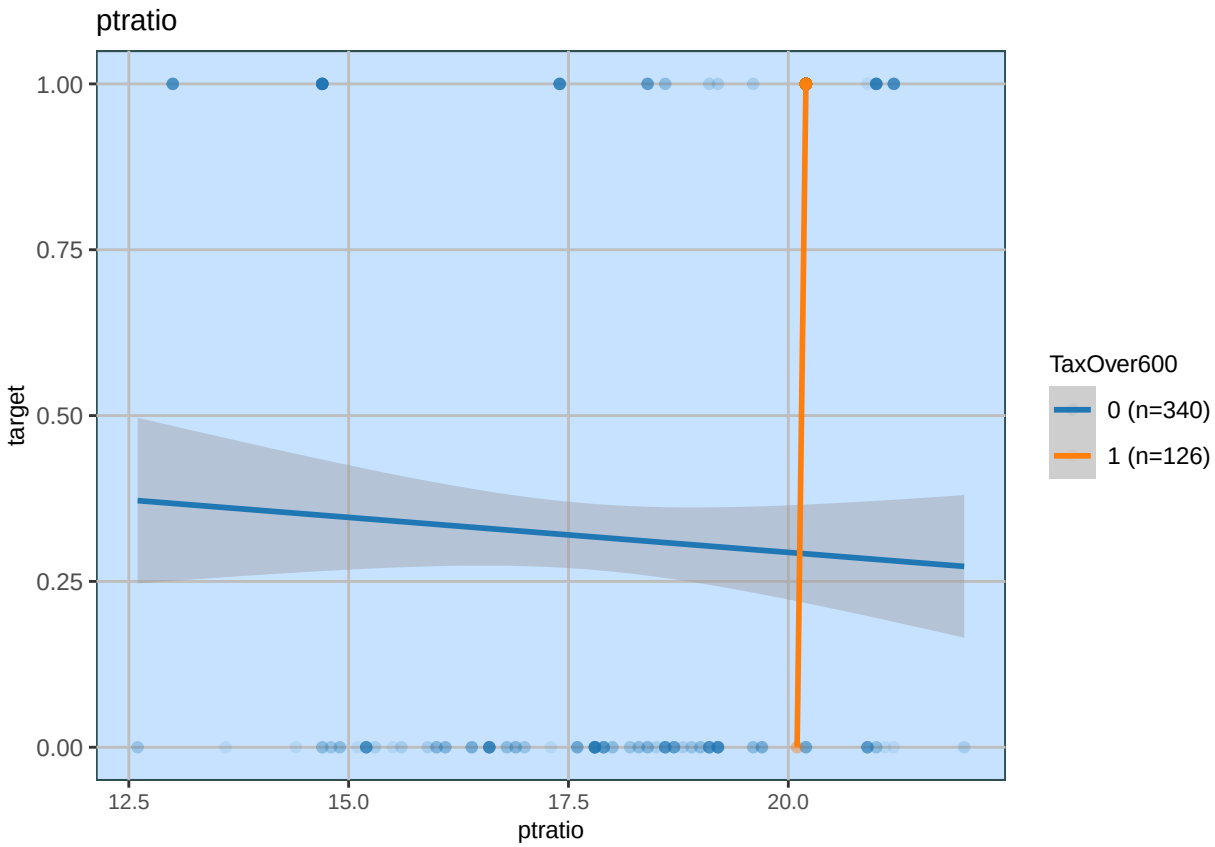
```
##
## [[8]]
```


rad
1.00
0.75
0.50
0.25
0.00
target
0
5
10
15
20
25
rad
TaxOver600
0 (n=340)
1 (n=126)

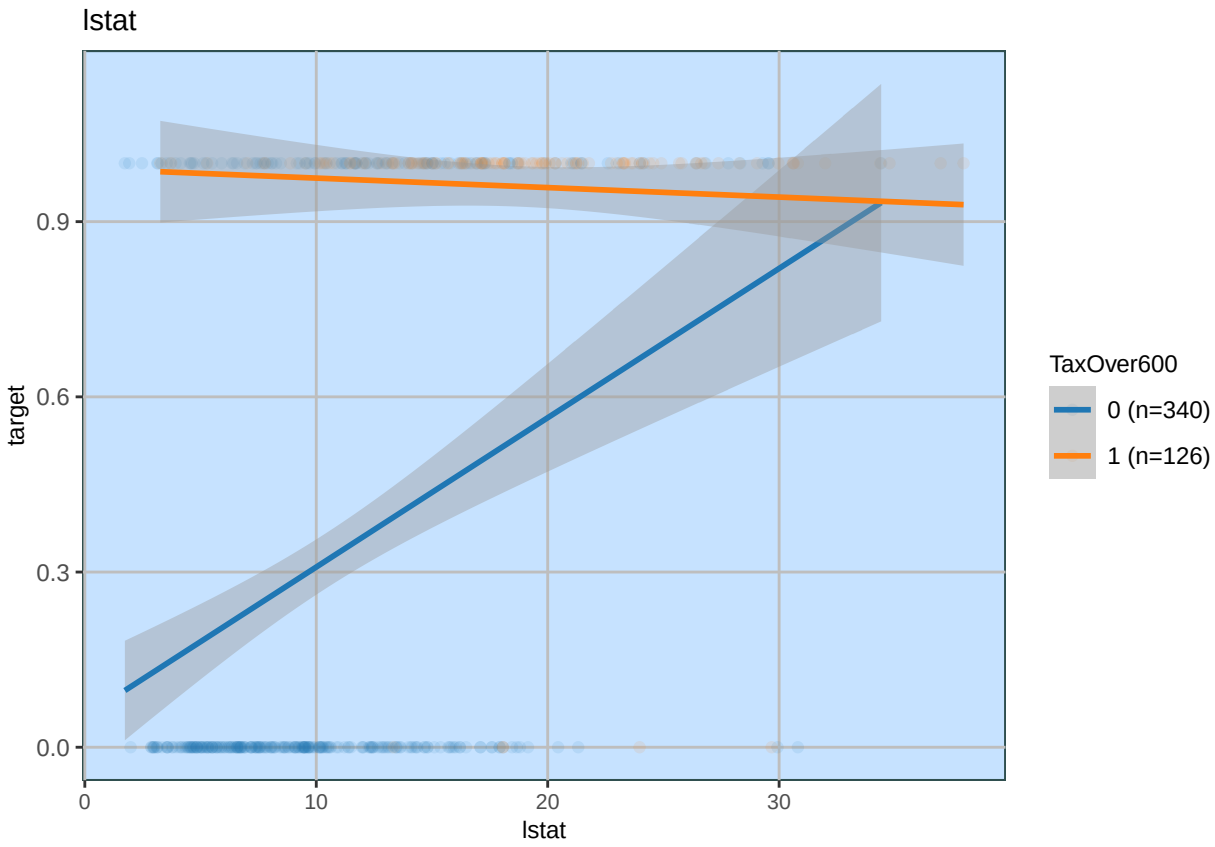
```
##
## [[9]]
```


tax
1.00
0.75
0.50
0.25
0.00
target
200
300
400
500
600
700
tax
TaxOver600
0 (n=340)
1 (n=126)

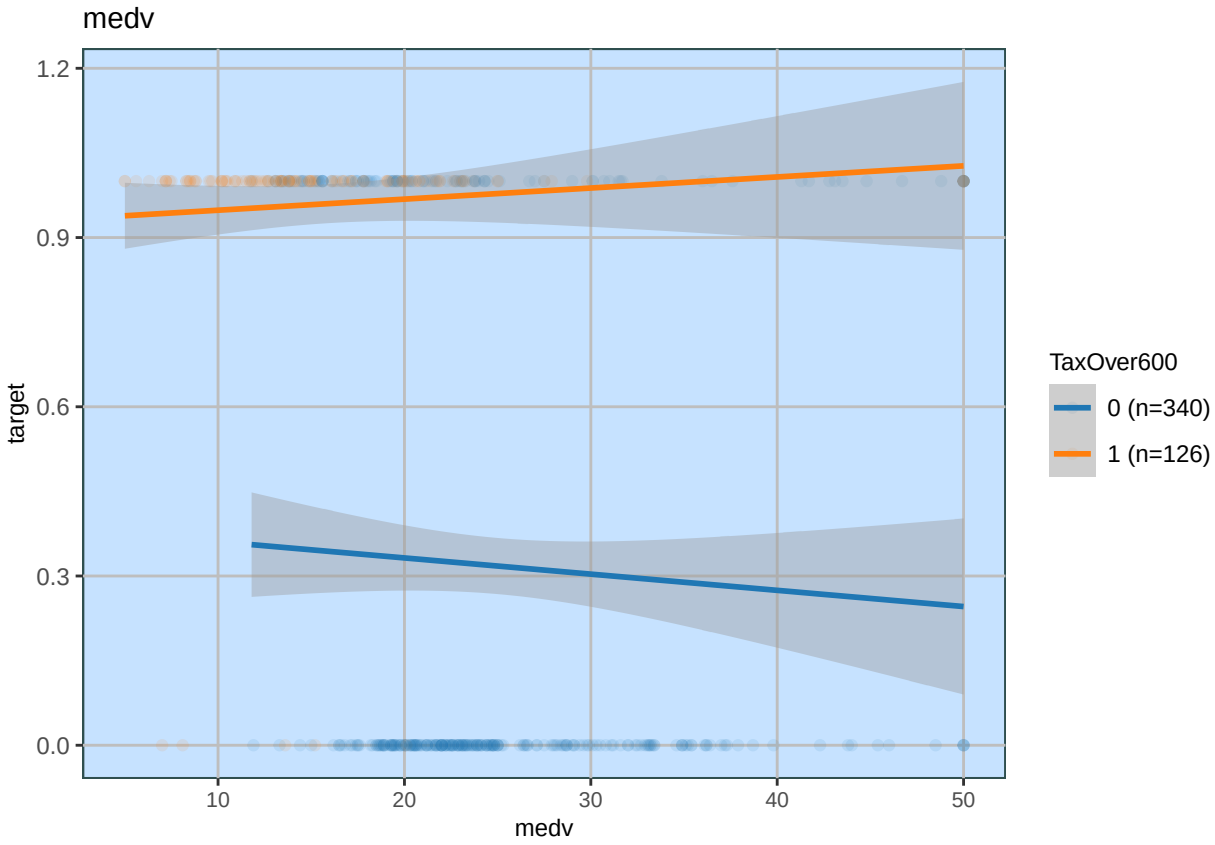
```
##
## [[10]]
```


ptratio
1.00
0.75
0.50
0.25
0.00
target
12.5
15.0
17.5
20.0
ptratio
TaxOver600
0 (n=340)
1 (n=126)

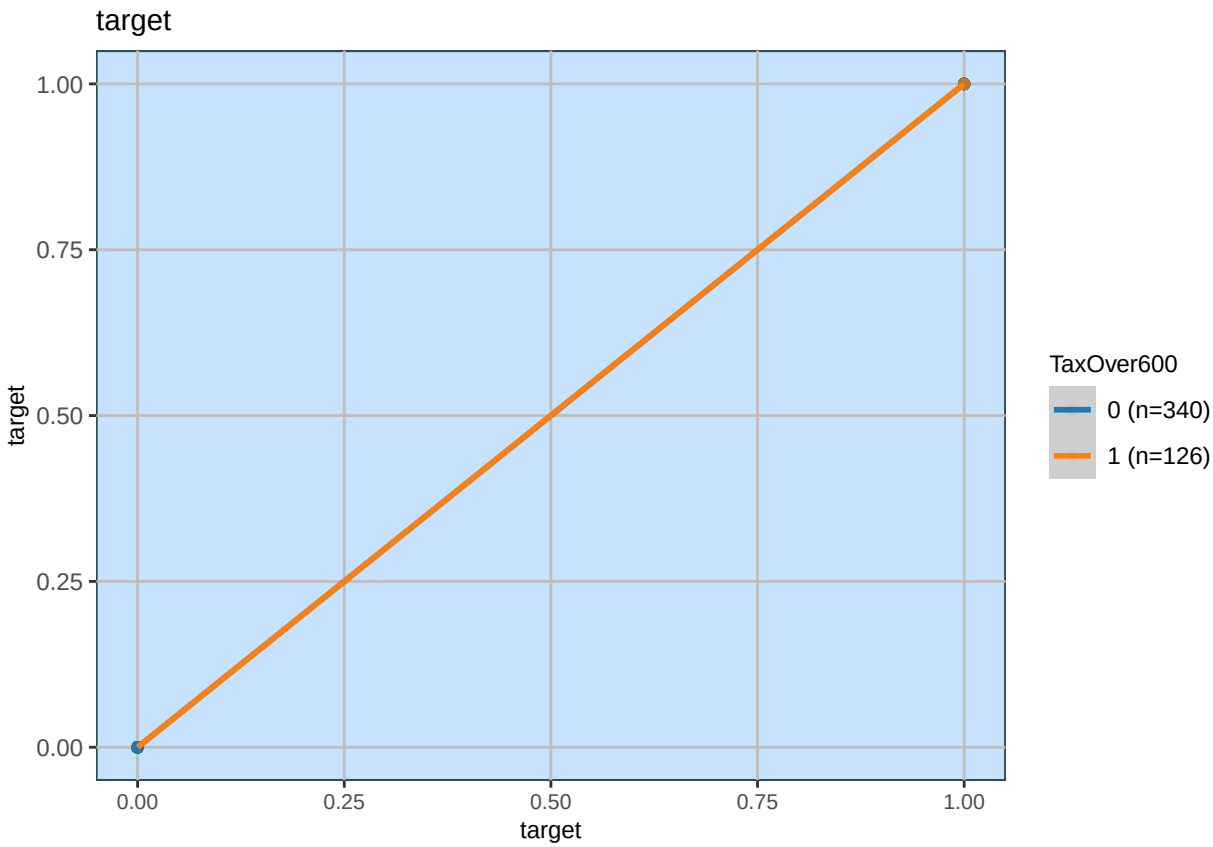
```
##
## [[11]]
```


lstat
target
0.0
0.3
0.6
0.9
0
10
20
30
lstat
TaxOver600
0 (n=340)
1 (n=126)

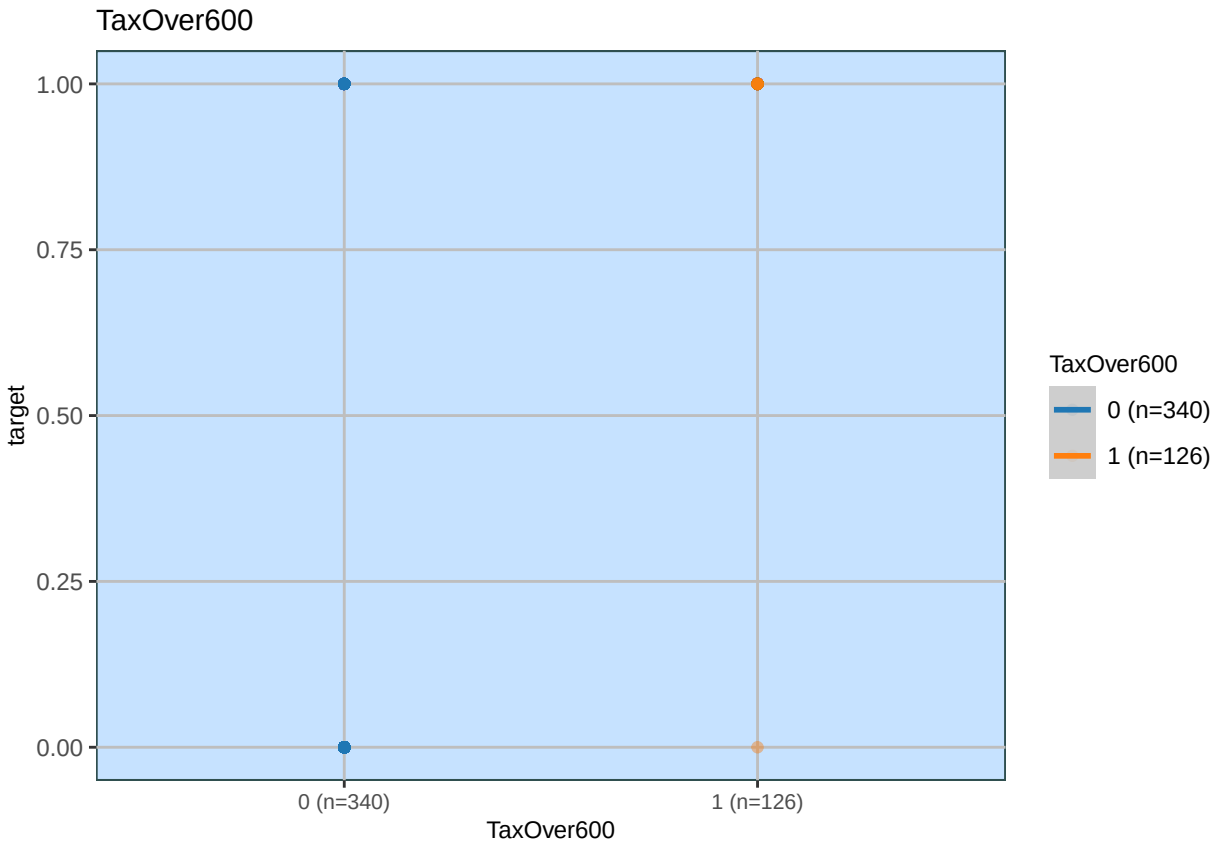
```
##
## [[12]]
```

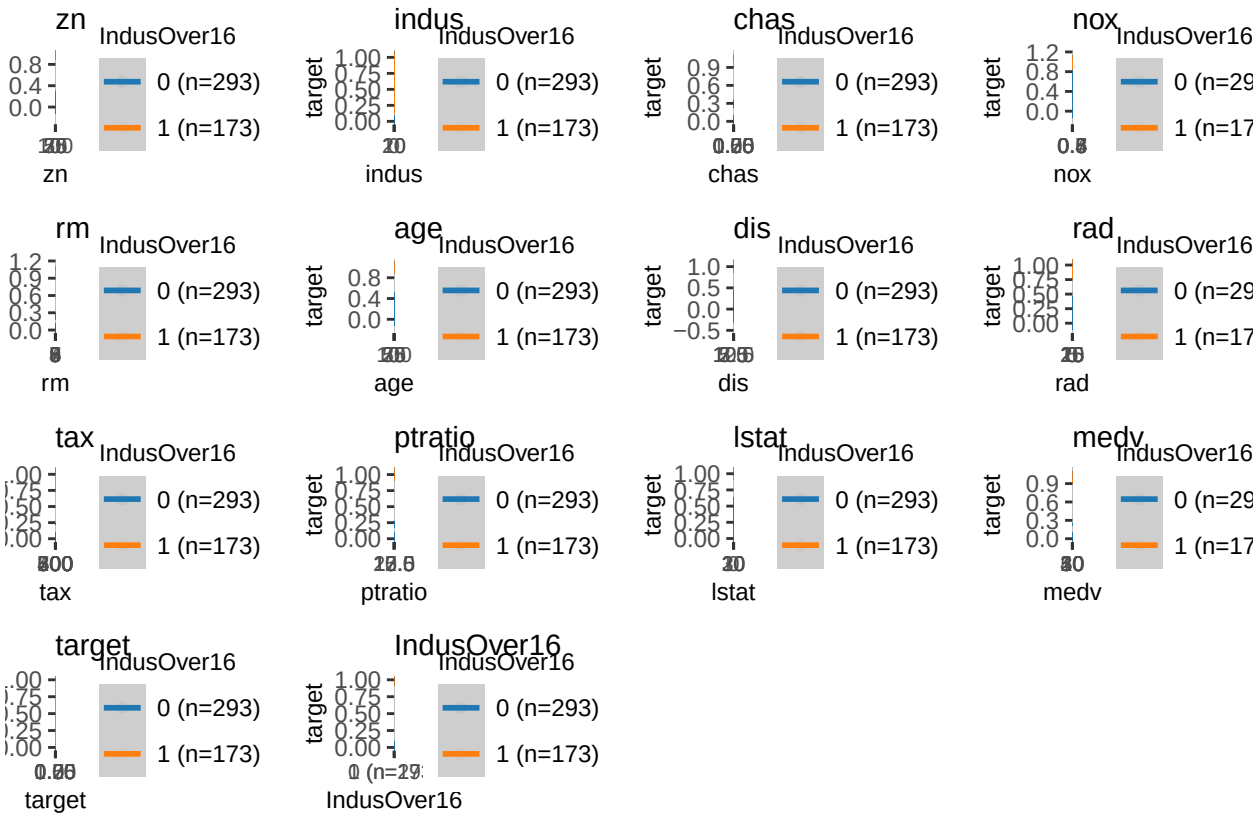

medv
1.2
0.9
0.6
0.3
0.0
target
10
20
30
40
50
medv
TaxOver600
0 (n=340)
1 (n=126)

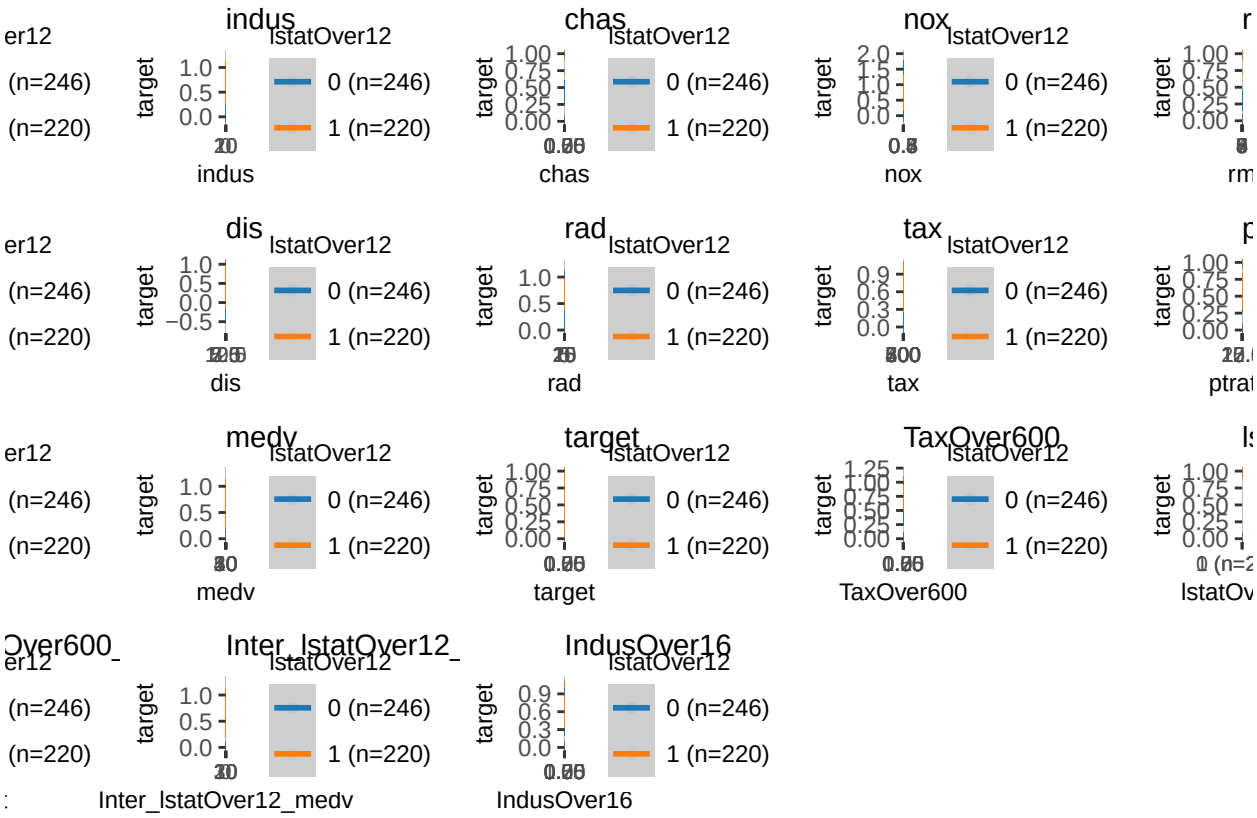
```
##
## [[13]]
```


target
1.00
0.75
0.50
0.25
0.00
target
0.00
0.25
0.50
0.75
1.00
target
TaxOver600
0 (n=340)
1 (n=126)

```
##
## [[14]]
```


TaxOver600
1.00
0.75
0.50
0.25
0.00
target
0 (n=340)
1 (n=126)
TaxOver600
TaxOver600
0 (n=340)
1 (n=126)

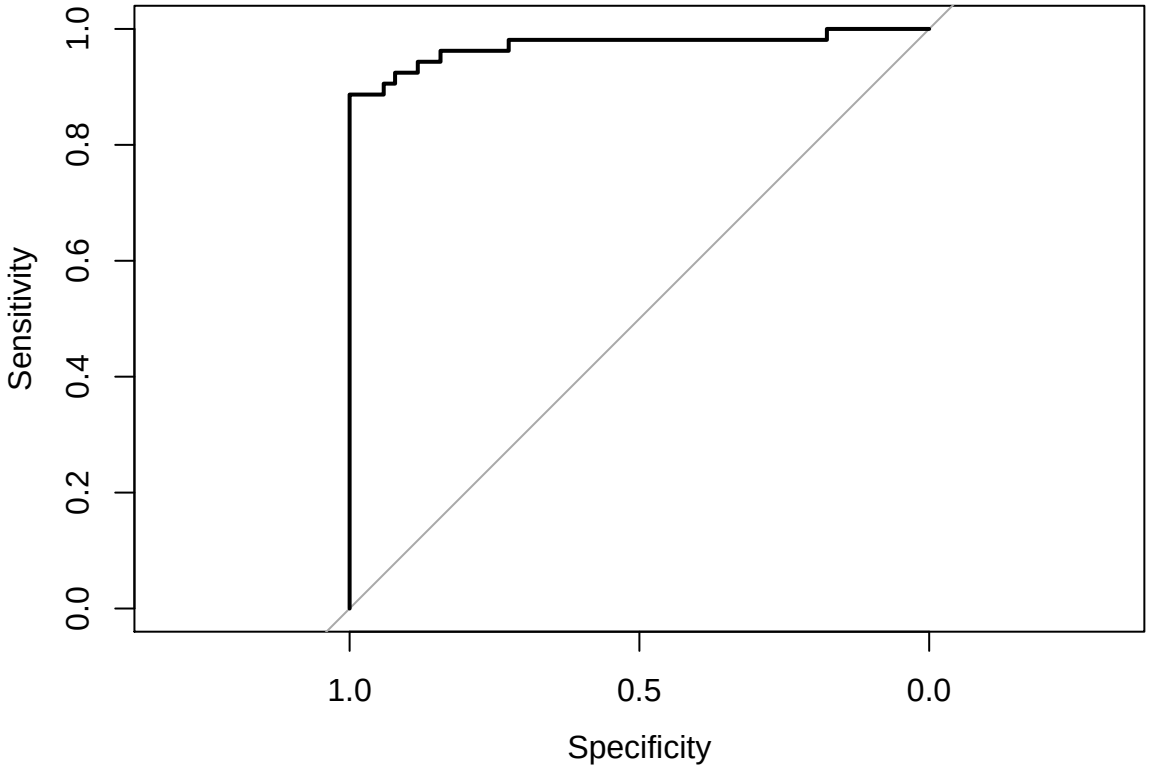

zn
indus
chas
nox
rm
age
dis
rad
tax
ptratio
lstat
medv
target
IndusOver16
0 (n=293)
1 (n=173)



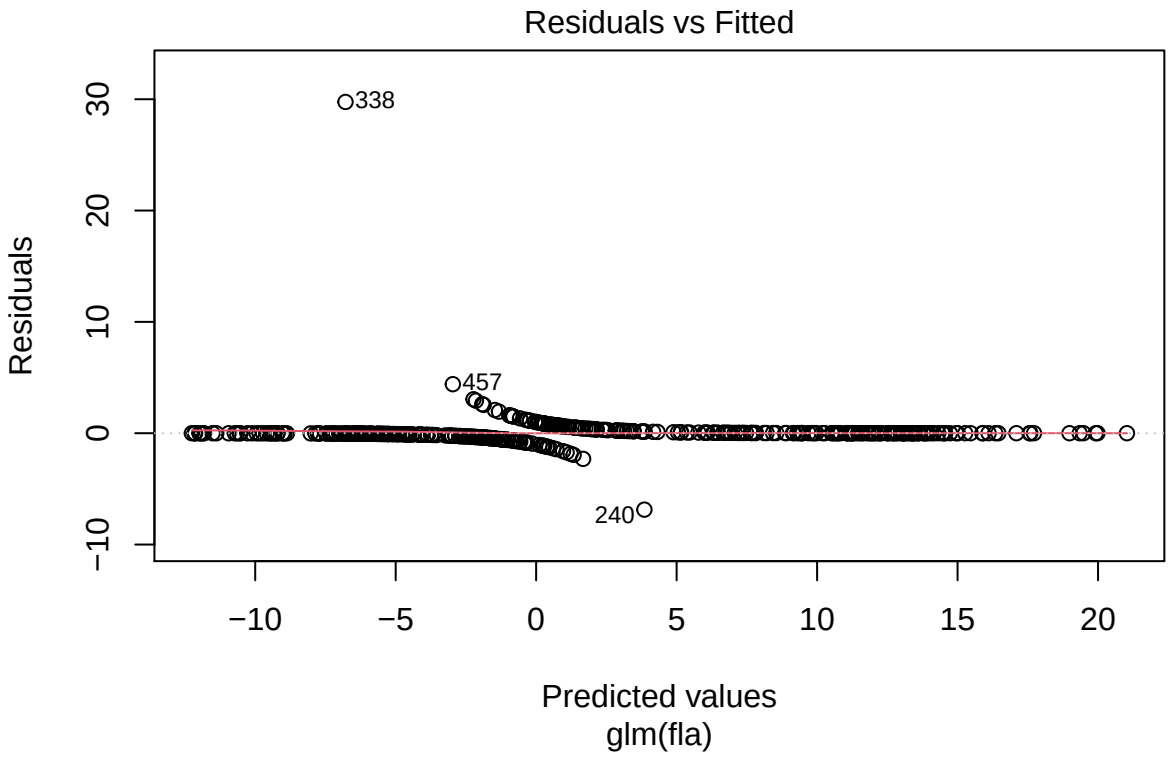

```
##
## Call:  glm(formula = fla, family = "binomial", data = df)
##
## Coefficients:
##            (Intercept)                      zn                   indus
##              -39.437128               -0.057955               -0.112958
##                   chas                     nox                      rm
##                1.373814               48.225024               -0.986275
##                    age                     dis                     rad
##                0.023800                0.678863                0.588571
##                    tax                 ptratio                   lstat
##               -0.001163                0.396786                0.136516
##                   medv              TaxOver600             lstatOver12
##                0.230639                -1.293461                3.463820
## Inter_taxOver600_lstat  Inter_lstatOver12_medv             IndusOver16
##               -0.273658               -0.237180                1.431433
##
## Degrees of Freedom: 465 Total (i.e. Null);  448 Residual
## Null Deviance:       645.9
## Residual Deviance: 176.8     AIC: 212.8
##
## Call:
## glm(formula = fla, family = "binomial", data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

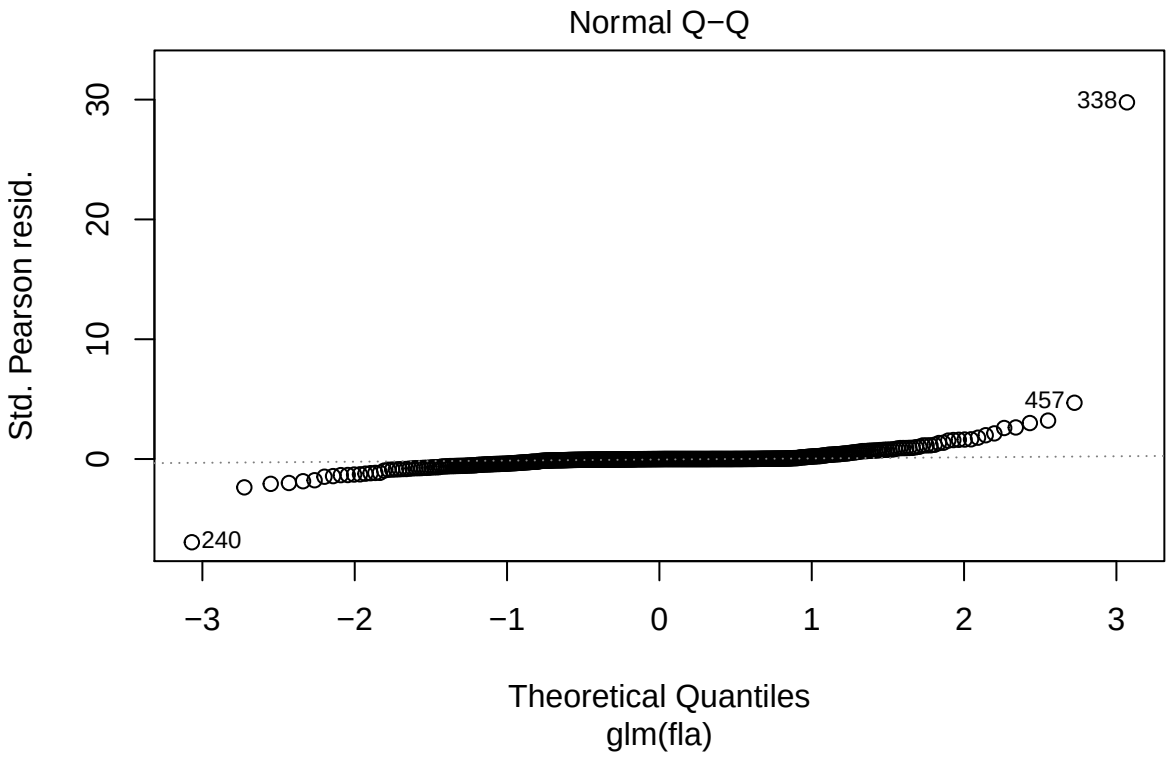
```
## -2.7834  -0.1444  -0.0034   0.0273   3.6843
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -39.437128   7.161435  -5.507 3.65e-08 ***
## zn                      -0.057955   0.033018  -1.755 0.079217 .
## indus                   -0.112958   0.098781  -1.144 0.252821
## chas                     1.373814   0.845214   1.625 0.104076
## nox                     48.225024   8.059928   5.983 2.19e-09 ***
## rm                      -0.986275   0.796630  -1.238 0.215694
## age                      0.023800   0.014601   1.630 0.103098
## dis                      0.678863   0.238467   2.847 0.004416 **
## rad                      0.588571   0.156077   3.771 0.000163 ***
## tax                     -0.001163   0.004025  -0.289 0.772548
## ptratio                  0.396786   0.136976   2.897 0.003770 **
## lstat                    0.136516   0.081492   1.675 0.093891 .
## medv                     0.230639   0.078249   2.947 0.003204 **
## TaxOver600              -1.293461  13.243964  -0.098 0.922199
## lstatOver12              3.463820   2.216992   1.562 0.118195
## Inter_taxOver600_lstat  -0.273658   0.678037  -0.404 0.686504
## Inter_lstatOver12_medv  -0.237180   0.103369  -2.294 0.021762 *
## IndusOver16              1.431433   1.488158   0.962 0.336109
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 645.88  on 465  degrees of freedom
## Residual deviance: 176.79  on 448  degrees of freedom
## AIC: 212.79
##
## Number of Fisher Scoring iterations: 10
##
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 45  4
##          1  6 49
##
##                Accuracy : 0.9038
##                  95% CI : (0.8303, 0.9529)
##     No Information Rate : 0.5096
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.8075
##
##  Mcnemar's Test P-Value : 0.7518
##
##             Sensitivity : 0.8824
##             Specificity : 0.9245
##          Pos Pred Value : 0.9184
##          Neg Pred Value : 0.8909
##              Prevalence : 0.4904
```

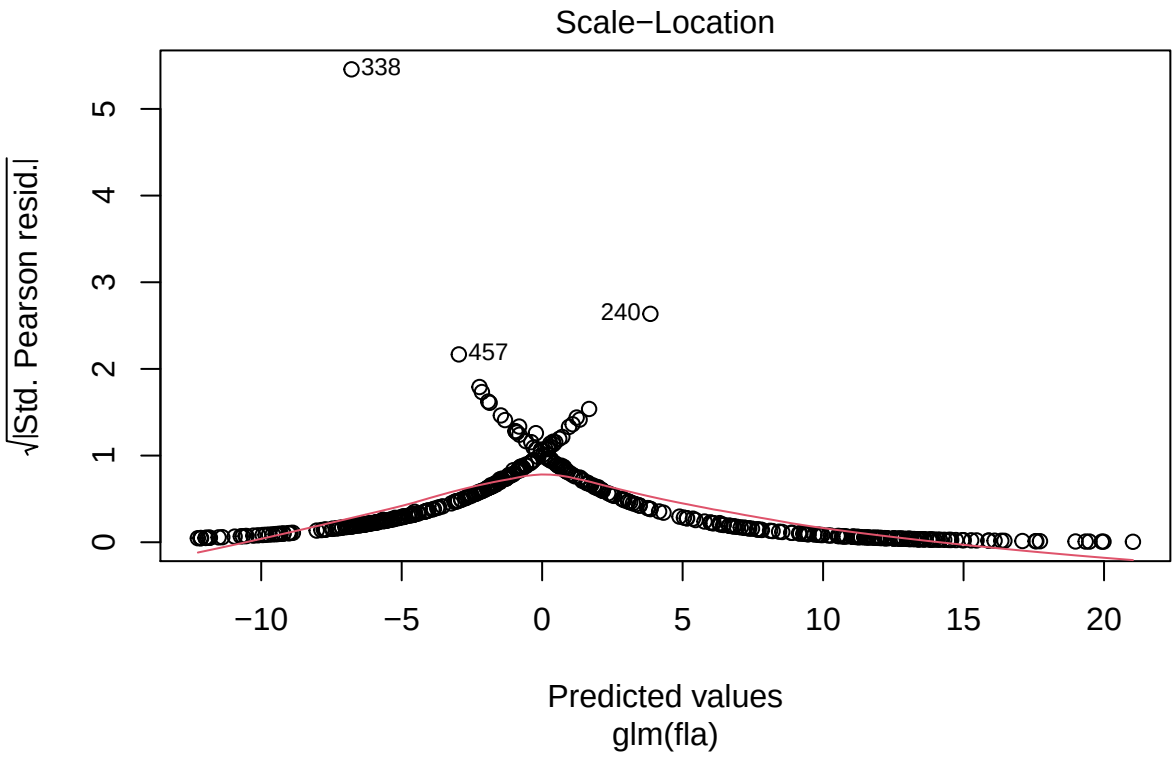
```
##          Detection Rate : 0.4327
##    Detection Prevalence : 0.4712
##       Balanced Accuracy : 0.9034
##
##        'Positive' Class : 0
##
```

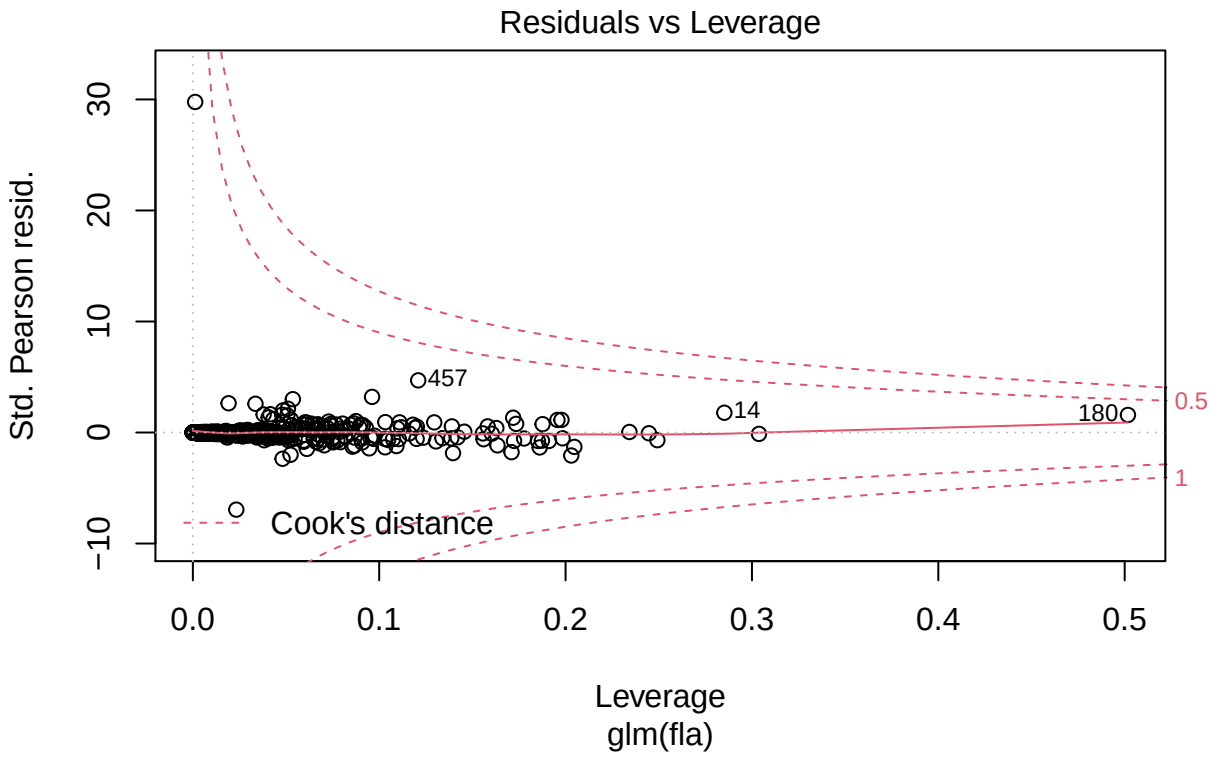



```
## [1] "AUC:  0.971513133555309"
##
## Call:
## roc.default(response = dfPred_raw$class, predictor = dfPred_raw$predict_reg,     plot = TRUE)
##
## Data: dfPred_raw$predict_reg in 51 controls (dfPred_raw$class 0) < 53 cases (dfPred_raw$class 1).
## Area under the curve: 0.9715
```


Residuals vs Fitted
Residuals
Predicted values
glm(fla)
338
457
240
30
20
10
0
−10
−10
−5
0
5
10
15
20


Normal Q−Q
Std. Pearson resid.
0
10
20
30
338
457
240
−3
−2
−1
0
1
2
3
Theoretical Quantiles
glm(fla)

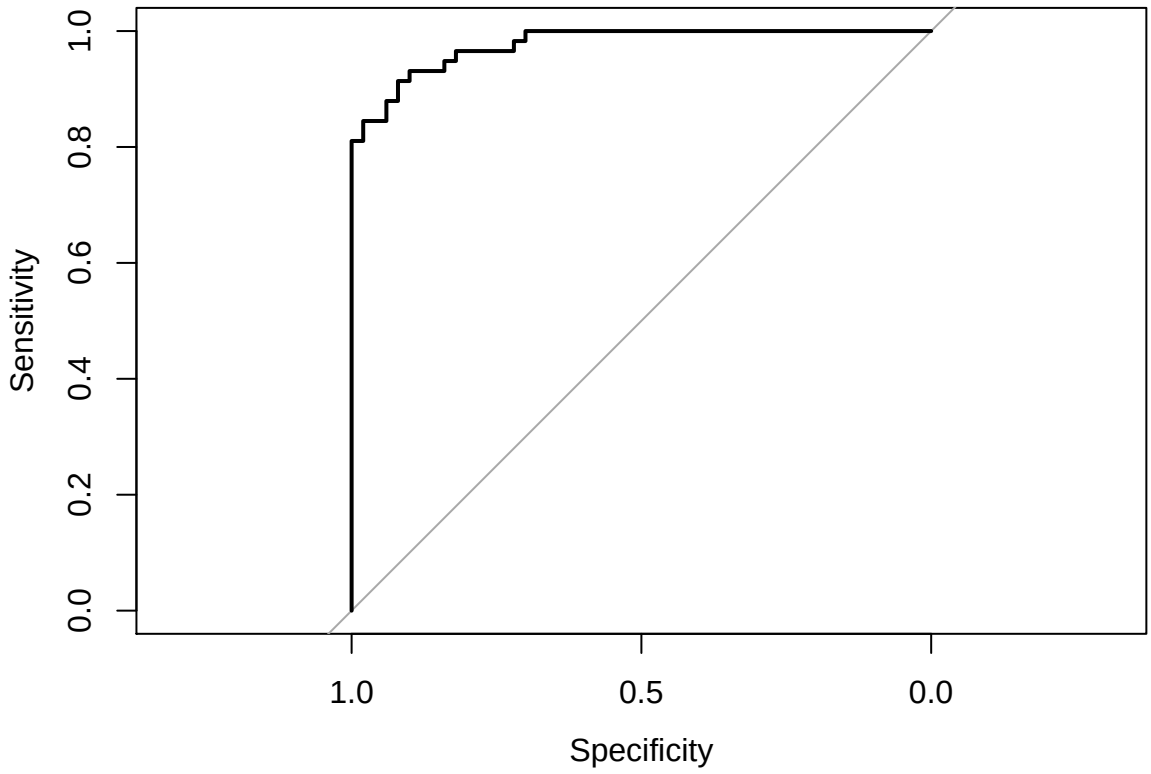

Scale−Location
338
240
457
√|Std. Pearson resid.|
0
1
2
3
4
5
−10
−5
0
5
10
15
20
Predicted values
glm(fla)



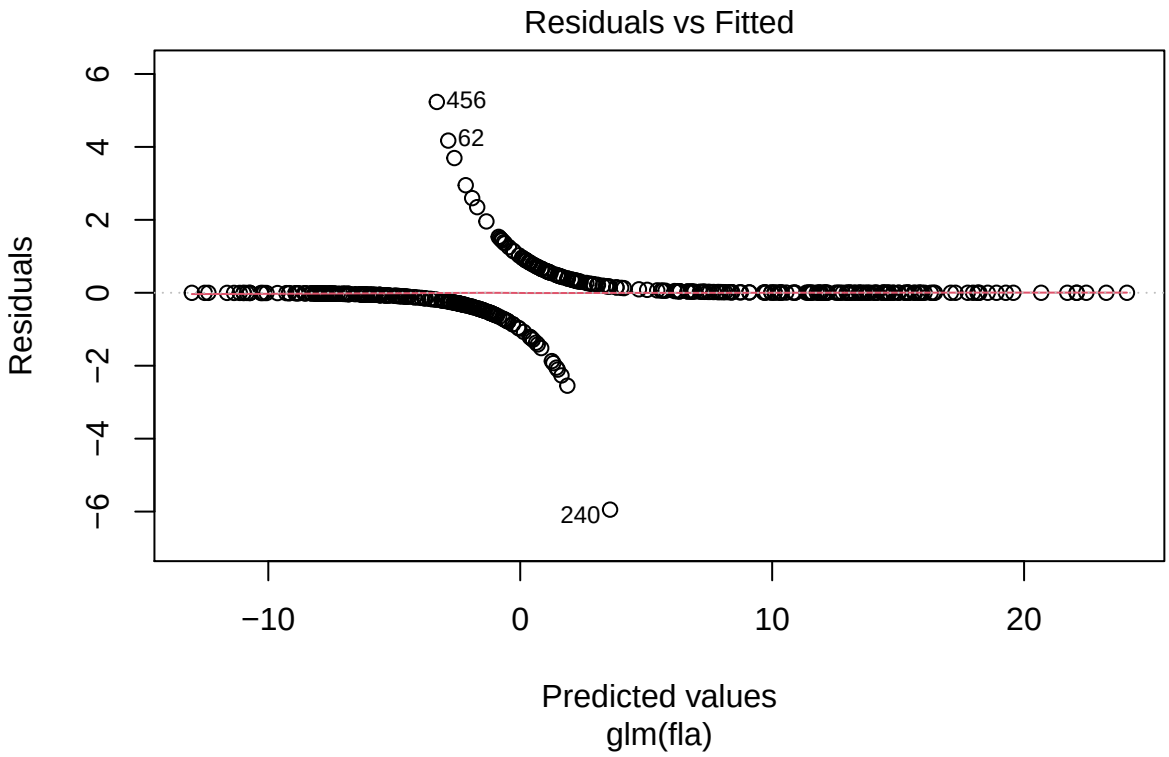

```
##
## Call:  glm(formula = fla, family = "binomial", data = df)
##
## Coefficients:
##            (Intercept)                      zn                   indus
##              -4.669e+01               -6.879e-01               -4.388e-02
##                   chas                     nox                      rm
##               1.170e+00                5.695e+01               -1.279e+00
##                    age                     dis                     rad
##               3.427e-02                9.827e-01                6.641e-01
##                    tax                 ptratio                   lstat
##              -7.533e-04                3.948e-01                1.623e-01
##                   medv              TaxOver600              lstatOver12
##               2.970e-01               -1.398e+00                3.418e+00
## Inter_taxOver600_lstat  Inter_lstatOver12_medv
##              -2.948e-01               -2.526e-01
##
## Degrees of Freedom: 464 Total (i.e. Null);  448 Residual
## Null Deviance:       644.5
## Residual Deviance: 162   AIC: 196
##
## Call:
## glm(formula = fla, family = "binomial", data = df)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
```

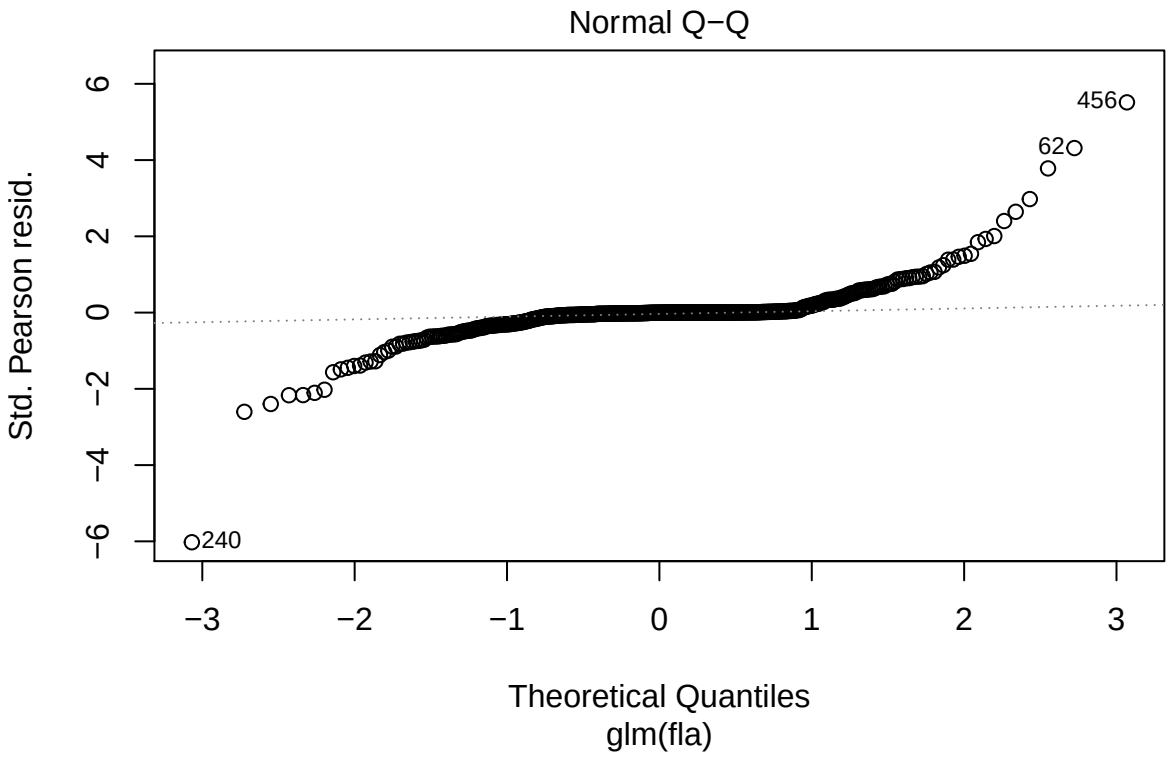
```
## -2.68080  -0.11809  -0.00483   0.01816   2.58674
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -4.669e+01  8.034e+00  -5.811 6.21e-09 ***
## zn                       -6.879e-01  3.002e-01  -2.292 0.021927 *
## indus                    -4.388e-02  5.699e-02  -0.770 0.441294
## chas                      1.170e+00  8.761e-01   1.336 0.181557
## nox                       5.695e+01  9.078e+00   6.273 3.54e-10 ***
## rm                       -1.279e+00  8.428e-01  -1.518 0.129035
## age                       3.427e-02  1.595e-02   2.149 0.031637 *
## dis                       9.827e-01  2.818e-01   3.487 0.000488 ***
## rad                       6.641e-01  1.731e-01   3.837 0.000125 ***
## tax                      -7.533e-04  4.271e-03  -0.176 0.859987
## ptratio                   3.948e-01  1.494e-01   2.643 0.008218 **
## lstat                     1.623e-01  8.504e-02   1.909 0.056248 .
## medv                      2.970e-01  8.394e-02   3.538 0.000403 ***
## TaxOver600               -1.398e+00  1.638e+01  -0.085 0.931975
## lstatOver12               3.418e+00  2.294e+00   1.489 0.136359
## Inter_taxOver600_lstat -2.948e-01  8.650e-01  -0.341 0.733264
## Inter_lstatOver12_medv -2.526e-01  1.071e-01  -2.360 0.018293 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 644.45  on 464  degrees of freedom
## Residual deviance: 161.98  on 448  degrees of freedom
## AIC: 195.98
##
## Number of Fisher Scoring iterations: 10
##
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 46  7
##          1  4 51
##
##                Accuracy : 0.8981
##                  95% CI : (0.8251, 0.948)
##     No Information Rate : 0.537
##     P-Value [Acc > NIR] : 5.356e-16
##
##                   Kappa : 0.796
##
##  Mcnemar's Test P-Value : 0.5465
##
##             Sensitivity : 0.9200
##             Specificity : 0.8793
##          Pos Pred Value : 0.8679
##          Neg Pred Value : 0.9273
##              Prevalence : 0.4630
##          Detection Rate : 0.4259
```

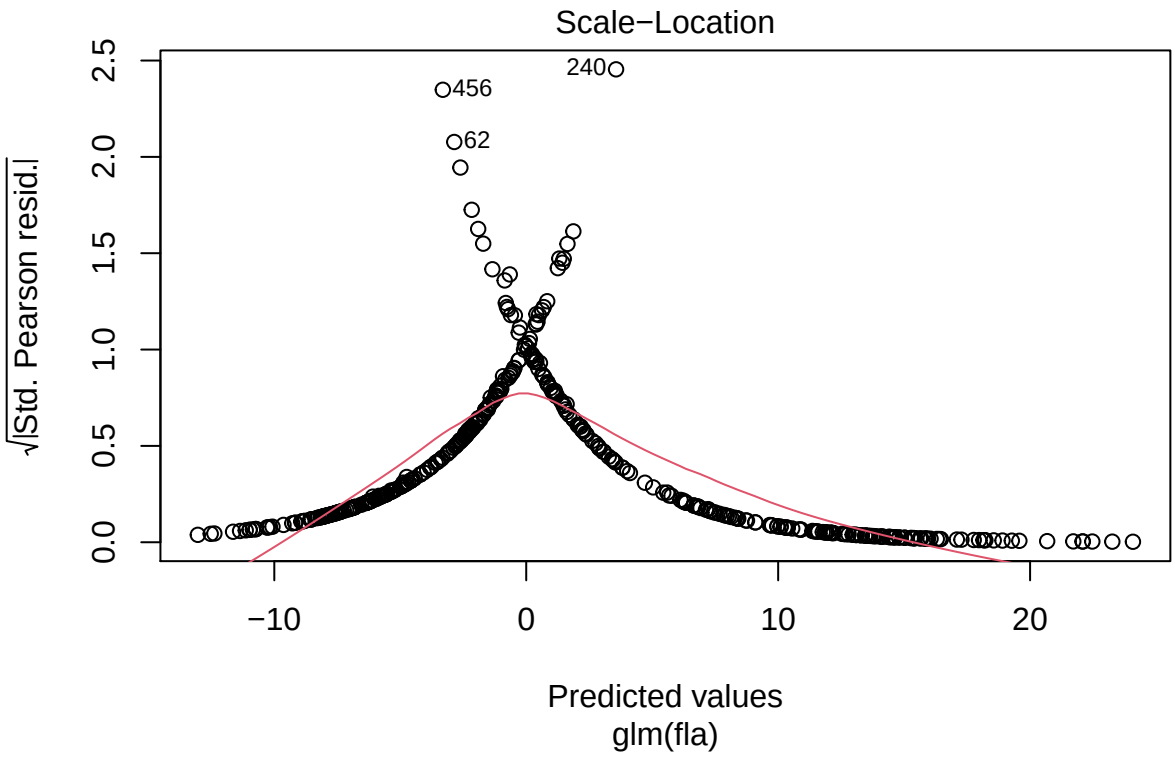
```
##    Detection Prevalence : 0.4907
##       Balanced Accuracy : 0.8997
##
##        'Positive' Class : 0
##
```

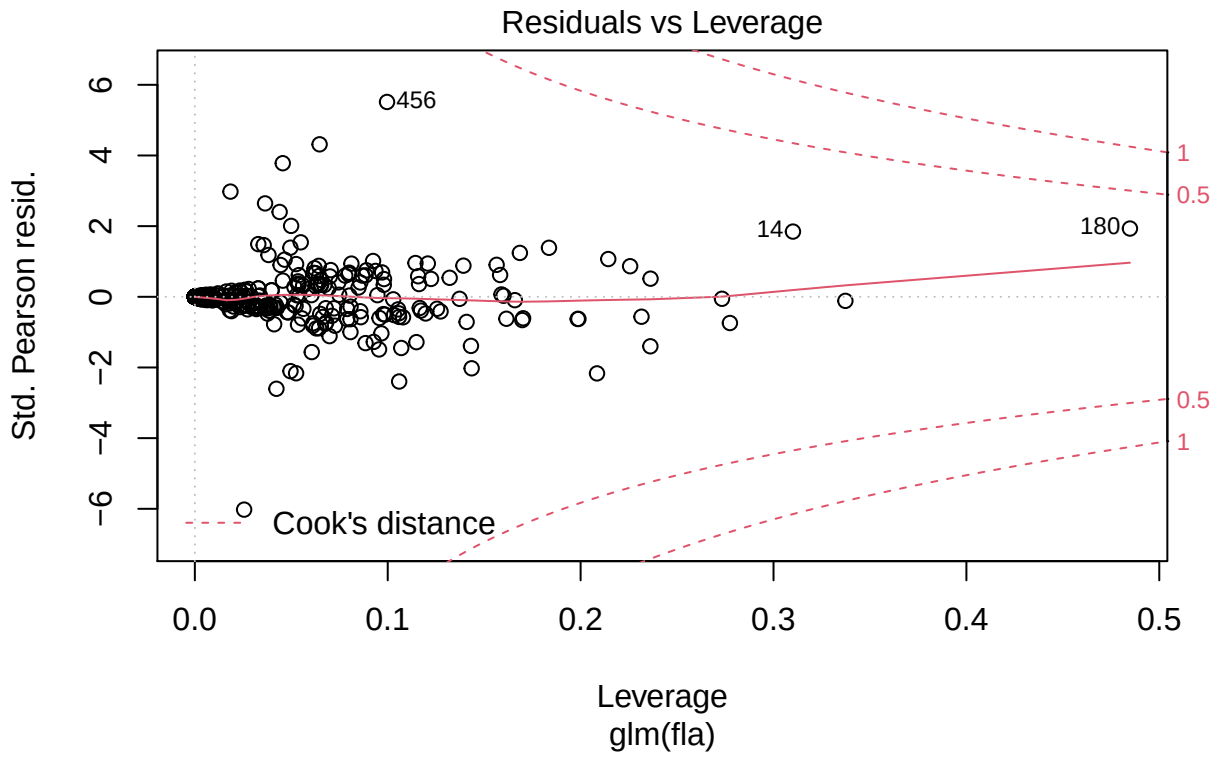



```
## [1] "AUC:  0.976896551724138"
##
## Call:
## roc.default(response = dfPred_raw$class, predictor = dfPred_raw$predict_reg,     plot = TRUE)
##
## Data: dfPred_raw$predict_reg in 50 controls (dfPred_raw$class 0) < 58 cases (dfPred_raw$class 1).
## Area under the curve: 0.9769
```


Residuals vs Fitted
Residuals
Predicted values
glm(fla)
456
62
240
6
4
2
0
−2
−4
−6
−10
0
10
20

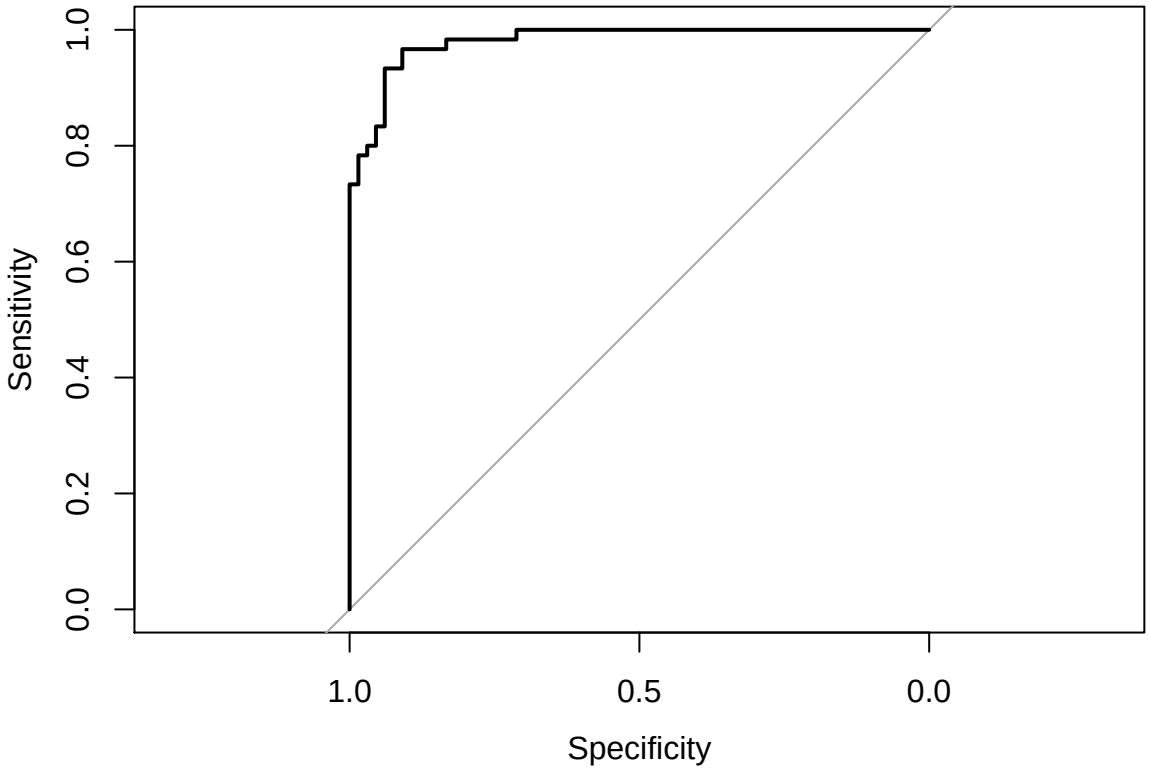

Normal Q−Q
Std. Pearson resid.
6
4
2
0
−2
−4
−6
456
62
240
−3
−2
−1
0
1
2
3
Theoretical Quantiles
glm(fla)


Scale−Location
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
2.5
240
456
62
−10
0
10
20
Predicted values
glm(fla)




```
##
## Call:  glm(formula = fla, family = "binomial", data = df)
##
## Coefficients:
## (Intercept)           zn          nox          age          dis          rad
##   -42.01114     -0.07453     44.50486      0.03490      0.77365      0.53383
##     ptratio         medv        indus  IndusOver16        inter
##     0.40204      0.13838     -0.05919    152.54358     -6.95711
##
## Degrees of Freedom: 464 Total (i.e. Null);  454 Residual
## Null Deviance:       644.5
## Residual Deviance: 174.5     AIC: 196.5
##
## Call:
## glm(formula = fla, family = "binomial", data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0340  -0.1419   0.0000   0.0000   3.1217
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -42.01114    7.17696  -5.854 4.81e-09 ***
## zn            -0.07453    0.03557  -2.095 0.036129 *
## nox           44.50486    7.85438   5.666 1.46e-08 ***
## age            0.03490    0.01153   3.027 0.002467 **
```

```
## dis            0.77365    0.22698   3.408 0.000653 ***
## rad            0.53383    0.13471   3.963 7.41e-05 ***
## ptratio        0.40204    0.13231   3.039 0.002377 **
## medv           0.13838    0.03959   3.495 0.000474 ***
## indus         -0.05919    0.09302  -0.636 0.524582
## IndusOver16  152.54358 9215.00483   0.017 0.986793
## inter         -6.95711  420.96877  -0.017 0.986814
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 644.45  on 464  degrees of freedom
## Residual deviance: 174.55  on 454  degrees of freedom
## AIC: 196.55
##
## Number of Fisher Scoring iterations: 20
##
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 62  7
##          1  4 53
##
##                Accuracy : 0.9127
##                  95% CI : (0.8492, 0.9556)
##     No Information Rate : 0.5238
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.8246
##
##  Mcnemar's Test P-Value : 0.5465
##
##             Sensitivity : 0.9394
##             Specificity : 0.8833
##          Pos Pred Value : 0.8986
##          Neg Pred Value : 0.9298
##              Prevalence : 0.5238
##          Detection Rate : 0.4921
##    Detection Prevalence : 0.5476
##       Balanced Accuracy : 0.9114
##
##        'Positive' Class : 0
##
```




```
## [1] "AUC:  0.980555555555556"
##
## Call:
## roc.default(response = dfPred_raw$class, predictor = dfPred_raw$predict_reg,     plot = TRUE)
##
## Data: dfPred_raw$predict_reg in 66 controls (dfPred_raw$class 0) < 60 cases (dfPred_raw$class 1).
## Area under the curve: 0.9806
```

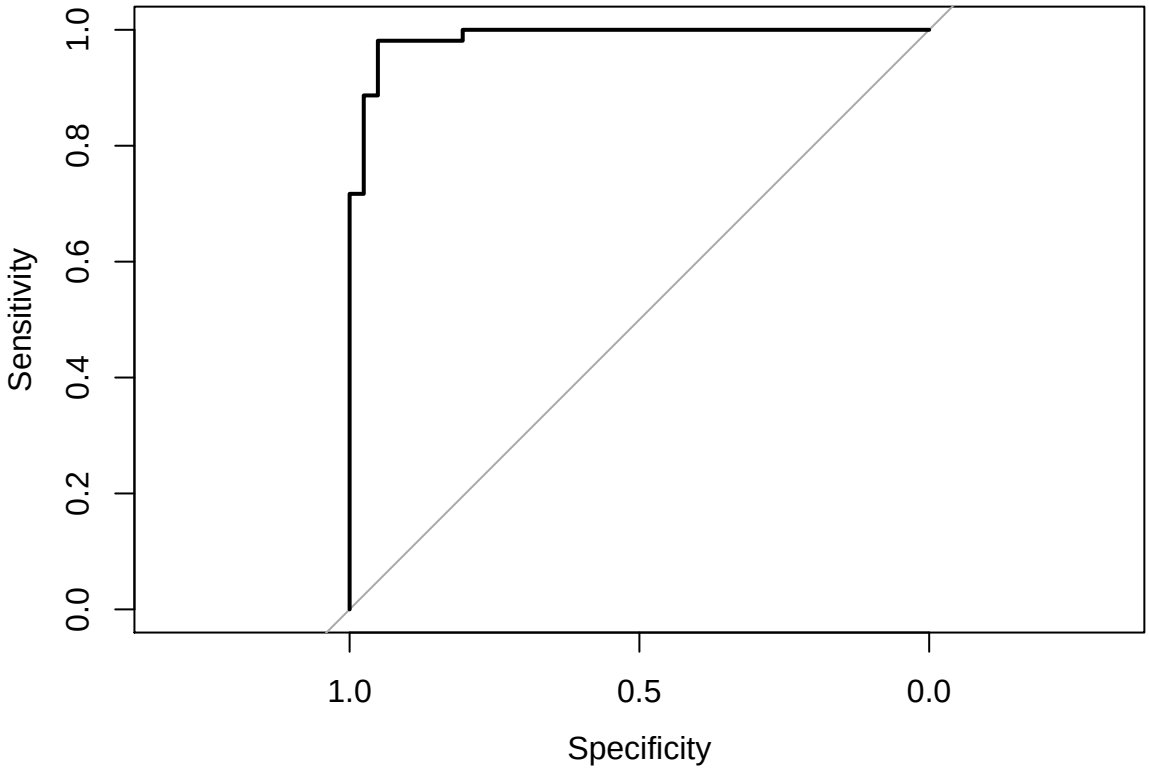
```
##
## Call:  glm(formula = fla, family = "binomial", data = df)
##
## Coefficients:
## (Intercept)           zn          nox          age          dis          rad
##   -42.01114     -0.07453     44.50486      0.03490      0.77365      0.53383
##     ptratio         medv        indus  IndusOver16        inter
##     0.40204      0.13838     -0.05919    152.54358     -6.95711
##
## Degrees of Freedom: 464 Total (i.e. Null);  454 Residual
## Null Deviance:       644.5
## Residual Deviance: 174.5     AIC: 196.5
```

Building Interactions

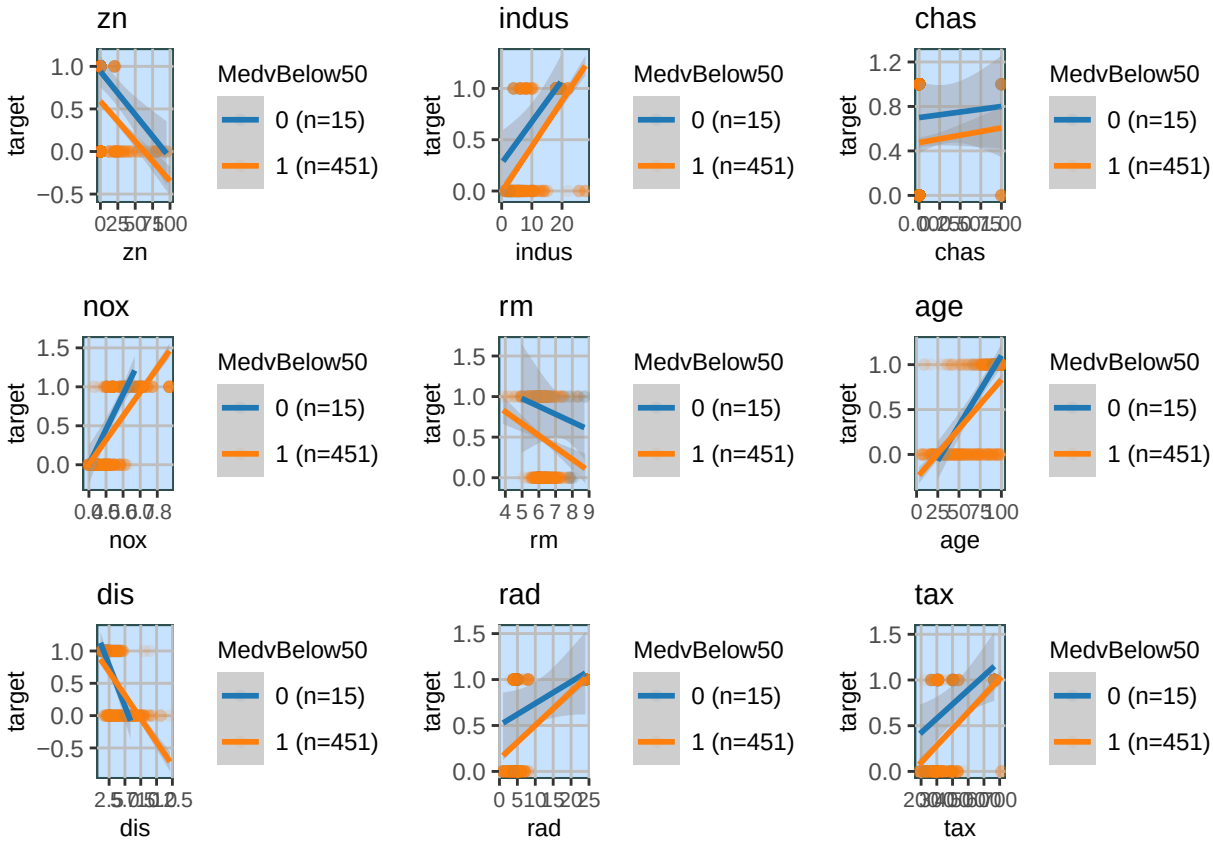
```
##
## Call:  glm(formula = fla, family = "binomial", data = df)
```

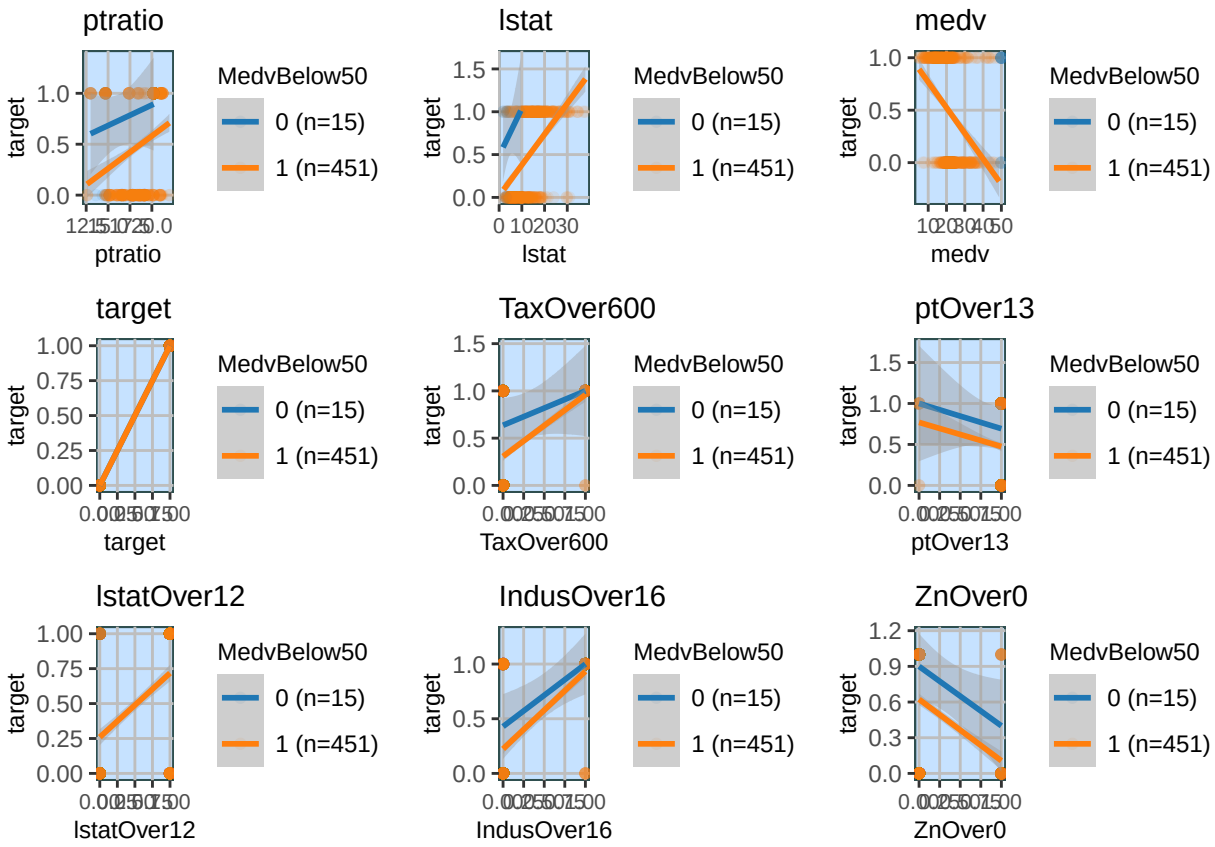
```
##
## Coefficients:
##   (Intercept)             zn          indus           chas            nox
##    -33.528688      -0.049437      -0.056815       0.964549      36.963196
##            rm            age            dis            rad            tax
##     -0.819963       0.038134       1.025748       0.648733       0.001652
##       ptratio          lstat           medv     TaxOver600       ptOver13
##      0.518286       0.171361       0.221716      -7.280268      -8.156292
##   lstatOver12    IndusOver16        ZnOver0  NoxOverPoint8    MedvBelow50
##     -1.372613       1.804373      -2.409817       7.538132       0.897657
##
## Degrees of Freedom: 465 Total (i.e. Null);  446 Residual
## Null Deviance:       645.9
## Residual Deviance: 172.8     AIC: 212.8
##
## Call:
## glm(formula = fla, family = "binomial", data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4794  -0.1493  -0.0023   0.0198   4.1463
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -33.528688   9.574531  -3.502 0.000462 ***
## zn             -0.049437   0.068585  -0.721 0.471027
## indus          -0.056815   0.103823  -0.547 0.584224
## chas            0.964549   0.846075   1.140 0.254275
## nox            36.963196   9.333723   3.960 7.49e-05 ***
## rm             -0.819963   0.831087  -0.987 0.323831
## age             0.038134   0.014766   2.583 0.009805 **
## dis             1.025748   0.292165   3.511 0.000447 ***
## rad             0.648733   0.164036   3.955 7.66e-05 ***
## tax             0.001652   0.004013   0.412 0.680551
## ptratio         0.518286   0.152180   3.406 0.000660 ***
## lstat           0.171361   0.077858   2.201 0.027739 *
## medv            0.221716   0.085963   2.579 0.009903 **
## TaxOver600     -7.280268   4.386502  -1.660 0.096975 .
## ptOver13       -8.156292   4.564175  -1.787 0.073934 .
## lstatOver12    -1.372613   0.676083  -2.030 0.042332 *
## IndusOver16     1.804373   1.467205   1.230 0.218771
## ZnOver0        -2.409817   1.772536  -1.360 0.173978
## NoxOverPoint8   7.538132 953.240826   0.008 0.993690
## MedvBelow50     0.897657   2.107985   0.426 0.670227
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 645.88  on 465  degrees of freedom
## Residual deviance: 172.77  on 446  degrees of freedom
## AIC: 212.77
##
## Number of Fisher Scoring iterations: 16
```

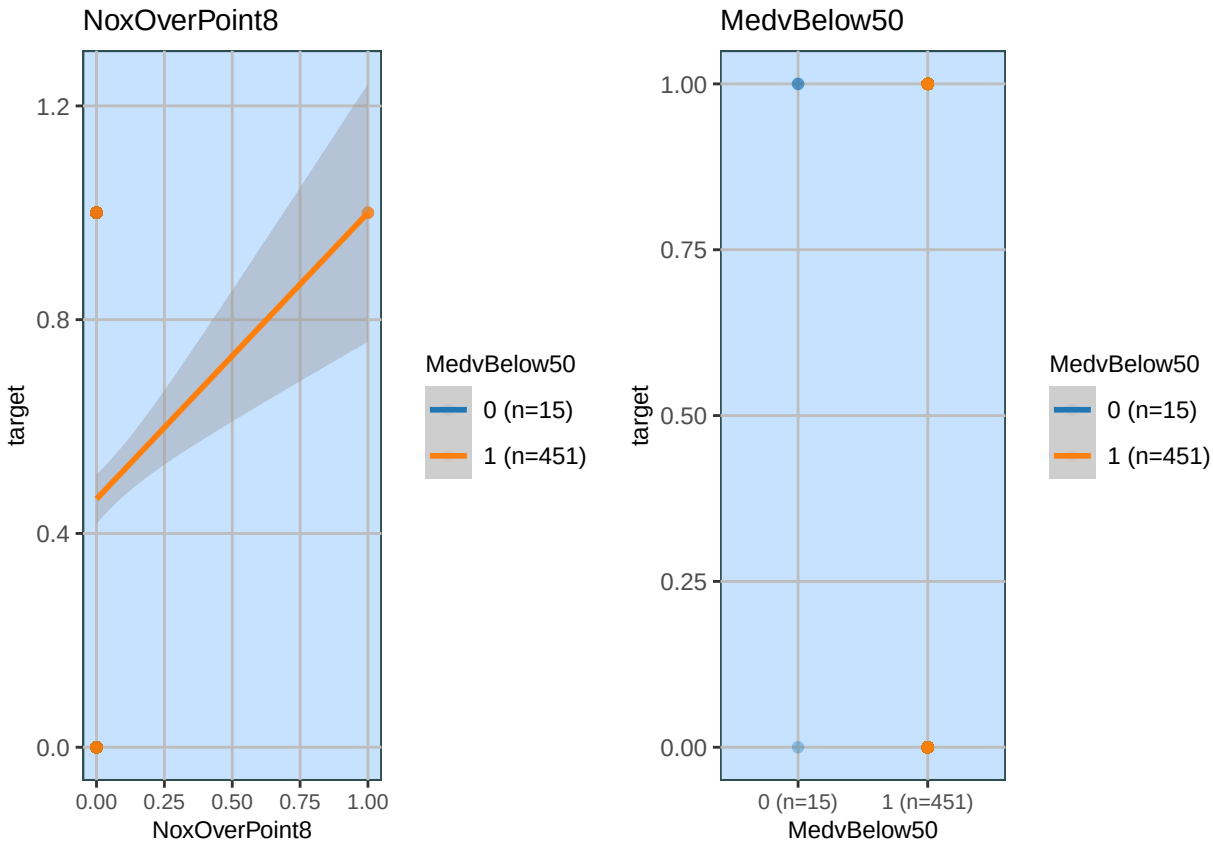
```
##
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 39  3
##          1  2 50
##
##                Accuracy : 0.9468
##                  95% CI : (0.8802, 0.9825)
##     No Information Rate : 0.5638
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.8922
##
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9512
##             Specificity : 0.9434
##          Pos Pred Value : 0.9286
##          Neg Pred Value : 0.9615
##              Prevalence : 0.4362
##          Detection Rate : 0.4149
##    Detection Prevalence : 0.4468
##       Balanced Accuracy : 0.9473
##
##        'Positive' Class : 0
##
```

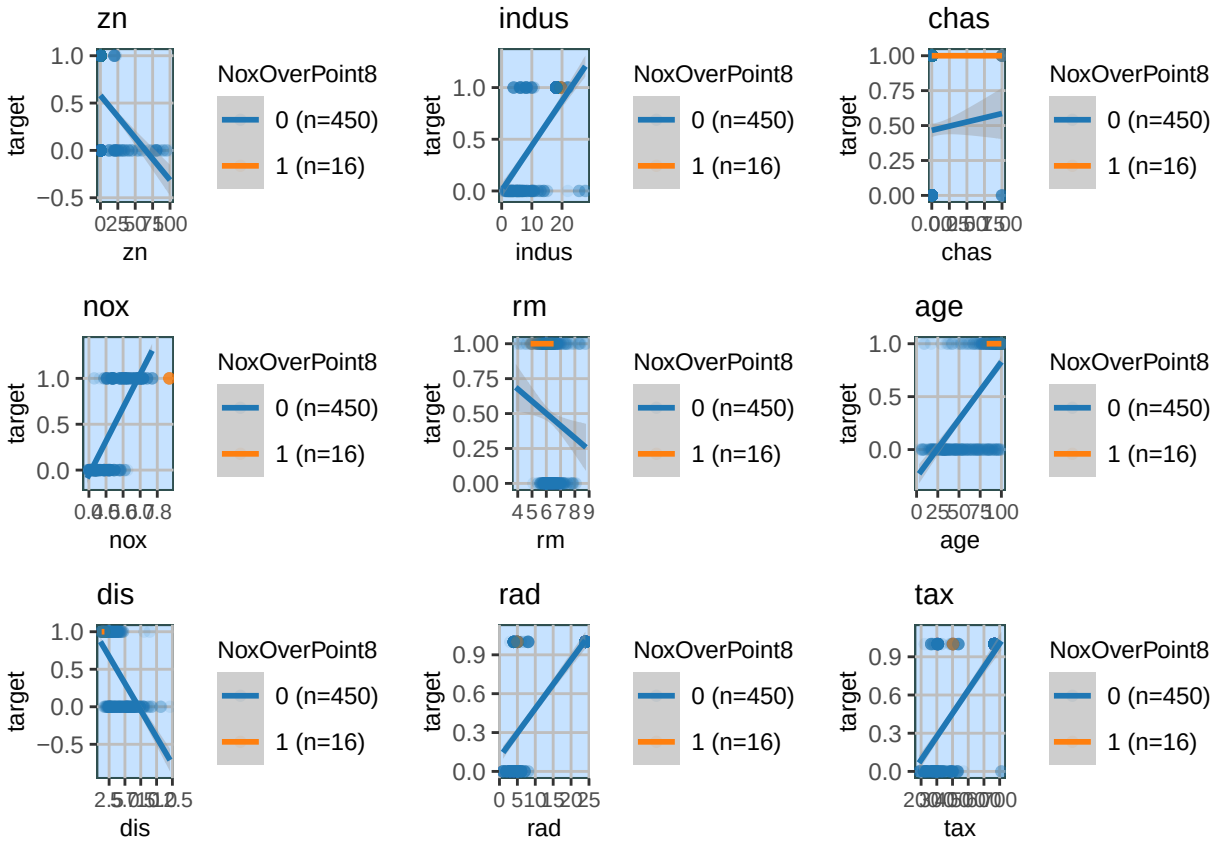


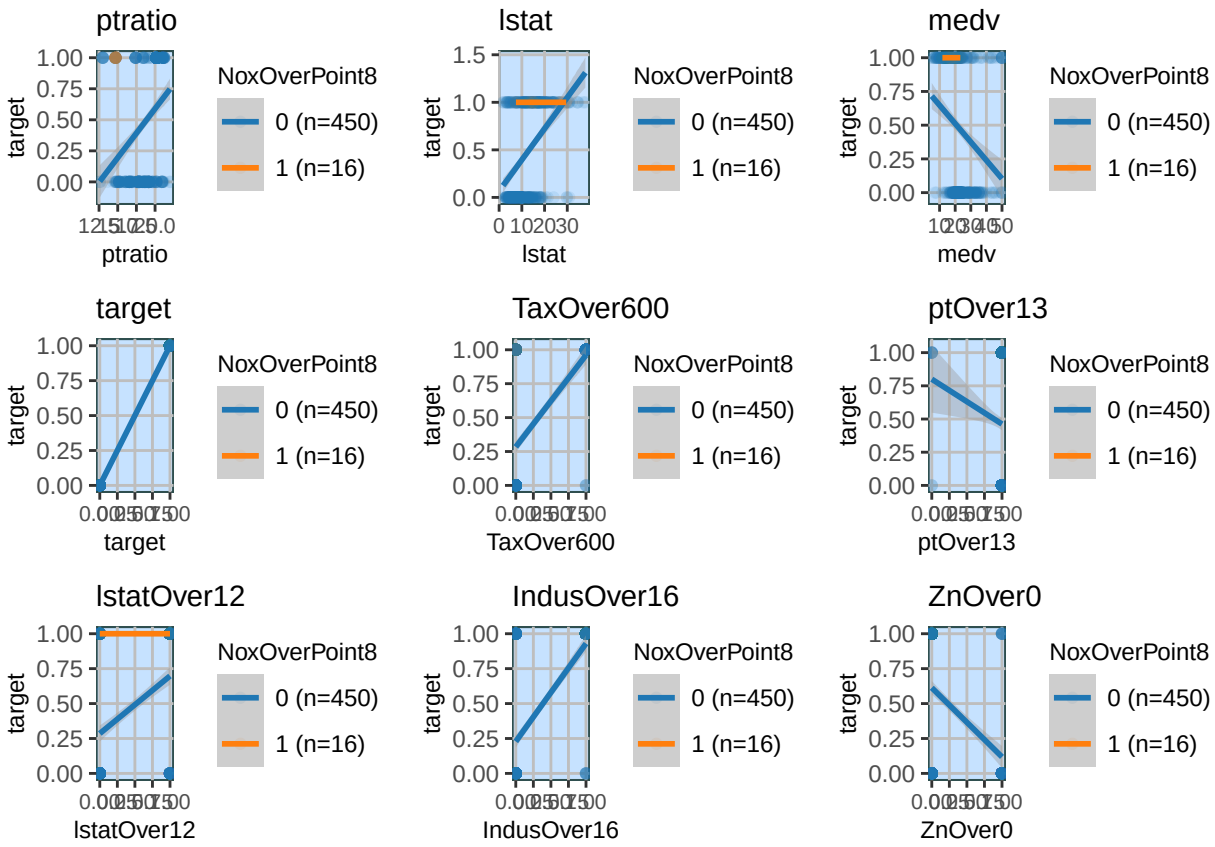

```
## [1] "AUC:  0.987574781408191"
##
## Call:
## roc.default(response = dfPred_raw$class, predictor = dfPred_raw$predict_reg,     plot = TRUE)
##
## Data: dfPred_raw$predict_reg in 41 controls (dfPred_raw$class 0) < 53 cases (dfPred_raw$class 1).
## Area under the curve: 0.9876

##
## Call:  glm(formula = fla, family = "binomial", data = df)
##
## Coefficients:
##   (Intercept)             zn          indus           chas            nox
##    -33.528688      -0.049437      -0.056815       0.964549      36.963196
##            rm            age            dis            rad            tax
##     -0.819963       0.038134       1.025748       0.648733       0.001652
##       ptratio          lstat           medv     TaxOver600       ptOver13
##      0.518286       0.171361       0.221716      -7.280268      -8.156292
##   lstatOver12    IndusOver16        ZnOver0  NoxOverPoint8    MedvBelow50
##     -1.372613       1.804373      -2.409817       7.538132       0.897657
##
## Degrees of Freedom: 465 Total (i.e. Null);  446 Residual
## Null Deviance:       645.9
## Residual Deviance: 172.8     AIC: 212.8
```

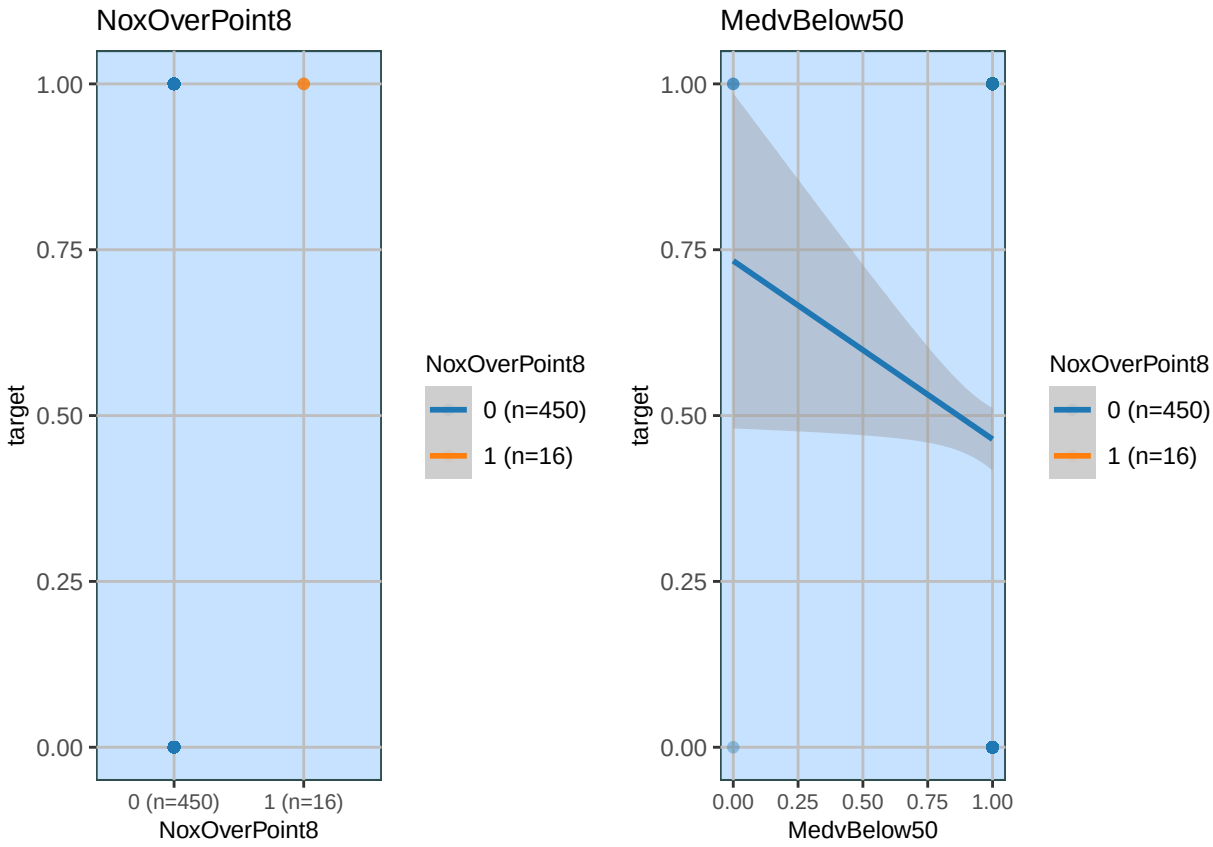

zn
indus
chas
nox
rm
age
dis
rad
tax
target
MedvBelow50
0 (n=15)
1 (n=451)

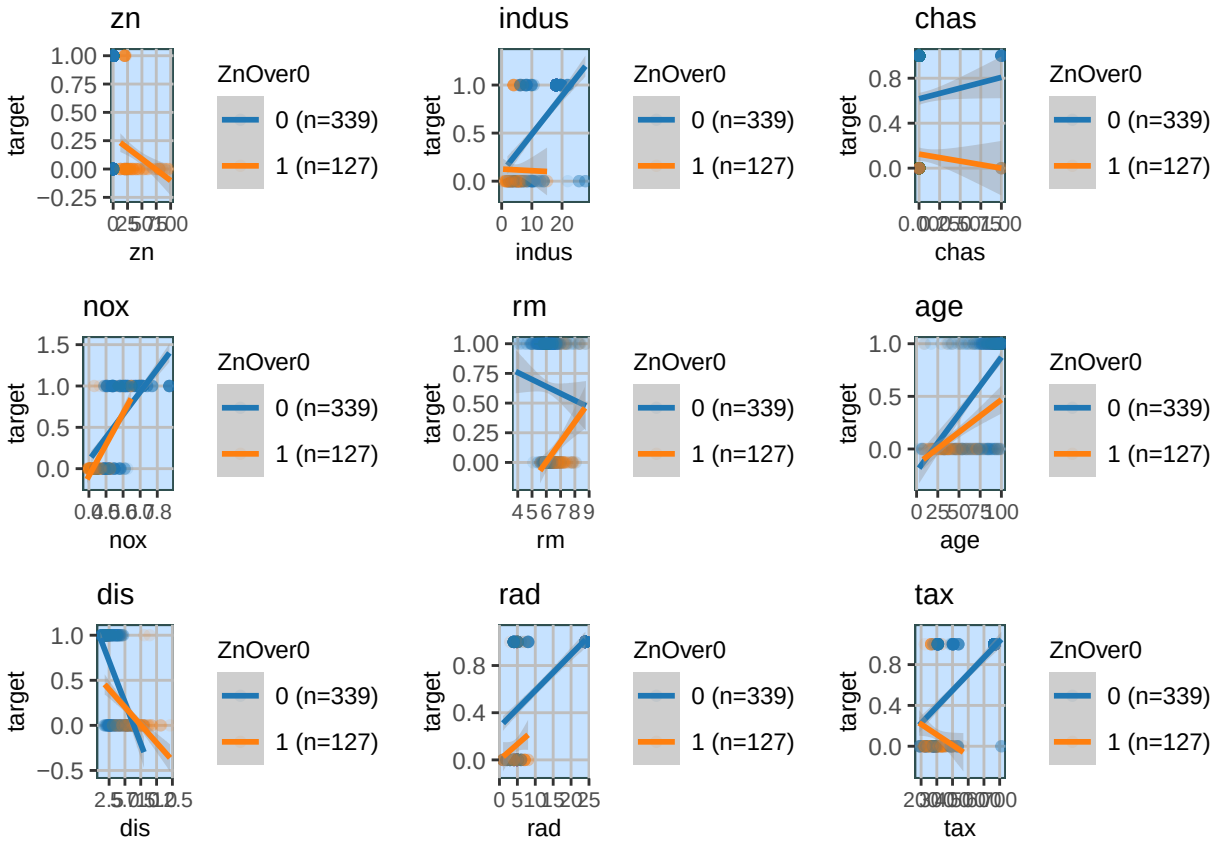

ptratio
target
ptratio
lstat
target
lstat
medv
target
medv
target
target
TaxOver600
target
TaxOver600
ptOver13
target
ptOver13
lstatOver12
target
lstatOver12
IndusOver16
target
IndusOver16
ZnOver0
target
ZnOver0
MedvBelow50
0 (n=15)
1 (n=451)

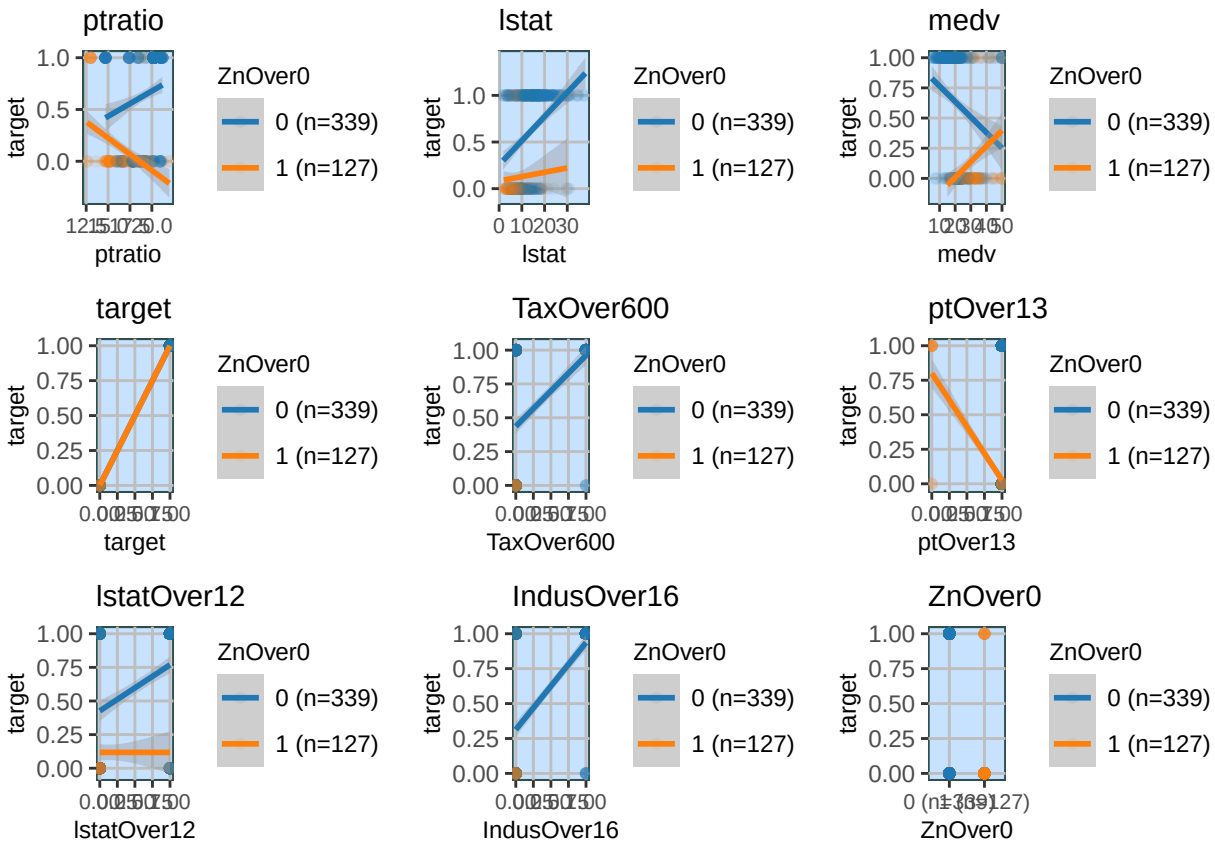

NoxOverPoint8
target
1.2
0.8
0.4
0.0
0.00
0.25
0.50
0.75
1.00
NoxOverPoint8
MedvBelow50
0 (n=15)
1 (n=451)
MedvBelow50
1.00
0.75
0.50
0.25
0.00
target
0 (n=15)
1 (n=451)
MedvBelow50
MedvBelow50
0 (n=15)
1 (n=451)

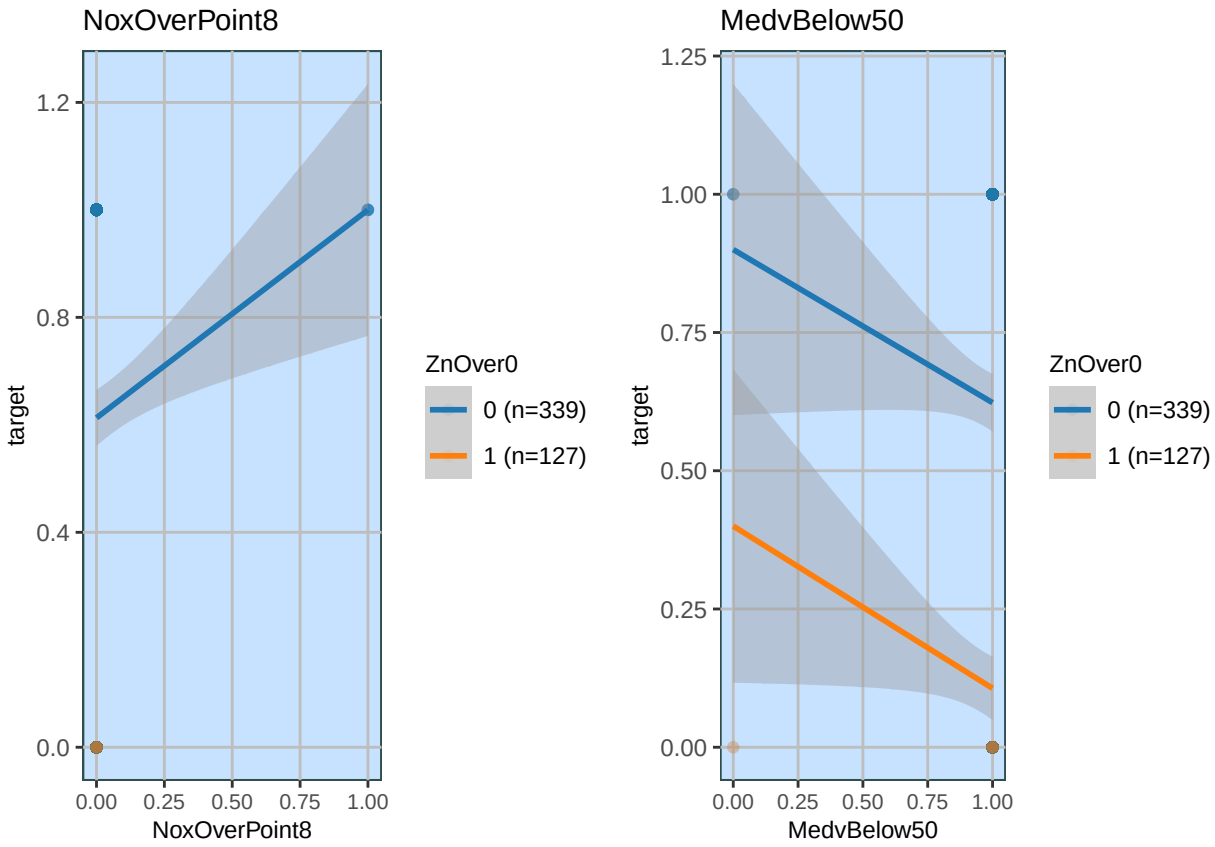

zn
target
zn
indus
target
indus
chas
target
chas
nox
target
nox
rm
target
rm
age
target
age
dis
target
dis
rad
target
rad
tax
target
tax
NoxOverPoint8
0 (n=450)
1 (n=16)

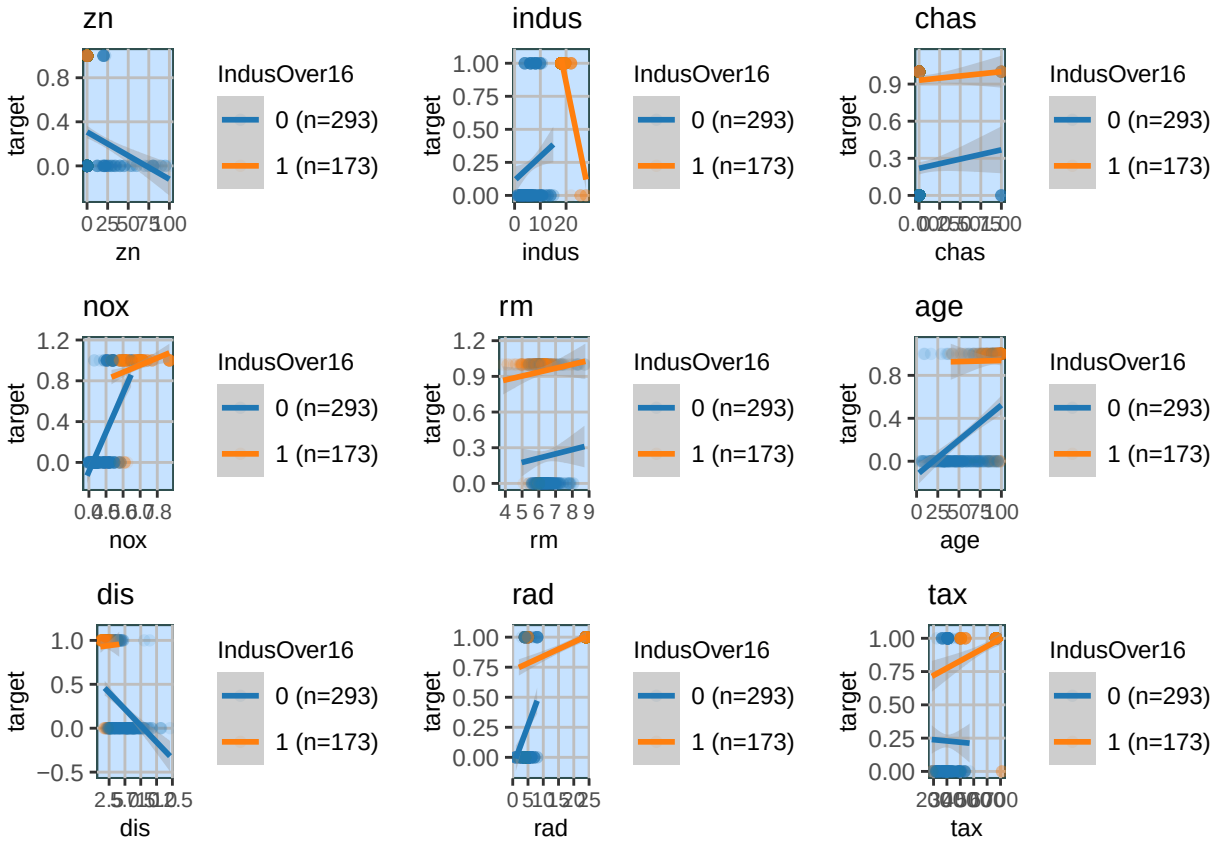

ptratio
target
ptratio
lstat
target
lstat
medv
target
medv
target
target
TaxOver600
target
TaxOver600
ptOver13
target
ptOver13
lstatOver12
target
lstatOver12
IndusOver16
target
IndusOver16
ZnOver0
target
ZnOver0
NoxOverPoint8
0 (n=450)
1 (n=16)

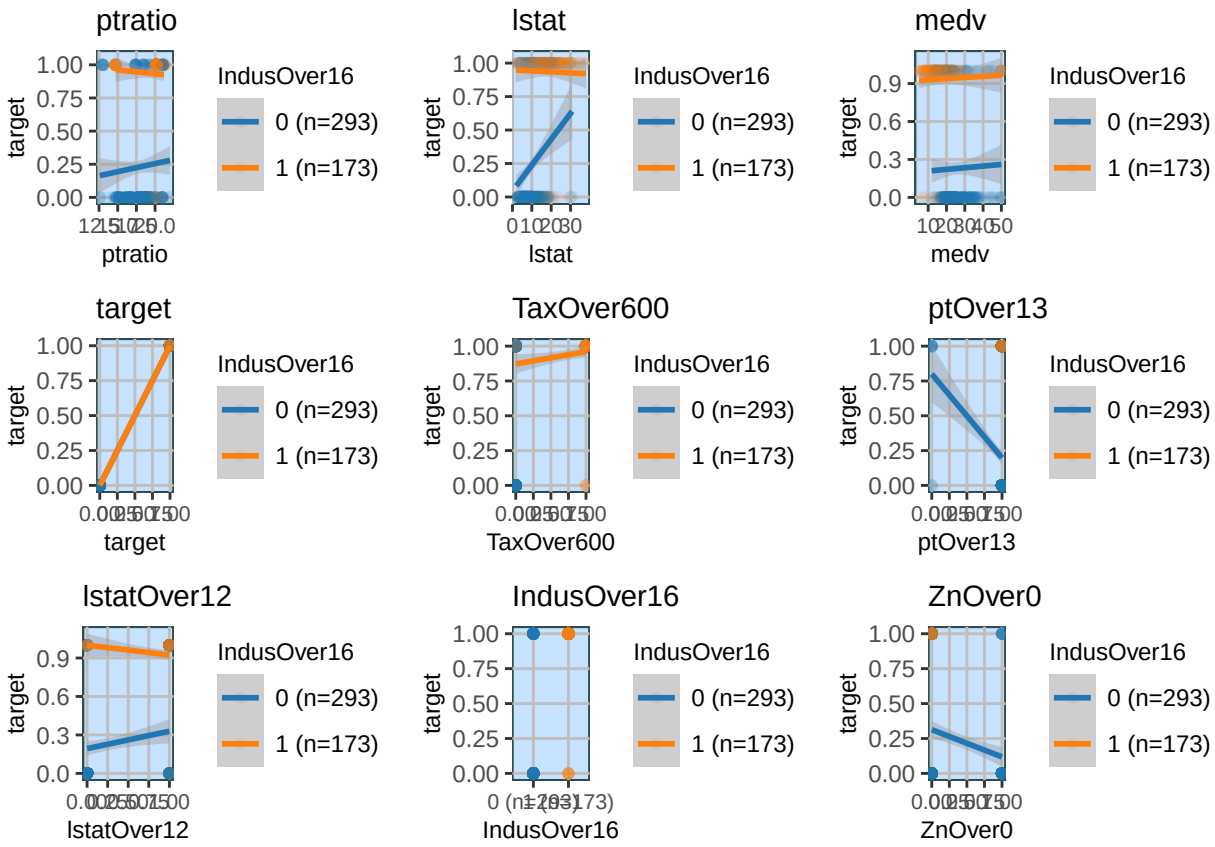

NoxOverPoint8
1.00
0.75
0.50
0.25
0.00
target
0 (n=450)
1 (n=16)
NoxOverPoint8
NoxOverPoint8
0 (n=450)
1 (n=16)
MedvBelow50
1.00
0.75
0.50
0.25
0.00
target
0.00
0.25
0.50
0.75
1.00
MedvBelow50
NoxOverPoint8
0 (n=450)
1 (n=16)

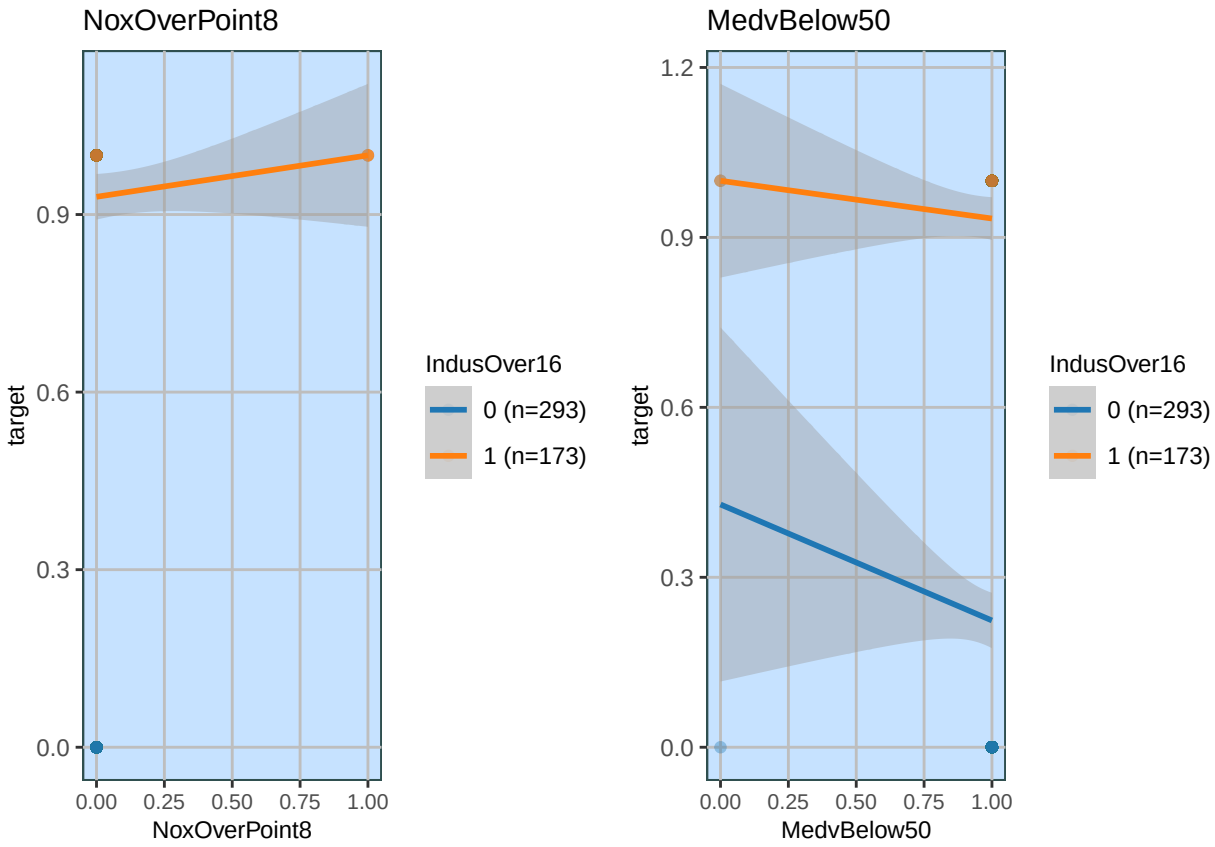

zn
target
zn
indus
chas
nox
rm
age
dis
rad
tax
ZnOver0
0 (n=339)
1 (n=127)

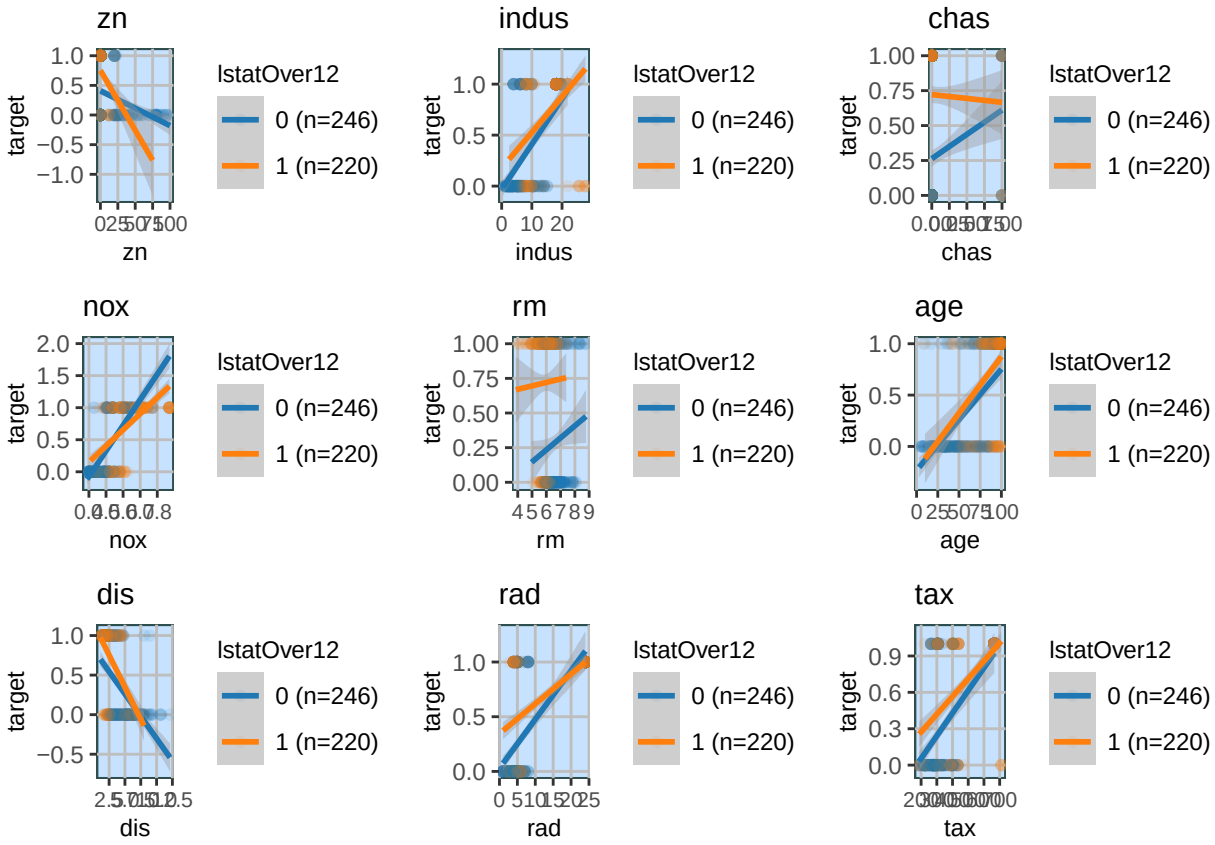

ptratio
target
ptratio
lstat
target
lstat
medv
target
medv
target
target
TaxOver600
target
TaxOver600
ptOver13
target
ptOver13
lstatOver12
target
lstatOver12
IndusOver16
target
IndusOver16
ZnOver0
target
ZnOver0
ZnOver0
0 (n=339)
1 (n=127)

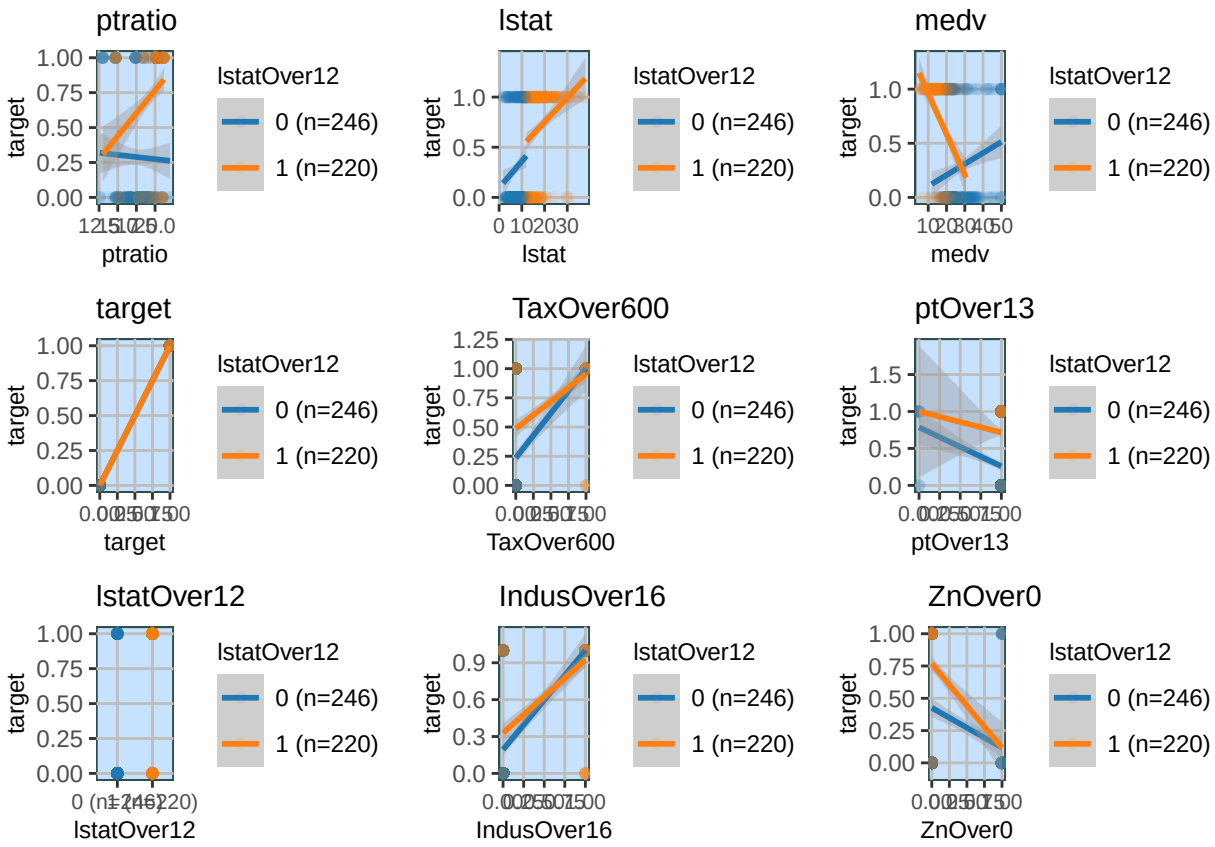

NoxOverPoint8
target
NoxOverPoint8
0.0
0.4
0.8
1.2
0.00
0.25
0.50
0.75
1.00
ZnOver0
0 (n=339)
1 (n=127)
MedvBelow50
target
MedvBelow50
0.00
0.25
0.50
0.75
1.00
1.25
ZnOver0
0 (n=339)
1 (n=127)

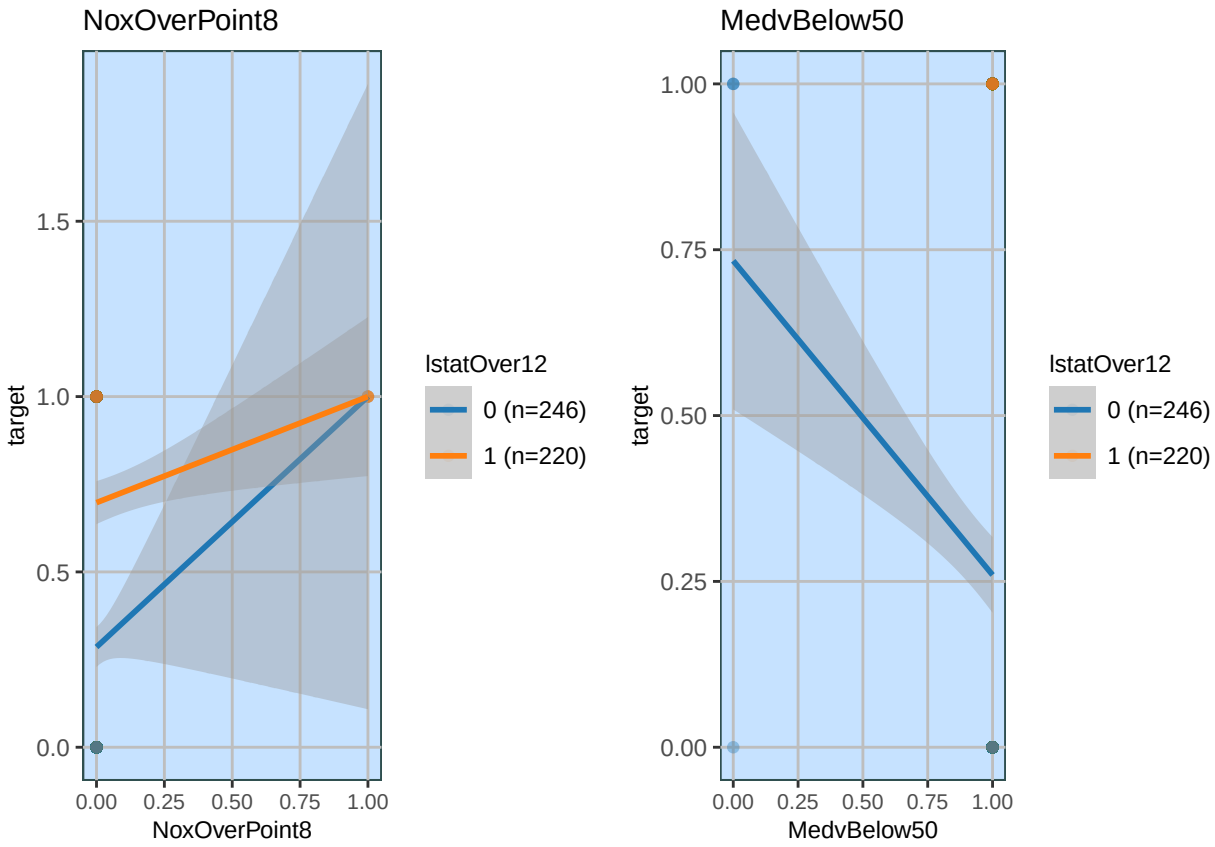

zn
target
zn
indus
chas
nox
rm
age
dis
rad
tax
IndusOver16
0 (n=293)
1 (n=173)

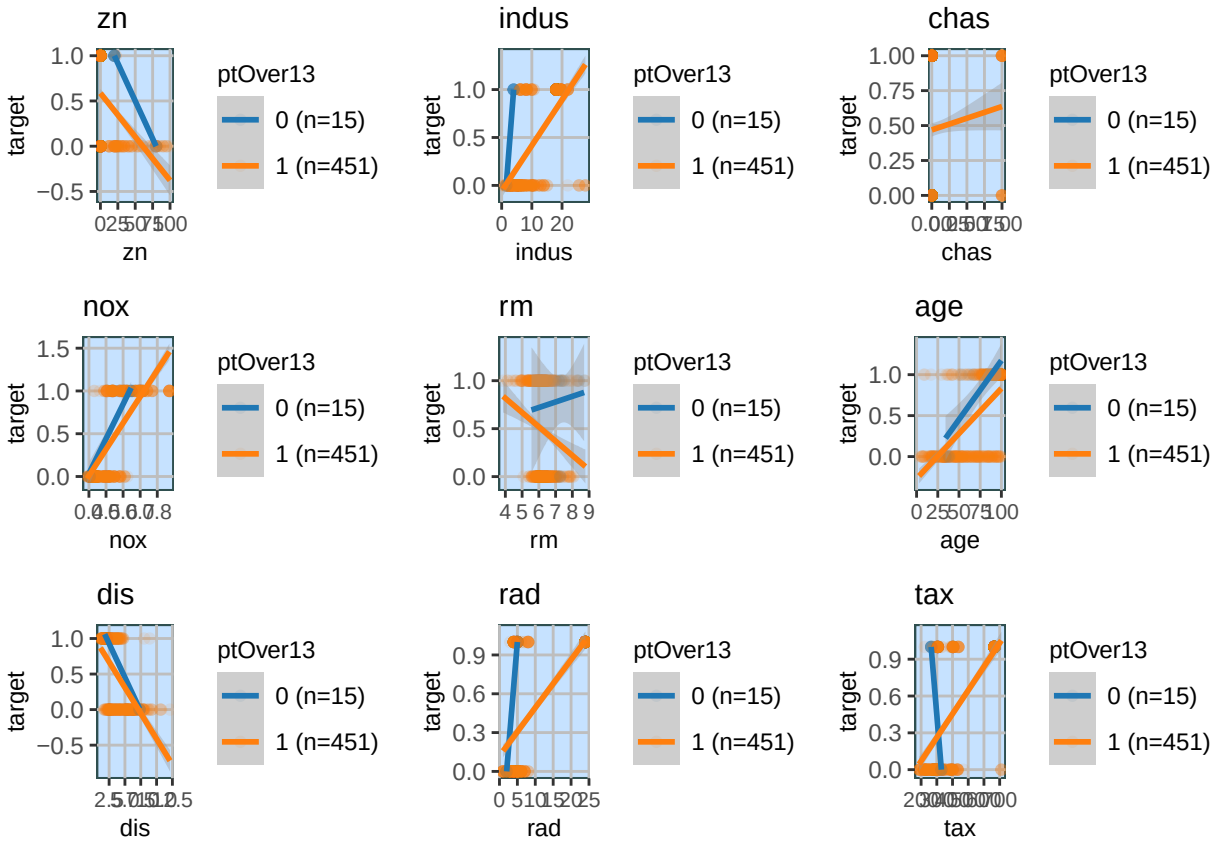

ptratio
target
ptratio
lstat
target
lstat
medv
target
medv
target
target
TaxOver600
target
TaxOver600
ptOver13
target
ptOver13
lstatOver12
target
lstatOver12
IndusOver16
target
IndusOver16
ZnOver0
target
ZnOver0
IndusOver16
0 (n=293)
1 (n=173)

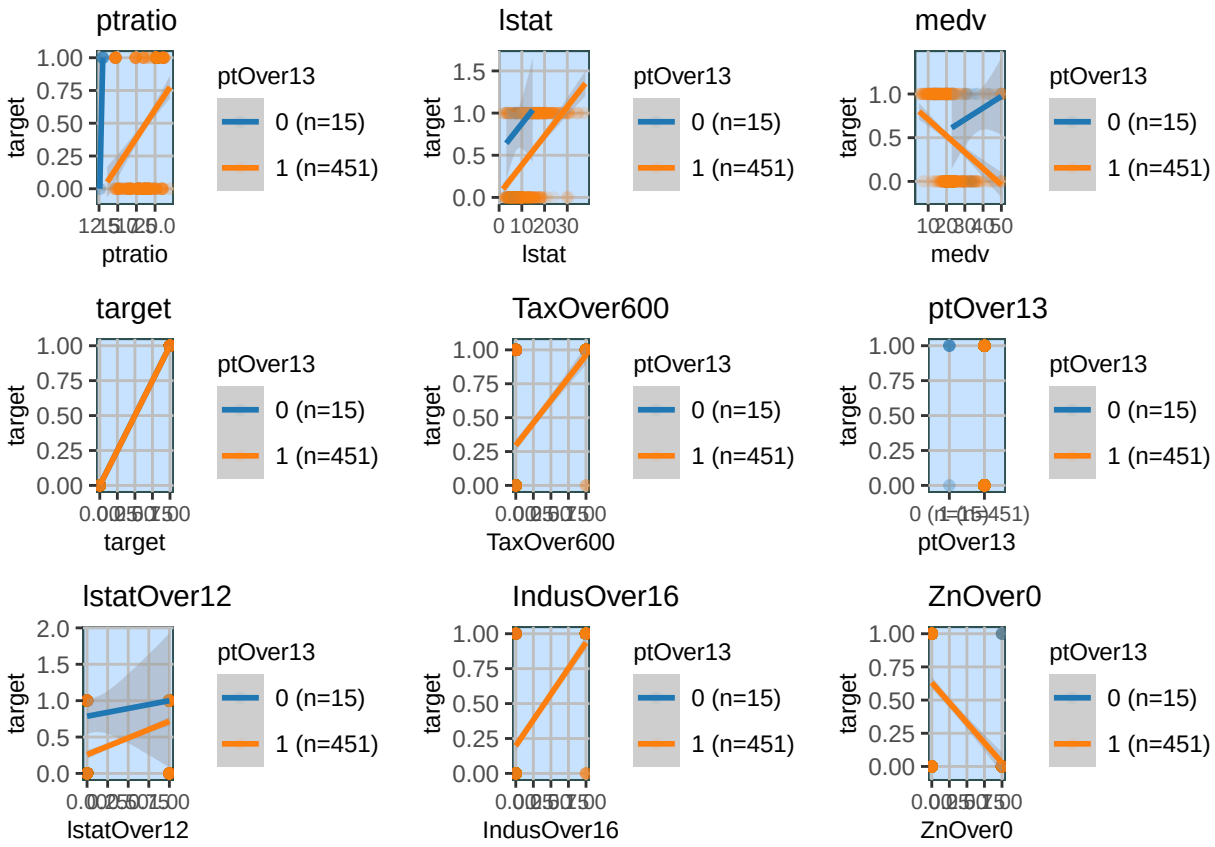

NoxOverPoint8
target
0.0
0.3
0.6
0.9
0.00
0.25
0.50
0.75
1.00
NoxOverPoint8
IndusOver16
0 (n=293)
1 (n=173)
MedvBelow50
target
0.0
0.3
0.6
0.9
1.2
0.00
0.25
0.50
0.75
1.00
MedvBelow50
IndusOver16
0 (n=293)
1 (n=173)

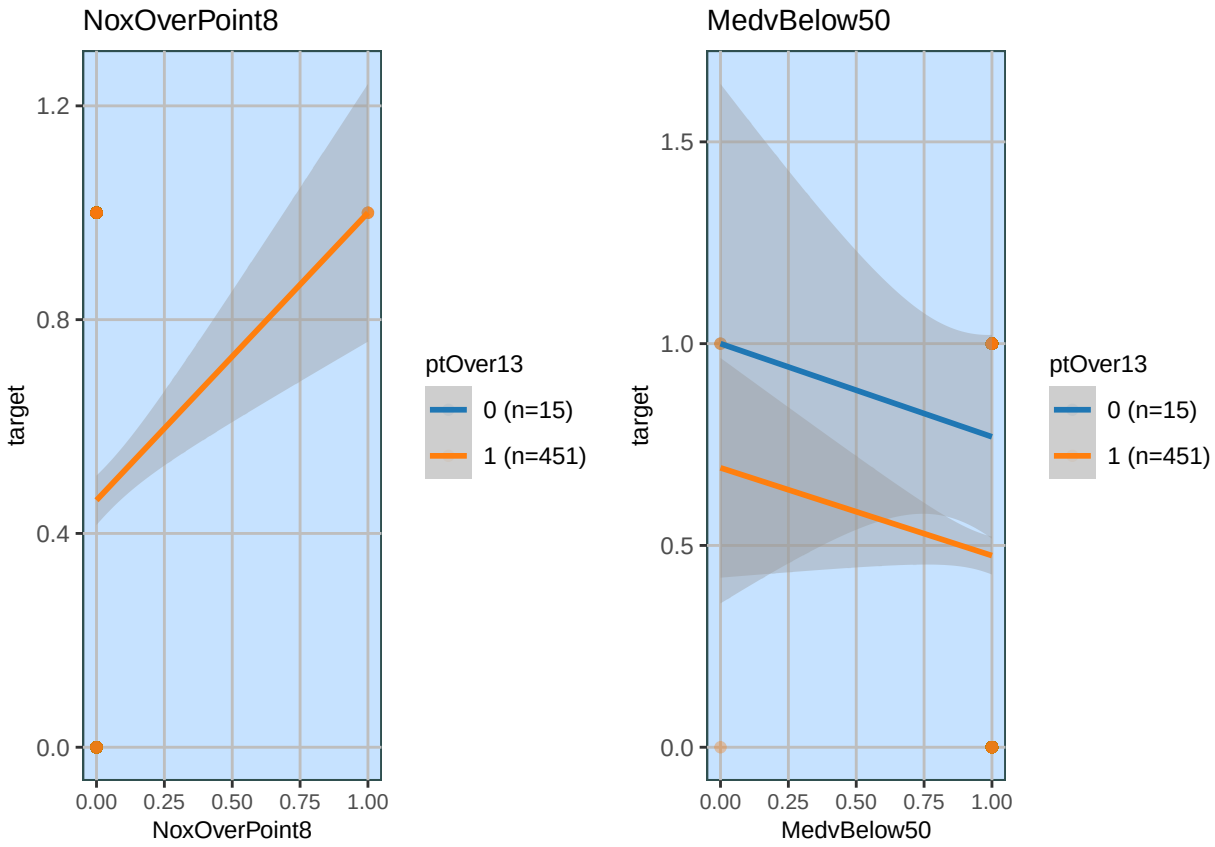

zn
indus
chas
nox
rm
age
dis
rad
tax
target
lstatOver12
0 (n=246)
1 (n=220)

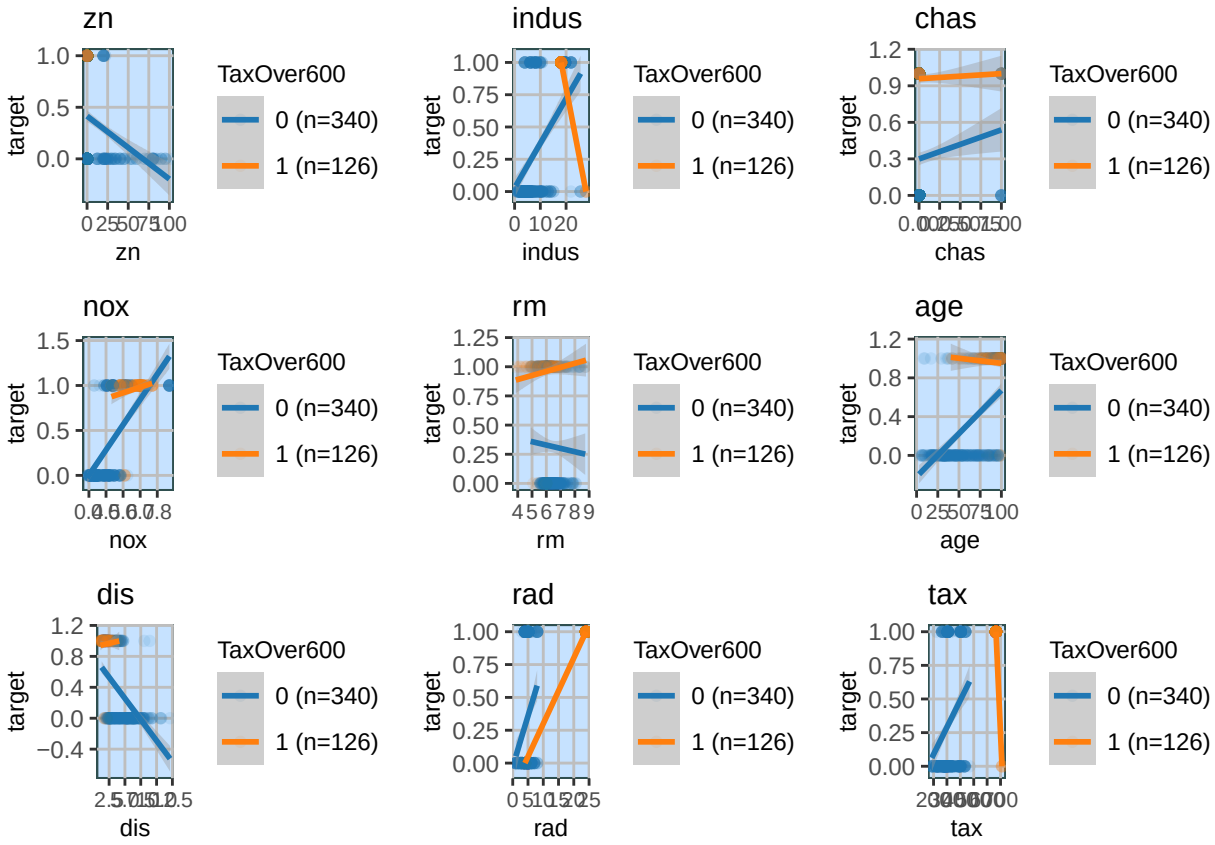

ptratio
lstat
medv
target
TaxOver600
ptOver13
lstatOver12
IndusOver16
ZnOver0
lstatOver12
0 (n=246)
1 (n=220)

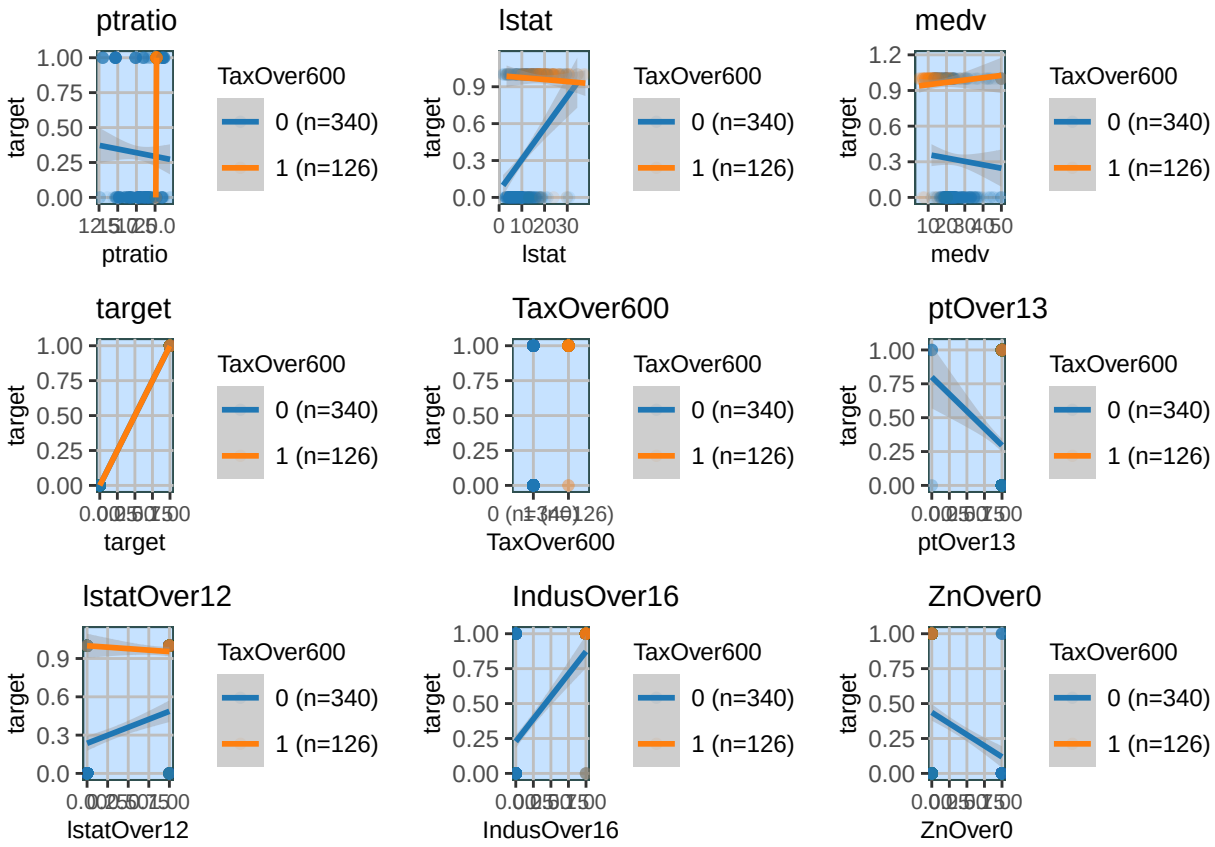

NoxOverPoint8
target
1.5
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00
NoxOverPoint8
lstatOver12
0 (n=246)
1 (n=220)
MedvBelow50
target
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
MedvBelow50
lstatOver12
0 (n=246)
1 (n=220)

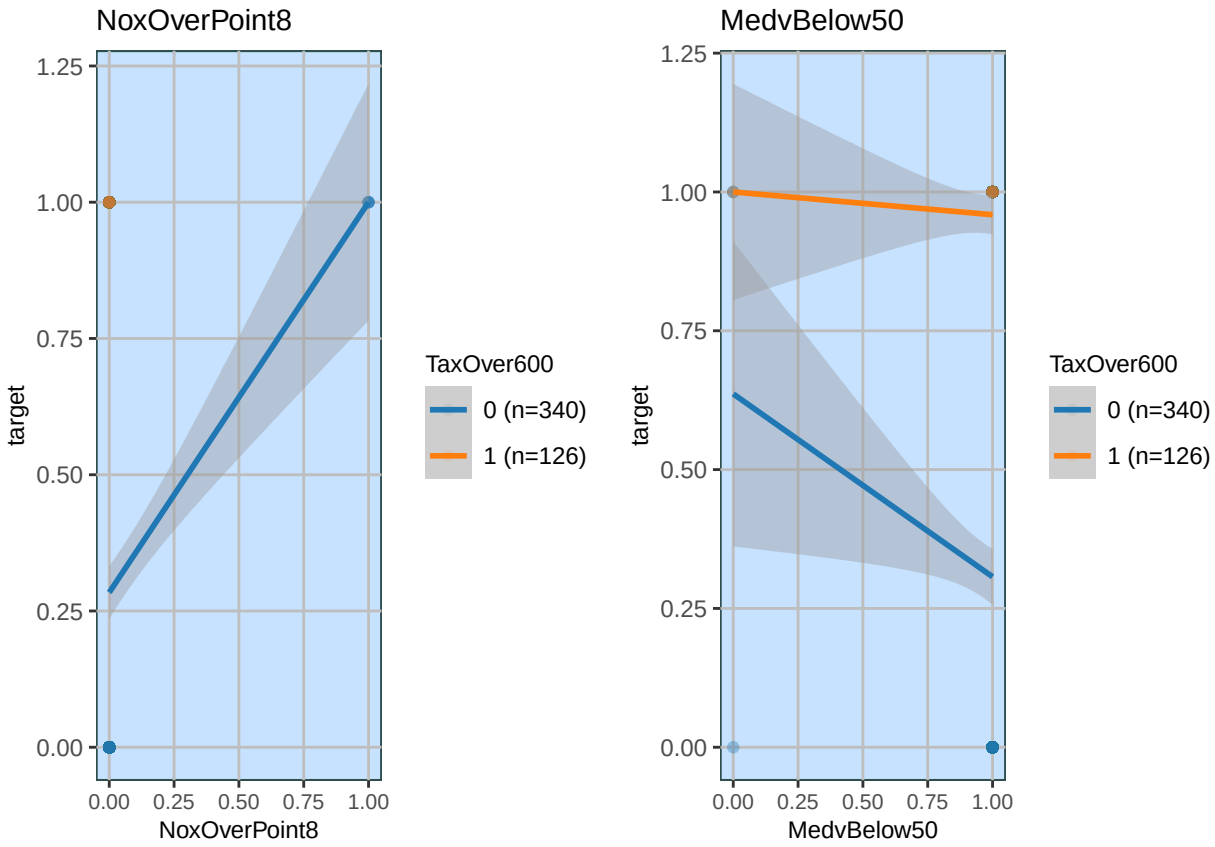

zn
target
zn
indus
target
indus
chas
target
chas
nox
target
nox
rm
target
rm
age
target
age
dis
target
dis
rad
target
rad
tax
target
tax
ptOver13
0 (n=15)
1 (n=451)


ptratio
target
ptratio
lstat
target
lstat
medv
target
medv
target
target
TaxOver600
target
TaxOver600
ptOver13
target
ptOver13
lstatOver12
target
lstatOver12
IndusOver16
target
IndusOver16
ZnOver0
target
ZnOver0
ptOver13
0 (n=15)
1 (n=451)


NoxOverPoint8
target
1.2
0.8
0.4
0.0
0.00
0.25
0.50
0.75
1.00
NoxOverPoint8
ptOver13
0 (n=15)
1 (n=451)
MedvBelow50
target
1.5
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00
MedvBelow50
ptOver13
0 (n=15)
1 (n=451)


zn
target
zn
indus
chas
nox
rm
age
dis
rad
tax
TaxOver600
0 (n=340)
1 (n=126)

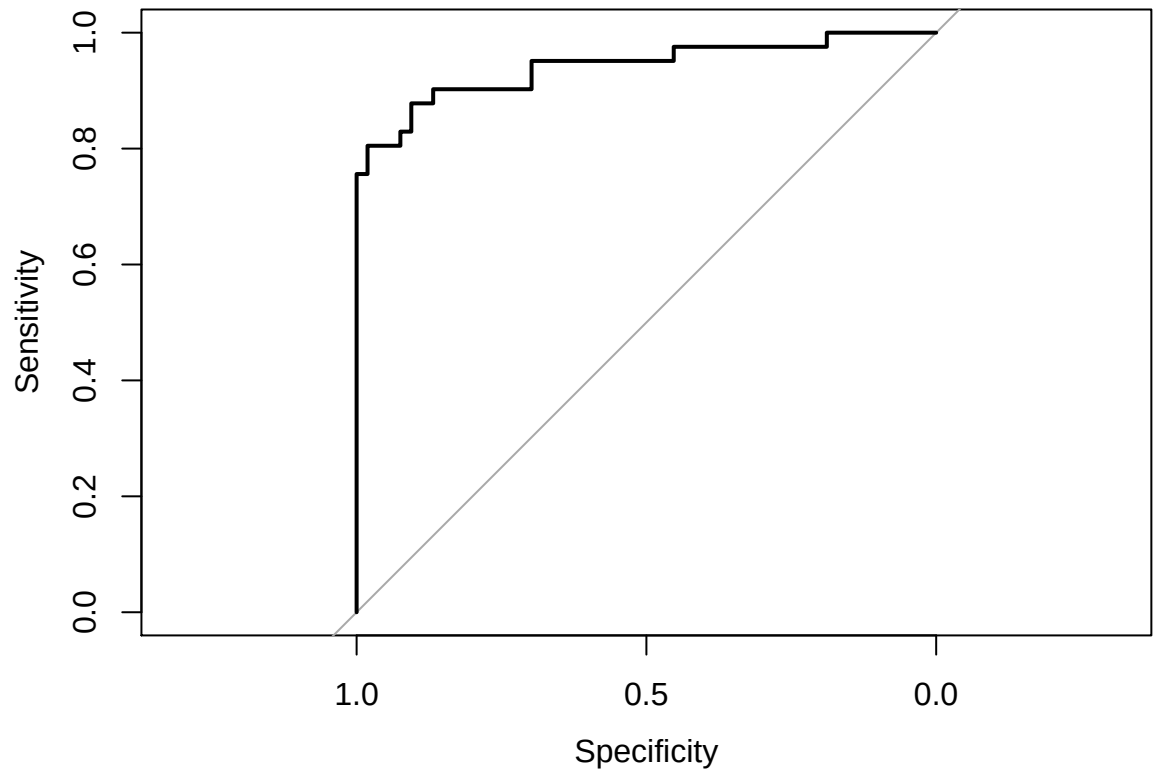

ptratio
target
ptratio
lstat
target
lstat
medv
target
medv
target
target
TaxOver600
ptOver13
target
ptOver13
lstatOver12
target
lstatOver12
IndusOver16
target
IndusOver16
ZnOver0
target
ZnOver0
TaxOver600
0 (n=340)
1 (n=126)




```
##
##       0   1
##   0 323  16
##   1 127   0

##
## Call:  glm(formula = fla, family = "binomial", data = df)
##
## Coefficients:
##          (Intercept)                   zn                indus
##            -2.612e+01           -3.788e-02           -4.792e-02
##                 chas                  nox                   rm
##             8.712e-01            3.727e+01           -1.104e+00
##                  age                  dis                  rad
##             3.522e-02            1.045e+00            6.873e-01
##                  tax              ptratio                lstat
##             4.949e-04            2.624e-01            1.671e-01
##                 medv           TaxOver600             ptOver13
##             2.033e-01            5.467e+01           -7.886e+00
##          lstatOver12          IndusOver16               ZnOver0
##            -1.057e+01            1.665e+00           -3.367e+00
##        NoxOverPoint8          MedvBelow50           inter_z_rm
##             1.147e+01            3.461e-01           -1.366e+00
##         inter_z_medv        inter_z_indus inter_lstatOver12_pt
##             2.546e-01            5.444e-01            4.796e-01
```

```
##         inter_tax_rm
##           -2.790e+00
##
## Degrees of Freedom: 465 Total (i.e. Null);  441 Residual
## Null Deviance:       645.9
## Residual Deviance: 166   AIC: 216
##
## Call:
## glm(formula = fla, family = "binomial", data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5992  -0.1641  -0.0001   0.0001   4.0374
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -2.612e+01  1.022e+01  -2.555 0.010621 *
## zn                   -3.788e-02  7.487e-02  -0.506 0.612936
## indus                -4.792e-02  1.051e-01  -0.456 0.648389
## chas                  8.712e-01  8.764e-01   0.994 0.320159
## nox                   3.727e+01  9.491e+00   3.927  8.6e-05 ***
## rm                   -1.104e+00  9.218e-01  -1.198 0.230863
## age                   3.522e-02  1.555e-02   2.265 0.023534 *
## dis                   1.045e+00  2.864e-01   3.650 0.000263 ***
## rad                   6.873e-01  1.955e-01   3.515 0.000439 ***
## tax                   4.949e-04  4.224e-03   0.117 0.906745
## ptratio               2.624e-01  2.097e-01   1.251 0.210900
## lstat                 1.671e-01  8.198e-02   2.039 0.041494 *
## medv                  2.033e-01  8.903e-02   2.283 0.022421 *
## TaxOver600            5.467e+01  1.496e+04   0.004 0.997085
## ptOver13             -7.886e+00  4.508e+00  -1.749 0.080247 .
## lstatOver12          -1.057e+01  5.131e+00  -2.060 0.039432 *
## IndusOver16           1.665e+00  1.578e+00   1.055 0.291230
## ZnOver0              -3.367e+00  1.182e+01  -0.285 0.775829
## NoxOverPoint8         1.147e+01  4.209e+03   0.003 0.997825
## MedvBelow50           3.461e-01  2.194e+00   0.158 0.874693
## inter_z_rm           -1.366e+00  2.641e+00  -0.517 0.604987
## inter_z_medv          2.546e-01  2.381e-01   1.069 0.285056
## inter_z_indus         5.444e-01  3.705e-01   1.469 0.141732
## inter_lstatOver12_pt  4.796e-01  2.662e-01   1.801 0.071659 .
## inter_tax_rm         -2.790e+00  8.118e+02  -0.003 0.997258
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 645.88  on 465  degrees of freedom
## Residual deviance: 165.97  on 441  degrees of freedom
## AIC: 215.97
##
## Number of Fisher Scoring iterations: 19
##
## Confusion Matrix and Statistics
##
```

```
##           Reference
## Prediction  0  1
##          0 48  7
##          1  5 34
##
##                Accuracy : 0.8723
##                  95% CI : (0.7876, 0.9323)
##     No Information Rate : 0.5638
##     P-Value [Acc > NIR] : 1.089e-10
##
##                   Kappa : 0.739
##
##  Mcnemar's Test P-Value : 0.7728
##
##             Sensitivity : 0.9057
##             Specificity : 0.8293
##          Pos Pred Value : 0.8727
##          Neg Pred Value : 0.8718
##              Prevalence : 0.5638
##          Detection Rate : 0.5106
##    Detection Prevalence : 0.5851
##       Balanced Accuracy : 0.8675
##
##        'Positive' Class : 0
##
```



```
## [1] "AUC:  0.941555453290382"
##
## Call:
## roc.default(response = dfPred_raw$class, predictor = dfPred_raw$predict_reg,     plot = TRUE)
##
## Data: dfPred_raw$predict_reg in 53 controls (dfPred_raw$class 0) < 41 cases (dfPred_raw$class 1).
## Area under the curve: 0.9416
```

```
##
## Call:  glm(formula = fla, family = "binomial", data = df)
##
## Coefficients:
##          (Intercept)                   zn                indus
##           -2.612e+01           -3.788e-02           -4.792e-02
##                 chas                  nox                   rm
##            8.712e-01            3.727e+01           -1.104e+00
##                  age                  dis                  rad
##            3.522e-02            1.045e+00            6.873e-01
##                  tax              ptratio                lstat
##            4.949e-04            2.624e-01            1.671e-01
##                 medv           TaxOver600             ptOver13
##            2.033e-01            5.467e+01           -7.886e+00
##          lstatOver12          IndusOver16              ZnOver0
##           -1.057e+01            1.665e+00           -3.367e+00
##        NoxOverPoint8          MedvBelow50           inter_z_rm
##            1.147e+01            3.461e-01           -1.366e+00
##         inter_z_medv        inter_z_indus  inter_lstatOver12_pt
##            2.546e-01            5.444e-01            4.796e-01
##         inter_tax_rm
##           -2.790e+00
##
## Degrees of Freedom: 465 Total (i.e. Null);  441 Residual
## Null Deviance:       645.9
## Residual Deviance: 166   AIC: 216
```