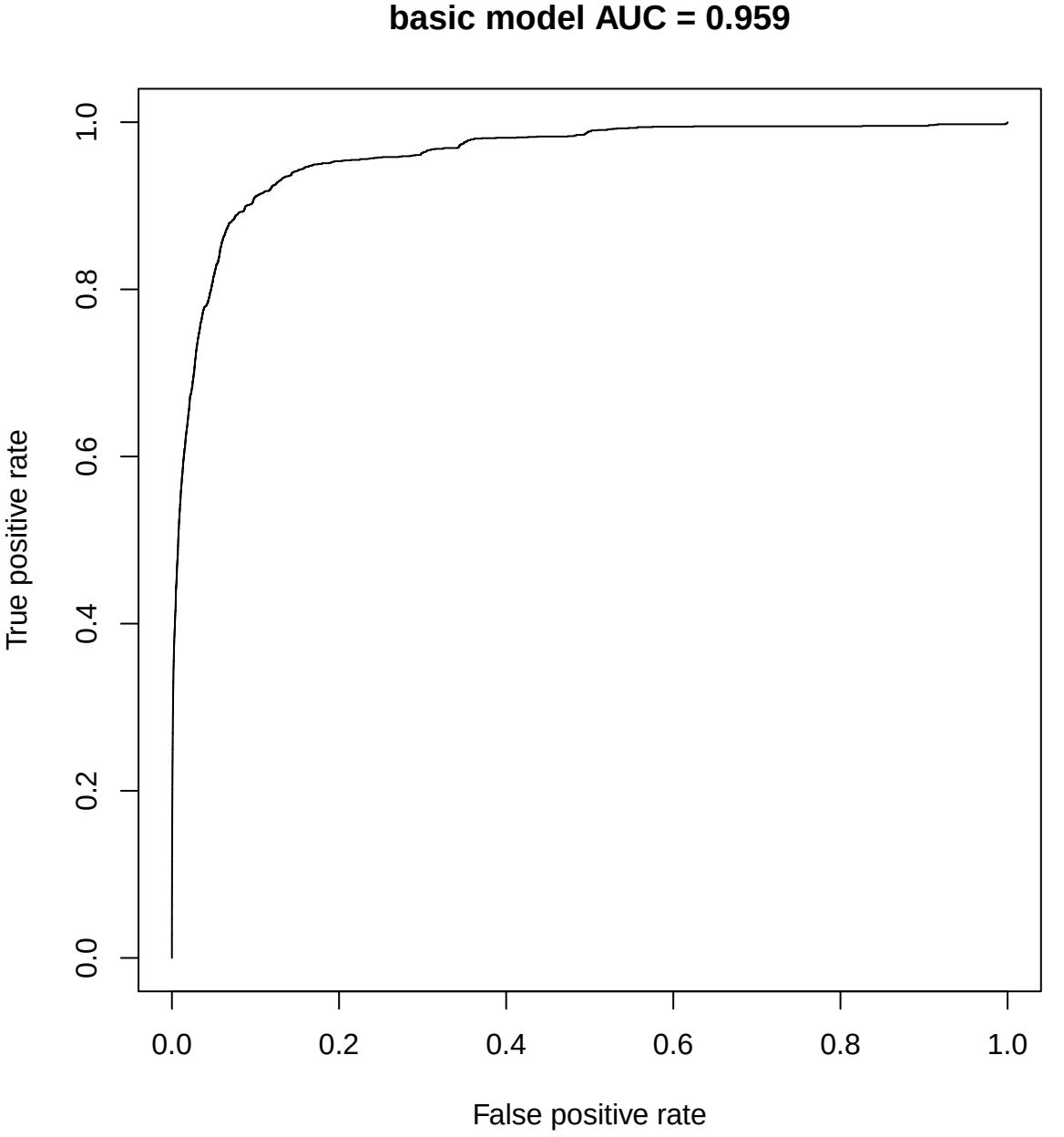
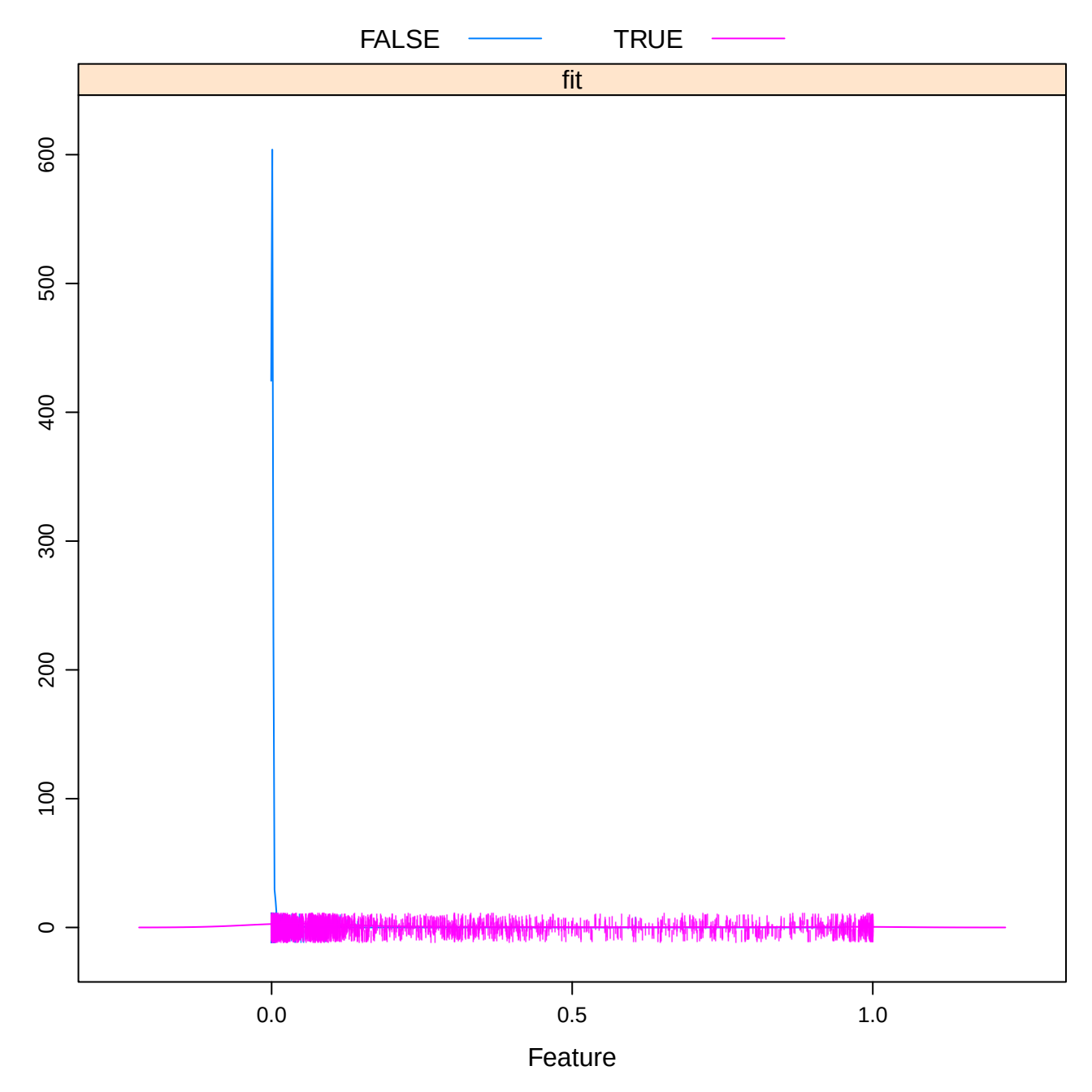
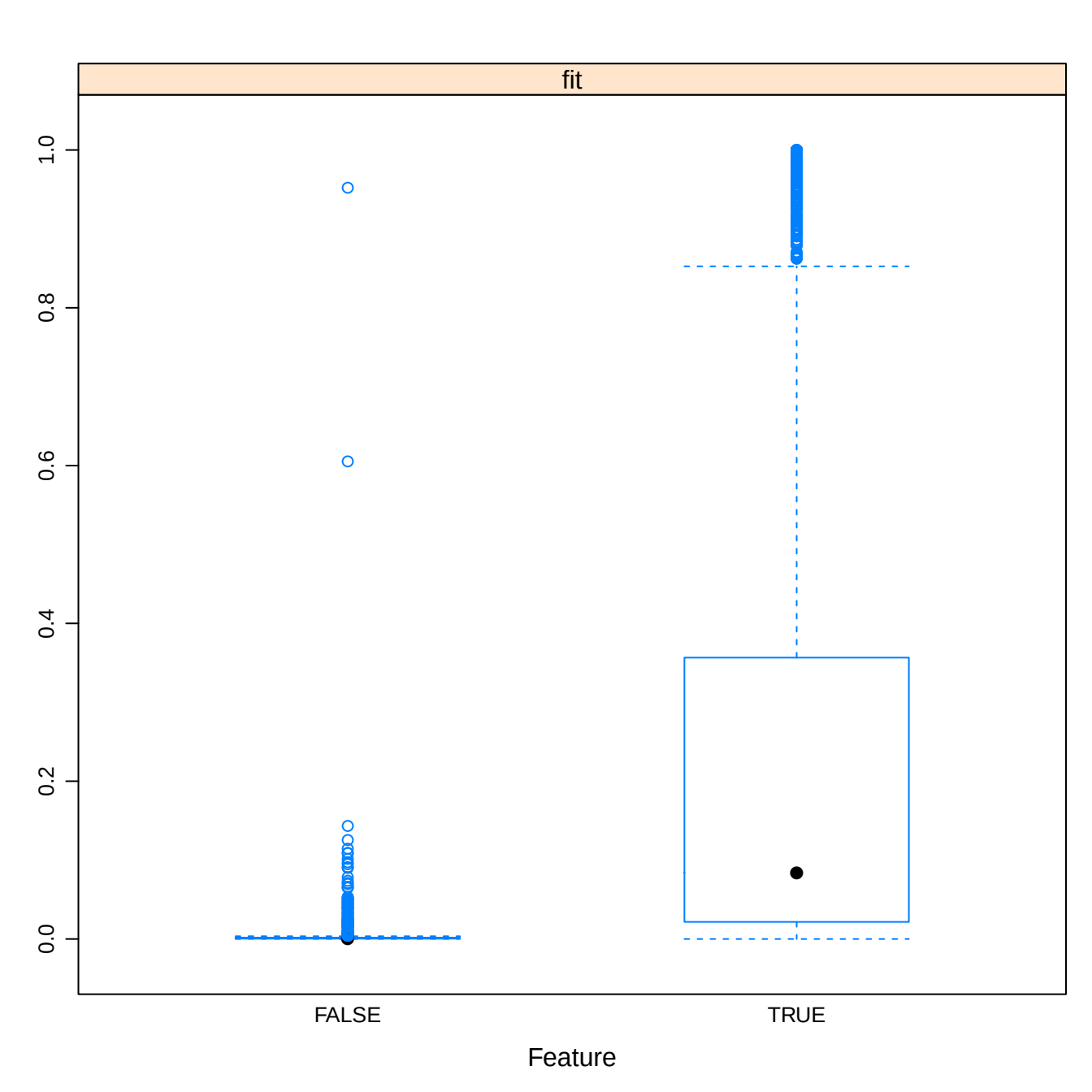
```
$project
[1] "DA"

$platform
[1] "google_play"

$dig
[1] 3

$dataFile
[1] "payer_model_DA_GP&iOS_mkt_2019-04-01_2019-06-30.rds"

$sampleSize
[1] 8e+05

$testSampleSplit
[1] 0.5

$seed
[1] 1
```

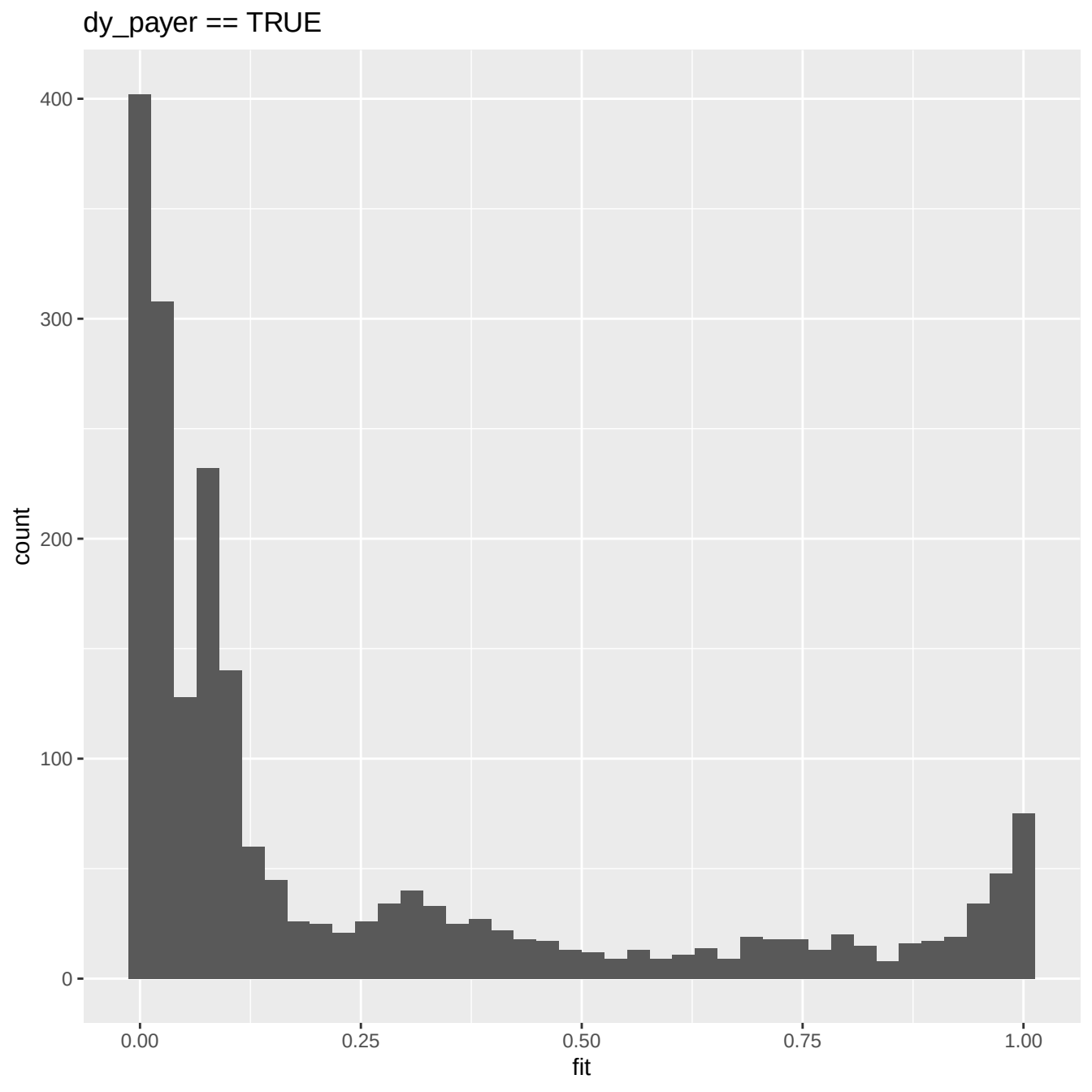
Transformation by Box−Cox together with scaling seems to slightly improve
the performance of the model. We should run CV to be sure, since
this is only one sample.
But transformation definitely has some potential, since we can see in the histogram
that the transformed fit better distinguishes between positives and negatives.

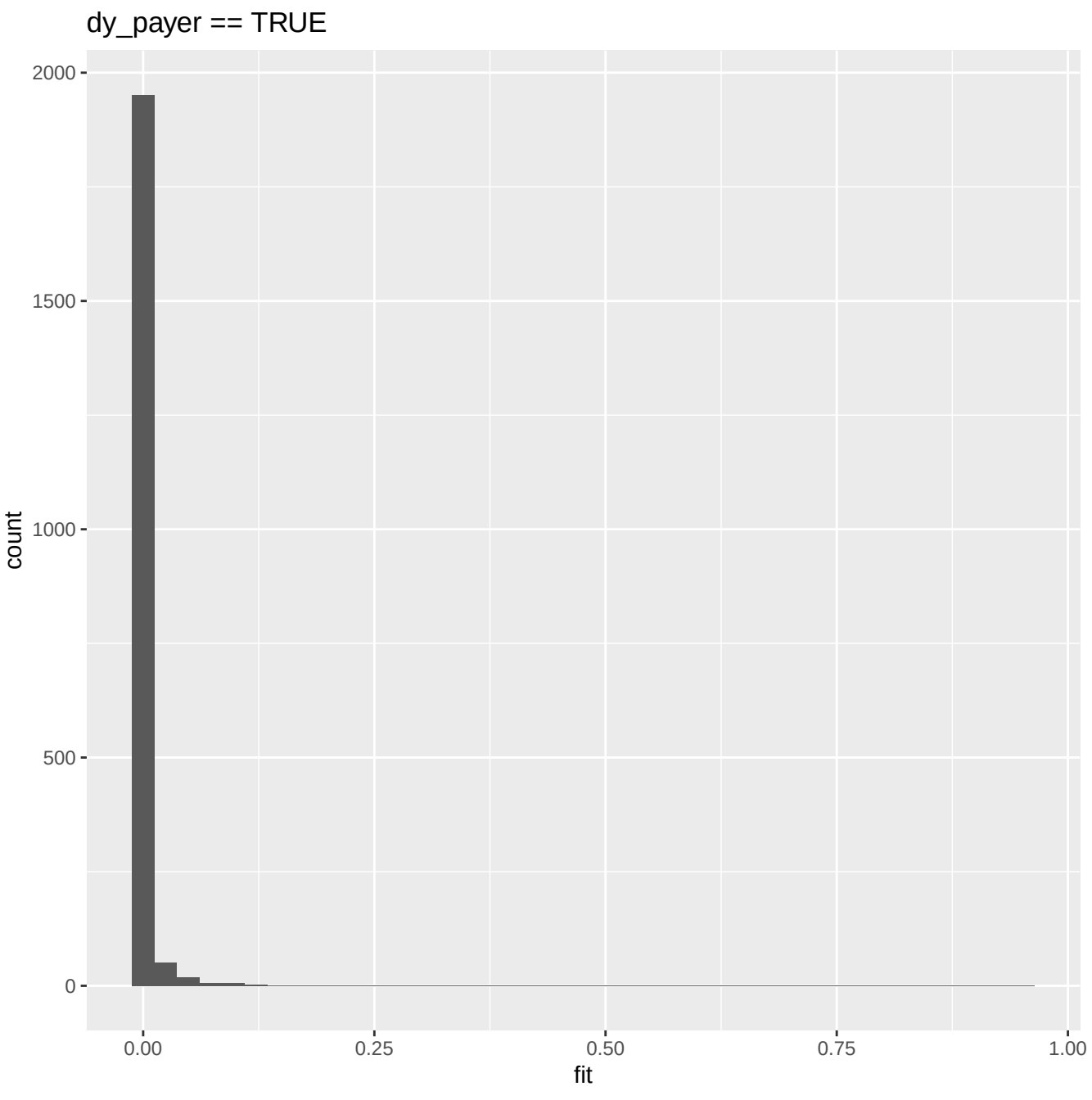

basic model AUC = 0.959
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

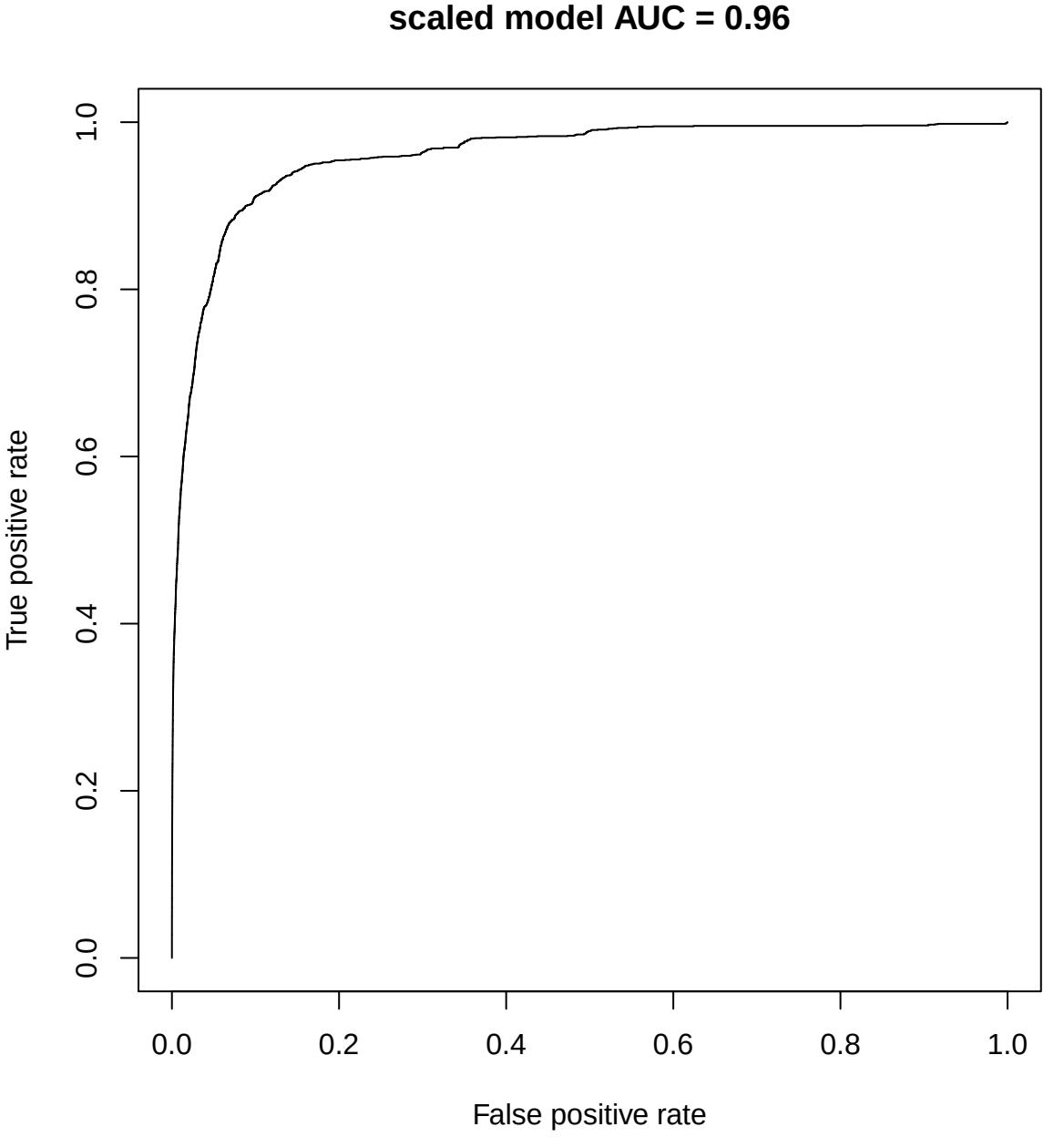
```
$note
[1] "no scaling or tranformation"

$rcd_optimal_cut_off
[1] 0.1312741

$rcd_optimal_confusion_matrix

         FALSE   TRUE
  FALSE 396739   1252
  TRUE    1222    787

$sensitivity
[1] 0.3859735

$precision
[1] 0.3917372

$relative_count_difference
[1] 0.9852869
```


FALSE
TRUE
fit
600
500
400
300
200
100
0
0.0
0.5
1.0
Feature

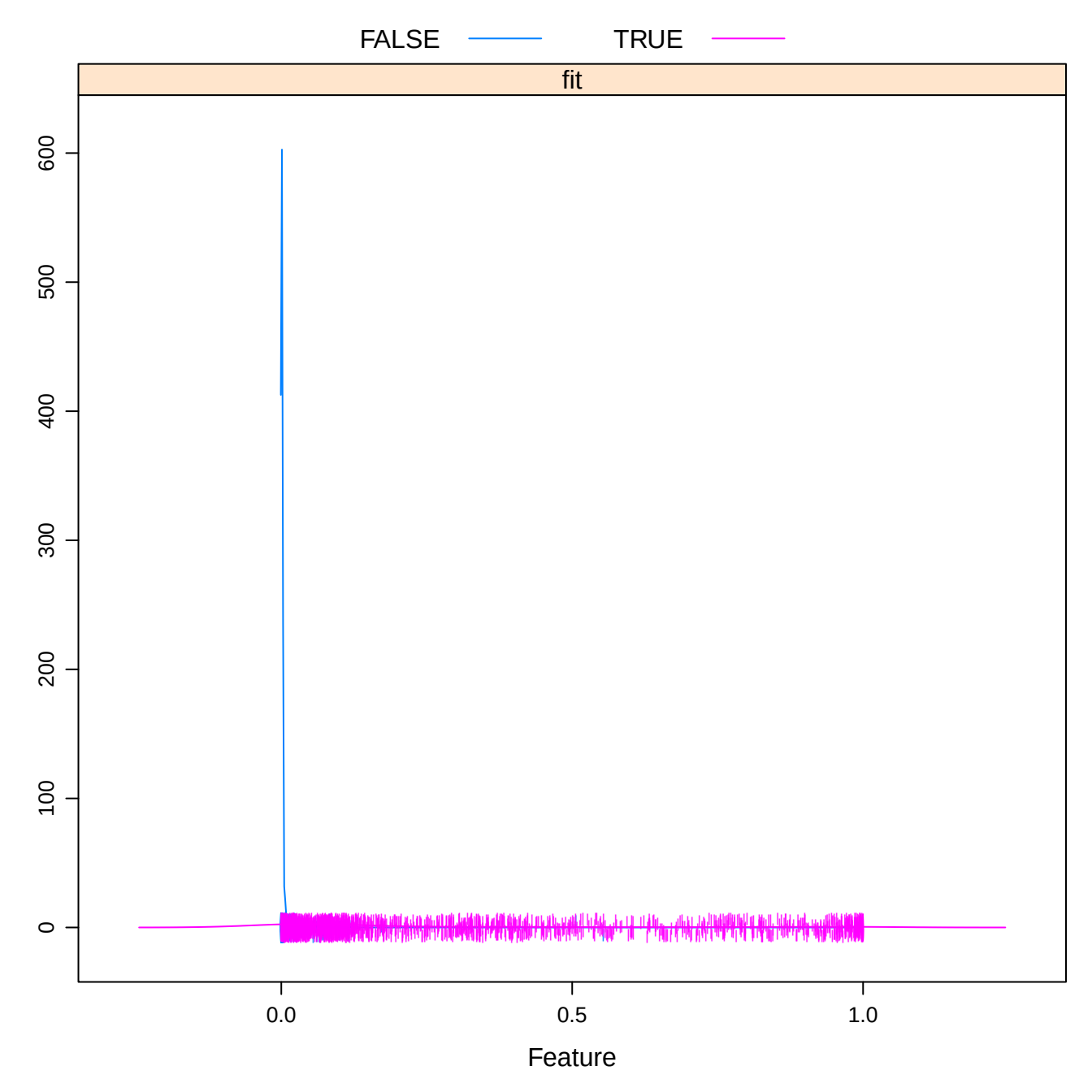

fit
1.0
0.8
0.6
0.4
0.2
0.0
FALSE
TRUE
Feature

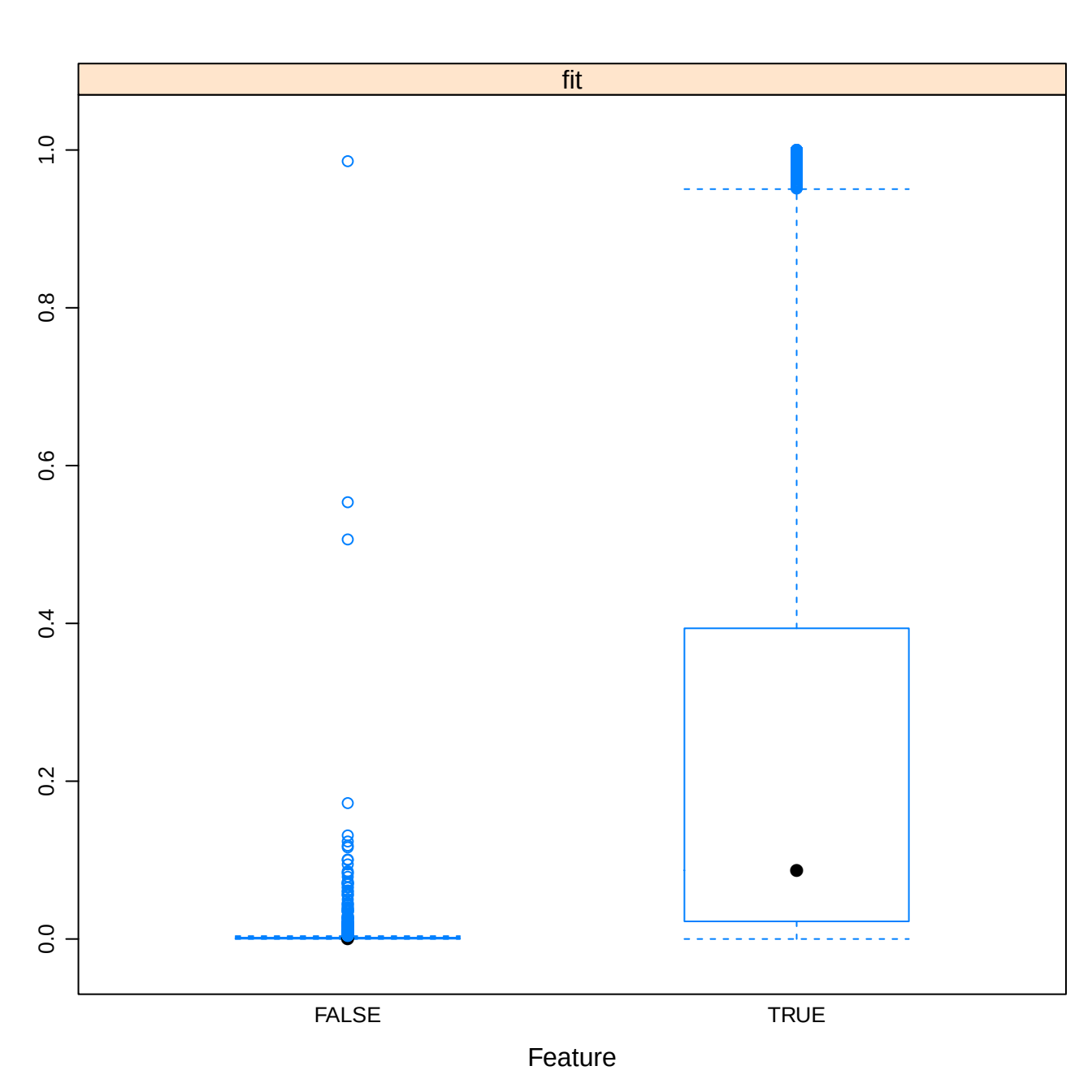

dy_payer == TRUE
count
400
300
200
100
0
0.00
0.25
0.50
0.75
1.00
fit

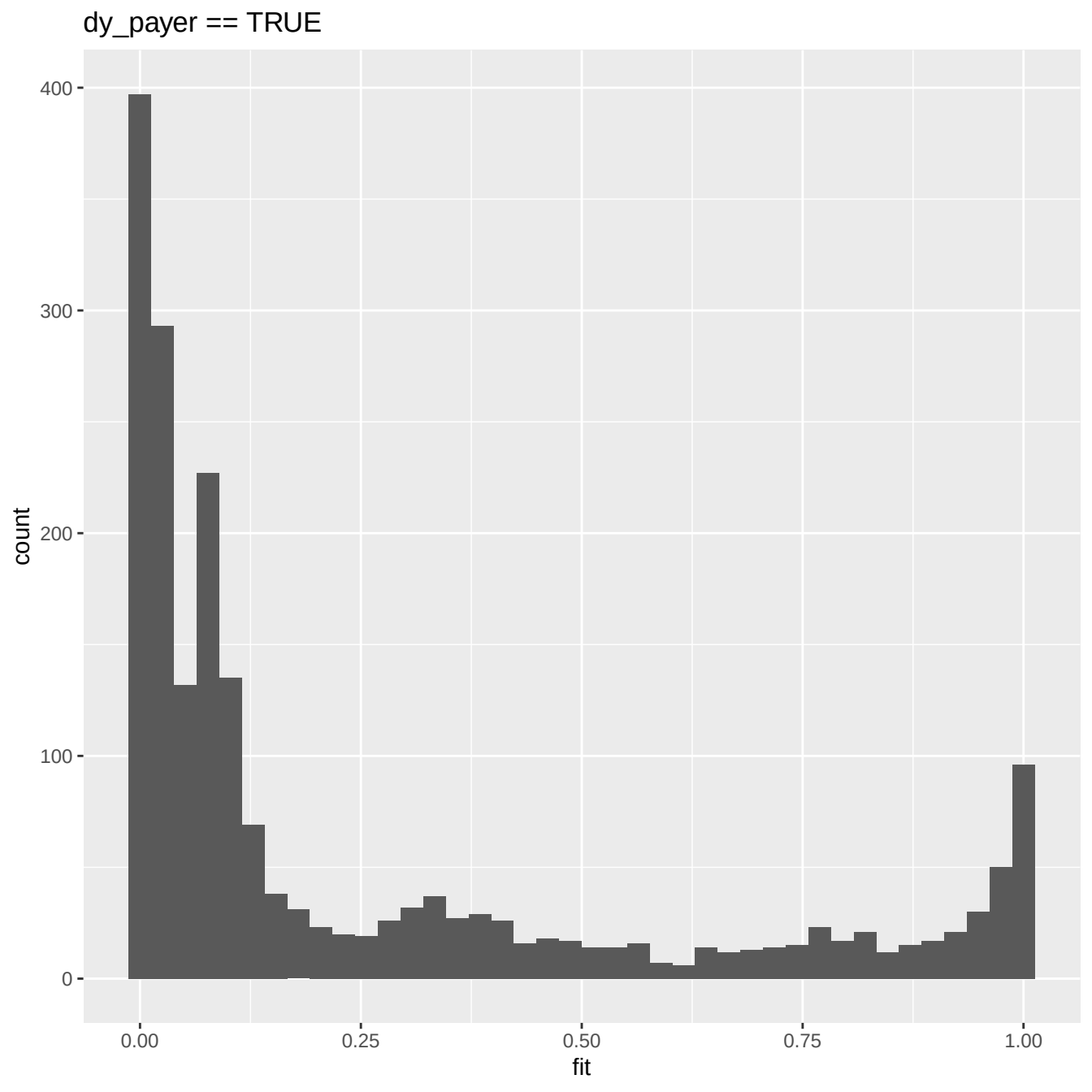

dy_payer == TRUE
2000
1500
1000
500
0
count
0.00
0.25
0.50
0.75
1.00
fit

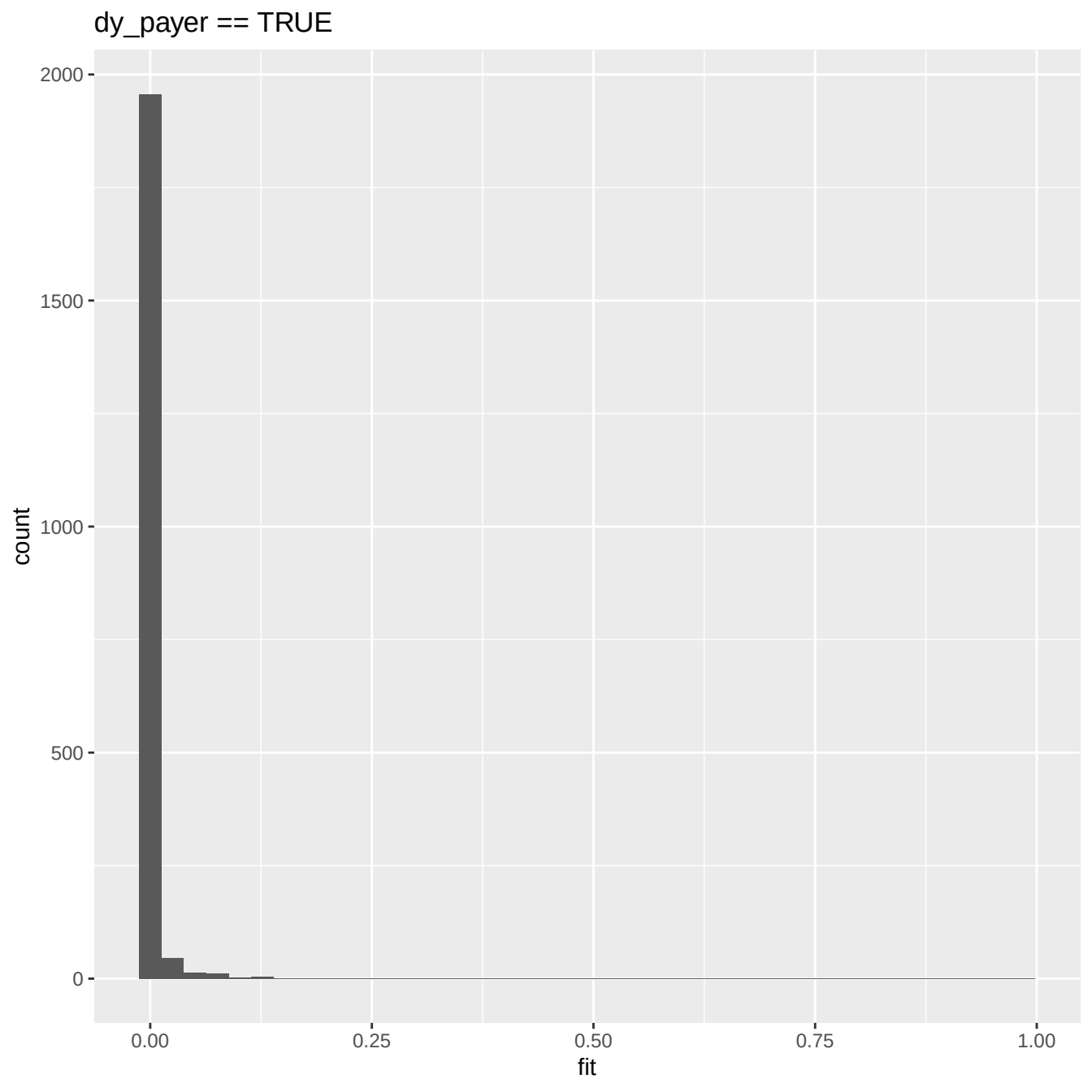

scaled model AUC = 0.96
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

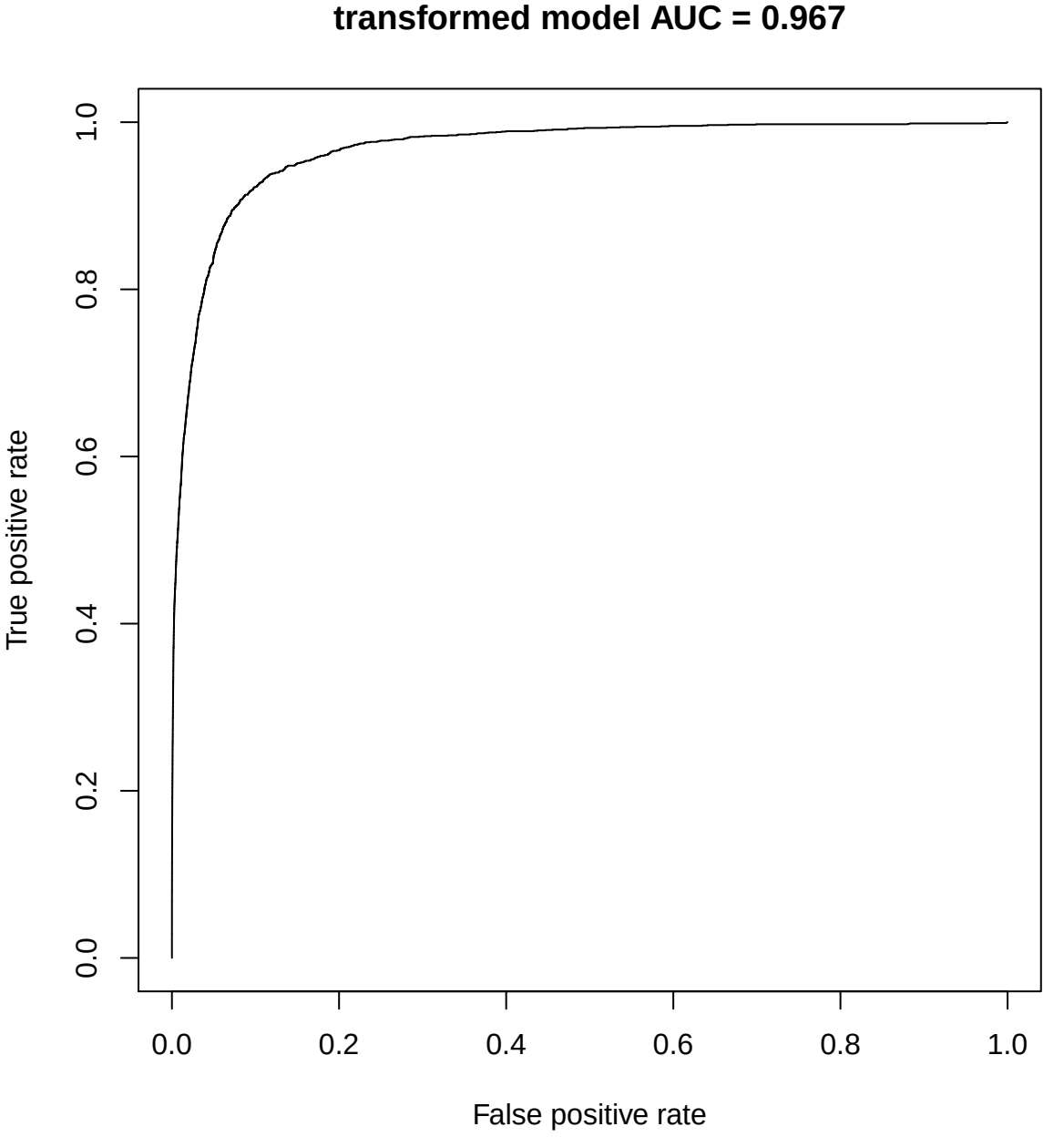
```
$note
[1] "scaled/standardized features"

$rcd_optimal_cut_off
[1] 0.1312741

$rcd_optimal_confusion_matrix

         FALSE   TRUE
  FALSE 396655   1235
  TRUE    1306    804

$sensitivity
[1] 0.3943109

$precision
[1] 0.3810427

$relative_count_difference
[1] 1.034821
```


FALSE
TRUE
fit
600
500
400
300
200
100
0
0.0
0.5
1.0
Feature

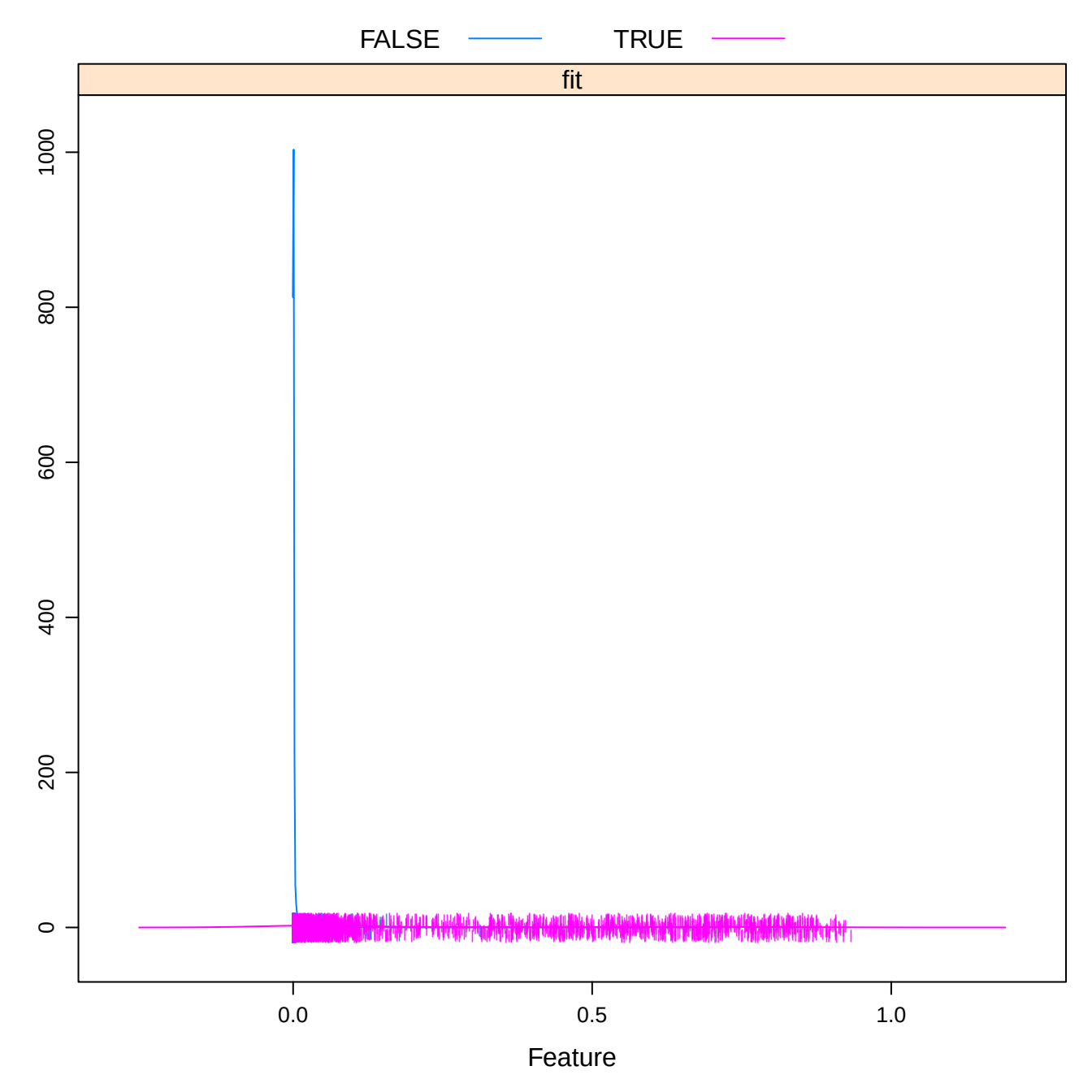

fit
1.0
0.8
0.6
0.4
0.2
0.0
FALSE
TRUE
Feature

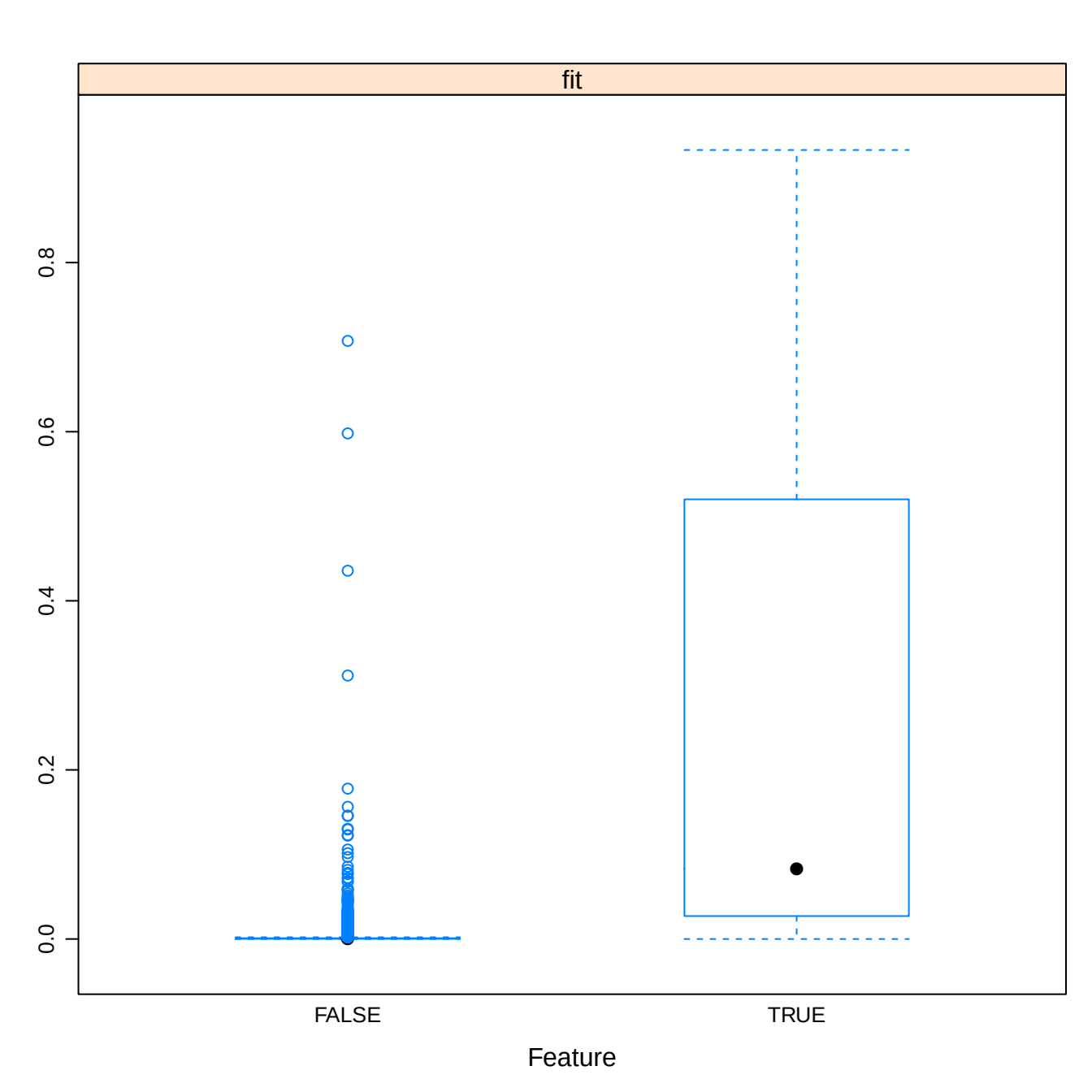

dy_payer == TRUE
400
300
200
100
0
count
0.00
0.25
0.50
0.75
1.00
fit

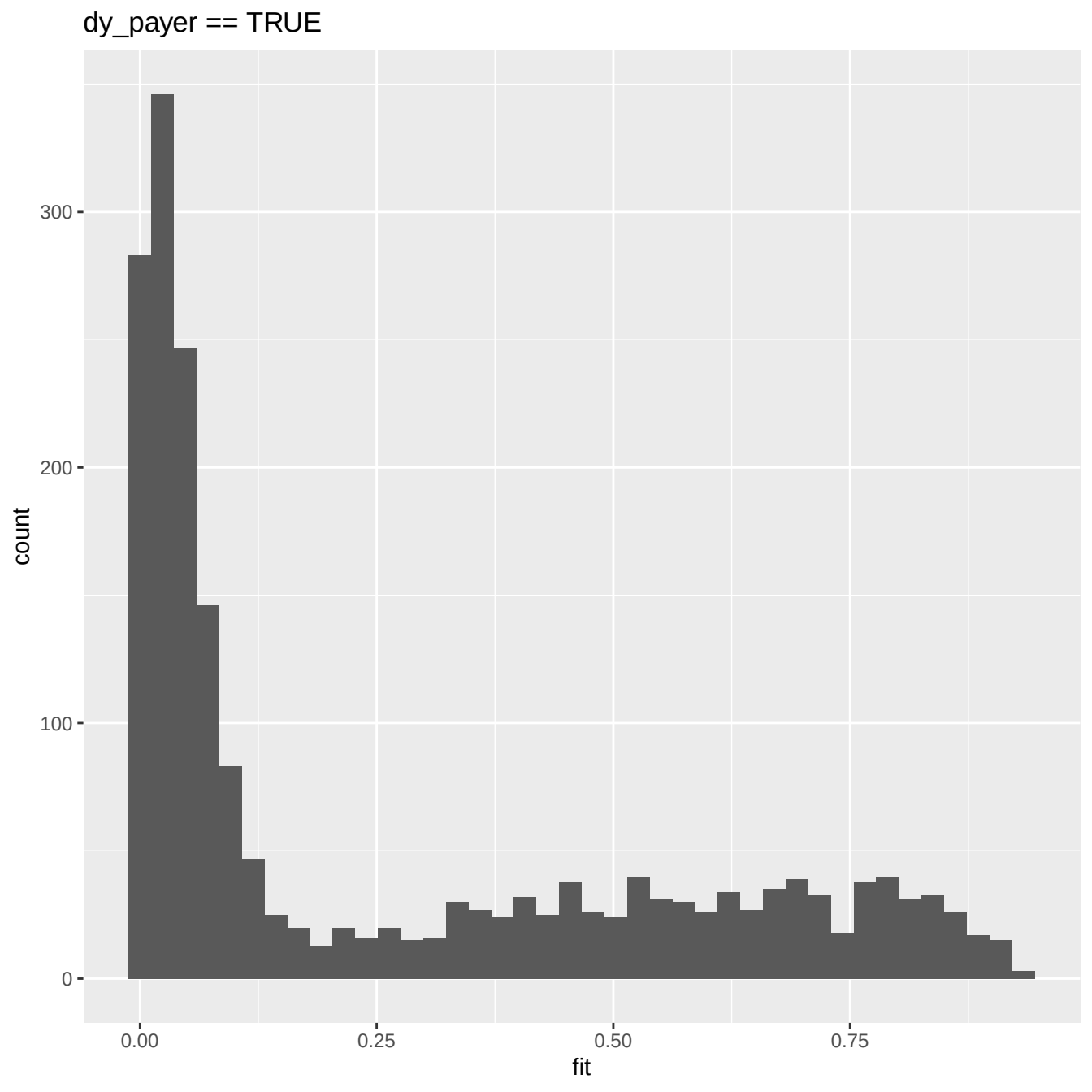

dy_payer == TRUE
2000
1500
1000
500
0
count
0.00
0.25
0.50
0.75
1.00
fit

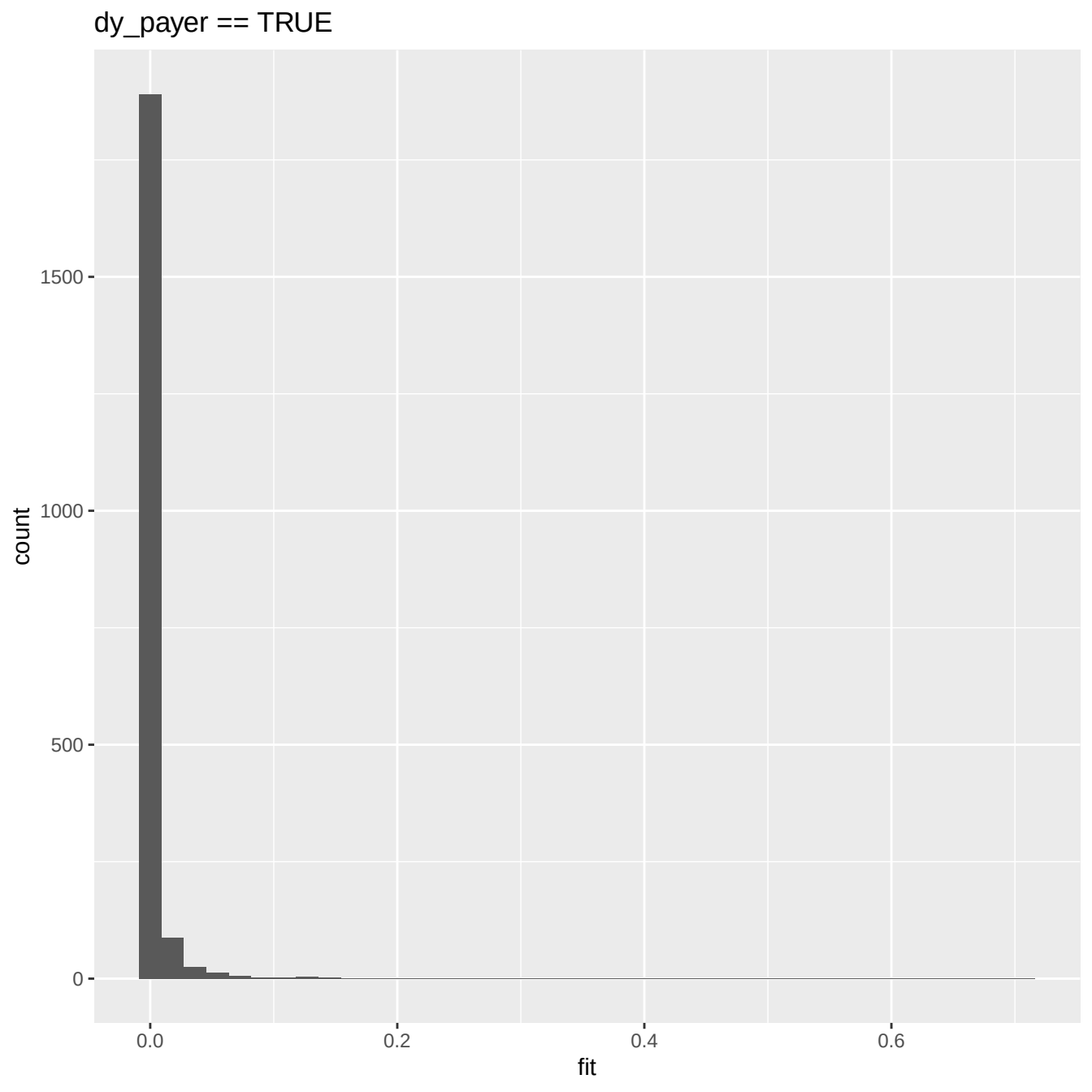

transformed model AUC = 0.967
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

```
$note
[1] "features transformed and scaled"

$rcd_optimal_cut_off
[1] 0.1596385

$rcd_optimal_confusion_matrix

         FALSE   TRUE
  FALSE 396779   1179
  TRUE    1182    860

$sensitivity
[1] 0.4217754

$precision
[1] 0.4211557

$relative_count_difference
[1] 1.001471
```


FALSE
TRUE
fit
1000
800
600
400
200
0
0.0
0.5
1.0
Feature


fit
0.0
0.2
0.4
0.6
0.8
FALSE
TRUE
Feature


dy_payer == TRUE
300
200
100
0
count
0.00
0.25
0.50
0.75
fit


dy_payer == TRUE
count
1500
1000
500
0
0.0
0.2
0.4
0.6
fit