

BankRuptcy

Batch 30 Group 05

September 02, 2017

Contents

Cute 03 - Bank Financial - Bankruptcy Prediction	1
Pre- Processing	1
Install Packages required	1
Understanding the data	2
Assign data types	11
Imputation of missing values	11
Normalising	12
Final dataframe	15
Outlier removal	15
Plots	16
To observe the correlation between the attributes	74
Build model	77
Predictions on train data	77
Predictions on train data	78

Cute 03 - Bank Financial - Bankruptcy Prediction

*Clear the environment.

```
rm(list = ls(all=TRUE))
```

Pre- Processing

- Set working directory and Read dataset

```
setwd("F:\\INSOFE\\MachineLearning\\Cute03\\Final")
bank_data <- read.table(file = "bankdata.csv", header = T, sep = ",")
```

Install Packages required

```
library(knitr)
library(DMwR)

## Loading required package: lattice
## Loading required package: grid
library(vegan)

## Loading required package: permute
```

```

## This is vegan 2.4-3
library(caret)

## Loading required package: ggplot2

##
## Attaching package: 'caret'

## The following object is masked from 'package:vegan':
##
##     tolerance

library(corrplot)
library(randomForest)

## randomForest 4.6-12

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin

library(forecast)
library(e1071)
library(Amelia)

## Loading required package: Rcpp

## ##
## ## Amelia II: Multiple Imputation
## ## (Version 1.7.4, built: 2015-12-05)
## ## Copyright (C) 2005-2017 James Honaker, Gary King and Matthew Blackwell
## ## Refer to http://gking.harvard.edu/amelia/ for more information
## ##

```

Understanding the data

```

str(bank_data)

## 'data.frame':    43004 obs. of  65 variables:
##   $ Attr1 : num  0.2006 0.2091 0.2487 0.0815 0.1873 ...
##   $ Attr2 : num  0.38 0.5 0.696 0.307 0.613 ...
##   $ Attr3 : num  0.396 0.472 0.267 0.459 0.23 ...
##   $ Attr4 : num  2.05 1.94 1.55 2.49 1.41 ...
##   $ Attr5 : num  32.35 14.79 -1.15 51.95 -7.31 ...
##   $ Attr6 : num  0.388 0 0 0.15 0.187 ...
##   $ Attr7 : num  0.2498 0.2583 0.3091 0.0927 0.1873 ...
##   $ Attr8 : num  1.331 0.996 0.437 1.866 0.631 ...
##   $ Attr9 : num  1.14 1.7 1.31 1.06 1.16 ...
##   $ Attr10: num  0.505 0.498 0.304 0.574 0.387 ...
##   $ Attr11: num  0.2498 0.2611 0.3126 0.0927 0.1873 ...
##   $ Attr12: num  0.66 0.517 0.642 0.302 0.331 ...

```

```

## $ Attr13: num 0.1666 0.1583 0.2444 0.0943 0.1218 ...
## $ Attr14: num 0.2498 0.2583 0.3091 0.0927 0.1873 ...
## $ Attr15: num 497 678 794 917 1133 ...
## $ Attr16: num 0.734 0.538 0.46 0.398 0.322 ...
## $ Attr17: num 2.63 2 1.44 3.25 1.63 ...
## $ Attr18: num 0.2498 0.2583 0.3091 0.0927 0.1873 ...
## $ Attr19: num 0.1494 0.152 0.2361 0.0714 0.1155 ...
## $ Attr20: num 43.4 88 73.1 79.8 57 ...
## $ Attr21: num 1.25 1.43 1.43 1.51 NA ...
## $ Attr22: num 0.214 0.248 0.303 0.116 0.198 ...
## $ Attr23: num 0.12 0.123 0.19 0.0628 0.1155 ...
## $ Attr24: num 0.477 NA NA 0.172 0.187 ...
## $ Attr25: num 0.505 0.395 0.289 0.574 0.387 ...
## $ Attr26: num 0.604 0.44 0.373 0.362 0.322 ...
## $ Attr27: num 1.458 88.444 86.011 0.941 1.414 ...
## $ Attr28: num 1.76 16.95 1.06 1.96 1.12 ...
## $ Attr29: num 5.94 3.69 4.37 4.65 4.14 ...
## $ Attr30: num 0.118 0.27 0.419 0.143 0.279 ...
## $ Attr31: num 0.1494 0.152 0.2382 0.0714 0.1155 ...
## $ Attr32: num 94.1 122.2 176.9 91.4 147 ...
## $ Attr33: num 3.88 2.99 2.06 3.99 2.48 ...
## $ Attr34: num 0.564 2.988 1.427 0.376 0.323 ...
## $ Attr35: num 0.214 0.206 0.316 0.116 0.198 ...
## $ Attr36: num 1.74 1.7 1.31 1.36 1.63 ...
## $ Attr37: num 593.3 NA 2.3 NA 11.2 ...
## $ Attr38: num 0.506 0.498 0.515 0.574 0.435 ...
## $ Attr39: num 0.128 0.121 0.241 0.089 0.122 ...
## $ Attr40: num 0.663 0.0864 0.322 0.4014 0.293 ...
## $ Attr41: num 0.0514 0.0644 0.074 0.0696 0.0967 ...
## $ Attr42: num 0.128 0.146 0.231 0.089 0.122 ...
## $ Attr43: num 114 199 166 181 142 ...
## $ Attr44: num 71 111.5 92.4 101 84.6 ...
## $ Attr45: num 1.01 0.51 0.948 0.287 0.739 ...
## $ Attr46: num 1.522 1.125 1.01 1.57 0.958 ...
## $ Attr47: num 49.4 100.1 96.4 84.3 65.9 ...
## $ Attr48: num 0.1853 0.2373 0.2918 0.0859 0.1881 ...
## $ Attr49: num 0.1109 0.1396 0.2229 0.0662 0.116 ...
## $ Attr50: num 2.04 1.94 1.08 2.49 1.3 ...
## $ Attr51: num 0.379 0.5 0.482 0.307 0.565 ...
## $ Attr52: num 0.258 0.335 0.485 0.25 0.403 ...
## $ Attr53: num 2.24 17.87 1.21 2.45 1.88 ...
## $ Attr54: num 2.25 17.87 2.05 2.45 2.12 ...
## $ Attr55: num 348690 2305 6333 20545 3187 ...
## $ Attr56: num 0.122 0.121 0.241 0.054 0.135 ...
## $ Attr57: num 0.397 0.42 0.818 0.142 0.484 ...
## $ Attr58: num 0.878 0.853 0.766 0.946 0.865 ...
## $ Attr59: num 0.00192 0 0.69484 0 0.12444 ...
## $ Attr60: num 8.42 4.15 4.99 4.57 6.4 ...
## $ Attr61: num 5.14 3.27 3.95 3.61 4.32 ...
## $ Attr62: num 82.7 107.3 134.3 86.4 127.2 ...
## $ Attr63: num 4.42 3.4 2.72 4.22 2.87 ...
## $ Attr64: num 7.43 60.99 5.21 5.55 7.9 ...
## $ target: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 ...

```

```

head(bank_data)

##      Attr1   Attr2   Attr3   Attr4   Attr5   Attr6   Attr7   Attr8   Attr9
## 1 0.200550 0.37951 0.39641 2.0472 32.3510 0.38825 0.249760 1.33050 1.1389
## 2 0.209120 0.49988 0.47225 1.9447 14.7860 0.00000 0.258340 0.99601 1.6996
## 3 0.248660 0.69592 0.26713 1.5548 -1.1523 0.00000 0.309060 0.43695 1.3090
## 4 0.081483 0.30734 0.45879 2.4928 51.9520 0.14988 0.092704 1.86610 1.0571
## 5 0.187320 0.61323 0.22960 1.4063 -7.3128 0.18732 0.187320 0.63070 1.1559
## 6 0.228220 0.49794 0.35969 1.7502 -47.7170 0.00000 0.281390 1.00830 1.9786
##      Attr10  Attr11  Attr12  Attr13  Attr14  Attr15  Attr16 Attr17
## 1 0.50494 0.249760 0.65980 0.166600 0.249760 497.42 0.73378 2.6349
## 2 0.49788 0.261140 0.51680 0.158350 0.258340 677.96 0.53838 2.0005
## 3 0.30408 0.312580 0.64184 0.244350 0.309060 794.16 0.45961 1.4369
## 4 0.57353 0.092704 0.30163 0.094257 0.092704 917.01 0.39803 3.2537
## 5 0.38677 0.187320 0.33147 0.121820 0.187320 1133.20 0.32211 1.6307
## 6 0.50206 0.286450 0.58691 0.148120 0.281390 620.14 0.58858 2.0083
##      Attr18  Attr19  Attr20 Attr21  Attr22  Attr23  Attr24 Attr25
## 1 0.249760 0.149420 43.370 1.2479 0.21402 0.119980 0.47706 0.50494
## 2 0.258340 0.152000 87.981 1.4293 0.24806 0.123040       NA 0.39542
## 3 0.309060 0.236100 73.133 1.4283 0.30260 0.189960       NA 0.28932
## 4 0.092704 0.071428 79.788 1.5069 0.11550 0.062782 0.17193 0.57353
## 5 0.187320 0.115530 57.045       NA 0.19832 0.115530 0.18732 0.38677
## 6 0.281390 0.142220 107.260 1.7278 0.28104 0.115350       NA 0.49970
##      Attr26  Attr27  Attr28 Attr29  Attr30  Attr31 Attr32 Attr33 Attr34
## 1 0.60411 1.45820 1.7615 5.9443 0.11788 0.149420 94.14 3.8772 0.56393
## 2 0.43992 88.44400 16.9460 3.6884 0.26969 0.152000 122.17 2.9876 2.98760
## 3 0.37282 86.01100 1.0627 4.3749 0.41929 0.238150 176.93 2.0630 1.42740
## 4 0.36152 0.94076 1.9618 4.6511 0.14343 0.071428 91.37 3.9948 0.37581
## 5 0.32211 1.41380 1.1184 4.1424 0.27884 0.115530 147.04 2.4823 0.32340
## 6 0.48181 55.51100 2.2359 4.3258 0.23382 0.144490 102.76 3.5519 3.42000
##      Attr35 Attr36  Attr37  Attr38  Attr39  Attr40  Attr41  Attr42
## 1 0.21402 1.7410 593.2700 0.50591 0.128040 0.662950 0.051402 0.128040
## 2 0.20616 1.6996       NA 0.49788 0.121300 0.086422 0.064371 0.145950
## 3 0.31565 1.3090 2.3019 0.51537 0.241140 0.322020 0.074020 0.231170
## 4 0.11550 1.3562       NA 0.57353 0.088995 0.401390 0.069622 0.088995
## 5 0.19832 1.6278 11.2470 0.43489 0.122310 0.293040 0.096680 0.122310
## 6 0.27566 1.9786 22.3120 0.51361 0.139320 0.078634 0.056701 0.142040
##      Attr43  Attr44  Attr45  Attr46  Attr47  Attr48  Attr49 Attr50 Attr51
## 1 114.42 71.050 1.00970 1.52250 49.394 0.185300 0.110850 2.0420 0.37854
## 2 199.49 111.510 0.51045 1.12520 100.130 0.237270 0.139610 1.9447 0.49988
## 3 165.51 92.381 0.94807 1.01010 96.372 0.291810 0.222930 1.0758 0.48152
## 4 180.77 100.980 0.28720 1.56960 84.344 0.085874 0.066165 2.4928 0.30734
## 5 141.62 84.574 0.73919 0.95787 65.936 0.188110 0.116010 1.2959 0.56511
## 6 147.84 40.578 0.39250 0.53744 124.630 0.269350 0.136130 1.6852 0.47944
##      Attr52  Attr53  Attr54  Attr55  Attr56  Attr57  Attr58 Attr59
## 1 0.25792 2.2437 2.2480 348690.0 0.121960 0.39718 0.87804 0.001924
## 2 0.33472 17.8660 17.8660 2304.6 0.121300 0.42002 0.85300 0.000000
## 3 0.48474 1.2098 2.0504 6332.7 0.241140 0.81774 0.76599 0.694840
## 4 0.25033 2.4524 2.4524 20545.0 0.054015 0.14207 0.94598 0.000000
## 5 0.40285 1.8839 2.1184 3186.6 0.134850 0.48431 0.86515 0.124440
## 6 0.28154 3.1209 3.1927 7616.8 0.139320 0.45457 0.85891 0.023002
##      Attr60 Attr61  Attr62 Attr63  Attr64 target
## 1 8.4160 5.1372 82.658 4.4158 7.4277       No
## 2 4.1486 3.2732 107.350 3.4000 60.9870       No

```

```

## 3 4.9909 3.9510 134.270 2.7185 5.2078      No
## 4 4.5746 3.6147 86.435 4.2228 5.5497      No
## 5 6.3985 4.3158 127.210 2.8692 7.8980      No
## 6 3.4028 8.9949 88.444 4.1269 12.2990      No

tail(bank_data)

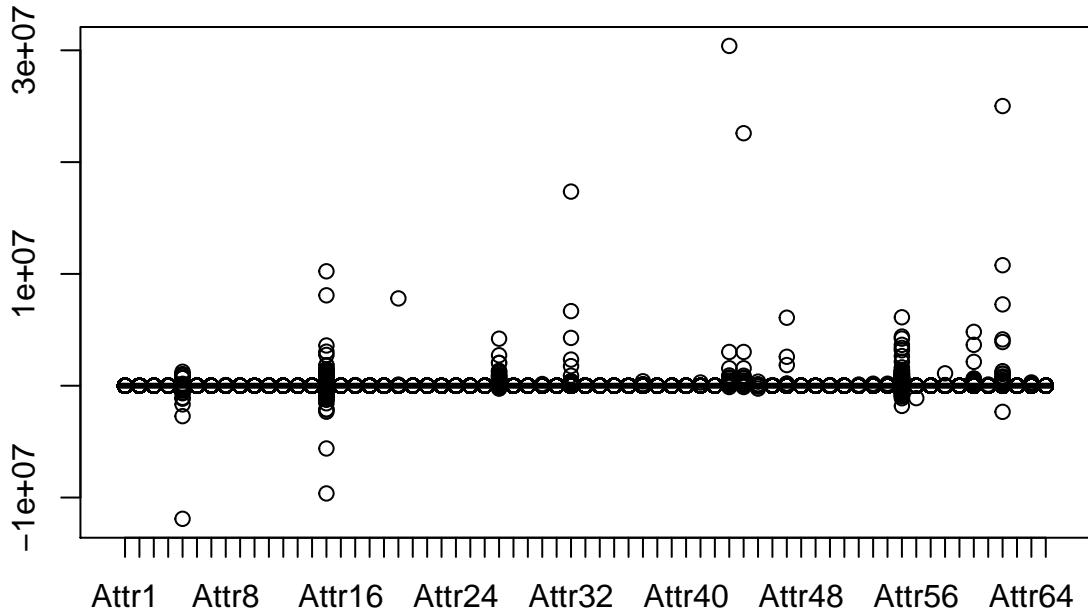
##          Attr1     Attr2     Attr3     Attr4     Attr5     Attr6     Attr7
## 42999 -0.283380 0.97794 -0.125140 0.64902 -56.126 -1.32980 -0.283380
## 43000  0.012898 0.70621  0.038857 1.17220 -18.907  0.00000  0.013981
## 43001 -0.578050 0.96702 -0.800850 0.16576 -67.365 -0.57805 -0.578050
## 43002 -0.179050 1.25530 -0.275990 0.74554 -120.440 -0.17905 -0.154930
## 43003 -0.108860 0.74394  0.015449 1.08780 -17.003 -0.10886 -0.109180
## 43004 -0.105370 0.53629 -0.045578 0.91478 -56.068 -0.10537 -0.109940
##          Attr8     Attr9     Attr10    Attr11    Attr12    Attr13    Attr14
## 42999  0.022514 1.39430  0.022018 -0.283380 -0.794780 -0.203240 -0.283380
## 43000  0.416000 1.67680  0.293790  0.041583  0.061959  0.023622  0.013981
## 43001 -0.403340 0.93979 -0.390040 -0.578050 -0.602160 -0.110220 -0.578050
## 43002 -0.260180 1.17490 -0.326590 -0.154930 -0.142840 -0.072940 -0.154930
## 43003  0.125310 0.84516  0.093224 -0.109180 -0.620380 -0.065652 -0.109180
## 43004  0.864600 0.95040  0.463670 -0.109940 -0.205570 -0.066747 -0.109940
##          Attr15    Attr16    Attr17    Attr18    Attr19    Attr20   Attr21
## 42999 -1259.60 -0.289770 1.02260 -0.283380 -0.203240 37.1960      NA
## 43000  6507.60  0.056089 1.41600  0.013981  0.008338 27.0920 1.03300
## 43001 -742.28 -0.491730 1.03410 -0.578050 -0.133990 3.2965 0.64770
## 43002 -3298.20 -0.110670 0.79665 -0.154930 -0.081350 37.0460 0.83104
## 43003 -5352.10 -0.068197 1.34420 -0.109180 -0.141280 26.2850 1.12100
## 43004 -2055.90 -0.177540 1.86470 -0.109940 -0.077072 47.1990 0.71351
##          Attr22    Attr23    Attr24    Attr25    Attr26    Attr27   Attr28
## 42999 -0.269220 -0.203240 -0.28338 -1.613300 -0.289770 -6.1342 -0.162830
## 43000  0.040401  0.007692 0.90184  0.083706  0.054556 1.4637 0.052831
## 43001 -0.534920 -0.133990 -0.57805 -0.390040 -0.491730      NA -0.952400
## 43002  0.176700 -0.094015 -0.15493 -0.326590 -0.129880      NA -1.442300
## 43003 -0.099950 -0.140860 -0.10918  0.093224 -0.067762      NA  0.019106
## 43004 -0.082947 -0.073868 -0.10994  0.463670 -0.169020      NA -0.089234
##          Attr29    Attr30    Attr31    Attr32 Attr33 Attr34 Attr35 Attr36
## 42999  4.1774  0.69780 -0.175650  96.479 4.8215 1.75790 -0.235400 1.47960
## 43000  4.9667  0.41853  0.024762 50.128 7.4000 2.36440  0.033819 1.67680
## 43001  4.0622  0.21864 -0.133990  76.330 4.7819 -0.55316 -0.534920 4.39120
## 43002  5.3244  0.65813 -0.081350 244.230 1.4945 0.14077  0.176700 1.95150
## 43003  4.8555  0.95181 -0.141280 70.252 5.1956 -0.13435 -0.099950 0.83553
## 43004  3.9698  0.37528 -0.077072 130.060 2.8064 -0.15467 -0.082947 1.44810
##          Attr37    Attr38    Attr39    Attr40 Attr41 Attr42 Attr43
## 42999  0.16297  0.57010 -0.168830 0.016216 -0.121080 -0.193090 59.065
## 43000  0.37850  0.66377  0.020169 0.022858 0.356510 0.024093 56.452
## 43001  17.03600 -0.38299 -0.123990 0.024771 -0.073529 -0.123990 11.451
## 43002  3.60650 -0.15597  0.092780 0.064095 0.214140 0.092780 141.650
## 43003  0.23908  0.66118 -0.129340 0.048976 -0.589300 -0.129340 86.348
## 43004 205.69000  0.46515 -0.058149 0.004456 -0.258460 -0.058149 124.570
##          Attr44    Attr45 Attr46 Attr47 Attr48 Attr49 Attr50
## 42999  21.8690 -1.99440 0.41473 30.168 -0.269220 -0.193090 0.23663
## 43000  29.3600  0.10363 0.62061 27.207 0.014771 0.008809 0.37454
## 43001  8.1548 -14.83600 0.12517 3.098 -0.637460 -0.147760 0.16455
## 43002 104.6100 -0.92630 0.56733 43.526 0.160680 0.084370 0.64421
## 43003  60.0640 -1.95610 0.77157 22.215 -0.158400 -0.204970 0.25733

```

```

## 43004 77.3740 -0.57124 0.56987 44.858 -0.097675 -0.068474 0.91225
## Attr51 Attr52 Attr53 Attr54 Attr55 Attr56 Attr57
## 42999 0.35655 0.20741 0.028649 0.74181 -1883.00 0.032555 -12.871000
## 43000 0.22564 0.13514 0.399440 0.90248 3599.10 0.020169 0.043904
## 43001 0.95997 0.20912 -0.463850 -0.45546 -9242.10 -0.064073 1.482000
## 43002 1.08460 0.66913 -1.706700 -0.81508 -58253.00 0.148880 0.548240
## 43003 0.17599 0.19247 0.115300 0.81772 1107.50 -0.183200 -1.167700
## 43004 0.53481 0.35632 0.907790 0.91069 -425.13 -0.052186 -0.227250
## Attr58 Attr59 Attr60 Attr61 Attr62 Attr63 Attr64 target
## 42999 0.97208 24.893000 9.8129 16.6910 93.338 3.9105 1.81430 Yes
## 43000 1.01220 1.259400 13.4720 12.4320 49.117 7.4313 2.27990 Yes
## 43001 1.06410 -0.018084 110.7200 44.7590 81.220 4.4940 5.13050 Yes
## 43002 0.85112 -0.522430 9.8526 3.4892 207.870 1.7559 9.95270 Yes
## 43003 1.18320 6.092400 13.8860 6.0769 83.122 4.3911 0.95575 Yes
## 43004 1.05220 0.003196 7.7332 4.7174 136.850 2.6672 2.79270 Yes
boxplot(bank_data)

```



```
names(bank_data)
```

```

## [1] "Attr1"  "Attr2"  "Attr3"  "Attr4"  "Attr5"  "Attr6"  "Attr7"
## [8] "Attr8"  "Attr9"  "Attr10" "Attr11" "Attr12" "Attr13" "Attr14"
## [15] "Attr15" "Attr16" "Attr17" "Attr18" "Attr19" "Attr20" "Attr21"
## [22] "Attr22" "Attr23" "Attr24" "Attr25" "Attr26" "Attr27" "Attr28"
## [29] "Attr29" "Attr30" "Attr31" "Attr32" "Attr33" "Attr34" "Attr35"
## [36] "Attr36" "Attr37" "Attr38" "Attr39" "Attr40" "Attr41" "Attr42"
## [43] "Attr43" "Attr44" "Attr45" "Attr46" "Attr47" "Attr48" "Attr49"

```

```

## [50] "Attr50" "Attr51" "Attr52" "Attr53" "Attr54" "Attr55" "Attr56"
## [57] "Attr57" "Attr58" "Attr59" "Attr60" "Attr61" "Attr62" "Attr63"
## [64] "Attr64" "target"

sum(is.na(bank_data))

## [1] 41037

summary(bank_data)

##      Attr1          Attr2          Attr3
## Min. :-463.8900  Min. :-430.8700  Min. :-479.9600
## 1st Qu.: 0.0034   1st Qu.: 0.2699   1st Qu.: 0.0215
## Median : 0.0497   Median : 0.4724   Median : 0.1968
## Mean   : 0.0408   Mean   : 0.5919   Mean   : 0.1137
## 3rd Qu.: 0.1297   3rd Qu.: 0.6891   3rd Qu.: 0.4036
## Max.  : 94.2800   Max.  : 480.9600  Max.  : 28.3360
## NA's   :8         NA's   :8       NA's   :8
##      Attr4          Attr5          Attr6
## Min.  :-0.40     Min. :-11903000  Min. :-508.4100
## 1st Qu.: 1.05     1st Qu.: -49    1st Qu.: 0.0000
## Median : 1.57     Median : -1     Median : 0.0000
## Mean   : 6.34     Mean   : -386   Mean   : -0.0576
## 3rd Qu.: 2.79     3rd Qu.: 51     3rd Qu.: 0.0877
## Max.  :53433.00   Max.  : 1250100  Max.  : 543.2500
## NA's   :133       NA's   :89     NA's   :8
##      Attr7          Attr8          Attr9
## Min. :-517.4800  Min. :-141.41   Min. :-3.496
## 1st Qu.: 0.0057   1st Qu.: 0.43   1st Qu.: 1.019
## Median : 0.0596   Median : 1.07   Median : 1.199
## Mean   : 0.0979   Mean   : 12.03  Mean   : 2.575
## 3rd Qu.: 0.1511   3rd Qu.: 2.60   3rd Qu.: 2.070
## Max.  : 649.2300  Max.  :53432.00  Max.  :9742.300
## NA's   :8         NA's   :93     NA's   :9
##      Attr10         Attr11         Attr12
## Min. :-479.9100  Min. :-463.8900  Min. :-6331.800
## 1st Qu.: 0.2946   1st Qu.: 0.0154   1st Qu.: 0.015
## Median : 0.5053   Median : 0.0754   Median : 0.172
## Mean   : 0.6049   Mean   : 0.1358  Mean   : 1.128
## 3rd Qu.: 0.7081   3rd Qu.: 0.1670   3rd Qu.: 0.586
## Max.  :1099.5000  Max.  : 681.5400  Max.  : 8259.400
## NA's   :8         NA's   :44     NA's   :133
##      Attr13         Attr14         Attr15
## Min. :-1460.600  Min. :-517.4800  Min. :-9632400
## 1st Qu.: 0.024   1st Qu.: 0.0058   1st Qu.: 222
## Median : 0.068   Median : 0.0596   Median : 846
## Mean   : 0.812   Mean   : 0.0980  Mean   : 1998
## 3rd Qu.: 0.135   3rd Qu.: 0.1511   3rd Qu.: 2229
## Max.  :13315.000 Max.  : 649.2300  Max.  :10236000
## NA's   :125       NA's   :8       NA's   :36
##      Attr16         Attr17         Attr18
## Min. :-6331.800  Min. :-0.41     Min. :-517.4800
## 1st Qu.: 0.073   1st Qu.: 1.45     1st Qu.: 0.0058
## Median : 0.245   Median : 2.11     Median : 0.0596
## Mean   : 1.407   Mean   : 13.19    Mean   : 0.1031

```

```

## 3rd Qu.: 0.663 3rd Qu.: 3.69 3rd Qu.: 0.1511
## Max. : 8259.400 Max. :53433.00 Max. : 649.2300
## NA's :94 NA's :93 NA's :8
## Attr19 Attr20 Attr21
## Min. :-1578.700 Min. : -29 Min. :-1325.000
## 1st Qu.: 0.004 1st Qu.: 15 1st Qu.: 0.908
## Median : 0.036 Median : 35 Median : 1.045
## Mean : 0.153 Mean : 245 Mean : 3.914
## 3rd Qu.: 0.091 3rd Qu.: 64 3rd Qu.: 1.204
## Max. : 9230.500 Max. :7809200 Max. :29907.000
## NA's :126 NA's :125 NA's :5836
## Attr22 Attr23 Attr24
## Min. :-431.5900 Min. :-1578.700 Min. :-463.8900
## 1st Qu.: 0.0000 1st Qu.: 0.002 1st Qu.: 0.0211
## Median : 0.0623 Median : 0.030 Median : 0.1550
## Mean : 0.1192 Mean : 0.136 Mean : 0.2747
## 3rd Qu.: 0.1501 3rd Qu.: 0.078 3rd Qu.: 0.3556
## Max. : 681.5400 Max. : 9230.500 Max. : 831.6600
## NA's :8 NA's :125 NA's :922
## Attr25 Attr26 Attr27
## Min. :-500.9300 Min. :-6331.800 Min. :-259010
## 1st Qu.: 0.1488 1st Qu.: 0.066 1st Qu.: 0
## Median : 0.3833 Median : 0.221 Median : 1
## Mean : 0.3605 Mean : 1.259 Mean : 1117
## 3rd Qu.: 0.6094 3rd Qu.: 0.598 3rd Qu.: 5
## Max. : 1353.3000 Max. : 8262.300 Max. : 4208800
## NA's :8 NA's :94 NA's :2750
## Attr28 Attr29 Attr30
## Min. :-3829.900 Min. :-0.8861 Min. : -6351.70
## 1st Qu.: 0.038 1st Qu.: 3.4906 1st Qu.: 0.08
## Median : 0.467 Median : 4.0077 Median : 0.22
## Mean : 6.068 Mean : 3.9993 Mean : 7.43
## 3rd Qu.: 1.502 3rd Qu.: 4.5137 3rd Qu.: 0.41
## Max. : 21701.000 Max. : 9.6983 Max. : 152860.00
## NA's :804 NA's :8 NA's :125
## Attr31 Attr32 Attr33
## Min. :-1495.600 Min. : -9296 Min. : -19.197
## 1st Qu.: 0.007 1st Qu.: 46 1st Qu.: 2.821
## Median : 0.043 Median : 78 Median : 4.629
## Mean : 0.174 Mean : 1066 Mean : 8.665
## 3rd Qu.: 0.101 3rd Qu.: 128 3rd Qu.: 7.816
## Max. : 9244.300 Max. :17364000 Max. :21944.000
## NA's :125 NA's :366 NA's :133
## Attr34 Attr35 Attr36
## Min. :-1696.000 Min. :-431.5900 Min. : -0.001
## 1st Qu.: 0.309 1st Qu.: 0.0060 1st Qu.: 1.105
## Median : 1.977 Median : 0.0605 Median : 1.648
## Mean : 5.443 Mean : 0.1120 Mean : 2.834
## 3rd Qu.: 4.569 3rd Qu.: 0.1500 3rd Qu.: 2.427
## Max. : 21944.000 Max. : 626.9200 Max. : 9742.300
## NA's :93 NA's :8 NA's :8
## Attr37 Attr38 Attr39
## Min. : -525.5 Min. : -479.9100 Min. : -7522.000
## 1st Qu.: 1.1 1st Qu.: 0.4194 1st Qu.: 0.004

```

```

## Median : 3.1 Median : 0.6117 Median : 0.037
## Mean : 102.5 Mean : 0.7026 Mean : -0.293
## 3rd Qu.: 11.4 3rd Qu.: 0.7713 3rd Qu.: 0.091
## Max. :398920.0 Max. :1099.5000 Max. : 2156.500
## NA's :18836 NA's :8 NA's :125
## Attr40 Attr41 Attr42
## Min. :-101.270 Min. :-1234.40 Min. :-1395.8000
## 1st Qu.: 0.053 1st Qu.: 0.03 1st Qu.: 0.0000
## Median : 0.177 Median : 0.09 Median : 0.0380
## Mean : 2.155 Mean : 7.79 Mean : -0.1454
## 3rd Qu.: 0.652 3rd Qu.: 0.21 3rd Qu.: 0.0921
## Max. :8007.100 Max. :288770.00 Max. : 2156.8000
## NA's :133 NA's :754 NA's :125
## Attr43 Attr44 Attr45
## Min. :-115870 Min. :-115870 Min. :-256230.0
## 1st Qu.: 66 1st Qu.: 35 1st Qu.: 0.0
## Median : 99 Median : 55 Median : 0.3
## Mean : 1082 Mean : 838 Mean : 14.7
## 3rd Qu.: 141 3rd Qu.: 80 3rd Qu.: 1.0
## Max. :30393000 Max. :22584000 Max. : 366030.0
## NA's :125 NA's :125 NA's :2133
## Attr46 Attr47 Attr48
## Min. :-101.26 Min. :-96 Min. :-542.5600
## 1st Qu.: 0.61 1st Qu.: 16 1st Qu.: -0.0383
## Median : 1.03 Median : 38 Median : 0.0185
## Mean : 5.46 Mean : 360 Mean : 0.0336
## 3rd Qu.: 1.91 3rd Qu.: 70 3rd Qu.: 0.1075
## Max. :53433.00 Max. :6084200 Max. : 623.8500
## NA's :134 NA's :296 NA's :9
## Attr49 Attr50 Attr51
## Min. :-9001.000 Min. :-0.05 Min. :-0.1866
## 1st Qu.: -0.027 1st Qu.: 0.77 1st Qu.: 0.1905
## Median : 0.011 Median : 1.22 Median : 0.3416
## Mean : -0.488 Mean : 5.86 Mean : 0.4850
## 3rd Qu.: 0.062 3rd Qu.: 2.20 3rd Qu.: 0.5352
## Max. : 178.890 Max. :53433.00 Max. :480.9600
## NA's :125 NA's :93 NA's :8
## Attr52 Attr53 Attr54
## Min. :-25.47 Min. :-3828.90 Min. :-3828.90
## 1st Qu.: 0.13 1st Qu.: 0.69 1st Qu.: 0.96
## Median : 0.21 Median : 1.21 Median : 1.38
## Mean : 6.52 Mean : 23.89 Mean : 24.78
## 3rd Qu.: 0.35 3rd Qu.: 2.23 3rd Qu.: 2.38
## Max. :88433.00 Max. :180440.00 Max. :180440.00
## NA's :300 NA's :804 NA's :804
## Attr55 Attr56 Attr57
## Min. :-1805200 Min. :-1108300.0 Min. :-1667.3000
## 1st Qu.: 29 1st Qu.: 0.0 1st Qu.: 0.0148
## Median : 1079 Median : 0.1 Median : 0.1200
## Mean : 7603 Mean : -26.5 Mean : -0.0100
## 3rd Qu.: 4929 3rd Qu.: 0.1 3rd Qu.: 0.2854
## Max. : 6123700 Max. : 293.1 Max. : 552.6400
## NA's :1 NA's :125 NA's :7
## Attr58 Attr59 Attr60

```

```

## Min. : -198.7 Min. : -327.970 Min. : -12
## 1st Qu.: 0.9 1st Qu.: 0.000 1st Qu.: 6
## Median : 1.0 Median : 0.006 Median : 10
## Mean : 30.3 Mean : 1.338 Mean : 452
## 3rd Qu.: 1.0 3rd Qu.: 0.236 3rd Qu.: 20
## Max. :1108300.0 Max. :23853.000 Max. :4818700
## NA's :82 NA's :7 NA's :2138
## Attr61 Attr62 Attr63
## Min. : -12.66 Min. : -2336500 Min. : -1.543
## 1st Qu.: 4.52 1st Qu.: 42 1st Qu.: 3.099
## Median : 6.65 Median : 71 Median : 5.091
## Mean : 17.11 Mean : 1509 Mean : 9.368
## 3rd Qu.: 10.43 3rd Qu.: 117 3rd Qu.: 8.612
## Max. :108000.00 Max. :25016000 Max. :23454.000
## NA's :101 NA's :125 NA's :133
## Attr64 target
## Min. : -10677.00 No :40921
## 1st Qu.: 2.18 Yes: 2083
## Median : 4.31
## Mean : 72.65
## 3rd Qu.: 9.83
## Max. :294770.00
## NA's :804



|         | No        | Yes   |
|---------|-----------|-------|
| Min.    | -10677.00 | 2083  |
| 1st Qu. | 2.18      | 4.31  |
| Median  | 72.65     | 9.83  |
| Mean    | 294770.00 | 804   |
| 3rd Qu. | 9.83      | 4.31  |
| Max.    | 294770.00 | 72.65 |
| NA's    | 804       | 101   |



## Check the missing values
sum(is.na(bank_data))

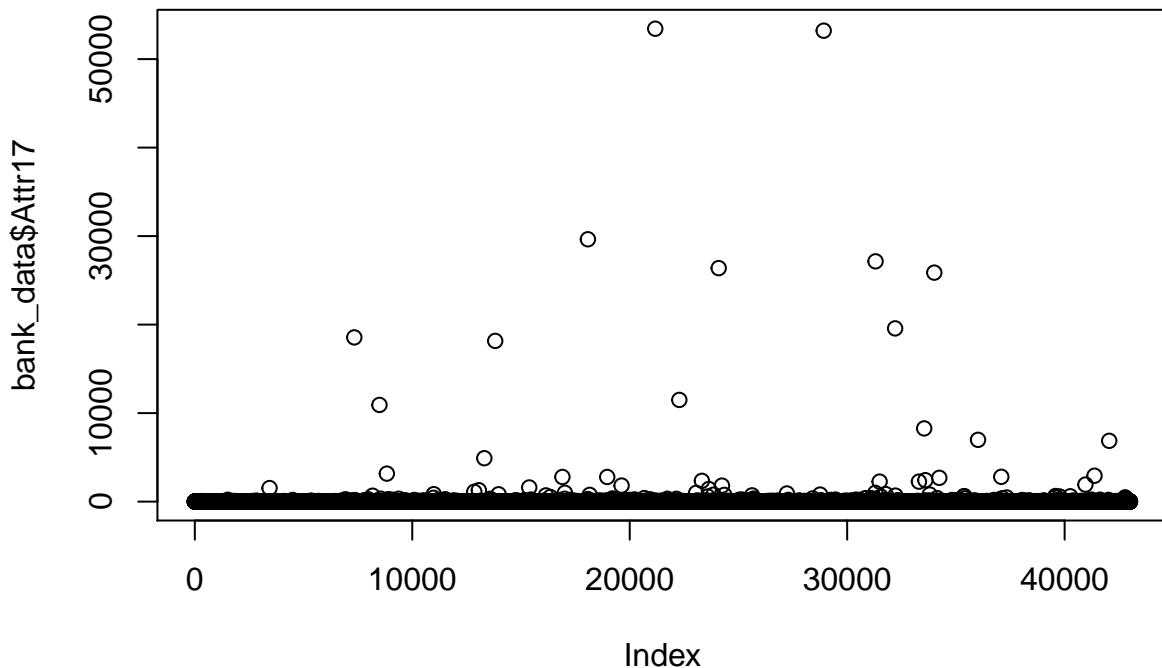
## [1] 41037

colSums(is.na(bank_data))

## Attr1 Attr2 Attr3 Attr4 Attr5 Attr6 Attr7 Attr8 Attr9 Attr10
## 8 8 8 133 89 8 8 93 9 8
## Attr11 Attr12 Attr13 Attr14 Attr15 Attr16 Attr17 Attr18 Attr19 Attr20
## 44 133 125 8 36 94 93 8 126 125
## Attr21 Attr22 Attr23 Attr24 Attr25 Attr26 Attr27 Attr28 Attr29 Attr30
## 5836 8 125 922 8 94 2750 804 8 125
## Attr31 Attr32 Attr33 Attr34 Attr35 Attr36 Attr37 Attr38 Attr39 Attr40
## 125 366 133 93 8 8 18836 8 125 133
## Attr41 Attr42 Attr43 Attr44 Attr45 Attr46 Attr47 Attr48 Attr49 Attr50
## 754 125 125 125 2133 134 296 9 125 93
## Attr51 Attr52 Attr53 Attr54 Attr55 Attr56 Attr57 Attr58 Attr59 Attr60
## 8 300 804 804 1 125 7 82 7 2138
## Attr61 Attr62 Attr63 Attr64 target
## 101 125 133 804 0

```

```
plot(bank_data$Attr17)
```



```
# Attr17 has many zero values  
#Attribute "Attr37" and "Attr21" has 18836 and 5836 NAs respectively
```

Assign data types

```
cat_attr <- c("target")  
num_attr <- setdiff(x = names(bank_data), y = cat_attr)  
data_cat <- bank_data[, cat_attr]  
data_num <- bank_data[, num_attr]  
data_num[, num_attr] <- data.frame(apply(data_num[, num_attr], 2, function(x) as.numeric(x)))
```

Imputation of missing values

```
data_num <- centralImputation(data = data_num)  
sum(is.na(data_num))  
## [1] 0
```

Normalising

```
data_num <- decostand(data_num, "standardize") #Using standardize method
summary(data_num)
```

```
##          Attr1          Attr2          Attr3
##  Min.   :-169.27809  Min.   :-73.51168  Min.   :-87.85874
##  1st Qu.: -0.01364   1st Qu.: -0.05485   1st Qu.: -0.01686
##  Median :  0.00324   Median : -0.02036   Median :  0.01522
##  Mean   :  0.00000   Mean   :  0.00000   Mean   :  0.00000
##  3rd Qu.:  0.03245   3rd Qu.:  0.01656   3rd Qu.:  0.05306
##  Max.   : 34.38581   Max.   : 81.84425   Max.   :  5.16499
##          Attr4          Attr5          Attr6
##  Min.   :-0.02272  Min.   :-193.65816  Min.   :-70.27166
##  1st Qu.: -0.01781 1st Qu.:  0.00547   1st Qu.:  0.00796
##  Median : -0.01606  Median :  0.00626   Median :  0.00796
##  Mean   :  0.00000  Mean   :  0.00000   Mean   :  0.00000
##  3rd Qu.: -0.01198 3rd Qu.:  0.00709   3rd Qu.:  0.02008
##  Max.   :180.28256  Max.   : 20.34567   Max.   : 75.10366
##          Attr7          Attr8          Attr9
##  Min.   :-91.35235  Min.   :-0.31490  Min.   :-0.10051
##  1st Qu.: -0.01627  1st Qu.: -0.02376  1st Qu.: -0.02576
##  Median : -0.00677  Median : -0.02245  Median : -0.02278
##  Mean   :  0.00000  Mean   :  0.00000  Mean   :  0.00000
##  3rd Qu.:  0.00938  3rd Qu.: -0.01931  3rd Qu.: -0.00837
##  Max.   :114.57161  Max.   :109.65055  Max.   :161.26387
##          Attr10         Attr11         Attr12
##  Min.   :-34.39450  Min.   :-88.39088  Min.   :-93.44273
##  1st Qu.: -0.02221  1st Qu.: -0.02291  1st Qu.: -0.01638
##  Median : -0.00713  Median : -0.01150  Median : -0.01407
##  Mean   :  0.00000  Mean   :  0.00000  Mean   :  0.00000
##  3rd Qu.:  0.00739  3rd Qu.:  0.00593  3rd Qu.: -0.00799
##  Max.   : 78.65719  Max.   :129.79866  Max.   :121.85138
##          Attr13         Attr14         Attr15
##  Min.   :-16.75563  Min.   :-91.35235  Min.   :-99.49457
##  1st Qu.: -0.00901  1st Qu.: -0.01627  1st Qu.: -0.01833
##  Median : -0.00850  Median : -0.00676  Median : -0.01189
##  Mean   :  0.00000  Mean   :  0.00000  Mean   :  0.00000
##  3rd Qu.: -0.00775  3rd Qu.:  0.00938  3rd Qu.:  0.00236
##  Max.   :152.65233  Max.   :114.57161  Max.   :105.68670
##          Attr16         Attr17         Attr18
##  Min.   :-92.13781  Min.   :-0.02779  Min.   :-90.95198
##  1st Qu.: -0.01937  1st Qu.: -0.02398  1st Qu.: -0.01710
##  Median : -0.01687  Median : -0.02262  Median : -0.00763
##  Mean   :  0.00000  Mean   :  0.00000  Mean   :  0.00000
##  3rd Qu.: -0.01081  3rd Qu.: -0.01942  3rd Qu.:  0.00844
##  Max.   :120.14037  Max.   :109.31673  Max.   :114.06745
##          Attr19         Attr20         Attr21
##  Min.   :-32.32462  Min.   :-0.00726  Min.   :-6.21731
##  1st Qu.: -0.00305  1st Qu.: -0.00607  1st Qu.: -0.01212
##  Median : -0.00240  Median : -0.00555  Median : -0.01160
##  Mean   :  0.00000  Mean   :  0.00000  Mean   :  0.00000
##  3rd Qu.: -0.00127  3rd Qu.: -0.00479  3rd Qu.: -0.01101
##  Max.   :188.97739  Max.   :207.32883  Max.   :139.94412
```

```

##      Attr22          Attr23          Attr24
## Min.   :-85.10776   Min.   :-32.56257   Min.   :-58.85545
## 1st Qu.: -0.02350   1st Qu.: -0.00274   1st Qu.: -0.03147
## Median : -0.01122   Median : -0.00217   Median : -0.01486
## Mean    : 0.00000   Mean    : 0.00000   Mean    : 0.00000
## 3rd Qu.: 0.00609   3rd Qu.: -0.00119   3rd Qu.: 0.00991
## Max.    :134.33626  Max.    :190.37098  Max.    :105.41942
##      Attr25          Attr26          Attr27
## Min.   :-44.81124   Min.   :-95.33249   Min.   :-7.64026
## 1st Qu.: -0.01892   1st Qu.: -0.01792   1st Qu.: -0.03073
## Median : 0.00204   Median : -0.01559   Median : -0.03070
## Mean    : 0.00000   Mean    : 0.00000   Mean    : 0.00000
## 3rd Qu.: 0.02225   3rd Qu.: -0.00995   3rd Qu.: -0.03060
## Max.    :120.94165  Max.    :124.35477  Max.    :123.62081
##      Attr28          Attr29          Attr30
## Min.   :-25.12801   Min.   :-5.91885   Min.   :-7.78273
## 1st Qu.: -0.03877   1st Qu.: -0.61611   1st Qu.: -0.00896
## Median : -0.03601   Median : 0.01017   Median : -0.00880
## Mean    : 0.00000   Mean    : 0.00000   Mean    : 0.00000
## 3rd Qu.: -0.02948   3rd Qu.: 0.62299   3rd Qu.: -0.00856
## Max.    :142.12008  Max.    : 6.90458   Max.    :187.07211
##      Attr31          Attr32          Attr33
## Min.   :-30.58800   Min.   :-0.11088   Min.   :-0.23332
## 1st Qu.: -0.00340   1st Qu.: -0.01083   1st Qu.: -0.04880
## Median : -0.00267   Median : -0.01049   Median : -0.03371
## Mean    : 0.00000   Mean    : 0.00000   Mean    : 0.00000
## 3rd Qu.: -0.00148   3rd Qu.: -0.00997   3rd Qu.: -0.00715
## Max.    :189.03888  Max.    :185.94272  Max.    :183.77703
##      Attr34          Attr35          Attr36
## Min.   :-14.01448   Min.   :-89.84196  Min.   :-0.04690
## 1st Qu.: -0.04222   1st Qu.: -0.02207   1st Qu.: -0.02861
## Median : -0.02849   Median : -0.01072   Median : -0.01962
## Mean    : 0.00000   Mean    : 0.00000   Mean    : 0.00000
## 3rd Qu.: -0.00720   3rd Qu.: 0.00791   3rd Qu.: -0.00674
## Max.    :180.70482  Max.    :130.44566  Max.    :161.11184
##      Attr37          Attr38          Attr39
## Min.   :-0.25714   Min.   :-34.19710  Min.   :-190.99868
## 1st Qu.: -0.02485   1st Qu.: -0.02015   1st Qu.:  0.00753
## Median : -0.02459   Median : -0.00647   Median :  0.00836
## Mean    : 0.00000   Mean    : 0.00000   Mean    : 0.00000
## 3rd Qu.: -0.02422   3rd Qu.: 0.00488   3rd Qu.:  0.00974
## Max.    :175.47654  Max.    : 78.18288  Max.    : 54.76742
##      Attr40          Attr41          Attr42
## Min.   :-1.8401    Min.   :-0.89159   Min.   :-87.00456
## 1st Qu.: -0.0373    1st Qu.: -0.00547   1st Qu.:  0.00903
## Median : -0.0351    Median : -0.00543   Median :  0.01140
## Mean    : 0.00000   Mean    : 0.00000   Mean    : 0.00000
## 3rd Qu.: -0.0267    3rd Qu.: -0.00535   3rd Qu.:  0.01476
## Max.    :142.4324    Max.    :207.28272  Max.    :134.46305
##      Attr43          Attr44          Attr45
## Min.   :-0.79187   Min.   :-1.05710  Min.   :-107.76024
## 1st Qu.: -0.00686   1st Qu.: -0.00725   1st Qu.: -0.00587
## Median : -0.00664   Median : -0.00707   Median : -0.00576
## Mean    : 0.00000   Mean    : 0.00000   Mean    : 0.00000

```

```

## 3rd Qu.: -0.00636 3rd Qu.: -0.00684 3rd Qu.: -0.00550
## Max.    :205.78601 Max.    :204.55583 Max.    : 153.92351
## Attr46          Attr47          Attr48
## Min.    :-0.36016 Min.    :-0.01368 Min.    :-108.28238
## 1st Qu.:-0.01632 1st Qu.:-0.01030 1st Qu.:-0.01433
## Median :-0.01491 Median :-0.00964 Median :-0.00300
## Mean    :0.00000 Mean   :0.00000 Mean   : 0.00000
## 3rd Qu.:-0.01194 3rd Qu.:-0.00868 3rd Qu.: 0.01475
## Max.    :180.33535 Max.    :183.32360 Max.    :124.49158
## Attr49          Attr50          Attr51
## Min.    :-198.70644 Min.    :-0.01913 Min.    :-0.12292
## 1st Qu.: 0.01016 1st Qu.:-0.01646 1st Qu.:-0.05389
## Median : 0.01099 Median :-0.01502 Median :-0.02624
## Mean   : 0.00000 Mean   :0.00000 Mean   : 0.00000
## 3rd Qu.: 0.01212 3rd Qu.:-0.01185 3rd Qu.: 0.00920
## Max.    : 3.96014 Max.    :173.19629 Max.    :87.94948
## Attr52          Attr53          Attr54
## Min.    :-0.04987 Min.    :-3.18959 Min.    :-3.17180
## 1st Qu.:-0.00992 1st Qu.:-0.01885 1st Qu.: -0.01925
## Median :-0.00978 Median :-0.01843 Median :-0.01890
## Mean   : 0.00000 Mean   :0.00000 Mean   : 0.00000
## 3rd Qu.:-0.00957 3rd Qu.:-0.01762 3rd Qu.: -0.01811
## Max.    :138.03093 Max.    :149.37702 Max.    :148.50901
## Attr55          Attr56          Attr57
## Min.    :-25.84053 Min.    :-207.35515 Min.    :-121.39698
## 1st Qu.:-0.10797 1st Qu.: 0.00494 1st Qu.: 0.00180
## Median :-0.09299 Median : 0.00495 Median : 0.00947
## Mean   : 0.00000 Mean   :0.00000 Mean   : 0.00000
## 3rd Qu.:-0.03812 3rd Qu.: 0.00496 3rd Qu.: 0.02151
## Max.    : 87.18169 Max.    : 0.05978 Max.    : 40.23897
## Attr58          Attr59          Attr60
## Min.    :-0.04276 Min.    :-2.68470 Min.    :-0.01397
## 1st Qu.:-0.00548 1st Qu.:-0.01091 1st Qu.: -0.01340
## Median :-0.00547 Median :-0.01086 Median :-0.01327
## Mean   : 0.00000 Mean   :0.00000 Mean   : 0.00000
## 3rd Qu.:-0.00546 3rd Qu.:-0.00898 3rd Qu.: -0.01297
## Max.    :206.99542 Max.    :194.45212 Max.    :152.09017
## Attr61          Attr62          Attr63
## Min.    :-0.05360 Min.    :-16.73518 Min.    :-0.08749
## 1st Qu.:-0.02265 1st Qu.:-0.01047 1st Qu.: -0.05019
## Median :-0.01882 Median :-0.01026 Median :-0.03423
## Mean   : 0.00000 Mean   :0.00000 Mean   : 0.00000
## 3rd Qu.:-0.01203 3rd Qu.:-0.00994 3rd Qu.: -0.00614
## Max.    :194.57602 Max.    :179.05094 Max.    :188.22139
## Attr64
## Min.    :-4.56286
## 1st Qu.:-0.02936
## Median :-0.02847
## Mean   : 0.00000
## 3rd Qu.:-0.02621
## Max.    :125.10428

```

Final dataframe

```
df_final <- cbind(data_num, data_cat)
```

Outlier removal

```
# Outliers removal more than 3 times standard deviation
for (i in (1:(length(df_final[, ]) - 1)))
{
  df_final <- df_final[!(df_final[, i] > 3),]
  df_final <- df_final[!(df_final[, i] < -3),]
}
# Attr 17 , Attr37, Attr21 has many null/zero values
df_final <- subset(x = df_final, select = -c(Attr17,Attr37, Attr14,Attr18,
                                              Attr21,Attr35,Attr25,Attr38,Attr51))
str(df_final)

## 'data.frame': 41877 obs. of  56 variables:
##   $ Attr1 : num  0.0614 0.0759 0.0149 0.0535 0.0684 ...
##   $ Attr2 : num  -0.01567 0.01773 -0.04847 0.00364 -0.016 ...
##   $ Attr3 : num  0.0656 0.0281 0.0632 0.0212 0.045 ...
##   $ Attr4 : num  -0.0148 -0.0161 -0.0129 -0.0166 -0.0155 ...
##   $ Attr5 : num  0.00651 0.00625 0.00712 0.00615 0.00549 ...
##   $ Attr6 : num  0.00796 0.00796 0.02868 0.03386 0.00796 ...
##   $ Attr7 : num  0.02831 0.037262 -0.000924 0.015775 0.032379 ...
##   $ Attr8 : num  -0.0226 -0.0237 -0.0208 -0.0233 -0.0226 ...
##   $ Attr9 : num  -0.01449 -0.02095 -0.02513 -0.02349 -0.00987 ...
##   $ Attr10: num  -0.00766 -0.02153 -0.00224 -0.01561 -0.00736 ...
##   $ Attr11: num  0.02389 0.03369 -0.0082 0.00983 0.02871 ...
##   $ Attr12: num  -0.00898 -0.00714 -0.01216 -0.01172 -0.00795 ...
##   $ Attr13: num  -0.00747 -0.00648 -0.00821 -0.00789 -0.00759 ...
##   $ Attr15: num  -0.01363 -0.01243 -0.01116 -0.00892 -0.01422 ...
##   $ Attr16: num  -0.0126 -0.0138 -0.0146 -0.0158 -0.0119 ...
##   $ Attr19: num  -1.64e-05 1.71e-03 -1.67e-03 -7.63e-04 -2.17e-04 ...
##   $ Attr20: num  -0.00415 -0.00454 -0.00436 -0.00497 -0.00363 ...
##   $ Attr22: num  0.025406 0.036158 -0.000727 0.0156 0.031907 ...
##   $ Attr23: num  -0.000252 0.001128 -0.001495 -0.000407 -0.000411 ...
##   $ Attr24: num  -0.0149 -0.0149 -0.0127 -0.0108 -0.0149 ...
##   $ Attr26: num  -0.0123 -0.0133 -0.0135 -0.0141 -0.0117 ...
##   $ Attr27: num  -0.0281 -0.0282 -0.0307 -0.0307 -0.0291 ...
##   $ Attr28: num  0.0719 -0.0321 -0.0262 -0.0317 -0.0244 ...
##   $ Attr29: num  -0.377 0.455 0.79 0.173 0.396 ...
##   $ Attr30: num  -0.00873 -0.00855 -0.00889 -0.00872 -0.00878 ...
##   $ Attr31: num  -0.000439 0.001323 -0.002087 -0.001185 -0.000593 ...
##   $ Attr32: num  -0.01002 -0.00944 -0.01035 -0.00976 -0.01023 ...
##   $ Attr33: num  -0.0475 -0.0552 -0.039 -0.0517 -0.0427 ...
##   $ Attr34: num  -0.0202 -0.033 -0.0417 -0.0421 -0.0166 ...
##   $ Attr36: num  -0.0188 -0.0252 -0.0244 -0.02 -0.0142 ...
##   $ Attr39: num  0.01051 0.01355 0.00969 0.01053 0.01096 ...
##   $ Attr40: num  -0.0367 -0.0325 -0.0311 -0.033 -0.0368 ...
##   $ Attr41: num  -0.00545 -0.00544 -0.00544 -0.00543 -0.00545 ...
##   $ Attr42: num  0.0181 0.0234 0.0146 0.0167 0.0179 ...
```

```

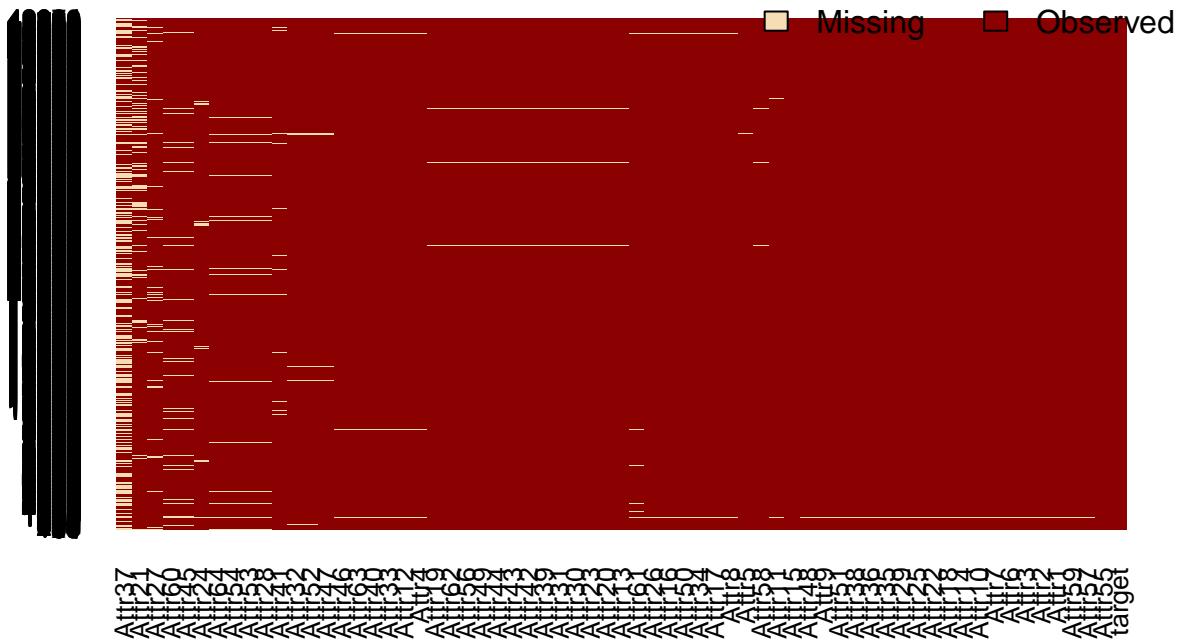
## $ Attr43: num -0.00596 -0.00619 -0.00609 -0.00635 -0.00631 ...
## $ Attr44: num -0.00656 -0.00673 -0.00665 -0.0068 -0.0072 ...
## $ Attr45: num -0.00566 -0.00548 -0.00576 -0.00557 -0.00571 ...
## $ Attr46: num -0.0146 -0.015 -0.0131 -0.0151 -0.0166 ...
## $ Attr47: num -0.00777 -0.00788 -0.00825 -0.0088 -0.00703 ...
## $ Attr48: num 0.0407 0.0515 0.0104 0.0308 0.0471 ...
## $ Attr49: num 0.0138 0.0157 0.0122 0.0133 0.0138 ...
## $ Attr50: num -0.0127 -0.0155 -0.0109 -0.0148 -0.0135 ...
## $ Attr52: num -0.00959 -0.00936 -0.00972 -0.00949 -0.00968 ...
## $ Attr53: num -0.00464 -0.01843 -0.0174 -0.01787 -0.01685 ...
## $ Attr54: num -0.00533 -0.01835 -0.01802 -0.0183 -0.01741 ...
## $ Attr55: num -0.075522 -0.018103 0.184485 -0.062949 0.000201 ...
## $ Attr56: num 0.00496 0.00498 0.00495 0.00496 0.00496 ...
## $ Attr57: num 0.0313 0.0603 0.0111 0.036 0.0338 ...
## $ Attr58: num -0.00549 -0.00551 -0.00547 -0.00549 -0.00549 ...
## $ Attr59: num -0.01091 -0.00524 -0.01091 -0.00989 -0.01072 ...
## $ Attr60: num -0.0134 -0.0134 -0.0134 -0.0134 -0.0135 ...
## $ Attr61: num -0.0249 -0.0237 -0.0243 -0.023 -0.0146 ...
## $ Attr62: num -0.01001 -0.00981 -0.01016 -0.00986 -0.01014 ...
## $ Attr63: num -0.0478 -0.0533 -0.0412 -0.0521 -0.042 ...
## $ Attr64: num -0.00441 -0.02809 -0.02794 -0.02695 -0.02508 ...
## $ target: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...

```

Plots

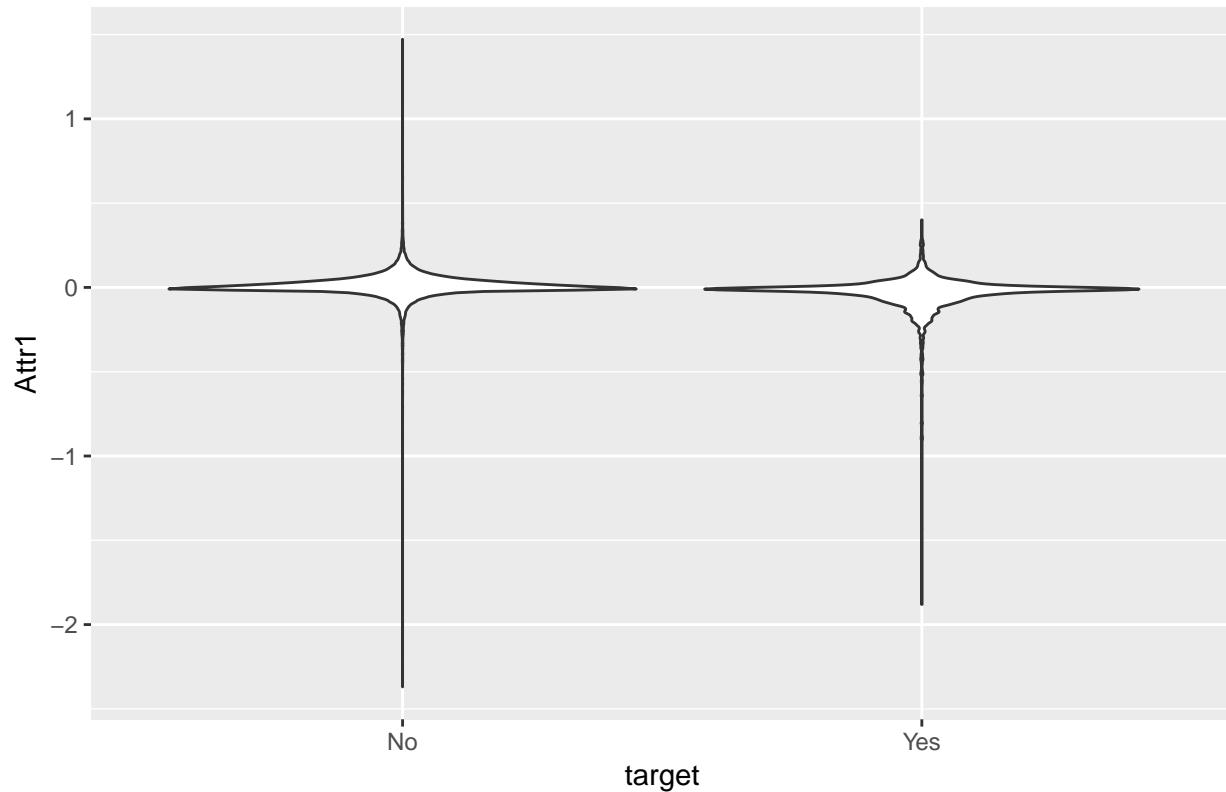
```
missmap(bank_data)
```

Missingness Map

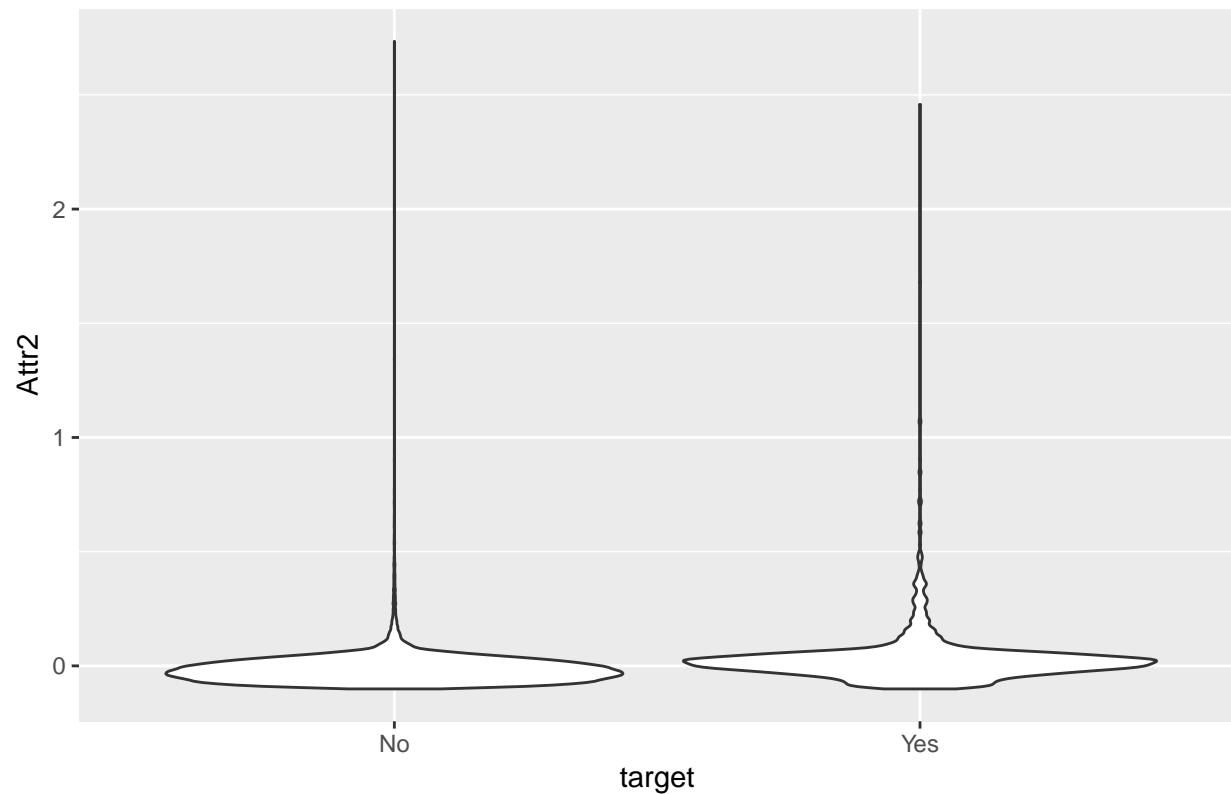


```
violin_plot <- function(x)
{
  if (is.numeric(df_final[, x]))
  {
    capture.output(
      ggplot(df_final, aes_string('target', x)) +
        geom_violin() +
        ggtitle(paste('Bankruptcy Status by', x))
    )
  }
}
cat_attr <- c("target")
num_attr <- setdiff(x = names(df_final), y = cat_attr)
for (i in 1:length(num_attr)) violin_plot(num_attr[i])
```

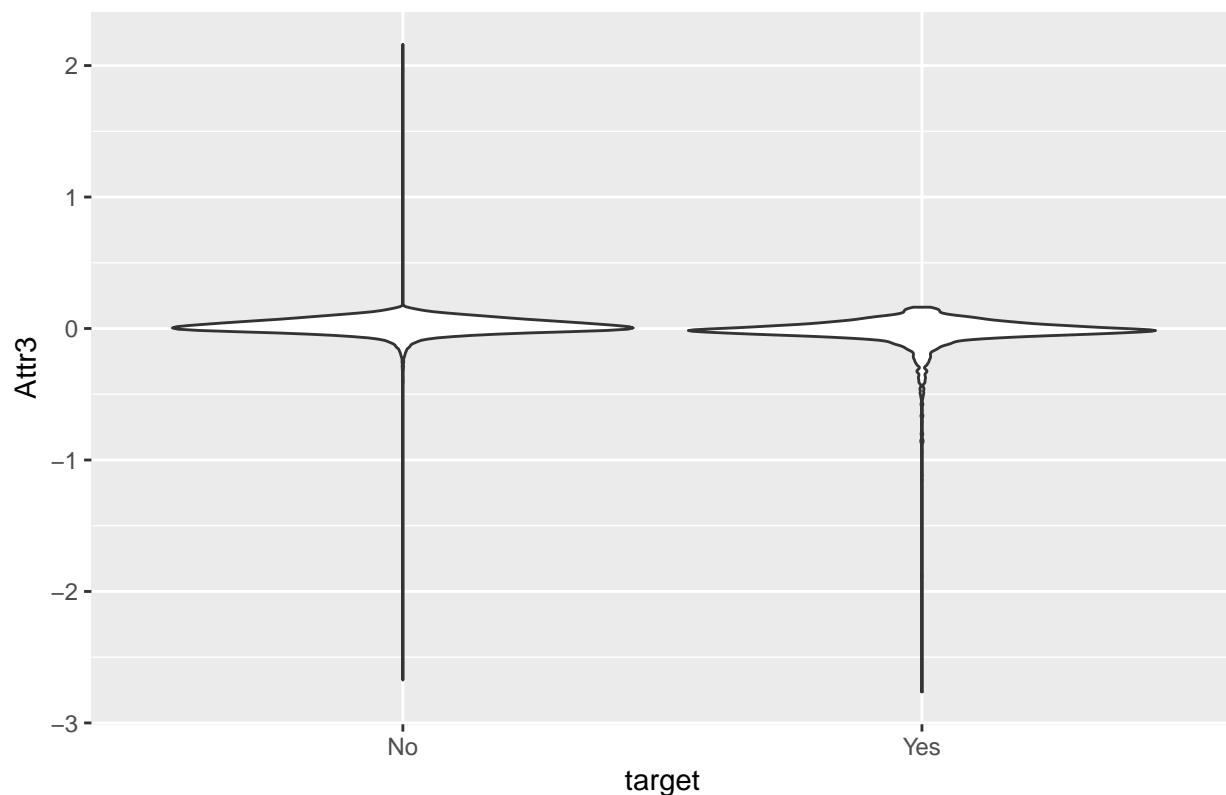
Bankruptcy Status by Attr1



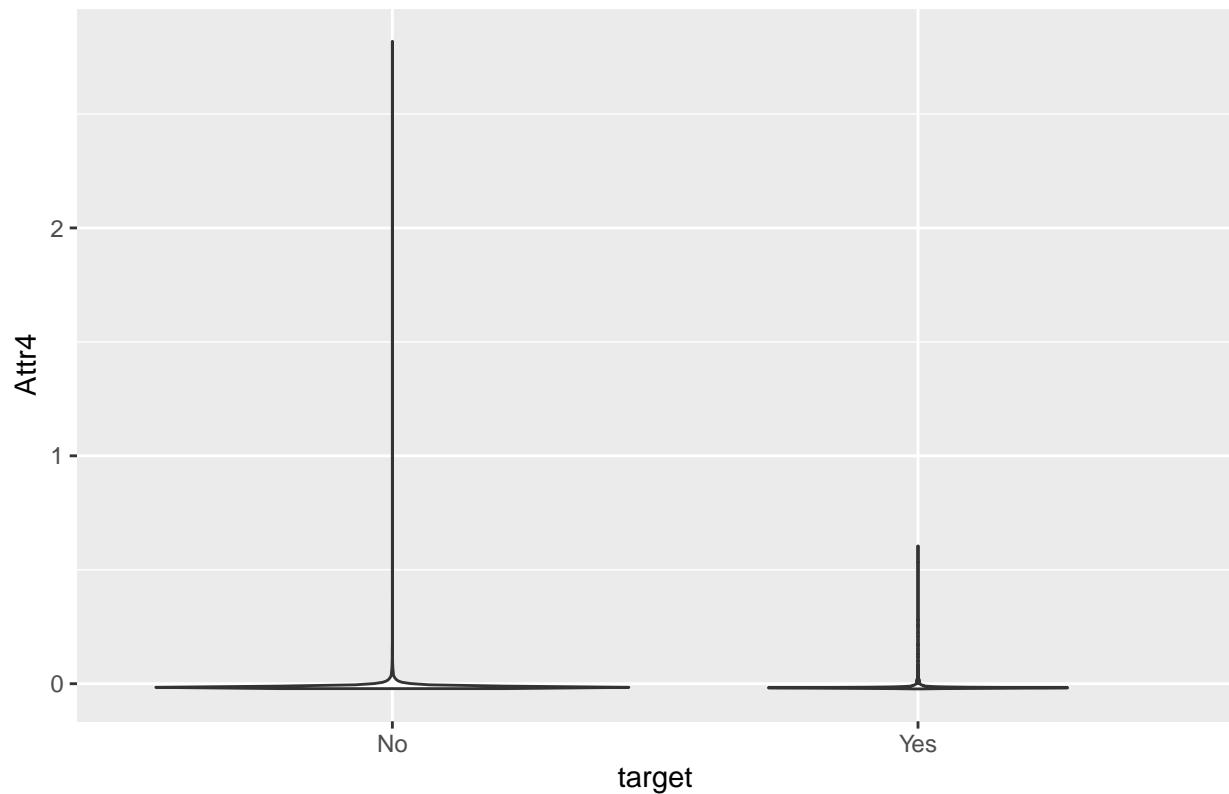
Bankruptcy Status by Attr2



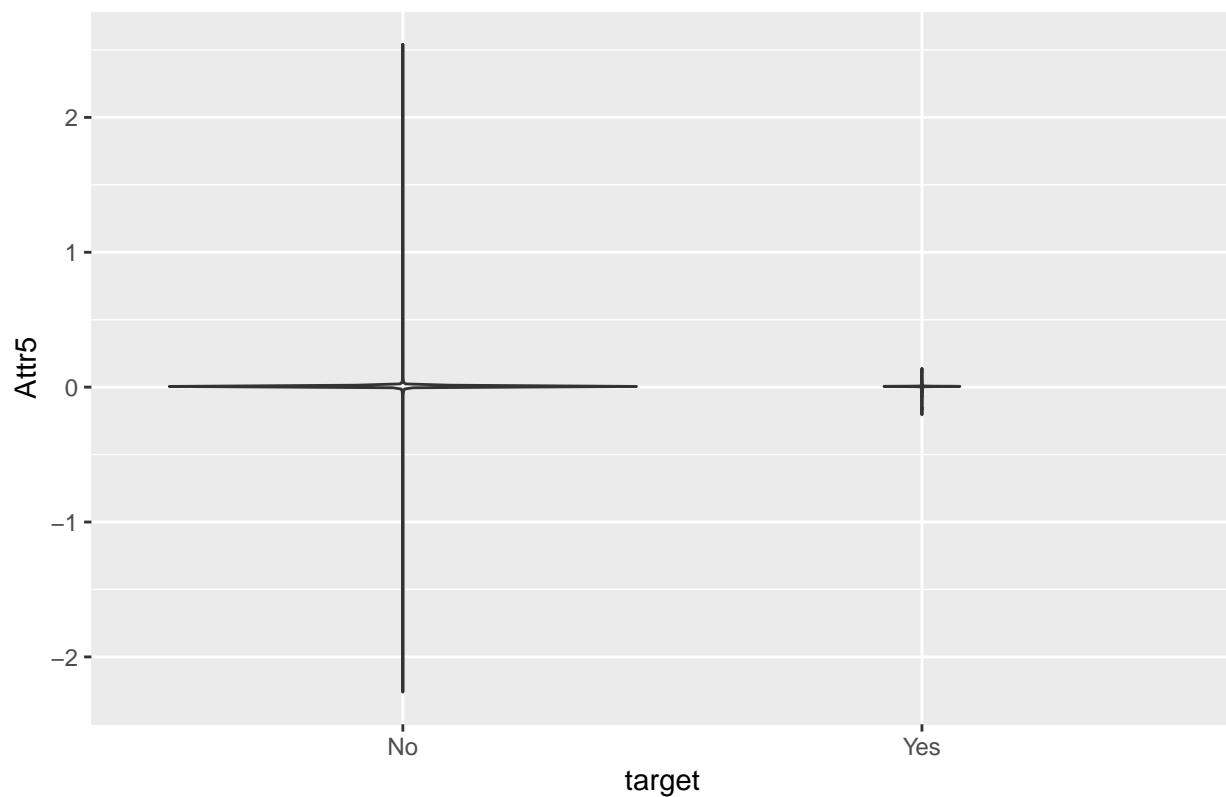
Bankruptcy Status by Attr3



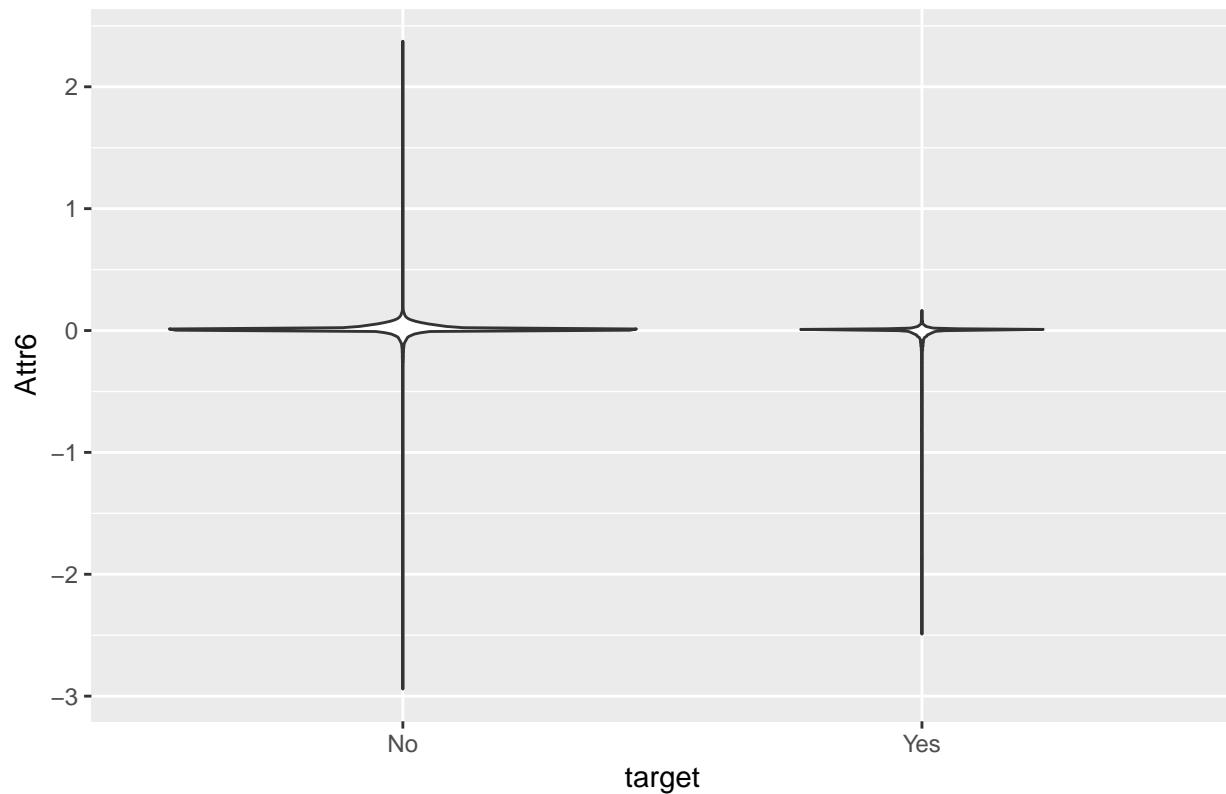
Bankruptcy Status by Attr4



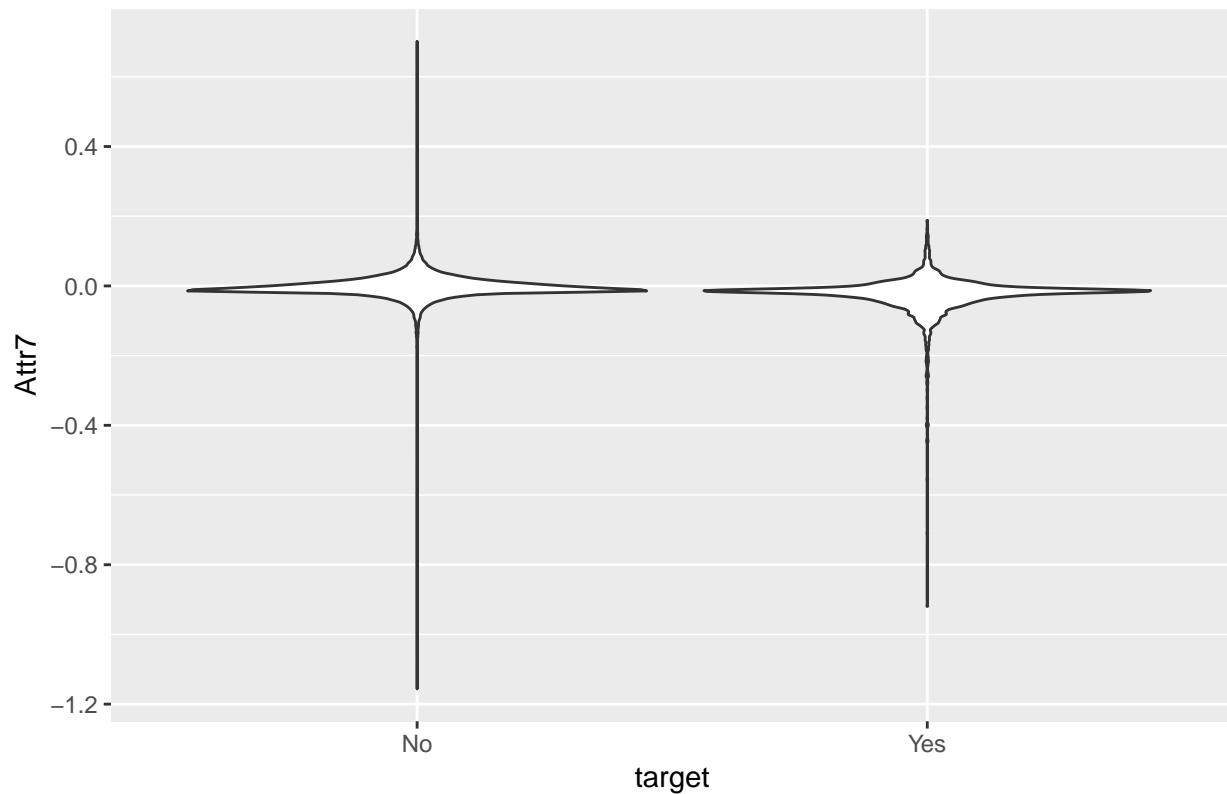
Bankruptcy Status by Attr5



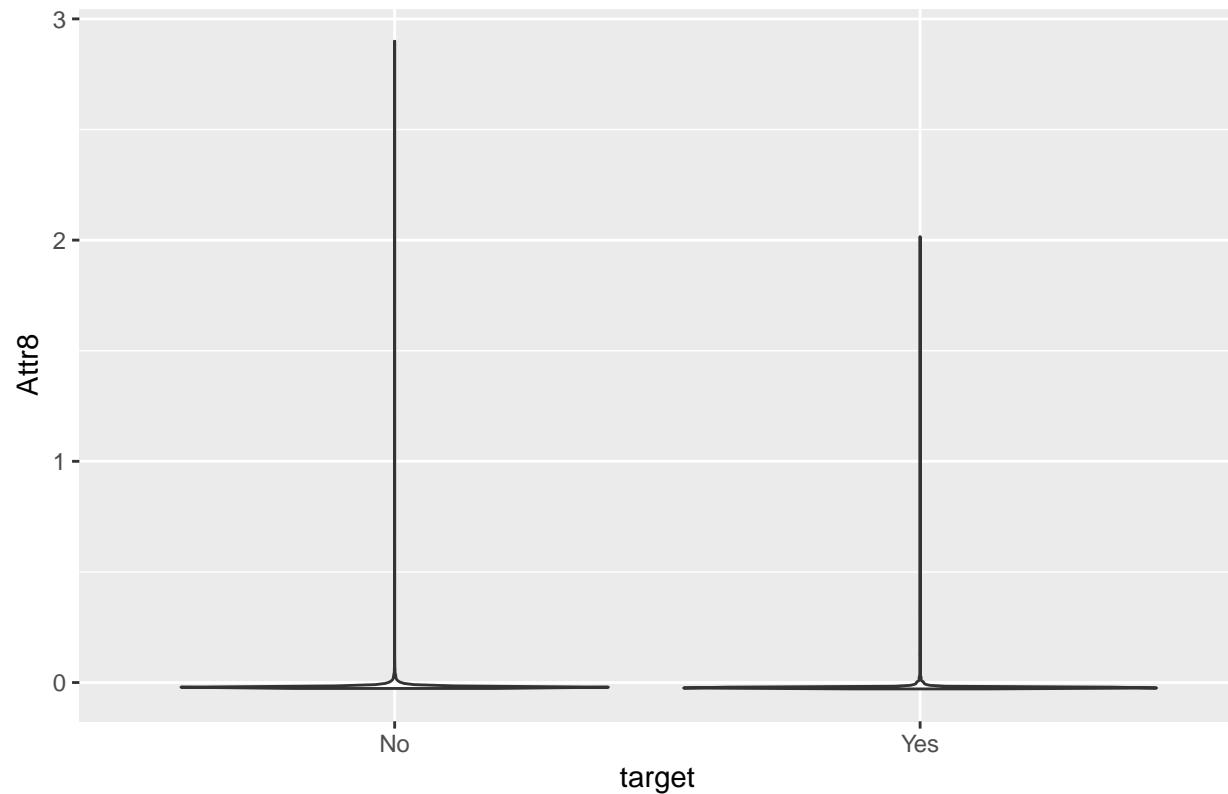
Bankruptcy Status by Attr6



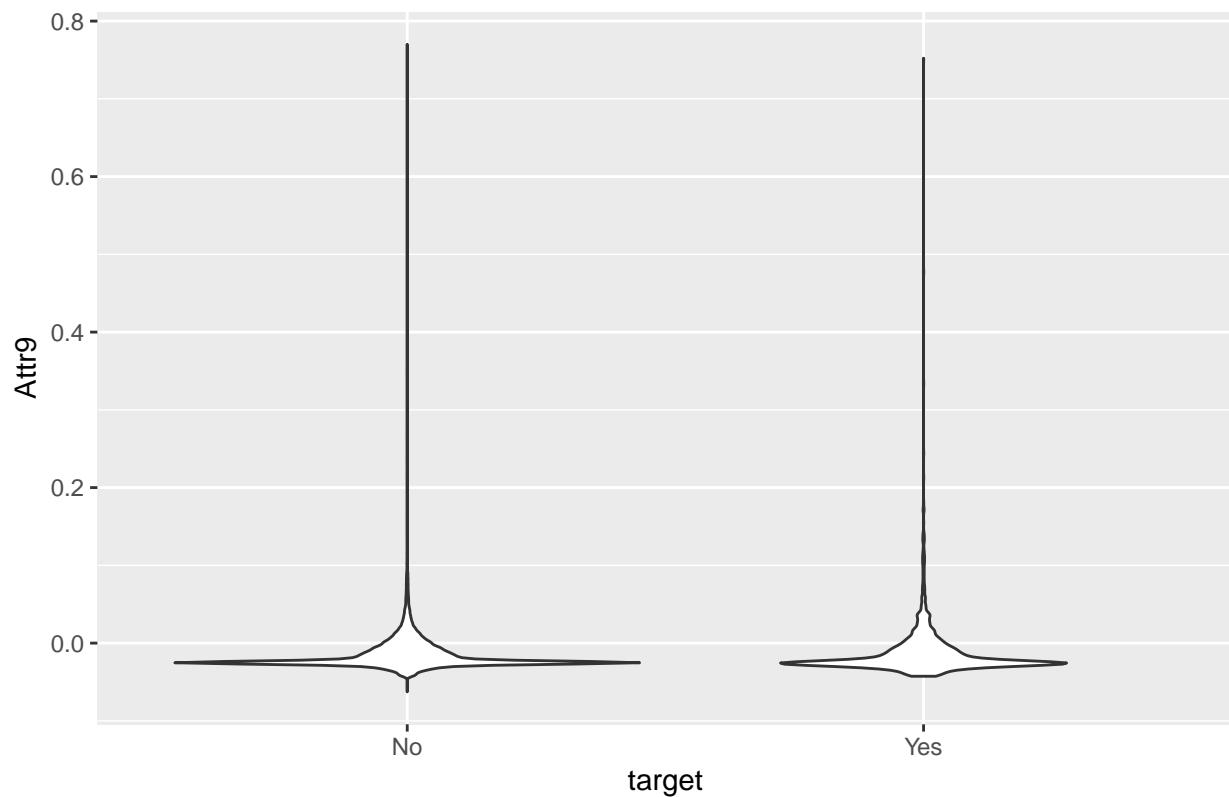
Bankruptcy Status by Attr7



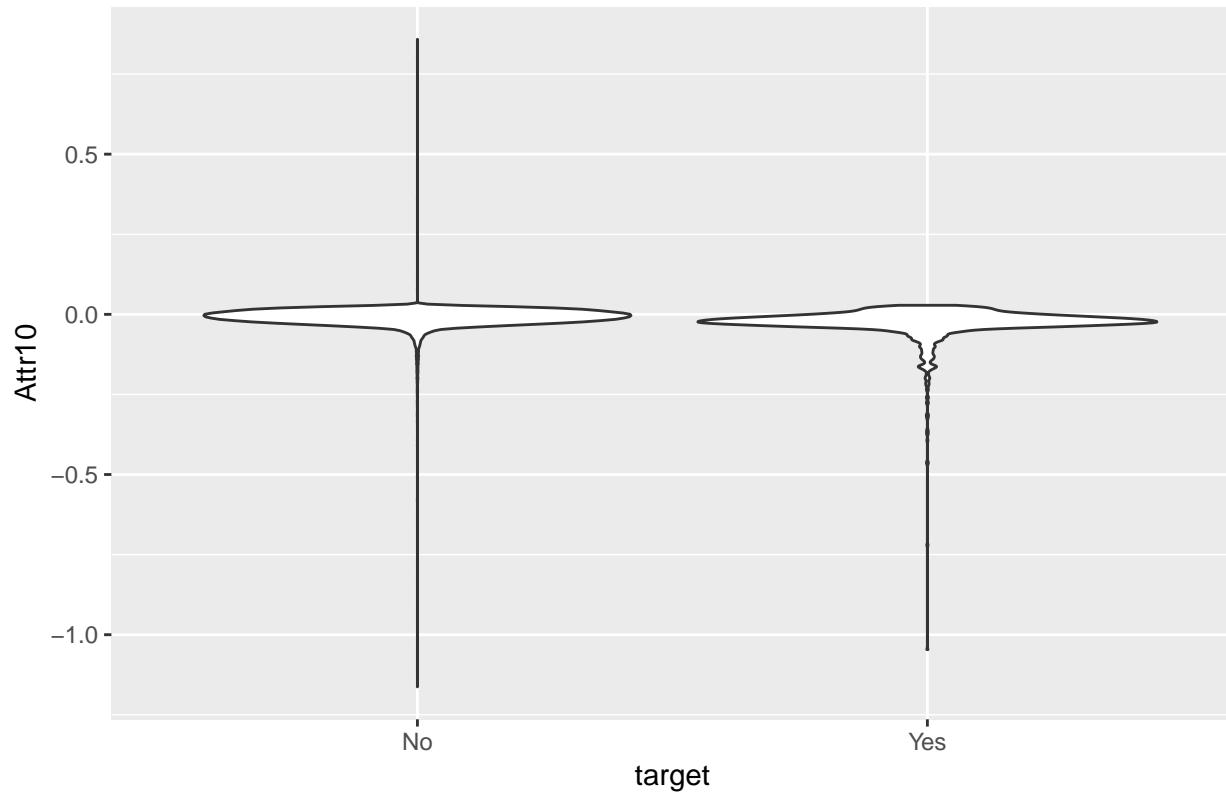
Bankruptcy Status by Attr8



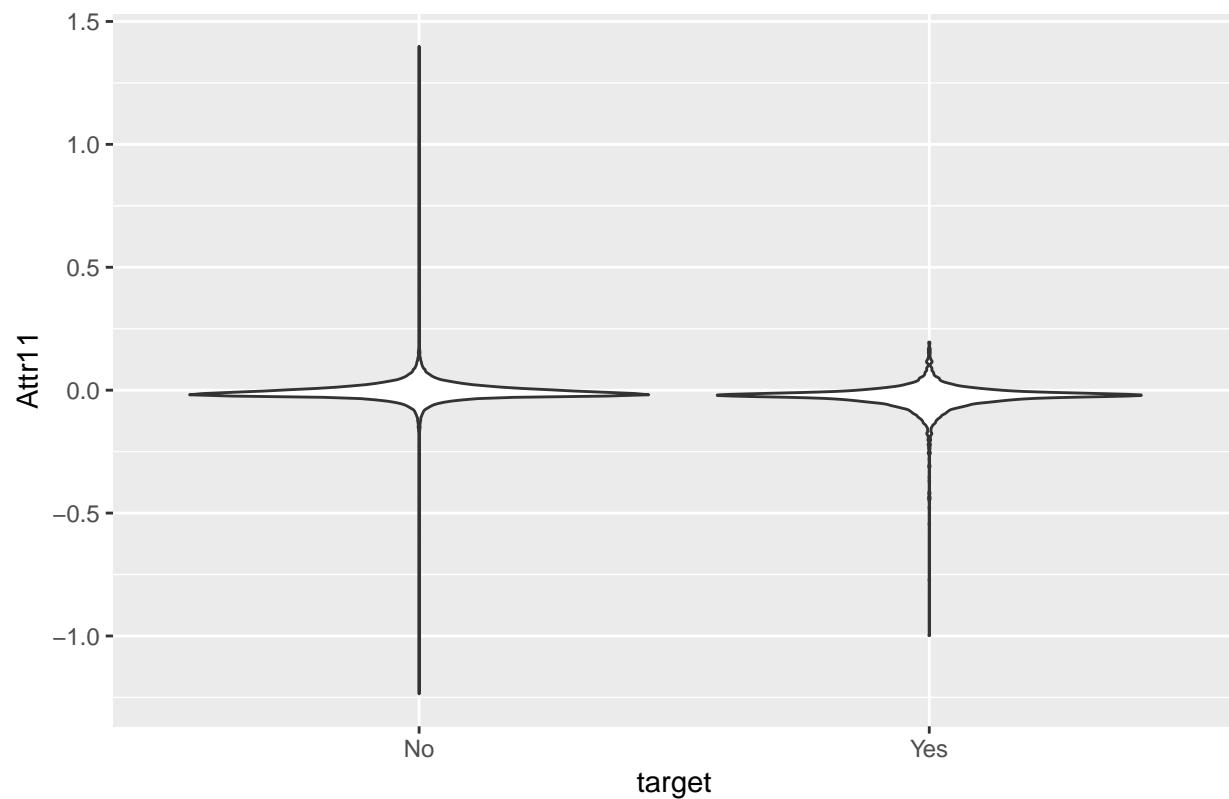
Bankruptcy Status by Attr9



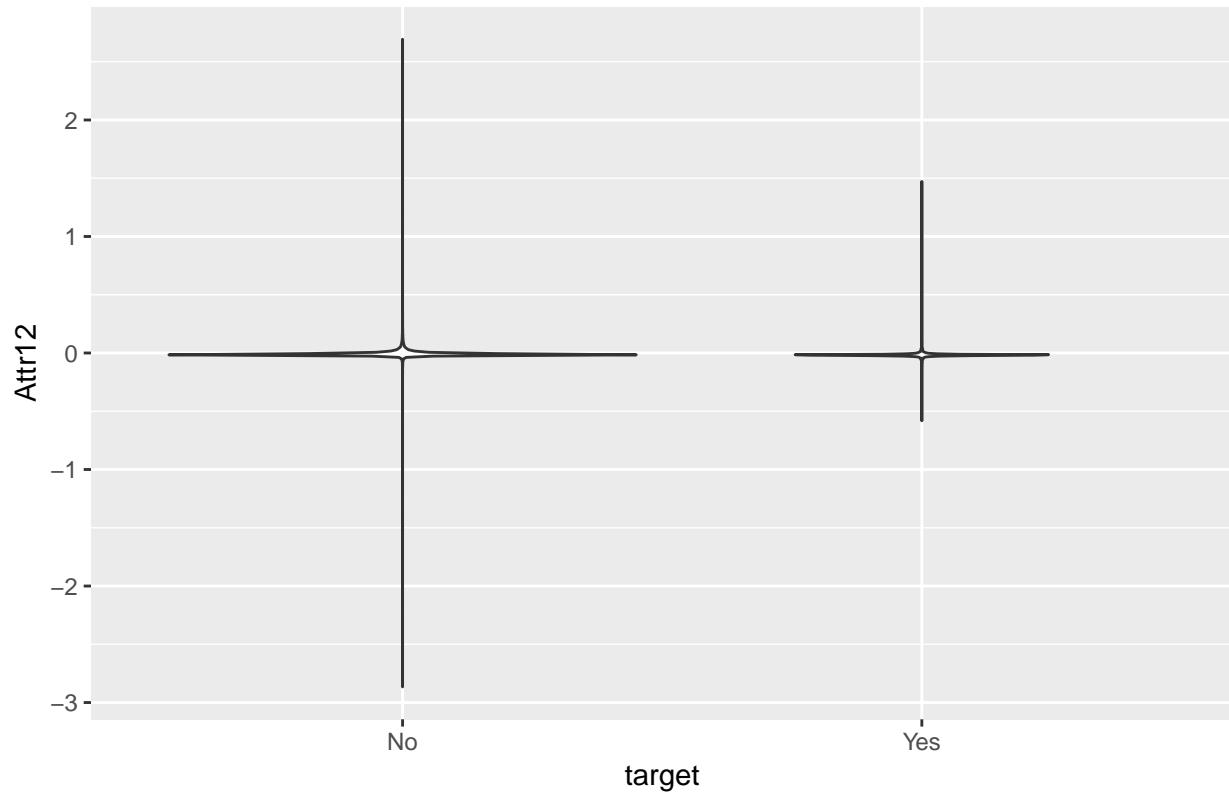
Bankruptcy Status by Attr10



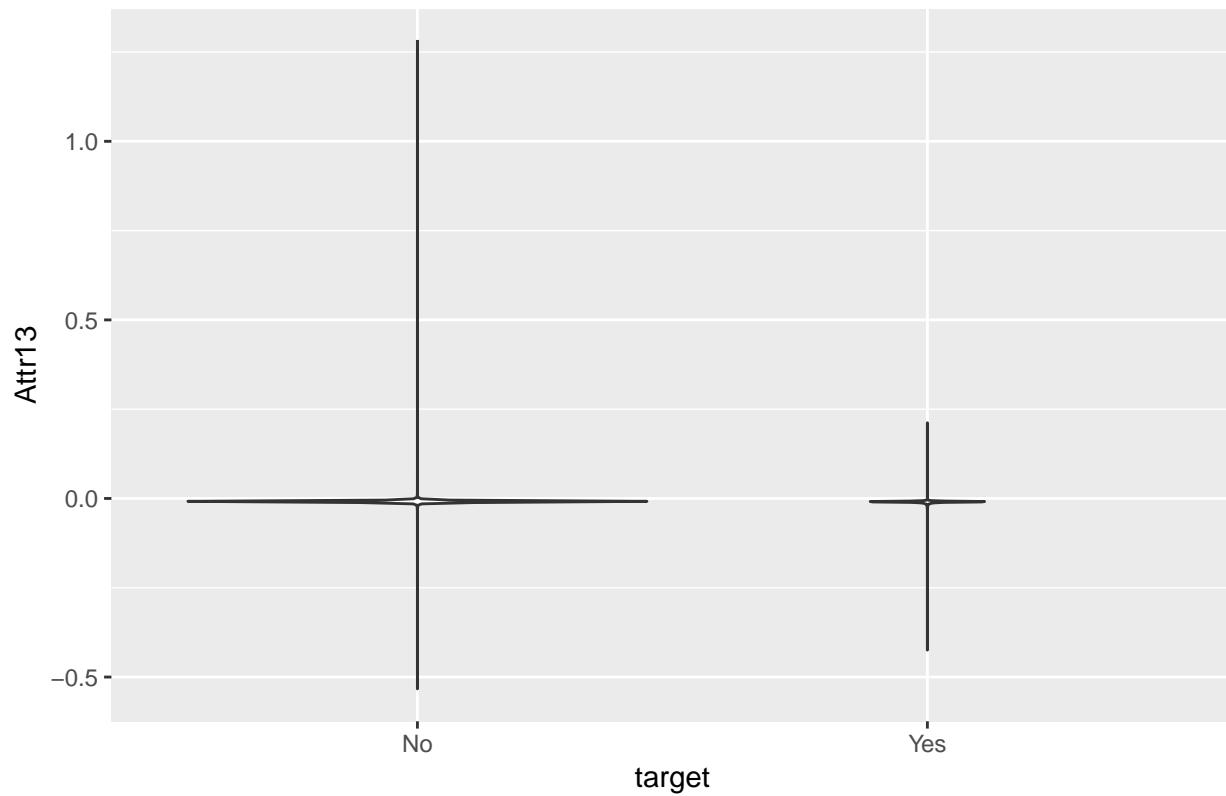
Bankruptcy Status by Attr11



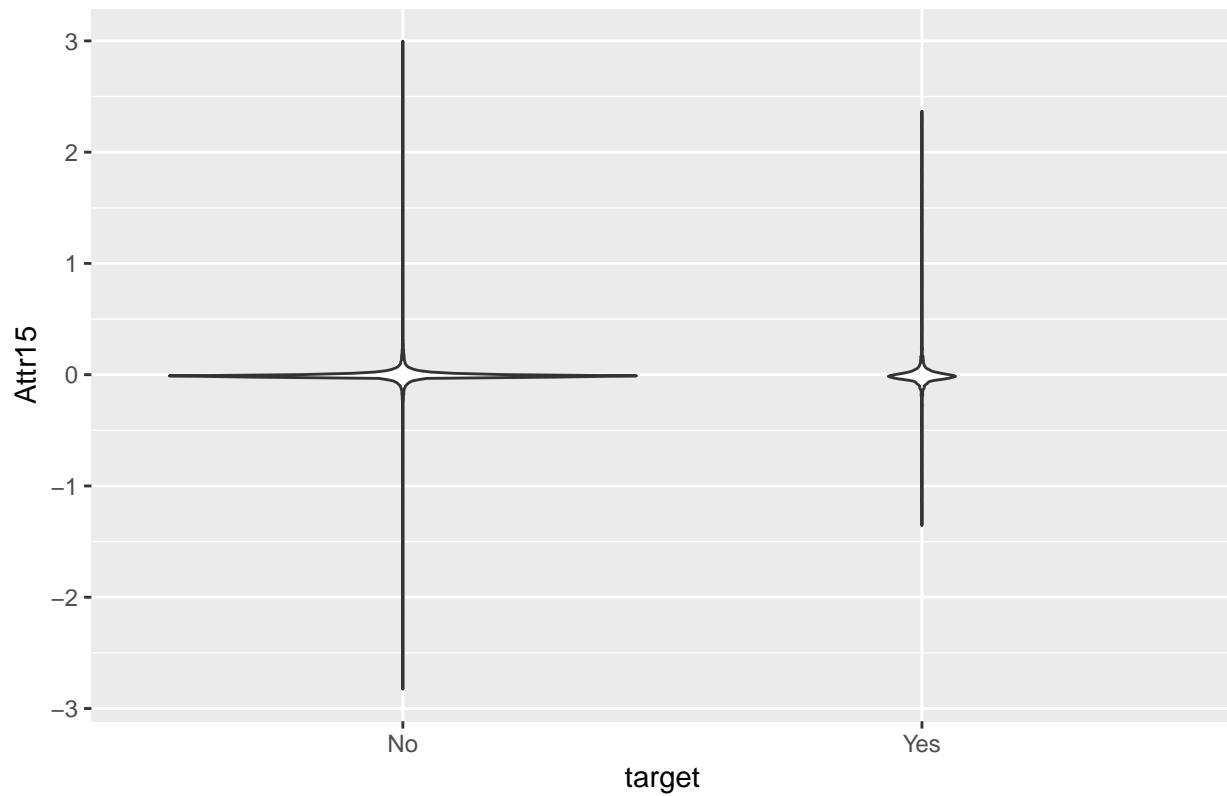
Bankruptcy Status by Attr12



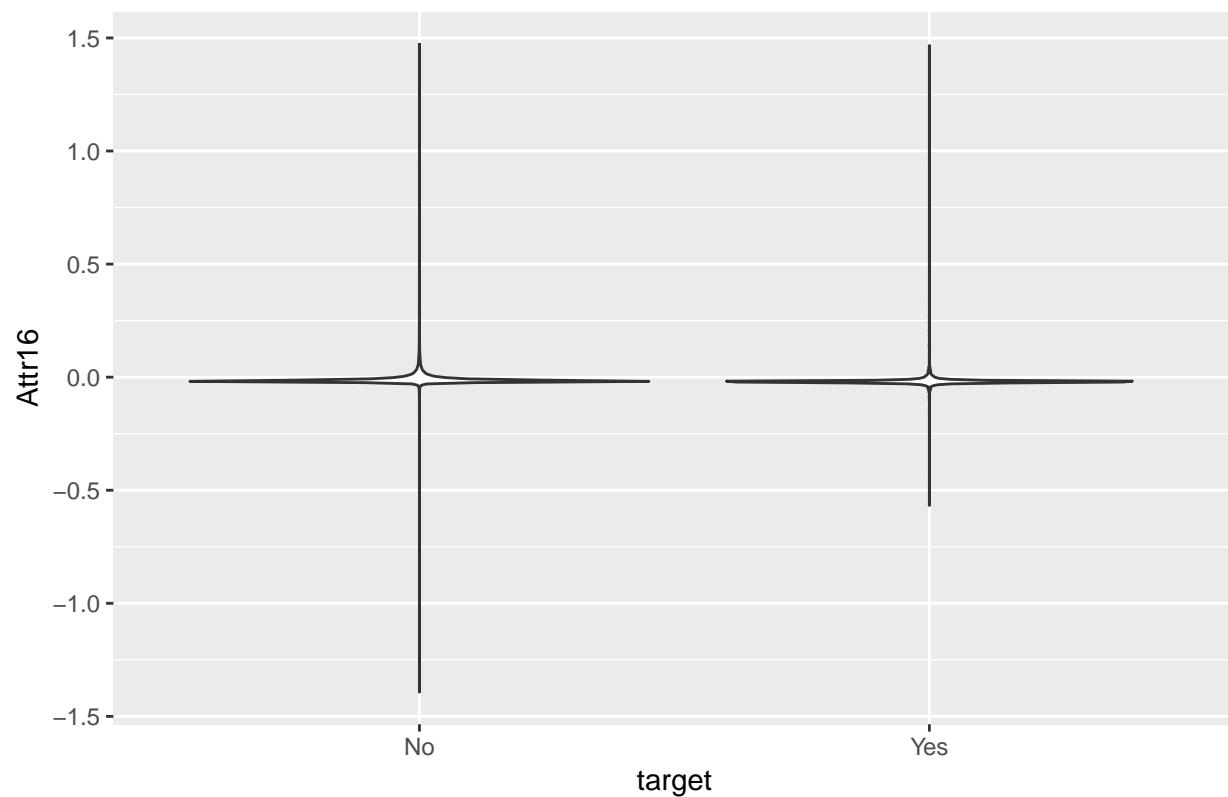
Bankruptcy Status by Attr13



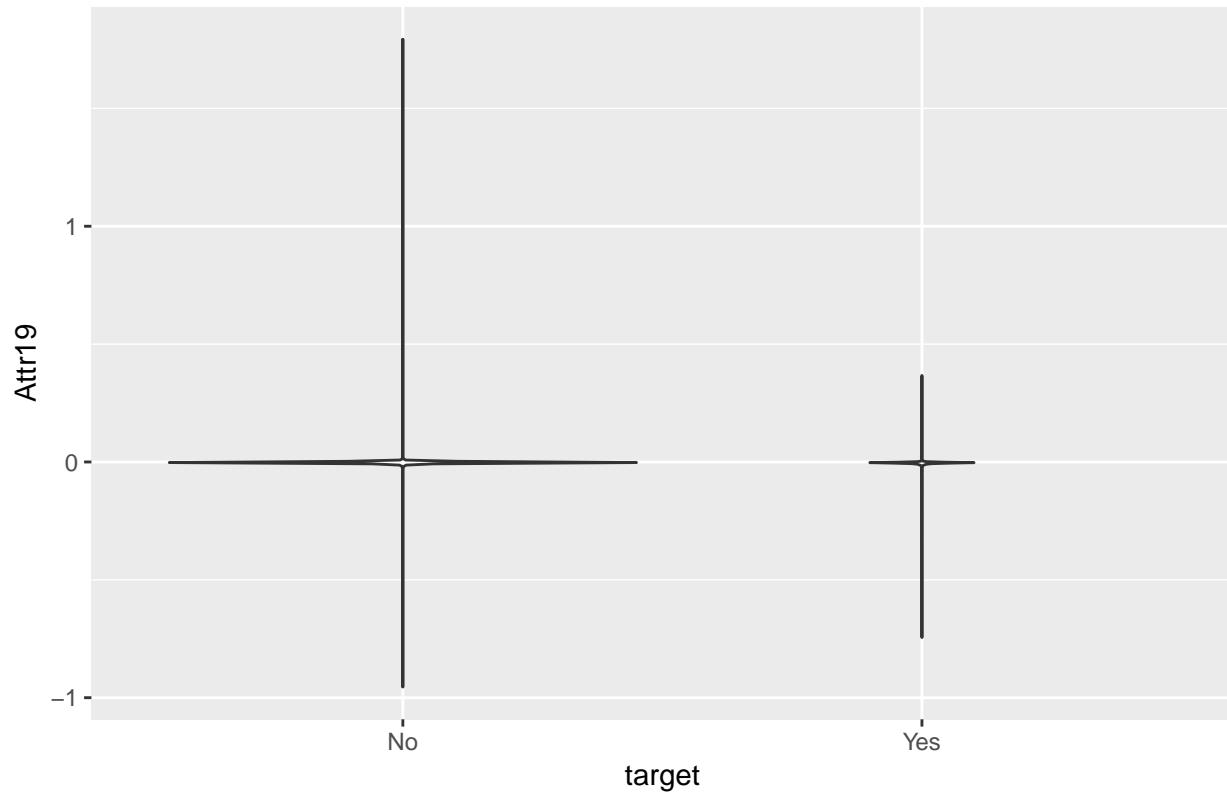
Bankruptcy Status by Attr15



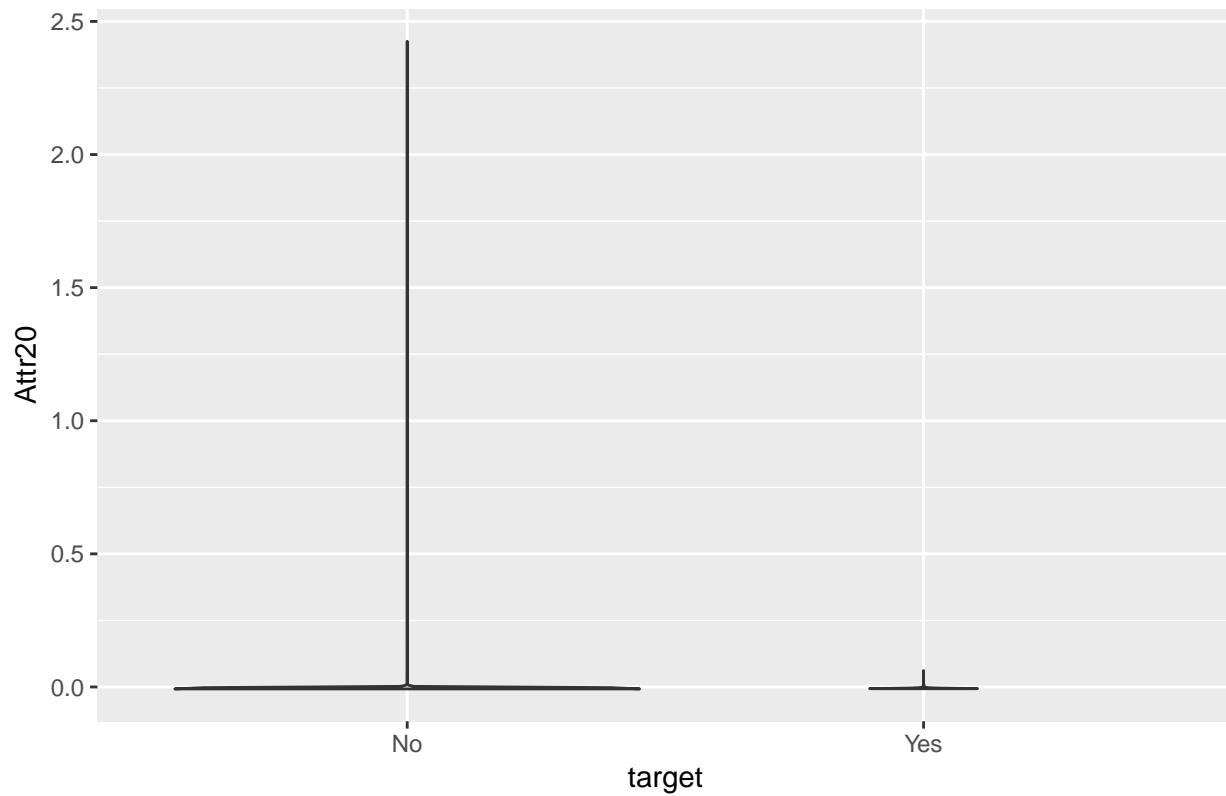
Bankruptcy Status by Attr16



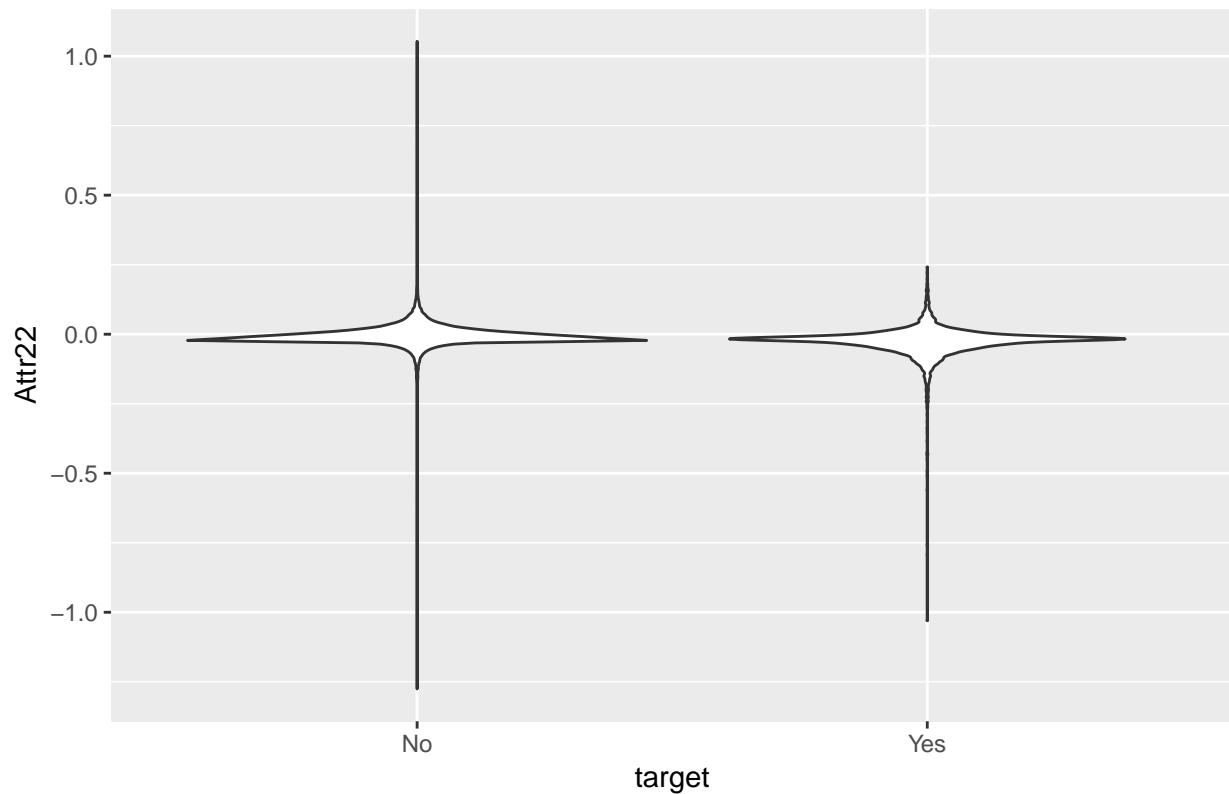
Bankruptcy Status by Attr19



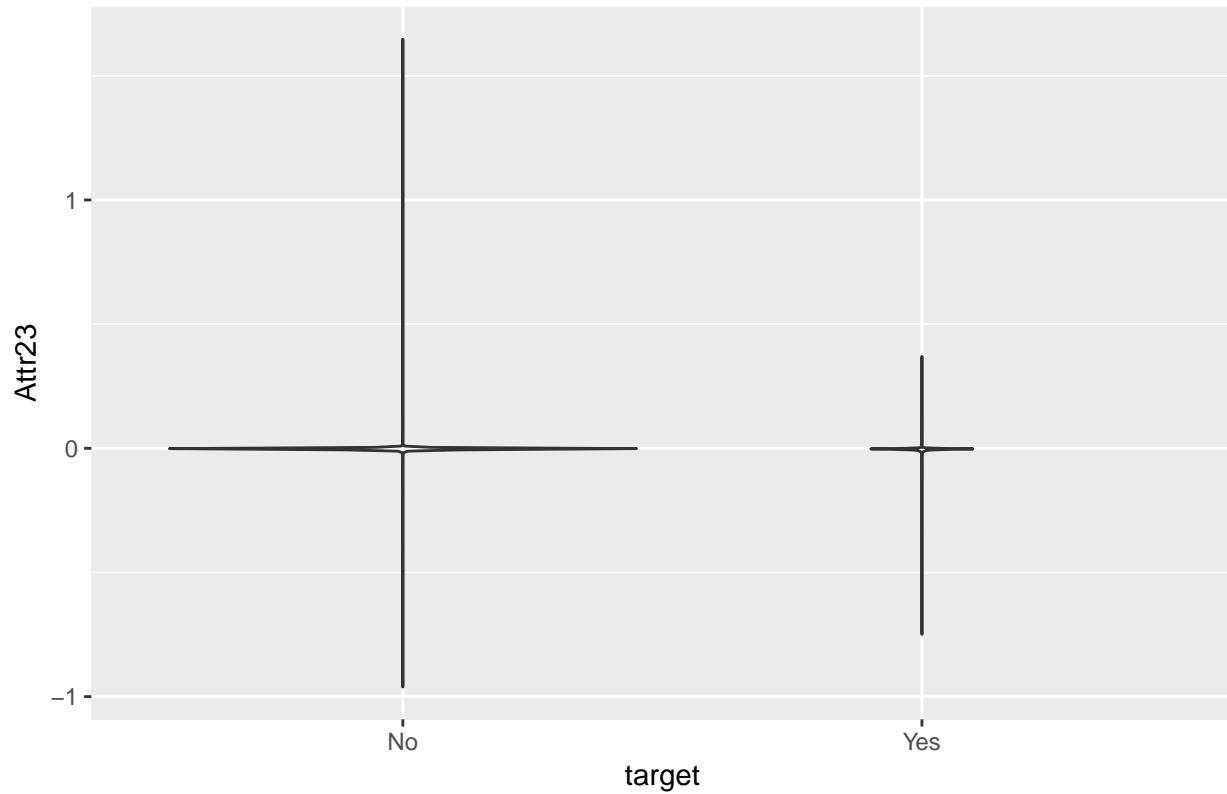
Bankruptcy Status by Attr20



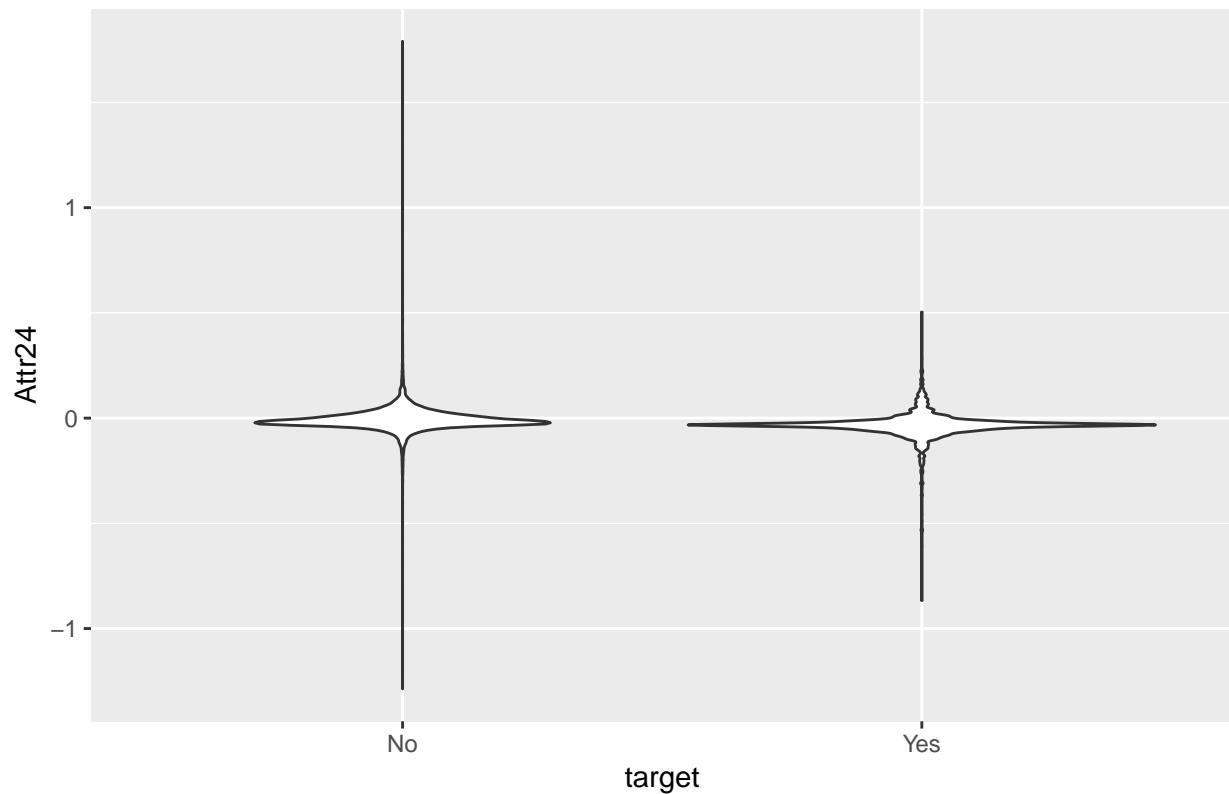
Bankruptcy Status by Attr22



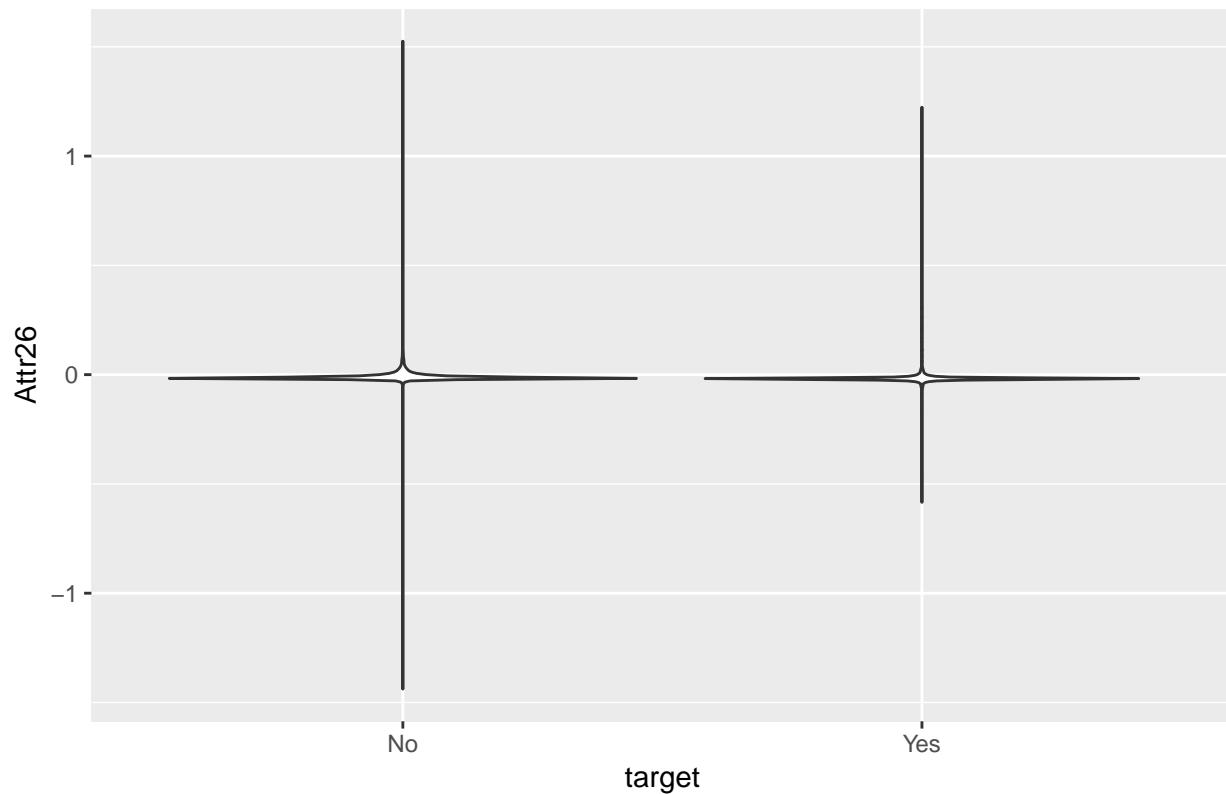
Bankruptcy Status by Attr23



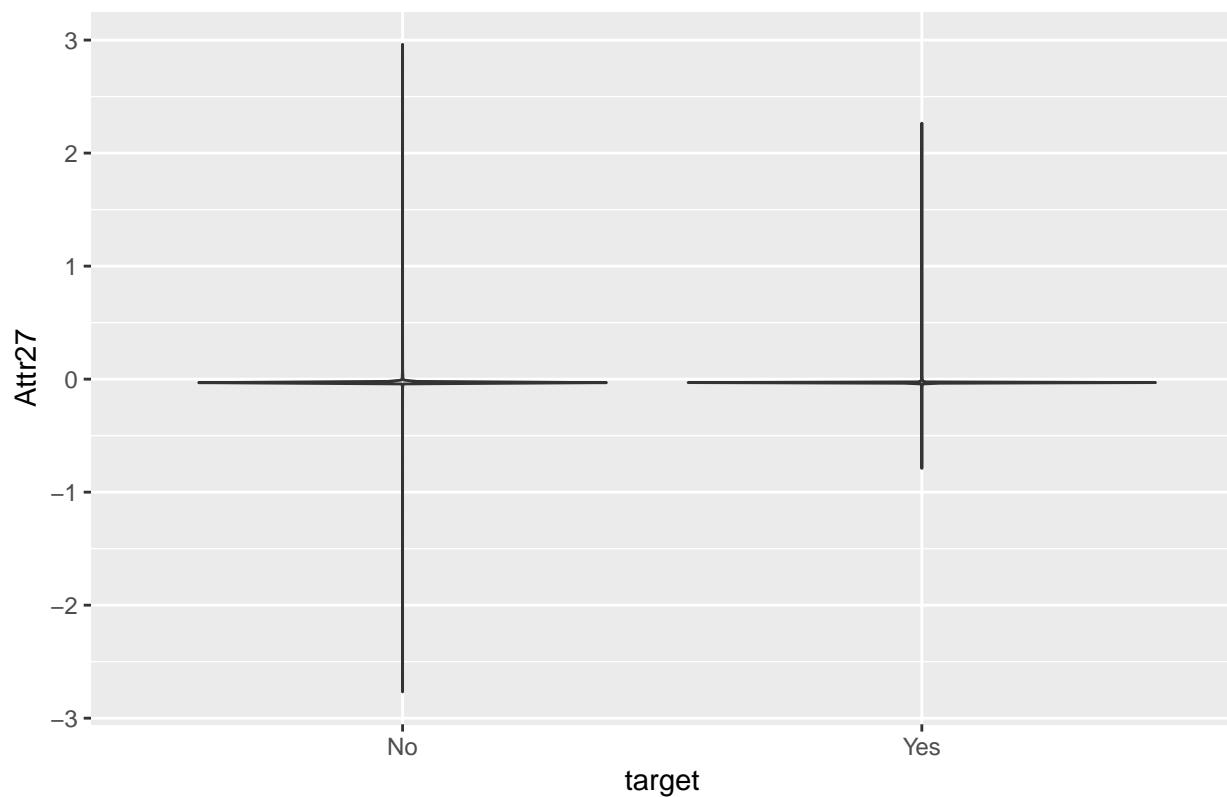
Bankruptcy Status by Attr24



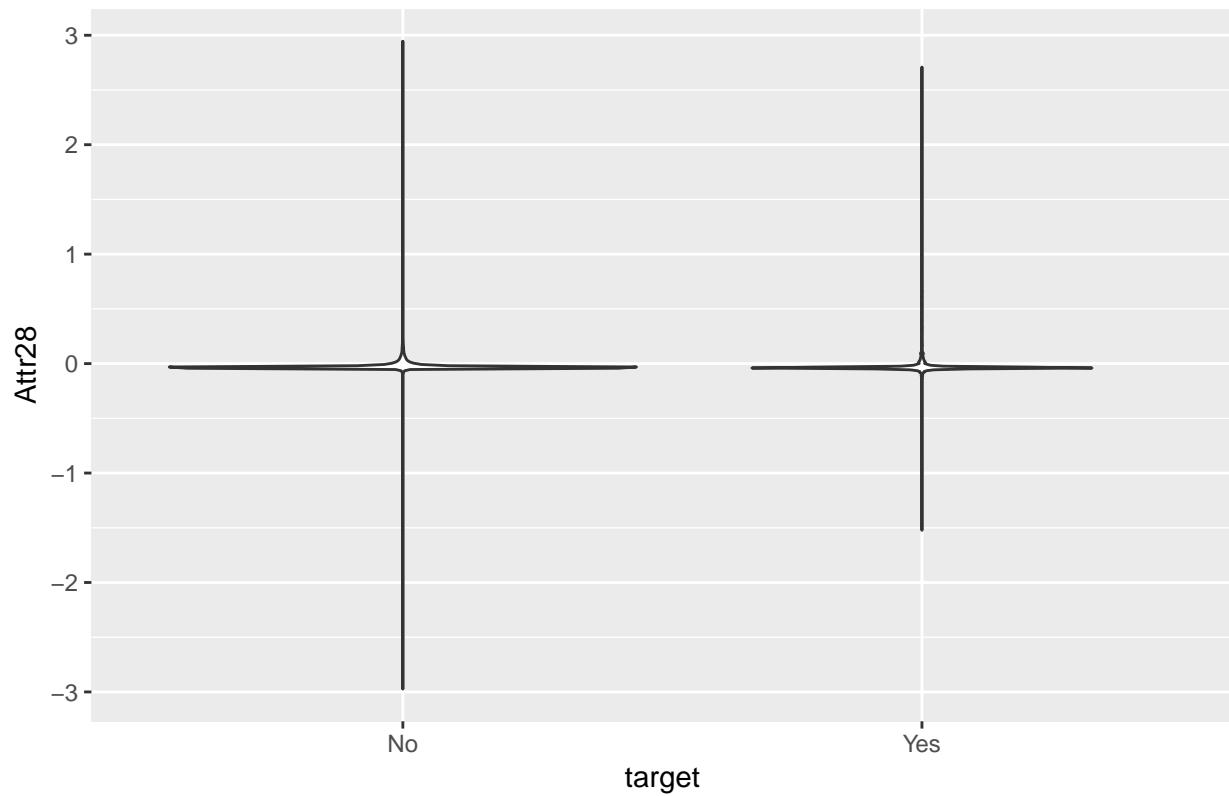
Bankruptcy Status by Attr26



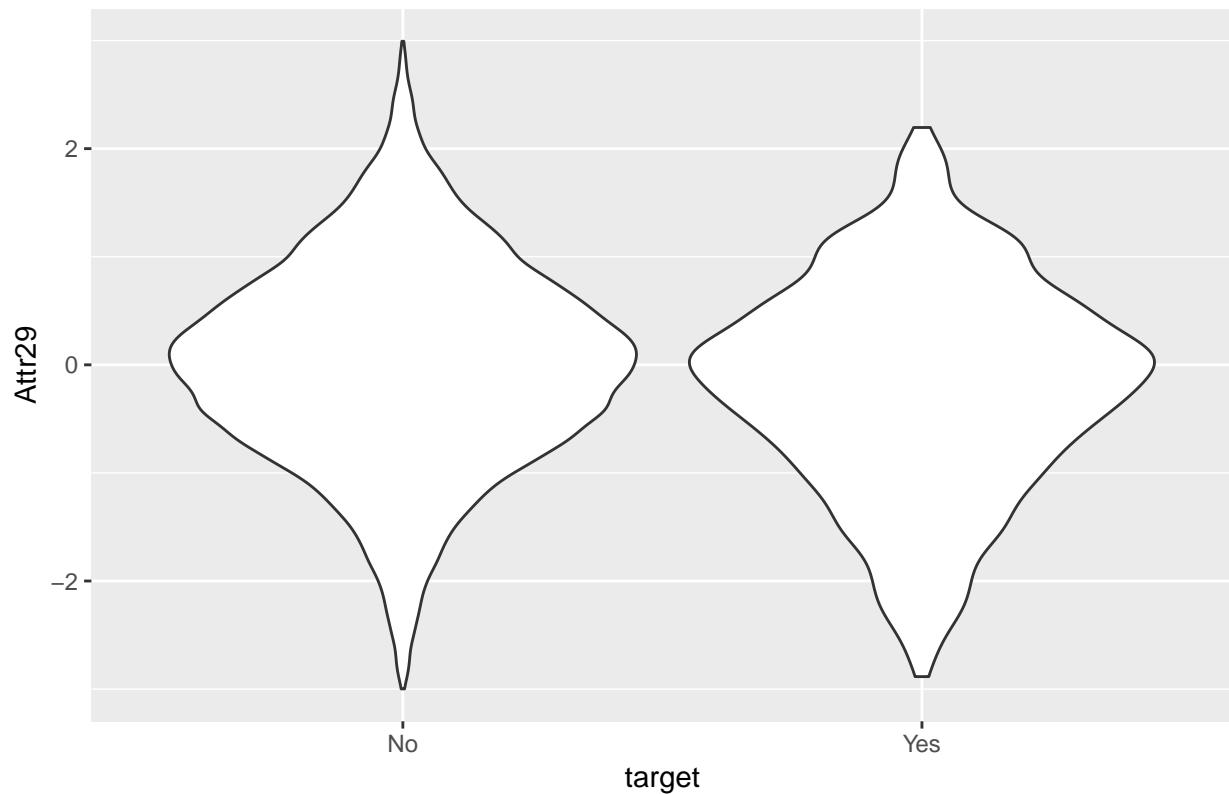
Bankruptcy Status by Attr27



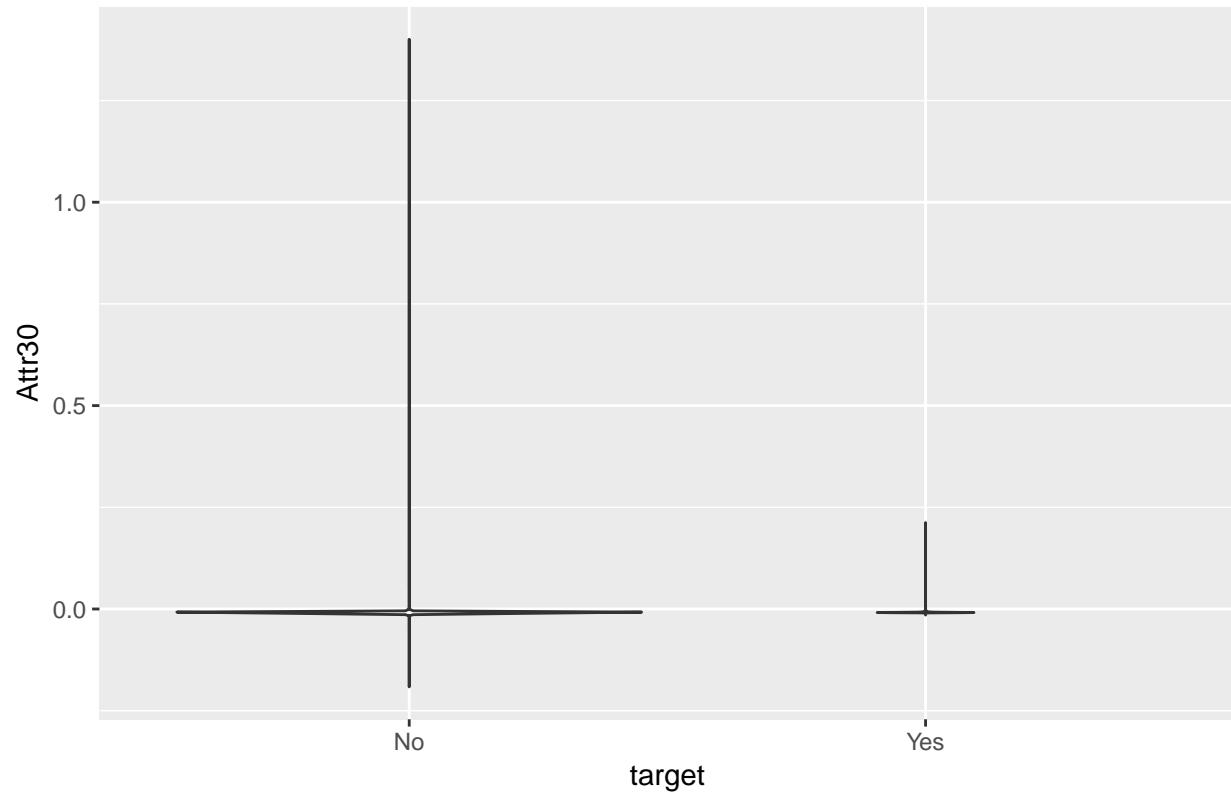
Bankruptcy Status by Attr28



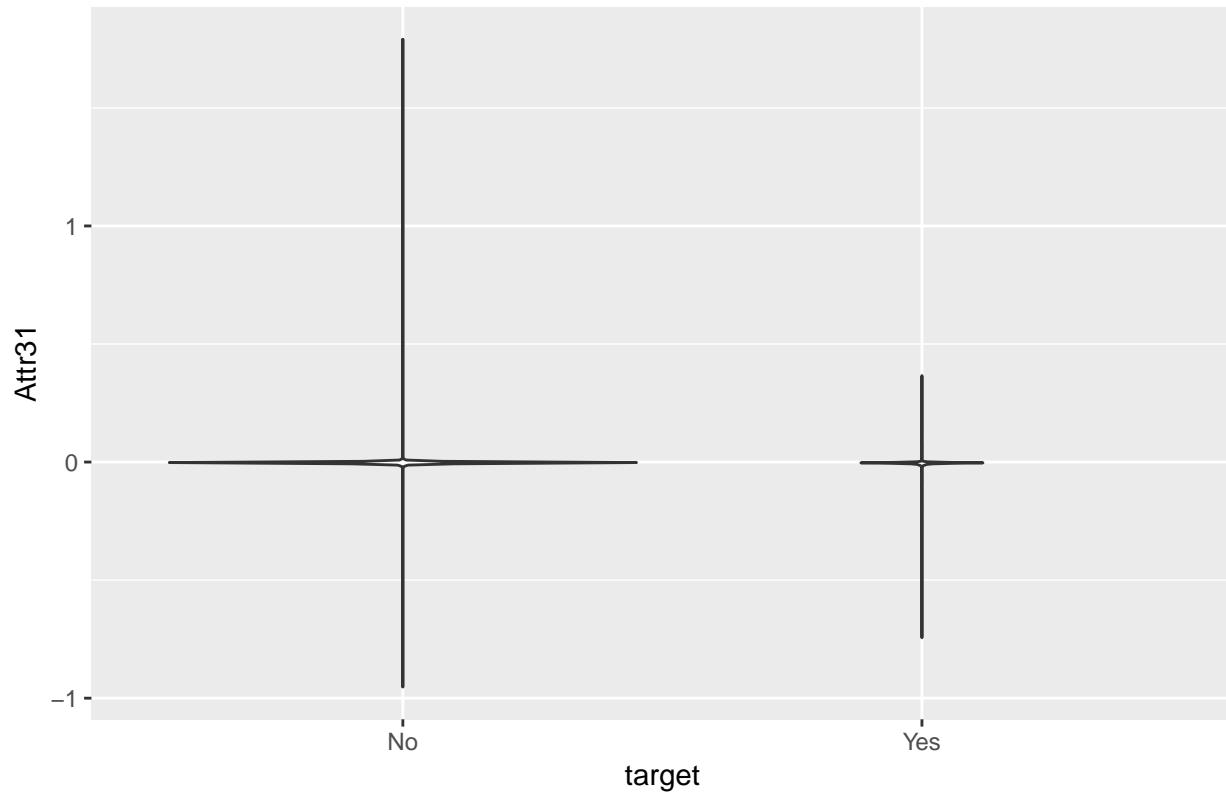
Bankruptcy Status by Attr29



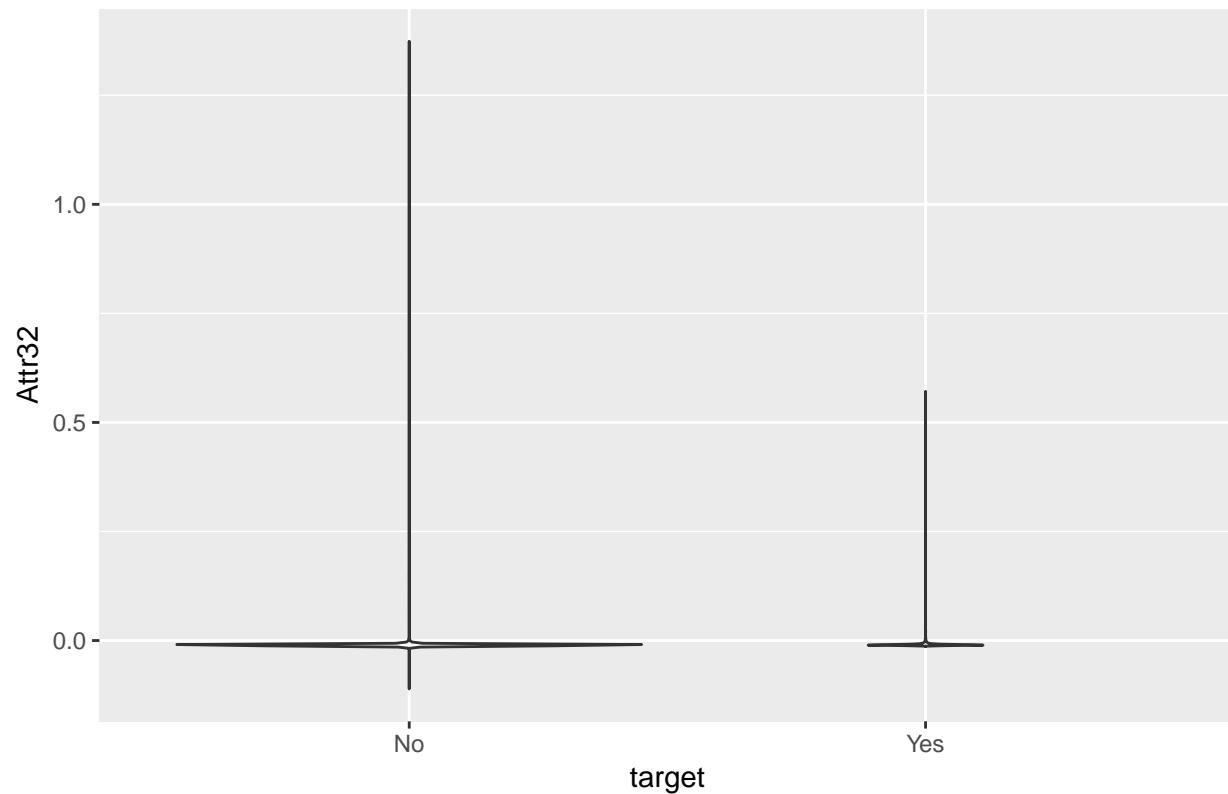
Bankruptcy Status by Attr30



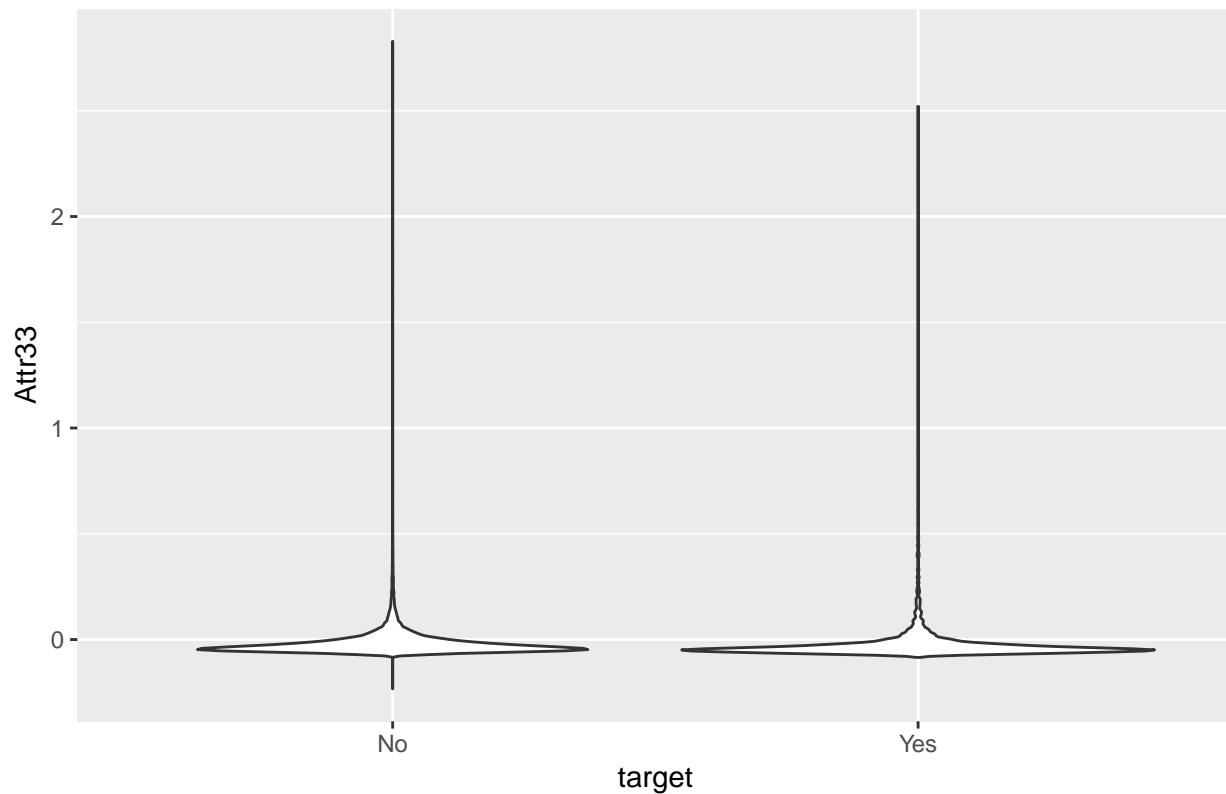
Bankruptcy Status by Attr31



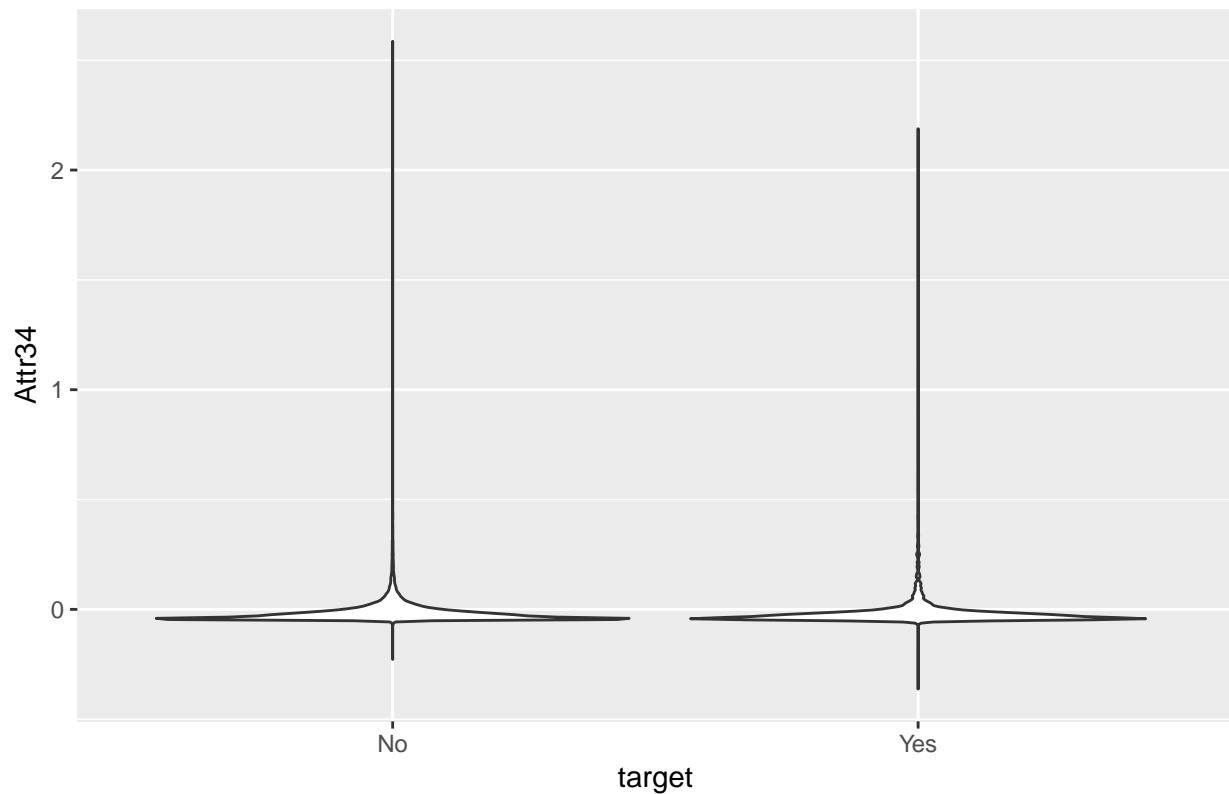
Bankruptcy Status by Attr32



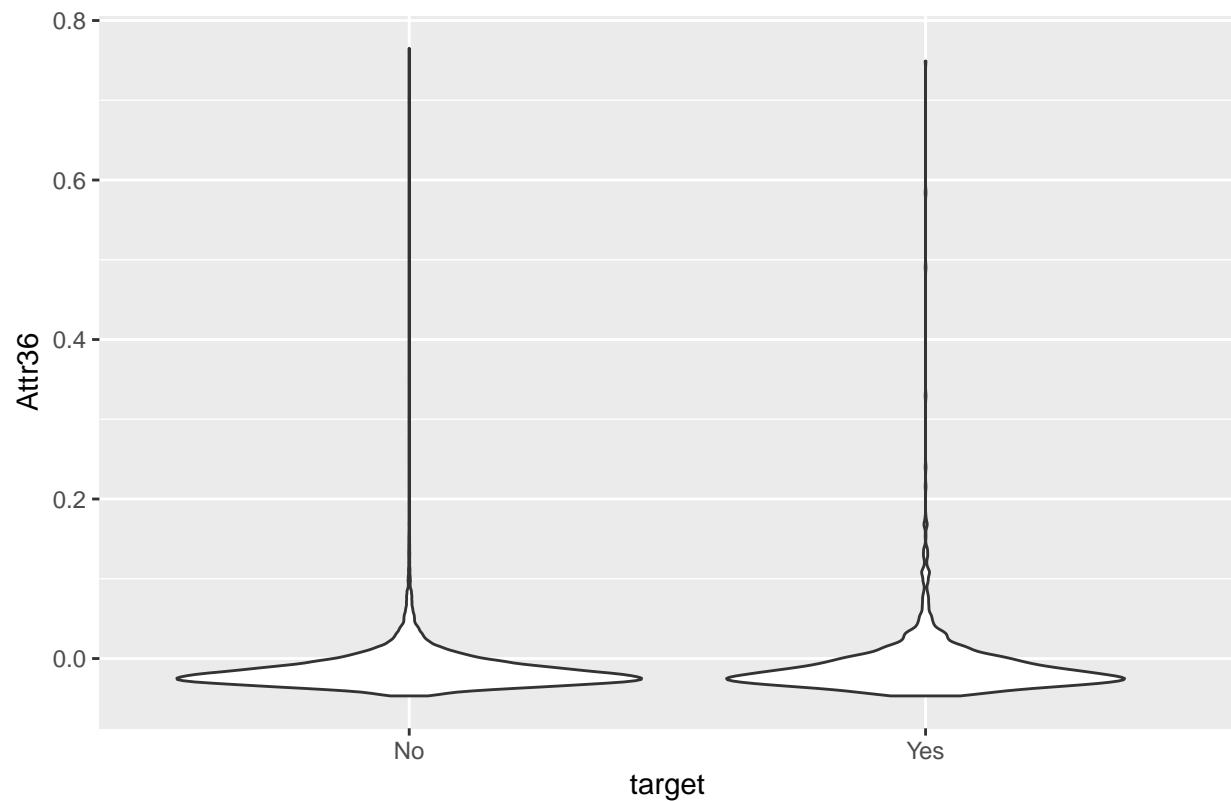
Bankruptcy Status by Attr33



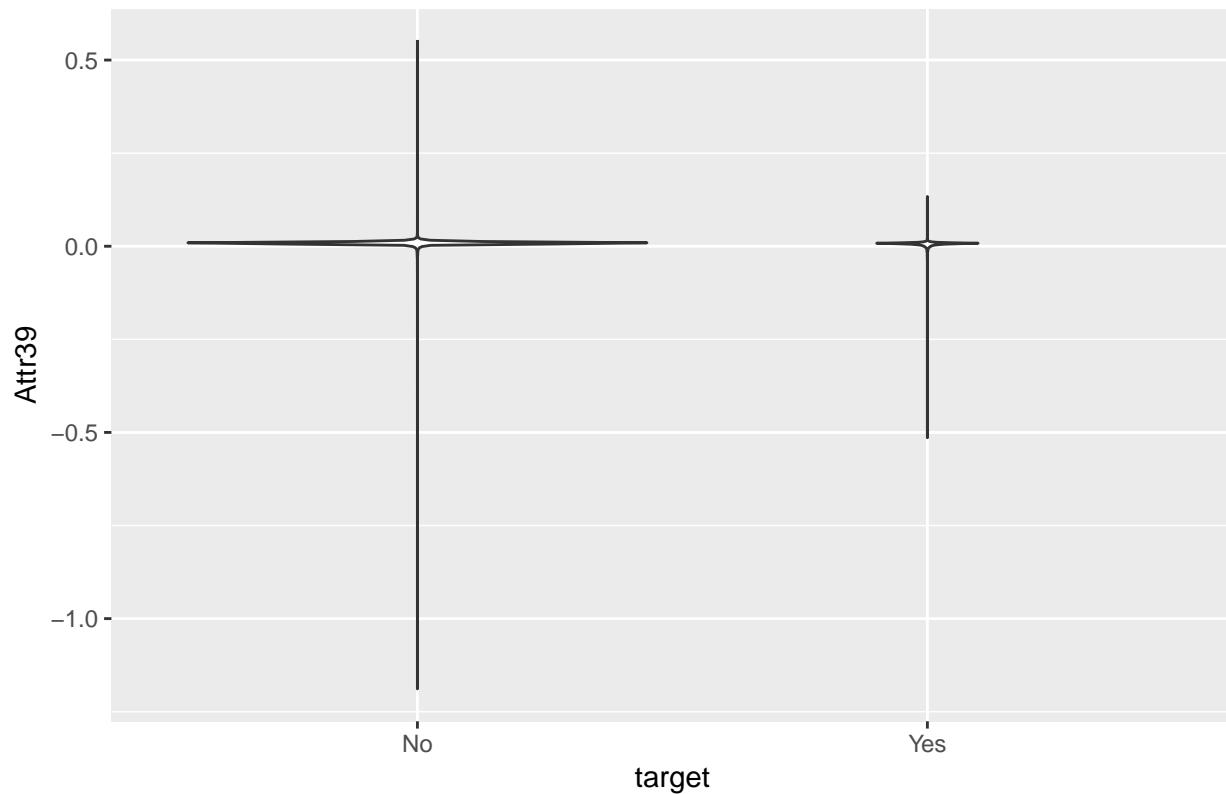
Bankruptcy Status by Attr34



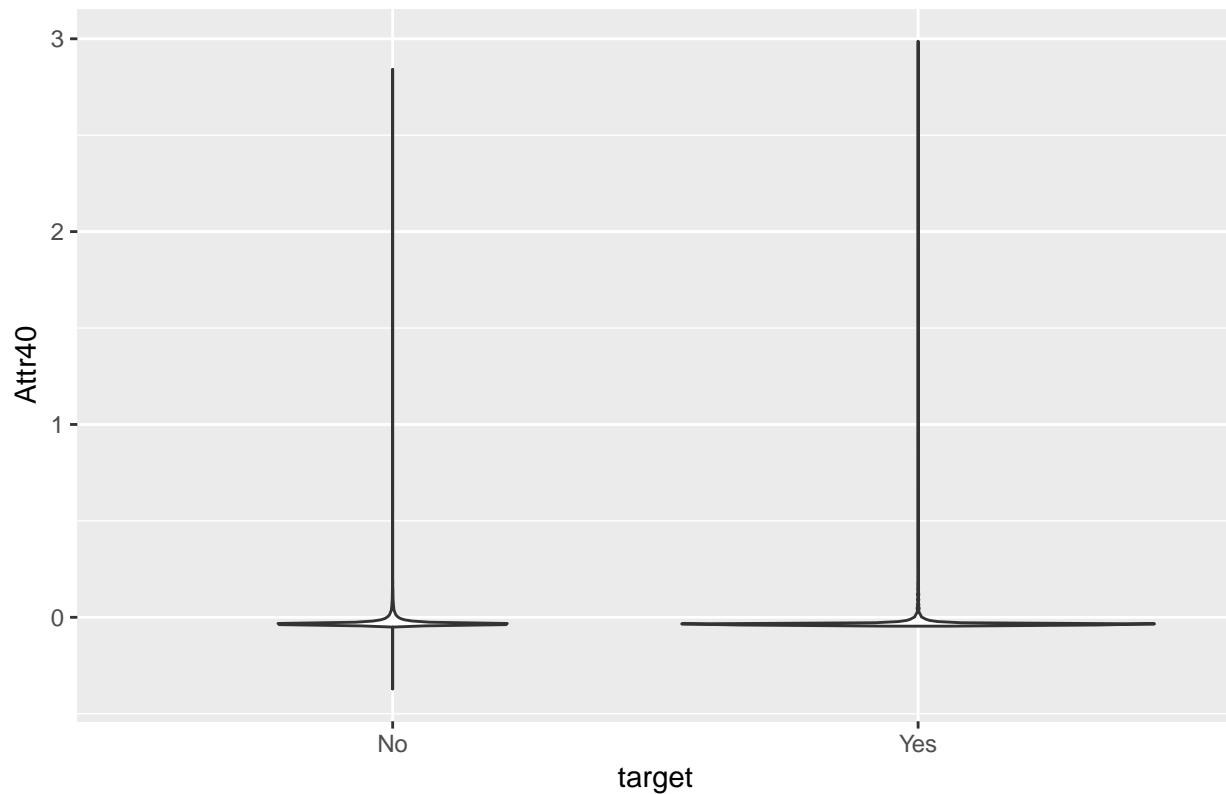
Bankruptcy Status by Attr36



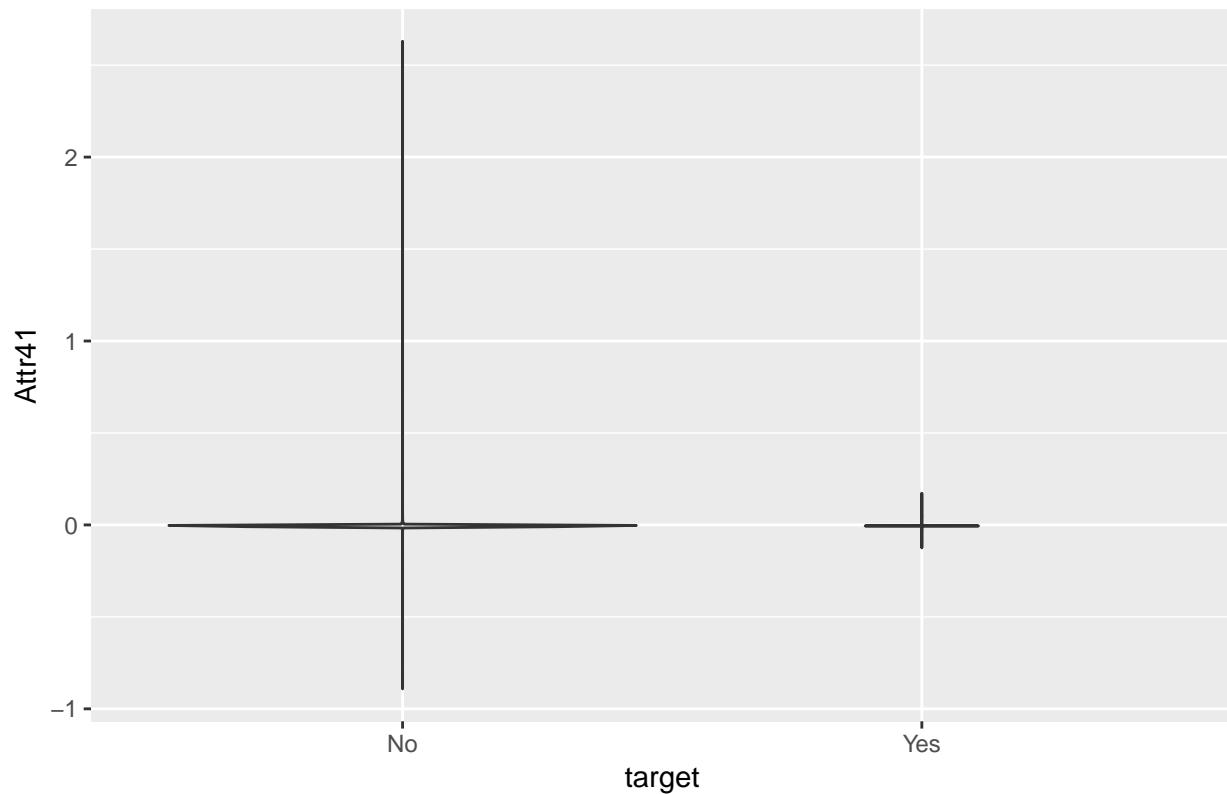
Bankruptcy Status by Attr39



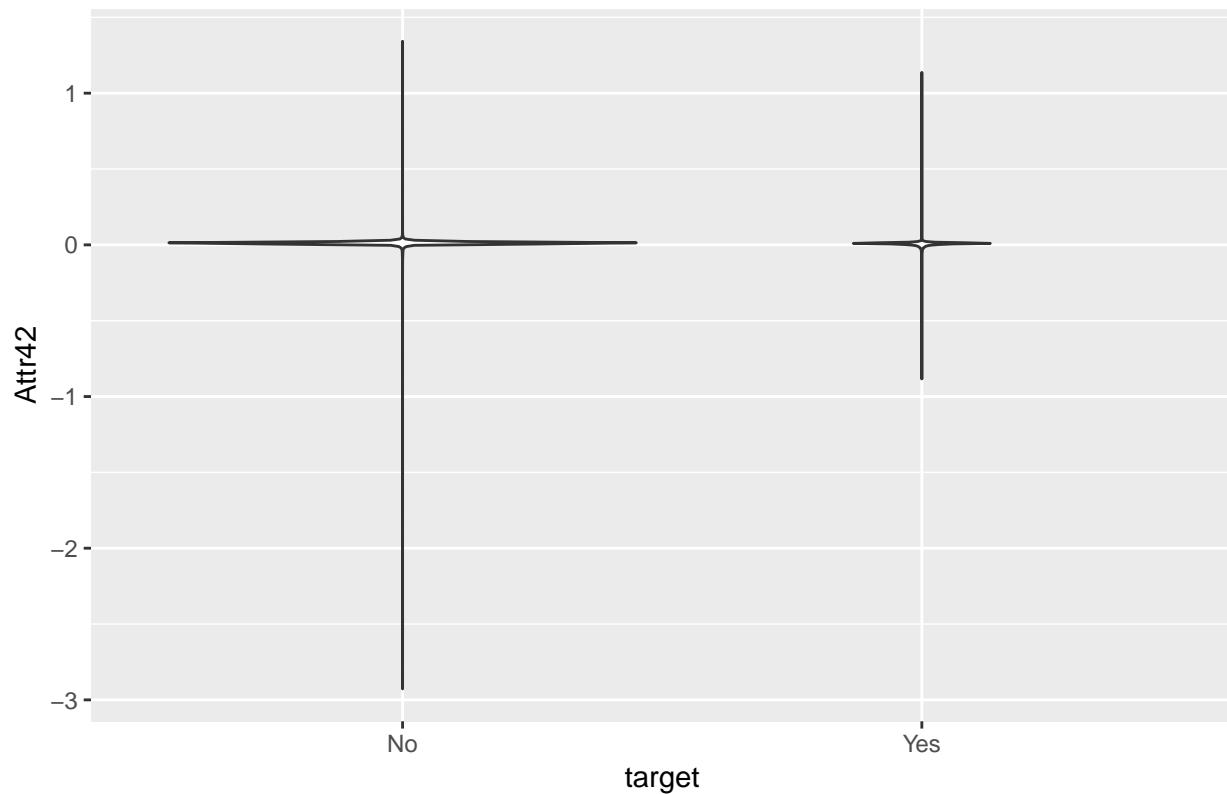
Bankruptcy Status by Attr40



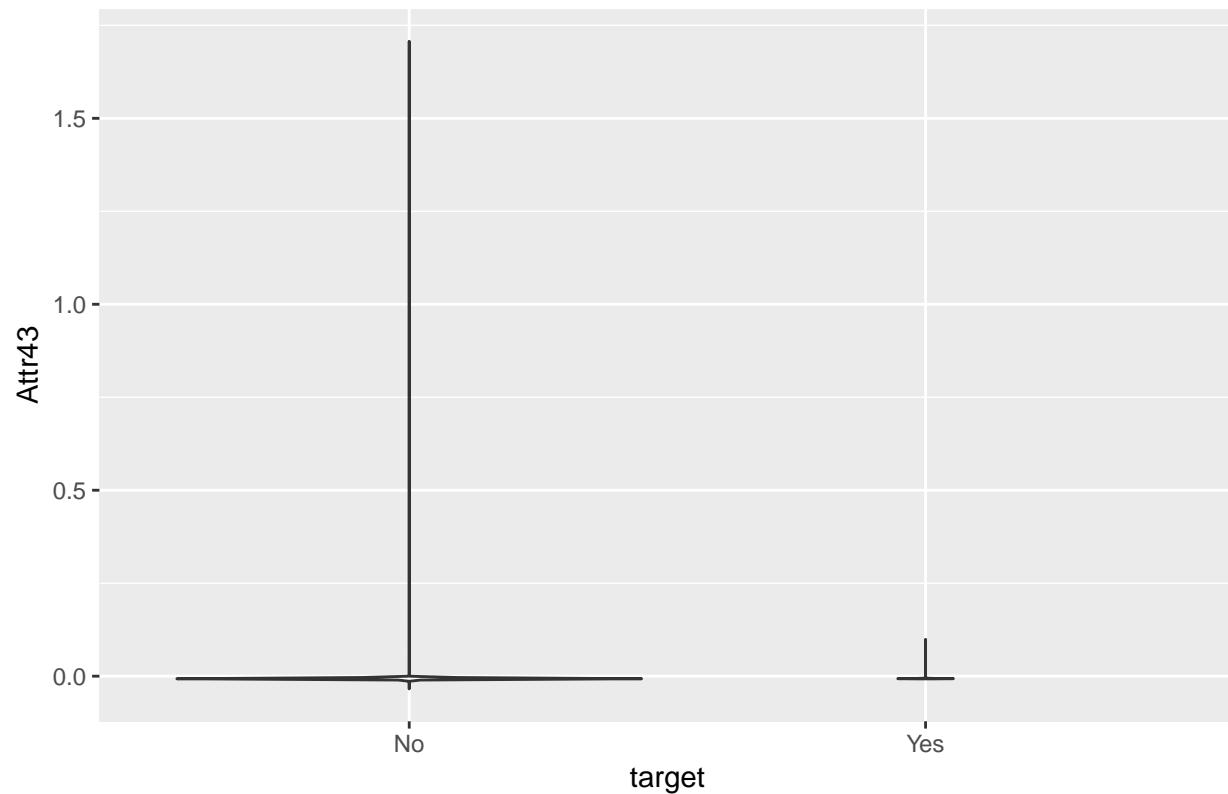
Bankruptcy Status by Attr41



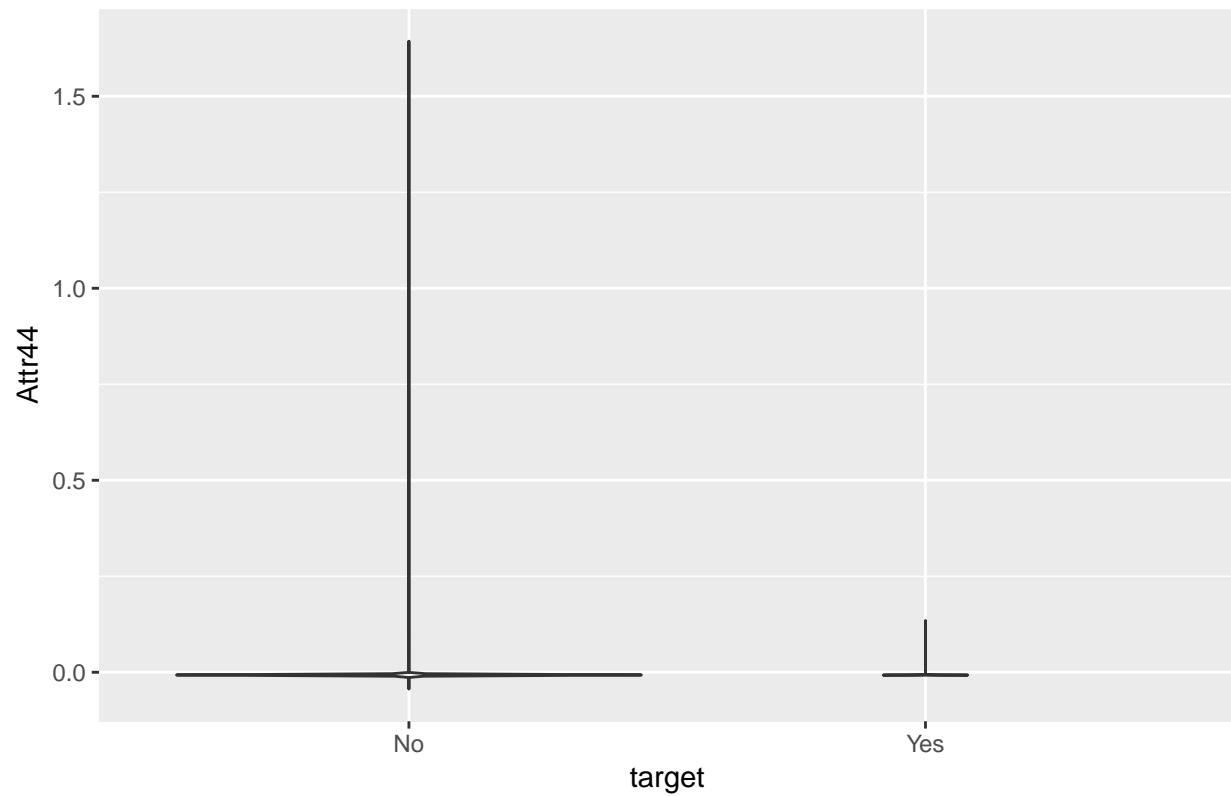
Bankruptcy Status by Attr42



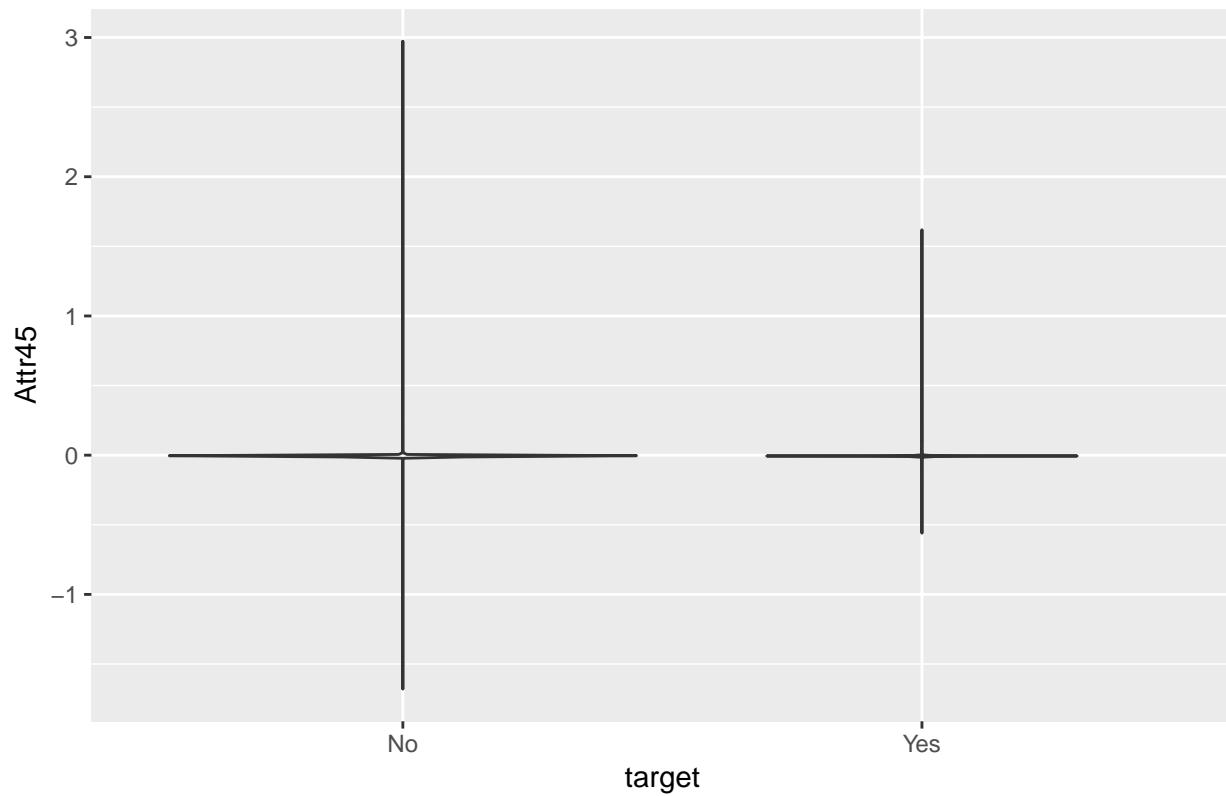
Bankruptcy Status by Attr43



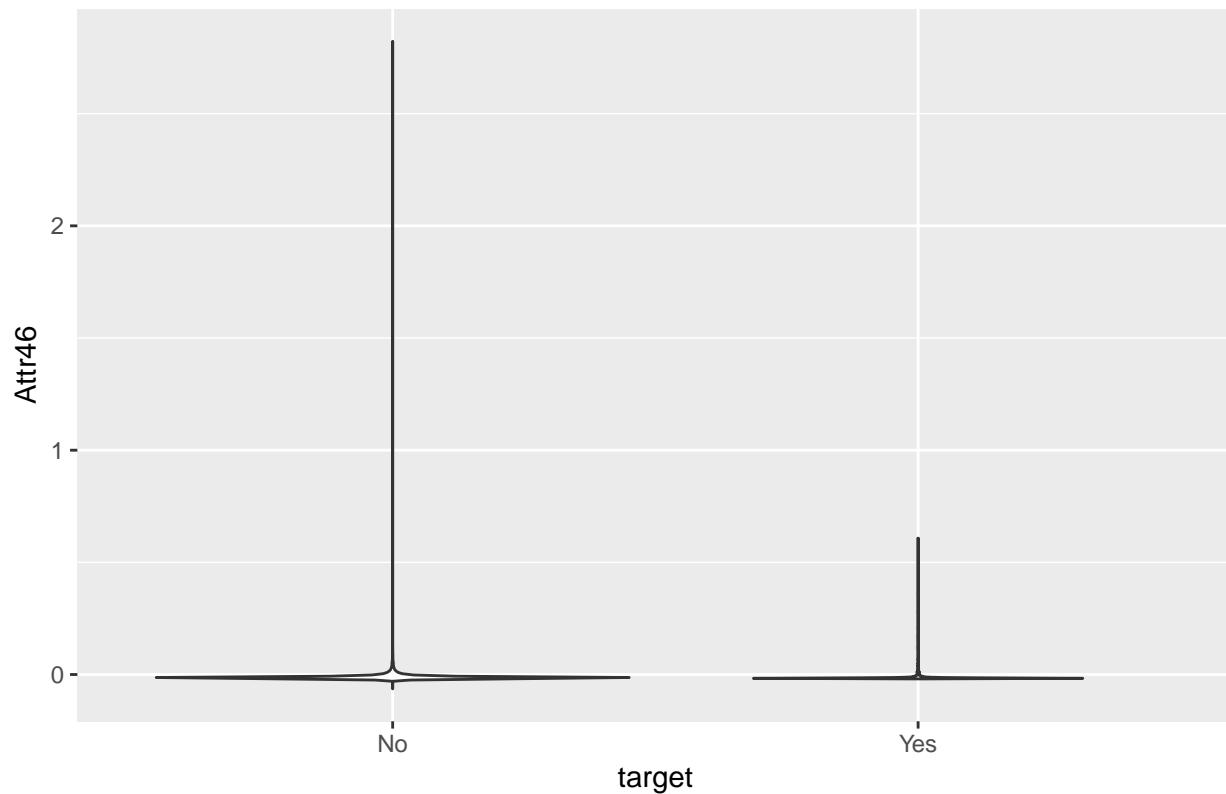
Bankruptcy Status by Attr44



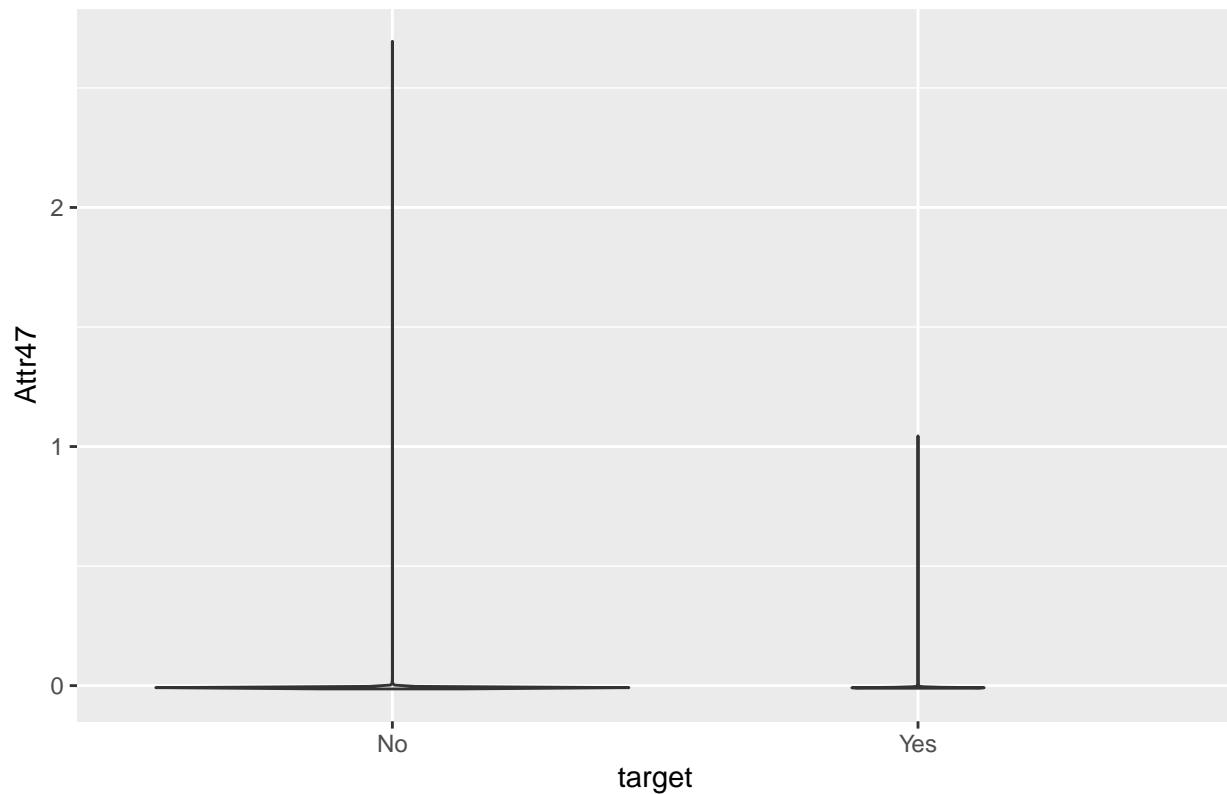
Bankruptcy Status by Attr45



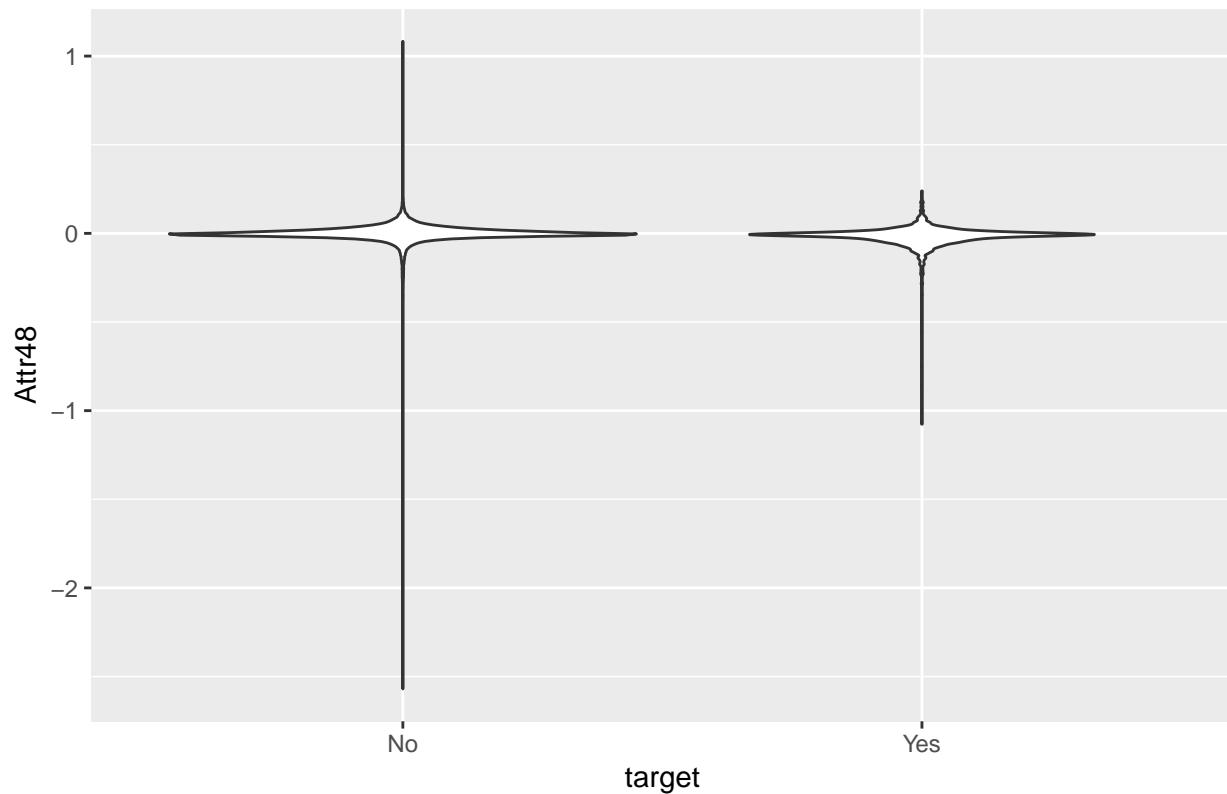
Bankruptcy Status by Attr46



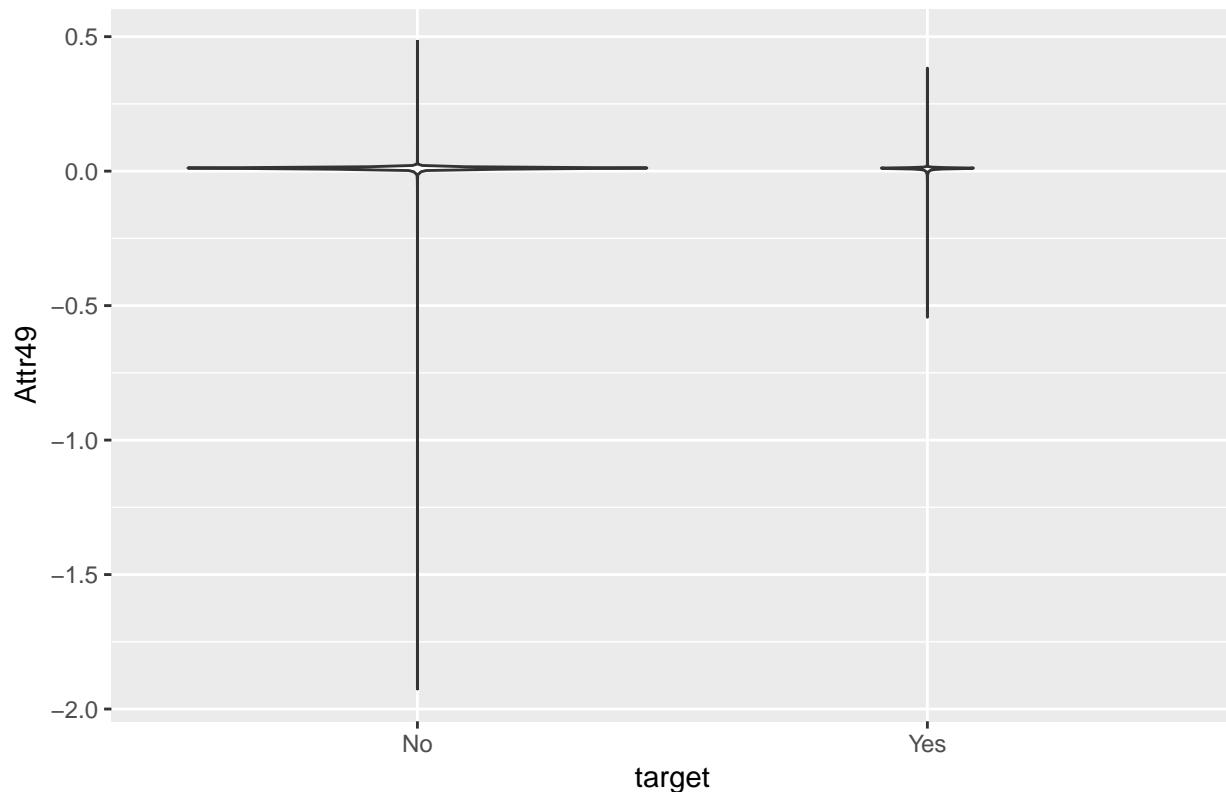
Bankruptcy Status by Attr47



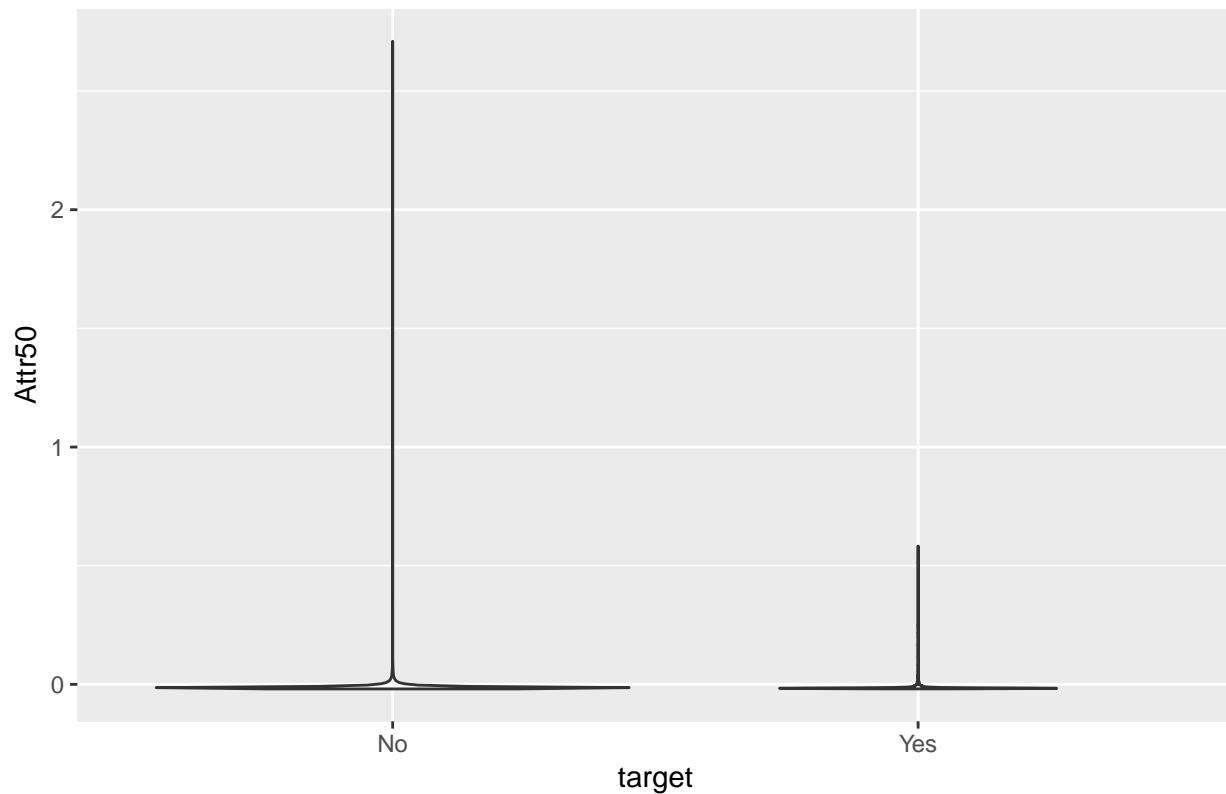
Bankruptcy Status by Attr48



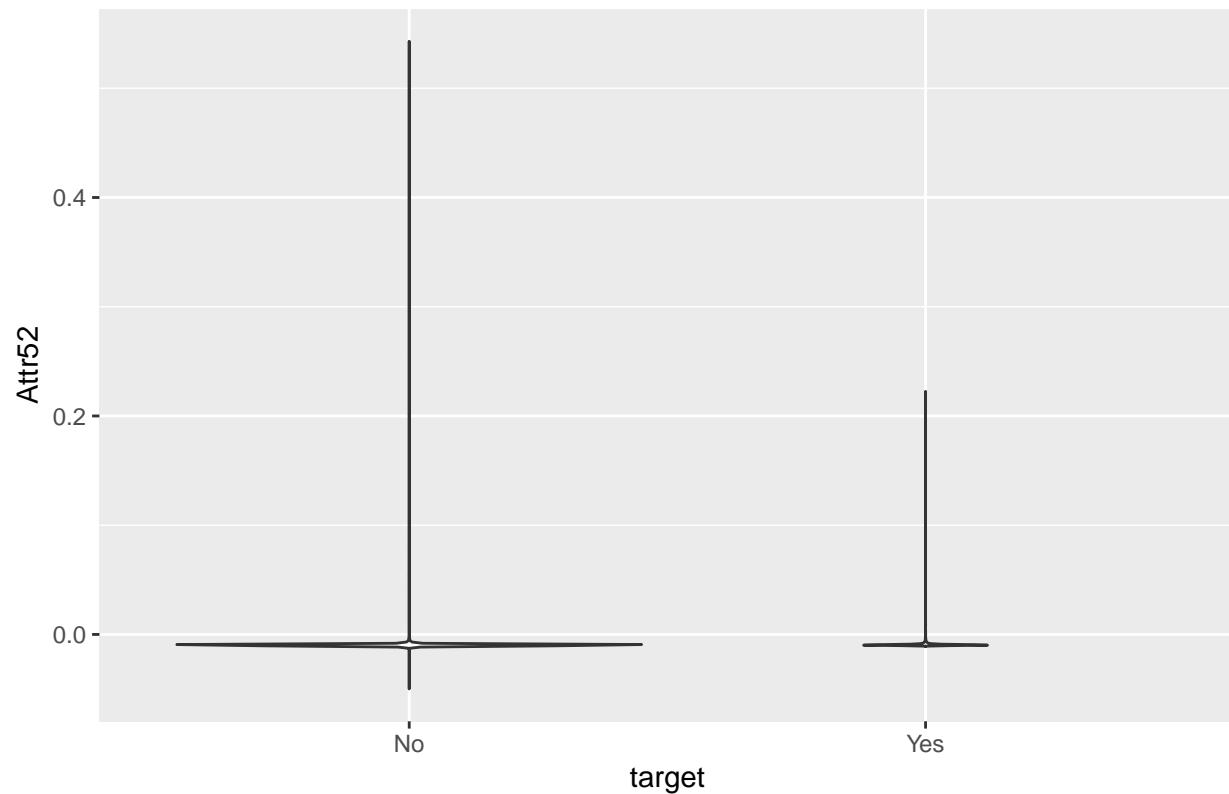
Bankruptcy Status by Attr49



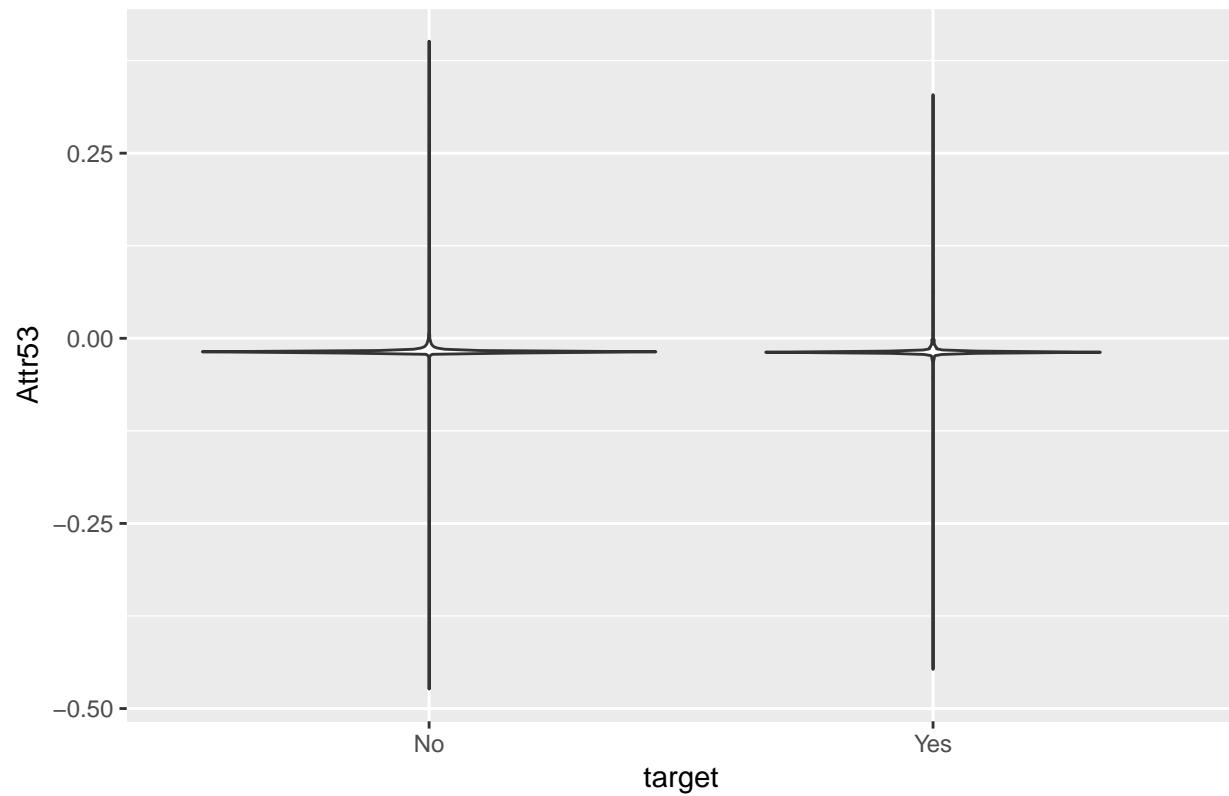
Bankruptcy Status by Attr50



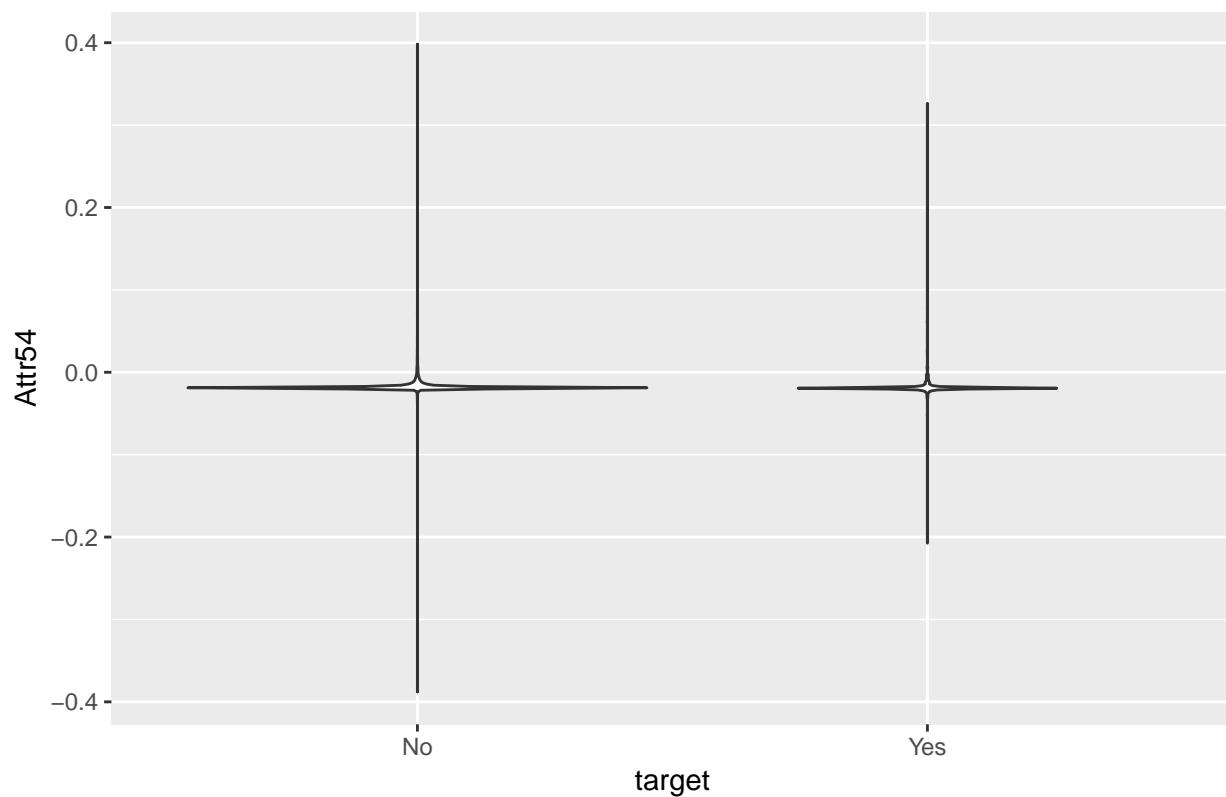
Bankruptcy Status by Attr52



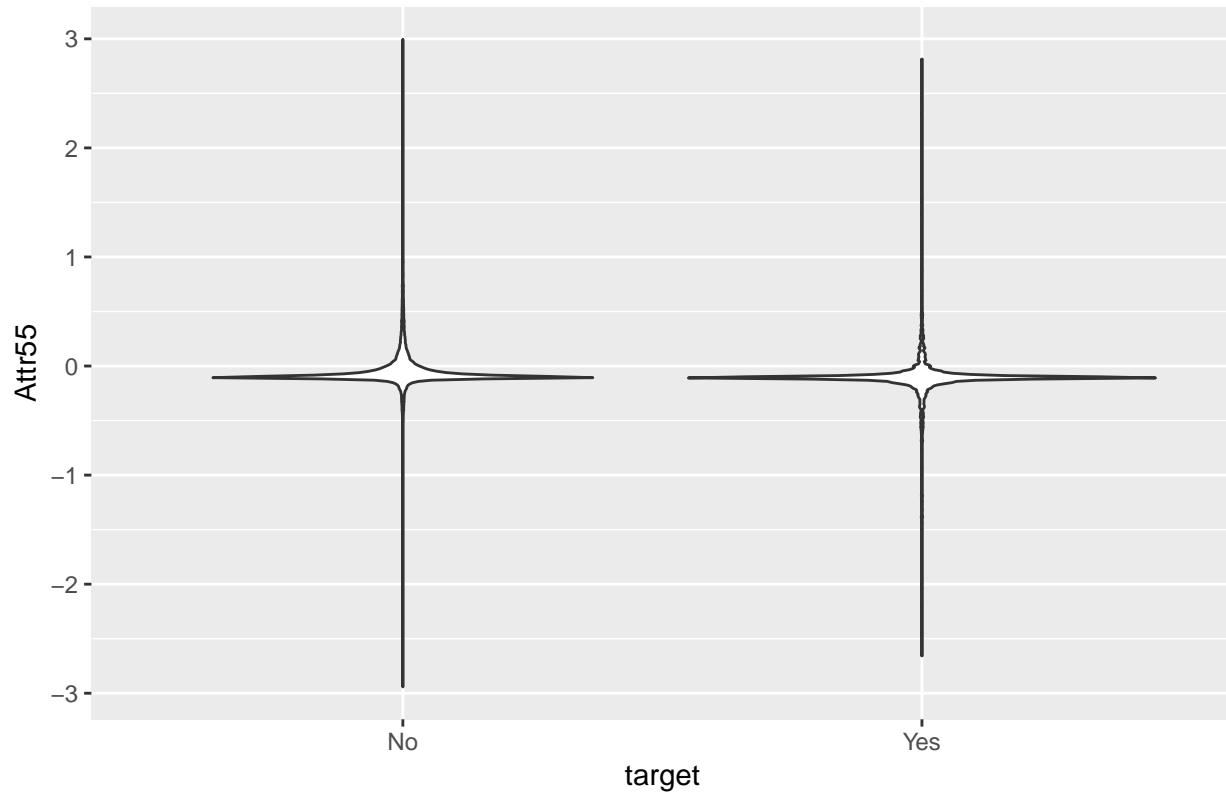
Bankruptcy Status by Attr53



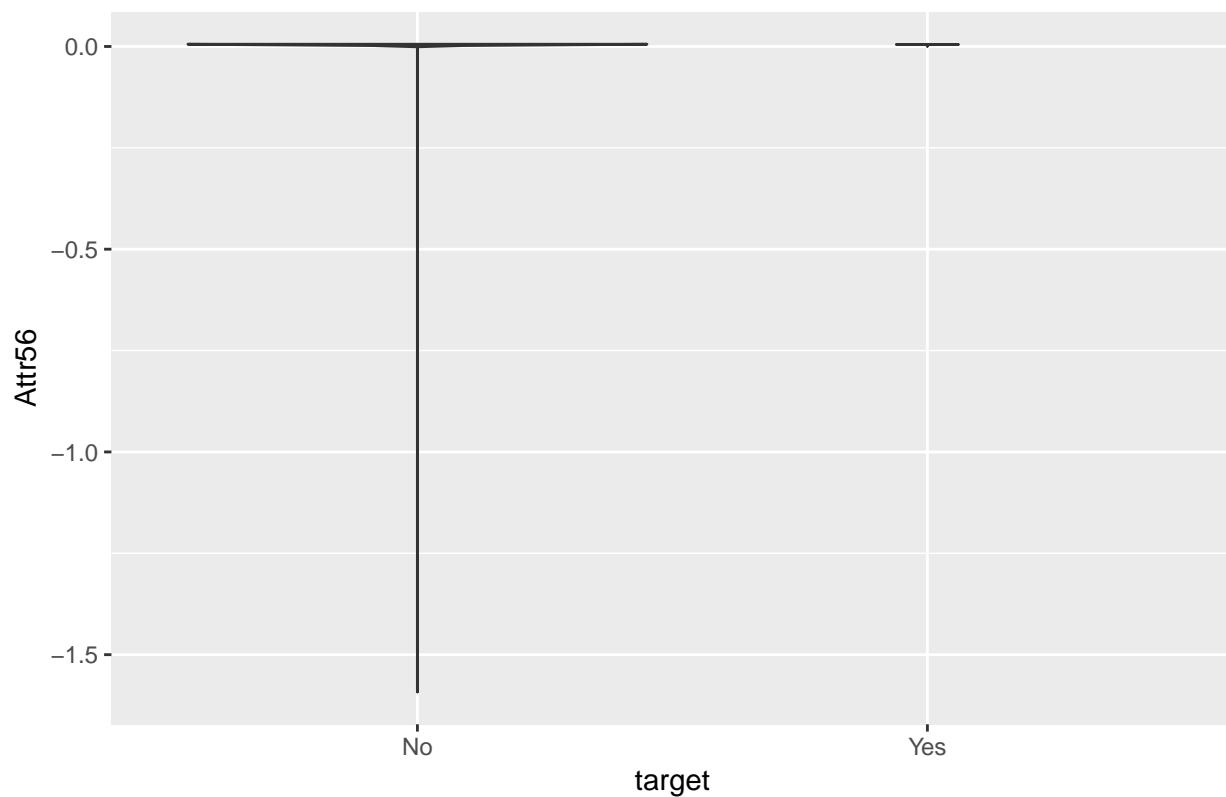
Bankruptcy Status by Attr54



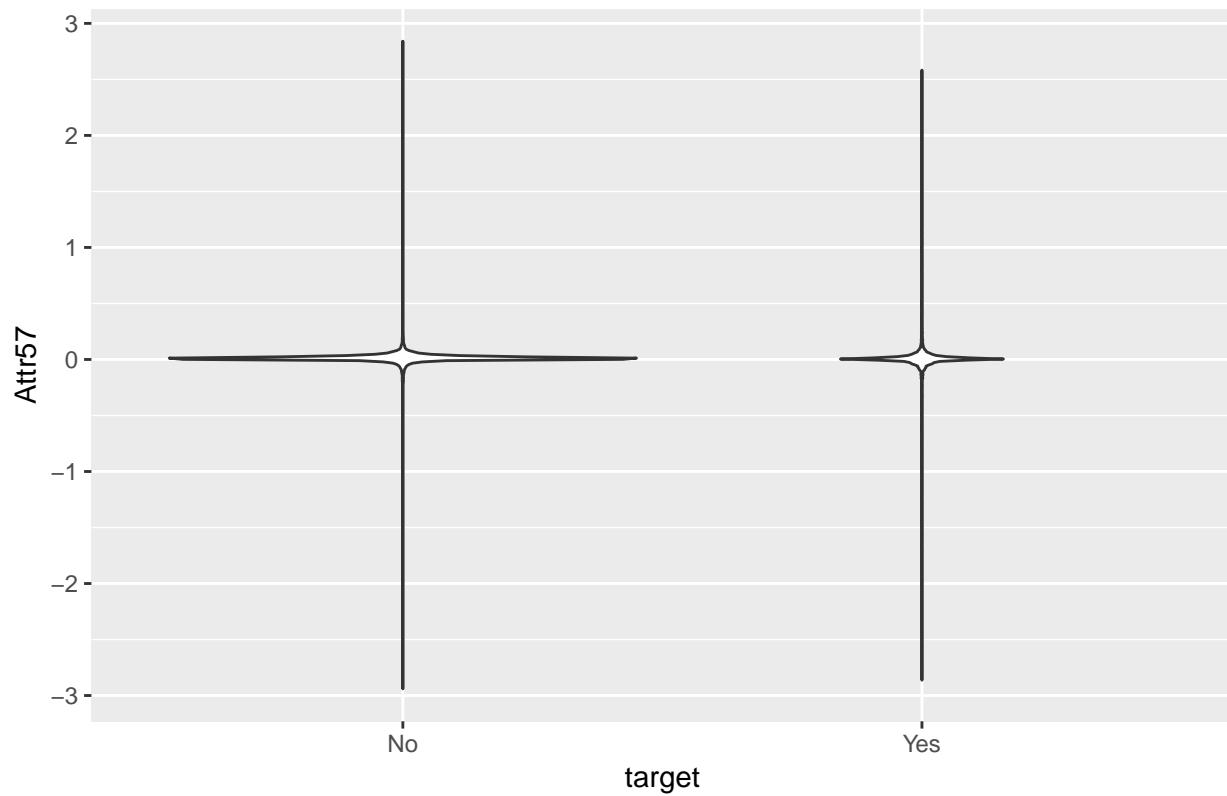
Bankruptcy Status by Attr55



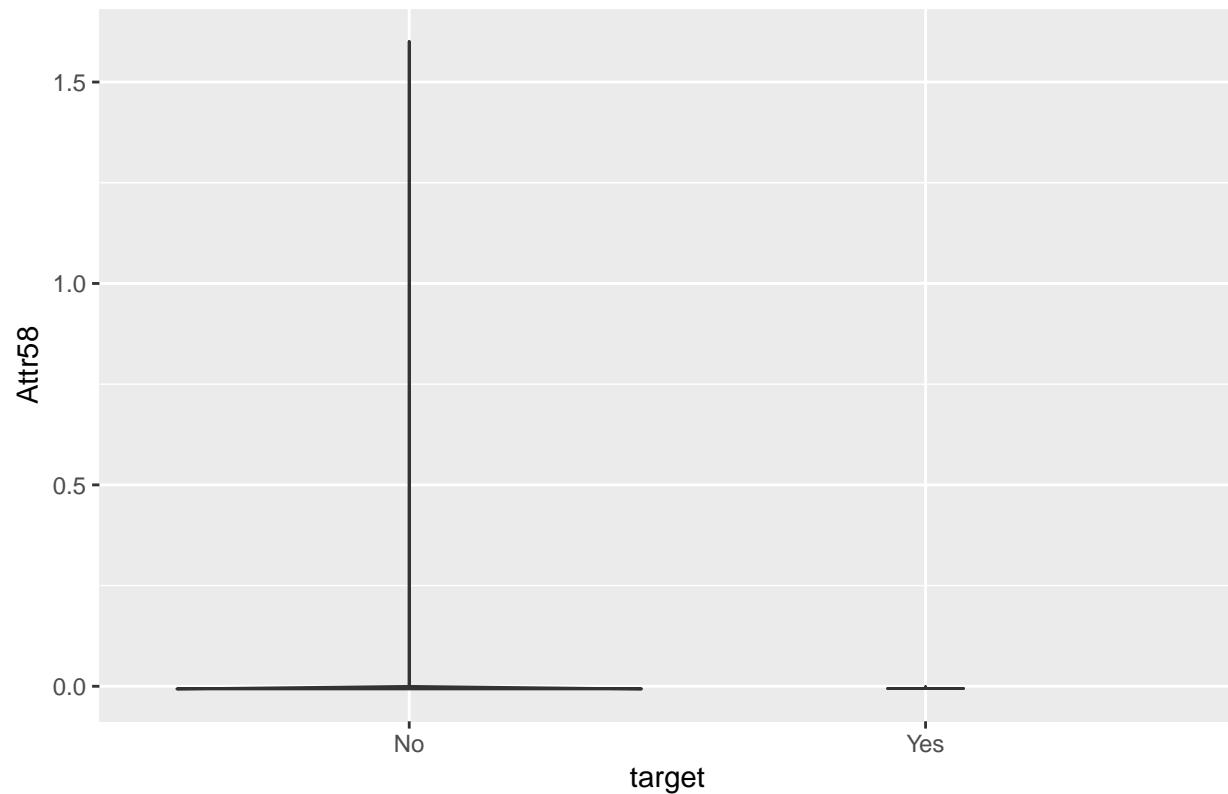
Bankruptcy Status by Attr56



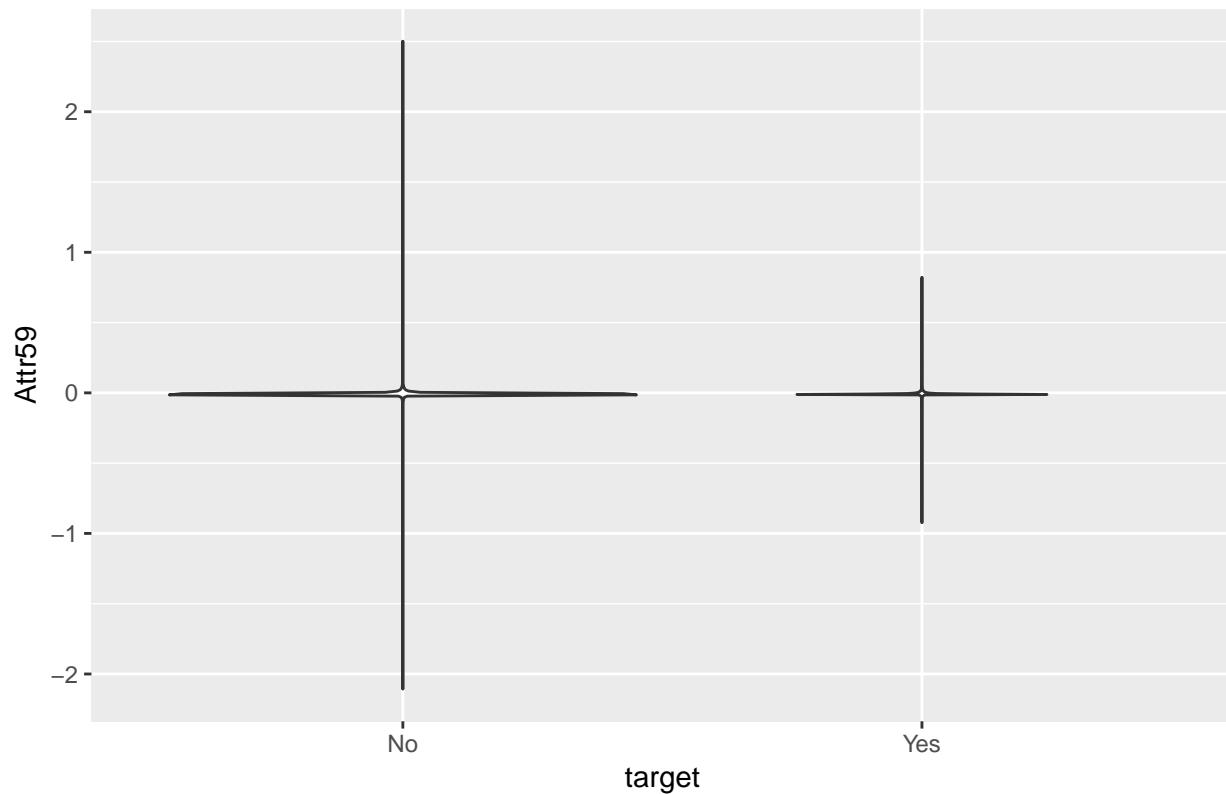
Bankruptcy Status by Attr57



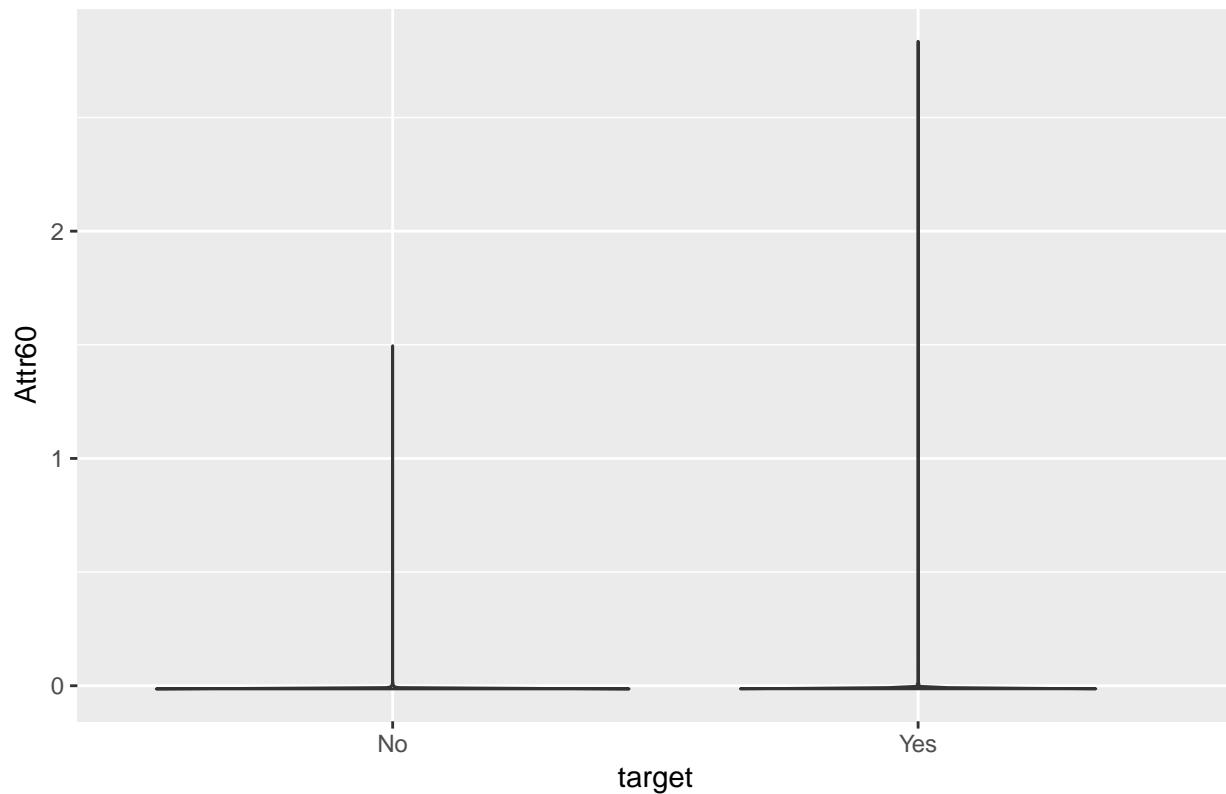
Bankruptcy Status by Attr58



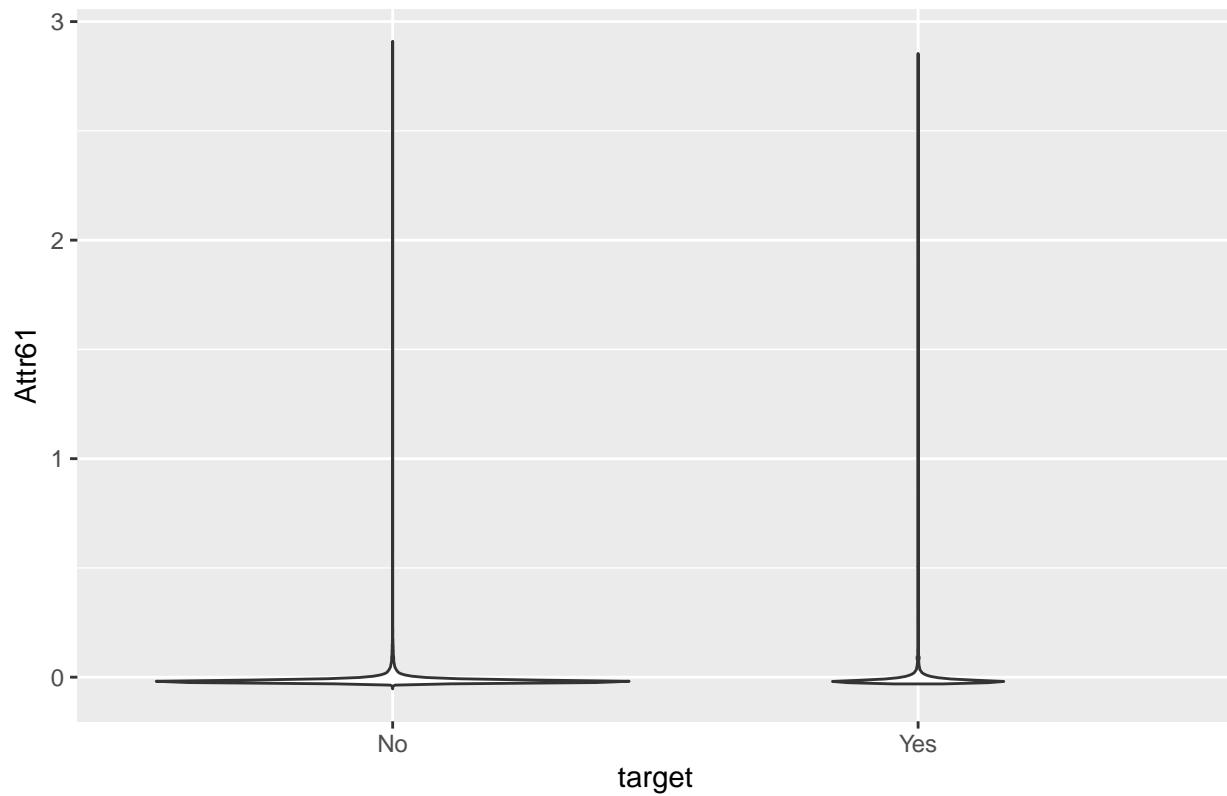
Bankruptcy Status by Attr59



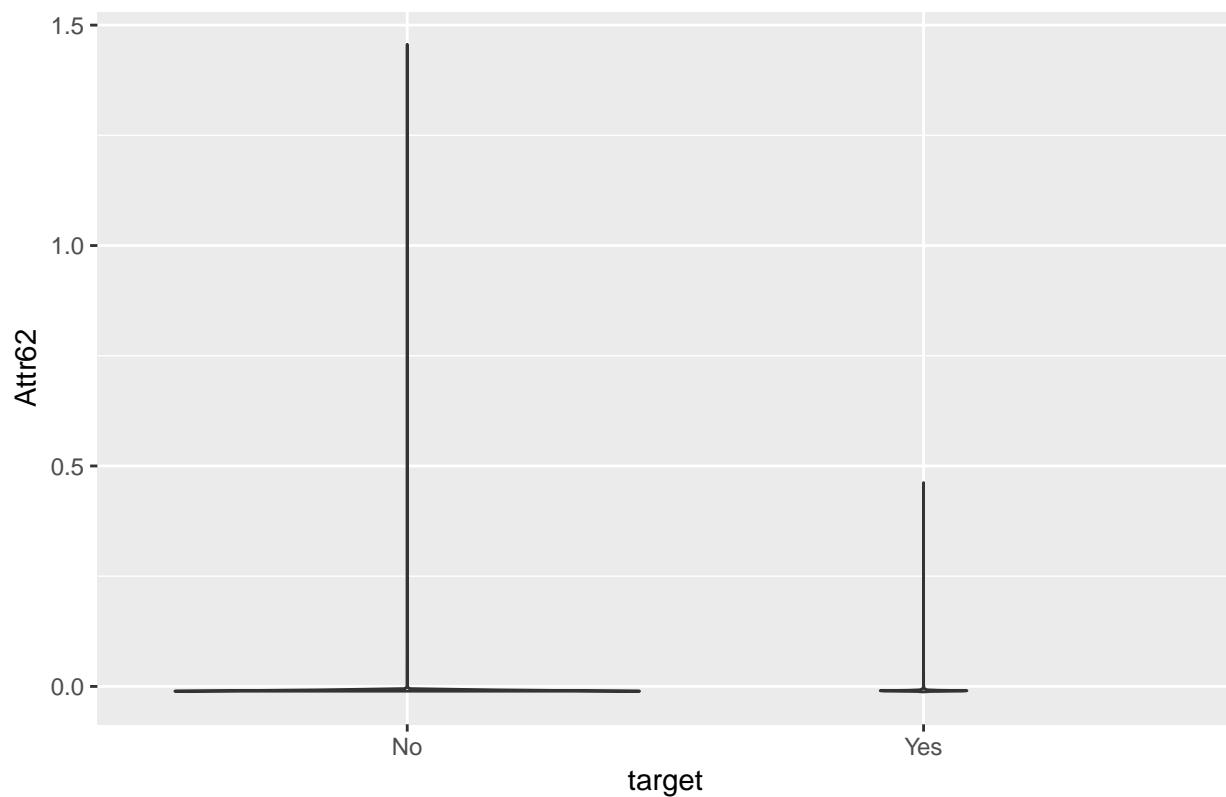
Bankruptcy Status by Attr60



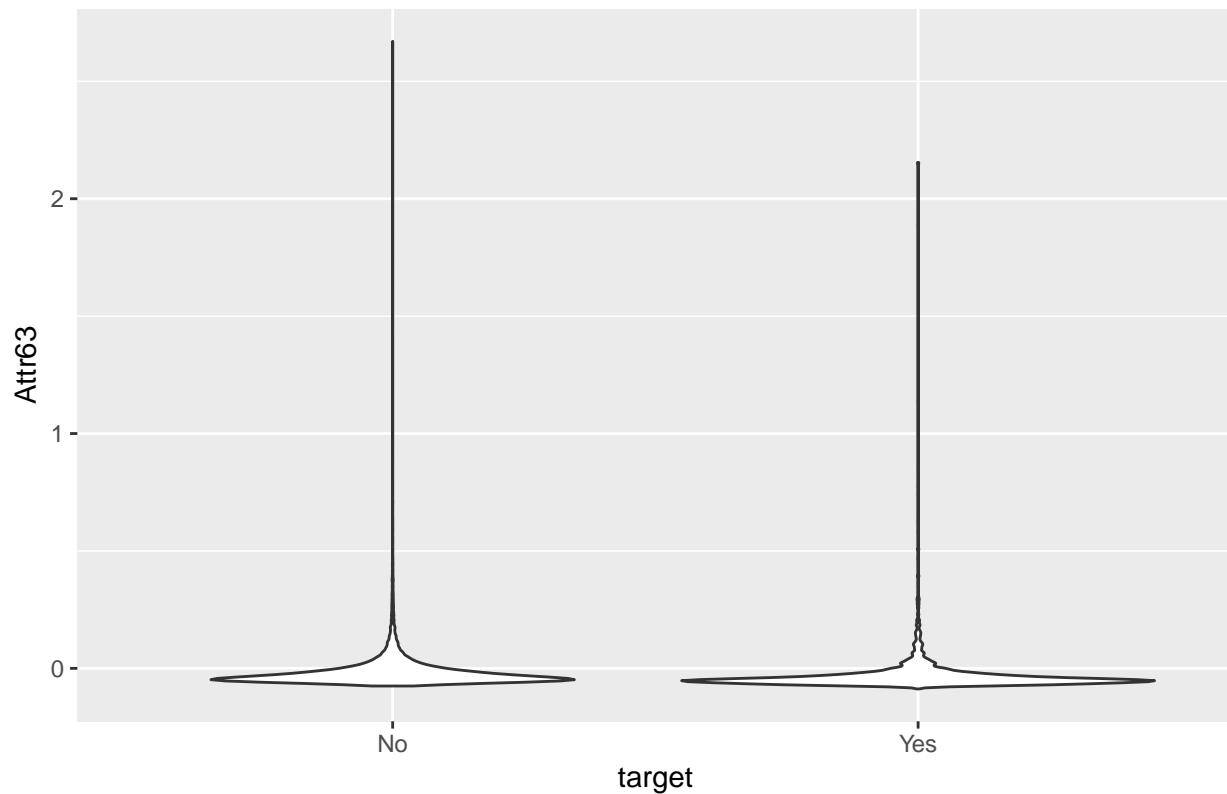
Bankruptcy Status by Attr61



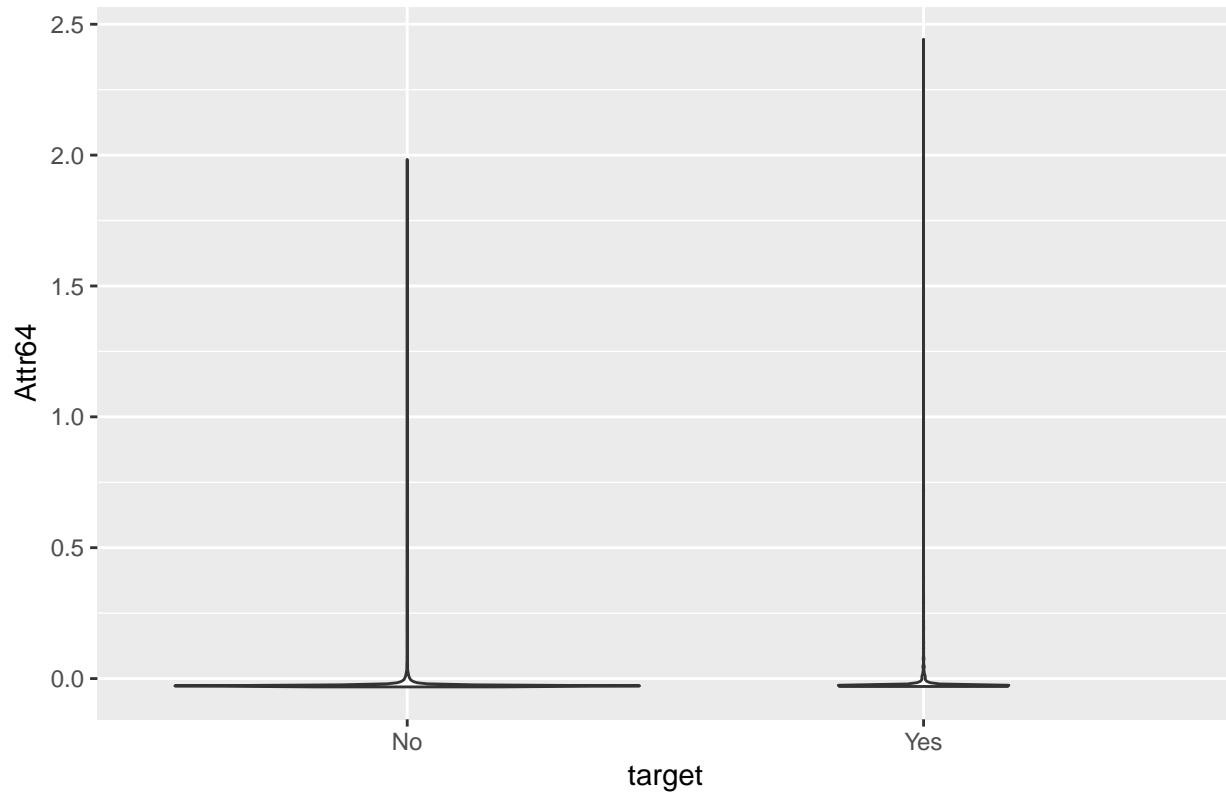
Bankruptcy Status by Attr62



Bankruptcy Status by Attr63



Bankruptcy Status by Attr64



* Conditional plots

```
str(df_final)
```

```
## 'data.frame': 41877 obs. of 56 variables:  
## $ Attr1 : num 0.0614 0.0759 0.0149 0.0535 0.0684 ...  
## $ Attr2 : num -0.01567 0.01773 -0.04847 0.00364 -0.016 ...  
## $ Attr3 : num 0.0656 0.0281 0.0632 0.0212 0.045 ...  
## $ Attr4 : num -0.0148 -0.0161 -0.0129 -0.0166 -0.0155 ...  
## $ Attr5 : num 0.00651 0.00625 0.00712 0.00615 0.00549 ...  
## $ Attr6 : num 0.00796 0.00796 0.02868 0.03386 0.00796 ...  
## $ Attr7 : num 0.02831 0.037262 -0.000924 0.015775 0.032379 ...  
## $ Attr8 : num -0.0226 -0.0237 -0.0208 -0.0233 -0.0226 ...  
## $ Attr9 : num -0.01449 -0.02095 -0.02513 -0.02349 -0.00987 ...  
## $ Attr10: num -0.00766 -0.02153 -0.00224 -0.01561 -0.00736 ...  
## $ Attr11: num 0.02389 0.03369 -0.0082 0.00983 0.02871 ...  
## $ Attr12: num -0.00898 -0.00714 -0.01216 -0.01172 -0.00795 ...  
## $ Attr13: num -0.00747 -0.00648 -0.00821 -0.00789 -0.00759 ...  
## $ Attr15: num -0.01363 -0.01243 -0.01116 -0.00892 -0.01422 ...  
## $ Attr16: num -0.0126 -0.0138 -0.0146 -0.0158 -0.0119 ...  
## $ Attr19: num -1.64e-05 1.71e-03 -1.67e-03 -7.63e-04 -2.17e-04 ...  
## $ Attr20: num -0.00415 -0.00454 -0.00436 -0.00497 -0.00363 ...  
## $ Attr22: num 0.025406 0.036158 -0.000727 0.0156 0.031907 ...  
## $ Attr23: num -0.000252 0.001128 -0.001495 -0.000407 -0.000411 ...  
## $ Attr24: num -0.0149 -0.0149 -0.0127 -0.0108 -0.0149 ...  
## $ Attr26: num -0.0123 -0.0133 -0.0135 -0.0141 -0.0117 ...  
## $ Attr27: num -0.0281 -0.0282 -0.0307 -0.0307 -0.0291 ...
```

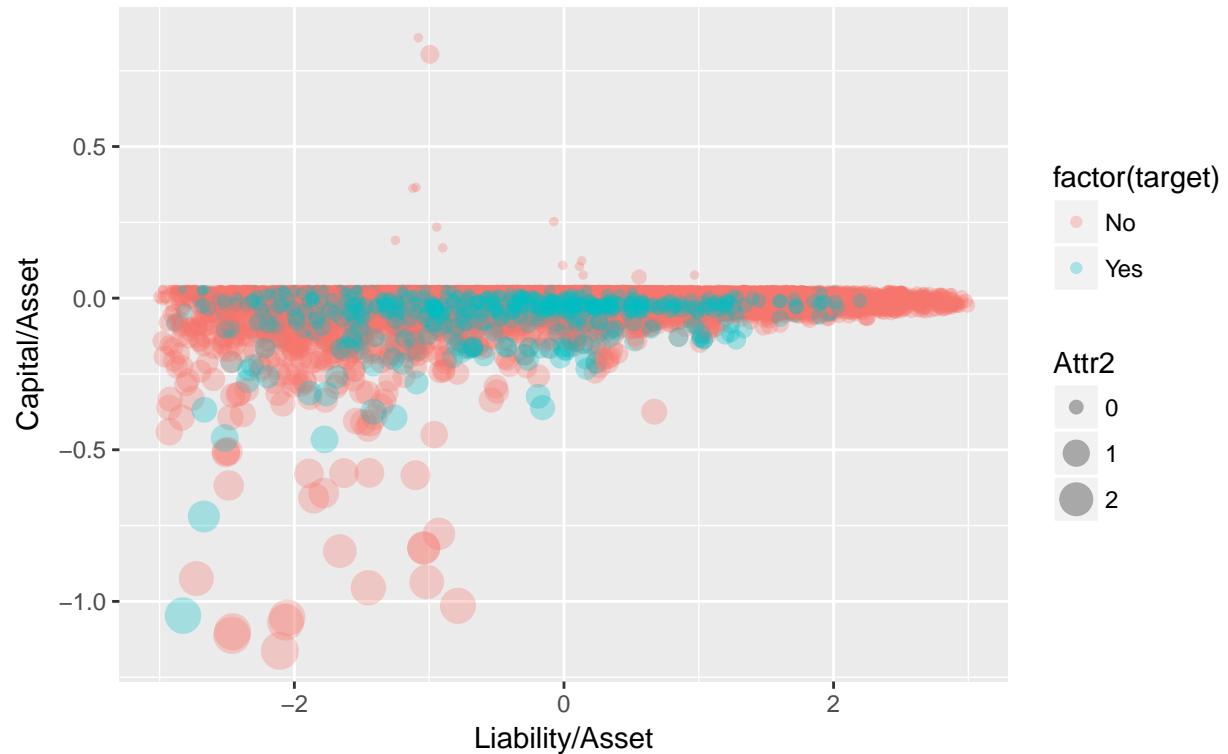
```

## $ Attr28: num  0.0719 -0.0321 -0.0262 -0.0317 -0.0244 ...
## $ Attr29: num  -0.377 0.455 0.79 0.173 0.396 ...
## $ Attr30: num  -0.00873 -0.00855 -0.00889 -0.00872 -0.00878 ...
## $ Attr31: num  -0.000439 0.001323 -0.002087 -0.001185 -0.000593 ...
## $ Attr32: num  -0.01002 -0.00944 -0.01035 -0.00976 -0.01023 ...
## $ Attr33: num  -0.0475 -0.0552 -0.039 -0.0517 -0.0427 ...
## $ Attr34: num  -0.0202 -0.033 -0.0417 -0.0421 -0.0166 ...
## $ Attr36: num  -0.0188 -0.0252 -0.0244 -0.02 -0.0142 ...
## $ Attr39: num  0.01051 0.01355 0.00969 0.01053 0.01096 ...
## $ Attr40: num  -0.0367 -0.0325 -0.0311 -0.033 -0.0368 ...
## $ Attr41: num  -0.00545 -0.00544 -0.00544 -0.00543 -0.00545 ...
## $ Attr42: num  0.0181 0.0234 0.0146 0.0167 0.0179 ...
## $ Attr43: num  -0.00596 -0.00619 -0.00609 -0.00635 -0.00631 ...
## $ Attr44: num  -0.00656 -0.00673 -0.00665 -0.0068 -0.0072 ...
## $ Attr45: num  -0.00566 -0.00548 -0.00576 -0.00557 -0.00571 ...
## $ Attr46: num  -0.0146 -0.015 -0.0131 -0.0151 -0.0166 ...
## $ Attr47: num  -0.00777 -0.00788 -0.00825 -0.0088 -0.00703 ...
## $ Attr48: num  0.0407 0.0515 0.0104 0.0308 0.0471 ...
## $ Attr49: num  0.0138 0.0157 0.0122 0.0133 0.0138 ...
## $ Attr50: num  -0.0127 -0.0155 -0.0109 -0.0148 -0.0135 ...
## $ Attr52: num  -0.00959 -0.00936 -0.00972 -0.00949 -0.00968 ...
## $ Attr53: num  -0.00464 -0.01843 -0.0174 -0.01787 -0.01685 ...
## $ Attr54: num  -0.00533 -0.01835 -0.01802 -0.0183 -0.01741 ...
## $ Attr55: num  -0.075522 -0.018103 0.184485 -0.062949 0.000201 ...
## $ Attr56: num  0.00496 0.00498 0.00495 0.00496 0.00496 ...
## $ Attr57: num  0.0313 0.0603 0.0111 0.036 0.0338 ...
## $ Attr58: num  -0.00549 -0.00551 -0.00547 -0.00549 -0.00549 ...
## $ Attr59: num  -0.01091 -0.00524 -0.01091 -0.00989 -0.01072 ...
## $ Attr60: num  -0.0134 -0.0134 -0.0134 -0.0134 -0.0135 ...
## $ Attr61: num  -0.0249 -0.0237 -0.0243 -0.023 -0.0146 ...
## $ Attr62: num  -0.01001 -0.00981 -0.01016 -0.00986 -0.01014 ...
## $ Attr63: num  -0.0478 -0.0533 -0.0412 -0.0521 -0.042 ...
## $ Attr64: num  -0.00441 -0.02809 -0.02794 -0.02695 -0.02508 ...
## $ target: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...

ggplot(df_final, aes(Attr29,Attr10)) + geom_point(aes(color = factor(target),
  size = Attr2), alpha = 0.3) + xlab('Liability/Asset') + ylab('Capital/Asset') +
  ggtitle('Relationship between Total Asset and Capital/Asset,
  target in colors and size as Liability/Asset')

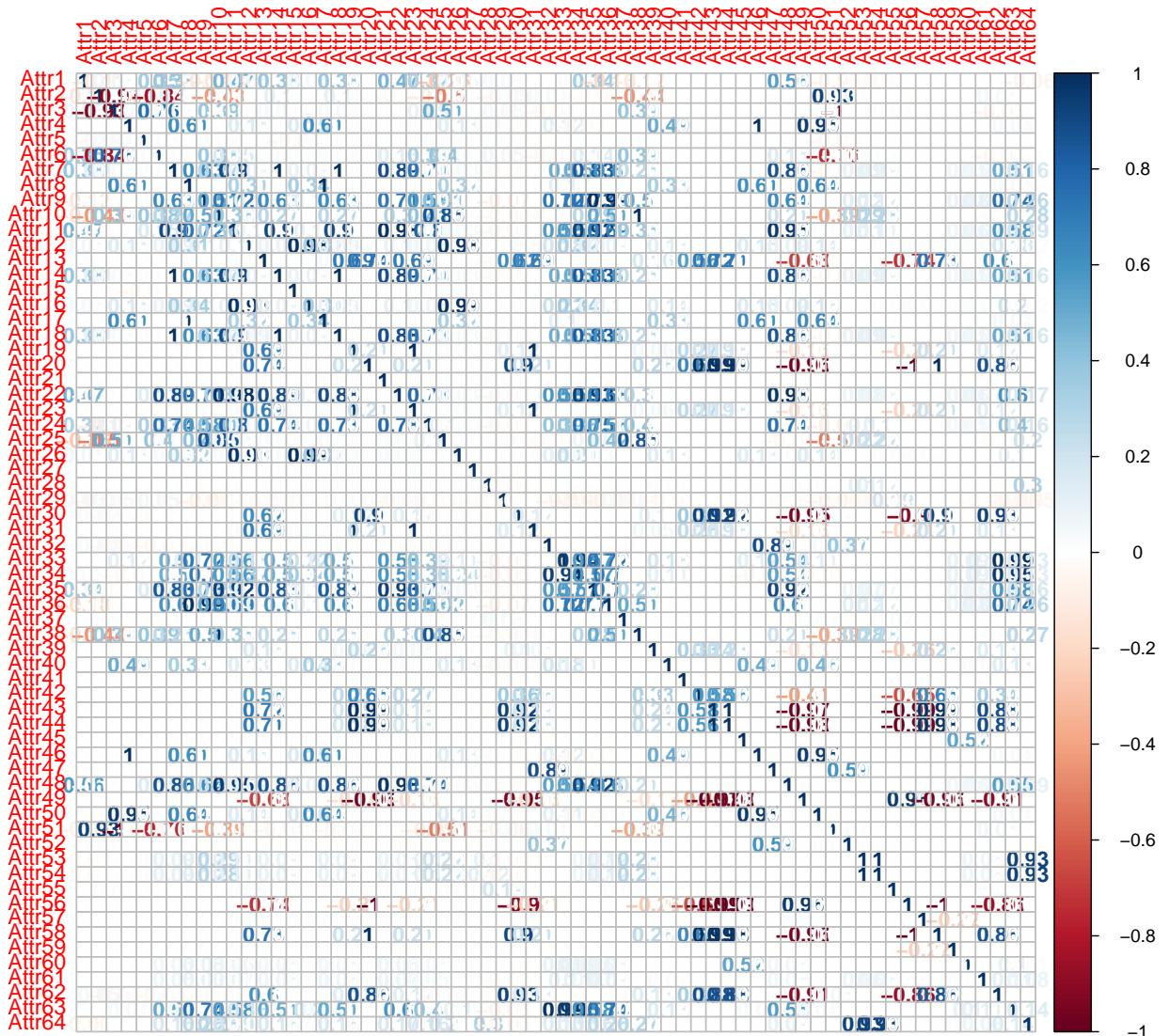
```

Relationship between Total Asset and Capital/Asset,
target in colors and size as Liability/Asset



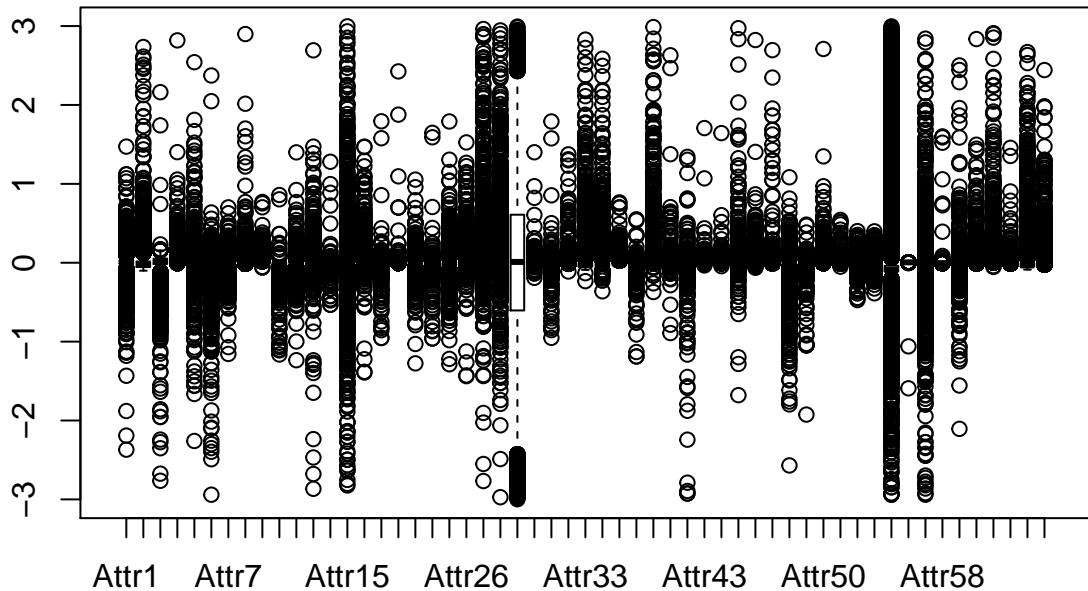
To observe the correlation between the attributes

```
corrplot(cor(data_num), method = "number")
```



- Numerical attributes box plots

```
boxplot(df_final[,-56])
```



```

# Data Partition
set.seed(1234)
train_rows = createDataPartition(df_final$target, p = 0.75, list = FALSE)
train_data <- df_final[train_rows,]
test_data <- df_final[-train_rows,]

sum(is.na(train_data))

## [1] 0
sum(is.na(test_data))

## [1] 0
prop.table(table(df_final$target))

##
##          No         Yes
## 0.95190677 0.04809323

prop.table(table(train_data$target))

##
##          No         Yes
## 0.95189277 0.04810723

prop.table(table(test_data$target))

##

```

```
##           No      Yes
## 0.9519488 0.0480512
```

Build model

- Build the Logistic Regression model using randomForest

```
fitRf <- randomForest(target ~ ., data=train_data, method="rf",
                       trControl=trainControl(method="cv", 10),
                       ntree=200)
print(fitRf)

##
## Call:
##   randomForest(formula = target ~ ., data = train_data, method = "rf",
##                 type of random forest: classification
##                 number of trees: 200
##   No. of variables tried at each split: 7
##
##       OOB estimate of error rate: 3.73%
## Confusion matrix:
##   No Yes class.error
## No 29807 91 0.003043682
## Yes 1082 429 0.716082065
```

Predictions on train data

- Predicting on train data
- Confusion matrix

```
pred_train <- predict(fitRf, train_data[, -56])
confusionMatrix(train_data$target, pred_train)

##
## Confusion Matrix and Statistics
##
##             Reference
## Prediction    No     Yes
##   No        29898     0
##   Yes        0     1511
##
##           Accuracy : 1
##           95% CI : (0.9999, 1)
##   No Information Rate : 0.9519
##   P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 1
##   Mcnemar's Test P-Value : NA
##
##           Sensitivity : 1.0000
##           Specificity : 1.0000
##   Pos Pred Value : 1.0000
##   Neg Pred Value : 1.0000
```

```

##           Prevalence : 0.9519
##           Detection Rate : 0.9519
##   Detection Prevalence : 0.9519
##           Balanced Accuracy : 1.0000
##
##           'Positive' Class : No
##

```

Predictions on train data

- Predicting on test data
- Confusion matrix

```

pred_test <- predict(fitRf, test_data[,-56])
confusionMatrix(test_data$target, pred_test)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No  Yes
##       No    9930   35
##       Yes    361  142
##
##           Accuracy : 0.9622
##           95% CI : (0.9583, 0.9657)
##   No Information Rate : 0.9831
##   P-Value [Acc > NIR] : 1
##
##           Kappa : 0.4027
##   Mcnemar's Test P-Value : <2e-16
##
##           Sensitivity : 0.9649
##           Specificity : 0.8023
##   Pos Pred Value : 0.9965
##   Neg Pred Value : 0.2823
##           Prevalence : 0.9831
##           Detection Rate : 0.9486
##   Detection Prevalence : 0.9519
##           Balanced Accuracy : 0.8836
##
##           'Positive' Class : No
##

```