Financial Analysis

Mingwei Wu

12/21/2020

Data Description

first_coloumn: Company represents sample companies

second_column: Time shows different time periods that data belongs to. Time series length varies between 1 to 14 for each company

third_column: Financial Distress is the target variable. if it is greater than 0.5 the company should be considered as healthy(0). otherwise, it would be regarded as financially distressed (1)

rest_column: the features denoted by x1 to x83. are some financial and non-financial characteristics of the sampled companies.

Goal

checking the dataframe which features are most indicative of financial distress?

what type of machine learning models perform best of this dataset.

```r
library(DataExplorer)
library(caret)
```

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

```
## Loading required package: ggplot2
```

```r
library(knitr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
## v purrr   0.3.4
```

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

```r
library(dplyr)
library(ggcorrplot)
library(gmodels)
library(ROSE)
```

```
## Loaded ROSE 0.0-3
```

```r
library(pROC)
```

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

```
##
## Attaching package: 'pROC'
```

```
## The following object is masked from 'package:gmodels':
##
##     ci
```

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

```r
library(plyr)
```

```
## ------------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## ------------------------------------------------------------------------------
```

```
##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

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

```r
library(randomForest)
```

```
## randomForest 4.6-14
```

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

```
##
## Attaching package: 'randomForest'
```

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

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

```r
library(caTools)
library(ggplot2) ## load the package in R
```

```r
data<-read.csv("Financial Distress.csv",header = TRUE)
head(data)
```

```
##   Company Time Financial.Distress      x1        x2      x3      x4       x5
## 1       1    1            0.0106360 1.2810   0.0229340 0.87454 1.21640  0.060940
## 2       1    2           -0.4559700 1.2700   0.0064542 0.82067 1.00490 -0.014080
## 3       1    3           -0.3253900 1.0529  -0.0593790 0.92242 0.72926  0.020476
## 4       1    4           -0.5665700 1.1131  -0.0152290 0.85888 0.80974  0.076037
## 5       2    1            1.3573000 1.0623   0.1070200 0.81460 0.83593  0.199960
## 6       2    2            0.0071875 1.0558   0.0819160 0.87949 0.68673  0.142630
##        x6      x7        x8        x9       x10     x11      x12      x13
## 1 0.188270 0.52510  0.018854  0.182790  0.0064489 0.85822  2.00580 0.125460
## 2 0.181040 0.62288  0.006423  0.035991  0.0017951 0.85152 -0.48644 0.179330
## 3 0.044865 0.43292 -0.081423 -0.765400 -0.0543240 0.89314  0.41220 0.077578
## 4 0.091033 0.67546 -0.018807 -0.107910 -0.0653160 0.89581  0.99490 0.141120
## 5 0.047800 0.74200  0.128030  0.577250  0.0940750 0.81549  3.01470 0.185400
## 6 0.043102 0.77198  0.119280  0.679730  0.0636050 0.81574  2.36460 0.120510
##      x14    x15       x16    x17       x18      x19       x20     x21    x22
## 1  6.9706 4.6512  0.050100 2.1984 0.0182650 0.024978 0.0272640 1.41730 9.5554
## 2  4.5764 3.7521 -0.014011 2.4575 0.0275580 0.028804 0.0411020 1.18010 7.2952
## 3 11.8900 2.4884  0.028077 1.3957 0.0125950 0.068116 0.0148470 0.81652 7.1204
## 4  6.0862 1.6382  0.093904 2.0588 0.0116010 0.094385 0.0144150 0.90391 7.9828
## 5  4.3938 1.6169  0.239210 3.0311 0.0068143 0.079346 0.0088763 1.02510 4.7463
## 6  7.2978 1.2609  0.207690 2.9089 0.0123070 0.087834 0.0159290 0.84185 3.8821
##        x23     x24      x25    x26     x27      x28      x29     x30    x31
## 1 0.148720 0.66995   214.760 12.641  6.4607 0.043835 0.204590 0.35179 8.3161
## 2 0.056026 0.67048    38.242 12.877  5.5506 0.265480 0.150190 0.41763 9.5276
## 3 0.065220 0.84827  -498.390 13.225 16.2540 0.416570 0.074149 0.36723 9.3513
## 4 0.125160 0.80478   -75.867 13.305  8.8950 0.083774 0.054098 0.54360 7.0909
## 5 0.266020 0.76770  1423.100 11.575 17.4880 0.620770 0.046907 0.56963 9.4861
## 6 0.239880 0.77264  1748.000 12.048 15.9330 0.604710 0.106850 0.59646 5.5914
```

```
##         x32      x33    x34      x35        x36     x37       x38       x39      x40
## 1 0.28922 0.76606 2.5825  77.400   0.0267220 1.63070 0.0150160 0.0054783 0.12730
## 2 0.41561 0.81699 2.6033  95.947   0.0075797 0.83754 0.0274250 0.0454340 0.13774
## 3 0.50356 0.91962 1.4931 144.670  -0.0664830 0.95579 0.0172700 0.0280590 0.10242
## 4 0.67133 0.93701 2.3533 219.750  -0.0170000 0.38335 0.0143270 0.2033700 0.10143
## 5 0.68143 0.94242 4.1296 222.650   0.1312300 0.25301 0.0081518 0.3530100 0.17612
## 6 0.86856 0.87851 3.8269 285.500   0.1004200 0.88660 0.0179220 0.3182900 0.17690
##        x41       x42     x43       x44     x45        x46     x47     x48      x49
## 1 9.6951  -0.73622 0.98559  0.180160 1.50060  0.0262240  7.0513 1174.90  5.3399
## 2 5.6035  -0.64385 1.30190  0.046857 1.00950  0.0078645  4.6022 1062.50  3.7389
## 3 9.4003 -14.03200 0.75746 -0.579760 0.57832 -0.0643730 11.9880  651.15 10.9340
## 4 5.7379   0.72205 1.39120 -0.150130 0.64508 -0.0177310  6.1114  703.04  5.7028
## 5 4.5088  -0.11339 1.05270  0.607660 0.25782  0.1313800  4.4151 2465.40  4.1408
## 6 5.6983   0.22828 0.68127  0.463080 0.35765  0.0931410  7.3356 2571.60  6.4112
##       x50    x51         x52      x53    x54       x55     x56      x57
## 1 0.85128 12.837  0.06173700 0.180900 209.87 -0.582550 0.47101 0.109900
## 2 0.94397 12.881 -0.00056494 0.056298 250.14 -0.474770 0.38599 0.369330
## 3 0.93478 12.909  0.04162500 0.047562 280.55 -1.000000 0.48844 0.053299
## 4 0.87484 13.094  0.10840000 0.101350 413.74  0.565000 0.34408 0.073356
## 5 0.73398 11.396  0.25031000 0.222370 315.34 -0.060101 0.20242 1.229100
## 6 0.76012 11.672  0.22394000 0.164730 412.79  0.102450 0.17945 0.611770
##          x58        x59     x60    x61     x62     x63     x64      x65     x66
## 1  0.0000000  0.0000000 0.22009 7.1241 15.3810  3.2702 17.8720  34.6920 30.087
## 2  0.0000000  0.0000000 0.00000 7.4166  7.1050 14.3210 18.7700 124.7600 26.124
## 3  0.0037854  0.0051907 0.00000 3.6373  7.0213  1.1538  9.8951   6.4467 30.245
## 4  0.0000366  0.0000453 0.00000 5.1442  9.9099  2.0408 -1.4903 -21.9070 34.285
## 5 -0.0024910 -0.0029800 0.22688 7.1241 15.3810  3.2702 17.8720  34.6920 30.087
## 6 -0.0112410 -0.0163690 0.22445 7.4166  7.1050 14.3210 18.7700 124.7600 26.124
##    x67    x68      x69 x70    x71 x72 x73     x74   x75    x76     x77 x78  x79
## 1 12.8 7991.4 364.9500 15.8 61.476   4  36  85.437 27.07 26.102 16.000  16  0.2
## 2 11.8 8322.8   0.1896 15.6 24.579   0  36 107.090 31.31 30.194 17.000  16  0.4
## 3 10.3 8747.0  11.9460 15.2 20.700   0  35 120.870 36.07 35.273 17.000  15 -0.2
## 4 11.5 9042.5 -18.7480 10.4 47.429   4  33  54.806 39.80 38.377 17.167  16  5.6
## 5 12.8 7991.4 364.9500 15.8 61.476   4  36  85.437 27.07 26.102 16.000  16  0.2
## 6 11.8 8322.8   0.1896 15.6 24.579   0  36 107.090 31.31 30.194 17.000  16  0.4
##   x80       x81 x82 x83
## 1  22  0.060390  30  49
## 2  22  0.010636  31  50
## 3  22 -0.455970  32  51
## 4  22 -0.325390  33  52
## 5  29  1.251000   7  27
## 6  29  1.357300   8  28
```

```
names(data)
```

```
##  [1] "Company"            "Time"              "Financial.Distress"
##  [4] "x1"                 "x2"                "x3"
##  [7] "x4"                 "x5"                "x6"
## [10] "x7"                 "x8"                "x9"
## [13] "x10"                "x11"               "x12"
## [16] "x13"                "x14"               "x15"
## [19] "x16"                "x17"               "x18"
## [22] "x19"                "x20"               "x21"
## [25] "x22"                "x23"               "x24"
```

```
## [28] "x25"              "x26"              "x27"
## [31] "x28"              "x29"              "x30"
## [34] "x31"              "x32"              "x33"
## [37] "x34"              "x35"              "x36"
## [40] "x37"              "x38"              "x39"
## [43] "x40"              "x41"              "x42"
## [46] "x43"              "x44"              "x45"
## [49] "x46"              "x47"              "x48"
## [52] "x49"              "x50"              "x51"
## [55] "x52"              "x53"              "x54"
## [58] "x55"              "x56"              "x57"
## [61] "x58"              "x59"              "x60"
## [64] "x61"              "x62"              "x63"
## [67] "x64"              "x65"              "x66"
## [70] "x67"              "x68"              "x69"
## [73] "x70"              "x71"              "x72"
## [76] "x73"              "x74"              "x75"
## [79] "x76"              "x77"              "x78"
## [82] "x79"              "x80"              "x81"
## [85] "x82"              "x83"
```

```r
names(data)[3] <- "financial_distress"
```

```r
head(data)
```

```
##   Company Time financial_distress      x1        x2      x3      x4        x5
## 1       1    1                   1  0.0106360 1.2810  0.0229340 0.87454 1.21640  0.060940
## 2       1    2                  -0.4559700 1.2700  0.0064542 0.82067 1.00490 -0.014080
## 3       1    3                  -0.3253900 1.0529 -0.0593790 0.92242 0.72926  0.020476
## 4       1    4                  -0.5665700 1.1131 -0.0152290 0.85888 0.80974  0.076037
## 5       2    1                   1.3573000 1.0623  0.1070200 0.81460 0.83593  0.199960
## 6       2    2                   0.0071875 1.0558  0.0819160 0.87949 0.68673  0.142630
##       x6      x7       x8         x9        x10     x11      x12      x13
## 1 0.188270 0.52510  0.018854  0.182790  0.0064489 0.85822  2.00580 0.125460
## 2 0.181040 0.62288  0.006423  0.035991  0.0017951 0.85152 -0.48644 0.179330
## 3 0.044865 0.43292 -0.081423 -0.765400 -0.0543240 0.89314  0.41220 0.077578
## 4 0.091033 0.67546 -0.018807 -0.107910 -0.0653160 0.89581  0.99490 0.141120
## 5 0.047800 0.74200  0.128030  0.577250  0.0940750 0.81549  3.01470 0.185400
## 6 0.043102 0.77198  0.119280  0.679730  0.0636050 0.81574  2.36460 0.120510
##       x14    x15      x16    x17       x18      x19       x20     x21    x22
## 1   6.9706 4.6512  0.050100 2.1984 0.0182650 0.024978 0.0272640 1.41730 9.5554
## 2   4.5764 3.7521 -0.014011 2.4575 0.0275580 0.028804 0.0411020 1.18010 7.2952
## 3  11.8900 2.4884  0.028077 1.3957 0.0125950 0.068116 0.0148470 0.81652 7.1204
## 4   6.0862 1.6382  0.093904 2.0588 0.0116010 0.094385 0.0144150 0.90391 7.9828
## 5   4.3938 1.6169  0.239210 3.0311 0.0068143 0.079346 0.0088763 1.02510 4.7463
## 6   7.2978 1.2609  0.207690 2.9089 0.0123070 0.087834 0.0159290 0.84185 3.8821
##       x23     x24      x25     x26     x27     x28       x29     x30    x31
## 1 0.148720 0.66995   214.760 12.641   6.4607 0.043835 0.204590 0.35179 8.3161
## 2 0.056026 0.67048    38.242 12.877   5.5506 0.265480 0.150190 0.41763 9.5276
## 3 0.065220 0.84827  -498.390 13.225  16.2540 0.416570 0.074149 0.36723 9.3513
## 4 0.125160 0.80478   -75.867 13.305   8.8950 0.083774 0.054098 0.54360 7.0909
## 5 0.266020 0.76770  1423.100 11.575  17.4880 0.620770 0.046907 0.56963 9.4861
## 6 0.239880 0.77264  1748.000 12.048  15.9330 0.604710 0.106850 0.59646 5.5914
```

```
##         x32      x33    x34      x35        x36     x37       x38       x39      x40
## 1 0.28922 0.76606 2.5825  77.400  0.0267220 1.63070 0.0150160 0.0054783 0.12730
## 2 0.41561 0.81699 2.6033  95.947  0.0075797 0.83754 0.0274250 0.0454340 0.13774
## 3 0.50356 0.91962 1.4931 144.670 -0.0664830 0.95579 0.0172700 0.0280590 0.10242
## 4 0.67133 0.93701 2.3533 219.750 -0.0170000 0.38335 0.0143270 0.2033700 0.10143
## 5 0.68143 0.94242 4.1296 222.650  0.1312300 0.25301 0.0081518 0.3530100 0.17612
## 6 0.86856 0.87851 3.8269 285.500  0.1004200 0.88660 0.0179220 0.3182900 0.17690
##       x41       x42     x43       x44     x45       x46     x47     x48     x49
## 1 9.6951  -0.73622 0.98559  0.180160 1.50060  0.0262240  7.0513 1174.90  5.3399
## 2 5.6035  -0.64385 1.30190  0.046857 1.00950  0.0078645  4.6022 1062.50  3.7389
## 3 9.4003 -14.03200 0.75746 -0.579760 0.57832 -0.0643730 11.9880  651.15 10.9340
## 4 5.7379   0.72205 1.39120 -0.150130 0.64508 -0.0177310  6.1114  703.04  5.7028
## 5 4.5088  -0.11339 1.05270  0.607660 0.25782  0.1313800  4.4151 2465.40  4.1408
## 6 5.6983   0.22828 0.68127  0.463080 0.35765  0.0931410  7.3356 2571.60  6.4112
##       x50    x51         x52      x53    x54       x55     x56      x57
## 1 0.85128 12.837  0.06173700 0.180900 209.87 -0.582550 0.47101 0.109900
## 2 0.94397 12.881 -0.00056494 0.056298 250.14 -0.474770 0.38599 0.369330
## 3 0.93478 12.909  0.04162500 0.047562 280.55 -1.000000 0.48844 0.053299
## 4 0.87484 13.094  0.10840000 0.101350 413.74  0.565000 0.34408 0.073356
## 5 0.73398 11.396  0.25031000 0.222370 315.34 -0.060101 0.20242 1.229100
## 6 0.76012 11.672  0.22394000 0.164730 412.79  0.102450 0.17945 0.611770
##          x58        x59     x60    x61     x62     x63     x64      x65    x66
## 1  0.0000000  0.0000000 0.22009 7.1241 15.3810  3.2702 17.8720  34.6920 30.087
## 2  0.0000000  0.0000000 0.00000 7.4166  7.1050 14.3210 18.7700 124.7600 26.124
## 3  0.0037854  0.0051907 0.00000 3.6373  7.0213  1.1538  9.8951   6.4467 30.245
## 4  0.0000366  0.0000453 0.00000 5.1442  9.9099  2.0408 -1.4903 -21.9070 34.285
## 5 -0.0024910 -0.0029800 0.22688 7.1241 15.3810  3.2702 17.8720  34.6920 30.087
## 6 -0.0112410 -0.0163690 0.22445 7.4166  7.1050 14.3210 18.7700 124.7600 26.124
##    x67    x68      x69 x70    x71 x72 x73     x74   x75    x76     x77 x78  x79
## 1 12.8 7991.4 364.9500 15.8 61.476   4  36  85.437 27.07 26.102 16.000  16  0.2
## 2 11.8 8322.8   0.1896 15.6 24.579   0  36 107.090 31.31 30.194 17.000  16  0.4
## 3 10.3 8747.0  11.9460 15.2 20.700   0  35 120.870 36.07 35.273 17.000  15 -0.2
## 4 11.5 9042.5 -18.7480 10.4 47.429   4  33  54.806 39.80 38.377 17.167  16  5.6
## 5 12.8 7991.4 364.9500 15.8 61.476   4  36  85.437 27.07 26.102 16.000  16  0.2
## 6 11.8 8322.8   0.1896 15.6 24.579   0  36 107.090 31.31 30.194 17.000  16  0.4
##   x80        x81 x82 x83
## 1  22   0.060390  30  49
## 2  22   0.010636  31  50
## 3  22  -0.455970  32  51
## 4  22  -0.325390  33  52
## 5  29   1.251000   7  27
## 6  29   1.357300   8  28
```

```r
data1<-data[complete.cases(data),]
dim(data1)
```
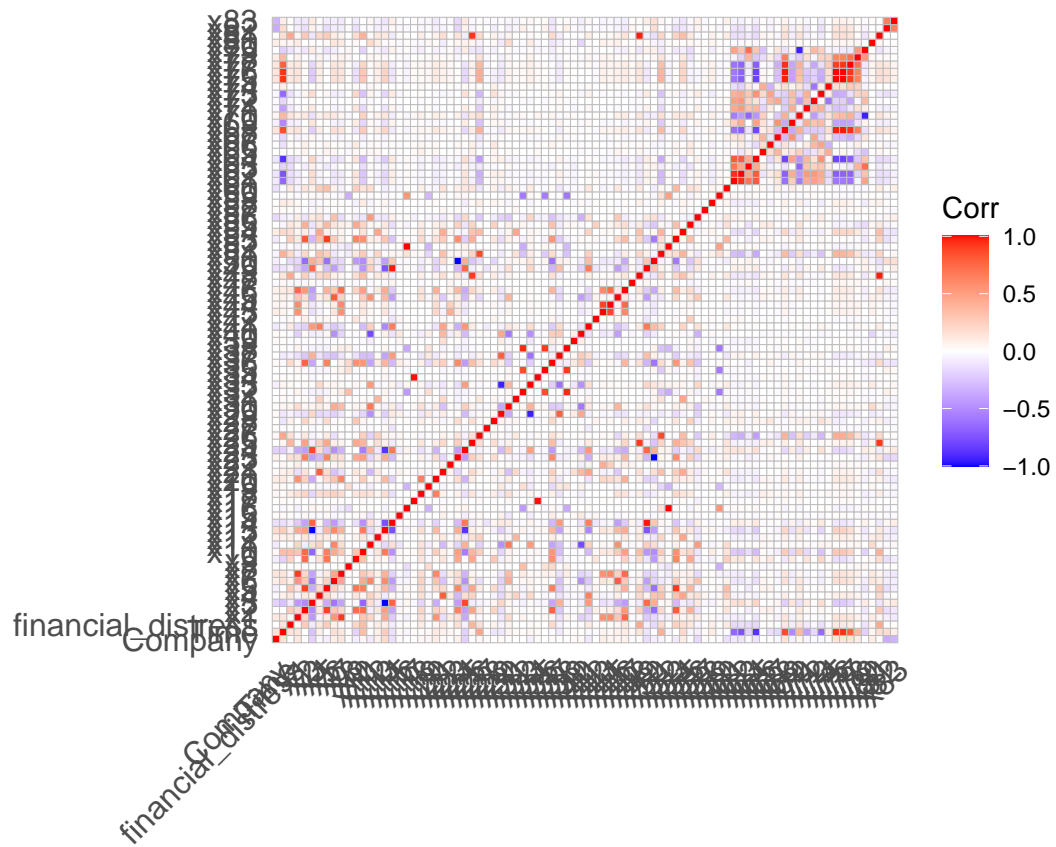
```
## [1] 3672    86
```

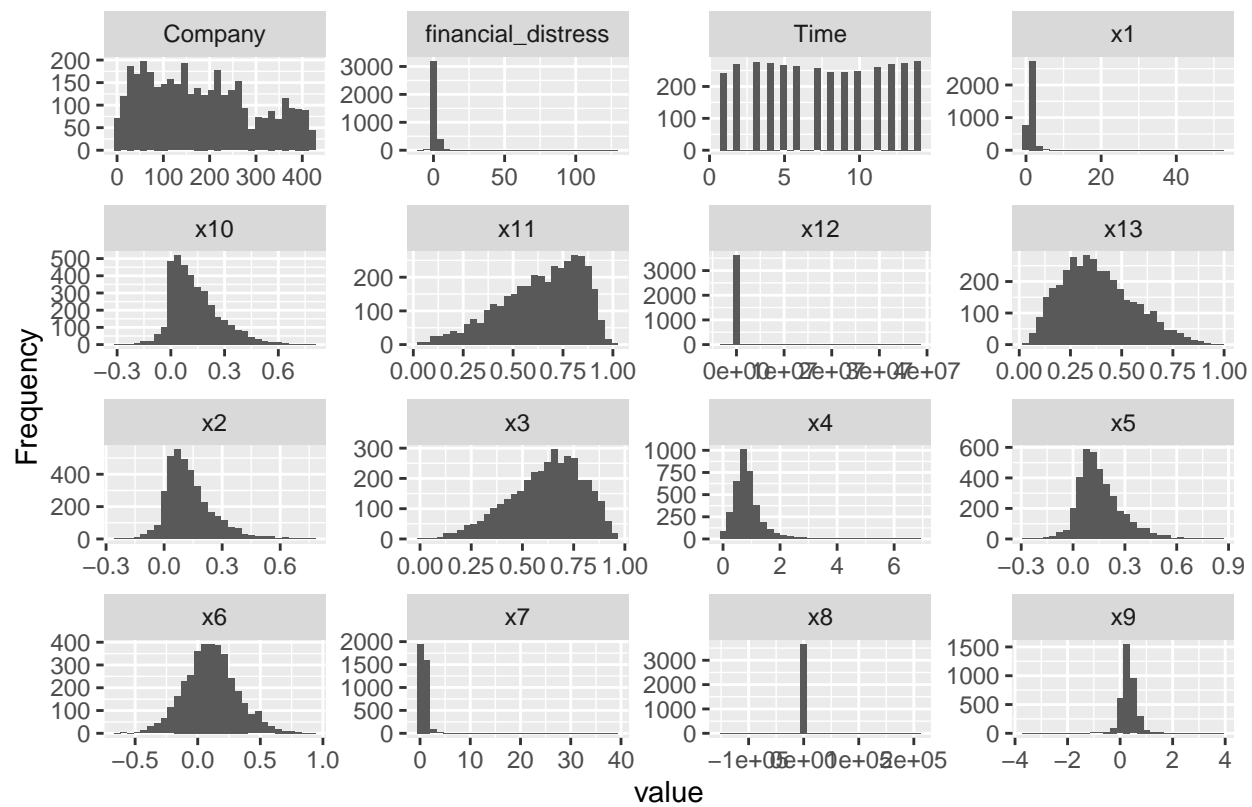**plot the relationship betweeen time and financial_distress**

```r
data1%>%
  ggplot(aes(Time,financial_distress))+geom_point(color="blue")
```
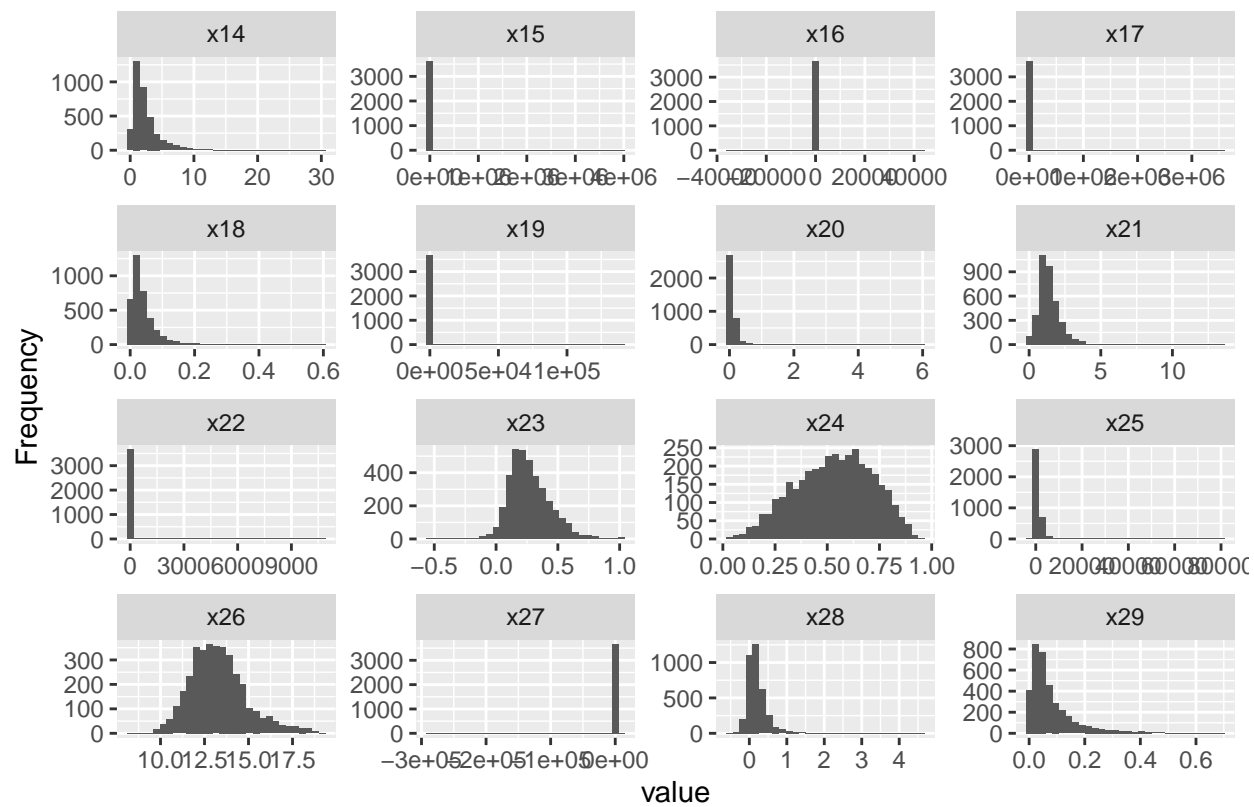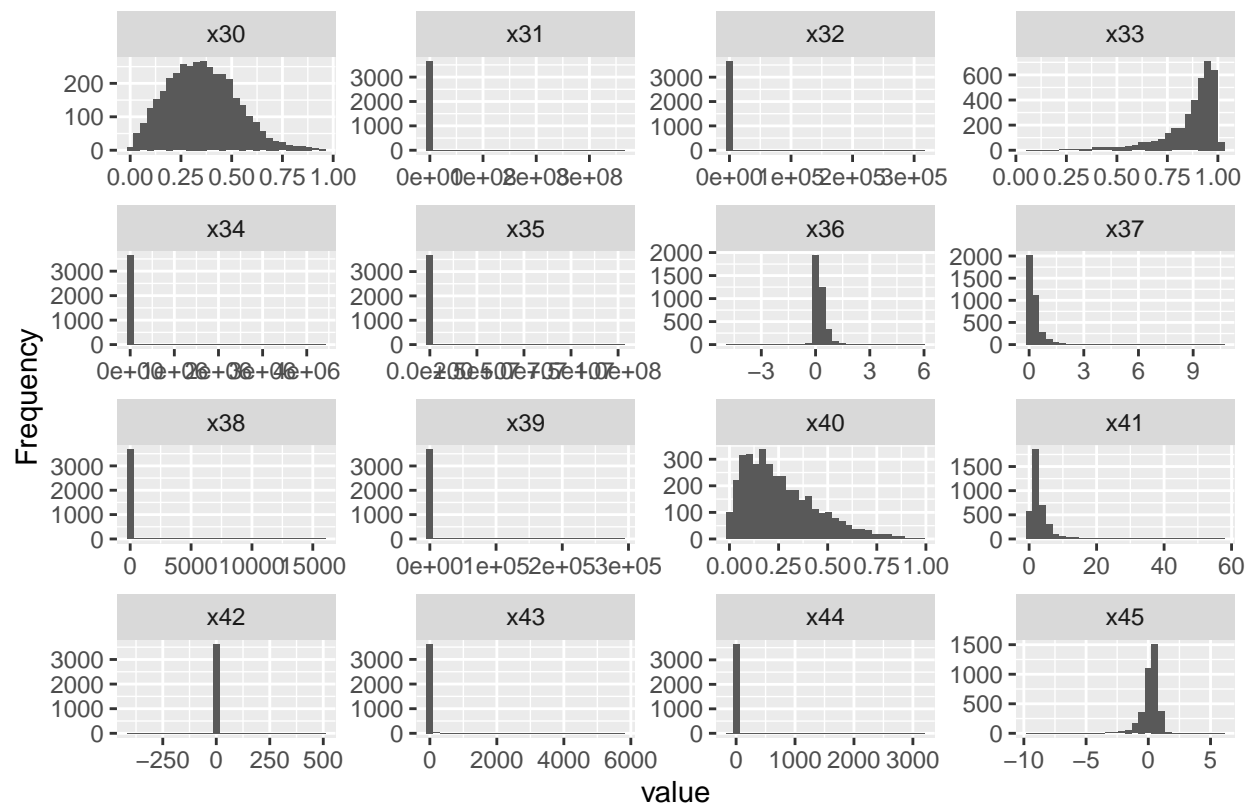
6
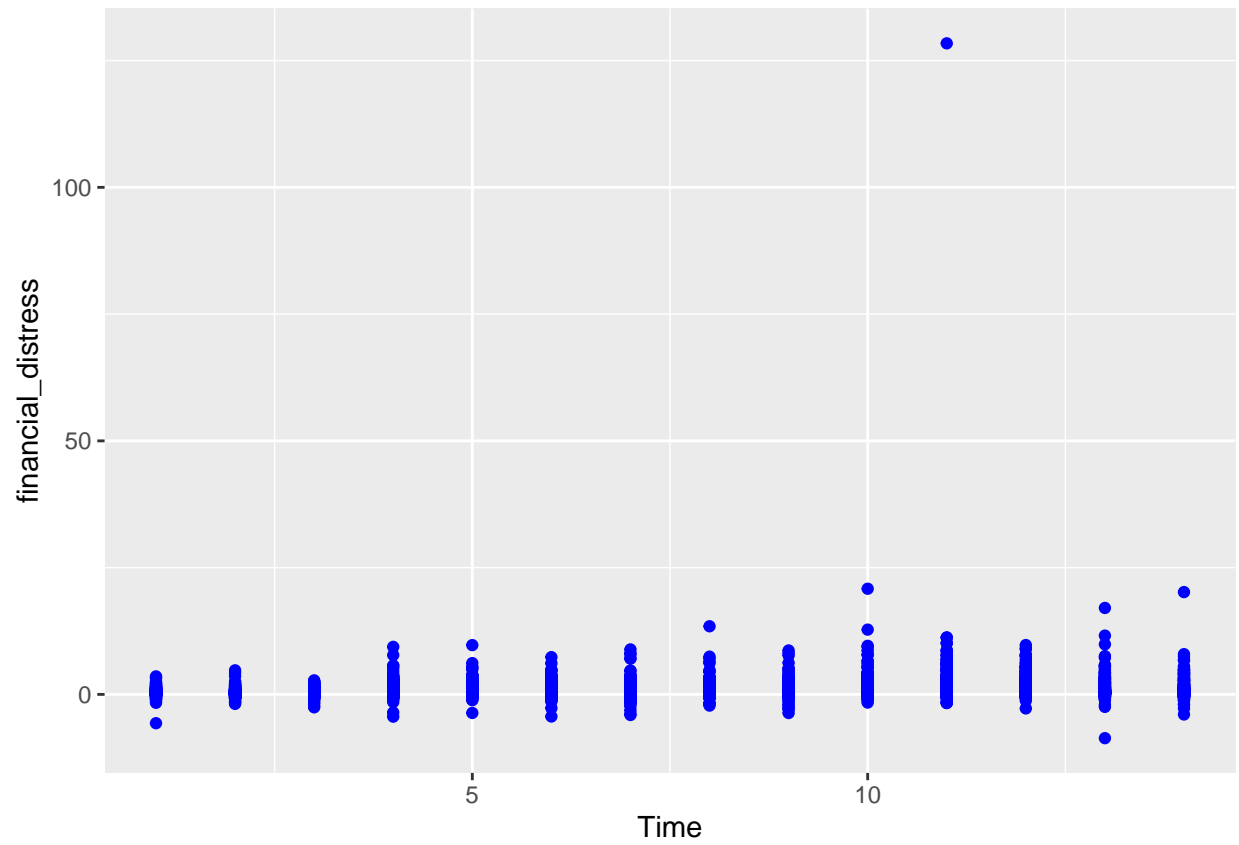
```
ggcorrplot(cor(data1))
```

```
plot_histogram(data1)
```

Frequency

value

Frequency

value

Frequency

value
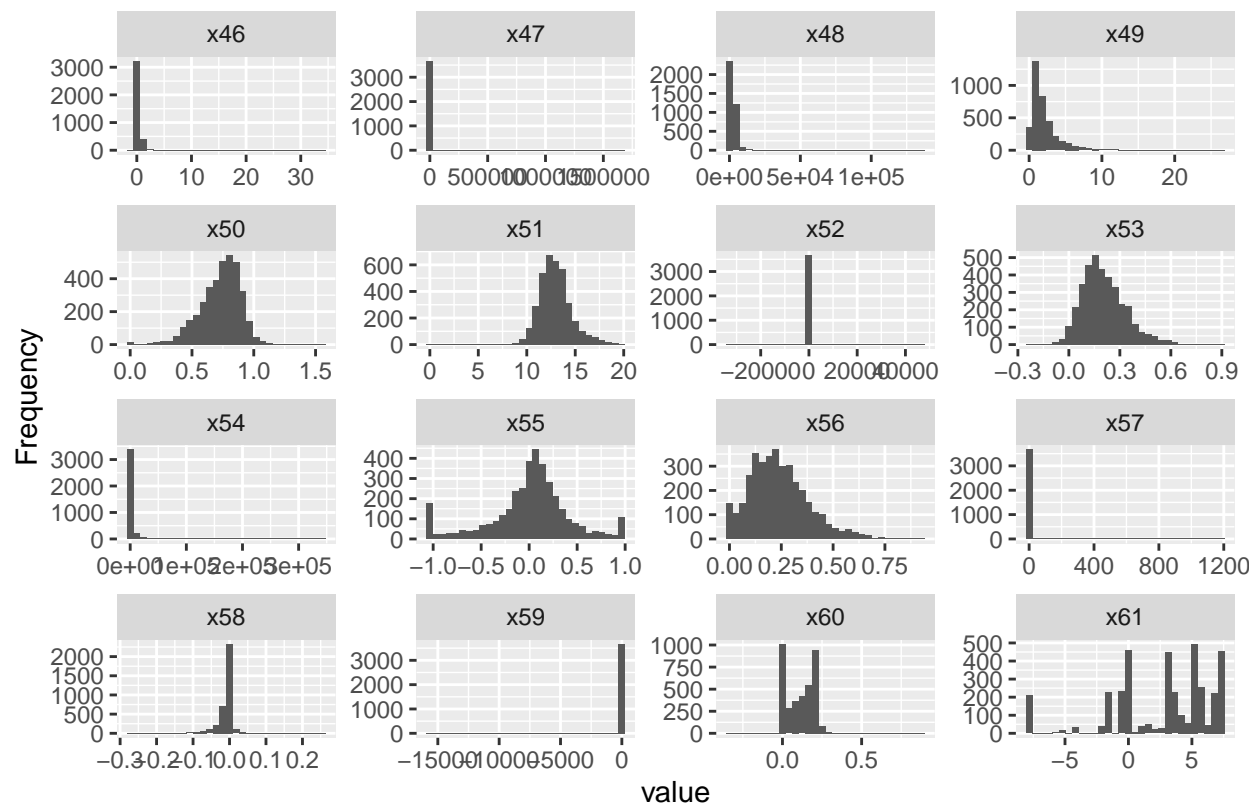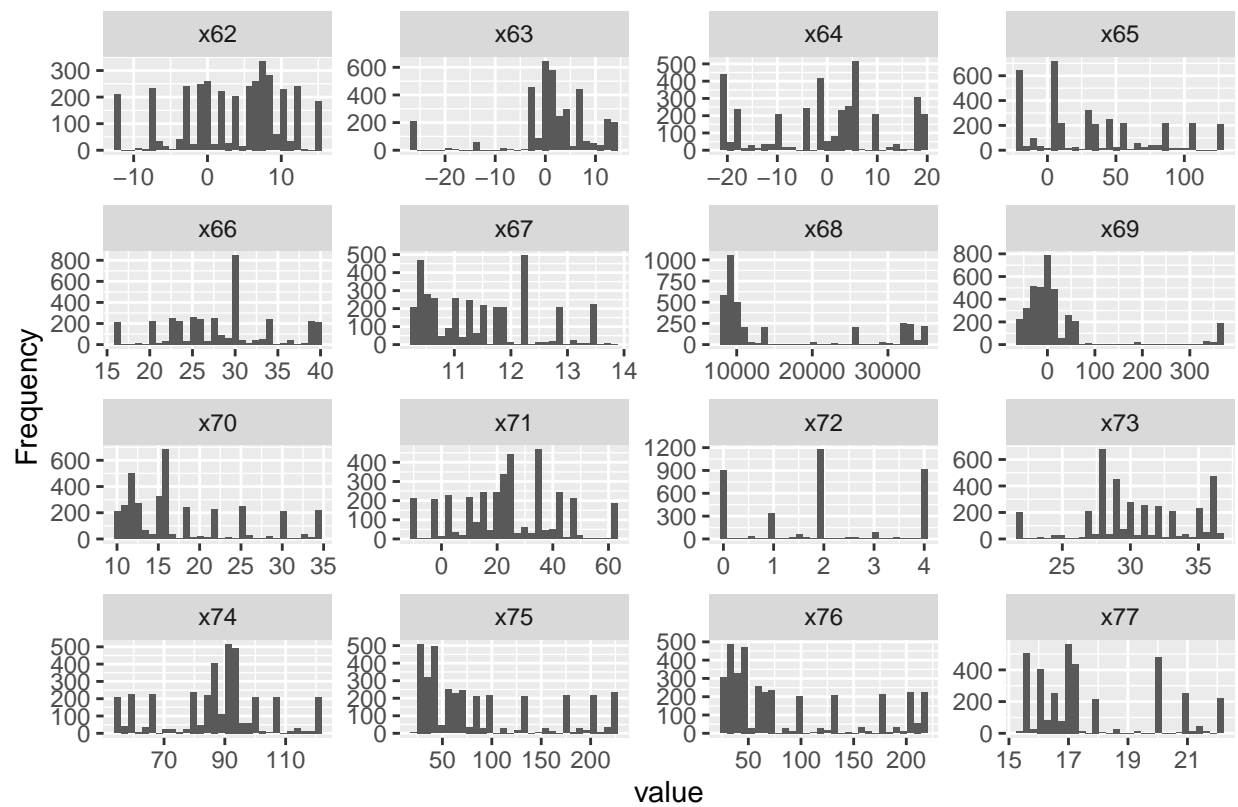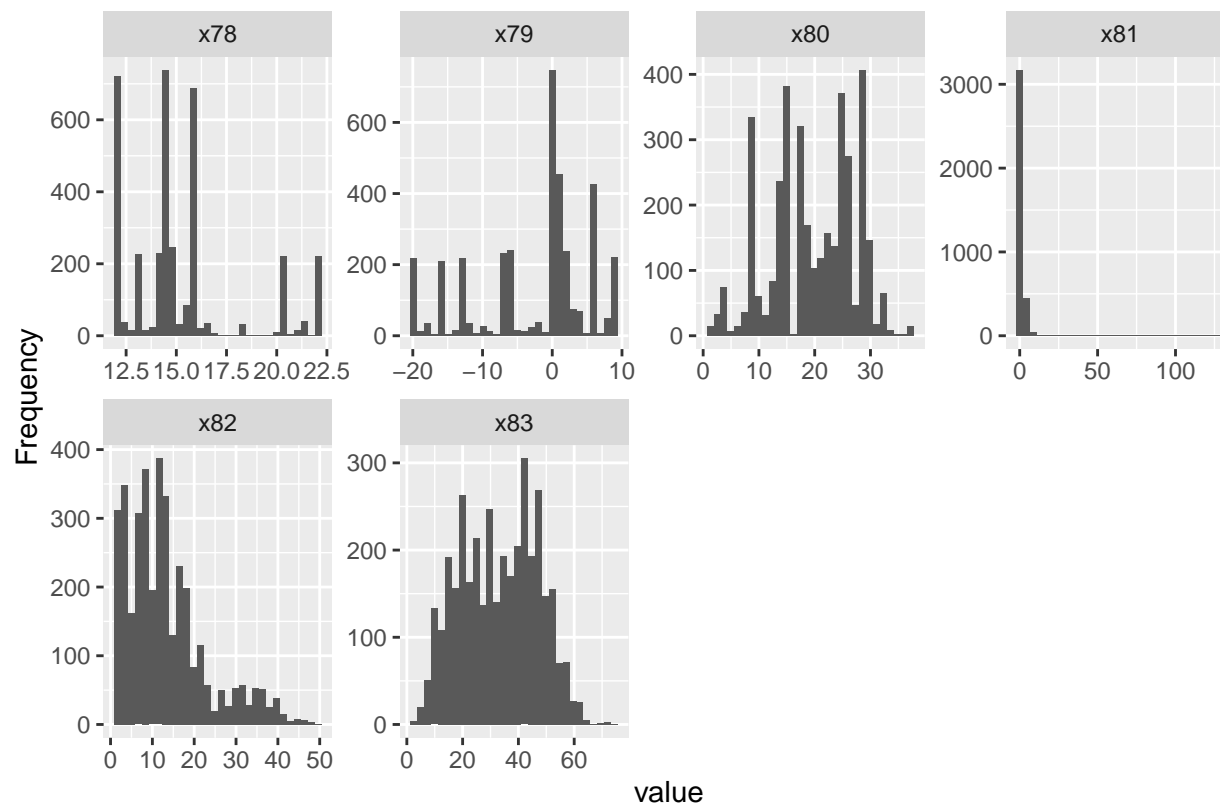
```r
summary(data1[1:3])
```

```
##     Company          Time          financial_distress
##  Min.   :  1.0   Min.   : 1.000   Min.   : -8.6317
##  1st Qu.: 80.0   1st Qu.: 4.000   1st Qu.:  0.1723
##  Median :168.0   Median : 7.000   Median :  0.5838
##  Mean   :182.1   Mean   : 7.528   Mean   :  1.0403
##  3rd Qu.:264.2   3rd Qu.:11.000   3rd Qu.:  1.3518
##  Max.   :422.0   Max.   :14.000   Max.   :128.4000
```

```r
summary(data1[4:86])
```

```
##        x1               x2                x3               x4
##  Min.   : 0.07517   Min.   :-0.2581   Min.   :0.01614   Min.   :0.000001
##  1st Qu.: 0.95214   1st Qu.: 0.0487   1st Qu.:0.50189   1st Qu.:0.552558
##  Median : 1.18360   Median : 0.1075   Median :0.63869   Median :0.775245
##  Mean   : 1.38782   Mean   : 0.1297   Mean   :0.61577   Mean   :0.868160
##  3rd Qu.: 1.50648   3rd Qu.: 0.1887   3rd Qu.:0.74943   3rd Qu.:1.039000
##  Max.   :51.95400   Max.   : 0.7494   Max.   :0.96790   Max.   :6.835600
##        x5               x6                x7               x8
##  Min.   :-0.2698   Min.   :-0.62775   Min.   : 0.03516   Min.   :-145000.00
##  1st Qu.: 0.0700   1st Qu.:-0.02775   1st Qu.: 0.43600   1st Qu.:      0.06
##  Median : 0.1318   Median : 0.10433   Median : 0.64187   Median :      0.14
##  Mean   : 0.1549   Mean   : 0.10672   Mean   : 0.78403   Mean   :     39.27
##  3rd Qu.: 0.2196   3rd Qu.: 0.23123   3rd Qu.: 0.89677   3rd Qu.:      0.27
```

14

```
## Max.    : 0.8585   Max.    : 0.92955   Max.    :38.83600   Max.    : 209000.00
##       x9               x10               x11               x12
## Min.    :-3.6112   Min.    :-0.31866   Min.    :0.02149   Min.    :-2620000
## 1st Qu.: 0.1577   1st Qu.: 0.03382   1st Qu.:0.50313   1st Qu.:       2
## Median : 0.3026   Median : 0.10727   Median :0.67085   Median :       5
## Mean    : 0.3326   Mean    : 0.13626   Mean    :0.63883   Mean    : 125273
## 3rd Qu.: 0.4840   3rd Qu.: 0.21002   3rd Qu.:0.80492   3rd Qu.:      15
## Max.    : 3.8102   Max.    : 0.76962   Max.    :0.99827   Max.    :38300000
##       x13              x14               x15               x16
## Min.    :0.0321   Min.    : 0.0164   Min.    :       0   Min.    :-35758.00
## 1st Qu.:0.2506   1st Qu.: 1.0075   1st Qu.:       2   1st Qu.:     0.09
## Median :0.3613   Median : 1.7677   Median :       4   Median :     0.18
## Mean    :0.3842   Mean    : 2.4383   Mean    :    8186   Mean    :   -25.04
## 3rd Qu.:0.4981   3rd Qu.: 2.9908   3rd Qu.:       8   3rd Qu.:     0.30
## Max.    :0.9839   Max.    :30.1520   Max.    :3960000   Max.    : 42180.00
##       x17              x18               x19               x20
## Min.    :       0   Min.    :0.0001016   Min.    :       0.00   Min.    :0.000127
## 1st Qu.:       2   1st Qu.:0.0132660   1st Qu.:       0.01   1st Qu.:0.024805
## Median :       2   Median :0.0283745   Median :       0.03   Median :0.053568
## Mean    :    2059   Mean    :0.0413796   Mean    :      77.55   Mean    :0.103999
## 3rd Qu.:       4   3rd Qu.:0.0522655   3rd Qu.:       0.07   3rd Qu.:0.108332
## Max.    :3540000   Max.    :0.5965500   Max.    :140000.00   Max.    :5.964600
##       x21              x22               x23               x24
## Min.    : 0.000003   Min.    :       0.000   Min.    :-0.5538   Min.    :0.01614
## 1st Qu.: 0.904230   1st Qu.:       1.926   1st Qu.: 0.1565   1st Qu.:0.40091
## Median : 1.262900   Median :       3.966   Median : 0.2476   Median :0.54250
## Mean    : 1.449663   Mean    :      14.195   Mean    : 0.2732   Mean    :0.53212
## 3rd Qu.: 1.741775   3rd Qu.:       7.780   3rd Qu.: 0.3687   3rd Qu.:0.67189
## Max.    :13.398000   Max.    :10721.000   Max.    : 1.0000   Max.    :0.93563
##       x25              x26               x27               x28
## Min.    :-3374.6   Min.    : 8.195   Min.    :-288000.00   Min.    :-0.49944
## 1st Qu.: 261.8   1st Qu.:12.163   1st Qu.:      -2.07   1st Qu.: 0.03521
## Median : 605.2   Median :13.139   Median :       2.94   Median : 0.15324
## Mean    : 996.0   Mean    :13.288   Mean    :     -77.48   Mean    : 0.20258
## 3rd Qu.: 1260.2   3rd Qu.:14.152   3rd Qu.:       6.91   3rd Qu.: 0.30075
## Max.    :79551.0   Max.    :19.106   Max.    :    9327.40   Max.    : 4.62540
##       x29              x30               x31               x32
## Min.    :0.00000   Min.    :0.002082   Min.    :       0   Min.    :     0.01
## 1st Qu.:0.02575   1st Qu.:0.227492   1st Qu.:       5   1st Qu.:     0.27
## Median :0.05271   Median :0.346125   Median :       9   Median :     0.43
## Mean    :0.08365   Mean    :0.354824   Mean    : 358087   Mean    :   379.87
## 3rd Qu.:0.10618   3rd Qu.:0.469145   3rd Qu.:      23   3rd Qu.:     0.65
## Max.    :0.69046   Max.    :0.948000   Max.    :361000000   Max.    :312000.00
##       x33              x34               x35               x36
## Min.    :0.04937   Min.    :       0   Min.    :       0   Min.    :-4.85620
## 1st Qu.:0.81875   1st Qu.:       2   1st Qu.:      42   1st Qu.: 0.07288
## Median :0.91109   Median :       4   Median :      92   Median : 0.16460
## Mean    :0.86314   Mean    :    2859   Mean    :   68878   Mean    : 0.24032
## 3rd Qu.:0.95702   3rd Qu.:       6   3rd Qu.:     156   3rd Qu.: 0.33656
## Max.    :1.00000   Max.    :4350000   Max.    :102000000   Max.    : 5.76610
##       x37              x38               x39               x40
## Min.    : 0.00000   Min.    :       0.000   Min.    :      -1.00   Min.    :0.0001702
## 1st Qu.: 0.06666   1st Qu.:       0.018   1st Qu.:      -0.01   1st Qu.:0.1141650
## Median : 0.15748   Median :       0.037   Median :       0.16   Median :0.2175350
```

```
##  Mean    : 0.31206   Mean    :   15.149   Mean    :  120.24   Mean   :0.2600971
##  3rd Qu.: 0.35431   3rd Qu.:    0.070   3rd Qu.:    0.34   3rd Qu.:0.3703675
##  Max.   :10.53600   Max.   :15772.000   Max.   :290000.00   Max.   :0.9777600
##       x41              x42                x43                x44
##  Min.   : 0.000   Min.   :-414.7100   Min.   :   0.089   Min.   : -56.719
##  1st Qu.: 1.318   1st Qu.:  -0.2685   1st Qu.:   0.949   1st Qu.:   0.188
##  Median : 2.157   Median :   0.0998   Median :   1.570   Median :   0.493
##  Mean   : 3.034   Mean   :   0.4273   Mean   :  11.527   Mean   :   3.321
##  3rd Qu.: 3.667   3rd Qu.:   0.5572   3rd Qu.:   2.942   3rd Qu.:   1.103
##  Max.   :56.913   Max.   : 482.6000   Max.   :5731.700   Max.   :3201.900
##       x45              x46              x47                x48
##  Min.   :-9.37690   Min.   :-1.07490   Min.   :      0.0   Min.   :   533.1
##  1st Qu.:-0.09719   1st Qu.: 0.06838   1st Qu.:      1.0   1st Qu.:  1418.9
##  Median : 0.28983   Median : 0.17498   Median :      1.8   Median :  1917.7
##  Mean   : 0.14924   Mean   : 0.29876   Mean   :   1490.9   Mean   :  2484.1
##  3rd Qu.: 0.59963   3rd Qu.: 0.35188   3rd Qu.:      3.1   3rd Qu.:  2812.5
##  Max.   : 6.05240   Max.   :34.06300   Max.   :1660000.0   Max.   :136000.0
##       x49              x50              x51              x52
##  Min.   : 0.0164   Min.   :0.0000   Min.   : 0.00   Min.   :-33563.00
##  1st Qu.: 0.8229   1st Qu.:0.6313   1st Qu.:11.84   1st Qu.:     0.11
##  Median : 1.4655   Median :0.7524   Median :12.80   Median :     0.21
##  Mean   : 2.1263   Mean   :0.7268   Mean   :12.93   Mean   :   -21.25
##  3rd Qu.: 2.6102   3rd Qu.:0.8435   3rd Qu.:13.80   3rd Qu.:     0.35
##  Max.   :26.3980   Max.   :1.5538   Max.   :19.81   Max.   : 46045.00
##       x53              x54              x55                x56
##  Min.   :-0.2252   Min.   :     0.0   Min.   :-1.000000   Min.   :0.0000
##  1st Qu.: 0.1168   1st Qu.:   443.7   1st Qu.:-0.154140   1st Qu.:0.1351
##  Median : 0.1867   Median :   962.6   Median : 0.047530   Median :0.2226
##  Mean   : 0.2073   Mean   :  3411.3   Mean   : 0.007702   Mean   :0.2374
##  3rd Qu.: 0.2833   3rd Qu.:  2176.0   3rd Qu.: 0.219235   3rd Qu.:0.3198
##  Max.   : 0.9046   Max.   :342000.0   Max.   : 1.000000   Max.   :0.9265
##       x57              x58                x59
##  Min.   :  -0.9987   Min.   :-0.279260   Min.   :-15649.000
##  1st Qu.:  -0.0411   1st Qu.:-0.012610   1st Qu.:    -0.017
##  Median :   0.0715   Median :-0.002939   Median :    -0.004
##  Mean   :   1.0271   Mean   :-0.010018   Mean   :    -6.869
##  3rd Qu.:   0.2734   3rd Qu.: 0.000000   3rd Qu.:     0.000
##  Max.   :1182.8000   Max.   : 0.247270   Max.   :     1.403
##       x60               x61              x62                x63
##  Min.   :-0.349060   Min.   :-7.71400   Min.   :-11.824   Min.   :-25.7360
##  1st Qu.: 0.005913   1st Qu.:-0.07535   1st Qu.: -0.711   1st Qu.: -0.5345
##  Median : 0.126035   Median : 3.63730   Median :  5.787   Median :  1.8883
##  Mean   : 0.113802   Mean   : 2.69474   Mean   :  3.456   Mean   :  1.2992
##  3rd Qu.: 0.196305   3rd Qu.: 5.72650   3rd Qu.:  8.480   3rd Qu.:  7.1562
##  Max.   : 0.863320   Max.   : 7.41660   Max.   : 15.381   Max.   : 14.3210
##       x64              x65               x66              x67
##  Min.   :-21.4110   Min.   :-21.907   Min.   :15.92   Min.   :10.30
##  1st Qu.: -9.3020   1st Qu.:  2.657   1st Qu.:23.91   1st Qu.:10.50
##  Median :  1.3445   Median : 28.286   Median :28.18   Median :11.30
##  Mean   : -0.7038   Mean   : 31.677   Mean   :28.25   Mean   :11.46
##  3rd Qu.:  5.8267   3rd Qu.: 57.368   3rd Qu.:30.25   3rd Qu.:12.20
##  Max.   : 18.7700   Max.   :124.760   Max.   :39.43   Max.   :13.85
##       x68            x69                x70              x71
##  Min.   : 7942   Min.   :-58.1220   Min.   :10.40   Min.   :-10.66
```

```
##  1st Qu.: 9042    1st Qu.:-25.8410    1st Qu.:11.90    1st Qu.: 10.13
##  Median : 9667    Median :  0.1896    Median :15.60    Median : 22.57
##  Mean   :15875    Mean   : 21.4724    Mean   :17.86    Mean   : 23.30
##  3rd Qu.:26059    3rd Qu.: 14.5290    3rd Qu.:21.50    3rd Qu.: 36.03
##  Max.   :34501    Max.   :364.9500    Max.   :34.70    Max.   : 61.48
##      x72              x73                x74                x75
##  Min.   :0.000    Min.   :22.00    Min.   : 54.81    Min.   : 24.32
##  1st Qu.:0.500    1st Qu.:28.00    1st Qu.: 79.95    1st Qu.: 39.80
##  Median :2.000    Median :30.00    Median : 90.00    Median : 66.12
##  Mean   :1.924    Mean   :30.41    Mean   : 86.84    Mean   : 91.92
##  3rd Qu.:4.000    3rd Qu.:33.00    3rd Qu.: 93.88    3rd Qu.:130.50
##  Max.   :4.000    Max.   :36.75    Max.   :120.87    Max.   :227.50
##      x76              x77              x78              x79
##  Min.   : 23.78    Min.   :15.25    Min.   :12.0    Min.   :-20.200
##  1st Qu.: 38.38    1st Qu.:16.00    1st Qu.:13.0    1st Qu.: -7.000
##  Median : 59.47    Median :17.00    Median :14.5    Median :  0.200
##  Mean   : 89.12    Mean   :17.78    Mean   :15.2    Mean   : -2.664
##  3rd Qu.:132.40    3rd Qu.:20.00    3rd Qu.:16.0    3rd Qu.:  2.100
##  Max.   :214.50    Max.   :22.00    Max.   :22.0    Max.   :  8.600
##      x80              x81                x82              x83
##  Min.   : 1.00    Min.   : -0.4992    Min.   : 1.00    Min.   : 2.00
##  1st Qu.:14.00    1st Qu.:  0.1899    1st Qu.: 6.00    1st Qu.:21.00
##  Median :20.00    Median :  0.5948    Median :11.00    Median :34.00
##  Mean   :19.71    Mean   :  1.1005    Mean   :13.12    Mean   :33.04
##  3rd Qu.:26.00    3rd Qu.:  1.3551    3rd Qu.:17.00    3rd Qu.:44.00
##  Max.   :37.00    Max.   :128.4000    Max.   :49.00    Max.   :74.00
```

```r
pairs(data1[1:3])
```
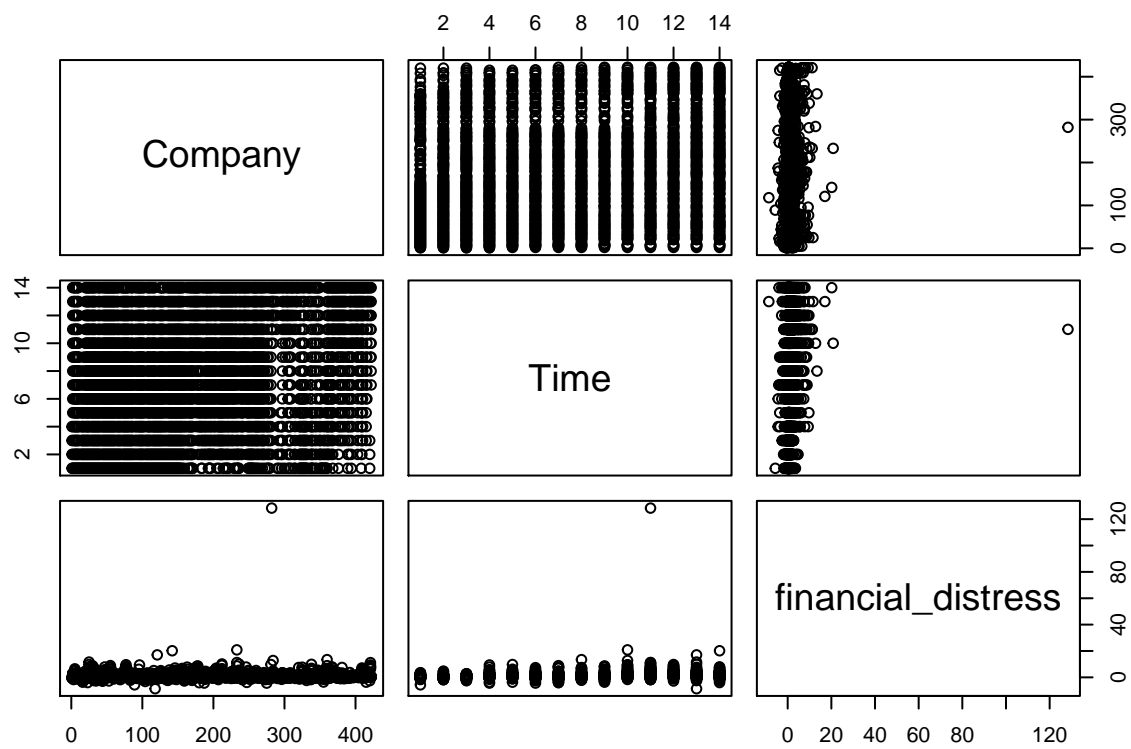
```
boxplot(data1[4:86]) # most outliers
```
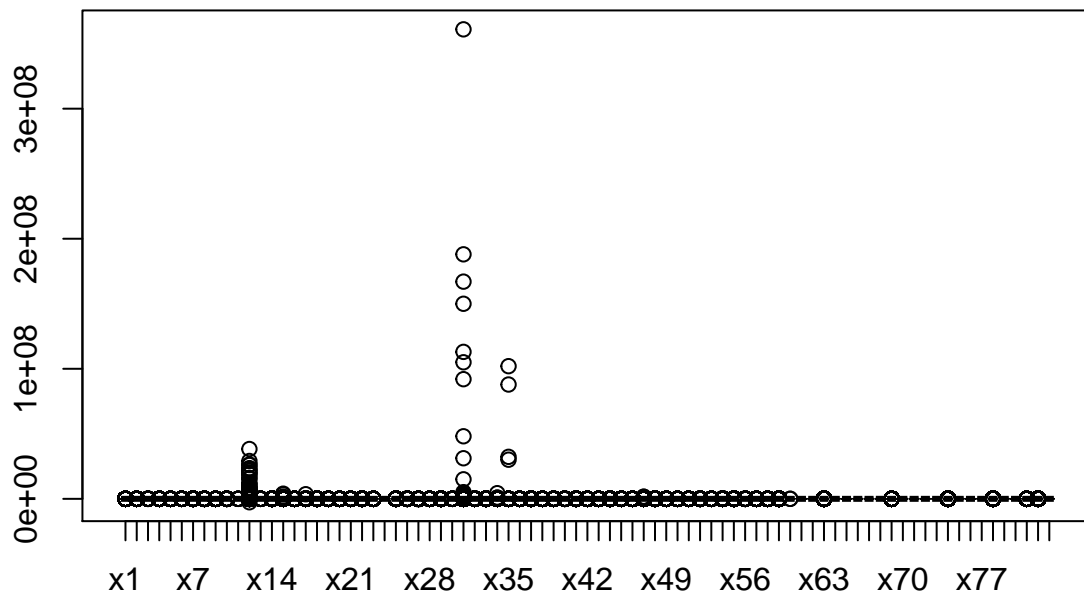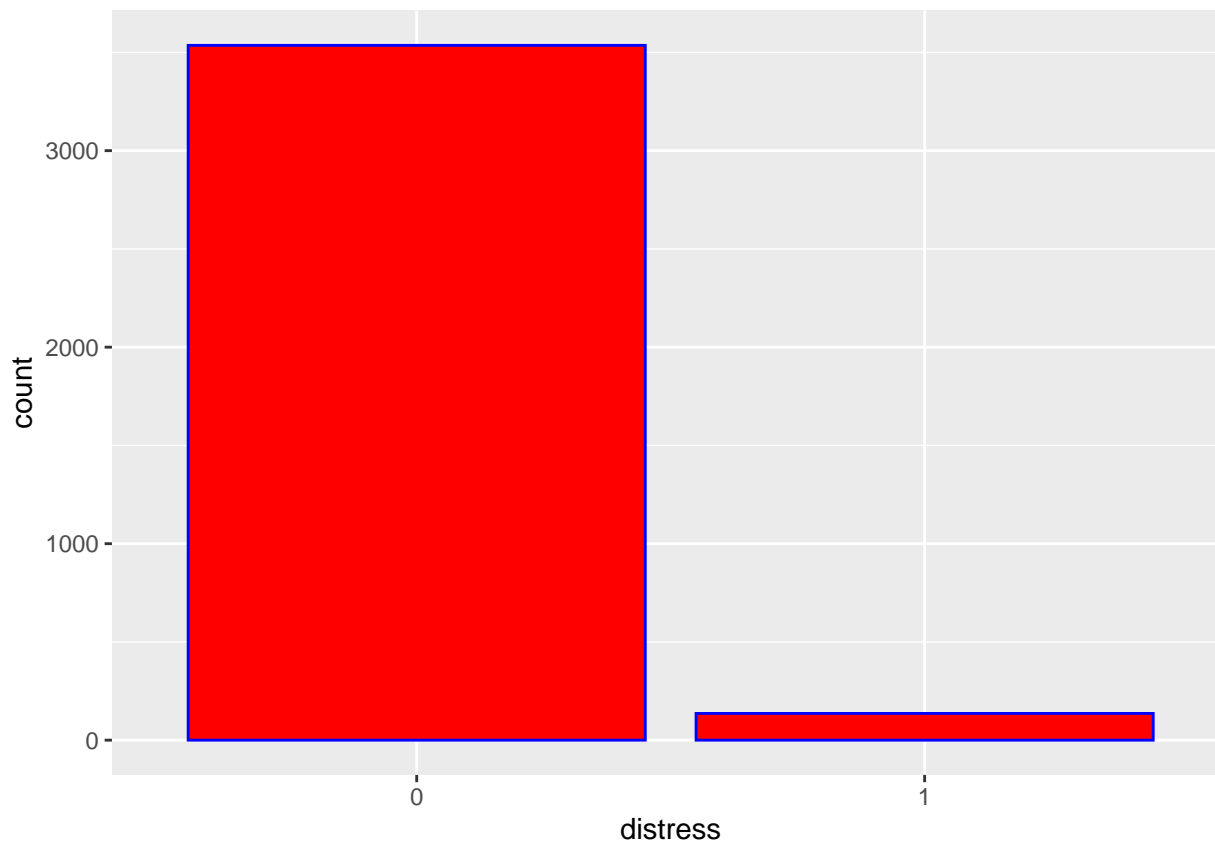
```
data1<-data1%>%
  mutate(distress = ifelse(financial_distress > -0.5 ,0,1))
data1%>%
  select(Company,Time,financial_distress,distress)%>%
  head()
```

```
##   Company Time financial_distress distress
## 1       1    1          0.0106360        0
## 2       1    2         -0.4559700        0
## 3       1    3         -0.3253900        0
## 4       1    4         -0.5665700        1
## 5       2    1          1.3573000        0
## 6       2    2          0.0071875        0
```

```
data1$distress<-as.factor(data1$distress)
```

```
data1%>%
  ggplot(aes(distress))+geom_bar(color="blue",
                                 fill="red")
```

```
glm<-glm(financial_distress~.,data=data1)
summary(glm)  #build up the logistic regression, and most column x is insignificant
```

```
##
## Call:
## glm(formula = financial_distress ~ ., data = data1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -34.470   -0.611   -0.068    0.477   82.003
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.295e+03  9.381e+03   0.351 0.725405
## Company      5.193e-04  4.215e-04   1.232 0.218028
## Time        -2.104e-01  2.322e-01  -0.906 0.365055
## x1           8.183e-02  9.150e-02   0.894 0.371206
## x2           8.053e+00  1.374e+00   5.861 5.00e-09 ***
## x3          -2.065e+04  8.603e+03  -2.401 0.016407 *
## x4          -1.530e+00  2.242e-01  -6.823 1.04e-11 ***
## x5           3.151e+00  9.634e-01   3.271 0.001083 **
## x6          -6.827e+03  7.735e+03  -0.883 0.377523
## x7           4.642e-02  1.431e-01   0.324 0.745662
## x8           1.627e-03  2.715e-03   0.599 0.549068
## x9           2.224e+00  2.263e-01   9.831  < 2e-16 ***
```

```
## x10           -4.654e-01  6.100e-01  -0.763 0.445573
## x11            6.829e+03  7.735e+03   0.883 0.377400
## x12            2.045e-07  2.890e-08   7.075 1.79e-12 ***
## x13            4.638e+02  1.763e+03   0.263 0.792446
## x14            1.466e+02  5.753e+02   0.255 0.798941
## x15            6.806e-07  3.175e-07   2.144 0.032104 *
## x16            7.840e-02  1.180e-01   0.664 0.506428
## x17            5.885e-06  3.904e-06   1.508 0.131759
## x18            2.676e+00  1.220e+00   2.194 0.028274 *
## x19            1.356e-03  2.191e-03   0.619 0.535945
## x20           -1.964e-01  3.153e-01  -0.623 0.533377
## x21           -9.442e-02  1.203e-01  -0.785 0.432744
## x22            8.027e-05  1.772e-04   0.453 0.650483
## x23           -1.636e+03  8.191e+03  -0.200 0.841709
## x24            1.429e+04  8.917e+03   1.603 0.109028
## x25           -2.238e-03  9.256e-05 -24.183  < 2e-16 ***
## x26            2.229e-01  1.137e-01   1.960 0.050067 .
## x27            1.081e-06  7.325e-06   0.148 0.882705
## x28            1.800e-01  1.351e-01   1.333 0.182676
## x29            2.112e+04  8.433e+03   2.504 0.012314 *
## x30           -1.326e+00  7.477e-01  -1.773 0.076293 .
## x31           -1.026e-08  4.925e-09  -2.084 0.037200 *
## x32           -2.147e-03  3.650e-03  -0.588 0.556555
## x33           -1.162e-02  1.203e+00  -0.010 0.992289
## x34           -4.774e-06  3.079e-06  -1.550 0.121148
## x35            1.663e-05  2.726e-05   0.610 0.541946
## x36            3.187e-01  2.913e-01   1.094 0.273943
## x37           -1.465e+02  5.753e+02  -0.255 0.799045
## x38           -1.360e-02  2.102e-02  -0.647 0.517565
## x39           -2.420e-07  7.496e-06  -0.032 0.974251
## x40           -8.905e-03  3.428e-01  -0.026 0.979275
## x41            3.742e-01  2.732e-02  13.700  < 2e-16 ***
## x42           -4.628e-03  2.537e-03  -1.825 0.068131 .
## x43            2.188e-04  6.497e-04   0.337 0.736339
## x44            4.157e-04  1.957e-03   0.212 0.831793
## x45           -1.452e-01  9.115e-02  -1.593 0.111241
## x46           -3.315e-01  1.600e-01  -2.072 0.038313 *
## x47           -3.759e-07  7.572e-07  -0.496 0.619600
## x48            6.540e-04  4.951e-05  13.211  < 2e-16 ***
## x49           -1.469e+02  5.753e+02  -0.255 0.798451
## x50           -1.633e+03  8.191e+03  -0.199 0.841969
## x51           -2.292e-01  1.112e-01  -2.060 0.039451 *
## x52           -9.247e-02  1.414e-01  -0.654 0.513027
## x53            3.096e+00  8.104e-01   3.820 0.000136 ***
## x54            7.962e-06  3.818e-06   2.085 0.037127 *
## x55           -2.719e-01  1.001e-01  -2.716 0.006640 **
## x56           -1.214e+00  7.283e-01  -1.667 0.095610 .
## x57            2.466e-03  1.343e-03   1.837 0.066350 .
## x58            2.146e+00  1.333e+00   1.610 0.107471
## x59            4.464e-02  7.446e-02   0.600 0.548872
## x60           -1.267e+00  4.810e-01  -2.635 0.008455 **
## x61           -3.226e+01  5.980e+01  -0.539 0.589625
## x62            8.076e+01  7.298e+01   1.107 0.268559
## x63           -1.623e+02  1.250e+02  -1.298 0.194210
```

```
## x64           7.696e+01  6.440e+01   1.195 0.232149
## x65           1.475e+01  9.494e+00   1.553 0.120442
## x66           6.957e+01  7.224e+01   0.963 0.335638
## x67          -1.099e+02  3.347e+02  -0.328 0.742651
## x68          -2.392e-01  2.832e-01  -0.845 0.398361
## x69          -3.192e+00  3.266e+00  -0.978 0.328364
## x70          -4.607e+02  5.182e+02  -0.889 0.374040
## x71           2.798e+01  3.154e+01   0.887 0.375031
## x72          -3.374e+02  5.221e+02  -0.646 0.518183
## x73          -7.216e+01  5.170e+01  -1.396 0.162866
## x74          -1.197e+01  1.536e+01  -0.779 0.435801
## x75           7.440e+01  4.772e+01   1.559 0.119057
## x76          -2.450e+01  5.267e+01  -0.465 0.641821
## x77          -1.340e+02  2.430e+02  -0.551 0.581484
## x78           5.792e+02  6.375e+02   0.909 0.363667
## x79          -3.837e+02  5.491e+02  -0.699 0.484792
## x80           4.565e-03  4.888e-03   0.934 0.350373
## x81           9.581e-01  6.816e-02  14.056  < 2e-16 ***
## x82           1.907e-03  5.191e-03   0.367 0.713305
## x83          -2.996e-04  3.415e-03  -0.088 0.930108
## distress1    -1.009e+00  2.013e-01  -5.012 5.66e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 4.193027)
##
##     Null deviance: 25823  on 3671  degrees of freedom
## Residual deviance: 15032  on 3585  degrees of freedom
## AIC: 15772
##
## Number of Fisher Scoring iterations: 2
```

```
glm1<-glm(financial_distress~Company+Time,data=data1)
summary(glm1) #removed the insignificant columns then make logistic regression
```

```
##
## Call:
## glm(formula = financial_distress ~ Company + Time, data = data1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -10.226   -0.799   -0.351    0.360  127.071
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.3992815  0.1070453   3.730 0.000194 ***
## Company      -0.0004507  0.0003744  -1.204 0.228733
## Time          0.0960421  0.0107800   8.909  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 6.889074)
##
##     Null deviance: 25823  on 3671  degrees of freedom
```

```
## Residual deviance: 25276  on 3669  degrees of freedom
## AIC: 17512
##
## Number of Fisher Scoring iterations: 2
```

```
glm1$coefficients
```

```
##   (Intercept)       Company          Time
##  0.3992815412 -0.0004506727  0.0960420632
```

```
set.seed(123)
sample<-createDataPartition(data1$distress,
                            p=0.7,
                            list=FALSE)
train<-data1[sample,]
test<-data1[-sample,] #split data into train and test by 70% and 30%
```

```
dim(train);dim(test)
```

```
## [1] 2572   87
```

```
## [1] 1100   87
```

```
balance<-ovun.sample(distress~.,data=train,
                     p=0.5,seed=1,
                     method="both")$data
summary(balance$distress)
```

```
##    0    1
## 1336 1236
```

```
rf1<-randomForest(as.factor(train$distress)~.,train,ntree=100)
rf1
```

```
##
## Call:
##  randomForest(formula = as.factor(train$distress) ~ ., data = train,     ntree = 100)
##               Type of random forest: classification
##                     Number of trees: 100
## No. of variables tried at each split: 9
##
##         OOB estimate of  error rate: 0.31%
## Confusion matrix:
##      0  1 class.error
## 0 2476  0  0.00000000
## 1    8 88  0.08333333
```

```
pred<-predict(rf1, test) #predictions
summary(pred)
```

```
##    0    1
## 1064   36
```

```
confusionMatrix(pred,as.factor(test$distress),positive="1") #matrix table
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 1060    4
##          1    0   36
##
##                Accuracy : 0.9964
##                  95% CI : (0.9907, 0.999)
##     No Information Rate : 0.9636
##     P-Value [Acc > NIR] : 2.738e-13
##
##                   Kappa : 0.9455
##
##  Mcnemar's Test P-Value : 0.1336
##
##             Sensitivity : 0.90000
##             Specificity : 1.00000
##          Pos Pred Value : 1.00000
##          Neg Pred Value : 0.99624
##              Prevalence : 0.03636
##          Detection Rate : 0.03273
##    Detection Prevalence : 0.03273
##       Balanced Accuracy : 0.95000
##
##         'Positive' Class : 1
##
```

```
roc<-predict(rf1,test,type="prob")
roc_all<-roc(test$distress, roc[,2])
```

```
## Setting levels: control = 0, case = 1
```

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

```
auc<-as.numeric(round(roc_all$auc,4))
plot(roc_all,lwd=2,col="blue")
text(0.6,0.7,"AUC=")
text(0.4,0.7,auc)
```