Teleco

Sangamesh

24 September 2018

PHASE 1: Reading the Dataset

PHASE 2 => Data Pre-Processing

Step 1: Checking for any Missing Data and Statistical Summary

```
## data
##
##  42  Variables      1000  Observations
## --------------------------------------------------------------------------
## region
##        n  missing distinct     Info     Mean      Gmd
##     1000        0        3    0.889    2.022   0.8888
##
## Value          1     2     3
## Frequency    322   334   344
## Proportion 0.322 0.334 0.344
## --------------------------------------------------------------------------
## tenure
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0       72        1    35.53    24.65        4        7
##      .25      .50      .75      .90      .95
##       17       34       54       66       70
##
## lowest :  1  2  3  4  5, highest: 68 69 70 71 72
## --------------------------------------------------------------------------
## age
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0       60    0.999    41.68    14.33       23       26
##      .25      .50      .75      .90      .95
##       32       40       51       59       64
##
## lowest : 18 19 20 21 22, highest: 73 74 75 76 77
## --------------------------------------------------------------------------
## marital
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2     0.75      495    0.495   0.5005
##
## --------------------------------------------------------------------------
## address
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0       50    0.998    11.55    10.92      0.0      1.0
##      .25      .50      .75      .90      .95
##      3.0      9.0     18.0     26.1     31.0
##
## lowest :  0  1  2  3  4, highest: 45 46 48 49 55
```

```
## ---------------------------------------------------------------------------
## income
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      218        1    77.53     75.3     18.0     21.0
##      .25      .50      .75      .90      .95
##     29.0     47.0     83.0    155.4    232.2
##
## lowest :    9   10   11   12   13, highest:  732  928  944 1131 1668
## ---------------------------------------------------------------------------
## ed
##        n  missing distinct     Info     Mean      Gmd
##     1000        0        5    0.946    2.671    1.369
##
## Value          1     2     3     4     5
## Frequency    204   287   209   234    66
## Proportion 0.204 0.287 0.209 0.234 0.066
## ---------------------------------------------------------------------------
## employ
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0       46    0.997    10.99    10.93        0        0
##      .25      .50      .75      .90      .95
##        3        8       17       25       31
##
## lowest :  0  1  2  3  4, highest: 41 43 44 45 47
## ---------------------------------------------------------------------------
## retire
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.134       47    0.047  0.08967
##
## ---------------------------------------------------------------------------
## gender
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.749      517    0.517   0.4999
##
## ---------------------------------------------------------------------------
## reside
##        n  missing distinct     Info     Mean      Gmd
##     1000        0        8    0.923    2.331     1.52
##
## Value          1     2     3     4     5     6     7     8
## Frequency    375   272   138   120    60    29     4     2
## Proportion 0.375 0.272 0.138 0.120 0.060 0.029 0.004 0.002
## ---------------------------------------------------------------------------
## tollfree
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.748      474    0.474   0.4991
##
## ---------------------------------------------------------------------------
## equip
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.711      386    0.386   0.4745
##
## ---------------------------------------------------------------------------
## callcard
```

```
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.655      678    0.678   0.4371
##
## ------------------------------------------------------------------------
## wireless
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.625      296    0.296   0.4172
##
## ------------------------------------------------------------------------
## longmon
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      425        1    11.72    9.559    2.998    3.645
##      .25      .50      .75      .90      .95
##    5.200    8.525   14.413   23.960   31.615
##
## lowest :  0.90  1.05  1.10  1.35  1.45, highest: 62.30 75.45 81.55 89.40 99.95
## ------------------------------------------------------------------------
## tollmon
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      162    0.855    13.27    16.92     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00     0.00    24.25    34.75    41.77
##
## lowest :   0.00   5.75   9.00   9.25   9.50, highest:  77.75  78.75  84.00  87.00 173.00
## ------------------------------------------------------------------------
## equipmon
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      310    0.769    14.22     19.2     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00     0.00    31.47    42.66    48.65
##
## lowest :  0.00 15.40 19.55 19.75 19.80, highest: 64.20 69.40 70.05 73.80 77.70
## ------------------------------------------------------------------------
## cardmon
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      176    0.967    13.78    14.67     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00    12.00    20.50    31.50    37.79
##
## lowest :   0.00   2.75   3.75   4.50   4.75, highest:  74.00  82.25  84.75  87.50 109.25
## ------------------------------------------------------------------------
## wiremon
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      263    0.651    11.58    17.76     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00     0.00    24.71    42.11    51.35
##
## lowest :   0.00  14.90  15.10  16.85  17.45, highest:  83.70  85.85  96.25 109.70 111.95
## ------------------------------------------------------------------------
## longten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      960        1    574.1      699    12.30    29.18
##      .25      .50      .75      .90      .95
##    90.14   285.48   755.02  1537.09  2144.79
```

```
##
## lowest :    0.90    1.05    1.10    1.35    1.45
## highest: 4333.00 5464.60 5988.50 6353.90 7257.60
## ---------------------------------------------------------------------------
## tollten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      473    0.854    551.3    823.1      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0      0.0    846.9   1838.4   2424.3
##
## lowest :    0.00    5.75   20.75   23.05   26.50
## highest: 4748.45 4905.85 4938.60 5850.25 5916.00
## ---------------------------------------------------------------------------
## equipten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      386    0.769    465.6    736.6      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0      0.0    579.5   1854.9   2435.6
##
## lowest :    0.00   15.40   17.25   22.60   22.80
## highest: 3925.50 4167.70 4758.05 4980.80 5028.65
## ---------------------------------------------------------------------------
## cardten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      338    0.967    605.8    775.2      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0    332.5    910.0   1505.5   2181.0
##
## lowest :    0.00    2.75    5.00   10.00   15.00
## highest: 4975.00 5115.00 5455.00 5520.00 7515.00
## ---------------------------------------------------------------------------
## wireten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      297    0.651    442.7    743.3      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0      0.0    316.5   1833.7   2727.6
##
## lowest :    0.00   10.45   16.85   20.95   24.20
## highest: 4399.60 4828.20 6132.20 6444.95 7856.85
## ---------------------------------------------------------------------------
## multline
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.748      475    0.475   0.4992
##
## ---------------------------------------------------------------------------
## voice
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.635      304    0.304   0.4236
##
## ---------------------------------------------------------------------------
## pager
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.579      261    0.261   0.3861
##
```

```
## ---------------------------------------------------------------------------
## internet
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.698      368    0.368   0.4656
##
## ---------------------------------------------------------------------------
## callid
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.749      481    0.481   0.4998
##
## ---------------------------------------------------------------------------
## callwait
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.749      485    0.485   0.5001
##
## ---------------------------------------------------------------------------
## forward
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2     0.75      493    0.493   0.5004
##
## ---------------------------------------------------------------------------
## confer
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2     0.75      502    0.502   0.5005
##
## ---------------------------------------------------------------------------
## ebill
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2      0.7      371    0.371   0.4672
##
## ---------------------------------------------------------------------------
## loglong
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      425        1    2.182    0.829    1.098    1.293
##      .25      .50      .75      .90      .95
##    1.649    2.143    2.668    3.176    3.454
##
## lowest : -0.105361  0.048790  0.095310  0.300105  0.371564
## highest:  4.131961  4.323470  4.401216  4.493121  4.604670
## ---------------------------------------------------------------------------
## logtoll
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      475      525      161        1     3.24   0.4583    2.621    2.757
##      .25      .50      .75      .90      .95
##    2.970    3.209    3.489    3.751    3.928
##
## lowest : 1.749200 2.197225 2.224624 2.251292 2.277267
## highest: 4.353499 4.366278 4.430817 4.465908 5.153292
## ---------------------------------------------------------------------------
## logequi
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      386      614      309        1    3.568   0.3162    3.121    3.201
##      .25      .50      .75      .90      .95
##    3.368    3.572    3.757    3.933    3.994
```

```
##
## lowest : 2.734368 2.972975 2.983153 2.985682 2.990720
## highest: 4.162003 4.239887 4.249209 4.301359 4.352855
## ---------------------------------------------------------------------------
## logcard
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      678      322      175        1    2.854   0.6274    1.946    2.169
##      .25      .50      .75      .90      .95
##    2.464    2.848    3.209    3.577    3.732
##
## lowest : 1.011601 1.321756 1.504077 1.558145 1.609438
## highest: 4.304065 4.409763 4.439706 4.471639 4.693639
## ---------------------------------------------------------------------------
## logwire
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      296      704      262        1    3.598    0.417    3.033    3.140
##      .25      .50      .75      .90      .95
##    3.334    3.595    3.862    4.083    4.192
##
## lowest : 2.701361 2.714695 2.824351 2.859340 2.873565
## highest: 4.427239 4.452602 4.566949 4.697749 4.718052
## ---------------------------------------------------------------------------
## lninc
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      218        1    3.957   0.8912    2.890    3.045
##      .25      .50      .75      .90      .95
##    3.367    3.850    4.419    5.046    5.448
##
## lowest : 2.197225 2.302585 2.397895 2.484907 2.564949
## highest: 6.595781 6.833032 6.850126 7.030857 7.419381
## ---------------------------------------------------------------------------
## custcat
##        n  missing distinct     Info     Mean      Gmd
##     1000        0        4    0.936    2.487    1.252
##
## Value          1     2     3     4
## Frequency    266   217   281   236
## Proportion 0.266 0.217 0.281 0.236
## ---------------------------------------------------------------------------
## churn
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.597      274    0.274   0.3982
##
## ---------------------------------------------------------------------------
```

So, we find missing data in 4 columns namely logtoll, logequi, logcard, logwire.

Since all the data that have missing values are of [log of the user], we will replace the missing values with mean of the data.

```
## data
##
##  42  Variables      1000  Observations
## ---------------------------------------------------------------------------
## region
```

```
##        n  missing distinct     Info     Mean      Gmd
##     1000        0        3    0.889    2.022   0.8888
##
## Value          1     2     3
## Frequency    322   334   344
## Proportion 0.322 0.334 0.344
## ---------------------------------------------------------------------------
## tenure
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0       72        1    35.53    24.65        4        7
##      .25      .50      .75      .90      .95
##       17       34       54       66       70
##
## lowest :  1  2  3  4  5, highest: 68 69 70 71 72
## ---------------------------------------------------------------------------
## age
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0       60    0.999    41.68    14.33       23       26
##      .25      .50      .75      .90      .95
##       32       40       51       59       64
##
## lowest : 18 19 20 21 22, highest: 73 74 75 76 77
## ---------------------------------------------------------------------------
## marital
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2     0.75      495    0.495   0.5005
##
## ---------------------------------------------------------------------------
## address
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0       50    0.998    11.55    10.92      0.0      1.0
##      .25      .50      .75      .90      .95
##      3.0      9.0     18.0     26.1     31.0
##
## lowest :  0  1  2  3  4, highest: 45 46 48 49 55
## ---------------------------------------------------------------------------
## income
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      218        1    77.53     75.3     18.0     21.0
##      .25      .50      .75      .90      .95
##     29.0     47.0     83.0    155.4    232.2
##
## lowest :    9   10   11   12   13, highest:  732  928  944 1131 1668
## ---------------------------------------------------------------------------
## ed
##        n  missing distinct     Info     Mean      Gmd
##     1000        0        5    0.946    2.671    1.369
##
## Value          1     2     3     4     5
## Frequency    204   287   209   234    66
## Proportion 0.204 0.287 0.209 0.234 0.066
## ---------------------------------------------------------------------------
## employ
##        n  missing distinct     Info     Mean      Gmd      .05      .10
```

```
##     1000        0       46    0.997    10.99    10.93        0        0
##      .25      .50      .75      .90      .95
##        3        8       17       25       31
##
## lowest :  0  1  2  3  4, highest: 41 43 44 45 47
## ------------------------------------------------------------------------
## retire
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.134       47    0.047  0.08967
##
## ------------------------------------------------------------------------
## gender
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.749      517    0.517   0.4999
##
## ------------------------------------------------------------------------
## reside
##        n  missing distinct     Info     Mean      Gmd
##     1000        0        8    0.923    2.331     1.52
##
## Value          1     2     3     4     5     6     7     8
## Frequency    375   272   138   120    60    29     4     2
## Proportion 0.375 0.272 0.138 0.120 0.060 0.029 0.004 0.002
## ------------------------------------------------------------------------
## tollfree
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.748      474    0.474   0.4991
##
## ------------------------------------------------------------------------
## equip
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.711      386    0.386   0.4745
##
## ------------------------------------------------------------------------
## callcard
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.655      678    0.678   0.4371
##
## ------------------------------------------------------------------------
## wireless
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.625      296    0.296   0.4172
##
## ------------------------------------------------------------------------
## longmon
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      425        1    11.72    9.559    2.998    3.645
##      .25      .50      .75      .90      .95
##    5.200    8.525   14.413   23.960   31.615
##
## lowest :  0.90  1.05  1.10  1.35  1.45, highest: 62.30 75.45 81.55 89.40 99.95
## ------------------------------------------------------------------------
## tollmon
##        n  missing distinct     Info     Mean      Gmd      .05      .10
```

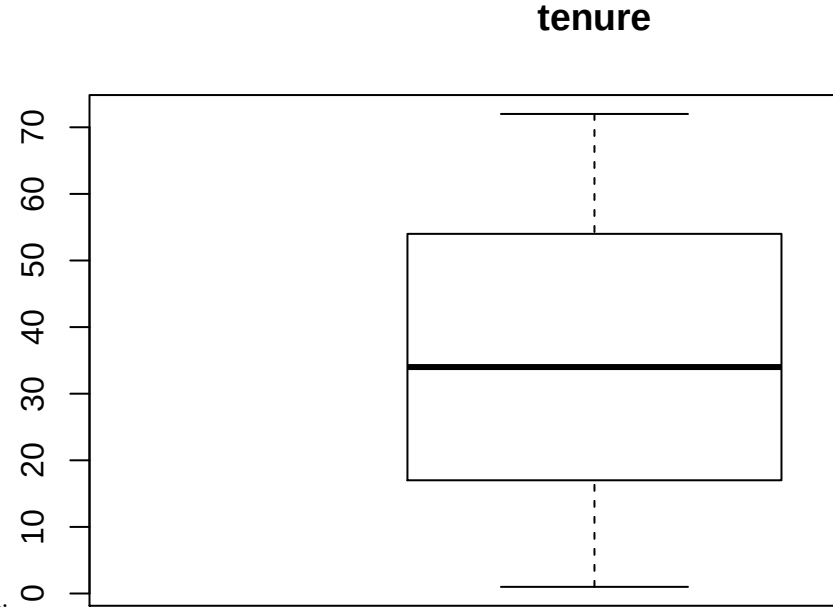
```
##     1000        0      162    0.855    13.27    16.92     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00     0.00    24.25    34.75    41.77
##
## lowest :   0.00   5.75   9.00   9.25   9.50, highest:  77.75  78.75  84.00  87.00 173.00
## ------------------------------------------------------------------------------
## equipmon
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      310    0.769    14.22     19.2     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00     0.00    31.47    42.66    48.65
##
## lowest :  0.00 15.40 19.55 19.75 19.80, highest: 64.20 69.40 70.05 73.80 77.70
## ------------------------------------------------------------------------------
## cardmon
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      176    0.967    13.78    14.67     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00    12.00    20.50    31.50    37.79
##
## lowest :   0.00   2.75   3.75   4.50   4.75, highest:  74.00  82.25  84.75  87.50 109.25
## ------------------------------------------------------------------------------
## wiremon
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      263    0.651    11.58    17.76     0.00     0.00
##      .25      .50      .75      .90      .95
##     0.00     0.00    24.71    42.11    51.35
##
## lowest :   0.00  14.90  15.10  16.85  17.45, highest:  83.70  85.85  96.25 109.70 111.95
## ------------------------------------------------------------------------------
## longten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      960        1    574.1      699    12.30    29.18
##      .25      .50      .75      .90      .95
##    90.14   285.48   755.02  1537.09  2144.79
##
## lowest :    0.90    1.05    1.10    1.35    1.45
## highest: 4333.00 5464.60 5988.50 6353.90 7257.60
## ------------------------------------------------------------------------------
## tollten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      473    0.854    551.3    823.1      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0      0.0    846.9   1838.4   2424.3
##
## lowest :    0.00    5.75   20.75   23.05   26.50
## highest: 4748.45 4905.85 4938.60 5850.25 5916.00
## ------------------------------------------------------------------------------
## equipten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      386    0.769    465.6    736.6      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0      0.0    579.5   1854.9   2435.6
##
```

```
## lowest :    0.00   15.40   17.25   22.60   22.80
## highest: 3925.50 4167.70 4758.05 4980.80 5028.65
## ---------------------------------------------------------------------------
## cardten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      338    0.967    605.8    775.2      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0    332.5    910.0   1505.5   2181.0
##
## lowest :    0.00    2.75    5.00   10.00   15.00
## highest: 4975.00 5115.00 5455.00 5520.00 7515.00
## ---------------------------------------------------------------------------
## wireten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      297    0.651    442.7    743.3      0.0      0.0
##      .25      .50      .75      .90      .95
##      0.0      0.0    316.5   1833.7   2727.6
##
## lowest :    0.00   10.45   16.85   20.95   24.20
## highest: 4399.60 4828.20 6132.20 6444.95 7856.85
## ---------------------------------------------------------------------------
## multline
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.748      475    0.475   0.4992
##
## ---------------------------------------------------------------------------
## voice
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.635      304    0.304   0.4236
##
## ---------------------------------------------------------------------------
## pager
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.579      261    0.261   0.3861
##
## ---------------------------------------------------------------------------
## internet
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.698      368    0.368   0.4656
##
## ---------------------------------------------------------------------------
## callid
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.749      481    0.481   0.4998
##
## ---------------------------------------------------------------------------
## callwait
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.749      485    0.485   0.5001
##
## ---------------------------------------------------------------------------
## forward
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2     0.75      493    0.493   0.5004
```

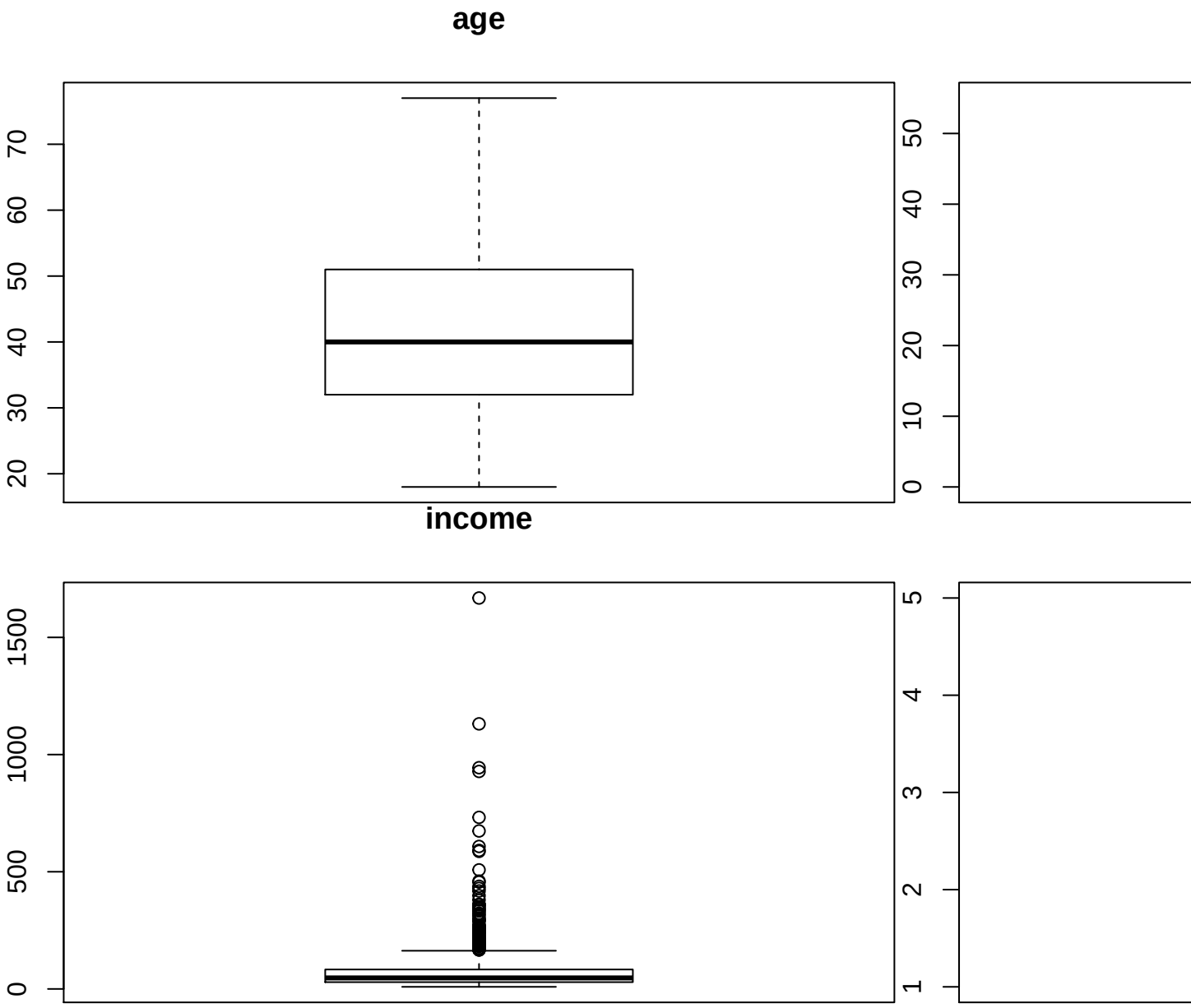
```
##
## ---------------------------------------------------------------------------
## confer
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2     0.75      502    0.502   0.5005
##
## ---------------------------------------------------------------------------
## ebill
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2      0.7      371    0.371   0.4672
##
## ---------------------------------------------------------------------------
## loglong
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      425        1    2.182    0.829    1.098    1.293
##      .25      .50      .75      .90      .95
##    1.649    2.143    2.668    3.176    3.454
##
## lowest : -0.105361  0.048790  0.095310  0.300105  0.371564
## highest:  4.131961  4.323470  4.401216  4.493121  4.604670
## ---------------------------------------------------------------------------
## logtoll
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      162    0.855     3.24   0.2626    2.757    2.918
##      .25      .50      .75      .90      .95
##    3.240    3.240    3.240    3.548    3.732
##
## lowest : 1.749200 2.197225 2.224624 2.251292 2.277267
## highest: 4.353499 4.366278 4.430817 4.465908 5.153292
## ---------------------------------------------------------------------------
## logequi
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      310    0.769    3.568   0.1544    3.240    3.372
##      .25      .50      .75      .90      .95
##    3.568    3.568    3.568    3.753    3.885
##
## lowest : 2.734368 2.972975 2.983153 2.985682 2.990720
## highest: 4.162003 4.239887 4.249209 4.301359 4.352855
## ---------------------------------------------------------------------------
## logcard
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      176    0.967    2.854   0.4806    2.079    2.277
##      .25      .50      .75      .90      .95
##    2.674    2.854    3.020    3.450    3.632
##
## lowest : 1.011601 1.321756 1.504077 1.558145 1.609438
## highest: 4.304065 4.409763 4.439706 4.471639 4.693639
## ---------------------------------------------------------------------------
## logwire
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      263    0.651    3.598     0.16    3.231    3.425
##      .25      .50      .75      .90      .95
##    3.598    3.598    3.598    3.740    3.939
##
```

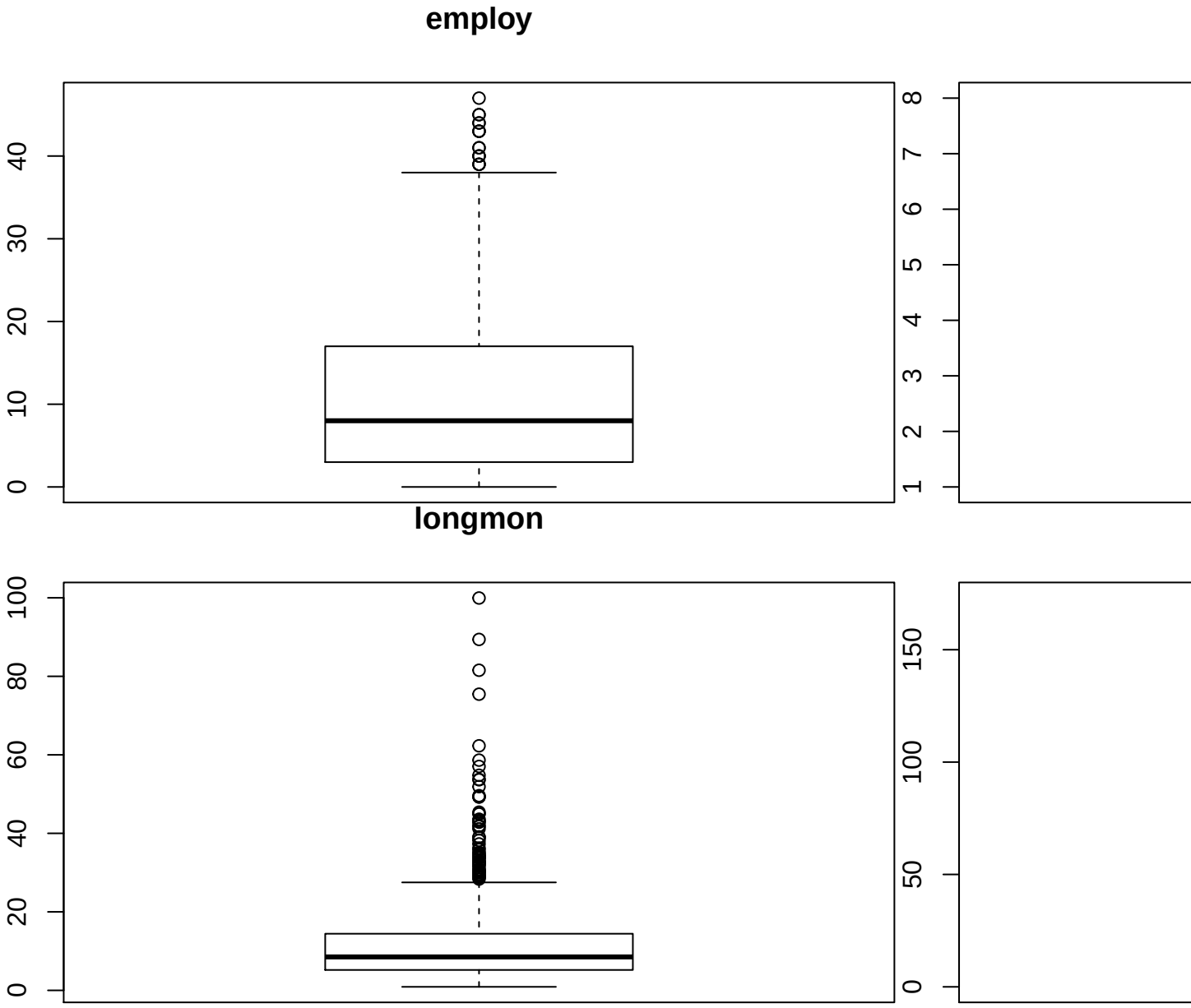
```
## lowest : 2.701361 2.714695 2.824351 2.859340 2.873565
## highest: 4.427239 4.452602 4.566949 4.697749 4.718052
## ---------------------------------------------------------------------------
## lninc
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1000        0      218        1    3.957   0.8912    2.890    3.045
##      .25      .50      .75      .90      .95
##    3.367    3.850    4.419    5.046    5.448
##
## lowest : 2.197225 2.302585 2.397895 2.484907 2.564949
## highest: 6.595781 6.833032 6.850126 7.030857 7.419381
## ---------------------------------------------------------------------------
## custcat
##        n  missing distinct     Info     Mean      Gmd
##     1000        0        4    0.936    2.487    1.252
##
## Value          1     2     3     4
## Frequency    266   217   281   236
## Proportion 0.266 0.217 0.281 0.236
## ---------------------------------------------------------------------------
## churn
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1000        0        2    0.597      274    0.274   0.3982
##
## ---------------------------------------------------------------------------
```

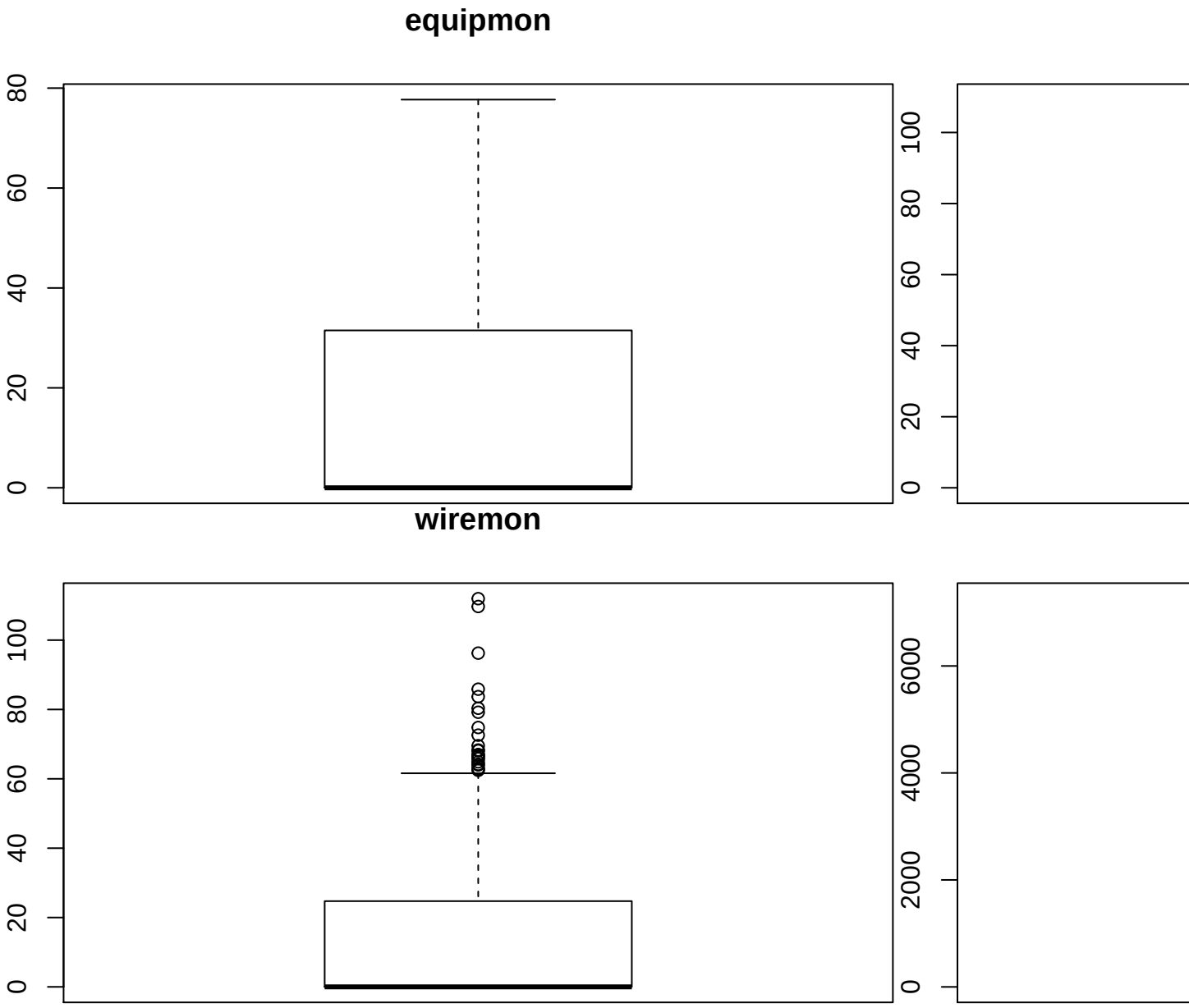
Result: So no missing values present now.

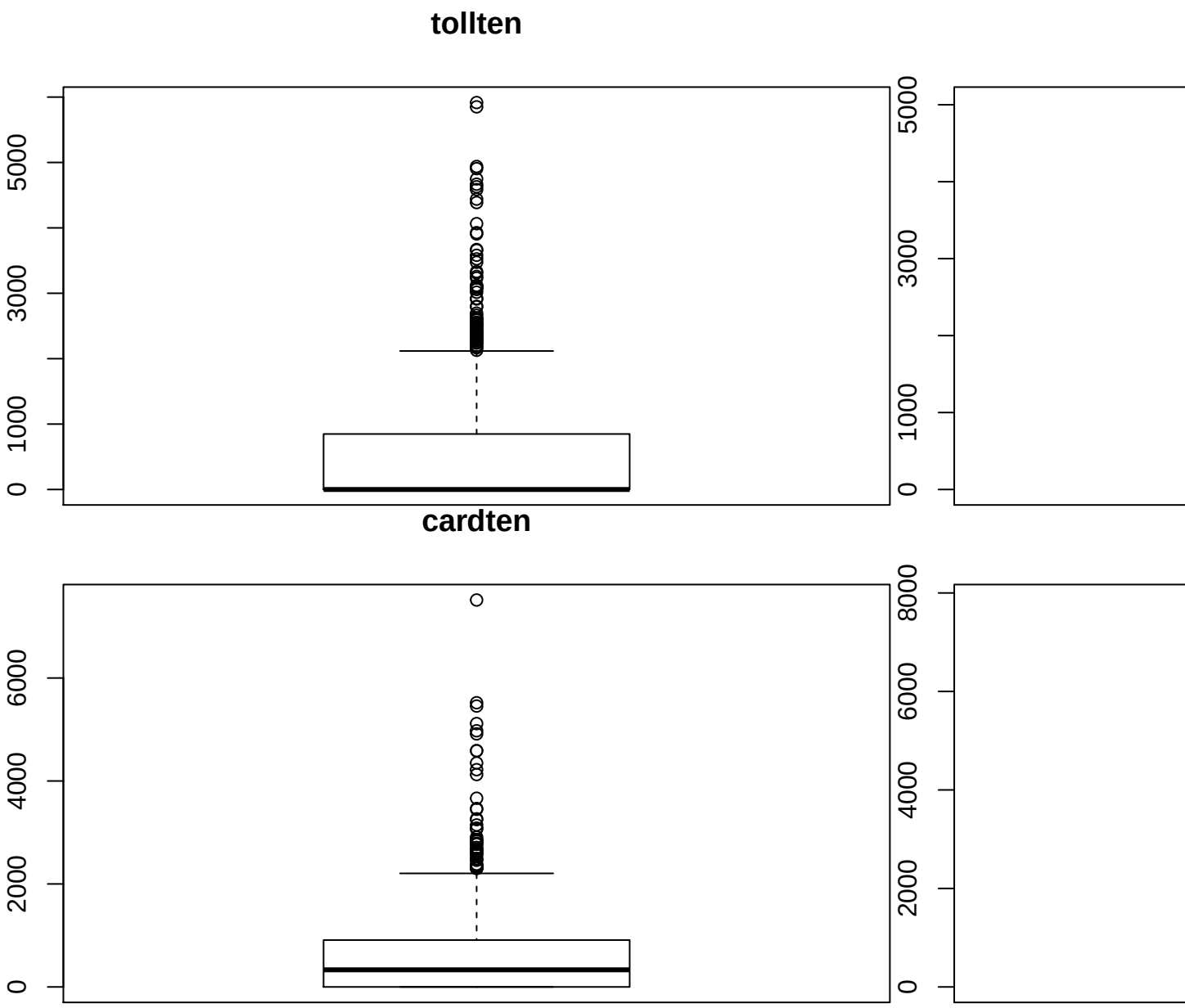


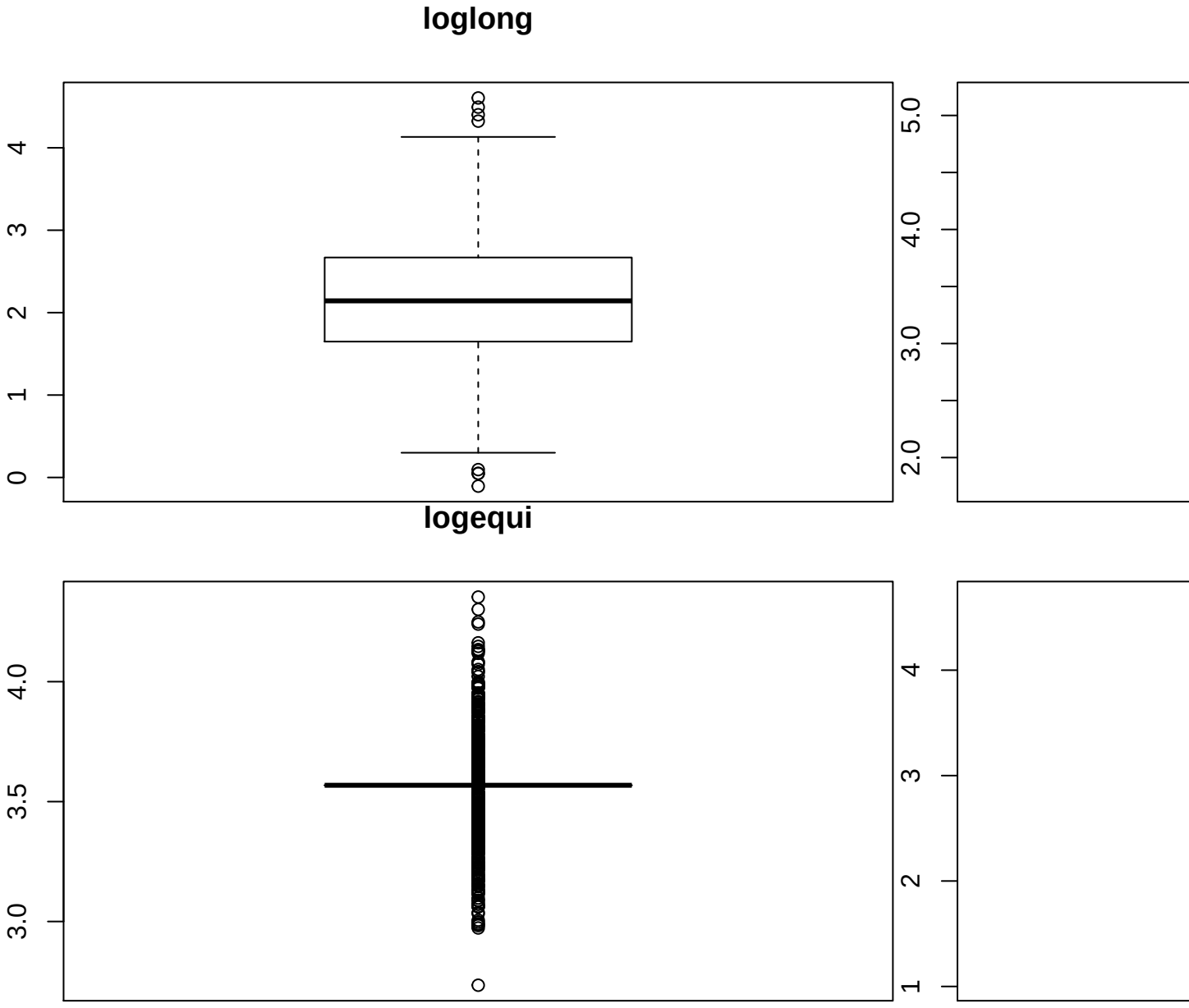
Step 2: Checking for any outliers by histogram method.

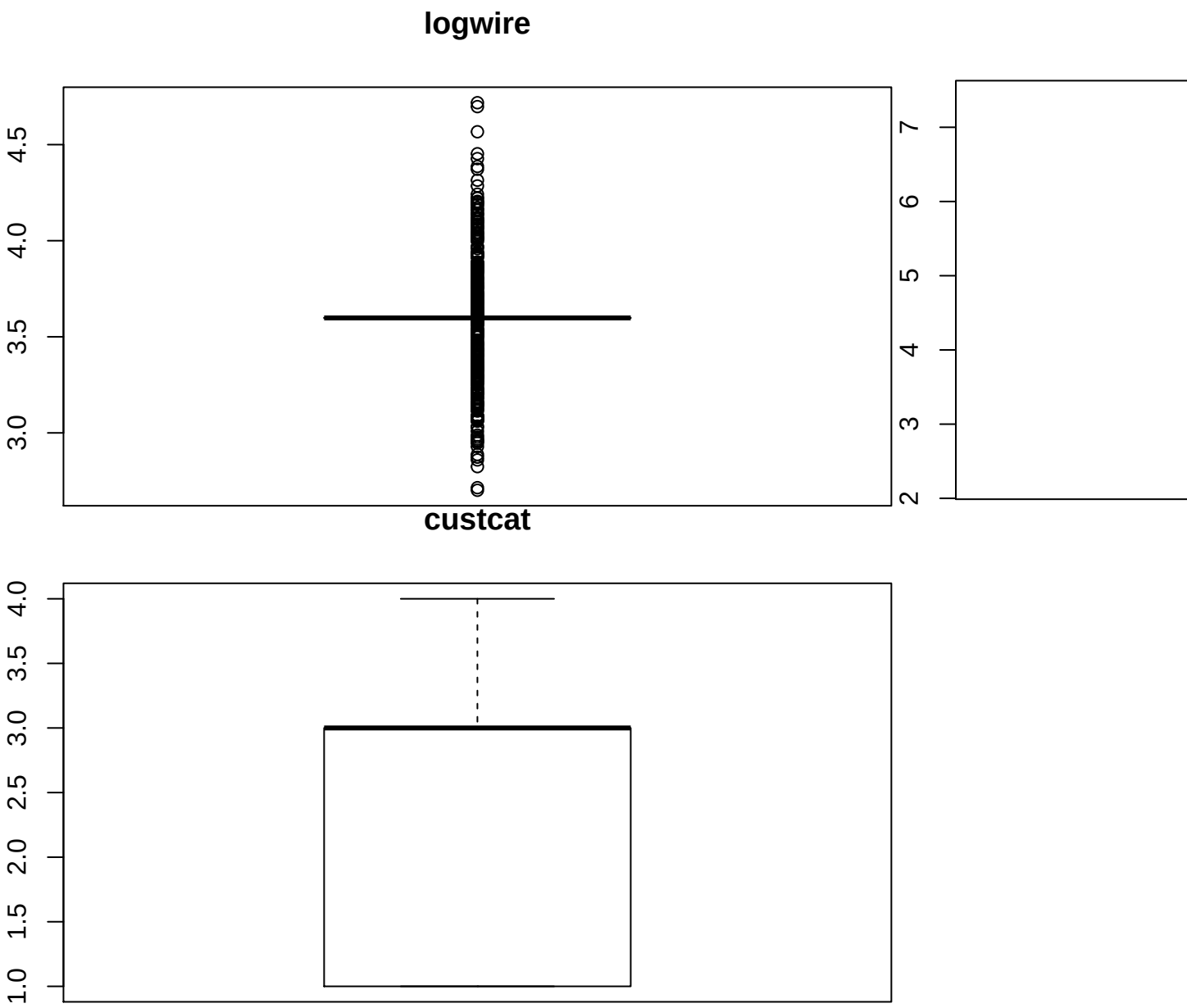

age
20
30
40
50
60
70
0
10
20
30
40
50
income
0
500
1000
1500
1
2
3
4
5


employ
0
10
20
30
40
1
2
3
4
5
6
7
8
longmon
0
20
40
60
80
100
0
50
100
150


equipmon
0
20
40
60
80
0
20
40
60
80
100
wiremon
0
20
40
60
80
100
0
2000
4000
6000

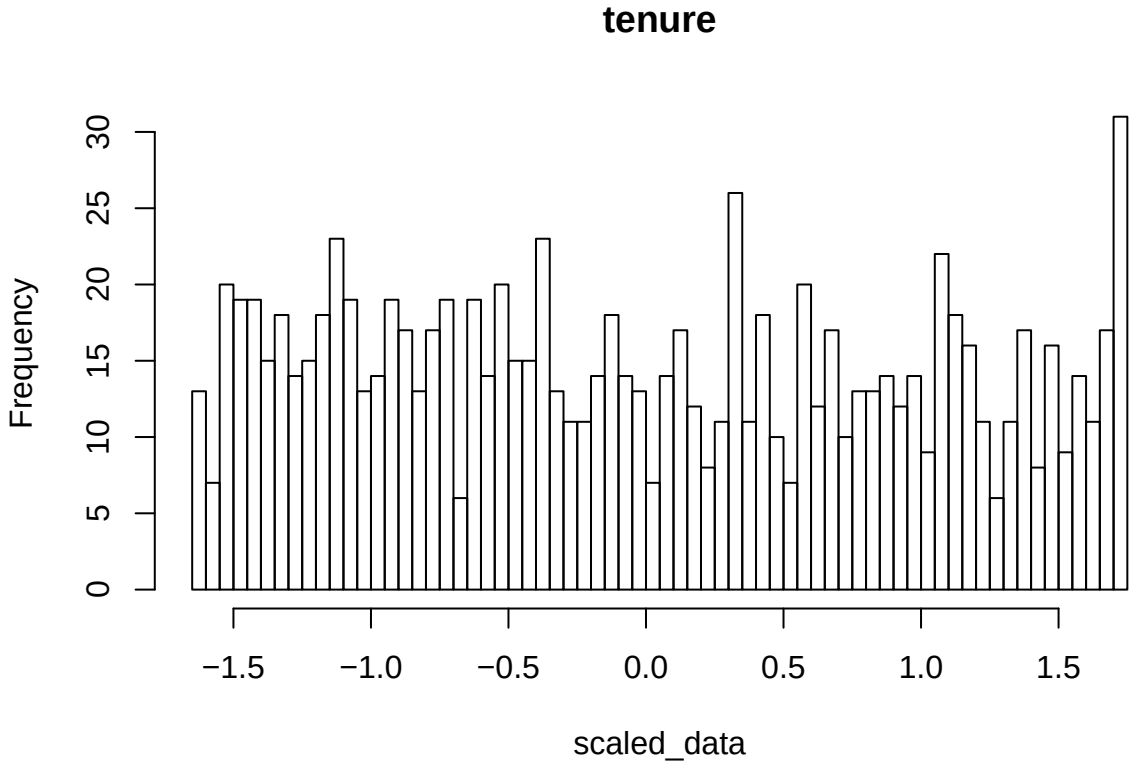

tollten
0
1000
3000
5000
cardten
0
2000
4000
6000
8000

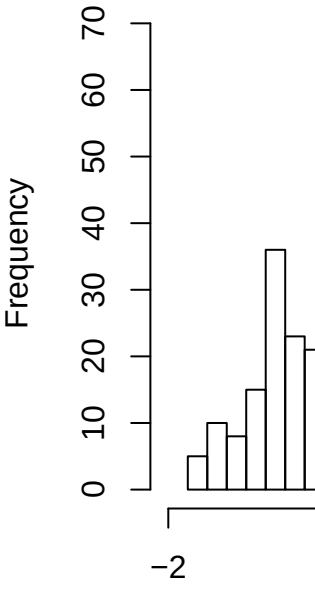

loglong
0
1
2
3
4
2.0
3.0
4.0
5.0
logequi
3.0
3.5
4.0
1
2
3
4

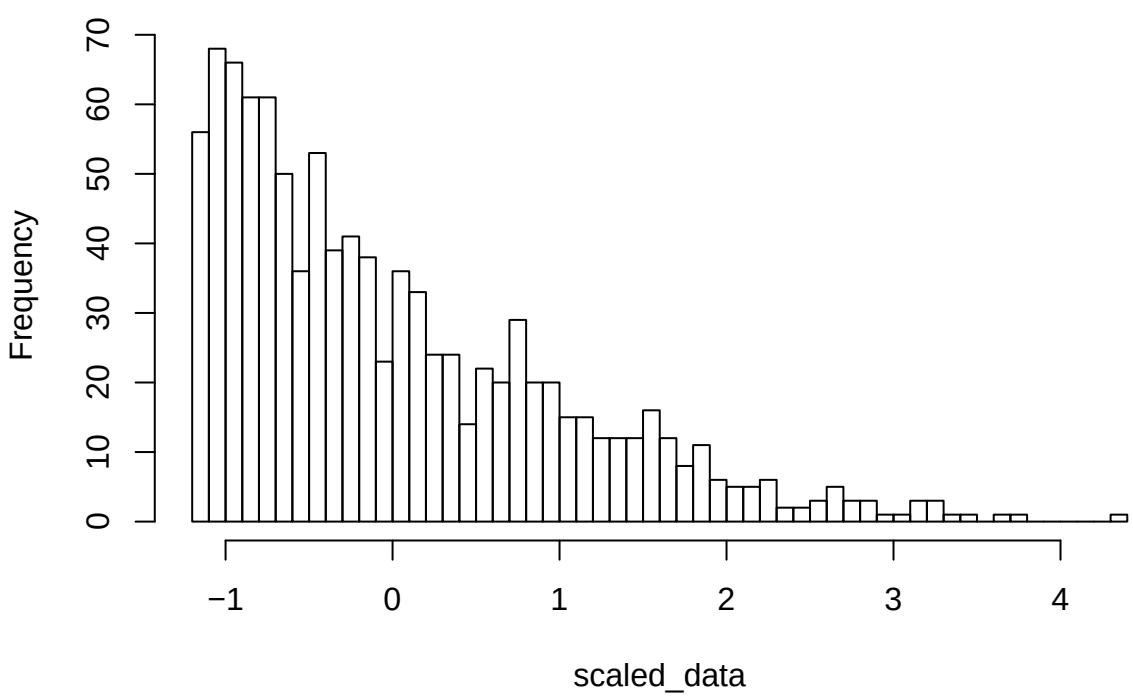


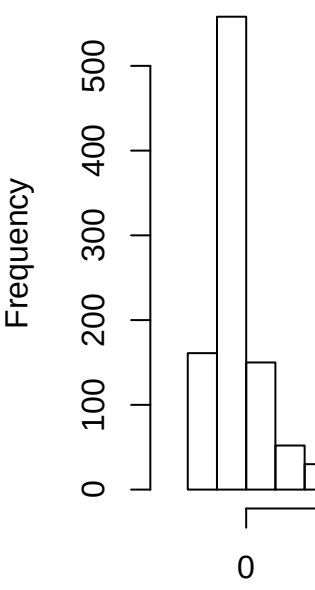
Result : There are outliers in many columns but it is not necessary that all the parameters of the same row are having extreme data, so we are not removing any data from it.

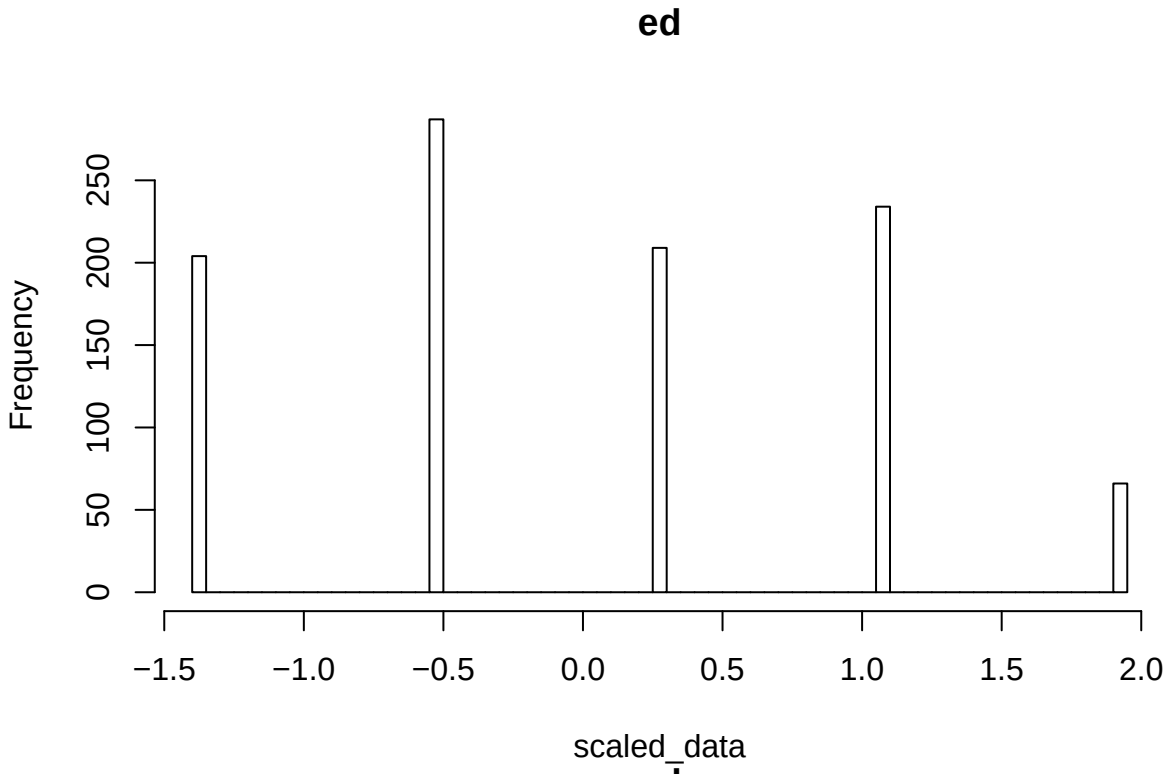
Step 3: Normality test

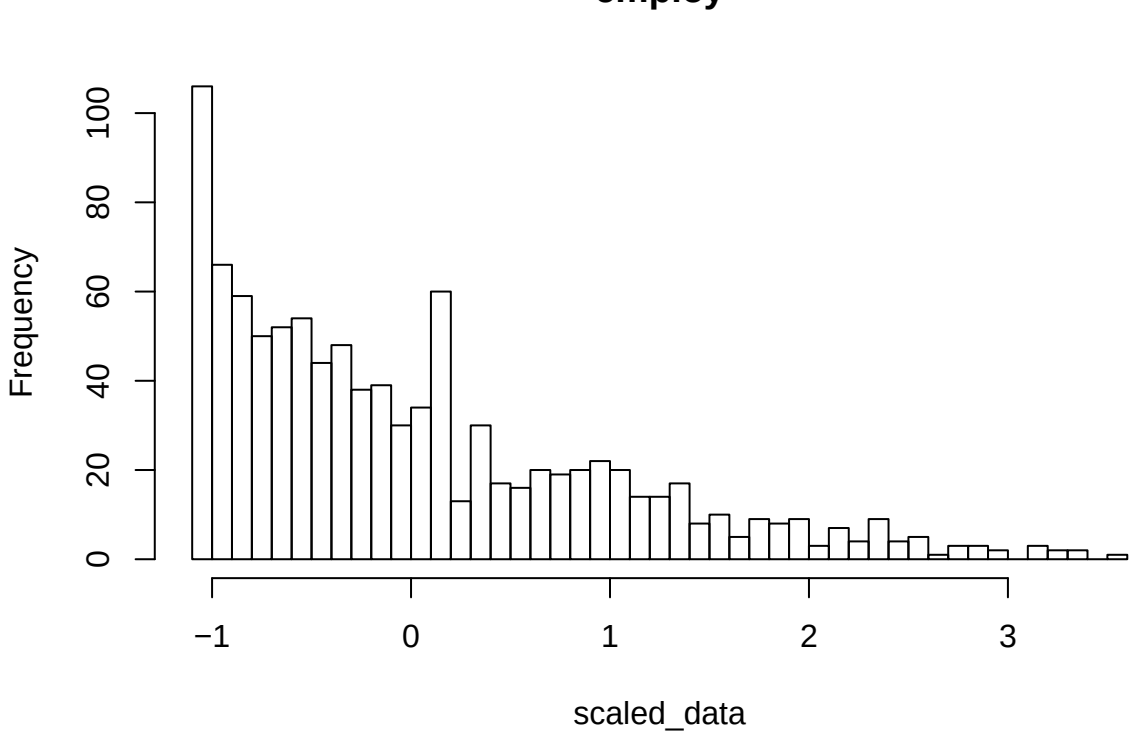

tenure
Frequency
0
5
10
15
20
25
30
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
scaled_data

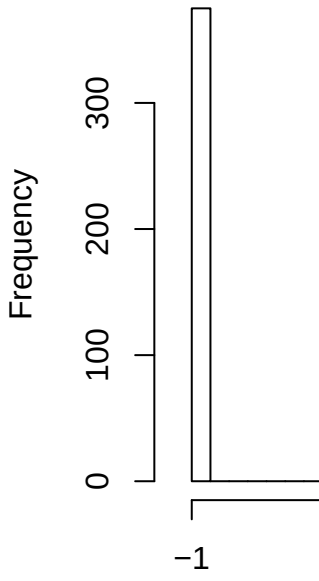

Frequency
0
10
20
30
40
50
60
70
−2

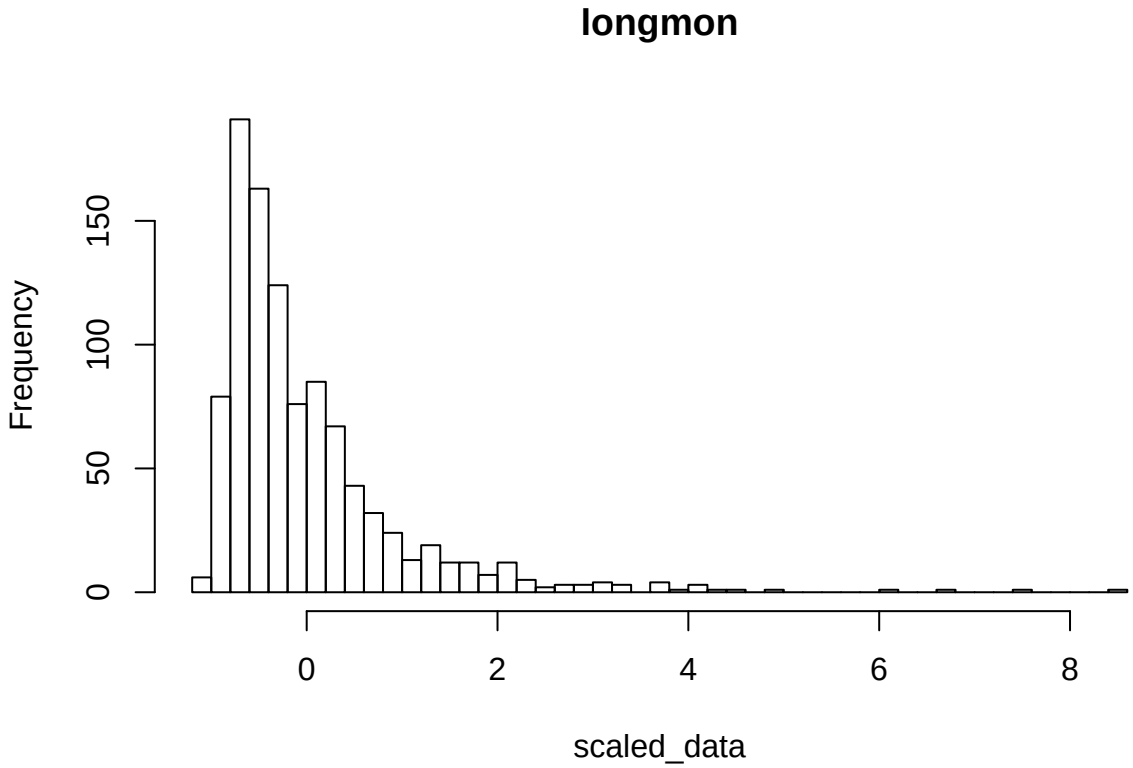

address
Frequency
0
10
20
30
40
50
60
70
−1
0
1
2
3
4
scaled_data

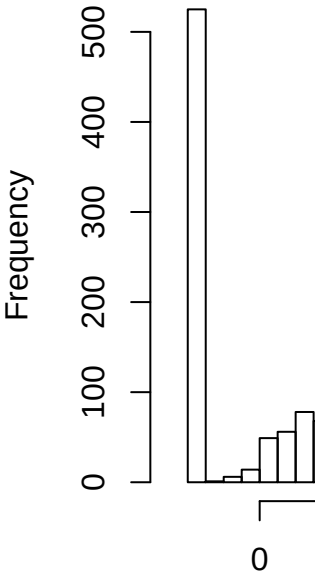

Frequency
0
100
200
300
400
500
0

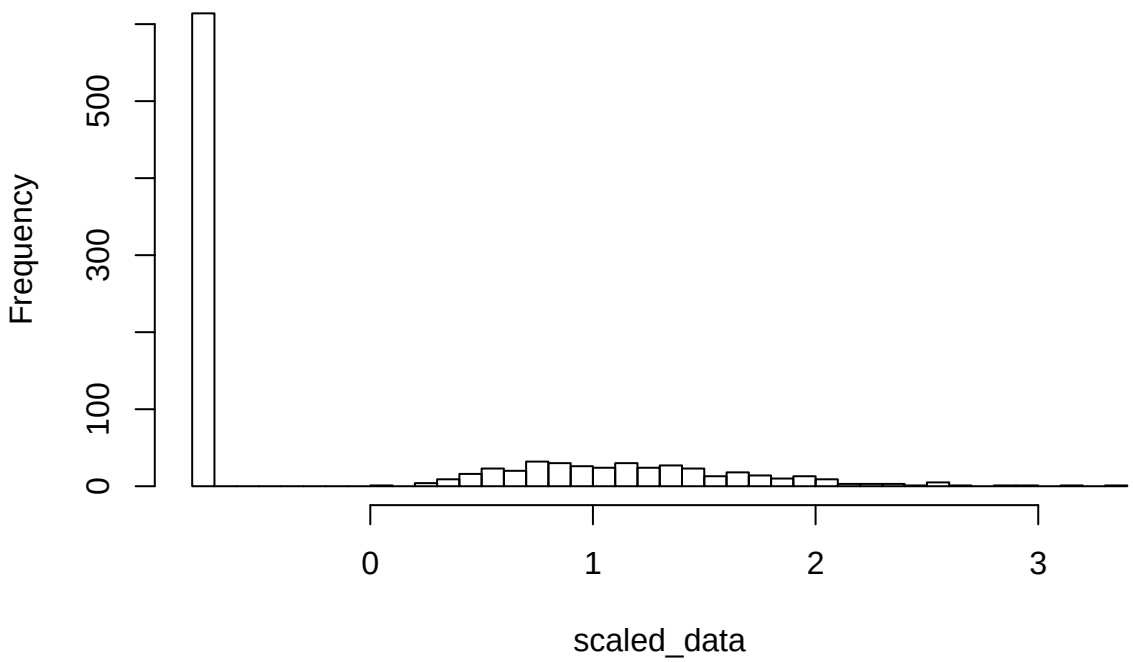

ed
Frequency
0
50
100
150
200
250
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
scaled_data

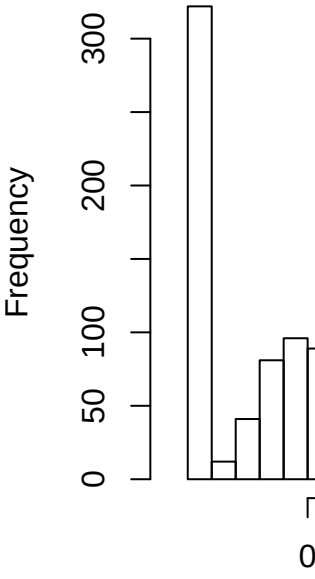

employ
Frequency
0
20
40
60
80
100
−1
0
1
2
3
scaled_data

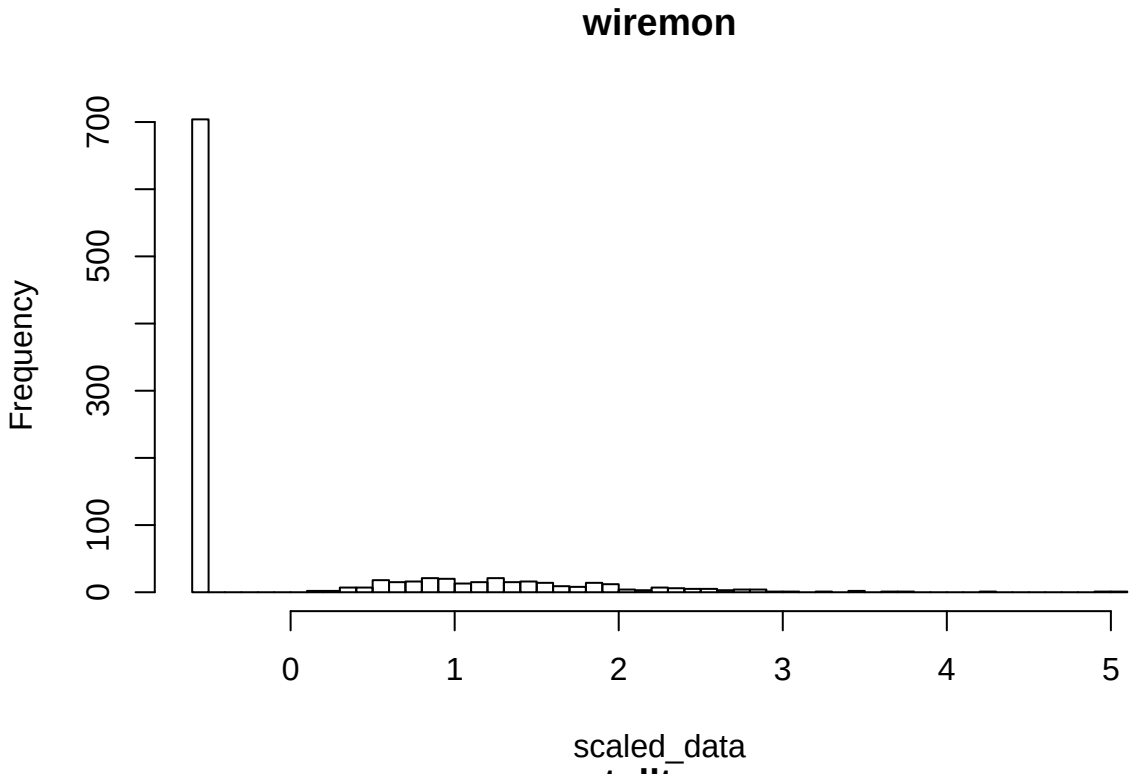

Frequency
0
100
200
300
−1

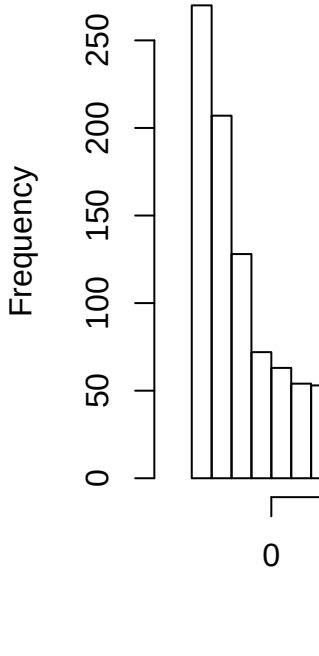

longmon
Frequency
0
50
100
150
0
2
4
6
8
scaled_data

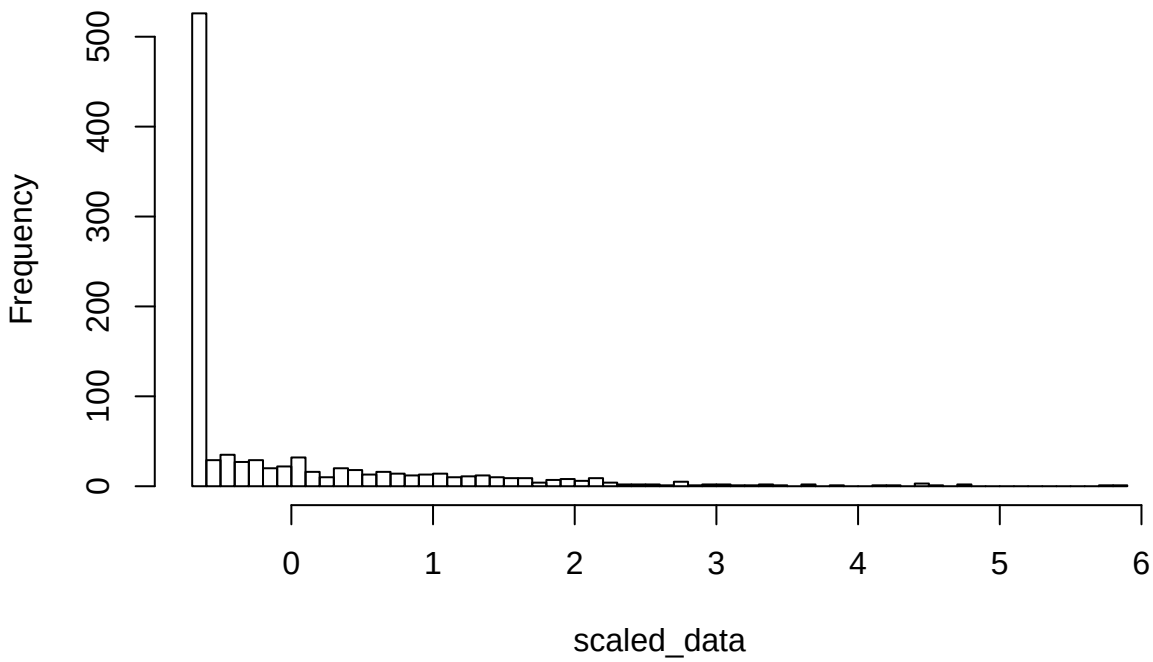

Frequency
0
100
200
300
400
500
0

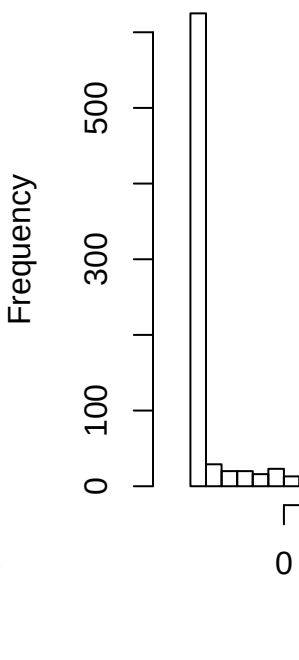

equipmon
Frequency
0
100
300
500
0
1
2
3
scaled_data

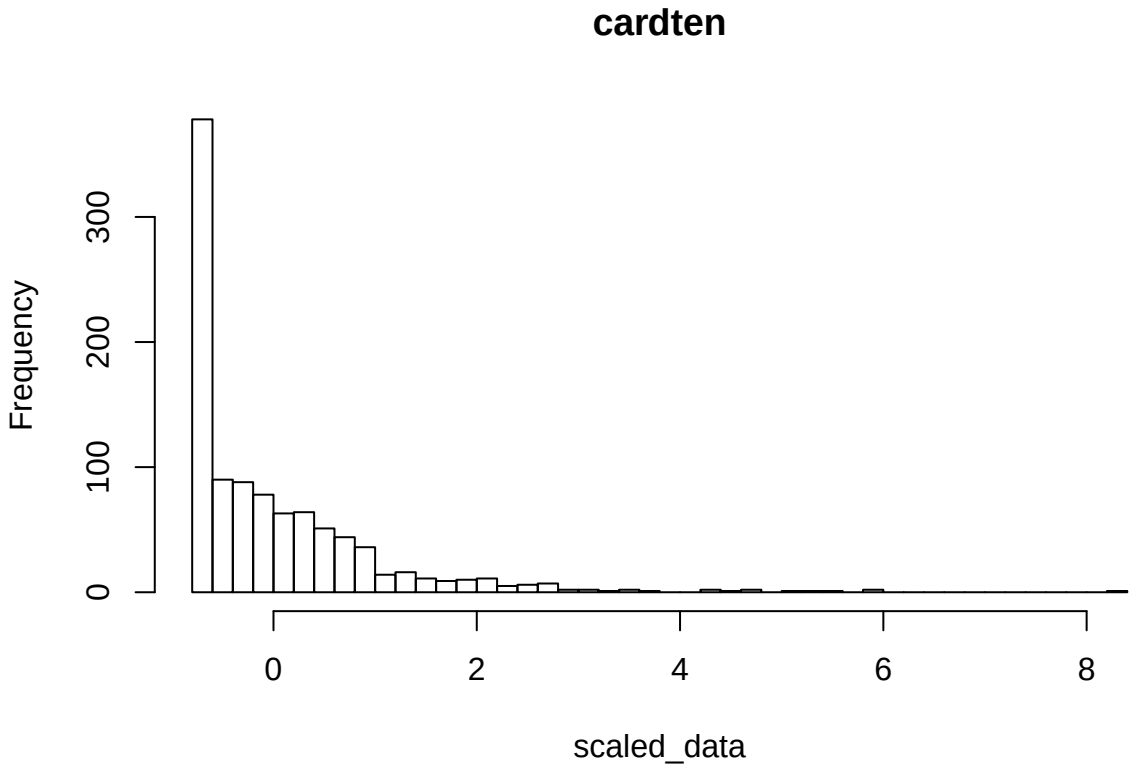

Frequency
0
50
100
200
300
0

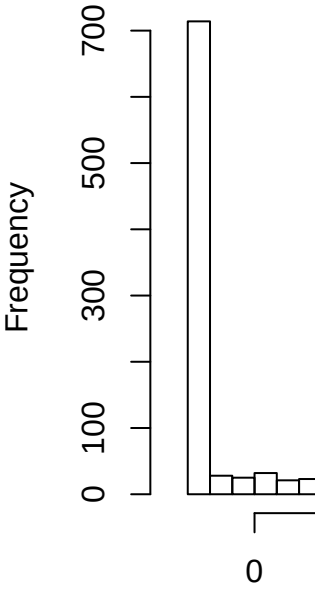

wiremon
Frequency
0
100
300
500
700
0
1
2
3
4
5
scaled_data

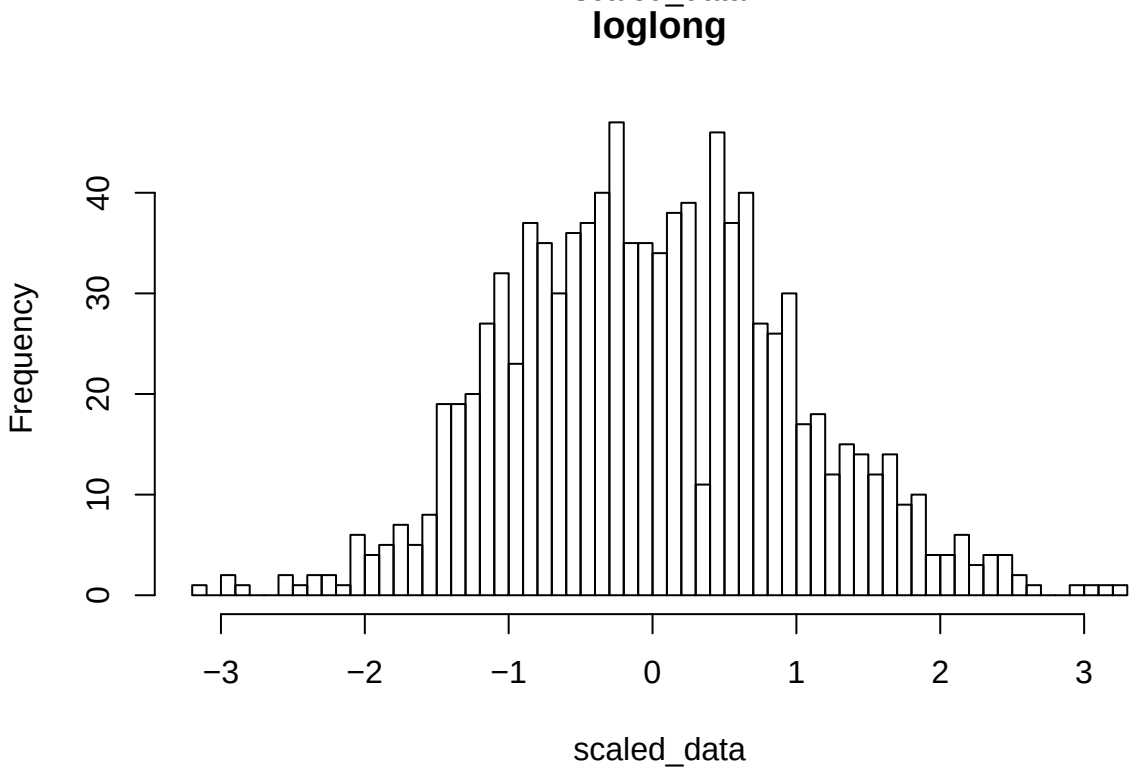

Frequency
0
50
100
150
200
250
0

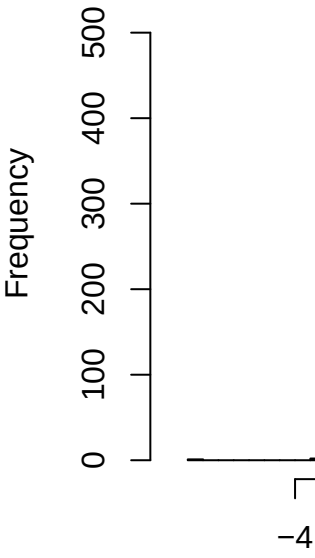

tollten
Frequency
0
100
200
300
400
500
0
1
2
3
4
5
6
scaled_data

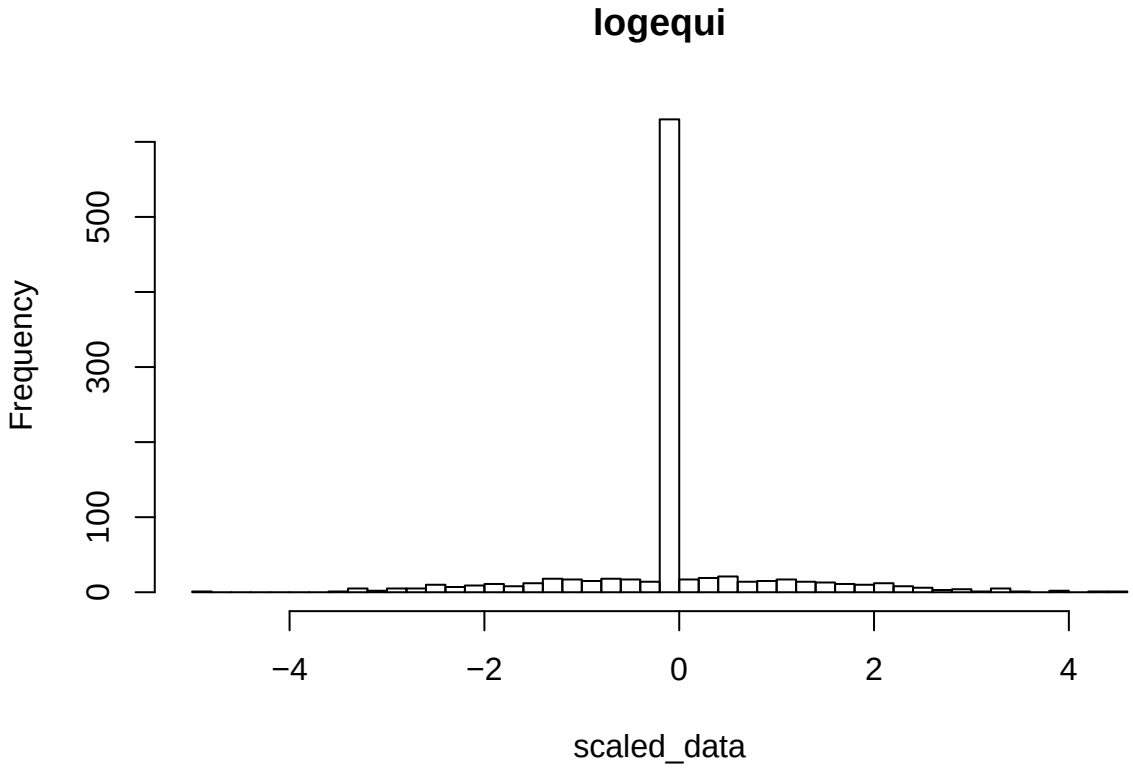

Frequency
0
100
300
500
0

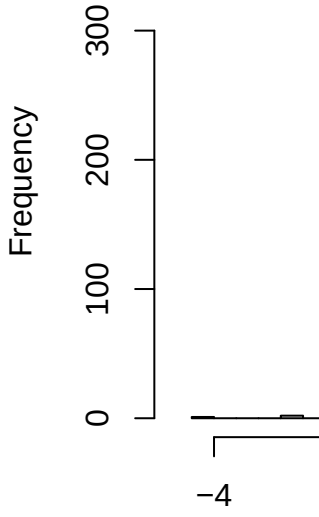

cardten
Frequency
0
100
200
300
0
2
4
6
8
scaled_data

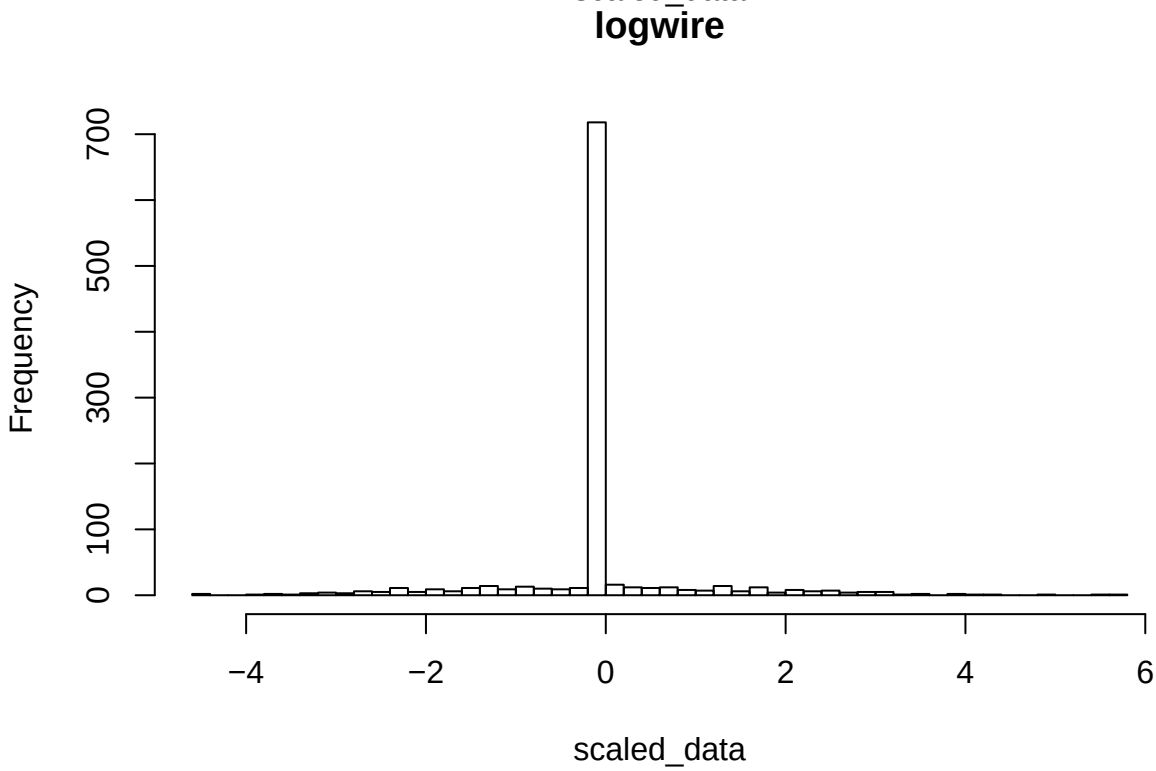

Frequency
0
100
300
500
700
0

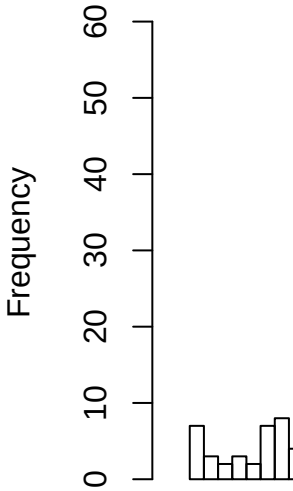

loglong
Frequency
0
10
20
30
40
−3
−2
−1
0
1
2
3
scaled_data

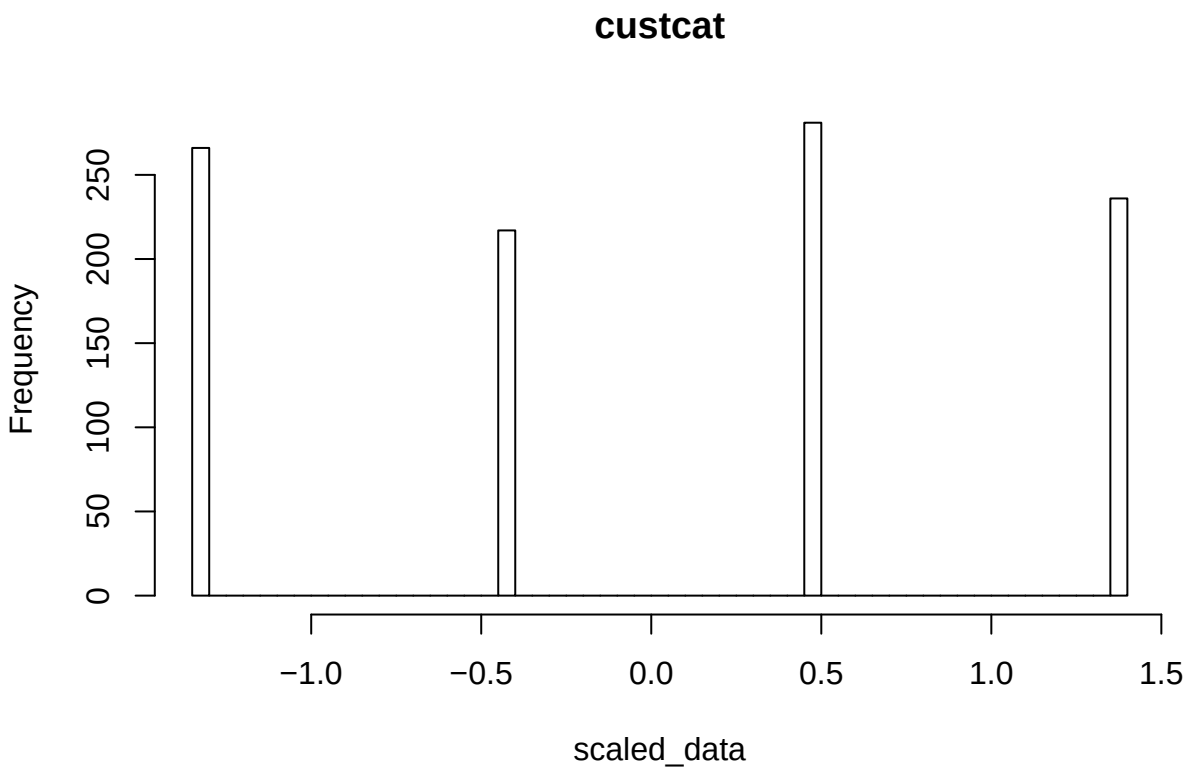

Frequency
0
100
200
300
400
500
−4


logequi
Frequency
0
100
300
500
−4
−2
0
2
4
scaled_data


Frequency
0
100
200
300
−4


logwire
Frequency
0
100
300
500
700
−4
−2
0
2
4
6
scaled_data


Frequency
0
10
20
30
40
50
60
−2


custcat
Frequency
0
50
100
150
200
250
−1.0
−0.5
0.0
0.5
1.0
1.5
scaled_data