

# 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
##

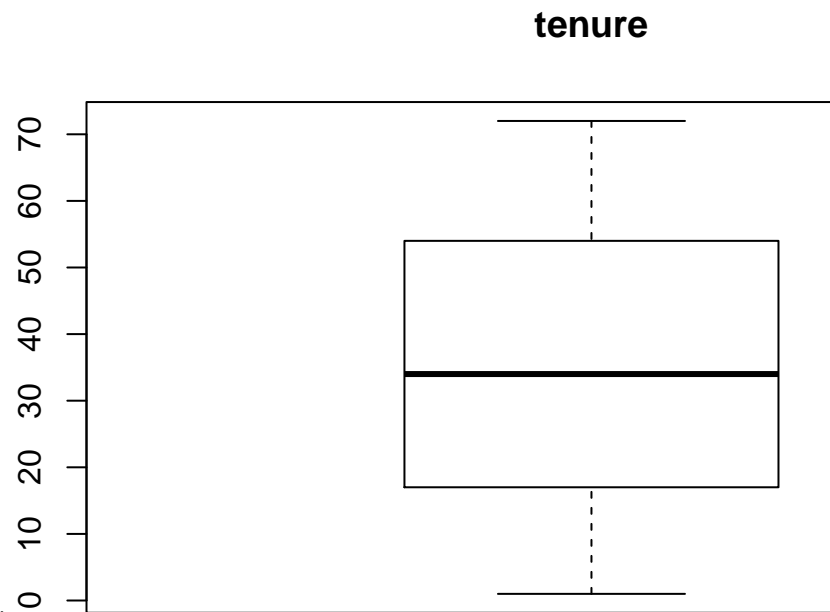
```

```

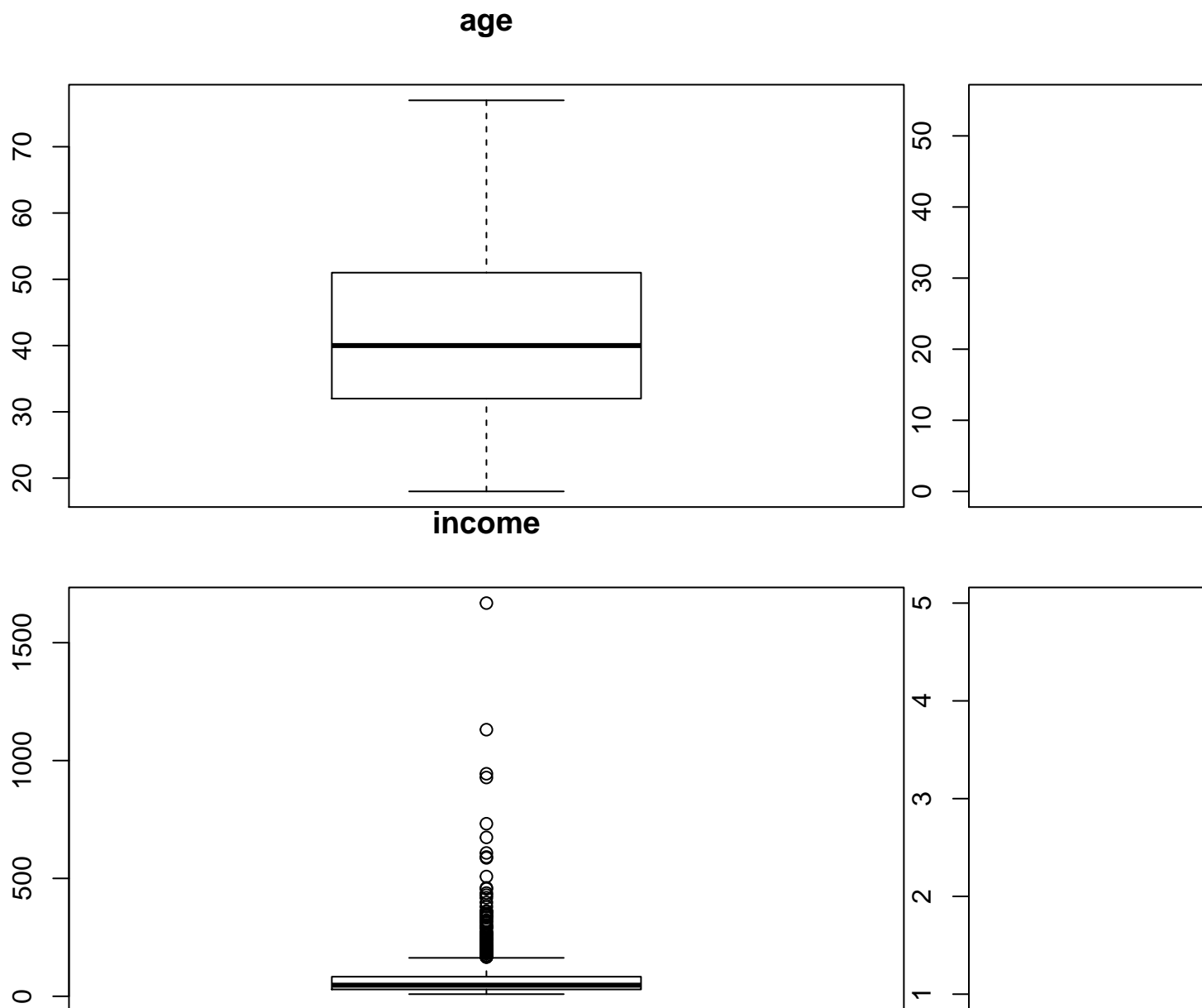
## lowest : 2.701361 2.714695 2.824351 2.859340 2.873565
## highest: 4.427239 4.452602 4.566949 4.697749 4.718052
## -----
## lninc
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1000         0       218         1     3.957    0.8912    2.890    3.045
##      .25      .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
##
## -----

```

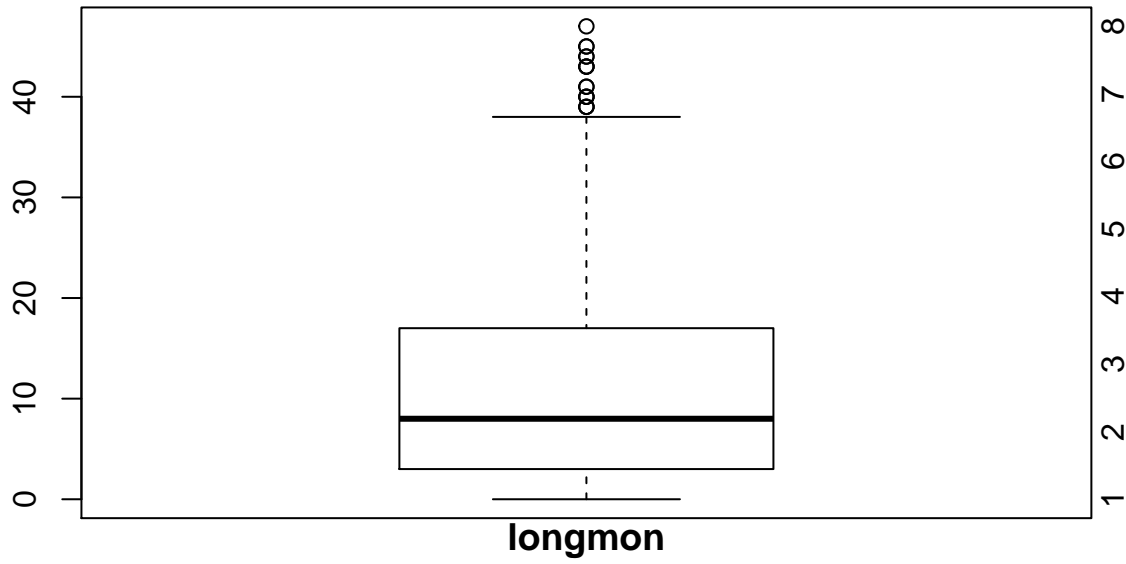
Result: So no missing values present now.



Step 2: Checking for any outliers by histogram method.

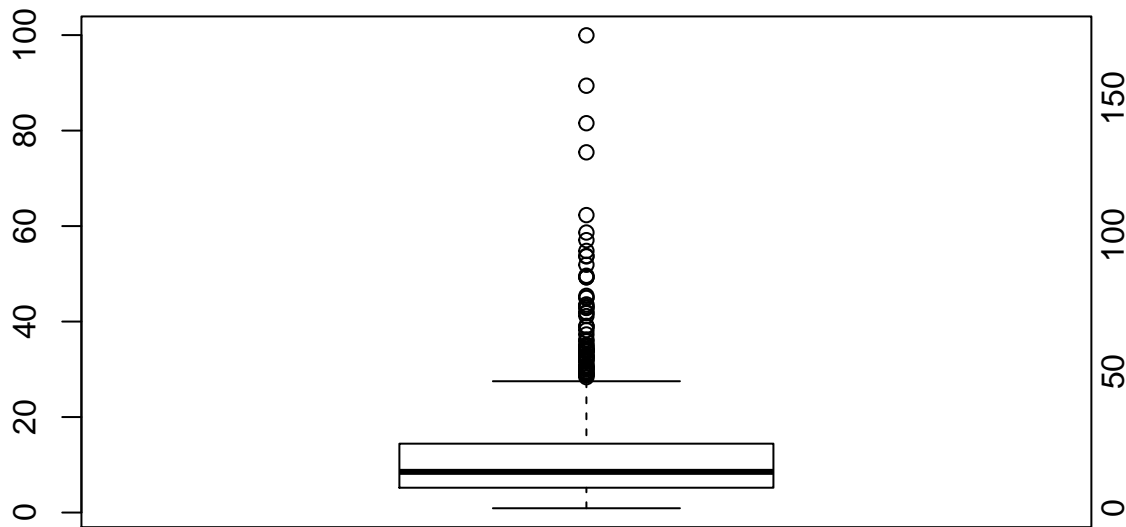


**employ**



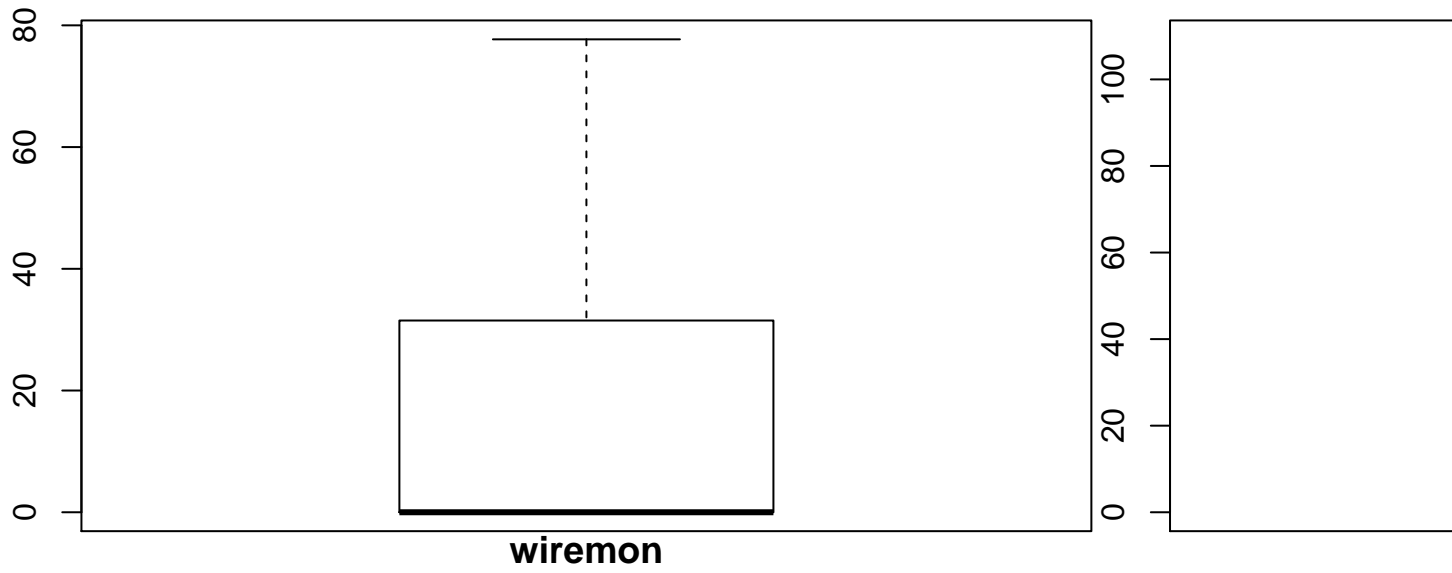
1 2 3 4 5 6 7 8

**longmon**

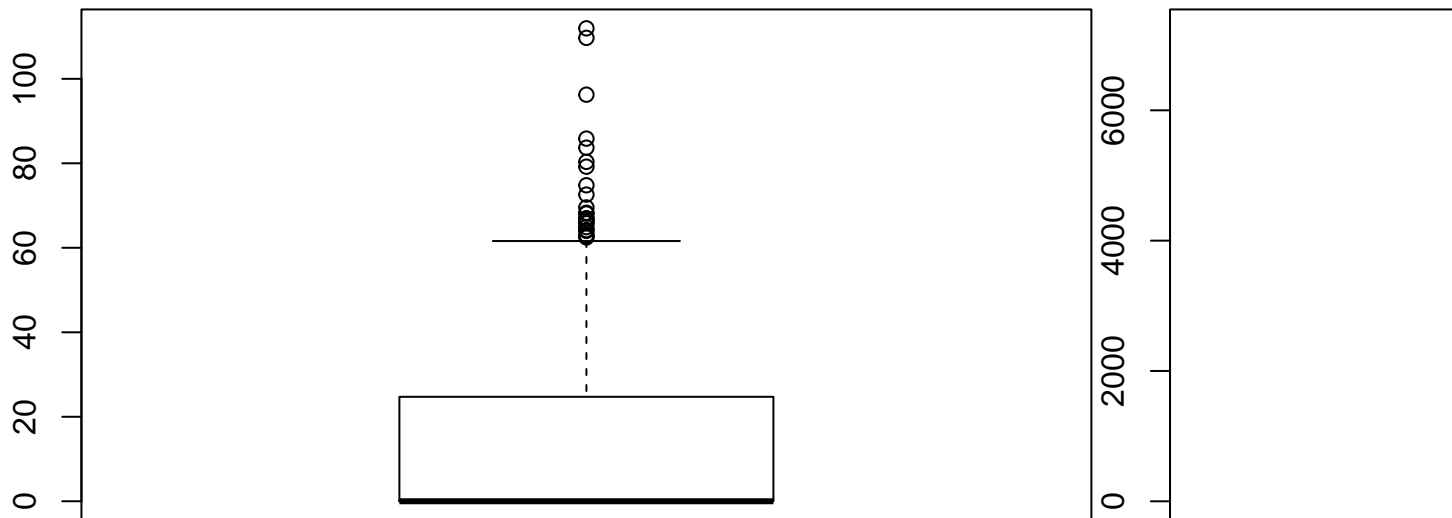


0 50 100 150

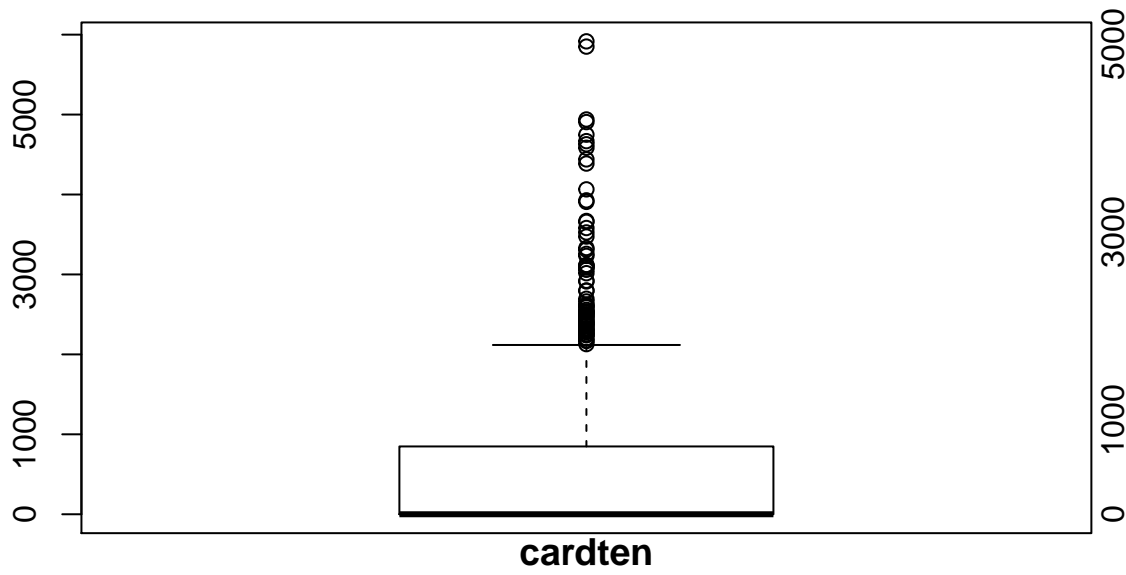
**equipmon**



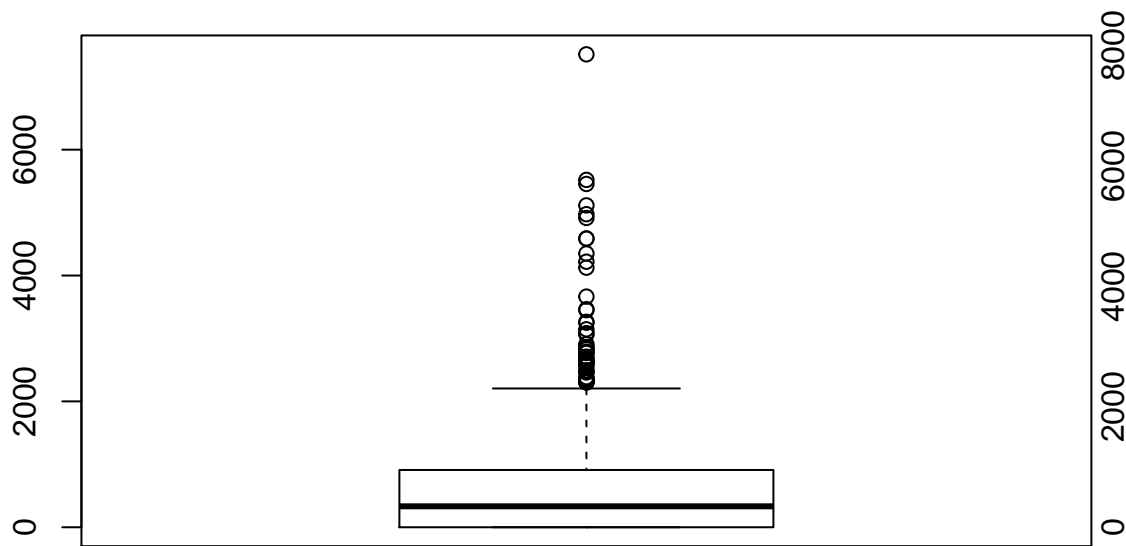
**wiremon**



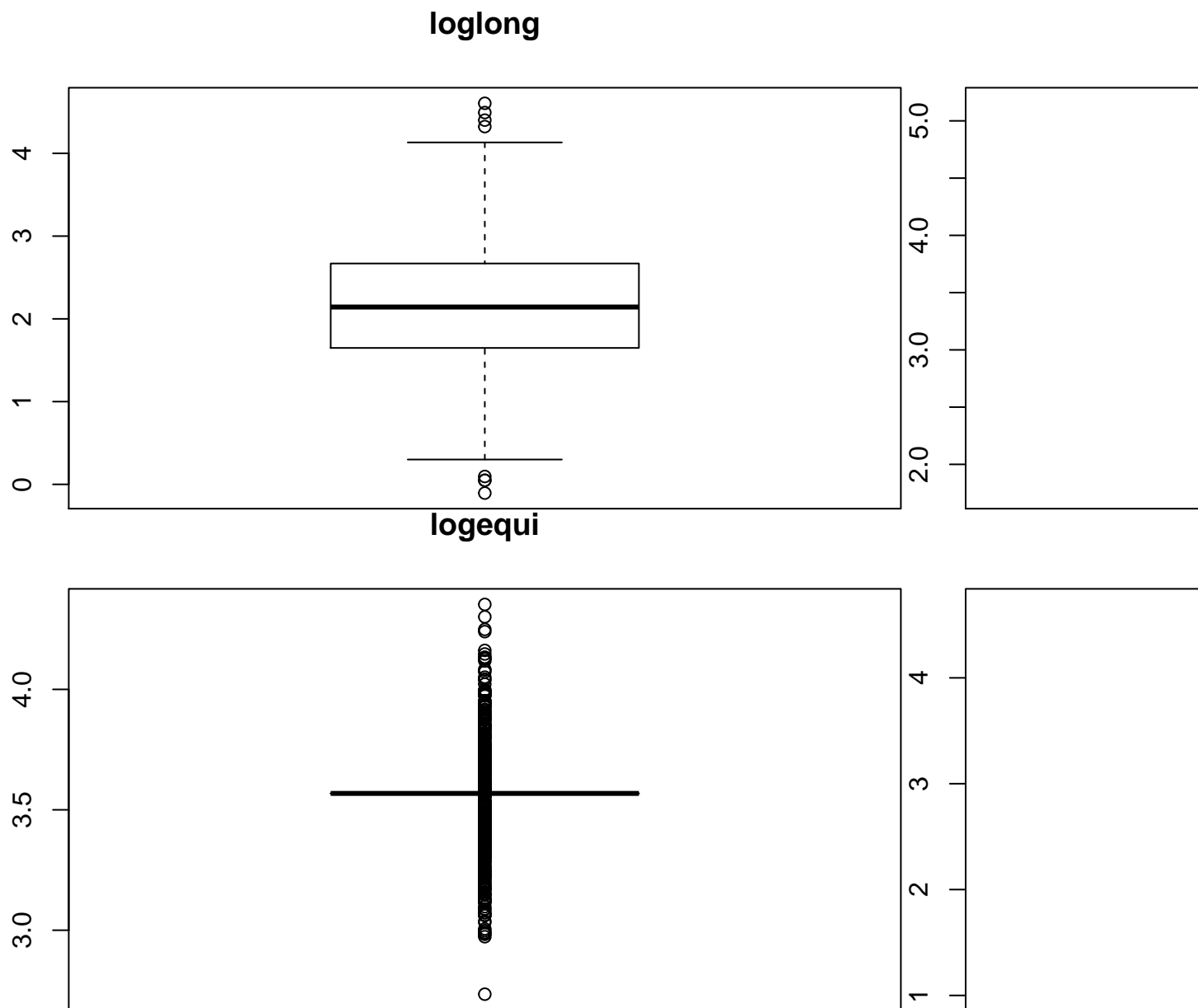
**tollten**

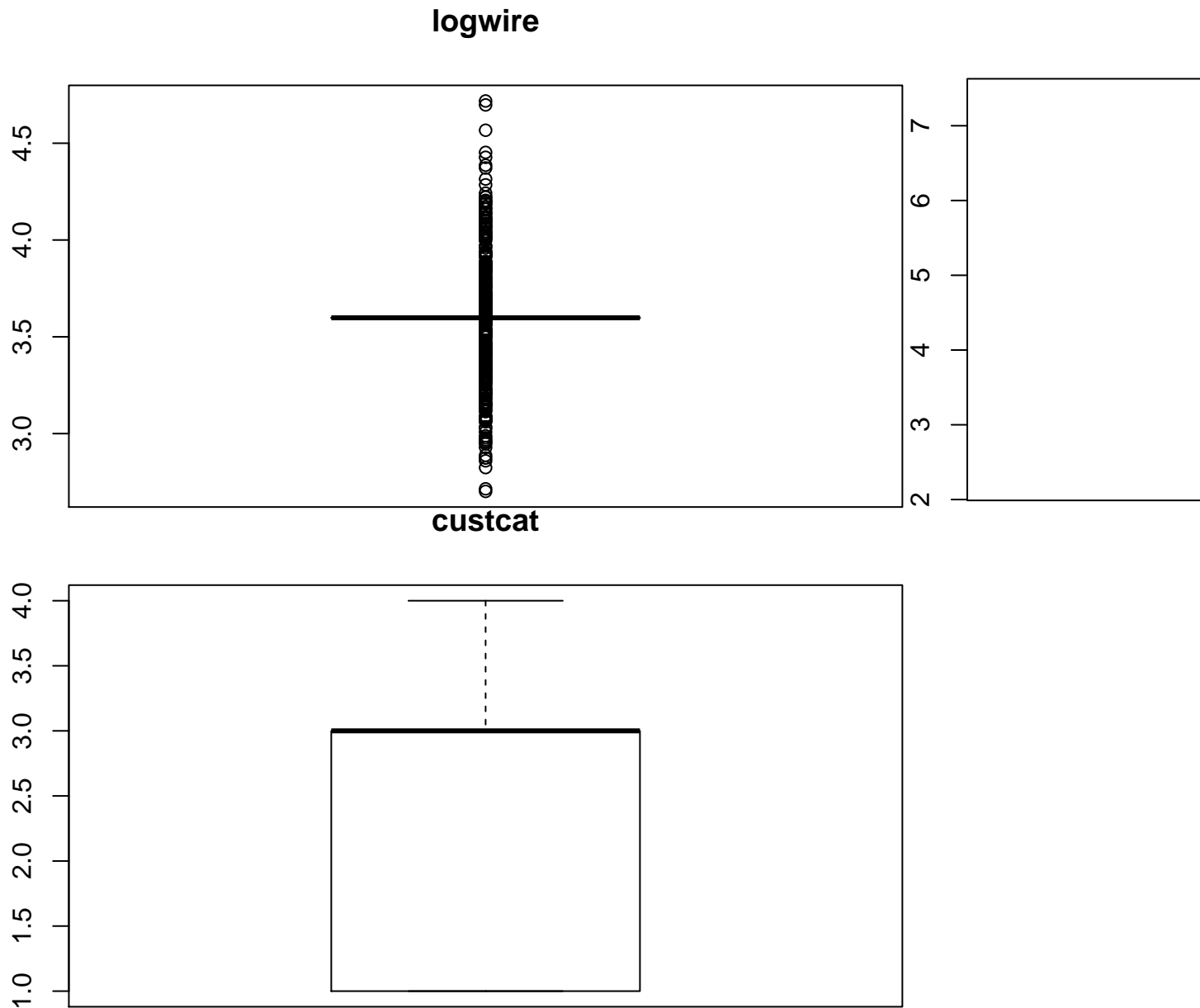


**cardten**





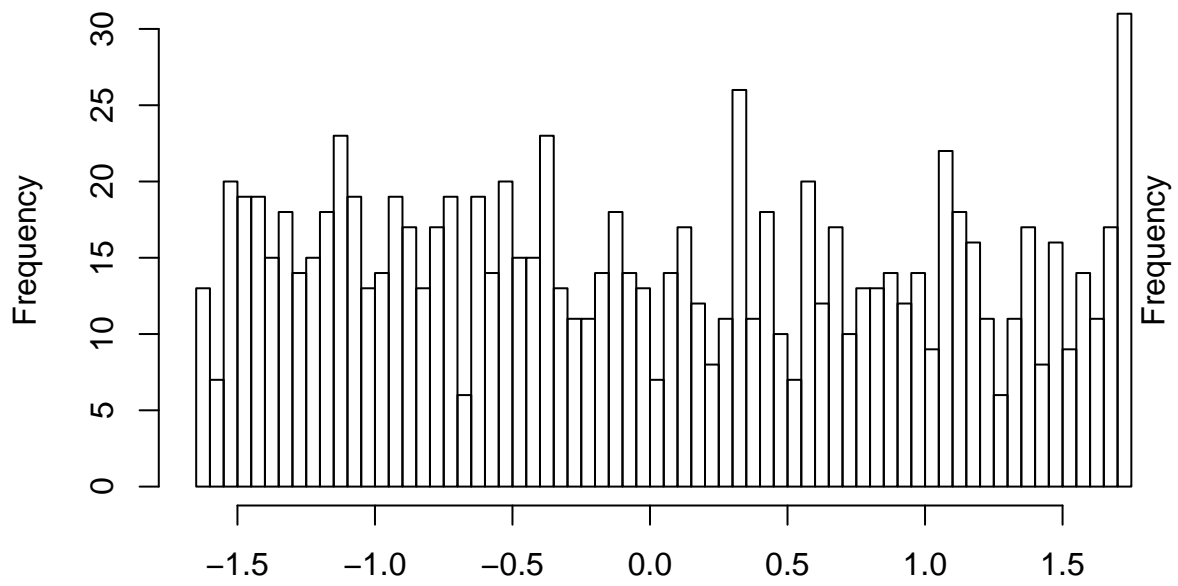




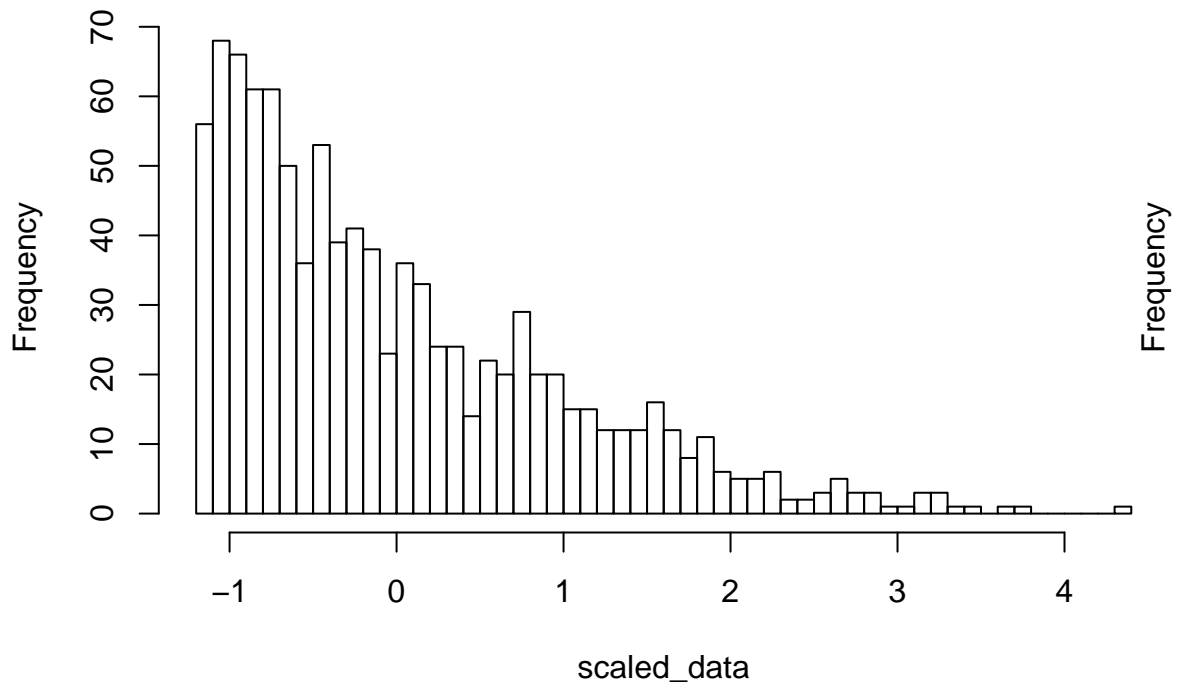
Result : There are outliers in many columns but it is not necessary that all the parameters of the same row are having extreme data, so we are not removing any data from it.

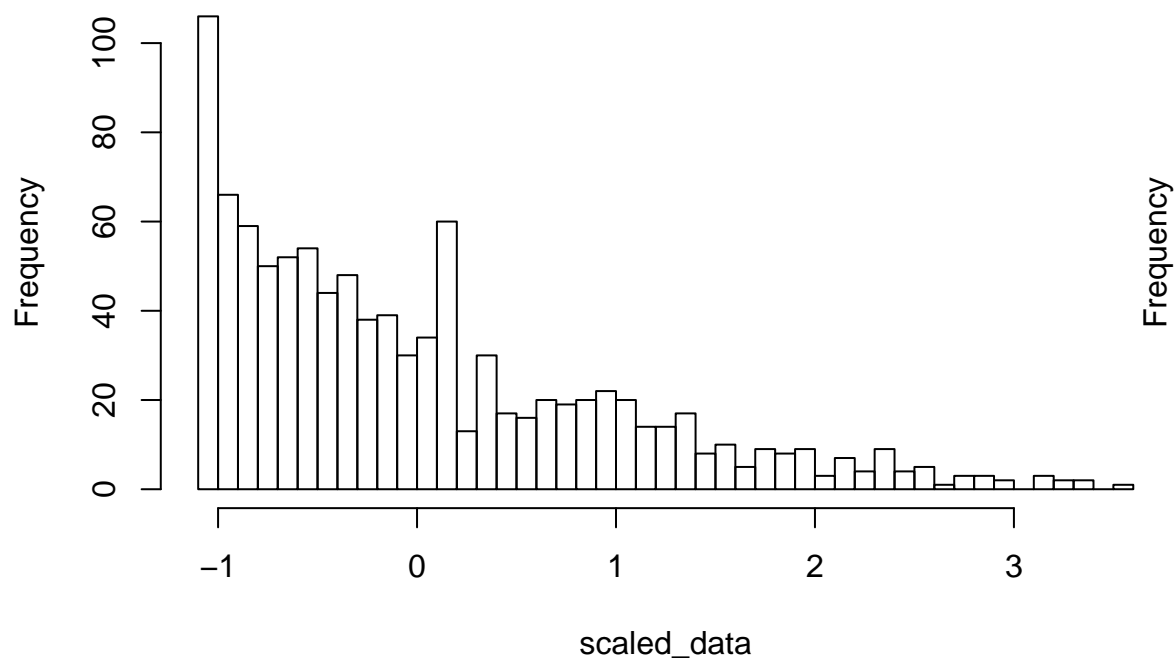
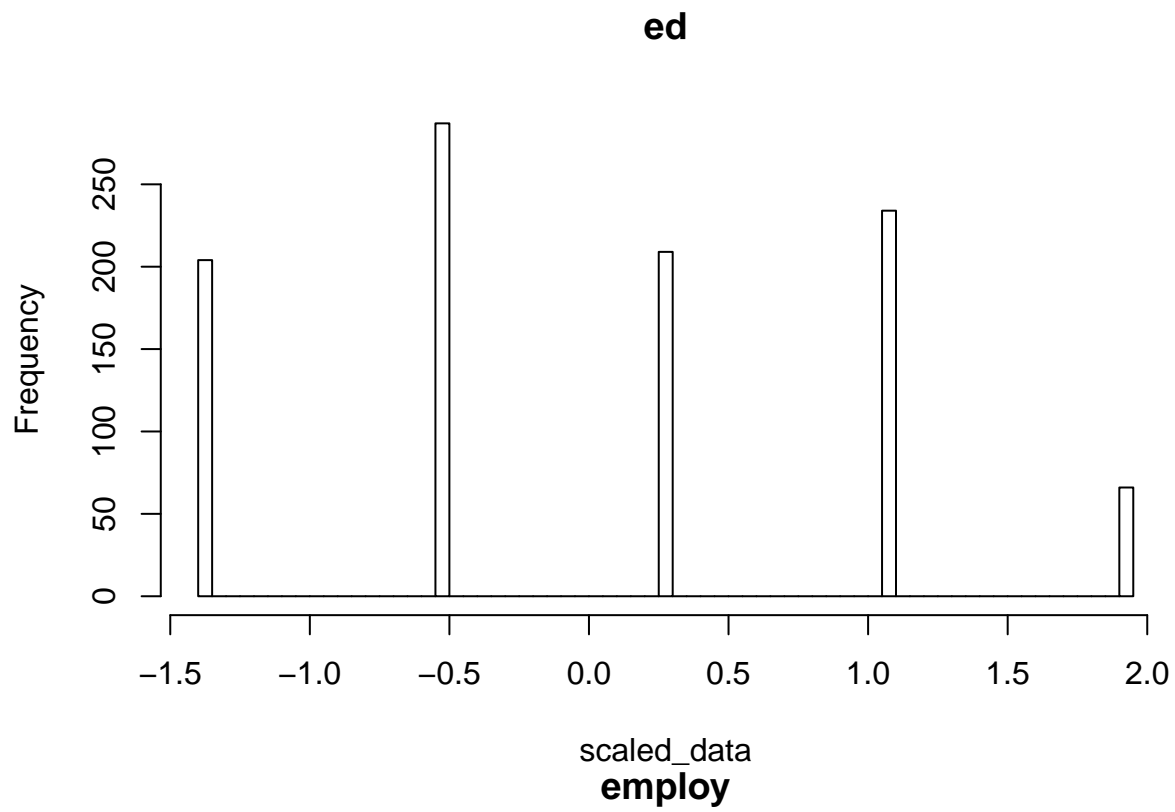
Step 3: Normality test

**tenure**

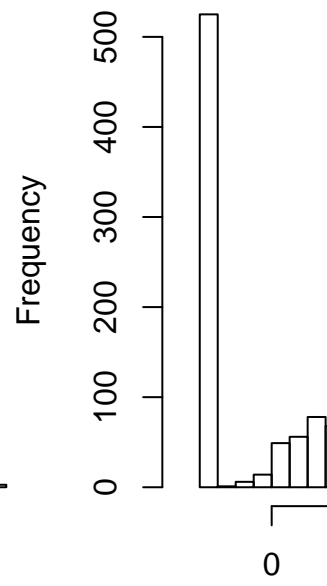
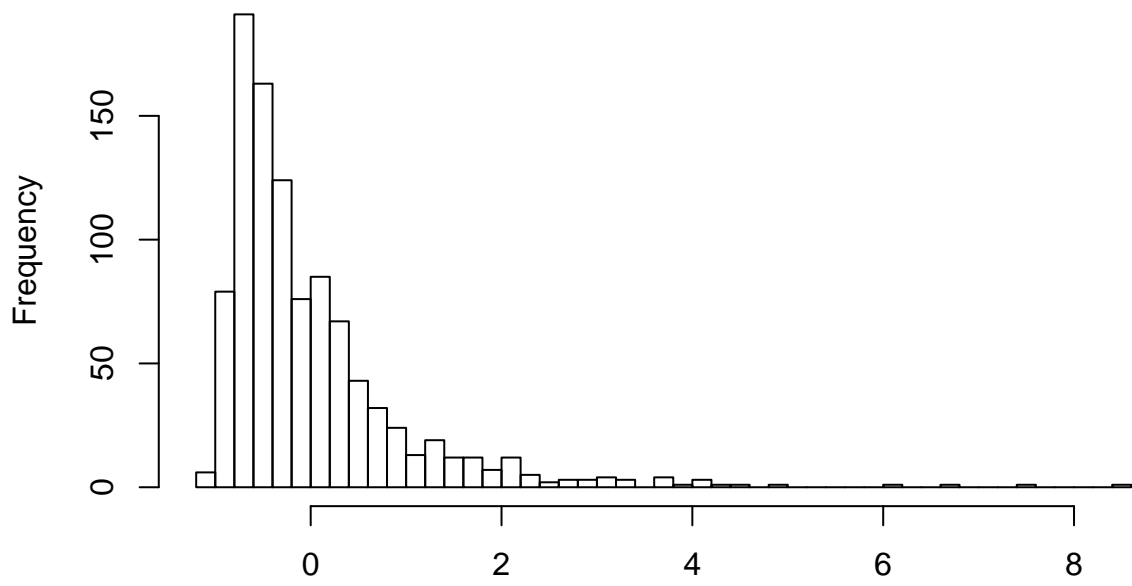


**address**

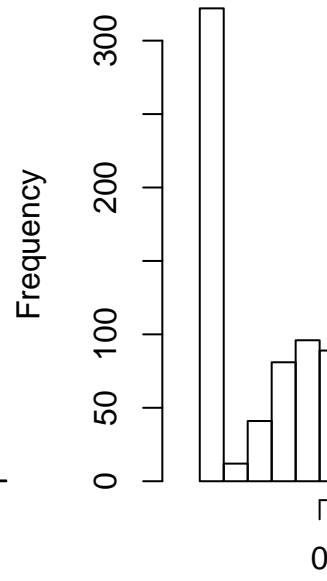
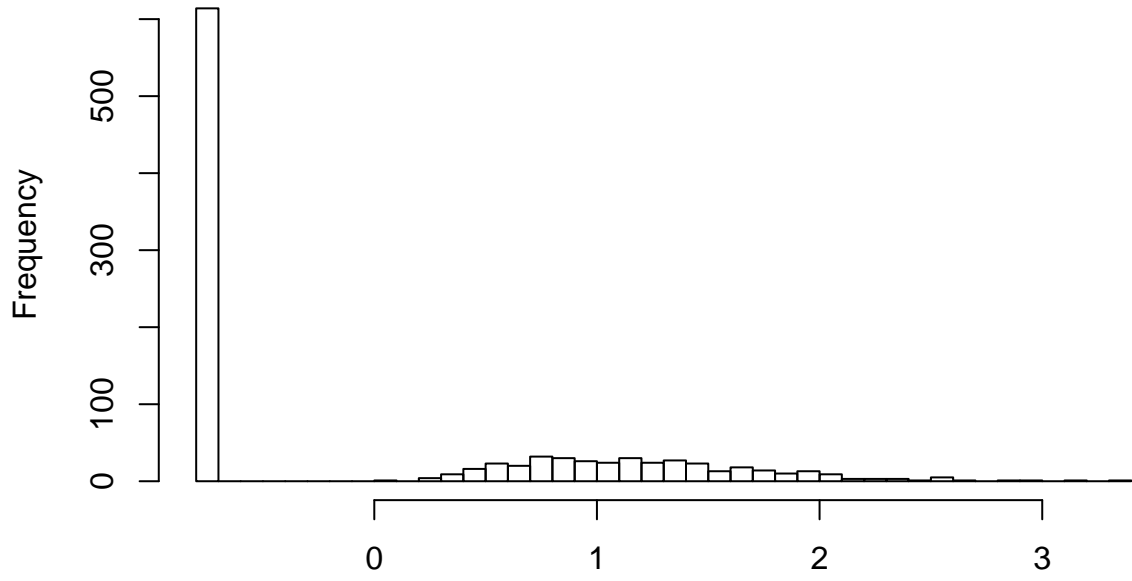




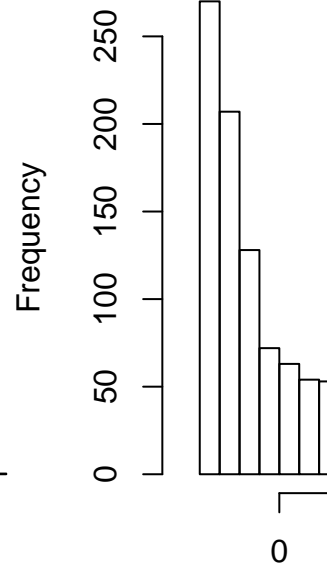
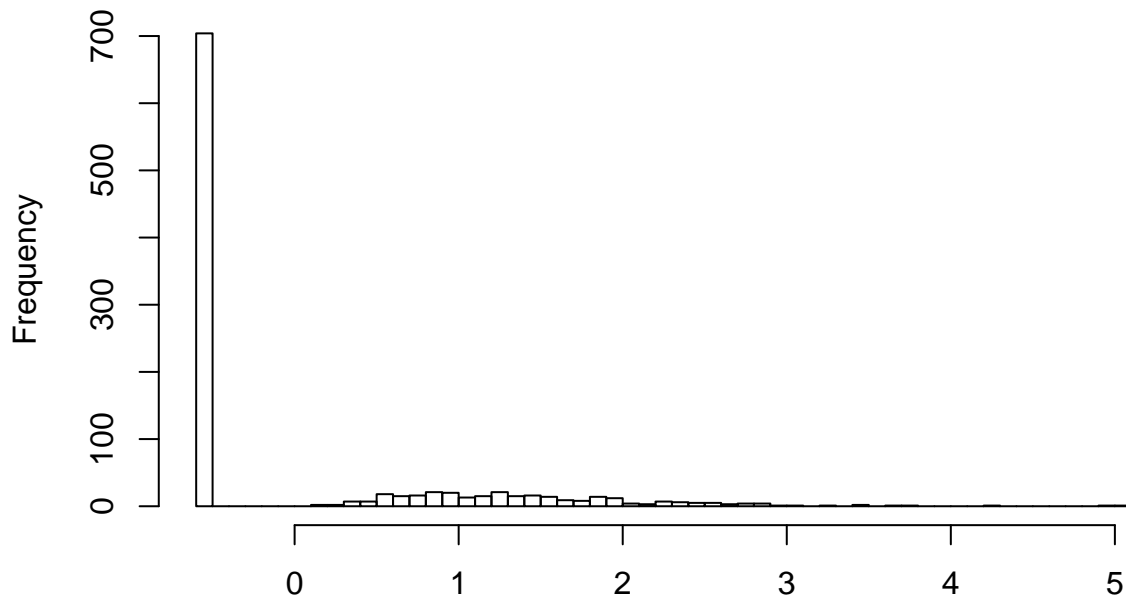
## longmon



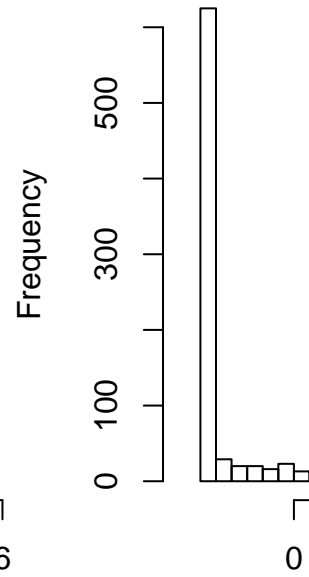
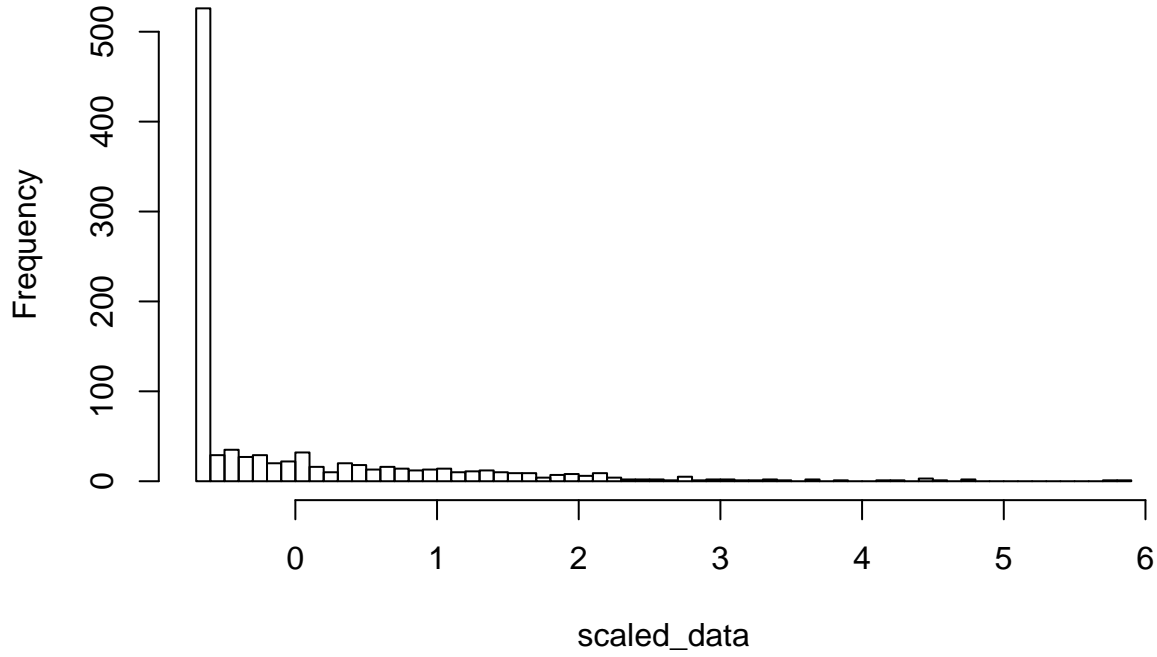
## equipmon



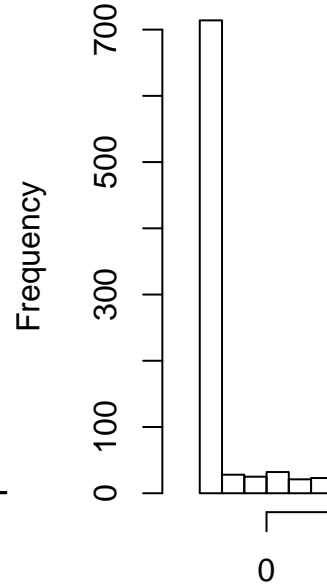
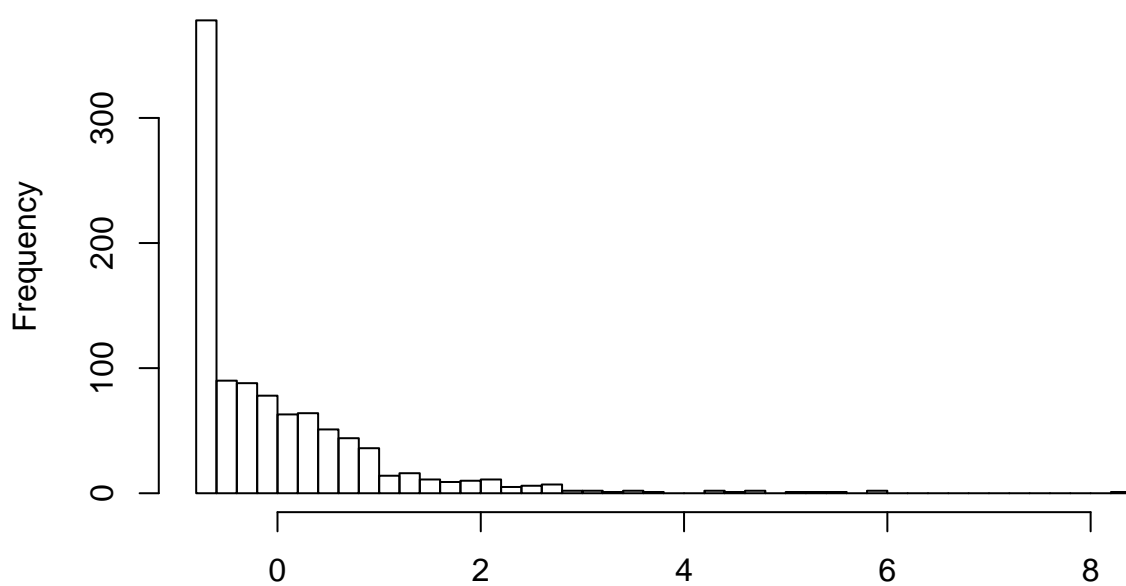
# wiremon



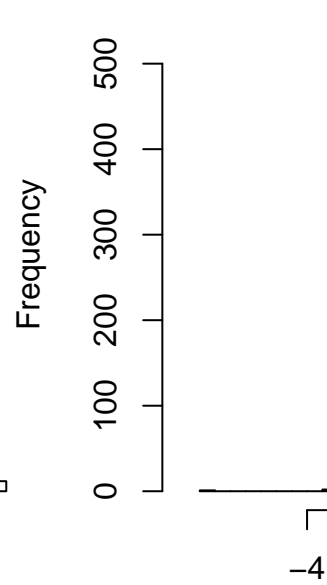
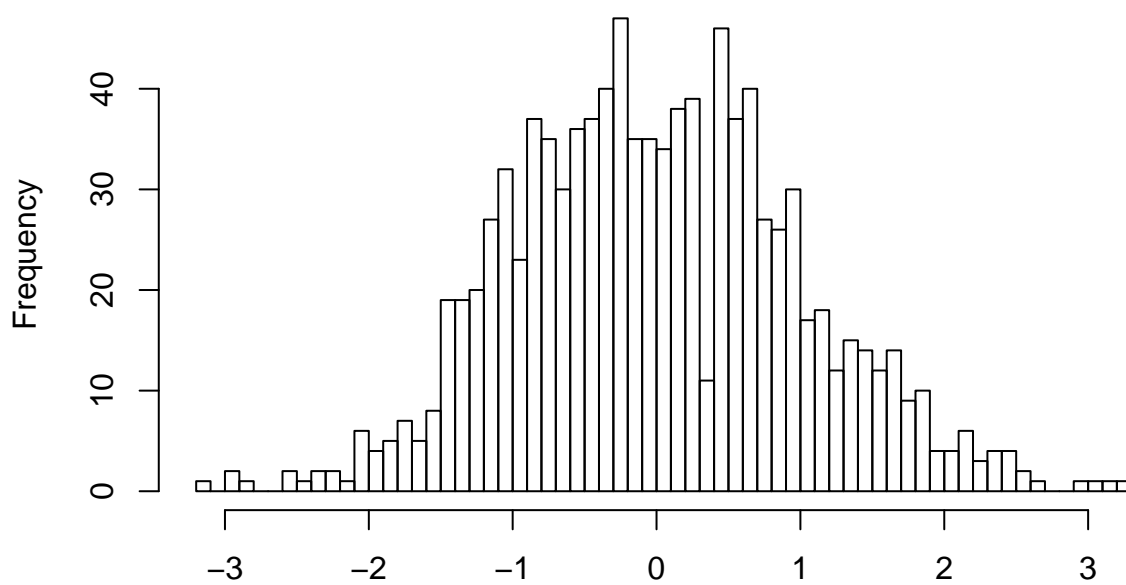
# tollten



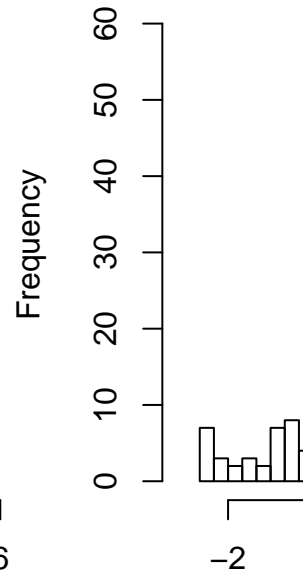
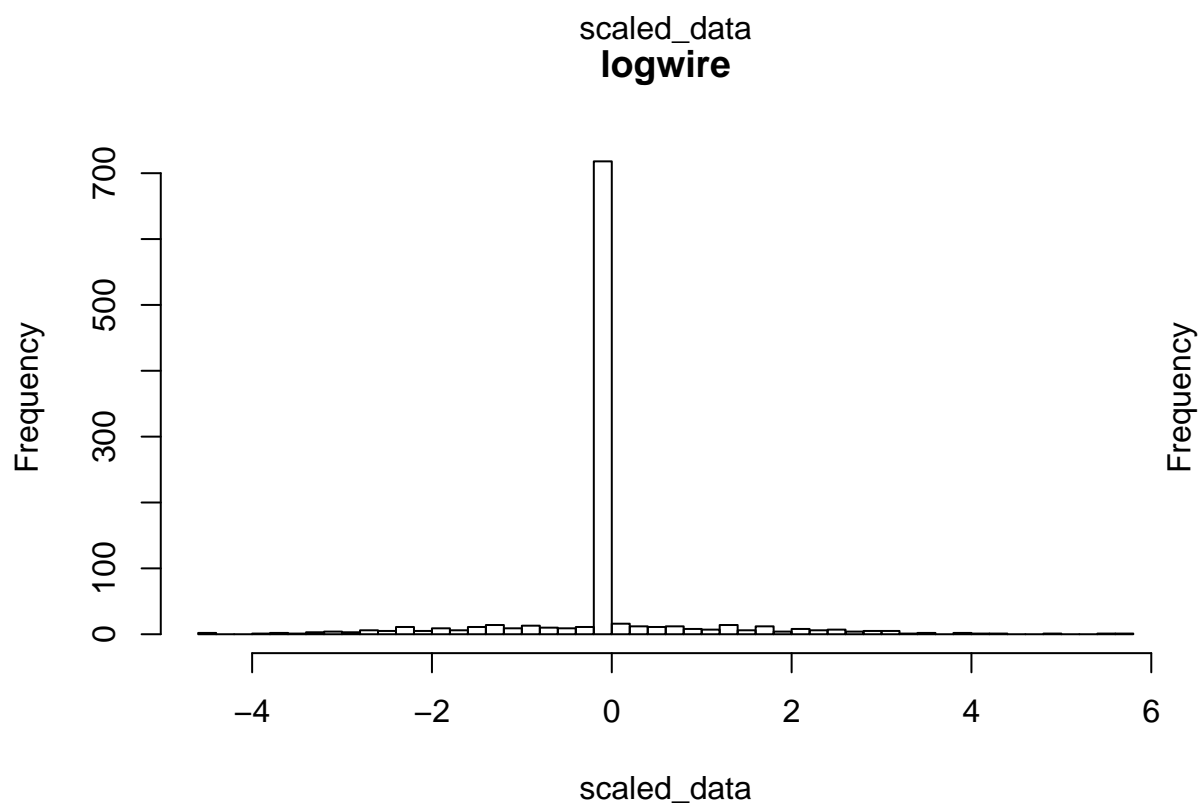
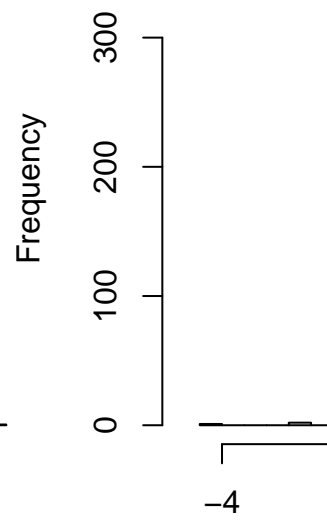
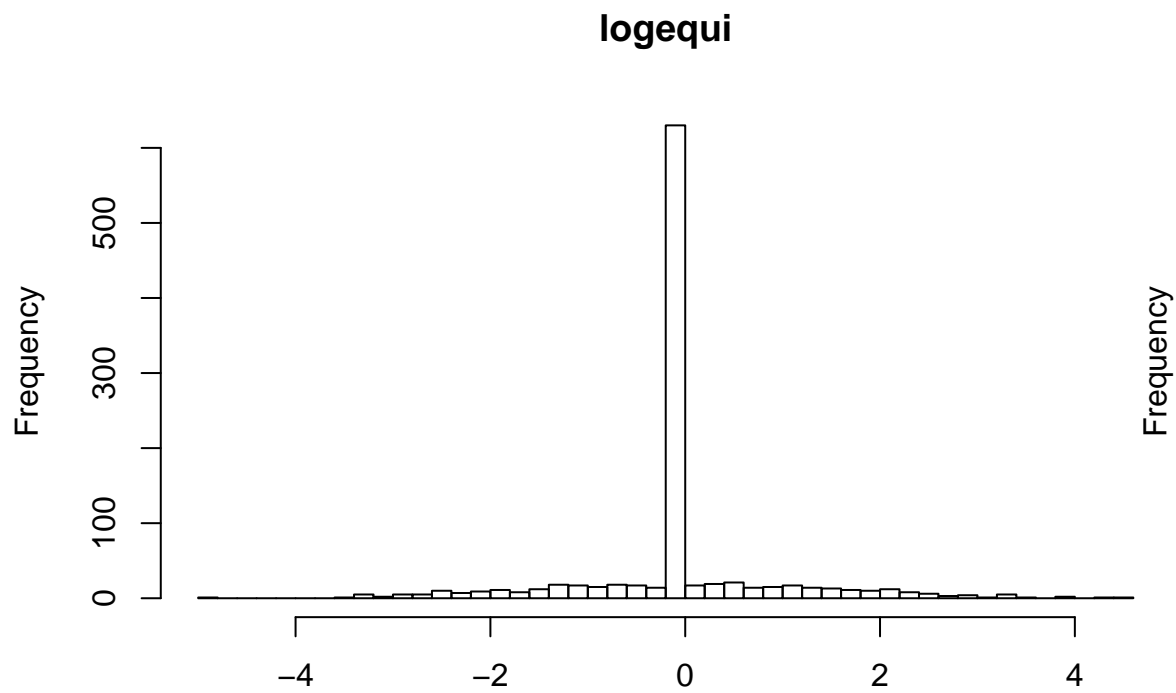
**cardten**



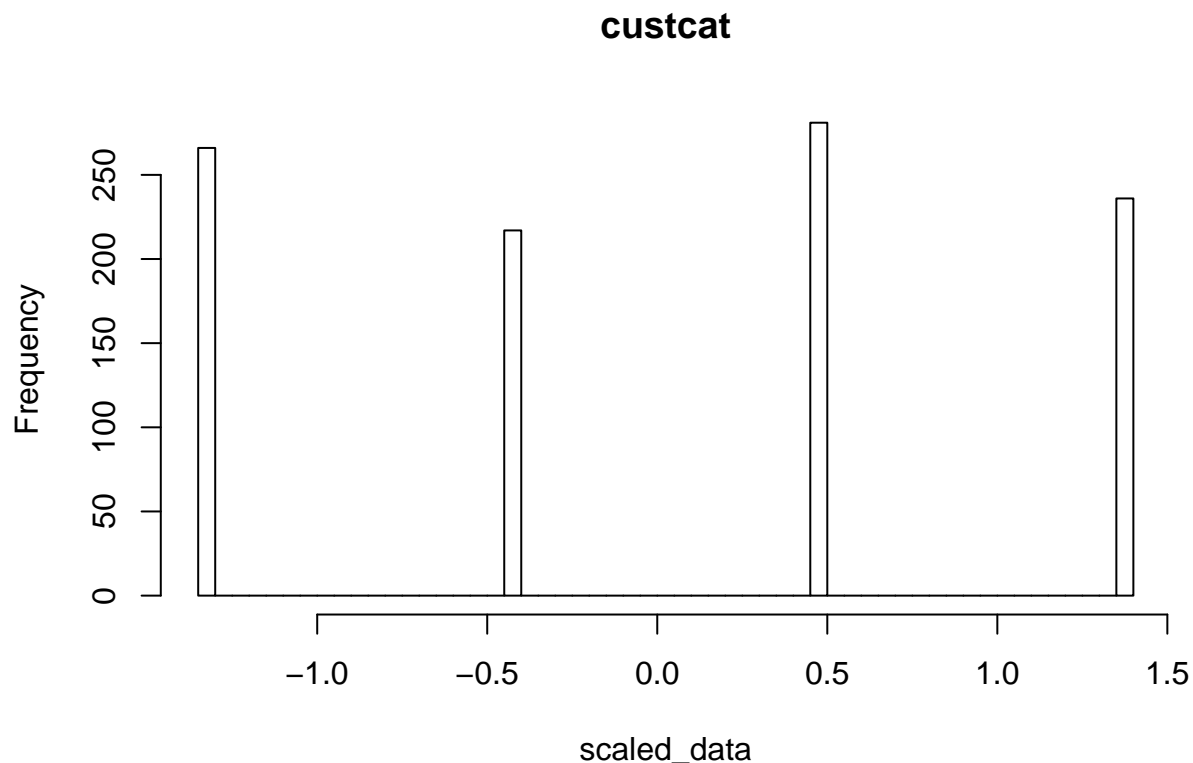
scaled\_data  
**loglong**



scaled\_data







## PCA on Scaled Data

```
##
## Attaching package: 'psych'
## The following object is masked from 'package:Hmisc':
##
##   describe
## The following objects are masked from 'package:ggplot2':
##
##   %+%, alpha
## Principal Components Analysis
## Call: principal(r = data, nfactors = 4, residuals = FALSE, rotate = "none",
##   scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##          PC1  PC2  PC3  PC4    h2  u2 com
## region    0.01 -0.02 -0.02 -0.10 0.0099 0.99 1.2
## tenure    0.55 -0.60  0.34  0.07 0.7865 0.21 2.6
## age       0.37 -0.54  0.14 -0.38 0.5875 0.41 2.8
## marital   0.12 -0.04  0.06  0.51 0.2820 0.72 1.2
## address   0.35 -0.53  0.18 -0.22 0.4798 0.52 2.4
## income    0.33 -0.19  0.06 -0.62 0.5308 0.47 1.8
## ed        0.15  0.50  0.32 -0.14 0.3951 0.60 2.1
## employ    0.42 -0.56  0.07 -0.44 0.6903 0.31 2.8
## retire    0.13 -0.31  0.02  0.00 0.1116 0.89 1.4
## gender    0.01 -0.02  0.04 -0.04 0.0035 1.00 2.4
## reside    0.01  0.13 -0.05  0.54 0.3152 0.68 1.1
```

```

## tollfree 0.60 0.14 -0.54 -0.06 0.6779 0.32 2.1
## equip 0.17 0.65 0.50 -0.05 0.7123 0.29 2.0
## callcard 0.58 -0.22 -0.11 0.09 0.4102 0.59 1.4
## wireless 0.63 0.55 -0.02 -0.04 0.6939 0.31 2.0
## longmon 0.50 -0.59 0.35 0.11 0.7318 0.27 2.7
## tollmon 0.70 -0.01 -0.45 0.03 0.7012 0.30 1.7
## equipmon 0.36 0.68 0.48 -0.02 0.8238 0.18 2.4
## cardmon 0.59 -0.30 0.03 0.34 0.5519 0.45 2.2
## wiremon 0.73 0.51 0.06 0.00 0.7927 0.21 1.8
## longten 0.52 -0.60 0.35 0.10 0.7610 0.24 2.7
## tollten 0.74 -0.21 -0.23 0.01 0.6435 0.36 1.4
## equipten 0.51 0.36 0.56 0.04 0.7076 0.29 2.7
## cardten 0.57 -0.49 0.25 0.30 0.7198 0.28 2.9
## wireten 0.74 0.27 0.18 0.01 0.6510 0.35 1.4
## multline 0.40 0.05 0.49 0.06 0.4009 0.60 2.0
## voice 0.55 0.48 -0.01 -0.02 0.5344 0.47 2.0
## pager 0.57 0.51 -0.03 -0.02 0.5814 0.42 2.0
## internet 0.17 0.60 0.40 -0.06 0.5547 0.45 2.0
## callid 0.60 0.17 -0.51 -0.04 0.6438 0.36 2.1
## callwait 0.60 0.13 -0.52 -0.02 0.6445 0.36 2.1
## forward 0.60 0.15 -0.48 0.02 0.6144 0.39 2.1
## confer 0.61 0.11 -0.46 -0.03 0.5916 0.41 1.9
## ebill 0.10 0.57 0.40 -0.11 0.5050 0.50 2.0
## loglong 0.54 -0.59 0.34 0.12 0.7619 0.24 2.7
## logtoll 0.43 -0.23 0.00 0.13 0.2536 0.75 1.7
## logequi 0.59 0.18 0.02 0.06 0.3875 0.61 1.2
## logcard 0.28 -0.19 0.11 0.37 0.2628 0.74 2.6
## logwire 0.39 0.02 0.18 0.08 0.1960 0.80 1.5
## lninc 0.41 -0.20 0.11 -0.60 0.5917 0.41 2.1
## custcat 0.76 0.32 -0.27 0.03 0.7560 0.24 1.6
## churn -0.15 0.46 0.04 -0.02 0.2341 0.77 1.2
##
## PC1 PC2 PC3 PC4
## SS loadings 9.82 6.51 3.78 2.18
## Proportion Var 0.23 0.15 0.09 0.05
## Cumulative Var 0.23 0.39 0.48 0.53
## Proportion Explained 0.44 0.29 0.17 0.10
## Cumulative Proportion 0.44 0.73 0.90 1.00
##
## Mean item complexity = 2
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is 0.06
## with the empirical chi square 6880.17 with prob < 0
##
## Fit based upon off diagonal values = 0.95

```

Result: As a result of PCA, Components having highest variability and which will show the same variation as complete data according to PCA are custcat, equipmon, equipten, marital