Example 2: real data report: data set for wage index calculations

Data set and main information needed

about: time-variables, modeled - variables, imputed-variables

For this report, a large sample from the wage data set is used, for speed and memory reasons but this can be easily run for the whole set.

Output of this report: pdf, html, word.

Plots are now static. We will produce soon interactive ones.

Dashboards are possible

Main packages and resources:

ggplot2, DataExplorer, funModeling, tabplot, forecast, tsfeatures, anomalous

view_data

Overview of main data-set characteristics

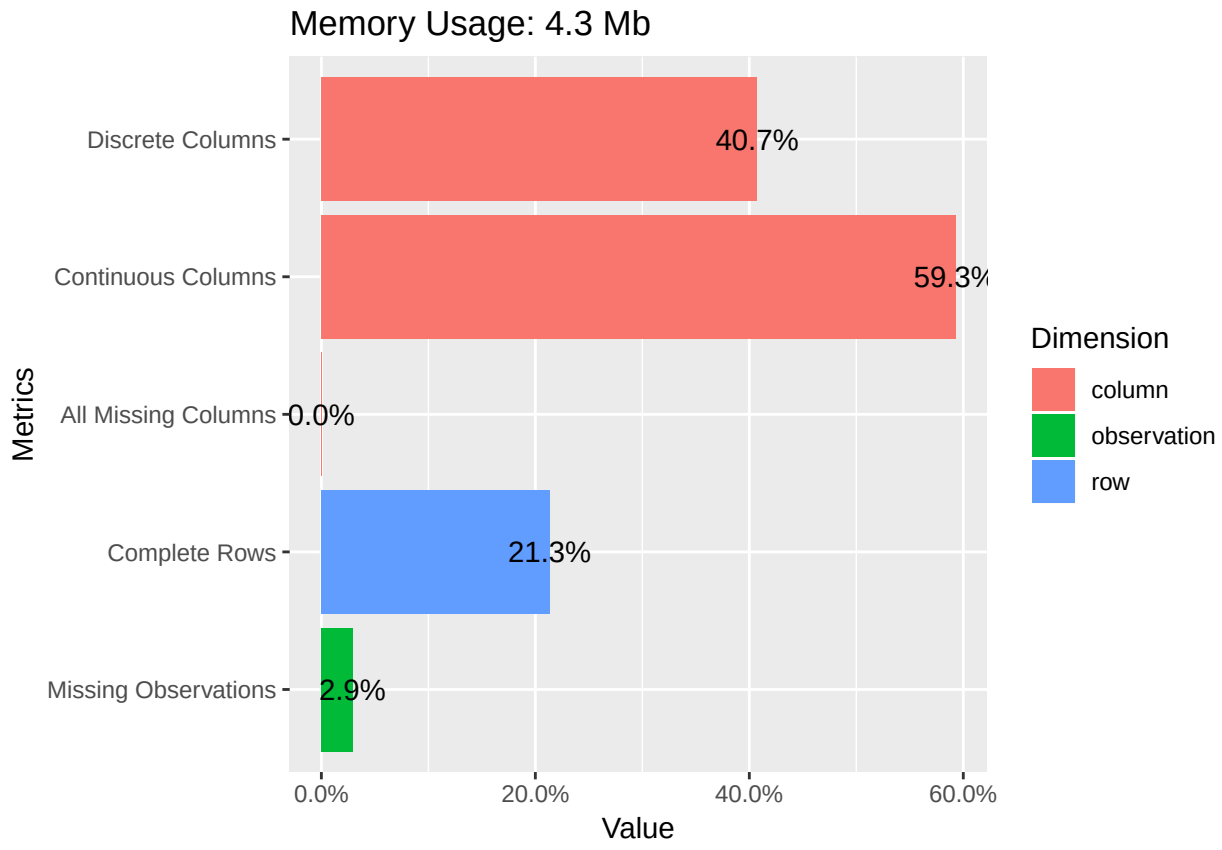
```
## df
##
##  27  Variables      27094  Observations
## ---------------------------------------------------------------------------
## IST
##        n  missing distinct
##    27094        0       28
##
## lowest : 38110 49310 52210 63990 81290, highest: 90040 91010 91020 91040 93110
## ---------------------------------------------------------------------------
## STA
##        n  missing distinct
##    27094        0      168
##
## lowest : 00000 11200 12100 12101 12230, highest: 92111 92121 93110 93120 93121
## ---------------------------------------------------------------------------
## ste
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    27094        0       38    0.961    487.1    319.4       20       20
##      .25      .50      .75      .90      .95
##      112      629      650      949      950
##
## lowest :   0   1   3   4   5, highest: 931 949 950 952 963
## ---------------------------------------------------------------------------
## BSRcell
##        n  missing distinct
##    27094        0        7
##
## Value       ADRIR    ASI    BHM   BSRB HSKOLI     KI   UTAN
## Frequency      24   5851   1705  11200    350   6828   1136
```

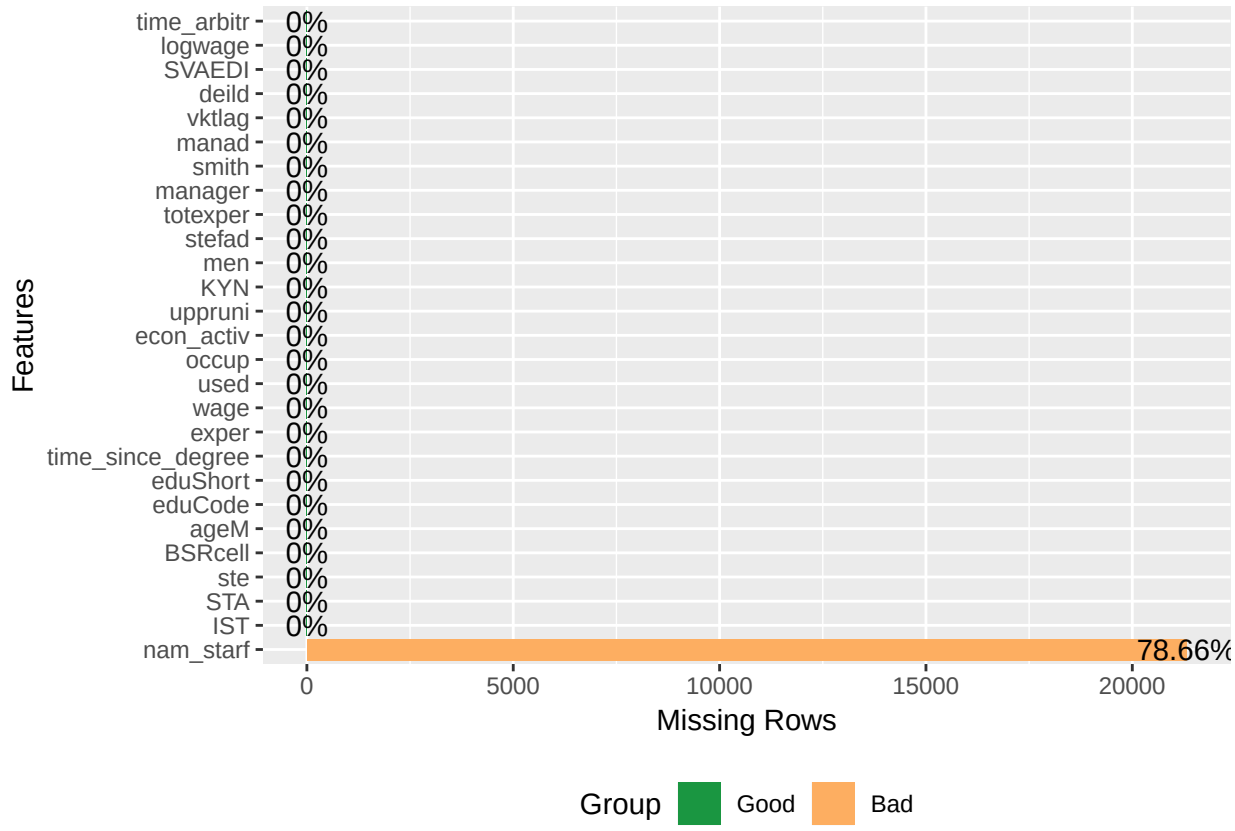
```
## Proportion  0.001  0.216  0.063  0.413  0.013  0.252  0.042
## ---------------------------------------------------------------------------
## ageM
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    27094        0      717        1    495.7      205    234.0    261.0
##      .25      .50      .75      .90      .95
##    336.2    495.0    648.0    738.0    774.0
##
## lowest : 162 163 164 165 166, highest: 906 919 923 932 971
## ---------------------------------------------------------------------------
## eduCode
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    27094        0       29    0.975    43.59    20.24       22       22
##      .25      .50      .75      .90      .95
##       25       36       61       62       72
##
## lowest : 11 20 21 22 23, highest: 60 61 62 71 72
## ---------------------------------------------------------------------------
## eduShort
##        n  missing distinct
##    27094        0        4
##
## Value      a_low   low   med vhigh
## Frequency   7258  8775  8725  2336
## Proportion 0.268 0.324 0.322 0.086
## ---------------------------------------------------------------------------
## time_since_degree
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    27094        0      683        1    189.2    192.1        3       13
##      .25      .50      .75      .90      .95
##       41      124      321      467      530
##
## lowest : -11 -10  -9  -8  -7, highest: 686 690 707 715 727
## ---------------------------------------------------------------------------
## exper
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    27094        0      538        1    99.18    80.56      4.0     11.0
##      .25      .50      .75      .90      .95
##     42.0     91.0    137.0    180.0    222.3
##
## lowest :   0   1   2   3   4, highest: 608 611 644 647 652
## ---------------------------------------------------------------------------
## wage
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    27094        0    17241        1     1859    748.7      979     1141
##      .25      .50      .75      .90      .95
##     1388     1765     2197     2791     3117
##
## lowest :  300.0000  324.8521  325.0000  325.0068  325.0069
## highest: 6337.4165 6407.1396 6410.8364 6702.1284 7345.3491
## ---------------------------------------------------------------------------
## used
##        n  missing distinct
##    27094        0        8
```

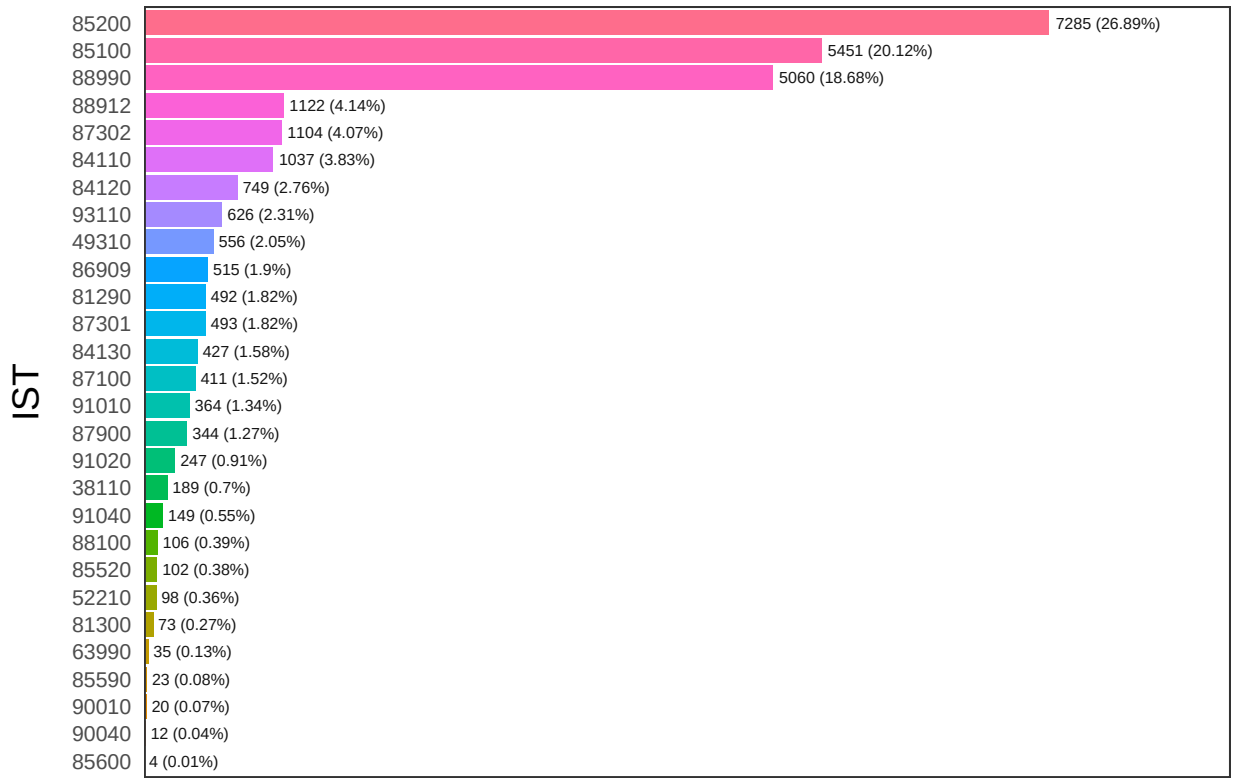
```
##
## Value          1     2     3     4     5     6     7     8
## Frequency  20954   142     9  2322   488   865   466  1848
## Proportion 0.773 0.005 0.000 0.086 0.018 0.032 0.017 0.068
## ---------------------------------------------------------------------------
## occup
##        n  missing distinct
##    27094        0        9
##
## Value          1     2     3     4     5     6   IDN  VERK     X
## Frequency   1488  8597  2332   797 10933    25    73  2766    83
## Proportion 0.055 0.317 0.086 0.029 0.404 0.001 0.003 0.102 0.003
## ---------------------------------------------------------------------------
## econ_activ
##        n  missing distinct
##    27094        0        8
##
## Value          E     H     J     N     O     P     Q     R
## Frequency    189   654    35   565  2213 12865  9155  1418
## Proportion 0.007 0.024 0.001 0.021 0.082 0.475 0.338 0.052
## ---------------------------------------------------------------------------
## uppruni
##        n  missing distinct
##    27094        0        2
##
## Value          0     1
## Frequency  24534  2560
## Proportion 0.906 0.094
## ---------------------------------------------------------------------------
## nam_starf
##        n  missing distinct    value
##     5783    21311        1        1
##
## Value         1
## Frequency  5783
## Proportion    1
## ---------------------------------------------------------------------------
## KYN
##        n  missing distinct     Info     Mean      Gmd
##    27094        0        2    0.583    1.736   0.3886
##
## Value          1     2
## Frequency   7153 19941
## Proportion 0.264 0.736
## ---------------------------------------------------------------------------
## men
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    27094        0       29    0.975    43.59    20.24       22       22
##      .25      .50      .75      .90      .95
##       25       36       61       62       72
##
## lowest : 11 20 21 22 23, highest: 60 61 62 71 72
## ---------------------------------------------------------------------------
## stefad
```

```
##        n  missing distinct     Info      Sum     Mean      Gmd
##    27094        0        2    0.099    26171   0.9659  0.06581
##
## ------------------------------------------------------------------------
## totexper
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    27094        0       44    0.999    20.19    12.51        2        5
##      .25      .50      .75      .90      .95
##       10       23       30       33       35
##
## lowest : -7 -6 -5 -4 -3, highest: 32 33 34 35 36
## ------------------------------------------------------------------------
## manager
##        n  missing distinct     Info      Sum     Mean      Gmd
##    27094        0        2    0.293     2976   0.1098   0.1956
##
## ------------------------------------------------------------------------
## smith
##        n  missing distinct     Info      Sum     Mean      Gmd
##    27094        0        2     0.02      186 0.006865  0.01364
##
## ------------------------------------------------------------------------
## manad
##        n  missing distinct     Info      Sum     Mean      Gmd
##    27094        0        2    0.149    25677   0.9477  0.09913
##
## ------------------------------------------------------------------------
## vktlag
##        n  missing distinct     Info      Sum     Mean      Gmd
##    27094        0        2    0.548     6524   0.2408   0.3656
##
## ------------------------------------------------------------------------
## deild
##        n  missing distinct
##    27094        0       13
##
## Value         38    49    52    63    81    84    85    86    87    88
## Frequency    189   556    98    35   565  2213 12865   515  2352  6288
## Proportion 0.007 0.021 0.004 0.001 0.021 0.082 0.475 0.019 0.087 0.232
##
## Value         90    91    93
## Frequency     32   760   626
## Proportion 0.001 0.028 0.023
## ------------------------------------------------------------------------
## SVAEDI
##        n  missing distinct    value
##    27094        0        1        1
##
## Value          1
## Frequency  27094
## Proportion     1
## ------------------------------------------------------------------------
## logwage
##        n  missing distinct     Info     Mean      Gmd      .05      .10
```

```
##    27094        0    17241        1    7.455   0.4271    6.887    7.040
##      .25      .50      .75      .90      .95
##    7.236    7.476    7.695    7.934    8.045
##
## lowest : 5.703782 5.783370 5.783825 5.783846 5.783846
## highest: 8.754226 8.765168 8.765745 8.810180 8.901823
## ------------------------------------------------------------------------------
## time_arbitr
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    27094        0      120        1    157.7    39.65      103      109
##      .25      .50      .75      .90      .95
##      128      158      187      205      211
##
## lowest :  97  98  99 100 101, highest: 212 213 214 215 216
## ------------------------------------------------------------------------------
```




time_arbitr
logwage
SVAEDI
deild
vktlag
manad
smith
manager
totexper
stefad
men
KYN
uppruni
econ_activ
occup
used
wage
exper
time_since_degree
eduShort
eduCode
ageM
BSRcell
ste
STA
IST
nam_starf
0%
78.66%
Features
0
5000
10000
15000
20000
Missing Rows
Group
Good
Bad



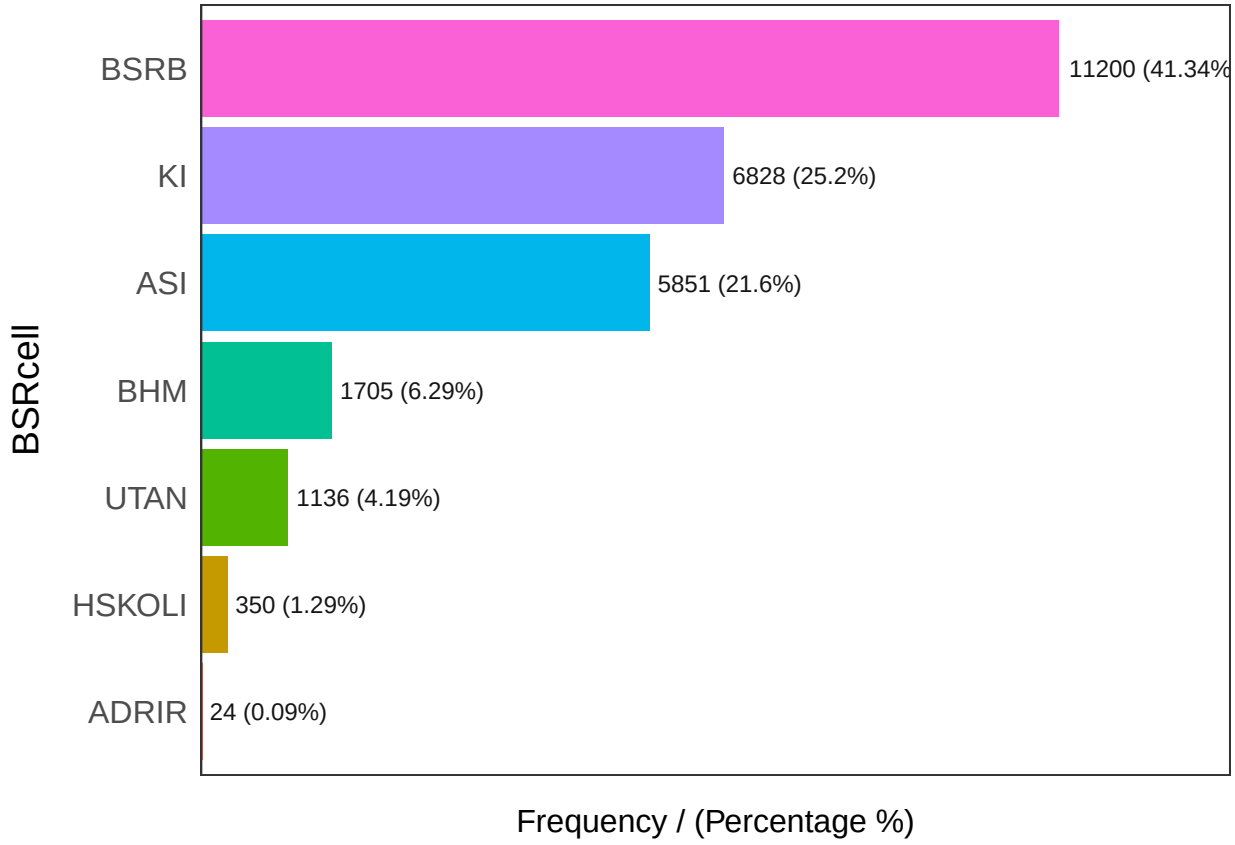

```
##       IST frequency percentage cumulative_perc
## 1   85200      7285      26.89           26.89
## 2   85100      5451      20.12           47.01
## 3   88990      5060      18.68           65.69
## 4   88912      1122       4.14           69.83
## 5   87302      1104       4.07           73.90
## 6   84110      1037       3.83           77.73
## 7   84120       749       2.76           80.49
## 8   93110       626       2.31           82.80
## 9   49310       556       2.05           84.85
## 10  86909       515       1.90           86.75
## 11  87301       493       1.82           88.57
## 12  81290       492       1.82           90.39
## 13  84130       427       1.58           91.97
## 14  87100       411       1.52           93.49
## 15  91010       364       1.34           94.83
## 16  87900       344       1.27           96.10
## 17  91020       247       0.91           97.01
## 18  38110       189       0.70           97.71
## 19  91040       149       0.55           98.26
## 20  88100       106       0.39           98.65
## 21  85520       102       0.38           99.03
## 22  52210        98       0.36           99.39
## 23  81300        73       0.27           99.66
## 24  63990        35       0.13           99.79
## 25  85590        23       0.08           99.87
```

```
## 26 90010        20       0.07           99.94
## 27 90040        12       0.04           99.98
## 28 85600         4       0.01          100.00
##
##       STA frequency percentage cumulative_perc
## 1   51310      5855      21.61           21.61
## 2   23310      3532      13.04           34.65
## 3   51330      2702       9.97           44.62
## 4   92110      1270       4.69           49.31
## 5   12290       934       3.45           52.76
## 6   23410       927       3.42           56.18
## 7   33300       825       3.04           59.22
## 8   23320       807       2.98           62.20
## 9   51690       625       2.31           64.51
## 10  51320       550       2.03           66.54
## 11  23321       537       1.98           68.52
## 12  32310       422       1.56           70.08
## 13  83230       381       1.41           71.49
## 14  24460       377       1.39           72.88
## 15  23420       355       1.31           74.19
## 16  91320       354       1.31           75.50
## 17  51220       336       1.24           76.74
## 18  41900       335       1.24           77.98
## 19  12390       253       0.93           78.91
## 20  24320       252       0.93           79.84
## 21  33400       244       0.90           80.74
## 22  24190       233       0.86           81.60
## 23  34600       213       0.79           82.39
## 24  51311       212       0.78           83.17
## 25  93120       208       0.77           83.94
## 26  23510       203       0.75           84.69
## 27  23311       192       0.71           85.40
## 28  22300       175       0.65           86.05
## 29  23590       159       0.59           86.64
## 30  51640       155       0.57           87.21
## 31  92111       148       0.55           87.76
## 32  12291       145       0.54           88.30
## 33  91610       132       0.49           88.79
## 34  42220       130       0.48           89.27
## 35  34370       117       0.43           89.70
## 36  24450       108       0.40           90.10
## 37  24120       106       0.39           90.49
## 38  42230       103       0.38           90.87
## 39  31200       101       0.37           91.24
## 40  51420        96       0.35           91.59
## 41  51390        95       0.35           91.94
## 42  34330        84       0.31           92.25
## 43  52210        84       0.31           92.56
## 44  00000        83       0.31           92.87
## 45  24130        81       0.30           93.17
## 46  24290        81       0.30           93.47
## 47  23421        78       0.29           93.76
## 48  51222        78       0.29           94.05
## 49  41210        75       0.28           94.33
```

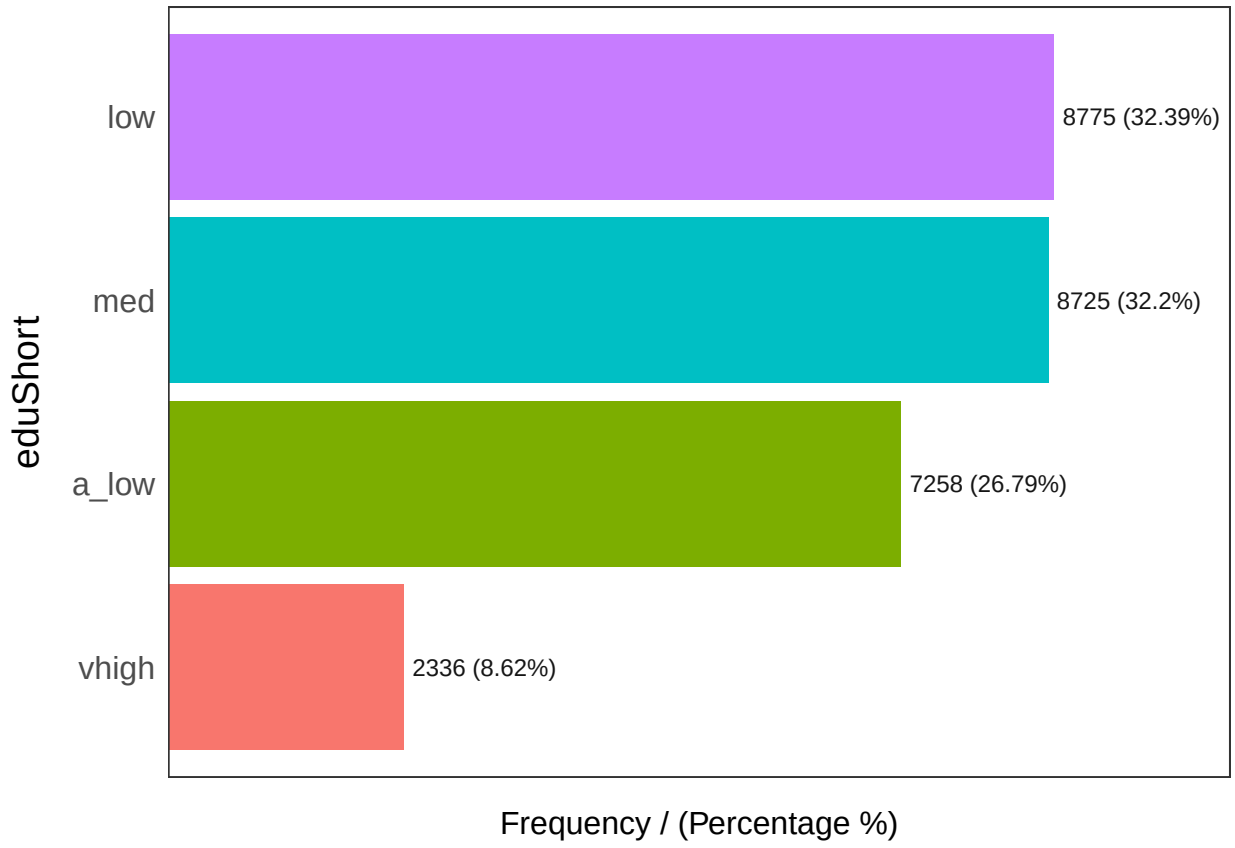
```
## 50  91530        71       0.26          94.59
## 51  31120        68       0.25          94.84
## 52  21420        67       0.25          95.09
## 53  34310        65       0.24          95.33
## 54  32220        60       0.22          95.55
## 55  51641        49       0.18          95.73
## 56  12100        47       0.17          95.90
## 57  21490        46       0.17          96.07
## 58  41410        42       0.16          96.23
## 59  12310        41       0.15          96.38
## 60  93121        38       0.14          96.52
## 61  24420        32       0.12          96.64
## 62  24310        31       0.11          96.75
## 63  41130        30       0.11          96.86
## 64  71242        29       0.11          96.97
## 65  72310        29       0.11          97.08
## 66  21310        28       0.10          97.18
## 67  91611        28       0.10          97.28
## 68  91420        27       0.10          97.38
## 69  51223        26       0.10          97.48
## 70  83310        24       0.09          97.57
## 71  21390        23       0.08          97.65
## 72  41150        22       0.08          97.73
## 73  21410        21       0.08          97.81
## 74  12320        20       0.07          97.88
## 75  51331        20       0.07          97.95
## 76  31520        19       0.07          98.02
## 77  42210        19       0.07          98.09
## 78  22301        18       0.07          98.16
## 79  34360        18       0.07          98.23
## 80  91520        17       0.06          98.29
## 81  21220        16       0.06          98.35
## 82  24321        16       0.06          98.41
## 83  34390        15       0.06          98.47
## 84  12391        14       0.05          98.52
## 85  22260        14       0.05          98.57
## 86  72422        13       0.05          98.62
## 87  24110        12       0.04          98.66
## 88  41310        12       0.04          98.70
## 89  61290        12       0.04          98.74
## 90  24440        11       0.04          98.78
## 91  34601        11       0.04          98.82
## 92  41901        11       0.04          98.86
## 93  51490        11       0.04          98.90
## 94  51691        11       0.04          98.94
## 95  22110        10       0.04          98.98
## 96  32311        10       0.04          99.02
## 97  61130        10       0.04          99.06
## 98  71243        10       0.04          99.10
## 99  31201         9       0.03          99.13
## 100 42110         9       0.03          99.16
## 101 42150         9       0.03          99.19
## 102 52211         8       0.03          99.22
## 103 12360         7       0.03          99.25
```

```
## 104 24191         7       0.03          99.28
## 105 24430         7       0.03          99.31
## 106 33200         7       0.03          99.34
## 107 34350         7       0.03          99.37
## 108 71240         7       0.03          99.40
## 109 72112         7       0.03          99.43
## 110 72330         7       0.03          99.46
## 111 83220         7       0.03          99.49
## 112 23411         6       0.02          99.51
## 113 34130         6       0.02          99.53
## 114 34361         6       0.02          99.55
## 115 51642         6       0.02          99.57
## 116 11200         5       0.02          99.59
## 117 12101         5       0.02          99.61
## 118 32221         5       0.02          99.63
## 119 32318         5       0.02          99.65
## 120 51130         5       0.02          99.67
## 121 72333         5       0.02          99.69
## 122 12260         4       0.01          99.70
## 123 12280         4       0.01          99.71
## 124 21210         4       0.01          99.72
## 125 51221         4       0.01          99.73
## 126 72130         4       0.01          99.74
## 127 74222         4       0.01          99.75
## 128 91620         4       0.01          99.76
## 129 12230         3       0.01          99.77
## 130 12311         3       0.01          99.78
## 131 24121         3       0.01          99.79
## 132 24210         3       0.01          99.80
## 133 31522         3       0.01          99.81
## 134 34440         3       0.01          99.82
## 135 61291         3       0.01          99.83
## 136 72420         3       0.01          99.84
## 137 12350         2       0.01          99.85
## 138 21411         2       0.01          99.86
## 139 21421         2       0.01          99.87
## 140 21480         2       0.01          99.88
## 141 22290         2       0.01          99.89
## 142 23520         2       0.01          99.90
## 143 24131         2       0.01          99.91
## 144 24451         2       0.01          99.92
## 145 24520         2       0.01          99.93
## 146 31510         2       0.01          99.94
## 147 32230         2       0.01          99.95
## 148 34351         2       0.01          99.96
## 149 34710         2       0.01          99.97
## 150 51131         2       0.01          99.98
## 151 51210         2       0.01          99.99
## 152 71230         2       0.01         100.00
## 153 12330         1       0.00         100.00
## 154 24410         1       0.00         100.00
## 155 24461         1       0.00         100.00
## 156 24510         1       0.00         100.00
## 157 31180         1       0.00         100.00
```

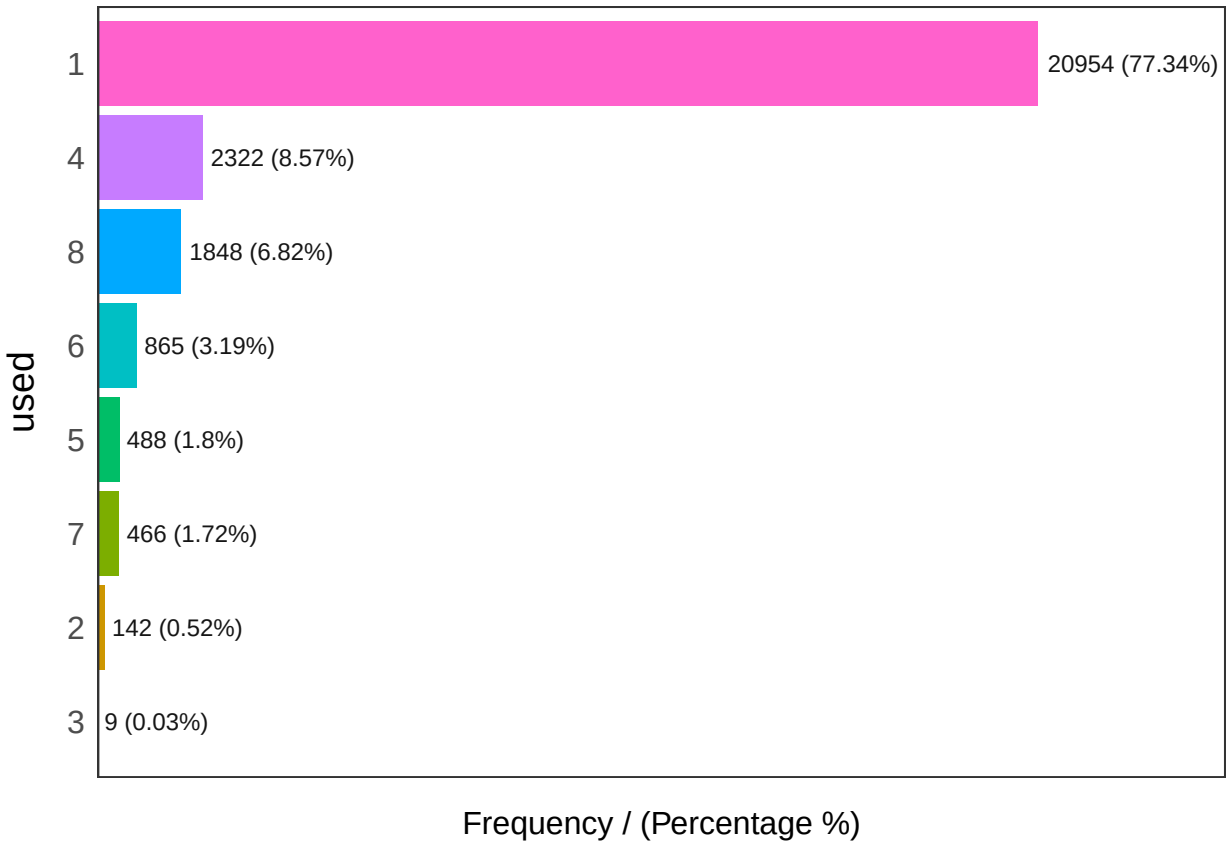
```
## 158 51321         1       0.00          100.00
## 159 71332         1       0.00          100.00
## 160 71333         1       0.00          100.00
## 161 71362         1       0.00          100.00
## 162 71370         1       0.00          100.00
## 163 72332         1       0.00          100.00
## 164 72423         1       0.00          100.00
## 165 72440         1       0.00          100.00
## 166 91000         1       0.00          100.00
## 167 92121         1       0.00          100.00
## 168 93110         1       0.00          100.00
```



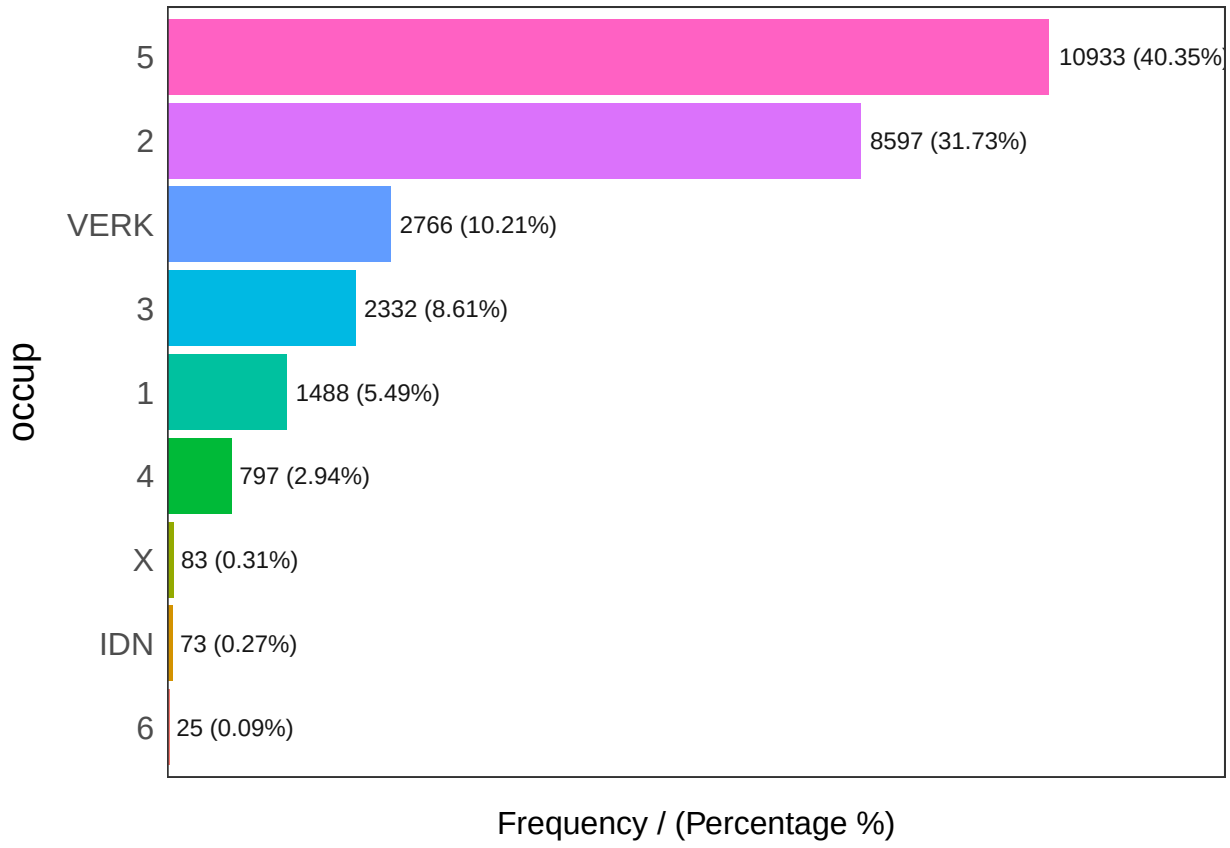

```
##   BSRcell frequency percentage cumulative_perc
## 1    BSRB     11200      41.34           41.34
## 2      KI      6828      25.20           66.54
## 3     ASI      5851      21.60           88.14
## 4     BHM      1705       6.29           94.43
## 5    UTAN      1136       4.19           98.62
## 6  HSKOLI       350       1.29           99.91
## 7   ADRIR        24       0.09          100.00
```



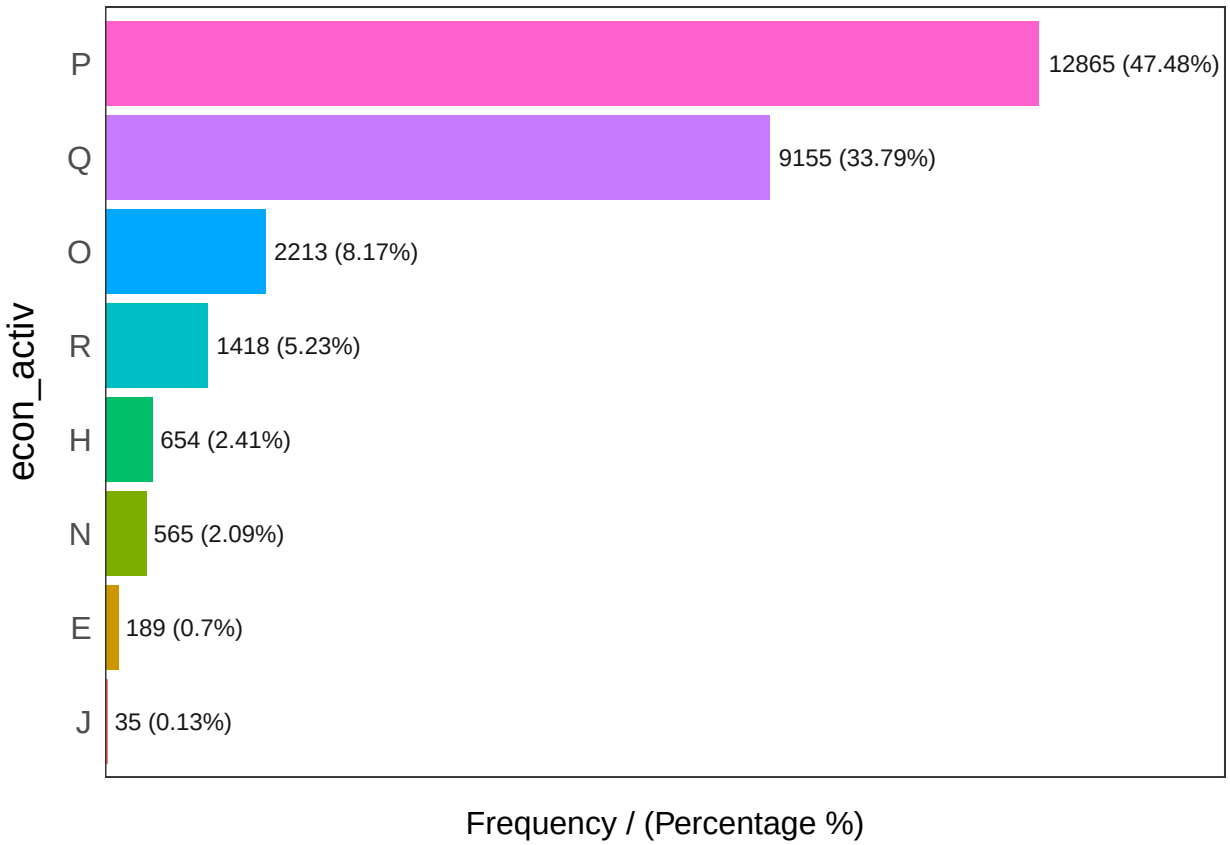

```
##   eduShort frequency percentage cumulative_perc
## 1      low      8775      32.39           32.39
## 2      med      8725      32.20           64.59
## 3    a_low      7258      26.79           91.38
## 4    vhigh      2336       8.62          100.00
```



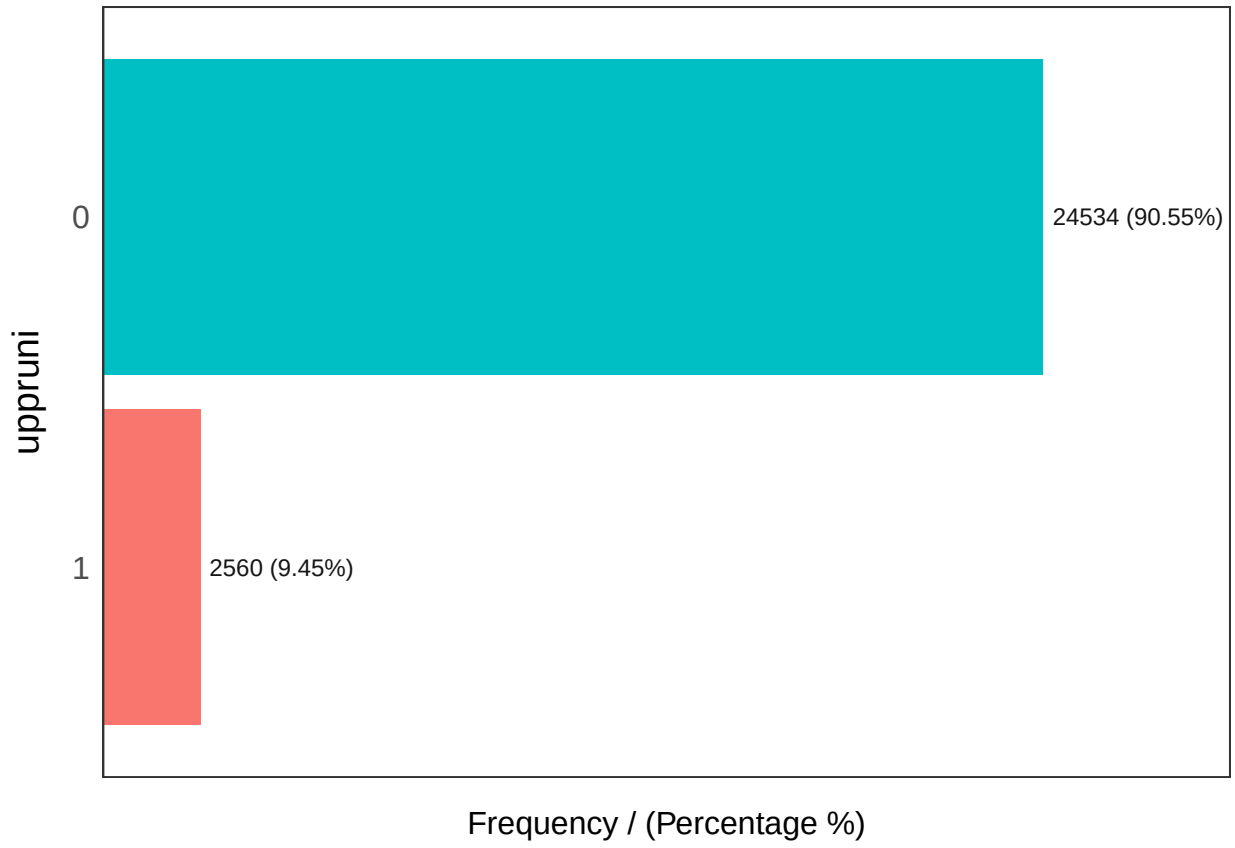

```
##   used frequency percentage cumulative_perc
## 1    1     20954      77.34           77.34
## 2    4      2322       8.57           85.91
## 3    8      1848       6.82           92.73
## 4    6       865       3.19           95.92
## 5    5       488       1.80           97.72
## 6    7       466       1.72           99.44
## 7    2       142       0.52           99.96
## 8    3         9       0.03          100.00
```



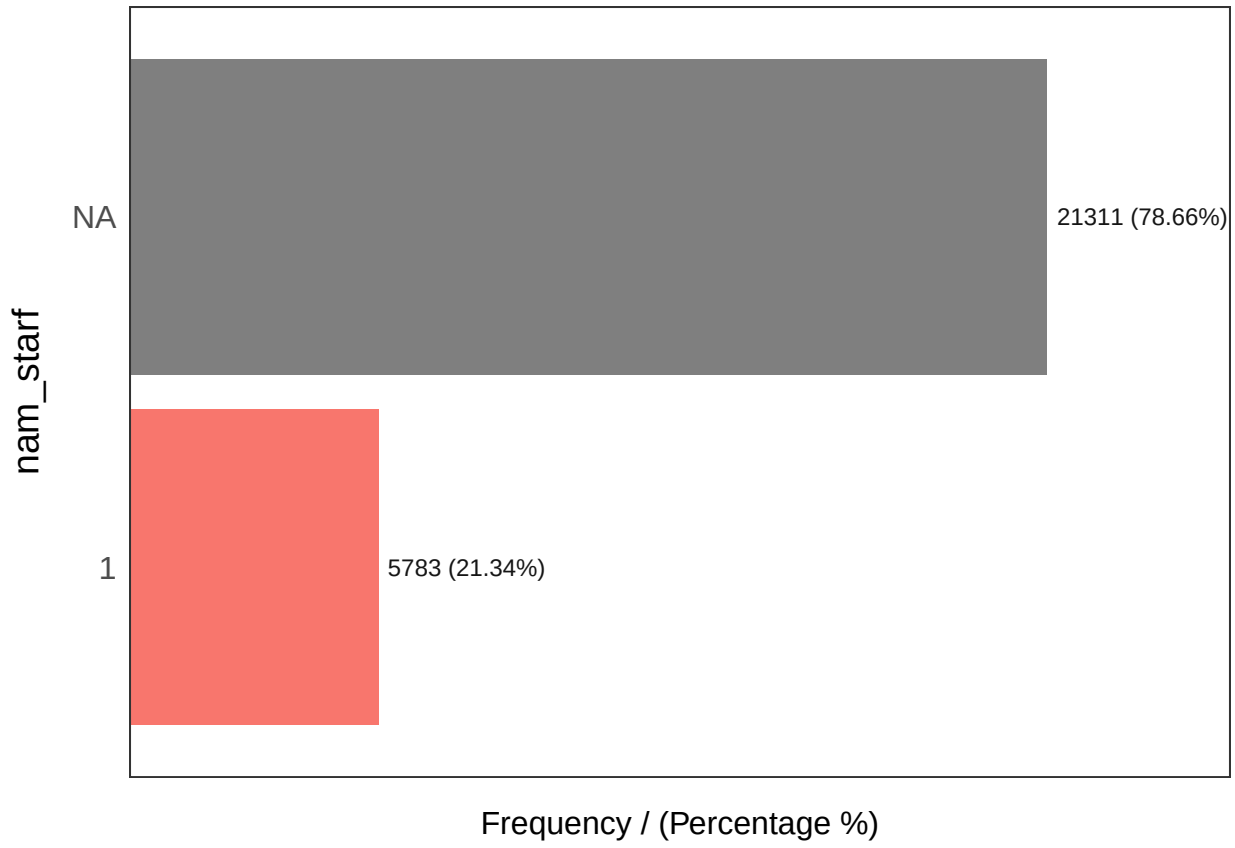

```
##   occup frequency percentage cumulative_perc
## 1     5     10933      40.35           40.35
## 2     2      8597      31.73           72.08
## 3  VERK      2766      10.21           82.29
## 4     3      2332       8.61           90.90
## 5     1      1488       5.49           96.39
## 6     4       797       2.94           99.33
## 7     X        83       0.31           99.64
## 8   IDN        73       0.27           99.91
## 9     6        25       0.09          100.00
```



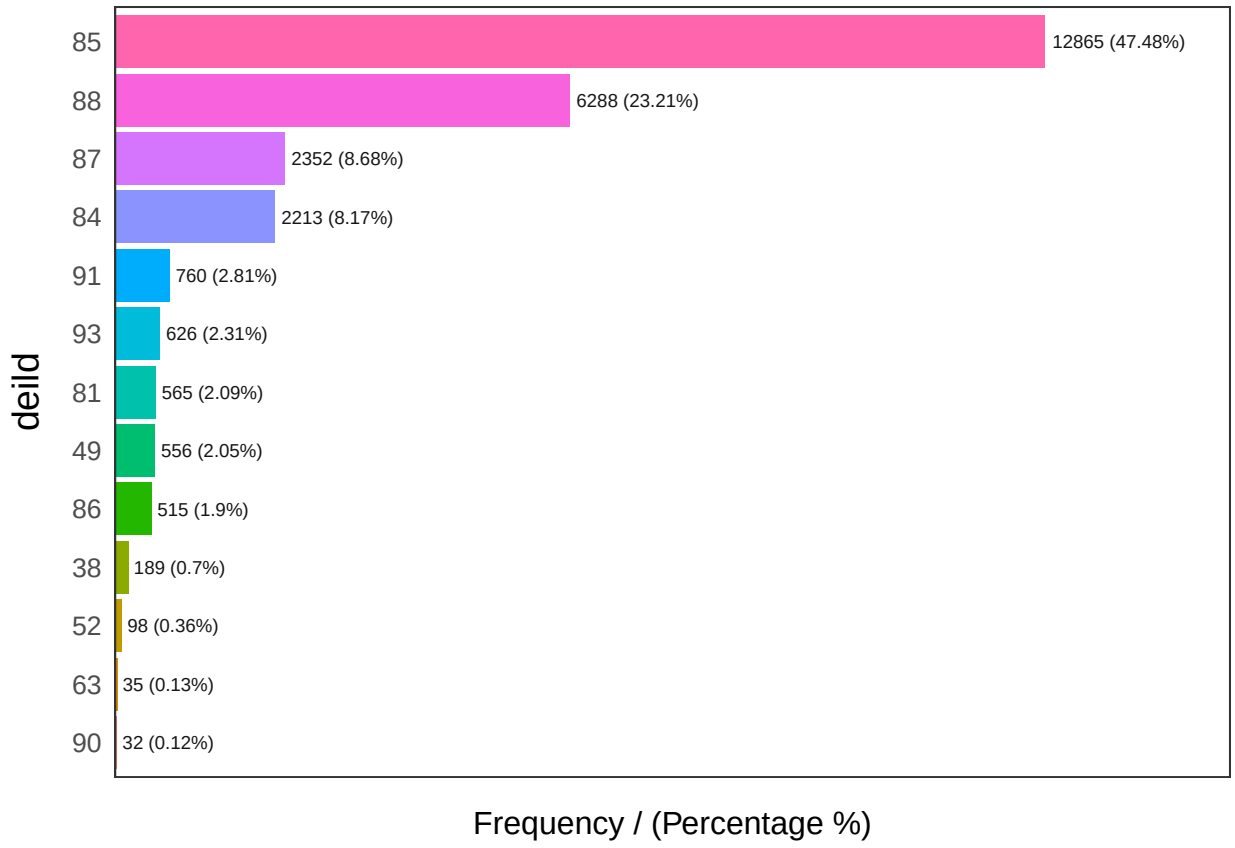

```
##   econ_activ frequency percentage cumulative_perc
## 1          P     12865      47.48           47.48
## 2          Q      9155      33.79           81.27
## 3          O      2213       8.17           89.44
## 4          R      1418       5.23           94.67
## 5          H       654       2.41           97.08
## 6          N       565       2.09           99.17
## 7          E       189       0.70           99.87
## 8          J        35       0.13          100.00
```



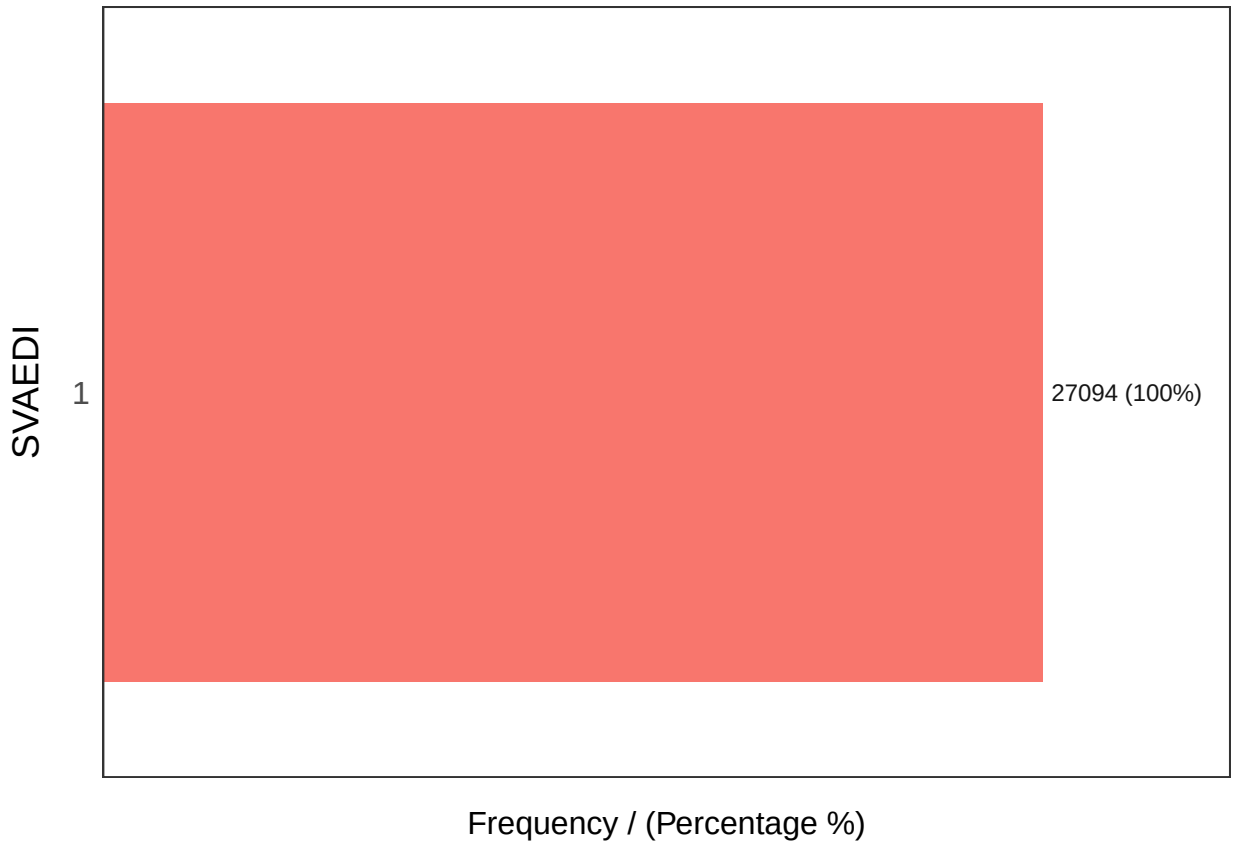

```
##   uppruni frequency percentage cumulative_perc
## 1       0     24534      90.55           90.55
## 2       1      2560       9.45          100.00
```




```
##   nam_starf frequency percentage cumulative_perc
## 1      <NA>     21311      78.66           78.66
## 2         1      5783      21.34          100.00
```

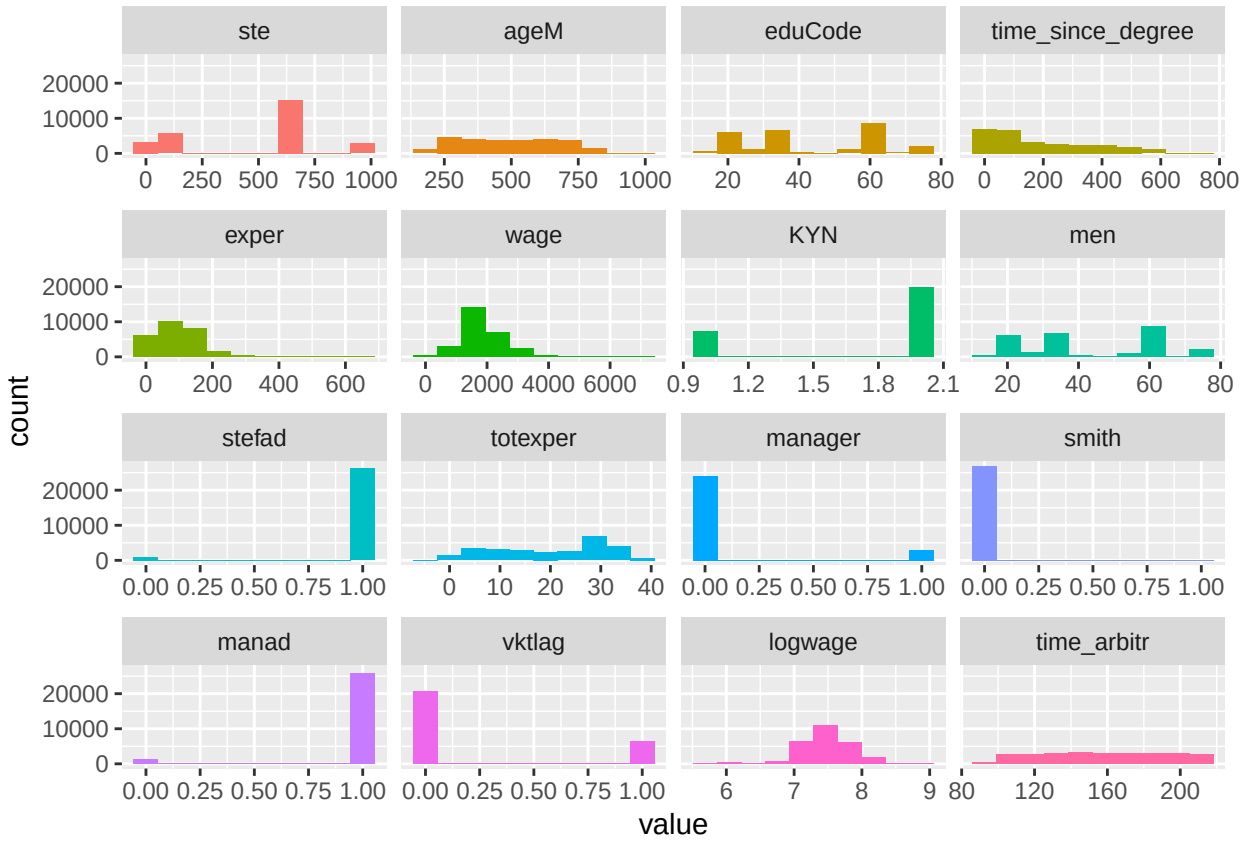


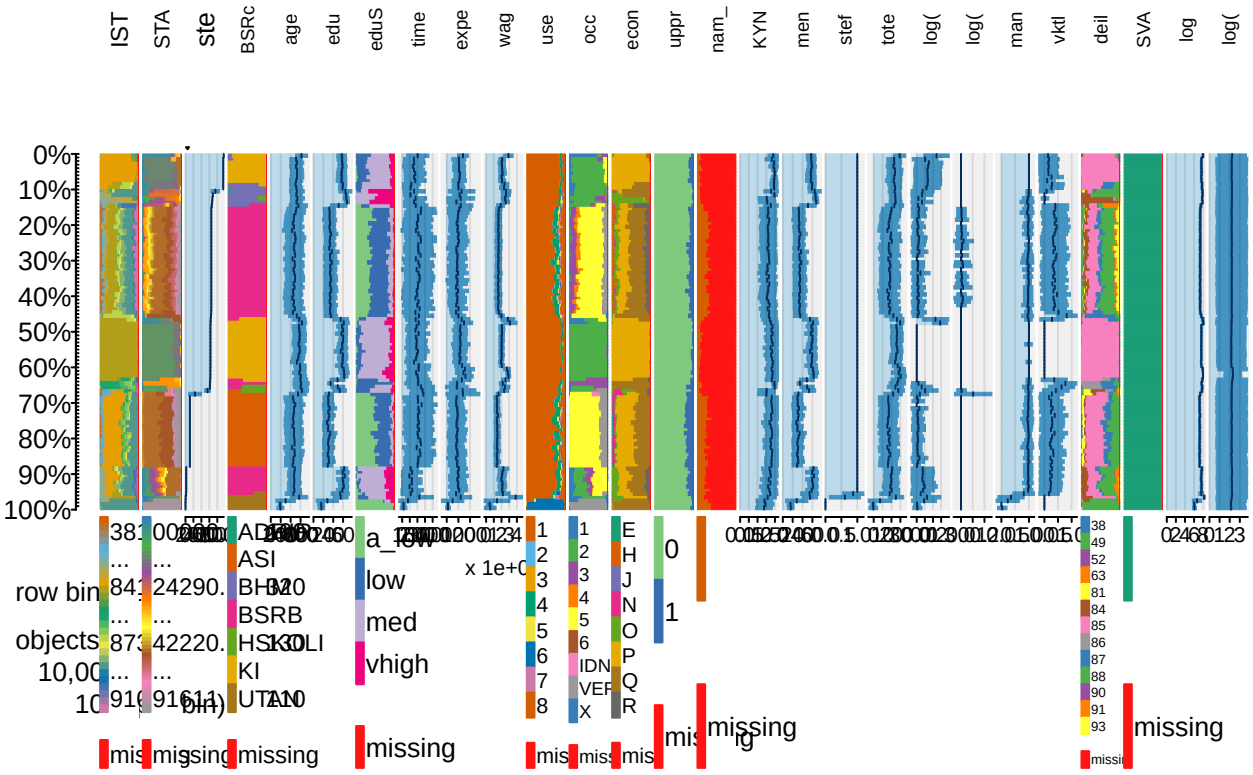

```
##    deild frequency percentage cumulative_perc
## 1     85     12865      47.48           47.48
## 2     88      6288      23.21           70.69
## 3     87      2352       8.68           79.37
## 4     84      2213       8.17           87.54
## 5     91       760       2.81           90.35
## 6     93       626       2.31           92.66
## 7     81       565       2.09           94.75
## 8     49       556       2.05           96.80
## 9     86       515       1.90           98.70
## 10    38       189       0.70           99.40
## 11    52        98       0.36           99.76
## 12    63        35       0.13           99.89
## 13    90        32       0.12          100.00
```

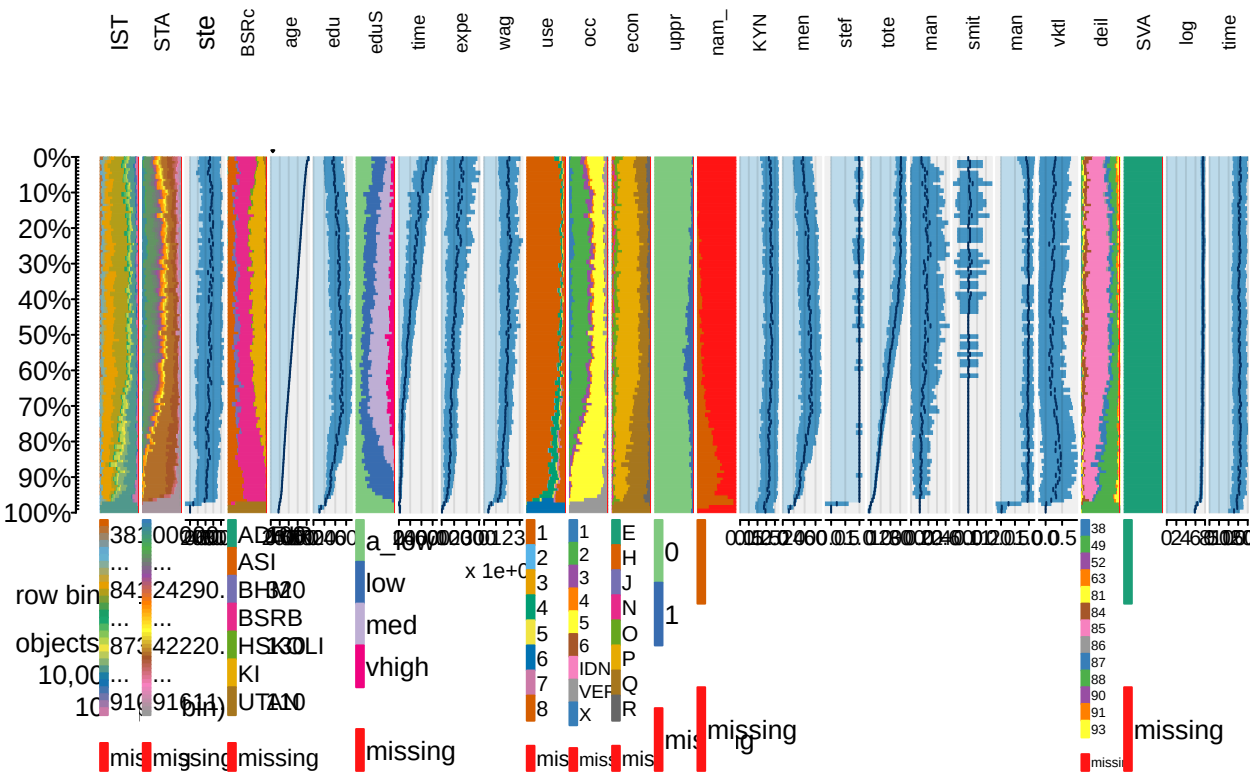


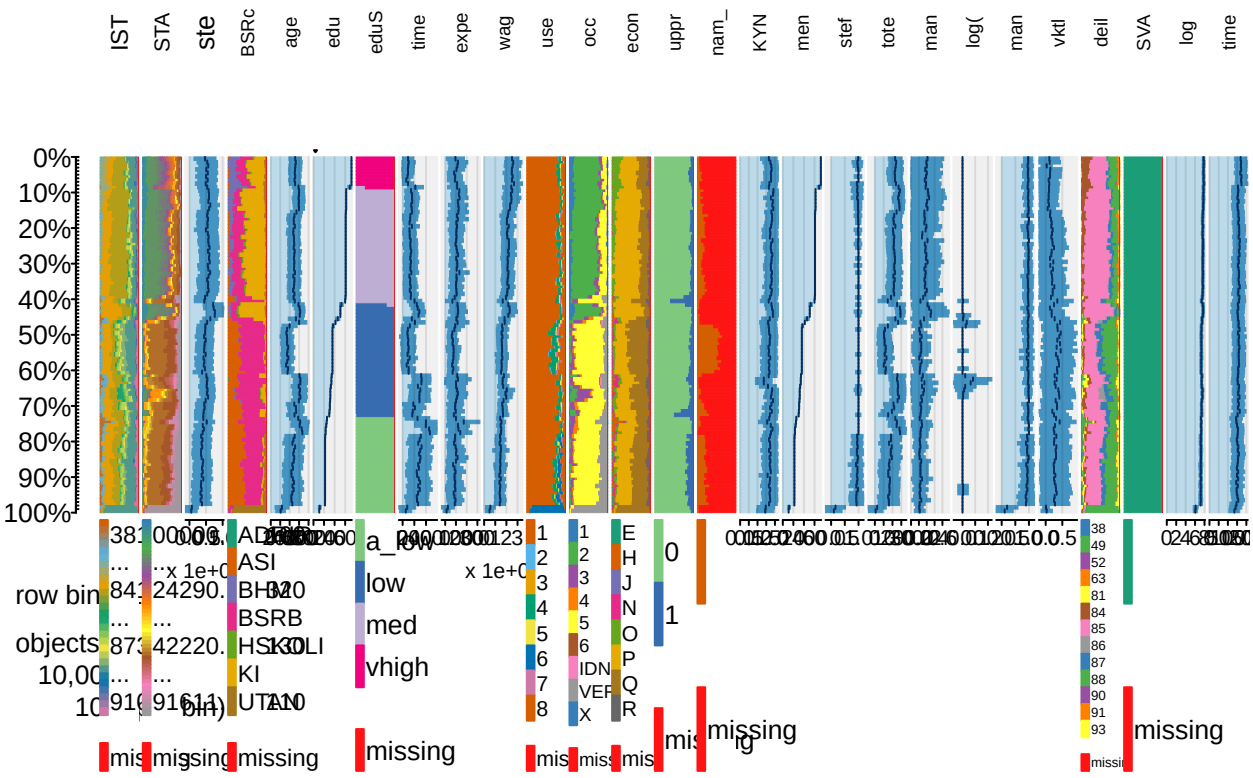

```
##   SVAEDI frequency percentage cumulative_perc
## 1      1     27094        100              100
```

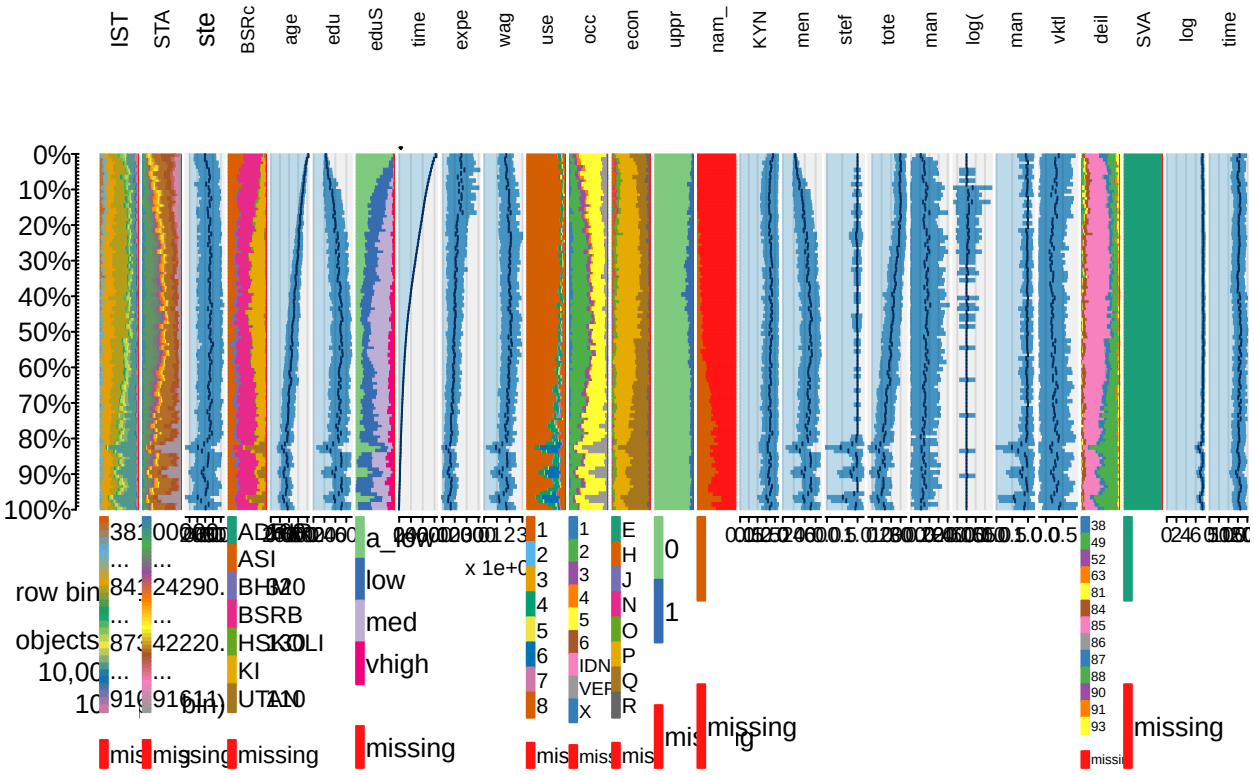
```
## [1] "Variables processed: IST, STA, BSRcell, eduShort, used, occup, econ_activ, uppruni, nam_starf, c
```

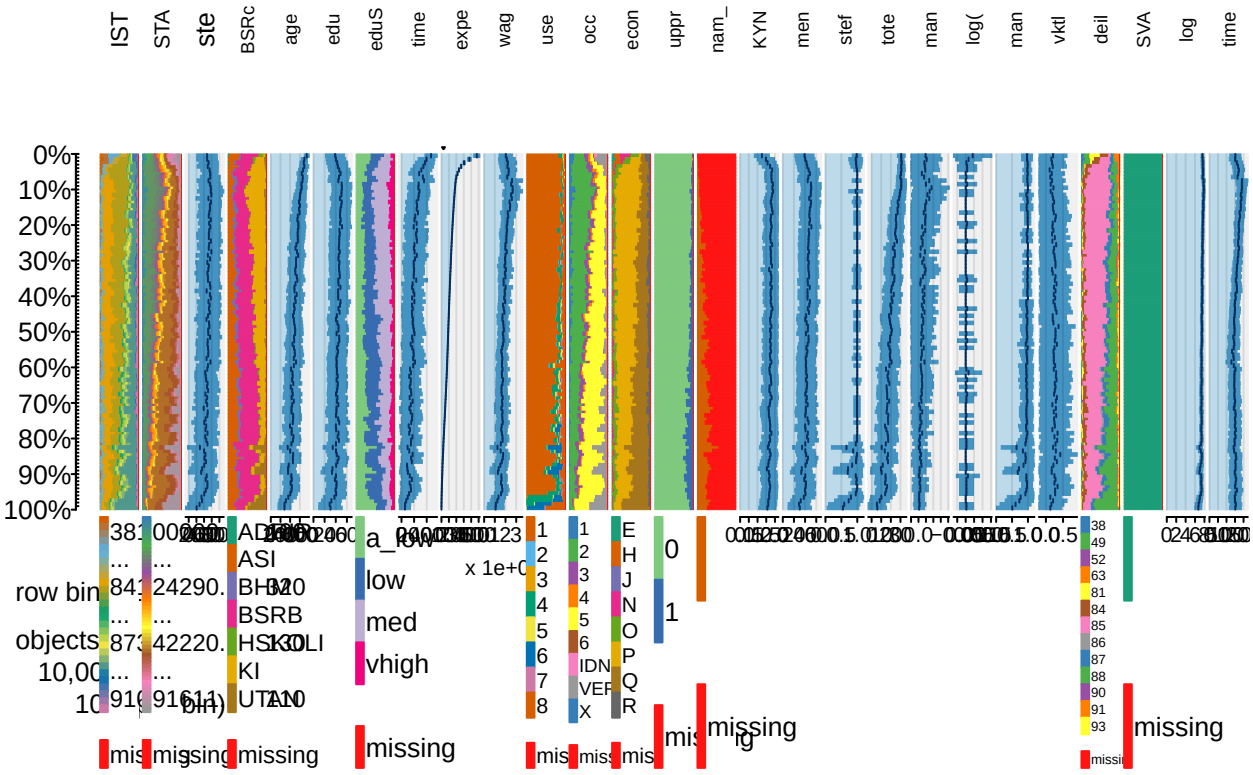

ste
ageM
eduCode
time_since_degree
exper
wage
KYN
men
stefad
totexper
manager
smith
manad
vktlag
logwage
time_arbitr
count
value

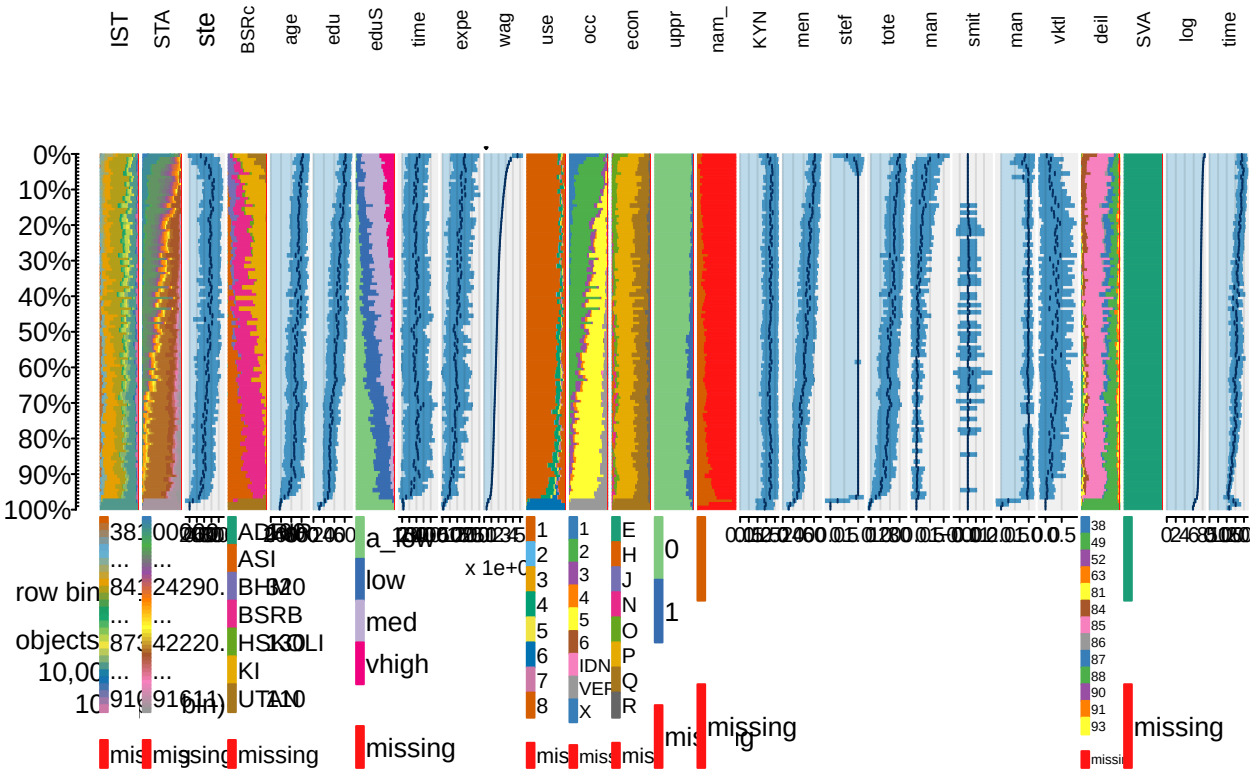

IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
log(
log(
man
vktl
deil
SVA
log
log(
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
low
med
vhigh
missing
0
1
missing
missing

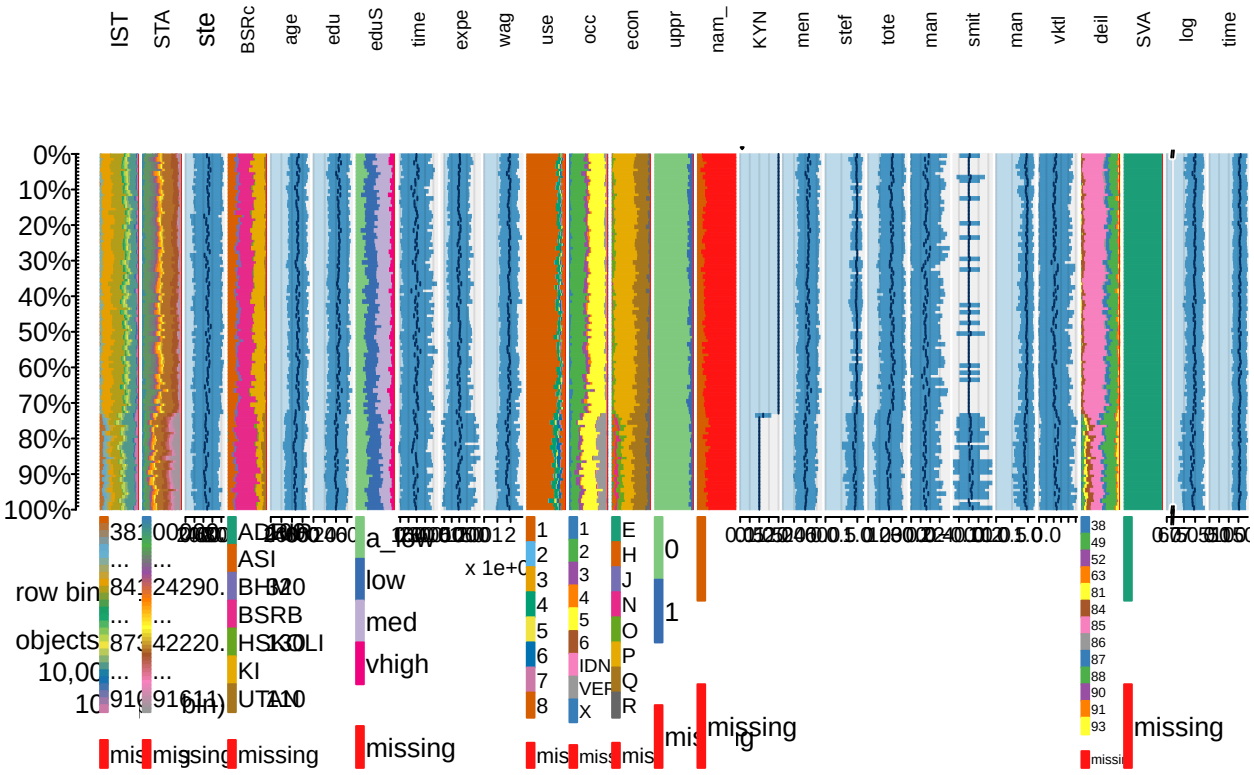

IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
smit
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
low
med
vhigh
missing
0
1
missing
missing

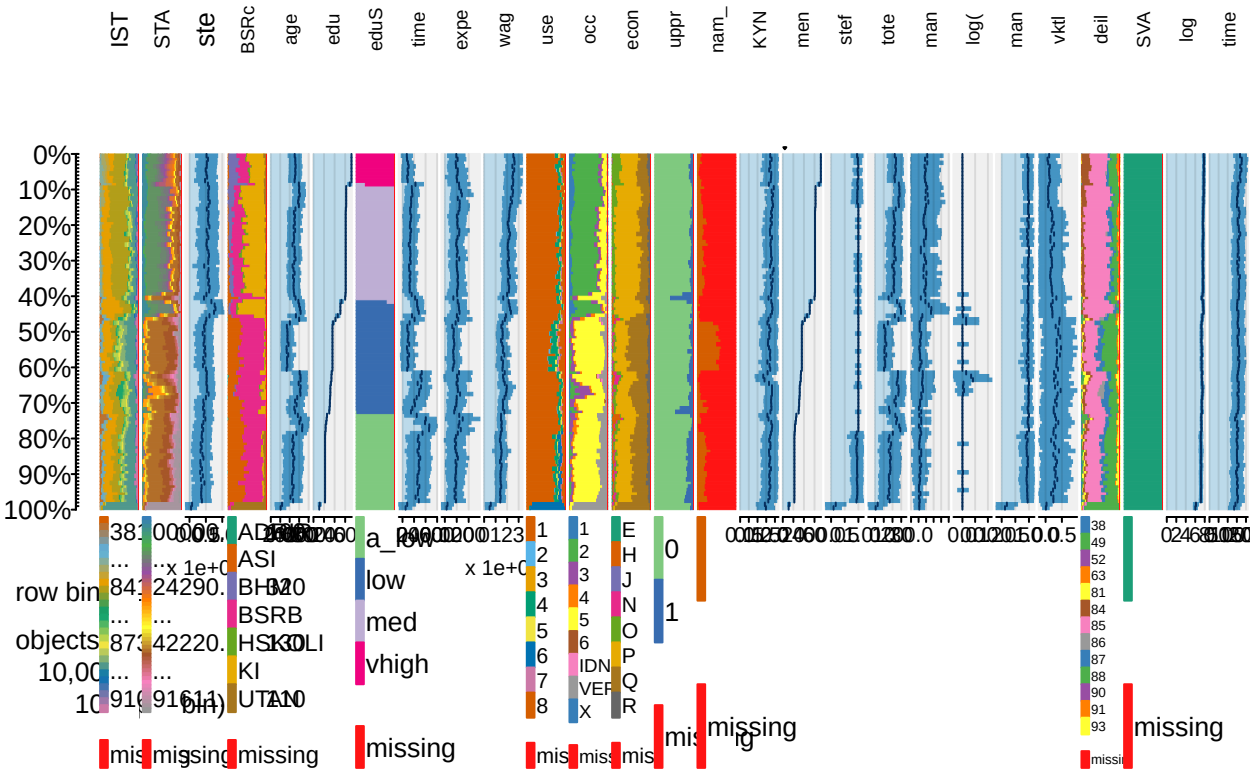

IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
log(
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
missing

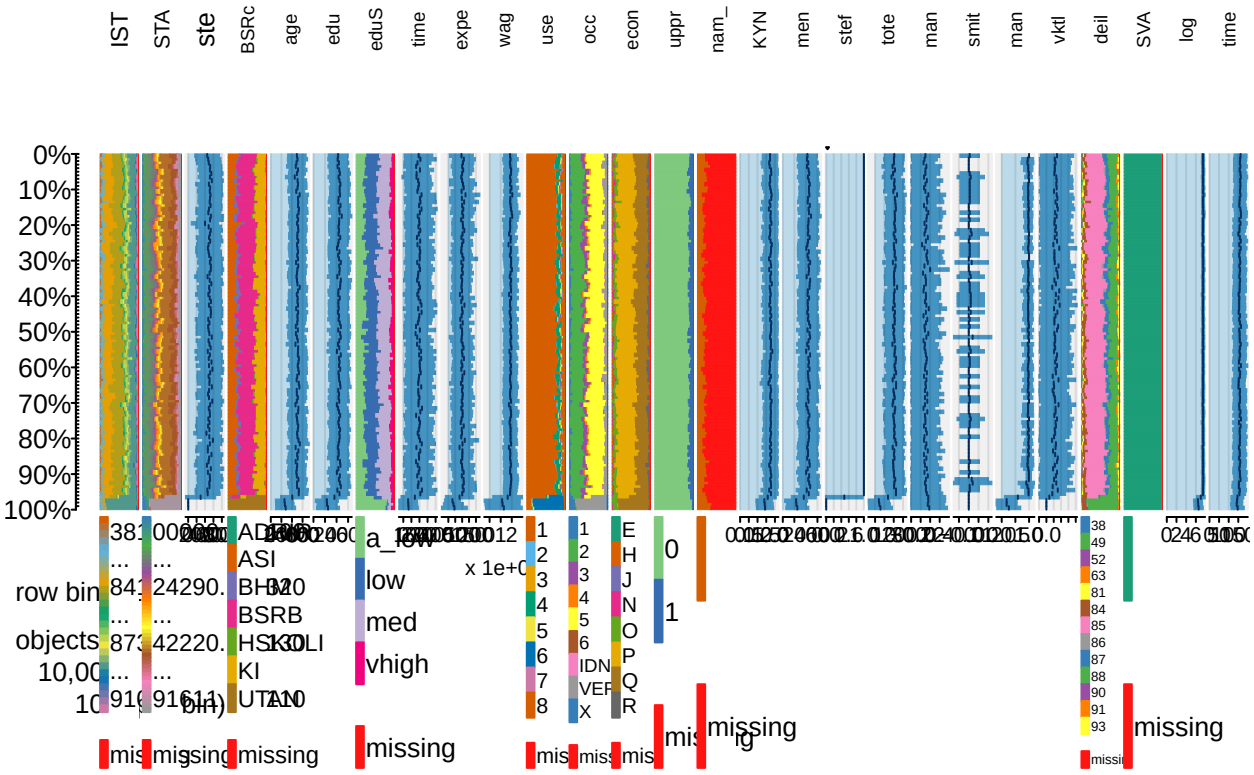

IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
log(
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
low
med
vhigh
missing
0
1
missing
missing

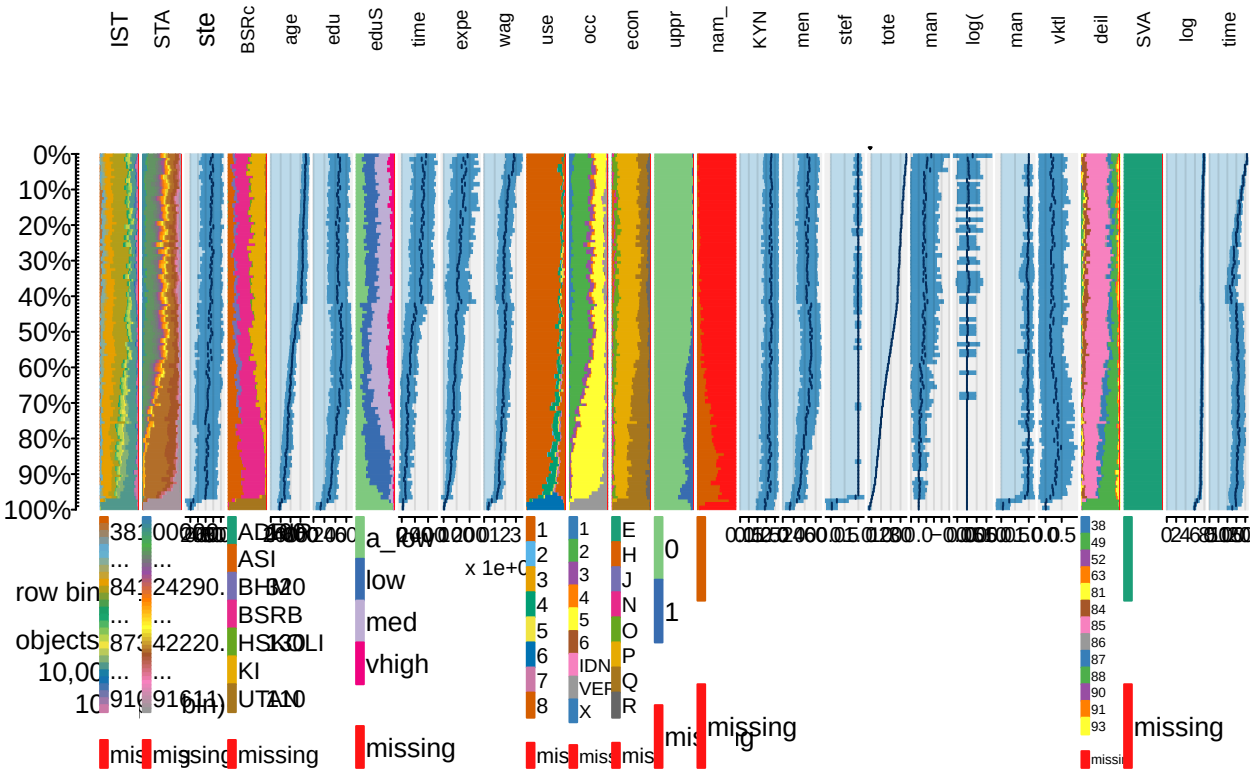

IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
log(
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
low
med
vhigh
missing
0
1
missing
missing

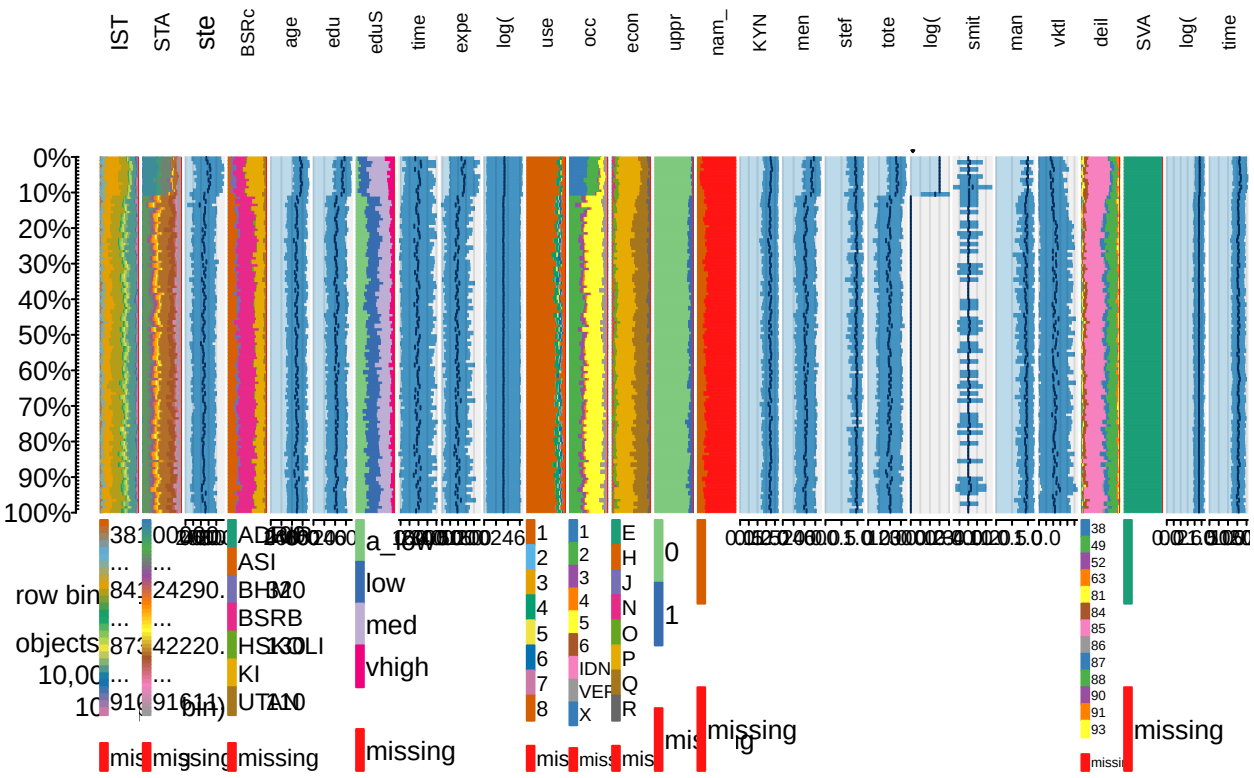

IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
smit
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
low
med
vhigh
missing
0
1
missing
missing

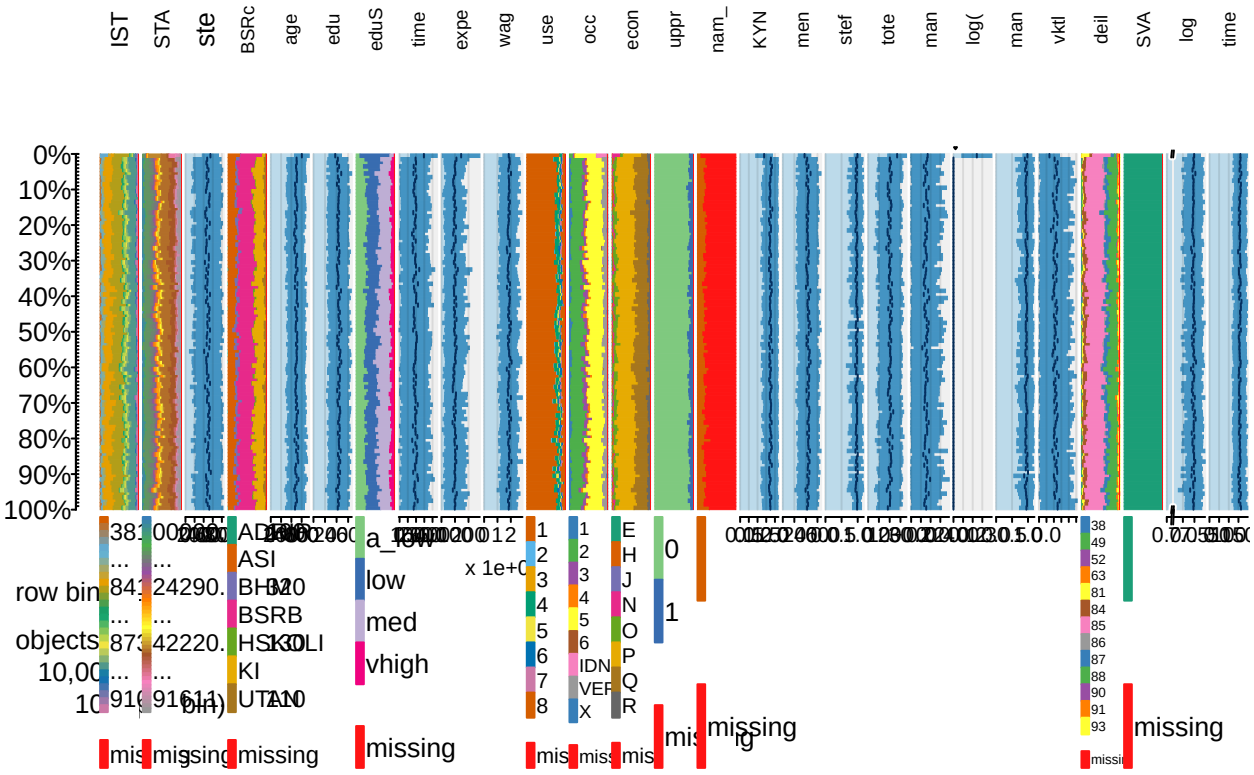

IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
smit
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
low
med
vhigh
missing
0
1
missing
missing

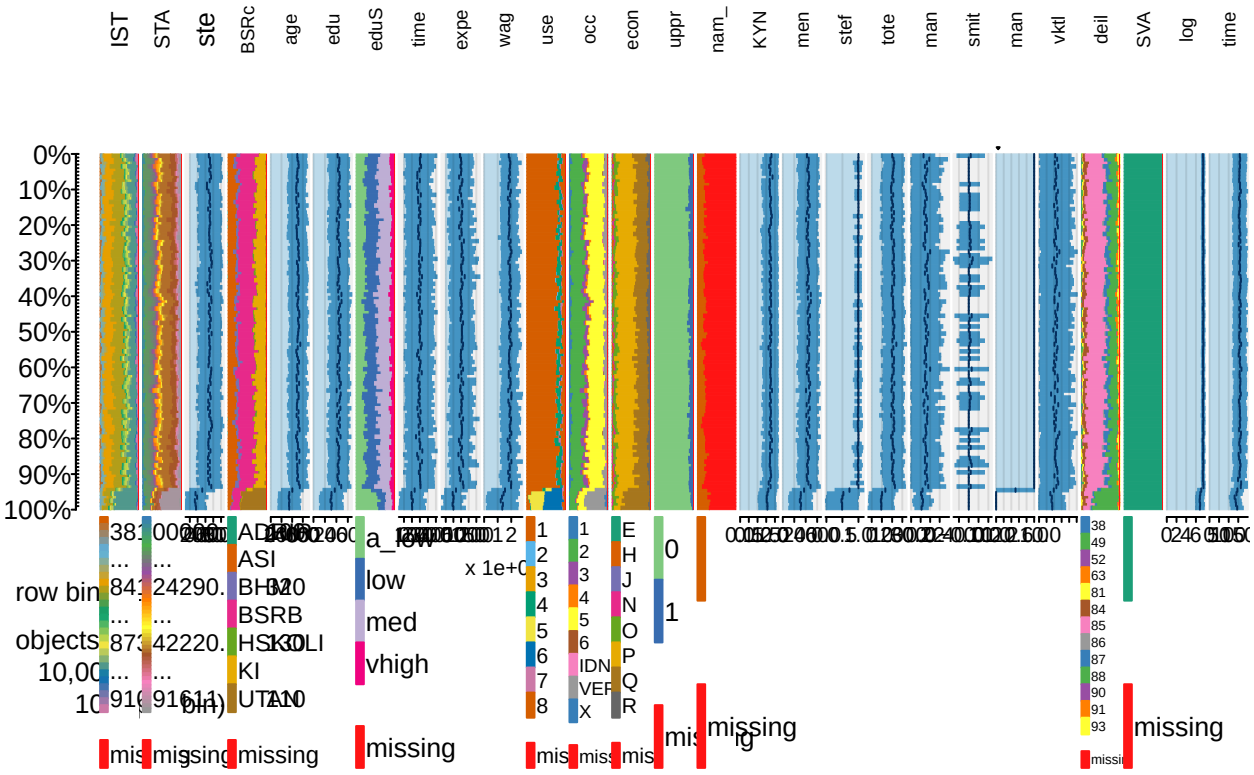

IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
log(
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
missing
low
med
vhigh
missing
missing
missing
missing
missing

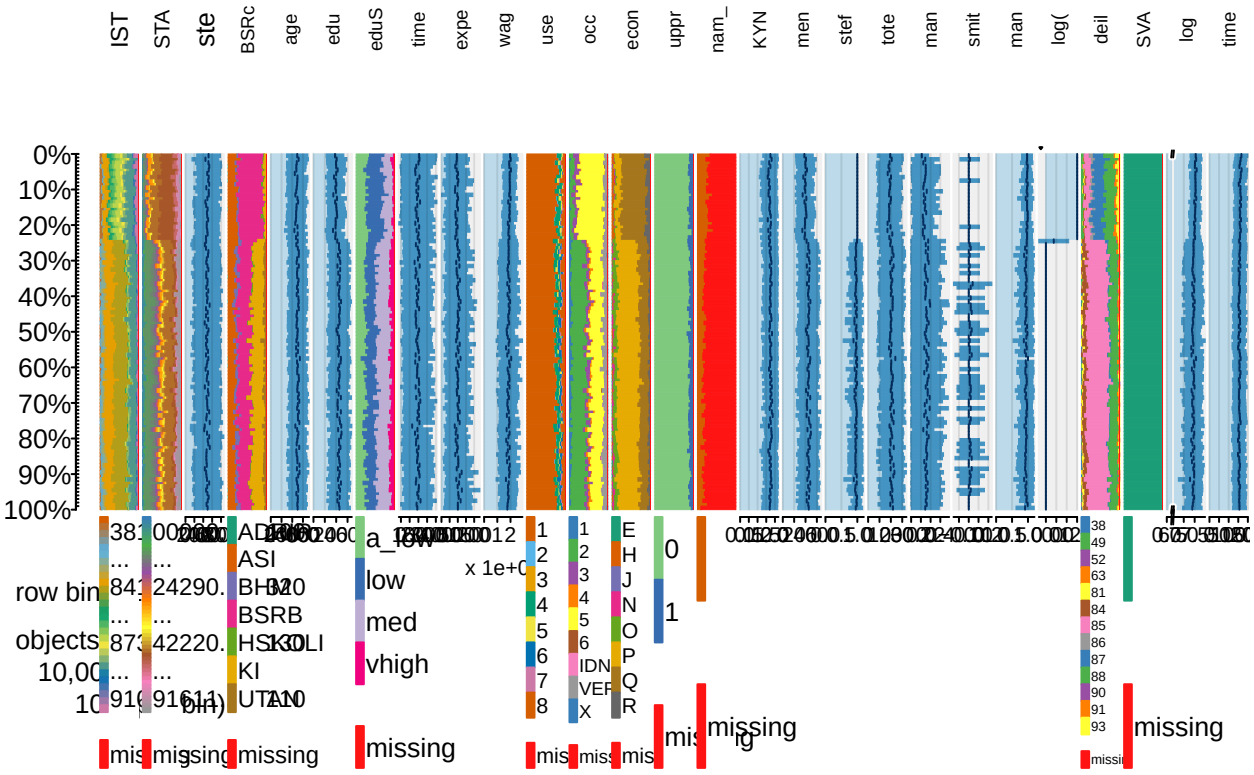

IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
smit
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
low
med
vhigh
missing
0
1
missing
missing

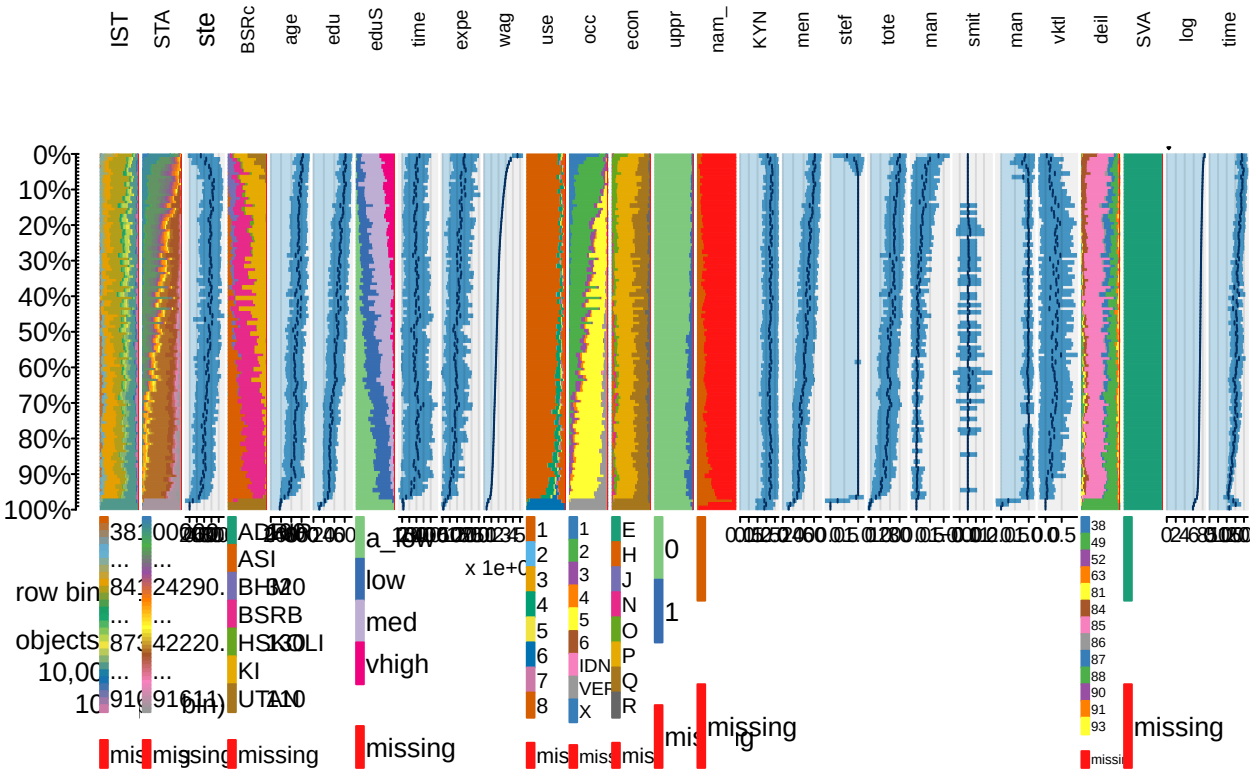

IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
log(
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
100%
row bin
objects
10,00
missing

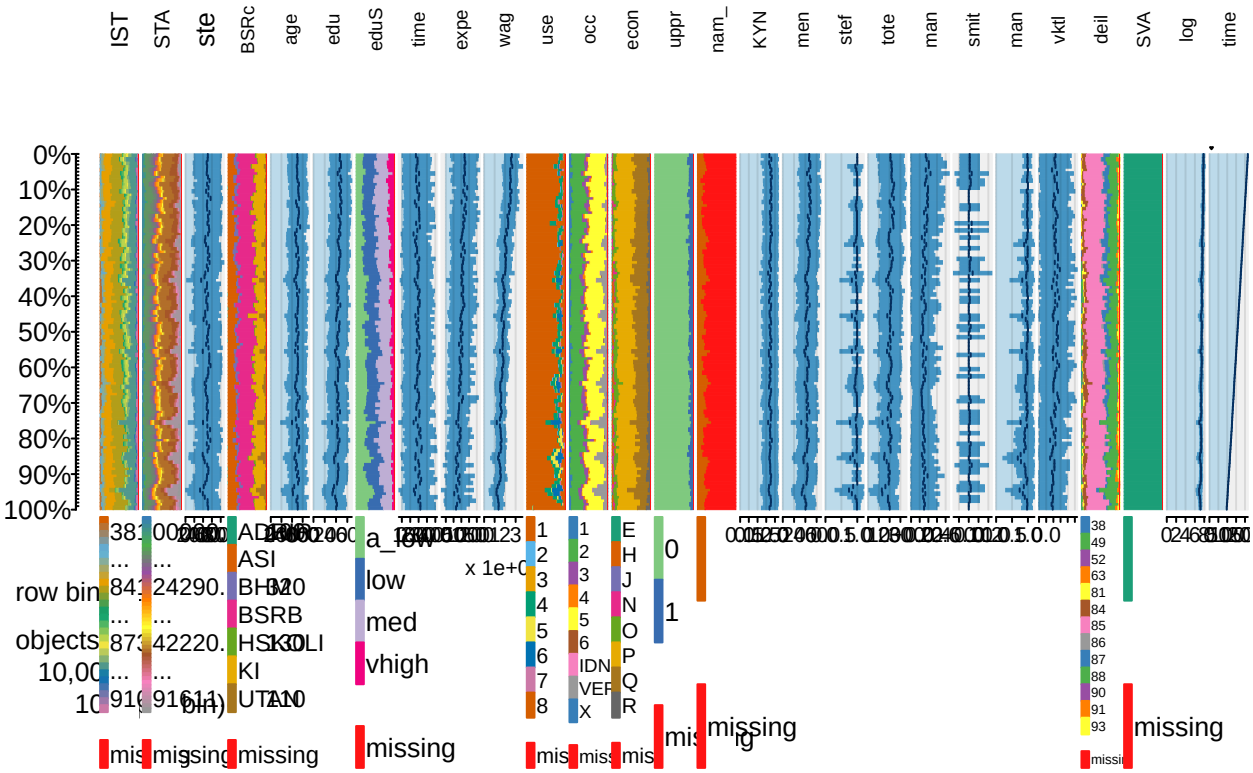

IST
STA
ste
BSRc
age
edu
eduS
time
expe
log(
use
occ
econ
uppr
nam_
KYN
men
stef
tote
log(
smit
man
vktl
deil
SVA
log(
time
0%
10%
20%
30%
100%
row bin
objects
10,00
10
low
med
vhigh
missing
0
1
missing
missing


IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
log(
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
100%
row bin
objects
10,00
10
low
med
vhigh
missing
0
1
missing
missing


IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
smit
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
low
med
vhigh
missing
0
1
missing
missing


IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
smit
man
log(
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
low
med
vhigh
missing
0
1
missing
missing


IST
STA
ste
BSRc
age
edu
eduS
time
expe
wag
use
occ
econ
uppr
nam_
KYN
men
stef
tote
man
smit
man
vktl
deil
SVA
log
time
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
row bin
objects
10,00
10
low
med
vhigh
missing
0
1
missing
missing
missing
missing
missing



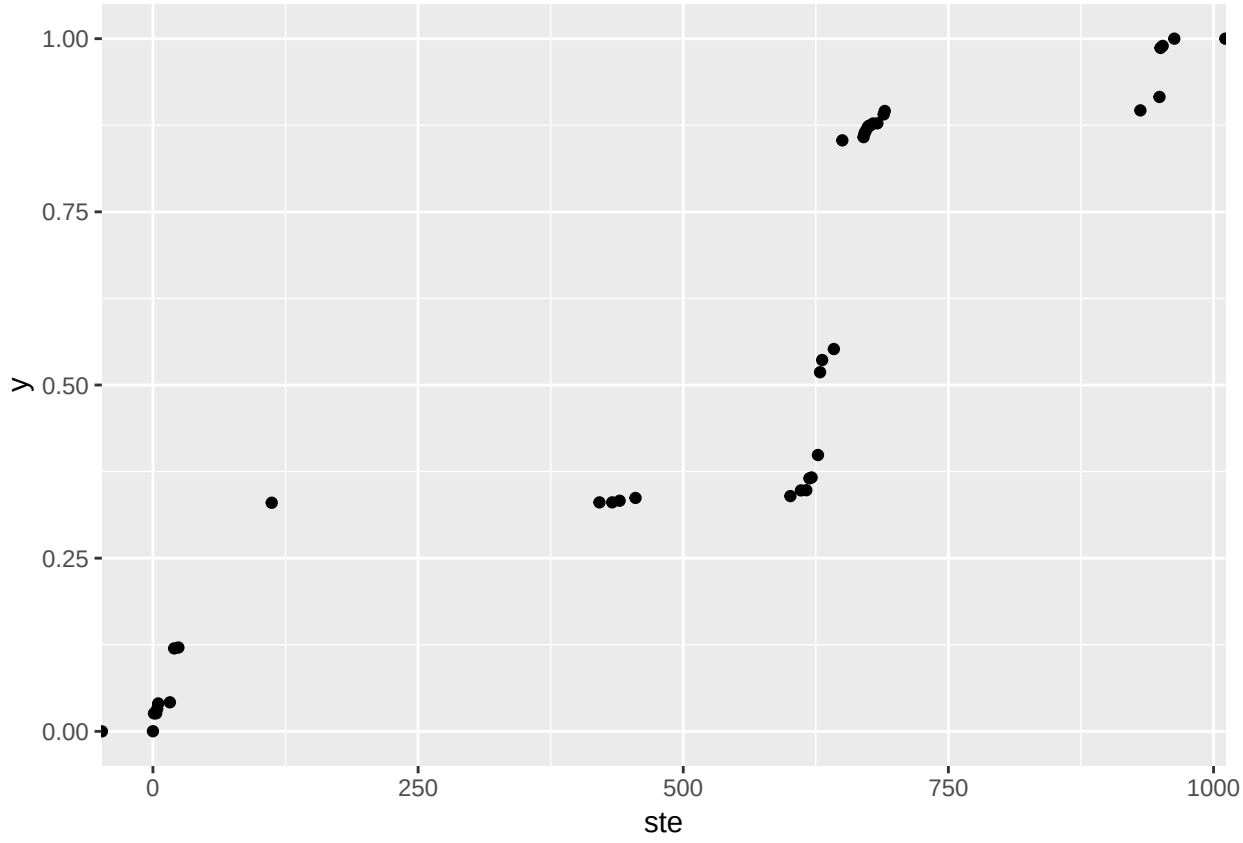

```
## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
## NULL
##
## [[4]]
## NULL
##
## [[5]]
## NULL
##
## [[6]]
## NULL
##
## [[7]]
## NULL
##
## [[8]]
## NULL
##
## [[9]]
## NULL
```

```
##
## [[10]]
## NULL
##
## [[11]]
## NULL
##
## [[12]]
## NULL
##
## [[13]]
## NULL
##
## [[14]]
## NULL
##
## [[15]]
## NULL
##
## [[16]]
## NULL
```

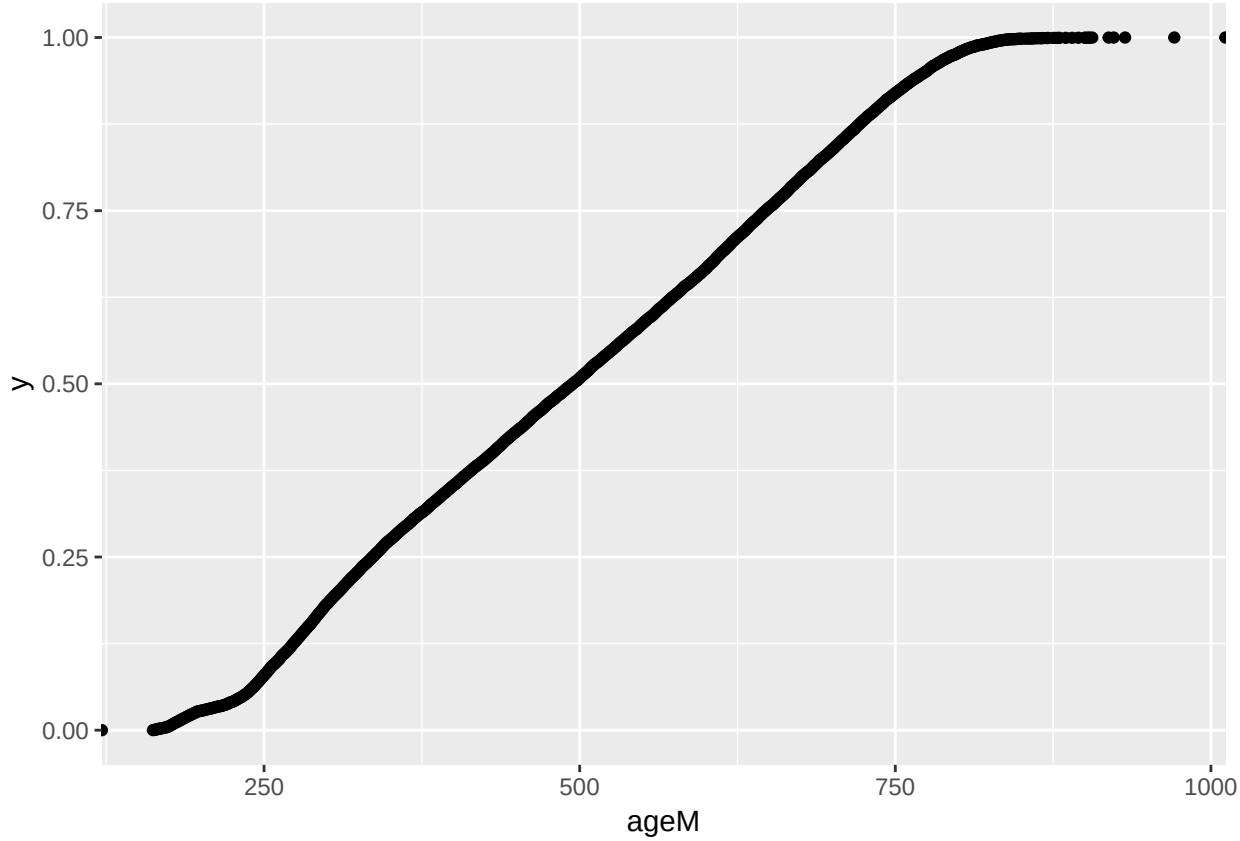
view_univar

Marginal cumulative distributions

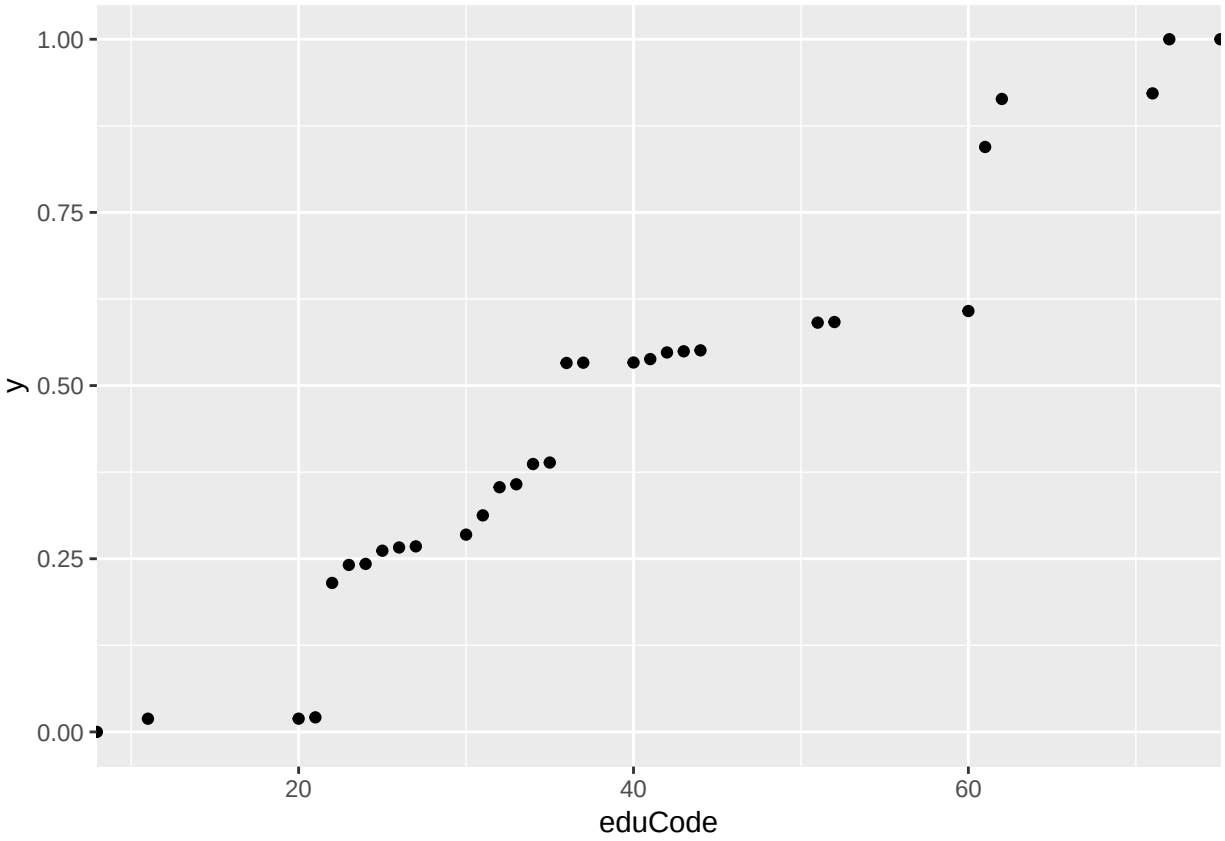
```
## [[1]]
```



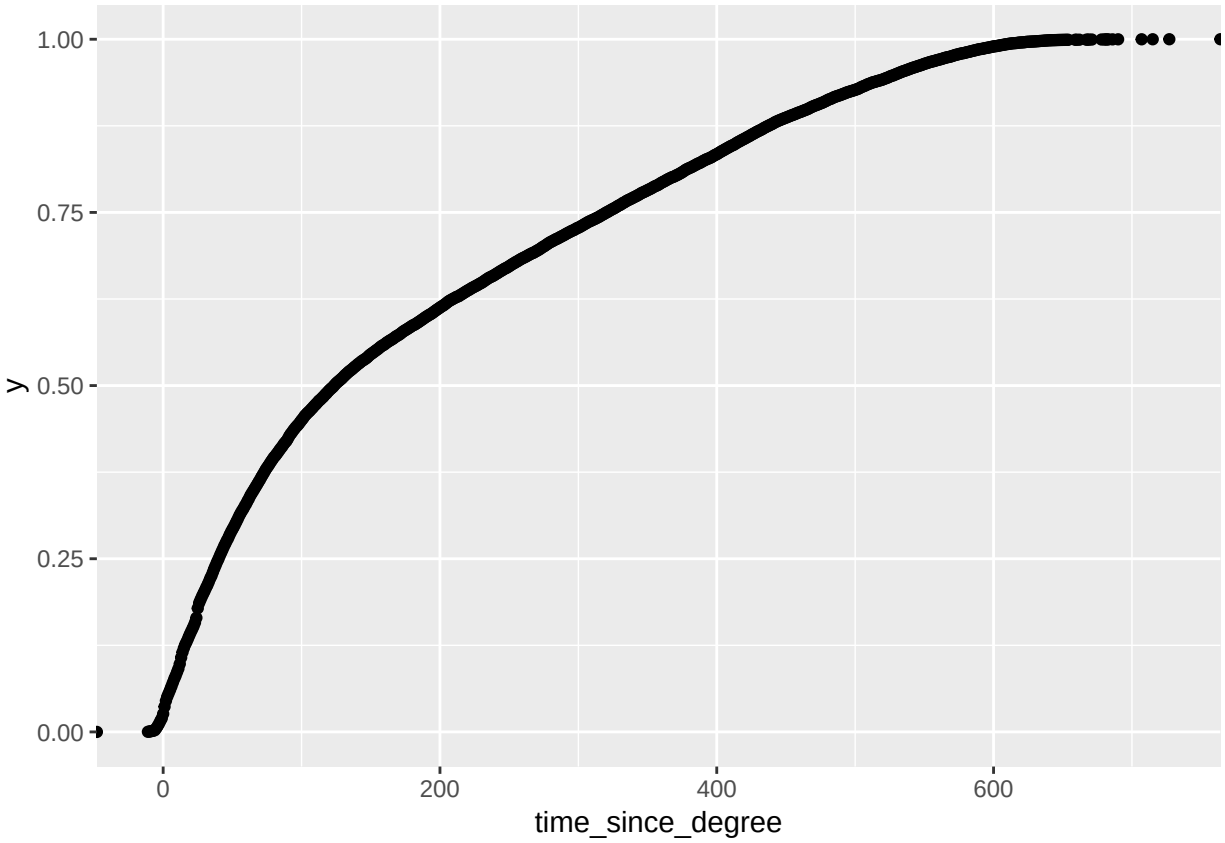
```
##
## [[2]]
```



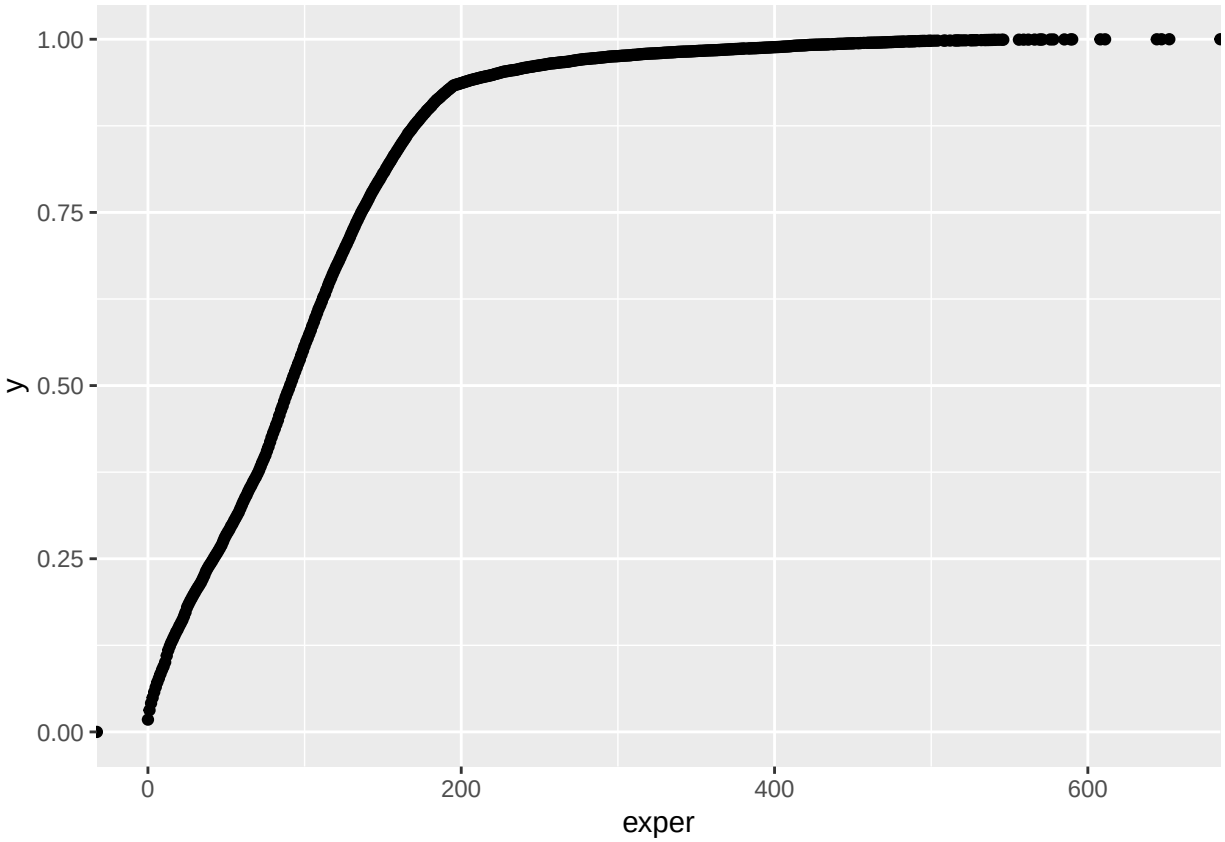

```
##
## [[3]]
```



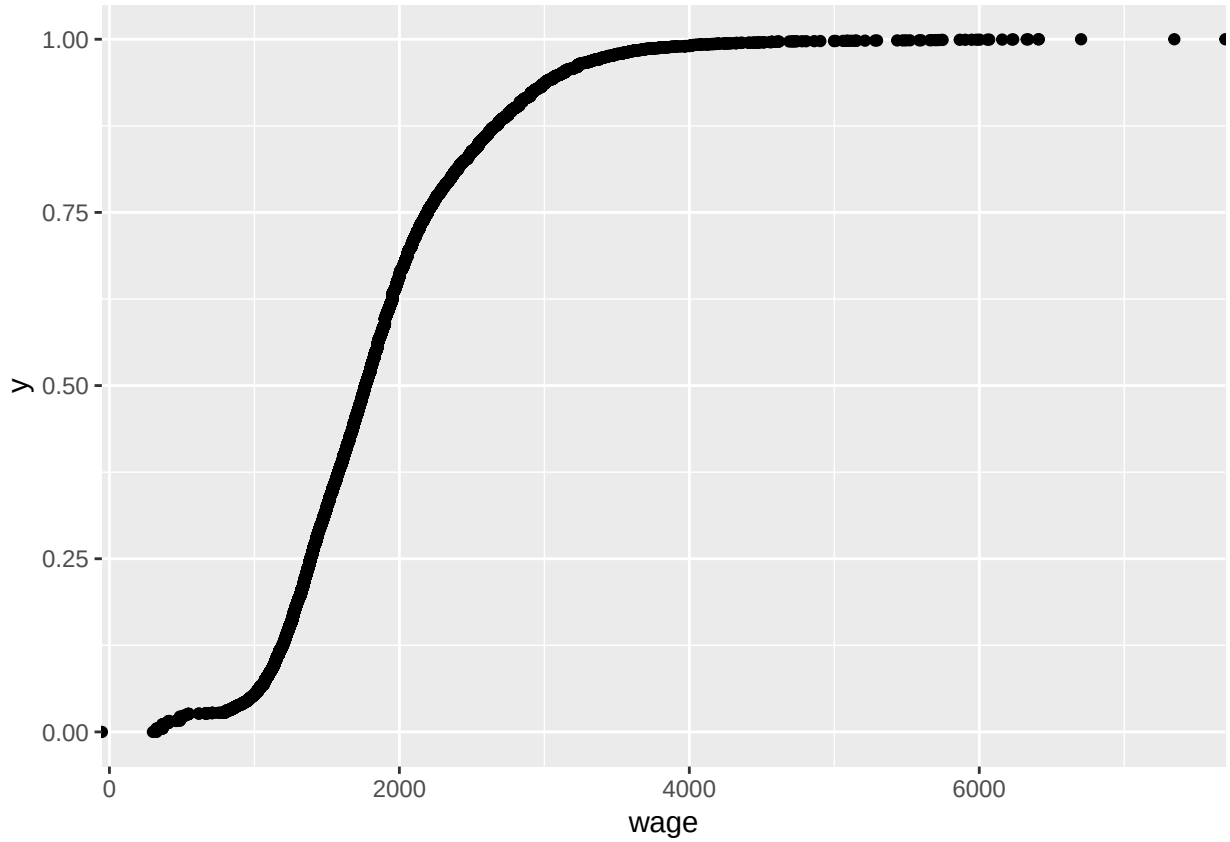

```
##
## [[4]]
```



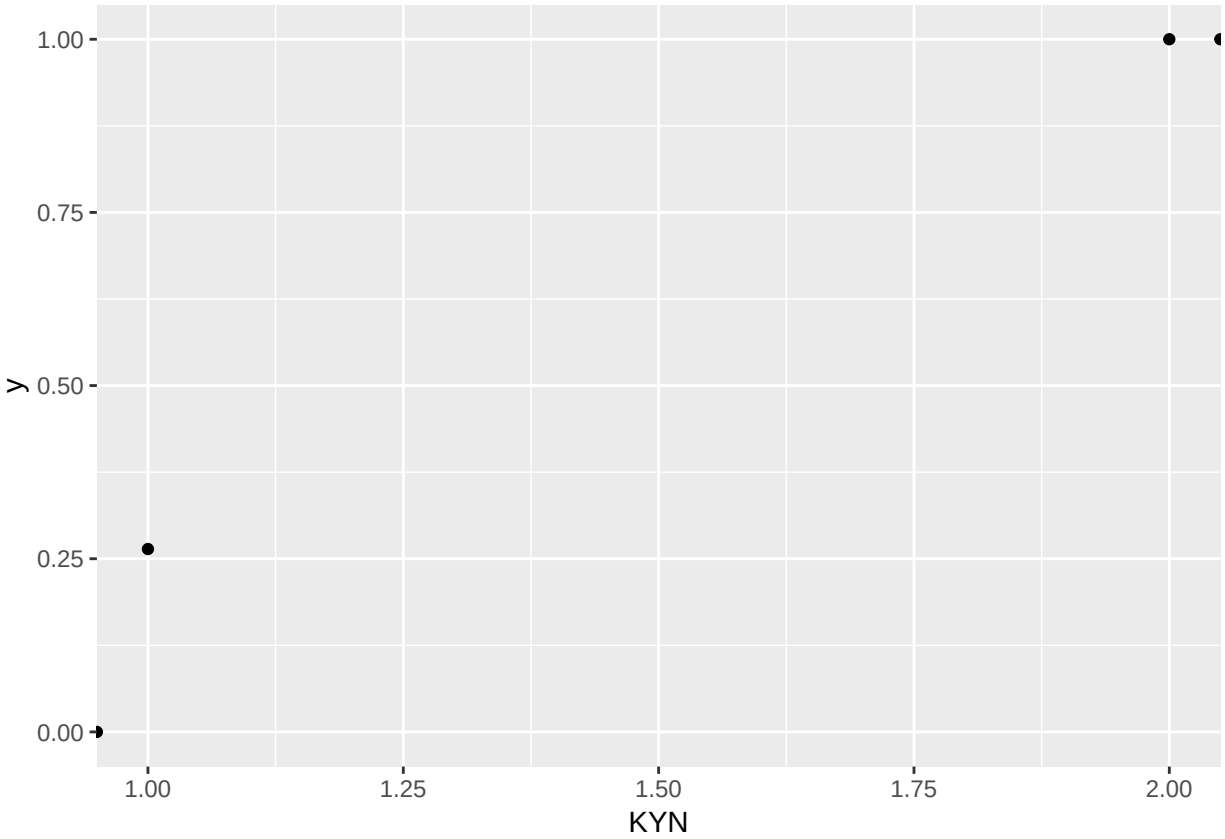

```
##
## [[5]]
```



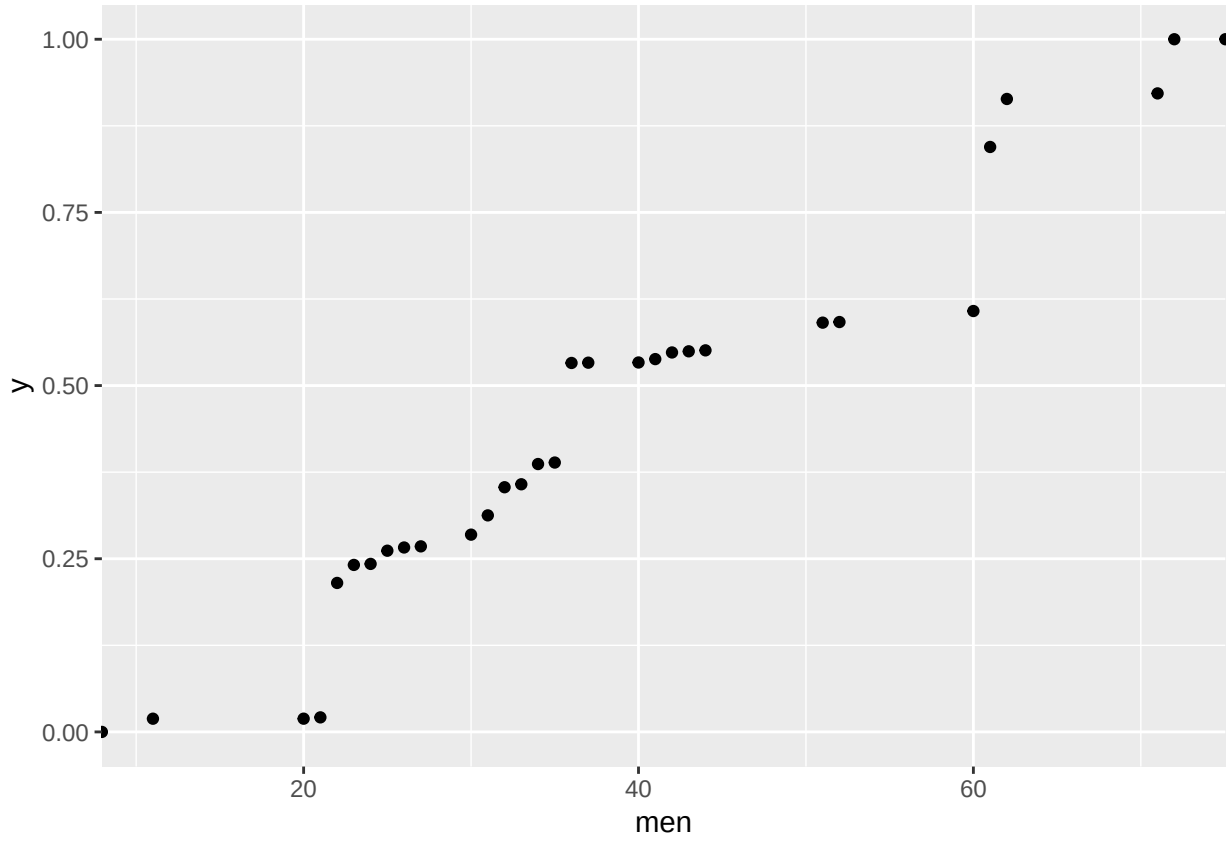

```
##
## [[6]]
```



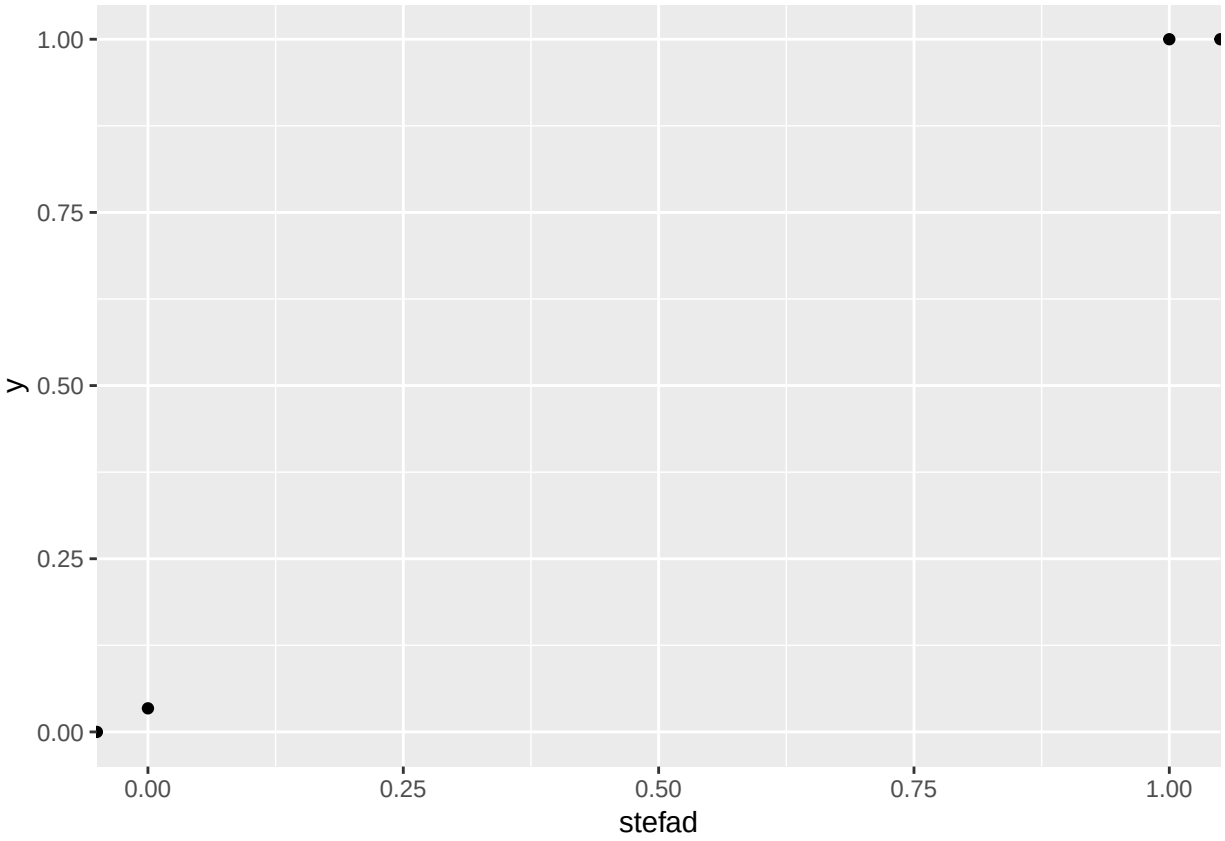

```
##
## [[7]]
```



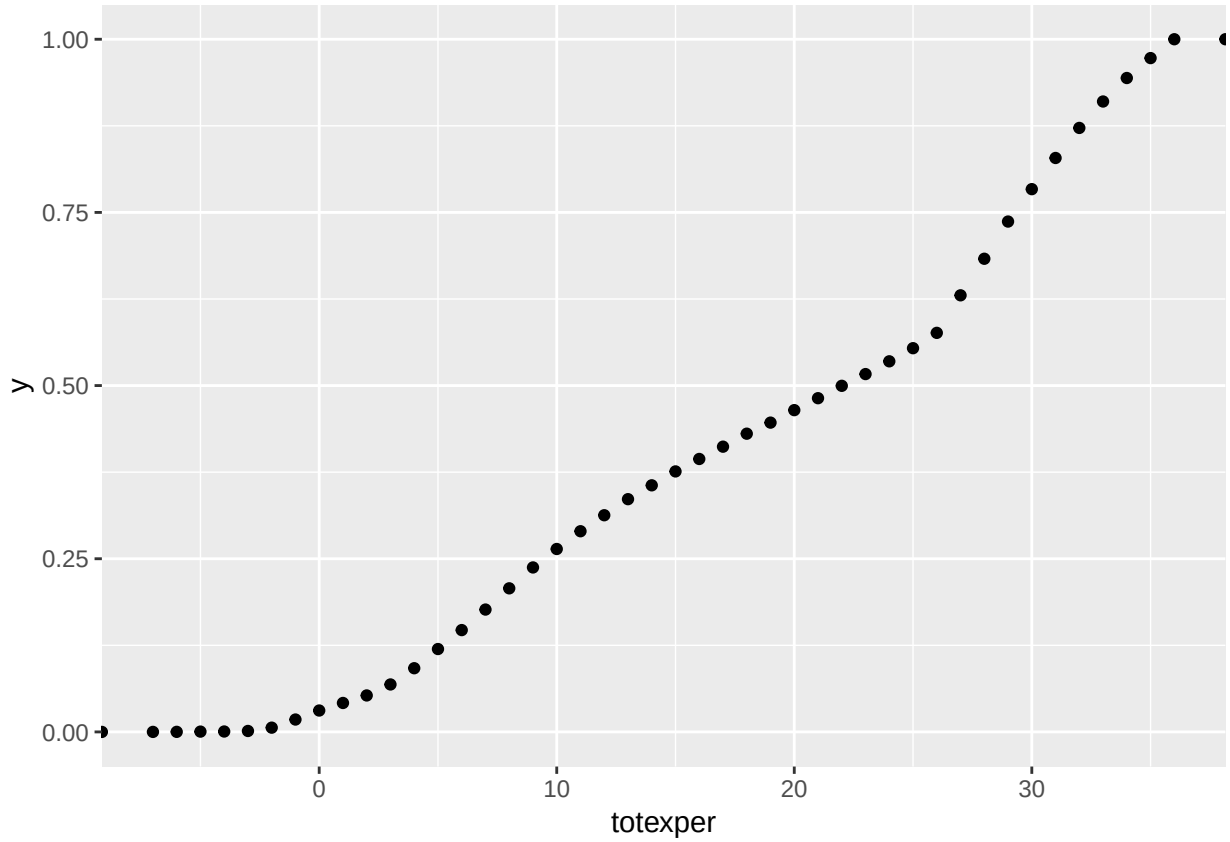

```
##
## [[8]]
```



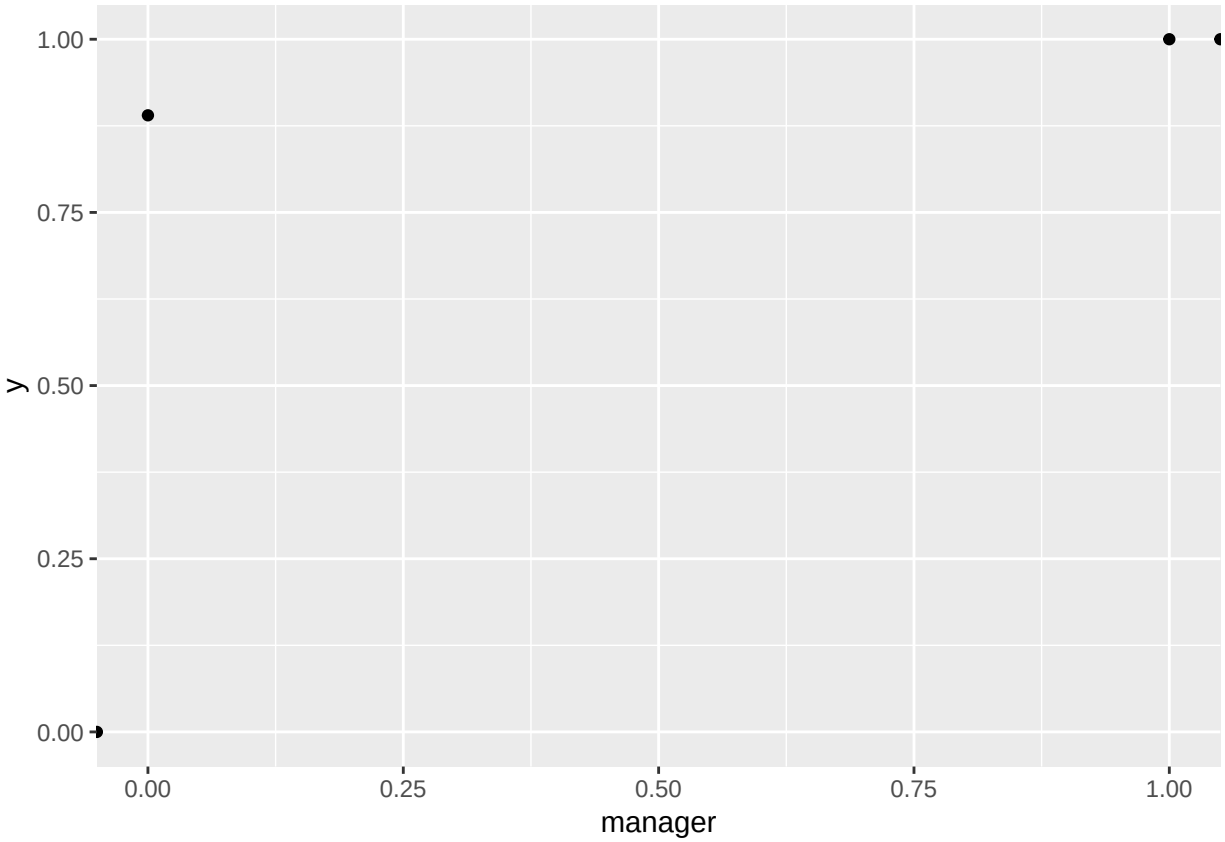

```
##
## [[9]]
```



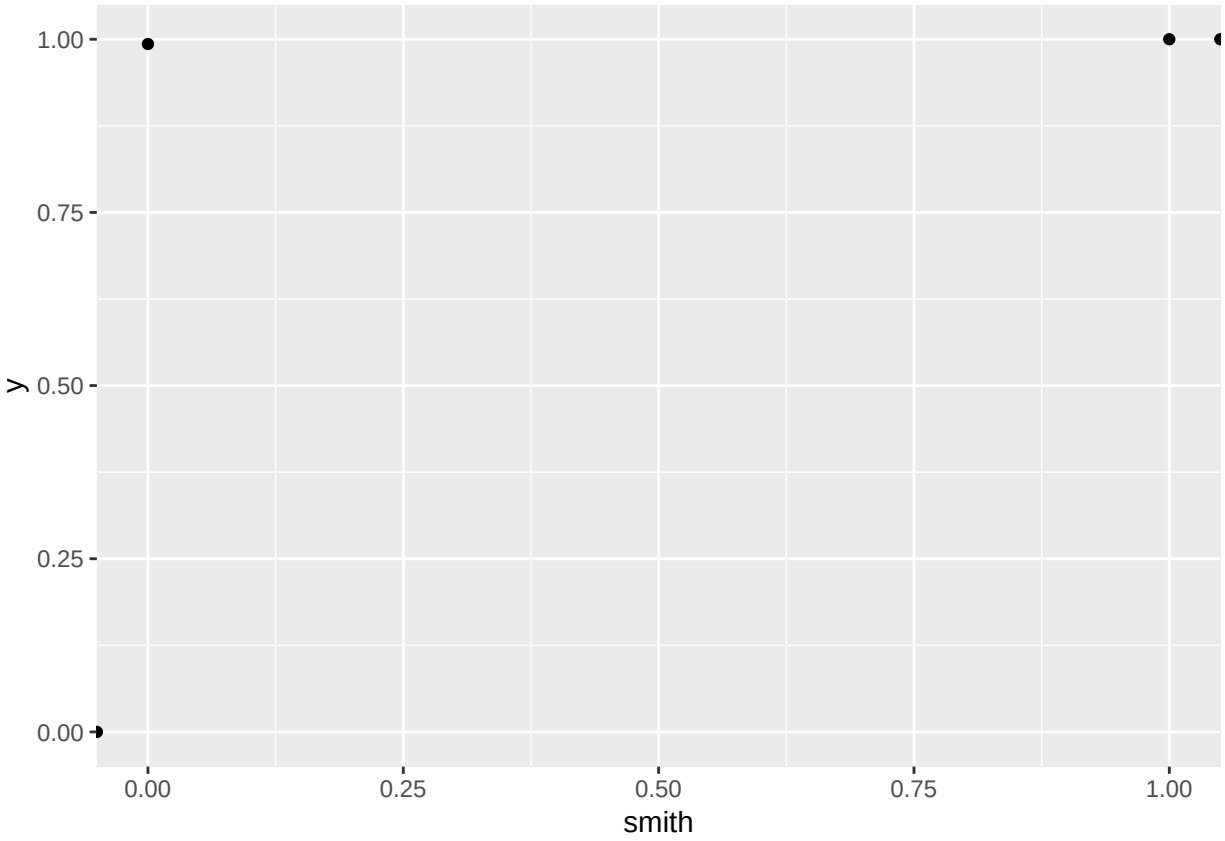

```
##
## [[10]]
```



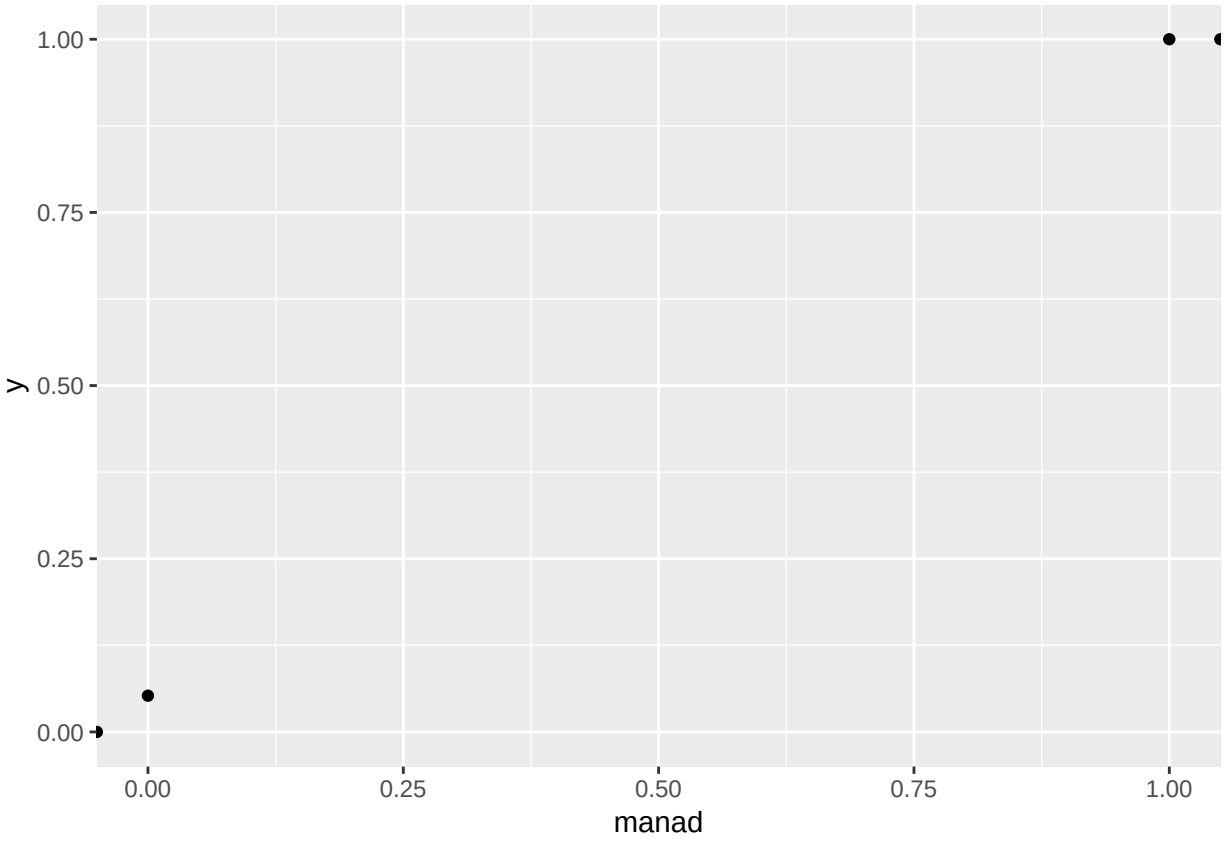

```
##
## [[11]]
```



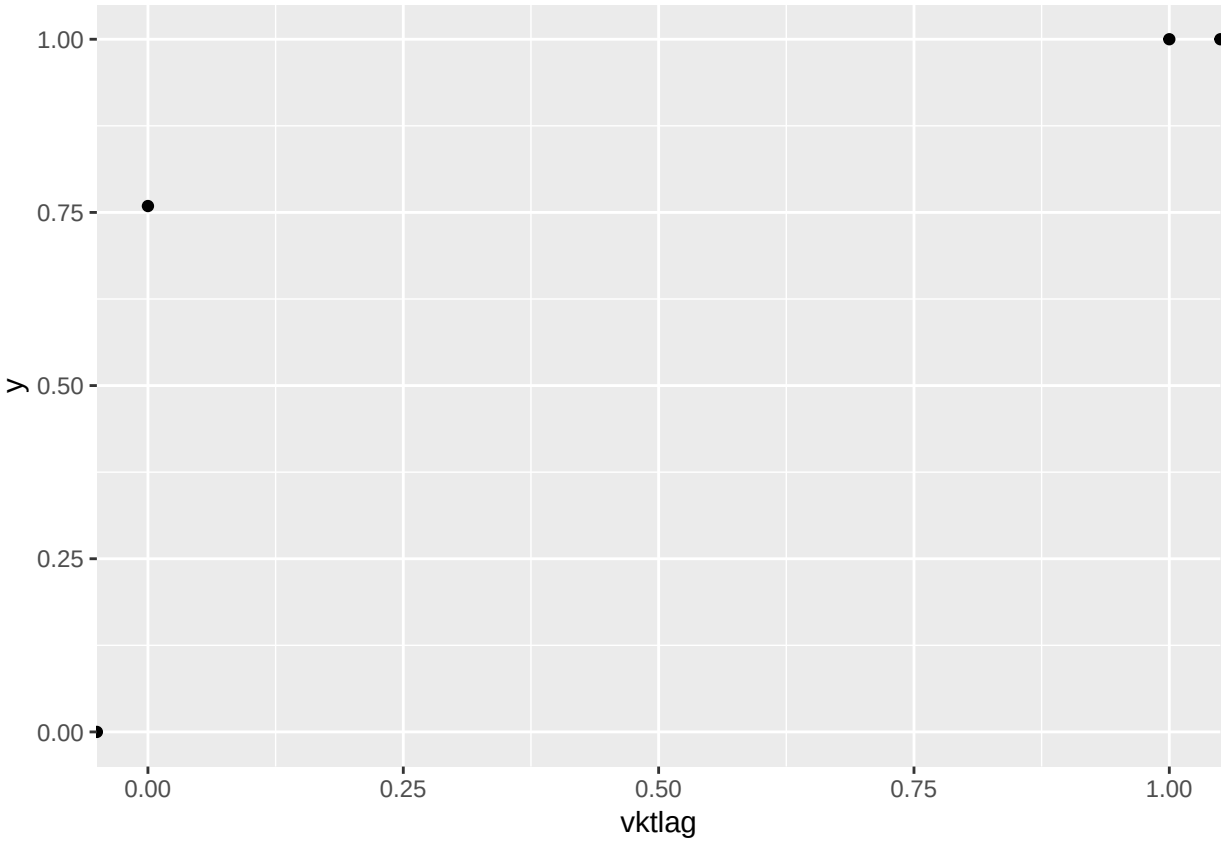

```
##
## [[12]]
```



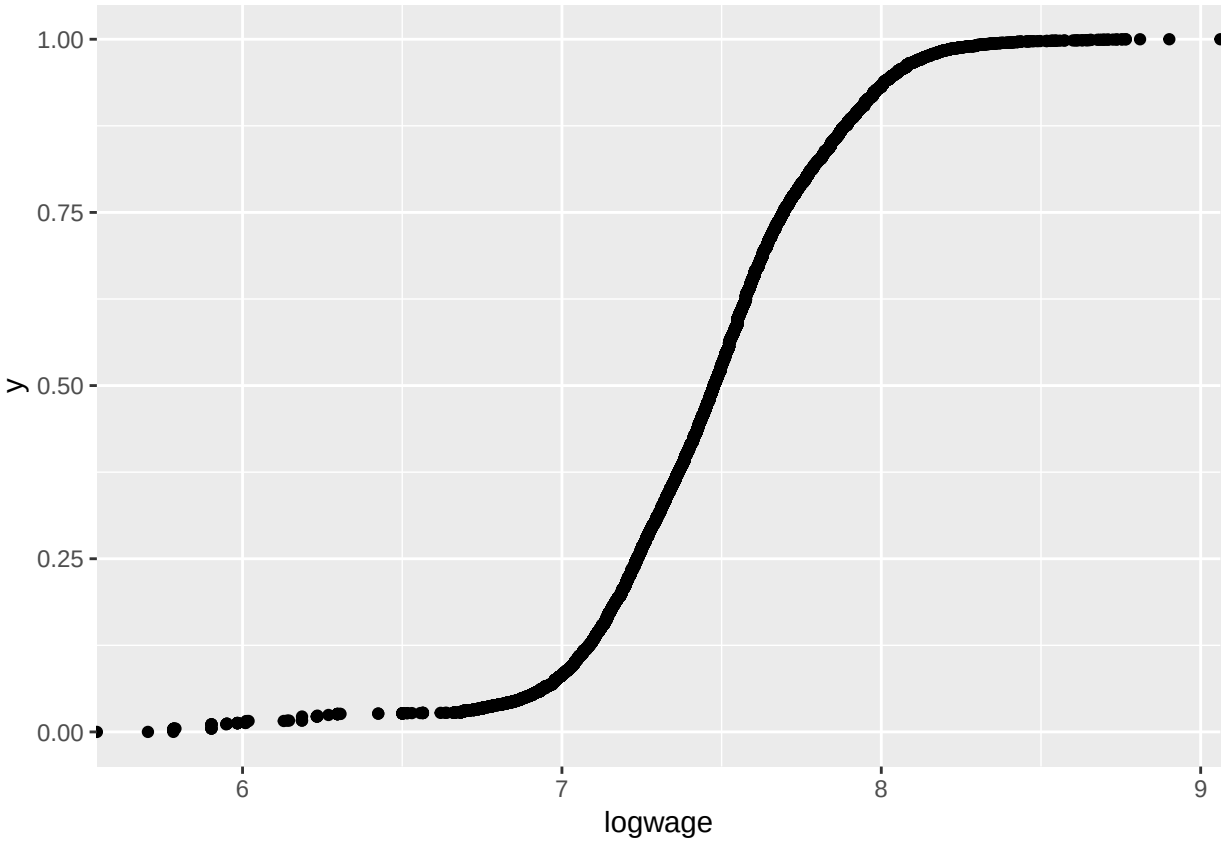

```
##
## [[13]]
```



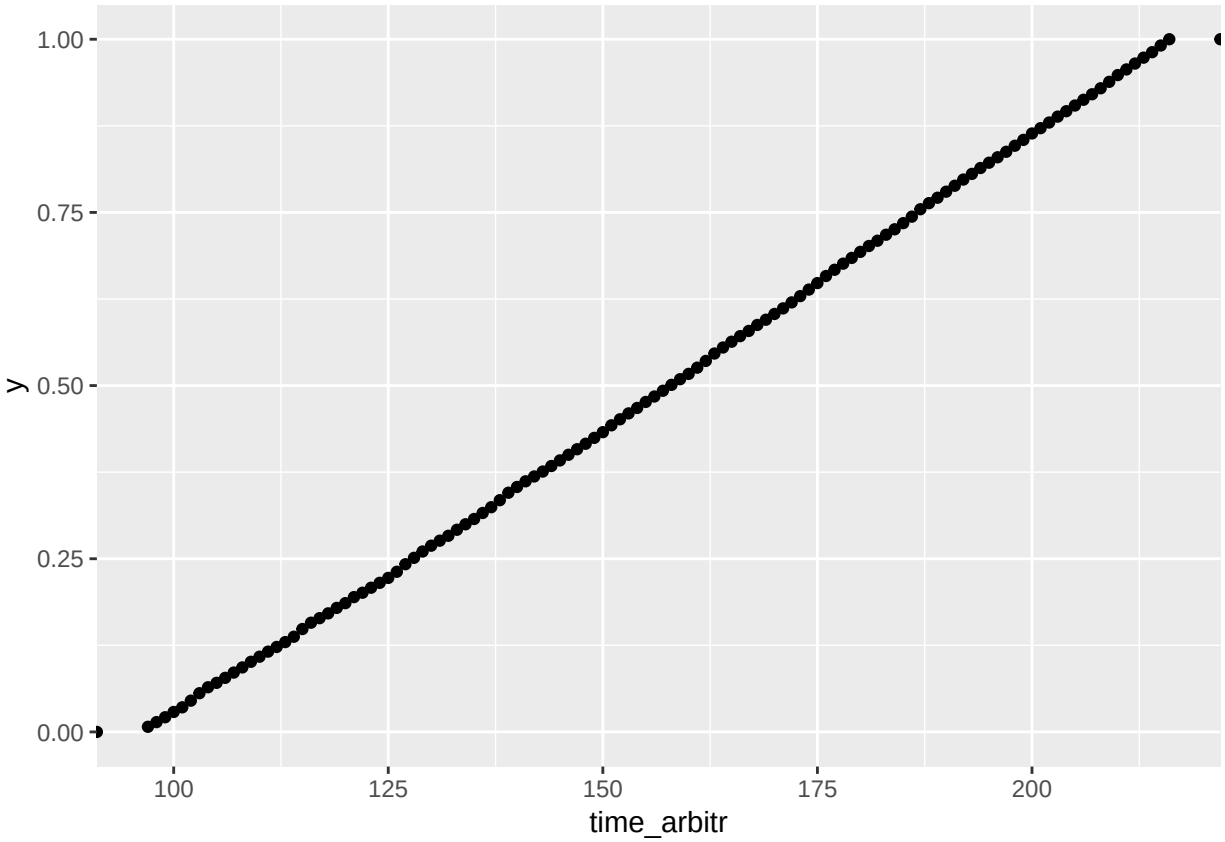

```
##
## [[14]]
```




```
##
## [[15]]
```




```
##
## [[16]]
```



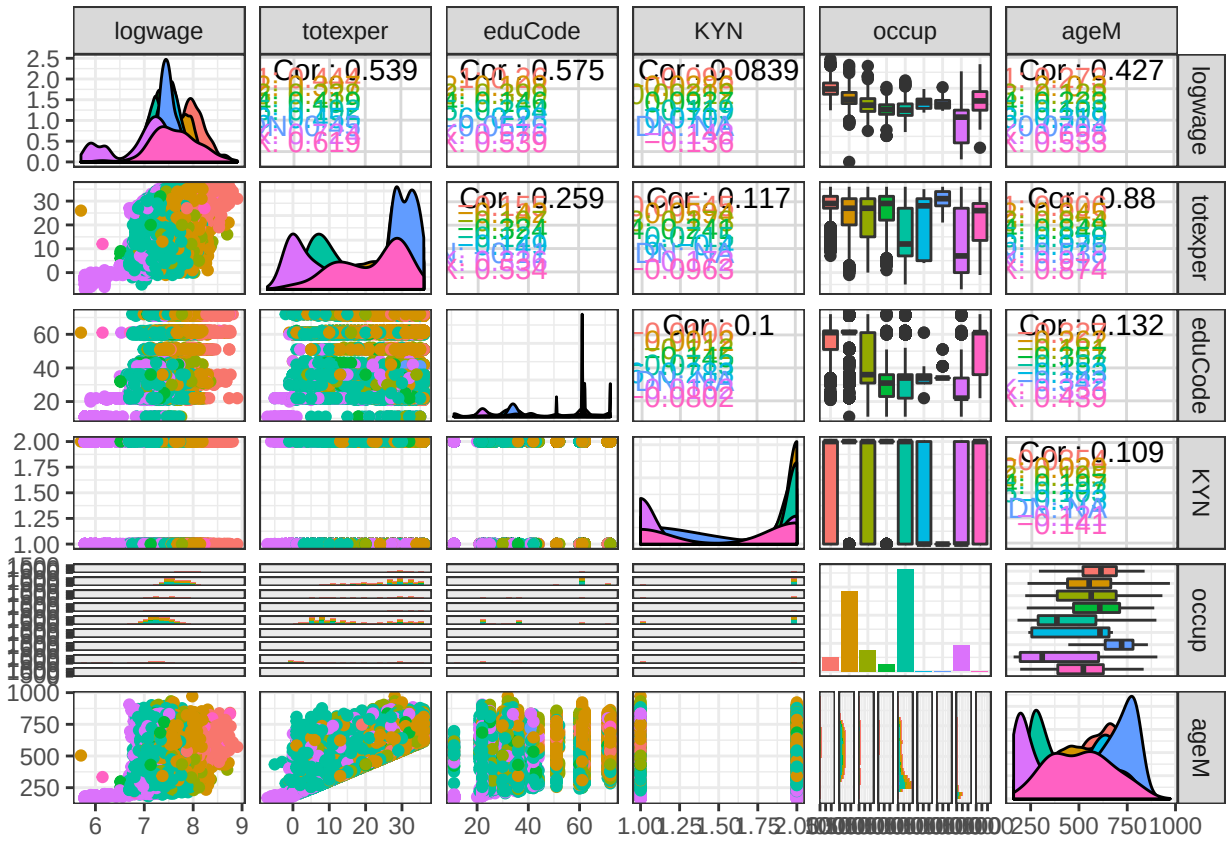
view_multivar

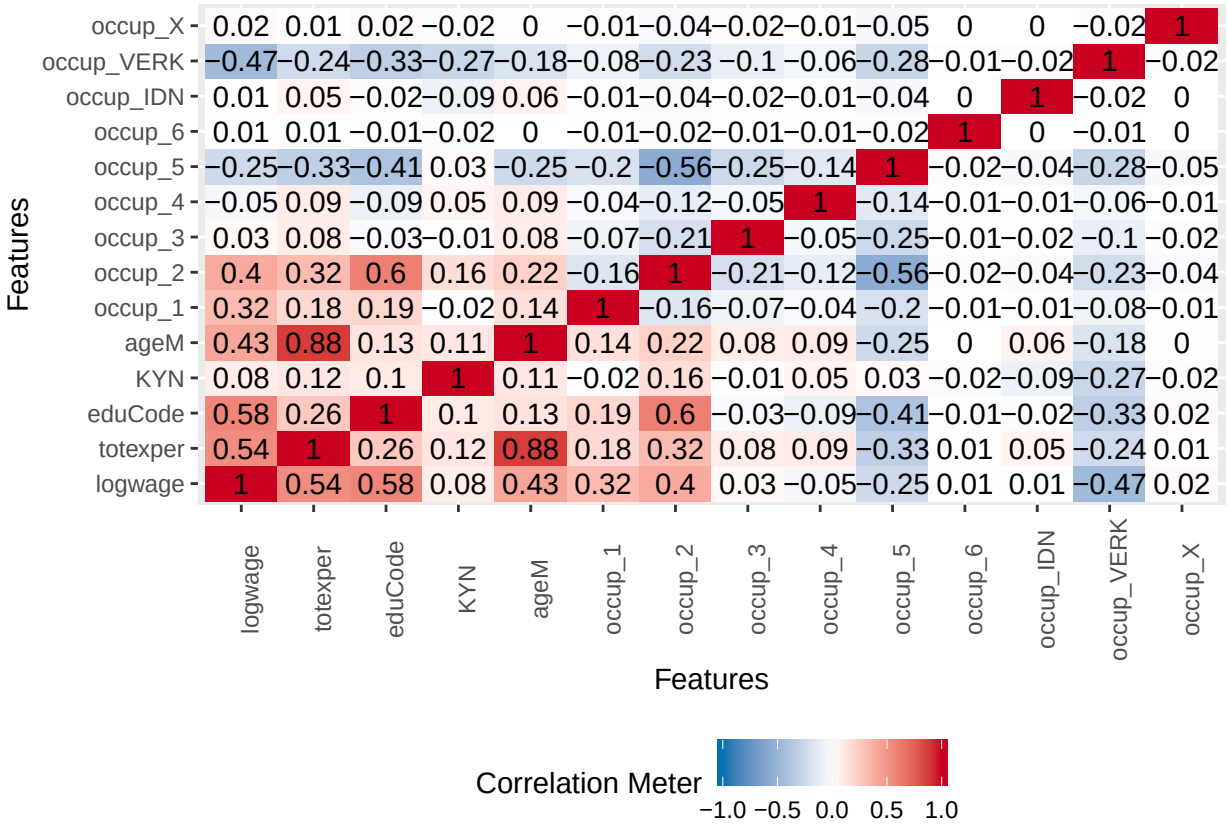
Pairwise bivariate distributions and correlation plots

Note that: printing is not yet "addapted" to size of data and paper

```
## [[1]]
```

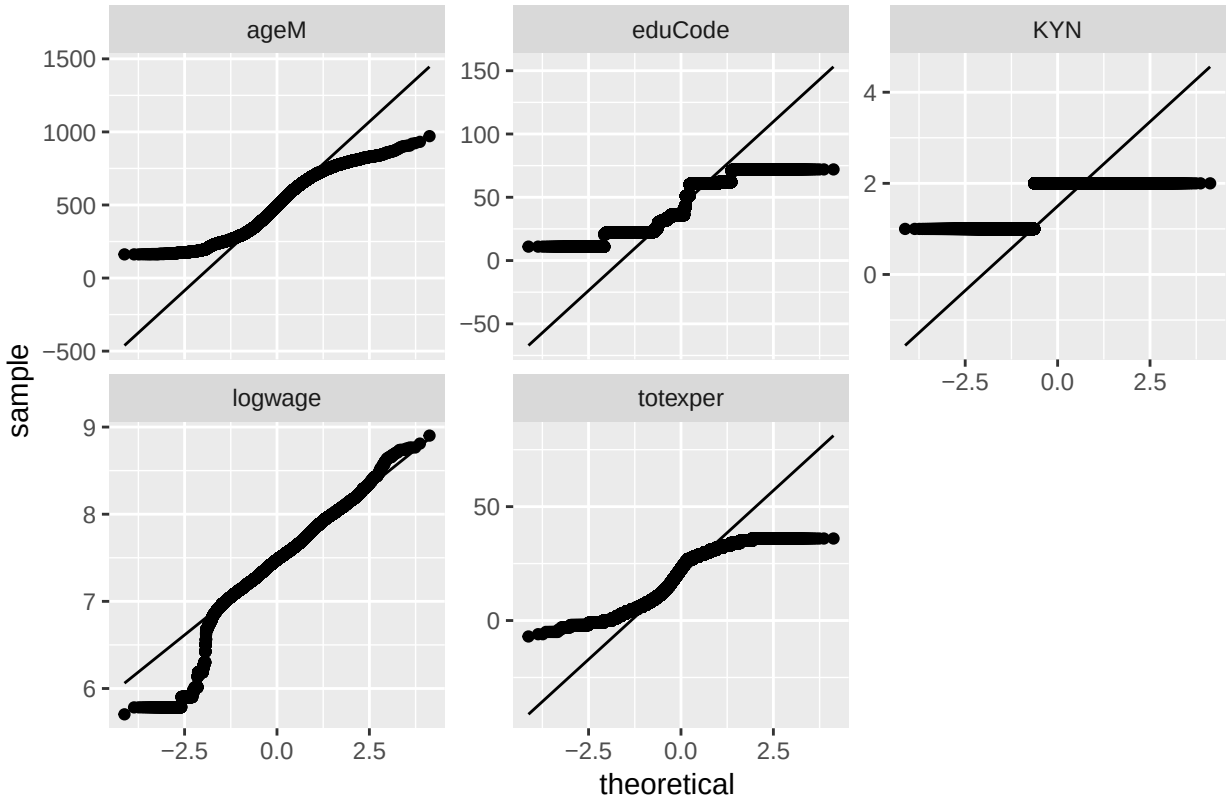
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

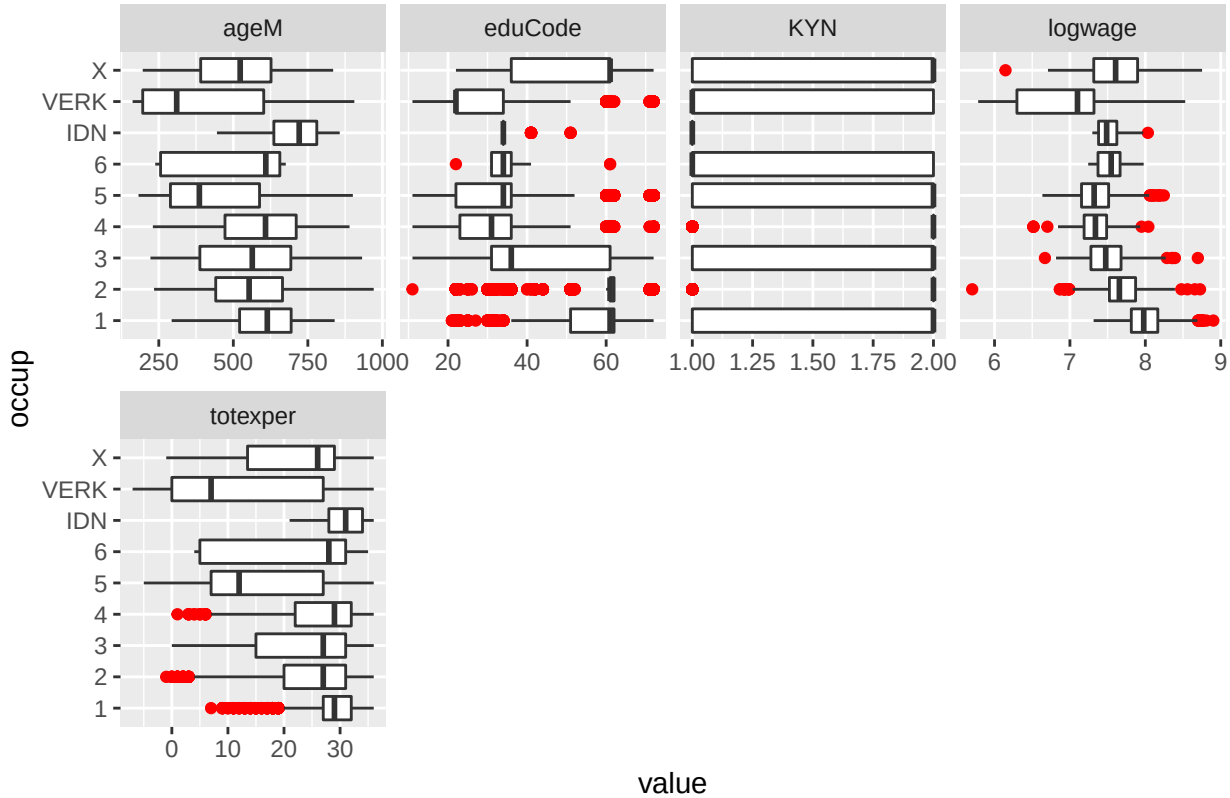

logwage
totexper
eduCode
KYN
occup
ageM
Cor : 0.539
Cor : 0.575
Cor : 0.0839
Cor : 0.427
Cor : 0.259
Cor : 0.117
Cor : 0.88
Cor : 0.1
Cor : 0.132
Cor : 0.109


occup_X 0.02 0.01 0.02 −0.02 0 −0.01 −0.04 −0.02 −0.01 −0.05 0 0 −0.02 1
occup_VERK −0.47 −0.24 −0.33 −0.27 −0.18 −0.08 −0.23 −0.1 −0.06 −0.28 −0.01 −0.02 1 −0.02
occup_IDN 0.01 0.05 −0.02 −0.09 0.06 −0.01 −0.04 −0.02 −0.01 −0.04 0 1 −0.02 0
occup_6 0.01 0.01 −0.01 −0.02 0 −0.01 −0.02 −0.01 −0.01 −0.02 1 0 −0.01 0
occup_5 −0.25 −0.33 −0.41 0.03 −0.25 −0.2 −0.56 −0.25 −0.14 1 −0.02 −0.04 −0.28 −0.05
occup_4 −0.05 0.09 −0.09 0.05 0.09 −0.04 −0.12 −0.05 1 −0.14 −0.01 −0.01 −0.06 −0.01
occup_3 0.03 0.08 −0.03 −0.01 0.08 −0.07 −0.21 1 −0.05 −0.25 −0.01 −0.02 −0.1 −0.02
occup_2 0.4 0.32 0.6 0.16 0.22 −0.16 1 −0.21 −0.12 −0.56 −0.02 −0.04 −0.23 −0.04
occup_1 0.32 0.18 0.19 −0.02 0.14 1 −0.16 −0.07 −0.04 −0.2 −0.01 −0.01 −0.08 −0.01
ageM 0.43 0.88 0.13 0.11 1 0.14 0.22 0.08 0.09 −0.25 0 0.06 −0.18 0
KYN 0.08 0.12 0.1 1 0.11 −0.02 0.16 −0.01 0.05 0.03 −0.02 −0.09 −0.27 −0.02
eduCode 0.58 0.26 1 0.1 0.13 0.19 0.6 −0.03 −0.09 −0.41 −0.01 −0.02 −0.33 0.02
totexper 0.54 1 0.26 0.12 0.88 0.18 0.32 0.08 0.09 −0.33 0.01 0.05 −0.24 0.01
logwage 1 0.54 0.58 0.08 0.43 0.32 0.4 0.03 −0.05 −0.25 0.01 0.01 −0.47 0.02
Features
logwage
totexper
eduCode
KYN
ageM
occup_1
occup_2
occup_3
occup_4
occup_5
occup_6
occup_IDN
occup_VERK
occup_X
Features
Correlation Meter
−1.0 −0.5 0.0 0.5 1.0

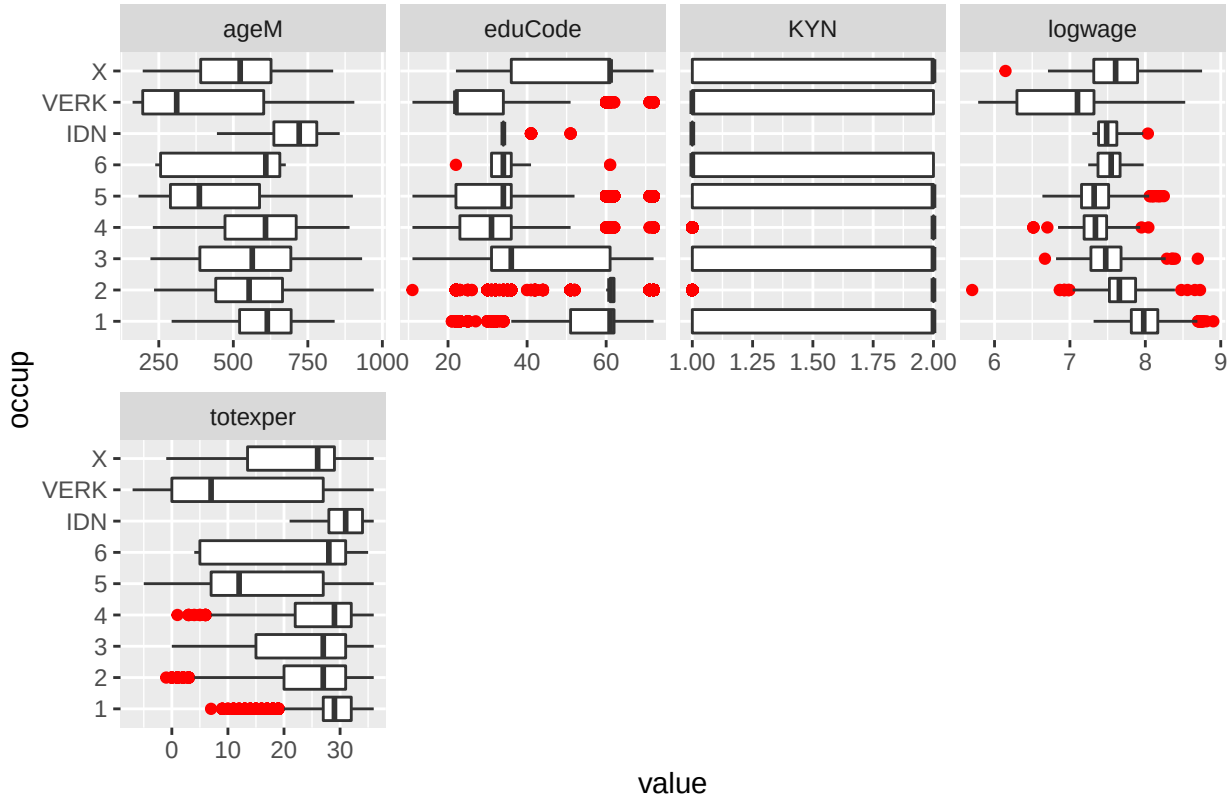
view_outliers

Plots and boxplots,limits based on Tukey and Hampel methods





```
## [[1]]
## [[1]]$page_1
```



	logwage
bottom_threshold	5.8581263054266
top_threshold	9.07230149588056

	totexper
bottom_threshold	-50
top_threshold	90

	eduCode
bottom_threshold	-83
top_threshold	169

	KYN
bottom_threshold	-2
top_threshold	5

	ageM

bottom_threshold	-599
top_threshold	1583.25

	logwage
bottom_threshold	6.45443641487166
top_threshold	8.49704241172387

	totexper
bottom_threshold	-17.0302
top_threshold	63.0302

	eduCode
bottom_threshold	-26.2692
top_threshold	98.2692

	KYN
bottom_threshold	2
top_threshold	2

	ageM
bottom_threshold	-198.8568
top_threshold	1188.8568

view_assoc

With validation rules mining potential.Under development.

view_clusters

Potentialy identifying unwanted structures or confirming known ones. Under development.

rev_variability

information theory based measures, for categorical variables

```
## [1] "var1" "var2" " "    " "    " "    " "
```

rev_ts

univariate and multivariate time series: detection of anomalous features, tests of stationarity and (auto/cross)-correlation

rev_model()

Model testing

check_assumptions()

Checking test or model assumptions aout data

reviewed

Reporting function