RProject

MariaRomanova

14 05 2019

Download required pachages

Download datasets

Tidy check

New subsetting

Correlation test

Check common variables

'X_CASCTL','X_PPAIR','X_UGUA24','X_UXGA24','X_UGUO24','X_UXGO24','X_UDG24','X_UXDG24'
SEX 1 - M, 2 - F

```
ipak <- function(pkg){
    new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
    if (length(new.pkg))
        install.packages(new.pkg, dependencies = TRUE)
    sapply(pkg, require, character.only = TRUE)
}
#if (!require('UpSetR'){devtools::install_github('hms-dbmi/UpSetR@fe2812c')})
required_libraries = c('sas7bdat','tidyr','dplyr','reshape2','plyr','ggplot2','UpSetR','data.table')
ipak(required_libraries)
```

```
## Loading required package: sas7bdat
## Loading required package: tidyr
## Loading required package: dplyr
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
## Loading required package: reshape2
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths

## Loading required package: plyr

## ------------------------------------------------------------------------------

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

## ------------------------------------------------------------------------------

##
## Attaching package: 'plyr'

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

## Loading required package: ggplot2

## Loading required package: UpSetR

## Loading required package: data.table

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:reshape2':
##
##     dcast, melt

## The following objects are masked from 'package:dplyr':
##
##     between, first, last

##   sas7bdat      tidyr      dplyr   reshape2       plyr    ggplot2
##       TRUE       TRUE       TRUE       TRUE       TRUE       TRUE
##     UpSetR data.table
##       TRUE       TRUE
```

```
analysis <- data.frame(sas7bdat::read.sas7bdat('Baseline/assay.sas7bdat'))
oldtnfa <- data.frame(sas7bdat::read.sas7bdat('Baseline/oldtnfa.sas7bdat'))
analysis_2 <- data.frame(sas7bdat::read.sas7bdat('Follow_up1/assay.sas7bdat'))
analysis_3 <- data.frame(sas7bdat::read.sas7bdat('Follow_up2/assay.sas7bdat'))
head(analysis)
```

```
##   CODE98 SITE SEX DATA_NAS X_DATEL X_AGEL X_VUOTO X_PIENO X_BUSTE X_INIZIO
## 1      1    1   2    -4464   14263     51     165    1290       0    06:30
## 2      3    1   2   -11529   14306     70     NaN     NaN     NaN
## 3      4    1   2   -14626   14306     79     163    2105       1    07:00
## 4      5    1   1    -2570   14333     46     164    1395       1    10:10
## 5      8    1   2   -14951   14397     80     172    1340       1    07:00
## 6      9    1   2   -14462   14186     78     147    1840       0    07:00
##   X_FINE X_TURINE X_QURINE   X_U_MIN X_UCREAT X_UCRE24 X_CL24 X_UCORSL
## 1  06:00     1410   1127.5 0.7996454       80    902.0   73.6    0.082
## 2               NaN      NaN       NaN      NaN      NaN    NaN       NaN
## 3  07:00     1440   1918.6 1.3323611       40    767.4   63.4    0.082
```

```
## 4  06:30     1220   1289.1 1.0566393      100   1289.1  136.6    0.086
## 5  06:30     1410   1170.9 0.8304255       50    585.5   64.3    0.096
## 6  07:00     1440   1677.9 1.1652083       50    839.0   60.9    0.051
##   X_UCOR24 X_UCA X_UCA24H X_UNA  X_UNA24H X_3_MH X_TELOP  X_TELOPR
## 1     92.4  6.14 6.922932 228.1 257.18581     76  325.33  46.53454
## 2      NaN  1.10      NaN  71.0       NaN     27     NaN       NaN
## 3    156.6  1.94 3.722026  43.0  82.49851     60  126.15  36.55115
## 4    110.5  1.93 2.487905 124.0 159.84469    178  664.90  63.13756
## 5    112.0  2.03 2.376986 140.0 163.93006     48  447.43 103.18284
## 6     84.8   NaN      NaN   NaN       NaN    129  114.77  28.19350
##   X_UTPEGE X_U_TPE X_U_TPEC X_UTPECC   X_UHEP X_UHEP24    X_UBPA X_UBPA24
## 1      NaN     NaN      NaN      NaN      NaN      NaN 4.8811069 5.620543
## 2      NaN     NaN      NaN      NaN      NaN      NaN       NaN      NaN
## 3   149.66  195.02   160.87   209.63 103.3648 79322.12 1.7517197 3.360849
## 4      NaN     NaN      NaN      NaN      NaN      NaN 6.3111305 9.602768
## 5      NaN     NaN      NaN      NaN      NaN      NaN 2.0214212 2.417241
## 6   397.16  473.37   488.20   581.89      NaN      NaN 0.8485329 1.423753
##   X_PPAIR X_CASCTL X_UGUA24 X_UXGA24 X_UGUO24 X_UXGO24 X_UDG24 X_UXDG24
## 1     NaN      NaN      NaN      NaN      NaN      NaN     NaN      NaN
## 2     NaN      NaN      NaN      NaN      NaN      NaN     NaN      NaN
## 3     121        1     1088    128.5    439.4    26.86    12.3    15.62
## 4     NaN      NaN      NaN      NaN      NaN      NaN     NaN      NaN
## 5     NaN      NaN      NaN      NaN      NaN      NaN     NaN      NaN
## 6     NaN      NaN      NaN      NaN      NaN      NaN     NaN      NaN
##   X_U_PH X_U_GLU X_U_PRO X_U_HB X_U_CC X_U_BIL X_U_URO X_U_NIT X_U_PS
## 1    5.5       0       0    0.0      0       0     0.2       0  1.019
## 2    5.0       0       0    0.2      0       0     0.2       0  1.006
## 3    5.0       0       0    0.0      0       0     0.2       0  1.007
## 4    6.0       0       0    0.0      0       0     0.2       0  1.118
## 5    7.0       0       0    0.0      0       0     0.2       0  1.010
## 6    5.0       0       0    0.0      0       0     0.2       0  1.009
##                                 X_U_SEDI X_USEDIA X_GLU X_BUN X_CREA
## 1                      ALCUNI LEUCOCITI       32    93    37   0.84
## 2                                                88    27   0.80
## 3                                               132    29   0.82
## 4                                                90    31   0.78
## 5                     NUMEROSI LEUCOCITI       20    70    34   0.62
## 6 LEUCOCITI E ABBONDANTE FLORA BATTERICA       12    98    35   0.88
##   X_URICO X_COLTOT X_COLHDL X_TRIGLI X_COLLDL    X_LP_A  X_OX_LDL X_NA X_K
## 1     4.8      196       80       63      103 40.527730  32.26055  144 4.2
## 2     4.7      234       48      148      156 20.830738  45.56022  143 4.9
## 3     5.6      224       46      197      139  7.927636  52.56838  143 4.0
## 4     2.4      222       46       75      161 18.210082  39.65529  137 4.0
## 5     3.5      291       48      209      201 59.735948 100.10414  139 3.8
## 6     7.2      213       50      209      121 60.256630  24.20469  144 4.5
##   X_CL X_CA X_MG X_MIO X_GOT X_GPT X_PALK X_GGT X_ALDO X_CPK X_LDH X_PT
## 1  113  9.1  2.1  33.0    14     8    116   113    1.6    52   282  6.6
## 2  110  9.8  2.1  23.0    18    17    235    26    3.5   114   369  7.1
## 3  110  9.2  2.1  96.0    17    20    245    21    1.6    33   288  7.4
## 4  109  9.4  2.5  26.0    17    28    276    55    2.5    56   274  7.5
## 5  109  9.0  2.3  53.0    17    16    242    27    3.2   103   357  7.3
## 6  105  9.5  2.1   0.1    21    11    163    16    0.1   120   301  7.1
##   X_ALB X_ALFA1 X_ALFA2 X_BETA X_GAMMA X_A_G X_ALF2M X_GB X_N_NEU X_N_LIN
## 1  60.5     2.4    12.8   10.4    13.9  1.53     159 6.97    4.58    1.95
```

```
## 2  54.3     2.8    11.9   11.3    19.7  1.19     168 6.18    3.30    1.65
## 3  55.1     2.8    11.2   13.2    17.7  1.23     167 4.37    3.39    0.71
## 4  55.4     3.1    12.2   13.0    16.3  1.24     161 8.04    4.65    2.80
## 5  57.2     2.6    11.1   13.4    15.7  1.34     233 6.11    3.95    1.69
## 6  56.0     2.8    12.0   12.7    16.5  1.27     115 7.20    3.59    2.87
##   X_N_MON X_N_EOS X_N_BAS X_P_NEU X_P_LIN X_P_MON X_P_EOS X_P_BAS X_GR
## 1    0.33    0.09    0.02    65.7    28.0     4.7     1.3     0.3 4.37
## 2    0.20    1.00    0.03    53.4    26.7     3.2    16.2     0.5 4.75
## 3    0.20    0.05    0.02    77.6    16.2     4.6     1.1     0.5 4.53
## 4    0.43    0.16    0.00    57.9    34.8     5.3     2.0     0.0 4.99
## 5    0.32    0.09    0.06    64.6    27.7     5.2     1.5     1.0 4.42
## 6    0.51    0.21    0.02    49.8    39.9     7.1     2.9     0.3 4.31
##   X_HB X_HCT X_VGM X_CNTME X_CNCME X_IDE X_PIAST X_V_PIAS X_VES X_FTIN
## 1 12.8  38.3  87.6    29.3    33.4  13.4     174     11.9    15     21
## 2 11.8  36.5  76.8    24.8    32.3  15.0     225     12.1    21     71
## 3 12.8  38.5  85.0    28.3    33.2  13.8     199     10.8    28     41
## 4 13.3  38.9  78.0    26.7    34.2  14.6     301     10.2    35    163
## 5 13.0  40.5  91.6    29.4    32.1  13.7     231     12.3    26    202
## 6 13.1  40.0  92.8    30.4    32.8  13.9     214     12.9    45    103
##   X_FERRO X_FIBRIN X_LS_WBC X_LS_LYM X_LS_CD4 X_LS_CD8 X_LS_T X_LS_B
## 1      52      355      NaN      NaN      NaN      NaN    NaN    NaN
## 2      91      323      NaN      NaN      NaN      NaN    NaN    NaN
## 3      36      357      NaN      NaN      NaN      NaN    NaN    NaN
## 4      39      487      NaN      NaN      NaN      NaN    NaN    NaN
## 5      86      363      NaN      NaN      NaN      NaN    NaN    NaN
## 6      47      378      NaN      NaN      NaN      NaN    NaN    NaN
##   X_LS_TDR X_LCD4_8 X_HOMCYS X_FOLICG X_FOLICM X_VIB12G X_VIB12M X_VITB6G
## 1      NaN      NaN     15.4      2.5     5.68      310   229.40    4.983
## 2      NaN      NaN     12.8      2.9     6.58      276   204.24    3.582
## 3      NaN      NaN      9.9      4.5    10.22      503   372.22    6.226
## 4      NaN      NaN      9.8      5.0    11.35      478   353.72    3.220
## 5      NaN      NaN     14.8      1.5     3.41      450   333.00    6.789
## 6      NaN      NaN      NaN      5.6    12.71      407   301.18   16.865
##   X_VITB6M X_GAMTOC X_ALFTOC X_GTOCRS X_ATOCRS   X_ACAROT  X_BCAROT
## 1    20.18 1.376502 26.69139 2.388278 32.88080 0.05551740 0.2972789
## 2    14.51 2.060130 31.95424 4.187345 43.38082 0.04141449 0.4800951
## 3    25.22 2.798211 28.21352 3.184386 36.87202 0.06490692 0.6465914
## 4    13.04 1.625111 31.62231 1.700746 30.15457 0.06108777 0.2768977
## 5    27.50 1.477774 28.95987 1.944763 30.45571 0.04439529 0.3465739
## 6    68.30 2.271205 30.62456      NaN      NaN        NaN       NaN
##     X_BCRYPT  X_LUTEIN  X_LYCOPN X_RETINL   X_ZEAXAN  X_SE X_25OH_D X_PTH
## 1 0.07577901 0.3884477 0.6746855 1.894085 0.06334074 82.54  46.4256  15.4
## 2 0.24434163 0.3880785 0.3923851 1.541614 0.04878450 72.31  46.1760  22.2
## 3 0.21518055 0.3114297 0.3362342 1.552421 0.05929734 98.58  19.2192  21.8
## 4 0.11749455 0.4203730 0.9793977 1.115228 0.12580248 89.38  29.2032  25.4
## 5 0.09868095 0.1939426 1.0460372 1.707053 0.03387666 69.63  31.2000  27.4
## 6        NaN       NaN       NaN      NaN        NaN   NaN  37.9392  14.1
##   X_TSH X_FT4 X_FT3 X_TIGF1 X_FRIGF1 X_IGFBP1 XIGFBP3C XIGFBP3M X_INSULN
## 1  1.50  1.42  4.11   94.51     0.57     8.67     3914  136.990    9.600
## 2  0.41  1.52  4.29   89.87     0.46      NaN      NaN      NaN   18.000
## 3  1.82  1.32  4.15   66.80     0.42    34.03     4143  145.005   15.000
## 4  2.49  1.51  4.17  149.88     0.52    12.35     4194  146.790   16.500
## 5  0.63  0.99  4.33   63.85     0.54    20.89     3265  114.275   17.207
## 6  2.05  1.63  4.47   81.84     0.58    44.34     5676  198.660   15.167
```

```
##   X_ADIPON X_RESIST X_CRP_LS X_CRP_HS X_GP130 X_IL6  X_IL6_EC  X_IL6R
## 1    14.38      6.4      4.0     3.06 297.614  1.12  2.448130  41.804
## 2    15.37      4.6      6.0     4.51 347.361  0.53  1.499784  88.177
## 3    10.79      4.7      6.0     4.44 264.653  3.23  5.540237 124.633
## 4     4.13      5.8     45.9    51.00 377.050 11.01 12.897282 146.043
## 5    20.68      4.5      6.0     1.70 296.345  1.94  3.705400  66.767
## 6    13.07      3.1      6.0      NaN 372.475  1.07  2.369180 146.295
##   X_IL10 X_IL1RA X_IL1B X_IL15 X_IL18 X_TGFB1 X_TNFA_M X_TNFAR1 X_TNFAR2
## 1   7.77  178.82   0.12  3.517 150.49 15650.0     2.94   750.33   1921.1
## 2   0.00  123.83   0.16  2.344 370.43 10952.0     5.50  1300.20   3054.5
## 3   0.00  120.30   0.08  1.959 376.52  3265.1     5.26  1569.40   3311.5
## 4   0.00  231.65   0.30  1.784 325.65 17221.0     4.02   957.74   1732.9
## 5   0.00  157.20   0.08  2.779 371.62  2596.7    57.58  1223.50   2890.8
## 6   0.00   23.46   0.16  2.894    NaN 20026.0     5.65  1259.30   2333.2
##   X_TRAIL X_IFNG_B X_IL8_B X_IL12_B X_MCP1_B X_MIP1B X_CORTIS X_CONMOL
## 1 190.330        0       0        0        0   35.69    10.30  284.280
## 2 105.060        0       0        0        0   89.92    12.48  344.448
## 3  75.496      NaN     NaN      NaN      NaN     NaN    11.35  313.260
## 4  71.309        0       0        0        0   46.97    23.48  648.048
## 5  78.494      NaN     NaN      NaN      NaN     NaN    14.67  404.892
## 6  87.255      NaN     NaN      NaN      NaN     NaN    18.53  511.428
##   X_DHEAS X_DHNMOL  X_CORTDH X_SHBG X_TESTO X_TENMOL  X_FREETS   XFREETSM
## 1   55.91 1515.161 0.1876236  51.50    0.55   1.9085 0.7646909 0.02653477
## 2   89.68 2430.328 0.1417290 142.40    0.56   1.9432 0.3464319 0.01202119
## 3   72.47 1963.937 0.1595061  56.20    0.93   3.2271 1.2219739 0.04240250
## 4  138.22 3745.762 0.1730083  57.58    5.52  19.1544 8.1823889 0.28392890
## 5   87.44 2369.624 0.1708676 111.18    0.73   2.5331 0.5541803 0.01923006
## 6  127.30 3449.830 0.1482473  89.07    0.75   2.6025 0.6905228 0.02396114
##     X_BIOATS  XBIOATSM   X_TSSHBG X_ESTDIO X_ESTDIM X_EPO X_STFRNM
## 1  16.695529 0.5793349 0.03705825    70.19     0.26  11.9     15.2
## 2   7.314773 0.2538226 0.01364607     3.50     0.01   NaN     14.4
## 3  27.217504 0.9444474 0.05742171     4.50     0.02  14.2     18.4
## 4 185.562351 6.4390136 0.33265717     8.39     0.03  14.1     33.6
## 5  12.627407 0.4381710 0.02278377     5.70     0.02   7.3     16.0
## 6  15.014967 0.5210194 0.02921859     7.11     0.03  10.0      NaN
##   X_STFRMG X_CLMIGG X_CLMIGM X_LEPTIN X_GHRELN  X_90K X_AGECML  X_ESRAGE
## 1     1.14        0      NaN 19.64128   0.3025 42.079 291.0440 0.3640160
## 2     1.08     1024       24 20.96080   0.2295    NaN 418.1063 0.5284556
## 3     1.38       64      NaN      NaN      NaN 40.848 340.9832 0.2415157
## 4     2.52       64      NaN  2.91808   0.3355 24.868 290.8089 0.4119085
## 5     1.20       16      NaN      NaN      NaN 21.610 314.9285 0.3436890
## 6      NaN        0      NaN      NaN      NaN 22.081 445.8912       NaN
##   X_CYSC  X_SCD14 X_CTX_1 X_PIIINP    X_ADMA    X_SDMA  X_L_ARG     X_HA
## 1   0.64 2068.471   0.289   3.3675 0.4993820 0.4860321 47.58898 1.986955
## 2   0.97 1991.847   0.292   5.1491 0.6378245 0.5092707 44.08726 2.295093
## 3   1.06 1669.264   0.376   3.7126 0.5883807 0.4177998 30.08037 2.385409
## 4   0.69 1702.040   0.496   3.6888 0.3772559 0.2674907 44.83352 1.758508
## 5   0.59 1486.313   0.562   4.1070 0.5735476 0.4993820 13.08840 1.530062
## 6   0.97 1555.233   0.234   4.8570 0.5438813 0.4514215 64.29392 3.437326
##        X_MMA X_C14_0  X_C14_0A X_C14_0B  X_C14_0C  X_C14_0D  X_C14_0E
## 1        NaN  109.77 0.7207537 15.27115 0.5989884  66.87309 0.7308496
## 2 0.14131820  415.66 1.3315573 35.08056 1.1158814 153.61942 1.3571730
## 3 0.10891064  290.73 1.7017677 57.28298 1.4358625 250.84484 1.7364408
## 4 0.06481511  249.58 1.0184661 24.04932 0.8402033 105.31311 1.0259564
```

```
## 5 0.11475463  528.80 1.8479515 64.73914 1.5595061 283.49572 1.8908075
## 6 0.14131820  219.43 1.1995909 39.24173 1.0045208 171.84136 1.2217634
##   X_C14_1  X_C14_1A X_C14_1B  X_C14_1C X_C14_1D  X_C14_1E X_C16_0 X_C16_0A
## 1   28.64 0.1880513 4.450034 0.1745460 19.65570 0.2148153 3621.75 23.78054
## 2   56.07 0.1796190 5.201134 0.1654434 22.97330 0.2029609 7743.38 24.80574
## 3   33.97 0.1988410 7.211643 0.1807680 31.85367 0.2205029 4706.92 27.55163
## 4   36.89 0.1505378 4.072820 0.1422908 17.98956 0.1752536 5828.58 23.78480
## 5   64.78 0.2263810 8.408839 0.2025612 37.14166 0.2477206 7197.58 25.15276
## 6   24.97 0.1365072 4.977788 0.1274228 21.98678 0.1563224 4393.61 24.01921
##    X_C16_0B X_C16_0C X_C16_0D X_C16_0E X_C16_1 X_C16_1A  X_C16_1B X_C16_1C
## 1  549.5168 21.55399 2143.030 23.42097  231.51 1.520103  34.46317 1.351767
## 2  711.3350 22.62693 2774.096 24.50815 1036.07 3.319027  95.26248 3.030215
## 3 1007.7396 25.26013 3930.028 27.20511  604.04 3.535706 129.92879 3.256812
## 4  612.6516 21.40401 2389.246 23.27595  380.65 1.553326  39.27817 1.372249
## 5  957.2995 23.06046 3733.319 24.89981 1069.39 3.737104 143.13214 3.447921
## 6  854.6518 21.87762 3333.009 23.69714  568.46 3.107685 110.81830 2.836758
##   X_C16_1D X_C16_1E X_C18_0 X_C18_0A X_C18_0B X_C18_0C  X_C18_0D X_C18_0E
## 1 135.4683 1.480520 1266.64 8.316803 192.2479 7.540640  675.8096 7.385858
## 2 374.4590 3.308213 3088.96 9.895413 281.4790 8.953596  989.4838 8.741738
## 3 510.7259 3.535434 1095.38 6.411730 234.7495 5.884262  825.2157 5.712449
## 4 154.3952 1.504112  949.75 3.875664 103.6430 3.620941  364.3364 3.549352
## 5 562.6258 3.752498 2646.01 9.246782 348.2667 8.389425 1224.2631 8.165364
## 6 435.6061 3.097087 1230.52 6.727068 239.1999 6.123106  840.8601 5.978375
##   X_C18_9 X_C18_9A  X_C18_9B X_C18_9C X_C18_9D X_C18_9E X_C18_7 X_C18_7A
## 1 3691.17 24.23635  580.2656 22.76006 2054.409 22.45244  196.16 1.287994
## 2 8579.62 27.48462  818.4302 26.03354 2897.623 25.59947  377.54 1.209441
## 3 4869.63 28.50404 1084.5191 27.18470 3839.701 26.57983  242.17 1.417525
## 4 7368.79 30.06997  802.4429 28.03469 2841.020 27.67711  246.15 1.004469
## 5 8259.69 28.86442 1143.0040 27.53391 4046.764 26.99036  360.80 1.260856
## 6 5654.63 30.91301 1142.9298 29.25704 4046.502 28.76995  280.80 1.535091
##   X_C18_7B  X_C18_7C  X_C18_7D  X_C18_7E X_C18_6 X_C18_6A X_C18_6B
## 1 30.77846 1.2072397 108.97003 1.1909230 4573.35 30.02878 784.0435
## 2 35.72512 1.1363845 126.48351 1.1174369 6602.19 21.14997 686.6163
## 3 52.96608 1.3276545 187.52448 1.2981138 3681.34 21.54847 894.3455
## 4 27.08520 0.9462667  95.89416 0.9341974 6657.18 27.16609 790.4044
## 5 49.01447 1.1807133 173.53394 1.1574047 5816.30 20.32572 878.1865
## 6 55.66154 1.4248398 197.06766 1.4011182 4059.78 22.19421 895.1212
##   X_C18_6C X_C18_6D X_C18_6E X_C18_3  X_C18_3A  X_C18_3B  X_C18_3C
## 1 30.75295 2795.169 30.54814   30.27 0.1987539  4.840135 0.1898472
## 2 21.84065 2447.834 21.62575  109.66 0.3512933 11.491512 0.3655348
## 3 22.41778 3188.403 22.07131   57.13 0.3344065 14.063137 0.3525084
## 4 27.61410 2817.846 27.45136   74.74 0.3049930  8.701343 0.3039960
## 5 21.15470 3130.796 20.88120  174.58 0.6100896 27.470690 0.6617436
## 6 22.91357 3191.169 22.68868   75.19 0.4110524 16.949209 0.4338707
##   X_C18_3D  X_C18_3E X_C20_0     X_C20_0A X_C20_0B    X_C20_0C  X_C20_0D
## 1 17.38435 0.1899919   21.49 0.1411041052 3.648961 0.143125097 11.675867
## 2 41.27416 0.3646425   19.16 0.0613786228 1.754958 0.055823671  5.615479
## 3 50.51068 0.3496536   10.10 0.0591196441 2.224283 0.055754162  7.117213
## 4 31.25268 0.3044625   22.78 0.0929587994 2.497456 0.087252803  7.991306
## 5 98.66670 0.6580690    0.22 0.0007688149 0.050000 0.001204454  0.160000
## 6 60.87662 0.4328226   10.57 0.0577846028 2.086319 0.053406195  6.675760
##      X_C20_0E X_C20_1  X_C20_1A X_C20_1B  X_C20_1C X_C20_1D  X_C20_1E
## 1 0.127604431   21.54 0.1414324 4.219801 0.1655155 13.59029 0.1485270
## 2 0.049610764   56.23 0.1801315 6.578573 0.2092586 21.18696 0.1871792
```

```
## 3 0.049267984   22.81 0.1335167 6.257881 0.1568609 20.15414 0.1395144
## 4 0.077851011   28.99 0.1183001 3.967809 0.1386221 12.77873 0.1244899
## 5 0.001067139   34.05 0.1189916 5.801572 0.1397545 18.68455 0.1246188
## 6 0.047463543   30.45 0.1664656 7.539422 0.1929963 24.28147 0.1726372
##   X_C20_2   X_C20_2A X_C20_2B   X_C20_2C  X_C20_2D   X_C20_2E X_C20_3
## 1   25.77 0.16920674 5.023316 0.19703218 16.283021 0.17795558  199.33
## 2   31.41 0.10062122 3.663915 0.11654588 11.876537 0.10492499  458.22
## 3   20.43 0.11958558 5.648135 0.14157687 18.308362 0.12673726  224.81
## 4   22.63 0.09234669 2.955071 0.10324034  9.578827 0.09331658  508.95
## 5   19.84 0.06933313 3.340908 0.08047939 10.829513 0.07222869  431.64
## 6   19.26 0.10529153 4.801004 0.12289748 15.562399 0.11064606  262.82
##   X_C20_3A X_C20_3B X_C20_3C X_C20_3D X_C20_3E X_C20_4 X_C20_4A X_C20_4B
## 1 1.308808 42.66134 1.673328 139.1885 1.521177  916.95 6.020726 207.1841
## 2 1.467897 59.35224 1.887942 193.6448 1.710783 2033.82 6.515296 278.8192
## 3 1.315910 67.96657 1.703659 221.7502 1.535037  938.32 5.492391 300.3402
## 4 2.076882 75.19032 2.626900 245.3187 2.389886 1566.20 6.391224 245.1447
## 5 1.508415 80.96051 1.950264 264.1447 1.761744 1445.88 5.052791 287.5654
## 6 1.436797 72.08587 1.845275 235.1900 1.672162 1170.80 6.400588 340.0873
##   X_C20_4C  X_C20_4D X_C20_4E X_C20_5  X_C20_5A X_C20_5B  X_C20_5C
## 1 8.126492  680.4959 7.437075   77.88 0.5113629 17.23412 0.6759830
## 2 8.868992  915.7812 8.090602  145.05 0.4646644 19.12123 0.6082294
## 3 7.528367  986.4669 6.828689   68.13 0.3987942 20.96793 0.5255848
## 4 8.564539  805.1772 7.844006  156.75 0.6396528 23.26304 0.8127333
## 5 6.927185  944.5079 6.299505  127.56 0.4457729 23.63721 0.5693986
## 6 8.705651 1117.0161 7.941798   61.54 0.3364299 17.73239 0.4539187
##   X_C20_5D  X_C20_5E X_C22_0   X_C22_0A X_C22_0B   X_C22_0C  X_C22_0D
## 1 56.97208 0.6226424   31.15 0.20453201 8.053865 0.31590099 23.648299
## 2 63.21043 0.5584417   27.09 0.08678220 4.054092 0.12895708 11.903898
## 3 69.31519 0.4798254   12.65 0.07404589 4.417548 0.11073083 12.971100
## 4 76.90231 0.7491795   29.50 0.12038124 5.092740 0.17792339 14.953644
## 5 78.13923 0.5211586   14.51 0.05070684 3.057541 0.07365335  8.977758
## 6 58.61923 0.4167729    8.64 0.04723358 2.564899 0.06565702  7.531229
##     X_C22_0E X_C22_1   X_C22_1A  X_C22_1B  X_C22_1C X_C22_1D  X_C22_1E
## 1 0.25845000   37.37 0.24537275 10.588252 0.4153086 31.27442 0.3417952
## 2 0.10516671   95.20 0.30497103 15.660705 0.4981531 46.25688 0.4086631
## 3 0.08979076   44.45 0.26018497 17.080163 0.4281336 50.44952 0.3492303
## 4 0.14567784   41.78 0.17049248  8.643294 0.3019680 25.52962 0.2487086
## 5 0.05987819   72.33 0.25276537 17.129427 0.4126321 50.59503 0.3374494
## 6 0.05354578   15.03 0.08216675  6.227834 0.1594219 18.39510 0.1307861
##   X_C22_6  X_C22_6A X_C22_6B X_C22_6C  X_C22_6D X_C22_6E X_C24_0
## 1   88.18 0.5789930 32.51965 1.275535  98.99479 1.081905   30.48
## 2  287.34 0.9204871 61.96848 1.971162 188.64154 1.666581   16.57
## 3  136.12 0.7967689 68.51631 1.717439 208.57413 1.443827    7.70
## 4  287.27 1.1722684 70.65994 2.468623 215.09966 2.095493   16.44
## 5  267.65 0.9353332 82.90112 1.997011 252.36368 1.683169   18.77
## 6  187.71 1.0261824 85.09076 2.178177 259.02929 1.841655    6.23
##     X_C24_0A  X_C24_0B   X_C24_0C  X_C24_0D   X_C24_0E X_C24_1   X_C24_1A
## 1 0.20013277 11.088967 0.43494838 30.082343 0.32876704   30.49 0.20019843
## 2 0.05308162  3.747942 0.11921871 10.167483 0.08982610   36.84 0.11801610
## 3 0.04507141  4.067879 0.10196598 11.035412 0.07639121   17.17 0.10050339
## 4 0.06708704  4.242539 0.14822021 11.509234 0.11212253   31.88 0.13009335
## 5 0.06559389  6.073449 0.14630379 16.476159 0.10988965   65.09 0.22746438
## 6 0.03405847  3.025962 0.07745944  8.208882 0.05836378   11.63 0.06357946
##    X_C24_1B  X_C24_1C X_C24_1D  X_C24_1E X_TFA_AR X_TFA_MG  X_TFA_MO
```

```
## 1 11.391226 0.4468041 31.07273 0.3395909 15229.89 2549.490  9150.048
## 2  8.410644 0.2675351 22.94236 0.2026876 31216.08 3143.753 11319.074
## 3  9.153968 0.2294545 24.96998 0.1728515 17084.00 3989.447 14445.919
## 4  8.336060 0.2912342 22.73891 0.2215217 24505.48 2862.322 10264.872
## 5 21.219689 0.5111627 57.88258 0.3860546 28615.47 4151.259 14993.368
## 6  5.718905 0.1463942 15.59990 0.1109127 18292.07 3906.512 14065.028
##    X_SFA_A  X_SFA_W  X_SFA_M X_MUFA_A X_MUFA_W X_MUFA_M X_PUFA_A X_PUFA_W
## 1 33.36387 30.58759 32.25250 27.81950 26.52124 26.16862 38.81663 42.89117
## 2 36.23395 33.00041 34.85167 32.79582 31.34053 31.02661 30.97022 35.65906
## 3 35.84336 32.84871 34.86945 34.15032 32.76438 32.29548 30.00632 34.38691
## 4 28.95936 26.27855 28.18691 33.19719 31.22732 30.88540 37.84345 42.49413
## 5 36.36456 33.23056 35.12681 34.68799 33.42866 32.99611 28.94745 33.34079
## 6 32.08494 29.20177 31.05665 36.00451 34.14487 33.83881 31.91055 36.65336
##   X_PUFA_M  X_OM3_A  X_OM3_MG  X_OM3_W  X_OM3_M  X_OM6_A  X_OM6_MG
## 1 41.57888 1.289110  54.59391 2.141366 1.894539 36.21871  996.2509
## 2 34.12172 1.736445  92.58123 2.944927 2.589665 27.76588  969.0994
## 3 32.83508 1.529970 103.54738 2.595532 2.273306 27.16044 1200.3338
## 4 40.92770 2.116914 102.62432 3.585352 3.149135 33.64966 1038.5041
## 5 31.87708 1.991196 134.00902 3.228154 2.862396 25.44784 1169.0928
## 6 35.10454 1.773665 119.77235 3.065967 2.691251 28.70009 1240.0095
##    X_OM6_W  X_OM6_M X_OM6_3A  X_OM6_3W  X_OM6_3M  X_AAEPAA X_AAEPAW
## 1 39.07647 38.16317 28.09591 18.248390 20.143775 11.773883 12.02174
## 2 30.82619 29.82128 15.99007 10.467559 11.515494 14.021510 14.58166
## 3 30.08772 29.02673 17.75228 11.592122 12.768510 13.772494 14.32379
## 4 36.28188 35.38868 15.89562 10.119474 11.237587  9.991707 10.53795
## 5 28.16237 27.25294 12.78018  8.723986  9.521022 11.334901 12.16579
## 6 31.74212 30.74112 16.18124 10.353053 11.422617 19.025024 19.17888
##    X_AADHAA X_AADHAW
## 1 10.398616 6.371045
## 2  7.078096 4.499372
## 3  6.893329 4.383485
## 4  5.452014 3.469359
## 5  5.402130 3.468776
## 6  6.237281 3.996759
```

```
sum(complete.cases(analysis))
```

```
## [1] 0
```

```
sum(complete.cases(oldtnfa))
```

```
## [1] 1312
```

```
analysis <- join(analysis,oldtnfa)
```

```
## Joining by: CODE98, SITE, SEX, DATA_NAS, X_DATEL, X_AGEL
```

```
df <- as.data.frame(apply(analysis,MARGIN = 2, FUN = function(x){
  gsub('NaN', NA,x)
}))
colSums(is.na(analysis))
```

```
##   CODE98     SITE      SEX DATA_NAS  X_DATEL   X_AGEL  X_VUOTO  X_PIENO
##        0        0        0        0        0        0       89       89
##  X_BUSTE X_INIZIO   X_FINE X_TURINE X_QURINE  X_U_MIN X_UCREAT X_UCRE24
##       88        0        0       94       94       94       94       94
##   X_CL24 X_UCORSL X_UCOR24    X_UCA X_UCA24H    X_UNA X_UNA24H   X_3_MH
```

```
##       94      113      113      160      169      159      169      712
##  X_TELOP X_TELOPR X_UTPEGE  X_U_TPE X_U_TPEC X_UTPECC   X_UHEP X_UHEP24
##      724      727      387      387      387      387      759      759
##   X_UBPA X_UBPA24  X_PPAIR X_CASCTL X_UGUA24 X_UXGA24 X_UGUO24 X_UXGO24
##      212      218     1123     1123     1142     1176     1143     1159
##  X_UDG24 X_UXDG24   X_U_PH  X_U_GLU  X_U_PRO   X_U_HB   X_U_CC  X_U_BIL
##     1165     1142      140      140      140      141      140      140
##  X_U_URO  X_U_NIT   X_U_PS X_U_SEDI X_USEDIA    X_GLU    X_BUN   X_CREA
##      140      140      140        0        0       17       17       17
##  X_URICO X_COLTOT X_COLHDL X_TRIGLI X_COLLDL   X_LP_A X_OX_LDL     X_NA
##       18       17       17       17       17       33       33       17
##      X_K     X_CL     X_CA     X_MG    X_MIO    X_GOT    X_GPT   X_PALK
##       17       17       17       17       19       17       17       17
##    X_GGT   X_ALDO    X_CPK    X_LDH     X_PT    X_ALB  X_ALFA1  X_ALFA2
##       17       19       17       17       19       18       18       18
##   X_BETA  X_GAMMA    X_A_G  X_ALF2M     X_GB  X_N_NEU  X_N_LIN  X_N_MON
##       18       18       18       17       25       25       25       25
##  X_N_EOS  X_N_BAS  X_P_NEU  X_P_LIN  X_P_MON  X_P_EOS  X_P_BAS     X_GR
##       25       25       25       25       25       25       25       25
##     X_HB    X_HCT    X_VGM  X_CNTME  X_CNCME    X_IDE  X_PIAST X_V_PIAS
##       25       25       25       25       25       25       25       27
##    X_VES   X_FTIN  X_FERRO X_FIBRIN X_LS_WBC X_LS_LYM X_LS_CD4 X_LS_CD8
##       31       17       17       28      827      828      824      823
##   X_LS_T   X_LS_B X_LS_TDR X_LCD4_8 X_HOMCYS X_FOLICG X_FOLICM X_VIB12G
##      823      823      823      824       22       52       52       52
## X_VIB12M X_VITB6G X_VITB6M X_GAMTOC X_ALFTOC X_GTOCRS X_ATOCRS X_ACAROT
##       52       49       49       37       37       31       31       31
## X_BCAROT X_BCRYPT X_LUTEIN X_LYCOPN X_RETINL X_ZEAXAN     X_SE X_25OH_D
##       31       31       31       31       31       31       32       66
##    X_PTH    X_TSH    X_FT4    X_FT3  X_TIGF1 X_FRIGF1 X_IGFBP1 XIGFBP3C
##       66      130       81       86       53      189      113      108
## XIGFBP3M X_INSULN X_ADIPON X_RESIST X_CRP_LS X_CRP_HS  X_GP130    X_IL6
##      108       53      133      133       18       30       73       16
## X_IL6_EC   X_IL6R   X_IL10  X_IL1RA   X_IL1B   X_IL15   X_IL18  X_TGFB1
##       23       15       19       15       16      153       31       26
## X_TNFA_M X_TNFAR1 X_TNFAR2  X_TRAIL X_IFNG_B  X_IL8_B X_IL12_B X_MCP1_B
##       66       65       65       61      109      110      110      109
##  X_MIP1B X_CORTIS X_CONMOL  X_DHEAS X_DHNMOL X_CORTDH   X_SHBG  X_TESTO
##      110       26       26       26       26       29       29       31
## X_TENMOL X_FREETS XFREETSM X_BIOATS XBIOATSM X_TSSHBG X_ESTDIO X_ESTDIM
##       31       42       42       42       42       35      113      113
##    X_EPO X_STFRNM X_STFRMG X_CLMIGG X_CLMIGM X_LEPTIN X_GHRELN    X_90K
##       96       52       52       40     1158      116      116      167
## X_AGECML X_ESRAGE   X_CYSC  X_SCD14  X_CTX_1 X_PIIINP   X_ADMA   X_SDMA
##       59       48      154      116      116      163       39       39
##  X_L_ARG     X_HA    X_MMA  X_C14_0 X_C14_0A X_C14_0B X_C14_0C X_C14_0D
##       39       39       55      161      161      161      161      161
## X_C14_0E  X_C14_1 X_C14_1A X_C14_1B X_C14_1C X_C14_1D X_C14_1E  X_C16_0
##      161      161      161      161      161      161      161      161
## X_C16_0A X_C16_0B X_C16_0C X_C16_0D X_C16_0E  X_C16_1 X_C16_1A X_C16_1B
##      161      161      161      161      161      161      161      161
## X_C16_1C X_C16_1D X_C16_1E  X_C18_0 X_C18_0A X_C18_0B X_C18_0C X_C18_0D
##      161      161      161      161      161      161      161      161
## X_C18_0E  X_C18_9 X_C18_9A X_C18_9B X_C18_9C X_C18_9D X_C18_9E  X_C18_7
```

```
##      161      161      161      161      161      161      161      161
## X_C18_7A X_C18_7B X_C18_7C X_C18_7D X_C18_7E  X_C18_6 X_C18_6A X_C18_6B
##      161      161      161      161      161      161      161      161
## X_C18_6C X_C18_6D X_C18_6E  X_C18_3 X_C18_3A X_C18_3B X_C18_3C X_C18_3D
##      161      161      161      161      161      161      161      161
## X_C18_3E  X_C20_0 X_C20_0A X_C20_0B X_C20_0C X_C20_0D X_C20_0E  X_C20_1
##      161      161      161      161      161      161      161      161
## X_C20_1A X_C20_1B X_C20_1C X_C20_1D X_C20_1E  X_C20_2 X_C20_2A X_C20_2B
##      161      161      161      161      161      161      161      161
## X_C20_2C X_C20_2D X_C20_2E  X_C20_3 X_C20_3A X_C20_3B X_C20_3C X_C20_3D
##      161      161      161      161      161      161      161      161
## X_C20_3E  X_C20_4 X_C20_4A X_C20_4B X_C20_4C X_C20_4D X_C20_4E  X_C20_5
##      161      161      161      161      161      161      161      161
## X_C20_5A X_C20_5B X_C20_5C X_C20_5D X_C20_5E  X_C22_0 X_C22_0A X_C22_0B
##      161      161      161      161      161      161      161      161
## X_C22_0C X_C22_0D X_C22_0E  X_C22_1 X_C22_1A X_C22_1B X_C22_1C X_C22_1D
##      161      161      161      161      161      161      161      161
## X_C22_1E  X_C22_6 X_C22_6A X_C22_6B X_C22_6C X_C22_6D X_C22_6E  X_C24_0
##      161      161      161      161      161      161      161      161
## X_C24_0A X_C24_0B X_C24_0C X_C24_0D X_C24_0E  X_C24_1 X_C24_1A X_C24_1B
##      161      161      161      161      161      161      161      161
## X_C24_1C X_C24_1D X_C24_1E X_TFA_AR X_TFA_MG X_TFA_MO  X_SFA_A  X_SFA_W
##      161      161      161      161      161      161      161      161
##  X_SFA_M X_MUFA_A X_MUFA_W X_MUFA_M X_PUFA_A X_PUFA_W X_PUFA_M  X_OM3_A
##      161      161      161      161      161      161      161      161
## X_OM3_MG  X_OM3_W  X_OM3_M  X_OM6_A X_OM6_MG  X_OM6_W  X_OM6_M X_OM6_3A
##      161      161      161      161      161      161      161      161
## X_OM6_3W X_OM6_3M X_AAEPAA X_AAEPAW X_AADHAA X_AADHAW   X_TNFA X_TNFA_E
##      161      161      162      162      163      163       17       29
```

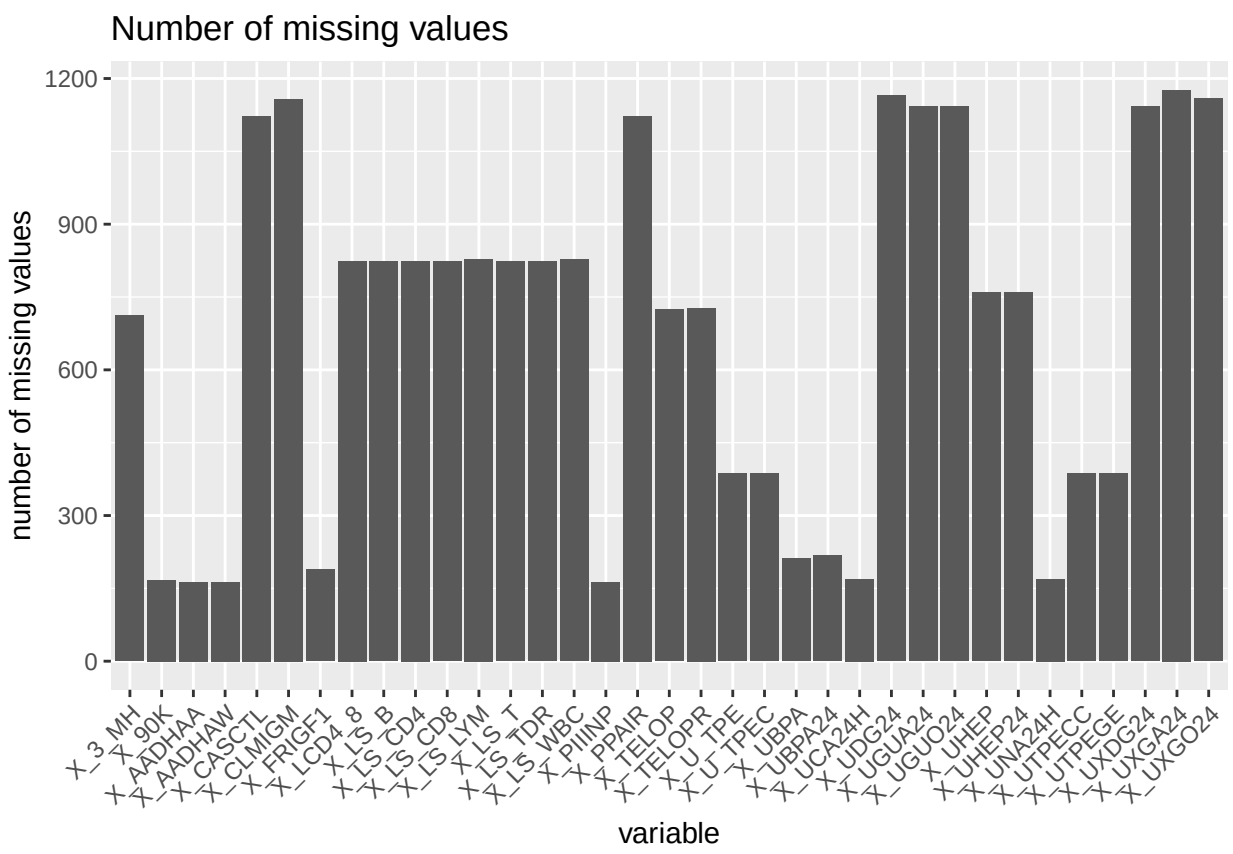
```
missing.values <- df %>%
    gather(key = "key", value = "val") %>%
    mutate(missing = is.na(val)) %>%
    group_by(key, missing) %>%
    dplyr::summarise(num.missing = n()) %>%
    filter(missing==TRUE) %>%
    select(-missing) %>%
    arrange(desc(num.missing))
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

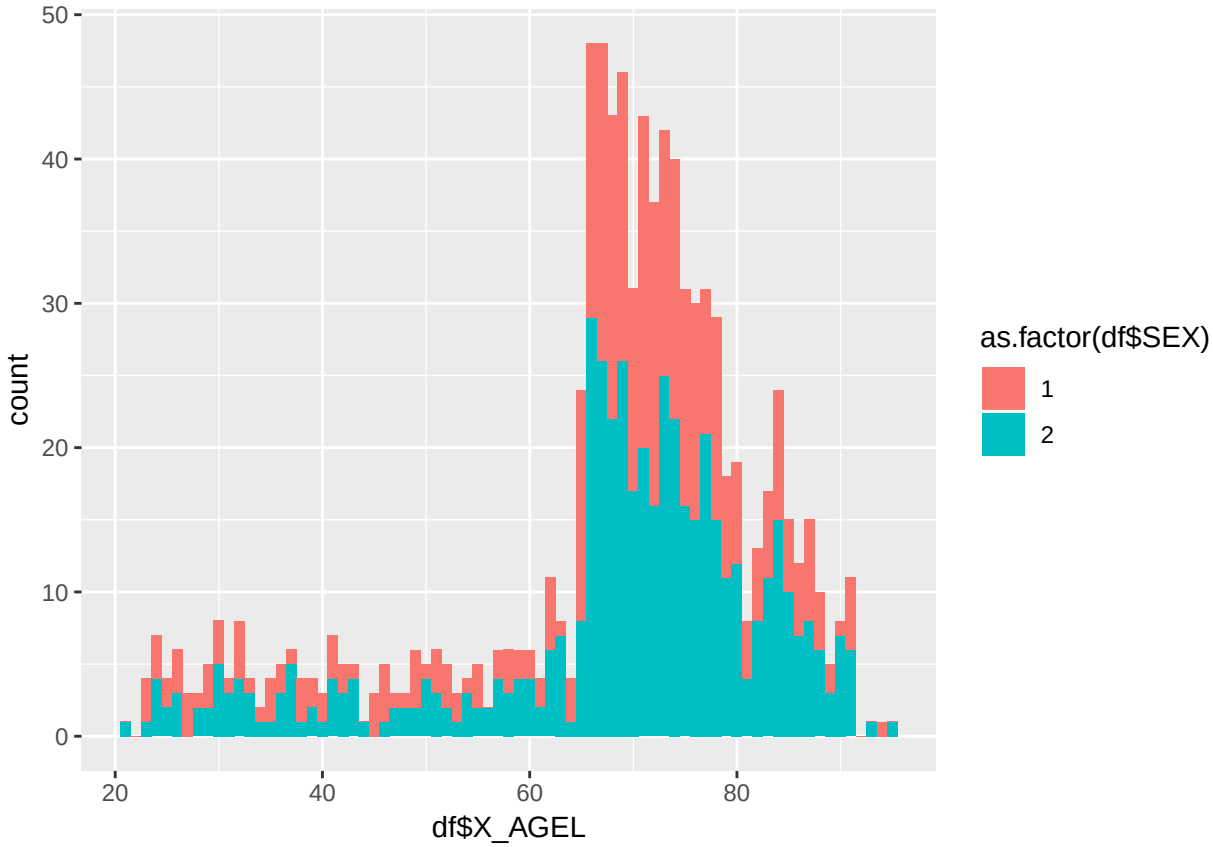
```
head(missing.values)
```

```
## # A tibble: 6 x 2
## # Groups:   key [6]
##   key      num.missing
##   <chr>          <int>
## 1 X_UXGA24        1176
## 2 X_UDG24         1165
## 3 X_UXGO24        1159
## 4 X_CLMIGM        1158
## 5 X_UGUO24        1143
## 6 X_UGUA24        1142
```

```
explore <- missing.values[1:35,]
explore %>%
  ggplot() +
    geom_bar(aes(x=key, y=num.missing), stat = 'identity') +
    labs(x='variable', y="number of missing values",
         title='Number of missing values') +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
vec <- missing.values[missing.values$num.missing < 96,1]
vec <- c('CODE98','SITE','SEX','X_AGEL',vec[['key']])
df <- select(analysis, vec)
df <- na.omit(df)
ggplot(df, aes(x=df$X_AGEL, fill=as.factor(df$SEX))) +
  geom_histogram(binwidth=1)
```




```
under_65 <- df[df$X_AGEL < 65,1:133]
sum(under_65$SEX == 1)
```

```
## [1] 93
```

```
sum(under_65$SEX == 2)
```

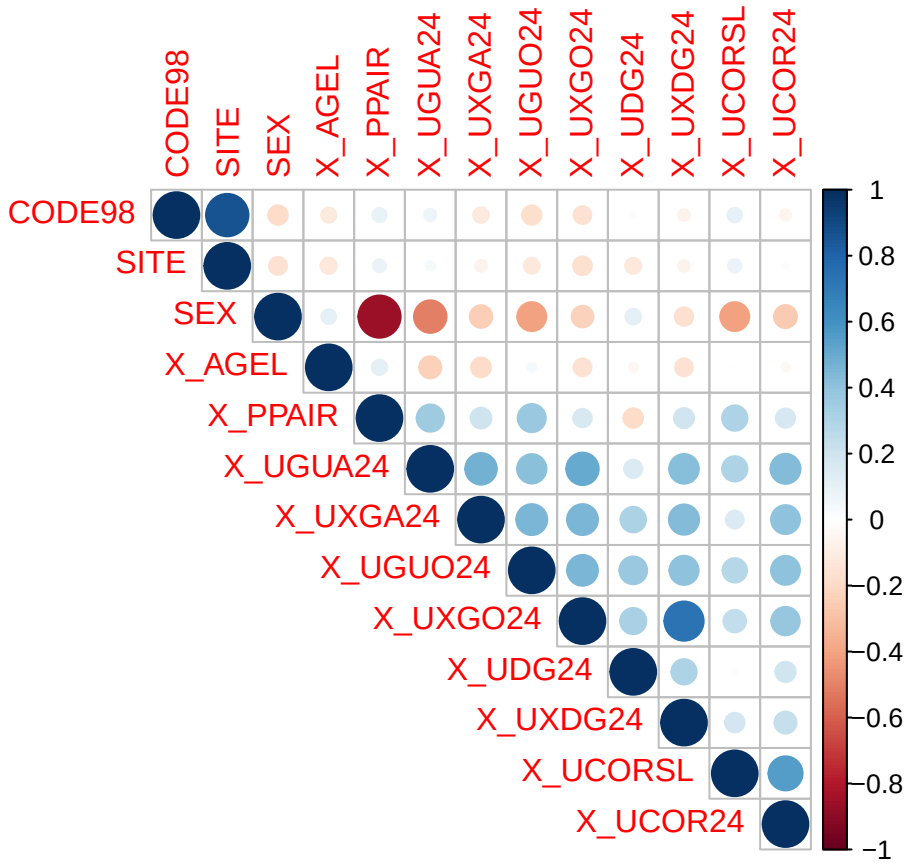
```
## [1] 111
```

```
under_65_cor <- as.data.frame(cor(under_65, method = 'spearman'))
under_65_cor <- as.data.frame(apply(under_65_cor,2,FUN= function(x){x*100}))
under_65_cor <- setDT(under_65_cor,keep.rownames = TRUE)
write.csv2(under_65_cor,'./under_65_cor.csv')
# Cor table phantasus and clustering (kmean and hierh)
```

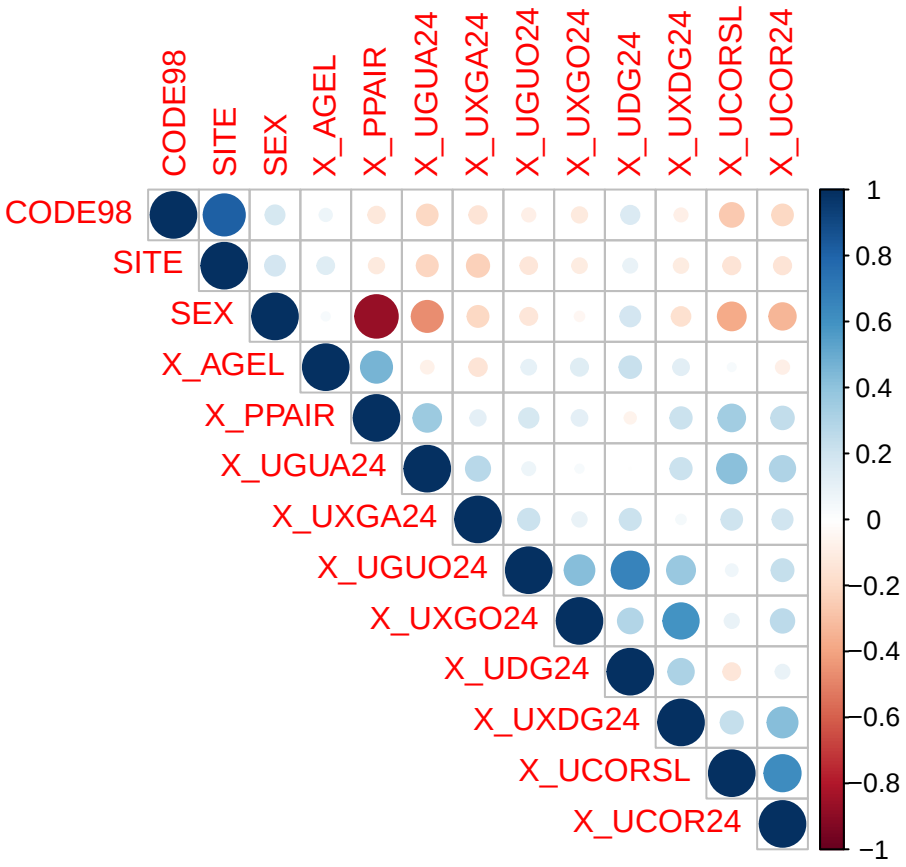
Oxidized guanine, guanosine and deoxyguanosine using Baseline 24-hour urine sample

X_CASCTL 1=case,0=control X_UGUO24 X_UXGA24

```
vec <- c('CODE98','SITE','SEX','X_AGEL','X_CASCTL','X_PPAIR','X_UGUA24','X_UXGA24','X_UGUO24','X_UXGO24
df <- select(analysis, vec)
df <- na.omit(df)
case <- df[df$X_CASCTL == 1,1:length(vec)]
control <- df[df$X_CASCTL == 0,1:length(vec)]
case <- case[,-which(colnames(case)=='X_CASCTL')]
corrplot::corrplot(cor(case,method = 'spearman'),method = c("circle"), type='upper')
```



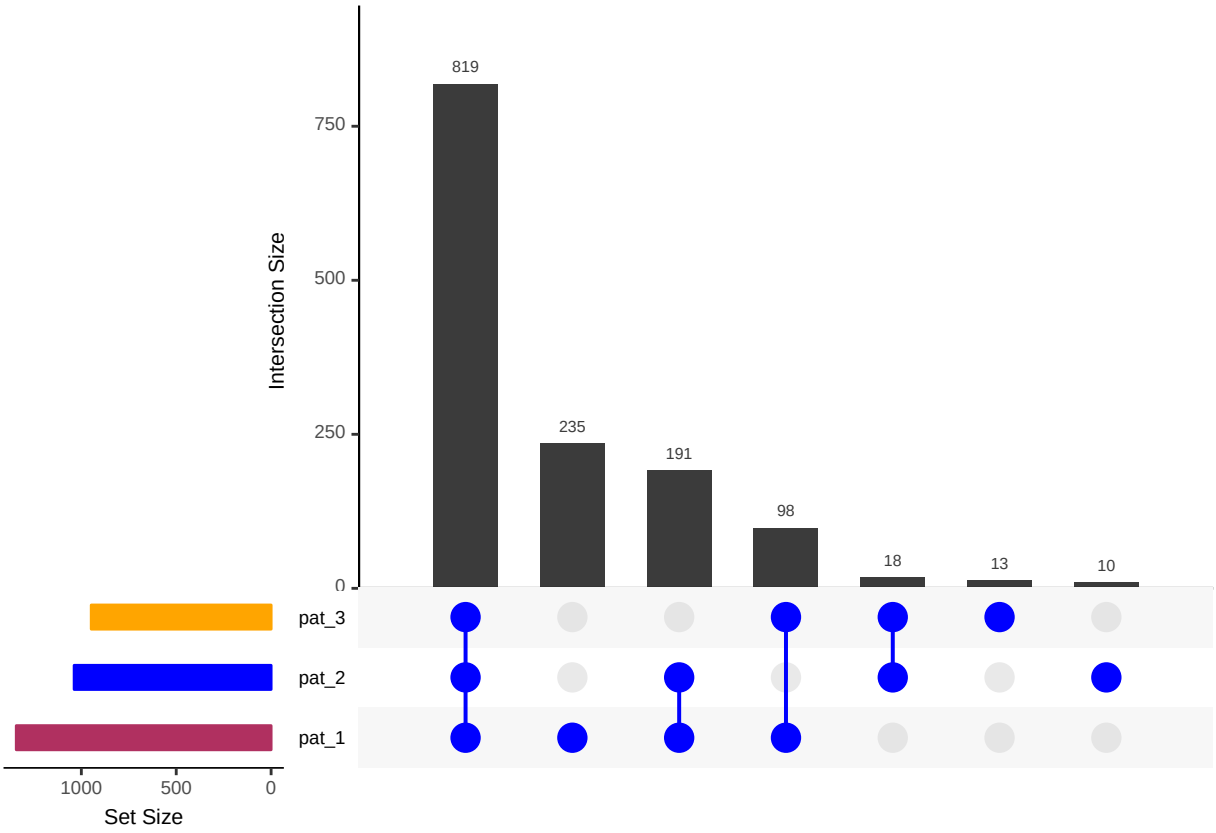

```
control <- control[,-which(colnames(control)=='X_CASCTL')]
corrplot::corrplot(cor(control,method = 'spearman'),method = c("circle"), type='upper')
```




```
combined <- 1:1806
common_patients <- data.frame(pat_1 = ifelse(combined %in% analysis$CODE98, 1, 0), pat_2 = ifelse(combi
head(common_patients)
```

```
##   pat_1 pat_2 pat_3
## 1     1     1     1
## 2     0     0     0
## 3     1     1     1
## 4     1     1     1
## 5     1     1     1
## 6     0     1     1
```

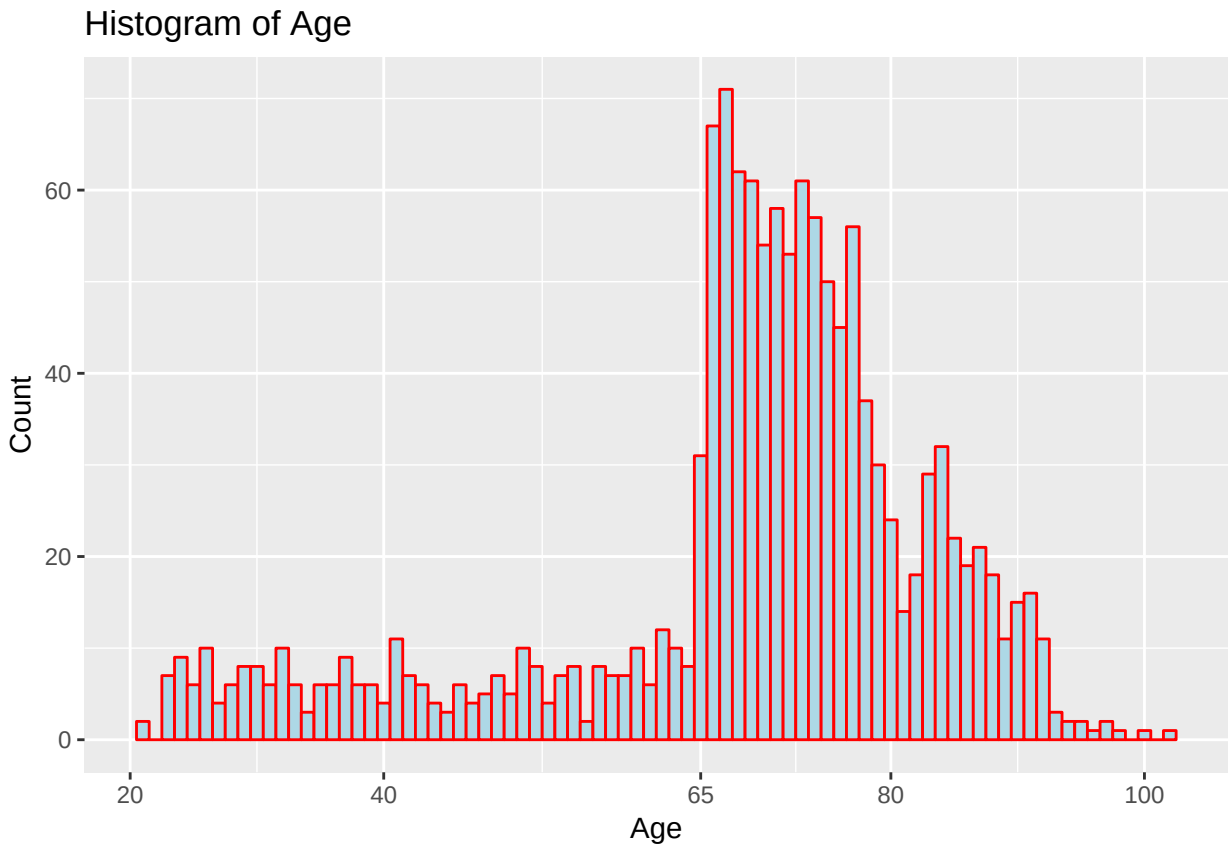
```
upset(common_patients,sets = c('pat_1','pat_2','pat_3'), order.by = "freq", empty.intersections = "on",
```




```
common_patients$id <- c(1:1806)
common_patients <- filter(common_patients, pat_1==1 & pat_2 ==1 & pat_3 == 1)
```

```
df <-  t(as.data.frame(lapply(colnames(analysis), FUN = function(x){
  gsub('X_', '',x)
})))
df1 <- t(as.data.frame(lapply(colnames(analysis_2), FUN = function(x){
  gsub('Y_', '',x)
})))
df2 <- t(as.data.frame(lapply(colnames(analysis_3), FUN = function(x){
  gsub('Z_', '',x)
})))
common_cols <- intersect(df,c(df1,df2))
```

```
qplot(analysis$X_AGEL,
      geom="histogram",
      binwidth = 1,
      main = "Histogram of Age",
      xlab = "Age",
      ylab = "Count",
      col=I("red"),
      fill=I("lightblue")) +
  scale_x_continuous(breaks=c(20,40,65,80,100))
```


Histogram of Age
Count
60
40
20
0
20
40
65
80
100
Age