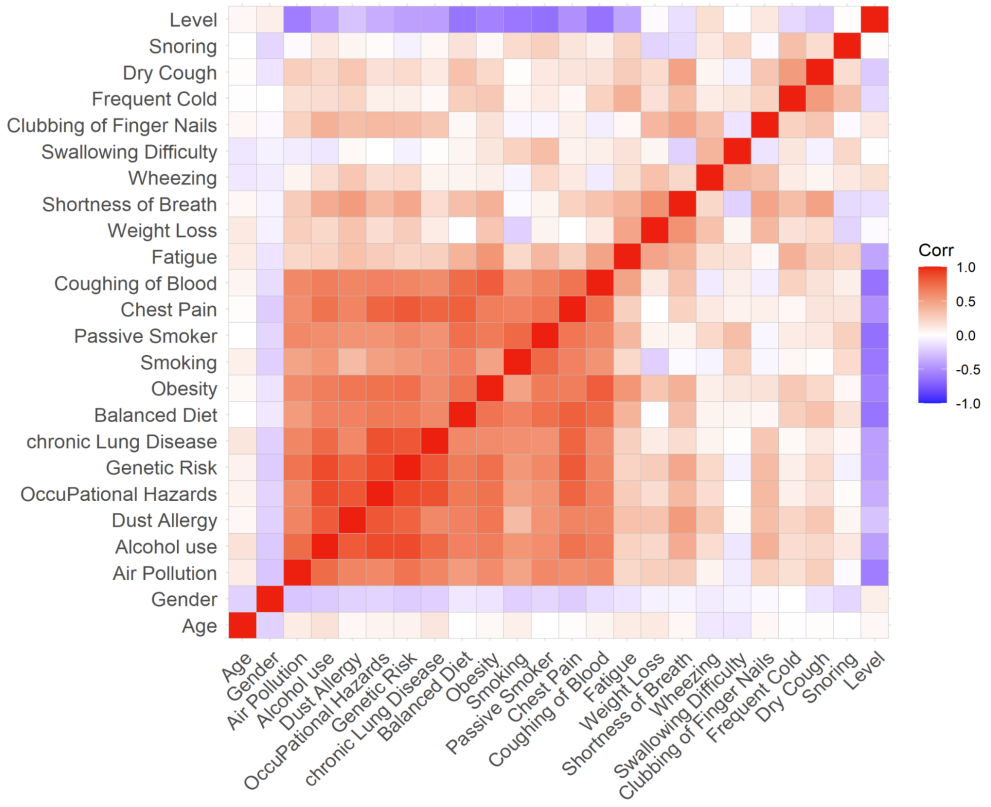
```
> library(readxl)
> DAAN888PCA <- read_excel("DAAN 888/DAAN888PCA.xlsx")
> View(DAAN888PCA)
> install.packages("corrr")
> library('corrr')
> install.packages("ggcorrplot")
> library(ggcorrplot)
> install.packages("FactoMineR")
> library("FactoMineR")
> str(DAAN888PCA)
tibble [1,000 × 24] (S3: tbl_df/tbl/data.frame)
 $ Age                     : num [1:1000] 33 17 35 37 46 35 52 28 35 46 ...
 $ Gender                  : num [1:1000] 1 1 1 1 1 1 2 2 2 1 ...
 $ Air Pollution           : num [1:1000] 2 3 4 7 6 4 2 3 4 2 ...
 $ Alcohol use             : num [1:1000] 4 1 5 7 8 5 4 1 5 3 ...
 $ Dust Allergy            : num [1:1000] 5 5 6 7 7 6 5 4 6 4 ...
 $ OccuPational Hazards    : num [1:1000] 4 3 5 7 7 5 4 3 5 2 ...
 $ Genetic Risk            : num [1:1000] 3 4 5 6 7 5 3 2 6 4 ...
 $ chronic Lung Disease    : num [1:1000] 2 2 4 7 6 4 2 3 5 3 ...
 $ Balanced Diet           : num [1:1000] 2 2 6 7 7 6 2 4 5 3 ...
 $ Obesity                 : num [1:1000] 4 2 7 7 7 7 4 3 5 3 ...
 $ Smoking                 : num [1:1000] 3 2 2 7 8 2 3 1 6 2 ...
 $ Passive Smoker          : num [1:1000] 2 4 3 7 7 3 2 4 6 3 ...
 $ Chest Pain              : num [1:1000] 2 2 4 7 7 4 2 3 6 4 ...
 $ Coughing of Blood       : num [1:1000] 4 3 8 8 9 8 4 1 5 4 ...
 $ Fatigue                 : num [1:1000] 3 1 8 4 3 8 3 3 1 1 ...
 $ Weight Loss             : num [1:1000] 4 3 7 2 2 7 4 2 4 2 ...
 $ Shortness of Breath     : num [1:1000] 2 7 9 3 4 9 2 2 3 4 ...
 $ Wheezing                : num [1:1000] 2 8 2 1 1 2 2 4 2 6 ...
 $ Swallowing Difficulty   : num [1:1000] 3 6 1 4 4 1 3 2 4 5 ...
 $ Clubbing of Finger Nails: num [1:1000] 1 2 4 5 2 4 1 2 6 4 ...
 $ Frequent Cold           : num [1:1000] 2 1 6 6 4 6 2 3 2 2 ...
 $ Dry Cough               : num [1:1000] 3 7 7 7 2 7 3 4 4 1 ...
 $ Snoring                 : num [1:1000] 4 2 2 5 3 2 4 3 1 5 ...
 $ Level                   : chr [1:1000] "Low" "Medium" "High" "High" ...
> colSums(is.na(DAAN888PCA))
                     Age                   Gender
                       0                        0
           Air Pollution              Alcohol use
                       0                        0
            Dust Allergy     OccuPational Hazards
                       0                        0
            Genetic Risk     chronic Lung Disease
                       0                        0
           Balanced Diet                  Obesity
                       0                        0
                 Smoking           Passive Smoker
                       0                        0
              Chest Pain        Coughing of Blood
```

```
                     0                      0
               Fatigue            Weight Loss
                     0                      0
  Shortness of Breath                Wheezing
                     0                      0
Swallowing Difficulty Clubbing of Finger Nails
                     0                      0
         Frequent Cold               Dry Cough
                     0                      0
               Snoring                  Level
                     0                      0
> numerical_data <- DAAN888PCA[,1:24]
> head(numerical_data)
# A tibble: 6 × 24
    Age Gender Air Pol…¹ Alcoh…² Dust …³ OccuP…⁴ Genet…⁵
  <dbl>  <dbl>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1    33      1         2       4       5       4       3
2    17      1         3       1       5       3       4
3    35      1         4       5       6       5       5
4    37      1         7       7       7       7       6
5    46      1         6       8       7       7       7
6    35      1         4       5       6       5       5
# … with 17 more variables:
#   `chronic Lung Disease` <dbl>,
#   `Balanced Diet` <dbl>, Obesity <dbl>,
#   Smoking <dbl>, `Passive Smoker` <dbl>,
#   `Chest Pain` <dbl>, `Coughing of Blood` <dbl>,
#   Fatigue <dbl>, `Weight Loss` <dbl>,
#   `Shortness of Breath` <dbl>, Wheezing <dbl>, …
# i Use `colnames()` to see all variable names

> DAAN888PCA$`Air Pollution` <- unclass(DAAN888PCA$`Air Pollution`)
> DAAN888PCA[sapply(DAAN888PCA, is.factor)] <-
data.matrix(DAAN888PCA[sapply(DAAN888PCA, is.factor)])
> numerical_data <- DAAN888PCA[,1:24]
> DAAN888PCA$Level <- as.numeric(as.factor(DAAN888PCA$Level))
> numerical_data <- DAAN888PCA[,1:24]
> data_normalized <- scale(numerical_data)
> head(data_normalized)
             Age     Gender Air Pollution Alcohol use
[1,] -0.34767419 -0.8194929   -0.90622555  -0.2148464
[2,] -1.68039750 -0.8194929   -0.41371167  -1.3596763
[3,] -0.18108378 -0.8194929    0.07880222   0.1667636
[4,] -0.01449337 -0.8194929    1.55634389   0.9299835
[5,]  0.73516350 -0.8194929    1.06383000   1.3115934
[6,] -0.18108378 -0.8194929    0.07880222   0.1667636
     Dust Allergy OccuPational Hazards Genetic Risk
[1,]   -0.0832983          -0.39851880   -0.7428307
[2,]   -0.0832983          -0.87294594   -0.2726847
```

```
[3,]    0.4215399           0.07590834    0.1974613
[4,]    0.9263780           1.02476262    0.6676073
[5,]    0.9263780           1.02476262    1.1377533
[6,]    0.4215399           0.07590834    0.1974613
     chronic Lung Disease Balanced Diet    Obesity
[1,]           -1.2875182    -1.1664563 -0.2188316
[2,]           -1.2875182    -1.1664563 -1.1600430
[3,]           -0.2055701     0.7066168  1.1929854
[4,]            1.4173520     1.1748851  1.1929854
[5,]            0.8763780     1.1748851  1.1929854
[6,]           -0.2055701     0.7066168  1.1929854
        Smoking Passive Smoker Chest Pain
[1,] -0.3798226    -0.94948547 -1.0692000
[2,] -0.7804794    -0.08435065 -1.0692000
[3,] -0.7804794    -0.51691806 -0.1920876
[4,]  1.2228045     1.21335160  1.1235810
[5,]  1.6234613     1.21335160  1.1235810
[6,] -0.7804794    -0.51691806 -0.1920876
     Coughing of Blood    Fatigue Weight Loss
[1,]        -0.3537942 -0.3813569  0.06571357
[2,]        -0.7656618 -1.2723778 -0.38748348
[3,]         1.2936760  1.8461953  1.42530473
[4,]         1.2936760  0.0641535 -0.84068053
[5,]         1.7055435 -0.3813569 -0.84068053
[6,]         1.2936760  1.8461953  1.42530473
     Shortness of Breath  Wheezing
[1,]          -0.9802691 -0.870259
[2,]           1.2078316  2.068151
[3,]           2.0830719 -0.870259
[4,]          -0.5426490 -1.359994
[5,]          -0.1050288 -1.359994
[6,]           2.0830719 -0.870259
     Swallowing Difficulty Clubbing of Finger Nails
[1,]            -0.3285789              -1.22401219
[2,]             0.9927841              -0.80526016
[3,]            -1.2094876               0.03224391
[4,]             0.1118754               0.45099594
[5,]             0.1118754              -0.80526016
[6,]            -1.2094876               0.03224391
     Frequent Cold  Dry Cough     Snoring       Level
[1,]    -0.8381985 -0.4183409  0.72829065  0.03953847
[2,]    -1.3839006  1.5433985 -0.62793030  1.23767385
[3,]     1.3446100  1.5433985 -0.62793030 -1.15859691
[4,]     1.3446100  1.5433985  1.40640112 -1.15859691
[5,]     0.2532058 -0.9087758  0.05018018 -1.15859691
[6,]     1.3446100  1.5433985 -0.62793030 -1.15859691
> corr_matrix <- cor(data_normalized)
```

```
> ggcorrplot(corr_matrix)
```




```
> data.pca <- princomp(corr_matrix)
> summary(data.pca)
Importance of components:
                          Comp.1    Comp.2     Comp.3
Standard deviation     1.2743968 0.5146625 0.39340328
Proportion of Variance 0.6851688 0.1117463 0.06529264
Cumulative Proportion  0.6851688 0.7969152 0.86220781
                           Comp.4     Comp.5     Comp.6
Standard deviation     0.30603987 0.27160325 0.24811255
Proportion of Variance 0.03951339 0.03112133 0.02597082
Cumulative Proportion  0.90172119 0.93284253 0.95881335
                           Comp.7      Comp.8
Standard deviation     0.15878140 0.131290266
Proportion of Variance 0.01063622 0.007271991
Cumulative Proportion  0.96944958 0.976721567
                           Comp.9     Comp.10
Standard deviation     0.113113316 0.109515439
Proportion of Variance 0.005397787 0.005059865
Cumulative Proportion  0.982119354 0.987179220
                          Comp.11     Comp.12
Standard deviation     0.100254065 0.081639412
Proportion of Variance 0.004240258 0.002811823
Cumulative Proportion  0.991419477 0.994231300
```

```
                          Comp.13    Comp.14
Standard deviation     0.064832814 0.05253261
Proportion of Variance 0.001773283 0.00116425
Cumulative Proportion  0.996004583 0.99716883
                            Comp.15      Comp.16
Standard deviation     0.0459403695 0.0456791441
Proportion of Variance 0.0008903838 0.0008802868
Cumulative Proportion  0.9980592169 0.9989395036
                            Comp.17      Comp.18
Standard deviation     0.0318911512 0.0238163998
Proportion of Variance 0.0004290705 0.0002392988
Cumulative Proportion  0.9993685741 0.9996078729
                           Comp.19      Comp.20
Standard deviation     0.019340250 0.0175228948
Proportion of Variance 0.000157802 0.0001295388
Cumulative Proportion  0.999765675 0.9998952137
                            Comp.21      Comp.22
Standard deviation     1.281070e-02 7.227000e-03
Proportion of Variance 6.923633e-05 2.203456e-05
Cumulative Proportion  9.999645e-01 9.999865e-01
                            Comp.23      Comp.24
Standard deviation     0.0056600507 8.793884e-09
Proportion of Variance 0.0000135154 3.262494e-17
Cumulative Proportion  1.0000000000 1.000000e+00
> data.pca$loadings[, 1:2]
                                Comp.1       Comp.2
Age                         0.001567193  0.025274667
Gender                     -0.108701976 -0.023027280
Air Pollution               0.264932937  0.086934765
Alcohol use                 0.278594941  0.123409058
Dust Allergy                0.236421182  0.178439439
OccuPational Hazards        0.279557964  0.095799250
Genetic Risk                0.290665569  0.131806029
chronic Lung Disease        0.276161438 -0.002227986
Balanced Diet               0.281550852 -0.084574271
Obesity                     0.268354807  0.035208735
Smoking                     0.254525191 -0.301694910
Passive Smoker              0.263447729 -0.232500007
Chest Pain                  0.296234579 -0.098412630
Coughing of Blood           0.285784714 -0.075442424
Fatigue                     0.104350680  0.067228883
Weight Loss                 0.006139002  0.399530494
Shortness of Breath         0.083709936  0.449799187
Wheezing                   -0.029356788  0.118332438
Swallowing Difficulty      -0.024239526 -0.277846030
Clubbing of Finger Nails    0.011191744  0.394180701
Frequent Cold               0.010929113  0.110950368
Dry Cough                   0.047107262  0.230229172
Snoring                    -0.005843746 -0.236231152
```

```
Level                        -0.262288698   0.120990732
```

```
> biplot <- princomp(data_normalized)
> par(mar = c(1, 1, 1, 1))
> plot(1:24)
> biplot(biplot)
```