Suicide

René Mayer

02.06.2015

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(foreign)
setwd('/home/rene/Dokumente/DirkRitter/')

s <- read.spss(file = "10 Suizidakten/Vgl. Suizid-Kontroll (incl. Daten).sav",
   use.value.labels = TRUE,  # SPSS variables with value labels into R factors with levels
   max.value.labels = Inf,   # can be any real number
      to.data.frame = TRUE,  # we want it to be an data.frame
       use.missings = TRUE  # recode SPSS set missings-code to NA<
)
```

```
## Warning in read.spss(file = "10 Suizidakten/Vgl. Suizid-Kontroll (incl.
## Daten).sav", : 10 Suizidakten/Vgl. Suizid-Kontroll (incl. Daten).sav:
## Unrecognized record type 7, subtype 17 encountered in system file

## Warning in read.spss(file = "10 Suizidakten/Vgl. Suizid-Kontroll (incl.
## Daten).sav", : 10 Suizidakten/Vgl. Suizid-Kontroll (incl. Daten).sav:
## Unrecognized record type 7, subtype 18 encountered in system file

## Warning in read.spss(file = "10 Suizidakten/Vgl. Suizid-Kontroll (incl.
## Daten).sav", : 10 Suizidakten/Vgl. Suizid-Kontroll (incl. Daten).sav:
## Unrecognized record type 7, subtype 24 encountered in system file

## re-encoding from UTF-8
```

```
names(s)
```

```
##  [1] "Nr"                "Gruppe"             "JVA"
##  [4] "Code"              "Alter_Inh"          "Alter_E"
##  [7] "Geschlecht"        "National"           "Sprache"
## [10] "Alt_Inhaft"        "Alt_End"            "Familie"
## [13] "Kind"              "Kinder"             "Bezieh"
## [16] "Besuche"           "Religion"           "Schule"
## [19] "Ausbild"           "Arbeitstätigkeit"   "kindheit"
## [22] "Mon_Haftbeg"       "Haftart"            "Verleg"
## [25] "Anz_Verl"          "Dau_Inha"           "Gesamtst"
## [28] "Indexdelikt"       "Hafterfa"           "Anzahl_Vo"
## [31] "Jurist_Dat"        "Verh_Haft"          "Arb_Haft"
## [34] "Unterbri"          "Inhaft_Dat"         "SPB"
## [37] "Erg_SPB"           "Erg_zsfs"           "Dat_SPB"
## [40] "Su_erhdat"         "SPB_TT"             "Absch_br"
```

```
## [43] "Grund_Sui"           "Beson_Sui"           "angekünd"
## [46] "Suizidvorgeschichte" "Suizidversuche"      "Anz_SV"
## [49] "SV_Wann"             "Sui_1_Wo"            "Suizidmeth_1"
## [52] "Suizidmeth_2"        "Tt_WT"               "Tt_Ft"
## [55] "Tt_FT_WE"            "TT_Jz"               "Tz_Az"
## [58] "PP_Auff"             "GA_vorh"             "Ergeb_GA"
## [61] "Psych_Erk"           "Anzahl_Erkr"         "Art_Erkrank"
## [64] "Art_Erkrank2"        "Mißbr_Such"          "Behandlung"
## [67] "Raucher"             "VISCI"               "SUR_1"
## [70] "SUR_2"               "SUR_3"               "SUR_4"
## [73] "SUR_5"               "SUR_6"
```

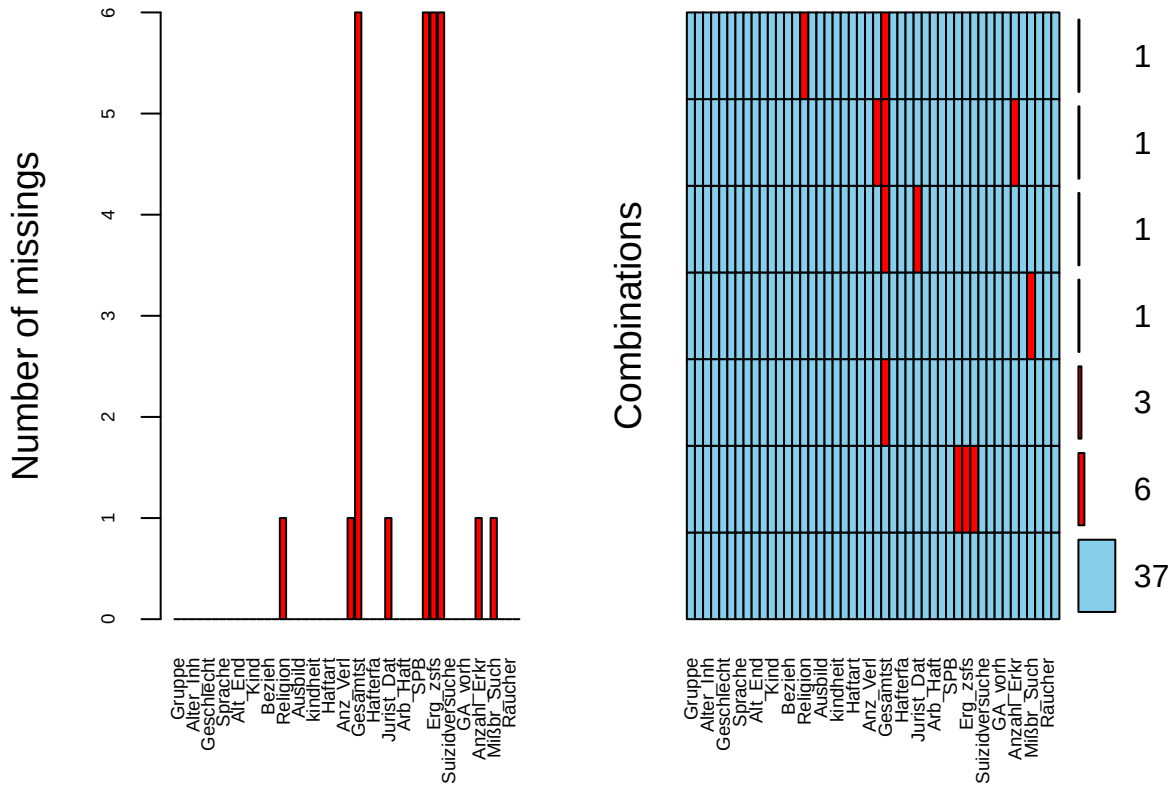
```
#xqplot(s, ask=T)

library(mice)
```

```
## Loading required package: Rcpp
## Loading required package: lattice
## mice 2.22 2014-06-10
```

```
library(VIM)
```

```
## Loading required package: colorspace
## Loading required package: grid
## Loading required package: data.table
## VIM is ready to use.
##  Since version 4.0.0 the GUI is in its own package VIMGUI.
##
##           Please use the package to use the new (and old) GUI.
##
## Suggestions and bug-reports can be submitted at: https://github.com/alexkowa/VIM/issues
##
## Attaching package: 'VIM'
##
## The following object is masked from 'package:datasets':
##
##     sleep
```

```
#md.pattern(s)

sg=s[,  !names(s) %in% c("Inhaft_Dat", "Dat_SPB", "Su_erhdat", "angekünd", "Absch_br", "Grund_Sui", "Bes
"Anz_SV", "SV_Wann", "Sui_1_Wo", "Suizidmeth_1", "Suizidmeth_2" , "Tt_WT", "Tt_Ft", "Tt_FT_WE", "TT_Jz"
"SUR_1","SUR_2", "SUR_3", "SUR_4", "SUR_5", "SUR_6" , "Su_erhdat" , "Dat_SPB" , "Inhaft_Dat", "Nr","Cod


# Nr 51 fehlt - evtl. einlesefehler
aggr(sg[-51,], prop = F, numbers = T, cex.axis=.6)
```




```
sg=sg[-51,]
# skalennivau: kontinuierlich und faktoriel, nicht ordinal - das wird ein Fehler sin, aber der ist nich
( o=sapply( sg, class  ) )
```

```
##           Gruppe              JVA        Alter_Inh          Alter_E
##         "factor"         "factor"         "numeric"        "numeric"
##       Geschlecht         National           Sprache       Alt_Inhaft
##         "factor"         "factor"         "factor"         "numeric"
##          Alt_End          Familie             Kind           Kinder
##        "numeric"         "factor"         "factor"         "numeric"
##           Bezieh          Besuche         Religion           Schule
##         "factor"         "factor"         "factor"         "factor"
##          Ausbild Arbeitstätigkeit          kindheit      Mon_Haftbeg
##         "factor"         "factor"         "factor"         "numeric"
##          Haftart           Verleg         Anz_Verl          Dau_Inha
##         "factor"        "numeric"        "numeric"        "numeric"
##         Gesamtst      Indexdelikt         Hafterfa        Anzahl_Vo
##        "numeric"         "factor"         "factor"         "numeric"
##       Jurist_Dat        Verh_Haft         Arb_Haft          Unterbri
##        "numeric"         "factor"         "factor"         "factor"
##              SPB          Erg_SPB         Erg_zsfs           SPB_TT
##         "factor"         "factor"         "factor"         "numeric"
##    Suizidversuche          PP_Auff          GA_vorh        Psych_Erk
##         "factor"         "factor"         "factor"         "factor"
##       Anzahl_Erkr      Art_Erkrank        Mißbr_Such       Behandlung
##         "numeric"         "factor"         "factor"         "factor"
##          Raucher             VISCI
##         "factor"         "factor"
```

```
# o[ order( o ) ]


d=md.pattern(sg)
d[8,]
```

```
##           Gruppe              JVA        Alter_Inh          Alter_E
##                0                0                0                0
##       Geschlecht         National          Sprache       Alt_Inhaft
##                0                0                0                0
##          Alt_End          Familie             Kind           Kinder
##                0                0                0                0
##           Bezieh          Besuche           Schule          Ausbild
##                0                0                0                0
## Arbeitstätigkeit         kindheit      Mon_Haftbeg          Haftart
##                0                0                0                0
##           Verleg         Anz_Verl      Indexdelikt         Hafterfa
##                0                0                0                0
##        Anzahl_Vo        Verh_Haft         Arb_Haft         Unterbri
##                0                0                0                0
##              SPB   Suizidversuche          PP_Auff          GA_vorh
##                0                0                0                0
##        Psych_Erk      Art_Erkrank       Behandlung          Raucher
##                0                0                0                0
##            VISCI         Religion         Dau_Inha       Jurist_Dat
##                0                1                1                1
##      Anzahl_Erkr       Mißbr_Such         Gesamtst          Erg_SPB
##                1                1                6                6
##         Erg_zsfs           SPB_TT
##                6                6               29
```

```
sg$Mißbr_Such = factor(as.character(sg$Mißbr_Such))

# wir haben aufgrund der unterschiedlichen skalennivaus mixed data ...
# Lösungen:   1. verwende polychorische cov [wieso nicht!, aber bedarf dummy coding aller factoren]
#                 2. FactoMineR kann mit mixed data umghen [vermtl. am leichtesten]
#                 3. cannonical correspondence analysis mit cca( . ~ gruppe, ...) [vermtl. am besten]

# setzte 2. im folgenden um
# FactoMineR
# install.packages('FactoMineR', dep=T)
library(FactoMineR)



# schneide alle Fragebogendaten weg, damit wir mit einem reien soziodemographischen Datensatz rechnenn
osg = sg[ , c(
 "Gruppe",
 "JVA",  "Geschlecht", "Sprache", "National",
 "Alter_Inh", "Alter_E", "Alt_Inhaft", "Alt_End",
 "Familie", "Kind",  "Bezieh", "Besuche",
 "Religion", "Schule", "Ausbild",
 "Haftart", "Arb_Haft", "Unterbri",  "Verh_Haft",
```

```
 "Anz_Verl", "Dau_Inha",
 "Indexdelikt", "Hafterfa",
 "SPB",  "Erg_SPB", "Erg_zsfs", "Suizidversuche", "Arbeitstätigkeit", "kindheit", "PP_Auff", "GA_vorh",
 "Anzahl_Erkr") ]

# > names(sg)[25]
# [1] "Gesamtst" - was soll das sein?

# wir mappen alle variablen auf variablengruppen und zwar so:
d=
data.frame(group= c(1, 4, 4, 4, 3,4,  2, 2, 14, 1 ),
          type= c("n", "n", "c", "n",  "n", "n",  "c",  "n",  "n",  "c"),
    name.group= c("suicide","demo", "time", "relations", "education", "haft", "delinq.c" , "delinq", "ri

# ergiebt folgende zuordnung
data.frame( d[rep(seq_len(nrow(d)), d$group),], sapply(osg, class), names(sapply(osg, class)))
```

```
##      group type name.group sapply.osg..class. names.sapply.osg..class..
## 1        1    n    suicide             factor                    Gruppe
## 2        4    n       demo             factor                       JVA
## 2.1      4    n       demo             factor                 Geschlecht
## 2.2      4    n       demo             factor                    Sprache
## 2.3      4    n       demo             factor                   National
## 3        4    c       time            numeric                  Alter_Inh
## 3.1      4    c       time            numeric                    Alter_E
## 3.2      4    c       time            numeric                 Alt_Inhaft
## 3.3      4    c       time            numeric                    Alt_End
## 4        4    n  relations             factor                    Familie
## 4.1      4    n  relations             factor                       Kind
## 4.2      4    n  relations             factor                     Bezieh
## 4.3      4    n  relations             factor                    Besuche
## 5        3    n  education             factor                   Religion
## 5.1      3    n  education             factor                     Schule
## 5.2      3    n  education             factor                    Ausbild
## 6        4    n       haft             factor                    Haftart
## 6.1      4    n       haft             factor                   Arb_Haft
## 6.2      4    n       haft             factor                   Unterbri
## 6.3      4    n       haft             factor                  Verh_Haft
## 7        2    c   delinq.c            numeric                    Anz_Verl
## 7.1      2    c   delinq.c            numeric                    Dau_Inha
## 8        2    n     delinq             factor                 Indexdelikt
## 8.1      2    n     delinq             factor                    Hafterfa
## 9       14    n       risk             factor                         SPB
## 9.1     14    n       risk             factor                     Erg_SPB
## 9.2     14    n       risk             factor                    Erg_zsfs
## 9.3     14    n       risk             factor              Suizidversuche
## 9.4     14    n       risk             factor           Arbeitstätigkeit
## 9.5     14    n       risk             factor                    kindheit
## 9.6     14    n       risk             factor                     PP_Auff
## 9.7     14    n       risk             factor                     GA_vorh
## 9.8     14    n       risk             factor                   Psych_Erk
## 9.9     14    n       risk             factor                 Art_Erkrank
## 9.10    14    n       risk             factor                  Mißbr_Such
```

```
## 9.11    14    n       risk            factor               Behandlung
## 9.12    14    n       risk            factor                  Raucher
## 9.13    14    n       risk            factor                    VISCI
## 10       1    c disorder.c           numeric               Anzahl_Erkr
```

```
res <- MFA(osg, group= c(1, 4, 4, 4, 3,4,  2, 2, 14, 1 ),
          type = c("n", "n", "c", "n",  "n", "n",  "c",  "n",  "n",  "c"),
    name.group = c("suicide","demo", "time", "relations", "education", "haft", "delinq.c" , "delinq", "
         graph = FALSE # bloß nicht, denn die plots sind nicht so schön
    )

# man sieht deulich, dass die Gruppenzugehörigkeitdurch die ersten beiden Dimensionen abgebildet wird
summary(res)
```

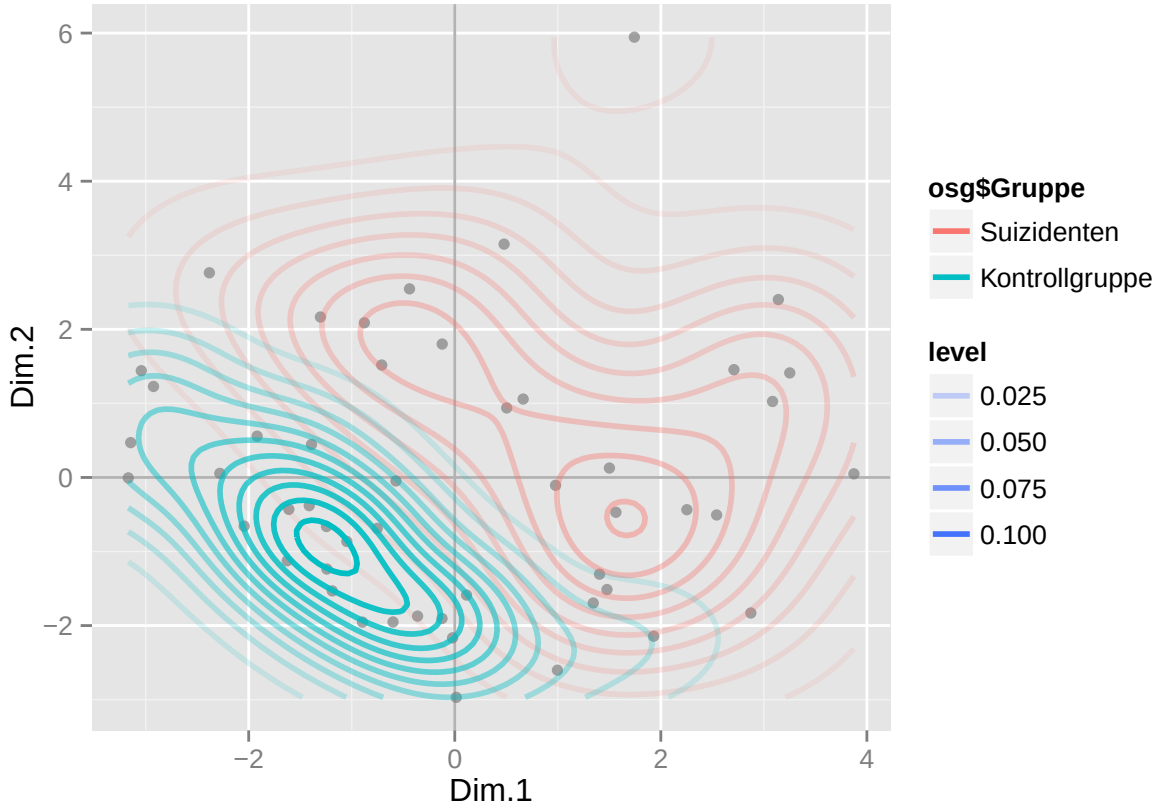
```
##
## Call:
## rmarkdown::render("dirk_suiceGroup_mfa_results.Rmd", encoding = "UTF-8")
##
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6
## Variance               3.381   3.038   2.676   2.343   2.169   1.937
## % of var.              7.244   6.510   5.735   5.021   4.648   4.151
## Cumulative % of var.   7.244  13.754  19.489  24.511  29.158  33.309
##                        Dim.7   Dim.8   Dim.9  Dim.10  Dim.11  Dim.12
## Variance               1.899   1.761   1.646   1.603   1.569   1.438
## % of var.              4.070   3.773   3.527   3.436   3.363   3.081
## Cumulative % of var.  37.380  41.153  44.679  48.115  51.478  54.560
##                       Dim.13  Dim.14  Dim.15  Dim.16  Dim.17  Dim.18
## Variance               1.393   1.286   1.173   1.149   1.094   1.046
## % of var.              2.984   2.757   2.513   2.462   2.345   2.242
## Cumulative % of var.  57.544  60.301  62.814  65.276  67.621  69.863
##                       Dim.19  Dim.20  Dim.21  Dim.22  Dim.23  Dim.24
## Variance               0.995   0.937   0.877   0.859   0.816   0.748
## % of var.              2.133   2.008   1.880   1.841   1.750   1.603
## Cumulative % of var.  71.996  74.004  75.883  77.725  79.474  81.077
##                       Dim.25  Dim.26  Dim.27  Dim.28  Dim.29  Dim.30
## Variance               0.703   0.682   0.615   0.584   0.564   0.532
## % of var.              1.507   1.462   1.319   1.251   1.209   1.140
## Cumulative % of var.  82.584  84.046  85.364  86.615  87.825  88.964
##                       Dim.31  Dim.32  Dim.33  Dim.34  Dim.35  Dim.36
## Variance               0.493   0.471   0.450   0.435   0.397   0.354
## % of var.              1.057   1.010   0.965   0.933   0.851   0.758
## Cumulative % of var.  90.021  91.031  91.997  92.929  93.781  94.539
##                       Dim.37  Dim.38  Dim.39  Dim.40  Dim.41  Dim.42
## Variance               0.346   0.337   0.277   0.265   0.217   0.209
## % of var.              0.742   0.723   0.593   0.568   0.464   0.448
## Cumulative % of var.  95.280  96.003  96.596  97.164  97.628  98.077
##                       Dim.43  Dim.44  Dim.45  Dim.46  Dim.47  Dim.48
## Variance               0.202   0.181   0.150   0.114   0.106   0.081
## % of var.              0.434   0.389   0.321   0.244   0.228   0.173
## Cumulative % of var.  98.511  98.899  99.221  99.465  99.692  99.866
##                       Dim.49
## Variance               0.063
```

```
## % of var.              0.134
## Cumulative % of var. 100.000
##
## Groups
##                     Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3
## suicide          |  0.398 11.761  0.158 |  0.268  8.807  0.072 |  0.005
## demo             |  0.169  5.003  0.007 |  0.362 11.928  0.030 |  0.435
## time             |  0.098  2.893  0.010 |  0.128  4.218  0.016 |  0.012
## relations        |  0.441 13.058  0.048 |  0.570 18.762  0.080 |  0.390
## education        |  0.158  4.662  0.006 |  0.379 12.484  0.032 |  0.661
## haft             |  0.503 14.878  0.064 |  0.557 18.343  0.078 |  0.121
## delinq.c         |  0.145  4.285  0.021 |  0.023  0.760  0.001 |  0.344
## delinq           |  0.332  9.810  0.022 |  0.315 10.372  0.020 |  0.311
## risk             |  0.733 21.683  0.143 |  0.392 12.907  0.041 |  0.261
## disorder.c       |  0.405 11.968  0.164 |  0.043  1.419  0.002 |  0.137
##                     ctr   cos2
## suicide           0.183  0.000 |
## demo             16.243  0.043 |
## time              0.446  0.000 |
## relations        14.574  0.037 |
## education        24.688  0.097 |
## haft              4.527  0.004 |
## delinq.c         12.863  0.119 |
## delinq           11.606  0.020 |
## risk              9.751  0.018 |
## disorder.c        5.120  0.019 |
##
## Individuals (the 10 first)
##                     Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3
## 1                |  1.344  1.069  0.063 | -1.695  1.892  0.100 | -0.745
## 2                |  2.874  4.886  0.238 | -1.830  2.206  0.097 | -1.030
## 3                | -1.304  1.006  0.026 |  2.166  3.090  0.073 | -0.773
## 4                |  2.542  3.823  0.113 | -0.507  0.169  0.004 | -0.935
## 5                |  3.251  6.255  0.112 |  1.411  1.312  0.021 |  6.292
## 6                | -0.709  0.297  0.010 |  1.518  1.517  0.046 | -0.098
## 7                |  1.564  1.448  0.077 | -0.471  0.146  0.007 | -0.675
## 8                | -0.878  0.456  0.017 |  2.088  2.872  0.095 | -0.113
## 9                | -0.122  0.009  0.000 |  1.801  2.136  0.057 |  0.440
## 10               |  3.141  5.835  0.139 |  2.402  3.798  0.082 |  1.633
##                     ctr   cos2
## 1                 0.415  0.019 |
## 2                 0.793  0.031 |
## 3                 0.447  0.009 |
## 4                 0.653  0.015 |
## 5                29.588  0.420 |
## 6                 0.007  0.000 |
## 7                 0.340  0.014 |
## 8                 0.009  0.000 |
## 9                 0.145  0.003 |
## 10                1.993  0.038 |
##
## Continuous variables
##                      Dim.1     ctr    cos2     Dim.2     ctr    cos2
## Alter_Inh        |  -3.590   0.800   0.106 |   4.388   1.331   0.159 |
```

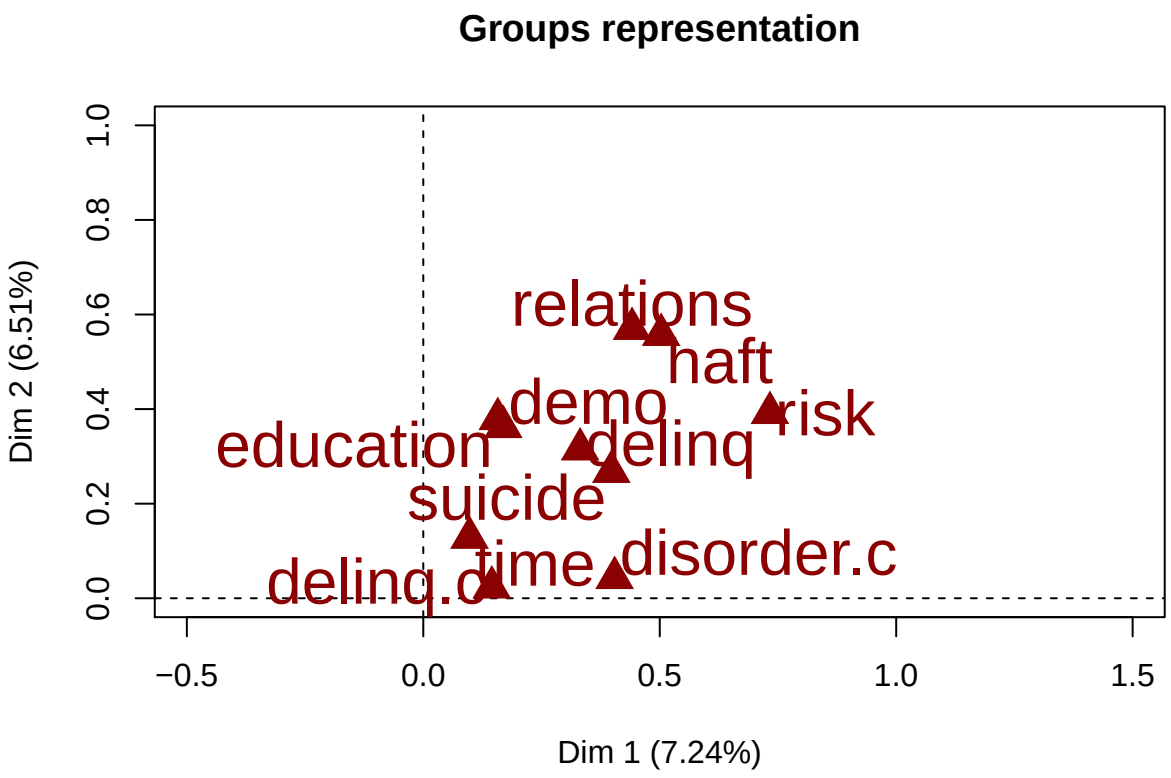
```
## Alter_E          |  -3.005   0.560   0.074 |   4.197   1.217   0.144 |
## Alt_Inhaft       |  -4.363   1.182   0.144 |   2.652   0.486   0.053 |
## Alt_End          |  -2.377   0.351   0.041 |   4.140   1.184   0.123 |
## Anz_Verl         |   0.163   0.000   0.068 |  -0.135   0.000   0.046 |
## Dau_Inha         | 170.497   4.285   0.145 | -68.075   0.760   0.023 |
## Anzahl_Erkr      |   0.545  11.968   0.405 |  -0.178   1.419   0.043 |
##                    Dim.3     ctr    cos2
## Alter_Inh         -1.012   0.080   0.008 |
## Alter_E           -0.396   0.012   0.001 |
## Alt_Inhaft        -0.822   0.053   0.005 |
## Alt_End           -1.956   0.300   0.027 |
## Anz_Verl           0.207   0.000   0.109 |
## Dau_Inha         262.853  12.863   0.344 |
## Anzahl_Erkr       -0.317   5.120   0.137 |
##
## Categories (the 10 first)
##                      Dim.1    ctr   cos2 v.test    Dim.2    ctr   cos2
## Suizidenten      |  1.159  5.880  0.524  4.414 |  0.901  4.403  0.317
## Kontrollgruppe   | -1.159  5.880  0.524 -4.414 | -0.901  4.403  0.317
## Bautzen          |  0.121  0.006  0.002  0.170 | -1.014  0.563  0.113
## Chemnitz         |  1.328  0.260  0.068  1.032 | -1.047  0.200  0.042
## Dresden          | -0.491  0.178  0.043 -0.935 |  0.611  0.341  0.067
## Leipzig          | -0.441  0.057  0.013 -0.495 |  0.582  0.124  0.023
## Regis-Breitingen | -0.035  0.000  0.000 -0.039 | -1.123  0.460  0.080
## Torgau           |  0.306  0.048  0.012  0.471 | -0.479  0.146  0.028
## Waldheim         |  0.065  0.002  0.000  0.091 | -0.732  0.293  0.048
## Zeithain         | -0.869  0.278  0.056 -1.102 | -0.382  0.066  0.011
##                    v.test    Dim.3    ctr   cos2 v.test
## Suizidenten         3.621 | -0.115  0.092  0.005 -0.490 |
## Kontrollgruppe     -3.621 |  0.115  0.092  0.005  0.490 |
## Bautzen            -1.505 | -0.451  0.143  0.022 -0.712 |
## Chemnitz           -0.859 | -0.103  0.003  0.000 -0.090 |
## Dresden             1.228 |  0.240  0.068  0.010  0.514 |
## Leipzig             0.690 | -0.991  0.461  0.066 -1.250 |
## Regis-Breitingen   -1.330 | -0.135  0.009  0.001 -0.170 |
## Torgau             -0.776 | -0.701  0.404  0.060 -1.210 |
## Waldheim           -1.086 |  1.857  2.428  0.311  2.934 |
## Zeithain           -0.511 |  0.785  0.362  0.045  1.119 |
```

```r
# eigenwerte zerfallen negativ exponentiell für 30% der varianz in den Daten brauchen wir 5 Dimensionen
# aber wir wollen ja nicht die varianez dieses Datensatzes aufklären, sondern die Gruppenzugehörigkeit!
# barplot(res$eig[,1],main="Eigenvalues",names.arg=1:nrow(res$eig))

# xy-coordinaten, d.h. Ladungen für jede Dimension an den eigenwerten relativiert
coord = data.frame(res$ind$coord)

library(ggplot2)

ggplot(data = coord, aes(x = Dim.1, y = Dim.2, group=osg$Gruppe)) +
  geom_hline(yintercept = 0, colour = "gray70") +
  geom_vline(xintercept = 0, colour = "gray70") +
  geom_point(colour = "gray50", alpha = 0.7) +
  stat_density2d(geom="density2d", aes(color = osg$Gruppe,alpha=..level..),
                 size=1,
```

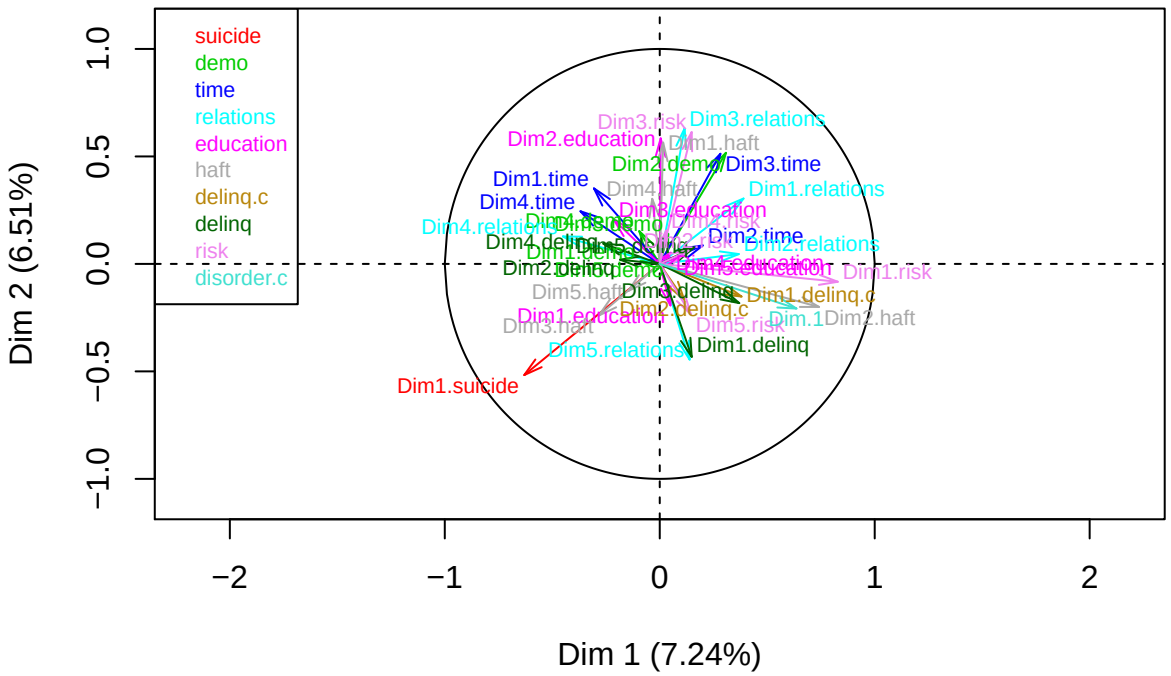
```
                contour=TRUE)
```



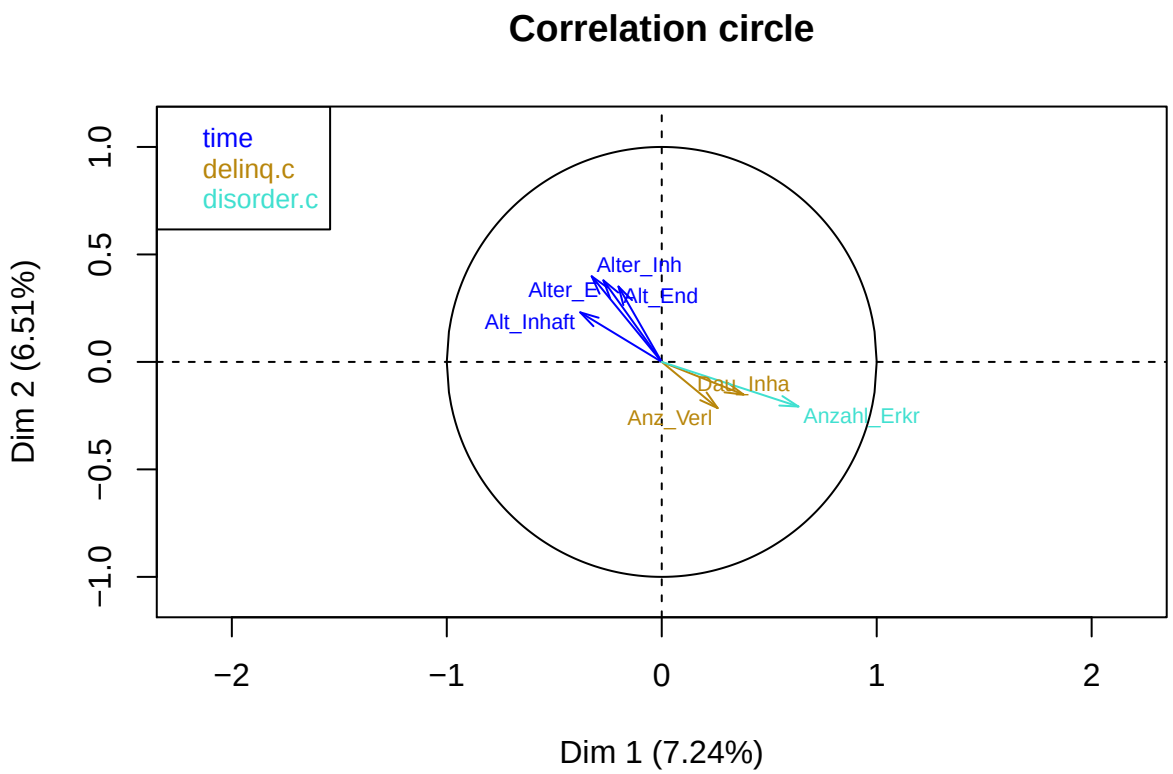

```
# zunächst die variablen-Gruppen
plot(res,choix="group",cex=2)
```




```
# welche skalen,, d.h.Ausprägungne (Dim.N) sind auf den Achsen gemappt?
# schutzfaktoren 'relations' vs. risikofaktoren bspw. 'Haft: 3', delinquenz dimension 1 scheint jedoch
plot(res,choix="axes",habillage="group",palette=palette(),cex=.7)
```



```r
# die kontinuierlichen variablen sind unabhängig von Suizid.
 plot(res,choix="var",habillage="group",palette=palette(),cex=.7)
```



Wenn wir die Achsen um 45° drehen können die Gruppen gut getrennt werden. oder gleich eine logisitsche ? evtl eine CCA, um denexplained variance zu bekommen? ada4::discrimin