

# Studies of Independent Variables

05.04.2022

## Bibliotheken laden, Hilfsfunktion

```
#library(ggplot2)      # moderne plots
library(GGally)

debug <- T             # debug printout
debug <- F             # kein debug printout
Log <- function(string) {
  if(debug){print(string)}
}
```

## For all MY Groups :

- Resistenzen.Rmd generated Resistenzen[Schicht].csv, read it in
- plot variables and calculate correlations

```
for( Schicht in c("GT8000","LE8000","U") ) {      # Greater Than 8000 / Less than or Equal to 8000 / Un-stratified
  FileIn <- paste( "Resistenzen",Schicht,".csv" , sep="" )
  Resistenzen <- read.csv(FileIn)

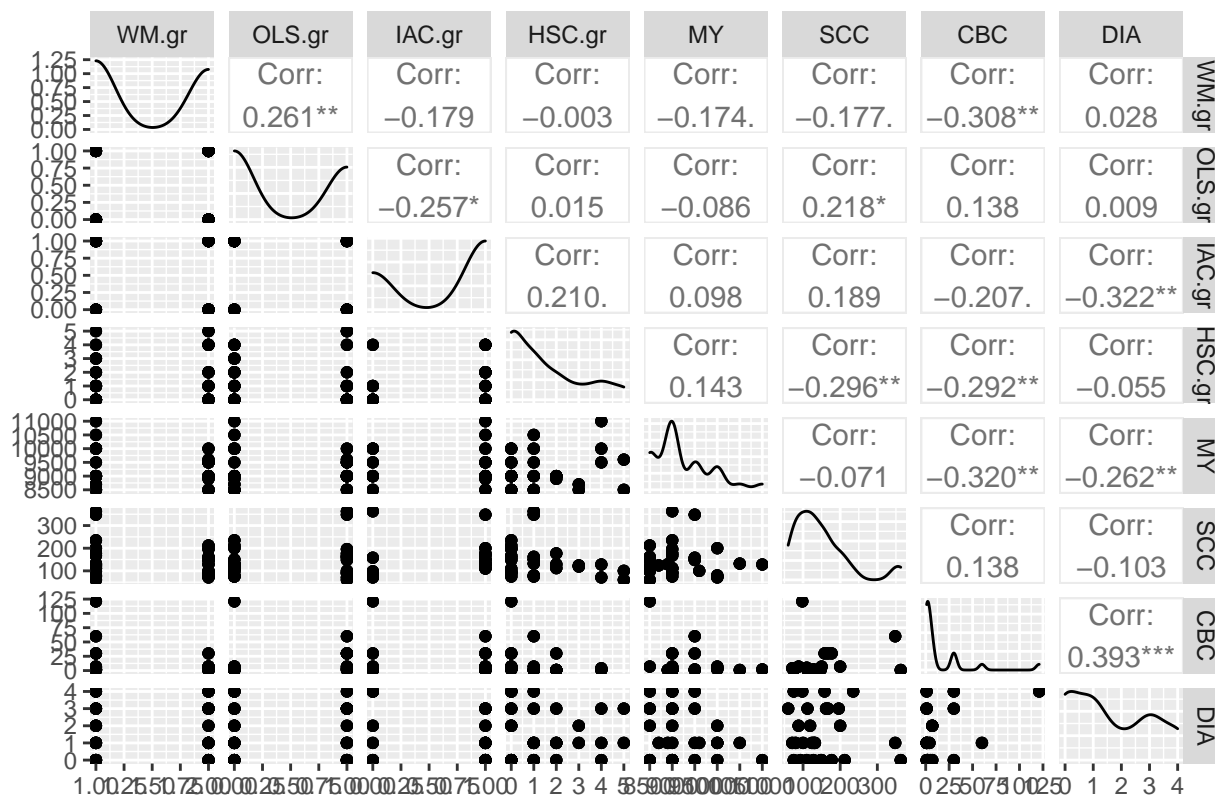
  # csv schreiben fügt vorne Index-Spalte an; diese entfernen :
  Resistenzen[,1] <- NULL

  if(debug){View(Resistenzen)}

  df <- data.frame(WM.gr = Resistenzen$WM.group,      # unabhängige Variablen extrahieren
                   OLS.gr = Resistenzen$OLS.group,  # Titel kürzen, sonst Platzprobleme ...
                   IAC.gr = Resistenzen$IAC.group,
                   HSC.gr = Resistenzen$HSC.group,
                   MY      = Resistenzen$MY,
                   SCC     = Resistenzen$SCC,
                   CBC     = Resistenzen$CBC,
                   DIA     = Resistenzen$DIA)

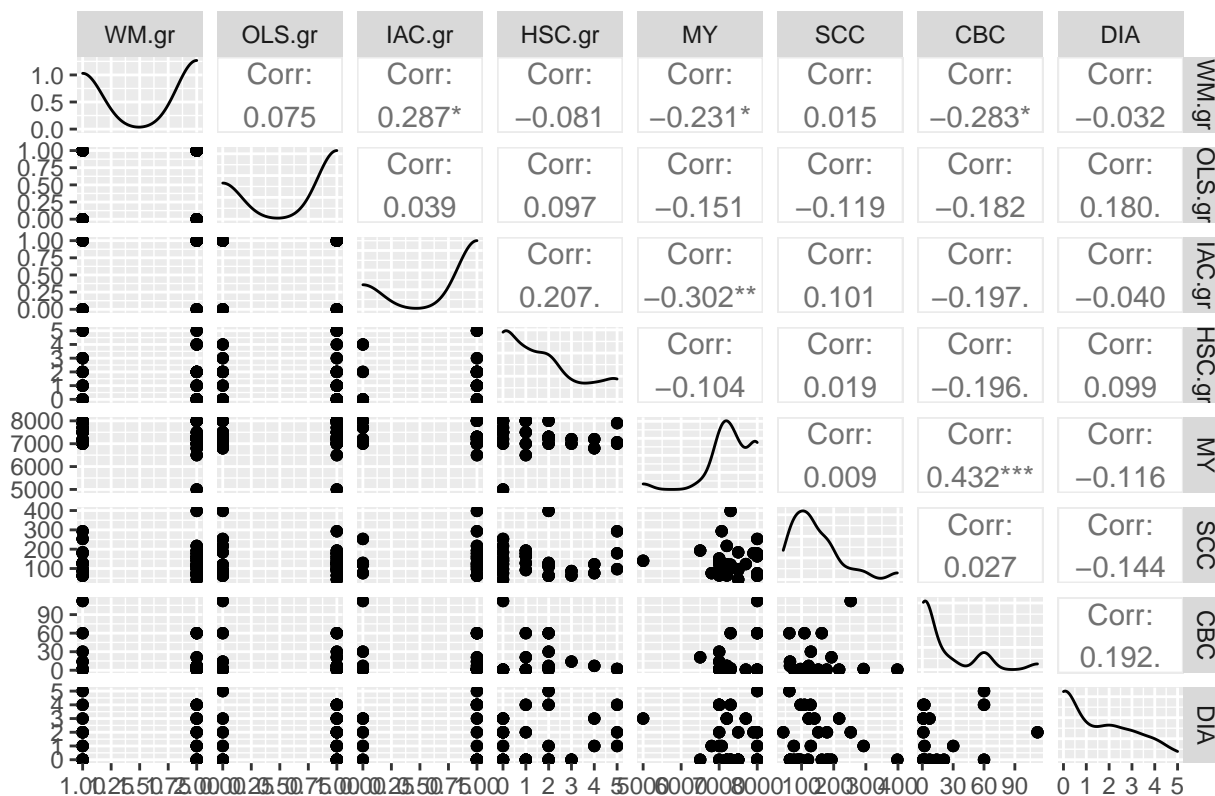
  #View(df)
  print(ggpairs(df, title = paste("group:",Schicht)))
  print("")
}
```

group: GT8000

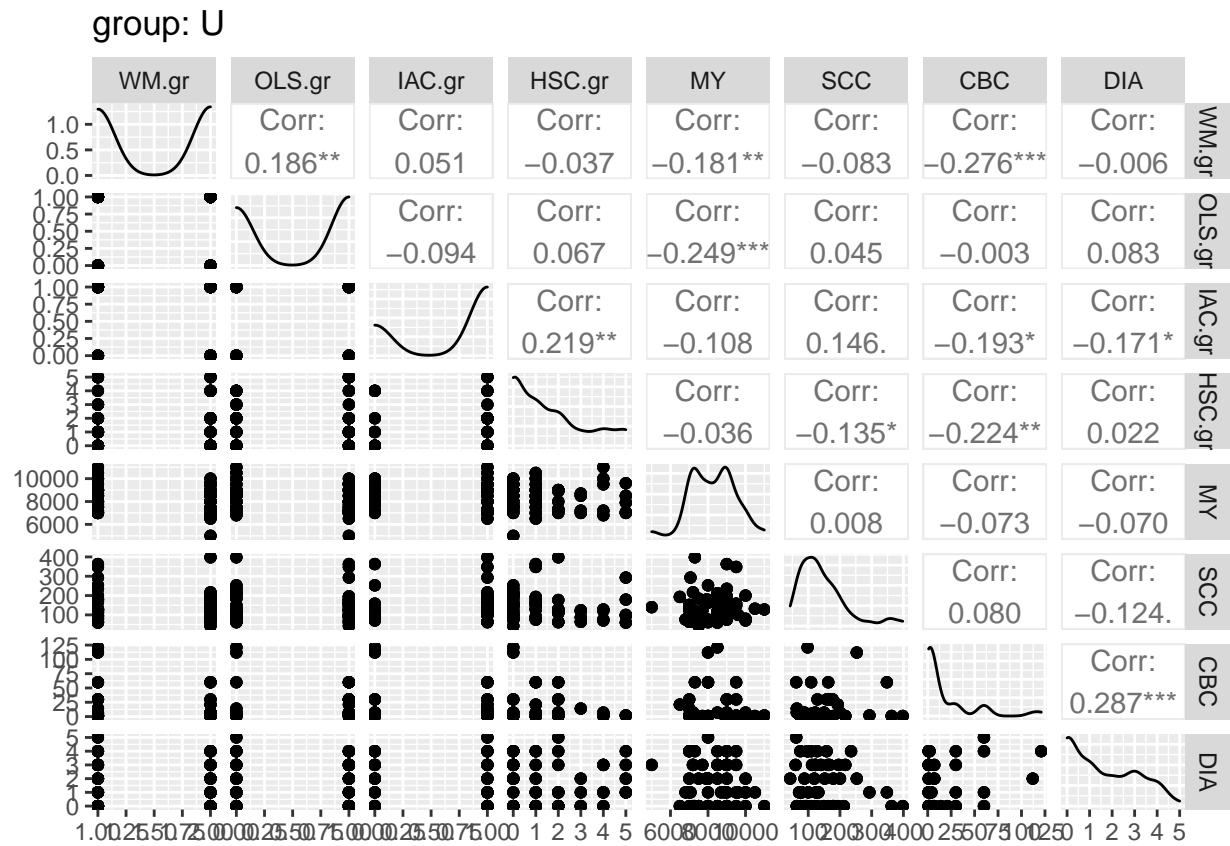


## [1] ""

group: LE8000



## [1] ""



## [1] ""

### Linear Depence?

As it implies multicollinearity, the logistic regression would be unreliable in its presence.

The maximum correlation magnitude amounts to

- 28.7% in the unstratified analysis
- 43.3% in the stratified analysis

These are not yet a strong linear correlations, they don't hint to collinearity problems in the multivariate logistic regression.

### Outliers?

Im wesentlichen sind nur die plots ohne diskrete Variablen gut zu interpretieren (für die anderen könnte man bessere Grafiken machen, hätte dann aber immer noch das Problem der beliebigen Kodierung).

In Histogrammen und Streuplots sehe ich einen Ausreisser mit MY = 5000, das ist Farm 32. Ist sie als problematisch bekannt?