Clustering

Zach Kontor

2025-03-18

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.1
```

```
##
## 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
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 4.4.2
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(FactoMineR)
```

```
## Warning: package 'FactoMineR' was built under R version 4.4.3
```

```
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.4.3
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```r
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.4.2
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
setwd("C:/Users/Zach/Documents/GitHub/Stress-Analysis")
data <- read.csv("StressData.csv")
data2 <- subset(data,select = -c(Occupation_Sales.Representative,Quality.of.Sleep_4,Occupation_Manager))
DataScaled <- scale(data)
for(col in names(data2)) {
  # Get unique values in the column (removing any NAs)
  unique_vals <- unique(data2[[col]])
  unique_vals <- unique_vals[!is.na(unique_vals)]

  # Check if the column is numeric and has exactly two unique values: 0 and 1 to convert to factor type
  if(is.numeric(data2[[col]]) && length(unique_vals) == 2 && all(unique_vals %in% c(0, 1))) {
    data2[[col]] <- as.factor(data2[[col]])
    cat("Converted", col, "to factor.\n")
  }
}
```

```
## Converted Gender to factor.
## Converted High_Blood_Pressure to factor.
## Converted Occupation_Accountant to factor.
## Converted Occupation_Doctor to factor.
## Converted Occupation_Engineer to factor.
## Converted Occupation_Lawyer to factor.
## Converted Occupation_Nurse to factor.
## Converted Occupation_Salesperson to factor.
## Converted Occupation_Scientist to factor.
## Converted Occupation_Software.Engineer to factor.
## Converted Occupation_Teacher to factor.
## Converted BMI.Category_Normal to factor.
## Converted BMI.Category_Underweight to factor.
## Converted BMI.Category_Obese to factor.
## Converted BMI.Category_Overweight to factor.
## Converted Sleep.Disorder_Insomnia to factor.
## Converted Sleep.Disorder_None to factor.
## Converted Sleep.Disorder_Sleep.Apnea to factor.
## Converted Quality.of.Sleep_5 to factor.
## Converted Quality.of.Sleep_6 to factor.
## Converted Quality.of.Sleep_7 to factor.
## Converted Quality.of.Sleep_8 to factor.
## Converted Quality.of.Sleep_9 to factor.
## Converted Stress.Level_3 to factor.
## Converted Stress.Level_4 to factor.
```

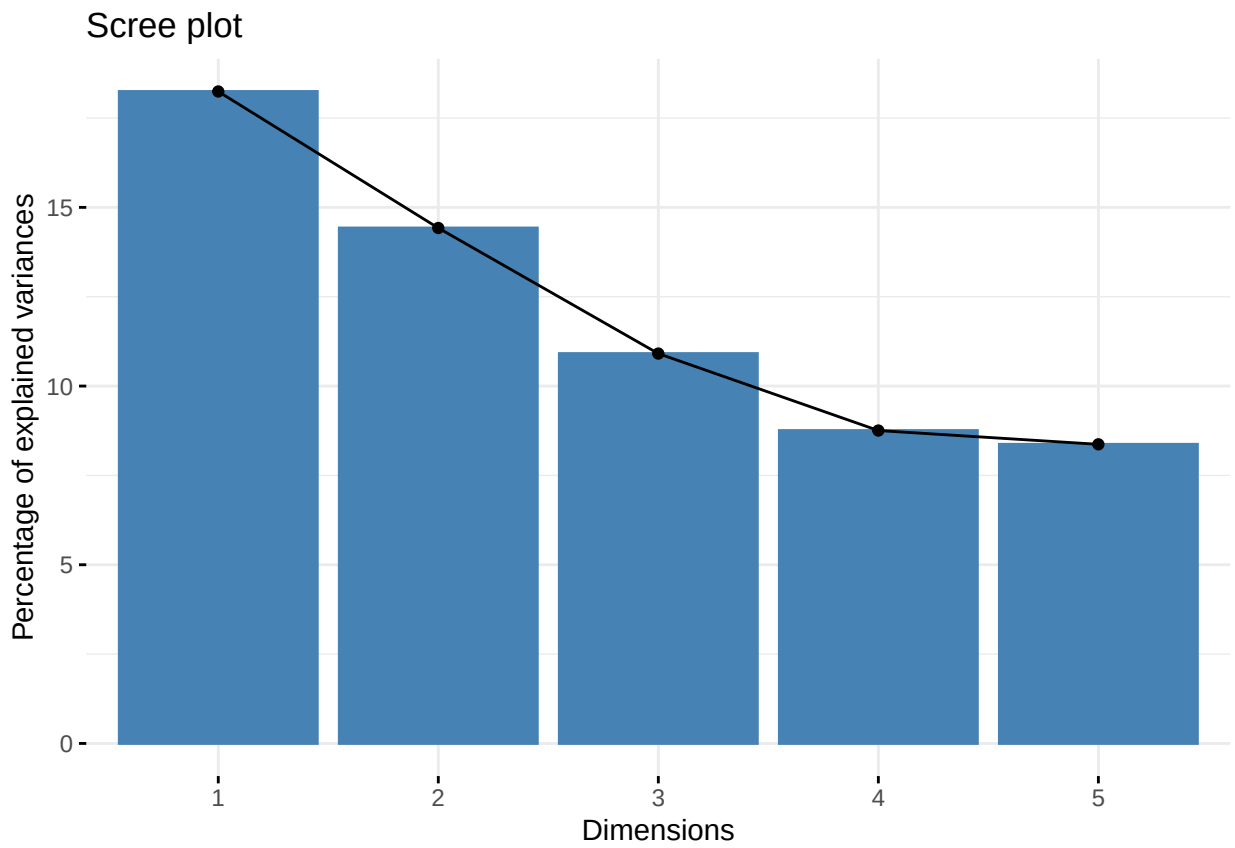
```
## Converted Stress.Level_5 to factor.
## Converted Stress.Level_6 to factor.
## Converted Stress.Level_7 to factor.
## Converted Stress.Level_8 to factor.
```

Categorical Variables Contribution

```
res.famd <- FAMD(data2, graph=FALSE)
eig.Vals <- get_eigenvalue(res.famd)
eig.Vals
```

```
##       eigenvalue variance.percent cumulative.variance.percent
## Dim.1   6.202424        18.242424                    18.24242
## Dim.2   4.903755        14.422808                    32.66523
## Dim.3   3.709183        10.909363                    43.57459
## Dim.4   2.976912         8.755625                    52.33022
## Dim.5   2.845847         8.370137                    60.70036
```

```
# After examining the cumulative variance we can see that five dimensions can explain 60% of the varian
ScreePlot1 <- fviz_screeplot(res.famd)
ScreePlot1
```



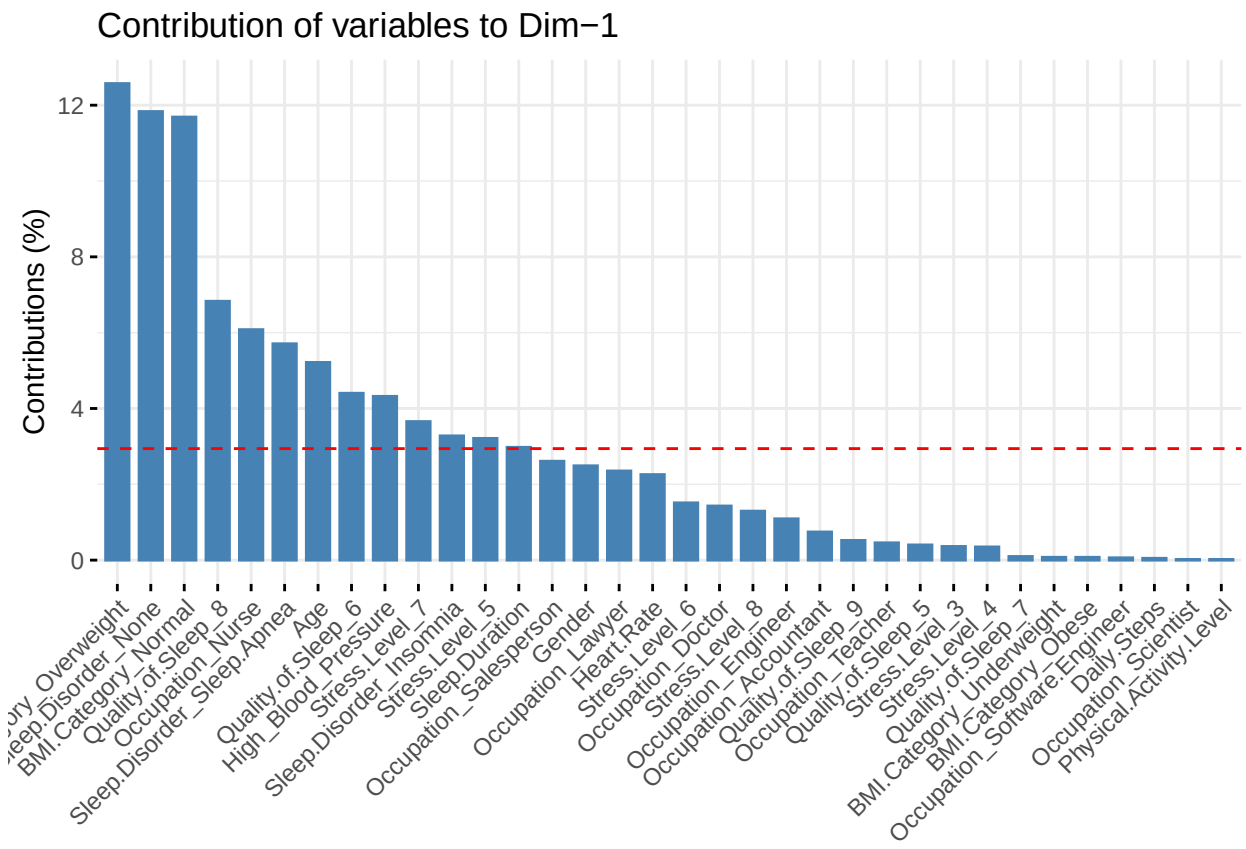
```r
## Getting varis from FAMD
var <- get_famd_var(res.famd)
head(var$coord)
```

```
##                                Dim.1        Dim.2        Dim.3       Dim.4
## Age                      0.323434442 0.5349693577 9.104483e-05 0.025162915
## Sleep.Duration           0.184542654 0.5923670011 9.407801e-03 0.002535919
## Physical.Activity.Level 0.000963602 0.0247633645 4.465158e-01 0.004875047
## Heart.Rate               0.139958326 0.2569620169 1.993052e-01 0.005380234
## Daily.Steps              0.002964138 0.0003757451 4.854828e-01 0.026570820
## Gender                   0.154324356 0.3429964707 2.093219e-02 0.180495489
##                                Dim.5
## Age                      0.002591964
## Sleep.Duration           0.017192486
## Physical.Activity.Level 0.219244082
## Heart.Rate               0.016112958
## Daily.Steps              0.185891621
## Gender                   0.073142295
```

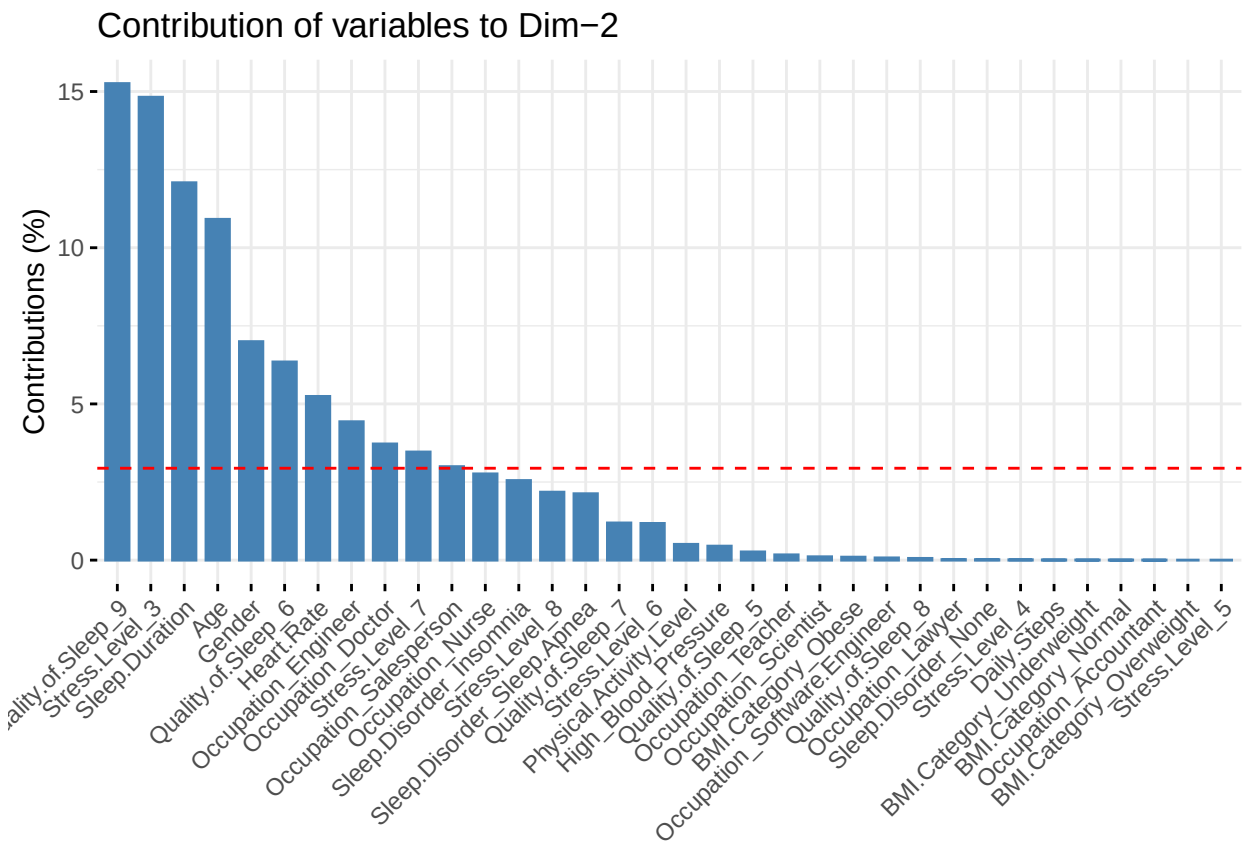
```r
#Vari Contributions to each dimension
head(var$contrib)
```

```
##                                Dim.1        Dim.2        Dim.3     Dim.4
## Age                      5.21464568 10.909382550  0.002454579 0.8452689
## Sleep.Duration           2.97533110 12.079866132  0.253635360 0.0851862
## Physical.Activity.Level 0.01553589  0.504987832 12.038116704 0.1637618
## Heart.Rate               2.25651009  5.240107500  5.373291754 0.1807320
## Daily.Steps              0.04778999  0.007662396 13.088670443 0.8925630
## Gender                   2.48812969  6.994568302  0.564334163 6.0631777
##                                Dim.5
## Age                      0.09107883
## Sleep.Duration           0.60412553
## Physical.Activity.Level 7.70400191
## Heart.Rate               0.56619205
## Daily.Steps              6.53203218
## Gender                   2.57014178
```

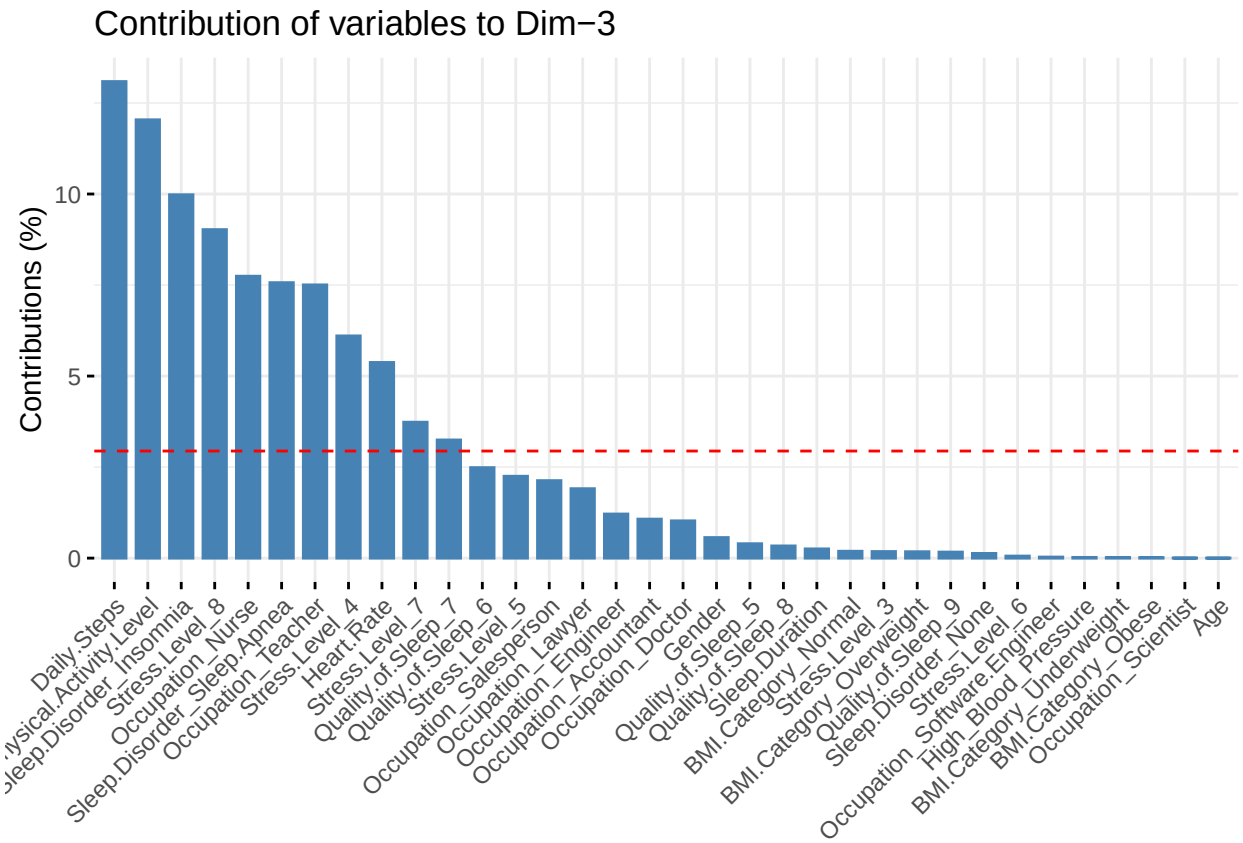
```r
#Plots
FigVAR <- fviz_famd_var(res.famd, repel = TRUE)
#contribution plots in accordance w dimensions
FigContrib1 <- fviz_contrib(res.famd, "var", axes = 1)
FigContrib2 <- fviz_contrib(res.famd, "var", axes = 2)
FigContrib3 <- fviz_contrib(res.famd, "var", axes = 3)
FigContrib4 <- fviz_contrib(res.famd, "var", axes = 4)
FigContrib5 <- fviz_contrib(res.famd, "var", axes = 5)
#Plots for categorical variable contribution to dimensions.
FigContrib1
```



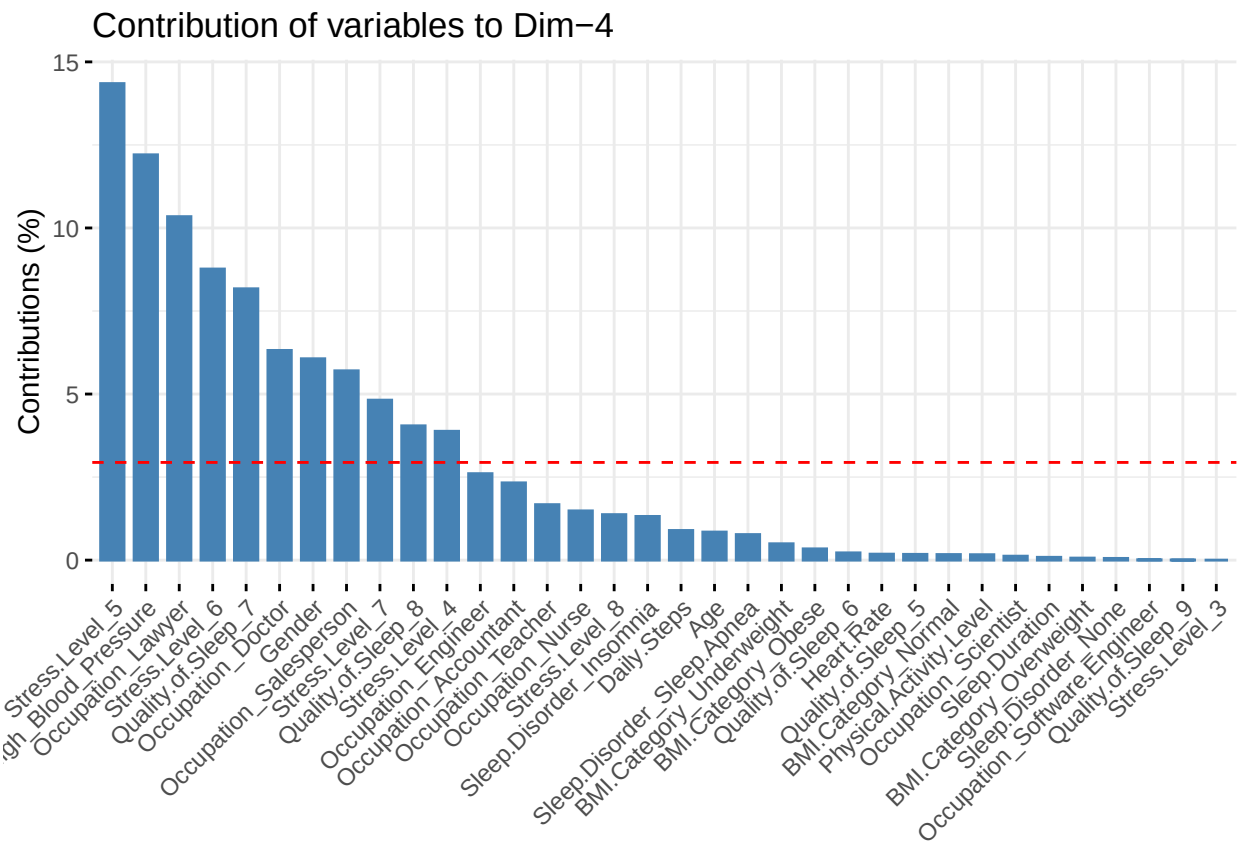

```
FigContrib2
```



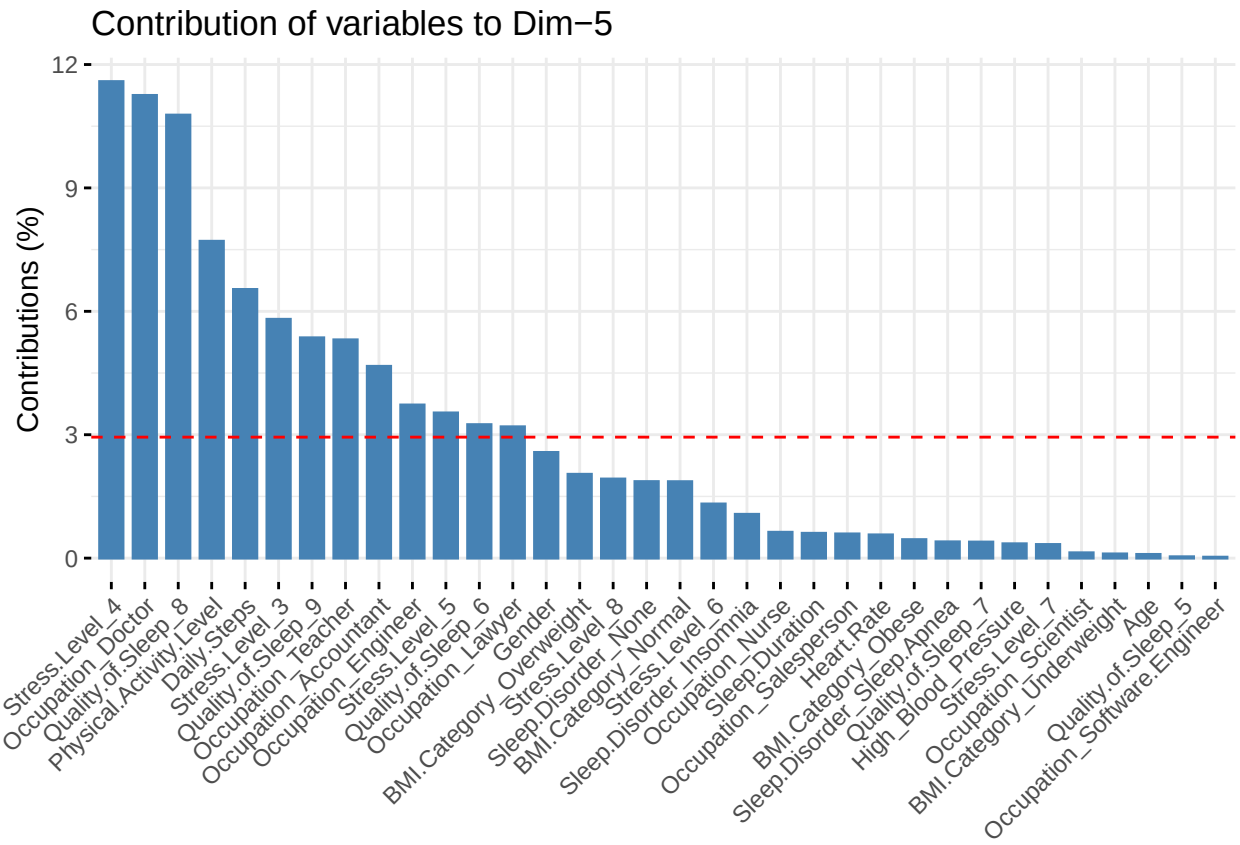

```
FigContrib3
```




```
FigContrib4
```

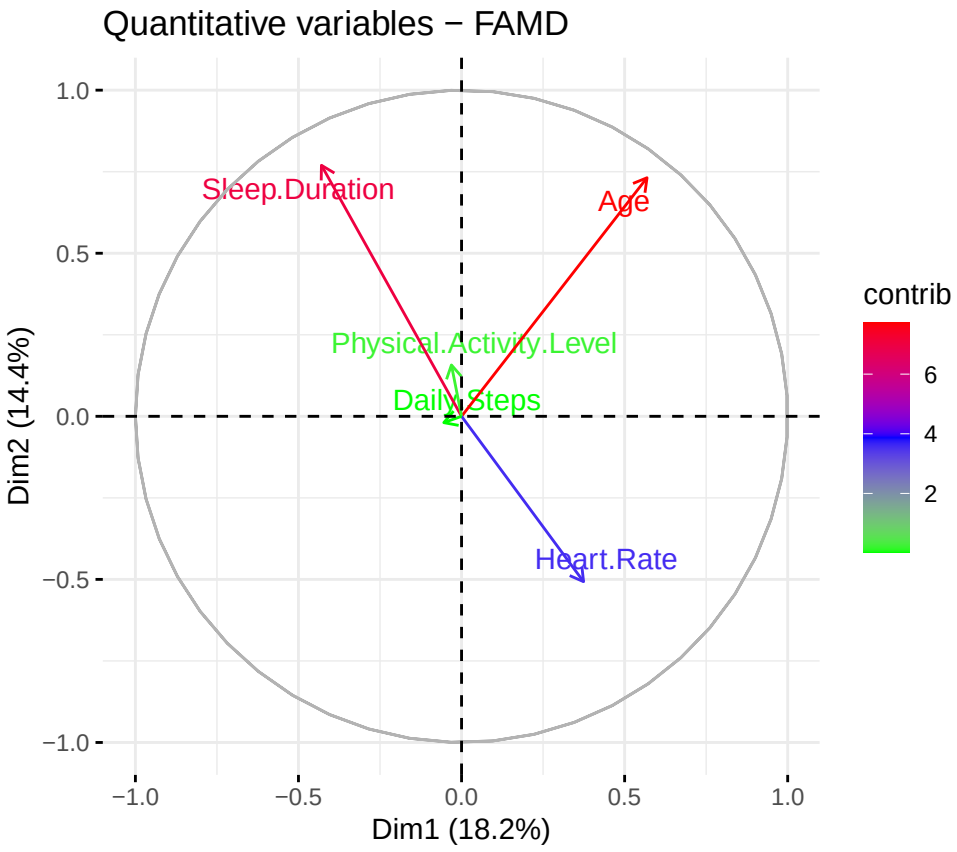



```
FigContrib5
```

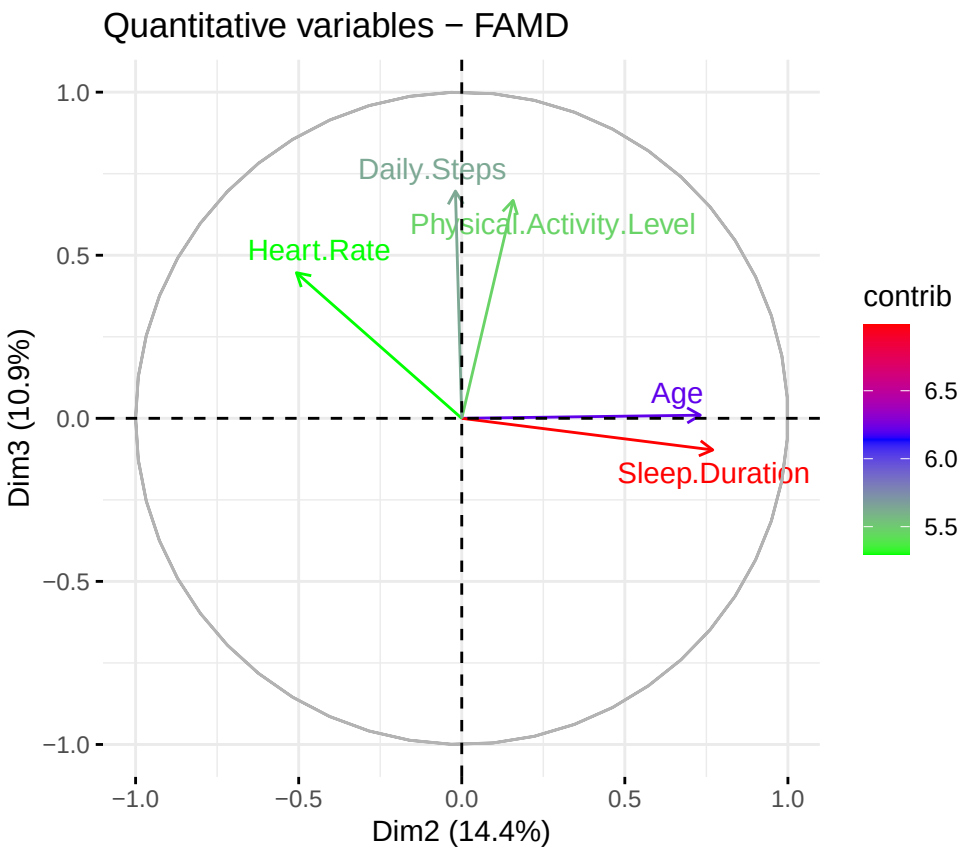


Quantitative Variables Contribution

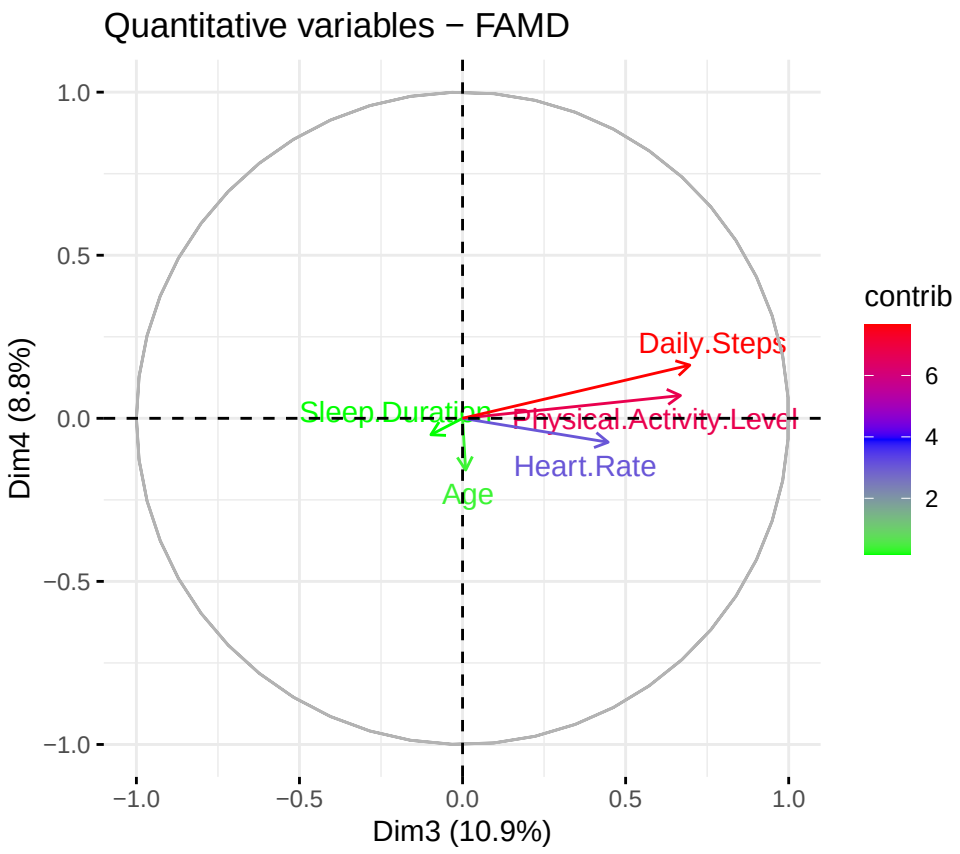
```
QuanVari1 <- fviz_famd_var(res.famd, "quanti.var", col.var = "contrib",
             gradient.cols = c("GREEN", "BLUE", "RED"),
             repel = TRUE)
QuanVari1
```



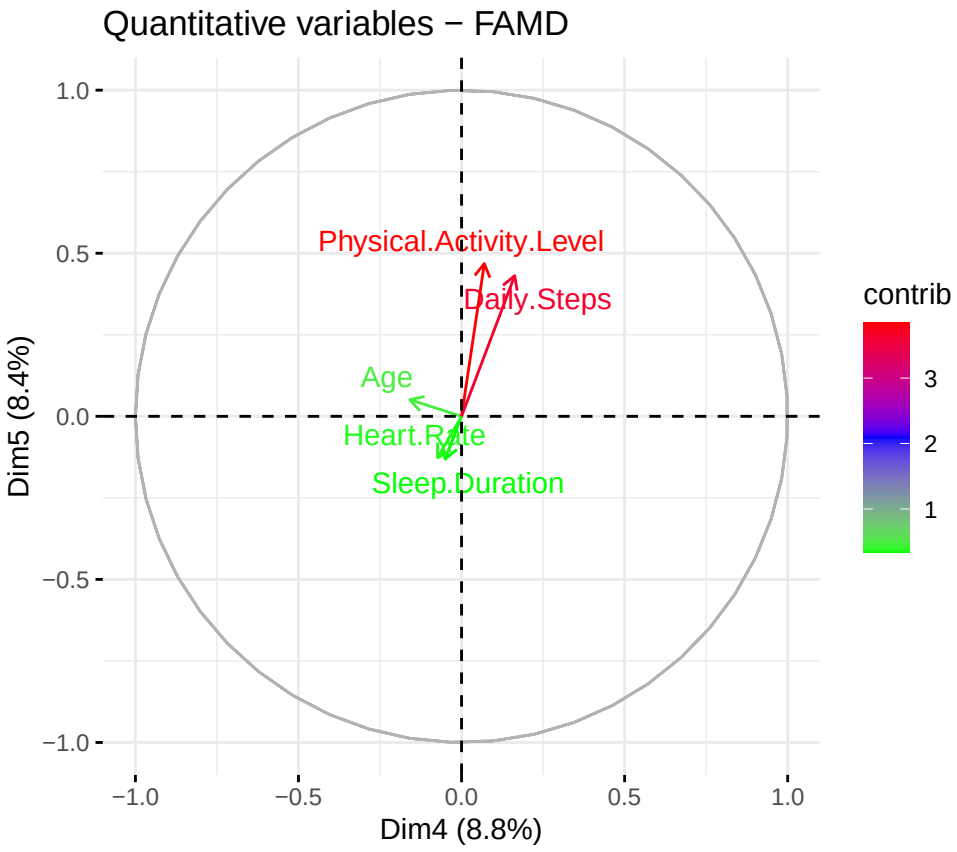

```
QuanVari2 <- fviz_famd_var(res.famd, "quanti.var", axes = c(2, 3),
                           col.var = "contrib",
                           gradient.cols = c("GREEN", "BLUE", "RED"),
                           repel = TRUE)
QuanVari2
```




```
QuanVari3 <- fviz_famd_var(res.famd, "quanti.var", axes = c(3, 4),
                           col.var = "contrib",
                           gradient.cols = c("GREEN", "BLUE", "RED"),
                           repel = TRUE)
QuanVari3
```




```
QuanVari4 <- fviz_famd_var(res.famd, "quanti.var", axes = c(4, 5),
                           col.var = "contrib",
                           gradient.cols = c("GREEN", "BLUE", "RED"),
                           repel = TRUE)
QuanVari4
```


Quantitative variables − FAMD
Physical.Activity.Level
Daily.Steps
Age
Heart.Rate
Sleep.Duration
contrib
3
2
1
1.0
0.5
0.0
−0.5
−1.0
Dim5 (8.4%)
Dim4 (8.8%)