Exploratory Factor Analysis

Zircon

2026-05-26

目录

1 这章在讲什么

主成分分析（PCA）和探索性因子分析（EFA）都是**降维**：把几十个高度相关的变量压缩成少数几个「主成分」/「因子」，让数据更可解释、更容易建模。两者数学很像（都用特征分解 / SVD），但**问题假设和目标不同**：

- **PCA**：纯几何变换——找一组互不相关的新坐标轴，让数据在第一根轴上方差最大、第二根次之……目标是**解释方差**。主成分本身没有「实质含义」，是数据自身的方向。
- **EFA**：假设有几个**潜在心理结构**（latent constructs）驱动了观测变量，目标是**推断这些结构是什么**。比如大五人格的 25 道题背后有 5 个潜在因子（开放性、尽责性、外向性、宜人性、神经质）。

PCA 的数学

对中心化（标准化）后的数据 $\mathbf{X}_{n\times p}$，计算协方差矩阵 $\mathbf{S} = \frac{1}{n-1}\mathbf{X}'\mathbf{X}$，做特征分解 $\mathbf{S} = \mathbf{V}\ \mathbf{V}'$。

特征向量 $\mathbf{V}$ 是新坐标轴的方向，特征值 λ_i 告诉你每根轴上有多少方差。按 λ 从大到小排序，保留前 k 个就完成了降维。

「保留多少 k」三种判据：(1) **Kaiser 准则** $\lambda_i > 1$（保留高于平均方差的成分）；(2) **碎石图肘部**（scree plot，看曲线突然变缓的位置）；(3) **平行分析**（parallel analysis，跟随机数据比较，更稳健，bruceR::EFA 默认就给）。

EFA 跟 PCA 的关键区别：EFA 引入「特有方差」（unique variance）——每个观测变量除了被共同因子解释，还有自己独有的方差（包括测量误差）。EFA 只解释「公因子之间的协方差」，PCA 解释「所有方差」。实务上对人格 / 智力量表做心理测量分析时用 EFA；对图像 / 基因表达做降维时用 PCA。

```r
rm(list=ls())
library("readxl")
library("psych")
library(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
```

```r
library(tidyr)
library(bruceR)
```

```
##
## bruceR (v0.8.10)
## BRoadly Useful Convenient and Efficient R functions
##
## Packages also loaded:
##   data.table   emmeans
##   dplyr        lmerTest
##   tidyr        effectsize
##   stringr      performance
##   ggplot2      interactions
##
## Main functions of `bruceR`:
## cc()             Describe()  TTEST()
## add()            Freq()      MANOVA()
## .mean()          Corr()      EMMEANS()
## set.wd()         Alpha()     PROCESS()
## import()         EFA()       model_summary()
## print_table()    CFA()       lavaan_summary()
##
## For full functionality, please install all dependencies:
## install.packages("bruceR", dep=TRUE)
##
## Online documentation:
## https://psychbruce.github.io/bruceR
```

```
##
## NEWS: A new version of bruceR (2026.1) is available (2026-01-29)!
##
```

```
## ************** Update **************
## install.packages("bruceR", dep=TRUE)
## ************************************
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(ggplot2)
library(GPArotation)
```

```
##
## Attaching package: 'GPArotation'
```

```
## The following objects are masked from 'package:psych':
##
##     equamax, varimin
```

```
currentData = psych::bfi
currentData = currentData %>% drop_na()
currentData = currentData[complete.cases(currentData),]
currentData = currentData[1:25]
```

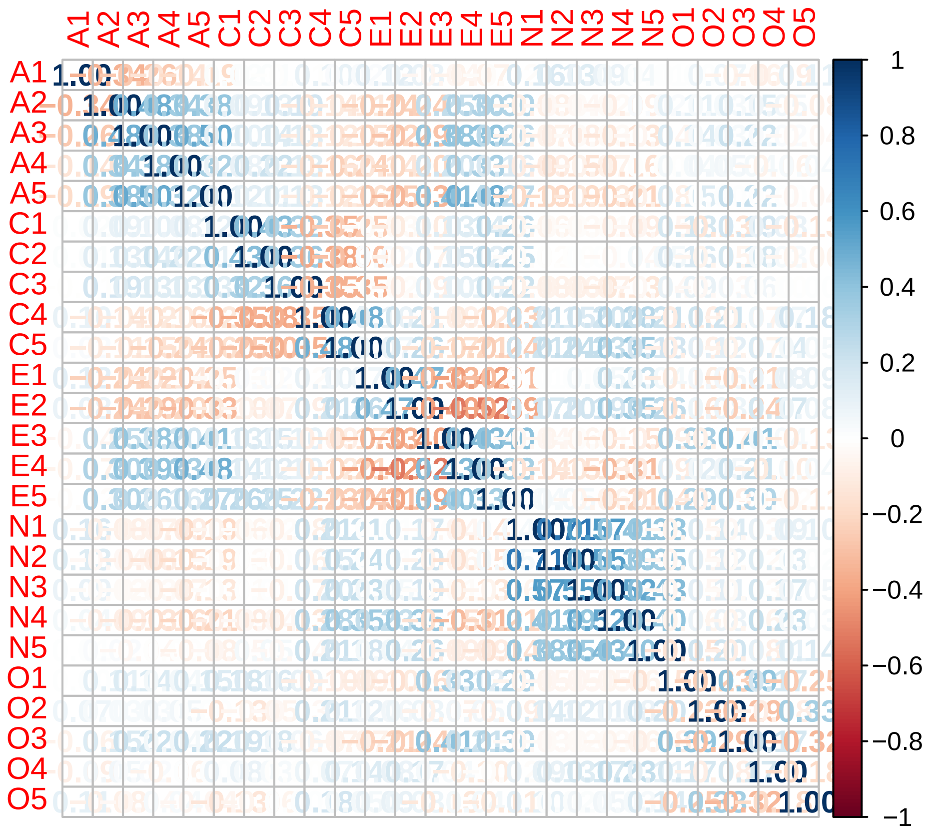
```
describe(currentData)
```

```
##    vars    n mean   sd median trimmed  mad min max range  skew kurtosis   se
## A1    1 2236 2.37 1.39      2    2.17 1.48   1   6     5  0.88    -0.17 0.03
## A2    2 2236 4.83 1.16      5    5.01 1.48   1   6     5 -1.15     1.14 0.02
## A3    3 2236 4.63 1.29      5    4.82 1.48   1   6     5 -1.03     0.56 0.03
## A4    4 2236 4.75 1.45      5    4.99 1.48   1   6     5 -1.09     0.23 0.03
## A5    5 2236 4.58 1.26      5    4.74 1.48   1   6     5 -0.88     0.24 0.03
## C1    6 2236 4.57 1.22      5    4.71 1.48   1   6     5 -0.89     0.42 0.03
## C2    7 2236 4.40 1.31      5    4.54 1.48   1   6     5 -0.77    -0.09 0.03
## C3    8 2236 4.32 1.29      5    4.44 1.48   1   6     5 -0.69    -0.11 0.03
## C4    9 2236 2.50 1.36      2    2.35 1.48   1   6     5  0.64    -0.56 0.03
## C5   10 2236 3.26 1.63      3    3.20 1.48   1   6     5  0.09    -1.23 0.03
## E1   11 2236 2.97 1.62      3    2.86 1.48   1   6     5  0.38    -1.07 0.03
```

```
## E2    12 2236 3.12 1.61      3    3.04 1.48   1   6     5  0.25    -1.13 0.03
## E3    13 2236 4.01 1.34      4    4.08 1.48   1   6     5 -0.48    -0.43 0.03
## E4    14 2236 4.43 1.46      5    4.60 1.48   1   6     5 -0.85    -0.27 0.03
## E5    15 2236 4.42 1.33      5    4.57 1.48   1   6     5 -0.81    -0.03 0.03
## N1    16 2236 2.91 1.56      3    2.80 1.48   1   6     5  0.39    -0.99 0.03
## N2    17 2236 3.49 1.53      4    3.48 1.48   1   6     5 -0.06    -1.07 0.03
## N3    18 2236 3.20 1.60      3    3.14 1.48   1   6     5  0.17    -1.18 0.03
## N4    19 2236 3.18 1.56      3    3.11 1.48   1   6     5  0.22    -1.06 0.03
## N5    20 2236 2.95 1.62      3    2.83 1.48   1   6     5  0.40    -1.05 0.03
## O1    21 2236 4.82 1.12      5    4.97 1.48   1   6     5 -0.91     0.47 0.02
## O2    22 2236 2.69 1.55      2    2.54 1.48   1   6     5  0.61    -0.77 0.03
## O3    23 2236 4.48 1.19      5    4.60 1.48   1   6     5 -0.79     0.40 0.03
## O4    24 2236 4.95 1.18      5    5.15 1.48   1   6     5 -1.26     1.26 0.02
## O5    25 2236 2.46 1.33      2    2.30 1.48   1   6     5  0.78    -0.16 0.03
```

1.1 Correlation Matrix

```
cor_matrix = cor(currentData)
corrplot(cor_matrix, method="number")
```



1.2 Data Adequacy

get the overall MSA from KMO; look at the bartlett test

```
KMO(cor(currentData))
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = cor(currentData))
## Overall MSA =  0.85
## MSA for each item =
##   A1   A2   A3   A4   A5   C1   C2   C3   C4   C5   E1   E2   E3   E4   E5   N1
## 0.74 0.83 0.87 0.87 0.90 0.84 0.79 0.85 0.82 0.86 0.84 0.88 0.89 0.88 0.89 0.78
##   N2   N3   N4   N5   O1   O2   O3   O4   O5
## 0.78 0.86 0.89 0.86 0.86 0.78 0.83 0.78 0.76
```

```r
cortest.bartlett(cor(currentData),nrow(currentData))
```

```
## $chisq
## [1] 16484.78
##
## $p.value
## [1] 0
##
## $df
## [1] 300
```

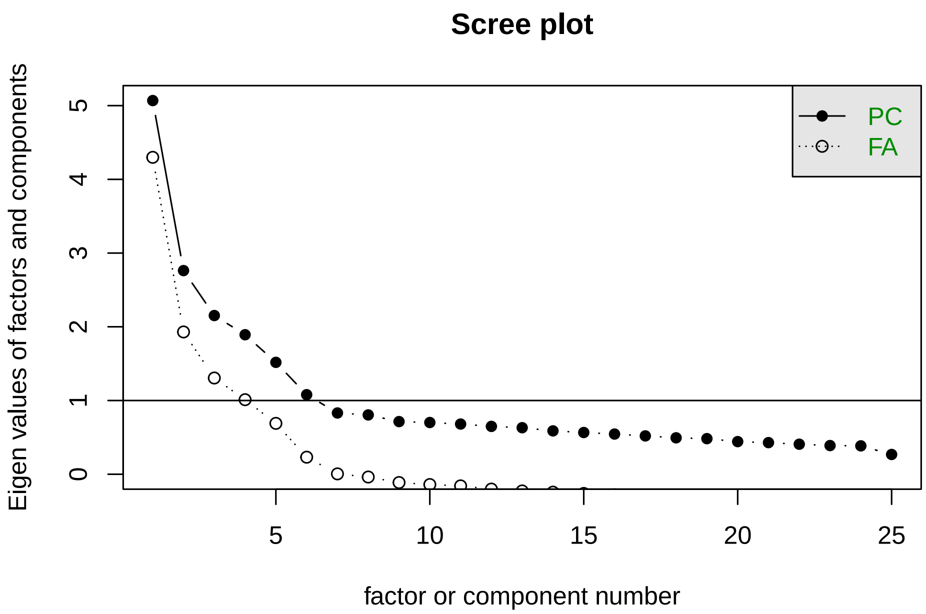
```r
det(cor_matrix)
```

```
## [1] 0.0006075253
```

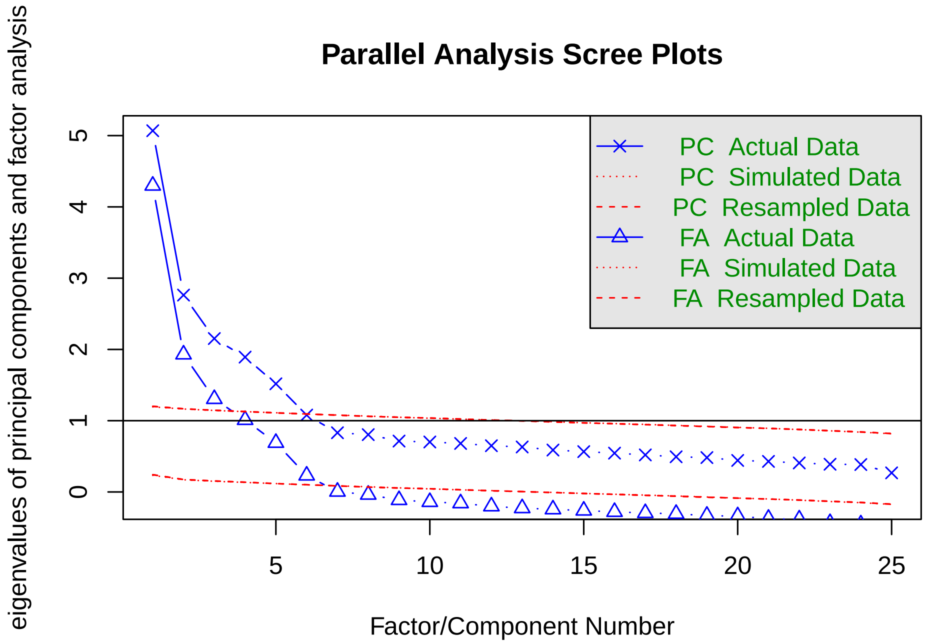
1.3 Determine num. of Factors

- To determine the number of factors, scree and parallel are two methods.
- Scree is obsolete; parallel suggests 3 factors here.

```r
scree(currentData)
```




```
fa.parallel(currentData,fm='pa',fa='both')
```




```
## Parallel analysis suggests that the number of factors =  6  and the number of compon
```

2 Factor Analysis

2.1 NULL Model

Take a look: without selecting the number of factors, without rotation.

Factoring Method "`fm`": * `minres`: do a minimum residual * `pa`: do the principal factor solution * `ml`: do a maximum likelihood factor analysis

The Proportion Var and Cumulative Var are the variance that can be explained by factors.

"SS loadings" is the sum of squared loadings for each factor; **>1** means the factor is worth keeping.

The model can be printed with the following results:

- Standardized loadings (pattern matrix) based upon correlation matrix
  - SS loadings
    - The eigenvalues, the sum of the squared loadings for each factor.
    - **Better to be >1**.
  - Proportion Var
    - The overall variance (for all variables/items) that can be explained by each factor.
  - Cumulative Var
  - Proportion Explained
    - The relative importance of each factor.
  - Cumulative Proportion
- Test of the hypothesis that selected numbers of factors are sufficient
- Measures of factor score adequacy
  - Correlation of (regression) scores with factors
  - Multiple R square of scores with factors
  - Minimum correlation of possible factor scores
- Data frame with loading matrix, communality, uniqueness, and item complexity.

Standardized loadings (pattern matrix) based upon correlation matrix is the most important part!

```
fa_none = fa(currentData,nfactors = ncol(currentData),
             fm='minres',rotate='none')
print(fa_none)
```

```
## Factor Analysis using method =  minres
## Call: fa(r = currentData, nfactors = ncol(currentData), rotate = "none",
##     fm = "minres")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      MR1   MR2   MR3   MR4   MR5   MR6   MR7   MR8   MR9  MR10  MR11  MR12
## A1 -0.22 -0.03  0.15  0.03 -0.42  0.31  0.07 -0.05  0.02  0.01 -0.02  0.10
## A2  0.47  0.31 -0.20  0.14  0.37 -0.23  0.14  0.15 -0.12  0.01 -0.05  0.07
## A3  0.53  0.31 -0.25  0.10  0.28  0.02  0.13 -0.12  0.09  0.00 -0.02 -0.12
## A4  0.41  0.13 -0.16  0.30  0.18  0.03 -0.01 -0.23 -0.06  0.23 -0.02  0.17
## A5  0.57  0.19 -0.27  0.04  0.18  0.15  0.02 -0.05  0.19 -0.10  0.07 -0.07
## C1  0.33  0.13  0.47  0.14  0.00  0.11 -0.09  0.19  0.17  0.03 -0.03  0.03
## C2  0.33  0.19  0.49  0.32  0.06  0.19 -0.10  0.06 -0.05  0.23 -0.03 -0.11
## C3  0.33  0.05  0.35  0.34  0.03  0.02  0.07  0.14 -0.07 -0.20  0.06  0.01
## C4 -0.47  0.13 -0.45 -0.24  0.02  0.27  0.08  0.06 -0.11  0.04  0.05  0.04
## C5 -0.50  0.16 -0.29 -0.31  0.12  0.08 -0.01  0.26  0.10  0.17 -0.08 -0.06
## E1 -0.43 -0.21  0.27  0.14  0.27  0.25  0.23 -0.14  0.06 -0.01 -0.09 -0.03
## E2 -0.63 -0.06  0.22  0.09  0.33  0.11  0.11 -0.01  0.03  0.03  0.20  0.02
## E3  0.55  0.34 -0.09 -0.20 -0.10  0.21  0.03 -0.09 -0.07 -0.11  0.01 -0.07
## E4  0.62  0.19 -0.34  0.08 -0.19  0.15 -0.16  0.03  0.15  0.01  0.02  0.12
## E5  0.53  0.30  0.11 -0.04 -0.23 -0.03  0.22  0.15 -0.11  0.02 -0.06  0.02
## N1 -0.43  0.65  0.06  0.09 -0.27 -0.13  0.13 -0.12  0.01  0.05  0.05 -0.03
## N2 -0.42  0.65  0.12  0.05 -0.22 -0.21  0.15  0.00  0.14  0.05  0.09 -0.01
## N3 -0.40  0.63  0.06  0.06 -0.01 -0.01 -0.13 -0.10  0.03 -0.11 -0.16  0.05
## N4 -0.54  0.42  0.09 -0.04  0.23  0.08 -0.11 -0.02 -0.06 -0.08 -0.22 -0.02
## N5 -0.36  0.44 -0.05  0.22  0.13  0.05 -0.27  0.02 -0.15 -0.03  0.18 -0.01
## O1  0.33  0.18  0.24 -0.36  0.02  0.16  0.14 -0.01 -0.09  0.01 -0.02  0.10
## O2 -0.20  0.09 -0.33  0.39 -0.05  0.15  0.10  0.15 -0.01 -0.01  0.01  0.01
```

```
## O3  0.40  0.28  0.19 -0.47  0.03  0.13 -0.05 -0.03 -0.08  0.05  0.11 -0.11
## O4 -0.09  0.24  0.18 -0.24  0.31  0.07  0.01  0.08  0.08 -0.08  0.06  0.18
## O5 -0.20 -0.05 -0.29  0.44 -0.15  0.21  0.06  0.06 -0.07 -0.05 -0.02 -0.06
##     MR13  MR14  MR15  MR16  MR17  MR18  MR19  MR20  MR21  MR22  MR23 MR24 MR25
## A1  0.05  0.17  0.00  0.08 -0.02 -0.02 -0.03  0.03  0.02 -0.01  0.00    0    0
## A2  0.07  0.07 -0.07  0.07  0.00  0.05 -0.01  0.02  0.00  0.03  0.00    0    0
## A3  0.01 -0.02  0.06  0.02 -0.08  0.01 -0.06  0.02  0.00 -0.05  0.00    0    0
## A4  0.02  0.02  0.07  0.00 -0.01 -0.03  0.04  0.00  0.00  0.00  0.01    0    0
## A5  0.04  0.09 -0.03 -0.05  0.06 -0.09  0.01 -0.03  0.03  0.02  0.00    0    0
## C1  0.08 -0.03 -0.03 -0.07 -0.13  0.00 -0.01  0.01 -0.04  0.01  0.01    0    0
## C2 -0.08 -0.02  0.01  0.03  0.04  0.00  0.02 -0.03  0.01  0.00 -0.01    0    0
## C3  0.09  0.01  0.12  0.12  0.02 -0.04  0.03 -0.02 -0.01  0.00  0.00    0    0
## C4  0.10 -0.06 -0.03  0.03 -0.08 -0.02  0.06 -0.05 -0.02 -0.02 -0.01    0    0
## C5  0.02  0.07  0.10  0.01  0.03  0.00 -0.01 -0.01  0.01  0.02  0.01    0    0
## E1  0.03 -0.05 -0.15  0.02  0.03  0.02  0.00 -0.02  0.00  0.02  0.01    0    0
## E2  0.01  0.10  0.06 -0.09  0.01  0.07  0.02  0.05 -0.01  0.00 -0.01    0    0
## E3 -0.14  0.06  0.04 -0.02 -0.02  0.05  0.05  0.00 -0.04  0.03  0.01    0    0
## E4  0.04 -0.06 -0.06  0.03  0.09  0.06  0.00  0.02 -0.03  0.00 -0.01    0    0
## E5  0.01  0.04 -0.07 -0.15  0.01 -0.03  0.04  0.02  0.03 -0.03  0.00    0    0
## N1  0.06 -0.10  0.03 -0.03 -0.01 -0.08 -0.03  0.02 -0.01  0.05 -0.01    0    0
## N2 -0.04  0.02 -0.03  0.07  0.06  0.05  0.03 -0.03 -0.01 -0.03  0.01    0    0
## N3  0.02  0.03  0.02 -0.01 -0.07  0.08  0.02 -0.03  0.05  0.01 -0.01    0    0
## N4  0.00  0.02 -0.01  0.01  0.08 -0.06  0.01  0.05 -0.04 -0.02  0.00    0    0
## N5  0.01  0.05 -0.08 -0.05  0.00 -0.02 -0.06 -0.03  0.00 -0.01  0.01    0    0
## O1 -0.02 -0.01  0.06 -0.04  0.06  0.03 -0.10 -0.05 -0.02  0.00  0.00    0    0
## O2 -0.21 -0.01 -0.05  0.06 -0.06 -0.03 -0.03  0.02  0.00  0.01  0.00    0    0
## O3  0.07 -0.07 -0.04  0.09 -0.01  0.02  0.01  0.06  0.04  0.00  0.01    0    0
## O4 -0.11 -0.13  0.04  0.00  0.00 -0.04  0.02  0.01  0.04  0.00  0.00    0    0
## O5  0.07 -0.12  0.09 -0.08  0.06  0.07  0.00  0.02  0.03  0.00  0.01    0    0
##      h2   u2 com
## A1 0.41 0.59 3.6
## A2 0.65 0.35 5.2
## A3 0.60 0.40 3.7
```

```
## A4 0.48 0.52 5.1
## A5 0.58 0.42 2.8
## C1 0.48 0.52 3.7
## C2 0.61 0.39 4.5
## C3 0.46 0.54 5.0
## C4 0.63 0.37 3.9
## C5 0.62 0.38 4.4
## E1 0.57 0.43 6.0
## E2 0.66 0.34 2.5
## E3 0.58 0.42 3.2
## E4 0.69 0.31 2.9
## E5 0.56 0.44 3.4
## N1 0.77 0.23 2.7
## N2 0.77 0.23 2.8
## N3 0.65 0.35 2.3
## N4 0.62 0.38 3.2
## N5 0.54 0.46 4.6
## O1 0.41 0.59 4.9
## O2 0.42 0.58 4.5
## O3 0.57 0.43 4.0
## O4 0.35 0.65 6.4
## O5 0.45 0.55 4.1
##
##                        MR1  MR2  MR3  MR4  MR5  MR6  MR7  MR8  MR9 MR10 MR11
## SS loadings           4.67 2.41 1.70 1.40 1.07 0.63 0.39 0.34 0.24 0.24 0.21
## Proportion Var        0.19 0.10 0.07 0.06 0.04 0.03 0.02 0.01 0.01 0.01 0.01
## Cumulative Var        0.19 0.28 0.35 0.41 0.45 0.47 0.49 0.50 0.51 0.52 0.53
## Proportion Explained  0.33 0.17 0.12 0.10 0.08 0.04 0.03 0.02 0.02 0.02 0.01
## Cumulative Proportion 0.33 0.50 0.62 0.72 0.80 0.84 0.87 0.89 0.91 0.93 0.94
##                       MR12 MR13 MR14 MR15 MR16 MR17 MR18 MR19 MR20 MR21 MR22
## SS loadings           0.16 0.13 0.12 0.10 0.10 0.07 0.05 0.03 0.02 0.02 0.01
## Proportion Var        0.01 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## Cumulative Var        0.54 0.54 0.55 0.55 0.56 0.56 0.56 0.56 0.56 0.56 0.56
```

```
## Proportion Explained  0.01 0.01 0.01 0.01 0.01 0.01 0.00 0.00 0.00 0.00 0.00
## Cumulative Proportion 0.95 0.96 0.97 0.98 0.99 0.99 0.99 1.00 1.00 1.00 1.00
##                       MR23 MR24 MR25
## SS loadings           0.00 0.00 0.00
## Proportion Var        0.00 0.00 0.00
## Cumulative Var        0.56 0.56 0.56
## Proportion Explained  0.00 0.00 0.00
## Cumulative Proportion 1.00 1.00 1.00
##
## Mean item complexity =  4
## Test of the hypothesis that 25 factors are sufficient.
##
## df null model =  300  with the objective function =  7.41 with Chi Square =  16484.7
## df of  the model are -25  and the objective function was  0
##
## The root mean square of the residuals (RMSR) is  0
## The df corrected root mean square of the residuals is  NA
##
## The harmonic n.obs is  2236 with the empirical chi square  0  with prob <  NA
## The total n.obs was  2236  with Likelihood Chi Square =  0  with prob <  NA
##
## Tucker Lewis Index of factoring reliability =  1.019
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##                                                    MR1  MR2  MR3  MR4  MR5  MR6
## Correlation of (regression) scores with factors   0.96 0.94 0.89 0.86 0.85 0.78
## Multiple R square of scores with factors          0.92 0.88 0.80 0.75 0.72 0.60
## Minimum correlation of possible factor scores     0.85 0.76 0.59 0.49 0.44 0.20
##                                                     MR7   MR8   MR9  MR10  MR11
## Correlation of (regression) scores with factors    0.70  0.66  0.61  0.59  0.59
## Multiple R square of scores with factors           0.49  0.43  0.37  0.35  0.35
## Minimum correlation of possible factor scores     -0.02 -0.13 -0.25 -0.30 -0.29
##                                                    MR12  MR13  MR14  MR15  MR16
```

```
## Correlation of (regression) scores with factors    0.50  0.47  0.46  0.43  0.42
## Multiple R square of scores with factors           0.25  0.22  0.21  0.19  0.18
## Minimum correlation of possible factor scores     -0.49 -0.57 -0.58 -0.62 -0.64
##                                                    MR17  MR18  MR19  MR20  MR21
## Correlation of (regression) scores with factors    0.39  0.36  0.26  0.23  0.19
## Multiple R square of scores with factors           0.15  0.13  0.07  0.05  0.04
## Minimum correlation of possible factor scores     -0.70 -0.75 -0.86 -0.90 -0.93
##                                                    MR22  MR23  MR24 MR25
## Correlation of (regression) scores with factors    0.17  0.06  0.01    0
## Multiple R square of scores with factors           0.03  0.00  0.00    0
## Minimum correlation of possible factor scores     -0.94 -0.99 -1.00   -1
```

For illustration purpose:

- If using 25 factors (here only 25 items), we can see how loading and proportion of variance decrease with factors;
- The cumulative variance explained will approach 1.

2.2 Screened Model

According to spree plot, take 6 factors.

Note that restrictions on factor number will have practical effect on the model, instead of just trimming the model to the reduced one.

```
fa6_none = fa(currentData, 6,fm='minres',rotate='none')
print(fa6_none)
```

```
## Factor Analysis using method =  minres
## Call: fa(r = currentData, nfactors = 6, rotate = "none", fm = "minres")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      MR1   MR2   MR3   MR4   MR5   MR6   h2   u2 com
## A1 -0.22 -0.03  0.15  0.04 -0.42  0.29 0.33 0.67 2.7
## A2  0.46  0.30 -0.20  0.12  0.34 -0.18 0.50 0.50 3.7
## A3  0.52  0.31 -0.25  0.09  0.27  0.02 0.51 0.49 2.8
```

```
## A4  0.40  0.12 -0.15  0.26  0.16  0.03 0.29 0.71 2.7
## A5  0.57  0.18 -0.26  0.03  0.17  0.13 0.47 0.53 2.0
## C1  0.33  0.12  0.44  0.15  0.00  0.10 0.35 0.65 2.4
## C2  0.33  0.18  0.45  0.32  0.06  0.17 0.48 0.52 3.5
## C3  0.32  0.05  0.32  0.33  0.04  0.04 0.32 0.68 3.1
## C4 -0.47  0.13 -0.45 -0.26  0.01  0.26 0.58 0.42 3.4
## C5 -0.49  0.15 -0.26 -0.30  0.10  0.05 0.43 0.57 2.7
## E1 -0.42 -0.20  0.24  0.13  0.25  0.22 0.40 0.60 3.8
## E2 -0.62 -0.06  0.21  0.08  0.32  0.12 0.56 0.44 1.9
## E3  0.55  0.34 -0.09 -0.19 -0.11  0.19 0.51 0.49 2.4
## E4  0.61  0.19 -0.33  0.08 -0.18  0.12 0.56 0.44 2.1
## E5  0.52  0.29  0.11 -0.02 -0.21 -0.03 0.41 0.59 2.1
## N1 -0.43  0.65  0.05  0.10 -0.25 -0.13 0.69 0.31 2.3
## N2 -0.41  0.63  0.11  0.06 -0.20 -0.19 0.67 0.33 2.3
## N3 -0.40  0.61  0.05  0.07  0.00 -0.01 0.54 0.46 1.8
## N4 -0.53  0.41  0.09 -0.04  0.22  0.08 0.51 0.49 2.4
## N5 -0.35  0.41 -0.05  0.20  0.12  0.05 0.35 0.65 2.7
## O1  0.33  0.18  0.24 -0.35  0.00  0.15 0.34 0.66 3.8
## O2 -0.20  0.09 -0.33  0.36 -0.04  0.14 0.31 0.69 3.1
## O3  0.40  0.28  0.19 -0.45  0.01  0.12 0.49 0.51 3.3
## O4 -0.09  0.24  0.18 -0.23  0.29  0.07 0.24 0.76 4.0
## O5 -0.20 -0.05 -0.30  0.42 -0.13  0.20 0.37 0.63 3.2
##
##                        MR1  MR2  MR3  MR4  MR5  MR6
## SS loadings           4.55 2.30 1.58 1.29 0.96 0.53
## Proportion Var        0.18 0.09 0.06 0.05 0.04 0.02
## Cumulative Var        0.18 0.27 0.34 0.39 0.43 0.45
## Proportion Explained  0.41 0.21 0.14 0.12 0.09 0.05
## Cumulative Proportion 0.41 0.61 0.75 0.87 0.95 1.00
##
## Mean item complexity =  2.8
## Test of the hypothesis that 6 factors are sufficient.
##
```

```
## df null model =  300  with the objective function =  7.41 with Chi Square =  16484.7
## df of  the model are 165  and the objective function was  0.37
##
## The root mean square of the residuals (RMSR) is  0.02
## The df corrected root mean square of the residuals is  0.03
##
## The harmonic n.obs is  2236 with the empirical chi square  500.78  with prob <  1.6e
## The total n.obs was  2236  with Likelihood Chi Square =  824.5  with prob <  3e-88
##
## Tucker Lewis Index of factoring reliability =  0.926
## RMSEA index =  0.042  and the 90 % confidence intervals are  0.039 0.045
## BIC =  -448.05
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                    MR1  MR2  MR3  MR4  MR5  MR6
## Correlation of (regression) scores with factors   0.95 0.92 0.86 0.83 0.80 0.71
## Multiple R square of scores with factors          0.90 0.84 0.74 0.69 0.65 0.50
## Minimum correlation of possible factor scores     0.80 0.68 0.49 0.38 0.29 0.01
```

However, the old big five model only has 5 factors, and we're to designate this number.

```
fa5_none = fa(currentData, 5,fm='minres',rotate='none')
print(fa5_none)
```

```
## Factor Analysis using method =  minres
## Call: fa(r = currentData, nfactors = 5, rotate = "none", fm = "minres")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      MR1   MR2   MR3   MR4   MR5   h2   u2 com
## A1 -0.21 -0.02  0.13  0.04 -0.36 0.20 0.80 2.0
## A2  0.45  0.29 -0.19  0.11  0.32 0.44 0.56 3.2
## A3  0.52  0.31 -0.26  0.09  0.29 0.53 0.47 2.9
## A4  0.40  0.12 -0.16  0.27  0.17 0.30 0.70 2.8
## A5  0.57  0.18 -0.26  0.02  0.18 0.46 0.54 1.9
## C1  0.33  0.12  0.44  0.15  0.00 0.34 0.66 2.3
```

```
## C2  0.32  0.18  0.44  0.32  0.06 0.44 0.56 3.2
## C3  0.32  0.05  0.32  0.34  0.04 0.33 0.67 3.1
## C4 -0.46  0.12 -0.42 -0.25  0.02 0.47 0.53 2.7
## C5 -0.49  0.15 -0.26 -0.31  0.10 0.43 0.57 2.7
## E1 -0.41 -0.20  0.24  0.13  0.24 0.34 0.66 3.1
## E2 -0.62 -0.06  0.21  0.08  0.32 0.55 0.45 1.8
## E3  0.54  0.33 -0.09 -0.19 -0.10 0.46 0.54 2.1
## E4  0.61  0.18 -0.33  0.07 -0.17 0.54 0.46 2.0
## E5  0.52  0.29  0.11 -0.02 -0.22 0.42 0.58 2.1
## N1 -0.43  0.64  0.05  0.10 -0.26 0.68 0.32 2.2
## N2 -0.41  0.62  0.10  0.06 -0.20 0.61 0.39 2.1
## N3 -0.40  0.62  0.05  0.07 -0.01 0.55 0.45 1.8
## N4 -0.53  0.41  0.08 -0.04  0.22 0.51 0.49 2.3
## N5 -0.35  0.42 -0.06  0.20  0.12 0.35 0.65 2.7
## O1  0.33  0.18  0.25 -0.34  0.01 0.32 0.68 3.4
## O2 -0.20  0.09 -0.34  0.35 -0.03 0.28 0.72 2.7
## O3  0.40  0.28  0.20 -0.45  0.02 0.48 0.52 3.1
## O4 -0.09  0.24  0.19 -0.23  0.30 0.24 0.76 3.9
## O5 -0.20 -0.05 -0.30  0.40 -0.12 0.31 0.69 2.7
##
##                        MR1  MR2  MR3  MR4  MR5
## SS loadings           4.53 2.28 1.55 1.26 0.92
## Proportion Var        0.18 0.09 0.06 0.05 0.04
## Cumulative Var        0.18 0.27 0.33 0.38 0.42
## Proportion Explained  0.43 0.22 0.15 0.12 0.09
## Cumulative Proportion 0.43 0.65 0.79 0.91 1.00
##
## Mean item complexity =  2.6
## Test of the hypothesis that 5 factors are sufficient.
##
## df null model =  300  with the objective function =  7.41 with Chi Square =  16484.7
## df of  the model are 185  and the objective function was  0.63
##
```

```
## The root mean square of the residuals (RMSR) is  0.03
## The df corrected root mean square of the residuals is  0.04
##
## The harmonic n.obs is  2236 with the empirical chi square  1046.45  with prob <  3.1
## The total n.obs was  2236  with Likelihood Chi Square =  1400.38  with prob <  1.6e-
##
## Tucker Lewis Index of factoring reliability =  0.878
## RMSEA index =  0.054  and the 90 % confidence intervals are  0.052 0.057
## BIC =  -26.42
## Fit based upon off diagonal values = 0.98
## Measures of factor score adequacy
##                                                    MR1  MR2  MR3  MR4  MR5
## Correlation of (regression) scores with factors   0.95 0.91 0.85 0.82 0.79
## Multiple R square of scores with factors          0.90 0.83 0.73 0.67 0.63
## Minimum correlation of possible factor scores     0.79 0.66 0.45 0.35 0.26
```

Obviously, with more factors taken into account, more cumulative variance will be explained.

3 Factor Rotation

Factor loadings listed for each variable. The square of it is the variation accounted for.

The hope is that all these correlation coefficients are close to (+/-)1 or 0, so we can easily interpret the meanings of these components (factors). We will accomplish this goal through rotation.

The relationship (correlation coefficient r) between the variable and the factor is shown. We can view it on a **Factor Plot**.

Use 2 factors for illustration purpose, to show how rotation works.

```
fa_unrotated = fa(currentData,nfactors = 2,fm='minres',rotate='none')
# print(fa_unrotated)
```

```r
fa_unrotated$loadings  #laodings, unrotated
```

```
##
## Loadings:
##    MR1    MR2
## A1 -0.207
## A2  0.439  0.282
## A3  0.511  0.301
## A4  0.394  0.125
## A5  0.563  0.189
## C1  0.311  0.119
## C2  0.300  0.166
## C3  0.308
## C4 -0.439
## C5 -0.473  0.128
## E1 -0.407 -0.200
## E2 -0.610
## E3  0.542  0.340
## E4  0.598  0.190
## E5  0.519  0.296
## N1 -0.432  0.621
## N2 -0.417  0.607
## N3 -0.414  0.626
## N4 -0.539  0.393
## N5 -0.354  0.407
## O1  0.315  0.168
## O2 -0.186
## O3  0.377  0.256
## O4         0.220
## O5 -0.189
##
##                  MR1   MR2
## SS loadings    4.392 2.193
```

```
## Proportion Var 0.176 0.088
## Cumulative Var 0.176 0.263
```

Now, rotate it (orthogonal and oblique).

Un-rotated factors are typically not very interpretable (most factors are correlated with many variables), though it has largest variance explained.

Factors are rotated to make them more meaningful and easier to interpret (each variable is associated with a minimal number of factors). Looking for “simple structure”.

The most popular rotational method is Varimax rotations (方差最大旋转）. Varimax use orthogonal rotations yielding uncorrelated factors/components.

Varimax attempts to maximize the variance that a single factor can account for (across all variables). This enhances the interpretability of the factors. (kind of simplification of interpretation)

```
fa_otho = fa(currentData,2,fm='minres',rotate='varimax')
print(fa_otho)
```

```
## Factor Analysis using method =  minres
## Call: fa(r = currentData, nfactors = 2, rotate = "varimax", fm = "minres")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      MR1   MR2    h2   u2 com
## A1 -0.19  0.09 0.044 0.96 1.4
## A2  0.52  0.01 0.272 0.73 1.0
## A3  0.59 -0.02 0.352 0.65 1.0
## A4  0.40 -0.10 0.171 0.83 1.1
## A5  0.58 -0.14 0.352 0.65 1.1
## C1  0.33 -0.06 0.111 0.89 1.1
## C2  0.34 -0.02 0.118 0.88 1.0
## C3  0.29 -0.12 0.098 0.90 1.3
## C4 -0.32  0.32 0.202 0.80 2.0
## C5 -0.33  0.36 0.240 0.76 2.0
## E1 -0.45  0.05 0.206 0.79 1.0
```

```
## E2 -0.55  0.26 0.377 0.62 1.4
## E3  0.64  0.00 0.409 0.59 1.0
## E4  0.61 -0.16 0.394 0.61 1.1
## E5  0.60 -0.02 0.357 0.64 1.0
## N1 -0.04  0.76 0.572 0.43 1.0
## N2 -0.03  0.74 0.542 0.46 1.0
## N3 -0.02  0.75 0.563 0.44 1.0
## N4 -0.25  0.62 0.444 0.56 1.3
## N5 -0.08  0.53 0.291 0.71 1.0
## O1  0.36 -0.02 0.127 0.87 1.0
## O2 -0.12  0.17 0.041 0.96 1.8
## O3  0.45  0.02 0.207 0.79 1.0
## O4  0.04  0.23 0.056 0.94 1.1
## O5 -0.18  0.06 0.038 0.96 1.2
##
##                        MR1  MR2
## SS loadings           3.77 2.81
## Proportion Var        0.15 0.11
## Cumulative Var        0.15 0.26
## Proportion Explained  0.57 0.43
## Cumulative Proportion 0.57 1.00
##
## Mean item complexity =  1.2
## Test of the hypothesis that 2 factors are sufficient.
##
## df null model =  300  with the objective function =  7.41 with Chi Square =  16484.7
## df of  the model are 251  and the objective function was  2.73
##
## The root mean square of the residuals (RMSR) is  0.09
## The df corrected root mean square of the residuals is  0.1
##
## The harmonic n.obs is  2236 with the empirical chi square  10195.98  with prob <  0
## The total n.obs was  2236  with Likelihood Chi Square =  6068.95  with prob <  0
```

```
##
## Tucker Lewis Index of factoring reliability =  0.57
## RMSEA index =  0.102  and the 90 % confidence intervals are  0.1 0.104
## BIC =  4133.13
## Fit based upon off diagonal values = 0.83
## Measures of factor score adequacy
##                                                    MR1  MR2
## Correlation of (regression) scores with factors   0.91 0.91
## Multiple R square of scores with factors          0.84 0.84
## Minimum correlation of possible factor scores     0.67 0.67
```

Notice that the loading matrix changes, with values moving aside either towards 0 or 1. However, the communality doesn't change much, and this result is quite natural.

As for rotation, Oblimin is the default in R, which is an oblique rotation. Varimax is the most popular orthogonal rotation. Other rotational methods include Quartimax (Orthogonal, each variable accounted by minimal number of factors, 简化对变量的解释), Equamax (Orthogonal), Promax (oblique, this one is fast!). Oblique rotation yields correlated factors.

```
fa_otho$loadings  # show the loadings after rotation
```

```
##
## Loadings:
##    MR1    MR2
## A1 -0.189
## A2  0.521
## A3  0.593
## A4  0.401 -0.104
## A5  0.577 -0.139
## C1  0.327
## C2  0.342
## C3  0.289 -0.120
## C4 -0.319  0.317
```

```
## C5 -0.332  0.359
## E1 -0.451
## E2 -0.554  0.265
## E3  0.639
## E4  0.608 -0.157
## E5  0.597
## N1         0.756
## N2         0.736
## N3         0.750
## N4 -0.248  0.619
## N5         0.533
## O1  0.356
## O2 -0.115  0.167
## O3  0.455
## O4         0.233
## O5 -0.183
##
##                  MR1   MR2
## SS loadings    3.771 2.814
## Proportion Var 0.151 0.113
## Cumulative Var 0.151 0.263
```

```
fa_otho$rot.mat  # show the rotation matrix
```

```
##           [,1]       [,2]
## [1,] 0.8470713 -0.5314793
## [2,] 0.5314793  0.8470713
```

```
fa_oblique = fa(currentData,2,fm='minres',rotate='promax')
print(fa_oblique)
```

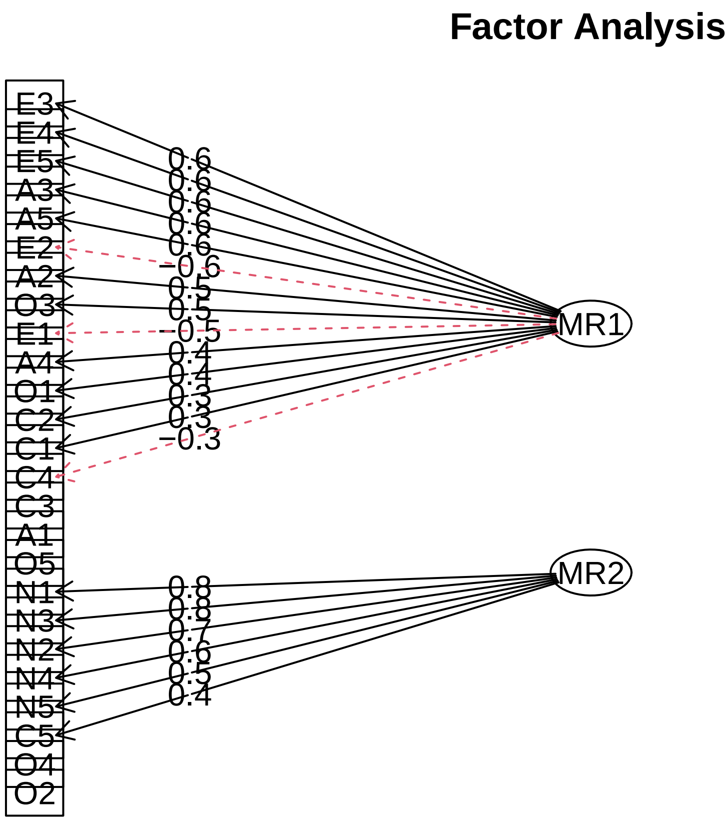
```
## Factor Analysis using method =  minres
## Call: fa(r = currentData, nfactors = 2, rotate = "promax", fm = "minres")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      MR1   MR2   h2   u2 com
```

```
## A1 -0.18  0.06 0.044 0.96 1.2
## A2  0.55  0.11 0.272 0.73 1.1
## A3  0.62  0.10 0.352 0.65 1.1
## A4  0.40 -0.03 0.171 0.83 1.0
## A5  0.58 -0.03 0.352 0.65 1.0
## C1  0.33  0.00 0.111 0.89 1.0
## C2  0.36  0.05 0.118 0.88 1.0
## C3  0.28 -0.07 0.098 0.90 1.1
## C4 -0.27  0.27 0.202 0.80 2.0
## C5 -0.28  0.31 0.240 0.76 2.0
## E1 -0.47 -0.04 0.206 0.79 1.0
## E2 -0.53  0.17 0.377 0.62 1.2
## E3  0.68  0.13 0.409 0.59 1.1
## E4  0.61 -0.04 0.394 0.61 1.0
## E5  0.63  0.10 0.357 0.64 1.0
## N1  0.12  0.79 0.572 0.43 1.0
## N2  0.12  0.77 0.542 0.46 1.0
## N3  0.14  0.79 0.563 0.44 1.1
## N4 -0.13  0.60 0.444 0.56 1.1
## N5  0.02  0.55 0.291 0.71 1.0
## O1  0.37  0.05 0.127 0.87 1.0
## O2 -0.09  0.15 0.041 0.96 1.6
## O3  0.48  0.11 0.207 0.79 1.1
## O4  0.09  0.25 0.056 0.94 1.3
## O5 -0.18  0.03 0.038 0.96 1.1
##
##                        MR1  MR2
## SS loadings           3.81 2.78
## Proportion Var        0.15 0.11
## Cumulative Var        0.15 0.26
## Proportion Explained  0.58 0.42
## Cumulative Proportion 0.58 1.00
##
```

```
##  With factor correlations of
##       MR1   MR2
## MR1  1.00 -0.37
## MR2 -0.37  1.00
##
## Mean item complexity =  1.2
## Test of the hypothesis that 2 factors are sufficient.
##
## df null model =  300  with the objective function =  7.41 with Chi Square =  16484.7
## df of  the model are 251  and the objective function was  2.73
##
## The root mean square of the residuals (RMSR) is  0.09
## The df corrected root mean square of the residuals is  0.1
##
## The harmonic n.obs is  2236 with the empirical chi square  10195.98  with prob <  0
## The total n.obs was  2236  with Likelihood Chi Square =  6068.95  with prob <  0
##
## Tucker Lewis Index of factoring reliability =  0.57
## RMSEA index =  0.102  and the 90 % confidence intervals are  0.1 0.104
## BIC =  4133.13
## Fit based upon off diagonal values = 0.83
## Measures of factor score adequacy
##                                                    MR1  MR2
## Correlation of (regression) scores with factors   0.92 0.92
## Multiple R square of scores with factors          0.85 0.85
## Minimum correlation of possible factor scores     0.70 0.70
```

Use diagram to show how variables relate to the factors.

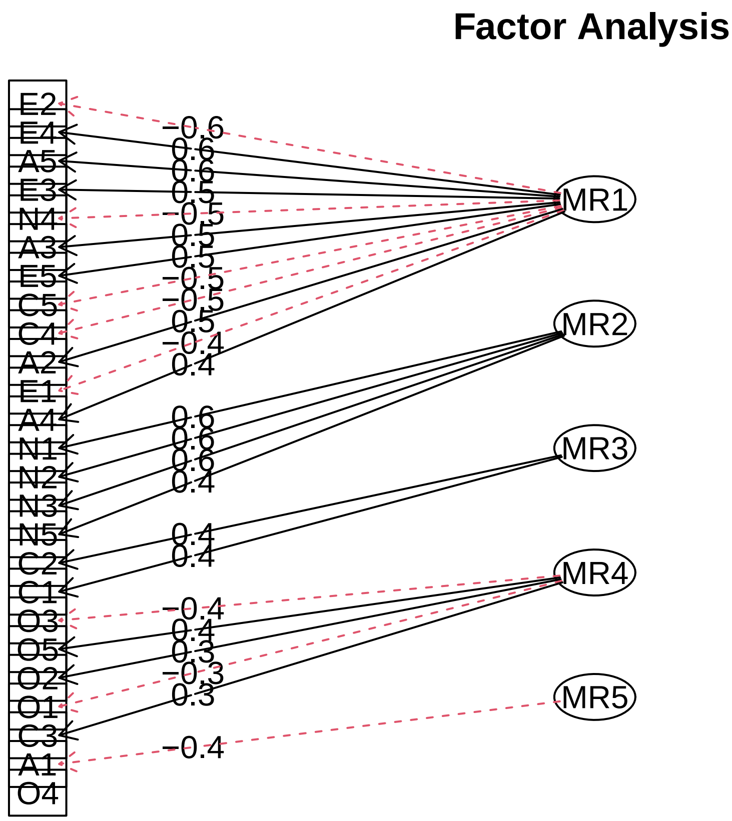
```
fa.diagram(fa_otho)
```



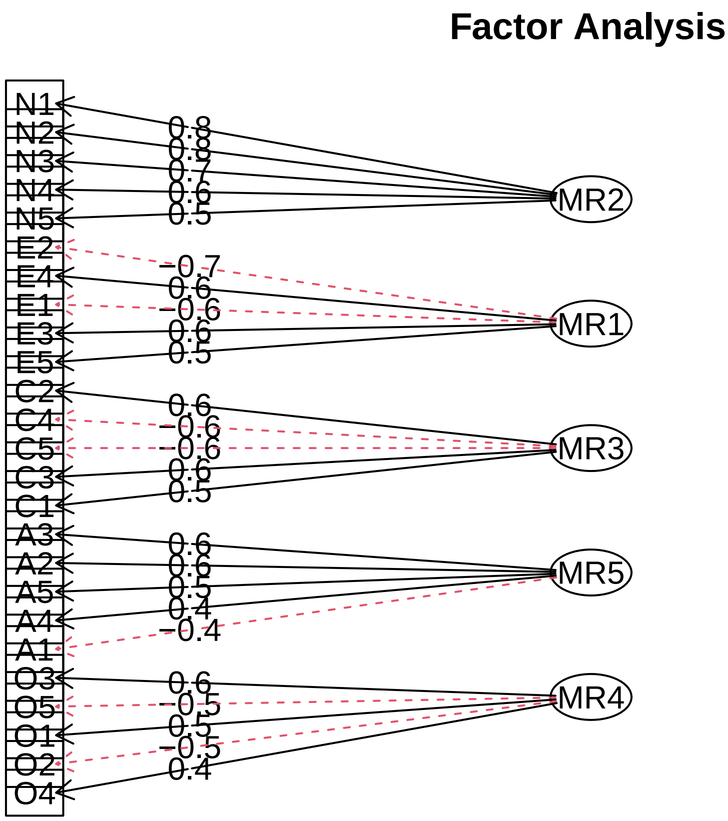
What if we use 5 factors, either rotated or not. To see whether big five really works here.

Note `fa.diagram` use r = 0.3 as a cut off, you can change it.

```
fa5_none = fa(currentData,5,fm='minres',rotate='none')
fa.diagram(fa5_none)
```

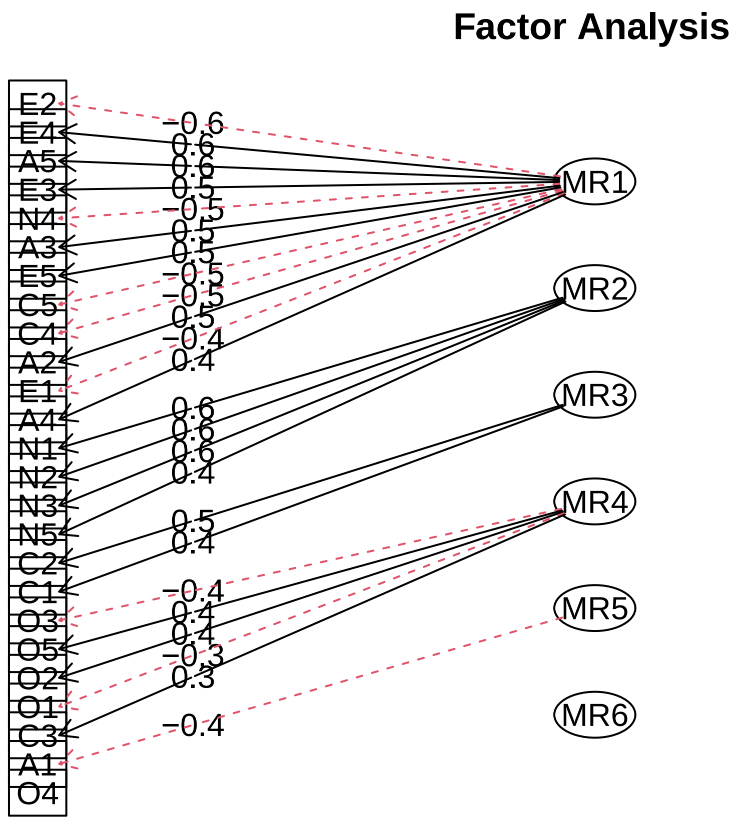



```r
fa5_otho = fa(currentData,5,fm='minres',rotate='varimax')
fa.diagram(fa5_otho)
```

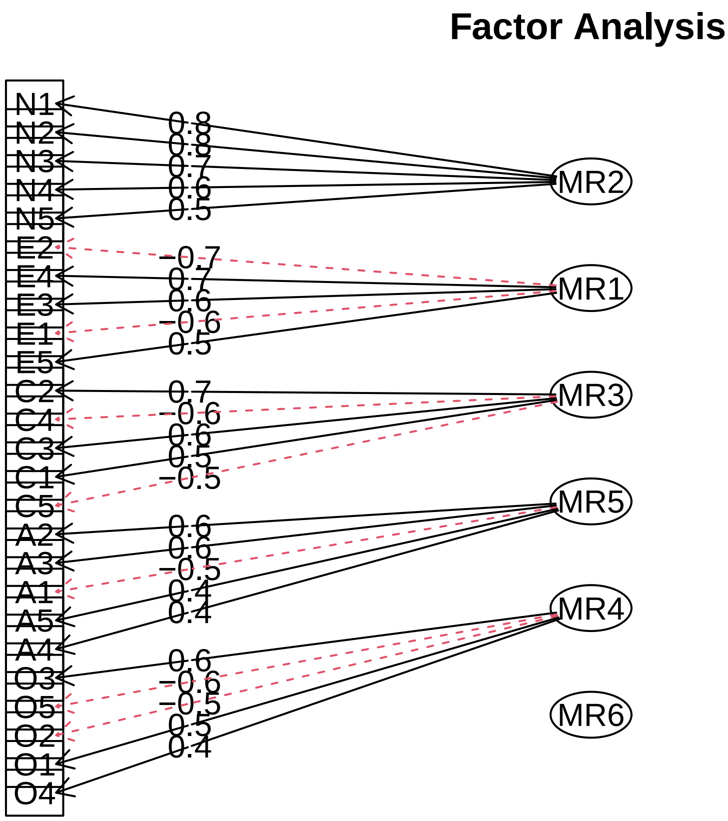


Interestingly, the 6-factor model does not work that well.

```
fa6_none = fa(currentData,6,fm='minres',rotate='none')
fa.diagram(fa6_none)
```




```
fa6_otho = fa(currentData,6,fm='minres',rotate='varimax')
fa.diagram(fa6_otho)
```



Scores, or the coefficient for the model: Factors = scores*X where X are the original data, Factors are the obtained latent variables, scores are their coefficients.

The scores here are for normalized data, which is the case for SPSS.

```
head(fa_otho$scores)
```

```
##             MR1         MR2
## 1  1.27436156 -0.05162498
## 2 -2.25649076  0.45934922
## 3 -0.09521188 -0.12571634
## 4 -0.70689909 -0.56922445
## 5 -0.18784942 -0.75916140
## 6  0.91231347  1.16628979
```

The below is for you to hand check the meaning of SS laoding, uniqueness and communality.

Communality, the variance that can be explained by factors, which should be high if the model is good.

```
# communality
apply(fa_otho$loadings^2,1,sum)
```

```
##         A1         A2         A3         A4         A5         C1         C2
## 0.04359507 0.27197456 0.35182910 0.17116362 0.35236109 0.11106610 0.11760829
##         C3         C4         C5         E1         E2         E3         E4
## 0.09774264 0.20204612 0.23964636 0.20559167 0.37742115 0.40892440 0.39419870
##         E5         N1         N2         N3         N4         N5         O1
## 0.35693755 0.57226220 0.54214564 0.56338639 0.44421810 0.29146687 0.12733686
##         O2         O3         O4         O5
## 0.04122094 0.20719293 0.05598365 0.03767935
```

```
# uniqueness
1 - apply(fa_otho$loadings^2,1,sum)
```

```
##        A1        A2        A3        A4        A5        C1        C2        C3
## 0.9564049 0.7280254 0.6481709 0.8288364 0.6476389 0.8889339 0.8823917 0.9022574
##        C4        C5        E1        E2        E3        E4        E5        N1
## 0.7979539 0.7603536 0.7944083 0.6225788 0.5910756 0.6058013 0.6430625 0.4277378
##        N2        N3        N4        N5        O1        O2        O3        O4
## 0.4578544 0.4366136 0.5557819 0.7085331 0.8726631 0.9587791 0.7928071 0.9440163
##        O5
## 0.9623207
```

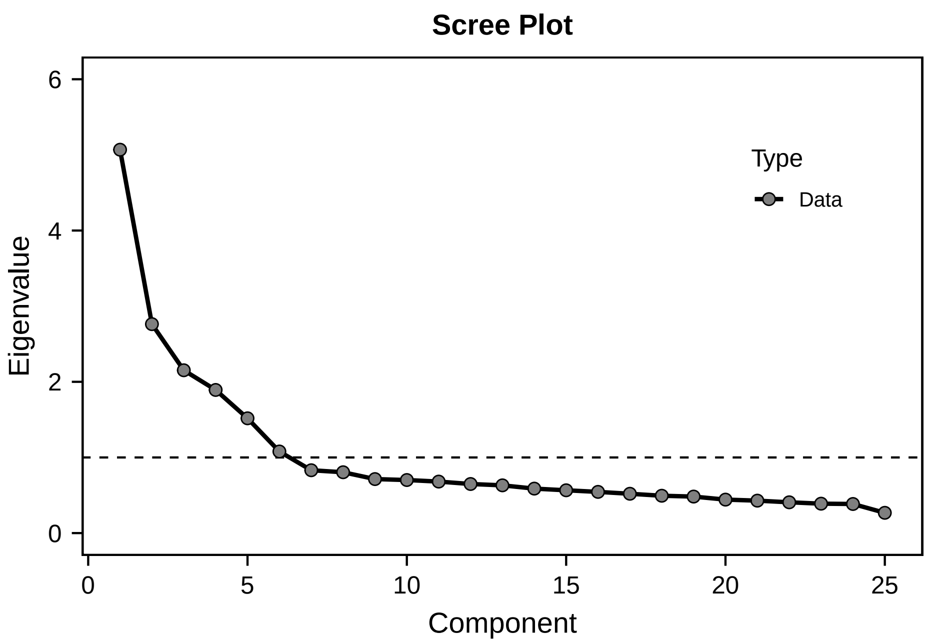
```
P = EFA(data = currentData,vars = names(currentData))
```

```
##
## Principal Component Analysis
##
## Summary:
## Total Items: 25
```

```
## Scale Range: 1 ~ 6
## Total Cases: 2236
## Valid Cases: 2236 (100.0%)
##
## Extraction Method:
## - Principal Component Analysis
## Rotation Method:
## - Varimax (with Kaiser Normalization)
##
## KMO and Bartlett's Test:
## - Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy: MSA = 0.847
## - Bartlett's Test of Sphericity: Approx.  ²(300) = 16484.78, p < 1e-99 ***
##
## Total Variance Explained:
##
##               Eigenvalue Variance % Cumulative % SS Loading Variance % Cumulative %
##
## Component 1        5.069     20.274       20.274      3.089     12.356       12.356
## Component 2        2.762     11.050       31.324      2.596     10.385       22.741
## Component 3        2.153      8.610       39.934      2.570     10.280       33.021
## Component 4        1.892      7.569       47.504      2.497      9.989       43.010
## Component 5        1.518      6.070       53.574      2.092      8.368       51.379
## Component 6        1.079      4.315       57.889      1.628      6.511       57.889
## Component 7        0.831      3.324       61.213
## Component 8        0.805      3.218       64.431
## Component 9        0.714      2.856       67.287
## Component 10       0.702      2.806       70.093
## Component 11       0.681      2.723       72.817
## Component 12       0.649      2.596       75.413
## Component 13       0.631      2.525       77.938
## Component 14       0.588      2.352       80.290
## Component 15       0.566      2.264       82.554
## Component 16       0.545      2.179       84.733
```

```
## Component 17       0.520      2.080       86.813
## Component 18       0.494      1.975       88.788
## Component 19       0.483      1.931       90.719
## Component 20       0.443      1.770       92.489
## Component 21       0.429      1.715       94.205
## Component 22       0.407      1.628       95.833
## Component 23       0.389      1.556       97.389
## Component 24       0.385      1.539       98.928
## Component 25       0.268      1.072      100.000
##
##
## Component Loadings (Rotated) (Sorted by Size):
##
##        RC2    RC5    RC3    RC6    RC1    RC4 Communality
##
## N1   0.836 -0.158 -0.041 -0.098 -0.026  0.065       0.741
## N2   0.835 -0.149 -0.021 -0.078 -0.033 -0.038       0.728
## N3   0.794 -0.005 -0.049  0.088  0.043  0.046       0.644
## N4   0.609 -0.027 -0.169  0.432  0.064 -0.002       0.591
## N5   0.609  0.147 -0.033  0.211 -0.096  0.188       0.483
## A2   0.058  0.740  0.117 -0.156  0.026 -0.112       0.602
## A3  -0.023  0.708  0.089 -0.184  0.224  0.079       0.600
## A1   0.072 -0.636  0.092 -0.064  0.229  0.436       0.664
## A5  -0.185  0.594  0.061 -0.227  0.322  0.165       0.574
## A4  -0.088  0.556  0.239 -0.111 -0.003  0.201       0.427
## C2   0.075  0.103  0.730  0.072  0.192  0.059       0.595
## C4   0.193 -0.072 -0.686  0.182  0.152  0.314       0.668
## C3  -0.032  0.125  0.678  0.005  0.003  0.045       0.478
## C1   0.014  0.006  0.650  0.014  0.253 -0.069       0.491
## C5   0.280 -0.067 -0.625  0.240  0.087  0.030       0.539
## E2   0.209 -0.192 -0.071  0.724 -0.157  0.033       0.636
## E1  -0.054 -0.211  0.092  0.720 -0.054  0.166       0.605
## E4  -0.148  0.388  0.080 -0.559  0.258  0.262       0.626
```

```
## E5   0.072  0.127  0.334 -0.518  0.332 -0.070       0.517
## O4   0.201  0.157 -0.047  0.434  0.405 -0.243       0.478
## O1  -0.070 -0.010  0.124 -0.049  0.683 -0.199       0.529
## O3   0.001  0.094  0.057 -0.174  0.669 -0.307       0.585
## E3  -0.016  0.274  0.071 -0.416  0.586  0.069       0.601
## O5   0.042 -0.017 -0.043  0.015 -0.261  0.699       0.561
## O2   0.164  0.134 -0.105  0.040 -0.204  0.641       0.510
##
## Communality = Sum of Squared (SS) Factor Loadings
## (Uniqueness = 1 - Communality)
```




```
P$eigenvalues
```

```
##             Eigenvalue Variance % Cumulative % SS Loading Variance %
## Component 1  5.0685162  20.274065     20.27406   3.089101  12.356404
## Component 2  2.7624793  11.049917     31.32398   2.596271  10.385084
## Component 3  2.1526230   8.610492     39.93447   2.569883  10.279533
## Component 4  1.8923330   7.569332     47.50381   2.497253   9.989012
## Component 5  1.5175329   6.070132     53.57394   2.092123   8.368493
```

```
## Component 6   1.0788293  4.315317   57.88925  1.627682  6.510729
## Component 7   0.8309057  3.323623   61.21288        NA        NA
## Component 8   0.8045002  3.218001   64.43088        NA        NA
## Component 9   0.7140883  2.856353   67.28723        NA        NA
## Component 10  0.7015381  2.806152   70.09338        NA        NA
## Component 11  0.6808421  2.723368   72.81675        NA        NA
## Component 12  0.6489735  2.595894   75.41265        NA        NA
## Component 13  0.6312563  2.525025   77.93767        NA        NA
## Component 14  0.5880320  2.352128   80.28980        NA        NA
## Component 15  0.5659652  2.263861   82.55366        NA        NA
## Component 16  0.5448396  2.179358   84.73302        NA        NA
## Component 17  0.5199335  2.079734   86.81275        NA        NA
## Component 18  0.4938686  1.975474   88.78823        NA        NA
## Component 19  0.4827362  1.930945   90.71917        NA        NA
## Component 20  0.4425003  1.770001   92.48917        NA        NA
## Component 21  0.4288706  1.715483   94.20466        NA        NA
## Component 22  0.4070974  1.628390   95.83305        NA        NA
## Component 23  0.3888753  1.555501   97.38855        NA        NA
## Component 24  0.3847626  1.539050   98.92760        NA        NA
## Component 25  0.2681008  1.072403  100.00000        NA        NA
##              Cumulative %
## Component 1      12.35640
## Component 2      22.74149
## Component 3      33.02102
## Component 4      43.01003
## Component 5      51.37853
## Component 6      57.88925
## Component 7            NA
## Component 8            NA
## Component 9            NA
## Component 10           NA
## Component 11           NA
## Component 12           NA
```

```
## Component 13          NA
## Component 14          NA
## Component 15          NA
## Component 16          NA
## Component 17          NA
## Component 18          NA
## Component 19          NA
## Component 20          NA
## Component 21          NA
## Component 22          NA
## Component 23          NA
## Component 24          NA
## Component 25          NA
```

```
P$loadings
```

```
##              RC2          RC5         RC3          RC6          RC1         RC4
## N1  0.836491887 -0.157590680 -0.04108418 -0.097998494 -0.025939731  0.06465551
## N2  0.834508658 -0.149156942 -0.02064085 -0.077864197 -0.032563963 -0.03816179
## N3  0.793711537 -0.005168488 -0.04888555  0.088043491  0.042752044  0.04602475
## N4  0.609322682 -0.027472354 -0.16871044  0.431528485  0.063980877 -0.00162981
## N5  0.609055668  0.147324483 -0.03340184  0.210953824 -0.095819511  0.18800144
## A2  0.057524885  0.739900333  0.11664100 -0.156447169  0.026011852 -0.11189172
## A3 -0.022773080  0.708278364  0.08884781 -0.183650315  0.224099530  0.07914837
## A1  0.072027071 -0.635726330  0.09156661 -0.063644039  0.229021008  0.43554813
## A5 -0.185158668  0.594493141  0.06074520 -0.227027404  0.322347467  0.16497047
## A4 -0.087699627  0.556153428  0.23937838 -0.110857744 -0.003473359  0.20080148
## C2  0.074779715  0.103231141  0.73004958  0.072391718  0.192269544  0.05935875
## C4  0.192541610 -0.072229949 -0.68596718  0.181892561  0.151801380  0.31422451
## C3 -0.031689921  0.125088185  0.67775704  0.005040871  0.003201663  0.04546638
## C1  0.013508483  0.006478096  0.64950421  0.013871224  0.252676838 -0.06868019
## C5  0.279840593 -0.067415850 -0.62465545  0.240249583  0.086582100  0.03010505
## E2  0.208799741 -0.191702723 -0.07061461  0.724424287 -0.157433230  0.03295199
## E1 -0.054138742 -0.211222539  0.09164581  0.720337059 -0.053606737  0.16560687
```

```
## E4 -0.147608529  0.387783331  0.07957953 -0.559071760  0.258448370  0.26180636
## E5  0.072113100  0.126769699  0.33384919 -0.518452171  0.332493323 -0.06969270
## O4  0.200561610  0.156760944 -0.04737357  0.434084755  0.404647132 -0.24277789
## O1 -0.069524090 -0.009900085  0.12373491 -0.048918290  0.683065226 -0.19916320
## O3  0.001108861  0.093768892  0.05719735 -0.174044696  0.669355360 -0.30726619
## E3 -0.016284826  0.273872428  0.07112138 -0.415697273  0.586162005  0.06888588
## O5  0.042186541 -0.016811945 -0.04314908  0.015058710 -0.261138655  0.69879394
## O2  0.164325447  0.134322459 -0.10524777  0.039920315 -0.203755777  0.64079428
##    Communality
## N1   0.7406983
## N2   0.7276581
## N3   0.6440922
## N4   0.5908051
## N5   0.4827964
## A2   0.6020388
## A3   0.6002833
## A1   0.6639237
## A5   0.5740604
## A4   0.4269226
## C2   0.5949527
## C4   0.6677060
## C3   0.4781088
## C1   0.4908351
## C5   0.5391727
## E2   0.6359953
## E1   0.6051297
## E4   0.6263965
## E5   0.5169277
## O4   0.4781532
## O1   0.5288790
## O3   0.5848060
## E3   0.6014649
## O5   0.5606573
```

```
## O2   0.5098498
```

```r
P$result
```

```
## Principal Components Analysis
## Call: psych::principal(r = data, nfactors = nfactors, rotate = rotation)
## Standardized loadings (pattern matrix) based upon correlation matrix
##      RC2   RC5   RC3   RC6   RC1   RC4   h2   u2 com
## A1  0.07 -0.64  0.09 -0.06  0.23  0.43 0.66 0.34 2.2
## A2  0.06  0.74  0.12 -0.16  0.03 -0.11 0.60 0.40 1.2
## A3 -0.02  0.71  0.09 -0.18  0.23  0.08 0.60 0.40 1.4
## A4 -0.09  0.56  0.24 -0.11  0.00  0.20 0.43 0.57 1.8
## A5 -0.19  0.59  0.06 -0.22  0.32  0.16 0.57 0.43 2.3
## C1  0.01  0.01  0.65  0.02  0.25 -0.07 0.49 0.51 1.3
## C2  0.07  0.10  0.73  0.07  0.19  0.06 0.59 0.41 1.2
## C3 -0.03  0.12  0.68  0.01  0.00  0.05 0.48 0.52 1.1
## C4  0.19 -0.07 -0.69  0.18  0.15  0.31 0.67 0.33 1.9
## C5  0.28 -0.07 -0.62  0.24  0.09  0.03 0.54 0.46 1.8
## E1 -0.05 -0.21  0.09  0.72 -0.06  0.16 0.61 0.39 1.3
## E2  0.21 -0.19 -0.07  0.72 -0.16  0.03 0.64 0.36 1.5
## E3 -0.02  0.27  0.07 -0.41  0.59  0.07 0.60 0.40 2.3
## E4 -0.15  0.39  0.08 -0.56  0.26  0.26 0.63 0.37 3.0
## E5  0.07  0.13  0.33 -0.52  0.34 -0.07 0.52 0.48 2.8
## N1  0.84 -0.16 -0.04 -0.10 -0.03  0.06 0.74 0.26 1.1
## N2  0.83 -0.15 -0.02 -0.08 -0.03 -0.04 0.73 0.27 1.1
## N3  0.79 -0.01 -0.05  0.09  0.04  0.05 0.64 0.36 1.0
## N4  0.61 -0.03 -0.17  0.43  0.06  0.00 0.59 0.41 2.0
## N5  0.61  0.15 -0.03  0.21 -0.10  0.19 0.48 0.52 1.7
## O1 -0.07 -0.01  0.12 -0.05  0.68 -0.20 0.53 0.47 1.3
## O2  0.16  0.13 -0.11  0.04 -0.20  0.64 0.51 0.49 1.5
## O3  0.00  0.09  0.06 -0.17  0.67 -0.31 0.58 0.42 1.6
## O4  0.20  0.16 -0.05  0.44  0.40 -0.25 0.48 0.52 3.4
## O5  0.04 -0.02 -0.04  0.01 -0.26  0.70 0.56 0.44 1.3
##
```

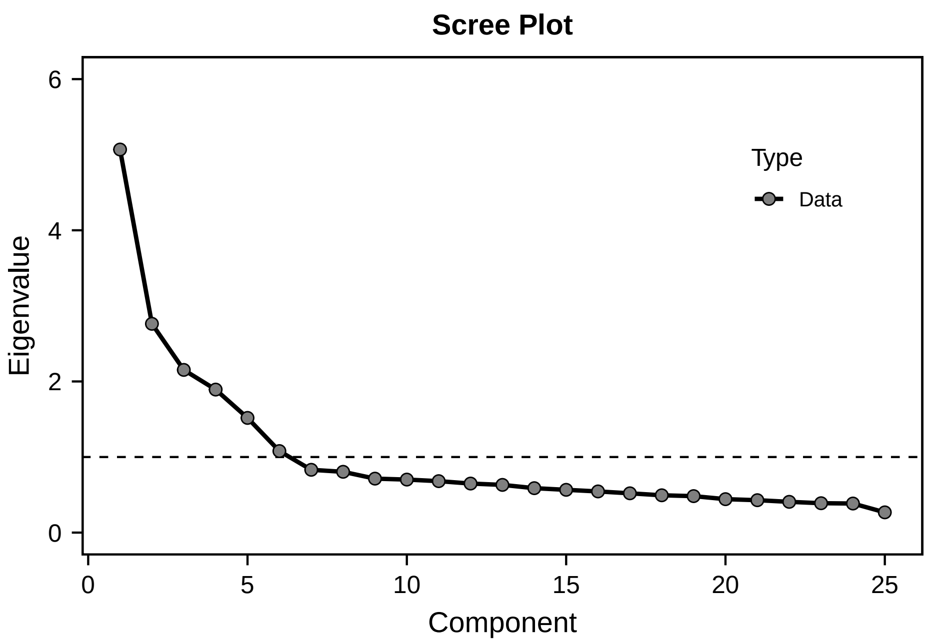
```
##                        RC2  RC5  RC3  RC6  RC1  RC4
## SS loadings           3.09 2.60 2.57 2.49 2.10 1.63
## Proportion Var        0.12 0.10 0.10 0.10 0.08 0.07
## Cumulative Var        0.12 0.23 0.33 0.43 0.51 0.58
## Proportion Explained  0.21 0.18 0.18 0.17 0.14 0.11
## Cumulative Proportion 0.21 0.39 0.57 0.74 0.89 1.00
##
## Mean item complexity =  1.7
## Test of the hypothesis that 6 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.05
##  with the empirical chi square  3727.83  with prob <  0
##
## Fit based upon off diagonal values = 0.94
```

```
P$result.kaiser
```

```
##
## Call: NULL
## Standardized loadings (pattern matrix) based upon correlation matrix
##      RC2   RC5   RC3   RC6   RC1   RC4   h2   u2
## A1  0.07 -0.64  0.09 -0.06  0.23  0.44 0.66 0.34
## A2  0.06  0.74  0.12 -0.16  0.03 -0.11 0.60 0.40
## A3 -0.02  0.71  0.09 -0.18  0.22  0.08 0.60 0.40
## A4 -0.09  0.56  0.24 -0.11  0.00  0.20 0.43 0.57
## A5 -0.19  0.59  0.06 -0.23  0.32  0.16 0.57 0.43
## C1  0.01  0.01  0.65  0.01  0.25 -0.07 0.49 0.51
## C2  0.07  0.10  0.73  0.07  0.19  0.06 0.59 0.41
## C3 -0.03  0.13  0.68  0.01  0.00  0.05 0.48 0.52
## C4  0.19 -0.07 -0.69  0.18  0.15  0.31 0.67 0.33
## C5  0.28 -0.07 -0.62  0.24  0.09  0.03 0.54 0.46
## E1 -0.05 -0.21  0.09  0.72 -0.05  0.17 0.61 0.39
## E2  0.21 -0.19 -0.07  0.72 -0.16  0.03 0.64 0.36
## E3 -0.02  0.27  0.07 -0.42  0.59  0.07 0.60 0.40
```

```
## E4 -0.15  0.39  0.08 -0.56  0.26  0.26 0.63 0.37
## E5  0.07  0.13  0.33 -0.52  0.33 -0.07 0.52 0.48
## N1  0.84 -0.16 -0.04 -0.10 -0.03  0.06 0.74 0.26
## N2  0.83 -0.15 -0.02 -0.08 -0.03 -0.04 0.73 0.27
## N3  0.79 -0.01 -0.05  0.09  0.04  0.05 0.64 0.36
## N4  0.61 -0.03 -0.17  0.43  0.06  0.00 0.59 0.41
## N5  0.61  0.15 -0.03  0.21 -0.10  0.19 0.48 0.52
## O1 -0.07 -0.01  0.12 -0.05  0.68 -0.20 0.53 0.47
## O2  0.16  0.13 -0.11  0.04 -0.20  0.64 0.51 0.49
## O3  0.00  0.09  0.06 -0.17  0.67 -0.31 0.58 0.42
## O4  0.20  0.16 -0.05  0.43  0.40 -0.24 0.48 0.52
## O5  0.04 -0.02 -0.04  0.02 -0.26  0.70 0.56 0.44
##
##                        RC2  RC5  RC3  RC6  RC1  RC4
## SS loadings           3.09 2.60 2.57 2.50 2.09 1.63
## Proportion Var        0.12 0.10 0.10 0.10 0.08 0.07
## Cumulative Var        0.12 0.23 0.33 0.43 0.51 0.58
## Proportion Explained  0.21 0.18 0.18 0.17 0.14 0.11
## Cumulative Proportion 0.21 0.39 0.57 0.74 0.89 1.00
```

```
P$scree.plot
```



4 Appendix

4.1 `psych::describe()`

5 容易踩的坑

PCA / EFA 最常见的几个误区

1. **不标准化就跑 PCA**：与聚类同理——方差大的变量会主导第一主成分。除非所有变量天然同量纲（如百分比、Z 分数），都应该先 `scale()` 或者用相关矩阵代替协方差矩阵跑 PCA（`prcomp(x, scale. = TRUE)`）。
2. **把 PCA 当 EFA 用**：用 `princomp` 或 `prcomp` 算出的「载荷」叫主成分载荷，本质是特征向量；EFA 的载荷是因子载荷，会随旋转方法变化。两者经过同一份数据可能数值很接近，但解释起来**完全不同**：主成分是「数据 → 成分」的投影，因子是「因子 →

数据」的反向解释。

3. **Kaiser 准则保留所有 $\lambda > 1$ 成分**：Kaiser ($\lambda > 1$) 经常会保留过多成分，特别是变量数 $p > 30$ 时。优先信赖**平行分析**（parallel analysis）和**碎石图**的肘部，Kaiser 当作下限参考。
4. **斜交旋转后看模式矩阵 / 结构矩阵搞混**：斜交旋转（如 promax / oblimin）后会产生两个载荷矩阵：**模式矩阵**（pattern matrix，回归系数）报告每个变量在每个因子上的「独特贡献」；**结构矩阵**（structure matrix，相关系数）报告每个变量和因子的「总相关」。解释因子用模式矩阵（剔除因子间相关之后的纯净贡献），别看错。正交旋转（varimax）两个矩阵相同。