HW 2 Data Mining PCA and Factor Analysis

Anushka Tak Chandan Manjunath

February 8, 2019

Contribution of each student

Anushka Tak 50%

Chandan Manjunath 50%

Signature

Student 1 : Anushka

Student 2 : Chandan

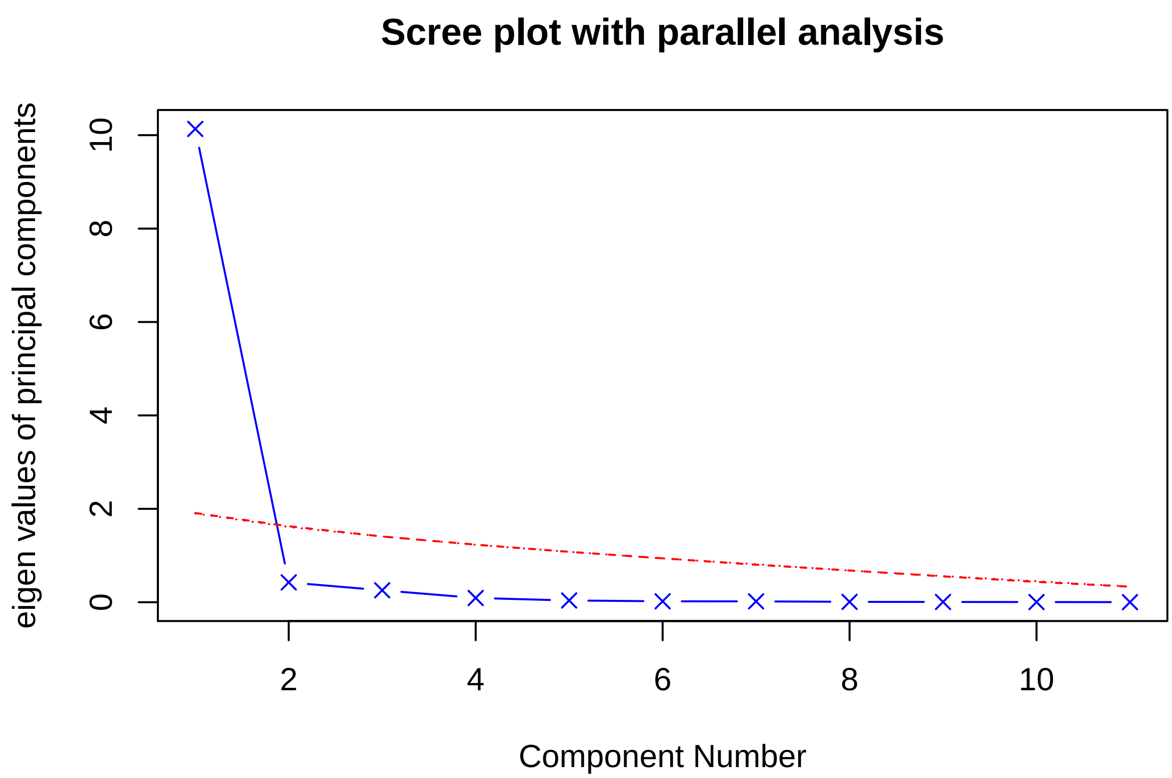
```
#install.packages("psych")
#install.packages("GPArotation")
#install.packages("readxl")

library(psych)
library(GPArotation)
library(readxl)
setwd("C:/Users/pc/Desktop/Spring2019/DM/hw2")
#View(USJudgeRatings)
```

TASK2

Problem 1

a) Determining number of components to extract Removing the CONT variable

```
fa.parallel(USJudgeRatings[,-1],fa="pc",n.iter=100,show.legend = FALSE,
           main="Scree plot with parallel analysis")
```




```
## Parallel analysis suggests that the number of factors =  NA  and the number of components =  1
```

```
#?fa.parallel()
```

b) Extracting components

```
pc<- principal(USJudgeRatings[,-1],nfactors=1)
pc
```

```
## Principal Components Analysis
## Call: principal(r = USJudgeRatings[, -1], nfactors = 1)
## Standardized loadings (pattern matrix) based upon correlation matrix
##       PC1   h2     u2 com
## INTG 0.92 0.84 0.1565   1
## DMNR 0.91 0.83 0.1663   1
## DILG 0.97 0.94 0.0613   1
## CFMG 0.96 0.93 0.0720   1
## DECI 0.96 0.92 0.0763   1
## PREP 0.98 0.97 0.0299   1
## FAMI 0.98 0.95 0.0469   1
## ORAL 1.00 0.99 0.0091   1
## WRIT 0.99 0.98 0.0196   1
## PHYS 0.89 0.80 0.2013   1
## RTEN 0.99 0.97 0.0275   1
```

```
##
##                    PC1
## SS loadings    10.13
## Proportion Var  0.92
##
## Mean item complexity =  1
## Test of the hypothesis that 1 component is sufficient.
##
## The root mean square of the residuals (RMSR) is  0.04
##  with the empirical chi square  6.21  with prob <  1
##
## Fit based upon off diagonal values = 1
```

c) Rotating components

```r
rc<-principal(USJudgeRatings[,-1], nfactors = 1,rotate="varimax")
#?principal()
rc
```

```
## Principal Components Analysis
## Call: principal(r = USJudgeRatings[, -1], nfactors = 1, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##       PC1   h2     u2 com
## INTG 0.92 0.84 0.1565   1
## DMNR 0.91 0.83 0.1663   1
## DILG 0.97 0.94 0.0613   1
## CFMG 0.96 0.93 0.0720   1
## DECI 0.96 0.92 0.0763   1
## PREP 0.98 0.97 0.0299   1
## FAMI 0.98 0.95 0.0469   1
## ORAL 1.00 0.99 0.0091   1
## WRIT 0.99 0.98 0.0196   1
## PHYS 0.89 0.80 0.2013   1
## RTEN 0.99 0.97 0.0275   1
##
##                    PC1
## SS loadings    10.13
## Proportion Var  0.92
##
## Mean item complexity =  1
## Test of the hypothesis that 1 component is sufficient.
##
## The root mean square of the residuals (RMSR) is  0.04
##  with the empirical chi square  6.21  with prob <  1
##
## Fit based upon off diagonal values = 1
```

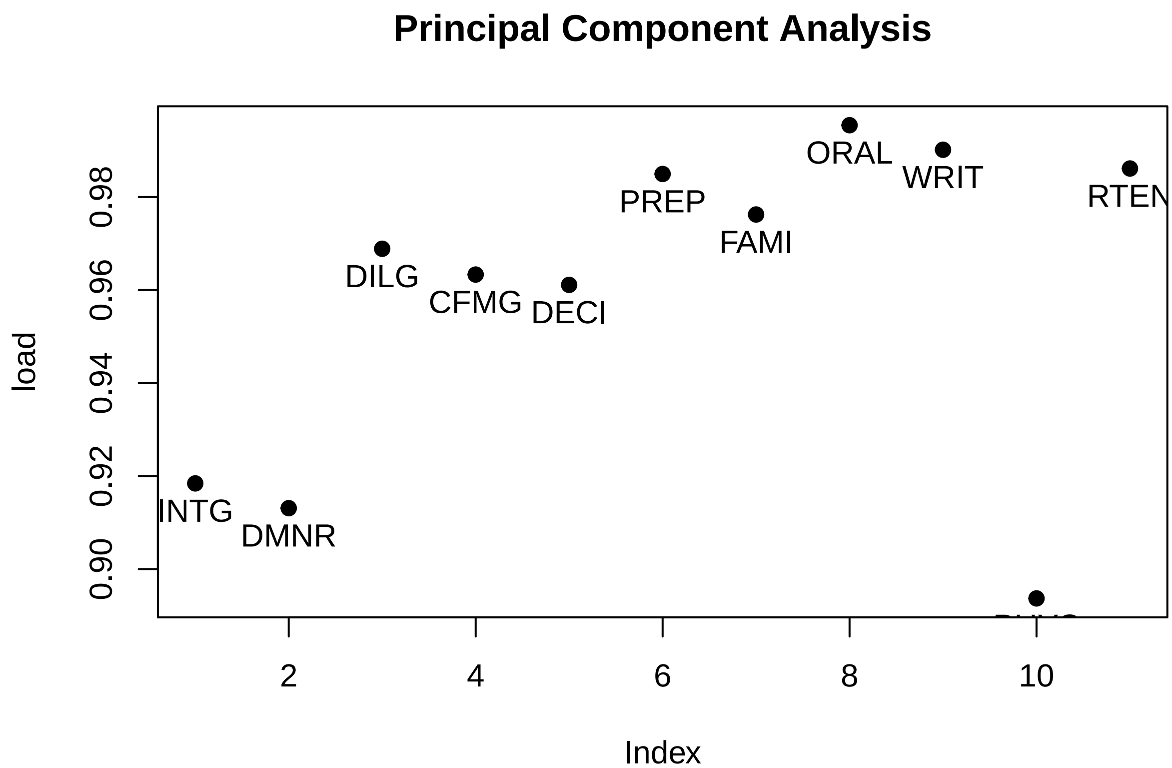
d) Computing Component scores

```r
pc<-principal(USJudgeRatings[,-1], nfactors = 1,scores=TRUE,rotate="varimax")
head(pc$scores)
```

```
##                         PC1
## AARONSON,L.H.  -0.1857981
## ALEXANDER,J.M.  0.7469865
## ARMENTANO,A.J.  0.0704772
## BERDON,R.I.     1.1358765
## BRACKEN,J.J.   -2.1586211
## BURNS,E.B.      0.7669406
```

e) Graphing an orthogonal solution

```r
#?factor.plot()
factor.plot(rc, labels = rownames(rc$loadings))
```



f) Interpretation

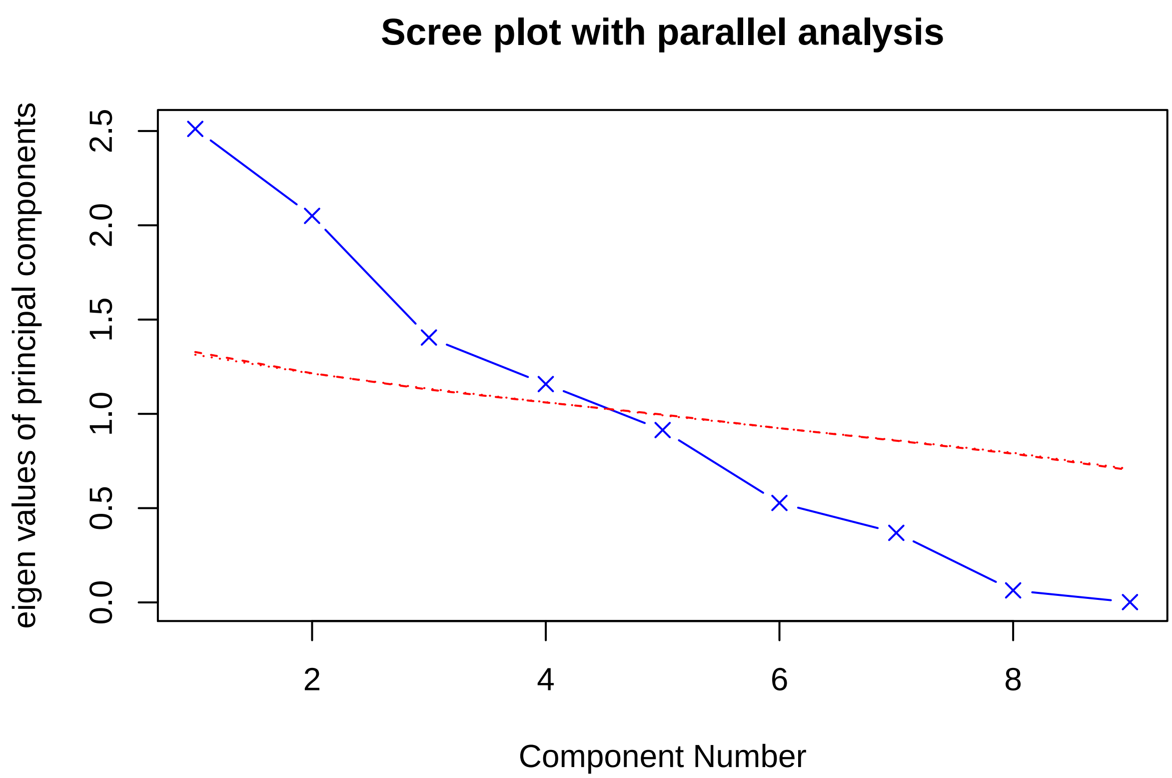
This dataset has has several attributes like demeanor,diligence, integrity, preparation to rate a US Judge. The dataset has 43 instances and has 12 attributes out of which we remove CONT , narrowing it down to 112 attributes. Running the fa.parallel function we find out that there is 1 principal component detected in the dataset and building a factor plot on it gives us the above correlation matrix diagram, which gives the domain expert a deep insight on how the components are classified and what varaibles are grouped together, this will further help us better understand how a Judge is rated.

Through this exercise we reduce a lot of correlated parameters into a few unrelated significant ones and conduct PCA in a very efficient manner.

Problem 2

a) Computing number of components to extract Removing ID variable and Class Variable

```
data2 <- read_excel("C:/Users/pc/Desktop/Spring2019/DM/hw2/Glass_Identification_Data.xlsx")

data2<-data2[-11]
fa.parallel(data2[,-1],fa="pc",n.iter=100,show.legend = FALSE,
            main="Scree plot with parallel analysis")
```




```
## Parallel analysis suggests that the number of factors =  NA  and the number of components =  4
```

```
#?fa.parallel()
```

b) Extracting components

```
pc2<- principal(data2[,-1],nfactors=4)
pc2
```

```
## Principal Components Analysis
## Call: principal(r = data2[, -1], nfactors = 4)
## Standardized loadings (pattern matrix) based upon correlation matrix
##       RC1   RC2   RC3   RC4   h2    u2 com
```

```
## RI  0.84 -0.07  0.15  0.47 0.95 0.051 1.7
## Na -0.06  0.22 -0.86  0.09 0.80 0.195 1.2
## Mg -0.35 -0.86  0.04  0.21 0.92 0.081 1.5
## Al -0.42  0.80  0.03  0.01 0.81 0.186 1.5
## Si -0.13  0.00 -0.02 -0.98 0.97 0.031 1.0
## K  -0.62  0.22  0.51  0.30 0.79 0.212 2.7
## CA  0.91  0.12  0.30  0.06 0.94 0.058 1.3
## Ba -0.01  0.72 -0.33  0.17 0.67 0.333 1.5
## Fe  0.12 -0.04  0.50  0.07 0.27 0.730 1.2
##
##                        RC1  RC2  RC3  RC4
## SS loadings           2.26 2.03 1.48 1.36
## Proportion Var        0.25 0.23 0.16 0.15
## Cumulative Var        0.25 0.48 0.64 0.79
## Proportion Explained  0.32 0.28 0.21 0.19
## Cumulative Proportion 0.32 0.60 0.81 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.08
##  with the empirical chi square  102.53  with prob <  7.4e-20
##
## Fit based upon off diagonal values = 0.92
```

c) Rotating components

```r
rc2<-principal(data2[,-1], nfactors = 4,rotate="varimax")
#?principal()
rc2
```

```
## Principal Components Analysis
## Call: principal(r = data2[, -1], nfactors = 4, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      RC1   RC2   RC3   RC4   h2    u2 com
## RI  0.84 -0.07  0.15  0.47 0.95 0.051 1.7
## Na -0.06  0.22 -0.86  0.09 0.80 0.195 1.2
## Mg -0.35 -0.86  0.04  0.21 0.92 0.081 1.5
## Al -0.42  0.80  0.03  0.01 0.81 0.186 1.5
## Si -0.13  0.00 -0.02 -0.98 0.97 0.031 1.0
## K  -0.62  0.22  0.51  0.30 0.79 0.212 2.7
## CA  0.91  0.12  0.30  0.06 0.94 0.058 1.3
## Ba -0.01  0.72 -0.33  0.17 0.67 0.333 1.5
## Fe  0.12 -0.04  0.50  0.07 0.27 0.730 1.2
##
##                        RC1  RC2  RC3  RC4
## SS loadings           2.26 2.03 1.48 1.36
## Proportion Var        0.25 0.23 0.16 0.15
## Cumulative Var        0.25 0.48 0.64 0.79
## Proportion Explained  0.32 0.28 0.21 0.19
## Cumulative Proportion 0.32 0.60 0.81 1.00
##
## Mean item complexity =  1.5
```

```
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.08
##  with the empirical chi square  102.53  with prob <  7.4e-20
##
## Fit based upon off diagonal values = 0.92
```

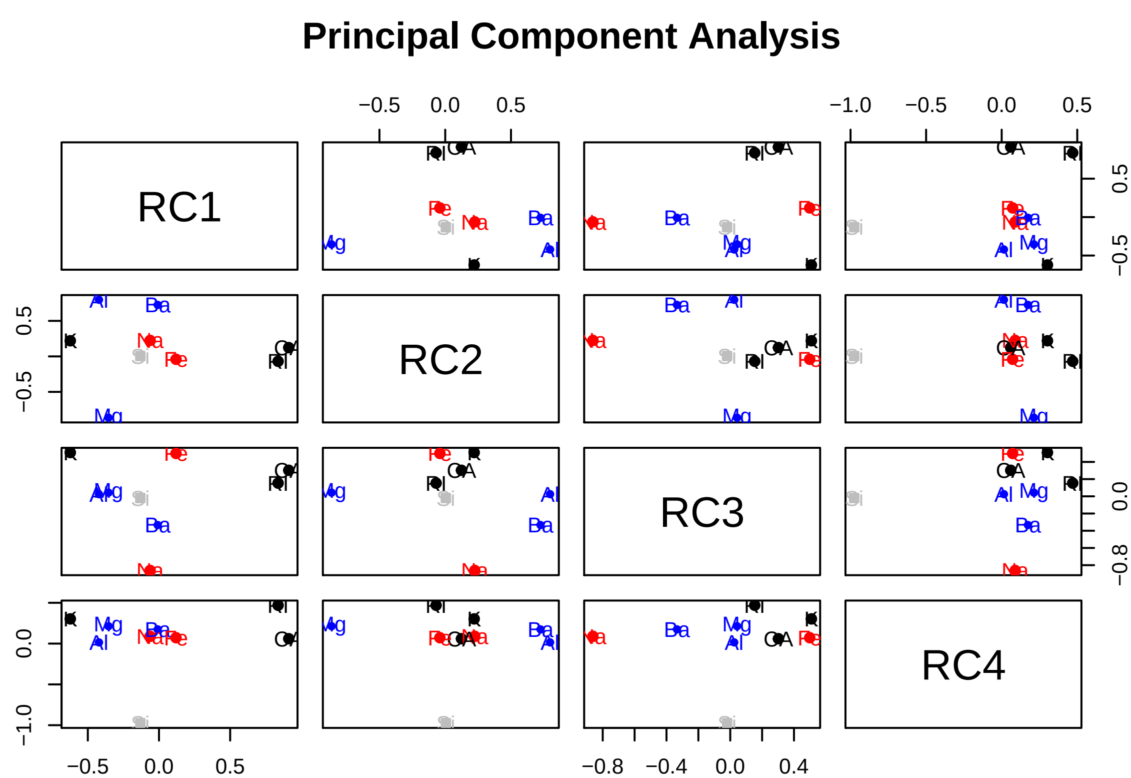
d)Computing Component scores

```r
pc2<-principal(data2[,-1], nfactors = 4,scores=TRUE,rotate="varimax")
head(pc2$scores)
```

```
##             RC1        RC2        RC3         RC4
## [1,]  0.2516834 -1.1257154 -0.8331376  1.14203433
## [2,] -0.5120556 -0.5823124 -0.7217195  0.07184681
## [3,] -0.6811108 -0.4417522 -0.4610237 -0.39146231
## [4,] -0.4363986 -0.6266048 -0.1520952  0.09532063
## [5,] -0.4446499 -0.6485935 -0.1947898 -0.37616223
## [6,] -0.7149524 -0.2237372  1.1926990 -0.41874608
```

e)Graphing orthogonal solution

```r
#?factor.plot()
factor.plot(rc2, labels = rownames(rc2$loadings))
```



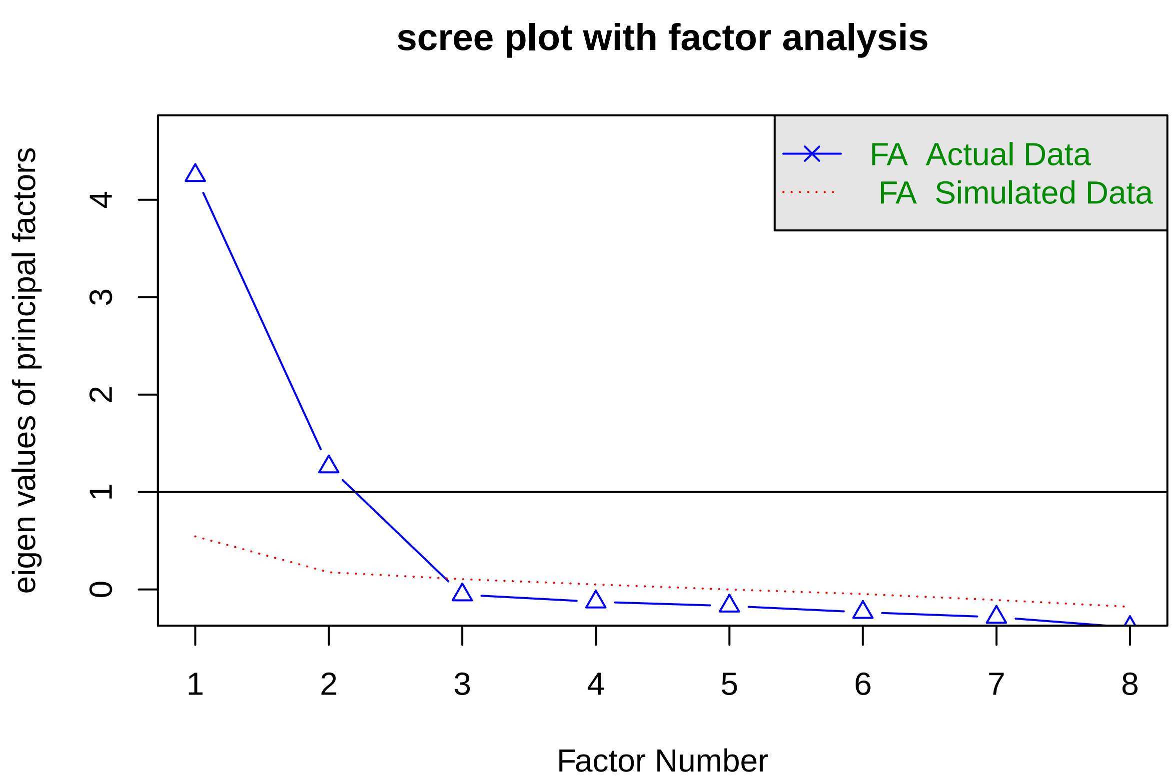
f) Interpretations This dataset studies the classification of types of glass and was motivated by criminological investigation. At the scene of the crime, the glass left can be used as evidence if it is correctly identified. The dataset has 214 instances and has 10 attributes out of which we remove ID and class , narrowing it down to 8 attributes. Running the fa.parallel function we find out that there 4 principal components detected in the dataset and building a factor plot on it gives us the above correlation matrix diagram, which gives the domain expert a deep insight on how the components are classified and what variables are grouped together, this will further help the criminology investigators in finding an insight on their evidence.

Problem3

Performing Factor Analysis

a)Determining number of components to extract

```
fa.parallel(Harman23.cor$cov, n.obs=Harman23.cor$n.obs, fa="fa", n.iter=100,
            main="scree plot with factor analysis")
```




```
## Parallel analysis suggests that the number of factors =  2  and the number of components =  NA
```

b) extract the components

```r
fa <- fa(Harman23.cor$cov, nfactors= 2, rotate = "none")
fa
```

```
## Factor Analysis using method =  minres
## Call: fa(r = Harman23.cor$cov, nfactors = 2, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                 MR1   MR2   h2   u2 com
## height         0.86 -0.32 0.84 0.16 1.3
## arm.span       0.85 -0.41 0.89 0.11 1.4
## forearm        0.81 -0.41 0.82 0.18 1.5
## lower.leg      0.83 -0.34 0.81 0.19 1.3
## weight         0.75  0.57 0.89 0.11 1.9
## bitro.diameter 0.63  0.49 0.64 0.36 1.9
## chest.girth    0.57  0.51 0.58 0.42 2.0
## chest.width    0.61  0.35 0.49 0.51 1.6
##
##                        MR1  MR2
## SS loadings           4.45 1.51
## Proportion Var        0.56 0.19
## Cumulative Var        0.56 0.74
## Proportion Explained  0.75 0.25
## Cumulative Proportion 0.75 1.00
##
## Mean item complexity =  1.6
## Test of the hypothesis that 2 factors are sufficient.
##
## The degrees of freedom for the null model are  28  and the objective function was  6.94
## The degrees of freedom for the model are 13  and the objective function was  0.26
##
## The root mean square of the residuals (RMSR) is  0.02
## The df corrected root mean square of the residuals is  0.03
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##                                                    MR1  MR2
## Correlation of (regression) scores with factors   0.98 0.94
## Multiple R square of scores with factors          0.96 0.89
## Minimum correlation of possible factor scores     0.93 0.78
```

C) Factor extraction with orthogonal and oblique rotation

```r
fa.varimax <- fa(Harman23.cor$cov, nfactors=2, rotate="varimax")
fa.varimax
```

```
## Factor Analysis using method =  minres
## Call: fa(r = Harman23.cor$cov, nfactors = 2, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                 MR1  MR2   h2   u2 com
## height         0.87 0.29 0.84 0.16 1.2
## arm.span       0.92 0.21 0.89 0.11 1.1
## forearm        0.89 0.19 0.82 0.18 1.1
## lower.leg      0.86 0.26 0.81 0.19 1.2
```

```
## weight         0.23 0.92 0.89 0.11 1.1
## bitro.diameter 0.18 0.78 0.64 0.36 1.1
## chest.girth    0.12 0.75 0.58 0.42 1.1
## chest.width    0.25 0.65 0.49 0.51 1.3
##
##                        MR1  MR2
## SS loadings           3.29 2.67
## Proportion Var        0.41 0.33
## Cumulative Var        0.41 0.74
## Proportion Explained  0.55 0.45
## Cumulative Proportion 0.55 1.00
##
## Mean item complexity =  1.1
## Test of the hypothesis that 2 factors are sufficient.
##
## The degrees of freedom for the null model are  28  and the objective function was  6.94
## The degrees of freedom for the model are 13  and the objective function was  0.26
##
## The root mean square of the residuals (RMSR) is  0.02
## The df corrected root mean square of the residuals is  0.03
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##                                                  MR1  MR2
## Correlation of (regression) scores with factors   0.97 0.95
## Multiple R square of scores with factors          0.94 0.91
## Minimum correlation of possible factor scores     0.88 0.82
```

```
fa.promax <- fa(Harman23.cor$cov, nfactors=2, rotate="promax")
fa.promax
```

```
## Factor Analysis using method =  minres
## Call: fa(r = Harman23.cor$cov, nfactors = 2, rotate = "promax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                  MR1   MR2   h2   u2 com
## height          0.88  0.06 0.84 0.16   1
## arm.span        0.96 -0.03 0.89 0.11   1
## forearm         0.93 -0.05 0.82 0.18   1
## lower.leg       0.88  0.03 0.81 0.19   1
## weight          0.00  0.94 0.89 0.11   1
## bitro.diameter -0.01  0.81 0.64 0.36   1
## chest.girth    -0.07  0.79 0.58 0.42   1
## chest.width     0.10  0.65 0.49 0.51   1
##
##                        MR1  MR2
## SS loadings           3.36 2.60
## Proportion Var        0.42 0.33
## Cumulative Var        0.42 0.74
## Proportion Explained  0.56 0.44
## Cumulative Proportion 0.56 1.00
##
##  With factor correlations of
##      MR1  MR2
## MR1 1.00 0.48
```

```
## MR2 0.48 1.00
##
## Mean item complexity =  1
## Test of the hypothesis that 2 factors are sufficient.
##
## The degrees of freedom for the null model are  28  and the objective function was  6.94
## The degrees of freedom for the model are 13  and the objective function was  0.26
##
## The root mean square of the residuals (RMSR) is  0.02
## The df corrected root mean square of the residuals is  0.03
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##                                                    MR1  MR2
## Correlation of (regression) scores with factors   0.98 0.96
## Multiple R square of scores with factors          0.96 0.93
## Minimum correlation of possible factor scores     0.92 0.85
```

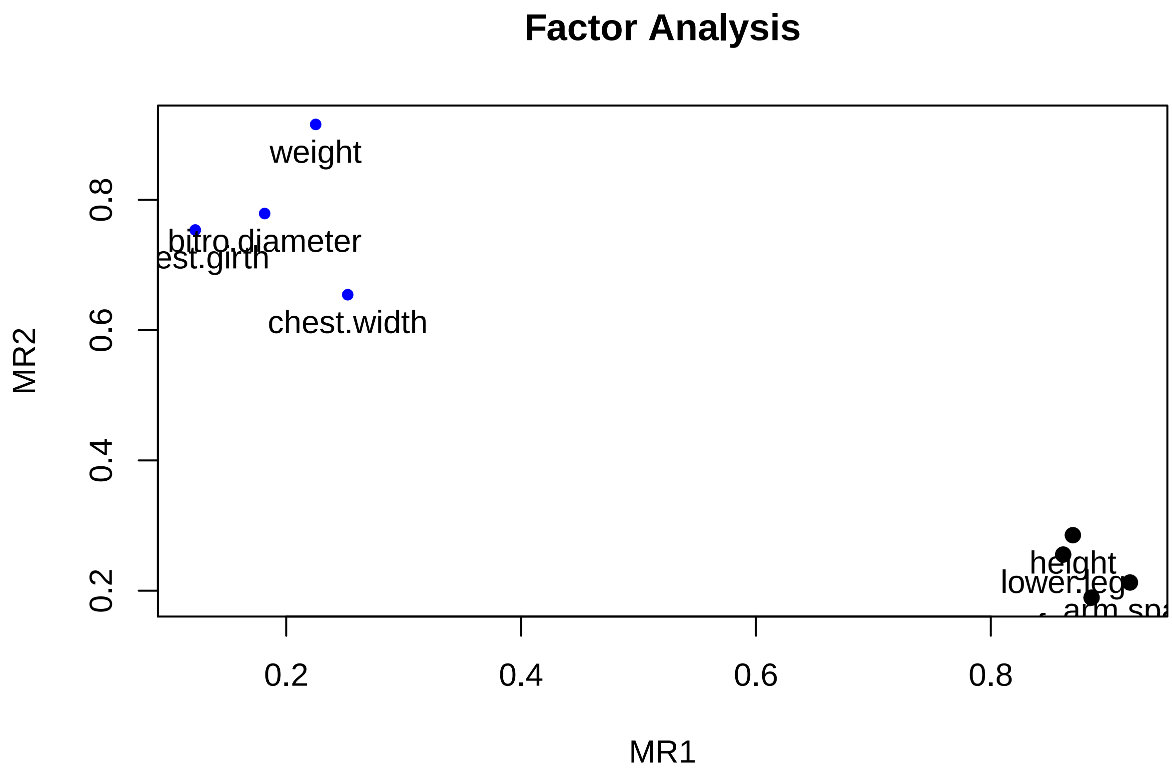
d) Factor scores

```
fa <- fa(Harman23.cor$cov, nfactors= 2, rotate = "varimax",score=TRUE)
(fa$weights)
```

```
##                       MR1         MR2
## height          0.26978933 -0.08743310
## arm.span        0.40153276  0.01394996
## forearm         0.23385047 -0.11827354
## lower.leg       0.21770758 -0.03983803
## weight         -0.15331340  0.71035858
## bitro.diameter -0.04193297  0.17190329
## chest.girth    -0.04161745  0.13432201
## chest.width    -0.03442273  0.11896061
```

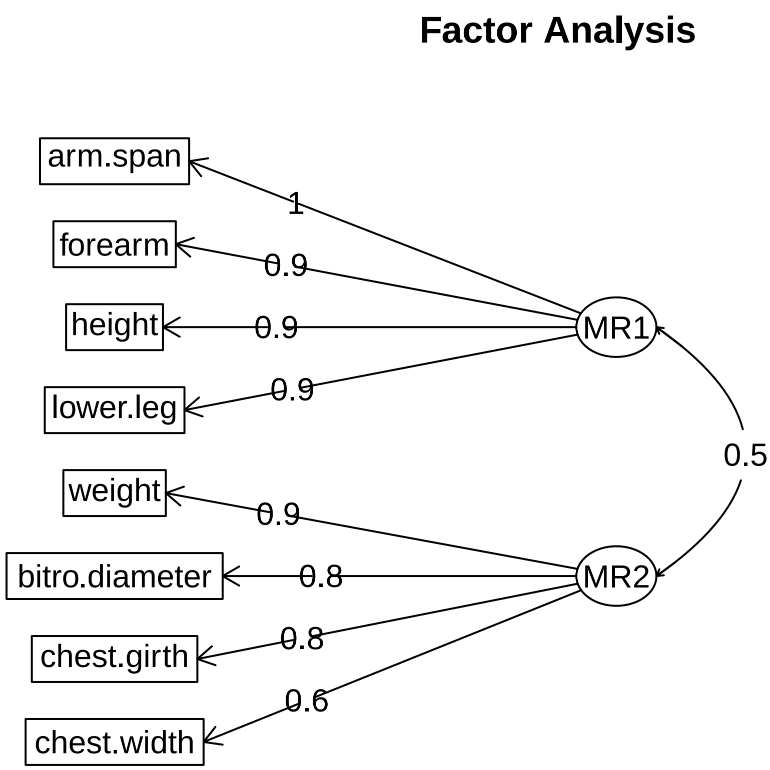
e)Orthogonal solution

```
factor.plot(fa.varimax, labels = rownames(fa.varimax$loadings))
```



f) Oblique Solutions

```
fa.diagram(fa.promax, labels = rownames(fa.promax$loadings))
```



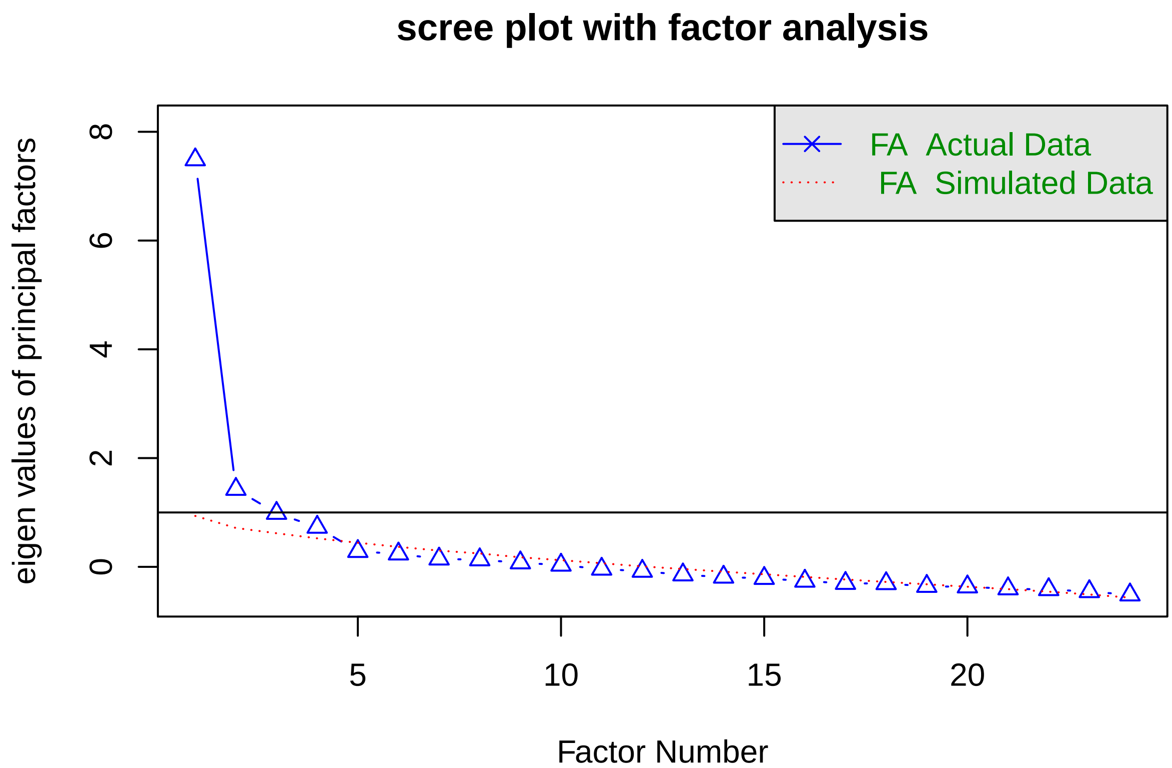
g) Interpretations

In this problem we explain the correlations among the given variables in the dataset by uncovering a smaller set of completely unrelated variables. Thus we find factors and then try to explain the correlations amongst all the variables given the dataset. In this problem we first find the factors i.e a smaller set of unrelated variables, rotate the variables and then compute their scores and then plot the results.

Problem 4

a) Determine the number of components to extract

```
fa.parallel(Harman74.cor$cov, n.obs=Harman74.cor$n.obs, fa="fa", n.iter=100, main="scree plot with facto
```




```
## Parallel analysis suggests that the number of factors =  4  and the number of components =  NA
```

b) Extract the components

```r
fa <- fa(Harman74.cor$cov, nfactors= 4, rotate = "none")
fa
```

```
## Factor Analysis using method =  minres
## Call: fa(r = Harman74.cor$cov, nfactors = 4, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                         MR1   MR2   MR3   MR4   h2   u2 com
## VisualPerception       0.60  0.03  0.38 -0.22 0.55 0.45 2.0
## Cubes                  0.37 -0.03  0.26 -0.15 0.23 0.77 2.2
## PaperFormBoard         0.42 -0.12  0.36 -0.13 0.34 0.66 2.3
## Flags                  0.48 -0.11  0.26 -0.19 0.35 0.65 2.0
## GeneralInformation     0.69 -0.30 -0.27 -0.04 0.64 0.36 1.7
## PargraphComprehension  0.69 -0.40 -0.20  0.08 0.68 0.32 1.8
## SentenceCompletion     0.68 -0.41 -0.30 -0.08 0.73 0.27 2.1
## WordClassification     0.67 -0.19 -0.09 -0.11 0.51 0.49 1.3
## WordMeaning            0.70 -0.45 -0.23  0.08 0.74 0.26 2.0
## Addition               0.47  0.53 -0.48 -0.10 0.74 0.26 3.1
## Code                   0.56  0.36 -0.16  0.09 0.47 0.53 2.0
## CountingDots           0.47  0.50 -0.14 -0.24 0.55 0.45 2.6
## StraightCurvedCapitals 0.60  0.26  0.01 -0.29 0.51 0.49 1.9
## WordRecognition        0.43  0.06  0.01  0.42 0.36 0.64 2.0
```

```
## NumberRecognition      0.39  0.10  0.09  0.37 0.31 0.69 2.2
## FigureRecognition      0.51  0.09  0.35  0.25 0.45 0.55 2.3
## ObjectNumber           0.47  0.21 -0.01  0.39 0.41 0.59 2.4
## NumberFigure           0.52  0.32  0.16  0.14 0.41 0.59 2.1
## FigureWord             0.44  0.10  0.10  0.13 0.23 0.77 1.4
## Deduction              0.62 -0.13  0.14  0.04 0.42 0.58 1.2
## NumericalPuzzles       0.59  0.21  0.07 -0.14 0.42 0.58 1.4
## ProblemReasoning       0.61 -0.10  0.12  0.03 0.40 0.60 1.1
## SeriesCompletion       0.69 -0.06  0.15 -0.10 0.51 0.49 1.2
## ArithmeticProblems     0.65  0.17 -0.19  0.00 0.49 0.51 1.3
##
##                        MR1  MR2  MR3  MR4
## SS loadings           7.65 1.69 1.22 0.92
## Proportion Var        0.32 0.07 0.05 0.04
## Cumulative Var        0.32 0.39 0.44 0.48
## Proportion Explained  0.67 0.15 0.11 0.08
## Cumulative Proportion 0.67 0.81 0.92 1.00
##
## Mean item complexity =  1.9
## Test of the hypothesis that 4 factors are sufficient.
##
## The degrees of freedom for the null model are  276  and the objective function was  11.44
## The degrees of freedom for the model are 186  and the objective function was  1.72
##
## The root mean square of the residuals (RMSR) is  0.04
## The df corrected root mean square of the residuals is  0.05
##
## Fit based upon off diagonal values = 0.98
## Measures of factor score adequacy
##                                                  MR1  MR2  MR3  MR4
## Correlation of (regression) scores with factors   0.97 0.91 0.87 0.79
## Multiple R square of scores with factors          0.94 0.82 0.75 0.62
## Minimum correlation of possible factor scores     0.89 0.65 0.50 0.24
```

```
fa$factors
```

```
## [1] 4
```

c) Factor extraction with orthogonal rotation

```
fa.varimax <- fa(Harman74.cor$cov, nfactors=4, rotate="varimax")
fa.varimax
```

```
## Factor Analysis using method =  minres
## Call: fa(r = Harman74.cor$cov, nfactors = 4, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                         MR1   MR3   MR2  MR4   h2   u2 com
## VisualPerception       0.15  0.68  0.20 0.15 0.55 0.45 1.4
## Cubes                  0.11  0.45  0.08 0.08 0.23 0.77 1.3
## PaperFormBoard         0.15  0.55 -0.01 0.11 0.34 0.66 1.2
## Flags                  0.23  0.53  0.09 0.07 0.35 0.65 1.5
## GeneralInformation     0.73  0.19  0.22 0.14 0.64 0.36 1.4
## PargraphComprehension  0.76  0.21  0.07 0.23 0.68 0.32 1.4
```

```
## SentenceCompletion     0.81  0.19  0.15 0.07 0.73 0.27 1.2
## WordClassification     0.57  0.34  0.23 0.14 0.51 0.49 2.2
## WordMeaning            0.81  0.20  0.05 0.22 0.74 0.26 1.3
## Addition               0.17 -0.11  0.82 0.16 0.74 0.26 1.2
## Code                   0.18  0.11  0.54 0.37 0.47 0.53 2.1
## CountingDots           0.02  0.20  0.71 0.09 0.55 0.45 1.2
## StraightCurvedCapitals 0.18  0.42  0.54 0.08 0.51 0.49 2.2
## WordRecognition        0.21  0.05  0.08 0.56 0.36 0.64 1.3
## NumberRecognition      0.12  0.12  0.08 0.52 0.31 0.69 1.3
## FigureRecognition      0.07  0.42  0.06 0.52 0.45 0.55 2.0
## ObjectNumber           0.14  0.06  0.22 0.58 0.41 0.59 1.4
## NumberFigure           0.02  0.31  0.34 0.45 0.41 0.59 2.7
## FigureWord             0.15  0.25  0.18 0.35 0.23 0.77 2.8
## Deduction              0.38  0.42  0.10 0.29 0.42 0.58 2.9
## NumericalPuzzles       0.18  0.40  0.43 0.21 0.42 0.58 2.8
## ProblemReasoning       0.37  0.41  0.13 0.29 0.40 0.60 3.0
## SeriesCompletion       0.37  0.52  0.23 0.22 0.51 0.49 2.7
## ArithmeticProblems     0.36  0.19  0.49 0.29 0.49 0.51 2.9
##
##                        MR1  MR3  MR2  MR4
## SS loadings           3.64 2.93 2.67 2.23
## Proportion Var        0.15 0.12 0.11 0.09
## Cumulative Var        0.15 0.27 0.38 0.48
## Proportion Explained  0.32 0.26 0.23 0.19
## Cumulative Proportion 0.32 0.57 0.81 1.00
##
## Mean item complexity =  1.9
## Test of the hypothesis that 4 factors are sufficient.
##
## The degrees of freedom for the null model are  276  and the objective function was  11.44
## The degrees of freedom for the model are 186  and the objective function was  1.72
##
## The root mean square of the residuals (RMSR) is  0.04
## The df corrected root mean square of the residuals is  0.05
##
## Fit based upon off diagonal values = 0.98
## Measures of factor score adequacy
##                                                    MR1  MR3  MR2  MR4
## Correlation of (regression) scores with factors   0.93 0.87 0.91 0.82
## Multiple R square of scores with factors          0.87 0.76 0.83 0.68
## Minimum correlation of possible factor scores     0.74 0.52 0.65 0.36
```

c) Factor extraction with oblique rotation

```r
fa.promax <- fa(Harman74.cor$cov, nfactors=4, rotate="promax")
fa.promax
```

```
## Factor Analysis using method =  minres
## Call: fa(r = Harman74.cor$cov, nfactors = 4, rotate = "promax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                          MR1   MR3   MR2   MR4   h2   u2 com
## VisualPerception       -0.08  0.78  0.06 -0.05 0.55 0.45 1.0
## Cubes                  -0.02  0.53 -0.02 -0.05 0.23 0.77 1.0
```

```
## PaperFormBoard          0.00  0.66 -0.16 -0.01 0.34 0.66 1.1
## Flags                   0.10  0.60 -0.03 -0.10 0.35 0.65 1.1
## GeneralInformation      0.79 -0.02  0.10 -0.05 0.64 0.36 1.0
## PargraphComprehension   0.82  0.00 -0.11  0.09 0.68 0.32 1.1
## SentenceCompletion      0.91 -0.02  0.03 -0.14 0.73 0.27 1.0
## WordClassification      0.54  0.22  0.11 -0.06 0.51 0.49 1.4
## WordMeaning             0.89 -0.02 -0.13  0.08 0.74 0.26 1.1
## Addition                0.09 -0.39  0.97 -0.01 0.74 0.26 1.3
## Code                    0.03 -0.11  0.53  0.29 0.47 0.53 1.7
## CountingDots           -0.15  0.09  0.81 -0.11 0.55 0.45 1.1
## StraightCurvedCapitals  0.00  0.38  0.55 -0.17 0.51 0.49 2.0
## WordRecognition         0.11 -0.15 -0.07  0.65 0.36 0.64 1.2
## NumberRecognition      -0.01 -0.03 -0.07  0.61 0.31 0.69 1.0
## FigureRecognition      -0.15  0.39 -0.14  0.54 0.45 0.55 2.2
## ObjectNumber            0.00 -0.15  0.11  0.66 0.41 0.59 1.2
## NumberFigure           -0.21  0.21  0.25  0.42 0.41 0.59 2.8
## FigureWord              0.01  0.16  0.07  0.32 0.23 0.77 1.6
## Deduction               0.27  0.35 -0.07  0.18 0.42 0.58 2.5
## NumericalPuzzles        0.00  0.34  0.38  0.03 0.42 0.58 2.0
## ProblemReasoning        0.26  0.33 -0.03  0.17 0.40 0.60 2.5
## SeriesCompletion        0.23  0.48  0.09  0.04 0.51 0.49 1.5
## ArithmeticProblems      0.27 -0.03  0.45  0.14 0.49 0.51 1.9
##
##                        MR1  MR3  MR2  MR4
## SS loadings           3.70 2.95 2.72 2.11
## Proportion Var        0.15 0.12 0.11 0.09
## Cumulative Var        0.15 0.28 0.39 0.48
## Proportion Explained  0.32 0.26 0.24 0.18
## Cumulative Proportion 0.32 0.58 0.82 1.00
##
##  With factor correlations of
##      MR1  MR3  MR2  MR4
## MR1 1.00 0.59 0.47 0.53
## MR3 0.59 1.00 0.53 0.59
## MR2 0.47 0.53 1.00 0.56
## MR4 0.53 0.59 0.56 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 4 factors are sufficient.
##
## The degrees of freedom for the null model are  276  and the objective function was  11.44
## The degrees of freedom for the model are 186  and the objective function was  1.72
##
## The root mean square of the residuals (RMSR) is  0.04
## The df corrected root mean square of the residuals is  0.05
##
## Fit based upon off diagonal values = 0.98
## Measures of factor score adequacy
##                                                    MR1  MR3  MR2  MR4
## Correlation of (regression) scores with factors   0.96 0.93 0.94 0.90
## Multiple R square of scores with factors          0.92 0.86 0.89 0.81
## Minimum correlation of possible factor scores     0.85 0.72 0.77 0.61
```

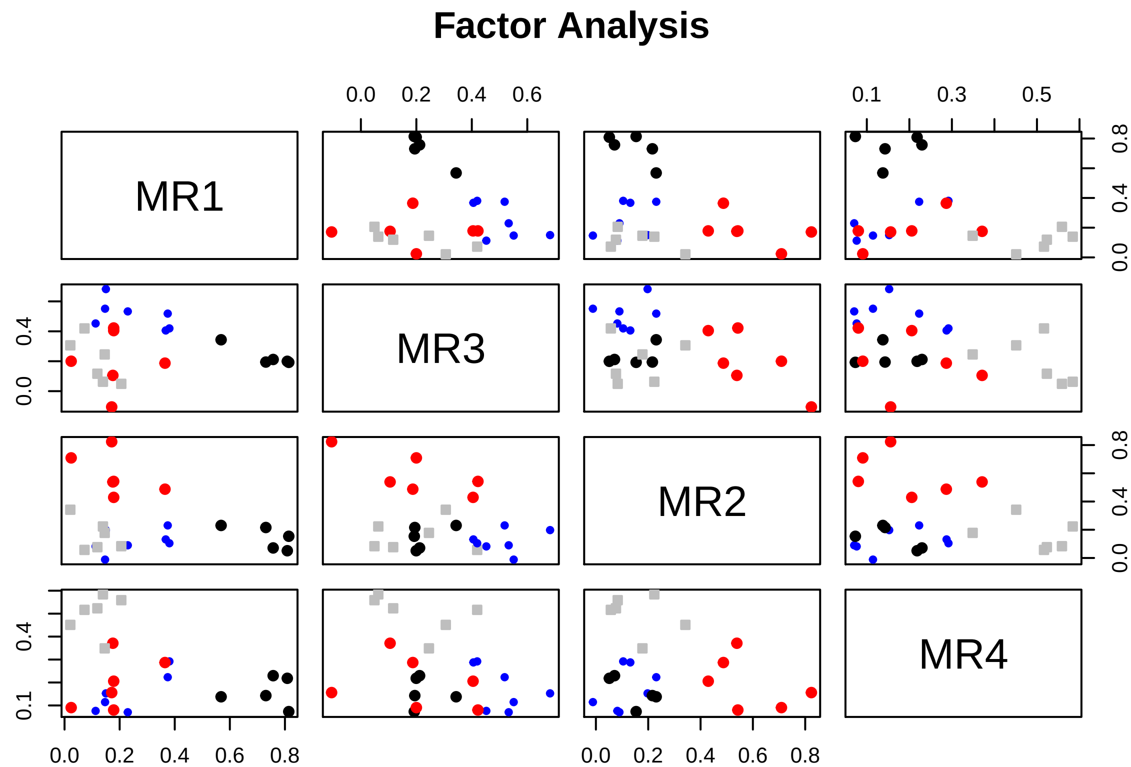
d) Factor scores

```r
fa <- fa(Harman74.cor$cov, nfactors= 4, rotate = "varimax",score=TRUE)
(fa$weights)
```

```
##                                 MR1         MR3          MR2          MR4
## VisualPerception       -0.0912528883  0.30864122  0.019076056 -0.059475135
## Cubes                  -0.0273878717  0.12425214 -0.003998151 -0.036046103
## PaperFormBoard         -0.0002664514  0.14435978 -0.021455124 -0.013910017
## Flags                  -0.0158992046  0.16513630 -0.018187085 -0.067188757
## GeneralInformation      0.1791509833 -0.01576399  0.006192041 -0.064803732
## PargraphComprehension   0.2087548894 -0.02967329 -0.089257772  0.053969292
## SentenceCompletion      0.3560671670 -0.07802378  0.009878325 -0.152657827
## WordClassification      0.0744517660  0.07646981  0.004918710 -0.042449858
## WordMeaning             0.3541762245 -0.12345571 -0.083274087  0.047708409
## Addition                0.0370184288 -0.28588788  0.530044962 -0.033006173
## Code                   -0.0152781994 -0.07727463  0.134310508  0.115144252
## CountingDots           -0.0568764212  0.04221931  0.241300880 -0.083059603
## StraightCurvedCapitals -0.0599618479  0.14994774  0.162459388 -0.110602297
## WordRecognition         0.0022332713 -0.07875443 -0.049514025  0.253344226
## NumberRecognition      -0.0305344160 -0.04164026 -0.040528450  0.215786525
## FigureRecognition      -0.0678653395  0.09799311 -0.077886130  0.223301180
## ObjectNumber           -0.0435663449 -0.07170630 -0.015111696  0.278035153
## NumberFigure           -0.0730633557  0.05023790  0.036233555  0.166892502
## FigureWord             -0.0300328554  0.02668857  0.002475198  0.083460720
## Deduction               0.0074291922  0.09386230 -0.040441038  0.055443652
## NumericalPuzzles       -0.0250346901  0.11130425  0.071040209 -0.001990422
## ProblemReasoning        0.0145135443  0.07473804 -0.014585388  0.057911235
## SeriesCompletion        0.0124897301  0.17157853 -0.007344283 -0.017896058
## ArithmeticProblems      0.0062265043  0.00390309  0.085098413  0.047726702
```

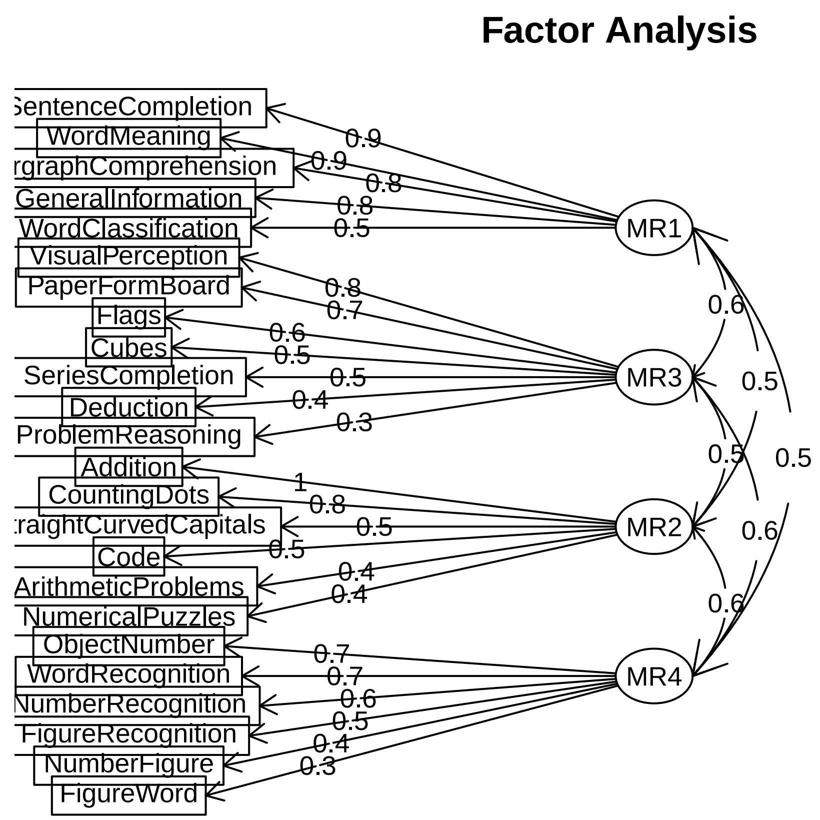
e) Orthogonal solution

```r
factor.plot(fa.varimax)
```



f) Oblique Solutions

```
fa.diagram(fa.promax, labels = rownames(fa.promax$loadings))
```



g) Interpretations

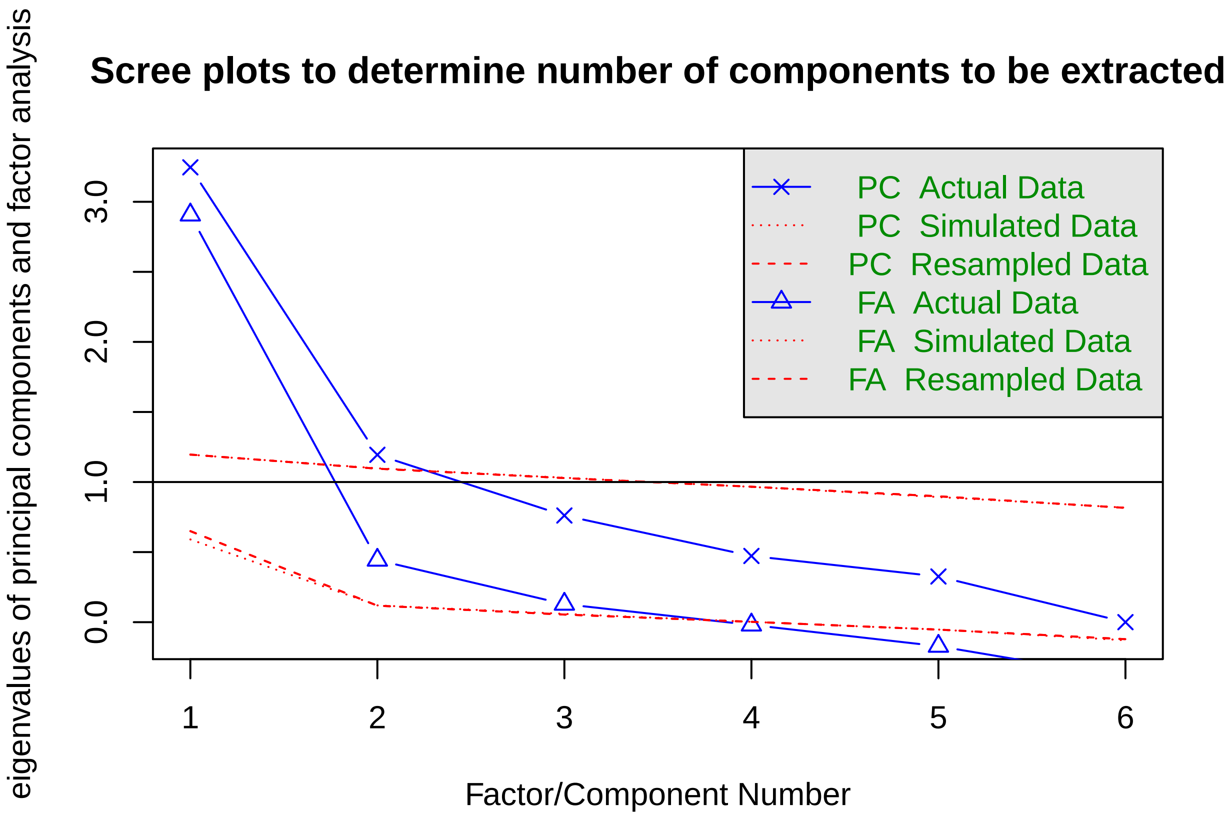
According to the Scree test with factor analysis we were able to analyze 4 factor that were to be extracted. In this problem we basically explain the correlations among the given variables in the dataset by uncovering a smaller set of completely unrelated variables. Thus we find factors and then try to explain the correlations amongst all the variables given the dataset.

In this problem we first find the factors i.e a smaller set of unrelated variables, rotate the variables and then compute their scores and then plot the results.

problem 5

Dropping Class variable a) Determining number of components to extract

```
verteb <- read_excel("C:/Users/pc/Desktop/Spring2019/DM/hw2/Vertebral_Column_Data.xlsx")

fa.parallel(verteb[1:6], fa="both", n.iter = 100, main = "Scree plots to determine number of components
```



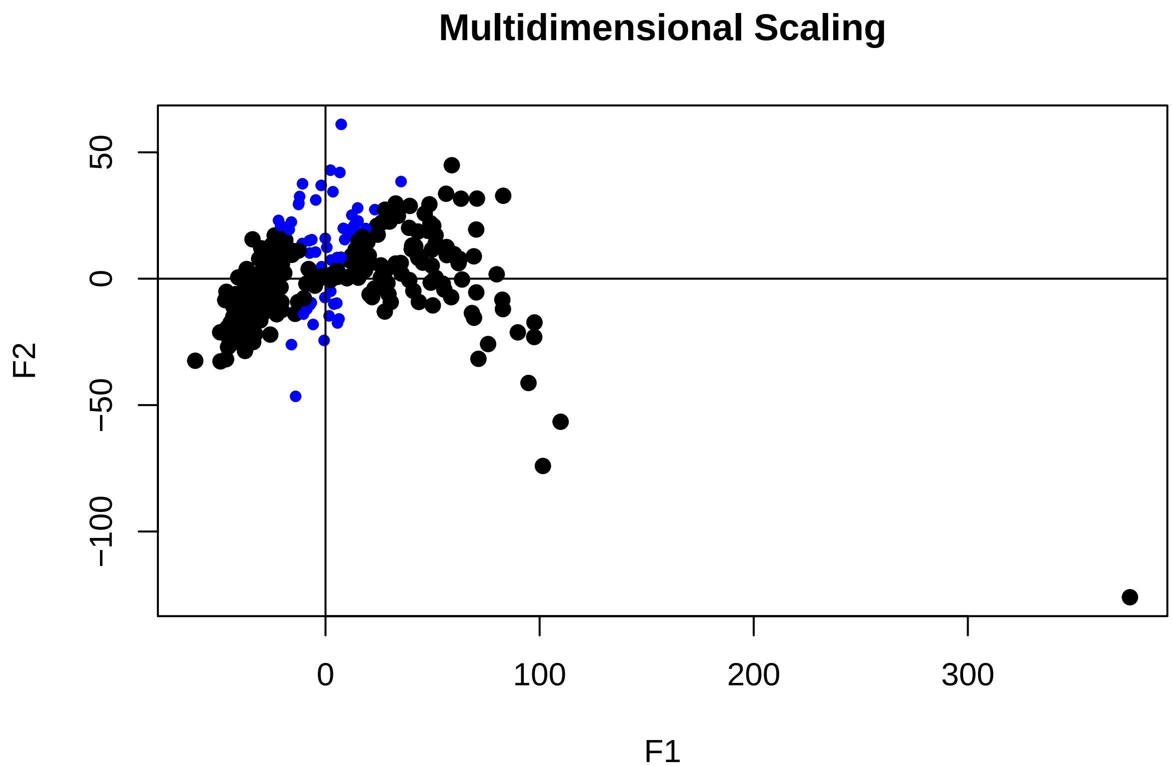

```
## Parallel analysis suggests that the number of factors =  3  and the number of components =  2
```

b) Perform Multidimensional scaling

```
dist<-dist(verteb[1:6], method = "euclidean", diag = FALSE, upper = FALSE, p=2)
cmd <- cmdscale(dist,k=2)
```

c) Orthogonal Solution

```
factor.plot(cmd, title = "Multidimensional Scaling", label="")
```



d) Interpretations

The vertebral dataset contains 6 variables and a class variable containing 3 sub-classes and a total of 310 instances. After deleting the class column we run the data through fa.parallel function, in fa argument we gave the argument as “both” which gives analysis of both factor analysis and parallel analysis. Factor analysis suggest us the number of component as 4 and parallel analysis as 3. Then we use cmdscale() function which does classical multidimensional scaling. Multidimensional scaling is one of several multivariate techniques that aim to reveal the structure of data set by plotting points in one or two dimesnions. We then develop a factor plot to the multidimensional scaled data.