Group15 - HW2

Hansika Karkera, Vrunda Shah

2/10/2020

Problem 01

```r
library(psych)
```

```
## Warning: package 'psych' was built under R version 3.6.2
```

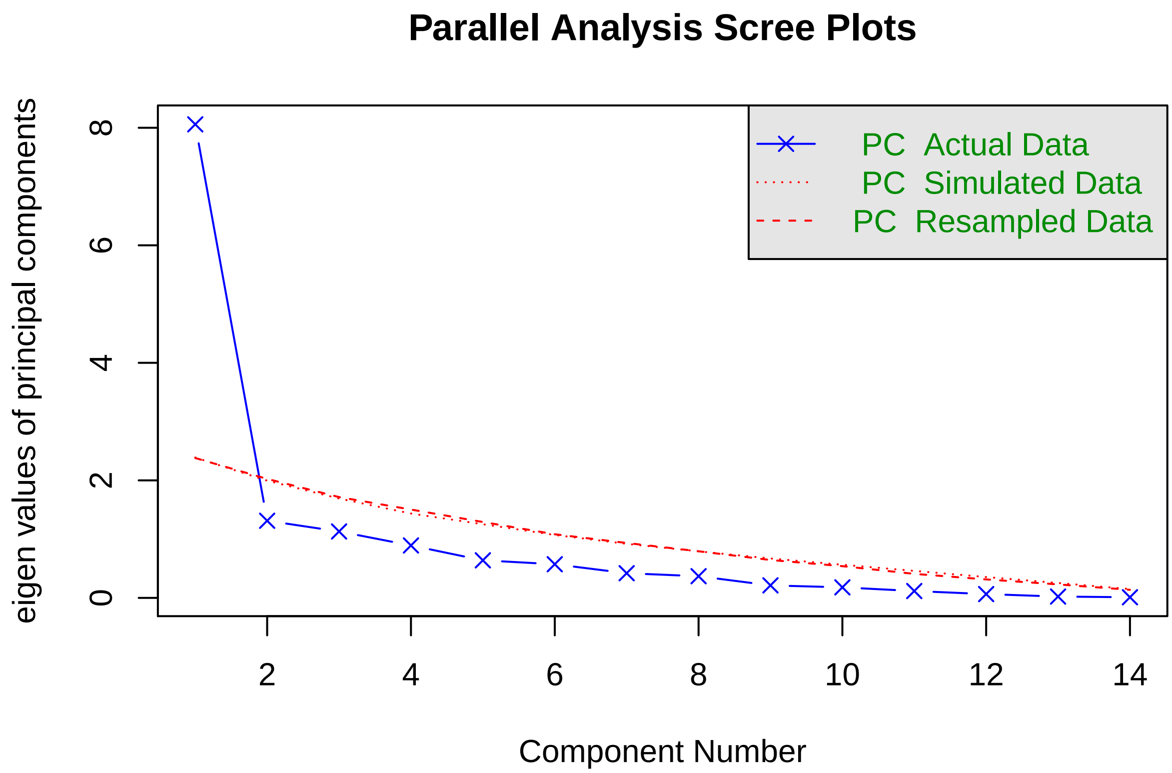
```r
library(readxl)
# read NHL.xlsx file
nhl_data <- read_xlsx("NHL.xlsx")
```

```
## New names:
## * `` -> ...1
```

```r
nhl_data <- nhl_data[,12:25]

# determines the number of components to extract
fa.parallel(nhl_data, fa = "pc")
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```




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

```r
# extracts the components
nhl_pc <- principal(nhl_data, nfactors = 1, rotate = "none")
nhl_pc
```

```
## Principal Components Analysis
## Call: principal(r = nhl_data, nfactors = 1, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          PC1   h2    u2 com
## p\rpc   0.97 0.94 0.057   1
## gg      0.83 0.69 0.308   1
## gag    -0.82 0.67 0.327   1
## five    0.92 0.84 0.162   1
## PPP     0.14 0.02 0.980   1
## PKP     0.69 0.48 0.519   1
## shots   0.59 0.34 0.656   1
## sag    -0.62 0.39 0.612   1
## sc1     0.81 0.66 0.338   1
## tr1     0.76 0.58 0.422   1
## lead1   0.81 0.65 0.351   1
## lead2   0.74 0.55 0.452   1
## wop     0.71 0.51 0.491   1
## wosp    0.86 0.73 0.267   1
##
```

```
##                  PC1
## SS loadings    8.06
## Proportion Var 0.58
##
## Mean item complexity =  1
## Test of the hypothesis that 1 component is sufficient.
##
## The root mean square of the residuals (RMSR) is  0.1
##  with the empirical chi square  52.29  with prob <  0.99
##
## Fit based upon off diagonal values = 0.97
```

```r
# rotates the principal components
nhl_pc_rotate <- principal(nhl_data, nfactors = 1)
nhl_pc_rotate
```

```
## Principal Components Analysis
## Call: principal(r = nhl_data, nfactors = 1)
## Standardized loadings (pattern matrix) based upon correlation matrix
##         PC1   h2    u2 com
## p\rpc  0.97 0.94 0.057   1
## gg     0.83 0.69 0.308   1
## gag   -0.82 0.67 0.327   1
## five   0.92 0.84 0.162   1
## PPP    0.14 0.02 0.980   1
## PKP    0.69 0.48 0.519   1
## shots  0.59 0.34 0.656   1
## sag   -0.62 0.39 0.612   1
## sc1    0.81 0.66 0.338   1
## tr1    0.76 0.58 0.422   1
## lead1  0.81 0.65 0.351   1
## lead2  0.74 0.55 0.452   1
## wop    0.71 0.51 0.491   1
## wosp   0.86 0.73 0.267   1
##
##                  PC1
## SS loadings    8.06
## Proportion Var 0.58
##
## Mean item complexity =  1
## Test of the hypothesis that 1 component is sufficient.
##
## The root mean square of the residuals (RMSR) is  0.1
##  with the empirical chi square  52.29  with prob <  0.99
##
## Fit based upon off diagonal values = 0.97
```

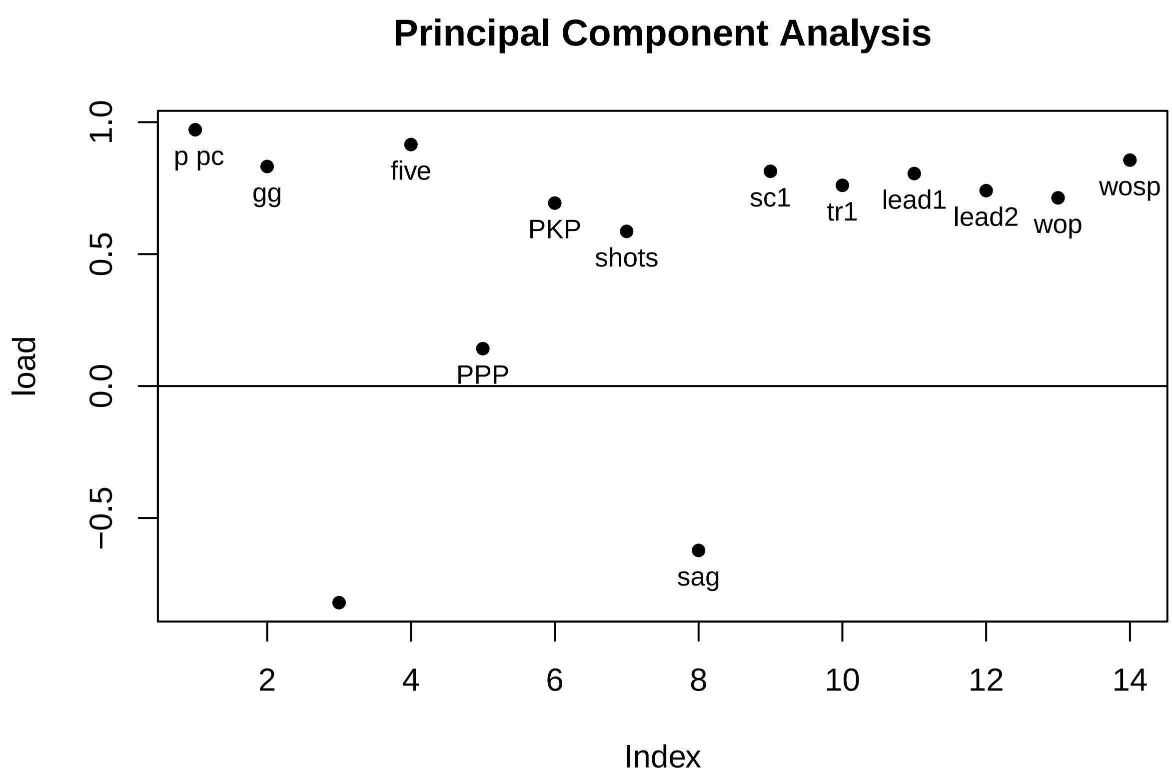
```r
# computes component scores
nhl_scores <- nhl_pc_rotate$scores
head(nhl_scores)
```

```
##            PC1
## [1,] 1.4163062
```

```
## [2,] 1.0620224
## [3,] 0.7363287
## [4,] 0.7558835
## [5,] 1.1940130
## [6,] 1.0912159
```

```r
# plots graph of an orthogonal solution
factor.plot(nhl_pc_rotate, cex = 0.8, labels = rownames(nhl_pc_rotate$loadings), pos = 1)
```

```
## Warning in text.default(load, labels, pos = pos, ...): font width unknown for
## character 0xd
```



INTERPRETATION :

To determine the number of components to be extracted, we use fa.parallel() which plots the Parallel Analysis Scree plot of the raw data. In this plot the eigenvalues are plotted against their component numbers and as per Kaiser-Hariss criterion of retaining components with eigenvalues greater than 1, we can conclude that the number of components to be extracted is one. Using principal() we extract the principal components and observe that PC1 has a negative correlation with gag and sag and a positive correaltion with rest of the features. The column labeled h2 and u2 contains the component communalities and uniqueness.Meaning, 69% of variance in gg is accounted for by PC1 and 31% is not. SS loadings contains the eigen values associated with the components and Proportion Var states the the principal component accounts for 58% of the variance in all the variables. Further on, we rotate the components but observe no difference in the result because only one component has been selected.Finally, we calculate the component scores and loading factor and go on to plot an orthogonal solution.

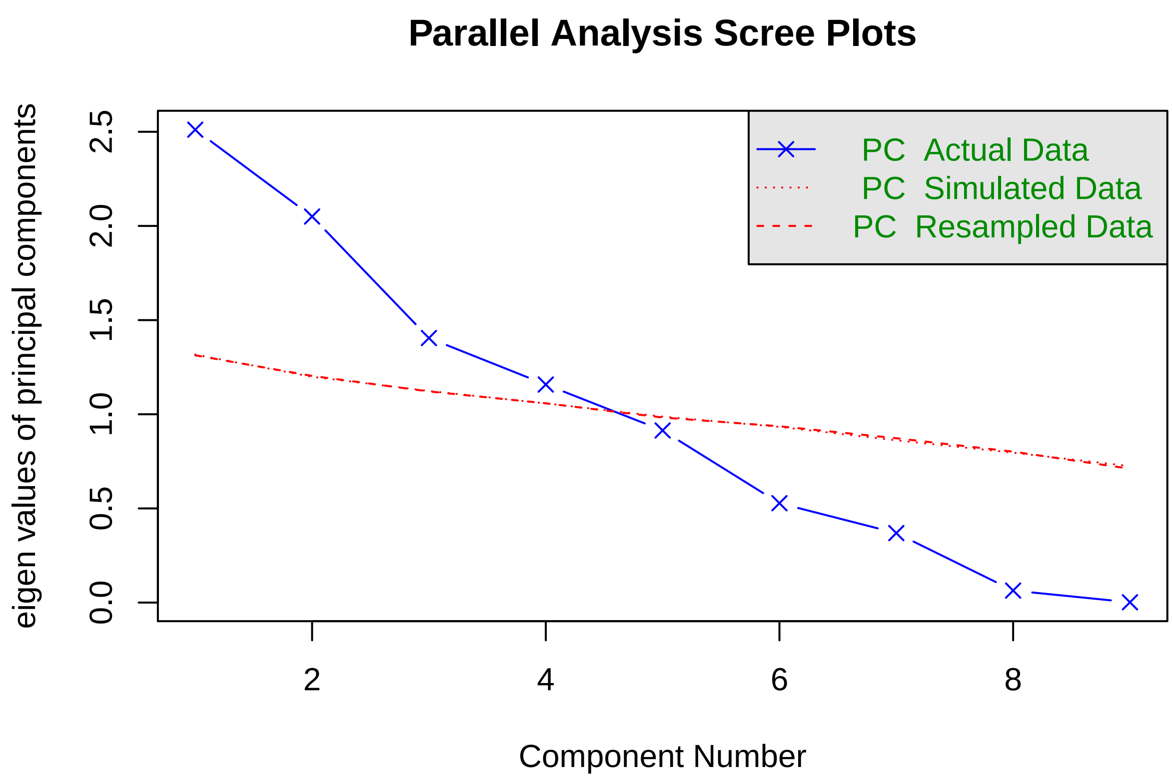
Problem 02

```r
library(psych)
library(readxl)
# read Glass Identification Data.xlsx file
glass_data <- read_xlsx("Glass Identification Data.xlsx")

# determines the number of components to extract
fa.parallel(glass_data[,2:10], fa = "pc")
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully
```




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

```r
# extracts the components
glass_pc <- principal(glass_data[,2:10], nfactors = 4, rotate = "none")
glass_pc
```

```
## Principal Components Analysis
## Call: principal(r = glass_data[, 2:10], nfactors = 4, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      PC1   PC2   PC3   PC4   h2    u2 com
## RI -0.86  0.41  0.10 -0.16 0.95 0.051 1.5
## Na  0.41  0.39 -0.46 -0.53 0.80 0.195 3.8
## Mg -0.18 -0.85  0.01 -0.41 0.92 0.081 1.5
## Al  0.68  0.42  0.39  0.15 0.81 0.186 2.5
## Si  0.36 -0.22 -0.54  0.70 0.97 0.031 2.7
## K   0.35 -0.22  0.79  0.04 0.79 0.212 1.6
## CA -0.78  0.49  0.00  0.30 0.94 0.058 2.0
## Ba  0.40  0.69  0.09 -0.14 0.67 0.333 1.7
## Fe -0.29 -0.09  0.34  0.25 0.27 0.730 3.0
##
##                        PC1  PC2  PC3  PC4
## SS loadings           2.51 2.05 1.40 1.16
## Proportion Var        0.28 0.23 0.16 0.13
## Cumulative Var        0.28 0.51 0.66 0.79
## Proportion Explained  0.35 0.29 0.20 0.16
## Cumulative Proportion 0.35 0.64 0.84 1.00
##
## Mean item complexity =  2.3
## 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
```

```r
# rotates the components
glass_rotate <- principal(glass_data[,2:10], nfactors = 4)
glass_rotate
```

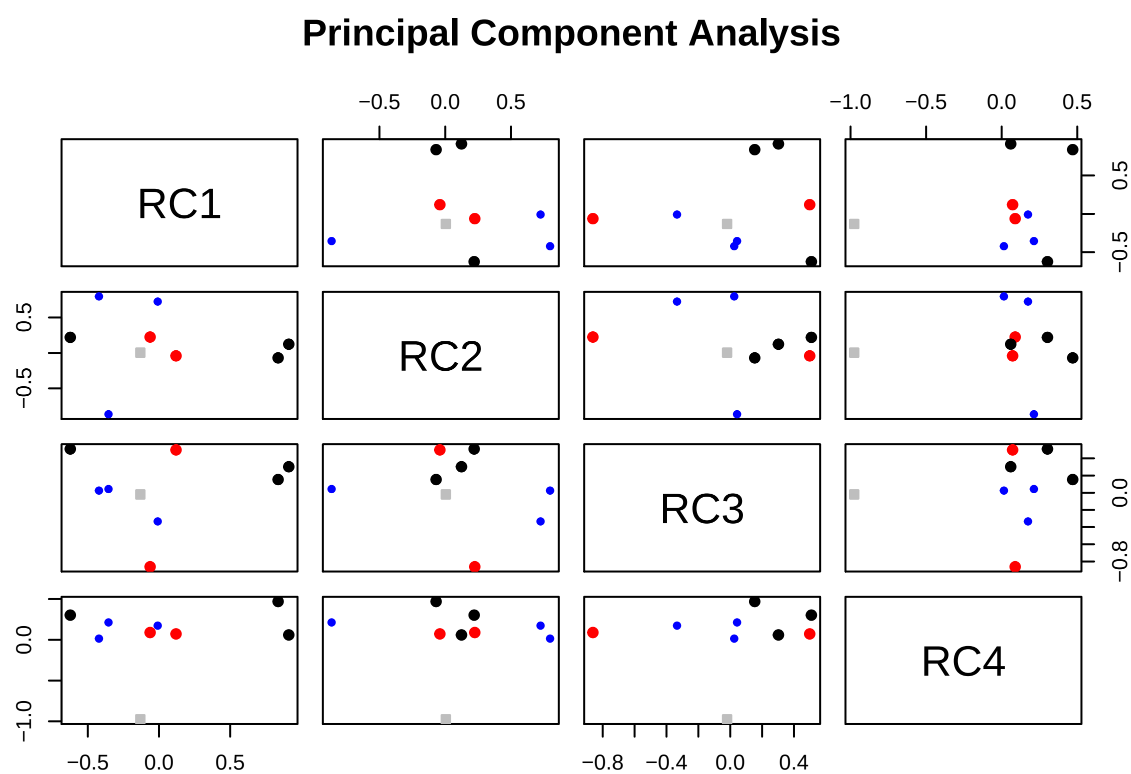
```
## Principal Components Analysis
## Call: principal(r = glass_data[, 2:10], nfactors = 4)
## Standardized loadings (pattern matrix) based upon correlation matrix
##      RC1   RC2   RC3   RC4   h2    u2 com
## RI  0.84 -0.07  0.15  0.47 0.95 0.051 1.7
## Na -0.06  0.22 -0.86  0.09 0.80 0.195 1.2
## 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
```

```
# computes the component score
glass_score <- glass_rotate$scores
head(glass_score)
```

```
##             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
```

```
# plots the graph of an orthogonal solution
factor.plot(glass_rotate)
```

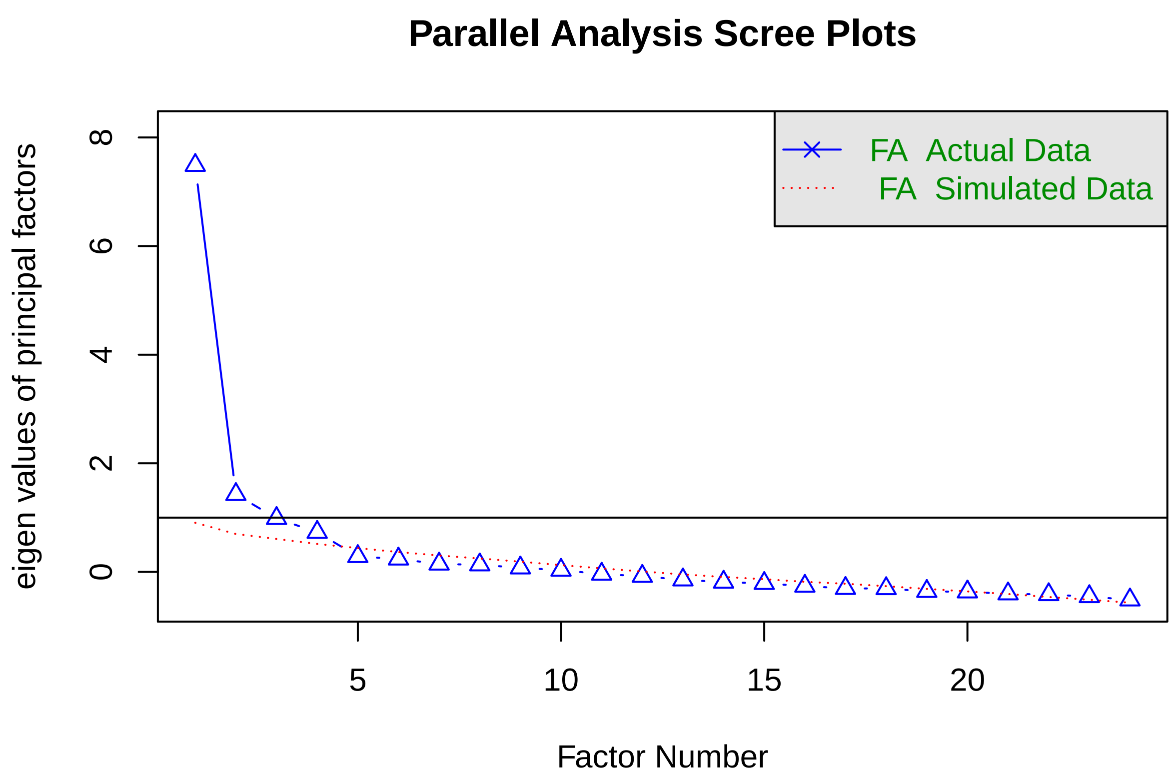


INTERPRETATION :

Using fa.parallel() we find the number of components to be extracted is 4. Using principal() we extract the principal components and observe that the 4 components together account for 95% of the variance in RI variable. The proportion variance reduces as the principal component increases. After rotating the components, we observe that RC1 is highly correalted with RI and CA, RC2 with Al and Ba, RC3 a combianation of K and Fe, RC4 with RI and K.
Further on, we rotate the components but observe no difference in the result because only one component has been selected.Finally, the correlation between the 4 components is shown the plot.

Problem 03

```r
# determines the number of components to extract
fa.parallel(Harman74.cor$cov, n.obs = Harman74.cor$n.obs, fa = "fa")
```




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

```r
# extracts the principal components
test_fa <- fa(r = Harman74.cor$cov, nfactors = 4, rotate = "none")
test_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
```

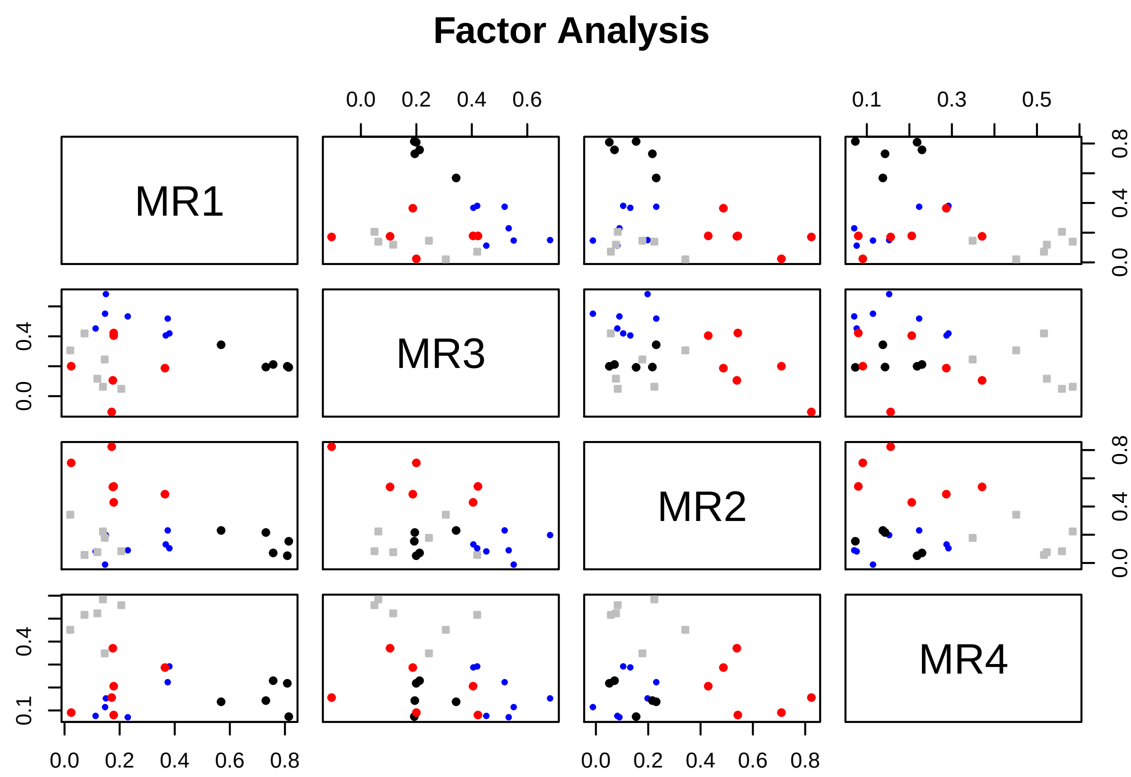
```r
# rotates the components
test_rotate <- fa(Harman74.cor$cov, nfactors = 4, rotate = "varimax")
test_rotate
```

```
## 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
```

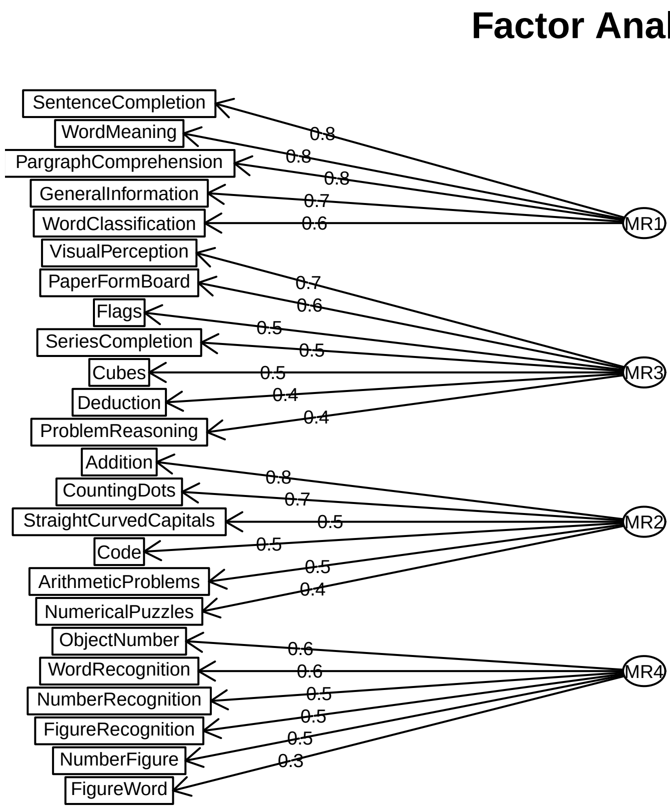
```r
# computes factor scores
factor.scores(Harman74.cor$cov, test_rotate)
```

```
## $scores
## NULL
##
## $weights
##                                MR1          MR3          MR2         MR4
## VisualPerception       -0.108560731  0.3638003297  0.021830043 -0.09131680
## Cubes                  -0.033127638  0.1470215827 -0.003579868 -0.05126393
## PaperFormBoard         -0.005193818  0.1679483822 -0.024596805 -0.02536148
## Flags                  -0.021213649  0.1960131974 -0.018199826 -0.09171444
## GeneralInformation      0.195336373 -0.0209818149  0.008294998 -0.08426651
## PargraphComprehension   0.225215420 -0.0448990761 -0.103652758  0.06638051
## SentenceCompletion      0.390814254 -0.0940553405  0.015604235 -0.19353462
## WordClassification      0.078410797  0.0881870459  0.005919246 -0.05960645
## WordMeaning             0.385469423 -0.1585444910 -0.097406069  0.06003563
## Addition                0.045521427 -0.3355928198  0.590020101 -0.05105587
## Code                   -0.018636073 -0.0977594786  0.143291766  0.13951977
## CountingDots           -0.063198794  0.0536824601  0.271156516 -0.11564496
## StraightCurvedCapitals -0.068888672  0.1809652548  0.184349441 -0.15217445
## WordRecognition        -0.001782629 -0.1071161105 -0.067298607  0.31823792
## NumberRecognition      -0.037431613 -0.0606070912 -0.055400742  0.27030867
## FigureRecognition      -0.082799689  0.1020944728 -0.098199518  0.27350898
## ObjectNumber           -0.052524482 -0.0992069908 -0.030138056  0.34762561
## NumberFigure           -0.086179188  0.0495228612  0.031458181  0.20159227
## FigureWord             -0.035967604  0.0264997952 -0.001645204  0.10134192
## Deduction               0.003129873  0.1049790901 -0.048775900  0.06372349
## NumericalPuzzles       -0.032056311  0.1288857633  0.077611016 -0.01273933
## ProblemReasoning        0.011105629  0.0821539043 -0.020216633  0.06640681
## SeriesCompletion        0.007430815  0.1990010401 -0.009262734 -0.03317640
## ArithmeticProblems      0.004041297  0.0001718559  0.091336945  0.05352096
##
## $r.scores
##              MR1          MR3          MR2          MR4
## MR1 1.000000e+00 1.334219e-15 3.524958e-15 1.280226e-15
## MR3 1.302669e-15 1.000000e+00 2.081668e-17 9.853229e-16
## MR2 3.527994e-15 8.593657e-17 1.000000e+00 2.602085e-16
## MR4 1.232738e-15 1.079161e-15 2.393918e-16 1.000000e+00
##
## $R2
## [1] 0.9312884 0.8694881 0.9075352 0.8208524
```

```r
# plots an orthogonal solution
factor.plot(test_rotate, cex = 0.7)
```




```r
# plots an oblique solution
fa.diagram(test_rotate, cex = 0.55)
```

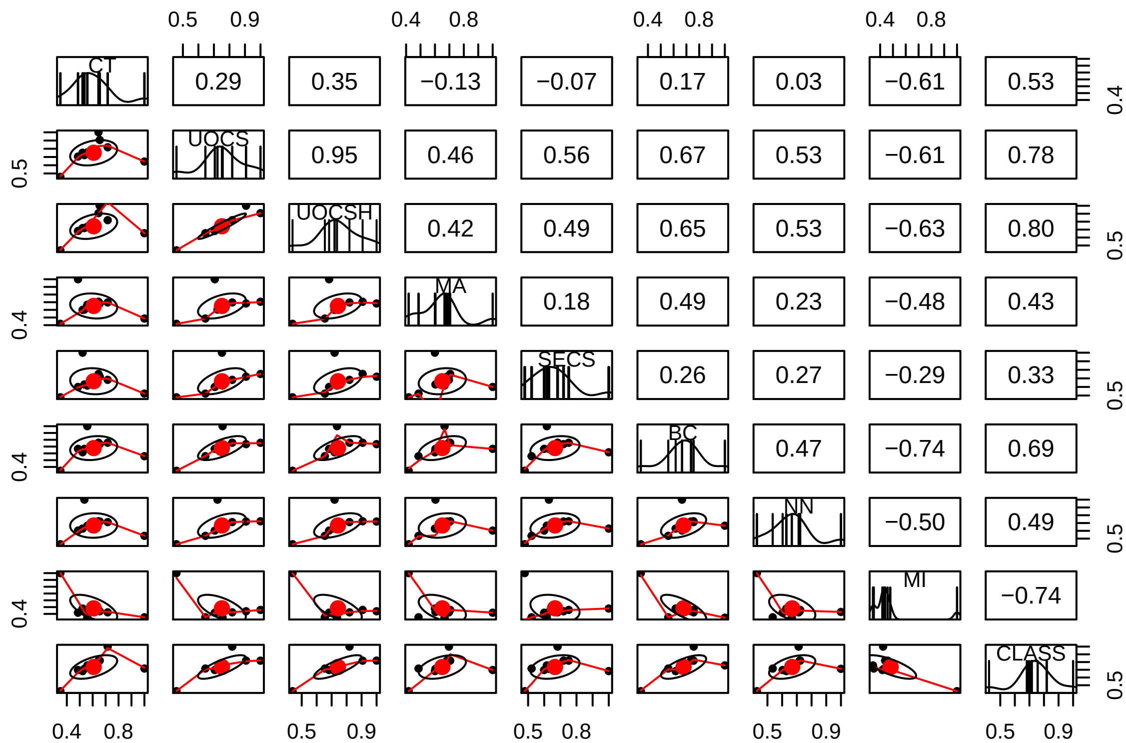


INTERPRETATION

Based on the Scree plot we determine the number of factors to be extracted as 4. Using fa() we extract the unrotated factors resulting in the minimal residual solution. Further we rotate the components using varimax oblique transformation. The factor plot contains data points from all the parameters. Each plot shows the combined correlation between the MRs from which the correlation between the factors is not very clear. fa.diagram() plots the oblique solution which shows that MR1 aligns with 5 components, MR2 with the next 7 components, MR3 with next 6 and MR4 with the last 6 elements.

Problem 04

```
breast_cancer <- read_xlsx("breast-cancer-wisconsin.xlsx")
breast_cancer$BN<- NULL
breast_cancer$ID <- NULL
# calculates the correlation matrix
cor_df <- cor(breast_cancer)
# visualizing the correlation matrix
pairs_panes <- pairs.panels(cor_df, method = "pearson")
```



Form the scatter plot implemented using pairs.panel we can see that all the variables are discreate in the dataset. Also most of the variables are positively correlated. Only MI variable is weekly correlated while all other variables are strongly correlated.

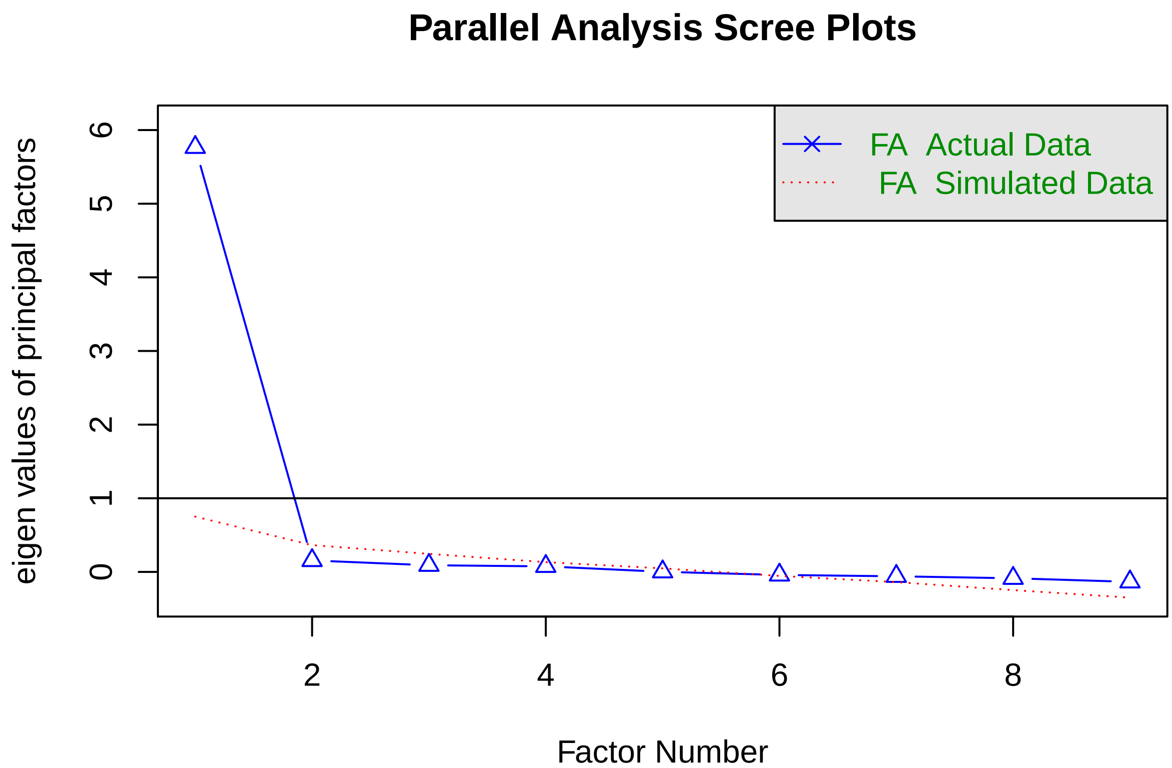
```
# determines number of components to be extracted
fa.parallel(cor_df, fa="fa")
```

```
## Warning in fa.parallel(cor_df, fa = "fa"): It seems as if you are using a
## correlation matrix, but have not specified the number of cases. The number of
## subjects is arbitrarily set to be 100
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```




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

```r
# extract the components
cancer_fa <- fa(cor_df,nfactors = 1, rotate = "none", fm = "pa")
cancer_fa
```

```
## Factor Analysis using method =  pa
## Call: fa(r = cor_df, nfactors = 1, rotate = "none", fm = "pa")
## Standardized loadings (pattern matrix) based upon correlation matrix
##        PA1   h2   u2 com
## CT    0.70 0.49 0.51   1
## UOCS  0.93 0.87 0.13   1
## UOCSH 0.92 0.84 0.16   1
## MA    0.76 0.58 0.42   1
## SECS  0.79 0.62 0.38   1
## BC    0.81 0.66 0.34   1
## NN    0.79 0.63 0.37   1
## MI    0.50 0.25 0.75   1
## CLASS 0.91 0.82 0.18   1
##
##                 PA1
## SS loadings    5.77
## Proportion Var 0.64
##
## Mean item complexity =  1
```

```
## Test of the hypothesis that 1 factor is sufficient.
##
## The degrees of freedom for the null model are  36  and the objective function was  7.63
## The degrees of freedom for the model are 27  and the objective function was  0.35
##
## The root mean square of the residuals (RMSR) is  0.03
## The df corrected root mean square of the residuals is  0.04
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##                                                    PA1
## Correlation of (regression) scores with factors   0.98
## Multiple R square of scores with factors          0.96
## Minimum correlation of possible factor scores     0.92
```

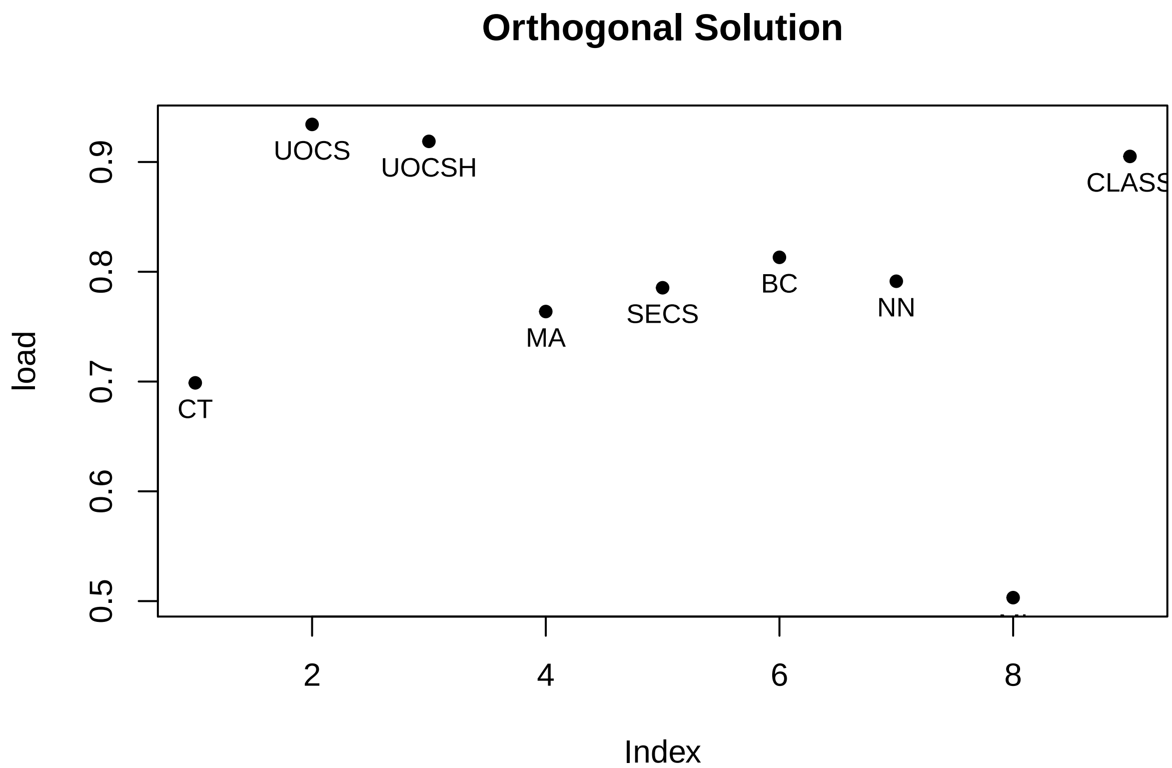
```
# rotating the factors
fa_varimax <- fa(cor_df, nfactors = 1, rotate = "varimax" , fm = "pa")
fa_varimax
```

```
## Factor Analysis using method =  pa
## Call: fa(r = cor_df, nfactors = 1, rotate = "varimax", fm = "pa")
## Standardized loadings (pattern matrix) based upon correlation matrix
##        PA1   h2   u2 com
## CT    0.70 0.49 0.51   1
## UOCS  0.93 0.87 0.13   1
## UOCSH 0.92 0.84 0.16   1
## MA    0.76 0.58 0.42   1
## SECS  0.79 0.62 0.38   1
## BC    0.81 0.66 0.34   1
## NN    0.79 0.63 0.37   1
## MI    0.50 0.25 0.75   1
## CLASS 0.91 0.82 0.18   1
##
##                 PA1
## SS loadings    5.77
## Proportion Var 0.64
##
## Mean item complexity =  1
## Test of the hypothesis that 1 factor is sufficient.
##
## The degrees of freedom for the null model are  36  and the objective function was  7.63
## The degrees of freedom for the model are 27  and the objective function was  0.35
##
## The root mean square of the residuals (RMSR) is  0.03
## The df corrected root mean square of the residuals is  0.04
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##                                                    PA1
## Correlation of (regression) scores with factors   0.98
## Multiple R square of scores with factors          0.96
## Minimum correlation of possible factor scores     0.92
```

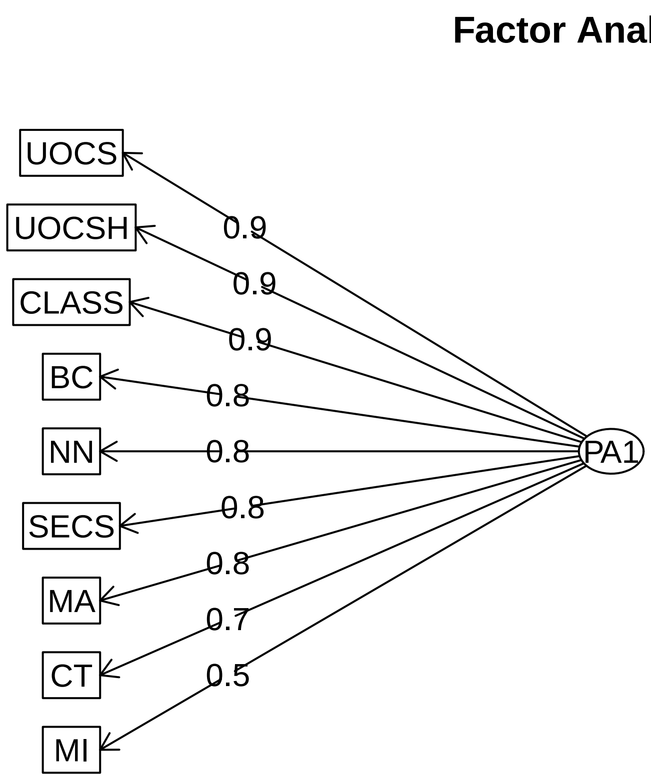
```r
# computing factor scores
factor.scores(cor_df, fa_varimax)
```

```
## $scores
## NULL
##
## $weights
##              PA1
## CT    0.02858932
## UOCS  0.28159394
## UOCSH 0.18593667
## MA    0.07296990
## SECS  0.08270340
## BC    0.08547128
## NN    0.09306812
## MI    0.03821781
## CLASS 0.24760259
##
## $r.scores
##     PA1
## PA1   1
##
## $R2
## [1] 0.9610708
```

```r
# plotting orthogonal solution
factor.plot(fa_varimax, title = "Orthogonal Solution", labels = rownames(fa_varimax$loadings), cex = 0.8
```




```
#plotting an oblique solution
fa.diagram(fa_varimax , labels = rownames(fa_varimax$loadings))
```



INTERPRETATION

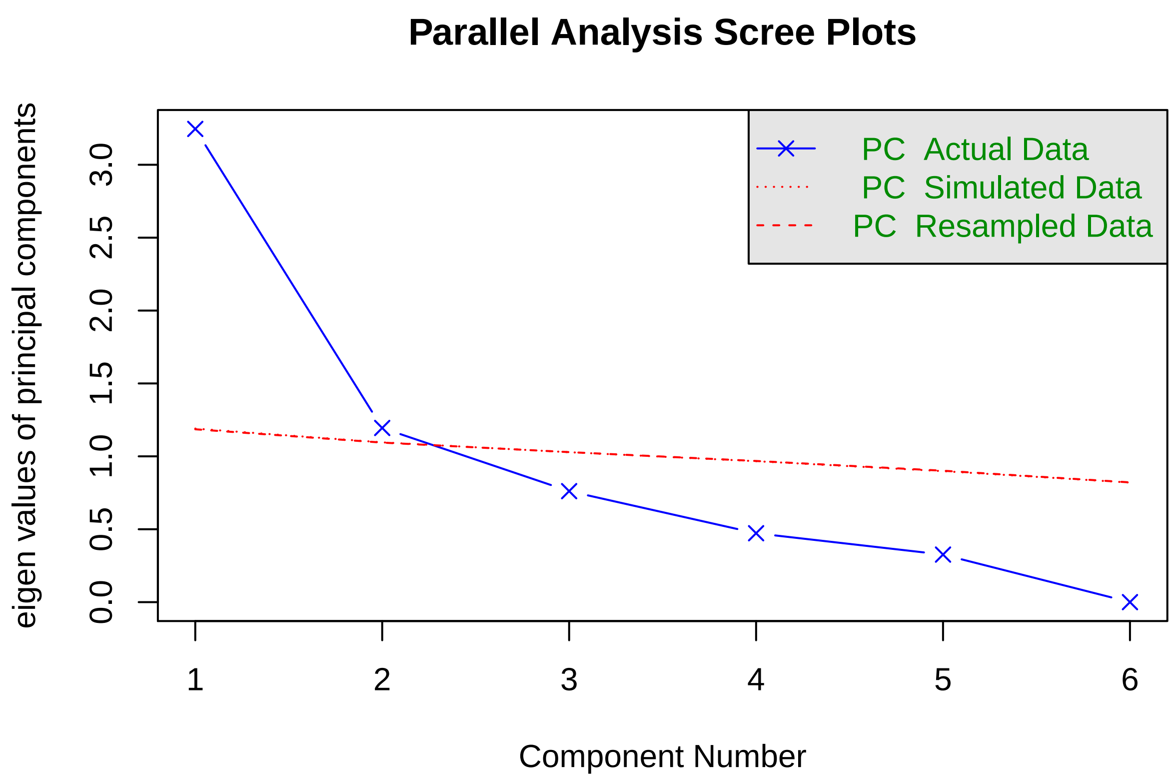
The dataset in Task 4 contains information about types of Breast cancer with 11 attributes. We first found the correlation matrix using cor() function and used that to plot scatter plot using pairs.panels() function to analyze the correlation. We used dataframe obtained from cor() function to find the number of components to be extracted. Using fa.parallel() gives us 1 component to be extracted. Further we use fa() function to extract the factors. By rotating the factors output remains same as we only have 1 factor. Finally correlation is examined using orthogonal graph and oblique solution using factor.diagram() function which shows relation of component with each variables.

Problem 05

```r
vertebral_data <- read_xlsx("Vertebral Column Data.xlsX")
# determines the number of components to be extracted
fa.parallel(vertebral_data[1:6], fa="pc" , n.iter = 100)
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully
```

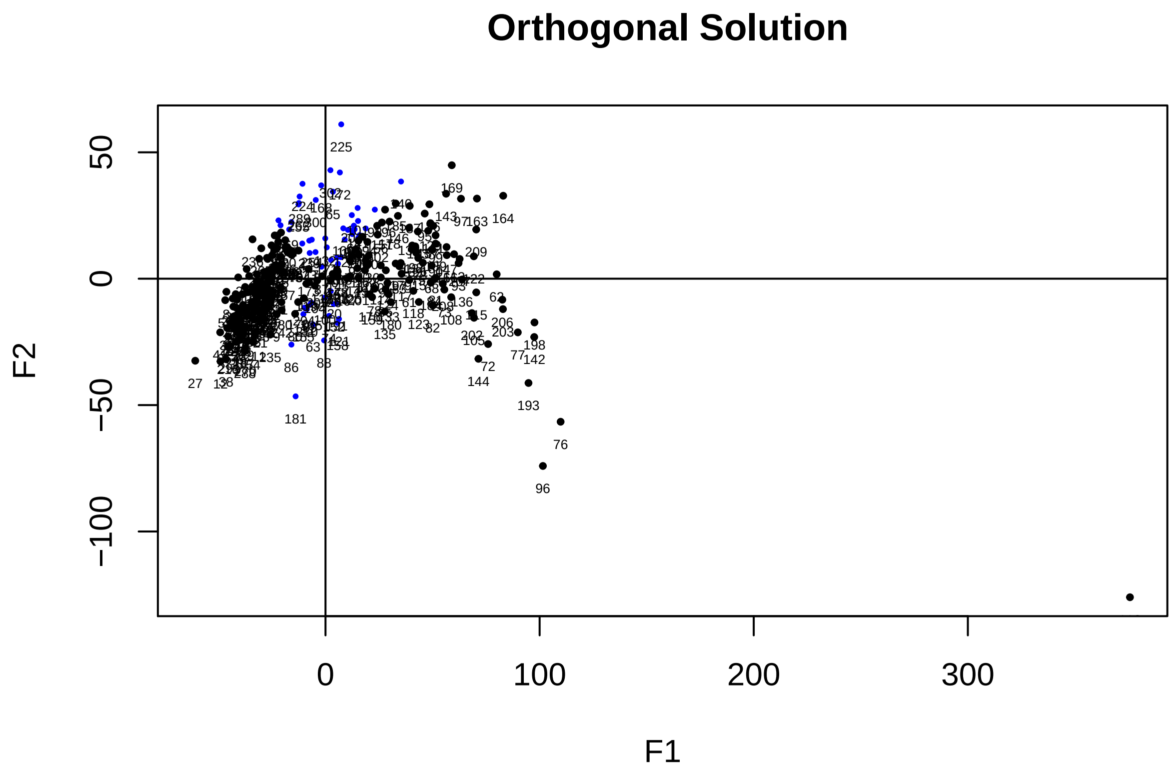



```
## Parallel analysis suggests that the number of factors =  NA  and the number of components =  2
```

```
# compute raw data matrix
dist <- dist(vertebral_data[,1:6], method = "euclidean")
# multidimensional scaling
cmd <- cmdscale(dist, k = 2)
head(cmd)
```

```
##            [,1]       [,2]
## [1,] -25.21264  13.204206
## [2,] -37.55028 -18.951621
## [3,] -21.95087  23.063614
## [4,] -10.84709  13.917984
## [5,] -27.73305  -7.589005
## [6,] -39.74800 -22.959841
```

```
# plotting an orthogonal solution
factor.plot(cmd, title = " Orthogonal Solution", cex = 0.4)
```



INTERPRETATION

The dataset used in task 5 is Vertebral Column, it contains 7 attributes. To determine the number of factors to be extracted we implemented fa.parallel() function from pysch package. This function suggests us to use 2 principal components for analysis. Next we applied dist() function to raw data which calculate Euclidean distance that represent dissimilarity in matrix. By using the distance, we conducted multidimensional scaling. Finally, correlation between the two components is shown with orthogonal plot. Here we have given the input of cmdscale function to plot a graph.