Exploratory Factor Analysis

Contents

```
library(mosaic)
library(ggplot2)
library(dplyr)
library(psych)

#install.packages("GPArotation")
library(GPArotation)
```

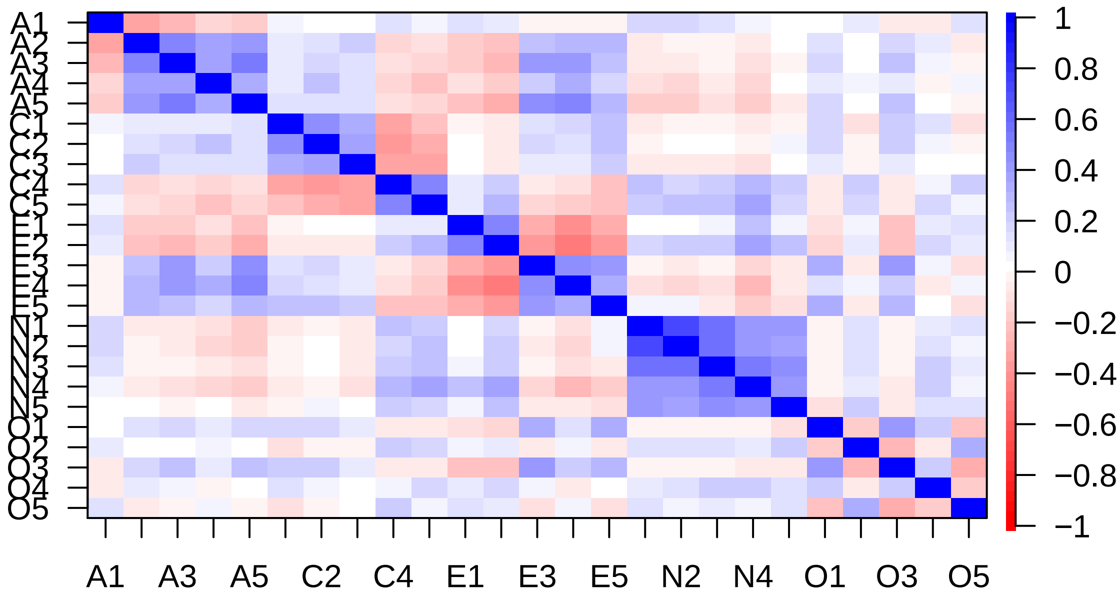
Exploratory Factor Analysis

Let's load the Big Five Inventory dataset.

```
data(bfi)
```

Correlation Heat Map

```
#correlation matrix of ALL items
r_matrix <- corr.test(select(bfi, -gender, -education, -age))$r

cor.plot(r_matrix)
```



Eigenvalues

A method for finding the number of factors called the Kaiser method involves taking the number of eigenvalues of the correlation matrix that are greater than 1. We can do that with the `eigen()` function.

```
sum(eigen(r_matrix)$values >1)
```

```
## [1] 6
```

Because there are 6, the Kaiser method would have us extract 6 factors.

Factor Analysis

We also want to ask for an oblique **rotation** because the factors may be correlated. Oblique rotation options: "promax", "oblimin", "simplimax", "bentlerQ, and "geominQ". The orthogonal rotation options: "none", "varimax", "quartimax", "bentlerT", and "geominT".

We also want to ask for Principal Axis Factoring, `fm="pa"`. The other **estimate techniques** available are `fm="minres"`, `fm=""wls"`, `fm="gls"` and `fm="ml"`. Read the documentation to find out more.

```
fa_6 <- fa(r_matrix, nfactors = 6, rotate = "oblimin", fm = "pa")

fa_6
```

```
## Factor Analysis using method =  pa
## Call: fa(r = r_matrix, nfactors = 6, rotate = "oblimin", fm = "pa")
## Standardized loadings (pattern matrix) based upon correlation matrix
```

```
##      PA2   PA3   PA1   PA5   PA4   PA6   h2   u2 com
## A1  0.09  0.08 -0.09 -0.56  0.06  0.30 0.34 0.66 1.7
## A2  0.04  0.08 -0.04  0.68  0.00 -0.05 0.50 0.50 1.1
## A3 -0.02  0.03 -0.12  0.61  0.07  0.11 0.51 0.49 1.2
## A4 -0.07  0.19 -0.06  0.39 -0.10  0.15 0.28 0.72 2.1
## A5 -0.16  0.01 -0.19  0.45  0.13  0.21 0.47 0.53 2.3
## C1  0.01  0.54  0.06 -0.06  0.19  0.07 0.34 0.66 1.3
## C2  0.07  0.67  0.14  0.02  0.11  0.17 0.49 0.51 1.3
## C3  0.01  0.55  0.06  0.08 -0.04  0.06 0.31 0.69 1.1
## C4  0.06 -0.64  0.09 -0.06  0.07  0.30 0.57 0.43 1.5
## C5  0.15 -0.54  0.18 -0.01  0.11  0.05 0.43 0.57 1.5
## E1 -0.13  0.11  0.59 -0.12 -0.08  0.09 0.39 0.61 1.3
## E2  0.05 -0.02  0.70 -0.07 -0.06  0.03 0.56 0.44 1.0
## E3  0.00  0.00 -0.34  0.15  0.40  0.21 0.48 0.52 2.8
## E4 -0.05  0.03 -0.53  0.20  0.04  0.29 0.55 0.45 1.9
## E5  0.15  0.27 -0.40  0.05  0.23  0.01 0.40 0.60 2.8
## N1  0.84  0.01 -0.10 -0.07 -0.05  0.00 0.68 0.32 1.0
## N2  0.83  0.02 -0.06 -0.04 -0.01 -0.07 0.66 0.34 1.0
## N3  0.67 -0.03  0.13  0.07  0.05  0.08 0.54 0.46 1.1
## N4  0.43 -0.13  0.42  0.08  0.10  0.05 0.49 0.51 2.4
## N5  0.44  0.00  0.23  0.18 -0.10  0.16 0.35 0.65 2.4
## O1 -0.05  0.07 -0.04 -0.05  0.57  0.03 0.35 0.65 1.1
## O2  0.11 -0.07  0.00  0.09 -0.36  0.36 0.29 0.71 2.4
## O3 -0.02  0.02 -0.09  0.03  0.65 -0.02 0.48 0.52 1.1
## O4  0.09 -0.02  0.35  0.15  0.37 -0.04 0.25 0.75 2.5
## O5  0.03 -0.02 -0.04 -0.04 -0.44  0.41 0.37 0.63 2.0
##
##                        PA2  PA3  PA1  PA5  PA4  PA6
## SS loadings           2.48 2.05 2.17 1.88 1.68 0.82
## Proportion Var        0.10 0.08 0.09 0.08 0.07 0.03
## Cumulative Var        0.10 0.18 0.27 0.34 0.41 0.44
## Proportion Explained  0.22 0.18 0.20 0.17 0.15 0.07
## Cumulative Proportion 0.22 0.41 0.60 0.77 0.93 1.00
##
##  With factor correlations of
##       PA2   PA3   PA1   PA5   PA4   PA6
## PA2  1.00 -0.18  0.25 -0.10  0.02  0.16
## PA3 -0.18  1.00 -0.21  0.19  0.19 -0.02
## PA1  0.25 -0.21  1.00 -0.30 -0.20 -0.08
## PA5 -0.10  0.19 -0.30  1.00  0.25  0.14
## PA4  0.02  0.19 -0.20  0.25  1.00  0.02
## PA6  0.16 -0.02 -0.08  0.14  0.02  1.00
##
## Mean item complexity =  1.7
## Test of the hypothesis that 6 factors are sufficient.
```

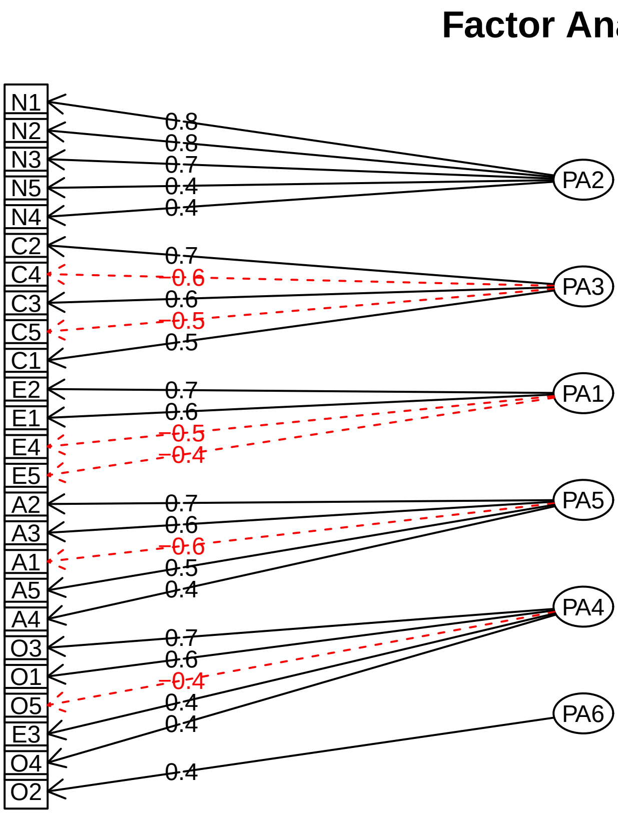
```
##
## The degrees of freedom for the null model are  300  and the objective function was  7
## The degrees of freedom for the model are 165  and the objective function was  0.37
##
## The root mean square of the residuals (RMSR) is  0.02
## The df corrected root mean square of the residuals is  0.03
##
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                   PA2  PA3  PA1  PA5  PA4
## Correlation of scores with factors              0.93 0.88 0.89 0.87 0.86
## Multiple R square of scores with factors        0.86 0.78 0.79 0.76 0.73
## Minimum correlation of possible factor scores  0.72 0.57 0.59 0.53 0.46
##                                                   PA6
## Correlation of scores with factors              0.77
## Multiple R square of scores with factors        0.59
## Minimum correlation of possible factor scores  0.17
```

We can see that the 6th factor is kind of a junk factor. But, pay close attention to the items that want to load on this factor. It would be a good idea to open up the dataset's help file and read those items, `?bfi`.

Factor Diagram

We can see a diagram of this factor structure with `fa.diagram()`.

```r
fa.diagram(fa_6)
```

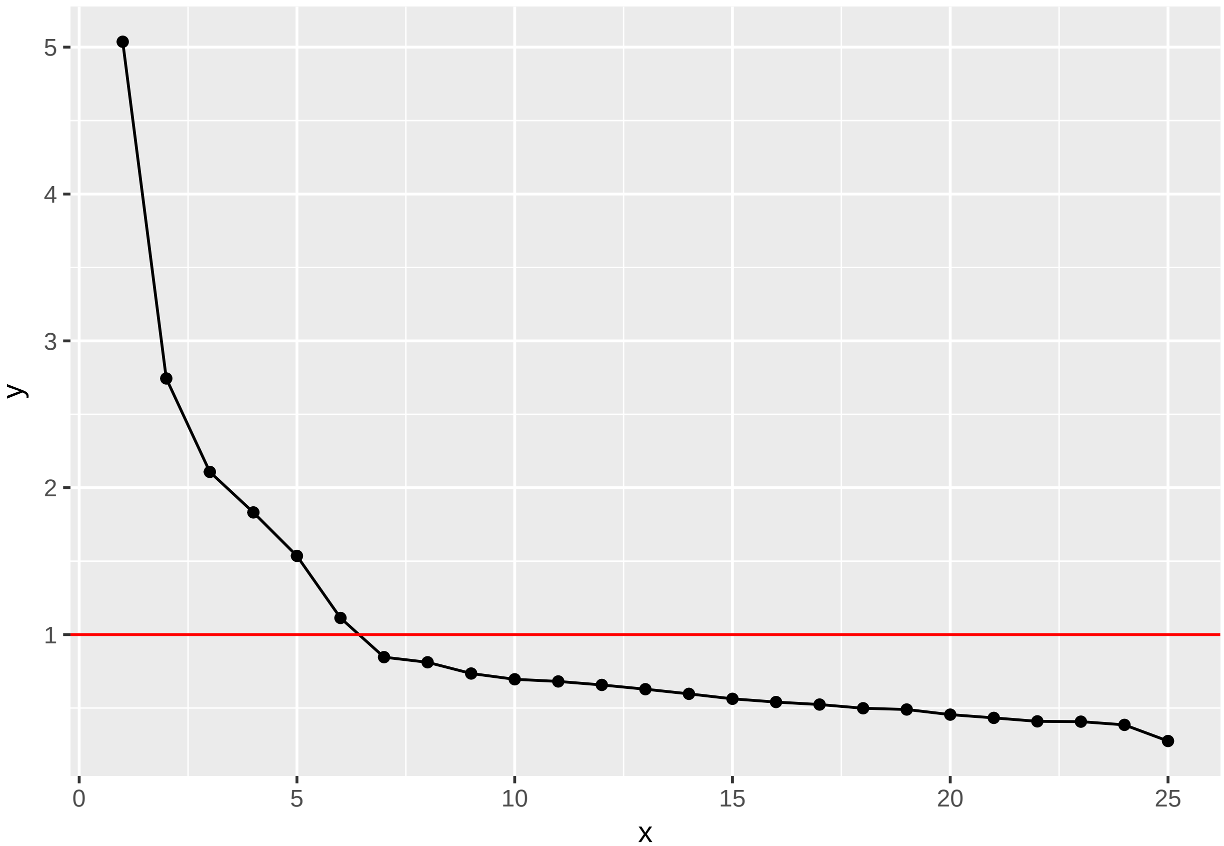


You can see that MR 6 only has the one openness to experience item, so maybe we should try 5 factors instead of 6.

Scree Plot

For more evidence, we could look at a scree plot—a plot of the eigenvalues.

```
plotdata <- data.frame(x = seq(1:25), y = (eigen(r_matrix)$values))

ggplot(plotdata, aes(x = x, y = y)) +
  geom_point() +
  geom_line() +
  geom_hline(yintercept = 1, color = "red")
```



Five-Factor Solution

Let's try the 5 factor solution.

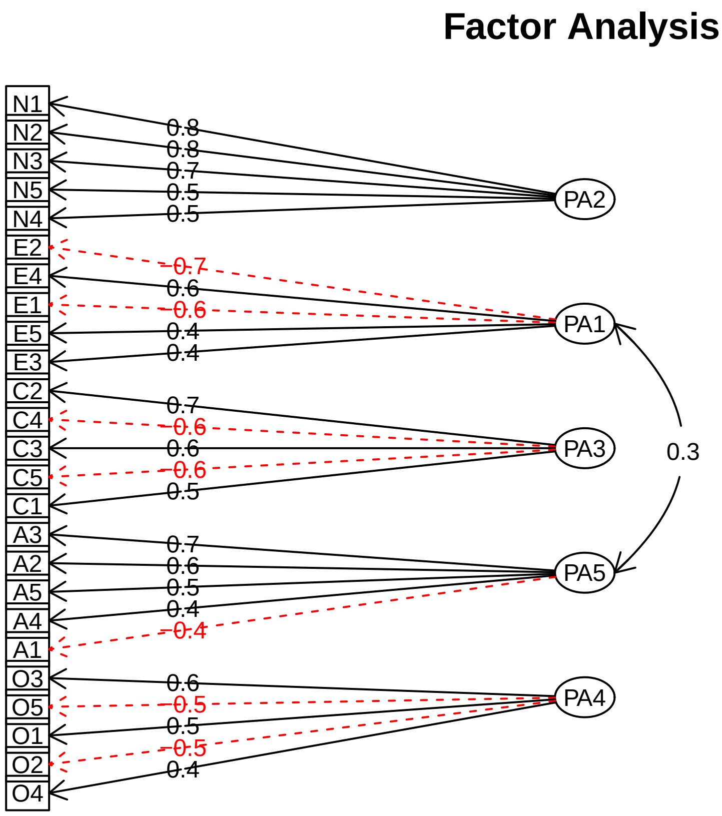
```
fa_5 <- fa(r_matrix, nfactors = 5, rotate = "oblimin", fm = "pa")

fa_5
```

```
## Factor Analysis using method =  pa
## Call: fa(r = r_matrix, nfactors = 5, rotate = "oblimin", fm = "pa")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      PA2   PA1   PA3   PA5   PA4   h2   u2 com
## A1  0.21  0.17  0.07 -0.41 -0.06 0.19 0.81 2.0
## A2 -0.02  0.00  0.08  0.64  0.03 0.45 0.55 1.0
## A3 -0.03  0.12  0.02  0.66  0.03 0.52 0.48 1.1
## A4 -0.06  0.06  0.19  0.43 -0.15 0.28 0.72 1.7
## A5 -0.11  0.23  0.01  0.53  0.04 0.46 0.54 1.5
## C1  0.07 -0.03  0.55 -0.02  0.15 0.33 0.67 1.2
## C2  0.15 -0.09  0.67  0.08  0.04 0.45 0.55 1.2
## C3  0.03 -0.06  0.57  0.09 -0.07 0.32 0.68 1.1
## C4  0.17  0.00 -0.61  0.04 -0.05 0.45 0.55 1.2
## C5  0.19 -0.14 -0.55  0.02  0.09 0.43 0.57 1.4
```

```
## E1 -0.06 -0.56  0.11 -0.08 -0.10 0.35 0.65 1.2
## E2  0.10 -0.68 -0.02 -0.05 -0.06 0.54 0.46 1.1
## E3  0.08  0.42  0.00  0.25  0.28 0.44 0.56 2.6
## E4  0.01  0.59  0.02  0.29 -0.08 0.53 0.47 1.5
## E5  0.15  0.42  0.27  0.05  0.21 0.40 0.60 2.6
## N1  0.81  0.10  0.00 -0.11 -0.05 0.65 0.35 1.1
## N2  0.78  0.04  0.01 -0.09  0.01 0.60 0.40 1.0
## N3  0.71 -0.10 -0.04  0.08  0.02 0.55 0.45 1.1
## N4  0.47 -0.39 -0.14  0.09  0.08 0.49 0.51 2.3
## N5  0.49 -0.20  0.00  0.21 -0.15 0.35 0.65 2.0
## O1  0.02  0.10  0.07  0.02  0.51 0.31 0.69 1.1
## O2  0.19  0.06 -0.08  0.16 -0.46 0.26 0.74 1.7
## O3  0.03  0.15  0.02  0.08  0.61 0.46 0.54 1.2
## O4  0.13 -0.32 -0.02  0.17  0.37 0.25 0.75 2.7
## O5  0.13  0.10 -0.02  0.04 -0.54 0.30 0.70 1.2
##
##                        PA2  PA1  PA3  PA5  PA4
## SS loadings           2.57 2.20 2.03 1.98 1.59
## Proportion Var        0.10 0.09 0.08 0.08 0.06
## Cumulative Var        0.10 0.19 0.27 0.35 0.41
## Proportion Explained  0.25 0.21 0.20 0.19 0.15
## Cumulative Proportion 0.25 0.46 0.66 0.85 1.00
##
##  With factor correlations of
##       PA2   PA1   PA3   PA5   PA4
## PA2  1.00 -0.21 -0.19 -0.04 -0.01
## PA1 -0.21  1.00  0.23  0.33  0.17
## PA3 -0.19  0.23  1.00  0.20  0.19
## PA5 -0.04  0.33  0.20  1.00  0.19
## PA4 -0.01  0.17  0.19  0.19  1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 5 factors are sufficient.
##
## The degrees of freedom for the null model are  300  and the objective function was  7
## The degrees of freedom for the model are 185  and the objective function was  0.65
##
## 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 = 0.98
## Measures of factor score adequacy
##                                                    PA2  PA1  PA3  PA5  PA4
## Correlation of scores with factors               0.92 0.89 0.88 0.88 0.84
## Multiple R square of scores with factors         0.85 0.79 0.77 0.77 0.71
```

```
## Minimum correlation of possible factor scores  0.70 0.59 0.54 0.54 0.42
```

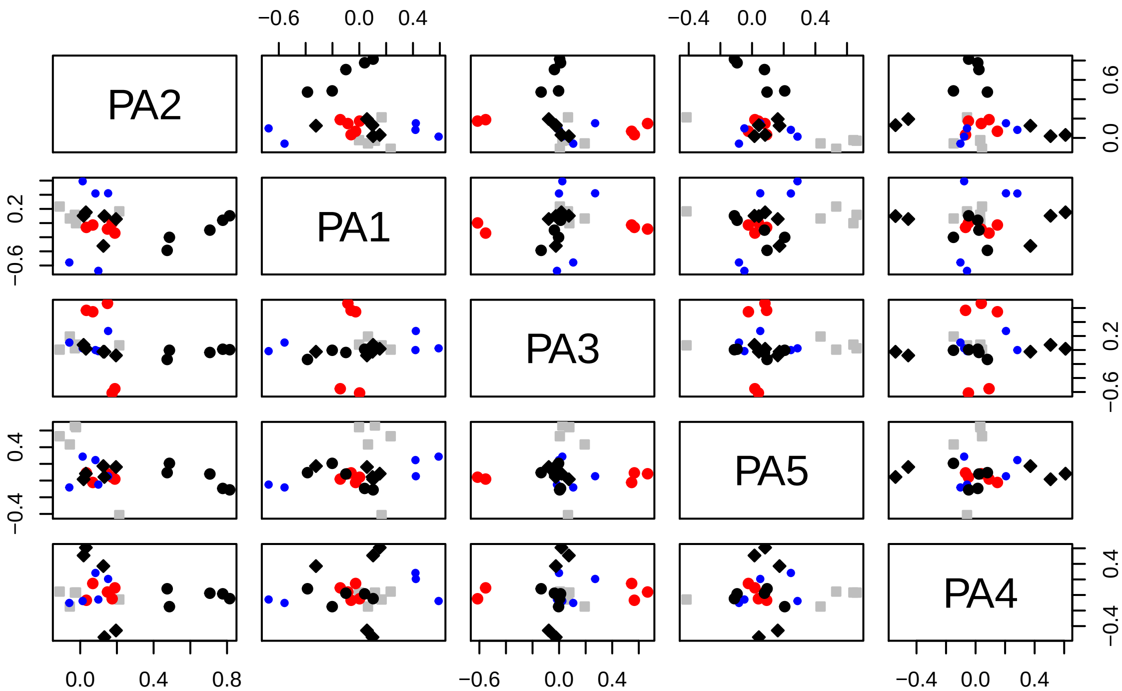
```r
fa.diagram(fa_5)
```



Lastly, you might want a factor loading plot. It is hard to interpret with so many factors, but with two factors it would look nice.

```r
plot(fa_5)
```

Factor Analysis



The items seem to be factoring as we would expect.

Principal Components Analysis

We might want to try some other estimation techniques to compare.

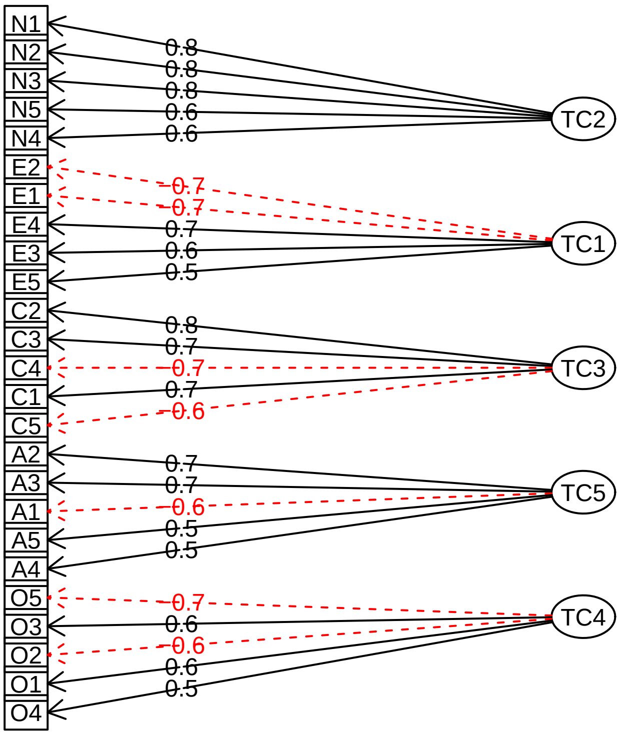
```
pca <- principal(r_matrix, 5, rotate = "oblimin")

pca
```

```
## Principal Components Analysis
## Call: principal(r = r_matrix, nfactors = 5, rotate = "oblimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      TC2   TC1   TC3   TC5   TC4   h2   u2 com
## A1  0.23  0.23  0.11 -0.64 -0.10 0.46 0.54 1.7
## A2  0.02  0.09  0.09  0.71  0.02 0.57 0.43 1.1
## A3  0.01  0.25  0.04  0.67  0.01 0.59 0.41 1.3
## A4 -0.05  0.10  0.23  0.51 -0.21 0.41 0.59 1.9
## A5 -0.10  0.36  0.02  0.54  0.01 0.54 0.46 1.8
## C1  0.07 -0.01  0.66 -0.06  0.17 0.47 0.53 1.2
## C2  0.16 -0.06  0.76  0.06  0.04 0.58 0.42 1.1
## C3  0.03 -0.08  0.69  0.09 -0.09 0.47 0.53 1.1
## C4  0.24  0.04 -0.67  0.02 -0.07 0.54 0.46 1.3
```

```
## C5  0.27 -0.11 -0.61  0.02  0.11 0.52 0.48 1.5
## E1 -0.03 -0.69  0.15 -0.04 -0.07 0.48 0.52 1.1
## E2  0.18 -0.73  0.00 -0.02 -0.02 0.60 0.40 1.1
## E3  0.10  0.59  0.00  0.20  0.26 0.53 0.47 1.7
## E4 -0.02  0.69  0.03  0.23 -0.13 0.61 0.39 1.3
## E5  0.14  0.54  0.30 -0.01  0.19 0.50 0.50 2.0
## N1  0.82  0.11  0.00 -0.15 -0.06 0.69 0.31 1.1
## N2  0.81  0.06  0.01 -0.13  0.01 0.66 0.34 1.1
## N3  0.79 -0.04 -0.02  0.04  0.02 0.63 0.37 1.0
## N4  0.59 -0.36 -0.12  0.10  0.10 0.57 0.43 1.9
## N5  0.61 -0.18  0.02  0.22 -0.18 0.48 0.52 1.7
## O1  0.05  0.22  0.09 -0.02  0.59 0.44 0.56 1.3
## O2  0.24  0.07 -0.07  0.15 -0.59 0.43 0.57 1.5
## O3  0.06  0.29  0.03  0.08  0.64 0.55 0.45 1.5
## O4  0.22 -0.32 -0.02  0.27  0.49 0.44 0.56 2.8
## O5  0.14  0.08 -0.01  0.00 -0.68 0.48 0.52 1.1
##
##                        TC2  TC1  TC3  TC5  TC4
## SS loadings           3.13 3.01 2.62 2.36 2.13
## Proportion Var        0.13 0.12 0.10 0.09 0.09
## Cumulative Var        0.13 0.25 0.35 0.45 0.53
## Proportion Explained  0.24 0.23 0.20 0.18 0.16
## Cumulative Proportion 0.24 0.46 0.66 0.84 1.00
##
##  With component correlations of
##       TC2   TC1   TC3   TC5   TC4
## TC2  1.00 -0.14 -0.13 -0.05 -0.01
## TC1 -0.14  1.00  0.20  0.22  0.09
## TC3 -0.13  0.20  1.00  0.14  0.12
## TC5 -0.05  0.22  0.14  1.00  0.10
## TC4 -0.01  0.09  0.12  0.10  1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 5 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.06
##
## Fit based upon off diagonal values = 0.92
```

```r
fa.diagram(pca)
```

Factor Analysis



Things look pretty similar with PCA, this is typical, but I would recommend using Principal Axis Factoring, `fm = "pa"`. Read more about this is Preacher & Maccallum (2003).