A presentation from the Telemetrics Lab

Telemetrics lab

Department of Psychology
Northwestern University
Evanston, Illinois USA

November, 2010

Outline

Introduction

- Factor analysis – several examples
  - Data from a correlation matrix
    - Simulated 2 factor data
    - Real data – Ability tests
  - Raw data
    - Simulated 2 factor data
    - Real data – 5 Personality dimensions

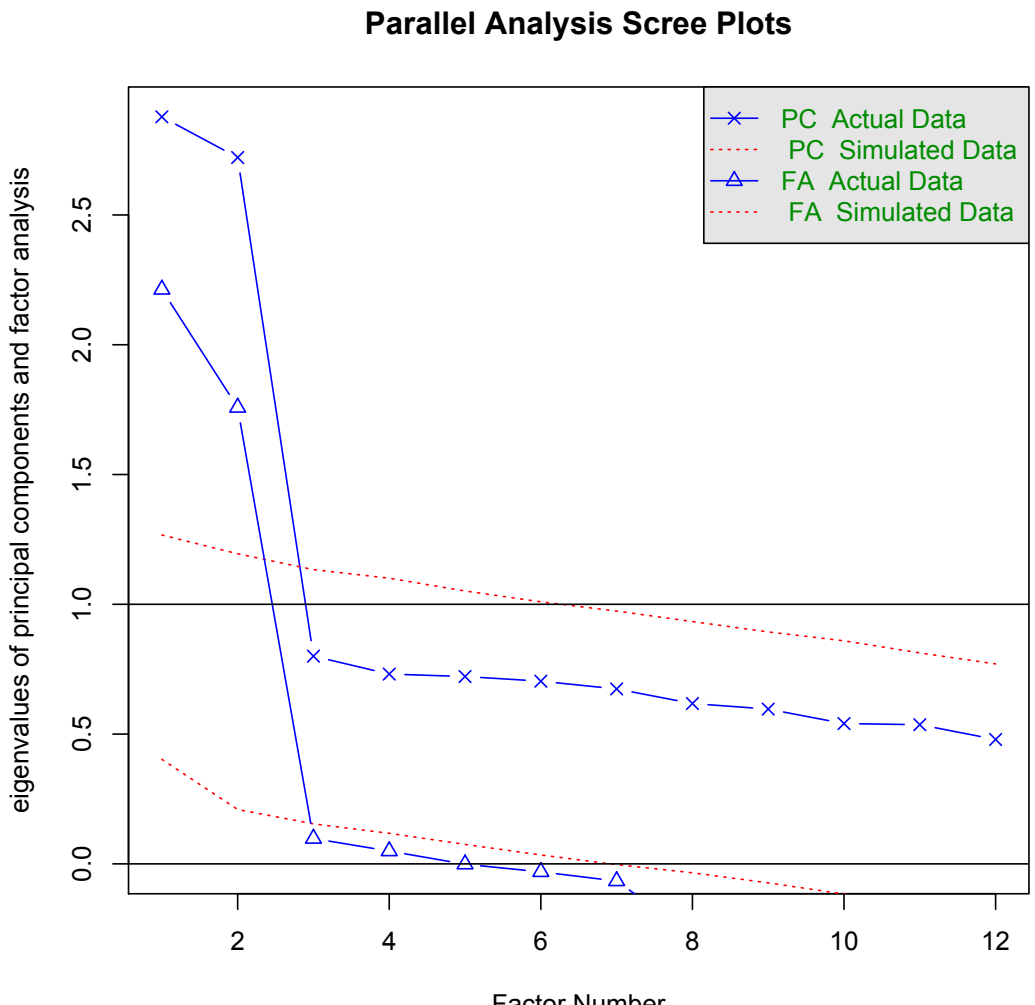
Simulate 2 factor data

Using the sim.item function

```
>  set.seed(42)  #to generate a reproducible example
>  my.data <- sim.item(12)
>  my.cor <- cor(my.data)
> round(my.cor,2)
       V1    V2    V3    V4    V5    V6    V7    V8    V9   V10   V11   V12
V1   1.00  0.36  0.38 -0.01  0.05  0.03 -0.35 -0.40 -0.41  0.06  0.02  0.01
V2   0.36  1.00  0.37 -0.04 -0.02  0.01 -0.37 -0.34 -0.36  0.07  0.03  0.01
V3   0.38  0.37  1.00 -0.01  0.01  0.01 -0.38 -0.39 -0.32  0.01  0.05 -0.11
V4  -0.01 -0.04 -0.01  1.00  0.34  0.37 -0.09  0.00  0.00 -0.33 -0.37 -0.31
V5   0.05 -0.02  0.01  0.34  1.00  0.35 -0.01  0.08  0.02 -0.32 -0.35 -0.30
V6   0.03  0.01  0.01  0.37  0.35  1.00 -0.05  0.11 -0.03 -0.39 -0.32 -0.33
V7  -0.35 -0.37 -0.38 -0.09 -0.01 -0.05  1.00  0.34  0.32 -0.04  0.02  0.08
V8  -0.40 -0.34 -0.39  0.00  0.08  0.11  0.34  1.00  0.39 -0.11 -0.12 -0.02
V9  -0.41 -0.36 -0.32  0.00  0.02 -0.03  0.32  0.39  1.00 -0.06 -0.01  0.00
V10  0.06  0.07  0.01 -0.33 -0.32 -0.39 -0.04 -0.11 -0.06  1.00  0.41  0.36
V11  0.02  0.03  0.05 -0.37 -0.35 -0.32  0.02 -0.12 -0.01  0.41  1.00  0.39
V12  0.01  0.01 -0.11 -0.31 -0.30 -0.33  0.08 -0.02  0.00  0.36  0.39  1.00
```

How many factors in my.cor

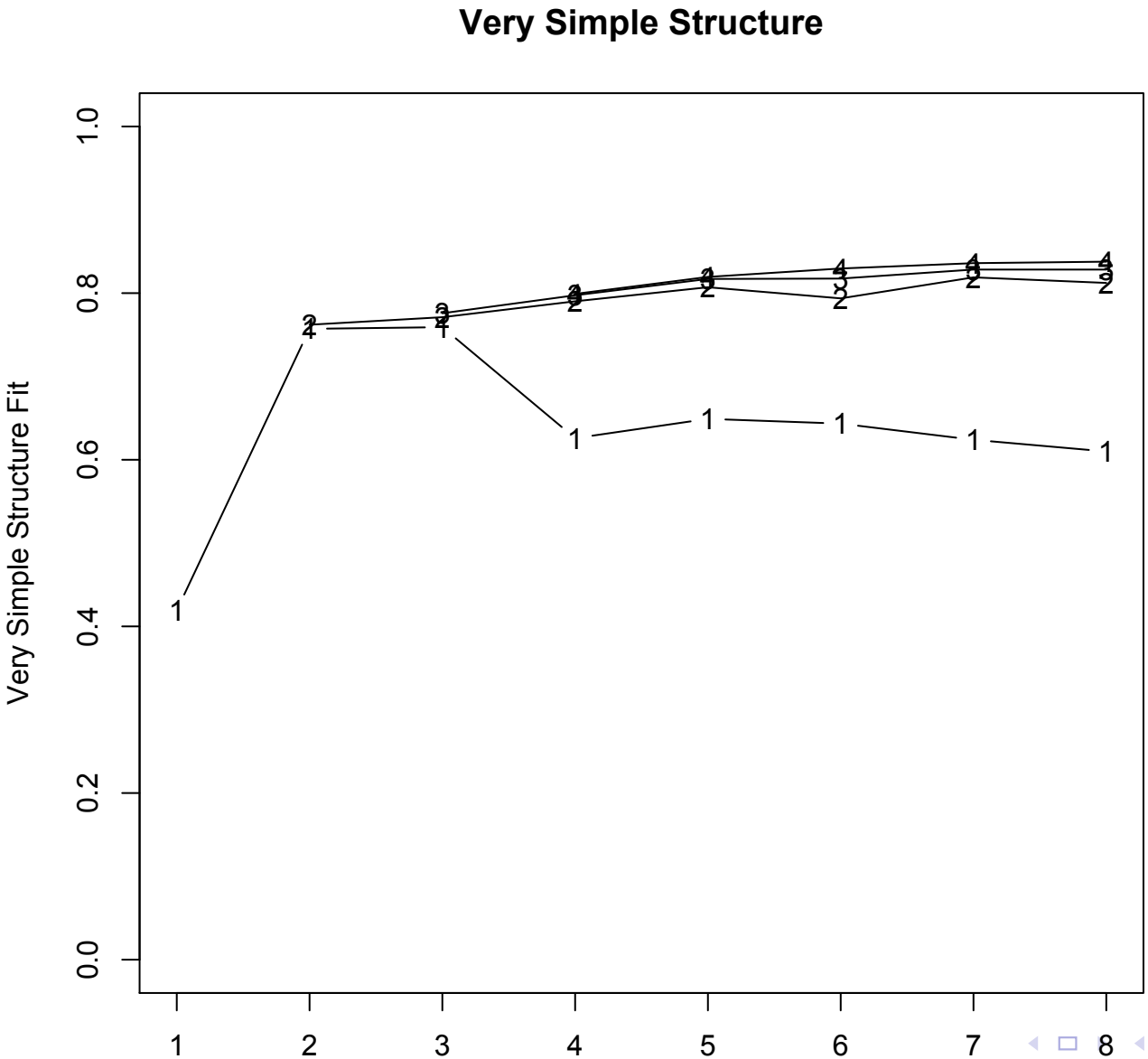
```
> fa.parallel(my.cor,n.obs=500)
Parallel analysis suggests that the number of factors =  2
                 and the number of components =  2
```



Try Very Simple Structure as well as MAP

```
> vss(my.cor,n.obs=500)

Very Simple Structure
Call: VSS(x = x, n = n, rotate = rotate, diagonal = diagonal, fm = fm,
    n.obs = n.obs, plot = plot, title = title)
VSS complexity 1 achieves a maximimum of 0.74  with  3  factors
VSS complexity 2 achieves a maximimum of 0.8  with  8  factors

The Velicer MAP criterion achieves a minimum of 0.02  with  2  factors

Velicer MAP
[1] 0.05 0.02 0.03 0.05 0.07 0.10 0.13 0.19

Very Simple Structure Complexity 1
[1] 0.39 0.74 0.74 0.63 0.70 0.66 0.58 0.57

Very Simple Structure Complexity 2
[1] 0.00 0.75 0.76 0.78 0.79 0.79 0.80 0.80
```

Examine the output



Extract 2 factors –part 1

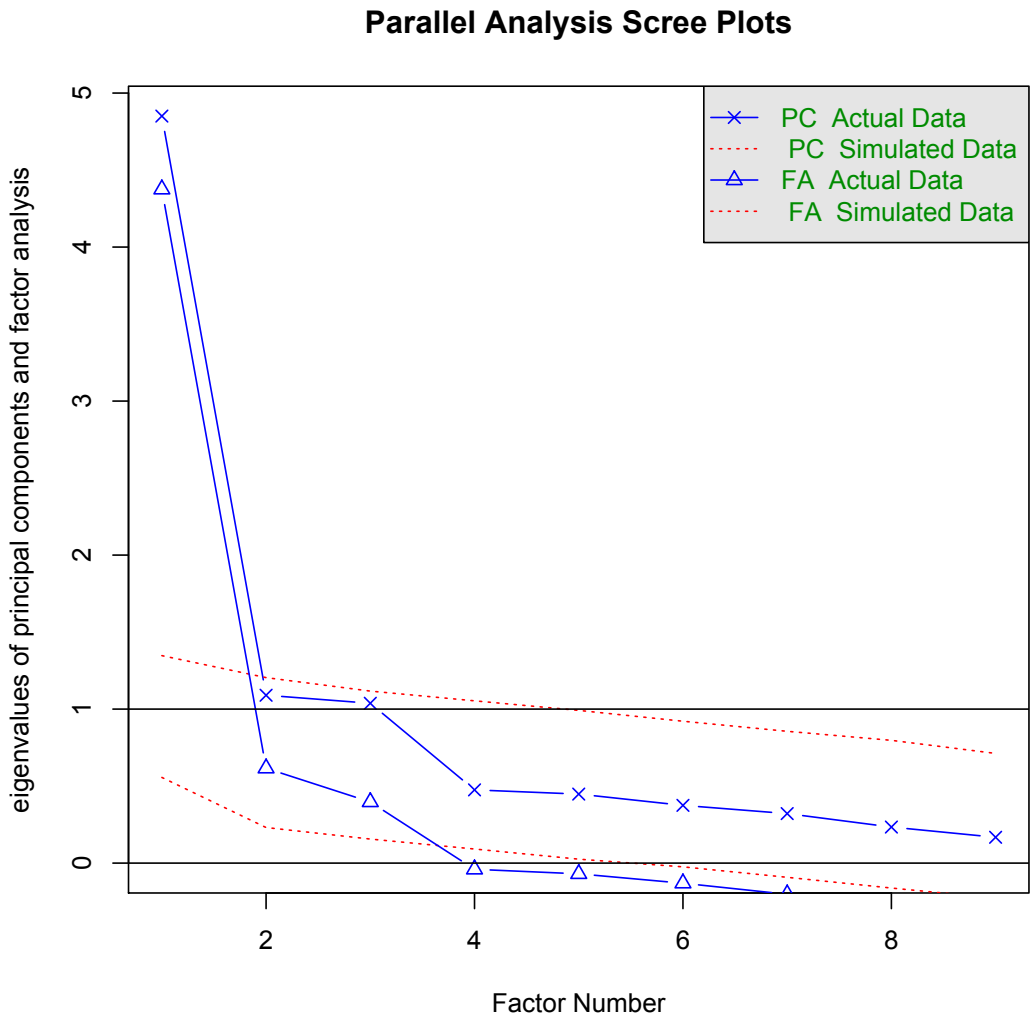
```
> fa(my.cor,2,n.obs=500)
Factor Analysis using method =  minres
Call: fa(r = my.cor, nfactors = 2, n.obs = 500)
Standardized loadings based upon correlation matrix
      MR1   MR2   h2   u2
V1   0.64 -0.02 0.41 0.59
V2   0.59  0.02 0.35 0.65
V3   0.61 -0.04 0.37 0.63
V4   0.03 -0.58 0.34 0.66
V5   0.01 -0.55 0.30 0.70
V6   0.03 -0.60 0.36 0.64
V7  -0.58  0.08 0.34 0.66
V8  -0.62 -0.10 0.40 0.60
V9  -0.59  0.00 0.35 0.65
V10  0.07  0.61 0.39 0.61
V11  0.03  0.63 0.39 0.61
V12 -0.06  0.57 0.33 0.67

                       MR1  MR2
SS loadings           2.21 2.12
Proportion Var        0.18 0.18
Cumulative Var        0.18 0.36
```

2 artificial factors part 2

```
 With factor correlations of
     MR1  MR2
MR1 1.00 0.04
MR2 0.04 1.00

Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are  66  and the objective function w
The degrees of freedom for the model are 43  and the objective function was  0.

The root mean square of the residuals is  0.02
The df corrected root mean square of the residuals is  0.03
The number of observations was  500  with Chi Square =  54.56  with prob <  0.1

Tucker Lewis Index of factoring reliability =  0.985
RMSEA index =  0.024  and the 90 % confidence intervals are  0.023 0.026
BIC =  -212.67
Fit based upon off diagonal values = 0.99
Measures of factor score adequacy
                                                MR1  MR2
Correlation of scores with factors             0.88 0.88
Multiple R square of scores with factors       0.78 0.77
Minimum correlation of possible factor scores  0.56 0.53
```

9 mental tests from Thurstone

```
data(bifactor)
fa.parallel(Thurstone,n.obs=213)
```



Exract 3 factors

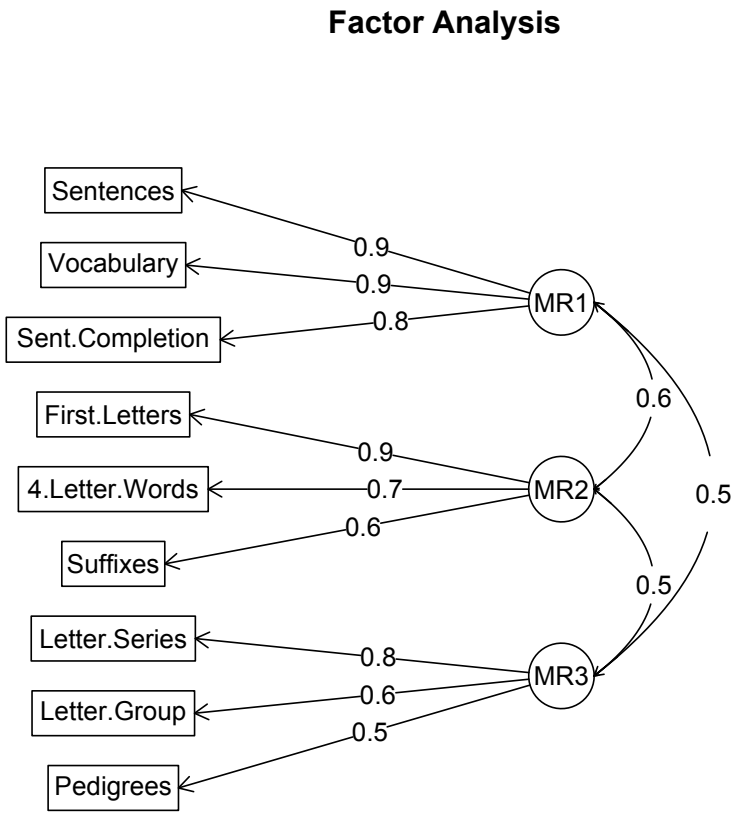
```
> fa3 <-  fa(Thurstone,3,n.obs=213)
> fa3
Factor Analysis using method =  minres
Call: fa(r = Thurstone, nfactors = 3, n.obs = 213)
Standardized loadings based upon correlation matrix
                   MR1   MR2   MR3   h2   u2
Sentences         0.91 -0.04  0.04 0.82 0.18
Vocabulary        0.89  0.06 -0.03 0.84 0.16
Sent.Completion   0.83  0.04  0.00 0.73 0.27
First.Letters     0.00  0.86  0.00 0.73 0.27
4.Letter.Words   -0.01  0.74  0.10 0.63 0.37
Suffixes          0.18  0.63 -0.08 0.50 0.50
Letter.Series     0.03 -0.01  0.84 0.72 0.28
Pedigrees         0.37 -0.05  0.47 0.50 0.50
Letter.Group     -0.06  0.21  0.64 0.53 0.47

                 MR1  MR2  MR3
SS loadings    2.64 1.86 1.50
Proportion Var 0.29 0.21 0.17
Cumulative Var 0.29 0.50 0.67
```

Thurstone 3 factors part 2

```
 With factor correlations of
     MR1  MR2  MR3
MR1 1.00 0.59 0.54
MR2 0.59 1.00 0.52
MR3 0.54 0.52 1.00

Test of the hypothesis that 3 factors are sufficient.

The degrees of freedom for the null model are  36  and the objective function w
The degrees of freedom for the model are 12  and the objective function was  0.

The root mean square of the residuals is  0
The df corrected root mean square of the residuals is  0.01
The number of observations was  213  with Chi Square =  2.82  with prob <  1

Tucker Lewis Index of factoring reliability =  1.027
RMSEA index =  0  and the 90 % confidence intervals are  0 0.023
BIC =  -61.51
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                MR1  MR2  MR3
Correlation of scores with factors             0.96 0.92 0.90
Multiple R square of scores with factors       0.93 0.85 0.81
```

A factor diagram

```
fa3 <-  fa(Thurstone,3,n.obs=213)
```

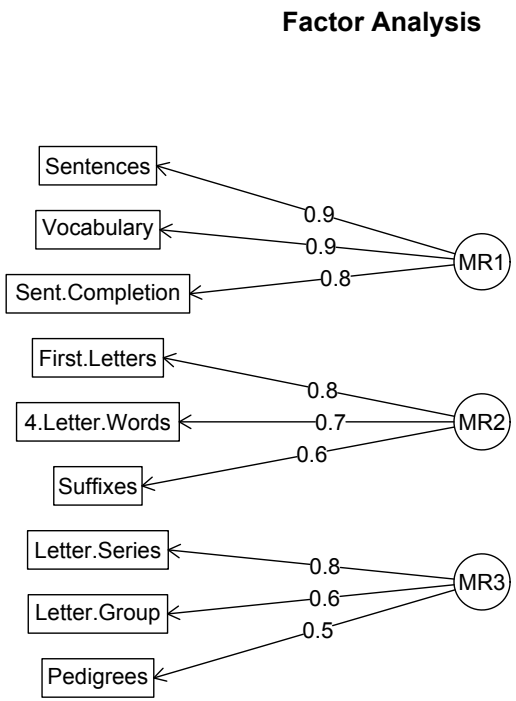


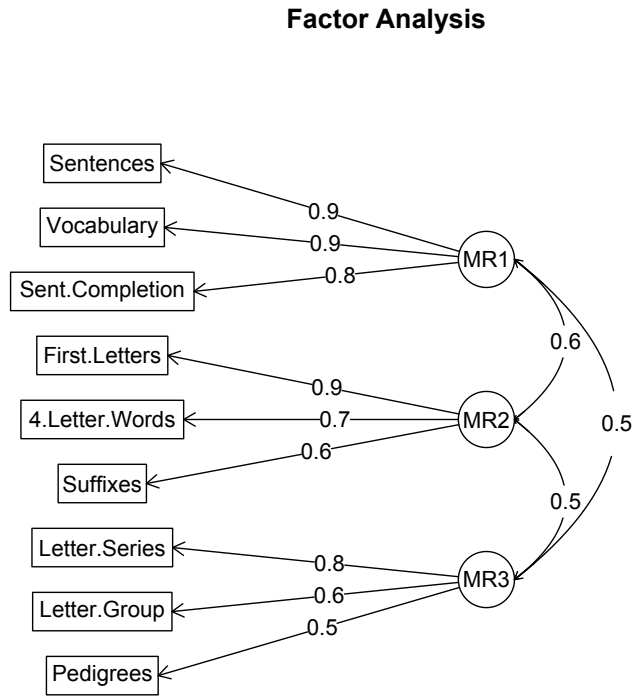
Thurstone, 3 factors Varimax rotated

```
> v3 <- fa(Thurstone,3,rotate="Varimax",n.obs=213)
> fa.diagram(v3)
> v3
Factor Analysis using method =  minres
Call: fa(r = Thurstone, nfactors = 3, n.obs = 213, rotate = "Varimax")
Standardized loadings based upon correlation matrix
                   MR1  MR2  MR3   h2   u2
Sentences         0.86 0.20 0.22 0.82 0.18
Vocabulary        0.85 0.27 0.18 0.84 0.16
Sent.Completion   0.80 0.24 0.19 0.73 0.27
First.Letters     0.29 0.78 0.20 0.73 0.27
4.Letter.Words    0.27 0.70 0.26 0.63 0.37
Suffixes          0.36 0.60 0.10 0.50 0.50
Letter.Series     0.28 0.18 0.78 0.72 0.28
Pedigrees         0.48 0.15 0.50 0.50 0.50
Letter.Group      0.20 0.32 0.62 0.53 0.47

                 MR1  MR2  MR3
SS loadings     2.73 1.78 1.48
Proportion Var  0.30 0.20 0.16
Cumulative Var  0.30 0.50 0.67
```

Compare the two solutions

```
> v3 <- fa(Thurstone,3,rotate="Varimax",n.obs=213)
> fa.diagram(v3)
```






```
> fa.diagram(v3)
```

R has many built in data sets

- data(bfi)
- 25 personality items from the Big 5
  - Collected as part of the SAPA project
- Thought to represent 5 dimensions
  - Agreeableness
  - Extraversion
  - Conscientiousness
  - Extraversion
  - Neuroticism

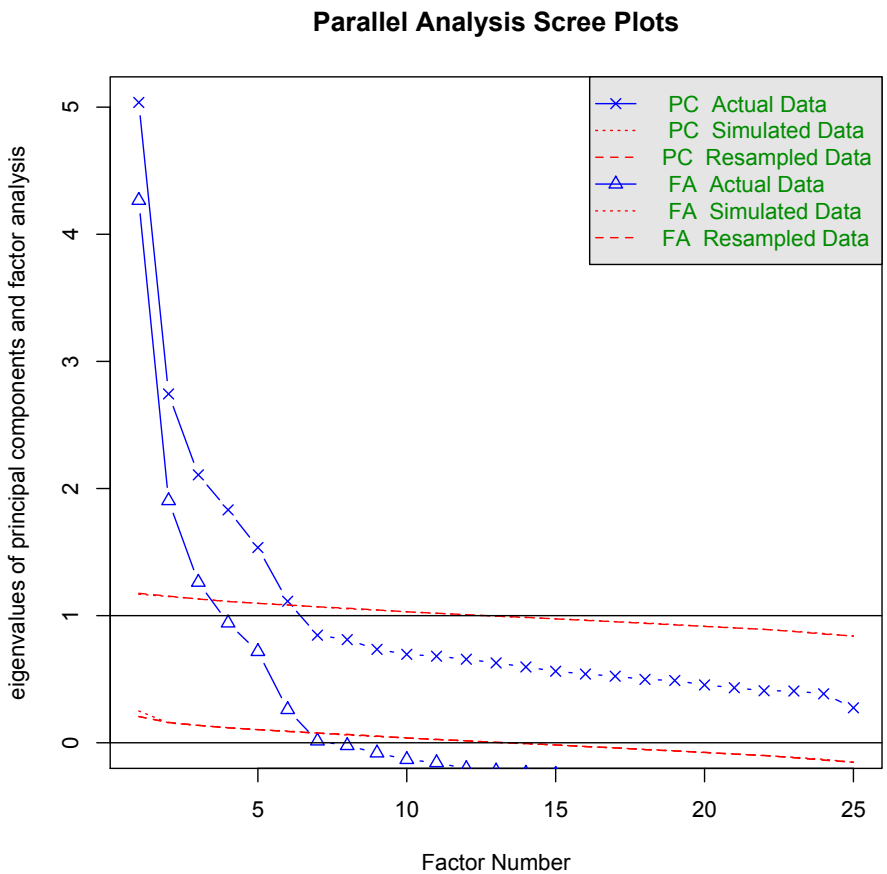
Describe the Big 5

```
> data(bfi)
> describe(bfi)
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
A1	1	2784	2.41	1.41	2	2.23	1.48	1	6	5	0.83	-0.31	0.03
A2	2	2773	4.80	1.17	5	4.98	1.48	1	6	5	-1.12	1.05	0.02
A3	3	2774	4.60	1.30	5	4.79	1.48	1	6	5	-1.00	0.44	0.02
A4	4	2781	4.70	1.48	5	4.93	1.48	1	6	5	-1.03	0.04	0.03
A5	5	2784	4.56	1.26	5	4.71	1.48	1	6	5	-0.85	0.16	0.02
C1	6	2779	4.50	1.24	5	4.64	1.48	1	6	5	-0.85	0.30	0.02
C2	7	2776	4.37	1.32	5	4.50	1.48	1	6	5	-0.74	-0.14	0.03
C3	8	2780	4.30	1.29	5	4.42	1.48	1	6	5	-0.69	-0.13	0.02
C4	9	2774	2.55	1.38	2	2.41	1.48	1	6	5	0.60	-0.62	0.03
C5	10	2784	3.30	1.63	3	3.25	1.48	1	6	5	0.07	-1.22	0.03
E1	11	2777	2.97	1.63	3	2.86	1.48	1	6	5	0.37	-1.09	0.03
E2	12	2784	3.14	1.61	3	3.06	1.48	1	6	5	0.22	-1.15	0.03
E3	13	2775	4.00	1.35	4	4.07	1.48	1	6	5	-0.47	-0.47	0.03
E4	14	2791	4.42	1.46	5	4.59	1.48	1	6	5	-0.82	-0.30	0.03
E5	15	2779	4.42	1.33	5	4.56	1.48	1	6	5	-0.78	-0.09	0.03
N1	16	2778	2.93	1.57	3	2.82	1.48	1	6	5	0.37	-1.01	0.03
N2	17	2779	3.51	1.53	4	3.51	1.48	1	6	5	-0.08	-1.05	0.03
N3	18	2789	3.22	1.60	3	3.16	1.48	1	6	5	0.15	-1.18	0.03
N4	19	2764	3.19	1.57	3	3.12	1.48	1	6	5	0.20	-1.09	0.03
N5	20	2771	2.97	1.62	3	2.85	1.48	1	6	5	0.37	-1.06	0.03
O1	21	2778	4.82	1.13	5	4.96	1.48	1	6	5	-0.90	0.43	0.02
O2	22	2800	2.71	1.57	2	2.56	1.48	1	6	5	0.59	-0.81	0.03
O3	23	2772	4.44	1.22	5	4.56	1.48	1	6	5	-0.77	0.30	0.02
O4	24	2786	4.89	1.22	5	5.10	1.48	1	6	5	-1.22	1.08	0.02
O5	25	2780	2.49	1.33	2	2.34	1.48	1	6	5	0.74	-0.24	0.03
gender	26	2800	1.67	0.47	2	1.71	0.00	1	2	1	-0.73	-1.47	0.01
education	27	2577	3.19	1.11	3	3.22	1.48	1	5	4	-0.05	-0.32	0.02
age	28	2800	28.78	11.13	26	27.43	10.38	3	86	83	1.02	0.56	0.21

How many factors?

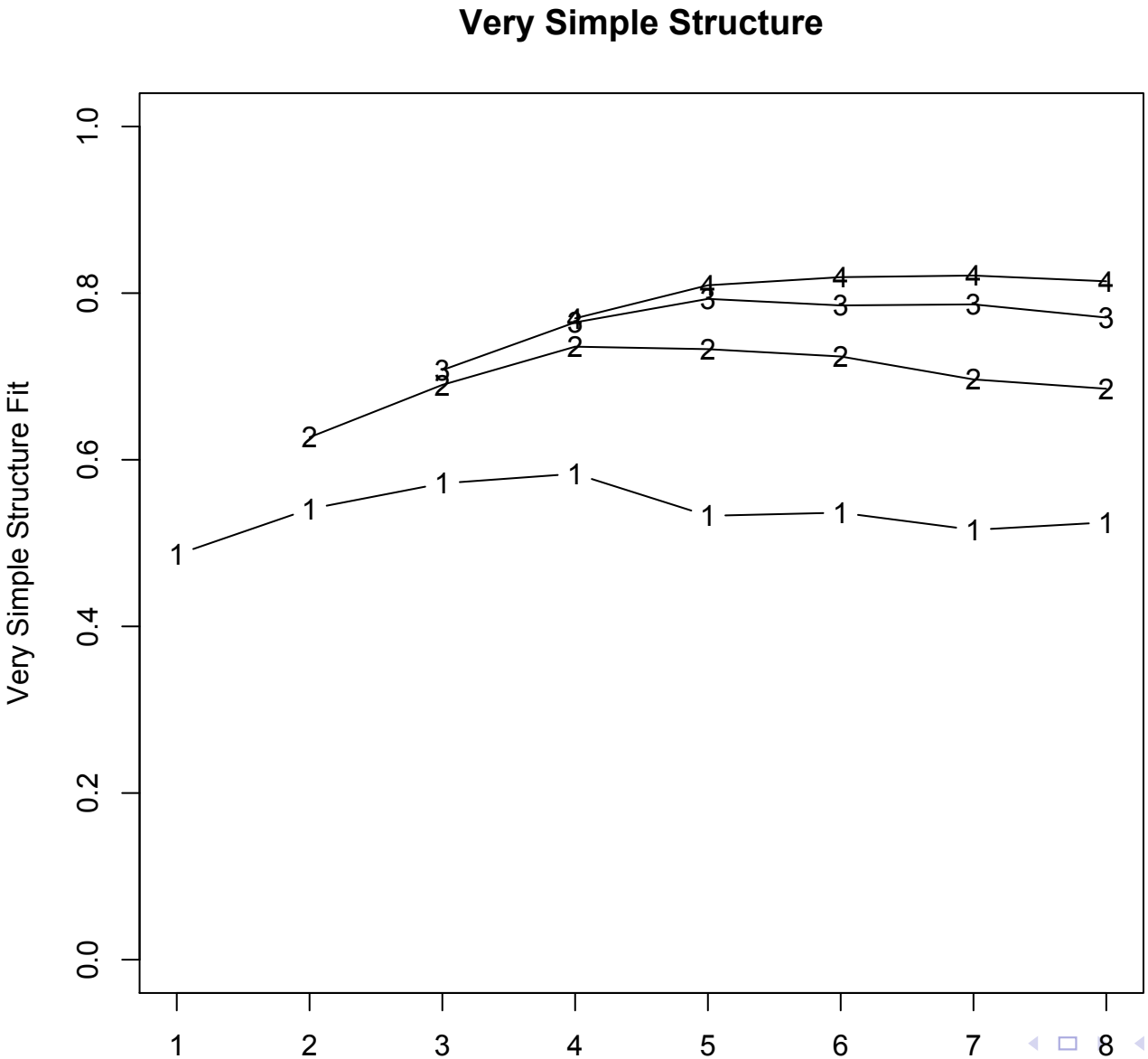
```
> fa.parallel(bfi[1:25]) #just  the items
Parallel analysis suggests that the number of factors =  6  and the number of c
```



How many factors part 2: VSS

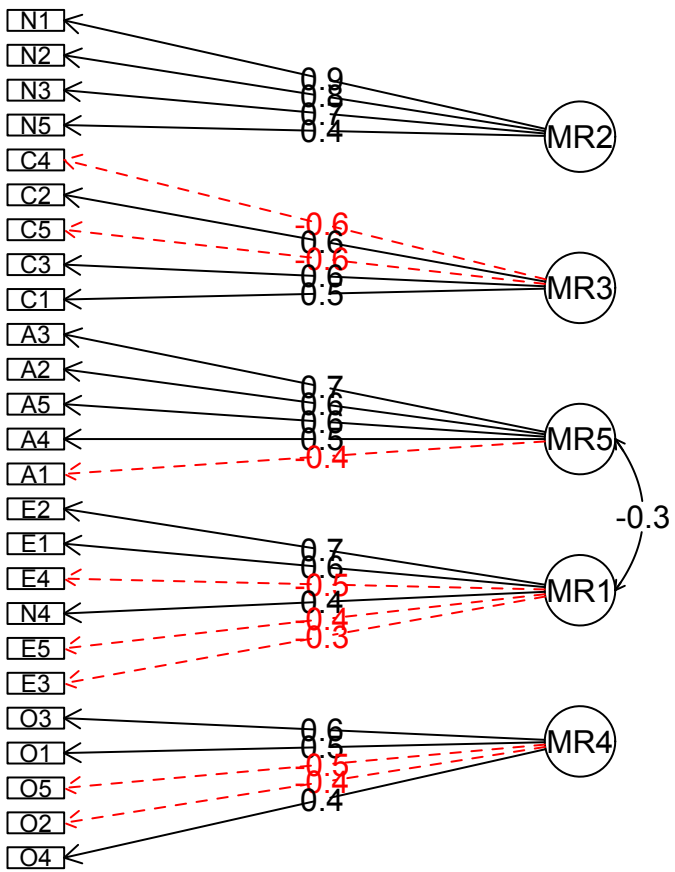
```
> VSS(bfi[1:25])

Very Simple Structure
Call: VSS(x = bfi[1:25])
VSS complexity 1 achieves a maximimum of 0.58  with  4  factors
VSS complexity 2 achieves a maximimum of 0.74  with  4  factors

The Velicer MAP criterion achieves a minimum of 0.01  with  5  factors

Velicer MAP
[1] 0.02 0.02 0.02 0.02 0.01 0.02 0.02 0.02

Very Simple Structure Complexity 1
[1] 0.49 0.54 0.57 0.58 0.53 0.54 0.52 0.52

Very Simple Structure Complexity 2
[1] 0.00 0.63 0.69 0.74 0.73 0.72 0.70 0.69
```

VSS plot


Very Simple Structure
Very Simple Structure Fit
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
8

Extract 5 factors from the BFI

```
> f5 <- fa(bfi[1:25],5)
fa.diagram(f5,main="Five factors of personality?")
```

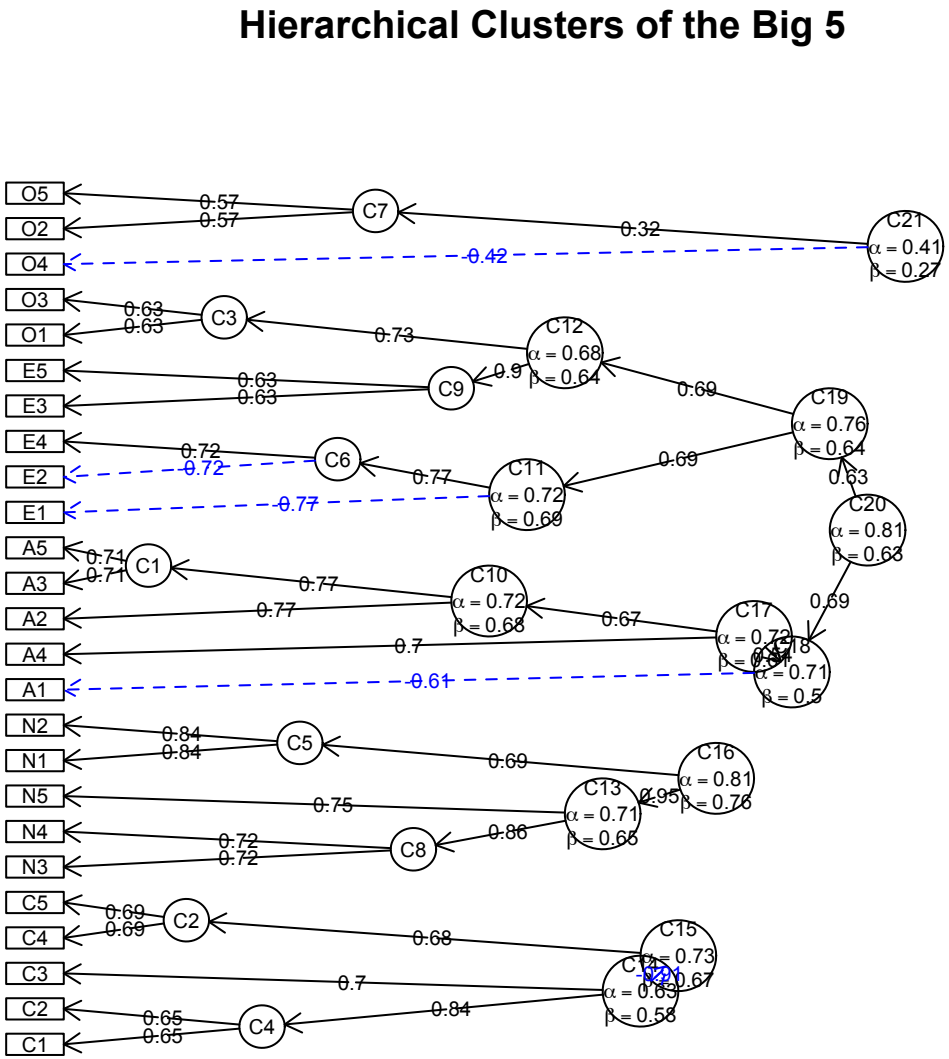


ICLUST of Big 5

```
> iclust(bfi[1:25])
ICLUST (Item Cluster Analysis

Purified Alpha:
 C20  C16  C15  C21
0.80 0.81 0.73 0.61

G6* reliability:
 C20  C16  C15  C21
0.82 0.81 0.72 0.61

Original Beta:
 C20  C16  C15  C21
0.63 0.76 0.67 0.27

Cluster size:
C20 C16 C15 C21
 10   5   5   5

With eigenvalues of:
C20 C16 C15 C21
3.8 3.0 2.6 1.9
```

```
Purified scale intercorrelati
reliabilities on diagonal
correlations corrected for at
     C20   C16   C15   C21
C20  0.80 -0.29 -0.40 -0.33
C16 -0.24  0.81  0.29  0.11
C15 -0.30  0.22  0.73  0.30
C21 -0.23  0.07  0.20  0.61
```

ICLUST as a graphic tree structure



Analyzing from an external file

- Data may reside on a local or a remote computer
- Option A: Using read.clipboard and its alternatives
  - Open the other other file using a text editor or spreadsheet program
  - Select all and copy (to the clipboard)
  - my.data <- read.clipboard() or my.data <- read.clipboard.csv() or read.clipboard.tab()
- Read the information directly
  - find the file and call it something fn <- file.choose()
  - Read in the data my.data <- read.table(fn, header=TRUE)
- Read from an SPSS file using the foreign package
  - library(foreign)
  - find the file and call it something fn <- file.choose()
  - my.data <- read.spss(fn,to.data.frame=TRUE)