Advanced Statistical Procedures

Data = Model + Residual

The basic data frame

DV	IV_1	IV_2	SV_1	SV_2	CV_1	CV_2	…	CV_N
Y_1	X_{11}	X_{21}	X_{31}	X_{41}				
…								
Y_n	X_{1n}	X_{2n}	X_{3n}	X_{4n}				

Types of models

- $Y = bX$ (X is continuous) regression
- $Y = bX$ (X has two levels) t-test
- $Y = bX$ (X has > 2 levels) F-test
- $Y = b_1X_1 + b_2X_2$ (X is continuous) multiple regression
- $Y = b_1X_1 + b_2X_2 + b_3X_1X_2$ (X is continuous) multiple regression with an interaction term
- $Y = b_1X_1 + b_2X_2 + b_3X_1X_2$ (X_i categorical) Analysis of Variance

Analysis of Covariance

- We want to remove the effect of some other predictor
- $Y = b_1X_1 + b_2X_2 + b_3X_1X_2 + Z$
  - (X_i categorical, Z continuous) Analysis of Covariance

The general linear model

- model = lm(y ~ x1 + x2 + x1*x2, data = my.data)
- But the product term is correlated with x1 and x2, and so we need to 0 center (subtract the mean) from each predictor
  - cen.data.df <- data.frame(scale (my.data,scale=FALSE))
  - model <- lm(y ~ x1 + x2 + x1*x2, data = cen.data.df)
  - summary(model)

Analysis of Variance

- If X_i are categorical, we can create a "factor" by
- X1cat <- as.factor(X1)
- X2cat <- as.factor(X2) ...
- and then
- model <- aov(y ~ x1cat + x2cat + x1cat*x2cat, data = my.data)
- summary(model)
- print(model.tables(model, "means"),digits=2)

Multivariate Analysis

- Suppose we have multiple predictors and we want to understand their structure.
- We can find the sum of all the predictors to get a total score, or we can find the sum of some subset of predictors to get total scores on subsets or factors of the data.
  - How many factors are there in the data?

Factor Analysis and Principal Components

- We try to model the correlations with fewer variables
- $R = FF' + U^2$ (Factor Analysis)
- $R = CC'$ (Components analysis)

9 mental tests from Harman-Holzinger

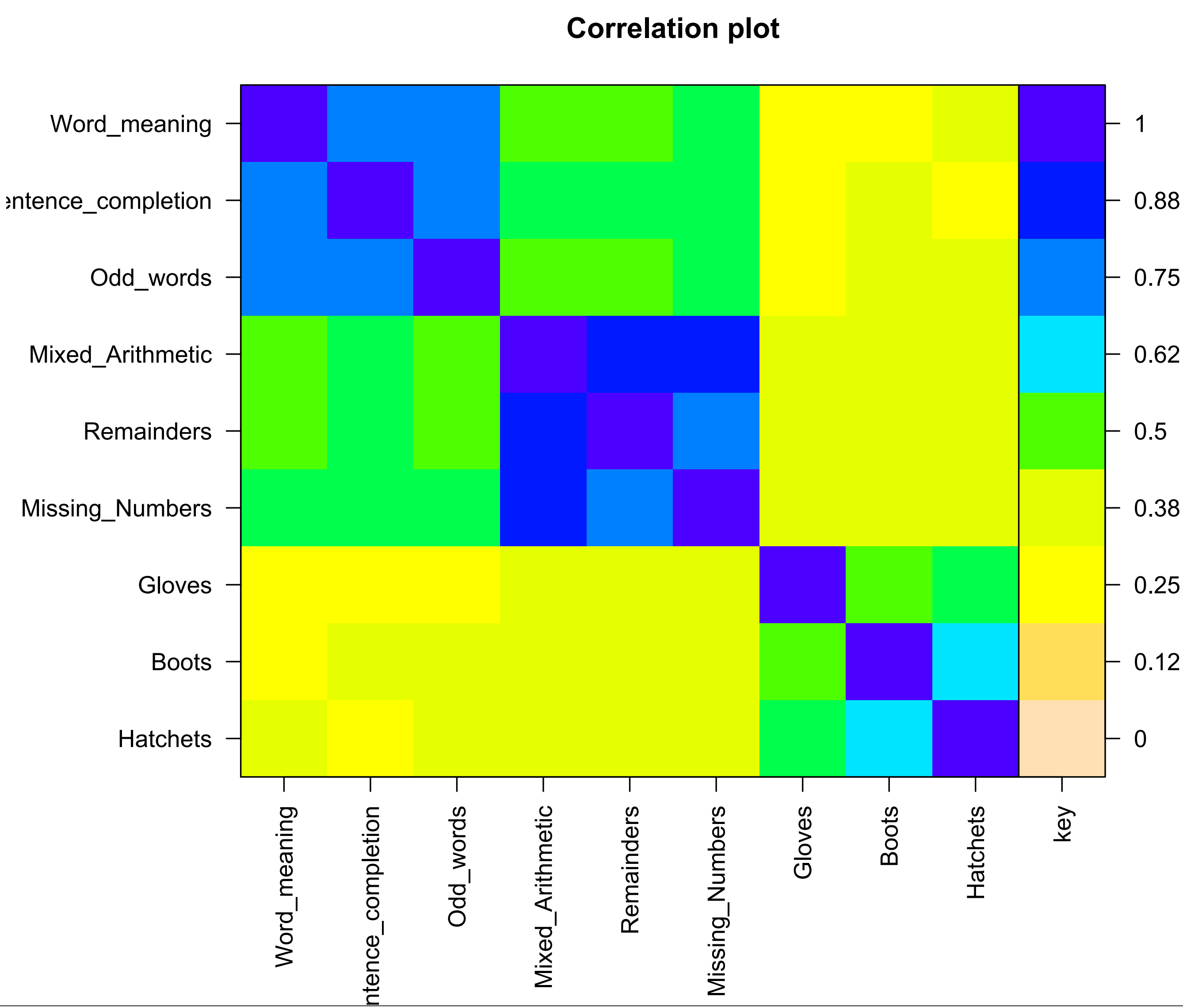
Harman.Holzinger

	V1	V2	V3	V4	V5	V6	V7	V8	V9
Word_meaning	1.00	0.75	0.78	0.44	0.45	0.51	0.21	0.30	0.31
Sentence_completion	0.75	1.00	0.72	0.52	0.53	0.58	0.23	0.32	0.30
Odd_words	0.78	0.72	1.00	0.47	0.48	0.54	0.28	0.37	0.37
Mixed_Arithmetic	0.44	0.52	0.47	1.00	0.82	0.82	0.33	0.33	0.31
Remainders	0.45	0.53	0.48	0.82	1.00	0.74	0.37	0.36	0.36
Missing_Numbers	0.51	0.58	0.54	0.82	0.74	1.00	0.35	0.38	0.38
Gloves	0.21	0.23	0.28	0.33	0.37	0.35	1.00	0.45	0.52
Boots	0.30	0.32	0.37	0.33	0.36	0.38	0.45	1.00	0.67
Hatchets	0.31	0.30	0.37	0.31	0.36	0.38	0.52	0.67	1.00

```
library(psych)
data(Harman,package="psych")
Harman.Holzinger
```

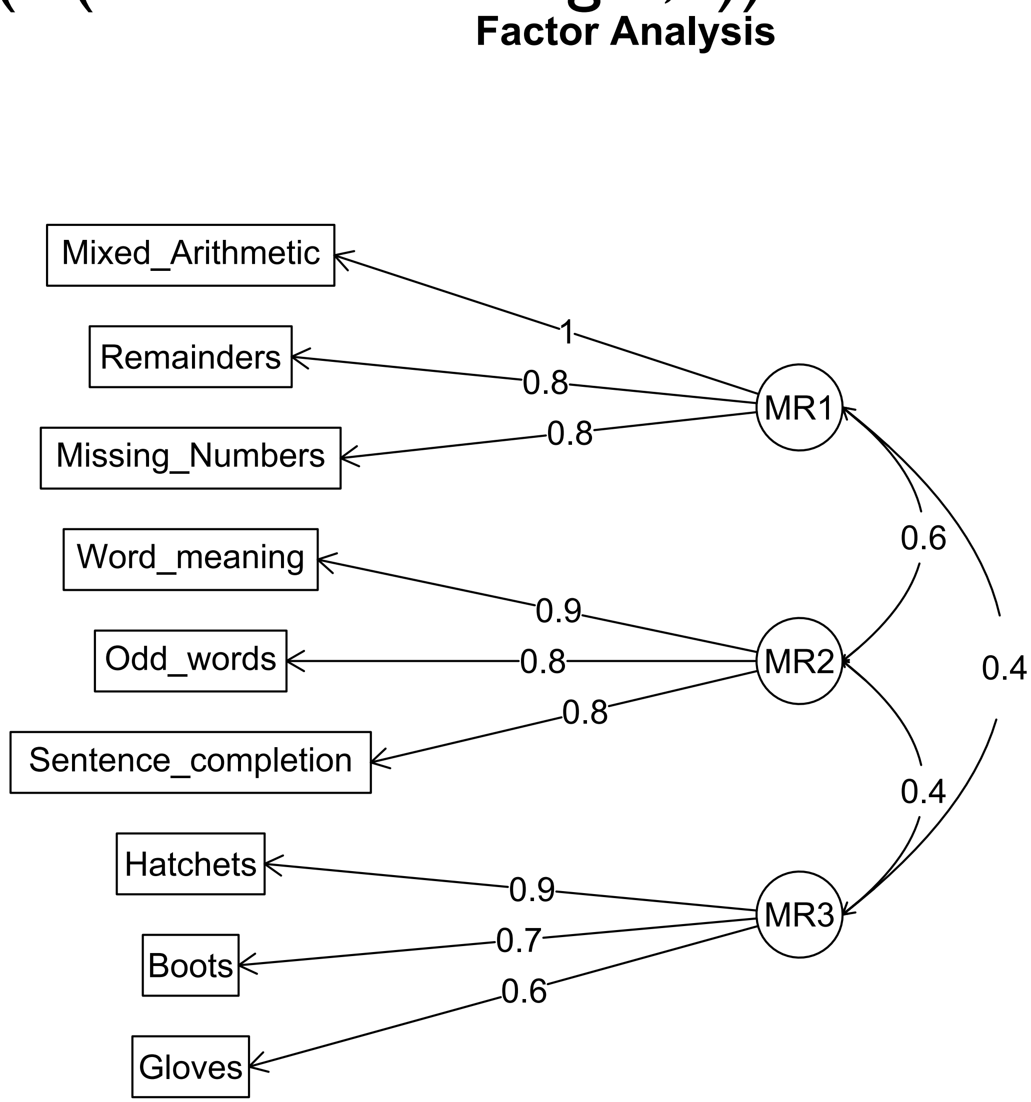
Graphical display

cor.plot(Harman.Holzinger,colors=TRUE)



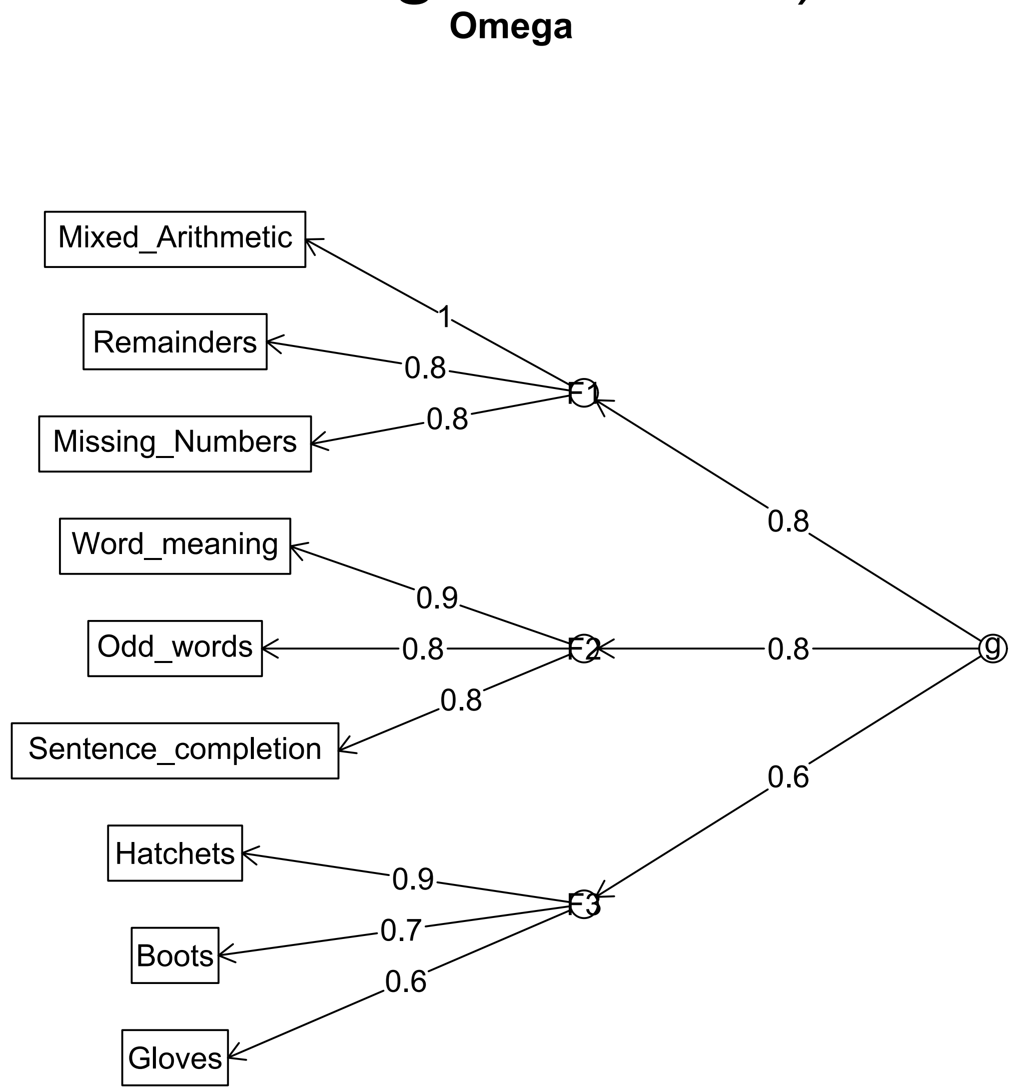
3 correlated factors

fa.diagram(fa(Harman.Holzinger,3))

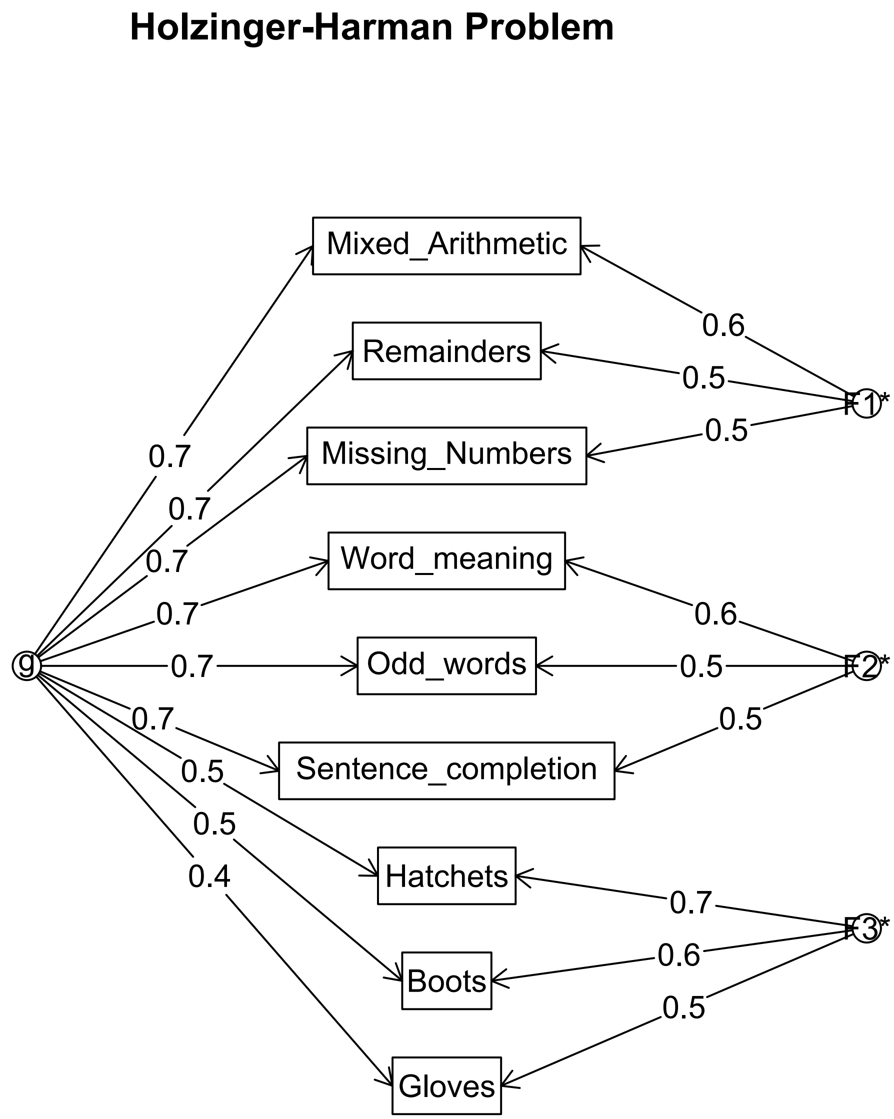


A hierarchical solution

omega(Harman.Holzinger,sl=FALSE)



A general factor solution



```
library(psych)
data(msq)
names(msq)
```

```
 [1] "MSQ_Time"     "active"       "afraid"       "alert"        "alone"        "angry"        "aroused"      "ashamed"      "astonished"   "at-ease"      "at-rest"
[12] "attentive"    "blue"         "bored"        "calm"         "clutched-up"  "confident"    "content"      "delighted"    "depressed"    "determined"   "distressed"
[23] "drowsy"       "dull"         "elated"       "energetic"    "enthusiastic" "excited"      "fearful"      "frustrated"   "full-of-pep"  "gloomy"       "grouchy"
[34] "guilty"       "happy"        "hostile"      "inspired"     "intense"      "interested"   "irritable"    "jittery"      "kindly"       "lively"       "lonely"
[45] "nervous"      "placid"       "pleased"      "proud"        "quiescent"    "quiet"        "relaxed"      "sad"          "satisfied"    "scared"       "scornful"
[56] "serene"       "sleepy"       "sluggish"     "sociable"     "sorry"        "still"        "strong"       "surprised"    "tense"        "tired"        "unhappy"
[67] "upset"        "vigorous"     "wakeful"      "warmhearted"  "wide-awake"   "anxious"      "idle"         "cheerful"     "inactive"     "tranquil"     "EA"
[78] "TA"           "PA"           "NegAff"       "Extraversion" "Neuroticism"  "Lie"          "Sociability"  "Impulsivity"  "MSQ_Round"    "scale"        "ID"
[89] "exper"        "condition"    "TOD"          "TOD24"
```

Form a new data set

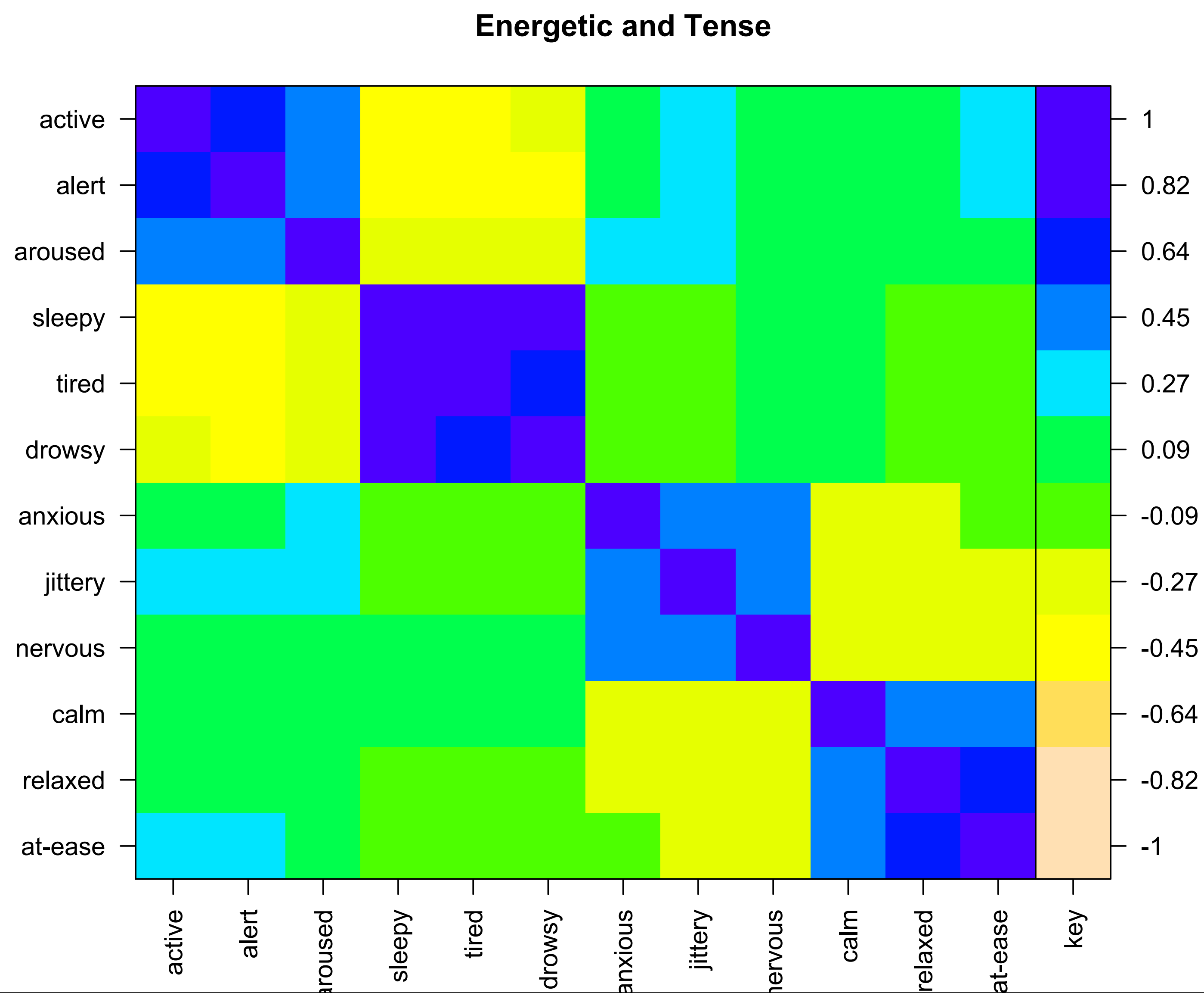
```
>  EA.TA <- msq[c
("active","alert","aroused","sleepy","tired","drowsy","anxious","jittery","nervous","calm","re
laxed","at-ease")]

>  round(cor(EA.TA,use="pairwise"),2)
```

	active	alert	aroused	sleepy	tired	drowsy	anxious	jittery	nervous	calm	relaxed	at-ease
active	1.00	0.62	0.60	-0.40	-0.42	-0.39	0.19	0.27	0.11	0.06	0.16	0.23
alert	0.62	1.00	0.53	-0.53	-0.53	-0.53	0.17	0.23	0.09	0.11	0.18	0.28
aroused	0.60	0.53	1.00	-0.33	-0.35	-0.32	0.22	0.29	0.17	0.01	0.09	0.15
sleepy	-0.40	-0.53	-0.33	1.00	0.81	0.85	-0.04	-0.12	0.02	0.03	-0.07	-0.12
tired	-0.42	-0.53	-0.35	0.81	1.00	0.78	-0.05	-0.12	0.01	0.01	-0.08	-0.14
drowsy	-0.39	-0.53	-0.32	0.85	0.78	1.00	-0.03	-0.11	0.02	0.05	-0.07	-0.10
anxious	0.19	0.17	0.22	-0.04	-0.05	-0.03	1.00	0.45	0.51	-0.25	-0.24	-0.19
jittery	0.27	0.23	0.29	-0.12	-0.12	-0.11	0.45	1.00	0.47	-0.28	-0.30	-0.22
nervous	0.11	0.09	0.17	0.02	0.01	0.02	0.51	0.47	1.00	-0.27	-0.28	-0.27
calm	0.06	0.11	0.01	0.03	0.01	0.05	-0.25	-0.28	-0.27	1.00	0.54	0.58
relaxed	0.16	0.18	0.09	-0.07	-0.08	-0.07	-0.24	-0.30	-0.28	0.54	1.00	0.61
at-ease	0.23	0.28	0.15	-0.12	-0.14	-0.10	-0.19	-0.22	-0.27	0.58	0.61	1.00

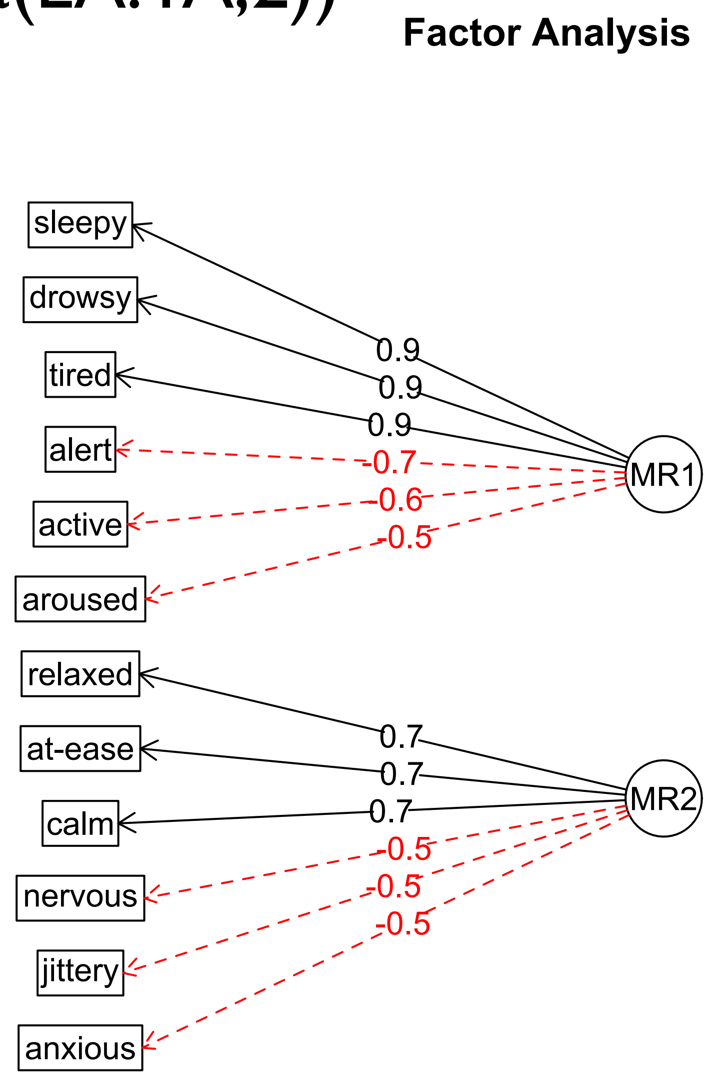
Graphical Displays show structure

```
cor.plot(cor(EA.TA,use="pairwise"),colors=TRUE,zlim=c(-1,1),main="Energetic and Tense")
```



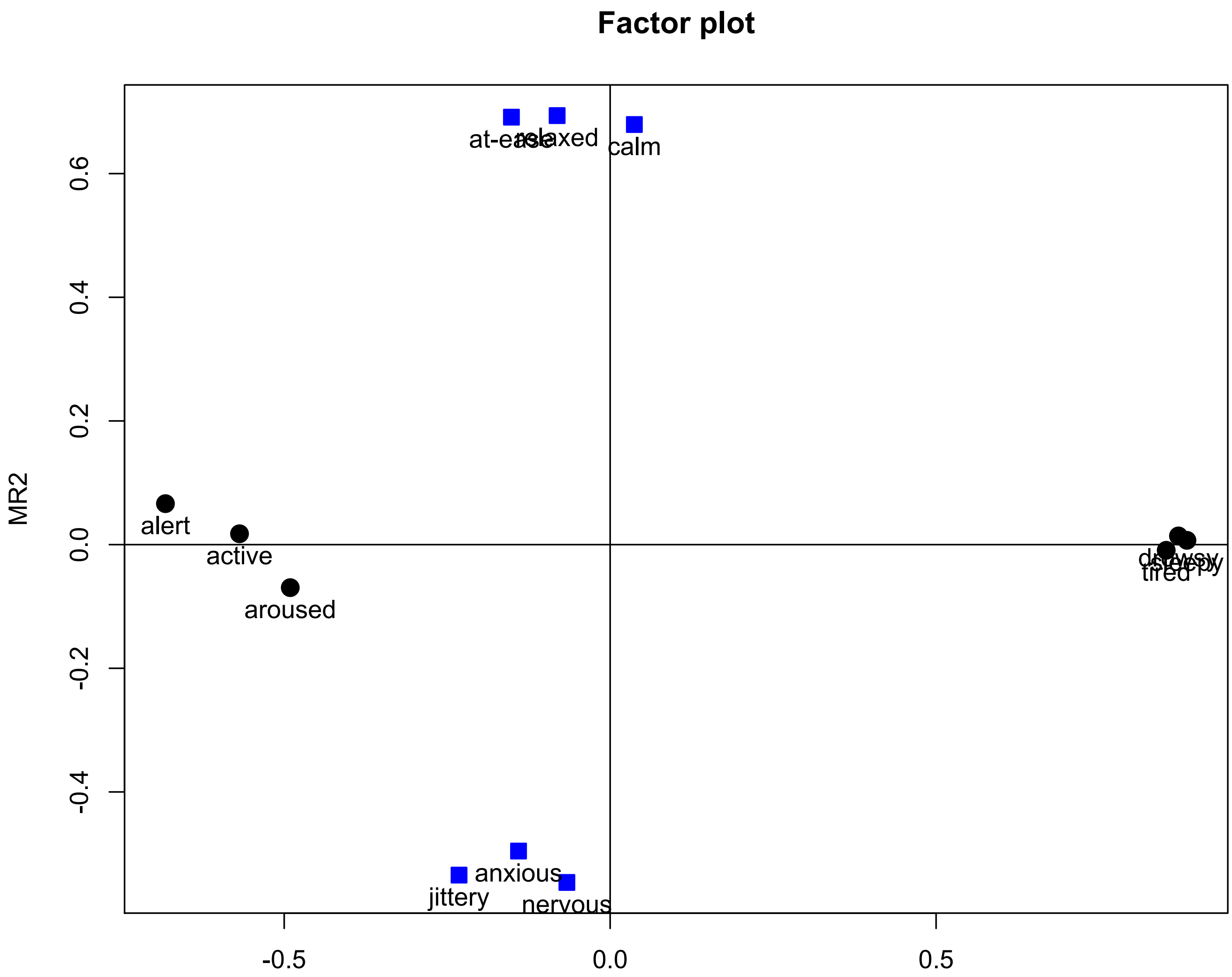
Graphical Display

fa.diagram(fa(EA.TA,2))



Another way to show it

factor.plot(fa(EA.TA,2),labels=colnames(EA.TA))



Score scales by using a keys vector

```
> keys <- make.keys(12,list(EA=c(1:3,-4,-5,-6),TA=c(7:9,-10,-11,-12)))
> keys
      EA TA
 [1,]  1  0
 [2,]  1  0
 [3,]  1  0
 [4,] -1  0
 [5,] -1  0
 [6,] -1  0
 [7,]  0  1
 [8,]  0  1
 [9,]  0  1
[10,]  0 -1
[11,]  0 -1
[12,]  0 -1
> ea.ta.scores <- score.items(keys,EA.TA)
```

Output of scoring program

```
> ea.ta.scores
Call: score.items(keys = keys, items = EA.TA)
(Unstandardized) Alpha:
        EA   TA
alpha 0.87 0.75
Average item correlation:
            EA   TA
average.r 0.54 0.34
 Guttman 6* reliability:
          EA   TA
Lambda.6 0.9 0.77
Scale intercorrelations corrected for
attenuation
 raw correlations below the diagonal, alpha
on the diagonal
 corrected correlations above the diagonal:
      EA    TA
EA  0.87 -0.02
TA -0.02  0.75
>
```

```
            EA    TA
active    0.65 -0.02
alert     0.73 -0.07
aroused   0.56  0.07
sleepy   -0.84  0.02
tired    -0.82  0.03
drowsy   -0.83  0.01
anxious   0.06  0.36
jittery   0.25  0.53
nervous   0.06  0.55
calm      0.01 -0.66
relaxed   0.14 -0.69
at-ease   0.22 -0.69
```

Additional output includes scores