Exploratory Factor Analysis & Visualization

Mengqian Lu

Factor Analysis

- A class of procedures to identify underlying dimensions explaining the correlation structure of variables
- Why FA? – Data Reduction and Data Summarization
- What special about FA? – understand the patterns of relationships among DVs, simultaneously discover how IDVs affect them
- Good for problems with too many correlated variables that need to be reduced to a manageable level
  - E.g. Psychologists use FA to understand the profile of a person by analyzing his/her lifestyle statements.
  - E.g. In market research, FA helps to identify customers' groups.

Exploratory Factor Analysis

- Exploratory Factor Analysis is used to determine the number of latent variables that are needed to explain the correlations among a set of observed variables.
- EFA is to discover the factor structure of a measure and to examine its internal reliability.
- Recommended when researchers have no hypotheses about the nature of the underlying factor structure of their measure.
- Three main decision points of EFA: (1) number of factors; (2) extraction method; (3) rotation method.

Some terms to clarify first

- A latent variable: not directly observable, but affect the response variable (manifest variable)
- Latent variable model: (1) represent the effect of unobservable covariates/factors; (2) account for the unobserved heterogeneity between subjects; (3) account for measurement errors – the latent variables represent the "true" outcomes and the manifest variables represent their "disturbed" versions; (4) summarize different measurements of the same unobservable characteristics (e.g. qualify-of-life)
- Factor analysis models (EFA,CFA) are latent variable models

Latent variable models

- Assume that large number of observed random variables X_1, …, X_p can be explained by a smaller set of unobservable (latent) underlying variables; e.g. PCA, EFA, Canonical Correlation Analysis (CCA), Structural Equation Modeling (SEM) and Independent Component Analysis (ICA)

In Exploratory Factor Analysis:

- Factors: (continuous) latent variables
- (Factor) indicators: observed variables (continuous, censored, binary, ordered/ordinal categorical, counts or mixture of these)

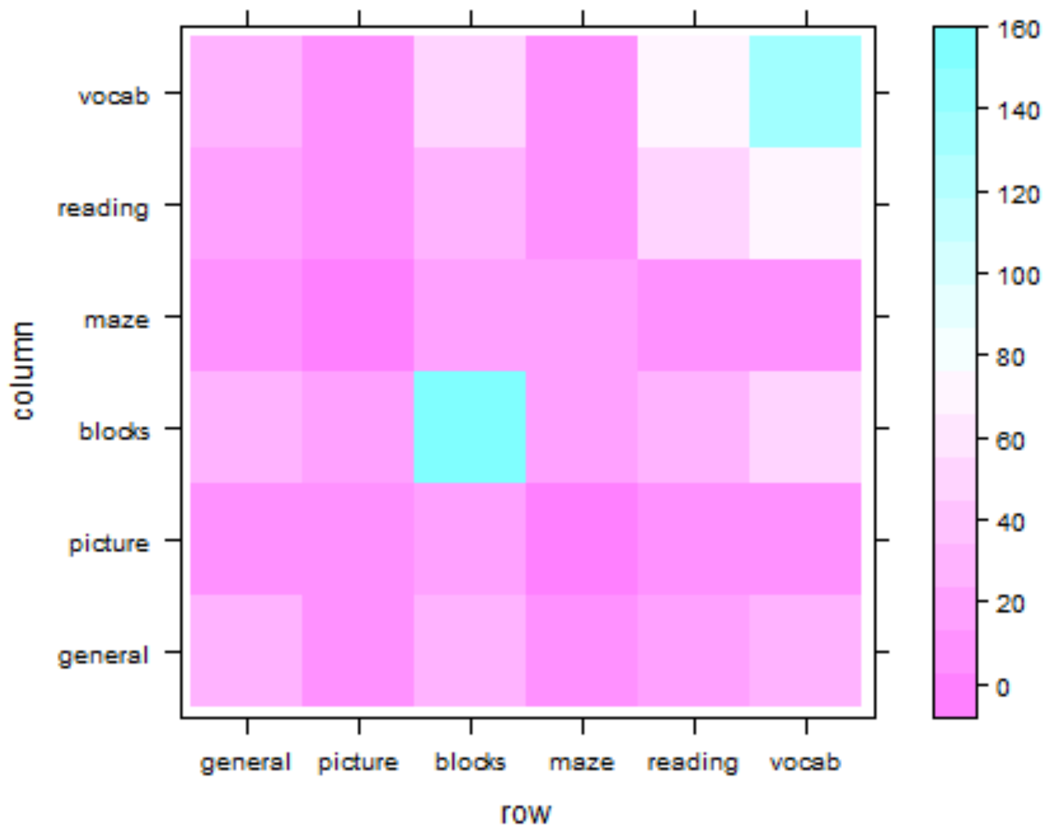
Example: Ability and Intelligence Tests

- Six intelligence tests (general, picture, blocks, maze, reading, vocabulary) were given to 112 individuals.
- The covariance matrix is given:

```
> ability.cov
$cov
        general picture  blocks   maze reading   vocab
general  24.641   5.991  33.520  6.023  20.755  29.701
picture   5.991   6.700  18.137  1.782   4.936   7.204
blocks   33.520  18.137 149.831 19.424  31.430  50.753
maze      6.023   1.782  19.424 12.711   4.757   9.075
reading  20.755   4.936  31.430  4.757  52.604  66.762
vocab    29.701   7.204  50.753  9.075  66.762 135.292

$center
[1] 0 0 0 0 0 0

$n.obs
[1] 112
```

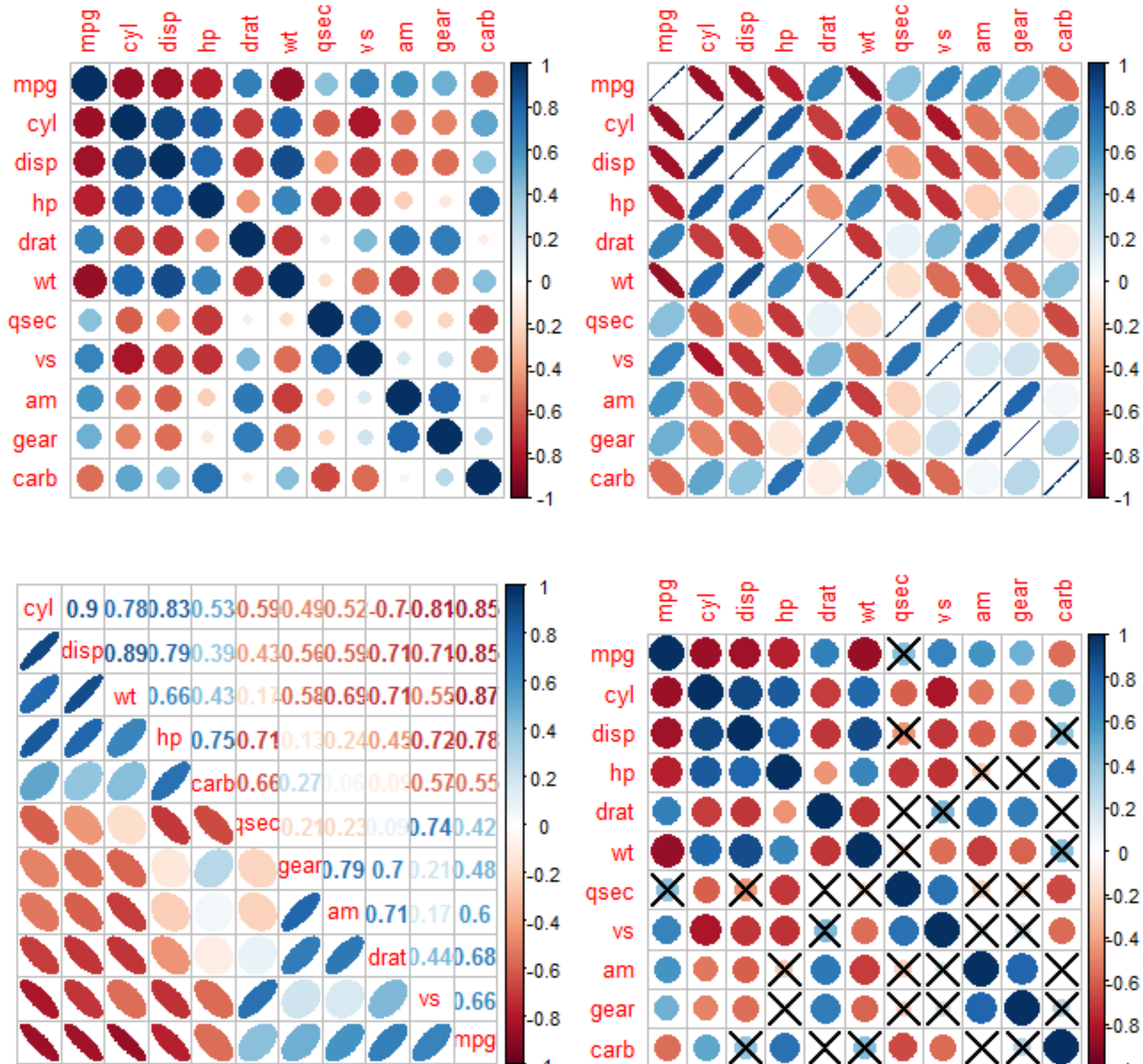
Many ways to visualize cov/cor max

```
library(lattice)

levelplot(ability.cov$cov)
```



Many ways to visualize cov/cor max (cont'd)

```
# Correlation matrix
library(corrplot)
M = cor(mtcars)
par(mfrow = c(2,2))
corrplot(M, method = "circle")
corrplot(M, method = "ellipse")
corrplot.mixed(M,lower="ellipse",upper='number',order="FPC")
# p.mat is the calculated p.value matrix
corrplot(M, p.mat = res1[[1]], sig.level = 0.01)
```



https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html

Example: Ability and Intelligence Tests

- Six intelligence tests (general, picture, blocks, maze, reading, vocabulary) were given to 112 individuals.
- The covariance matrix is given:

```
> ability.cov
$cov
        general picture  blocks   maze reading  vocab
general  24.641   5.991  33.520  6.023  20.755 29.701
picture   5.991   6.700  18.137  1.782   4.936  7.204
blocks   33.520  18.137 149.831 19.424  31.430 50.753
maze      6.023   1.782  19.424 12.711   4.757  9.075
reading  20.755   4.936  31.430  4.757  52.604 66.762
vocab    29.701   7.204  50.753  9.075  66.762 135.292

$center
[1] 0 0 0 0 0 0

$n.obs
[1] 112
```

Question: Can you find one or two "summary" variable(s)/factor(s) describing some general concept of intelligence out of this correlation?

Example: Ability and Intelligence Tests

- One-factor Model for each individual i:

$$x_{1i} = \lambda_1 f_i + u_{1i}$$
$$x_{2i} = \lambda_2 f_i + u_{2i}$$
$$\ldots$$
$$x_{6i} = \lambda_6 f_i + u_{6i}$$

  - f: common factor ('ability')
  - u: random disturbance s.t. each test
  - λ: factor loadings, give weights/importance of f on x_j

- Underlying assumption: $u_1, u_2, \ldots, u_6$ are uncorrelated → x_1, x_2, x_3 are conditionally uncorrelated given f

$$x_{1i} - \lambda_1 f_i = u_{1i}; x_{2i} - \lambda_2 f_i = u_{2i}, \ldots; x_{6i} - \lambda_6 f_i = u_{6i}$$
$$\mathrm{corr}(u_{1i}, u_{2i}) = 0 = \mathrm{corr}(x_{1i} - \lambda_1 f, x_{2i} - \lambda_2 f)$$

Factor Models

- One-factor model for all subjects:

$$\boldsymbol{X}_i = \lambda F_i + \boldsymbol{U}_i, \quad 1 \le i \le n$$
$$X_{ji} = \lambda_j F_i + U_{ji}, \quad 1 \le j \le p, 1 \le i \le n.$$

  where X_{ji} is the j^{th} random variable in a p-dimensional random vector $\boldsymbol{X}_i$, the i^{th} subject.

- F_i is the univariate factor variable defined for each subject i
- $\lambda = (\lambda_1, ..., \lambda_p)^T$ is the factor loadings explain the relation between the factor F and the observed $\boldsymbol{X}$.
- U_{ij} is the error term.

Matrix form: $\mathbf{X}_{pxn} = \boldsymbol{\lambda}_{px1} \mathbf{F}_{1xn} + \mathbf{U}_{pxn}$

Example: Ability and Intelligence Tests

- General model for one individual:

$$x_1 = \mu_1 + \lambda_{11} f_1 + \ldots + \lambda_{1q} f_q + u_1$$

$$\ldots$$

$$x_p = \mu_p + \lambda_{p1} f_p + \ldots + \lambda_{pq} f_q + u_p$$

$$\mathrm{cov}(x_i) = \Sigma = \Lambda\Lambda^T + \Psi$$

To be determined from x:

1. q : no. of common factors
2. Factor loadings Λ
3. Ψ, the cov($\mathbf{u}$)
4. Factor scores f

- Matrix notation for one individual:

$$x = \mu + \Lambda f + u$$

- Matrix notation for *n* individuals:

$$x_i = \mu + \Lambda f_i + u_i \quad (i = 1, \ldots, n)$$

Factor Models (Cont'd)

- Generalized q-factor model:

$$X_{ji} = \mu_j + \sum_{l=1}^{q} \lambda_{jl} F_{li} + U_{ji}, 1 \le j \le p, 1 \le i \le n$$

Or

$$\mathbf{X}_{\mathbf{i}} = \boldsymbol{\mu} + \boldsymbol{\Lambda}_{(\mathbf{p}\times\mathbf{q})} \mathbf{F}_{\mathbf{i}(\mathbf{q}\times\mathbf{1})} + \mathbf{U}_{\mathbf{i}}, 1 \le i \le n$$

where X_{ji} is the j^{th} random variable in a p-dimensional random vector $\boldsymbol{X}_i$, the i^{th} subject.

- $\boldsymbol{\mu} \in \mathbb{R}^p$ constant
- $\boldsymbol{\Lambda}_{(\mathbf{p}\times\mathbf{q})}$ is factor loading matrix, explaining the relation between the factors in $\boldsymbol{F}_i$ and the observed $\boldsymbol{X}_i$
- Errors $\boldsymbol{U}_i$ with $E(\boldsymbol{U}_i) = 0$; $Cov(\boldsymbol{U}_i) = diag(\psi_1, \ldots, \psi_p) = \Psi$
- $Cov(\boldsymbol{F}_i, \boldsymbol{U}_i) = \mathbf{0}$, **or** $Cov(u_j, f_s) = 0$ for all j,s

Assumptions

Convention: factors are scaled, otherwise Λ and μ are not well determined.

- $\mathbf{F}_{\mathbf{i}(\mathbf{q}\times\mathbf{1})}$ is a q-variate random vector with $E(\boldsymbol{F}_i) = 0$; $Cov(\boldsymbol{F}_i) = I_q$

FA: Covariance matrix

$$x = \mu + \Lambda f + u \Leftrightarrow \mathrm{cov}(x) = \Sigma = \Lambda\Lambda^T + \Psi$$

- Factor model is essentially a particular structure imposed on covariance matrix

Still remember the decomposition of covariance matrix in PCA?

$$\boxed{\begin{aligned} Cov(x) &= \frac{1}{n}\sum_i x_i x_i^T = C_x \\ C_x &= U\Lambda U^T \end{aligned}}$$

- Rotations: PCs
- Scores: describe relation between individuals and PCs

FA: $\boxed{\mathrm{cov}(x) = \Sigma = \Lambda\Lambda^T + \Psi}$

FA: Covariance matrix

$$x = \mu + \Lambda f + u \Leftrightarrow \mathrm{cov}(x) = \Sigma = \Lambda\Lambda^T + \Psi$$

- Factor model is essentially a particular structure imposed on covariance matrix
- Instead of spectral decomposition in PCA, we look for a way to split up the variances:

$$\mathrm{var}(x_j) = \sigma_j^2 = \mathrm{var}\left(\sum_{l=1}^{q} \lambda_{jl}^2 F_l\right) + \mathrm{var}(u_j) = \sum_{l=1}^{q} \lambda_{jl}^2 + \psi_j$$

$\sum_{l=1}^{q} \lambda_{jl}^2$	Describe the "Communality", variance due to common factors
ψ_j	Describe the "uniqueness", variance specific to X_j

- "Heywood case" in FA, when $\psi_j \leq 0$

(Fabrigar et al.)

Factor Analysis in R: `factanal()`

data "ability.cov" in R

general	a non-verbal measure of general intelligence using Cattell's culture-fair test
picture	a picture-completion test
blocks	block design
maze	mazes
reading	reading comprehension
vocab	vocabulary

Ref: Bartholomew, D. J. (1987) and Bartholomew, D. J. and Knott, M. (1990)

In R: `factanal()` perform maximum-likelihood factor analysis on a covariance matrix or data matrix

Example: Ability and Intelligence Tests

In R: `factanal()` perform maximum-likelihood factor analysis on a covariance matrix or data matrix

Maximum Likelihood Estimation (MLE):

Assume X_i follows Multivariate Normal Distribution (MVN)

The log-likelihood with parameters μ, Λ, Ψ:

$$ll = \log(L) = -\frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)$$

$$= -\frac{n}{2}\log(|\Lambda\Lambda^T + \Psi|) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)$$

μ is estimated with sample mean, thus choose Λ, Ψ to maximize the log-likelihood

Some notes for reference

- Likelihood function: a product of likelihood functions across all observations

$$L(\mathbf{x}|\mu) = f(X_1|\mu) \times f(X_2|\mu) \times \ldots \times f(X_N|\mu)$$

$$L(\mathbf{x}|\mu) = \prod_{i=1}^{N} f(X_i|\mu) = \left(\frac{1}{2\pi(1)}\right)^{N/2} \exp\left(-\frac{\sum_{i=1}^{N}(X_i-\mu)^2}{2(1)^2}\right)$$

- The value of μ that maximizes f(**X**|μ) is the MLE of population mean, e.g. here μ is the sample mean
- For unknown mean **μ** and variance **Σ**, the likelihood function is:

$$L(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\left(\mathbf{x}_i-\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}\right)/2\right)$$

- More commonly, it is written as

$$L(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}}\frac{1}{|\Sigma|^{n/2}} \exp\left(-\sum_{i=1}^{N}\left(\mathbf{x}_i-\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}\right)/2\right)$$

Number of factors

- MLE approach for estimation provides test:

H_q: q-factor model is sufficient; vs. *H_u: Σ is unconstrained*

- Practical strategy:
  - Start with a small value of q (e.g. q=1) and check hypothesis test; increase q by 1 at a time until some H_q is not rejected

R codes for Intelligence Tests Example

```
ability.FA = factanal(factors = 1, covmat=ability.cov)
update(ability.FA, factors=2)
update(ability.FA, factors=2, rotation="promax")
```

```
> ability.FA
Call:
factanal(factors = 1, covmat = ability.cov)

Uniquenesses:
general picture blocks maze reading vocab
  0.535   0.853  0.748  0.910  0.232  0.280

Loadings:
        Factor1
general 0.682
picture 0.384
blocks  0.502
maze    0.300
reading 0.877
vocab   0.849

               Factor1
SS loadings      2.443
Proportion Var 0.407

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 75.18 on 9 degrees of freedom.
The p-value is 1.46e-12
```

```
> update(ability.FA, factors=2)

Call:
  factanal(factors = 2, covmat = ability.cov)

Uniquenesses:
  general picture  blocks    maze reading   vocab
0.455   0.589   0.218   0.769   0.052   0.334

Loadings:
        Factor1 Factor2
general 0.499   0.543
picture 0.156   0.622
blocks  0.206   0.860
maze    0.109   0.468
reading 0.956   0.182
vocab   0.785   0.225

                 Factor1 Factor2
SS loadings      1.858   1.724
Proportion Var   0.310   0.287
Cumulative Var   0.310   0.597

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 6.11 on 4 degrees of freedom.
The p-value is 0.191
```

Hypothesis can not be rejected; for simplicity, we thus use two factors

```
> update(ability.FA, factors=2)

Call:
  factanal(factors = 2, covmat = ability.cov)

Uniquenesses:
  general picture  blocks    maze reading   vocab
0.455   0.589   0.218   0.769   0.052   0.334

Loadings:
        Factor1 Factor2
general 0.499   0.543
picture 0.156   0.622
blocks  0.206   0.860
maze    0.109   0.468
reading 0.956   0.182
vocab   0.785   0.225

                Factor1 Factor2
SS loadings       1.858   1.724
Proportion Var    0.310   0.287
Cumulative Var    0.310   0.597

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 6.11 on 4 degrees of freedom.
The p-value is 0.191
```

Spatial reasoning

Verbal intelligence

```
> update(ability.FA, factors=2)
```

Default rotation = 'varimax'

```
Call:
  factanal(factors = 2, covmat = ability.cov)

Uniquenesses:
  general picture  blocks    maze reading   vocab
0.455   0.589   0.218   0.769   0.052   0.334

Loadings:
        Factor1 Factor2
general 0.499   0.543
picture 0.156   0.622
blocks  0.206   0.860
maze    0.109   0.468
reading 0.956   0.182
vocab   0.785   0.225

               Factor1 Factor2
SS loadings      1.858   1.724
Proportion Var   0.310   0.287
Cumulative Var   0.310   0.597

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 6.11 on 4 degrees of freedom.
The p-value is 0.191
```

Spatial reasoning

Verbal intelligence

Not clear

Interpretation of factors is generally debatable

```
> update(ability.FA, factors=2, rotation="promax")

Call:
  factanal(factors = 2, covmat = ability.cov, rotation = "promax")

Uniquenesses:
  general picture  blocks   maze    reading  vocab
   0.455    0.589   0.218   0.769   0.052   0.334

Loadings:
        Factor1 Factor2
general  0.364   0.470
picture          0.671
blocks           0.932
maze             0.508
reading  1.023
vocab    0.811

                        actor1 Factor2
SS loadings      1.853   1.807
Proportion Var   0.309   0.301
Cumulative Var   0.309   0.610

Factor Correlations:
          Factor1 Factor2
Factor1   1.000   0.557
Factor2   0.557   1.000

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 6.11 on 4 degrees of freedom.
The p-value is 0.191
```

Spatial reasoning

Verbal intelligence

BETTER? But wait, orthogonality is not ensured, factors are correlated, what accounts for the relationship between the factors?

Scale invariance of factor analysis

- Suppose $y_j = c_j x_j$ or in matrix: $y = Cx$, C is a diagonal matrix, e.g. units conversion, we already have $\boxed{\Sigma_x = \Lambda_x \Lambda_x^T + \Psi_x}$
- Thus, $Cov(y) = C\Sigma_x C^T = C(\Lambda_x \Lambda_x^T + \Psi_x)C^T$ K-factor model holds for x

$$= (C\Lambda_x)(C\Lambda_x)^T + C\Psi_x C^T = \Lambda_y \Lambda_y^T + \Psi_y$$

i.e. loadings and uniquenesses are the same if expressed in new units

- In many applications, the search for loadings Λ and for specific variance Ψ is done by the decomposition of the correlation matrix of X rather than the covariance matrix Σ. This corresponds to a factor analysis of a linear combination of X, i.e., $Y = D^{-1}(X - \mu)$, $D=diag(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{pp})$. As you find the Λ_y and Ψ_y, then we have $\Lambda_x = D\Lambda_y$ and $\Psi_x = D\,\Psi_y D$

Scale invariance of factor analysis (Cont'd)

- Using *cov* or *cor* gives basically the same result
- Common practice: use correlation matrix or scale input data (this is done in factanal())

Rotational invariance of factor analysis

- The factor loadings are not unique!
  - Suppose that $G^T G = I$ (G: orthogonal matrix), transform f to $f^* = G^T f$, Λ to $\Lambda^* = \Lambda G$, for X^*
  - This yields the same model:

$$X^* = \Lambda^* f^* + u = (\Lambda G)(G^T f) + u = \Lambda f + u = X$$

$$\Sigma^* = \Lambda^* \Lambda^{*T} + \Psi = (\Lambda G)(\Lambda G)^T + \Psi = \Lambda\Lambda^T + \Psi = \Sigma$$

  - The rotated model is equivalent for explaining the covariance matrix
  - Usage: use rotation that makes interpretation of loadings easy or look for rotation that make more practice sense
  - Most popular rotation: **Varimax** – each factor has few large and many small loadings

How to find unique solution?

- Options 1: Have $\mathbf{M} = \boldsymbol{\Lambda}\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ be diagonal, with diagonal elements in descending order of magnitude → ordered orthogonal factors with descending contributions (same logic as PCA)
- Potential problems:
  1. Variables may have substantial loadings on 1+ factor
  2. From the 2nd factor, they often turn to be bipolar, i.e., a mixture of positive and negative loadings -> hard to interpret

- Possible solution: rotate the loadings, but there has been debate over this, in my opinion, there is nothing wrong to introduce domain knowledge

How to find unique solution? (Cont'd)

- Options 2: Factor rotation such that (1) each variable is highly loaded on at most one factor; and/or (2) all factor loadings are either large positive or near zero, with as few intermediate values as possible, thus we split original variables into disjoint sets, each set is associated with a single factor.

→ ***Simple Structure*** *[Thurstone, 1931]*

1. *Each row or the factor loading matrix should contain at least one zero.*
2. *Each column of the loading matrix should contain at least k zeros.*
3. *Every pair of columns of the loading matrix should contain several variables whose loadings vanish in one column but not in the other.*
4. *If the number of factors is four or more, every pair of columns should contain a large number of variables with zero loadings in both columns.*
5. *Conversely, for every pair of columns of the loading matrix only a small number of variables should have non-zero loadings in both columns.*

(IAMA by Everitt and Hothorn, page 145)

```
> update(ability.FA, factors=2, rotation="promax")

Call:
  factanal(factors = 2, covmat = ability.cov, rotation = "promax")

Uniquenesses:
  general picture  blocks  maze   reading  vocab
   0.455   0.589   0.218   0.769   0.052   0.334

Loadings:
        Factor1 Factor2
general  0.364   0.470
picture          0.671
blocks           0.932
maze             0.508
reading  1.023
vocab    0.811

                 Factor1 Fact
SS loadings      1.853   1.80
Proportion Var   0.309   0.30
Cumulative Var   0.309   0.61

Factor Correlations:
        Factor1 Factor2
Factor1   1.000   0.557
Factor2   0.557   1.000

Test of the hypothesis that 2
The chi square statistic is 6
The p-value is 0.191
```

Simple Structure

1. Each row or the factor loading matrix should contain at least one zero.

2. Each column of the loading matrix should contain at least k zeros.

3. Every pair of columns of the loading matrix should contain several variables whose loadings vanish in one column but not in the other.

4. If the number of factors is four or more, every pair of columns should contain a large number of variables with zero loadings in both columns.

5. Conversely, for every pair of columns of the loading matrix only a small number of variables should have non-zero loadings in both columns.

Factor rotation

- *varimax*, *promax* and more …
- *Still remember? → we may have to abandon this restriction now*

Convention: factors are scaled, otherwise Λ and μ are not well determined.

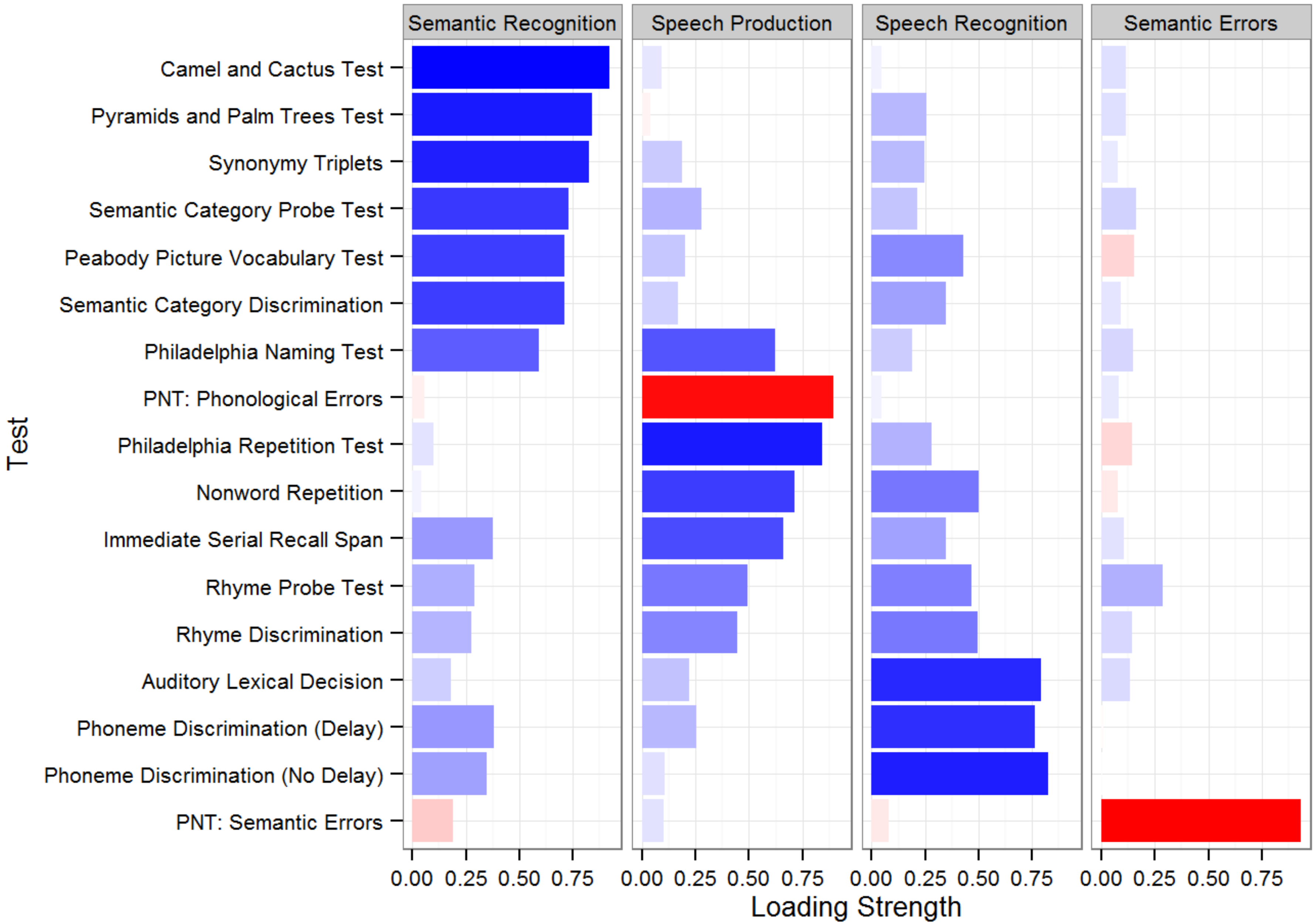
$\boldsymbol{F}_{i(k\times 1)}$ is a k-variate random vector with $E(\boldsymbol{F}_i) = 0$; $Cov(\boldsymbol{F}_i)=\mathbf{I}_k$

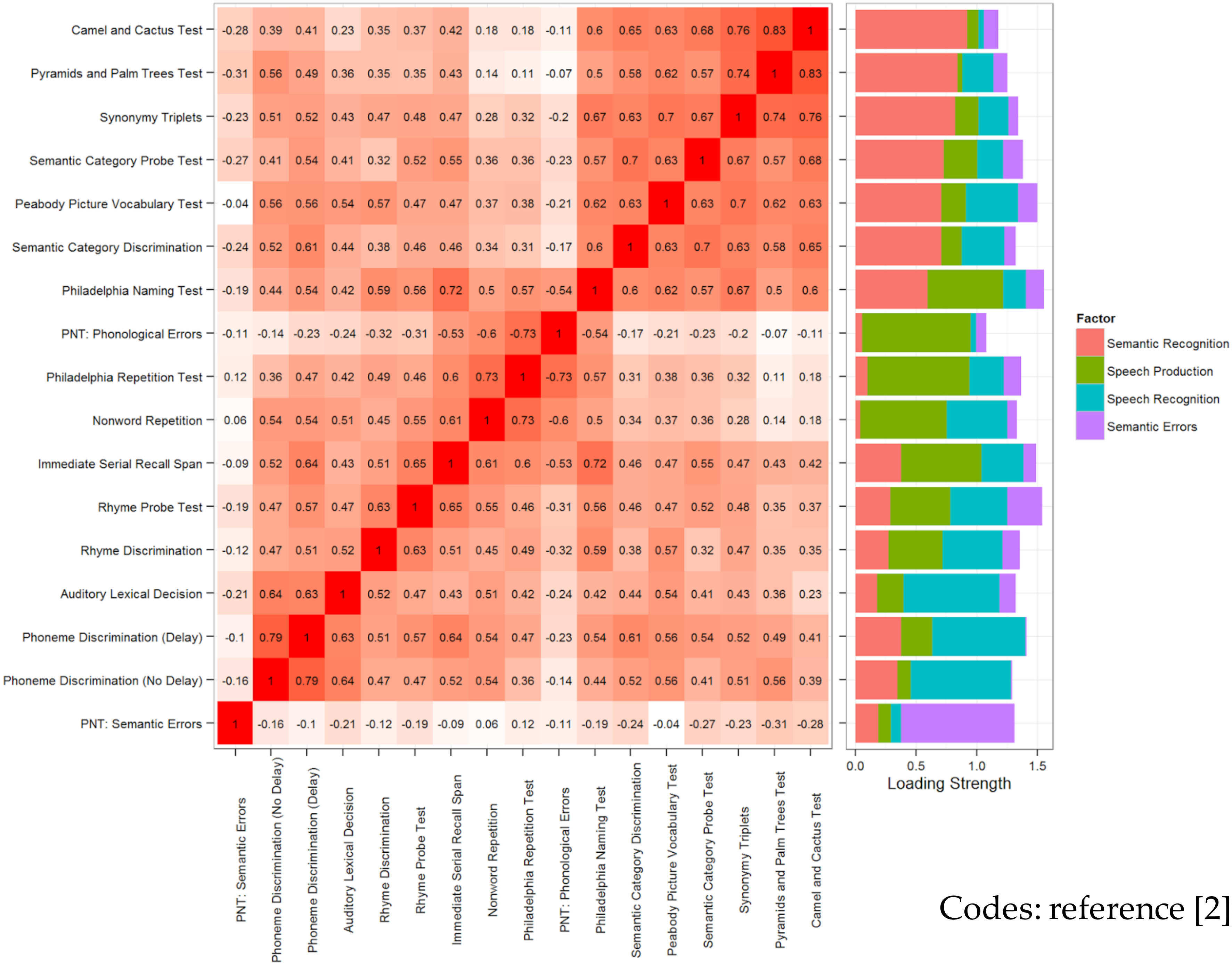
Two main types of rotation:

- Orthogonal rotation: restrict the rotated factors to be uncorrelated
  - Post-multiply the original matrix of loadings by an **orthogonal matrix**
  - After rotation, $Cov(\boldsymbol{F}_i)=\boldsymbol{I}_k$ still holds
- Oblique rotation: allow correlated factors
  - Post-multiply the original matrix of loadings by an **matrix (no longer constrained to be orthogonal)**
  - After rotation, $Cov(\boldsymbol{F}_i)$ has unit elements on its diagonal, but no restrictions on the off-diagonal elements

So which one to use, Orthogonal or Oblique?

- If objective is to "best fit" data, oblique is better; if objective is generalizability, then choose orthogonal to avoid: "what accounts for the relationship between the factors?"
  - An orthogonal rotation is simple, an oblique rotation introduce a matrix of factor correlation to consider.
  - *Varimax* is orthogonal rotation; *promax* is oblique rotation.
  - Factor rotation is still controversial, as it is criticized as – possible allowing one to impose on the data a preconceived structure.


Semantic Recognition
Speech Production
Speech Recognition
Semantic Errors
Camel and Cactus Test
Pyramids and Palm Trees Test
Synonymy Triplets
Semantic Category Probe Test
Peabody Picture Vocabulary Test
Semantic Category Discrimination
Philadelphia Naming Test
PNT: Phonological Errors
Philadelphia Repetition Test
Nonword Repetition
Immediate Serial Recall Span
Rhyme Probe Test
Rhyme Discrimination
Auditory Lexical Decision
Phoneme Discrimination (Delay)
Phoneme Discrimination (No Delay)
PNT: Semantic Errors
Test
0.00 0.25 0.50 0.75
Loading Strength



Codes: reference [2]

Factor scores

- Scores are assumed to be random variables
- Scores are predicted values for each subject (e.g. individuals who took the intelligence tests)
- They represent the original data in a reduced dimension for further analysis or modeling (PCs from PCA also can be used to develop a model for prediction using original data information)
- They are likely to be more reliable than the observed variable values – "noises" eliminated
- The factor score is a "pure" measure of a latent variable, while an observed value may be unclear due to the "noises"

Factor scores – estimation

- Scores are assumed to be random variables, assume normality, the conditional distribution of **f** given x is:

$$N(\Lambda^\top \Sigma^{-1} \mathbf{x}, (\Lambda^\top \Psi^{-1} \Lambda + \mathbf{I})^{-1}).$$

- Two methods to do estimation:
  1. "Bartlett" – Assume *f* as fix (Maximum-likelihood estimate) – Bartlett's weighted least-squares scores, e.g. factanal(~v1+v2+v3+v4+v5+v6, factors = 3, scores = "Bartlett")$scores
  2. "Thompson " – Assume *f* as random (Bayesian estimate), e.g. factanal(~v1+v2+v3+v4+v5+v6, factors = 3, scores = "regression")$scores
  - No big difference in practice

Example

```
> v1 = c(1,1,1,1,1,1,1,1,1,1,3,3,3,3,3,4,5,6)
> v2 = c(1,2,1,1,1,1,2,1,2,1,3,4,3,3,3,4,6,5)
> v3 = c(3,3,3,3,3,1,1,1,1,1,1,1,1,1,1,5,4,6)
> v4 = c(3,3,4,3,3,1,1,2,1,1,1,1,2,1,1,5,6,4)
> v5 = c(1,1,1,1,1,3,3,3,3,3,1,1,1,1,1,6,4,5)
> v6 = c(1,1,1,2,1,3,3,3,4,3,1,1,1,2,1,6,5,4)
> m1 = cbind(v1,v2,v3,v4,v5,v6)
> factanal(m1, factors = 3,
+          scores = "Bartlett")$scores
        Factor1    Factor2    Factor3
 [1,] -0.9039949 -0.9308984  0.9475392
 [2,] -0.8685952 -0.9328721  0.9352330
 [3,] -0.9082818 -0.9320093  0.9616422
 [4,] -1.0021975 -0.2529689  0.8178552
 [5,] -0.9039949 -0.9308984  0.9475392
 [6,] -0.7452711  0.7273960 -0.7884733
 [7,] -0.7098714  0.7254223 -0.8007795
 [8,] -0.7495580  0.7262851 -0.7743704
 [9,] -0.8080740  1.4033517 -0.9304636
[10,] -0.7452711  0.7273960 -0.7884733
[11,]  0.9272282 -0.9307506 -0.8371538
[12,]  0.9626279 -0.9327243 -0.8494600
[13,]  0.9229413 -0.9318615 -0.8230509
[14,]  0.8290256 -0.2528211 -0.9668378
[15,]  0.9272282 -0.9307506 -0.8371538
[16,]  0.4224366  2.0453079  1.2864761
[17,]  1.4713902  1.2947716  0.5451562
[18,]  1.8822320  0.3086244  1.9547752
```

FA vs. PCA

- PCA targets explaining variances, FA targets explaining correlations
- PCA is exploratory and without assumptions;

  FA is based on statistical model with assumptions
- First few PCs will be same regardless of q;

  First few factors of FA depend on q
- FA: Orthogonal rotation of factor loadings are equivalent.

  This does not hold in PCA
- More mathematically:

  PCA: $x=\mu+\Gamma_1 z_1+\Gamma_2 z_2=\mu+\Gamma_1 z_1+e$

  Assume we only keep the PCs in Γ_1

  FA: $x=\mu+\Lambda f+u$

  Cov(u) is diagonal by assumption; Cov(e) is not
- ! Both PCA and FA only useful if input data is correlated !

More references

1. Plot correlation matrix: https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html
2. The two visualization examples can be found at: https://rstudio-pubs-static.s3.amazonaws.com/74109_5a0572586ce34b668a9280fd026602da.html