# Factor Analysis with R

By: Jovita Monteiro, Nikitha Fernandes, Prof. Mohsin Ahmed

MBA students of JKSHIM, India 574110

Email: Jovita.Monteiro.3 @gmail.com, Jackline.Niks @gmail.com

moshahmed@gmail.com

# Agenda

- Introduction to factor Analysis
- Details:
  - FA in R.
  - Scree plots
  - Eigen values
- Conclusion
- Question and Answers
- References

# Introduction: Factor Analysis

- Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors.
- It is used in market research to analyse surveys.

– From http://en.wikipedia.org/wiki/Factor_analysis

# Introduction: SPSS vs R

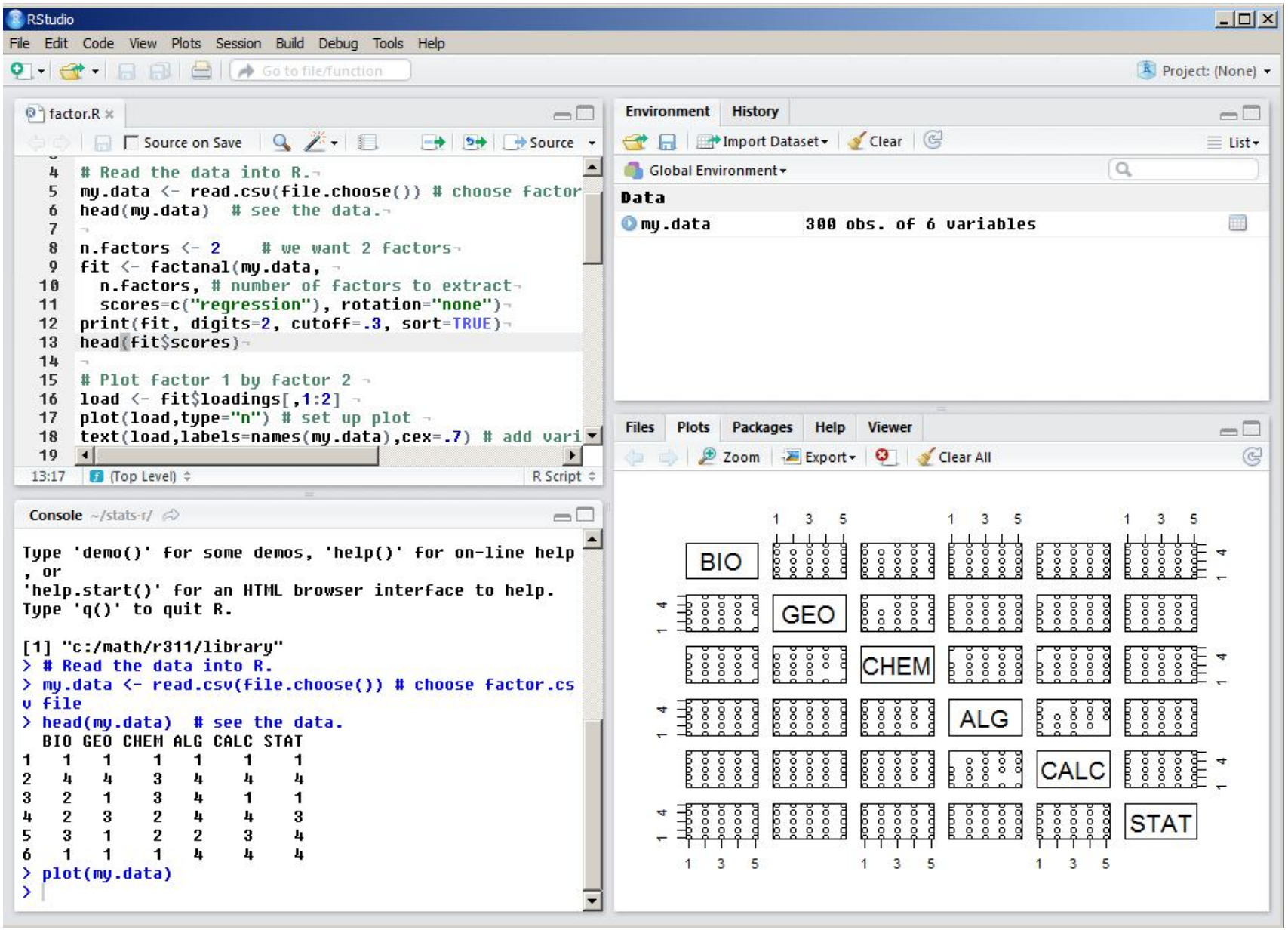
- Factor Analysis (FA) is typically done in SPSS, a statistical software package from IBM.
- In this talk we show how similar analysis can be done in R, a free and sophisticated statistical programming language.
- One must keep in mind, that R allows complete control on its FA methods, while SPSS may hide the details from the user.

# Details: Read the data into R

```
# Read the data into R.
my.data <- read.csv(file.choose()) # choose factor.csv file
head(my.data)  # see the data.
```

```
> # Read the data into R.
> my.data <- read.csv(file.choose()) # choose factor.cs
v file
> head(my.data)  # see the data.
  BIO GEO CHEM ALG CALC STAT
1   1   1    1   1    1    1
2   4   4    3   4    4    4
3   2   1    3   4    1    1
4   2   3    2   4    4    3
5   3   1    2   2    3    4
6   1   1    1   4    4    4
```

# R Studio


RStudio
File Edit Code View Plots Session Build Debug Tools Help
Project: (None)
factor.R
Source on Save
Source
4 # Read the data into R.
5 my.data <- read.csv(file.choose()) # choose factor
6 head(my.data) # see the data.
8 n.factors <- 2 # we want 2 factors
9 fit <- factanal(my.data,
10 n.factors, # number of factors to extract
11 scores=c("regression"), rotation="none")
12 print(fit, digits=2, cutoff=.3, sort=TRUE)
13 head(fit$scores)
15 # Plot factor 1 by factor 2
16 load <- fit$loadings[,1:2]
17 plot(load,type="n") # set up plot
18 text(load,labels=names(my.data),cex=.7) # add vari
13:17 (Top Level) R Script
Environment History
Import Dataset Clear List
Global Environment
Data
my.data 300 obs. of 6 variables
Files Plots Packages Help Viewer
Zoom Export Clear All
Console ~/stats-r/
Type 'demo()' for some demos, 'help()' for on-line help
, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[1] "c:/math/r311/library"
> # Read the data into R.
> my.data <- read.csv(file.choose()) # choose factor.cs
v file
> head(my.data) # see the data.
BIO GEO CHEM ALG CALC STAT
1 1 1 1 1 1 1
2 4 4 3 4 4 4
3 2 1 3 4 1 1
4 2 3 2 4 4 3
5 3 1 2 2 3 4
6 1 1 1 4 4 4
> plot(my.data)
BIO
GEO
CHEM
ALG
CALC
STAT

# Call factanal

```
n.factors <- 2   # we want 2 factors
fit <- factanal(my.data,
      n.factors, # number of factors to extract
      scores=c("regression"), rotation="none")
print(fit, digits=2, cutoff=.3, sort=TRUE)

Call: factanal(x = my.data, factors = n.factors,
   scores = c("regression"), rotation = "none")

Uniquenesses:
 BIO  GEO CHEM  ALG CALC STAT
0.25  0.37  0.25    0.37  0.05 0.71
```

(continued on next slide)

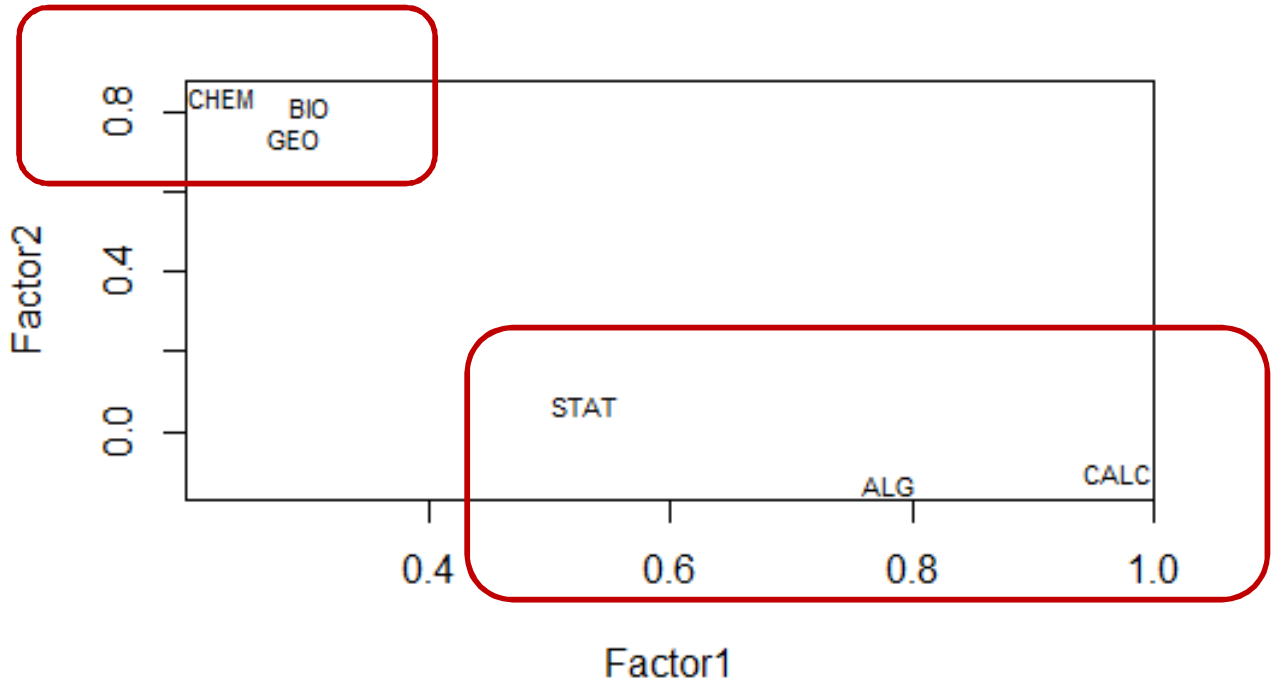
# Loadings of 2 factors

```
Loadings:
     Factor1 Factor2
ALG    0.78
CALC   0.97
STAT   0.53
BIO    0.30    0.81
GEO            0.74
CHEM           0.84
```

```
                Factor1   Factor2
SS loadings        2.06    1.93
Proportion Var     0.34    0.32
Cumulative Var     0.34    0.66
Test of the hypothesis that 2 factors are
  sufficient.
The chi square statistic is 2.94 on 4 degrees of
  freedom.
The p-value is 0.568
```

# Plot the factors

```
# Plot factor 1 by factor 2
load <- fit$loadings[,1:2]
plot(load,type="n") # set up plot
text(load,labels=names(my.data),cex=.7) # add variable names
```



# Factanal varimax

```
# Read the data into R.
text(load,labels=names(my.data),cex=.7) # add variable names
fit <- factanal(my.data,  n.factors,
     rotation="varimax") # 'varimax' is an ortho rotation
load <- fit$loadings[,1:2]
load

Factor1    Factor2
BIO  0.85456336 0.13257053
GEO  0.77932854 0.13455074
CHEM 0.86460737 0.05545425
ALG  0.03133486 0.79070534
CALC 0.09671653 0.97107765
STAT 0.16998499 0.50612151
```

# Scree plot

# packages for scree plot

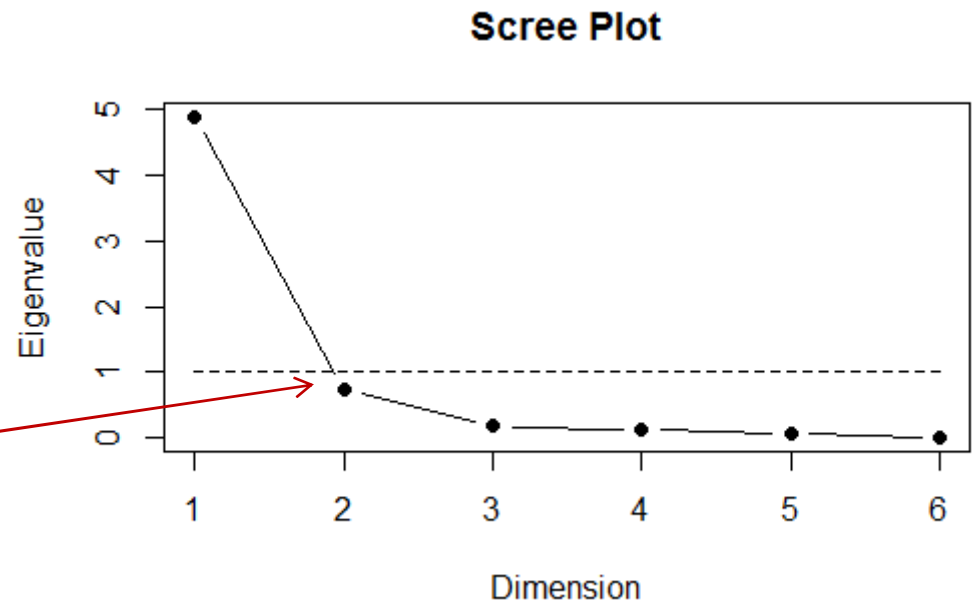
```
# Do this once to download from the web.
install.packages("psych")
install.packages("GPArotation")
install.packages("psy")

# Load the libraries
library(psych)
library(psy)
```

# Scree plot with fa(..)

```
solution <- fa(r = cor(my.data), nfactors = 2,
   rotate = "oblimin", fm = "pa")
plot(solution, labels=names(my.data),cex=.7,
   ylim=c(-.1,1))
scree.plot(fit$correlation)
```

2 factors are needed



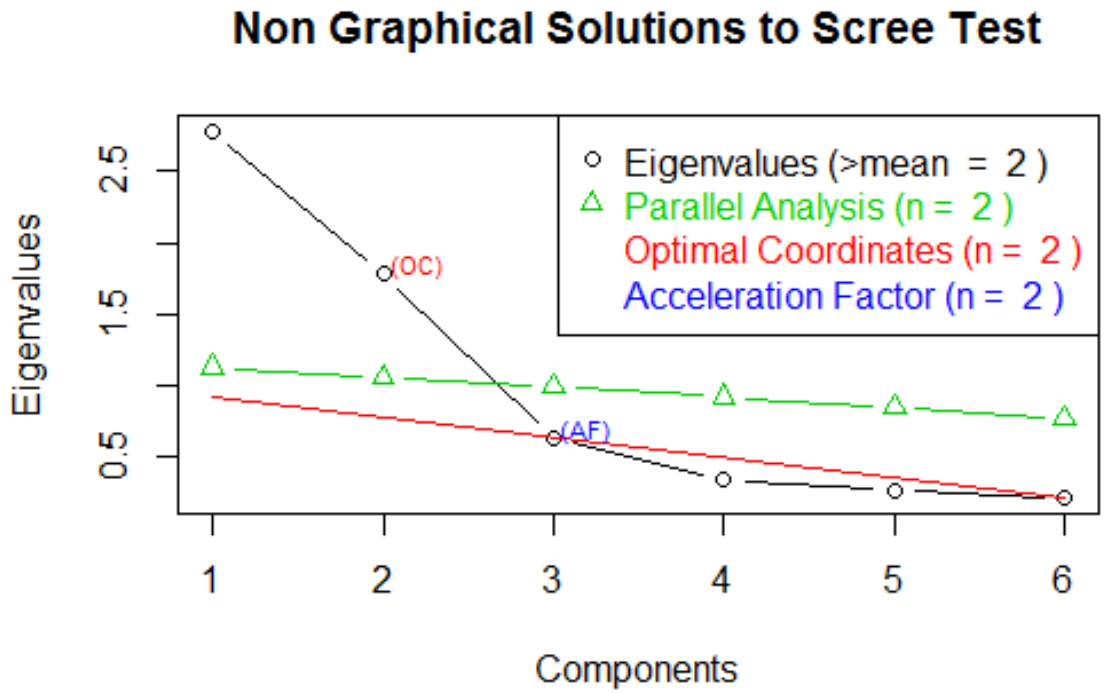
# Eigen values in R

# Get factors from Eigen values

```
# Determine Number of Factors to Extract
# install.packages("nFactors")
library(nFactors)
ev <- eigen(cor(my.data)) # get eigenvalues
```

# Determine factors without plot

```
ap <- parallel( subject=nrow(my.data),
   var=ncol(my.data),rep=100,cent=.05)
nS <- nScree(x=ev$values, aparallel=ap$eigen$qevpea)
plotnScree(nS)
```



# Conclusion

- We have shown how to find number of factors using scree plot and eigen vectors in R.
- We also showed which factors are the principal components (factor1 and factor2).
- We recommend MBA students use R for factor analysis and other statistical analysis, as knowing R is a valuable skill asset in management.

# References


1. http://en.wikipedia.org/wiki/ Factor_analysis
2. PCA http://www.statsoft.com/ Textbook/Principal-Components-Factor-Analysis
3. SPSS factor analysis http://www.ats.ucla.edu/stat/spss/output/factor1.htm
4. http://stats.stackexchange.com/questions/1576/what-are-the-differences-between-factor-analysis-and-principal-component-analysi

# Question and Answers

- Please email us your unresolved questions or suggestions:
  - Jovita.Monteiro.3 @ gmail.com
  - Jackline.Niks @ gmail.com
  - Mosh.Ahmed @ gmail.com