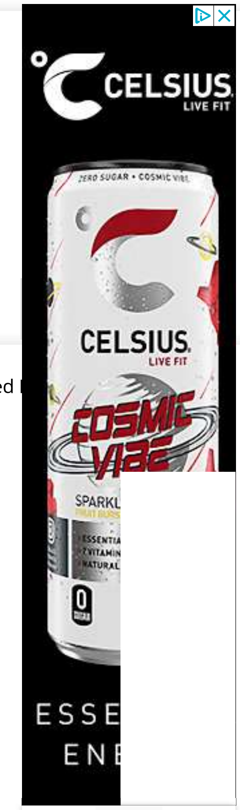
STHDA

Statistical tools for high-throughput data analysis



Search...

Articles - Principal Component Methods in R: Practical Guide

FAMD - Factor Analysis of Mixed Data in R: Essentials

kassambara | *24/09/2017* | *76283* | *Comments (3)* | *Principal Component Methods in R: Practical Guide* | *Multivariate Analysis*

Factor analysis of mixed data (**FAMD**) is a principal component method dedicated to analyze a data set containing both quantitative and qualitative variables (Pagès 2004). It makes it possible to analyze the similarity between individuals by taking into account a mixed types of variables. Additionally, one can explore the association between all variables, both quantitative and qualitative variables.

Roughly speaking, the FAMD algorithm can be seen as a mixed between principal component analysis (PCA) (Chapter @ref(principal-component-analysis)) and multiple correspondence analysis (MCA) (Chapter @ref(multiple-correspondence-analysis)). In other words, it acts as PCA quantitative variables and as MCA for qualitative variables.

Quantitative and qualitative variables are normalized during the analysis in order to balance the influence of each set of variables.

In the current chapter, we demonstrate how to compute and visualize factor analysis of mixed data using *FactoMineR* (for the analysis) and *factoextra* (for data visualization) *R* packages.

Contents:

The Book:

Practical Guide to Principal Component Methods in R

Computation

R packages

Install required packages as follow:

```
install.packages(c("FactoMineR", "factoextra"))
```

Load the packages:

```
library("FactoMineR")
library("factoextra")
```

Data format

We'll use a subset of the `wine` data set available in *FactoMineR* package:

```
library("FactoMineR")
data(wine)
df <- wine[,c(1,2, 16, 22, 29, 28, 30,31)]
head(df[, 1:7], 4)
```

```
##           Label Soil Plante Acidity Harmony Intensity Overall.quality
## 2EL      Saumur Env1   2.00    2.11    3.14      2.86            3.39
## 1CHA     Saumur Env1   2.00    2.11    2.96      2.89            3.21
## 1FON Bourgueuil Env1   1.75    2.18    3.14      3.07            3.54
## 1VAU     Chinon Env2   2.30    3.18    2.04      2.46            2.46
```

To see the structure of the data, type this:

```
str(df)
```

The data contains 21 rows (wines, `individuals`) and 8 columns (`variables`):

- The first two columns are factors (`categorical variables`): `label` (Saumur, Bourgueil or Chinon) and `soil` (Reference, Env1, Env2 or Env4).
- The remaining columns are numeric (`continuous variables`).

✔ The goal of this study is to analyze the characteristics of the wines.

R code

The function *FAMD*() [*FactoMiner* package] can be used to compute FAMD. A simplified format is :

- `ncp`: the number of dimensions kept in the results (by default 5)
- `sup.var`: a vector indicating the indexes of the supplementary variables.
- `ind.sup`: a vector indicating the indexes of the supplementary individuals.
- `graph` : a logical value. If TRUE a graph is displayed.

To compute FAMD, type this:

```
library(FactoMineR)
res.famd <- FAMD(df, graph = FALSE)
```

The output of the *FAMD*() function is a list including :

```
print(res.famd)
```

```
## *The results are available in the following objects:
##
##   name          description
## 1 "$eig"        "eigenvalues and inertia"
## 2 "$var"        "Results for the variables"
## 3 "$ind"        "results for the individuals"
## 4 "$quali.var"  "Results for the qualitative variables"
## 5 "$quanti.var" "Results for the quantitative variables"
```

Visualization and interpretation

We'll use the following *factoextra* functions:

- `get_eigenvalue(res.famd)`: Extract the eigenvalues/variances retained by each dimension (axis).
- `fviz_eig(res.famd)`: Visualize the eigenvalues/variances.
- `get_famd_ind(res.famd)`: Extract the results for individuals.
- `get_famd_var(res.famd)`: Extract the results for quantitative and qualitative variables.
- `fviz_famd_ind(res.famd)`, `fviz_famd_var(res.famd)`: Visualize the results for individuals and variables, respectively.

In the next sections, we'll illustrate each of these functions.

To help in the interpretation of FAMD, we highly recommend to read the interpretation of principal component analysis (Chapter (**???**)(principal-component-analysis)) and multiple correspondence analysis (Chapter (**???**)(multiple-correspondence-analysis)). Many of the graphs presented here have been already described in our previous chapters.

Eigenvalues / Variances

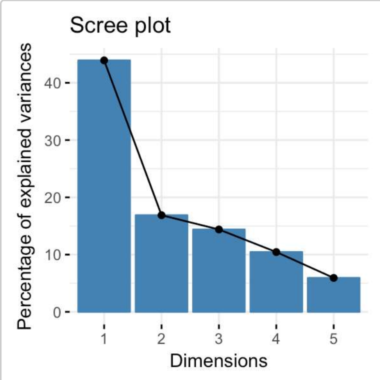
The proportion of variances retained by the different dimensions (axes) can be extracted using the function `get_eigenvalue()` [*factoextra* package] as follow:

```
library("factoextra")
eig.val <- get_eigenvalue(res.famd)
head(eig.val)
```

```
##       eigenvalue variance.percent cumulative.variance.percent
## Dim.1      4.832            43.92                        43.9
## Dim.2      1.857            16.88                        60.8
## Dim.3      1.582            14.39                        75.2
## Dim.4      1.149            10.45                        85.6
## Dim.5      0.652             5.93                        91.6
```

The function `fviz_eig()` or `fviz_screeplot()` [*factoextra* package] can be used to draw the scree plot (the percentages of inertia explained by each FAMD dimensions):

```
fviz_screeplot(res.famd)
```



Graph of variables

All variables

The function get_mfa_var() [in *factoextra*] is used to extract the results for variables. By default, this function returns a list containing the coordinates, the cos2 and the contribution of all variables:

```
var <- get_famd_var(res.famd)
var
```

```
## FAMD results for variables
##  ===================================================
##    Name       Description
## 1 "$coord"   "Coordinates"
## 2 "$cos2"    "Cos2, quality of representation"
## 3 "$contrib" "Contributions"
```

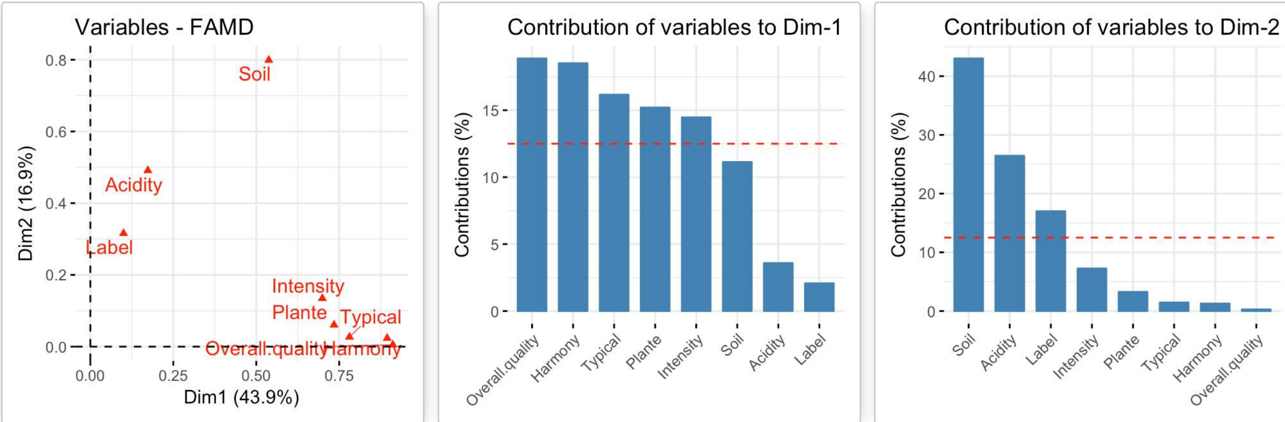
The different components can be accessed as follow:

```
# Coordinates of variables
head(var$coord)
# Cos2: quality of representation on the factore map
head(var$cos2)
# Contributions to the  dimensions
head(var$contrib)
```

The following figure shows the correlation between variables - both quantitative and qualitative variables - and the principal dimensions, as well as, the contribution of variables to the dimensions 1 and 2. The following functions [in the *factoextra* package] are used:

- fviz_famd_var() to plot both quantitative and qualitative variables
- fviz_contrib() to visualize the contribution of variables to the principal dimensions

```
# Plot of variables
fviz_famd_var(res.famd, repel = TRUE)
# Contribution to the first dimension
fviz_contrib(res.famd, "var", axes = 1)
# Contribution to the second dimension
fviz_contrib(res.famd, "var", axes = 2)
```



The red dashed line on the graph above indicates the expected average value, If the contributions were uniform. Read more in chapter (Chapter @ref(principal-component-analysis)).

From the plots above, it can be seen that:

- variables that contribute the most to the first dimension are: Overall.quality and Harmony.
- variables that contribute the most to the second dimension are: Soil and Acidity.

Quantitative variables

To extract the results for quantitative variables, type this:

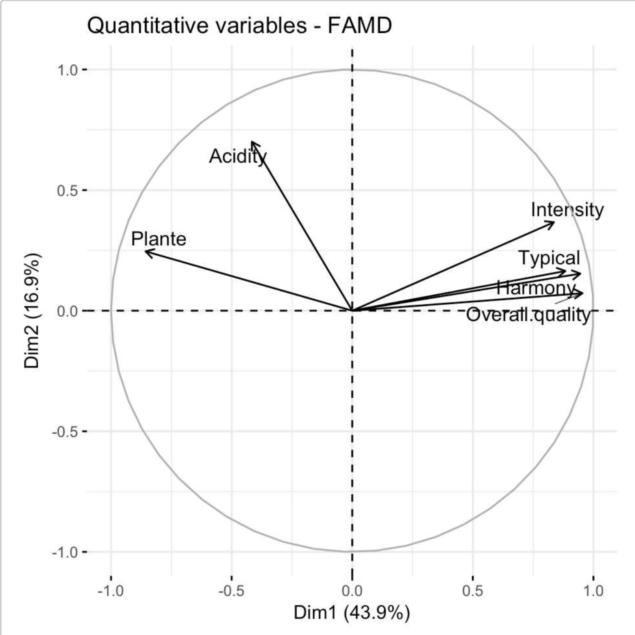
```
quanti.var <- get_famd_var(res.famd, "quanti.var")
quanti.var
```

```
## FAMD results for quantitative variables
##  ===================================================
##   Name       Description
## 1 "$coord"   "Coordinates"
## 2 "$cos2"    "Cos2, quality of representation"
## 3 "$contrib" "Contributions"
```

In this section, we'll describe how to visualize quantitative variables. Additionally, we'll show how to highlight variables according to either i) their quality of representation on the factor map or ii) their contributions to the dimensions.

The R code below plots quantitative variables. We use repel = TRUE, to avoid text overlapping.

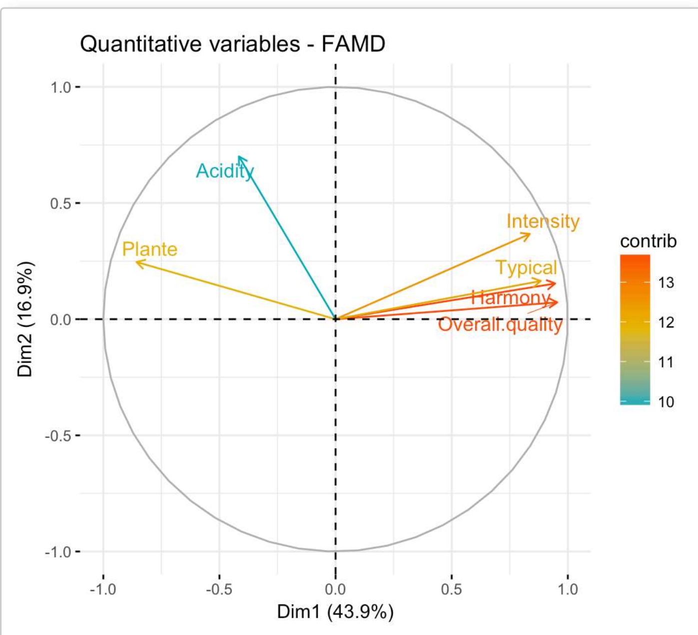
```
fviz_famd_var(res.famd, "quanti.var", repel = TRUE,
             col.var = "black")
```



Briefly, the graph of variables (correlation circle) shows the relationship between variables, the quality of the representation of variables, as well as, the correlation between variables and the dimensions. Read more at PCA (Chapter @ref(principal-component-analysis)), MCA (Chapter @ref(multiple-correspondence-analysis)) and MFA (Chapter @ref(multiple-factor-analysis)).

The most contributing quantitative variables can be highlighted on the scatter plot using the argument `col.var = "contrib"`. This produces a gradient colors, which can be customized using the argument `gradient.cols`.

```
fviz_famd_var(res.famd, "quanti.var", col.var = "contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE)
```



Similarly, you can highlight quantitative variables using their cos2 values representing the quality of representation on the factor map. If a variable is well represented by two dimensions, the sum of the cos2 is closed to one. For some of the items, more than 2 dimensions might be required to perfectly represent the data.

```
# Color by cos2 values: quality on the factor map
fviz_famd_var(res.famd, "quanti.var", col.var = "cos2",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE)
```

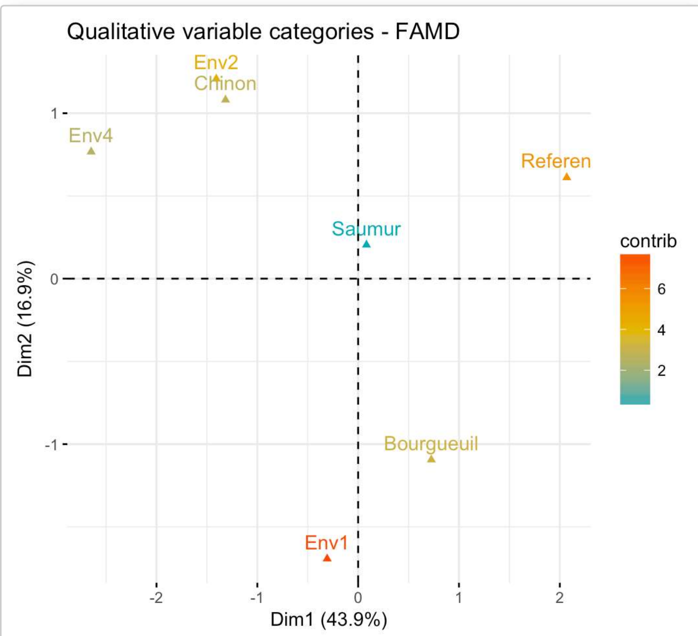
Like quantitative variables, the results for qualitative variables can be extracted as follow:

```
quali.var <- get_famd_var(res.famd, "quali.var")
quali.var
```

```
## FAMD results for qualitative variable categories
## ===================================================
##   Name       Description
## 1 "$coord"   "Coordinates"
## 2 "$cos2"    "Cos2, quality of representation"
## 3 "$contrib" "Contributions"
```

To visualize qualitative variables, type this:

```
fviz_famd_var(res.famd, "quali.var", col.var = "contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07")
             )
```



The plot above shows the categories of the categorical variables.

Graph of individuals

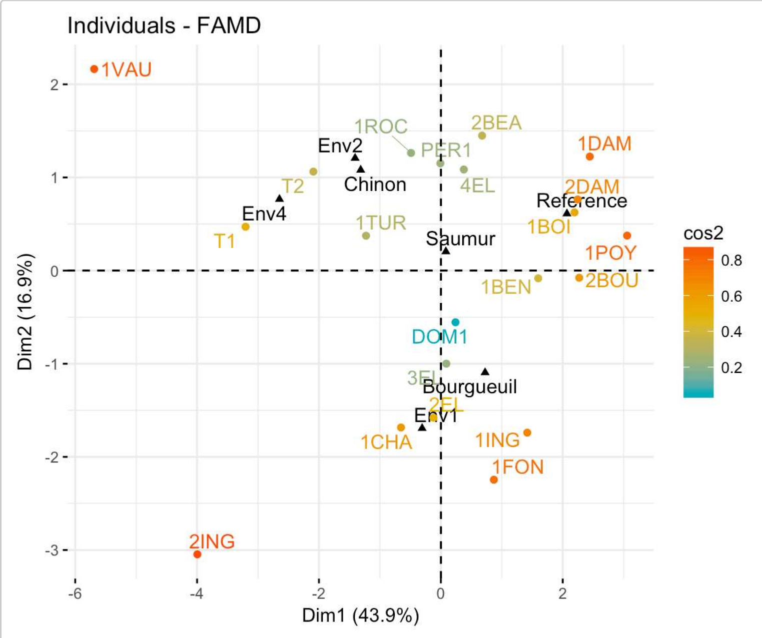
To get the results for individuals, type this:

```
ind <- get_famd_ind(res.famd)
ind
```

```
## FAMD results for individuals
## ===================================================
##   Name       Description
## 1 "$coord"   "Coordinates"
## 2 "$cos2"    "Cos2, quality of representation"
## 3 "$contrib" "Contributions"
```

To plot individuals, use the function `fviz_mfa_ind()` [in *factoextra*]. By default, individuals are colored in blue. However, like variables, it's also possible to color individuals by their cos2 and contribution values:

```
fviz_famd_ind(res.famd, col.ind = "cos2",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
```

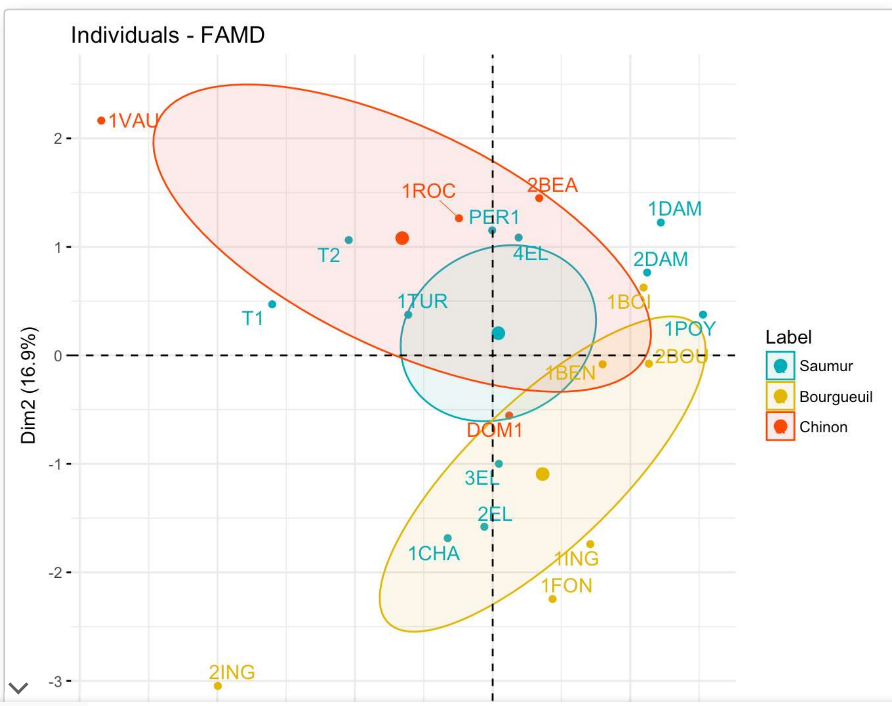


In the plot above, the qualitative variable categories are shown in black. Env1, Env2, Env3 are the categories of the soil. Saumur, Bourgueuil and Chinon are the categories of the wine Label. If you don't want to show them on the plot, use the argument `invisible = "quali.var"`.

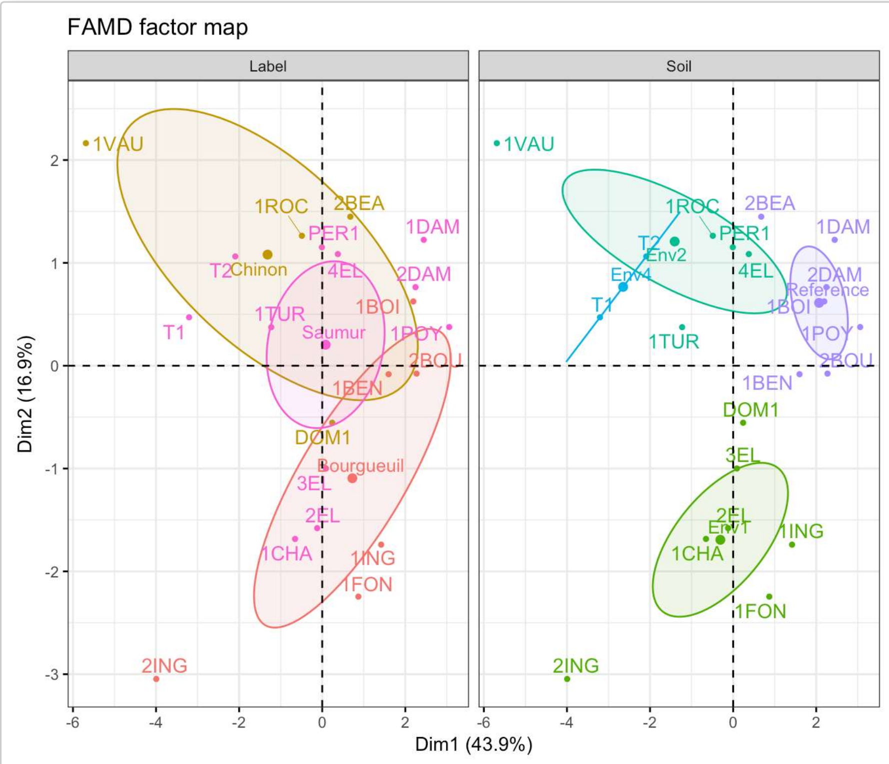
Individuals with similar profiles are close to each other on the factor map. For the interpretation, read more at Chapter @ref(multiple-correspondence-analysis) (MCA) and Chapter @ref(multiple-factor-analysis) (MFA).

Note that, it's possible to color the individuals using any of the qualitative variables in the initial data table. To do this, the argument `habillage` is used in the `fviz_famd_ind()` function. For example, if you want to color the wines according to the supplementary qualitative variable "Label", type this:

```
fviz_mfa_ind(res.famd,
             habillage = "Label", # color by groups
             palette = c("#00AFBB", "#E7B800", "#FC4E07"),
             addEllipses = TRUE, ellipse.type = "confidence",
             repel = TRUE # Avoid text overlapping
             )
```



```
fviz_ellipses(res.famd, c("Label", "Soil"), repel = TRUE)
```



Alternatively, you can specify categorical variable indices:

```
fviz_ellipses(res.famd, 1:2, geom = "point")
```

Summary

The factor analysis of mixed data (FAMD) makes it possible to analyze a data set, in which individuals are described by both qualitative and quantitative variables. In this article, we described how to perform and interpret FAMD using FactoMineR and factoextra R packages.

Further reading

Factor Analysis of Mixed Data Using FactoMineR (video course). https://goo.gl/64gY3R

References

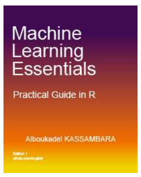
Pagès, J. 2004. "Analyse Factorielle de Donnees Mixtes." *Revue Statistique Appliquee* 4: 93–111.

★★★★★ 1 Note

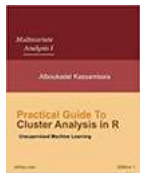
Enjoyed this article? Give us 5 stars (just above this text block)! Reader needs to be STHDA member for voting. I'd be very grateful if you'd help it spread by emailing it to a friend, or sharing it on Twitter, Facebook or Linked In.

Show me some love with the like buttons below... Thank you and please don't forget to share and comment below!!

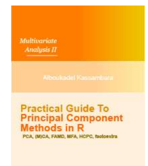
Recommended for You!



Machine Learning Essentials: Practical Guide in R

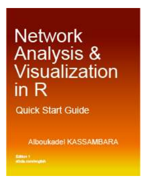


Practical Guide to Cluster Analysis in R



Practical Guide to Principal Component Methods in R

R Graphics Essentials for Great Data Visualization



Network Analysis and Visualization in R

More books on R and data science

Recommended for you

This section contains best data science and self-development resources to help you on your path.

Coursera - Online Courses and Specialization

Data science

- Course: Machine Learning: Master the Fundamentals by Standford
- Specialization: Data Science by Johns Hopkins University
- Specialization: Python for Everybody by University of Michigan
- Courses: Build Skills for a Top Job in any Industry by Coursera
- Specialization: Master Machine Learning Fundamentals by University of Washington
- Specialization: Statistics with R by Duke University
- Specialization: Software Development in R by Johns Hopkins University
- Specialization: Genomic Data Science by Johns Hopkins University

Popular Courses Launched in 2020

- Google IT Automation with Python by Google
- AI for Medicine by deeplearning.ai
- Epidemiology in Public Health Practice by Johns Hopkins University
- AWS Fundamentals by Amazon Web Services

Trending Courses

- The Science of Well-Being by Yale University
- Google IT Support Professional by Google

- Psychological First Aid by Johns Hopkins University
- Graphic Design by Cal Arts

Books - Data Science

Our Books

- Practical Guide to Cluster Analysis in R by A. Kassambara (Datanovia)
- Practical Guide To Principal Component Methods in R by A. Kassambara (Datanovia)
- Machine Learning Essentials: Practical Guide in R by A. Kassambara (Datanovia)
- R Graphics Essentials for Great Data Visualization by A. Kassambara (Datanovia)
- GGPlot2 Essentials for Great Data Visualization in R by A. Kassambara (Datanovia)
- Network Analysis and Visualization in R by A. Kassambara (Datanovia)
- Practical Statistics in R for Comparing Groups: Numerical Variables by A. Kassambara (Datanovia)
- Inter-Rater Reliability Essentials: Practical Guide in R by A. Kassambara (Datanovia)

Others

- R for Data Science: Import, Tidy, Transform, Visualize, and Model Data by Hadley Wickham & Garrett Grolemund
- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems by Aurelien Géron
- Practical Statistics for Data Scientists: 50 Essential Concepts by Peter Bruce & Andrew Bruce
- Hands-On Programming with R: Write Your Own Functions And Simulations by Garrett Grolemund & Hadley Wickham
- An Introduction to Statistical Learning: with Applications in R by Gareth James et al.
- Deep Learning with R by François Chollet & J.J. Allaire
- Deep Learning with Python by François Chollet

! You are not authorized to post a comment

Galpaccru 09/22/2021 at 17h35
Member

Good evening, there is a Problem / BUG with this. I went to the repository to pull a issue too.

The problem comes when you have multiple factors that shares level names.
I have made a link on reddit to see if i can find help : https://www.reddit.com/r/rstats/comments/pt9dyo/problems_with_factominerfactoextra_famd_when/

Here its a reproducible example

Quotation :

```
arm = c("long","short","long","long")
leg = c("short","short","short", "long")
value = c(1,2,3,4)
ej = data.frame(arm,as.factor(leg),as.factor(value))
res.famd_ej <- FAMD(ej, graph= FALSE)
fviz_famd_var(res.famd_ej,"quanti.var",repel=TRUE,
col.var="cos2", gradient.cols = c("blue","yellow","red"))
```

I hope i can find some help soon 😁, cause this its an important and interesting feature to add to factoextra.

And to show that the same code with the dataset wine works -cause they dont share levelnames-

Quotation :

```
res.famd_wine <- FAMD(wines, graph= FALSE)
fviz_famd_var(res.famd_wine,"quanti.var",repel=TRUE,
col.var="cos2", gradient.cols = c("blue","yellow","red"))
```

#920

Riad 02/01/2020 at 15h18
Member

Thank you sir for the great explanation. It did really benefited me and my colleagues.

How can I reconstruct transformed individuals? I have seen in the documents of FactoMineR package, reconst() function can be used to recover the data from PCA, MFA or CA objects but not FAMD. Also, it accepts classes not coordinates.

Many thanks,
Riad

#845

Visitor 09/20/2018 at 09h35
Visitor

Hi sir,

I want to ask question
I dont really understand what is ncp used for? I know it represent the number of dimension, but what dimension? It is similar like loading in Factor analysis?

FAMD (base, **ncp = 5,** sup.var = NULL, ind.sup = NULL, graph = TRUE)

#610

Sign in

Login

Login

Password

Password

Auto connect

Sign in

Register

Forgotten password

Welcome!

Want to Learn More on R Programming and Data Science?
Follow us by Email

Subscribe
by FeedBurner

Click to see our collection of resources to help you on your path...

Course & Specialization

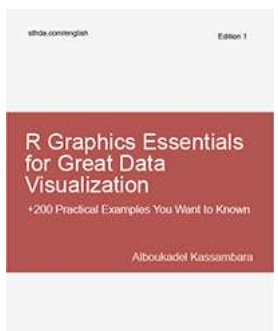
Recommended for You (on Coursera):

- Course: Machine Learning: Master the Fundamentals
- Specialization: Data Science
- Specialization: Python for Everybody
- Course: Build Skills for a Top Job in any Industry
- Specialization: Master Machine Learning Fundamentals
- Specialization: Statistics with R
- Specialization: Software Development in R
- Specialization: Genomic Data Science

See More Resources

- factoextra
- survminer
- ggpubr
- ggcorrplot
- fastqcr

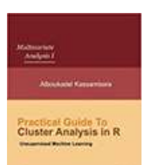
Our Books



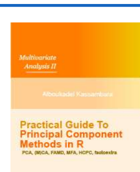
R Graphics Essentials for Great Data Visualization: 200 Practical Examples You Want to Know for Data Science

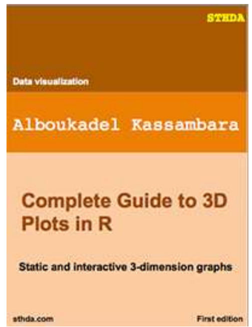
★ NEW!!

3D Plots in R



Practical Guide to Cluster Analysis in R





Datanovia: Online Data Science Courses

R-Bloggers

Newsletter Email