

Welcome!

Want to Learn More on R Programming and Data Science?

Follow us by Email

Subscribe

by FeedBurner

on Social Networks



[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](#)



Articles - Principal Component Methods in R: Practical Guide

CA - Correspondence Analysis in R: Essentials

[kassambara](#) | 24/09/2017 | 56984 | [Comments \(6\)](#) | [Principal Component Methods in R: Practical Guide](#) | [Multivariate Analysis, Categorical data analysis](#)

Correspondence analysis (CA) is an extension of principal component analysis (Chapter [@ref\(principal-component-analysis\)](#)) suited to explore relationships among qualitative variables (or categorical data). Like principal component analysis, it provides a solution for summarizing and visualizing data set in two-dimension plots.

Here, we describe the simple correspondence analysis, which is used to analyze frequencies formed by two categorical data, a data table known as *contingency table*. It provides factor scores (coordinates) for both row and column points of contingency table. These coordinates are used to visualize graphically the association between row and column elements in the contingency table.

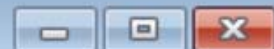
When analyzing a two-way contingency table, a typical question is whether certain row elements are associated with some elements of column elements. Correspondence analysis is a geometric approach for visualizing the rows and columns of a two-way contingency table as points in a low-dimensional space, such that the positions of the row and column points are consistent with their associations in the table. The aim is to have a global view of the data that is useful for interpretation.

In the current chapter, we'll show how to compute and interpret correspondence analysis using two R packages: i) *FactoMineR* for the analysis and ii) *factoextra* for data visualization. Additionally, we'll show how to reveal the most important variables that explain the variations in a data set. We continue by explaining how to apply correspondence analysis using supplementary rows and columns. This is important, if you want to make predictions with CA. The last sections of this guide describe also how to filter CA result in order to keep only the most contributing variables. Finally, we'll see how to deal with outliers.

Contents:

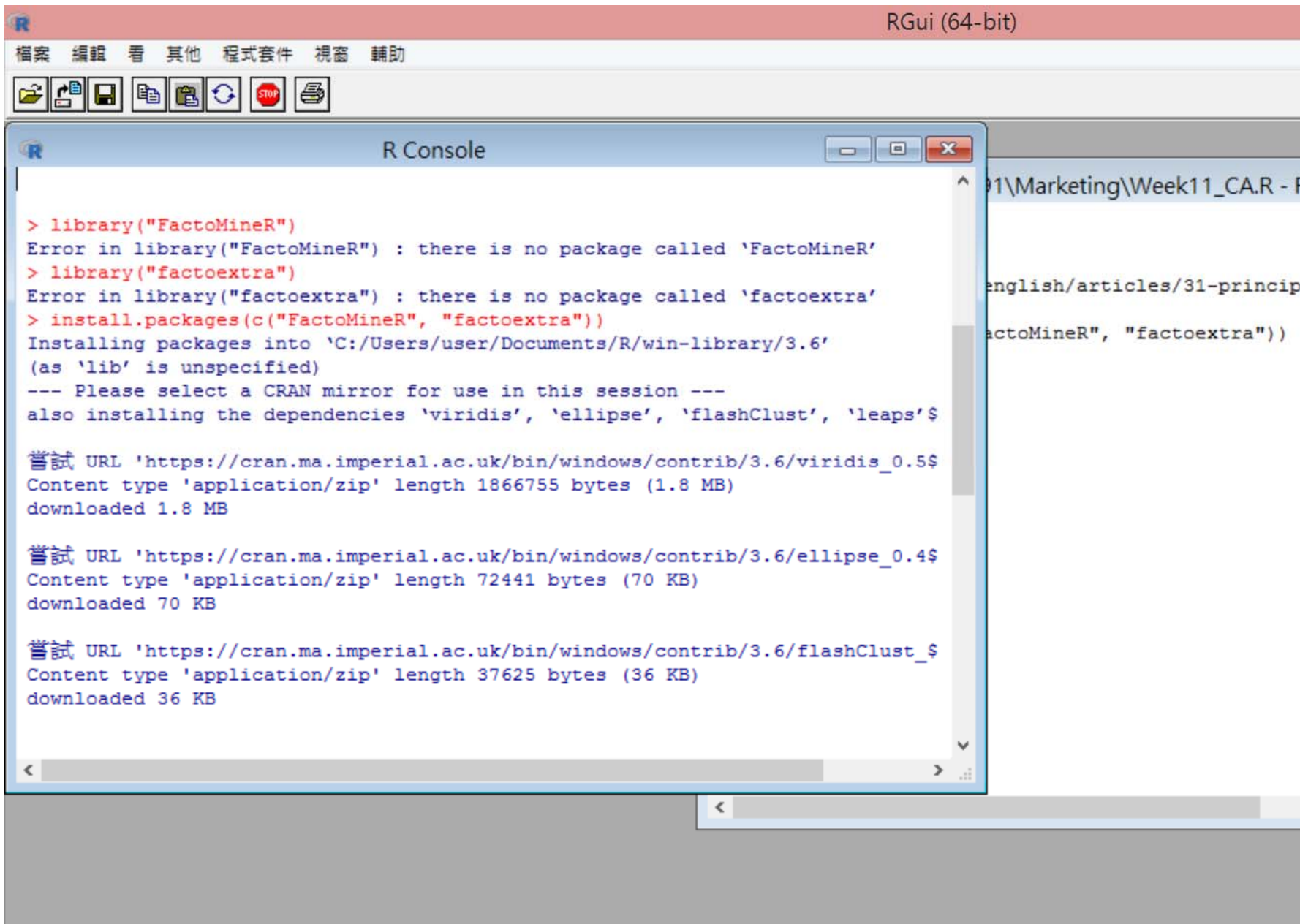


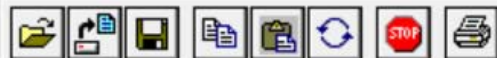
D:\SCU\1091\Marketing\Week11_CA.R - R 編輯器



```
# http://www.sthda.com/english/articles/31-principal-component-methods-in-r-pract  
  
# install.packages(c("FactoMineR", "factoextra"))  
library("FactoMineR")  
library("factoextra")  
  
|
```







R Console

```
Content type 'application/zip' length 3880446 bytes (3.7 MB)
downloaded 3.7 MB
```

```
嘗試 URL 'https://cran.ma.imperial.ac.uk/bin/windows/contrib/3.6/FactoMineR_$
Content type 'application/zip' length 3759342 bytes (3.6 MB)
downloaded 3.6 MB
```

```
嘗試 URL 'https://cran.ma.imperial.ac.uk/bin/windows/contrib/3.6/factoextra_$
Content type 'application/zip' length 418195 bytes (408 KB)
downloaded 408 KB
```

```
package 'viridis' successfully unpacked and MD5 sums checked
package 'ellipse' successfully unpacked and MD5 sums checked
package 'flashClust' successfully unpacked and MD5 sums checked
package 'leaps' successfully unpacked and MD5 sums checked
package 'scatterplot3d' successfully unpacked and MD5 sums checked
package 'dendextend' successfully unpacked and MD5 sums checked
package 'FactoMineR' successfully unpacked and MD5 sums checked
package 'factoextra' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
      C:\Users\user\AppData\Local\Temp\Rtmp6T1AdM\downloaded_packages
```

```
> |
```





```
R Console
> data(housetasks)
> str(housetasks)
'data.frame': 13 obs. of 4 variables:
 $ Wife      : int  156 124 77 82 53 32 33 12 10 13 ...
 $ Alternating: int  14 20 11 36 11 24 23 46 51 13 ...
 $ Husband   : int   2 5 7 15 1 4 9 23 75 21 ...
 $ Jointly   : int   4 4 13 7 57 53 55 15 3 66 ...
> housetasks
      Wife Alternating Husband Jointly
Laundry    156          14         2         4
Main_meal  124          20         5         4
Dinner      77          11         7        13
Breakfast   82          36        15         7
Tidying     53          11         1        57
Dishes      32          24         4        53
Shopping    33          23         9        55
Official    12          46        23        15
Driving     10          51        75         3
Finances    13          13        21        66
Insurance    8           1        53        77
Repairs     0           3       160         2
Holidays    0           1         6       153
> |
```

01\Marketing
english/arti
factoMineR",



```
R Console
Driving      10      51      75      3
Finances    13      13      21      66
Insurance     8       1      53      77
Repairs      0       3     160      2
Holidays    0       1       6     153

> library("gplots")
Error in library("gplots") : there is no package called 'gplots'
> # 1. convert the data as a table
> dt <- as.table(as.matrix(housetasks))
> # 2. Graph
> balloonplot(t(dt), main="housetasks", xlab="", ylab="",
+             label = FALSE, show.margins = FALSE)
Error in balloonplot(t(dt), main = "housetasks", xlab = "", ylab = "", :
 沒有這個函數 "balloonplot"
> library("gplots")
Error in library("gplots") : there is no package called 'gplots'
> install.packages("gplots")
錯誤: 未預期的 ')' in "install.packages()"
> install.packages("gplots")
Installing package into 'C:/Users/user/Documents/R/win-library/3.6'
(as 'lib' is unspecified)
also installing the dependencies 'gtools', 'caTools'
```

housetasks

| | Wife | Alternating | Husband | Jointly |
|-----------|------|-------------|---------|---------|
| Laundry | ● | • | • | • |
| Main_meal | ● | • | • | • |
| Dinner | ● | • | • | • |
| Breakfast | ● | • | • | • |
| Tidying | ● | • | • | ● |
| Dishes | ● | • | • | ● |
| Shopping | ● | • | • | ● |
| Official | • | ● | • | • |
| Driving | • | ● | ● | • |
| Finances | • | • | • | ● |
| Insurance | • | • | ● | ● |
| Repairs | | • | ● | • |
| Holidays | | • | • | ● |

CA - Biplot

