

CA with FactoMineR and factoextra

Basics

CA (Correspondence Analysis) is a descriptive technique, that explains dependency between two categorical variable. It is usually applied to contingency tables — decomposes the chi-squared statistic into orthogonal factors, but may be applied to any table with nonnegative values.

It is conceptually similar to principal component analysis. The preprocessed contingency table is decomposed (Generalized Singular Value Decomposition) into a orthogonal profiles of rows and columns.

The Game of Thrones

This example uses the Kaggle dataset about battles in *Game of Thrones* (see <http://bit.ly/2qJufQ2>). The **battles** table presents number of battles in which given House (row) was an attacking army with a given attacking strategy (column).

`head(battles)`

House	Battle_Type				
	ambush	pitched battle	raiding	siege	
Baratheon	0	2	0	3	
Frey	1	0	0	2	
Lannister	1	6	0	2	
Stark	5	3	0	0	
Tully	3	0	0	0	

Use the **FactoMineR::CA()** function for correspondence analysis.

```
library("FactoMineR")
res.ca <- CA(battles, graph = FALSE)
summary(res.ca)
```

The chi square of independence between the two variables is equal to 25.82556 (p-value = 0.103814).

Eigenvalues	Dim.1	Dim.2	Dim.3
Variance	0.362	0.201	0.099
% of var.	54.617	30.360	15.024
Cumulative % of var.	54.617	84.976	100.000

Rows	Iner*1000				Dim.1	ctr	cos2
Baratheon		113.333		0.861	26.281	0.839	
Bolton		45.299		-0.508	5.481	0.438	
Frey		78.632		0.227	1.099	0.051	
Greyjoy		83.333		-0.074	0.307	0.013	
...							

Columns	Iner*1000				Dim.1	ctr	cos2
ambush		217.236		-0.747	59.386	0.989	
pitched battle		160.897		0.427	16.766	0.377	
raiding		99.359		-0.122	0.106	0.004	
siege		184.701		0.579	23.742	0.465	
...							

This cheatsheet presents functions from **FactoMineR** package (Francois Husson, Julie Josse, Sebastien Le, Jeremy Mazet, <http://factominer.free.fr/>) in version 1.35 and **factoextra** package (Alboukadel Kassambara, Fabian Mundt, <http://www.sthda.com/english/rpkgs/factoextra/>) in version 1.0.4

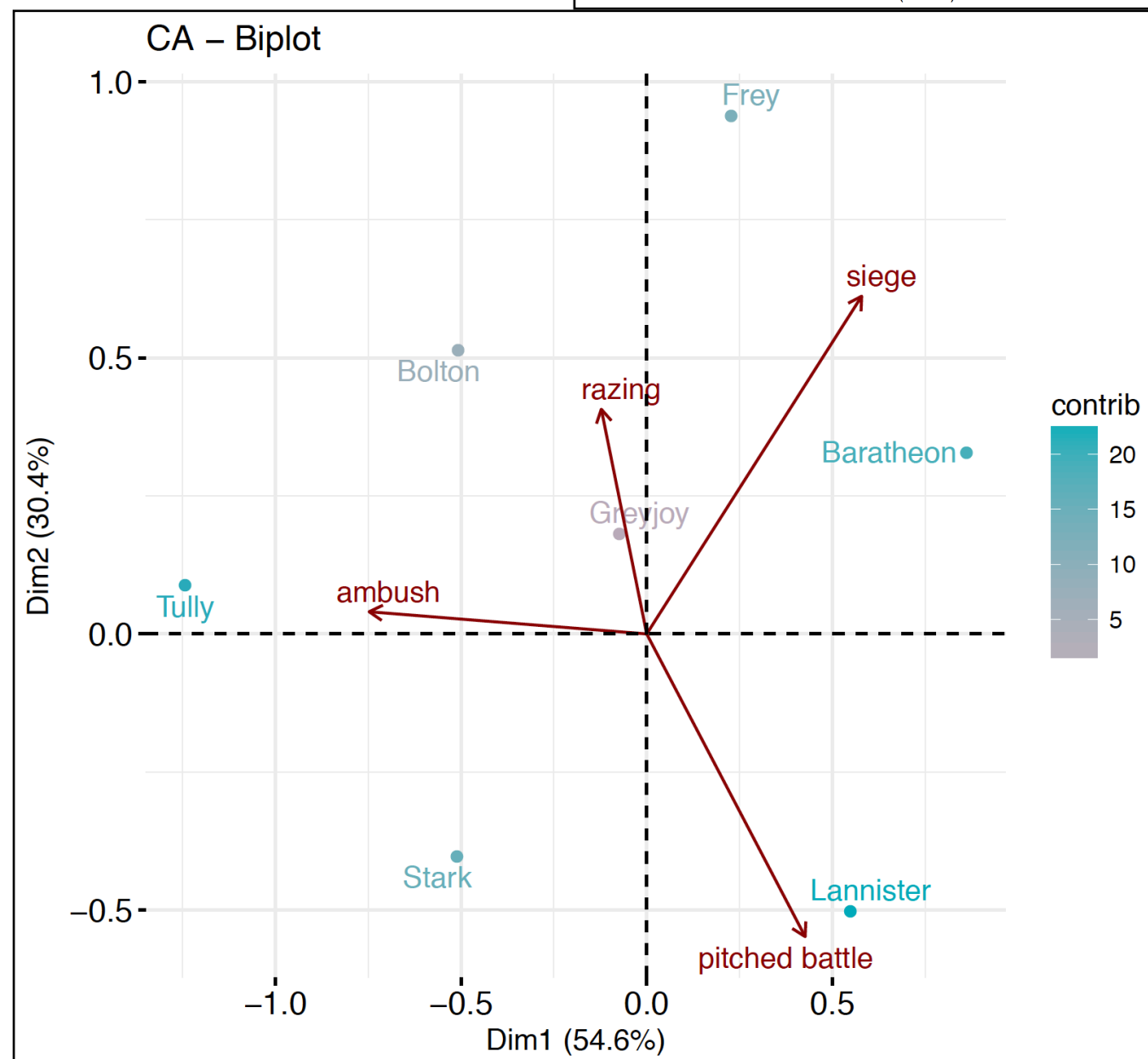
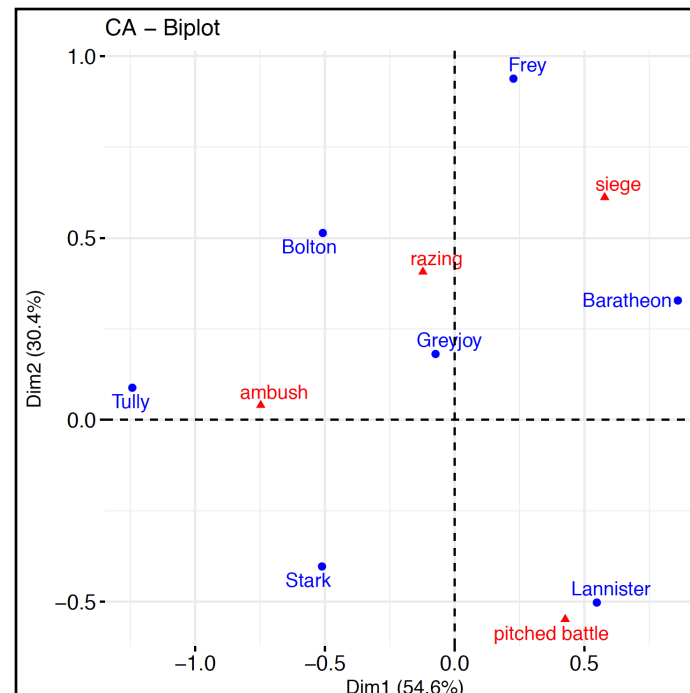
FactoMineR (for multivariate data analysis) and factoextra (for visualisation of CA results)

Biplots

Use the **factoextra::fviz_ca_biplot()** function to plot the biplot for a given model. Use **fviz_ca_row()** or **fviz_ca_col()** if you are interested only in profiles of rows/columns.

Consult the help page for numerous parameters of the **fviz_ca_biplot()** function, that allows to customise of biplot plots.

```
# Basic biplot, results in the right plot ->
fviz_ca_biplot(res.ca)
# Customised biplot, results in the bottom plot
fviz_ca_biplot(res.ca, repel = TRUE,
  arrow = c(FALSE, TRUE),
  col.row = "contrib",
  col.col = "red4",
  gradient.cols = c("#BBAFBB", "#00AFBB"))
```

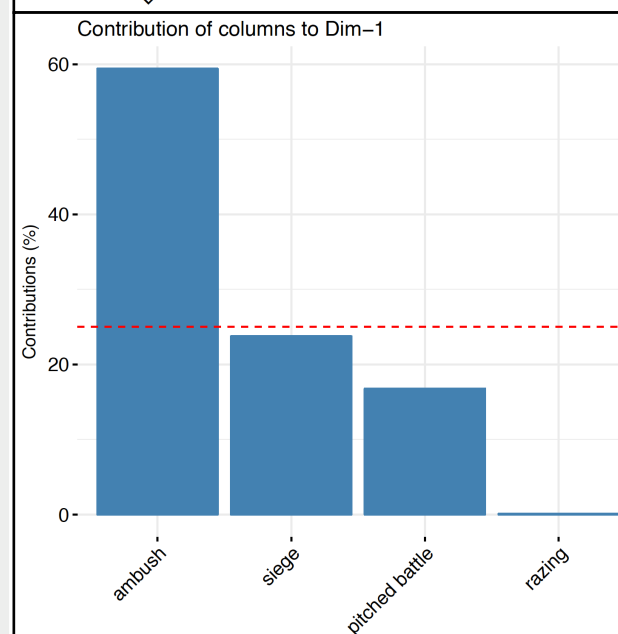
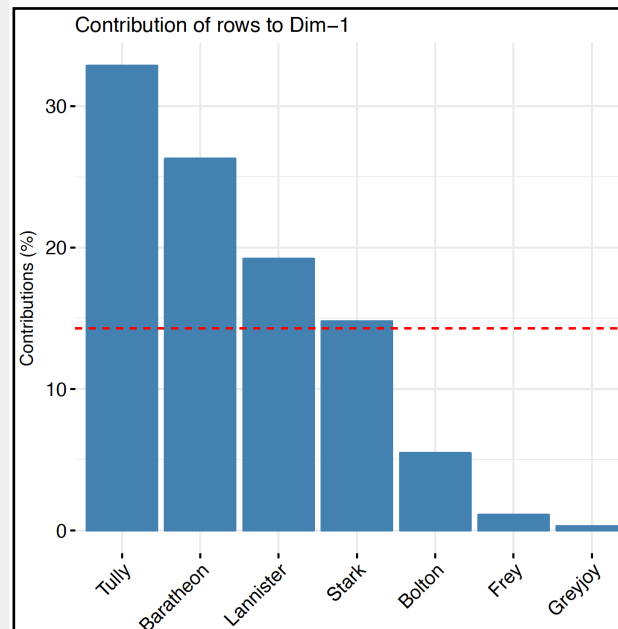


Row/Column contributions plot

Use the **factoextra::fviz_contrib()** function to plot contributions of selected dimension (rows or columns) onto a selected axis.

Numerical values for contributions may be extracted with **get_ca_row()** or **get_ca_col()** functions.

```
# Row contributions to Dimension 1
fviz_contrib(res.ca, choice = "row", axes = 1)
# Column contributions to Dimension 1
fviz_contrib(res.ca, choice = "col", axes = 1)
# Numeric summaries
get_ca_col(res.ca)
Correspondence Analysis - Results for columns
=====
Name      Description
1 "$coord" "Coordinates for the columns"
2 "$cos2"  "Cos2 for the columns"
3 "$contrib" "contributions of the columns"
4 "$inertia" "Inertia of the columns"
```



CC BY Przemysław Biecek <http://github.com/pbiecek>
<https://creativecommons.org/licenses/by/4.0/>