

Running a Correspondence Analysis (CA) from a raw data table with XLSTAT and plot a 3D representation with XLSTAT-3D Plot

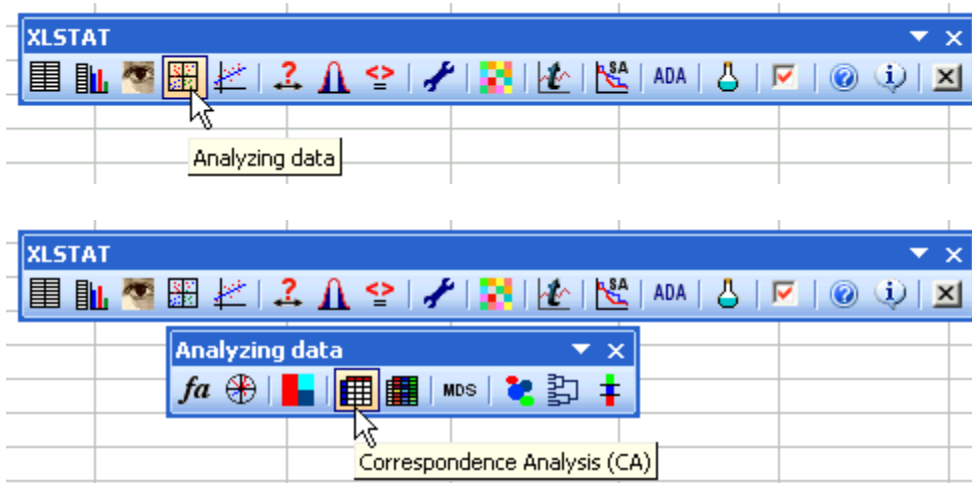
[demoCorrespondenceAnalysis2.xls](#)

Dataset for running a Correspondence Analysis from a raw data table

An Excel sheet with both the data and the results can be downloaded by clicking [here](#). The data correspond to the list of foreigner soccer player in premier league and their nationality. We want to study the distribution of the foreign players in the English clubs.

Setting up a Correspondence Analysis from a raw data table

Once XLSTAT is open, select the **Analyzing data / Correspondence analysis** command, or click on the corresponding button of the **Analyzing Data** toolbar (see below).



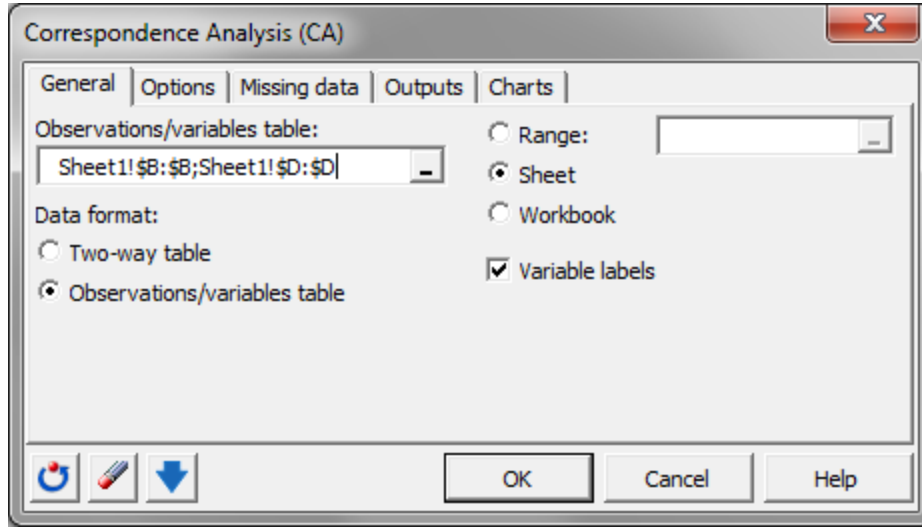
Once you have clicked on the button, the Correspondence analysis dialog box appears.

In the field **Observations/variables table**, select the columns **Club** and **Region** on the Excel sheet.

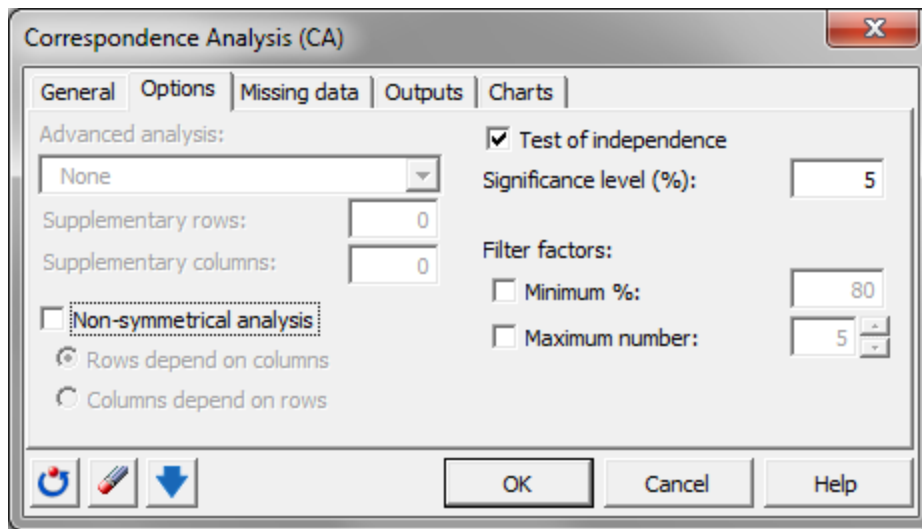
The data are in an Observations/variables format, tick the corresponding option.

As the names of the columns are included, the **Variable labels** option should be selected as well.

Choose the **Sheet** option for the output.

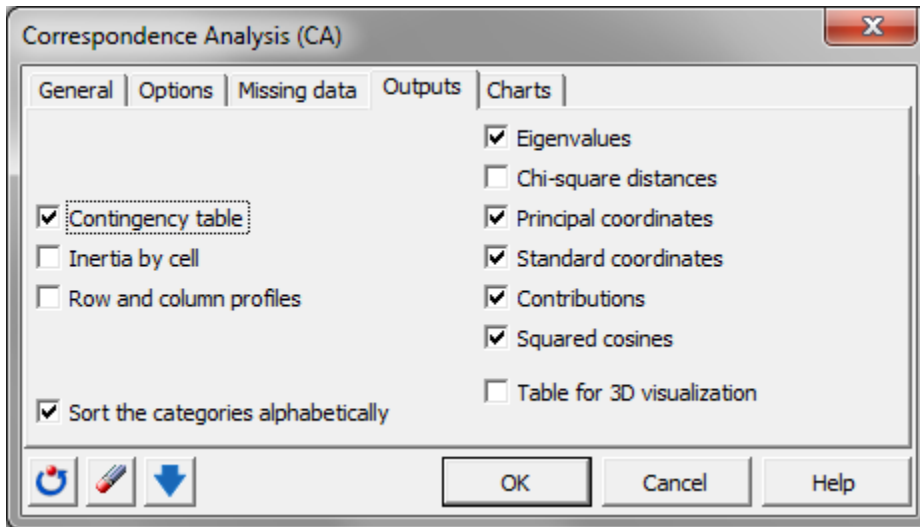


On the tab **Options** tick the **Test of independence** and leave the significance level to 5.



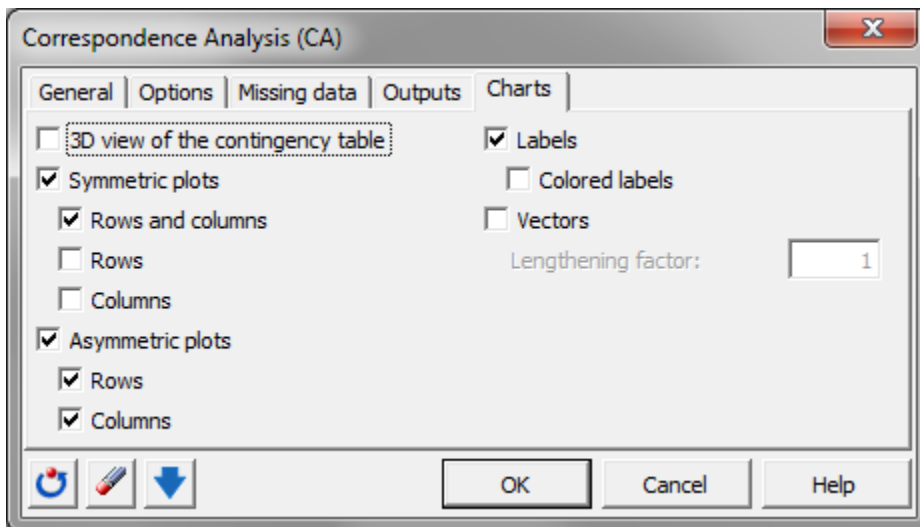
In the **Outputs** section, select the following options:

- Contingency table
- Eigenvalues
- Principal coordinates
- Standard coordinates
- Contributions
- Squared cosines



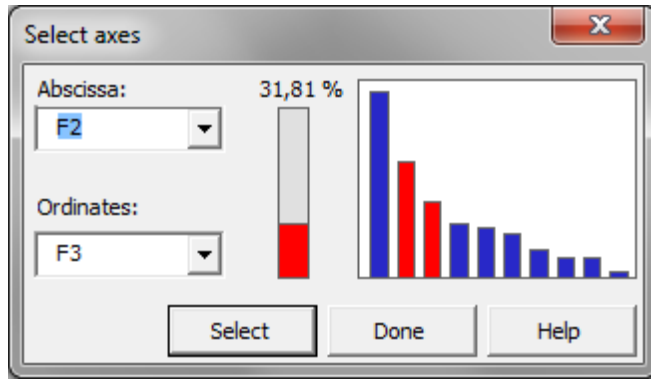
Go to the last tab **Charts** and enable the:

- Symmetric plots
- Asymmetric plots
- Labels



Click on **OK**.

As the model needs more than two factors. Click first on **Select** to select the plot F1-F2. Then change the **Abscissa** to F2. It will change the **Ordinates** to F3. Click again on **Select**. This way we will have two plots: F1-F2 and F2-F3. Click on **Done**.



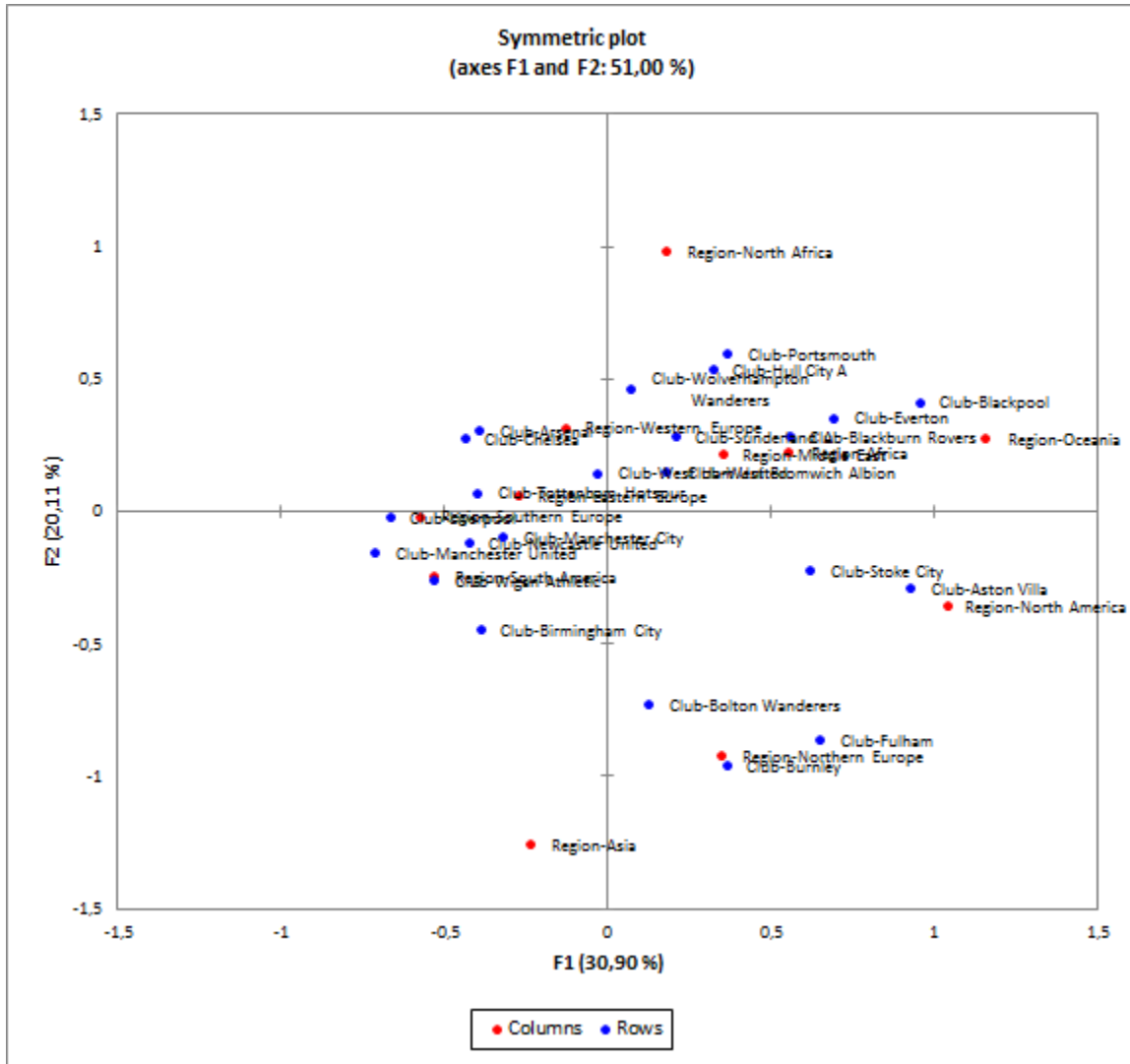
Interpreting the results of a Correspondence Analysis

The first result is the contingency table and then the test of independence between the rows and columns.

The p-value of 0.008 is inferior to 5% thus the null hypothesis should be rejected. This means that the distribution of nationality is not random in the UK clubs.

Test of independence between the rows and the columns:	
Chi-square (Observed value)	273,215
Chi-square (Critical value)	255,602
DF	220
p-value	0,008
alpha	0,05
Test interpretation:	
H0: The rows and the columns of the table are independent.	
Ha: There is a link between the rows and the columns of the table.	
As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.	
The risk to reject the null hypothesis H0 while it is true is lower than 0,05.	

Then you have the symmetric plots. From the first plot you can see that the clubs such as Aston Villa and Stoke City have more North-American players than the rest of the teams. In the same way, Burney has a lot of Northern European players.



Creating a 3-D plot for the a Correspondence Analysis results

We will now do a plot in 3-dimensions to have a better representation of the points.

First we will make a table containing both the first 3 principal coordinates for the clubs and geographic areas and the sum of the cosines for those 3 factors.

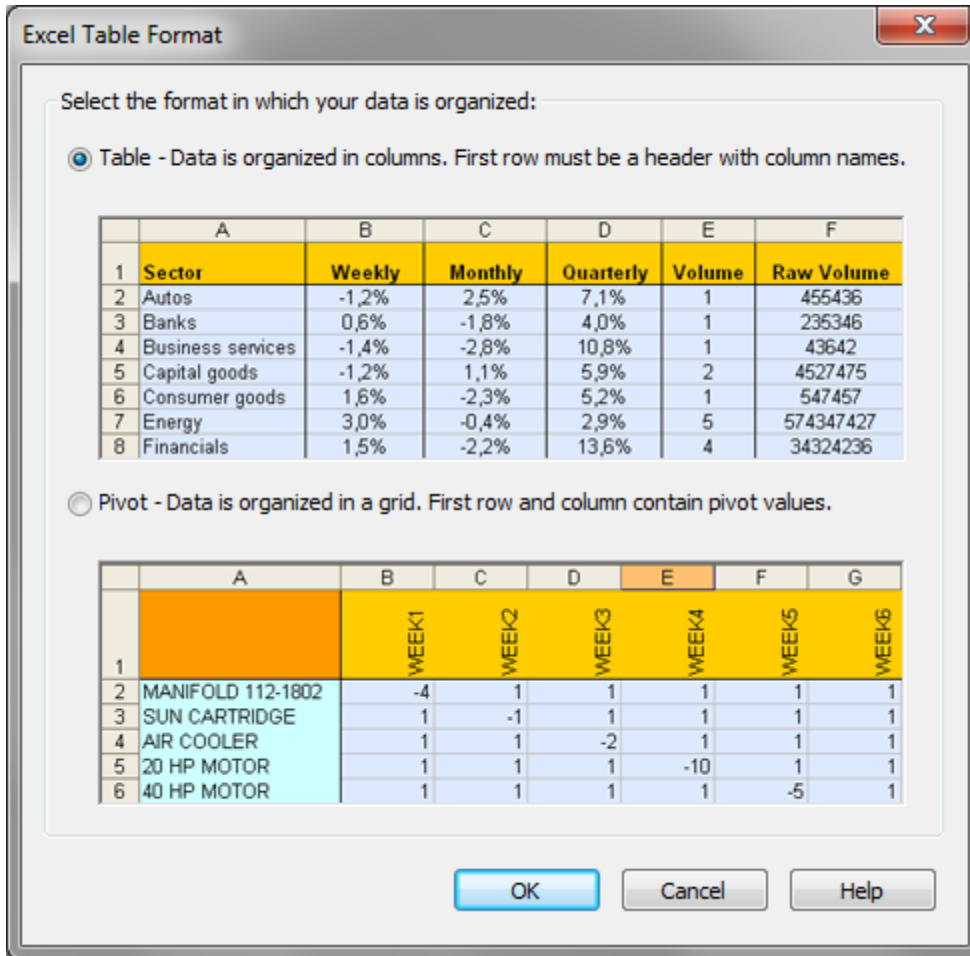
The sum of the cosines for the 3 factors give an idea of how well is represented the sample in the 3-D space.

Add a last column to have the information about the rows and columns. The rows are the clubs and the columns the regions. Make a category variable with R and C to describe each sample.

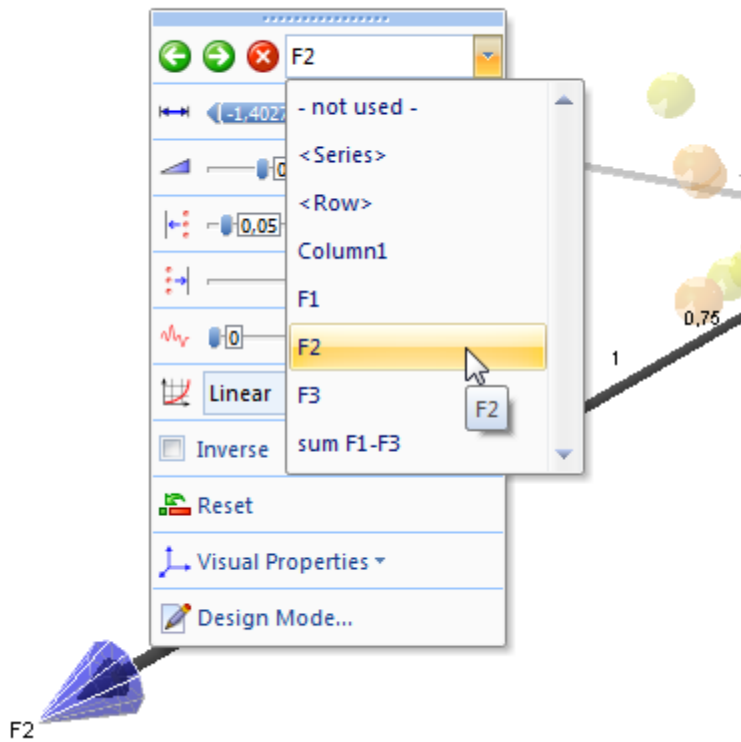
	F1	F2	F3	sum cos ² F1-F2	Rows/Column
Club-Arsenal	-0,392	0,304	-0,105	0,574293227	R
Club-Aston Villa	0,929	-0,293	0,282	0,531941297	R
Club-Birmingham City	-0,389	-0,445	-0,067	0,579242975	R
Club-Blackburn Rovers	0,557	0,285	-0,002	0,573023162	R
Club-Blackpool	0,955	0,413	0,681	0,728883637	R
Club-Bolton Wanderers	0,123	-0,731	0,087	0,675735368	R
Club-Burnley	0,366	-0,961	-0,273	0,723578625	R
Club-Chelsea	-0,437	0,276	0,022	0,567392076	R
Club-Everton	0,693	0,353	0,326	0,724790025	R
Club-Fulham	0,650	-0,866	-0,306	0,941493687	R
Club-Hull City A	0,322	0,538	-0,035	0,554851928	R
Club-Liverpool	-0,665	-0,024	-0,405	0,687760626	R
Club-Manchester City	-0,323	-0,095	-0,117	0,543605066	R
Club-Manchester United	-0,712	-0,158	0,383	0,777550953	R
Club-Newcastle United	-0,425	-0,119	-0,098	0,709232928	R
Club-Portsmouth	0,369	0,594	-0,905	0,871529738	R
Club-Stoke City	0,618	-0,221	-0,080	0,497919588	R
Club-Sunderland A	0,209	0,281	0,432	0,544325199	R
Club-Tottenham Hotspur	-0,398	0,070	0,409	0,369587381	R
Club-West Bromwich Albion	0,182	0,150	0,043	0,101006673	R
Club-West Ham United	-0,032	0,145	0,204	0,318023456	R
Club-Wigan Athletic	-0,534	-0,261	0,279	0,542051688	R
Club-Wolverhampton Wanderers	0,070	0,463	-0,021	0,253461418	R
Region-Africa	0,555	0,220	-0,212	0,701909442	C
Region-Asia	-0,238	-1,257	0,350	0,437709516	C
Region-Eastern Europe	-0,274	0,059	0,258	0,304251431	C
Region-Middle East	0,352	0,217	0,566	0,271089666	C
Region-North Africa	0,181	0,982	-1,073	0,705046055	C
Region-North America	1,040	-0,356	0,365	0,718711463	C
Region-Northern Europe	0,351	-0,925	-0,439	0,930032821	C
Region-Oceania	1,156	0,277	0,489	0,576222934	C

Select the full table and go to the menu **Visualizing data** and select the option **XLSTAT-3DPlot**.

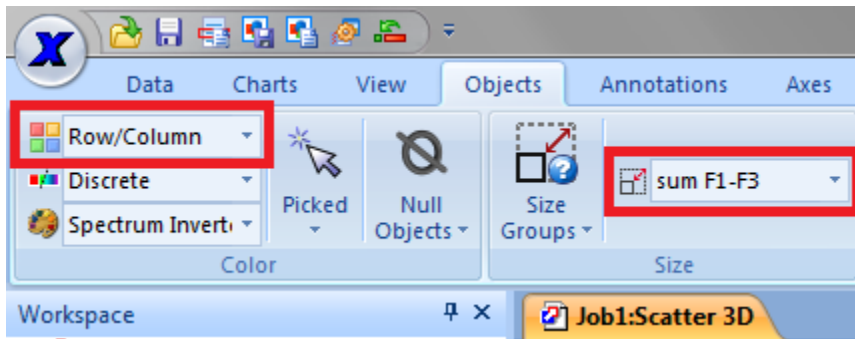
When prompt select the format of your data as **Table**.



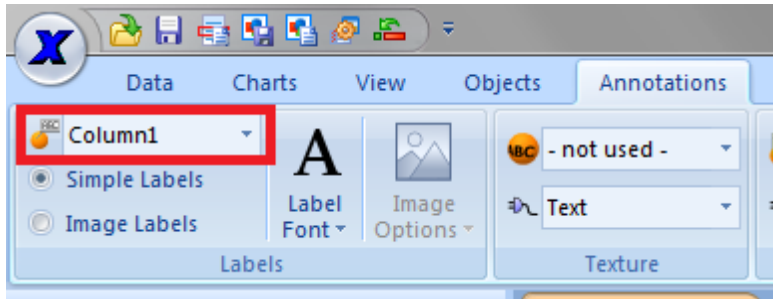
You will need to specify the axes. Do so by a right click and select in the dropdown the appropriate variable to use. For the 3 axes we utilize: F1, F2 horizontally and F3 vertically. You also need to set the size of the axis so as to have an orthonormal plot. For example use for all the axes : -1.5 and 1.5 as limits.



For the color and size of the dot you can use the sum of cosines. Go to the tab **Objects** and modify the color and size sections.



Finally we can add the labels by going into the tab **Annotations** and selecting "Column1" as the label.



Here is your 3-D representation.

