

Creating a Parallel Coordinates Visualization with XLSTAT

[demoPCor.xls](#)

Dataset to create a Parallel Coordinates plot

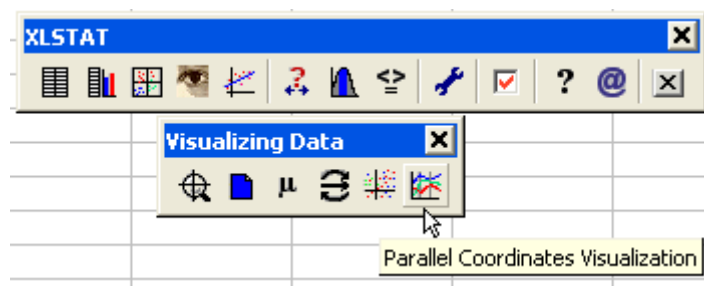
An Excel sheet with both the data and the results can be downloaded by clicking [here](#).

The data used in this tutorial have been extracted from a 1994 survey by the [American Census Bureau](#). The data set is such that half of the observations corresponds to individuals with a revenue below 50k\$, and the other half to individuals with a revenue greater than 50k\$. For all the individuals in the sample, the country of origin is the USA.

Our goal is to visualize if some of the descriptors (Age, Number of years of study, Race, Sex, Hours-per-week) influence the Revenue of the individuals.

Setting up the Parallel Coordinates plot dialog box

Once XLSTAT-Pro is activated, select the **XLSTAT / Visualizing data / Parallel Coordinates** command, or click on the corresponding button of the **Visualizing Data** toolbar (see below).

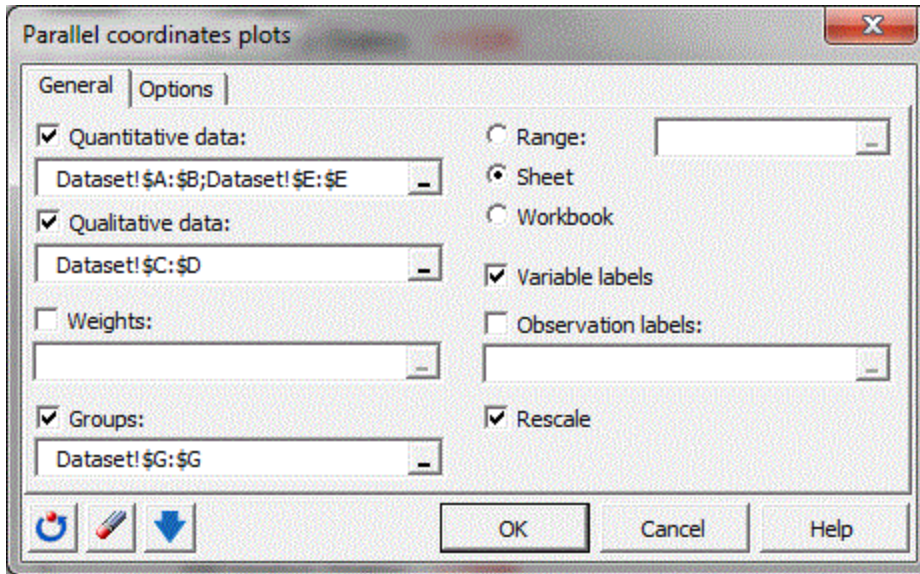


Once you have clicked on the button, the dialog box appears.

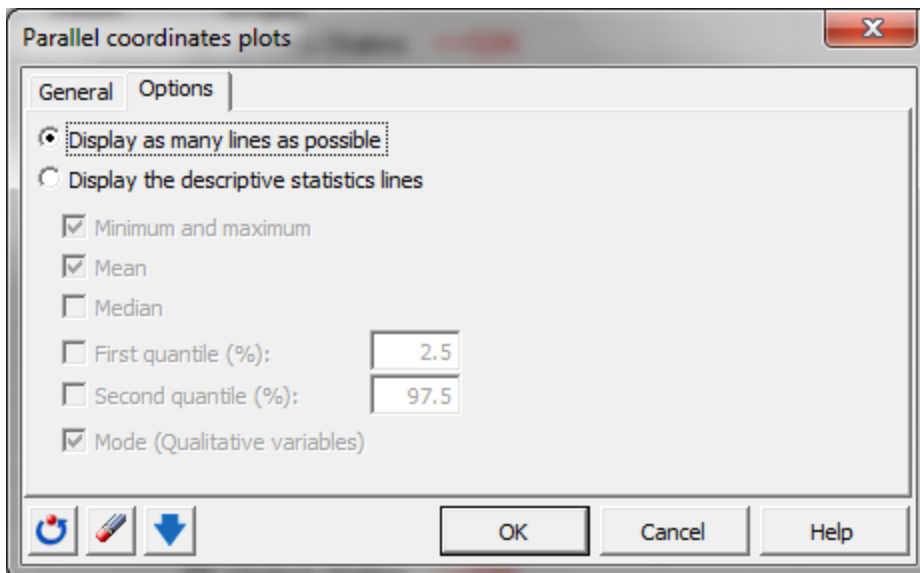
Select the data on the Excel sheet. This tool accepts that you mix numerical and nominal variables. The **Groups** information is used to color the lines.

We activated the **Mean lines** option to let XLSTAT display for each group a line that corresponds to the mean of the quantitative variables and to the mode of the nominal variables.

The **Rescale** option allows you to compare how the data are distributed for all the variables and facilitates the visualization.



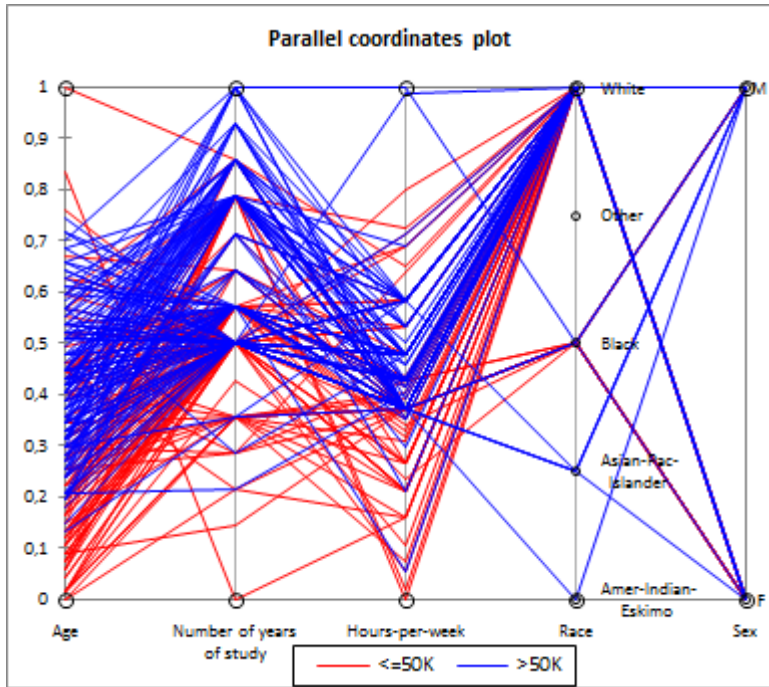
Move on to the **Chart** tab where you can decide how the plot will look like. Select the option **Display as many lines as possible**.



Then, after you have clicked on the **OK** button, a chart is displayed on a new Excel sheet (because the Sheet option has been selected for outputs).

Interpreting a Parallel Coordinates plot

In the Parallel Coordinates plot you can see that being a white man, in the upper bracket of ages, with a higher number of years of study, and a high level of hours worked per week increases the likelihood of having a revenue above 50k\$. However we notice that the number of hours is not very discriminant as the means of the two groups (50k\$) are close.



This video shows you how to run this tutorial.

http://www.youtube.com/watch?feature=player_embedded&v=efnYO_PrX38