

GrapheR – User manual

Maxime Hervé

For any question, comment or suggestion: maxime.herve@laposte.net

Contents

1	Introduction	3
2	Use	4
2.1	Launching and dataset loading	4
2.2	Choice of the type of graph to draw	7
2.3	Available types of graph	7
2.3.1	Histogram	7
2.3.2	Box-and-whisker plot	10
2.3.3	Bar plot	14
2.3.4	Pie chart	19
2.3.5	Curve	23
2.3.6	Scatter plot	27
2.4	Dealing with graphics devices	32
2.5	Toolbar	33
2.5.1	Add a vertical line	34
2.5.2	Add a horizontal line	34
2.5.3	Add any line	35
2.5.4	Add a theoretical distribution curve	35
2.5.5	Add text	36
2.5.6	Add <i>p-values</i>	36
2.5.7	Save a graph	38
3	Add a new language to GrapheR	40

Chapter 1

Introduction

The GrapheR package is a multiplatform user interface for drawing highly customizable graphs in R. It aims to be a valuable help to quickly draw publishable graphs without any knowledge of R commands.

Six kinds of graphs are available:

- histogram
- box-and-whisker plot
- bar plot
- pie chart
- curve
- scatter plot

GrapheR was built on the `tcltk` package, and consequently bugs can happen if R is configured in the MDI (Multiple-Document Interface) mode. It is recommended to configure it in SDI (Single-Document Interface) mode before starting GrapheR.

GrapheR needs to function at least the 2.10.0 version of R and two additional packages: `tcltk` and `mgcv`. Under Mac OS X, Tcl/Tk must have already been installed (for more information see [here](#)).

Chapter 2

Use

2.1 Launching and dataset loading

At launch, GrapheR proposes to choose the user language (Figure 2.1). In the 1.0 version, English and French are available. It is not possible to change this choice later, unless closing and relaunching the interface by the `GrapheR.begin()` function.



Figure 2.1: Choice of user language (display under Windows 7)

The next window permits loading of the dataset which will be used to draw the graphs (Figure 2.2).

The upper frame contains options to be chosen before loading the dataset:

- file type: the 1.0 version of GrapheR deals with `.txt` and `.csv` files
- column separator: white space (space, tab or entry), comma or semicolon. If in doubt, open the file in the notepad to check
- decimal separator: period or comma. If in doubt, open the file in the notepad to check

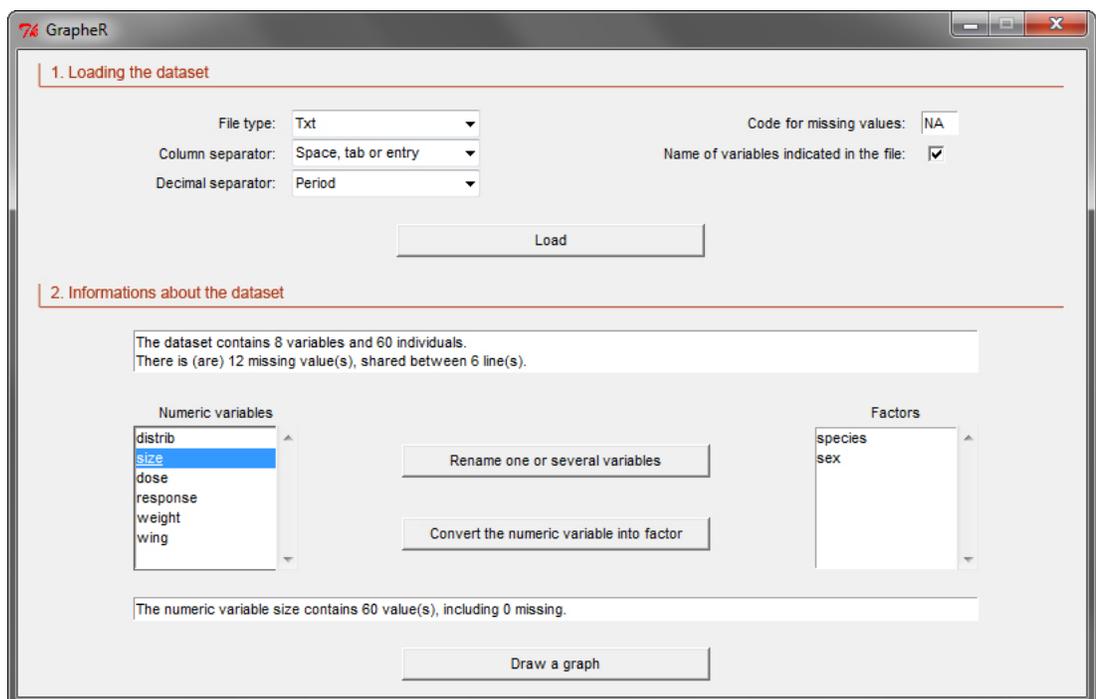


Figure 2.2: Loading and modification of the dataset

- code for missing values: they are coded as NA (Not Available) in R. It is important to correctly fill in this box to let R interpret these values as missing
- name of variables indicated in the file: tick it if the columns of the dataset have a title

The lower frame gives information about the dataset once it has been loaded. Some modifications are possible if necessary:

- rename one or several variables (Figure 2.3): this is useful if the dataset does not contain the name of the variables (in this case R gives as names V1, V2,...). Click on the variable to be renamed in the list, change it and click on “Rename”. Once all modifications have been made, click on “Ok”. Respect the basic rules of R: do not begin the variable name by a number, do not insert spaces and avoid accents

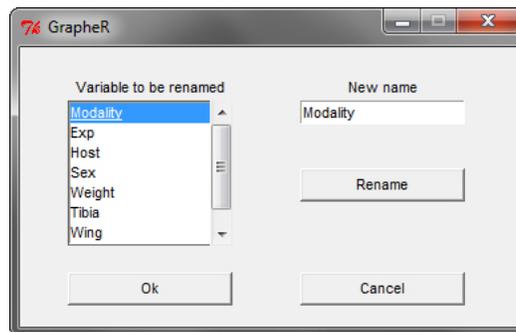


Figure 2.3: Rename one or several variables

- convert a numeric variable into factor (Figure 2.4): this conversion is useful if a factor of the dataset is coded numerically (ex: binary factor 0/1 or ordinal factor; R recognizes them as numeric) or if it is necessary to group values of a numeric variable into classes. Several options are available:
 - no grouping: each value of the variable to transform becomes a level of the factor (the number in brackets corresponds to the number of levels obtained if this option is chosen)
 - grouping by classes of same length: each class interval will have the same length (ex : [0-9], [10-19]...). The number of classes is defined by the cursor
 - grouping by classes of same size: each class interval will contain the same number of individuals. The number of classes is defined by the cursor

Once the dataset is loaded and ready, click on “Draw a graph”.

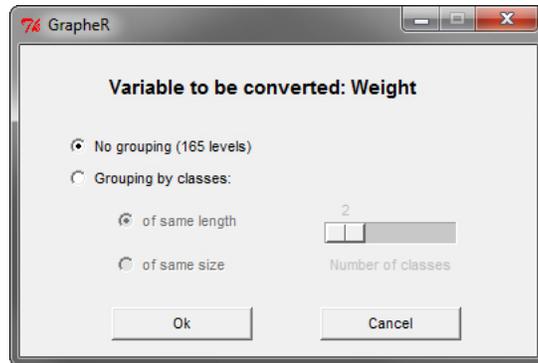


Figure 2.4: Convert a numeric variable into factor

2.2 Choice of the type of graph to draw

Just click on the desired kind of graph (Figure 2.5). It is possible to come back to the last window to load another dataset or modify the already loaded one.

2.3 Available types of graph

2.3.1 Histogram

The window contains four frames corresponding to four parameter categories (Figure 2.6):

- general parameters:
 - variable to be represented: has to be numeric
 - factor: it is possible to represent values of the numeric variable corresponding to a level of a given factor. In this case choose a factor, then the desired level. If no factor is chosen the histogram will represent all values of the numeric variable
 - type of histogram: the histogram can represent either strength, frequency or density of each interval. Only “densities” type allows to draw the distribution curve of the data
- title of the graph:
 - title: optional, the graph will not have a title if this box is empty
 - color
 - size
- axes:

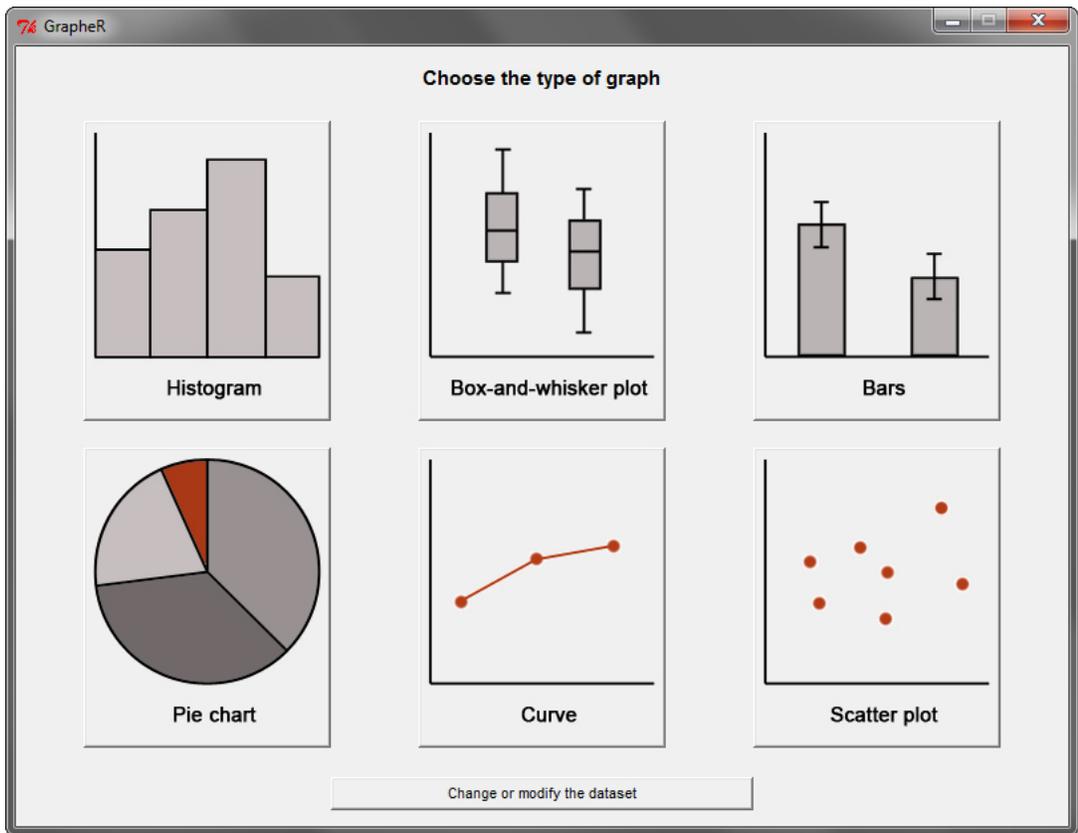


Figure 2.5: Choice of the type of graph to draw

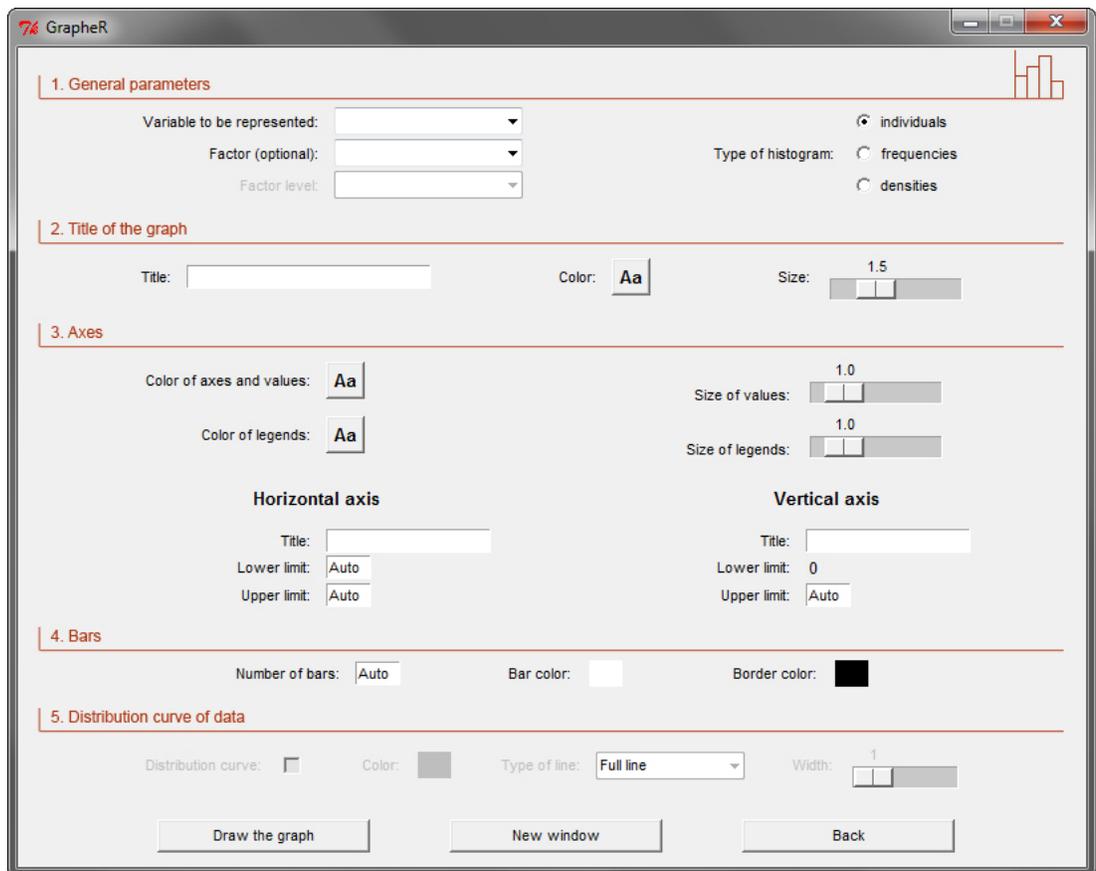


Figure 2.6: Histogram

- color of axes and axis values, size of axis values, color of legends (name of axes) and size of legends: these four parameters are the same for horizontal and vertical axes
- title of each axis
- lower and upper limits of axes: the value of these parameters can be changed (for decimal numbers, the decimal separator is a period) or left on “Auto”. The lower limit of the vertical axis is necessarily 0
- bars:
 - number of bars: the value of this parameter has to be an entire number or left on “Auto”. In the case of an automatically calculated value, the Sturges rule is used. Due to the grouping method of \mathbb{R} , the number of bars can be slightly different from the given number
 - color of bars
 - color of bar borders
- distribution curve of the data:
 - representation: the curve can be added only if the histogram is drawn in densities
 - color of line
 - type of line: full, large dashed or fine dashed
 - width of line

If all parameters are left on the default value (except the variable to be represented), the graph looks like the Figure 2.7.

2.3.2 Box-and-whisker plot

The window contains five frames corresponding to five parameter categories (Figure 2.8):

- general parameters:
 - variable to be represented: has to be numeric
 - factor delimiting boxes
 - orientation of the graph: corresponds to the orientation of boxes, vertical or horizontal
- title of the graph:
 - title: optional, the graph will not have a title if this box is empty
 - color
 - size

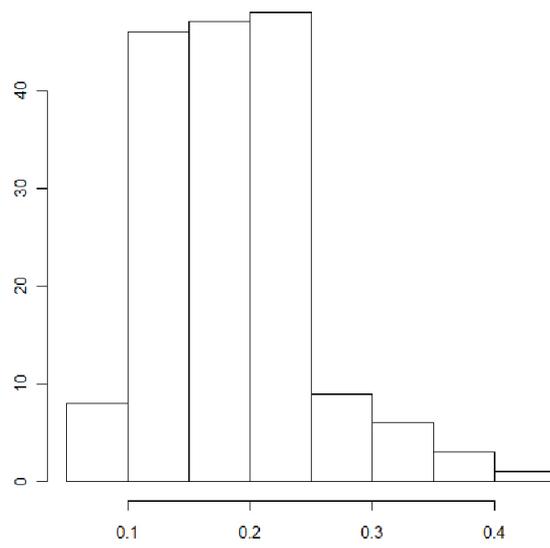


Figure 2.7: Default histogram

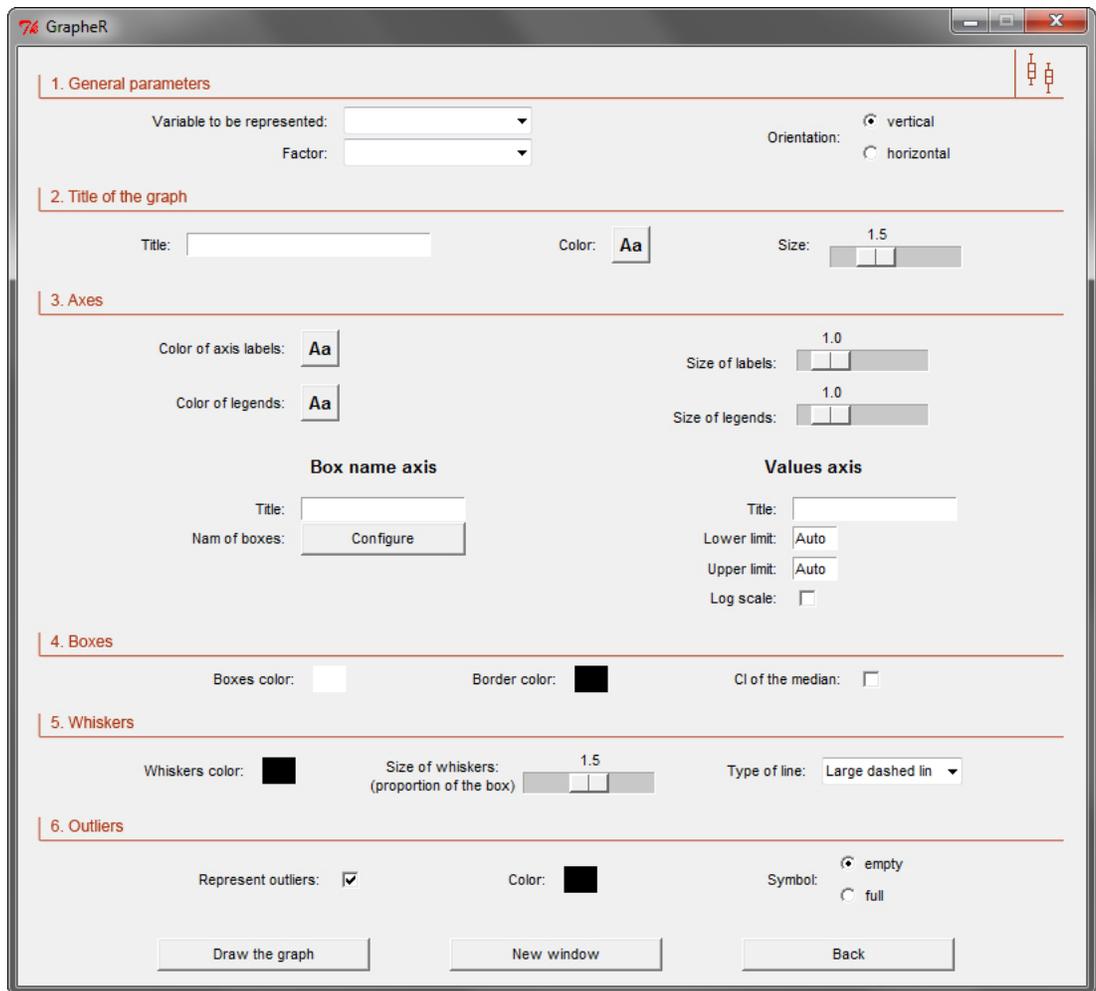


Figure 2.8: Box-and-whisker plot

- axes:
 - color of axis labels (numeric values and box names), size of axis labels, color of legends (name of axes) and size of legends: these four parameters are the same for horizontal and vertical axes
 - title of each axis
 - name of boxes: by default these are levels of the chosen factor, but it is possible to change them (which does not modify the name of the factor levels) (Figure 2.9)

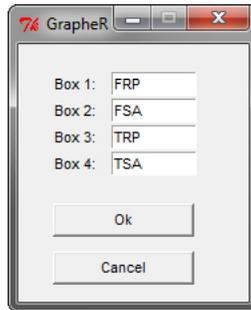


Figure 2.9: Name of boxes

- lower and upper limits of the values axis (horizontal or vertical depending on the orientation of the graph): the value of these parameters can be changed (for decimal numbers, the decimal separator is a period) or left on “Auto”
- logarithmic scale of the values axis
- boxes :
 - color of boxes
 - color of borders
 - confidence interval of the median: represented as notches in boxes around the median
- whiskers:
 - color
 - size (in proportion of the interquartile, *i.e.* the length of the corresponding box)
 - type of line : full, large dashed or fine dashed
- outliers (*i.e.* individuals beyond the whisker extremity):
 - representation

- color
- symbol: represented individually as circles, empty or full

If all parameters are left on the default value (except the variable to be represented and the factor), the graph looks like the Figure 2.10

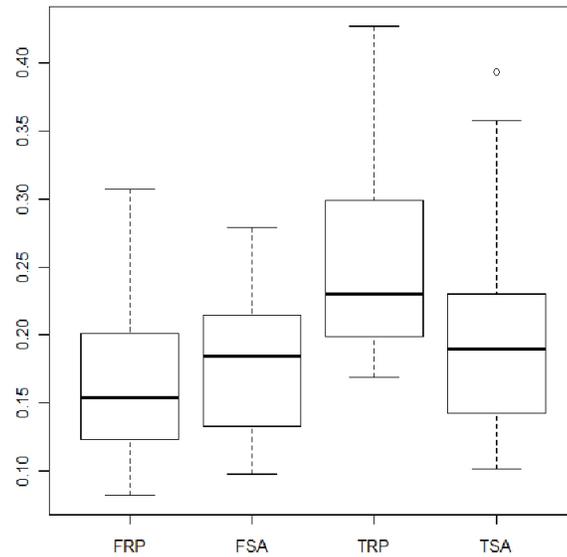


Figure 2.10: Default box-and-whisker plot

2.3.3 Bar plot

The window contains six frames corresponding to six parameter categories (Figure 2.11):

- general parameters:
 - variable to be represented: can be numeric or factorial
 - factor 1: obligatory, it delimits groups to be compared
 - factor 2: optional, it can be defined only if the variable to be represented is numeric
 - type of graph: if the variable to be represented is numeric, bars can represent sums or means of its values by level of the factor 1 (and

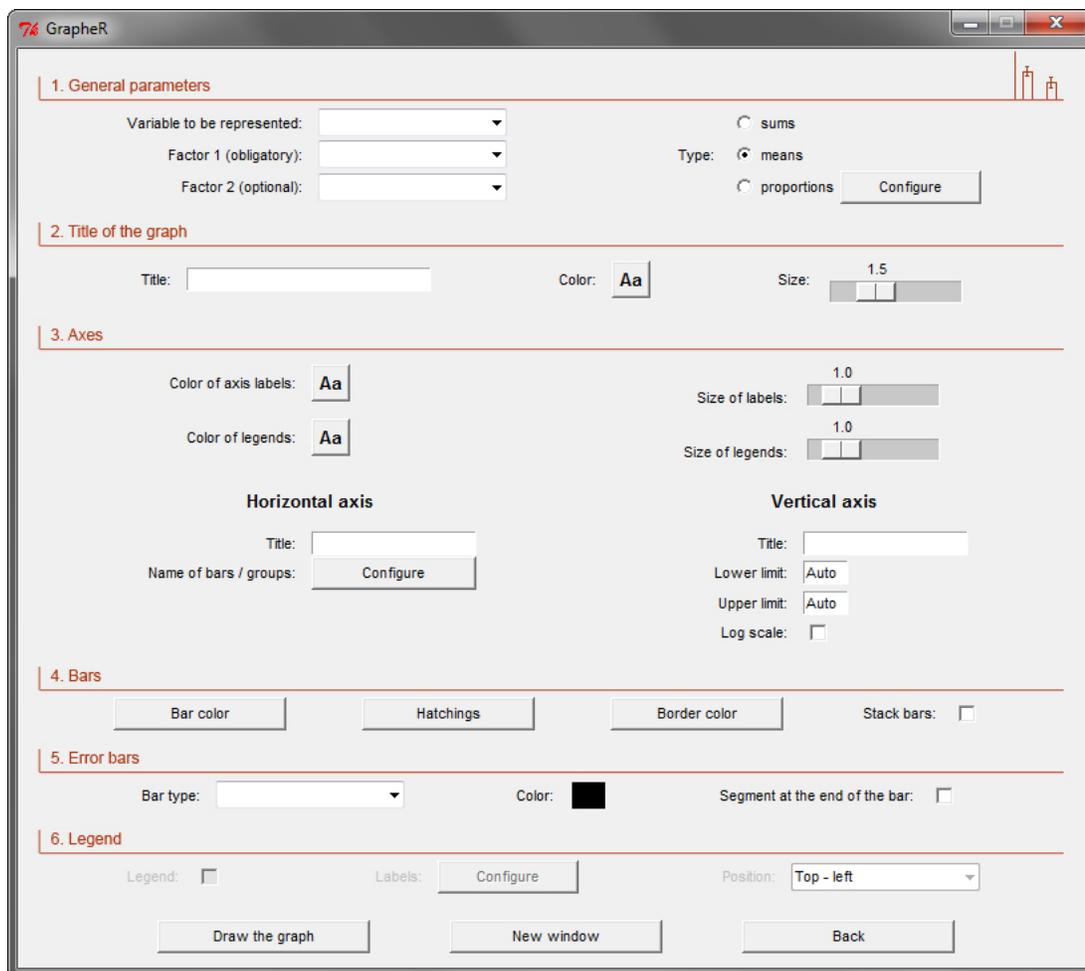


Figure 2.11: Bar plot

eventually of the factor 2). If the variable to be represented is a factor, bars can represent proportions of levels of this factor by level of the factor 1. It is possible to choose the proportions to be represented by clicking on “Configure” (Figure 2.12). No error bar can be drawn if the “sums” type is chosen

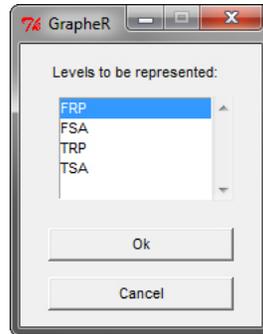


Figure 2.12: Proportions to be represented

- title of the graph:
 - title: optional, the graph will not have a title if this box is empty
 - color
 - size
- axes:
 - color of axis labels (numeric values and bar names), size of axis labels, color of legends (name of axes) and size of legends: these four parameters are the same for horizontal and vertical axes
 - title of each axis
 - name of bars: by default these are levels of the factor 1, but it is possible to change them (which does not modify the name of the factor levels) (Figure 2.13)
 - lower and upper limits of the vertical axis: the value of these parameters can be changed (for decimal numbers, the decimal separator is a period) or left on “Auto”
 - logarithmic scale of the vertical axis
- bars:
 - color of bars: if factor 2 is not defined all bars have the same color (Figure 2.14), otherwise a different color can be attributed to each level of the factor 2 (Figure 2.15)

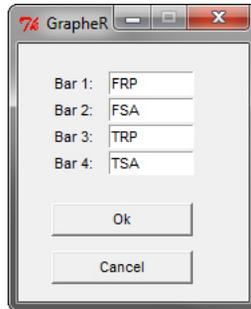


Figure 2.13: Name of bars (if factor 2 is not defined) or groups (if it is)

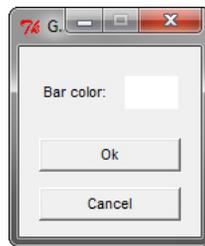


Figure 2.14: One color of bars

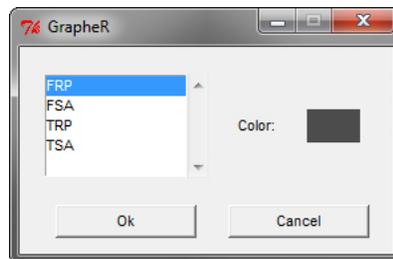


Figure 2.15: Several colors of bars

- hatchings inside bars: optional, if factor 2 is not defined all bars have the same pattern (Figure 2.16), otherwise a different pattern can be attributed to each level of the factor 2 (Figure 2.17). R does not represent hatchings if the scale of the ordinate axis is logarithmic

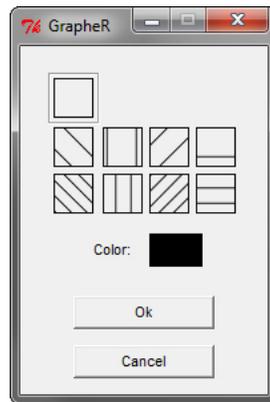


Figure 2.16: One pattern

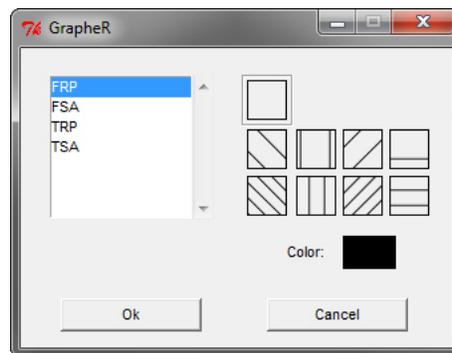


Figure 2.17: Several patterns

- color of borders: see color of bars
- stack bars: if the factor 2 is not defined this option has no effect, otherwise for each level of the factor 1 all bars corresponding to levels of the factor 2 are stacked. No error bar can be represented if this option is chosen
- error bars:
 - type of bars: if bars of the graph represent means, error bars can be standard deviations, standard errors or confidence intervals. If bars

- of the graph represent proportions, error bars can only be confidence intervals
- color
- segment at the end of the bars: if this option is not chosen error bars are represented only by a vertical segment
- legend:
 - representation: a legend can be added to the graph only if the variable to be represented is a factor (and if several proportions are represented) or if the factor 2 is defined. The legend uses the color of the bars, but R doesn't represent hatchings, even if bars contain it
 - labels: by default these are levels of the factor 2, but it is possible to change them (which doesn't modify the name of the factor levels). It is also possible to add a title to the legend (Figure 2.18)

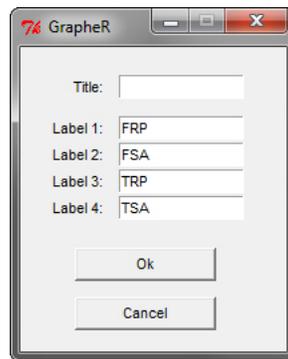


Figure 2.18: Legend

- position: position of the legend on the graph. If the legend overlays on bars think to modify limits of the vertical axis

If all parameters are left on default value (except the variable to be represented and the two factors), the graph looks like the Figure 2.19

2.3.4 Pie chart

The window contains five frames corresponding to five parameter categories (Figure 2.20):

- general parameters:
 - variable to be represented: can be numeric or factorial

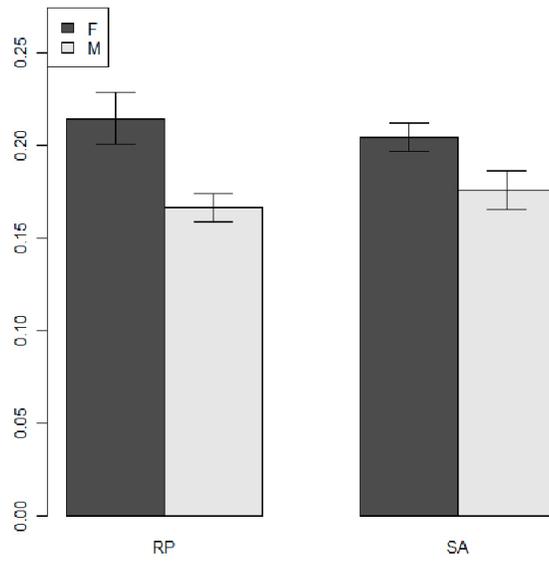


Figure 2.19: Default bar plot

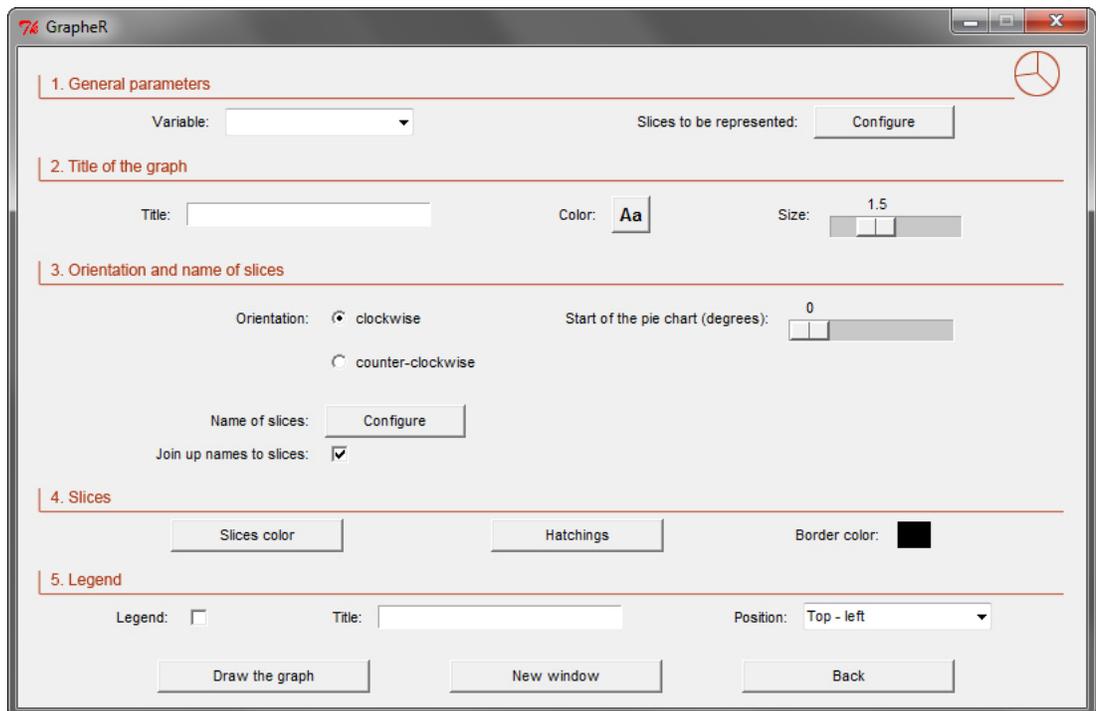


Figure 2.20: Pie chart

- slices to be represented: if the variable to be represented is numeric, slices correspond to the proportion of each value compared with the sum of all values. If the variable to be represented is a factor, slices correspond to the proportion of different levels. It is possible to choose the slices to be represented (the choice window looks like the Figure 2.12, except that by default all possible slices are selected)
- title of the graph:
 - title: optional, the graph will not have a title if this box is empty
 - color
 - size
- orientation and name of slices:
 - orientation: slices can be represented clockwise or counter-clockwise
 - start of the pie chart: the first slice starts by default at 12 o'clock (0 degree), but it is possible to change this value. Degrees are added according to the rotation direction of the pie chart
 - name of slices: by default these are levels of the variable to be represented if it is a factor or entire numbers from 1 if it is numeric, but it is possible to change them (which does not modify the name of the factor levels) (Figure 2.21)

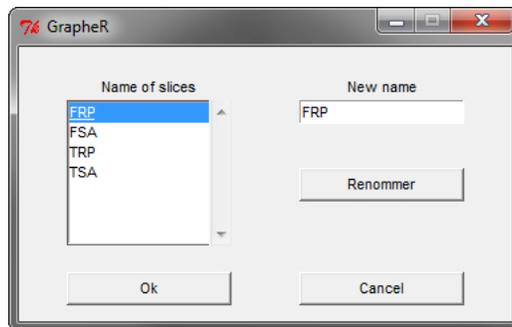


Figure 2.21: Name of slices

- join up names to slices: if this option is chosen names of slices are represented just around the graph, otherwise they are not represented
- slices:
 - color of slices: a different color can be attributed to each slice (the choice window looks like the Figure 2.15)
 - hatchings: a different patten can be attributed to each part (the choice window looks like the Figure 2.17)

- color of border: the border is the same for all the graph
- legend:
 - representation: the legend uses the color of the slices, but **R** does not represent hatchings, even if slices contain it
 - title : optional
 - position

If all parameters are left on default value (except the variable to be represented), the graph looks like the Figure 2.22

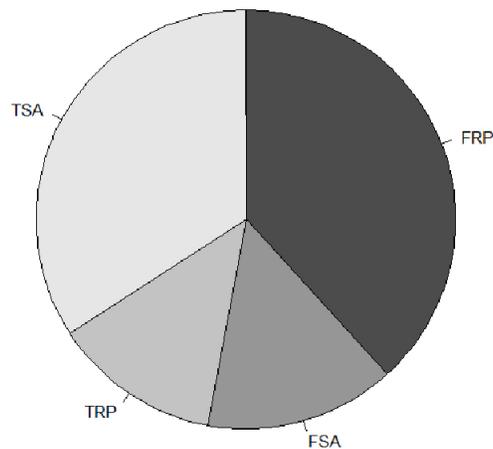


Figure 2.22: Default pie chart

2.3.5 Curve

The window contains six frames corresponding to six parameter categories (Figure 2.23):

- general parameters:
 - variable: variable of the horizontal axis, has to be numeric

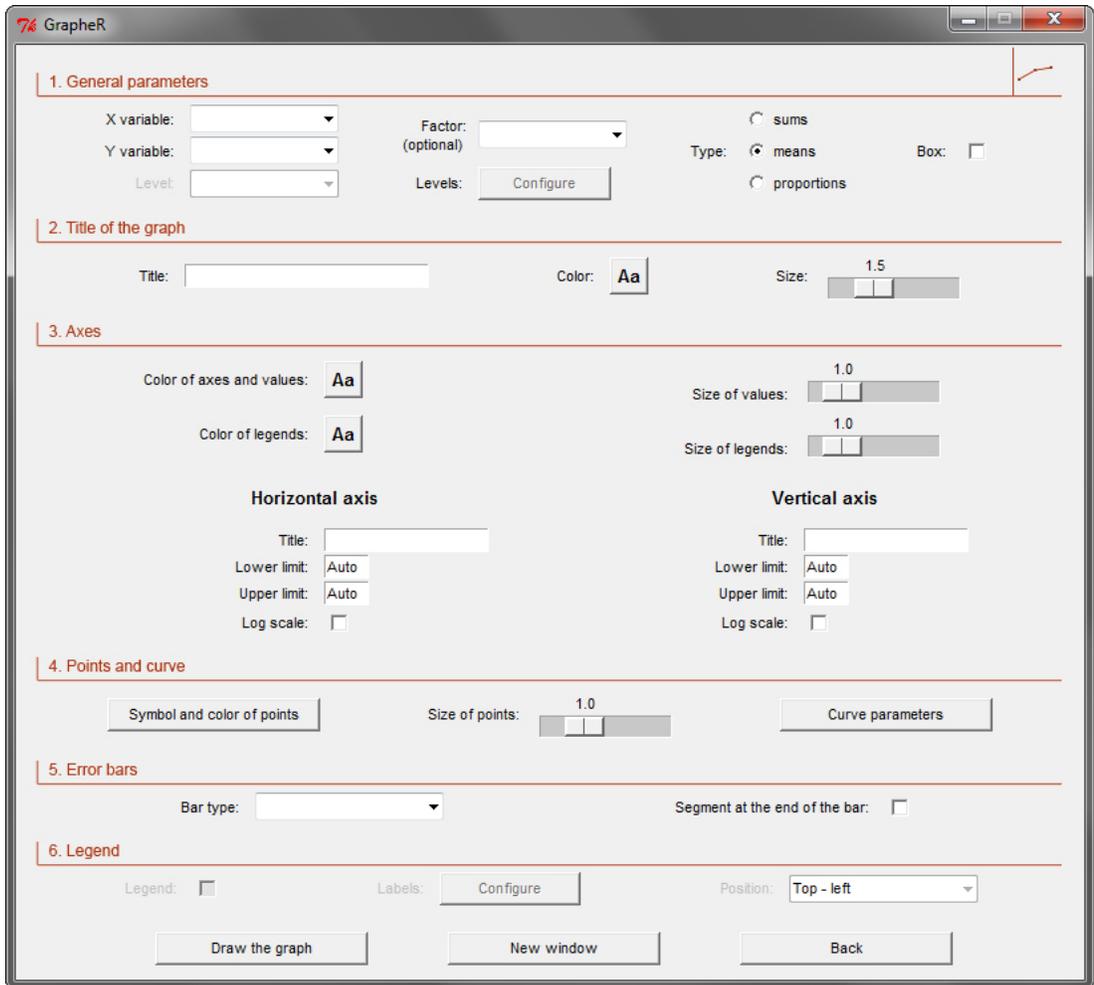


Figure 2.23: Curve

- Y variable: variable of the vertical axis, can be numeric or factorial. If this variable is a factor, only one level can be represented
- factor: optional, several curves can be drawn corresponding to values of variables X and Y for each chosen level of the factor. The levels to be represented can be selected by clicking on “Configure” (the choice window looks like the Figure 2.12)
- type of graph: if the Y variable is numeric, points of the curve can represent sums or means of values of this variable by value of the X variable (and eventually of the factor). If the Y variable is a factor, points of the curve represent the proportion of the chosen level of Y variable by value of the X variable (and eventually of the factor). No bar can be drawn if the “sums” type is chosen
- box: a box is drawn around the graph, passing by the axes
- title of the graph:
 - title: optional, the graph will not have a title if this box is empty
 - color
 - size
- axes:
 - color of axes, size of axis values, color of legends (name of axes) and size of legends: these four parameters are the same for horizontal and vertical axes
 - title of each axis
 - lower and upper limits of axes: the value of these parameters can be changed (for decimal numbers, the decimal separator is a period) or left on “Auto”
 - logarithmic scales of axes
- points and curve:
 - symbol and color of points: a different symbol and color of points can be attributed by level of the factor (hence by curve) (Figure 2.24). If the factor is not defined only one symbol and one color have to be chosen (Figure 2.25)
 - size of points: whatever the number of curves, all points have the same size
 - parameters of the curve: it is possible to customize the type of curve to represent (points only, curve only, points and curve or sticks), the type of line (full, large dashed or fine dashed) and the width of the line (Figure 2.26). If the factor is defined, these parameters can be different by level of the factor (hence by curve) (Figure 2.27)

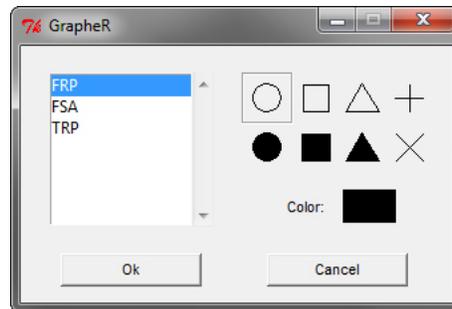


Figure 2.24: Several symbols of points

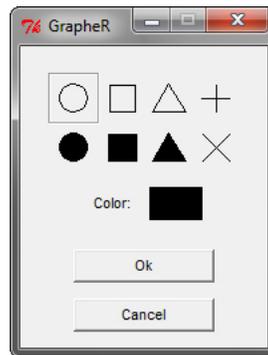


Figure 2.25: One symbol of points

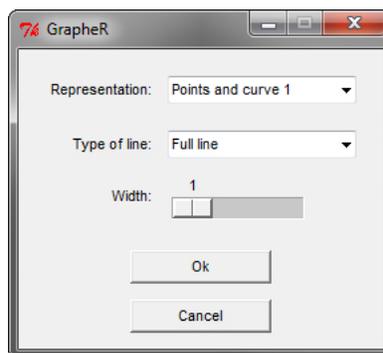


Figure 2.26: One type of curve

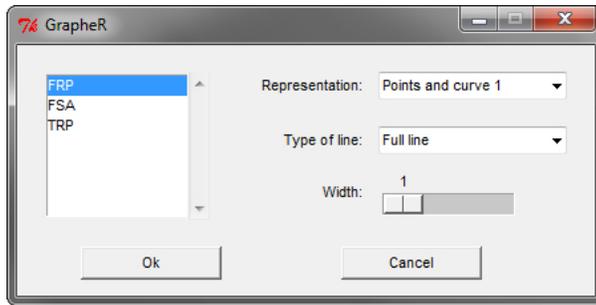


Figure 2.27: Several types of curve

- error bars:
 - type of bars: if the points of the curve represent means, error bars can be standard deviations, standard errors or confidence intervals. If the points of the curve represent proportions, error bars can only be confidence intervals. The color of the bars is the same as the corresponding curve
 - segment at the end of the bars: if this option is not chosen error bars are represented only by a vertical segment
- legend:
 - representation : a legend can be added to the graph only if the factor is defined
 - labels: by default these are levels of the factor, but it is possible to change them (which does not modify the names of the factor levels). It is also possible to add a title to the legend (the configuration window looks like the Figure 2.18)
 - position: position of the legend on the graph. If the legend overlays on curves think to modify limits of the vertical axis

If all parameters are left on default value (except the variables X and Y), the graph looks like the Figure 2.28

2.3.6 Scatter plot

The window contains five frames corresponding to five parameter categories (Figure 2.29):

- general parameters:
 - X variable: variable of the horizontal axis, has to be numeric
 - Y variable: variable of the vertical axis, has to be numeric

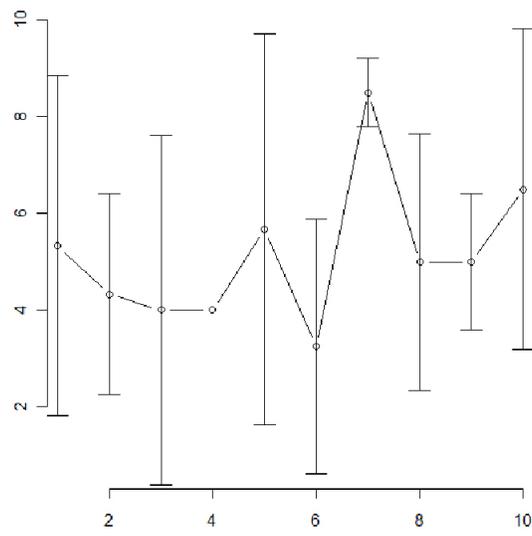


Figure 2.28: Default curve

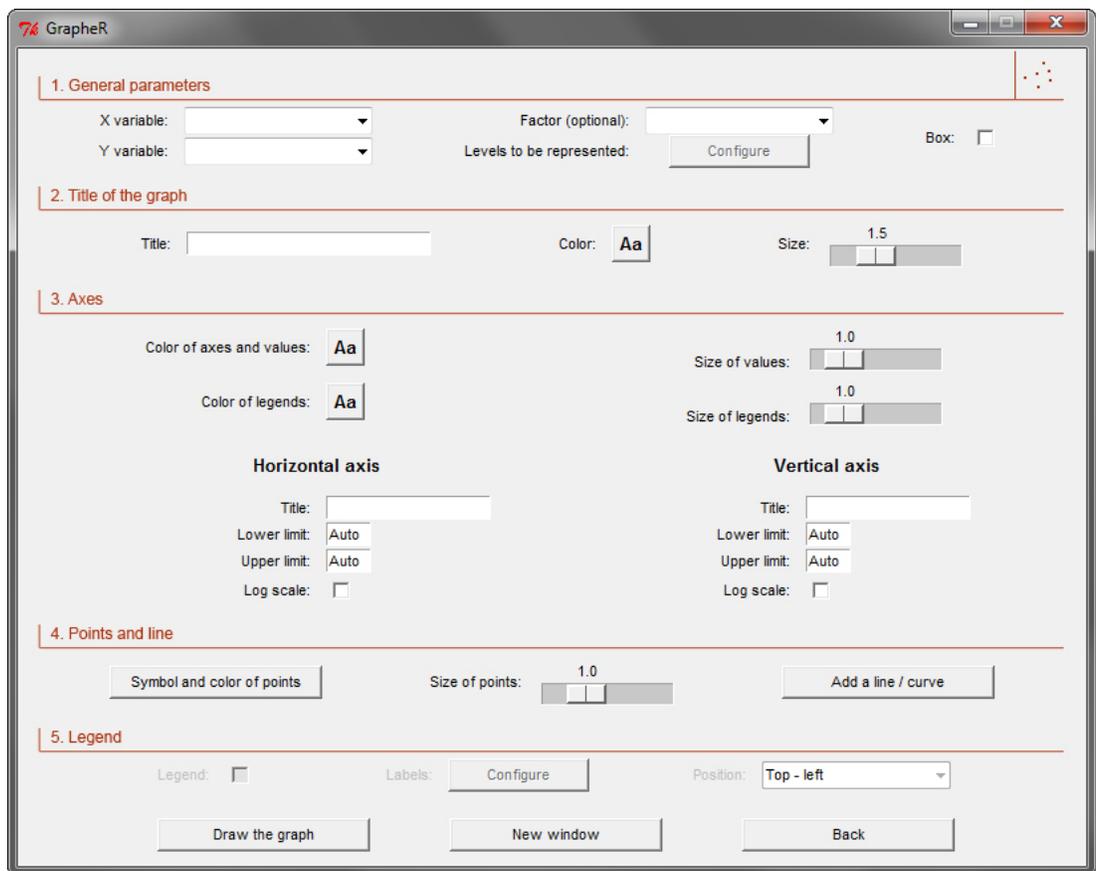


Figure 2.29: Scatter plot

- factor: optional, several plots can be drawn corresponding to values of the X and Y variables for each chosen level of the factor. The levels to be represented can be selected by clicking on “Configure” (the choice window looks like the Figure 2.12)
- box: a box is drawn around the graph, passing by the axes
- title of the graph:
 - title: optional, the graph will not have a title if this box is empty
 - color
 - size
- axes:
 - color of axes, size of axis values, color of legends (name of axes) and size of legends: these four parameters are the same for horizontal and vertical axes
 - title of each axis
 - lower and upper limits of axes: the value of these parameters can be changed (for decimal numbers, the decimal separator is a period) or left on “Auto”
 - logarithmic scales of axes
- points and line:
 - symbol and color of points: a different symbol and color of points can be attributed by level of the factor (hence by plot) (the configuration window looks like the Figure 2.24). If the factor is not defined only one symbol and one color have to be chosen (the configuration window looks like the Figure 2.25)
 - size of points: whatever the number of plots, all points have the same size
 - add a line / curve: a straight line or a curve can be drawn on the plot. The different types of lines available are:
 - * horizontal
 - * vertical
 - * of known equation (first degree)
 - * of linear least squares regression
 - * of linear least rectangles regression
 - * of quadratic regression
 - * tendency curve

The type of line (full, large dashed or fine dashed) and its width can be defined. The color of the line is the same as the points of the graph (Figure 2.30). If the factor is defined, a line can be drawn by level of the factor (hence by plot) (Figure 2.31)

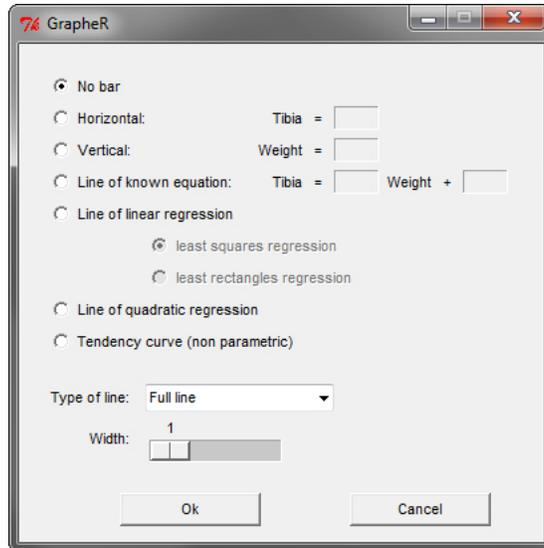


Figure 2.30: Add a line to the scatter of data points

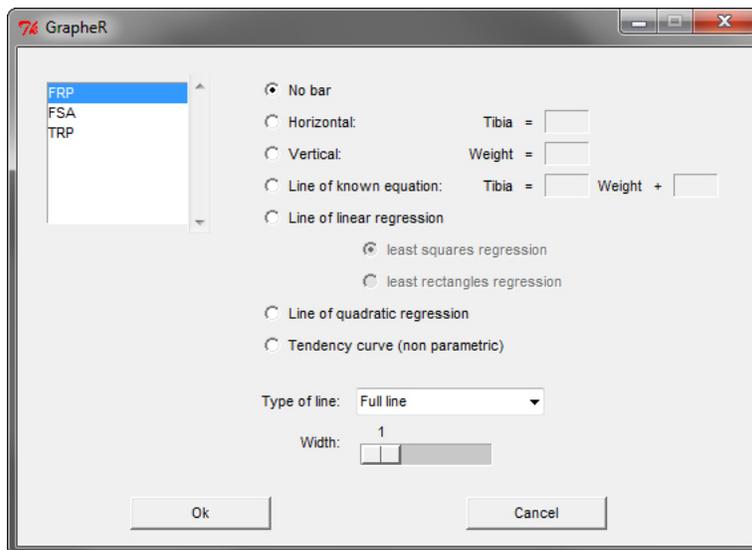


Figure 2.31: Add several lines to the scatter of data points

- legend:
 - representation : a legend can be added to the graph only if the factor is defined
 - labels: by default these are levels of the factor, but it is possible to change them (which does not modify the name of the factor levels). It is also possible to add a title to the legend (the configuration window looks like the Figure 2.18)
 - position: position of the legend on the graph. If the legend overlays on curves think to modify limits of the vertical axis

If all parameters are left on default values (except the X and Y variables), the graph looks like Figure 2.32

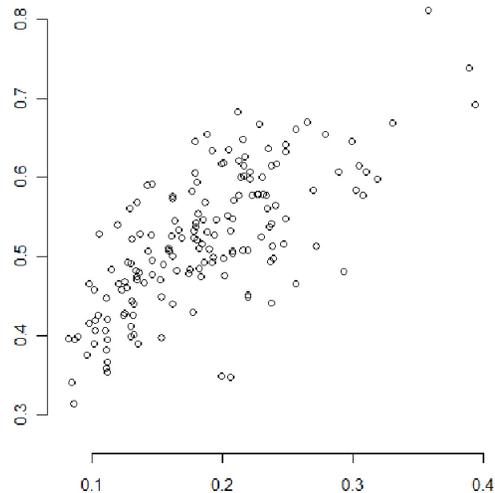


Figure 2.32: Default scatter plot

2.4 Dealing with graphics devices

It is possible to draw graphs in different windows, and/or to draw several graphs in the same one. To do this click on the “New window” button, at the bottom of each of the six parameter windows. The dialog box which opens allows to

define how many graphs have to be drawn in the device to be created, and the background color of this device (Figure 2.33).

It is possible to draw up to 16 graphs in the same device, shared between four rows and four columns. However, how big is the number of graphs to be drawn, how small is the space allocated to each.

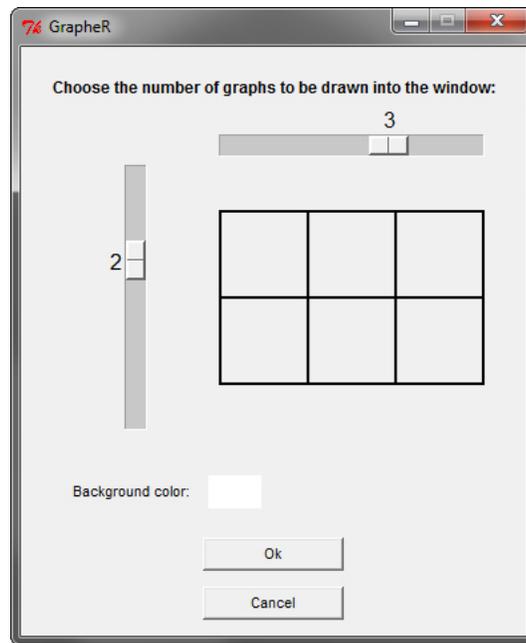


Figure 2.33: Creation of a new window

2.5 Toolbar

Once a graph is drawn, a toolbar automatically opens. Elements can be added to the graph, not all being available depending on the type of graph chosen (Figure 2.34). Elements that can be added are:

- a vertical line (all graph types except pie chart)
- a horizontal line (all graph types except pie chart)
- any straight line (only on curves and scatter plots)
- a theoretical distribution curve (only on histograms drawn in densities)
- text (all graph types)
- *p-values* (only on bar plots)

The toolbar also enable to save drawn graph(s).



Figure 2.34: Toolbar

2.5.1 Add a vertical line

It is possible to manually enter the abscissa line, or click on the graph to define it using the “Select with the mouse” button (Figure 2.35). An abscissa selected with the mouse can always be manually modified.

The type of line (full, large dashed or fine dashed), its width and its color can be defined.

Click on “Draw” to draw the line. Once all lines are drawn click on “Close”.

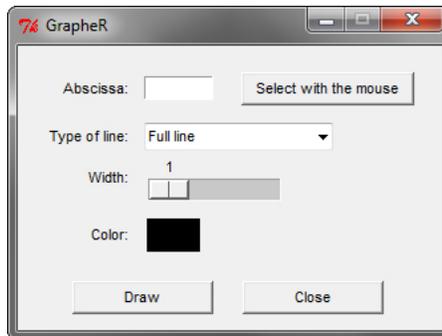


Figure 2.35: Add a vertical line

2.5.2 Add a horizontal line

It is possible to manually enter the ordinate line, or click on the graph to define it using the “Select with the mouse” button (Figure 2.36). An ordinate selected with the mouse can always be manually modified.

The type of line (full, large dashed or fine dashed), its width and its color can be defined.

Click on “Draw” to draw the line. Once all lines are drawn click on “Close”.

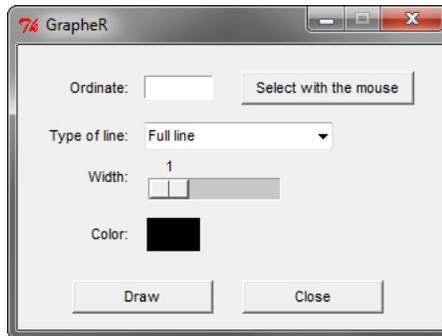


Figure 2.36: Add a horizontal line

2.5.3 Add any line

It is possible to manually enter the equation line (first degree only), or click on the graph to define it using the “Select 2 points with the mouse” button (Figure 2.37). The line will pass through these two points. The equation defined by this way can always be manually modified.

The type of line (full, large dashed or fine dashed), its width and its color can be defined.

Click on “Draw” to draw the line. Once all lines are drawn click on “Close”.

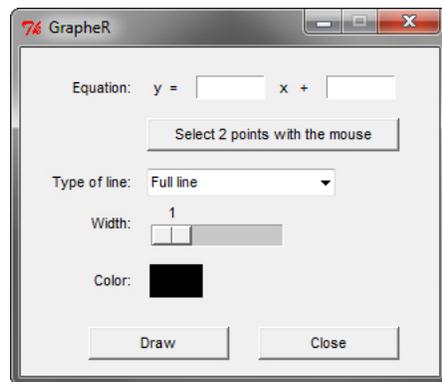


Figure 2.37: Add any line

2.5.4 Add a theoretical distribution curve

On histograms drawn in densities, it is possible to add a theoretical distribution curve with known parameters (Figure 2.38). The different types of distribution available are:

- discrete laws of probability:
 - binomial law
 - Poisson law
 - negative binomial law
 - geometric law
 - hypergeometric law
- continuous laws of probability:
 - normal law
 - gamma law
 - exponential law
 - χ^2 law
 - Fisher-Snedecor law
 - Student’s law
- other distributions of probability:
 - Mann-Whitney distribution
 - Wilcoxon sign rank distribution

The type of line (full, large dashed or fine dashed), its width and its color can be defined.

Click on “Draw” to draw the curve. Once all curves are drawn click on “Close”.

2.5.5 Add text

It is possible to manually enter the text coordinates, or click on the graph to define them using the “Select with the mouse” button (Figure 2.39). Coordinates defined by this way can always be manually modified. Text will be centered on these coordinates.

The size and the color of the text can be defined.

Click on “Draw” to draw the line. Once all lines are drawn click on “Close”.

2.5.6 Add *p-values*

Two types of representation are available (Figure 2.40). In both cases the text, its size and its color can be defined. In both cases too, these parameters have to be entered before selecting bars to be compared. The way of selecting these bars is different depending on the type of representation:

- for the first type (left), click on “Select” then on the two bars to be compared

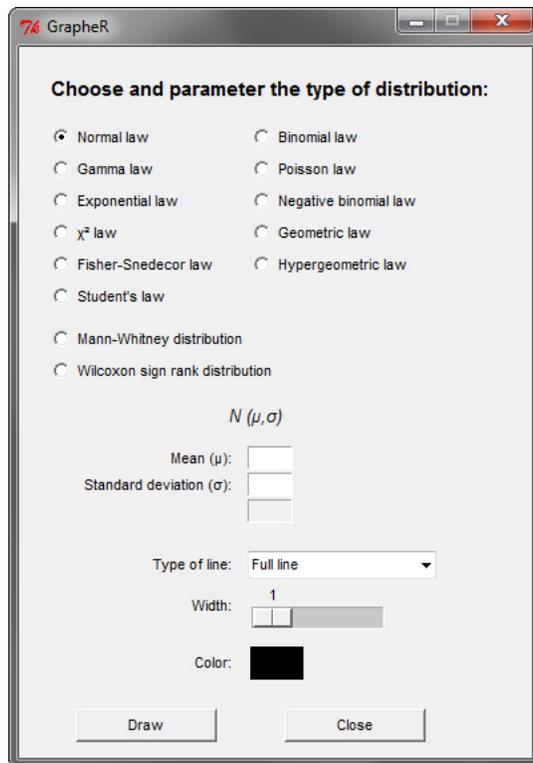


Figure 2.38: Add a theoretical distribution curve

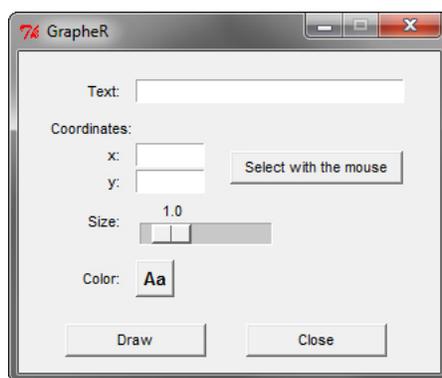


Figure 2.39: Add text

- for the second type (right), click on “Select” then on the two extreme bars (at left and at right) of the group to be compared

Once all the p-values are added click on “Close”.

Note: this tool was designed for graphs drawn without logarithmic scale on ordinate. It is otherwise not optimal.

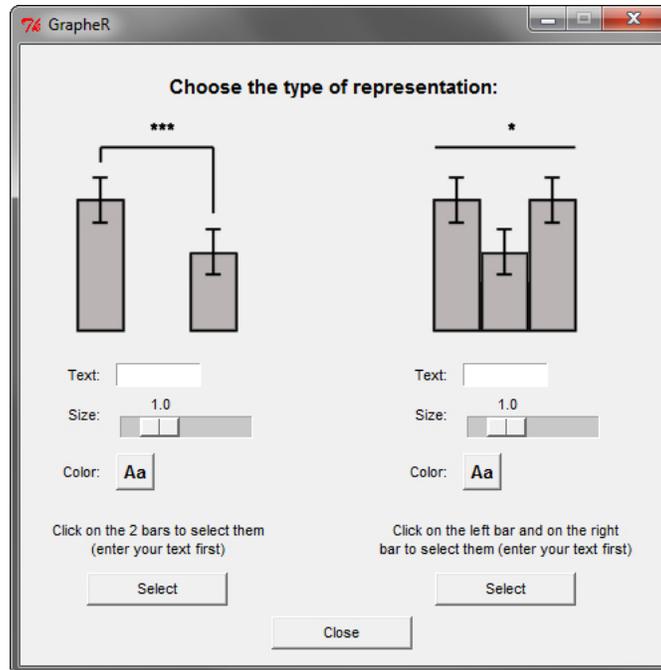


Figure 2.40: Add *p-values*

2.5.7 Save a graph

Choose the graphics device to be saved with the drop-down menu (the window number corresponds to the number automatically allocated by R, under the guise “R Graphics: Device *number*”). Choose the extension of the file to be created and its width (in pixels; height is automatically calculated according to the width) (Figure 2.41). Once all parameters are defined, click on “Save”.

Note: only the .jpg format is recognized, not .jpeg.

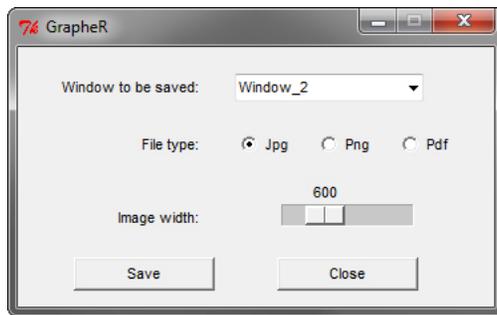


Figure 2.41: Save a graph

Chapter 3

Add a new language to GrapheR

Extend GrapheR in a new language is very easy. You only have two things to do:

- get the file named `Language.en.csv` in the `lang` directory of the package. All words written in the interface come from this file. Just translate scrupulously each line (be careful of the spaces before or after words, they are not randomly placed) and save the file in `.csv` extension
- translate this user manual. You can just write a classic Microsoft Word or Open Office Writer file, or if you are `LATEX` familiar ask me for the original `.tex` file

After doing your translation, send me the two files and I will finish the job ;)