

Chapter 945

Histograms

Introduction

This section describes the options that are available for the appearance of a histogram. A set of all these options can be stored as a template file which can be retrieved later.

Histogram Tab

These options specify the characteristics of the histogram.

Format

These options specify the format of the histogram outline and interior.

Interior Lines

This is the smallest value of the axis. If left blank, a nice value is determined from the data.

Bins

These options specify the appearance of the histogram's bins (divisions or classes). This includes setting the number of bins or the width of the bins.

Auto

This option indicates that the number of bins is to be determined from the data. Click to set the number of bins (intervals or classes) to *Auto* (meaning automatic).

The program finds a number of bins that will:

- 1) result in nice numbers for the bin boundaries, and
- 2) display the general shape of the data distribution.

Note that as the number of data values increases, so does the number of bins.

When a discrete dataset (one with only a few unique values) is detected, the number of bins will be set to the number of unique values, regardless of the number of rows.

Treat as Discrete if the Number of Uniques

This option is only used when the Auto option is selected for the number of bins.

This option sets the maximum number of unique values that a dataset can have and still be treated as discrete. When a dataset is discrete, the bins are set to surround the discrete values. Note that only (nearly) consecutive integers are considered. Datasets with fractional discrete values receive no special treatment.

Histograms

Number

Click to set the number of bins to *Number*. This option forces the program to use the number of bins specified to the right.

If a minimum and/or maximum is specified, they will be used. If not, an appropriate minimum and maximum values are determined from the data.

Min

Optional minimum for the bin boundaries. If this value is blank, a nice minimum based on the data is used.

This minimum is used to construct the bins, but it does not necessarily set the minimum for the axis or the axis lines. Those are set elsewhere.

Max

Optional maximum for the bins boundaries. If this value is blank, a nice maximum based on the data is used.

This maximum is used to construct the bins, but it does not necessarily set the maximum for the axis or the axis lines. Those are set elsewhere.

Bin Width

Click to set the number of bins to *Bin Width*. This option sets the space between bin boundaries to the amount entered to the right (or a power of ten multiple of the value). Beginning at the minimum, bins are added until the maximum of the data is reached.

For example, suppose the minimum is set to 20 and the maximum is set to 44. Further suppose that the interval is 10. Bins would be $10 \leq X < 20$, $20 \leq X < 30$, $30 \leq X < 40$, and $40 \leq X \leq 50$.

Note that the boundaries of the bins are closed on the left and open on the right, except for the last bin which is closed on both sides.

Bin Width Scaling

This option determines whether the interval is scaled so that between 4 and 20 ticks occur.

- **Scale interval by power of 10**

The interval is scaled (multiplied) by an appropriate power of 10.

- **Unscaled - Use this interval**

Use this interval without scaling.

Bin Width Min

Optional minimum for the bins. If this value is blank, a nice minimum based on the data is determined. This minimum is used to construct the bins, but it does not necessarily set the minimum for the axis or the axis lines. Those are set elsewhere.

Overlays Tab

These options specify objects that can enhance or replace the histogram. These include the density trace, frequency polygon, and normal density.

Density Trace

The density trace is a smoothed version of the histogram. It is described elsewhere.

Interval Width as a Percent of Range

The interval width controls the smoothness of the density trace. The wider the interval, the smoother the curve.

Since the entering the actual interval desired would require a different entry for each case, a relative interval width is specified as the percentage of the overall range of the data.

The appearance of the density trace depends a great deal on your entry here. Remember that as you increase this width, more data are processed at each calculation point which results in more and more time needed to complete the plot.

Number of Calculation Points

The number of values along the X axis at which the Y value is calculated. As you increase the number, the line gets more and more precise. Usually, 50 calculation points will suffice. Sometimes you will need more to show an accurate depiction of the relationship.

Frequency Polygon

The frequency polygon replaces the histogram bars with a line connecting the top midpoint of each bar.

Normal Density

A normal density can be displayed over the plot. This object depends solely on the mean and standard deviation of the data.

Number of Calculation Points

This is the number of values along the X axis at which the Y value is calculated. As you increase the number, the line gets more and more precise. Usually, 50 calculation points will suffice. Sometimes you will need more to show an accurate depiction of the relationship.

X and Y Axis Tabs

These options specify the characteristics of the lines, labels, and tick marks along the X and Y axes.

Boundaries

These boundaries set the minimum and maximum of the plot. Data values outside these boundaries are ignored. When set, these boundaries override any other boundaries.

Minimum

This is the smallest value of the axis. If left blank, a nice value is determined from the data.

Maximum

This is the largest value of the axis. If left blank, a nice value is determined from the data.

Corners

Click to bring up a window that lets you extend the axis a little further. This is usually done so that the X and Y tick mark labels are not too close together or so that data points do not fall directly on the axis line. The value is the number of pixels the axis end is shifted.

Note that this does not change scale of the graph. It only shifts the location of the axis.

Scale (Y axis only)

Scale

Indicate whether the scale of the Y axis is in terms of percents, proportions, or frequencies (counts).

Cumulative Scale

Checking this option displays the cumulative frequencies (or percentages) instead of the individual frequencies (or percentages).

The summing occurs as you move from left to right.

Axis Labels

These options specify the characteristics of the axis label.

Top, Bottom or Left, Right

These options indicate whether the label is to be shown along the top and/or bottom (left and/or right) axes.

Histograms

Text

Enter the label text here.

Text replacement codes can be used to create labels that will be changed at runtime. These codes are replaced by appropriate text (such as variable names) when a plot is generated. The codes are

{X} is replaced by the name of the X axis variable.

{Y} is replaced by the name of the Y axis variable.

Font

The font size, color, and style of the label may be modified by pressing the font button (A) to the right of the text.

Layout

The exact position of the text is set by pressing the layout button (Layout) to the right of the font button.

The text is printed inside of an imaginary box of appropriate size. The Layout window sets the alignment, the margins, and the shift of the text.

The **alignment** designates whether the text is centered, right-justified, or left-justified within the box.

The **margin** indicates the number of pixels displayed above and below the text.

The **shift** indicates the number of pixels that the text is shifted in the X and/or Y directions. This is used to make small adjustments to the text placement.

Lines

These options specify the characteristics of the axis lines. Note that the axis lines can be specified separately from the axis boundaries.

Top, Bottom or Left, Right

These options indicate whether a line is displayed for the top and/or bottom (left and/or right) axes.

Line Format Button

Pressing this button displays the Axis Line Format window that sets the format (color, width, and pattern) of the axis line. Click the left side of the button to load the Line Format window. Click the right side of the button to select a format from various preset line formats.

Crosses X (or Y) Axis at

Specifies the point at which this axis crosses the opposite axis.

- **{blank}**

The axis is placed at the edge of the plot region.

- **Numeric Value**

The axis is placed at this value of the opposite axis. For example, if the X axis has both positive and negative values, you might want to enter '0' here for the Y axis so that it crosses the X axis at 0.

Histograms

Axis Line Min and Max

These values are used to set a specific minimum and maximum for the axis line. When entered, the axis line is only visible in that region. Note that other axis objects such as the tick marks and tick reference labels might extend beyond the end of the visible axis line.

Possible choices are

- **{blank}**
The axis line is extended to the edge of the axis.
- **Value**
The axis line is extended to the specified value.
- **Data**
If *Data* is entered, the axis line is extended to the exact minimum (or maximum) of the data.

Major Ticks

These options specify the characteristics of the major (primary) ticks, including their placement and format.

Top Marks, Bottom Marks or Left Marks, Right Marks

These options indicate whether major tick marks are displayed along the top and/or bottom (left and/or right) axes.

Format

Clicking this button allows the color, size, and position of the tick marks to be set.

Tick Number and Spacing

Clicking this button allows the number and spacing of the ticks to be set.

Top Labels, Bottom Labels or Left Labels, Right Labels

These options indicate whether the tick reference labels are displayed at the corresponding ticks.

Font

Pressing the font button (A) lets the font size, color, and style of the tick reference label to be modified.

Layout

The exact position of the tick reference labels are set by pressing the layout button (Layout). The labels are displayed inside an imaginary box. The Layout window sets the alignment, rotation angle, margins, and shift of the reference labels.

The **alignment** designates whether the reference labels are centered, right-justified, or left-justified within the box.

The **rotation angle** designates whether the reference labels are rotated through a specified angle.

The **margin** indicates the number of pixels displayed above and below the reference labels.

Histograms

The **shift** indicates the number of pixels that the reference labels are shifted in the X and/or Y directions. This is used to make small adjustments to the text placement.

Decimals

Specify the number of decimal places displayed in the tick label.

- **Auto**

If *Auto* is selected, the number of decimal places is determined from the data values.

Minor Ticks

These options specify the characteristics of the minor ticks, including their placement and format. Note that no reference labels are available for minor ticks.

Top Marks, Bottom Marks or Left Marks, Right Marks

These options indicate whether minor tick marks are displayed along the top and/or bottom (left and/or right) axes.

Format

Clicking this button allows the color, size, and position of the minor tick marks to be set.

Tick Number and Spacing

Clicking this button allows the number and spacing of the minor ticks to be set.

Grid Line Tab

These options specify the characteristics of various grid lines the X and Y axes.

From X Axis

These grid lines extend vertically from the X axis.

At Major Ticks

Check this option to extend grid lines vertically from the major ticks along the X axis. The check box is followed by a Line Format button that sets the characteristics of the line.

At Minor Ticks

Check this option to extend grid lines vertically from the minor ticks along the X axis. The check box is followed by a Line Format button that sets the characteristics of the line.

Histograms

At These Values

Check this option to extend grid lines vertically from the minor ticks at the values in the list. The check box is followed by a text box in which the values are entered. The text box is followed by a Line Format button that sets the characteristics of the line.

- **List**

You can enter a list separated by blanks such as

10 20 25 30 40

- **Range**

You can enter a range of values using the *xxx TO yyy BY zzz* syntax such as

10 to 100 by 10

The value list is followed by a Line Format button that sets the characteristics of the line.

From Y Axis

These grid lines extend horizontally from the Y axis.

At Major Ticks

Check this option to extend grid lines horizontally from the major ticks along the Y axis. The check box is followed by a Line Format button that sets the characteristics of the line.

At Minor Ticks

Check this option to extend grid lines horizontally from the minor ticks along the Y axis. The check box is followed by a Line Format button that sets the characteristics of the line.

At These Values

Check this option to extend grid lines horizontally from the minor ticks at the values in the list. The check box is followed by a text box in which the values are entered. The text box is followed by a Line Format button that sets the characteristics of the line.

- **List**

You can enter a list separated by blanks such as

10 20 25 30 40

- **Range**

You can enter a range of values using the *xxx TO yyy BY zzz* syntax such as

10 to 100 by 10

The value list is followed by a Line Format button that sets the characteristics of the line.

Titles Tab

These options set the titles of the plot. Up to two titles may be specified on each of the four sides of the plot.

Titles along the X Axes

These options specify the characteristics of the title displayed at the top or bottom of the plot.

Top 1, Top 2, Bottom 1, or Bottom 2

These options indicate whether the title is shown along the top and/or bottom of the plot.

Text

Enter the title text here.

Text replacement codes can be used to create labels that will be changed at runtime. These codes are replaced by appropriate text (such as variable names) when a plot is generated. The codes are

{T} by the program generated plot title.

{X} is replaced by the name of the X axis variable.

{Y} is replaced by the name of the Y axis variable.

{G} by the grouping variable's name.

{Z} by the third variable's name.

{A} by the intercept.

{B} by the slope.

{R} by the R squared value.

For example, you would enter

$\{Y\} = \{A\} + (\{B\})\{X\}$

to display the linear regression equation.

Font (A)

The font size, color, and style of the title may be modified by pressing the font button (A) to the right of the text.

Layout

The exact position of the text is set by pressing the layout button (Layout) to the right of the font button.

The text is printed inside of an imaginary box of appropriate size. The Layout window sets the alignment, the margins, and the shift of the text.

The **alignment** designates whether the text is centered, right-justified, or left-justified within the box.

The **margin** indicates the number of pixels displayed above and below the text.

The **shift** indicates the number of pixels that the text is shifted in the X and/or Y directions. This is used to make small adjustments to the text placement.

Titles along the Y Axes

These options specify the characteristics of the title displayed at the left or right of the plot. The details are identical to those for the X Axes (above).

Legend Tab

These options display the characteristics and position to the plot legend.

General Tab

This tab displays options about the location, title, and appearance of the box containing the legend.

Show Legend

This check box specifies whether to display the legend.

Show Legend If

This setting indicates if the legend is displayed when there is only one group.

Margin Outside the Legend Frame

These options specify the margins around the edges of the legend.

Titles

These options specify one or two titles for the legend.

Frame and Fill

These options set the properties of the legend's frame (outline) and its fill (interior color).

Layout Tab

This tab displays options about the layout (arrangement) of the entries within the legend.

Arrangement

These options control the number of columns in the legend and the direction in which the legend entries are displayed.

Margins Inside the Legend

These options control the margins between the various portions of the legend.

Spacing

These options control the margins (space) between the entries

Legend Entry Order

These options control order in which the entries are shown.

Lines Tab

This tab displays options about the display of the overlay lines in the legend.

Overlay Line Labels

These options control the format of the text portion of the entry for each overlay line.

Overlay Line Size

These options control the format of the line portion of the entry for each overlay line.

Customize Individual Legend Entries

These options let you change the format and position of specific overlay lines in the legend.

Extra Lines Tab

This tab displays options about the display of the extra lines in the legend.

Extra Line Labels

These options control the format of the text portion of the entry for each extra line.

Extra Line Size

These options control the format of the line portion of the entry for each extra line.

Customize Individual Legend Entries

These options let you change the format and position of specific extra lines in the legend.

Extras Tab

The options on this tab let you add extra lines, symbols, and notes to the plot.

Extra Lines

This option lets you add an unlimited number of extra lines to a plot. To add a line, specify a line format and the beginning and ending points of the line.

Notes

This option lets you add an unlimited number of notes (text) to a plot. To add a phrase, specify the text, its format, and the beginning location on the plot.

Border Plots Tab

These options let you display any of several reference plots in the border of the plots. These plots include the box plot and the dot plot.

Plot Order

The order in which the plots are displayed in the margin is set by the Plot Order list.

Box Plot

A box plot can be placed in the margin of the X axis. Box plots emphasize the middle of the data.

Dot Plot

A dot plot can be placed in the margin of the X axis. Dot plots emphasize the individual characteristics of the data values.

Colors and Size Tab

These options specify a background color for the whole plot (background) and for the data region (interior). They also specify size and margins of the plot.

Background

This option specifies the background outline and fill color.

Interior

This option specifies the interior fill color. This is the color of the portion of the plot on which the data points are plotted.

Size

This option specifies the width and height (in pixels) of the interior area of the plot. The legend, border plots, tick marks, titles, and labels are added to the outside of this area. Thus, the total width and height of the plot cannot be specified.

Margin

This option a margin for each edge of the plot (in pixels).

Painting Order Tab

The options under this tab control the order in which items are placed on the plot. Items at the top of the list are placed on the plot first. They may be obscured the items that are plotted later.

Example Data Tab

Often you will need to set up the format of a histogram before the data is available. This option lets you see how your options appear on a set of random data.

You control the number of data points and the data limits of the X variable.

Pressing the New Data button generates a new set of random data.