GraphPad PRISM®

Version 4.0

Step-by-Step Examples

Histograms and Frequency Distributions[1]

This article shows how to create histograms—bar graph depictions of frequency distributions. We'll cover two alternative approaches:

- Tabulate and graph frequency data that you have already computed from your raw data, or
- Give Prism the raw data itself and let it compute and graph the frequencies automatically.

The following special techniques are included in this article:

- Reformatting histograms to increase readability, including offsetting and angling baseline labels
- Choosing to show "relative" vs "cumulative" frequencies on histograms
- Superimposing a fitted Gaussian curve onto a histogram

Histograms from Pre-Computed Frequency Data

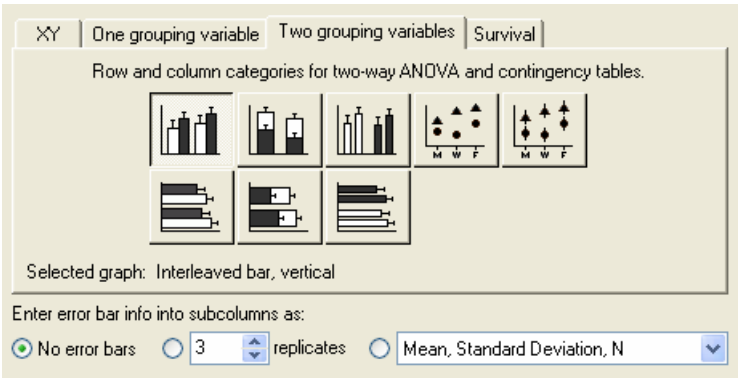
In this section, we'll discuss two methods for constructing histograms when you have already computed the frequency values and you intend to enter them manually. Histograms may be either bar graphs or XY graphs.

Bar-Graph Histogram

Suppose we've measured the heights, in inches, of 55 male adults and we record the number of values in each of nine categories, or "bins", each representing a different 2-inch range.

In the Welcome dialog, choose **Type of graph**. Select the **Two grouping variables** tab and then the first thumbnail view.

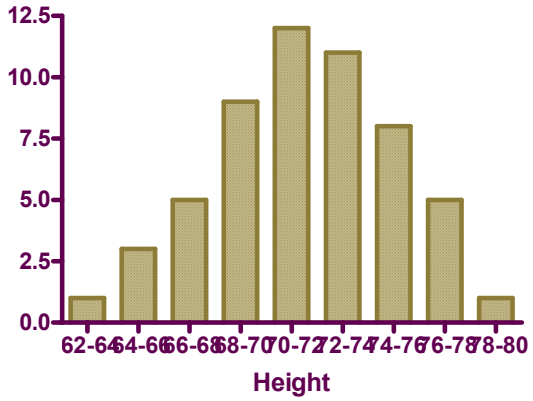
[1] Adapted from: Miller, J.R., *GraphPad Prism Version 4.0 Step-by-Step Examples*, GraphPad Software Inc., San Diego CA, 2003. *Step-by-Step Examples* is one of four manuals included with Prism 4. All are available for download as PDF files at www.graphpad.com. While the directions and figures match the Windows version of Prism 4, all examples can be applied to Apple Macintosh systems with little adaptation. We encourage you to print this article and read it at your computer, trying each step as you go. Before you start, use Prism's **View** menu to make sure that the Navigator and all optional toolbars are displayed on your computer.



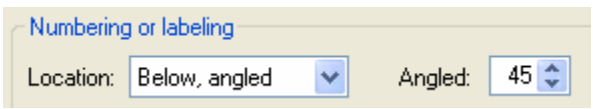
Enter the data and column headings as follows:

	X Labels	A
	Height	Frequency
	X	Y
1	62-64	1
2	64-66	3
3	66-68	5
4	68-70	9
5	70-72	12
6	72-74	11
7	74-76	8
8	76-78	5
9	78-80	1

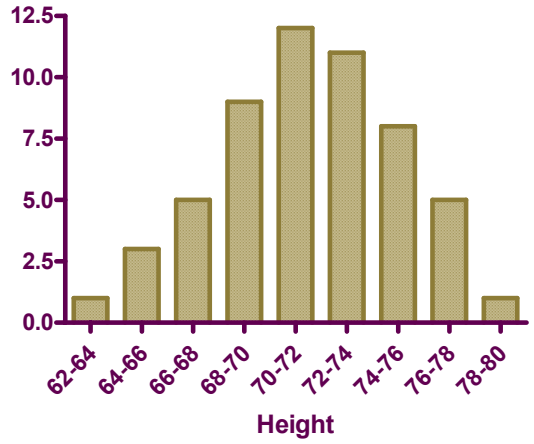
The histogram appears as soon as you click the yellow **Graphs** tab on the toolbar.



You can rotate the bar labels to avoid overlapping—double-click the baseline to open the **Format Axes** dialog, select the **X axis** tab, and adjust the **Numbering or labeling** settings as follows:



Below is the graph, with a few additional formatting changes. The baseline title “Height” has been lowered to increase readability. To do this, place your cursor just above or below the title so that it becomes a double-headed arrow, then click and drag downward.



This is a quick and easy way to make a histogram for which you have already computed the frequency values. Because the bar labels are text (even if that “text” takes the form of numbers), and since you can change the angle of the labels, you have versatility in labeling. If you don’t want angled labels but need to create a little extra clearance space between labels, select the baseline and then either drag one of the ends to make it longer or choose **Change… Selected Text…** to change the font size. You can also increase readability by omitting some of the labels at regular intervals (which wouldn’t work well here, but might work where the labels are single numbers designating bin “centers”). To omit a label, simply omit the X-column entry for that row, entering the Y value only.

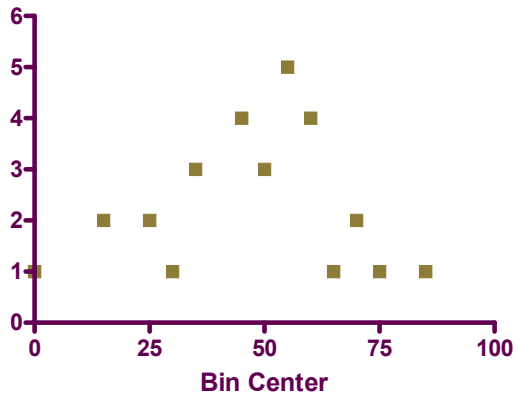
XY-Graph Histogram

Here is another way to make a histogram using pre-computed frequencies. It’s a bit more work, but it gives more versatility in labeling, allows you to use irregularly spaced bin centers if desired, and allows you to superimpose a line graph, if desired.

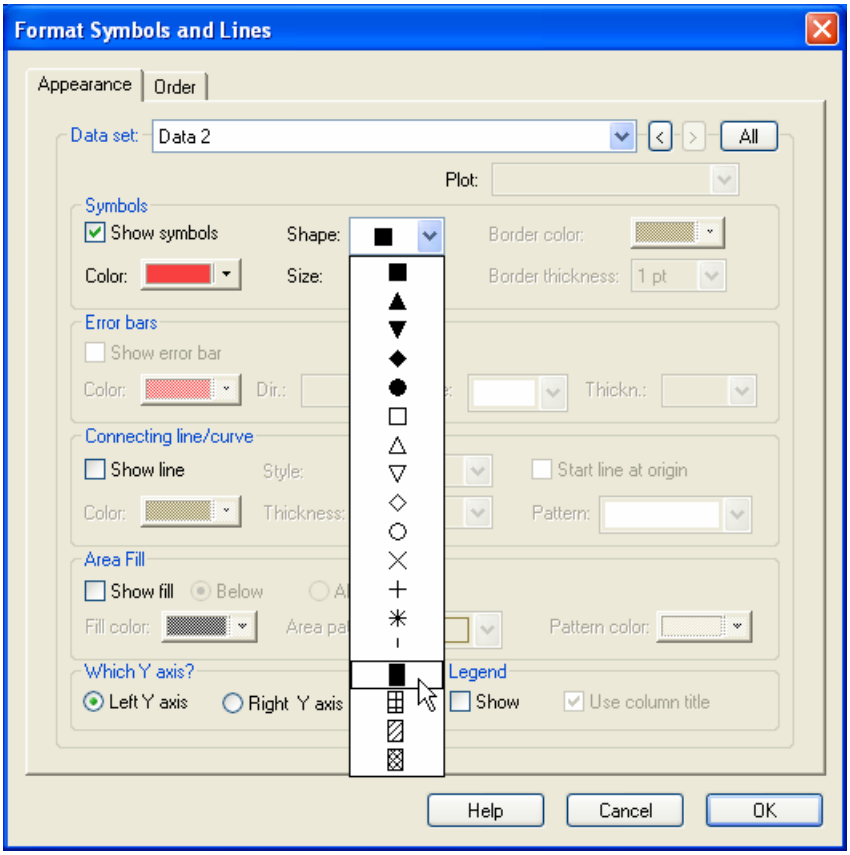
Click the yellow **Data** tab, then the **New** button. Choose **New Data Table (+Graph)**. In the **Create New Table** dialog, select **Create new table (choose X and Y format).** Format the X column for **Numbers (XY Graph)** and the Y columns for **A single column of values**. Enter this data:

	X Values	A
	Bin Center	Frequency
	X	Y
1	0	1
2	15	2
3	25	2
4	30	1
5	35	3
6	45	4
7	50	3
8	55	5
9	60	4
10	65	1
11	70	2
12	75	1
13	85	1

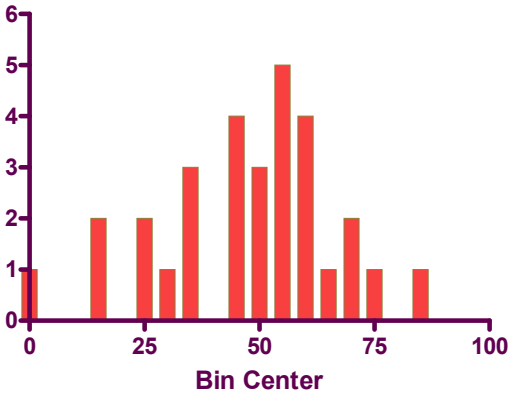
Click the yellow **Graphs** tab on the toolbar.



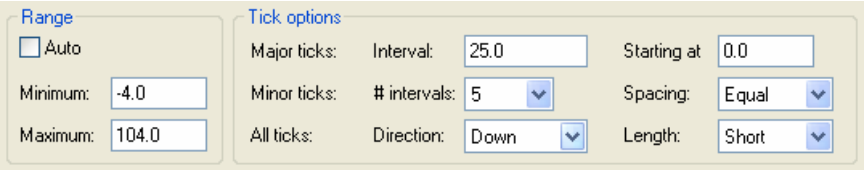
You could immediately change this XY plot to a bar graph, but instead we'll just change the point symbols to "spikes", which produces a pseudo bar graph. Since the graph is still an XY plot, we'll have the X-axis formatting latitude that comes with that graph type. Click **Change.. Symbols & Lines**. Under the **Appearance** tab, change the symbol shape to one of the last four choices to produce "spikes" (bars).



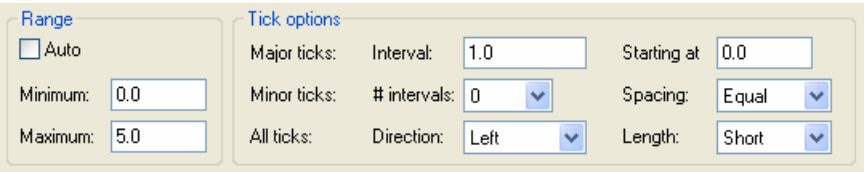
Increase the symbol size, which will widen the "bars", if desired.



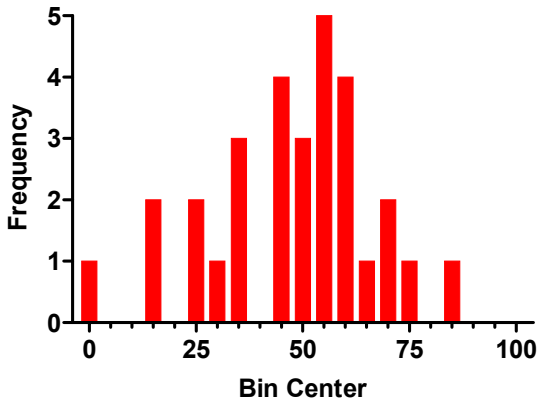
This results in easier-to-read (not crowded) bin center labeling, which you can adjust further by double-clicking the baseline, verifying that the **X axis** tab is selected, and making changes to the **Range** and **Tick options** settings (don't forget to deselect the **Auto** checkbox). The following settings will add an "offset" between the Y axis and the first bar, while keeping the bin labels centered. Note that we are also adding minor ticks by selecting **5** intervals from the **Minor ticks** drop-down list:



Let's also make the Y axis labeling more useful. Double-click on the Y axis. Make sure that in the resulting **Format Axes** dialog, the **Y axis** tab is selected. Change the **Range** and **Tick options** settings as follows:

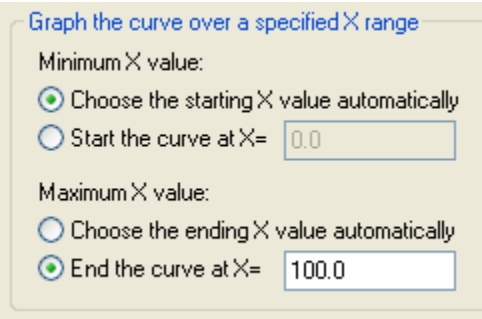


In the graph below, we've also edited the Y-axis title.

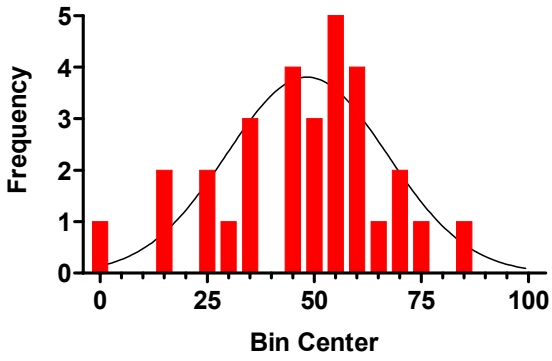


Adding an Ideal Gaussian Distribution Line

To superimpose a normal-distribution line fitted to your data, click **Analyze... Curves & regression... Nonlinear regression (curve fit)** . In the **Parameters: Nonlinear Regression (Curve Fit)** dialog, choose **Gaussian distribution** from the list of **Classic equations**. The curve, by default, will only be plotted over the X range of the input data, so let's extend the rightward extent of the line. Select the **Range** tab and make the following settings:



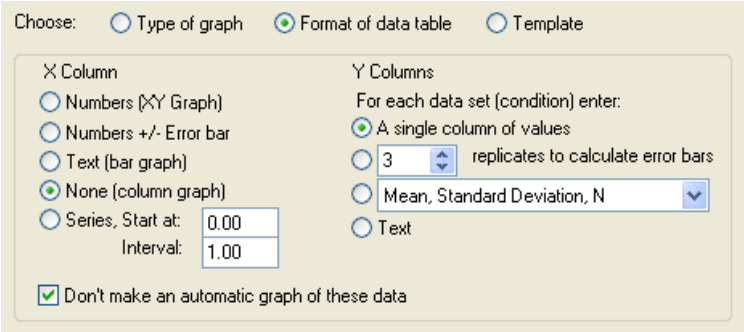
Click **OK** to fit the curve.



Histograms from Raw Data, using Automatic Frequency Computation

Prism can provide a frequency distribution from raw data and draw the histogram automatically.

In the Welcome dialog, indicate that you will specify the **Format of data table** directly.



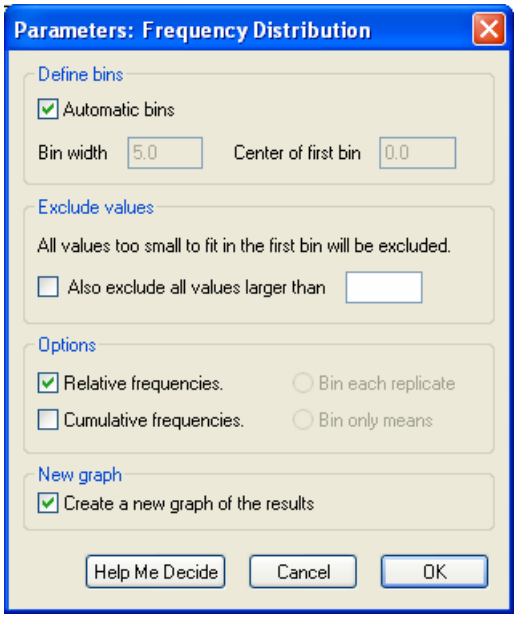
In the resulting data table, enter the raw data values into a *single* Y column.

	A
	Data Set-A
	Y
1	30
2	17
3	22
4	19
5	23
6	17
7	22
8	8

Following are all the values for this example—but be sure to *put all 50 numbers into one column*, as shown above.

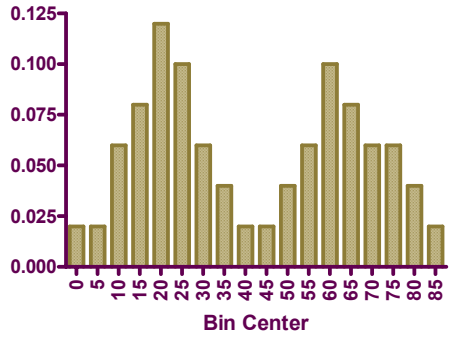
30	17	22	19	23	17	22	8	19	49
37	28	61	18	60	46	74	69	4	61
23	71	66	24	42	78	64	60	70	63
83	74	2	31	57	20	80	23	15	57
16	24	9	58	67	10	54	35	52	76

Click **Analyze**, and choose **Frequency distribution** from the **Statistical analyses** list. The **Parameters: Frequency Distribution** dialog box appears.



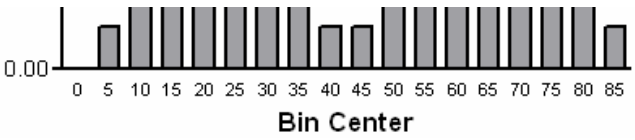
In the **Define bins** section, you can choose to define the way Prism will sort the data into bins or you can allow Prism to do it automatically. Under **Exclude values**, note that Prism excludes from the analysis all values below those fitting the first bin. You can additionally choose to exclude all values larger than a value that you designate. Under **Options**, you have the option to choose **Relative frequencies** to have Prism plot the fraction of the total

number of values contained in each bin, rather than the absolute number of values themselves. You also have the option to display a cumulative distribution, whereby each bar shows the number or fraction of value falling *in or below* that bin. If you make the settings shown in the figure above (don't forget **Create a new graph of the results**), the following line graph is produced (click the yellow **Graphs** tab):

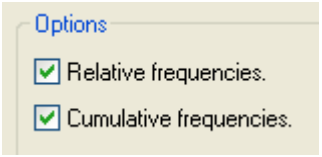


As we pointed out earlier, you can drag the end of the baseline to create more room for the labels,

and you can reduce the size of the labels (select baseline and choose **Change... Selected Text...**).



Finally, if we had chosen to display **Cumulative** relative frequencies (you can back up and try these options by clicking **Change... Analysis Parameters...**),



the graph would resemble the one below. In the illustration below, we made some additional formatting changes.