

# Violin Plot Statlet



Revised: 10/9/2017



Summary .....	1
Data Input.....	3
Statlet .....	4
Calculations.....	5

## Summary

The *Violin Plot Statlet* displays data for a single quantitative sample using a combination of a box-and-whisker plot and a nonparametric density estimator. It is very useful for visualizing the shape of the probability density function for the population from which the data came.

**Sample StatFolio:** *violinplot.sgp*

## Sample Data

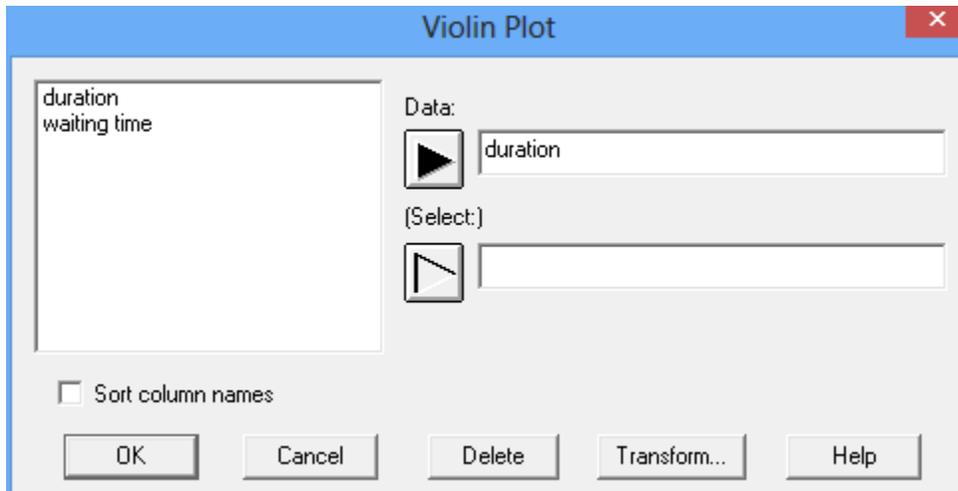
The file *old faithful.sgd* contains measurements of the duration of consecutive eruptions of the Old Faithful Geyser in Yellowstone National Park. The first several rows of that file are shown below:

duration	waiting time	subsequent duration
3.600	79	1.800
1.800	54	3.333
3.333	74	2.283
2.283	62	4.533
4.533	85	2.883
2.883	55	4.700
4.700	88	3.600
3.600	85	1.950
1.950	51	4.350
4.350	85	1.833
...	...	...

*Duration* measures the length of each eruption in minutes. *Waiting time* is the number of minutes until the next eruption occurred. *Subsequent duration* is the length of the following eruption in minutes.

## Data Input

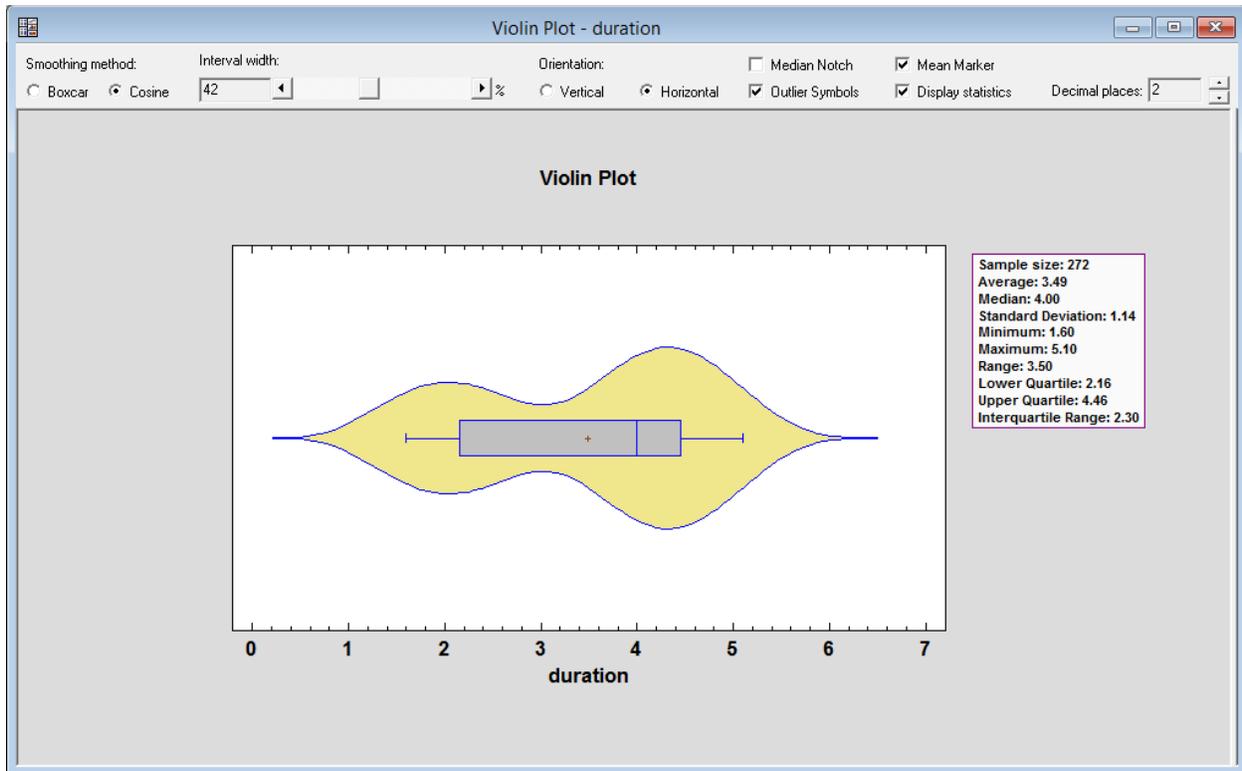
The data input dialog box requests the name of the column containing the data values to be plotted:



- **Data:** name of a numeric column containing the data to be plotted.
- **Select:** optional subset selection.

## Statlet

The output of this procedure is displayed in a dynamic Statlet window:



The bimodal nature of the data is clearly visible.

The first 2 controls on the toolbar determine how the density estimate is calculated. Basically, the estimate is created by counting the number of observations that fall within a window of fixed width moved across the range of the data.

- **Smoothing method:** the desired weighting function. The boxcar function weights all values within the window equally. The cosine function gives decreasing weight to observations further from the center of the window. The default selection is determined by the setting on the *EDA* tab of the *Preferences* dialog box accessible from the *Edit* menu.
- **Interval Width:** the width of the window  $h$  within which observations affect the estimated density, as a percentage of the range covered by the x-axis.  $h = 60\%$  is not unreasonable for a small sample but may not give as much detail as a smaller value in larger samples.

The next several controls specify options for the box-and-whisker plot:

- **Orientation:** the orientation of the plot, corresponding to the direction of the whiskers.

- **Median Notch:** if selected, a notch will be added to the plot showing an approximate  $100(1-\alpha)\%$  confidence interval for the median at the default system confidence level (set on the *General* tab of the *Preferences* dialog box on the *Edit* menu).
- **Outlier Symbols:** if selected, indicates the location of outside points.
- **Mean Marker:** if selected, shows the location of the sample mean as well as the median.

Two other fields are also useful:

- **Display statistics:** determines whether summary statistics are included in the righthand margin of the plot.
- **Decimal places:** controls the number of decimal places shown for the calculated statistics.

## Calculations

For information on how the nonparametric density estimate is calculated, refer to the PDF file titled *Distribution Fitting (Uncensored Data)*. For information on the options for the box-and-whisker plot, refer to the PDF file titled *Box-and-Whisker Plot*.