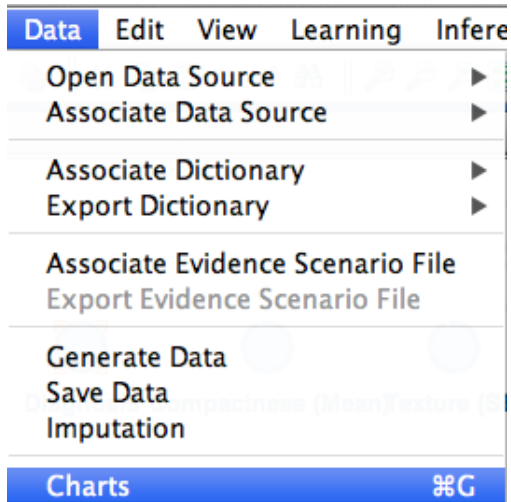


Box Plot (5.1)

Context

Data | Charts



New: Box Plot

A *Box Plot* (also known as a *box-and-whisker plot*) is a classical tool used in *Descriptive Statistics* to analyze the distributions of numerical variables.

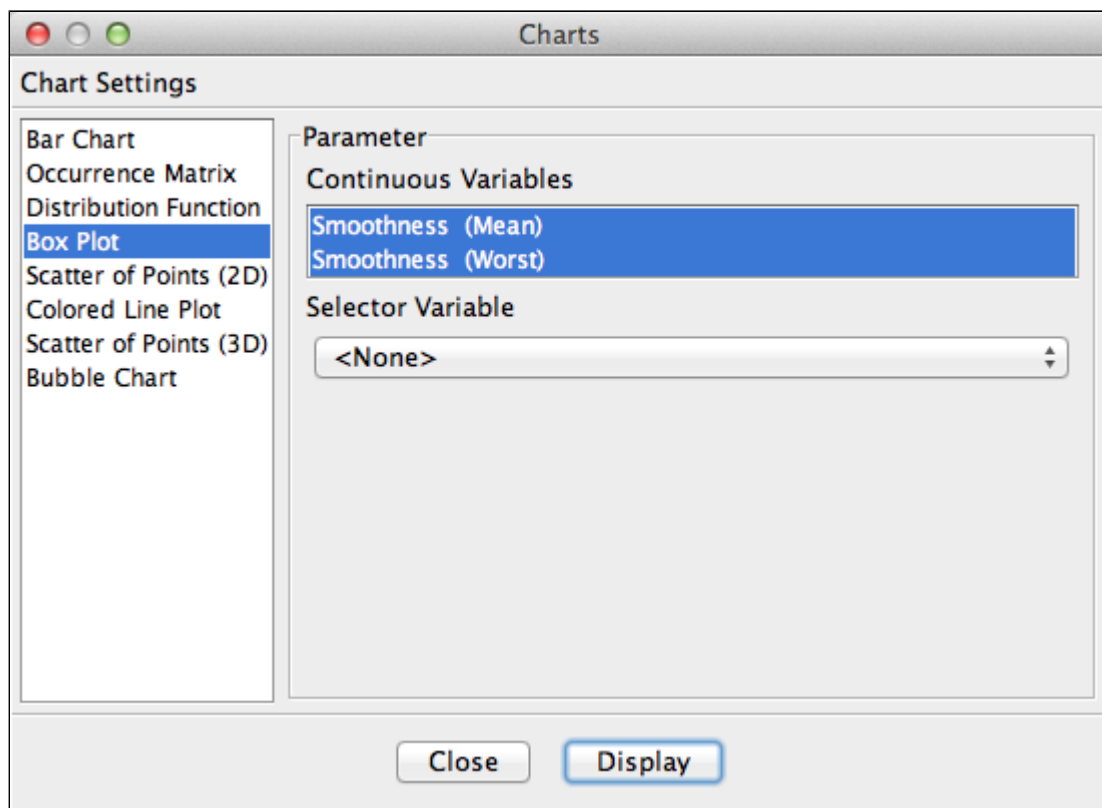
It shows:

- the *First Quartile* Q_1
- the *Median* Q_2
- the *Last Quartile* Q_3
- the *Minimum*, i.e. the lowest datum still within 1.5 of the Interquartile Range ($IQR = Q_3 - Q_1$) of the first quartile
- the *Maximum*, i.e. the highest datum still within 1.5 IQR of the last quartile
- the first *Notch*, $Q_2 - 1.58 \times IQR / \sqrt{N}$, where N is the number of observations
- the last *Notch*, $Q_2 + 1.58 \times IQR / \sqrt{N}$. Notches are useful in offering a rough guide to significance of difference of medians.
- the *Mean*
- *Lower Suspect* values that are within the *Minimum* and $Q_1 - 3 \times IQR$
- *Upper Suspect* values that are within the *Maximum* and $Q_3 + 3 \times IQR$
- *Lower Extreme* values that are inferior to $Q_1 - 3 \times IQR$
- *Upper Extreme* values that are superior to $Q_3 + 3 \times IQR$

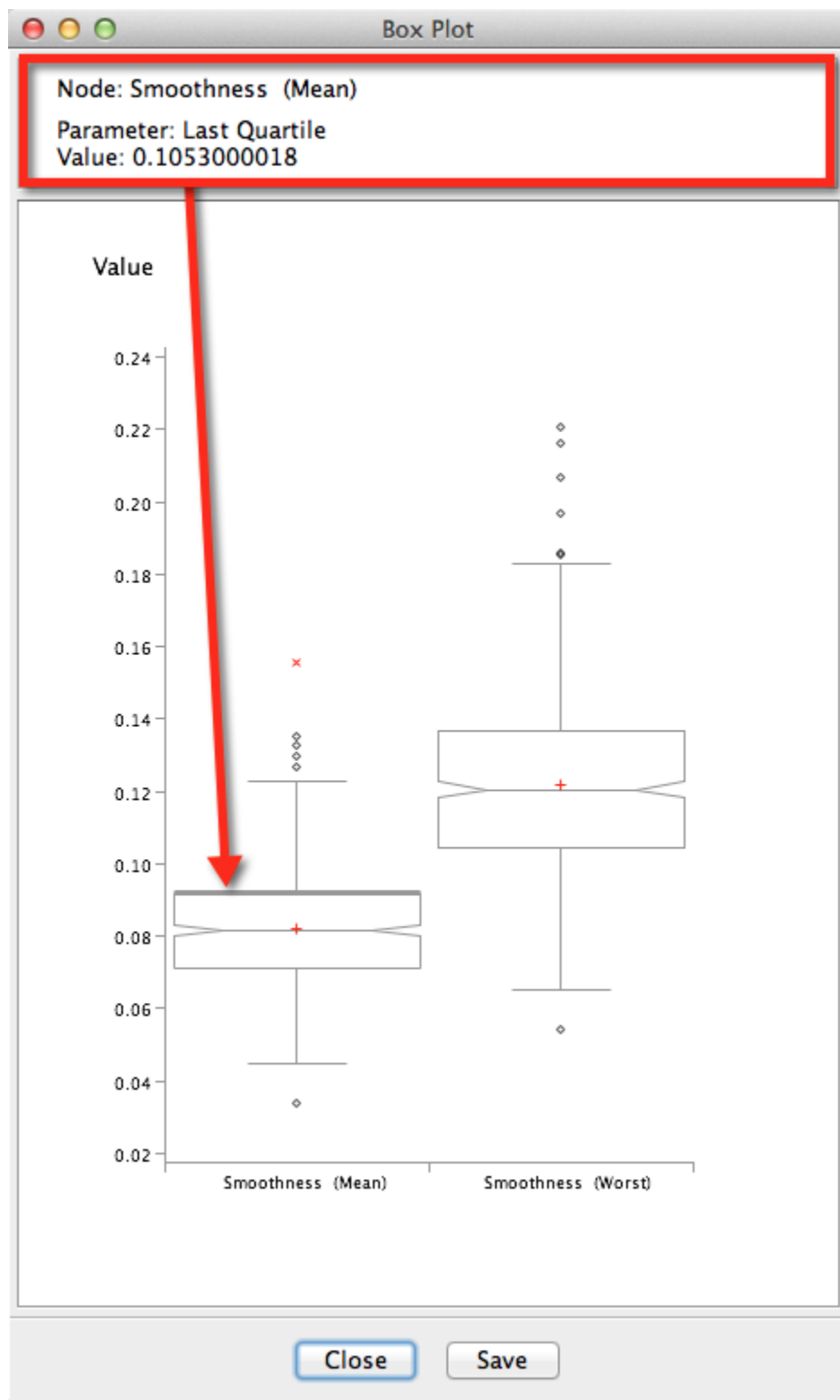
Example

Without Selector Variable

You select the variable to be analyzed in the **Parameter** window.

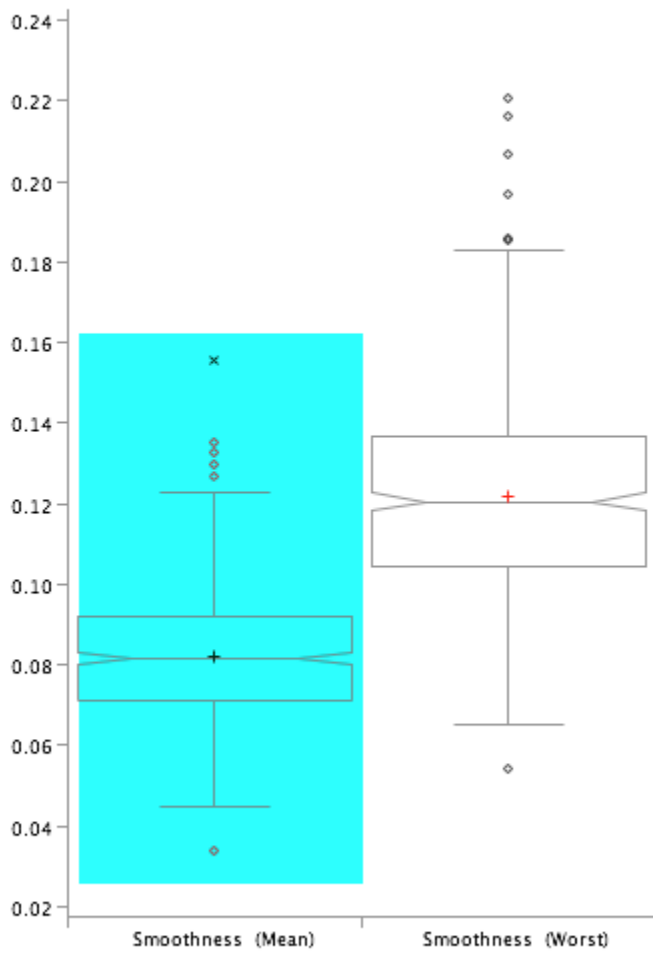


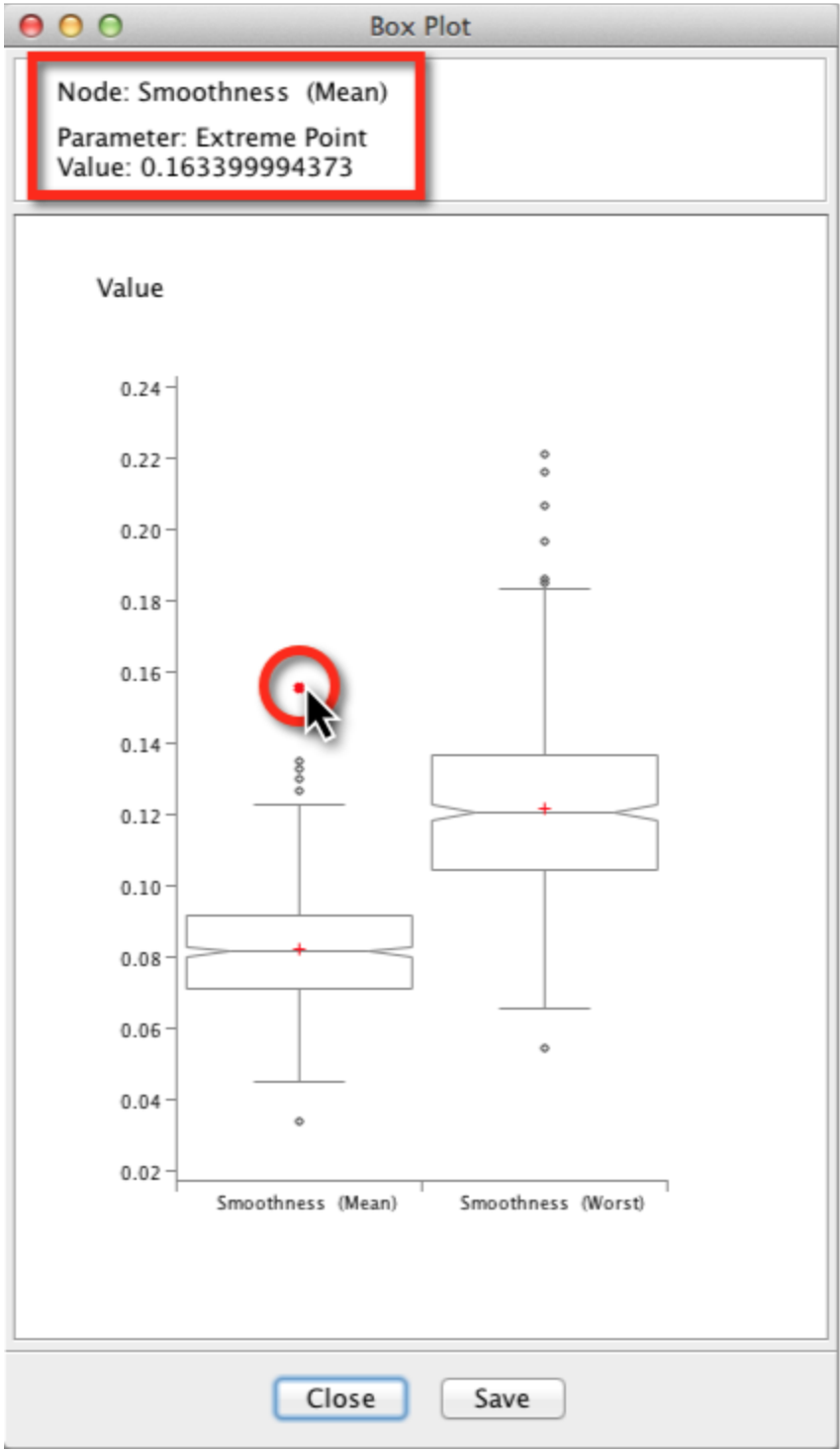
Upon clicking on Display, the box plots are generated. Hovering over each element returns its description and numerical value in the upper part of the window.

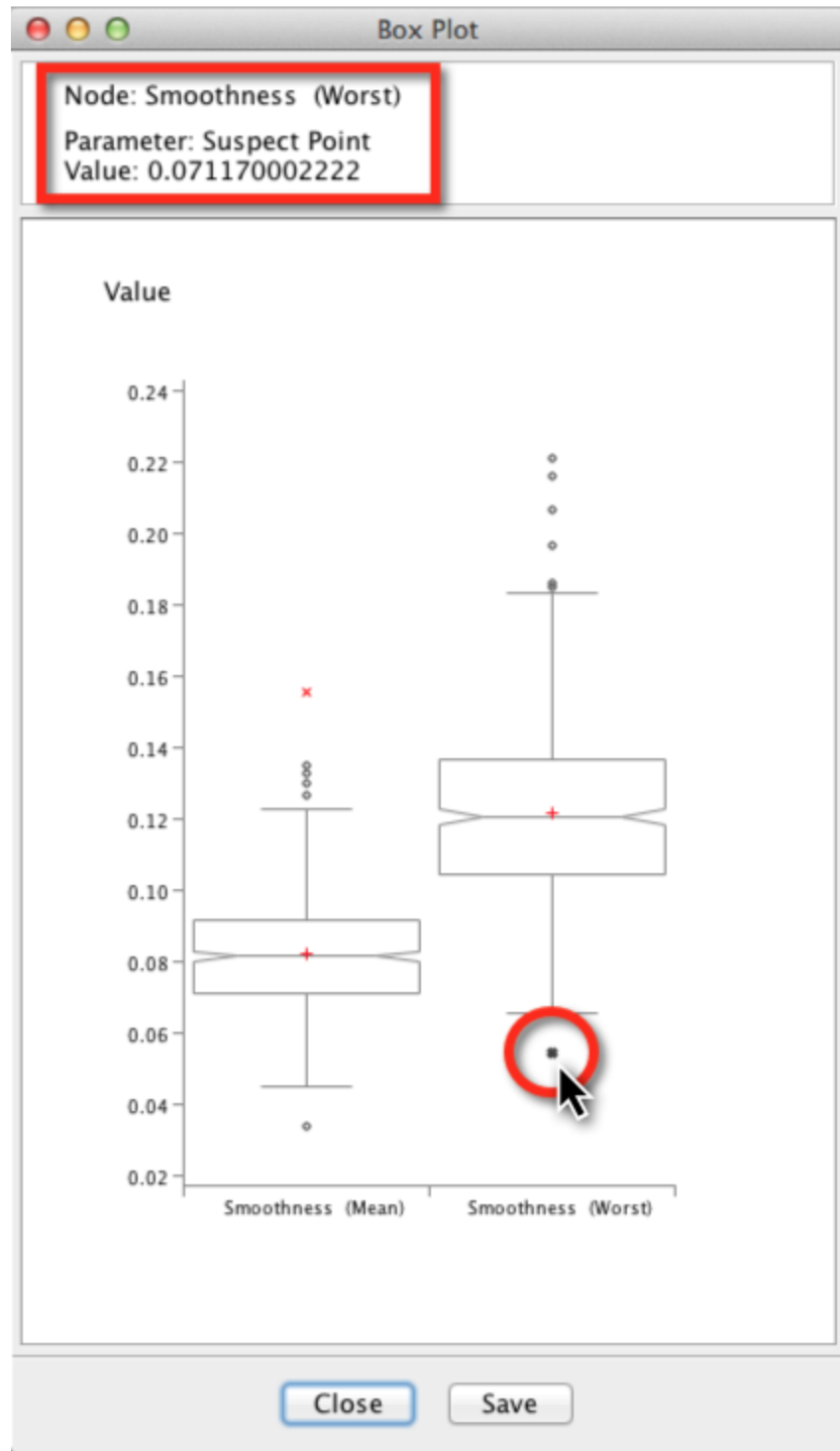


You can also zoom in vertically by selecting the y-range as illustrated below. Double-clicking on the graph returns the default view.

Value







Clicking on a **Suspect Point** or **Extreme Point** (as indicated in the header) brings up a table with the corresponding data record.

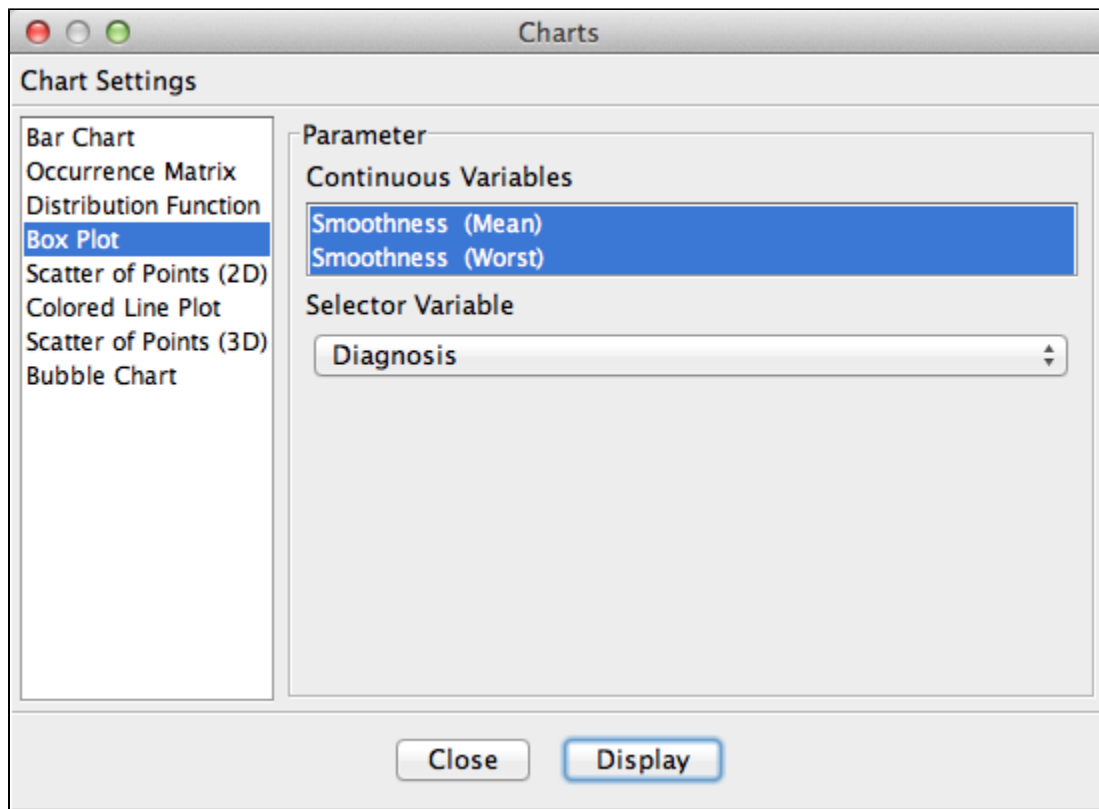
The figure is a 'Corresponding Data' window. It contains a table with the following data:

Diagnosis	Radius (Me...	Texture (M...	Perimeter (...	Area (Mean)	Smoothne...	Compactn...	Concavity...	Concave P...	Symmetry...
M	17.99	10.38	122.8	1,001	0.118	0.278	0.3	0.147	0.242

An 'OK' button is located at the bottom right of the window.

With a Selector Variable

This option is useful for comparing the conditional distributions of a variable given the states of a **Selector Variable**.



The graph below allows to compare the distributions of two variables, *Smoothness (Mean)* and *Smoothness (Worst)*, given the two states of the variable *Diagnosis (B and M)*.

Value

