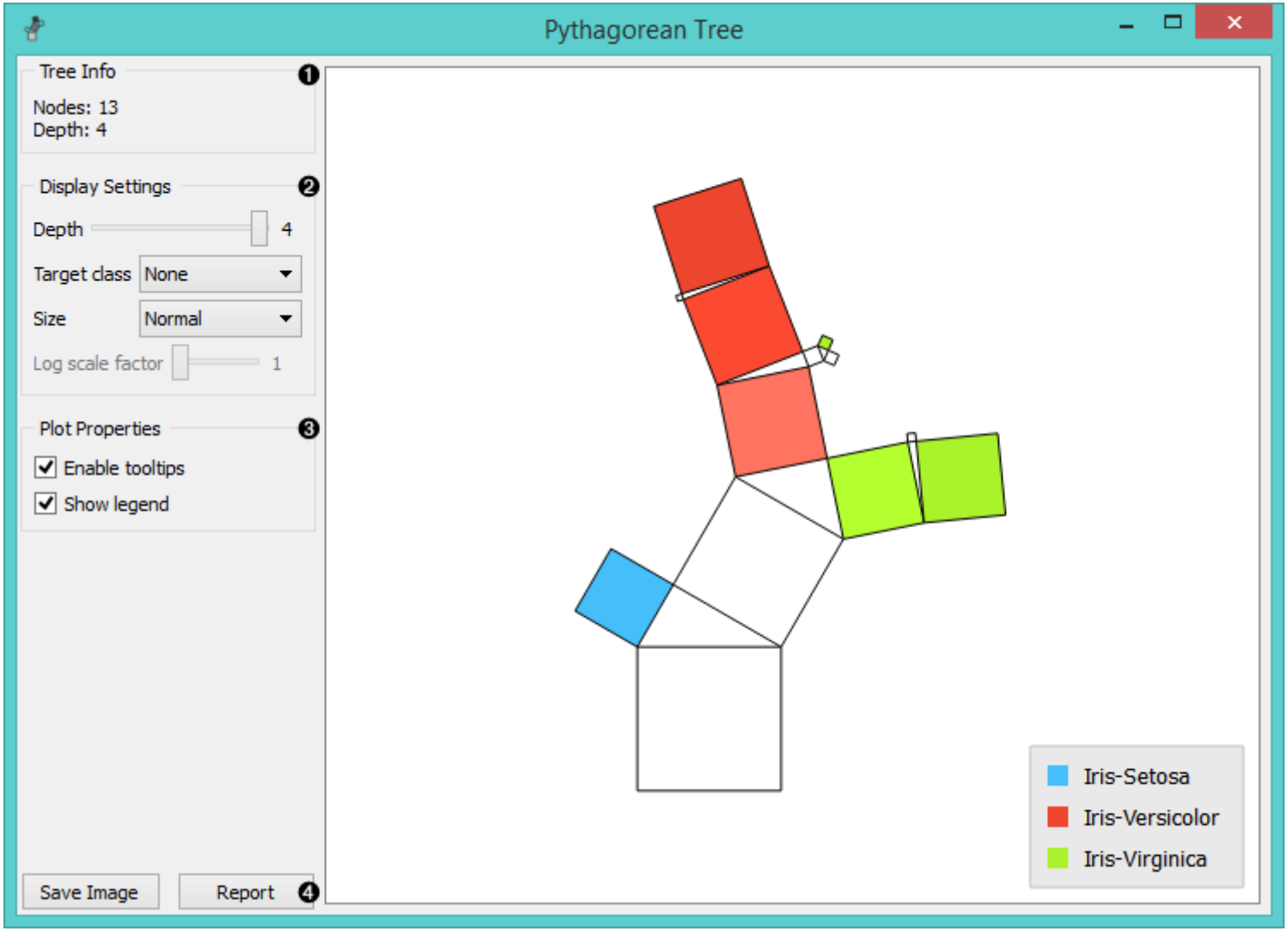
Pythagorean Tree

Pythagorean tree visualization for classification or regression trees.

Inputs

- Tree: tree model
- Selected Data: instances selected from the tree

Pythagorean Trees are plane fractals that can be used to depict general tree hierarchies as presented in an article by Fabian Beck and co-authors. In our case, they are used for visualizing and exploring tree models, such as Tree.

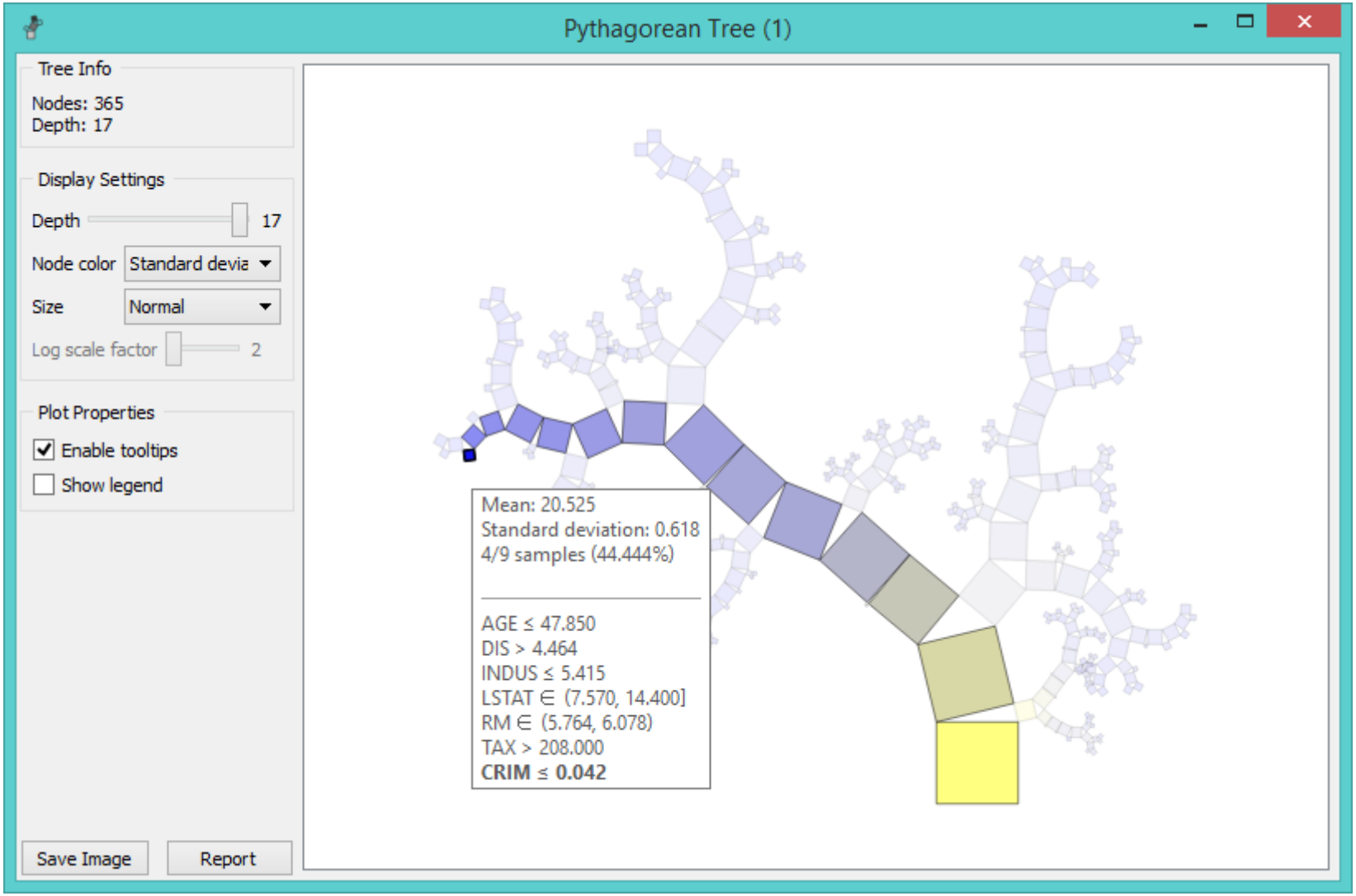


1. Information on the input tree model.
2. Visualization parameters:
   - *Depth*: set the depth of displayed trees.
   - *Target class* (for classification trees): the intensity of the color for nodes of the tree will correspond to the probability of the target class. If *None* is selected, the color of the node will denote the most probable class.
   - *Node color* (for regression trees): node colors can correspond to mean or standard deviation of class value of the training data instances in the node.
   - *Size*: define a method to compute the size of the square representing the node. *Normal* will keep node sizes correspond to the size of training data subset in the node. *Square root* and *Logarithmic* are the respective transformations of the node size.

- *Log scale factor* is only enabled when *logarithmic* transformation is selected. You can set the log factor between 1 and 10.

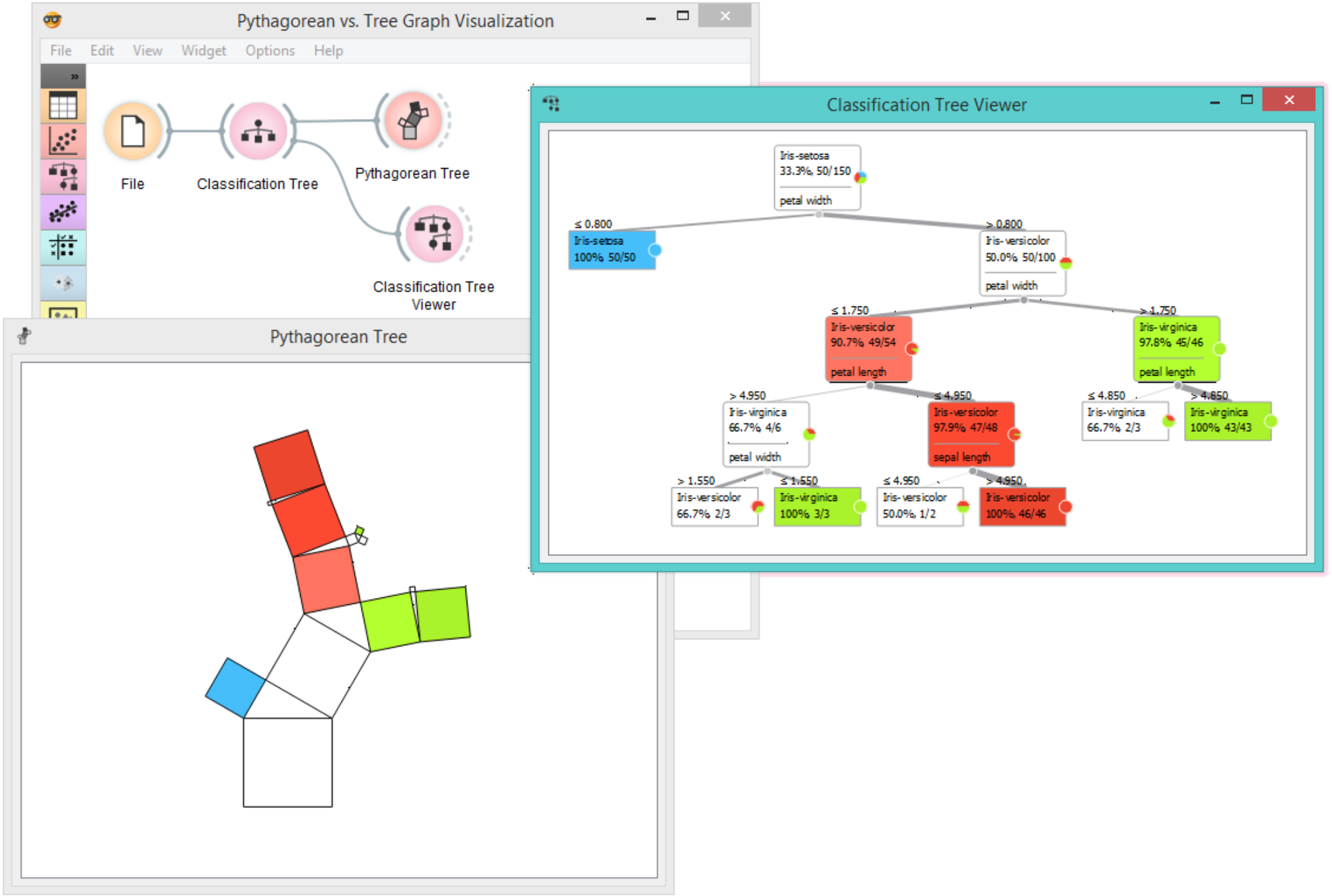
3. Plot properties:
   - *Enable tooltips*: display node information upon hovering.
   - *Show legend*: shows color legend for the plot.
4. Reporting:
   - *Save Image*: save the visualization to a SVG or PNG file.
   - *Report*: add visualization to the report.

Pythagorean Tree can visualize both classification and regression trees. Below is an example for regression tree. The only difference between the two is that regression tree doesn’t enable coloring by class, but can color by class mean or standard deviation.

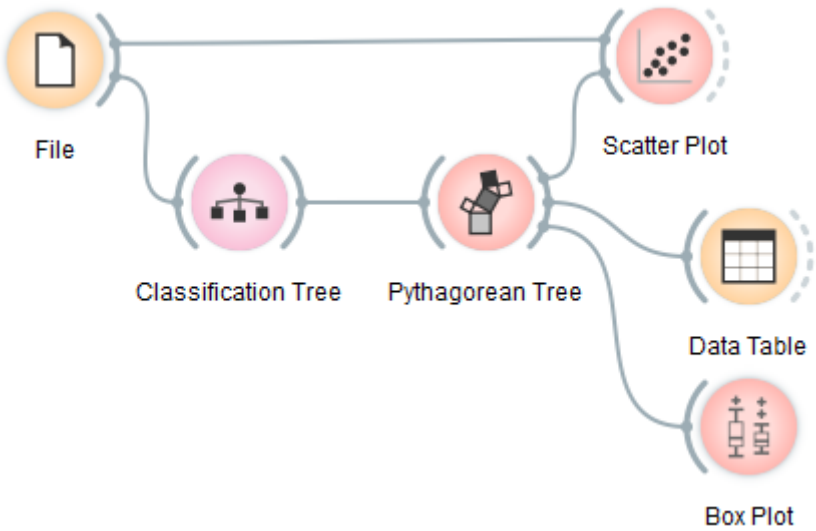


Example

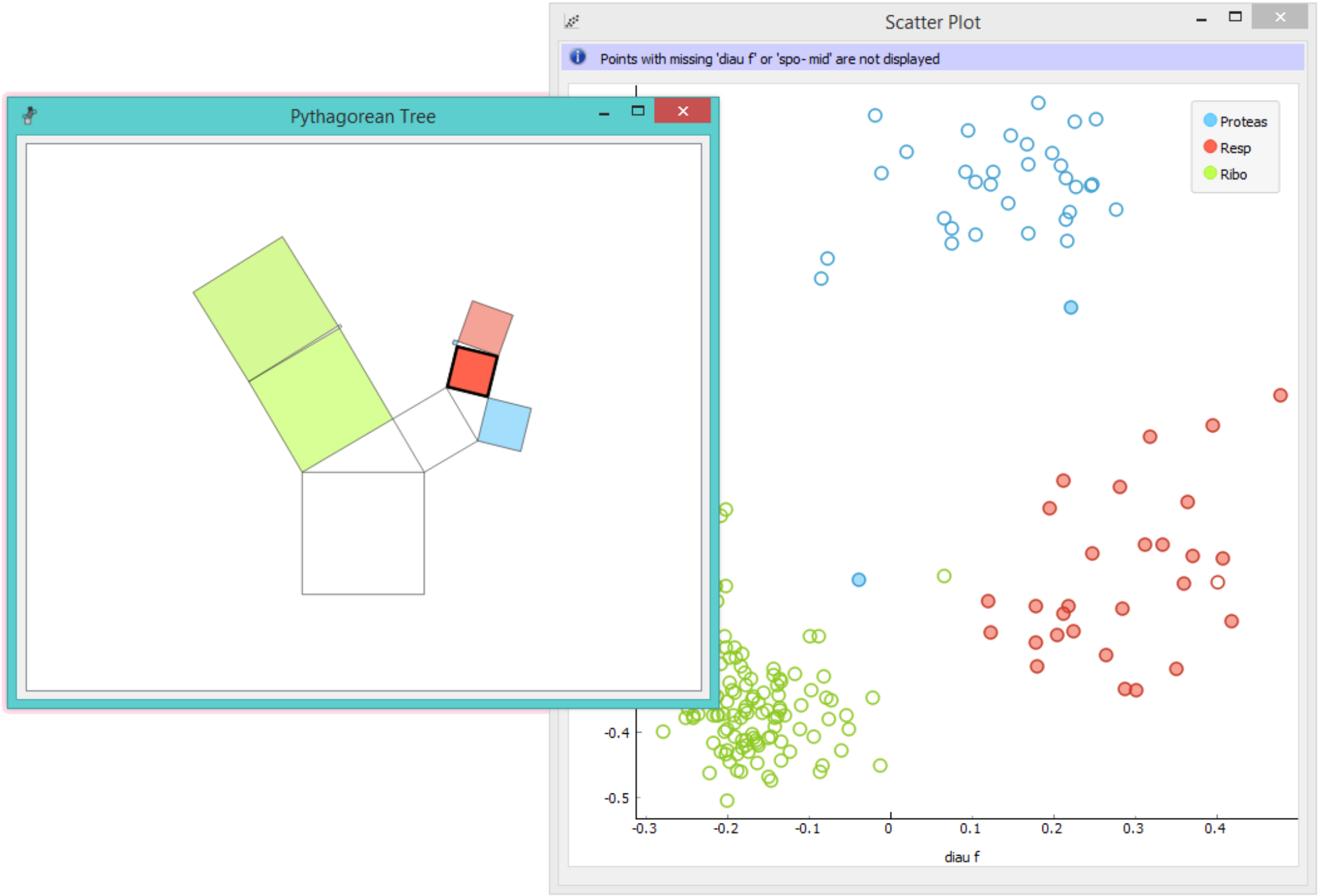
The workflow from the screenshot below demonstrates the difference between Tree Viewer and Pythagorean Tree. They can both visualize Tree, but Pythagorean visualization takes less space and is more compact, even for a small Iris flower dataset. For both visualization widgets, we have hidden the control area on the left by clicking on the splitter between control and visualization area.


Pythagorean vs. Tree Graph Visualization
File Edit View Widget Options Help
File
Classification Tree
Pythagorean Tree
Classification Tree Viewer
Pythagorean Tree
Classification Tree Viewer
Iris-setosa
33.3%, 50/150
petal width
≤ 0.800
Iris-setosa
100%, 50/50
> 0.800
Iris-versicolor
50.0%, 50/100
petal width
≤ 1.750
Iris-versicolor
90.7%, 49/54
petal length
> 1.750
Iris-virginica
97.8%, 45/46
petal length
> 4.950
Iris-virginica
66.7%, 4/6
petal width
≤ 4.950
Iris-versicolor
97.9%, 47/48
sepal length
≤ 4.850
Iris-virginica
66.7%, 2/3
> 4.850
Iris-virginica
100%, 43/43
> 1.550
Iris-versicolor
66.7%, 2/3
≤ 1.550
Iris-virginica
100%, 3/3
≤ 4.950
Iris-versicolor
50.0%, 1/2
> 4.950
Iris-versicolor
100%, 46/46

Pythagorean Tree is interactive: click on any of the nodes (squares) to select training data instances that were associated with that node. The following workflow explores these feature.



The selected data instances are shown as a subset in the Scatter Plot, sent to the Data Table and examined in the Box Plot. We have used brown-selected dataset in this example. The tree and scatter plot are shown below; the selected node in the tree has a black outline.


Pythagorean Tree
Scatter Plot
Points with missing 'diau f' or 'spo- mid' are not displayed
Proteas
Resp
Ribo
-0.4
-0.5
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
diau f

References

Beck, F., Burch, M., Munz, T., Di Silvestro, L. and Weiskopf, D. (2014). Generalized Pythagoras Trees for Visualizing Hierarchies. In IVAPP ‘14 Proceedings of the 5th International Conference on Information Visualization Theory and Applications, 17-28.