Data Mining Tools

- Data mining serves the primary purpose of discovering patterns among large volumes of data.

- And transforming data into more refined/actionable information.

- This technique utilizes
  - Specific algorithms
  - Statistical analysis
  - Artificial intelligence
  - Database systems.

- It aims to extract information from huge data sets and convert it into an understandable structure for future use.

Most Popular Data Mining Tools

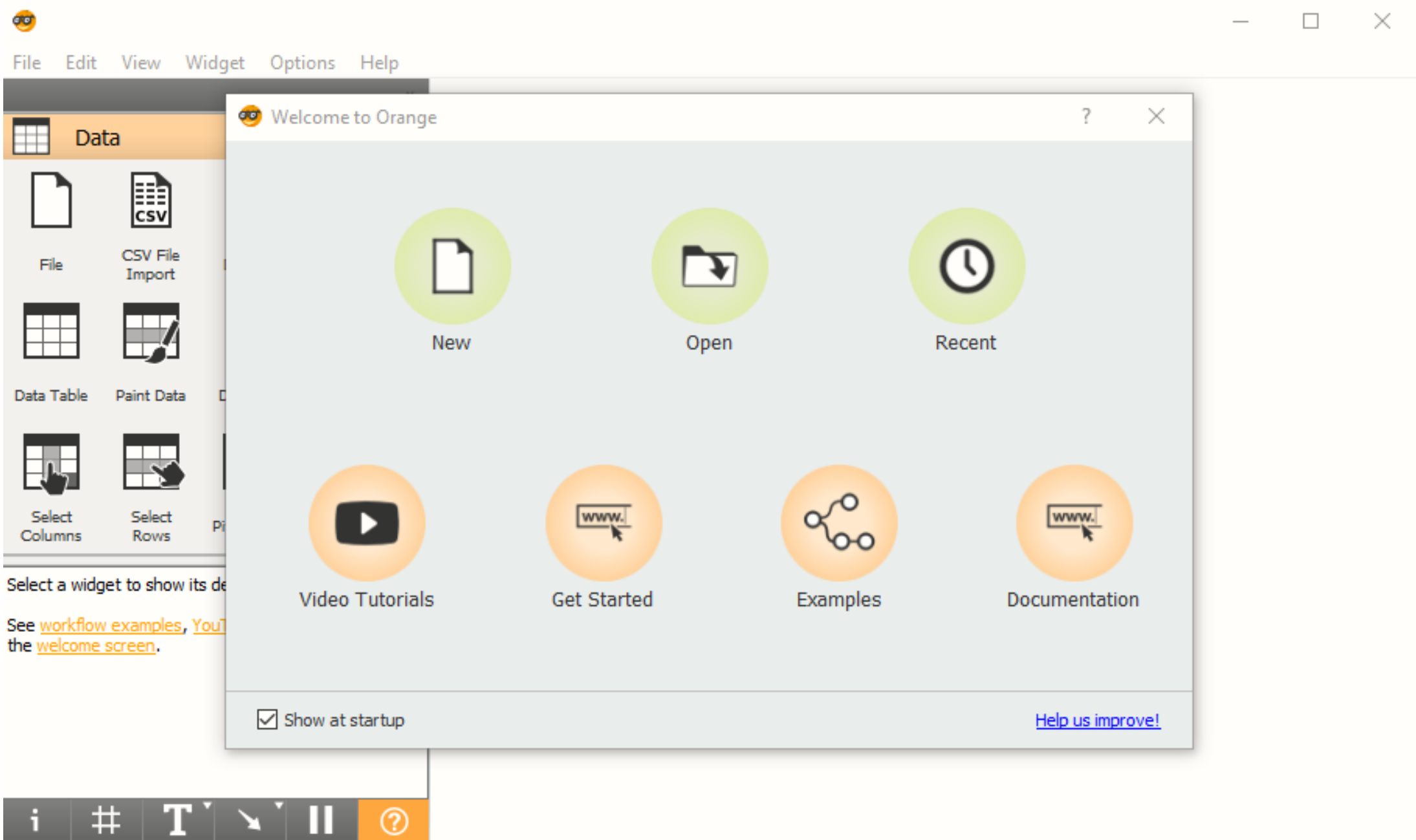
Tools	Introduction	Availability
Orange	Orange is a perfect software suite for machine learning & data mining. It best aids the data visualization and is a component based software. It has been written in Python computing language.	Open source
Xplenty	Xplenty provides a platform that has functionalities to integrate, process, and prepare data for analytics.	Licensed tools.
Rapid Miner	Rapid Miner is one of the best predictive analysis system. It is written in JAVA programming language.	Open source
Weka	Also known as Waikato Environment is a machine learning software developed at the University of Waikato in New Zealand. It is best suited for data analysis and predictive modeling	Free software
Oracle Data Mining	Oracle data mining software provides excellent data mining algorithms for data classification, prediction, regression	License

- It is a component-based software.
- The components of orange are called ‘widgets'.
- These widgets range from data visualization & pre-processing to an evaluation of algorithms and predictive modeling.

Widgets offer major functionalities like

- Showing data table and allowing to select features
- Reading the data
- Training predictors and to compare learning algorithms
- Visualizing data elements etc.

https://orange.biolab.si/


File Edit View Widget Options Help
Data
File
CSV File Import
Data Table
Paint Data
Select Columns
Select Rows
Select a widget to show its de
See workflow examples, YouT
the welcome screen.
Welcome to Orange
New
Open
Recent
Video Tutorials
Get Started
Examples
Documentation
Show at startup
Help us improve!

Data

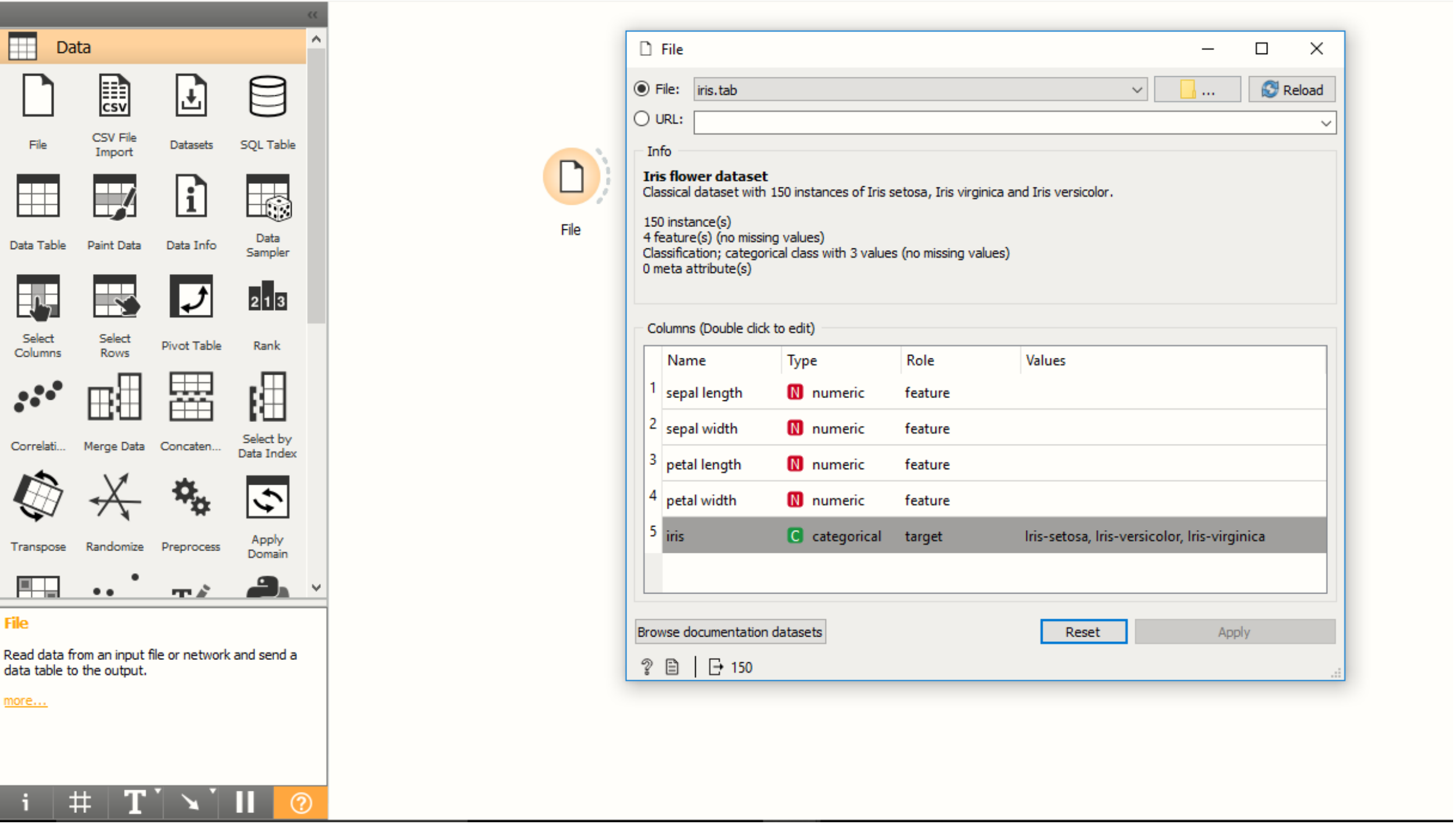
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concaten...
Select by Data Index
Transpose
Randomize
Preprocess
Apply Domain

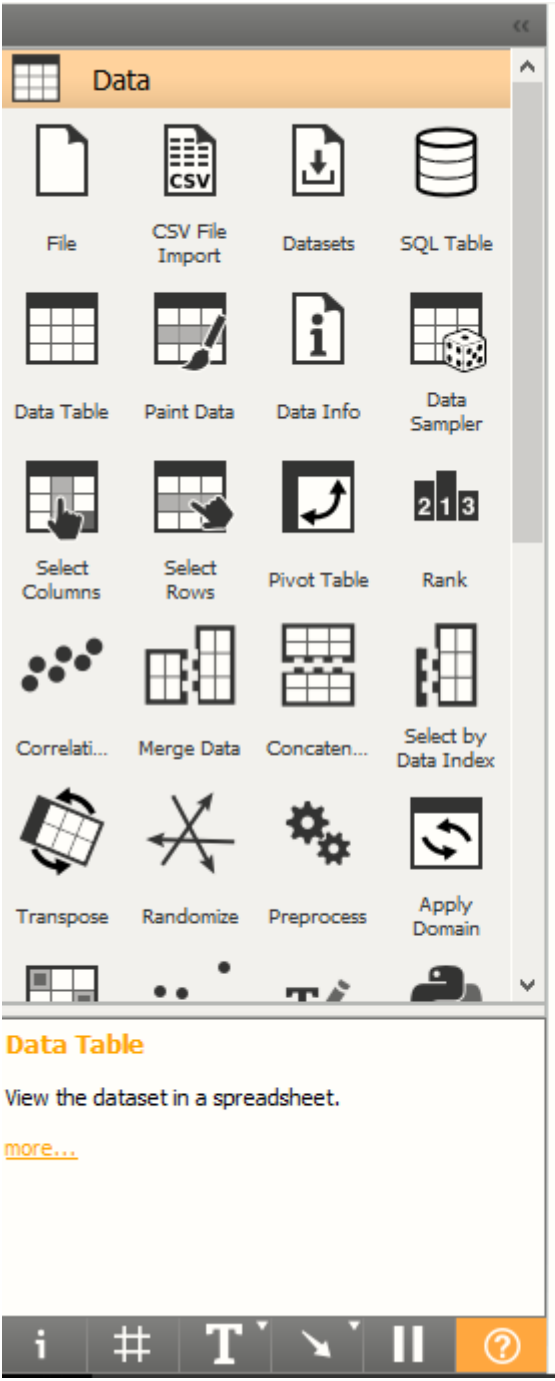
Select a widget to show its description.

See workflow examples, YouTube tutorials, or open the welcome screen.

Drag and drop file to canvas
or click on file or any other data tools you want to use

File


Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concaten...
Select by Data Index
Transpose
Randomize
Preprocess
Apply Domain
File
Read data from an input file or network and send a data table to the output.
more...
File
File:
iris.tab
...
Reload
URL:
Info
Iris flower dataset
Classical dataset with 150 instances of Iris setosa, Iris virginica and Iris versicolor.
150 instance(s)
4 feature(s) (no missing values)
Classification; categorical class with 3 values (no missing values)
0 meta attribute(s)
Columns (Double click to edit)
Name
Type
Role
Values
1 sepal length numeric feature
2 sepal width numeric feature
3 petal length numeric feature
4 petal width numeric feature
5 iris categorical target Iris-setosa, Iris-versicolor, Iris-virginica
Browse documentation datasets
Reset
Apply
150



Data Table

Variables

- [x] Show variable labels (if present)
- [x] Visualize numeric values
- [x] Color by instance classes

Selection

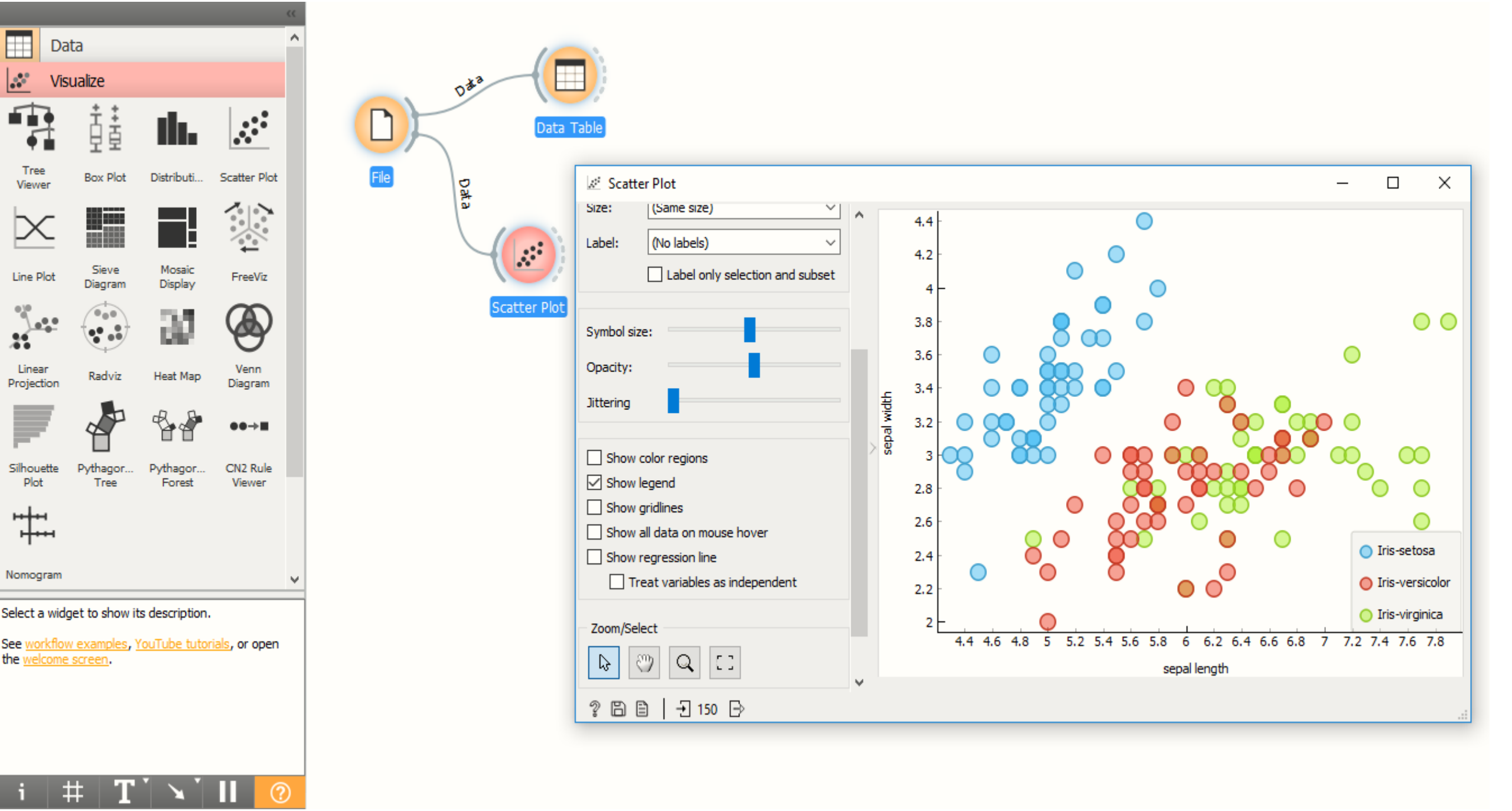
- [x] Select full rows

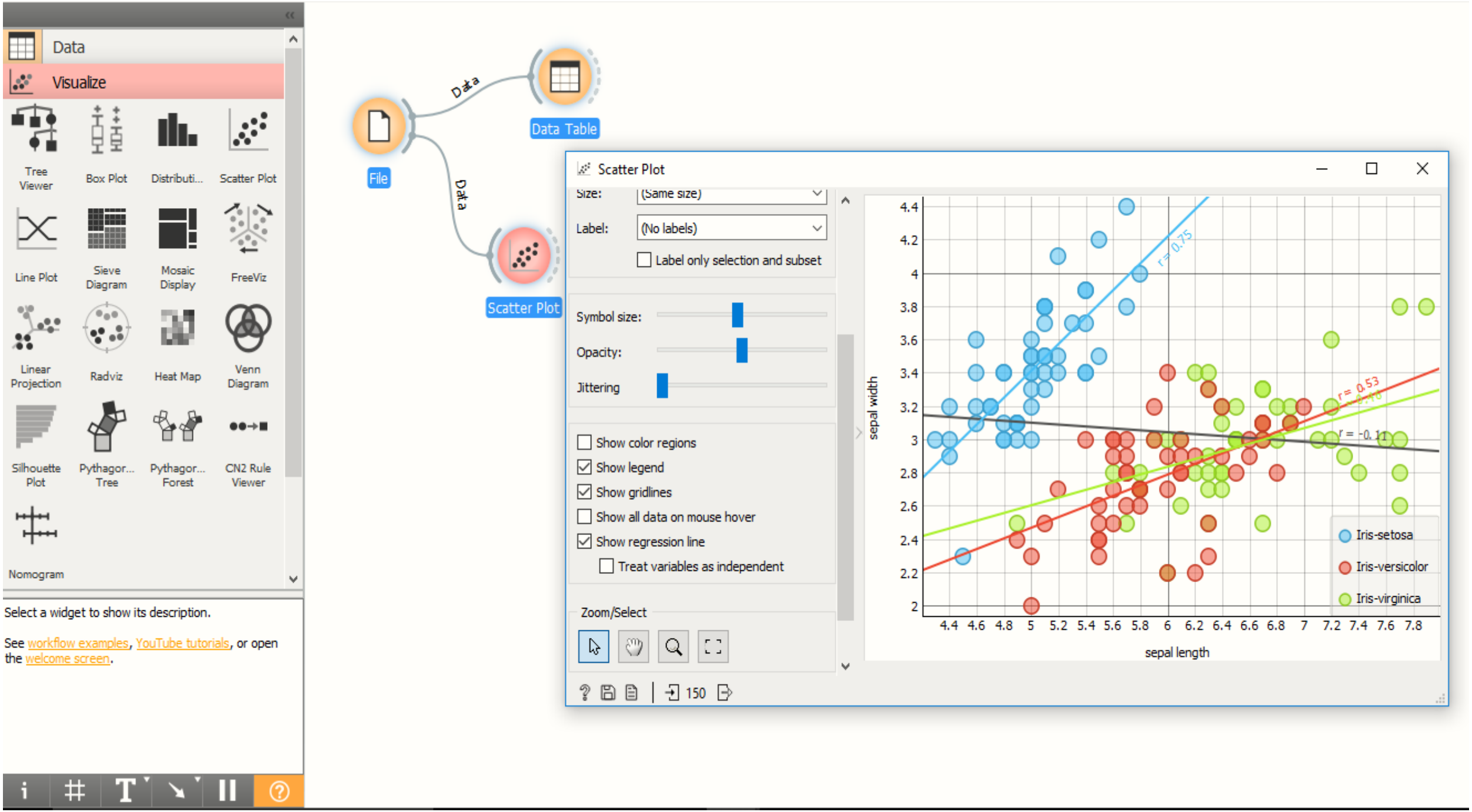
Restore Original Order

- [x] Send Automatically

150

	iris	sepal length	sepal width	petal length	petal width
1	Iris-setosa	5.1	3.5	1.4	0.2
2	Iris-setosa	4.9	3.0	1.4	0.2
3	Iris-setosa	4.7	3.2	1.3	0.2
4	Iris-setosa	4.6	3.1	1.5	0.2
5	Iris-setosa	5.0	3.6	1.4	0.2
6	Iris-setosa	5.4	3.9	1.7	0.4
7	Iris-setosa	4.6	3.4	1.4	0.3
8	Iris-setosa	5.0	3.4	1.5	0.2
9	Iris-setosa	4.4	2.9	1.4	0.2
10	Iris-setosa	4.9	3.1	1.5	0.1
11	Iris-setosa	5.4	3.7	1.5	0.2
12	Iris-setosa	4.8	3.4	1.6	0.2
13	Iris-setosa	4.8	3.0	1.4	0.1
14	Iris-setosa	4.3	3.0	1.1	0.1


Data
Visualize
Tree Viewer
Box Plot
Distributi...
Scatter Plot
Line Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Silhouette Plot
Pythagor... Tree
Pythagor... Forest
CN2 Rule Viewer
Nomogram
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
Data
Data
File
Data Table
Scatter Plot
Scatter Plot
Size:
(Same size)
Label:
(No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering
Show color regions
Show legend
Show gridlines
Show all data on mouse hover
Show regression line
Treat variables as independent
Zoom/Select
150
sepal width
sepal length
Iris-setosa
Iris-versicolor
Iris-virginica


Data
Visualize
Tree Viewer
Box Plot
Distributi...
Scatter Plot
Line Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Silhouette Plot
Pythagor... Tree
Pythagor... Forest
CN2 Rule Viewer
Nomogram
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
File
Data
Data Table
Data
Scatter Plot
Scatter Plot
Size:
(Same size)
Label:
(No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering
Show color regions
Show legend
Show gridlines
Show all data on mouse hover
Show regression line
Treat variables as independent
Zoom/Select
150
sepal width
sepal length
r = 0.75
r = 0.53
r = -0.11
Iris-setosa
Iris-versicolor
Iris-virginica