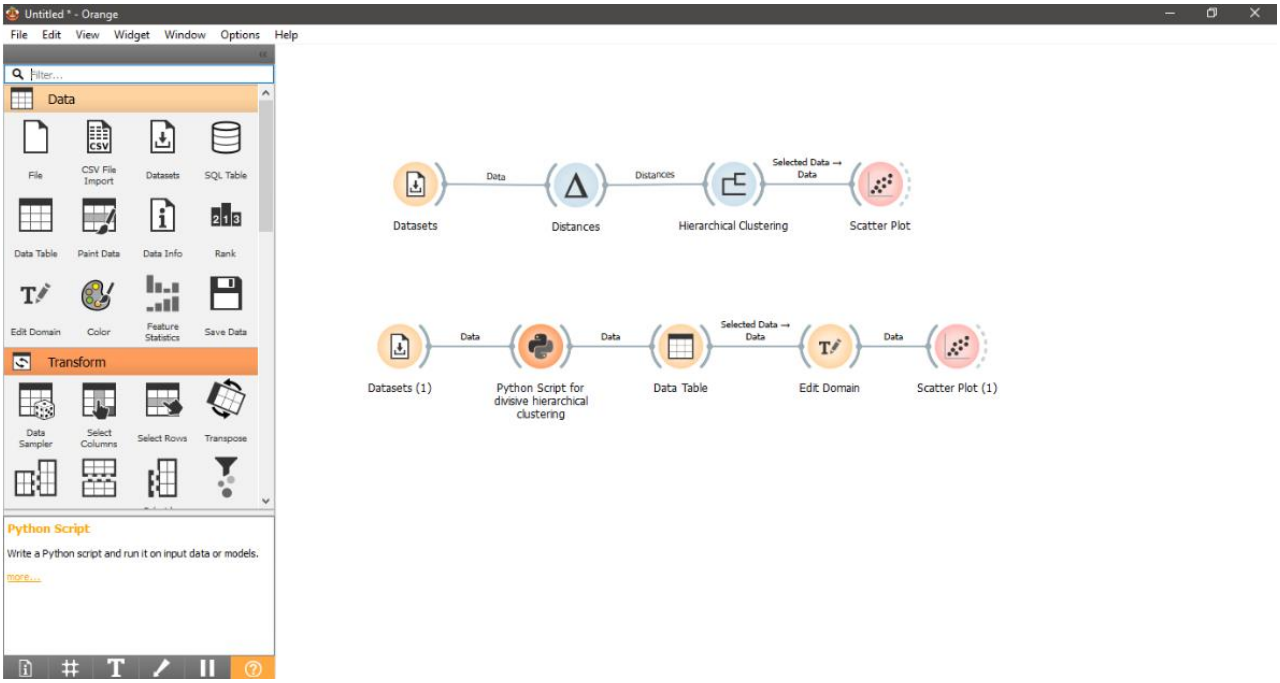
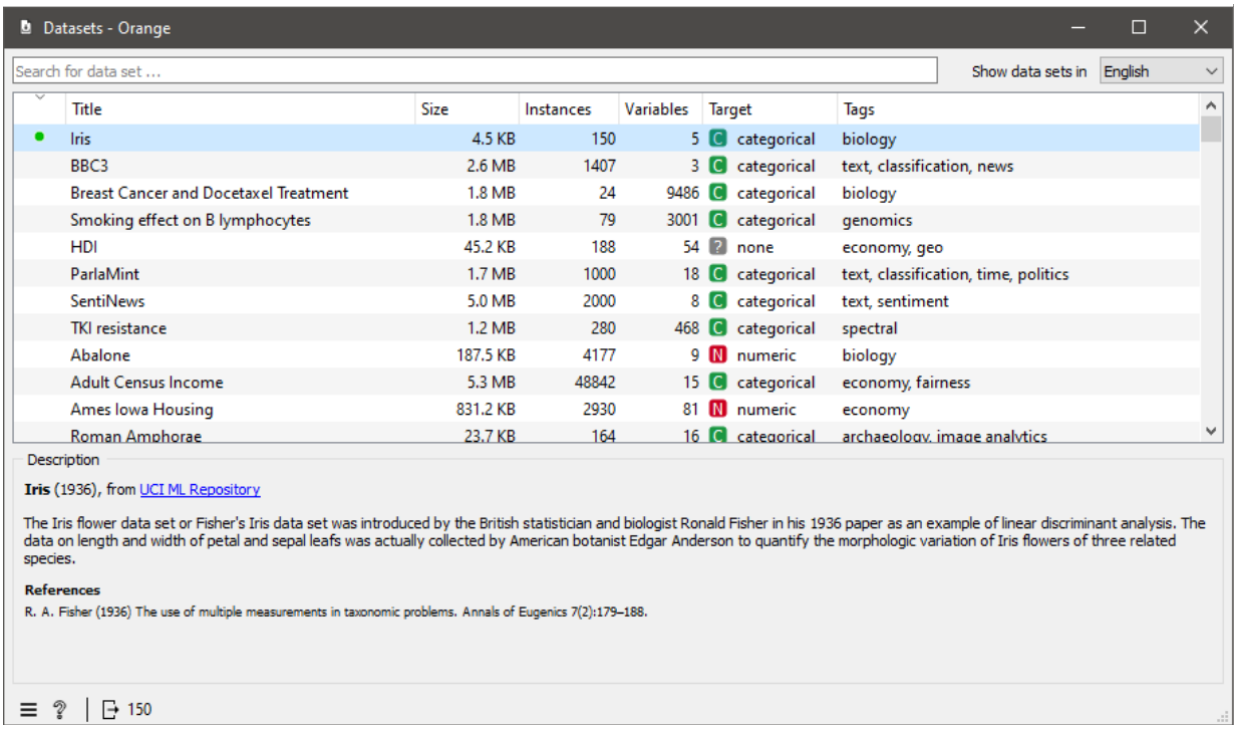
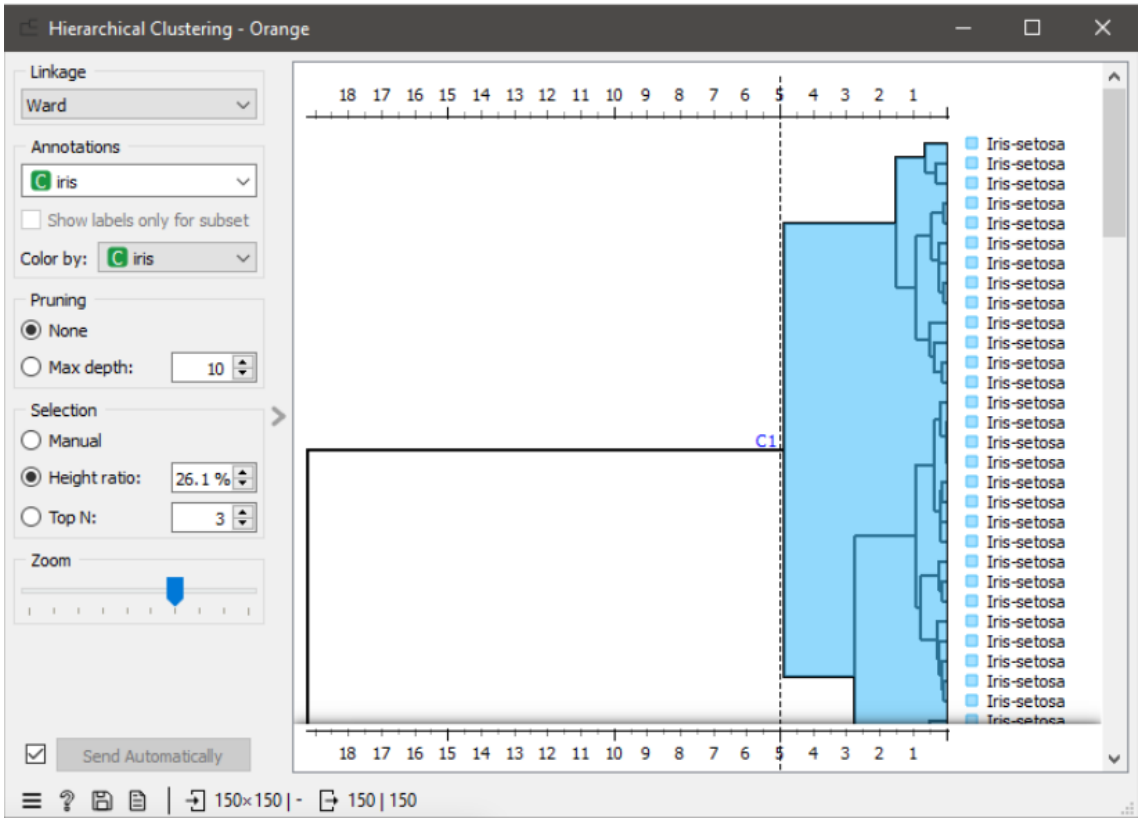
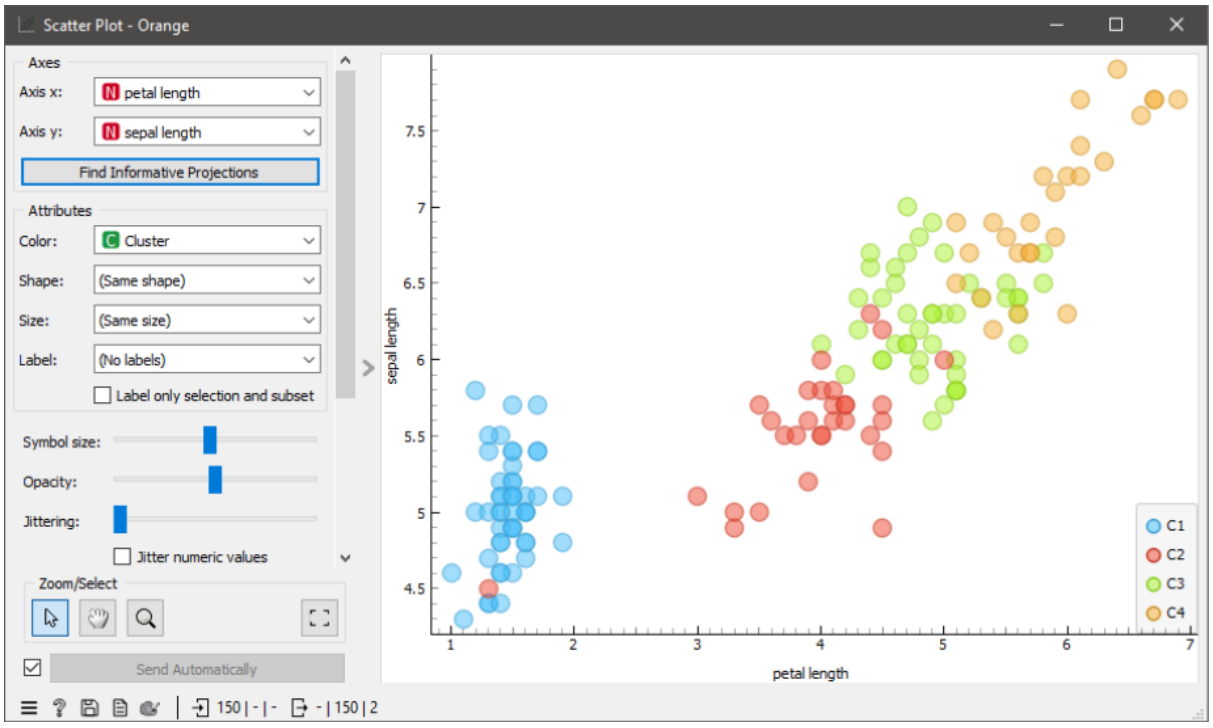
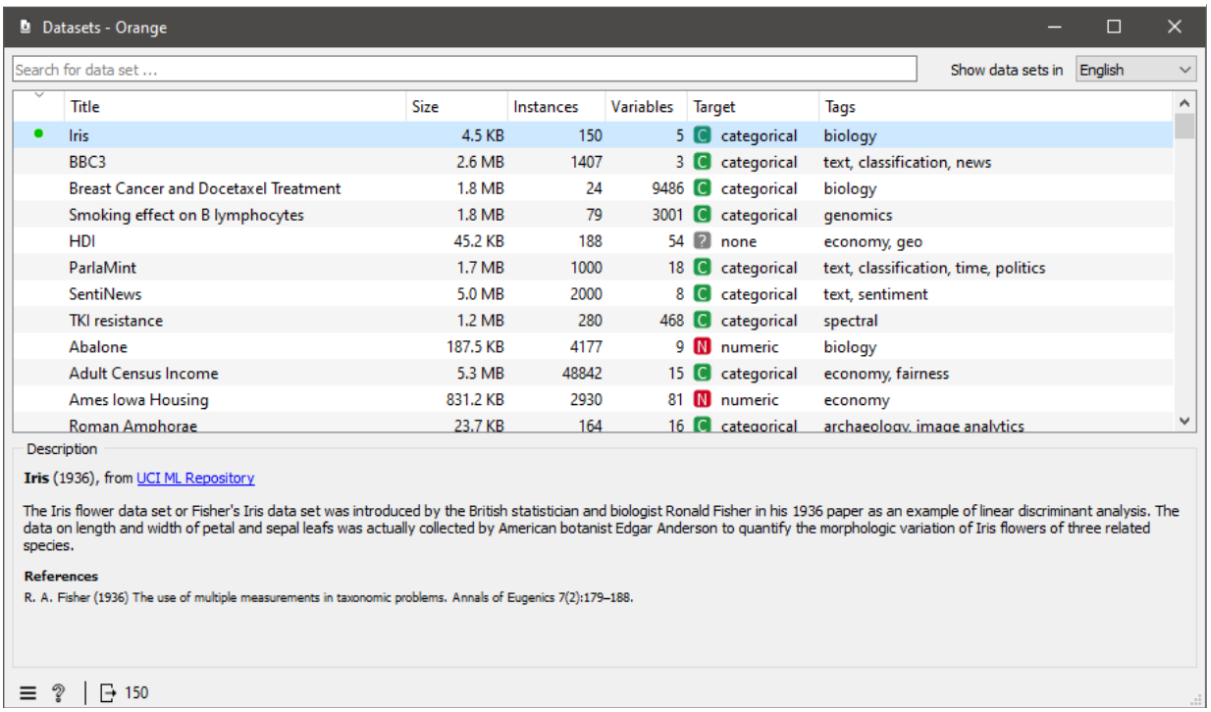
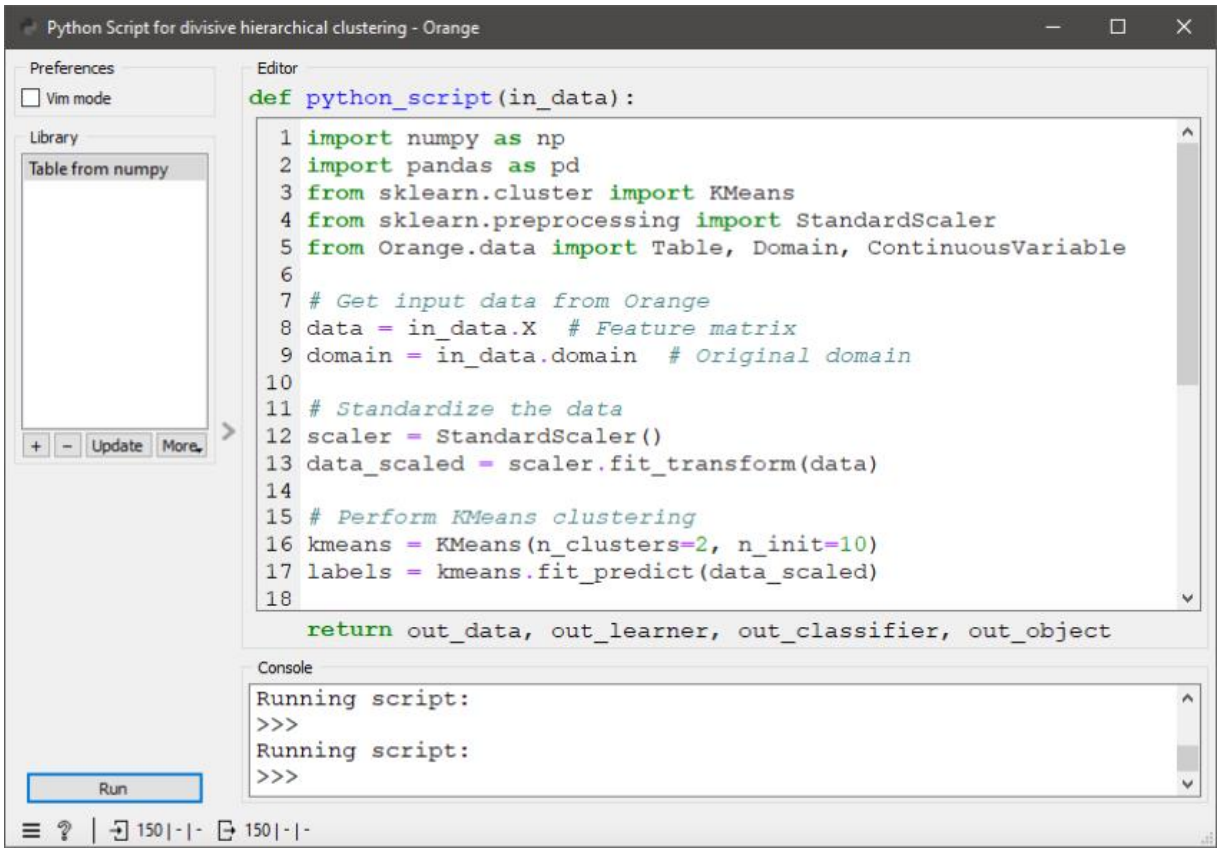

Untitled * - Orange
File Edit View Widget Window Options Help
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Rank
Edit Domain
Color
Feature Statistics
Save Data
Transform
Data Sampler
Select Columns
Select Rows
Transpose
Python Script
Write a Python script and run it on input data or models.
more...
Data
Distances
Selected Data → Data
Datasets
Distances
Hierarchical Clustering
Scatter Plot
Data
Data
Selected Data → Data
Data
Datasets (1)
Python Script for divisive hierarchical clustering
Data Table
Edit Domain
Scatter Plot (1)

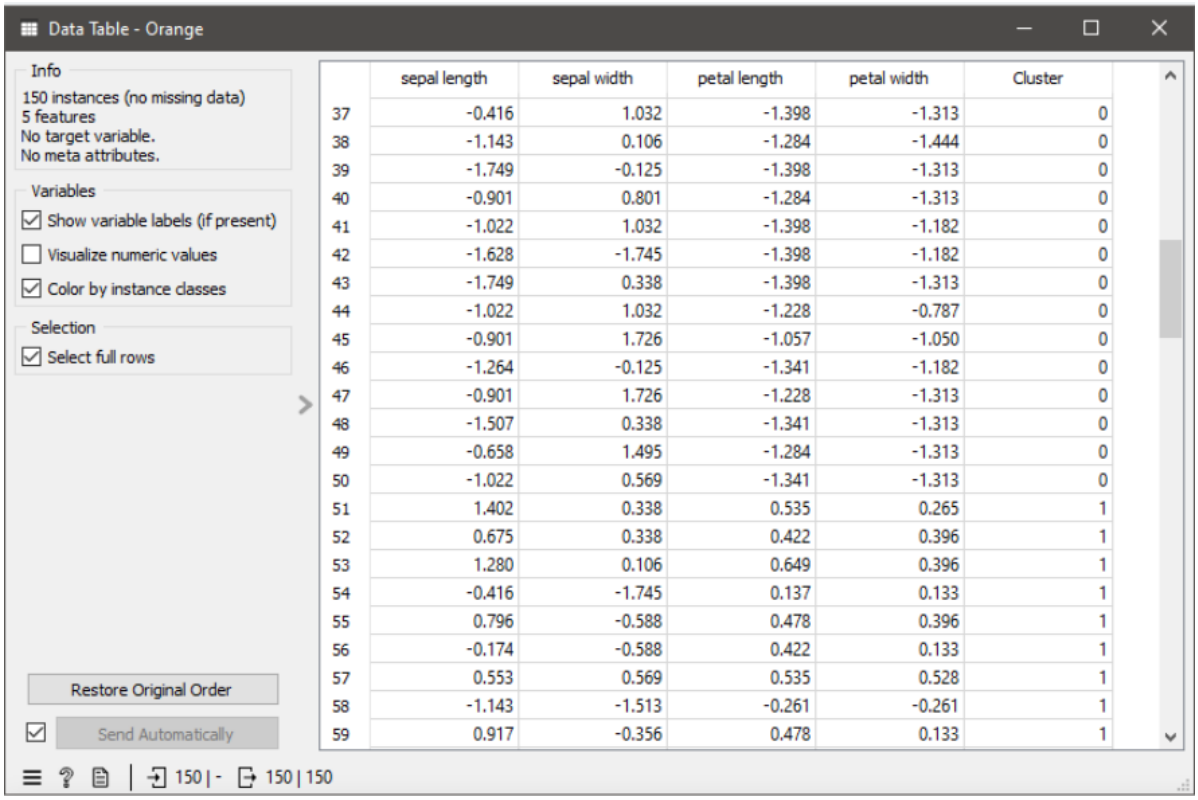

Datasets - Orange
Search for data set ...
Show data sets in English
Title Size Instances Variables Target Tags
Iris 4.5 KB 150 5 categorical biology
BBC3 2.6 MB 1407 3 categorical text, classification, news
Breast Cancer and Docetaxel Treatment 1.8 MB 24 9486 categorical biology
Smoking effect on B lymphocytes 1.8 MB 79 3001 categorical genomics
HDI 45.2 KB 188 54 none economy, geo
ParlaMint 1.7 MB 1000 18 categorical text, classification, time, politics
SentiNews 5.0 MB 2000 8 categorical text, sentiment
TKI resistance 1.2 MB 280 468 categorical spectral
Abalone 187.5 KB 4177 9 numeric biology
Adult Census Income 5.3 MB 48842 15 categorical economy, fairness
Ames Iowa Housing 831.2 KB 2930 81 numeric economy
Roman Amphorae 23.7 KB 164 16 categorical archaeology, image analytics
Description
Iris (1936), from UCI ML Repository
The Iris flower data set or Fisher's Iris data set was introduced by the British statistician and biologist Ronald Fisher in his 1936 paper as an example of linear discriminant analysis. The data on length and width of petal and sepal leafs was actually collected by American botanist Edgar Anderson to quantify the morphologic variation of Iris flowers of three related species.
References
R. A. Fisher (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics 7(2):179–188.
150


Hierarchical Clustering - Orange
Linkage
Ward
Annotations
iris
Show labels only for subset
Color by: iris
Pruning
None
Max depth: 10
Selection
Manual
Height ratio: 26.1 %
Top N: 3
Zoom
Send Automatically
18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
C1
Iris-setosa
150×150 | -
150 | 150

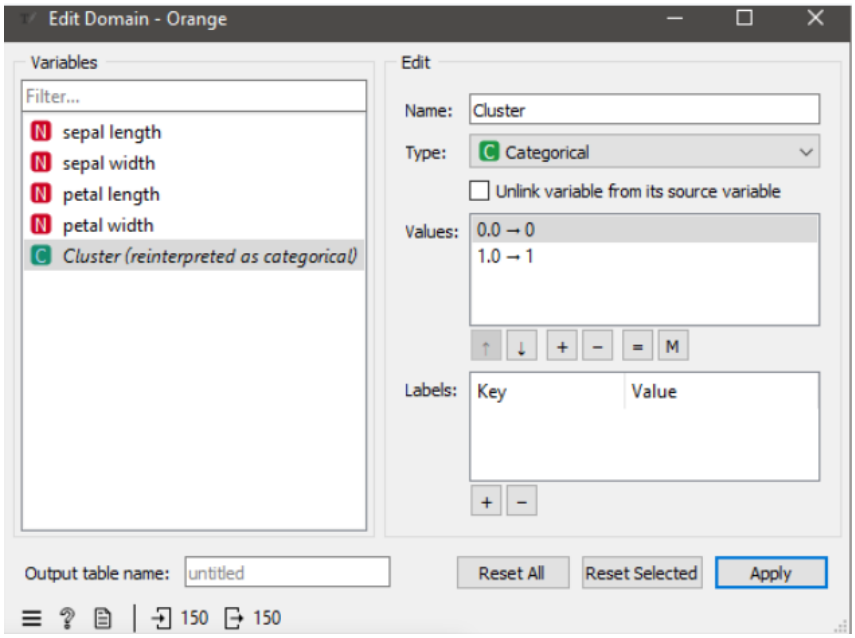

Scatter Plot - Orange
Axes
Axis x: petal length
Axis y: sepal length
Find Informative Projections
Attributes
Color: Cluster
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Jitter numeric values
Zoom/Select
Send Automatically
sepal length
petal length
C1
C2
C3
C4
150 | - | -
- | 150 | 2

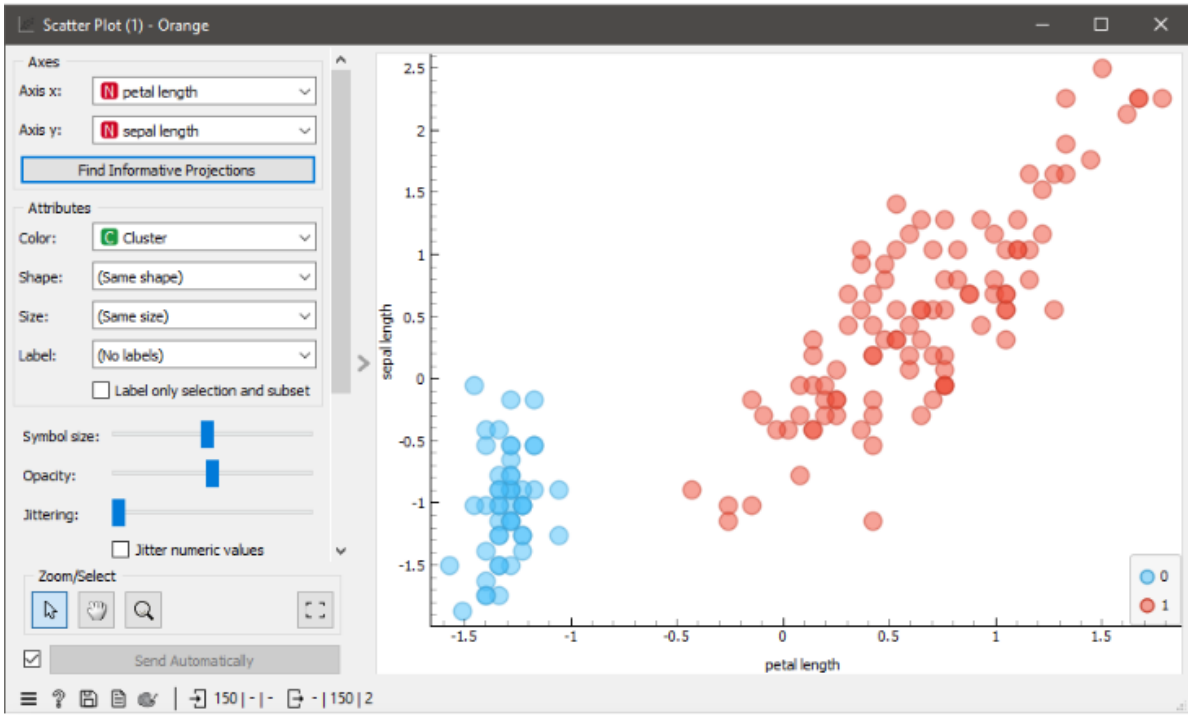

Datasets - Orange
Search for data set ...
Show data sets in
English
Title
Size
Instances
Variables
Target
Tags
Iris 4.5 KB 150 5 categorical biology
BBC3 2.6 MB 1407 3 categorical text, classification, news
Breast Cancer and Docetaxel Treatment 1.8 MB 24 9486 categorical biology
Smoking effect on B lymphocytes 1.8 MB 79 3001 categorical genomics
HDI 45.2 KB 188 54 none economy, geo
ParlaMint 1.7 MB 1000 18 categorical text, classification, time, politics
SentiNews 5.0 MB 2000 8 categorical text, sentiment
TKI resistance 1.2 MB 280 468 categorical spectral
Abalone 187.5 KB 4177 9 numeric biology
Adult Census Income 5.3 MB 48842 15 categorical economy, fairness
Ames Iowa Housing 831.2 KB 2930 81 numeric economy
Roman Amphorae 23.7 KB 164 16 categorical archaeology, image analytics
Description
Iris (1936), from UCI ML Repository
The Iris flower data set or Fisher's Iris data set was introduced by the British statistician and biologist Ronald Fisher in his 1936 paper as an example of linear discriminant analysis. The data on length and width of petal and sepal leafs was actually collected by American botanist Edgar Anderson to quantify the morphologic variation of Iris flowers of three related species.
References
R. A. Fisher (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics 7(2):179–188.
150


Python Script for divisive hierarchical clustering - Orange
Preferences
Vim mode
Library
Table from numpy
+
-
Update
More
Editor
def python_script(in_data):
1 import numpy as np
2 import pandas as pd
3 from sklearn.cluster import KMeans
4 from sklearn.preprocessing import StandardScaler
5 from Orange.data import Table, Domain, ContinuousVariable
6
7 # Get input data from Orange
8 data = in_data.X # Feature matrix
9 domain = in_data.domain # Original domain
10
11 # Standardize the data
12 scaler = StandardScaler()
13 data_scaled = scaler.fit_transform(data)
14
15 # Perform KMeans clustering
16 kmeans = KMeans(n_clusters=2, n_init=10)
17 labels = kmeans.fit_predict(data_scaled)
18
return out_data, out_learner, out_classifier, out_object
Console
Running script:
>>>
Running script:
>>>
Run
150 | - | -
150 | - | -



	sepal length	sepal width	petal length	petal width	Cluster
37	-0.416	1.032	-1.398	-1.313	0
38	-1.143	0.106	-1.284	-1.444	0
39	-1.749	-0.125	-1.398	-1.313	0
40	-0.901	0.801	-1.284	-1.313	0
41	-1.022	1.032	-1.398	-1.182	0
42	-1.628	-1.745	-1.398	-1.182	0
43	-1.749	0.338	-1.398	-1.313	0
44	-1.022	1.032	-1.228	-0.787	0
45	-0.901	1.726	-1.057	-1.050	0
46	-1.264	-0.125	-1.341	-1.182	0
47	-0.901	1.726	-1.228	-1.313	0
48	-1.507	0.338	-1.341	-1.313	0
49	-0.658	1.495	-1.284	-1.313	0
50	-1.022	0.569	-1.341	-1.313	0
51	1.402	0.338	0.535	0.265	1
52	0.675	0.338	0.422	0.396	1
53	1.280	0.106	0.649	0.396	1
54	-0.416	-1.745	0.137	0.133	1
55	0.796	-0.588	0.478	0.396	1
56	-0.174	-0.588	0.422	0.133	1
57	0.553	0.569	0.535	0.528	1
58	-1.143	-1.513	-0.261	-0.261	1
59	0.917	-0.356	0.478	0.133	1




Scatter Plot (1) - Orange
Axes
Axis x:
petal length
Axis y:
sepal length
Find Informative Projections
Attributes
Color:
Cluster
Shape:
(Same shape)
Size:
(Same size)
Label:
(No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Jitter numeric values
Zoom/Select
Send Automatically
150 | - | -
- | 150 | 2
sepal length
petal length
0
1