

# Getting Started

## Download and Install

[Download](#) Orange distribution package and run the installation file on your local computer. [Here](#) is a step-by-step installation guide, that we recommend you to follow.

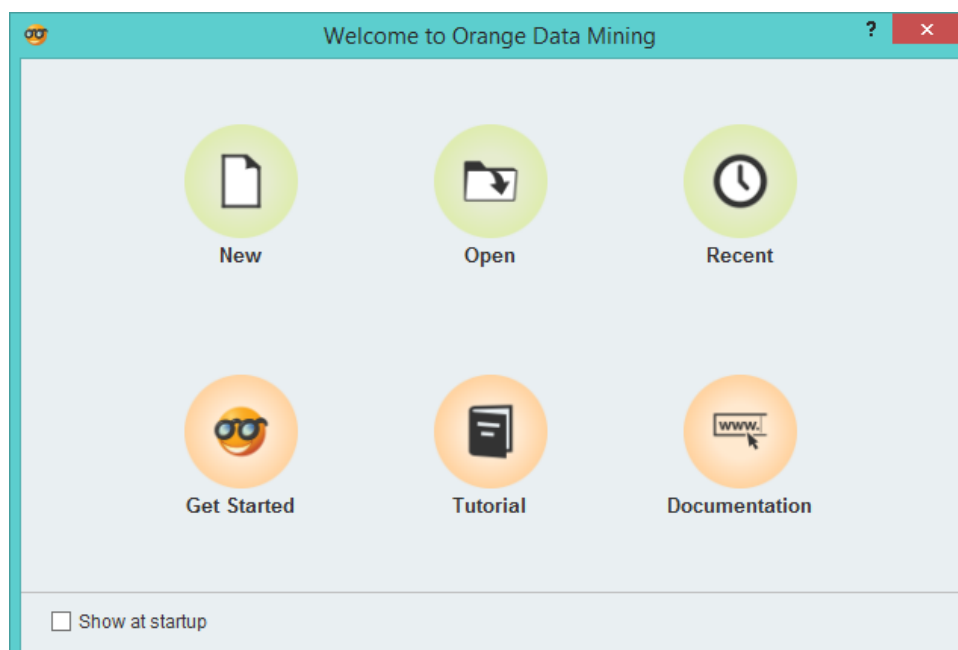
## Run

Locate Orange program icon. It is probably on your desktop (Win, Linux) or in the Applications folder (Mac). Double click on the icon to run Orange.



## Welcome to Orange

At the start, Orange opens a welcome screen. From here you can create new data mining workflows or browse through the ones you have already created. If you are running Orange for the first time, start by clicking on the Tutorial icon to browse through tutorial workflows.



# Tutorials

From the tutorials window, select any of the preloaded data mining workflows. Here, we will choose the one with hierarchical clustering.

The screenshot shows the 'Tutorials' window in Orange3. The main content area displays a workflow diagram for 'Hierarchical Clustering'. The workflow starts with a 'File' widget (brown icon) connected to a 'Data Table' widget (orange icon). The 'Data Table' widget is connected to a 'Distance' widget (blue icon with a triangle). The 'Distance' widget is connected to two other widgets: 'Distance Map' (blue icon with a heatmap) and 'Hierarchical Clustering' (blue icon with a dendrogram). The 'Hierarchical Clustering' widget is connected to a 'Data Table (!)' widget (orange icon). Red arrows point to each widget with descriptive text: 'Read the data. Try this schema with the brown-selected set (from data sets that come with Orange).' points to the File widget; 'Compute the distances between the data samples.' points to the Distance widget; 'Hierarchically cluster the data.' points to the Hierarchical Clustering widget; 'Visualize the data distances in a heat map.' points to the Distance Map widget. Text on the right side of the diagram says: 'Choose any part of the clustering dendrogram. Then, observe the selected data in a data table, or in any other analysis widget. Open both Hierarchical Clustering and Data Table (!) widget to turn this schema into interactive data analysis.'

**Hierarchical Clustering**

The workflow clusters the data items in Iris data set by first estimating the distances between data instances. Distance matrix is passed to hierarchical clustering, which renders the dendrogram. Select different parts of the dendrogram to further analyze the corresponding data.

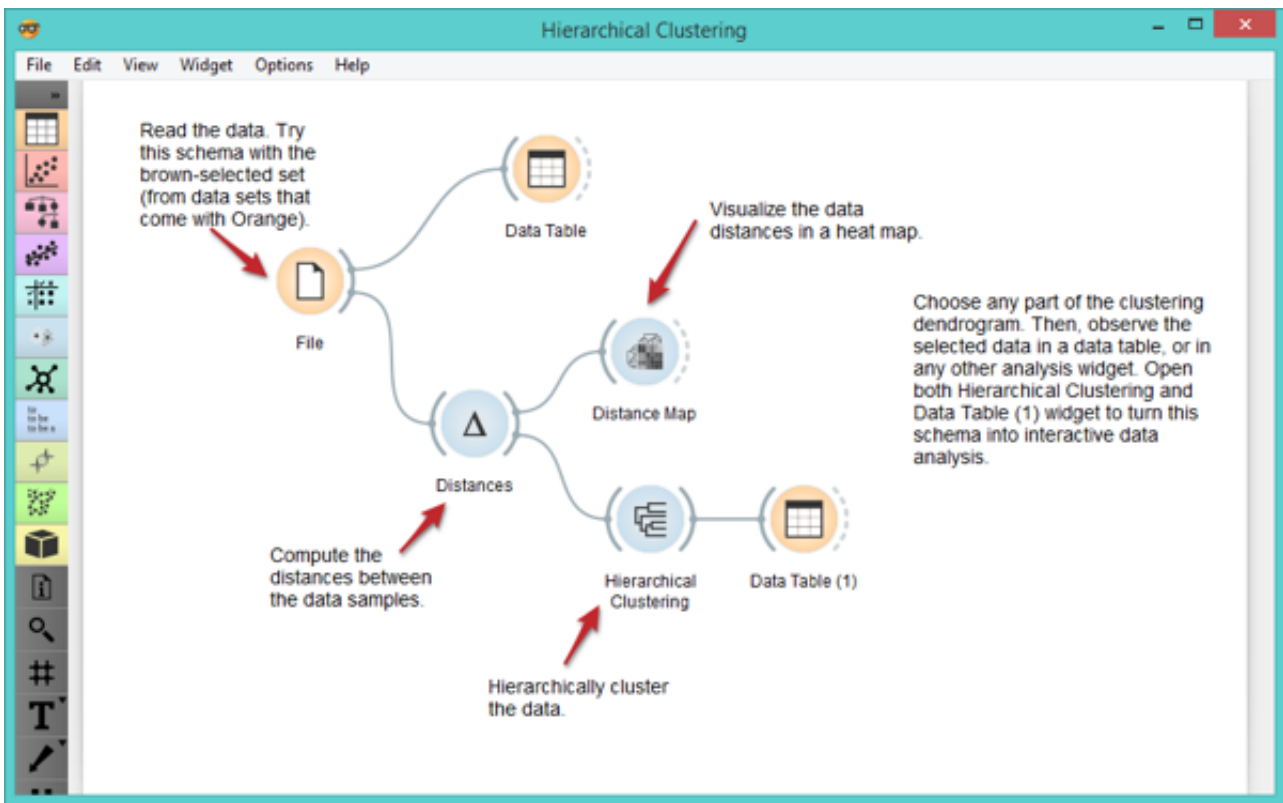
Notice that hierarchical clustering can only handle small data sets, that is, those that contain only a small number of data instances. For larger data sets the distance matrix may get too big and may not fit in the memory. An alternative method that can consider such data sets is k-means clustering.

**Path:** d:\orange\orange3\Orange\canvas\application\tutorials\310-clustering.ows

Classification Tree   Hierarchi...   Cross-Validation   Collective Matrix ...   Exploring the Lat...   Matrix

Open   Cancel

Selected tutorial will open in Orange canvas. In Orange, data mining workflows consist of computational components called widgets. Widgets do all the work and exchange information. They can communicate through channels. In the workflow below, the File widget sends its data to the Data Table widget and Distance widget, which, in turn, communicates the computed distances to two other widgets in the workflow.

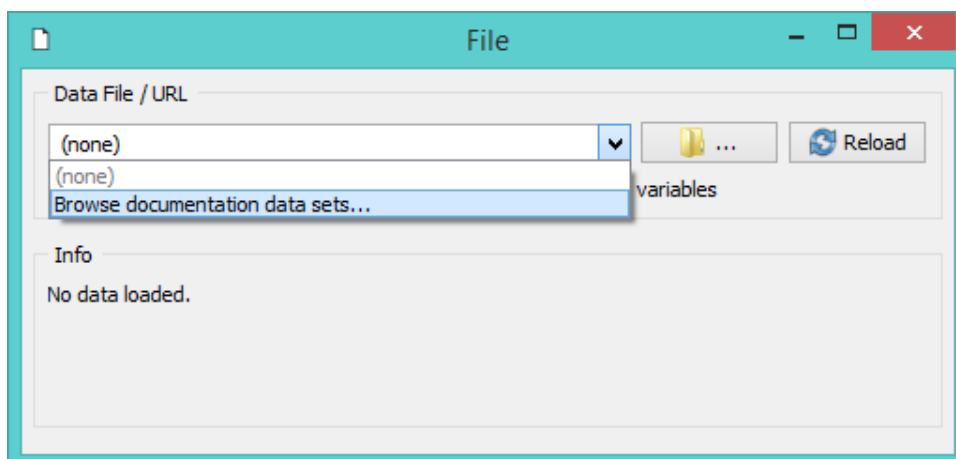


Any data mining starts with the data. In our hierarchical clustering schema, the File widget reads the data from the file on your computer and sends the data to other widgets.



File

Double click on the File widget icon to open it. Select "Browse documentation data sets..." and from the list of pre-installed data files chose *iris.tab*.



The File widget will now read the [the famous data set on 150 Iris flowers](#),

and send it to the workflow. The changes will propagate through the workflow updating its widgets. Close the window of the File widget and double click on the Data Table widget to open it. This displays the data that we have just read.

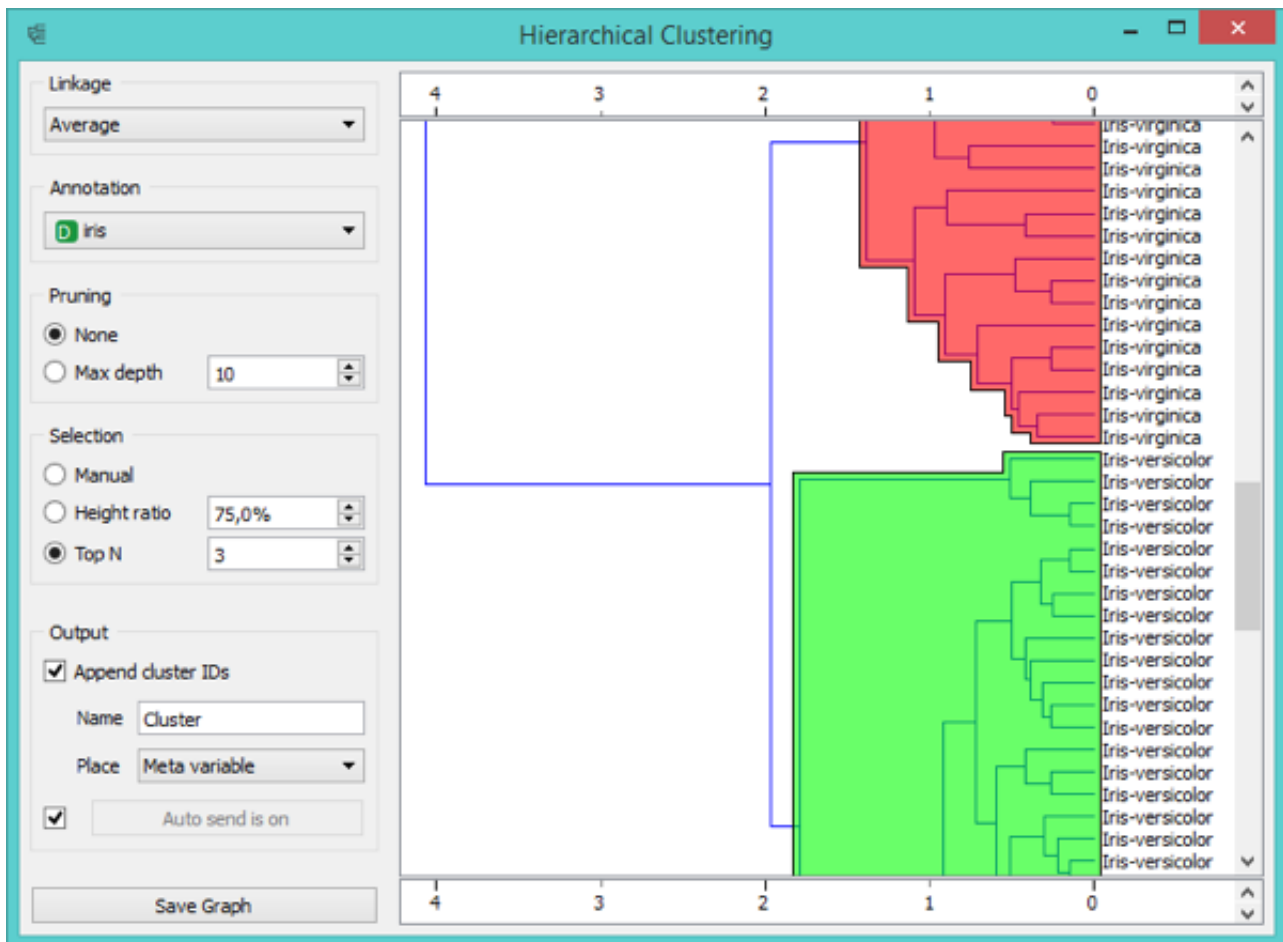
The screenshot shows a 'Data Table' widget with the following data:

	sepal length	sepal width	petal length	petal width	iris
1	5.100	3.500	1.400	0.200	Iris-setosa
2	4.900	3.000	1.400	0.200	Iris-setosa
3	4.700	3.200	1.300	0.200	Iris-setosa
4	4.600	3.100	1.500	0.200	Iris-setosa
5	5.000	3.600	1.400	0.200	Iris-setosa
6	5.400	3.900	1.700	0.400	Iris-setosa
7	4.600	3.400	1.400	0.300	Iris-setosa
8	5.000	3.400	1.500	0.200	Iris-setosa
9	4.400	2.900	1.400	0.200	Iris-setosa
10	4.900	3.100	1.500	0.100	Iris-setosa
11	5.400	3.700	1.500	0.200	Iris-setosa
12	4.800	3.400	1.600	0.200	Iris-setosa
13	4.800	3.000	1.400	0.100	Iris-setosa
14	4.300	3.000	1.100	0.100	Iris-setosa
15	5.800	4.000	1.200	0.200	Iris-setosa

The sidebar on the left contains the following sections:

- Info:** 150 instances (no missing values), 4 features (no missing values), Discrete class with 3 values (no missing values), No meta attributes. Includes a 'Restore Original Order' button.
- Variables:**  Show variable labels (if present),  Visualize continuous values,  Color by instance classes. Includes a 'Set colors' button.
- Selection:**  Select full rows.
- Auto send is on

Open and close other widgets to see what they do. In this workflow, the most interesting widget is Hierarchical Clustering that displays clustering results. Scroll through the dendrogram - the tree-based rendering of the clustering - to check if the algorithm correctly identified the three species of Iris.

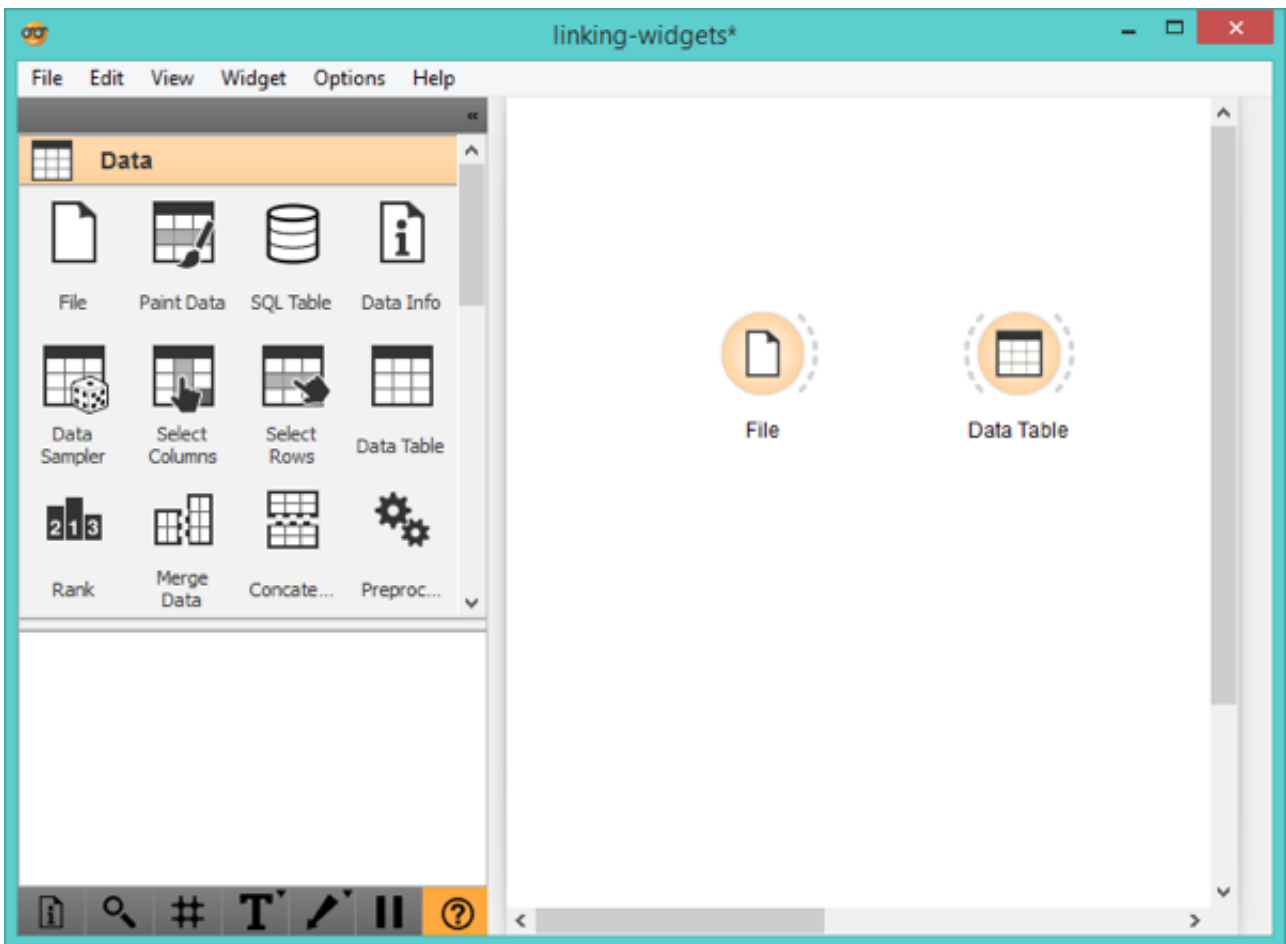


You may now open other tutorials (from the Help menu choose Tutorials). Or create a workflow of your own.

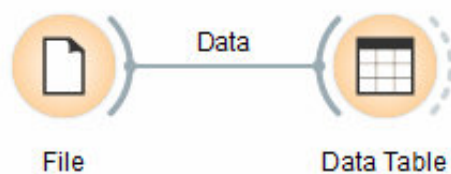
## Your Own Workflow

We first need to start with an empty canvas. Click on New in Orange's welcome screen, or, if Orange is already running, choose New from the File menu.

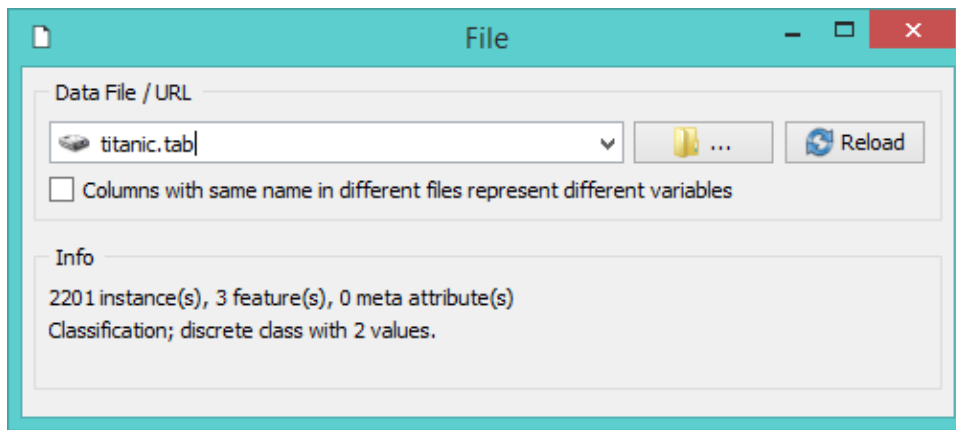
We will explore the data on passengers of the HMS Titanic and develop a model to predict the probability of survival based on the passenger's traveling class, gender and age. Let us start by placing the File and Data Table widgets on the canvas.



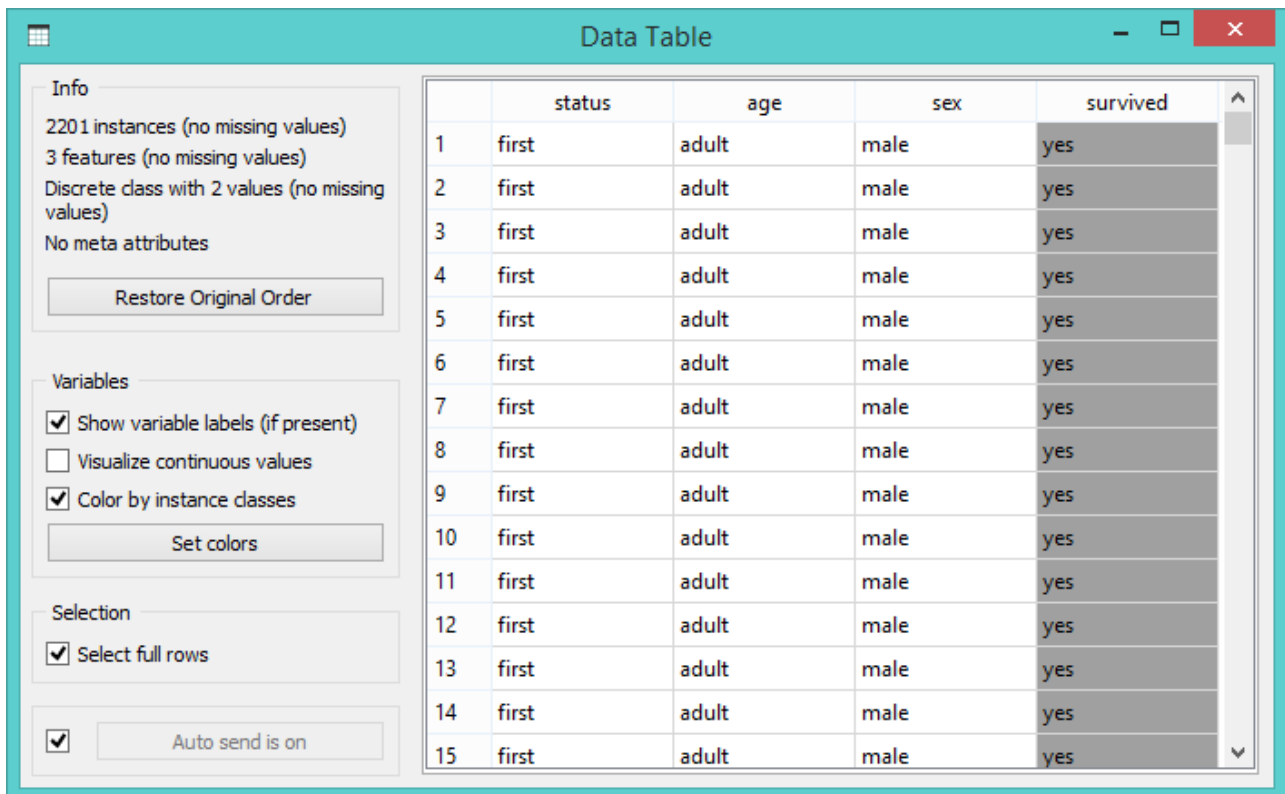
We would like the File widget to read the data and send it to the Data Table for inspection. We need to connect these two widgets to establish a communication between them. Click on the dashed line beside the File widget and drag the line to the Data Table.



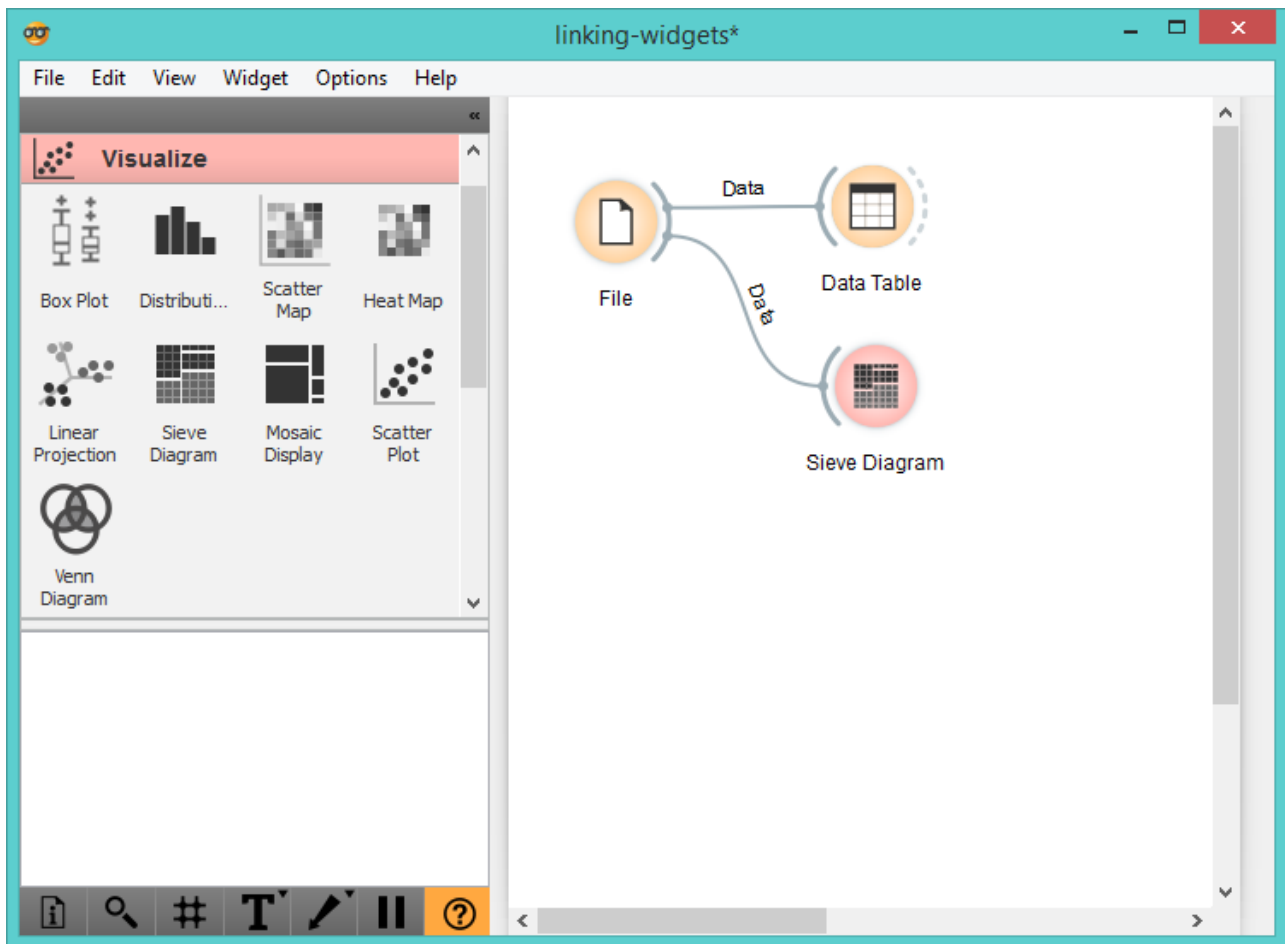
To load the data, open the File widget (double click on its icon), select "Browse documentation data sets" from the Data File box and choose `titanic.tab`.



The widget automatically transferred the loaded data to all the connected widgets. Check this by opening the Data Table widget.

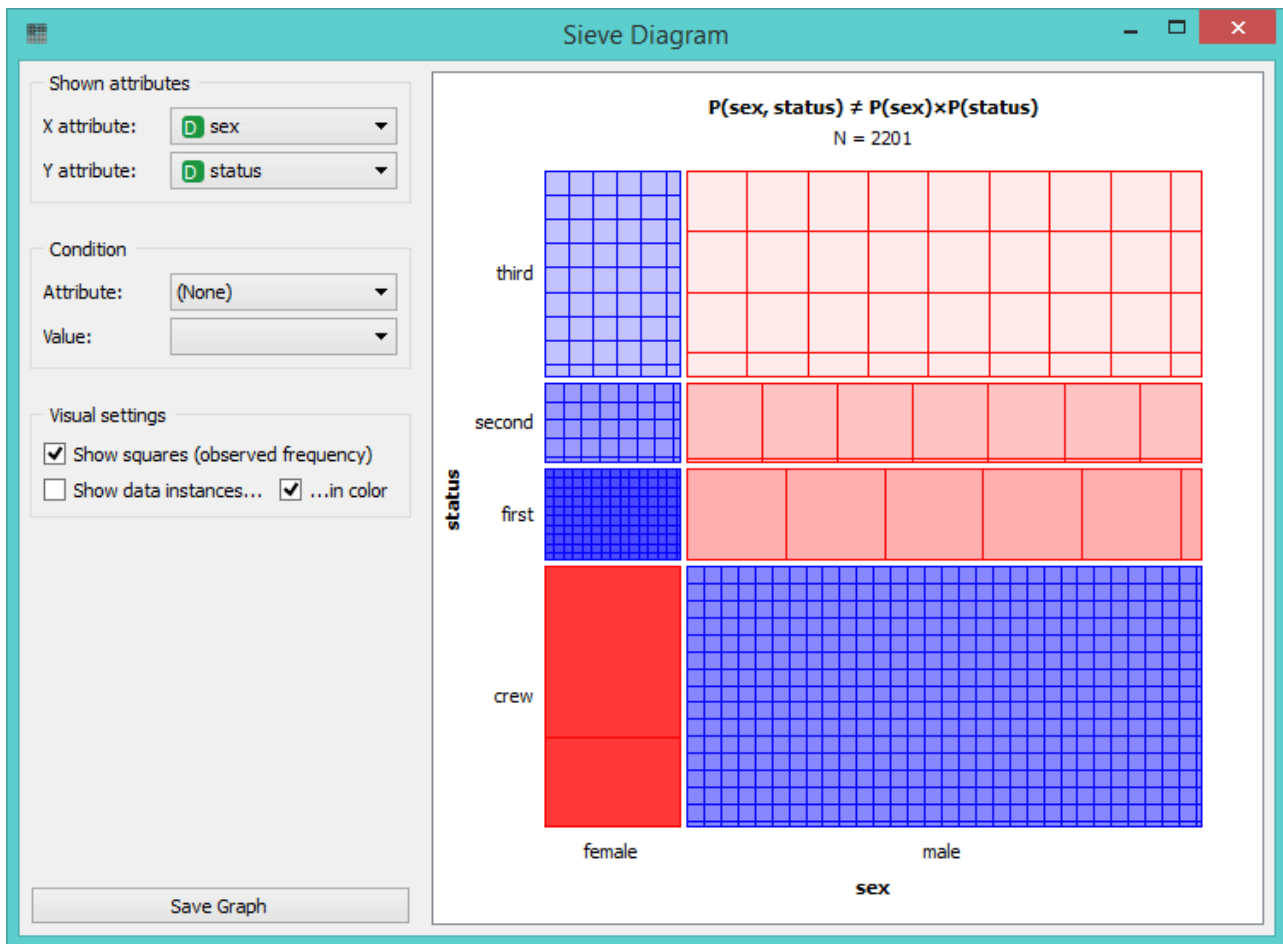


Our aim is to inspect survival probabilities for the passengers of Titanic by age, sex and status. Place Sieve Diagram on the canvas and connect it to the File widget.



Double click on the Sieve Diagram widget to visualize actual survival probabilities against expected ones. Play with attribute combinations to get the best visualization. Here's a hint: sex and status will give you the most interesting results.





The lowest survival probability is estimated for adult males traveling in the third class and the highest for females from the first class. How about the crew? Who had the highest probability of survival? Hover over the diagram to see the information.

You have now learned how to place widgets on the canvas, connect them to make workflows, read the data and visualize it. Consider exploring other widgets and their combinations, or load some data of your own and see how Orange can help you in the analysis.