

AND COLLABORATORS

IG QUASAR

Copyright © 2020 Biolab and Collaborators

PUBLISHED BY BIOLAB

TUFTE-LATEX.GOOGLECODE.COM

Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at <http://www.apache.org/licenses/LICENSE-2.0>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

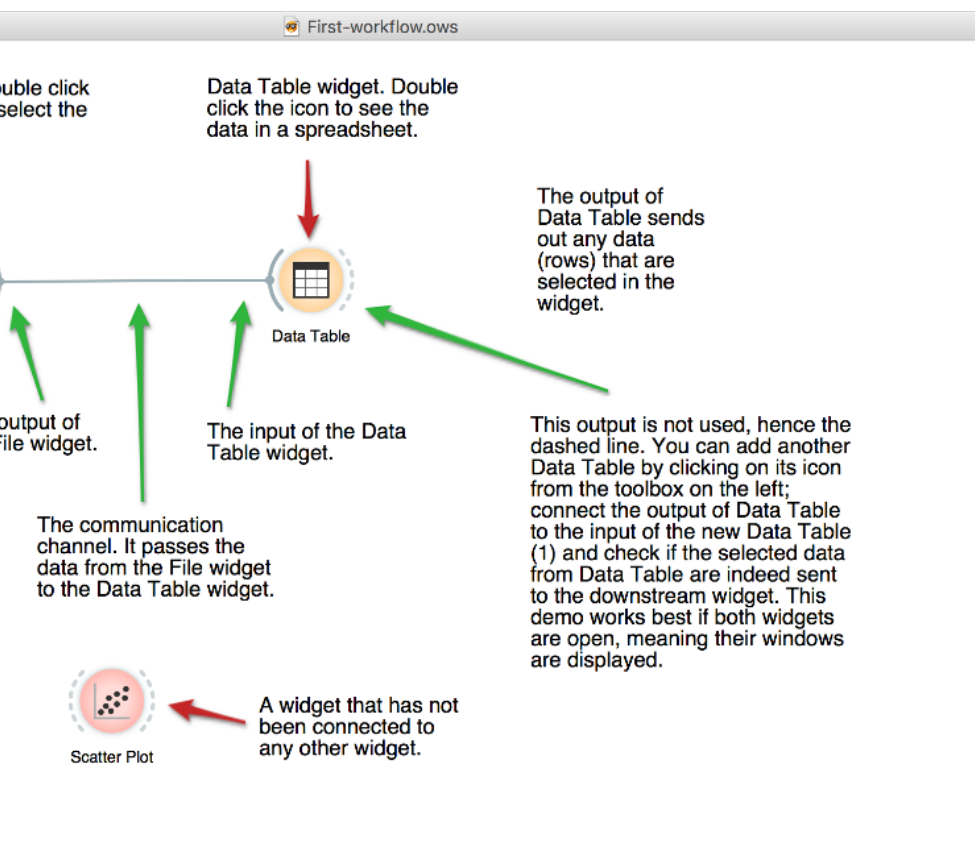
*First printing, November 2020*

# *Contents*



# Workflows in Quasar

QUASAR WORKFLOWS consist of components that read, process and visualize data. We call them “widgets”. Widgets are placed on a drawing board (the “canvas”). Widgets communicate by sending information along a communication channel. Output from one widget is used as input to another.

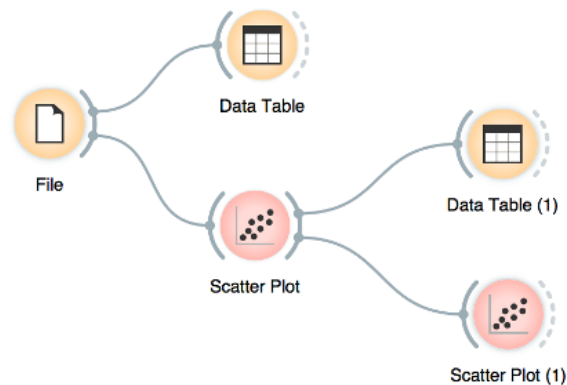


two connected  
without con-  
a widget ap-  
ne inputs ap-

We construct workflows by dragging widgets onto the canvas and connecting them by drawing a line from the transmitting widget to the receiving widget. The widget's outputs are on the right and the inputs on the left. In the workflow above, the File widget sends data to the Data Table widget.

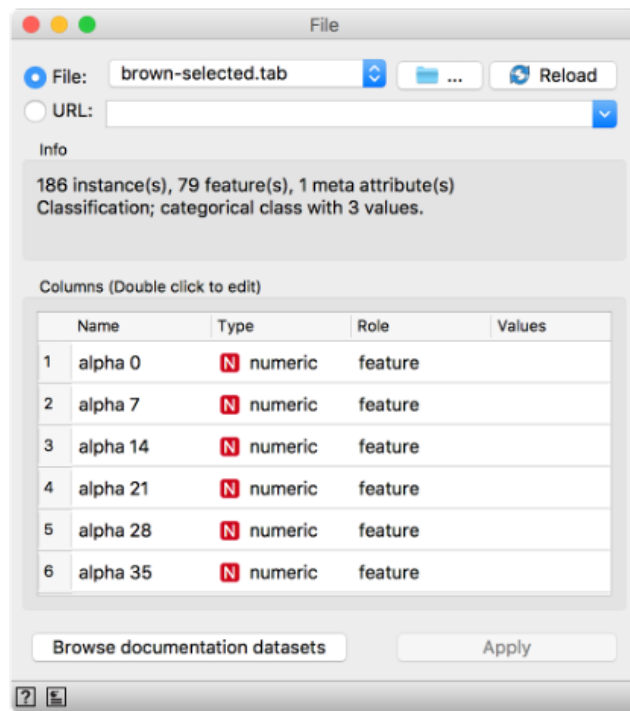
Start by constructing a workflow that consists of a File widget, two Scatter Plot widgets and two Data Table widgets:

Workflow with a File widget that reads data from disk and sends it to the Scatter Plot and Data Table widget. The Data Table renders the data in a spreadsheet, while the Scatter Plot visualizes it. Selected data points from the plot are sent to two other widgets: Data Table (1) and Scatter Plot (1).



The File widget reads data from your local disk. Open the File widget by double clicking its icon. Quasar comes with several pre-loaded data sets. From these (“Browse documentation data sets...”), choose *brown-selected.tab*, a yeast gene expression data set.

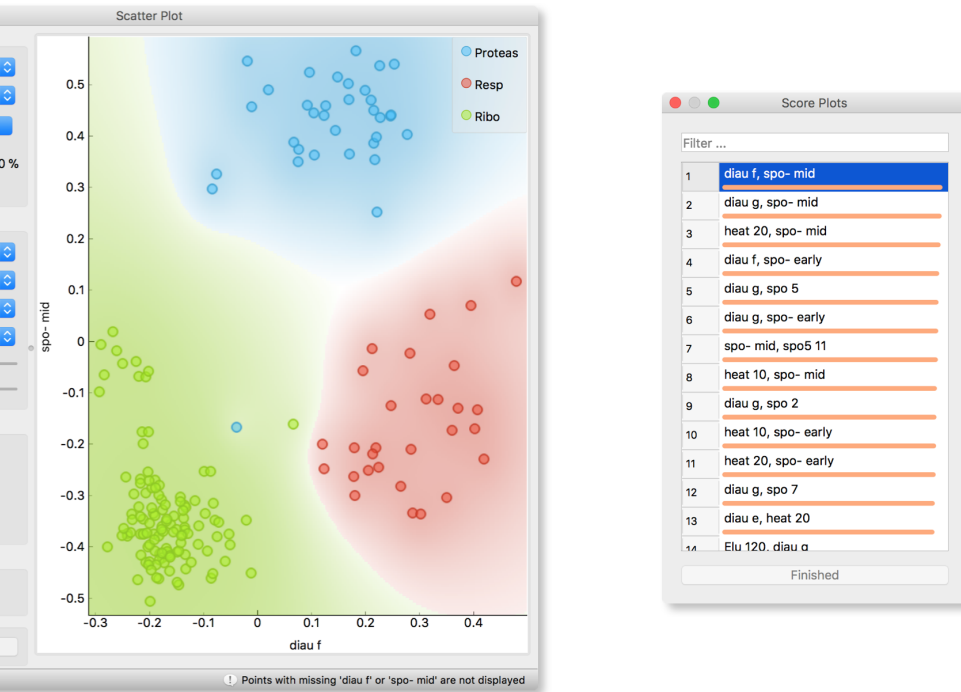
Quasar workflows often start with a File widget. The brown-selected data set comprises of a 186 rows (genes) and 81 columns. Out of the 81 columns, 79 contain gene expressions of baker’s yeast under various conditions, one column (marked as a “meta attribute”) provides gene names, and one column contains the “class” value or gene function.



After you load the data, open the other widgets. In the Scatter Plot widget, select a few data points and watch as they appear in widget Data Table (1). Use a combination of two Scatter Plot widgets, where the second scatter plot shows a detail from a smaller region selected in the first scatter plot.

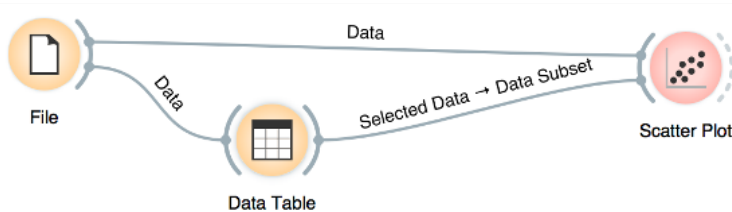
The following is more of a side note, but it won’t hurt. Namely, the scatter plot for a pair of random features does not provide much information on gene function. Does this change with a different choice

of feature pairs in the visualization? Rank projections (the button on the top left of the Scatter Plot widget) can help you find a good feature pair. How do you think this works? Could the suggested pairs of features be useful to a biologist?

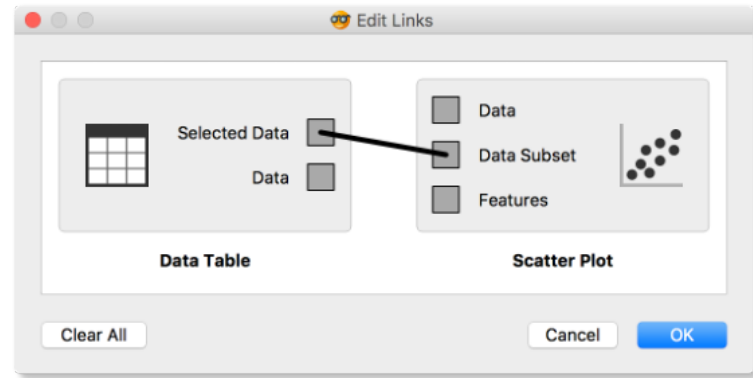


We can connect the output of the Data Table widget to the Scatter Plot widget to highlight the chosen data instances (rows) in the scatter plot.

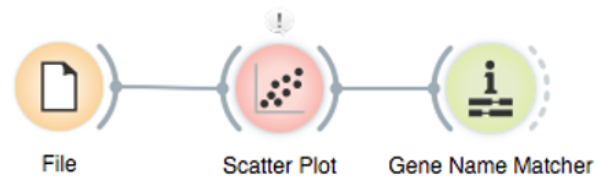
switched on  
el names be-  
ferences.



How does Quasar distinguish between the primary data source and the data selection? It uses the first connected signal as the entire data set and the second one as its subset. To make changes or to check what is happening under the hood, double click on the line connecting the two widgets.



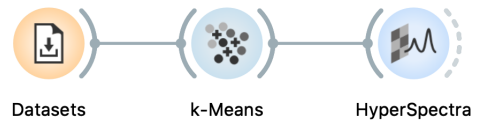
The rows in the data set we are exploring in this lesson are gene profiles. We could perhaps use widgets from the Bioinformatics add-on to get more information on the genes we selected in any of the Quasar widgets.



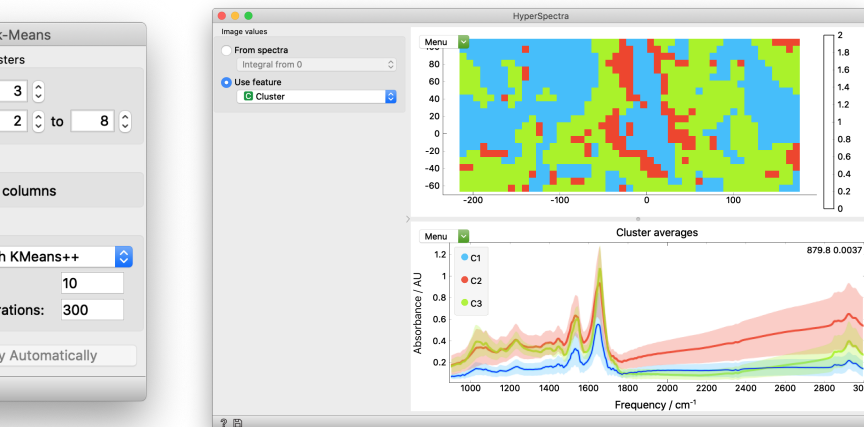
Quasar comes with a basic set of widgets for data input, preprocessing, visualization and modeling. For other tasks, like text mining, network analysis, and bioinformatics, there are add-ons. Check them out by selecting Add-ons... from the Options menu.



# Clustering Spectral Images

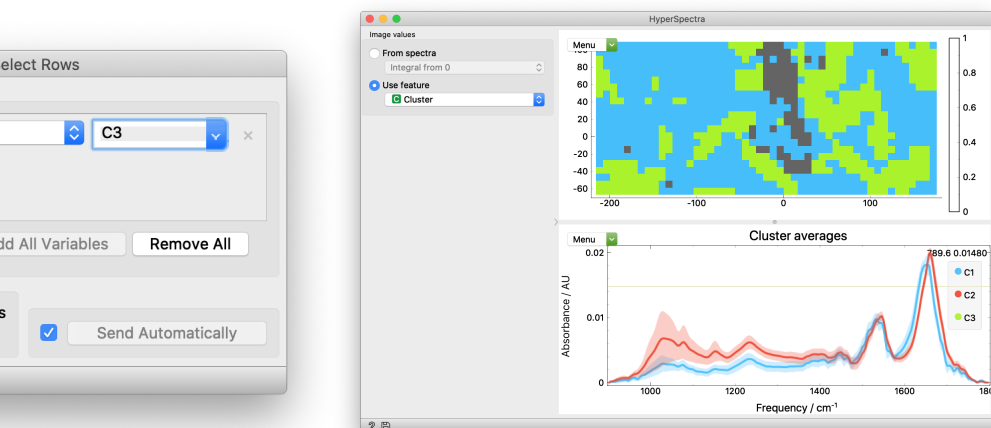
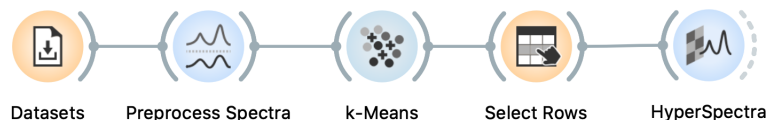


We have already seen hierarchical clustering. Another clustering algorithm, k-Means, is much faster for data with lots of rows, like images, which contain a row (a spectrum) for each pixel. Still, for the liver-cirrhosis data, both approaches are fast. Here, we use *k-Means* with  $k=3$  clusters.



s wrong pre-  
.

We see no meaningful clusters. Therefore, we need to preprocess the data. If we do it well, we see that a cluster corresponds to the background. We could remove it with the *Select Rows* widget.



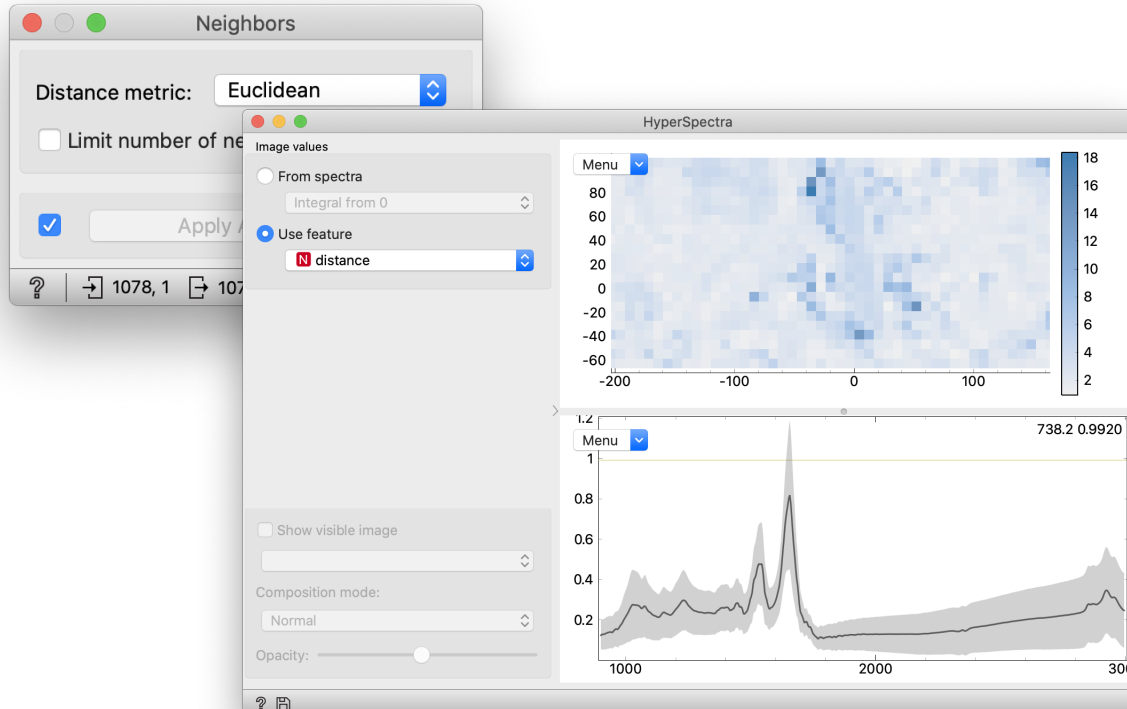
s wrong pre-  
.

## Visualize spectral distances

LET'S CONSIDER A SPECTRUM, or any other data entry for that matter, as a point in a multidimensional space. We can define distance metrics between these points and visualize the distance values from one another or from a selected reference point or reference spectrum. By doing so, we can explore how similar our measurements are to a selected reference. We can do this on a series of spectra or even on hyperspectral maps!

If you build the workflow shown above this paragraph in Quasar you will be able to explore various distance metrics. First, let's use the '*Liver cirrhosis - spectral image*' dataset provided by the *Datasets* widget, then calculate the *Euclidean distances* from the average spectrum with the *Neighbors* widget and visualize them in *Hyperspectra*.

Can you reproduce the results below? Pay attention to the color scheme.



Explore different distance metrics, inspect distances in a *Data Table* widget. Don't forget, you can select points on the top map and see the corresponding spectra on the bottom in *Hyperspectra*.

**Lecturer note.** Possibility for discussion of the general mathematical properties of distance functions. See [https://en.wikipedia.org/wiki/Metric\\_\(mathematics\)](https://en.wikipedia.org/wiki/Metric_(mathematics))

