НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Internet Studies Lab, Department of Applied Mathematics and Business Informatics

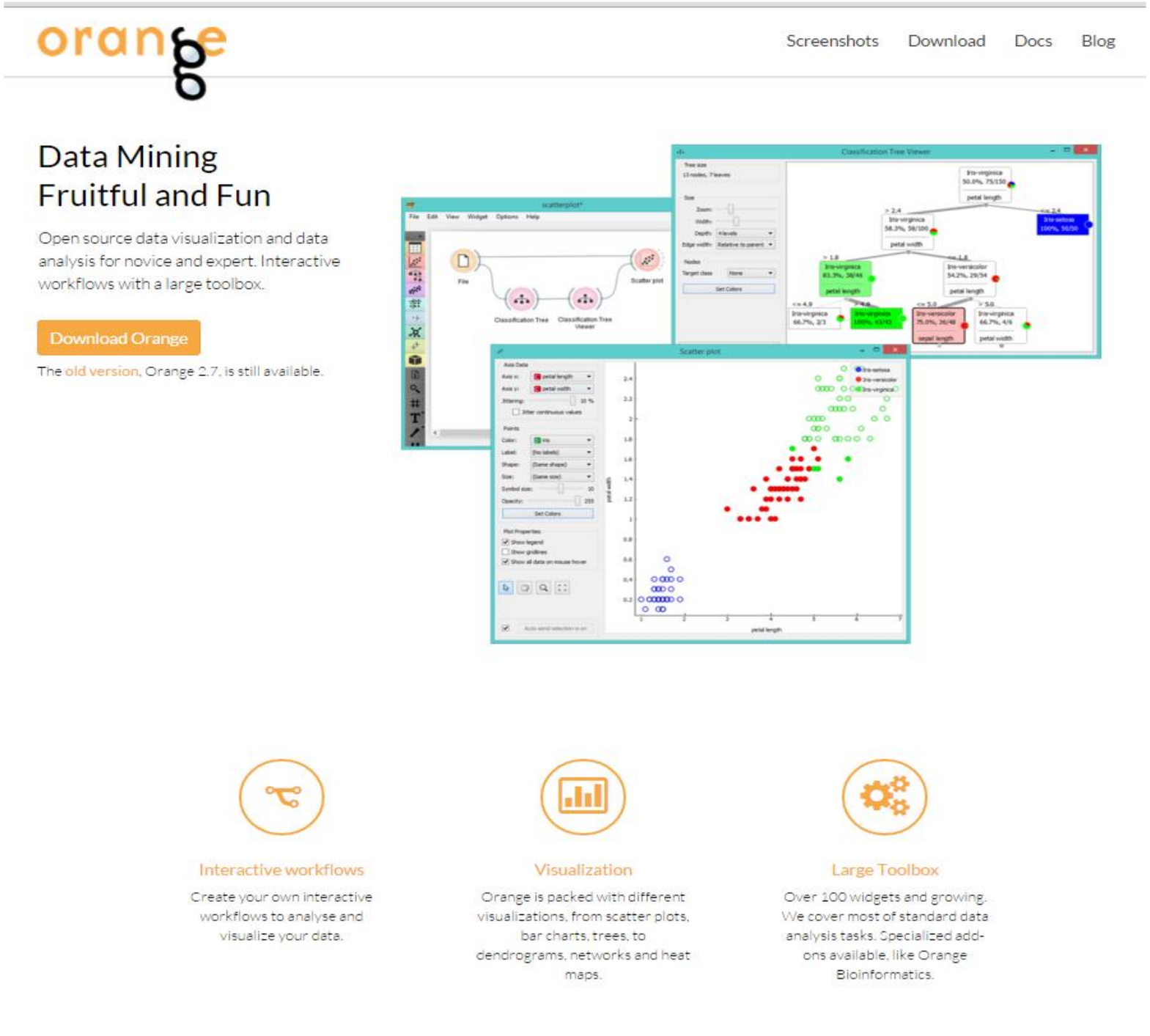
INTRODUCTION TO ORANGE SOFTWARE

Анализ баз данных в публичном управлении

Кольцов С.Н.

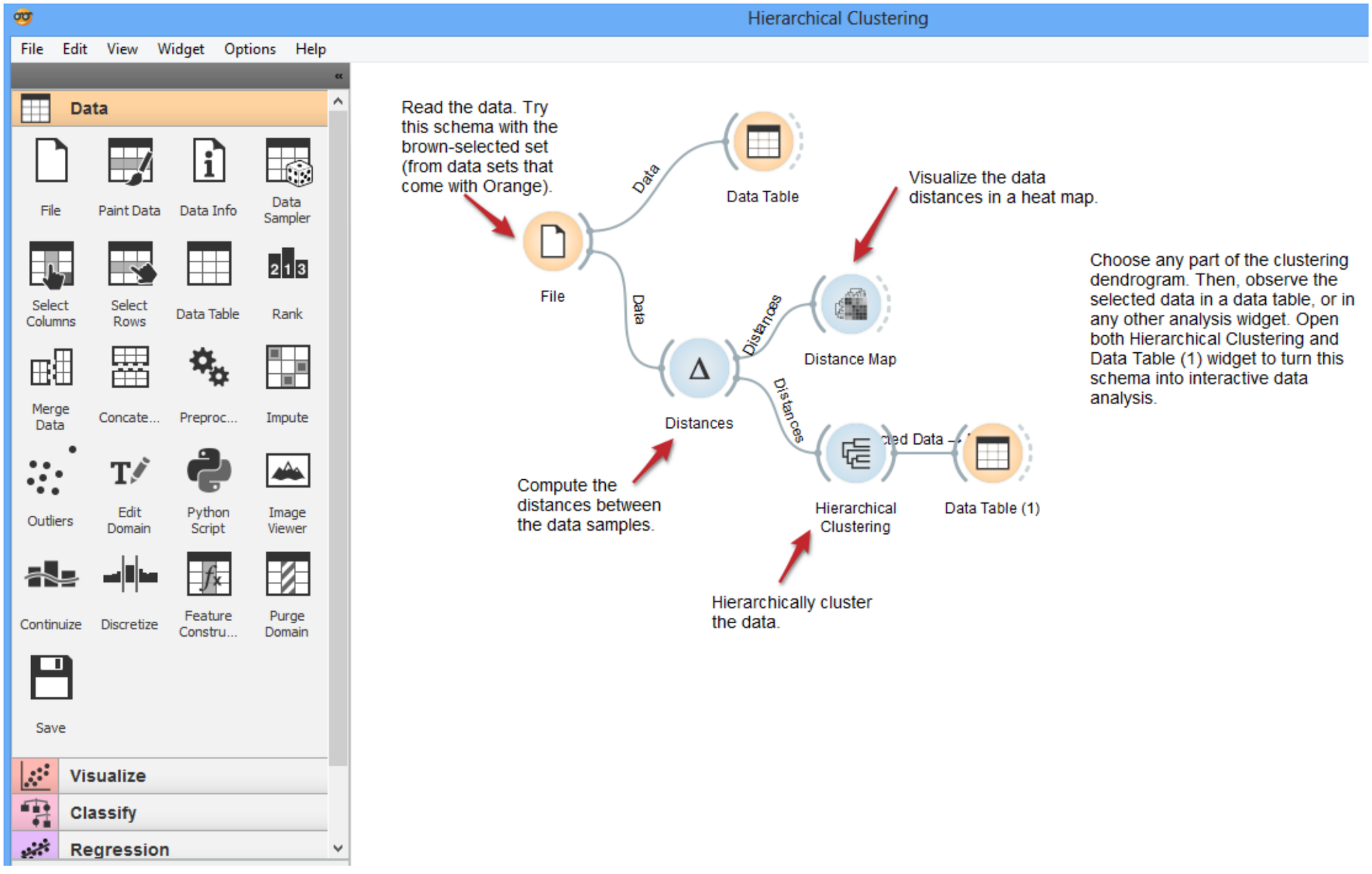
Saint Petersburg, 07.09.2018

ORANGE DATA MINING

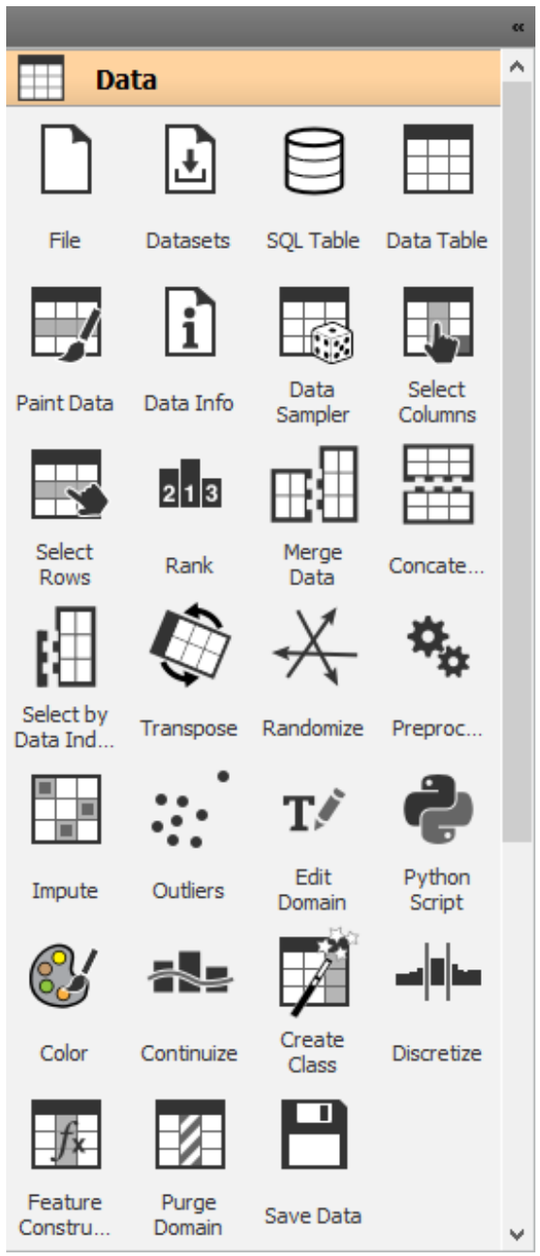


Можно скачать по адресу: http://orange.biolab.si/

ORANGE CONSISTS OF A CANVAS INTERFACE ONTO WHICH THE USER PLACES WIDGETS AND CREATES A DATA ANALYSIS WORKFLOW



ORANGE CONSISTS OF A CANVAS INTERFACE ONTO WHICH THE USER PLACES WIDGETS AND CREATES A DATA ANALYSIS WORKFLOW



Data: widgets for data input, data filtering, sampling, imputation, feature manipulation and feature selection

Widgets ответственный за открытие файлов с данными.

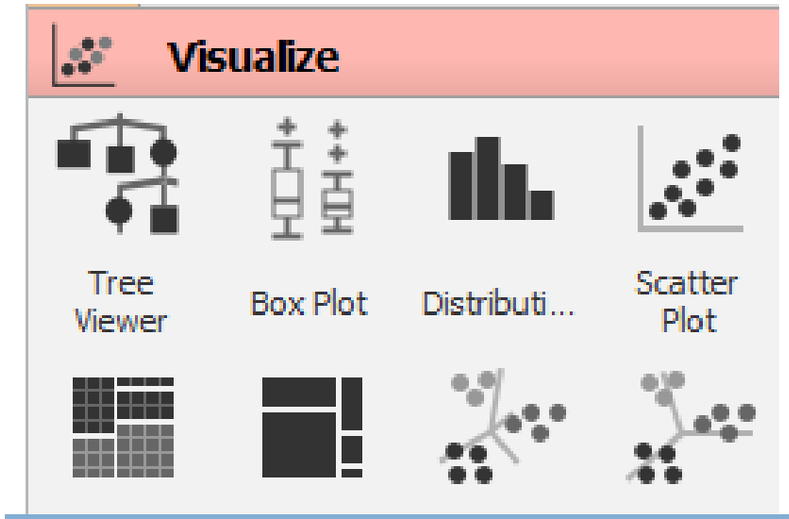
Widgets ответственный за визуализацию данных

Widgets позволяет выбирать колонки для анализа

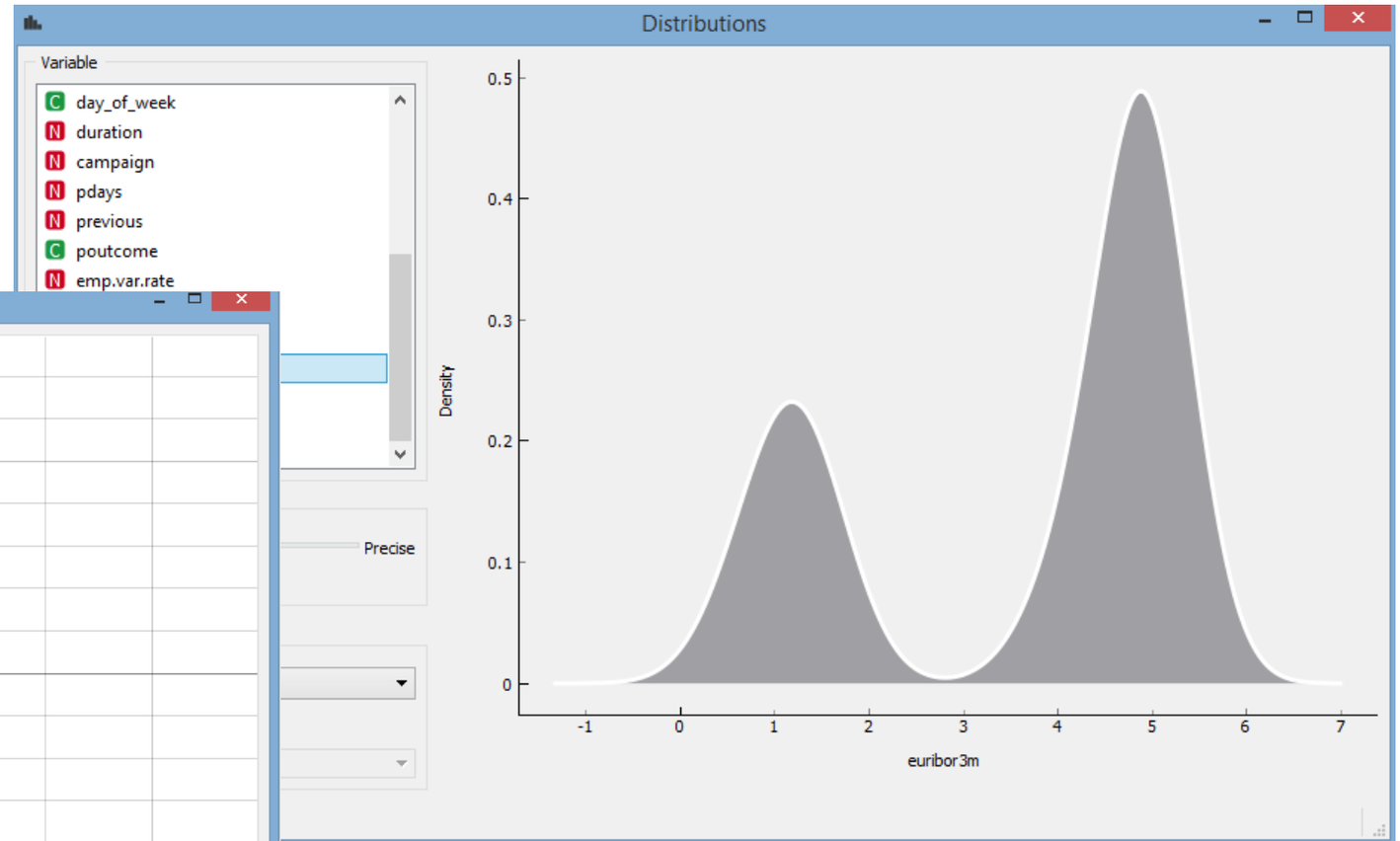
Widgets позволяет выбирать строки для анализа

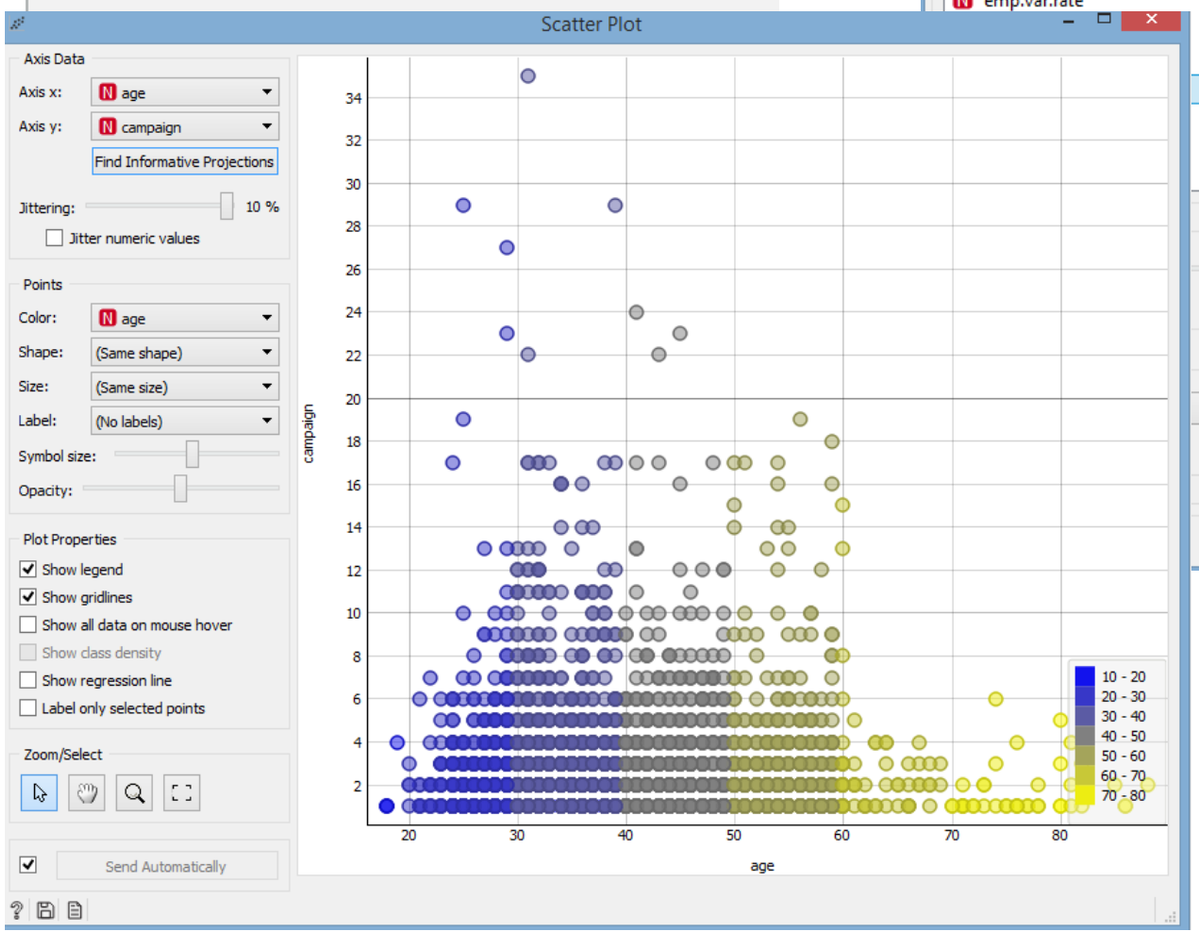
Widgets позволяет выбирать датасеты из предустановленного набора

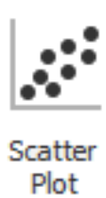
ORANGE : VISUALIZE WIDGETS



Widgets ответственный за рисование распределений по данным

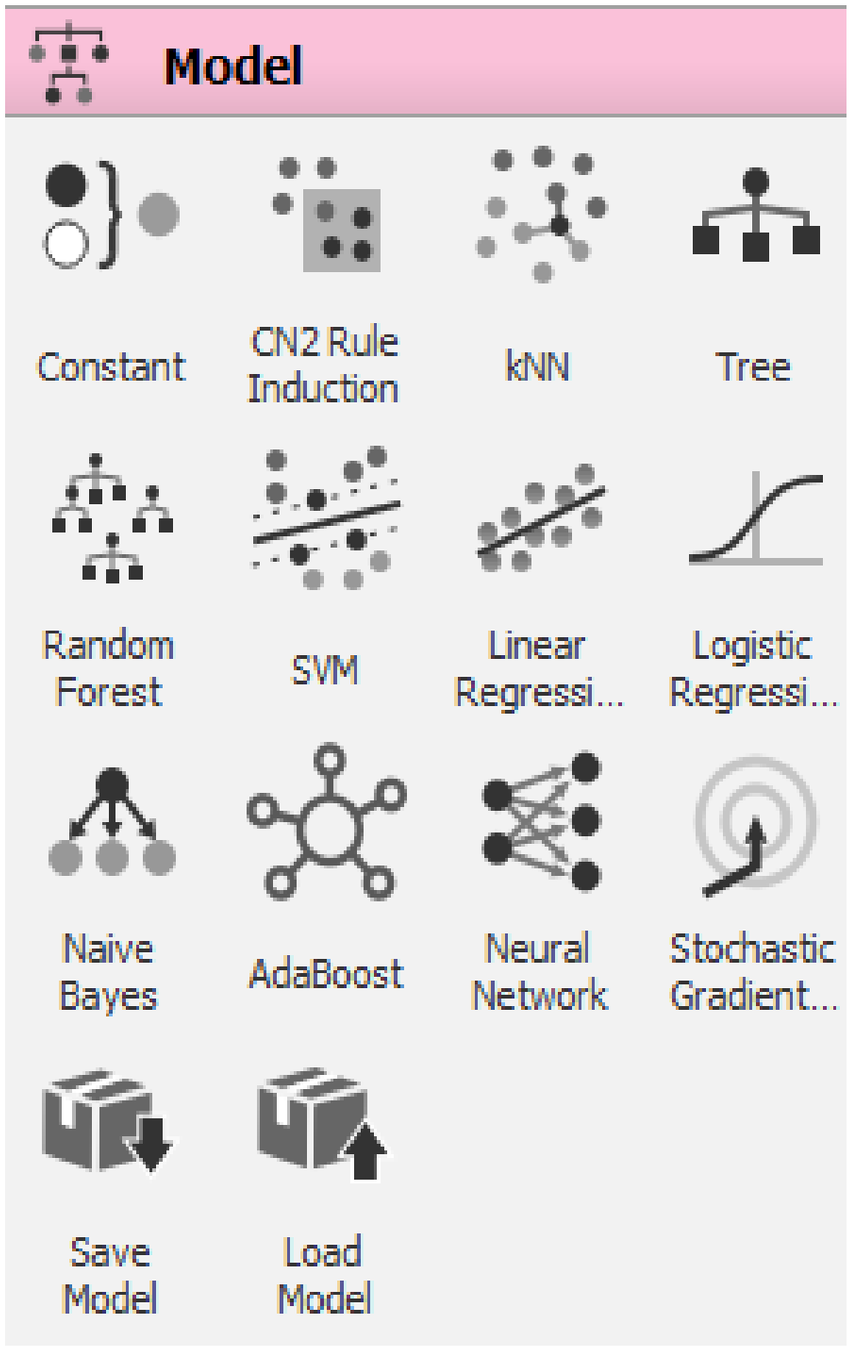






Widgets ответственный за рисование распределений по данным

ORANGE : MODELS



Naive Bayes

Widgets отвечающий за наивный **байесовский классификатор**.

SVM

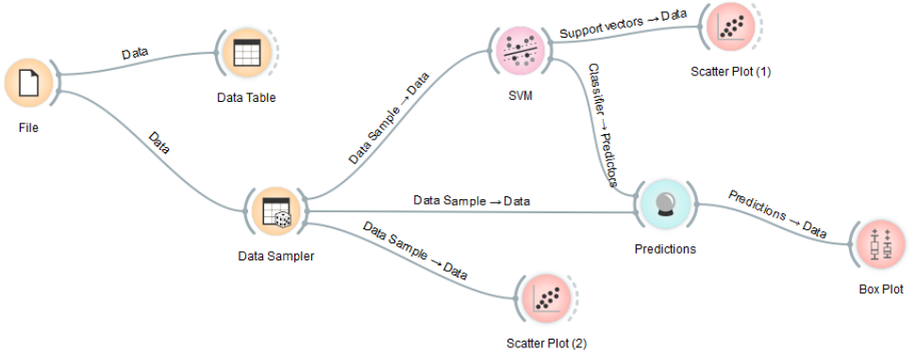
Widgets осуществляющий классификацию на основе **Метода опорных векторов.**

Logistic Regressi...

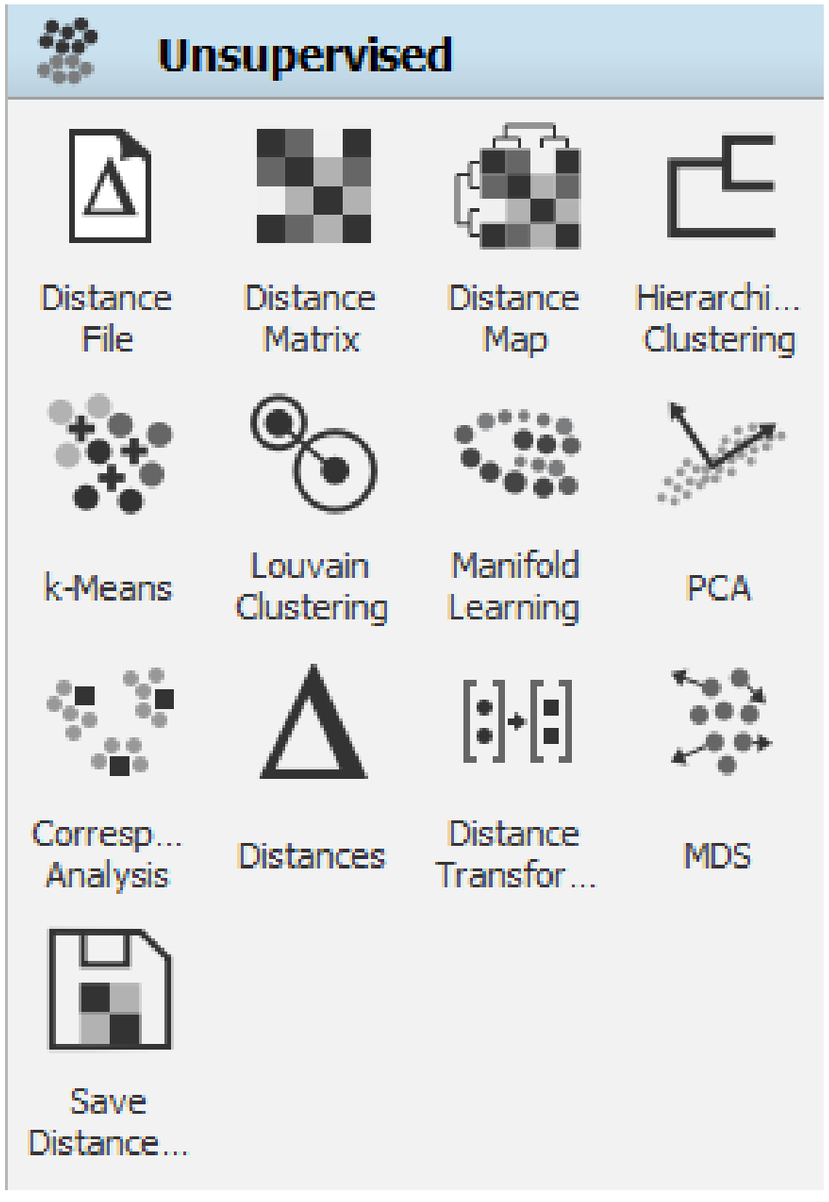
Widgets осуществляющий классификацию на основе **логистической регресии.**

Nearest Neighbors

Widgets производящий классификацию по **методу ближайших соседей.**

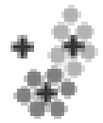


ORANGE : UNSUPERVISED



Hierarchi... Clustering

Widgets отвечающий за **иерархическую кластеризацию.**

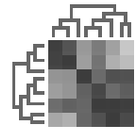


k-Means

Widgets позволяющий проводить кластеризацию **методом K-Means**.

PCA

Widgets позволяющий проводить **principal component analysis**



Distance Map

Widgets визуализирующий матрицу расстояний.

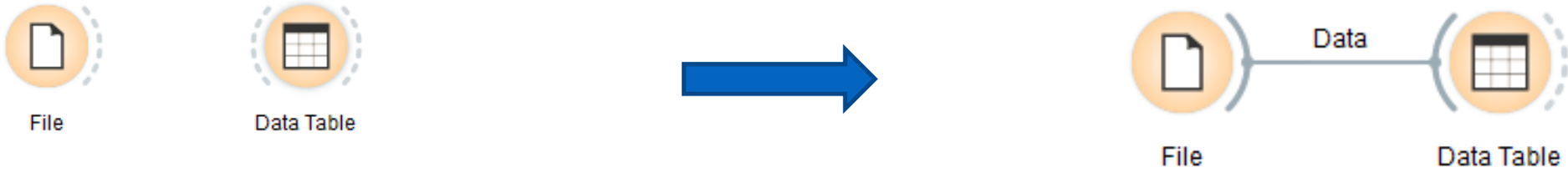
Louvain Clustering

Widgets позволяющий провести кластеризацию сети при помощи луванского алгоритма.

ORANGE : КАК ЗАГРУЗИТЬ И ПРОСМОТРЕТЬ ДАННЫЕ

1. Запускаем Orange и создаем пустой проект.
2. Открываем опцию ‘Data’, и кликаем на widget ‘File’
3. Выбираем widget ‘Data Table’.

Соединяем ‘File’
‘Data Table’ линией.



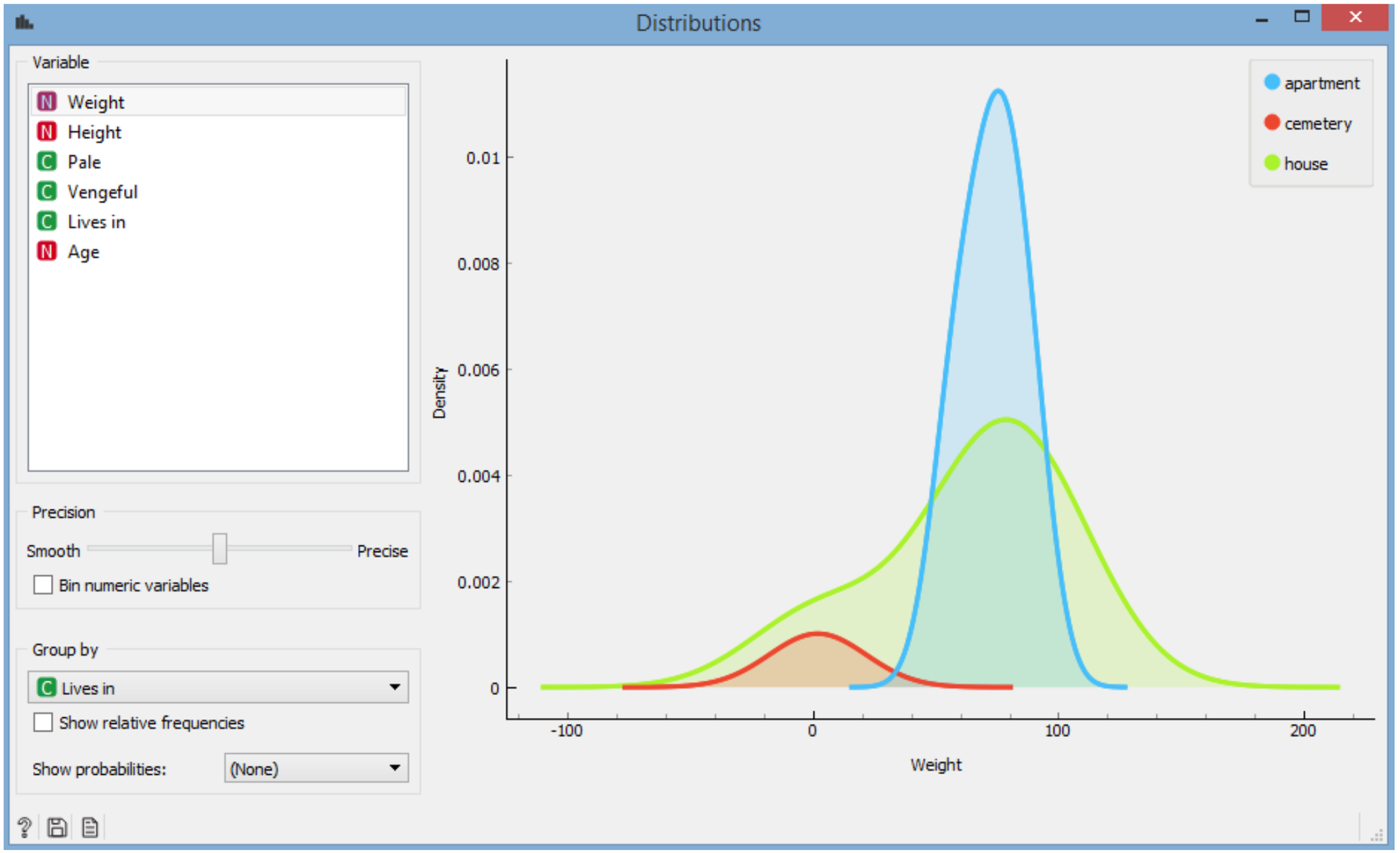
widgets ‘File’ – отвечает за то какой файл нужно загрузить. Кликаем на нем и указываем имя файла, например, **Paranormal distribution.csv.** Если все нормально, то что бы посмотреть что находится в этом файле нужно кликнуть на widgets ‘Data Table’. В результате получим следующее:

	Weight	Height	Pale	Vengeful	Lives in	Age	Name
1	55.000	167.000	yes	no	house	45.000	Anne
2	89.000	180.000	no	no	apartment	19.000	Jack
3	78.000	175.000	yes	yes	apartment	43.000	Mark
4	2.000	166.000	yes	yes	house	50.000	Ghoul
5	70.000	150.000	yes	yes	house	32.000	Marie
6	102.000	190.000	no	no	house	44.000	Clark
7	60.000	155.000	yes	no	apartment	61.000	Rebecca
8	2.000	190.000	yes	yes	cemetery	120.000	Spirit
9	79.000	161.000	no	yes	apartment	21.000	Helen
10	81.000	181.000	no	yes	house	37.000	Nick
11	63.000	171.000	no	no	apartment	35.000	Sarah
12	92.000	178.000	yes	no	house	41.000	Oliver
13	78.000	168.000	yes	no	house	28.000	Valerie
14	2.000	171.000	yes	yes	house	92.000	Phantom
15	72.000	185.000	yes	yes	apartment	58.000	Bill
16	54.000	153.000	yes	no	apartment	18.000	Claire
17	90.000	178.000	no	no	house	47.000	Ian

ORANGE : КАК ПОСТРОИТЬ РАСПРЕДЕЛЕНИЯ ДАННЫХ НА ГРАФИКЕ

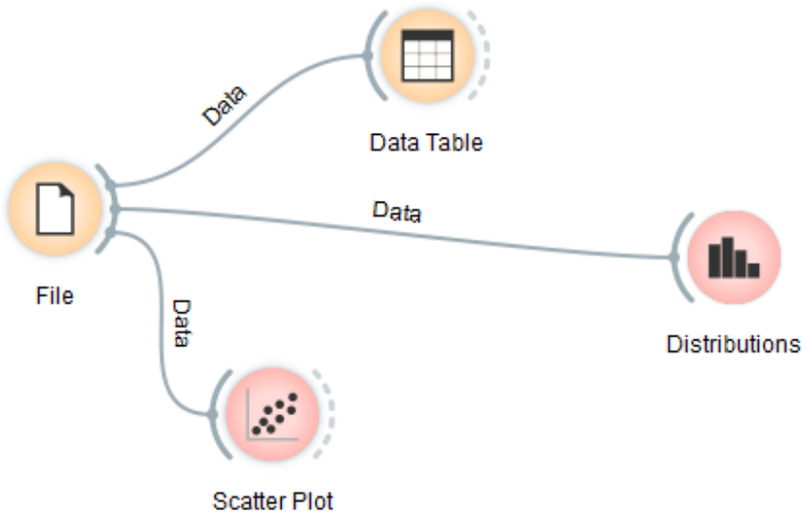
К существующему проекту добавляем widgets ‘Distribution’. Он находится в опции ‘Visualize’. Соединяем widget ‘File’ с widget ‘Distribution’.

Можно объединить несколько графиков на графике. На данном примере показаны распределения веса в зависимости от места проживания.

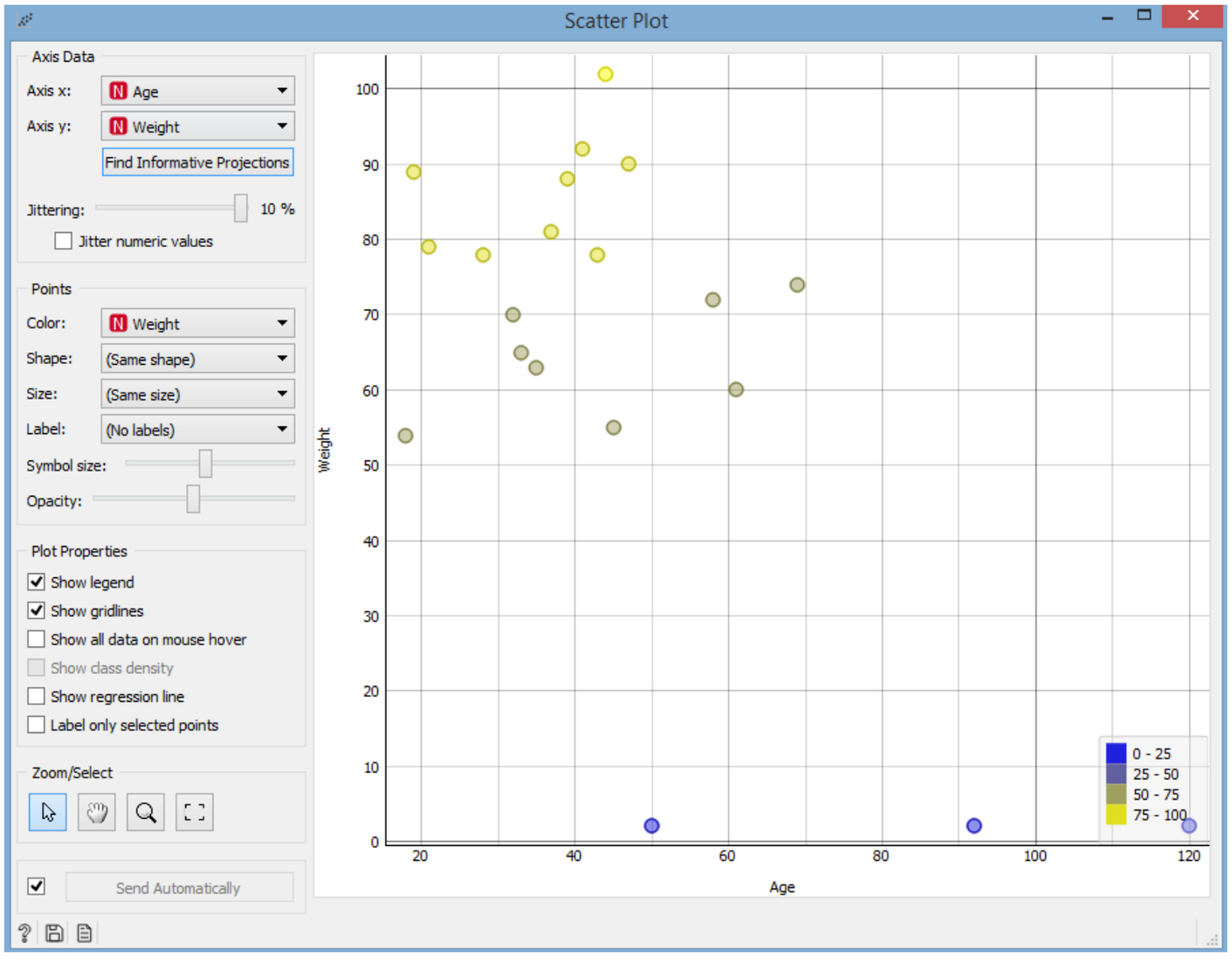


ORANGE : КАК ПОСТРОИТЬ SCATTERING PLOT

К существующему проекту добавляем widgets ‘Scatter plot’. Он находится в опции ‘Visualize’. Соединяем widget ‘Scatter plot’ с widget ‘File’.



Для того что бы построить график нужно кликнуть на widget ‘Scatter plot’. В появившемся окне указать какие данные будут использоваться для осей X,Y и по какой переменной раскрашивать данные.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

https://linis.hse.ru/

Phone: +7 (911) 981 9165

Email: skoltsov@hse.ru