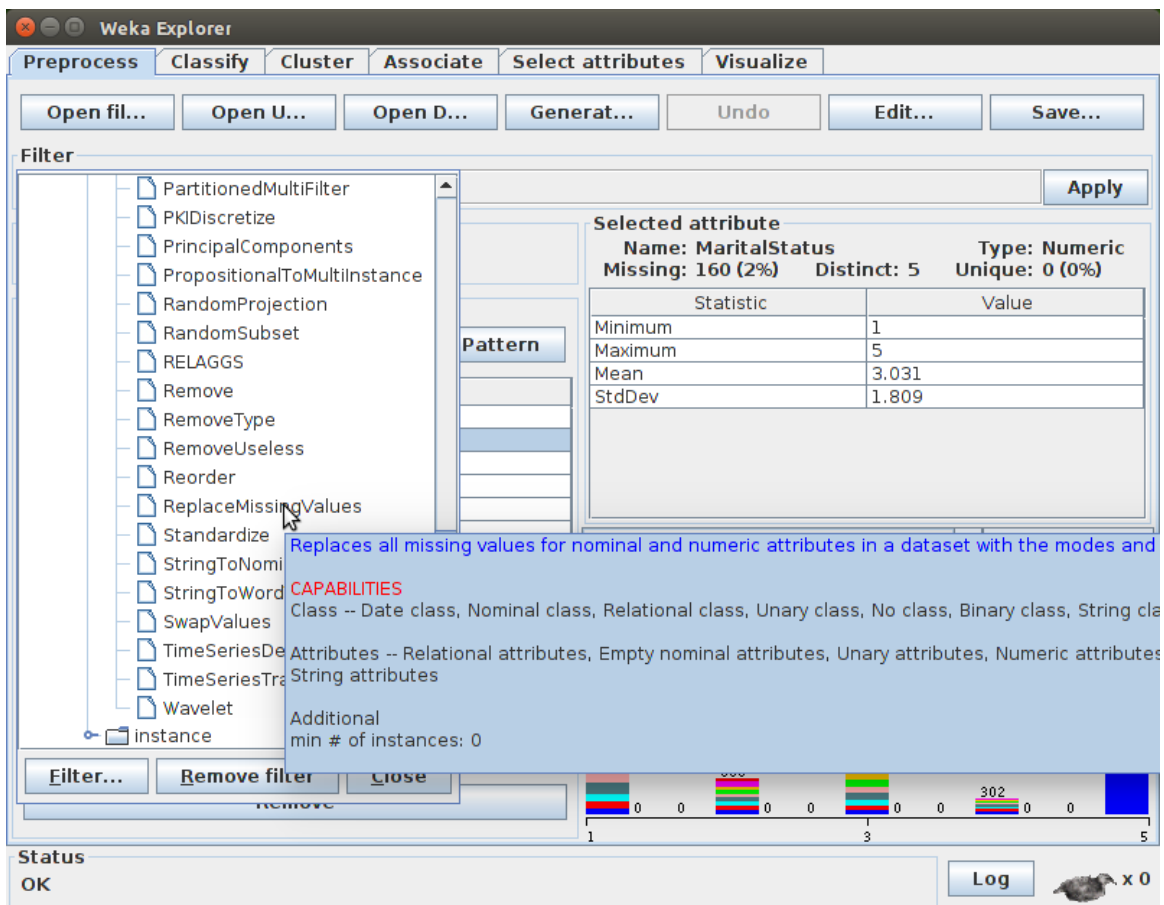


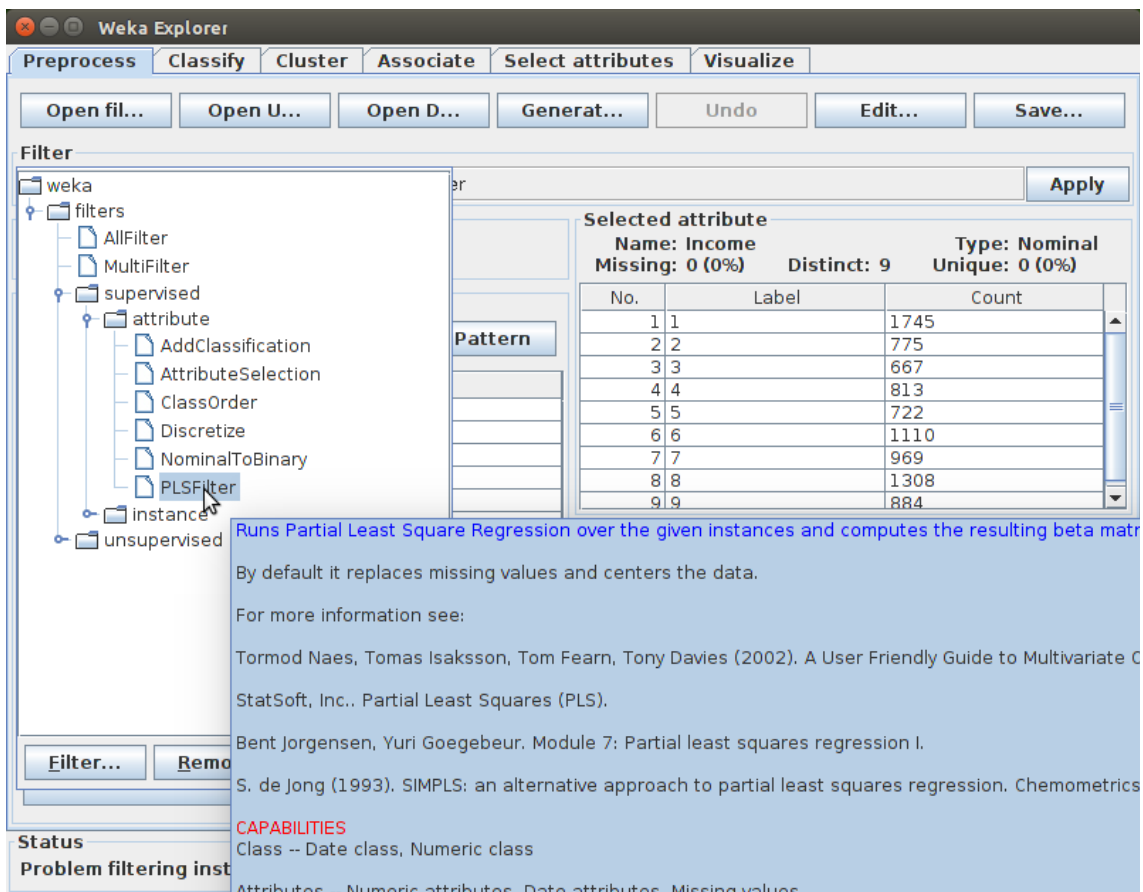
## 1. Pre-Process Missing Values

Preprocess(tab)

filter(choose)>unsupervised>attribute>**ReplaceMissingValues**



supervised>attribute>**PLSfilter**(cannot handle multi-valued nominal class)

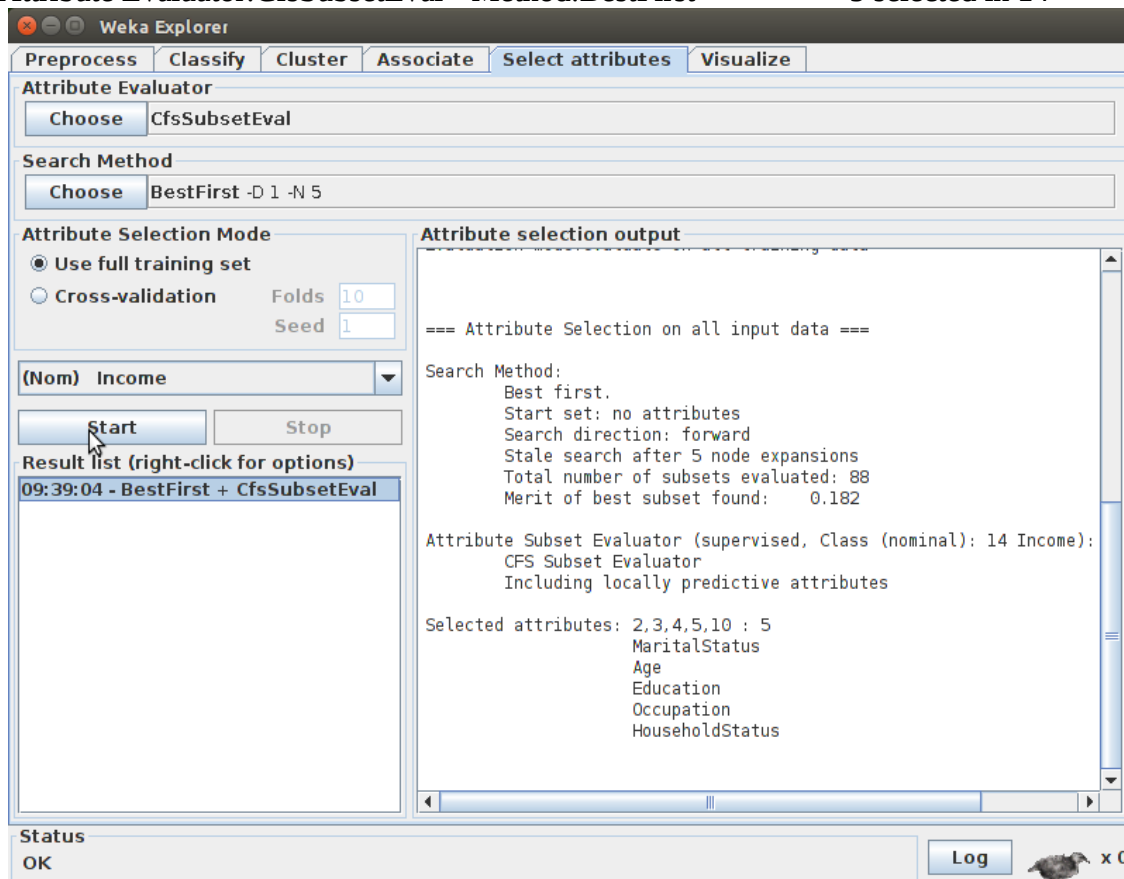


## 2. Attribute Filter Option

Select Attributes(tab)

Attribute Evaluator:CfsSubsetEval Method:BestFirst

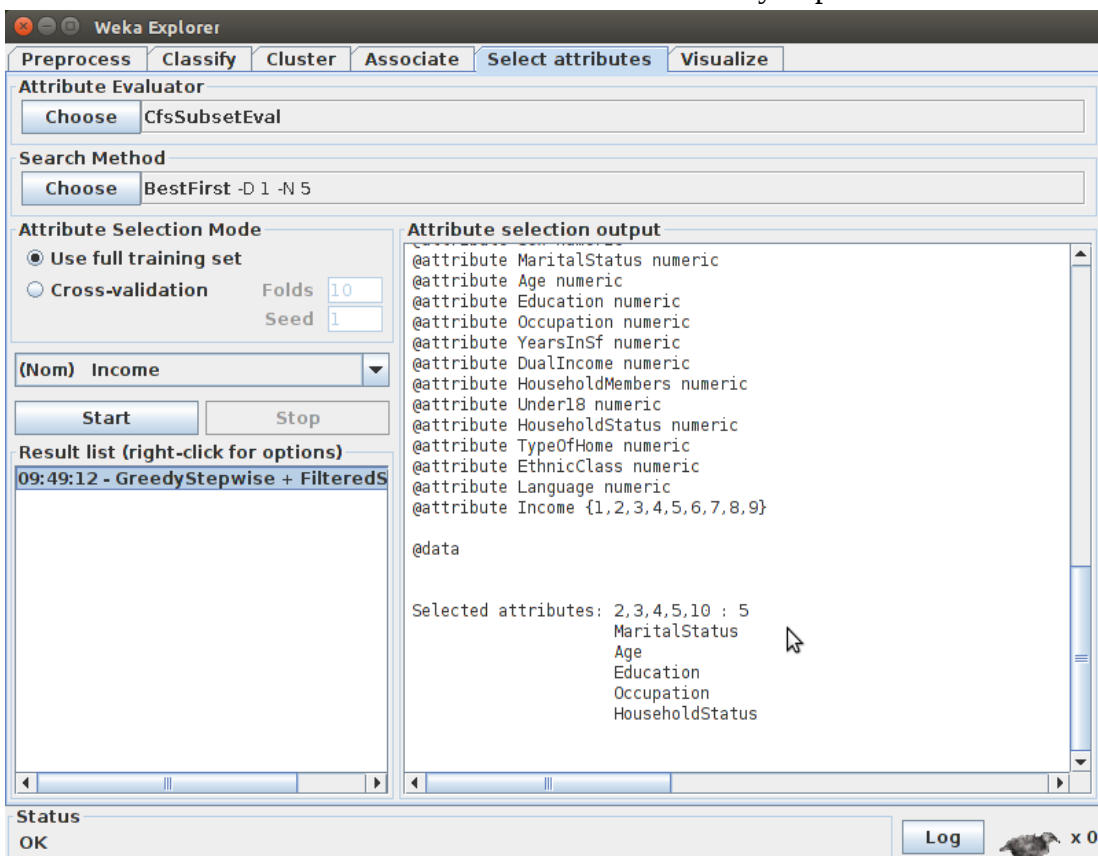
5 selected in 14



Attribute Evaluator:FilteredSubsetEval

Method:GreedyStepWise

5 selected in 14



### 3. Discretization

Preprocess(tab)

Supervised>attribute>Discretize

The screenshot shows the Weka Explorer interface with the 'Preprocess' tab selected. The 'Filter' pane on the left shows the 'Discretize' filter under 'Supervised' > 'Attribute'. The 'Selected attribute' pane shows 'Name: Sex', 'Missing: 0 (0%)', 'Distinct: 2', and 'Type: Numeric'. The 'Statistic' pane shows 'Minimum: 1', 'Maximum: 2', 'Mean: 1.547', and 'StdDev: 0.498'. The 'Pattern' pane is empty. The 'Status' pane shows 'min # of instances: 0'.

An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes

Discretization is by Fayyad & Irani's MDL method (the default).

For more information, see:

Usama M. Fayyad, Keki B. Irani: Multi-interval discretization of continuousvalued attributes for classification Intelligence, 1022-1027, 1993.

Igor Kononenko: On Biases in Estimating Multi-Valued Attributes. In: 14th International Joint Conference on Artificial Intelligence, 1995.

**CAPABILITIES**

Class -- Nominal class, Binary class

Attributes -- Relational attributes, Missing values, String attributes, Date attributes, Nominal attributes, U Numeric attributes

The screenshot shows the Weka Explorer interface with the 'Preprocess' tab selected. The 'Filter' pane on the left shows the 'Discretize' filter under 'Supervised' > 'Attribute'. The 'Selected attribute' pane shows 'Name: Income', 'Missing: 0 (0%)', 'Distinct: 1', and 'Type: Nominal'. The 'Statistic' pane shows 'Minimum: 1', 'Maximum: 2', 'Mean: 1.547', and 'StdDev: 0.498'. The 'Pattern' pane is empty. The 'Status' pane shows 'min # of instances: 0'.

Relation: marketing-weka.filters.unsupervised.attribute.Repl...

Relation: marketing-weka.filters.unsupervised.attribute.Repl...

Unsupervised>attribute>Discretize

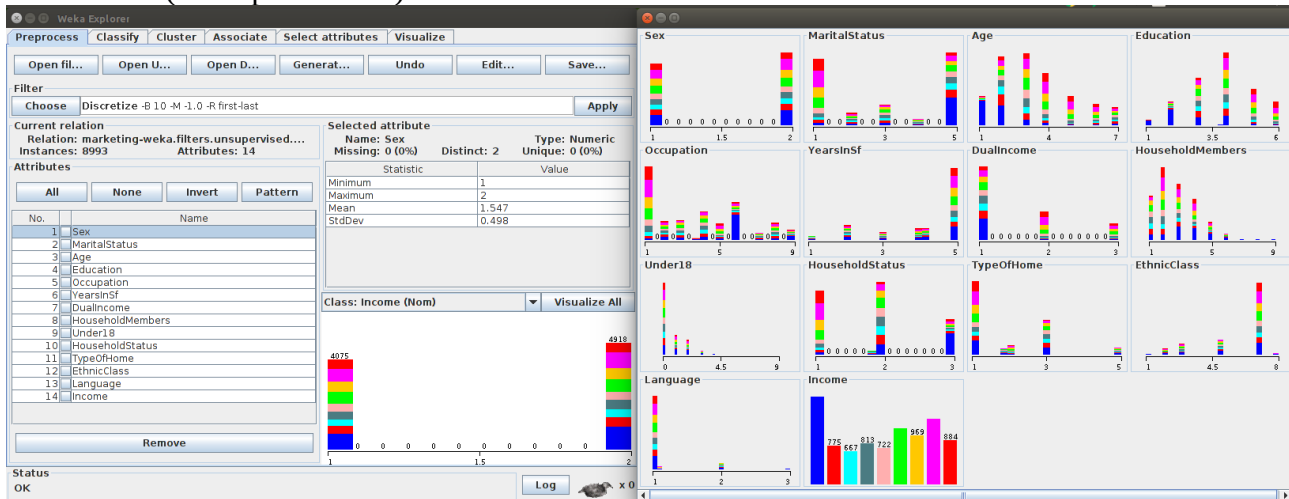
The screenshot shows the Weka Explorer interface with the 'Preprocess' tab selected. The 'Filter' pane on the left shows the 'Discretize' filter under 'Unsupervised' > 'Attribute'. The 'Selected attribute' pane shows 'Name: Income', 'Missing: 0 (0%)', 'Distinct: 2', and 'Type: Nominal'. The 'Statistic' pane shows 'Minimum: 1', 'Maximum: 2', 'Mean: 1.547', and 'StdDev: 0.498'. The 'Pattern' pane is empty. The 'Status' pane shows 'min # of instances: 0'.

Relation: marketing-weka.filters.unsupervised.attribute.Repl...

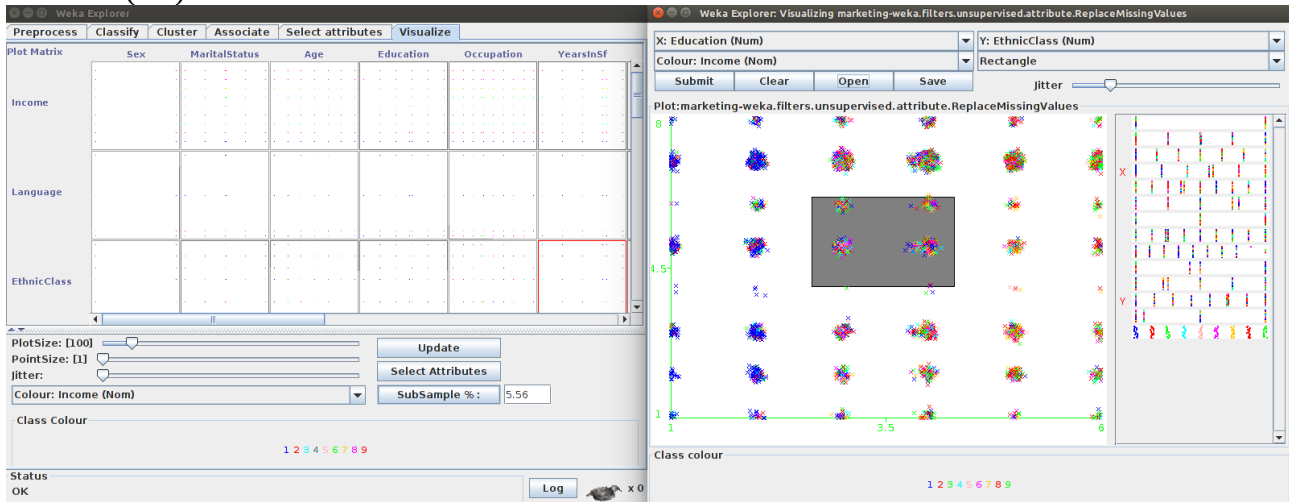
Relation: marketing-weka.filters.unsupervised.attribute.Repl...

## 4. Visualize

Visualize all(in Preprocess tab)



## Visualize(tab)



(LEFT img)A matrix of all attributes against all attributes is provided (with default coloring with CLASS attribute), options to play with the visualization are provided at the bottom

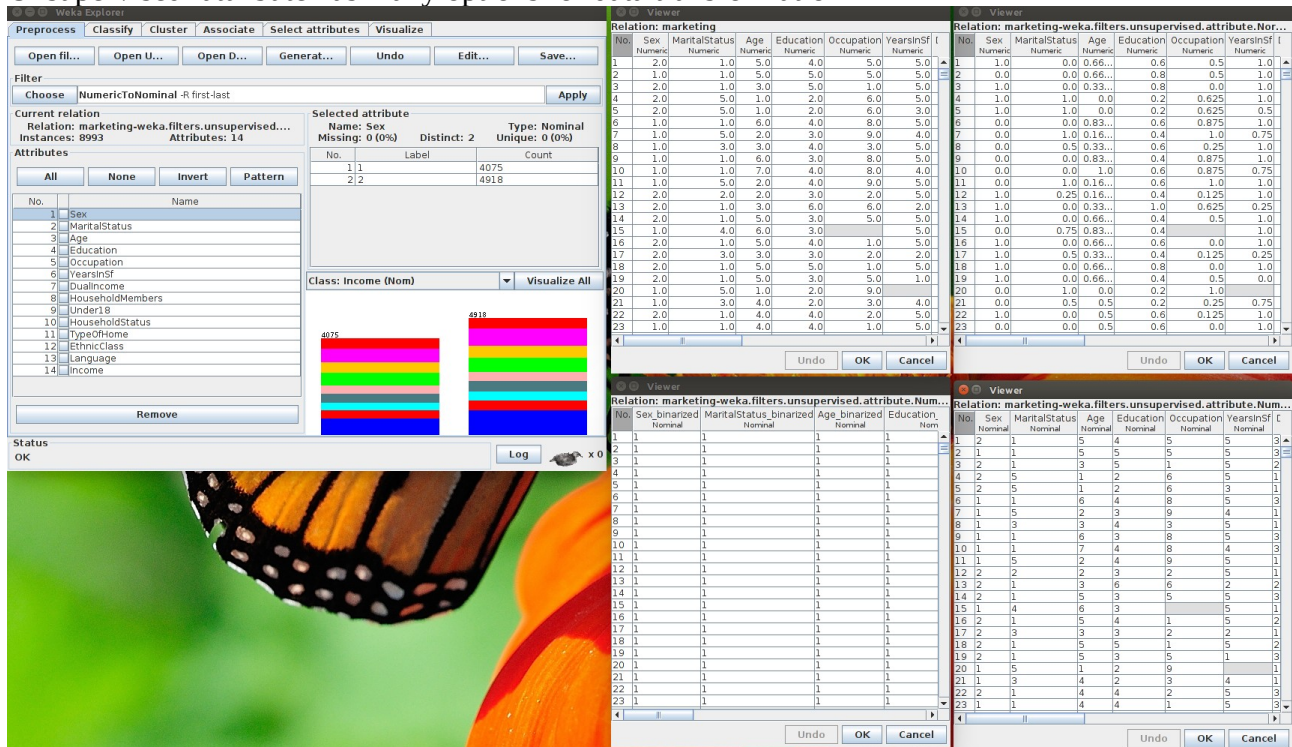
(RIGHT img) Selecting any section gives a closer look onto the data, with flexible options, like hanging X, Y axes, selecting a region (from Select Instance - (here used Rectangle))



## 5. Data Transformation

Preprocess(tab)

Unsupervised>attribute has many options for data transformation



- 1.Original
- 2.Normalize
- 3.NumericToBinary
- 4.NumericToNominal

## 6. Attribute Selection

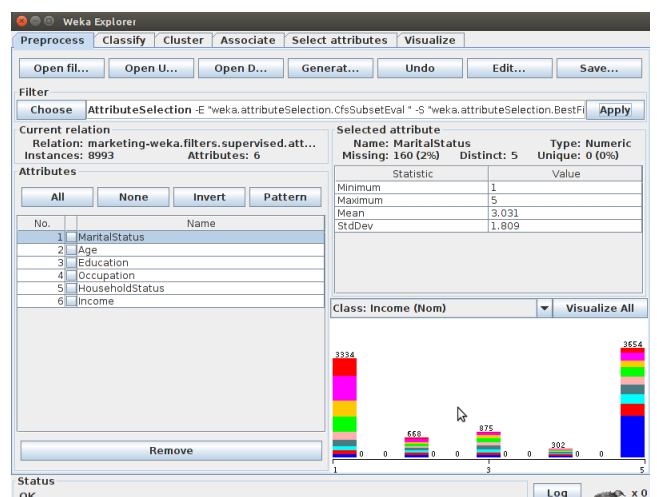
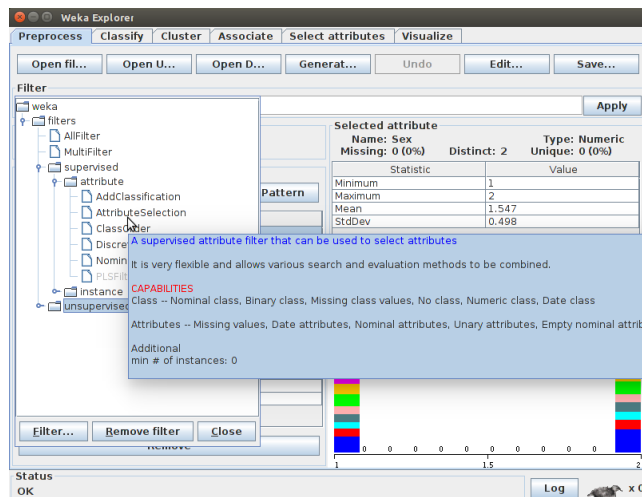
Preprocess(tab)

supervised.attribute.AttributeSelection

Default

Evaluator:CfsSubsetEval

method:BestFit



Final set of attributes have only 6 in number