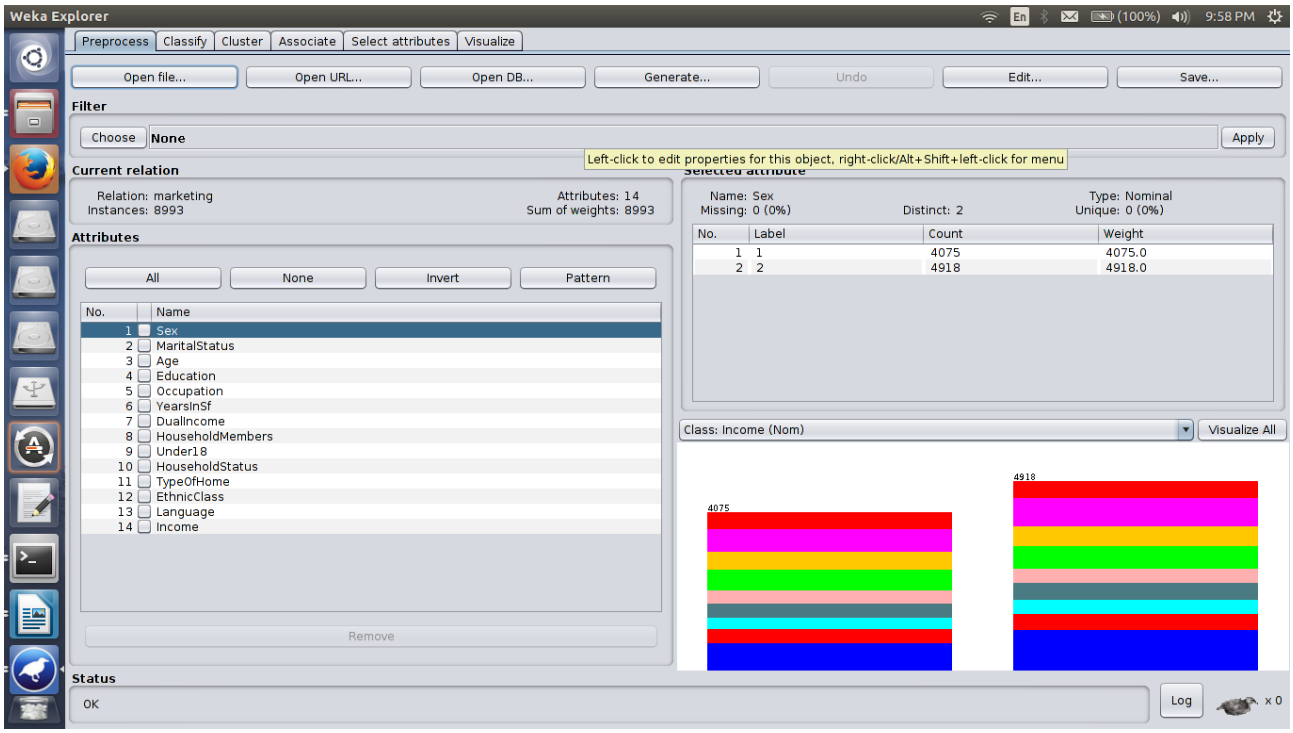
AMMANAMANCHI SAI KARTHIK

B150310CS

Data Set on WEKA:

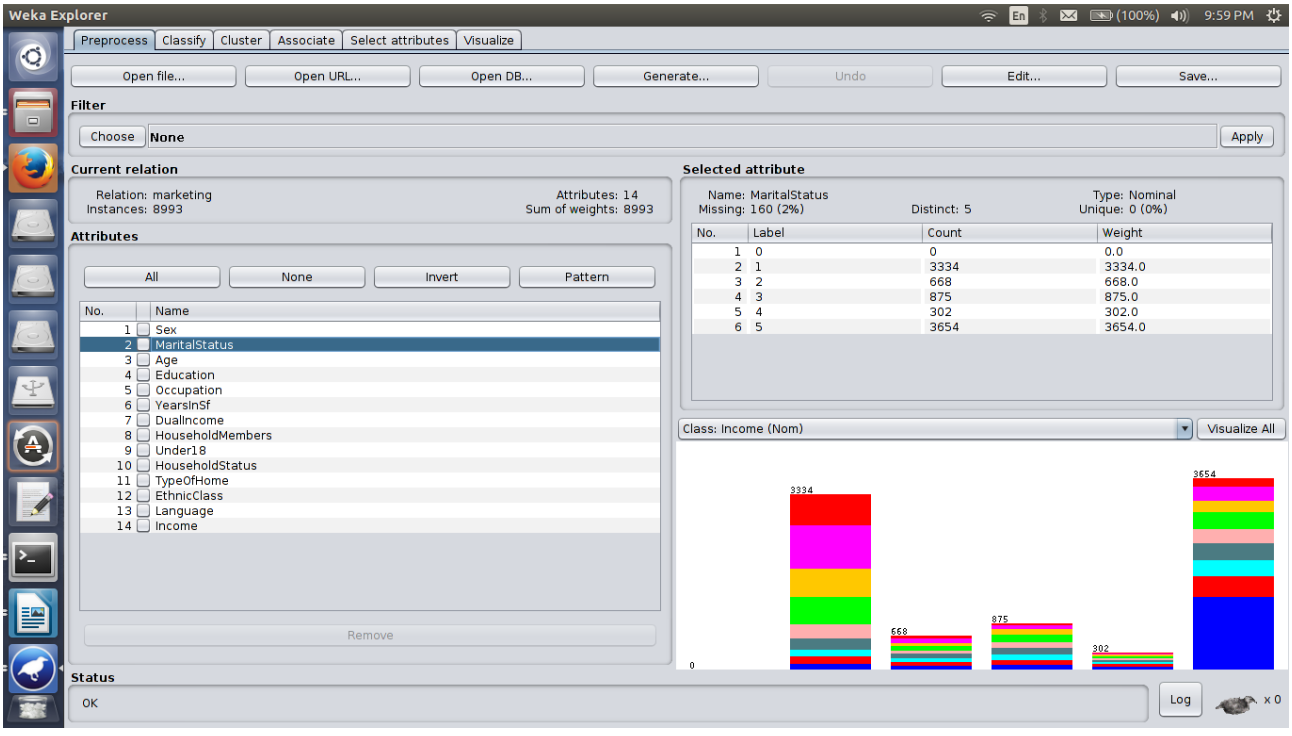


Replacing missing values:

Step 1:

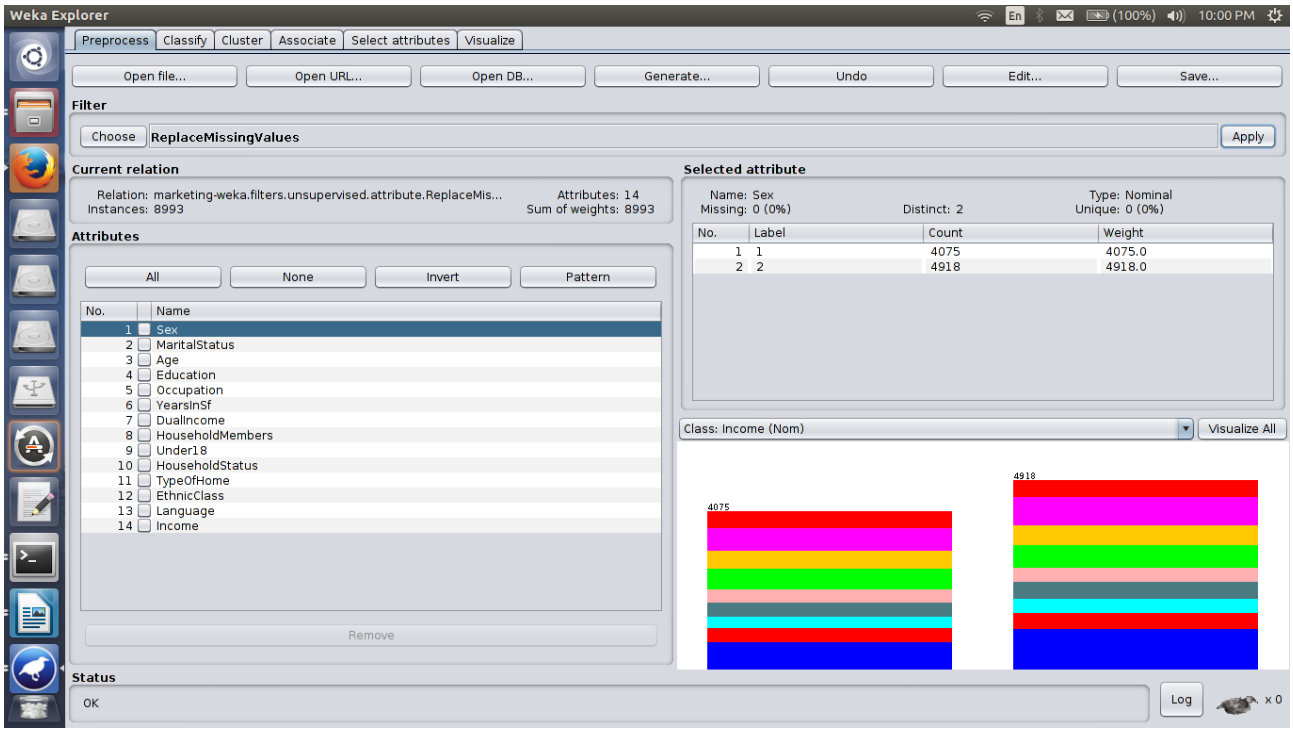
This is the original data set opened in weka.

For attribute marital status we see that there are 160 missing values.



Step 2:

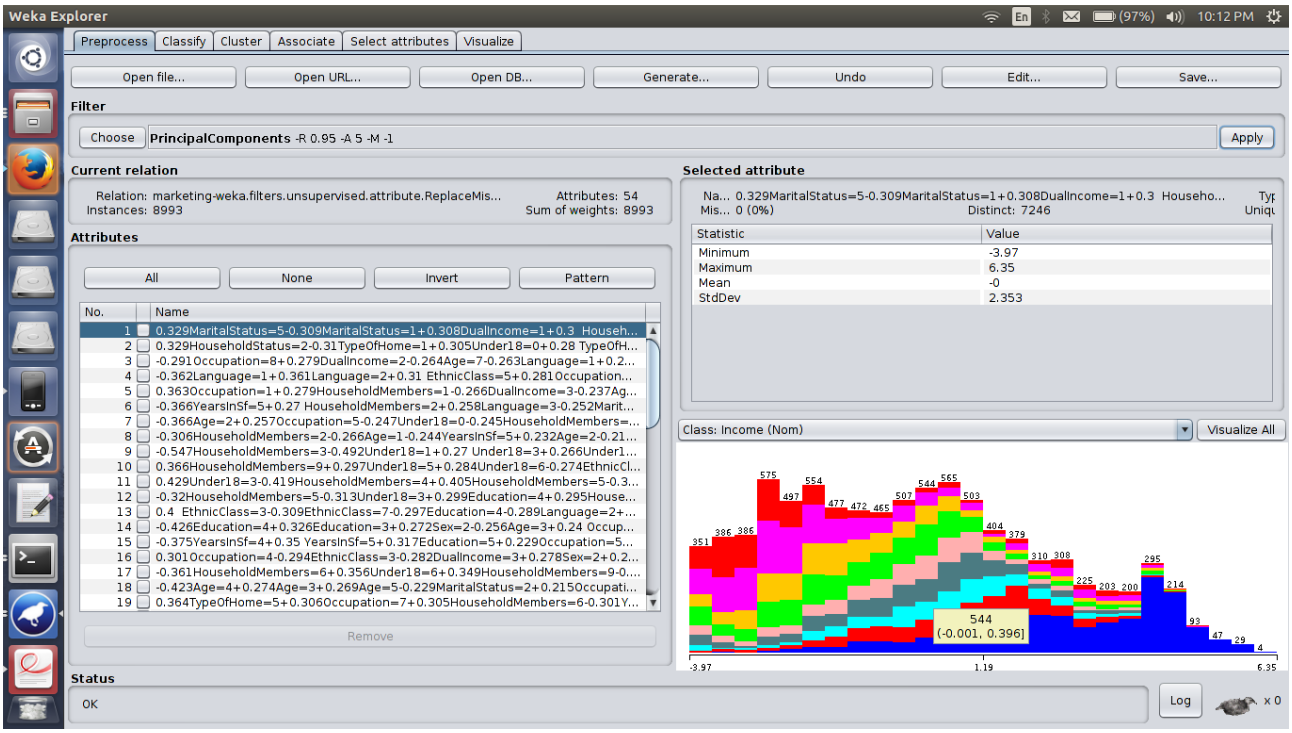
By using the replace missing values option in the unsupervised filters of attributes the missing values are replaced by the mean or mode of the attribute.



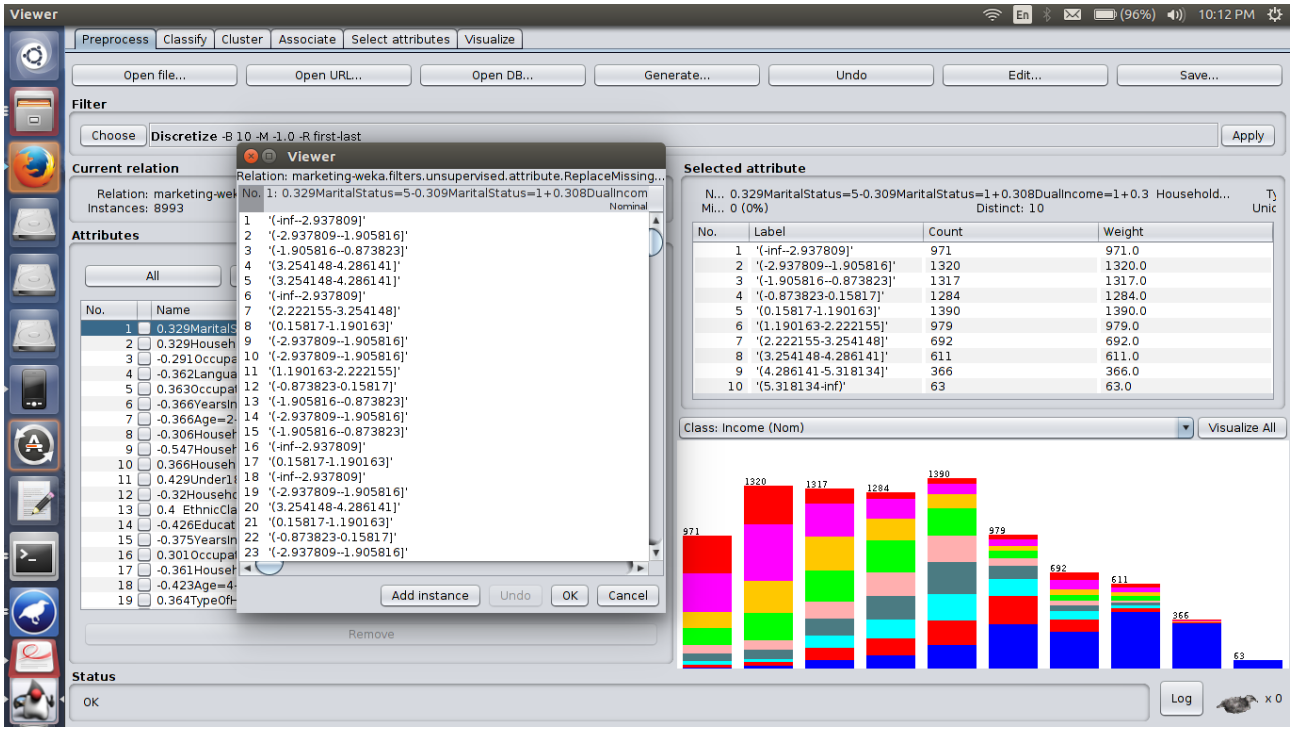
Attribute filter option of weka:

Weka provides many attribute filter options one among it is principal components shown below.

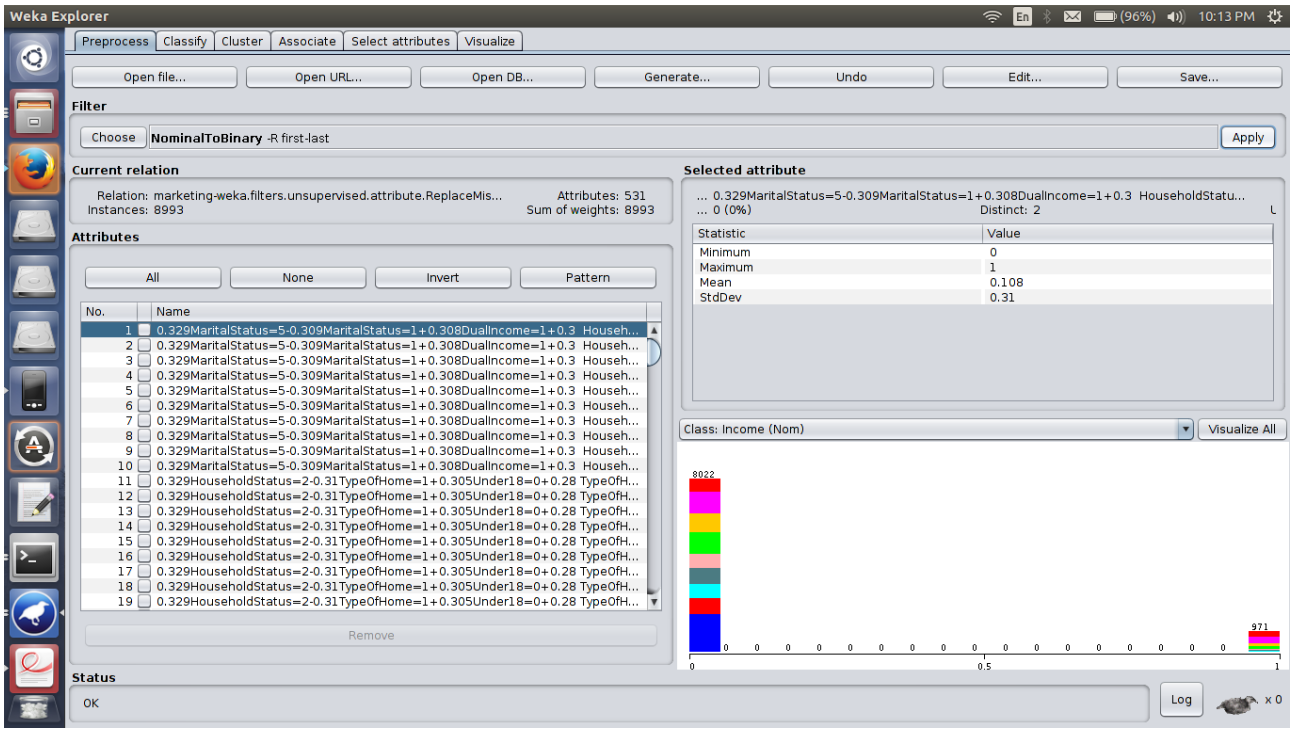
PCA is a dimensionality reduction technique.



Discretization is another attribute filter option in weka .

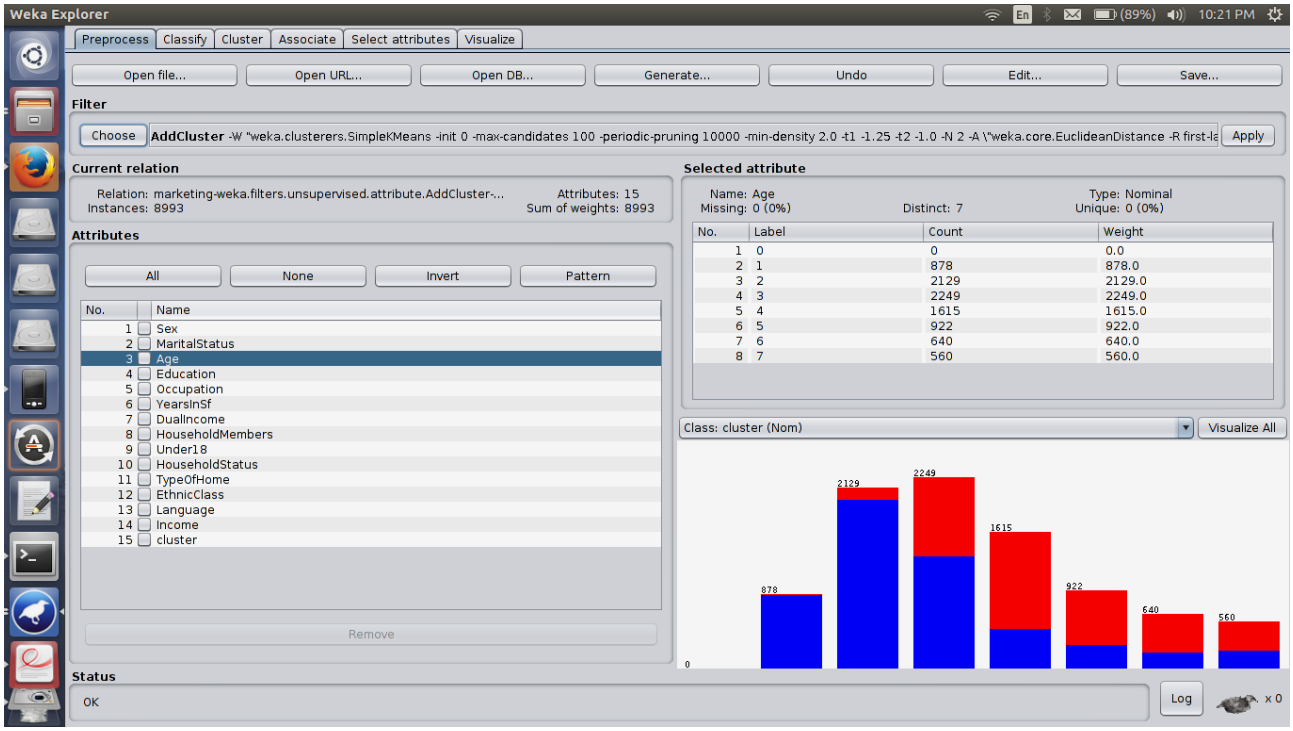


Changing nominal attribute to binary attribute is also supported by weka.



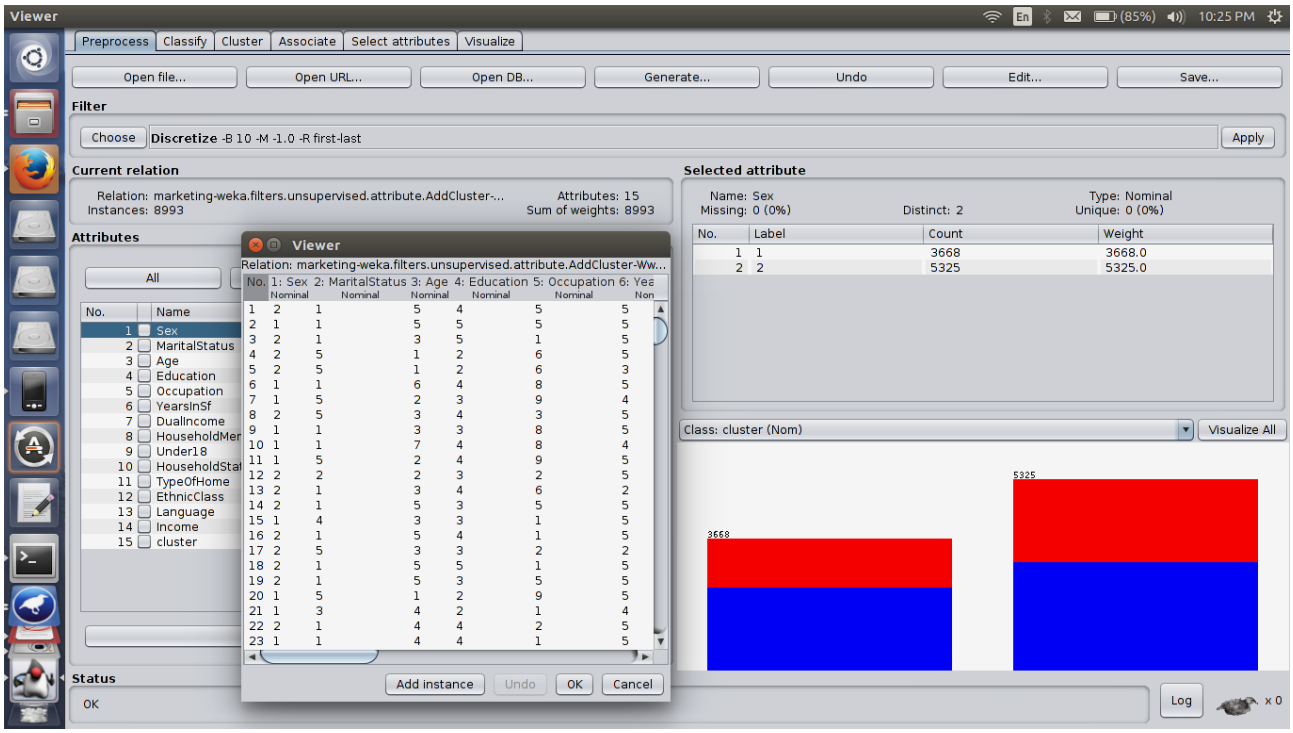
we observe that the attribute has been changed from nominal containing 10 attributes to binary .

Clustering



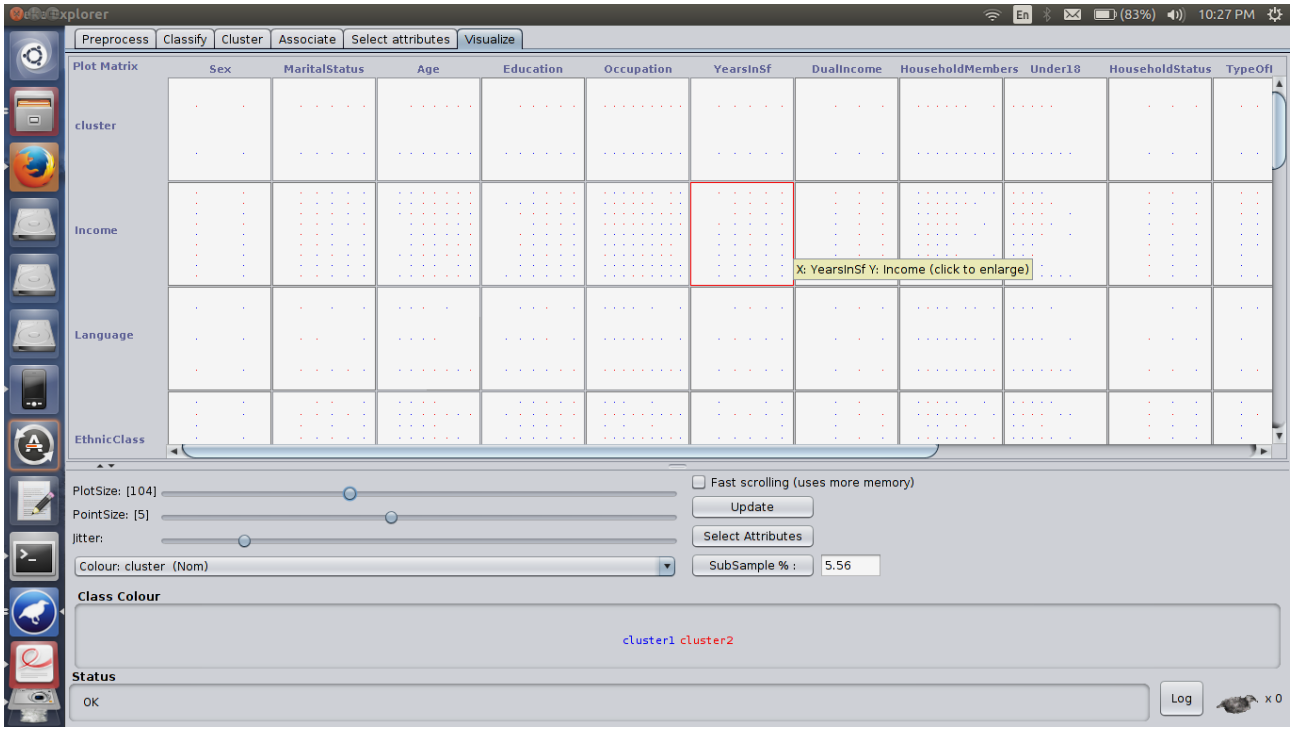
Comparing with the discretization plot we can observe that the attributes are clustered.

Discretization by WEKA:

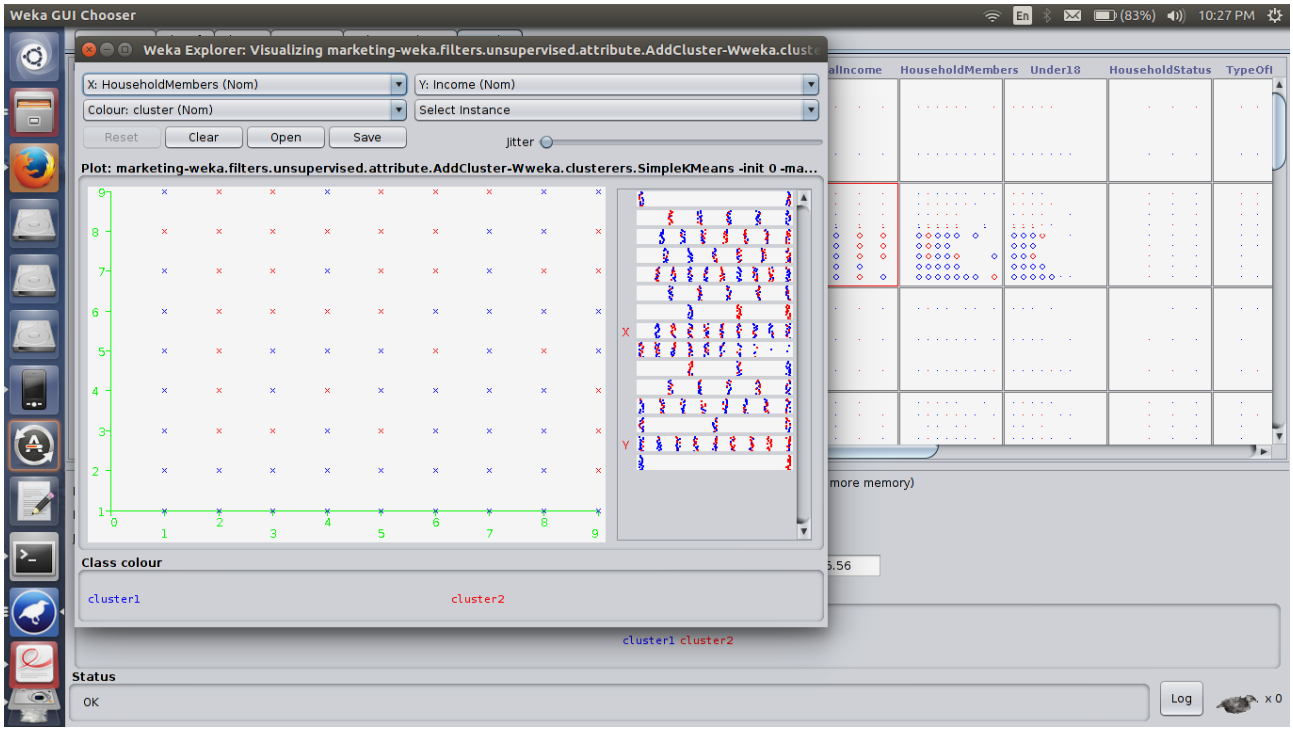


All the values of the data are discretized by this option as we can see in the dialogue box.

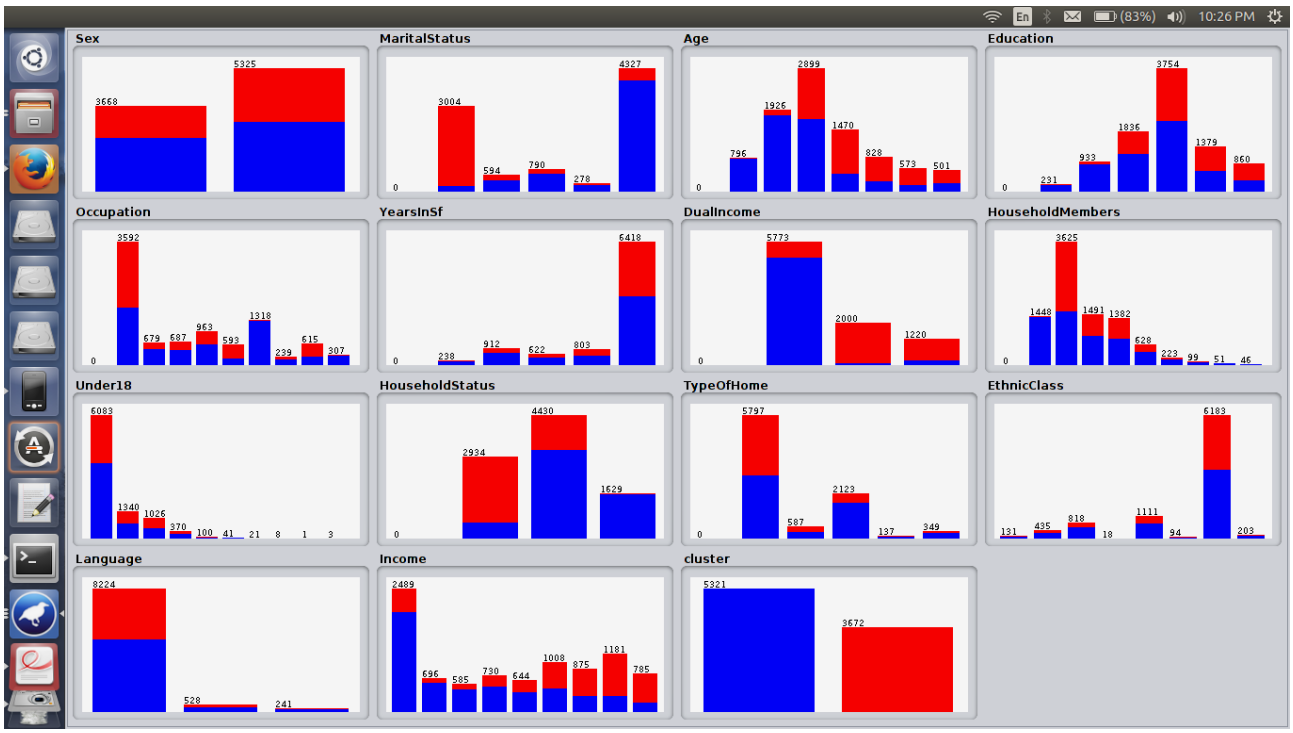
Visualization techniques of WEKA:



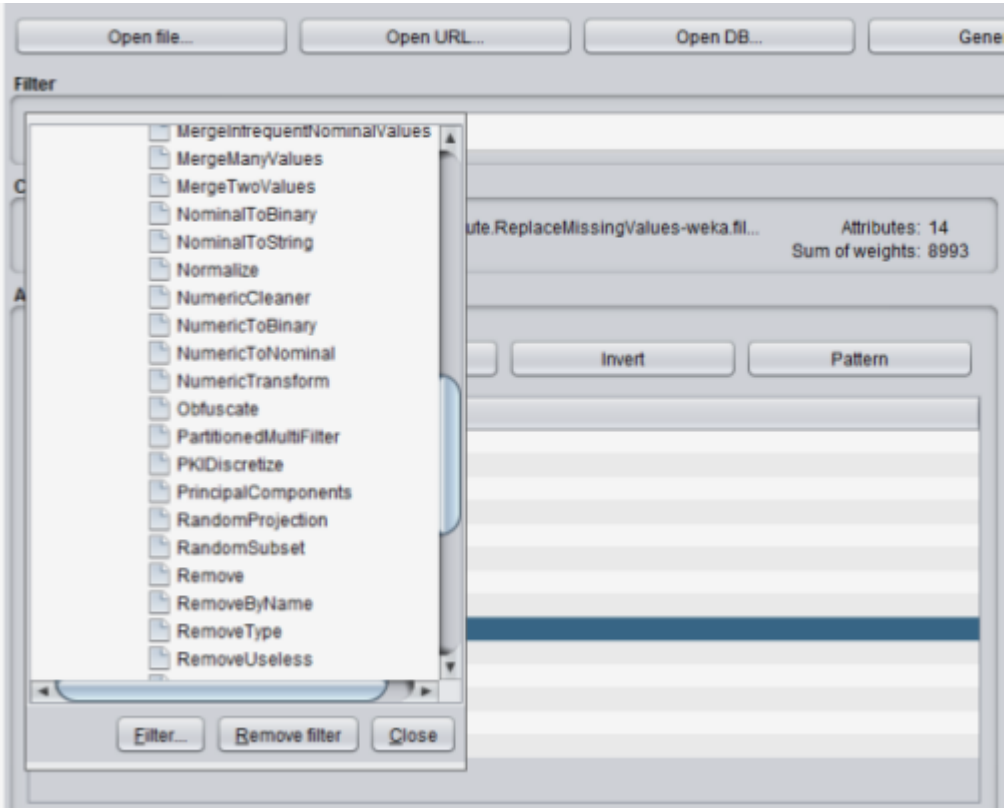
The above screenshot is the scatter plot of the given data.



This is the scatter plot of HouseholdMembers on x-axis and income on y-axis.



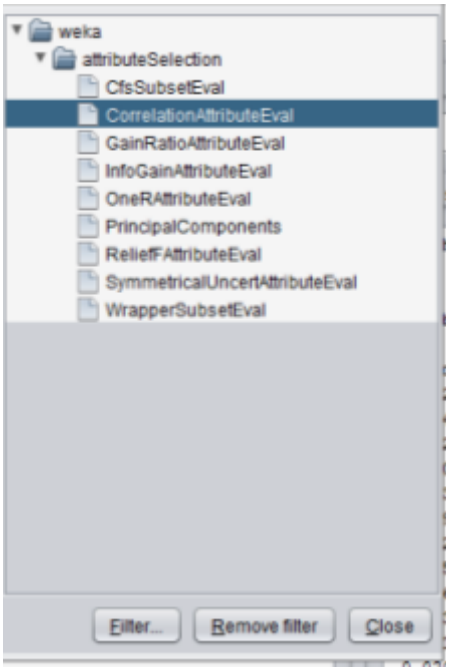
Data transformations supported by weka:



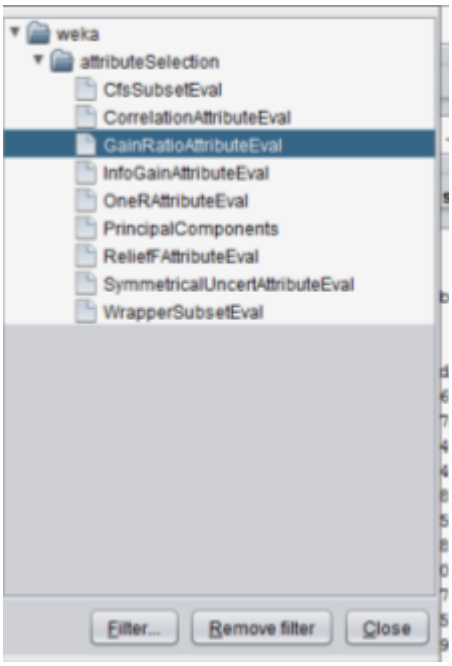
nominal to binary,nominal to string ,numeric to nomial etc.are some of the data transformations.

Attribute selection features of weka:

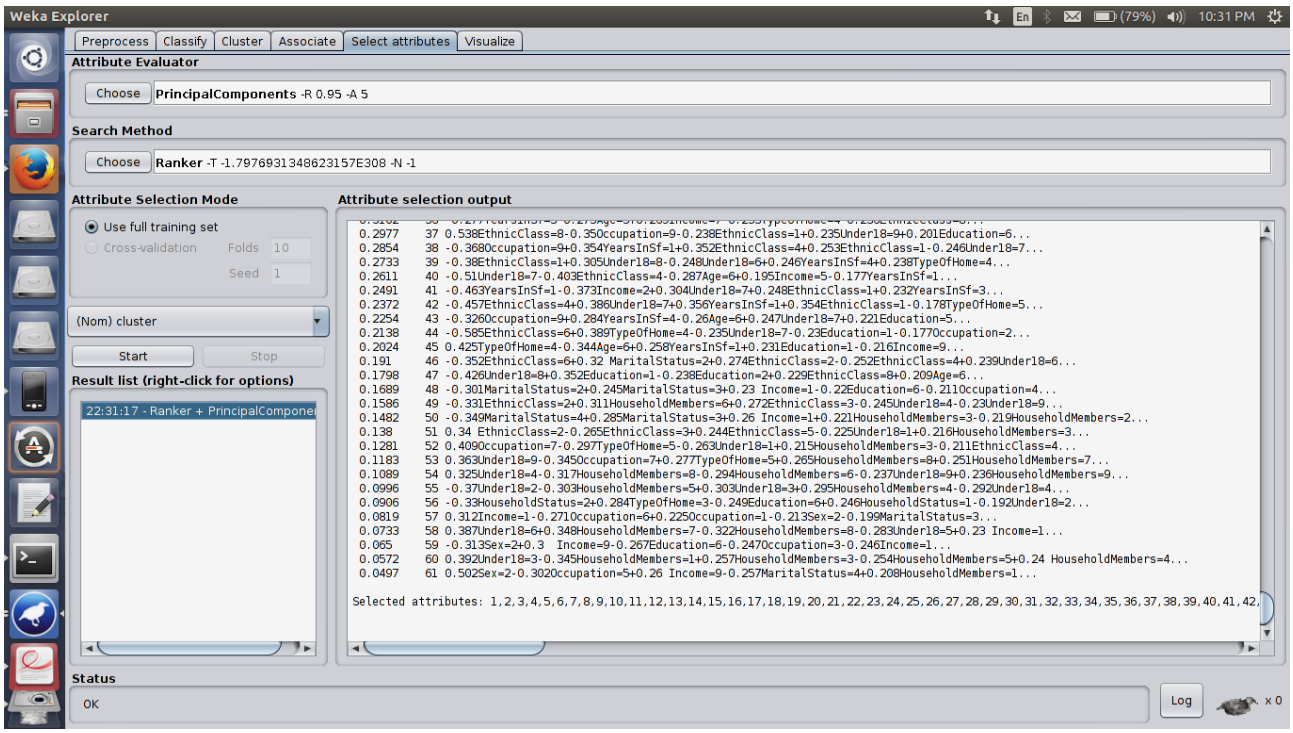
correlation attribute evaluation



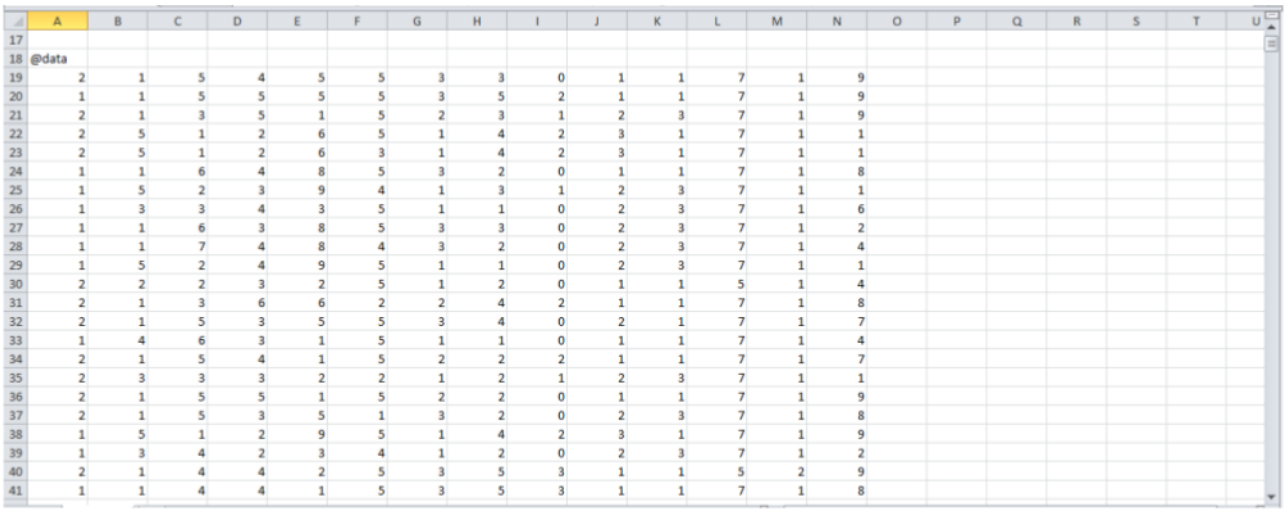
gain ratio attribute evaluation



Principal Components



Exporting to Excel:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
17														
18	@data													
19	2	1	5	4	5	5	3	3	0	1	1	7	1	9
20	1	1	5	5	5	5	3	5	2	1	1	7	1	9
21	2	1	3	5	1	5	2	3	1	2	3	7	1	9
22	2	5	1	2	6	5	1	4	2	3	1	7	1	1
23	2	5	1	2	6	3	1	4	2	3	1	7	1	1
24	1	1	6	4	8	5	3	2	0	1	1	7	1	8
25	1	5	2	3	9	4	1	3	1	2	3	7	1	1
26	1	3	3	4	3	5	1	1	0	2	3	7	1	6
27	1	1	6	3	8	5	3	3	0	2	3	7	1	2
28	1	1	7	4	8	4	3	2	0	2	3	7	1	4
29	1	5	2	4	9	5	1	1	0	2	3	7	1	1
30	2	2	2	3	2	5	1	2	0	1	1	5	1	4
31	2	1	3	6	6	2	2	4	2	1	1	7	1	8
32	2	1	5	3	5	5	3	4	0	2	1	7	1	7
33	1	4	6	3	1	5	1	1	0	1	1	7	1	4
34	2	1	5	4	1	5	2	2	2	1	1	7	1	7
35	2	3	3	3	2	2	1	2	1	2	3	7	1	1
36	2	1	5	5	1	5	2	2	0	1	1	7	1	9
37	2	1	5	3	5	1	3	2	0	2	3	7	1	8
38	1	5	1	2	9	5	1	4	2	3	1	7	1	9
39	1	3	4	2	3	4	1	2	0	2	3	7	1	2
40	2	1	4	4	2	5	3	5	3	1	1	5	2	9
41	1	1	4	4	1	5	3	5	3	1	1	7	1	8

This is the pre-processed data set exported to excel.