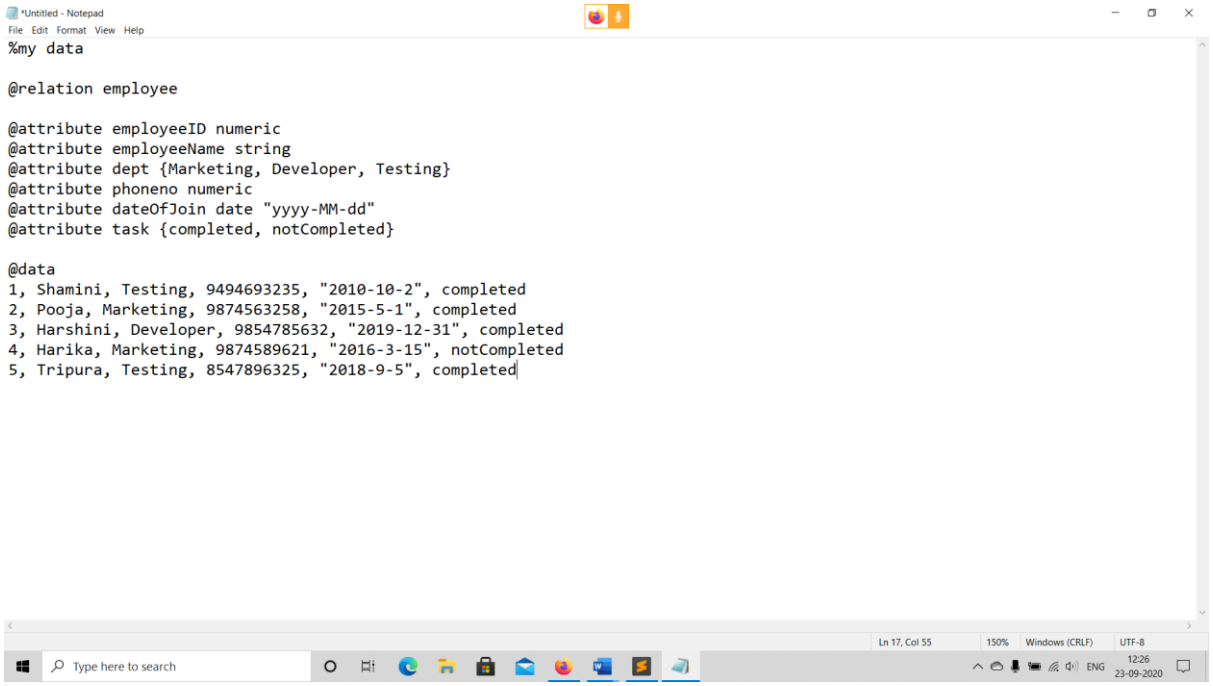
DATA MINING LAB

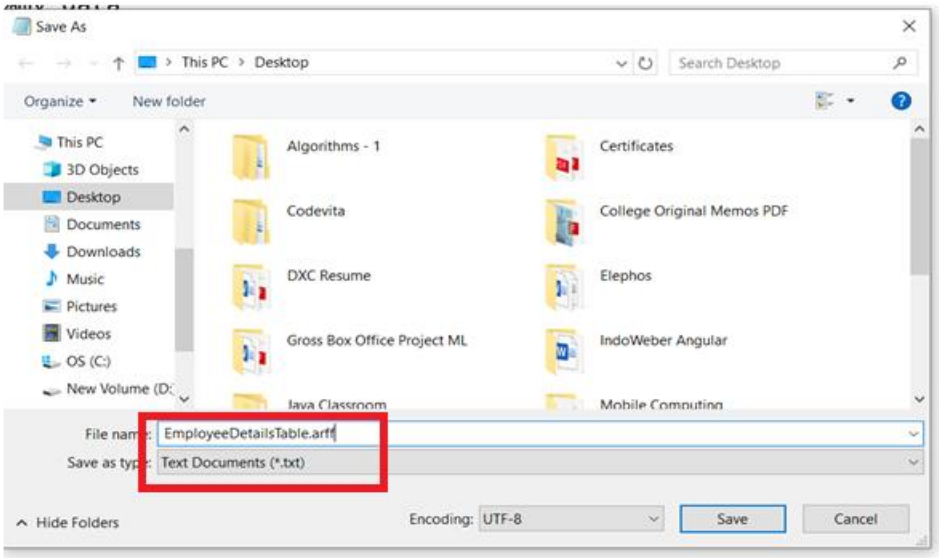
1)Steps to create .arff file.

Step 1: Open a notepad

Step 2: According to the format of .arff file, write the code in notepad.



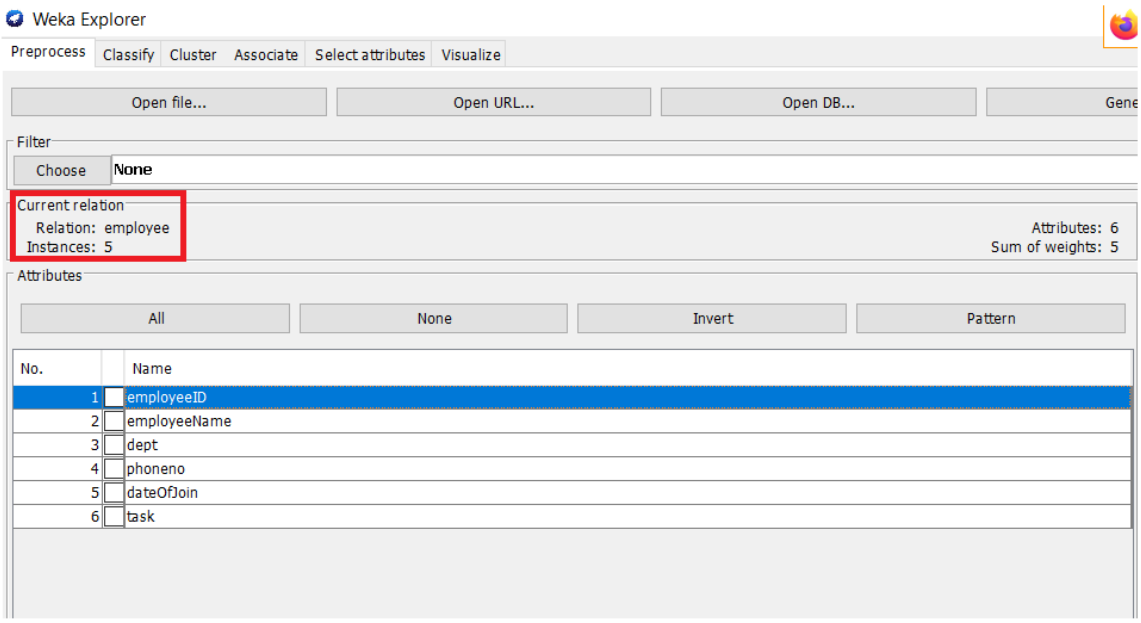
Step 3: After completion of code, save the file with .arff extension at a particular location.



Step 4: Open the created .arff file.

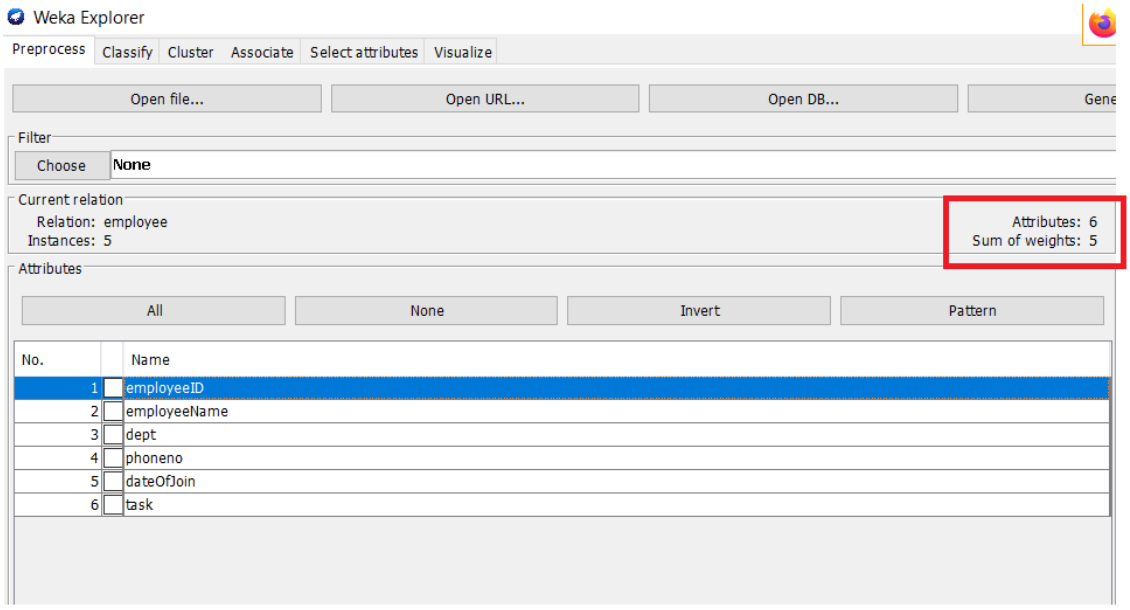
Observations in created file:

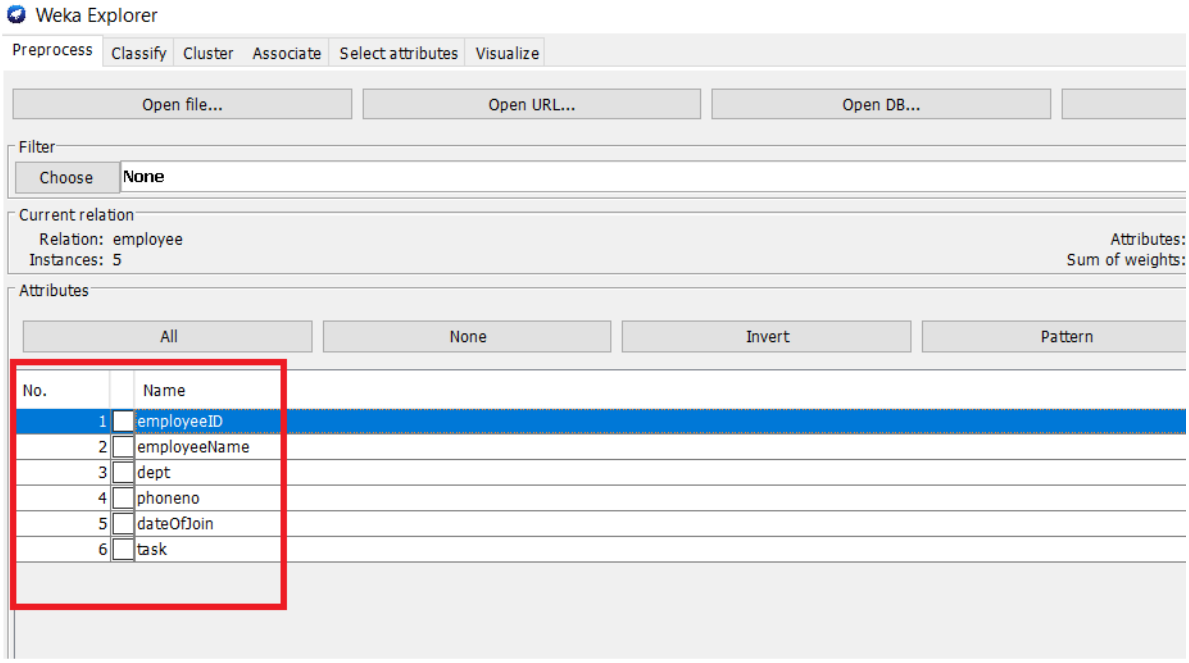
Relation represent the current table name



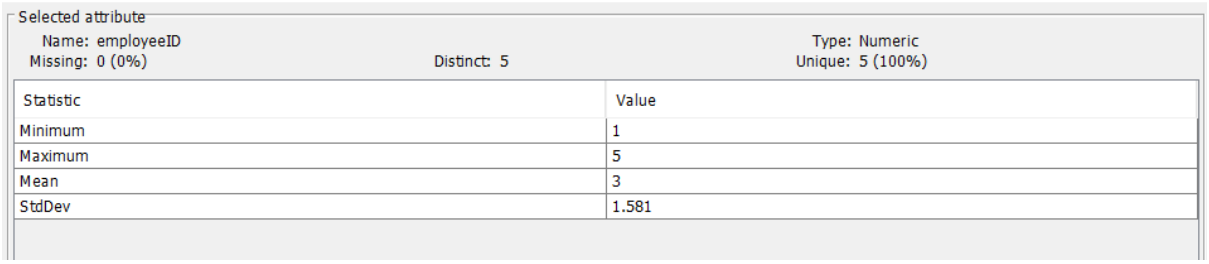
Employee is the name of the relation(table) created.

Attributes represent the number of (attributes)columns present in the relation(table) created.

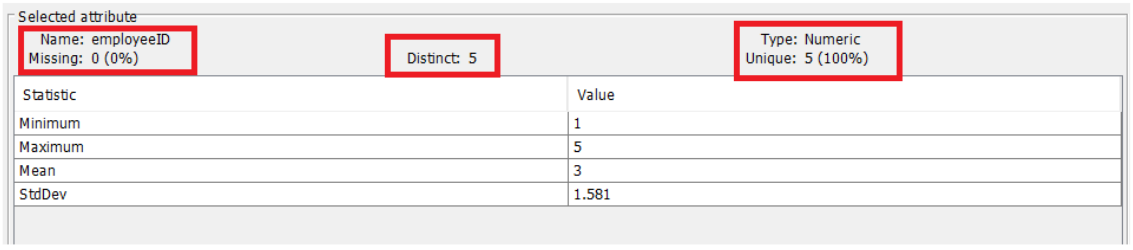




The above red box represents the names of the attributes created in employee table.

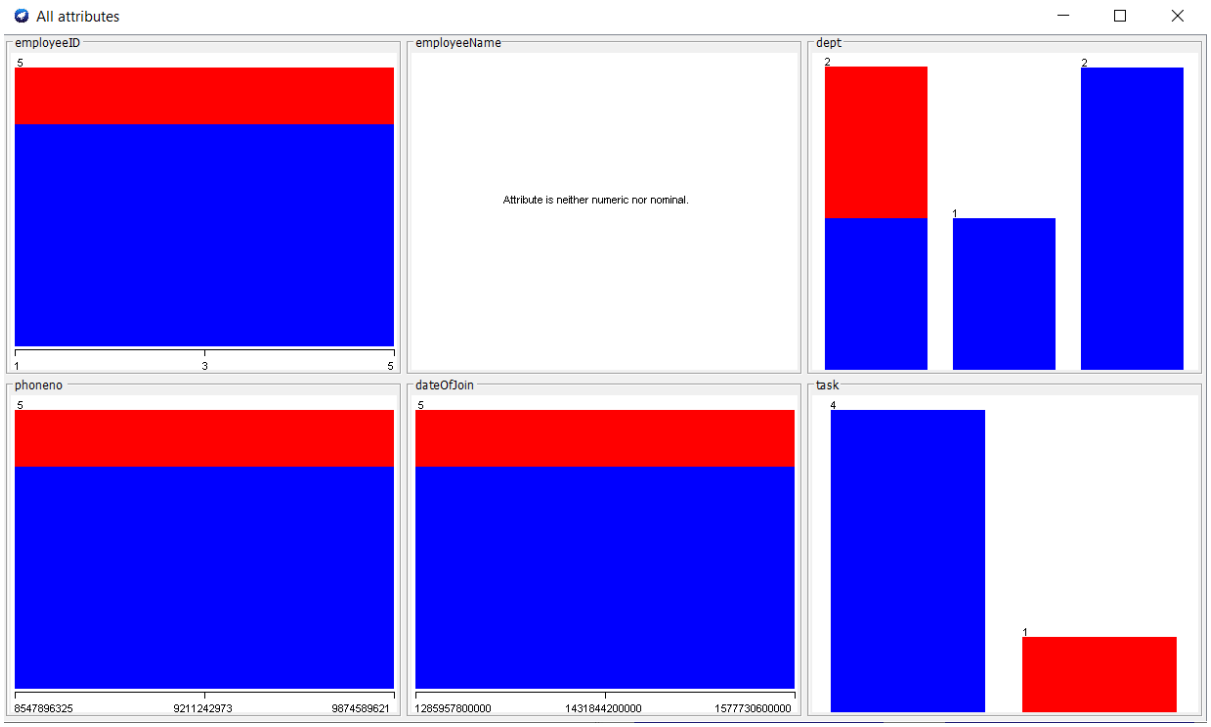


Minimum, maximum, mean, standard deviation represent the values of that particular attribute selected(if its type is numeric).



Name represent the attribute name, missing value represent the number of missing values of that particular attribute, distinct represent the number of unique values present in that attribute, type represent the attribute type(i.e, numeric, nominal, date, string)

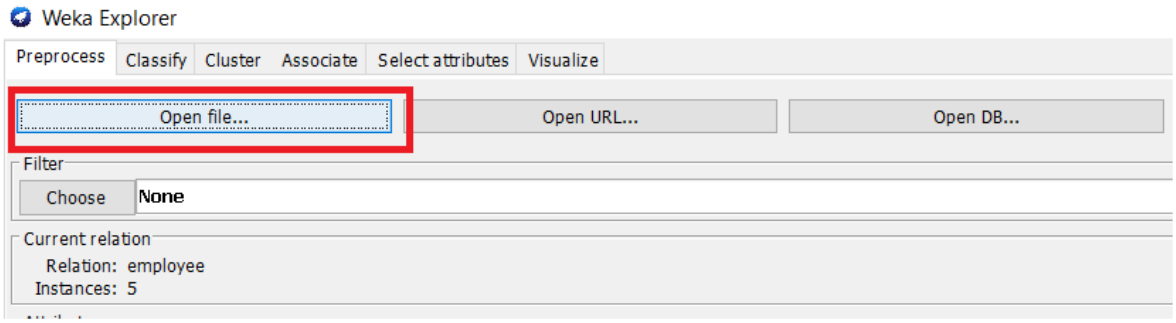
Histogram Plot:



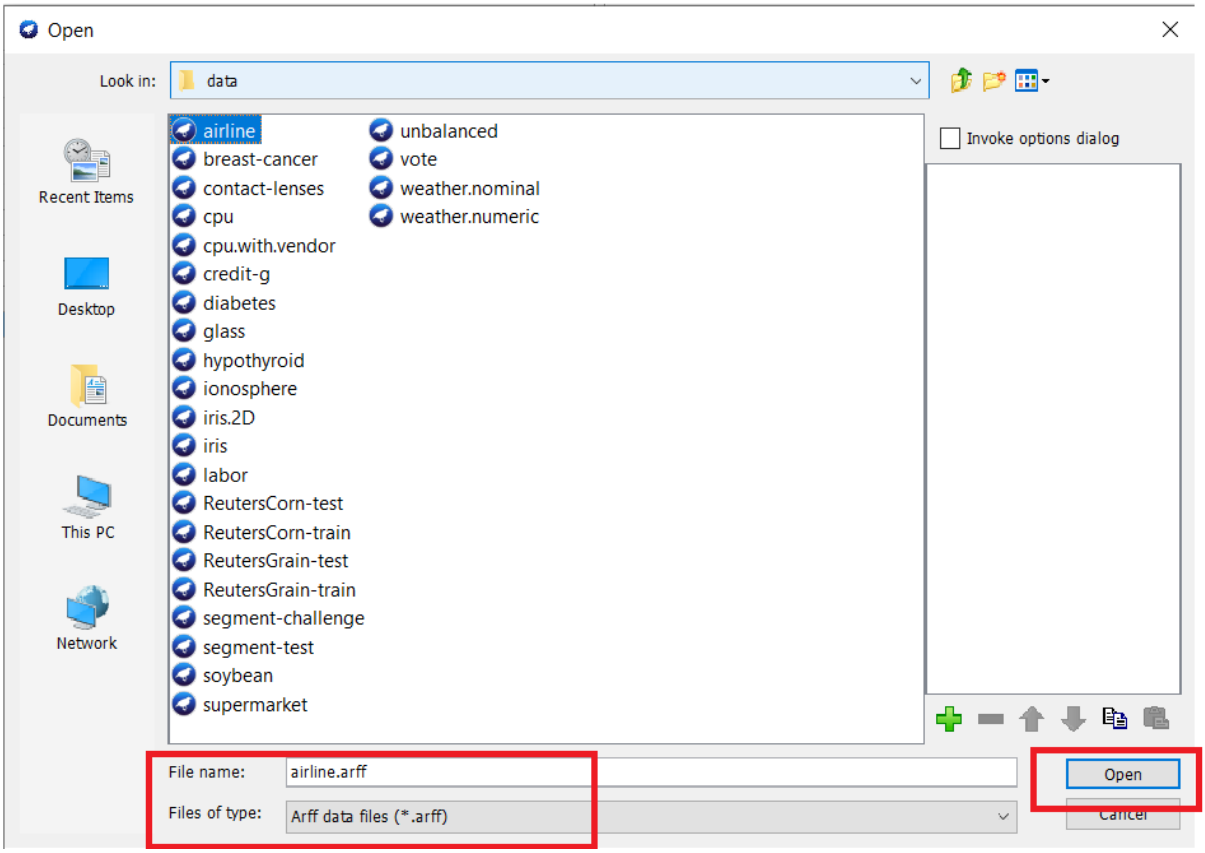
As the type of employee name is string, we cannot plot the graph for it. For all the remaining types graphs can be plotted.

2)Exploring in-built datasets.

a)Loading the dataset.

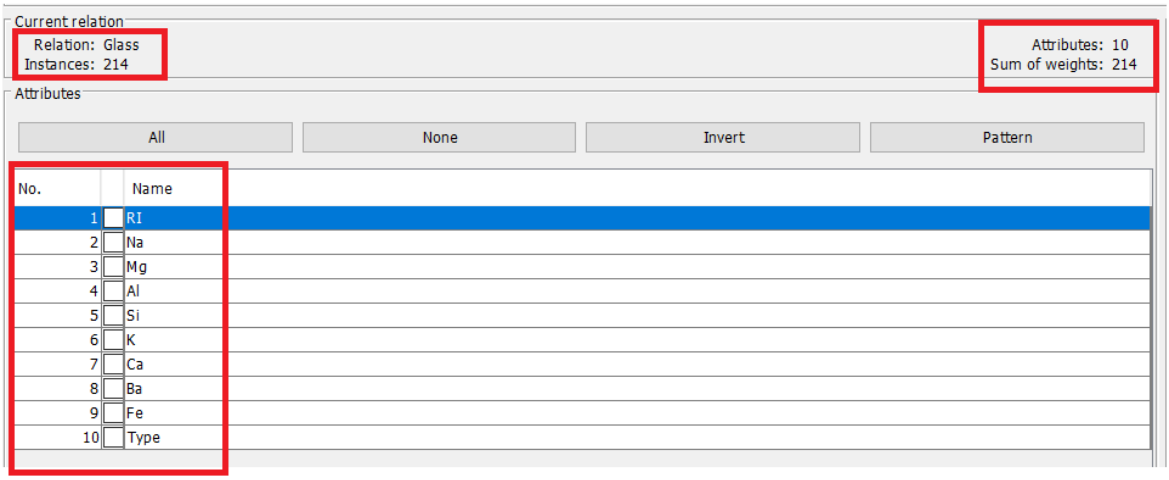


Click on the open file, to open a new dataset. It opens a new window where a .arff file need to be selected from the directories and then open it.



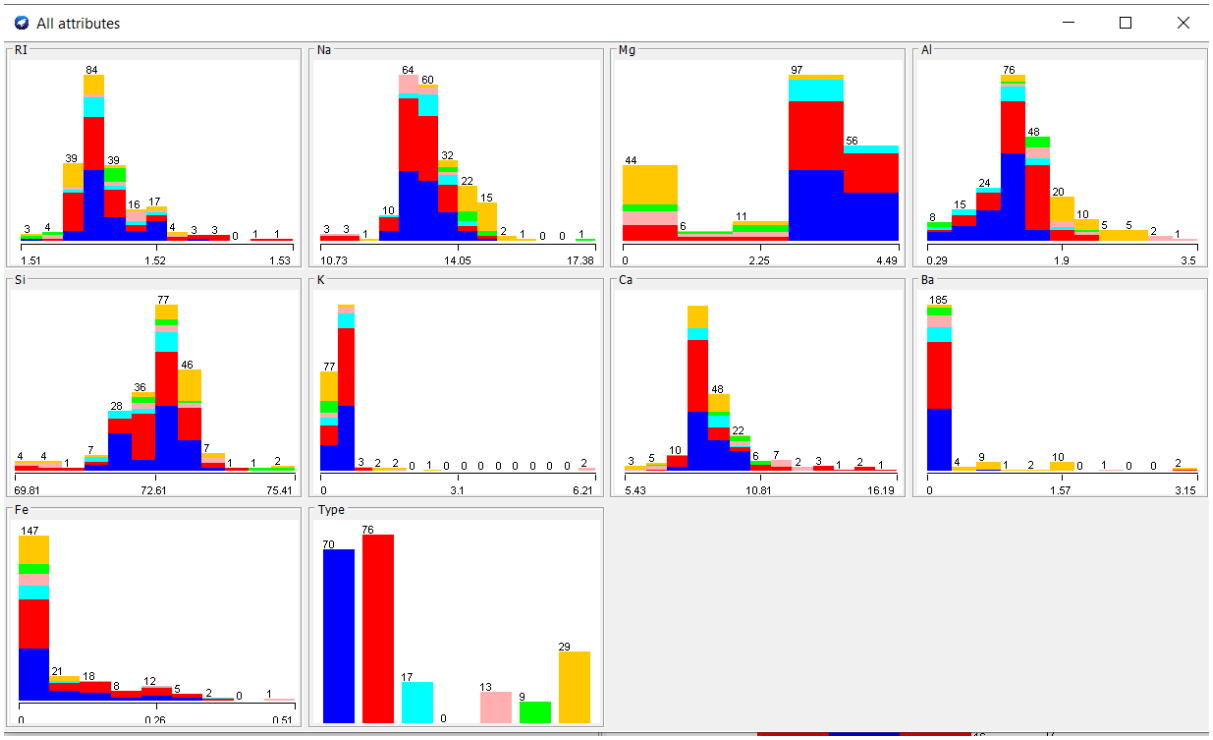
b)Observations in Glass dataset

The name of the relation is “Glass”, it consist of 214 records with 10 attributes(columns). The names of 10 columns are RI, Na, Mg, Al, Si, K, Ca, Ba, Fe, Type with all having type as numeric, except Type attribute having nominal type.



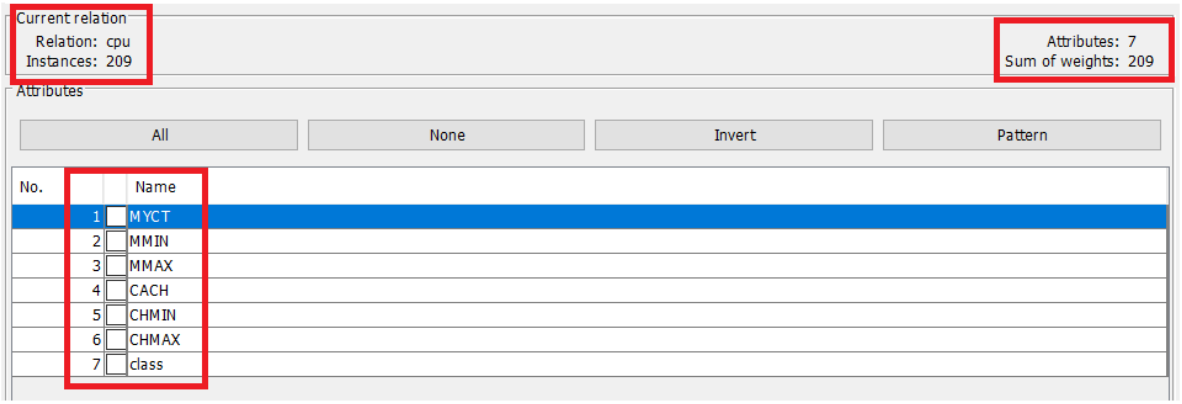
Attribute name	Attribute Type
Ri	Numeric
Na	Numeric
Mg	Numeric
Al	Numeric
Si	Numeric
K	Numeric
Ca	Numeric
Ba	Numeric
Fe	Numeric
Type	Nominal

Histogram Plot:



c)Observations in cpu dataset

The name of the relation is “cpu”, consist of 209 records with 7 attributes.



Attribute Name	Attribute Type
MYCT	Numeric
MMIN	Numeric
MMAX	Numeric
CACH	Numeric
CHMIN	Numeric
CHMAX	Numeric
class	Numeric

Histogram Plot: