1. **Exploration of Student dataset :**

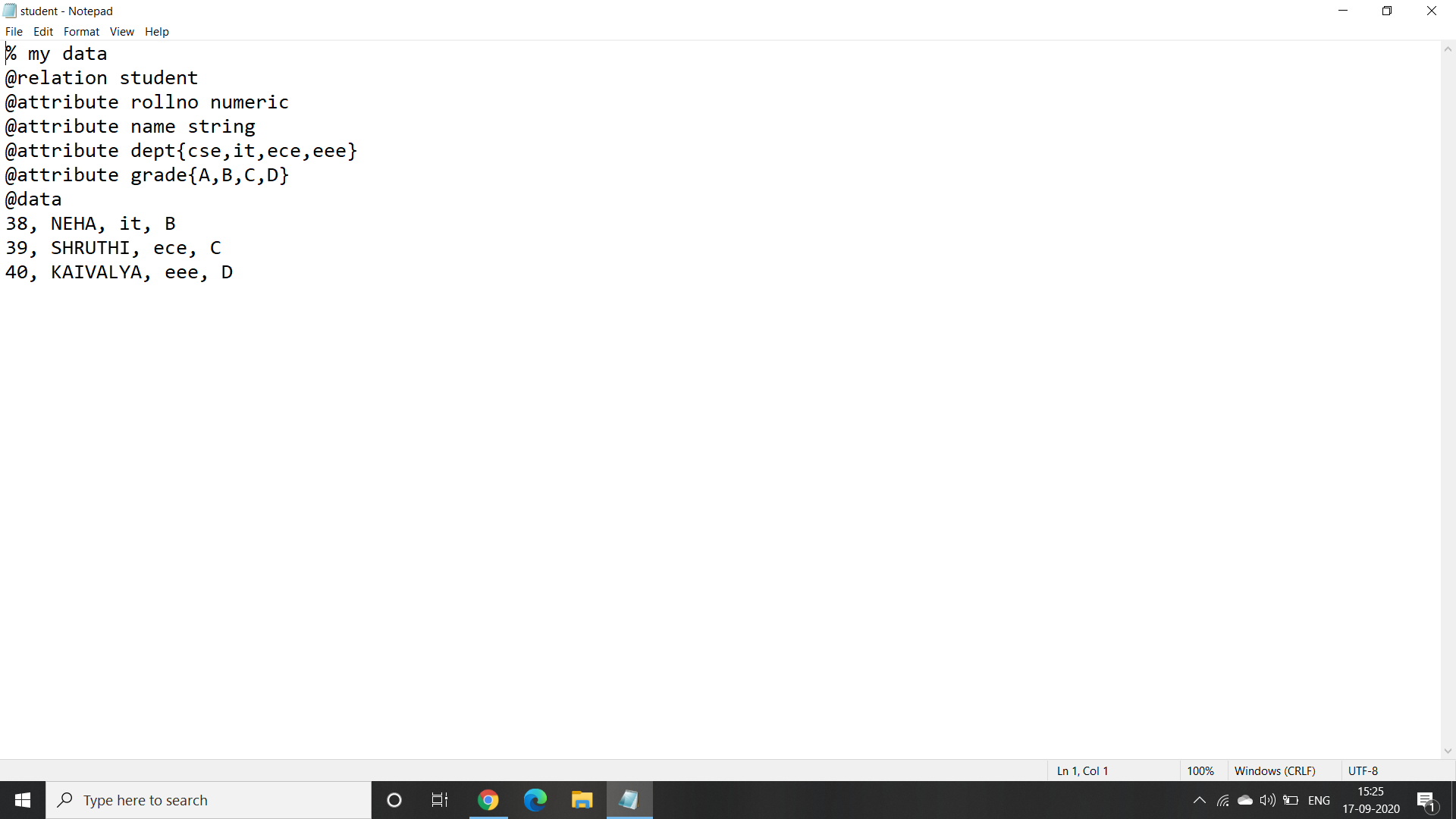
Creation of ARFF FILE :

Step 1 :

Open Notepad

Step 2 :

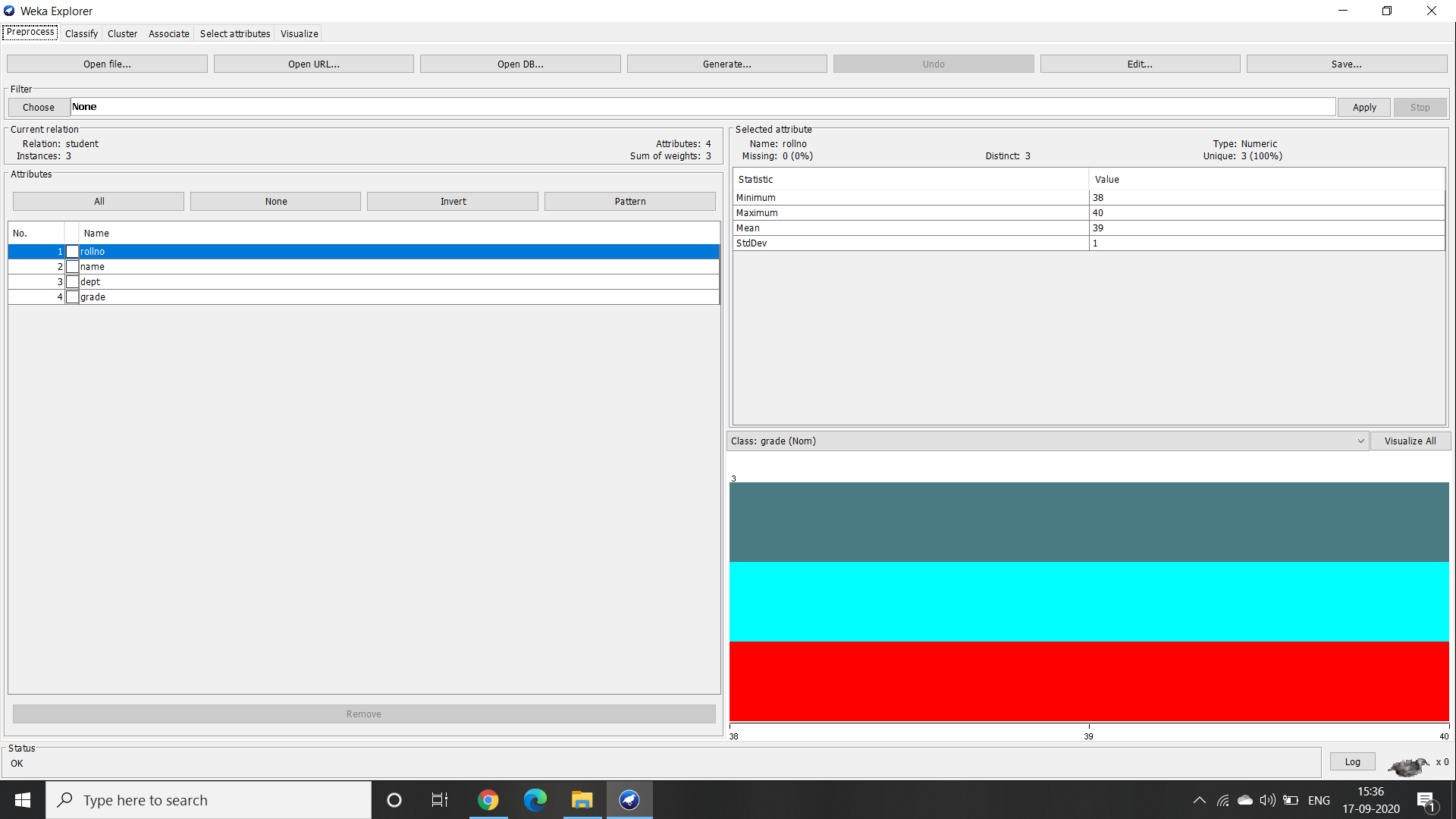
Create your own dataset. Here in the screenshot below a sample student dataset has been created.



=> Save the file as “ student.arff ”.

Step 3:

Now opening the file in weka.



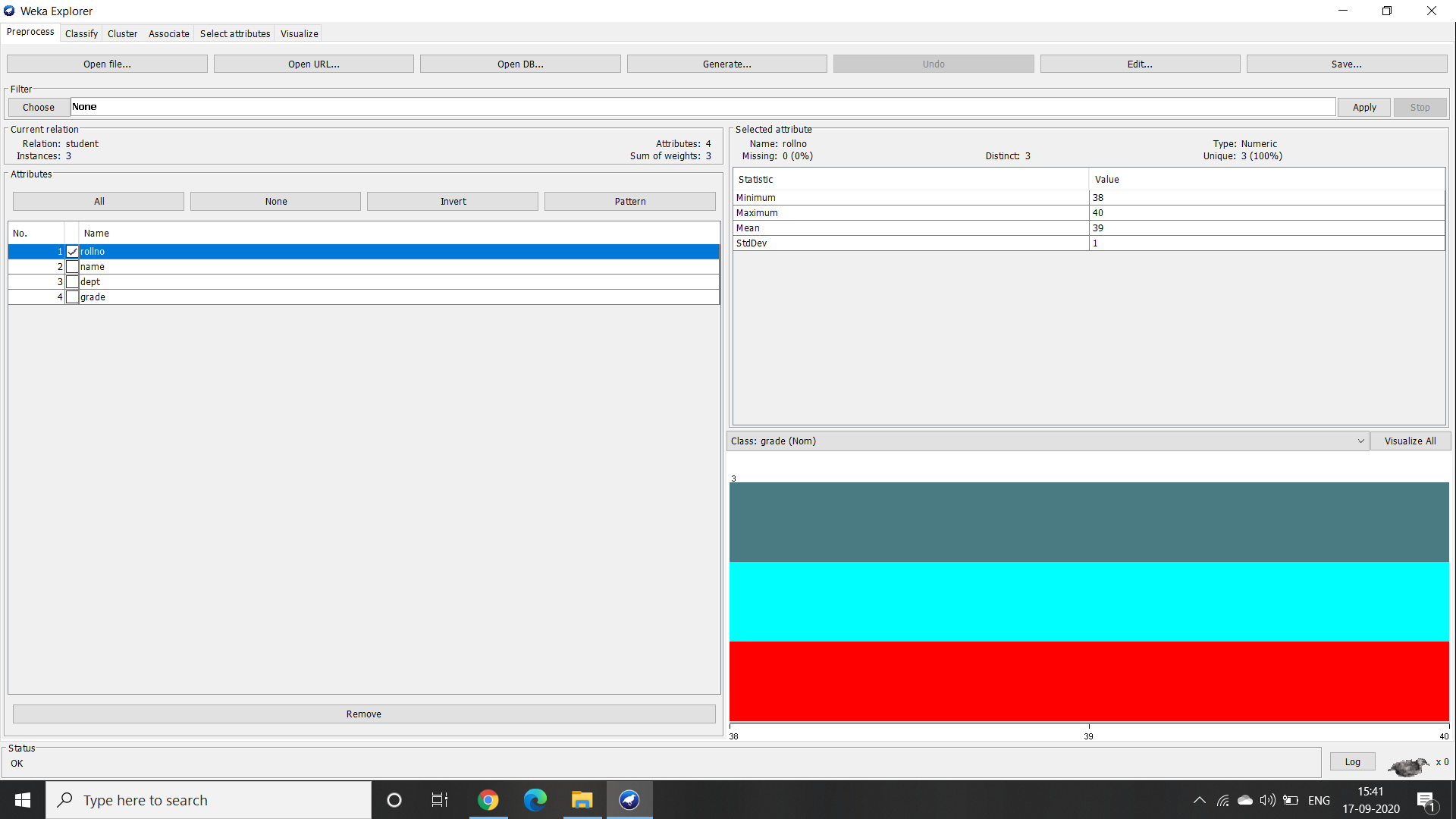
We can see that there are four attributes namely :

1. Roll number
2. Name
3. Department
4. Grade

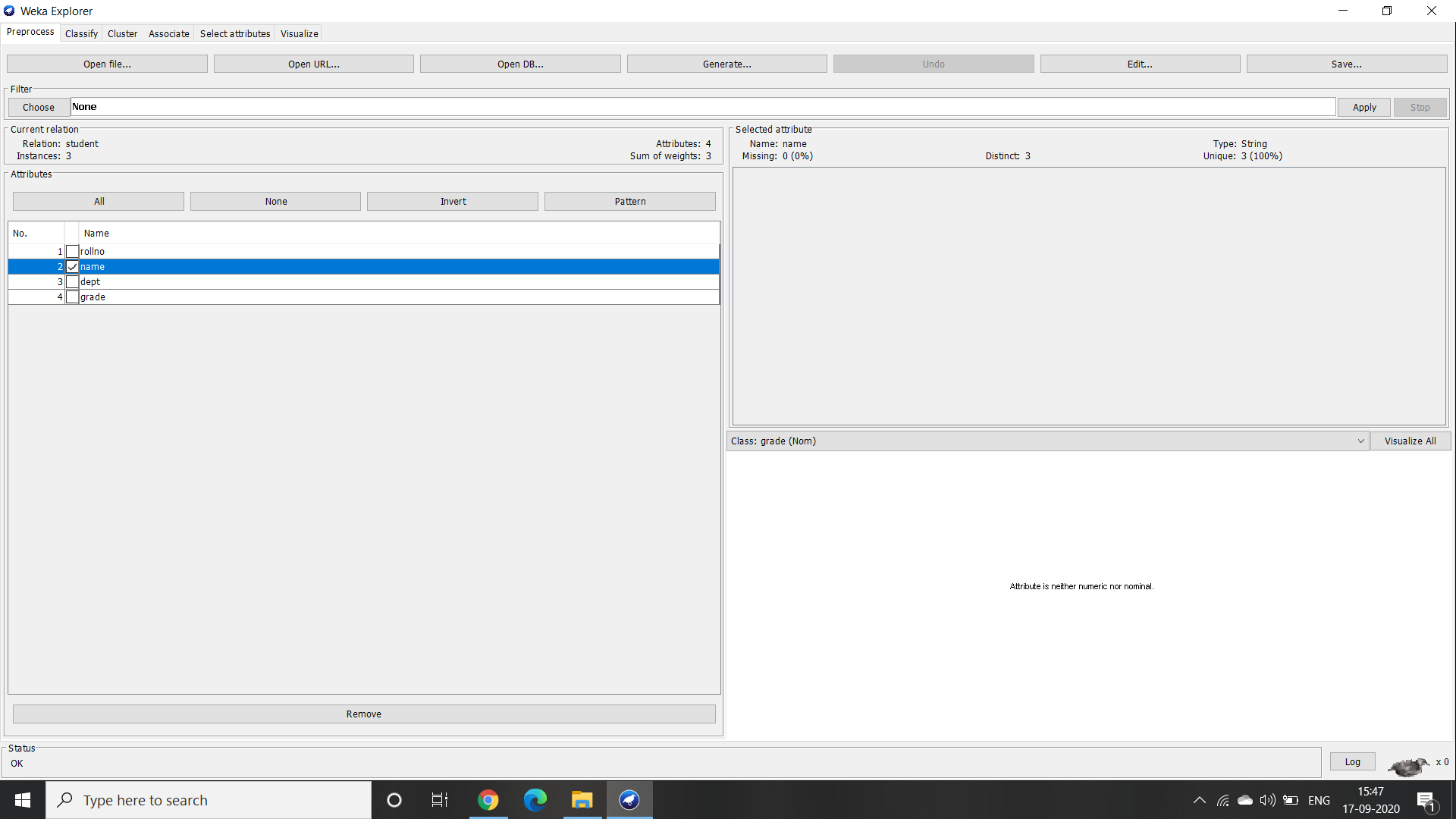
Step 4 :

Analysing the selected attributes :

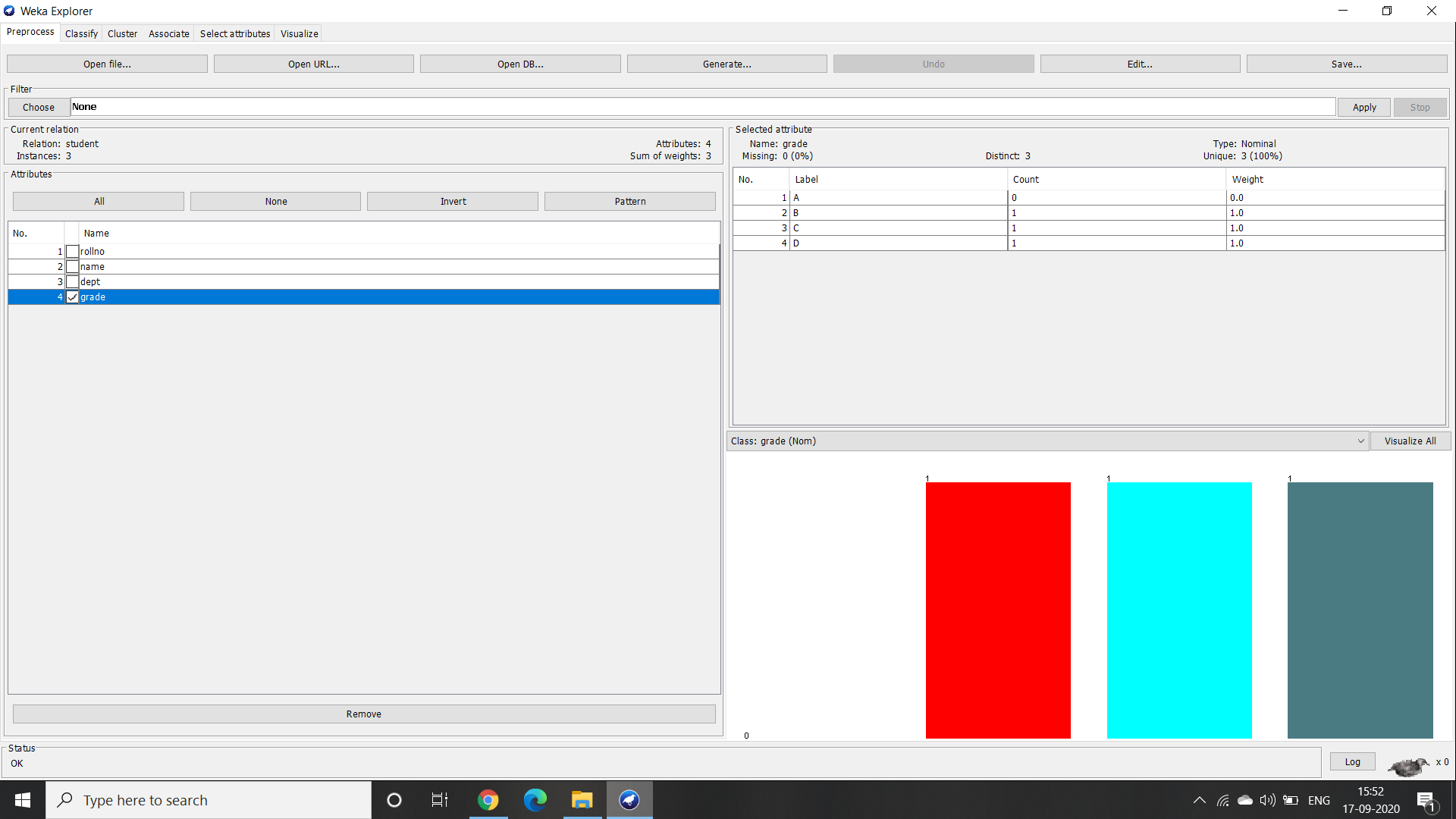
1. Roll Number(numeric) : We can see that the dataset has 40 as its highest roll number, 38 as the least and 39 is the median.



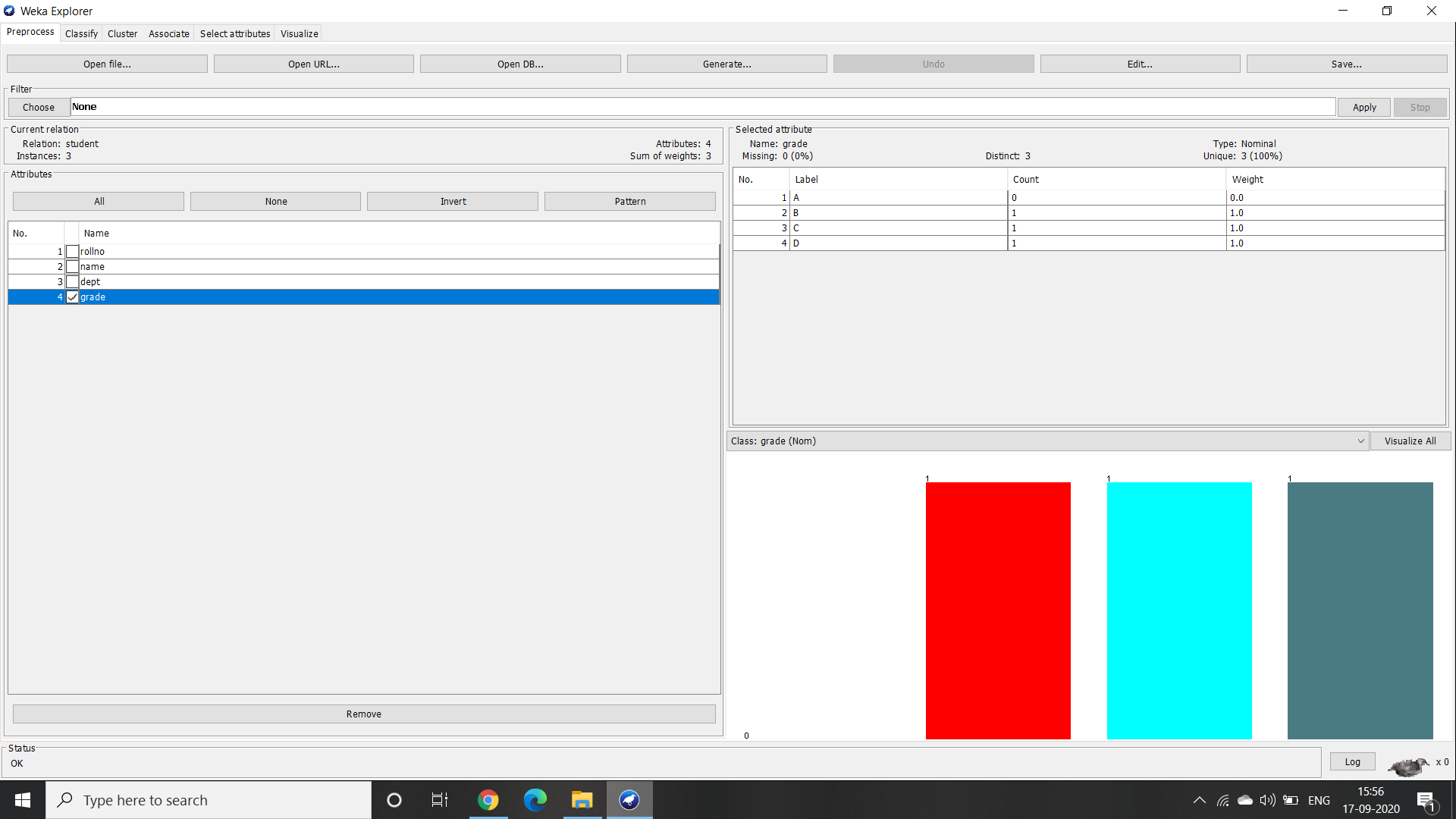
1. Name(string) : When name attribute is selected a message gets displayed saying that attribute is neither numeric nor nominal.



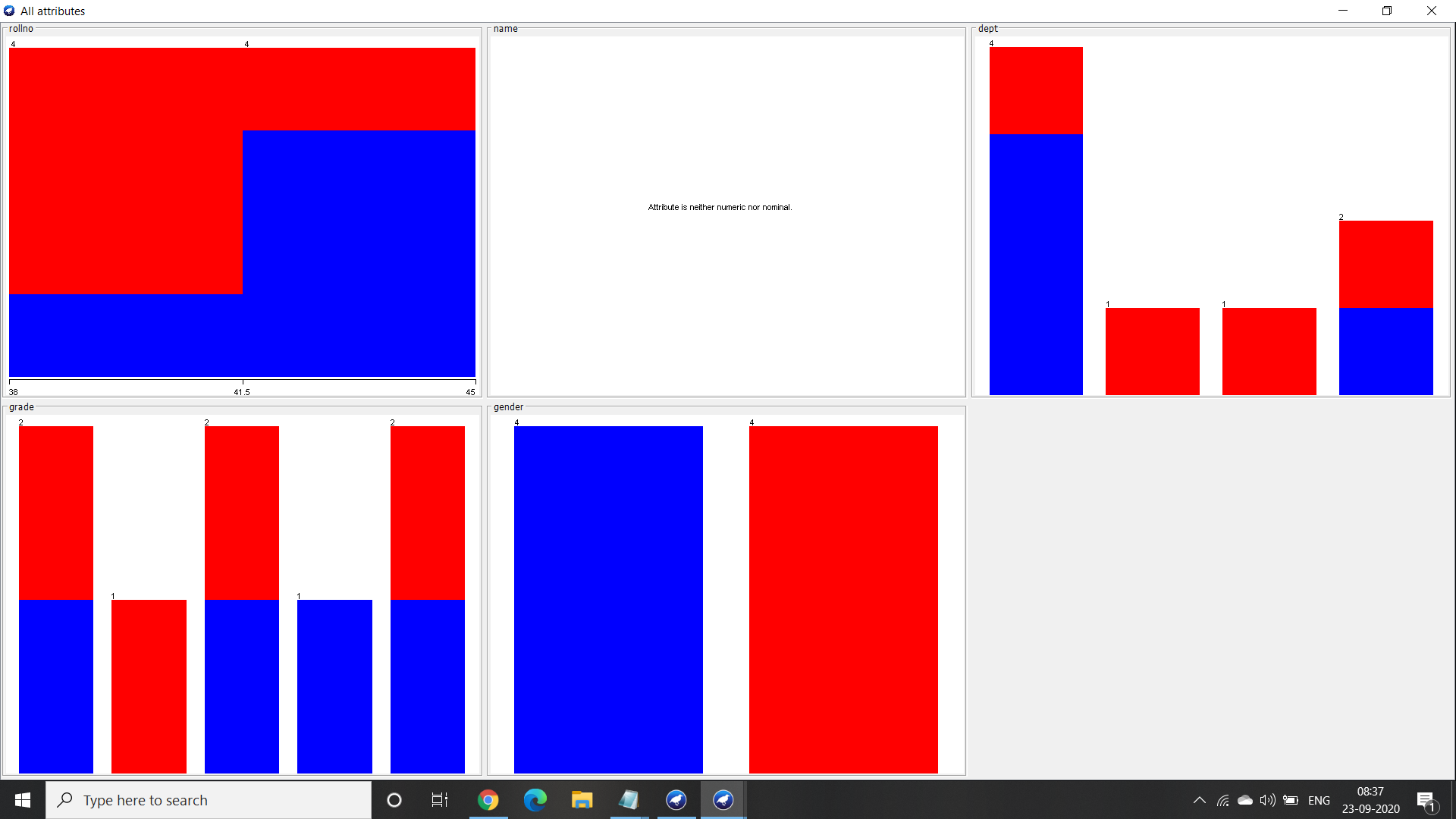
1. Department(nominal) : When a dept attribute is selected the conclusions that can be drawn are there is no person in the dataset with dept ‘cse’. Remaining dept has got one member each.



1. Grade(nominal) : We can observe that there is no person with ‘A’ grade. Distinct attribute’s count is three which implies that every person has got a unique grade.



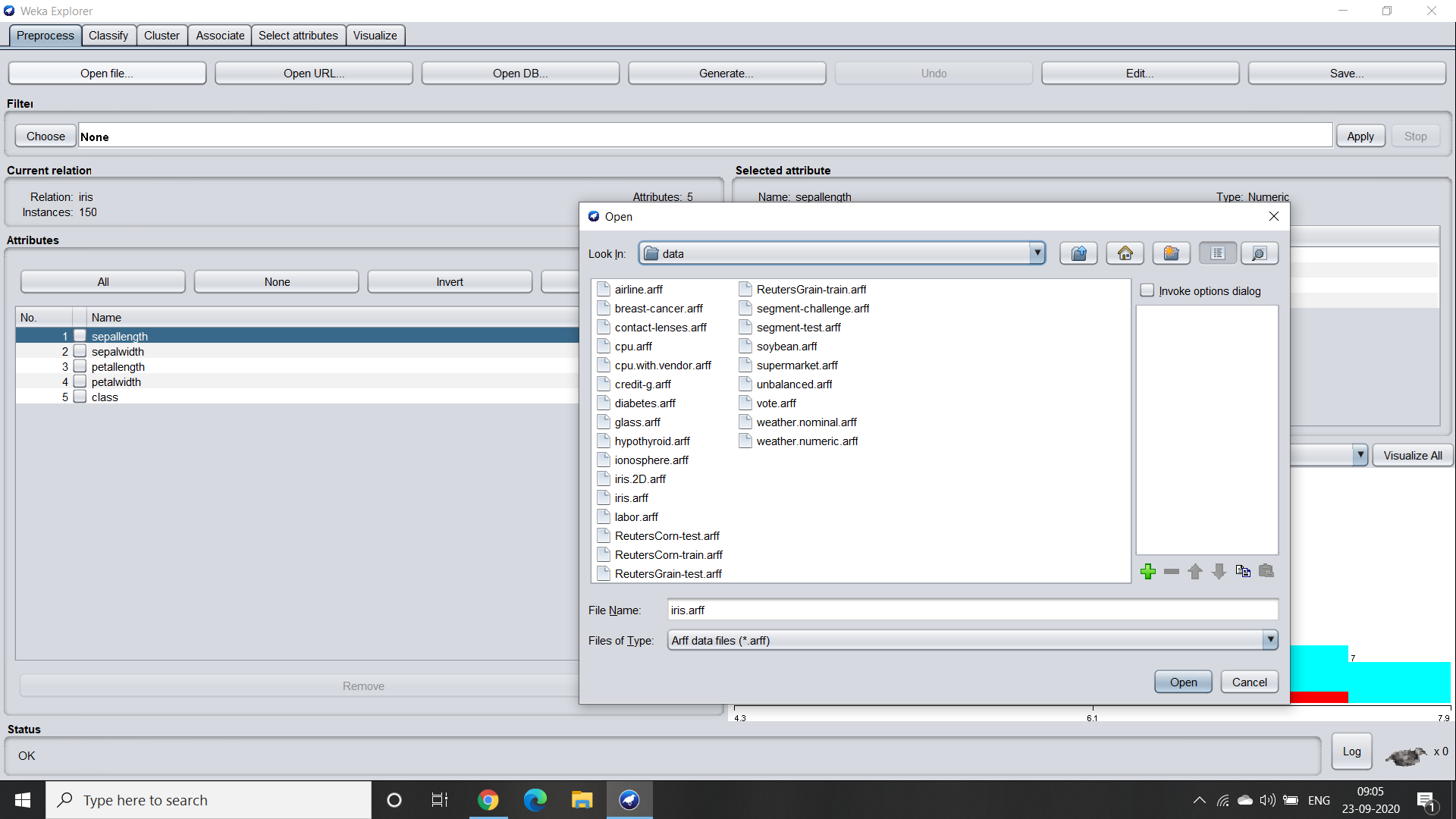
1. Missing Attribute : kjdgnkvsjkdsfkfsnjsgfjksdfvksdjgksndgknsdjgnksdgnksngksdngksngnsdkgnksdgsdngksdgksdngksngskdgngks
2. Visualising the records present in the student dataset.



The above picture shows

**2. Exploration of Iris dataset :**

Step 1 : Click on the open file and search for the data folder in weka. Now select the iris dataset. The total number of instances are 150.



Step 2 : This dataset has got 5 attributes namely :

Sepal length

Sepal width

Petal length

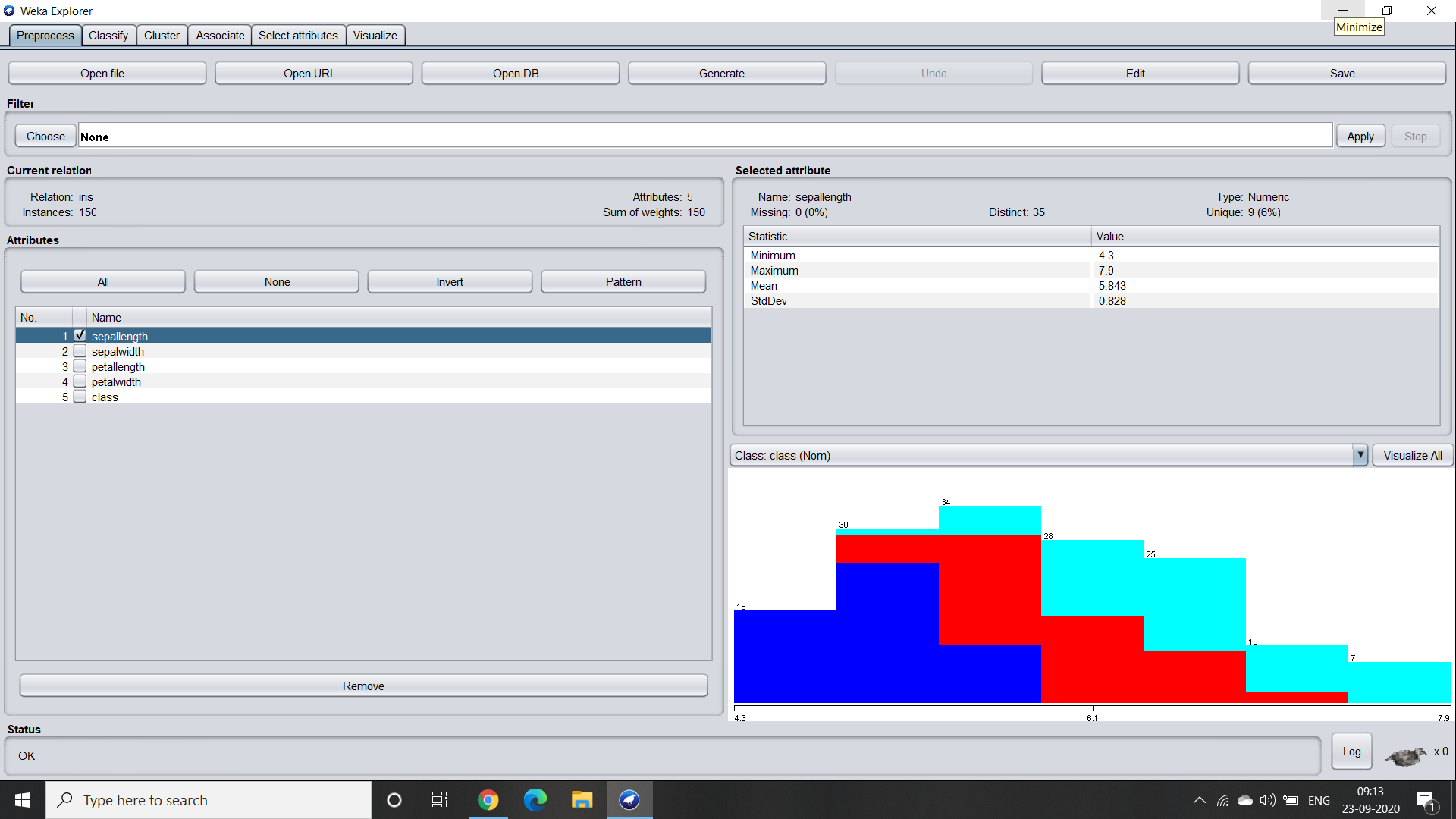
Petal width

Class

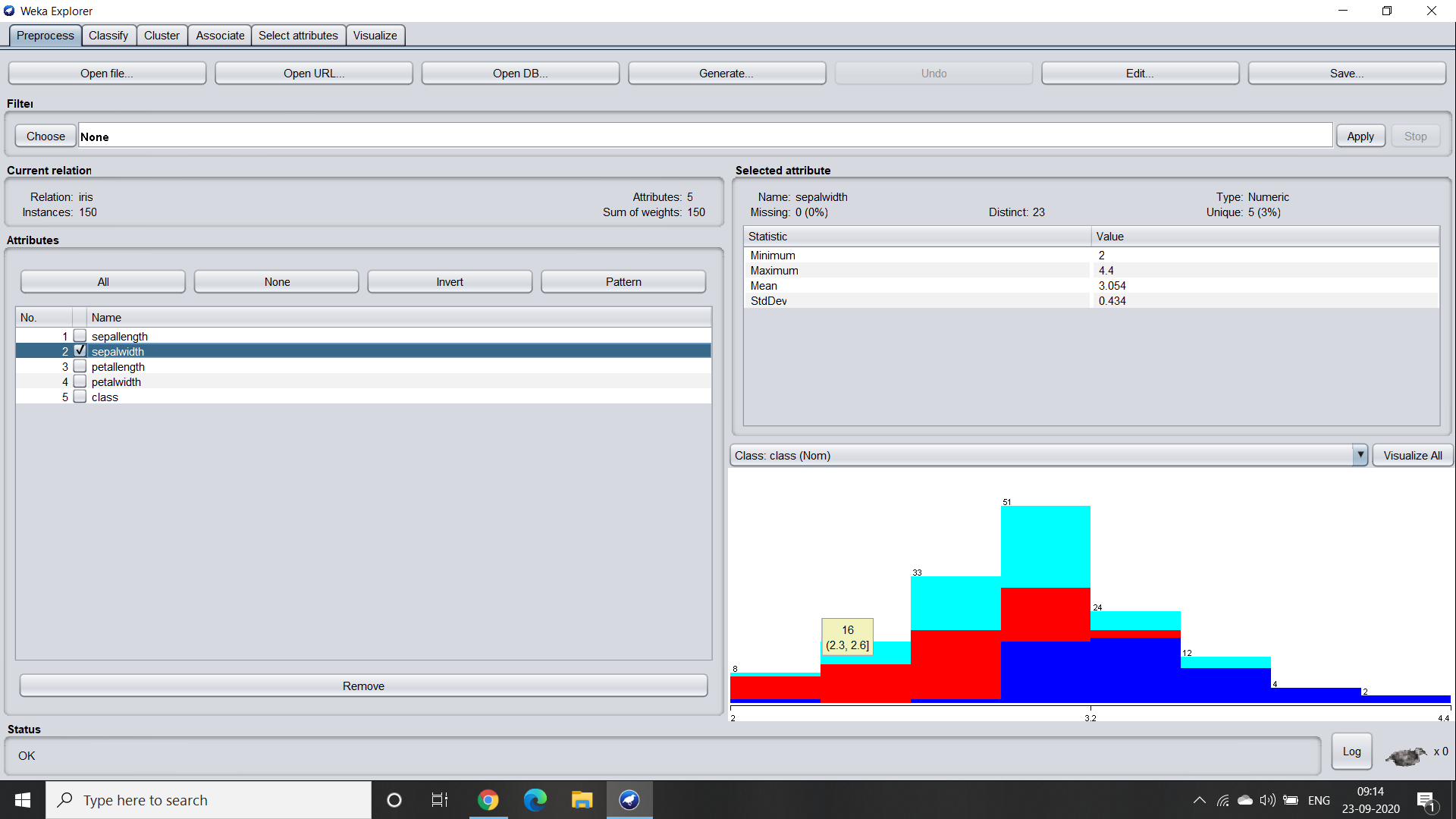
Step 3 :

Analysing the attributes :

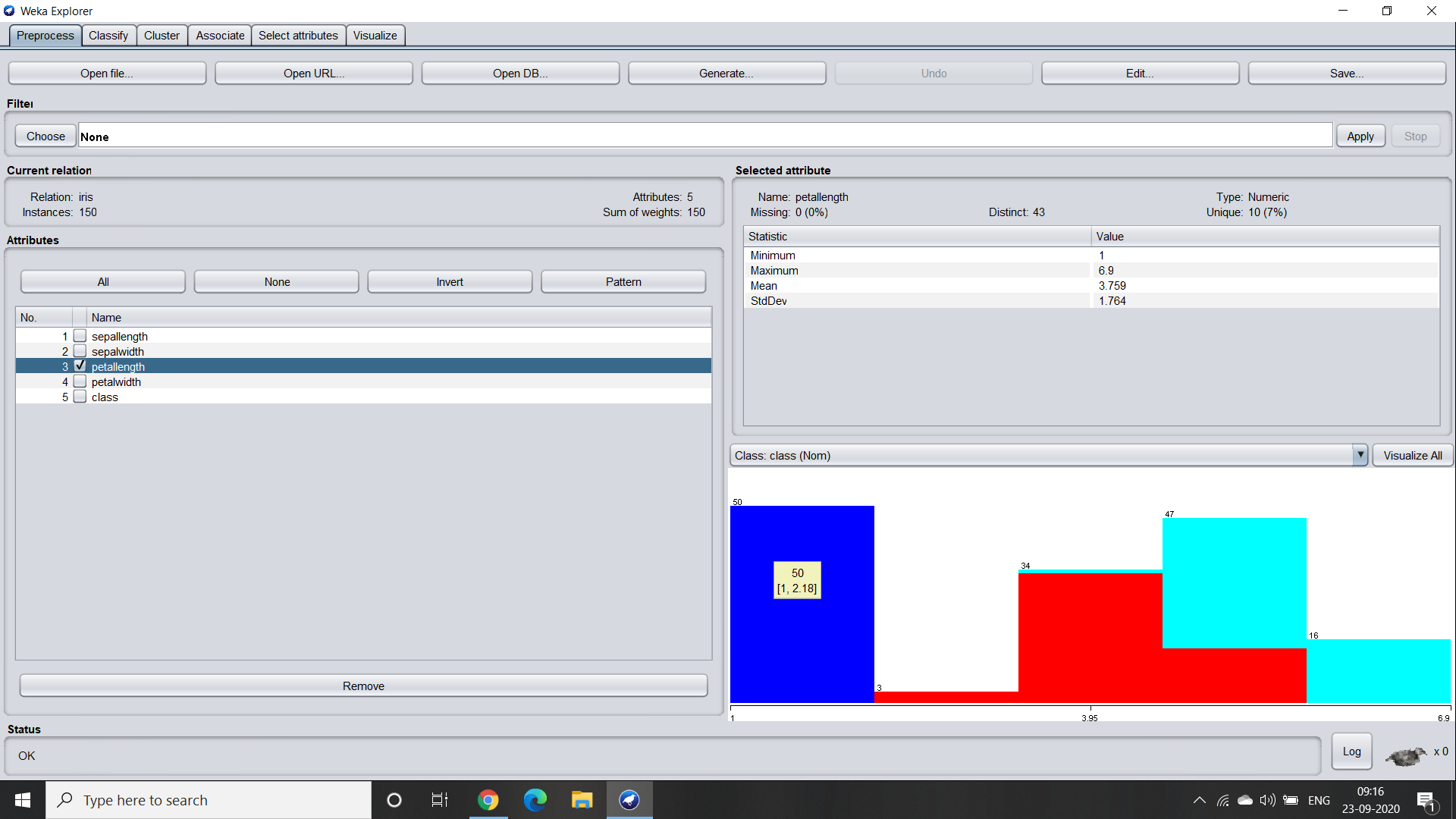
1. Sepal length (numeric) - There are few statistics calculated for the sepal length. Where, the distinct count is 35 and the unique percentage is 6.



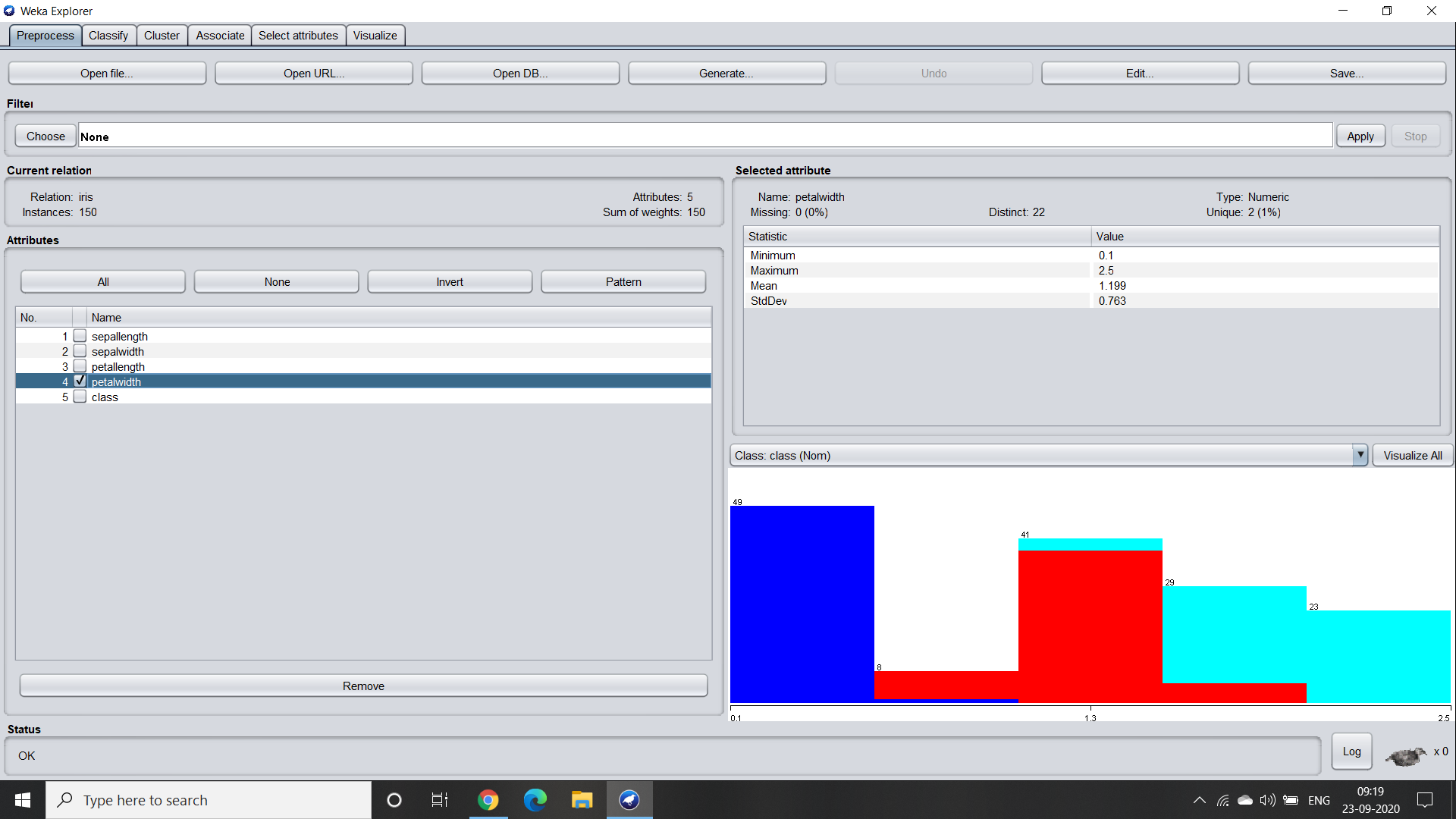
1. Sepal width (numeric) - There are few statistics calculated for the sepal length. Where, the distinct count is 23 and the unique percentage is 3.



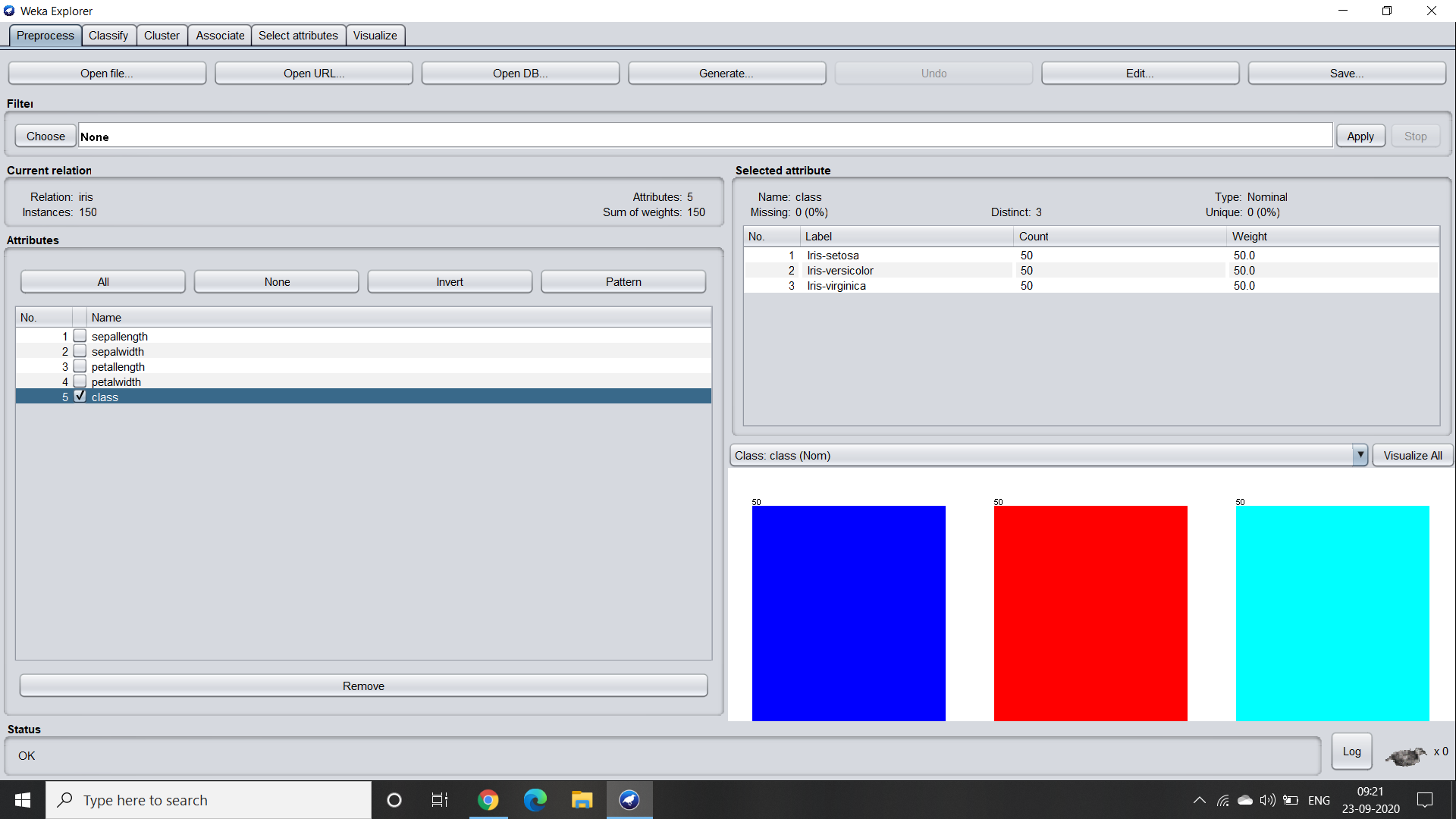
1. Petal length (numeric) - Similarly, for petal length also few statistics calculated for the sepal length. Where, the distinct count is 43 and the unique percentage is 7.



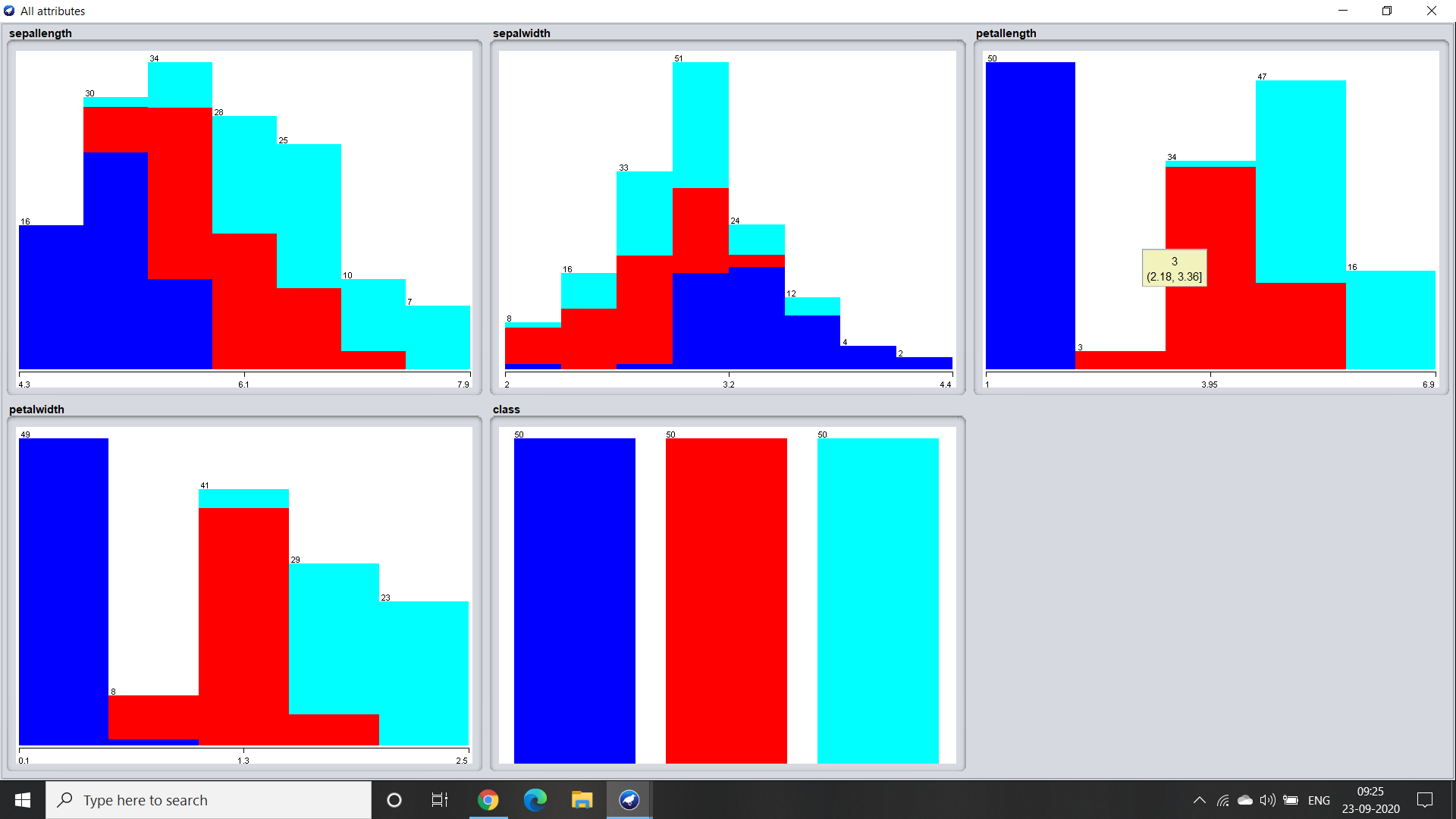
1. Petal width (numeric) - The petal width also has few statistics calculated for the sepal length. Where, the distinct count is 22 and the unique percentage is 1.



1. Class (nominal) - All the flowers fall under three categories that are setosa, virginica and versicolor.



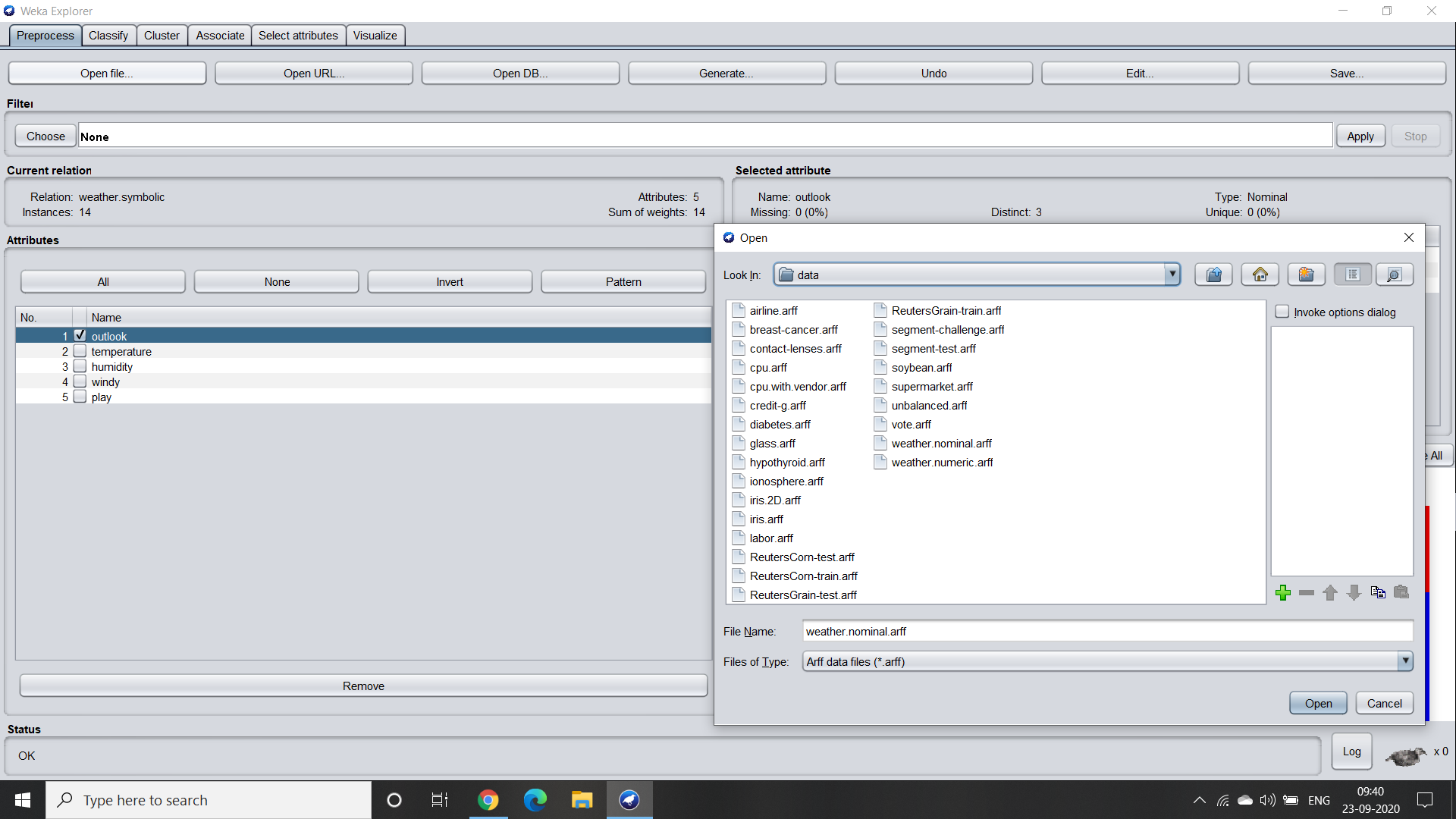
1. Visualisation of the dataset :



From the picture we can see that petal length of setosa is unique whereas the other two classes have some commonalities in their lengths. While considering the sepal length, width and petal width all the classes have some similarities in their sizes.

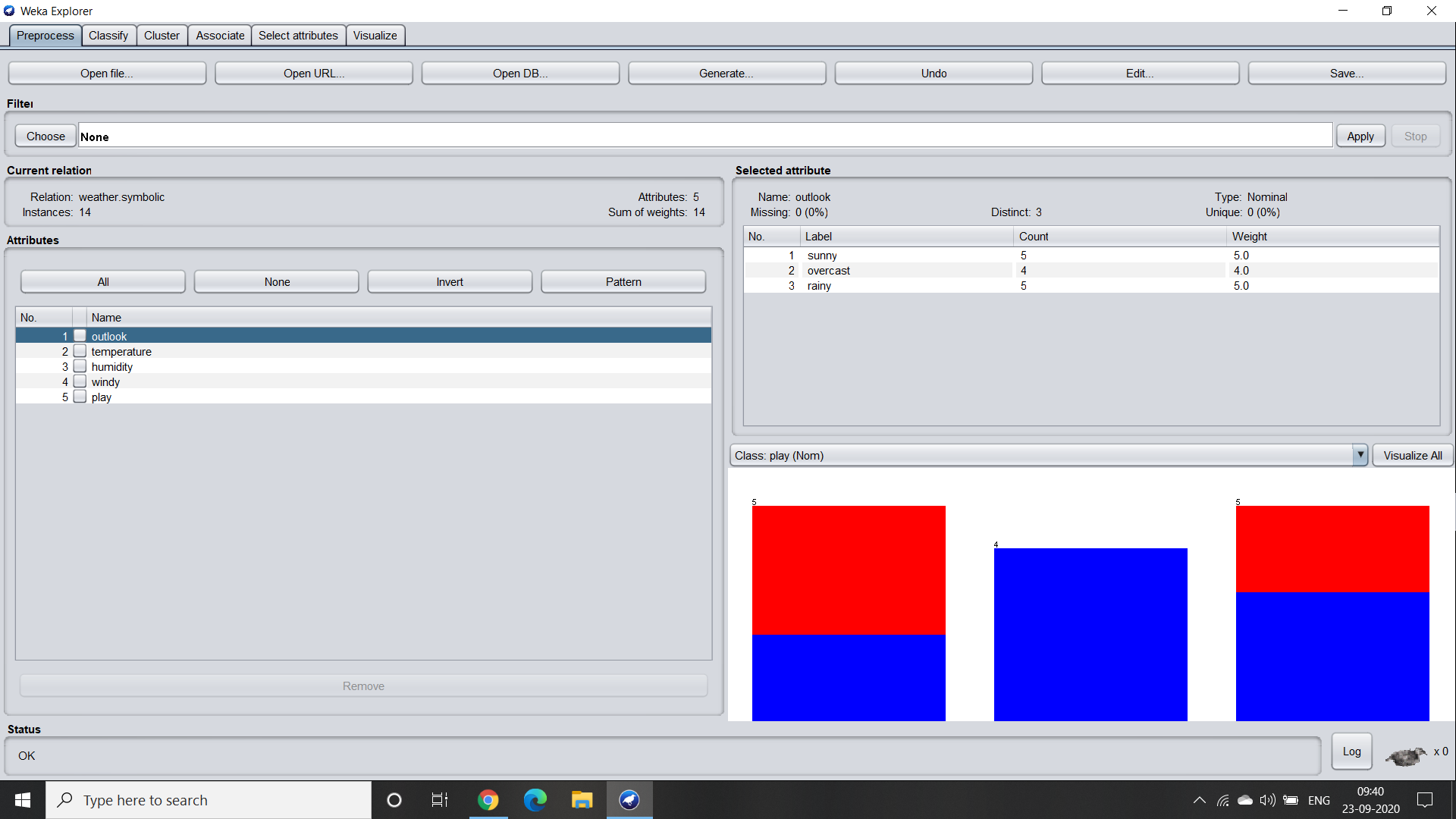
**3. Exploration of weather-nominal dataset :**

Step 1 : Click on the open file and search for the data folder in weka. Now select the weather nominal dataset. Total number of instances are 14.



Step 2 : This dataset has got 5 attributes namely :

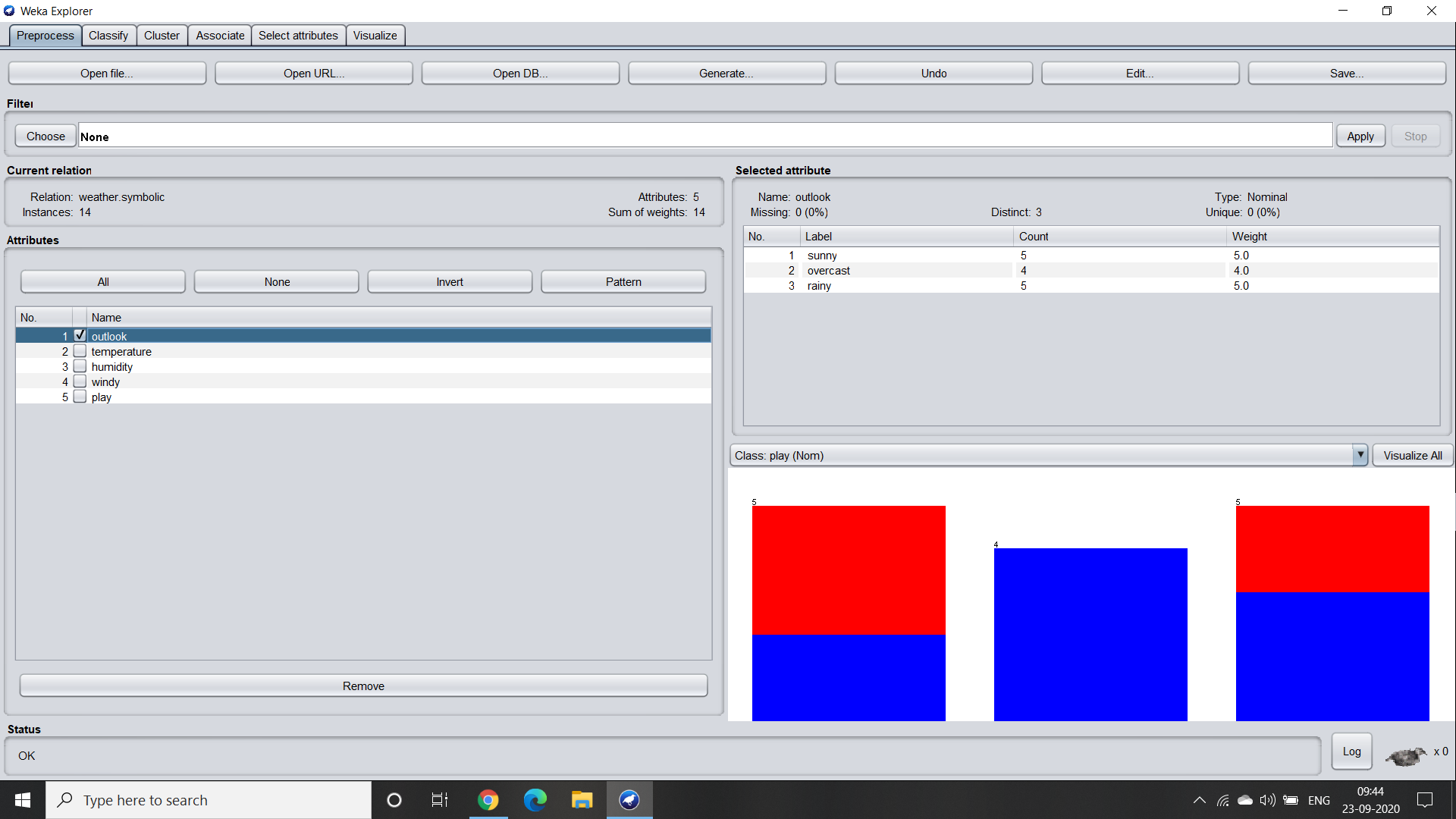
1. Outlook
2. Temperature
3. Humidity
4. Windy
5. Play



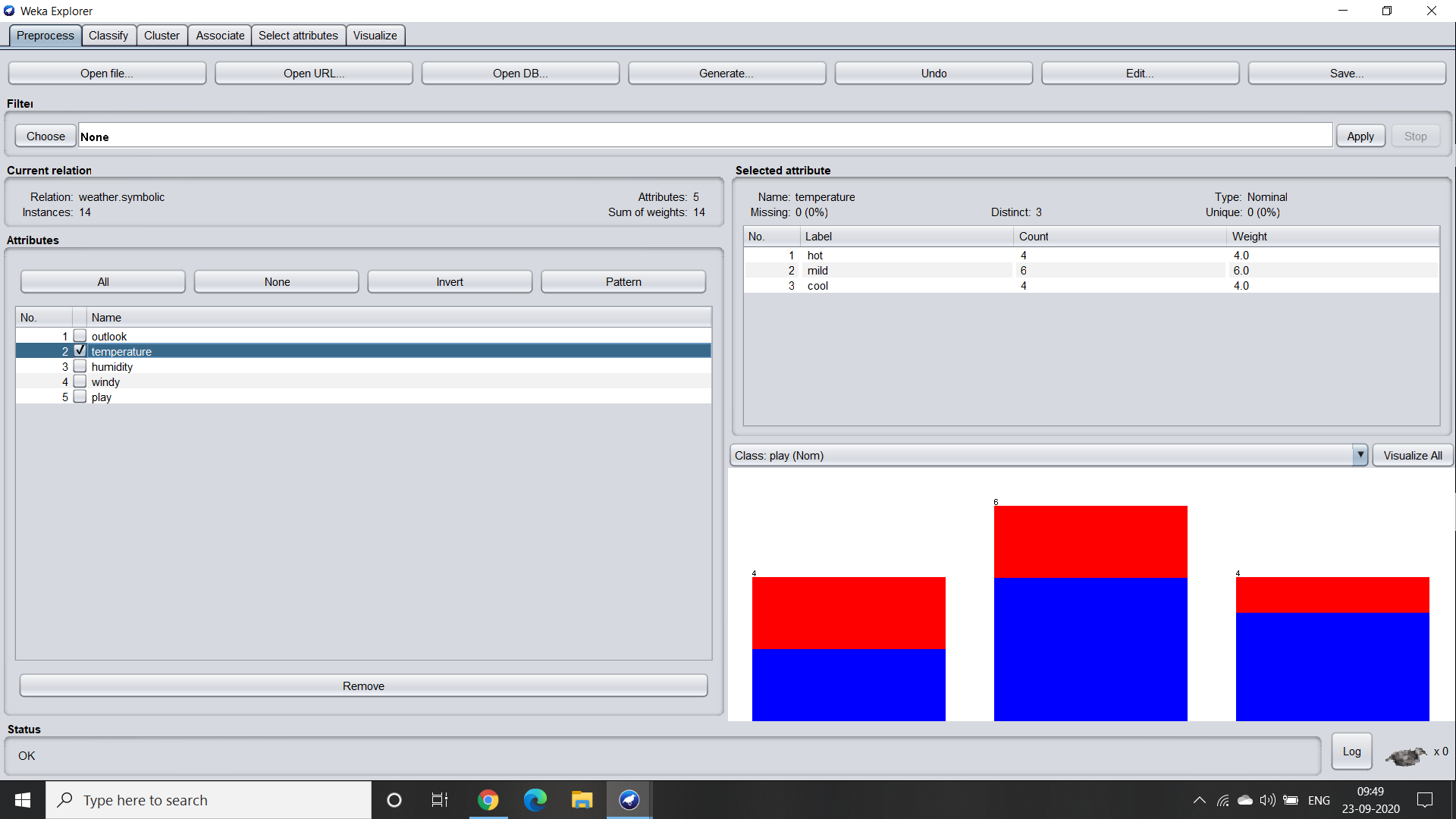
Step 3 :

Analysing the attributes :

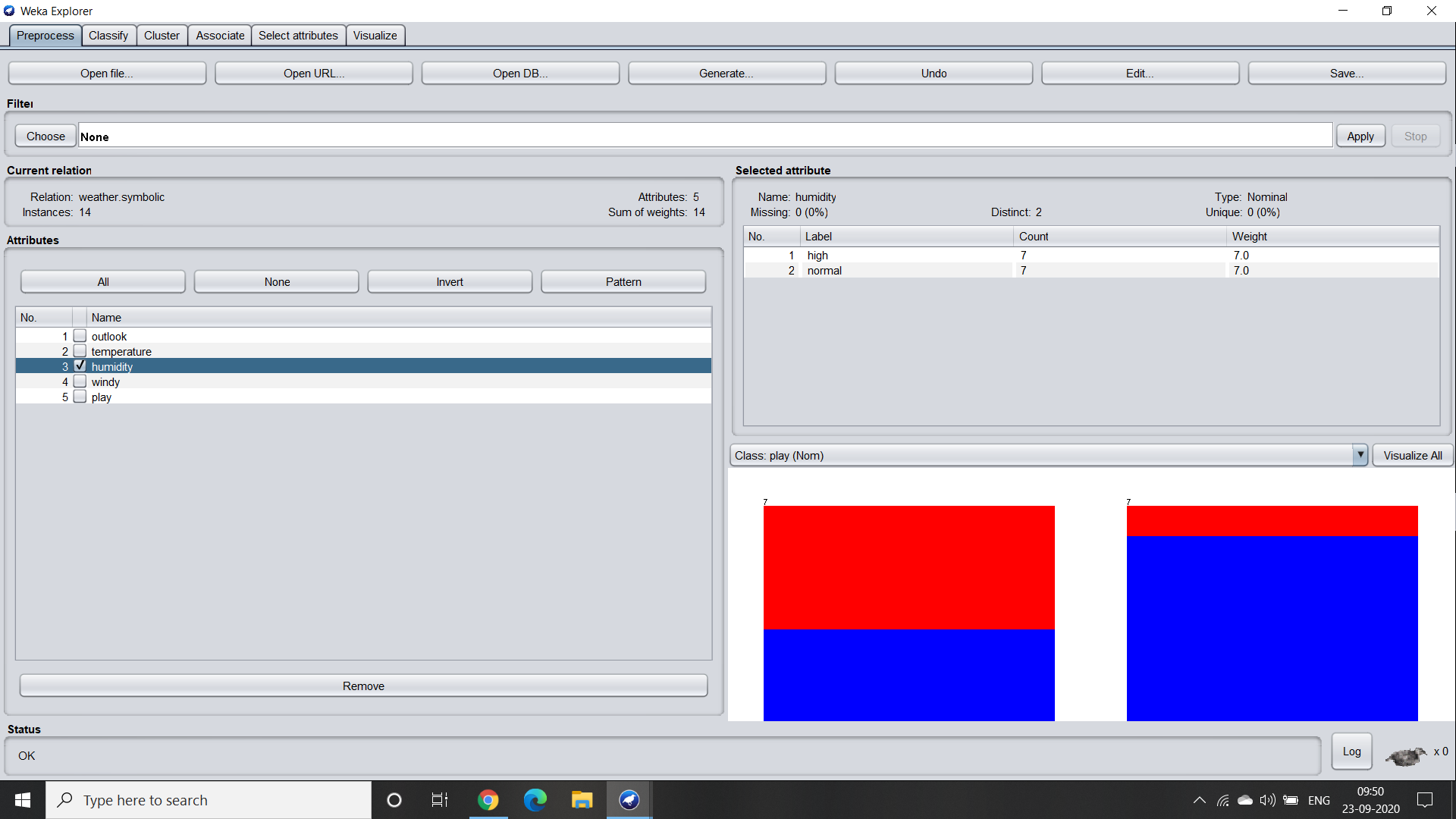
1. Outlook - There are three distinct labels namely sunny, overcast and rainy.



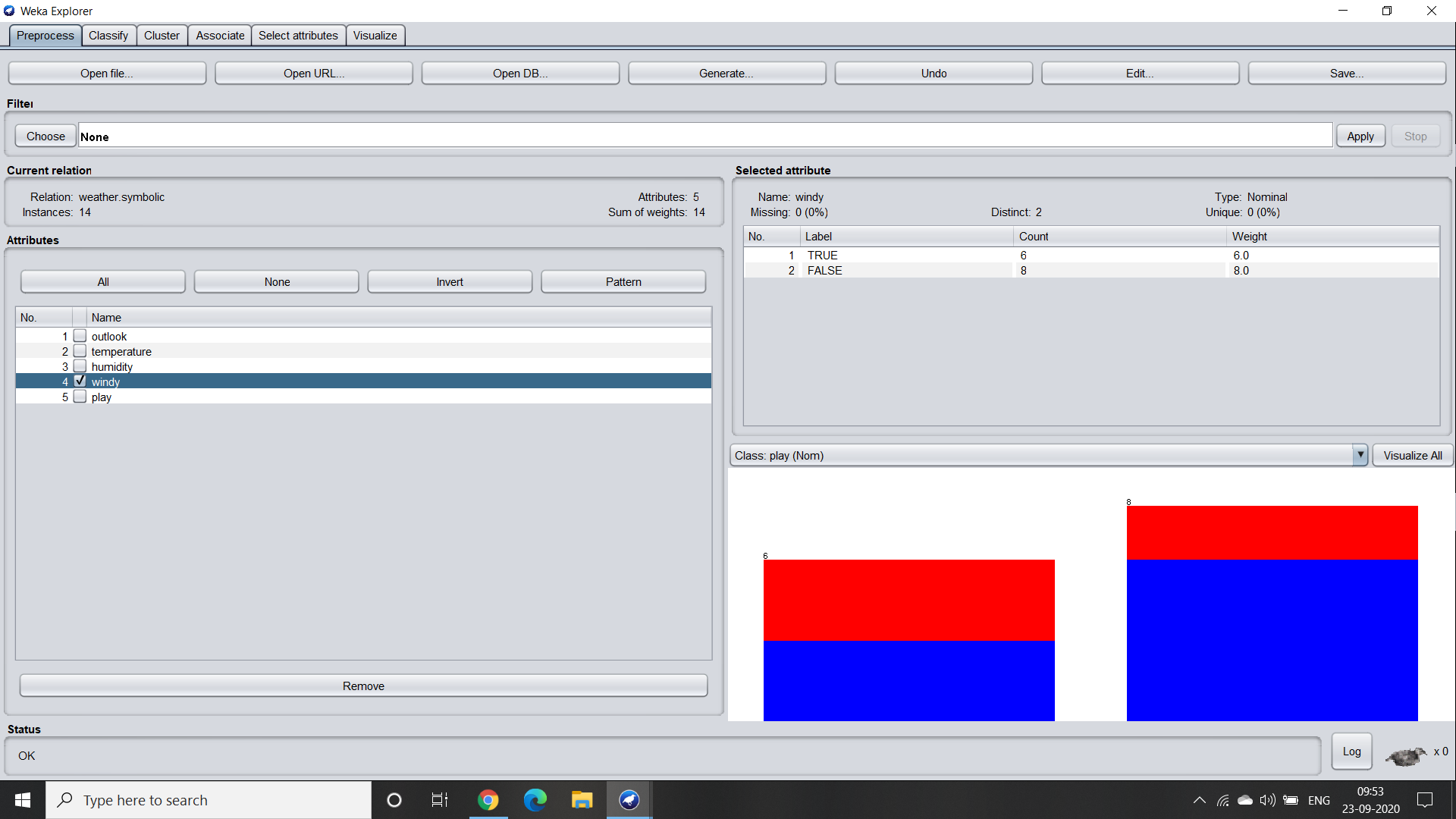
1. Temperature - This attribute also has three distinct labels namely hot, mild and cool.



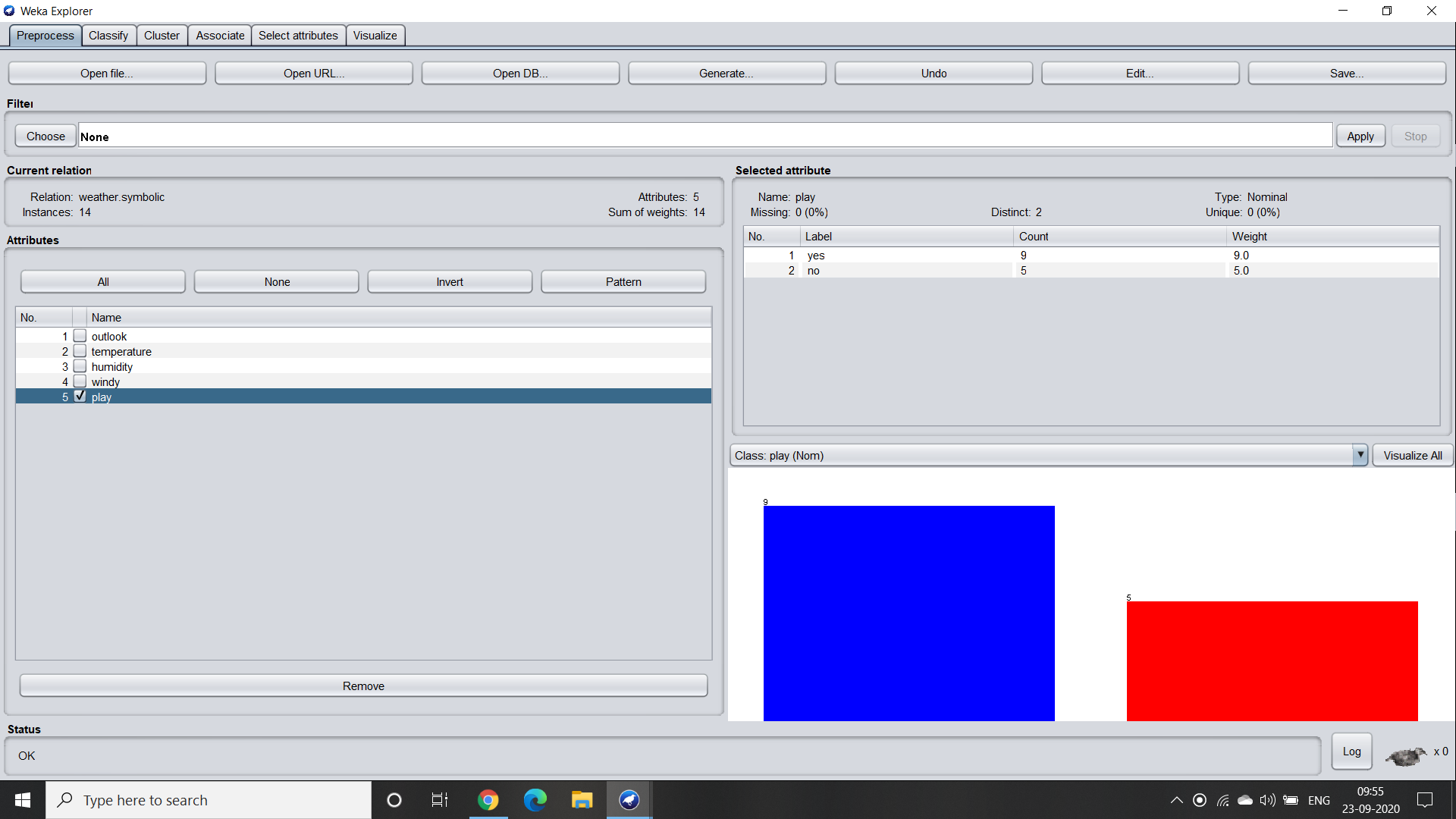
1. Humidity - In this there are two distinct labels such as high and normal with a equal number of count 7.



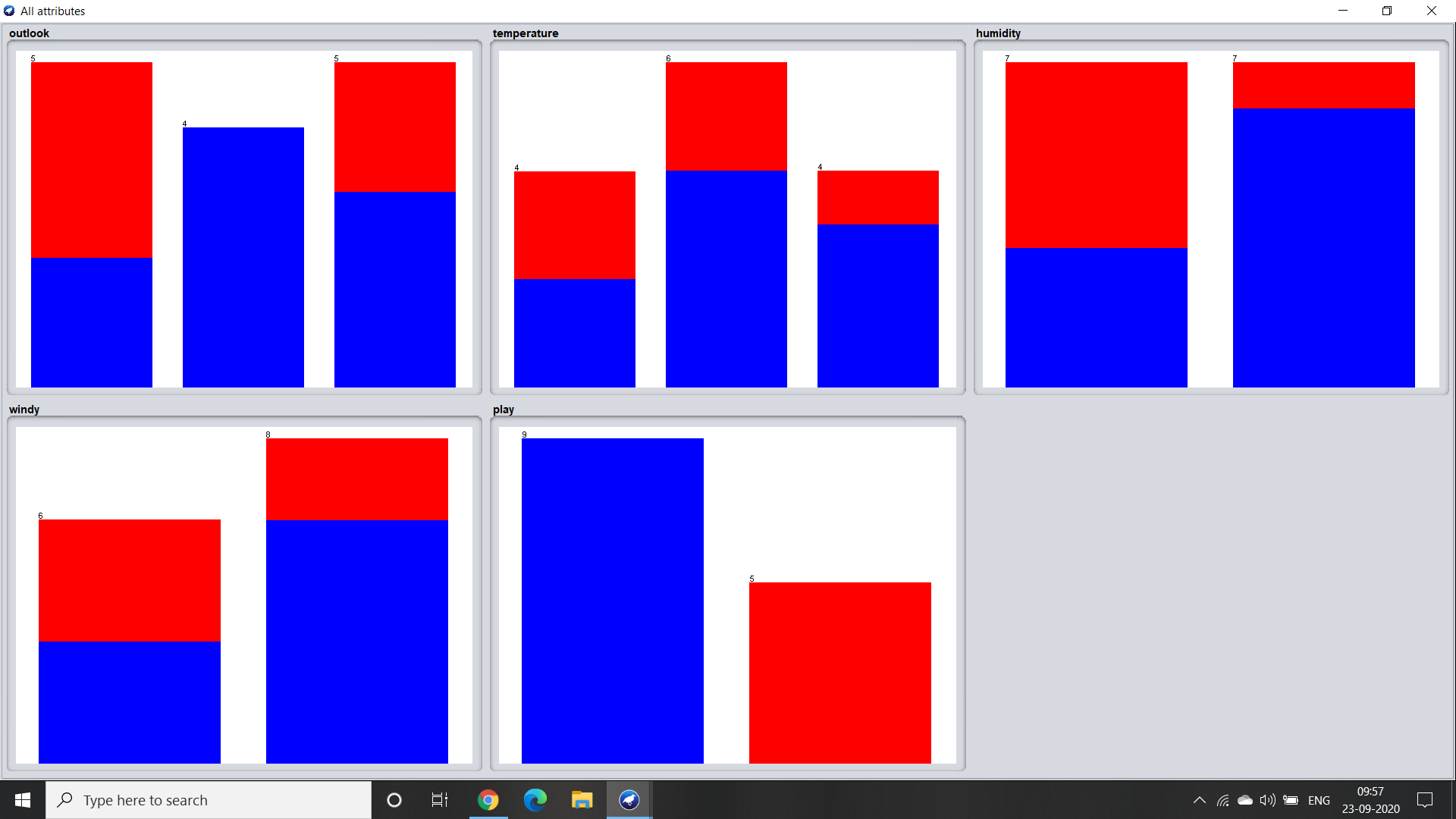
1. Windy - This particular attribute describes whether the day was windy or not. Hence, the labels are True or False.



1. Play - This attribute describes how likely one can go out and play. There are two labels that are yes or no.



1. Visualisation of the dataset :



From the above picture we can infer that in most of the cases although there are variations observed in the temperature yet the day still remains cloudy.