

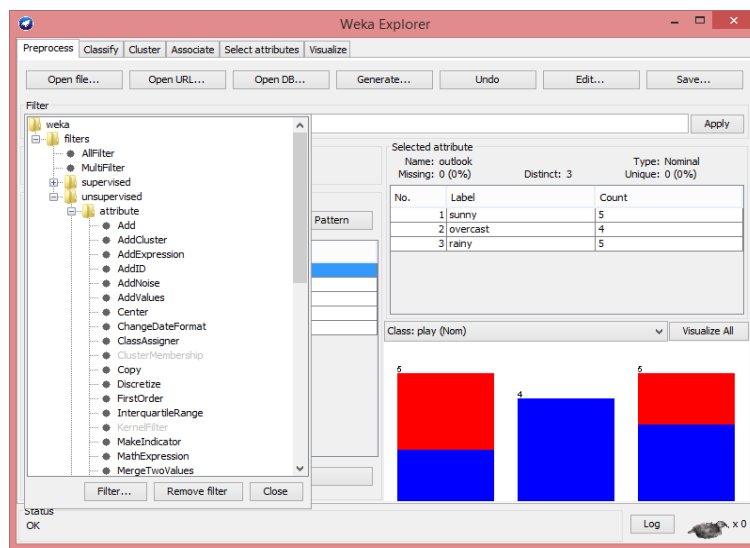
Lab Exercise One

Data Preprocessing with WEKA Explorer

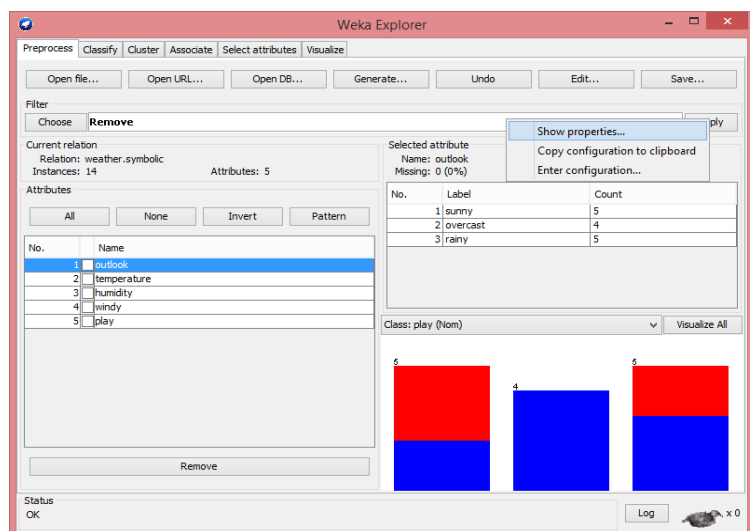
Using Filters

Unsupervised Attribute Filter – Remove: This filter deletes specified attributes from a dataset. The same effect can be achieved more easily by selecting the relevant attributes using the tick boxes and pressing the **Remove** button.

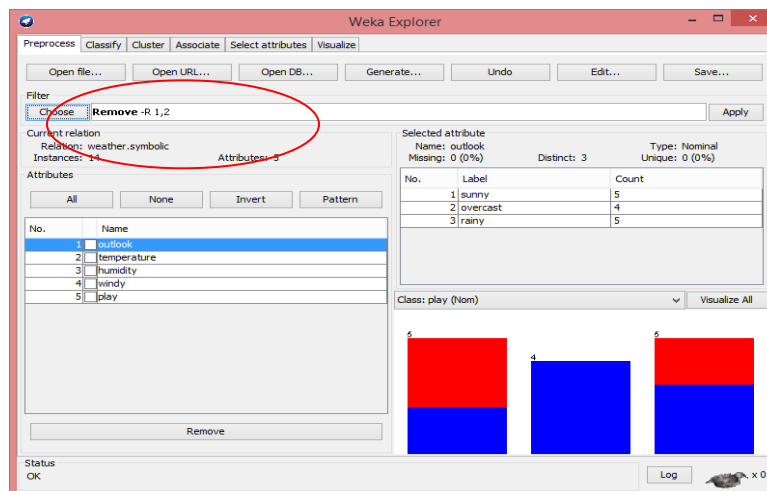
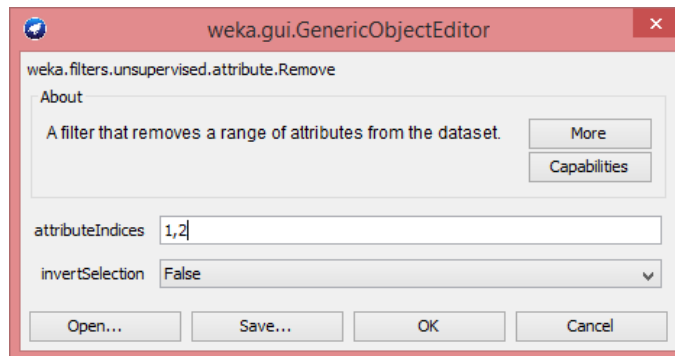
1. Open a data file, such as *weather.nominal* dataset.
2. Click Choose button under Filter. Click filters -> unsupervised -> attribute -> Remove.



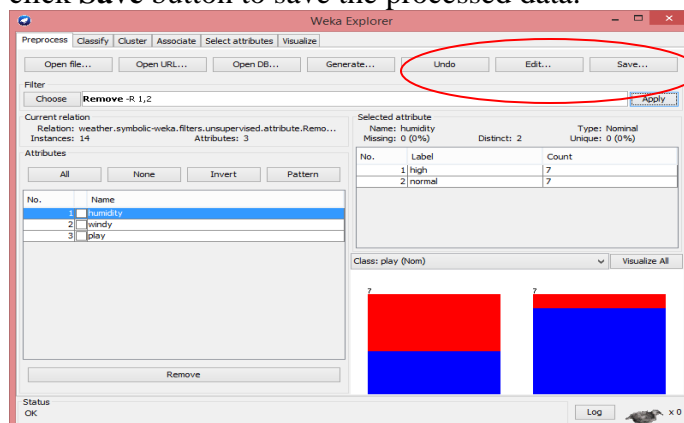
3. Right click the Remove box, choose Show Properties.



4. There are two options for Remove filter. One option is **attributeIndices** that specifies the range of attributes to be removed. (We set **1,2** in this example – outlook and temperature). The other option is **invertSelection** that determines whether the filter select or deletes attributes (We set **False**, which means to remove rather than retain them). Then you click **OK**.



5. Click **Apply** button next to the box of **Remove** filter, first two attributes are removed from the dataset, only three are left. You could click **Undo** button to reverse the filtering operation and restore the original dataset. You could also click **Save** button to save the processed data.



Handling missing data

Unsupervised Attribute Filter – ReplaceMissingValues: This filter replaces all missing values for nominal and numeric attributes with the modes for nominal attributes and means for numeric attributes of the training data.

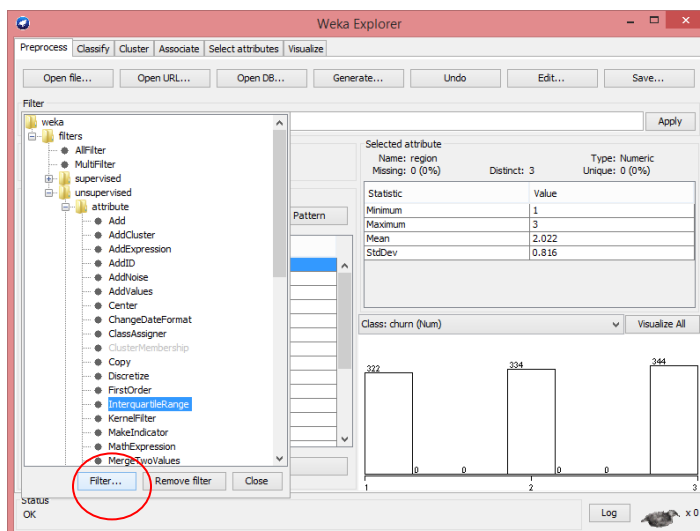
1. Open the dataset – **weather.numeric**. Click **Edit** button to view the raw data. You could find that two attributes have missing values.

Relation: weather

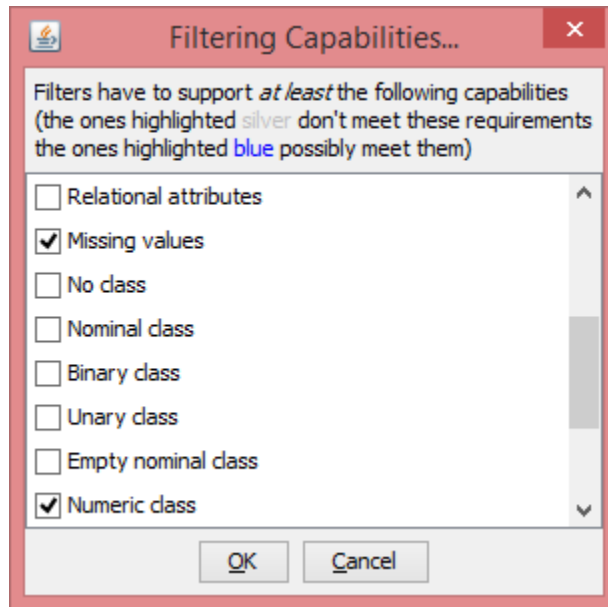
No.	outlook Nominal	temperature Numeric	humidity Numeric	windy Nominal	play Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy		96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny	69.0		FALSE	yes
10	rainy	75.0	80.0	FALSE	yes
11	sunny	75.0	70.0	TRUE	yes
12	overcast	72.0	90.0	TRUE	yes
13	overcast	81.0	75.0	FALSE	yes
14	rainy	71.0	91.0	TRUE	no

Buttons: Undo, OK, Cancel

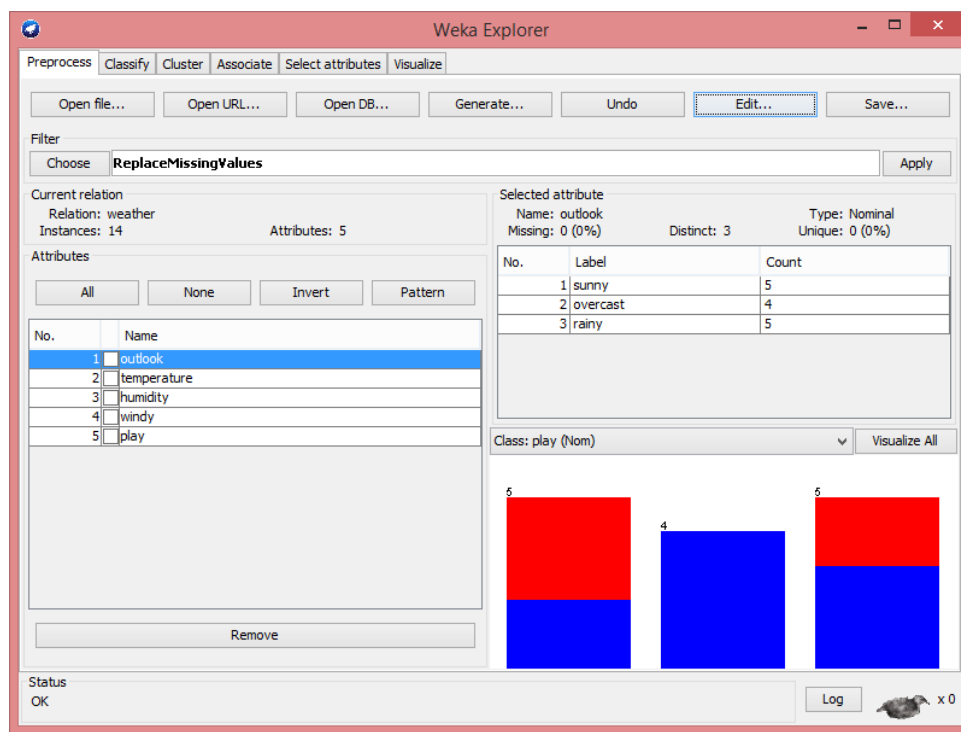
2. Click **Choose** button under **Filter**. Click **Filter** button at the bottom of the drop-down window.

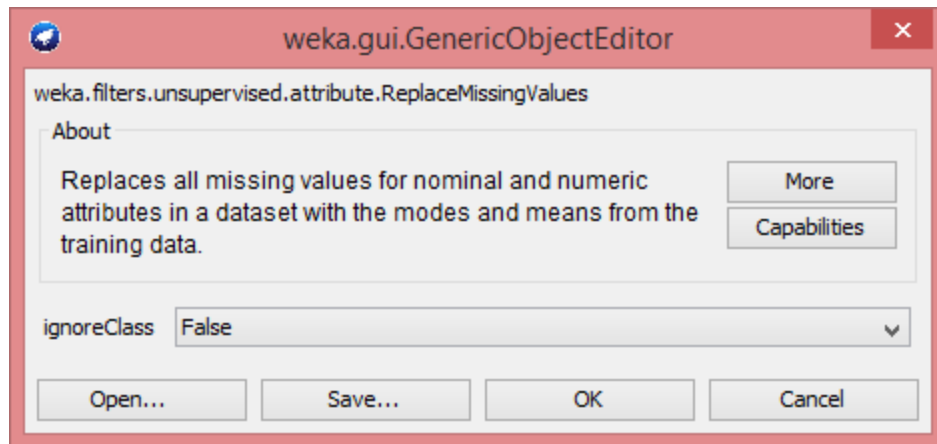


3. A window called Filtering Capabilities opens. This window shows what kind of attributes that filters support. Make sure that only **Numeric Attributes**, **Missing values** and **Numeric Class** are checked. Click **OK**.



4. Choose **ReplaceMissingValues** filter from the drop down list. Then left-click the box of the filter to show the properties window. Click **OK**.





5. Click **Apply** button at the end of the filter box. Then click **Edit** button to check the processed data set – missing values are filled. If you want to save the modified data, please click **Save** button on the main window. Choose a different file name to save it.

Viewer

Relation: weather-weka.filters.unsupervised.attribute.ReplaceMi...

No.	outlook Nominal	temperature Numeric	humidity Numeric	windy Nominal	play Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	73.846153...	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny	69.0	82.538...	FALSE	yes
10	rainy	75.0	80.0	FALSE	yes
11	sunny	75.0	70.0	TRUE	yes
12	overcast	72.0	90.0	TRUE	yes
13	overcast	81.0	75.0	FALSE	yes
14	rainy	71.0	91.0	TRUE	no

Undo OK Cancel