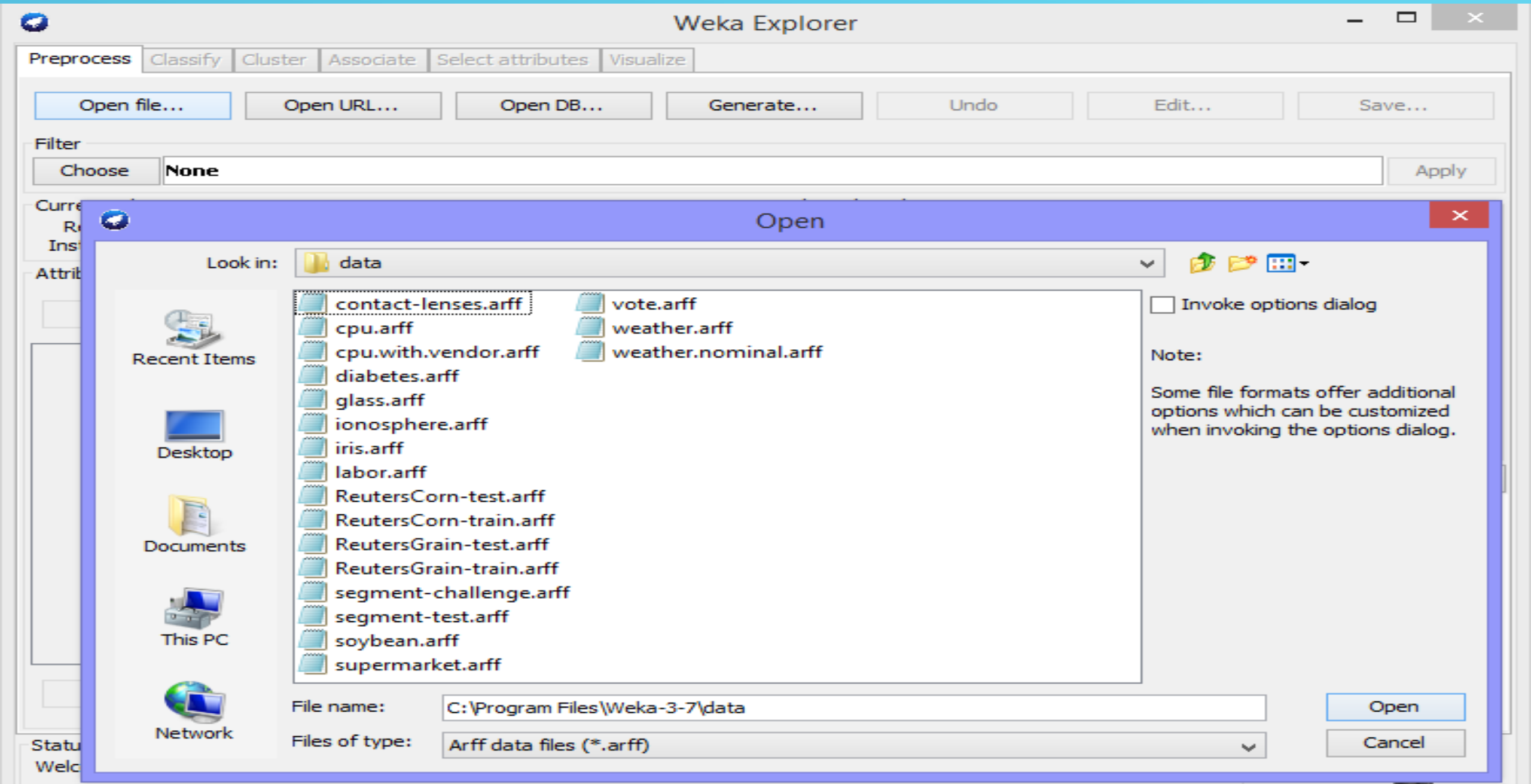
WEKA DATA MINING TOOL

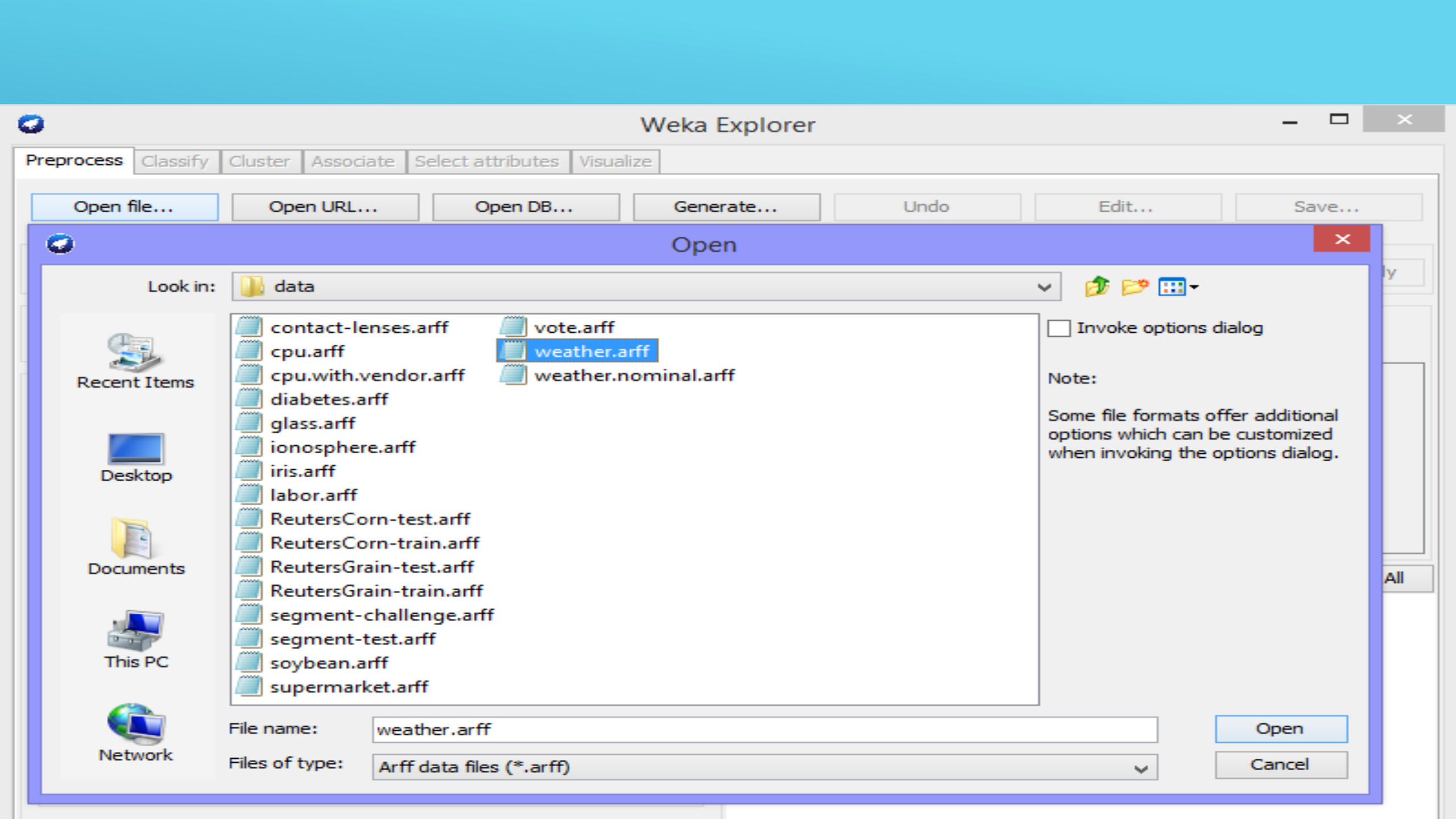
Data Loading, appending, merging and Preprocessing

- Preparing data for analyzing in WEKA

```
1  @relation weather
2
3  @attribute outlook {sunny, overcast, rainy}
4  @attribute temperature real
5  @attribute humidity real
6  @attribute windy {TRUE, FALSE}
7  @attribute play {yes, no}
8
9  @data
10 sunny,85,85,FALSE,no
11 sunny,80,90,TRUE,no
12 overcast,83,86,FALSE,yes
13 rainy,70,96,FALSE,yes
14 rainy,68,80,FALSE,yes
15 rainy,65,70,TRUE,no
16 overcast,64,65,TRUE,yes
17 sunny,72,95,FALSE,no
18 sunny,69,70,FALSE,yes
19 rainy,75,80,FALSE,yes
20 sunny,75,70,TRUE,yes
21 overcast,72,90,TRUE,yes
22 overcast,81,75,FALSE,yes
23 rainy,71,91,TRUE,no
```

WEKA TOOL GUI


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Open
Look in:
data
contact-lenses.arff
cpu.arff
cpu.with.vendor.arff
diabetes.arff
glass.arff
ionosphere.arff
iris.arff
labor.arff
ReutersCorn-test.arff
ReutersCorn-train.arff
ReutersGrain-test.arff
ReutersGrain-train.arff
segment-challenge.arff
segment-test.arff
soybean.arff
supermarket.arff
vote.arff
weather.arff
weather.nominal.arff
Invoke options dialog
Note:
Some file formats offer additional options which can be customized when invoking the options dialog.
Recent Items
Desktop
Documents
This PC
Network
File name:
C:\Program Files\Weka-3-7\data
Files of type:
Arff data files (*.arff)
Open
Cancel


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Open
Look in:
data
contact-lenses.arff
cpu.arff
cpu.with.vendor.arff
diabetes.arff
glass.arff
ionosphere.arff
iris.arff
labor.arff
ReutersCorn-test.arff
ReutersCorn-train.arff
ReutersGrain-test.arff
ReutersGrain-train.arff
segment-challenge.arff
segment-test.arff
soybean.arff
supermarket.arff
vote.arff
weather.arff
weather.nominal.arff
Invoke options dialog
Note:
Some file formats offer additional options which can be customized when invoking the options dialog.
Recent Items
Desktop
Documents
This PC
Network
File name:
weather.arff
Files of type:
Arff data files (*.arff)
Open
Cancel

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter

Choose | None | Apply

Current relation

Relation: weather
Instances: 14
Attributes: 5
Sum of weights: 14

Attributes

All | None | Invert | Pattern

No.		Name
1	☐	outlook
2	☐	temperature
3	☐	humidity
4	☐	windy
5	☐	play

Remove

Selected attribute

Name: outlook
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

Class: play (Nom) | Visualize All

- Appending of files in WEKA

D:\TEMP\Appendweather1.ARFF - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

Appendweather1.ARFF × | Appendweather2.ARFF × | append.ARFF

```
1  @relation weather
2
3  @attribute outlook {sunny, overcast, rainy}
4  @attribute temperature numeric
5  @attribute humidity numeric
6  @attribute windy {TRUE, FALSE}
7  @attribute play {yes, no}
8
9  @data
10 sunny,85,85,FALSE,no
11 sunny,80,90,TRUE,no
12 overcast,83,86,FALSE,yes
13 rainy,70,96,FALSE,yes
14 rainy,68,80,FALSE,yes
15 rainy,65,70,TRUE,no
16 overcast,64,65,TRUE,yes
17 sunny,72,95,FALSE,no
18 sunny,69,70,FALSE,yes
19 rainy,75,80,FALSE,yes
20 sunny,75,70,TRUE,yes
21 overcast,72,90,TRUE,yes
22 overcast,81,75,FALSE,yes
23 rainy,71,91,TRUE,no
24
```

D:\TEMP\Appendweather2.ARFF - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

Appendweather1.ARFF × | Appendweather2.ARFF × | append.ARFF

```
1  @relation weather
2
3  @attribute outlook {sunny, overcast, rainy}
4  @attribute temperature numeric
5  @attribute humidity numeric
6  @attribute windy {TRUE, FALSE}
7  @attribute play {yes, no}
8
9  @data
10 sunny,85,85,FALSE,no
11 sunny,80,90,TRUE,no
12 overcast,83,86,FALSE,yes
13 rainy,70,96,FALSE,yes
14 rainy,68,80,FALSE,yes
15 rainy,65,70,TRUE,no
16 overcast,64,65,TRUE,yes
17 sunny,72,95,FALSE,no
18 sunny,69,70,FALSE,yes
19 rainy,75,80,FALSE,yes
20 sunny,75,70,TRUE,yes
21 overcast,72,90,TRUE,yes
22 overcast,81,75,FALSE,yes
23 rainy,71,91,TRUE,no
24
```

- After Appending of files the contents in the destination file in WEKA

weather.arff | merge.arff | test.arff | training1.arff | training2.arff | Append.arff

```
@relation weather

@attribute outlook {sunny, overcast, rainy}
@attribute temperature real
@attribute humidity real
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no
```

- Merging of files in WEKA

D:\TEMP\Mergeweather1.ARFF - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

Mergeweather1.ARFF | Mergeweather2.ARFF | Merge.ARFF

```
1  @relation weather
2
3  @attribute outlook {sunny, overcast, rainy}
4  @attribute temperature numeric
5  @attribute humidity numeric
6  @attribute windy {TRUE, FALSE}
7  @attribute play {yes, no}
8
9  @data
10 sunny,85,85,FALSE,no
11 sunny,80,90,TRUE,no
12 overcast,83,86,FALSE,yes
13 rainy,70,96,FALSE,yes
14 rainy,68,80,FALSE,yes
15 rainy,65,70,TRUE,no
16 overcast,64,65,TRUE,yes
17 sunny,72,95,FALSE,no
18 sunny,69,70,FALSE,yes
19 rainy,75,80,FALSE,yes
20 sunny,75,70,TRUE,yes
21 overcast,72,90,TRUE,yes
22 overcast,81,75,FALSE,yes
23 rainy,71,91,TRUE,no
24
```

D:\TEMP\Mergeweather2.ARFF - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences H

Mergeweather1.ARFF | Mergeweather2.ARFF | Merge

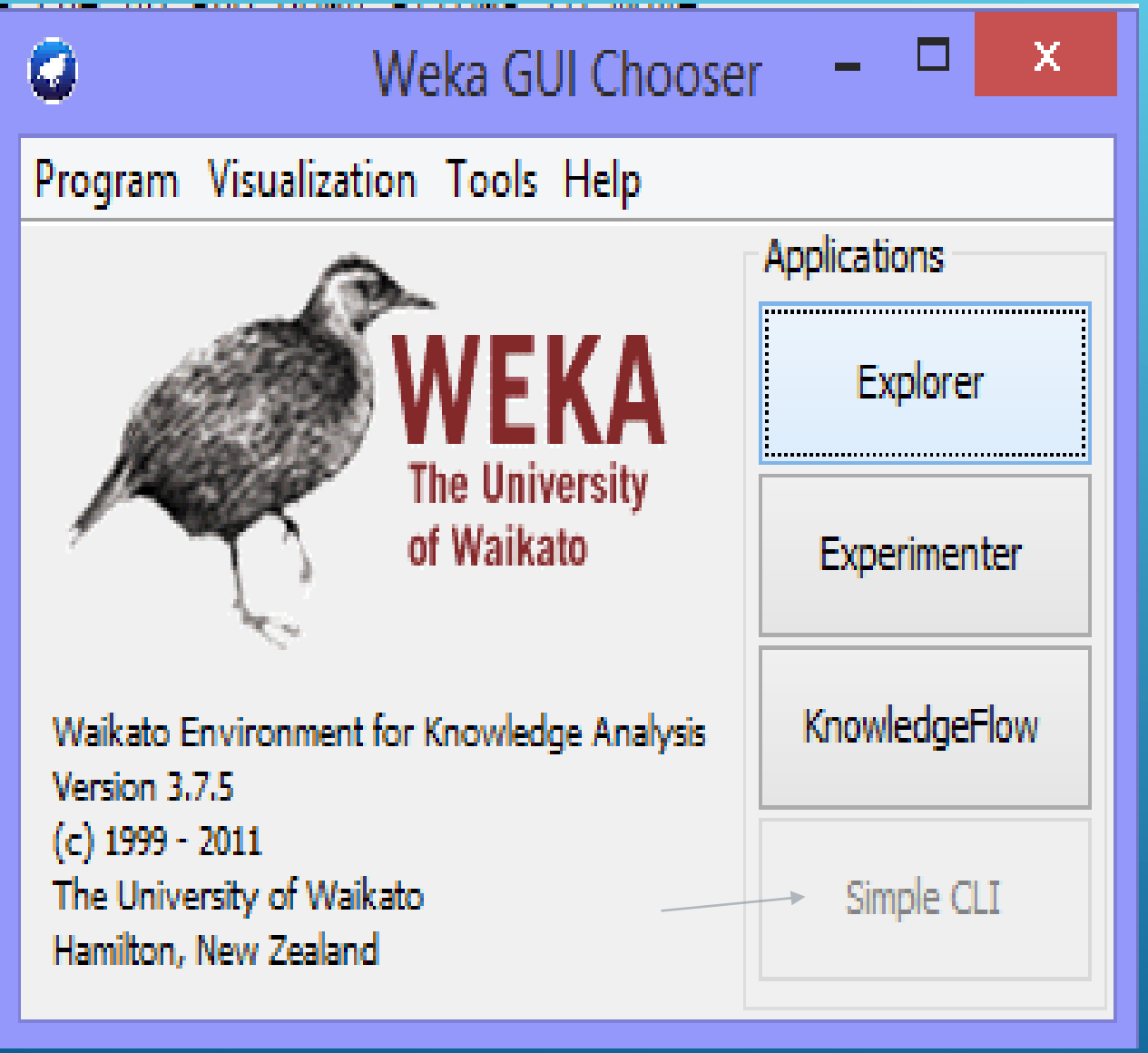
```
1  @relation weather
2
3  @attribute outlook1 {sunny, overcast, rainy}
4  @attribute temperature2 numeric
5  @attribute humidity2 numeric
6  @attribute windy2 {TRUE, FALSE}
7  @attribute play2 {yes, no}
8
9  @data
10 sunny,85,85,FALSE,no
11 sunny,80,90,TRUE,no
12 overcast,83,86,FALSE,yes
13 rainy,70,96,FALSE,yes
14 rainy,68,80,FALSE,yes
15 rainy,65,70,TRUE,no
16 overcast,64,65,TRUE,yes
17 sunny,72,95,FALSE,no
18 sunny,69,70,FALSE,yes
19 rainy,75,80,FALSE,yes
20 sunny,75,70,TRUE,yes
21 overcast,72,90,TRUE,yes
22 overcast,81,75,FALSE,yes
23 rainy,71,91,TRUE,no
24
```

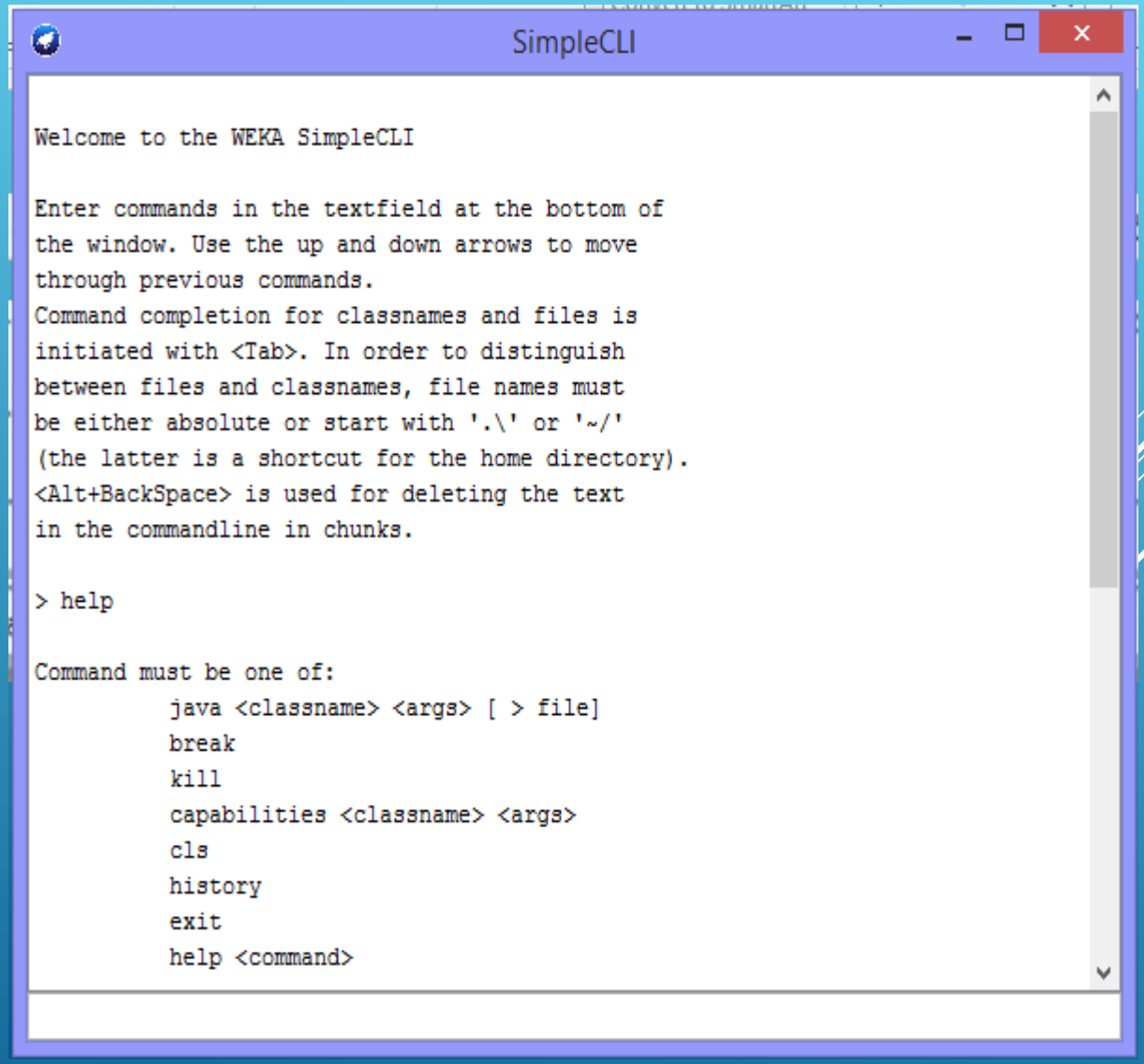
- Merging of files in WEKA

weather.arff | merge.arff | test.arff | training1.arff | training2.arff

```
1   @relation weather
2
3   @attribute outlook {sunny, overcast, rainy}
4   @attribute temperature real
5   @attribute humidity real
6   @attribute windy {TRUE, FALSE}
7   @attribute play {yes, no}
8   @attribute outlook2 {sunny, overcast, rainy}
9   @attribute temperature2 real
10  @attribute humidity2 real
11  @attribute windy2 {TRUE, FALSE}
12  @attribute play2 {yes, no}
13
14  @data
15  sunny,85,85,FALSE,no,sunny,85,85,FALSE,no
16  sunny,80,90,TRUE,no,sunny,80,90,TRUE,no
17  overcast,83,86,FALSE,yes,overcast,83,86,FALSE,yes
18  rainy,70,96,FALSE,yes,rainy,70,96,FALSE,yes
19  rainy,68,80,FALSE,yes,rainy,68,80,FALSE,yes
20  rainy,65,70,TRUE,no,rainy,65,70,TRUE,no
21  overcast,64,65,TRUE,yes,overcast,64,65,TRUE,yes
22  sunny,72,95,FALSE,no,sunny,72,95,FALSE,no
23  sunny,69,70,FALSE,yes,sunny,69,70,FALSE,yes
24  rainy,75,80,FALSE,yes,rainy,75,80,FALSE,yes
25  sunny,75,70,TRUE,yes,sunny,75,70,TRUE,yes
26  overcast,72,90,TRUE,yes,overcast,72,90,TRUE,yes
27  overcast,81,75,FALSE,yes,overcast,81,75,FALSE,yes
28  rainy,71,91,TRUE,no,rainy,71,91,TRUE,no
```

- Commands for merging and appending files WEKA SCLI OPTION





- Commands for merging and appending files WEKA SCLI OPTION in windows OS.

```
> java weka.core.Instances append D:\TEMP\Appendweather1.ARFF D:\TEMP\Appendweather2.ARFF > D:\TEMP\append.ARFF

Finished redirecting output to 'D:\TEMP\append.ARFF'.
```

```
java weka.core.Instances append D:\TEMP\Appendweather1.ARFF D:\TEMP\Appendweather2.ARFF > D:\TEMP\append.ARFF
```

```
> java weka.core.Instances merge D:\TEMP\Mergeweather1.ARFF D:\TEMP\Mergeweather2.ARFF > D:\TEMP\Merge.ARFF

Finished redirecting output to 'D:\TEMP\Merge.ARFF'.
```

```
java weka.core.Instances merge D:\TEMP\Mergeweather1.ARFF D:\TEMP\Mergeweather2.ARFF > D:\TEMP\Merge.ARFF
```

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter

Choose | None | Apply

Current relation

Relation: weather
Instances: 14
Attributes: 5
Sum of weights: 14

Attributes

All | None | Invert | Pattern

No.		Name
1	☐	outlook
2	☐	temperature
3	☐	humidity
4	☐	windy
5	☐	play

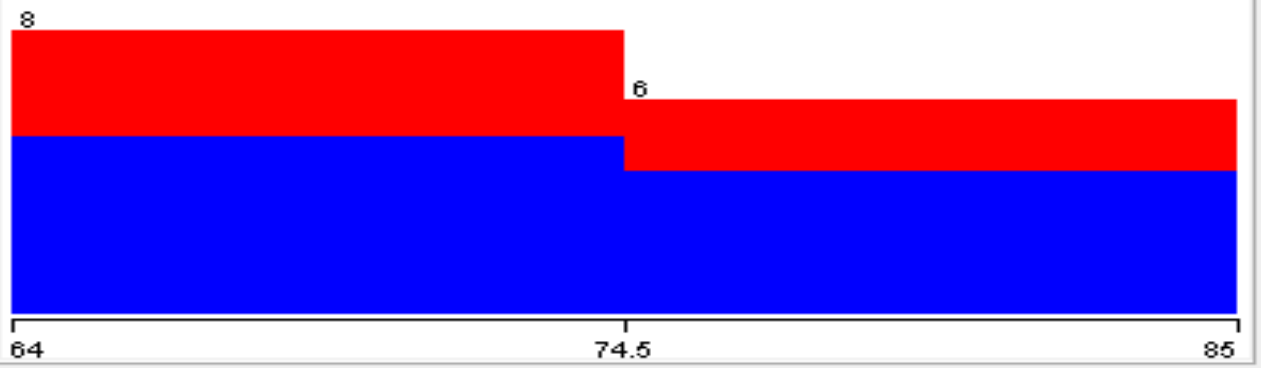
Remove

Selected attribute

Name: temperature
Missing: 0 (0%)
Distinct: 12
Type: Numeric
Unique: 10 (71%)

Statistic	Value
Minimum	64
Maximum	85
Mean	73.571
StdDev	6.572

Class: play (Nom) | Visualize All



Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter

Choose **None** Apply

Current relation

Relation: weather
Instances: 14
Attributes: 5
Sum of weights: 14

Attributes

All | None | Invert | Pattern

No.		Name
1	☐	outlook
2	☐	temperature
3	☐	humidity
4	☐	windy
5	☐	play

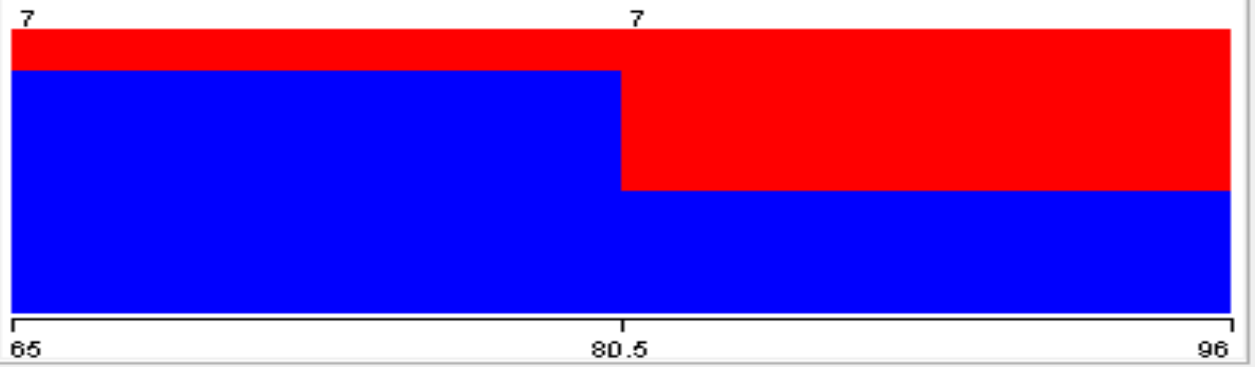
Remove

Selected attribute

Name: humidity
Missing: 0 (0%)
Distinct: 10
Type: Numeric
Unique: 7 (50%)

Statistic	Value
Minimum	65
Maximum	96
Mean	81.643
StdDev	10.285

Class: play (Nom) | Visualize All



Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter

Choose | None | Apply

Current relation

Relation: weather
Instances: 14
Attributes: 5
Sum of weights: 14

Attributes

All | None | Invert | Pattern

No.		Name
1	☐	outlook
2	☐	temperature
3	☐	humidity
4	☐	windy
5	☐	play

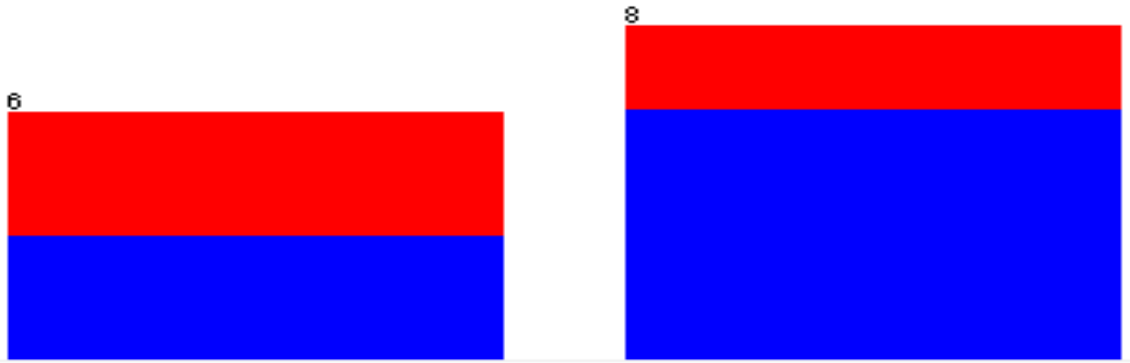
Remove

Selected attribute

Name: windy
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	TRUE	6	6.0
2	FALSE	8	8.0

Class: play (Nom) | Visualize All



Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter

Choose | None | Apply

Current relation

Relation: weather
Instances: 14
Attributes: 5
Sum of weights: 14

Attributes

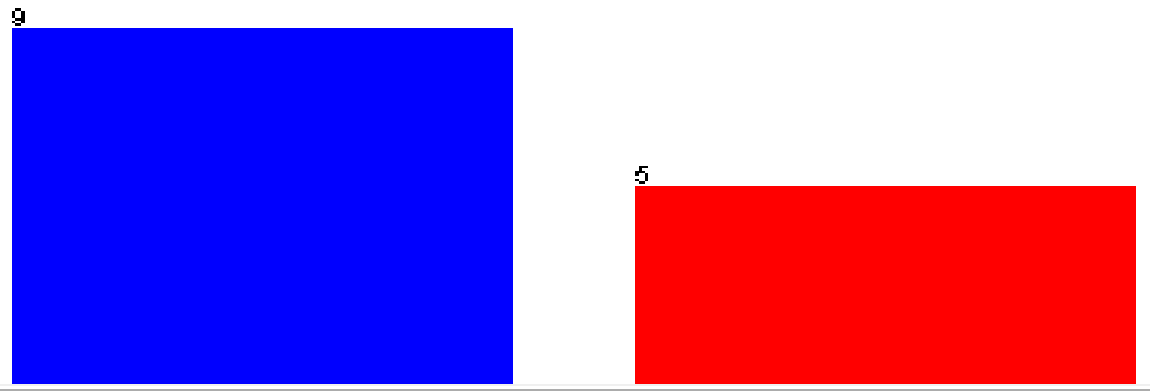
All | None | Invert | Pattern

No.		Name
1	☐	outlook
2	☐	temperature
3	☐	humidity
4	☐	windy
5	☐	play

Remove

Selected attribute

Name: play
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	yes	9	9.0
2	no	5	5.0

Class: play (Nom) | Visualize All