An Introduction to Weka with Demos

Tomasz Oliwa

Ph.D., Computer Science, The University of Georgia, USA

Dipl.-Inform., University of Koblenz-Landau, Germany

Lecture in CSCI/ARTI 8950 Machine Learning
Slides at: http://www.cs.uga.edu/~tomasz/weka2013fall/

September 5, 2013

The University of Georgia

Weka Software

The Weka machine learning/data mining suite:

- Rich collection of preprocessing, (un)/supervised learning and evaluation algorithms
- Accessible GUI
- Java, Free software (GPL): `http://www.cs.waikato.ac.nz/~ml/weka/`

Attribute-Relation File Format (ARFF)

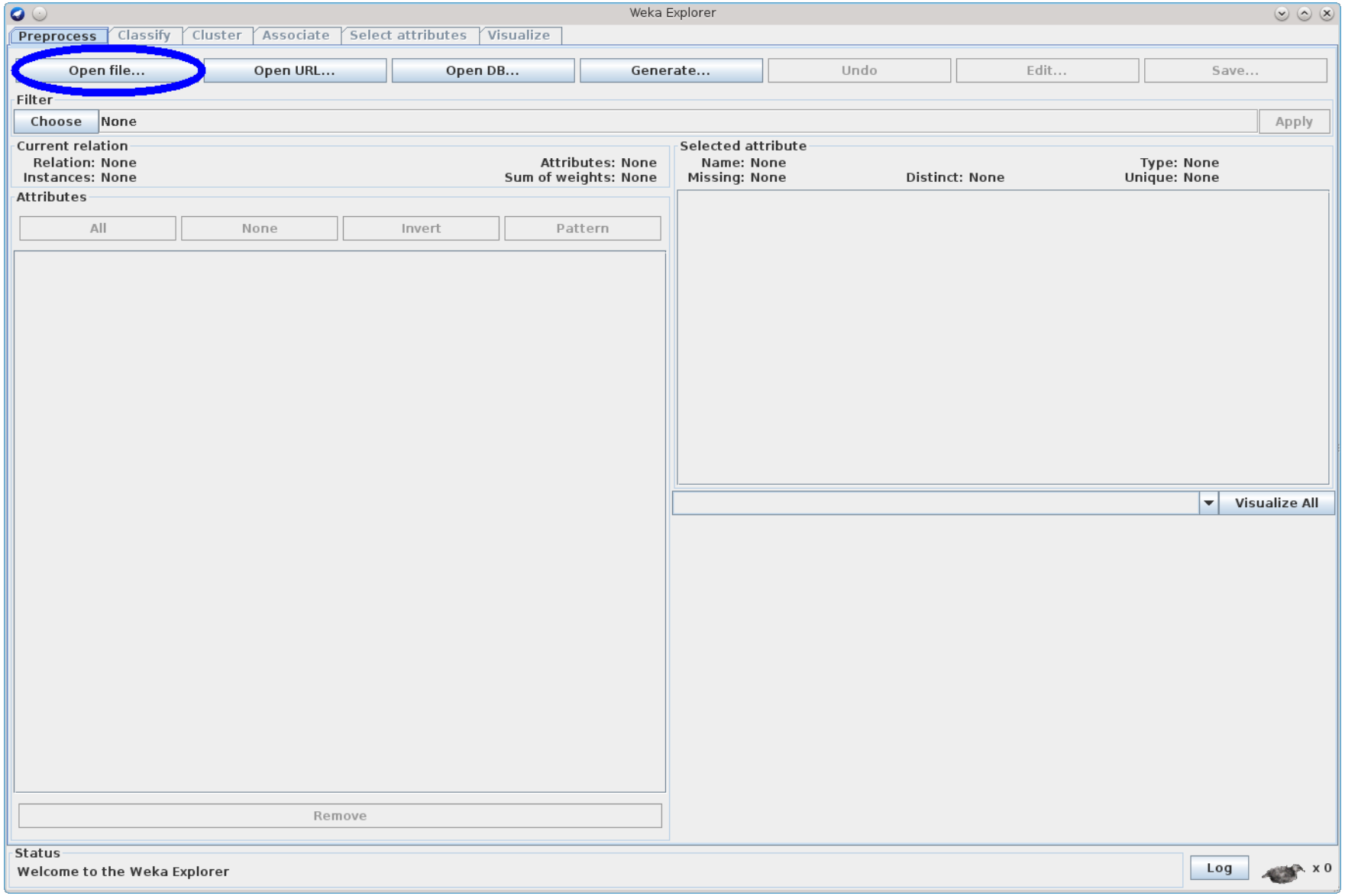
```
@relation weather

@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
```

Weka Modes

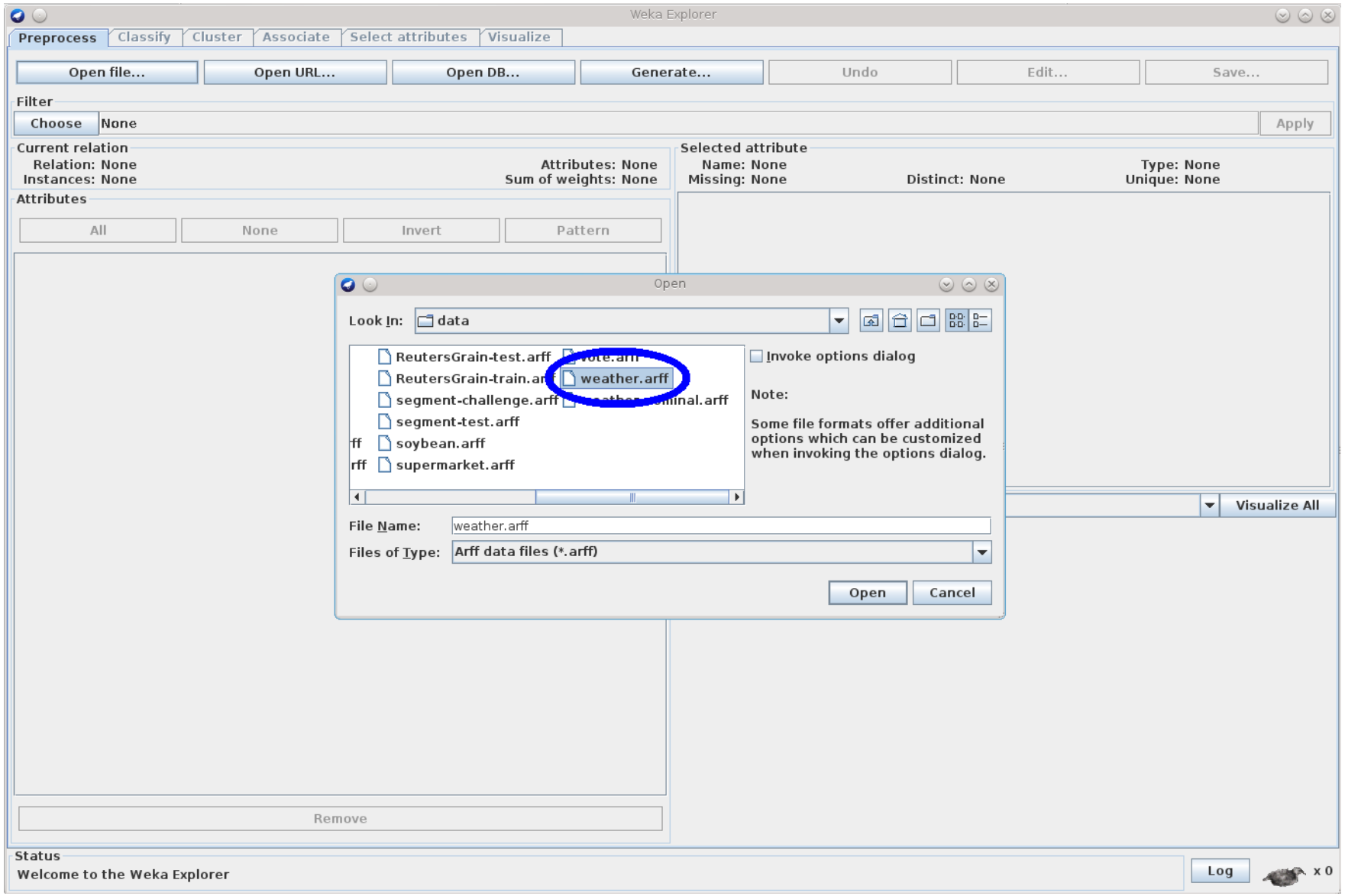
1. **Explorer:** Direct application and visualizations
2. **Experimenter:** Run & compare algorithms
3. **Knowledge Flow:** Visual composition of workflow
4. **Simple CLI:** Command line interface
5. **Source code:** Weka classes in Java code

Weka GUI Chooser

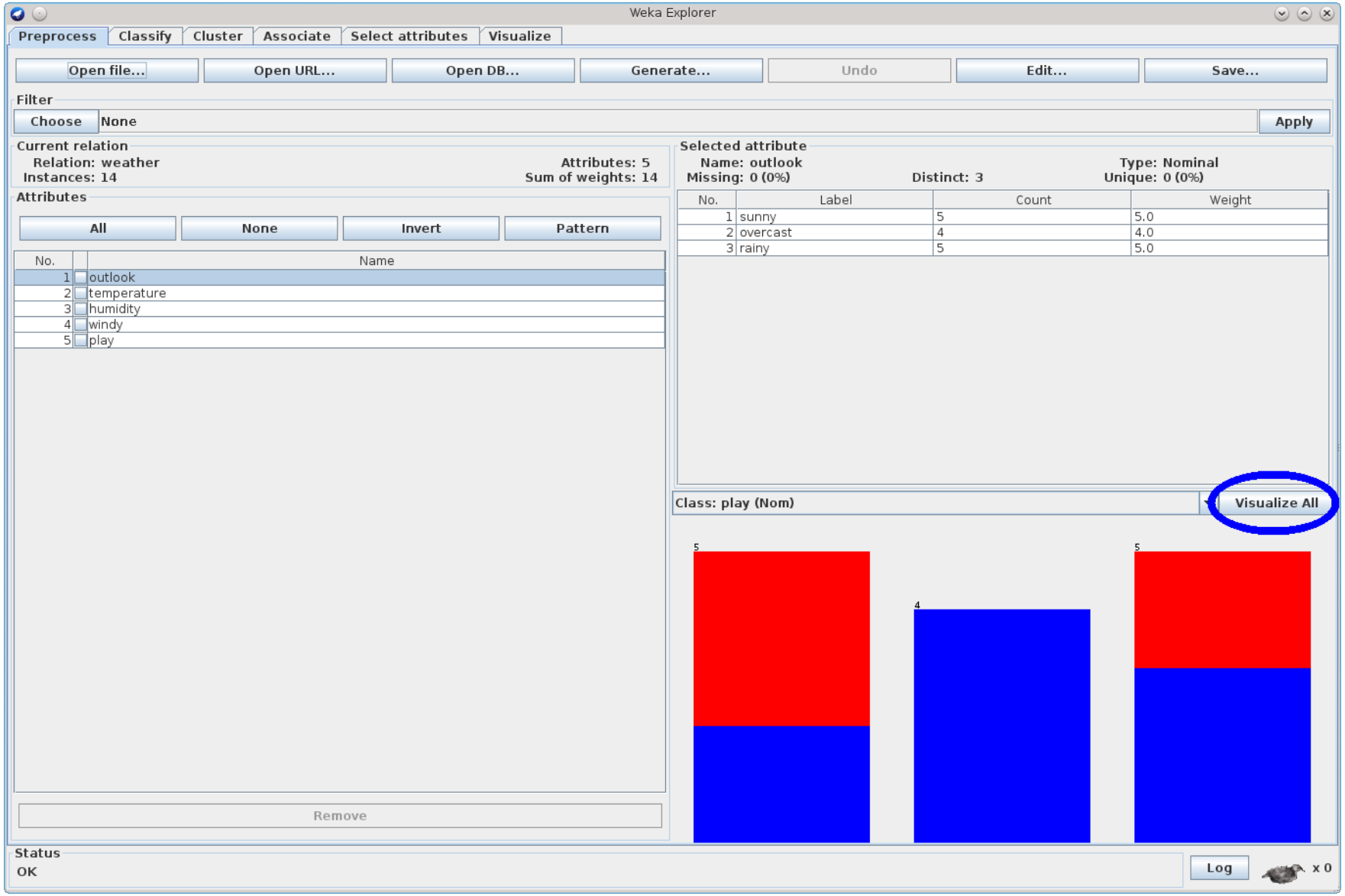
Preprocessing



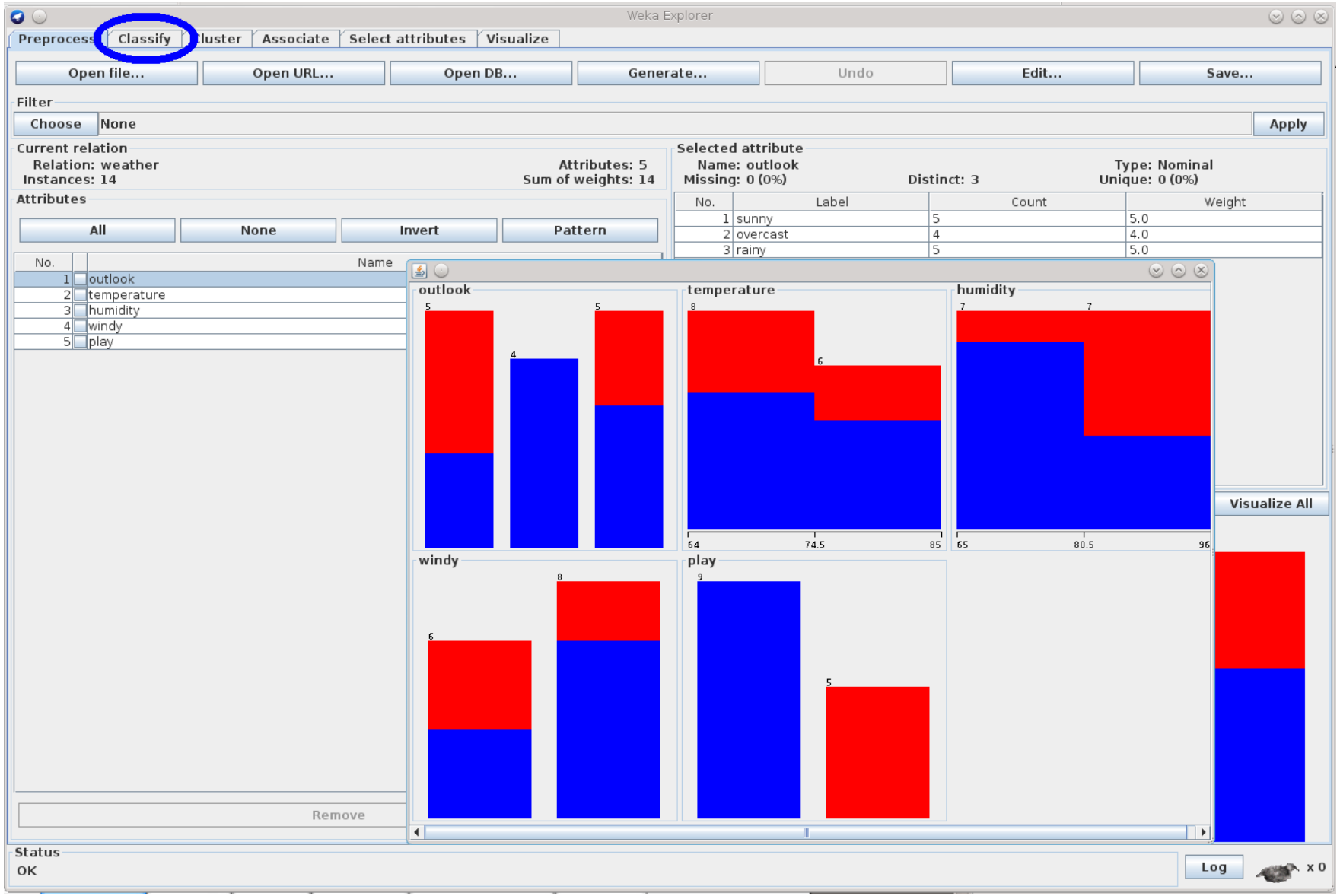
Preprocessing - Load Data



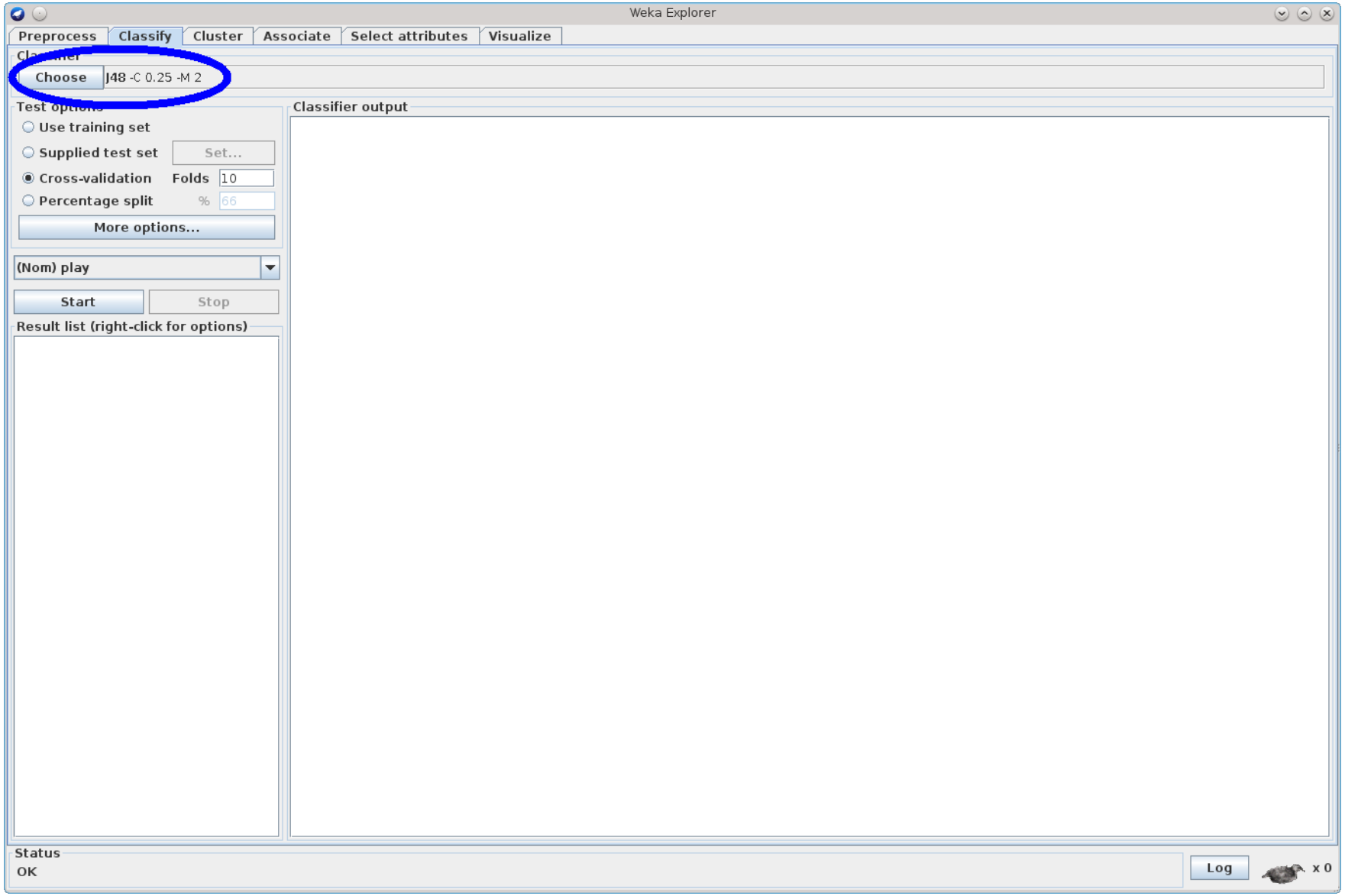
Preprocessing - Data Overview



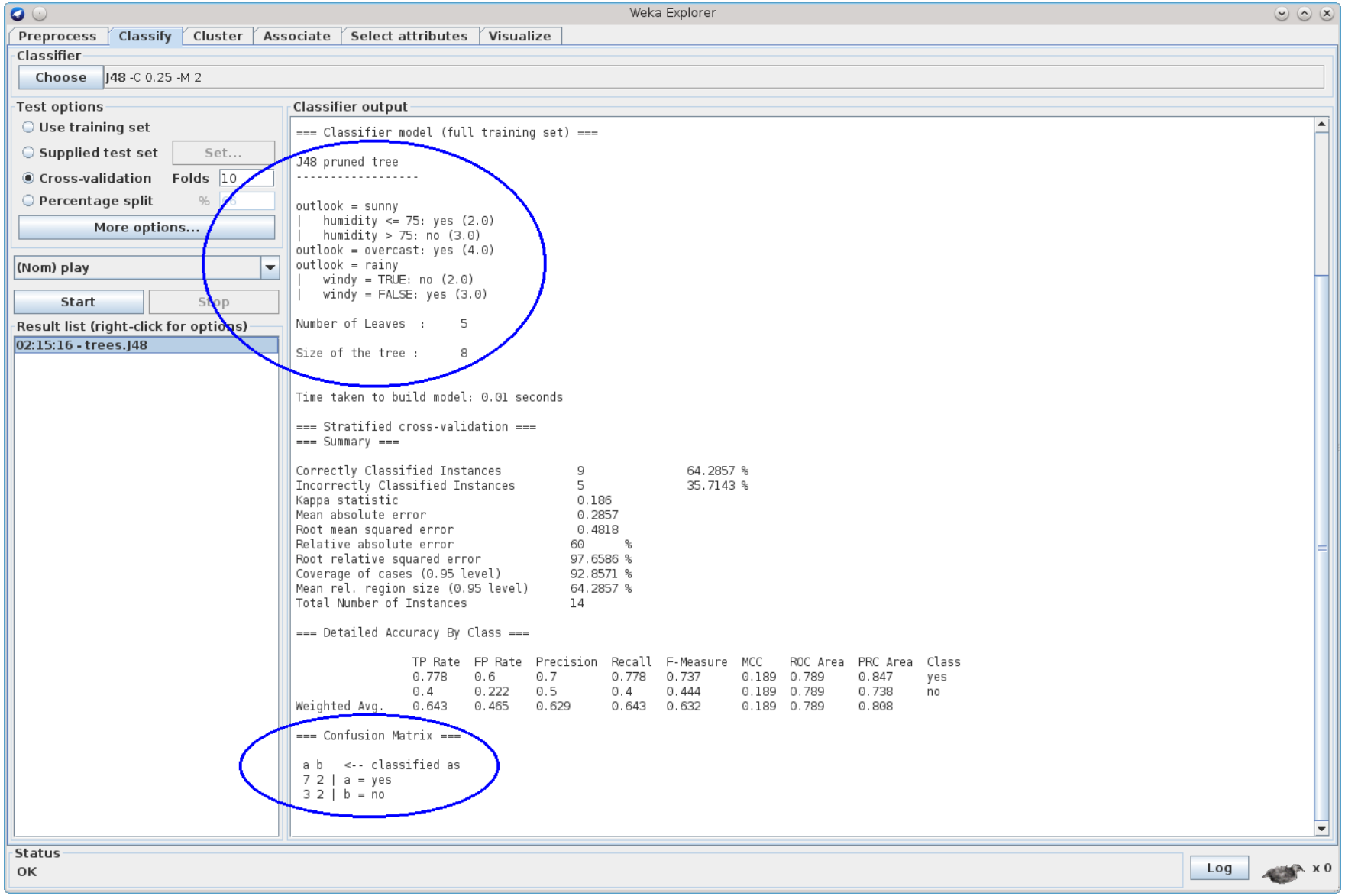
Preprocessing - Visualize All



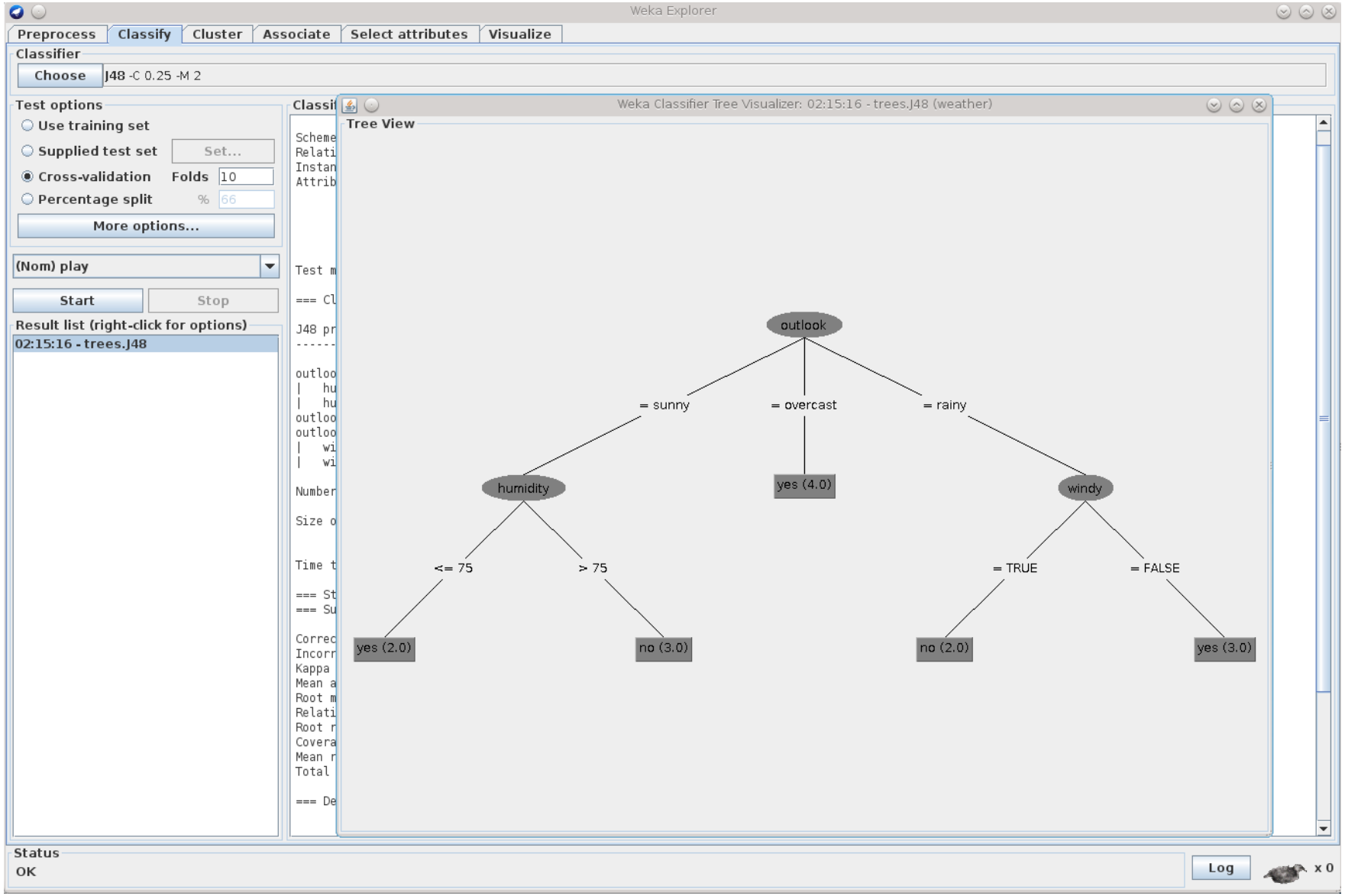
Classification



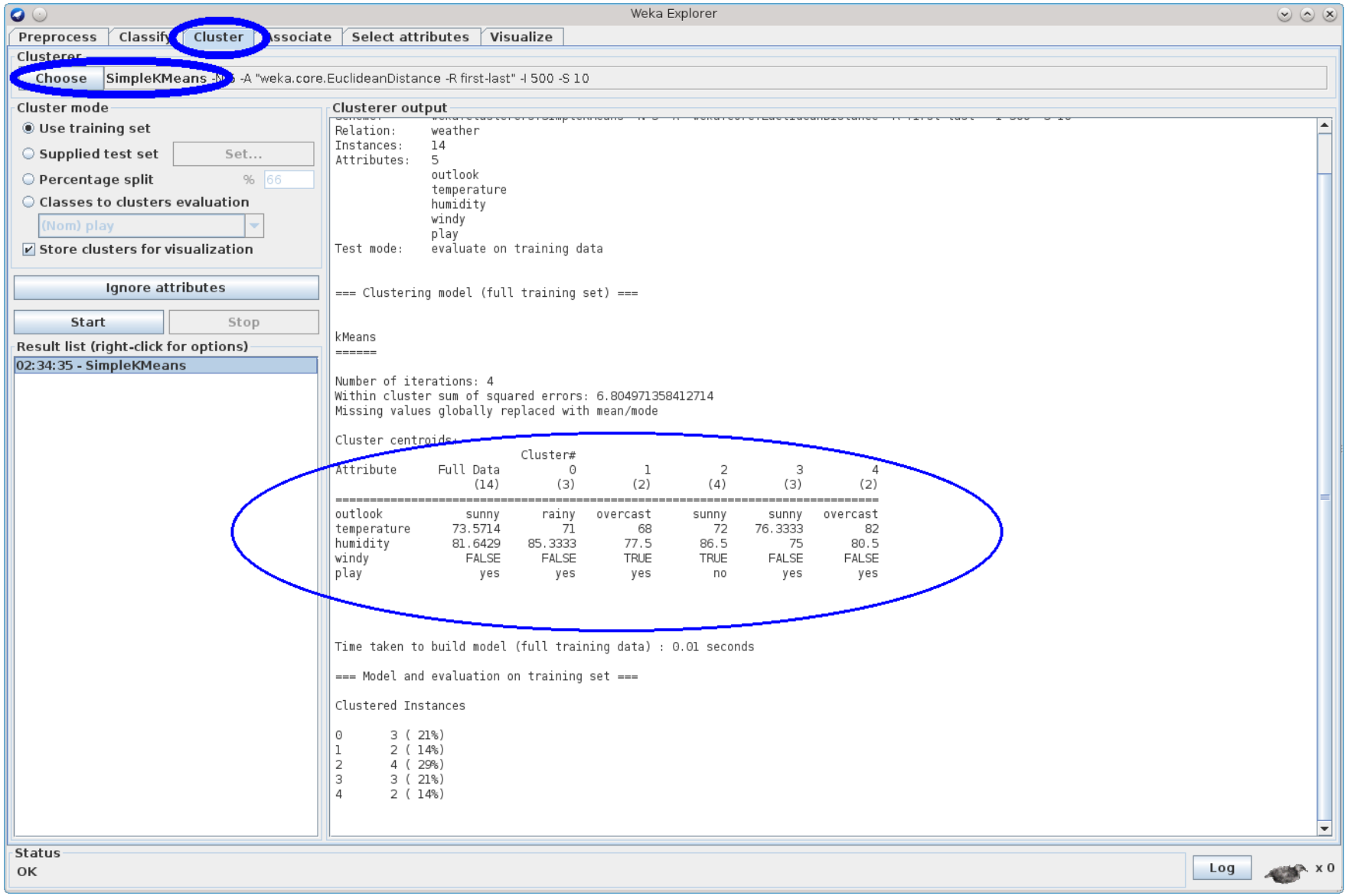
Classification Results



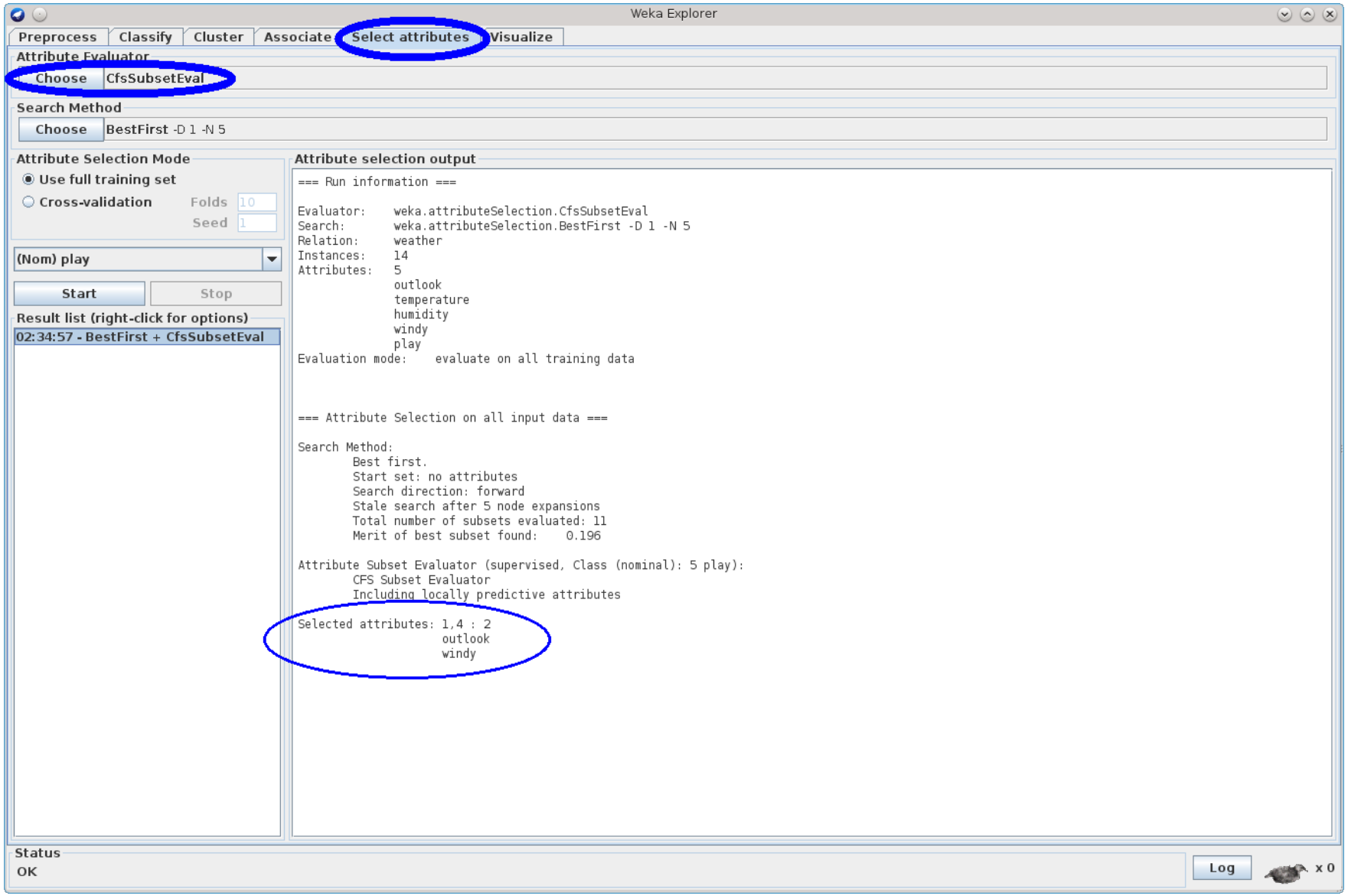
Classification Results - DT



Clustering Results



Attribute Selection Results



Demos

Demo time:

- Preprocessing
- Classification
- Regression
- Clustering
- Feature Selection
- Experimenter
- Source code import

Take Home Message

Weka is a powerful machine learning suite:

- **Explorer**: Intuitive suite with result visualizations
- **Experimenter**: Batch perform statistical tests on results
- **Source code**: Directly import Weka classes in your programs

Questions

- Thanks for your attention!
- Do you have any questions?

References and Tutorials

- Weka: `http://www.cs.waikato.ac.nz/~ml/weka/`
- Weka Wiki: `http://weka.wikispaces.com/`

IBM Weka Tutorials:

- Introduction and Regression:
  `http://www.ibm.com/developerworks/opensource/library/os-weka1/index.html`
- Classification and clustering:
  `http://www.ibm.com/developerworks/opensource/library/os-weka2/index.html`
- Nearest Neighbor and server-side library:
  `http://www.ibm.com/developerworks/opensource/library/os-weka3/index.html`