

# Weka: Introduction

**Cristóbal Romero,**

Córdoba University, Campus Universitario de Rabanales, 14071, Córdoba, Spain  
cromero@uco.es,

# WEKA: The bird

**Weka is a bird found only on the islands of New Zealand.**



*Copyright: Martin Kramer (mkramer@wxs.nl)*

# WEKA: Introduction

- A collection of open source DM/ML algorithms (pre-processing, classifiers, clustering, association rule).
- Created by researchers at the University of Waikato in New Zealand
- Weka is open source software in JAVA issued under the GNU License.
- Download software from <http://www.cs.waikato.ac.nz/ml/weka/>

# Main Features

- 49 data preprocessing tools
- 76 classification/regression algorithms
- 8 clustering algorithms
- 3 algorithms for finding association rules
- 15 attribute/subset evaluators + 10 search algorithms for feature selection

# WEKA:: ARRF file format

Attribute-Relation File Format (ARFF): Header section and Data section

```
@relation heart-disease-simplified
@attribute age numeric
@attribute sex { female, male}
@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}
@attribute cholesterol numeric
@attribute exercise_induced_angina { no, yes}
@attribute class { present, not_present}

@data
63,male,typ_angina,233,no,not_present
67,male,asympt,286,yes,present
67,male,asympt,229,yes,present
38,female,non_anginal,?,no,not_present
...
```

**relation name**

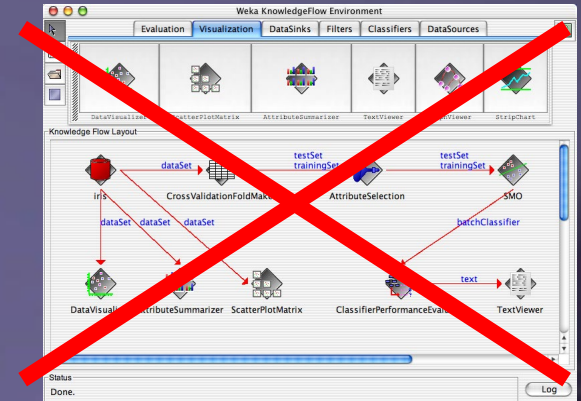
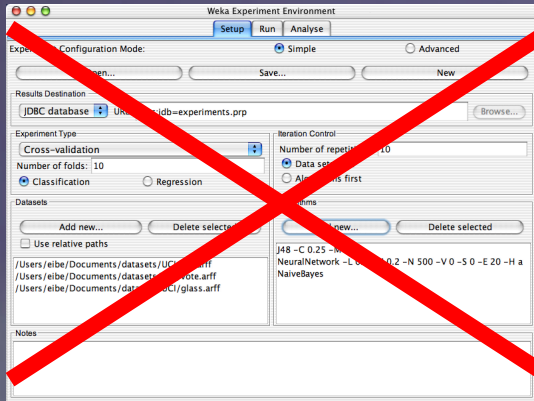
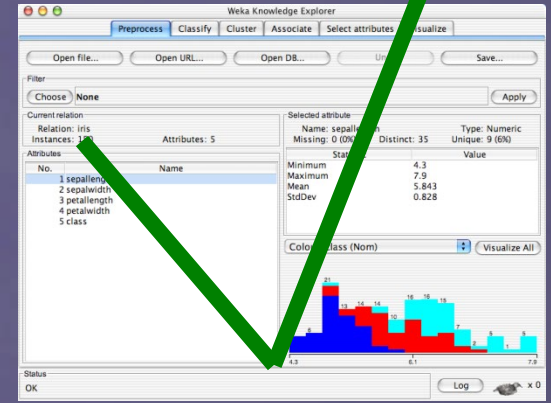
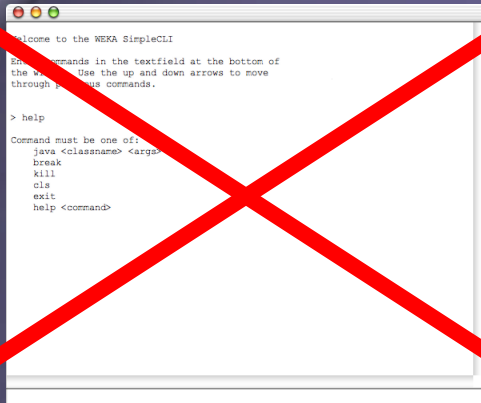
**numeric attribute**

**nominal attribute**

**data**

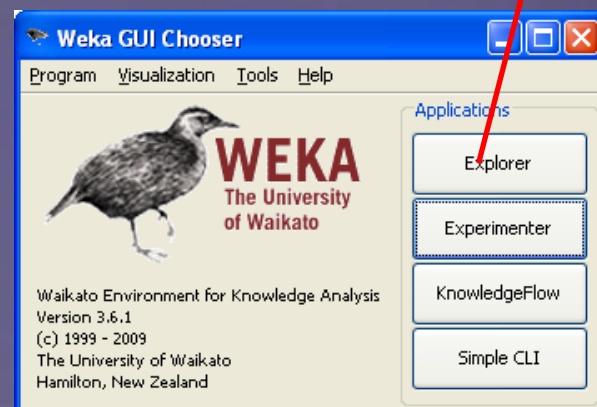
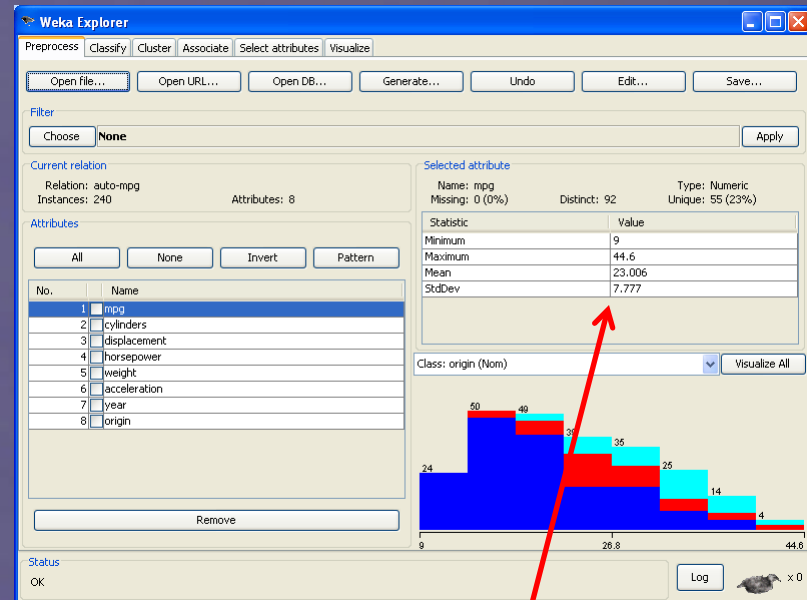
A more thorough description is available here  
<http://www.cs.waikato.ac.nz/~ml/weka/arff.html>

# WEKA: Interface



# Explorer Interface

- Explorer
  - pre-processing the data
  - building “classifiers”
  - clustering data
  - finding associations
  - attribute selection
  - data visualization



# Limitations of WEKA

- Traditional algorithms need to have all data in main memory
  - ==> big datasets are an issue
  - Solution:
    - Incremental schemes
    - Stream algorithms
- MOA “**M**assive **O**nline **A**nalysis”  
(not only a *flightless* bird, but also *extinct!*)



# References and Resources

- WEKA website:

<http://www.cs.waikato.ac.nz/~ml/weka/index.html>

- WEKA Data Mining Book:

Ian H. Witten and Eibe Frank,  
Data Mining: Practical Machine Learning  
Tools and Techniques (Third Edition)

- WEKA Wiki:

<http://weka.wikispaces.com/>

