ALGONQUIN COLLEGE

CST8390 - LAB
BUSINESS INTELLIGENCE & DATA ANALYTICS

Week 1

LAB 1 - Data exploration and Weka - Iris dataset

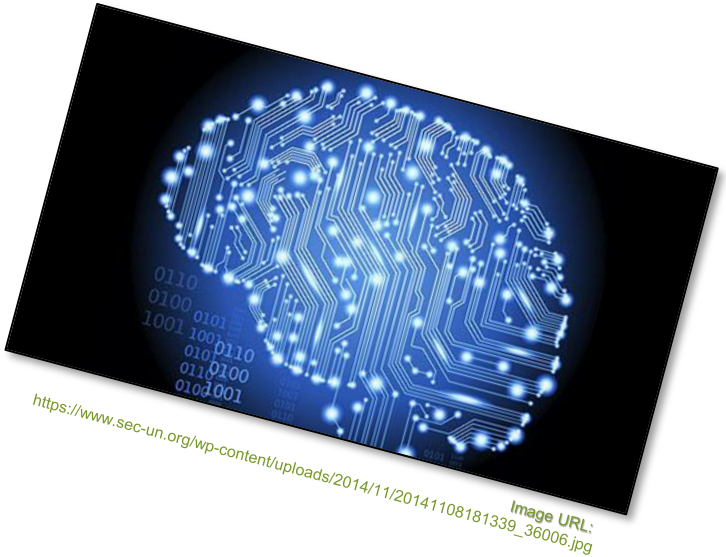
Lab 1 – Weka and Data

PART I

- General view
- Steps

PART II

- Expected results
- Doubts…



CST8390 - Lab
Business intelligence & data analytics

Lab 1 – Weka and Data

Part I – About Weka

General View

- The labs will use WEKA:
  - Main site: https://www.cs.waikato.ac.nz/ml/weka/
- What is WEKA?
  - Acronym for *Waikato Environment for Knowledge Analysis*
- Why to use it?

Weka is an **open source** machine learning software that can be accessed through a **graphical user interface**, standard terminal applications, or a Java API. It is widely used for teaching, research, and industrial applications and contains a plethora of **built-in tools** for standard machine learning tasks.

More about the Tool

- Weka = Machine Learning without Programming
  - **Q1:** What are the advantages x disadvantages about this?
- Models can be used, built, and evaluated in Java, R or Python
  - **Q2:** What is the purpose for this?

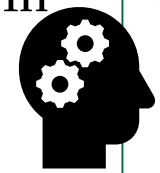


Image URL:https://i0.wp.com/opensourceforu.com/wp-content/uploads/2014/03/Weka-bird.gif?fit=500%2C298&ssl=1

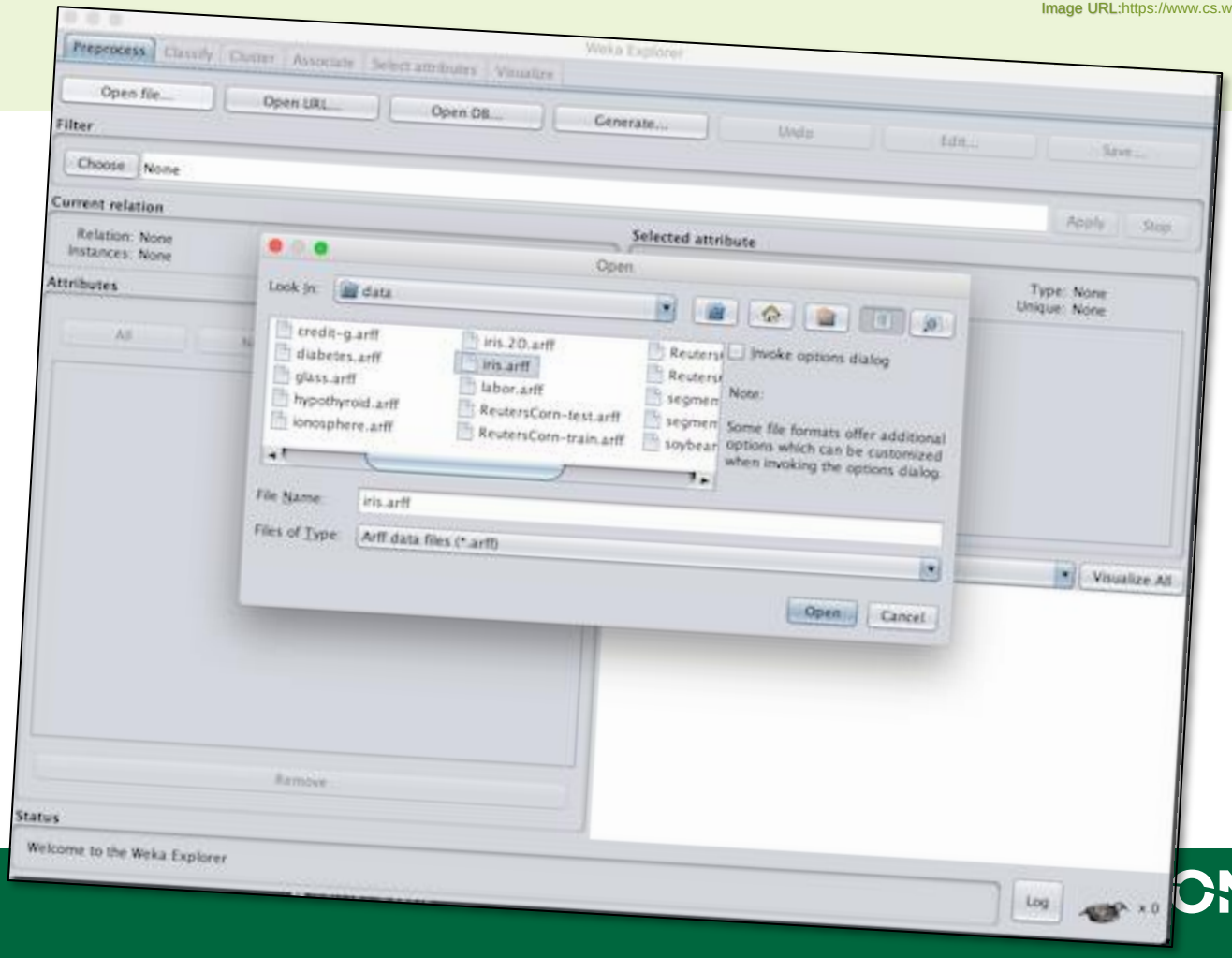
Remember: Weka is just one option that you can use in your professional life.

General 3 steps (https://www.cs.waikato.ac.nz/ml/weka/)

- **Open a dataset**:
  - First, we open the dataset that we would like to evaluate.
- **Choose a classifier**:
  - Second, we select a learning algorithm to use, e.g., the J48 classifier, which learns decision trees.
- **Evaluate predictive accuracy**:
  - Finally, we run a 10-fold cross-validation evaluation and obtain an estimate of predictive performance.

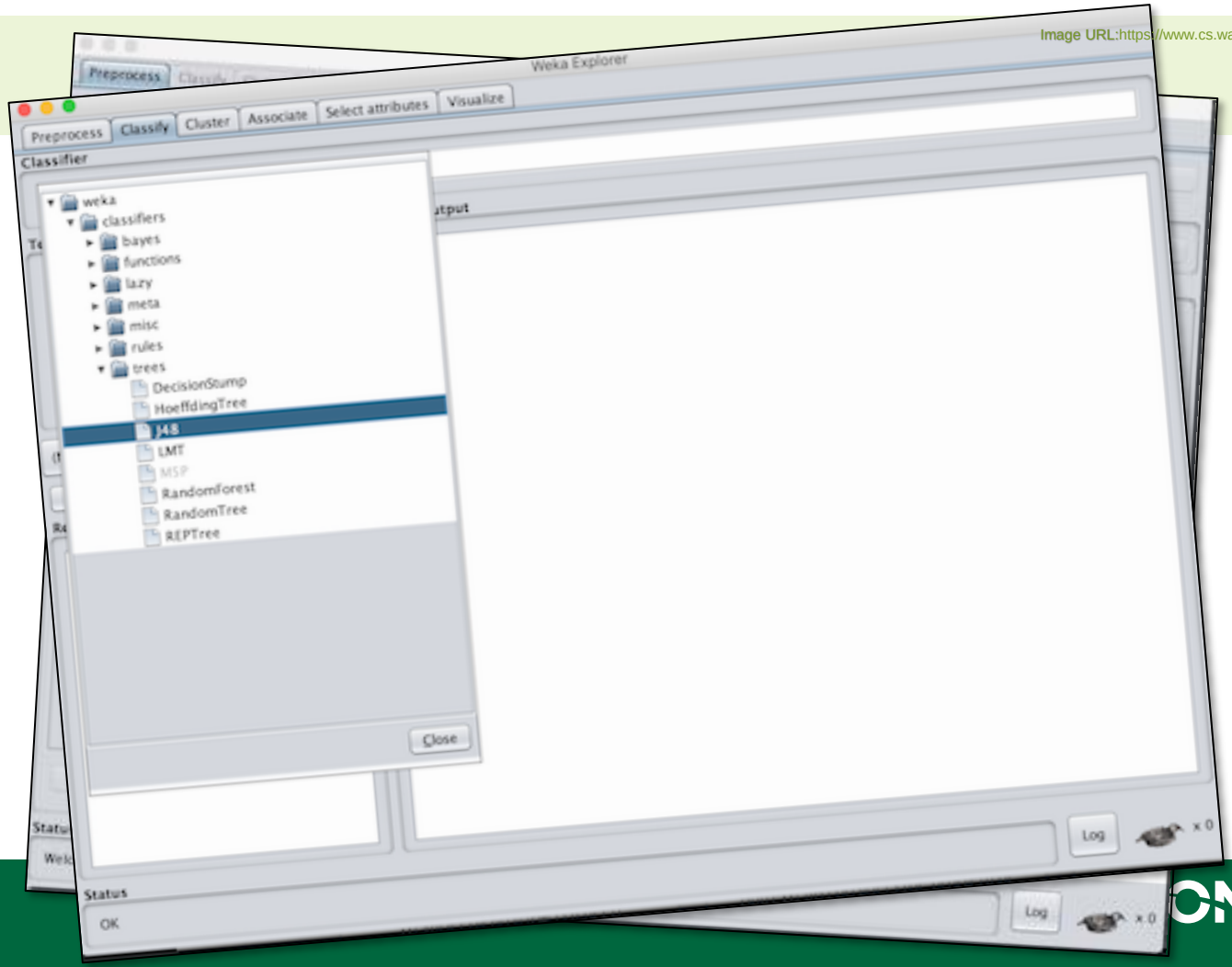
S1

Image URL:https://www.cs.waikato.ac.nz/ml/weka/



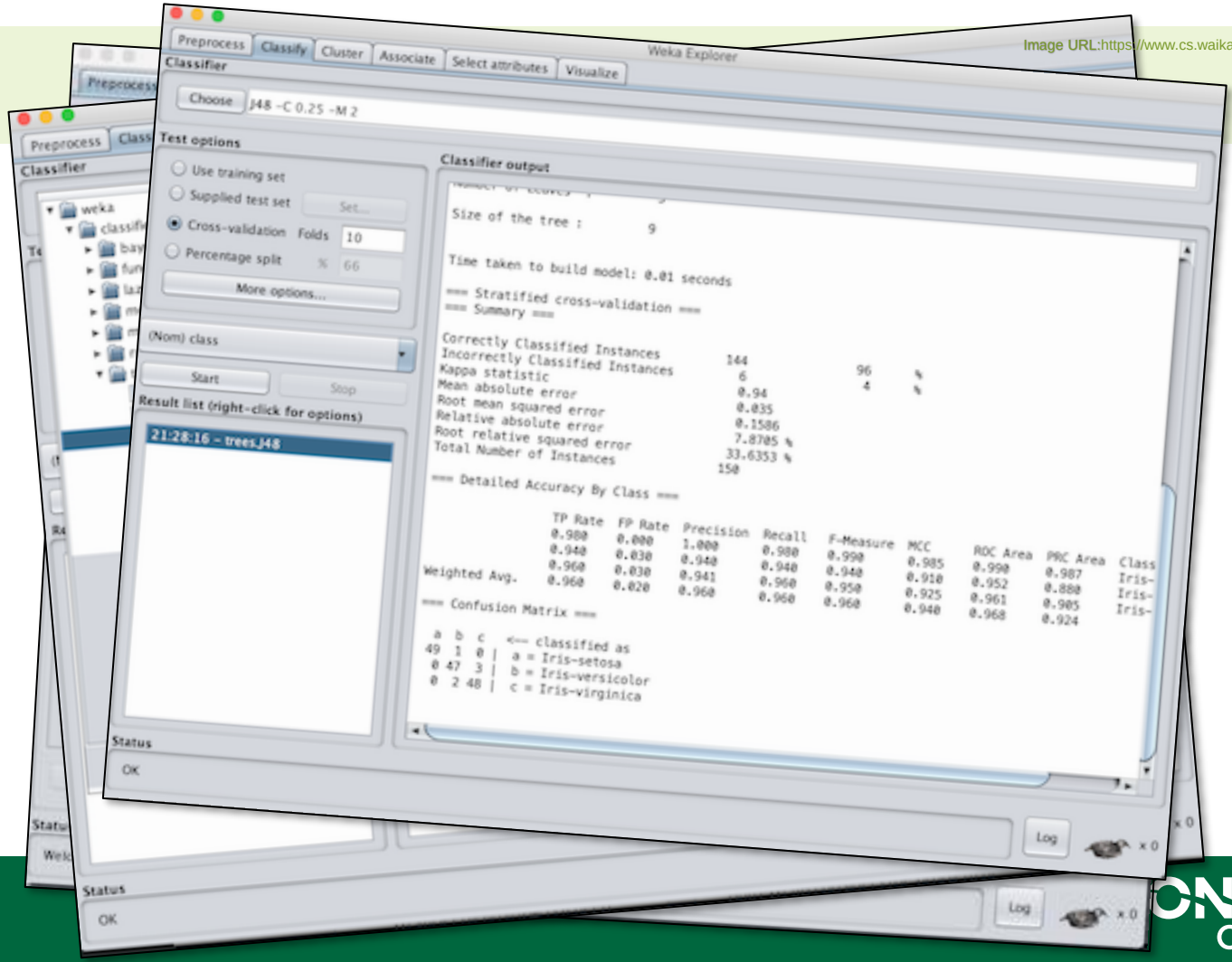
S2

Image URL:https://www.cs.waikato.ac.nz/ml/weka/



S3

Image URL:https://www.cs.waikato.ac.nz/ml/weka/



Note

- **S1:** Sometimes you need to create your own dataset!

  Do not worry about the format: it is a CSV file. The problem is how to prepare the data.

- **S2:** Choose the algorithm is not so easy!

  It depends on the **problem** that you are trying to solve. You need to understand it perfectly.

- S3: Evaluate is the main role of a business analyst.

  The key is understanding how the numbers can be interpreted in order to **answer** the original problem.

CST8390 - Lab
Business intelligence & data analytics

Lab 1 – Weka and Data

Part II – Practice

General View

- About all labs

1. You will do the practice **step-by-step**, following the instructions;
2. During this, you have some **questions** to be answer;
3. Do not forget to do your own **analysis** about you have obtained;
4. In the end, you need to **show** your own results and analysis to professor.

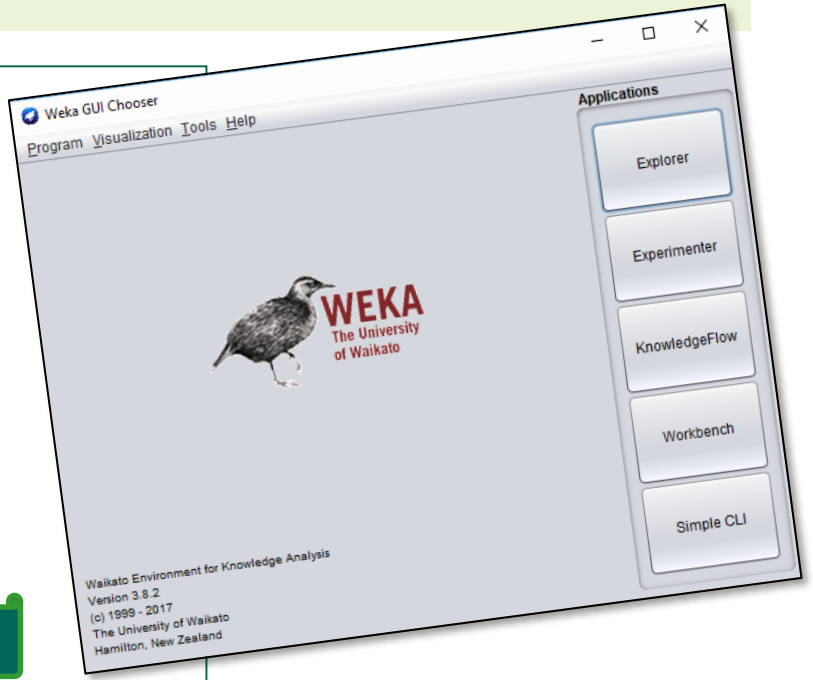
Image URL:
https://images.shiksha.com/mediadata/images/1533291262php6tLOhO.jpeg

Tip: Be sure that the professor has marked your grade.

First Activities

- Download, install and test WEKA;
- Open the Iris dataset (iris.arff);
- Try to familiarize with the format:
  - Comments – attributes – data;
- Explore the Weka panels and basic information from the dataset.

Tip: Be sure that your Weka version is 3.8 or later.

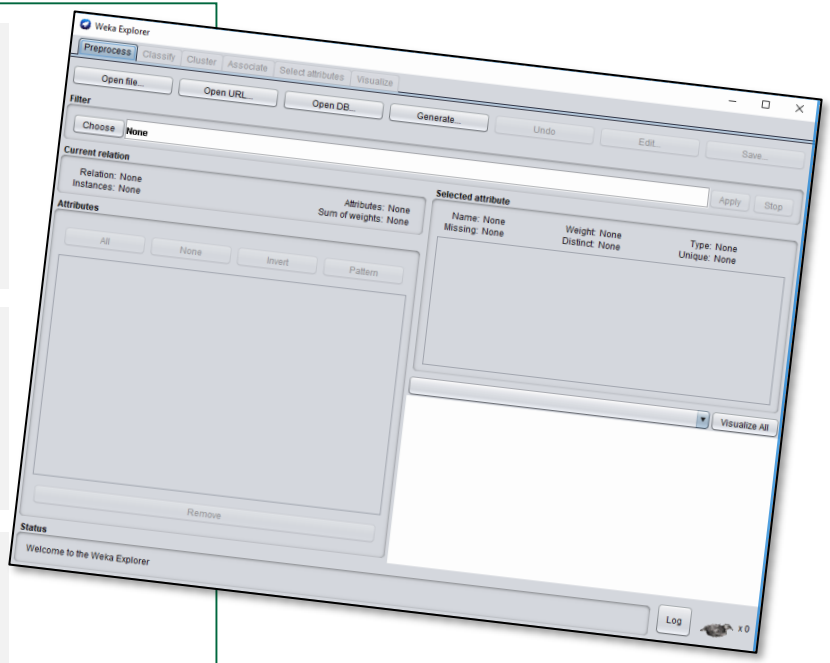


Weka Introduction

Demo

Step-by-step

1. Download and install Weka. You can find it here: http://www.cs.waikato.ac.nz/ml/weka/downloading.html
2. Open Weka and have a look at the interface. It is an open-source project written in Java from the University of Waikato.
3. Click on the Explorer button on the right side:



Step-by-step

4. Check different tabs to familiarize with the tool.

5. Weka comes with a number of small datasets. Those files are located at C:\Program Files\Weka-3-8 (If it is installed at this location. Or else, search for Weka-3-8 to find the installation location). In this folder, there is a subfolder named ‘data’. Open that folder to see all files that comes with Weka.

Step-by-step

6. For easy access, copy the folder ‘data’ and paste it in your ‘Documents’ folder.

Tip: Copy the original data to another folder and use it in the labs.

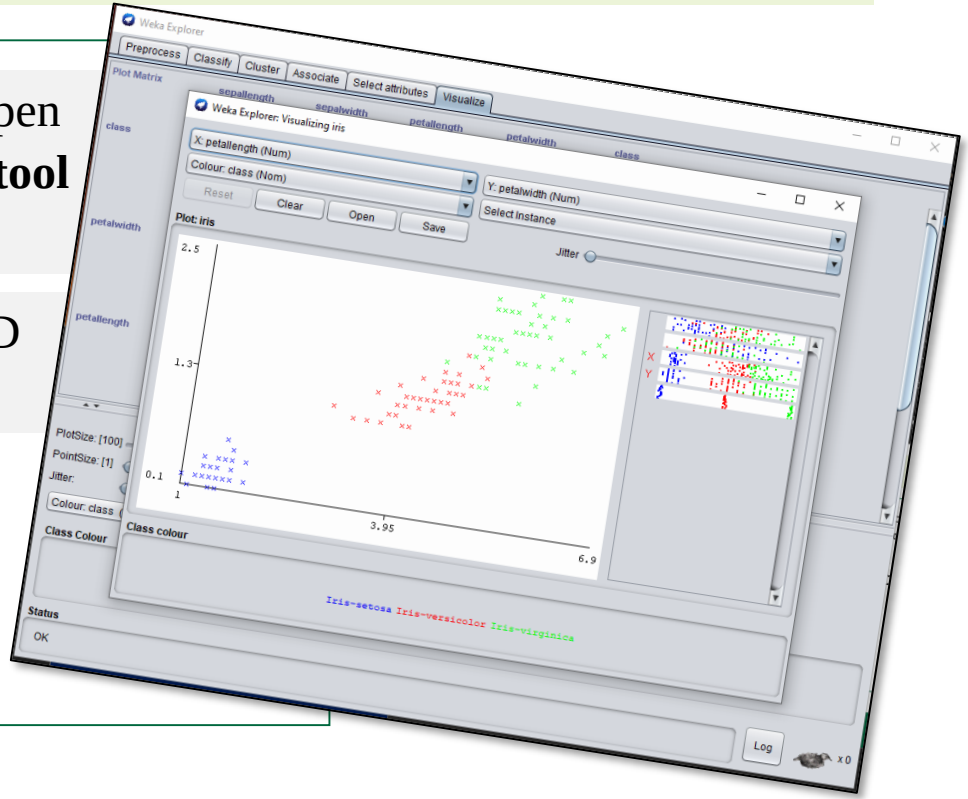
7. In this lab, we will work with the dataset Iris. To open Iris dataset, click on ‘Open file’ in the ‘Preprocess tab’. From your ‘data’ folder, select **iris.arff** and hit open.

Image URL:
https://www.gardendesign.com/pictures/images/900x705Max/site_3/iris-louisiana-black-gamecock-iris-beardless-louisiana-iris-shutterstock-com_12592.jpg

Step-by-step

8. To know more about the iris dataset, open **iris.arff** in Notepad++ or in a **similar tool** and read the comments.

9. Click on visualize tab to see various 2D visualizations of the dataset.

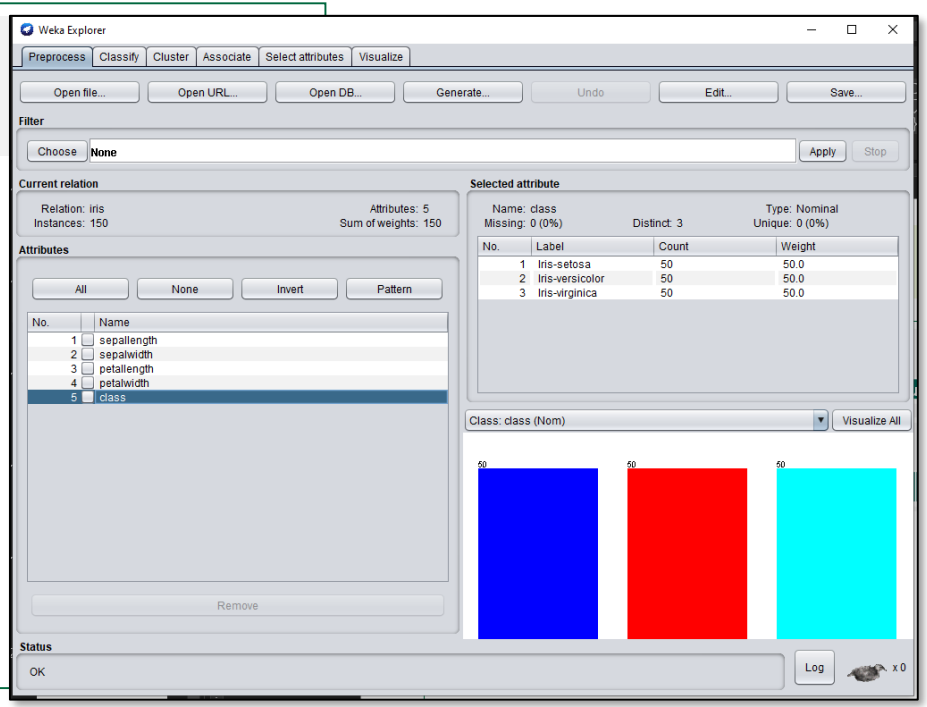


*The **Jitter** allows you to see some point that are hidden or in other words it is a random displacement applied to x and y values to separate points that lie on top of one another.*

Step-by-step

10. Fill this table:

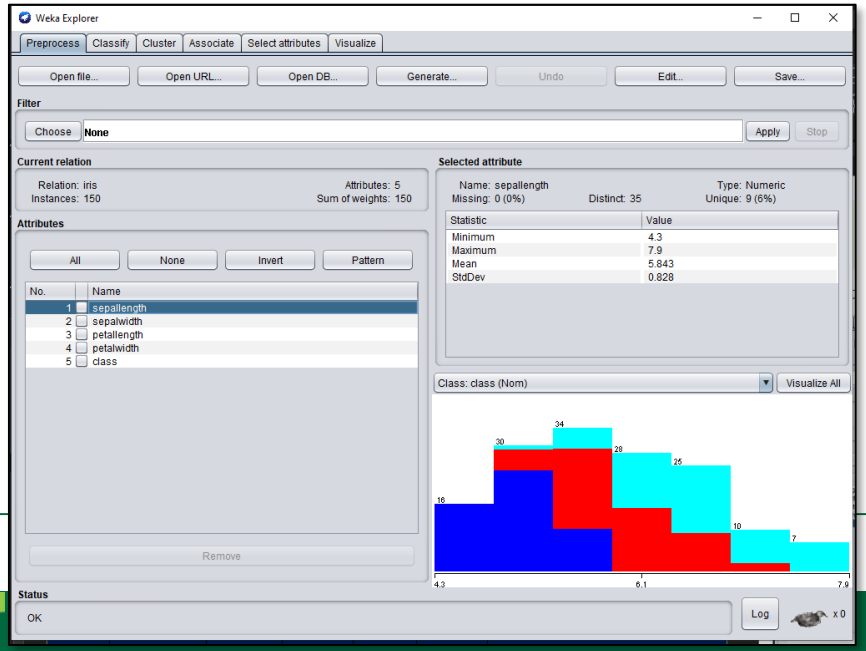
Flower Type	Count



Step-by-step

11. Fill this table:

Attribute	Minimum	Maximum	Mean	StdDev



Get the marks…

In order to get the credit for this lab:

- Show the Iris file in Weka
- Fill in the tables for questions 10 & 11

Image URL:
https://storage.googleapis.com/tune-me-bucket/images/A-kiss-for-good-marks.width-280.jpegquality-70.jpg

What about my analysis?

- At any lab, perform your own analysis;
- For instance, in this first lab, you can think:
  - About statistics (ex: what is the maximum standard deviation);
  - How to perform your own values (ex: using a spreadsheet such as Excel);
  - Understand and change the ARFF file;
  - **Use your creativity...**

Tip: The better idea comes from your own analysis. It is your differential.

Image URL: https://www.marketingdirecto.com/wp-content/uploads/2018/01/ciencia-datos.jpg

Open questions…

- Before we start, do you have any doubt / question?

az.allevents.in/events3/banners/26b150363d5757da8578fa6a1481585a368f12d4247adfe95837e6ea6c5ab2af-rimg-w1200-h549-gmir.jpg?v=1569691155

Image URL: https://cdn-

See you…

- Remember:
  - Labs require practice and it is ok committing errors and learning with them.
  - Do not forget to show your results…
  - Any questions, let me know…

sousap@algonquincollege.com

Image URL: https://thumbs.gfycat.com/MaleFrigidBull-size_restricted.gif

Thank you for your attention!