

**Lab 1: Data Exploration**

1. **Loading + Exploration – *algae***
  - a. Load the data.
  - b. How many instances does this dataset have?
  - c. How many attributes? (Nominal, Binary and Numerical)
  - d. Explore the attribute *season*.
    - i. What is its type?
    - ii. What is its domain?
    - iii. Explore its distribution.
  - e. Explore the *pH* attribute.
    - i. What is its type?
    - ii. What is its domain and range?
    - iii. How many missing values?
    - iv. How is its distribution? – Histograms, boxplots
    - v. What about outliers?
  - f. Explore pH and Ammonium together.
    - i. Are they in the same scale? Discuss implications.
  - g. Look for dependent variables by inspecting scatter plots.

**EVALUATION:**

**Show histogram for attribute chlorophyll. Are you confident about its distribution?**