**Lesson Plan: Introduction to Support Vector Machines**

**Rational:** Classification of information is an important tool to understanding biological diversity. Here we introduce support vector machines as a way of classifying data. Often there are large amounts of data that are represented by many different measurements, the use of SVM helps the researcher classify unknown observations.

**Student Learning Objectives**

1. Understand how to implement SVM in R
2. Understand the goals of classification
3. Interpret information regarding variables that impact classification

**Materials Needed for Lesson**

* Laptop

**Anticipatory Set up time:** Gauge how comfortable students are with the material with a series of warm-up questions, for example:

1. How do we classify things?
2. How many different types of data do we have?
3. What are some alternative ways of classifying things?

**Direct instruction and guided practice**

* Student Activity 1: Explore the bean data
  + Load in bean data into R
  + Create a visualization of the data
  + Share your visualization and code with the class
* Student Activity 2: Create a support vector machine to classify bean genotypes
  + Load in bean data into R
  + Explore and annotate the R-markdown SVM code
  + Create a confusion table
  + Describe the results to your group members
* Student Activity 3: Class discussion
  + What other problems can the classification method be used for?
  + What other types of data could be used?
  + How does this help influence how we explore the landscape?

**Independent Practice**

Use the Maize leaf shape data to create a new confusion matrix to classify genotypes

**Output:** Students will bring a confusion matrix of classification to the next class

**Follow up:** At the beginning of the next class, have students share their thoughts and what they learned during the lesson and how well their classification worked.