**Lab 9**

For this lab, we will return to the Idaho Soils dataset, and use all of the soil characteristics measured together in a PCA analysis. 30 pts total.

1. Use the data wrangling code (lines 24-26) from “Idaho soils example.R” (you can find this in the “Mixed models 2” week) to separate the “Sample\_ID” column into three columns: “Type”, “Source”, and “PlantID”.
2. **Our question is: What is the variation in soil characteristics measured among different plant species sources?** We will use all of the soil characteristics together our response matrix (OM, P1, P2, BICARB, K, Mg, Ca, pH, Nitrate, CEC). How many dimensions does our response have (2 pts)?

* Our response has 10 different dimensions which would not be able to be visualized since it is beyond 3.

1. Use ggpairs() to examine correlations among all of the response variables in this dataset (OM, P1, P2, BICARB, K, Mg, Ca, pH, Nitrate, CEC). Color the output by species “Source” from which the soil sample was taken. This plot is going to be pretty big, so you might have to zoom in to see all the different colors and numbers.
   1. What is the coefficient of correlation between Nitrate and OM (organic matter) overall, and for soils from each of the plant species sources (2 pts)?

* Coeff of correlation between Nitrate and OM = 0.341
* ARTR: 0.436, HECO: 0.278, POSE: 0.416, PSSP: 0.134
  1. What is the coefficient of correlation between Ca and CEC (cation exchange capacity) overall, and for soils from each of the plant species sources (2 pts)?
* Coeff of correlation between Ca and CEC = 0.974
* ARTR: 0.988, HECO: 0.980, POSE: 0.992, PSSP: 0.959

1. Analyze all response variables in a PCA using prcomp()
   1. Which variable has the largest amount of contribution to (or loading on) PC1 (2 pts)?

* Mg as its value is the highest.
  1. Which variable has the smallest amount of contribution to (or loading on) PC1 (2 pts)?
* pH as its value is the most negative (smallest)
  1. Chart, histogram

     Description automatically generatedMake a scree plot of the principle components using fviz\_eig(). Include plot here (2 pts).
  2. How much % of total variation do PC1 and PC2 combined explain (2pts)?
* The total variation explained by PC1 and PC2 is 70.1%

1. Make a biplot of your PCA results. Color the points according to species source using fviz\_pca\_biplot().
   1. Include the plot here (4 pts)

Chart

Description automatically generated

* 1. What are some patterns and conclusions you can draw from this biplot analysis (4 pts)? For example, in what ways do the soils from PSSP (*Pseudoregnaria spicata*) differ from the soils from other plants?
* I see that soils from HECO, POSE, and ARTR cluster in relationship the response variables relatively similarly, whereas there seems to be some relationship of Mg, BICARB, OM, K, and P1 and PSSP. Mg and K seem to have a positive correlation with PSSP soils in comparison to weak effect on the other species. Overall there is a lot of overlap between the species making it difficult to discern differences using a PCA alone.

1. Since this is the first time we will use a dataset different from the in-class example for the lab, please also turn in your R code as supplemental material. Turn in the .R file (or the R markdown file) instead of copy-pasting your code into word or something. 8 pts.