Goal: Introduce your dataset, perform structured data exploration, and describe patterns in your variables to prepare for modeling later in the semester.

Assignment Components (Total: 50 points)

1. **Study Context and Dataset Description (10 points)**

- Describe the broader purpose of the study the data were collected for.
    - The data was collected to study species that are drought tolerant, or drought avoidant based on different species and location. We aimed to see if ecosystems drive relation of plant to drought. Comparing some plant functional traits to leaf spectral value, we aim to find which leaf spectral value correlates strongly with each functional trait and this is further divided into functional traits that identify drought tolerant vs drought avoidant species and see if both behaviors have same spectral signatures.
- Describe your specific dataset:

Trait/Variables I will use	Data source	Definition of variables	Sampling context
Leaf thickness (LT; mm)	Digital caliper	general thickness, toughness	Measured at 3 different locations in the leaf and averaged to get LT
Water potential (Mpa)	Scholander type pressure chamber	Leaf water relation	Measured predawn which Measures the plant's water status after overnight equilibration, reflecting near-maximum hydration and mid-day which Captures the plant's water status during peak transpiration, indicating daily water stress.
Leaf water content (LWC; g)	Weighed the leaf fresh weight and leaf dry weight	The proportion of water in a leaf relative to its fresh mass, indicating the leaf's hydration status.	We sampled also at predawn and midday. **Predawn leaf water content:** Reflects the leaf's maximum hydration after overnight water uptake and minimal transpiration. **Midday leaf water content:** Reflects the leaf's hydration under peak transpiration and daily water stress.

Diameter by breast height (DBH; m)	Measuring tape	This is the stem diameter used to quantify size of the tree.	Measured once.
Tree height (m)	Laser scanner	The vertical distance from the base of a tree (ground level) to the top of its highest point.	Measured once.
Leaf dry mass per area (LMA; m^2/g)	Scanned the leaf for leaf area using LI-3100C Area Meter and dried weight was weighed with a measuring scale	The dry mass of a leaf divided by its surface area, indicating leaf density and structural investment.	Measured by collecting fully expanded leaves, midday and predawn, to compare leaf structural traits across species, environmental conditions.
Stress level	Physical inspection of plant stress following a protocol from USDA	A quantitative measure of the physiological strain a plant experiences which range from 0 to 10. 0 means dead/and very stressed out. And 10 means very healthy.	Assessed through physical inspection of plant stress following a USDA protocol.
Vapor pressure deficit (VPD; kpa)	Measured the leaf sample temperature using a laser scanner temperature gun. Measured environmental temperature and relative humidity using a sensor. This value is plugged into a formular which gives us VPD for each plant level and predawn and mid-day	The difference between the amount of moisture in the air and the maximum amount of moisture the air can hold at a given temperature, indicating the atmospheric demand for water.	Calculated using air temperature and relative humidity, often measured once the leaf is taken down to assess the evaporative demand on plants.
Precipitation	Remote sensor	Total amount of water (rain, snow, sleet, or hail) falling over a specific period.	Measured in millimeters at designated stations, recorded daily
Leaf spectral values	Filed collected using leaf spectral clip from ASD	Reflectance or absorbance of light by a leaf at specific	Measured on individual leaves using a spectroradiometer, at

		wavelengths, indicating physiological and biochemical properties	predawn or midday; values recorded for multiple spectral bands.

- Clearly state:
    - What **hypotheses** you may test later
        - Using the field samples, we aim to clearly group our species from each plot/ecosystem into drought tolerant vs drought avoidant.
        - Study if our ecosystem/plot drive relation of plant to drought
        - Investigate which spectral range correlates strongly with each functional trait that is also the driver of our plant drought behavior.

2. **Attribute Types and Summary Statistics (10 points)**

- Identify the **attribute type** (e.g., continuous, ordinal, binary, nominal) of each variable.

Trait/Variables I will use	attribute type	Summary
Leaf thickness (LT; mm)	Numerical (continuous)	Mean:0.34 Median:0.33 Variance:2.9e-03 Percentiles: 0.43
Water potential (Mpa)	Numerical (continuous)	Mean:-1.181 Median: -195 Variance:8.5e-02 Percentiles:-0.72
Leaf water content (LWC; g)	Numerical (continuous)	Mean: 2.427 Median: 2.445 Variance:1.9e-01 Percentiles:3.11
Diameter by breast height (DBH; m)	Numerical (continuous)	Mean:0.2417 Median:0.2500 Variance:5.8e-03 Percentiles:0.36
Tree height (m)	Numerical (continuous)	Mean: 12.12 Median:11.95 Variance:8.0e+00 Percentiles:16.42

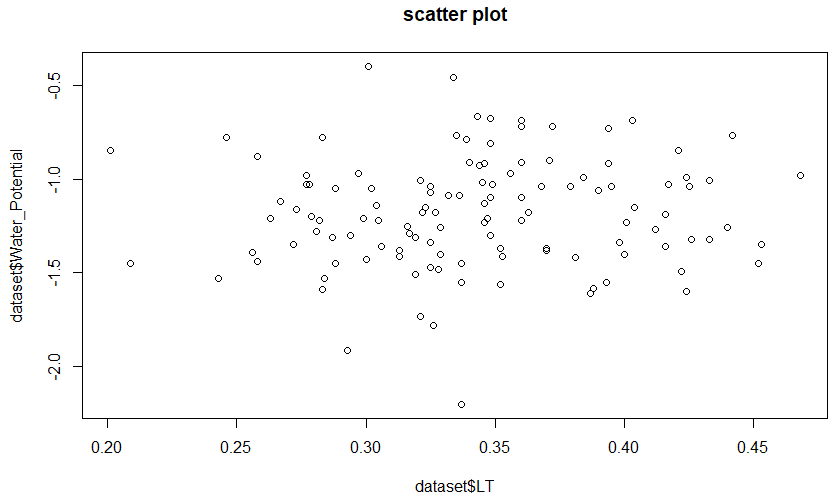
Leaf dry mass per area (LMA; m^2/g)	Numerical (continuous)	Mean: 0.01222 Median: 0.01240 Variance:1.3e-05 Percentiles:0.018
Stress level	Ordinal	Mean: 1.875 Median:2.000 Variance:6.1e-01 Percentiles:3.0
Vapor pressure deficit (VPD; kpa)	Numerical (continuous)	Mean:1.848 Median:1.895 Variance:2.2e-01 Percentiles:2.60
Precipitation	Numerical (continuous)	Mean:919.4 Median:928.0 Variance:2.6e+04 Percentiles:1196.10
Leaf spectral values	Numerical (discrete)	Mean:2884 Median:2947 Variance:1.3e+06 Percentiles:4733.5
Species	Nominal	4 different species. A=29 B=35 C=26 D=30
Plot	Nominal	9 different plots.

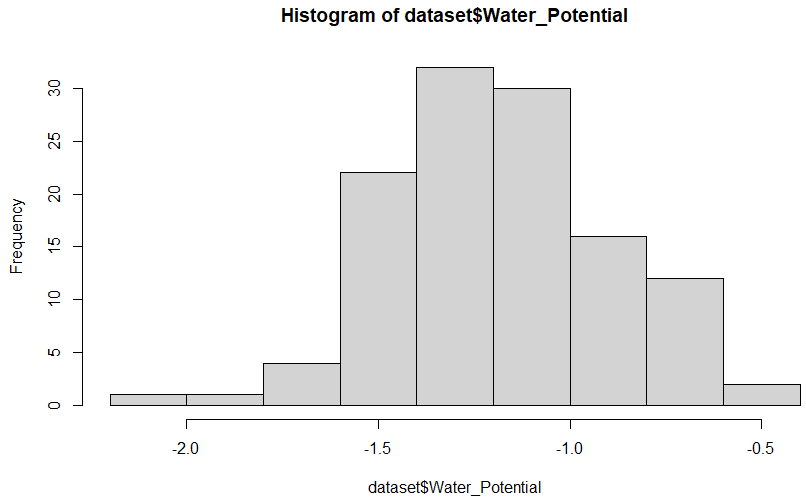
Data Exploration Using Zuur's 8-Step Protocol (20 points)

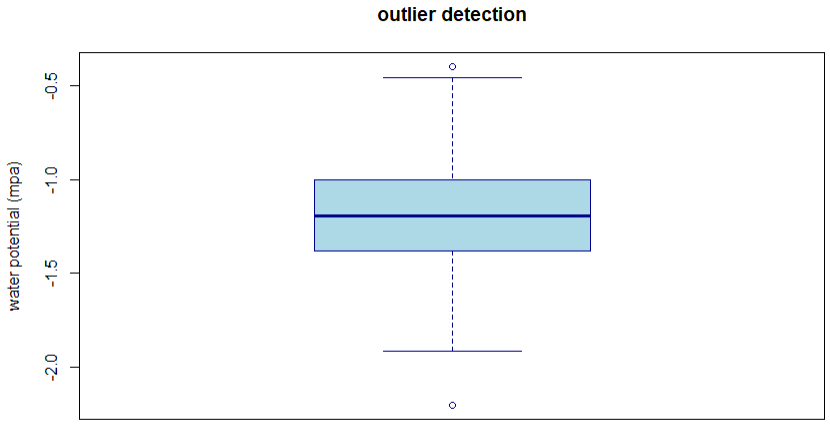
- Work through all 8 data exploration steps as discussed in class and from the *Zuur et al.* reading.
    - No outliers
    - There is a variance in the dataset. I used the boxplot to check for variance between water potential across our plots and species.
    - The water potential is not normally distributed. The histogram returned a right skewed curve.
    - No zero distribution in our dataset water potential
    - The dataset doesn't show any colleranity but it is important to note that this is a pseudo dataset.

- No interaction between my X and Y.
- There is an interaction between water potential, LT and species.
- The observations (our response variable) are independent.

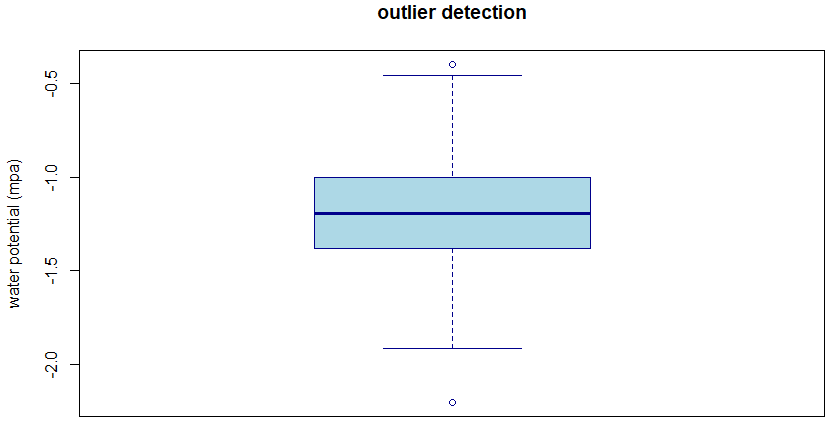
- Support with graphs and appropriate statistics:
  - Histograms, boxplots, scatterplots



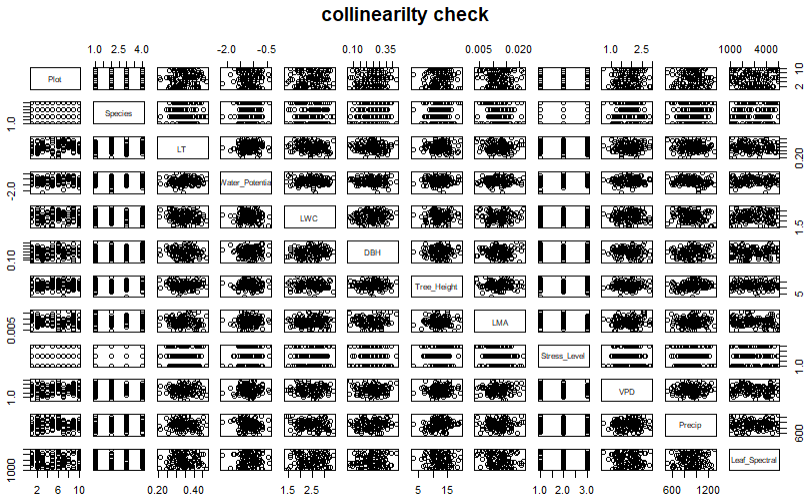




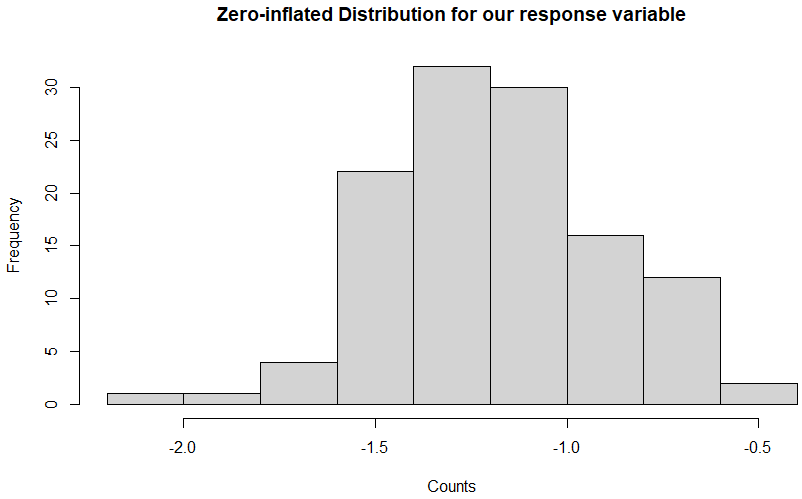
- Outlier detection



-
- Collinearity checks



-
- Missing data considerations
    - No missing data



-

Highlight Key Patterns (5 points)

- Describe the **main patterns or surprises** found in your data.
    - From what I am seeing, I am not seeing any clear pattern in the dataset. But it is good to point out that this is a pseudo dataset and the result will definitely vary when I try with my main dataset.
- Relate patterns back to your research questions or model expectations.
    - Based on the result, we cant use leaf functional traits to relate to plant drought behavior.