 **📖 Project**  | Statistical Analysis with R Using the CO2 Dataset

OBJECTIVE: In this assignment, you will apply basic statistical analysis techniques using the R programming language on the CO2 dataset (make sure you use the CO2 dataset and not the co2 dataset, they are different). This dataset details CO2 uptake in grass plants under different environmental conditions. Your tasks will include data exploration, visualization, hypothesis testing with a t-test, and examining correlations.

DATA OVERVIEW: The CO2 dataset contains observations from an experiment on the cold tolerance of the grass species Echinochloa crus-galli. Variables include treatment types, CO2 uptake, concentration, temperature, and more. This rich dataset allows for comprehensive statistical analysis.

INSTRUCTIONS: In your own R script file, please complete the following tasks:

1. **Data Exploration**: Familiarize yourself with the dataset using the head() and summary() functions in R.
2. **Data Visualization**: Create visualizations to understand the distributions and relationships in the data:
3. **t-Test**: Conduct a t-test to compare the mean CO2 uptake between two treatment groups. Clearly state your hypothesis, perform the test, and interpret the results.
4. **Correlation Analysis**: Calculate and interpret the correlation coefficients between CO2 uptake and other numeric variables in the dataset.

**— Deliverables**

* An R Script with comments explaining your findings, insights, and any patterns observed during your exploration.
* Create a slide presentation (e.g., PowerPoint, Google Slides) with the results of the tasks, emphasizing the interpretation of your findings.
  + Take a look at [SlidesGo](https://slidesgo.com/) and [Slides Carnival](https://www.slidescarnival.com/) for fantastic, free slide templates!
* Submit your presentation as a PDF.
* You will prepare a 5-10 minute presentation and submit a recording of your presentation. This practice is valuable for simulating live presentations, a crucial skill in data analytics.

**— Rubric (Total: 20 Points)**

#### **1. Data Exploration (4 points)**

* **Excellent (4 points):** Comprehensive examination and description of the dataset, with detailed summary statistics.
* **Good (3 points):** Adequate exploration with some detailed observations and summary statistics.
* **Adequate (2 points):** Basic exploration with general observations and minimal statistics.
* **Poor (0-1 points):** Limited or incorrect exploration with missing or inaccurate statistics.

#### **2. Data Visualization (4 points)**

* **Excellent (4 points):** Visualizations are insightful, accurately represented, and creatively displayed.
* **Good (3 points):** Competent visualizations that correctly represent the data.
* **Adequate (2 points):** Basic visualizations that display data without errors but lack creativity.
* **Poor (0-1 points):** Incorrect or misleading visualizations.

#### **3. t-Test Execution and Interpretation (4 points)**

* **Excellent (4 points):** Correct application of the t-test, with a well-articulated hypothesis and thorough interpretation of results.
* **Good (3 points):** Correct t-test application with a clear hypothesis and adequate interpretation.
* **Adequate (2 points):** Basic t-test application with some errors in hypothesis formulation or interpretation.
* **Poor (0-1 points):** Incorrect t-test application with significant errors or misinterpretations.

#### **4. Correlation Analysis (4 points)**

* **Excellent (4 points):** Precise calculation of correlation with deep analysis and interpretation of its implications.
* **Good (3 points):** Correct correlation calculation with basic interpretation of results.
* **Adequate (2 points):** Some errors in correlation analysis with limited interpretation.
* **Poor (0-1 points):** Incorrect or failed correlation analysis.

#### **5. Clarity and Presentation of Report (4 points)**

* **Excellent (4 points):** The script and presentation is exceptionally well-organized, with clear, concise, and professional presentation.
* **Good (3 points):** Clear presentation with minor organizational or clarity issues.
* **Adequate (2 points):** Understandable presentation with some issues in clarity and organization.
* **Poor (0-1 points):** Difficult to understand due to poor organization and presentation.