Plant Growth Data Analysis

**Introduction**

This repository contains R code and documentation for the analysis of plant growth data. The analysis includes fitting a multiple linear regression model to predict plant growth rates based on several environmental factors. Additionally, key assumptions of the linear model are assessed, and relationships between variables are visualized using various plots and charts.

**Dataset**

The dataset used in this analysis is stored in a CSV file named "plant\_growth\_data.csv."

The dataset includes the following variables:

* plant\_id: Identifier for each plant.
* sunlight\_exposure: Sunlight exposure level.
* soil\_ph: Soil pH level.
* precipitation: Precipitation amount.
* average\_temperature: Average temperature.
* growth\_rate: The growth rate of the plants.

**Analysis Steps**

The R code provided in the analysis can be broken down into several key steps:

1. **Data Loading and Model Fitting**

Load the plant growth data from the CSV file.

Fit a multiple linear regression model to predict growth rate based on environmental variables.

1. **Model Summary**

Print a summary of the linear regression model, including coefficients, standard errors, t-values, p-values, and important model statistics.

1. **Assumptions of the Linear Model**

* Check for multicollinearity among predictor variables using VIF (Variance Inflation Factor).
* Assess the normality of residuals using the Shapiro-Wilk normality test.
* Examine homoscedasticity (constant variance of residuals) using a plot.

1. **Interpretation of Coefficients**

Extract coefficients and their statistical significance from the model summary and present them in a tabular format.

1. **Visualization**

Visualize relationships between variables using the following plots:

* Scatter plot of growth rate vs. sunlight exposure.
* Scatter plot matrix (pairs plot) for all variables.
* Bar plot showing the average growth rate by soil pH.
* Heatmap illustrating the correlation between variables.

**Dependencies**

This analysis uses R as the programming language.

Several R packages are required for data manipulation, visualization, and statistical analysis, including ggplot2, car, knitr, and GGally. Make sure to install these packages if not already installed.

**Running the Analysis**

To replicate the analysis:

1. Ensure you have R installed on your system.
2. Install the required R packages using the following commands:

***install.packages("ggplot2")***

***install.packages("car")***

***install.packages("knitr")***

***install.packages("GGally")***

1. Clone or download this repository to your local machine.
2. Open the R script containing the analysis code in your preferred R development environment (e.g., RStudio).
3. Run the code step by step to perform data analysis, modeling, and visualization.

**Conclusion**

This analysis provides insights into the relationships between plant growth and environmental factors, including sunlight exposure, soil pH, precipitation, and temperature. The linear regression model helps in understanding how these variables collectively affect plant growth rates. Additionally, diagnostic checks and visualizations enhance the interpretation of the data and model results.

For questions or further assistance, please contact [Your Name] at [Your Email Address].