R for Plant Nutrition Research

T.S. Murrell, Ph.D.

2023-12-15

Table of contents

Preface

This is a Quarto book. To learn more about Quarto books visit https://quarto.org/docs/books.

1 License

2 License

This work is licensed under Creative Commons CC BY-NC-ND 4.0 DEED, Attribution-NonCommercial-NoDerivs 4.0 International. You are free to copy and redistribute the material in any medium or format under the following terms:

3 Introduction

This book is intended to be a living compendium of procedures used in the field of plant nutrition. Because it is housed on GitHub, it is open-source and open-access, allowing interested experts to contribute, either by submitting additional procedures or by editing existing ones. We hope that, over time, this work becomes a source of best practices operationalized routinely by students, staff, and faculty in educational organizations, as well as by professionals in business and industry.

We dedicated ourselves to using only software that is open-access. This decision is meant to democratize this guidance for everyone with access to only a computer and the internet. Software applications with hidden fees and paywalls were avoided. All content was produced using Quarto for books, and using GitHub as a repository for all book content, including figures and data sets. Editing was performed in RStudio Desktop, and all code chunks were written in R. When topics related to spreadsheets were addressed, we used Google Sheets.

We hope you find this resource useful. If you see ways to improve it, please contact info@apni.net.

4 Preparing External Data

In this chapter, we discuss how to prepare external files that you have access to and can edit. We limit our discussion to spreadsheet files, such as Microsoft Excel and Google Sheets, since they are widely used in plant nutrition research.

We discuss best practices for:

- organizing data
- labeling columns
- denoting missing values
- using data validation rules

4.1 Organizing Data

R requires that data be "tidy" According to (Wickham, Vaughan, and Girlich 2023), tidy data are structured so that:

- Each column is a unique variable
- Every row is an observation for each variable; if no observation exists for a given variable, it is acceptable to use an entry that denotes a missing value
- Every cell contains only one value

An example of a spreadsheet with tidy data is shown below ().

Summary

No summary currently exists, but will be written after the intended content has been created.

Acknowledgements

The authors with to acknowledge the many farmers, agricultural service providers, university students, faculty, and staff, as well as the statisticians and data scientists who have shared their time, expertise, insights, and guidance over the past many years. All of these interactions have created the vision for this work.

quarto publish gh-pages

References

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2023. *Tidyr: Tidy Messy Data.* https://tidyr.tidyverse.org.