CS506: Data Wrangling and Management

Norah Jones

2025-07-24

Table of contents

Preface

Welcome to CS506: Data Wrangling and Management. This course introduces graduate students to data wrangling and management using R and the Tidyverse ecosystem. Students will learn to import, manipulate, clean, and visualize data with a strong emphasis on practical applications and reproducible workflows. The course will utilize the free textbook *R for Data Science* by Hadley Wickham and Garrett Grolemund.

Course Objectives: Upon successful completion of the course, students will be able to:

- Develop an understanding of R and the Tidyverse ecosystem

- Import structured and unstructured data into R

- Clean and transform data using `dplyr`, `tidyr`, and other core Tidyverse packages

- Visualize data effectively using `ggplot2`

- Conduct exploratory data analysis (EDA)

- Apply data wrangling techniques to real-world datasets

Textbook: *R for Data Science* by Hadley Wickham & Garrett Grolemund (Available for free: https://r4ds.had.co.nz/)

Software Requirements:

- R (https://www.r-project.org/)

- RStudio (https://posit.co/downloads/)

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```

2 Summary

In summary, this book has no content whatsoever.

```r
1 + 1
```

```
[1] 2
```

References

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.