Importing and working with data in R

Frederick Anyan

11 October 2024

Workshop objectives:

- Importing data from diverse sources using RStudio, base R functions, and popular libraries (`readr`, `haven`, `googlesheets4`, etc.).
- Key techniques for cleaning, transforming, and manipulating datasets using `dplyr` and other tools.
- Best practices for handling missing data, reshaping datasets, and performing complex transformations using chaining and piping.
- How to save and export data in multiple formats, ensuring compatibility with other tools and software.

By the end of this workshop, you will gain practical skills to confidently import, handle and manipulate data in R. No prior experience with R is required, but familiarity with basic programming concepts is helpful.

Workshop outline:

1. **Introduction**

- **Objectives of the workshop**
  - Data importation and manipulation in R
  - Different data sources and formats
  - Common libraries for data importation and manipulation

---

2. **Data Importation Techniques**

- **A. Importing CSV Files**
  - Point-and-click method via RStudio
  - Base R method: `read.csv()`
  - `readr` package: `read_csv()`, `read_tsv()`, `read_csv2()`, and `read_delim()`
  - `data.table` method: `fread()`

- **B. Importing SPSS Files**
  - Importing SPSS files: `haven` package: `read_spss()` and `read_sav()`

- **C. Importing Excel and Google Sheet Files**
  - `readxl` package: `read_excel()`
  - `openxlsx` package: `read.xlsx()`
  - Importing from Google Sheets: `googlesheets4` Packagae: `gs4_get()` and `read_sheet()`

---

3. **Data Manipulation**

- **A. Base R vs. Tidyverse**
  - Differences between base R and tidyverse approaches
  - Introduction to `dplyr` for data manipulation
  - Chaining with pipes (`%>%`) for cleaner workflows

- **B. Exploring Data**
  - Viewing data: `head()`, `tail()`, and `summary()`
  - Handling missing data: `na.omit()` and `is.na()`
  - Renaming columns, filtering, and selecting columns

- **C. Data Transformation**
  - Adding or transforming columns: `mutate()`
  - Grouping and summarizing data: `group_by()` and `summarize()`
  - Sorting data: `arrange()`
  - Merging datasets: `merge()`

- **D. Reshaping Data**
  - Pivoting data: `pivot_longer()`, `pivot_wider()`

---

4. **Data Output or Saving**

- **A. Writing data to**
  - Saving .csv format `write.csv()` and `write_csv2()`
  - Saving data as SPSS formats `write_sav()`
