1.3.2: Data Wrangling

(Asynchronous-Online)

Session Objectives

1. To read in data.
2. To subset data.
3. To select, create, and modify variables.
4. To filter the observations according to a certain condition.
5. To explore and summarize data.
6. To run descriptive statistics.

Key points to cover: 1. Import data. 2. Introduce to tidyverse. 3. Summarize data and run descriptive statistics. 4. Introduce other resources (e.g., books, blogs, or websites) trainees can refer to.

---

0. Prework - Before You Begin

Install Packages

Before you begin, please go ahead and install the following packages - these are all on CRAN, so you can install them using the RStudio Menu Tools/Install Packages interface:

- haven
- readr
- readxl
- dplyr

---

1. To read in data.

Begin with a NEW Project

add steps and screenshot on starting with a new project...

importing data

exploring builtin datasets - see environment also - including data with packages

---

2. To subset data.

3. To select, create, and modify variables.

---

4. To filter the observations according to a certain condition.

———————————————————————

5. To explore and summarize data.

base r - summary stats

making tables

gtsummary package

other packages - arsenal, tableone and Rmarkdown formatting of tables... get inspired

table competitions

---

6. To run descriptive statistics.

References

R Core Team. 2024. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Other Helpful Resources

[Other Helpful Resources](#)