

Learning Goals

- Load external data into R using `read.csv()` or the RStudio Import Dataset interface.
- Examine the structure of a dataset using `str()` and determine which variables should be converted to factors.
- Use the `table()` function to summarize and visualize categorical variables and their combinations with barplots.
- Use logical vectors to filter data and compute summaries like means for specific subgroups.
- Use the `aggregate()` function to compute summary statistics for quantitative variables grouped by one or more categorical variables.

Key Functions

For each of the following functions below, write down a brief definition of what it does and a basic example.

- `read.csv()`:
- `as.factor()`:
- `table()`:
- `barplot()`:
- `aggregate()`:

Key Concepts

1. What is the difference between loading data using `read.csv()` and using the RStudio Import Dataset interface?

2. What kinds of variables should usually be converted to factors?
3. How is a two-way table different from a one-way table?
4. What is the purpose of the `aggregate()` function?
5. Why does it matter whether the variable we group by in `aggregate()` is categorical or quantitative?

Practice Problems

1. Use the small dataset below to answer the following questions.

Student	Major	Year	Study Hours
A	Data	Junior	12
B	Math	Senior	15
C	Data	Junior	10
D	Biology	First-Year	8
E	Math	Junior	18
F	Biology	Senior	11

- (a) Write R code that creates a two-way table showing the relationship between **Major** and **Year**. Then explain what the table tells us.
- (b) Write R code that uses `aggregate()` to calculate the average **Study Hours** by **Major**. Then explain which major studies the most on average in this small dataset.