

# BioR — Challenge: Replicates vs Barplots (dplyr + ggplot2)

Goal: summarise replicates correctly and compare a misleading plot vs a better plot.

## Context

You ran a treatment experiment with multiple replicates across sites (or batches). Your task is to summarise the data and produce two plots: a barplot of means (bad on purpose) and a plot that shows the raw replicates (better).

## What to submit

- summary\_table.csv
- fig1\_bad.png (barplot of means only)
- fig2\_good.png (raw replicates + mean  $\pm$  SE)
- One script: scripts/replicates\_vs\_barplots.R

## Data (import from the folder)

```
dat_treat <- read.csv("dat_treat.csv", header=TRUE)
```

## Tasks

- 1) Inspect the dataset (glimpse/summary). Identify how many rows are missing length\_mm.
- 2) Create a summary table using group\_by(treatment, site) and summarise(mean, sd, se, n).
  - Use na.rm = TRUE for mean/sd and compute n as number of non-missing observations.
- 3) Export the summary as summary\_table.csv.
- 4) Create fig1\_bad.png: a barplot of means (from the summary table).
  - Add clear axis labels + units (Length [mm]) and a clean theme.
- 5) Create fig2\_good.png: show raw replicates + mean  $\pm$  SE.
  - Option A: jittered points + mean point + error bars (recommended).
  - Option B: boxplot + points.
- 6) In comments: Explaining why the barplot can be misleading.

## Hints

- SE (standard error) = sd / sqrt(n). Use n = number of non-missing values.
- For fig2\_good, use stat\_summary() or join your summary table back onto raw data.
- Save plots with: ggsave('fig1\_bad.png', plot = p1, width = 7, height = 4, dpi = 300).