BioR — Challenge: Enzyme kinetics (Michaelis–Menten without fitting)

Goal: summarise replicate kinetics data, flag outliers, and visualise results with ggplot2.

Scenario

You performed an enzyme kinetics assay measuring reaction rate at multiple substrate concentrations. Each concentration was measured with several technical replicates. You tested two enzyme variants (WT and MutA) with and without an inhibitor. Instead, you will tidy and summarise the data, flag potential outliers using an IQR rule, and produce a clear figure showing mean trends and uncertainty. Your task is NOT to fit a Michaelis–Menten model.

Learning targets

- dplyr: group_by() + summarise() for mean, SD, and n (with missing data).
- QC: outlier flag using a simple IQR rule within each concentration group.
- ggplot2: raw points + mean trend + uncertainty (SD error bars or ribbon).

Deliverables

- kinetics_summary.csv (mean_rate, sd_rate, n per variant × inhibitor × conc_uM)
- kinetics_outliers.csv (rows flagged as outliers + thresholds used)
- kinetics_plot.png (raw points + mean ± SD, faceted by variant and/or inhibitor)
- scripts/enzyme_kinetics_challenge.R

Data (import)

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

Tasks

- A) Inspect the dataset: how many rows are missing rate?
- B) Summarise mean ± SD by variant × inhibitor × conc_uM (compute n as non-missing rates).
- C) Outliers (QC): within each group compute Q1, Q3, IQR; lower = Q1 - 1.5×IQR; upper = Q3 + 1.5×IQR; flag rows outside [lower, upper].
- D) Export: kinetics_summary.csv and kinetics_outliers.csv.
- E) Plot: raw points + mean trend + SD error bars (or ribbon), faceted by variant and/or inhibitor; label axes with units.
- F) In comments: 2–3 lines on what outliers might mean in lab data.

Hints

- Use na.rm=TRUE in mean() and sd().
- Use sum(!is.na(rate)) for n (not n()).
- Ribbon approach: ymin = mean_rate - sd_rate; ymax = mean_rate + sd_rate.
- Optional: scale_x_log10() for wide concentration ranges.

- Save with ggsave('kinetics_plot.png', width=10, height=6, dpi=300).

Bonus (optional, no fitting)

- Vmax approx = max(mean_rate) per condition (variant × inhibitor).
- Km approx = concentration closest to Vmax/2 (per condition).