

# STA 440 Case 2: Code Script

Olivia Fu, Christina Lee, Eunice Lee

## Package Import & Data Cleaning

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
library(tidyr)
library(readr)
library(ggplot2)
```

```
data <- read_csv("J02.csv", col_names = FALSE, show_col_types = FALSE)
```

```
# extract substrate and dose
substrate <- as.character(data[1, ])
dose <- as.character(data[2, ])

# fill in substrate for each column
substrate <- trimws(sub("\\s*\\(.*\\)$", "", substrate))
substrate <- na_if(substrate, "")
```

```
substrate <- na_if(substrate, "NA")
substrate <- fill(tibble(Substrate = substrate),
                  Substrate, .direction = "down")$Substrate

# build new column names
new_names <- ifelse(tolower(dose) == "subject", "Subject",
                    paste0(substrate, ".", dose))
data <- data[-c(1,2), ]
names(data) <- make.unique(new_names, sep = "_")

# solve type issue
val_cols <- setdiff(names(data), "Subject")
data[val_cols] <- lapply(data[val_cols], function(x) {
  if (is.character(x)) parse_number(x) else as.numeric(x)
})

# make the longer table
data_cleaned <- data |>
  pivot_longer(
    cols = -Subject,
    names_to = c("substrate", "dose"),
    names_pattern = "^([^.]+)\\.\\.(.+)$", # split once at first dot
    values_to = "Value"
  ) |>
  mutate(
    group = ifelse(grepl("^NT", Subject, ignore.case=TRUE), "NT", "Tg")
  )
```

## Exploratory Data Analysis

## Trend Exploration & Visualization

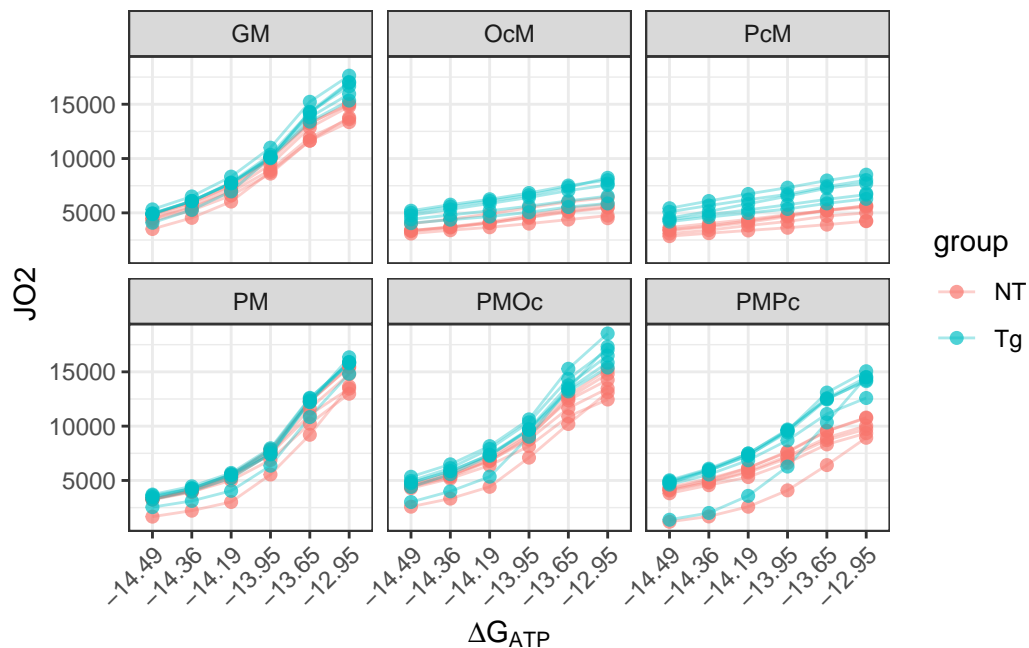
```
# treat dose as ordered categorical levels
# group by NT/Tg to see the difference
dose_levels <- c("-14.49", "-14.36", "-14.19", "-13.95", "-13.65", "-12.95")

data_plot <- data_cleaned |>
  # don't include basal in the plot
  filter(dose != "Basal") |>
  mutate(dose = factor(dose, levels = dose_levels, ordered = TRUE))
```

```
ggplot(data_plot, aes(dose, Value,
                      group = interaction(Subject, substrate),
                      color = group)) +
  geom_line(alpha = 0.35) +
  geom_point(alpha = 0.7, size = 1.8) +
  facet_wrap(~ substrate) +
  labs(x = expression(Delta*G[ATP]), y = "J02") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

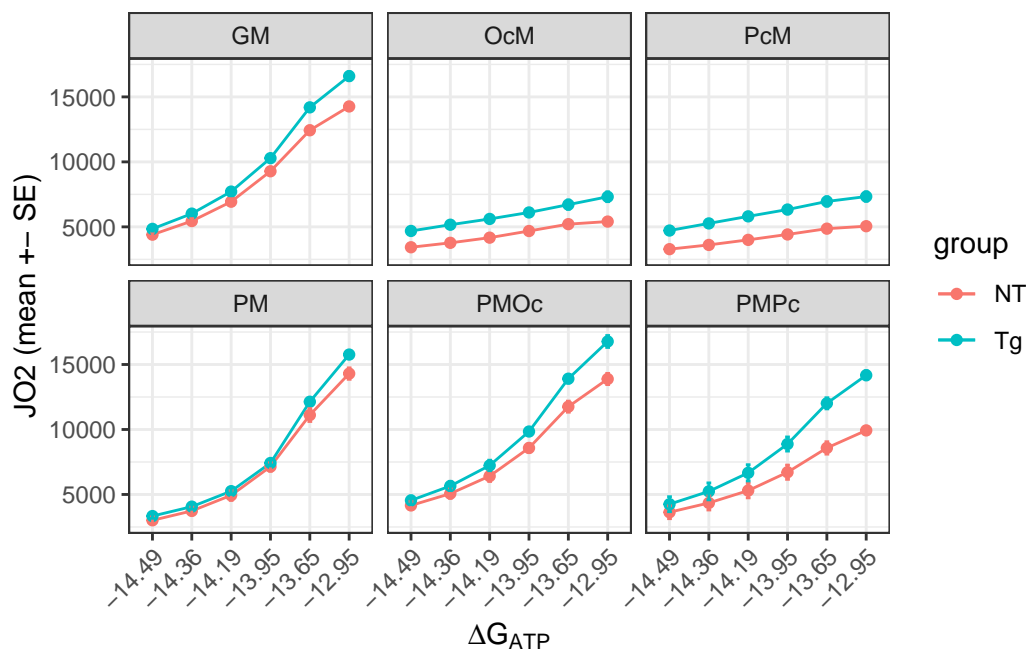
Warning: Removed 20 rows containing missing values or values outside the scale range (`geom\_line()`).

Warning: Removed 20 rows containing missing values or values outside the scale range (`geom\_point()`).



```
# calculate mean and se for each group
sum_df <- data_plot |>
  group_by(substrate, group, dose) |>
  summarise(mean = mean(Value, na.rm = TRUE),
            se = sd(Value, na.rm = TRUE)/sqrt(n()),
            .groups = "drop")
```

```
ggplot(sum_df, aes(dose, mean, color = group, group = group)) +
  geom_line() +
  geom_point() +
  geom_errorbar(aes(ymin = mean - se,
                    ymax = mean + se,
                    width = 0.12)) +
  facet_wrap(~ substrate) +
  labs(x = expression(Delta*G[ATP]),
       y = "JO2 (mean +- SE)") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



## Model Fitting & Testing