

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

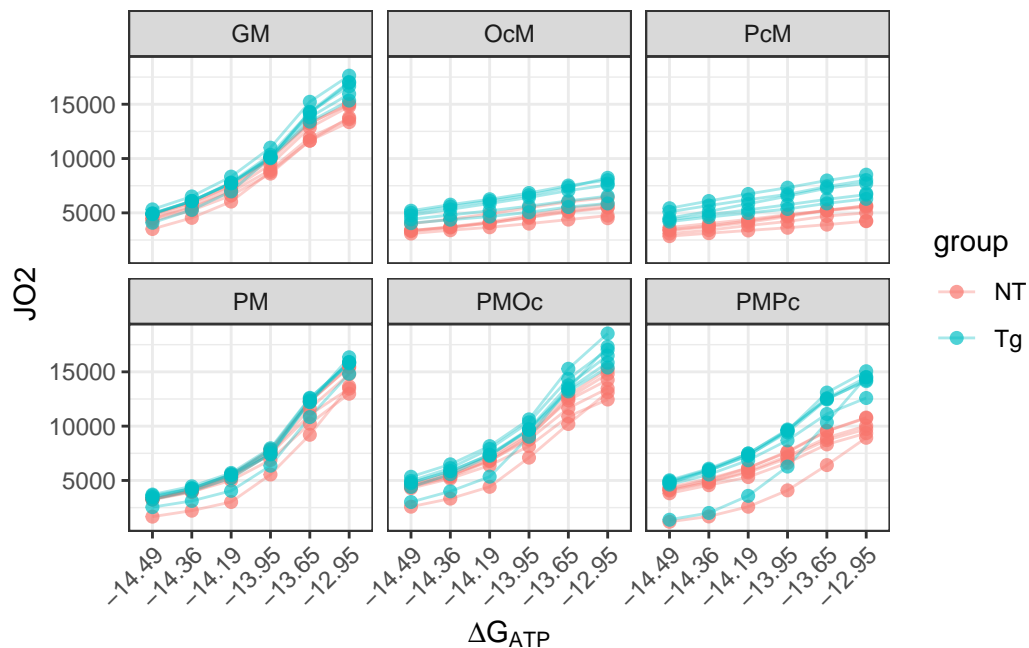
```
# treat dose as ordered categorical levels
# group by NT/Tg to see the difference
dose_levels <- c("-14.49", "-14.36", "-14.19", "-13.95", "-13.65", "-12.95")

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))
```

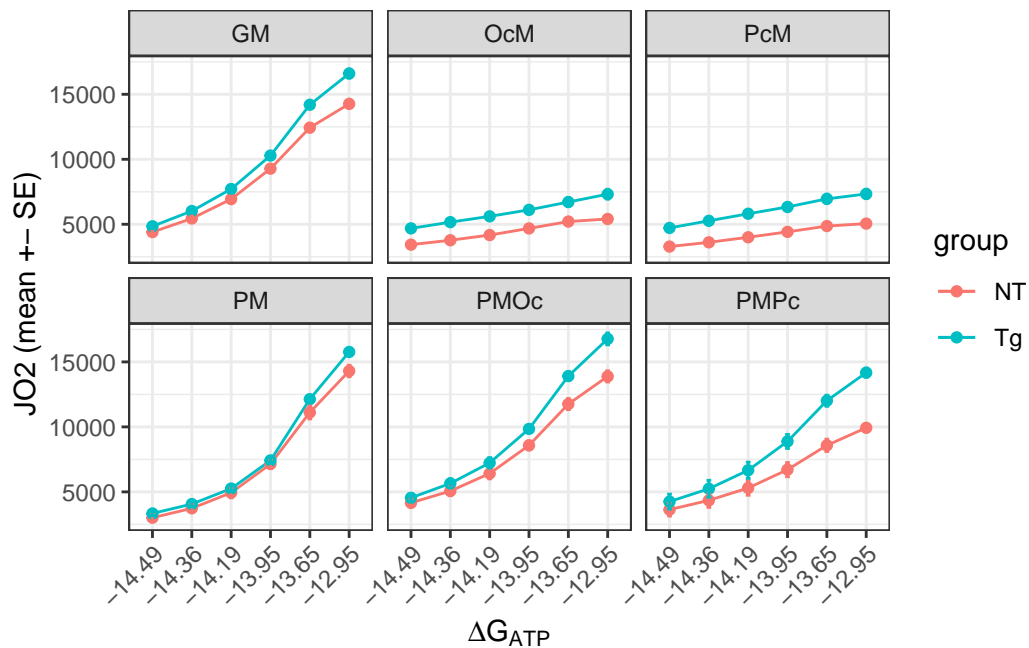
Warning: Removed 20 rows containing missing values or values outside the scale range (`geom_line()`).

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

```
# data frame for LME
library(dplyr)

dose_levels <- c("-14.49", "-14.36", "-14.19", "-13.95", "-13.65", "-12.95")

data_lme <- data_cleaned %>%
  # exclude basal for modeling
  filter(dose != "Basal") %>%
  mutate(group = factor(group, levels = c("NT", "Tg"))),
```

```

    substrate = factor(substrate),
    dose = factor(dose, levels = dose_levels, ordered = TRUE),
    dose_num = as.numeric(dose),
    Subject = factor(Subject)
  )

```

```

# linear mixed effects model: all possible interaction terms
library(lme4)

```

Loading required package: Matrix

Attaching package: 'Matrix'

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

expand, pack, unpack

Warning in check_dep_version(): ABI version mismatch:
 lme4 was built with Matrix ABI version 2
 Current Matrix ABI version is 1
 Please re-install lme4 from source or restore original 'Matrix' package

```

lme_all <- lmer(
  Value ~ group * substrate * dose + (1 + dose_num | Subject),
  data = data_lme,
  REML = TRUE,
  control = lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e5))
)

summary(lme_all)

```

Linear mixed model fit by REML ['lmerMod']
 Formula: Value ~ group * substrate * dose + (1 + dose_num | Subject)
 Data: data_lme
 Control: lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))

REML criterion at convergence: 5356.7

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.9474	-0.5456	0.0383	0.4960	2.4295

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subject	(Intercept)	1050353	1024.9	
	dose_num	24718	157.2	-0.73
Residual		215963	464.7	

Number of obs: 412, groups: Subject, 12

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	8793.66	307.41	28.605
groupTg	1149.72	434.75	2.645
substrate0cM	-4603.14	114.95	-40.043
substratePcM	-4844.10	114.95	-42.139
substratePM	-1419.88	109.54	-12.963
substratePM0c	-483.10	109.54	-4.410
substratePMPc	-2375.62	109.54	-21.688
dose.L	8679.60	328.77	26.400
dose.Q	1155.00	189.72	6.088
dose.C	-667.64	189.72	-3.519
dose^4	-480.15	189.72	-2.531
dose^5	-102.32	189.72	-0.539
groupTg:substrate0cM	385.46	162.57	2.371
groupTg:substratePcM	762.99	162.57	4.693
groupTg:substratePM	-526.46	154.91	-3.399
groupTg:substratePM0c	199.49	154.91	1.288
groupTg:substratePMPc	970.56	154.91	6.265
groupTg:dose.L	1582.98	464.95	3.405
groupTg:dose.Q	481.93	268.31	1.796
groupTg:dose.C	15.08	268.31	0.056
groupTg:dose^4	-122.71	268.31	-0.457
groupTg:dose^5	-112.49	268.31	-0.419
substrate0cM:dose.L	-6668.55	278.10	-23.979
substratePcM:dose.L	-6872.51	278.10	-24.712
substratePM:dose.L	974.30	268.31	3.631
substratePM0c:dose.L	-208.31	268.31	-0.776
substratePMPc:dose.L	-3229.92	268.31	-12.038
substrate0cM:dose.Q	-1056.24	277.94	-3.800
substratePcM:dose.Q	-1085.39	277.94	-3.905
substratePM:dose.Q	1397.85	268.31	5.210
substratePM0c:dose.Q	308.18	268.31	1.149

substratePMPc:dose.Q	-408.76	268.31	-1.524
substrateOcM:dose.C	582.47	279.79	2.082
substratePcM:dose.C	630.71	279.79	2.254
substratePM:dose.C	367.31	268.31	1.369
substratePMOc:dose.C	157.84	268.31	0.588
substratePMPc:dose.C	384.78	268.31	1.434
substrateOcM:dose^4	445.76	280.99	1.586
substratePcM:dose^4	475.13	280.99	1.691
substratePM:dose^4	-98.82	268.31	-0.368
substratePMOc:dose^4	19.37	268.31	0.072
substratePMPc:dose^4	253.87	268.31	0.946
substrateOcM:dose^5	109.43	281.36	0.389
substratePcM:dose^5	95.97	281.36	0.341
substratePM:dose^5	-118.43	268.31	-0.441
substratePMOc:dose^5	-29.14	268.31	-0.109
substratePMPc:dose^5	54.89	268.31	0.205
groupTg:substrateOcM:dose.L	-1117.65	393.30	-2.842
groupTg:substratePcM:dose.L	-868.89	393.30	-2.209
groupTg:substratePM:dose.L	-658.23	379.44	-1.735
groupTg:substratePMOc:dose.L	524.43	379.44	1.382
groupTg:substratePMPc:dose.L	1591.23	379.44	4.194
groupTg:substrateOcM:dose.Q	-405.33	393.06	-1.031
groupTg:substratePcM:dose.Q	-570.49	393.06	-1.451
groupTg:substratePM:dose.Q	87.27	379.44	0.230
groupTg:substratePMOc:dose.Q	95.74	379.44	0.252
groupTg:substratePMPc:dose.Q	157.19	379.44	0.414
groupTg:substrateOcM:dose.C	121.89	395.68	0.308
groupTg:substratePcM:dose.C	-14.64	395.68	-0.037
groupTg:substratePM:dose.C	58.76	379.44	0.155
groupTg:substratePMOc:dose.C	-34.07	379.44	-0.090
groupTg:substratePMPc:dose.C	-230.35	379.44	-0.607
groupTg:substrateOcM:dose^4	134.78	397.38	0.339
groupTg:substratePcM:dose^4	78.86	397.38	0.198
groupTg:substratePM:dose^4	-79.79	379.44	-0.210
groupTg:substratePMOc:dose^4	-24.14	379.44	-0.064
groupTg:substratePMPc:dose^4	-73.57	379.44	-0.194
groupTg:substrateOcM:dose^5	103.03	397.90	0.259
groupTg:substratePcM:dose^5	86.25	397.90	0.217
groupTg:substratePM:dose^5	-61.58	379.44	-0.162
groupTg:substratePMOc:dose^5	58.72	379.44	0.155
groupTg:substratePMPc:dose^5	47.10	379.44	0.124

Correlation matrix not shown by default, as $p = 72 > 12$.
 Use `print(x, correlation=TRUE)` or
`vcov(x)` if you need it

```
# exclude higher-order polynomial terms for dose from model above
data_lme_L <- data_lme %>%
  mutate(dose_num = as.numeric(as.character(dose)),
         dose_numc = scale(dose_num, center = TRUE, scale = FALSE))

lme_all_L <- lmer(
  Value ~ group * substrate * dose_numc + (1 + dose_num | Subject),
  data = data_lme_L,
  REML = TRUE,
  control = lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e5))
)

summary(lme_all_L)
```

Linear mixed model fit by REML ['lmerMod']
 Formula: Value ~ group * substrate * dose_numc + (1 + dose_num | Subject)
 Data: data_lme_L
 Control: lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))

REML criterion at convergence: 6463.9

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.5597	-0.5420	-0.1208	0.3791	3.2690

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subject	(Intercept)	40120992	6334.1	
	dose_num	219210	468.2	0.99
Residual		750501	866.3	

Number of obs: 412, groups: Subject, 12

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	8793.66	328.98	26.730
groupTg	1149.72	465.25	2.471
substrateOcM	-4608.96	214.23	-21.514
substratePcM	-4848.86	214.23	-22.634

substratePM	-1419.88	204.19	-6.954
substratePMOc	-483.10	204.19	-2.366
substratePMPC	-2375.62	204.19	-11.634
dose_numc	6698.31	338.31	19.799
groupTg:substrateOcM	384.16	302.97	1.268
groupTg:substratePcM	757.75	302.97	2.501
groupTg:substratePM	-526.46	288.77	-1.823
groupTg:substratePMOc	199.49	288.77	0.691
groupTg:substratePMPC	970.56	288.77	3.361
groupTg:dose_numc	1292.30	478.45	2.701
substrateOcM:dose_numc	-5217.75	403.07	-12.945
substratePcM:dose_numc	-5368.61	403.07	-13.319
substratePM:dose_numc	1075.45	394.77	2.724
substratePMOc:dose_numc	-72.37	394.77	-0.183
substratePMPC:dose_numc	-2473.60	394.77	-6.266
groupTg:substrateOcM:dose_numc	-930.75	570.02	-1.633
groupTg:substratePcM:dose_numc	-810.09	570.02	-1.421
groupTg:substratePM:dose_numc	-472.56	558.28	-0.846
groupTg:substratePMOc:dose_numc	412.35	558.28	0.739
groupTg:substratePMPC:dose_numc	1208.17	558.28	2.164

Correlation matrix not shown by default, as $p = 24 > 12$.

Use `print(x, correlation=TRUE)` or
`vcov(x)` if you need it