Lab 4: Tidying your dataset

Riley Payung

Online 2D1

Prune the dataset

Exercise 1

```
brauer2 <- brauer %>% select(NAME,"G0.05":"U0.3")
```

Enforcing one observation per row

Exercise 2

```
brauer3 <- brauer2 %>% gather(G0.05:U0.3,key = "sample", value="expression")
```

Enforcing one column per variable and one value per cell

Exercise 3

```
brauer4 <- brauer3 %>%
  separate(
    col=sample,
    into=combine("nutrient","rate"),
    sep=1,
    convert=TRUE
  )
```

Exercise 4

```
  brauer5 <- brauer4 %>%
    separate(
      col = NAME,
      into = combine("gene_name", "biological_process", "molecular_function", "systematic_id", "numbe
r"),
      sep="\\|\\|",
      convert = TRUE
    )
```

Exercise 5

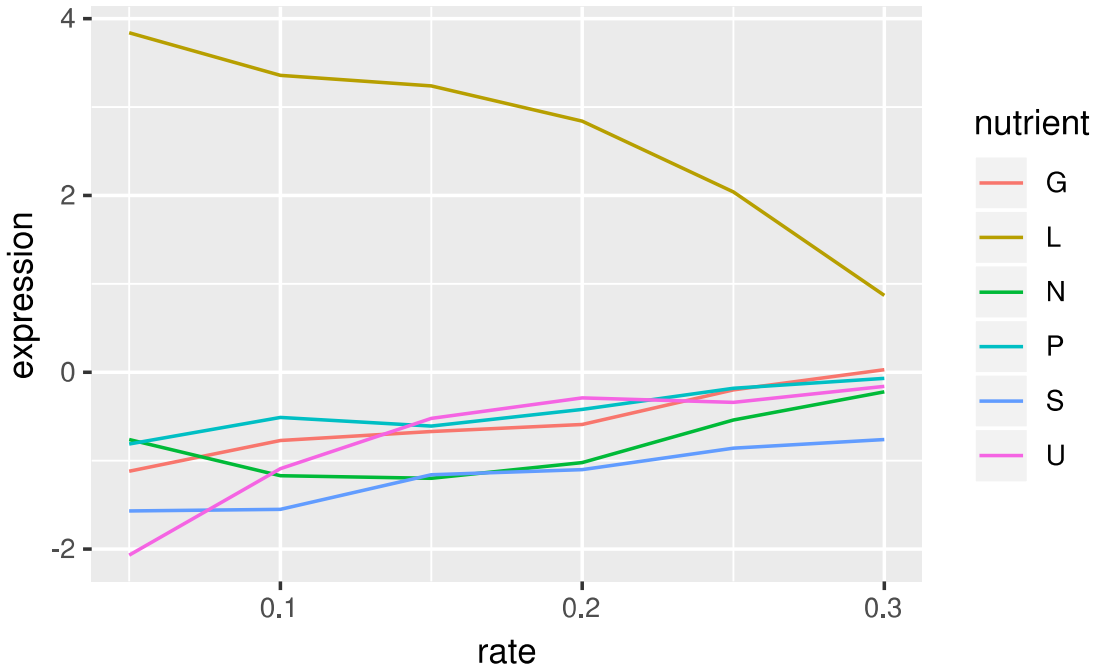
```
  brauer5 <- brauer5 %>%
    mutate_at(
      vars(gene_name:systematic_id),
      str_trim
    )
```

Visualizations using the tidy dataset

Exercise 6

```
  filtered <- brauer5 %>% filter(str_detect(gene_name,"LEU1"))

  ggplot(filtered) +
    geom_line(mapping = aes(x = rate, y = expression, color=nutrient))
```


4
2
0
-2
expression
0.1
0.2
0.3
rate
nutrient
G
L
N
P
S
U