

# Lab 7: Tidying your dataset

jiwon mok

2022-02-23

---

## Prune the dataset

### Exercise 1

```
brauer2 <- brauer %>% select(-GID, -YORF, -GWEIGHT)
```

## Enforcing one observation per row

### Exercise 2

```
brauer3 <- brauer2 %>% pivot_longer(  
  cols = G0.05:U0.3, names_to = "sample", values_to = "expression"  
)
```

## Enforcing one column per variable and one value per cell

### Exercise 3

```
brauer4 <- brauer3 %>% separate(  
  sample,  
  into = c("nutrient", "rate"), sep = 1  
)
```

### Exercise 4

```
brauer5 <- brauer4 %>% separate (  
  NAME,  
  into = c("gene_name", "biological_process", "molecular_function", "systematic_id", "number")  
  sep = "\\|\\|\\|\\|"  
)
```

### Exercise 5 -> white space

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

## Visualizations using the tidy dataset

### Exercise 6

```
brauer_tidy %>%  
  filter(gene_name == "LEU1") %>%  
  ggplot()+  
  geom_line(aes(x= rate, y=expression, color=nutrient, group = nutrient))
```

