Lecture 02 Demo

PSTAT 100: Summer Session A, 2025

Instructor: Ethan P. Marzban

First we load our necessary packages.

```
library(tidyverse)
library(reshape2)
```

Now, we read in the data. Remember that you will need to set the working directory first!

```
enrollments <- read.csv("data/enrollments.csv")
```

First Task: Extract Lower-Div Courses

We begin by extracting only the lower-division courses. There are a couple of different ways to do this; I'll do so using the following two-step procedure:

1) Create a vector that stores the names of the lower-div courses
2) Filter the `enrollments` dataframe to return only entries whose `Course` value is one of the lower-div courses

```
## Step 1: create a vector to store the names of the lower-div courses
lwr_div <- c("PSTAT 5A", "PSTAT 5LS", "PSTAT 5H", "PSTAT 8", "PSTAT 10")

## Step 2: Filter
lwr_div_courses <- enrollments %>% filter(Course %in% lwr_div)
```

Second Task: Melt

Now, we need to melt our `lwr_div_courses` dataframe. As a reminder:

- There are two colvars: `Course` and `Title`
- We want the remaining column headers to become values of a single `Quarter` variable
- We want the values to concactenate into a single `Enrollments` variable

Using the `melt()` function from the `reshape2` package:

```r
lwr_div_molten <- lwr_div_courses %>%
  melt(
    id.vars = c("Course", "Title"),
    variable.name = "Quarter",
    value.name = "Enrollment"
  )

lwr_div_molten %>% head()
```

```
     Course               Title Quarter Enrollment
1  PSTAT 5A UNDERSTANDING DATA     S25        167
2  PSTAT 5H         STATISTICS     S25         11
3 PSTAT 5LS STAT LIFE SCIENCES     S25        355
4   PSTAT 8 TRANS DS PROB STAT     S25         70
5  PSTAT 10 DATA SCIENCE PRINC     S25        235
6  PSTAT 5A UNDERSTANDING DATA     W25        222
```

Third Task: Generate a Plot

```r
lwr_div_molten %>% ggplot(aes(x = Quarter, y = Enrollment)) +
  geom_point(aes(colour = Course)) +
  geom_line(aes(colour = Course, group = Course)) +
  theme_minimal() +
  scale_x_discrete(limits = rev) +
  ggtitle("Lower Div Enrollments over Time")
```