

---

# Lecture 7: Data Wrangling with dplyr

## In-Class Activities

Answer the following questions in an RMarkdown document and discuss it with the people around you (no need to write the problem down, just give it a header (## Problem 1, etc.):

1. Using the `data_210_census` dataset in the `MSMU`, use `dplyr` to accomplish the following goals:
  - (a) Display just the observations which have no degree and a salary greater than 82,000 or a graduate degree and a salary greater than 140,000
  - (b) Arrange the previous example by their degree, then their gender, then their age (in descending order)
  - (c) Rename the gender column to “sex”
  - (d) Display all of the columns except for the “sex”
  - (e) Create a new column which calculates their monthly salary
  - (f) Create 3 new columns and display only 4 columns. They should display the age, how many years they are from retirement (65), how much they are saving for retirement ( $\text{salary} \times 0.07$ ), and their height in feet
  
2. Use the “house” dataset in the “openintro” library to answer the following questions:
  - (a) Keep all of the columns except for the last 4 (other, vac, del, res)
  - (b) Alter the column names to where they make sense (read the documentation!)
  - (c) Look at just the house “session”s which occurred after 1900
  - (d) Determine which “session”s has the biggest difference in members between the first and second political parties
  - (e) Determine how many times since 1900 the Democrats have been in control and how many times since 1900 the Republicans have been in control