

---

# Lecture 8: Grouped and Bivariate Analysis

## 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. For the following question you will be using the “mlb\_players\_18” dataset in the “open-intro” library. All of these parts should follow each other, but do them in individual pieces to make sure it does what you want.
  - (a) Remove the columns so the only ones that are selected are a) name, b) position, c) games, d) AB, e) H f) HR
  - (b) Alter the dataset so that only the players with more than 100 ABs are showing
  - (c) Create a new column that labels them as “Power” if they hit 20 or more HRs and “No Power” if they hit less than 20 HRs
  - (d) Create a new column that labels them either Outfield (ending with F), Catcher (C), or Infield (everything else) [Hint: ifelse() and grepl() might be beneficial]
  - (e) You can now save this dataframe to a new variable name and then do the following questions individually
  - (f) Determine the average number of Hits (H) that a player gets depending on their power level and position category (the categories you just made).
  - (g) Write a 6-Number summary (min, q1, q2, mean, q3, max) along with the number of values in each group for power-level and position category
  - (h) Surprise me with a (non-scatter plot) Bi-variate visualization of your choosing using ggplot()
2. Looking at the “state\_stats” in the “usdata” library, answer the following questions:
  - (a) Remove some columns so only the state, population totals (for both 2000 and 2010), the homeownership rate, poverty rate, income, murders, and unemployment are present
  - (b) Order the states from largest population change to smallest population change
  - (c) How many of them had a larger population in 2010? How many had 10,000 more people? 100,000 more people? 1,000,000 more people? Try to do all 3 in one command
  - (d) Visualize the homeownership rate and the poverty rate using only states with an income of greater than \$24,000