PAIA 730: Data Driven Management

Prof. Jesse Lecy

**Lab 03 – Combining Data Sets**

Turn in via Blackboard by start of class as PDF or webpage complete file

In this lab you will merge data in order to generate insights. The questions may pertain to any dataset in the Lahman package.

1. Which fielding position had the highest average pay in 2002? Right field (RF), center field (CF), or left field (LF)? Use the FieldingPost table (statistics over the full season), not the Fielding table (statistics organized by stints).
2. What is the average salary of a player for the Oakland A’s in 2002? The Boston Red Sox that same year?
3. Which city (birth city) has produced the most baseball hall-of-famers?
4. OPTIONAL: How many players went to college in Los Angeles, San Diego, or San Francisco? This one is tricky because the new Lahman package includes one row of data for each year at college in the CollegePlaying table, so a merge will create up to four rows of data for each player. Before counting the cities, you need to apply the unique() function to your data to remove duplicates.

Submit your result as a **PDF or HTML document** on BlackBoard. You will need to “knit” your .rmd file first.