PAIA 730: Data Driven Management

Prof. Jesse Lecy

**Lab 04 – Describing Data**

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

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

1. What is the average pay of each fielding position in 2002? Use the Fielding dataset.
2. Has the median batting average gone down over time? What about the top batting average each season? Use bat.ave = hits / at bats, **and only use data for individuals that have at least ten at bats.**# to plot your results

median.by.year <- tapply( X=bat.ave, INDEX=Batting$yearID, FUN=median, na.rm=T )

years <- as.numeric( names( median.by.year ) )

plot( years, bat.ave.by.year, type="b" )

1. Is there a relationship between batting average and pay in 2002? Create five groups for batting average with equal numbers of players in each group, and calculate the average salary for each group.
2. OPTIONAL: Is there a relationship between a team’s budget and the number of games they win in a season in 2002? Create three equal member groups for each variable and a table of the relationship.
3. OPTIONAL: What cities do the Atlanta Braves recruit from most?

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