**Goal:** build simple linear regressions and make some predictions using the real-world baseball game data. You will also have the chance to review some dplyr functions to clean the data and make some visualizations using ggplot. You do not need to fully understand rules of the baseball games in order to complete the lab.

**Notice:**

1. **geom\_boxjitter()**: combination of geom\_box() with geom\_jitter(). geom\_jitter() will automatically shits the points horizontally or vertically by some amount so that the records with the same value will not overlap with each other. Nothing special in terms of using the function. Just specify the arguments in ggplot() and then add geom\_jitter()
2. You will be asked to type the linear equations in latex. For those of you who are not familiar with latex, there is a code provided in the rmd file. Simply replace it with b0 and b1 with the values you guys obtain from the linear regression.
3. For exercise 10, please refer to the details on the side of the lab instruction to correctly get the statistics from the website.

* Runs scored & strikeouts: standard batting table
* Runs allowed & strikeouts allowed: standard pitching table

1. For exercise 11-14, you will use a categorical variable. Pay attention to the data type of the variable.

**f <- factor(c("a", "b", "c", "d"), levels = c("a", "b", "c", "d"))**

**fct\_relevel(f, "a", after = 4) = fct\_relevel(f, "d", after = 0)**