# Lab 1.1 – Computing the league-wide ERA and slugging percentage

In this lab, we will start to explore the effect of the introduction of the [designated hitter](https://en.wikipedia.org/wiki/Designated_hitter), by comparing the league-wide *earned run average* (ERA) and *slugging percentage* for each year in the modern era of baseball.

\*\*\*Aggregate before calculating ERA

\*\*\*should have 2 csv files as outputs

\*\*\* ERA should be 3-4 (smaller the better)

\*\*\* slugging should be .4 or .5 (bigger the better)

**Tasks.**

1. Using the Wikipedia page for the DH linked above, outline changes to the DH rule in each league.



1. Download the CSV version of the most recent Lahman database from <http://seanlahman.com/>
2. Create a data repository for these data and unzip the Lahman database into a data folder.
3. Create a new Python notebook in the root of your data repository called comparing\_league\_wide\_ERA.ipynb

**Comparing league-wide ERA – Part 1**

The [earned run average (ERA)](https://en.wikipedia.org/wiki/Earned_run_average) is computed using the following formula

where we multiply by 27 because there are 27 outs in a regular game (no extra innings due to a tie). Consider Pitching.csv file. This file contains, for each season/stint, pitching statistics for each pitcher. We will focus on the ERA, which measures the average number of runs allowed by each team’s pitchers over a 9-inning game, with a smaller number indicating better pitching + defense. Note that a smaller ERA indicates better pitching.

**Tasks.**

1. Compute the league-wide ERA for each league (AL & NL) for each year in the modern era (1946+). Note that you will need to group and aggregate the totals before computing the yearly league-wide ERAs.
2. Create a categorical variable that represents the various states of the DH rule (see **Task 1**).
3. Write the resulting table to a new CSV file in the data folder.

**Slugging percentage.**

Like ERA, the [Slugging percentage](https://en.wikipedia.org/wiki/Slugging_percentage) is an overall measure of batting effectiveness. Suppose that we also want to compare the leagues by this metric relative to the various states of the DH rule. Repeat steps like those outlined above, but this time compute the league-wide slugging percentage using the Batting.csv file. As before, the end-goal is a CSV file containing the results saved to the data folder.

A math equations and numbers

Description automatically generated with medium confidence

**Deliverables.** Submit completed copy of this document that includes

1. A screenshot of the head of each final table, and

|  |
| --- |
| League ERA with DH Rule |
| SLG |

1. The link to your data repository.

<https://github.com/ShoSato-047/DH_comparisons>