CSc 496, Homework #1: Correlation of offensive statistics to runs

Due date: Tuesday, September 10th, 2024. **No late assignments will be accepted.**

For the 1911, 1956, 1982, and 2023 seasons, develop one linear regression for each of the following MLB offensive statistics (both leagues together) to runs scored: hits, batting average, on-base percentage, slugging percentage, on-base percentage plus slugging percentage (OPS), and adjusted on-base percentage plus slugging percentage (OPS+). That is, for each linear regression, the dependent variable is runs scored, but the the independent variable is hits (first regression), batting average (second regression), etc. For the input data, use Baseball Reference (e.g., `https://www.baseball-reference.com/leagues/majors/2023.shtml`). You can manually grab data with the "share and export" link. Your program should output the correlation coefficient for each regression.

Make the program configurable so that any season could be used (in baseball reference format). Specifically, your program should take one input argument, the file, with no error checking needed. The output should be five lines, similar or identical to: `Correlation coefficient for R/G vs H: X`, where `X` is whatever your program finds.

In addition, please plot each of the five correlations (including the points and the regression line). You do not need to put team name labels on them, but please put them all in one file. (An AI tool will know how to do that.)

You should submit your assignment on lectura using the `turnin` command; for this program, use the assignment name `csc496-f24-hw1`. Call the file `hw1.py`.