BIS/Retrosheet BIP user notes

The data files are either .csv or .xlsx, and should be converted to the form suitable for statistical analysis in the statistics package of your choice (.dta for Stata).

The analysis script is a .do file. But it is annotated sufficiently to enable translation to other programming languages.

The analyses use both unstandardized and season-standardized versions of the variables. The former are used for plotting in order to make the relationships intuitively graspable. The latter are used for regression estimates of the impact of the pitching, fielding, and BIP variables on runs allowed per season; doing so models how relevant metrics influence *season-specific* variance (as opposed to across-season variance, which is not a quantity of much relevance). It also is an alternative to multi-level modeling or time-period regressions for removing extraneous variability associated with shifting game conditions over time. See the discussion in Schell, M. J., Baseball's All-Time Best Hitters: How Statistics Can Level the Playing Field (Princeton University Press 2013); Schell, M. J., Baseball's All-Time Best Sluggers: Adjusted Batting Performance from Strikeouts to Home Runs (Princeton University Press 2016).