

How Has Major League Baseball Changed Over Time?

Exploring the Lahman's Baseball Database

Frank Driscoll, Cam Garfield, Cole Guerin, Hannah Hartnett
Statistical Graphics, Colby College

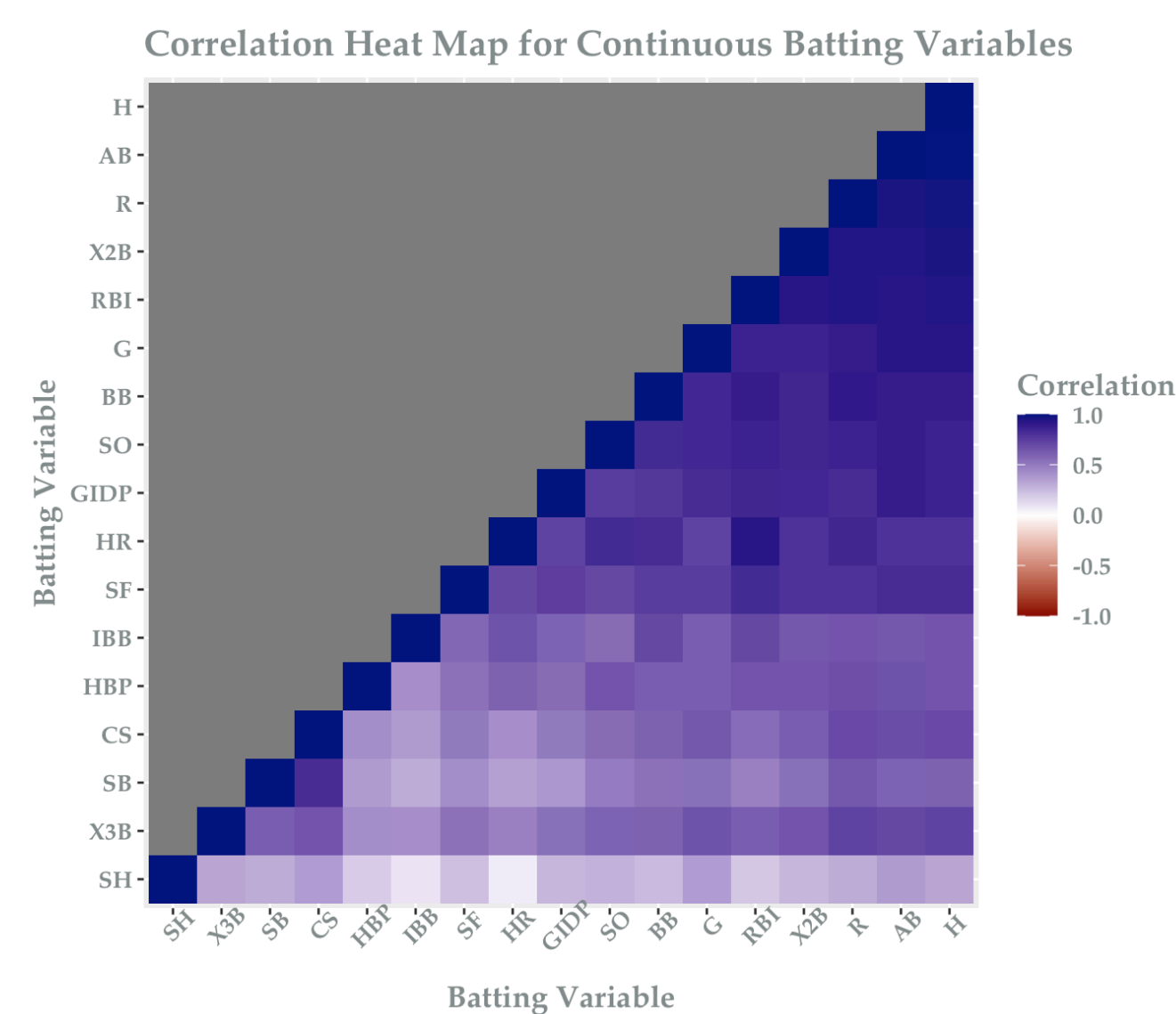
Introduction

The Major League Baseball association has served as America's pastime's highest level of competition for almost 120 years. We dive into the in-depth historical database that chronicles the sport in order to examine how hitting has changed over time.

The Database

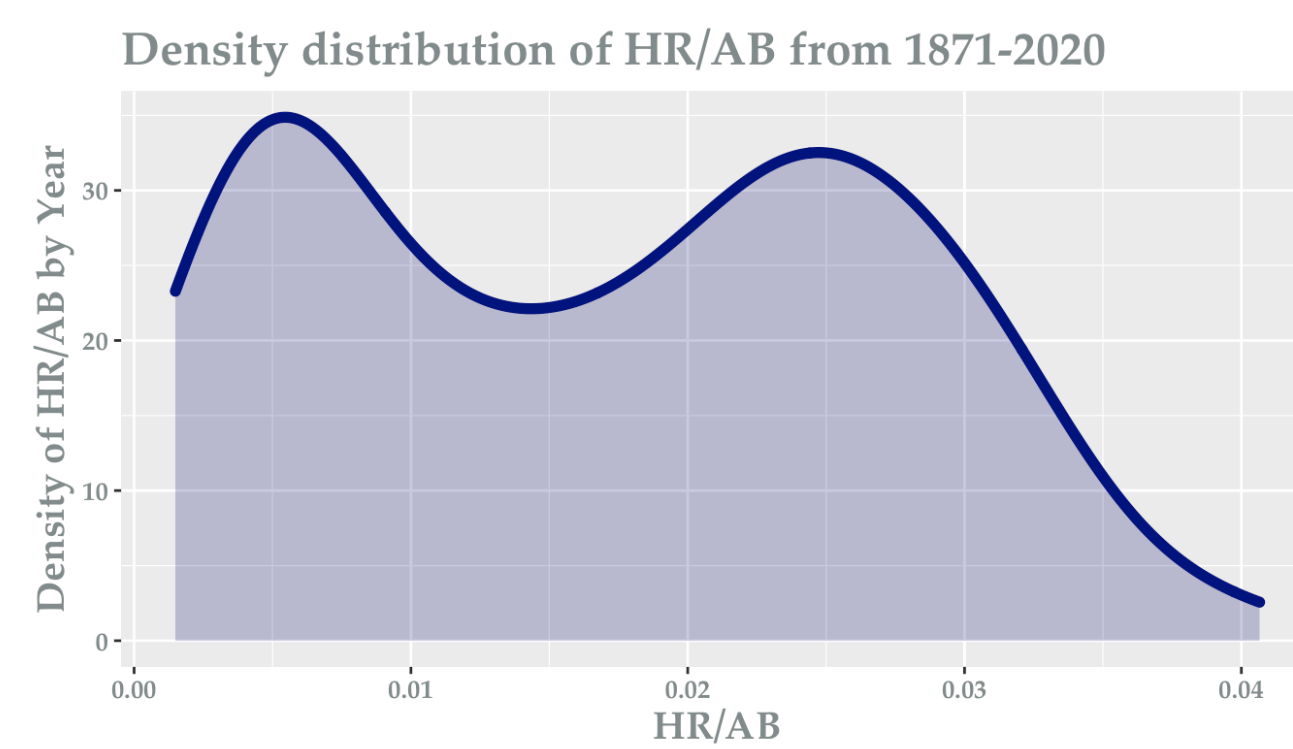
Lahman's Baseball Database is part of the Comprehensive R Archive Network. The updated version of the database contains complete batting and pitching statistics from 1871 to 2020, plus fielding statistics, standings, team stats, managerial records, post-season data, and more.

Looking at Continuous Hit Data



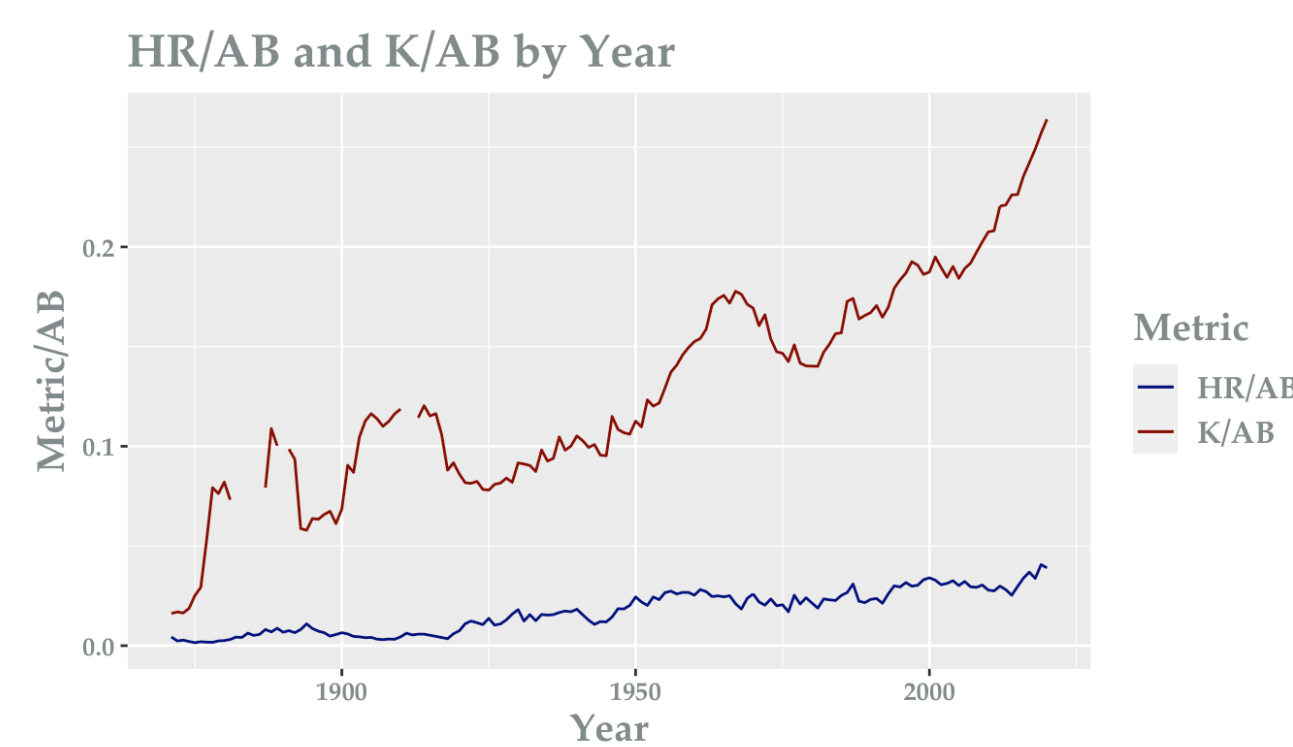
We began with a high-level analysis of correlations between variables. Unsurprisingly, the number of games, hits, at bats, runs, and runs batted in are very highly positively correlated. There are no negatively correlated pairs of variables, though sacrifice hits have essentially no association with either home runs or intentional walks.

Exploring HRs per AB



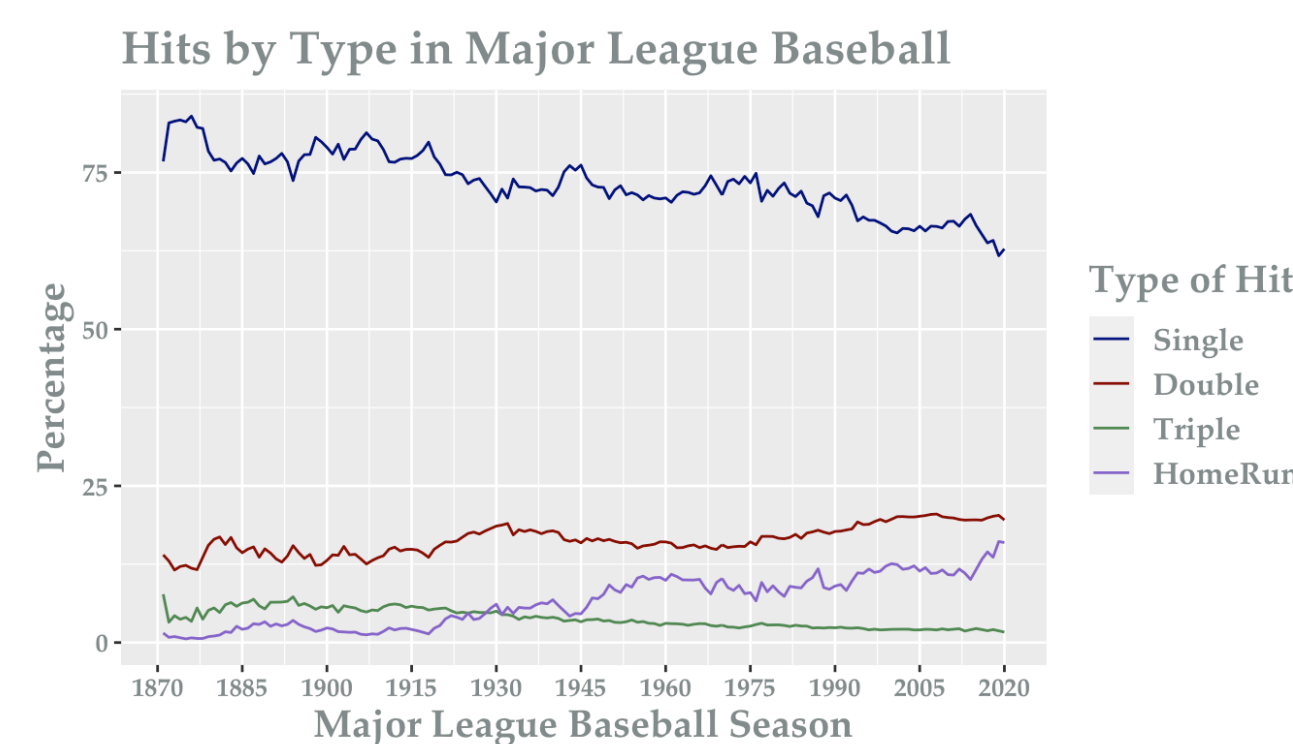
The distribution of HR/AB by year has a two peaked distribution, with the first at around 0.05– a high concentration of low HR/AB numbers and the second at around 0.025 representing a concentration of higher HR/AB values. There is a decent right skew representing the few players with higher HR/AB numbers.

What influences number of home runs hit?



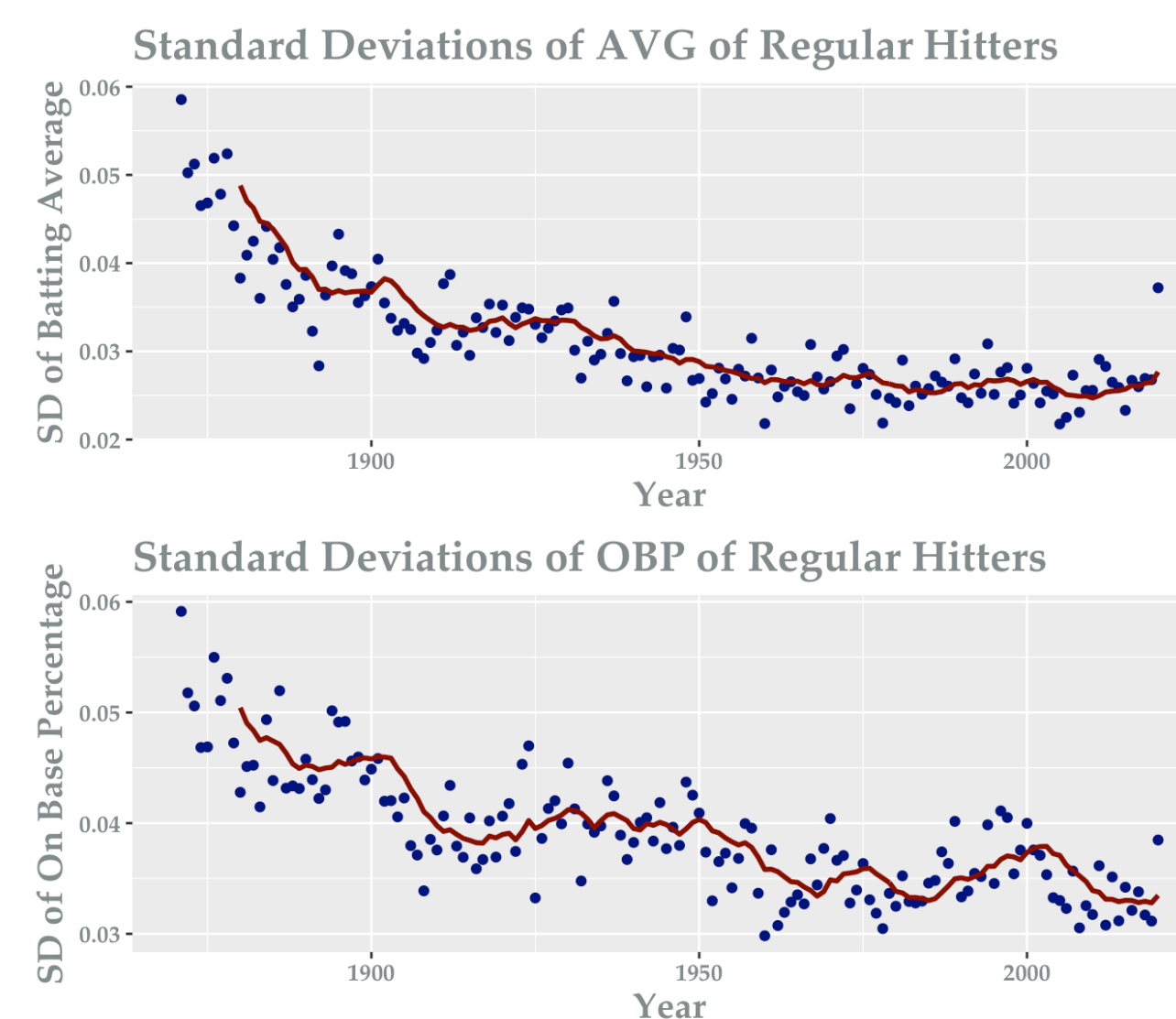
Over the years, both home runs and strikeouts have increased, indicating a transition from small ball baseball where pitchers throw to contact, to high power baseball where stronger hitters try to hit home runs and pitchers with increased velocity go for strikeouts.

Exploring Hit Type



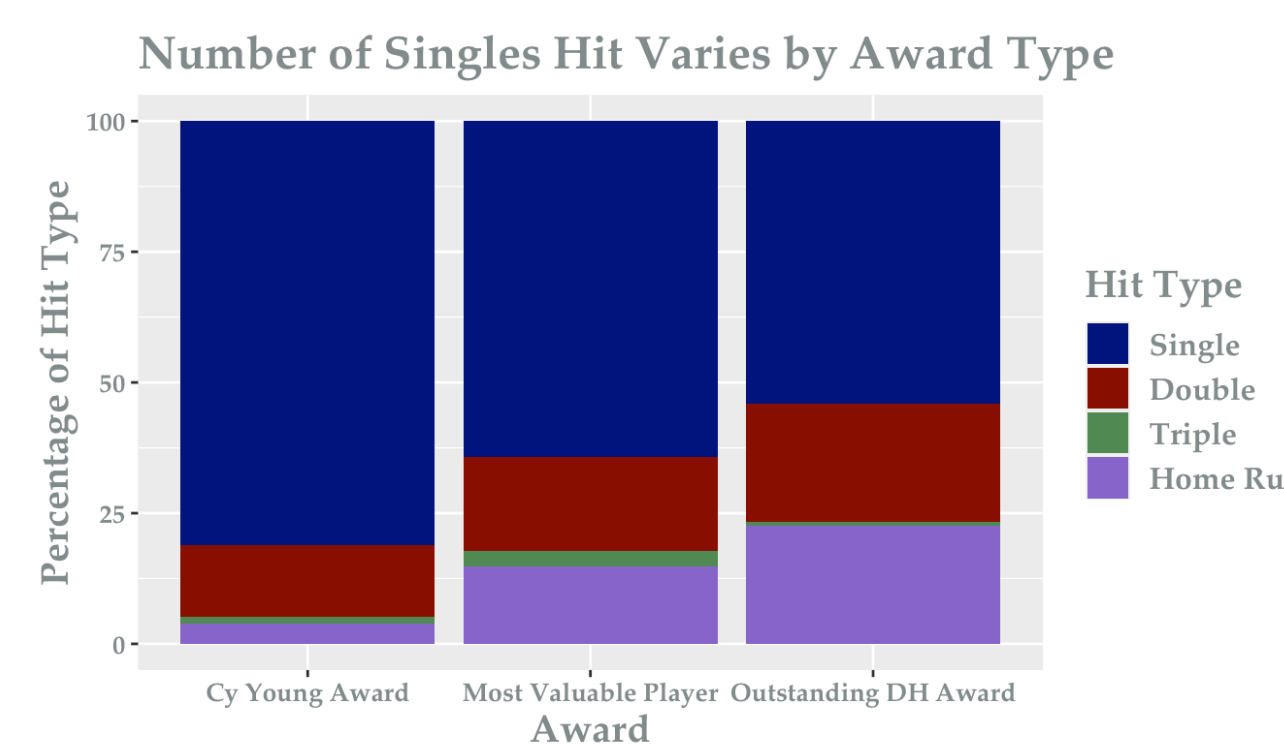
Similarly, we see an increase in homeruns and doubles hit over time, corresponding with a decrease in singles and triples. There is a particularly large increase in home runs and decrease in singles in the last ten years.

Batting Average getting Lower?



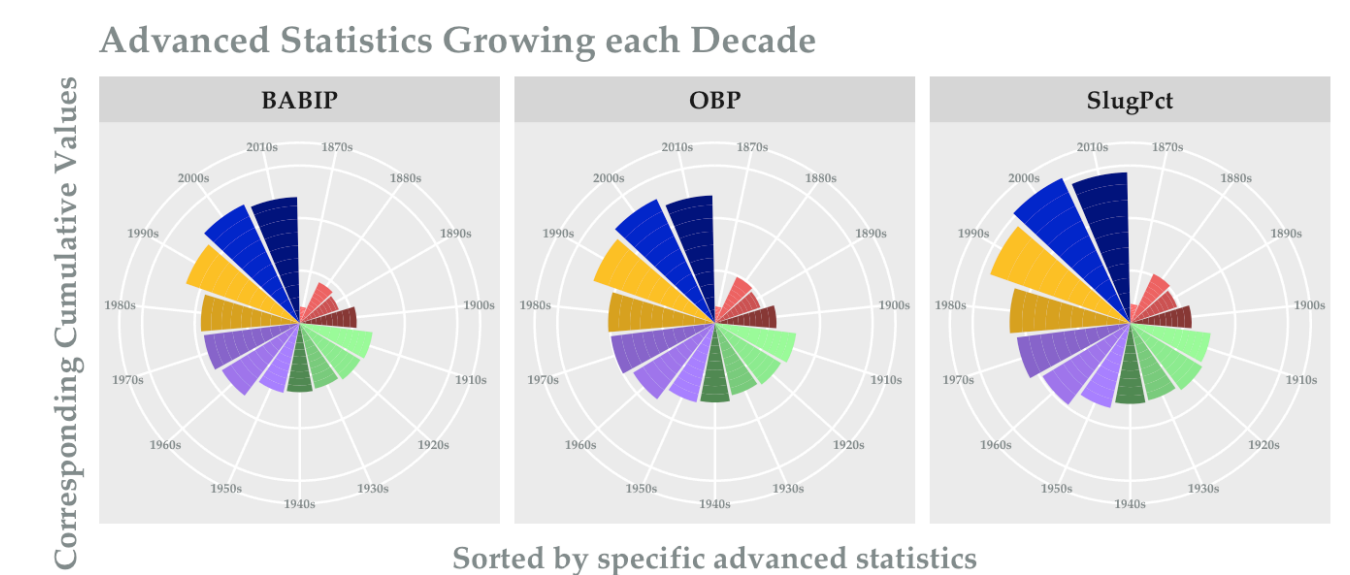
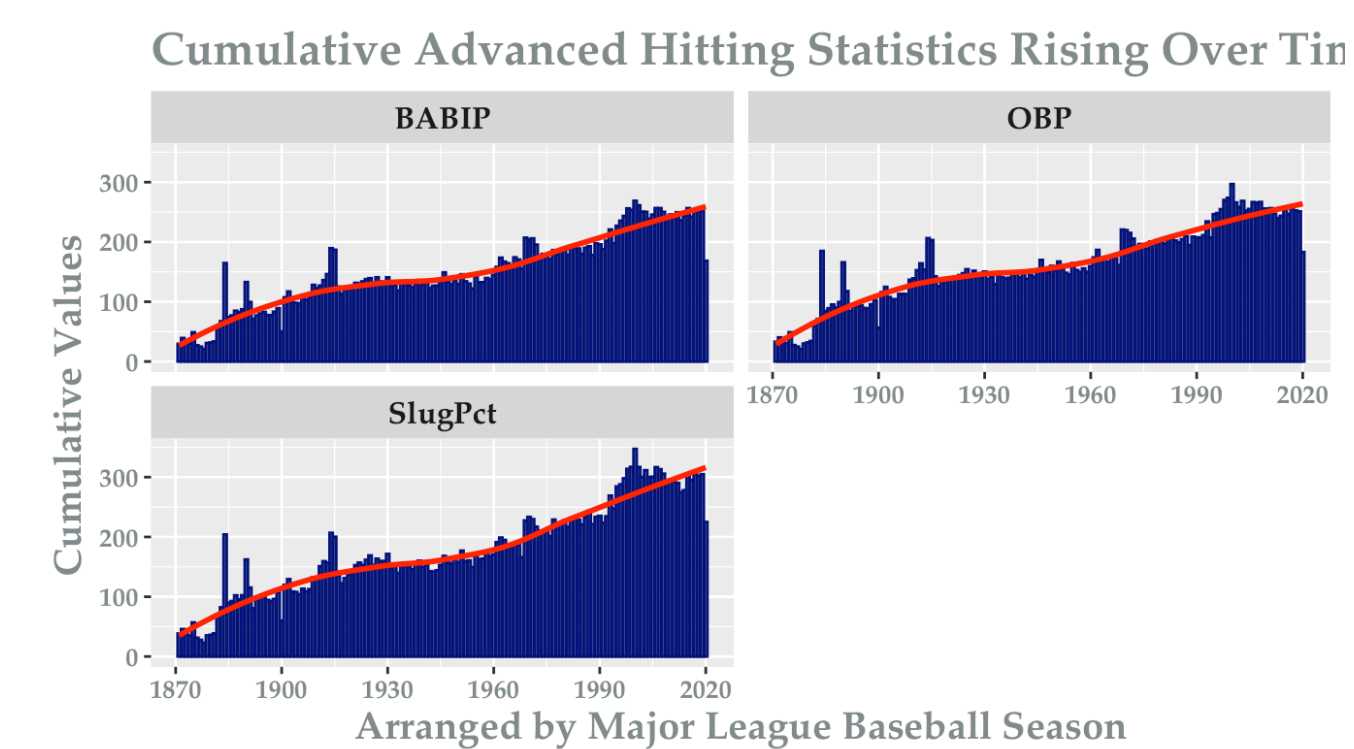
A recent study theorized that the disappearance of the 0.400 batting average in baseball is caused by the steady decrease of batting averages among “regular” players overtime, perhaps related to the earlier claim of pitchers with high velocity attempting more strikeouts. Great hitters are moving closer to the average because of this smaller variation, so less baseball athletes have been able to reach the 0.400 batting average mark. From the top plot, we can see that the standard deviation of batting average for so called “regular” players has decreased over time, but it has been very consistent in the last 60 years. A different batting statistic, that has seen more of a drop recently, is the standard deviation for on base percentage which is shown in bottom plot.

Awards and Hitting Relationship



Unsurprisingly, the players who have received awards specific to hitting, like Outstanding Designated Hitter, have much higher home run totals and fewer singles than nonspecific awards, like MVP. Those players who have received awards unrelated to hitting, like the Cy Young Award for pitching, have the fewest collective home runs and the most singles.

Hitting Statistics Rising



Throughout the years, baseball's advanced statistics, including batting average on balls in play (BABIP), on-base percentage (OBP) and slugging percentage (SlugPct), a measure of total bases gained by a batter per at bat, have become increasingly important in understanding a player's contribution and value to the game. As such, these players and outcomes have become more highly valued within the game.

Sources

Lahman's Baseball Database

<http://www.seanlahman.com/baseball-archive/statistics/>

Stephen Jay Gould. Full House: The Spread Of Excellence From Plato To Darwin (New York: Harmony Books, 1996).