Major League Baseball Hall of Fame Prediction Model

Christopher I. Moussa

[redacted]

University of California, Riverside

Data Science

Abstract

Baseball lies not only in its timeless traditions but also continuously evolves and improves. The sport's Hall of Fame (HoF), a shrine of its most excellent players, managers, and icons in the sport, has remained a cherished, yet complex, tradition. This research project seeks to revolutionize and modernize the evaluation process by introducing a data-driven approach that centralizes four advanced statistical metrics: Offensive Wins Above Replacement (oWAR), Defensive Wins Above Replacement (dWAR), Weighted On-Base Average (w/OBA), and Weighted Runs Created Plus (wRC+). This study challenges conventional notions of Hall of Fame candidacy that have for years relied on outdated and flawed statistics. Through statistical analysis, I look to uncover intriguing insights based on data aggregation of the current body of Hall of Fame position players, build a predictive model for current Hall of Fame candidates, and analyze past elections of rejected Hall of Fame candidates. This research aims to represent a progressive step toward more transparent, inclusive, and data-informed Hall of Fame selection processes for future candidates. By embracing modern statistical analysis and an open discourse about position-specific standards, this research aims to preserve the Hall of Fame's status as a shrine to excellence while at the same time introducing new ways to look at contemporary players and make decisive and objective comparisons to those considered the greatest to ever walk on the field. As the complexities of Hall of Fame induction are navigated, this research hopes to serve as a testament to the enduring spirit of baseball—an endeavor where both numbers and narratives converge to honor those who have left an unforgettable mark on the sport.

Introduction

Sports are one of the great pastimes in our lives, providing us with year-round exciting, gripping entertainment. Among these, baseball stands as an enduring fixture in American culture, intertwining its historical significance with present-day relevance. The sport encapsulates complexities that blend into a beautiful yet intricate tapestry, decorated with moments that resonate through time. Joe Dimaggio's 56-game hitting streak in 1941, Willie Mays' over-the-shoulder catch in Game 1 of the 1954 World Series, and Hank Aaron's 715th career home run in 1974 are etched into the records of baseball history.

At the heart of these moments lies the essence of baseball greatness and legacy - a narrative that extends beyond the confines of the playing field to captivate the hearts of fans. It is this collective investment in the players, teams, and the sport itself that propels baseball into the realm of cultural phenomenon. In recognition of the unparalleled contributions of exceptional players, the Hall of Fame (HoF) was inaugurated in 1936. Its walls enshrine greatness of players, managers, and sports icons, honoring those who have left an indelible mark on its history. The first HoF class: Ty Cobb, Walter Johnson, Christy Mathewson, Babe Ruth, and Honus Wagner pioneered the standard for which future players would forever be measured, and it has grown to house over two hundred of the greatest players to ever step on the field.

However, the pathway to the Hall of Fame is not without its imperfections. The selection process, primarily entrusted to the Baseball Writers' Association of America (BBWAA), hinges on a multidimensional assessment of a candidate's "record, playing ability, integrity, sportsmanship, character, and contributions to the team(s) on which the player played" (BBWAA Election Rules, n.d.). This comprehensive evaluation attempts to capture the player's holistic

impact both on and off the field. Yet, the subjective interpretation of these criteria introduces a potential bias that could unintentionally alter a candidate's fate (Mills & Salaga, 2011).

In this context, the motivation for this project emerges: an exploration of an objective framework for evaluating a candidate's playing ability - the cornerstone of their Hall of Fame eligibility. This research constructs a predictive model that consolidates a candidate's performance into a single numerical value for equitable and objective comparison. By doing so, it seeks to address instances where incomplete or skewed perspectives may sway the trajectory of a player's enshrinement. Through a blend of statistical analysis, historical context, and a commitment to transparency, this project provides a more robust and unbiased foundation for Hall of Fame considerations.

Project Overview

This project is broken up into three components: 1) formulation of a quantifiable measurement for a Hall of Fame player using four advanced offensive and defensive statistical measurements, 2) data collection and organization of current Hall of Fame players and current and past Hall of Fame candidates, and 3) analytic methods for both Hall of Fame players and current and past Hall of Fame candidates. Each project part builds off the progress made in the previous part and is used to summarize a player’s baseball career while also providing an easy comparison to current players and HoF candidates.

The formulation of a quantifiable measurement will break down the justification for choosing each advanced offensive and defensive statistical measurement and how it relates to a player’s career-long success, their relation to analyzing player performance over the sport’s one hundred-plus year existence, and their relevance to how the numerical value is calculated.

The data collection component will discuss how data was gathered, organized, and used to perform the statistical analysis of both HoF players and HoF candidates. It will explain why the data was organized in the way that it was and how data visualization helped expose both patterns and holes in the Hall of Fame's current structure.

The last component, analysis performed on current and previous HoF candidates, will use the same methodology for numerical calculation for current hall of famers for comparison and prediction. For current HoF candidates, use of the numerical grade will help make predictions on a candidate's chances of being successfully elected to the Hall of Fame. For previous candidates, the grade calculator will be used to highlight where voters might have missed the mark in terms of leaving a particular candidate off of a given ballot. While a player's career statistics are not the only prerequisite for their election, it often is the predominant factor referenced when discussing the player's case for election. This component will include examples of candidates who might be *justifiably* excluded from election to the Hall of Fame and of candidates who might be *unjustifiably* excluded from election.

An important clarification must be made about the dataset used. This project focuses solely on Hall of Fame candidates and Hall of Fame **positional players**. A **positional player** is defined as a player who primarily plays any position besides a pitcher. The reason for this choice is the statistical measurements used to evaluate a player are vastly different between positional players and pitchers. Different statistical emphases on player, team and league value are used for these two positional categories, and it is unhelpful to create a single grade calculator that includes *both* sets of emphases for positional players and pitchers. For example, a positional player's contributions are primarily summed up by their *offensive* contributions, while a pitcher's contributions are summed up by their *defensive* contributions. Because of this stark distinction,

this project excludes data collection, grade formulation, and analysis of Hall of Fame pitchers and Hall of Fame pitcher candidates. However, the analytic methods employed in this project can be applied to create a similar analysis for pitchers exclusively.

Significance of Project

The Hall of Fame holds a revered position in the world of sports, honoring exceptional players who have made permanent marks on the game, and induction into the HoF is a recognition of an athlete's enduring impact, talent, and contributions to the sport. The process of selecting Hall of Fame candidates, however, is often a subject of debate and speculation, with differing opinions on the criteria that should be prioritized. Because of this debate, this project looks to refine this selection process and contribute to its evolution; it is important for a number of different reasons.

Unbiased Evaluation

Traditional methods of evaluating Hall of Fame candidates have often relied on popular perceptions, traditional (and flawed) statistics, and subjective opinions of voters. These set a dangerous precedence to create incomplete, and sometimes false, narratives (Parsons & Stern, 2012). The proposed prediction model in this project introduces an evidence-based and unbiased approach that utilizes a number of advanced metrics. By leveraging objective and normalized data, this model provides an unbiased and quantifiable assessment of how candidates measure up against Hall of Famers of all eras.

Incorporating Advanced Metrics

The advent of advanced statistics in baseball over the last twenty years has revolutionized player evaluation, offering insights into performance that go beyond traditional metrics. Offensive and Defensive Wins Above Replacement (oWAR and dWAR, respectively), Weighted

Runs Created Plus (wRC+), Weighted On Base Average (w/OBA), and others provide a comprehensive understanding of a player's contributions on both offense and defense. This project leverages these advanced metrics to develop a comprehensive evaluation system that captures the multifaceted nature of a player's impact on the game, both individually and in relation to their peers across the league and across different eras. This project can continue to be expanded in the future by introducing additional advanced statistics to the numerical grade calculation as well as adjusting the weight that each statistic has in the calculation.

Enhancing Transparency

The prediction model not only provides an evaluation framework but also enhances the transparency of the selection process. Clearly defining the parameters and factors used in the model demystifies a lot of the concerns about the ambiguity of the selection process and provides a clearer understanding of the criteria driving the decisions. This increased transparency fosters productive and informed discussions about the merits of individual players and the evolving nature of the sport.

Contributing to the Analytics Landscape

Analytics across the sports landscape has gained prominence across various disciplines, from player performance optimization, team success, fan engagement strategies, and more. The investment in advanced technological methods and studies in order to predict player success and evaluate past performances has grown significantly, particularly in the baseball industry (Sun et al., 2023). This project contributes to the wider landscape of sports analytics by showcasing the application of advanced statistical techniques in a real-world context. The methodology developed here can potentially serve as a template for evaluating candidates in other sports or even guide decision-making in player development and team management.

Literature

Previous Research

In the realm of baseball analytics and the evaluation of Hall of Fame candidates, several notable works have explored the shift from traditional statistics to advanced analytics over the past twenty-to-thirty years, the challenges faced by the organizations and media, and the nuances of player evaluation. The following peer-reviewed journals and books have provided valuable insights into these topics, shaped the landscape of modern player assessment, and influenced the approach taken in this research.

"Data analytics effects in major league baseball" by Ramy Elitzur

This journal describes the "Moneyball" revolution that baseball organizations across the country invested in in the last twenty to thirty years along with the benefits reaped as a result of the investment. Organizations who spent resources on building a robust analytics department sought success in constructing their rosters, developing younger players, and acquiring players via free agency and trades to build successful and competitive rosters. Not only do analytics provide organizations with an insight to evaluate current player performance and their future trajectories as players, but they allow for detailed analysis upon completion of their career, which is where this project focuses.

"Smart Baseball" by Keith Law

Keith Law's *Smart Baseball* delves into the revolutionary impact of advanced analytics on player evaluation, emphasizing their superiority over traditional statistics and how they do a better job describing a player's impact on any given game. Law dismantles the flaws inherent in outdated metrics like Batting Average, Runs Batted In, and Fielding Percentage, and introduces a new era of data-driven analysis that has transcended the boundaries of the game. The book also

examines the intricacies of Hall of Fame voting and highlights the emergence of data science divisions within baseball organizations. This work serves as a foundational resource that underscores the importance of adopting modern statistical methodologies to accurately assess player performance.

"Ahead of the Curve" by Brian Kenny

Ahead of the Curve writes about a number of anecdotal narratives and gives critical insights into the challenges confronted by baseball organizations, media, and Hall of Fame voters when assessing players using traditional statistics. The book, like *Smart Baseball*, advocates for the replacement of outdated metrics with advanced analytics, particularly challenging misconceptions surrounding end-of-year awards and Hall of Fame voting. Brian Kenny's exploration of data science's role in informed organizational decision-making also resonates with the growing influence of analytics in modern baseball operations.

"The New Bill James Historical Baseball Abstract" by Bill James

Historical Baseball Abstract is a monumental work that combines a vivid narrative of American baseball history with intricate statistical breakdowns of players, teams, and seasons. Bill James' application of stats-based methods (like Run Shares) to rank the top 100 players at each position has significantly shaped the way players' legacies are evaluated. The comprehensive approach to assessing players provides inspiration for the present project, which seeks to leverage advanced metrics to create a predictive model for Hall of Fame candidacy. Like this project, James explains the importance of using quantifiable measurements to create an unbiased, objective look at player evaluation.

"A Fan's Guide to Baseball Analytics" by Anthony Castrovince

Anthony Castrovince introduces readers to the foundational principles of sabermetrics, which is an essential aspect of modern player evaluation. The book serves as an introductory course on advanced statistics, bridging the gap between traditional metrics and modern analytics. By introducing readers to a range of advanced statistical tools used to measure player performance, Castrovince's work aligns with the goal of this project to create an evidence-based predictive model for Hall of Fame candidacy.

While the works I've mentioned above have contributed significantly to the discourse surrounding player evaluation and the transformation of baseball analytics, this research differentiates itself by focusing explicitly on the creation of a predictive model for Hall of Fame candidacy. By synthesizing advanced offensive and defensive statistics, this project seeks to offer an objective and quantitative assessment of players' contributions. This distinctive approach builds upon the foundations laid by these influential texts, contributing to the ongoing evolution of player assessment in the context of the Hall of Fame. It is important to recognize pioneers in the sabermetric revolution like Bill James, Keith Law, Brian Kenny, and others who have recognized why using advanced statistical methods and techniques to evaluate players leads to greater on-field success. This project is an homage to the work they have done.

Findings and Unanswered Questions

The existing body of literature underscores the limitations of relying on traditional statistics to evaluate a player's career. It is widely acknowledged that a shift towards more advanced analytics is essential to achieve a comprehensive and unbiased assessment for a player, and has even been used to retroactively analyze players from older generations. Additionally, a

consensus has emerged among experts in the field that the incorporation of memories, anecdotes, and outdated metrics can hinder the objective evaluation of a player's candidacy.

However, while the literature has pinpointed the shortcomings of legacy metrics, it falls short in providing a comprehensive blueprint for quantifying HoF candidates' careers. A notable gap in the discourse is the lack of specific guidelines on which advanced statistics to employ and how to structure the analysis to inform HoF voting decisions on a consistent basis. Several critical questions emerge as a result of this research gap:

- How many "traditional" statistics, previously used to assess HoF eligibility, should be retained in the evaluation of new candidates, if any at all? Is there a threshold that amounts to their significance in the modern context?
- When evaluating a player's case for HoF induction within a specific position, which advanced statistics should be prioritized? Do certain metrics hold greater (or less) relevance for particular positions? How should the comparisons be made between players of varying positions, such as catchers and shortstops?
- How should the statistics used for the calculation be divided between offensive and defensive metrics when considering position players? Can an agreement be made on the appropriate weights to apply in each category in determining a player's overall candidacy?

Addressing these unanswered questions is important in bridging the divide between acknowledging the shortcomings of traditional statistics, which were previously used to determine a Hall of Fame candidate’s eligibility, and implementing a comprehensive, data-driven framework for evaluating new HoF candidates. This research project endeavors to address these

gaps by proposing a prediction model that leverages a combination of advanced offensive and defensive metrics to establish a more informed and transparent approach to Hall of Fame considerations. The statistics used in this framework build off of the traditional metrics that were used in years past in order to keep them relevant when comparing new HoF candidates to existing hall of famers.

By investigating these inquiries and proposing a quantitative model that integrates advanced statistics, this research aims to contribute to the evolving landscape of player assessment in the realm of the Hall of Fame candidacy. In doing so, it aspires to provide a more equitable and evidence-based foundation for evaluating the contributions of players and their rightful places in baseball history.

Methodology

Approach

The approach for creating the quantifiable measurement for a given Hall of Famer and Hall of Fame candidate is straightforward. The grade *G* for a player is a sum of products, where each product is career statistic multiplied by a decimal weight that represents its importance in relation to the overall grade. The following equation describes how this project's Hall of Fame grade *G* is calculated:

$$G = (oWAR \times 0.4) + (dWAR \times 0.3) + (w/OBA \times 0.15) + (wRC + \times .15)$$

The result, *G*, is based on four key statistics: *Offensive Wins Above Replacement* (oWAR), *Defensive Wins Above Replacement* (dWAR), career average *Weighted On-Base Average* (w/OBA), and career average *Weighted Runs Created Plus* (wRC+). Each of these metrics provides unique insights into a player's offensive and defensive contributions, culminating in a comprehensive, quantifiable, and comparable assessment of their career. The following sections

describe each statistic in more detail, how it builds off of traditional statistics, and why it is a good statistical component of a player's career.

oWAR

Offensive Wins Above Replacement (oWAR) is a comprehensive measure of a position player's offensive contributions. Wins Above Replacement (WAR) fundamentally quantifies a player's value compared to a hypothetical "replacement" player, accounting for multiple offensive facets of the game. In the context of oWAR, the emphasis is solely on a player's offensive impact, making it a valuable tool for evaluating batting performance. The calculation of oWAR is based on constructing an average value derived from various player performances across the league. This average forms the baseline for comparison - hence, the name, "replacement player." The oWAR of a player is determined by contrasting their offensive contributions to this league-wide average and calculating the resulting surplus. The higher this surplus, the more above-average a player is compared to the rest of the league. When summed over a player's entire career, it becomes measurable how much more (or less) valuable a player is compared to his counterparts.

oWAR extends beyond traditional offensive metrics by including a range of factors that contribute to run production. This holistic approach accommodates the comprehensive evaluation of a player's offensive capabilities beyond mere Batting Average, home run counts, On-Base Percentage, and Slugging Percentage, factors that were (and still are) frequently used to analyze and compare player performance. oWAR is a more comprehensive gauge than a number of traditional offensive statistics, providing a reliable measure of a player's excellence at the plate and on the basepaths.

One of the most important factors of oWAR in relation to this project is it allows for a standardized comparison of players across eras and positions, offering a uniform metric that adapts to the evolving statistical landscape of baseball and how elite players are measured. It tackles multiple obstacles: 1) its inclusion in the calculation of the overall grade provides a solid foundation for analyzing a player's offensive contributions, and 2) it eliminates the potential bias associated with defensive performance. Because of its importance in player measurement, it receives the highest weight in this project's formula with a weight of 0.40.

dWAR

Defensive Wins Above Replacement (dWAR) is a specialized metric focused solely on quantifying a player's defensive contributions, considering both defensive statistics while also adjusting for the position played. In contrast to oWAR, dWAR zeros in on the player's fielding performance, providing insights into their ability to prevent runs through defensive excellence. But, like oWAR, a vital component of this calculation is the establishment of a "replacement-level" player for defense, which corresponds to a league average. dWAR gauges how much better a player's defensive contributions are compared to a hypothetical replacement player. This statistic also creates a single value that encompasses a number of different defensive measurements which enables a standardized comparison across both different positions and different eras.

Incorporating dWAR into the evaluation process allows for a comprehensive understanding of a player's defensive value, beyond the scope of traditional fielding statistics like Fielding Percentage. This metric also sheds light on a player's ability to impact the game through their defensive prowess, a dimension often underrepresented in traditional statistical analysis.

Separating offensive and defensive metrics by utilizing both oWAR and dWAR enables a granular analysis of a player's overall impact. This approach serves to underscore the diverse skill sets required across different positions in baseball.

The weight chosen for dWAR in the calculation for a player is less than the weight chosen for oWAR. This project uses a weight of 0.30 for a player's career total dWAR for a number of reasons. It is widely acknowledged that the voting body, entrusted with the responsibility of Hall of Fame induction, tends to emphasize offensive contributions to a greater extent than defensive contributions, particularly for positional players. This recognition prompted a deliberate choice to assign different weights to oWAR and dWAR in order to reflect this historical voting tendency and preference. By assigning a slightly lower weight to dWAR, this methodology recognizes the overarching sentiment that offensive performance often generates greater attention when determining a player's Hall of Fame eligibility. This approach seeks to align the calculation of a player's grade *G* with the broader context of both historical and current Hall of Fame voting trends while still capturing the pivotal role that defensive excellence plays in the overall evaluation.

w/OBA

Weighted On-Base Average (w/OBA) is an evolved version of the traditional On-Base Percentage (OBP) statistic, introducing a dynamic evaluation of a player's ability to reach base and contribute to their team's offensive performance. Unlike OBP, which treats all methods of reaching base equally, w/OBA assigns different values to each offensive event based on its potential to influence the scoring of runs. The calculation of w/OBA involves assigning distinct weights to various offensive events based on their projected impact on runs scored. For instance, a home run is weighted more heavily than a single, which reflects its capacity to generate a

greater number of runs. As a specific example, according to Major League Baseball's official website, in 2014, a home run was worth 2.101 times on base, while a walk was worth less than a base, at 0.69 (Weighted on-base average (WOBA): Glossary, n.d.). These weighted values are then aggregated to produce the player's w/OBA, encapsulating their ability to create scoring opportunities. The following formula describes how w/OBA is calculated for a given player, taken from Major League Baseball's official glossary:

$$w/OBA = \frac{(BB \times BB_{factor}) + (HBP \times HBP_{factor}) + (1B \times 1B_{factor}) + (2B \times 2B_{factor}) + (3B \times 3B_{factor})(HR \times HR_{factor})}{AB + BB_{unintentional} + SF + HBP}$$

Using a player's career average w/OBA introduces a higher level of granularity to the assessment of a player's offensive capabilities, in particular to their contributions for their respective team. Because reaching base is a diverse event, w/OBA is more capable of accurately portraying a player's ability to contribute to producing runs.

An important note about this statistic is its ability to differentiate good players from great players *from season to season*. The values for each offensive event are adjusted seasonally, which accounts for the evolving landscape of baseball and the changing significance of various offensive events over time. For example, in a season where home runs are a frequent offensive event, the weight associated with hitting a home run is less than in a season where home runs are few and far between.

The utilization of w/OBA in the calculation of each player's grade augments the analysis by emphasizing the distinct impact of each offensive event on a team's overall performance. In this project, the weight associated with a player's career average w/OBA stands at 0.15. It is a useful comprehensive parameter for evaluating a player's offensive excellence and thus is

included in our analysis of current Hall of Famers and in our predictions of Hall of Fame candidates. w/OBA enriches the project's ability to provide an objective and comprehensive assessment of a player's career (Castrovince, 2020).

wRC+

Weighted Runs Created Plus (wRC+), similar to w/OBA, is a comprehensive measure of a player's offensive contributions, quantifying their ability to generate runs through a combination of on-base skills and power hitting. The statistic takes into account both a player's ability to reach base as well as their proficiency in producing extra-base hits, which offers a holistic understanding of their offensive impact.

$$wRC = \left(\frac{(wRAA\ per\ PA + league\ runs\ per\ PA) + (league\ runs\ per\ PA - ballpark\ factor \times league\ runs\ per\ PA)}{league\ wRC\ per\ plate\ appearance}\right) \times 100$$

The formula above, taken from Major League Baseball's official glossary, describes the process for calculating a player's wRC+. It encapsulates a player's ability to reach base and produce extra-base hits divided by their total opportunities at the plate. What sets wRC+ apart from many other traditional offensive statistics, however, is its *normalization*, which ensures that the league average is always established at 100. This feature eliminates the influence of external factors, such as the ballpark played in or the historical era, enabling direct comparisons across players and time periods. A ballpark that is notoriously difficult to hit in, like Oracle Park in San Francisco, California, is adjusted in order to measure the same as a ballpark that is very friendly to hitters, like Great American Ballpark, located in Cincinnati, Ohio. The interpretation of wRC+ is also relatively simple to understand. A wRC+ of 150 denotes that the player is performing 50% better than the league average, and any wRC+ below 100 denotes a player who is performing below the league average.

A statistic like wRC+ is especially useful for this project because we are comparing players from different ballparks *and* from different time periods. wRC+ introduces an element of normalization that enables the equitable evaluation of players from unequal eras, accounting for the variations in offensive conditions and ballpark dimensions. It does a particularly good job of eliminating external factors, facilitating a balanced assessment of a player's offensive prowess and ensuring that the performance of a player is evaluated within the context of the broader baseball landscape.

The incorporation of wRC+ into the calculation of each player's grade enriches the analysis by providing a normalized framework for evaluating offensive production. This metric is weighted in this project's grade calculation at 0.15.

Data Collection

The data used for this project was collected from two reputable and comprehensive sports reference and encyclopedia sites: Baseball-Reference and Fangraphs. Both sites comprise historical player data and game logs dating back as early as the 1870's. Both provided the required data for the three main categories of players analyzed in this project: elected hall of famers, current Hall of Fame candidates, and past rejected Hall of Fame candidates.

Data for all of these categories, along with their visualizations, were aggregated into a Google Sheet file, which itself is made up of a number of sheets. The following subsections detail each sheet and what it holds.

Hall of Fame Data

This sheet houses every currently elected Hall of Fame position player, sorted by position. It contains some rudimentary data, such as the player's career length and how many games they played, along with some interesting, but unused, career totals, like number of hits

and home runs. It also contains the data used to calculate the numerical grade *G* developed by this project: oWAR, dWAR, w/OBA, and wRC+. As an example, take this truncated entry for Harmon Killebrew, a Hall of Fame first baseman that played from 1954 to 1975:

Name	Position	oWAR	dWAR	w/OBA	wRC+	HoF grade
Harmon Killebrew	1B	71.5	-18.7	0.389	142	**44.35**

The first thirteen columns in the sheet are all fetched from Baseball-Reference. The next two are fetched from Fangraphs. Finally, the "HoF grade" holds the result of the numerical grade *G* calculation defined in the *Approach* section.

At the bottom of this sheet, averages are calculated for each labeled position: catchers (C), first basemen (1B), second basemen (2B), third basemen (3B), shortstops (SS), outfielders (OF), and designated hitters (DH). For this project, all outfield positions (left field, center field, and right field) were combined into just one position because a majority of outfielders, Hall of Fame-elected or not, play multiple outfield positions over the course of their careers; thus, the positions are referred to as one position.

A number of the players elected to the Hall of Fame also switched primary positions at some point in their careers. To handle this case, the position assigned to each player in this project was the position they started the most games at in their career. This assignment was verified by Fangraphs.

Projected 2024 Hall of Fame Candidate Data

For the 2024 Hall of Fame candidates, a new sheet was created that housed the same career statistics for each player, also sorted by position. At the time of this writing, there are twenty two projected candidates that will appear on the 2024 Hall of Fame ballot; thus, the sheet

has twenty two rows, one for each player. Unlike the first sheet, however, it contains two extra columns that describe the candidate's relation among the existing body of elected hall of famers *at the candidate's respective position*.

relation to average HoF at position	percentile among all HoF at position

Using the average positional HoF grade calculated in the **HoF data** sheet, a difference *D* is taken between the candidate's grade and the average positional grade:

$$D = G_{candidate} - \mu(G_H)$$

This is placed in the first column. The second column calculates a percentile *P* among the existing body of hall of famers at the candidate's respective position. It sorts the list of current hall of famers at that position according to their hall of fame grade *G*, inserts the candidate's grade among this list, and calculates the percentile using the PERCENTRANK function. The full formula for *P* is described by the following:

P=ROUND(PERCENTRANK({avg. HoF grade at position;candidate's HoF grade}, candidate's HoF grade) * 100)

Lastly, this sheet contains a **chances of making HoF** column, which makes a prediction about a given candidate's chances to be successfully elected into the Hall of Fame. This prediction is based on the values calculated in the aforementioned **relation to average HoF at position** and **percentile among all HoF at position** columns. Both of these values are analyzed to create a prediction with confidence as high as eighty percent and as low as five percent.

Details and examples about how the prediction model calculates a candidate's chances of successful election are covered in the *Analytic Methods* section.

Rejected Hall of Fame Candidate Data

This sheet, like the **Project 2024 Hall of Fame Candidate** sheet, contains all of the same information except for the **chances of making HoF** column since these candidates are no longer eligible for election by the BBWAA. It contains a number of candidates that were accurately denied election into the HoF according to the prediction model as well as a number of players who clearly stood out (even amongst the current body of hall of famers) but were not elected. Like the previously mentioned sheet, after the rejected player's HoF grade is calculated, their relation to the average hall of famer at their respective position as well as their percentile rank among all existing hall of famers at their respective position is also calculated.

Name	Position	oWAR	dWAR	w/OBA	wRC+	HoF grade	relation	percentile
Lou Whitaker	2B	67.7	16.3	0.353	118	**49.72**	**3.70**	**75**

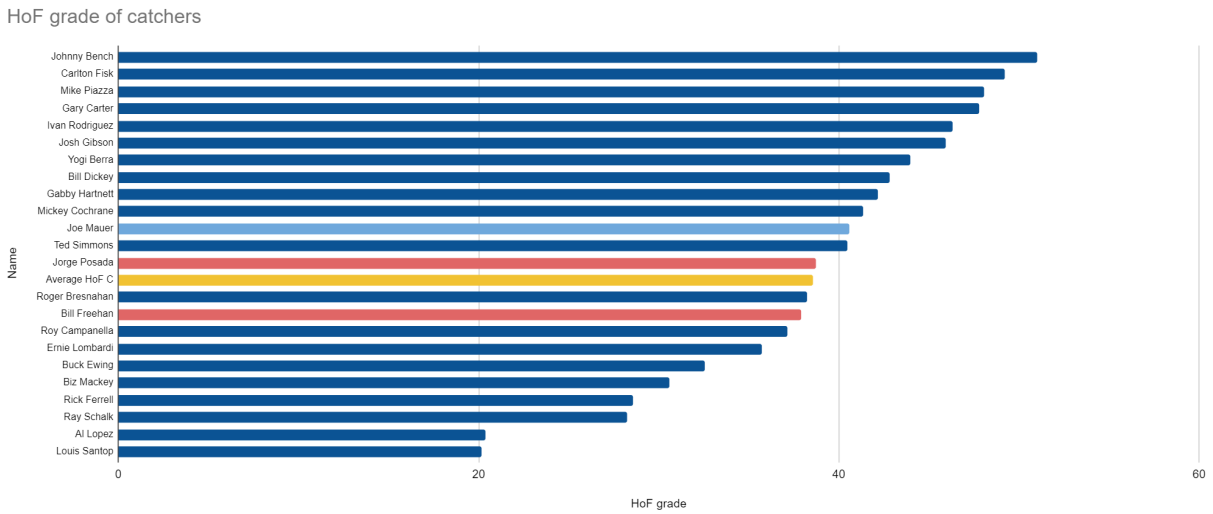
Positional Data

There are additional sheets that separate all of the players in this sheet (hall of famer, hall of fame candidate, or rejected hall of fame candidate) by their respective primary position. Currently eligible Hall of Fame candidates are marked with an *italicized player name*, and rejected candidates are marked with an underlined player name. These sheets are used mainly for the visualizations and graphs created in the **visualizations** sheet.

Visualizations

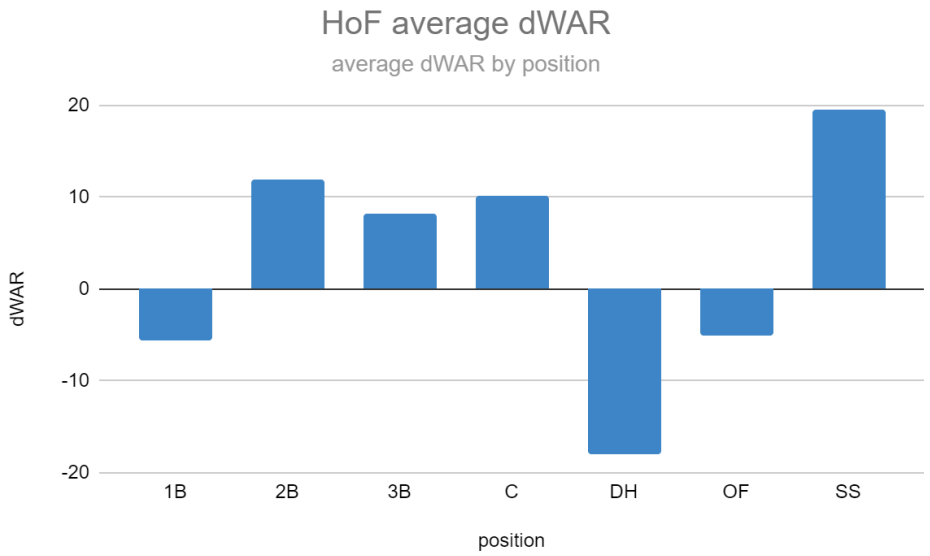
This sheet contains graphs and visualizations for the data collected and organized throughout the Google Sheet file. The graphs use sorted data organized into a hidden sheet called **tmp**. The reason this sheet is hidden is because data from elected hall of famers, candidates, and rejected candidates are all intertwined and sorted by their HoF grade. This sorted data is used to

create multiple kinds of graphs and visualizations, such as an ordered list of all catchers who were included in this project:



All elected hall of famers are marked with the dark blue bar. The “average hall of famer” at their position is denoted with the gold bar. Currently eligible candidates are marked with the light blue bar. Rejected candidates are marked with the red bar.

Also in this visualization sheet are averages for the different offensive and defensive statistics used to calculate the HoF grade for every elected hall of famer, organized by position.



As an example, the above bar graph shows the average dWAR at every position in the Hall of Fame. Shortstops rank the highest among all positions at 19.5, while designated hitters rank the lowest with an average dWAR of -18.

Analytic Methods

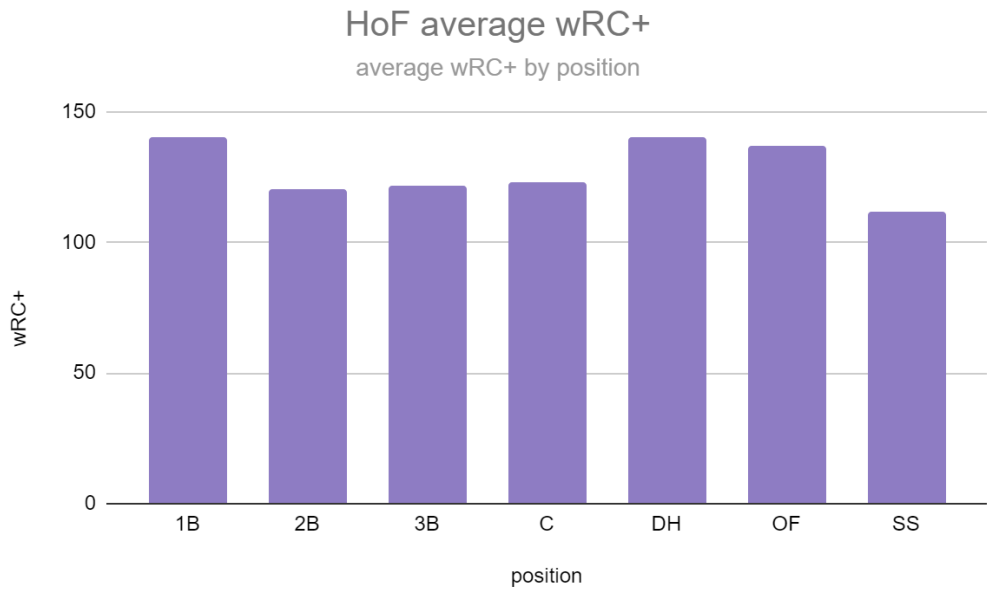
In the previous section, I stepped through each category of player analyzed for this project. I'll now go through each category and describe the analyses performed.

Hall of Famers

One of the refinements made on the calculated grade *G* of every elected Hall of Fame positional player was the aggregation of this grade by position. This allowed for discovery of averages of each statistic used in the Hall of Fame grade calculator as well as for comparison.

One of the key findings during the analysis phase of this research project pertained to the disparities in the thresholds for Hall of Fame candidacy among different positions. In one case, it was observed that the average career wRC+ for hall of fame first basemen exceeded that of hall

of fame second basemen by approximately 20 points, which suggests a significant discrepancy in offensive production between these positions.



However, delving deeper into the average positional data revealed a more interesting outcome. Despite first basemen exhibiting significantly higher career wRC+ and w/OBA values on average, their average overall grade *G* for Hall of Fame induction was three points lower than that of second basemen: 43.32 for first basemen compared to 46.02 for second basemen. Even more extreme was the offensive disparity between HoF shortstops and HoF first basemen. Although the average career wRC+ and w/OBA for shortstops were 38 and 48 points lower, respectively, shortstops also ranked close to two points higher in their average overall grade compared to first basemen.

This outcome confirmed one hypothesis made before starting this project: different positions, in the eyes of voters, have been analyzed differently depending on the position they played. Certain contributions from a player have been measured differently based on the position they played. Shortstops were (and, for the most part, still are) not known for their massive

run-producing prowess on the offensive side, but rather the runs they prevented by playing elite defense. They are often smaller in size compared to some of the other positions played on the field, but find meaningful ways to help their team win, like getting on base and stealing bases. First basemen and designated hitters, on the other hand, have been given more slack by organizations in terms of their defensive proficiency in favor of more responsibility when they are up to bat. Typically known as larger, but slower, athletes, these two positions are centered around driving in runs and hitting for extra bases.

There are exceptions to every position, as there were very large shortstops who were elite hitters and smaller first basemen who were elite defenders. Cal Ripken Jr. is perhaps the most prominent example. Ripken stood at six foot four and weighed around 200 pounds, an extreme outlier in terms of size compared to the rest of major league shortstops in the 1980s and 90s. While he was an elite defender throughout the majority of his career with a career dWAR of 35.7, the highest among all hall of fame shortstops, his overall HoF grade is significantly boosted by the fact that he is one of the greatest hitters to ever play the position, posting a career w/OBA of .346. Of all elected shortstops, he leads them all in home runs, hitting 431 of them.

Ernie Banks is one of the few HoF first basemen to have played a significant number of games at another position (Banks played over 1100 games at shortstop), but is one of only four elected first baseman with a positive career dWAR. Banks was a decent hitter, but not an elite one compared to his positional HoF peers, ranking in the bottom percentile in terms of his offensive metrics.

With this analysis comes an interesting raised question: **should voters use distinct thresholds when evaluating the candidacy of players from different positions for the Hall of Fame?** This question expounds upon one of the conventional approaches of applying a

somewhat uniform standard across all positions and underscores the necessity of considering the unique demands and roles associated with each position on the field.

The intriguing juxtaposition of higher offensive output from certain positions and the apparent higher (and lower) standards for a player's HoF induction prompts a deeper evaluation of how we assess the worthiness of players across positions. This aspect of the analysis underscores the complex interplay between offense and defense in the context of Hall of Fame candidacy and hints at potential biases or historical voting trends that may influence these standards.

Current Hall of Fame Candidates

The **Data Collection** section outlined how additional calculations were performed on a given current hall of fame candidate based on the overall Hall of Fame grade *G* that was found for each candidate. These two additional calculations were:

1) The difference *D* between the candidate's grade *G* and the average grade $\mu(G_H)$ of an elected hall of famer at the candidate's respective primary position, and
2) The percentile rank *P* of the candidate's grade *G* among a sorted list of grades *G* for all elected hall of famers at the player's respective position.

Using these two additional calculations, a probability table was created describing the probability of a player's successful election into the Hall of Fame:

	G >= $\mu(G_H)$ - 2	G >= $\mu(G_H)$ - 3	G >= $\mu(G_H)$ - 5	G >= $\mu(G_H)$ - 7	G >= $\mu(G_H)$ - 9
> 45th percentile	80%	65%	50%	35%	20%
<= 45th percentile	65%	50%	35%	20%	5%

This table provides answers to two questions:

1) Is the candidate being evaluated at least nine points below, or at around 20% of the average hall of famers' grade at his primary position?
2) Does the candidate rank among the current list of elected hall of famers at his position in at least the 45th percentile?

The relation between these two follow-up calculations can be strongly linked to a player's actual voting success in their official election. If a candidate fulfills both conditions at the highest level, i.e. the player is at least two points below the average HoF grade $\mu(G_H)$ at their primary position *and* the player ranks in at least the 45th percentile of all elected hall of famers at their primary position, then a strong case can be made that the player's chances of being elected stand at around 80%. The farther a candidate's grade *G* falls below the average $\mu(G_H)$, the lower chances a candidate has to be elected.

The table used to make these measurements were compared with the voting results of last year's 2023 Hall of Fame ballot. The following subsections outline a few of the 2024 MLB HoF candidates that positively showcase this model's accuracy.

Todd Helton.

In the 2023 Hall of Fame voting, Todd Helton received 72.2% of votes, falling just short of election, which currently requires 75% of votes from members of the BBWAA. According to the HoF grade calculator used in this project, Helton's grade *G* is 41.51, which sits at just 1.81 points below the average grade of all current HoF first basemen. He also ranks in the 48th percentile of all HoF first basemen. This satisfies both conditions for the highest probability of being elected into the HoF according to our table, putting his chances of being elected at around 80%. 2024 will be his 6th year on the ballot, where his percentage of votes has steadily increased

every year: 16.5%, 29.2%, 44.9%, 52.0%, and 72.2%. In 2024, this model is confident that he will cross the 75% threshold and be enshrined into the Hall of Fame.

Andruw Jones.

Andruw Jones is another interesting case that validates the prediction model constructed. In 2023, Jones' 6th year on the ballot, he received only 58.1% of votes, a ways off from the 75% he still needs to be elected into the HoF. Andruw Jones' HoF grade *G* is 39.94, which is 4.14 points below the average grade of a HoF outfielder. He also ranks at the 48th percentile of all current HoF outfielders, plenty ahead of a number of great players already elected. At first glance, Jones looks like a somewhat strong candidate for election. His percentage of votes, however, have been disappointing: 7.3%, 7.5%, 19.4%, 33.9%, 41.4%, and 58.1%. At this rate, it will be a somewhat close call whether Jones will receive 75% of votes by the time he reaches his 10th (and final) year of eligibility on the ballot. This also strongly validates the project's model of outputting his chances of election at around 50%.

Alex Rodriguez.

Alex Rodriguez is a case that specifically highlights why the highest percentage of confidence of the prediction model is capped at 80%. Rodriguez was one of the great shortstops and third basemen of the early 21st century, enjoying extremely successful careers at both positions. Rodriguez has a staggering HoF grade *G* of 70.45, placing him second among all current HoF shortstops, around seven points below Honus Wagner, who ranks first with a grade of 77.86. Rodriguez clearly satisfies both conditions of the prediction model and seems like a lock for HoF election. However, his voting results after just two years on the ballot tell a much different story: 34.3% and 35.7%. This is not a significant increase and places many doubts on

whether Alex Rodriguez will actually ever get elected to the HoF. This is where a multitude of other factors come in.

Alex Rodriguez served a year-long suspension in the 2014 season for his use of performance-enhancing drugs, a penalty that cost him around $25 million and, now most likely, his chances of being elected to the HoF (Sanchez, 2014). For the voting committee, this is oftentimes a strong enough case to deny a player's entry into the HoF, regardless of the statistics they put up in their careers, before *and* after their suspension. So, while the prediction model gives Rodriguez an incredibly strong chance to be elected to the HoF, factors outside of statistical measurement hinder his HoF future; thus, the model includes the 20% chance that he *does not* get elected.

Rejected Hall of Fame Candidates

There are a number of players who never reached the 75% threshold of election into the Hall of Fame that deserve their own discussion. This project includes a list of players who are no longer eligible for a number of different reasons, and was constructed with intentions of including three types of players:

1) players who were justly left out of the Hall of Fame,
2) players who might have been wrongly left out of the Hall of Fame based on misleading traditional statistics, and
3) players who statistically made more than enough of a case to have been elected but were not chosen because of factors other than statistical performance.

Dale Murphy.

Dale Murphy played for eighteen years and had a long and considerably successful career as an outfielder for the Atlanta Braves, Philadelphia Phillies, and Colorado Rockies. He earned a

number of prestigious accolades throughout his career, winning two Most Valuable Player, five Gold Glove, and four Silver Slugger awards. Using the advanced statistics in this project, however, his HoF grade *G* stands at 35.42, 8.66 points below the average grade of a HoF outfielder. He ranks in just the 25th percentile of all current HoF outfielders, a rank seemingly too low for election. According to the project's model, his chances of successful election stood at just 5%. This is close to the actual voting results he received over his 15 years on the ballot. He received as high as 23.2% of votes in his second year on the ballot, but only received an average of 13.92% of votes.

Lou Whitaker.

Lou Whitaker also had a long career in the major leagues as a second baseman, enough so that he is worthy to be included in our list of candidates that probably deserved a second look at election. He played for nineteen years, building up a career oWAR and dWAR total of 67.7 and 16.3, respectively. He only produced a negative dWAR in two of those seasons: his rookie season (-0.2) and his final season (-0.4). His HoF grade stands at 49.72, which is 3.70 points *above* the average grade of a HoF second baseman. He also ranks in the 75th percentile of all HoF second baseman, 6th out of the twenty total second baseman currently elected. His HoF voting results, however, tell a very different story. His sole year on the ballot in 2001 produced a voting total of just 2.9%, low enough to remove him from further consideration of the Hall of Fame. Our prediction model would have placed his chances of being elected at 80%; however, Whitaker was probably unjustly neglected based on a tunnel vision approach of using traditional statistics; his career batting average was a modest .276 and only hit 244 home runs over the span of his nineteen seasons. Whitaker went unconsidered for enshrinement after just one year on the ballot.

Barry Bonds.

Perhaps the most famous case of someone being left out of the Hall of Fame, Barry Bonds stands as possibly the greatest player ever to not be elected. His career could be split into two different chapters: the first, a seven year career with the Pittsburgh Pirates, a span in which he won two Most Valuable Player awards, two All-Star selections, three Gold Glove awards, and three Silver Slugger awards, and the second, a fifteen year career with the San Francisco Giants, where he won four more MVP awards, twelve more All-Star selections, five more Gold Gloves, and nine more Silver Sluggers. He has the most home runs, walks, and intentional walks of any player who has ever lived. His HoF grade stands at 85.74, over forty points above the average grade for a HoF outfielder. In fact, he ranks just second behind Babe Ruth among all current Hall of Fame outfielders and all current Hall of Fame position players. The prediction model constructed in this project would have placed his chances of election at 80%, an extremely confident prediction that he would have been elected a Hall of Famer. His voting results paint a different and bleak picture, receiving as low as 36.2% of the vote in his first year on the ballot and peaking at just 66% of the vote, falling nine percentage points shy of being elected, a significant margin. Like the case of Alex Rodriguez, more factors need mentioning to tell his whole story.

Barry Bonds was also caught for his performance-enhancing drug use in November of 2007, where he was indicted on perjury (Trial of Barry Bonds, n.d.). It was a national case that sparked plenty of media attention, a space where Bonds was already looked at pretty unfavorably. His relationship with the media had never been positive, and thus, when his career ended in 2007 and he walked away from the game, his image had been tainted by his personality and his criminalities, enough so to fully prevent him from being elected to the HoF by the very

people he had a poor relationship with. Although his career numbers tell a fascinating story of one of the greatest players to ever play the sport, he will never have the accolade of being enshrined among the greatest.

Like Alex Rodriguez, his case is one of the primary reasons why the prediction model in this project is capped at 80%. For every candidate, there are anecdotes, relationships, and images that affect his legacy when all is said and done. For better or for worse, they affect the way voters think about players' careers and their case for election into the Hall of Fame. Election is not always a simple calculation and it raises questions about both the integrity and the responsibility of both the Hall of Fame and Major League Baseball.

Results

Different Statistical Thresholds for Each Position

One of the more interesting results found while gathering, aggregating, and processing the HoF data was the discovery of different statistical thresholds based on the position played by a player throughout their careers, which confirmed a hypothesis before starting this project. I will go over some of the more interesting discoveries here.

The highest average HoF grade belonged to second basemen at 46.02; the lowest average HoF grade belonged to catchers at 38.57.

The average HoF designated hitter had the lowest average dWAR at -18; this is somewhat expected because part of the reason designated hitters stay at this position is because they are great hitters but are unreliable defenders; they also have the second-highest average oWAR at 61.2. The average HoF first baseman and outfielder both average negative dWAR at -5.4 and -5.1, respectively. We can conclude here that first basemen do not have a very high threshold for elite defense as a measure of HoF candidacy, as they are more relied on for their high-impact

offense. Outfielders play a more important role defensively for their teams, but, somewhat surprisingly, the importance of their defensive accomplishments are far outweighed by their offensive contributions. The highest average oWAR among all positions belong to outfielders at 62.5.

Shortstops and second basemen have the highest average dWAR among all positions at 19.2 and 12, respectively. This is an expected result, as both positions are incredibly important defensively, and most shortstops and second basemen elected into the HoF are heavily known for their defensive excellence throughout their careers.

The general conclusion to be made here is that there are different expectations for HoF candidates depending on the primary position of a player in his career.

Different Statistical Thresholds for Each Era

Only somewhat related to the project, but another interesting discovery was the threshold for HoF candidacy, irregardless of position, increasing as the game has matured. The sport has been around for more than a hundred years, and with that, the average player has also gotten better. Thus, the standards for a Hall of Fame career have also risen. We are unlikely to see a player get elected to the Hall of Fame that does not have a HoF grade at least in the 45th percentile of all current hall of famers at their respective position. This is shown by the HoF grade visualizations broken down by position.

As more time goes on, we will likely see a further increase in statistical thresholds for newer Hall of Fame candidates along with less successfully elected players *if traditional statistics continue to be used to determine a player's career contributions*. This is a result of the contemporary player development strategy in baseball. Hitters are taught and trained to make less weak contact and instead focus on two outcomes: hitting for power, or getting on base via

the walk. With the shift towards this philosophy comes a third and statistically unfavorable outcome: the strikeout. Thus, most modern day players, when analyzed at the end of their careers, look feeble compared to their Hall of Fame counterparts using traditional statistics like batting average. There are a number of other factors that are outside the scope of this project, so they will be excluded here, but for the sake of both the longevity and the validity of the Hall of Fame, it is important for voters to come to an understanding of looking at a more comprehensive and complex picture when analyzing a modern day candidate's career to avoid falling into early dismissal of their legitimacy using flawed and outdated statistics.

Strong Predictions on Future Candidates Can Be Made with Just Four Variables

The prediction model presented in this project does a good job of predicting the success of future Hall of Fame candidates. According to the model, only seven players out of a potential ballot of 22 players have an 80% chance of being elected. Of those seven players, the one candidate who has already been on the ballot previously (2024 will be Todd Helton's 6th year on the ballot) received 72.2% of the 75% needed to be elected, which is well within range of being elected on this next ballot.

Two of the seven candidates listed at 80% of successful election will be on the 2024 ballot for the first time in their careers: Joe Mauer and Adrian Beltre. Their careers are more than strong enough to make a case for the Hall of Fame and they are backed up by the prediction model, each with a HoF grade above the average at their respective positions and both at least in the 50th percentile of their respective positions among current hall of famers.

Multiple players were labeled with a 50% chance of election to the HoF according to the prediction model that were in line with actual voting results of previous years. Andruw Jones' voting trend has gone up steadily in his six years on the ballot, but is still a ways away from the

75% threshold. Thus, his 50% election prediction is in line with how his voting trends will pan out in his remaining four years of eligibility on the Hall of Fame ballot. Bobby Abreu is also labeled at just a 50% chance of election, and his voting results have solidified this claim, never peaking above 16% of votes in four years of eligibility. His candidacy is likely a coin flip as to whether he will ever be elected to the HoF.

External Factors Still Prove Significant Uncertainty in Voting Results

Three other players of the seven who were listed at an 80% probability of being elected face external issues regarding their candidacy: performance-enhancing drug use and a cheating scandal, both of which have severely affected their chances of successful elections. Regardless of the elite statistical careers put up by Manny Ramirez, Alex Rodriguez, and Carlos Beltran, their candidacy for the Hall of Fame is in severe jeopardy; these are examples of why the highest percentage prediction placed on a candidate is 80%.

Other famously rejected candidates, like Pete Rose and Barry Bonds, are no longer eligible for election by the BBWAA because of the same external troubles that cost them their relationships with both Major League Baseball and the writers, the folks ultimately responsible for voting and determining their place in the Hall of Fame. Even though their careers were among the greatest, we will likely never see them enshrined because of their troubles with performance-enhancing drugs and gambling scandals. It is a tough balance for both the fans and the sport itself, as there is a complicated balance of delivering a sensational product on the field for people to consume while also sticking to a rigorous ethos that claims an even playing field (Von Burg & Johnson, 2009).

Conclusion

As baseball has evolved into a wonderfully complex sport over time, so has the process of evaluating players for the Hall of Fame. This research project aimed to create a more objective and data-driven approach to assess the careers of hall of fame candidates by incorporating just four advanced statistics: oWAR, dWAR, w/OBA, and wRC+. It sought to provide a more nuanced understanding of player contributions beyond traditional metrics. Through analysis and the development of a comprehensive grading system, this project addresses some of the challenges present in Hall of Fame voting. Intriguing insights were uncovered that could have profound implications for the future of Hall of Fame induction.

This project represents a step toward a more transparent, objective, and data-informed Hall of Fame selection process. It highlights the importance of leveraging advanced analytics and creating a holistic view of player performance to provide a fair assessment of players' careers, regardless of the era that they played in or the position they took on the field.

Regardless of the methods and analysis performed in this project, a critical question remains: *what does it truly mean to be a Hall of Famer*? The answer is a dynamic and evolving one, deeply rooted in the rich tapestry of baseball history. By embracing modern statistical analysis and creating an open discourse about position-specific standards and historical comparisons, we can be sure that the Hall of Fame remains a shrine to excellence that includes athletes of both past and present as well as serving as a testament to the ever-changing nature of America's pastime.

In the years to come, as baseball continues to evolve, so too will the criteria for enshrinement in the Hall of Fame. This research project looks to be a humble contribution to

such ongoing dialogue, providing a blueprint for a more equitable and data-driven approach to honor those who have left an unforgettable mark on the sport cherished by so many.

References

BBWAA election rules. BBWAA Election Rules | Baseball Hall of Fame. (n.d.). https://baseballhall.org/hall-of-famers/rules/bbwaa-rules-for-election

Castrovince, A. (2020). *A Fan's Guide to Baseball Analytics: Why WAR, WHIP, WOBA, and Other Advanced Sabermetrics Are Essential to Understanding Modern Baseball.* Skyhorse Publishing Company, Incorporated.

Elitzur, R. (2020). *Data analytics effects in major league baseball. Omega (Oxford), 90*, 102001–. https://doi.org/10.1016/j.omega.2018.11.010

James, B. (2003). *The New Bill James Historical Baseball Abstract.* Free Press.

Kenny, B. (2016). *Ahead of the Curve: Inside the Baseball Revolution*. Simon & Schuster.

Law, K. (2018). *Smart baseball : the story behind the old stats that are ruining the game, the new ones that are ruining it, and the right way to think about baseball (First William Morrow paperback edition.)*. William Morrow, an imprint of HarperCollins Publishers.

Mills, B. & Salaga, S. (2011). Using Tree Ensembles to Analyze National Baseball Hall of Fame Voting Patterns: An Application to Discrimination in BBWAA Voting. *Journal of Quantitative Analysis in Sports, 7*(4). https://doi.org/10.2202/1559-0410.1367

Parsons, N., & Stern, M. (2012). There's No Dying in Baseball: Cultural Valorization, Collective Memory, and Induction into the Baseball Hall of Fame. *Sociology of Sport Journal, 29*(1), 62–88. https://doi.org/10.1123/ssj.29.1.62

Sanchez, R. (2014, January 13). Alex Rodriguez suspended for entire 2014 season. *CNN*. https://www.cnn.com/2014/01/11/us/alex-rodriguez-suspended/index.html

Sun, H., Lin, T., & Tsai, Y. (2023). Performance prediction in major league baseball by long short-term memory networks. *International Journal of Data Science and Analytics, 15*(1), 93–104. https://doi.org/10.1007/s41060-022-00313-4

Trial of Barry Bonds. *Trial of Barry Bonds - BR Bullpen*. (n.d.). https://www.baseball-reference.com/bullpen/Trial_of_Barry_Bonds#:~:text=On%20November%2015th%2C%20Bonds%20was,He%20pleaded%20not%20guilty.

Von Burg, R., & Johnson, P. (2009). Yearning for a Past that Never Was: Baseball, Steroids, and the Anxiety of the American Dream. *Critical Studies in Media Communication, 26*(4), 351–371. https://doi.org/10.1080/15295030903176641

Weighted on-base average (WOBA): Glossary. *MLB.com*. (n.d.). https://www.mlb.com/glossary/advanced-stats/weighted-on-base-average

Weighted runs created plus (wRC+): Glossary. *MLB.com* (n.d.). https://www.mlb.com/glossary/advanced-stats/weighted-runs-created-plus

Appendix

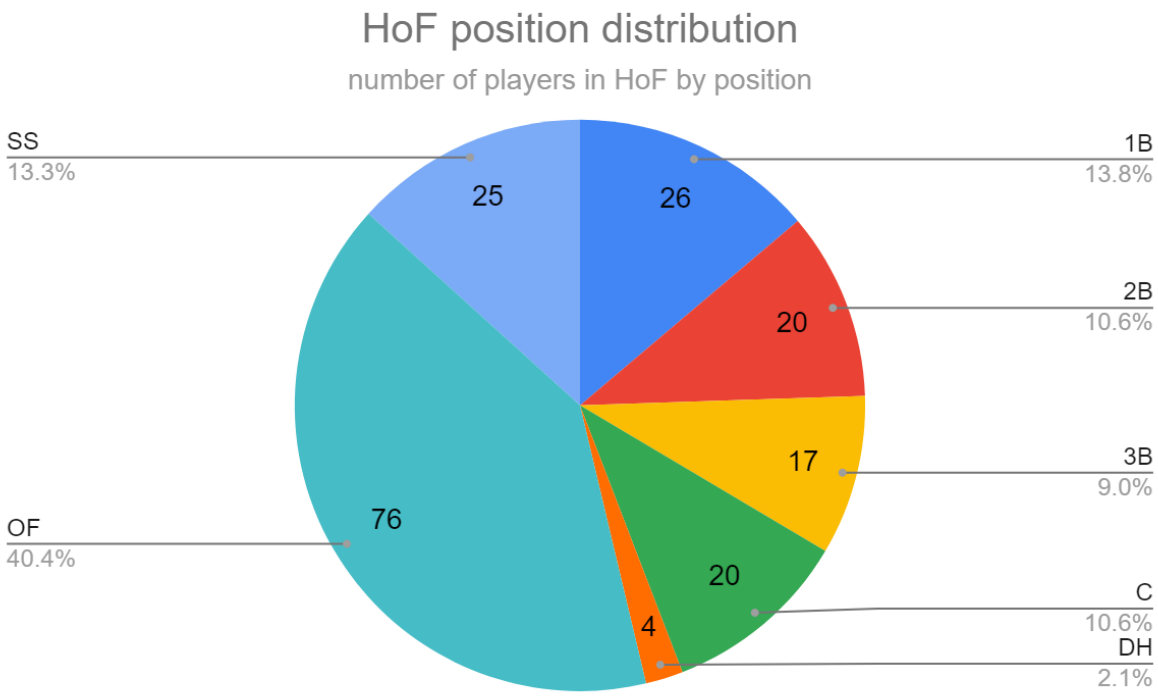


figure 1: a breakdown of the different elected Hall of Famers, grouped by primary position.

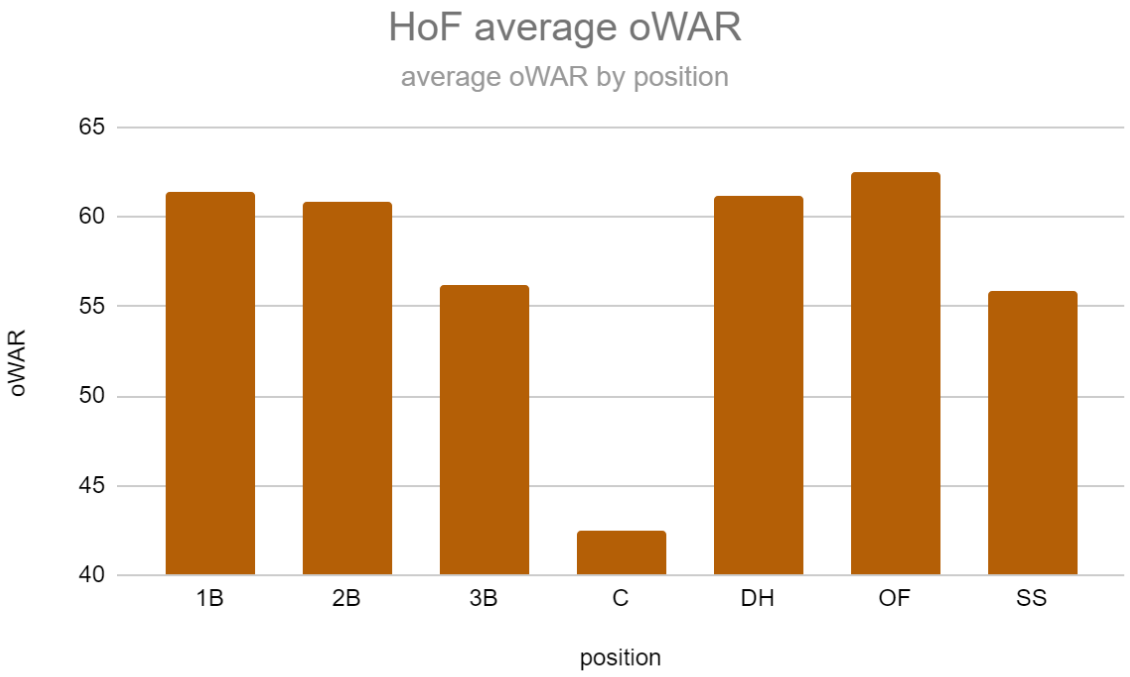


figure 2: average oWAR of an elected Hall of Fame player, grouped by primary position.

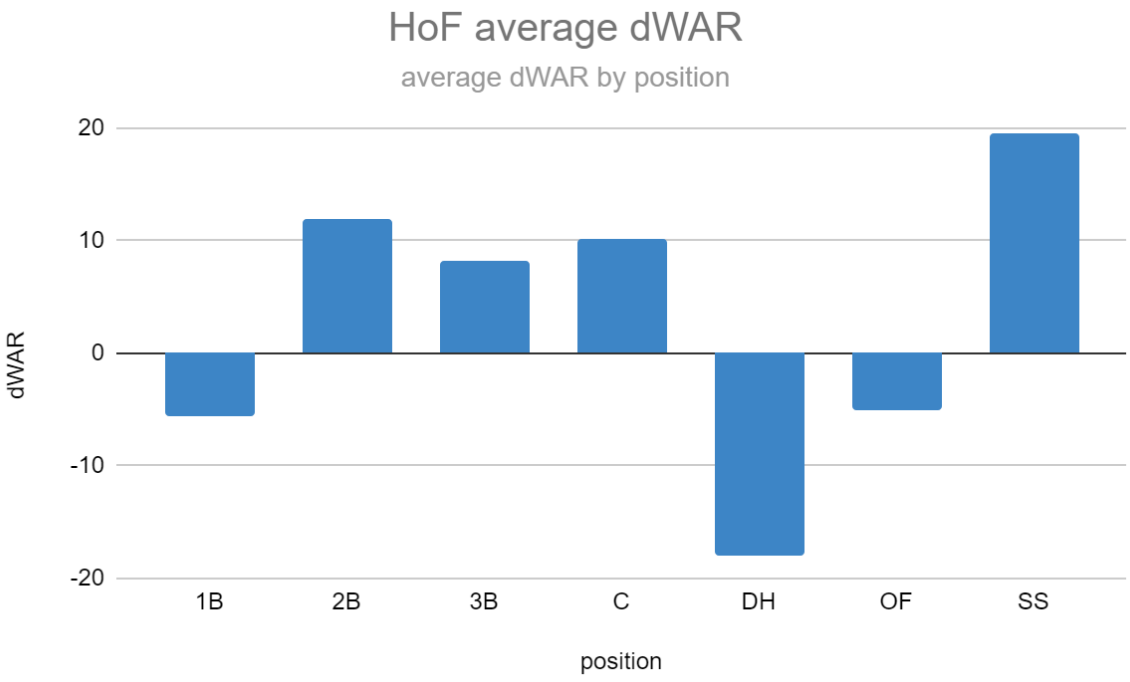


figure 3: average dWAR of an elected Hall of Fame player, grouped by primary position.

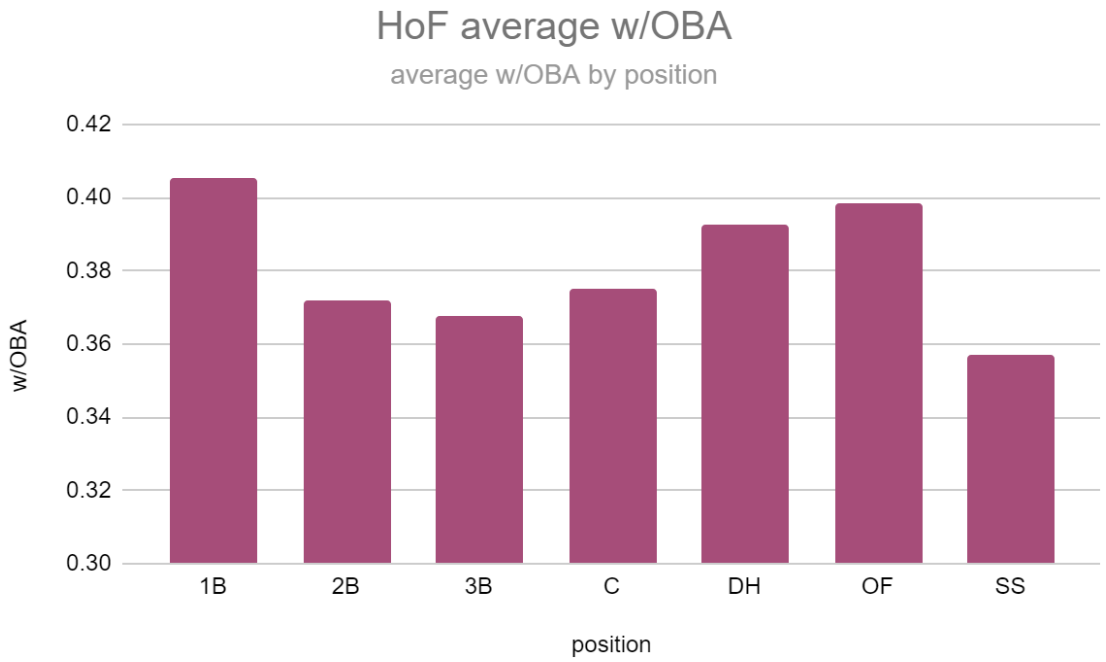


figure 4: average w/OBA of an elected Hall of Fame player, grouped by primary position.

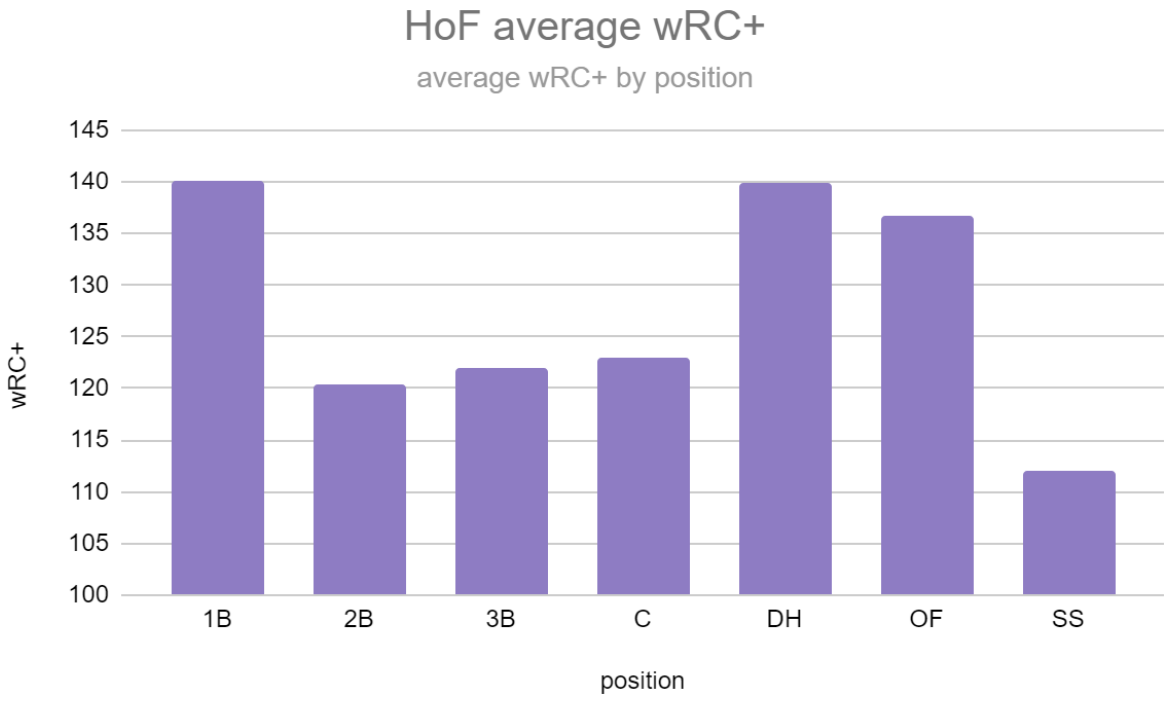


figure 5: average wRC+ of an elected Hall of Fame player, grouped by primary position.

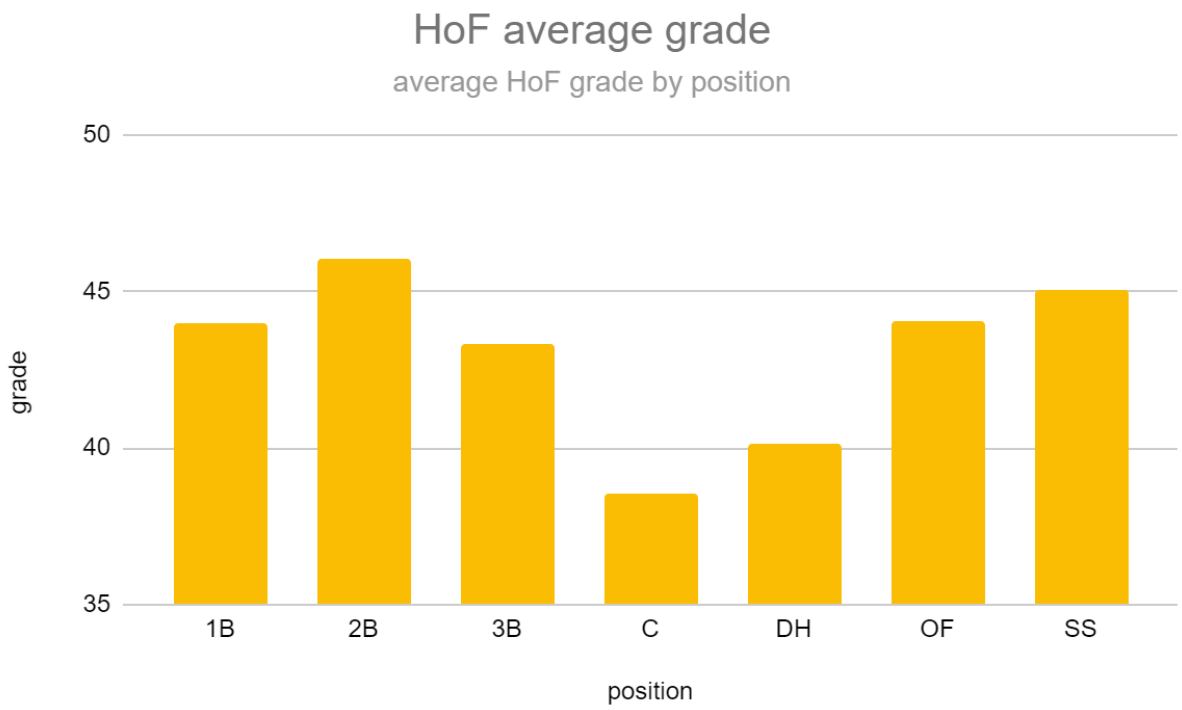


figure 6: average Hall of Fame grade of an elected Hall of Famer, grouped by primary position.