EXPLORING BASEBALL ANALYTICS THROUGH DATA VISUALIZATION

ABSTRACT

In this project, we use diverse data visualizations to illuminate the baseball statistics in many facets of the sport. We source data from trusted platforms, including Baseball Savant, Baseball Reference, and Spotrac, scrutinize trends spanning the past twenty-three years, identify statistical leaders from the most recent season in 2023, and extract valuable insights on player performance and rankings, encompassing both offense and defense metrics. Furthermore, we examine team payrolls and overall production and reveal the complex relationship between performance and financial investment. While this project does not aim to provide an infallible blueprint for athlete evaluation and team building, it gives insight into one of the myriad approaches teams can adopt to evaluate players and build successful strategies for the future of baseball.

Keywords: Baseball, data visualization, sports analytics, sports science.

INTRODUCTION

Baseball, America's first national pastime, has evolved dramatically, driven by changes in player performance, team strategies, and decision-making processes. To gain deeper insights into this dynamic sport, researchers and analysts have turned to the wealth of baseball statistics and data available today. The emergence of data-driven analysis, propelled by advancements in technology and data collection methods, has revolutionized how we evaluate player performance and team strategies in Major League Baseball (MLB).

The integration of advanced metrics, such as launch angles, exit velocities, Wins Above Replacement (WAR), and On-Base Plus Slugging (OPS), has allowed for a more multifaceted understanding of player contributions (Baseball Savant, 2023). This data-driven approach empowers teams to make informed decisions regarding player valuation, free agency, and in-game tactics (Barnes, 2019; Lee, 2022; Mizels, 2022; Romero, 2021).

In this paper, we delve into a comprehensive analysis of baseball data, examining various aspects of the game's evolution and player performance (Baseball Reference, 2023). Through systematically exploring statistics and trends, we aim to provide a broader perspective on how baseball analytics is transforming the sport. We also explore the financial side of the game, looking at the cost of player acquisitions and team payrolls (Spotrac, 2023). By the end of this analysis, we hope to offer a more informed understanding of how data analysis and sabermetrics are shaping the modern world of professional baseball (Albert, 2010).

BASEBALL GAME ANALYTICS METRICS

Baseball game analytics relies on an abundance of metrics to assess the performance of both individual players and teams. These metrics provide valuable insights into various facets of the game. Here, we will explore key offensive and defensive metrics, as well as individual player and team performance metrics

used in baseball analytics.

Offensive Metrics

Batting Average (BA): BA measures a player's ability to make hits, calculated as the number of hits divided by the number of at-bats. XBA refers to the Expected Batting Average.

Exit Velocity (EV): EV assesses the speed of the ball coming off the bat when a batter makes contact with the pitch. It is used to gauge a hitter's power and the quality of their contact with the ball.

Home Runs (HR): HR represents hits that clear the fences in fair territory, allowing runners to score.

Isolated Power (ISO): ISO measures a hitter's raw power by considering only extra-base hits and their types.

Launch Angle (LA): LA indicates the trajectory at which a baseball is launched when a batter hits a pitch.

On-Base Percentage (OBP): OBP assesses a player's capability to reach base safely, calculated as the number of times a player reaches base (via hits, walks, or being hit by a pitch) divided by plate appearances.

On-Base Plus Slugging (OPS): OPS combines OBP and SLG to provide an overall measure of a player's offensive performance.

Plate Appearance (PA): PA is a batter's turn at the plate, encompassing every interaction between a batter and a pitcher, and it is used to determine qualifications for statistical leaderboards, unlike at-bats.

Slugging Percentage (SLG): SLG quantifies a player's power-hitting ability, calculated as the total number of bases from hits divided by at-bats.

Strikeouts (K): K results from a pitcher throwing three strikes, causing the batter to be out and preventing them from reaching first base. Strikeout Percentage (K%) measures the frequency of strikeouts per plate appearance for a batter.

Walks (BB): BB occurs when a pitcher throws four balls outside the strike zone, enabling the batter to advance to first base.

Weighted On-base Average (wOBA): wOBA assigns values to each method of reaching base based on its contribution to projected runs scored. For instance, a double is assigned a higher value than a single.

Defensive Metrics

Outs Above Average (OAA): OAA quantifies a player's defensive contribution by measuring the number of outs made beyond what an average player would achieve. This metric is a valuable tool for evaluating defensive prowess.

Individual Player's Value Metrics

Wins Above Replacement (WAR): WAR is a comprehensive metric that assesses a player's value in all aspects of the game. It calculates how many more wins a player is worth compared to a replacement-level player at the same position. For position players, it considers batting, baserunning, fielding, and adjustments, while for pitchers, different WAR computations may use various statistics like RA9 or FIP. In the end, WAR quantifies each player's value in terms of specific numbers of wins.

Team Performance Metrics

WAR Accumulated: This metric totals the WAR accumulated by all players on a team's roster. It provides a measure of the team's overall performance by considering the contributions of individual players in terms of wins added to the team's total.

In baseball analytics, these metrics serve as the cornerstone for evaluating player and team performance, helping teams make informed decisions and fans gain a deeper understanding of the game.

LITERATURE REVIEW

Baseball analytics has gained lots of popularity over the years, and it started with the beginning of stat tracking in general in the 19th century (Barnes, 2019). Statistics have allowed people who understand the numbers to compare and contrast players based on their performance and have definitive answers as to which players are better than others. No longer do people have to rely on the "eye-test," where spectators go off of anecdotal and personal evidence of games attended to vouch for one player over another. Baseball analytics has taken it further and created stats encompassing all parts of the sport and even tried to create predictive models with advanced analytics for possible future performance of players. Nowadays, there are too many stats that attempt to quantify player performance (Mizels, 2022), so looking to narrow them down to the best stats to use in models can be quite a difficult task.

When choosing which impactful stats to look at to try and visualize the best players of today, we need to see which offensive and defensive stats correlate best to scoring runs and winning baseball games. There have been findings that show that On-Base Plus Slugging (OPS) is strongly and positively correlated with runs scored, indicating that OPS is a good predictor of runs scored (Albert, 2010). It gives legitimacy to why OPS is used (along with other stats like K% and Launch Angle) as our comparison amongst qualified batters below.

As for defensive stats, the newly created stat of Outs Above Average (OAA) is being used mainly because of its intuitive nature (Tango, 2020). OAA considers how far the fielder needs to range to reach a ball and how far the putout throw should be. According to Tango (2020), errors, zone rating, and UZR/DRS are previous iterations of OAA. It still needs to be a perfect science but believed to be the best at this point in time to measure defensive prowess.

In the context of evaluating overall team performance, a streamlined approach is using a statistical metric that comprehensively encapsulates the multifaceted contributions of individual players. Hence, the selection of Wins Above Replacement (WAR) stands as a prominent choice in this regard. WAR is a summary performance statistic that shows how a player contributes to wins above a replacement player and normalizes comparisons between players of different positions (Conforti, 2022). Conforti (2022) leverages the Wins Above Replacement (WAR) metric to assess the potential of draft picks. In addition, WAR serves as a valuable tool for quantifying roster production, offering a numerical representation of player contributions. Ultimately, WAR accumulation emerges as a pivotal metric for evaluating the comparative

performance of teams in the 2023 season relative to their respective payroll disbursements.

DATA AND DESCRIPTIVE STATISTICS

Data Sources

Our primary data sources include Baseball Savant, which provides detailed information on player performance in 2023 and trends in Exit Velocity and Launch Angle from 2015 to 2023. Additionally, we collected data from Baseball Reference, offering a comprehensive historical perspective with statistics over twenty-three years, including trends in slash lines, Three True Outcomes per game, and team-accumulated Wins Above Replacement (WAR). Lastly, we derived team financial data, specifically team payrolls, from Spotrac, enriching our analysis with insights into the financial aspects of Major League Baseball.

Data Dictionary

In the following, we provide a comprehensive overview of the variables and their statistical characteristics for 136 qualified players (by 3.1 PA per team game, about 502 for a full season) and the 30 MLB team descriptive statistics in the year 2023. These data dictionaries encompass various key metrics, including variable names, counts, means, standard deviations, minimum and maximum values, and percentile information, all of which are essential for understanding and interpreting the data associated with player and team performance.

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
Barrel Batted Rate (BB)	136	9.09	4.03	0.6	6.2	8.65	11.7	19.6
Batting Avg (BA)	136	0.26	0.03	0.2	0.24	0.26	0.27	0.35
Exit Velocity Avg (EV)	136	89.73	2.04	85	88.3	89.6	91.2	94.6
Strikeouts (K%)	136	20.93	5.23	5.7	17.2	21.05	23.75	32.7
Launch Angle Avg (LA)	136	13.12	4.36	1.5	10.5	13.1	16.12	23.9
On Base Percent (OBP)	136	0.33	0.03	0.27	0.32	0.33	0.35	0.42
On-Base Plus Slugging (OPS)	136	0.78	0.08	0.58	0.72	0.78	0.82	1.07
Slugging Percent (SLG)	136	0.45	0.06	0.28	0.41	0.45	0.48	0.66
Woba	136	0.34	0.03	0.26	0.32	0.34	0.35	0.44
Xba	136	0.26	0.02	0.21	0.24	0.26	0.27	0.36
Xiso	136	0.18	0.06	0.06	0.14	0.18	0.21	0.34
Xobp	136	0.33	0.03	0.28	0.31	0.33	0.35	0.44
Xslg	136	0.44	0.07	0.3	0.4	0.44	0.48	0.66
Xwoba	136	0.34	0.03	0.27	0.32	0.33	0.36	0.46

Table 1 2023 Qualified Players Descriptive Statistics

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
WAR Total	30	30.61	12.13	4.1	24.58	31.05	40.4	51.7
2023 Total Payroll	30	164043516	73802312.8	52742755	97476675.5	159139296	226281088	343556413
26-Man Payroll	30	106168142	56294711.1	30953011	58100345	96976044.5	153561358	219238779
Buried	30	3679328.9	2584160.84	972671	1789241.75	3217933.5	4059026.75	11683673
Retained	30	24078089.8	28365164.3	2037878	9854591.25	17111143	32691432.3	158207683
Roster	30	27.4	3.1	11	28	28	28	28
Win%	30	0.5	0.08	0.31	0.46	0.51	0.55	0.64

Table 2 2023 MLB Team Descriptive Statistics

DATA VISUALIZATION AND DISCUSSION

Offensive Trends 2000-2023

In this section, we explore the shifting offensive trends spanning from 2000 to 2023. This period has witnessed significant changes in the game's dynamics, particularly affecting key offensive statistics. We will look closer at the decreasing Batting Average (BA) and its relationship with the evolution of hitting mechanics, as illustrated in Figures 1 and 2. Additionally, we analyze the concurrent increase in Home Runs (HR) and Strikeouts (SO), providing insights into the factors driving this trend, as detailed in Figure 3. Furthermore, we examine the stable statistic for Walks (BB) and its importance amidst these transformations, as outlined in the same figure. Together, these trends offer valuable insights into These trends provide valuable insights into shifting offensive strategies and player performance over twenty-three years.

Decreasing Batting Average (BA) (Figure 1&2) - During the past twenty-three years, a notable shift has been observed in the emphasis on increasing launch angle and exit velocity with a focus on optimizing the mechanics of each swing. Consequently, this evolution has reduced reliance on adaptable swing choices and a corresponding decline in the inclination to simply ‘put the ball in play.’ These alterations in hitting approaches are a central factor contributing to the gradual decrease in batting averages over time.

Increasing Home Run (HR) and Strikeout (SO) (Figure 3) - The first part of the 2000s was marked by the crackdown on steroid usage, which explains the decrease from 2000 to about 2014. However, the advent of changes to how hitters approach at-bats began to counteract this, and we saw a rise in home runs again while also aiding in the strikeout totals going way up. Since 2000, we have seen an increase of about two and a half strikeouts per game as a result of this new hitting philosophy.

No change BB or Walk (Figure 3) - Surprisingly, one TTO (Three True Outcomes) has stayed relatively the same. Walks have slightly decreased over the past twenty-three years, which makes sense with the slight lowering of On Base Percentage (OBP) over the same period. New hitting philosophy continues to value Walks as an important statistic, so its statistical relevance remains; the fact that it has not changed much over the years is also noteworthy.

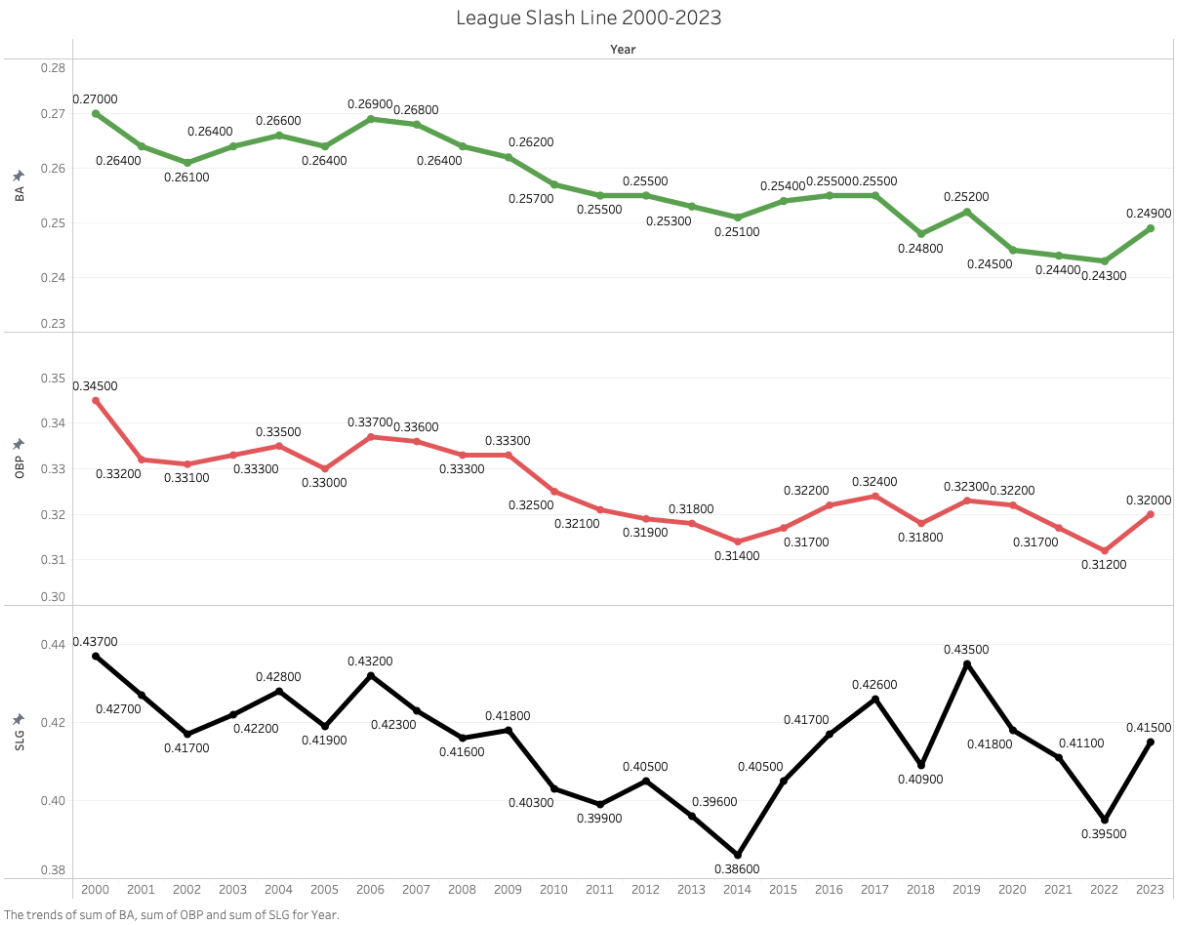


Figure 1 League Slash Line 2000-2023

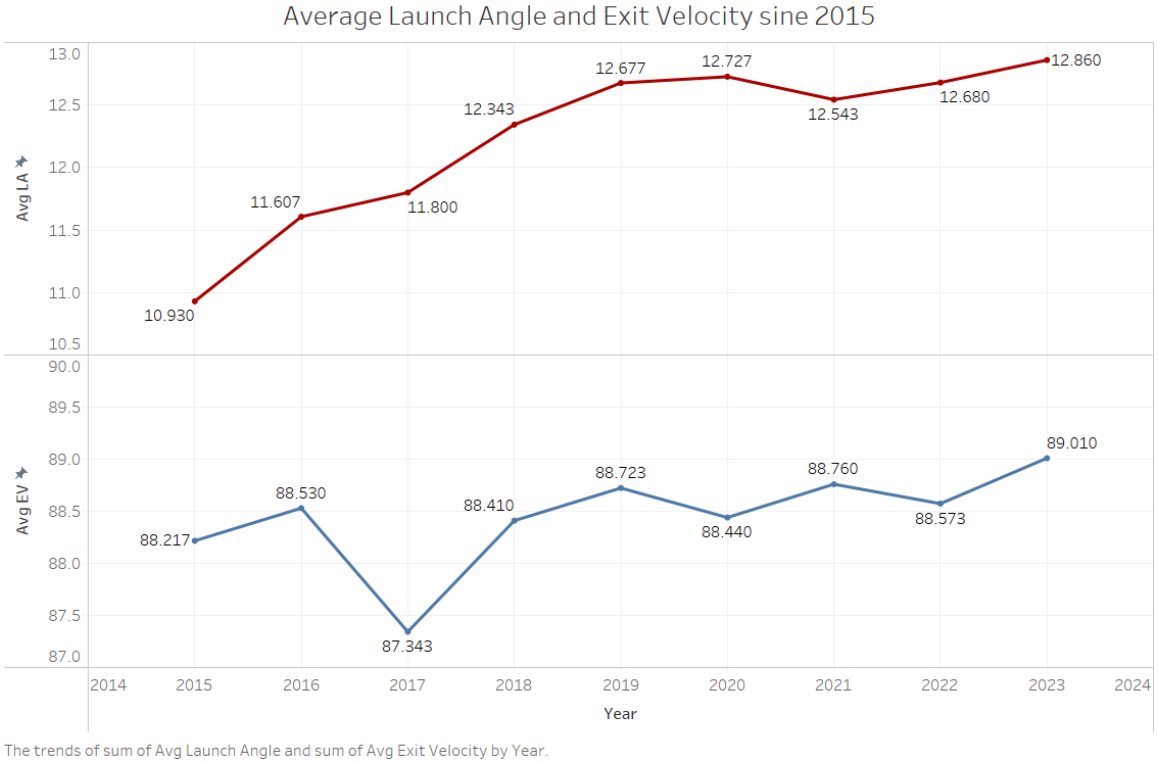


Figure 2 Launch Angle and Exit Velocity Trends 2015-2023

Figure 3 Three True Outcomes 2000-2023

2023 Trends

In this section, we examine key trends that have shaped baseball in 2023, offering valuable insights into player performance and the factors at play. We start by investigating the positive correlation between Exit Velocity and Slugging Percentage (SLG) illustrated in Figure 4. This relationship underscores the impact of a player's exit velocity on their ability to accumulate more bases per hit, subsequently influencing their slugging percentage. Figure 5 presents league leaders in OPS (On Base + Slugging) and their associated Strikeout Percentage and Launch Angle Average, revealing diverse performance metrics. Additionally, we identify the league leader in Outs Above Average (OAA), a crucial defensive metric that measures a player's ability to execute challenging plays. This analysis, presented in Figure 6, showcases the proficiency of different players in the field. These trends collectively provide a comprehensive overview of player performance and the changing dynamics of the game in 2023.

Figure 4 shows the positive correlation between Exit Velocity and Slugging Percentage (SLG). Typically, as a hitter looks to increase their exit velocity, it will give them more of an opportunity to collect more bases per hit than if they hit the ball with less exit velocity. Exit velocity is not the catch-all stat that relates to the increase in slugging, as many other factors influence slugging percentage, like a player's speed, what ballpark they play in, and the fielding prowess of the defense. However, in general, there is an opportunity for more bases to be collected the harder a ball is hit.

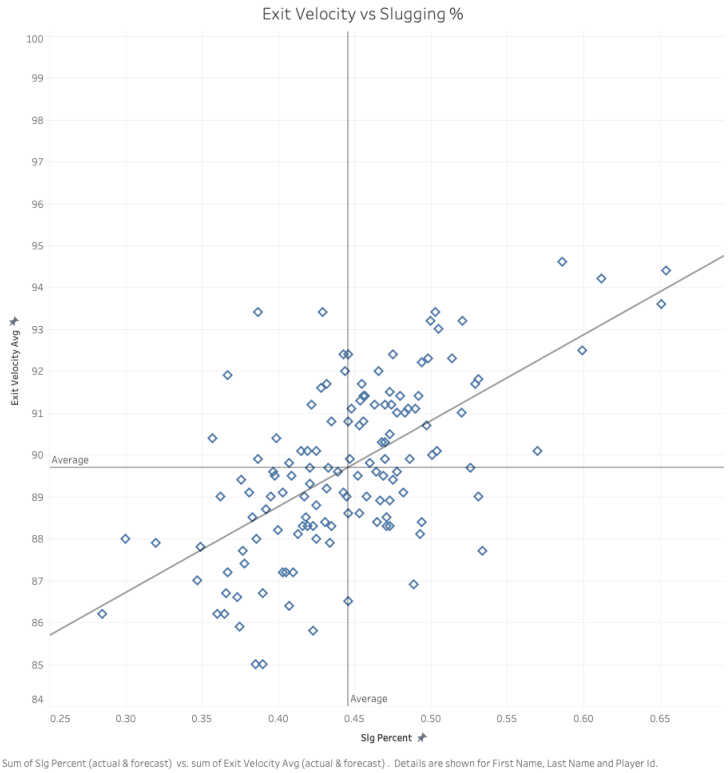
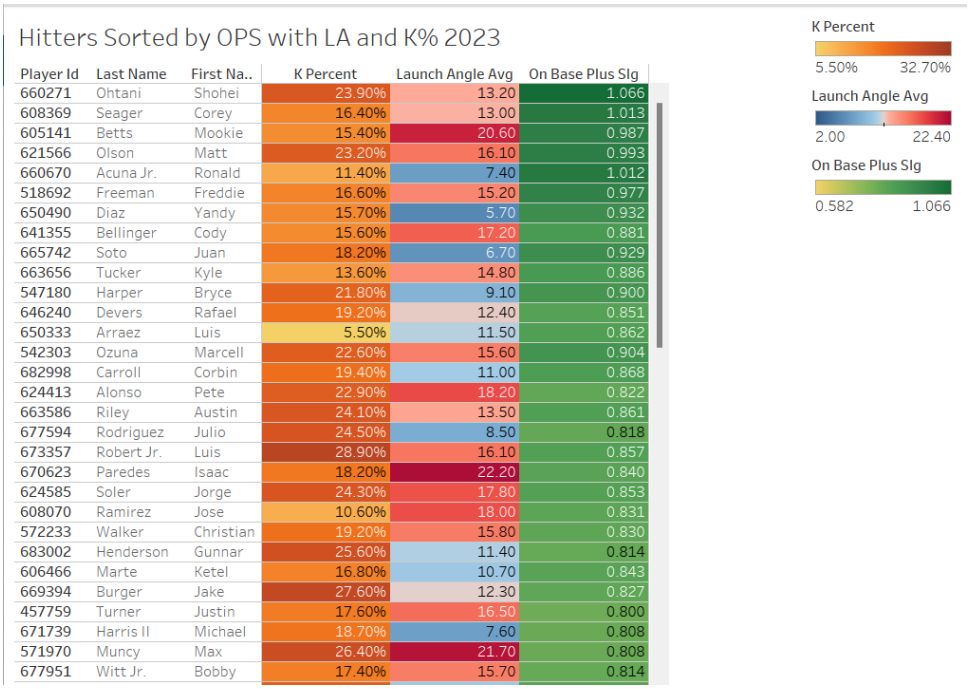


Figure 4 Positive Correlation between Exit Velocity and Slugging Percentage

On Base plus Slugging (OPS) encapsulates well what a hitter is doing All Plate (AP) appearances and is a great way to measure performance amongst peers. Strikeout percentage and Launch Angle Average give a glimpse into the other dimensions of the player, whether they put the ball in play often, and at what average launch angle these hits are. There are many ways a hitter can have a high OPS, and these metrics do a decent job of showing how effective they are. For example, Shohei Ohtani on this list leads the pack with an OPS of 1.066 with a strikeout percentage over twenty percent and a launch angle of about 13 degrees. Compare that with Luis Arraez, who has had success this season with an .862 OPS, but a strikeout rate of 5.7% and a launch angle slightly lower at 11.4 degrees. This shows the parity between approaches with similar amounts of success, with Arraez doing it more by being a hitting machine while Ohtani sells out more for power and more strikeouts. Arraez is an anomaly in this chart, as seen by the trends of way higher strikeout rates around him, coinciding with the change in the philosophy of hitting over the years. Arraez would be considered an "old-school" approach hitter, while the majority of all the other league leaders, Ohtani included, have bought into the new way of hitting.

Figure 5 shows the top 30 hitters ranked by Outs Above Average (OPS), Launch Angle (LA), and Strikeout Percentage (K%). Outs Above Average (OAA) is one of the better defensive metrics that encapsulates how well a player makes a play and how difficult it is to make such a play. It deals with the range necessary to make a play, moving in any direction, grades the difficulty, and awards positive or negative points depending on the result of the play. OAA plays based on which direction a fielder needs to move in order to make a play, whether it be in, back, left, or right. Then, if the play is on the ground in the infield, a putout at first is required as well. That is factored into the calculation of how hard the play is overall. Each direction has its grade, and all numbers are totaled up to get the final OAA total for each individual player on all defensive plays made over the course of a season. Figure 6 lists the top 30 qualified players in OAA, which shows who are great defenders. Players like Dansby Swanson of the Chicago Cubs grade out very well on defense by OAA, whereas a player like Enrique "Kike" Hernandez grades out very poorly.



Player Id	Last Name	First Na..	K Percent	Launch Angle Avg	On Base Plus Slg
660271	Ohtani	Shohei	23.90%	13.20	1.066
608369	Seager	Corey	16.40%	13.00	1.013
605141	Betts	Mookie	15.40%	20.60	0.987
621566	Olson	Matt	23.20%	16.10	0.993
660670	Acuna Jr.	Ronald	11.40%	7.40	1.012
518692	Freeman	Freddie	16.60%	15.20	0.977
650490	Diaz	Yandy	15.70%	5.70	0.932
641355	Bellinger	Cody	15.60%	17.20	0.881
665742	Soto	Juan	18.20%	6.70	0.929
663656	Tucker	Kyle	13.60%	14.80	0.886
547180	Harper	Bryce	21.80%	9.10	0.900
646240	Devers	Rafael	19.20%	12.40	0.851
650333	Arraez	Luis	5.50%	11.50	0.862
542303	Ozuna	Marcell	22.60%	15.60	0.904
682998	Carroll	Corbin	19.40%	11.00	0.868
624413	Alonso	Pete	22.90%	18.20	0.822
663586	Riley	Austin	24.10%	13.50	0.861
677594	Rodriguez	Julio	24.50%	8.50	0.818
673357	Robert Jr.	Luis	28.90%	16.10	0.857
670623	Paredes	Isaac	18.20%	22.20	0.840
624585	Soler	Jorge	24.30%	17.80	0.853
608070	Ramirez	Jose	10.60%	18.00	0.831
572233	Walker	Christian	19.20%	15.80	0.830
683002	Henderson	Gunnar	25.60%	11.40	0.814
606466	Marte	Ketel	16.80%	10.70	0.843
669394	Burger	Jake	27.60%	12.30	0.827
457759	Turner	Justin	17.60%	16.50	0.800
671739	Harris II	Michael	18.70%	7.60	0.808
571970	Muncy	Max	26.40%	21.70	0.808
677951	Witt Jr.	Bobby	17.40%	15.70	0.814

Figure 5 Top 30 Hitters Ranked by OPS, LA, and K%

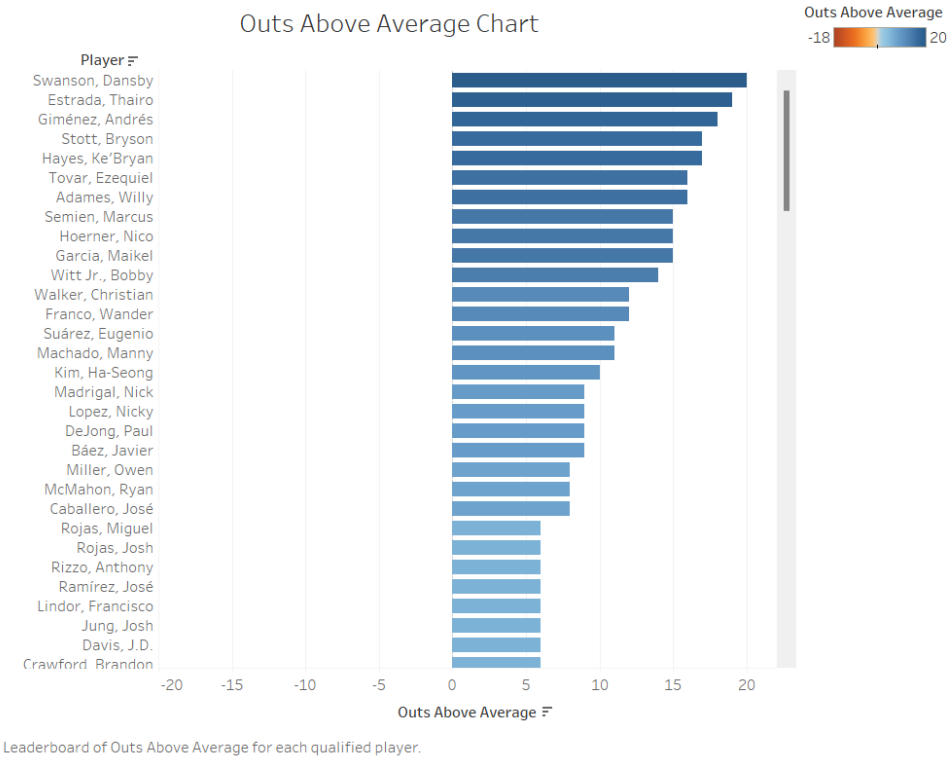


Figure 6 Top 30 Players Ranked by OAA

Team Payroll 2023

Figure 7 visualizes the differences in payroll amongst the 30 teams in the MLB. It shows which team owners are more willing to spend on players and which teams are looking to create value elsewhere and skimping on free agents. The New York Mets, far and away, spent the most on their roster this year, while teams like the Oakland A's and Baltimore Orioles are actively tanking or putting together young rosters that have not reached service time to warrant high salaries.

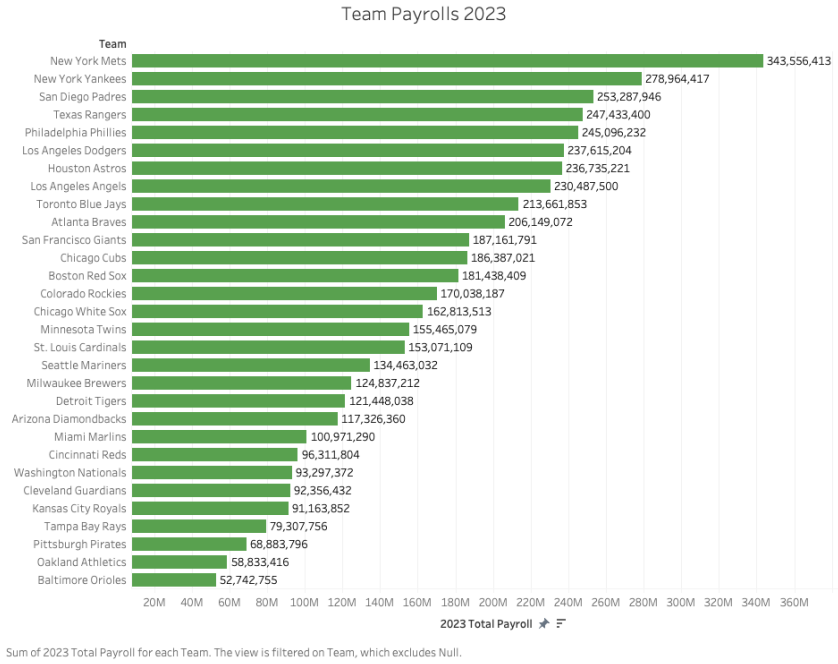


Figure 7 Team Payrolls 2023

Team Performance and Payroll

Figure 8 highlights various team spending strategies with their Wins Above Replacement (WAR) performance. This scatterplot is divided into four quadrants, each representing a distinct group of teams based on their spending efficiency and WAR production. Teams in Quadrant I are big spenders, but they are getting their money's worth, for the most part. Teams in Quadrant II are spending lots of money but getting less-than-average wins above replacement compared to others. Teams in Quadrant III are spending less and paying the price for it in low WAR totals. Teams in Quadrant IV have teams that cut corners but still are getting excess value out of their roster.

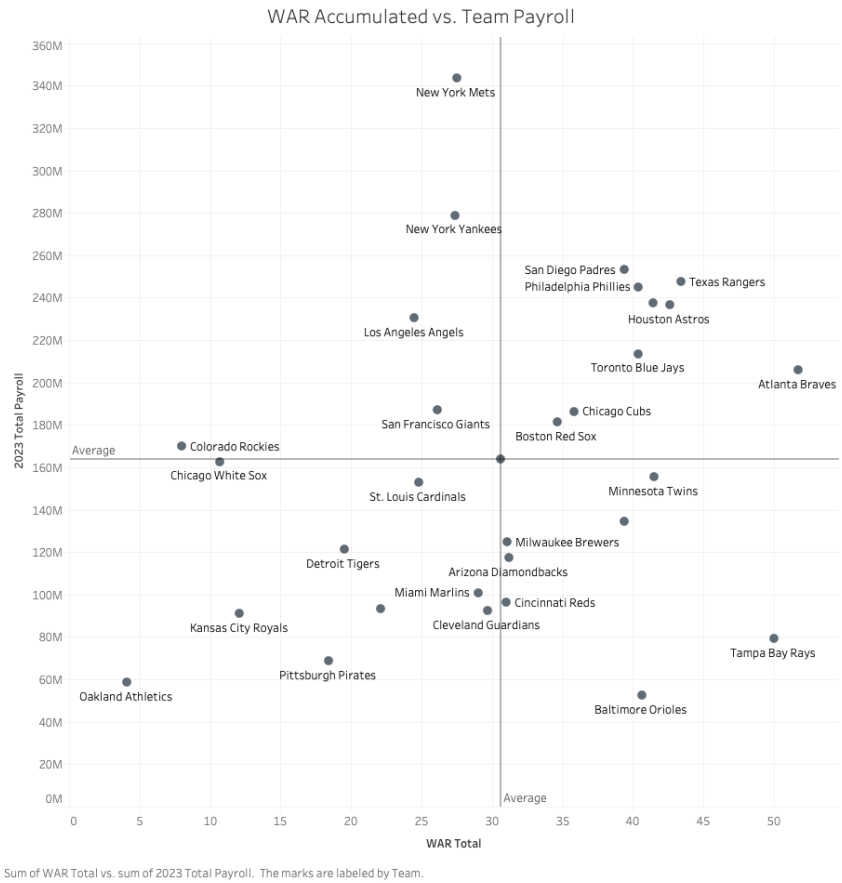


Figure 8 WAR Accumulated and Team Payroll

CONCLUSION

Our data visualization, including charts, graphs, and scatter plots, is designed to offer a comprehensive understanding of real-time game strategy and a player's value within the context of each baseball team's competitive goals in their respective divisions. These data visualizations offer valuable insights for teams seeking to make informed decisions and strategic choices in the evolving world of baseball. By studying the diverse array of metrics and trends we have presented, organizations can better navigate the complexities of player valuation, free agency, and in-game tactics and gain a deeper understanding of the multifaceted dynamics that influence the performance and cost of individual players and the team.

REFERENCES

Albert, J. (2010). Sabermetrics: The past, the present, and the future. Mathematics and Sports, 3–14. https://doi.org/10.5948/upo9781614442004.002

Barnes, S. L., & Bjarnadóttir, M. V. (2019). Baseball Analytics: Advancing to Prescriptive Analytics in the Major League Baseball Front Office. INFORMS Transactions on Education, 19(3), 143-145. https://doi.org/10.1287/ited.2018.0201ca

Baseball-Reference. (n.d.-a). 2023 MLB team position performance by wins above replacement. https://www.baseball-reference.com/leagues/team_compare.cgi?request=1&year=2023&lg=MLB&stat=WAR

Baseball-Reference. (n.d.-b). Major League batting year-by-year averages. https://www.baseball-reference.com/leagues/majors/bat.shtml

Conforti, C. M., Crotin, R. L., & Oseguera, J. (2022). Major League Draft Wars: An Analysis of wins above replacement in player selection. Journal of Sports Analytics, 8(1), 77–84. https://doi.org/10.3233/jsa-200586

Lee, J. S. (2022). Prediction of pitch type and location in baseball using ensemble model of deep neural networks. Journal of Sports Analytics, 8, 115-126. https://doi.org/10.3233/JSA-200559

Mizels, J., Erickson, B., & Chalmers, P. (2022). Current State of Data and Analytics Research in Baseball. In J. Dines and C. Camp (Section Editors), Current Reviews in Musculoskeletal Medicine, 15, 283–290. https://doi.org/10.1007/s12178-022-09695-8

MLB.com. (2023, October 30). Baseball metrics. https://www.mlb.com/glossary

Romero, F. P., Lozano-Murcia, C., Lopez-Gomez, J. A., Sanchez-Herrera, E. A., & Sanchez-Lopez, E. (2021, February 26). A data-driven approach to predicting the most valuable player in a game. Computational and Mathematical Methods in Medicine. https://doi.org/10.1002/cmm4.1155

Spotrac.com. (n.d.). MLB 2023 payroll tracker. https://www.spotrac.com/mlb/payroll/

Tango, T. (n.d.). History of the fielding. http://tangotiger.com/images/uploads/History_of_the_Fielding.pdf

Baseball Savant. (n.d.). Statcast custom leaderboards. https://baseballsavant.mlb.com/leaderboard/custom?year=2023&type=batter&filter=&sort=7&sortDir=desc&min=q&selections=k_percent%2Cbatting_avg%2Cslg_percent%2Con_base_percent%2Con_base_plus_slg%2Cba%2Cxslg%2Cwoba%2Cxwoba%2Cxobp%2Cxiso%2Cexit_velocity_avg%2Claunch_angle_avg%2Cbarrel_batted_rate%2C&chart=true&x=xba&y=xba&r=no&chartType=beeswarm

Baseball Savant. (n.d.-b). Statcast outs above average leaderboard. https://baseballsavant.mlb.com/leaderboard/outs_above_average?type=Fielder&startYear=2023&endYear=2023&split=yes&team=&range=year&min=q&pos=if&roles=&viz=hide