An Analysis of the 2025 MLB Statcast Dataset: Count Effects, Tunneling, and Pitch Sequencing

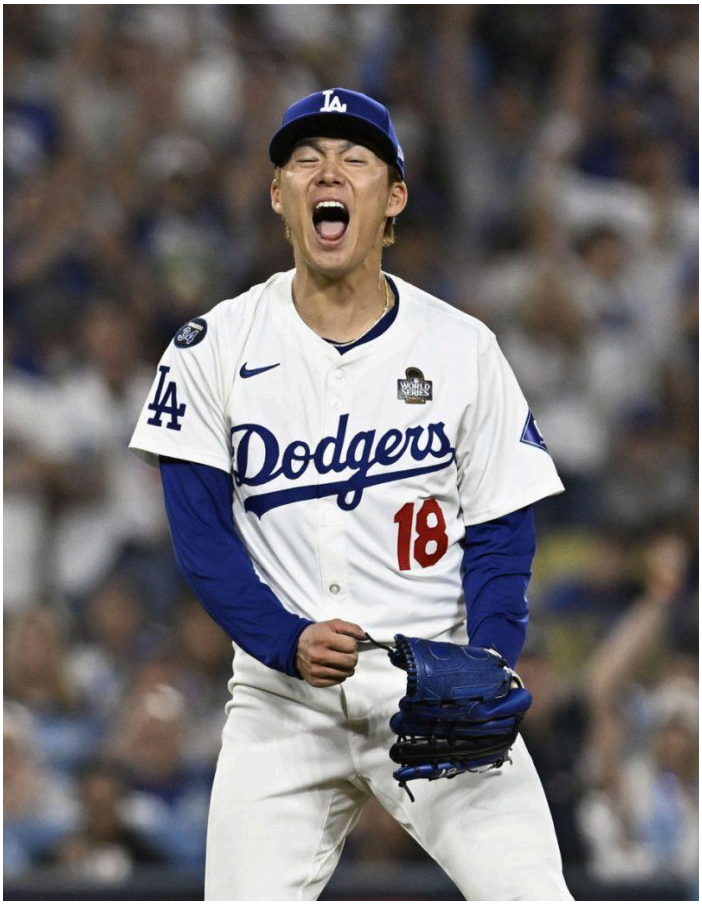
Leonardo C. Gentile

Williams College Statistics Department

Independent Study

Advisor: Professor Richard De Veaux

Spring 2026



Introduction

In every professional baseball league around the world, advances in pitch design through the use of tracking technology and game management along with the implementation of statistical models and game theory have revolutionized how front offices and coaching staffs approach each at-bat, inning, and game. The implementation of modern biomechanical data into statistical models, as well as specialized training programs and fatigue-management regimes, have enabled managers to curtail their starting pitchers' workloads in favor of high-effort, specialized relievers whose primary focus is striking out a particular handedness of batter. This allows managers to keep their starters' pitch counts low—preserving arm, shoulder, and elbow health for future outings—and to rotate in fresh, high-velocity arms to deal with the more threatening parts of the batting order.

Across baseball, pitch velocities are also increasing as pitchers' arsenals are deepening. Statcast is MLB's advanced tracking technology that uses high-speed cameras and Doppler radar installed in all 30 MLB stadiums to track each and every precise ball and player movement. While even those with little knowledge about baseball know about fastballs and curveballs, according to an article written by David Adler for MLB, "Statcast currently classifies 14 unique pitch types…[and the] number of big-league pitchers throwing five-plus, even six-plus of those pitch types, is at a record high for the decade-plus of Statcast tracking." [1] Currently in MLB, there are more than 125 pitchers with a pitch arsenal of six or more pitches, compared to a mere 50 pitchers just a decade prior. According to another article written by Christopher Boan for RotoWire using FanGraphs data on the six most-used pitches across MLB from 2002 to 2025,

[1] https://www.mlb.com/news/mlb-pitch-arsenals-are-bigger-than-ever-in-2025

the average pitch velocity among these pitches has increased by 5.98% since velocity tracking began.[2]

Another pitching strategy gaining traction around baseball is pitch tunneling: the concept of making different pitch types look identical out of the pitcher's hand and throughout their flight path until the ball is 23.8 feet from home plate—the point at which the batter must decide whether or not to swing—so as to mask pitch type and increase pitch deceptibility. When done effectively across multiple pitch types, pitch tunneling shrouds a hitter's swing decision in uncertainty, theoretically leading to more whiffs (swings-and-misses). As pitchers continue to make advancements in pitch design and delivery, increase their velocity, alternate their pitch mixes based on specific batter matchups and batter handedness, and produce increasingly distinct movement profiles across their pitch types with an added tunneling effect, the job of a modern-day hitter has never been more challenging. Batters now face an endless array of problems every time they step inside the batter's box.

Both pitchers and batters, however, stand to gain from understanding and modeling which pitch might be coming next. On the one hand, optimizing pitch sequencing with respect to location and pitch type could usher in an era of pitcher dominance and helpless hitters. While there are other external factors at play that are league-dependent—such as NPB's (Nippon Professional Baseball, Japan's highest level of baseball) current "dead ball era," caused by balls "not flying as well as in past years," which has stranded offenses at historic lows despite marginal pitching advancements—there remains a clear motive for either side to secure a

[2] https://www.rotowire.com/baseball/article/mlb-pitch-speed-and-usage-2002-to-2025-94262

sustained competitive advantage through the use of data optimizations.[3] [4] From a business standpoint, being able to accurately anticipate what pitch may be coming next given the previous pitches in an at-bat could lead to increased offensive output, more wins and, thus, more ticket sales as well as higher TV and media ratings.

Away from the privacy and confidentiality of MLB clubhouses and front offices, we can already see organizations making unconventional changes to the way they think about pitching. For the last nine games of the 2025 MLB season, the Miami Marlins began calling pitches from the dugout as opposed to the traditional approach of letting the catcher call the game, despite being locked into a tight Wild Card race with the New York Mets. This was an unconventional method that the Marlins had been experimenting with throughout their minor league system in 2025 with great success, as their systemic overhaul—led by President of Baseball Operations Peter Bendix (hired November 2023) and manager Clayton McCullough (hired November 2024)—continues to seek new analytical advantages with a focus on long-term success. Under this pitch calling approach in 2025, the Marlins' Triple-A affiliate, the Jacksonville Jumbo Shrimp, posted the lowest ERA (earned run average) across all of Triple-A by a significant margin at 3.73. They ranked second in WHIP (walks plus hits per innings pitched) at 1.281 and in RA9 (runs allowed per nine innings) at 4.19, and compiled the best record in the International League, culminating in a National Championship victory over the Scranton/Wilkes-Barre RailRiders (NYY Triple-A affiliate). Though many may argue that taking the game out of the pitcher's and catcher's hands will lead to miscommunication and a lack of trust between the

[3] https://japanball.com/japanese-baseball-news/npb-newsletter/pacific-league-midseason-check-in/?srsltid=AfmBOoqsLy_rcwJ-txaXjHc0BvXIZRCcez8efRWDvNy7bDGkrU-PELWZ

[4] https://jballallen.com/japans-dead-ball-is-likely-a-mistake/

battery-mates, only time will tell whether this pitch-calling method leads to actionable, on-field improvements for the Marlins and any teams that follow suit.

This independent study aims to answer numerous research questions regarding the effectiveness of the Marlins' new pitch-calling method, the effect of game state, tunneling, and pitch sequencing on many outcome-based metrics, while also providing case studies of specific pitches, game states, and pitchers along the way. First, we will assess the results of the Marlins' top-down dugout pitch-calling system by conducting a series of analyses on both pitch-level and sequence-level data from the Marlins' 2025 season, paying specific attention to their last nine games and comparing the relative outcomes of the two time frames (before and after the transition). We will then determine, given a specific count, location, or pitch type, what the optimal pitch to throw is—in terms of location and type—for maximizing the chances of a whiff. Whiff rate is one of my personal favorite statistics and one I rely on heavily when scouting international prospects, as it is a fantastic metric for measuring a pitcher's ability to beat their opponent based on "stuff" alone, without the influence of external factors such as defensive alignment and varying field dimensions. We will also consider the effect of the at-bat count on leverage, pitch selection (fastball vs. breaking ball usage), run value, launch angles, exit velocities, and quality of contact. Much of this count effects analysis will be based on and expand upon the work of Jim Albert in his 2023 blog titled "Count Effects." [5]

In the next analysis, we examine the effect of pitch tunneling to understand whether pitchers who cluster their arsenals more tightly at the swing-commitment point generate higher strikeout rates (K%), higher whiff rates (Whiff%), lower BABIP (Batting Average on Balls In Play), lower hard-hit rate (exit velocity greater than 95 mph), and how this trend has developed

[5] https://bayesball.github.io/BLOG/Count_Effects.html

over time. This analysis is inspired by the work done by Jon Roegele in his 2014 article "The Effects of Pitch Sequencing." [6] Lastly, we aim to understand the nuances of pitch sequencing with regard to consecutive pitch pairs, location, and pitch types, as discussed by John Z. Clay in his 2023 SABR article "Strategic Pitch Location: The Role of Two-Pitch Sequences in Pitching Success." [7] Additionally, we will compare the accuracy and predictive capabilities of a neural network with added context and game states against a first-order Markov chain with respect to predicting pitch-level outcomes such as whiff rate.

For this independent study, I have chosen Yoshinobu Yamamoto's 2025 season—and more specifically his September 6th game against the Baltimore Orioles at Camden Yards—as my case study of choice. Having spent the last two summers scouting South Korean and Japanese pitchers in Asia, Yamamoto represents the most successful pitching transition from Asia to MLB in recent memory. In 2025 alone, he won his second World Series and first World Series MVP with the Los Angeles Dodgers and was selected to both his first All-Star game and All-MLB First Team. On September 6th, 2025, Yamamoto pitched 8.2 nearly flawless innings against the Baltimore Orioles. He threw 112 pitches, recording ten strikeouts against just two walks and, notably, no hits. With one out remaining in his no-hit bid—and a 3-0 Dodger victory nearly in hand—Yamamoto dealt a 95 mph cutter to Orioles second baseman Jackson Holiday, who sent the pitch 362 feet to right field for a solo home run to spoil the gem. This was the last batter Yamamoto faced, and the Orioles went on to win in walk-off fashion, 4-3. Yamamoto achieved a game score of 86, tied for 17th best in MLB in 2025.

[6] https://tht.fangraphs.com/the-effects-of-pitch-sequencing/

[7] https://sabr.org/journal/article/strategic-pitch-location-the-role-of-two-pitch-sequences-in-pitching-success/

Literature Reviews

Baseball has long been an exciting source of inspiration for statisticians, and there has been a rich wealth of sabermetrics literature produced since the late 1970s detailing the statistical advancements that have transformed the sport. Sabermetrics is the empirical analysis of baseball through advanced analytics and a notable portion of its literature touches on count effects, pitch tunneling, and pitch sequencing. The three works most relevant to this independent study are "Count Effects" by Jim Albert (2023), "The Effects of Pitch Sequencing" by Jon Roegele (2014, published in The Hardball Times on FanGraphs), and "Strategic Pitch Location: The Role of Two-Pitch Sequences in Pitching Success" by John Z. Clay (2023, published in SABR's Baseball Research Journal).

"Count Effects" by Jim Albert

"Count Effects" by Jim Albert is a compilation of blog posts in which he explores the 2015 MLB Statcast pitch dataset in R, focusing on the importance of the ball-strike count at each point in an at-bat. He details many useful, "off the shelf" techniques and visualizations from his exploratory data analysis and poses numerous potential research questions.

In the first section, Albert creates a heat map and a tree diagram to visualize how the run value of a plate appearance changes for each ball-strike count, classifying counts as hitter-friendly (positive average run value), pitcher-friendly (negative average run value), or neutral. Upon speaking with Albert on the phone, he encouraged me to explore this idea further while accounting for inning, runners on base, and number of outs. He also defines leverage—or importance—of a specific count as the absolute difference in run value between adding another strike and adding another ball. He claims that 0-0 counts are low leverage and that counts such as 2-2 and 3-2 are high leverage. This distinction is even more pronounced now with the Automated Ball-Strike (ABS) Challenge System in use in MLB to start the 2026 season, which allows the umpire's original call of a pitch to be reviewed and overturned upon a challenge by either the batter, pitcher, or catcher—potentially switching a walk into a strikeout in extreme cases—and completely flipping the momentum of an inning. Albert claims that both better pitchers and hitters have great success in high-leverage situations. He later uses the tree approach to show how better pitchers spend more time in pitcher-friendly counts, with most of their count states associated with negative run values, whereas the best hitters in MLB display positive run values for most counts, reflecting extreme productivity in virtually all situations. Albert also models the value of plate appearances passing through specific counts using a Markov chain, with the end of the plate appearance (walk, strikeout, or ball in play) representing the absorbing states. He

compares the transition probabilities of 3-time Cy Young Award Winner and 2014 NL MVP Clayton Kershaw (LAD) against those of all 2015 MLB starters who made at least 30 starts, finding that Kershaw was exceptionally skilled at reaching and maintaining pitcher-friendly counts and achieving first-pitch strikes.

He explores the effect of the first pitch by comparing average run values, noting a clear advantage in an at-bat when a hitter reaches a 1-0 count as opposed to a 0-1 count. He remarks that first-swing behavior is highly player-dependent, and that in 2015, run value peaked when hitters swung at the first pitch between roughly 25% and 30% of the time. Albert also visualizes how hitters' launch angles and exit speeds increase and "hot zones" expand as the count becomes more hitter-friendly. Lastly, Albert demonstrates pitchers threw fastballs roughly 50–60% of the time in neutral counts, compared to 90% when behind in the count and below 50% when ahead.

Albert's article is particularly effective in its clear application of foundational methods to explain widely-accepted baseball logic. Reproducing these results eleven years later and identifying which trends have held and which new ones have emerged is especially valuable—and achievable thanks to Albert's foundation. What I found most useful was his visualizations (tree graphs, heat maps, scatterplots, contour maps) and provided code, as well as his ability to find statistical backing for well-established baseball intuitions, such as hitters being more likely to swing in both two-strike counts and in hitter-friendly counts.

"The Effects of Pitch Sequencing" by Jon Roegele

Roegele's "The Effects of Pitch Sequencing" centers on the effect of pitch tunneling and measuring the impact of pitch sequencing on swinging-strike percentage. He first introduces context for his paper and discusses the effect of specialized pitch design, advances in biomechanics, as well as the increased use of shorter, high-effort outings from relievers on pitch velocity and horizontal movement profiles.

He first explores how often the second pitch in a two-pitch sequence generates a whiff. To do this, he constructed a transition matrix mapping the distance between consecutive pitches at the swing/no-swing decision point (what I define as tunnel distance) against the distance between those same pitches as they cross home plate (what I define as plate distance) from all two-pitch sequences during the 2013 and 2014 MLB regular seasons. The matrix revealed that consecutive pitches that initially look quite different (large tunnel distance) but end up crossing home plate in the same location (small plate distance) can be effective. More importantly, Roegele found that the most effective two-pitch sequences— or the most reliable sources of swings-and-misses—are those in which consecutive pitches are relatively close together at the swing decision point (small tunnel distance) but arrive at very different locations at home plate (large plate distance). In other words, the less separation two pitches have at the swing decision point, the less plate separation they need to generate elevated swing-and-miss rates. Replicating these results with 2025 MLB data would be extremely valuable for pinpointing the continued effect of tunneling over the past decade.

Roegele also found that most two-pitch sequences falling within the high-whiff band involved two different pitch types—most commonly four-seam fastballs followed by sliders and vice versa. He observes that the last pitch in a sequence is more effective, on average, than when

considering each individual pitch in isolation, and that nearly all hitters perform worse against tunneled sequences compared to non-tunneled ones. Roegele leaves readers with a natural follow-up question: do pitchers whose entire arsenals cluster more tightly at the swing-commitment point generate higher strikeout rates or suppress contact more effectively?

"Strategic Pitch Location: The Role of Two-Pitch Sequences in Pitching Success" by John Z. Clay

In "Strategic Pitch Location: The Role of Two-Pitch Sequences in Pitching Success," Clay refers to the sequence of strategic decisions a pitcher makes throughout an at-bat as sequential pitch behavior. His paper explores the process through which pitchers use strategic decisions about pitch location within the strike zone to retire batters. It is almost intuitive that a key factor in determining future pitch selection is the success of previous behaviors within a defined time frame—such as earlier in the same game or in prior at-bats against a particular hitter. The more success a behavior has generated within a given context, the more likely a pitcher is to reuse it in future situations of similar circumstances. Clay used pitch-by-pitch data for every American League pitcher who threw more than 1,450 pitches during the 2019 MLB season, resulting in a final sample of 87 pitchers, primarily starters.

Clay first quantifies two-pitch location-sequence behaviors for each pitcher using a 14×14 transition matrix (196 total transition variables) containing the probability for each possible state-to-state transition and then partitions the pitchers into five groups based on similarities in those behaviors using k-means, hierarchical, and spectral clustering algorithms. He then tests for differences across 19 performance metrics using one-way ANOVAs, a statistical test used to determine the difference between the means of three or more independent groups.

Among most performance metrics, no statistically significant differences were found between the five groups. However, I believe having no statistically significant results is different from having no meaningful on-field implications, especially when considering how statistically powered each analysis might be. However, significant differences did emerge across seven categories, including spin rate, pitch velocity, average exit velocity allowed, batting average

against (BAA), expected batting average against (xBA), ERA+, and zone contact rate. Significant differences in spin rate and velocity suggest that pitching mechanics like pitch grip and arm slot may influence underlying sequencing patterns. Additionally, significant differences in average-exit-velocity allowed imply that certain location sequencing patterns may be more effective at limiting hard contact, such as starting an at-bat off in the zone and finishing it low-and-away. Differences in BAA, xBA, and ERA+ suggest that two-pitch location sequencing may be directly associated with run prevention—hugely important information for pitching staffs and front offices.

Clay then focuses on Group 1 (worst) and Group 4 (best). Group 1 performed poorly relative to the others, having the fewest career accolades as well as significantly lower spin rates, lower velocity, and higher batting averages against. Group 1 also displayed the most predictable state transition behavior among all 5 groups, having the lowest probability of using over 33% of the possible transitions. Group 4 demonstrated the highest performance with the most career accolades and the highest career earnings as well as displayed the highest probability for a wide range of possible state transitions. This suggests that these pitchers varied their two-pitch location sequences more than other groups—likely making their behavior less predictable to hitters and contributing to greater success.

Finally, Clay conducts a case study on Justin Verlander's third career no-hitter (September 1st, 2019). Verlander threw 120 pitches, producing 148 total state transitions distributed across 83 unique sequences. Notably, 52 of the 83 sequences were used only once, and no sequence was used more than five times. Verlander effectively avoided pitch-location predictability, which likely contributed to the struggles of the Toronto Blue Jays that afternoon. Of the 28 batters he faced, the first pitch to 13 of them was a strike in Location 3, 11, or 12, and 16 of the 28 plate

appearances ended in Locations 5, 9, or 14. During his no-hitter, Verlander demonstrated situational variance—tactically unpredictable in some states while reusing transitions in other situations, continuing behaviors that had already proven successful. This balance between unpredictability and selective repetition appears to have been key. In my own analysis, I will attempt to replicate and extend many of Clay's results using 2025 Statcast data and explore whether his findings have maintained their relevance since 2019.

Methodology

This independent study analyzes pitch-level data from the 2025 MLB season obtained from Baseball Savant's Statcast system. The raw dataset contains pitch-by-pitch records for all MLB games in 2025 (737,186 total pitches), each recorded with pitch type, velocity, spin rate, movement vectors, plate coordinates, pitch outcome, launch angle, exit velocity, and event descriptions including the pitcher, batter, teams, inning, score, and many other variables. The four analyses that follow—the Miami Marlins dugout pitch-calling experiment, count effects, pitch tunneling, and pitch sequencing—all utilize and build upon this dataset while applying distinct filtering criteria, feature engineering steps, and statistical methods appropriate to their respective research questions.

Miami Marlins Analysis

The Marlins analysis evaluates the effectiveness of the team's experiment of calling pitches from the dugout during the final nine games of the 2025 regular season (September 19th onward). Because Statcast does not explicitly include a team identifier at the pitch level, Marlins pitches were isolated using inning-half logic: pitches thrown in the top half of the inning were attributed to the home team's pitching staff and pitches thrown in the bottom half of the inning were attributed to the away team's pitching staff. I filtered for every inning in which the Marlins were pitching and was able to extract all Marlins pitches across the 2025 season. I then partitioned the study period into two windows: games before September 19th (153 games, 23,369 pitches) and games on or after September 19th (9 games, 1,208 pitches), when the experiment began. This allowed for direct comparison of pitching patterns before and after implementation. I included only Marlins pitchers who threw more than 200 pitches during the 2025 season and appeared in the nine-game window, which helped remove position players who were occasionally brought in to pitch in blowouts, such as Javier Sanoja. While it is true that a pitcher could have debuted in the 9-game window and thrown less than 200 pitches (excluding them), anyone with 200 pitches across a full season is already a very marginal arm and their inclusion likely would not be meaningful. Additionally, I only considered pitch types with more than 100 pitches in both the before and after period to make the sample sizes reasonably large enough to detect meaningful differences.

Pitch mix proportions were computed for each pitcher and each window to quantify arsenal-level shifts, and a faceted bar chart of pitch usage by pitcher was produced to visualize the changing composition of arsenals. Team-wide summary statistics—average delta run expectancy per pitch, team-wide whiff rate, hard-hit rate, and batting average against—were also

calculated for each window. Additionally, a box plot of per-pitch run value in both periods was produced, along with a table and bar chart of average run value broken down by pitch type and period. Delta run expectancy is defined as the exact change in a team's expected runs in a given inning before and after a specific pitch. Given that over 730,000 pitches were thrown over the 2025 season and runs scored on only a fraction of those pitches, the impact of any single pitch on run expectancy is often very small. A two-sample t-test on average delta run expectancy per pitch was used to determine whether the post-September 19th improvement in run value was statistically distinguishable from random chance.

Next, a sequential pitch dataset was built by examining consecutive pitch pairs. Marlins pitch sequencing behavior was analyzed using transition matrices. For each window in the matrix, consecutive pitch pairs within each plate appearance were extracted and the empirical probability of transitioning from one pitch type to another was computed. The two transition matrices were compared to identify which transitions became more or less common during the post-September 19th period. These results were printed in a table and organized in a bar chart showing the 15 pitch sequences that changed the most. These changes were further contextualized by examining pitch-type usage proportions broken down by count (0-0, hitter-friendly, pitcher-friendly, neutral), allowing identification of specific count situations in which dugout strategy most noticeably altered pitch selection.

Lastly, to control for confounding variables, a multinomial logistic regression was fit to predict the probability of each pitch outcome (whiff, called strike, foul, ball in play, or ball) on each pitch as a function of pitch type, velocity, horizontal and vertical movement, ball count, strike count, and a binary indicator for the dugout-calling window (1 = post-September 19th, 0 = prior). Ball was used as the reference category. I also computed a Bonferroni-adjusted p-value to

control the error rate across all coefficients in the model. With 5 outcome variables and numerous predictors, we will likely produce many coefficients, some of which will be significant by random chance, and by applying the correction, we can ensure that any surviving predictor represents strong evidence of a before/after difference. If any predictors were originally significant and then dropped out after the correction then they were likely noise inflated by multiple comparisons. The key output of this model is the risk ratio for the period indicator on the whiff outcome. An OLS regression on run value using the same covariates was also fit, with particular attention to the magnitude and direction of the before/after coefficient.

Count Effects Analysis

The count effects analysis explores, given a specific count, location, or pitch type, what the optimal pitch to throw is for maximizing the chances of a whiff (a swing and a miss). We will also consider the effect of the at-bat count on pitch selection (fastball vs. breaking ball usage), run value, launch angles, exit velocities, and quality of contact. To begin, I computed three necessary datasets: average run expectancy by count state, batted-ball contact stats by count state, and a modeling dataset that bins pitch locations into zones and encodes ball and strike counts as integers. For visualization purposes, delta run expectancy values are scaled by 100 and displayed as runs per 100 pitches to improve readability and interpretability. For regression models and statistical tests, we still use the raw delta run expectancy values.

The first visualization replicates Jim Albert's count tree graph for the entire 2025 MLB season, where each node is a count state, the y-axis is the mean run value of plate appearances passing through it, edges connect counts via ball/strike transitions, and edge thickness is proportional to visit frequency (0-1, 1-0 thickest and 3-2 thinnest). We can interpret these tree graphs such that moving from any count down-and-right (more strikes) improves the pitcher's situation and moving left (more balls) helps the hitter. I then computed the leverage of each count as $|RV(+strike) - RV(+ball)|$, with higher leverage indicating a larger swing in expected outcome. For most counts, this involved moving from one count to another and referencing the run expectancy; however for boundary counts (two strikes or three balls) and most notably full counts (3-2), terminal run values from strikeout/walk outcomes were used.

I then repeated this leverage calculation segmented by three game-state variables: runners on base, outs, and inning groupings. Three heat maps for each runners-on-base permutation show count leverage broken down by each of these variables to try and determine which game

situations make our standard count more consequential. Following this, I computed a weighted leverage metric for all pitchers with 950 or more pitches during the 2025 season—defined as the sum of leverage times delta run expectancy divided by the sum of leverage for each count. I also defined a clutch metric computed by subtracting the weighted leverage run value from the average run value, to try and identify whether pitchers perform differently when counts matter more. A pitcher with a negative clutch score performs better in higher-leverage counts (meaning his run expectancy is very low). I then created bar charts for the top 40 and bottom 40 pitchers by leverage-weighted run value and clutch score, along with a sorted table of all qualified pitchers including their standard run value, leverage-weighted run value, clutch score, leverage hard-hit rate, and whiff rate.

Next, I recreated Albert's bar chart and heat map of average run value across all 12 ball-strike counts, colored by count type (hitter-friendly, pitcher-friendly, or neutral). I also reproduced Albert's tree visualization, this time for Yoshinobu Yamamoto (LAD) and Bobby Witt Jr. (KC), where each node is a count state and the edge thickness is proportional to how often plate appearances visited that count. I then created seven bar charts—one each for the seven most common pitch types (four-seam fastball, changeup, curveball, cutter, sinker, slider, sweeper)—showing when each pitch is most commonly thrown by count.

I also produced a dedicated fastball-percentage chart across all 12 counts league-wide, combining four-seam fastballs, cutters, and sinkers under one umbrella (Statcast now designates two-seam fastballs as sinkers), as fastballs are the most commonly thrown pitch and thus have a special importance in terms of understanding when they are most likely to be thrown. Next, I took a random subset of pitchers who threw more than 2,000 pitches in 2025 and replicated Albert's fastball-by-count analysis both at the group level and individually, using a connected

line plot per pitcher colored by team. I also created a line chart of whiff rates by count type and pitch type as well as a 3×3 whiff rate by zone and count type heat map. A logistic regression (binomial GLM) predicting whiff probability from count, pitch type, zone, and velocity was fit and statistically significant odds ratios reported. Lastly, three contact-quality bar plots were produced (hard-hit rate, average launch angle, and average exit velocity by count state), along with a 2D density plot of launch angle versus exit velocity faceted by count type.

Pitch Tunneling Analysis

The pitch tunneling analysis contains the most computationally intensive data preparation in this study. To calculate each pitch's (x, z) coordinates 23.8 feet from home plate—the tunneling point—I use the standard kinematic equation $position = initial\,position + velocity \times t + ½ \times acceleration \times t^2$, drawing on the Statcast variables of release_pos_x, release_pos_z, vx0, vz0, ax, and az that are available for each pitch where *t* represents the time that has elapsed since the ball was at its initial position. Unlike Roegele, who uses 21 feet from release or around 33 feet from the plate when assuming a 6 foot extension off the mound, we will use the 23.8 feet from home plate that was identified by Jeff Long, Jonathan Judge, and Harry Pavlidis at Baseball Prospectus, "based on 175 milliseconds and a league-average fastball." [8] [9] The Baseball Prospectus has a more concrete computation behind their 23.8 designation than Roegele, and we will thus consider how much of a difference this change will make in our findings.

A natural question to consider is the effect of arm slot and spin rate and how they might interplay with the tunneling effect, as both affect the way the batter perceives each pitch. The position formula traces the physical flight path of the ball through the air on the way to home plate, and the acceleration fields that it uses (ax, az) are not just gravity, but also the net aerodynamic accelerations measured at release which incorporate the Magnus effect and drag. This is to say that spin rate is baked into ax and az. The Magnus effect is a phenomenon where a spinning object, such as a four-seam fastball, travels through air curves away from its straight path because the spin creates a high-pressure zone on one side and a low-pressure force on the

[8] https://tht.fangraphs.com/pitch-tunneling-is-it-real-and-how-do-pitchers-actually-pitch/
[9] https://www.baseballprospectus.com/news/article/31030/prospectus-feature-introducing-pitch-tunnels/

other side, creating a force towards the low-pressure side. Fastballs with high spin rate tend to have a "rising" effect relative to the batter's expectation, even though it is of course dropping—just less than the batter may expect.

As for the arm slot, Statcast has an arm angle variable that we can include as a predictor in our later analysis. Arm angle is calculated using Statcast cameras to determine the geometric angle of a pitcher's arm upon release. A pitcher like Trey Yesavage (TOR) has a very over-the-top release and thus an arm slot of 64 degrees, whereas his teammate Tyler Rodgers (TOR) is a submariner who has an arm slot of -61 degrees, below his knees. Two pitches thrown from different arm slots will look different to the batter almost immediately and make true tunneling impossible regardless of the math, so it is important for pitchers to maintain the same arm slot across all of their pitches.

I first paired consecutive pitches within each plate appearance, and for each pair, tunnel distance (separation at the tunnel point, in inches), plate distance (separation at the plate, in inches), and tunnel ratio (plate distance divided by tunnel distance) were computed. Then, pitcher-level tunnel scores and outcome statistics (K%, Whiff%, BABIP, hard-hit rate) are also computed, along with a pair-level, league-wide summary of the 15 tightest pitch-type transitions. The first visualization I made was a faceted histogram showing the distribution of mean tunnel distance, mean plate distance, and mean tunnel ratio among qualified starters.

Arguably the most important visualization in this independent study is the "Roegele Matrix"—a replication of Jon Roegele's tunnel distance vs. plate distance matrix from 2014. This matrix bins pitch pairs by tunnel and plate distance, coloring each cell by whiff rate on the second pitch. After this, my first instinct was to create scatterplots of mean tunnel ratio, mean tunnel distance, and mean plate distance against average release speed per pitcher colored by

whiff rate to get a sense of the interaction between pitcher's velocity and their ability to tunnel their pitches. I then made scatterplots for each of the mean tunnel ratio, mean tunnel distance, and mean plate distance against BABIP, K%, whiff rate, and hard hit rate for all qualified starters (2,000+ pitches), each with a linear trend line.

I also created a plot of whiff rate on the second pitch as a function of tunnel ratio, with point size proportional to sample size, and a bar chart of the 15 pitch-type pair combinations with the lowest mean tunnel distance, color-coded by whiff rate. A bar chart of the top 20 pitchers by mean tunnel ratio shaded by whiff rate was also created. Lastly, separate OLS regression models were fit predicting K%, BABIP, hard-hit rate, and whiff rate from mean tunnel distance, mean plate distance, and mean tunnel ratio (with extended models including arm slot and spin rate).

Lastly, a case study of Yoshinobu Yamamoto's tunneling profile was conducted. This included a scatterplot of his pitch pairs in the tunnel vs. plate distance space colored by whiff outcome, a by-pitch-pair summary of his tunnel and plate distances, a density heat map of where his pitches pass through the swing decision point faceted by pitch type, as well as a comparison of his mean tunnel distance and ratio against the league distribution. Yamamoto was also compared against two other pitchers—Jacob deGrom (TEX) and Erick Fedde (STL/ATL/MIL) who are considered elite and poor tunnelers, respectively—using a summary table of outcome statistics, density plots of tunnel distance distributions, and faceted scatterplots of tunnel vs. plate distance colored by whiff rate. Additionally, Erick Fedde recently spent the 2023 season in the Korean Baseball Organization (KBO) with the NC Dinos after a rough start to his career in MLB. He revitalized his career in South Korea, being named an All-Star, posting a 20-6 record with 209 strikeouts, winning the pitching triple crown, and winning the Choi Dong-won award

(KBO's equivalent to the Cy Young Award, as well as KBO MVP, before signing an MLB deal with the St. Louis Cardinals in 2024.

Pitch Sequencing Analysis

My final analysis explores pitch sequencing, beginning with a transition matrix setup. Pitches are mapped to the 14 Statcast zones (a 3×3 strike zone grid plus four chase zones), and consecutive pitches within each plate appearance are paired as (from_zone, to_zone) tuples. A 14×14 transition probability matrix is constructed for each qualified pitcher (2,000+ pitches), and these matrices were then pivoted into a feature matrix with 196 columns for clustering. Compared to Clay's analysis, I use a higher pitch threshold (2,000 as opposed to Clay's 1,450 pitches) and include the entire MLB rather than only the American League, as I was primarily interested in starting pitchers who achieved more than 30 starts.

The first visualization is a 15×15 league-wide mean pitch location transition probability matrix across all qualified starters (including an extra row and column for the first pitch of the plate appearance). The transition probabilities represent the probability of a pitch being in that zone given that the last pitch came from some other (or the same zone) and is shaded according to the frequency with which they occur. K-means clustering is applied for k = 2 through k = 8, producing an elbow plot with an inflection point at k = 5. All qualified pitchers are then assigned to one of five clusters, and one-way ANOVAs are run across the clusters on seven performance metrics (spin rate, velocity, exit velocity, BAA, K%, whiff rate, hard-hit rate). Bar charts with ±1 standard error bars and a cluster summary table are also produced. Box plots of each cluster's Shannon entropy—which quantifies sequencing unpredictability—are presented, with another ANOVA testing for cluster differences in entropy. An example of maximum entropy is something totally random like a uniform distribution where each observation has an equal chance of being chosen, and something with entropy near 0 is a completely predictable outcome, like a pitcher throwing in the same location every time.

Following Clay, I then compare the best and worst clusters by BAA, tabulating average velocity, spin rate, BAA, K%, and exit velocity, and produce a velocity vs. BAA scatter plot colored by cluster. Pitch-type profiles (velocity, spin rate, movement, and whiff rate) for pitchers in the two extreme clusters are also tabulated.

In his article, Clay characterizes each pitcher using a transition matrix derived from their two-pitch location sequence behavior using a discrete time Markov chain. In this analysis, I fit two logistic regression models to predict the whiff outcome: one first order model using the current from-zone → to-zone transition, as well as a second-order model that also incorporates the previous transition in the sequence. I then compared the models using a likelihood ratio test and made a table that reports log-likelihoods, degrees of freedom, AUC on held-out test sets for both models.

To address Clay's recommendation that future studies employ neural networks rather than first-order Markov chains, an 80/20 train-test split is constructed. Features included: from-zone, to-zone, pitch number in the at-bat, ball count, strike count, first-order Markov transition probability for that zone pair, and pitch type. A single-hidden-layer neural network is trained to predict whether the next pitch will be a whiff, and evaluated against the Markov baseline using accuracy and AUC, with a ROC curve comparing both models. A second neural network with extended predictors—velocity, movement, batter handedness, previous pitch type, pitch number in game, spin rate, and arm slot angle—is also evaluated on the 20% holdout test set. At this point in the analysis, I felt it was appropriate to check the whiff class balance for the entire dataset, Yamamoto’s 2025 season, as well as his September 6th game against the Orioles.

Because whiffs are relatively rare events, the training data is heavily imbalanced with many more non-whiffs than whiffs. To solve this problem, we ran a third neural network with non-whiff rows undersampled down to match the whiff count using the same extended parameters so the models are directly comparable. While second-order Markov chains encode more information than first-order Markov chains and are also simpler than neural networks, they still exclude a lot of interesting context behind each pitch. A minimal improvement from first to second order would suggest that simply extending the chain is not the right course of action.

To evaluate the models, we made ROC curves as well as a league-wide permutation importance table that randomly shuffles each feature 10 times, re-predicting on the shuffled data each time and computing the AUC drop. Because neural networks are black box models, which produce predictions with no explanations, we can use permutation importance to help shed light on which predictors actually matter. If a feature is genuinely important then removing it from the model should hurt performance, however if removing it made no difference then the feature is not meaningfully contributing. Reporting importance across ten shuffles with standard deviations gives us a sense of whether the rankings are stable or noisy, and allow us to make specific interpretive claims in the future.

Much like Clay's case study of Justin Verlander's third no-hitter, a case study of Yamamoto's September 6th game against the Baltimore Orioles is conducted. Transition frequency tables are constructed for Yamamoto's pitch-type pairs across the full 2025 season, before September 6th, after September 6th, on September 6th alone, and split by batter handedness. These are displayed as raw counts, with percentage tables, bar plots, as well as before/after comparison plots with direction indicators. Location zone labels (High/Mid/Low x Inside/Middle/Outside) are appended to each pitch record and sequences are plotted as

"PitchType @ Zone → PitchType @ Zone" pairs. For the September 6th game itself, a game-specific 15×15 zone transition matrix and heat map are built, unpredictability metrics are computed and benchmarked against Verlander's 2019 no-hitter, and first-pitch and terminal-pitch zone distributions are visualized.

To score each pitch in the game by its whiff probability, I fit a pitcher-specific logistic regression model that was trained exclusively on Yamamoto's 2025 season as opposed to our entire dataset. This decision was motivated by a problem of unequal distributions: the league-wide neural network from earlier in the analysis was trained on over 700,000 pitches with an 11% whiff rate whereas Yamamoto had a 20% whiff rate during his September 6th game. As a result, applying a model calibrated to 11% whiffs to a game with 20% whiffs produced systemically underestimated probabilities. Like the neural network, the pitcher-specific logistic regression uses from-zone, to-zone, ball count, strike count, pitch type, release speed, horizontal movement, and vertical movement as predictors and is trained on all Yamamoto's 2025 pitches with NAs removed. I also used the zone + count league-wide neural network model as a direct comparison to the logistic regression and the first-order Markov chain. A table reports the predicted probability of each model against the actual game whiff rate, and ROC curves compare all three models on the September 6th holdout game. The logistic regression coefficients are also reported as odds ratios to provide a fully interpretable account of which sequences and characteristics drive whiff probability for Yamamoto specifically. This methodology could also easily be applied to any pitcher in MLB.

Findings and Results

Miami Marlins Analysis

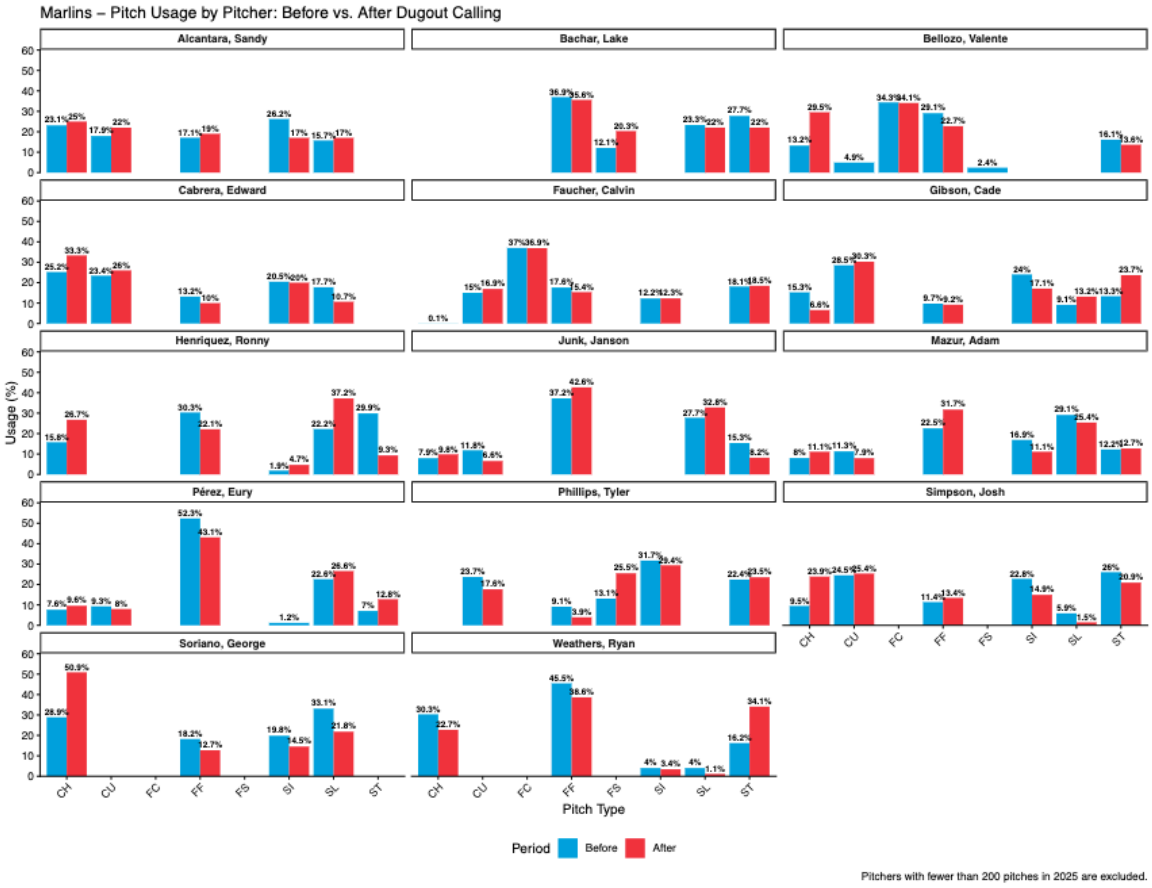


Figure 1: Pitch type usage for each Marlins pitcher before (blue) and after (red) the dugout pitch-calling change. Pitchers with fewer than 200 pitches excluded.

Across the 14 Marlins pitchers who threw at least 200 pitches during the 2025 season and appeared in the post-September 19th window, pitch mix shifts were visible at the individual level. 2022 NL Cy Young Award winner Sandy Alcantara notably reduced his sinker usage by over 9 percentage points after the switch, opting instead to throw more curveballs. Ryan Weathers, who was acquired by the New York Yankees in January, reduced both his four-seam fastball and changeup usage while more than doubling his sweeper usage—a pitch that

complements his arsenal very well with a completely different horizontal movement profile. In 2025, Weathers held batters to a .174 batting average and 39.1% whiff rate with his sweeper, per Baseball Savant. Weathers is off to an excellent start to the 2026 season as the Yankees have wisely opted to continue building on this late-season change made by the Marlins organization. The first Marlins pitcher to start a game under the new system was Janson Junk, who pitched seven innings of one-run ball against the Rangers on September 19th and saw a drop in both his curveball and sweeper usage in favor of his fastball and slider. Edward Cabrera, a rising star recently traded to the Chicago Cubs during the 2026 offseason, reduced his slider usage by about 8 percentage points in favor of his one-of-a-kind changeup, which often reaches 95–97 mph. These individual-level changes suggest that the Marlins' coaching staff was actively reshaping each pitcher's arsenal on an individualized basis rather than applying a sweeping team-wide strategy.

Across all Marlins pitchers during this nine-game window, the raw numbers tell an encouraging story. Compared to the rest of the season, the Marlins saw their team-wide whiff rate increase from 22.08% to 26.46% and their batting average against drop from .250 to .221. Conversely, their hard-hit rate ticked up incrementally from 27.54% to 28.94%. Average run value per pitch moved slightly in a favorable direction for the Marlins, from 0.0014 to 0.0006 (lower is better for pitchers). When considering individual pitch types with more than 100 pitches thrown in the post-change window, curveballs had a lower average run value after the transition (more effective), while sweepers, sliders, changeups, four-seam fastballs, and sinkers had a higher average run value (less effective), as reflected in the table and bar chart.

A two-sample t-test on run value per pitch returned a p-value of 0.913, confirming that the overall difference was not statistically distinguishable from random chance. This was an

expected result: nine games and approximately 1,200 pitches—compared to 153 games and 23,369 pitches—provides very limited statistical power to detect small true effects.

A multinomial logistic regression was fit to predict the probability of each pitch outcome—whiff, called strike, foul, ball in play, or ball (reference)—while controlling for pitch type, velocity, horizontal and vertical movement, ball count, and strike count. The risk ratio for the dugout-calling indicator on the whiff outcome was 1.12 (95% CI: 0.92–1.36, $p = 0.247$). This falls short of the conventional $\alpha = 0.05$ threshold for significance (not surviving the Bonferroni correction either) and the confidence interval crosses 1 on the low end. The direction is nonetheless positive and consistent with the raw whiff rate increase observed in the summary statistics, suggesting a plausible but unconfirmed effect of approximately 12% higher relative risk of inducing a whiff during the dugout-calling stretch compared to a ball outcome, after accounting for pitch type and count. The $p = 0.247$ result is thus more hypothesis-generating than it is confirmatory. An OLS run value regression produced a similarly directional but nonsignificant result (coefficient on before period: 0.0007, $p = 0.914$).

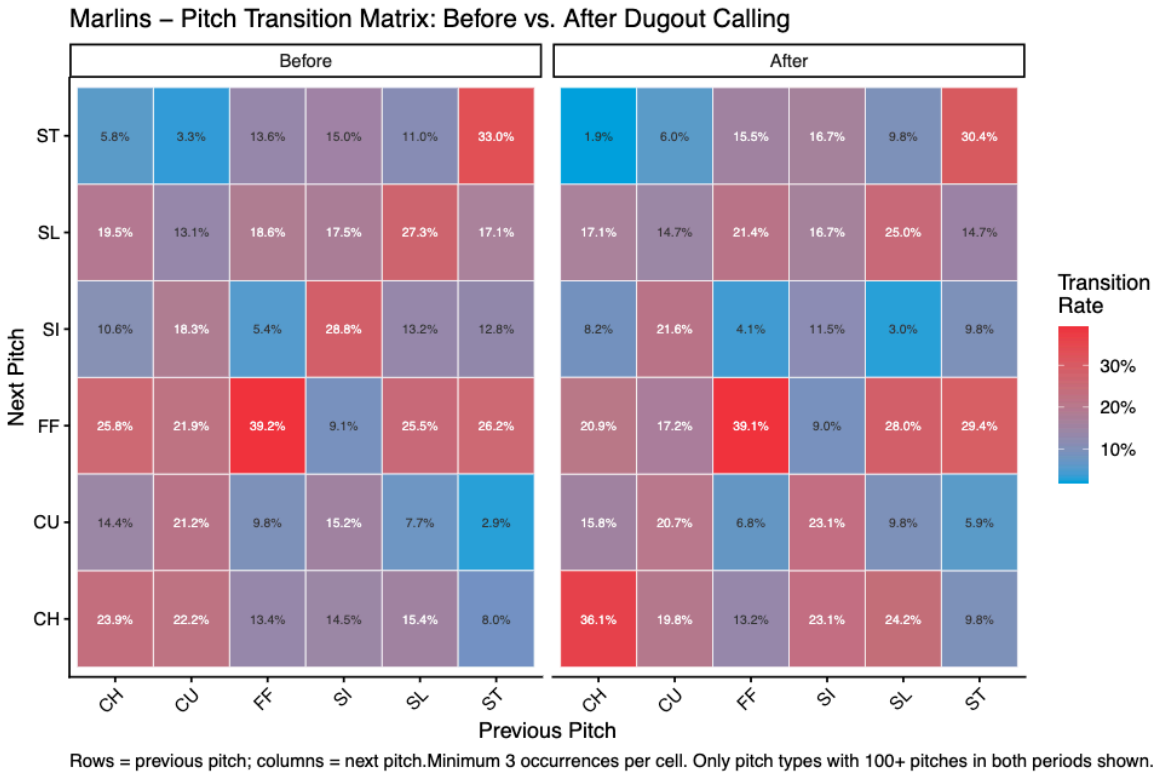


Figure 5: Pitch transition probability matrix for Marlins pitchers before and after the dugout calling change. Color intensity reflects transition rate. Only pitch types with 100+ pitches in both periods shown.

The transition matrix reveals meaningful changes in how pitches were sequenced after September 19th. The data suggests the dugout was making deliberate, situation-specific decisions about what pitches and zones to repeat and which to use more sporadically. The most dramatic sequence increases were seen with change-up-to-change-up (+12.2%) and slider-to-change-up (+8.8%), indicating that the Marlins began setting up change-ups far more aggressively with other change-ups and sliders, indicating a shift toward sequences that mask pitch type and mess with hitter's timing. Sinker-to-change-up (+8.6%) and sinker-to-curveball (+7.9%) both increased sharply as well, suggesting a deliberate effort to sequence pitches that play well off each other's movement and location profiles. On the flip side, the sinker-to-sinker transition dropped by 17.3% (the largest decrease in the Marlins dataset), with the

slider-to-sinker also seeing a decrease of 10.1%. This implies that the coaching staff was steering Marlins' pitchers away from sequences that leave their sinker vulnerable to hard contact. This is emphasized in the per-pitcher outcome plot which shows the average run value per pitch by qualified Marlins pitchers before and after the transition being completely different across the board. When considering count-level plots, pitcher-friendly counts saw a noticeable increase in the number of change-ups thrown, signaling that the Marlin's front office may be targeting hitter's timing as a way of generating whiffs. Following the switch, we see the most pronounced usage shifts in 0-0 and 2-2 counts, suggesting the dugout strategy particularly focused on shaping at-bat openings and two-strike situations—where sequencing logic has the most leverage and impact.

This analysis confirms that the dugout system produced no detectable negative outcomes on pitch results and positive directionality in numerous important pitch metrics such as run value and whiff rate. It also signaled that the changes to pitch repertoire were noticeable at the individual level. It will be interesting to see if these changes persist into next season, even with certain players on new teams and new additions to the roster, such as closer Pete Fairbanks, starter Chris Paddack, and reliever Osvaldo Bido.

Count Effects Analysis

For the count effects analysis, I created a count label, count type, and an indicator for whether a pitch resulted in a whiff, swing, or hard hit (launch speed ≥ 95 mph) for each observation in the dataset.

Run Value by Count

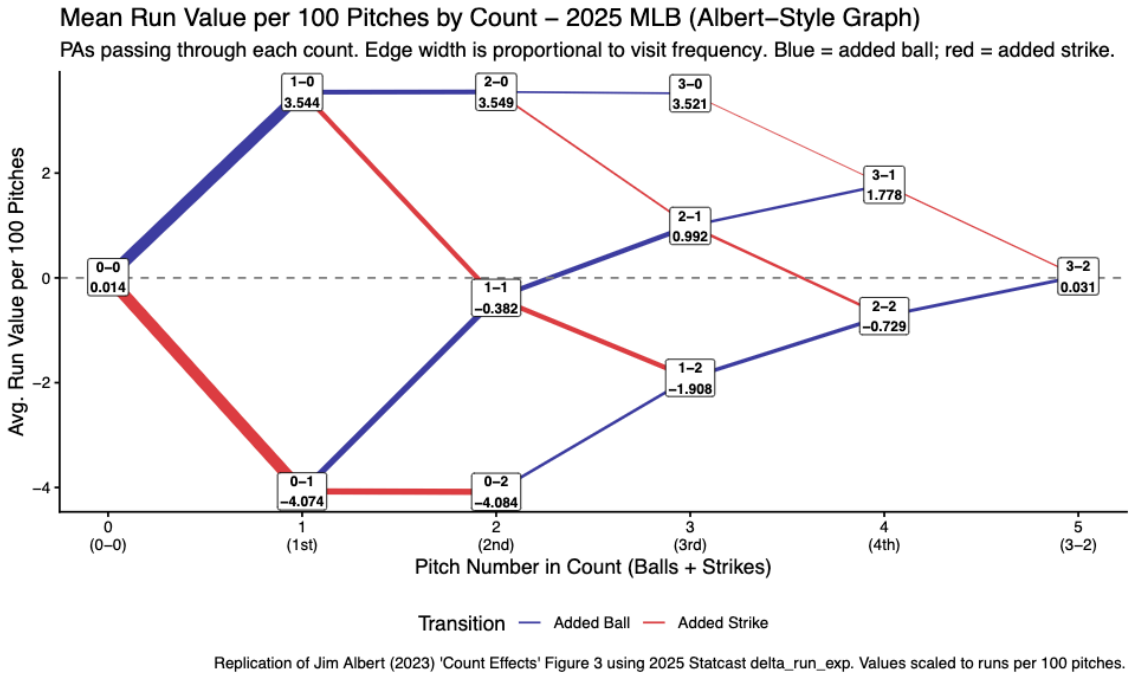


Figure 12: Mean run value per 100 pitches for all 2025 MLB plate appearances passing through each ball-strike count. Each node is a count state; y-axis shows average run value of PAs that visited that count. Blue edges = added ball; red edges = added strike. Line width is proportional to the fraction of PAs visiting each count. Replicates Albert (2023) Figure using 2025 Statcast data.

The first visualizations I made analyzed leverage and included a tree graph of the mean delta run expectancy per 100 pitches for all 2025 MLB plate appearances passing through each ball-strike count as well as the mean true leverage by count, inning, and number of outs, with a different heat map for each permutation of runners on base. The tree graph showed exactly what we would expect: there is a positive delta run expectancy per 100 pitches for counts with more

balls than strikes across MLB and a negative delta run expectancy per 100 pitches for even counts as well as counts with more strikes than balls.

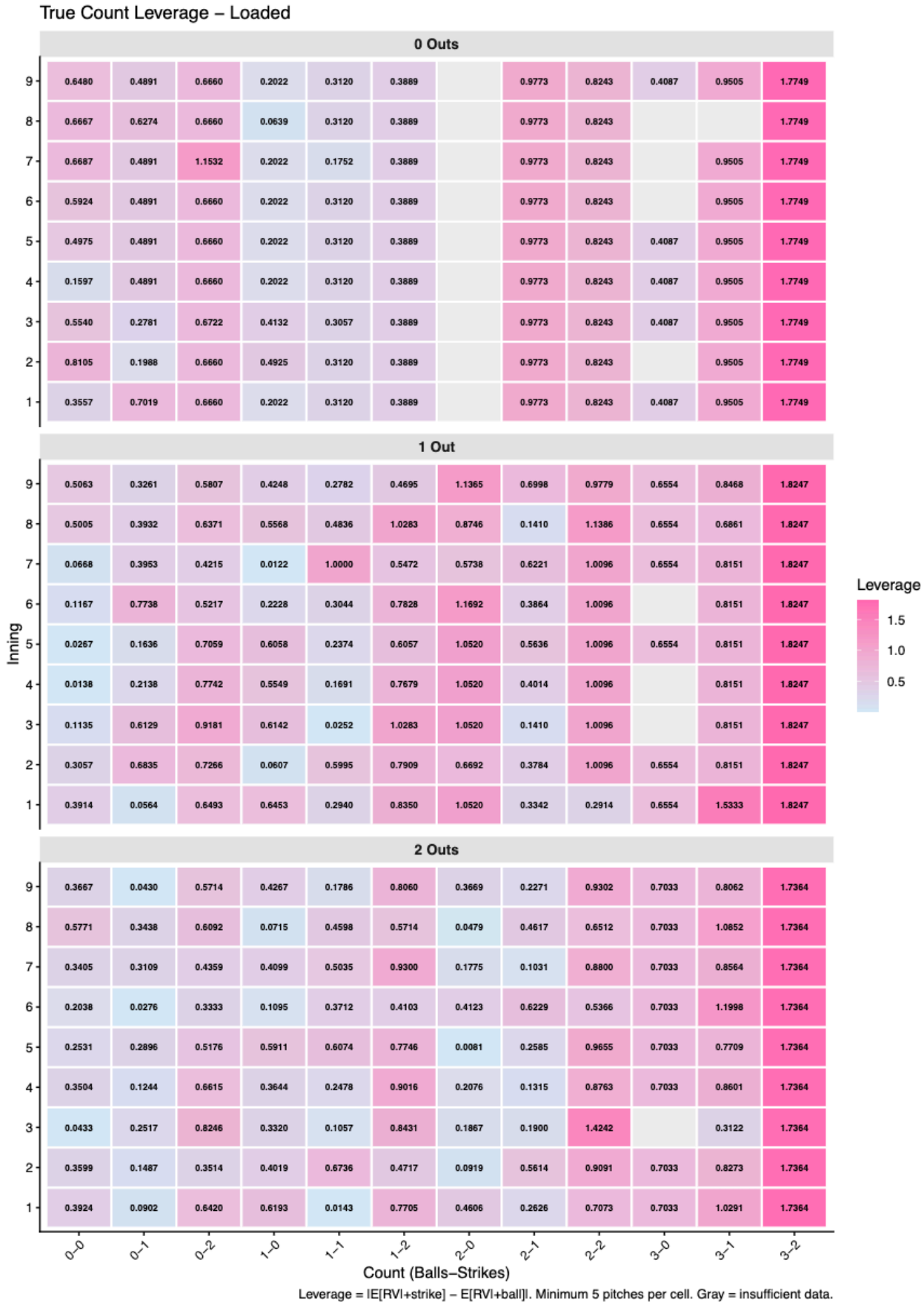


Figure 20: True count leverage — bases loaded.

The leverage heat map consistently shows leverage increasing as the count gets deeper and the game nears its conclusion, with the top leverage situation being full counts with runners in scoring position. The highest leverage situation was a 3-2 count, 1 out with the bases loaded in any inning (1.8247). I suspect that 1 out had more leverage than 2 outs in this game state as adding a strike with 3-2, 2 outs would decrease the run expectancy all the way to 0, whereas with one out, recording an out leaves one more to be secured. When I weighted MLB pitchers performances by their true count leverage and clutch score, the bar charts revealed that the top relievers in baseball—specifically closers—such as Aroldis Chapman (BOS), Mason Miller (SD), Robert Suárez (SD now ATL), and Carlos Estévez (KC), as well as ace starters such as Tarik Skubal (DET), Zack Wheeler (PHI), Paul Skenes (PIT), and Garrett Crochet (BOS) all came up in our leaderboard of either Leverage-Weighted Run Value score, Clutch score, or both, highlighting their dominance in high pressure situations. On the flip side, many of the worst pitchers with respect to Leverage-Weighted Run Value score and Clutch score were Colorado Rockies pitchers. This may be amplified in part by Coors Field's higher altitude, which has historically contributed to worse outcomes for Rockies' pitchers as the thinner air and lack of air resistance causes batted balls to travel farther.

I then averaged the change in delta run expectancy per 100 pitches per count and visualized this in two ways: a bar chart for analyzing the overall trend and a heat map for spatial intuition. The heat map makes it easy to visualize how adding either a strike or a ball influences the average delta run expectancy of the count. The bar chart and heat map show that a 0-2 count has the lowest average delta run expectancy (-0.01) among all pitcher-friendly counts and that a 1-2 count has a higher delta run expectancy (0.02). For hitter-friendly counts, the bar chart and heat map show that a 2-0 count has the highest average delta run expectancy (0.07) of all

hitter-friendly counts. There is also a clear trend of delta run expectancy decreasing from a 3-0 count with each additional strike that is thrown. Neutral counts (0-0, 1-1, 2-2) show a higher average delta run expectancy in favor of hitters. In conclusion, I have reproduced the results found by Jim Albert in his blog "Count Effects" where he used data from the 2015 season to estimate the run value of each count. The results were mostly consistent in the earlier counts, while new trends show that counts with three balls (3-1, 3-2) are more likely to result in a strikeout or put-out than they were 15 years ago, as suggested by their average negative delta run expectancy. Like Albert did for the 2015 Cy Young Award winners Jake Arrieta (CHC) and Dallas Keuchel (HOU), I made a graph for 2025 World Series MVP Yoshinobu Yamamoto (LAD) and it was clear that plate appearances passing through high leverage counts have, on average, negative run value. I also made a graph for Bobby Witt Jr. (KC), and found that an elite hitter such as himself should have more counts associated with positive run value than negative.

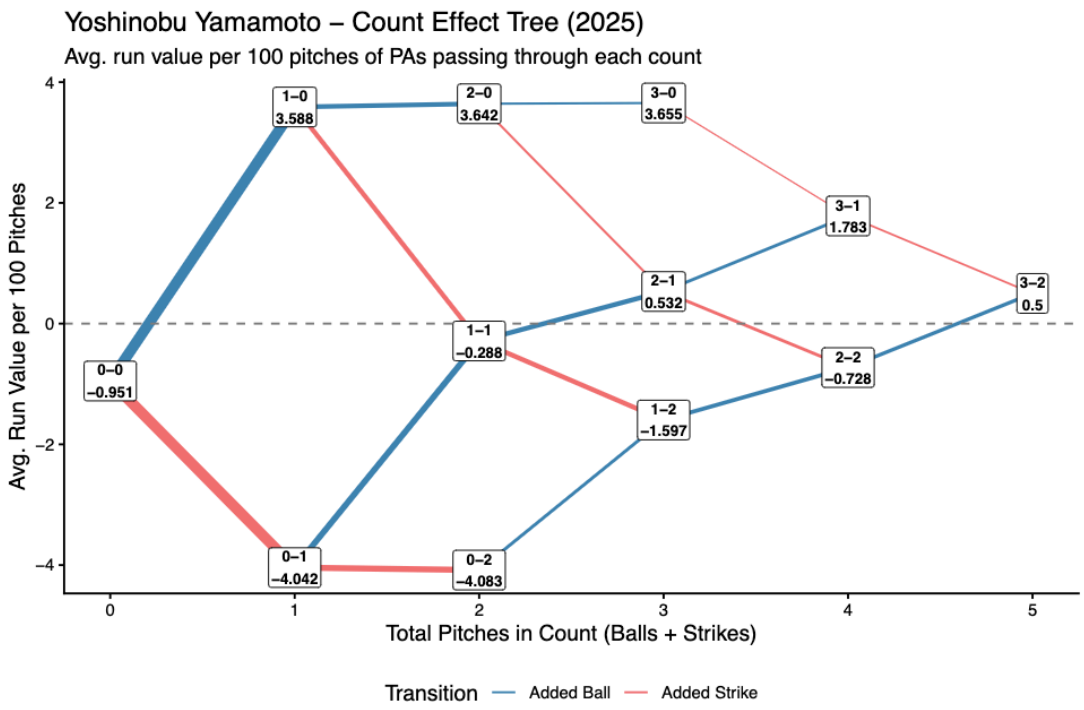


Figure 25: Count effect graph for Yoshinobu Yamamoto (2025). Y-axis = average run value per 100 pitches of PAs passing through that count. Blue edges = added ball; red edges = added strike. Line width proportional to fraction of PAs visiting each count.

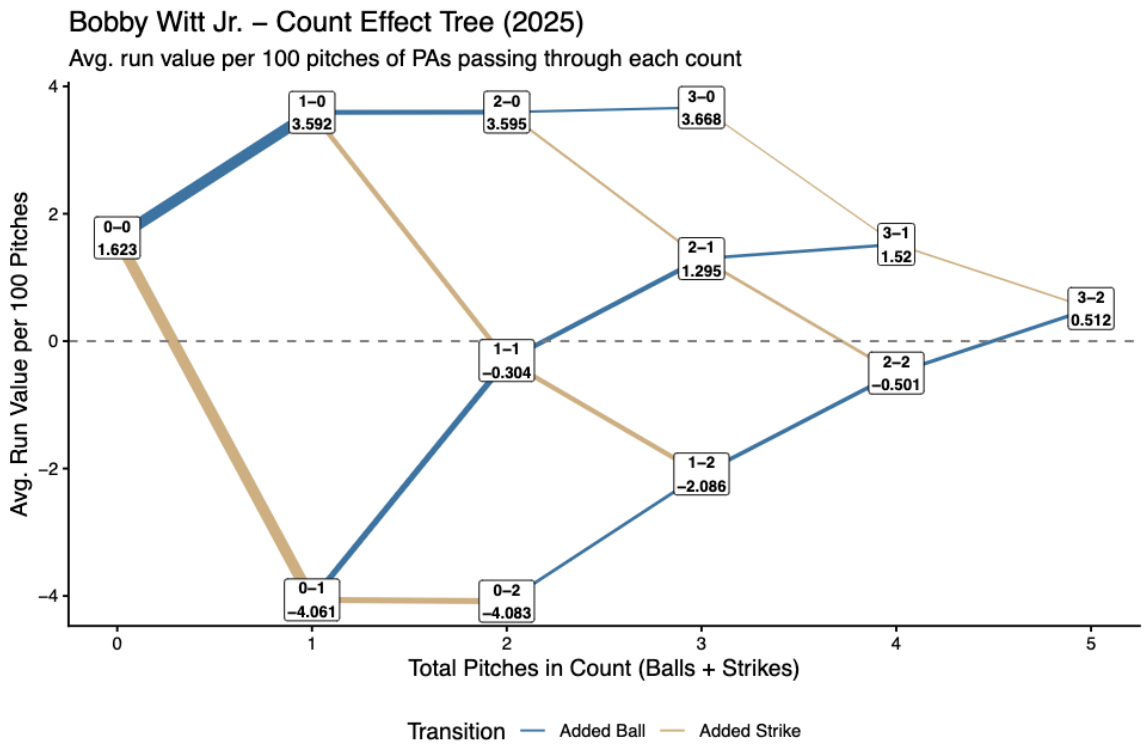


Figure 26: Count effect graph for Bobby Witt Jr. (2025). Y-axis = average run value per 100 pitches of PAs passing through that count. Blue edges = added ball; red edges = added strike. Line width proportional to fraction of PAs visiting each count.

Next, I looked at which pitches are thrown the most in each count. When I recreated Albert's results using all pitchers in MLB in 2025, I found that in 0-0 counts fastball usage was around 60.5%, in 1-1 counts it was around 50.2%, and in 2-2 counts it was around 48.3%. In 3-0 counts, fastball usage sat at about 93.9%, and in pitcher-friendly counts, fastball usage was consistently below 50% but no lower than 42.1%. This is extremely consistent with what Albert found in 2019.

I then considered 20 pitchers who threw more than 2,000 pitches in the 2025 season and examined their fastball usage trends at both the group and individual levels. At the group level, results were consistent with the full-league findings and with Albert's results. Overall, these twenty pitchers had an average fastball usage of around 61% in 0-0 counts, 53% in 1-1 counts,

and 49% in 2-2 counts. In 3-0 counts, average fastball usage was around 95%, and in pitcher-friendly counts, fastball usage was consistently below 53% but no lower than 45%. We know that fastball usage varies from pitcher to pitcher, and at the individual level we can see that 19 out of 20 pitchers in our subset had an average fastball usage above 50% in hitter-friendly counts, while just 6 out of 20 had an average fastball usage above 50% in pitcher-friendly counts. When analyzed at the individual level, it becomes clear that all 20 pitchers tend to throw more fastballs in hitter-friendly counts and decrease their average fastball usage when the count tips in their favor. The overall trend consistent across both Albert's and my analysis is that when behind in counts, pitchers generally lean toward throwing fastballs (which have very predictable movement patterns and are easiest to control), while when ahead in counts, pitchers lean toward throwing more offspeed and breaking pitches to try to get a batter to chase a pitch outside of the strike zone. This is confirmed by the faceted pitch type chart, which shows that changeup, slider, curveball, and sweeper usage collapse in hitter-friendly counts.

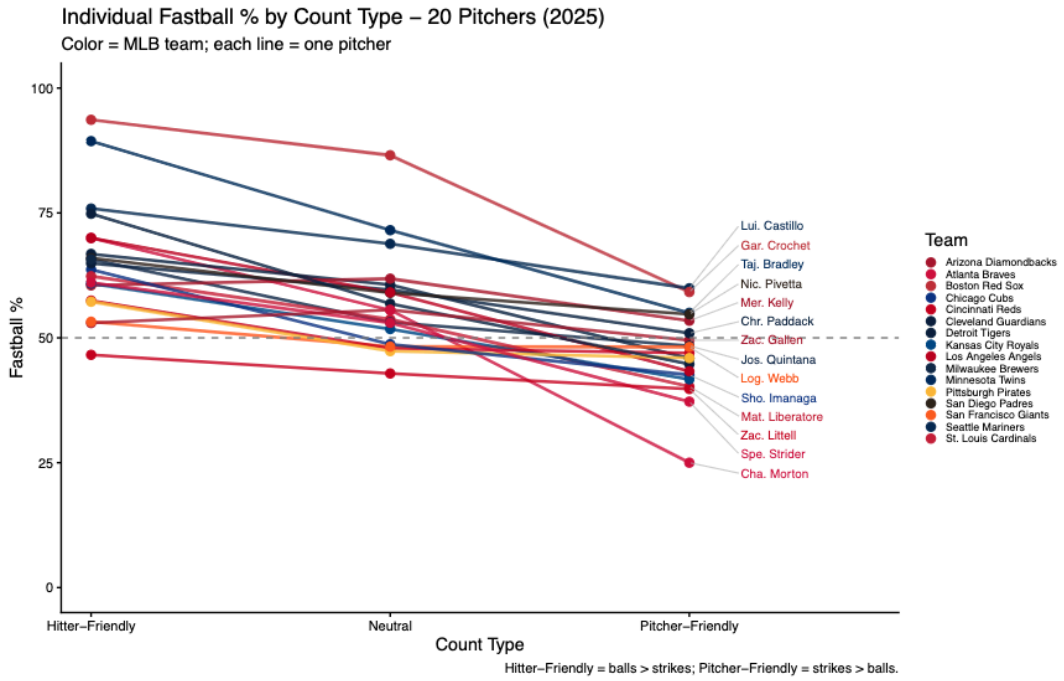


Figure 30: Individual fastball percentage by count type for 20 selected pitchers (2000+ pitches). Each line = one pitcher; color = team.

I then conducted an analysis on whiff rates by count and pitch type, visualized using a line chart. Using a line chart helps answer whether the effectiveness of each pitch type changes depending on the count. The main takeaway from the line chart is that breaking pitches—namely curveballs (29.6%), changeups (29.3%), sliders (31.4%), and sweepers (30.1%)—had the highest whiff rates across almost every count, with their peak coming around 2-0. Four-seam fastballs (18.5%) and sinkers (12.0%) had the lowest overall whiff rates, with the exception of sweepers showing a spike north of 39% in 2-0 counts. Overall, there is a noticeable dip in whiff rate between 2-0 counts and 2-1 and 2-2 counts as hitters get more selective with the possibility of a strikeout increasingly looming.

I then divided the strike zone into a 3×3 grid using raw plate coordinates to examine whether location mattered more or less depending on the count situation and saw that across

hitter-friendly, pitcher-friendly, and neutral counts, the middle of the strike zone consistently generates the fewest swings-and-misses (12.5-13.6%), while low-and-outside consistently generated the most (49.5-53.6%). In neutral and pitcher-friendly counts, low and high in the zone are associated with higher whiff rates as pitchers try to induce chases and mess with the batter's eye levels. In hitter-friendly counts, we see fewer whiffs in the middle of the zone compared to the other two count situations, as pitchers are trying to locate within the zone (generally with fastballs, as we recently established), which gives hitters better pitches to swing at. I then ran a logistic regression model with a Bonferroni correction to predict whether a swing results in a whiff or not. The strongest predictor of whiff probability was a pitch low-and-outside in the zone (OR = 1.155), meaning that location raises whiff odds by about 15.5% compared to the baseline zone (high-and-inside). The next three most significant predictors were curveball (OR = 1.140), sweeper (OR = 1.128), and slider (OR = 1.096), whose horizontal and vertical movement profiles make them harder to hit compared to the baseline pitch type (change-up). Additionally, release speed was significant (OR = 1.026 per mph), meaning that every additional mph of velocity increases whiff odds by about 2.6%—so a fastball thrown at 98 mph vs. 90 mph has roughly a 21% increase in whiff odds from velocity alone, holding all other predictors constant. Many predictors also reduced the odds of a whiff, such as a pitch being in the middle of the zone (OR = 0.192) or a pitch being a four-seam fastball (OR = 0.589) or sinker (OR = 0.345). The number of strikes decreases whiff odds (OR = 0.843) more than the number of balls (OR = 0.943), as hitters naturally get more aggressive with two strikes.

I then analyzed the quality of contact as the count changes to see if there was any significant difference in exit velocity, hard-hit rate, or launch angle depending on the count situation. Overall, exit velocity was remarkably stable across all counts (~81-90 mph), with

hitter-friendly counts such as 3-0 (89.6 mph) and 3-1 (86.5 mph) producing slightly higher average exit velocities. Higher exit velocities are associated with harder hit balls that have higher likelihoods of getting down for hits. Launch angle shows a much clearer count effect, with hitter-friendly counts producing higher launch angles (19-24°) while pitcher-friendly and neutral counts produced flatter contact (16-17°). Higher launch angles are typically attributed to line drives, fly balls, and homeruns. Hard-hit rate is most dramatic, with 3-0 counts producing a hard-hit rate above 40% compared to 0-2 counts at just 20.1%. This suggests that pitchers in hitter-friendly counts give up more hard contact as they swing more freely at pitches. Lastly, I made a two-dimensional density plot of launch angle and exit velocity by count type. The hitter-friendly panel shows a brighter spot at higher exit velocities and positive launch angles, while the pitcher-friendly panel shows a dimmer spot shifted toward lower exit velocities and flatter launch angles, suggesting that pitchers are better at suppressing quality contact when ahead in the count.

Pitch Tunneling Analysis

My next analysis focuses on effect of pitch tunneling and trying to determine whether pitchers who cluster their arsenals more tightly at the swing-commitment point generate higher strikeout rates (K%) and lower BABIP (Batting Average on Balls In Play) against compared to those who do not. The faceted league-wide distribution histograms indicate that the mean tunnel distance, mean plate distance, and mean tunnel ratio among qualified starters are approximately 6.16 inches, 18.41 inches, and 3.78, respectively, with all three metrics displaying roughly a normal distribution.

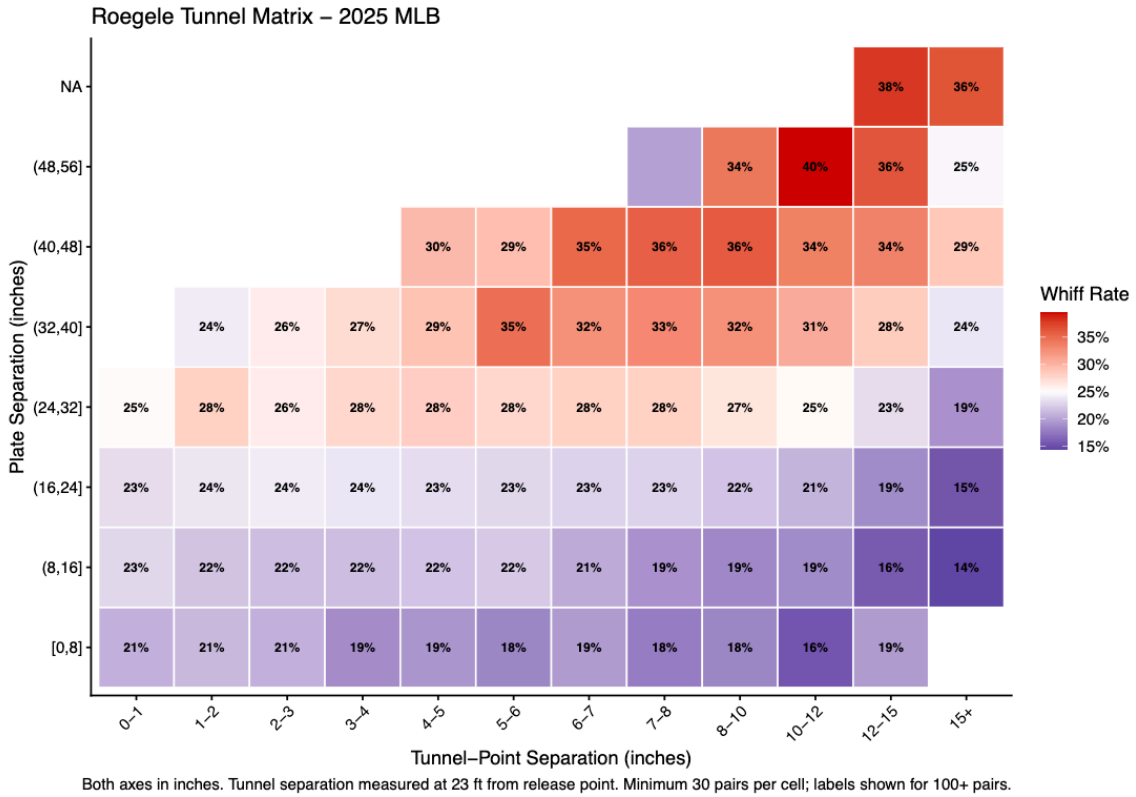


Figure 37: Roegele tunnel matrix (2025 MLB): tunnel-point separation (x-axis) vs. plate separation (y-axis) for consecutive same-PA pitch pairs. Color = whiff rate on second pitch. Minimum 30 pairs per bin. Labels shown for cells with 100+ pairs.

The most important visualization I made was a recreation of Roegele's tunnel matrix. Like Roegele's matrix, there is a noticeable band of elevated whiff rates. This band's location in the top-right corner of the matrix suggests that the combination of moderate tunnel distance and high plate distance makes for the "ideal tunneling zone" and that for a given tunnel distance, whiff rate increases as plate separation increases. The highest whiff cell (40%, top-right area) occurs when pitches are relatively similar but not identical at the swing-decision point (tunnel distance of 10-12 inches) but end up far apart at the plate (plate distance of 48-56 inches). The matrix also shows that high plate separation is highly important to contributing to high whiff rates, rather than just very tight tunnel distances. The matrix also supports Roegele's finding that the closer consecutive pitches are at the swing decision point, the less plate separation they need to generate elevated swing-and-miss rates. This is evidenced by tunnel distances of 5-6 inches having a 35% whiff rate with 32-40 inches of plate separation compared to tunnel distances of 12-15 inches having a 36% whiff rate with 48-56 inches of plate separation.

The next three visualizations were scatterplots of mean tunnel ratio, mean tunnel distance, and mean plate distance against average release speed and colored by whiff rate with a loess regression line and pitcher labels. Their plots did not reveal any noticeable clustering of whiff rates; however, there was a negative linear trend between mean tunnel distance and average release speed, suggesting that as mean tunnel distance increases, average velocity decreases. Then, for each of our tunnel metrics (mean tunnel ratio, mean tunnel distance, and mean plate distance), I made faceted scatterplots of K% whiff rate, BABIP, and hard hit rate and I noticed clear linear trends throughout the figures. For tunnel distance, there was a negative linear trend for K% and whiff rate, suggesting that as tunnel distance increases and pitches become easier to distinguish, K% and whiff rate decrease. There was a positive linear

relationship for tunnel distance and BABIP as well as hard hit rate, suggesting that as tunnel distance increases, batters had higher batting averages and made better contact. There was a positive linear trend for plate distance and whiff rate, suggesting that as plate distance increases and pitches end up farther away at the plate, whiff rate increases. Similarly to our tunnel distance results, our tunnel ratio plots show the same linear trends and suggest that pitches that are close together at the tunneling point but farther at the plate have higher whiff rates, K%, lower BABIP, and lower hard hit rate. Together, these visualizations provide consistent evidence that effective pitch tunneling—characterized by moderate tunnel distance and large plate difference—is

associated with higher whiff rates and weaker contact.

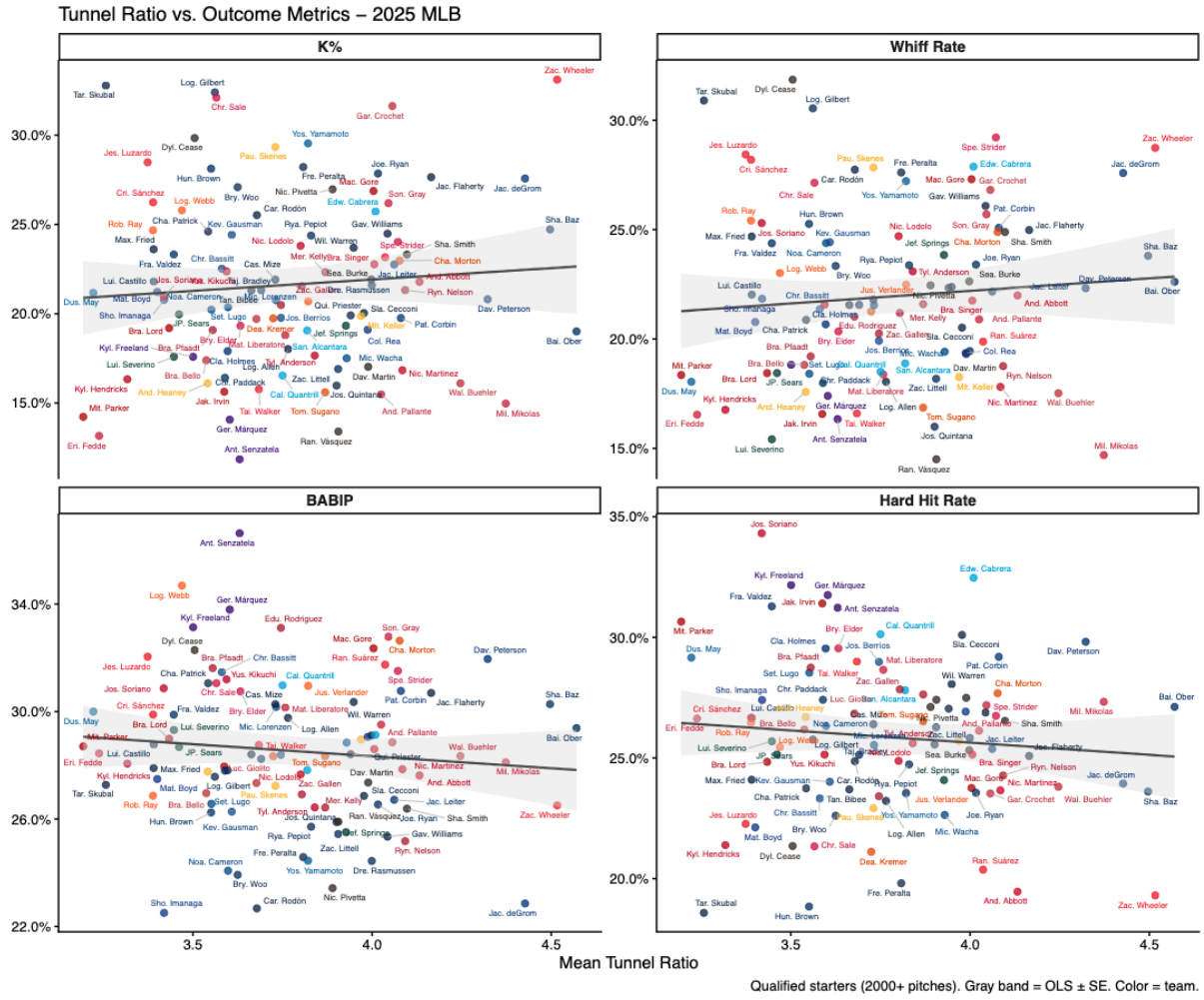


Figure 43: Mean tunnel ratio vs. K%, whiff rate, BABIP, and hard-hit rate for qualified starters.

Next, I binned the range of the tunnel ratio, computed the whiff rate for each bin, and plotted it as a line with point size proportional to sample size as a direct test of whether Roegele's framework can actually predict swing-and-miss outcomes. The plot suggests that the relationship is nonlinear with a peak at [4,5] before declining. This suggests the presence of an optimal tunneling range rather than a strictly increasing effect and that pitches with around four times as much plate distance as tunnel distance are generally the most effective at producing

swings-and-misses. There is also likely some confounding of pitch type, as the low ratio bins may be dominated by same-pitch sequences because they land in similar plate locations as opposed to cross-pitch pairs.

The next visualization answers the question of what pitch type combinations tunnel well across all MLB. Based on the visualization, the tightest tunnelers are same-pitch pairs such as four-seam fastball-to-four-seam fastball and sinker-to-sinker however, their whiff rates are low hinting that throwing the same pitch twice in a row doesn't fool batters, regardless of how similar the two pitches look. Same pitch pairs such as change-up-to-change-up, slider-to-slider, and splitter-to-splitter have higher whiff rates but perhaps as a byproduct of their movement profiles. On the other hand, cross-type pairs such as sinker-to-changeup, sinker-to-splitter, or cutter-to-slider appear to have a farther tunnel distance but carry meaningfully higher whiff rates, suggesting that the deception effect only kicks in when pitch types diverge enough at the plate. This is exactly what Roegele found in 2014: most of the sequences within the high-whiff band involved different pitch types. The top sequences in Roegele's analysis contained loosely defined pitch types in a time when Statcast wasn't nearly as precise as it is today, and thus we can say that our sequence types are also broadly consistent with the ones that Roegele found in 2014. I also made a horizontal bar chart of the top 20 pitchers by mean tunnel ratio colored by whiff rates to find whose pitches diverge the most at the plate relative to how similar they looked at the decision point.

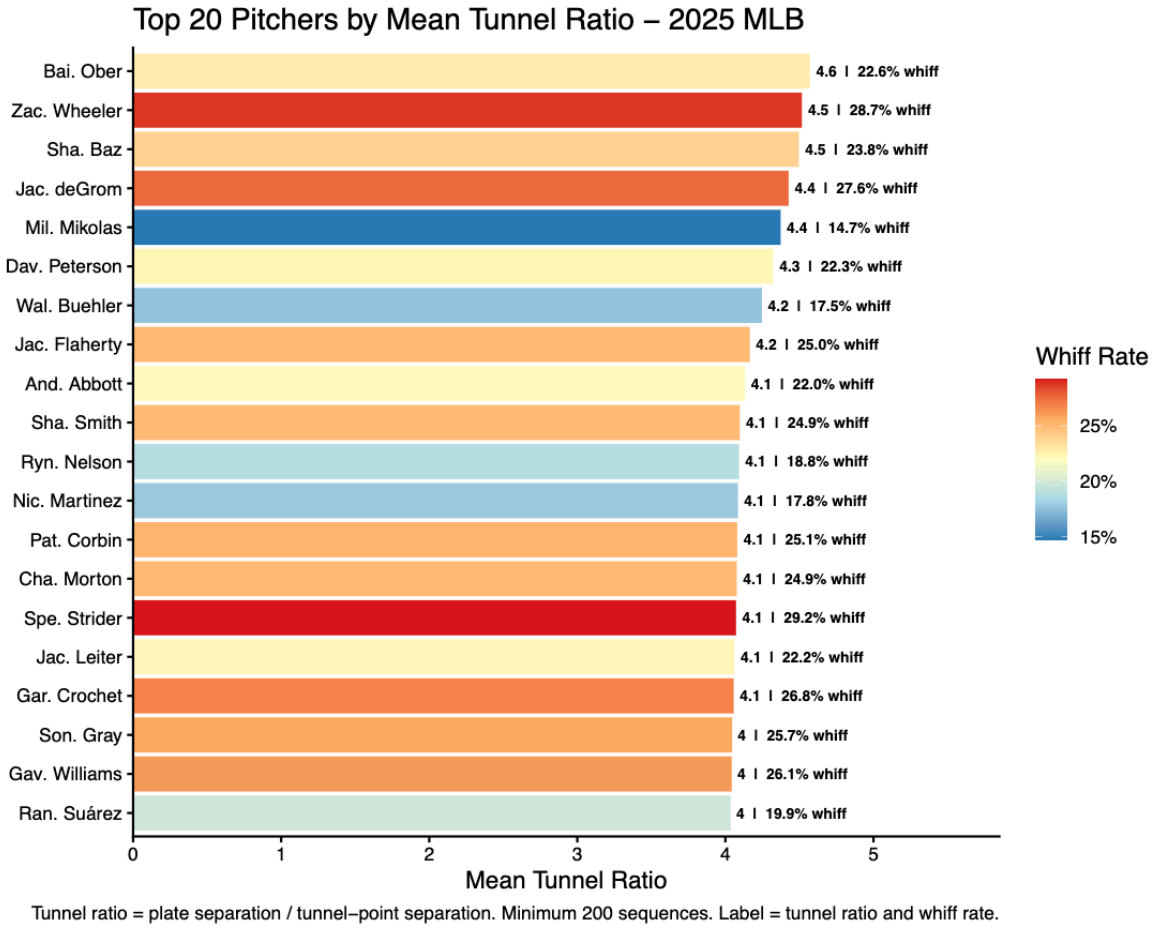


Figure 46: Top 20 pitchers by mean tunnel ratio. Higher ratio = more plate divergence relative to tunnel-point similarity. Color = whiff rate (red = high, blue = low). Labels show tunnel ratio and whiff rate.

The top 20 pitchers in terms of mean tunnel ratio include a mix of elite starters with high whiff rates like Zack Wheeler (PHI), Jacob deGrom (TEX), Spencer Strider (ATL), and Garrett Crochet (BOS) as well as soft-tossers with lower whiff rates like Bailey Ober (MIN), Nick Martinez (TB), and Ranger Suarez (PHI now BOS) suggesting that there is some sort of relationship between tunneling effectiveness and velocity.

To gauge whether pitcher-level tunneling metrics predict outcomes at the arsenal level, three separate groups of OLS models were fit—one for each of mean tunnel distance, mean plate distance, and mean tunnel ratio—predicting K%, whiff rate, BABIP, and hard hit rate.

Additionally, a joint model using all three predictors was also fit. Prior to fitting the models, VIF diagnostics on the joint model confirmed minimal collinearity between the three predictors (VIF: tunnel distance = 2.25, plate distance = 1.76, tunnel ratio = 1.82) despite the conceptual relationship that exists between tunnel distance, plate distance, and tunnel ratio. None of these values exceed the conventional threshold of 5, however fitting the predictors separately ensures the cleanest interpretation of each metric's independent association with outcomes.

The model comparison table reveals a consistent pattern. Overall, tunnel distance explained very little variance and never achieved significance at the $\alpha = 0.05$ level (K%: R-squared = 0.0205, p = 0.151; whiff rate: R-squared = 0.0139, p = 0.239; BABIP: R-squared = 0.0095, p = 0.330; hard hit rate: R-squared = 0.0230, p = 0.128). Tunnel ratio alone performed very similarly across all four outcome variables, with its p-values ranging from 0.317 to 0.420 and its R-squared values ranging from 0.0065 to 0.0100. Among all three predictors, plate distance performed the worst. When fit alone, plate distance displayed R-squared values no higher than 0.0251 and p-values no smaller than 0.11, which were associated with whiff rate.

Across all 12 extended models, which included either tunnel distance, plate distance, or tunnel ratio along with arm slot and spin rate, only a few predictors were ever significant. For the models using tunnel ratio or plate distance along with arm slot and spin rate to predict K%, arm slot was significant at the $\alpha = 0.05$ level with p-values of 0.03685 and 0.03155 respectively with both their confidence intervals being fully negative, suggesting that an decrease in arm slot is associated with an increase in K%, holding all other predictors constant. The negative sign signals that pitchers with side arm or submarine release points tend to generate slightly elevated K%. This pattern is perfectly exhibited by future hall of fame left-hander Chris Sale for the Atlanta Braves, who has an arm angle of 8° and a K% of 32.5% in 2025 which is solidly in the

top 5% of MLB among qualified pitchers. Similarly, righty closer Edwin Díaz of the New York Mets (now Los Angeles Dodgers) had an arm angle of 19° and an astronomical K% of 38% which ranks 3rd in MLB among pitchers who threw more than 50 innings in 2025.

These results provide a more complete answer to Roegele's open-ended question than the original joint model did. We can say that no tunneling metric reliably predicts K%, BABIP, whiff rate, or hard hit rate at the pitcher-season level in the 2025 MLB Statcast dataset. The arm slot finding is the most substantial result in the regression block and suggests that future tunneling analyses could control or subset for delivery angle, as it seems to independently influence strikeout rate in ways that overlap with the deception mechanism tunneling is intended to capture.

Next, I isolated just Yoshinobu Yamamoto's tunneling sequences and plotted every consecutive pitch pair as a point in the tunnel distance vs. plate distance space, colored Dodger blue for non-whiffs and Dodger red for whiffs. From the plot we can see that the bulk of the data clusters in the lower-left meaning most of his consecutive pairs are tight at both the tunnel point and the plate, hinting that same-pitch sequences like four-seam fastball-to-four-seam fastball dominate (his most used sequence with 228 occurrences) his repertoire. The cloud has a positive slope, which is expected as high tunnel separation typically leads to higher plate separation. Based on the plot, his whiffs are spread out and not clustered in the lower-right as we had hoped, meaning that tunneling isn't the primary driver of his whiffs as much as his pure "stuff quality." The next visualization aggregates Yamamoto's pairs to the pitch type level then plots each pair in the mean tunnel distance vs. mean plate distance space, answering the question of which of Yamamoto's specific pitch type combinations are his most deceptive with respect to whiffs. We will again find these combinations in the low-tunnel distance, high-plate distance corner. His

most deceptive pitch pair is his cutter-to-slider sequence (sample size of just 10), which has a mean tunnel distance of 3.77 inches and a mean plate distance of 11.56 inches, with a whiff rate of a staggering 80%, likely due to the small sample size. Next, his cutter-to-cutter sequence is also very deceptive with a mean tunnel distance of 3.97 inches and a mean plate distance of 10.99 inches, but has a whiff rate of just 7.1% (sample size of 23). His four-seam fastball-to-four-seam fastball sequence has a mean tunnel distance of 4.66 inches and a mean plate distance of 13.64 inches with a 21.1% whiff rate which is above the league average whiff rate of 19.3% for the same sequence. His sinker-to-splitter pair is more effective with a whiff rate of 39.51% and a mean tunnel distance and plate distance of 4.24 inches and 15.53 inches, respectively (48 occurrences).

A heat map of Yamamoto's tunnel point density plots shows all of Yamamoto's pitch types clustering in roughly the same vertical band (~4.5 feet Tunnel Z) at the 23-foot decision point, which confirms consistent release point disguise among pitches. Had they been more spread out, we could have assumed that Yamamoto's whiff generation likely comes from another source. Making two side-by-side histograms of the full distribution of qualified starters, one for tunnel distance and the other for tunnel ratio, we can compare Yamamoto to the MLB average. We find that Yamamoto's tunnel stats are essentially league average with his mean tunnel distance and ratio both being a little higher than league average. He sports an average tunnel distance of 6.184 inches compared to the MLB average 6.164 inches and a tunnel ratio of 3.821 compared to the MLB average of 3.778, which rank 49th and 45th among MLB pitchers with 500+ pitches respectively. His elite status and outcomes (K% 29.54%, 7th; BABIP 0.245, 8th) are not driven entirely by tunneling as measured here. His whiff rate of 27.2% ranks in the 87th percentile which is a signifier of elite swing-and-miss stuff. The small difference between

our calculation for whiff rate (27.2%) and Baseball Savant's calculation (28.9%) is because we are considering pitches that are part of a consecutive same-PA pair meaning the first pitches of every PA are dropped which slightly alters the denominator.

We can also compare Yamamoto to two other pitchers, one who is an elite tunneler (Jacob deGrom, Texas Rangers) and one who is a very loose tunneler (Erick Fedde, St. Louis Cardinals/Atlanta Braves/Milwaukee Brewers now Chicago White Sox). We find that deGrom has the tightest average tunnel distance (4.97 inches) and has more similar outcomes to Yamamoto (27.6% K%, 27.6% whiff%, 0.231 BABIP), while Fedde has the loosest average tunnel distance (6.75 inches) and much worse results compared to the other two pitchers (13.16% K%, 16.54% whiff%, 0.286 BABIP). This comparison suggests that tunneling has an important effect when coupled with quality of individual pitch types: deGrom and Yamamoto both tunnel well and have elite stuff while Fedde tunnels moderately and lacks the pitch quality to capitalize.

Table 20: Yamamoto percentile ranks among pitchers with 500+ pitch sequences (2025).

Metric	Yamamoto Value	MLB Avg	Yamamoto Rank	N Pitchers	Percentile
Tunnel Dist (in)	6.184	6.164	49	102	48.0
Tunnel Ratio	3.821	3.778	45	102	44.1
K%	0.295	0.216	7	102	94.1
Whiff%	0.272	0.219	14	102	87.3
BABIP	0.244	0.285	8	102	93.1

Table 21: Tunneling profile comparison: Yamamoto (average tunneler), deGrom (tight), Fedde (loose). Stats computed directly from 2025 pitch data.

Pitcher	Team	Tunnel Dist (in)	Tunnel Ratio	K%	Whiff%	BABIP	Pitches
Eri. Fedde	St. Louis Cardinals	6.754	3.2	13.2%	16.5%	0.286	2369
Yos. Yamamoto	Los Angeles Dodgers	6.184	3.8	29.5%	27.2%	0.245	2830
Jac. deGrom	Texas Rangers	4.975	4.4	27.6%	27.6%	0.231	2613

Pitch Sequencing Analysis

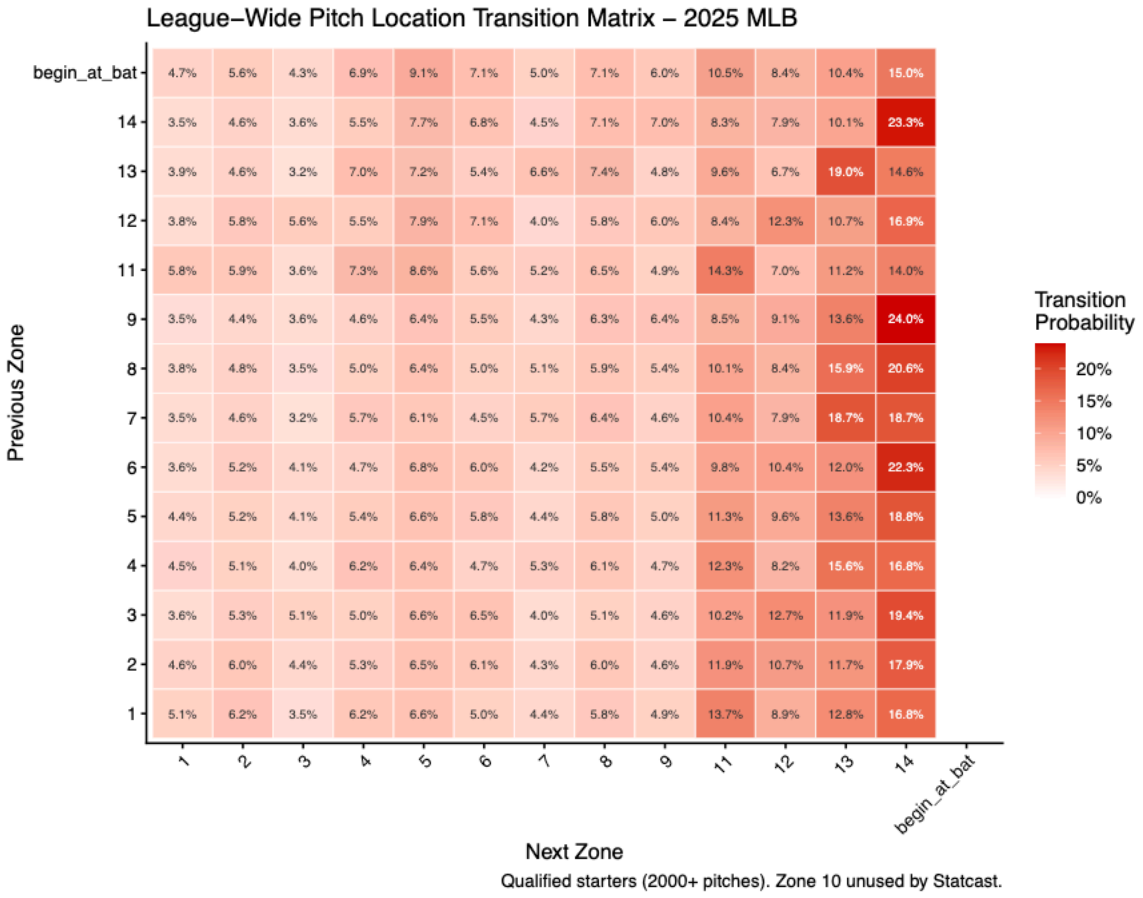


Figure 53: League-wide mean pitch location transition probability matrix across all qualified starters (2000+ pitches, 2025 MLB). Zones 1-9 = strike zone; 11-14 = chase zones; begin_at_bat = first pitch of PA.

Using the 104 qualifying starting pitchers (2000+ pitches in 2025), the league-wide transition matrix heat map shows that pitchers tend to repeat zones and specific attack patterns, with zone 14 being the most common next zone from almost every previous zone, conveying a league-wide tendency to expand the zone low-and-away. In fact, all chase zones (zones 11, 12, 13, and 14) were consistently the most followed-up zones from any previous zone. Besides the chase zones, the most common zone for a pitcher to begin an at-bat is in the middle of the strike zone (zone 5). Applying k-means clustering to each pitcher's flattened 196-dimensional

transition matrix and plotting an elbow plot from k = 2 through k = 8, we find a clear inflection point at k = 5, identical to Clay's original finding. The five-cluster structure in pitcher location sequencing thus persists across seasons. Cluster 1 is headlined by Spencer Strider (ATL), Dylan Cease (SD now TOR), and Jacob deGrom (TEX). Cluster 2 (7 pitchers) includes all lefties with the exception of Yoshinobu Yamamoto (LAD). Cluster 3 is entirely lefties featuring Garrett Crochet (BOS), Chris Sale (ATL), Max Fried (NYY), and Shota Imanaga (CHC). Cluster 4 features aces like Bryan Woo (SEA), Zack Wheeler (PHI), and Tarik Skubal (DET). Group 5 includes other aces such as Freddy Peralta (MIL now NYM), Logan Webb (SD), and Paul Skenes (PIT). It appears that handedness of a pitcher plays a role in their sequencing behavior, with same handedness pitchers sharing similar sequencing patterns.

One-way ANOVA tests show that the three metrics which proved significant with spin rate ($p = 0.0173$), velocity ($p = 0.0269$), and whiff rate ($p = 0.0425$) suggesting that the average whiff rate, velocity, and spin rate of pitchers differs meaningfully depending on which cluster they belong to. It also suggests that clustering picks up on stuff quality rather than pure sequencing behaviors. We can see from our bar graphs that the clusters vary meaningfully in average spin rate, suggesting the groupings partially reflect differing movement profiles among groups, although the average spin rate of cluster 1 was the highest (2353.004). The average velocity of the clusters was between 87 and 89 mph with cluster 4 having the highest average velocity (89.664). Lastly, there were some noticeable differences in whiff rate between the groups, with cluster 2 having by far the highest whiff rate (25.7%), although this may be due to its smaller size (n=7). It is important to note that even though K%, exit velocity, hard-hit rate, and BAA do not differ significantly, tests for significance were not originally created to explain baseball phenomena and differences may still be relevant on the field even if not statistically

significant.

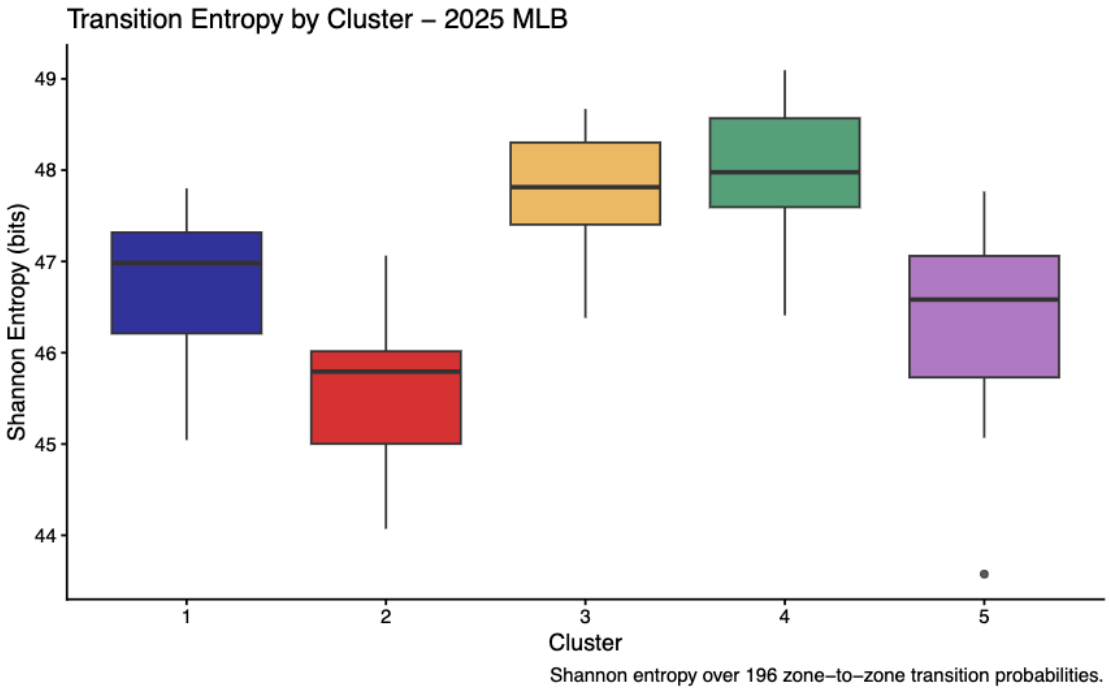


Figure 56: Shannon entropy of each pitcher's 196-element transition distribution by cluster. Higher entropy = more varied, less predictable sequencing.

Shannon entropy is highly significant across clusters with clusters 3 and 4 having the highest average entropy (47.781 and 47.994 bits, respectively) and cluster 2 having the lowest (45.566 bits). This means that pitchers like Yamamoto and Framber Valdez (HOU now DET) are actually more predictable in location sequencing than average, but compensate for this with their exceptional pitch quality. This matches Clay's finding that the highest-performing group wasn't necessarily the most unpredictable.

Next, I plotted the pitchers on velocity vs BAA, colored by cluster to try and replicate Clay's "Group 1 vs. Group 4" comparison. There is a clear negative linear trend between average pitch velocity and BAA, suggesting that as velocity increases BAA correspondingly decreases; however, there is no clear grouping of pitchers from different clusters

albeit cluster 3 pitchers having slightly lower average velocities. The best cluster in terms of BAA is cluster 2 (0.233), anchored by Carlos Rodón (NYY), Christopher Sánchez (PHI), Framber Valdez , Jesus Luzardo (PHI), and Yoshinobu Yamamoto, who all had strong seasons last year. The worst cluster in terms of BAA is cluster 5 (0.247), anchored by Paul Skenes, Logan Webb, and Kevin Gausman (TOR) who had strong seasons despite the struggles of their other peers in cluster 5. This is somewhat surprising and may be a result of the small sample size cluster 2 compared to cluster 5 as their average velocities, spin rate, and exit velocities are all roughly the same. The difference between the two clusters in terms of BAA is likely a result of pitch quality, movement profiles, and how they utilize their arsenal although the small sample size gives this comparison limited statistical power. When we break down the pitch metrics for each clusters' 8 most used pitches, we see that cluster 5 has much less induced vertical break and lower whiff rates on most pitches compared to cluster 2.

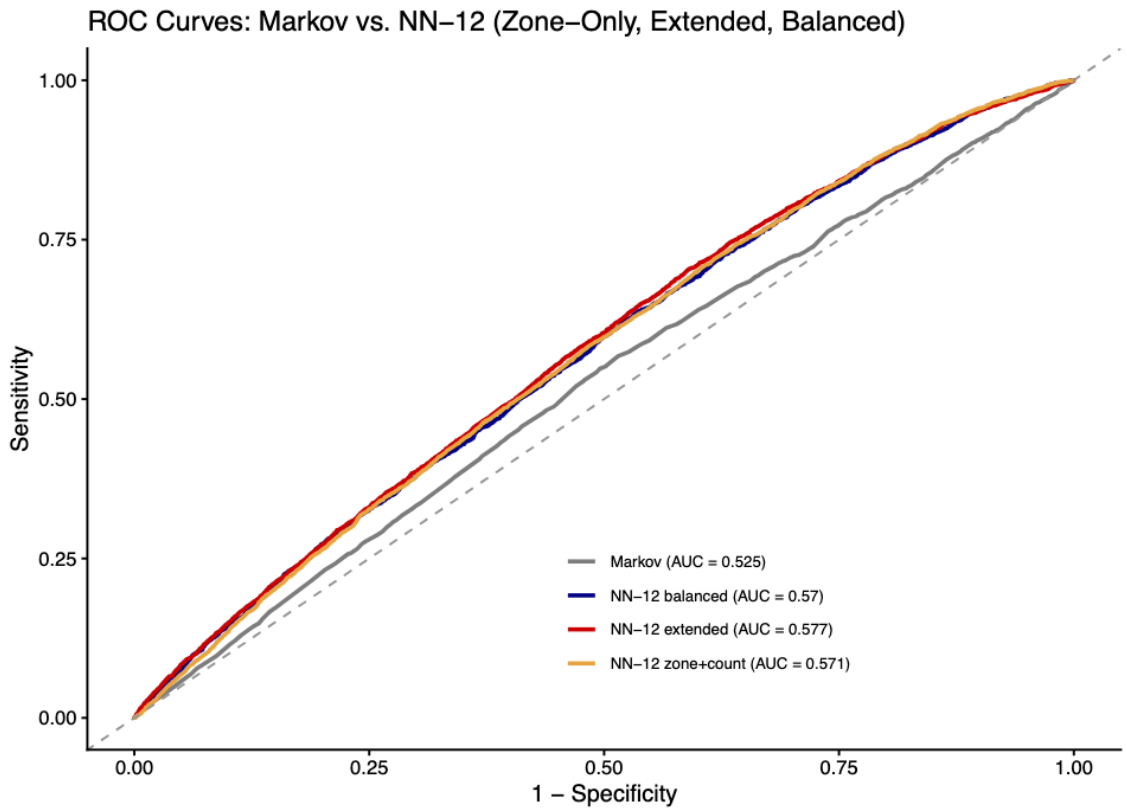
The next section of the analysis answers Clay's call for readers to employ neural networks as opposed to first-order Markov chains and determine whether they can outperform Markov chains with respect to predicting whether the next pitch will be a whiff. The first-order Markov chain only captures the probability of transitioning between two adjacent zones, ignoring sequence depth, count context, and pitch type, while a neural network includes all of these as features, directly addressing the limitations Clay previously identified. When I checked the whiff class balance, I found that out of 727,197 pitches in our dataset, there were only 80,157 whiffs (11.02%). When looking at the whiff class balance for Yamamoto's 2025 season, I found that he generated 359 whiffs out of a possible 2,830 pitches thrown (12.69%). I also found that for his September 6th game he had 22 whiffs out of a possible 112 pitches (19.64%).

Although the likelihood ratio test between first-order and second-order Markov chains was statistically significant (LRT = 212.92, $p < 0.001$), the AUC's suggest (1st-Order Markov AUC 0.5394 vs. 2nd-Order Markov AUC 0.5260) that the second-order Markov chain did not meaningfully improve predictive accuracy on the holdout set despite being an extension of the first-order dependence structure. The neural network has 12 nodes and one hidden layer using features including current zone, previous zone, count, pitch type, and the Markov transition probabilities. The first-order Markov baseline simply uses the empirical transition probabilities as the prediction. Both the neural network and first-order Markov baseline used a 80/20 train-test split. The first-order Markov had a 51.1% accuracy compared to the neural network's 91.4% and had a AUC of 0.5252 compared to the neural network's 0.5713. The accuracy gap is misleading as whiffs are relatively uncommon so a model that predicts no whiffs could achieve high accuracy based on the random split of the whiff distribution in the train and test sets. A better metric of performance is the AUC score, which tells us that the neural network with count context, pitch type, and sequence depth provide a modest but real improvement and add predictive value beyond just the pure zone transition probabilities; however, this improvement is still marginally better than completely random guessing. The extended neural network had a slightly higher AUC at 0.5771 and almost the exact same accuracy (91.0%) suggesting that our original predictors were accounting for most of the variation in our model.

Upon consulting these neural networks two-way tables, it became clear that the huge class imbalance (dataset only consisted of 11% whiffs) was training the model to predict everything as a non-whiff and thus achieve very high accuracy. When we undersampled the non-whiffs down to match the whiff count and used the same extended parameters, we could see that the model became more calibrated and was genuinely trying to distinguish between the two

classes. In this case, our accuracy dropped sharply (54.8%) because the model is now more uncertain rather than completely biased. The AUC also dropped to 0.5704, however this means that while the balanced model actually ranked pitches by whiff likelihood better overall when compared to the Markov model, it fell short of the imbalanced models' performance. When using the balanced dataset in JMP and using more tours, the AUC of the test set was 0.5836 which would have been the best among all of our models. Additionally, JMP's profiler suggests that the most important predictors of whiff are release speed and number of pitches in the at-bat (which are both intuitive as the faster a pitch is the harder it is to hit and the more pitches in an at-bat there are the more opportunities there are to swing-and-miss), as well as mean tunnel distance, hence our scatterplots in our pitch tunneling analysis. When we consult our permutation feature importance table, we see that in the zone + count model, the most important predictors are to_zone, balls, and strikes. In the extended model, balls strikes, to zone, and release speed are most important, and in the balanced model the most important predictor is the number of pitches in the plate appearance.

This result validates Clay's critique of first-order Markov chains leaving information on the table. In conclusion, the additional gain achieved by neural networks with added context and features is real but not enormous. Determining if there are any predictors left out of the model that might improve results would be an obvious next step, although whiffs might be rare enough that there is no way to improve the model without wildly increasing complexity.



al whiff/no–whiff sample. Extended features: velocity, pfx_x, pfx_z, batter handedness, previous pitch type, pitch number in game.

Figure 58: ROC curves for the Markov baseline, zone-only NN-12, extended NN-12, and balanced NN-12 (trained on equal whiff/no-whiff sample).

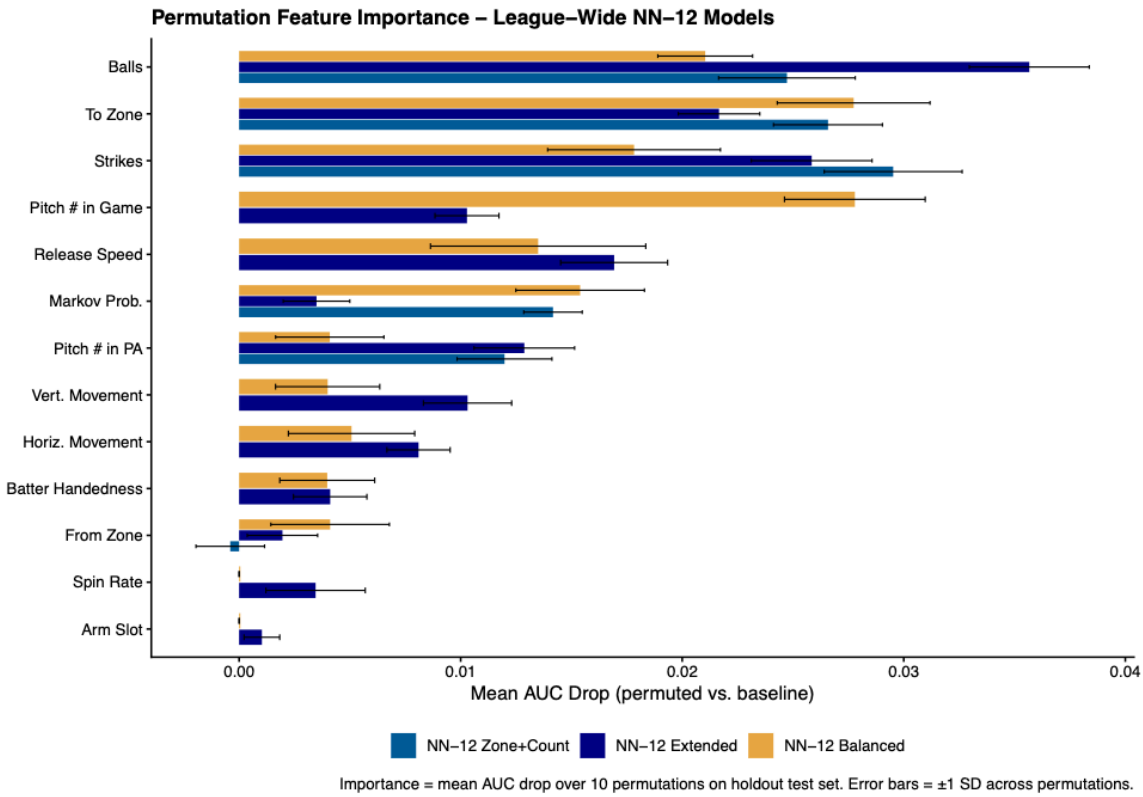


Figure 59: Permutation feature importance for the three NN-12 models. Importance = mean drop in AUC when a feature is randomly permuted 10 times on the respective test set. Higher = more important. Zone-only model uses its own feature set; extended and balanced share the full feature set.

Yoshinobu Yamamoto Pitch Sequencing Case Study

We can clearly see that four-seam fastball-to-four-seam fastball is Yamamoto's single most used sequence at 228 total instances in 2025. While his four-seam fastball-to-splitter and splitter-to-four-seam fastball are next at around 209 and 188 occurrences respectively, which speaks to the core of Yamamoto's tunneling strategy laid out in the previous analysis. Yamamoto's four-seam fastball and splitter look nearly identical at the tunneling point and are thrown at roughly the same arm angle, so alternating them is maximally deceptive and a solid strategy and it makes sense why these three sequences account for over 625 transitions and roughly a third of his consecutive pitch pairings. These three pairings were also the three most used on September 6th against the Orioles (10, 9, and 8 occurrences respectively). He also established the curveball as an effective weapon as his next most frequently used combos are four-seam fastball-to-curveball (163) and curveball-to-four-seam fastball (155), which require the batter to adjust to different vertical movement profiles and alternating eye levels.

The handedness table is especially telling as we see that against left-handed hitters, Yamamoto leans more heavily on his core four-seam fastball-splitter combinations. Four-seam fastball-to-four-seam fastball runs at 12.3% vs LHH compared to 9.2% vs RHH, four-seam fastball-to-splitter runs at 11.9% vs LHH compared to 7.8% vs RHH, and splitter-to-four-seam fastball runs at 11.4% vs LHH compared to 6.3% vs RHH. Against righties, he throws significantly more sinkers and sliders. He throws his four-seam fastball-to-sinker 5% of the time to RHH but only 0.3% of the time to LHH. He throws his sinker-to-splitter 4.2% of the time to RHH vs 0.2% of the time to LHH. Finally, he throws a sinker-to-curveball 3.3% of the time vs RHH vs 0.1% of the time vs LHH. Essentially, his sinker is reserved for right handed batters, which makes sense as his sinker has a lot of armside movement and creates challenging looks for

righties. Against lefties, he opts for mainly his fastball and splitter which tunnel well and play on the outside corner for lefties.

Yamamoto's top pitch sequence usage noticeably changed after his September 6th start against the Orioles. He decreased his four-seam fastball-to-cutter usage from 6% to 2.9%. His sinker-to-curveball usage jumped from 2.6% to 7.1%. His sinker-to-sinker usage dropped from 7.5% to 5.8%, and his cutter-to-four-seam fastball usage dropped from 3.4% to 1.6%. When broken down to just the individual pitch level, we see that Yamamoto increased his curveball usage and splitter usage by 4.5% and 2.3% respectively and decreased his four-seam fastball usage and cutter usage by 3.4% and 2.8% respectively after September 6th. The figure with Yamamoto's pitch sequences with location zone suggest that his sequencing is almost entirely a low-zone story, with his most common sequence by a good margin being splitter low-middle to splitter low-middle at approximately 5% usage.

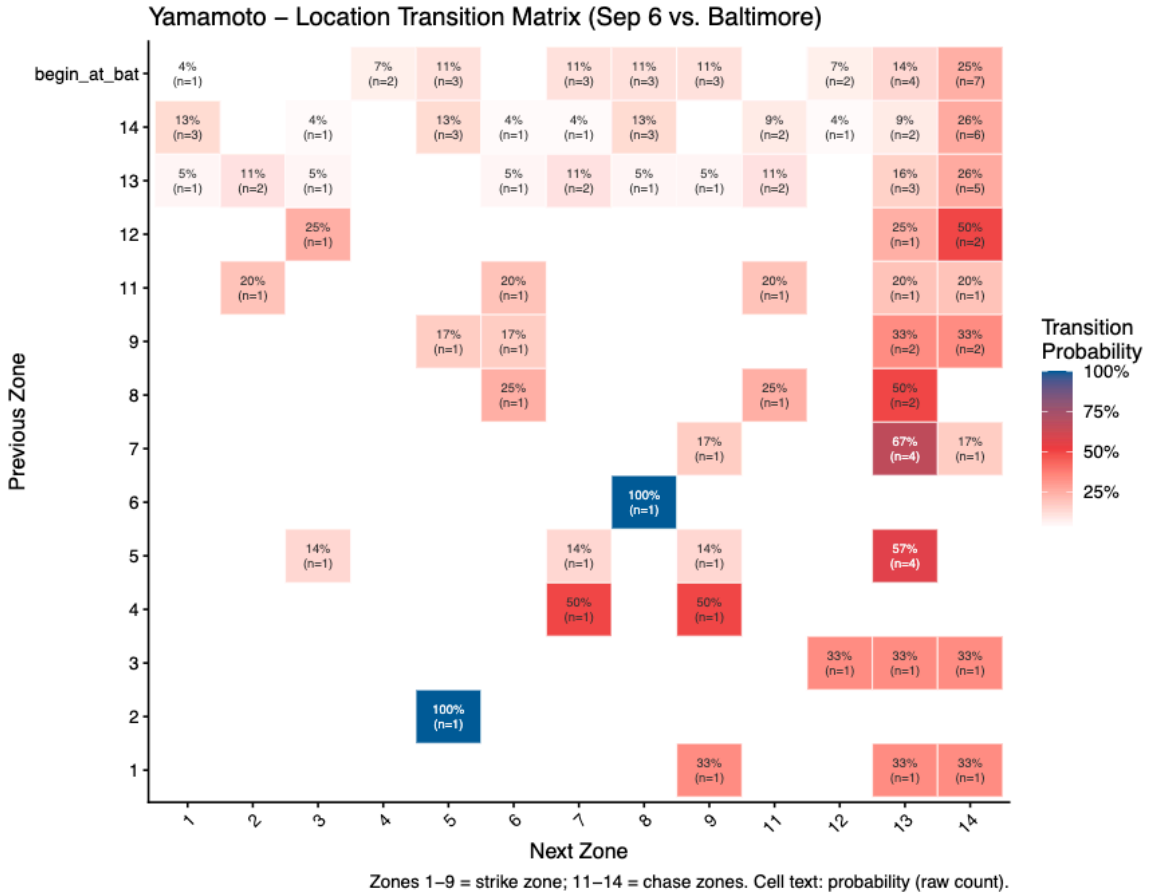


Figure 63: Yamamoto's location transition matrix for the September 6 game vs. Baltimore. Only non-zero transitions shown. Each cell shows transition probability and raw count (n).

During the outing on September 6th against the Orioles, Yamamoto made 112 zone transitions, using 61 unique from-to zone pairs. 59% of those pairs appeared exactly once and a maximum of 7 repetitions was the most among any single transition. In Clay's case study of Verlander's 2019 no-hitter benchmark, Verlander had 148 two-pitch sequences and 83 unique transitions. 52 out of the 83 (62.65%) two-pitch sequences thrown by Verlander were only used once, and no sequence was used more than five times. Compared to Verlander, Yamamoto falls slightly below his uniqueness threshold and has a higher maximum repeat count, meaning he

relied on a few favored transitions more often than Verlander did but overall his sequencing variety provided optimal unpredictability.

When looking at Yamamoto's location transition matrix on September 6th, we can see that it is actually surprisingly sparse with most cells being empty, which serves as a visual confirmation of the 59% single use pairs. Zones 13 and 14, the low right and left chase zones and the primary splitter destination for RHH, is the most frequent destination from most previous zones. Yamamoto's plot is spread out and he displays unpredictable usage patterns that keep hitters on their toes, never reusing a specific location transition pair more than 5 times in the game. The most common zone's Yamamoto used to begin the at-bat (first pitches) were zones 14 (25%), 13 (14.3%) as well as zones 5, 7, 8, and 9 (all 10.7%). His most common pitch to begin an at-bat against a RHH was a curveball (n=5) and a four-seam fastball against a LHH (n=6). His curveball usage explains why his most common starting zones are low chase zones. Terminal pitches (final pitch in the at-bat) are clustered in zones 13 (21.4%), 6 , and 8 (both 14.3%) with his most common pitch to end the at-bat against a RHH was a four-seam fastball (n=7) and a splitter against a LHH (n=5). The high splitter usage against lefties explains the high frequency of zone 13 visits. Verlander's no-hitter saw 57% of at-bats ending in zones 5, 9, or 14, broadly consistent with Yamamoto's philosophy of finishing down in the zone.

To score each pitch in the September 6th game by its whiff probability, three models were compared: the first-order Markov baseline, the league-wide zone + count NN-12, and a pitcher-specific logistic regression model trained on Yamamoto's 2025 season data. The use of different models stems from the differing distributions that support the league-wide model and the pitcher-specific model. The neural network was trained on over 700,000 pitches and has an 11% league-wide whiff rate, while Yamamoto's September 6th game had a 20% whiff rate.

Because we are calculating the predicted probability of a whiff on the next pitch, whiffs on the terminal pitch of the at-bat are excluded leading to an actual whiff rate of 17% for the September 6th game. The calibration table makes this abundantly clear when we see that the Markov baseline's mean predicted probability was 0.108, the league-wide NN's was 0.091, and the pitcher-specific logistic regression was 0.123—the only model whose average prediction approaches the actual 17% game whiff rate. This is not necessarily a point against the use of neural networks but rather speaks to the nature of the domain mismatch and suggests that a neural network may have been overly complex for capturing the patterns in this data. A model trained on the league's average patterns has no way of knowing that Yamamoto induces nearly twice as many whiffs as the average MLB pitcher, or that his specific pitch mix, stuff quality and sequencing tendencies substantially shift the baseline probability.

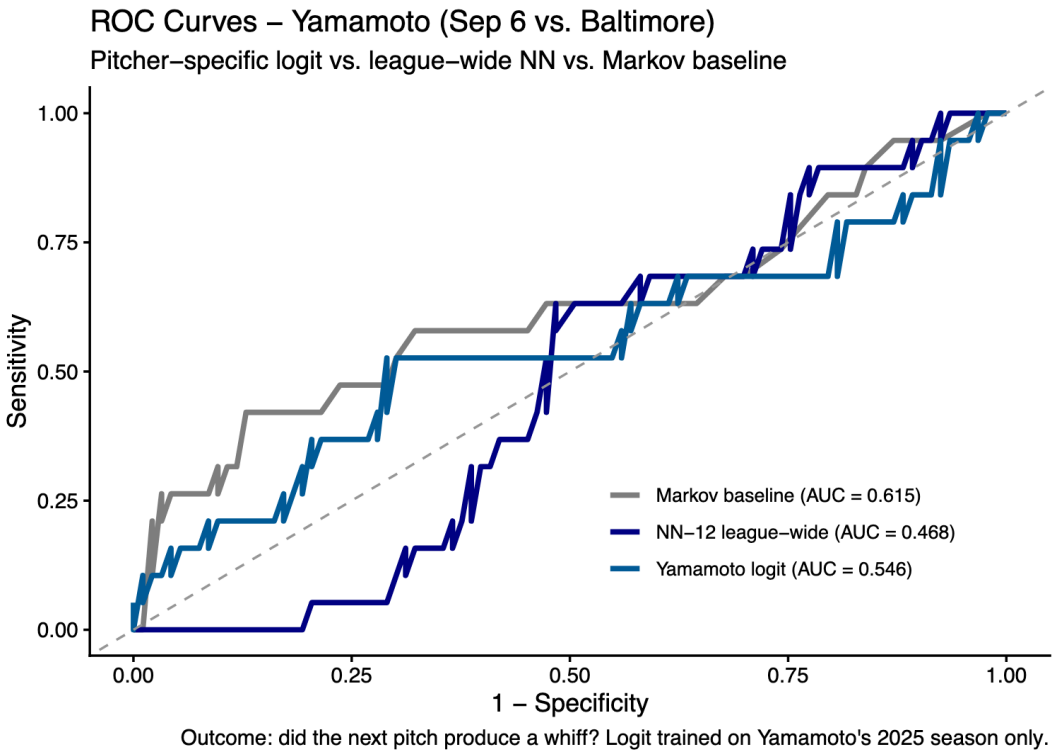


Figure 65: ROC curves for the Markov baseline, league-wide NN-12, and pitcher-specific logistic model applied to Yamamoto's September 6 start vs. Baltimore. The logit was trained on Yamamoto's own 2025 season data and is better calibrated to his actual whiff rate.

The ROC curve reflects this calibration advantage. The pitcher-specific logit achieves much higher AUC on the September 6th holdout (0.546), outperforming the NN-12 (0.468) but not the Markov baseline (0.615). The Markov baseline's performance may be due to the generally small sample of whiffs and the overcomplexity of the other models. The logit's odds are also fully interpretable: release speed is the strongest predictor (OR = 1.118, p = 0.037), meaning each additional mile per hour in velocity increases the whiff odds by approximately 11.8%. This makes sense as Yamamoto's fastball consistently sits at 96 miles per hour. Vertical movement (pfx_z) has a negative odds ratio, reflecting that pitches with more downward break—primarily his splitter—are harder to make contact with when the batter was previously late on a high fastball but easier to identify when thrown in isolation. This coefficient

interpretations are only possible because the model was trained on Yamamoto's own repertoire and not the entire league population.

When examining his top 15 pitches ranked by logit whiff probability, the highest rated whiff is a first-pitch slider thrown low to zone 14 in the 9th inning (17.7%). This same pitch was given a 15% chance by the Markov model and just an 11.1% chance by the NN-12. The logit model is capturing that Yamamoto opening an at-bat with a slider low and away—given his specific velocities and movement profiles—is historically associated with elevated swing-and-miss rates for him specifically, context the Markov model's raw zone transition probability can not encode. This pitch, however, did not produce a whiff illustrating that even well-calibrated probabilistic rankings will not perfectly predict individual outcomes. The rest of the top 15 are dominated by splitters and sinkers and pitches to zones 13 and 14, consistent with Yamamoto's terminal-pitch tendencies described earlier.

Discussion

This independent study examined four interconnected questions in modern pitch analysis: the effect of calling pitches from the dugout on game outcomes, the effect of count on pitch selection and contact quality, whether pitch tunneling drives measurable improvements in strikeout rate, BABIP, and whiff rate, and whether neural networks can outperform first-order Markov chains in predicting pitch-level outcomes.

The Miami Marlins analysis successfully identified visible pitch-mix shifts at the individual level, with deliberate changes to arsenals observable for young stars like Edward Cabrera and Ryan Weathers, as well as established veterans such as Sandy Alcantara. These changes worked both to diversify movement profiles and to amplify the effect of each pitcher's best underused offering.

While the sample size was small and statistical power was low, the results still were encouraging with the Marlins' team-wide whiff rate increased and their BAA decreased. The transition matrix shifts highlight deliberate, situation specific shifts in pitch sequencing decisions after September 19th rather than passive arsenal usage. The count-level plots further hint that the Marlins are intentionally optimizing how they open and close at-bats, with a particular emphasis on 0-0 and 2-2 counts.

The Marlins analysis represents perhaps the most practically consequential thread of this research, and I look forward to reproducing these results at the end of the 2026 season using a full year of Marlins data from their dugout pitch-calling era, with the goal of identifying statistically significant patterns. While the nine-game sample lacks the statistical power to confirm or deny the effectiveness of dugout pitch calling, the directionality of the relevant

metrics points in the same direction, highlighting more whiffs, a lower batting average against, and deliberate sequencing changes consistent with modern tunneling principles. Moreover, there are several confounders that limit causal interpretation. Opponent quality cuts in both directions here: the Marlins faced the Texas Rangers, Philadelphia Phillies, and New York Mets in the last 9 games of the season respectively. The Mets and Phillies are divisional rivals of the Marlins and had already matched up with them 16 prior times each that season. Additionally, while the Phillies had clinched the second seed in the National League and a first round bye in the playoffs, the Mets and Marlins were locked in a tight playoff race for the last wildcard spot. Likewise, the Rangers missed the playoffs but were in playoff contention late into the season. These were not weak opponents, which means that the observed improvements in batting average against and whiff rate can not be dismissed as a product of facing inferior lineups. However, this also introduces a different confounder: all three opponents had strong motivation to perform in late-September (either to make the playoffs or secure a first-round bye and home field advantage), making the Marlins' results against them a particularly demanding test that may not generalize well to a full-season sample. Additional uncontrolled factors include late-season callups, which can alter a pitching staff's composition and workload distribution, as well as the possibility that individual pitchers were likely managing fatigue or pushing their arms to try and make the playoffs with the season winding down. Any of these could independently shift pitch mix and outcomes on a game-to-game and even matchup-to-matchup basis in ways unrelated to the dugout-calling system.

The count effects analysis successfully replicates and extends Jim Albert's findings from the 2015 season to the 2025 season. Fastball usage rises sharply in hitter-friendly counts, breaking balls and offspeed pitches dominate pitcher-friendly counts, and that low-and-outside

remains the premier whiff-generating location across all count types. The new finding for 2025 is that three ball counts (3-1, 3-2) are more likely to result in a strikeout or put-out than they were 15 years ago, as hinted at by their average negative delta run expectancy. This suggests that modern pitchers have become meaningfully better at converting deep counts into strikeouts—a trend consistent with the higher velocity and arsenal expansion documented in the introduction.

The leverage-weighted analysis added another dimension: top relievers in baseball, such as Aroldis Chapman (BOS), Mason Miller (SD), Robert Suárez (SD now ATL), and Carlos Estévez (KC), and ace starters such as Tarik Skubal (DET), Zack Wheeler (PHI), Paul Skenes (PIT), Christopher Sánchez (PHI), Hunter Brown (HOU), Garrett Crochet (BOS), and Yoshinobu Yamamoto (LAD) all rank in the top 40 of qualified pitchers for either their Leverage-Weighted Run Value score, Clutch score, or both, validating their dominance is not limited to low pressure counts but holds precisely when counts matter more.

The pitch tunneling analysis also successfully replicated and extended Jon Roegele's findings from "The Effects of Pitch Sequencing." The Roegele matrix confirmed the ideal tunneling zone—low-to-moderate tunnel distance combined with high plate distance—still generates the highest whiff rates, and cross-pitch pairs remain far more effective than same-pitch pairs at any given tunnel proximity. The matrix shows that for a given tunnel distance, whiff rate increases as plate separation increases and that the closer consecutive pitches are at the swing decision point, the less plate separation they need to generate elevated swing-and-miss rates. These findings have major implications for MLB pitchers, specifically that by ingraining the tunneling effect into their repertoire they should be able to increase the deceptiveness of their pitches when considered in a sequence, which will naturally lead to more whiffs. If pitchers throw pitches that have similar arm slots, but very large plate distances than theoretically adding

a tunneling effect would make these pitches harder to identify and increase their effectiveness. The direction of the scatterplot relationships is consistent with Roegele's hypothesis: smaller tunnel distance correlates with higher K% and whiff rates and with lower BABIP and hard-hit rate.

However, at the pitcher-season level, none of the three tunneling metrics—tunnel distance, plate distance, and tunnel ratio—achieved statistical significance in any of the OLS models fit separately for each predictor. This suggests that tunneling is more actionable as a sequence level tool than as a pitcher-wide summary statistic: the effect is real in the transition matrix but averages out across an entire arsenal when pitchers considerably vary how and when they deploy each sequence type. The regressions did, however, show that arm slot was a statistically significant predictor of K% with a middling negative directionality, suggesting pitchers with lower arm angles—side-armers and submariners—tend to generate slightly elevated strikeout rates. This likely represents an independent depiction mechanism distinct from geometric tunneling, and future work should control or subset by delivery angle before drawing conclusions about tunneling's effect on strikeouts.

The Yamamoto case study illustrates the main limitation of tunneling metrics cleanly. His tunnel distance and tunnel ratio are both essentially league average, yet he ranks very highly among qualified pitchers in both K% and whiff rate. His success derives from elite pitch quality—velocity, movement, and command—rather than geometric deception. The comparison to deGrom (tighter tunnel distance, similar elite outcomes) and Fedde (looser tunnel distance, substantially worse outcomes) further supports the conclusion that tunneling amplifies pitch quality but cannot substitute for it.

The final pitch analysis was done on pitch sequencing. This analysis creates 5 distinct clusters of pitchers whose sequencing behavior and zone transition patterns are most similar and the ANOVA tests showed that spin rate, velocity, and whiff rate differed meaningfully depending on which cluster they belong to. Shannon entropy is highly significant across clusters with clusters 3 and 4 having the highest average entropy and cluster 2 having the lowest. This matches Clay's finding that the highest-performing group wasn't necessarily the most unpredictable.

We then employ neural networks to determine whether they can outperform Markov chains with respect to predicting whether the next pitch will be a whiff. The basic neural network as well as the extended predictor neural network had high accuracies as they were incorrectly classifying every pitch as a non-whiff due to the extreme class imbalance, however, their AUC scores were above that of the Markov chain. The balanced neural network had lower accuracy and AUC compared to the imbalanced models, however it also outperformed the Markov chain. It is important to note that the all AUC values across every model tested remained only marginally above the 0.50 baseline of random guessing. The practical ceiling on predictability here is likely low: whiffs occur only on 11% on pitches league-wide, the outcome depends heavily on real time physical execution and batter reaction that no sequence model can reasonably observe, and the most important predictors are noisy proxies for what actually determines swing-and-miss. The improvement from adding count context and sequence depth over the Markov baseline is real but small, and future work should treat AUCs in this range as evidence that useful signals exist rather than as evidence that the model is ready for operational use.

In Yamamoto's case study, we saw his heavy reliance on four-seam fastballs and splitters, whose permutations were his most used sequences in 2025. This makes sense as his fastball and

splitter have nearly the exact same tunnel location and arm slot, meaning they are maximally deceptive. These three pairings were also the three most used on September 6th against the Orioles. This is also Yamamoto's primary strategy against lefties, where against righties he throws far more sinkers and sliders. Following his stellar outing, Yamamoto also changed the frequency with which he threw certain sequences for the rest of the season, throwing fewer cutters and more curveballs. Yamamoto's sequencing on September 6th was optimally unpredictable, making 112 zone transitions, using 61 unique from-to zone pairs. 59% of those pairs appeared exactly once and a maximum of 7 repetitions was the most among any single transition. He worked effectively to expand the zone low and away throughout the game, finishing at-bats against lefties with splitters and against righties with fastballs.

When the neural network was applied to Yamamoto's September 6th game to rank pitches by whiff probability, the central finding was that the pitcher-specific logistic regression model substantially outperformed the league-wide NN-12 in terms of both calibration and AUC. The whiff rate of the September 6th game was 20% (or 17% by our definition) while the league-wide whiff rate was 11%. Applying a league-wide model to score a specific pitcher's game is a fundamentally misspecified problem. Pitcher-specific models, even relatively simple ones, will outperform more complex league-wide architectures when the pitcher's distribution differs substantially from the training population, such as Yamamoto's elite whiff rate and more generally, elite pitchers whose outcomes are by definition outliers relative to the league average.

Taken together, these analyses describe a layered, mutually reinforcing system. Count determines what pitch a hitter expects; tunneling determines whether they can distinguish the pitch from its alternatives in time to act; and sequencing determines whether the entire at-bat structure is predictable or not. Future research should examine these three dimensions

simultaneously, ideally with individual batter-pitcher matchup data to control for opponent quality.

Conclusion

In conclusion, this independent study analyzing the effectiveness of the Marlins new pitch-calling method, the effect of game state, tunneling, and pitch sequencing on many outcome based metrics as well as providing case studies of specific pitches, game states, and pitchers answered many of our original research questions.

First, we assessed the results of the Marlin's top-down dugout pitch-calling system by conducting a series of analyses on both pitch-level and sequence-level data from the Marlins 2025 season, paying specific attention to their last 9 games of the season and comparing the relative outcomes of the two time-frames (before and after the transition). Overall there were no detectable negative outcomes on the Marlins' results, there was positive directionality in numerous important pitch metrics such as run value and whiff rate, and noticeable changes to pitch repertoire for many prominent pitchers that will be interesting to track over the course of the 2026 season. The primary constraint of the Miami Marlins analysis is the sample size disparity. Nine games is simply insufficient to draw conclusions from and there were also uncontrolled confounders such as opponent quality, park factor, and input from front office executives. For a properly powered study, we should redo this analysis after the Marlins (or any other team that adopts this methodology) use this system for an entire season.

Next, we looked at, given a specific count, location, or pitch type, what the optimal pitch to throw for maximizing the chances of a whiff. We also considered the effect of the at-bat count on leverage, pitch selection (fastball vs breaking ball usage), run value, launch angles, exit velocities and quality of contact. The results of the count effects analysis was mostly consistent with the results found by Jim Albert in his original blog, with the new results that counts with

three balls are more likely to result in a an out than they were 15 years ago and that the top relievers in baseball are those that have the lowest leverage-weighted average delta run expectancy per pitch and clutch scores, meaning that they perform better than your average pitcher in high leverage situations. An extension of this analysis that would be eye-opening would be to consider the effect of pitch count on the effectiveness of different pitch types throughout the game. An example of this could be how does the number of fastballs thrown by a pitcher in a game affect their effectiveness as the game progresses.

In the pitch tunneling analysis we tried to understand whether or not pitchers who cluster their arsenals more tightly at the swing-commitment point generate higher strikeout rates, higher whiff rates, lower BABIP, lower hard-hit rate. The pitch tunnel analysis was fruitful in identifying a strong correlation between whiff rate and various tunneling metrics. Overall, it was found that pitchers whose consecutive pitches cluster more tightly at the 23.8-foot swing decision point (low tunnel distance) and have large separation by the time they've reached the plate (high plate distance) are associated with higher whiff rates. Additionally, for a given tunnel distance, larger plate separation generates higher whiff rates. When considering tunnel ratio (plate distance / tunnel distance) we see that elite tunnelers (higher tunnel ratio) have a higher whiff rate when coupled with higher average velocities. This analysis suggests that tunneling is an effective tool that can and should be employed by pitchers to increase the deceptiveness of their pitches, and is a tool that has the capacity to increase whiff rates when paired with above average velocity. In the regression analysis, we also identified arm angle as a significant predictor of K%. A potential extension of this analysis would be to examine whether different bucketed ranges of arm angles enhance or diminish the effects of pitch tunneling, particularly

when two relief pitchers on the same team with drastically different arm angles appear consecutively.

The Yamamoto case study threads through all four analyses and provides a concrete picture of how count effects, pitch tunneling, and pitch sequencing interact for an individual elite pitcher. In the count effects analysis, we showed that Yamamoto has on average a negative run value for plate appearances passing through higher leverage counts, showing that he performs well in big moments and high pressure situations. In the tunneling analysis, we saw that his metrics are essentially league average, confirming that his outcomes derive from elite pitch quality and less so from geometric deception. In the sequencing analysis, his cluster assignment and low Shannon entropy confirm that his sequencing behavior is actually not as unpredictable as previously thought but that predictability is not exploited because of the exceptional quality of his pitches. His September 6th game saw 59% of his transition pairs appear exactly once.

The most methodologically significant finding from the Yamamoto neural network scoring is the distribution shift problem. League-wide models trained on 11% whiff rates are poorly calibrated for an elite pitcher whose whiff rate is closer to 20%. The pitcher-specififc logistic regression—a far simpler model—outperfromed the league-wide NN-12 in terms of calibration and AUC by training on Yamamoto's own season data. This suggests that future pitch-scoring applications should prioritize pitcher-specific calibration over model complexity and that league-wide models may be limited by their own complexity when the goal is to evaluate an individual pitcher's specific outing.

Works Cited

Adler, David. “MLB Pitch Arsenals Are Bigger Than Ever in 2025.” MLB.com, 25 Mar. 2025, https://www.mlb.com/news/mlb-pitch-arsenals-are-bigger-than-ever-in-2025.

Albert, Jim. “Count Effects.” Bayesball Blog, 2023, https://bayesball.github.io/BLOG/Count/_Effects.html.

Allen, Jim B. “Japan’s Dead Ball Is Likely a Mistake.” jballallen.com, n.d., https://jballallen.com/japans-deadball-is-likely-a-mistake/.

Boan, Chris. “MLB Pitch Speed and Usage: 2002–2025.” RotoWire, 15 Jan. 2025, https://www.rotowire.com/baseball/article/mlb-pitch-speed-and-usage-2002-to-2025-94262.

Clay, J. Z. “Strategic Pitch Location: The Role of Two-Pitch Sequences in Pitching Success.” SABR Baseball Research Journal, vol. 52, no. 1, 2023, https://sabr.org/journal/article/strategic-pitch-location-the-role-of-two-pitch-sequences-in-pitching-success/.

Judge, Jonathan, Harry Pavlidis, and Jeff Long. “Prospectus Feature: Introducing Pitch Tunnels.” Baseball Prospectus, 17 Aug. 2015, https://www.baseballprospectus.com/news/article/31030/prospectus-feature-introducing-pitch-tunnels/.

Long, Jeff. “Pitch Tunneling: Is It Real and How Do Pitchers Actually Pitch?” The Hardball Times, FanGraphs, 2015, https://tht.fangraphs.com/pitch-tunneling-is-it-real-and-how-do-pitchers-actually-pitch/.

"Pacific League Midseason Check-In." JapanBall, Aug. 2025, https://japanball.com/japanese-baseball-news/npb-newsletter/pacific-league-midseason-check-in.

Roegele, Jon. "The Effects of Pitch Sequencing." The Hardball Times, FanGraphs, 16 Apr. 2014, https://tht.fangraphs.com/the-effects-of-pitch-sequencing/.

"Statcast CSV Documentation." Baseball Savant / Statcast, MLB Advanced Media, 2025, https://baseballsavant.mlb.com/csv-docs.