

The Future of Challenge Strategy: Insights from ABS in Triple-A

Maxfield Lane and Owen Riley

We explore the triumphs and pitfalls of teams' use of the ABS challenge system in 2024 in Triple-A, providing a sneak peek of MLB's future

The automated ball-strike (ABS) challenge system has finally (sort of) hit the bigs! Spring Training has been many fans' first introduction to the system, in which hitters, catchers, and pitchers can challenge the result of pitches, triggering a ruling by an automated system that uses pitch data to determine the true location of the pitch and make the proper call. It's increasingly clear that the question with ABS is when, not if, it will be implemented in real MLB games. That, plus all the media hype around it that we're shamelessly slipstreaming, makes it the perfect time to dive into the Triple-A data from 2024 to conduct a wholesale analysis on challenge usage. The ability to correct an umpire's call is a powerful tool. As soon as challenges are properly introduced, organizations that use them thoughtfully and efficiently will reap substantial benefits.

So, here's our first attempt at a playbook for organizations to use when considering challenge strategy. By outlining success rates and usage by situation, pitch type, pitch location, and other leverage factors, we've assembled an 8-pack of ABS insights. As is our MO, we'll spoil the basics here before delving deeper.

- 1. Hitters and Batteries Challenge with Roughly Equal Frequency and Success**
- 2. Teams Hoard Challenges, Then Scramble to Get Rid of Them**
- 3. Teams are Too Cautious!**
- 4. Don't Get Tilted! Fear of Failure Leads to Bad Decisions**
- 5. Leverage Matters**
- 6. Hitters Take the *High* Road, Keep Away from the "*Slots*"**
- 7. Changeups Flummox Catchers**
- 8. Good Eyes, a Good Challenger Does Not Make**

A Quick (or not) Note on Methodology and the Challenge System

Our data consists of most Triple-A games post-June 25, 2024, when both Triple-A leagues went full challenge system. In this period, the Pacific Coast League (PCL) capped challenges at three failed challenges per game, while the International League (IL) had a two failed challenge per game limit. If successful, a team could theoretically challenge infinitely; only failures count toward a team’s allotment. For the dataset, we combined Retrosheet data, which contained challenge info, with Statcast data, which we utilized for pitch location data, player data, and a host of other data points. Six games failed to pull from Statcast, and there are extremely rare occasions in which rows failed to match properly. We also excluded all pitches under 70 mph, as position player pitcher eephus pitches had a disproportionate effect on our dataset.

When evaluating pitches that were not challenged, but perhaps could have been, we utilized a ball-radius-sized buffer zone (1.44 inches), only calling pitches that were outside of this buffer zone as “missed opportunities to challenge.” This was necessary due to differences in measurement mode between Statcast and the ABS system. While the ABS system tracks pitches at the center of the plate (8.5 inches from the front and tip of the plate), Statcast looks at pitches as they cross the front of the plate. This can lead to mismatches in expected calls, as, for instance, a curveball with a substantial break may be a strike at the front of the plate but too low by the time it reaches the middle ABS zone. Fortunately, we were able to establish the rough bounds of this grey area and removed the vast majority of mismatches from our “missed opportunity” pool.

Now, let’s get to our 8-pack of ABS findings.

1. Hitters and Batteries Challenge with Roughly Equal Frequency and Success

We’ll start things off with the broadest possible questions: How are challenges split between batteries and hitters, and how successful are challenges?

Challenge Breakdown by Challenging Side

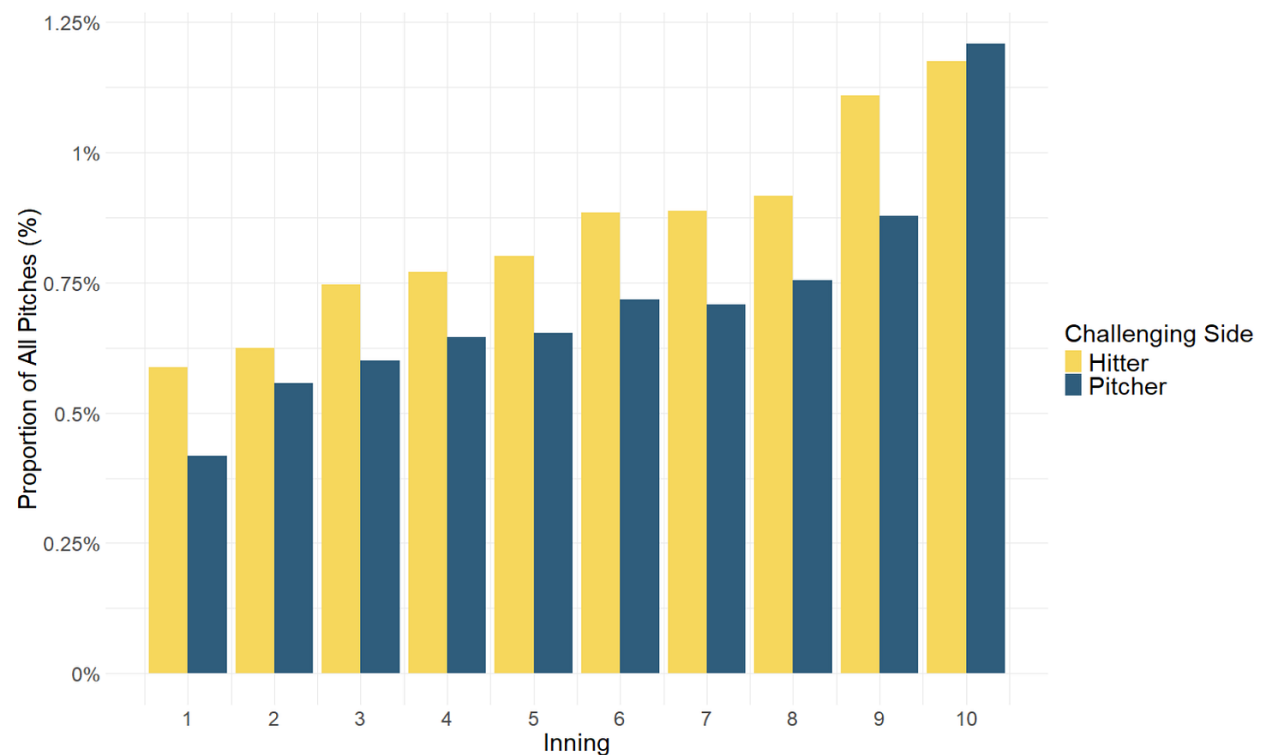
Challenging Side	Success Count	Failure Count	Success Rate	Proportion of Challenges
Catcher/Pitcher	1149	1069	51.8%	47.6%
Hitter	1323	1420	48.2%	52.4%
Overall	2472	2489	49.8%	100%

Batteries challenge slightly less than hitters, accounting for 47.6% of total challenges. For both sides, the success rate on challenges is right around 50%, with catchers/pitchers succeeding on just over half of their attempts and hitters on just below half. We were a bit surprised by this result, as we expected catchers' better perspectives (catchers make the vast majority of battery challenges) to yield vastly higher success rates than their hitting peers, but this is not the case. While the 3.6 ppt advantage catchers hold is statistically significant at the $p < .05$ level ($p = .01237$), it's hardly a massive gulf in hit rate.

2. Teams Hoard Challenges, Then Scramble to Get Rid of Them

One of the clearest trends in existing challenge data is that the frequency of challenges increases as the game continues:

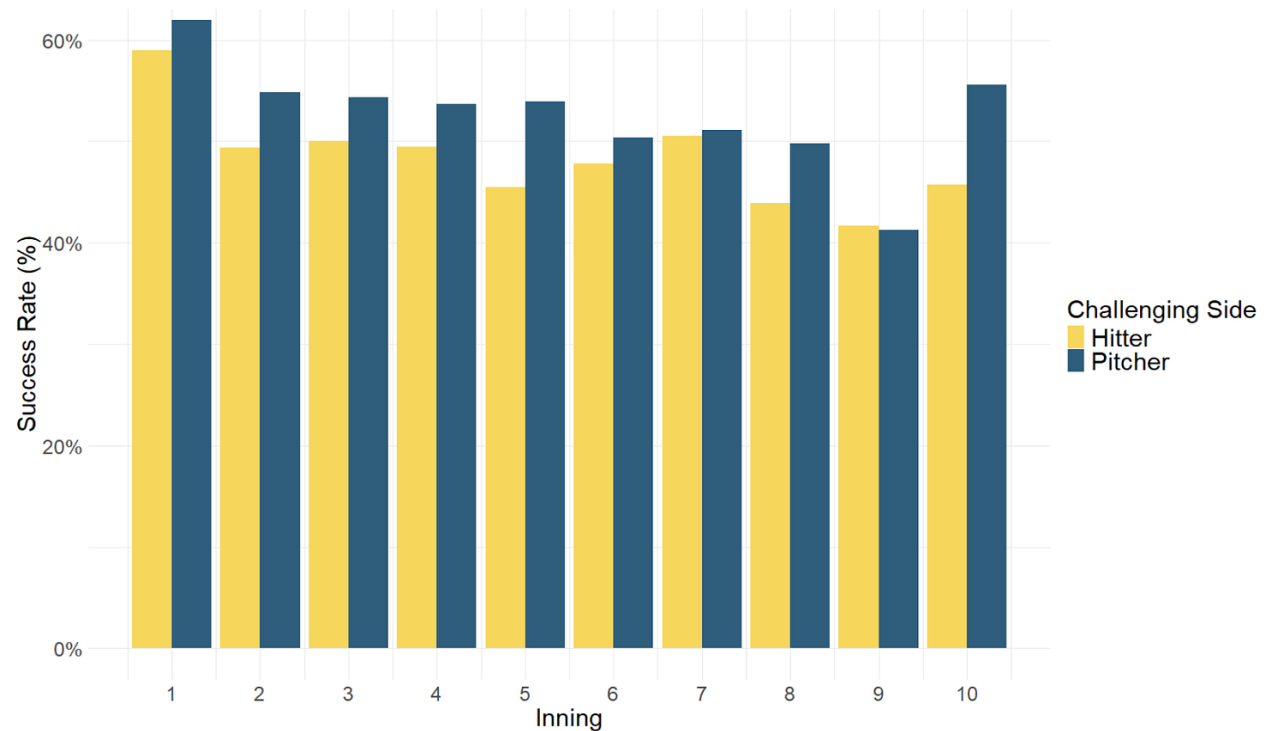
Challenges by Inning as a Proportion of All Pitches



Teams love to save their challenges, challenging around twice as often in the ninth inning as in the first, and steadily increase the frequency of their challenges as the innings go by. Again, this aligns with prior expectations, as teams are likely saving their challenges for key situations and are forced into late-game aggression if they're sitting on spare challenges.

Later in the game, flush with challenges and often in higher leverage situations, teams appear more willing to risk being wrong in making a challenge:

Challenge Success Rate by Inning



Success rates on both sides of the ball in the ninth inning hover just above 40%, compared to around 60% in the first. We like calling these last-ditch, usually unsuccessful challenges “vanity challenges.”

3. Teams are Too Cautious!

If you’re a regular Down on the Farmer, [you’ve heard this one before](#). In theory, saving challenges for high-leverage situations may be advisable; however, the data below illustrate that teams are hoarding challenges to their detriment.

Challenges Per-Team Per-Game Usage by League

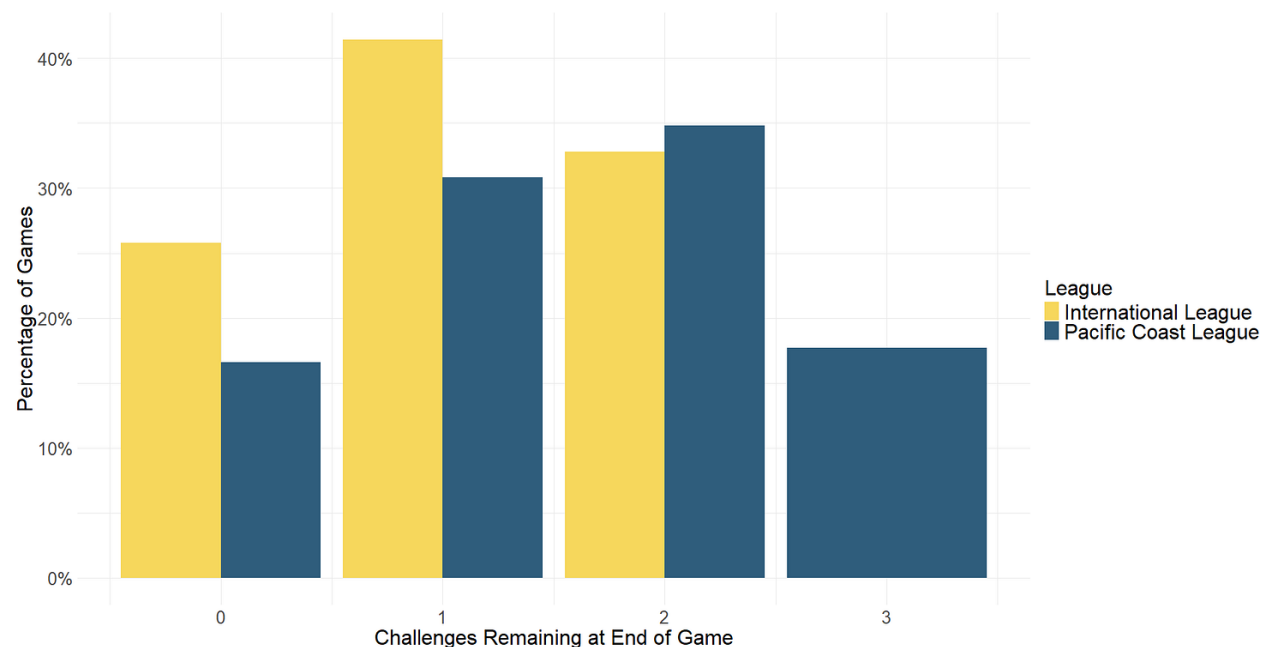
League	Mean	Median	SD	Min	Max
PCL (3 failures P/G)	2.94	3.00	1.43	1.00	10.00
IL (2 failures P/G)	2.18	2.00	1.11	1.00	7.00

**The big winner is the Sugar Land Space Cowboys, who went 8 for 10 in the only 10 challenge game. Even took a challenge home!*

Those means/medians are low! Even if every challenge was unsuccessful, teams are *still* below their total challenge allocation, indicating immense challenge waste. If teams finish games without utilizing all their challenge opportunities, there's a high likelihood that they're missing opportunities to challenge. The below chart illuminates this idea

(NOTE: this chart originally included only home teams in error, and the chart and metrics have been adjusted accordingly. There was a < 1% shift in the data):

Challenges Remaining at Conclusion of Game by League



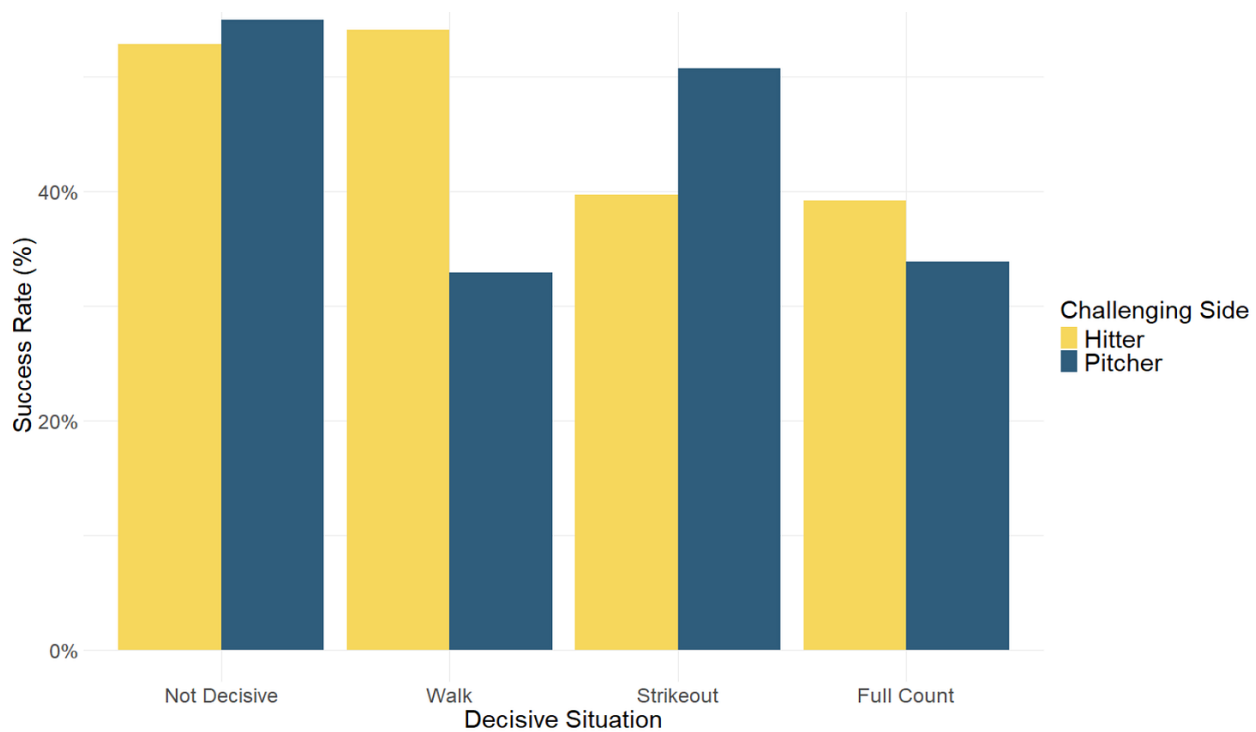
Shockingly, IL teams only utilize all their challenges in 25.8% of their games, while PCL teams are even more stingy, only using all their challenges in 16.6% of contests. In fact, teams are more likely to end a game without failing a challenge than running out of challenges, a testament to a team's willingness to save their opportunities. Even if missing a challenge [causes embarrassment and shame](#), teams have wiggle room to be more aggressive in their challenge philosophies.

One possible team-friendly explanation is that teams already take advantage of all possible challenge opportunities, but the data refutes this notion. Excluding successful challenges, IL contests average 2.53 unchallenged incorrect calls per game in which the call was wrong by more than the radius of a baseball (1.44 in), while PCL games average 2.24 of these blown calls per game. While it is difficult to get more granular with missed calls due to the Statcast zone measuring the front of the plate, while the ABS utilizes the spot 8.5 inches from the front of the plate, there's still a clear indication that teams are leaving meat on the bone when it comes to challenge usage. Teams can't be perfect but should be less afraid to fail in their challenge efforts, especially in early-game situations.

4. Don't Get Tilted! Fear of Failure Leads to Bad Decisions

Two of the three true outcomes are in play in ABS challenge situations, and these "decisive calls" in which a ball four and/or a strike three is in play offer high-leverage opportunities to change the game. Speaking from personal experience (as we all know the best evidence is anecdotal evidence), as a hitter, I hated striking out more than I liked walking, and as a pitcher and catcher, I hated allowing a walk more than I liked getting a K. Unfortunately for our challengers, they seem to be reacting similarly, with challenge success rates plummeting when failure is on the line, as displayed below:

Challenge Success Rates by Situation Type



When players are challenging for a good result (P/C for a K, Hitter for a BB) their success rates are similar to those in lower-leverage situations. However, when challenging to avoid a bad result (walks for P/C and Hitters for a K), their success rate drops by 15%-20%. These results aren't just bad luck; they're a result of worse challenge decisions.

For example, when hitters challenge to attempt to win a walk, their failed challenges have .89 inches of in-zone cushion. When challenging to avoid a strikeout, hitters are far worse, with their failed challenges averaging 1.20 inches of in-zone cushion, around a third of an inch worse than on walk bids. The same is true of catchers, who average a miss distance of 0.62 inches on failed strikeout challenges and a miss distance of 0.84 inches on failed walk challenges.

While we still lack the mind-reading machine we mentioned in our baserunning article (we're working on it), it seems realistic that frustration with an AB ending with a bad outcome promotes more aggressive lower-quality challenges. But, even though these decisions may be anger-driven, is more aggressive necessarily a bad thing?

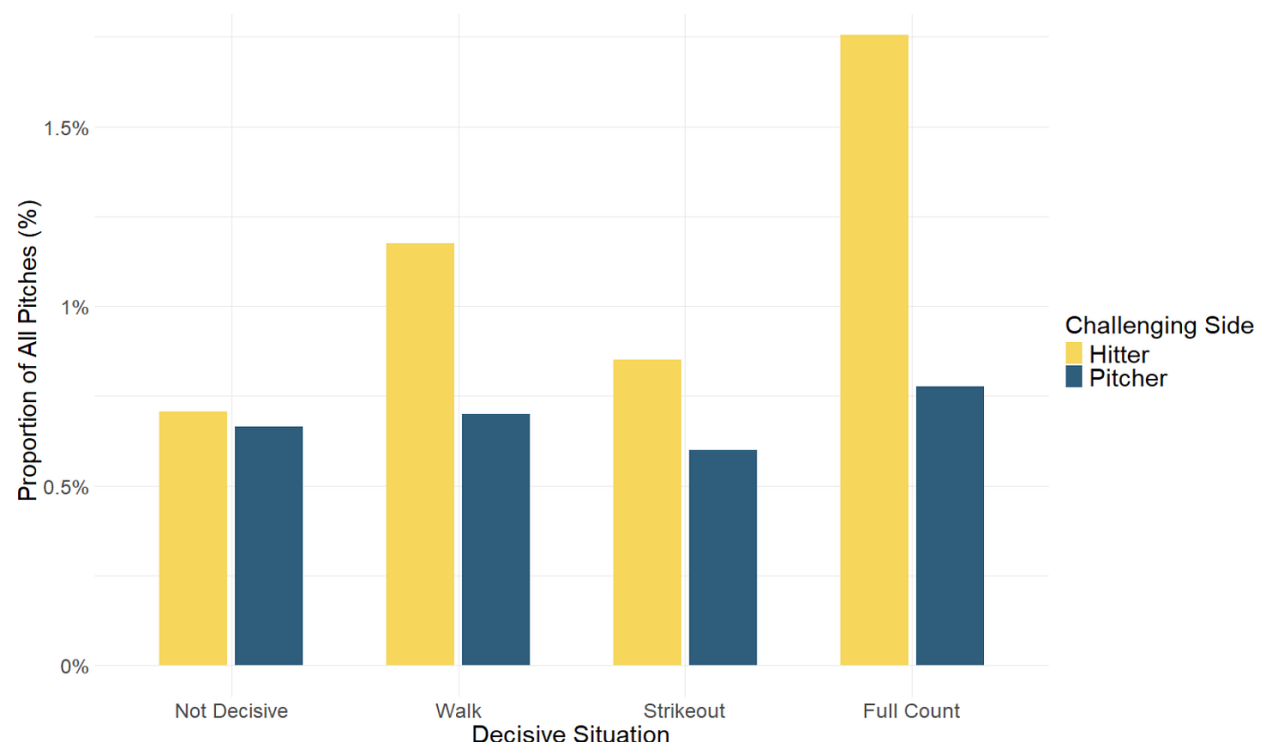
5. Leverage Matters

Despite the lower success rates, it's important to remember that these decisive situations have far higher leverage when compared with their non-decisive peers, potentially making lower success rates worth their heightened reward.

Based on [work by Sunyvale inspired by Tom Tango \(Tangotiger\) on run expectancy by situation](#), with the bases empty and no outs, going from 1-2 to 2-1 or vice versa via a successful challenge causes a 0.13 run swing in run expectancy. However, going from 3-2 to a BB via a successful challenge ups run expectancy by .30 runs, and 3-2 to a K via a successful challenge decreases run expectancy by 0.34 runs. Full counts in this situation have the most significant shifts, accounting for a swing of 0.64 runs with the bases empty and no outs. The leverage differences between these situations are stark; it would take around five successful 1-2 to 2-1 challenges to be of similar worth to a full count challenge in this situation.

To the teams' credit, they do a solid job of taking advantage of blown calls in decisive situations, with only 0.39 unchallenged, decisive missed calls by a ball radius or more per IL game and 0.28 per PCL game. In key situations, teams should be risk-happy and should challenge any close calls, especially in full counts. Teams already do some of this, as shown below:

Challenge Frequency by Situation Type



Teams challenge full counts almost twice as frequently as in non-decisive situations, clearly understanding the high leverage of those high-impact pitches. Interestingly, nearly all of the variance in challenge frequency is hitter-driven, with pitchers rarely shifting their challenge behavior in high-leverage situations. Despite hitters challenging far more than pitchers in full counts, they're still more successful than their pitching peers (39.2% vs. 33.8%).

A great way to track challenge effectiveness on a team basis would be to monitor the number of unchallenged decisive blown calls above some distance threshold, like our ball radius standard. Keeping this percentage as close to zero as possible should be the goal!

We'll expand on the run expectancy aspect of challenges in a future piece, as this area demands a more comprehensive analysis. Still, we wanted to highlight the importance of leverage in challenge decision-making.

6. Hitters Take the *High* Road, Keep Away from the "Slots"

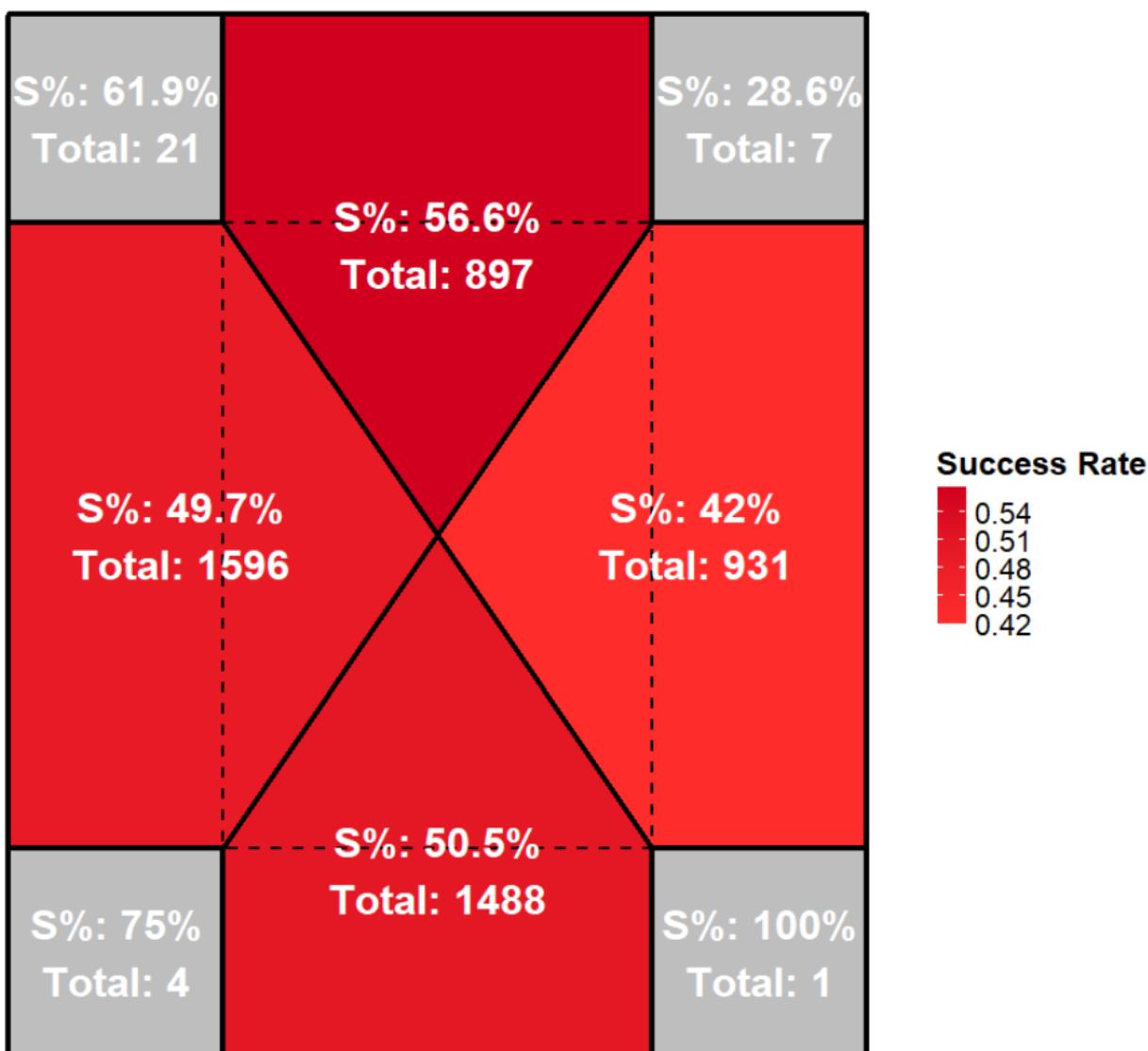
This isn't just life advice coming from a Reno native! Below is a table with challenge success rates by "closest edge," which is the closest side of the strike zone to a given pitch for both balls and strikes, adjusted by handedness. 33 of these pitches missed the zone on two planes, but the closest is listed here.

Challenges by Closest Edge

Closest Edge	Success %	Count
High	56.6%	897
Inside	42.0%	931
Outside	49.7%	1596
Low	50.5%	1488

And here's the same chart, visualized to the strike zone (right = inside):

Strike Zone by the Four Closest Edges (right = inside)



High pitches are overturned around 15% more frequently than inside pitches and around 7% more than inside and low pitches, making challenging the high ones and avoiding the inside ones an advisable strategy. The inside vs. outside pitch disparity matches our expectations, as umpires align themselves with the inside part of the plate and have a vision advantage in this area (called the “slot”). Umps call what they can see, and inside pitches are directly in their line of sight.

Here’s a similar breakdown, but accounting for pitcher and hitter challenges:

Challenges by Closest Edge and Challenging Side

Closest Edge	Challenging Side	Success %	Count
High	Hitter	60.2%	517
High	Pitcher	51.7%	389
Inside	Hitter	41.0%	462
Inside	Pitcher	43.0%	472
Outside	Hitter	46.2%	904
Outside	Pitcher	54.5%	712
Low	Hitter	46.8%	851
Low	Pitcher	55.6%	638

Notably, the overall elevated success rate on high pitches is nearly exclusively hitter-driven, whereas pitchers can find similar success on low and inside pitches. Inside pitches have a similarly low success rate. We were a bit surprised by both hitters' success on high pitches and pitchers' success on low ones, as catchers seem to be better at framing low balls and worse at high pitches, but this potential difference in deception doesn't surface in the challenge results.

Exposing players to this data could be a great way to encourage positive challenge behavior; if hitters know that they can experience higher success challenging high strike calls, they may feel more inclined to challenge these pitches and hopefully bump up their overall accuracy.

7. Changeups Flummox Catchers

Next, we'll bring pitch type into the mix. Is there any difference in fastballs vs. offspeed? See for yourself below, ordered by success rate:

Challenges by Pitch Type (Challenge Count >= 50)

Pitch Type	Success %	Challenge Count	Challenge Rate by Pitch
Cutter	56.1%	446	1.41%
Sweeper	51.8%	220	1.45%
Curveball	51.4%	385	1.61%
Splitter	51.0%	51	1.03%
Sinker	50.3%	889	1.82%
4-Seam Fastball	49.0%	1700	1.57%
Slider	48.3%	818	1.33%
Changeup	45.8%	382	1.01%

Pitch type, for the most part, has a limited effect on challenge success, with the middle six pitches on the chart falling within a 3.5% range. The two notable exceptions are the cutter, which has a 4.3% higher success rate than the second-highest pitch, and the changeup, which has by far the lowest success rate of all pitches. Further analysis shows little difference between hitter and pitcher success rate on cutter challenges (56.4% hitter success, 55.5% pitcher success), but a large difference in changeup challenge success, with hitters successfully challenging 48.9% of changeups while catchers only succeeded on 42.2% of challenges.

The heightened success on the cutter has us stumped; we're not sure why players are great at judging these pitches when compared to 4-seamers and sinkers. Location doesn't seem to be a factor, as cutters have similar location profiles to curveballs, which have a more normal success rate. Any theories are welcome!

The low success rate with changeups and large pitcher/catcher vs. hitter difference indicates a clear flaw in how catchers perceive the zone (they're bad at judging low, outside, and high changeups). Are catchers deceiving themselves with their good framing work on late-fading changeups? Are they thrown off by the last-second movement that drags the pitch out of the zone before reaching the ABS zone?

8. Good Eyes a Good Challenger Does Not Make

Who's got their challenge chops, and who's an umpire ego inflater? It stands to reason that players with high walk rates and low strikeout rates are likely to be the best challengers, as

their improved understanding of the zone should lead to better challenge decisions. Shockingly, this is untrue! With hitters with three or more challenge attempts and catchers with 10 or more attempts, there was no statistically significant correlation between walk rate or strikeout rate and challenge success rate (p-values of .785 for BB% and .136 for K% for catchers, .169 for BB% and .404 for K% for catchers). This was a great surprise to us, but looking at the best hitter challengers backs up the absence of a correlation:

Top Hitter Challengers by Success Rate (Challenge Count >= 10)

Name	Successful Count	Total Challenges	Success Rate	Missed Opportunities (> 1.44 in)	2024 K% (All Levels)	2024 BB% (All Levels)
Nick Sogard	9	10	90.0%	4	21.9%	13.5%
Spencer Torkelson	13	16	81.2%	8	29.0%	11.4%
P.J. Higgins	12	15	80.0%	7	24.6%	7.9%
Nick Allen	8	11	72.7%	7	12.6%	11.2%
Samad Taylor	7	10	70.0%	8	26.0%	11.3%

[Spencer Torkelson](#), a player known for his swing-and-miss tendencies, was possibly the best challenger of the bunch, succeeding in over 80% of his 16 chances. [Nick Sogard](#), who's a far more patient swinger, was the best percentage-wise, nailing nine of his 10 challenges, while [P.J. Higgins](#), a grizzled veteran catcher, also hit 80% of his challenges. Higgins seems to have his challenge vision calibrated well, as he's also atop the catchers' list!

Top Catcher Challengers by Success Rate (Challenge Count >= 10)

Name	Successful Count	Total Challenges	Success Rate	Missed Opportunities (> 1.44 in)	2024 K% (All Levels)	2024 BB% (All Levels)
P.J. Higgins	9	12	75.0%	7	24.6%	7.9%
Sandy León	9	12	75.0%	5	28.6%	17.2%
Cody Roberts	20	27	74.1%	10	31.9%	10.6%
Luis Campusano	8	11	72.7%	2	14.0%	6.5%
Dillon Dingler	10	14	71.4%	9	23.5%	8.5%

[Cody Roberts](#) is a standout here, succeeding at a high rate with a high volume of challenges. Just missing the list was [Bryan Lavastida](#), who nailed 27 of 38 challenges. [Chad Wallach](#) led our sample in overruns, registering 29 in 53 attempts. The list was striking to us as it includes veteran glue guys (Higgins and [León](#)), D-first talents (Roberts and [Dingler](#)), and hit-over-field types ([Campusano](#)). The diversity of player types in both the hitter and catcher lists further enforces the idea that challenge success is a skill that has different demands than other aspects of pitch recognition.

Areas for Expansion

In the future, our primary goal is to find a way to accurately transpose the Statcast zone onto the further back on the plate ABS zone. Right now, we have a decent number of mismatched pitches: Statcast strikes that drop below or outside the zone before reaching the ABS plane or Statcast balls that fall down or into the zone as ABS picks them up. Achieving this task will allow for K-Zone-style visualizations and a more concrete evaluation of missed opportunities, as we can judge players not only on their ball-radius-sized missed challenge opportunities.

We also want to better expand on the idea of run/ win expectancy. By adding this to the dataset on a pitch-by-pitch basis, we should be able to critique the effect of challenges on winning games in a way that we cannot with the current data. It could also be interesting, if possible, to evaluate umpires. Are some umpires biased to certain tendencies that challenges could exploit? Do umpires improve their zones over the season or within games based on the public shame of players winning challenges?

Final Thoughts

Unlike our other analyses of [baserunning](#) and the effect of [swings on stolen base success](#), there were no longstanding conventions or strategies to interrogate here. Challenging is new for everyone, and at least as of yet people have not siloed themselves into philosophical camps on how to best use them.

But an ideological split is likely on the horizon. It's not hard to see how it might fall along already established lines of data nerds vs. old school analytics skeptics. Already, more traditionally-minded commentators have lambasted their teams for challenging too early, doing their best to establish new rules of thumb that mirror long-established catch phrases like "don't make the first or last out at third." Nerds like us are already firing up minutely calibrated analyses of challenge attempt run expectancies. Our prediction: pretty soon, there will be a battle between these factions about how aggressive to be in challenging, with the analysts (as in sending runners home from second) arguing for more aggressive use of challenges that considers leverage (baserunners, count, score, inning) rather than simply inning alone.

For now, our basic general advice to teams constructing a challenge strategy is relatively simple: challenge in high-leverage decisive counts, don't fret too much about saving your challenges, challenge high pitches as a hitter, avoid challenging changeups as a battery, and avoid challenging inside pitches altogether.

As always, thanks for reading!