

**Intro**

Wanting to define and measure “deception” in MLB, I researched blog posts, articles, and box scores and ultimately decided that the most important aspects of deception are being able to throw different pitches from the same arm slot (not tipping and disguising similar pitches), and having unexpected movement. I call these two metrics ‘release point consistency’ and unexpected movement.

**Data**

I used the pybaseball package to pull data from statcast. I chose to do this because I already had 2020-22 loaded up from another project, so it was simple to just get ‘23 as well and go from there. The data is well documented too, so it was easy to figure out what everything meant. https://baseballsavant.mlb.com/csv-docs#home\_team

**“Release Point Consistency” Metric Methodology**

I calculated consistency by getting variances for each pitch type by the pitcher, weighted by the actual distribution of thrown pitch types by pitcher.

The leader of this ranking was Tony Watson. Taken from the Philidelphia Inquirer: “Since I’m on the first base side of the rubber, the arm slot comes behind the left-handed hitter. A little deception,” Watson said. “I’ve always had deception. It’s a healthier delivery. Knock on wood, it’s kept me going.” This checks out that he would have the best rating here, since he intentionally “lowered his arm slot to throw from a three-quarter delivery”

This part did not include a model so I didn’t use any cross-validation or error metrics. I made sure to filter so only pitchers with 100+ pitchers were considered. There wasn’t a ton of variance between pitchers, but this later proved to be the second most helpful metric that I came up with.

A graph with blue lines

Description automatically generated

**“Unexpected Movement Scores” Methodology & Model**

Next, to create “unexpected movement scores” in both the horizontal and vertical directions, I decided to get predictions for each pitcher and each of their pitch types based on similar pitchers/pitches, and then compare their actual pitch movement to these predictions. The pitchers who vary the most from their predictions would have the most unexpected movement.

To accomplish this I did a k nearest neighbors regression analysis.

In cleaning this data, I flipped all the movements for horizontal movement from lefties so they could be compared to right-handed hitters.

I found hardly any correlation between release point and movement, but once I broke each pitch out by hit type, there was a little bit that felt like I could run with it. Here you see the general trends vs a few of the pitch\_type trends.

A group of blue dots

Description automatically generatedA screenshot of a graph

Description automatically generated

I then took the average movement (x and z) for each pitcher by pitch type.

To deal with low pitch counts, I consolidated a few pitch types; for example, I treated FA as FF (generic fastballs as four-seamers), SV as SL (slurves as sliders). In comparing rmse for varying levels of K (1-100), I found that for most pitches, rmse began to taper around the k= 10/20 range. Even though I got slightly different k recommendations from my tests, I decided to use k=15 because I wasn’t losing much with rmse and it was simpler.

A graph of different pitch types

Description automatically generated

After running each pitch\_type model for vertical and horizontal movement and obtaining predictions, I took the absolute value of the difference between prediction and actual movement, weighted those values by thrown pitch distribution for each pitcher and then came out with two metrics: unexpected movement in the x direction, and unexpected movement in the z direction.

The most unexpected vertical movement on a pitch was Aaron Civale’s CS (I believe slow curve). From an article:("Civale's ability to pound the zone and stay ahead in the count has been enormously effective. It's this deceptive simplicity that has been the key to Civale's success thus far.” This didn’t speak to his slow curve but alludes to him being deceptive, which is nice.

The most unexpected horizontal movement on a pitch was Matt Wisler’s changeup. He hardly ever threw it though, so it’s important to weigh and sum all the pitch types.

**Interpretation of top unexpected movement pitchers**

After weighing each pitch type for pitchers and getting one main score for horizontal and vertical movement, the top pitchers were Andrew Vasquez (x) and César Valdez (z).

Andrew Vasquez quote: “His curve and fastball come out in the same arm slot, making it very deceiving for a hitter to pick up on what pitch is coming. They’re deathly afraid of getting caught with a surprise fastball they can’t catch up with after having to sit back for the breaking ball.”

César Valdez: Quote: “The first time we faced him, the hitters were like, 'This is not your typical changeup,’” Red Sox hitting coach Tim Hyers told FanGraphs. “It’s almost like a unique curveball, because it gets to home plate and just dives. And at times, it can dive both ways; it can break in or break out.”

If his curve “just dives” that makes sense that he ranks highest for unexpected vertical movement.

**Final Rankings**

To obtain my final leaderboard for most deceptive pitchers, I did MLR on the three metrics I had obtained: consistency and unexpected movement (vertical and horizontal). I decided to use Called Plus Swinging Strike Percent (based on findings from another researcher) as the response variable because if a pitcher is deceptive, they are generating swings and misses.

I took the coefficients for the three metrics, used those as weights for each metric, and calculated a final measure to determine the most deceptive pitchers. The weights were:

Horizontal movement: 0.0277

Vertical movement: 0.0103

Release Point Consistency 0.0092

The numbers don’t mean much alone, but comparing them to each other we learn that unexpected horizontal movement made the most impact on CSW and deception( about 2x the others), then vertical movement and consistency.

Final ranking results are as follows:

A table with numbers and a number on it

Description automatically generated

I touched on Andrew Vasquez earlier, so it’s no surprise he’s up here.

Tyler Rogers, #2 overall, is a submariner, making his movement hard to track.A baseball player throwing a ball

Description automatically generated

#3, Ben Rowen, former Texas Ranger, didn’t have great numbers in ‘21, but understandably tops this list with his unorthodox delivery:

A baseball player throwing a ball

Description automatically generated

**Next Steps/Weakness/Ways to improve:**

If I were to continue working on this project, here is what I would do:

* Analyze results to find pitchers that top the list with normal deliveries, or at least, pitchers that aren’t already well-known for being deceptive.
* Run this model on minor leaguers, amateurs; anyone we have data for.
* Run model, grouping pitchers by the team they played for, and then getting team rankings. Or, look at changes in player rankings/scores from team to team, to see how well our pitching staff might be doing with helping pitchers develop deception.
* Double down on standardizing metrics so I can compare players better.
* Scrape https://www.mlb.com/player/albert-abreu-656061 or another source to get photos for the leaderboard. Not super important but could help
* Venture into predictability based on pitch type and pitch count. Just didn't get there on this project
* Mess around with release point y and finger length, does that help?

My understanding, methodology, model strategies, and interpretations were inspired by the following sources:

-<https://www.theringer.com/mlb/2021/9/28/22695180/yusmeiro-petit-deception-pitching-delivery-velocity-biomechanics-invisiball>

-<https://towardsdatascience.com/quantifying-pitcher-deception-7fb2288661c8>

-<https://blogs.fangraphs.com/an-attempt-to-quantify-pitcher-deception/>

-<https://www.batterypower.com/2023/4/26/23699430/atlanta-braves-player-analysis-bryce-elder-pitching-stuff-bamboozling-he-can-throw-a-baseball>

-<https://jon2anderson.medium.com/quantifying-pitcher-fastball-deception-760832ec14ab>

-<https://thinkbluepc.com/2021/09/04/a-closer-look-at-andrew-vasquez/>

-<https://mlb.com/news/cesar-valdez-s-dead-fish-changeup-analyzed>

-<http://englandtribe.blogspot.com/2019/08/aaron-civale-2019.html>

-<https://neptune.ai/blog/knn-algorithm-explanation-opportunities-limitations#:~:text=To%20select%20the%20value%20of,candidate%20for%20our%20K%20value>

-<https://www.reddit.com/r/baseball/comments/jbcjoc/what_pitch_is_labelled_cs/>

Again, **Python Code:**

<https://colab.research.google.com/drive/1tGmuIVW0iIwrHEui2gxLnZ0yf4njgUC3?usp=sharing>