Big Red Machine Learning

Shane Hauck, Jack Hinde, Anthony Vicinanzo, Jacob Weber, and Dave Yount

Introduction: In the context of modern baseball, the analysis of pitcher performance through advanced metrics has become paramount for strategic decision-making. A big part of this advancement is related to pitchers and their defined roles. This study introduces an innovative approach to quantify pitcher fatigue, leveraging the rich data provided by the Statcast system. By focusing on detailed pitch-by-pitch data, we aim to uncover how fatigue affects various pitching metrics, potentially impacting a pitcher's performance within a game. Our methodology employs statistical models to analyze changes in key pitching metrics, offering a comprehensive view of fatigue. This research not only enhances our understanding of pitcher fatigue but also provides valuable insights for optimizing pitcher utilization and performance strategy in professional baseball.

To complement our analytical framework, we also created a Shiny application, enabling a dynamic exploration of pitcher performance data. This tool allows for interactive analysis, offering users the ability to examine and identify pitchers who may benefit from role adjustments based on their fatigue profiles and performance metrics. Integrating this application into our study enriches the decision-making process, providing a practical platform for applying our findings to real-world baseball strategy and pitcher management.

Link to App: bigredmachinelearning.com

Methods: In our study, we developed a novel approach to quantify pitcher fatigue using detailed pitch-by-pitch data from the Statcast system, referred to as "savant" data. This methodological framework was designed to assess how various pitching metrics change with fatigue, potentially influencing a pitcher's performance throughout a game.

For the core of our fatigue analysis, we identified key response variables that could be direct indicators of pitcher fatigue, such as release speed, release positions, pitch movement metrics, velocity components, acceleration components, effective speed, release spin rate, and release extension. These metrics were chosen for their potential to reflect changes in a pitcher's mechanics or effectiveness due to fatigue.

To quantify fatigue for each pitcher and pitch type, we employed linear models, fitting a separate model for each combination of pitcher, pitch type, and response variable. The linear models were structured to predict the chosen response variable as a function of pitch number within a game appearance, under the hypothesis that changes in pitch characteristics throughout an appearance might indicate fatigue. Specifically, we utilized the linear model formula: lm(response~pitch_number_appearance), with the slope coefficient (β_1) representing the fatigue effect for each response variable. This coefficient indicates how much the response variable changes with each successive pitch, serving as a direct quantifier of fatigue.

After calculating the fatigue coefficient for each response variable, we aggregated these coefficients to derive a comprehensive fatigue rating for each pitcher by pitch type. This was achieved by analyzing the distribution of fatigue coefficients across all response variables and calculating percentile rankings. The mean percentile across all variables provided an overall fatigue rating, while individual response variables allowed us to explore fatigue effects on specific aspects of pitching performance.

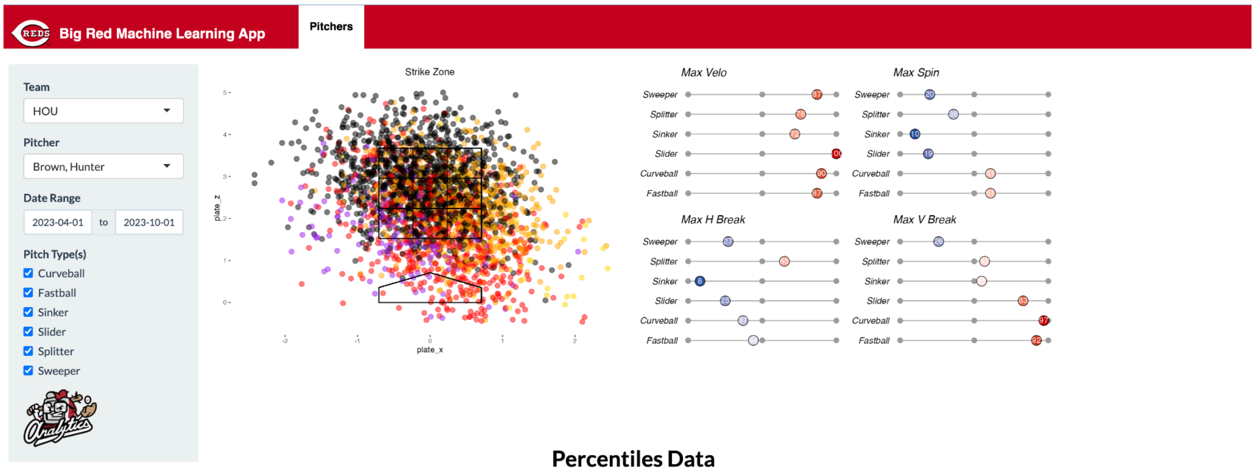
Our methodology offers a nuanced view of pitcher fatigue, leveraging high-resolution data to uncover subtle trends that traditional metrics might overlook. By examining fatigue on a pitch-by-pitch basis and across a broad spectrum of pitching metrics, we can gain insights into the multifaceted nature of fatigue and its impact on pitching performance.

To further enhance our analysis and practical application of pitcher fatigue and performance optimization, we developed a comprehensive Shiny application. This application facilitates the detailed analysis of pitchers, allowing users to interactively assess various metrics and fatigue ratings to identify pitchers who may be better suited for different roles within the team. By integrating our fatigue quantification models into an interactive platform, we provide coaches, analysts, and decision-makers with a powerful tool to make informed, data-driven decisions about pitcher utilization, aiming to optimize team performance and pitcher longevity.

Discussion & Results:

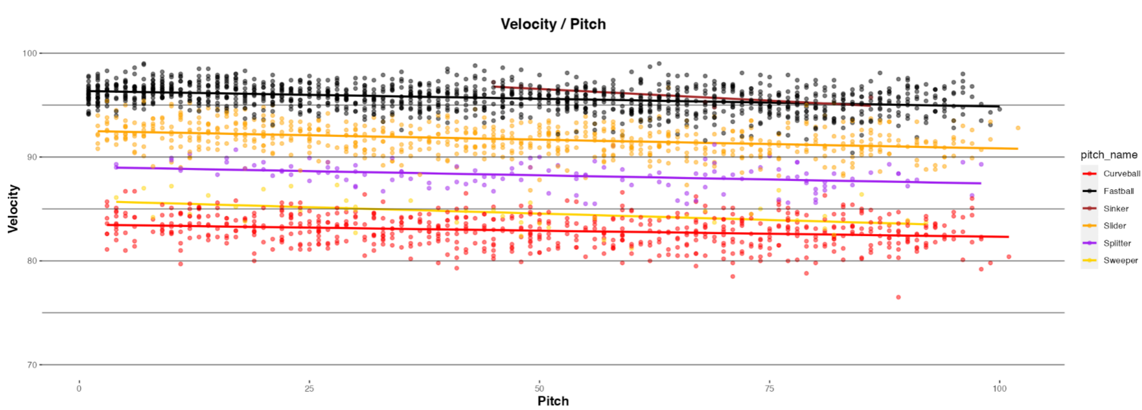
With our results, we have found Huascar Brazoban (MIA) to be fit to become a starting pitcher and Hunter Brown (HOU) to become a relief pitcher.

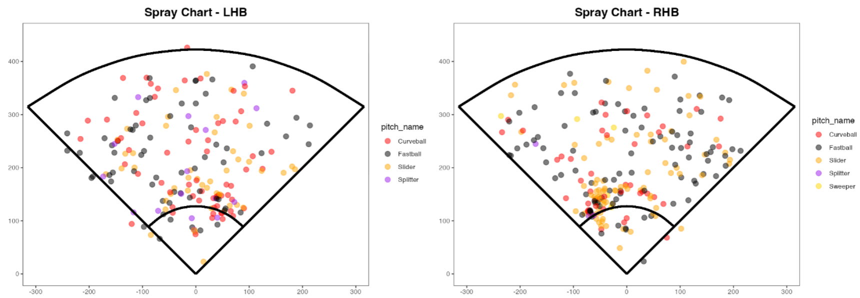
Hunter Brown: is a better fit as a reliever, despite being a young arm and a highly touted prospect, his velocity takes a dip from the start of the game around the fifty-pitch mark. This drops him from the 88th fastball percentile to around the 50th. However, he has many key metrics that profile him as an elite bullpen arm. He is in the 79th percentile fastball velocity, an elite 87th percentile groundball rate, a plus 75th percentile strikeout rate, and an exceptionally low flyball rate at 17%. While Brown has primarily been a starter throughout the regular season, but he has worked in relief during the playoffs and found more success. Across his 7 outings in this year's playoffs, he stranded 89% of runners, had a 2.91 xFIP, and a WHIP of 1.

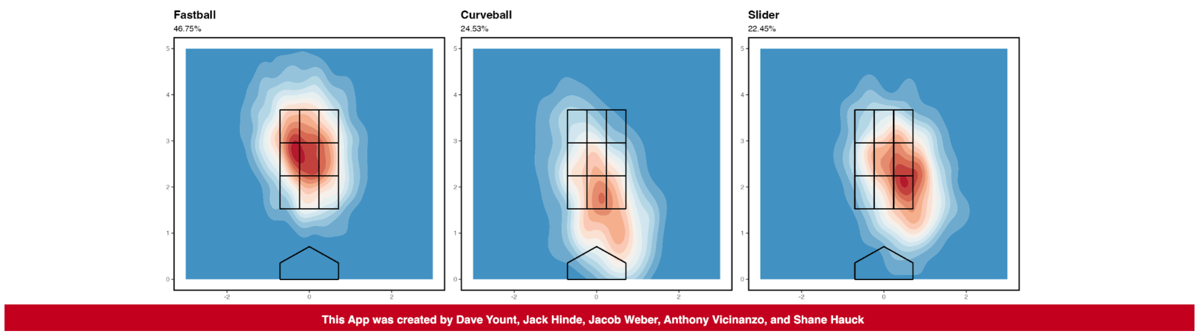


Percentiles Data

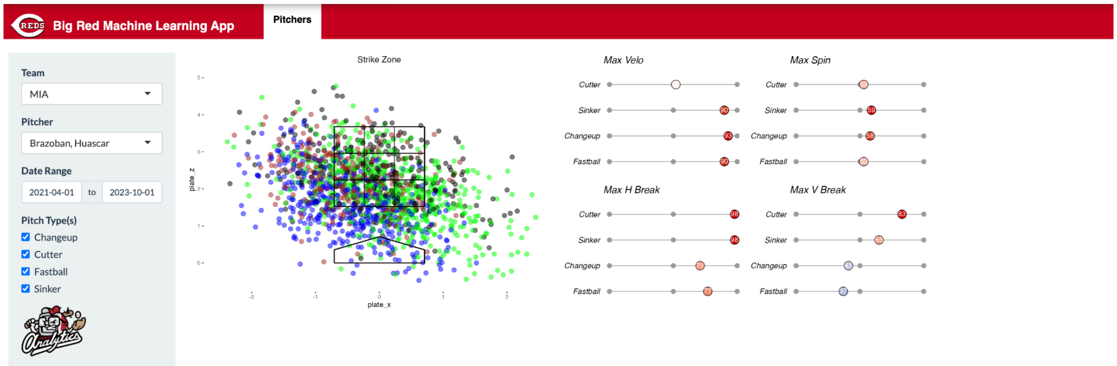
pitch_name	Max Velo	Max Spin	Max H Break	Max V Break	Usage %	Max_Velo_Rank	Max_Spin_Rank	Max_HBreak_Rank	Max_VBreak_Rank
Fastball	99	2536	1	1.9	46.1	109/819	320/819	456/819	65/819
Curveball	86.7	2746	1	1.8	23.75	47/449	177/449	284/449	14/449
Slider	96.7	2411	0.6	1	23.49	1/684	551/684	513/684	113/684
Splitter	91.2	1668	1.5	1.1	4.78	23/94	60/94	33/94	40/94
Sweeper	87.3	2573	1.5	0.6	1.69	27/215	171/215	157/215	160/215
Sinker	97.2	2193	1.2	1.2	0.18	156/556	503/556	514/556	249/556







Huascar Brazoban: has been successful as a reliever, however, he actually would be more valuable as a starter. In our model, we saw that when Huascar has thrown multiple innings his velocity has climbed throughout the outing. The model showed that his cutter is one of the best pitchers in terms of break across the big leagues as it is in the 98th percentile for Horizontal break in the 83rd percentile for Vertical break. Brazoban's advanced metrics from Savant further back up our findings. In 2023, he ranked in the 97th percentile for average exit velocity against, 94th percentile in chase rate, 86th percentile in groundball and whiff rate, and in 87th percentile in average fastball velocity.



Percentiles Data

pitch_name	Max Velo	Max Spin	Max H Break	Max V Break	Usage %	Max_Velo_Rank	Max_Spin_Rank	Max_HBreak_Rank	Max_VBreak_Rank
Fastball	99.3	2507	1.3	1.6	20.02	170/1760	835/1760	413/1760	1112/1760
Changeup	92.8	2135	1.7	1	27.08	94/1427	600/1427	410/1427	849/1427
Sinker	99	2459	2.1	1.4	16.62	117/1207	493/1207	27/1207	423/1207
Cutter	91.5	2573	1.2	1.5	36.28	299/617	289/617	10/617	104/617