Quantifying The Effect of Pitching Metrics

Max Brown

May 20, 2022

Introduction

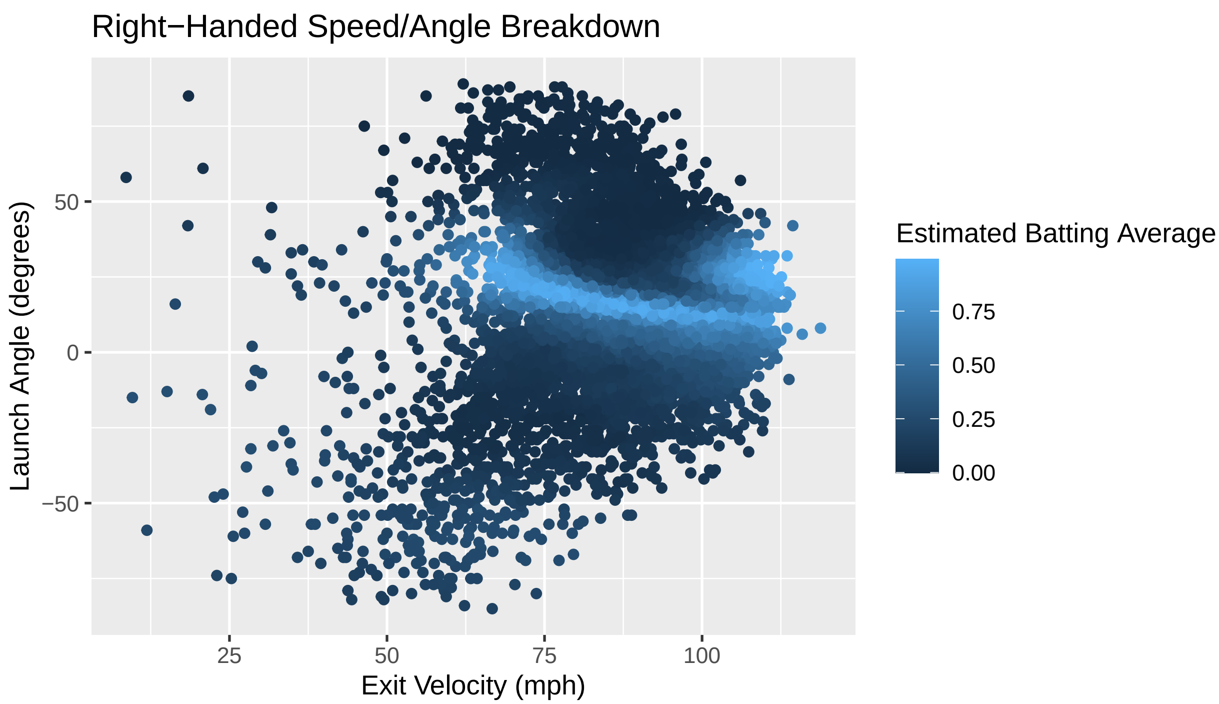
Throughout the establishment of baseball statistics, the perception of pitchers has surrounded around their earned run average and walk rate. Recently, there has been a focus on new techniques, such as using pitcher and pitch specific statistics to determine the most efficient way to evaluate pitchers. In 2015, Mike Sonne created a metric titled "Stuff+" in which he used a pitcher's change in velocity and break distance of off-speed pitches to determine a value relative to other pitchers (https://fantasy.fangraphs.com/using-the-stuff-metric-as-an-injury-identification-tool/). The following year, he used the stuff metric to discover that Marco Estrada recently saw a dip in his rating from his previous starts, releasing an article that the Blue Jays pitcher might be dealing with an injury. Sonne ended up being correct as it was announced that Estrada was suffering from a herniated disc. My report is inspired by Sonne's work, but takes a different approach. Similar to Driveline's December 2021 study of this new "Stuff" metric, this report will focus closer on individual pitches and how they compare within each pitcher and across different pitchers (https://www.drivelinebaseball.com/2021/12/what-is-stuff-quantifying-pitches-with-pitch-models/). However, instead of using a formula of percentiles for different metrics to determine a final value, my methodology will use a results-based approach to determine how certain pitching metrics alter the odds of an ideal result for the pitcher.

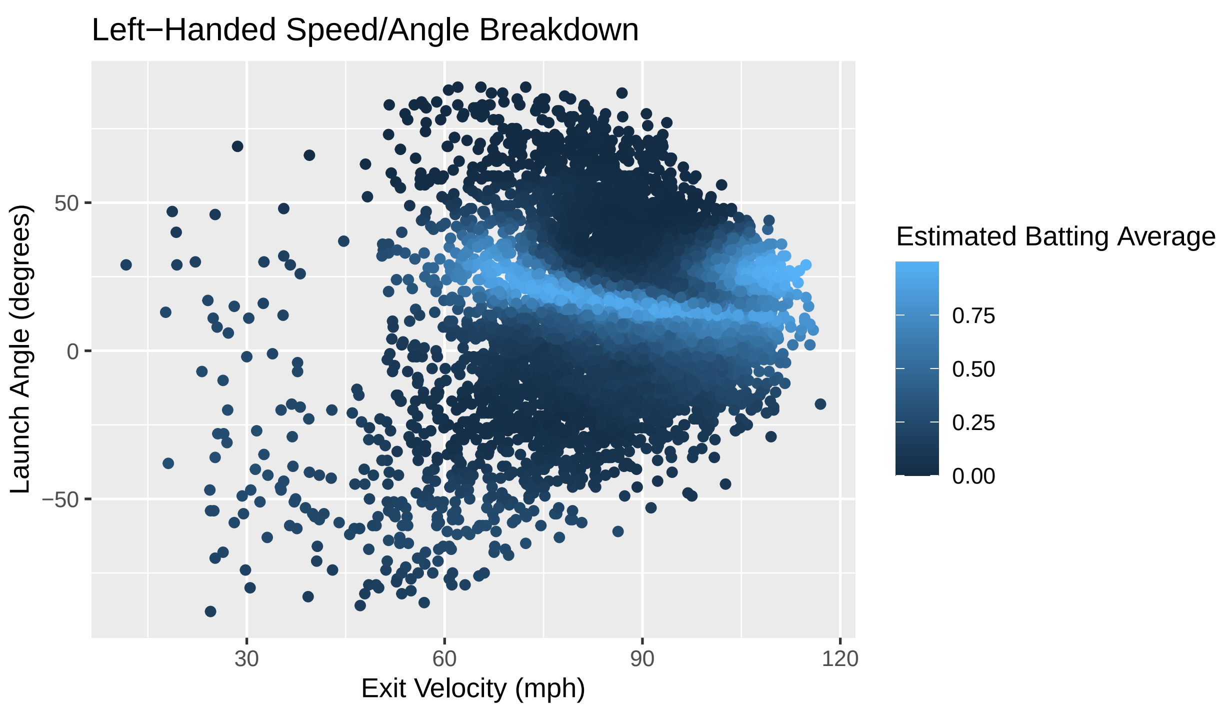
The data used for this analysis was extracted from baseballsavant.mlb.com. It consists of over 70000 pitches throughout the 2022 regular season from Opening Day to July 30th. To narrow the focus of this study, the data set consists of only breaking pitches, with a primary focus on sliders, knuckle curves, and curveballs. These indications are determined based on the "pitch type" in Baseball Savant's data. The main pitching variables are:

"release_speed" - Speed of baseball once released from the pitcher's hand (mph) "release_spin_rate" - Spin of the baseball once released from the pitcher's hand (rpm) "pfx_x" - Horizontal movement of baseball during the duration of the pitch "pfx_z" - Vertical movement of baseball during the duration of the pitch

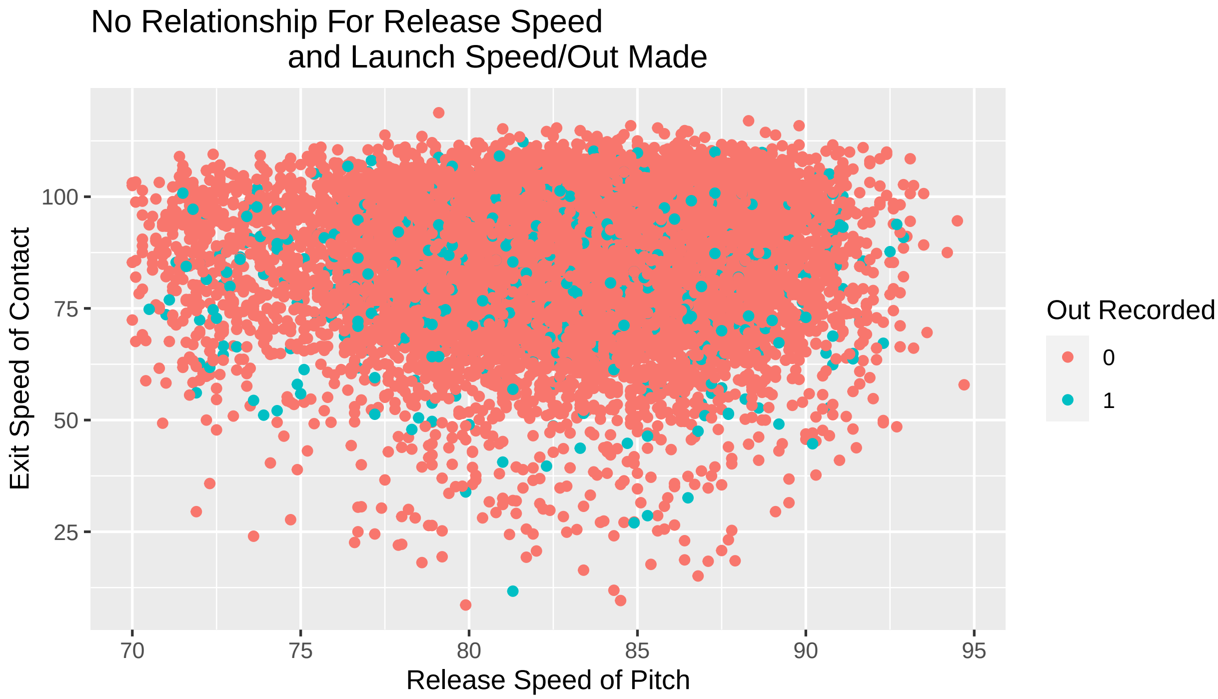
Statcast and Baseball Savant records pfx_x and pfx_z in feet. To evaluate how the change in each of these variables affects our response, converting to inches is reasonable. Additionally, the data sets do not have a minimum requirement of pitches, meaning that there are most likely position players included. To remove position players and pitchers with very little major league experience throwing breaking balls, a minimum requirement of 30 pitches will be set for analysis.

General EDA





The graphs above show an almost identical distribution of batted balls between left-handed and right-handed pitchers. It is also worth noting that estimated batting average, a measure utilized by Statcast to approximate the probability of a batted ball resulting in a hit, is heavily affected by both exit velocity and launch angle.



After filtering out batted balls hit less than 70 mph, it is apparent that there is not a clear relationship between pitch speed, exit velocity, and an out being recorded. This visual shows that the game of baseball can't solely be determined from speed, as television broadcasts often use these statistics to broadly evaluate the quality of a pitch.

Righty EDA

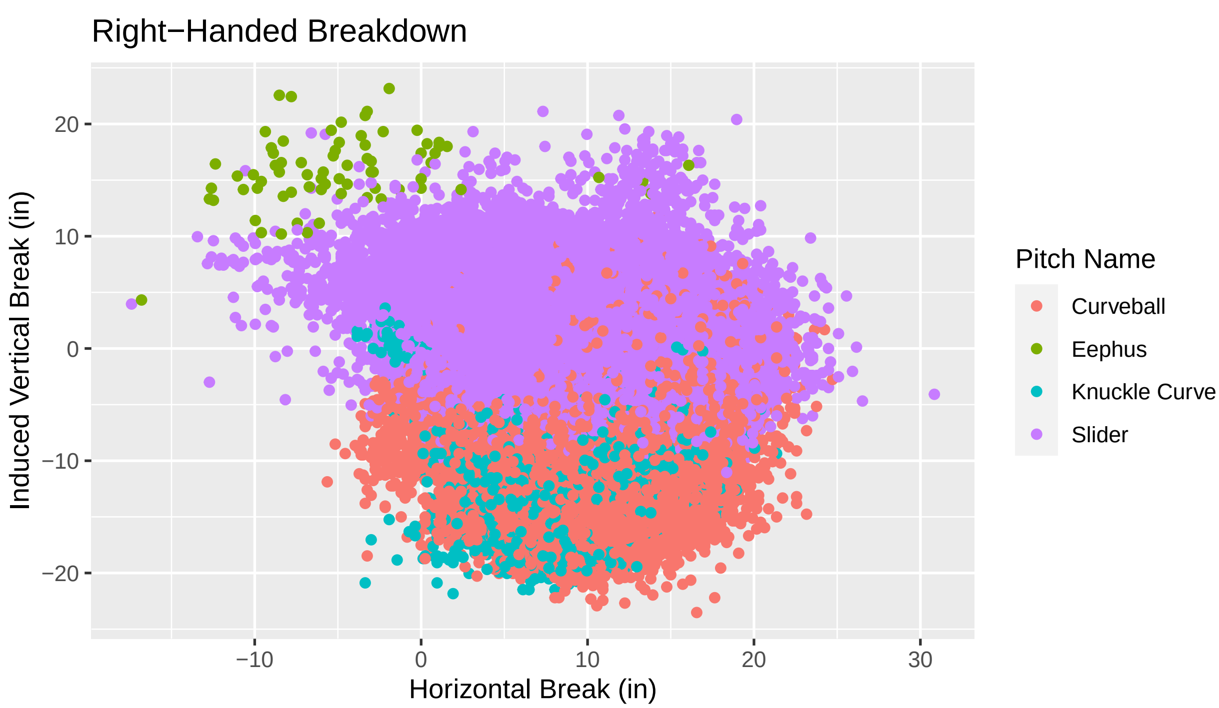


Table 1: Average RH Pitch Type Metrics

pitch_name	release_speed	release_spin	hor_break	vert_break	tilt
Curveball	79.37	2572.90	9.61	-9.69	43.66
Eephus	52.98	1248.92	-4.54	15.86	211.78
Knuckle Curve	81.52	2584.29	8.09	-10.79	47.93
Slider	84.95	2455.40	6.93	1.97	108.24

pitch_type	pitch_name	n
CS	Curveball	15
CU	Curveball	9136
EP	Eephus	72
KC	Knuckle Curve	2593
SL	Slider	26538

Lefty EDA

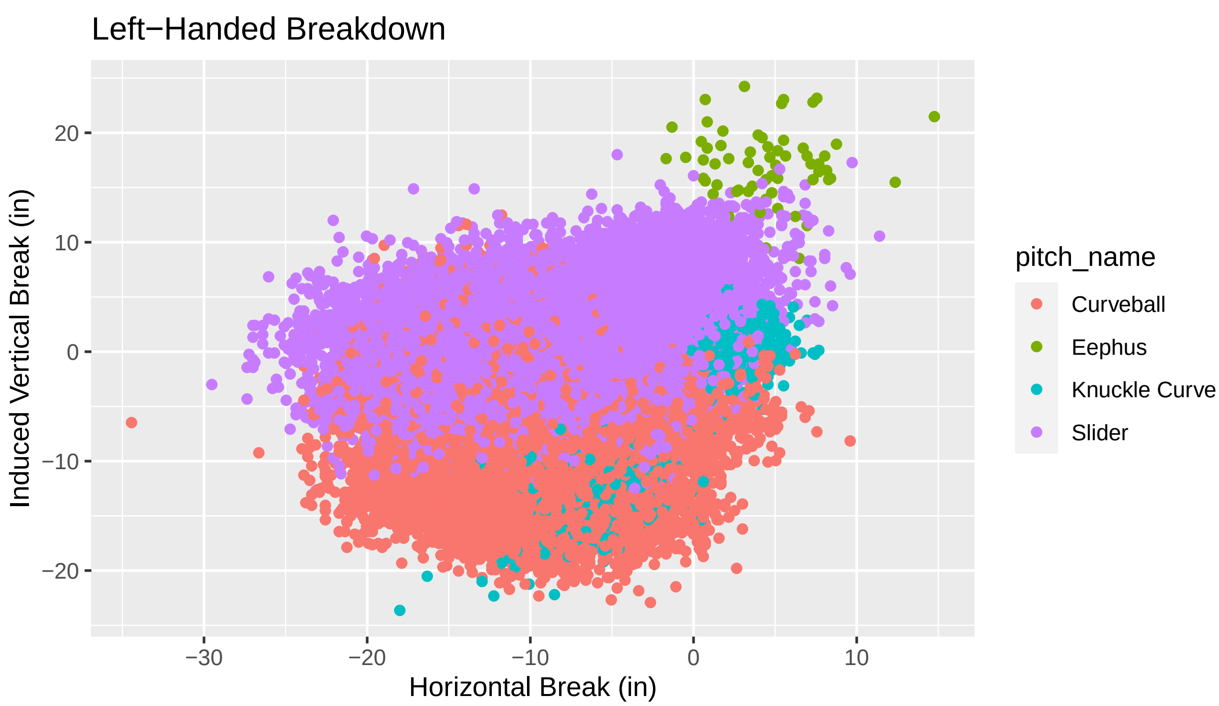


Table 3: Average LH Pitch Type Metrics

pitch_type	release_speed	release_spin	hor_break	vert_break	tilt
CS	65.45	2354.13	-18.82	-10.63	309.67
CU	78.27	2469.73	-9.10	-8.63	314.17
EP	45.14	1224.42	4.53	17.15	157.29
KC	78.95	2290.82	-3.07	-7.78	248.41
SL	84.04	2365.92	-6.48	1.77	245.49

pitch_type	pitch_name	n
CS	Curveball	15
CU	Curveball	11213
EP	Eephus	62
KC	Knuckle Curve	1398
SL	Slider	23245

After looking at summary statistics for both right and left handed pitch types, it makes the most sense to proceed for sliders, knuckle-curves, and curveballs. For slow curves (listed as "CS") and eephus pitches, there is not a large enough sample size to provide useful insight.

Methodology

In creating a response variable I decided to simplify the result of the pitch into an ideal and unideal scenario for the pitcher. Any pitch not put into play and results in a strike is an ideal result for the pitch. Additionally, any ball put in play with an expected batting average of less than .311, the average estimated batting average for batted balls in the entire dataset, is considered an ideal result. These ideal results are assigned as 1 and all other outcomes are given a 0, creating a binary response variable. Using this response and given the correlation in pitches between pitchers, a multi-level approach is needed with each individual pitch as the first level observational unit and the pitchers as the second level observational unit. To maximize the amount of predictor variables that can be included into the model for each pitch type, a level of significance of .2 will suffice.

Creating Right Handed Models

Curveball

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	0.3093	0.7170	0.4314	0.6662
fixed	NA	release_speed	-0.0038	0.0086	-0.4480	0.6541
fixed	NA	hor_break	0.0128	0.0057	2.2554	0.0241
fixed	NA	ind_vert_break	-0.0034	0.0053	-0.6476	0.5173
fixed	NA	mean_centered_spinrate	0.0004	0.0002	2.2990	0.0215
ran_pars	player_name	sd__(Intercept)	0.2014	NA	NA	NA

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	0.0294	0.0569	0.5178	0.6046
fixed	NA	hor_break	0.0138	0.0053	2.6069	0.0091
fixed	NA	mean_centered_spinrate	0.0004	0.0002	2.3809	0.0173
ran_pars	player_name	sd__(Intercept)	0.2030	NA	NA	NA

Knuckle-Curve

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	1.2660	1.4850	0.8526	0.3939
fixed	NA	release_speed	-0.0130	0.0172	-0.7522	0.4519
fixed	NA	hor_break	0.0148	0.0119	1.2433	0.2138

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	ind_vert_break	0.0078	0.0112	0.6969	0.4859
fixed	NA	mean_centered_spinrate	-0.0003	0.0003	-0.9872	0.3235
ran_pars	player_name	sd__(Intercept)	0.2117	NA	NA	NA

Slider

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	1.4916	0.5176	2.8819	0.0040
fixed	NA	release_speed	-0.0147	0.0060	-2.4674	0.0136
fixed	NA	hor_break	-0.0015	0.0035	-0.4381	0.6613
fixed	NA	ind_vert_break	0.0069	0.0037	1.8602	0.0629
fixed	NA	mean_centered_spinrate	0.0003	0.0001	2.9590	0.0031
ran_pars	player_name	sd__(Intercept)	0.2145	NA	NA	NA

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	1.3766	0.4440	3.1007	0.0019
fixed	NA	release_speed	-0.0135	0.0053	-2.5711	0.0101
fixed	NA	ind_vert_break	0.0072	0.0037	1.9342	0.0531
fixed	NA	mean_centered_spinrate	0.0003	0.0001	3.0154	0.0026
ran_pars	player_name	sd__(Intercept)	0.2142	NA	NA	NA

Creating Left Handed Models

Curve

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	-0.2307	0.7200	-0.3204	0.7487
fixed	NA	release_speed	0.0050	0.0088	0.5676	0.5703
fixed	NA	hor_break	-0.0075	0.0053	-1.3981	0.1621
fixed	NA	ind_vert_break	0.0036	0.0045	0.7970	0.4255
fixed	NA	mean_centered_spinrate	0.0004	0.0002	2.3189	0.0204
ran_pars	player_name	sd__(Intercept)	0.1568	NA	NA	NA

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	0.1373	0.0536	2.5617	0.0104
fixed	NA	hor_break	-0.0064	0.0051	-1.2528	0.2103
fixed	NA	mean_centered_spinrate	0.0004	0.0002	2.3606	0.0182
ran_pars	player_name	sd__(Intercept)	0.1630	NA	NA	NA

Knuckle-Curve

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	1.2660	1.4850	0.8526	0.3939
fixed	NA	release_speed	-0.0130	0.0172	-0.7522	0.4519

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	hor_break	0.0148	0.0119	1.2433	0.2138
fixed	NA	ind_vert_break	0.0078	0.0112	0.6969	0.4859
fixed	NA	mean_centered_spinrate	-0.0003	0.0003	-0.9872	0.3235
ran_pars	player_name	sd__(Intercept)	0.2117	NA	NA	NA

Slider

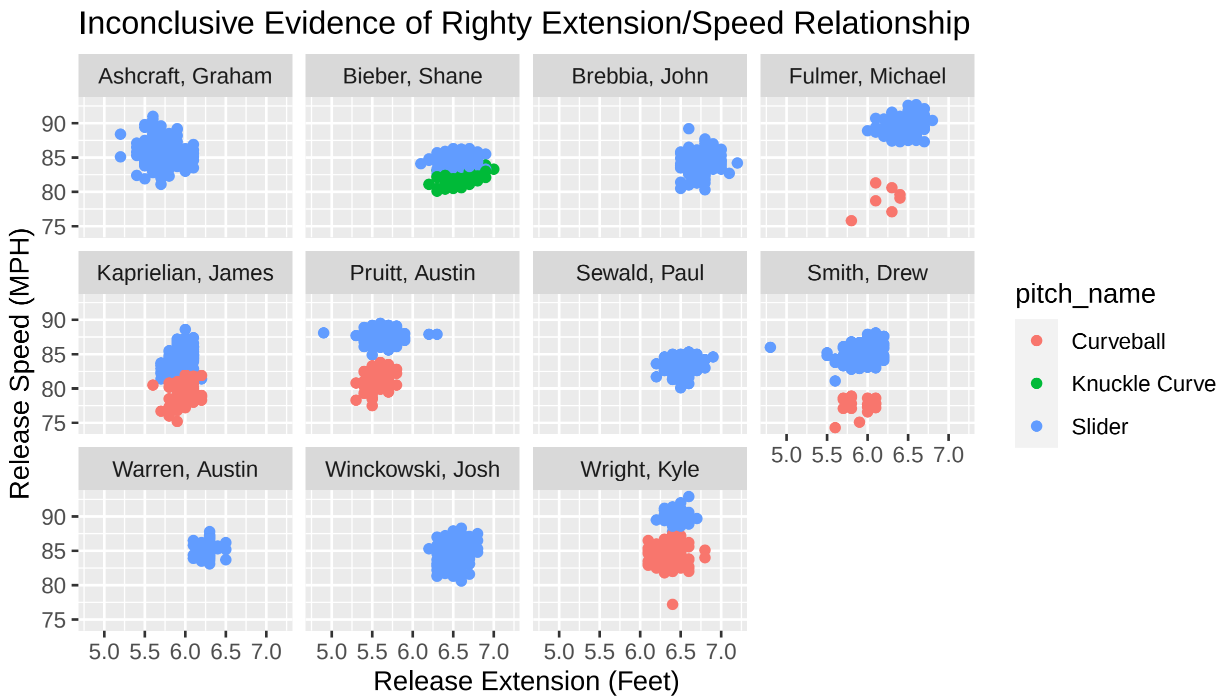
effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	0.4623	0.5566	0.8304	0.4063
fixed	NA	release_speed	-0.0032	0.0065	-0.4837	0.6286
fixed	NA	hor_break	-0.0053	0.0038	-1.4111	0.1582
fixed	NA	ind_vert_break	0.0147	0.0042	3.5115	0.0004
fixed	NA	mean_centered_spinrate	0.0002	0.0001	1.6091	0.1076
ran_pars	player_name	sd__(Intercept)	0.1478	NA	NA	NA

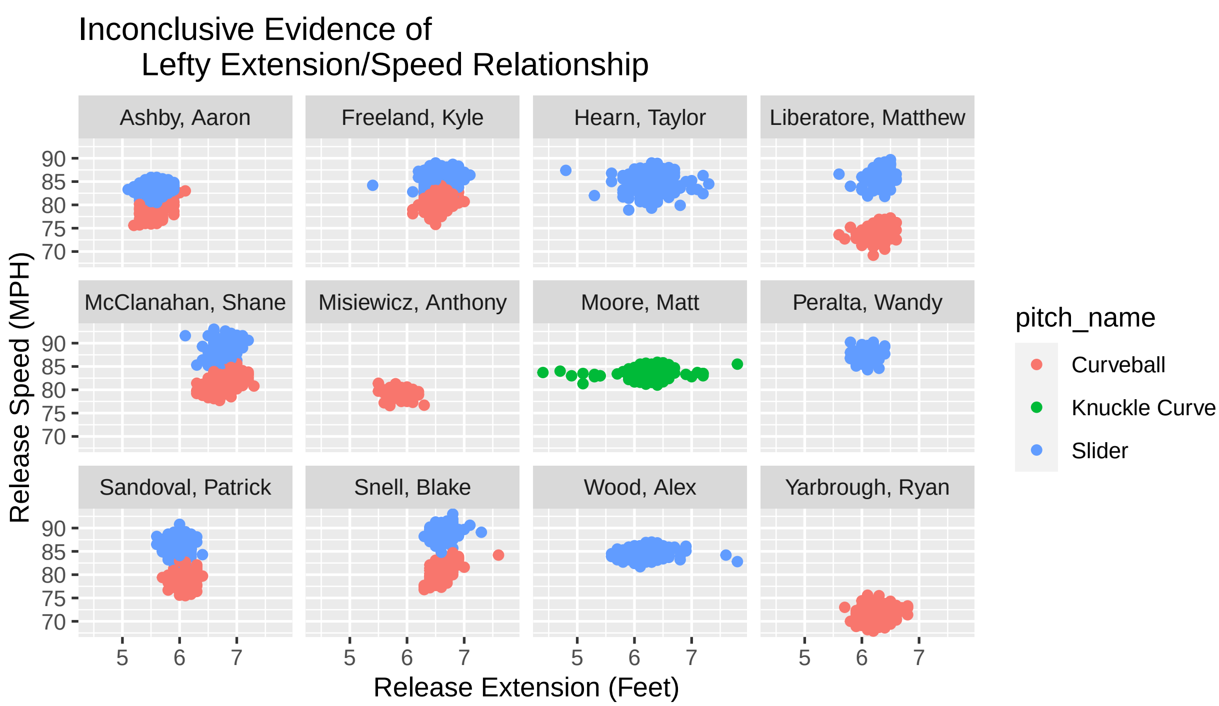
Table 14: Top Ten Average Pitch Ratings By Player

effect	group	term	estimate	std.error	statistic	p.value
fixed	NA	(Intercept)	0.1934	0.0324	5.9705	0.0000
fixed	NA	hor_break	-0.0062	0.0032	-1.9249	0.0542
fixed	NA	ind_vert_break	0.0143	0.0041	3.5003	0.0005
fixed	NA	mean_centered_spinrate	0.0002	0.0001	1.7248	0.0846
ran_pars	player_name	sd__(Intercept)	0.1448	NA	NA	NA

player_name	mean_rating	pitches
Rogers, Tyler	111.400	136
Romo, Sergio	107.242	63
Smith, Joe	106.851	40
Cimber, Adam	106.034	106
Sandlin, Nick	105.354	69
Thompson, Ryan	105.178	77
O'Day, Darren	105.144	74
Young, Danny	104.977	36
Maton, Phil	104.955	111
Martinez, Seth	104.817	62

Appendix


Inconclusive Evidence of Righty Extension/Speed Relationship
Ashcraft, Graham
Bieber, Shane
Brebbia, John
Fulmer, Michael
Kaprielian, James
Pruitt, Austin
Sewald, Paul
Smith, Drew
Warren, Austin
Winckowski, Josh
Wright, Kyle
Release Speed (MPH)
90
85
80
75
5.0 5.5 6.0 6.5 7.0
Release Extension (Feet)
pitch_name
Curveball
Knuckle Curve
Slider


Inconclusive Evidence of
Lefty Extension/Speed Relationship
Ashby, Aaron
Freeland, Kyle
Hearn, Taylor
Liberatore, Matthew
McClanahan, Shane
Misiewicz, Anthony
Moore, Matt
Peralta, Wandy
Sandoval, Patrick
Snell, Blake
Wood, Alex
Yarbrough, Ryan
Release Speed (MPH)
90
85
80
75
70
5 6 7
Release Extension (Feet)
pitch_name
Curveball
Knuckle Curve
Slider