Nick Dazell
Project Start Date: 3/14/23
Last Updated: 4/18/23

ALL EXPERIMENTS WILL BE CONDUCTED ON DATA FROM THE 2022 MLB SEASON UNLESS OTHERWISE NOTED

Disclaimer: This document was created to serve as a personal note sheet documenting my thought processes and findings throughout the course of this project. It is incomplete and not meant to serve as a formal write up or publication of scientifically rigorous results.

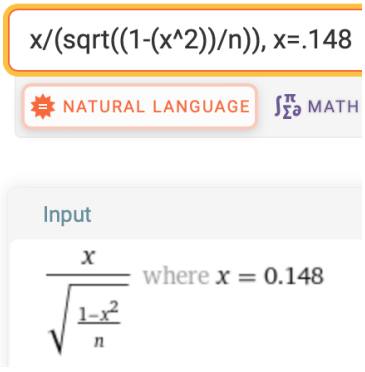
Useful Links

DataSources:
https://baseballsavant.mlb.com/leaderboard/pitch-movement?year=2022&team=&min=q&pitch_type=ALL&hand=L&x=diff_x_hidden&z=diff_z_hidden
https://baseballsavant.mlb.com/leaderboard/custom?year=2022&type=pitcher&filter=&sort=2&sortDir=asc&min=100&selections=p_formatted_ip,p_home_run,p_k_percent,p_bb_percent,on_base_plus_slg,p_era,p_opp_batting_avg,xba,wobadiff,exit_velocity_avg,launch_angle_avg,&chart=false&x=p_home_run&y=p_home_run&r=no&chartType=beeswarm

Great Tool For Custom Sabermetrics Tables from FanGraphs:
https://www.fangraphs.com/leaders.aspx?pos=all&stats=pit&lg=all&qual=0&type=c,4,5,11,7,8,13,-1,36,37,40,43,44,48,51,-1,240,-1,6,332,45,62,-1,59,139,139&season=2022&month=0&season1=2022&ind=0&team=0,ss&rost=0&age=0&filter=&players=0&startdate=&enddate=

My adjusted CSV in google sheets:
https://docs.google.com/spreadsheets/d/1PU5NmH90DsJP3xQ9FgZ5lNmn97UoH7f9_ETSOBvOMio/edit#gid=1727602173

Formula used to calculate T Score:
https://www.scribbr.com/statistics/pearson-correlation-coefficient/



Tool used to calculate p values:
https://www.socscistatistics.com/pvalues/tdistribution.aspx

Intro

Project Motivation: I initially wanted to incorporate the mountaineering concept of peak prominence by calculating mathematical prominence of different pitches in a given pitcher's arsenal. I was interested in this idea because prominence quantifies a peak's perceived height amongst other peaks in the area. Translated to pitch data, I thought this idea could show how a pitch is perceived compared to other pitches that have a similar movement and velocity profile (i.e. Comparing the movement magnitude of pitcher X's curveball to all other curveballs in baseball) I then realized that if we are comparing a pitcher X's curveball to all other curveballs, the important consideration isn't the next best curveball. A more telling metric would be the separation between pitcher X's curveball and the league average curveball (which is what a hitter has the most training with). This brought me to my next question: Is it better to have a naive increase in movement and/or velocity for a given pitch? Or could that increase actually bring some pitchers closer to league average and therefore make them less effective? In other words, is there a strict positive correlation between velocity/movement and performance? Or will performance actually decrease with an increase in velocity/movement until we pass league average thresholds, at which time it will level out and begin to increase as we pull further away from league average?

Initial Predictions: In recent years, technologies like Rapsodo, TrackMan, and Hawk-Eye have given us access to A LOT of pitching data. Through analysis of this data, it has become generally accepted that fastballs with above league average spin rates should be thrown at the top of the strike zone (these pitches will "rise" more) and these pitchers can generally be more effective because their ball moves differently from the average pitcher. The same is true on the opposite end of the spectrum, with low spin rate fastballs being more effective in the bottom of the zone because of their tendency to "sink". Most coaches and players would agree that it is better to have a spin rate that diverges from league average as much as possible. It has also long been postulated that hitters succeed against pitchers whose pitches are thrown at "hitting speed" (not too fast and **not too slow**). **TODO**: Look for research papers into this theory. I predict that this idea may be true for velocity, where slower could actually be better if it is further from league average. However, I think that more movement on a pitch will lead to better results, even if it is closer to league average.

Note: This is an incomplete study that is only pseudo-scientific. Although we will (hopefully) be able to see trends and correlations between these variables, the pitcher-hitter relationship is complex and it is impossible to control for the other variables that affect performance over the course of a season. It is very likely that pitchers who are close to league average in velocity/movement find a way to be successful by manipulating a number of other variables. (Think Nestor Cortes manipulating timing or Tyler Rogers manipulating release angle).

SQL Code

SQL Code to combine first and last name columns:

```
CREATE TABLE "performData22" as
SELECT CONCAT(first_name, ' ', last_name) AS full_name, formatted_ip, era, home_run,
k_percent, bb_percent, on_base_plus_slg as opp_ops, opp_batting_avg as baa, wobadiff as
woba_diff, exit_velocity_avg as ev_avg, launch_angle_avg as la_avg
FROM "performanceData2022"
```

SQL Code to combine pitch data tables and performance data tables:

```
SELECT curveballs2.full_name, curveballs2.pitch_type_name, curveballs2.avg_speed,
curveballs2.vert_break, curveballs2.vert_vs_league_pct, curveballs2.horiz_break,
curveballs2.horiz_vs_league_pct, "performData22".formatted_ip as ip, "performData22".era,
"performData22".k_percent as k_pct, "performData22".bb_percent as bb_pct,
"performData22".opp_ops, "performData22".baa, "performData22".woba_diff as
"wOBA-xwOBA", "performData22".ev_avg, "performData22".la_avg
FROM curveballs2
INNER JOIN "performData22"
ON curveballs2.full_name = "performData22".full_name;
```

Of the 291 pitchers with data in the changeups2 table, 262 had performanceData…
Of the 236 pitchers with data in the curveballs2 table, 215 had performanceData…
Of the 136 pitchers with data in the cutters2 table, 127 had performanceData…
Of the 409 pitchers with data in the fastballs2 table, 367 had performanceData…
Of the 256 pitchers with data in the sinkers2 table, 225 had performanceData…
Of the 350 pitchers with data in the sliders2 table, 320 had performanceData…
Of the 2 pitchers with data in the slurves2 table, 0 had performanceData…
Of the 34 pitchers with data in the splitters2 table, 29 had performanceData…
Of the 24 pitchers with data in the sweepers2 table, 21 had performanceData…

SQL Code to get a row of Pearson correlation values (velo):

```
SELECT CORR(avg_speed, era)as "avg_speed vs era", CORR(avg_speed, k_pct) as
"avg_speed vs k rate", CORR(avg_speed, bb_pct) as "avg_speed vs bb rate",
CORR(avg_speed, opp_ops) as "avg_speed vs ops", CORR(avg_speed, baa) as "avg_speed
vs baa", CORR(avg_speed, ev_avg) as "avg_speed vs ev", CORR(avg_speed, la_avg) as
"avg_speed vs la"
FROM ch_perform
```

SQL Code to get a row of Pearson correlation values (vert_break):

```
SELECT CORR(vert_break, era)as "vert_break vs era", CORR(vert_break, k_pct) as
"vert_break vs k rate", CORR(vert_break, bb_pct) as "vert_break vs bb rate", CORR(vert_break,
```

opp_ops) as "vert_break vs ops", CORR(vert_break, baa) as "vert_break vs baa", CORR(vert_break, ev_avg) as "vert_break vs ev", CORR(vert_break, la_avg) as "vert_break vs la"
FROM ch_perform

SQL Code to get a row of Pearson correlation values (horiz_break):
SELECT CORR(horiz_break, era)as "horiz_break vs era", CORR(horiz_break, k_pct) as "horiz_break vs k rate", CORR(horiz_break, bb_pct) as "horiz_break vs bb rate", CORR(horiz_break, opp_ops) as "horiz_break vs ops", CORR(horiz_break, baa) as "horiz_break vs baa", CORR(horiz_break, ev_avg) as "horiz_break vs ev", CORR(horiz_break, la_avg) as "horiz_break vs la"
FROM ch_perform

Test Results

Results - Full Range

CHANGEUP FULL TEST RESULTS (n=262):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.115	0.321	0.119	-0.191	-0.184	0.067	-0.156
Velo pVal	0.063	<0.00001	0.001	0.002	0.003	0.280	0.011
significant?	No	Yes	Yes	Yes	Yes	No	Yes
Vert corr	-0.064	-0.062	-0.135	-0.031	0.071	-0.198	-0.225
Vert pVal	0.302	0.317	0.029	0.617	0.252	0.001	0.0002
significant?	No	No	Yes	No	No	Yes	Yes
Horiz corr	-0.114	0.204	0.091	-0.139	-0.123	-0.204	-0.071
Horiz pVal	0.065	0.0009	0.142	0.024	0.047	0.0009	0.252
significant?	No	Yes	No	Yes	Yes	Yes	No

CUTTER FULL TEST RESULTS (n=127):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.250	0.443	0.103	-0.302	-0.314	-0.133	-0.208
Velo pVal	0.005	<0.00001	0.249	0.0006	0.0003	0.136	0.019
significant?	Yes	Yes	No	Yes	Yes	No	Yes
Vert corr	0.145	-0.252	-0.154	0.145	0.202	0.176	-0.035
Vert pVal	0.104	0.004	0.084	0.104	0.023	0.048	0.696
significant?	No	Yes	No	No	Yes	Yes	No
Horiz corr	-0.075	0.034	-0.092	-0.072	-0.034	0.011	0.174
Horiz pVal	0.402	0.705	0.304	0.421	0.705	0.902	0.050
significant?	No	No	No	No	No	No	No

CURVEBALL FULL TEST RESULTS (n=215):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.041	0.361	0.264	-0.119	-0.149	0.189	-0.206
Velo pVal	0.550	<0.00001	0.00009	0.082	0.029	0.005	0.002
significant?	No	Yes	Yes	No	Yes	Yes	Yes
Vert corr	-0.069	-0.187	-0.071	-0.015	0.031	-0.117	0.010
Vert pVal	0.314	0.006	0.300	0.827	0.651	0.087	0.884
significant?	No	Yes	No	No	No	No	No
Horiz corr	-0.007	0.009	-0.111	0.015	0.041	-0.166	0.016
Horiz pVal	0.919	0.896	0.105	0.827	0.550	0.015	0.815
significant?	No	No	No	No	No	Yes	No

FASTBALL FULL TEST RESULTS (n=367):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.193	0.476	0.201	-0.268	-0.253	0.075	-0.100
Velo pVal	0.0002	<0.00001	0.0001	<0.00001	<0.00001	0.152	0.056
significant?	Yes	Yes	Yes	Yes	Yes	No	No
Vert corr	0.100	-0.352	-0.158	0.133	0.218	-0.191	-0.335
Vert pVal	0.056	<0.00001	0.002	0.011	0.00003	0.0002	<0.00001
significant?	No	Yes	Yes	Yes	Yes	Yes	Yes
Horiz corr	0.010	0.076	0.011	0.011	-0.026	-0.011	-0.023
Horiz pVal	0.849	0.146	0.834	0.834	0.619	0.834	0.660
significant?	No	No	No	No	No	No	No

SINKER FULL TEST RESULTS (n=225):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.121	0.417	0.338	-0.249	-0.283	-0.022	-0.258
Velo pVal	0.070	<0.00001	<0.00001	0.0002	0.00002	0.742	0.00009
significant?	No	Yes	Yes	Yes	Yes	No	Yes
Vert corr	-0.034	-0.199	-0.159	0.018	0.152	-0.060	-0.306
Vert pVal	0.612	0.003	0.017	0.788	0.023	0.370	<0.00001
significant?	No	Yes	Yes	No	Yes	No	Yes
Horiz corr	-0.069	0.087	-0.041	-0.078	0.009	-0.122	-0.255
Horiz pVal	0.303	0.194	0.541	0.244	0.894	0.068	0.0001
significant?	No	No	No	No	No	No	Yes

SLIDER FULL TEST RESULTS (n=320):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.106	0.259	-0.009	-0.179	-0.142	0.148	-0.109
Velo pVal	0.058	<0.00001	0.873	0.001	0.011	0.008	0.051
significant?	No	Yes	No	Yes	Yes	Yes	No
Vert corr	0.126	-0.198	0.058	0.175	0.173	-0.011	-0.025
Vert pVal	0.024	0.0004	0.301	0.002	0.002	0.845	0.656
significant?	Yes	Yes	No	Yes	Yes	No	No
Horiz corr	-0.090	0.018	0.066	-0.067	-0.082	-0.299	-0.029
Horiz pVal	0.108	0.748	0.239	0.232	0.143	<0.00001	0.606
significant?	No	No	No	No	No	Yes	No

SPLITTER FULL TEST RESULTS (n=29):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.177	0.360	0.043	-0.147	-0.048	0.206	-0.584
Velo pVal	0.359	0.055	0.824	0.447	0.804	0.284	0.0009

significant?	No	No	No	No	No	No	Yes
Vert corr	0.060	-0.185	0.113	-0.039	-0.139	-0.255	0.116
Vert pVal	0.757	0.337	0.559	0.841	0.472	0.182	0.549
significant?	No	No	No	No	No	No	No
Horiz corr	0.068	-0.196	-0.251	0.194	0.378	0.157	-0.329
Horiz pVal	0.726	0.308	0.189	0.313	0.043	0.416	0.081
significant?	No	No	No	No	Yes	No	No

SWEEPER FULL TEST RESULTS (n=21):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.261	0.403	0.180	-0.391	-0.381	0.001	-0.091
Velo pVal	0.253	0.070	0.435	0.080	0.088	0.997	0.695
significant?	No	No	No	No	No	No	No
Vert corr	0.290	-0.629	-0.105	0.501	0.618	0.298	-0.313
Vert pVal	0.202	0.002	0.651	0.021	0.003	0.189	0.167
significant?	No	Yes	No	Yes	Yes	No	No
Horiz corr	-0.163	-0.009	-0.125	0.159	0.253	-0.132	-0.173
Horiz pVal	0.480	0.969	0.589	0.491	0.268	0.569	0.453
significant?	No	No	No	No	No	No	No

average velocities according to PITCHf/x data:

vFA (Four Seam and Unclassified Fastballs) Yes - 93.8

vFT (Two Seam Fastballs) No Data

vFC (Cutters) Yes - 89.0

vFS (Split Fingers) Yes - 87.2

vFO (Forkballs) No Data

vSI (Sinkers) Yes - 93.3

vSL (Sliders) Yes - 84.6

vCU (Curveballs) Yes - 79.1

vKC (Knuckle-Curves) Yes - 81.3

vEP (Ephesuses) Yes - 50.0

vCH (Changeups) Yes - 85.3

vSC (Screwballs) No Data

vKN (Knuckleballs) Yes - 55.4

vUN (Unknowns) No Data

vSW (Sweepers) Yes - 81.5 Calculated from my dataset*

Results - Velo Splits

EXPERIMENT B: CORRELATIONS, SIGNIFICANCE TESTS BROKE DOWN BY O/U LEAGUE AVG VELO, HB, VB...

CHANGEUP - BELOW AVG VELO (n=111):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.071	0.183	0.078	-0.072	-0.109	0.263	0.047
Velo pVal	0.459	0.055	0.416	0.452	0.255	0.005	0.624
Significant?	No	No	No	No	No	Yes	No

CHANGEUP - ABOVE AVG VELO (n=149):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.155	0.279	-0.004	-0.190	-0.178	-0.093	-0.242
Velo pVal	0.059	0.001	0.962	0.020	0.030	0.259	0.003
Significant?	No	Yes	No	Yes	Yes	No	Yes

CUTTER - BELOW AVG VELO (n=74):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.200	0.214	0.075	-0.123	-0.173	-0.127	0.116
Velo pVal	0.087	0.067	0.525	0.296	0.141	0.281	0.325
Significant?	No	No	No	No	No	No	No

CUTTER - ABOVE AVG VELO (n=51):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.296	0.272	-0.043	-0.358	-0.273	-0.339	-0.599
Velo pVal	0.035	0.053	0.765	0.010	0.052	0.015	<0.00001
Significant?	Yes	No	No	Yes	No	Yes	Yes

CURVEBALL - BELOW AVG VELO (n=102):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	0.153	0.052	0.309	0.110	-0.001	0.241	-0.024
Velo pVal	0.125	0.604	0.002	0.271	0.992	0.015	0.811
Significant?	No	No	Yes	No	No	Yes	No

CURVEBALL - ABOVE AVG VELO (n=109):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.271	0.510	0.280	-0.320	-0.319	-0.074	-0.209
Velo pVal	0.004	<0.00001	0.003	0.001	0.001	0.444	0.029
Significant?	Yes	Yes	Yes	Yes	Yes	No	Yes

FASTBALL - BELOW AVG VELO (n=177):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.063	0.292	0.257	-0.105	-0.174	0.139	0.111
Velo pVal	0.405	0.0001	0.001	0.164	0.021	0.065	0.141
Significant?	No	Yes	Yes	No	Yes	No	No

FASTBALL - ABOVE AVG VELO (n=184):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.202	0.355	-0.016	-0.224	-0.177	0.025	-0.189
Velo pVal	0.006	<0.00001	0.829	0.002	0.016	0.737	0.010
Significant?	Yes	Yes	No	Yes	Yes	No	Yes

SINKER - BELOW AVG VELO (n=113):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg

Velo corr	-0.046	0.195	0.196	-0.069	-0.093	0.159	-0.042
Velo pVal	0.629	0.038	0.037	0.468	0.327	0.092	0.659
Significant?	No	Yes	Yes	No	No	No	No

SINKER - ABOVE AVG VELO (n=107):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.035	0.255	0.298	-0.158	-0.204	-0.188	-0.244
Velo pVal	0.720	0.008	0.002	0.104	0.035	0.053	0.011
Significant?	No	Yes	Yes	No	Yes	No	Yes

SLIDER - BELOW AVG VELO (n=147):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	0.037	0.196	0.097	0.015	-0.057	0.209	0.095
Velo pVal	0.656	0.017	0.242	0.857	0.493	0.011	0.252
Significant?	No	Yes	No	No	No	Yes	No

SLIDER - ABOVE AVG VELO (n=172):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.292	0.280	-0.147	-0.331	-0.272	-0.090	-0.229
Velo pVal	0.0001	0.0002	0.054	<0.00001	0.0003	0.240	0.003
Significant?	Yes	Yes	No	Yes	Yes	No	Yes

SPLITTER - BELOW AVG VELO (n=12):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	0.329	0.117	-0.226	0.314	0.378	0.204	-0.545
Velo pVal	0.296	0.717	0.480	0.320	0.226	0.525	0.067
Significant?	No	No	No	No	No	No	No

SPLITTER - ABOVE AVG VELO (n=17):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	0.041	0.093	-0.040	0.146	0.204	0.116	-0.515

Velo pVal	0.876	0.722	0.879	0.576	0.432	0.658	0.034
Significant?	No	No	No	No	No	No	Yes

SWEEPER - BELOW AVG VELO (n=10):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	0.186	0.473	0.281	-0.240	-0.534	0.014	0.527
Velo pVal	0.607	0.167	0.432	0.504	0.112	0.969	0.118
Significant?	No	No	No	No	No	No	No

SWEEPER - ABOVE AVG VELO (n=11):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
Velo corr	-0.084	0.068	-0.581	-0.178	0.040	-0.047	-0.177
Velo pVal	0.806	0.843	0.061	0.600	0.907	0.891	0.602
Significant?	No	No	No	No	No	No	No

Break averages (in my somewhat limited dataset):

CH
HB - 14.3
VB - 31.4

CT
HB - 2.8
VB - 27.0

CV
HB - 8.9
VB - 53.9

FB
HB - 7.6
VB - 15.5

SK
HB - 14.7
VB - 22.8

SL
HB - 6.7
VB - 36.7

SP
HB - 11.0
VB - 32.1

SW
HB - 15.5
VB - 41.4

Results - HB Splits

CHANGEUP - BELOW AVG HB (n=111):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	-0.069	0.046	-0.060	-0.052	-0.077	-0.067	0.168
HB pVal	0.472	0.631	0.531	0.588	0.422	0.485	0.078
Significant?	No	No	No	No	No	No	No

CHANGEUP - ABOVE AVG HB (n=142):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	-0.255	0.237	0.161	-0.262	-0.292	-0.280	0.003
HB pVal	0.002	0.005	0.056	0.002	0.0004	0.0007	0.972
Significant?	Yes	Yes	No	Yes	Yes	Yes	No

CUTTER - BELOW AVG HB (n=68):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	0.002	-0.057	-0.205	0.081	0.107	0.027	0.380
HB pVal	0.987	0.644	0.093	0.512	0.385	0.827	0.001
Significant?	No	No	No	No	No	No	Yes

CUTTER - ABOVE AVG HB (n=58):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	-0.061	-0.011	-0.021	-0.113	-0.085	0.039	0.011
HB pVal	0.649	0.935	0.876	0.398	0.526	0.771	0.935
Significant?	No	No	No	No	No	No	No

CURVEBALL - BELOW AVG HB (n=112):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	0.041	-0.064	-0.080	-0.002	0.042	-0.080	-0.145
HB pVal	0.668	0.502	0.402	0.983	0.660	0.402	0.127
Significant?	No	No	No	No	No	No	No

CURVEBALL - ABOVE AVG HB (n=99):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	0.067	0.040	0.032	0.069	0.060	-0.057	0.067
HB pVal	0.510	0.694	0.753	0.497	0.555	0.575	0.510
Significant?	No	No	No	No	No	No	No

FASTBALL - BELOW AVG HB (n=188):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	0.071	0.040	-0.058	0.089	0.021	0.086	0.101
HB pVal	0.333	0.586	0.429	0.224	0.775	0.241	0.168
Significant?	No	No	No	No	No	No	No

FASTBALL - ABOVE AVG HB (n=175):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	0.014	0.099	0.068	-0.048	-0.062	-0.104	-0.111
HB pVal	0.854	0.192	0.371	0.528	0.415	0.171	0.144
Significant?	No	No	No	No	No	No	No

SINKER - BELOW AVG HB (n=94):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg

HB corr	-0.027	-0.158	-0.082	0.024	0.086	0.134	-0.104
HB pVal	0.796	0.128	0.432	0.819	0.410	0.198	0.318
Significant?	No	No	No	No	No	No	No

SINKER - ABOVE AVG HB (n=128):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	-0.003	0.083	-0.006	-0.048	-0.029	-0.155	-0.054
HB pVal	0.973	0.352	0.947	0.591	0.745	0.081	0.545
Significant?	No	No	No	No	No	No	No

SLIDER - BELOW AVG HB (n=194):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	0.020	-0.050	-0.031	0.026	0.035	-0.061	0.037
HB pVal	0.782	0.489	0.668	0.719	0.628	0.398	0.609
Significant?	No	No	No	No	No	No	No

SLIDER - ABOVE AVG HB (n=125):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	-0.082	0.104	-0.062	-0.091	-0.090	-0.185	0.041
HB pVal	0.364	0.248	0.492	0.313	0.318	0.039	0.650
Significant?	No	No	No	No	No	Yes	No

SPLITTER - BELOW AVG HB (n=9):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	-0.150	-0.447	-0.456	0.167	0.600	0.527	-0.626
HB pVal	0.700	0.228	0.217	0.668	0.088	0.145	0.071
Significant?	No	No	No	No	No	No	No

SPLITTER - ABOVE AVG HB (n=19):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	0.342	0.062	-0.063	0.374	0.464	0.541	-0.271

HB pVal	0.152	0.801	0.798	0.115	0.045	0.017	0.262
Significant?	No	No	No	No	Yes	Yes	No

SWEEPER - BELOW AVG HB (n=12):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	-0.359	0.342	0.118	-0.190	-0.161	-0.040	-0.131
HB pVal	0.252	0.277	0.715	0.554	0.617	0.901	0.685
Significant?	No	No	No	No	No	No	No

SWEEPER - ABOVE AVG HB (n=9):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
HB corr	-0.031	0.238	-0.094	0.172	0.111	-0.501	0.449
HB pVal	0.937	0.538	0.810	0.658	0.776	0.169	0.226
Significant?	No	No	No	No	No	No	No

Results - VB Splits

CHANGEUP - BELOW AVG VB (n=139):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.181	0.262	0.202	-0.265	-0.262	-0.340	-0.031
VB pVal	0.033	0.002	0.017	0.002	0.002	0.00004	0.717
Significant?	Yes	Yes	Yes	Yes	Yes	Yes	No

CHANGEUP - ABOVE AVG VB (n=121):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.049	0.158	-0.015	-0.021	-0.008	-0.072	-0.097
VB pVal	0.594	0.083	0.870	0.819	0.931	0.433	0.290
Significant?	No	No	No	No	No	No	No

CUTTER - BELOW AVG VB (n=61):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.046	0.135	-0.075	0.032	0.040	-0.048	0.163
VB pVal	0.725	0.299	0.565	0.807	0.760	0.713	0.209
Significant?	No	No	No	No	No	No	No

CUTTER - ABOVE AVG VB (n=65):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.075	0.034	-0.092	-0.072	-0.034	0.011	0.174
VB pVal	0.553	0.788	0.466	0.569	0.788	0.931	0.166
Significant?	No	No	No	No	No	No	No

CURVEBALL - BELOW AVG VB (n=114):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	0.026	-0.029	-0.059	0.056	0.056	-0.204	-0.038
VB pVal	0.784	0.759	0.533	0.554	0.554	0.029	0.688
Significant?	No	No	No	No	No	Yes	No

CURVEBALL - ABOVE AVG VB (n=101):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.032	0.147	-0.173	-0.036	0.015	-0.056	0.087
VB pVal	0.750	0.142	0.084	0.721	0.882	0.578	0.387
Significant?	No	No	No	No	No	No	No

FASTBALL - BELOW AVG VB (n=214):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	0.038	0.023	-0.056	0.071	0.054	-0.042	-0.081
VB pVal	0.580	0.738	0.415	0.301	0.432	0.541	0.238
Significant?	No	No	No	No	No	No	No

FASTBALL - ABOVE AVG VB (n=145):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.069	0.264	0.157	-0.126	-0.196	0.033	0.111
VB pVal	0.410	0.001	0.059	0.131	0.018	0.693	0.184
Significant?	No	Yes	No	No	Yes	No	No

SINKER - BELOW AVG VB (n=132):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.069	0.036	0.003	-0.097	-0.028	-0.075	-0.301
VB pVal	0.432	0.682	0.973	0.269	0.750	0.392	0.0005
Significant?	No	No	No	No	No	No	Yes

SINKER - ABOVE AVG VB (n=92):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.096	0.214	-0.074	-0.121	-0.047	-0.134	-0.125
VB pVal	0.363	0.041	0.483	0.251	0.657	0.203	0.235
Significant?	No	Yes	No	No	No	No	No

SLIDER - BELOW AVG VB (n=160):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.058	0.112	0.258	-0.096	-0.130	-0.248	-0.049
VB pVal	0.466	0.158	0.001	0.227	0.101	0.002	0.538
Significant?	No	No	Yes	No	No	Yes	No

SLIDER - ABOVE AVG VB (n=159):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.187	0.074	-0.110	-0.161	-0.156	-0.400	-0.006
VB pVal	0.018	0.354	0.167	0.043	0.050	<0.00001	0.940
Significant?	Yes	No	No	Yes	Yes	Yes	No

SPLITTER - BELOW AVG VB (n=16):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg

VB corr	0.338	-0.282	0.187	0.454	0.507	0.079	-0.229
VB pVal	0.200	0.290	0.488	0.077	0.045	0.771	0.394
Significant?	No	No	No	No	Yes	No	No

SPLITTER - ABOVE AVG VB (n=13):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.350	-0.012	-0.678	-0.140	0.259	0.309	-0.438
VB pVal	0.241	0.969	0.011	0.648	0.393	0.304	0.134
Significant?	No	No	Yes	No	No	No	No

SWEEPER - BELOW AVG VB (n=10):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.543	0.553	0.139	-0.321	-0.246	-0.139	-0.321
VB pVal	0.105	0.097	0.702	0.366	0.493	0.702	0.366
Significant?	No	No	No	No	No	No	No

SWEEPER - ABOVE AVG VB (n=11):

	era	k_pct	bb_pct	opp_ops	baa	ev_avg	la_avg
VB corr	-0.077	-0.070	-0.158	0.306	0.496	-0.467	0.148
VB pVal	0.822	0.838	0.643	0.360	0.121	0.148	0.664
Significant?	No	No	No	No	No	No	No

Conclusions

I think it makes sense to analyze the results by pitch. This way we can see the performance correlations for all 3 metrics (velo, vb, hb) for each pitch over all the data juxtaposed with the split data.

Changeups

Velocity

Through a ***full range of velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity changeups and an increase in:

- k_pct
- bb_pct

As well as a decrease in:

- opp_ops
- baa
- la_avg

At ***below league average velocities***, a statistically significant correlation ($p<0.05$) was found between higher velocity changeups and an increase in:

- ev_avg

At ***above league average velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity changeups and an increase in:

- k_pct

As well as a decrease in:

- opp_ops
- baa
- la_avg

HB

Through a ***full range of HB***, statistically significant correlations ($p<0.05$) were found between higher HB changeups and an increase in:

- k_pct

As well as a decrease in:

- opp_ops
- baa
- ev_avg

At ***below league average HB***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average HB***, a statistically significant correlation ($p<0.05$) was found between higher HB changeups and an increase in:

- k_pct

As well as a decrease in:

- era

- opp_ops
- baa
- ev_avg

VB

Through a ***full range of VB***, statistically significant correlations ($p<0.05$) were found between higher VB changeups and a decrease in:

- bb_pct
- ev_avg
- la_avg

At ***below league average VB***, statistically significant correlations ($p<0.05$) were found between higher VB changeups and an increase in:

- k_pct
- bb_pct

As well as a decrease in:

- era
- opp_ops
- baa
- ev_avg

At ***above league average VB***, no statistically significant correlations ($p<0.05$) were found.

Cutters

Velocity

Through a ***full range of velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity cutters and an increase in:

- k_pct

As well as a decrease in:

- era
- opp_ops
- baa
- ev_avg
- la_avg

At ***below league average velocities***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity cutters and a decrease in:

- era
- opp_ops
- ev_avg
- la_avg

HB

Through a ***full range of HB***, no statistically significant correlations ($p<0.05$) were found.

At ***below league average HB***, a statistically significant correlation ($p<0.05$) was found between higher HB cutters and an increase in:

- la_avg

At ***above league average HB***, no statistically significant correlations ($p<0.05$) were found.

VB

Through a ***full range of VB***, statistically significant correlations ($p<0.05$) were found between higher VB cutters and an increase in:

- baa
- ev_avg

As well as a decrease in:

- k_pct

At ***below league average VB***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average VB***, no statistically significant correlations ($p<0.05$) were found.

Curveballs

Velocity

Through a ***full range of velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity curveballs and an increase in:

- k_pct
- bb_pct
- ev_avg

As well as a decrease in:

- baa
- la_avg

At ***below league average velocities***, a statistically significant correlation ($p<0.05$) was found between higher velocity curveballs and an increase in:

- bb_pct
- ev_avg

At ***above league average velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity curveballs and an increase in:

- k_pct
- bb_pct

As well as a decrease in:

- era
- opp_ops
- baa
- la_avg

HB

Through a ***full range of HB***, statistically significant correlations ($p<0.05$) were found between higher HB curveballs and a decrease in:

- ev_avg

At ***below league average HB***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average HB***, no statistically significant correlations ($p<0.05$) were found.

VB

Through a ***full range of VB***, statistically significant correlations ($p<0.05$) were found between higher VB curveballs and a decrease in:

- k_pct

At ***below league average VB***, a statistically significant correlation ($p<0.05$) was found between higher VB curveballs and a decrease in:

- ev_avg

At ***above league average VB***, no statistically significant correlations ($p<0.05$) were found.

Fastballs

Velocity

Through a ***full range of velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity fastballs and an increase in:

- k_pct
- bb_pct

As well as a decrease in:

- era
- opp_ops
- baa

At ***below league average velocities***, a statistically significant correlation ($p<0.05$) was found between higher velocity fastballs and an increase in:

- k_pct
- bb_pct

As well as a decrease in:

- baa

At ***above league average velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity fastballs and an increase in:

- k_pct

As well as a decrease in:

- era
- opp_ops
- baa
- la_avg

HB

Through a ***full range of HB***, no statistically significant correlations ($p<0.05$) were found.

At ***below league average HB***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average HB***, no statistically significant correlations ($p<0.05$) were found.

VB

Through a ***full range of VB***, no statistically significant correlations ($p<0.05$) were found.

At ***below league average VB***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average VB***, statistically significant correlations ($p<0.05$) were found between higher VB fastballs and an increase in:

- k_pct

As well as a decrease in:

- baa

Sinkers

Velocity

Through a ***full range of velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity sinkers and an increase in:

- k_pct
- bb_pct

As well as a decrease in:

- opp_ops
- baa
- la_avg

At ***below league average velocities***, a statistically significant correlation ($p<0.05$) was found between higher velocity sinkers and an increase in:

- k_pct
- bb_pct

At ***above league average velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity sinkers and an increase in:

- k_pct
- bb_pct

As well as a decrease in:

- baa
- la_avg

HB

Through a ***full range of HB***, statistically significant correlations ($p<0.05$) were found between higher HB sinkers and a decrease in:

- la_avg

At ***below league average HB***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average HB***, no statistically significant correlations ($p<0.05$) were found.

VB

Through a ***full range of VB***, statistically significant correlations ($p<0.05$) were found between higher VB sinkers and an increase in:

- k_pct
- bb_pct

As well as a decrease in:

- baa
- la_avg

At ***below league average VB***, a statistically significant correlation ($p<0.05$) was found between higher VB sinkers and a decrease in:

- la_avg

At ***above league average VB***, statistically significant correlations ($p<0.05$) were found between higher VB sinkers and an increase in:

- k_pct

Sliders

Velocity

Through a ***full range of velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity sliders and an increase in:

- k_pct
- ev_avg

As well as a decrease in:

- opp_ops
- baa
- la_avg

At ***below league average velocities***, a statistically significant correlation ($p<0.05$) was found between higher velocity sliders and an increase in:

- k_pct
- ev_avg

At ***above league average velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity sliders and an increase in:

- k_pct

As well as a decrease in:

- era
- opp_ops
- baa
- la_avg

HB

Through a ***full range of HB***, statistically significant correlations ($p<0.05$) were found between higher HB sliders and a decrease in:

- ev_avg

At ***below league average HB***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average HB***, statistically significant correlations ($p<0.05$) were found between higher HB sliders and a decrease in:

- ev_avg

VB

Through a ***full range of VB***, statistically significant correlations ($p<0.05$) were found between higher VB sliders and an increase in:

- k_pct

As well as a decrease in:

- era
- opp_ops
- baa

At ***below league average VB***, a statistically significant correlation ($p<0.05$) was found between higher VB sliders and an increase in:

- bb_pct

As well as a decrease in:

- ev_avg

At ***above league average VB***, statistically significant correlations ($p<0.05$) were found between higher VB sliders and a decrease in:

- era
- opp_ops
- baa
- ev_avg

Splitters

Velocity

Through a ***full range of velocities***, no statistically significant correlations ($p<0.05$) were found.

At ***below league average velocities***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average velocities***, statistically significant correlations ($p<0.05$) were found between higher velocity splitters and a decrease in:

- la_avg

HB

Through a ***full range of HB***, statistically significant correlations ($p<0.05$) were found between higher HB splitters and an increase in:

- baa

At ***below league average HB***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average HB***, statistically significant correlations ($p<0.05$) were found between higher HB splitters and an increase in:

- baa
- ev_avg

VB

Through a ***full range of VB***, no statistically significant correlations ($p<0.05$) were found.

At ***below league average VB***, a statistically significant correlation ($p<0.05$) was found between higher VB splitters and an increase in:

- baa

At ***above league average VB***, statistically significant correlations ($p<0.05$) were found between higher VB splitters and a decrease in:

- bb_pct

Sweepers

Velocity

Through a ***full range of velocities***, no statistically significant correlations ($p<0.05$) were found.

At ***below league average velocities***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average velocities***, no statistically significant correlations ($p<0.05$) were found.

HB

Through a ***full range of HB***, no statistically significant correlations ($p<0.05$) were found.

At ***below league average HB***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average HB***, no statistically significant correlations ($p<0.05$) were found.

VB

Through a ***full range of VB***, statistically significant correlations ($p<0.05$) were found between higher VB sweepers and an increase in:

- opp_ops
- baa

As well as a decrease in:

- k_pct

At ***below league average VB***, no statistically significant correlations ($p<0.05$) were found.

At ***above league average VB***, no statistically significant correlations ($p<0.05$) were found.

Extrapolations (TLDR)

What the results seem to suggest EVEN WITHOUT EXPLICIT QUANTITATIVE PROOF
Very important to recognize that **these results are suggested from correlations** and may be presented in a way that implies causation.

Changeups

Higher velo changeups are better overall, however a velo increase in below avg velo changeup is detrimental. The benefits appear when increasing above league average.
The same is true for horizontal break, with benefits only appearing for the above average group.
A strange trend for vertical break: overall beneficial to increase VB, but the splits imply that it is only beneficial when VB is below average. Essentially implies that league average VB is ideal.

Cutters

Increase in velo is good overall, but benefits appear when increasing above league average.
Very little significant findings for HB, surprising for cutters. Full range suggests that lower VB is better across all ranges.

Curveballs

Increase in velo is good overall, but benefits appear when increasing above league average.
Some evidence (but not a lot) suggesting that higher HB is better across all HB ranges.
Interesting result that increase in VB is correlated with lower k_pct, probably because of less velo. Otherwise no VB trends to report, which in itself is interesting since such an emphasis is placed on increasing curveball VB.

Fastballs

Unsurprisingly, velo increase is beneficial across all velo ranges. Notably, velo increase is still beneficial when below league avg. HB is insignificant. Increasing VB is beneficial only if already above average.

Sinkers

Velo increase is beneficial across all ranges but is more beneficial to the above average group. HB is insignificant. VB increase is beneficial across all ranges.

Sliders

Velo increase is beneficial across all ranges, especially above average. Some (but not a lot) of evidence suggesting increase in HB is beneficial, especially in above league average situations. VB is beneficial across the board, but the true impact occurs in the above average group.

Splitters

Very little evidence supporting benefits from increase in velo. Some (but not a lot) of evidence that HB increase is detrimental. Very little impact evident by VB increase. Small amount of splitter data makes findings unreliable anyways.

Sweepers

No evidence suggesting velo or HB impacts performance. Evidence suggests that VB increase is detrimental, but small sample size makes sweeper findings unreliable.

Summary

Although there does seem to be evidence that velocity increase is most beneficial for pitches that are already above league average, a velocity increase in below average cases are rarely harmful. Instead of considering league average the ideal “hitting speed” for hitters, maybe it would make more sense to think of league average velocity as a threshold at which further velocity gains begin to make a pitcher significantly more effective.

When it comes to pitch movement, it is hard to draw such a broad conclusion. Performance impact seems to vary by pitch for both horizontal and vertical break. Even in cases such as horizontal break for cutters/sliders and vertical break for curveballs, it is hard to see definitive proof that more is always better. This could be partially due to the fact that we aren’t controlling for velocity. We know that velocity is a strong predictor of performance, and due to physics, a curveball that has more vertical break will likely be slower (more time for break to occur due to gravity). In future studies, it may be interesting to try to establish pitch movement scores by normalizing over velocity. I think it is hard to imagine that a naive increase in movement without a decrease in velocity would negatively impact performance, but evidence for that should be found in a future study.