Untitled

David Scolari, Harrison Snell, Brandon Williams

5/8/2022

Introduction

Hitting a baseball has been shown to be one of the most difficult tasks in all of sports. Anybody who has gone to a batting cage and tried the fastest machine for fun knows how hard it can be to hit a normal fastball. That does not even include the possibilities of breaking balls and off-speed pitches. Batters have a fraction of a fraction of a second to recognize a pitch, determine if that pitch will be a ball or a strike, and set their swing into motion. Considering the human body can only move so fast to get the bat over the plate, the batter has to make their decision moments after the ball leaves the pitcher's hand. At that point, even some professional baseball players have to guess on whether they are facing a changeup or a fastball or a curveball. Any additional information the batter can have is crucial. The goal of our model is to provide some of that additional information.

Pitch predicting is a very important part of baseball. So important that teams will do almost anything for insight on what the opposing team's pitcher is going to throw. Other than illicit methods, hitters are trained to use information on a pitcher's previous pitch throwing behvior to guess which pitch is coming next. We can model this deciscion making process with machine learning techniques.

By modeling pitch prediction, we can learn what information is relevant to a pitcher's pitch selection. We can also better understand what makes a pitcher more unpredictable than other. It's gonna be good

In the following report will talk about why pitches are different, share evidence that pitchers have different strategies from at bat to at bat, and describe the engineering of our model, which is basically a trashcan.

Why do we care about a pitch?

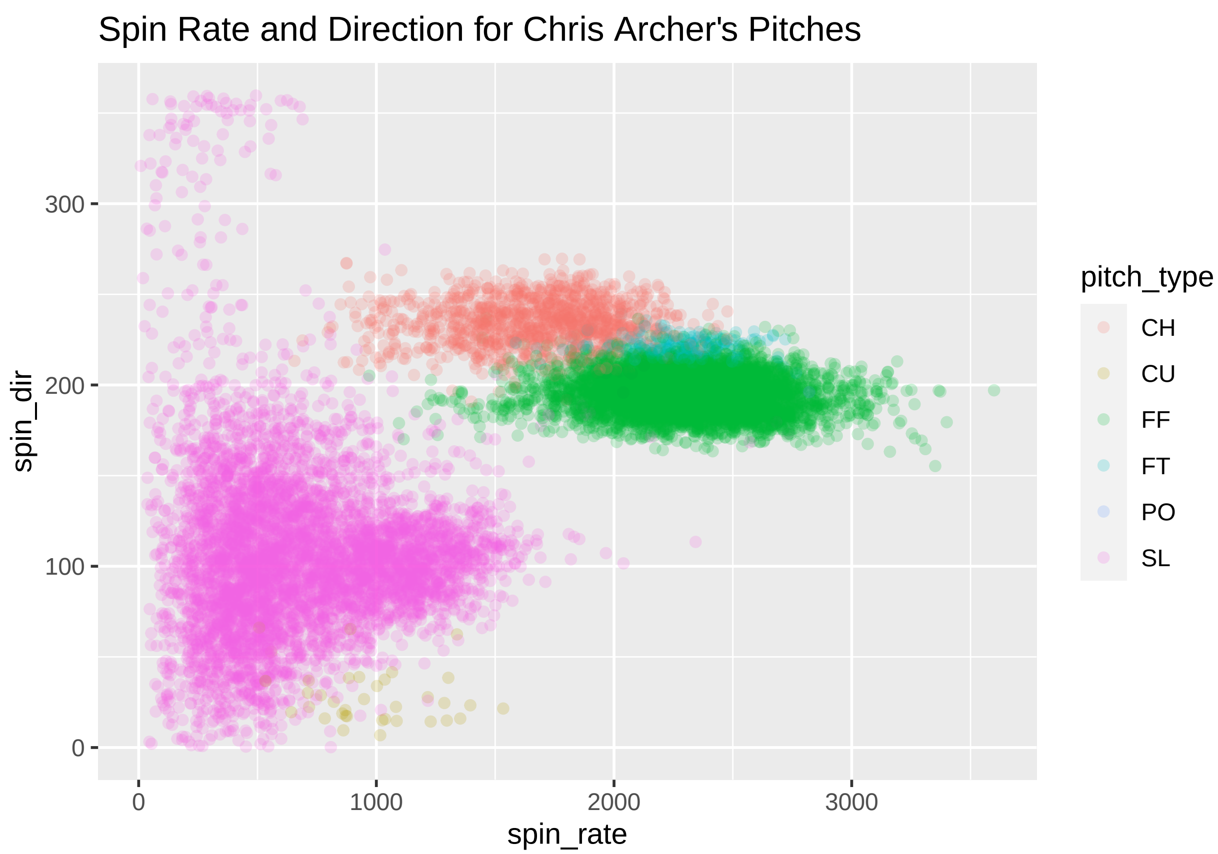
Not all pitches are made the same. From sliders to cutters to knuckleballs, each pitch tends to have its own characteristics that define it.

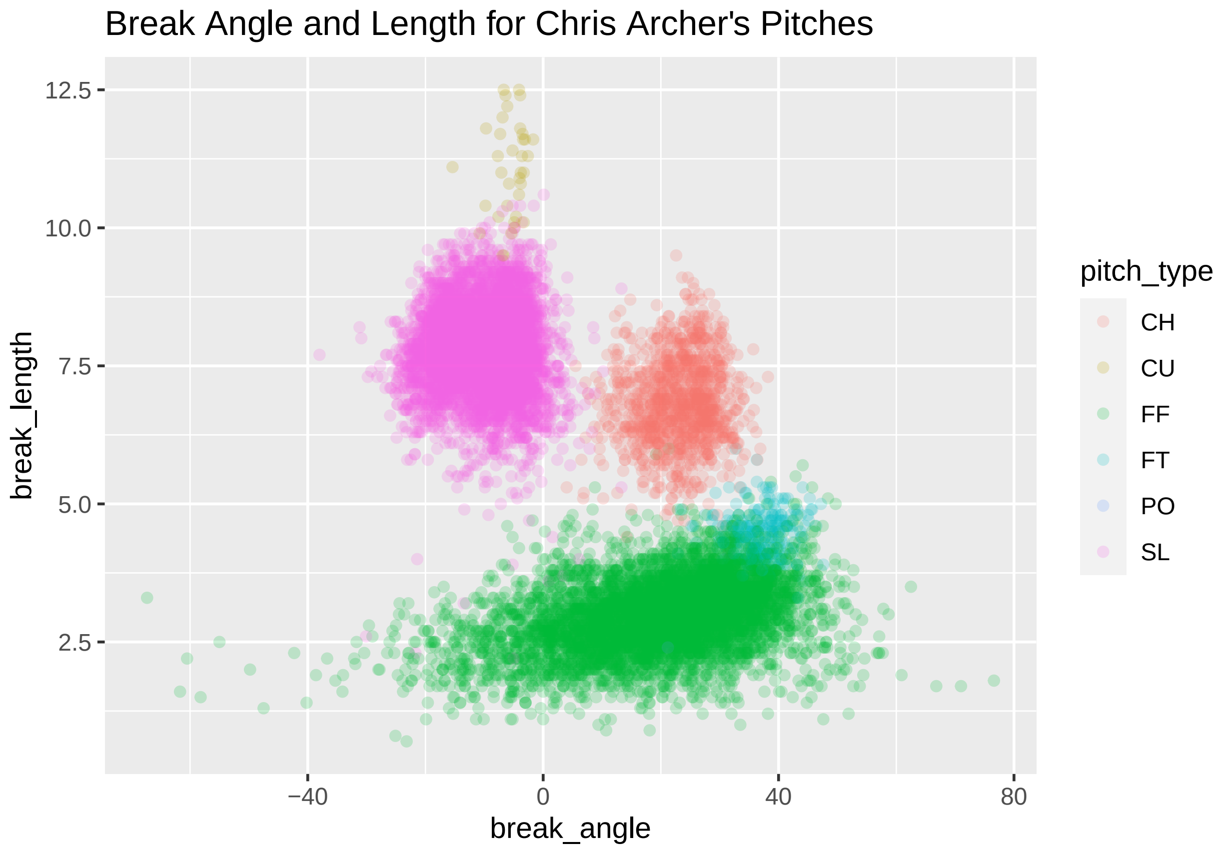
Table 1: Averages of Different Pitch Characteristics

pitch_type	px	pz	start_speed	end_speed	spin_rate	spin_dir	break_angle	break_length	break_y
CH	-0.2004531	1.859088	85.81363	79.19324	1734.5396	212.37185	8.859414	7.694970	23.82537
CU	0.0756174	1.802014	78.16450	72.29356	1446.6248	89.21560	-9.948970	12.569182	23.83973
EP	0.1831130	2.022092	66.51596	61.32254	1381.7363	44.09603	-9.033333	16.907512	23.85540
FA	-0.1840000	1.768333	88.90000	80.60000	2448.7507	148.70967	-7.066667	7.500000	23.70000
FC	0.2211626	2.310223	88.54320	81.99196	1018.4097	164.45461	-5.357321	6.093412	23.85148

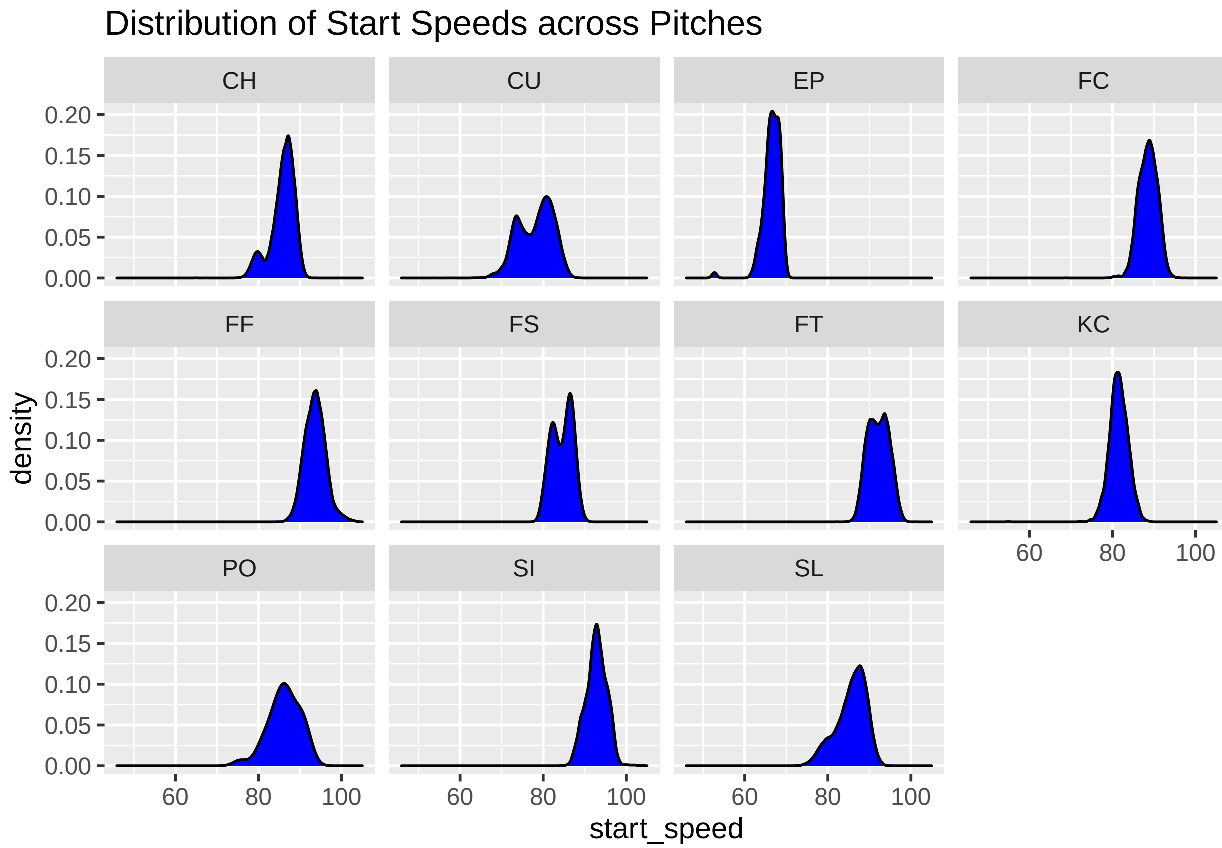
pitch_type	px	pz	start_speed	end_speed	spin_rate	spin_dir	break_angle	break_length	break_y
FF	0.0071570	2.638026	93.62794	85.90142	2175.7277	193.63051	13.491285	3.721318	23.79359
FS	-0.3771798	1.915900	84.56765	78.60294	1483.1980	222.07379	17.104412	6.652941	23.85882
FT	-0.0227988	2.351487	92.15093	84.67758	2129.6691	169.48873	-5.557804	5.749812	23.80294
KC	0.0839310	1.709555	81.50706	75.23154	1318.3440	89.54775	-6.516319	11.551069	23.82959
PO	0.9680839	4.114041	86.21714	79.08000	1928.4078	210.20437	21.037143	5.171429	23.79143
SI	-0.1321975	2.309338	92.81275	85.41892	2008.5825	216.36221	21.267940	5.879977	23.80938
SL	0.2318519	1.866506	85.58814	79.35203	901.0771	162.50623	-3.270639	8.079149	23.85377
UN	3.4876667	3.531333	41.86667	39.80000	823.2617	218.88567	-1.166667	31.800000	23.96667

From the table, we essentially see the obvious that is pitches are different. On average we find that fastballs start at almost 94 mph and changeups start at 86 mph. The average break length of a fastball is about 4 inches, which is not too bad, the point that a batter sees a fastball leave the pitcher's hand will most likely cross following a straight line. Compare that to a curveball with almost 13 inches of break and we start to see why knowing the next pitch can be helpful.

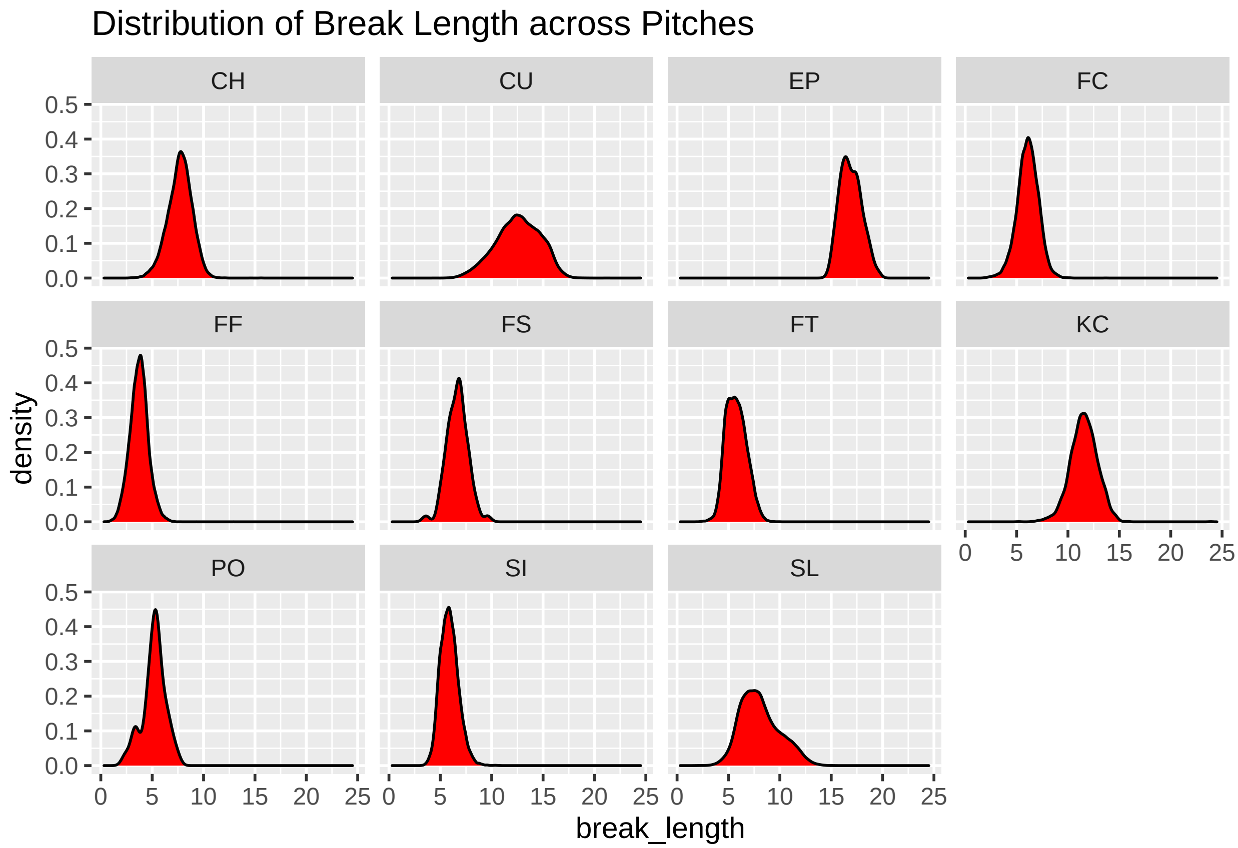




The plots of Chris Archer's pitches give us a glimpse of where one type of pitch becomes another. For Archer, we see that his slider tends to have a relatively low spin rate, with his fastballs having higher spin rates and his changeups in between. As we would expect, Archer's fastballs have little break. His slider however, has a large but variable amount of break and always tends to break in the same direction. These visuals help us understand where a pitch may be going. If we can predict the next pitch is a slider, then we can start to understand the direction and magnitude of break in the pitch.



Distribution of Start Speeds across Pitches
CH
CU
EP
FC
FF
FS
FT
KC
PO
SI
SL
density
0.20
0.15
0.10
0.05
0.00
60
80
100
start_speed



Mostly as an exposition into our data, we found the graphs of distributions of break length and start speed to be interesting. They provide a more throrough comparison of pitch types because every pitcher is slightly different. Some pitchers may have a lot more movement on their sinker than others. Similarly, Aroldis Chapman throwing 105 mph fastballs is certainly not the norm in the league. Getting to see the variability of break length and speeds begins to paint the picture of why some pitchers are better than others.