Context: I was tasked with determining an order of preference for three pitchers based on some given data, and explaining my results and approach.

To achieve the goal of determining an order of preference for the three pitchers in the data subset, I took a look at several statistics and created some graphics. For simplicity, I will refer to the pitchers with IDs 516888, 517386, and 570026 as Pitcher A, Pitcher B, and Pitcher C, respectively. I have ranked the pitchers in the following order:

1. ID 570026 (Pitcher C)
2. ID 516888 (Pitcher A)
3. ID 517386 (Pitcher B)

I used Python to perform my analysis. For several statistics it was useful to separate each pitcher's data further by MLB, AAA, or SS. This chart gives the sample size of each category.

Table: Number of Pitches in Dataset

	MLB	**AAA**	**SS**
Pitcher A	190	172	0
Pitcher B	151	284	0
Pitcher C	26	126	28

AAA was the most comparable in terms of sample size across the three pitchers, so I tried to make most of my comparisons there. The sample size for Pitchers A and B are similar for the MLB, but it is difficult to draw any conclusions based on Pitcher C's performance in SS or MLB. While Pitcher B has the largest total sample size, Pitcher A has the most recorded in the MLB, the highest level of play. Pitcher C has the fewest recorded total (180) at less than half that of Pitcher A (362) or Pitcher B (435).

A few stats first came to mind when I started to think about how pitcher performance is measured, including ERA and K/BB. Because the dataset doesn't provide exact information on which runs are earned vs. unearned or the exact game situation when the reliever entered, I couldn't calculate ERA easily. However, I used "runs_scored_on_play" as a really rough substitute for earned runs, just to see where each pitcher stood. Obviously, this is a super rough number and I only used it as the most basic of starting points. See the table "Some Basic Pitching Stats, Separated by Level of Play" below for results.

Pitcher C didn't have any walks recorded for MLB or SS. The ERA estimate is still pretty useless at best, but since we have the runs scored on each play available, it's worth taking a look. In the MLB, Pitcher A does the best, but in AAA, Pitcher C does much better than A and B. For K/BB, Pitcher B does the best in the MLB and AAA. Overall, I didn't draw any lasting conclusion from this information alone because the advantages were so evenly split across unreliable data.

ERA is a crowd favorite, but to evaluate a pitcher on his own merits, we have to be able to "subtract out" the effect of his defense and luck. I decided to calculate FIP in hopes of getting a better grasp on the pitchers' true capabilities. FIP is nice because it's easily read on the same scale as ERA. Also, FIP has been shown to predict future ERA better than ERA itself does, because it isolates factors that the pitcher can control and attempts to discard those components

Table: Some Basic Pitching Stats, Separated by Level of Play

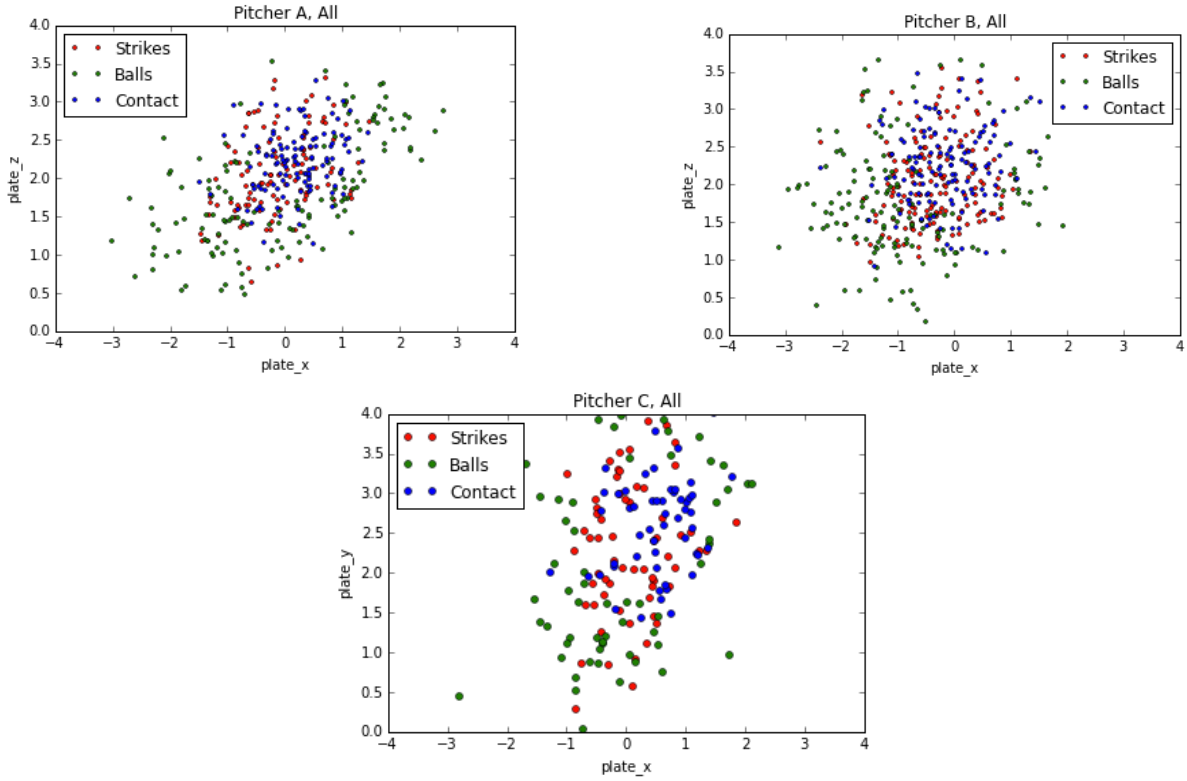
MLB	**ERA Estimate**	**K/BB**
Pitcher A	3.97	4.0
Pitcher B	9.39	4.5
Pitcher C	6.75	-
AAA	**ERA Estimate**	**K/BB**
Pitcher A	4.85	3.33
Pitcher B	5.52	6.75
Pitcher C	0.0	2.2
SS	**ERA Estimate**	**K/BB**
Pitcher A	-	-
Pitcher B	-	-
Pitcher C	0.0	-

that he cannot control. In AAA ball, Pitchers A, B, and C had FIPs of 3.05, 2.11, and 2.27, respectively, versus the lefty hitters. I calculated their WHIPs as well: 1.0, 1.16, and 0.78. Judging by FIP, the ranking should be BCA, but by WHIP it would be CAB. Either way, Pitcher C appeared to be better than Pitcher A at this point. Batting average against is another important metric because it directly measures how the batters hit against the pitcher. In AAA, Pitcher A's batting average against versus lefties was .189; Pitcher B's was .206; and Pitcher C's was .167. I interpreted this as more confirmation that Pitcher B should be ranked last. In the MLB, Pitcher A's batting average against versus lefties was .196 and Pitcher C's was 0.500. Obviously, .500 is terrible, but as I mentioned earlier, it is very hard to compare A and C on their MLB stats.

It's crucial for a pitcher to throw more than one type of pitch, just as it is for a pitcher to have good fastball velocity. These can be compared for the three pitchers regardless of level of play. Pitcher A pretty much only throws fastballs and sliders. Pitcher B throws a good mix of fastballs, curveballs, and "Other". Pitcher C throws mostly fastballs and sliders, with a smattering of curveballs. I think having three pitches is better than having two, in general. Pitcher A's average fastball velocity is 87.37 mph. Pitcher B's is 81.73 mph, and Pitcher C's is 94.35 mph. This is where I began to take issue with Pitcher B. Pitcher C had the most variation in pitch speed by a long shot, and variation helps with deception. Another point scored for Pitcher C.

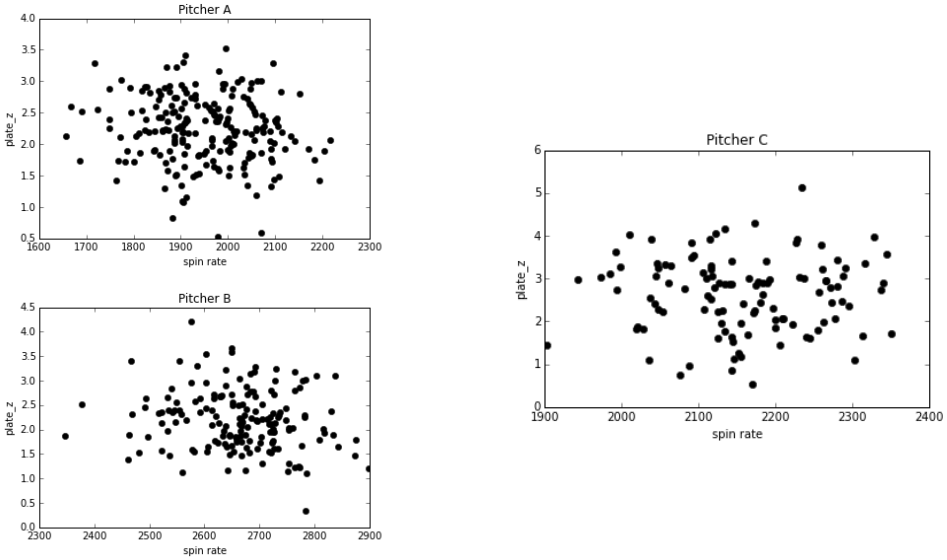
I took a look at contact rate on pitches in the zone, and found that MLB hitters made contact with Pitcher A about 48% of the time. They made contact with Pitcher B about 55% of the time. The less contact on balls in the zone, the more strikes, and the more strikes, the better. However, AAA hitters made contact with Pitcher B about 40.2% of the time and with Pitcher A about 57.14% of the time, so it flip-flopped with level of play. To get more information on each pitcher's pitches specifically, I plotted where the pitches were being thrown. The hope here was to glean some sort of information about where each pitcher gets hitters to miss, and where each pitcher is getting hit the most. See the three graphs below (Pitch Type and Location).

Graphs: Pitch Type and Location



Where green and red are mixed, the pitcher is getting batters to swing at stuff outside the zone. You can see that Pitchers B and C are getting swings and misses at pitches thrown low and away. Most contact is happening inside, especially for Pitcher B. Pitcher A doesn't seem to be as deceptive as Pitcher C on balls that cross the plate low and away.

Graphs: Fastballs – Vertical Location Across Plate vs. Spin Rate



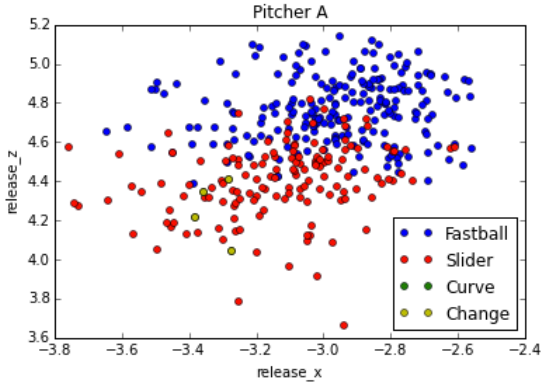
As for off-speed pitches, a higher spin rate is almost always better. I calculated each pitcher's average spin rate on their off-speed stuff. Pitcher A averages 2,445 rpm, Pitcher B averages 2,943 rpm, and Pitcher C averages 2,336 rpm. Here, Pitchers A and C are pretty close to each other, but Pitcher B wins out. A pitcher who throws at a high spin rate will do well to stay high in the zone with his fastballs; a pitcher who throws at a lower spin rate will do better low in the zone with his fastballs. The set of graphs below (Fastballs – Vertical Location Across Plate vs. Spin Rate) shows each pitcher's fastball location in the vertical direction vs. the pitch's spin rate. They were constructed with the hope of observing a relationship between vertical pitch location as it crosses the plate and spin rate for fastballs. Pitchers A and C are pretty well spread out, but Pitcher B slightly tends to be lower in the zone when his fastballs have a higher spin rate. Unfortunately for Pitcher B, this is opposite of what is usually effective. On the subject of fastballs, I checked out extension on fastball pitches. Pitcher C gets a bit closer to the plate and reaches full extension more often than A and B, which gives him another advantage.

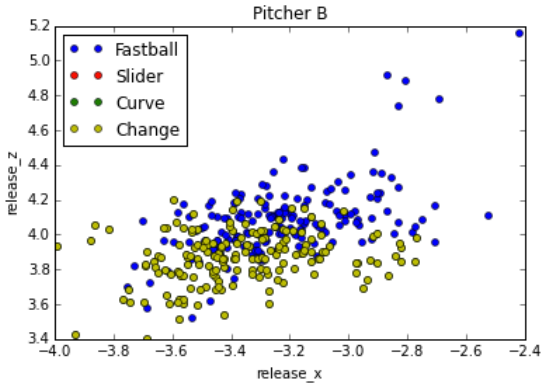
By definition, a pitcher does his best work when the batter can't figure out what he is doing. As I mentioned above, pitchers need to be deceptive. Part of being deceptive is throwing several types of pitches, and keeping batters on their toes. A pitcher needs to throw different pitches from the same release point so that the batter doesn't know what's coming. I generated three scatter plots (Release Point by Pitch Type) to see where each pitcher releases each type of pitch. A deceptive pitcher will have lots of overlap. This is very important and actually gives us information on what to expect versus righty hitters as well. The type of clustering apparent from Pitchers A and B is dangerous. It actually begins to look like pitches are being tipped. Contrast that with Pitcher C, whose release points are nicely mixed regardless of pitch type.

After I worked on those main performance indicators, there were a few little interesting things to look at as well. One item of particular interest to me was count situations. Who throws more first pitch strikes? Pitchers B and C throw first pitch strikes about 50% of the time, while Pitcher A throws them only 37% of the time. This isn't necessary to do, but it certainly doesn't hurt to get ahead in the count from the beginning. In fact, Pitcher B stays ahead in the count more often (78% of the time he pitches) than Pitchers A and C (about 70%).

As a last point, I inspected the exit speeds of pitches hit by these lefty batters. The table below summarizes my findings. In general, a higher exit speed signals trouble. Bear in mind also that because Pitcher C throws a lot harder than Pitchers A and B, the exit speeds off of his pitches will be faster, even if the hitter doesn't quite get ahold of the ball. Of course, it depends on launch angle as well, and that is an important component. Couple this information with the next table containing information on groundball percentage and fly ball rate for each pitcher. I chose to compare values from AAA because all pitchers have a decent number of samples for that level as opposed to MLB and SS. Even though Pitcher C has the highest exit speed in the MLB, despite his high fastball velocity he still has a lower average exit speed than Pitcher A in AAA. Pitcher C's exit speed average also deviates the most from his average velocity. This means hitters have a more difficult time sending his pitches out at high velocities. And, even though Pitcher C's exit speed is highest of the three in MLB, his fly ball rate is by far the lowest, and his groundball rate is the median.

Graphs: Release Point by Pitch Type





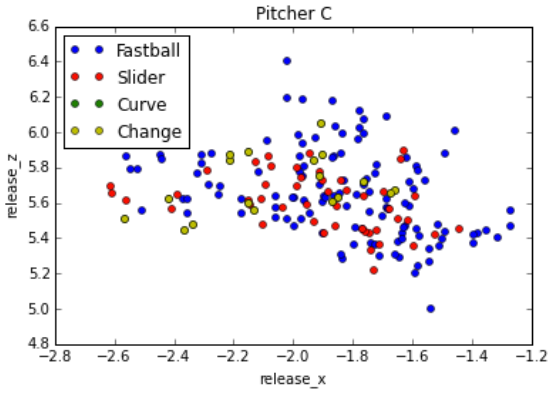


Table: Average Exit Speeds

Average Exit Speed (mph)	**MLB**	**AAA**
Pitcher A	82.22	81.49
Pitcher B	79.59	72.69
Pitcher C	84.98	77.01

Table: Groundball and Flyball Rates, MLB

	GB%	**FB%**
Pitcher A	45.28	18.87
Pitcher B	14.29	15.87
Pitcher C	23.3	6.67