Analysis of Statcast data from a random MLB Game

Ruslan Davtian

December 16, 2018

Data Description:

The dataset Data Sample.csv represents pitch by pitch baseball game details for a game between Team 1 and Team 2. There are 292 rows which represent pitches in the ball game. The columns of the dataset include pitch number, plate appearance of inning, pitch number of plate appearance, batter, pitcher, result of pitch, result of plays, pitch type, pitch location, etc. My main objective for this analysis is to understand the game flow and how the game was played using the dataset without needing to watch the game. First, I wanted to create a box score to analyze both teams and dive into individual players. For players, I compared the starting pitchers and bullpen as well as hitters for both teams.

Below, is a new table aggreated using the official data set from Data Sample.csv file. For each team and each pitcher, I calculated the total pitches thrown, total innings thrown (IP), total runs given up (not ERA due to possible unearned runs), total strikeouts (K), total walks (BB), number of ground balls (GB), number of line drives (LD), and number of fly balls (FB). From this table, both starting pitchers had good games but the starting pitcher for Team 2 pitched slightly better as the pitcher went 6.1 innings pitched with 9 strikeouts, 1 walk, and only 2 runs given up on 101 pitches. For team 1, their starter went 5.1 innings with only 4 strikeouts while giving up 3 runs on 81 pitches. Also, the starter for team 2 gave up almost twice less line drives (6) compared to the starter for team 1 (11). For team 2, the starter and bullpen had better pitching performances than the pitchers from team 1. I need this table below to create a box score with starting pitchers information.

PitcherTeam	Pitcher	Total_Pitches	IP	Runs_Given_Up	K	BB	GB	LD	FB
Team 1	Pitcher 1	81	5.3333333	3	4	0	5	11	1
Team 2	Pitcher 1	101	6.3333333	2	9	1	7	6	3
Team 1	Pitcher 2	15	0.6666667	1	1	1	0	1	0
Team 2	Pitcher 2	14	1.3333333	0	0	1	2	0	0
Team 1	Pitcher 3	3	0.3333333	0	0	0	1	0	0
Team 2	Pitcher 3	20	1.3333333	1	1	0	2	2	1
Team 1	Pitcher 4	9	0.6666667	0	1	0	1	0	0
Team 1	Pitcher 5	10	0.3333333	0	0	0	0	1	0
Team 1	Pitcher 6	22	0.3333333	0	1	2	2	1	0
Team 1	Pitcher 7	17	0.3333333	1	1	0	0	2	1

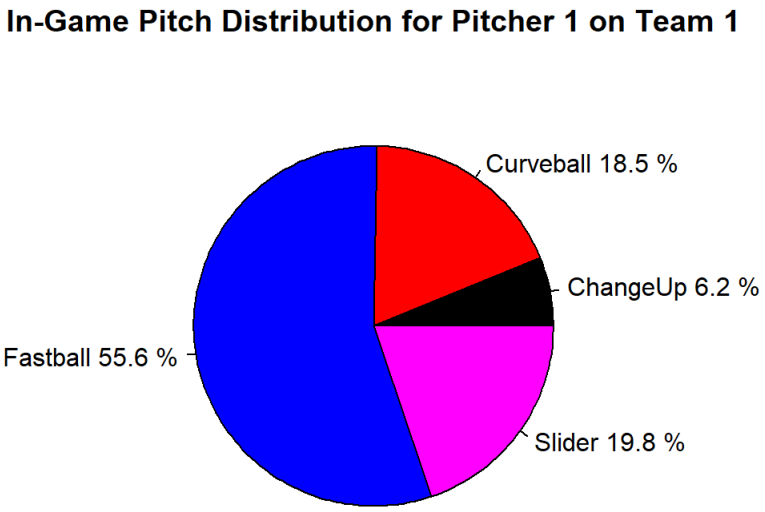
Below are two tables that summarize the ball game and pitchers for both teams. The first table is the official box score which summarizes the game. Team 2 beat Team 1 5-3 and team 2 had better offensive numbers across the board from hits to on-base percentage to slugging percentage. The starting pitcher information is included with the box score as well. The starter for team 2 pitched a quality start while the starter for team 1 pitched a good game but not a quality start. Let's take a quick look at the overall performances of each bullpen

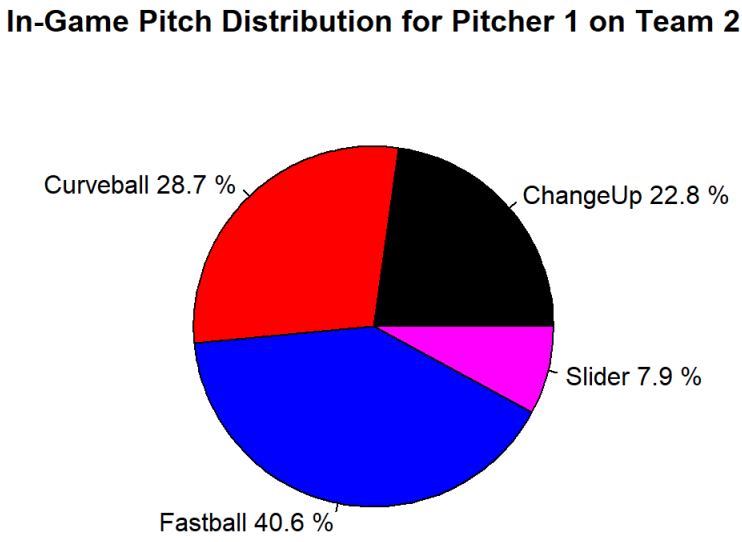
From the second table, it is clear that the bullpen for team 1 struggled mor and had to work longer than the bullpen for team 2 since team 1's starter only lasted 5.1 innings. The bullpen for team 1 threw 76 pitches total while allowing 2 runs and 4 hits. However, they did strikeout 4 batters which has some influence with the large pitch count. For team 2, the starter pitched well and the bullpen got quick outs while only alowing one hit (Solo HomeRun). From these tables, team 2 played a better game compared to team 1 with better offensive and pitching numbers and it shows in the final score.

Team	Score	Hits	AVG	OBP	SLG	OPS	Starting_Pitchers	IP	Runs_Given_Up	K	BB
Team 1	3	9	0.243	0.270	0.378	0.648	Pitcher 1	5.33	3	4	0
Team 2	5	13	0.310	0.357	0.524	0.881	Pitcher 1	6.33	2	9	1

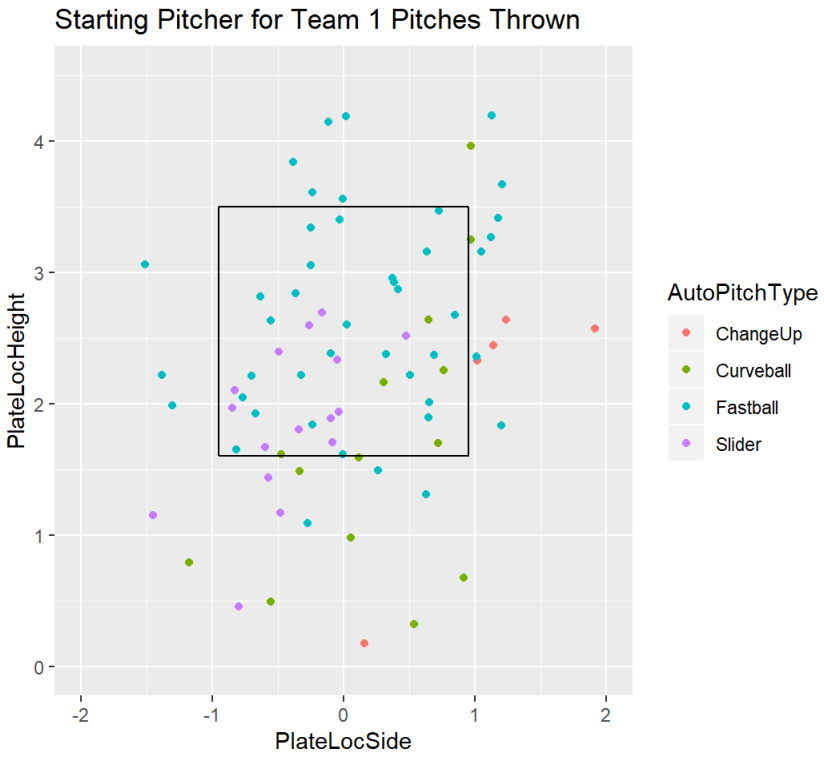
Team	Pitches_Thrown	Runs_Allowed	Hits	K
Team 1	76	2	4	4
Team 2	34	1	1	1

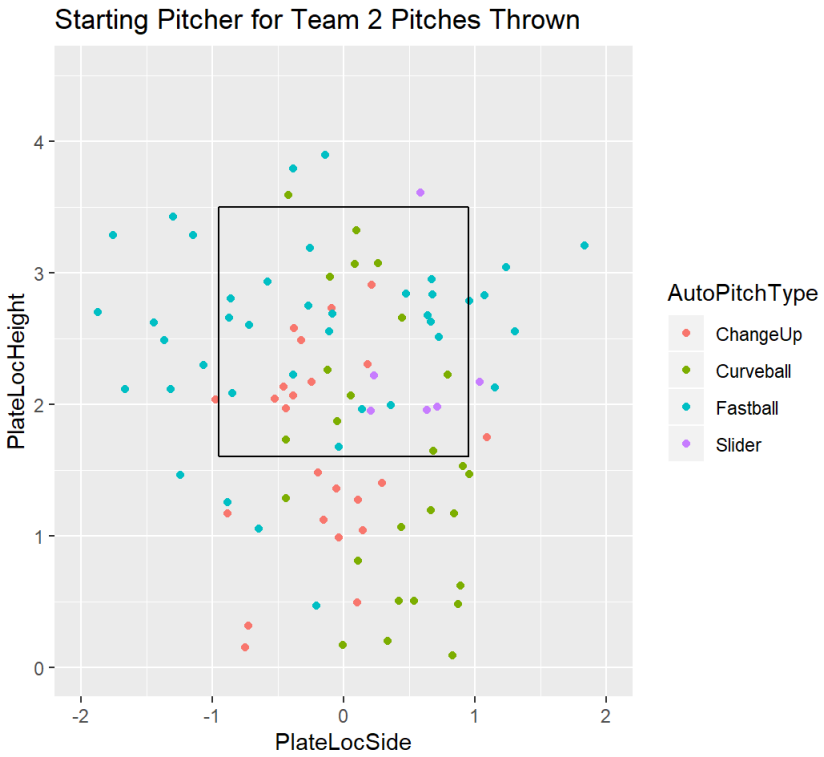
After understanding the overall pitching and offensive number of both teams, let's look deeper into each pitchers pitch distribution. I created a table that displays for each pitcher and each pitch type, the number of pitches, average velocity, average exist speed of the bat, averge distance hit, and average verticle break on pitches. For simplicity purposes, I will focus on the strting pitchers due to more body of work in this game than any bullpen pitcher. Below, are pitch distribution pie charts for both starters. Both starters threw the same pitches (Fastball, ChangeUp, Slider, Curveball) but the starter for team 1 threw fastballs more often the team 2's starter (55.6% vs 40.6%). Also, team 2's starter mixed in ChangeUp's and Curveball's more often than team 1's starter. From the box score, we know that team 2's starter had a better performance than team 1's starter and from the pitch distribution chart, we see that the starter from team 2 had more variety with his pitches than team 1's starter.



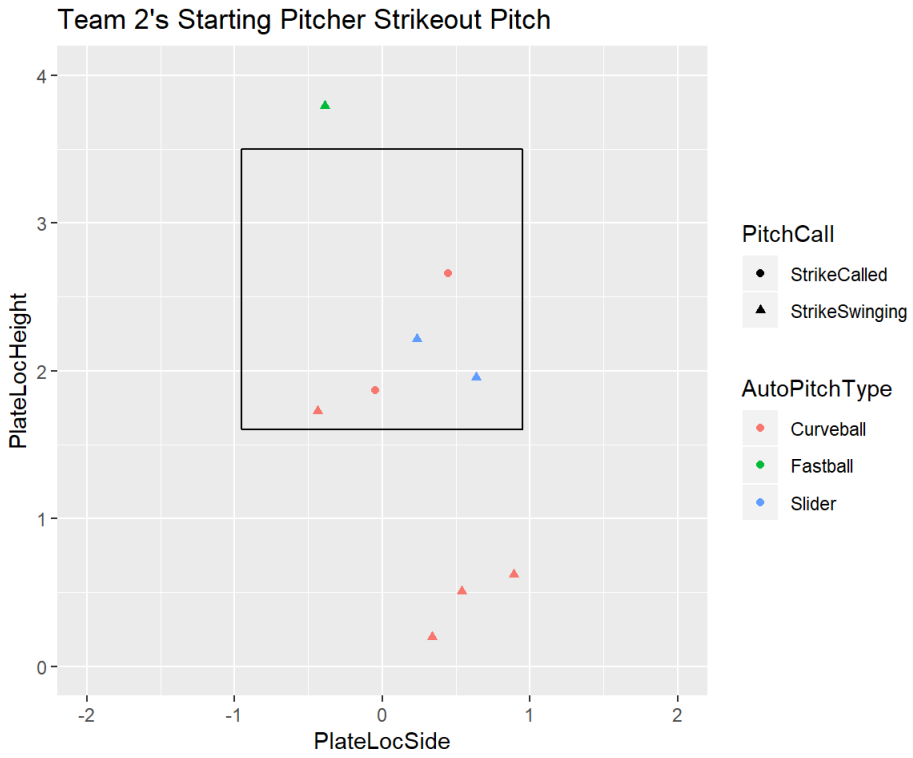


The pitches for both starting pitchers are plotted below. The box represents the strikezone as defined by fan graphs in the following web address (https://www.fangraphs.com/tht/strike-zone-fact-vs-fiction/). From the plotted pitche, we see that Team 1's starter pitched more consistenly inside or near the strikezone than Team 2's starter. However, leaving more percentage of pitches in the strikezone allows for hitters to connect on pitches more often which is what happened. For team 2, their starter threw more off-speed pitches (ChangeUp, Curveball) below the strikezone which can be associated with the 9 strikeouts that team 2's starter threw.

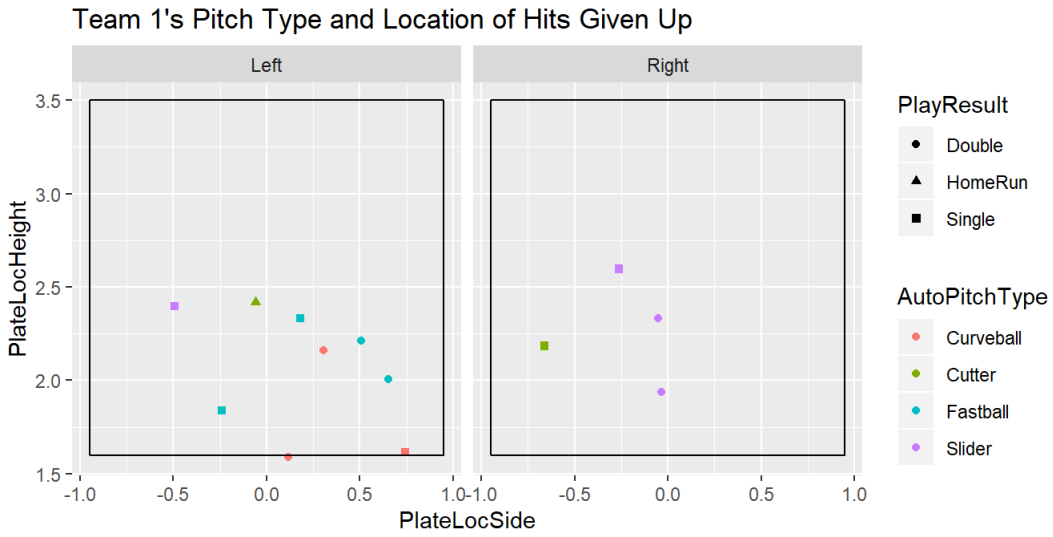


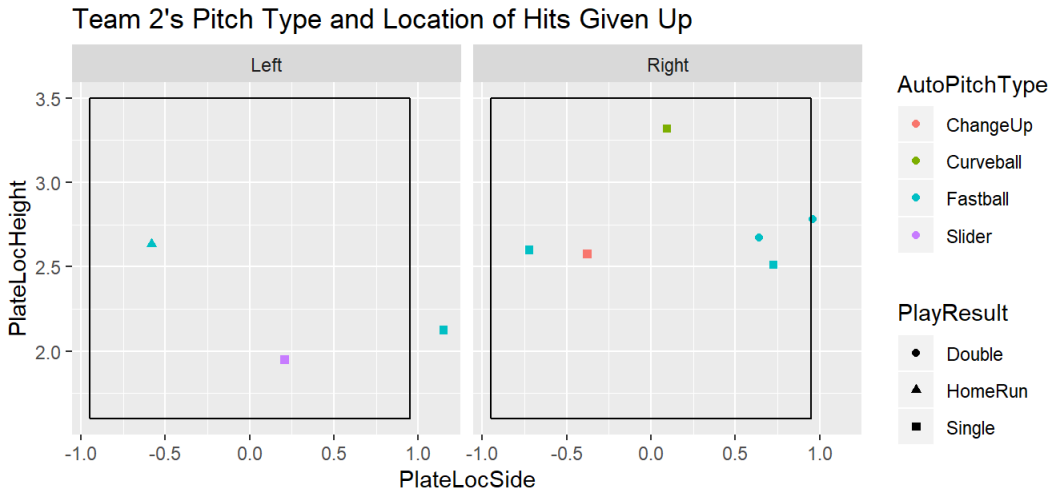


Since team 2's starter had 9 strikeouts, I wanted to look at his closing pitches on those 9 strikeouts. Of his 9 strikeouts, 7 were swinging strikeouts with 4 of them outside the strikezone.



Instead of only looking at pitches that resulted in strikeouts, we can also view the locations of pitches by both teams that resulted in hits. Below, are two plots for each team where the pitches that ended as hits are plotted with its play result (single, double, or homerun) and pitch type. The plots are divided by left handed hitters and right handed hitters. Team 1's pitchers gave up more hits and the pitches were mostly middle and low which is surprising because I would expect more balls up in the zone. Team 2 pitchers perfomed better than those from team 1 but team 2 did leave some pitches up in the zone for mistakes. Both teams had one homerun that were on fastballs in the middle or middle-in of the zone. There are more interesting analysis that can be done with regard to pitchers but I would also like to touch on the hitters in this game.



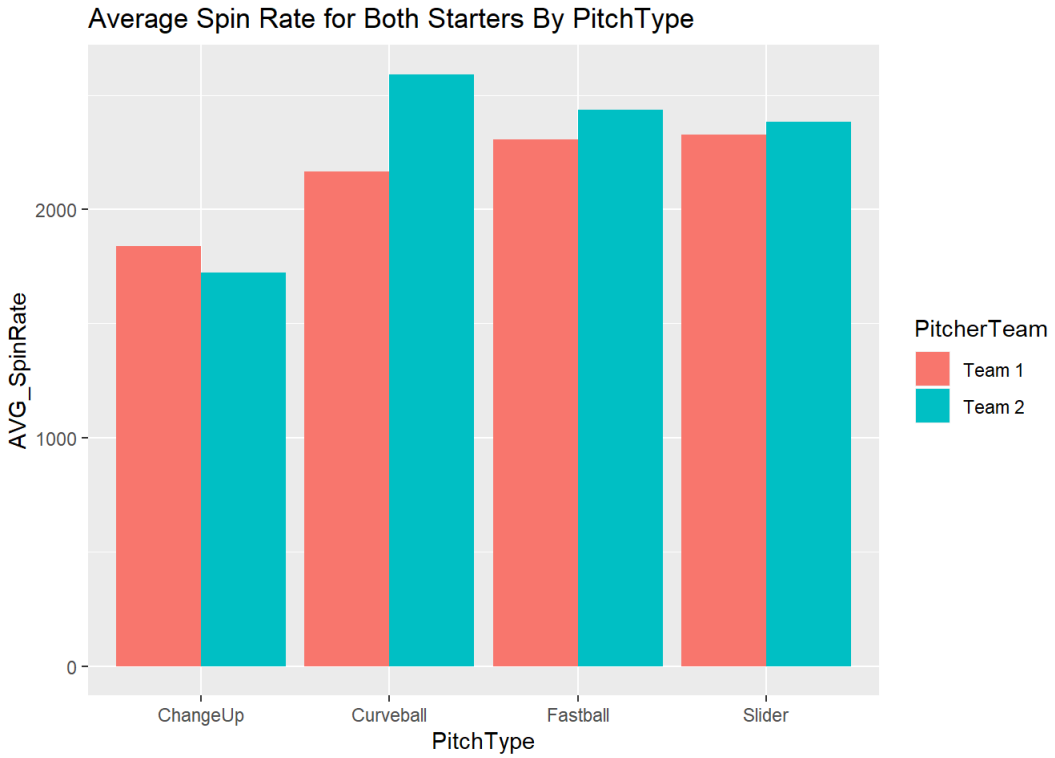


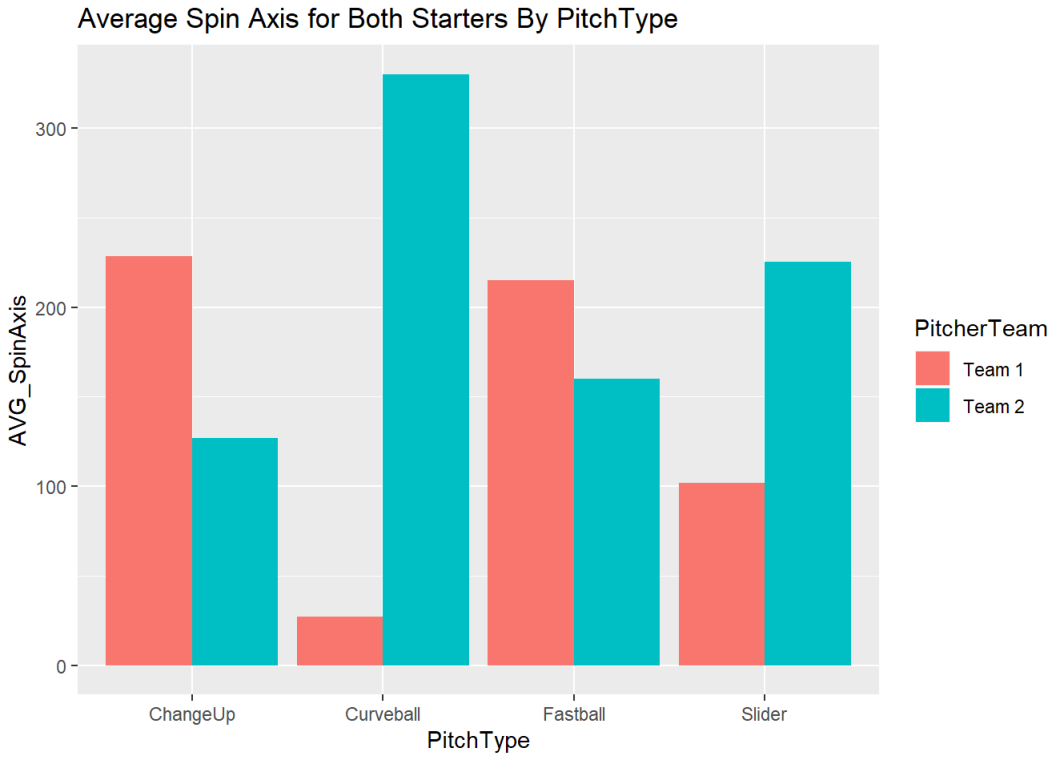
A table is shown below that shows every batter for each team ordered by number of hits. The first few rows represent players that produced offensively in terms of hits. For team 2, batter 3 went 3 for 4 with a single and 2 doubles but no runs batted in. Also, the first 3 hitters for team 2 combined for 7 hits (3 singles, 4 doubles) and 3 rbi's. The front of the lineup produced and set the table for the rest of the lineup. For team 1, there best offensive producers in this game were batters 2 and 7 in the lineup. They had a combined 4 hits and 1 rbi. Both teams had solo homeruns hit by team 1, batter 10 (pinch-hitter most likely) and team 2, batter 9. Also, looking at average exit speed, batter 9 on team 2 had the largest average exit speed and went 2 for 2 with a single and homerun. Interesting that the bottom of the lineup for team 1 produced while the top of the lineup for team 2 produced.

BatterTeam	Batter	PA	Hits	Singles	Doubles	HomeRun	RBI	K	BB	AVG_ExitSpeed	AVG
Team 2	Batter 3	4	3	1	2	0	0	1	0	95.79809	0.750
Team 1	Batter 2	3	2	2	0	0	0	1	0	90.28838	0.667
Team 1	Batter 7	2	2	1	1	0	1	1	0	96.67279	1.000
Team 2	Batter 1	3	2	1	1	0	0	1	0	88.94506	0.667
Team 2	Batter 2	3	2	1	1	0	2	0	0	77.70451	0.667
Team 2	Batter 9	2	2	1	0	1	1	0	0	101.16189	1.000
Team 1	Batter 10	1	1	0	0	1	1	0	0	99.40658	1.000
Team 1	Batter 3	3	1	1	0	0	0	1	1	90.59557	0.333
Team 1	Batter 6	2	1	1	0	0	0	1	0	74.10363	0.500
Team 1	Batter 8	2	1	0	1	0	1	1	0	96.26935	0.500
Team 1	Batter 9	2	1	1	0	0	0	0	0	87.87578	0.500
Team 2	Batter 4	4	1	0	1	0	1	2	0	97.70015	0.250
Team 2	Batter 5	3	1	1	0	0	1	2	0	70.58120	0.333
Team 2	Batter 6	3	1	1	0	0	0	1	2	81.03401	0.333
Team 2	Batter 7	3	1	0	1	0	0	1	1	88.86545	0.333
Team 1	Batter 1	3	0	0	0	0	0	2	0	96.50466	0.000
Team 1	Batter 11	1	0	0	0	0	0	0	0	70.70645	0.000
Team 1	Batter 4	4	0	0	0	0	0	1	1	80.59769	0.000

BatterTeam	Batter	PA	Hits	Singles	Doubles	HomeRun	RBI	K	BB	AVG_ExitSpeed	AVG
Team 1	Batter 5	4	0	0	0	0	0	2	0	74.03274	0.000
Team 2	Batter 10	1	0	0	0	0	0	0	0	84.63922	0.000
Team 2	Batter 8	4	0	0	0	0	0	0	0	73.09405	0.000

Lastly, I chose two statisitics such as average spin rate and average spin axis to show differences between the two starting pitchers on how they release the ball per pitch type. For average spin rate, there doesn't seem to be much difference between the two starting pitchers but for average spin axis, there is a clear difference between the two pitchers on the the ball is rotating on their pitches, especially curveballs.





Further analysis can be made at plotting the pitchers pitches when ahead of the count to see every pitcher's pitch tendency when ahead of the count. Also, further analysis can be done with comparing the starting pitchers on average velocity, average exit speed, average distance hit, and average verticlal, horizontal breaks on pitches to compare there pitching styles and strengths for this game. Lastly, further analysis can be made for hitters with the given information in the data. However, due to time, I had to pick and choose which analysis I would do. an give someone a good snapshot of the baseball game without having to watch it.