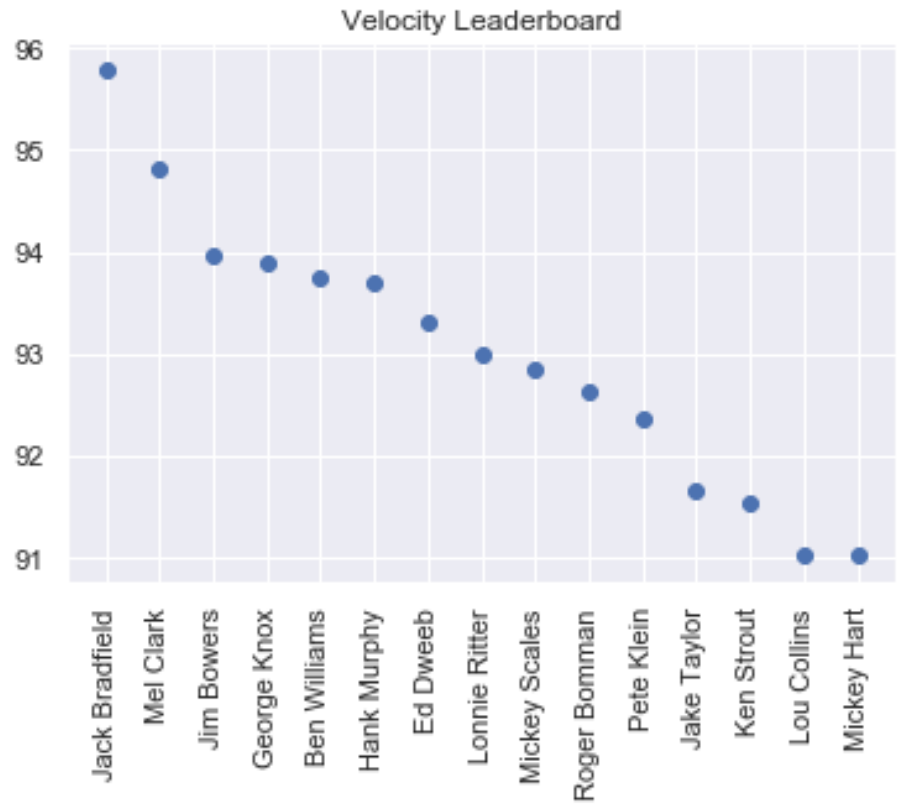
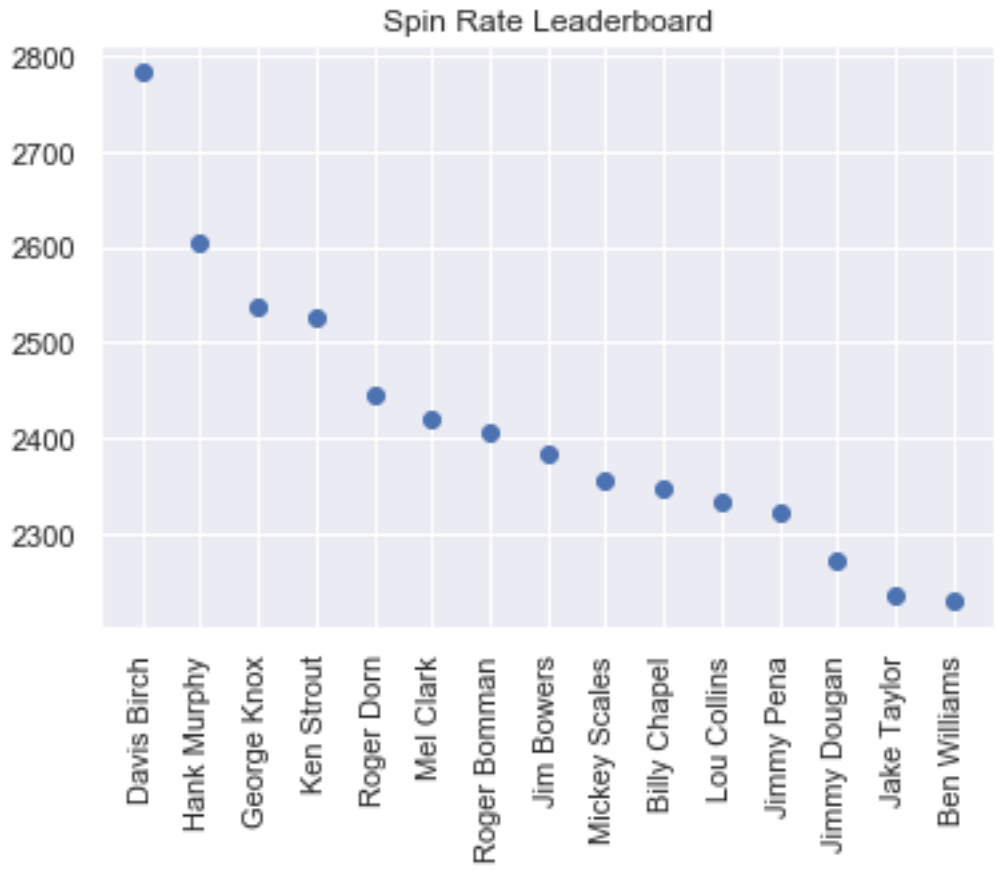
The first step is to clean the data. To do that, I removed any pitches that had missing data. Then, looking at the description of the dataframe, I could see that there weren't any obviously impossible pitches. The mins and maxes are all within possible ranges.

	velo	brk_x	brk_z	spin_rate	extension
count	18383.000000	18383.000000	18383.000000	18383.000000	18383.000000
mean	87.366620	1.229463	5.792375	2240.527961	6.059642
std	5.819996	6.664969	5.399473	301.600108	0.515799
min	67.872000	-25.529000	-13.924000	757.084000	2.787300
25%	82.988000	-4.614000	2.719000	2075.833000	5.714700
50%	88.799000	2.124000	6.891000	2231.496000	6.031400
75%	91.923000	6.784500	9.474500	2428.510000	6.385750
max	99.309000	32.247000	21.141000	3284.972000	9.015100

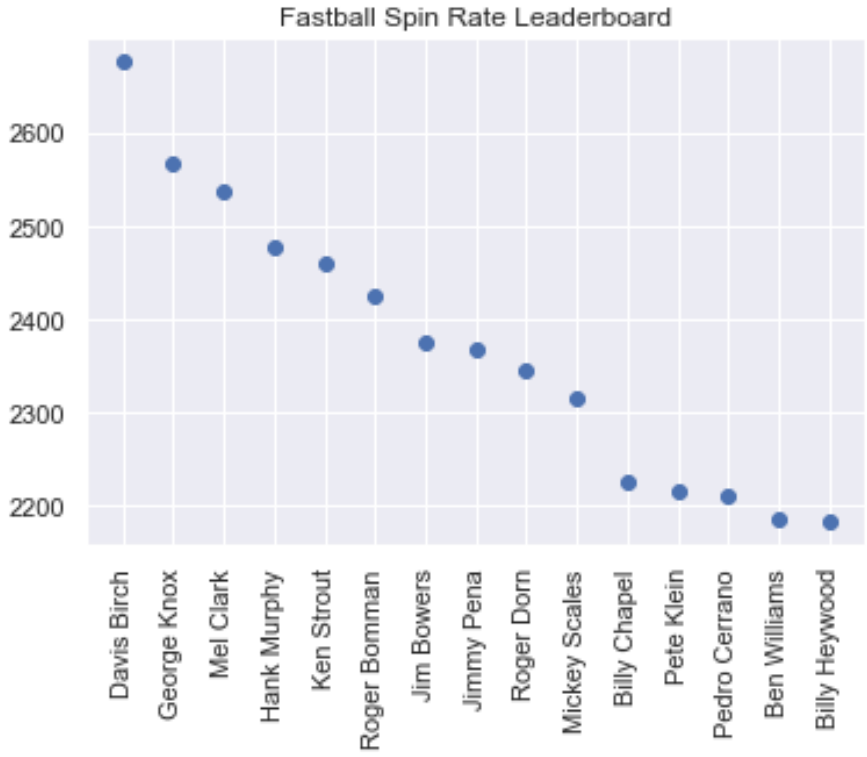
For my velocity leaderboard, I decided to only look at average fastball velocity. While other velocities are important, as well as differential between fastball and offspeed velocities, fastball velocity is the most telling. I used average fastball velocity instead of max because it is more important to tell where a pitcher sits than where he can throw as hard as possible.

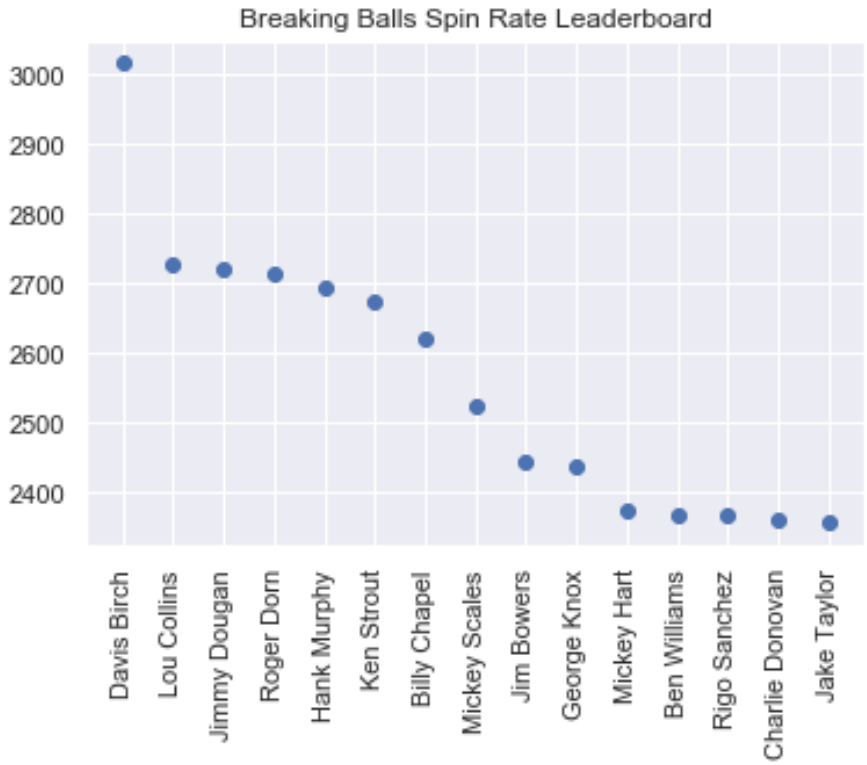


Because different pitches have different ranges of spin rates, this leaderboard is not as straightforward. Changeups are often purposely thrown with less spin, so I removed them from the data set. Fastballs and breaking balls still have different ranges, but since they are generally thrown with purposeful spin I averaged them both.

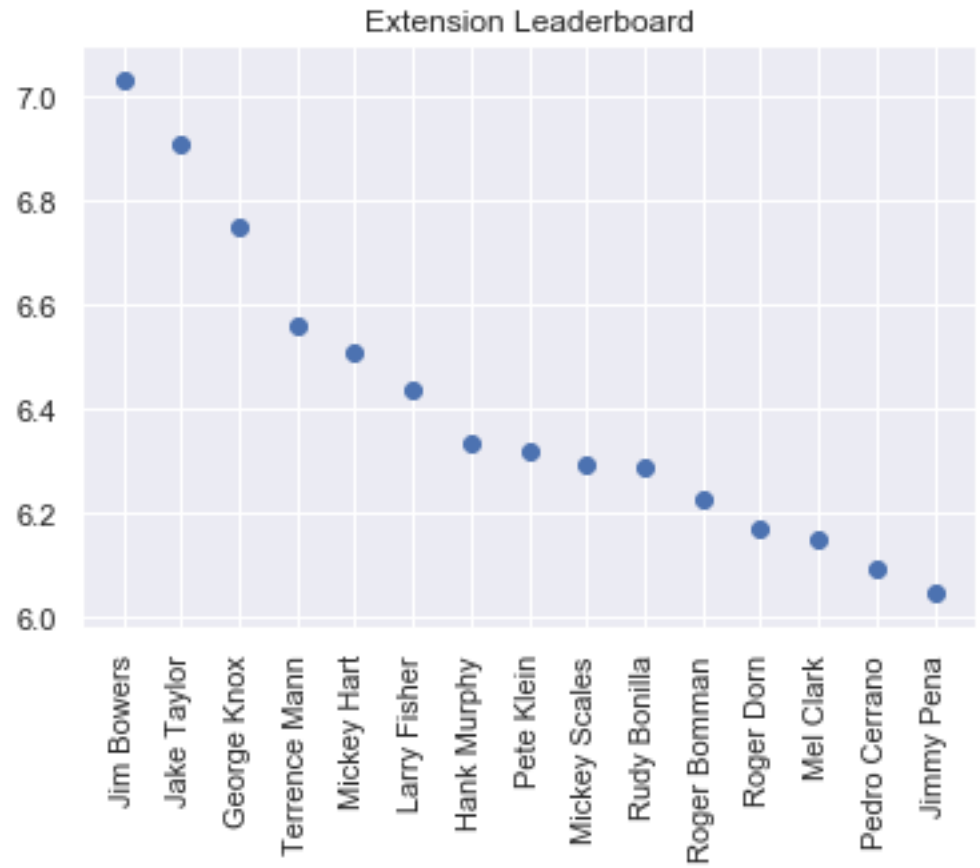


I also created spin rate separate leaderboards for fastballs and breaking balls.



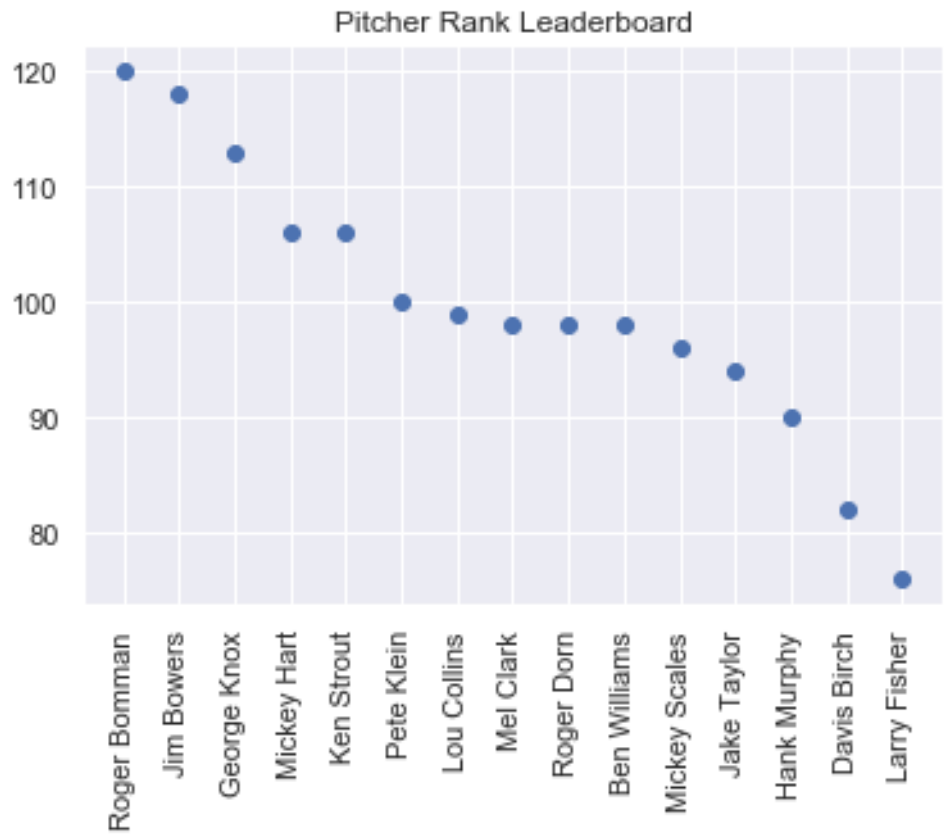


The ideal extension should be consistent across all pitches to keep release point the same. For that reason, I judged extension off of all pitches.



For each variable, I ranked the pitchers on a point system from 1-30. The top of every variable was worth 30 points, next was 29, etc. I added in similar rankings for horizontal break and vertical break of offspeed pitches using their absolute values to account for direction. I only used offspeed pitches because break is more important for them than for fastballs. Using this point system for all five variables, I ended up with these final rankings.

Name	extension	ext_rank	velo	velo_rank	spin_rate	spin_rank	final_rank	abs_brk_x	brk_x_rank	abs_brk_z	brk_z_rank	new_final_rank
Roger Bomman	6.224722	20.0	92.617665	21.0	2401.370182	26.0	67.0	8.092867	29.0	5.941812	24.0	120.0
Jim Bowers	7.029275	30.0	93.953211	28.0	2385.027560	25.0	83.0	3.374655	7.0	8.557298	28.0	118.0
George Knox	6.750164	28.0	93.894410	27.0	2482.358803	27.0	82.0	6.970616	25.0	2.498839	6.0	113.0
Mickey Hart	6.508680	26.0	91.006673	16.0	2169.144541	16.0	58.0	6.181909	22.0	6.622182	26.0	106.0
Ken Strout	5.929800	13.0	91.519024	18.0	2493.658700	28.0	59.0	7.061799	27.0	4.369372	20.0	106.0
Pete Klein	6.319681	23.0	92.345581	20.0	2141.269149	12.0	55.0	5.581963	18.0	7.859259	27.0	100.0
Lou Collins	5.810429	8.0	91.013019	17.0	2325.191554	22.0	47.0	8.537877	30.0	5.132229	22.0	99.0
Mel Clark	6.151364	18.0	94.803957	29.0	2289.718779	20.0	67.0	4.890037	16.0	3.509285	15.0	98.0
Roger Dorn	6.173496	19.0	88.726590	6.0	2365.436237	24.0	49.0	6.987727	26.0	5.866968	23.0	98.0
Ben Williams	5.934379	15.0	93.737541	26.0	2142.238885	13.0	54.0	6.259428	23.0	4.731335	21.0	98.0
Mickey Scales	6.293415	22.0	92.845043	22.0	2330.276718	23.0	67.0	4.257642	13.0	3.580014	16.0	96.0
Jake Taylor	6.906651	29.0	91.662316	19.0	2209.462741	18.0	66.0	4.524617	14.0	3.387042	14.0	94.0
Hank Murphy	6.333295	24.0	93.685198	25.0	2542.473506	29.0	78.0	2.310521	4.0	2.505776	8.0	90.0
Davis Birch	5.559418	2.0	86.078755	1.0	2774.550073	30.0	33.0	5.866864	20.0	9.061641	29.0	82.0
Larry Fisher	6.439176	25.0	90.807464	15.0	1936.101671	2.0	42.0	4.591286	15.0	4.271203	19.0	76.0



With more time and data there is much more I would like to explore. I would clean the data more carefully, looking for outliers or pitches with data combinations that don't make sense instead of general mins and maxes. I would also like to create a model for pitching success with more accurate weights based on Statcast data of MLB pitchers. This would be more likely to accurately predict success.