

In order to create useful leaderboards, it was necessary to filter the data and examine it for irregularities or errors. I first searched for potential duplicate data points and eliminated them. The chances of two or more pitches being identical across every provided measurement are astronomically low, so I will assume that any identical pitches have accidentally been included twice in the data set. I then searched for any missing values. I found that there were a few missing statistics, such as numerous missing extension values for Hank Murphy, but not enough to seriously affect any projections. I next divided the data into three separate data sets, one for each pitch type that has been measured. This is a necessary step when determining the three measures that leaderboards will be provided for. If the data is not split, measurements such as velocity will not be very useful, as a pitcher's velocity should greatly differ based on the pitch that he is throwing. I was then able to compute the average velocities, spin rates, and extensions for each pitcher across the three newly created data sets. After attaining these averages, I was able to narrow down the field to the top 15 in each category and displayed them in a bar chart.

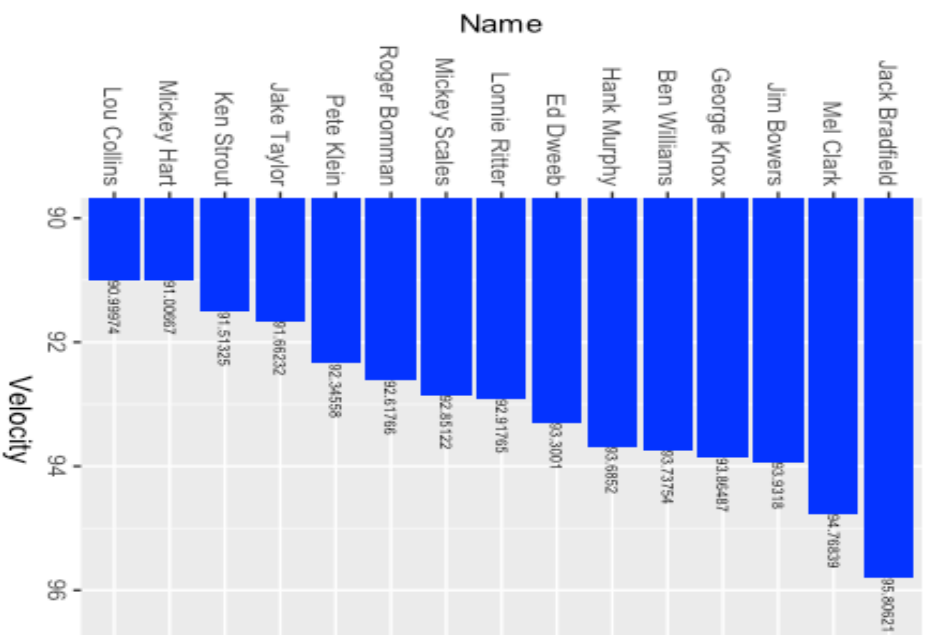
The velocity leaderboard showed that there were a few pitchers that stood above the rest across all three pitches. Notably Jack Bradfield was in the top 3 for all three pitches, leading in both fastball and breaking ball velocity. George Knox also was near the top of all three charts, leading in changeup velocity. With these leaderboards, one can also tell that the average velocity for a fastball is substantially higher than that for breaking balls or changeups. The spin rate leaderboard shows a much different group of leaders. While Jack Bradfield dominated the top of the velocity leaderboard, he does not have a notable presence on the spin rate leaderboards, only appearing towards the bottom of the changeup spin rate board. The most impressive pitcher regarding spin rate is Davis Birch, as he easily bests the field in all three pitches. Besides this, the top of the fastball spin rate leaderboard is very similar to the top of the fastball velocity leaderboard, which makes sense since there is generally a positive linear relationship between the two variables. However, the leaderboards for breaking balls and changeups are quite different, which may be indicative of the many different kinds of breaking balls and changeups that a pitcher can throw. For example, spin rates of sliders and 12-6 curveballs are likely very different, while some pitchers may intentionally slow the spin rate of their changeup to create a greater disparity between that and their fastball. The extension leader board is most similar to the velocity leaderboard, although Jack Bradfield is again no longer among the leaders. These leaderboards are instead dominated by Jim Bowers and Jake Taylor, who have very similar measures and hold the top two spots for fastballs and breaking balls. With changeups, Taylor is virtually tied with Pete Klein and Mickey Scales, although Bowers is not on the list, as he does not throw a changeup.

I then used these leaderboards, as well as other manipulations of the given data, to create a ranking of the top 15 pitchers. I decided to focus on two major aspects of pitching (velocity and movement) and chose statistics that show different sides of these two aspects. I would've also liked to incorporate location in these rankings, but no data was provided that would allow me to do this. When determining which aspects of velocity to include, I decided to not focus on just pure pitch velocity but rather all aspects of a pitch that could impact a batter's timing. I felt it was necessary to start with fastball velocity, as this is the most raw and basic measure of how hard a pitcher throws. Rather than including changeup velocity, I chose to include the difference in speed between a pitcher's changeup and fastball. If a pitcher has a

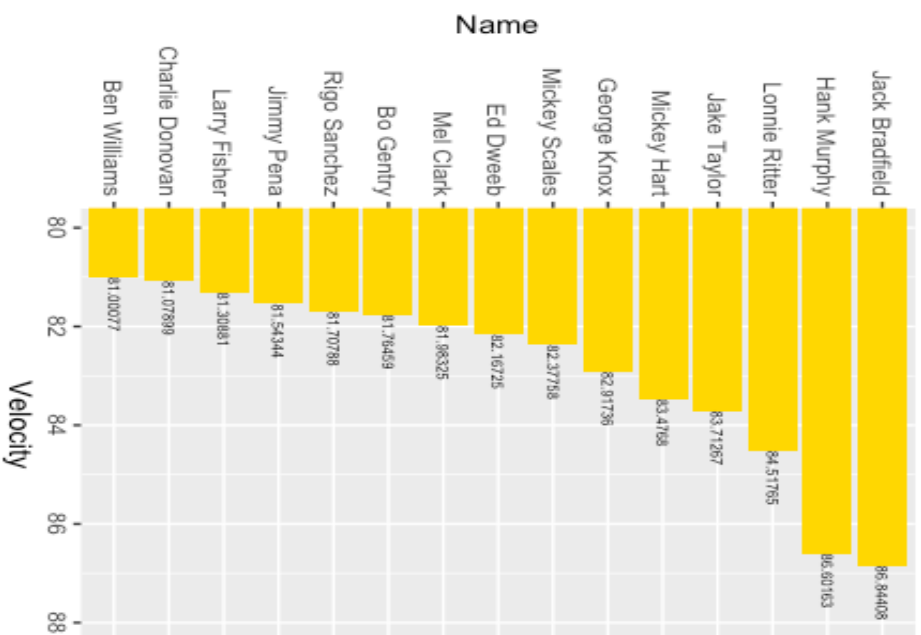
greater difference in velocity, he is more likely to induce swings and misses. Although pitchers can be effective with changeups that have similar velocities to their fastballs, these pitchers tend to induce more ground balls with their changeup, rather than swings and misses. A swing and miss is more valuable than a ground ball, so I rated pitchers with greater differences between the two pitches as the most valuable. I included extension as the third measure of velocity. Pitchers with higher extension measures release the ball closer to home plate, which reduces the time that a batter has to pick up the ball and increases the perceived velocity. Therefore, if two pitchers throw a fastball at the same velocity, the one that has the better extension will appear to be faster. I only included the extension statistics for fastballs, as more extension on fastballs would have the greatest effect. To determine movement, I normalized the three spin rates that were previously shown on the leaderboards. Since spin rate and movement are directly correlated, if two pitches have the same velocity the one with the higher spin rate is expected to have more break. In order to determine the true average break for each pitcher, it was necessary to eliminate the bias that increased velocity may cause. To do this I calculated the average Bauer unit for each pitcher, which is a normalized spin rate that is found by dividing spin rate by velocity. By using this measure, I did not need to include the measures for horizontal and vertical break that were provided, as they would be redundant. The last obstacle was determining how to weigh velocity vs movement, but in the end I determined that both were equally valuable when scouting pitchers at the college level.

In order to create a total value score for each pitcher, I ranked each pitcher across all six measures. If a pitcher led in a measure he got 30 points, while a pitcher that was last received 1 point. At this point, I also had to manually enter values for Jim Bowers, who did not have a changeup and therefore was not on any leaderboards that included changeup measures. Since he does not throw the pitch he received a 0 for any measures that involved changeups. I then added the six individual scores to create one overall score, which I then used to rank the top 15 pitchers. With this scoring system, the maximum value would be 180 points, which would indicate that a pitcher would top the leaderboard in all six categories. In the end, Hank Murphy stood above the rest with a score of 137. When comparing my leaderboard to the individual leaderboards that were constructed earlier, it is apparent that my system favors players that are well rounded over those that dominate in specific categories. For example, Hank Murphy, the overall leader, appears on nearly every leaderboard but is only in the top 3 in one category, breaking ball velocity. On the other hand, Jack Bradfield, who leads in fastball and breaking ball velocity, is not even in the overall top 15 and Davis Birch, who leads in each spin rate measure, barely cracks the list. Overall, the top 15 list that was created does an excellent job of showing the best, most well-rounded pitchers, but it could be improved if location measures were included.

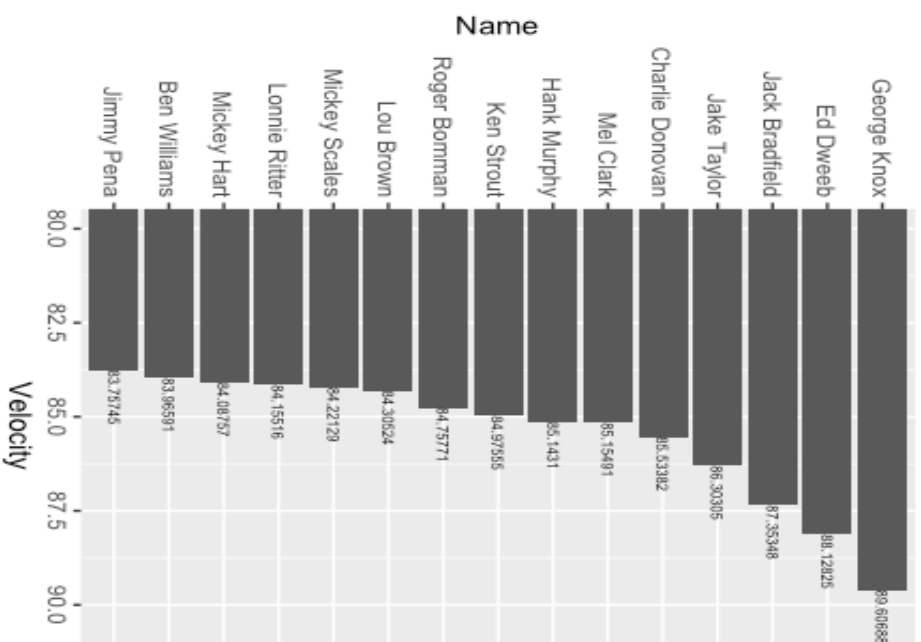
Fastball



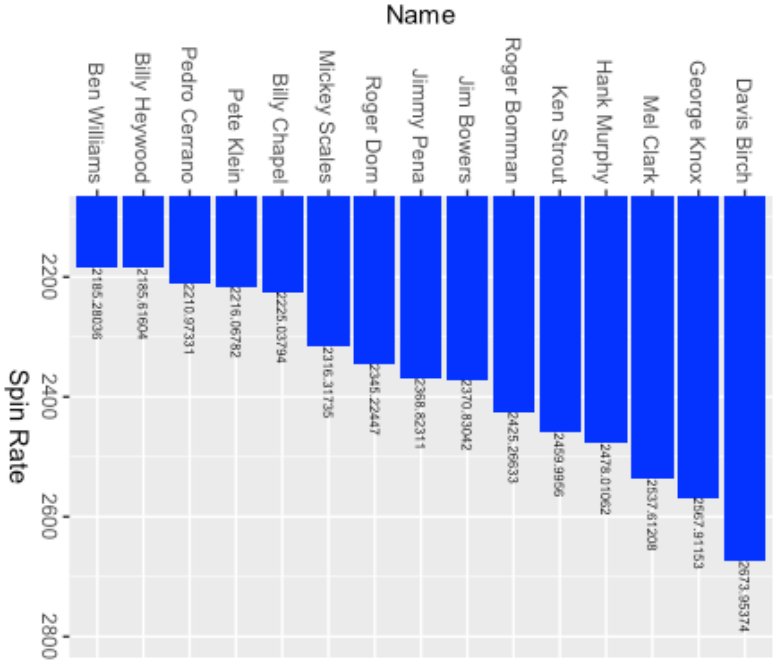
Velocity Leaderboard Breaking Ball



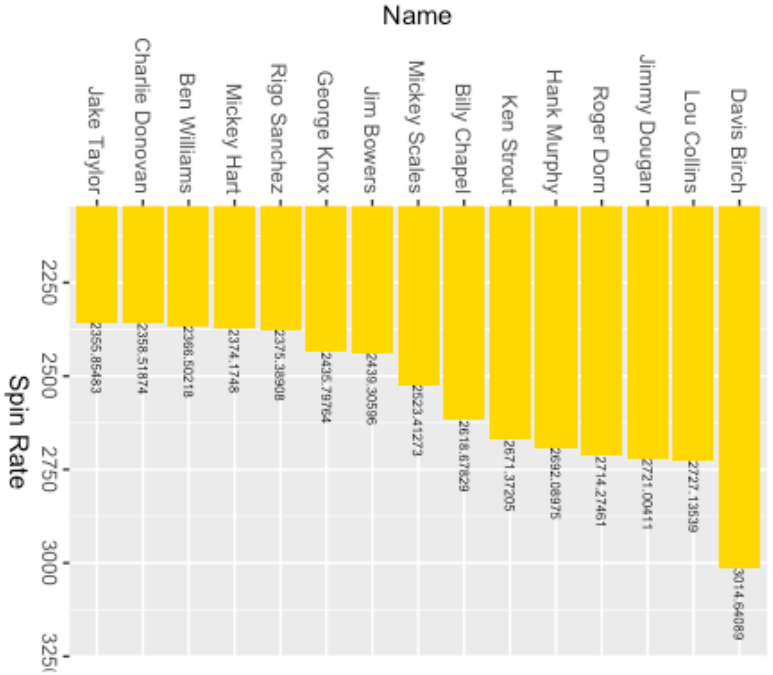
Changeup



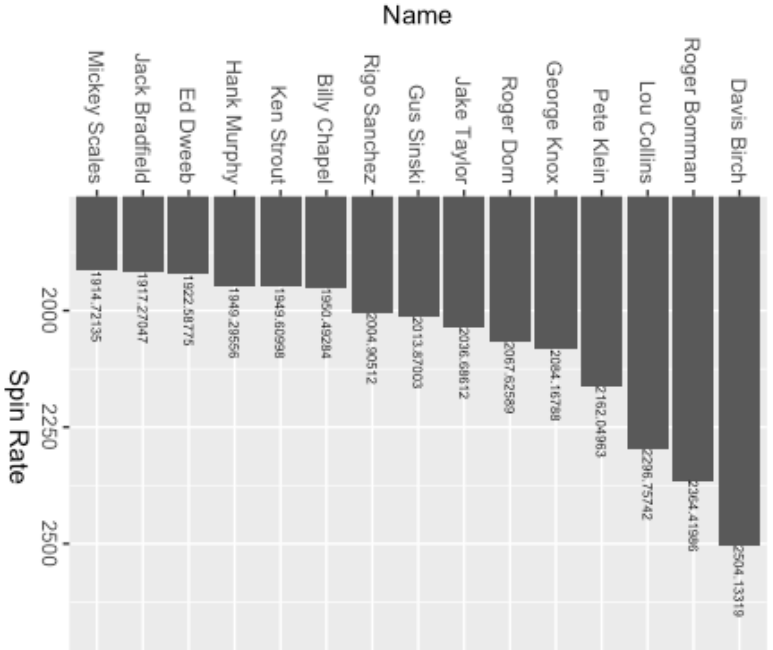
Fastball



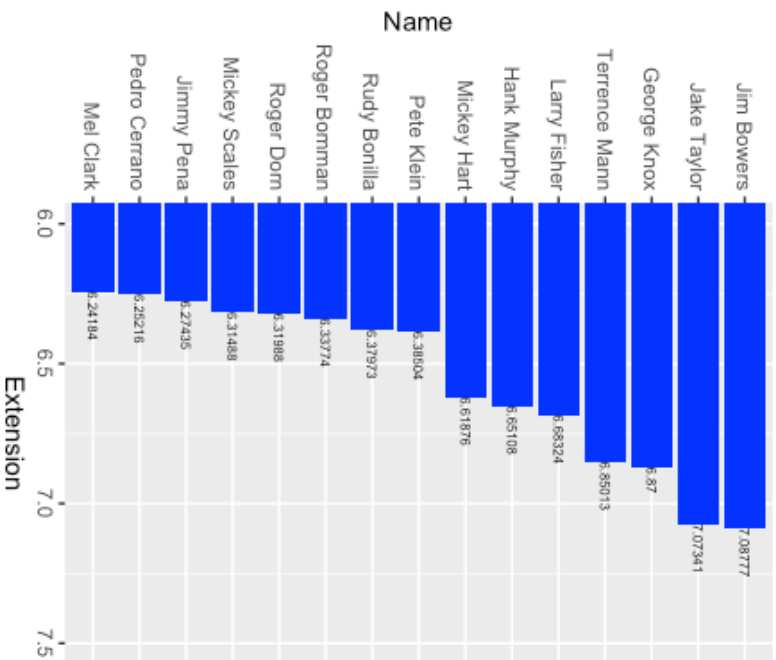
Spin Rate Leaderboard
Breaking Ball



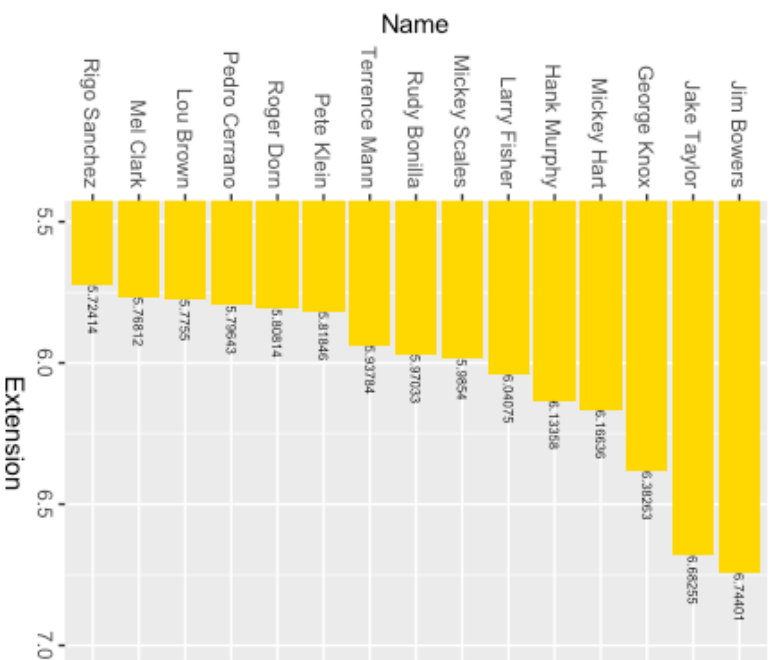
Changeup



Fastball



Extension Leaderboard Breaking Ball



Changeup

