What is Wrong with Devin Williams?

Chris Callender

May 23, 2025

1 Abstract

Devin Williams is having a historically bad start to the 2025 season. At the time of writing this, he just blew a 2-1 lead in the 9th inning, and has been removed from the closer role. This is very surprising given that he has been one of, if not the best, relievers for the past several years. Let's examine Williams' metrics and pitch data to determine a possible cause.

Note: The data was gathered on April 27, 2025.

2 Introduction

We will start by gathering Devin Williams' career data, such as ERA, FIP, WAR, and K%. We will observe his stats throughout his career by looking at time series graphs. We will see which statistics are correlated with each other and if they are statistically significant. We will also determine which metrics are most valuable for pitchers and see what variables affect them.

3 Methods

Here are some statistics and metrics that you may not be familiar with:

3.1 FIP

FIP stands for fielding independent pitching.

$$\frac{13*HR+3*(BB+HBP)-2*K}{IP} + \text{FIP Constant}$$

The FIP constant puts FIP onto the same scale as the entire league's ERA.

FIP is similar to ERA, but only focuses on strikeouts, walks, hit by pitches, and home runs, since these are what a pitcher has the most control over. It removes balls in play, except home runs, since the results are determined by the fielders. For example, a ground ball can be a hit if hit to a bad fielder, but it may be an out if hit to a good fielder. The pitcher can't control the fielders on his team, so FIP removes this from consideration.

3.2 wOBA

wOBA stands for weighted on base average.

$$wOBA = \frac{.69\times uBB+.72\times HBP+.89\times 1B+1.27\times 2B+1.62\times 3B+2.10\times HR}{AB+BB-IBB+SF+HBP}$$

Note: The factors that each outcome is multiplied by, i.e. .69, .72, .89, ..., vary by year, but they are always relatively similar.

wOBA measures hitting ability by assigning a weight to each possible plate appearance outcome. For example, a home run has more weight than a single. We can think of wOBA as a better version of OPS, since it combines aspects of OBP and SLG. That is, it considers walks like OBP, but it also weights hits accordingly, similar to SLG. However, it is better than OPS since it weights each outcome in one metric, rather than just adding two metrics together.

3.3 xwOBA

xwOBA stands for expected wOBA. It is calculated using exit velocity, launch angle, and sprint speed. Every batted ball is given single, double, triple, and home run probability. It is calculated the same way as wOBA, i.e., it uses the same formula. Outcomes are based on contact, not ballpark, weather, or defence.

3.4 xERA

xERA stands for expected ERA. It is xwOBA converted to the ERA scale. This is a great metric, since it considers every possible way a pitcher can allow a base runner, unlike FIP, but is based on statcast data.

3.5 xFIP

xFIP stands for expected FIP. It is similar to FIP, but uses projected home run rate based on league average home runs per fly ball, instead of actual home runs allowed.

$$\text{Home Run Rate} = \frac{\text{Fly Balls}}{\text{League Average Rate of HRs per Fly Ball}}$$

Pitchers have more control over how many fly balls they allow than how many of those fly balls are home runs. That is why xFIP uses projected home run rate. xFIP uses the same formula as FIP, but substitutes home runs with how many home runs he should have allowed given the number of fly balls he induced.

3.6 Summary

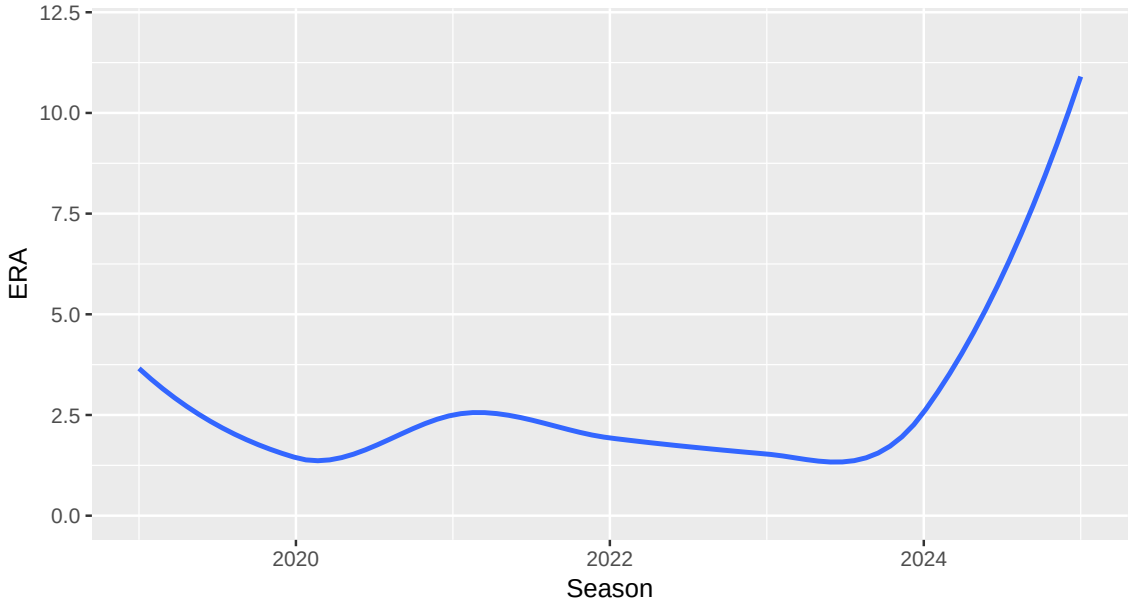
I think that ERA, xERA, FIP, and xFIP are the most valuable metrics to assess pitcher performance, since they are directly impacted by runs allowed, and the goal of a pitcher is to prevent the other team from scoring runs. ERA is the average number of earned runs a pitcher allows in 9 innings, and xERA determines how many runs the pitcher is expected to allow based on factors such as hit probability and catch probability. FIP is similar to ERA, but it only accounts for things a pitcher can control, such as walks, strikeouts, and home runs. It is not affected by fielding, since the pitcher can not control whether he has good fielders on his team. xFIP adjusts FIP based on expected home run values based on league average home run per fly ball rate.

We will determine which metrics are statistically significant and correlated with ERA, xERA, FIP, and xFIP.

4 Results

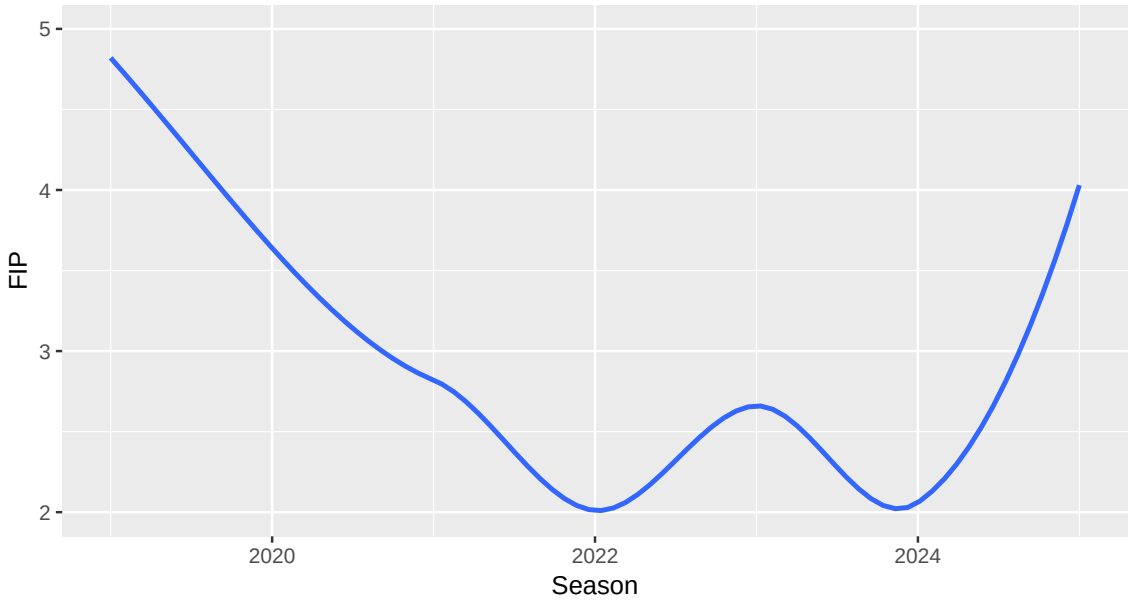
4.1 ERA, FIP, xERA, and xFIP

4.1.1 ERA



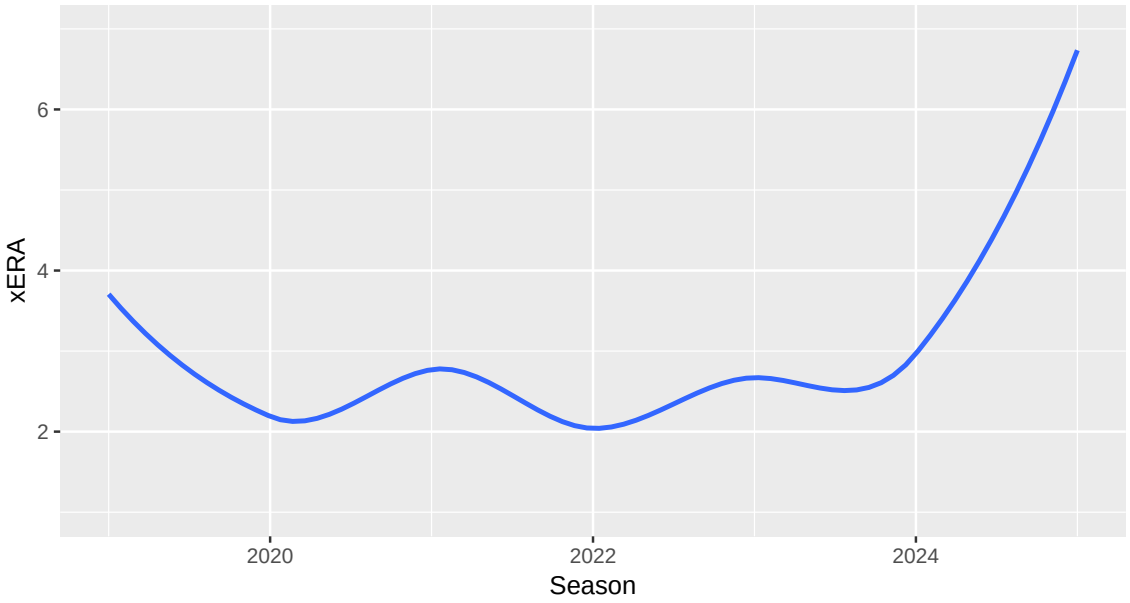
We can see that Williams' ERA has always been very low, especially the past few seasons. However, it has greatly increased this season. That means he is giving up a higher rate of earned runs per nine innings.

4.1.2 FIP



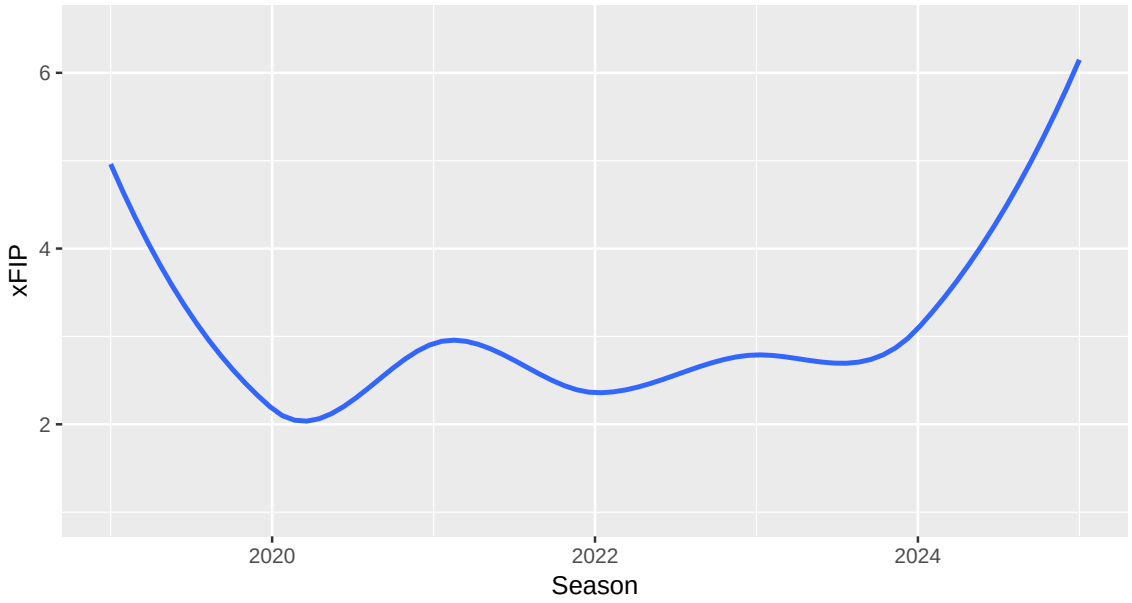
This graph is interesting because Williams' FIP was worse during his rookie season than it is this season. However, he only pitched 13.2 innings in 2019. But since then his FIP has been under 3 every season.

4.1.3 xERA



Williams' xERA this season is by far the worst of his career. This is concerning, since it is based on statcast data and probability, implying that he is giving up a lot of high hit probabilities. However, it is interesting that his xERA of 6.96 is much lower than his ERA of 11.25, implying that he is getting unlucky this season.

4.1.4 xFIP



Williams' xFIP this season is the highest of his career, but it also was high his rookie season.

4.2 Significance and correlation

An event is statistically significant if its result is unlikely to be due to chance.

We will conduct t-tests to determine significance with $\alpha = .05$.
H_0 : True correlation $= 0$.
H_A : True correlation $\neq 0$.
If $p \leq .05$, then we reject the null hypothesis, meaning the true correlation is not zero. Otherwise, we fail to reject the null hypothesis, meaning that we can not say for sure that the true correlation is not equal to zero.

To determine correlation, we will calculate r, the coefficient of correlation. Two events are correlated if they occur simultaneously. That is, as one increases or decreases, the other increases or decreases. Two events are positively correlated if as one increases the other also increases, or if one decreases and the other also decreases. Two events are negatively correlated if as one increases the other decreases.

4.2.1 Exit Velocity

ERA: $p = .00292$, $r = .9242$. This means that exit velocity has a statistically significant impact on ERA, and the two also have a high positive correlation. That is, the relationship between exit velocity and ERA is unlikely to be due to chance. So, when exit velocity increases, we can expect to see an increase in ERA.

FIP: $p = .02832$, $r = .8068$. This means that exit velocity has a statistically significant impact on FIP, and the two also have a high positive correlation.

xERA: $p = .0003338$, $r = .9685$. This means that exit velocity has a statistically significant impact on xERA, and the two also have a high positive correlation.

xFIP: $p = .002574$, $r = .9280$. This means that exit velocity has a statistically significant impact on xFIP, and the two also have a high positive correlation.

4.2.2 Ground Ball Rate

ERA: $p = .05667$, $r = -.7409$. This means that ground ball rate does not have a statistically significant impact on ERA. The two have a relatively high negative correlation. That is, we can not say for certain that ground ball rate and ERA have a relationship that is not due to chance. However, it seems that as ground ball rate increases, ERA decreases.

FIP: $p = .01758$, $r = -.8416$. This means that ground ball rate has a statistically significant impact on FIP, and the two also have a high negative correlation.

xERA: $p = .03055$, $r = -.8005$. This means that ground ball rate has a statistically significant impact on xERA, and the two also have a high negative correlation.

xFIP: $p = .009953$, $r = -.8748$. This means that ground ball rate has a statistically significant impact on xFIP, and the two also have a high negative correlation.

4.2.3 Strikeout Rate

ERA: $p = .02097$, $r = -.8295$. This means that strikeout rate has a statistically significant impact on ERA, and the two also have a high negative correlation.

FIP: $p = .0005818$, $r = -.9605$. This means that strikeout rate has a statistically significant impact on FIP, and the two also have a very high negative correlation.

xERA: $p = .007262$, $r = -.8900$. This means that strikeout rate has a statistically significant impact on xERA, and the two also have a high negative correlation.

xFIP: $p = .00004353$, $r = -.9861$. This means that strikeout rate has a statistically significant impact on xFIP, and the two also have a very high negative correlation.

4.2.4 Chase Rate

ERA: $p = .05499$, $r = -.7444$. This means that chase rate does not have a statistically significant impact on ERA. The two also have a relatively high negative correlation.

FIP: $p = .009806$, $r = -.8755$. This means that chase rate has a statistically significant impact on FIP, and the two also have a high negative correlation.

xERA: $p = .02818$, $r = -.8072$. This means that chase rate has a statistically significant impact on xERA, and the two also have a high negative correlation.

xFIP: $p = .004028$, $r = -.9135$. This means that chase rate has a statistically significant impact on xFIP, and the two also have a high negative correlation.

4.2.5 Contact Rate

ERA: $p = .02827$, $r = .8069$. This means that contact rate has a statistically significant impact on ERA, and the two also have a high positive correlation.

FIP: $p = .001977$, $r = .9353$. This means that contact rate has a statistically significant impact on FIP, and the two also have a high positive correlation.

xERA: $p = .01237$, $r = .8630$. This means that contact rate has a statistically significant impact on xERA, and the two also have a high positive correlation.

xFIP: $p = .0002081$, $r = .9739$. This means that contact rate has a statistically significant impact on xFIP, and the two also have a very high positive correlation.

4.2.6 Swing and Miss Rate

ERA: $p = .01414$, $r = -.8553$. This means that swing and miss rate has a statistically significant impact on ERA, and the two also have a high negative correlation.

FIP: $p = .003049$, $r = -.9228$. This means that swing and miss rate has a statistically significant impact on FIP, and the two also have a high negative correlation.

xERA: $p = .004857$, $r = -.9067$. This means that swing and miss rate has a statistically significant impact on xERA, and the two also have a high negative correlation.

xFIP: $p = .0000256$, $r = -.9888$. This means that swing and miss rate has a statistically significant impact on xFIP, and the two also have a very high negative correlation.

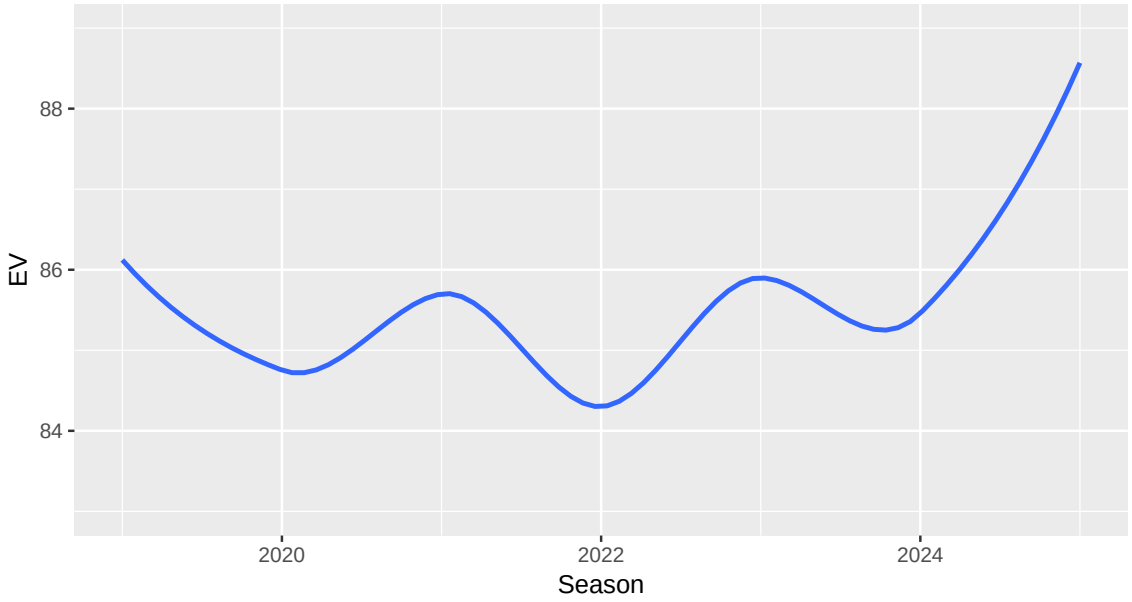
4.2.7 Summary

This shows that getting strikeouts and limiting contact are the keys to success. If a batter does make contact, the pitcher wants it to be a weak ground ball. Chasing pitches out of the zone leads to more strikeouts ($p = .008155$, $r = .8846$) and weak hit balls ($p = .04944$, $r = -.7557$). More swing and misses lead to more strikeouts ($p = .000105$, $r = .9802$). We will call these six metrics the key metrics for a pitcher's success.

4.3 Key Metrics Analysis

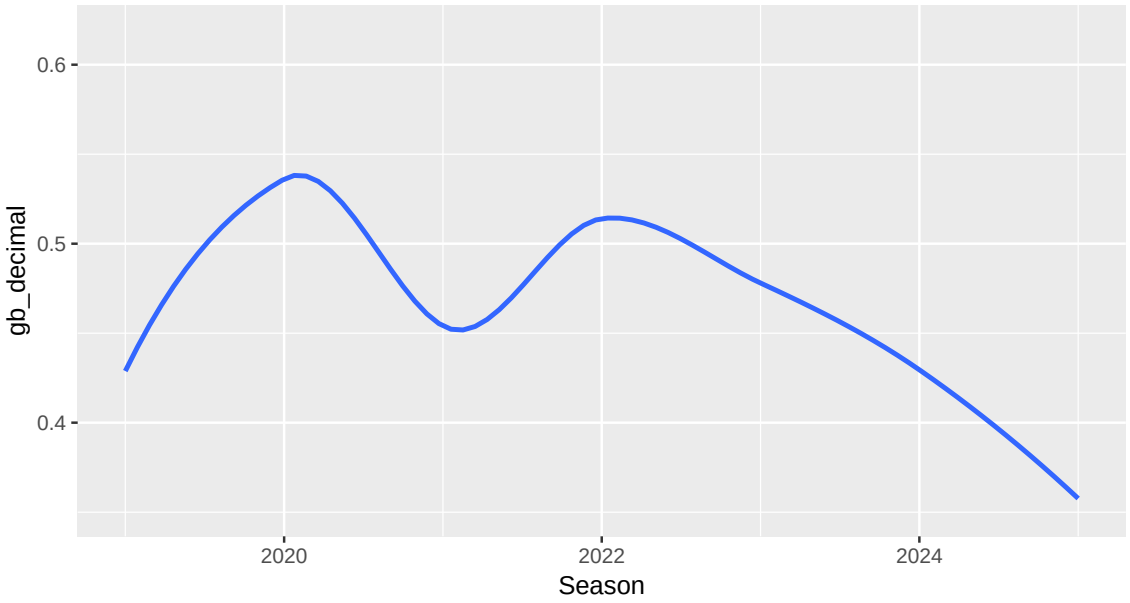
Let's analyze Williams' data for these key metrics.

4.3.1 Exit Velocity



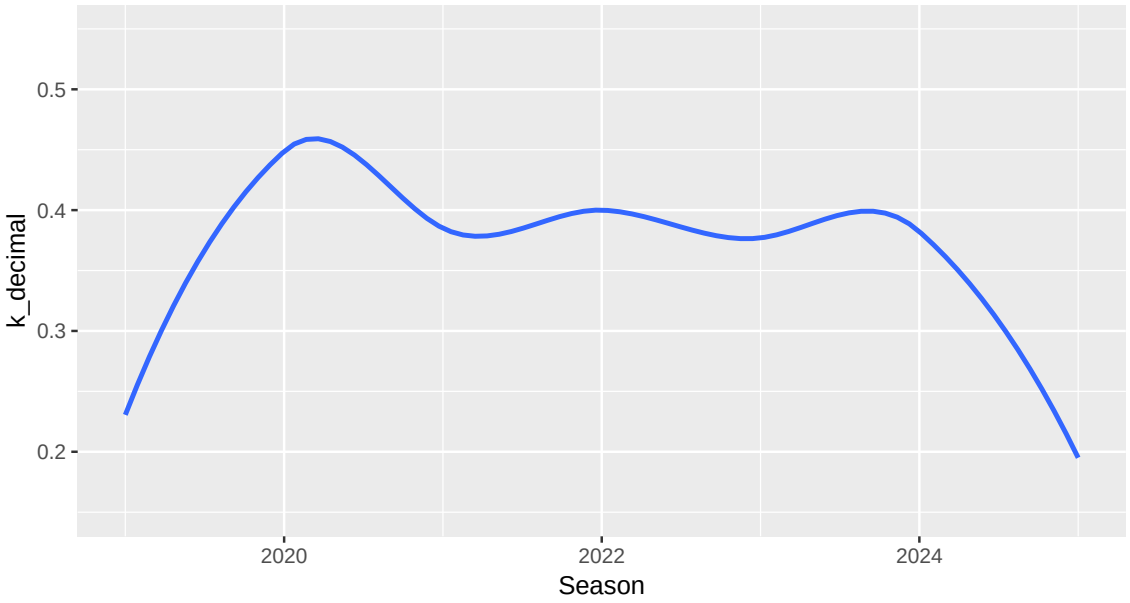
Williams' exit velocity this season is the highest of his career, but it tends to fluctuate by season. That means batters are hitting the ball harder and making better contact against Williams this season.

4.3.2 Ground Ball Rate



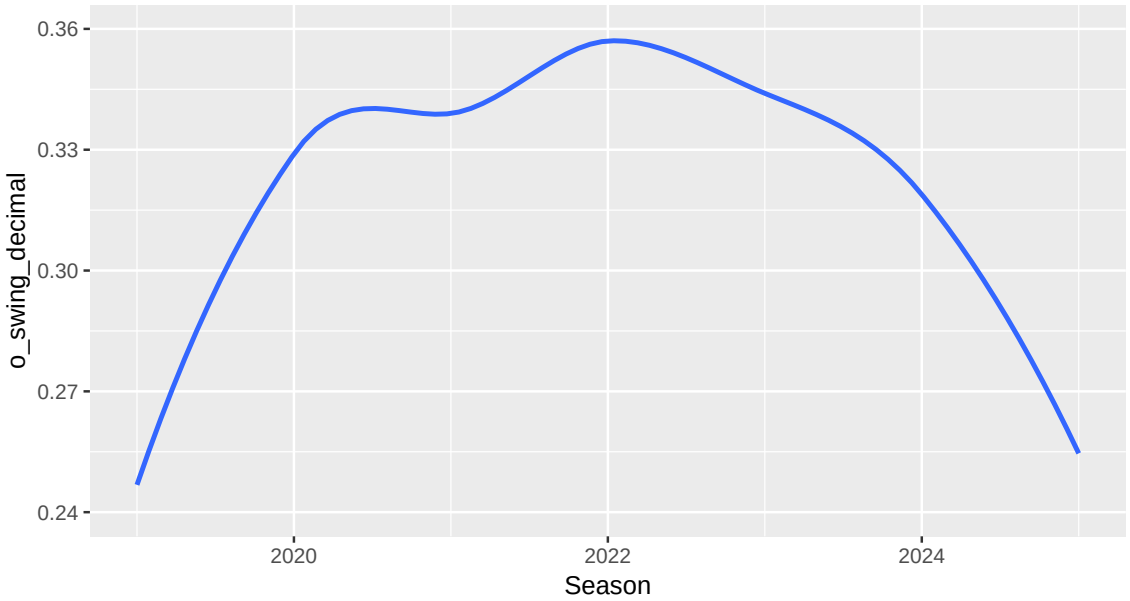
His ground ball rate is the lowest of his career. He usually excels at getting ground balls and weak hits.

4.3.3 Strikeout Rate



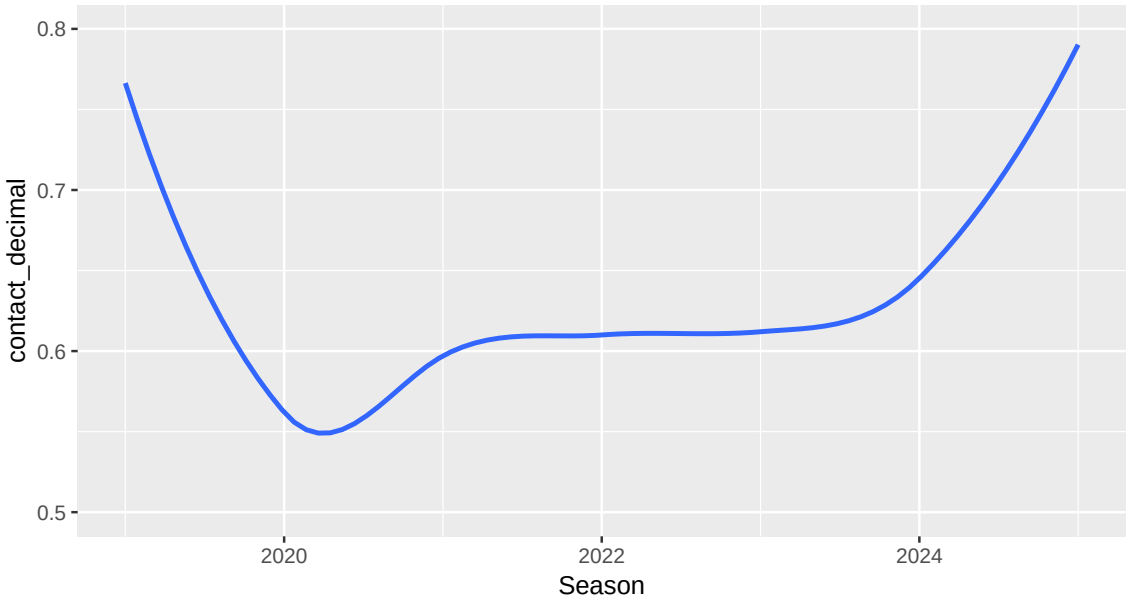
Williams' strikeout rate is the lowest of his career, but it was also very low his rookie season. Williams' ability to get strikeouts is one of the reasons he is so effective, so this is a bit concerning.

4.3.4 Chase Rate



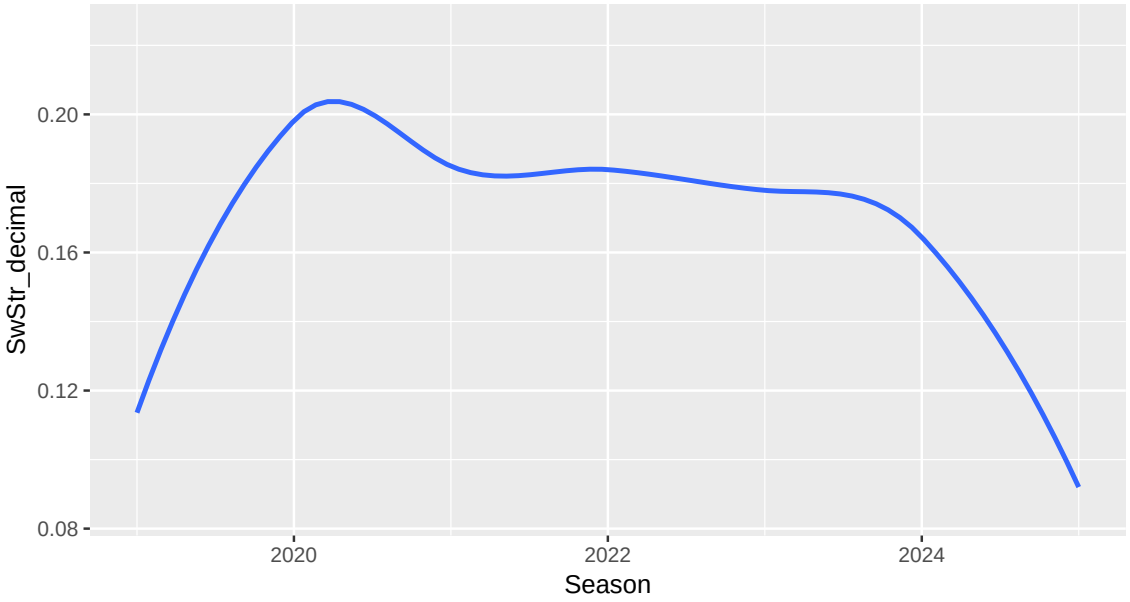
His chase rate is very low, but it was lower his rookie season. This is related to his decrease in strikeouts. He usually gets a lot of chases on his changeup out of the zone. This leads to strikeouts. However, that is not happening this season. The low chase rate his rookie season would also explain the low strikeout rate that season, too.

4.3.5 Contact Rate



Williams' contact rate is the highest of his career. This may be due to the lower chase and strikeout rates. Players are striking out less and putting the ball in play against Williams a lot more this season.

4.3.6 Swing and Miss Rate



Williams' swing and miss rate is also the lowest of his career, which explains the low strikeout rate and high contact rate.

5 Conclusion

Devin Williams is having a very bad season so far, but I still think that he is a great pitcher. He has only pitched 8 innings so far, so that is a very small sample size. It also may take some time to adjust to his new team, since he has been a Brewer his whole career. His rookie season, where he pitched just 13.2 innings was not great, but he became one of the best relievers in baseball after that. I think Devin Williams will return to being an elite reliever.

I think the most valuable metrics for a pitcher are ERA, FIP, xERA, and xFIP, but I think the most valuable of those four is xERA. It includes all types of outcomes, unlike FIP, but it is not as affected by fielding as ERA, since it is based on statcast data and hit probability.

The key to success for a pitcher is getting batters to chase pitches out of the zone, leading to weaker hit balls, more swing and misses, and more strikeouts.

6 References

1. "Devin Williams - Stats - Pitching",
   https://www.fangraphs.com/players/devin-williams/15816/stats?position=P.

2. "Fielding Independent Pitching (FIP) - Glossary",
   https://www.mlb.com/glossary/advanced-stats/fielding-independent-pitching.

3. "Weighted On-Base Average (wOBA) - Glossary",
   https://www.mlb.com/glossary/advanced-stats/weighted-on-base-average.

4. "Expected Weighted On-Base Average (xwOBA) - Glossary",
   https://www.mlb.com/glossary/statcast/expected-woba.

5. "Expected ERA (xERA) - Glossary",
   https://www.mlb.com/glossary/statcast/expected-era.

6. "Expected Fielding Independent Pitching (xFIP) - Glossary",
   https://www.mlb.com/glossary/advanced-stats/expected-fielding-independent-pitching

7 Appendix

Data tables and R code can be viewed on GitHub.
https://github.com/ChrisCallender46/Devin-Williams-Data