STA 160 Final: A Five-Year Performance Review of Clayton Kershaw

By: Theodore Lam, Patrick Manson, Brandon Sparks, Zhanyi Qiu

Introduction

In recent years, the availability of advanced metrics has revolutionized the way we analyze and understand player performance in Major League Baseball (MLB). One of the primary sources of these advanced metrics is Statcast, a state-of-the-art tracking technology installed in every MLB stadium since 2015. Statcast provides a wealth of data, including pitch velocity, spin rate, and movement, among many others. This project focuses on the evolution of MLB pitchers over a five-year period, using data from Statcast to track and analyze changes in their performance metrics.

Clayton Kershaw

Clayton Kershaw is a name synonymous with pitching excellence in Major League Baseball. Since his debut with the Los Angeles Dodgers in 2008, Kershaw has established himself as one of the most dominant and consistent pitchers in the game. With multiple Cy Young Awards, an MVP, and numerous All-Star selections, Kershaw's career is a testament to his exceptional skill, work ethic, and competitive spirit.

Known for his devastating curveball, pinpoint control, and ability to adjust his pitching strategy, Kershaw has maintained an elite level of performance throughout his career. His dominance on the mound is reflected in his career statistics, including a consistently low ERA, high strikeout rates, and numerous wins. Despite dealing with injuries in recent years, Kershaw has adapted his pitching style to remain effective against hitters in an evolving league.

Project Scope

This project aims to analyze Clayton Kershaw's pitching performance over a five-year period from 2015 to 2019. By leveraging Statcast data, we will examine key pitching metrics such as type of pitch, spin rate, and horizontal/vertical movement of pitches. The goal is to identify trends and changes in Kershaw's performance, understand how these metrics have evolved, and assess their impact on his overall effectiveness as a pitcher. By comparing these metrics year-over-year, we aim to identify trends and factors contributing to Kershaw's success or decline. This analysis will provide a comprehensive view of how Kershaw has adapted his strategies and mechanics over time to maintain or improve his performance, despite the challenges posed by injuries and evolving hitter strategies.

Average Launch Angle

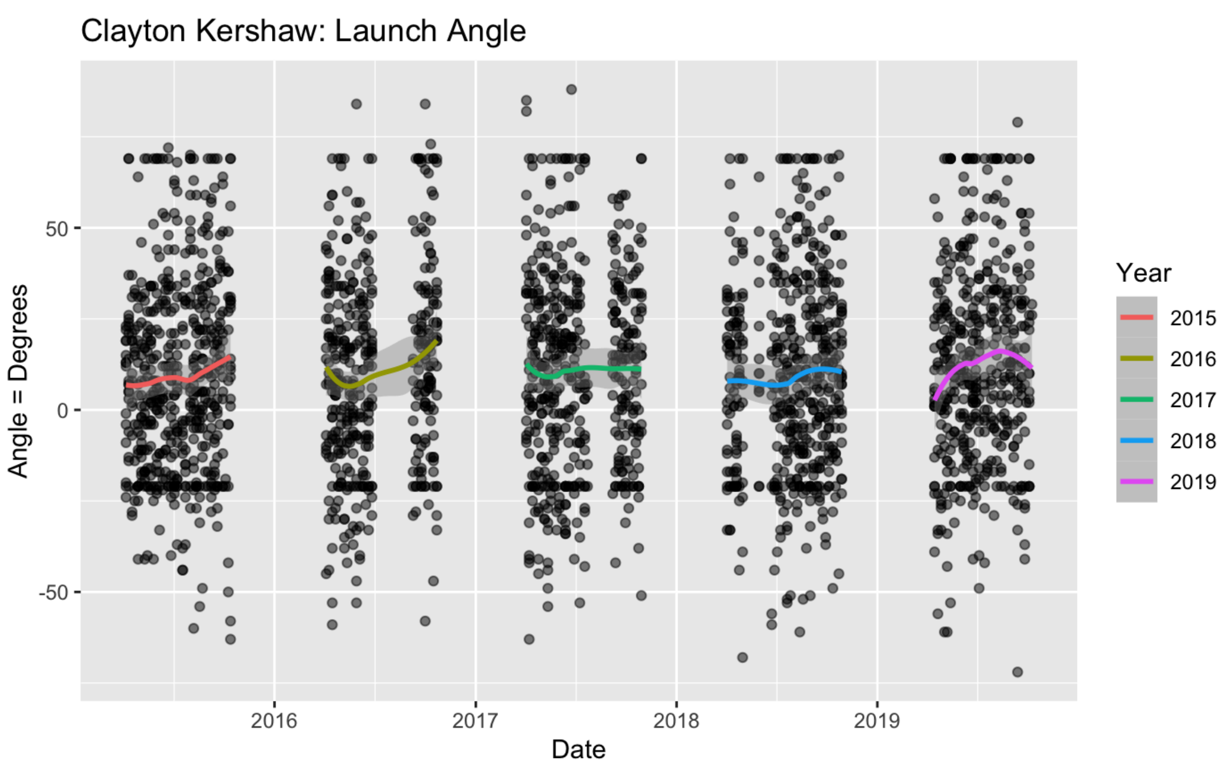
Launch angle is a critical metric that measures the vertical angle at which the ball leaves the bat after contact. Higher launch angles typically indicate that the ball is hit more upwards, resulting in fly balls and potential home runs, whereas lower angles suggest more ground balls and line drives. Analyzing Kershaw's average launch angle over the years reveals important trends in how batters have been making contact with his pitches.

Analysis

In 2015, the average launch angle for balls hit off Kershaw was 9.2 degrees. This relatively low angle indicated that batters were hitting more ground balls and line drives. The following year, 2016, saw an increase in the average launch angle to 10.8 degrees. This rise suggests that batters were starting to hit Kershaw's pitches at higher angles, potentially leading to more fly balls.

In 2017, the average launch angle slightly decreased to 10.5 degrees, maintaining a similar trend of higher contact compared to 2015. Interestingly, in 2018, the launch angle returned to the 2015 level at 9.2 degrees, indicating a shift back to more ground balls and line drives. However, 2019 saw a significant increase to 12.3 degrees, suggesting that batters were hitting Kershaw's pitches with even higher angles, leading to more fly balls and potential home runs. These fluctuations in launch angle over the years may reflect changes in Kershaw's pitching strategies, variations in his pitch effectiveness, or adjustments by batters in their approach against him.

The average launch angle against Kershaw's pitches fluctuated between 9.2 and 12.3 degrees from 2015 to 2019, indicating changes in how batters made contact with his pitches, with higher angles suggesting more fly balls and potential home runs in certain years.



Batted Ball Speed (Exit Velocity)

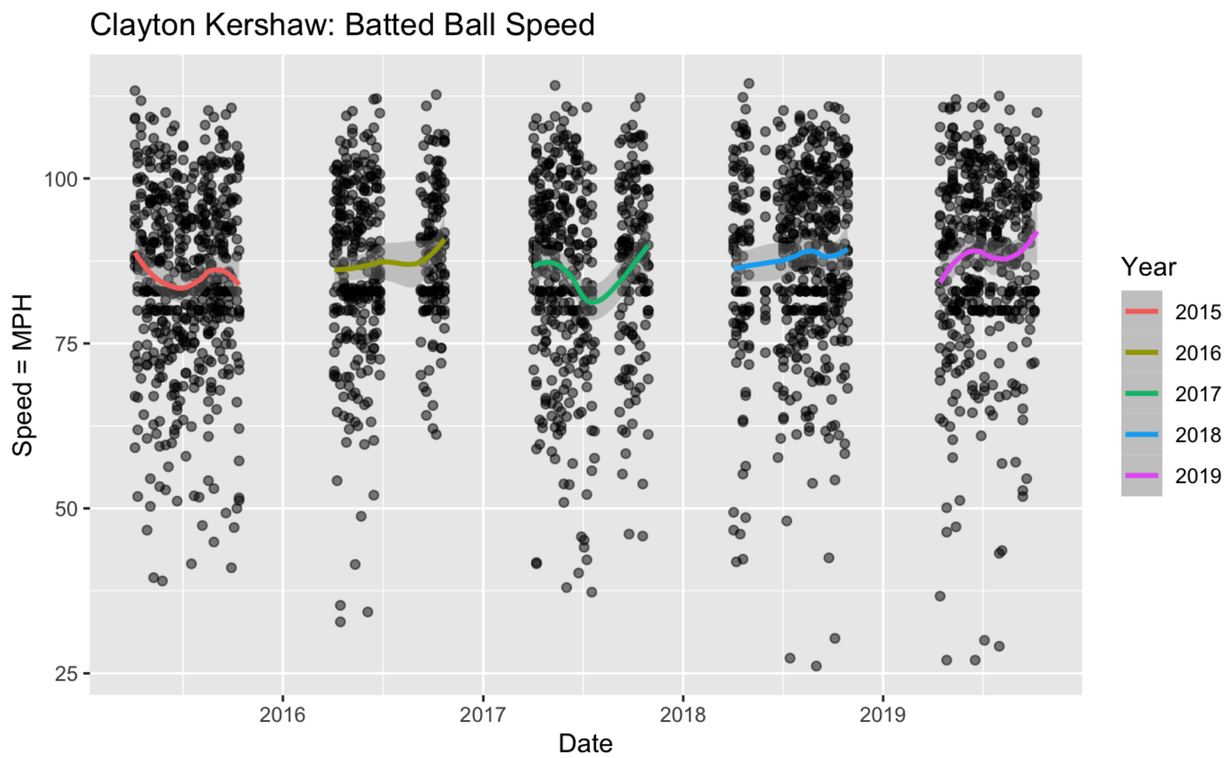
Batted ball speed, or exit velocity, measures the speed of the ball as it comes off the bat after being hit. Higher exit velocities generally indicate harder-hit balls, which are more likely to result in hits, extra-base hits, and home runs. Examining Kershaw's average batted ball speed provides insights into how well batters have been making contact with his pitches.

Analysis

In 2015, the average batted ball speed against Kershaw was 85.2 mph. This relatively moderate speed suggested that batters were not hitting Kershaw's pitches particularly hard. However, in 2016, there was an increase to 87.4 mph, indicating that batters were making harder contact. The average batted ball speed slightly decreased to 85.9 mph in 2017, although it remained higher than in 2015. This indicated a reduction in hard contact compared to the previous year but still higher than the earlier period.

The trend continued with another increase in 2018, reaching 88.0 mph. By 2019, the average batted ball speed was 88.1 mph, suggesting continued hard contact by batters against Kershaw's pitches. The overall increase in batted ball speed over the years indicates that batters have been increasingly able to make solid contact with Kershaw's pitches, potentially leading to more hits and runs scored against him.

From 2015 to 2019, the average batted ball speed against Kershaw increased from 85.2 mph to 88.1 mph, indicating that batters were making progressively harder contact with his pitches over the years.



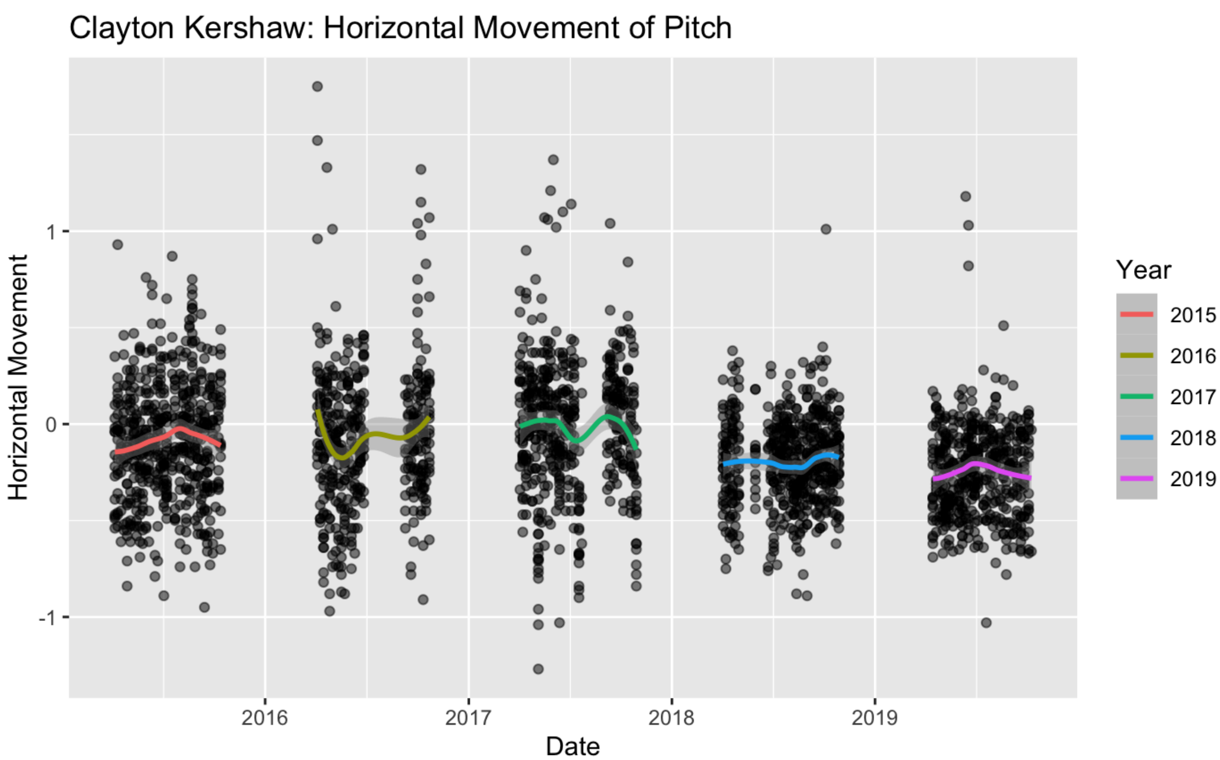
Movement of Pitch

In baseball, the movement of a pitch is a fundamental aspect that significantly impacts a pitcher's performance and a batter's ability to make contact. Pitch movement refers to the deviation of the baseball from its initial straight-line trajectory as it travels towards home plate. This movement is a result of various physical forces and properties, such as spin, velocity, and the interaction with air resistance.

Pitch movement can be categorized into two primary types: horizontal and vertical. Horizontal movement describes the side-to-side deviation, often creating a "sweeping" effect that can deceive batters. Pitches like sliders and cutters are known for their pronounced horizontal movement. Vertical movement, on the other hand, refers to the up-and-down deviation, influenced by gravity and the Magnus effect. This can cause pitches like curveballs to drop sharply or fastballs to appear to rise relative to the batter's expectation.

Horizontal Movement of Pitch

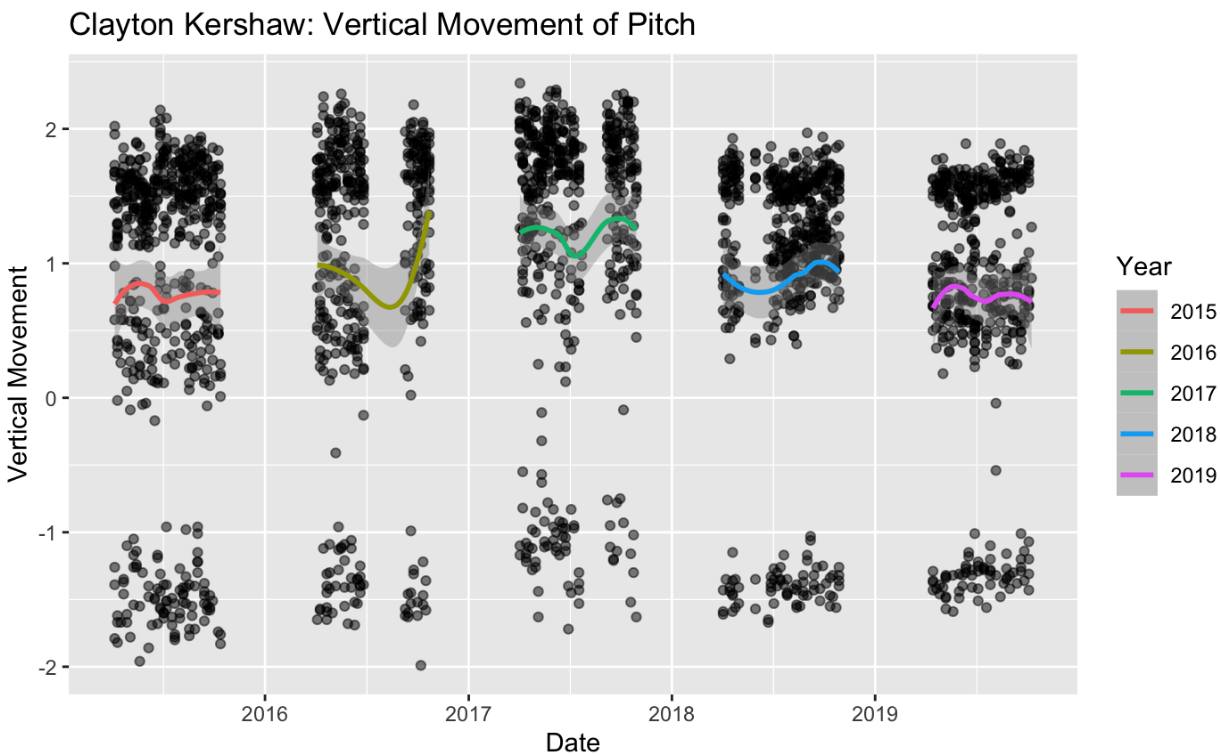
In 2015, the average horizontal movement of Kershaw's pitches was -0.09 feet, indicating slight movement towards the left-handed batter. This movement was fairly consistent in 2016, with an average of -0.08 feet. However, in 2017, the horizontal movement decreased significantly to -0.01 feet, suggesting less side-to-side deviation. This trend reversed in 2018, with horizontal movement increasing to -0.19 feet, and further to -0.25 feet in 2019, indicating a greater deviation towards left-handed batters. The increase in horizontal movement in the later years suggests that Kershaw may have adjusted his pitching mechanics or grip to achieve greater movement, potentially making his pitches more deceptive.



Vertical Movement of Pitch

Vertical movement, on the other hand, showed notable fluctuations over the years. In 2015, the average vertical movement was 0.78 feet, indicating a moderate rise on Kershaw's pitches. This increased to 0.97 feet in 2016, and further to 1.24 feet in 2017, suggesting even more rise or less drop, making pitches harder to hit squarely. However, the trend reversed in 2018, with the vertical movement decreasing to 0.92 feet, and further to 0.77 feet in 2019, similar to 2015 levels. The significant rise in vertical movement in 2016 and 2017 suggests adjustments that Kershaw made to increase the effectiveness of his pitches. The decrease in 2018 and 2019 might indicate changes in his pitching approach or a natural decline in his pitch characteristics.

Kershaw's pitches showed increased horizontal movement from 2017 to 2019, indicating greater side-to-side deviation. Vertical movement peaked in 2016 and 2017, suggesting increased pitch rise due to backspin during those years, but decreased in 2018 and 2019, indicating less rise or more drop in his pitches.

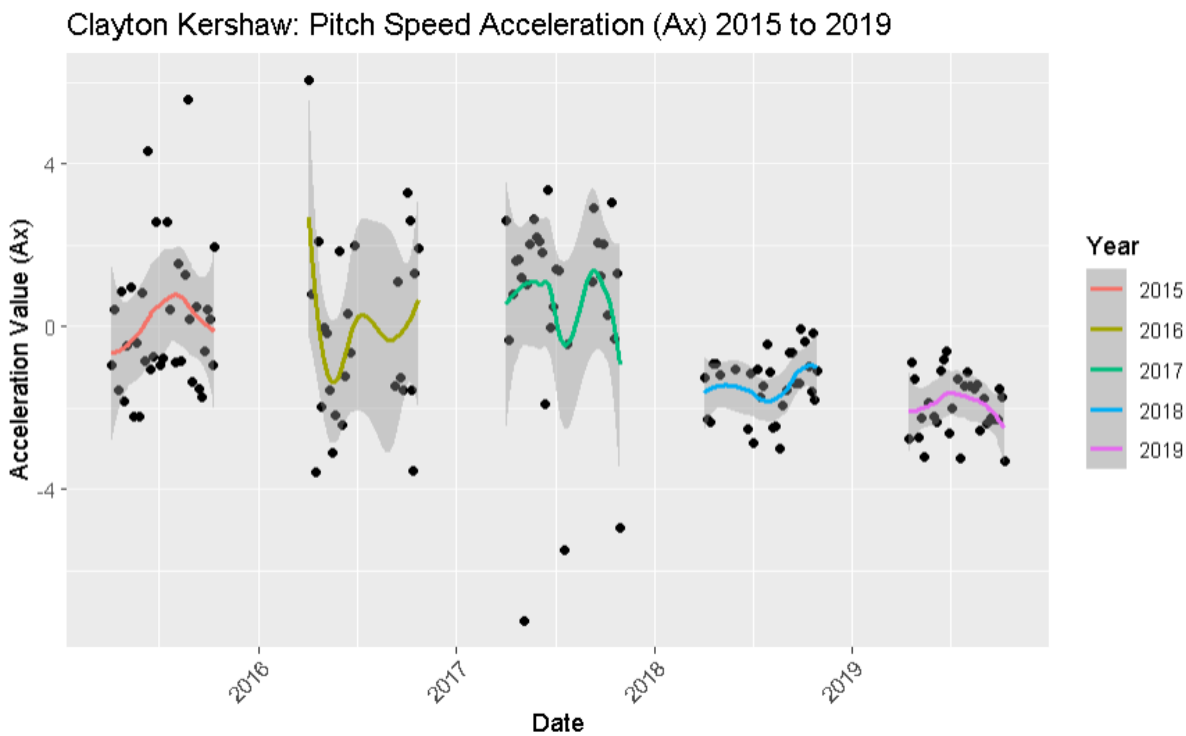


Coordinate of pitch acceleration

The coordinate of pitch acceleration plays a crucial role in understanding and enhancing a pitcher's performance and strategy. Pitch acceleration, which measures the rate of change of the pitch's velocity, provides deep insights into the physical dynamics of a pitch. By analyzing the acceleration data, coaches and analysts can discern subtle variations in a pitcher's mechanics that might affect the ball's trajectory, velocity, and ultimately, its effectiveness. Furthermore, pitch acceleration data can be used to refine pitch selection and sequencing during games, enhancing the pitcher's ability to deceive batters and improving overall team performance. Understanding pitch acceleration coordinates enables teams to gain a competitive edge through precise and data-driven decision-making.

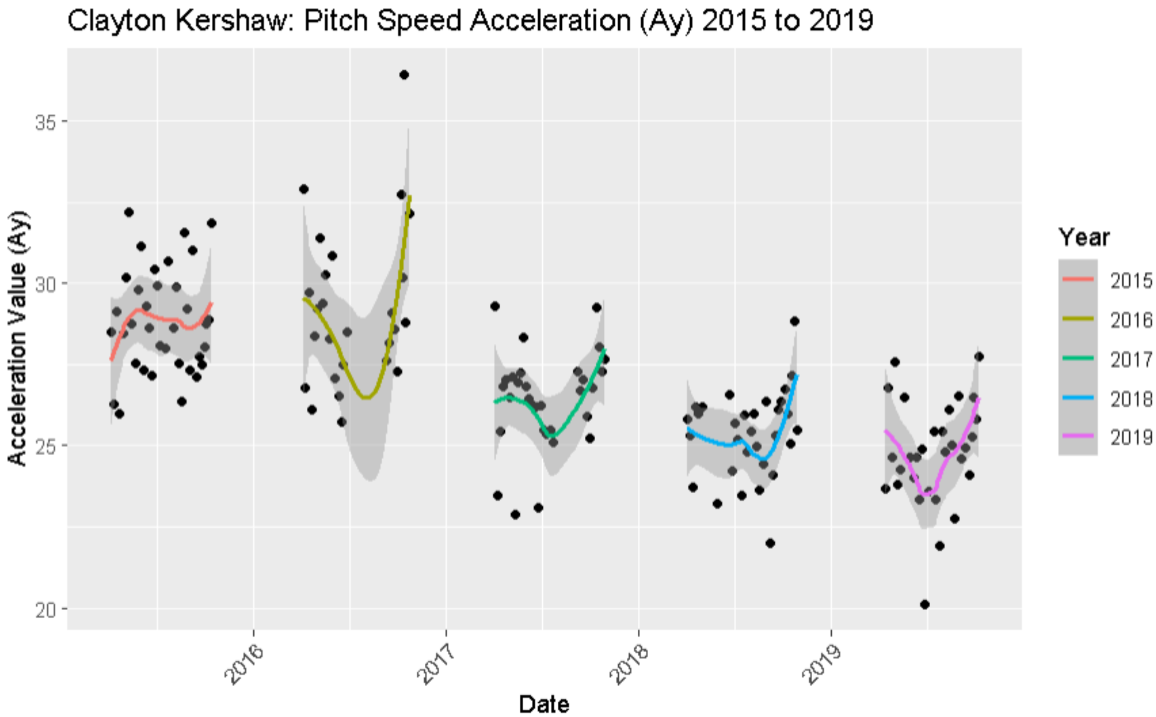
X coordinate of pitch acceleration

The x-direction acceleration shows surprising changes in Clayton Kershaw's pitch acceleration from 2015 to 2019. His pitch accelerations fluctuated between approximately -4 and 4 feet per second instead of showing a trend in improvement or even decline. In 2015 and 2016, the acceleration values include both positive and negative numbers, suggesting a very inconsistent year. The trend line for 2017 is more centralized around zero, indicating a year of more balanced lateral movement. However, in 2018 and 2019, the values predominantly trend towards the negative, reflecting a slight shift in Kershaw's pitch dynamics. This fluctuation in lateral acceleration components could be attributed to various factors, such as adjustments in his approach to different batters or physical condition over the seasons.



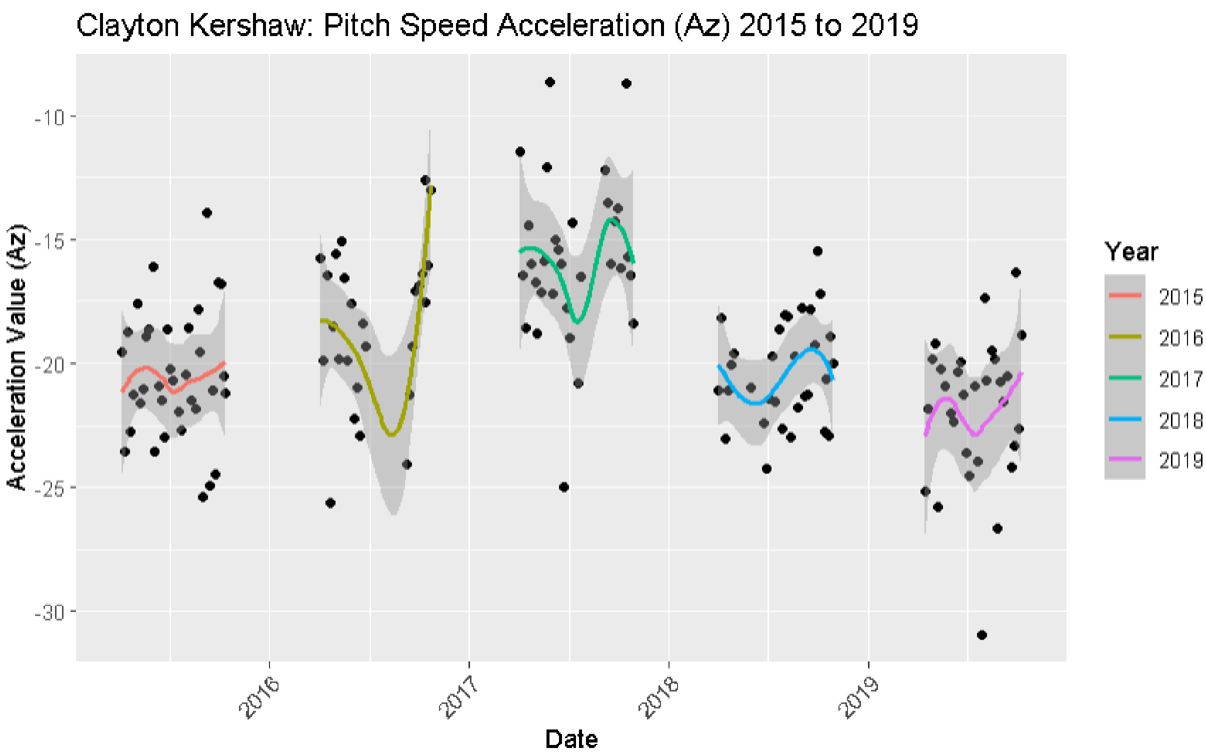
Y coordinate of pitch acceleration

Unlike x-direction acceleration, the y direction acceleration shows a more visible trend. The data points range between 20 and 35 feet per second, with a noticeable decline in the yearly averages. In 2015 and 2016, Kershaw's pitches displayed higher y-direction accelerations, often between 30 and 35. However, a significant drop is observed in 2017, with average values descending to the 25 to 30 range. This downward trend continues into 2018 and 2019, where the values further decrease to around 20 to 25. The steady decline in y-direction acceleration may indicate adjustments in Kershaw's pitching mechanics, possibly aiming for greater control or compensating for changes in physical condition. It could also reflect strategic changes in his approach to pitching, focusing more on pitch movement and control rather than speed alone. This consistent decrease in acceleration suggests a deliberate change in pitching style which is most likely a response to the natural evolution of his career or an adaptation to the competition around him.



Z coordinate of pitch acceleration

The z-direction acceleration values range from -30 to -10 feet per second, reflecting vertical acceleration components. In 2015 and 2016, Kershaw's pitches are characterized by relatively higher vertical accelerations, clustered around -15 to -20. However, in 2017, there is a noticeable shift towards more negative values, with averages moving to the -20 to -25 range. This shift indicates a change in the downward acceleration component of his pitches. Interestingly, from 2018 onwards, there is a slight increase in the vertical acceleration values, trending back towards -20 feet per second. This pattern suggests an initial decline in vertical acceleration, possibly due to adjustments in Kershaw's pitching mechanics or physical changes. The randomness of Kershaw's pitching speed acceleration in the z-direction shows Kershaw's lack of need for training in this statistic in favor of other pitching styles that would benefit him more.

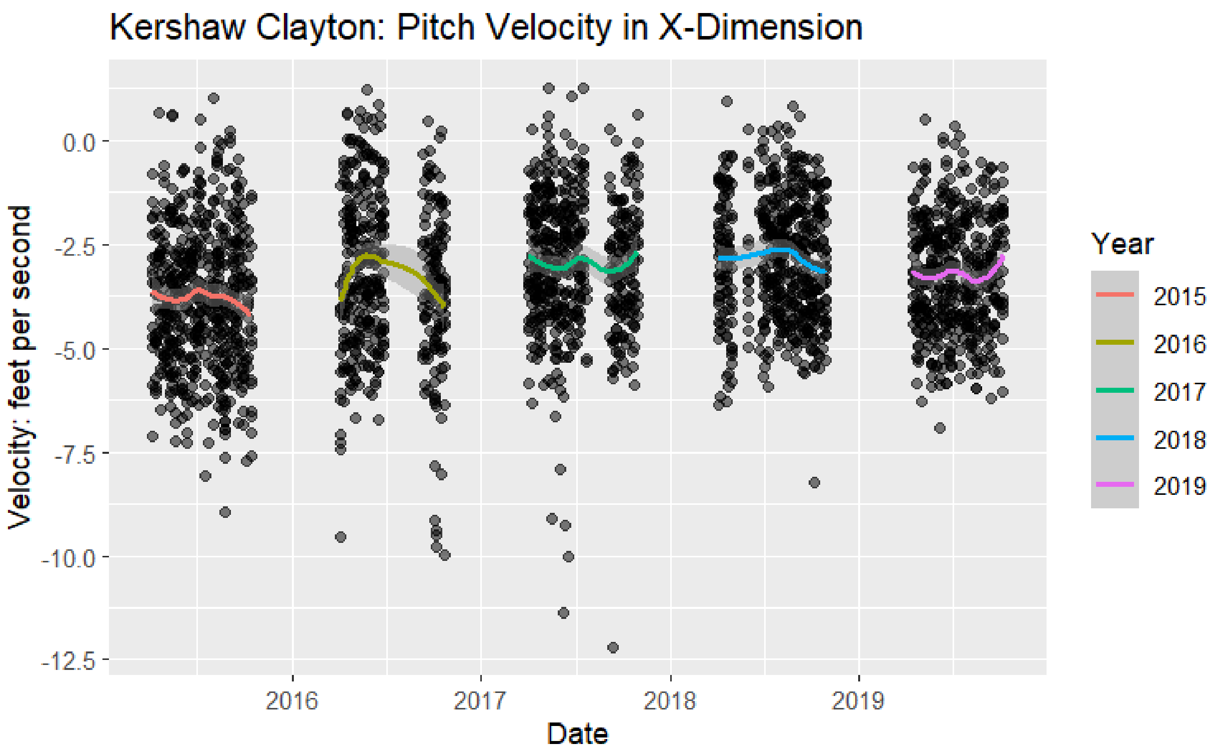


Pitch Velocity

Pitch velocity is the speed of a baseball thrown by a pitcher to the home plate. It is measured in X, Y, and Z coordinates to analyze various aspects of a pitcher's behavior. Pitch velocity can be used to analyze a pitcher's pitching performance and strategy adjustment.

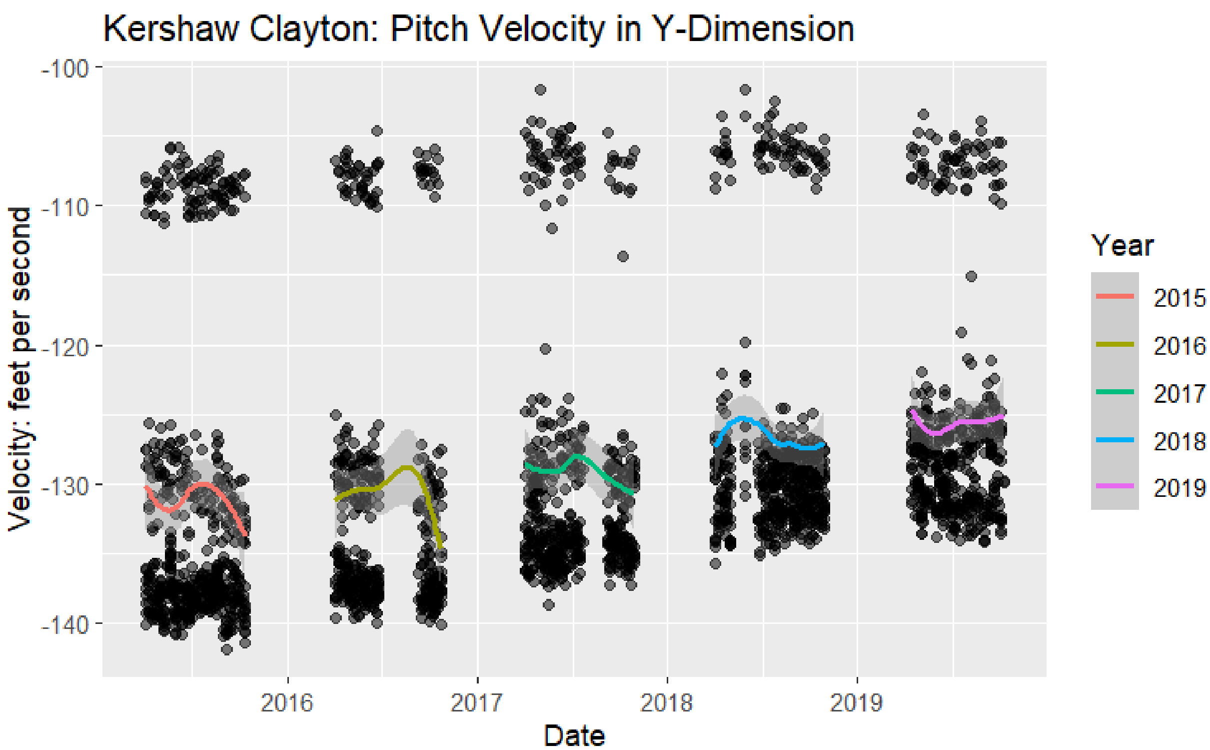
X-Coordinate of Pitch Velocity

The x-coordinate pitch velocity represents the lateral movement of the pitch. Kershaw's average x-coordinate pitch velocity fluctuates from -3.79 to -2.83 feet per second across the time period we chose to analyze. In 2015, Kershaw's average x-coordinate pitch velocity was -3.79 feet per second. For a left-handed pitcher like Kershaw, the ball moves more toward the left, which is a batter's outer half, essentially away from the batter. It suggests that Kershaw used a trickier pitching technique for batters to suffer to hit. However, in 2016, there was a noticeable increase to -3.25 feet per second. The less negative value suggests a reduction in lateral movement compared to 2015. This could be a strategic adjustment; Kershaw considered taking advantage of speed and precision over control. This suggests that Kershaw is relying on faster balls to make the batter unable to hit. This trend continues for the next two years. In 2017 and 2018, the x-coordinate pitch velocity further increased to -2.96 and -2.83. It was the peak velocity from 2015 to 2019, suggesting that Kershaw constantly evolves his pitching technique for a fastball. However, in 2019, the velocity decreased to -3.24 feet per second, suggesting that Kershaw adjusts the lateral movement of the pitch for a better balance between speed and control.



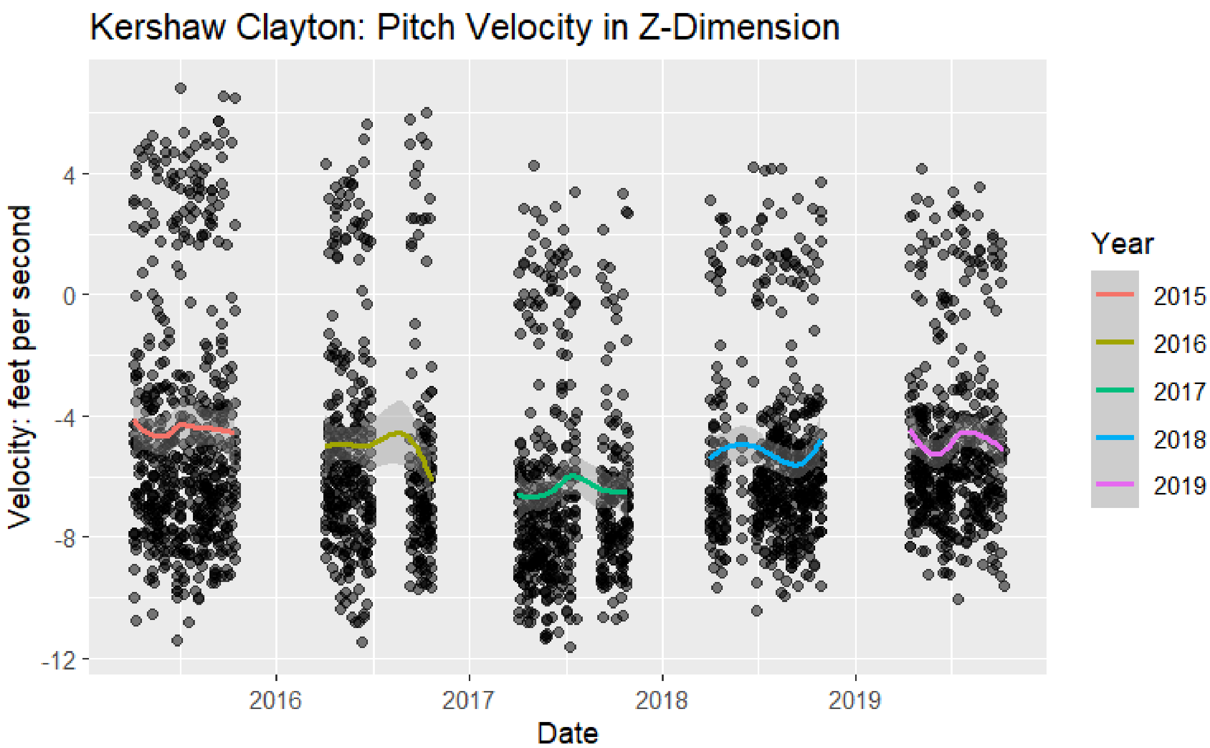
Y-Coordinate of Pitch Velocity

The y-coordinate pitch velocity represents the speed of the pitch towards the home plate. Kershaw's average y-coordinate pitch velocity increased from -131.18 to -125.71 feet per second in the time period we are analyzing. In 2015, Kershaw's average y-coordinate pitch velocity was -131.18. This value indicates a high pitch speed toward home plate. However, in 2016, there was a slight increase to -130.98. The less negative value indicates a marginally faster pitch speed compared to 2015. This could be a strategic adjustment; Kershaw considered taking advantage of speed over pitch movements. This suggests that Kershaw relies on faster balls to make it harder for the batter to react. This trend continues for the next four years. In 2017, the y-coordinate pitch velocity further increased to -129.18. The continued slight improvement suggests an ongoing improvement and a strategic emphasis on throwing faster pitches. In 2018 and 2019, the velocity further increased to -126.85 and -125.71. It was the peak velocity from 2015 to 2019, suggesting that Kershaw constantly evolves his pitching technique for a fastball. This trend toward higher y-coordinate pitch velocity shows Kershaw's ability to throw fast pitches consistently and evolve into a power pitcher that relies on his formidable fastballs.



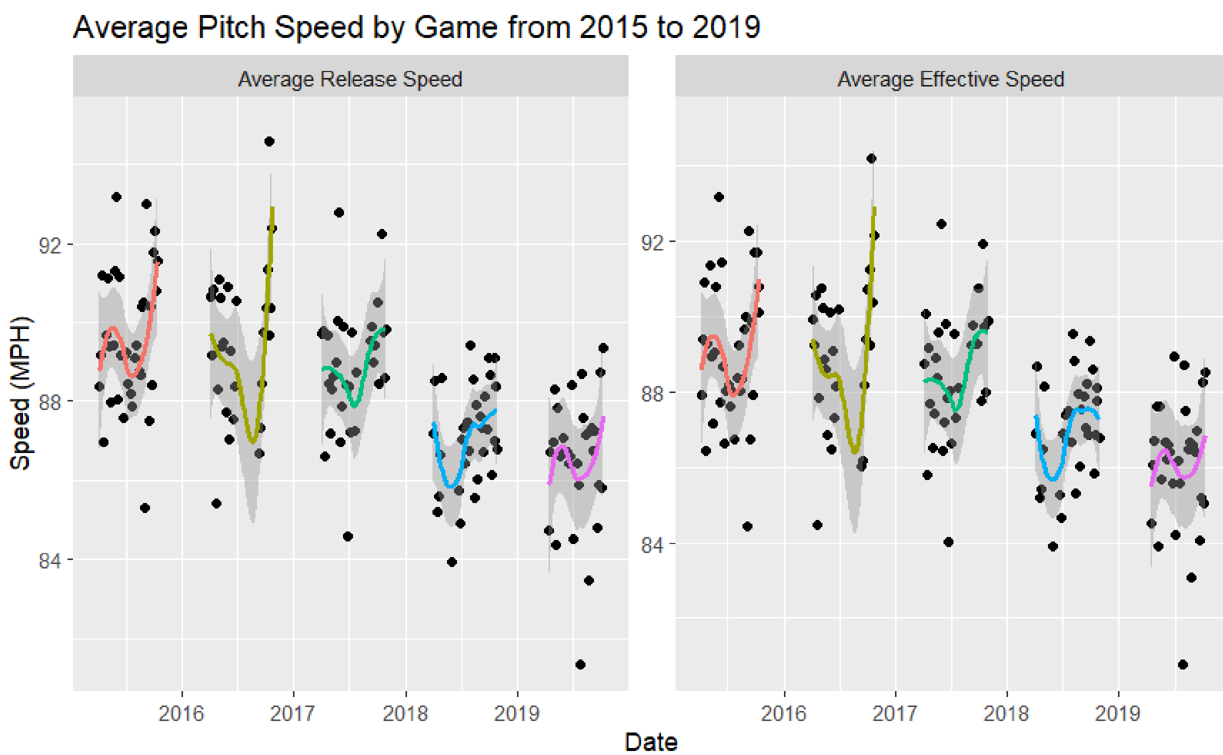
Z-Coordinate of Pitch Velocity

The z-coordinate pitch velocity represents the vertical movement of the pitch. Kershaw's average z-coordinate pitch velocity fluctuates from -4.46 to -6.47 feet per second across the time period we are analyzing. In 2015, Kershaw's average z-coordinate pitch was -4.46. This negative value indicates significant downward movement of the pitch. Pitches with greater downward movement can be more difficult for batters to hit solidly, which indicates that Kershaw emphasized breaking pitches like curveballs in 2015. However, in 2016, there was a slight decrease to -5.11. The higher negative value suggests a marginally decreased downward movement of Kershaw's pitches compared to 2015. This could be a strategic adjustment; Kershaw considered enhancing the power of his breaking pitch while maintaining a high speed. This suggests that Kershaw considered developing more pitch techniques. In 2017, the z-coordinate pitch velocity further decreased to -6.47. It was the peak velocity from 2015 to 2019 suggests that Kershaw has more precision over his powerful breaking pitch. The continued slight improvement between 2015 and 2017 suggests an ongoing improvement and a strategic emphasis on better precision with breaking pitches and maintaining a high speed. In 2018, the z-coordinate pitch velocity increased to -5.30, suggesting another strategic adjustment. This adjustment could reflect a shift towards balancing pitch movement between speed and precision. In 2019, the z-coordinate pitch velocity further increased to -4.88, suggesting a strategic shift away from extreme vertical movement. This strategic adjustment between 2018 to 2019 could reflect a shift from a sharp downward breaking pitch to a more diverse mix pitch.



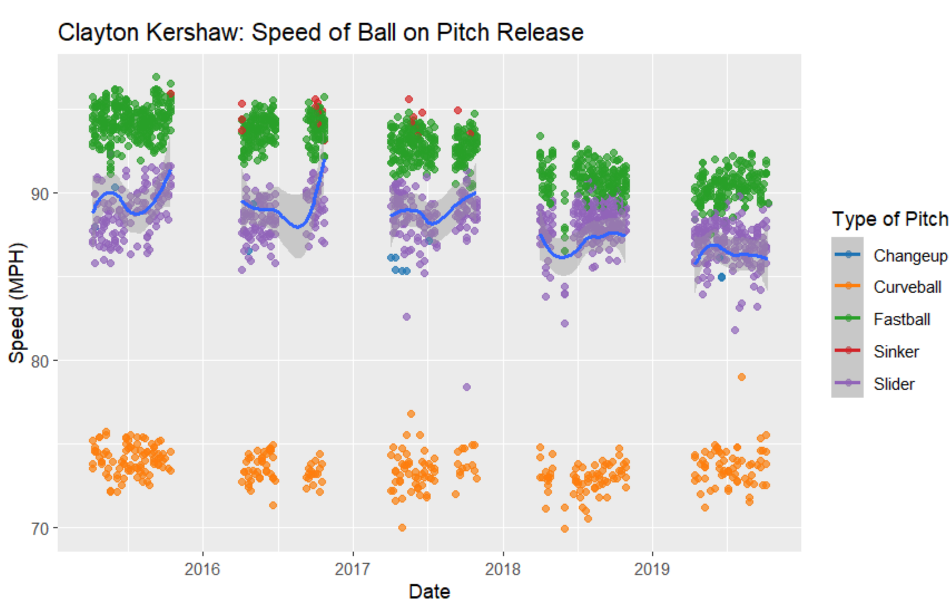
Overall Speed of Pitches

When discussing the overall speed, we need to break down the speed of the pitch into its two components: the speed of the pitch as it is thrown, and the speed of the ball when it crosses the home plate. In general, there should not be a lot of variance in the difference between these two variables, and we find that on average there is not a massive difference between these two stats, with an average difference of 0.32 miles per hour (mph). This is also reflected when looking at the game averages for these two statistics.

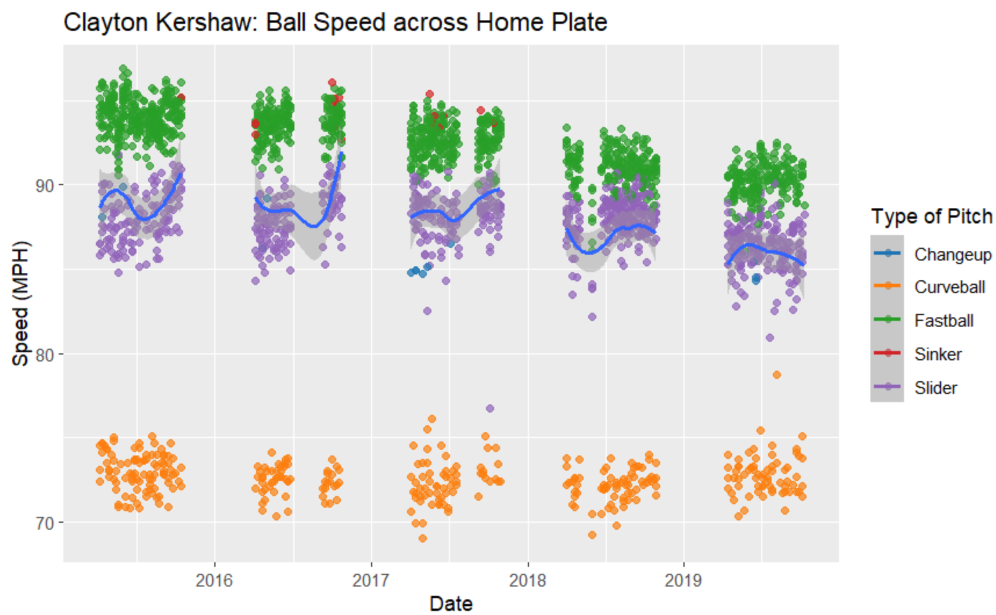


Analysis

The average release speed of Kershaw's pitches seems to have a downward trend over the years, starting at an average of 89.5 in 2015 and an average of 86.4 in 2019. We see the biggest change in speed around 2018 with a 1.8 mph difference. At the same time, we see that Kershaw



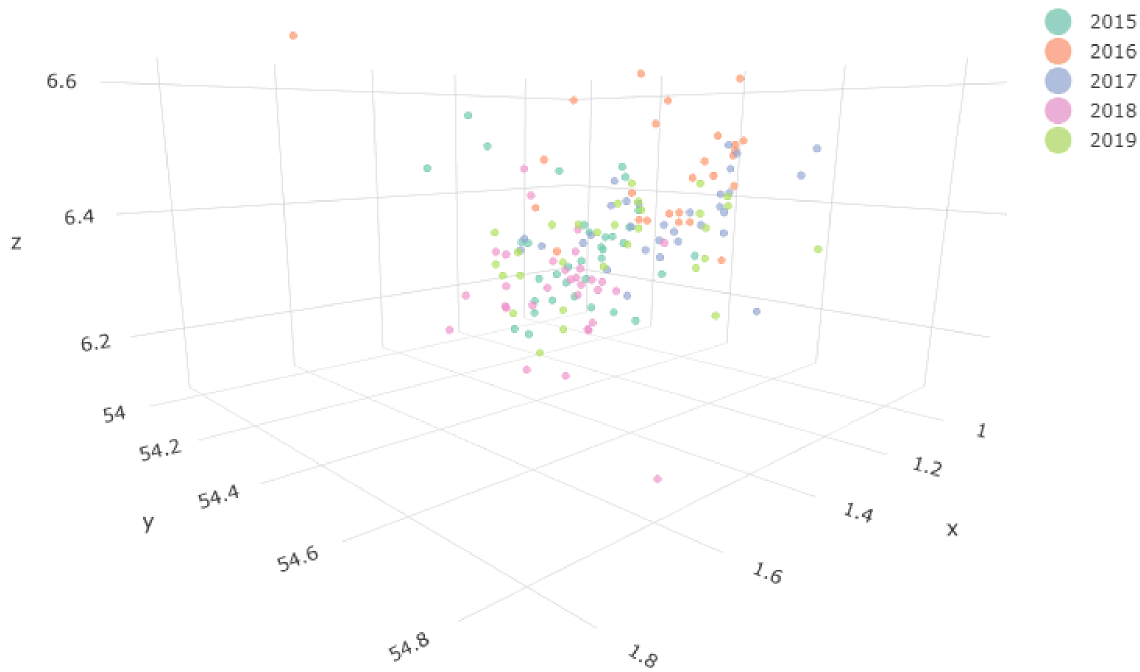
also stops throwing sinkers around this time, which could indicate that this was a planned move as his pitches seem to be a lot more consistent. When analyzing the individual years, there does appear to be a slight trend of Kershaw pitching faster in the later half of the season, seen in both the breakdown of all pitches and the average pitch speed per game. Additionally, when we look at the type of pitch thrown, there are clear groups between the different pitches in terms of speed.



Given the low variance between the release and effective speed, there are not any noticeable discrepancies between the plot for the release speed of all pitches and the effective speed for all pitches. There is slightly more spread in the pitch type groupings, but there are many variables that could cause differing speeds at the home plate, ultimately leading to a larger variance.

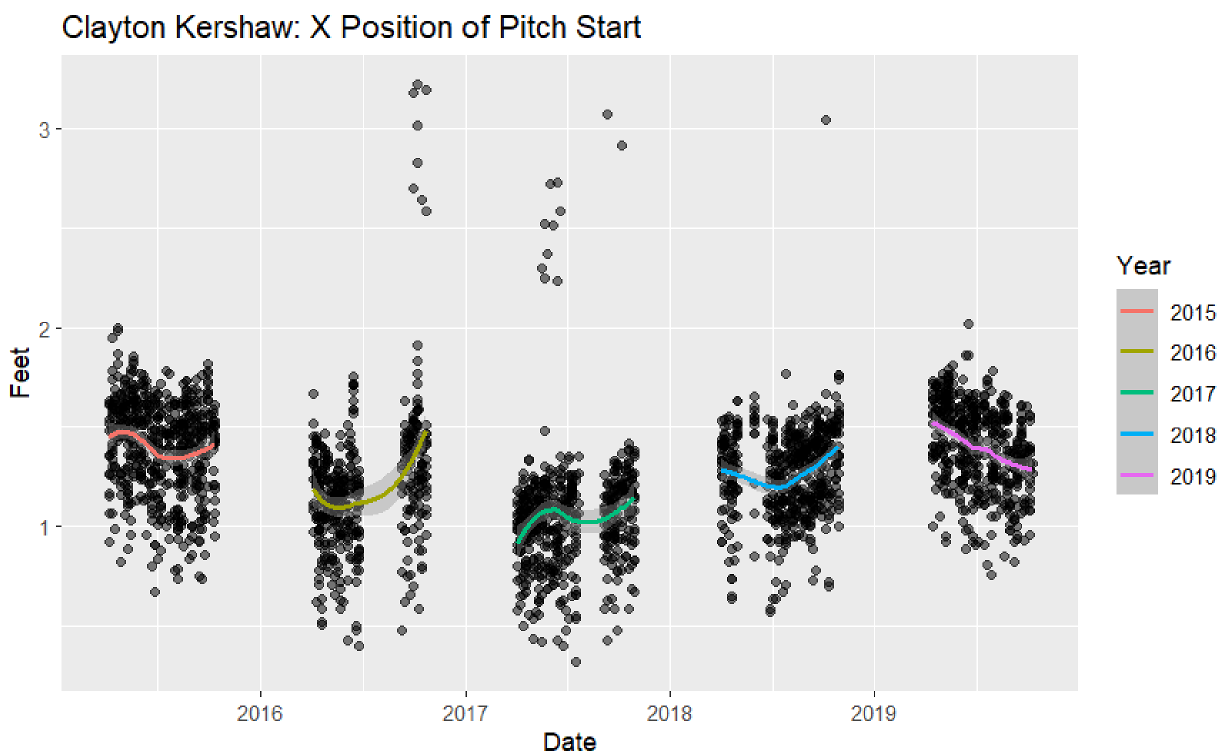
Pitch Release Position

We were able to obtain the x, y, and z distances for where a pitch was thrown in relation to the catcher. Upon first analysis, we noticed that the y distance, the distance from home plate to the pitcher's mound, was only the same value in 2015 and 2016. We assume that this data was not recorded at the time, and they simply used the average distance from the home plate to the pitcher's mound for this data. As such, we decided to keep the data as it was technically not inaccurate, and was approximately the same as the averages from the other years. We can see this as when viewing the average position per game on a 3D plot, as there isn't any obvious clustering of the 2015 and 2016 averages when viewed in a 3D space.



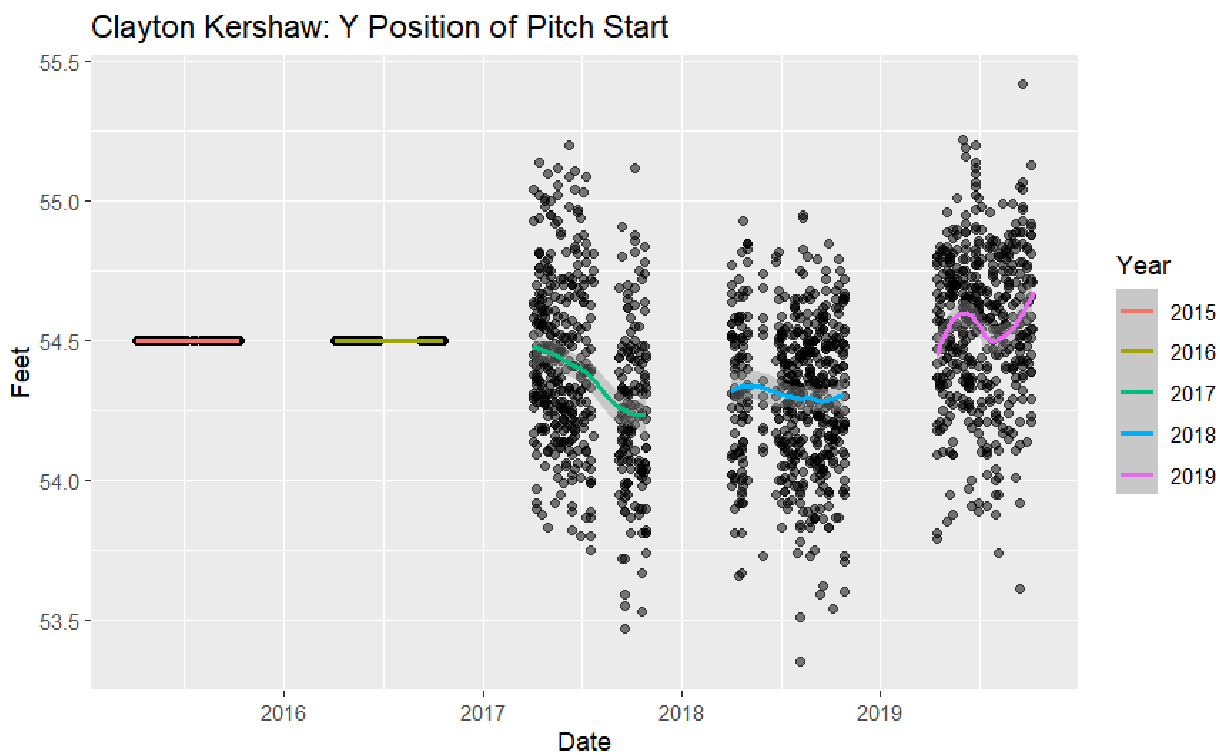
X Axis Starting Position

The starting position of the pitch on the X axis indicates how far away the pitch was in perspective to the catcher's center on a lateral axis. We see that in general most of Kershaw's pitches are within two feet of the catcher, with 2017 being the best year for him as the average was the lowest at 1.05 feet. Interestingly, Kershaw seems to start with a higher average X position, which hits a minimum in 2017 before returning to its initial average in 2019.



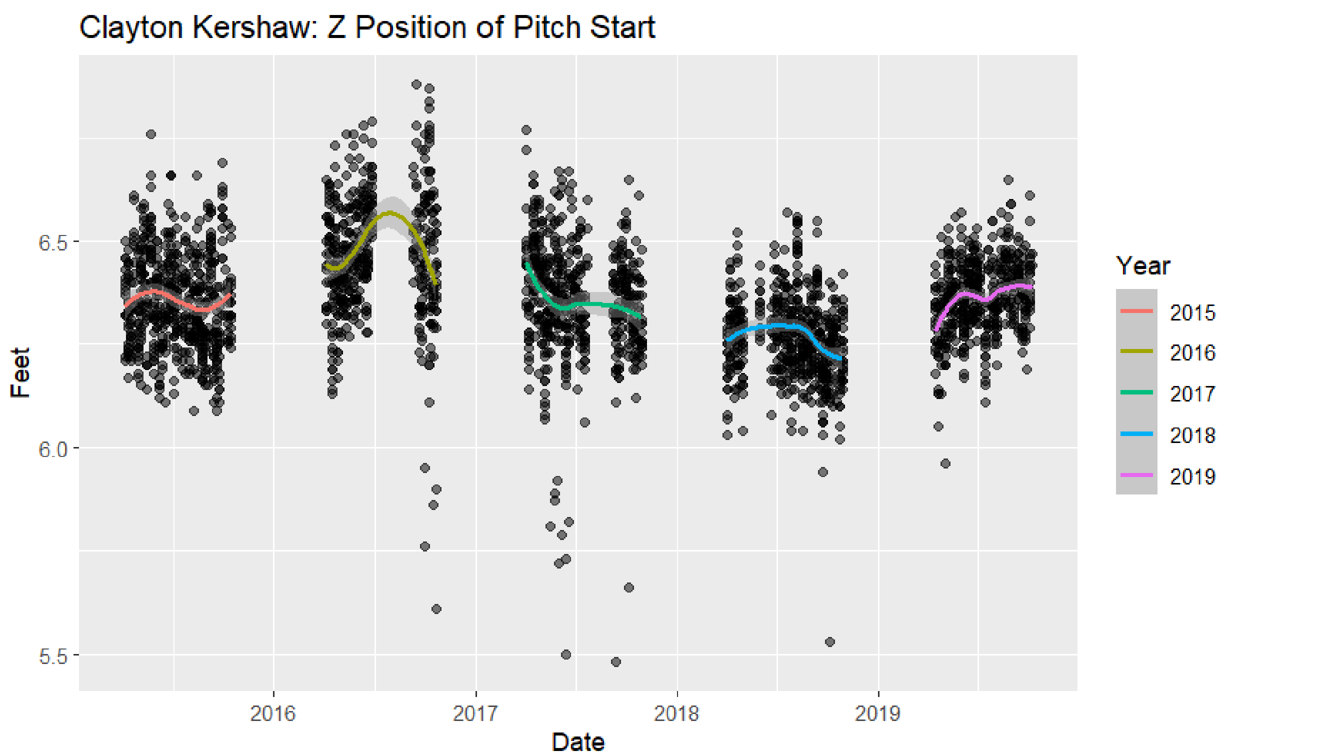
Y Axis Starting Position

As previously mentioned, we noticed that there was an issue with 2015 and 2016's data collection. Interestingly, we can see that the averages for the following years are approximately the same, if not a bit closer to the home plate than these two years. Another interesting point to notice is that for the years with detailed data collection, the variance appears to be within a foot, but it appears to be very spread out.



Z Axis Starting Position

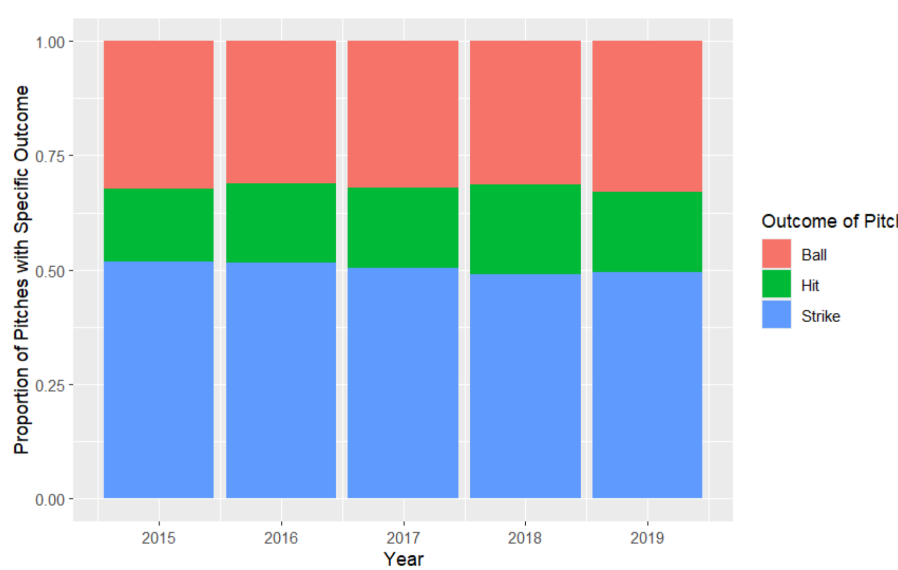
This is tracking how high or low the pitch was thrown. Like the other statistics in this section, there appears to be a lot of variance between the individual pitches, but there is a general convergence around 6.25 feet. Given Kershaw's height of 6 feet and 4 inches, we can conclude that on average he is throwing his pitches slightly below his head, around where his face is. This is normal for most pitchers, with only a few throwing underhand or submarine style. Interestingly, we see that there are low pitches with a position of under 6 feet, primarily in the end of 2016 season and the start of the 2017 season. There also doesn't seem to be any correlation between the type of pitch thrown and the height it was thrown at, so we are unsure why this trend happened for just that period of time.



Analysis of Pitch Results

To analyze the effect of these variables and quantify Kershaw's performance over the years, we decided to look at the number of balls, strikes, and hits he had over the 2015-2019 period and the outcomes of each batter-pitcher interaction.

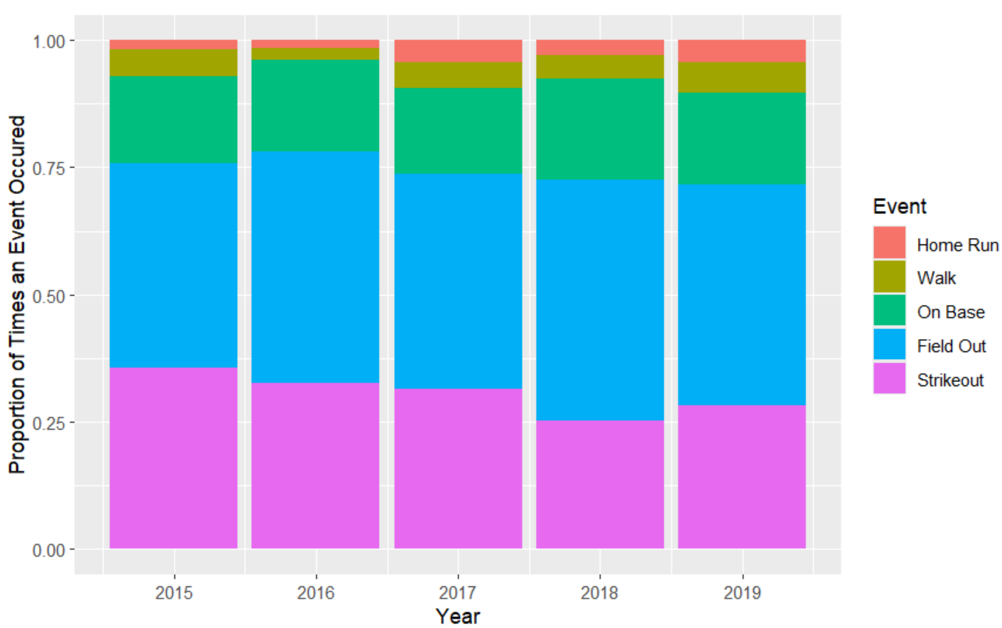
We started by looking at the proportion of balls, strikes, and hits. We saw that in general the proportion of balls, strikes, and hits doesn't change too much over the years. Approximately half of Kershaw's pitches are strikes across the years, and we see that



balls tend to make up over half of pitches that were not a strike.

However, just knowing the result of a pitch isn't enough to determine Kershaw's performance, as a hit could result in a field out, or after Kershaw threw 2 strikes the batter is able to get a home run. Due to the potential of events like this happening, we decided to analyze some of the things that happened due to his pitches. Namely, we looked at when his pitches resulted in a home run, a walk, a strikeout, a field out, and when it resulted in a batter getting on base. There are other edge cases that we did not include here, such as force outs, double plays, and field errors. We feel confident in removing them as they only make up 0.06% of all events we analyzed, and feel that they would make minimal difference to our analysis.

We find that nearly 75% of Kershaw's interactions with batters result in an out of some form, with field outs making up the majority of those outs. That is not to say that Kershaw does not have a decent amount of strikeouts, as they do make up approximately 30% of the events recorded, though we see that that number is noticeably lower in 2018 and 2019. Thus, we



conclude that in general, Kershaw's performance was at its best in 2015, and slightly declined in 2016 and 2017. It then took a larger dip in 2018 before a return to form in 2019.

We also noticed through our analysis that across all the years a batter made it to base approximately 17% of the time. Since this is a significant number, we decided to also analyze how many bases a batter would generally get off of a hit. Looking at our plot, we see that a vast majority of times that a batter makes it onto a base, they get a double. This is relatively consistent across years. While this does not inherently reflect Kershaw's overall performance, it does play into the amount of runs that he may have given up, even if preventing someone from getting a double is a fielder's responsibility.

Discussion on Factors Attributed to Kershaw's Performance

Based on the previous analysis of variables, we noticed that there appears to be a negative trend between the average Y pitch velocity and the number of strikeouts recorded. There also appears to be a slight positive correlation between the average release and effective pitch speed and the number of strikeouts. There is also a slight positive correlation between the x and y pitch accelerations and the number of strikeouts, but it appears to be stronger with the y axis pitch acceleration. These are likely not the only factors that play into the end result of a pitch, but from the variables that we analyzed these appear to be the ones with obvious trends.

Conclusion

In summary, this analysis of Clayton Kershaw's performance from 2015 to 2019 using advanced Statcast metrics reveals insights into the evolution of his pitching strategy and effectiveness. Over the five-year period, key metrics such as average launch angle, batted ball speed, pitch movement, pitch acceleration, and pitch velocity have demonstrated notable trends that reflect both Kershaw's adaptations and the responses of opposing batters.

The fluctuations in average launch angle and batted ball speed indicate that batters have increasingly made harder contact with Kershaw’s pitches, resulting in more fly balls and potential home runs in certain years. Our analysis of pitch movement revealed that while Kershaw increased the horizontal movement of his pitches in the later years, vertical movement showed significant rise in 2016 and 2017, followed by a decline in 2018 and 2019, suggesting potential strategic adjustments.

The pitch acceleration data provided insights into the subtle variations in Kershaw's mechanics, with each of the x, y, and z coordinates having different trends. The overall decrease in y-direction acceleration and the fluctuating z-direction acceleration suggest adjustments aimed at maintaining control and adapting to physical changes.

Kershaw’s pitch velocity data showed a trend towards faster pitches over the years, reflecting a strategic emphasis on speed. The consistency in his pitch release position further highlighted his precision and ability to adapt his mechanics to sustain performance.

Despite the overall decline in certain performance metrics, Kershaw continued to have a strong track record, with a relatively high proportion of field outs and strikeouts he had over the years. This shows that Kershaw is able to adapt his strategies and mechanics, even as he faced challenges from injuries and evolving hitter strategies.

In conclusion, Clayton Kershaw's performance over these five years demonstrates the dynamic nature of his pitching style, as it had a wide variance over the years. Ultimately, this shows that even though Kershaw may have had challenges, he was able to adapt and stay a strong pitcher with relatively consistent performance over the years.

Code Appendix

```
library(baseballr)
library(dplyr)
library(ggplot2)
library(reshape2)
library(zoo)

# Clayton Kershaw's player ID
player_id <- 477132

# Updated functions in the package `baseballr`
csv_from_url <- function(...){
  data.table::fread(...)
}

make_baseballr_data <- function(df, type, timestamp){
  out <- df %>%
       tidyr::as_tibble()
  class(out) <- c("baseballr_data","tbl_df","tbl","data.table","data.frame")
  attr(out,"baseballr_timestamp") <- timestamp
  attr(out,"baseballr_type") <- type
  return(out)
}

# @title
statcast_search <- function(start_date = Sys.Date() - 1, end_date = Sys.Date(),
playerid = NULL, player_type = "pitcher", ...) {
  # Check for other user errors.
  if (start_date <= "2015-03-01") {
       message("Some metrics such as Exit Velocity and Batted Ball Events have only
been compiled since 2015.")
  }
  if (start_date < "2008-03-25") {
       stop("The data are limited to the 2008 MLB season and after.")
       return(NULL)
  }
  if (start_date == Sys.Date()) {
       message("The data are collected daily at 3 a.m. Some of today's games may not
be included.")
  }
  if (start_date > as.Date(end_date)) {
       stop("The start date is later than the end date.")
       return(NULL)
  }

  playerid_var <- ifelse(player_type == "pitcher", "pitchers_lookup%5B%5D",
"batters_lookup%5B%5D")

  vars <- tibble::tribble(
       ~var, ~value,
       "all", "true",
       "hfPT", "",
```

```
    "hfAB", "",
    "hfBBT", "",
    "hfPR", "",
    "hfZ", "",
    "stadium", "",
    "hfBBL", "",
    "hfNewZones", "",
    "hfGT", "R%7CPO%7CS%7C&hfC",
    "hfSea", paste0(lubridate::year(start_date), "%7C"),
    "hfSit", "",
    "hfOuts", "",
    "opponent", "",
    "pitcher_throws", "",
    "batter_stands", "",
    "hfSA", "",
    "player_type", player_type,
    "hfInfield", "",
    "team", "",
    "position", "",
    "hfOutfield", "",
    "hfRO", "",
    "home_road", "",
    playerid_var, ifelse(is.null(playerid), "", as.character(playerid)),
    "game_date_gt", as.character(start_date),
    "game_date_lt", as.character(end_date),
    "hfFlag", "",
    "hfPull", "",
    "metric_1", "",
    "hfInn", "",
    "min_pitches", "0",
    "min_results", "0",
    "group_by", "name",
    "sort_col", "pitches",
    "player_event_sort", "h_launch_speed",
    "sort_order", "desc",
    "min_abs", "0",
    "type", "details"
) %>%
    dplyr::mutate(pairs = paste0(.data$var, "=", .data$value))

if (is.null(playerid)) {
    vars <- vars %>%
    dplyr::filter(!grepl("lookup", .data$var))
}

url_vars <- paste0(vars$pairs, collapse = "&")
url <- paste0("https://baseballsavant.mlb.com/statcast_search/csv?", url_vars)

tryCatch(
    {
    suppressMessages(
    suppressWarnings(
    payload <- csv_from_url(url, encoding ="UTF-8")
    )
```

```
      )
      },
      error = function(cond) {
      message(cond)
      stop("No payload acquired")
      },
      warning = function(cond) {
      message(cond)
      }
  )

  colos <- c("pitch_type", "game_date",
             "release_speed", "release_pos_x", "release_pos_z",
             "player_name", "batter", "pitcher",
             "events", "description", "spin_dir",
             "spin_rate_deprecated", "break_angle_deprecated",
             "break_length_deprecated", "zone", "des",
             "game_type", "stand", "p_throws",
             "home_team", "away_team", "type",
             "hit_location", "bb_type", "balls",
             "strikes", "game_year", "pfx_x",
             "pfx_z", "plate_x", "plate_z",
             "on_3b", "on_2b", "on_1b", "outs_when_up",
             "inning", "inning_topbot", "hc_x",
             "hc_y", "tfs_deprecated", "tfs_zulu_deprecated",
             "fielder_2", "umpire", "sv_id",
             "vx0", "vy0", "vz0", "ax",
             "ay", "az", "sz_top", "sz_bot",
             "hit_distance_sc", "launch_speed", "launch_angle",
             "effective_speed", "release_spin_rate",
             "release_extension", "game_pk", "pitcher_1",
             "fielder_2_1", "fielder_3", "fielder_4",
             "fielder_5", "fielder_6", "fielder_7",
             "fielder_8", "fielder_9", "release_pos_y",
             "estimated_ba_using_speedangle", "estimated_woba_using_speedangle",
             "woba_value", "woba_denom", "babip_value",
             "iso_value", "launch_speed_angle", "at_bat_number",
             "pitch_number", "pitch_name", "home_score",
             "away_score", "bat_score", "fld_score",
             "post_away_score", "post_home_score",
             "post_bat_score", "post_fld_score", "if_fielding_alignment",
             "of_fielding_alignment", "spin_axis",
             "delta_home_win_exp", "delta_run_exp")

  colNumber <- ncol(payload)
  if(length(colos) != colNumber){
      colos <- colos[1:colNumber]  # Adjust the length of column names to match the
payload
      message("New stats detected! baseballr will be updated soon to properly
identify these stats")
  }

  names(payload) <- colos
```

```
  payload <- payload %>%
       make_baseballr_data("MLB Baseball Savant Statcast Search data from
baseballsavant.mlb.com", Sys.time())
  return(payload)
}

# Fetch data for Clayton Kershaw from 2015 to 2019
kershaw_15 <- statcast_search(start_date = "2015-03-31", end_date = "2015-10-31",
playerid = player_id, player_type = 'pitcher')
kershaw_16 <- statcast_search(start_date = "2016-03-31", end_date = "2016-10-31",
playerid = player_id, player_type = 'pitcher')
kershaw_17 <- statcast_search(start_date = "2017-03-31", end_date = "2017-10-31",
playerid = player_id, player_type = 'pitcher')
kershaw_18 <- statcast_search(start_date = "2018-03-31", end_date = "2018-10-31",
playerid = player_id, player_type = 'pitcher')
kershaw_19 <- statcast_search(start_date = "2019-03-31", end_date = "2019-10-31",
playerid = player_id, player_type = 'pitcher')

# Combine data for all years
kershaw_data <- bind_rows(kershaw_15, kershaw_16, kershaw_17, kershaw_18, kershaw_19)
%>%
  mutate(Year = as.factor(substr(game_date, 1, 4))) %>%
  filter(type == "X") %>%
  filter(launch_speed != 0)

# Calculate average launch angle and batted ball speed by game date
kershaw_grpd <- kershaw_data %>%
  group_by(game_date) %>%
  summarise(
       `Average Launch Angle` = mean(launch_angle, na.rm = TRUE),
       `Average Batted Ball Speed` = mean(launch_speed, na.rm = TRUE)) %>%
  ungroup() %>%
  melt(id=c("game_date")) %>%
  mutate(Year = as.factor(substr(game_date,1,4)))

head(kershaw_grpd)

# Calculate average launch angle and batted ball speed by year
kershaw_avg_yr <- kershaw_data %>%
  group_by(Year) %>%
  summarise(
       angle = round(mean(launch_angle, na.rm = TRUE),1),
       speed = round(mean(launch_speed, na.rm = TRUE),1))

kershaw_avg_yr

# Plot data
kershaw_grpd %>%
  ggplot(aes(game_date, value)) +
  geom_point() +
  stat_smooth(aes(group = Year, color = Year),formula = y~x, method='loess') +
  facet_wrap(~variable, scales = "free_y") +
  ggtitle("Clayton Kershaw: 2015 to 2019") +
  ylab("Angle = Degrees, Speed = MPH") +
```

```
  xlab("Date")

kershaw_data %>%
  ggplot(aes(game_date, launch_angle)) +
  geom_point(alpha = 0.5) +
  stat_smooth(aes(group = Year, color = Year), formula = y~x, method='loess')+
  ggtitle("Clayton Kershaw: Launch Angle") +
  ylab("Angle = Degrees") +
  xlab("Date")

kershaw_data %>%
  ggplot(aes(game_date, launch_speed)) +
  geom_point(alpha = 0.5) +
  stat_smooth(aes(group = Year, color = Year), formula = y~x, method='loess')+
  ggtitle("Clayton Kershaw: Batted Ball Speed") +
  ylab("Speed = MPH") +
  xlab("Date")

# Calculate average horizontal and vertical movement by game date
kershaw_pfx <- kershaw_data %>%
  group_by(game_date) %>%
  summarise(
       `Average Horizontal Movement` = mean(pfx_x, na.rm = TRUE),
       `Average Vertical Movement` = mean(pfx_z, na.rm = TRUE)) %>%
  ungroup() %>%
  melt(id=c("game_date")) %>%
  mutate(Year = as.factor(substr(game_date,1,4)))

# Calculate average horizontal and vertical movement by year
kershaw_avg_pfx <- kershaw_data %>%
  group_by(Year) %>%
  summarise(
       pfx_x = round(mean(pfx_x, na.rm = TRUE),2),
       pfx_z = round(mean(pfx_z, na.rm = TRUE),2))

kershaw_avg_pfx

# Plot horizontal and vertical movement
kershaw_pfx %>%
  ggplot(aes(game_date, value)) +
  geom_point() +
  stat_smooth(aes(group = Year, color = Year),formula = y~x, method='loess') +
  facet_wrap(~variable, scales = "free_y") +
  ggtitle("Clayton Kershaw: Pitch Movement 2015 to 2019") +
  ylab("Horizontal Movement, Vertical Movement") +
  xlab("Date")

kershaw_data %>%
  ggplot(aes(game_date, pfx_x)) +
  geom_point(alpha = 0.5) +
  stat_smooth(aes(group = Year, color = Year), formula = y~x, method='loess')+
  ggtitle("Clayton Kershaw: Horizontal Movement of Pitch") +
  ylab("Horizontal Movement") +
  xlab("Date")
```

```
kershaw_data %>%
  ggplot(aes(game_date, pfx_z)) +
  geom_point(alpha = 0.5) +
  stat_smooth(aes(group = Year, color = Year), formula = y~x, method='loess')+
  ggtitle("Clayton Kershaw: Vertical Movement of Pitch") +
  ylab("Vertical Movement") +
  xlab("Date")
kershaw_speed_acceleration <- kershaw_data %>%
  group_by(game_date) %>%
  summarise(
    `Average Ax` = mean(ax, na.rm = TRUE),
    `Average Ay` = mean(ay, na.rm = TRUE),
    `Average Az` = mean(az, na.rm = TRUE)) %>%
  ungroup() %>%
  melt(id = c("game_date")) %>%
  mutate(Year = as.factor(substr(game_date, 1, 4)))


ggplot(subset(kershaw_speed_acceleration, variable == "Average Ax"), aes(game_date,
value)) +
  geom_point() +
  stat_smooth(aes(color = Year), formula = y ~ x, method = 'loess') +
  ggtitle("Clayton Kershaw: Pitch Speed Acceleration (Ax) 2015 to 2019") +
  ylab("Acceleration Value (Ax)") +
  xlab("Date") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Graph for Ay
ggplot(subset(kershaw_speed_acceleration, variable == "Average Ay"), aes(game_date,
value)) +
  geom_point() +
  stat_smooth(aes(color = Year), formula = y ~ x, method = 'loess') +
  ggtitle("Clayton Kershaw: Pitch Speed Acceleration (Ay) 2015 to 2019") +
  ylab("Acceleration Value (Ay)") +
  xlab("Date") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Graph for Az
ggplot(subset(kershaw_speed_acceleration, variable == "Average Az"), aes(game_date,
value)) +
  geom_point() +
  stat_smooth(aes(color = Year), formula = y ~ x, method = 'loess') +
  ggtitle("Clayton Kershaw: Pitch Speed Acceleration (Az) 2015 to 2019") +
  ylab("Acceleration Value (Az)") +
  xlab("Date") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
# Calculate average velocities by game
Kershaw_X_velocity <- kershaw_data %>%
  group_by(game_date) %>%
  summarise(
      avg_velocity_x = mean(vx0, na.rm = TRUE)) %>%
  ungroup() %>%
  melt(id = c("game_date")) %>%
```

```
  mutate(Year = as.factor(substr(game_date, 1, 4)))

Kershaw_Y_velocity <- kershaw_data %>%
  group_by(game_date) %>%
  summarise(
       avg_velocity_y = mean(vy0, na.rm = TRUE))%>%
  ungroup() %>%
  melt(id = c("game_date")) %>%
  mutate(Year = as.factor(substr(game_date, 1, 4)))

Kershaw_Z_velocity <- kershaw_data %>%
  group_by(game_date) %>%
  summarise(
       avg_velocity_z = mean(vz0, na.rm = TRUE)) %>%
  ungroup() %>%
  melt(id = c("game_date")) %>%
  mutate(Year = as.factor(substr(game_date, 1, 4)))

# Display the first few rows of the reshaped data
head(Kershaw_X_velocity)
head(Kershaw_Y_velocity)
head(Kershaw_Z_velocity)

Kershaw_avg_velocity_yr <- kershaw_data %>%
  group_by(Year) %>%
  summarise(
       avg_velocity_x = mean(vx0, na.rm = TRUE),
       avg_velocity_y = mean(vy0, na.rm = TRUE),
       avg_velocity_Z = mean(vz0, na.rm = TRUE))

Kershaw_avg_velocity_yr
```

```
Kershaw_X_velocity %>%
  ggplot(aes(game_date, value)) +
  geom_point() +
  stat_smooth(aes(group = Year, color = Year),formula = y~x, method='loess') +
  facet_wrap(~variable, scales = "free_y") +
  ggtitle("Kershaw Clayton: Average velocity in X-dimension") +
  ylab("Velocity: feet per second") +
  xlab("Date")

Kershaw_Y_velocity %>%
  ggplot(aes(game_date, value)) +
  geom_point() +
  stat_smooth(aes(group = Year, color = Year),formula = y~x, method='loess') +
  facet_wrap(~variable, scales = "free_y") +
  ggtitle("Kershaw Clayton: Average velocity in Y-dimension") +
  ylab("Velocity: feet per second") +
  xlab("Date")
```

```
Kershaw_Z_velocity %>%
  ggplot(aes(game_date, value)) +
  geom_point() +
  stat_smooth(aes(group = Year, color = Year),formula = y~x, method='loess') +
  facet_wrap(~variable, scales = "free_y") +
  ggtitle("Kershaw Clayton: Average velocity in Z-dimension") +
  ylab("Velocity: feet per second") +
  xlab("Date")



kershaw_data %>%
  ggplot(aes(game_date, vx0)) +
  geom_point(alpha = 0.5) +
  stat_smooth(aes(group = Year, color = Year), formula = y~x, method='loess')+
  ggtitle("Kershaw Clayton: Pitch Velocity in X-Dimension") +
  ylab("Velocity: feet per second") +
  xlab("Date")

kershaw_data %>%
  ggplot(aes(game_date, vy0)) +
  geom_point(alpha = 0.5) +
  stat_smooth(aes(group = Year, color = Year), formula = y~x, method='loess')+
  ggtitle("Kershaw Clayton: Pitch Velocity in Y-Dimension") +
  ylab("Velocity: feet per second") +
  xlab("Date")

kershaw_data %>%
  ggplot(aes(game_date, vz0)) +
  geom_point(alpha = 0.5) +
  stat_smooth(aes(group = Year, color = Year), formula = y~x, method='loess')+
  ggtitle("Kershaw Clayton: Pitch Velocity in Z-Dimension") +
  ylab("Velocity: feet per second") +
  xlab("Date")

# Combine data for all years and apply new filter
kershaw_data <- bind_rows(kershaw_15, kershaw_16, kershaw_17, kershaw_18, kershaw_19)
%>%
  mutate(Year = as.factor(substr(game_date, 1, 4))) %>%
  filter(type == "X") %>%
  filter(launch_speed != 0) %>%
  filter(pitch_type != "")
# Calculate average release and effective speed by game date
kershaw_avgspeeds <- kershaw_data %>%
  group_by(game_date) %>%
  summarise(
       `Average Release Speed` = mean(release_speed, na.rm = TRUE),
       `Average Effective Speed` = mean(effective_speed, na.rm = TRUE)) %>%
  ungroup() %>%
  melt(id=c("game_date")) %>%
  mutate(Year = as.factor(substr(game_date,1,4)))

head(kershaw_avgspeeds)
```

```
kershaw_avg_pitch_speeds <- kershaw_data %>%
  group_by(game_date) %>%
  summarise(
       `Average Release Speed` = mean(release_speed, na.rm = TRUE),
       `Average Effective Speed` = mean(effective_speed, na.rm = TRUE)) %>%
  ungroup() %>%
  melt(id=c("game_date")) %>%
  mutate(Year = as.factor(substr(game_date,1,4)))

# Calculate average release and effective speed speed by year
kershaw_avg_speed_yr <- kershaw_data %>%
  group_by(Year) %>%
  summarise(
       release = round(mean(release_speed, na.rm = TRUE),1),
       efffective = round(mean(effective_speed, na.rm = TRUE),1))

kershaw_avg_speed_yr

kershaw_avg_speed_diff <- kershaw_data %>%
  group_by(Year) %>%
  summarise(
       difference = round(mean(release_speed - effective_speed, na.rm = TRUE),1))

kershaw_avg_speed_diff

# Plot data
kershaw_avgspeeds %>%
  ggplot(aes(game_date, value)) +
  geom_point() +
  stat_smooth(aes(group = Year, color = Year),formula = y~x, method='loess') +
  facet_wrap(~variable, scales = "free_y") +
  ggtitle("Average Pitch Speed by Game from 2015 to 2019") +
  ylab("Speed (MPH)") +
  xlab("Date") + theme(legend.position = "none")

#plot release speed per year
#for adding color to the dots -
https://stackoverflow.com/questions/68457505/ggplot-geom-point-color-based-on-a-variab
le
kershaw_data %>%
  ggplot(aes(game_date, release_speed, color = pitch_type)) +
  geom_point(alpha = 0.7) +
  stat_smooth(aes(group = Year), formula = y~x, method='loess')+
  ggtitle("Clayton Kershaw: Speed of Ball on Pitch Release") +
  ylab("Speed (MPH)") +
  xlab("Date") +
  guides(color = guide_legend(title = "Type of Pitch")) +
  scale_color_manual(labels = c("Changeup", "Curveball", "Fastball", "Sinker",
"Slider"), values = c("#1f77b4", "#ff7f0e", "#2ca02c", "#d62728", "#9467bd"))

#plot effective speed per year
kershaw_data %>%
  ggplot(aes(game_date, effective_speed, color = pitch_type)) +
```

```r
  geom_point(alpha = 0.7) +
  stat_smooth(aes(group = Year), formula = y~x, method='loess')+
  ggtitle("Clayton Kershaw: Ball Speed across Home Plate") +
  ylab("Speed (MPH)") +
  xlab("Date") +
  guides(color = guide_legend(title = "Type of Pitch")) +
  scale_color_manual(labels = c("Changeup", "Curveball", "Fastball", "Sinker",
"Slider"), values = c("#1f77b4", "#ff7f0e", "#2ca02c", "#d62728", "#9467bd"))

# Calculate average horizontal and vertical movement by game date
kershaw_release <- kershaw_data %>%
  group_by(game_date) %>%
  summarise(
       `Average Pitch Release (X-axis)` = mean(release_pos_x, na.rm = TRUE),
       `Average Pitch Release (Y-axis)` = mean(release_pos_y, na.rm = TRUE),
       `Average Pitch Release (Z-axis)` = mean(release_pos_z, na.rm = TRUE)) %>%
  ungroup() %>%
  melt(id=c("game_date")) %>%
  mutate(Year = as.factor(substr(game_date,1,4)))

# Calculate average horizontal and vertical movement by year
kershaw_avg_release <- kershaw_data %>%
  group_by(Year) %>%
  summarise(
       x = round(mean(release_pos_x, na.rm = TRUE),2),
       y = round(mean(release_pos_y, na.rm = TRUE),2),
       z = round(mean(release_pos_z, na.rm = TRUE),2))

kershaw_avg_release


#Plot release positions
kershaw_data %>%
  ggplot(aes(game_date, release_pos_x)) +
  geom_point(alpha = 0.5) +
  stat_smooth(aes(group = Year, color = Year), formula = y~x, method='loess')+
  ggtitle("Clayton Kershaw: X Position of Pitch Start") +
  ylab("Feet") +
  xlab("Date")

kershaw_data %>%
  ggplot(aes(game_date, release_pos_y)) +
  geom_point(alpha = 0.5) +
  stat_smooth(aes(group = Year, color = Year), formula = y~x, method='loess')+
  ggtitle("Clayton Kershaw: Y Position of Pitch Start") +
  ylab("Feet") +
  xlab("Date")

kershaw_data %>%
  ggplot(aes(game_date, release_pos_z)) +
  geom_point(alpha = 0.5) +
  stat_smooth(aes(group = Year, color = Year), formula = y~x, method='loess')+
  ggtitle("Clayton Kershaw: Z Position of Pitch Start") +
  ylab("Feet") +
```

```
  xlab("Date")

### 3D graph for starting pitches
library(plotly)

#https://stackoverflow.com/questions/5890584/how-to-reshape-data-from-long-to-wide-for
mat

dimensional_data <- reshape(kershaw_release, idvar = "game_date", timevar =
"variable", direction = "wide")[-c(3,5)]
colnames(dimensional_data) <- c("game_date", "x","y","z","Year")


# Generate a 3D scatter plot with color grouping for averages
plot_ly(dimensional_data, x = ~x, y = ~y, z = ~z, color = ~Year,
             type = "scatter3d", mode = "markers", alpha = 0.95, size = 2)

# Generate a 3D scatter plot with color grouping for all pitches
plot_ly(kershaw_data, x = ~release_pos_x, y = ~release_pos_y, z = ~release_pos_z,
color = ~Year, type = "scatter3d", mode = "markers", alpha = 0.8, size = 1) %>%
layout(title = "Clayton Kershaw: Starting Position of Pitches",scene = list(
             xaxis = list(title = "X"),
             yaxis = list(title = "Y"),
             zaxis = list(title = "Z")))
#create dataframe with coutns for ball, strikes, and hits
kershaw15_results = table(unlist(kershaw_15$type))
rownames(kershaw15_results) = c("Ball", "Strike", "Hit")
kershaw15_results["Year"] = 2015

kershaw16_results = table(unlist(kershaw_16$type))
rownames(kershaw16_results) = c("Ball", "Strike", "Hit")
kershaw16_results["Year"] = 2016

kershaw17_results = table(unlist(kershaw_17$type))
rownames(kershaw17_results) = c("Ball", "Strike", "Hit")
kershaw17_results["Year"] = 2017

kershaw18_results = table(unlist(kershaw_18$type))
rownames(kershaw18_results) = c("Ball", "Strike", "Hit")
kershaw18_results["Year"] = 2018

kershaw19_results = table(unlist(kershaw_19$type))
rownames(kershaw19_results) = c("Ball", "Strike", "Hit")
kershaw19_results["Year"] = 2019
#convert tables into dataframe and reshape from wide to long
pitch_results =
data.frame(rbind(kershaw15_results,kershaw16_results,kershaw17_results,kershaw18_resul
ts,kershaw19_results))
pitch_results = reshape(pitch_results, direction = "long", varying =
list(names(pitch_results)[1:3]), v.names = "Count", idvar = c("Year"), timevar =
"Type", times = c("Ball", "Strike", "Hit"))

#plot data as stacked bar plot
```

```
ggplot(data = pitch_results, aes(fill = Type, y = Count, x = Year)) +
geom_bar(position="fill", stat="identity") + ylab("Proportion of Pitches with Specific
Outcome") + guides(fill=guide_legend(title="Outcome of Pitch"))
#calculate the percent of events that will be removed by only looking at
homerun/walk/strikeout/fieldout/batter makes it to base
total_removed =
sum(table(unlist(kershaw_15$events))[2:length(table(unlist(kershaw_15$events)))])-sum(
table(unlist(kershaw_15$events))[c("home_run", "walk", "strikeout", "field_out",
"double", "single", "triple")]) +
sum(table(unlist(kershaw_16$events))[2:length(table(unlist(kershaw_16$events)))])-sum(
table(unlist(kershaw_16$events))[c("home_run", "walk", "strikeout", "field_out",
"double", "single", "triple")]) +
sum(table(unlist(kershaw_17$events))[2:length(table(unlist(kershaw_17$events)))])-sum(
table(unlist(kershaw_17$events))[c("home_run", "walk", "strikeout", "field_out",
"double", "single", "triple")]) +
sum(table(unlist(kershaw_18$events))[2:length(table(unlist(kershaw_18$events)))])-sum(
table(unlist(kershaw_18$events))[c("home_run", "walk", "strikeout", "field_out",
"double", "single", "triple")]) +
sum(table(unlist(kershaw_19$events))[2:length(table(unlist(kershaw_19$events)))])-sum(
table(unlist(kershaw_19$events))[c("home_run", "walk", "strikeout", "field_out",
"double", "single", "triple")])

total =
sum(table(unlist(kershaw_15$events))[2:length(table(unlist(kershaw_15$events)))])+sum(
table(unlist(kershaw_16$events))[2:length(table(unlist(kershaw_16$events)))])+sum(tabl
e(unlist(kershaw_17$events))[2:length(table(unlist(kershaw_17$events)))])+sum(table(un
list(kershaw_18$events))[2:length(table(unlist(kershaw_18$events)))])+sum(table(unlist
(kershaw_19$events))[2:length(table(unlist(kershaw_19$events)))])

percent_removed = total_removed/total
percent_removed



#create extract the data for events and join double/single/triple into one column;
keep the original values as a different table
kershaw15_events = table(unlist(kershaw_15$events))[c("home_run", "walk", "strikeout",
"field_out", "double", "single", "triple")]
kershaw15_bases = kershaw15_events[5:7]
kershaw15_events = kershaw15_events[1:4]
kershaw15_events["on_base"] = sum(kershaw15_bases)
kershaw15_events["Year"] = 2015
kershaw15_bases["Year"] = 2015

kershaw16_events = table(unlist(kershaw_16$events))[c("home_run", "walk", "strikeout",
"field_out", "double", "single", "triple")]
kershaw16_bases = kershaw16_events[5:7]
kershaw16_events = kershaw16_events[1:4]
kershaw16_events["on_base"] = sum(kershaw16_bases)
kershaw16_events["Year"] = 2016
kershaw16_bases["Year"] = 2016

kershaw17_events = table(unlist(kershaw_17$events))[c("home_run", "walk", "strikeout",
"field_out", "double", "single", "triple")]
```

```
kershaw17_bases = kershaw17_events[5:7]
kershaw17_events = kershaw17_events[1:4]
kershaw17_events["on_base"] = sum(kershaw17_bases)
kershaw17_events["Year"] = 2017
kershaw17_bases["Year"] = 2017

kershaw18_events = table(unlist(kershaw_18$events))[c("home_run", "walk", "strikeout",
"field_out", "double", "single", "triple")]
kershaw18_bases = kershaw18_events[5:7]
kershaw18_events = kershaw18_events[1:4]
kershaw18_events["on_base"] = sum(kershaw18_bases)
kershaw18_events["Year"] = 2018
kershaw18_bases["Year"] = 2018

kershaw19_events = table(unlist(kershaw_19$events))[c("home_run", "walk", "strikeout",
"field_out", "double", "single", "triple")]
kershaw19_bases = kershaw19_events[5:7]
kershaw19_events = kershaw19_events[1:4]
kershaw19_events["on_base"] = sum(kershaw19_bases)
kershaw19_events["Year"] = 2019
kershaw19_bases["Year"] = 2019

#convert the tables into dataframes, and then convert the dataframes from wide to long
pitch_events =
data.frame(rbind(kershaw15_events,kershaw16_events,kershaw17_events,kershaw18_events,k
ershaw19_events))
pitch_events = reshape(pitch_events, direction = "long", varying =
list(names(pitch_events)[1:5]), v.names = "Count", idvar = c("Year"), timevar =
"Type", times = c("Home Run", "Walk", "Strikeout", "Field Out", "On Base"))

pitch_bases =
data.frame(rbind(kershaw15_bases,kershaw16_bases,kershaw17_bases,kershaw18_bases,kersh
aw19_bases))
pitch_bases = reshape(pitch_bases, direction = "long", varying =
list(names(pitch_bases)[1:3]), v.names = "Count", idvar = c("Year"), timevar = "Type",
times = c("Single", "Double", "Triple"))

#plot the data
ggplot(data = pitch_events, aes(fill = factor(Type,rev(c("Strikeout", "Field Out", "On
Base", "Walk", "Home Run"))), y = Count, x = Year)) + geom_bar(position="fill",
stat="identity") + ylab("Proportion of Times an Event Occured") +
guides(fill=guide_legend(title="Event"))


ggplot(data = pitch_bases, aes(fill = factor(Type, c("Triple", "Double","Single")), y
= Count, x = Year)) + geom_bar(position="fill", stat="identity") + ylab("Proportion of
Times an Event Occured") + guides(fill=guide_legend(title="Event"))

#calculate overall average percentages for each event
events = kershaw15_events +  kershaw16_events +  kershaw17_events + kershaw18_events +
kershaw19_events
events = events[1:5]
event_total = sum(events)
paste0("Average Homeruns:",events["home_run"]/event_total)
```

```
paste0("Average Walks:",events["walk"]/event_total)
paste0("Average Strikeouts:",events["strikeout"]/event_total)
paste0("Average Field outs:",events["field_out"]/event_total)
paste0("Average On Base:",events["on_base"]/event_total)
```