Trevor Story's June 2025 Resurgence: A Data-Driven Look

Jackson Leary

July 2025

1 Introduction

If you have paid attention to the MLB in the last decade or so, you are more than likely aware of who Trevor Story is. A two-time Silver Slugger Award winner and two-time All-Star during his tenure on the Colorado Rockies, Story was a force to be reckoned with in Denver. In March 2022, however, Story signed a massive six-year $140 million deal with the Boston Red Sox, where he is currently playing his fourth season as a member of the team.

So far, Story's time in Boston has been plagued by injuries, which has contributed to what many fans view as underwhelming, and undeserving of his current contract. Story started off the 2025 season hot, before his numbers started to slowly decline through the months of April and May. But just when trade rumors started to circle and fans felt like giving up on him, a switch flipped. Story's performance since the start of June has been nothing short of extraordinary, and his hitting statistics have quickly risen. As a Red Sox fan myself, watching this resurgence made me want to take a deeper look at Story's season thus far. I will aim to investigate whether Story's performance after the month of May significantly improved, and what might have caused it.

2 Data and Methodology

2.1 Data Collection

For this project, I used RStudio to obtain Statcast data for Trevor Story's 2025 season using the `baseballr` package, which can pull stats from MLB's public Statcast database. The dataset that I will be working with includes

all plate appearances from March 27 through July 13, 2025, Boston's final game before the All-Star break. Event outcomes (e.g., single, home run, strikeout), batted ball characteristics (e.g., exit velocity, launch angle), pitch types, and other variables are included.
To explore my question, I chose to divide the data into two different periods:

- **Before:** Games played before June 1, 2025
- **After:** Games played on or after June 1, 2025

This cutoff was chosen as a result of both my (and others') observations. And while it may seem somewhat arbitrary, which in some ways it is, it also allows us to observe whether there was any notable change in performance around midway through the first half of the season.

2.2 Variables

There were 10 different metrics that I chose to use in order to evaluate different aspects of hitting:

2.2.1 Traditional Metrics

- **AVG:** Batting average (Hits / At-Bats)
- **SLG:** Slugging percentage (Total Bases / At-Bats)
- **OPS:** On-base plus slugging (OBP + SLG)
- **OPS+:** OPS scale with respect to the league average OPS (normalized, was not tested)
- **BABIP:** Batting average on balls in play
- **K% / BB%**: Strikeout and base-on-balls (walk) percentage

2.2.2 Statcast Metrics

- **EV:** Average exit velocity (mph)
- **LA:** Average launch angle (degrees)
- **HardHit%:** % batted balls hit $\geq$ 95 mph

Note that only plate appearances with valid event outcomes were used in the calculations for traditional metrics, and only batted balls were used for Statcast metrics.

2.3 Statistical Tests

In order to determine whether there were significant differences in performance between the two periods, different statistical tests were used depending on the type of statistic being evaluated:

- **Proportion tests** (`prop.test`) were used for proportion or rate-based statistics such as AVG, BABIP, HardHit%, K%, and BB%. This was done using the raw counts, such as H and AB.
- **Two-sample t-tests** (`t.test`) were used for continuous (and normally distributed) variables, specifically EV and LA, which were calculated at the game level. The tests were performed after aggregating the data by game date to ensure that there was a sufficient number of observations in each group for analysis.
- Graphical observations as well as a Shapiro-Wilk test indicated that both SLG and OPS were not normally distributed ($p < 0.05$). For these variables, a Wilcoxon Rank-Sum test was used.
- For variables with only one observation per period, specifically OPS+, no statistical test was performed, but rather differences were descriptively noted.

The analyzes were performed using the following R packages: `baseballr`, `dplyr`, `ggplot2`, and `janitor`.

2.4 Pitch Types

To further analyze what may be causing Story's improvement, I looked at different pitch types and grouped them into the following categories:

- **4-Seam Fastball**
- **Curveball:** Includes both Curveball and Knuckle Curve
- **Offspeed:** Includes Changeup and Split-Finger
- **Sinker**
- **Cutter**
- **Slider**

- **Sweeper**

Pitch types with limited usage (e.g., Eephus, Slurve, Forkball) were excluded from the data before grouping, allowing me to assess Story's performance against each type.

3 Results

3.1 Overview of Offensive Performance Before and After June 1

After performing the relevant tests on the data, the following results were obtained (specific statistics and p-values will be provided in a table below):

3.1.1 Traditional Offensive Performance

Analyzing batting average (AVG) revealed a significant increase from the pre- to post-June period (0.216 to 0.319). A two-sample proportion test resulted in a p-value that was significant ($p < 0.05$). Similarly, slugging percentage (SLG) and on-base plus slugging (OPS) appear to have increased post-June, though the differences according to the Wilcoxon Rank-Sum tests appear to be marginally significant ($0.05 < p < 0.1$).

The rest of the traditional metrics (BABIP, K%, BB%), while all showing improvement, did not reveal any significant changes between the two periods ($p > 0.1$).

Because OPS+ is only available as a single value per period, no formal statistical tests could be performed. However, an approximately 129% increase in OPS+ was observed.

3.1.2 Statcast Performance Evaluation

Average exit velocity (EV) showed a marginally significant increase in the second period ($0.05 < p < 0.1$), indicating a slight improvement in quality of contact. However, both launch angle (LA) and HardHit% exhibited no significant change between the two periods ($p > 0.1$).

Metric	Pre-June Value	Post-June Value	P-value
AVG	0.216	0.319	0.037*
SLG	0.326	0.556	0.057†
OPS	0.576	0.911	0.064†
OPS+	79	181	N/A
BABIP	0.288	0.376	0.191
K%	0.316	0.230	0.111
BB%	0.044	0.053	0.853
EV	83.2	85.6	0.069†
LA	19.1	21.1	0.434
HardHit%	0.469	0.486	0.884

Table 1: Table summarizing Trevor Story's different offensive metrics, pre-June and post-June values, and p-values. Statistical significance is as follows: $0.05 < p < 0.1(^{\dagger}), \quad p < 0.05(^{*})$

.

3.1.3 Metrics Figures and Table

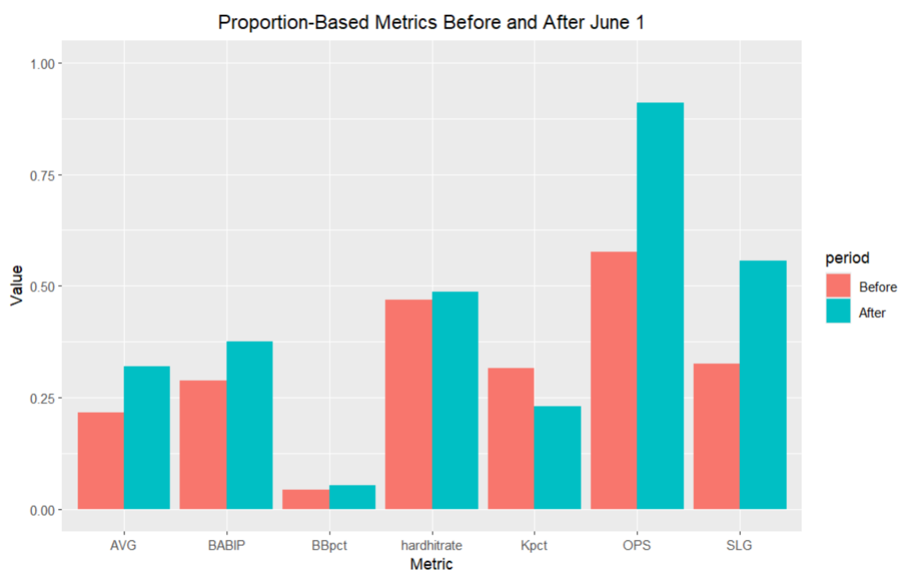

Figure 1: Trevor Story's proportion-based metrics before and after June 1.

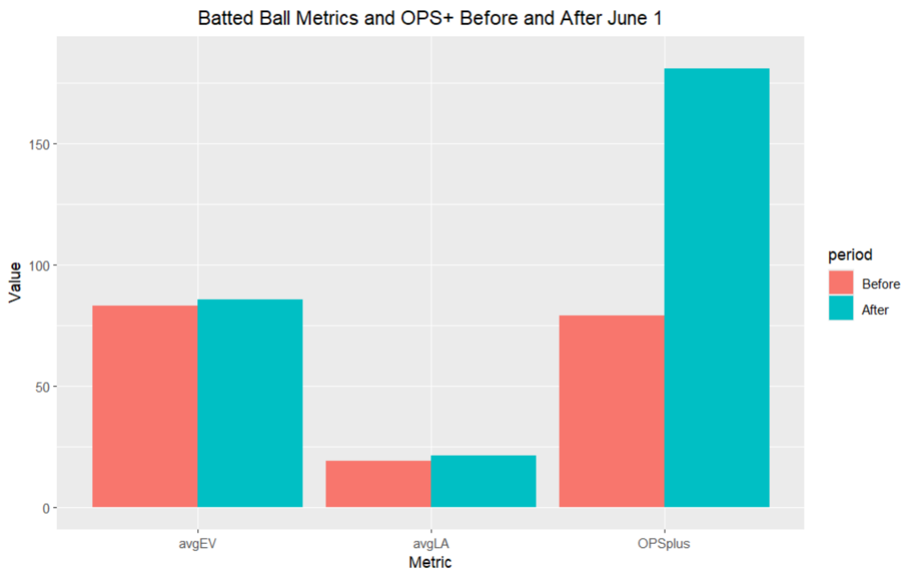

Figure 2: Trevor Story's batted ball metrics and OPS+ before and after June 1.

3.2 Performance Against Different Pitch Types

To determine whether Story's improvement was a result of improved ability to hit certain pitch types, I compared his offensive metrics across multiple pitch categories, before and after June 1. A table and bar plot summarizing the results is provided below:

Pitch Group	Period	# Pitches	AVG	SLG	HardHit%	EV	LA
4-Seam Fastball	Before	326	0.143	0.157	0.18	82.3	31.3
	After	173	0.326	0.442	0.29	86.4	31.8
Curveball	Before	75	0.294	0.471	0.21	81.8	16.6
	After	73	0.056	0.056	0.33	83.9	9.4
Slider	Before	150	0.250	0.444	0.25	83.1	10.1
	After	84	0.318	0.864	0.35	89.4	12.4
Offspeed	Before	94	0.222	0.444	0.32	84.0	6.2
	After	46	0.286	0.429	0.25	84.3	7.9
Sinker	Before	130	0.226	0.226	0.38	85.7	5.8
	After	98	0.481	0.778	0.37	85.4	17.1
Cutter	Before	46	0.300	0.300	0.29	83.9	31.8
	After	33	0.455	0.545	0.31	86.7	15.2
Sweeper	Before	75	0.211	0.526	0.19	81.1	18.4
	After	26	0.333	1.333	0.29	84.7	52.57

Table 2: Trevor Story's offensive statistics, and number of each pitch type faced, Before and After June 1.

Figure 3: Trevor Story's batting average vs. major pitch types, compared before and after June 1.

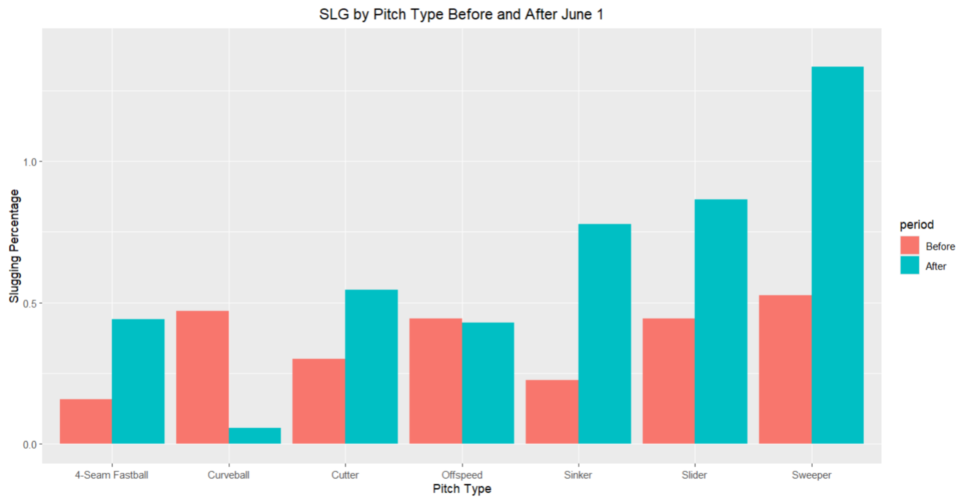

Figure 4: Trevor Story's slugging percentage vs. major pitch types, compared before and after June 1.

4 Discussion

4.1 Key Findings

When analyzing Story's statistics and metrics, there appears to have been an offensive resurgence starting in June 2025. Traditional offensive metrics such as AVG, SLG, and OPS all exhibit notable improvements, with AVG being statistically significant and SLG and OPS on the verge of significance. These findings suggest that there is some evidence to support that this improvement is not merely random, but rather that it is reflective of a shift in performance.

While the Statcast data did not show any large changes in LA or HardHit%, EV showed a marginally significant improvement. This suggests that Story's quality of contact might be improving. A lack of significance in LA also implies that his batted ball profile, and likely his swing path, has not changed much.

Though looking purely at the offensive metrics helps provide insight into whether Story has improved, analyzing these metrics with regards to different pitch types allowed us to gain more meaningful insight. Major improvements appear against just about every pitch type except for curveballs, especially 4-seam fastballs, sliders, and sinkers, which are some of the most common pitch types thrown in modern baseball. Both his AVG and SLG against 4-seams more than doubled, with increases in HardHit% and EV as well. Against sliders, his SLG jumped from 0.444 to 0.864, which is especially remarkable.

While most of his stats showed improvement, there appears to have been a sharp decline in performance against curveballs. His AVG dropped from 0.294 to 0.056, and his SLG went from 0.471 to 0.056 as well, both of which are especially poor. While this could be due to a small sample size, pitchers throwing more deceptive curveballs, or an improvement in location by pitchers, I believe that this decline, as well as his improvements against other pitches, can be explained by one simple thing.

4.2 My Theory for What's Behind Story's Improvement

"Dead red" is a term that refers to a hitter anticipating a fastball before the pitch is thrown and trying to attack that pitch. I believe that much of Story's success can be explained by a shift in his hitting strategy to a more "dead red" approach. His increased performance against the three different types of fastballs (4-seam, cutter, and sinker) are likely a result of an increased aggression towards higher-velocity pitches early in the count. Additionally, curveballs tend to be the slowest pitch type in most pitchers arsenals, slower than sliders and most offspeed pitches. Story's decline in his ability to hit pitchers' slowest pitch further suggests a possible shift in his attack strategy, and an increase in the frequency of curveballs thrown to Story suggests other pitchers may be catching on to this weakness. If you are trying to target higher-velocity pitches, you are more prone to being early on and out in front of slower-velocity breaking balls.

4.3 Limitations

My analysis has many different areas with room for error. Arguably the largest one is the relatively short time span for the period after June 1. While the differences I observed in my analysis are certainly compelling, a longer time frame would strengthen my confidence in these results and my interpretations of them. Additionally, other contextual variables such as opposing pitchers (especially pitcher handedness), ballpark effects (not included in my OPS+ calculations), and more, were not accounted for in my research, and may have also influenced my results.

Nevertheless, these findings suggest that Story's improvement is real, and is largely driven by an improved approach against key pitch types. Whether other factors such as improved pitch recognition, or simply the fact that he is fully healthy again played a role, is beyond the scope of my analysis. However, it would certainly warrant future investigation.

5 Conclusion

Trevor Story's offensive performance since June 2025 is something to appreciate, especially when looking at his performance earlier in the season, as well as previous seasons with Boston. This report looked at a combination of both traditional and advanced metrics, and found significant improvement in batting average, as well as a strong upwards trend in, slugging percentage, OPS, and average exit velocity.

Story's improvements in performance were especially noticeable against different types of fastballs, and sliders, all pitches which he seemed to struggle with in the earlier parts of the season. This suggests an improved offensive approach, though whether it is mechanical, mental, or a combination of both, remains to be seen.

While these results are promising, continued monitoring to see if this success can be sustained over a longer period of time will provide more confidence in Story's return to his former Colorado self.

Ultimately, this study demonstrates the value of advanced pitch tracking data in evaluating baseball players' performances, and also highlights the culture shift in Major League Baseball to a more data-driven decision-making process. Story's June turnaround offers not only hope for loyal Red Sox fans, but also gives a reminder as to how advanced metrics can reveal the hidden factors behind many parts of baseball, as the sport continues to enter a new era.

Appendix A: R Code Repository Link

The full R code used for data collection, manipulation, analysis, and visualization can be accessed at:
`https://github.com/Jaxleary/Trevor-Story-Analysis`

A static version is available in Appendix B as well.

Appendix B: R Code

The entirety of my R code is available starting on the next page.

```
# Install/load necessary packages
  library(baseballr)
  library(dplyr)
  library(lubridate)
  library(janitor)
  library(ggplot2)
  library(devtools)
  library(tidyr)
  library(patchwork)
  library(rvest)
  devtools::install_github("BillPetti/baseballr")


# Compile and format Trevor Story's data for analysis


  ## Retrieve statcast data
      storyID <- 596115
      roughstoryData <- statcast_search(
        start_date = "2025-03-27",
        end_date = "2025-07-15",
        playerid = storyID,
        player_type = "batter"
      )

  ## Split data up into pre-/post-June 1st
      storyData <- roughstoryData %>%
        mutate(period = if_else(game_date < as.Date("2025-06-01"), "Before", "After"))


  ## First, look at standard batting stats (AVG, SLG, OPS, OPS+, etc.)
      ### Count hits, total bases, at bats, etc. to calculate statistics after filtering non-
outcome pitches
        storyBatting <- storyData %>%
          filter(!is.na(events)) %>%
          mutate(
            hit = events %in% c("single", "double", "triple", "home_run"),
            bases = case_when(
              events == "single" ~ 1,
              events == "double" ~ 2,
              events == "triple" ~ 3,
              events == "home_run" ~ 4,
              .default = 0
            ),
            atbat = events %in% c("single", "double", "home_run", "strikeout", "field_out",
                                  "force_out", "double_play", "field_error",
"grounded_into_double_play"),
            plateapp = events %in% c("single", "double", "home_run", "strikeout", "field_out",
                                     "force_out", "double_play", "field_error",
"grounded_into_double_play", "walk",
                                     "hit_by_pitch", "sac_bunt"),
            walk = events == "walk",
            homerun = events == "home_run",
            k = events == "strikeout"
          ) %>%
          group_by(period) %>%
          summarize(
            H = sum(hit, na.rm = TRUE),
            TB = sum(bases, na.rm = TRUE),
            AB = sum(atbat, na.rm = TRUE),
            PA = sum(plateapp, na.rm = TRUE),
            BB = sum(walk, na.rm = TRUE),
            HR = sum(homerun, na.rm = TRUE),
            K = sum(k, na.rm = TRUE),
```

```
          SF = 0,
          AVG = H/AB,
          SLG = TB/AB,
          OBP = (H+BB)/(AB+BB), # Using simple formula, does not include HBP and sac
          OPS = OBP+SLG,
          BABIP = (H-HR)/(AB-HR-K+SF),
          Kpct = K/(AB+BB),
          BBpct = BB/(AB+BB),
        )

      ### Calculate OPS+ (simplified, not adjusted for park effects)
        leaguestats <- fg_bat_leaders(
          startseason = 2025,
          endseason = 2025,
          stats = "bat",
          ind = 1,
          qual = 1,   #### Stats only for batters with at least 1 plate appearance
          pos = "np"
        )
        leagueOBP = mean(leaguestats$OBP, na.rm = TRUE)
        leagueSLG = mean(leaguestats$SLG, na.rm = TRUE)
        storyBatting <- storyBatting %>%
          mutate(OPSplus = round(100*(((storyBatting$OBP/leagueOBP)+
(storyBatting$SLG/leagueSLG))-1)))

  ## Obtain statcast metrics
    storyStatcast <- storyData %>%
      filter(!is.na(launch_speed)) %>%
      mutate(
        hardhit = launch_speed >= 95,
        battedball = events %in% c("single", "double", "triple", "home_run", "groundout",
                                   "flyout", "field_out","force_out", "double_play",
                                   "field_error", "grounded_into_double_play")
      ) %>%
      group_by(period) %>%
      summarize(
        avgEV = mean(launch_speed, na.rm = TRUE),
        avgLA = mean(launch_angle, na.rm = TRUE),
        battedballsnum = sum(battedball, na.rm = TRUE),
        hardhitnum = sum(hardhit & battedball, na.rm = TRUE),
        hardhitrate = hardhitnum/battedballsnum
      )

  ## Compile all data into one set
    storyTotals <- storyBatting %>%
      left_join(storyStatcast, by = "period") %>%
      arrange(desc(period))



  ## Create bar plots to visualize differences in data
    storysmalltotals <- storyTotals %>%
      select(period, AVG, SLG, OPS, BABIP, hardhitrate, Kpct, BBpct) %>%
      pivot_longer(-period, names_to = "Metric", values_to = "Value") %>%
      mutate(period = factor(period, levels = c("Before", "After"))) %>%
      ggplot(aes(x = Metric, y = Value, fill = period)) +
      ggtitle("Proportion-Based Metrics Before and After June 1") +
      theme(plot.title = element_text(hjust = 0.5)) +
      geom_col(position = "dodge") +
      scale_y_continuous(limits = c(0,1))
    storybigtotals <- storyTotals %>%
      select(period, OPSplus, avgEV, avgLA) %>%
      pivot_longer(-period, names_to = "Metric", values_to = "Value") %>%
      mutate(period = factor(period, levels = c("Before", "After"))) %>%
      ggplot(aes(x = Metric, y = Value, fill = period)) +
```

```
      ggtitle("Batted Ball Metrics and OPS+ Before and After June 1") +
      theme(plot.title = element_text(hjust = 0.5)) +
      geom_col(position = "dodge") +
      scale_y_continuous(limits = c(0,185))
    storysmalltotals
    storybigtotals



# Analyze compiled data using relevant statistical tests and analytical methods to answer research
questions

    storyGamestats <- storyData %>%    ### Compiles data and groups by game to be able to perform
t-tests
      filter(!is.na(events)) %>%
      mutate(
        hit = events %in% c("single", "double", "triple", "home_run"),
        bases = case_when(
          events == "single" ~ 1,
          events == "double" ~ 2,
          events == "triple" ~ 3,
          events == "home_run" ~ 4,
          .default = 0
        ),
        atbat = events %in% c("single", "double", "triple", "home_run", "groundout", "strikeout",
"flyout", "field_out",
                              "force_out", "double_play", "field_error",
"grounded_into_double_play"),
        walk = events == "walk",
        homerun = events == "home_run",
        k = events == "strikeout"
      ) %>%
      group_by(game_date, period) %>%
      summarize(
        avgEV = mean(launch_speed, na.rm = TRUE),
        avgLA = mean(launch_angle, na.rm = TRUE),
        H = sum(hit, na.rm = TRUE),
        TB = sum(bases, na.rm = TRUE),
        AB = sum(atbat, na.rm = TRUE),
        BB = sum(walk, na.rm = TRUE),
        HR = sum(homerun, na.rm = TRUE),
        K = sum(k, na.rm = TRUE),
        SF = 0,
        AVG = H/AB,
        SLG = TB/AB,
        OBP = (H+BB)/(AB+BB), # Using simple formula, does not include HBP and sac
        OPS = OBP+SLG,
        BABIP = (H-HR)/(AB-HR-K+SF),
        Kpct = K/(AB+BB),
        BBpct = BB/(AB+BB),
        .groups = "drop"
      ) %>%
      filter(!is.na(OPS))


  ## Question: Is Story's improvement since June 1st real and significant?

    ### Perform different tests and collect results into data frame
      propResults <- data.frame(
        Metric = c("AVG", "SLG", "OPS", "BABIP", "EV", "LA", "HardHit%", "K%", "BB%"),
        P_Value = c(
          prop.test(storyBatting$H, storyBatting$AB, alternative = "two.sided")$p.value,
          wilcox.test(SLG ~ period, data = storyGamestats)$p.value,
          wilcox.test(OPS ~ period, data = storyGamestats)$p.value,
          prop.test((storyBatting$H-storyBatting$HR), (storyBatting$AB-storyBatting$K-
```

```
storyBatting$HR))$p.value,
          t.test(launch_speed ~ period, data = storyData)$p.value,
          t.test(launch_angle ~ period, data = storyData)$p.value,
          prop.test(storyTotals$hardhitnum, storyTotals$battedballsnum)$p.value,
          prop.test(storyTotals$K, storyTotals$PA)$p.value,
          prop.test(storyTotals$BB, storyTotals$PA)$p.value

        )
      )




  ## Question: Is he doing more damage vs certain pitch types now vs before?

    ### Extract data against different pitch types
      storyPitchStats <- storyData %>%
        mutate(
          pitch_group = case_when(
            pitch_name == "4-Seam Fastball" ~ "4-Seam Fastball",
            pitch_name %in% c("Curveball", "Knuckle Curve") ~ "Curveball",
            pitch_name %in% c("Changeup", "Split-Finger") ~ "Offspeed",
            pitch_name == "Sinker" ~ "Sinker",
            pitch_name == "Cutter" ~ "Cutter",
            pitch_name == "Slider" ~ "Slider",
            pitch_name == "Sweeper" ~ "Sweeper",
            TRUE ~ NA_character_
          )
        ) %>%
        filter(!is.na(pitch_group), !is.na(pitch_name), !is.na(events)) %>%
        mutate(
          hit = events %in% c("single", "double", "triple", "home_run"),
          bases = case_when(
            events == "single" ~ 1,
            events == "double" ~ 2,
            events == "triple" ~ 3,
            events == "home_run" ~ 4,
            .default = 0
          ),
          atbat = events %in% c("single", "double", "home_run", "strikeout", "field_out",
                               "force_out", "double_play", "field_error",
"grounded_into_double_play"),
          plateapp = events %in% c("single", "double", "home_run", "strikeout", "field_out",
                                  "force_out", "double_play", "field_error",
"grounded_into_double_play", "walk",
                                  "hit_by_pitch", "sac_bunt"),
          walk = events == "walk",
          homerun = events == "home_run",
          k = events == "strikeout"
        ) %>%
        group_by(period, pitch_group) %>%
        summarize(
          Pitches = n(),
          avgEV = mean(launch_speed, na.rm = TRUE),
          avgLA = mean(launch_angle, na.rm = TRUE),
          hardhitrate = mean(launch_speed >= 95, na.rm = TRUE),
          H = sum(hit, na.rm = TRUE),
          TB = sum(bases, na.rm = TRUE),
          AB = sum(atbat, na.rm = TRUE),
          BB = sum(walk, na.rm = TRUE),
          HR = sum(homerun, na.rm = TRUE),
          K = sum(k, na.rm = TRUE),
          SF = 0,
          AVG = H / AB,
```

```
      SLG = TB / AB,
      OBP = (H + BB) / (AB + BB),  # Approximate OBP (not including HBP or sac flies)
      OPS = OBP + SLG,
      BABIP = (H - HR) / (AB - HR - K + SF),
      Kpct = K / (AB + BB),
      BBpct = BB / (AB + BB),
      .groups = "drop"
    )




### Create bar plots to visualize performance vs. different pitch types

  #### Plot for AVG
  storyPitchAvgplot <- storyPitchStats %>%
    mutate(period = factor(period, levels = c("Before", "After"))) %>%
    ggplot(aes(x = pitch_group, y = AVG, fill = period)) +
    geom_col(position = "dodge") +
    ggtitle("AVG by Pitch Type Before and After June 1") +
    theme(plot.title = element_text(hjust = 0.5)) +
    labs(x = "Pitch Type", y = "Batting Average") +
    scale_y_continuous(limits = c(0, .6))
storyPitchAvgplot

#### Plot for SLG
storyPitchSlgplot <- storyPitchStats %>%
  mutate(period = factor(period, levels = c("Before", "After"))) %>%
  ggplot(aes(x = pitch_group, y = SLG, fill = period)) +
  geom_col(position = "dodge") +
  ggtitle("SLG by Pitch Type Before and After June 1") +
  theme(plot.title = element_text(hjust = 0.5)) +
  labs(x = "Pitch Type", y = "Slugging Percentage") +
  scale_y_continuous(limits = c(0, 1.4))
storyPitchSlgplot

#### Plot for HardHit%
storyPitchHHplot <- storyPitchStats %>%
  mutate(period = factor(period, levels = c("Before", "After"))) %>%
  ggplot(aes(x = pitch_group, y = hardhitrate, fill = period)) +
  geom_col(position = "dodge") +
  ggtitle("Hard Hit Rate by Pitch Type Before and After June 1") +
  theme(plot.title = element_text(hjust = 0.5)) +
  labs(x = "Pitch Type", y = "Hard Hit Rate") +
  scale_y_continuous(limits = c(0, 0.5))
storyPitchHHplot
```