

Final Report

Nathan Butler

2025-05-15

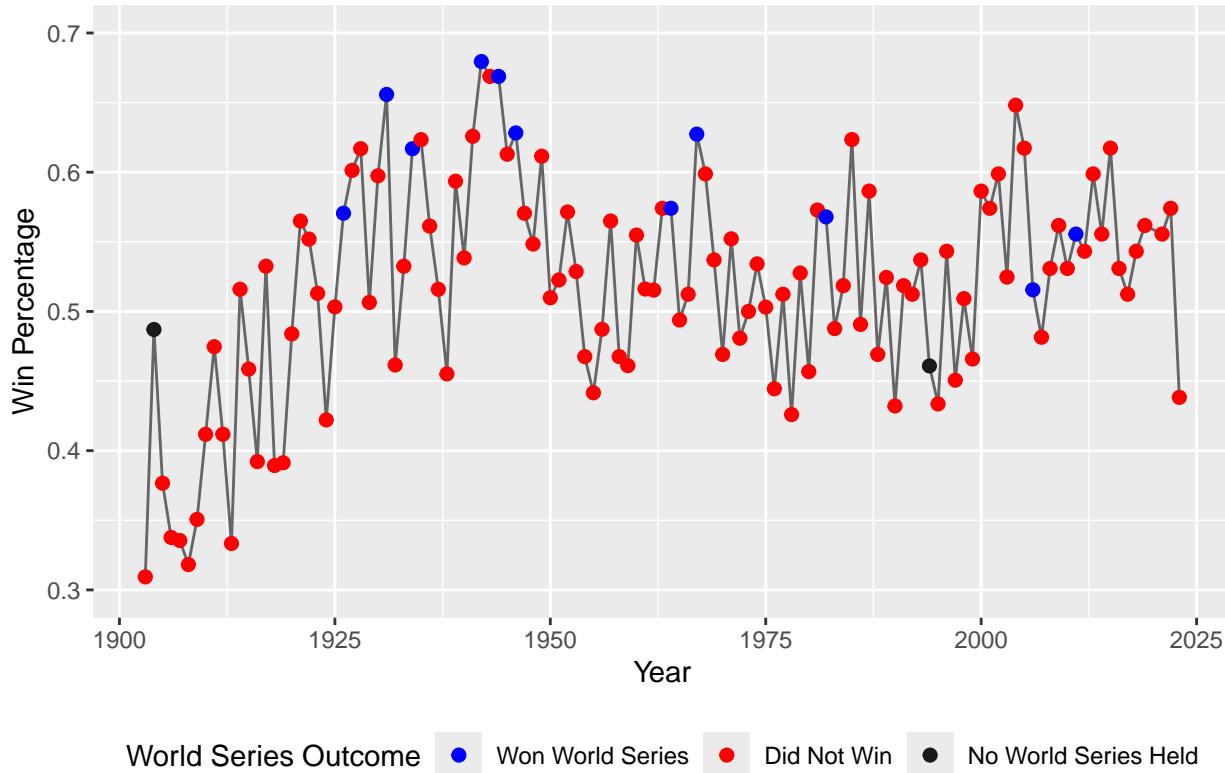
```
# Load Required Libraries
library(Lahman)
library(tidyverse)
library(readr)
```

Presentation can be found here https://drive.google.com/file/d/1dggvbSPEj7rPAOu4CkwCW_2JyYOacyvv/view?usp=drive_link

Background

The St. Louis Cardinals have historically been a team that is at the top of the standings. They are always competitive and have a large list of Hall of Famers that have played for them. Cardinals have the second most World Series Wins (11) among all teams, except for the Yankees and have dominated in the division. For example here is a plot that shows all of their wins by the year.

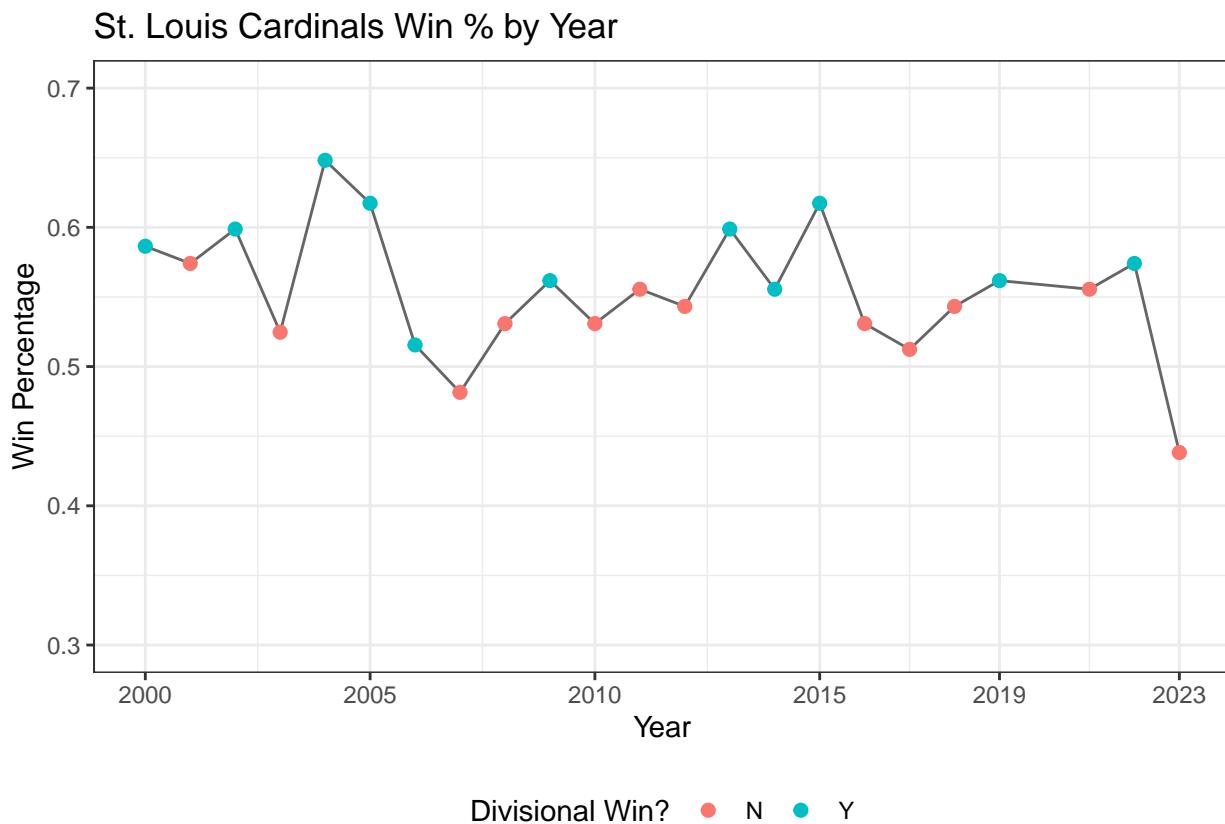
St. Louis Cardinals Win % by Year



As you can see there is history in the Cardinals for being a good team and having lots of wins. Specifically if

we look at the recent Cardinals (since 2000), and see how they fared in the division we will see that they have been successful.

```
ggplot(Cards %>% filter(yearID >= 2000), aes(x = yearID, y = Wpct)) +
  geom_line(color = "gray40") +
  geom_point(aes(color = DivWin), size = 2) +
  scale_y_continuous(limits = c(0.3, 0.7)) +
  scale_x_continuous(breaks = c(2000, 2005, 2010, 2015, 2019, 2023)) +
  labs(
    title = "St. Louis Cardinals Win % by Year",
    x = "Year",
    y = "Win Percentage",
    color = "Divisional Win?"
  ) +
  theme_bw() +
  theme(legend.position = "bottom")
```



As you can see the Cardinals have been very successful in their division in the past. In 2023 the Cardinals finished dead last in their division while winning it the previous year, and had one of the worst records in team history. The 2023 season was a change in the way the Cardinals played and viewed the team for the future for many fans. What some people thought might just be an off couple of months turned into one of the worst displays of baseball in Cardinals history that led them to their worst record in the 21st century and further from the playoffs than they thought.

In 2023 there was what seemed like an anomaly to people that are not Cardinals fans, but to Cardinals fans it was an exposure that could be seen coming for years.

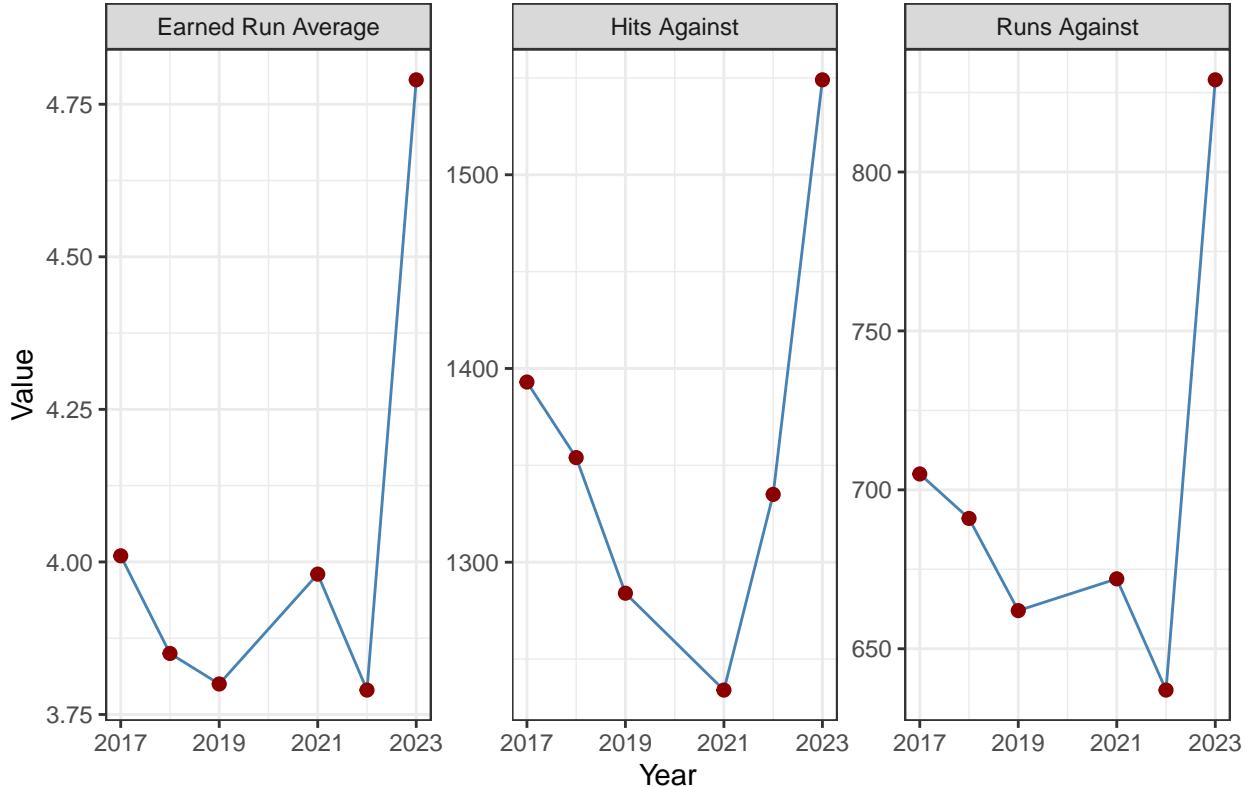
Research Question

The point of this paper is to look at factors that led to the Cardinals downfall of 2023. We will look specifically at the pitching downfall that led to a major difference production for the team. Looking at this plot below you will see the difference in the years and how pitching was a major difference in the season.

```
cards_long <- Cards %>%
  filter(yearID >= 2017 & yearID != 2020) %>%
  select(yearID, HA, RA, ERA) %>%
  pivot_longer(cols = c(HA, RA, ERA), names_to = "Stat", values_to = "Value") %>%
  mutate(Stat = recode(Stat,
    "HA" = "Hits Against",
    "RA" = "Runs Against",
    "ERA" = "Earned Run Average"))

# Plot
ggplot(cards_long, aes(x = yearID, y = Value)) +
  geom_line(group = 1, color = "steelblue") +
  geom_point(color = "darkred", size = 2) +
  facet_wrap(~ Stat, scales = "free_y") +
  scale_x_continuous(limits = c(2017, 2023), breaks = seq(2017, 2023, by = 2)) +
  labs(
    title = "Cardinals Pitching Statistics (2017–2023, Excluding 2020)",
    x = "Year",
    y = "Value"
  ) +
  theme_bw()
```

Cardinals Pitching Statistics (2017–2023, Excluding 2020)



All of these categories really shot up during this year and it led to one of the worst records in Cardinals history. When looking at this tibble you can see how just how different the pitching performance was during this time.

```
Cards %>%
  filter(yearID >= 2017 & yearID != 2020) %>%
  select("Year" = yearID, W, L, Rank, R, RA, ERA, HRA, SOA)

##   Year W L Rank R RA ERA HRA SOA
## 1 2017 83 79    3 761 705 4.01 183 1351
## 2 2018 88 74    3 759 691 3.85 144 1337
## 3 2019 91 71    1 764 662 3.80 191 1399
## 4 2021 90 72    2 706 672 3.98 152 1225
## 5 2022 93 69    1 772 637 3.79 146 1177
## 6 2023 71 91    5 719 829 4.79 179 1215
```

All categories appear to be much worse than in the past, especially the Runs Against. To know that it is a bad number and not just common among the league we will look at Runs scored for 2023 for all Teams.

```
library(mlbplotR)

lahman_to_mlbplotr <- tibble::tribble(
  ~teamID, ~team_abbr,
  "LAA", "LAA",
  "ARI", "AZ",
  "ATL", "ATL",
  "BAL", "BAL",
  "BOS", "BOS",
  "CHN", "CHC",
  "CHA", "CWS",
  "CIN", "CIN",
  "CLE", "CLE",
  "COL", "COL",
  "DET", "DET",
  "HOU", "HOU",
  "KCA", "KC",
  "LAN", "LAD",
  "MIA", "MIA",
  "MIL", "MIL",
  "MIN", "MIN",
  "NYA", "NYY",
  "NYN", "NYM",
  "OAK", "OAK",
  "PHI", "PHI",
  "PIT", "PIT",
  "SDN", "SD",
  "SEA", "SEA",
  "SFN", "SF",
  "SLN", "STL",
  "TBA", "TB",
  "TEX", "TEX",
  "TOR", "TOR",
  "WAS", "WSH"
)

# Get 2022 and 2023 teams data
```

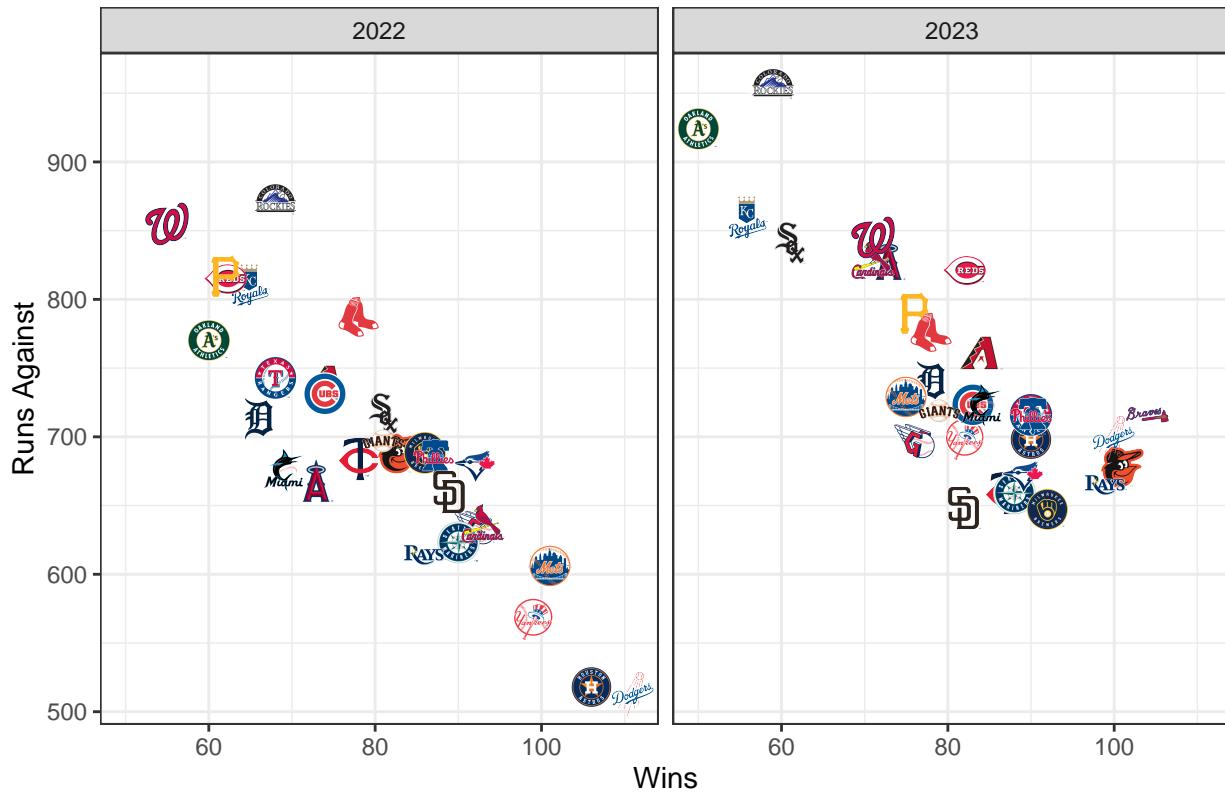
```

teams_recent <- Lahman::Teams %>%
  filter(yearID %in% c(2022, 2023)) %>%
  left_join(lahman_to_mlbplotr, by = "teamID") %>%
  filter(!is.na(team_abbr)) %>%
  mutate(year = factor(yearID))

# Plot with facet_wrap by year
ggplot(teams_recent, aes(x = W, y = RA)) +
  geom_mlb_logos(aes(team_abbr = team_abbr), width = 0.08) +
  facet_wrap(~ year) +
  labs(
    title = "Runs Against vs Wins (2022 & 2023)",
    x = "Wins",
    y = "Runs Against"
  ) +
  theme_bw()

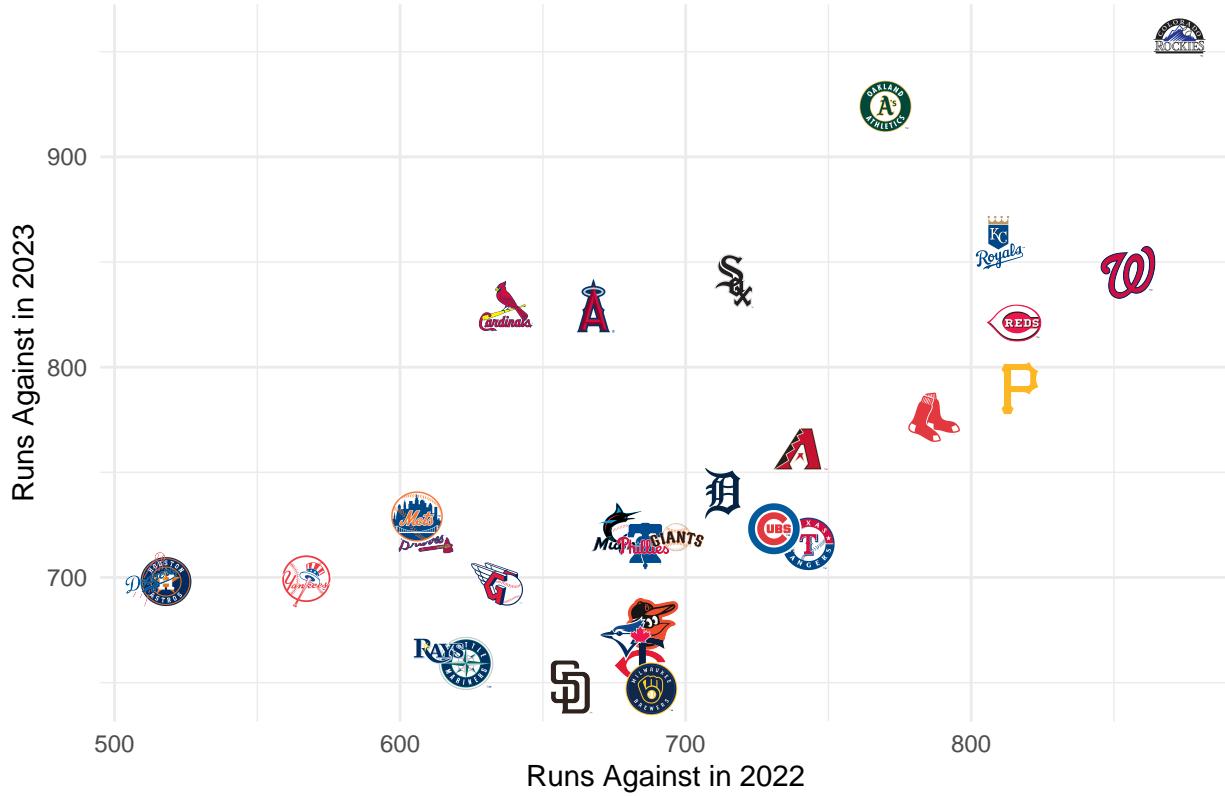
```

Runs Against vs Wins (2022 & 2023)

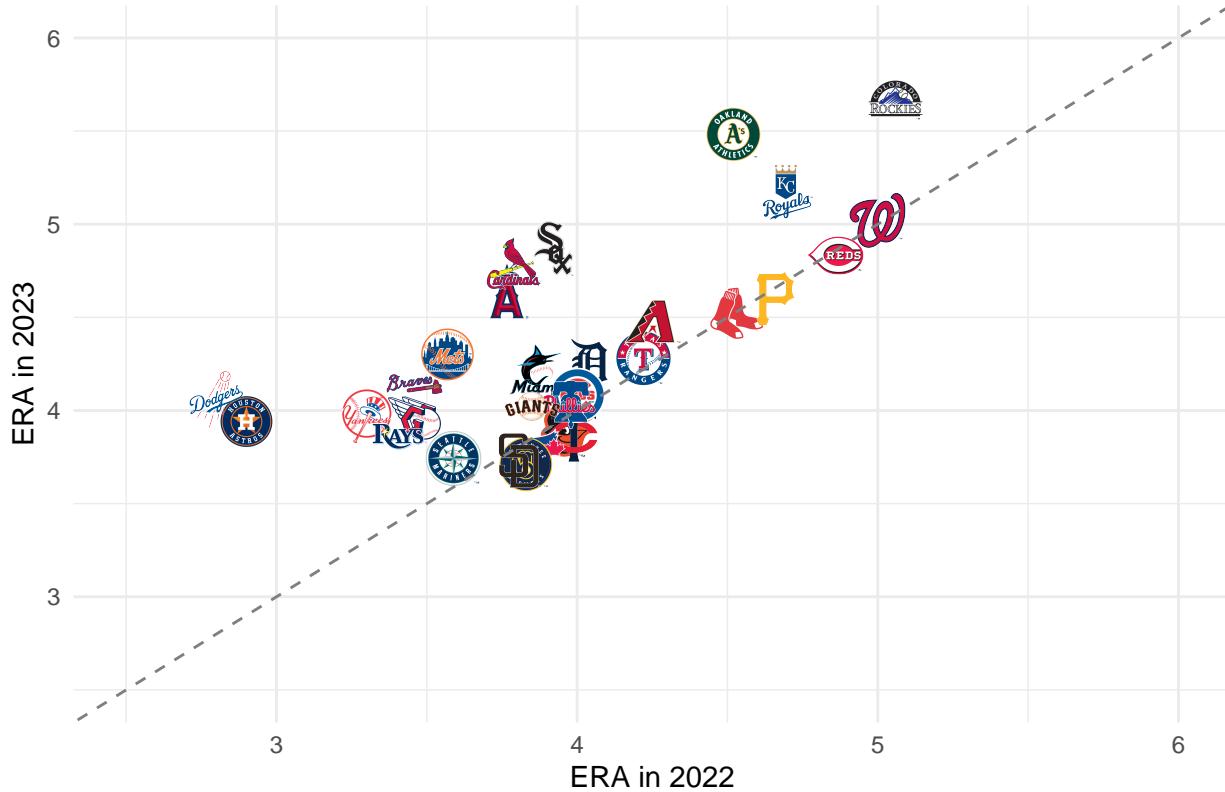


From this plot you can see that the Cardinals moved to a completely different region of the graph. Most of the teams that had low wins and high runs against stayed the same, while the Cardinals moved into a region unlike most of the other teams. If we now look into differences in Runs Against and ERA over the years you will see the similar pattern of the Cardinals moving into a territory of bad baseball.

2023 vs 2022 Runs Against by Team



2023 vs 2022 ERA by Team



Now as you can see almost every teams ERA increased from 2022-2023, but if you notice which teams had the largest increase are some of the worst teams in baseball for the 2023 season. The Cardinals were trending in a bad direction if they were grouped in a category with the Athletics, Angels, and White Sox.

Cardinals Roster

Between 2022 and 2023 there were some major losses to the Cardinals. They lost Albert Pujols and Yadier Molina to retirement. These are both future Hall of Famers that were two of the best players to ever put on a Cardinal uniform and even baseball uniform in general. The rest of the roster stayed pretty similar except for a couple of additions. The team added some young rookies and signed Willson Contreras from the Cubs as Molina's replacement. Less than 2 months into the season, the Cardinals announced they were benching Contreras for his bad play as a catcher. The front office blamed Contreras for the decline in pitching performance rather than blaming the pitchers for not being good. The starting rotation had 4 pitchers older than 31 years old and that included Wainwright who 40 and could not throw over 89 MPH. This also included Miles Mikolas who statistically has been one of the worst pitchers in the MLB since 2022 with leading the MLB in Hits Against while having a very low strikeout rate. He consistently has finished near the top of ERA ratings, but consistently gets a lot of starts while not producing like the 'Ace' he was paid to be.

This is why we will look into the Statcast data to find any information we can about the team and what really changed. I believe there will large differences in the way that the pitchers were pitching and how they were going about their game plans.

Statcast Data

We will look into the 2021-2023 pitching data from Statcast. This will give us 2 years of baseline performance before a bad year where we can look into the differences. We will compare the differences between the Cardinals past years and the differences between better teams during those years to see how they compare.

Load in the Data

```
# Load in Statcast Data
stat2021 <- readRDS('statcast2021.RData')
stat2022 <- readRDS('statcast2022.RData')
stat2023 <- readRDS('statcast2023.RData')

PLAYERIDMAP <- read_csv("C:/Users/qbcar/github/Stat430Baseball/PLAYERIDMAP.csv")

cards_2021 <- stat2021 %>%
  filter(home_team == 'STL' | away_team == 'STL') %>%
  left_join(PLAYERIDMAP %>% select(IDPLAYER, TEAM, MLBID, MLBNAME),
            by = c("pitcher_id" = 'MLBID'))

cards_2022 <- stat2022 %>%
  filter(home_team == 'STL' | away_team == 'STL') %>%
  left_join(PLAYERIDMAP %>% select(IDPLAYER, TEAM, MLBID, MLBNAME),
            by = c("pitcher_id" = 'MLBID'))

cards_2023 <- stat2023 %>%
  filter(home_team == 'STL' | away_team == 'STL') %>%
  left_join(PLAYERIDMAP %>% select(IDPLAYER, TEAM, MLBID, MLBNAME),
            by = c("pitcher_id" = 'MLBID'))

cards_pitch21 <- cards_2021 %>%
```

```

filter((home_team == 'STL' & half_inning == 'top') |
       (away_team == 'STL' & half_inning == 'bottom'))
cards_pitch22 <- cards_2022 %>%
  filter((home_team == 'STL' & half_inning == 'top') |
       (away_team == 'STL' & half_inning == 'bottom'))
cards_pitch23 <- cards_2023 %>%
  filter((home_team == 'STL' & half_inning == 'top') |
       (away_team == 'STL' & half_inning == 'bottom'))

```

The following code creates a function that makes a strikezone.

```

library(ggdensity)

# create geom to visualize the strikezone
geom_strikezone = function(sz_top = 3.8, sz_bot = 1.1) {
  plate_width = 17 + 2 * (9 / pi)
  sz_left = -(plate_width / 2) / 12
  sz_right = (plate_width / 2) / 12
  strikezone = data.frame(
    x = c(sz_left, sz_left, sz_right, sz_right, sz_left),
    y = c(sz_bot, sz_top, sz_top, sz_bot, sz_bot)
  )
  geom_path(
    mapping = aes(.data$x, .data$y),
    data = strikezone,
    linewidth = 0.5,
    linetype = 1,
    color = "black"
  )
}
location_func <- function(pitchers){
  location_plot = {
    ggplot(pitchers) +
      aes(x = plate_x, y = plate_z) +
      xlim(c(-4, 4)) +
      ylim(c(-2, 6)) +
      xlab("") +
      ylab("") +
      theme_bw() +
      coord_fixed() +
      facet_wrap(vars(pitch_name))
  }

  location_plot +
    geom_hdr(
      method = method_kde(h = 0.75),
      probs = c(0.90, 0.70, 0.50, 0.40, 0.30, 0.20),
      show.legend = FALSE,
      aes(fill = after_stat(probs)),
      alpha = 0.75
    ) +
    scale_fill_brewer(
      palette = "RdBu",
      direction = -1
    )
}

```

```

    ) +
  geom_strikezone()
}

```

Cardinals Pitchers

One of the first things we will look into is pitching usage. Looking into the players they were pitching for the Cardinals and see what is the difference in mindset of using certain pitches. So first we will look at the overall usage trends for each year by the pitch type.

```

cards_perc21 <- cards_pitch21 %>%
  group_by(pitch_name) %>%
  summarise(tot = n()) %>%
  mutate(usage = round(tot / sum(tot), digits = 4)) %>%
  arrange(desc(usage)) %>%
  mutate(year = 2021)

cards_perc22 <- cards_pitch22 %>%
  group_by(pitch_name) %>%
  summarise(tot = n()) %>%
  mutate(usage = round(tot / sum(tot), digits = 4)) %>%
  arrange(desc(usage)) %>%
  mutate(year = 2022)

cards_perc23 <- cards_pitch23 %>%
  group_by(pitch_name) %>%
  summarise(tot = n()) %>%
  mutate(usage = round(tot / sum(tot), digits = 4)) %>%
  arrange(desc(usage)) %>%
  mutate(year = 2023)

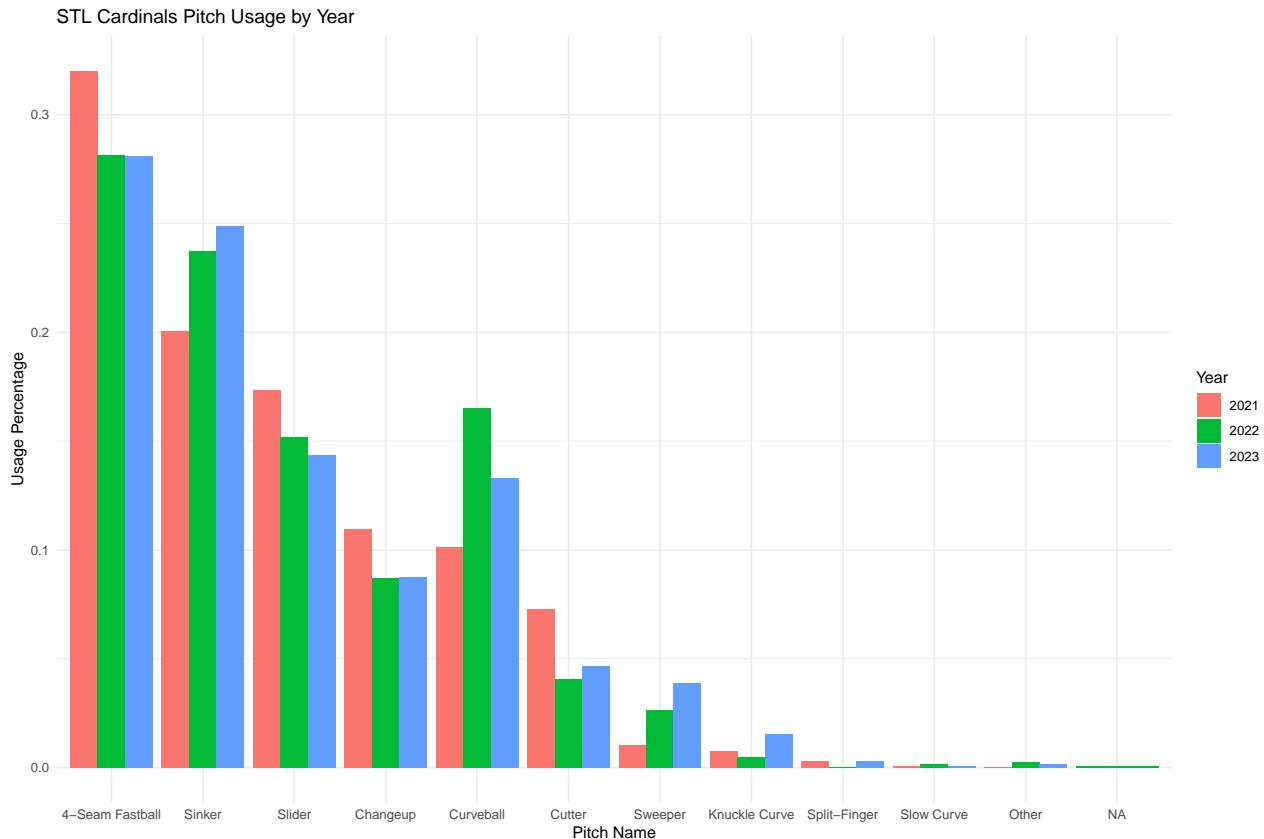
cards_all <- bind_rows(cards_perc21, cards_perc22, cards_perc23)

pitch_order <- cards_perc21 %>%
  arrange(desc(usage)) %>%
  pull(pitch_name)

cards_all$pitch_name <- factor(cards_all$pitch_name, levels = pitch_order)

# Step 3: Plot
ggplot(cards_all, aes(x = pitch_name, y = usage, fill = factor(year))) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "STL Cardinals Pitch Usage by Year",
    x = "Pitch Name",
    y = "Usage Percentage",
    fill = "Year"
  ) +
  theme_minimal()

```



As you can see there was a decrease in the usage of 4-Seam, Slider, and Changeup while there was an increase in the usage of Sinker and Curveball. This seems to be a change in who was on the mound and the idea of attack within each pitcher. I know that there was a change in the pitching coach between 2022 and 2023 so that could indicate some of the arsenal changes.

In the past the Cardinals have always been a pitch team that does not get a lot of strikeouts. Seeing that over 50% of pitches are Fastballs of some sort really shows the idea that the Cardinals do not throw off speed a lot compared to other teams.

We will specifically look at the Starting Pitchers and look into their stats to see what they were like.

```
# Summarize RA by starters only
starters_RA <- Lahman::Pitching %>%
  filter(yearID >= 2021 & teamID == 'SLN' & GS > 0) %>%
  group_by(yearID) %>%
  summarise(RA_starters = sum(R))

# Get total team RA from Teams dataset
team_RA <- Lahman::Teams %>%
  filter(yearID >= 2021 & teamID == 'SLN') %>%
  select(yearID, RA) %>%
  rename(RA_team = RA)

# Combine into one table
RA_comparison <- left_join(starters_RA, team_RA, by = "yearID")
RA_comparison %>% mutate(Run_percent = RA_starters / RA_team)

## # A tibble: 3 x 4
##   yearID RA_starters RA_team Run_percent
##       <dbl>      <dbl>    <dbl>        <dbl>
## 1     2021        0.00     0.00        0.00
## 2     2022        0.00     0.00        0.00
## 3     2023        0.00     0.00        0.00
```

```

##      <int>      <int>      <int>      <dbl>
## 1    2021        439       672     0.653
## 2    2022        463       637     0.727
## 3    2023        548       829     0.661

```

As you can see the Starters account for around 66% of all runs scored on the Cardinals. Therefore we will look into some of the data to see if there are any noticeable reasons for the increase in runs against in 2023. First we will need to figure out who the pitchers are and then look into their differences.

```

starter_years <- Lahman::Pitching %>%
  filter(yearID %in% 2021:2023, teamID == "SLN", GS > 0) %>%
  select(playerID, yearID) %>%
  distinct()

# Collapse years into a comma-separated string per player
starter_years_collapsed <- starter_years %>%
  group_by(playerID) %>%
  summarise(years_started = paste(sort(unique(yearID)), collapse = ", ")) %>%
  ungroup()

# Join with Players table to get names
starter_names <- Lahman::People %>%
  select(playerID, nameFirst, nameLast)

# Combine names with starter years
cards_starters_named <- starter_years_collapsed %>%
  left_join(starter_names, by = "playerID") %>%
  mutate(player_name = paste(nameFirst, nameLast)) %>%
  select(player_name, years_started) %>%
  arrange(player_name)

cards_starters_named

## # A tibble: 22 x 2
##   player_name      years_started
##   <chr>            <chr>
## 1 Adam Wainwright 2021, 2022, 2023
## 2 Andre Pallante   2022
## 3 Carlos Martinez 2021
## 4 Dakota Hudson   2021, 2022, 2023
## 5 Daniel Ponce de Leon 2021
## 6 Drew Rom         2023
## 7 J. A. Happ       2021
## 8 Jack Flaherty   2021, 2022, 2023
## 9 Jake Woodford   2021, 2022, 2023
## 10 Johan Oviedo   2021, 2022
## # i 12 more rows

```

We will first analyze the players that pitched in 2023 and also pitched in 2022 to really understand the differences in their pitching.

```

# Filter Cardinals pitchers from 2021-2023
cards_pitchers <- Lahman::Pitching %>%
  filter(yearID %in% 2021:2023, teamID == "SLN" & GS > 0) %>%
  select(playerID, yearID) %>%
  distinct()

```

```

# Identify pitchers who pitched in 2023
pitched_2023 <- cards_pitchers %>%
  filter(yearID == 2023) %>%
  pull(playerID)

# Identify pitchers who also pitched in 2022
pitched_other_years <- cards_pitchers %>%
  filter(yearID %in% c(2022)) %>%
  pull(playerID)

# Get intersection of both sets
multi_year_pitchers <- intersect(pitched_2023, pitched_other_years)

# Get names and years pitched for these players
multi_year_pitcher_years <- cards_pitchers %>%
  filter(playerID %in% multi_year_pitchers) %>%
  group_by(playerID) %>%
  summarise(years_pitched = paste(sort(unique(yearID)), collapse = ", ")) %>%
  ungroup()

# Add player names
pitcher_names <- Lahman::People %>%
  select(playerID, nameFirst, nameLast)

multi_year_pitchers_named <- multi_year_pitcher_years %>%
  left_join(pitcher_names, by = "playerID") %>%
  mutate(player_name = paste(nameFirst, nameLast)) %>%
  select(player_name, years_pitched) %>%
  arrange(player_name)

multi_year_pitchers_named

## # A tibble: 9 x 2
##   player_name      years_pitched
##   <chr>            <chr>
## 1 Adam Wainwright 2021, 2022, 2023
## 2 Dakota Hudson    2021, 2022, 2023
## 3 Jack Flaherty    2021, 2022, 2023
## 4 Jake Woodford    2021, 2022, 2023
## 5 Jordan Montgomery 2022, 2023
## 6 Matthew Liberatore 2022, 2023
## 7 Miles Mikolas    2021, 2022, 2023
## 8 Steven Matz       2022, 2023
## 9 Zack Thompson     2022, 2023

```

As you can see there is about 9 starters that pitched in both years. So we will do a look into what changed between these pitchers and how their seasons differed.

```

# Step 1: Filter Cardinals starters from 2021-2023
cards_pitchers <- Lahman::Pitching %>%
  filter(yearID %in% 2021:2023, teamID == "SLN", GS > 0) %>%
  select(playerID, yearID) %>%
  distinct()

```

```

# Step 2: Get 2023 and 2022 pitchers
pitched_2023 <- cards_pitchers %>%
  filter(yearID == 2023) %>%
  pull(playerID)

pitched_2022 <- cards_pitchers %>%
  filter(yearID == 2022) %>%
  pull(playerID)

# Step 3: Intersect 2023 and 2022 pitchers
multi_year_pitchers <- intersect(pitched_2023, pitched_2022)

# Step 4: Extract yearly stats for each qualifying pitcher
cards_stats <- Lahman::Pitching %>%
  filter(playerID %in% multi_year_pitchers, teamID == "SLN", yearID %in% 2022:2023) %>%
  select(playerID, yearID, ERA, R, W, L, H, GS, SO)

# Step 5: Add player names
pitcher_names <- Lahman::People %>%
  select(playerID, nameFirst, nameLast)

cards_stats_named <- cards_stats %>%
  left_join(pitcher_names, by = "playerID") %>%
  mutate(player_name = paste(nameFirst, nameLast)) %>%
  select(player_name, yearID, ERA, R, W, L, H, GS, SO) %>%
  arrange(player_name, yearID)

cards_stats_named

##          player_name yearID   ERA    R    W    L    H  GS   SO
## 1      Adam Wainwright  2022 3.71  80  11  12 192 32 143
## 2      Adam Wainwright  2023 7.40  89   5  11 151 21  55
## 3      Dakota Hudson  2022 4.45  71   8   7 141 26  78
## 4      Dakota Hudson  2023 4.98  46   6   3  88 12  45
## 5      Jack Flaherty  2022 4.25  18   2   1  36  8  33
## 6      Jack Flaherty  2023 4.43  56   7   6 116 20 106
## 7      Jake Woodford  2022 2.23  13   4   0  43  1  24
## 8      Jake Woodford  2023 6.23  34   2   3  61  8  29
## 9      Jordan Montgomery  2022 3.11  24   6   3  56 11  61
## 10     Jordan Montgomery  2023 3.42  54   6   9 116 21 108
## 11     Matthew Liberatore  2022 5.97  23   2   2  42  7  28
## 12     Matthew Liberatore  2023 5.25  42   3   6  66 11  46
## 13     Miles Mikolas  2022 3.29  81  12  13 170 32 153
## 14     Miles Mikolas  2023 4.78 110   9  13 226 35 137
## 15     Steven Matz  2022 5.25  28   5   3  50 10  54
## 16     Steven Matz  2023 3.86  48   4   7 108 17  98
## 17     Zack Thompson  2022 2.08   9   1   1  20  1  27
## 18     Zack Thompson  2023 4.48  35   5   7  69  9  72

```

After looking into this a little more we can see that Jake Woodford and Zack Thompson both only started in 1 game in 2022, so we should exclude them from this as their data will not enough to compare.

```

cards_stats_named <- cards_stats_named %>%
  filter(!player_name %in% c("Jake Woodford", "Zack Thompson"))
cards_stats_named

```

```

##          player_name yearID   ERA    R    W    L    H   GS   SO
## 1      Adam Wainwright  2022 3.71  80  11  12 192 32 143
## 2      Adam Wainwright  2023 7.40  89  5 11 151 21 55
## 3      Dakota Hudson  2022 4.45  71  8  7 141 26 78
## 4      Dakota Hudson  2023 4.98  46  6  3 88 12 45
## 5      Jack Flaherty  2022 4.25  18  2  1 36 8 33
## 6      Jack Flaherty  2023 4.43  56  7  6 116 20 106
## 7      Jordan Montgomery  2022 3.11  24  6  3 56 11 61
## 8      Jordan Montgomery  2023 3.42  54  6  9 116 21 108
## 9      Matthew Liberatore  2022 5.97  23  2  2 42 7 28
## 10     Matthew Liberatore  2023 5.25  42  3  6 66 11 46
## 11     Miles Mikolas  2022 3.29  81 12 13 170 32 153
## 12     Miles Mikolas  2023 4.78 110  9 13 226 35 137
## 13     Steven Matz  2022 5.25  28  5  3 50 10 54
## 14     Steven Matz  2023 3.86  48  4  7 108 17 98

```

So the pitchers we will look into are Adam Wainwright, Dakota Hudson, Jack Flaherty, Jordan Montgomery, Matthew Liberatore, Miles Mikolas, and Steven Matz.

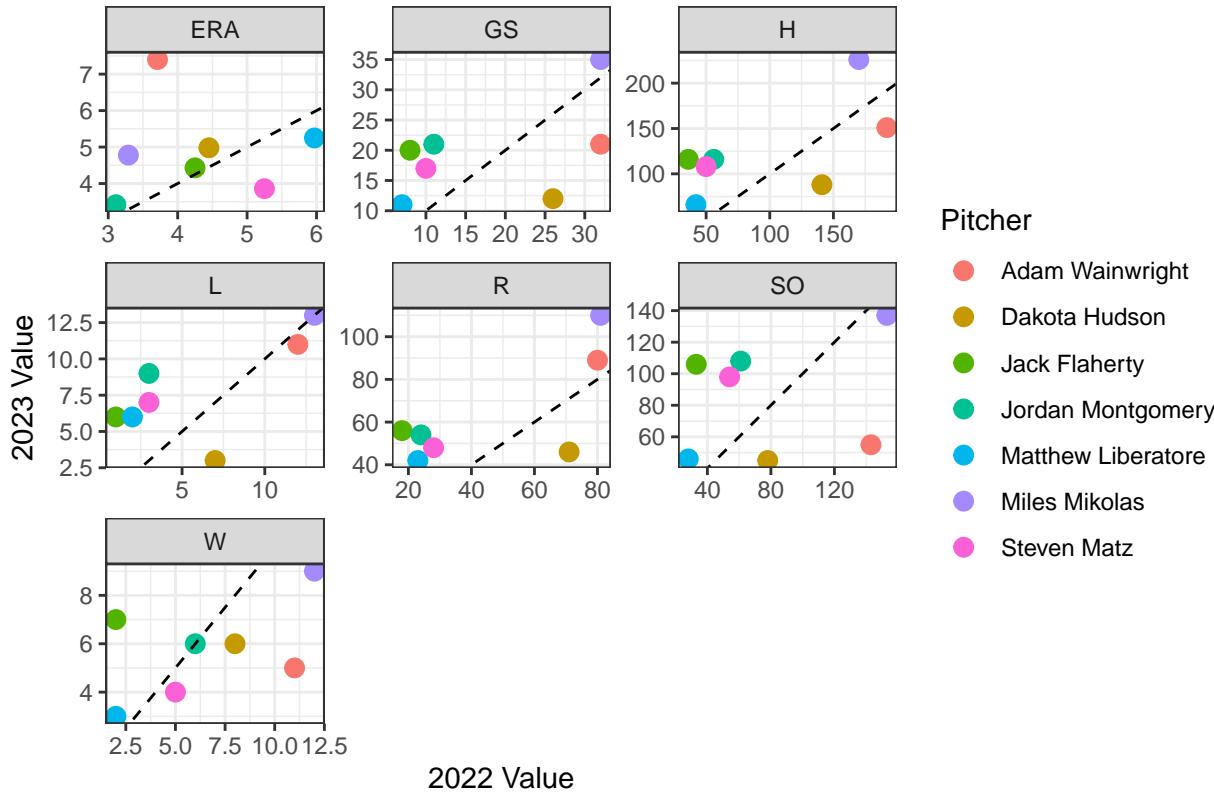
```

# Reshape the data to long format
cards_plot_data <- cards_stats_named %>%
  pivot_longer(cols = c(ERA, R, W, H, L, GS, SO),
               names_to = "stat",
               values_to = "value") %>%
  pivot_wider(names_from = yearID, values_from = value, names_prefix = "year_") %>%
  filter(!is.na(year_2022) & !is.na(year_2023)) # Ensure both years are present

# Plot
ggplot(cards_plot_data, aes(x = year_2022, y = year_2023, color = player_name)) +
  geom_point(size = 3) +
  geom_abline(slope = 1, intercept = 0, linetype = "dashed") +
  facet_wrap(~stat, scales = "free") +
  labs(
    title = "Pitching Stats: 2022 vs 2023 by Pitcher",
    x = "2022 Value",
    y = "2023 Value",
    color = "Pitcher"
  ) +
  theme_bw()

```

Pitching Stats: 2022 vs 2023 by Pitcher



As you can see from the plot above almost all categories are higher in 2023 for these starters. It is interesting to notice that Miles Mikolas stats are not that differing from past years, which showcases that he stayed about the same except in RA and ERA. There appears to be an increase all around of runs against, which is what we will research farther into factors that might have caused them to be hit off of.

It is also interesting that Adam Wainwright pitched in less games, but had more runs against. Looking into the differences in runs against by pitchers shows just how the ERA differs and why they were trended in a bad direction.

Statcast Cardinals Starters

Now let's use Statcast data on Cardinals Starters to find any differences. We will first look at each pitchers arsenal and see how that differed over the years.

```
target_pitchers <- cards_stats_named$player_name

# Combine years
cards_pitch_all <- bind_rows(
  cards_pitch22 %>% mutate(year = 2022),
  cards_pitch23 %>% mutate(year = 2023)
)

# Filter to target pitchers only
pitchers_arenal <- cards_pitch_all %>%
  filter(MLBNAME %in% target_pitchers) %>%
  filter(!is.na(pitch_name)) %>%
  group_by(MLBNAME, year, pitch_name) %>%
  summarise(n = n(), .groups = "drop") %>%
```

```

arrange(MLBNAME, year, desc(n))

pitcher_usage <- pitchers_arsenal %>%
  group_by(MLBNAME, year) %>%
  mutate(usage = round(n / sum(n), 4)) %>%
  ungroup()
pitcher_usage

## # A tibble: 78 x 5
##   MLBNAME      year pitch_name     n  usage
##   <chr>        <dbl> <chr>       <int>  <dbl>
## 1 Adam Wainwright 2022 Curveball    960  0.306
## 2 Adam Wainwright 2022 Sinker      874  0.279
## 3 Adam Wainwright 2022 Cutter      784  0.250
## 4 Adam Wainwright 2022 4-Seam Fastball 290  0.0926
## 5 Adam Wainwright 2022 Changeup    189  0.0603
## 6 Adam Wainwright 2022 Slow Curve   33   0.0105
## 7 Adam Wainwright 2022 Slider      3    0.001
## 8 Adam Wainwright 2023 Sinker      560  0.314
## 9 Adam Wainwright 2023 Curveball    546  0.306
## 10 Adam Wainwright 2023 Cutter     404  0.226
## # i 68 more rows

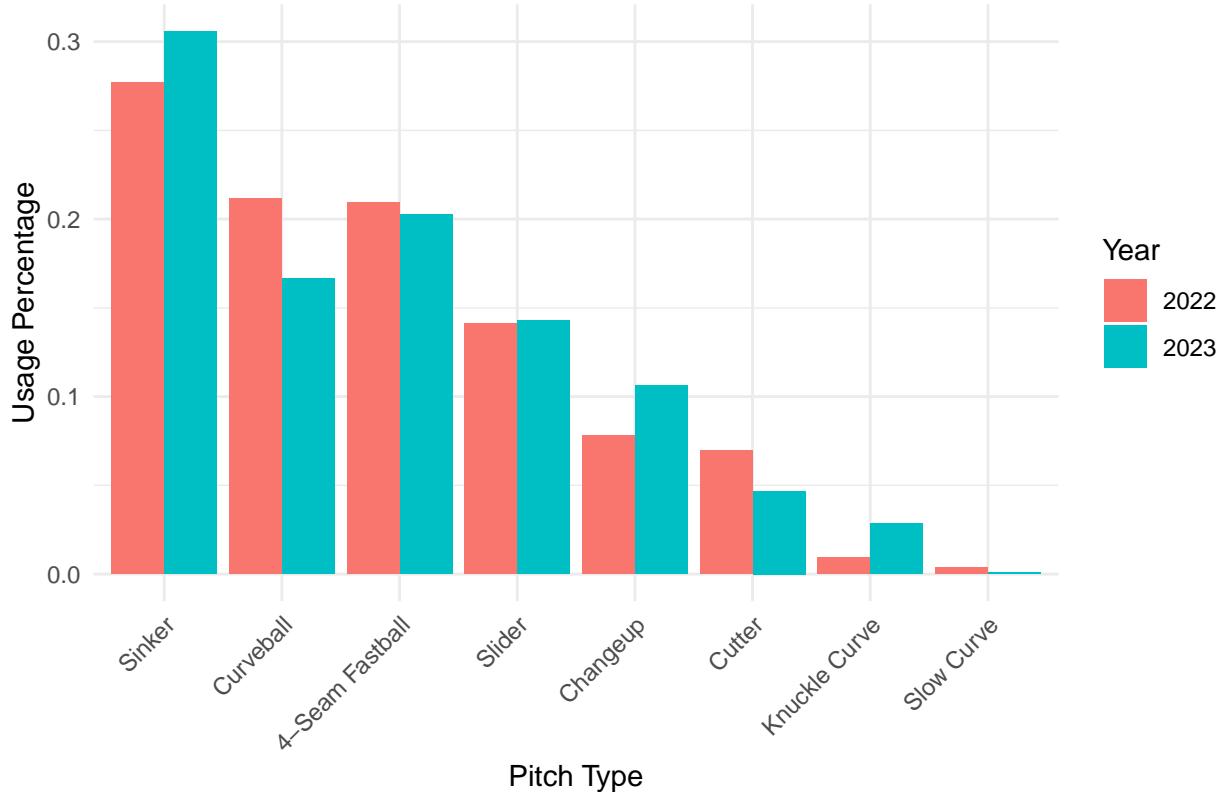
total_usage <- pitcher_usage %>%
  group_by(year, pitch_name) %>%
  summarise(tot = sum(n), .groups = "drop")
total_usage <- total_usage %>%
  group_by(year) %>%
  mutate(usage = round(tot / sum(tot), 4)) %>%
  ungroup()
pitch_order <- total_usage %>%
  filter(year == 2022) %>%
  arrange(desc(usage)) %>%
  pull(pitch_name)

total_usage$pitch_name <- factor(total_usage$pitch_name, levels = pitch_order)

ggplot(total_usage, aes(x = pitch_name, y = usage, fill = factor(year))) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Pitch Usage by Year (Filtered Starters)",
    x = "Pitch Type",
    y = "Usage Percentage",
    fill = "Year"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

Pitch Usage by Year (Filtered Starters)



Here you can find that starters use Sinker considerably more and started to decrease in Curveball.

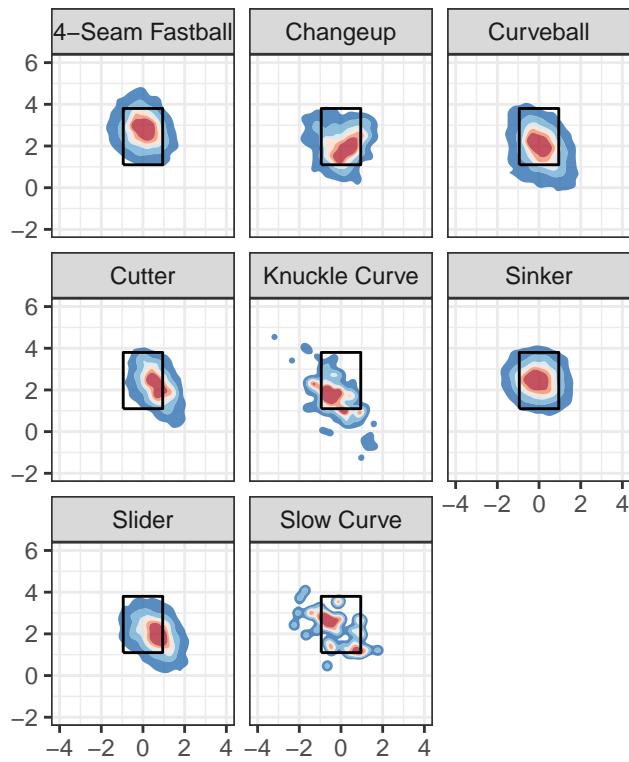
Pitch Location

Since we know that they like to throw certain pitches more, lets look at locations of where these pitchers are going.

```
Cards_starters_sc <- cards_pitch_all %>%
  filter(MLBNAME %in% target_pitchers) %>%
  filter(!is.na(pitch_name))

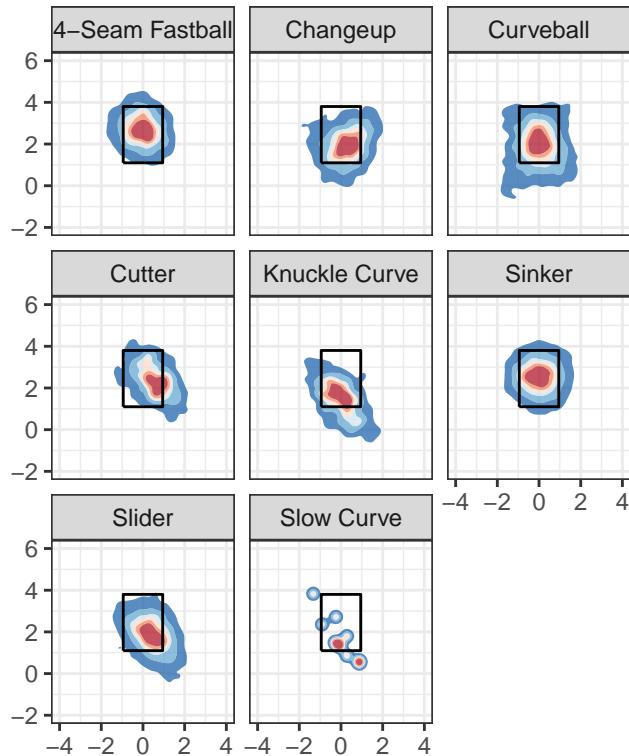
location_func(Cards_starters_sc %>% filter(year == 2022)) +
  labs(title = '2022 Pitching Heat Maps')
```

2022 Pitching Heat Maps



```
location_func(Cards_starters_sc %>% filter(year == 2023)) +  
  labs(title = '2023 Pitching Heat Maps')
```

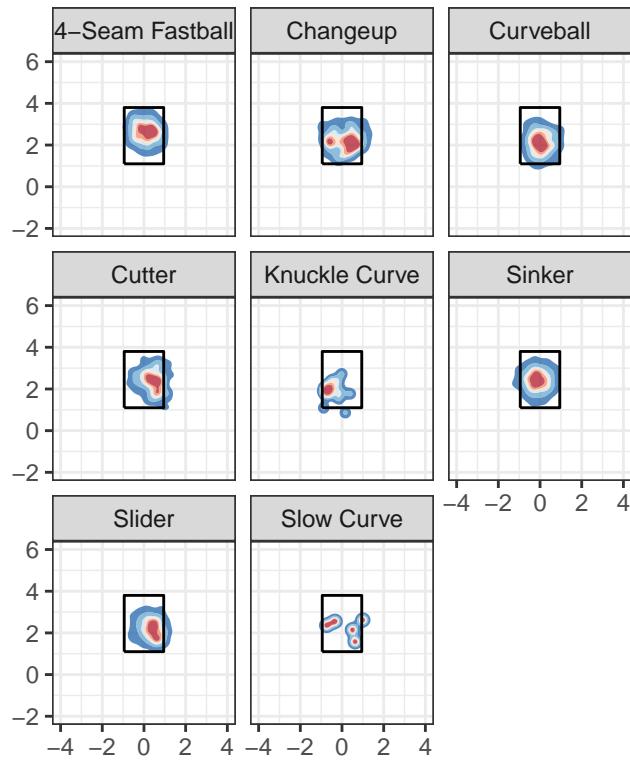
2023 Pitching Heat Maps



As you can see true heat maps do not tell a full story of actually what is happening in the games. I want to separate hits vs nonhits to see what the pitches looked like when they were being put into play.

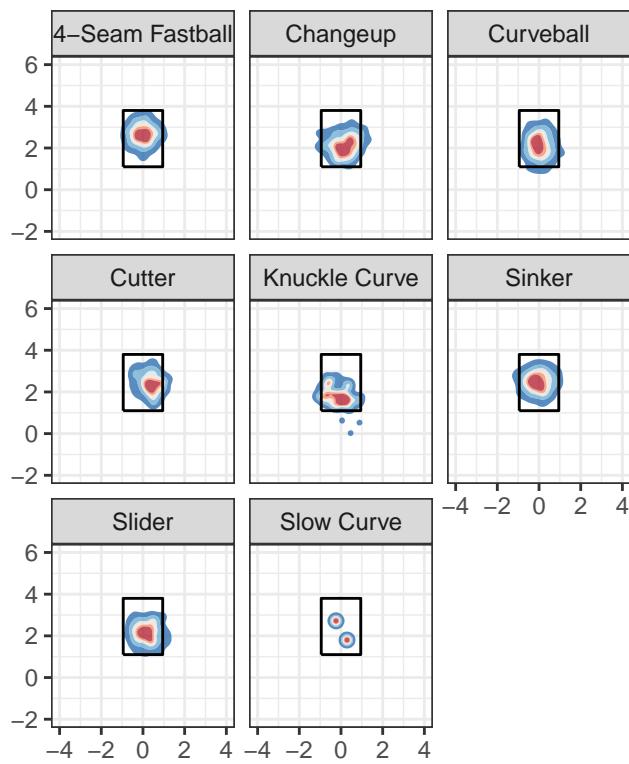
```
play22 <- Cards_starters_sc %>% filter(year == 2022 & type == 'X')
location_func(play22) +
  labs(title = '2022 Pitching Heat Maps Balls in Play')
```

2022 Pitching Heat Maps Balls in Play



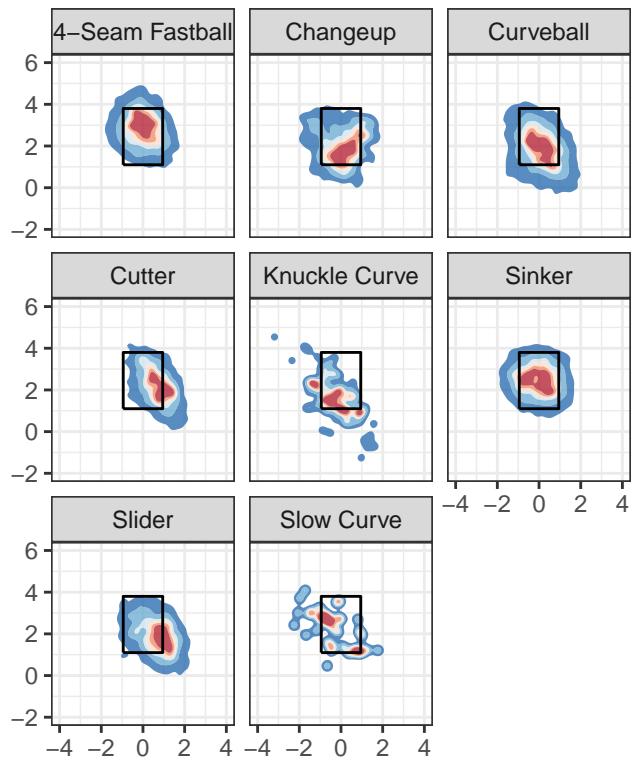
```
play23 <- Cards_starters_sc %>% filter(year == 2023 & type == 'X')
location_func(play23) +
  labs(title = '2023 Pitching Heat Maps Balls in Play')
```

2023 Pitching Heat Maps Balls in Play



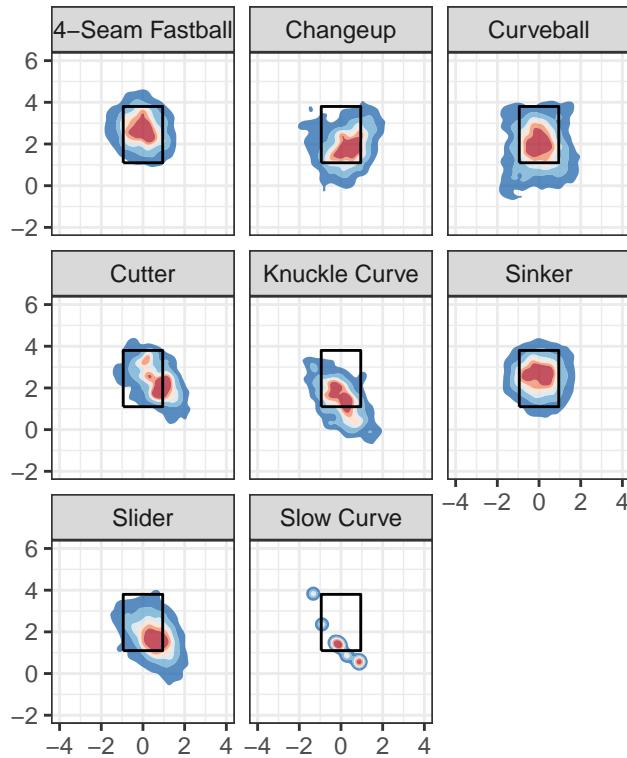
```
miss22 <- Cards_starters_sc %>% filter(year == 2022 & type != 'X')
location_func(miss22) +
  labs(title = '2022 Pitching Heat Maps')
```

2022 Pitching Heat Maps



```
miss23 <- Cards_starters_sc %>% filter(year == 2023 & type != 'X')
location_func(miss23) +
  labs(title = '2023 Pitching Heat Maps')
```

2023 Pitching Heat Maps



The heat maps are definitely a lot tighter for Balls in Play. The Cardinals do pitch for contact and you can see the differences in how they are just throwing strikes to make sure that they can try and get outs. We can see that the Cardinals do throw a lot of pitches over the plate and care more about throwing strikes than anything else. Since there does not seem to be a large change in location of the heatmaps, lets take a look at the effectiveness of each pitch type and what it was doing.

Effects of Each Pitch

We will now look into the details of each pitch type and also what were there any differences in pitch statistics such as spin rate, movement, and other factors.

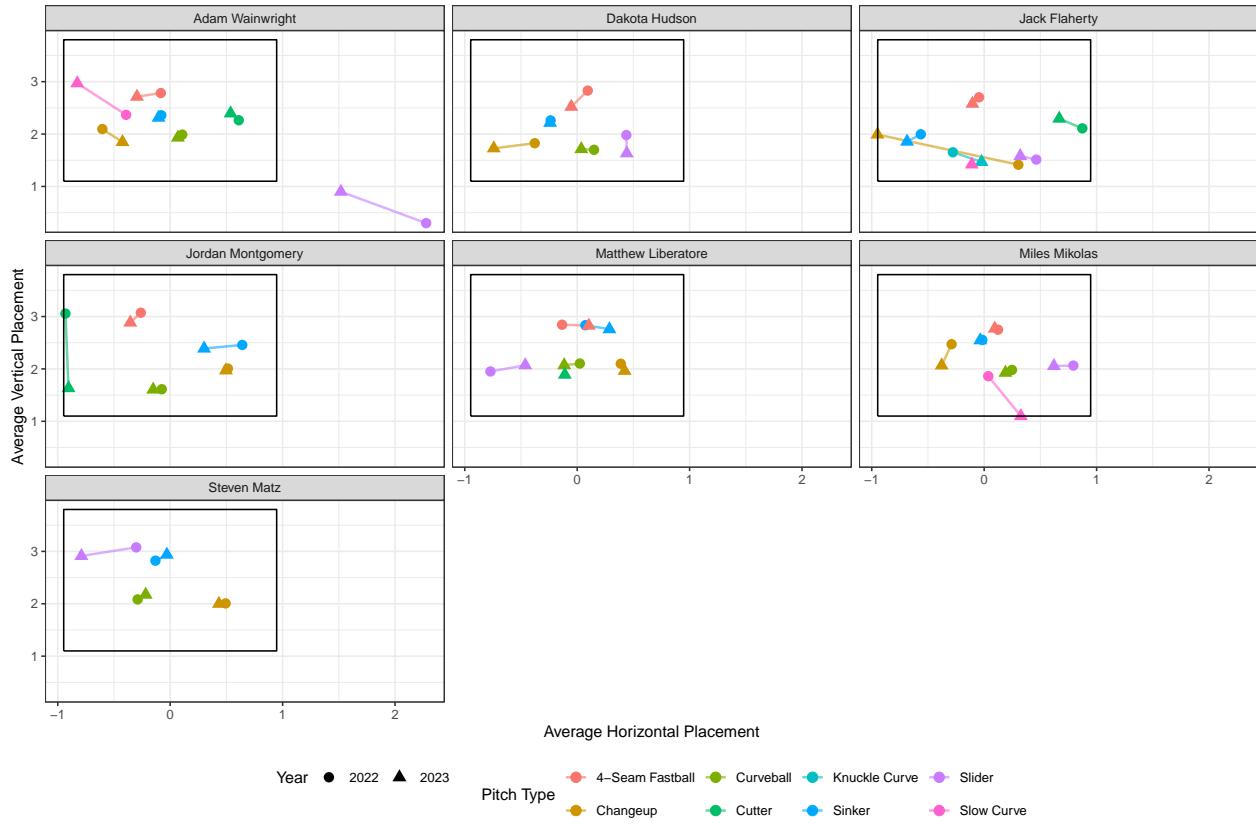
```
# Find Averages of all pitch stats
cards_sc <- Cards_starters_sc %>%
  group_by(MLBNAME, year, pitch_name) %>%
  summarise(
    avg_spin = mean(release_spin_rate, na.rm = TRUE),
    avg_velo = mean(release_speed, na.rm = TRUE),
    avg_x = mean(plate_x, na.rm = TRUE),
    avg_z = mean(plate_z, na.rm = TRUE),
    avg_horizontal = mean(pfx_x, na.rm = TRUE),
    avg_vertical = mean(pfx_z, na.rm = TRUE),
    avg_striket = mean(strike_zone_top),
    avg_strikeb = mean(strike_zone_bottom),
    .groups = "drop"
  )

ggplot(cards_sc, aes(x = avg_x, y = avg_z, color = pitch_name)) +
  geom_line(aes(group = interaction(MLBNAME, pitch_name)), linewidth = 0.8, alpha = 0.6) +
  geom_point(aes(shape = factor(year)), size = 3) +
```

```

geom_strikezone() +
facet_wrap(~MLBNAME) +
theme_bw() +
theme(legend.position = 'bottom') +
labs(
  x = "Average Horizontal Placement",
  y = "Average Vertical Placement",
  color = "Pitch Type",
  shape = "Year"
)

```

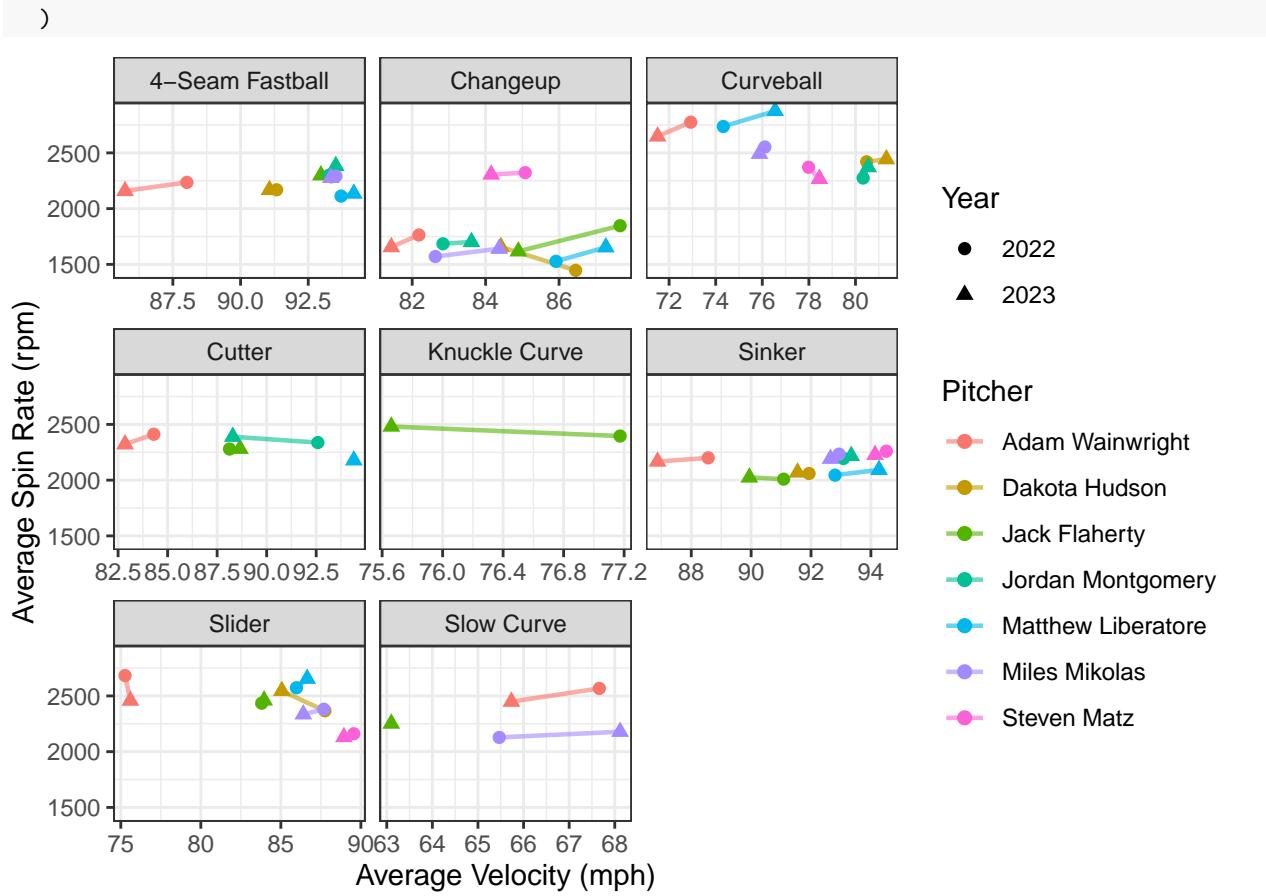


The plot above shows the differences in average pitch location between years. As you can see most of the pitchers average location for each pitch moved towards the middle or towards the inside of right handed hitter. These locations are the from the catchers perspective and thus can show the difference in where a catcher calls a pitch. For example most pitches were on the left side and you can see where the majority of the pitches were moving towards the middle. Looking at Flaherty you can see major changes in changeup location and Matz has huge differences in slider location. These changes in location most likely have to do with changes in strategy of game calling by these pitchers and how they want to attack hitters.

```

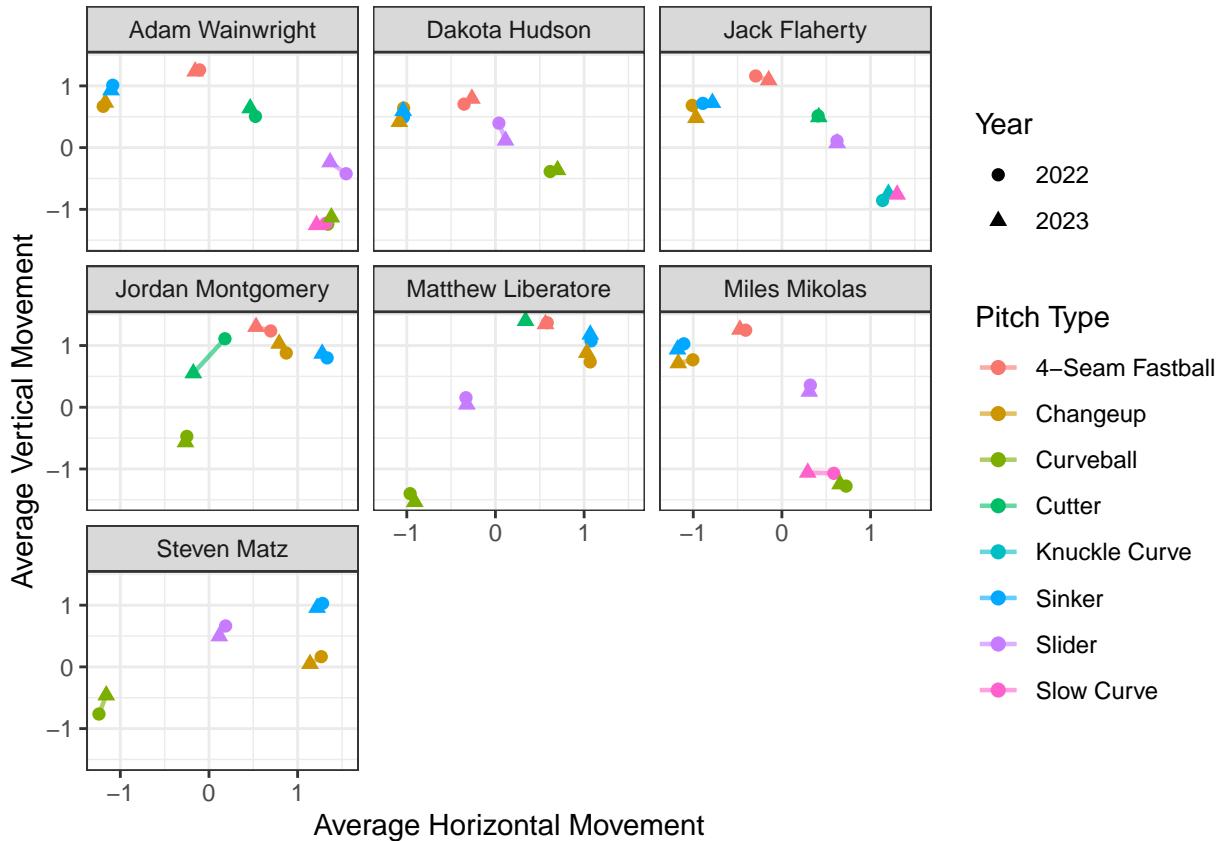
ggplot(cards_sc, aes(x = avg_velo, y = avg_spin, color = MLBNAME)) +
  geom_line(aes(group = interaction(MLBNAME, pitch_name)), linewidth = 0.8, alpha = 0.6) +
  geom_point(aes(shape = factor(year)), size = 2) +
  facet_wrap(~pitch_name, scales = 'free_x') +
  theme_bw() +
  labs(
    x = "Average Velocity (mph)",
    y = "Average Spin Rate (rpm)",
    color = "Pitcher",
    shape = "Year"
)

```



This plot above shows the differences in velocity and spin rate by pitch type. As you can see almost every pitcher except Mikolas has a decrease in velocity. This might not seem severe with the average dropping around 1-2 MPH, but is not a direction you want to be with a starting rotation that is aging and is seeing differences in velocity. Looking deeper you can see that there also appears to a slight decrease in spin rate, which also shows that pitches were not moving as much as others should be. This leads to a decrease in tunneling and can make the difference between hitters picking up the pitch earlier out of your hand. When pitches start to become more recognizable and you are just trying to pump strikes then you can see hitters start to recognize pitches better.

```
ggplot(cards_sc, aes(x = avg_horizontal, y = avg_vertical, color = pitch_name)) +
  geom_line(aes(group = interaction(MLBNAME, pitch_name)), linewidth = 0.8, alpha = 0.6) +
  geom_point(aes(shape = factor(year)), size = 2) +
  facet_wrap(~MLBNAME) +
  theme_bw() +
  labs(
    x = "Average Horizontal Movement",
    y = "Average Vertical Movement",
    color = "Pitch Type",
    shape = "Year"
  )
```



This plot shows the horizontal movement and vertical movement differences for each pitcher and pitch type. There is not really any significant differences when looking at averages of the movements of pitches. This would lead me to believe that the pitchers were not having less movement, but rather slower spin and velocity. The ball was still moving, not a lot compared to other pitchers who throw more off speed pitches, but still moving like they normally do.

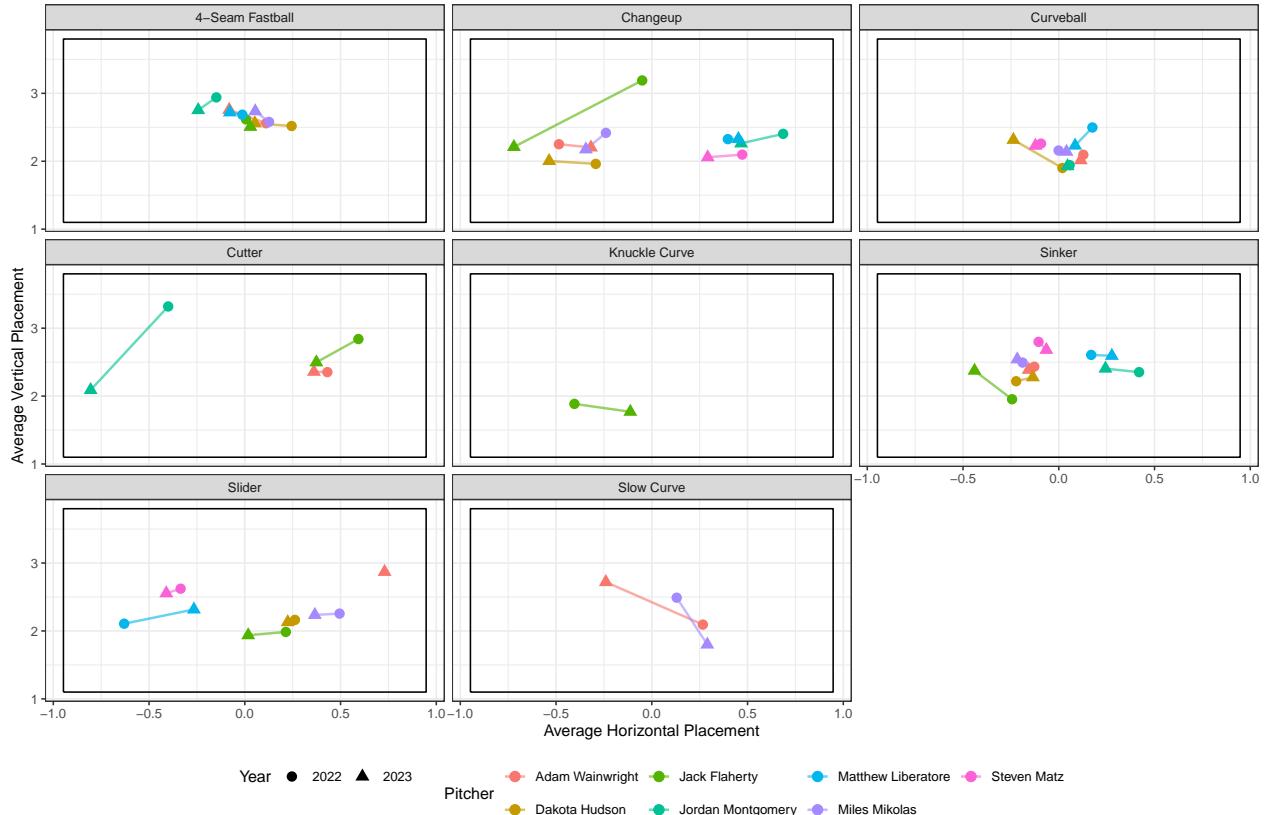
```
# Find Averages of all pitch stats in play
cards_play <- Cards_starters_sc %>%
  filter(type == 'X') %>%
  group_by(MLBNAME, year, pitch_name) %>%
  summarise(
    avg_spin = mean(release_spin_rate, na.rm = TRUE),
    avg_velo = mean(release_speed, na.rm = TRUE),
    avg_x = mean(plate_x, na.rm = TRUE),
    avg_z = mean(plate_z, na.rm = TRUE),
    avg_horizontal = mean(pfx_x, na.rm = TRUE),
    avg_vertical = mean(pfx_z, na.rm = TRUE),
    avg_striket = mean(strike_zone_top),
    avg_strikeb = mean(strike_zone_bottom),
    .groups = "drop"
  )

ggplot(cards_play, aes(x = avg_x, y = avg_z, color = MLBNAME)) +
  geom_line(aes(group = interaction(MLBNAME, pitch_name)), linewidth = 0.8, alpha = 0.6) +
  geom_point(aes(shape = factor(year)), size = 3) +
  geom_strikezone() +
  facet_wrap(~pitch_name) +
```

```

theme_bw() +
theme(legend.position = 'bottom') +
labs(
  x = "Average Horizontal Placement",
  y = "Average Vertical Placement",
  color = "Pitcher",
  shape = "Year"
)

```

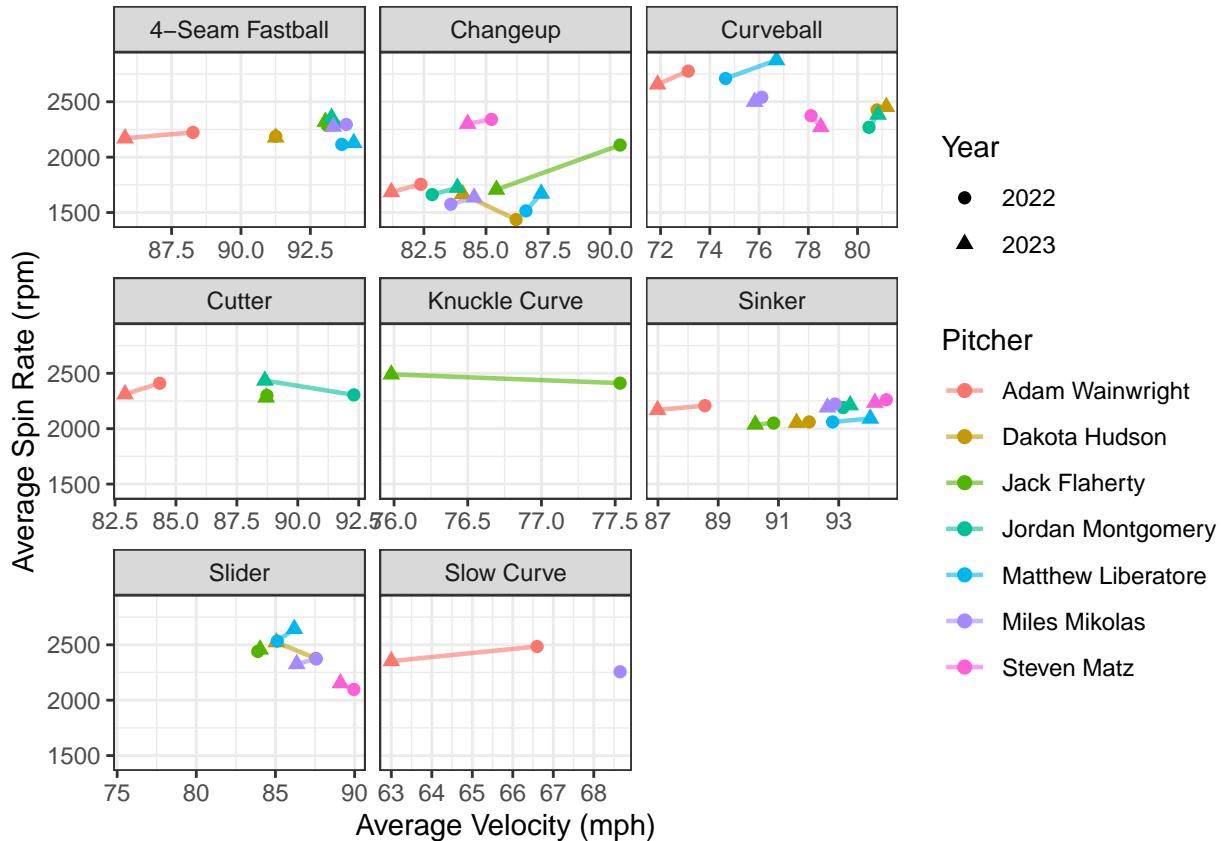


Now looking at balls in play you can see that there were definite differences compared to previous years. For the fast ball you can see that the 4 seam stayed about the same with being middle middle, while the sinkers moved towards the left side of home plate for the catcher. These small differences in trying to hit their spot shows that the differences can make a huge impact on hits.

```

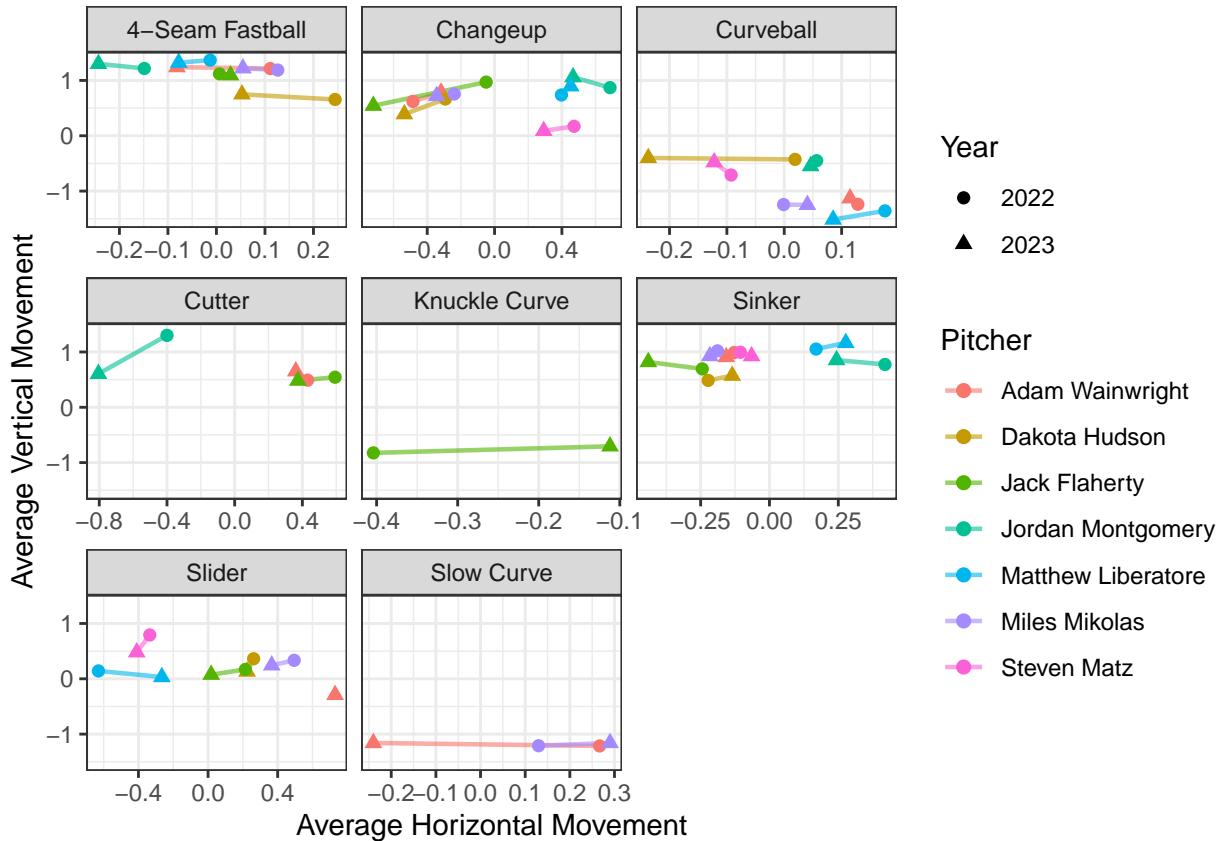
ggplot(cards_play, aes(x = avg_velo, y = avg_spin, color = MLBNAME)) +
  geom_line(aes(group = interaction(MLBNAME, pitch_name)), linewidth = 0.8, alpha = 0.6) +
  geom_point(aes(shape = factor(year)), size = 2) +
  facet_wrap(~pitch_name, scales = 'free_x') +
  theme_bw() +
  labs(
    x = "Average Velocity (mph)",
    y = "Average Spin Rate (rpm)",
    color = "Pitcher",
    shape = "Year"
)

```



When looking deeper into balls in play you can see the major differences in the spin rates and velocity. Looking at Adam Wainwright's fast balls, he was averaging around 87 for his sinker while having average spin rates. For the rest of the pitchers you can that changeups all decreased in their spin rates causing the ball to be able to pick up easier and thus more easily hit. These decreases in pitching velocity and spin rates did likely cause the ball to be hit more as these trends show that the pitchers were pitching slower with less spin on the ball.

```
ggplot(cards_play, aes(x = avg_x, y = avg_vertical, color = MLBNAME)) +
  geom_line(aes(group = interaction(MLBNAME, pitch_name)), linewidth = 0.8, alpha = 0.6) +
  geom_point(aes(shape = factor(year)), size = 2) +
  facet_wrap(~pitch_name, scales = 'free_x') +
  theme_bw() +
  labs(
    x = "Average Horizontal Movement",
    y = "Average Vertical Movement",
    color = "Pitcher",
    shape = "Year"
  )
```



Finally looking into balls in play movements we can see the differences between how these pitchers moved the ball. We already mentioned the decrease in spin rates for pitchers, so it is no surprise that for sliders we see the movements trending toward no movement. The decreased spin rates of the changeup and slider lead to less horizontal movement and vertical movement on pitches that are supposed to be moving during the pitch. These are breaking balls and off speed that do not seem to move. Looking at the sinker which the Cardinals seem to love so much, you can see that the movement is very small with pitches only moving 0.25 ft compared to cutters and changeups moving far more. A sinker needs to have run for the pitch to fool the batters eyes and miss judge the ball. It is weird that sinkers appear to be having positive vertical movement when they are supposed to run arm side and fall slightly. The Cardinals seem to throw standard two seams that sometimes tend to rise more than they run. You can see with spin rates and velocity dropping that it is not surprising that they do not run as much as they would probably like to.

Conclusion

Looking at all of the data that came from the Cardinals 2023 season, it appears that the Cardinals got ‘caught’ and ‘exposed’ during the 2023 season. There were some rule changes that drastically hurt the Cardinals idea of pitching contact style rather than strikeouts. The ban of the shift was a big one that I will show below.

```
Cards_starters_sc %>%
  group_by(year, if_fielding_alignment) %>%
  summarise(tot = n(), .groups = 'drop')
```

```
## # A tibble: 8 x 3
##   year if_fielding_alignment tot
##   <dbl> <chr>                <int>
## 1 2022 Infield shade          295
## 2 2022 Infield shift         3773
```

## 3	2022 Standard	6385
## 4	2022 Strategic	1248
## 5	2023 Infield shade	3076
## 6	2023 Standard	8680
## 7	2023 Strategic	1169
## 8	2023 <NA>	184

As you can see the Cardinals had to go from using the shift on 3773 pitches for their starters to not being able to use it a single time. They had to rely on an infield shade, which is just moving the second basemen or shortstop closer to second base, but that is not the same as shifting the infield to having more than 2 infielders on a side of second base. The Cardinals outdated ideas of pitch for contact were exposed during this season when there was not a way to defend the same way. The Cardinals gave up more hits than they had in recent history and that is probably partially due to these factors.

Looking back at the Statcast data you can see an aging rotation that had decreased velocity, spin rates, and movement on pitches. This roster was already old before the season started and there were worries about whether the old guys could do it, but it just seems like they were getting exposed with all of their flaws about their mindset of pitching and execution of pitches. Although there were lots of other problems with this team such as a lack of power and consistent hitting, the pitching was by far the worst problem on that field and was something that was not addressed. The pitching staff was one of the worst in the MLB with very low strikeout numbers, high runs against, and being one of the oldest rotations in the league.

You can see from all the graphs that the pitching continued to be worse than any previous year before it. Many people blamed the loss of Yadier Molina for why the pitching was so bad. There were rumors that he was saving the pitchers and making them look better than they ever were. There are numbers to back that claim, but it still does not account for the lack of execution in pitches that were thrown. Yadier Molina did not make the pitchers velocities stay high while he was catcher. All he did was tell them what to throw and where to throw it, but the pitchers still had good velocities and spin rates. During the 2023 season almost all pitchers saw some sort of loss of velocity or spin rate on their pitches resulting in a lack of execution and declining performance. The problems with location and game calling do play a factor, but it still does not account for multiple MPH losses on fastballs and less movement on their pitches than in previous years.

There is still a lot of data to explore and other aspects of this team to look at, but looking at the starting pitchers data, it was easily seen that they were performing worse and some causes that could have led to it. I do not believe it was Willson Contreras's fault that the pitching was so bad at the start of the season, he was a scapegoat for the pitchers and there is data to show that they were not as good.