Joe Musgrove Pitching Data

Cory Gargan

2022-12-08

Taking a Closer Look at Joe Musgrove's Spinrate

Did he use a foreign substance to increase spinrate, and if so, did it have a significant impact on his Post Season? During the 2022 MLB Playoffs vs. the NY Mets, the Umpires were sent by Mets skipper Buck Showalter to check Pitcher Joe Musgrove for the use of a foreign substance.

The Umpires quick check of the pitcher's ears resulted in a dismissal of the allegation, but was that truly the case?

Note: To create my environment I loaded the following libraries:

- `tidyverse`

- `skimr`

- `janitor`

- `gridExtra`

Along with the dataset:

- All Data was collected from MLB Baseball Savant
- `MLB_Musgrove-Joe_Reg-Post-Season_Pitch-Data`
    - This dataset can be found in the github project folder

```
library(tidyverse)
library(skimr)
library(janitor)
library(gridExtra)
spinrate <- read_csv("~/Documents/Data-Analysis/Github/Portfolio-Projects/01-MLB-Spinrate-Data-in-R/MLB_
```

Average Spinrate, Regular vs NY Mets Post Season Game (10/09/2022)

When looking at the Data there was a clear increase in Post Season Spinrate.

- 4-Seam Fastball +106.71 RPM

- Changeup +176.64 RPM

- Curveball +171.15 RPM

- Cutter +116.67 RPM

- Sinker +105.55 RPM

- **Slider +249.19 RPM**

```r
# Game in Question: Looking at Spinrates, Regular Season vs  Post Season Mets Game (10/09/22)
reg_slider_sr <- spinrate %>%
  group_by(season_type, pitch_name) %>%
  filter(season_type == 'Regular Season') %>%
  summarize(avg_spinrate = format(round(mean(release_spin_rate), 2), nsmall=2))
reg_slider_sr
```

```
## # A tibble: 6 x 3
## # Groups:   season_type [1]
##   season_type    pitch_name       avg_spinrate
##   <chr>          <chr>            <chr>
## 1 Regular Season 4-Seam Fastball  2559.32
## 2 Regular Season Changeup         1971.61
## 3 Regular Season Curveball        2721.77
## 4 Regular Season Cutter           2581.22
## 5 Regular Season Sinker           2435.65
## 6 Regular Season Slider           2714.75
```
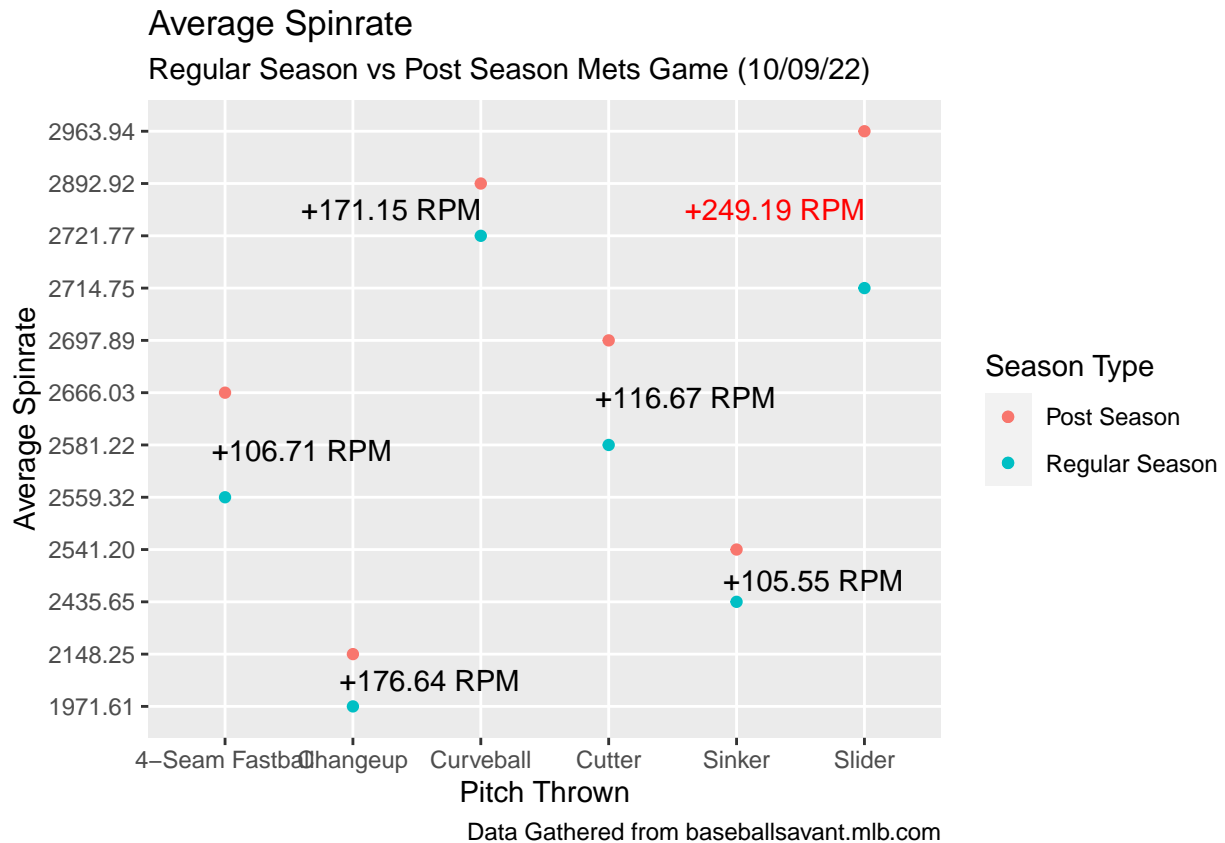
```r
mets_slider_sr <- spinrate %>%
  group_by(season_type, pitch_name) %>%
  filter(game_date == '10/9/2022') %>%
  summarize(avg_spinrate = format(round(mean(release_spin_rate), 2), nsmall=2))
mets_slider_sr
```

```
## # A tibble: 6 x 3
## # Groups:   season_type [1]
##   season_type pitch_name       avg_spinrate
##   <chr>       <chr>            <chr>
## 1 Post Season 4-Seam Fastball  2666.03
## 2 Post Season Changeup         2148.25
## 3 Post Season Curveball        2892.92
## 4 Post Season Cutter           2697.89
## 5 Post Season Sinker           2541.20
## 6 Post Season Slider           2963.94
```

```r
comb <- data.frame(reg_slider_sr, mets_slider_sr)

# Graphing the above Spinrates
ggplot(comb) +
  geom_point(aes(pitch_name, avg_spinrate, color = season_type)) +
  geom_point(aes(pitch_name.1, avg_spinrate.1, color = season_type.1)) +
  labs(title='Average Spinrate', subtitle='Regular Season vs Post Season Mets Game (10/09/22)',
       color='Season Type', x='Pitch Thrown', y='Average Spinrate',
       caption='Data Gathered from baseballsavant.mlb.com') +
  annotate('text', x=5.3, y=10.5, label='+249.19 RPM', color = 'red') +
  annotate('text', x=2.3, y=10.5, label='+171.15 RPM') +
  annotate('text', x=1.6, y=5.9, label='+106.71 RPM') +
  annotate('text', x=2.6, y=1.5, label='+176.64 RPM') +
  annotate('text', x=4.6, y=6.9, label='+116.67 RPM') +
  annotate('text', x=5.6, y=3.4, label='+105.55 RPM')
```

Data Gathered from baseballsavant.mlb.com

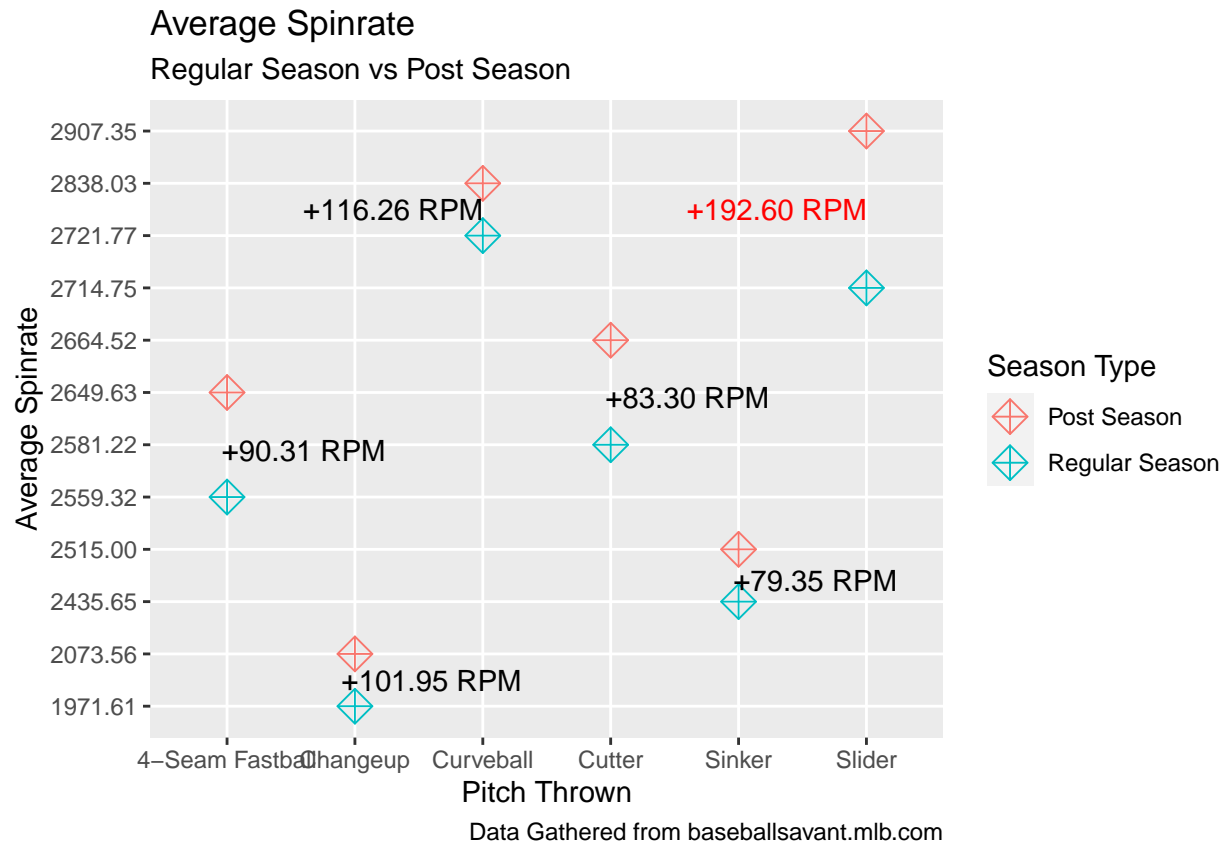**Average Spinrate, Regular vs Post Season**

When looking at the Data there was a clear increase in Post Season Spinrate.

- 4-Seam Fastball +90.31 RPM

- Changeup +101.95 RPM

- Curveball +116.26 RPM

- Cutter +83.30 RPM

- Sinker +79.35 RPM

- **Slider +192.60 RPM**

```r
# Average Spinrates by Pitch, Regular vs Post Season
avg_spin <- spinrate %>%
  group_by(season_type, pitch_name) %>%
  summarize(avg_spinrate = format(round(mean(release_spin_rate), 2), nsmall=2)) %>%
  arrange(desc(season_type))

# Scatter Plot: Average Spinrate by Pitch, Regular vs Post Season
ggplot(avg_spin, aes(pitch_name, avg_spinrate, color = season_type)) +
  geom_point(size = 4, shape = 9) +
  labs(title='Average Spinrate', subtitle='Regular Season vs Post Season',
       color='Season Type', x='Pitch Thrown', y='Average Spinrate',
```

```
        caption='Data Gathered from baseballsavant.mlb.com') +
annotate('text', x=5.3, y=10.5, label='+192.60 RPM', color = 'red') +
annotate('text', x=2.3, y=10.5, label='+116.26 RPM') +
annotate('text', x=1.6, y=5.9, label='+90.31 RPM') +
annotate('text', x=2.6, y=1.5, label='+101.95 RPM') +
annotate('text', x=4.6, y=6.9, label='+83.30 RPM') +
annotate('text', x=5.6, y=3.4, label='+79.35 RPM')
```

Average Spinrate
Regular Season vs Post Season

Data Gathered from baseballsavant.mlb.com

Average Speed MPH, Regular vs Post Season

When looking at the Data there was also a increase in Post Season Pitch Speed.

- 4-Seam Fastball +0.67 MPH

- **Changeup +1.58 MPH**

- Curveball +0.55 MPH

- Cutter +0.71 MPH

- Sinker +1.14 MPH

- Slider +0.16 MPH

```
# Average Speed MPH by Pitch, Regular vs Post Season
avg_speed <- spinrate %>%
  group_by(season_type, pitch_name) %>%
```

```
    summarize(average_mph = format(round(mean(effective_speed, na=TRUE), 2), nsmall=2)) %>%
    arrange(desc(season_type))

# Scatter Plot: Average Speed MPH by Pitch, Regular vs Post Season
ggplot(avg_speed, aes(pitch_name, average_mph, color = season_type)) +
  geom_point(size = 4) +
  labs(title='Average Pitch Speed, MPH', subtitle='Regular Season vs Post Season',
       color='Season Type', x='Pitch Thrown', y='Average Speed, MPH',
       caption='Data Gathered from baseballsavant.mlb.com')
```

Average Pitch Speed, MPH
Regular Season vs Post Season

Data Gathered from baseballsavant.mlb.com
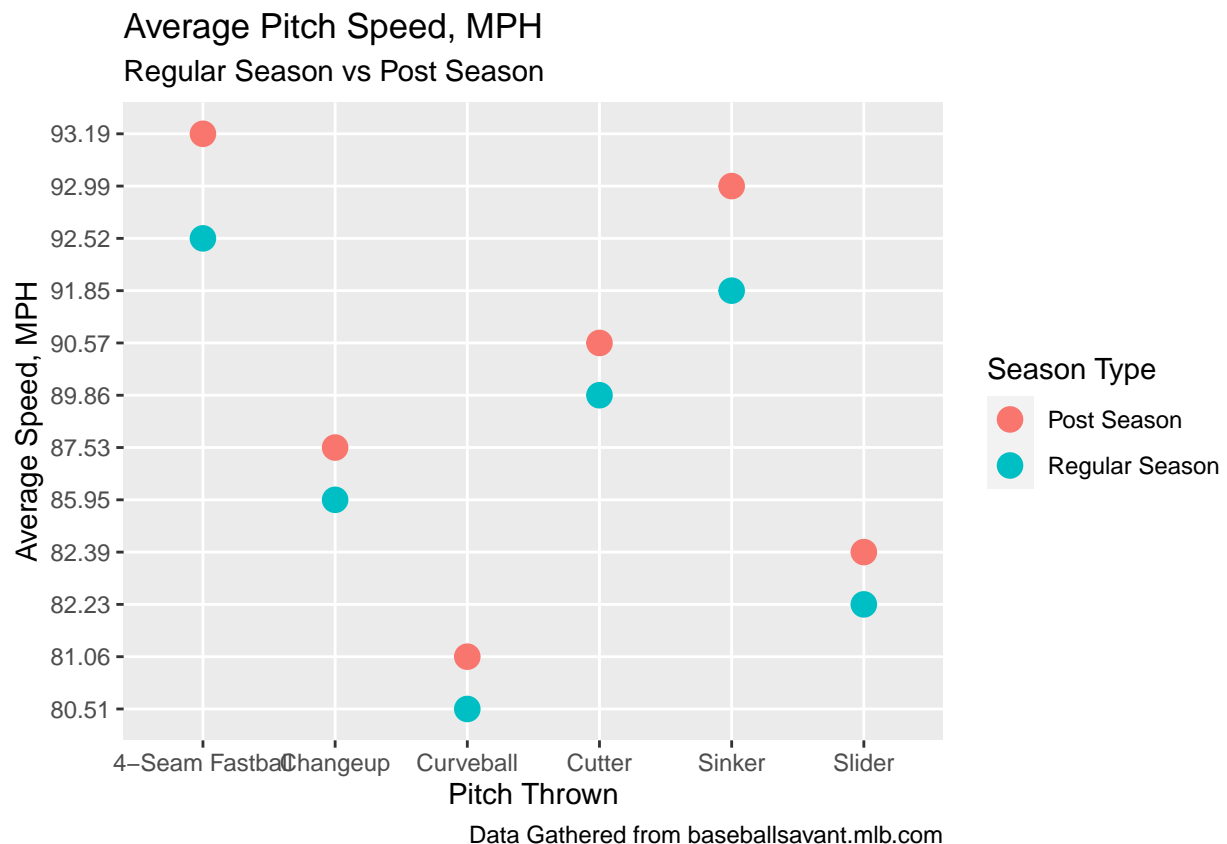
With Two Strikes: Percent Thrown for Strikeout, Regular vs Post Season

The Strikeout percentages decreased for all pitches, **except** the Slider.

- 4-Seam Fastball -0.02%

- Changeup -9.72%

- Curveball -5.81%

- Cutter -0.71%

- Sinker -14.81%

- **Slider +2.41%**

```
# With Two Strikes: Percent Thrown for Strikeout by Pitch, Regular vs Post Season
two_strikes <- spinrate %>%
  group_by(season_type, pitch_name) %>%
  filter(strikes == 2) %>%
  summarize(strikeouts = sum(at_bat_outcome == 'strikeout', na.rm=TRUE),
            times_thrown_2strikes = sum(strikes == 2),
            strikeout_pct = format(round((strikeouts/times_thrown_2strikes)*100, 2), nsmall=2)) %>%
  arrange(desc(season_type))

# Column Chart: With Two Strikes: Percent Thrown for Strikeout by Pitch, Regular vs Post Season
ggplot(two_strikes, aes(pitch_name, strikeout_pct, fill = season_type)) +
  geom_col(position = 'dodge') +
  labs(title='With Two Strikes: Percent Thrown for Strikeout', subtitle='Regular Season vs Post Season'
       fill='Season Type', x='Pitch Thrown', y='Strikeout Percentage',
       caption='Data Gathered from baseballsavant.mlb.com') +
  annotate('text', x=5.3, y=11, label='Slider Was The') +
  annotate('text', x=5.3, y=10.6, label='Only Increase')
```

With Two Strikes: Percent Thrown for Strikeout
Regular Season vs Post Season

Data Gathered from baseballsavant.mlb.com
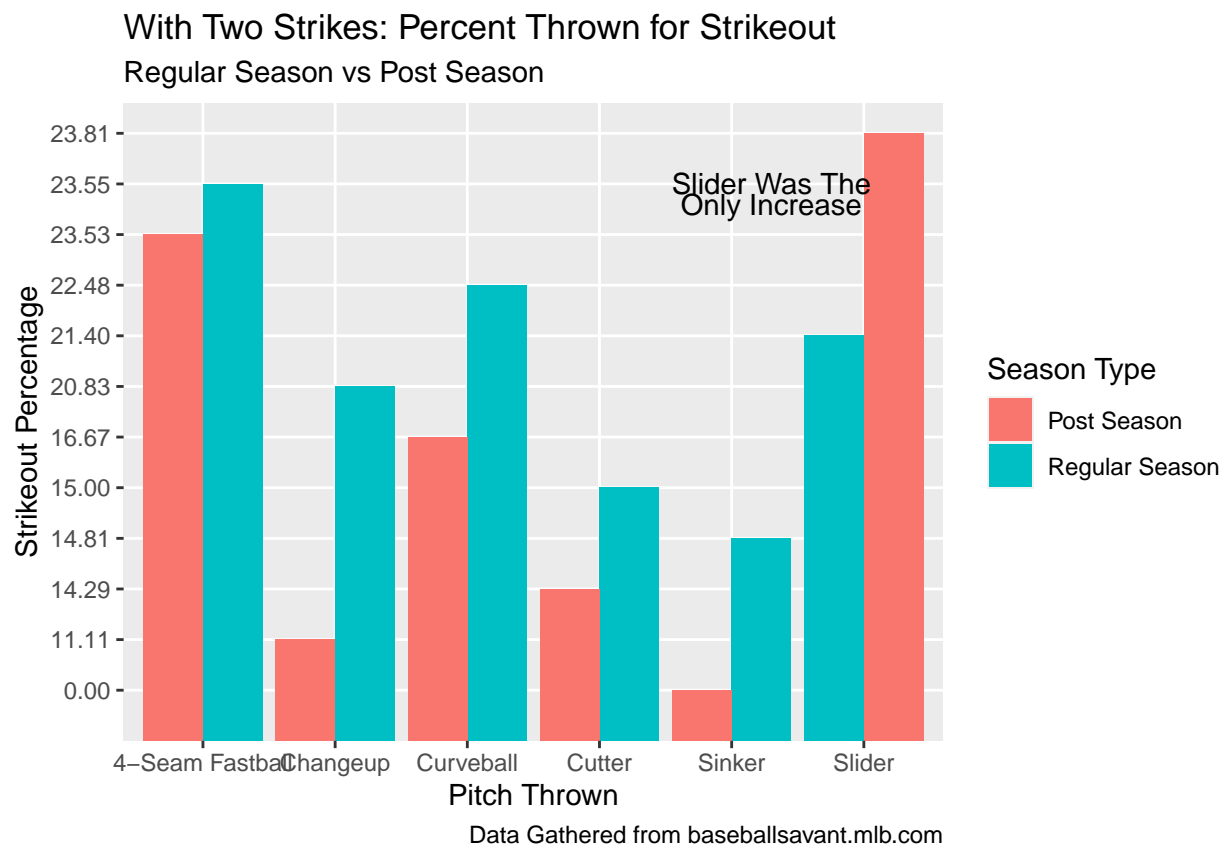
The Slider: Putting it under the Microscope

The Slider was the most improved pitch in regard to increased Spinrate and Strikeout Percentage.
The Speed also slightly increased, but not what I would consider a significant amount.

```
# Slider Strikeout Data
slider_strike <- spinrate %>%
  group_by(season_type, pitch_name) %>%
```

```r
  filter(strikes == 2, pitch_name == 'Slider') %>%
  summarize(strikeouts = sum(at_bat_outcome == 'strikeout', na.rm=TRUE),
            times_thrown_2strikes = sum(strikes == 2),
            strikeout_pct = format(round((strikeouts/times_thrown_2strikes)*100, 2), nsmall=2)) %>%
  arrange(desc(season_type))

# Slider Spinrate Data
slider_sr <- spinrate %>%
  group_by(season_type, pitch_name) %>%
  filter(pitch_name == 'Slider') %>%
  summarize(avg_spinrate = format(round(mean(release_spin_rate), 2), nsmall=2)) %>%
  arrange(desc(season_type))

# Plotting Strikeout & Spinrate Data
plot1 <- ggplot(slider_strike, aes(pitch_name, strikeout_pct, fill = season_type)) +
  geom_col(position = 'dodge') +
  labs(title='Two Strikes: Strikeout Percentage', subtitle='Regular Season vs Post Season',
       fill='Season Type', x='Pitch Thrown', y='Strikeout Percentage',
       caption='Data Gathered from baseballsavant.mlb.com')

plot2 <- ggplot(slider_sr, aes(pitch_name, avg_spinrate, fill = season_type)) +
  geom_col(position = 'dodge') +
  labs(title='Average Spinrate, Slider', subtitle='Regular Season vs Post Season',
       fill='Season Type', x='Pitch Thrown', y='Average Spinrate',
       caption='Data Gathered from baseballsavant.mlb.com')

# Combining the Two Graphs Side-by-Side
grid.arrange(plot1, plot2, ncol=2)
```
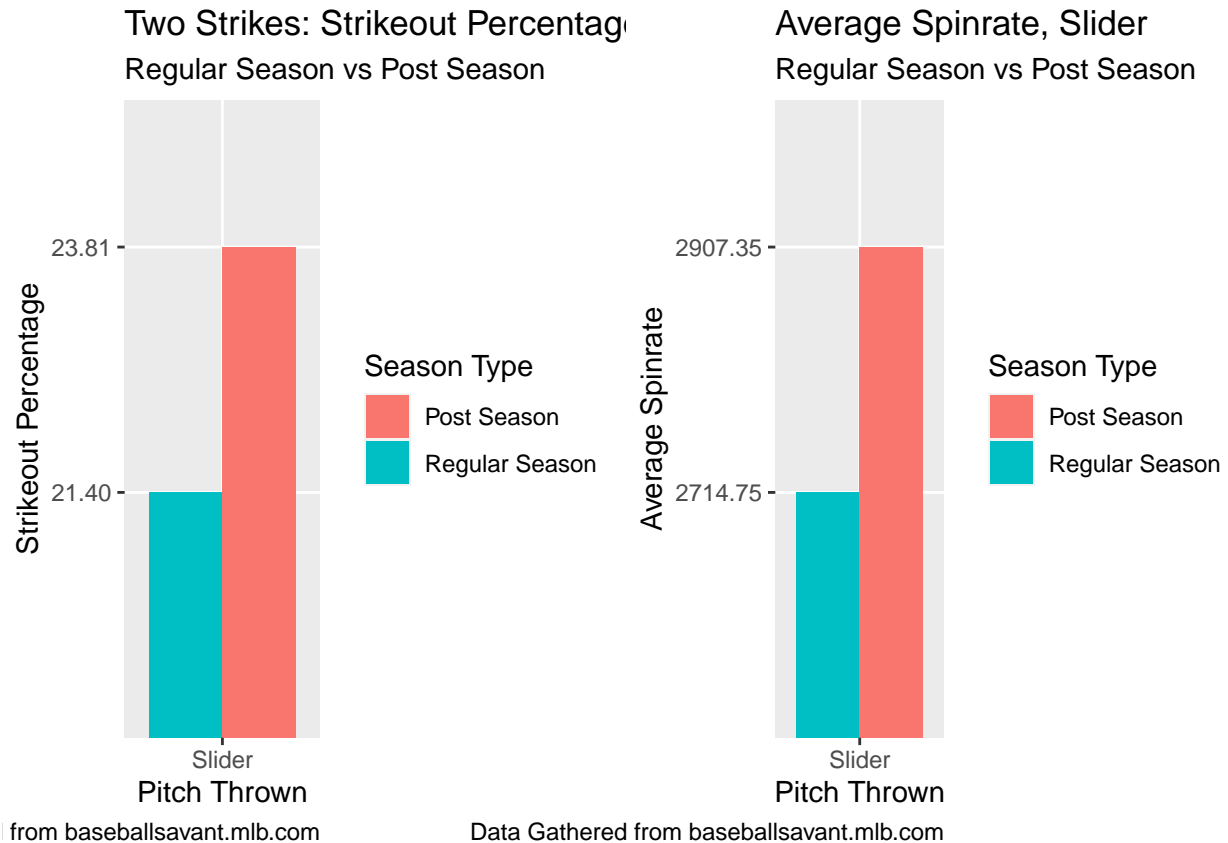
Two Strikes: Strikeout Percentage

Regular Season vs Post Season

from baseballsavant.mlb.com

Average Spinrate, Slider

Regular Season vs Post Season

Data Gathered from baseballsavant.mlb.com

In Conclusion:

Although there was never an official investigation into the pitcher's use of a foreign substance during the game, I believe the data tells a different story.

I am of the opinion that the increases in Spinrate and Speed across all pitches points towards the use of a foreign substance on the ball.

In particular with the Slider, the significant Spinrate increase made it a more effective pitch based on the Post Season Strikeout Percentages.

We may never find out the truth, but the data does not lie!