Quantifying The "Different Look":

Measuring The Success of Altering Pitcher Release Angle From PA to PA

Presented By: Connor Dwyer, Niko Konstantellis
Peter Majors, Meg McAlary

Research Question:

Do higher degrees of separation in arm angles from preceding plate appearances contribute to reduced offensive output?

Sports Analytics Society

Hypothesis:

Teams with a higher mean average release angle difference from plate appearance to plate appearance are better at reducing offensive output.

Definitions:

- **Release Angle**: the vertical degree of the ball as the pitcher is releasing it (0° = Overhead, 90° = Sidearm)

- **wOBA**: a rate statistic which credits hitters for the value of outcomes relative to their run-scoring environment, best measure of offensive production

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

Data Acquisition:

- Used the **mlb_pbp()** function from the **baseballr** package to pull all pitch-by-pitch data from the 2021 season (746,106 pitches)
- Inner joined pitch-by-pitch data on data frame of release angles of all pitches from the 2021 season

High Level Methodology:

- Goal - Determine if there is any correlation between a separation in arm angles and offensive output (scaled wOBA)
- Methodology - Determine Pearson R and P value for three correlation scenarios
    - Overall effect of release angle difference on scaled wOBA
    - Effect of release angle difference on scaled wOBA where difference <55 degrees
    - Effect of release angle differences on scaled wOBA by team

Data Preparation (1):

- Created a primary key for each plate appearance **“pa_pk”** using:
  - game id, batter id, pitcher id, and at-bat index (within game).

- Created a **“pa_order”** column that denotes the order of plate appearance within each game for each batter using and joined on main data frame:
  - game id, batter id, and pitch start time of one pitch in plate appearance

```r
#Create PA Order Column By Estimated Starting Time Of Pitch
 mlb_pbp_2021_pa <- mlb_pbp_2021_pa %>%
   group_by(game_pk.x, matchup.batter.id) %>%
   mutate(pa_order = order(about.startTime), decreasing = FALSE)
```

Data Preparation (2):

- Found average release angle for each plate appearance grouped by **“pa_pk”**
  - Named this column **“avg_relangle_pa”** and joined on main data frame

- Created binary columns in main data frame based on whether certain events took place, used this in eventual calculation of wOBA

```python
#Create Binary Columns For Each Input In The wOBA Formula
pbp['single_binary'] = [1 if x == 'Single' else 0 for x in pbp['result.event']]
pbp['double_binary'] = [1 if x == 'Double' else 0 for x in pbp['result.event']]
pbp['triple_binary'] = [1 if x == 'Triple' else 0 for x in pbp['result.event']]
pbp['homerun_binary'] = [1 if x == 'Home Run' else 0 for x in pbp['result.event']]

pbp['walk_binary'] = [1 if x == 'Walk' else 0 for x in pbp['result.event']]
pbp['hbp_binary'] = [1 if x == 'Hit By Pitch' else 0 for x in pbp['result.event']]
pbp['sacf_binary'] = [1 if x == 'Sac Fly' else 0 for x in pbp['result.event']]
pbp['intentional_binary'] = [1 if x == 'Intent Walk' else 0 for x in pbp['result.event']]
```

Sports Analytics Society

Data Preparation (3):

- Classified events based on results as at-bats or plate appearances for calculation of wOBA (AB, HBP, uBB, IBB, Sac Fly), removed non-instances (cut 6,201 rows)
- Found last pitch of plate appearance and only kept those rows as to ensure plate appearance events are non-duplicates (189,165 plate appearances (rows remaining))

```
#Group Each At-Bats by Its pk And Find The Max For Each Group By Pitch Number
last_pitchpa = pbp.groupby(['pa_pk'], sort = False)['pitchNumber'].max()
last_pitchpa = pd.DataFrame(last_pitchpa)

#Inner Join The Last Pitches Of PA On The Total Data Frame
pbp_pa = pd.merge(pbp, last_pitchpa, on = 'pa_pk', how = 'inner')

#Create New Column Showing Whether The Pitch Columns Are The Same
pbp_pa['same'] = np.where(pbp_pa['pitchNumber_x'] == pbp_pa['pitchNumber_y'], 1, 0)

#Filter The Main Dataframe By That Condition
pbp_pa = pbp_pa.loc[pbp_pa['same'] == 1]
```

Data Preparation (4):

- Calculated **wOBA** for all plate appearances using 2021 linear weights

```
#Caclulate wOBA Based On Above Binary Inputs
wBB, wHBP, w1B, w2B, w3B, wHR = .692, .772, .879, 1.242, 1.568, 2.007

pbp_pa['wOBA'] = ((pbp_pa['walk_binary'] * wBB) + (pbp_pa['hbp_binary'] * wHBP) + (pbp_pa['single_binary'] * w1B) +
                  (pbp_pa['double_binary'] * w2B) + (pbp_pa['triple_binary'] * w3B) + (pbp_pa['homerun_binary'] * wHR)) /
                  (pbp_pa['ab_binary'] + (pbp_pa['walk_binary'] - pbp_pa['intentional_binary']) + pbp_pa['sacf_binary'] +
                   pbp_pa['hbp_binary'])
```

- Calculated **"wOBA+"** for each plate appearance, which is how much better a given plate appearance's wOBA is than average wOBA, scaled so that 100 is league average

```
#Find The League Average wOBA
wOBA_avg = pbp_pa['wOBA'].mean()

#Create wOBA+ For Each PA
pbp_pa['wOBA+'] = (pbp_pa['wOBA'] / wOBA_avg) * 100
```

Data Preparation (5):

- Scale "wOBA+" to 100 based on bias given between plate appearances & wOBA
  - Since wOBA+ increases with plate appearances, and so does average release angle difference, we isolated our independent variable as best we could

```
#Scale wOBA+ In Each PA Back To League Average (Remove PA Bias)
pbp_pa_order['wOBA+_scaled'] = ""

#Initialize i and j
i = 1

#While Loop To Scale Back wOBAs
while i <= len(pbp_pa_order):
    if pbp_pa.at[i, 'pa_order'] == 1:
        pbp_pa.at[i, 'wOBA+_scaled'] = pbp_pa.at[i, 'wOBA+'] * 1.04956608
    elif pbp_pa.at[i, 'pa_order'] == 2:
        pbp_pa.at[i, 'wOBA+_scaled'] = pbp_pa.at[i, 'wOBA+'] * 1.0007180
    elif pbp_pa.at[i, 'pa_order'] == 3:
        pbp_pa.at[i, 'wOBA+_scaled'] = pbp_pa.at[i, 'wOBA+'] * 1.00422351
    elif pbp_pa.at[i, 'pa_order'] == 4:
        pbp_pa.at[i, 'wOBA+_scaled'] = pbp_pa.at[i, 'wOBA+'] * 0.97448902
    elif pbp_pa.at[i, 'pa_order'] == 5:
        pbp_pa.at[i, 'wOBA+_scaled'] = pbp_pa.at[i,'wOBA+'] * 0.85051935
    elif pbp_pa.at[i, 'pa_order'] == 6:
        pbp_pa.at[i, 'wOBA+_scaled'] = pbp_pa.at[i, 'wOBA+'] * 0.85708535
    elif pbp_pa.at[i, 'pa_order'] == 7:
        pbp_pa.at[i, 'wOBA+_scaled'] = pbp_pa.at[i, 'wOBA+'] * 0.7282736
    i = i + 1
```

pa_order	wOBA+
1	95.277469
2	99.928249
3	99.579425
4	102.617883
5	117.575220
6	116.674495
7	137.311032

pa_order	relangle_diff_prevPA
1	0.000000
2	10.060561
3	13.685594
4	15.986429
5	17.098366
6	16.642297
7	23.455804

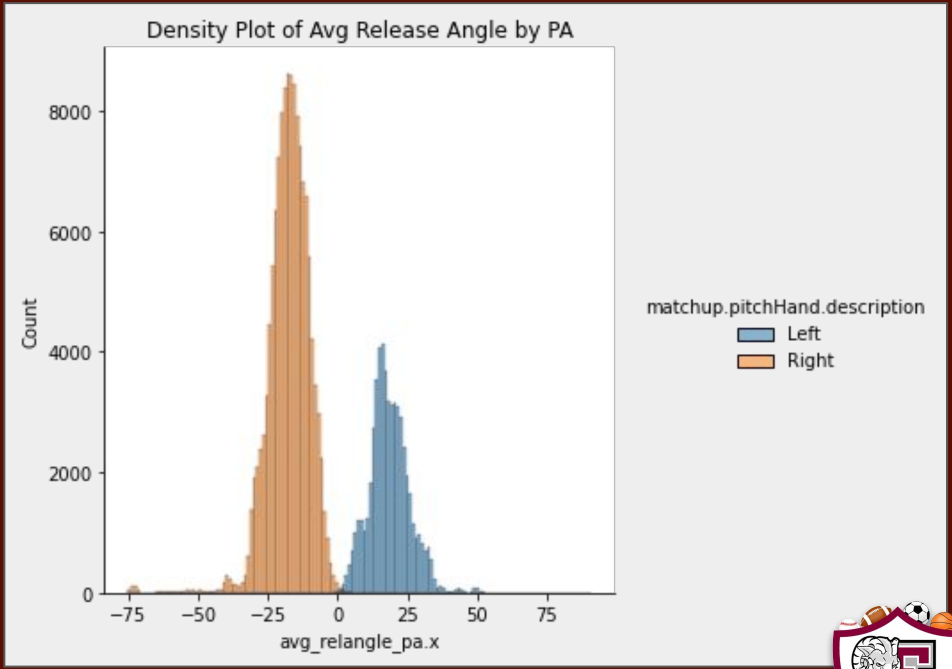
Data Preparation (6):

- Order game ids, batter ids, and “pa_order” in ascending order and calculate difference between average release angle preceding and current plate appearance
  - Used the absolute value of this difference, did not factor direction of change

	game_pk.x.x	matchup.batter.id.x	pa_order	pa_pk	avg_relangle_pa.x	relangle_diff_prevPA
0	632169	446334	1	632169-446334-592346-8	-16.598744	
1	632169	446334	2	632169-446334-592346-24	-16.503940	0.094804
2	632169	446334	3	632169-446334-641386-48	7.642938	24.146878
3	632169	446334	4	632169-446334-571710-64	-27.250800	34.893738
4	632169	453568	1	632169-453568-657277-3	-22.630430	0
...	...	...	...	...	...	...
189160	660938	657077	4	660938-657077-605280-24	-11.334337	4.727603
189161	660938	666915	1	660938-666915-543037-15	-16.467126	0
189162	660938	666915	2	660938-666915-605280-30	-8.957674	7.509452
189163	660938	666915	3	660938-666915-642528-47	-21.320397	12.362724
189164	660938	666915	4	660938-666915-643338-62	-13.902478	7.417919

Exploratory Data Analysis (LHP vs RHP):

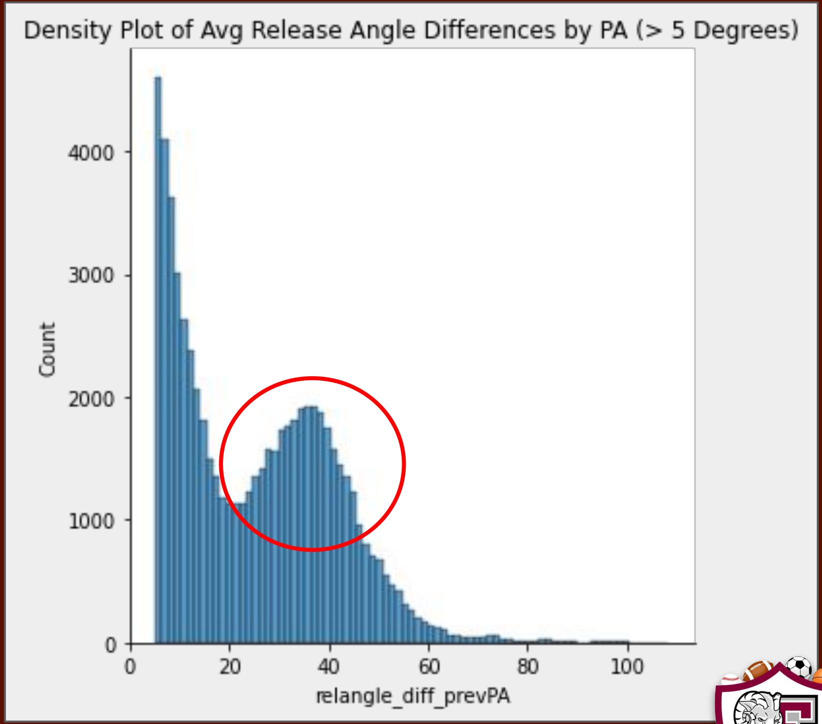
- **LHP**:
  - Avg. Rel. Angle: 18.57
  - Standard Dev: 7.32
- **RHP**:
  - Avg. Rel. Angle: -17.57
  - Standard Dev.: 7.66
- Lefties typically throw from a lower slot, have less variance



Exploratory Data Analysis (RA Differences):

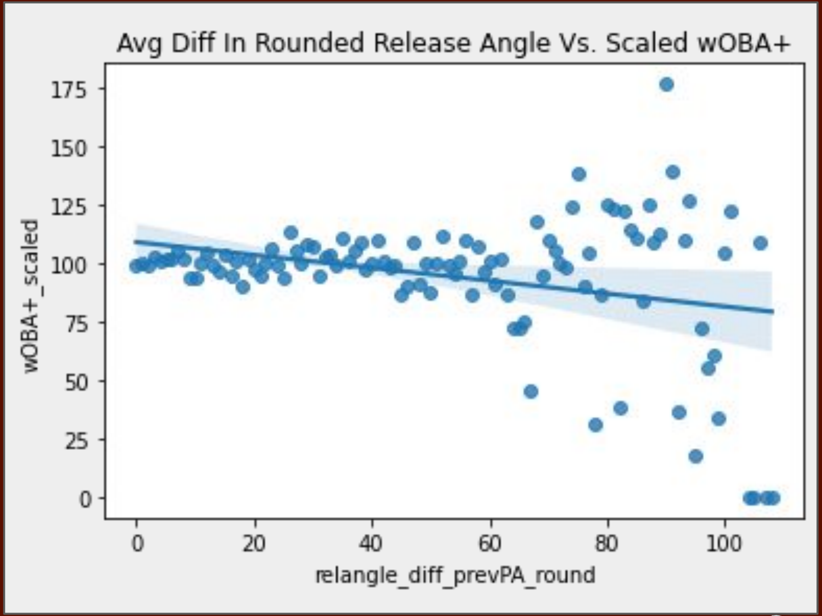
- Most differences in release angle are extremely small (< 5 degrees)
    - Consecutive PA, same pitcher
    - Many pitchers with similar RA

- Area in the circle represents trend of managers using opposite handed pitchers, or pitchers with significantly different release angles, consecutively



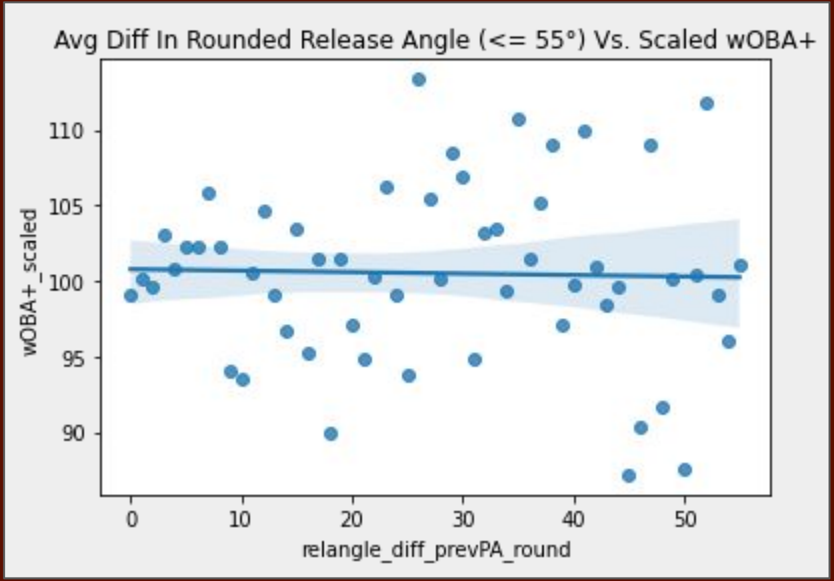
Effect of Release Angle Differences (Total)

- In the context of all possible release angle differences, there is a very slight correlation
  - Pearson R: .30, P-Value: .0017
- Observations further to the right have fewer PA, much less reliable



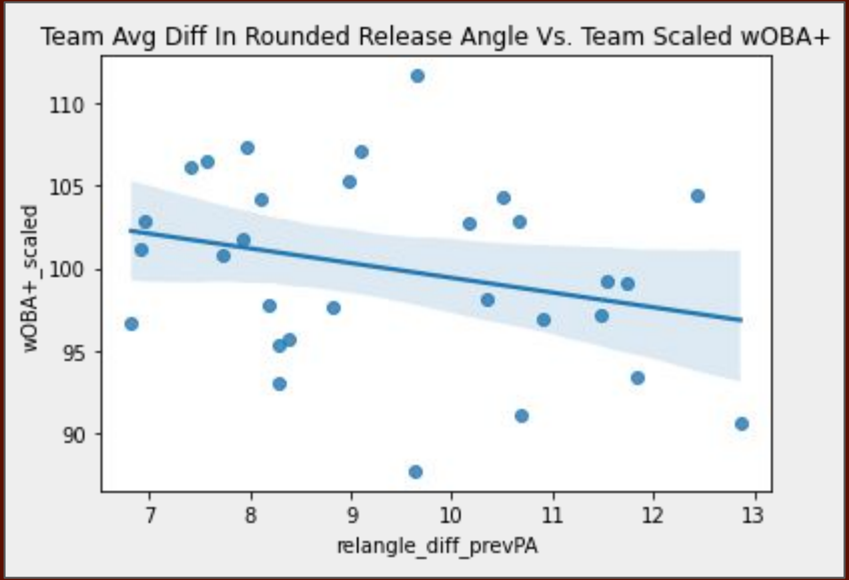
Effect of Release Angle Differences (<55°)

- Removed observations with more than 55° of difference between previous PA
  - 189,165 to 187,399 (0.1% Dec.)
- No correlation whatsoever
  - Pearson R: -.027, P-Value: .8417
- Suggests that any statistical significance was being "carried" by largest differences



Effect of Release Angle Differences (By Team)

- Explored how teams utilize release angle differences, examined how this may play into pitching staff success or failure
- Allows us to contextualize wOBA+
- Limited correlation, unreliable regression
    - Pearson R: -.28, P-Value: .1358



Release Angle Differences (Top 5 vs. Bottom 5)

- Examined top 5 and bottom 5 teams in terms of release angle difference
  - Top 5: 97.43 Scaled wOBA+
  - Bottom 5: 102.67 Scaled wOBA+
- 2 of the top 4 pitching teams are in the top 3 of average release angle difference
- The extremes tell a different story!

	fielding_team	relangle_diff_prevPA	wOBA+_scaled
0	Houston Astros	6.819697	96.686683
1	Philadelphia Phillies	6.919084	101.158995
2	Kansas City Royals	6.960549	102.836476
3	Colorado Rockies	7.418060	106.117520
4	Pittsburgh Pirates	7.569527	106.552760
25	San Diego Padres	11.546512	99.201899
26	Seattle Mariners	11.730976	99.067537
27	Tampa Bay Rays	11.842931	93.432930
28	Los Angeles Angels	12.438276	104.430653
29	San Francisco Giants	12.863579	90.605254

Our Conclusions

- Although there is not a particularly strong league wide correlation overall between release angle differences and wOBA+, the team data reveals interesting conclusions.
- Teams who appear to have focused on utilizing and sequencing pitchers with larger differences in arm angles (Giants, Angels, Rays, etc.) tended to hold opponents to a lower offensive output.
- Here, it took the most extreme examples to prove that there is some validity to drastically changing arm angles from PA to PA.

Remaining Questions

- How can we combine these findings with other variables, such as changes in velocity from PA to PA to build a model which reduces offensive output in subsequent plate appearances?
- Should teams spend more capital on pitchers with more unique deliveries or is there simply not enough already effective pitchers with unique deliveries to make the strategy worth it?
    - In other words, have the Rays and Giants already taken all the most talented unique throwers for 2022? After all, our research shows that it seems to be an “all or nothing” type of strategy.