OLYAS: A New and Improved Game Score Metric

By Jake Federman, Eshan Mehere, Aidan Resnick, and Maxwell Resnick

Opening Case Example: Blind Resume

Start A	Start B
7 IP	3.2 IP
2 H	9 H
0 ER	10 ER
0 BB	2 BB
10 Ks	5 Ks
Game Score: 88	Game Score: 4

Opening Case Example: Blind Resume

Start A	Start B
23% Barrel%	0% Barrel%
39% HardHit%	33% HardHit%
73% Contact%	71% Contact%
46% Zone%	49% Zone%
61% F-Strike%	68% F-Strike%
16% SwStr%	16% SwStr%
0.952 Opponent Multiplier	1.003 Opponent Multiplier

OLYAS

O - Opponent

L - League

Y - Year

A - Adjusted

S - Score

Agenda

1. Discuss the current Game Score metric: its composition, its flaws, etc.
2. Explain our process of building OLYAS
   a. Why and how we built the metric the way we did
3. Analyze OLYAS and its relationships to other statistics; compare these relationships to those of Game Score
4. Revisit case examples to see OLYAS in action
5. Analyze and explore potential uses for OLYAS as a long-term statistic
6. Draw conclusions, think about applications, and speculate about potential next steps

Background: What is Game Score?

1. Game Score is a pitcher evaluation metric designed by Bill James.
2. Calculation:
    a. Start with 50 points
    b. Each out recorded: +1 point
    c. Each inning pitched after the fourth inning: +2 points
    d. Each strikeout recorded: +1 point
    e. Each hit allowed: -2 points
    f. Each earned run allowed: -4 points
    g. Each unearned run allowed: -2 points
    h. Each walk allowed: -1 point

Background: Game Score 2.0

1. In 2014, Tom Tango created a revised version of Game Score that provides different linear weights for each variable.
2. Calculation:
    a. Start at 40
    b. Each out: +2
    c. Each K: +1
    d. Each BB: -2
    e. Each H: -2
    f. Each R: -3
    g. Each HR: -6

Background: What is the issue with Game Score?

1. Arbitrary values/starting points
2. Metrics used are poor indicators of performance
3. Fails to account for context (does not *isolate* true performance)

Background: What does it mean to isolate performance?

1. Our goal is to truly boil our game score down to what the pitcher can control (K, BB, Quality of Contact).
   a. Statcast and zone metrics:
      i. Exit Velocity, Barrel %, Chase Rate, F-Strike %, etc.
2. In the past, metrics such as FIP and xwOBA have attempted to, and in large part, succeeded in isolating the pitcher's performance.
   a. However, we wanted to take it a step further by accomplishing this task on a start-to-start basis.

Process: Data

1. Dataset consists of 19,384 rows
    a. Each row represents an individual start since 2015 (when Statcast was implemented).
2. Rows contain:
    a. Categorical Information (date, opponent)
    b. Basic box score stats (innings pitched, hits, runs, walks, strikeouts, game score, etc.)
    c. Advanced Statcast and zone metrics (Exit Velocity, Launch Angle, Barrel%, O-Swing %, Z-Swing%, SwStr%, etc.)

Data obtained from Fangraphs.com

Process: Multivariate Regression Model

- After we compiled our dataset, we ran a linear regression model to determine how well thirteen metrics predict FIP-.

Statcast Metrics	Zone Metrics
• Exit Velocity • Launch Angle • Barrel% • HardHit%	• O-Swing% • Z-Swing% • Swing% • O-Contact% • Z-Contact% • Contact% • Zone% • F-Strike% • SwStr%

Why these metrics?

1. We DO care about *performance*:
    a. How well the pitcher actually pitches in a given start
2. We DO NOT care about *outcome*:
    a. How well the pitcher’s defense plays
    b. The dimensions/altitude of the pitcher’s ballpark

Process: Adjusting for Innings

1. Once we had the regression-predicted values for each of the 19,384 starts, we found that the *best* start since 2015 was a Clayton Kershaw near-perfect two-inning start.
2. While he pitched well, we realized that we did not reward pitchers for recording more outs.
3. Hence, we multiplied each start by (Outs Recorded /27).
    a. For negative values, we multiplied each start by (27/Outs Recorded).
4. To top it off, we adjusted the values proportionally to make 100 the average value (the FIP+ scale).

Our Revised Metric

1. At this point, we had a functioning metric that uses advanced metrics to determine a pitcher’s true performance
2. The results clearly supported our goal, which was to credit pitcher for having stronger Statcast and zone metrics instead of result-oriented statistics.
3. However, we had an epiphany…

Lucas Giolito’s No-Hitter

1. The second-best start in our dataset per our new metric was Lucas Giolito’s no-hitter in 2020 against the Pittsburgh Pirates.
2. While he undeniably had strong isolated metrics, we recognized that his no-hitter came against the worst team in baseball.
3. Should this be taken into account? Absolutely.

https://www.flickr.com/photos/8432336@N08/48055007196/in/photolist-2gdstqQ-2gAe8AK-GvkGng-Hp2hWb-2gAc7i2-2gAbQuU-GvkGWH-FA3RBD-GvkGCr-G6ceCQ-2gAdyb7-G6ceQJ-J58woh-Gn9Qru-J8f9gP-HGzUCG-HcjtFz-HGEazs-Hcdr9y-HYoWuf-HcjSjR-HGE1D1-HYjNHw-HGEx2L-J1Gcvv-J5d9T3-HGEvSw-J1BvGv-HcjLjV-HYoCMS-J5cYF9-HcjxxH-Jdvdxk

Process: Context-Based Adjustments

1. Since Game Score factors in just outcomes and not context, it can lead to misleading conclusions.
   a. A start that earns a 100 Game Score against the 2020 San Diego Padres should be evaluated differently than a start that earns a 100 Game Score against the 2020 Pittsburgh Pirates.
   b. A start that earns a 100 in 2015 is less impressive than a start that earns a 100 in 2020, for the run environments were very different.
2. By adjusting for league, year, and opponent, we grasped a more accurate picture of each start in our dataset.

Process: Context-Based Adjustments

1. First, we adjusted the starts in each league according to the year.
    a. Scaled the metric so the average start for a given year in a given league is rated 100
2. Then, we adjusted for opponent.
    a. Assigned each team a multiplier for each year
        i. Calculated by dividing the league average of our metric by a team's average (against) in a given year
    b. Then we multiplied the score of every start against a given team by that team's multiplier.
    c. We used a multiplied adjustment instead of an added adjustment to match the nature of the Plus system to which we scaled our statistic.

Results

dollars_olyas_summary

player	date	year	team	opp	IP	H	E	BI	SO	Barrel%	HardHit%	Zone%	Z-Swing%	LA	Contact%	game_score	OLYAS	lgyr_multiplier	team_multiplier
Jacob deGrom	9/6/2020	2020	NYM	PHI	7.00	3.	1	2.	12	0.08	0.25	0.43	0.85	5.20	0.47	76.00	414.78	1.25	1.45
Justin Verlander	5/1/2018	2018	HOU	NYY	8.00	3.	0	0.	14	0.00	0.31	0.53	0.63	29.60	0.60	96.00	379.56	1.08	1.23
Lucas Giolito	2019-08-21	2019	CHW	MIN	9.00	3.	0	0.	12	0.06	0.28	0.46	0.66	21.60	0.63	102.00	376.41	1.09	1.32
Chris Sale	4/27/2017	2017	BOS	NYY	8.00	8.	2	0.	10	0.00	0.35	0.55	0.72	6.10	0.63	74.00	365.36	1.04	1.16
Antonio Senzatel	2020-08-29	2020	COL	SDP	7.00	7.	0	1.	3.0	0.00	0.44	0.33	0.64	1.00	0.77	70.00	363.93	1.25	1.70
Gerrit Cole	9/8/2019	2019	HOU	SEA	8.00	1.	1	0.	15	0.00	0.20	0.47	0.56	5.30	0.44	94.00	362.88	1.09	1.03
Matt Shoemaker	2016-05-21	2016	LAA	BAL	7.10	3.	0	0.	12	0.00	0.29	0.62	0.68	11.40	0.57	90.00	359.53	1.01	1.18
Andrew Heaney	2020-09-03	2020	LAA	SDP	7.00	3.	0	2.	6.0	0.00	0.47	0.51	0.73	11.50	0.78	80.00	358.61	1.25	1.70
Max Scherzer	8/25/2016	2016	WSN	BAL	8.00	2.	0	0.	10	0.00	0.38	0.43	0.73	19.10	0.63	93.00	357.39	1.01	1.18
Charlie Morton	5/30/2019	2019	TBR	MIN	7.00	4.	2	0.	6.0	0.00	0.39	0.49	0.56	6.80	0.70	76.00	349.85	1.09	1.32
Gerrit Cole	4/29/2018	2018	HOU	OAK	6.20	6.	3	0.	12	0.00	0.39	0.48	0.60	14.70	0.58	71.00	345.51	1.08	1.26
Lucas Giolito	2020-08-25	2020	CHW	PIT	9.00	0.	0	1.	13	0.00	0.14	0.48	0.73	21.60	0.52	107.00	344.49	1.25	0.76
Jacob deGrom	8/23/2019	2019	NYM	ATL	7.00	4.	1	1.	13	0.00	0.25	0.32	0.71	10.10	0.55	82.00	342.48	1.05	1.20
Lucas Giolito	2019-05-23	2019	CHW	HOU	9.00	4.	0	1.	9.0	0.00	0.48	0.58	0.66	16.80	0.81	95.00	341.40	1.09	1.22
Masahiro Tanaka	6/23/2017	2017	NYY	TEX	8.00	3.	0	2.	9.0	0.00	0.19	0.36	0.69	3.30	0.60	88.00	340.97	1.04	1.12
Patrick Corbin	4/4/2018	2018	ARI	LAD	7.10	1.	0	1.	12	0.00	0.27	0.36	0.51	1.30	0.54	90.00	340.91	1.01	1.31
Luis Severino	5/2/2018	2018	NYY	HOU	9.00	5.	0	1.	10	0.00	0.36	0.56	0.68	3.40	0.77	92.00	340.07	1.08	1.13
Justin Verlander	8/3/2018	2018	HOU	LAD	7.20	4.	1	1.	14	0.00	0.23	0.47	0.65	23.70	0.60	81.00	333.13	1.01	1.31
Max Scherzer	9/8/2018	2018	WSN	CHC	9.00	9.	3	0.	11	0.00	0.17	0.42	0.68	12.60	0.63	76.00	332.55	1.01	0.97
Corey Kluber	6/19/2017	2017	CLE	BAL	9.00	3.	0	0.	11	0.00	0.32	0.43	0.52	-2.00	0.66	100.00	331.76	1.04	0.97
Max Fried	2020-07-30	2020	ATL	TBR	6.20	3.	1	1.	7.0	0.00	0.20	0.47	0.60	8.00	0.69	77.00	329.68	1.28	1.17

The 21 best starts per OLYAS

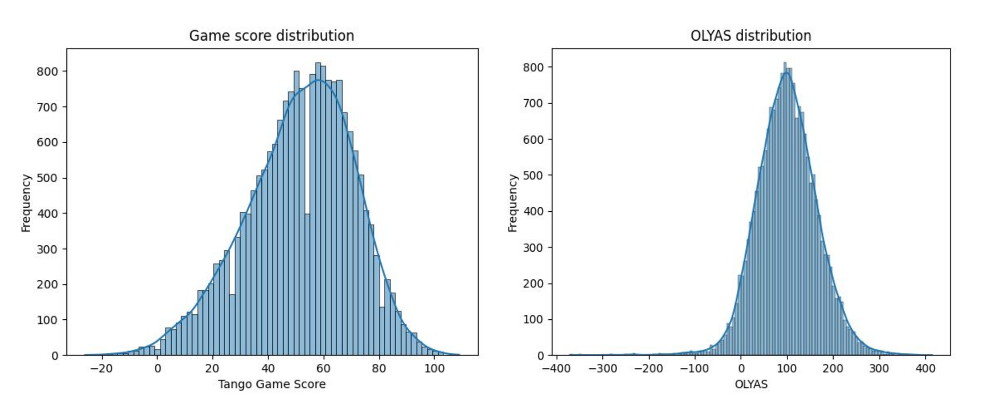
dollars_olyas_summary

player	date	year	team	opp	IP	H	E	BI	SO	Barrel%	HardHit%	Zone%	Z-Swing%	LA	Contact%	game_score	OLYAS	lgyr_multiplier	team_multiplier
Gio Gonzalez	6/25/2018	2018	WSN	TBR	1.00	3.	6	5.	2.0	0.50	1.00	0.27	0.59	18.90	0.68	6.00	-370.73	1.08	0.84
Danny Salazar	2017-09-05	2017	CLE	CHW	0.20	1.	4	2.	1.0	0.50	0.50	0.27	0.71	17.20	0.67	22.00	-365.11	1.04	0.85
Jose Berrios	5/16/2016	2016	MIN	DET	0.20	3.	7	4.	1.0	0.50	0.75	0.28	0.82	8.70	0.79	4.00	-348.40	1.01	1.08
Jordan Zimmerm	4/11/2018	2018	DET	CLE	0.20	0.	0	0.	0.0	0.50	1.00	0.14	1.00	13.10	1.00	44.00	-346.06	1.08	1.29
Lance McCullers	9/4/2020	2020	HOU	LAA	0.00	2.	3	3.	0.0	0.50	0.50	0.22	0.75	18.40	0.75	17.00	-304.68	1.28	1.10
James Paxton	7/12/2018	2018	SEA	LAA	0.20	3.	3	0.	1.0	0.50	1.00	0.41	0.86	17.80	0.89	18.00	-295.67	1.08	1.14
Jaime Garcia	9/26/2016	2016	STL	CIN	1.00	4.	2	0.	0.0	0.33	0.33	0.29	0.67	19.40	1.00	19.00	-282.57	0.95	0.93
Dan Straily	6/19/2018	2018	MIA	SFG	1.10	4.	4	2.	1.0	0.43	0.71	0.49	0.53	32.80	0.86	11.00	-282.12	1.01	0.89
Max Fried	2020-09-23	2020	ATL	MIA	1.00	3.	2	0.	0.0	0.33	0.67	0.41	0.89	12.60	1.00	23.00	-279.78	1.25	0.80

The 9 worst starts per OLYAS

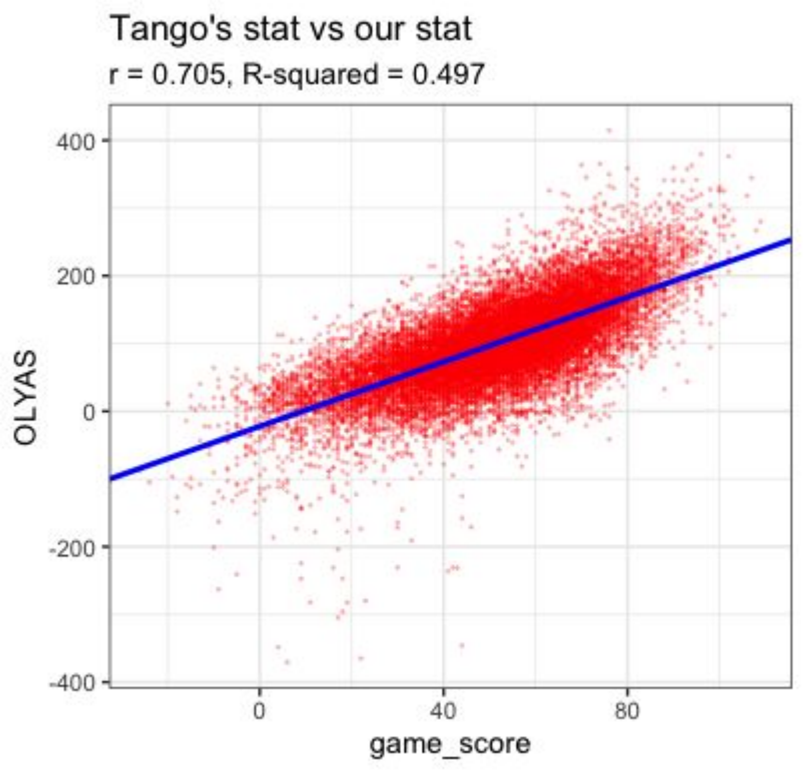
Analysis: How does OLYAS compare to Game Score?

Distribution:



	game_score	OLYAS
count	19383.000000	19383.000000
mean	51.504205	100.492176
std	19.339647	65.365361
min	-26.000000	-370.732762
25%	39.000000	58.290149
50%	53.000000	99.542082
75%	65.000000	141.364560
max	109.000000	414.778222

Bivariate comparison:



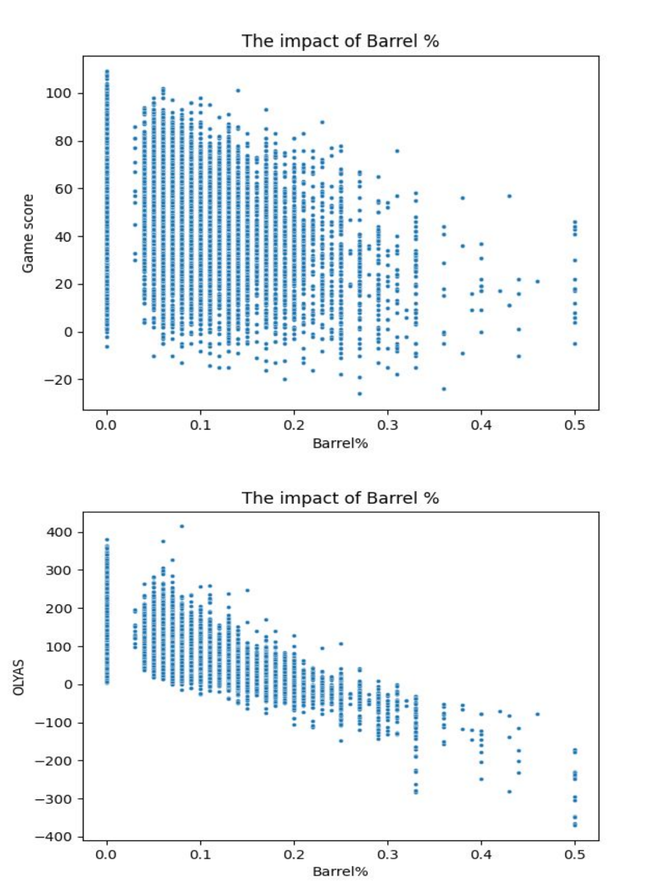
Analysis: How does OLYAS compare to Game Score?

Correlations between different statistics and game_score and OLYAS

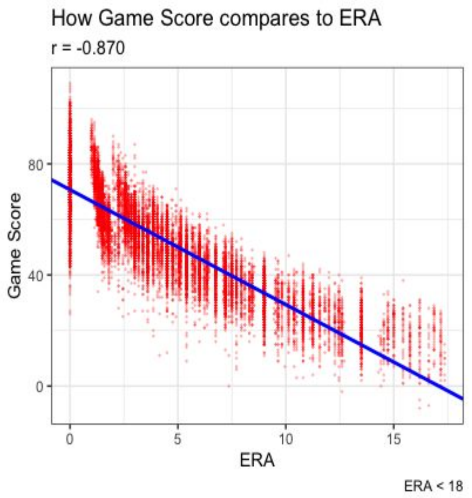
	game_score	OLYAS
game_score	1.000000	0.705162
OLYAS	0.705162	1.000000
Barrel%	-0.435303	-0.728235
HardHit%	-0.332944	-0.414438
LA	-0.064354	-0.144536
Contact%	-0.280368	-0.379337
Zone%	0.065536	0.124818
Z-Swing%	-0.085596	-0.081725

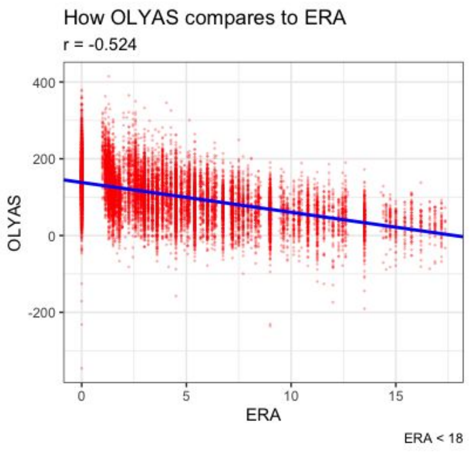
OLYAS has a stronger relationship with each of these key performance indicators (KPI) than Game Score does.

One KPI of particular importance: Barrel%



Analysis: How does OLYAS compare with other common stats?

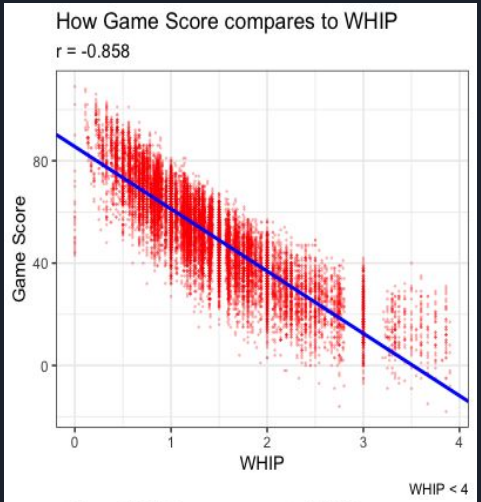


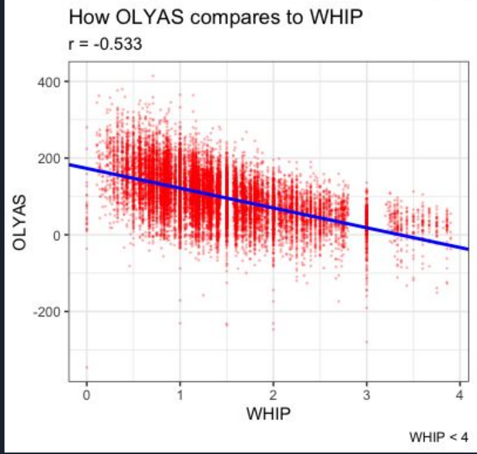


OLYAS has a decent correlation with commonly used stats like ERA and WHIP.

However, the correlation between those stats OLYAS is lower than for Game Score

This trend was expected, as Game Score is constructed with outcomes, while OLYAS evaluates the pitcher's Isolated performance.

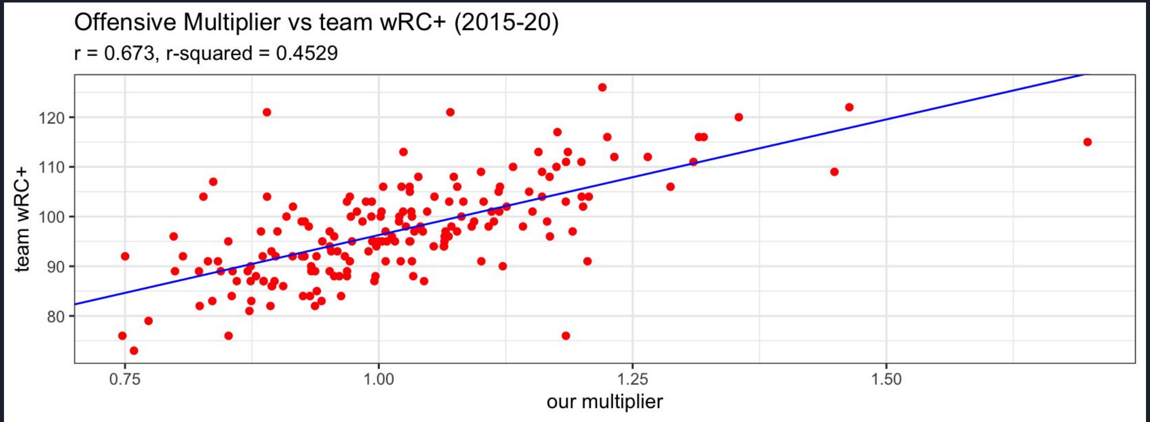




Analysis: The Opponent Adjustment Multiplier

The opponent adjustment proved to have a major impact on OLYAS.

Since it is so impactful, it is good to know that our multipliers correlate well with established metrics like team wRC+.



Analysis: Revisiting Opening Case Example

Start A: JA Happ (TOR), 5/16/2018 vs NYM
Start B: Yu Darvish (TEX), 7/26/2017 vs MIA
Happ OLYAS: 32.6
Darvish OLYAS: 112.35
Happ Z_Discrepancy: -2.93
Darvish Z_Discrepancy: 2.638 (8th largest since 2015)

Key Statistics: 23% vs 0% Barrel%, 61% vs 68% F-Strike%, 0.952 vs 1.003 Opponent Multiplier

Start A	Start B
7 IP	3.2 IP
2 H	9 H
0 ER	10 ER
0 BB	2 BB
10 Ks	5 Ks
Game Score: 88	Game Score: 4

Start A	Start B
23% Barrel%	0% Barrel%
39% HardHit%	33% HardHit%
73% Contact%	71% Contact%
46% Zone%	49% Zone%
61% F-Strike%	68% F-Strike%
16% SwStr%	16% SwStr%
0.952 Opponent Multiplier	1.003 Opponent Multiplier

Analysis: Revisiting the Lucas Giolito Case Example

Start A

- **9 IP**
- **3 H Allowed**
- **0 ER**
- **0 BB**
- **12 K**

Start B

- **9 IP**
- **0 H Allowed (NO HITTER!!)**
- **0 ER**
- **1 BB**
- **13 K**

Which was better?

Analysis: Giolito Case Example

- Start A: Against 2019 Twins
    - 116 wRC+ (2nd in MLB)
    - 120.4 Off (3rd in MLB)
    - 307 HR (1st in MLB)
    - 1.32 Multiplier
    - 101-61 record
- Start B: Against 2020 Pirates
    - 73 wRC+ (Last in MLB)
    - -76.8 Off (Last in MLB)
    - 59 HR (28th in MLB)
    - 0.759 Multiplier
    - On pace for 51-111 record

Start A OLYAS: 376.41
Start B OLYAS: 344.49

Using OLYAS as a long-term metric

Top 10 seasons in average OLYAS since 2015

1. Corey Kluber (CLE), 2017 (185.21)
2. German Marquez (COL), 2020 (182.03)
3. Kyle Hendricks (CHC), 2020 (181.35)
4. Jacob deGrom (NYM), 2018 (177.75)
5. Yu Darvish (CHC), 2020 (176.77)
6. Zack Wheeler (PHI), 2020 (176.71)
7. Clayton Kershaw (LAD), 2016 (174.95)
8. Gerrit Cole (HOU), 2019 (173.9)
9. Shane Bieber (CLE), 2020 (173.55)
10. Clayton Kershaw (LAD), 2015 (173.1)

German Marquez

1. Perhaps the most surprising pitcher on that list is Colorado Rockies pitcher German Marquez.
2. While Marquez is commonly recognized as a valuable pitcher, basic box score statistics do not do him justice, considering his unconventional situation.
3. Marquez pitches half of his starts in the hitter friendly Coors Field.
    a. Coors Field contains the highest park factor by a substantial margin.
4. Moreover, he pitched in a division with two of the league’s most potent offenses.
    a. His average opponent possessed a 1.084 multiplier (1.00 is average).

Conclusions and Takeaways

1. Game score is an imperfect metric due to its inability to control for confounding factors.
2. Our version of Game Score, OLYAS, does a better job of isolating the pitcher's performance in a game through the use of expected statistics, quality of contact measures, pitch data, and adjustments for league, year, and quality of opponent, rather than just outcomes.
3. OLYAS lines up fairly well with lots of KPIs but is different enough to reveal new information that previously wouldn't have been unearthed (i.e German Marquez).

Applications

How can fans, reporters, and other evaluators use OLYAS? :

1. Assess how well a pitcher performed on a given day.
2. Rolling average OLYAS can be another indicator of performance just as ERA, WHIP, FIP, xERA, etc.

Pitchers, coaches, player developers:

1. Maintain a rolling OLYAS database.
2. Create tables like the ones in the Happ vs Darvish case example to find which component a player is excelling or not doing well in.
3. Manipulate OLYAS formula to determine how a pitcher can get better.

Potential Next Steps

1. Introduce a Game Score/OLYAS model for hitters
2. Build a Shiny App with instantly-updated OLYAS
3. Create a Pitcher Elo rank based on OLYAS
4. Implement more refined Statcast metrics such as xwOBA

Thank You!!!

Any questions?

About the presenters:
Aidan Resnick (18), Maxwell Resnick (18), Jake Federman (17), and Eshan Mehere (16) are students at Horace Mann School in Bronx, NY. All four students attended Wharton Moneyball Academy, hold leadership positions in the Horace Mann Sports Analytics Club, and compete on varsity sports teams. Aidan, Maxwell, Jake, and Eshan hope to one day work in the sports analytics industry, specifically within baseball.