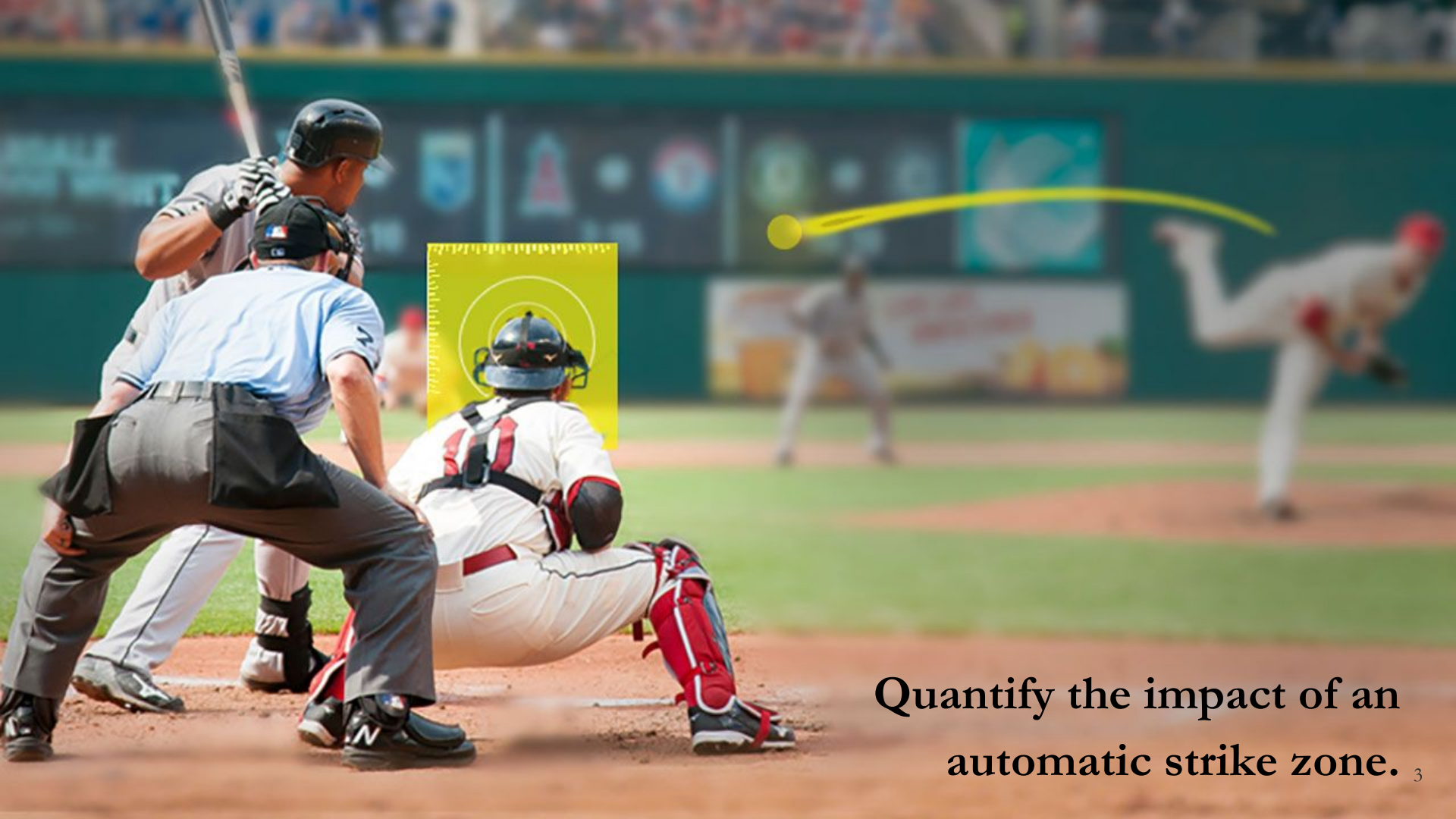
Final Presentation

MLB: Automatic Strike Zone

Devon Goetz, Zack Kopstein, Simran Pabla, Claire Wichman

May 12, 2020



Quantify the impact of an automatic strike zone.

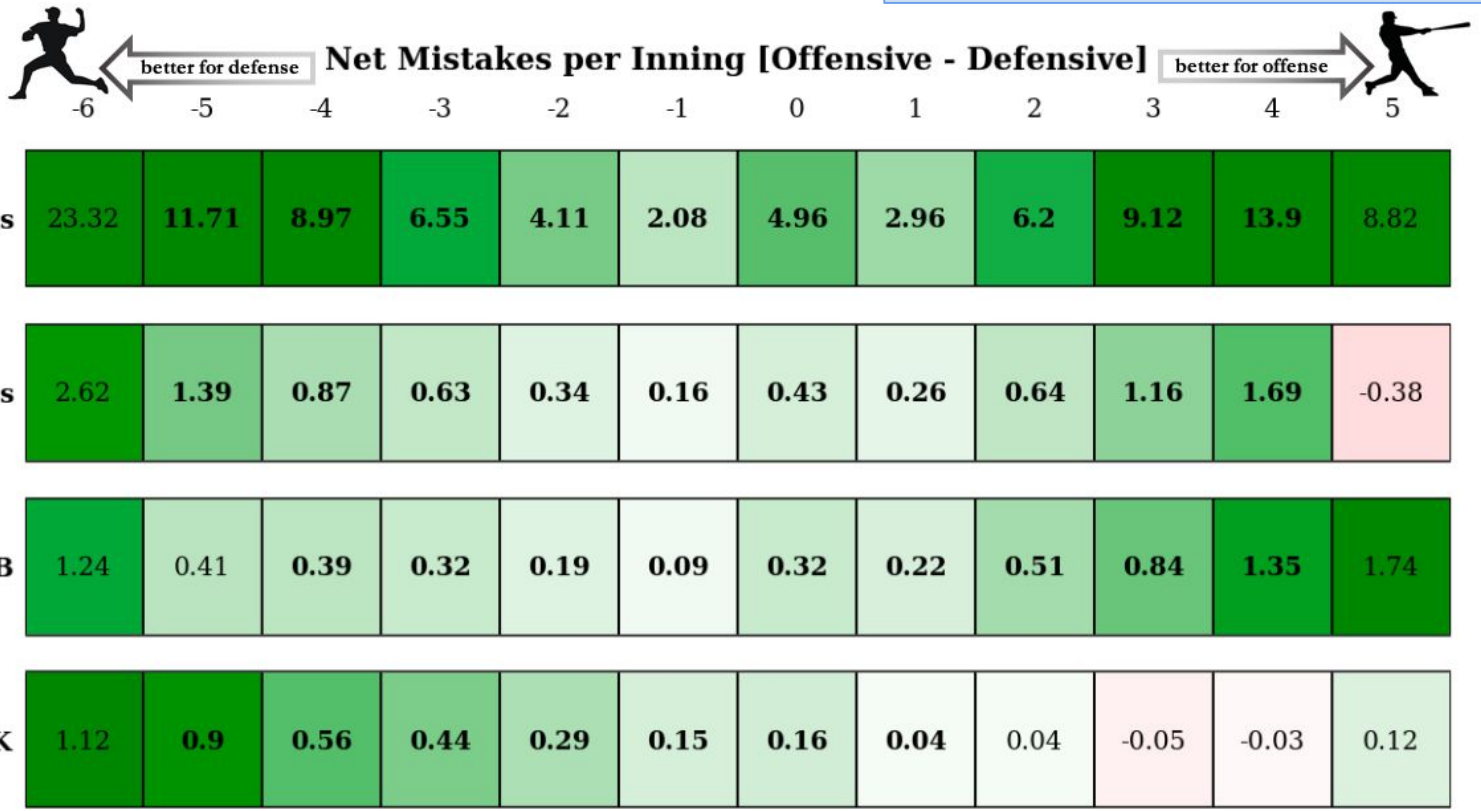
Approach and Methodology

- Classify mistakes
- Compare existing innings
- Create predictive algorithm
- Analyze differences

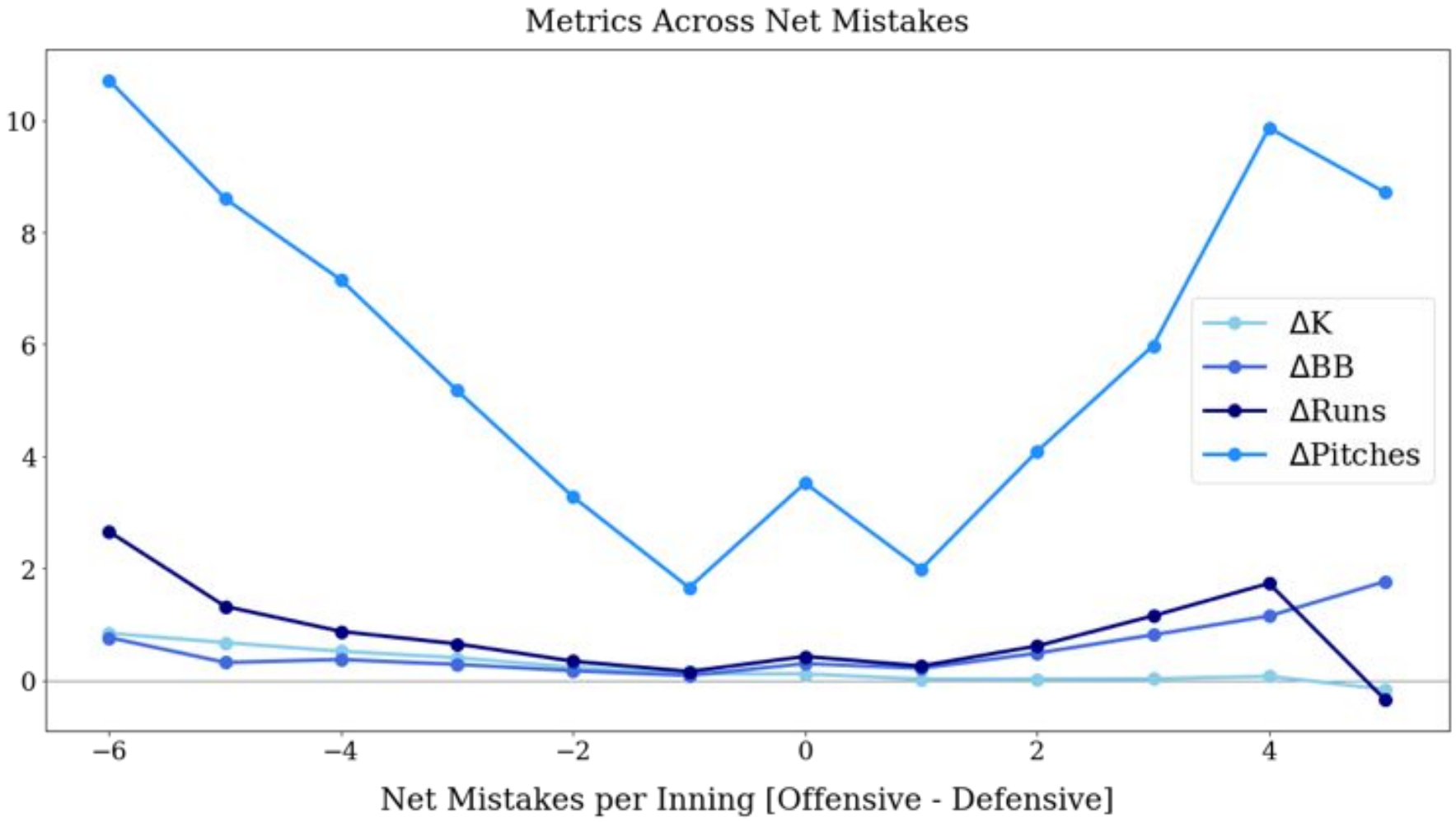
Approach and Methodology

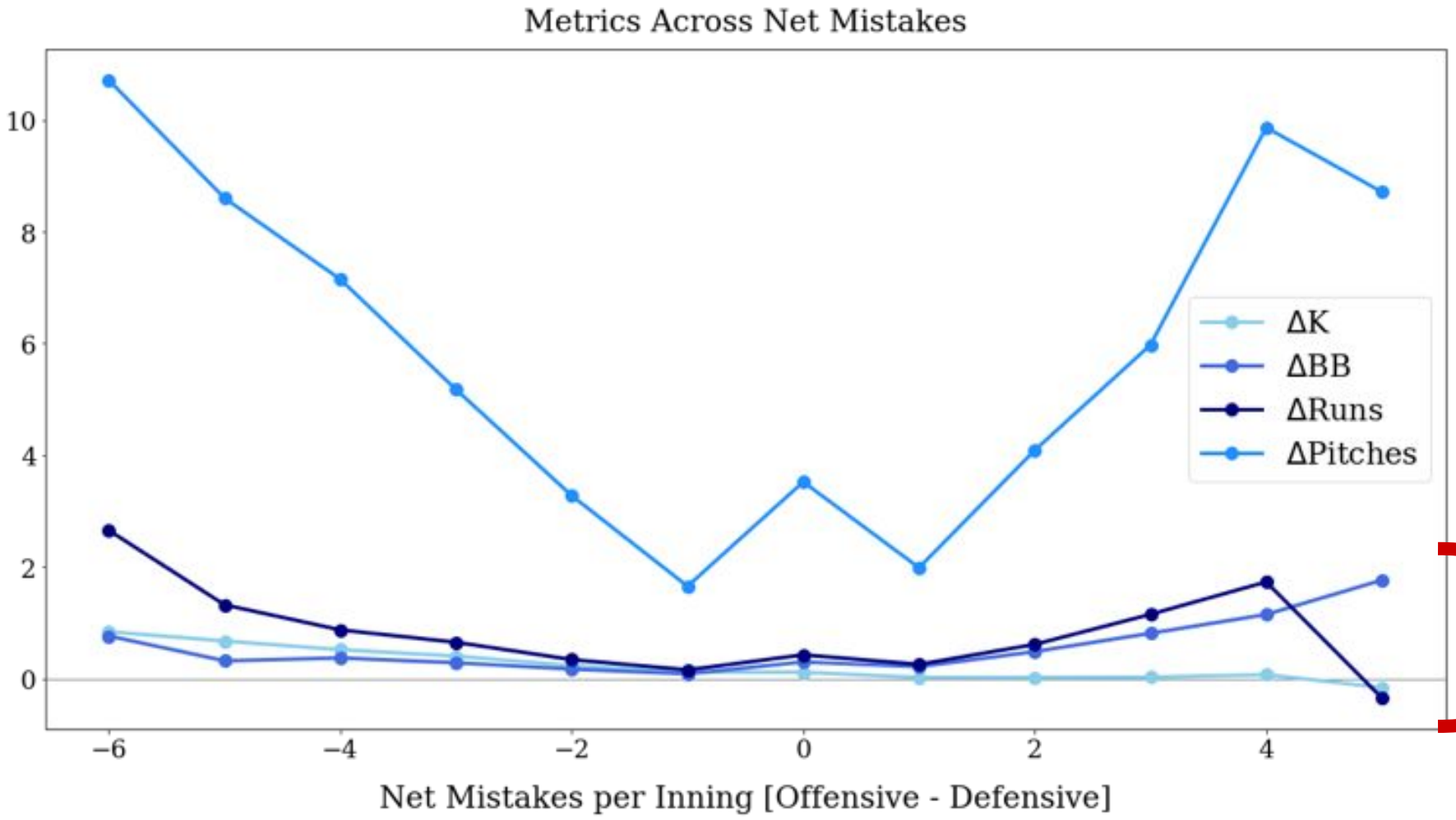
- **Classify mistakes**
- **Compare existing innings**
- Create predictive algorithm
- Analyze differences

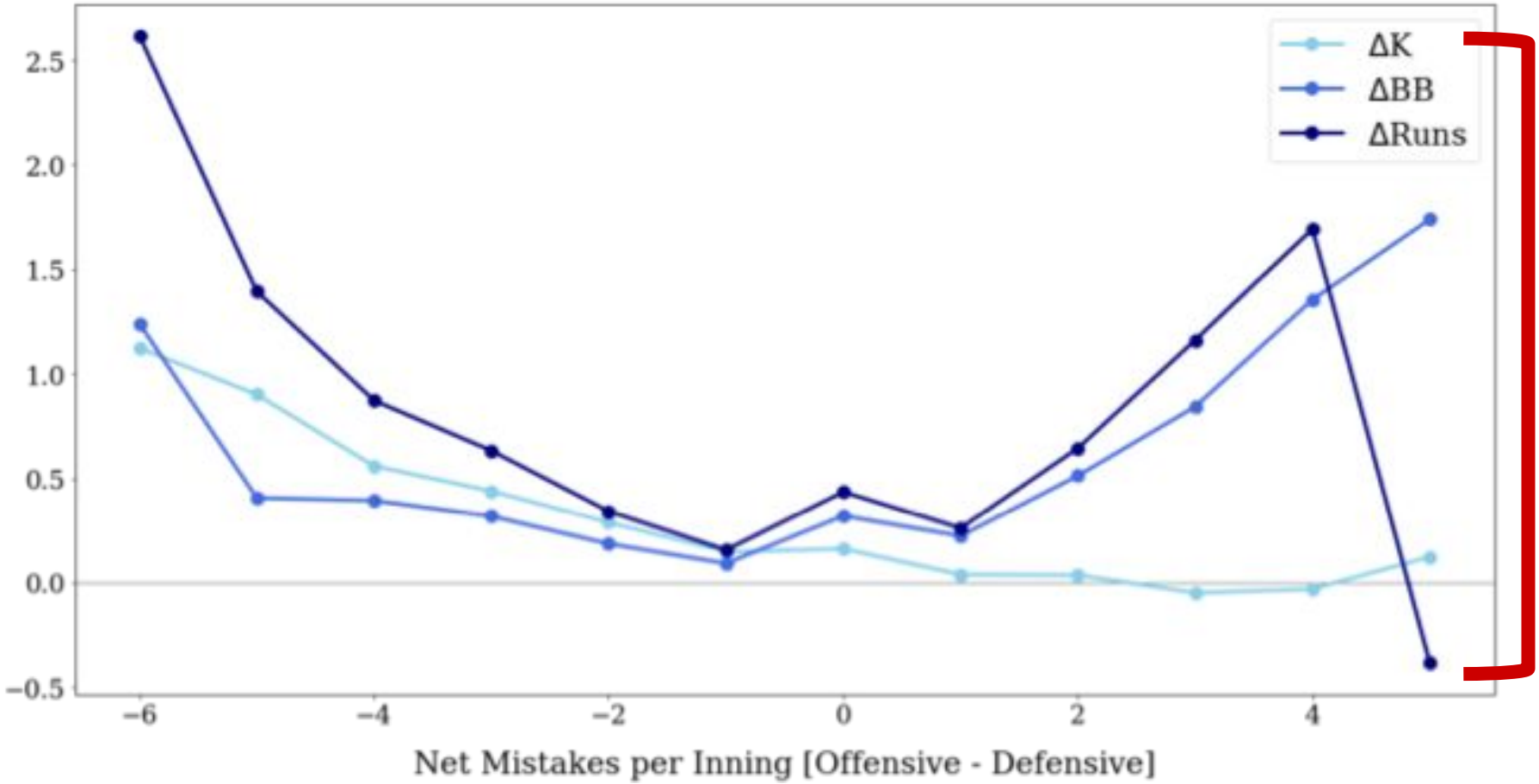
pitches are driving our metrics



Note: Δ values calculated as [X Mistake Inning Average - 0 Mistake Inning Average]
Non-bolded statistics have a p-value of greater than 0.05, and thus have been deemed statistically insignificant.

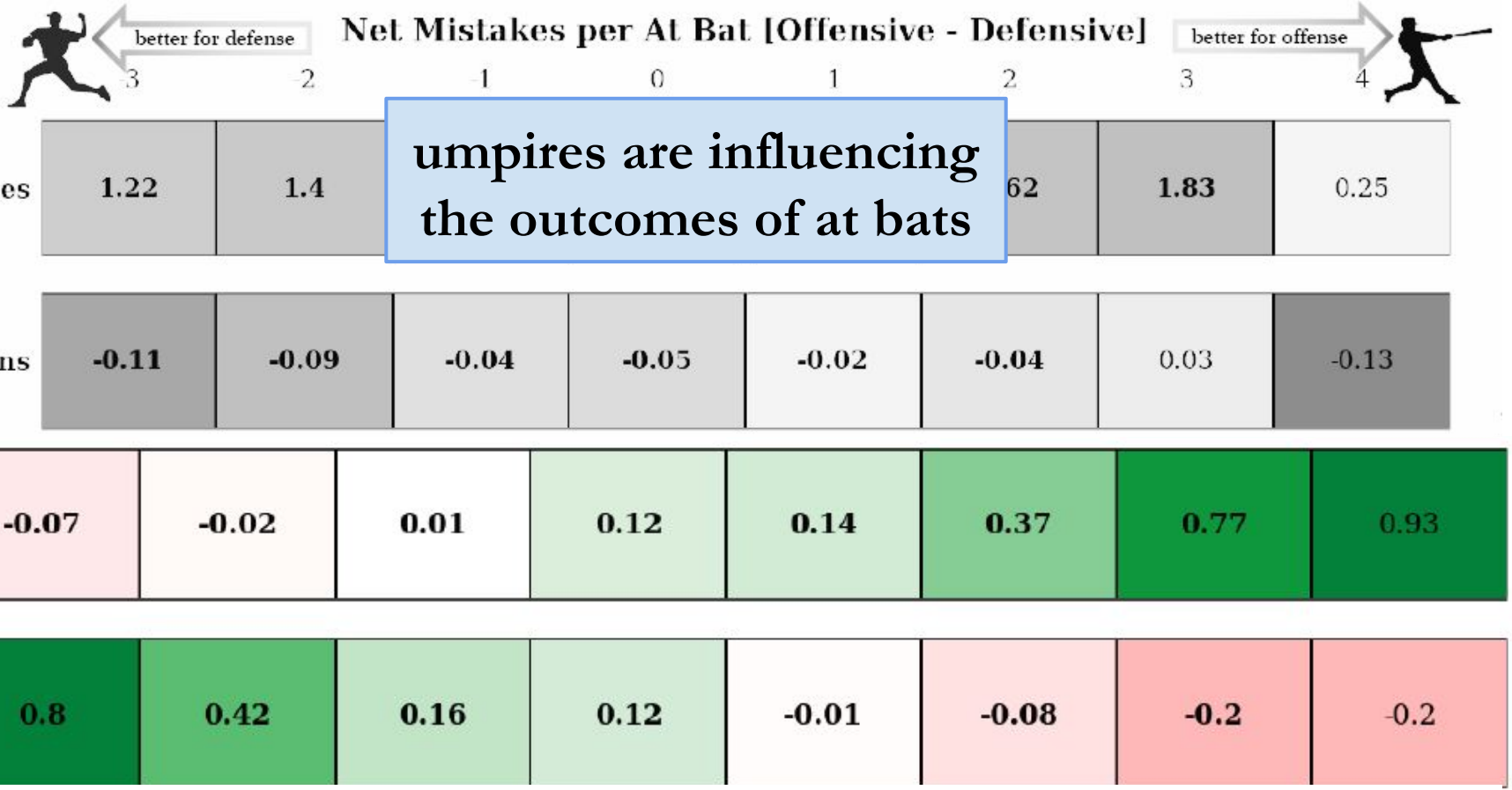

Metrics Across Net Mistakes
10
8
6
4
2
0
−6
−4
−2
0
2
4
ΔK
ΔBB
ΔRuns
ΔPitches
Net Mistakes per Inning [Offensive - Defensive]


Metrics Across Net Mistakes
10
8
6
4
2
0
−6
−4
−2
0
2
4
ΔK
ΔBB
ΔRuns
ΔPitches
Net Mistakes per Inning [Offensive - Defensive]


ΔK
ΔBB
ΔRuns
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−6
−4
−2
0
2
4
Net Mistakes per Inning [Offensive - Defensive]

	-3	-2	-1	0	1	2	3
ΔPitches	**1.22**	**1.4**	**0.91**	**1.75**	**1.14**	**1.62**	**1.83**
ΔRuns	**-0.11**	**-0.09**	**-0.04**	**-0.05**	**-0.02**	**-0.04**	0.03
ΔBB	**-0.07**	**-0.02**	**0.01**	**0.12**	**0.14**	**0.37**	**0.77**
ΔK	**0.8**	**0.42**	**0.16**	**0.12**	**-0.01**	**-0.08**	**-0.2**

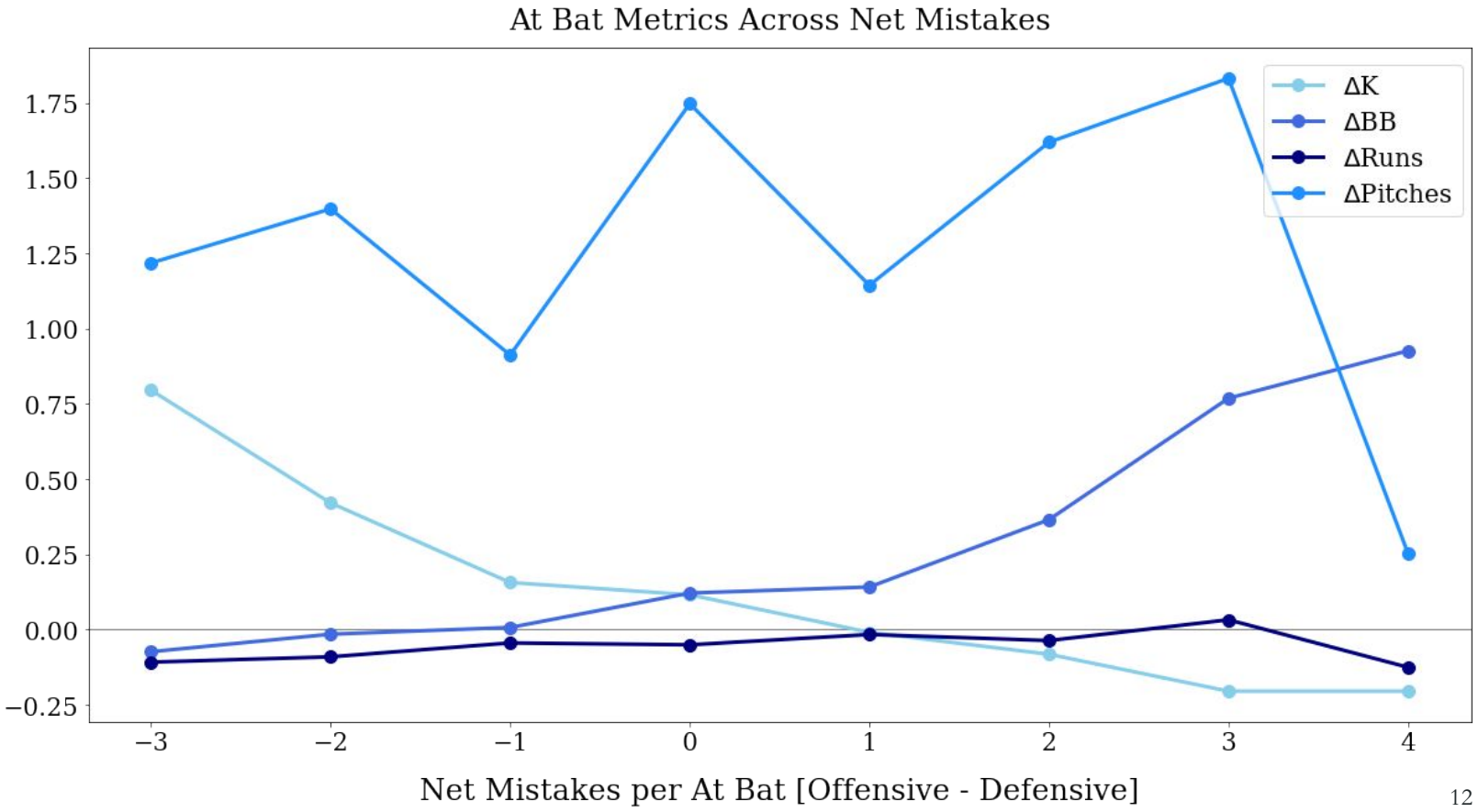
Note: Δ values calculated as [X Mistake Inning Average - 0 Mistake Inning Average]
Non-bolded statistics have a p-value of greater than 0.05, and thus have been deemed statistically insignificant.

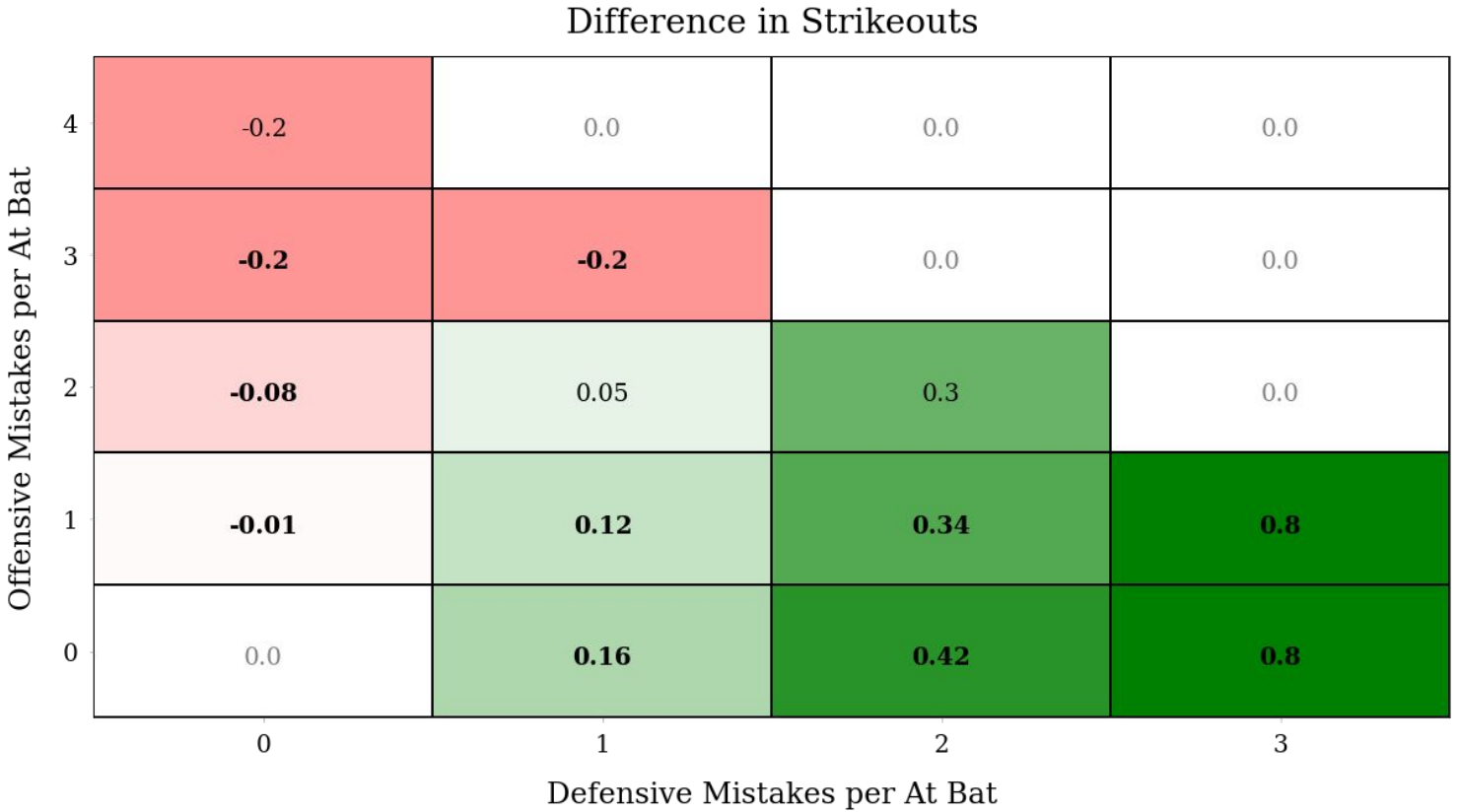


Note: Δ values calculated as [X Mistake Inning Average - 0 Mistake Inning Average]

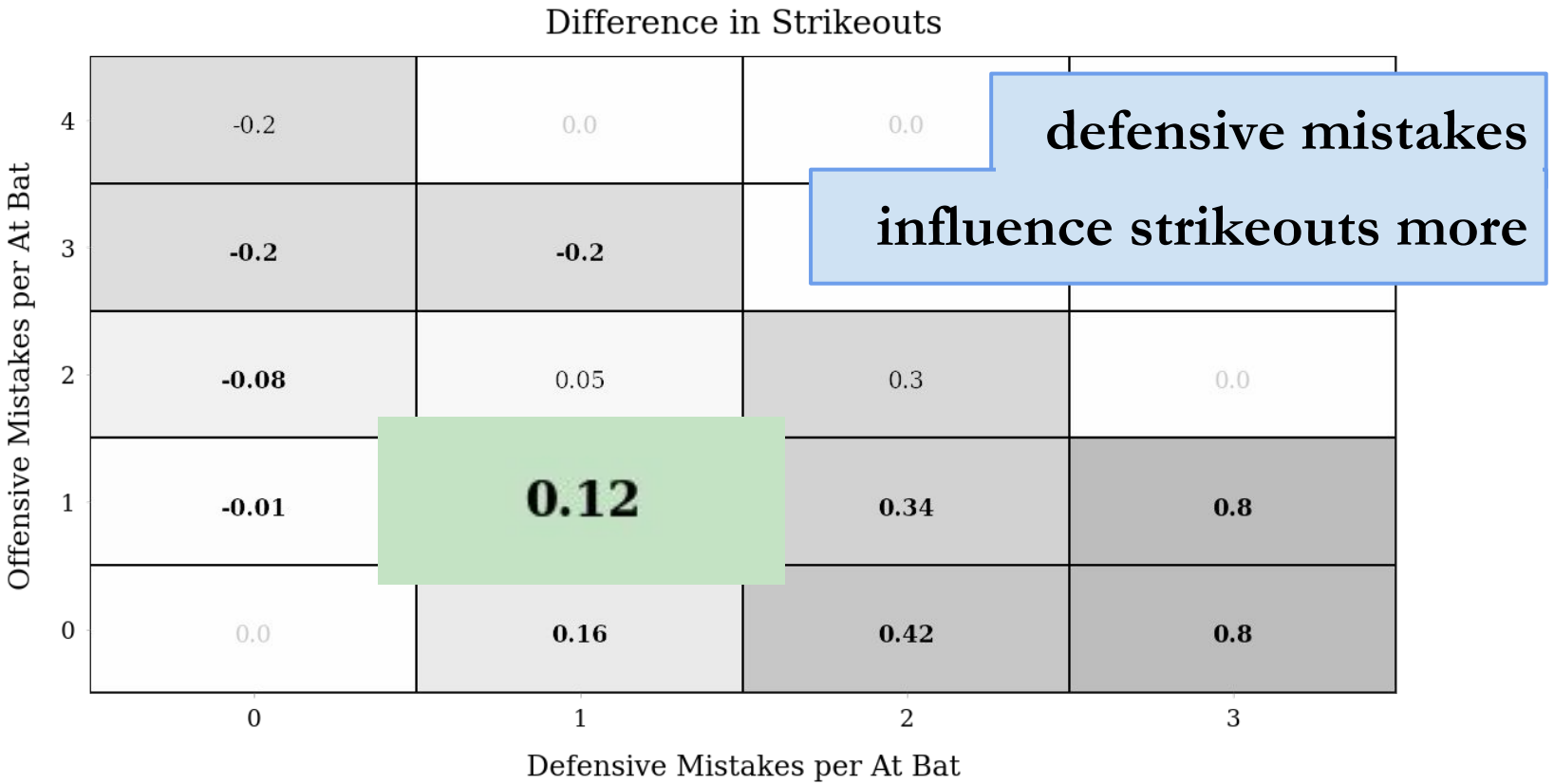
Non-bolded statistics have a p-value of greater than 0.05, and thus have been deemed statistically insignificant.

favorable mistakes have a greater impact

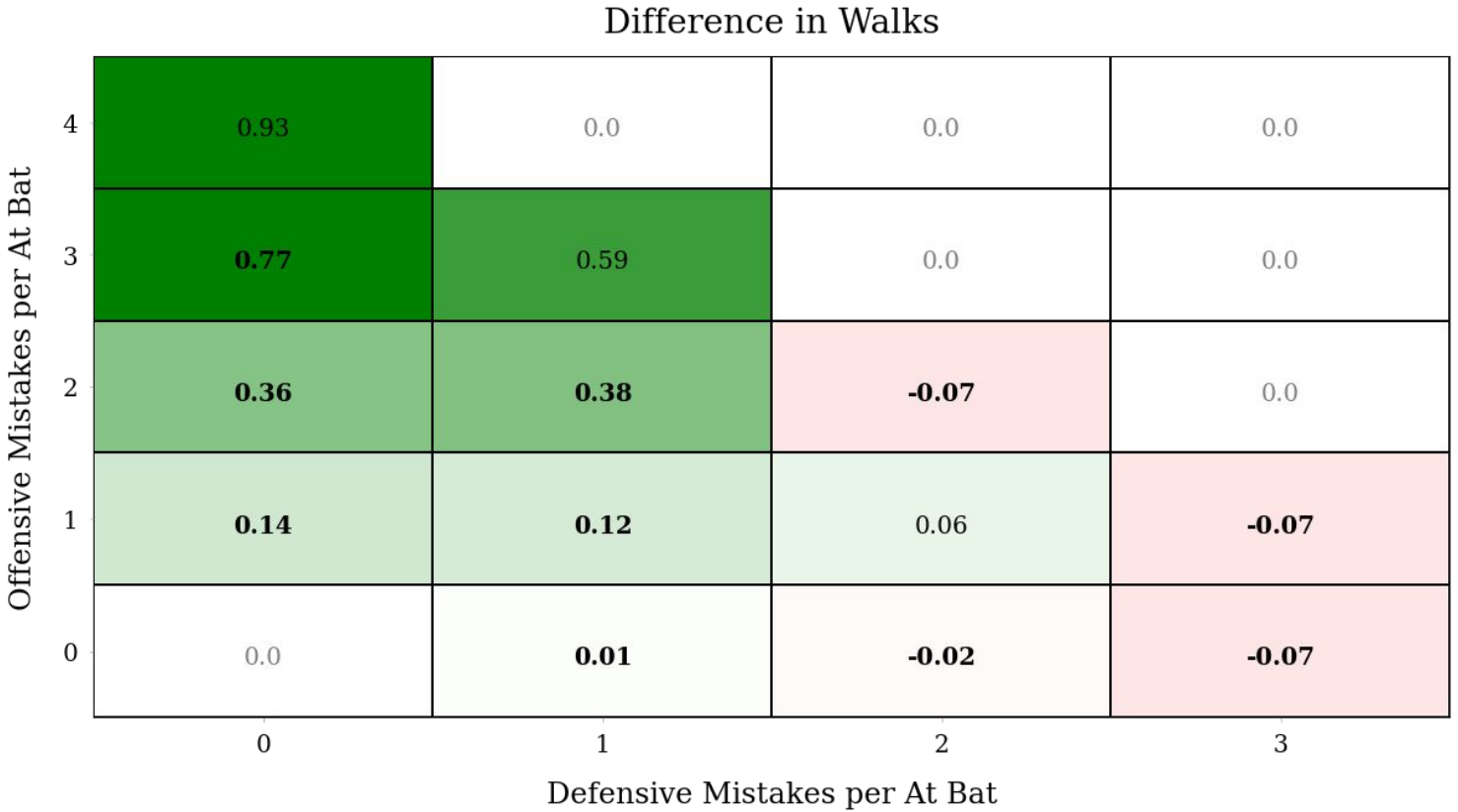




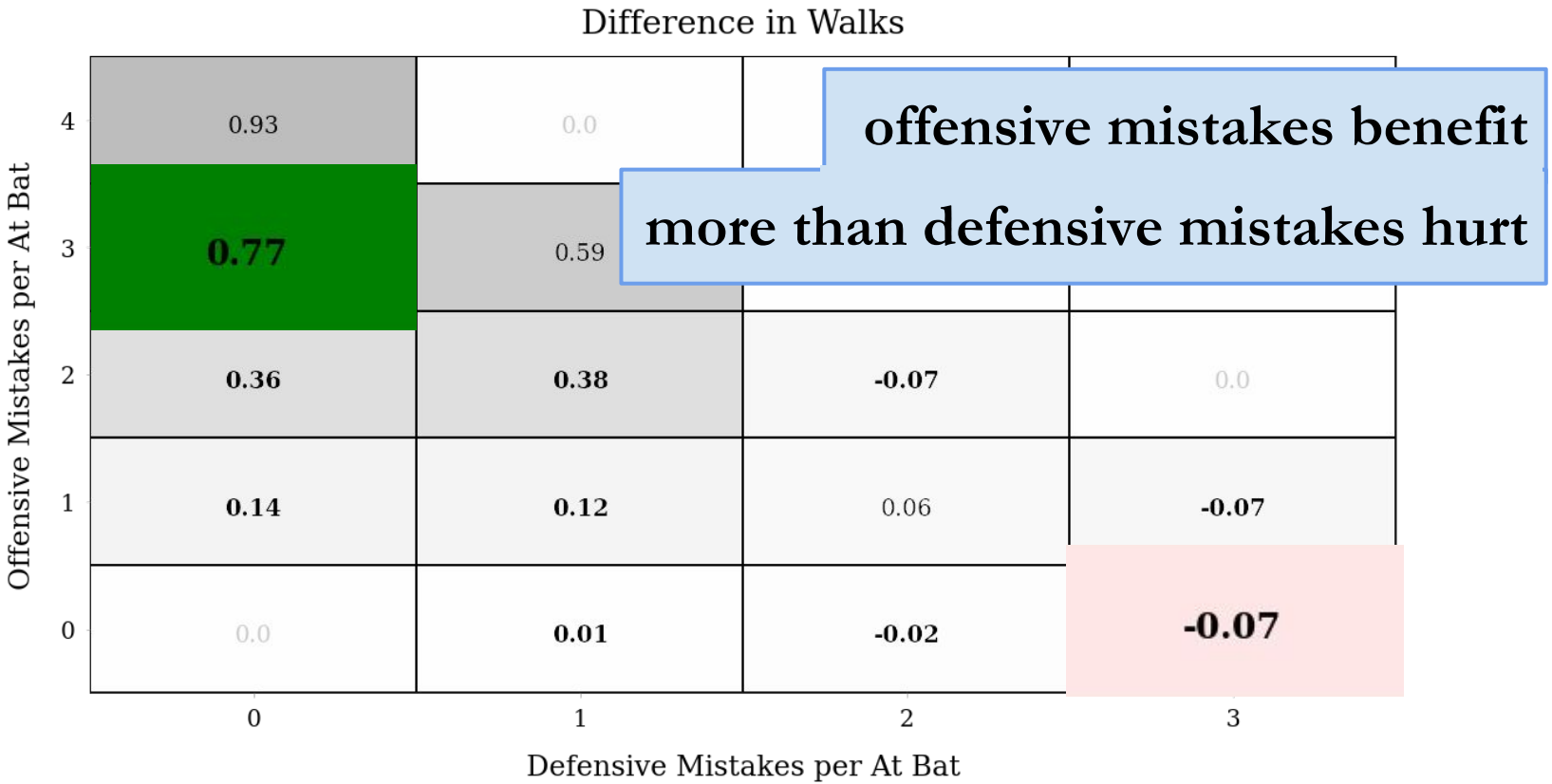
Note: Δ values calculated as [X Mistake At Bat Average - 0 Mistake At Bat Average]
Non-bolded statistics have a p-value of greater than 0.05, and thus have been deemed statistically insignificant.



Note: Δ values calculated as [X Mistake At Bat Average - 0 Mistake At Bat Average]
Non-bolded statistics have a p-value of greater than 0.05, and thus have been deemed statistically insignificant.



Note: Δ values calculated as [X Mistake At Bat Average - 0 Mistake At Bat Average]
Non-bolded statistics have a p-value of greater than 0.05, and thus have been deemed statistically insignificant.



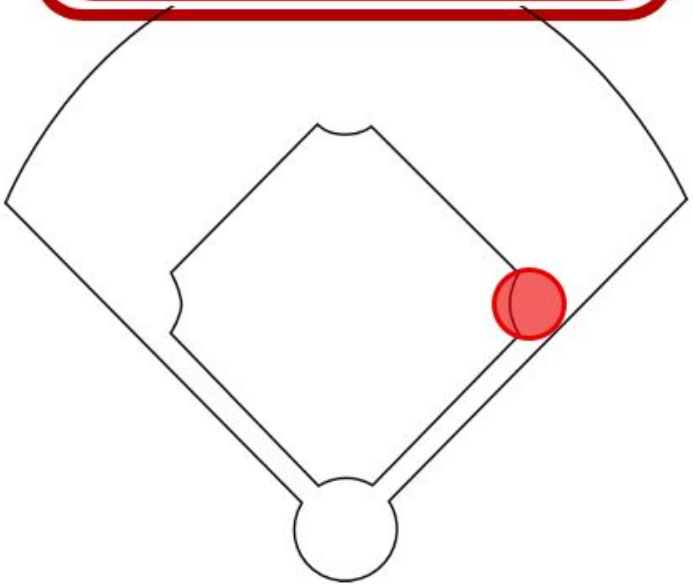
Note: Δ values calculated as [X Mistake At Bat Average - 0 Mistake At Bat Average]
Non-bolded statistics have a p-value of greater than 0.05, and thus have been deemed statistically insignificant.

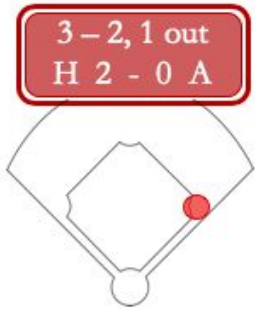
Approach and Methodology

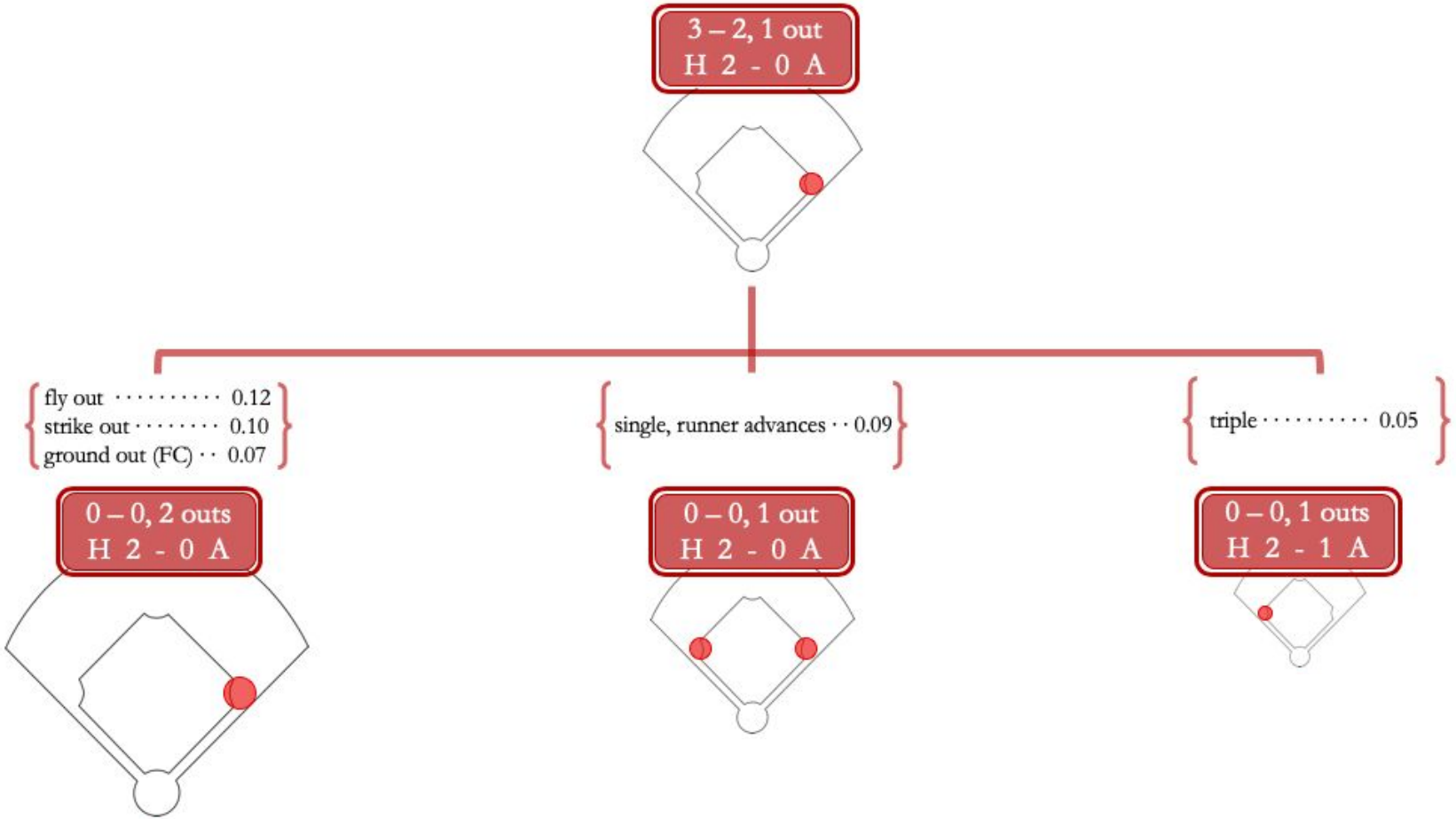
- Classify mistakes
- Compare existing innings
- **Create predictive algorithm**
- **Analyze differences**

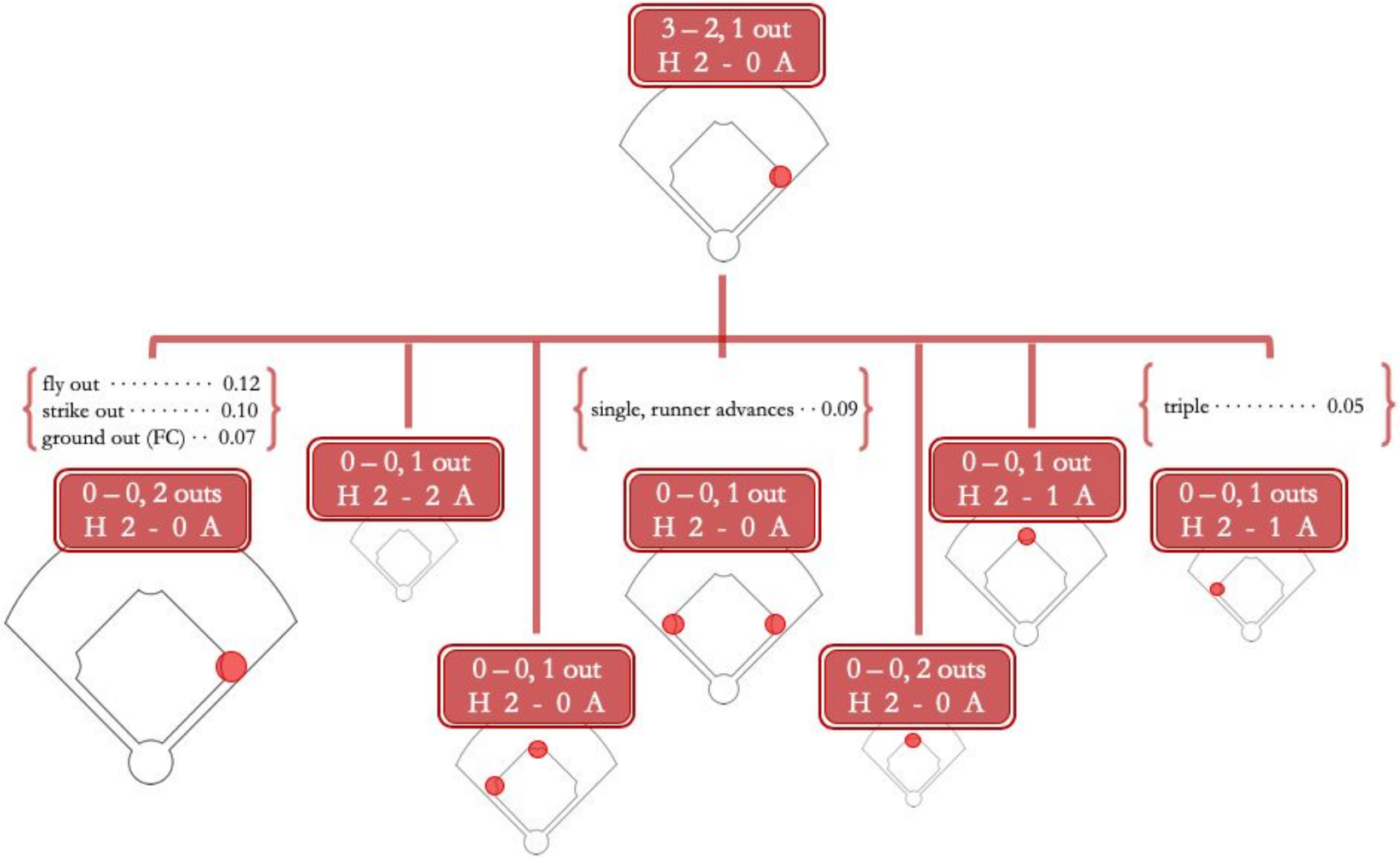
3 – 2, 1 out

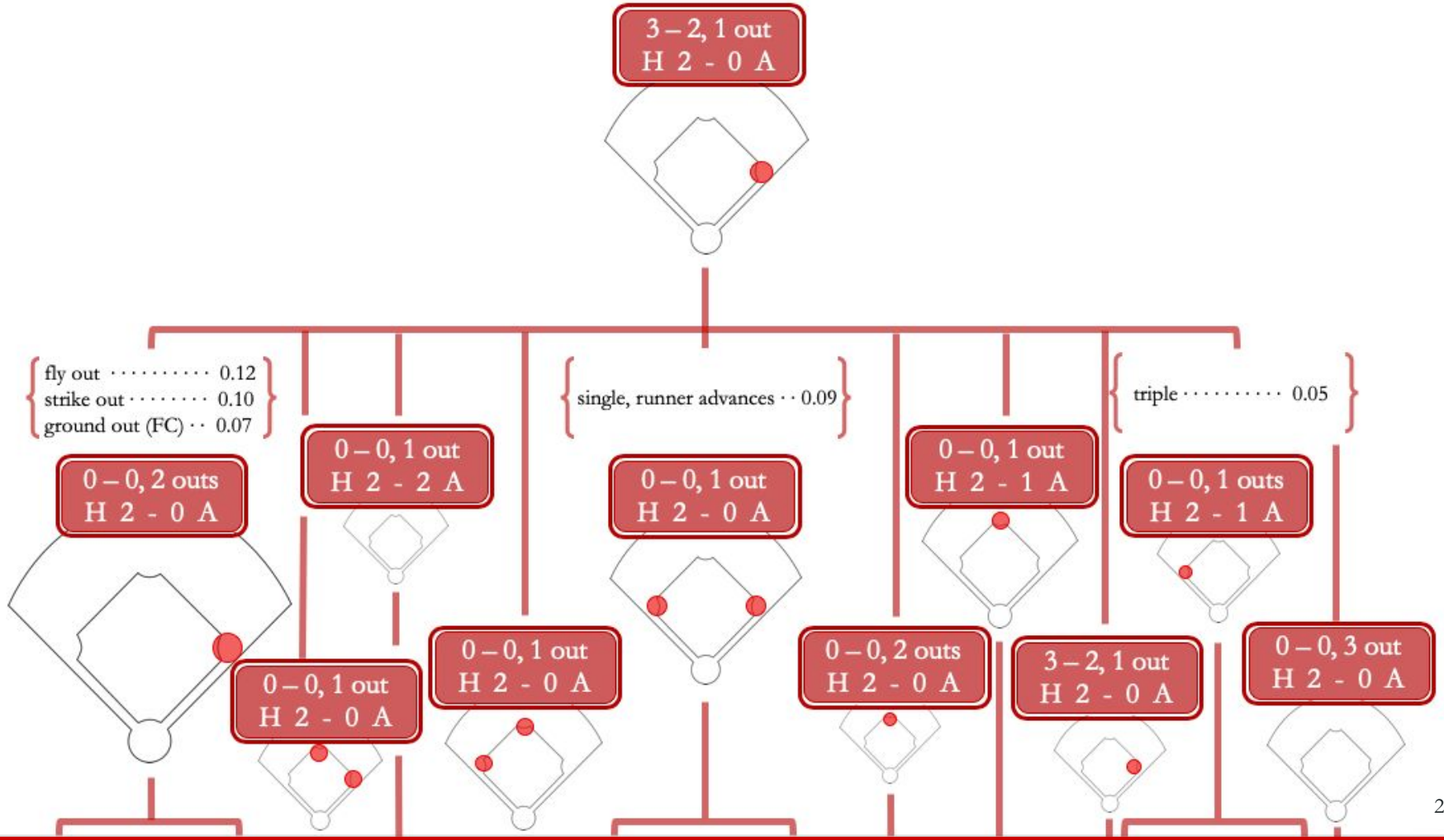
H 2 - 0 A

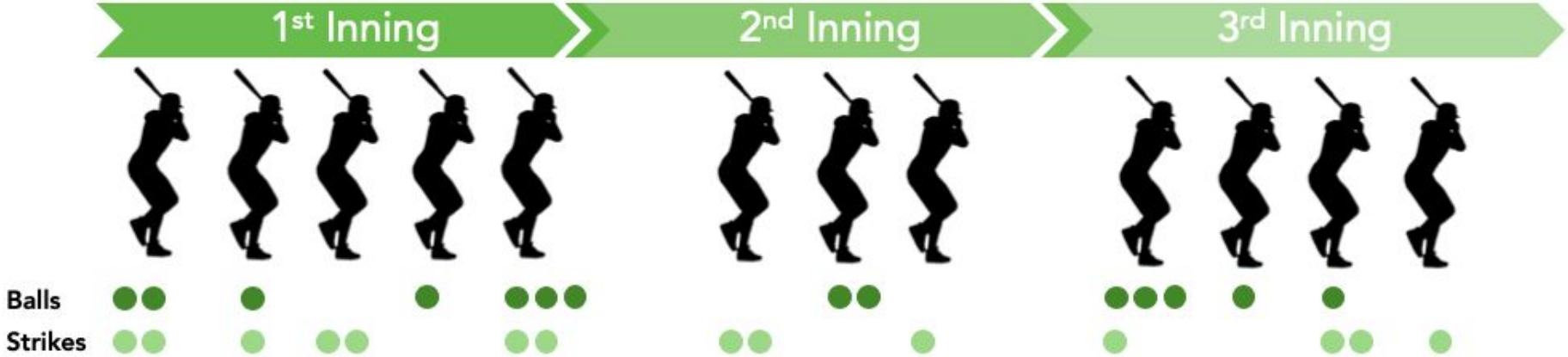


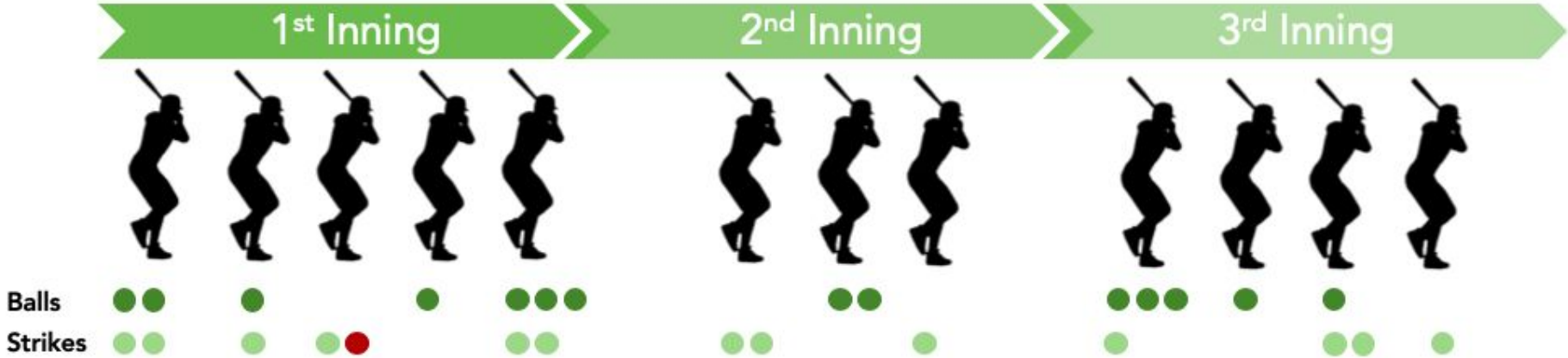

3 – 2, 1 out
H 2 - 0 A

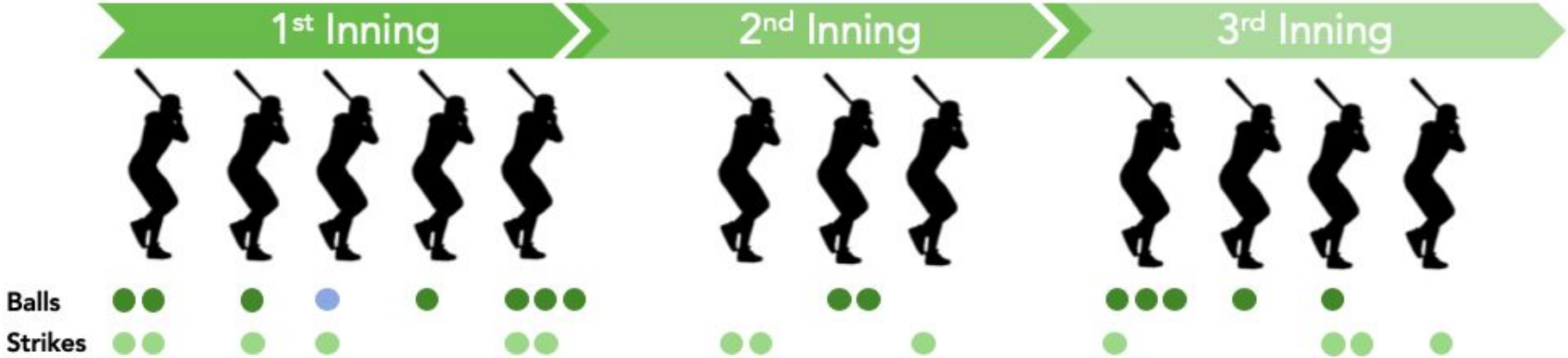

3 – 2, 1 out
H 2 - 0 A
fly out ····· 0.12
strike out ····· 0.10
ground out (FC) ·· 0.07
single, runner advances ·· 0.09
triple ····· 0.05
0 – 0, 2 outs
H 2 - 0 A
0 – 0, 1 out
H 2 - 0 A
0 – 0, 1 outs
H 2 - 1 A


3 – 2, 1 out
H 2 - 0 A
fly out · · · · · · · · · · 0.12
strike out · · · · · · · · 0.10
ground out (FC) · · 0.07
0 – 0, 2 outs
H 2 - 0 A
0 – 0, 1 out
H 2 - 2 A
0 – 0, 1 out
H 2 - 0 A
single, runner advances · · 0.09
0 – 0, 1 out
H 2 - 0 A
0 – 0, 2 outs
H 2 - 0 A
0 – 0, 1 out
H 2 - 1 A
triple · · · · · · · · · · 0.05
0 – 0, 1 outs
H 2 - 1 A


3 – 2, 1 out
H 2 - 0 A
fly out ····· 0.12
strike out ····· 0.10
ground out (FC) ·· 0.07
0 – 0, 2 outs
H 2 - 0 A
0 – 0, 1 out
H 2 - 2 A
0 – 0, 1 out
H 2 - 0 A
0 – 0, 1 out
H 2 - 0 A
single, runner advances ·· 0.09
0 – 0, 1 out
H 2 - 0 A
0 – 0, 1 out
H 2 - 1 A
0 – 0, 2 outs
H 2 - 0 A
3 – 2, 1 out
H 2 - 0 A
triple ····· 0.05
0 – 0, 1 outs
H 2 - 1 A
0 – 0, 3 out
H 2 - 0 A


1st Inning
2nd Inning
3rd Inning
Balls
Strikes

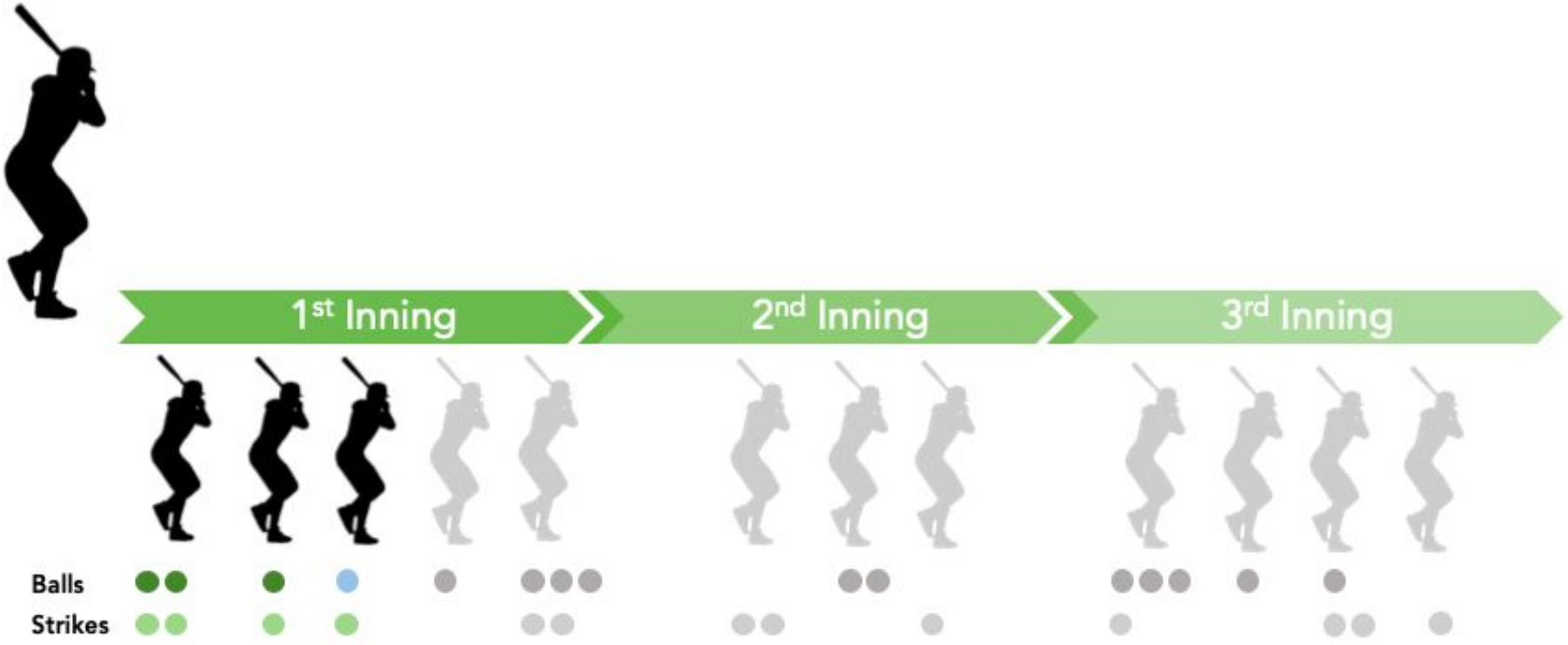

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

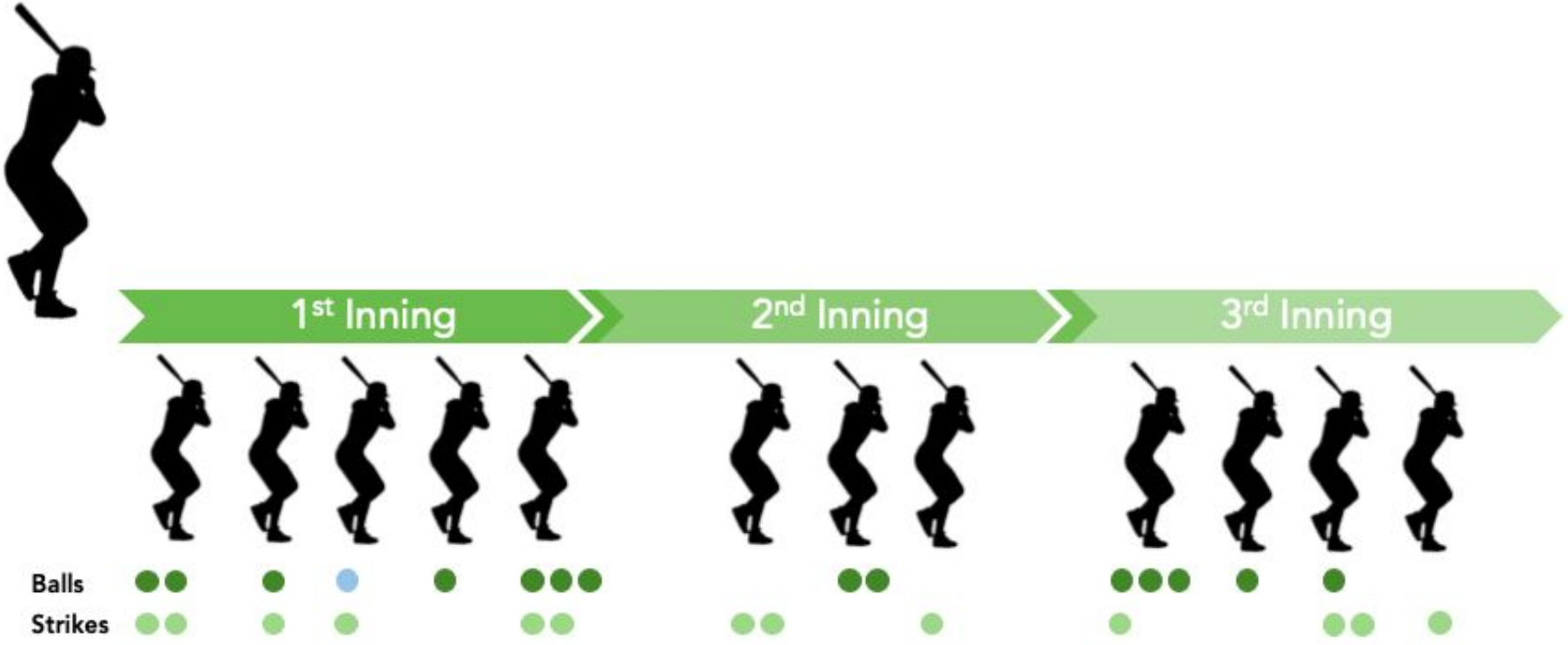

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

COLORADO
27

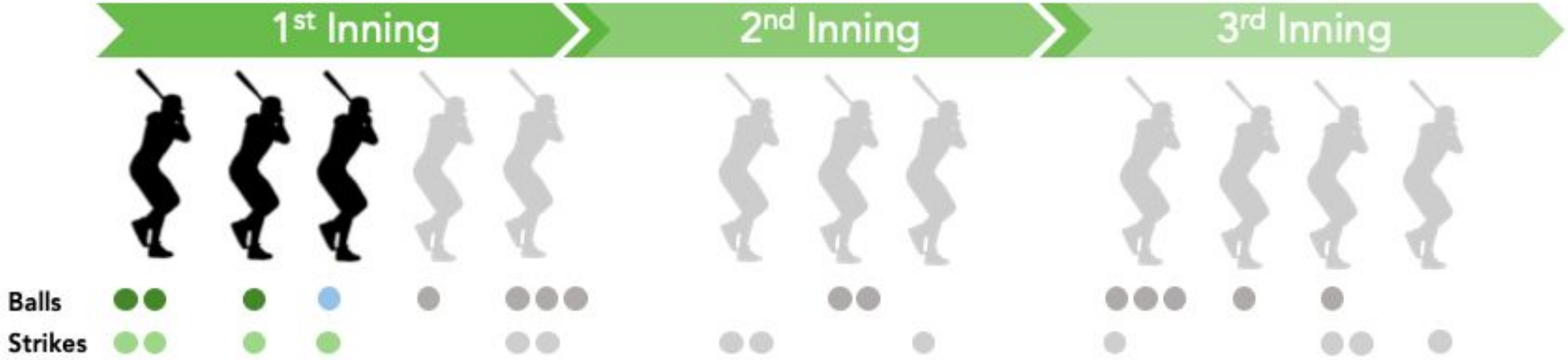
COLORADO
27

SOX

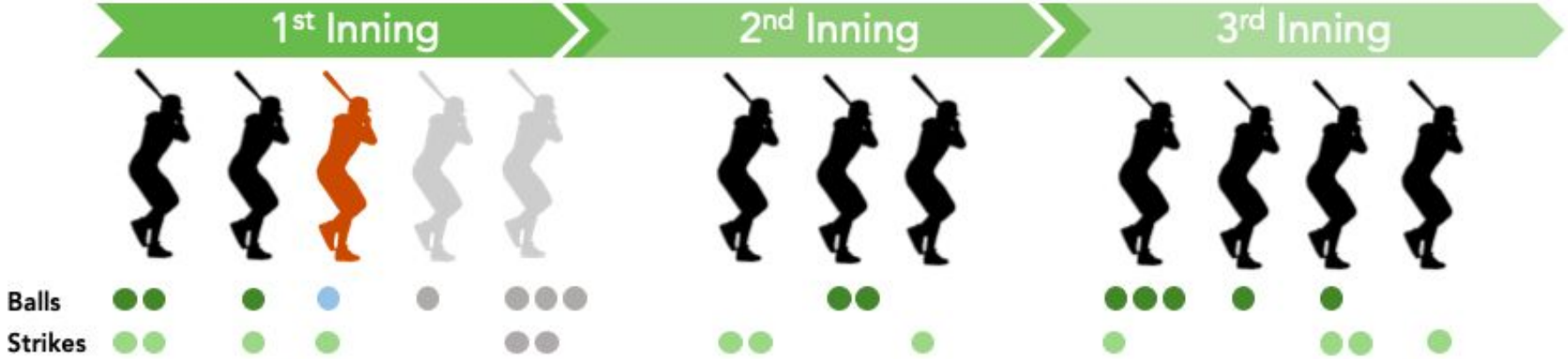

1st Inning
2nd Inning
3rd Inning
Balls
Strikes


1st Inning
2nd Inning
3rd Inning
Balls
Strikes

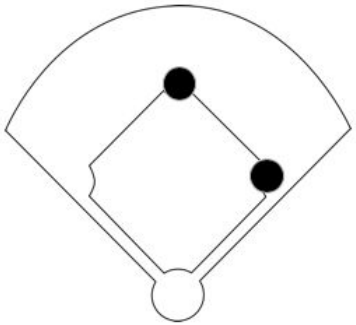
COLORADO
27

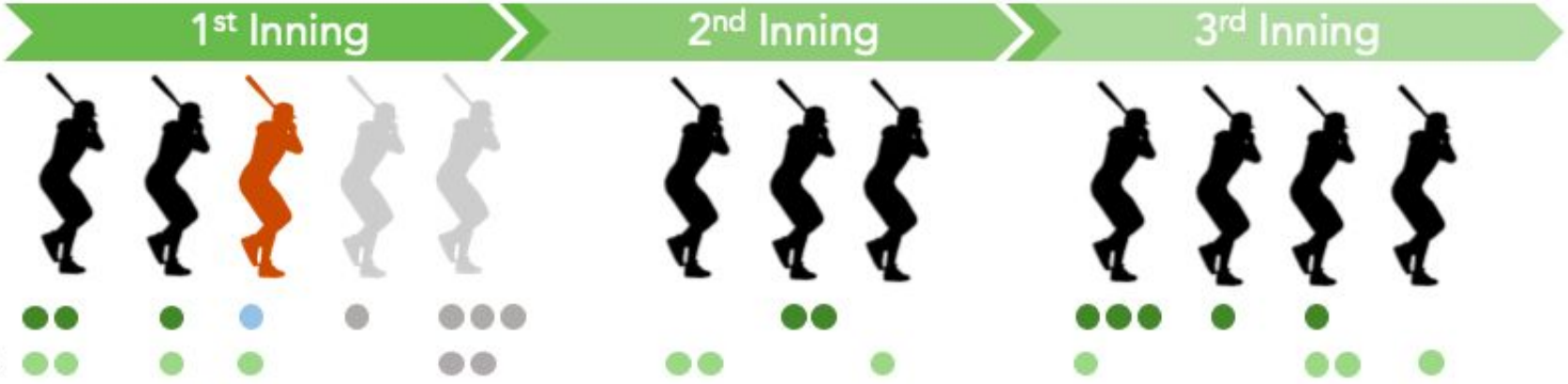

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

COLORADO
27

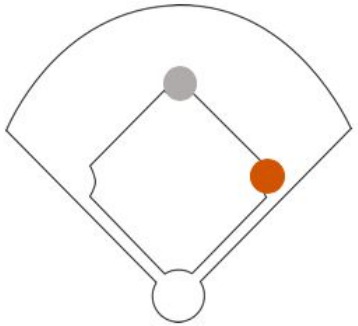

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

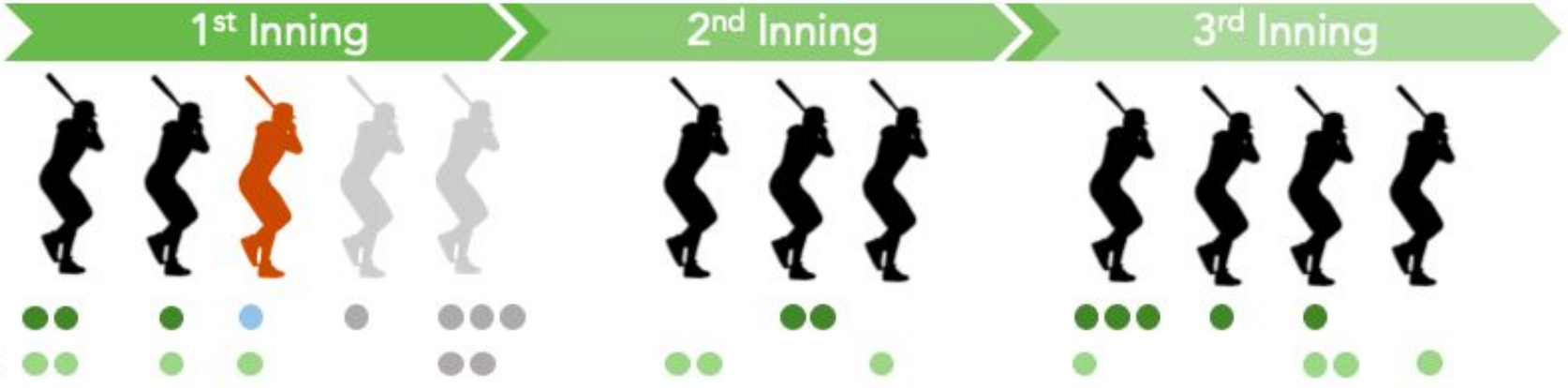
COLORADO
27



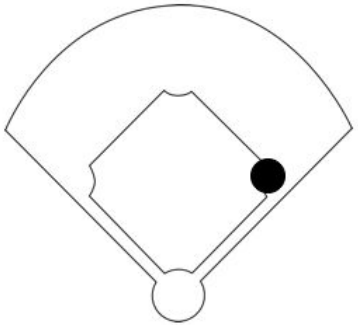

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

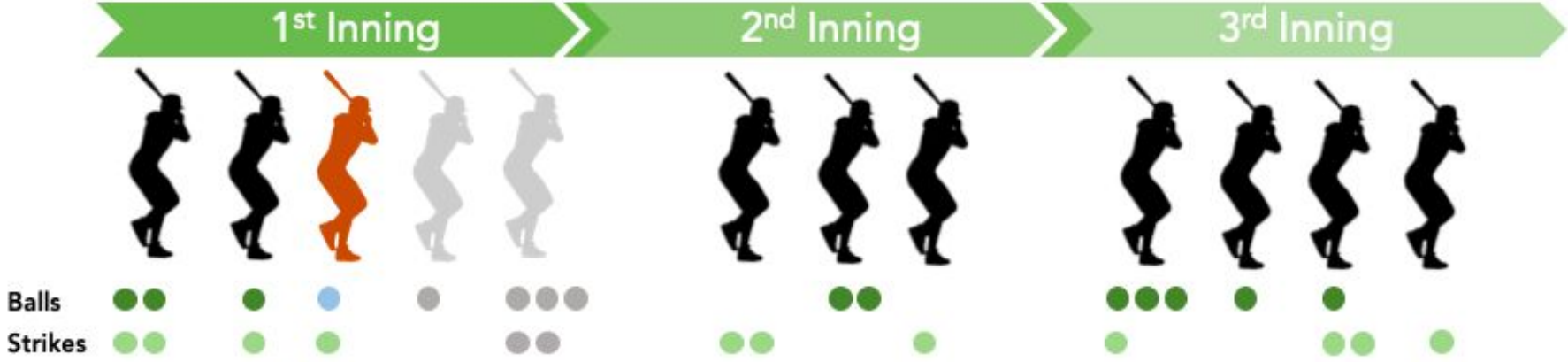
COLORADO
27

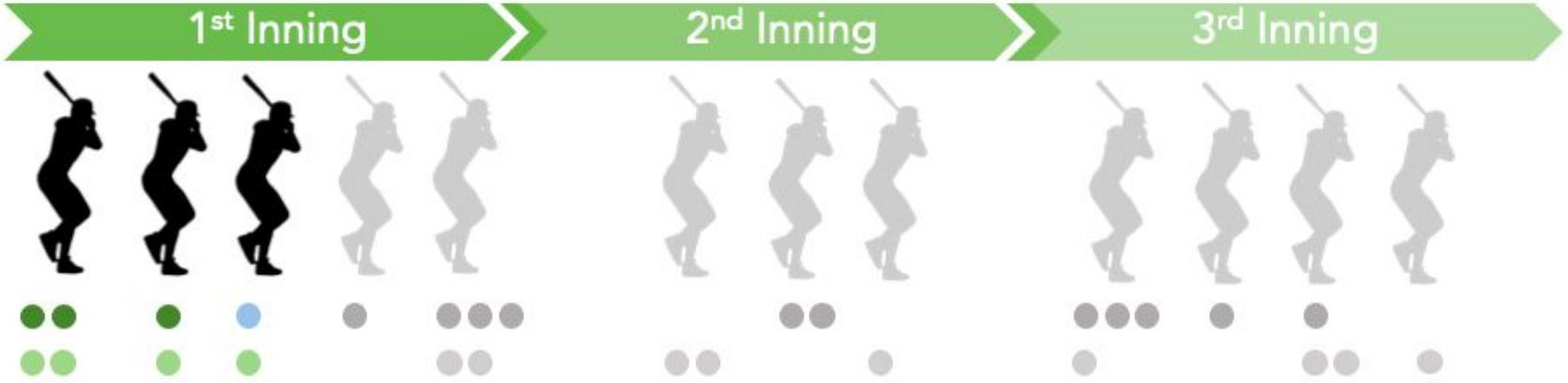


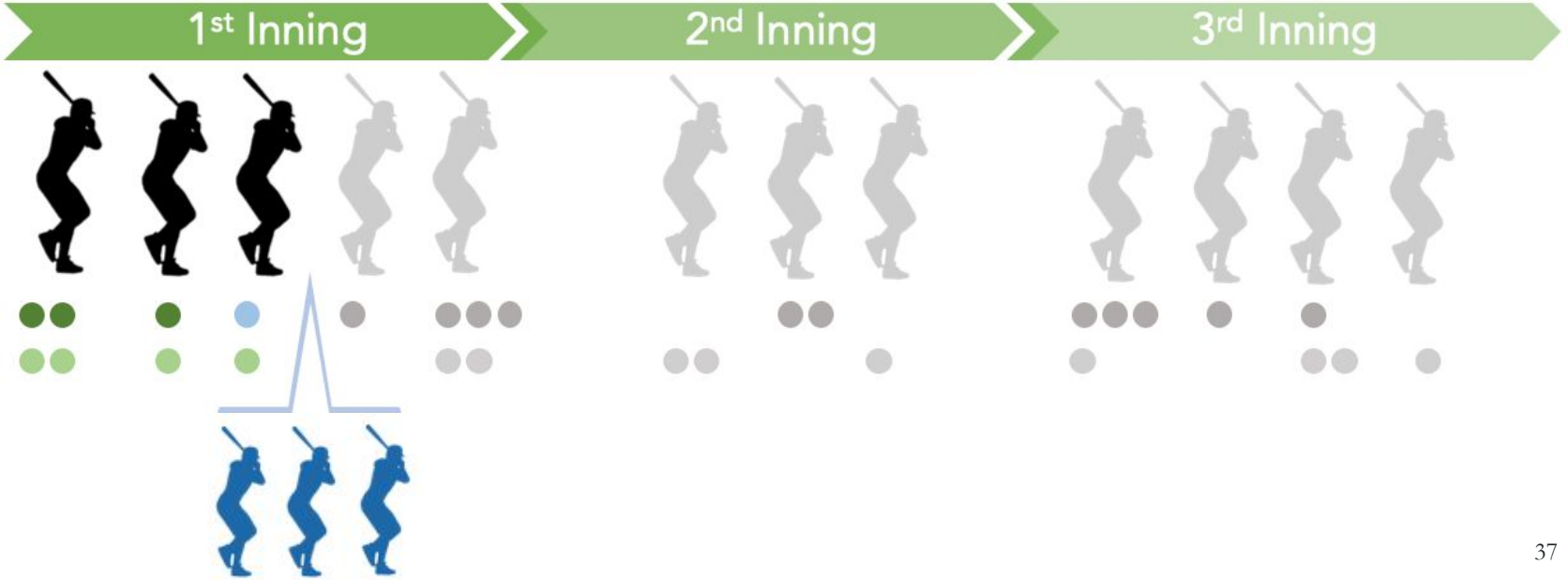

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

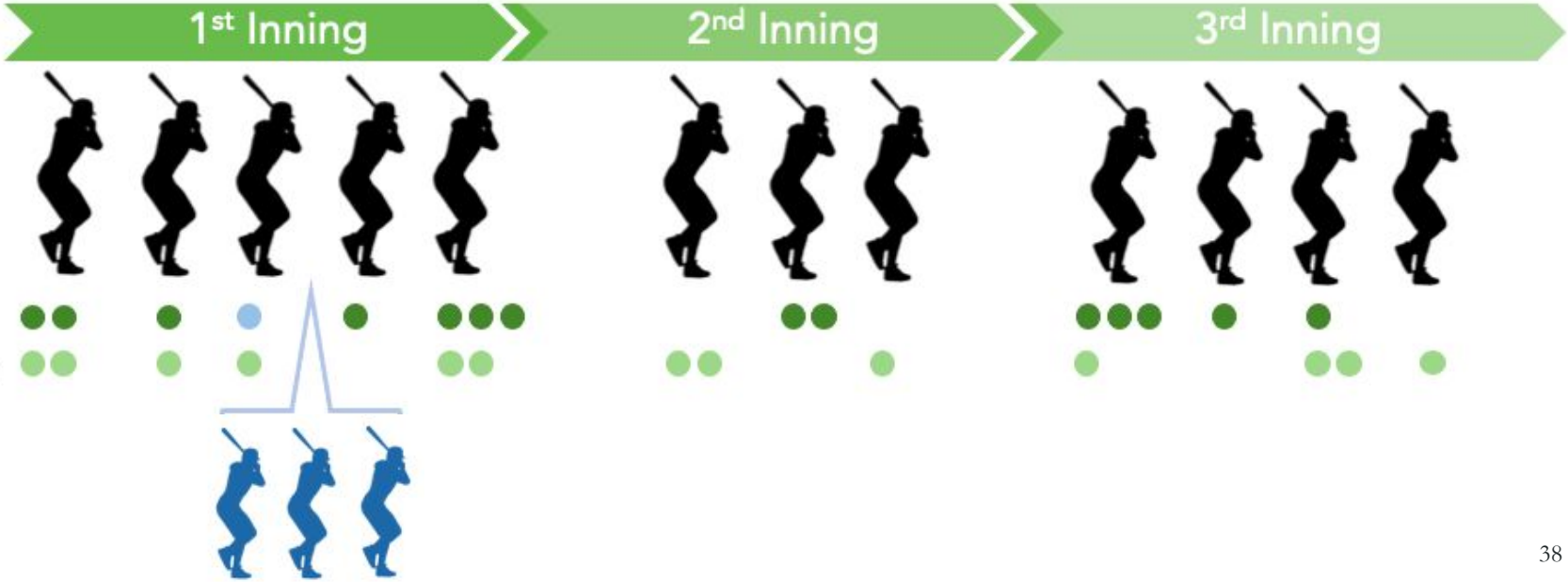
COLORADO
27

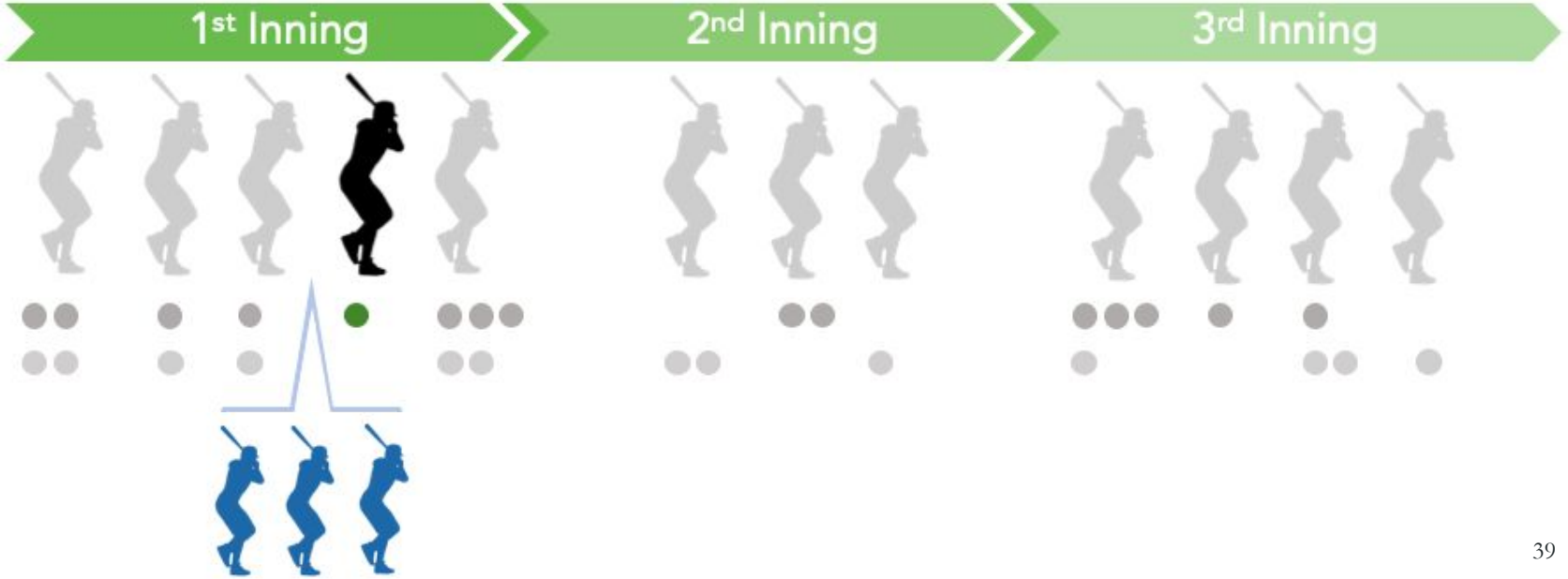


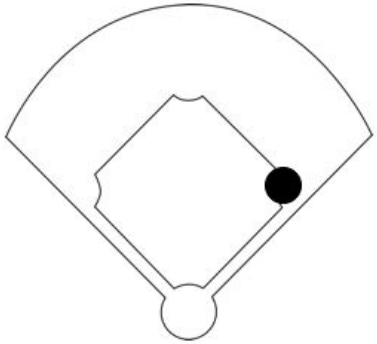

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

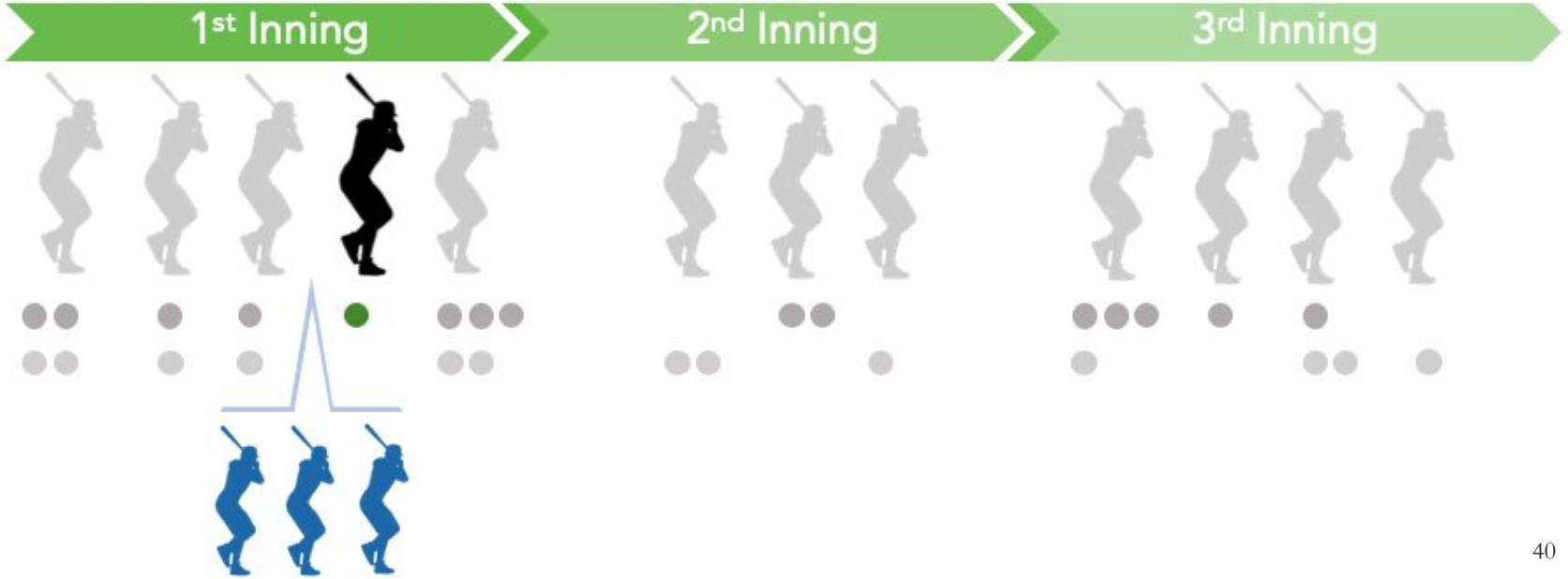

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

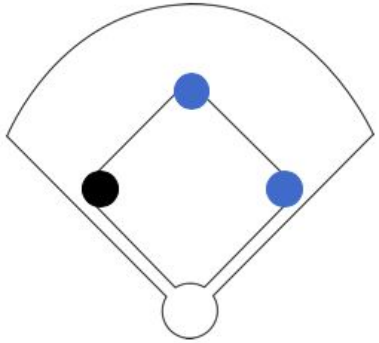

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

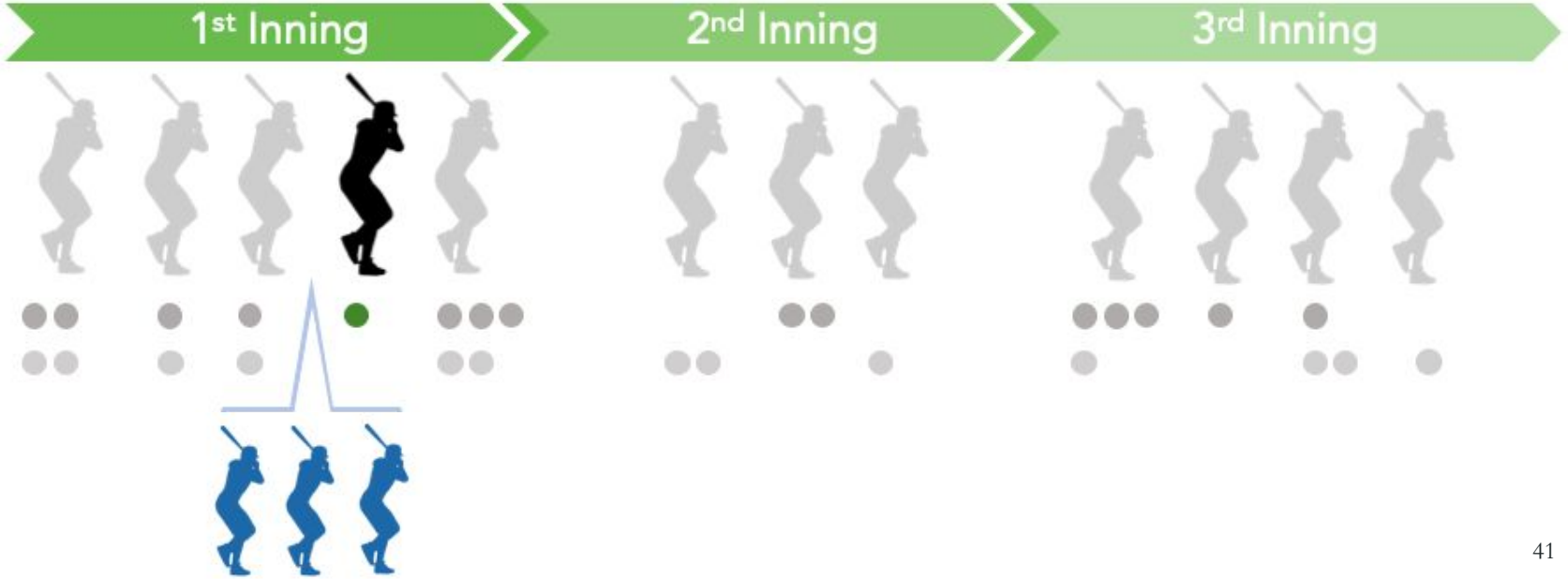

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

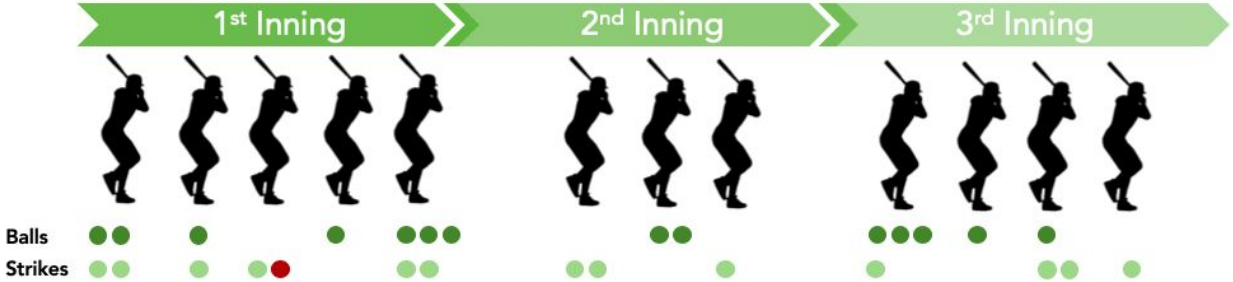

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

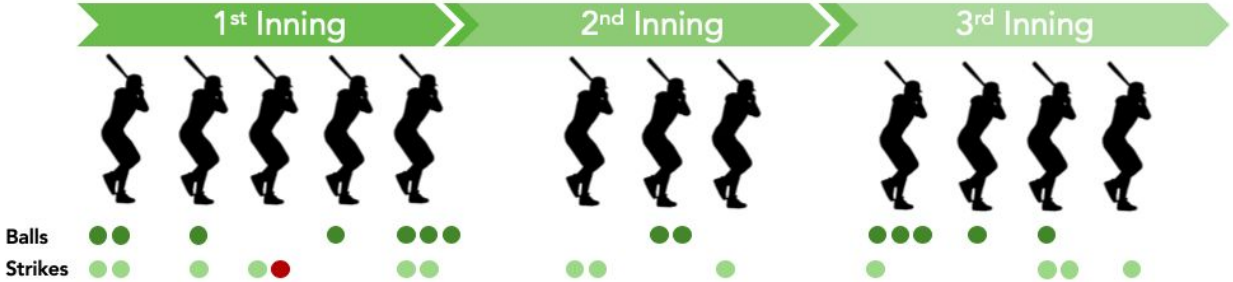


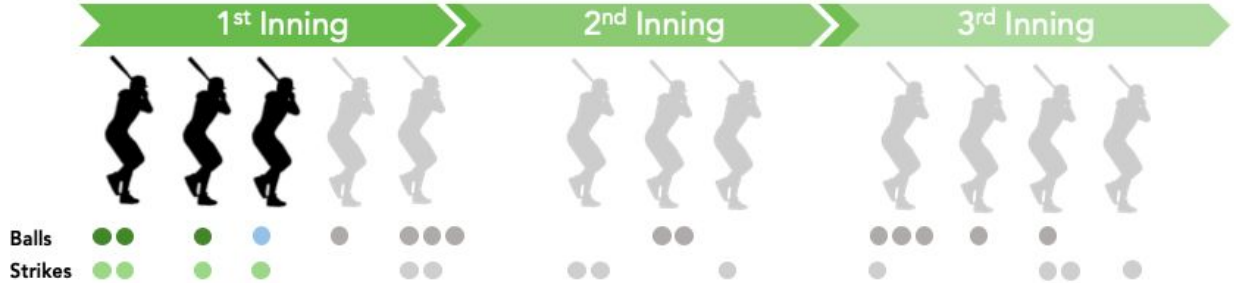

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

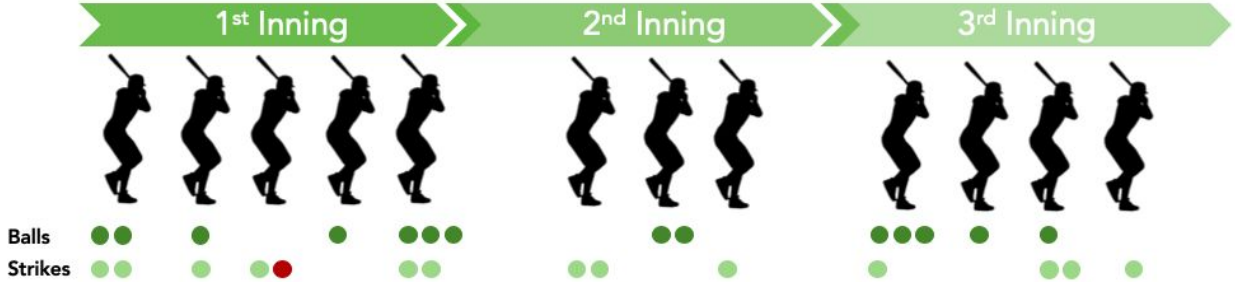


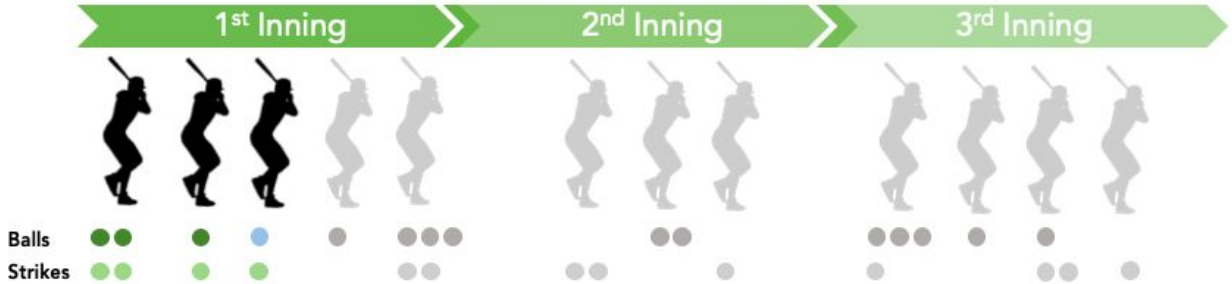

1st Inning
2nd Inning
3rd Inning
Balls
Strikes

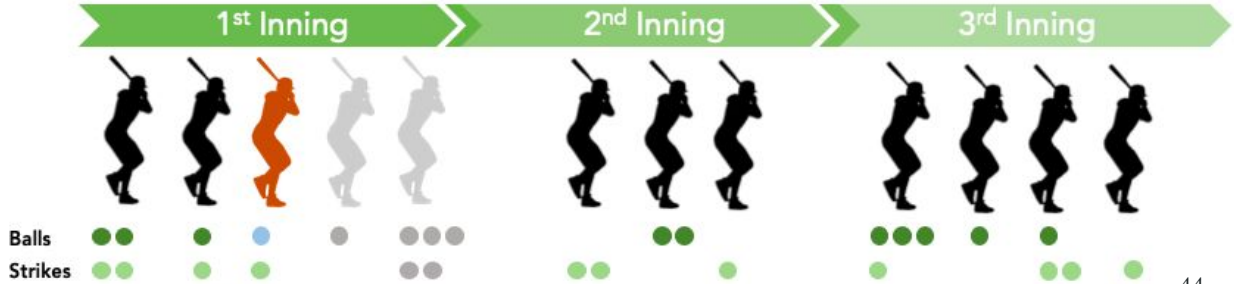

1st Inning
2nd Inning
3rd Inning
Balls
Strikes


1st Inning
2nd Inning
3rd Inning
Balls
Strikes

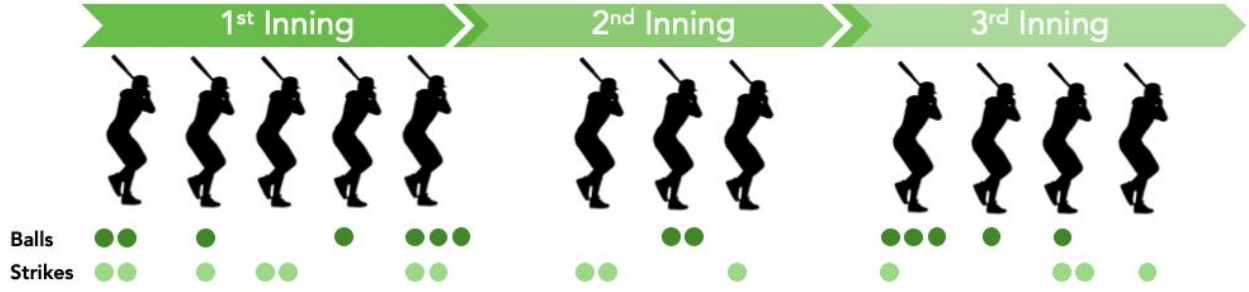

1st Inning
2nd Inning
3rd Inning
Balls
Strikes


1st Inning
2nd Inning
3rd Inning
Balls
Strikes


1st Inning
2nd Inning
3rd Inning
Balls
Strikes


1st Inning
2nd Inning
3rd Inning
Balls
Strikes

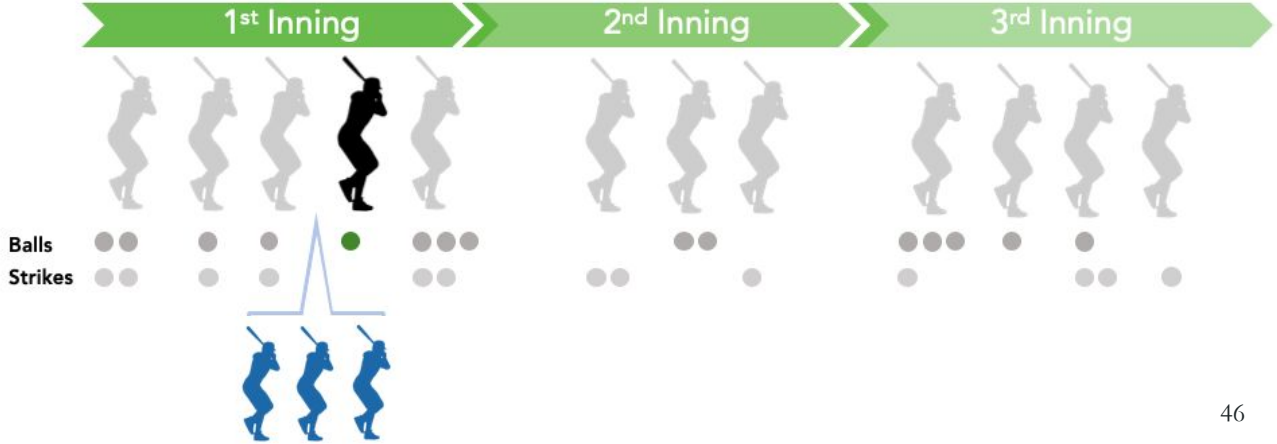
Predictive Model Results

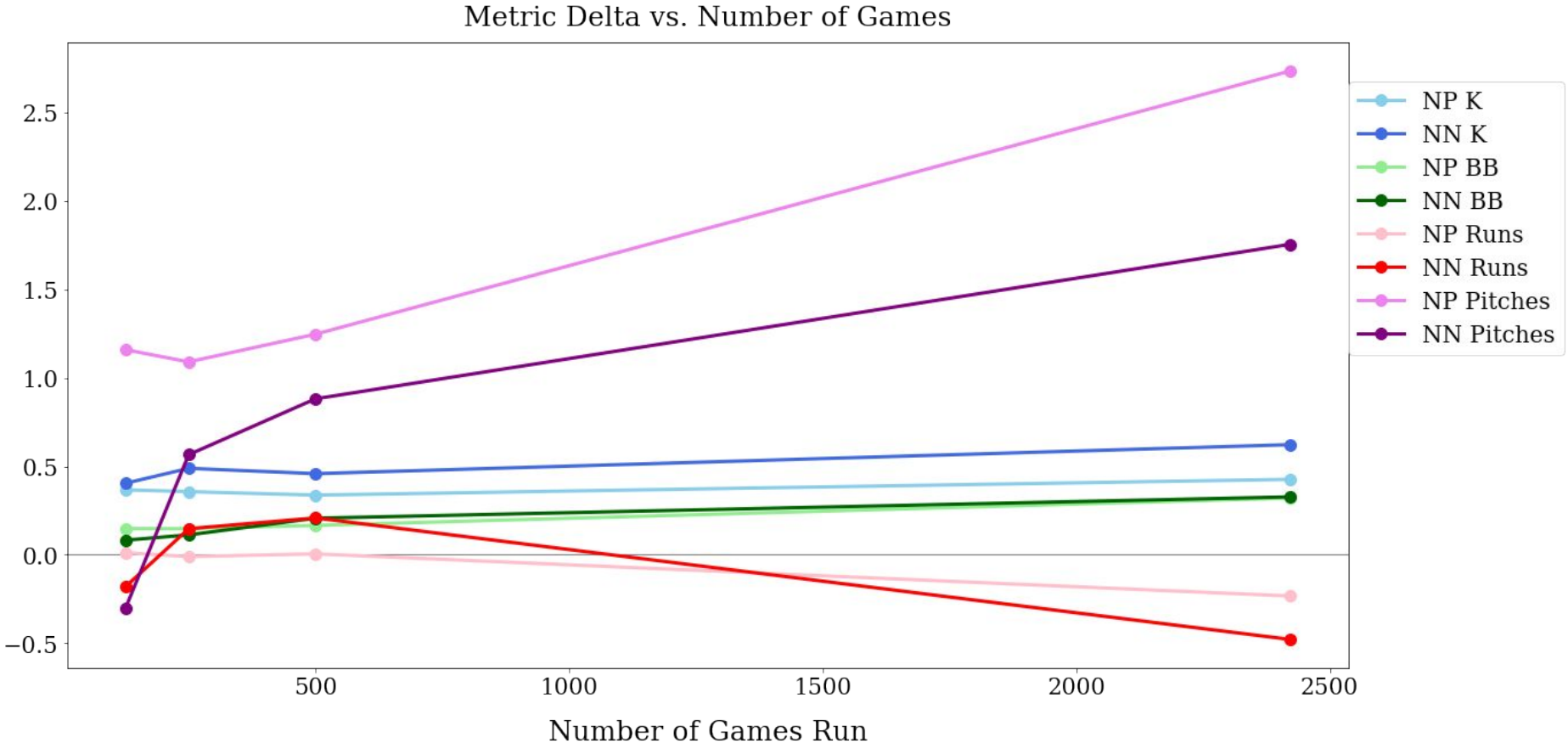


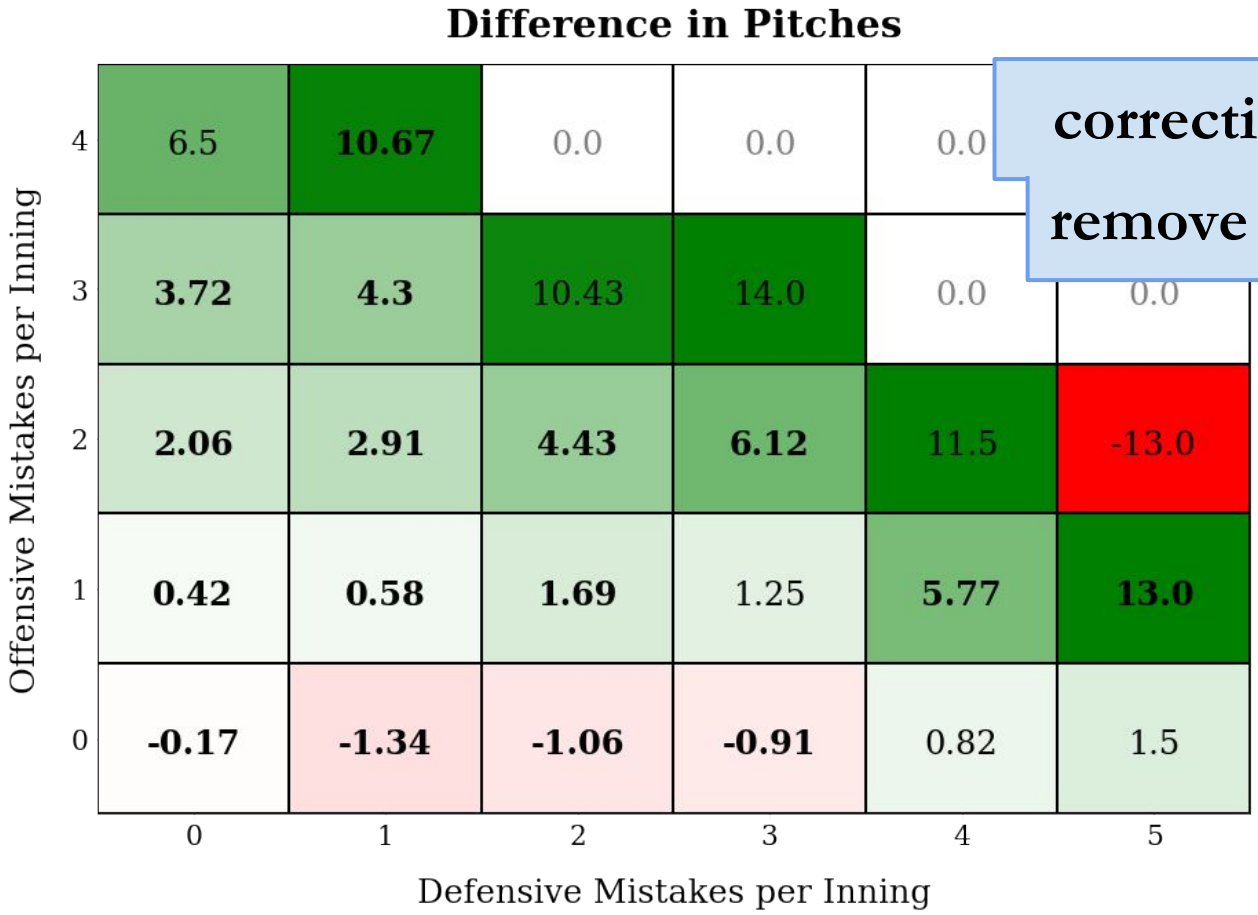
Original Innings

vs.

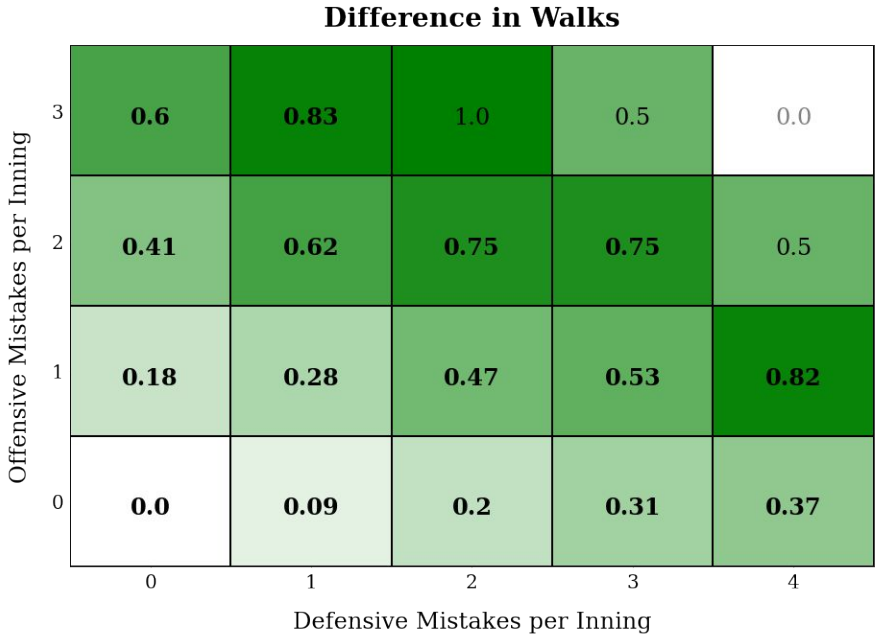
Corrected Innings




Metric Delta vs. Number of Games
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
500
1000
1500
2000
2500
Number of Games Run
NP K
NN K
NP BB
NN BB
NP Runs
NN Runs
NP Pitches
NN Pitches

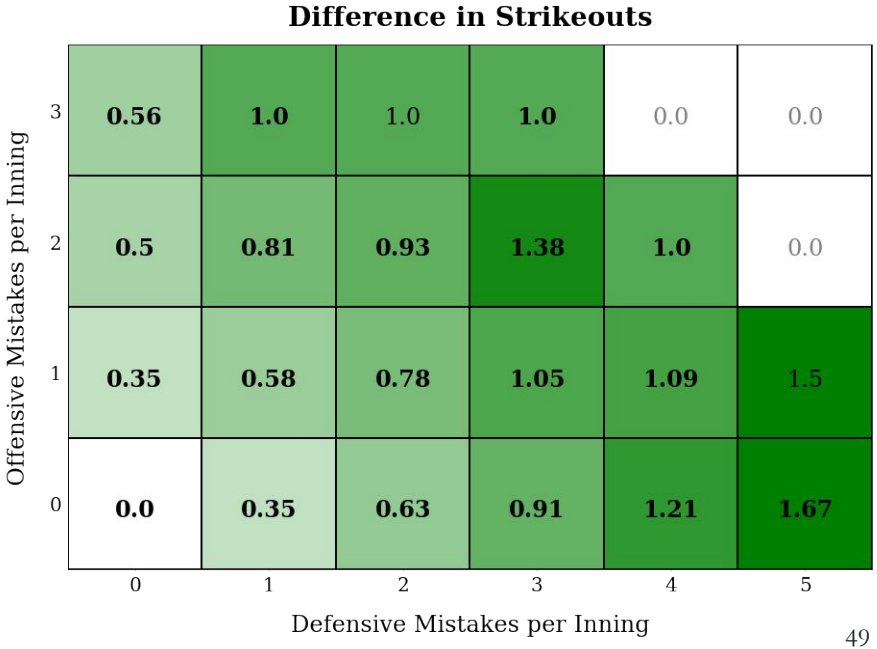


correcting mistakes can remove pitch advantage

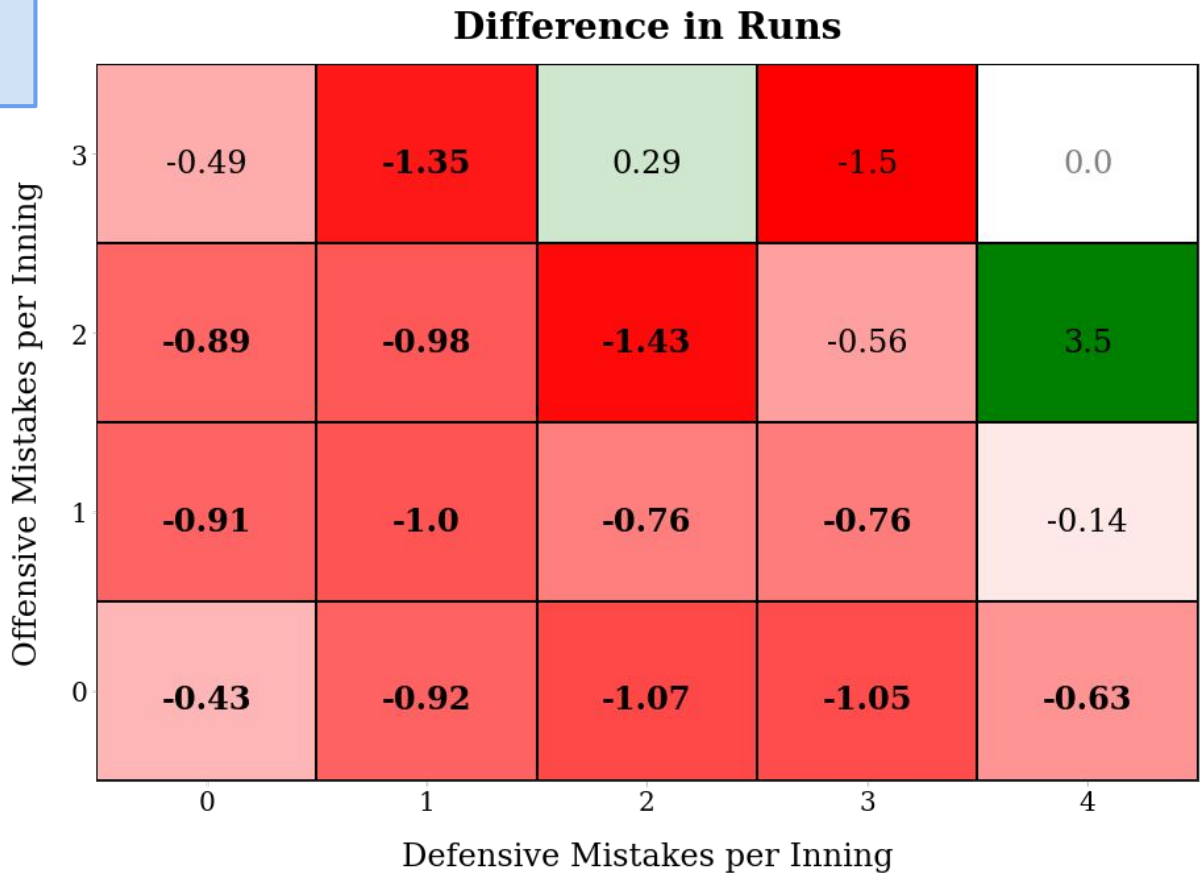


similar decrease in walks across all corrections

fewer strikeouts when correcting defensive mistakes



consistent increase in runs scored



Mets
20

Future Work

Future Work

- Full implementation
- Game manipulation
  - Size and shape
  - Consistency
  - Desired outcomes

What do we want the game of baseball to look like?

questions?