The Value of Statcast in MLB

By: Devin Maiello

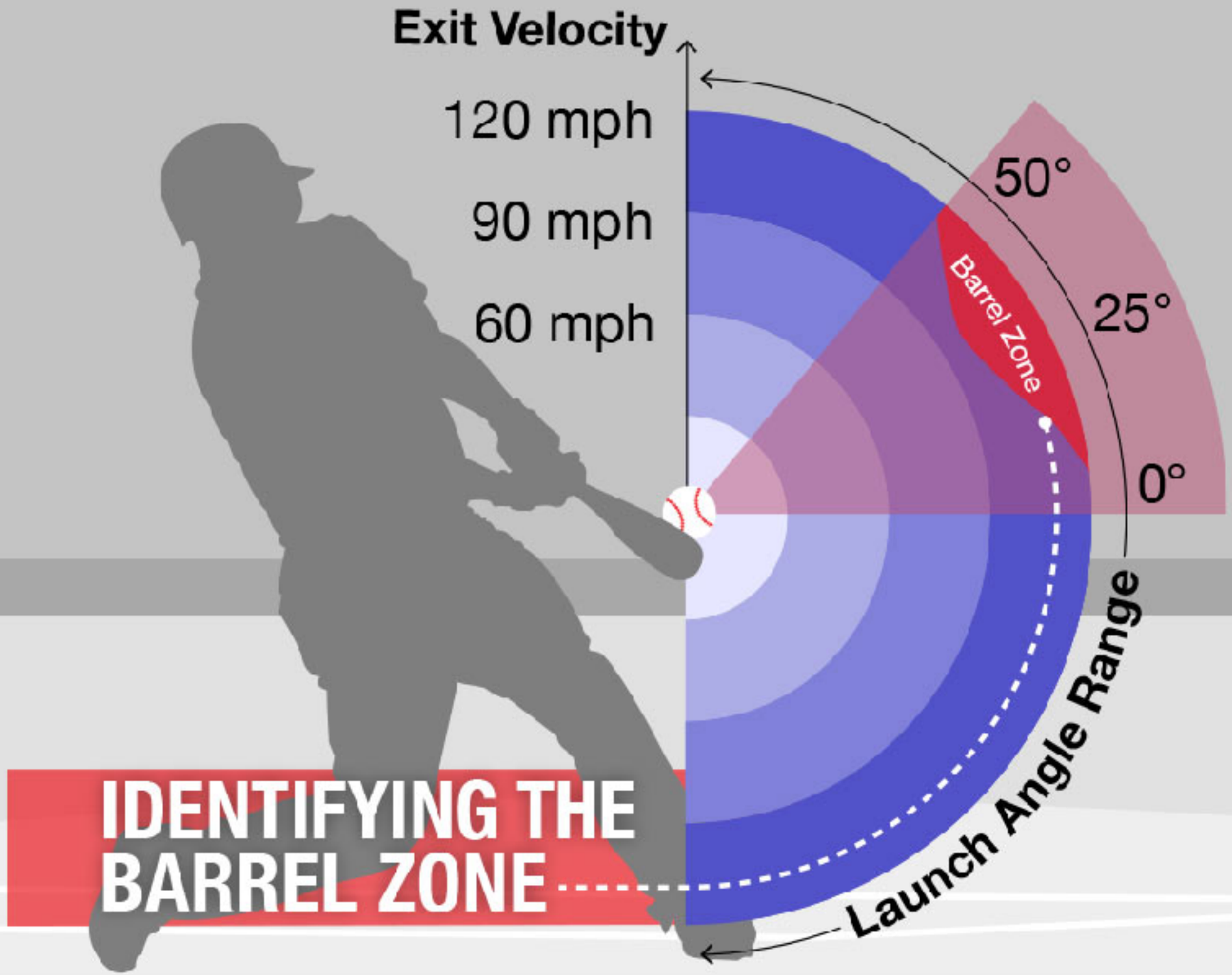
What is Statcast?

- Really powerful cameras
- Used to measure plays from more perspectives
- Launch Speed/Angle off bat, etc.
- Started being used in MLB in 2015

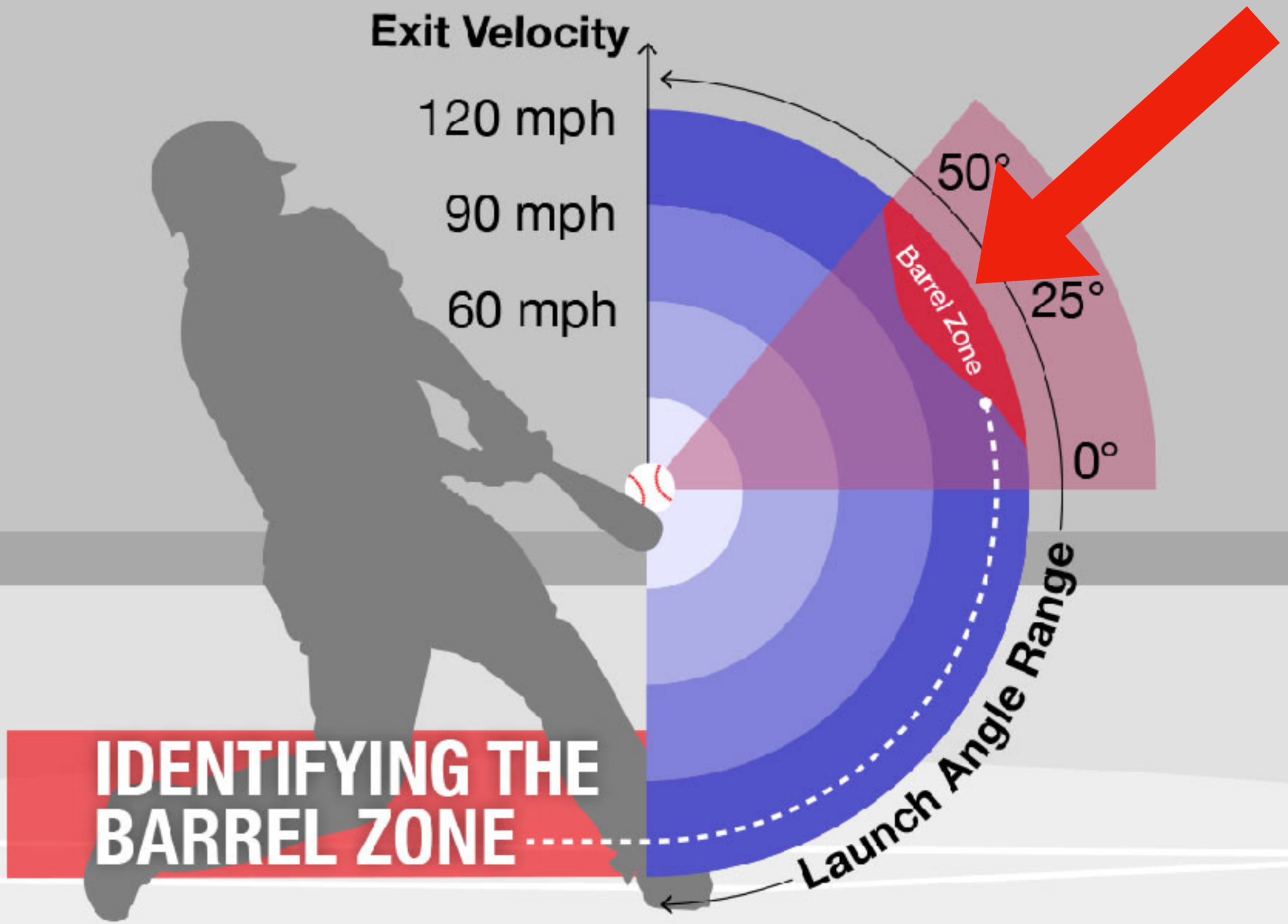
STAT CAST
POWERED BY: aws

Statcast In Action

Angle, Velocity, Barrels

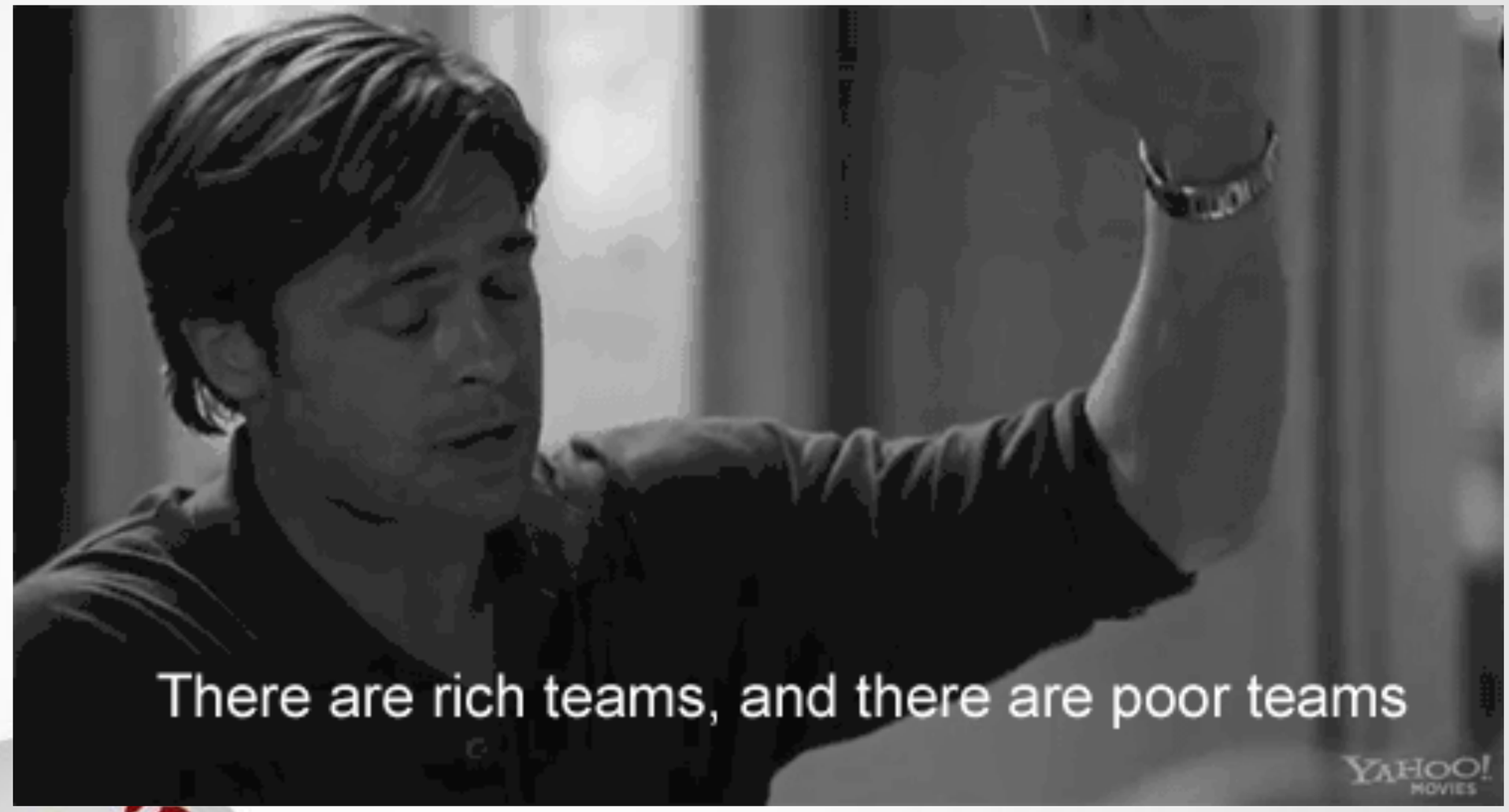


Angle, Velocity, Barrels



MLB Problem

- Measuring players' talent/value is hard
  - MLB keeps track of a lot of stats…probably too many for General Managers
  - GMs have tens (sometimes hundreds) of millions of dollars to field a team
  - GMs want to field a winning team
  - Use data to figure out how to spend money efficiently for Wins
  - (obligatory "money-ball" reference … sorry)


There are rich teams, and there are poor teams
YAHOO! MOVIES

Can Statcast Data Help?

Are players who hit the ball hard and in the barrel range <u>consistently</u> also the most valuable?

Player Value = WAR

- Wins Above Replacement (Our Target)
  - How many more wins does this guy add to the team compared to some bench scrub?

WAR = (Batting Runs + Base Running Runs + Fielding Runs + Positional Adjustment + League Adjustment +Replacement Runs) / (Runs Per Win)

Regression Problem

- Can we look at an anonymous player's Statcast numbers and be able to predict their value (WAR) accurately?

Regression Problem

- Can we model player value (WAR) <u>accurately</u> using Statcast data?
  - <u>More accurately</u> than traditional baseball stats?

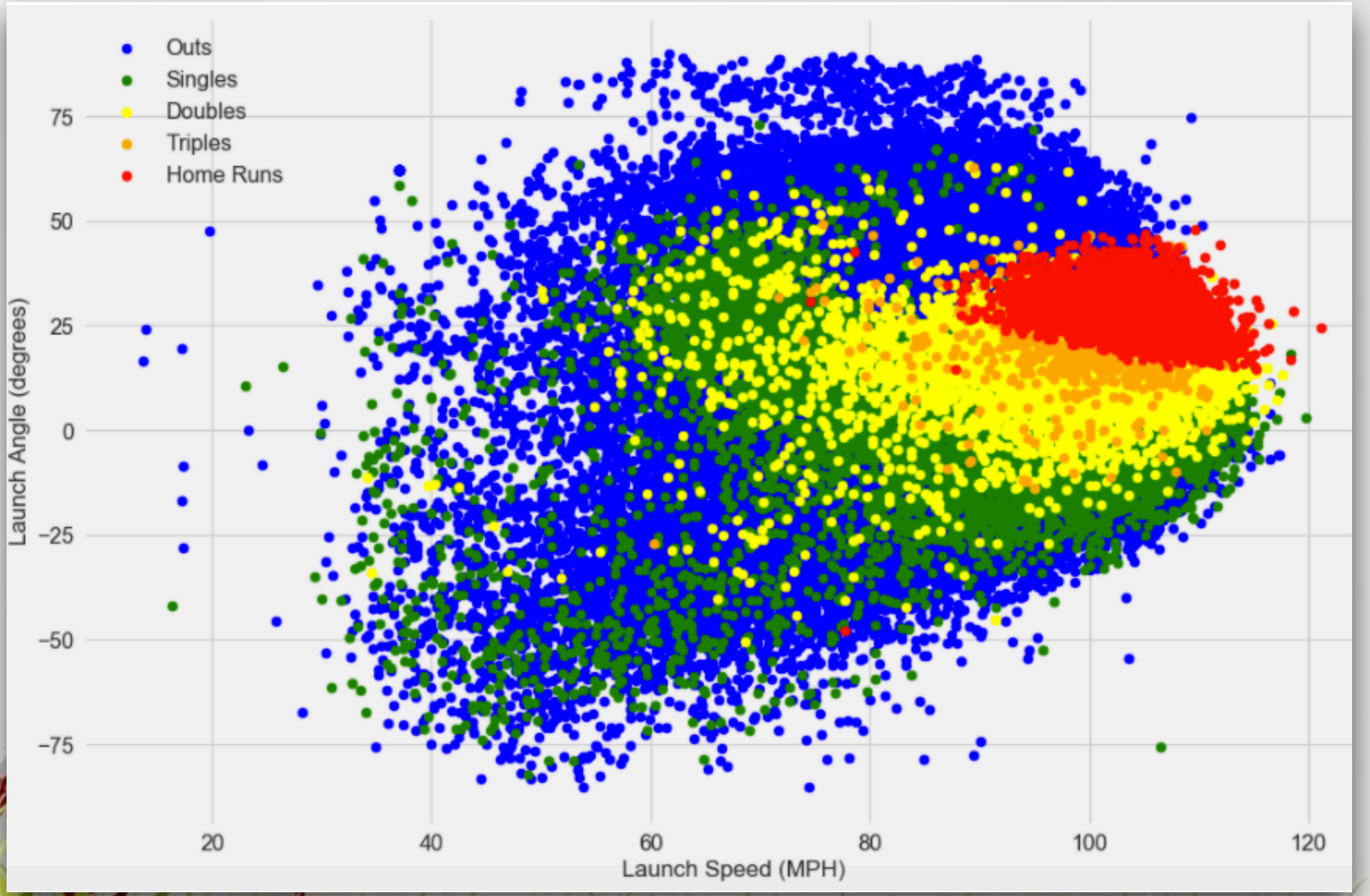
(Batting Average, RBIs, Home Runs, etc.)

Data Sources

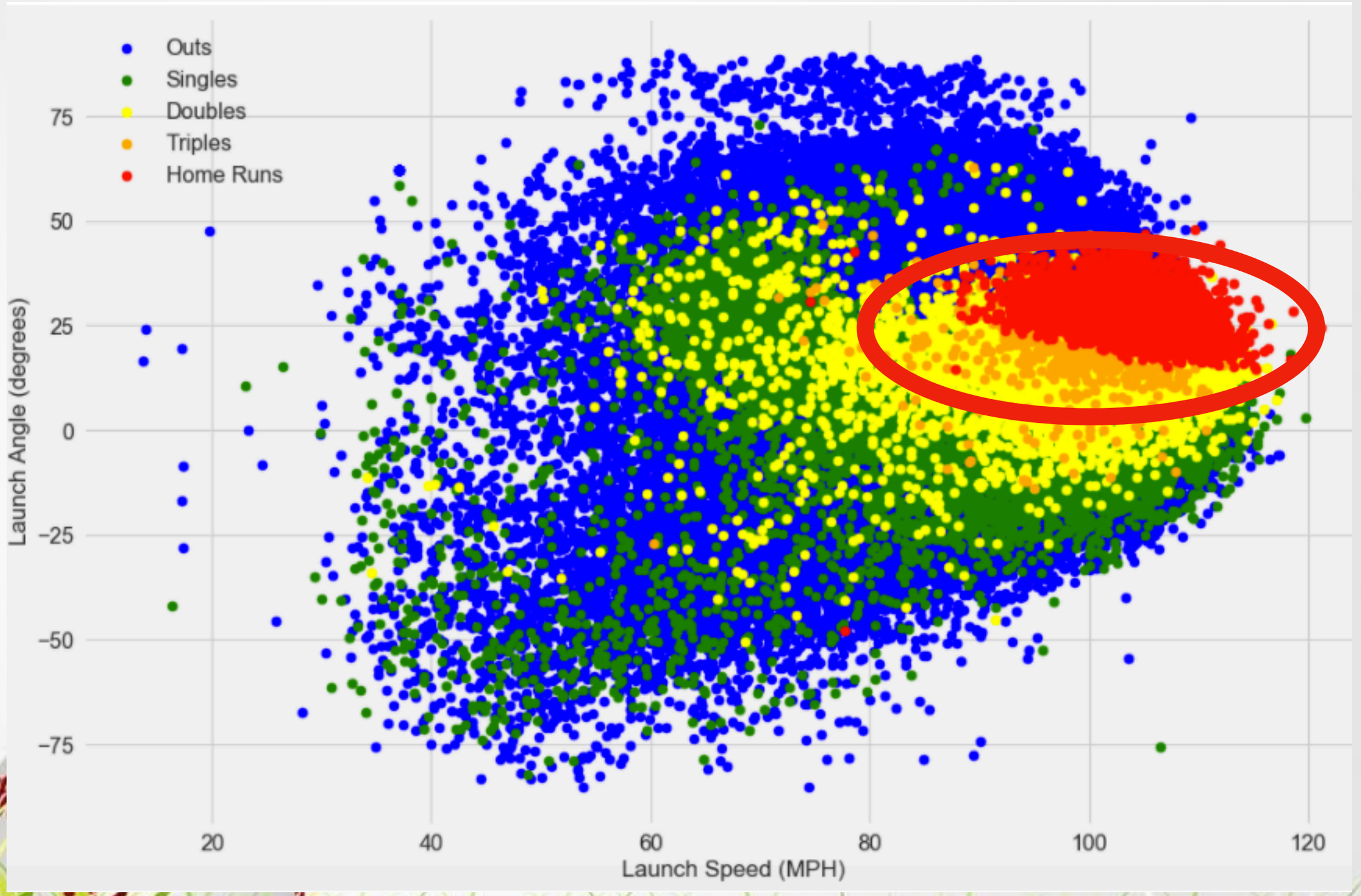
- Scraped with Beautiful Soup:
  - Fangraphs (traditional + some more advanced stats)
  - Baseball-Savant (Statcast data)
    - 2015 - 2016 (~150 players qualified per year)

Data Viz / EDA

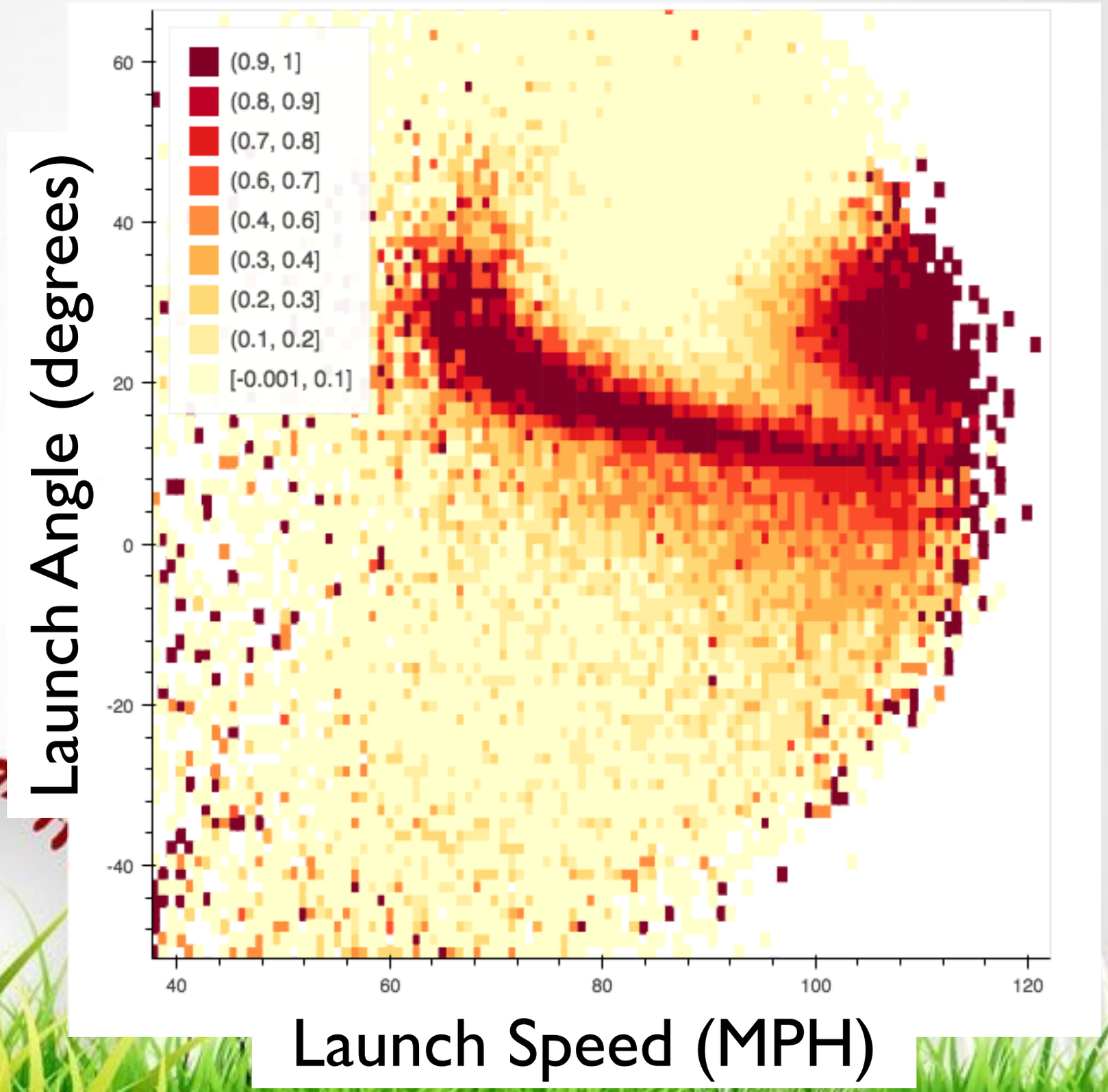
Hits and Outs from Launch Speed vs. Launch Angle



Hits and Outs from Launch Speed vs. Launch Angle

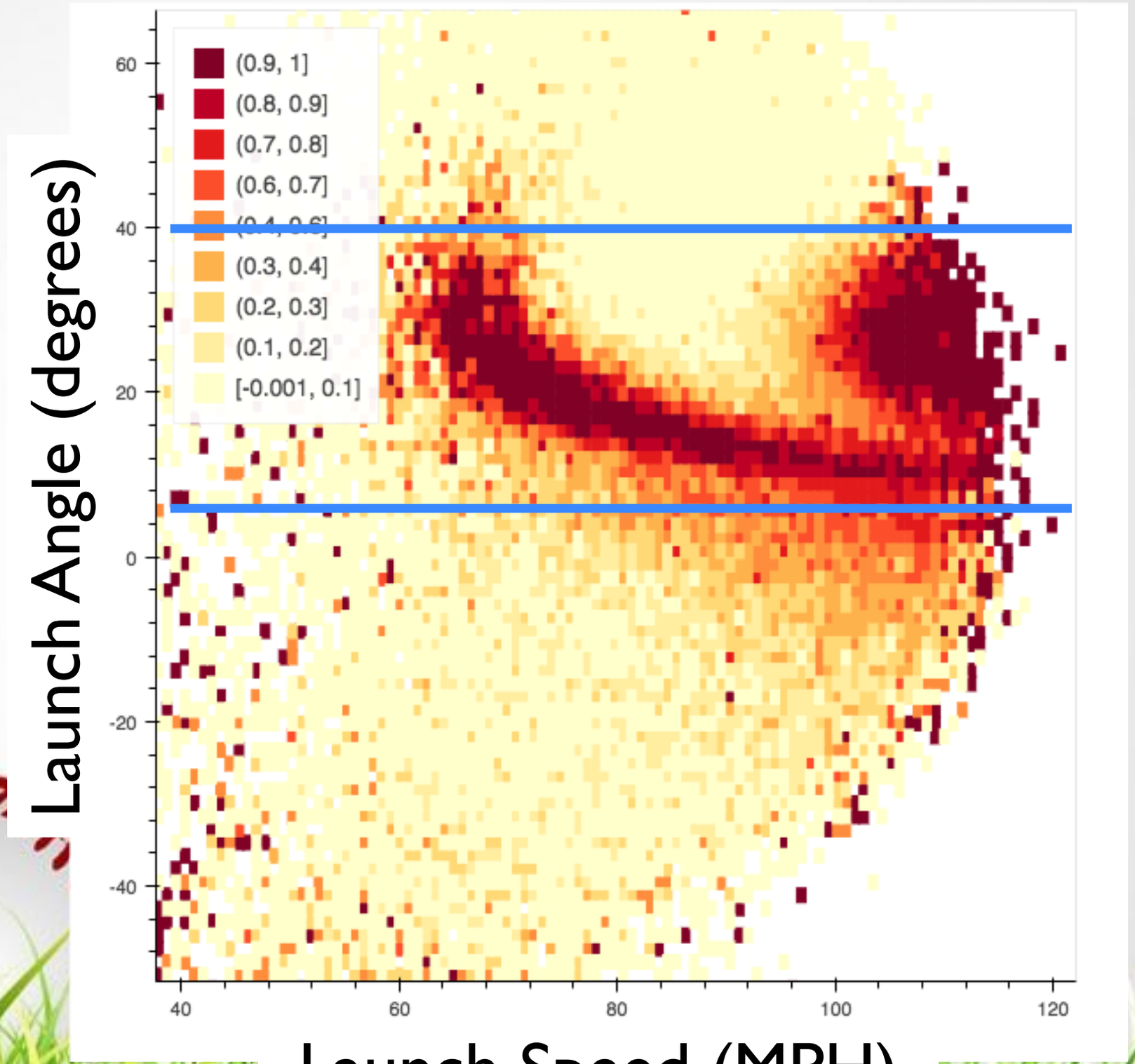


Statcast Hit Distribution



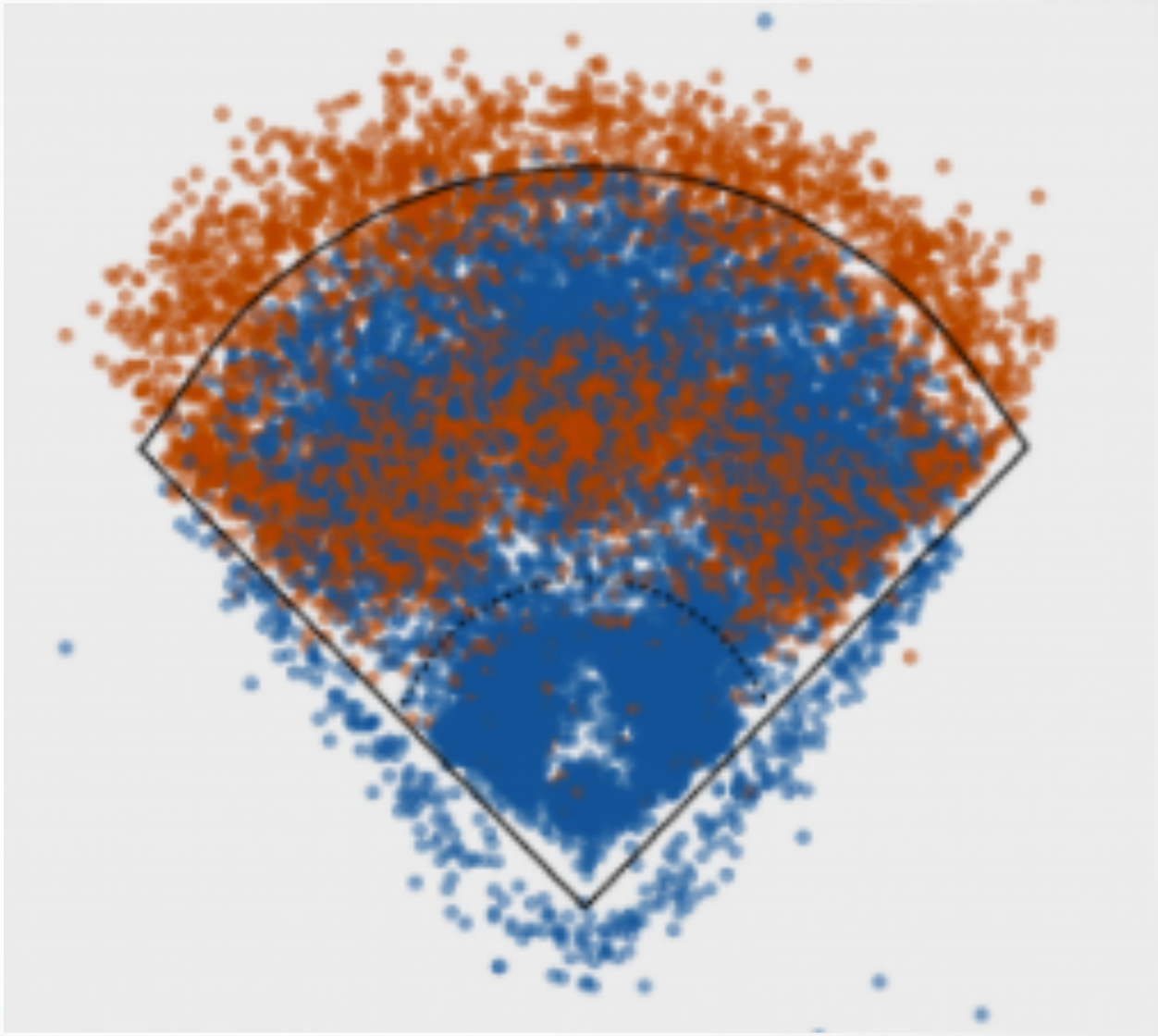
Bokeh Library

Statcast Hit Distribution



Bokeh Library

Explains that Distribution



- Hits
- Outs

Correlations - Traditional Stats

	WAR
Batting Average	0.560
Home Runs	0.406
Runs	0.713
RBI	0.424
Stolen Bases	0.202

Correlations - Advanced Stats

	WAR
OPS	0.730
ISO	0.457
BABIP	0.417
Fly Ball%	0.151
HR / Fly Ball	0.318

Correlations - Statcast … !?

	WAR
Avg Launch Angle	0.119
Avg Launch Speed	0.282
Barrels / Batted Ball	0.322
Total Barrels	0.404

Why!?

Example of Why

AVG Vel: 96 mph
Barrel: 26.4%

AVG Vel: 85 mph
Barrel: 6.5%

Aaron Judge
6’ 7”
280 lbs
1st Place WAR

Aaron Judge
6’ 7”
280 lbs
1st Place WAR

Jose Altuve
5’ 6”
165 lbs
2nd Place WAR
!!!!

More Why

- Statcast is mostly a measure of power
- Valuable players also:
  - Work walks / only swing at good pitches to hit
  - Don’t strike out too much
  - Are good fielders (defense)

Statcast doesn’t measure these things well!

Correlations - Statcast to SLG

	Slugging
Avg Launch Angle	0.322
Avg Launch Speed	0.672
Barrels / Batted Ball	0.748
Total Barrels	0.808

Modeling

Modeling Strategy

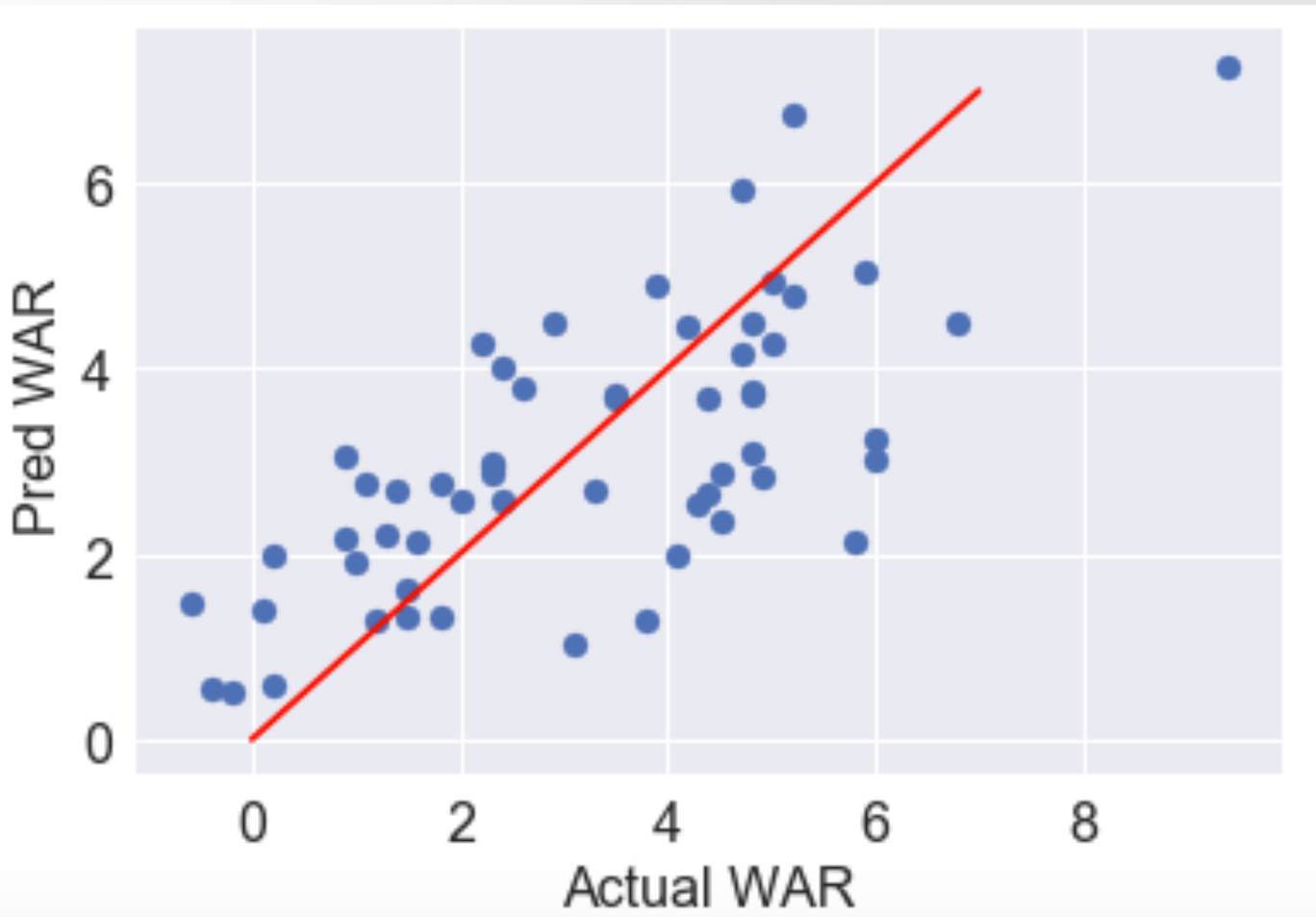
- Create regression models combining features and compare the scores:
  - Traditional (old-school, less informative)
  - Advanced (post money-ball era, more informative)
  - Statcast

Traditional (AVG, R, etc.)

- Linear Regression
- Cross Val Avg:

~0.4

Actual vs, Predicted

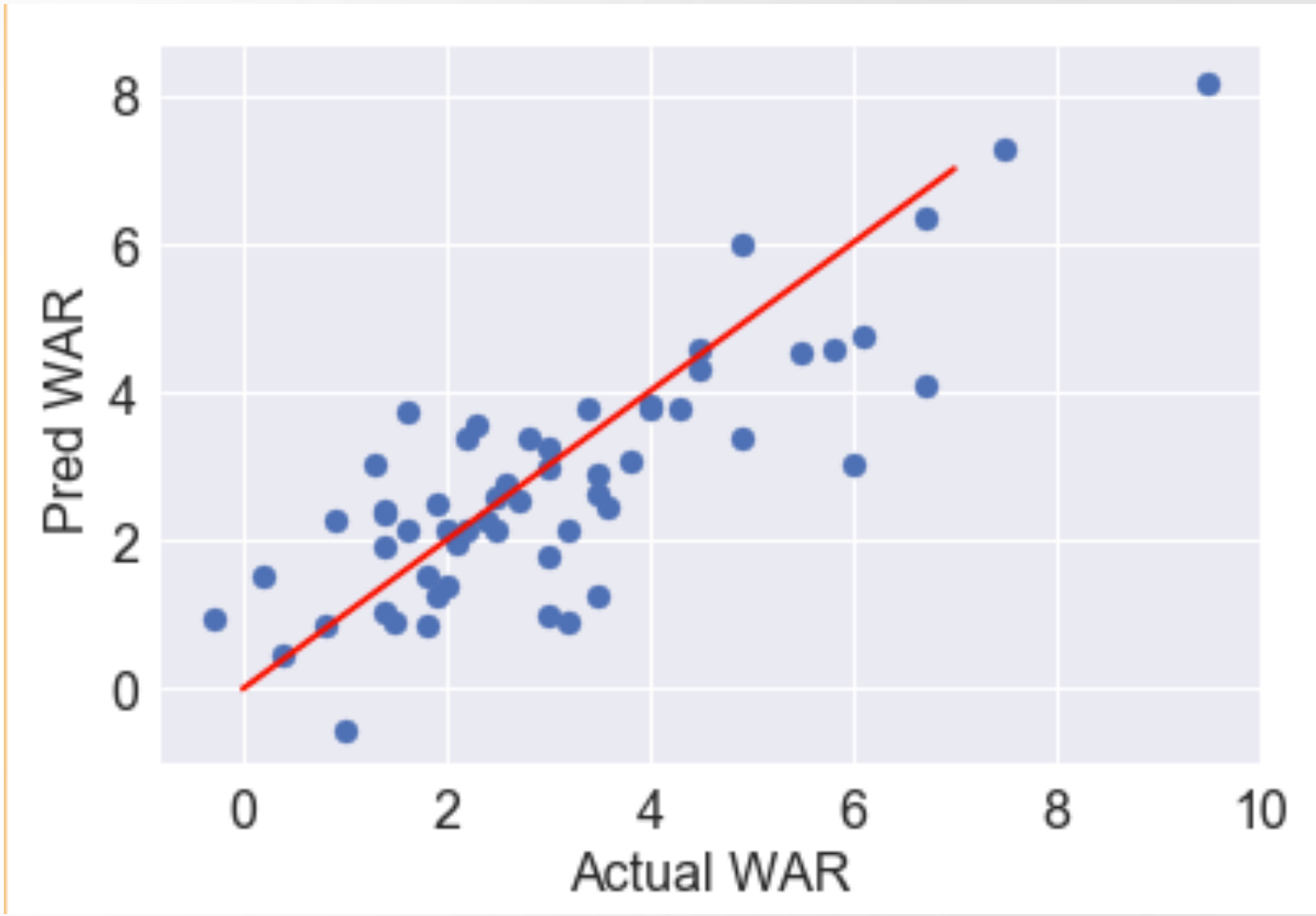


Advanced (OPS, FB, etc.)

- Linear Regression
- Cross Val Avg:

~0.518

Actual vs, Predicted

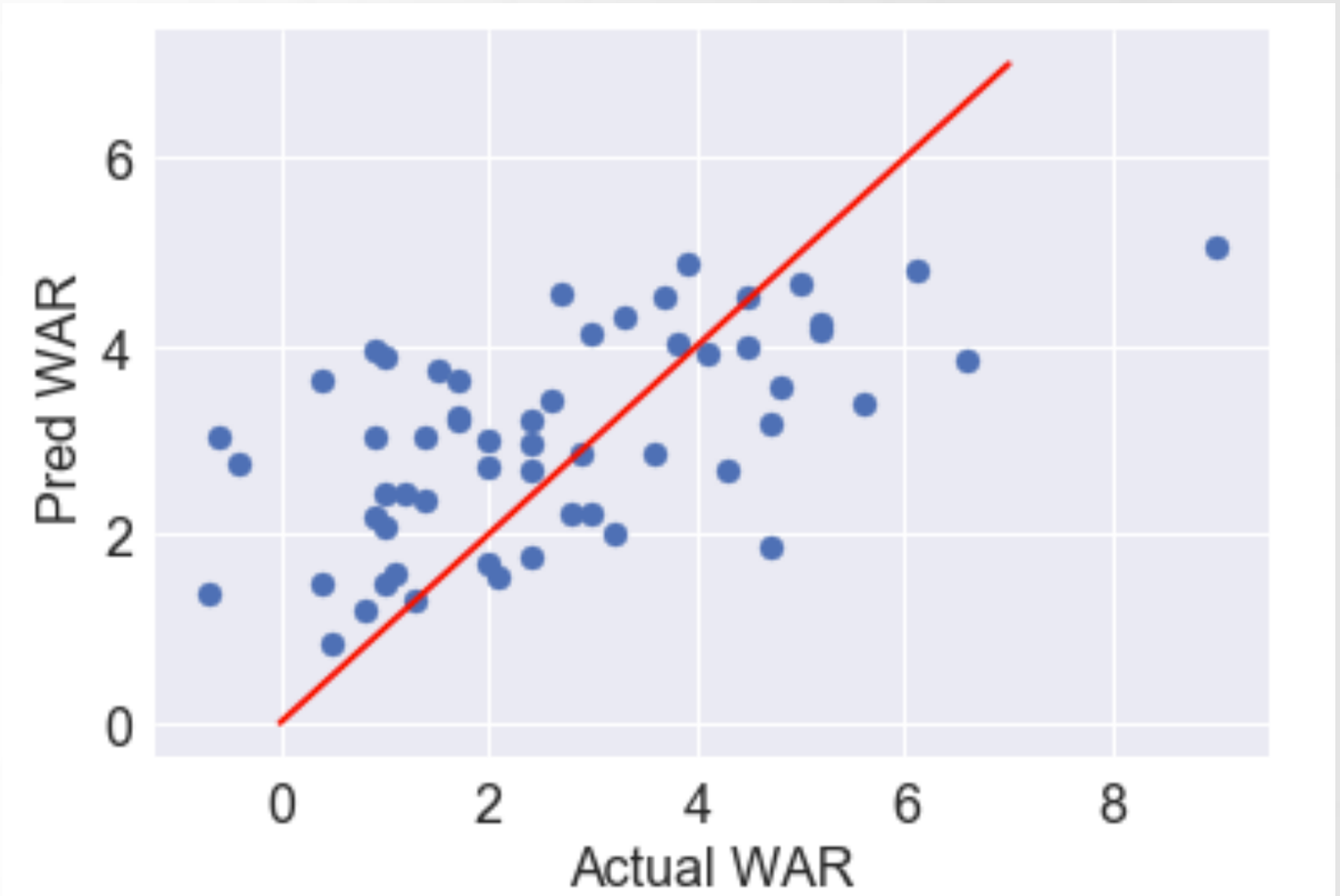


Statcast

- Linear Regression
- Cross Val Avg:

~0.25 :(

Actual vs, Predicted

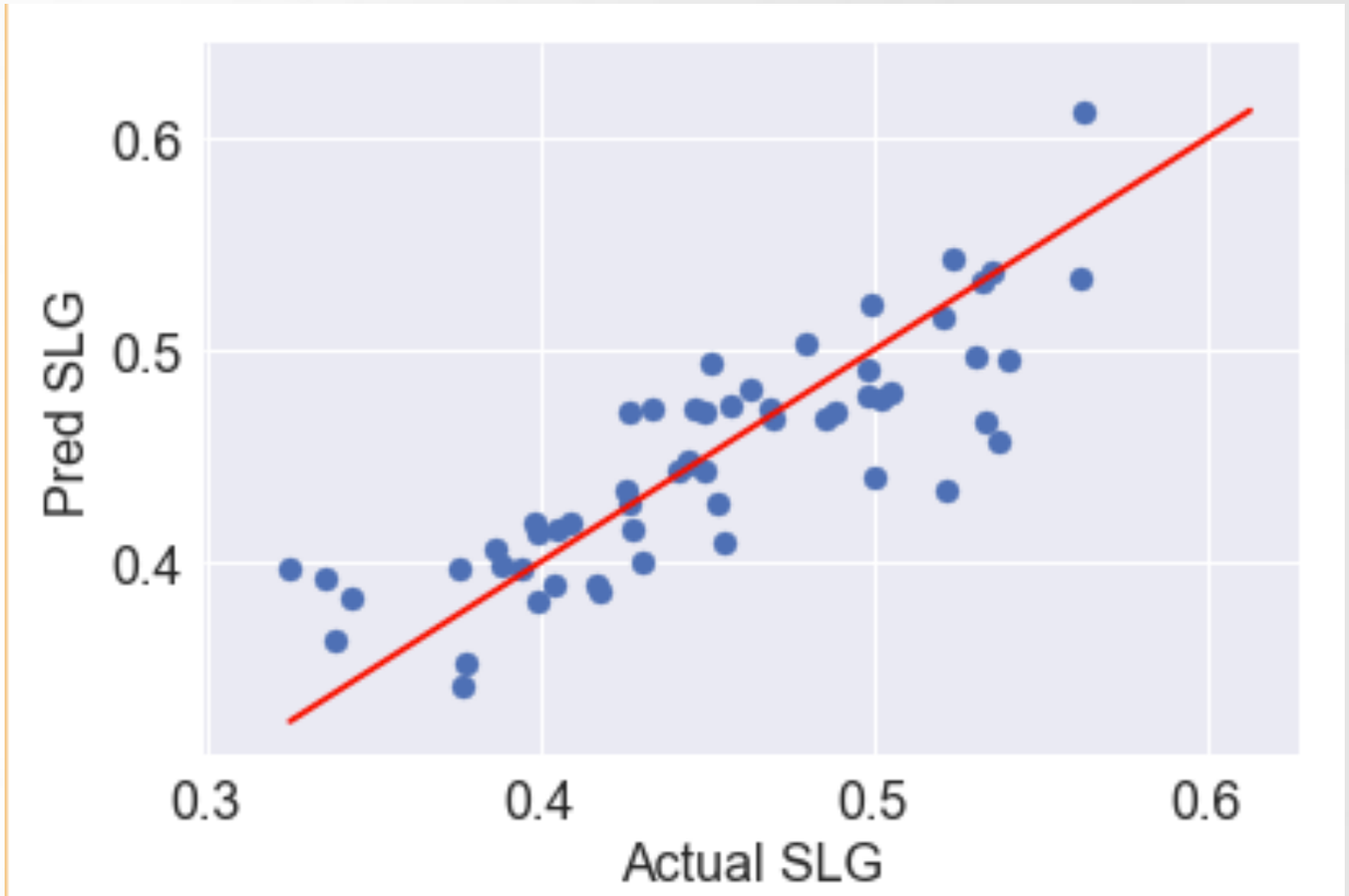


Statcast - SLG

- Linear Regression
- Cross Val Avg:

  ~0.5 :)

Actual vs, Predicted



Conclusion

- Baseball involves a lot of randomness, can't expect a nearly perfect model
- Statcast is interesting, but we (a GM) will also need many more features to measure player value
- Strong measure of a player's slugging ability, so there is some value to Statcast!
- Maybe better for player by player analysis rather than trying to generalize over the entire league

Thanks! Questions?

See my notebooks/code on my GitHub

username: *maielld1*