



# MLB Defensive Alignments

Sam Bornstein and Paul Sirena

KC FRANCO | 3B

KC MONDESI | SS

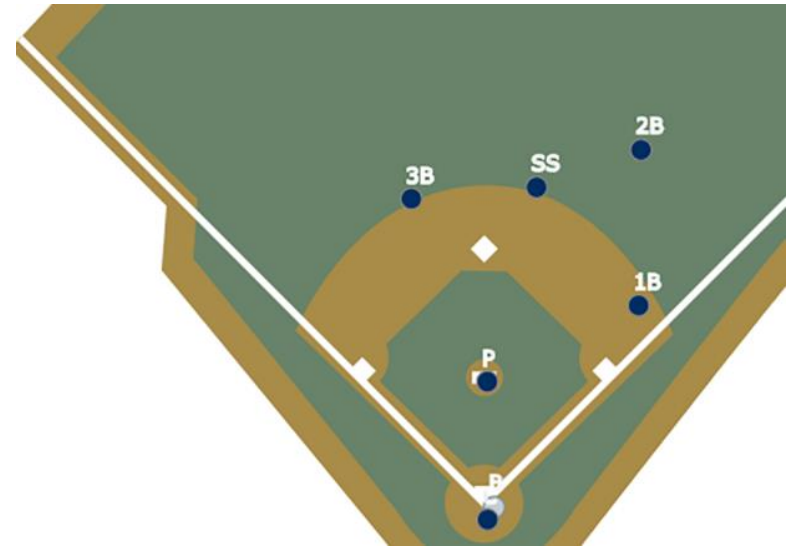
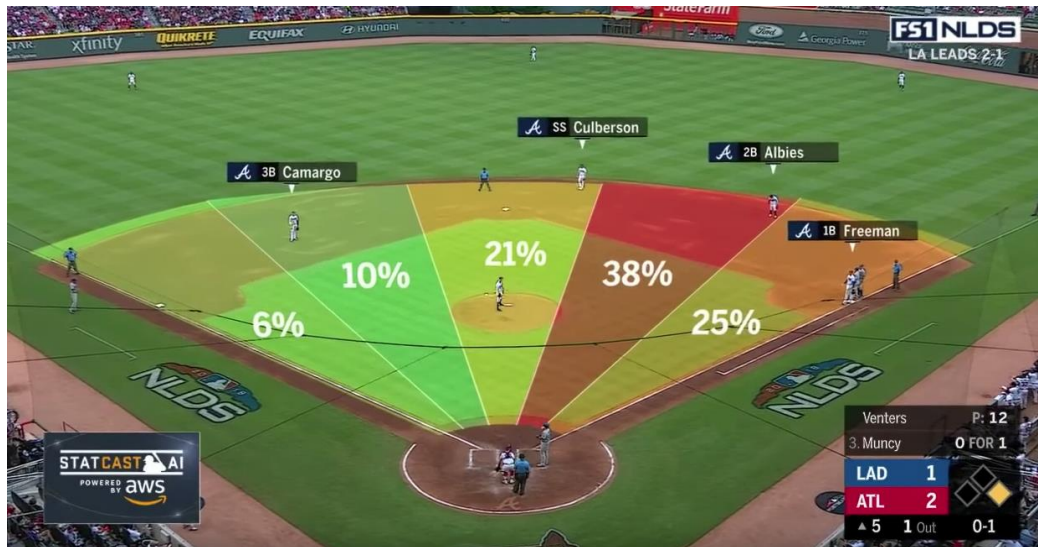
KC MERRIFIELD | 2B

KC O'HEARN | 1B

Montgomery		P: 45
KC	1	◆◆
DET	5	
▼ 2	2 Outs	1-2

# Background

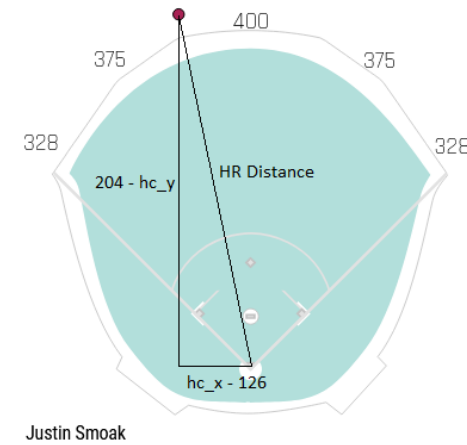
- Plate Appearances with Shift:
  - 2016: 13.7%
  - 2020: 33.6%





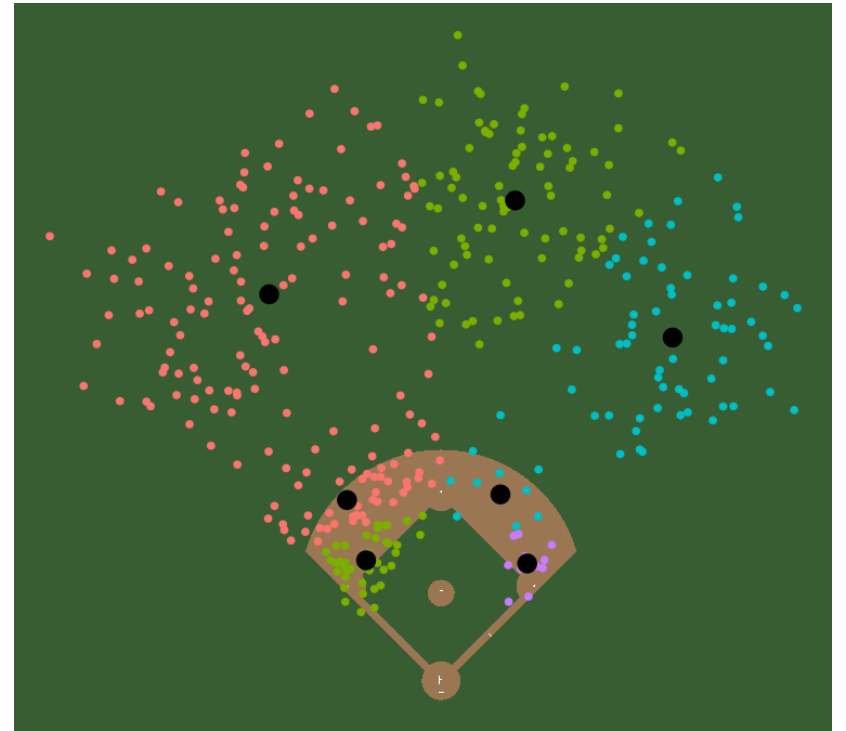
# Preparation

- Baseball Savant
- Modeling Techniques
- Variables
  - Exit Velocity, Launch Angle
  - Hit Coordinates (x, y), Hit Type, and Result



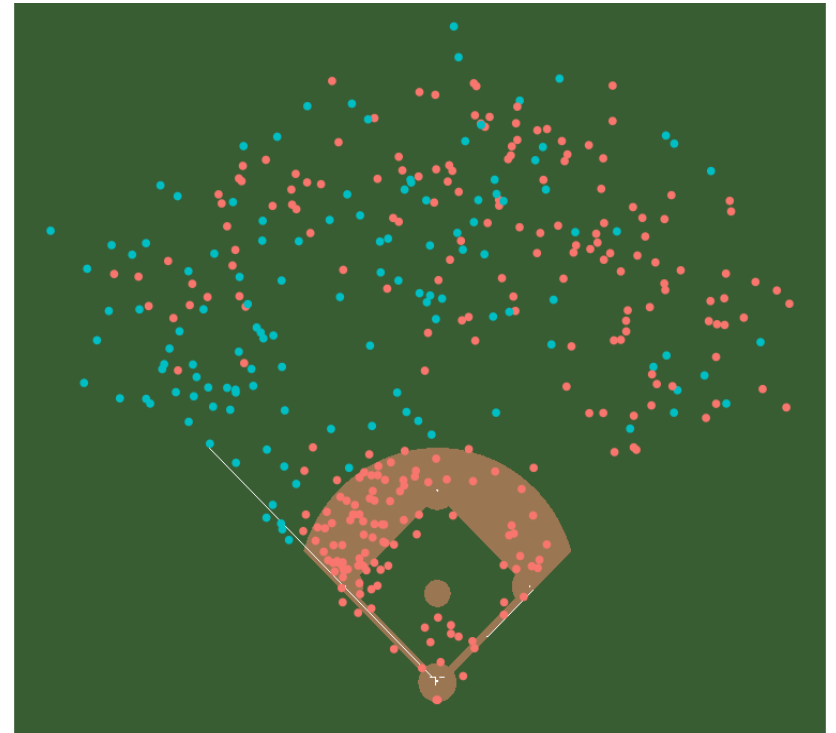
# Bisecting K-Means

- Hit location
- Seven clusters
  - Four for infielders, three for outfielders
  - Centroid



# Random Forest

- Hit probability (xBA)
- Hit location, exit velocity, launch angle
- No clusters



# Examples

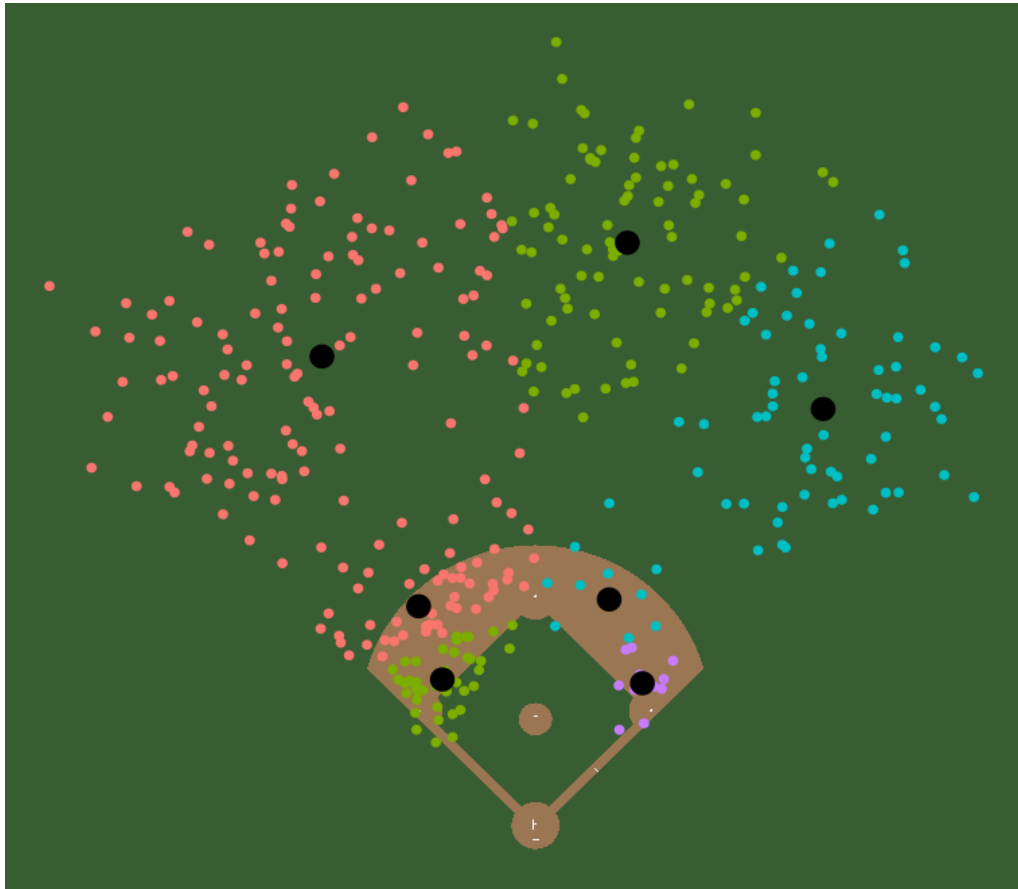


Mike Trout

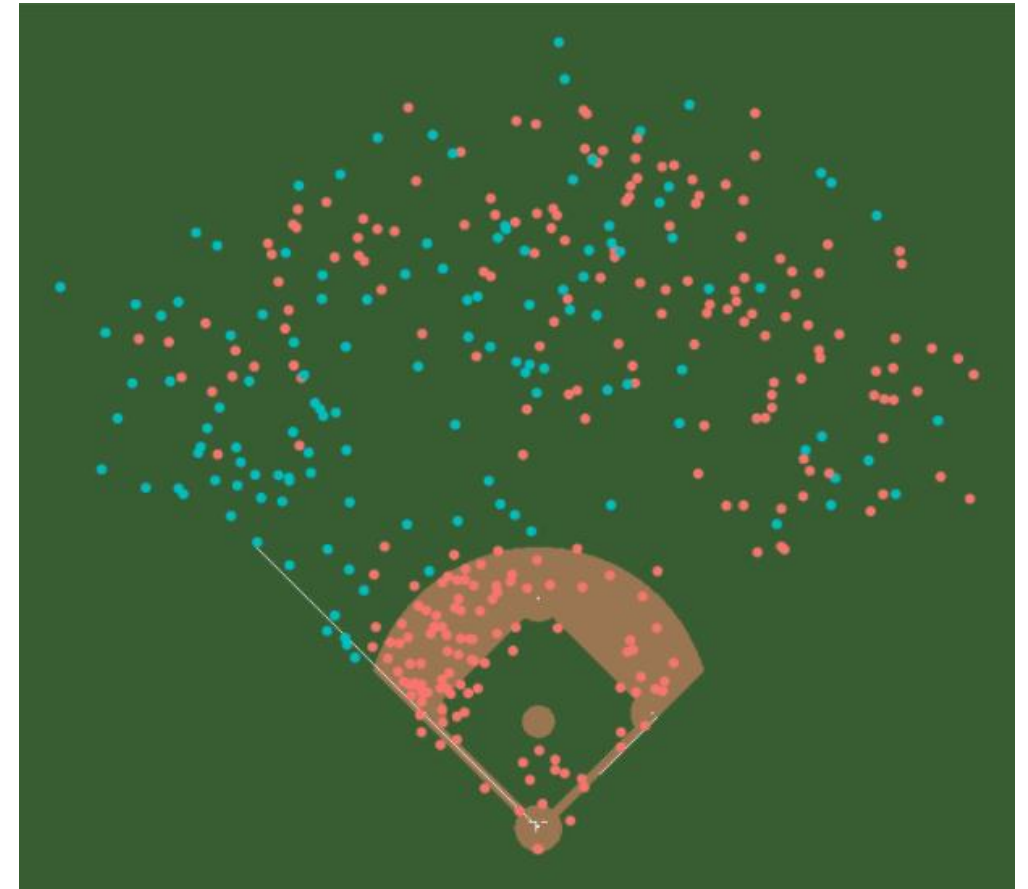


Wilson Ramos

# Mike Trout

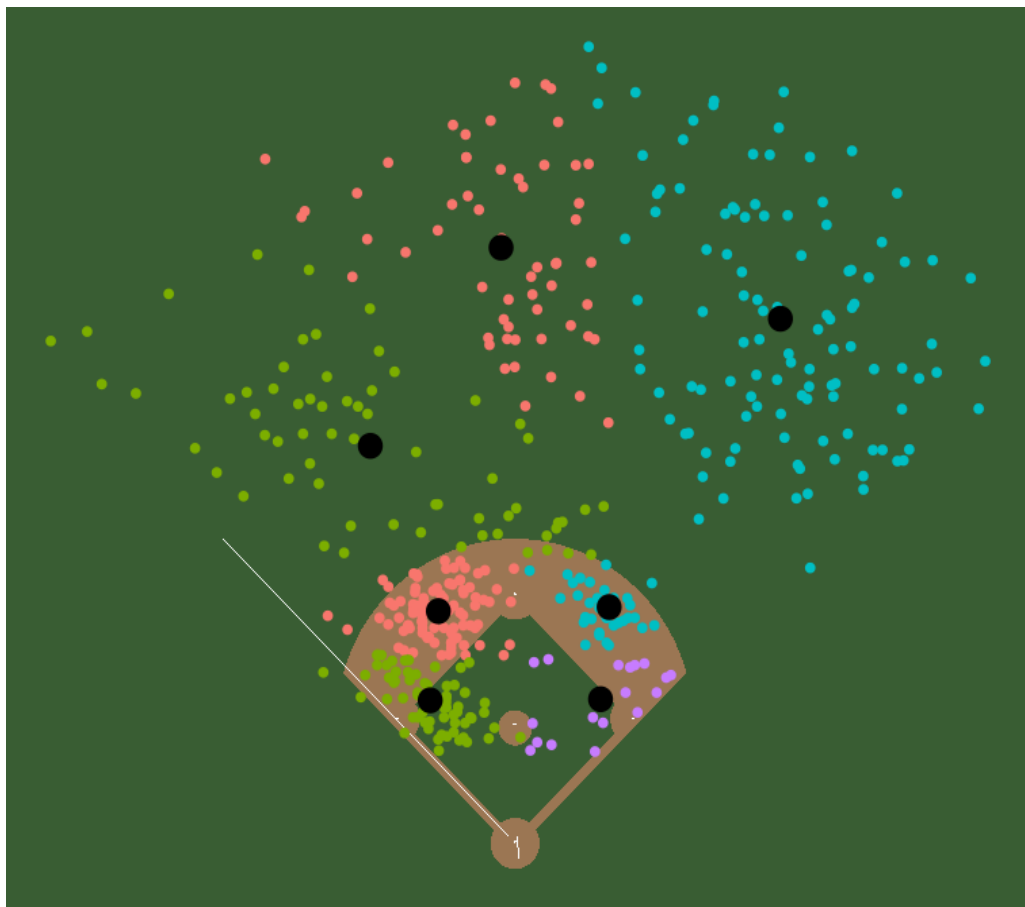


Bisecting K-Means

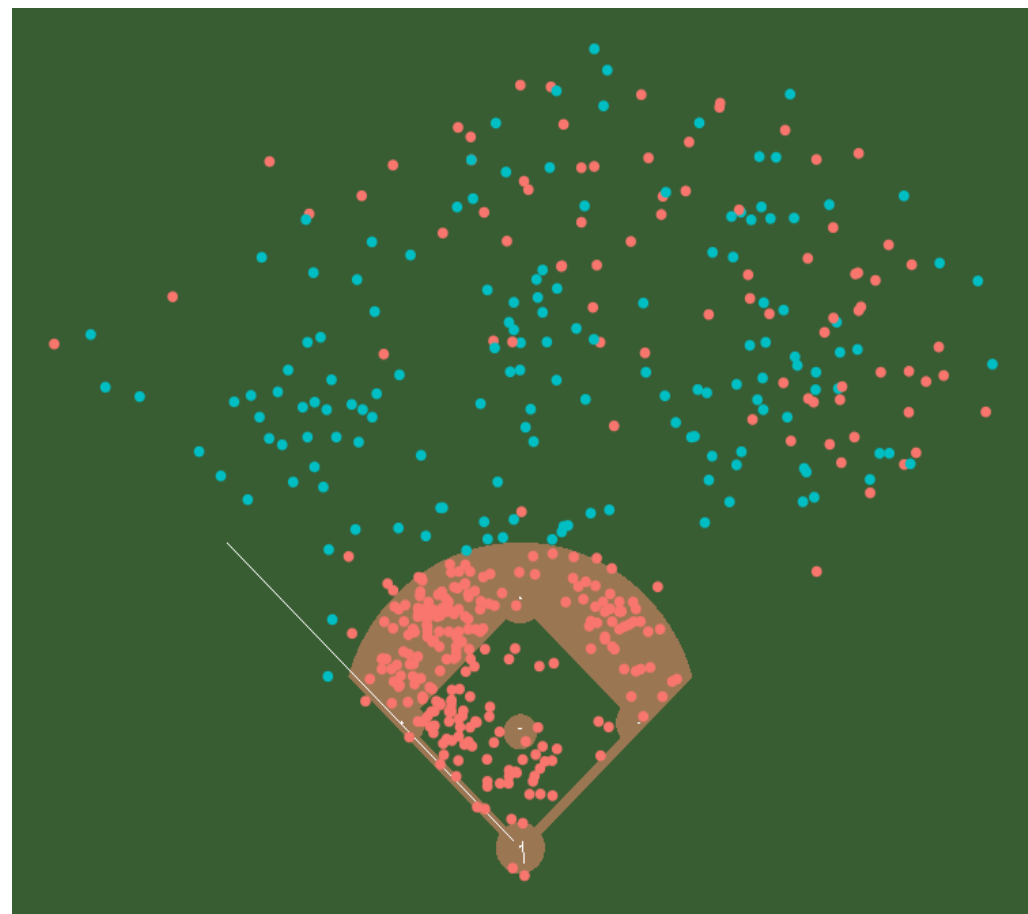


Random Forest

# Wilson Ramos



Bisecting K-Means



Random Forest



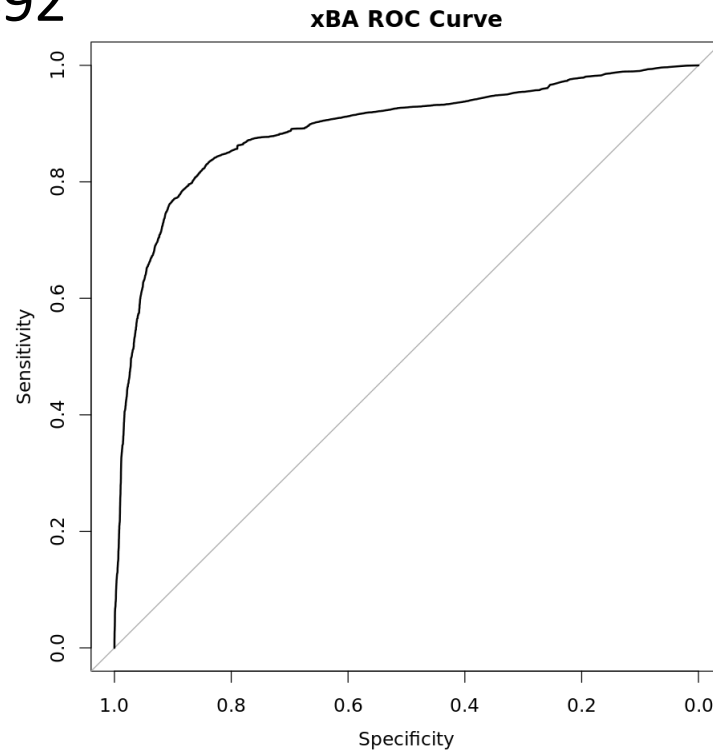
# Evaluation

- Bisecting K-Means: promising

- Random Forest:

Accuracy	Precision	Recall
.859	.789	.763

- ROC Curve: AUC = .8892



# Conclusions

- Very valuable
- Yet to be implemented
- Room for improvement
  - Sprint speed, alternate alignments, etc.



# Bibliography

Ross Drucker (2021). sportyR: Plot Scaled 'ggplot' Representations of Sports Playing Surfaces. R package version 1.0.1.

<https://CRAN.R-project.org/package=sportyR>

Willman, Daren. “Baseball Savant: Trending MLB Players, Statcast and Visualizations.” *Baseball Savant*, MLB Advanced Media, 2 May 2021, [baseballsavant.mlb.com/](https://baseballsavant.mlb.com/).