



# Clustering Swing Behavior Within MLB Hitters

Joseph Dante Maurice & Tai Fowler  
Rice University, Pomona College

## Background

- ❖ MLB and Baseball Savant released new bat-tracking metrics, including: bat speed, swing length, swing path tilt, attack angle, and attack direction.



- ❖ Much exploratory work has focused on identifying different swing archetypes between players.
  - We are interested in swing distributions for **individual players**.
- ❖ Players may have different types of swings.
  - Two-Strike Approach
  - Sac-Fly Opportunity
  - RISP

## Data

- ❖ We have swing data for every pitch of the 2024 MLB season.
  - Removed partial/checked swings<sup>1</sup>
  - Data scraped from Baseball Savant using sabRmetrics package

Sample of Swing Metrics (2024)

Batter	Bat Speed	Swing Length	Swing Path Tilt	Attack Angle	Attack Direction
Altuve, Jose	70.3	8.3	29.9	17.3	-5.3
Judge, Aaron	73.5	8.2	40.0	16.4	-2.2
Ohtani, Shohei	72.6	7.2	31.1	-1.9	1.9

## Acknowledgements

Special thanks to our advisor, Bryant Davis, as well as Quang Nguyen, Dr. Yurko, and our program TA's for their guidance and support.

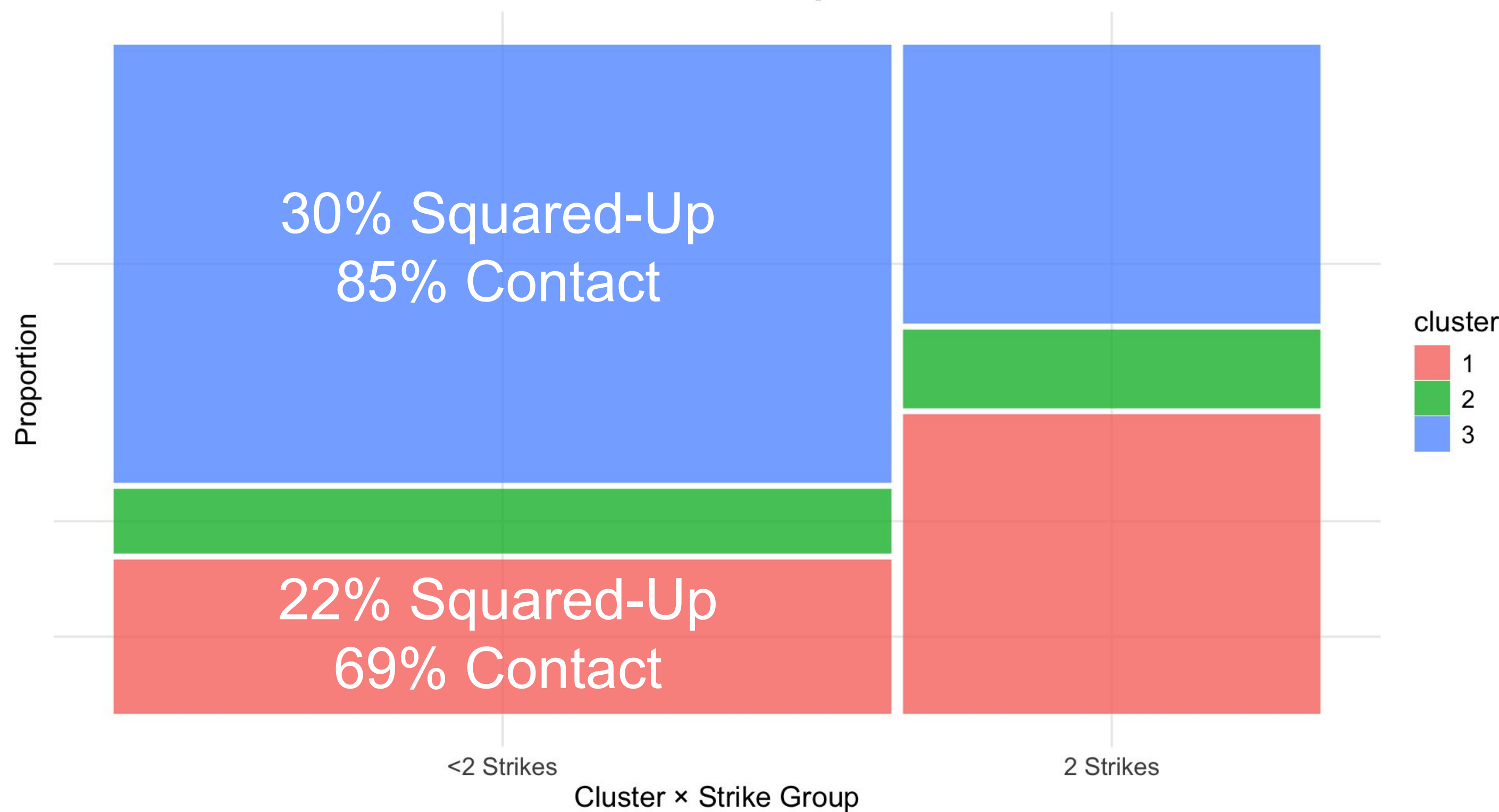
## Methods

- ❖ We used a Gaussian mixture model (GMM) to assign each swing a player took into a **cluster** based on the five new swing-tracking features.
- ❖ We considered reducing dimensionality using Principal Component Analysis (PCA) before clustering but found that the clustering and resulting visualizations were similar.
  - Decided to use the raw features to preserve interpretability.

## Results

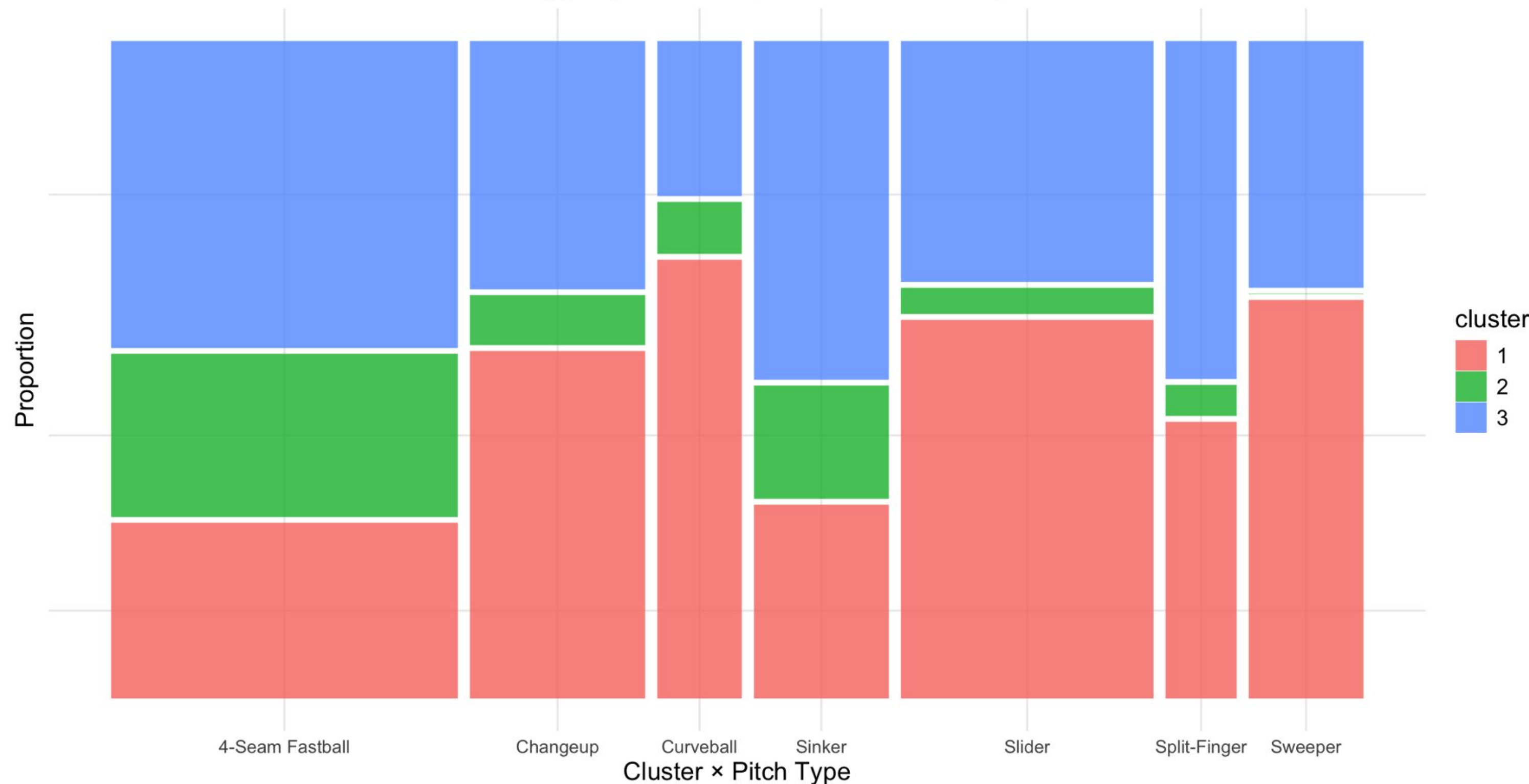
- ❖ After applying GMM clustering to each hitter's swings, a trend emerged: **hitters tend to use "worse" swings in two-strike counts**.
  - These findings challenge the assumption that players adopt a more disciplined, contact-oriented approach with two strikes.

Jose Altuve - Cluster vs. Strike Count Group



- ❖ Importantly, by "worse" swing, we do not imply that hitters are consciously choosing less effective swings.
  - Contextual factors (pitch type, location, etc) appear to induce swings from the lower-performing regions of a player's swing distribution.

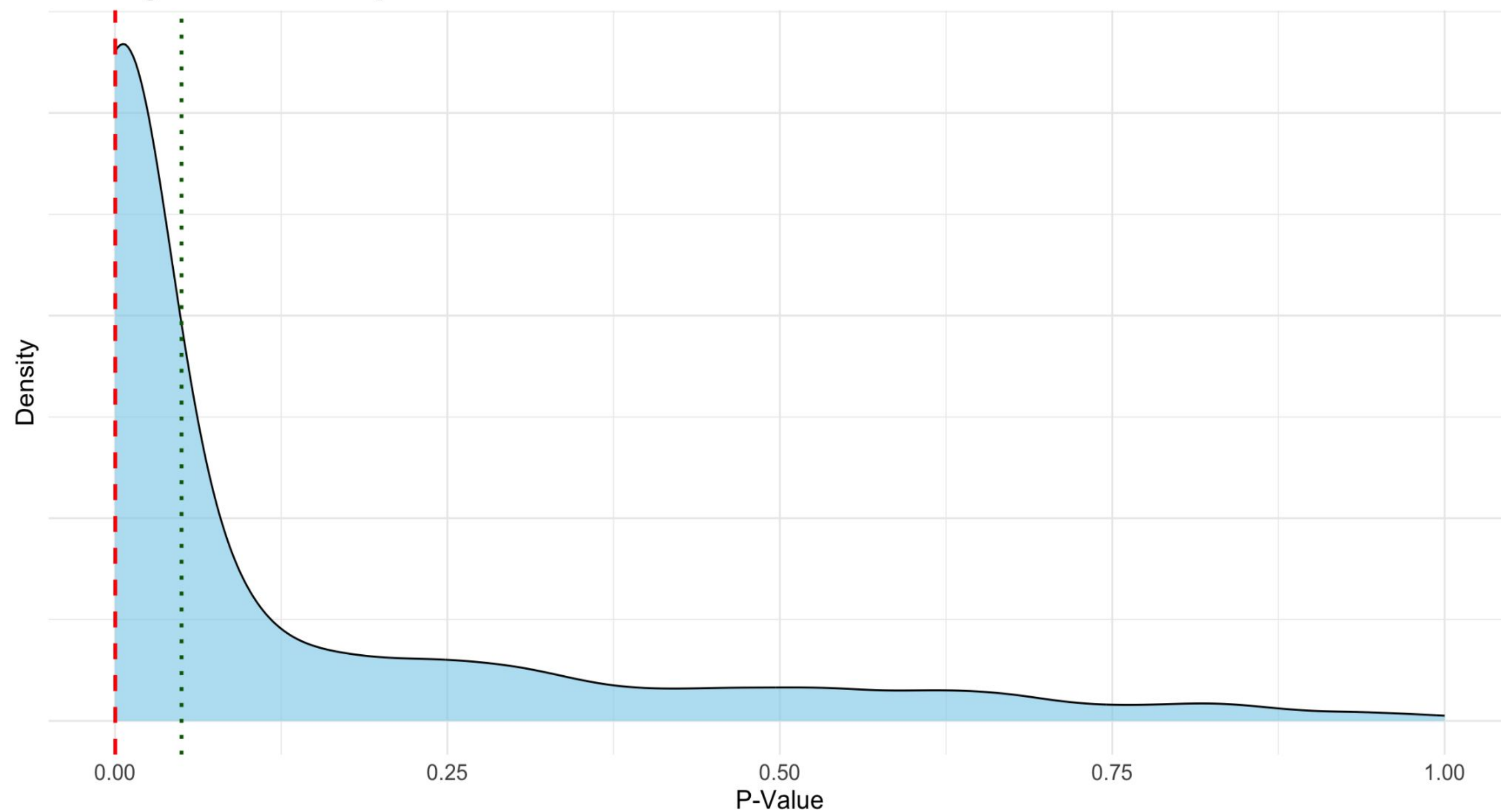
Jose Altuve - Cluster vs. Pitch Type (2 Strikes, Min 20 Pitches)



## More Results

- ❖ Our clustering doesn't suggest that players have several entirely distinct swing mechanics.
  - We've captured **different regions** within each player's overall swing distribution.
  - This allows us to examine when different swing types emerge and begin to explore why these shifts occur.
- ❖ We conducted a 2D Chi-Square Test on cluster and strikes for every hitter with at least 420 swings (median) to generalize our results.

Density Plot of Chi-Square P-Values



- ❖ There's a statistically significant dependence between swing cluster and strikes for 96/307 hitters after accounting for multiple testing.

## Discussions

- ❖ Pitchers and coaches could use these two-strike plots and use them to keep hitters off balance.
- ❖ This framework can be extended to other situations where a hitter's swing may change.
- ❖ In the future, we'd like to develop a mixed effect model to predict contact given swing based on pitcher pitch characteristics and hitter swing characteristics.

## References

1. Powers, S. (2025). sabRmetrics: Query {statsapi,baseballsavant}.mlb.com and fit fundamental sabermetric models. <https://github.com/saberpowers/sabRmetrics/tree/v1.0.4>