

# Take Me Out To The Ball Game

## An Analysis of the Home Run Surge in Major League Baseball

Leocadie Haguma, Rosalia Hernandez, Jacob Rojas

RUSIS @ Oregon State University

August 23, 2018



# Table of Contents

**1** History

**2** Previous Work

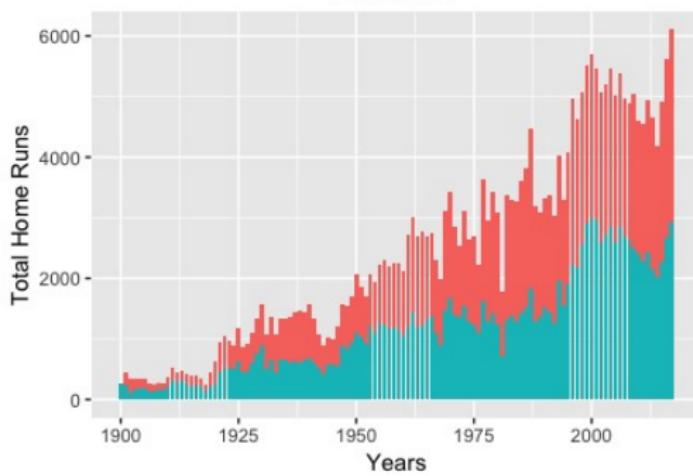
**3** Our Work

**4** Conclusions



# History

## Home Runs Over the Years Since 1900



## Historical Events (Getz 2009) Notable Rule Changes :

- 1920's : Babe Ruth
- 1945 - 1950 : World War II
- 1969 : Pitcher's Mound Lowered 5 inches
- 1973 : Pitcher's designated batter
- 1981 : Strike Canceled 713 Games
- 1994 : The "Great Strike"
- 2000's : Steroids
- 2015 : ?

## Our Problem

There is an increasing trend of home runs within Major League Baseball between the 2015 to 2017 seasons. We plan to explore why and how this trend occurred.



Oregon State  
University

# Speculation



**Justin Verlander**   
@JustinVerlander



All I'm saying is I don't care if balls are juiced (seriously). We're all using the same ball so it's a fair field. My issue is I don't like being lied to. I knew something was different. Century old records are being broken and numbers are skewed.

7:07 PM - Mar 1, 2018



14.8K



2,891 people are talking about this



Oregon State  
University

# Official Investigation

Report of the Committee Studying Home Run Rates in Major League Baseball (Albert, 2018)

The Scientists (Albert et. al 2018) :

- Jim Albert, Professor of Statistics at Bowling Green State University
- Jay Bartroff, Professor of Mathematics and the Graduate Vice-Chair for Statistics at USC
- Roger Blandford, Luke Blossom Professor in the School of Humanities at Standford University
- Dan Brooks, Researcher and Co-Organizer of Saberseminar
- Josh Derenski, Statistics PhD Student in the USC Marshall School of Business
- Larry Goldstein, Professor of Mathematics at the University of Southern California
- Anette (Peko) Hosoi, Neil and Jane Pappalardo Professor of Mechanical Engineering and Mathematics at MIT
- Gary Lorden, Professor Emeritus of Mathematics at Caltech
- Alan Nathan (Chair), Professor Emeritus of Physics at the University of Illinois at Urbana-Champaign
- Lloyd Smith, Professor in the School of Mechanical and Materials Engineering at Washington State University



Oregon State  
University

# Terms

## Launch Angle

The angle from when the ball hits the bat and flies into the outfield.

## Exit Velocity

The initial speed of the ball immediately after being hit by the bat.

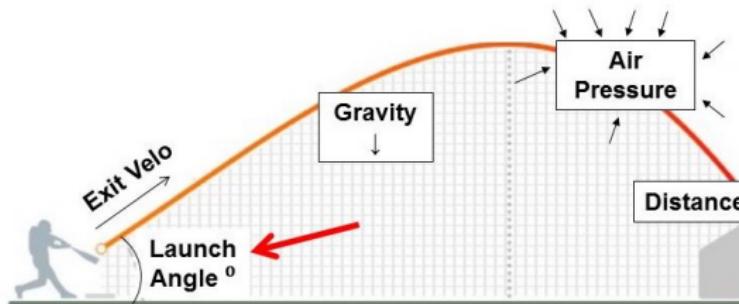


FIGURE – Image by Daren Willman, Director of R&D at MLB

## Red Zone

The majority of home runs are within the parameters of an exit velocity between 90 to 115 mph, and 15 to 45 degree launch angle.

# Official Investigation

## Their initial thoughts about home runs

- Concentration of home runs among players
- Concentration of home runs among ballparks
- Changes in pitching strategy
- Changes in batting strategy
- Characteristics of home runs

## Committee Findings

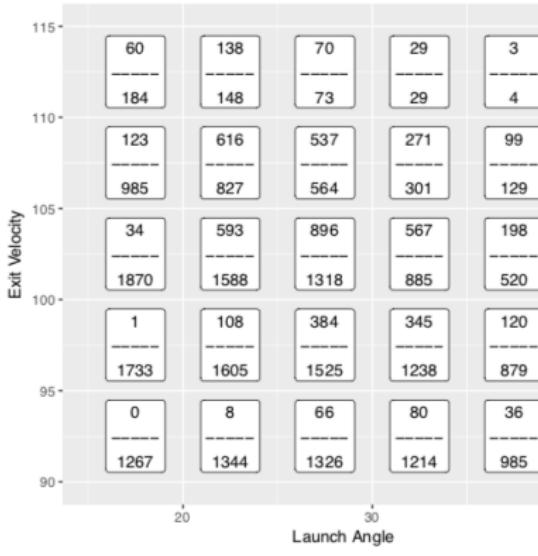
- The probability of hitting a ball into the "Red Zone" did not change.
- The probability of hitting a ball into the "Red Zone" and making a home run did change.



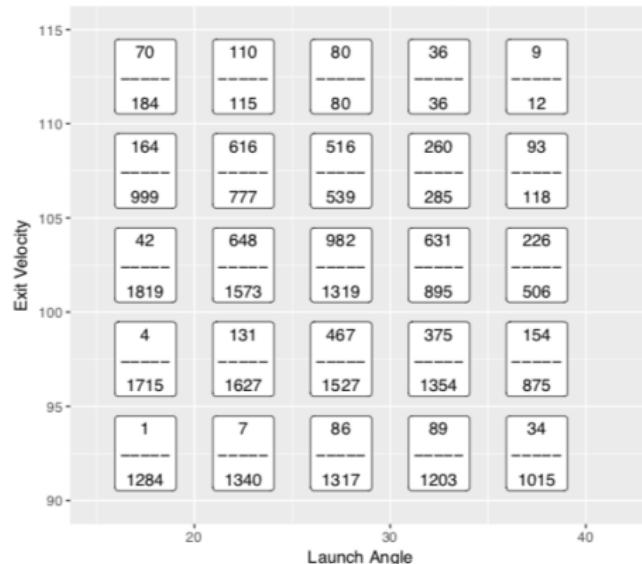
Oregon State  
University

# The Red Zone

HR / N for Period y16



HR / N for Period y17



Images by The Committee of Scientists studying the home run surge in the MLB



# Official Committee

## Exploring the factors that affect the carry of a baseball

### Temperature :

- Examined home run rates against temperature for open stadiums and for closed stadiums.
- They found similar trends through the seasons examined (2015-2017).
- Conclusion : The home run surge could not be explained by the effect of temperature.

### The Ball :

- Examining the manufacturing process of MLB baseballs required a trip to Rawlings Sporting Goods, the official MLB baseball manufacturer, located in Costa Rica.
- Rawlings Sporting Goods measure the baseball's weight, circumference, compression, coefficient of restitution(COR).
- Need more information



Oregon State  
University

# Official Committee

## More testing

- Sports Science Laboratory at Washington State University
  - UMass Lowell Baseball Research Center 2013-2017
  - MLB, authenticated, 2012-2017
  - Impact testing, aerodynamic testing
  - Other Properties : seam heights, center of gravity, surface roughness
- Analysis of StatCast Trajectories

## Conclusions

The baseball has shown changes in its aerodynamic properties that are not currently being measured or recorded. Since the baseballs are being made within the parameters set by the MLB, there is not evidence that supports the claim that the baseball has physically changed (Albert, 2018).

## Some Recommendation

- Implement new parameters to measure and monitor the characteristics of the baseballs that affect carry.
- Monitor the baseball storage climate.



# Fivethiryeight.com

## The Researchers

- A data journalism blog created by Nate Silver who is an American statistician.
- Long history of predictive analytics successes in sports and politics.

## What did the researchers do ?

- Used a Computed Tomography (CT) scan to X-ray baseballs
  - This looked at the density of the rubber lining of the ball
- Used a Thermogravimetric testing for a Chemical Composition of the ball.
  - This looked at identifying the chemical changes in the rubber lining



Oregon State  
University

# Findings

## Discoveries

- The baseballs used by the MLB prior to 2015 were more dense and weighed more than the ones being used now.

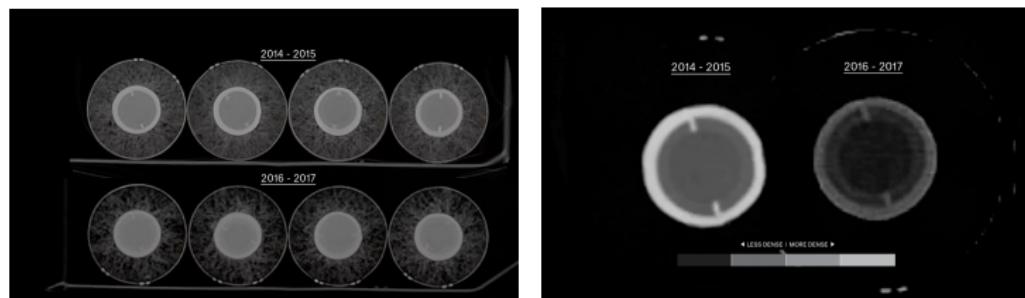


Image by [fivethirtyeight.com](http://fivethirtyeight.com)



Oregon State  
University

# Fivethiryeight.com Conclusions

## Findings

- The chemical composition of the core of the baseball changed
- The baseball was "bouncier"
- The rubber portion of the core was less dense
- The entire baseball was now 0.5 grams lighter

## Conclusion

- These changes to the ball could only account for 25% increase in home runs
- The surge was a 46% increase home runs



Oregon State  
University

## Our Story

We believe the most unexplored area of the surge is the players themselves.



Oregon State  
University

# Difficulty in Collecting Data

## Initial data collection

- Initial data gathering conducted on MLB stats website. Interface built for viewing, not scraping.
- Gathering data through MLB proved to be inefficient.



# Data Collection

## Baseball Savant

- Statcast measurement
- Pitchfx data (Release speed, Type of pitch, etc.)
- Hitfx data (Exit velocity, Launch angle, etc. )

## Lahman Database

- General MLB information
- Age, Height, weight
- Teams, Leagues (NL and AL)



# Data Organization

## Joining Databases

- Statcast pitching data back to 2004. Over 8 million total pitches.
- Lahman data of player information dating back to beginning of baseball.
- Joining of databases to have complete picture. Using unique batter IDs and names in combination with *Baseballr* player lookup function, databases could be joined.



Oregon State  
University

## Good Data

*"Data is good, only if it is good data"* – Dr. Javier Rojo



Oregon State  
University



# Bryce Harper's Home run



## Players

Our investigation focused on how the players themselves were affecting the increase in home runs. We decided to look at the following variables for this :

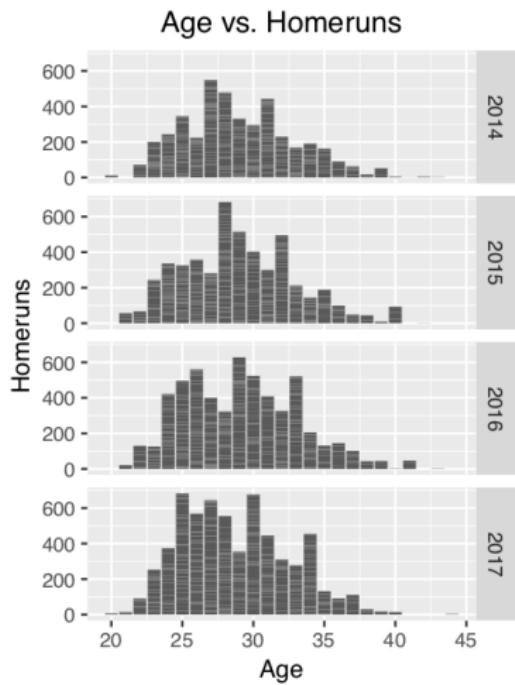
- Age
- Height and Weight
- Pitchers
- New MLB players (Rookies)



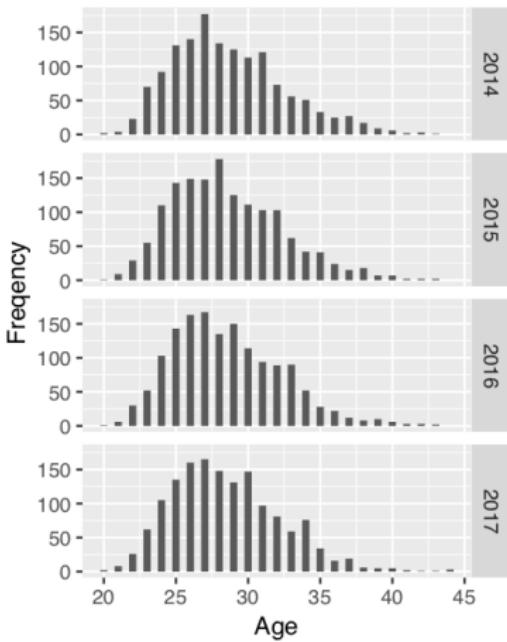
Oregon State  
University

# Age

Age Distributions per Season



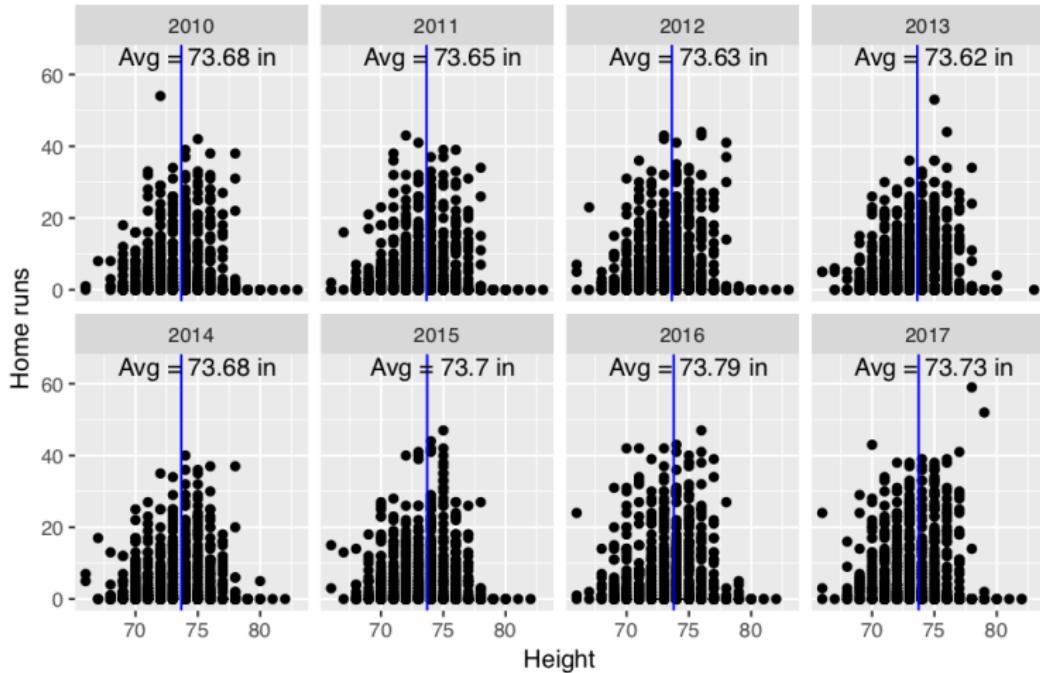
Age Frequency



Oregon State  
University

# Heights of All MLB Players

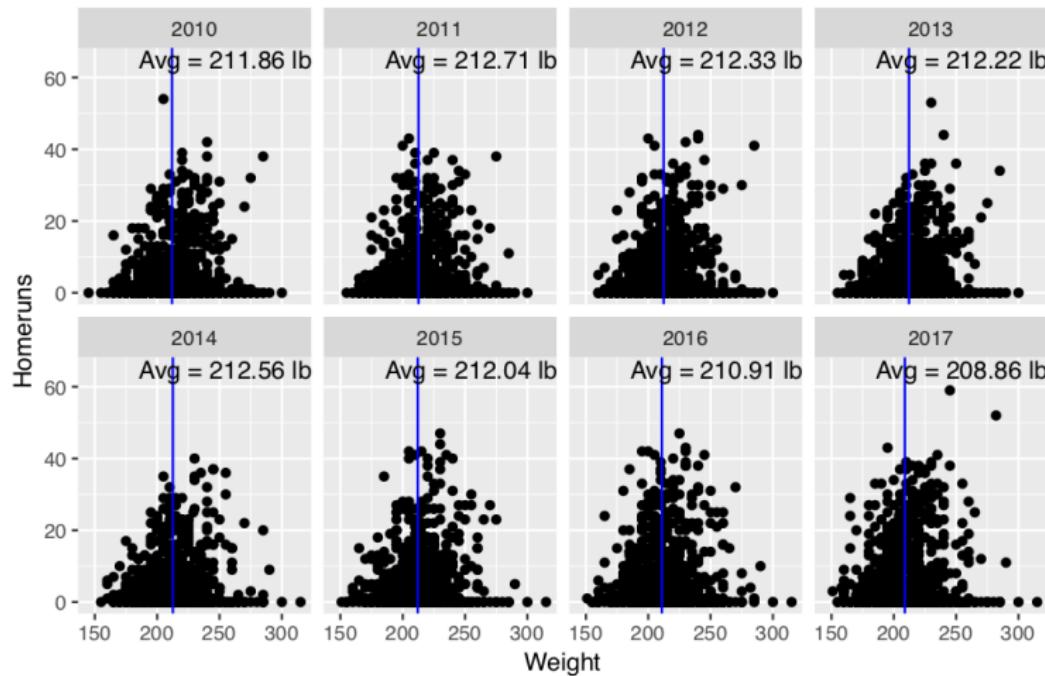
Home runs by Height



Oregon State  
University

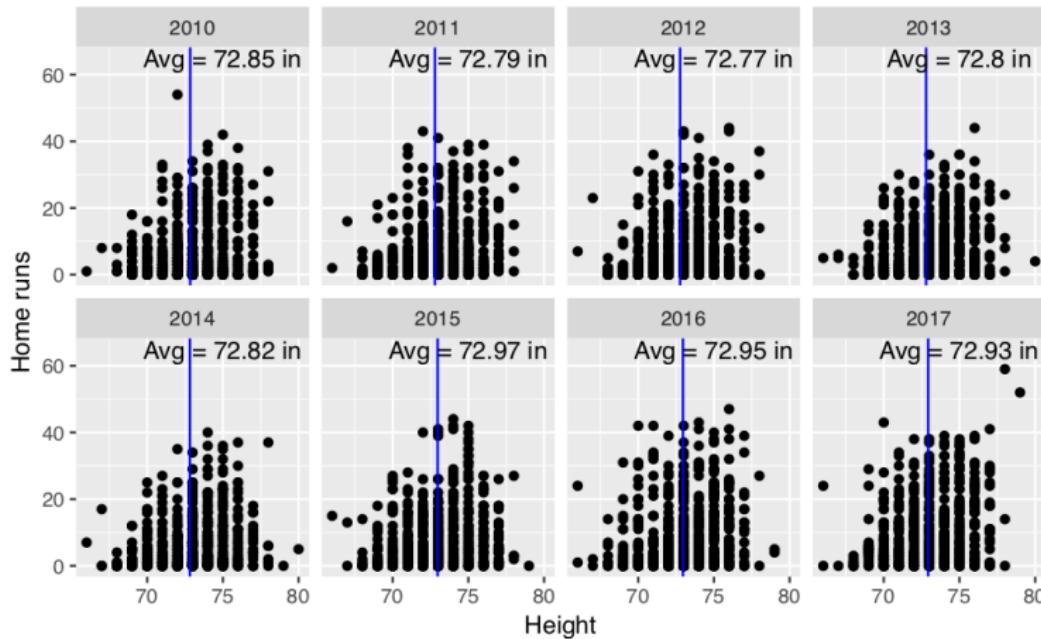
# Weights of All MLB Player

Home runs by Weight



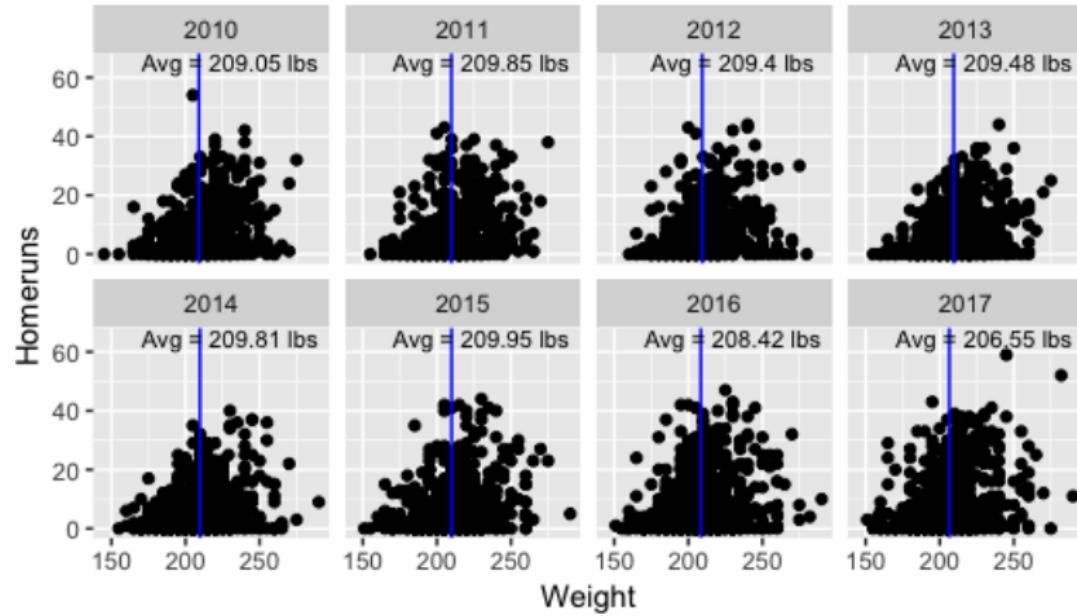
# Heights of MLB Batters

Home runs by Height  
without pitchers



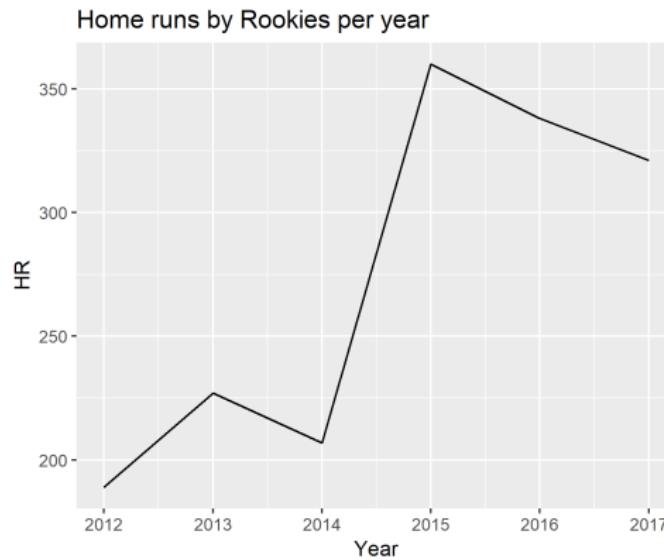
# Weights of MLB Batters

Home runs by Weight  
without pitchers



# All New MLB Players (Rookies)

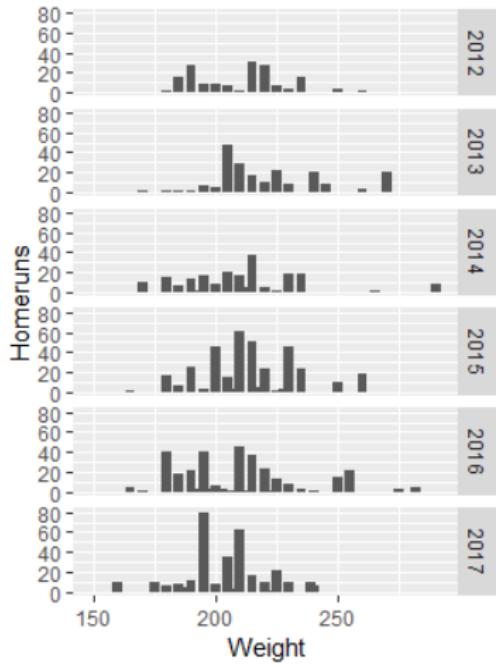
Is the incoming class of MLB hitters affect by the home run surge ?



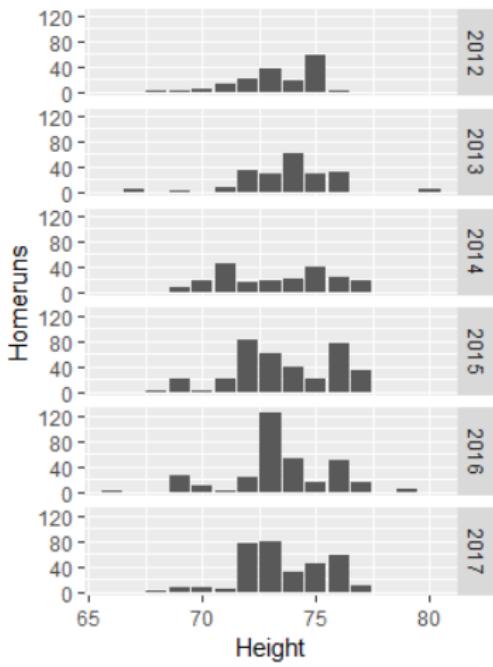
# New MLB Players (Rookies) cont...

Rookie Distributions

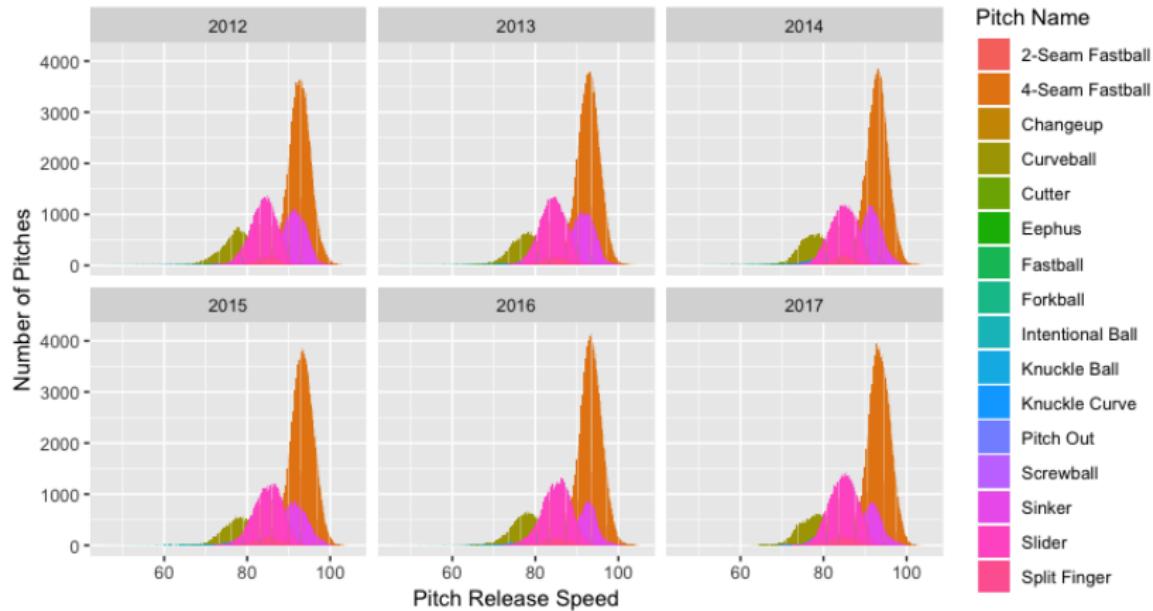
Home runs vs. Weight



Home runs vs. Height



# Pitchers



Oregon State  
University

# Logistic Regression

## Logistic Regression

- A logistic regression allows us to analyze the impact of a group of variables on a response variable (Home run = 0,1).

## Variables :

- Release Speed
- Launch Speed
- Weight
- Height
- Stand
- Age
- Pitch throw
- Interaction of height and weight
- Multiple logistic regressions : Pre/Post 2015, Post 2015 with launch speed



Oregon State  
University

# Significance

## Pre-2015 Logistic Regression

	Estimate	Std. Error	z value	Pr(> z )	Significance
Release Speed	-0.0063463	0.0012061	-5.262	<<0.001	***
Weight	0.0094586	0.0005333	17.737	<<0.001	***
Height	0.0500789	0.0049122	10.195	<<0.001	***
Age	0.0116417	0.0025300	4.601	<<0.001	***
Stand_R	0.0908204	0.0147309	6.165	<<0.001	***
Pitcher_R	0.0441629	0.0162320	2.721	0.00651	**

## Significance codes

■ 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '' 1



Oregon State  
University

# Significance cont...

## Post-2015 Logistic Regression

	Estimate	Std. Error	z value	Pr(> z )	Significance
Release Speed	-0.0058135	0.0013788	-4.216	<<0.001	***
Weight	0.0090255	0.0005521	16.347	<<0.001	***
Height	0.0403999	0.0053456	7.558	<<0.001	***
Age	-0.0038989	0.0022585	-1.726	0.0843	.
Stand_R	0.0703056	0.0167903	4.187	<<0.001	***
Pitcher_R	0.0792483	0.0189381	4.185	<<0.001	***

## Significance codes

- 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '' 1



# Significance cont...

Post-2015 Logistic Regression With Added Variables					
	Estimate	Std. Error	z value	Pr(> z )	Significance
Release Speed	-3.763e-02	1.518e-03	-24.789	«<0.001	***
Launch Speed	1.960e-01	1.485e-03	131.957	«<0.001	***
Weight	1.633e-02	1.520e-02	1.075	0.282563	
Height	4.729e-02	4.350e-02	1.087	0.277006	
Age	1.711e-02	2.452e-03	6.979	«0.001	***
Stand_R	-6.822e-02	1.823e-02	-3.741	«0.001	***
Pitcher_R	8.883e-02	2.056e-02	4.320	«0.001	***
Weight :Height	-2.417e-04	2.070e-04	-1.168	0.242937	

## Significance codes

- 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



Oregon State  
University

# Conclusions

## Summary

- History of home runs in MLB tell us a story
- Widespread speculations
- Previous work
  - Official MLB committee : changes in drag coefficient
  - fivethirtyeight : changes in chemical composition
- Our work revealed that the change in players had no effect on the surge in home runs.
- Our logistic regression emphasized that the surge affected all players and that launch speed is a better proxy for strength.

## Further Analysis

- Further investigation to validate the change in significance of the handedness of the pitcher could provide more information on the home run surge.



Oregon State  
University

# Acknowledgment

This work was supported through NSA grant number H98230-18-1-0009 and NSF grant DMS-1731082, to Dr. Javier Rojo PI, through the RUSIS @ Oregon State University.



Oregon State  
University