

# "Statcast – the New Metrics in Baseball"

Jim Albert  
Bowling Green State University

April 7, 2018

# My background



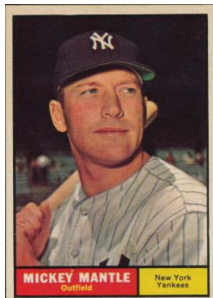
- Statistician by training



- Phillies fan

# Growing Up in Philly

- Played Little League baseball (no soccer)



- Collected baseball cards
- Played baseball games

# Strat-O-Matic Baseball



# Enjoyed Reading Bill James




# Bill James *Baseball Abstracts*

- Start with a good baseball question
- Collect relevant data
- Explore
- What has been learned?

# Fast Forward to 2018



- : Open-source statistics programming language
- Statcast and PitchFX: New sources of data
- Great way to learn data science

# PitchFX

- Starting in 2006, cameras were installed in each stadium to track pitches
- Many variables for each pitch are recorded
- Pitch speed, pitch movement, pitch location
- Outcome (called ball or strike, ball in play, etc)



# PitchFX data

- All of this data is freely downloadable
- Relatively large dataset
- For June of a recent season, I have a data frame of 120,000 pitches and 49 variables.

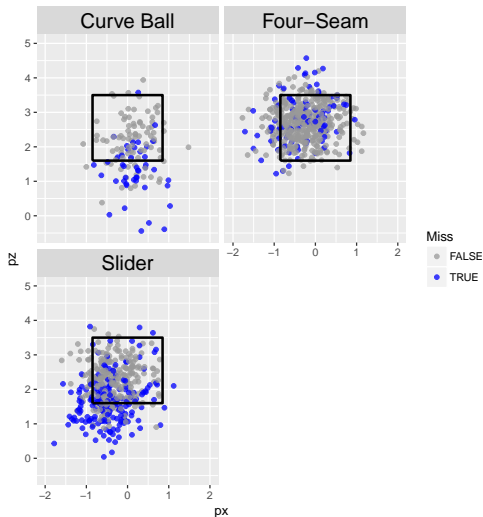
# Clayton Kershaw



# Types of Pitches

- Fastball (two-seam, four-seam, etc.)
- Off-speed (curve ball, slider, changeup)
- A good pitcher typically has at least 2-3 good pitches

# Clayton Kershaw - Locations and Outcomes of 3 Pitches



## Wikipedia:

- Major League Baseball's (MLB) Statcast is a high-speed, high-accuracy, automated tool developed to analyze player movements and athletic abilities.
- Statcast was introduced to all thirty MLB stadiums in 2015.
- Some of this data is publicly available

# What Does StatCast Measure?

- Pitch information (like PitchFX)
- Movements of every player on the field
- Measurements off of the bat (exit velocity, launch angle, and horizontal angle)

# Measuring Fielding

- In the old days, one measured fielding by the fielding percentage

$$PCT = \frac{PO + A}{PO + A + E}$$

- Some players are regarded as great fielders since they have a high fielding percentage
- Derek Jeter received 5 Golden Glove Awards

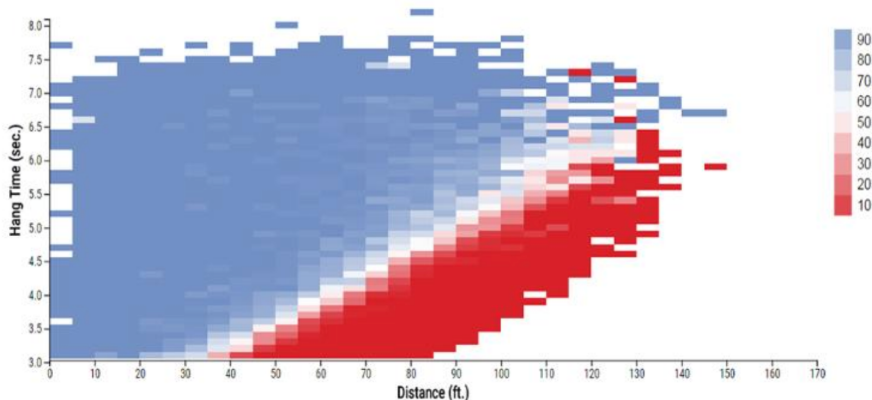
# StatCast Has Changed Things

- Making a outfield catch depends on (1) distance to the location and (2) time the ball is in the air (opportunity)
- Can estimate probability of making a catch
- A great catch is one where the probability of catch is small



# Catch Probability

## Outfield Catch Probability (Based on Distance Needed / Opportunity Time)



# Hitting

- Batter's objective is to get a base hit
- Different types of batted balls (liner, ground ball, fly ball, and pop up)
- Some of these types are more desirable
- Want to hit the ball hard
- Location of batted ball important

# Definition of Batted Ball Type

## Launch Angle (LA)

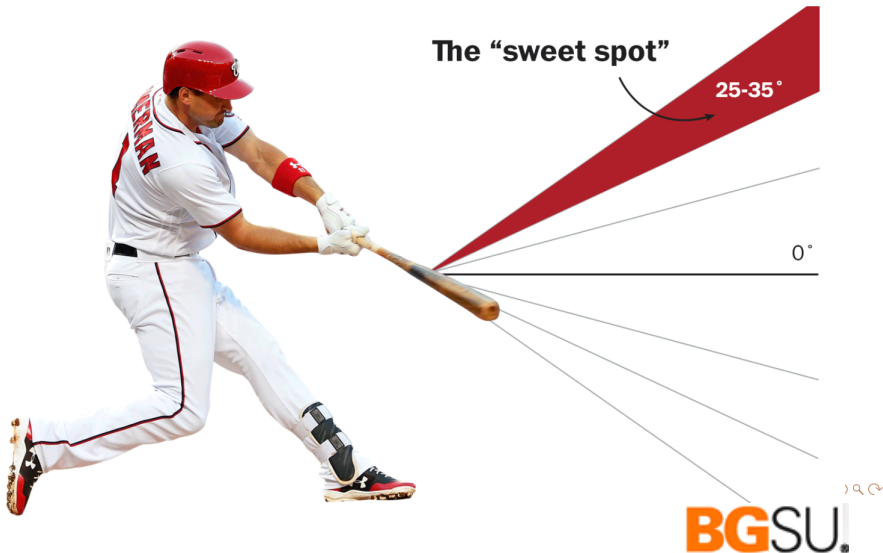
### Definition

Launch Angle represents the vertical angle at which the ball leaves a player's bat. Launch Angle (aLA) is calculated by dividing the sum of all Launch Angles by

As a guideline, here are the Launch Angles for different types of contact:

- Ground ball: Less than 10 degrees
- Line drive: 10-25 degrees
- Fly ball: 25-50 degrees
- Pop up: Greater than 50 degrees

# Launch Angle



# Focus on Hits on Balls in Play

- Hit the ball in the air (positive launch angle)
- But not too high (popup)

## Focus on Question

For a given exit velocity, what launch angle maximizes the chance of a base hit?

# Start with 2017 Statcast Batting Data

- 735,817 pitches
- Most pitches are not put “in-play”
- 79 variables collected for each pitch

# Filter

- Only consider pitches where the ball is put in play
- Exclude home runs
- Only consider batted balls where launch angle is positive



# Fix an Exit Velocity

- Optimal launch angle depends on the exit velocity
- Only consider batted balls close to that velocity

# Fix Exit Velocity at 90 mpg

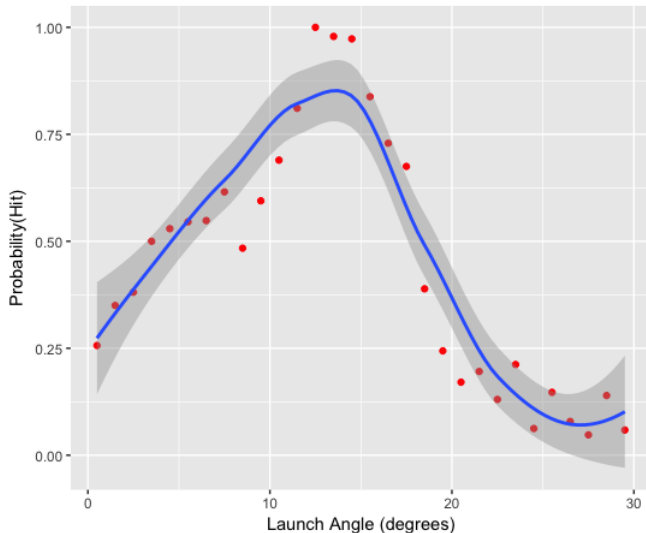
- Bin the launch angles in the groups  $(0, 1)$ ,  $(1, 2)$ ,  $(2, 3)$ ,  
...
- Find the number of batted balls and hits in each bin
- Compute the probability of hit in each bin

# Binned Results

	LA	N	H	Probability
	<fct>	<int>	<dbl>	<dbl>
1	(0,1]	39	10.	0.256
2	(1,2]	40	14.	0.350
3	(2,3]	42	16.	0.381
4	(3,4]	38	19.	0.500
5	(4,5]	34	18.	0.529
6	(5,6]	44	24.	0.545
7	(6,7]	31	17.	0.548
8	(7,8]	39	24.	0.615
9	(8,9]	31	15.	0.484
10	(9,10]	37	22.	0.595

# Plot Launch Angle Against Hit Probability

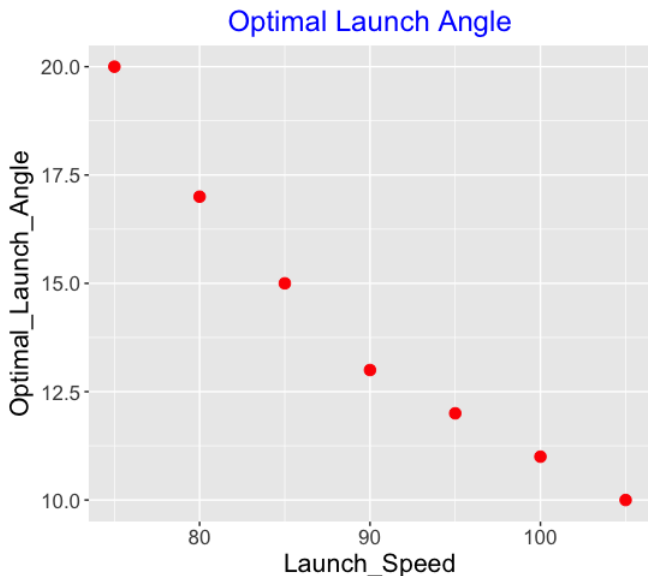
## Optimal Launch Angle?



# What Have We Learned?

- When balls are hit at 90 mpg, probability of hit is maximized at a launch angle of 14 degrees
- But this optimal launch angle depends on the speed of batted ball

# Dependence of optimal launch angle on speed of batted ball



# This Week's Blog Post

## Exploring Baseball Data with R

[Home](#)[Authors](#)[Analyzing Baseball Data with R](#)[R-bloggers](#)

### In Search of the Optimal Launch Angle


By [Jim Albert](#) on March 31, 2018 | [Edit](#)

### Carnegie-Mellon Workshop on April 7

I am excited about participating in the first [CMU Baseball Analytics Workshop](#) next Saturday. This workshop will provide an opportunity for students to learn about baseball analytics and do some R programming on baseball data. The participants will be visiting **BGSU** and meeting with Dan Fox, director of Baseball Informatics for the Pirates. (If you are in the area,

# The R script on GitHubGist

**GitHubGist**  [All gists](#)

 **bayesball / optimal\_angle.R**  
Created 14 hours ago

[Code](#) [Revisions 1](#)

R code to find optimal launch angles from Statcast data

 **optimal\_angle.R**

```
1 library(tidyverse)
2
3 # read in the Statcast data for the 2017 season
4
5 sc <- read_csv("statcast2017.csv")
6
7 # only consider balls put in play
```



# Statcast has changed hitting

- Player accounts suggest Statcast data has replaced traditional metrics.
- For example, on the first day of spring training, Tampa Bay Rays hitters are told they will be measured by batted-ball exit velocity, not batting average.
- Kris Bryant credits his improved performance in 2016 with changes he made in the off-season to adjust the launch angle of his hits.

# Analytics and Sports

- More teams are building analytics departments
- Need expertise in all aspects of data science
- Yes, there are jobs in baseball and other sports

# Looking Ahead

- Teams are still learning from the Statcast data
- What makes an effective base stealer?
- What variables are important besides speed?
- How you measure these "other" variables?

# The Student

- Baseball is a great way of learning data science
- Tons of good data
- Pose a question and explore the relevant data to answer
- Use graphs to effectively communicate findings