

# Final Project Proposal & Dataset

**Tom Weichle**

August 21, 2018

# Data Set

- BaseballSavant on [MLB.com](https://www.mlb.com/savant)
  - A site dedicated to providing player matchups, **Statcast** metrics, and advanced statistics in a simple and easy-to-view way
- Statcast:
  - A state-of-the-art tracking technology (radar and HD camera system), capable of measuring previously unquantifiable aspects of the game
  - New in 2015 to all 30 MLB ballparks
  - Statcast database contains files including measurements and metrics from each game throughout the baseball season

# Outcomes for Supervised Learning (Predictive Modeling)

- Home Run (HR)
  - Among the most important events in games; great statistic for evaluating a hitter's power
- Extra-base Hit (XBH)
  - Any hit that is not a single, meaning doubles, triples and home runs; good statistic to evaluate an offensive player's power -- and in some cases, his speed
- Single (1B)
  - Most common type of hit in baseball

Note: Outcomes will be predicted using Classification because they are categorical measures.

# Predictors

- Exit Velocity
  - How fast, in miles per hour, a ball was hit by a batter
- Launch Angle
  - How high, in degrees, a ball was hit by a batter
- Pitch Velocity
  - How hard, in miles per hour, a pitch is thrown
- Pitch Type
  - What Type of pitch thrown to defeat a batter (e.g., fastball, curveball, changeup, slider)



*Questions?*