TruMedia 2017 MLB Hackathon

Project Setup

- Inside the zip file you've been provided MLB pitch by pitch data for 2014 through 2016.
- The data has been provided in two formats: JSON and CSV.
- There is one file per format per season. As these files are extremely large, there have been sample files for the 2016 World Series also provided. We recommend that you explore these files first. Once you've determined what data you're interested in, we recommend that you transform and prune the full season files to better suite your application or research needs.

Data Format

- The data contains one row per pitch.
- Most columns are self-explanatory. New columns for this year's hackathon are: probCalledStrike, umpireId, and umpire. Below are the columns that need additional documentation:
    - releaseVelocity - Pitch velocity in MPH
    - spinRate - pitch spin rate in RPM
    - spinDir - from the catcher's perspective, the angle (from 0 to 360) between the the pole around which the ball is rotating and the positive x-axis
    - px - distance in feet from the horizontal center of the plate as the ball crosses the front plane of the plate, negative values are inside to RHB.
    - pz - height in feet above the ground as the ball crosses the front plane of the plate
    - szt - height in feet above the ground of the rulebook top of the strike zone for the batter
    - szb - height in feet above the ground of the rulebook bottom of the strike zone for the batter
    - x0 - distance in feet from the center of the plate at the first measurement (50 ft from home plate), negative values are inside to RHB.
    - y0 - distance in feet from the tip of home plate at the first measurement (50 ft from home plate)
    - z0 - height in feet off the ground at the first measurement (50 ft from home plate)
    - vx0 - velocity in FPS along the X axis of the field at the first measurement
    - vy0 - velocity in FPS along the Y axis of the field at the first measurement, negative values are heading towards home plate
    - vz0 - velocity in FPS along the Z axis of the field at the first measurement
    - ax - acceleration of the ball along the X axis of the field, estimated as a constant for the flight of the pitch
    - ay - acceleration of the ball along the Y axis of the field, estimated as a constant for the flight of the pitch
    - az - acceleration of the ball along the Z axis of the field, estimated as a constant for the flight of the pitch

- battedBallAngle - angle in degrees of the batted ball, -45 is down the left field line, 45 is down the right field line
- battedBallDistance - distance in feet from home plate to where the ball was fielded
- pitchResult
    - SS - Swinging Strike
    - SL - Strike Looking
    - F – Foul
    - FT - Foul Tip
    - FB - Foul Bunt
    - MB - Missed Bunt
    - B – Ball
    - BID - Ball in Dirt
    - HBP - Hit By Pitch
    - IB - Interntional Ball
    - PO - Pitch Out
    - IP - Ball in Play
    - AS - Automatic Strike
    - AB - Automatic Ball
    - CI - Catcher Interference
    - UK – Unknown
- pitchType
    - CH – Changeup
    - CU – Curveball
    - FA – Fastball
    - FT - Two Seamer
    - FF - Four Seamer
    - FC – Cutter
    - SL – Slider
    - FS – Splitter
    - SI – Sinker
    - FO – Forkball
    - KN – Knuckleball
    - KC - Knuckle Curve
    - SC – Screwball
    - GY – Gyroball
    - EP – Eephus
    - PO – Pitchout
    - IN - Intentional Ball
    - AB - Automatic Ball
    - AS - Automatic Strike
    - UN – Unknown
- *(NEW)* probCalledStrike - Odds (0-1) that the pitch would be called a strike if taken, based on TruMedia's model.
- paResult

        - S – Single
        - D – Double
        - T – Triple
        - HR – Homerun
        - BB – Walk
        - IBB - Intentional Walk
        - HBP - Hit by Pitch
        - IP_OUT - In Play Out
        - K – Strikeout
        - FC - Fielder's Choice
        - DP - Double Play
        - TP - Triple Play
        - SH - Sacrifice Bunt
        - SF - Sacrifice Fly
        - ROE - Reached on Error
        - SH_ROE - Sacrifice Bunt ROE
        - SF_ROE - Sacrifice Fly ROE
        - BI - Batter Interference
        - CI - Catcher Interference
        - FI - Fielder Interference
        - NO_PLAY - No Play (ex: Runner Out)
    - battedBallType
        - PU - Pop Up
        - FB - Fly Ball
        - GB - Ground Ball
        - LD - Line Drive
        - BPU - Bunt Pop Up
        - BGB - Bunt Ground Ball
        - UK – Unknown

Project Output

- For analysts, we'd like to receive a PDF file documenting your finding and including any visualizations that you have produced.
- For developers, you can choose to host (we've included a document on hosting on github) your project and provide a public URL. Alternatively, you can send us a zip of static web files that we will host on AWS S3. If you send us a zip, please be sure to include an index.html file and make sure that your file paths are relative as your app will not be hosted at the root of the server.