

WEATHER AND BASEBALL

W200 - Project 2 - Nov 2019

Team Members

Alyssa Augsburger
Henry Bazakas
John Lee
Derek Topper

Team Github Repository

Project2_Augsburger_Bazakas_Lee_Topper

Summary

We plan to use basic pitch information from 2018 MLB pitch tracking data to identify how weather affects the game of baseball. We'd like to see how weather affects different aspects of baseball games such as characteristics of pitches (velocity, spin rate), attendance, home run likelihoods, ejection likelihoods, etc.

Dataset

Link to raw data: <https://drive.google.com/open?id=1Z4rKCPvN36HCfcr-9IZAfUWpY4p6Th4r>

Original dataset location: <https://www.kaggle.com/pschale/mlb-pitch-data-20152018#games.csv>

Description: The raw data contains 2015-2018 MLB regular seasons data and we plan to only use the 2018 data. Our dataset will consist of pitch-level data for every pitch thrown during the 2018 MLB regular season. Each row of the data represents a single pitch or game or player, depending on the file. Size of files range from 45 KB to 98 MB after we cleaned the dataset and filtered out all irrelevant variables.

The 5 raw/filtered datasets that will be used are listed below:

- atbats.csv (11 columns)
- ejections.csv (10 columns)
- games.csv (17 columns)
- PitchingData2018.csv (20 columns)
- player_names.csv (3 columns)

Initial Plots and Data Exploration

There are many variables that we wish to explore to generate insight on how batting, pitching, and miscellaneous factors are affected by weather (temperature, wind speed, and type of weather). The plots shown below demonstrate a few of these variables and the quantity of data available.

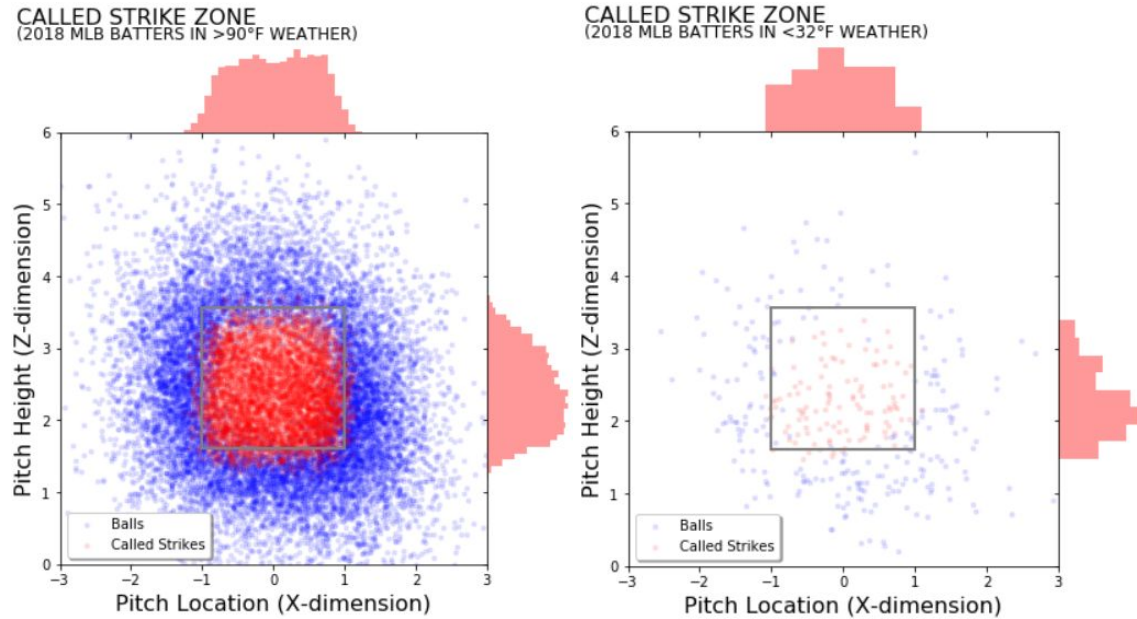


Figure 1a (left) and 1b (right) - Called Strike Zone of MLB Batters

Here are plots showing the called strike zones of MLB batters from the 2018 dataset. Figure 1a and 1b are filtered by weather conditions of greater than 90°F and less than 32°F , respectively. Even after filtering to show only extreme temperatures, there are still a large quantity of pitch data available in the dataset for us to analyze.

Range of Temperatures by Weather Type (Oakland Coliseum, Year 2018)

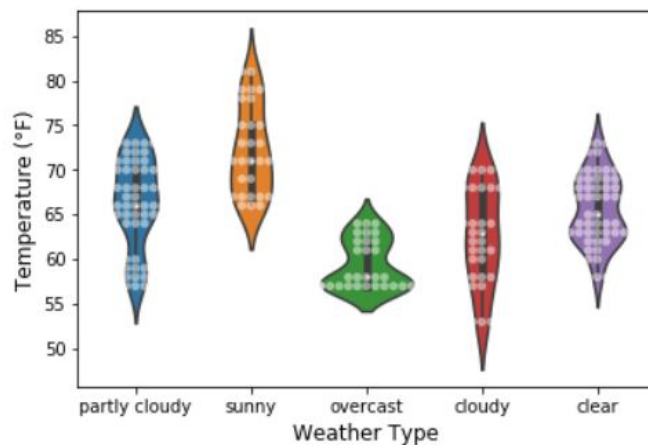


Figure 2 - Temperature Range versus Weather Type

Figure 2 shows the range of temperatures at Oakland Coliseum at different weather type conditions (sunny, overcast, etc...). Even though this is only at the Oakland Coliseum (a mild tempered baseball stadium), there is a wide range of temperatures and different weather types.

Other variables we'd like to explore include pitch speed (as it leaves the pitcher's hand and crosses the batter's box), pitch spin and direction, break angle, break length, break position,

pitch type, attendance, game length, and game delay. With these variables, we hope to answer more generalized questions such as: What is the optimal weather conditions to score the most runs or to maximize pitcher performance? Or how does the weather affect game time and attendance to best allocate stadium employee resources? We believe that weather should have a key impact on the game of baseball and we'd like to measure that.

Supplemental Dataset

We do not plan to use any supplemental data, as we believe a sufficient amount of data is included in the dataset that we will be using. The 5 datasets we will be using include a sufficient amount of data and about several different aspects of each baseball game and the weather.

Approach

In order to answer our guiding question, we will look at various aspects of the game of baseball and examine how each is affected by weather conditions. We plan to organize our report into three sections, each with a series of subquestions. Each section will analyze a different way that weather affects baseball.

The sections are:

1. Effects of Weather on Batting
2. Effects of Weather on Pitching
3. Other Effects of Weather on Baseball

In section one, we will look into how outcomes of batted balls vary depending on weather conditions, and how managers and players can best strategize for a given weather condition. In section two, we will compare pitch effectiveness in different weather conditions, looking at pitch speed, movement, and spin rate. We will use this data to analyze which pitches should be thrown more or less frequently depending on weather conditions. Furthermore, we will look into weather's effect on how many pitches starters throw and the effect of weather delays on pitcher performance. In section three we will look at a variety of miscellaneous questions regarding weather's effect on baseball. Among them are weather's effects on attendance, game length, and team performance. By examining all three of these sections, we hope to obtain a variety of insights into how weather affects baseball games and how teams should respond accordingly.