Final Project Writeup: Pitchf/x Data

USING PITCHF/X DATA TO SORT AND ANALYZE PITCH BREAK AND CALLED STRIKES

PAUL VARJAN, STEVE PELKEY, Kent owen

# DATA SET and areas of analysis

The primary dataset used in this analysis is the Pitchf/x dataset owned and maintained by Major League Baseball Advanced Media, which tracks the speeds and trajectories of every baseball pitched in every Major League Baseball game. All 2015 data were joined with weather data from a separate source for every game. We have two primary areas of interest:

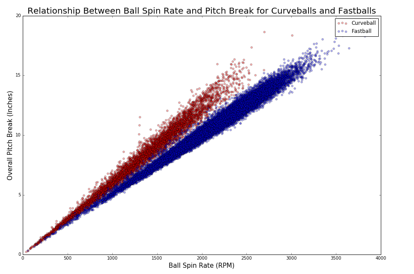
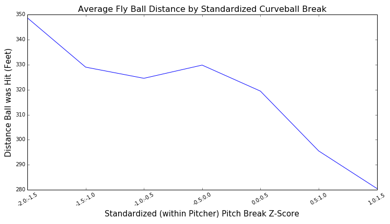
1. **EXPLORATION OF TOTAL PITCH BREAK BASED ON SPIN RATE, VELOCITY, STADIUM ALTITUDE, WEATHER**
2. **EXPLORATION OF HOME FIELD IMPACT, WEATHER IMPACT ON UMPIRE CALLED STRIKES**

# 1) EXPLORATION OF TOTAL PITCH BREAK BASED ON SPIN RATE, VELOCITY, STADIUM ALTITUDE

## SPIN RATE AND TOTAL BREAK OVERVIEW AND IMPORTANCE

In professional baseball, pitchers put spin the ball at delivery to alter the overall trajectory of the pitch. The change in trajectory is determined mainly by the [Magnus effect](https://en.wikipedia.org/wiki/Magnus_effect) – a fundamental principle of aerodynamics. This phenomenon states that the rotation of any cylinder or ball causes an angular deflection of the surrounding air depending on the type of spin (e.g. backspin, topspin, and sidespin). Therefore, pitchers vary the spin rate, angle, and overall velocity of the ball to produce over 10 unique pitch types (Fastball, Curveball, Changeup, etc.) that are categorized by their speed and movement along the x and z axes. The relative x/z movement of the pitch is called **pitch break**. **Figure A** depicts the relationship between ball spin and break for two pitch types and nicely demonstrates the Magnus effect. The slope differences between curveball and fastball are due differences in velocities of pitch type.

**Figure B**



**Figure A**

Since the Magnus effect depends on the turbulent wake created in the airflow surrounding a ball, it logically follows that the deflective force of the air might significantly vary by meteorological factors such as air density, temperature, or humidity. We will be further exploring this relationship because breaking balls (which are typically slower pitches that do not travel straight as they approach the batter, such as a Curveball) with less break result in good bat contact, demonstrated by **Figure B**. As such, weather may have impact on batting outcomes purely by modifying total break of the pitch.

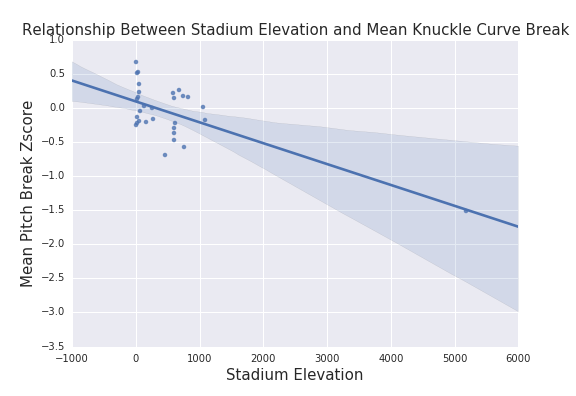
## Stadium Elevation

Of the 30 stadiums in major league baseball, the average elevation is 517 feet above sea level; however, Coors Field in Denver, Colorado is a notable outlier at 5,183 feet. Since higher elevations have considerably thinner air, it is hypothesized that the thin air will exert less force upon the ball causing less pitch break.

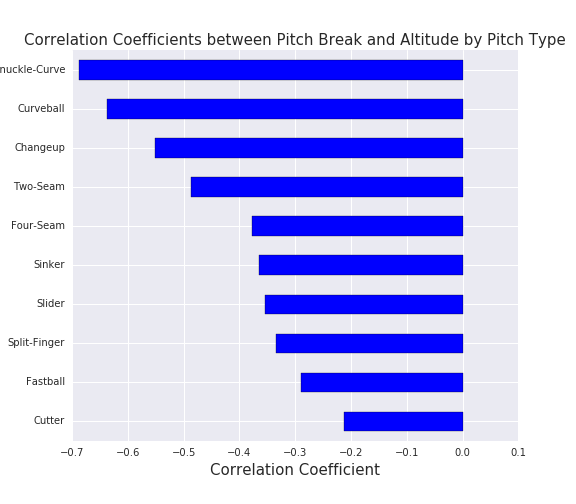
It is at first difficult to reasonably infer that the relative lack of pitch break at Coors Field is actually due to altitude rather than other factors unique to Colorado. For example, a potential confounding factor is the frequency at which the Colorado pitching rotation pitches at home. The Colorado Rockies are one of the worst teams in baseball, and have difficulty attracting pitching talent.

In order to determine the relationship between stadium elevation and pitch break, z-scores were calculated for pitch movement according to pitch type for each unique pitcher. In other words, the break of every pitch is compared to that pitcher’s average respective pitch break in an attempt to minimize the effects of between-pitcher variation.

After computing the mean pitch break z-score for each stadium, an inverse relationship with stadium altitude is easily apparent (see **Figure C**). In fact, the average break of a knuckle curve thrown at Coors Field is 1.5 standard deviations below the mean, which corresponds to roughly 4.4 inches. The overall correlation between stadium elevation and mean pitch break is -.54, but the exact correlation coefficient varies by pitch type (**see figure D**). Additionally, overall correlation between average standardized pitch break for home and away pitching is a very tight .94 making this point moot.



**Figure C**



**Figure D**

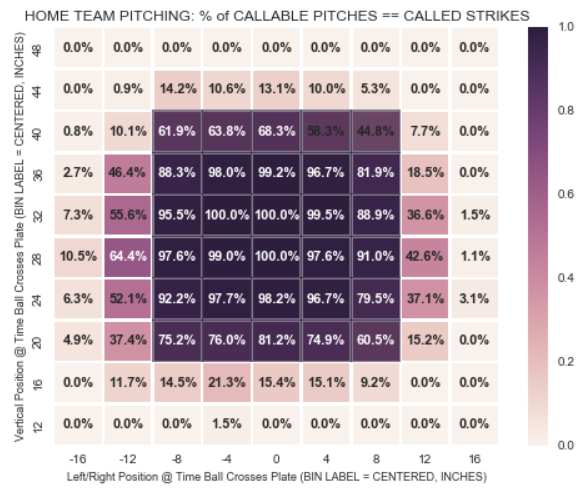
## Weather IMPACT ON PITCH BREAK

*IN-PROGRESS…*

# 2) Exploration of home field impact, weather impact on umpire called strikes

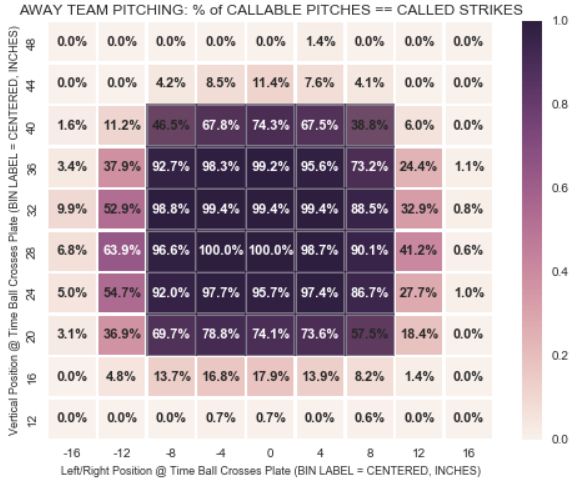
## HOME-FIELD IMPACT ON UMPIRE STRIKE CALLING

In Major League Baseball, there is a distinct advantage to playing at a team’s “Home” stadium. In fact, from 1920 onwards, the home team has won 54% of MLB games (<http://www.baseball-reference.com/blog/archives/9916>), despite having equal talent to the away team (when compared across large enough sample sizes). Though umpires are *trained* to call a consistent game, we have been exploring whether there could actually be a mental factor to calling a baseball game, and whether umpires are influenced to call more strikes by the home crowd. An umpire is meant to call a consistent “Strike Zone” of called balls v. called strikes, but if they are influenced by the crowd or other home factors in any way, we will see material differences in their Called Strike rates.



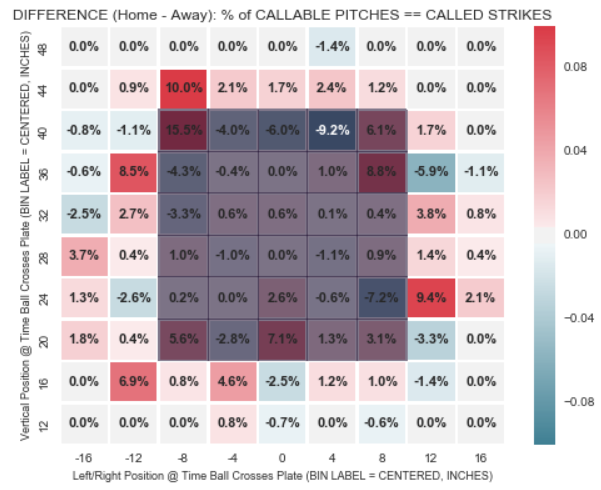
**Figure E (Home Team Pitching)**

First, we filtered the data to be only on “Callable” pitches (called balls and called strikes), removing any data where the batter made a swing (appx. 50% of the data). However, we cannot just take the percentage of called strikes for when a home team is pitching v. when an away team is pitching on this subset to derive any meaningful conclusions, as there may be other factors at play. For example, if a pitcher is better at pitching into the strike zone at home compared to when he is away, fully unbiased umpiring would still show a higher percentage of called strikes for the home team. As such, we need to analyze the ***percentage of called strikes by exact location*** as the ball crosses the plate in order to remove for overlapping home/away pitcher performance impact. To do this, we “binned” the location of each pitch by 4x4 inch sections of when the ball crosses the plate, then brought in the rate of called strikes by each location bin.



**Figure F (Away Team Pitching)**

Note that in each of our charts, we mapped the average strike zone size over top of the bins (dark grey box) for a visual representation of the strike zone. We are also only showing just two extra bins outside of the strike zone, as pitches more than 2 bins (8 inches) outside of the average strike zone are rarely called strikes for either the home or away team, and as such have minimal impact on analysis. See **Figure E (Home Team Pitching), Figure F (Away Team Pitching), and Figure G (Difference between Home and Away Teams)** for a comparison of these bins.



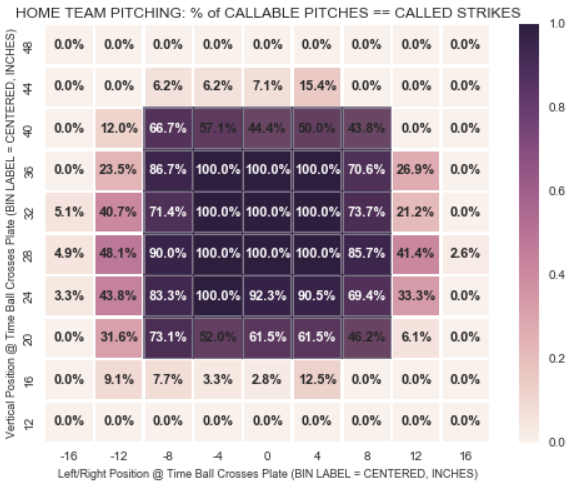
**Figure G (Difference Between Home & Away)**

If umpires were uninfluenced by whether a home v. away team is pitching, we would see minimal differences by bin that average near 0.0% **in Figure G**. However, in Figure G, you can visually see more red than blue bins, indicating that the home team gets strikes called at a higher rate. In fact, as a % of total pitches per bin, the home pitcher gets:

* 0.1% more strikes called in the inner bins of the standards strike zone
* 0.6% more strikes called in the outermost bins *within* the standard strike zone
* 1.7% more strikes called *one bin outside* the standard strike zone
* 0.2% more strikes called *two bins outside* the standard strike zone
* 0.1% more strikes called *two bins outside* the standard strike zone

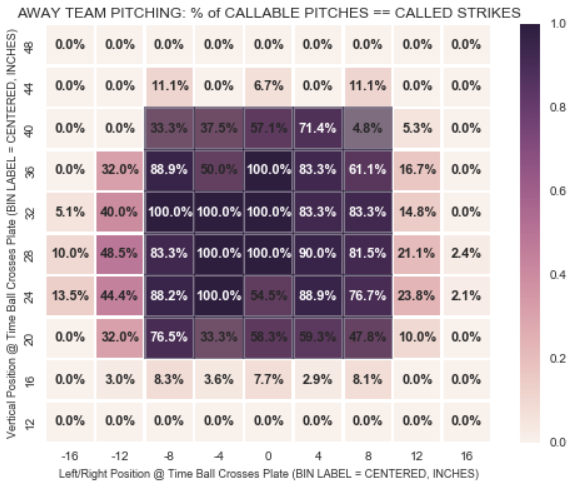
Overall, as a % of callable pitches, there are 0.61% more strikes called for the home pitcher than the away pitcher! As 50.2% of pitches are callable pitches, this equates to 0.31% of pitches in the full data set that that the home team will receive an extra advantage of getting a called strike on. While this may not seem especially material at f, this has a significant impact over the course of a game, and especially over the course of the season. Separate research has shown the value of an additional strike called to be worth ~0.14 runs (<http://www.baseballprospectus.com/article.php?articleid=22934>). As such, over the course of a typical baseball game of ~290 pitches (<http://www.baseball-reference.com/blog/archives/7533>), the impact of an umpire is approximately 290 pitches/game\*0.31% difference in called strikes per pitch\*0.14 run value per additional called strike = **0.12** runs per game. As such, umpire impact in itself is enough to give a legitimate home field advantage! Extrapolating this further, over the course of a 164-game season, a team can expect to score **20** more runs playing at home than playing away purely due to intentional or unintentional umpire bias.

**Figure H (Home Team Pitching, 2 STRIKES)**



Furthermore, we wanted to explore when the pressure and importance increases on a callable pitch whether the umpire actually gives an even greater bias towards the home team. We did this by exploring the subset of when there were two strikes in the count; meaning one more strike and the batter strikes out.

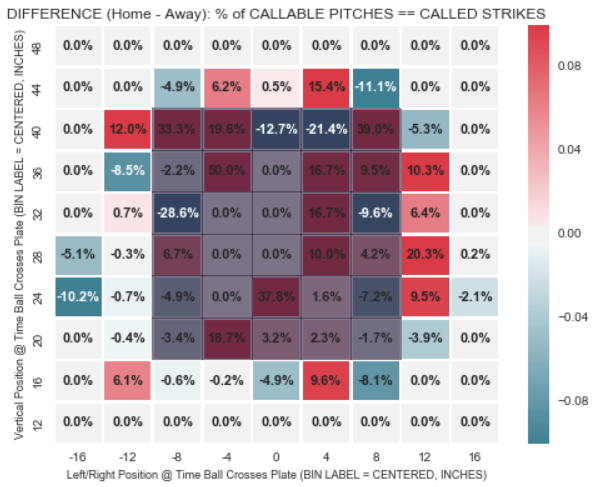
Almost unsurprisingly, when there is increased pressure and importance on the call, umpires actually get more biased. In fact, as a % of total pitches per bin, with two strikes in the count the home pitcher gets:



**Figure I (Away Team Pitching, 2 STRIKES)**

* 10.5% more strikes called in the inner bins of the standards strike zone
* 1.8% more strikes called in the outermost bins *within* the standard strike zone
* 2.0% more strikes called *one bin outside* the standard strike zone
* -0.7% more strikes called *two bins outside* the standard strike zone
* 0.0% more strikes called *two bins outside* the standard strike zone

Overall, as a % of callable pitches, the home team gets a 2.0% advantage of called strikes when there are already two strikes in the count, greatly increasing the chance of an opposing strikeout.



**Figure J (Difference between Home and Away Team Pitching, 2 STRIKES)**

## WEATHER IMPACT ON UMPIRE STRIKE CALLING (“WIMPIEST UMPIRE”)

*IN-PROGRESS…*