Analyzing Catcher Defensive Index

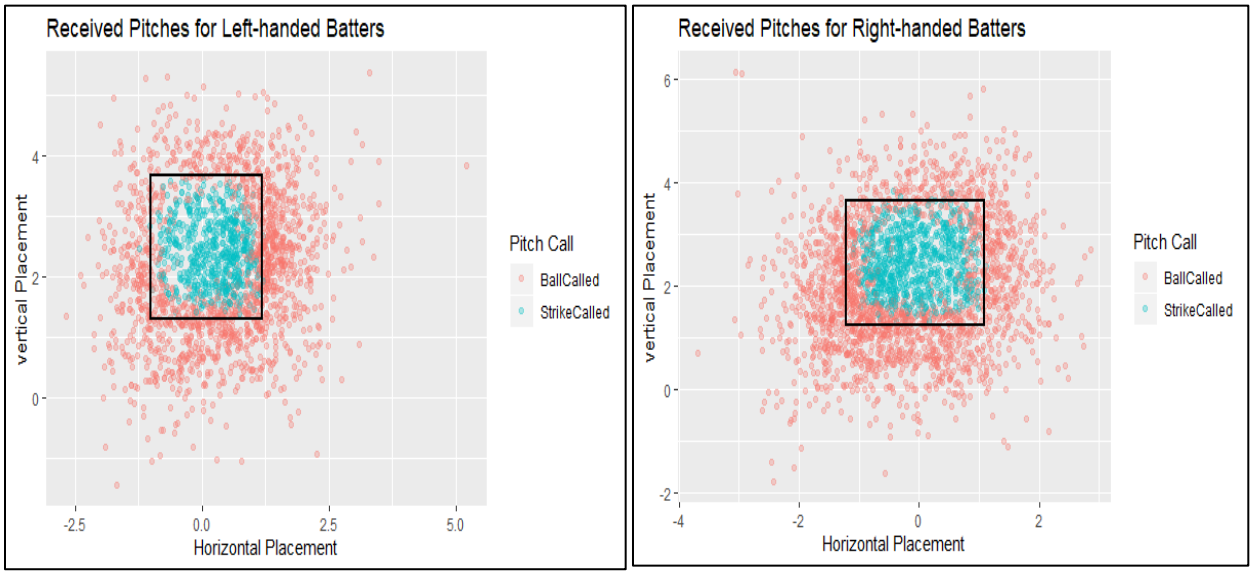
Daniel Lee

I. Introduction

Catcher is fundamentally an important position, for it being not only the commander to the pitchers, but also in various other ways such as being able to transform the borderline balls to borderline strikes, being a good communicator to a pitcher and catching steals, etc. Although the contextual information may be needed for the pitch f/x data to be utilized to its full potential and generate insights about the catcher's defensive ability, there are exploratory data analysis and certain implications that can be posed from the pitch data. In this report, I explore a few key characteristics of the new Mariners catcher (ID – f06c9fdf) through the pitches that he received throughout the season.

II. Pitch Framing

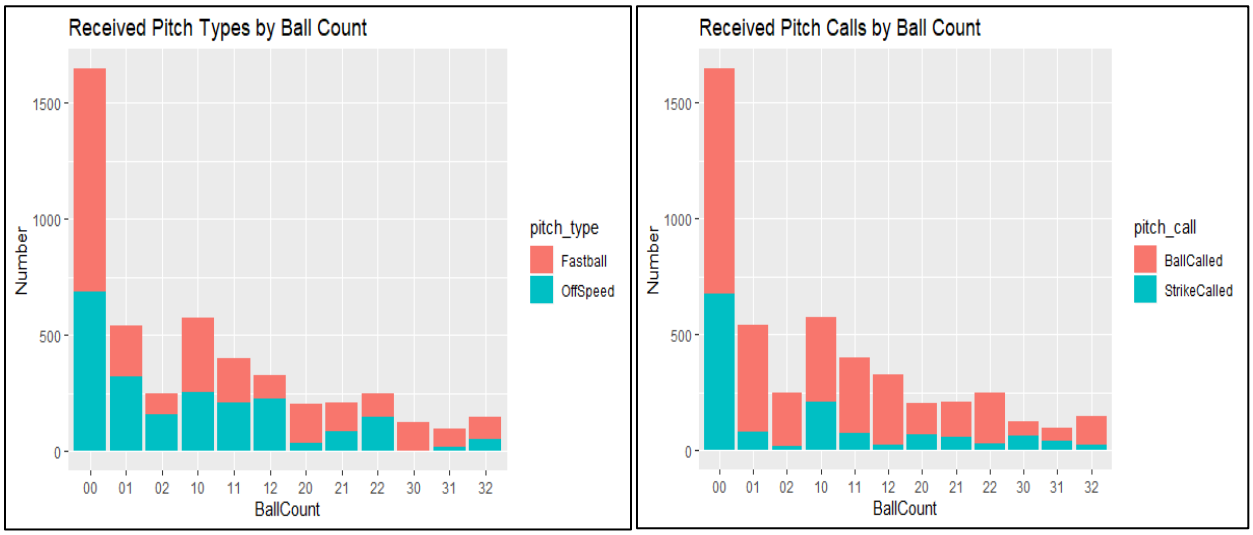
Pitch framing is the ability of a catcher to generate strike calls on what are seemingly borderline balls. To explore the framing ability of the catcher, it is important to generate the virtual strike zones for respective umpires and batter-handedness. We already have this from the continued data frame from Part I – the "BorderInclusion" variable is a binary variable that is assigned if the pitch falls in between the maximum and minimum height and width boundaries. Below is the result of the plot of received pitches for the catcher for left-handed batters and right-handed batters:



The “Zone” is constructed from taking the mean boundaries of all the umpires involved with the catcher’s received pitches. Since the zone is constructed from the mean of the maximum and minimum values for width and height, we have to acknowledge that the actual strike zone is narrower than what is appeared to be on the plots and that this is only for comparative analysis on an exploratory level building block. To comment about the visuals at a first glance, we can see that there are very few cases of balls outside of the respective zone being called strikes. This may imply that the catcher is not so strong in pitch framing.

For further detailed and more precise analysis however, there needs to be a more structured approach where we can train the individual strike zone with the batter’s height factor considered with neural network model. With this we can assign, rather than binary border inclusivity, a probabilistic figure to a given pitch being called strike or not. From there, a following algorithm may be applied: for every ball (w/ P percentage) that was called strike, P*strike – P*ball.

III. Pitches by Ball Count



To comment about the positive implications, the pitch types are relatively evenly distributed between fastball and off-speed pitches on breakeven counts. This means that there is a lot of variance in terms of the type of pitch a batter may get on a count and can give confusion to the opponent’s plan of attack. The pitch counts are skewed to the right, which means that the counts are usually kept short until the batter gets out via in-play.

To comment about the negative implications, the pitch types get quite predictable and routine under high-pressure situation such as 2-0, 3-0, 3-1, or 3-2; the catcher prefers fastball much more than the off-speed pitch. This may seem like an obvious preference, but several off-speed pitches are proven to be affective to catch batters off-guard. With more data readily available, the variety of pitch types, steal catches, strikes under high pressure situation such as right after runs given, straight balls and men on base. Higher variance in pitch patterns for every circulation of batter-lineup implies unpredictability and the efficiencies of off-speed pitches implies the pitch count efficiency.