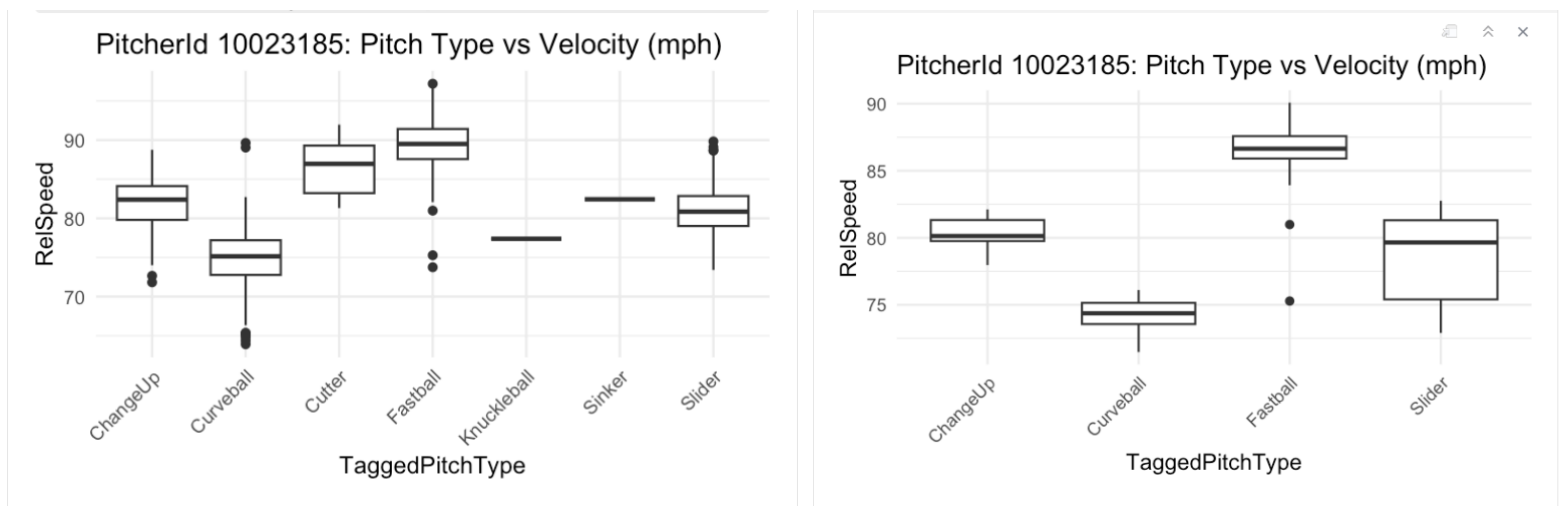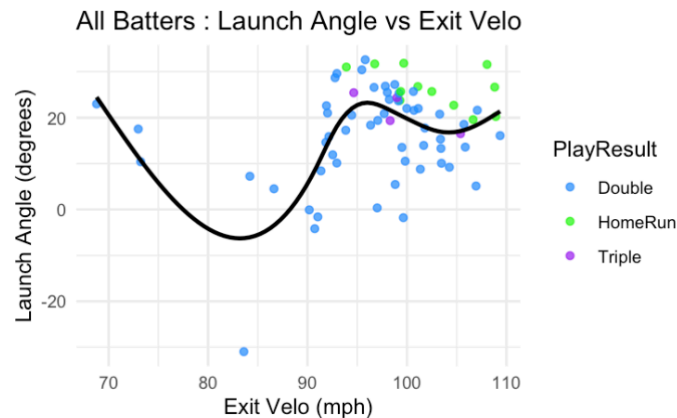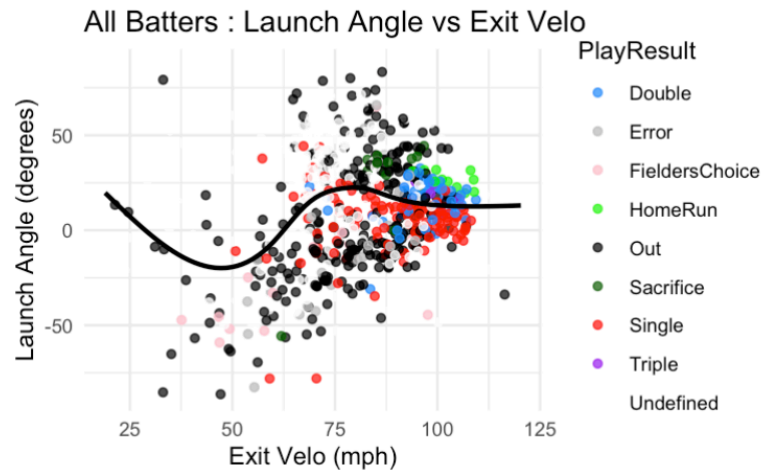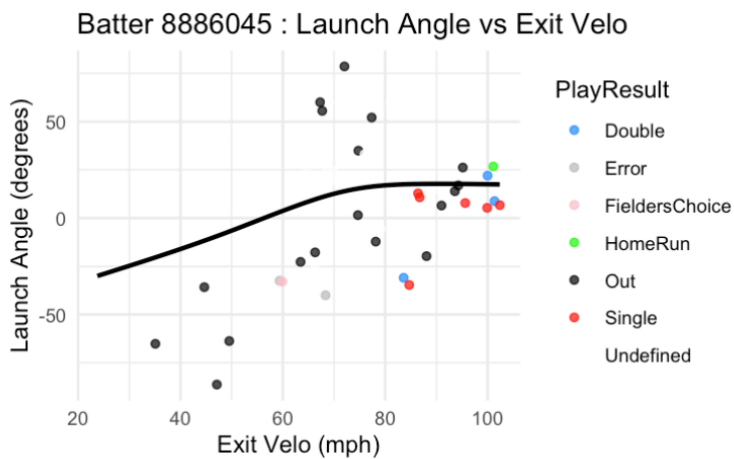I chose to create data visuals and draw conclusions from the Track_Combo.csv data set on collegiate baseball data.

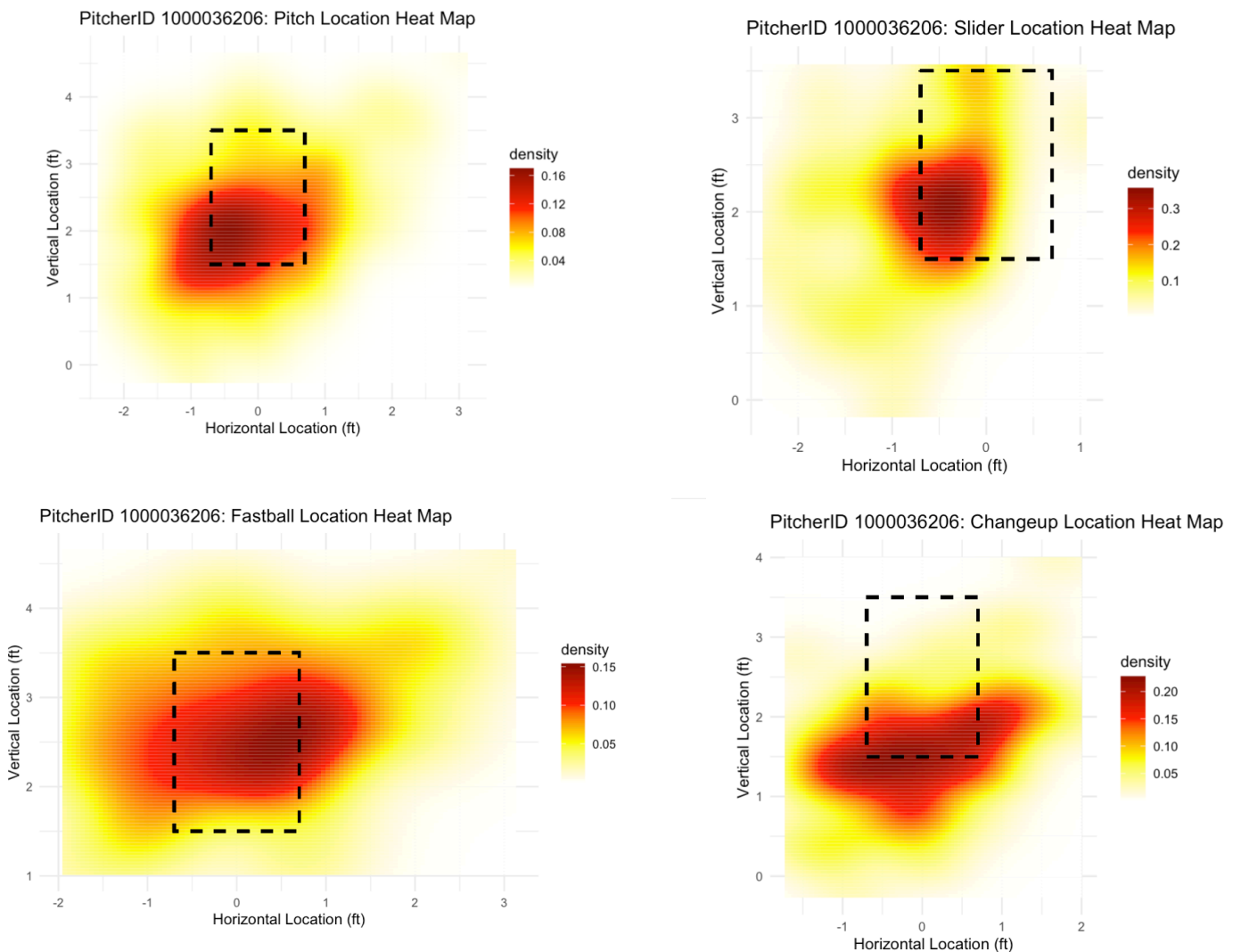# Inference 1 : Pitcher 10023185 Velocity by Pitch Type

The box and whiskers plot above separates PitcherId 10023185's pitches into the following splits: ChangeUp, Curveball, Fastball, and Slider. Then for each category, it creates a box representing the IQR (middle 50 percentile) of that respective pitch's speeds, with the whiskers representing the minimum-25%, and 75%-maximum and the dots representing outliers. For the ChangeUp, we see a very short box meaning that this pitch was thrown with a very short range of differing speeds. A majority of the time, the changeup clocked in around 79-82 mph, with the median coming around 80 mph. We see a similar trend for both the curveball and fastball, as well. A majority of the curveballs were clocked in between 73-75 mph with the median being roughly 74 mph. A majority of the fastballs were clocked in between 86-87.5 mph with the median falling around 87 mph. However, we do notice that there are 2 outliers in which his fastball was thrown at around 75 mph and 81 mph. Lastly, his slider came in at a wide range of speeds, with the middle 50 percentile of pitches getting measured between 75-81 mph. If we wanted to compare this to the rest fo the league, we can see that Pitcher 10023185's changeup falls around the 25-50th percentile for speed of changeups resulting in a strike. Pitcher 10023185's curveball also falls around the 25-50th percentile for speed of curveballs resulting in a strike. Pitcher 10023185's fastball falls around the 12.5-25th percentile for speed of fastballs resulting in a strike. Lastly, Pitcher 10023185's slider falls around the 5-30th percentile for speed of sliders resulting in a strike.

Inference 2 : Batter 8886045 Launch Angle vs Exit Velocity

I created a scatter plot with a generalized additive model regression line using the data from BatterID 8886045. On the y-axis, I am tracking Launch Angle and comparing it to Exit Velocity on the x-axis. Some of the data points were labeled as "Undecided". I am not sure if this was to resemble foul balls, or if there was an error in the tracking, but regardless I whited-out these data points since we are not aware of the actual outcome. I also color-coded the known outcomes, separated by singles, doubles, triples, home runs, ROE, and outs. The GAM regression line gives us the overall trend of the data without needing any parametric assumptions. After looking at BatterID 8886045's outcomes, I decided to expand this to the entire data set to look for a common trend. Lastly, I transitioned my focus to only include data points that resulted in an extra base hit, since that could be a realistic approach for a coach - to increase extra base hits. After doing so, I noticed that there is a strong, positive relationship between exit velocity and launch angle from around (82.5 mph, 95 mph) and (-2.5º, 22º). The tighter range from (92.5 mph, 95 mph) and (19º, 22º) seems to resemble the optimal swing pattern until diminishing or even negative returns are noticed. Depedning on a coach's focus, we see that a majority of the home runs hit fall above the 22º swing path; however, if we look at the entire batter data, we can see that isolating to swings with a launch angle greater than 22º results in a sacrifice or an out more often than it does a home run.

Inference 3 : Pitcher 1000036206 Evaluation via Heat Map

PitcherID 1000036206: Pitch Location Heat Map

PitcherID 1000036206: Slider Location Heat Map

PitcherID 1000036206: Fastball Location Heat Map

PitcherID 1000036206: Changeup Location Heat Map

I created a heat map for PitcherID 1000036206's pitches. Areas that are colored in to be darker red represent locations where the pitch is more frequently thrown while locations that are lighter yellow represent less dense spots. When evaluating PitcherID 1000036206, I noticed that this pitcher seems to keep his pitches (specifically the slider and changeup) in the ideal part of the zone. Thus, if a batter is able to read the spin and laces of the pitches coming in and identifies it as a slider, they can most likely predict that the slider will finish in the low and away section of the strike zone unless it is left hanging. On the other hand, if the laces are spinning similarly to that of a changeup, it appears that the pitch may result in a ball more often than not, so it may be more successful to let the pitch go with the alternate mindset of fouling it off in emergencies. Lastly, we can see that PitcherID 1000036206 favors to throwing his fastball inside to a righty and outside to a lefty, but with a lot more variation. ***Note: This graph is meant to represent the directional POV of a pitcher, not a batter.