JARED HARNICK, JONATHAN MANTA, & TEJ SUKLIKAR

How to get a good WHIFF

Analyzing different locations in the strike zone where batters make contact or miss. Specifically fastballs, sliders, curveballs, and changeups.

Background

In our project, we used batters' swing and miss data from the 2019 season from Statcast. From there we constructed a heatmap for our predictions and results using logistic regression and comparing our prediction with the actual data.

Hypothesis

Can a logistic regression model predict where MLB pitchers should throw a baseball to have the most likelihood to get a swinging strike?

Data & References with Code

REFERENCES: STATCAST.COM, BASEBALLSAVANT.MLB.COM

What did our group exclude?

We filtered our data to include fastballs, curveballs, sliders, and changeups as these are the most common types of pitches.

How did we test error?

We used **Brier scores** to test the **accuracy** of our models and all of our models ended up having a score **between .2 and .25.**

Logistic Regression Models We Created

```
logit_fb<- glm(Call~plate_x+plate_z,family=binomial,data=fb)
logit_fbz<- glm(Call~plate_z+plate_x+release_speed+release_pos_x,family=binomial,data=fb)
logit_fbx<- glm(Call~plate_x+plate_z+release_pos_x+release_pos_z,family=binomial,data=fb)
logit_fbv<- glm(Call~plate_x+plate_z+pfx_x+pfx_z+release_pos_x+release_pos_z,family=binomial,da

fb_pred<-
    fb%>%
    add_predictions(model=logit_fb, type="response", var="Prediction")%>%
    add_predictions(model=logit_fbz, type="response", var="Prediction2")%>%
    add_predictions(model=logit_fbx, type="response", var="Prediction3")%>%
    add_predictions(model=logit_fbv, type="response", var="Prediction4")

fb_pred

summarise(fb_pred,
        phatfb_all=mean((Prediction-Call)^2),
        phatfb_z=mean((Prediction2-Call)^2),
        phatfb_x=mean((Prediction3-Call)^2),
        phatfb_v=mean((Prediction4-Call)^2))
```
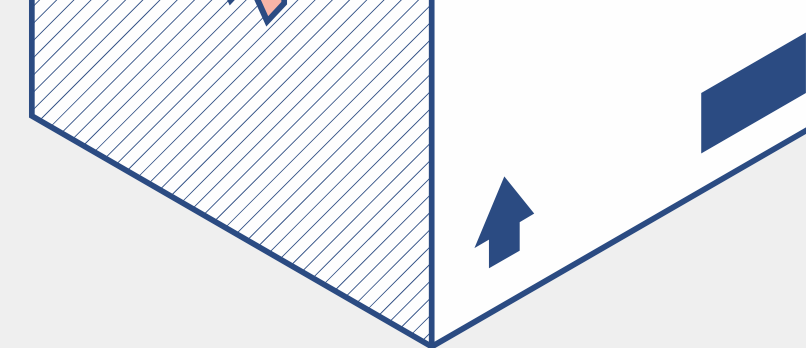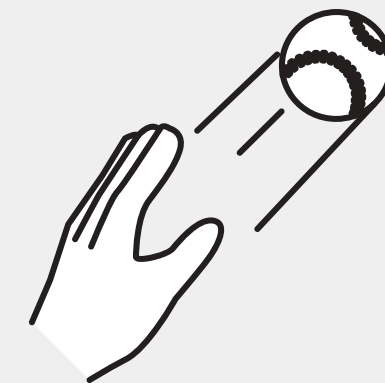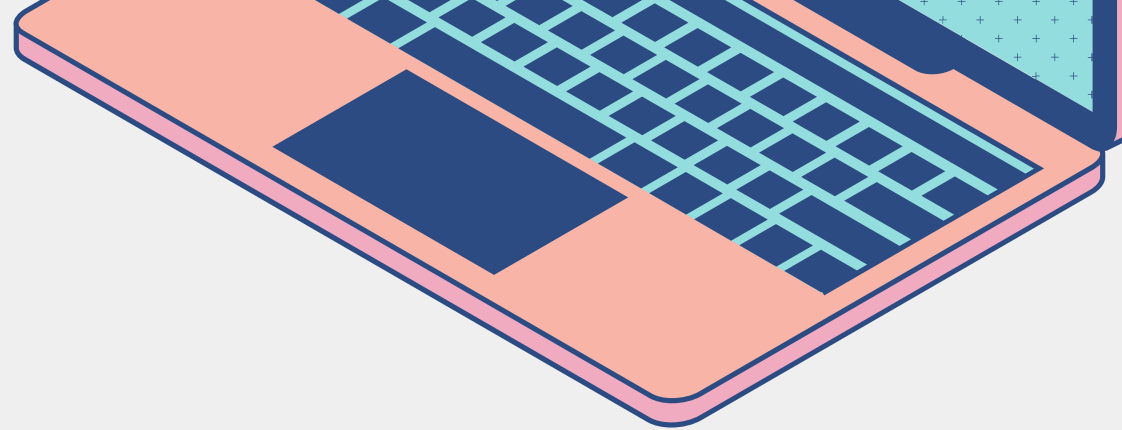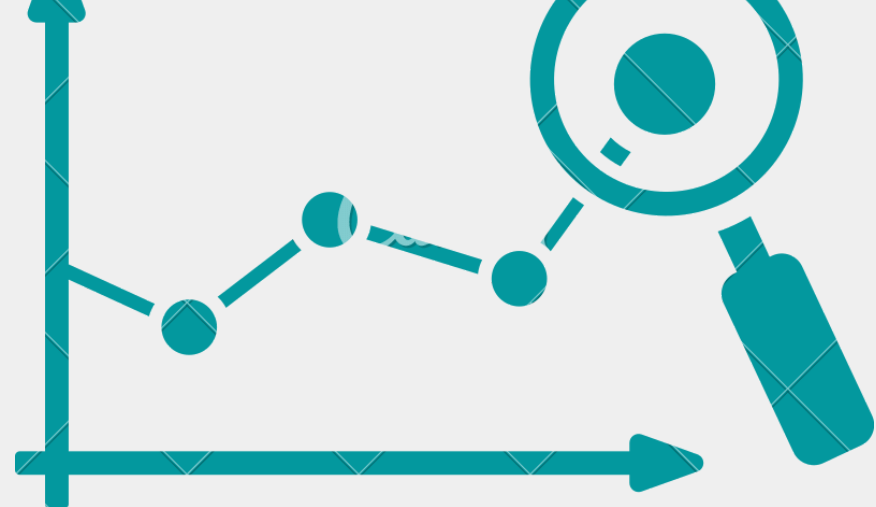
Right Handed Pitchers vs. Right Handed Batters

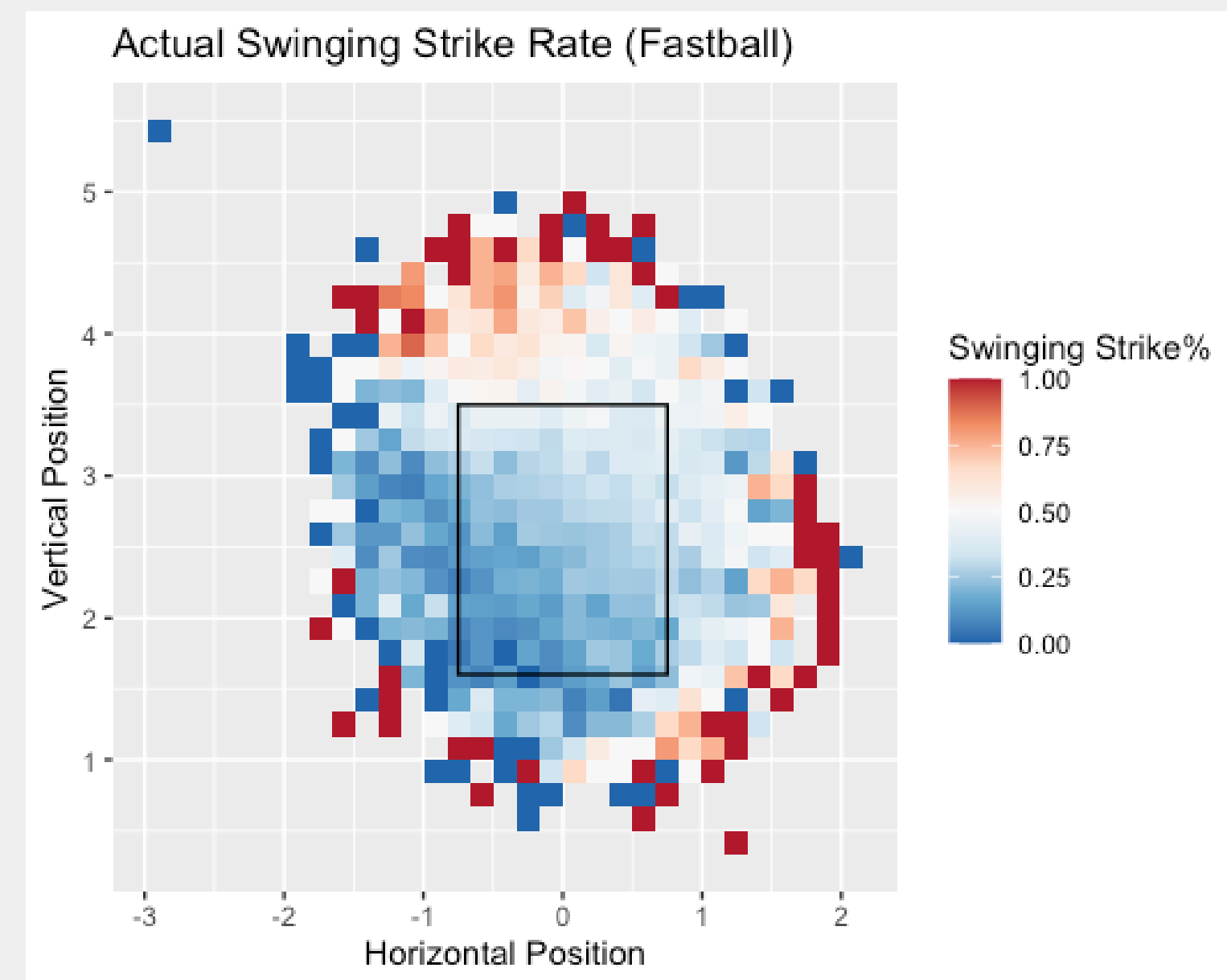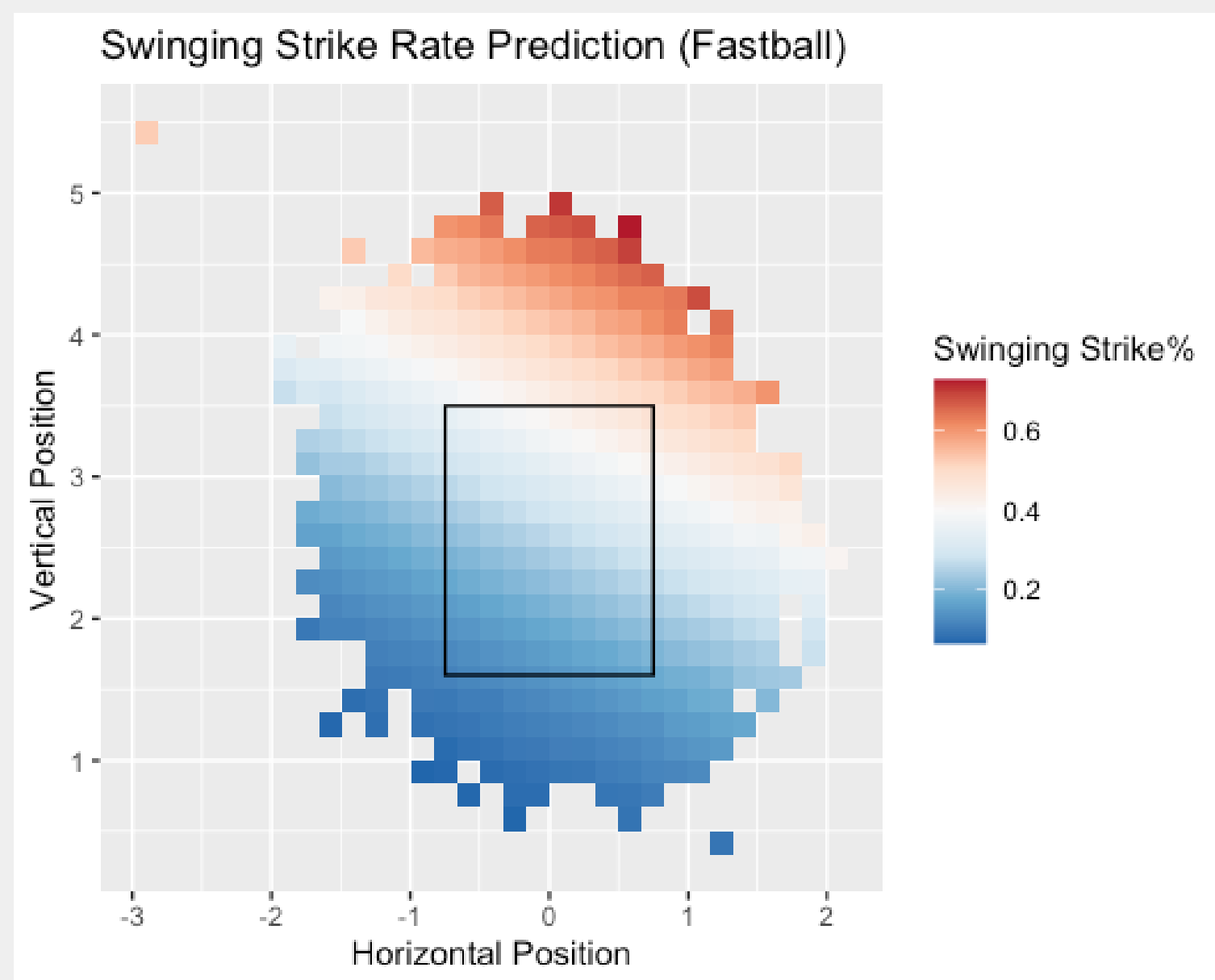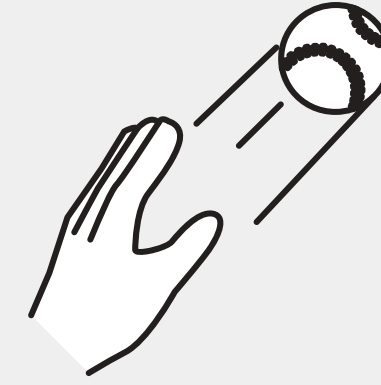# Right Handed Pitchers vs. Left Handed Batters

Fastball
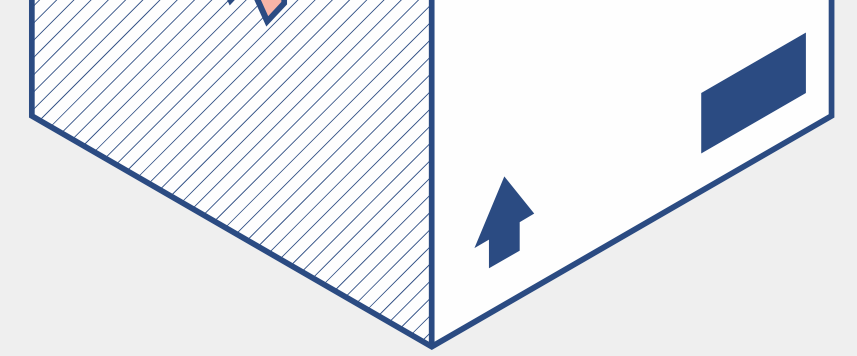
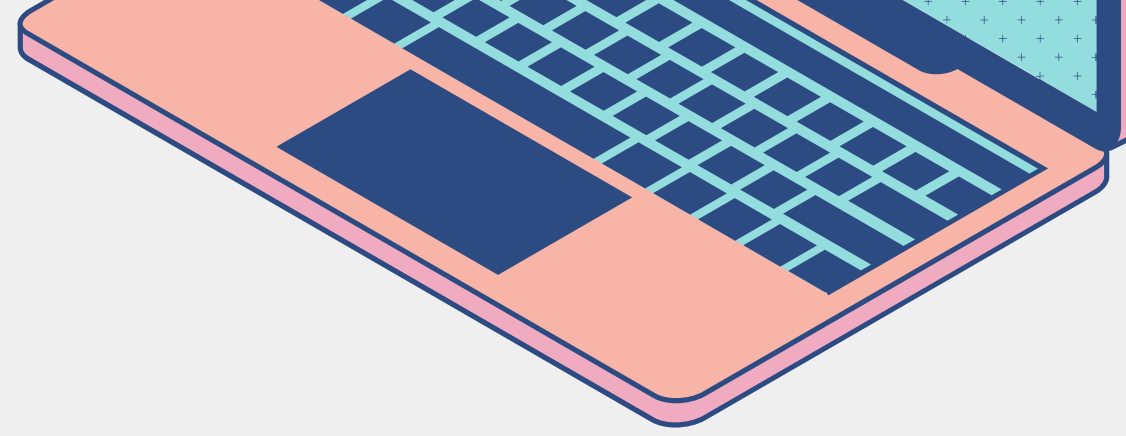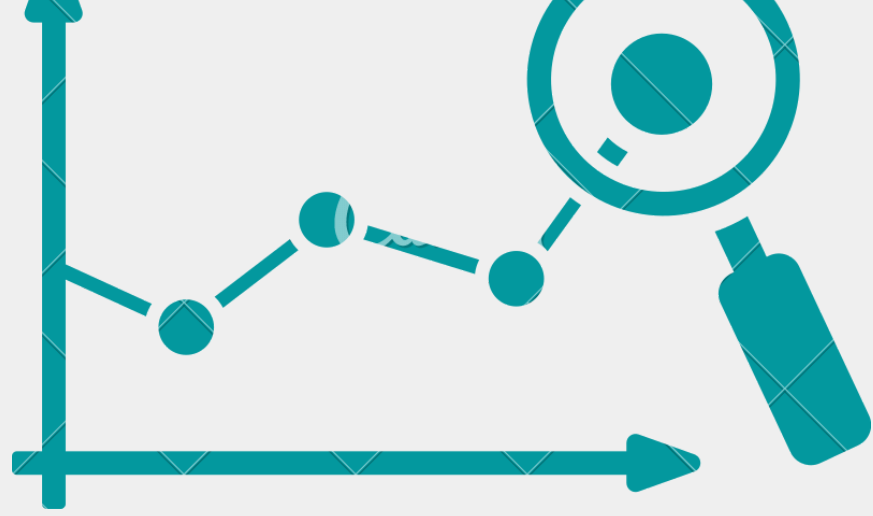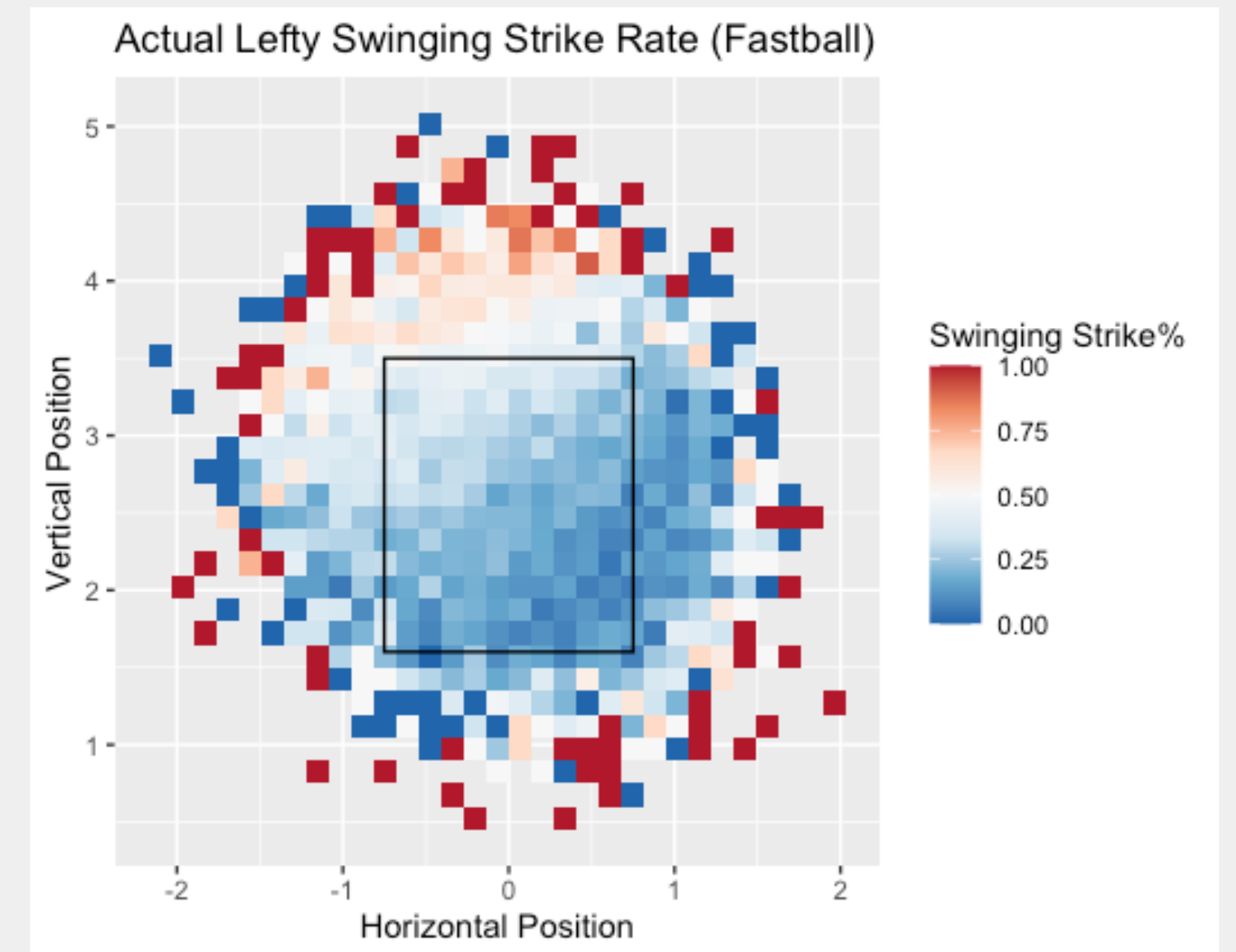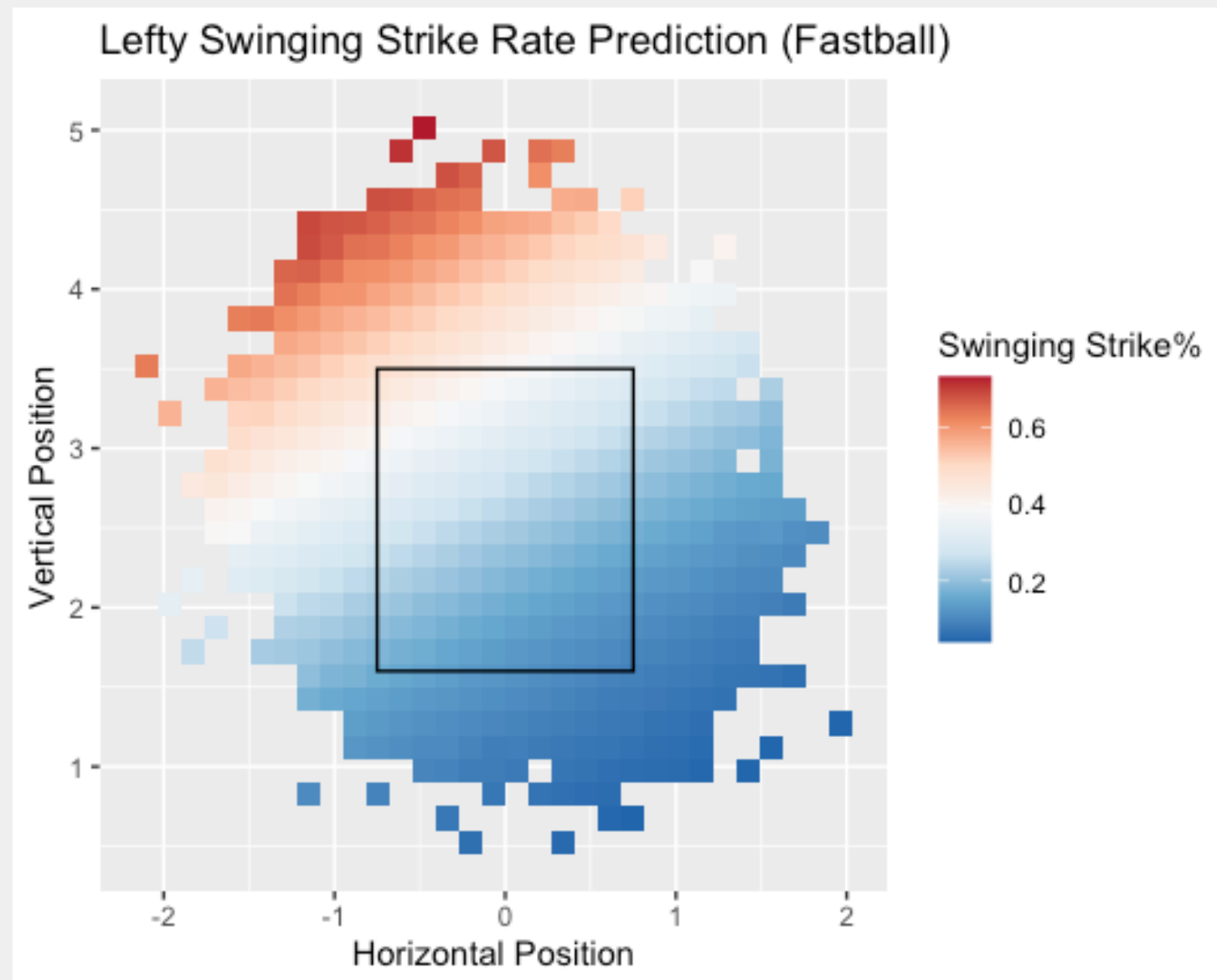# Fastball (Rvs.R)

Prediction

Actual

Fastball (Rvs.L)

Prediction

Actual

Lefty Swinging Strike Rate Prediction (Fastball)

Actual Lefty Swinging Strike Rate (Fastball)

Slider

Slider (Rvs.R)

Prediction

Actual

Swinging Strike Rate Prediction (Slider)

Vertical Position vs Horizontal Position

Swinging Strike%
- 0.75
- 0.50
- 0.25

Actual Swinging Strike Rate (Slider)

Vertical Position vs Horizontal Position

Swinging Strike%
- 1.00
- 0.75
- 0.50
- 0.25
- 0.00

Slider (Rvs.L)

Prediction

Actual

Curveball

Curveball (Rvs.R)

Prediction

Actual

Curveball (Rvs.L)

Prediction

Actual

Changeup

Changeup (Rvs.R)

Prediction

Swinging Strike Rate Prediction (Changeup)

Actual

Actual Swinging Strike Rate (Changeup)
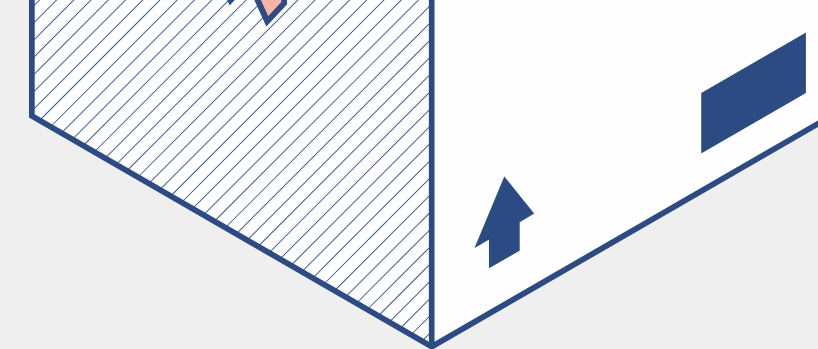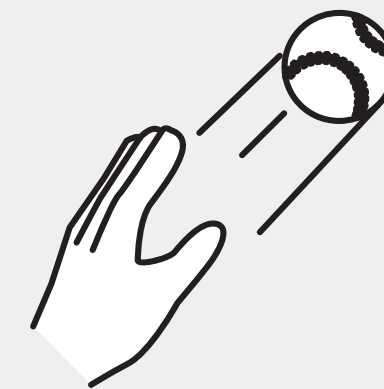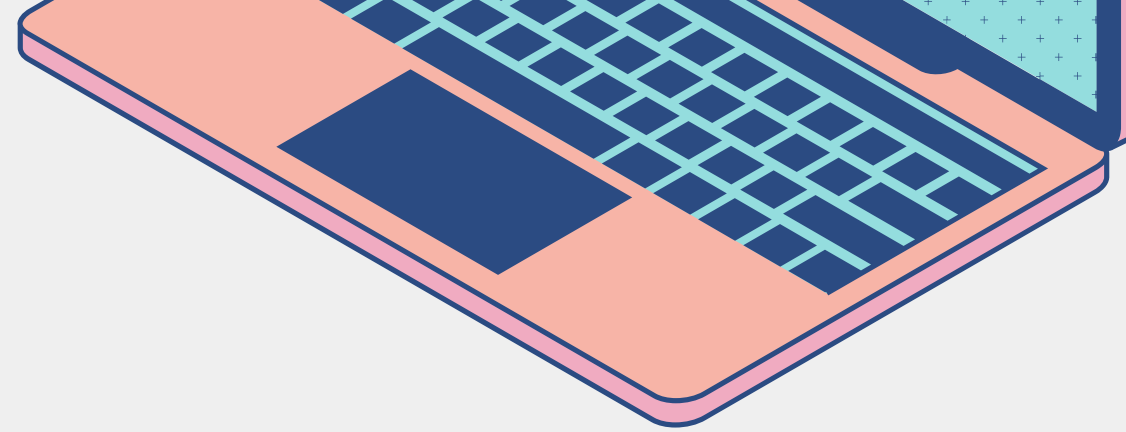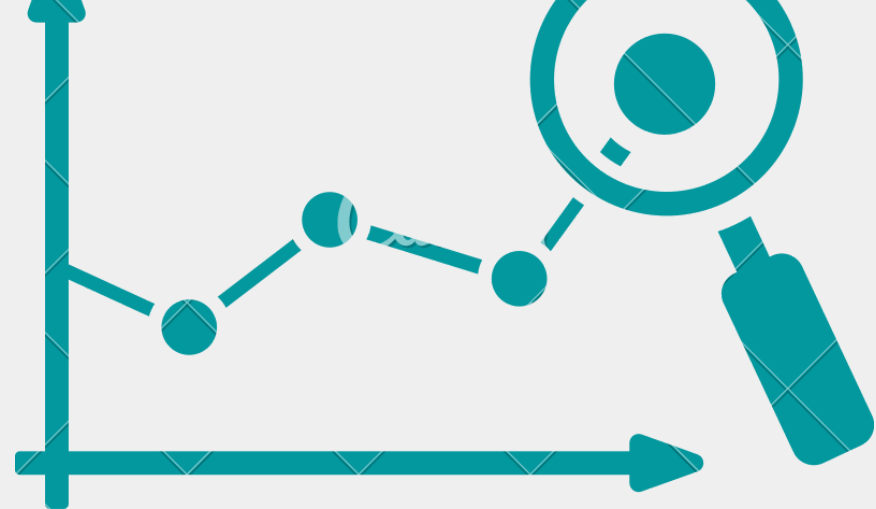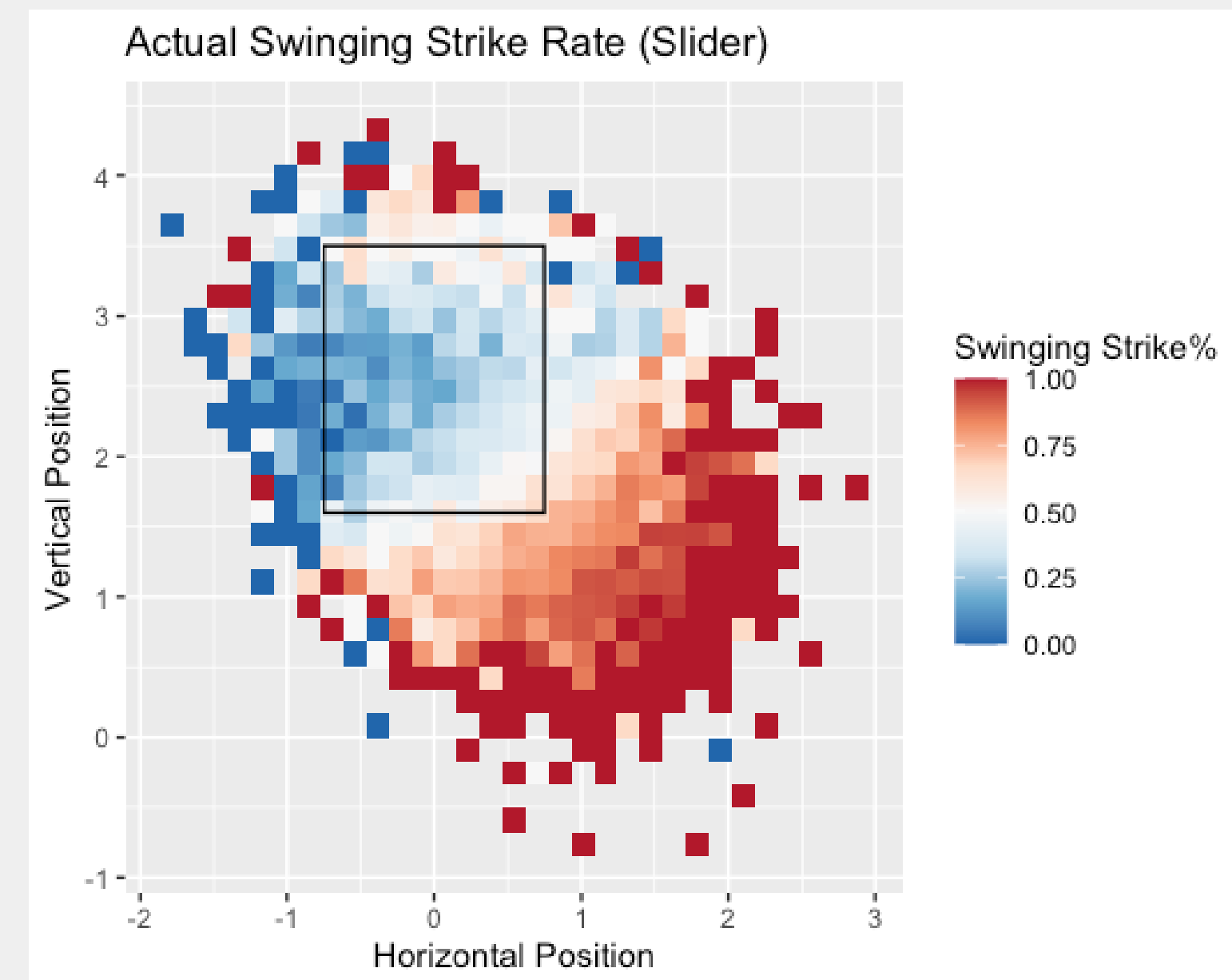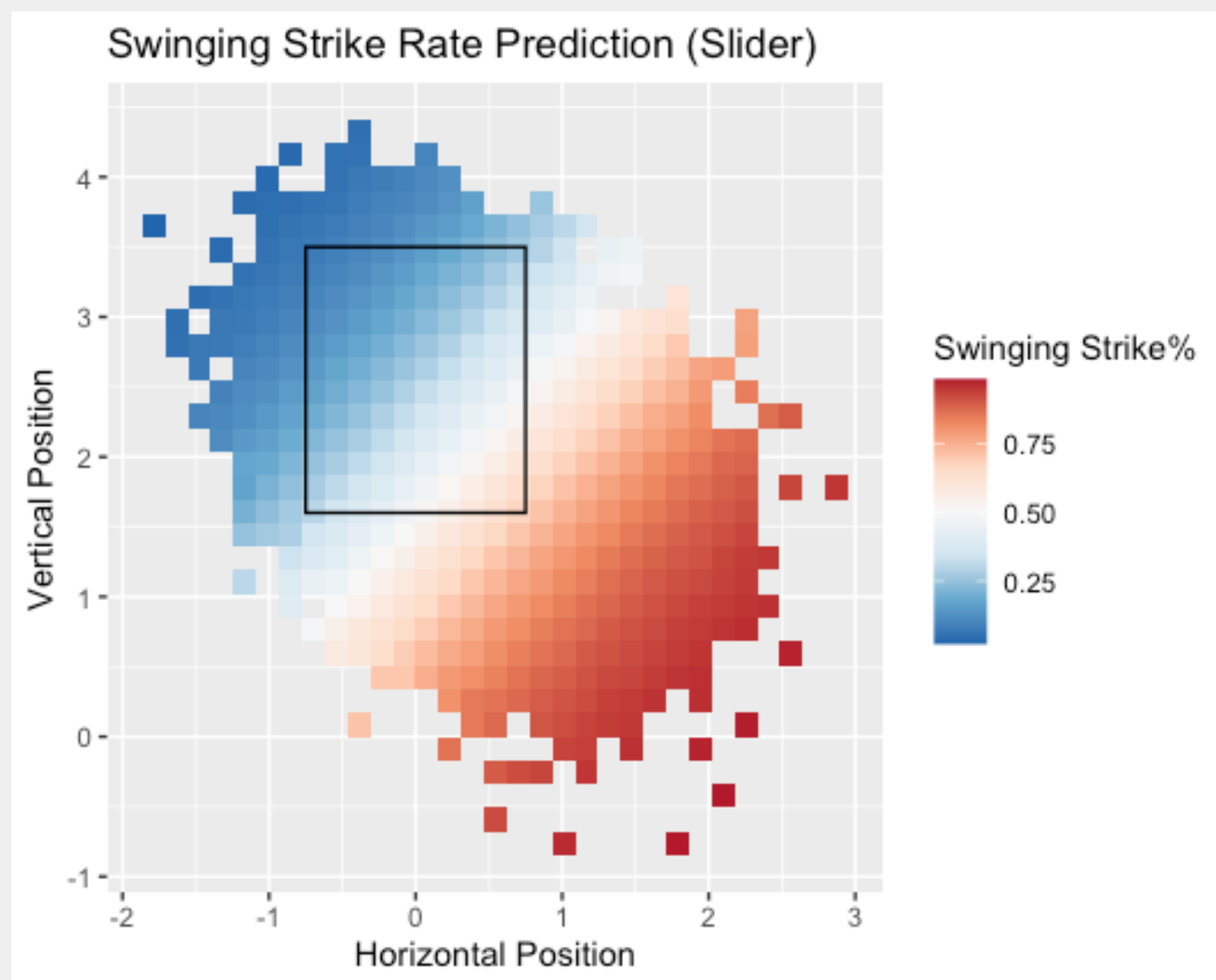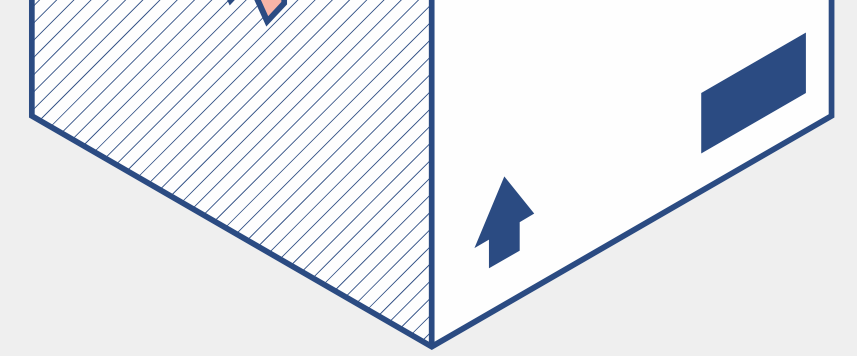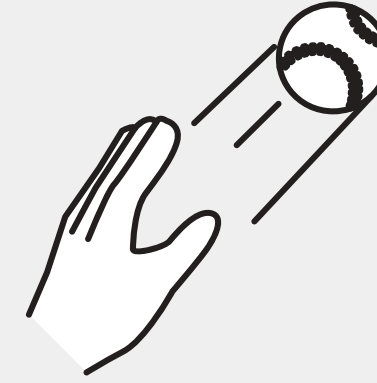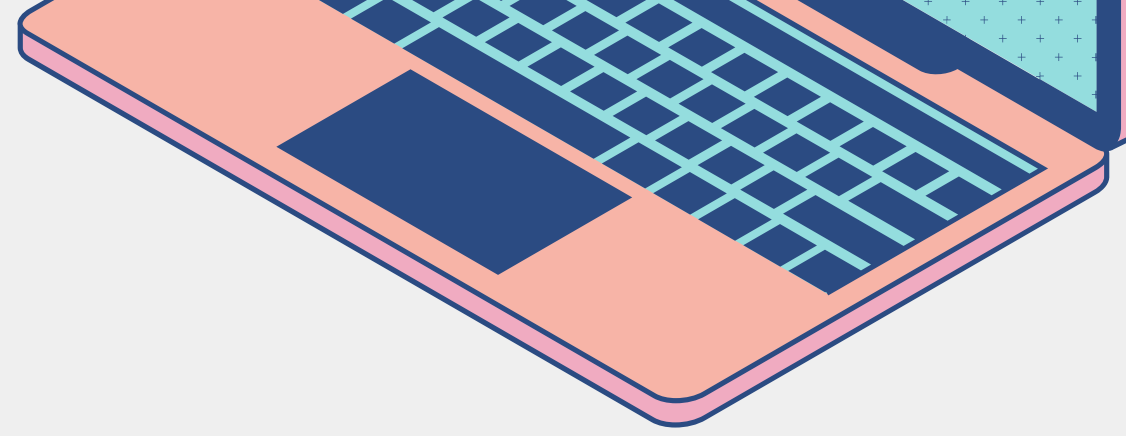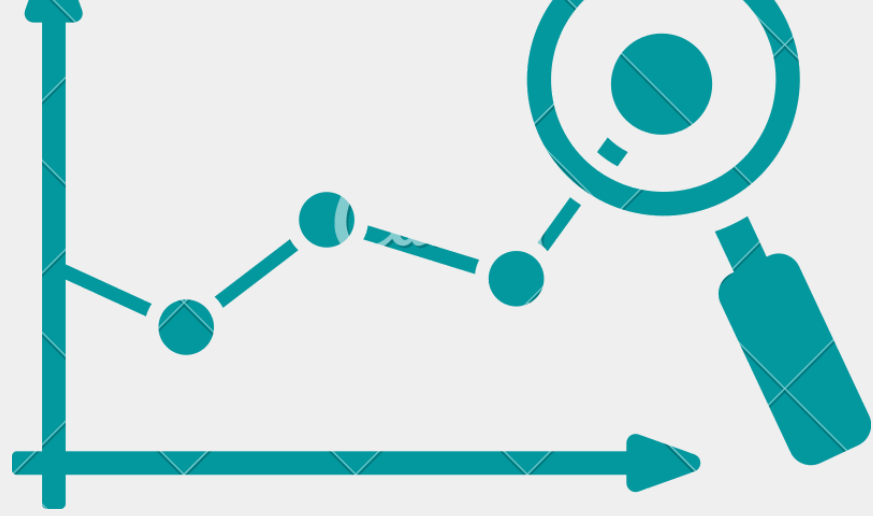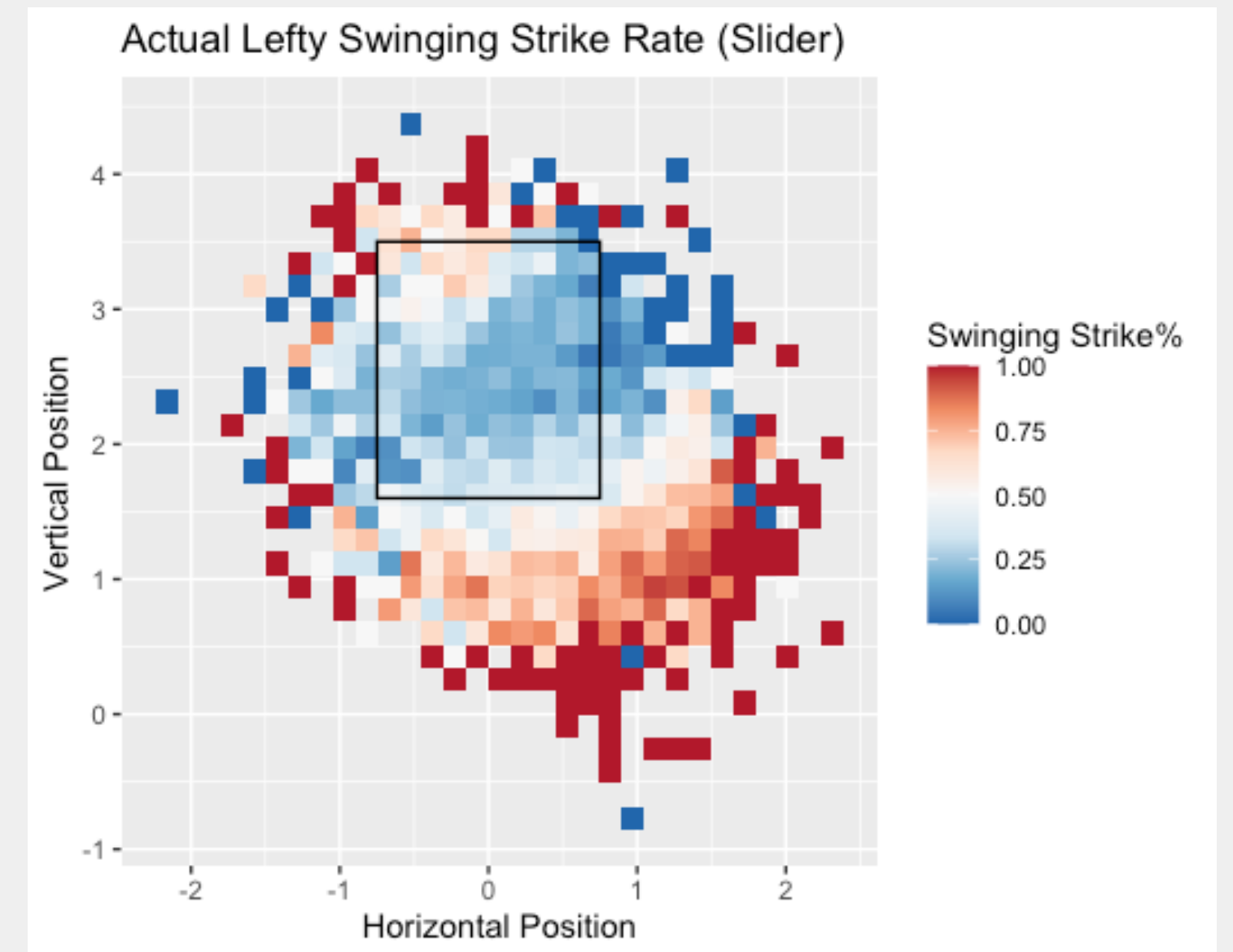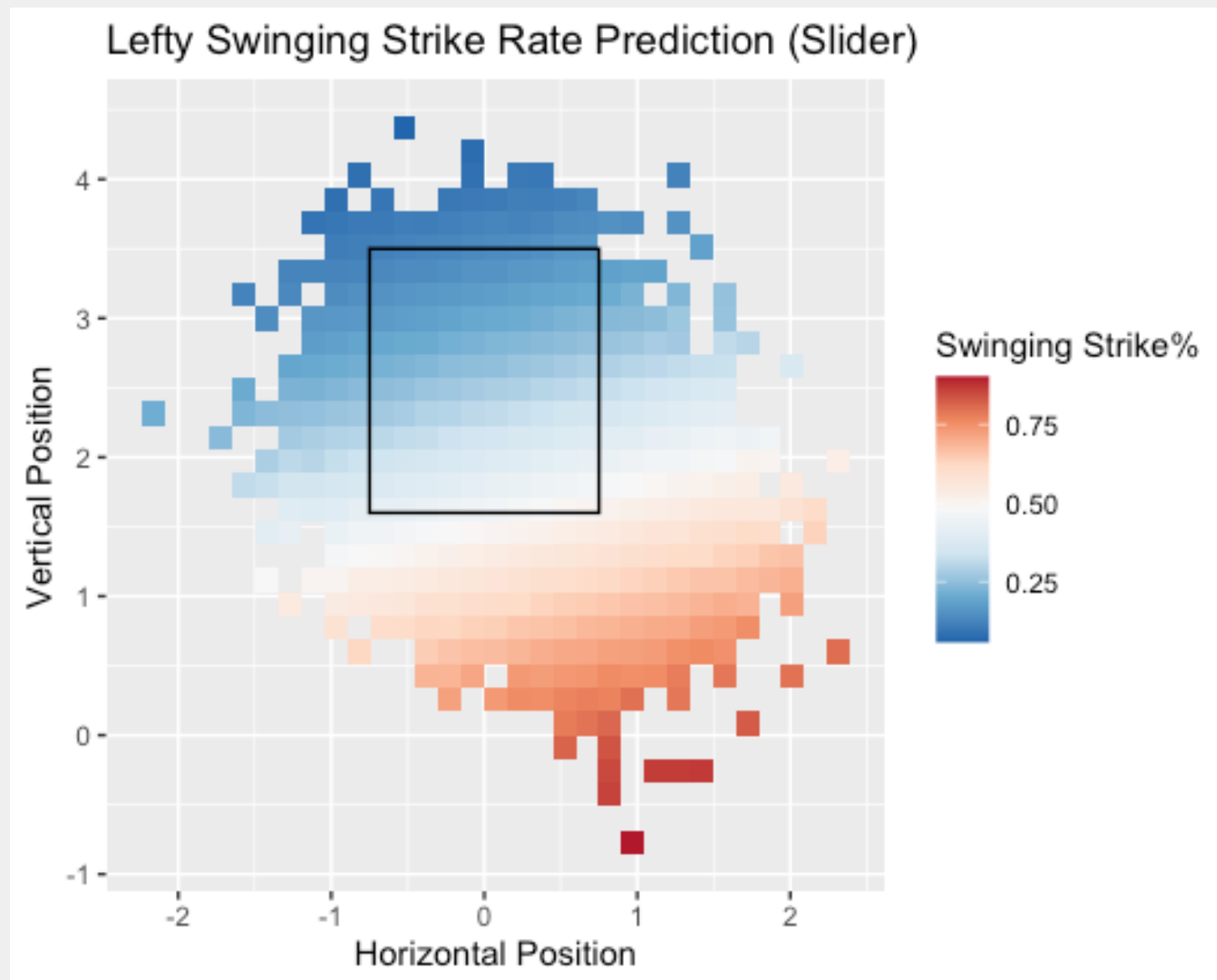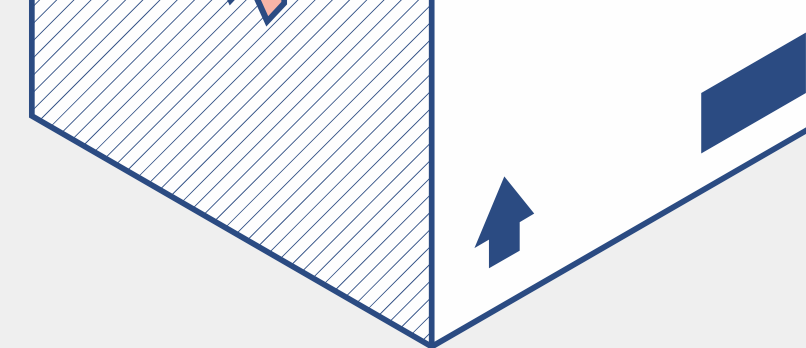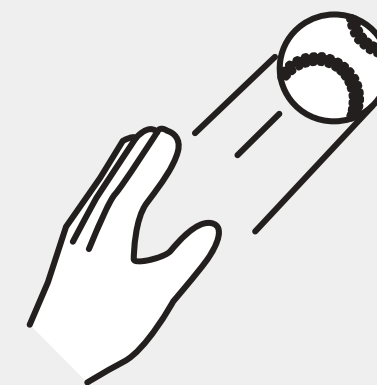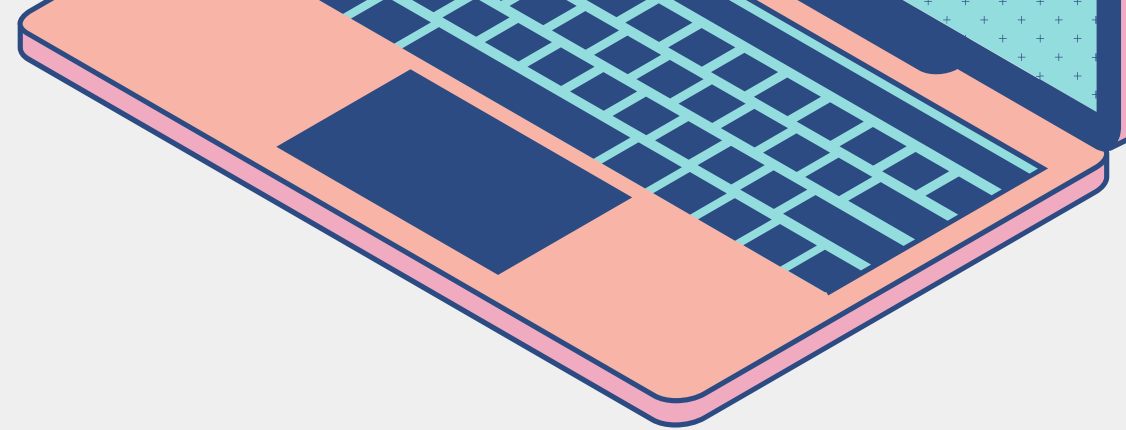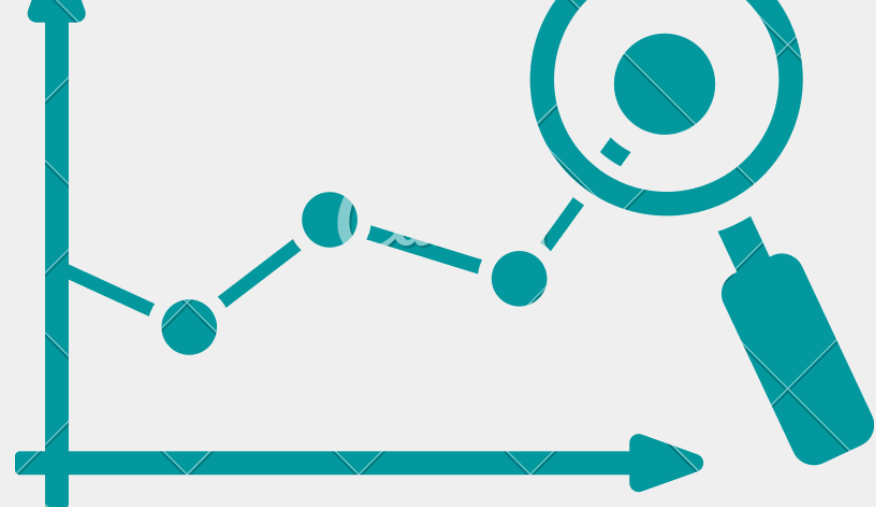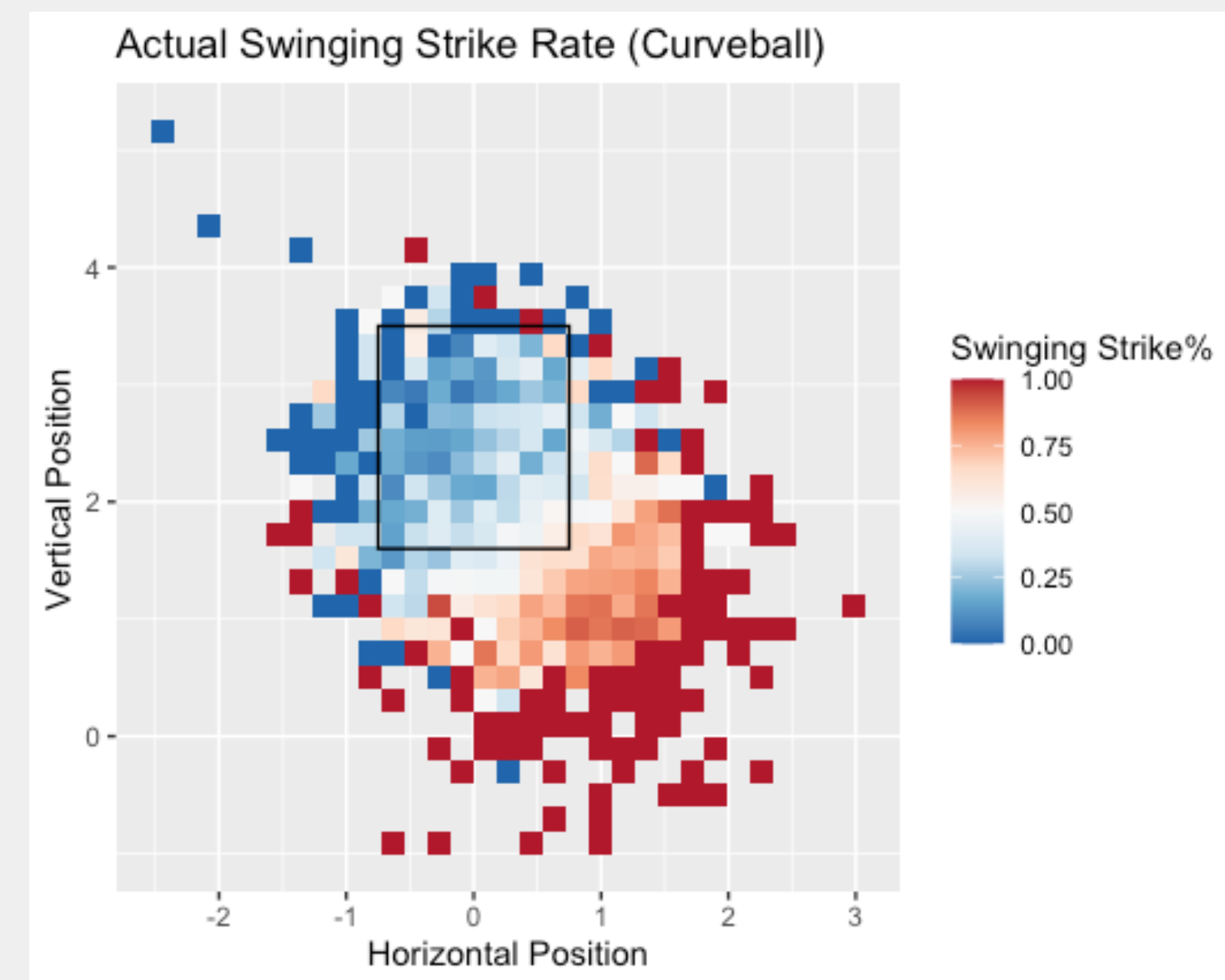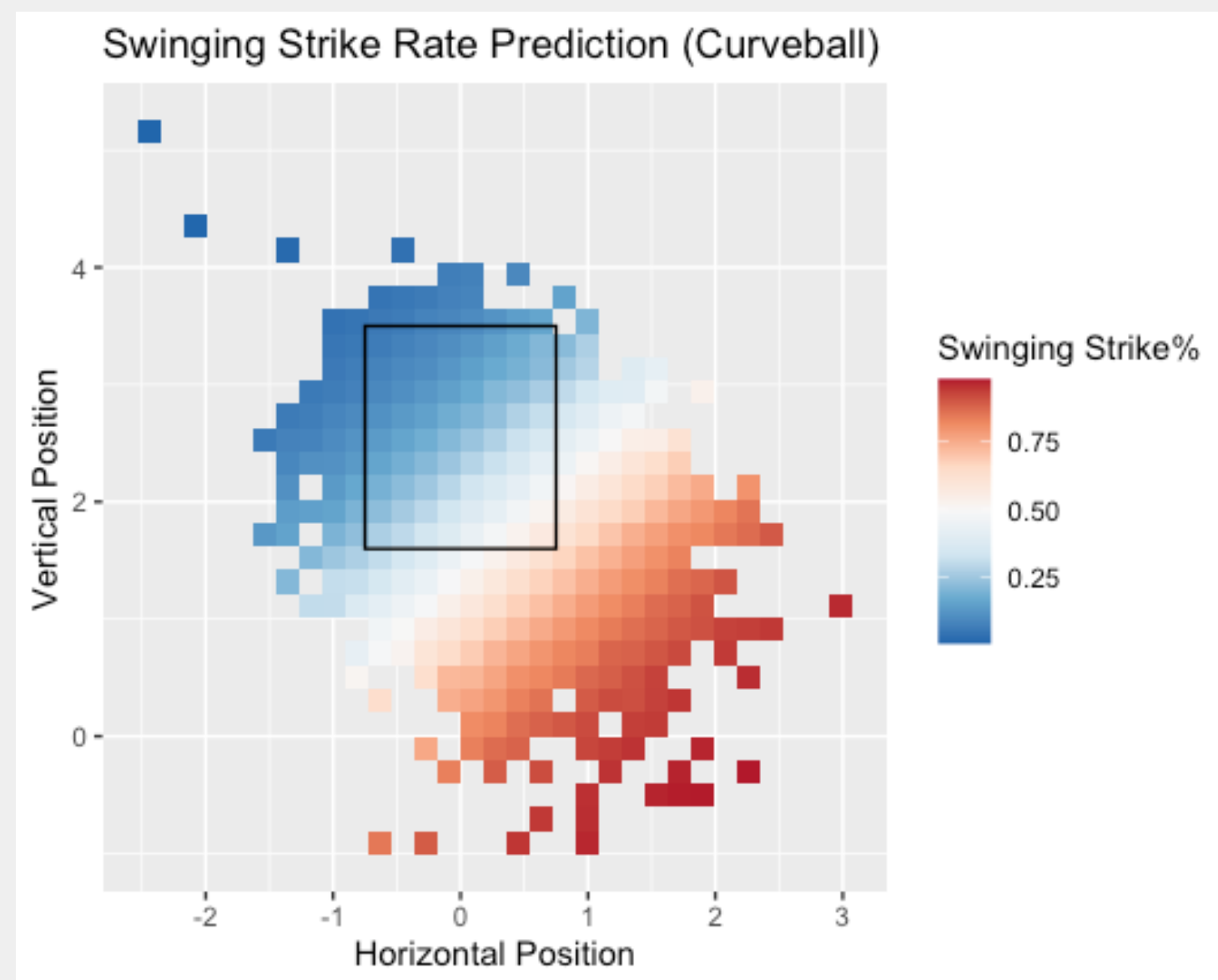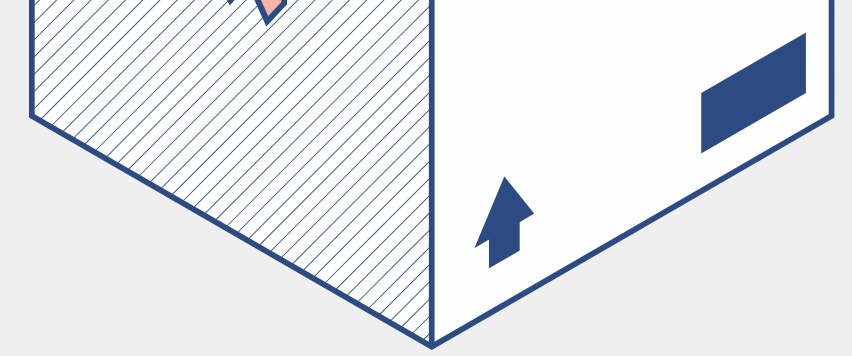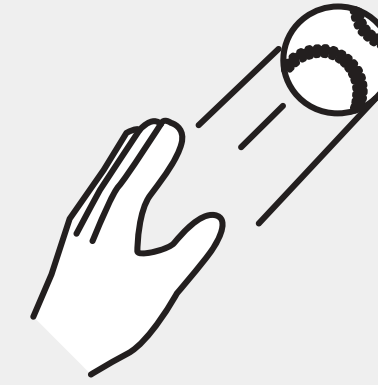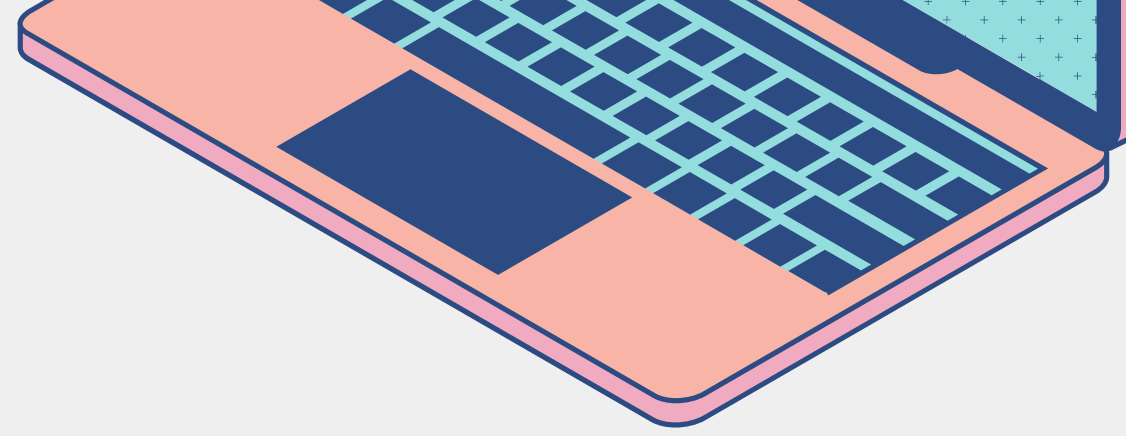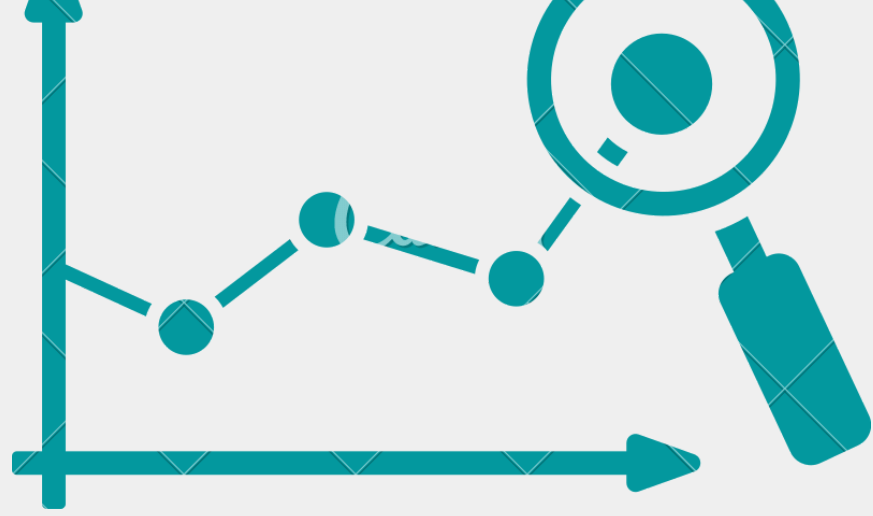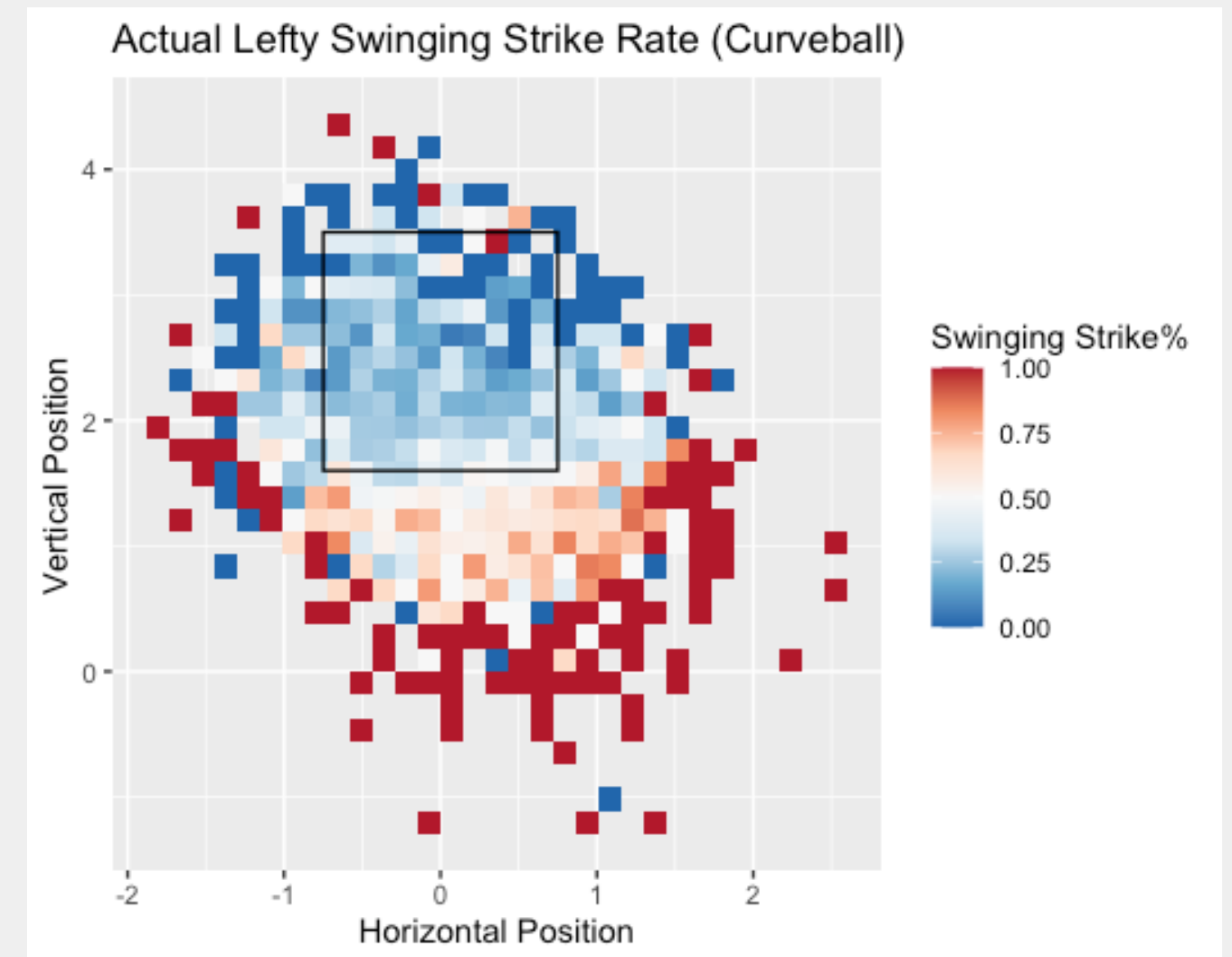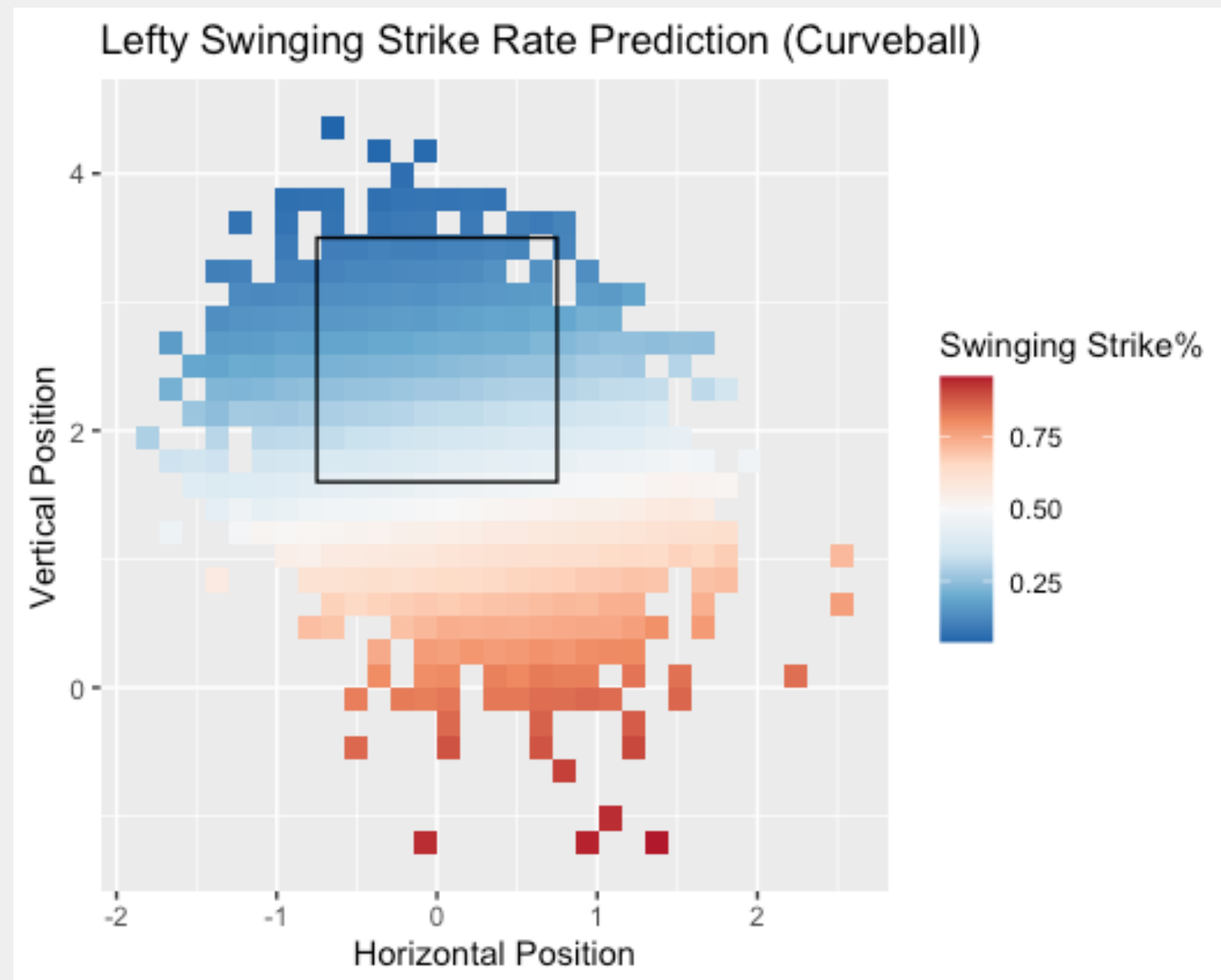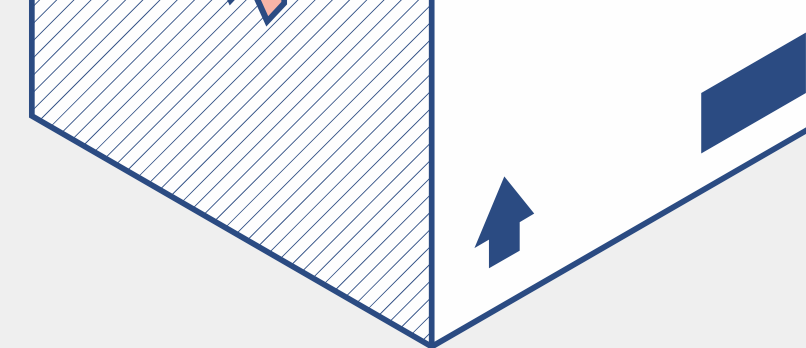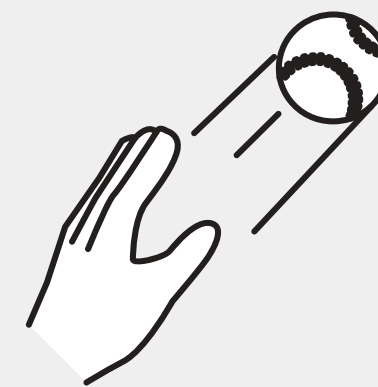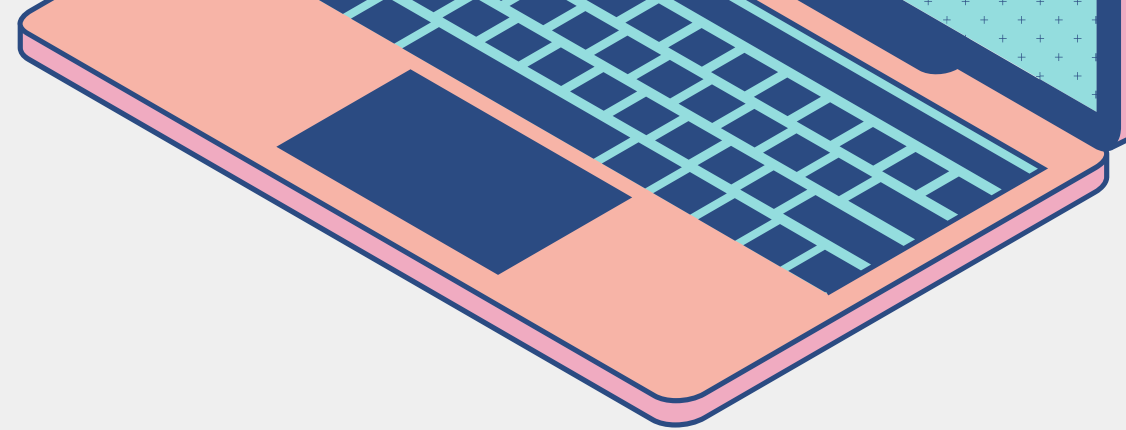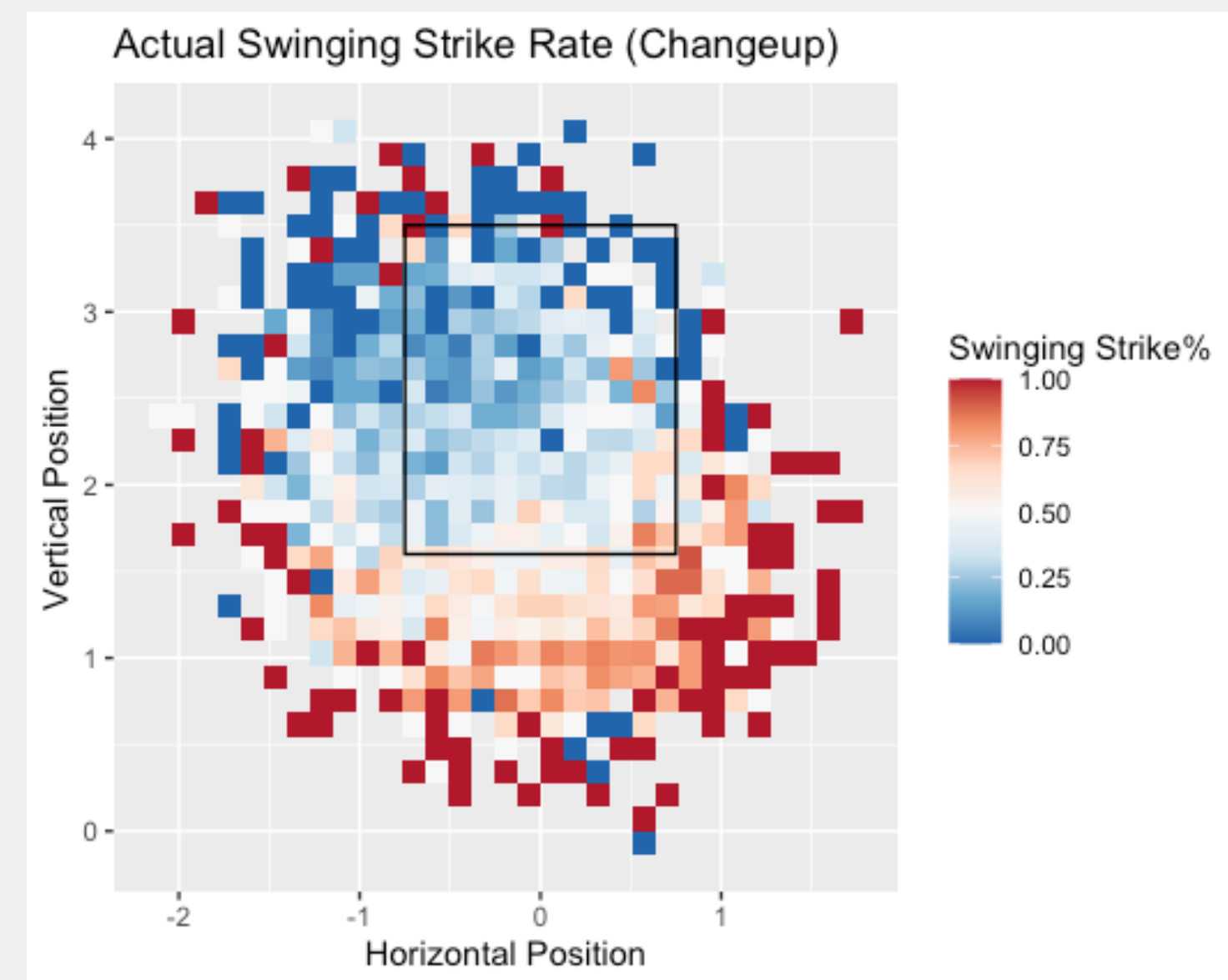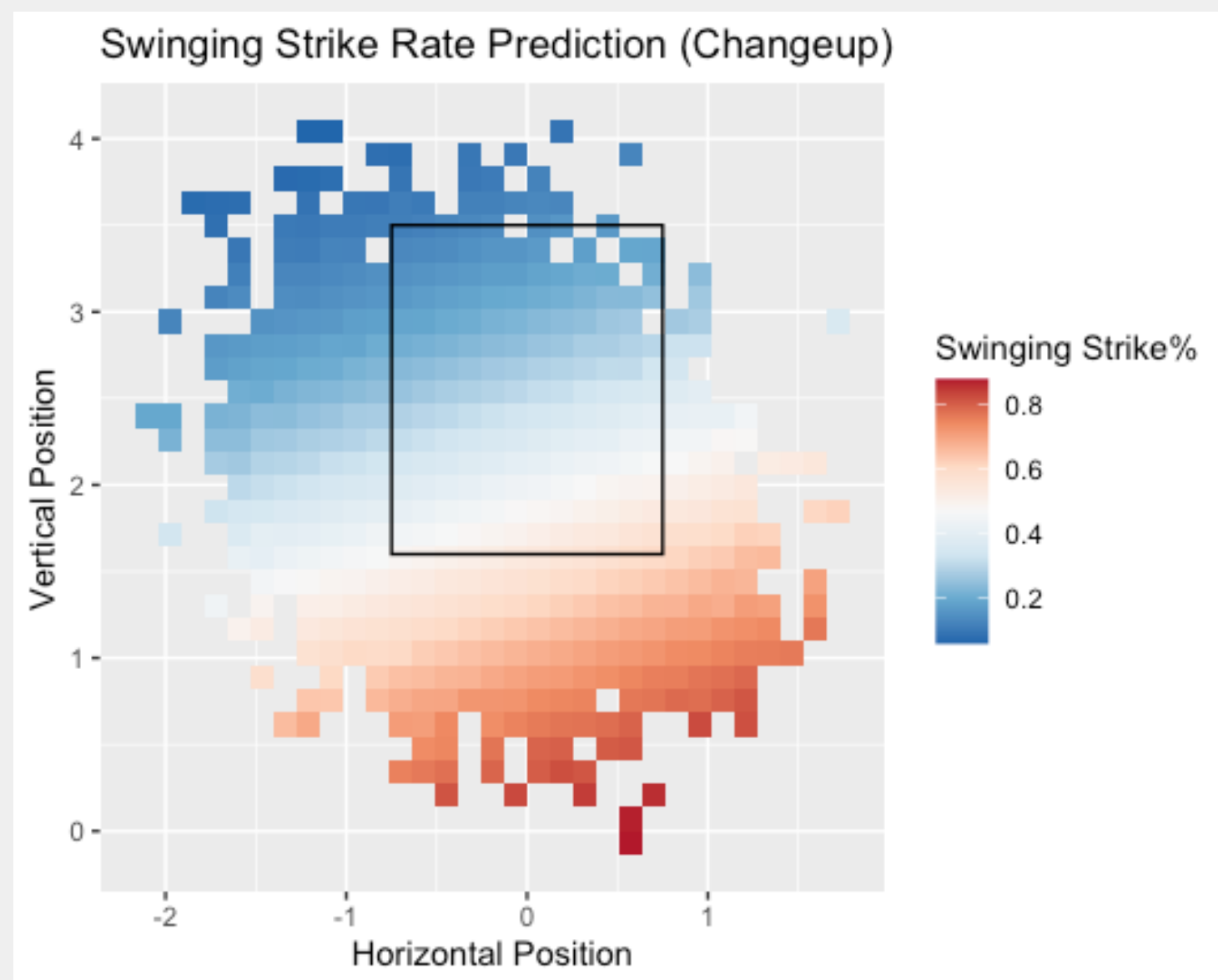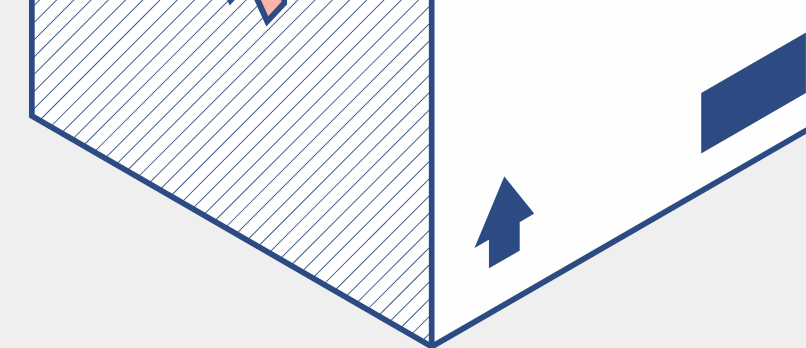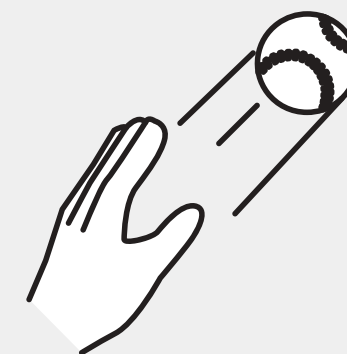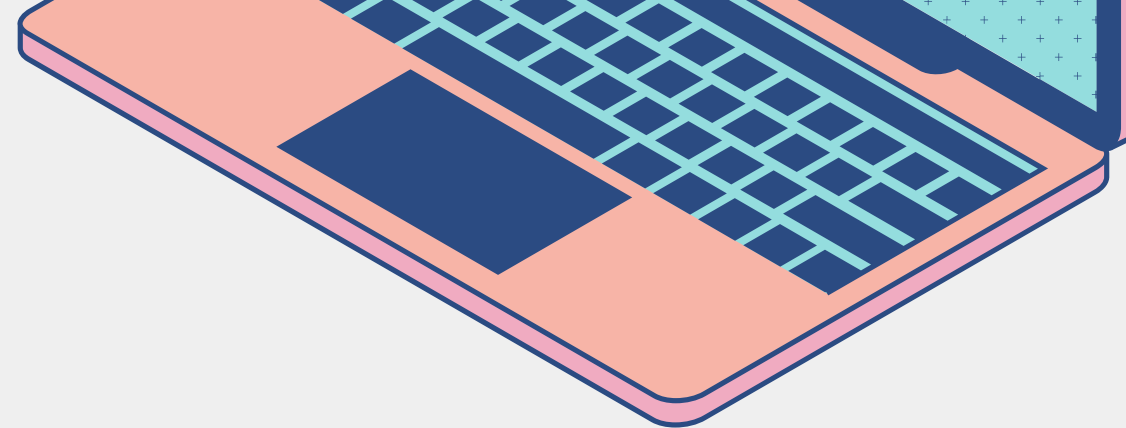
Changeup (Rvs.L)

Prediction

Actual
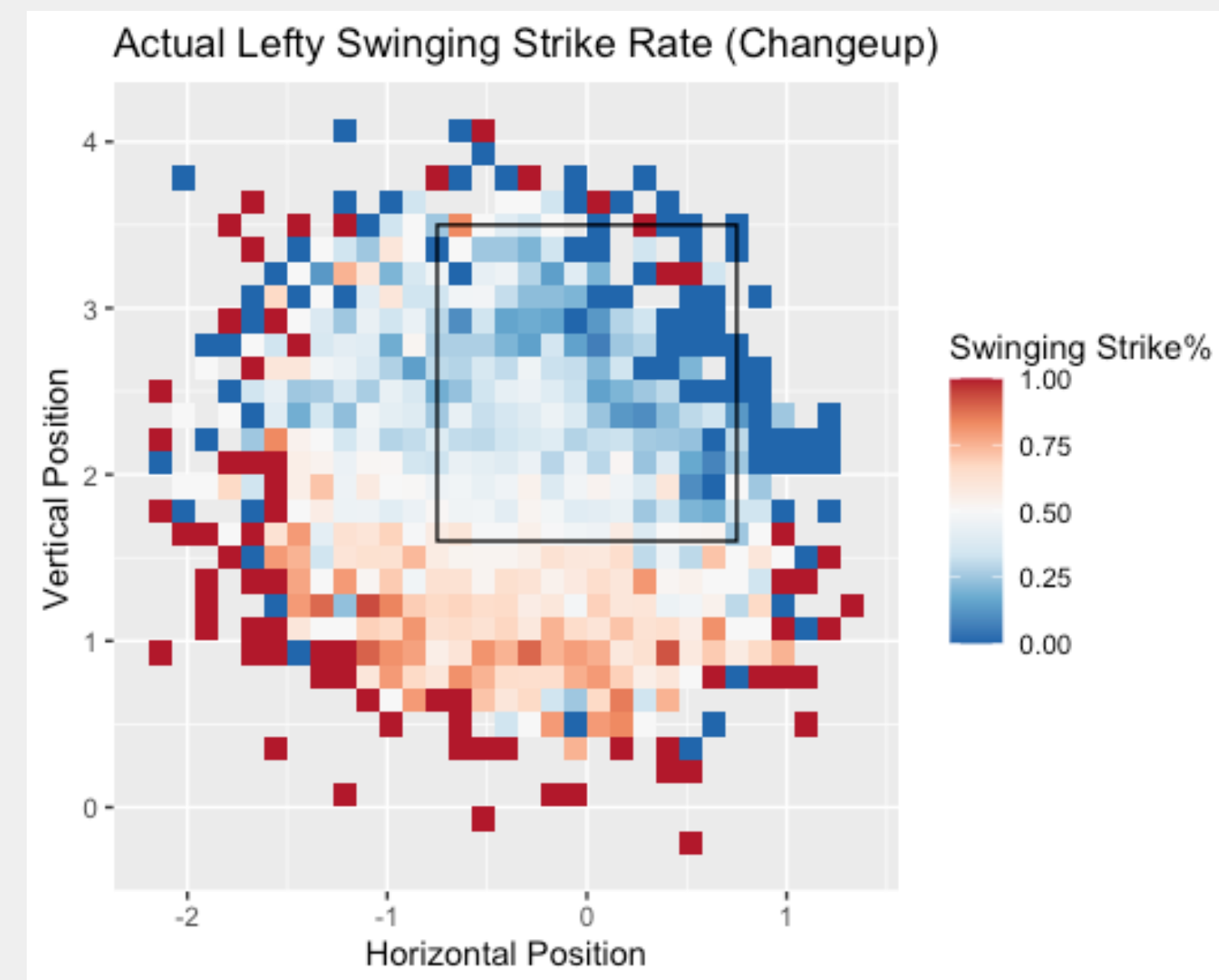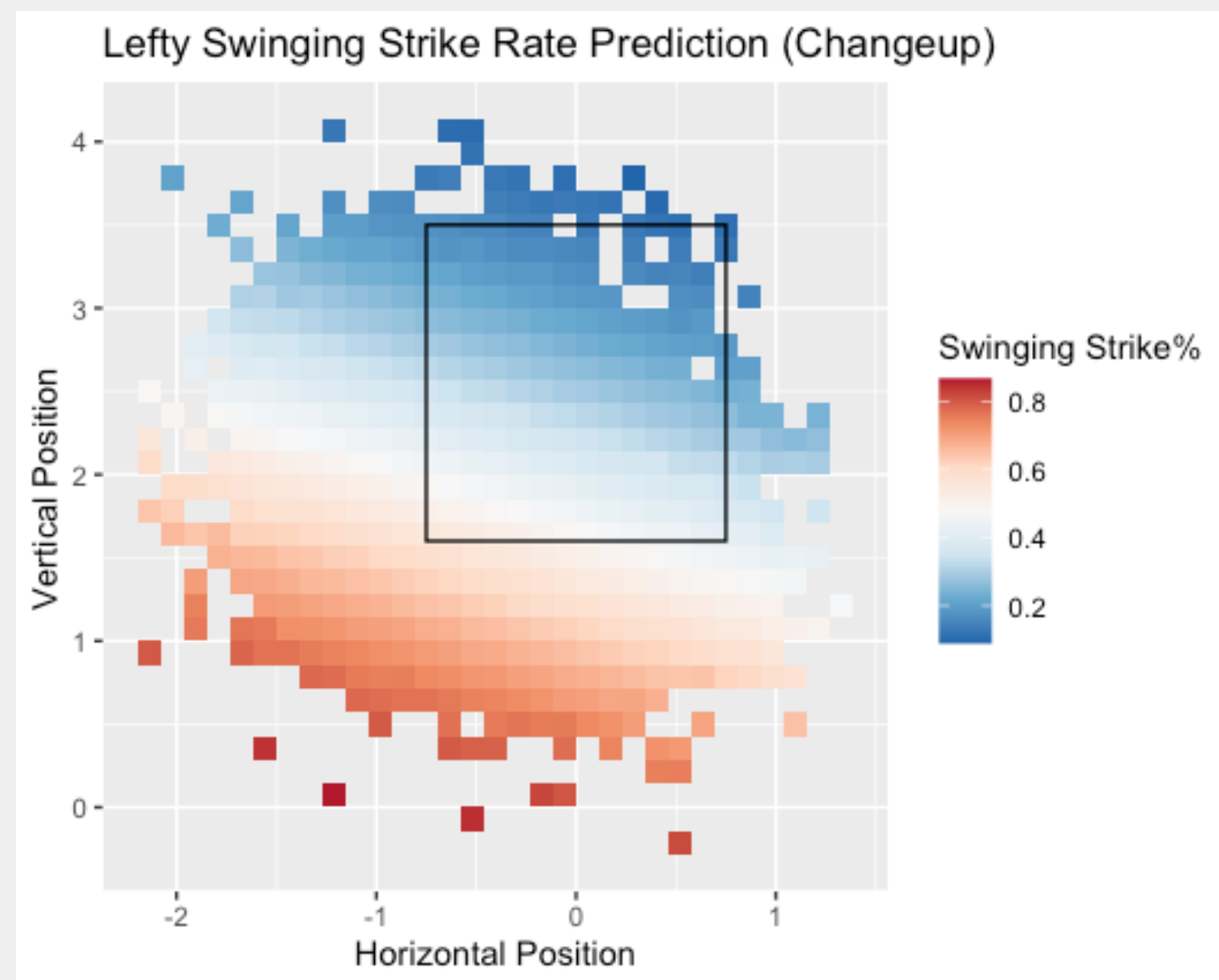
Possible Improvements

We could have improved our analysis by *analyzing high percentage locations of where batters don't swing to see where pitchers can be even more effective*. If we can identify pitch locations which have a low swing percentage and a high called strike percentage, we can suggest that these locations are where the pitcher can be most successful as he wouldn't have to worry about batters even swinging. *We could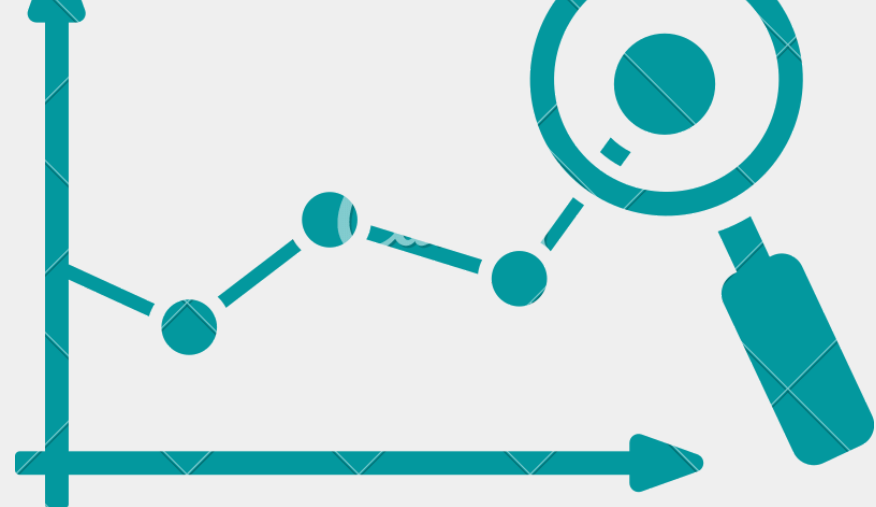 have also seen what types of plays (ground out, pop fly, etc.) certain pitches result in* to help pitchers know what to throw in certain situations. We can combine this with the research we have already done to *give pitchers insights into what pitches to throw and where to throw them in certain situations.*

Conclusion

What did we learn?

We learned that *under high leverage situations* (bases loaded no outs), *pitchers should refrain from overusing a changeup and even a fastball as they result in higher contact rates* than curveballs and sliders, helping bring men in.

The pitches that result in higher swinging strikeouts rates are the curveball and slider, in comparison to the other analysed pitches.

The final conclusion we made from our data is where the most effective locations are for pitchers to pitch to get a swinging strikeout and how these locations differ by pitch.

We also learned the locations that batters should swing to make contact.

Do you have
any questions?