Fang Analysis

May 31, 2023

1 Zachary Inn

[1]:
```
import warnings
warnings.filterwarnings('ignore')

import pandas as pd
import numpy as np
from plotnine import *

from sklearn.linear_model import LinearRegression # Linear Regression Model
from sklearn.preprocessing import StandardScaler #Z-score variables
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score␣
 ↪#model evaluation
from sklearn.model_selection import train_test_split # simple TT split cv
from heapq import nsmallest
```

#Import all datasets

[2]:
```
Sep = pd.read_csv("https://raw.githubusercontent.com/ZacharyInn/
 ↪WreckingYardData/main/CSVFiles/FangSeptember.csv")

Oct = pd.read_csv("https://raw.githubusercontent.com/ZacharyInn/
 ↪WreckingYardData/main/CSVFiles/FangOctober.csv")

Nov = pd.read_csv("https://raw.githubusercontent.com/ZacharyInn/
 ↪WreckingYardData/main/CSVFiles/FangNovember.csv")

Jan = pd.read_csv("https://raw.githubusercontent.com/ZacharyInn/
 ↪WreckingYardData/main/CSVFiles/FangJanuary.csv")

Feb = pd.read_csv("https://raw.githubusercontent.com/ZacharyInn/
 ↪WreckingYardData/main/CSVFiles/FangFebruary.csv")

Mar = pd.read_csv("https://raw.githubusercontent.com/ZacharyInn/
 ↪WreckingYardData/main/CSVFiles/FangMarch.csv")
```

```
Apr = pd.read_csv("https://raw.githubusercontent.com/ZacharyInn/
 ↪WreckingYardData/main/CSVFiles/FangApril.csv")

May = pd.read_csv("https://raw.githubusercontent.com/ZacharyInn/
 ↪WreckingYardData/main/CSVFiles/FangMay.csv")
```

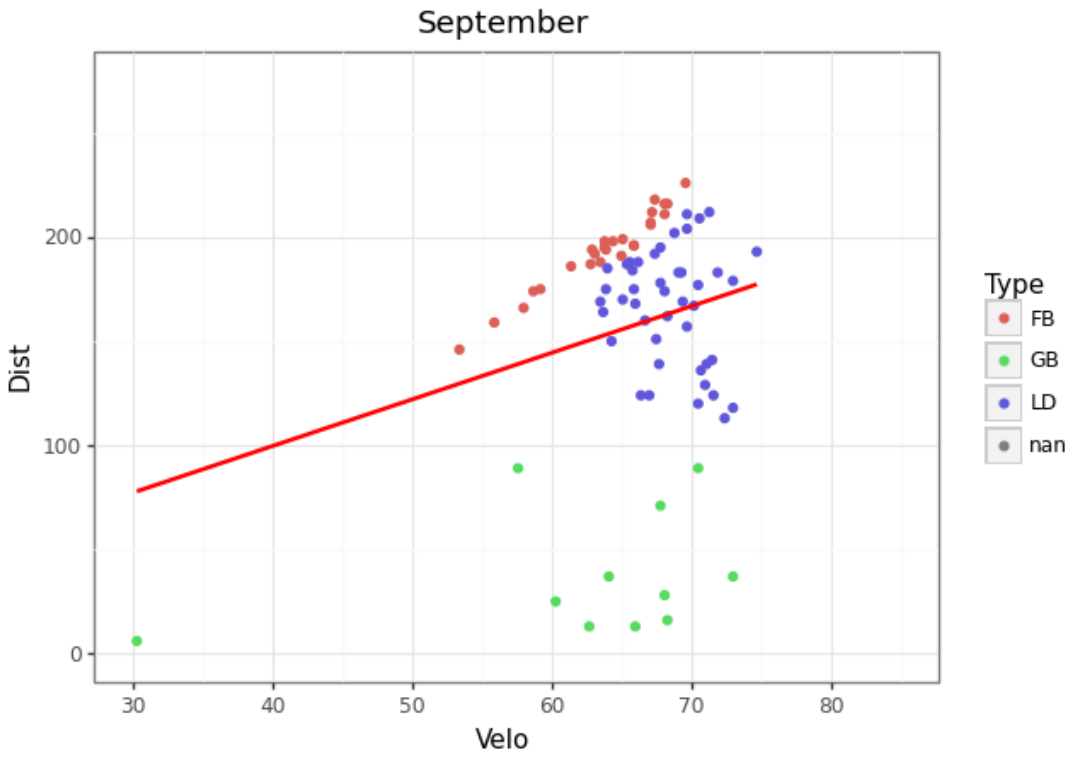
#Introduction

The concept of these reports are usually born out of a question posed to myself. More often than not, there is an unexplainable discepancy between a player's performance in practice and in game. Oftentimes, there are mechanical issues that are pointed out by the Wrecking Yard coaching staff that they are curious about. For this project, no question was asked; rather I would have to find ones to ask myself and look for some sort of meaning within the data. These datasets are stats taken from HitTrax during lessons of a player named Fang. There is one data set taken from each month from September 2022 to May 2023 with the exception of December.

This report was created with the intent to find a flaw in this 13 year old player's game. I was given no prior knowledge or video of this player and cycled through data until I found something that was worth expanding upon. This report will take you through my thought process and my findings. I had no leads when starting this project so I just started where I began in my last report, graphing based on exit velocity and distance.

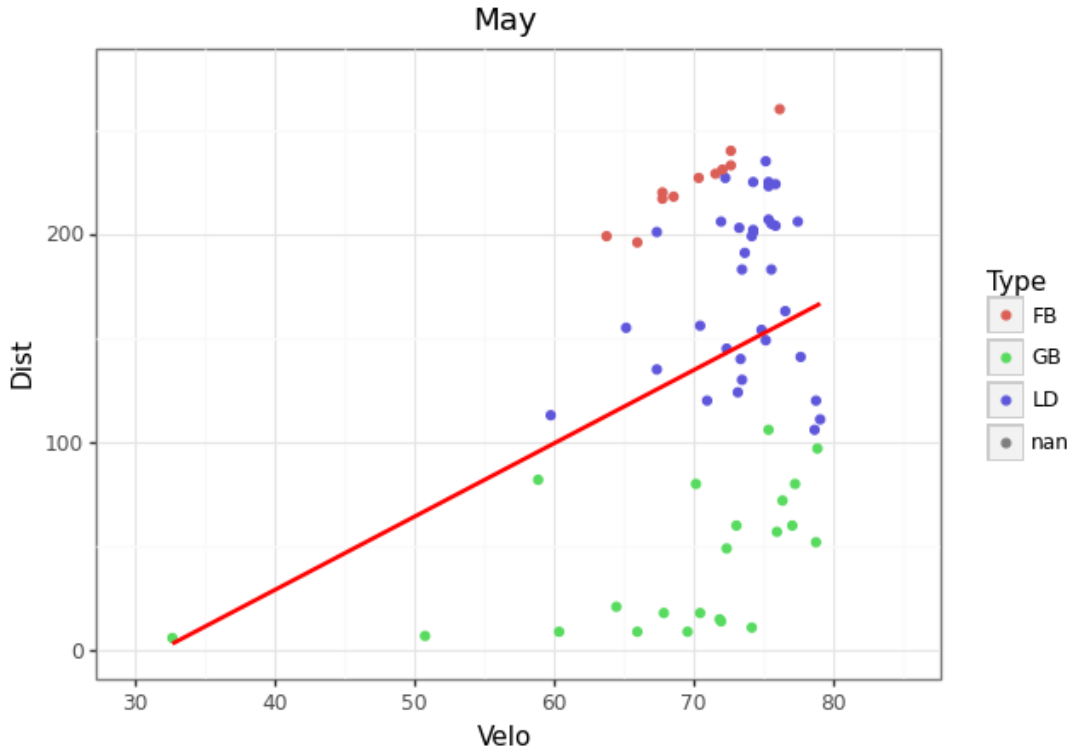
#Graphs comparing the exit velocity and distance of each pitch, colored by each outcome

[3]:
```
ggplot(Sep, aes(x = "Velo", y = "Dist", color='Type')) + geom_point() +␣
 ↪theme_bw() + geom_smooth(method='lm',  se=False, color = "red") + xlim(30,␣
 ↪85) + ylim(0, 275) + ggtitle("September")
```



[3]:
```
<ggplot: (8755056917877)>
```

[4]:
```
ggplot(May, aes(x = "Velo", y = "Dist", color='Type')) + geom_point() +␣
↪theme_bw() + geom_smooth(method='lm',  se=False, color = "red") + xlim(30,␣
↪85) + ylim(0, 275) + ggtitle("May")
```



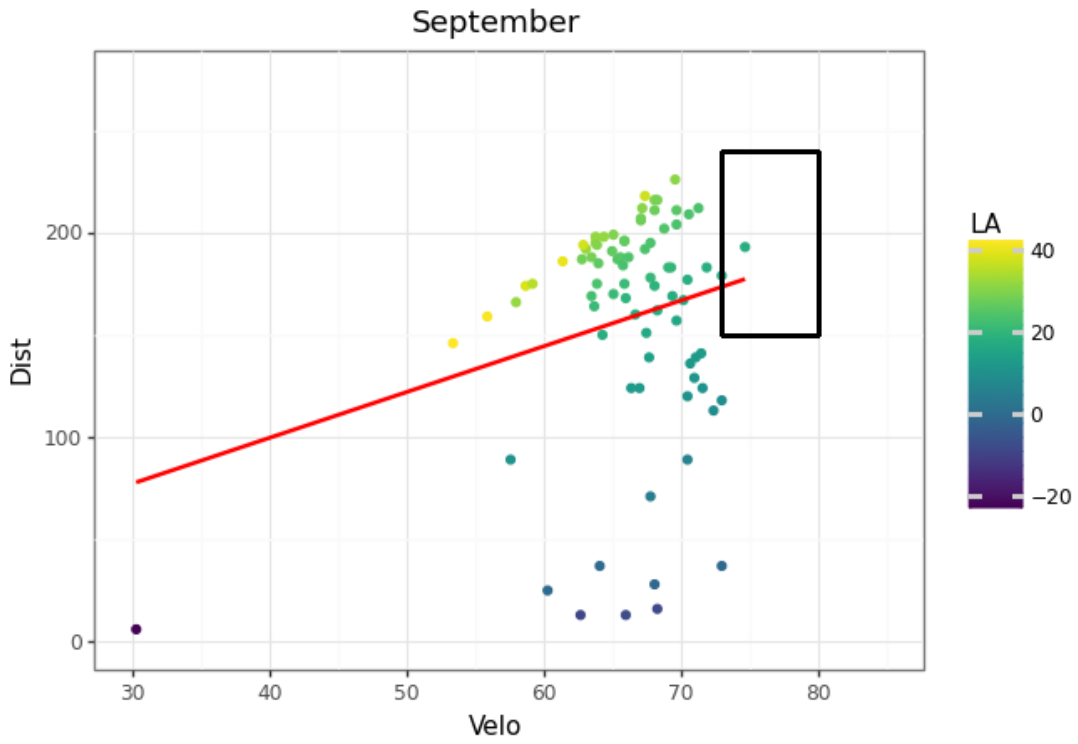
[4]:
```
<ggplot: (8755054651540)>
```

Fang's Results

Fang's max exit velocity and average exit velocity increased significantly from September 2022 to May 2023. In both graphs, there was a positive linear relationship between exit velocity and distance. There wasn't much to make of these graphs so I decided to look at the time inbetween these months to see of there was anything worth noting.

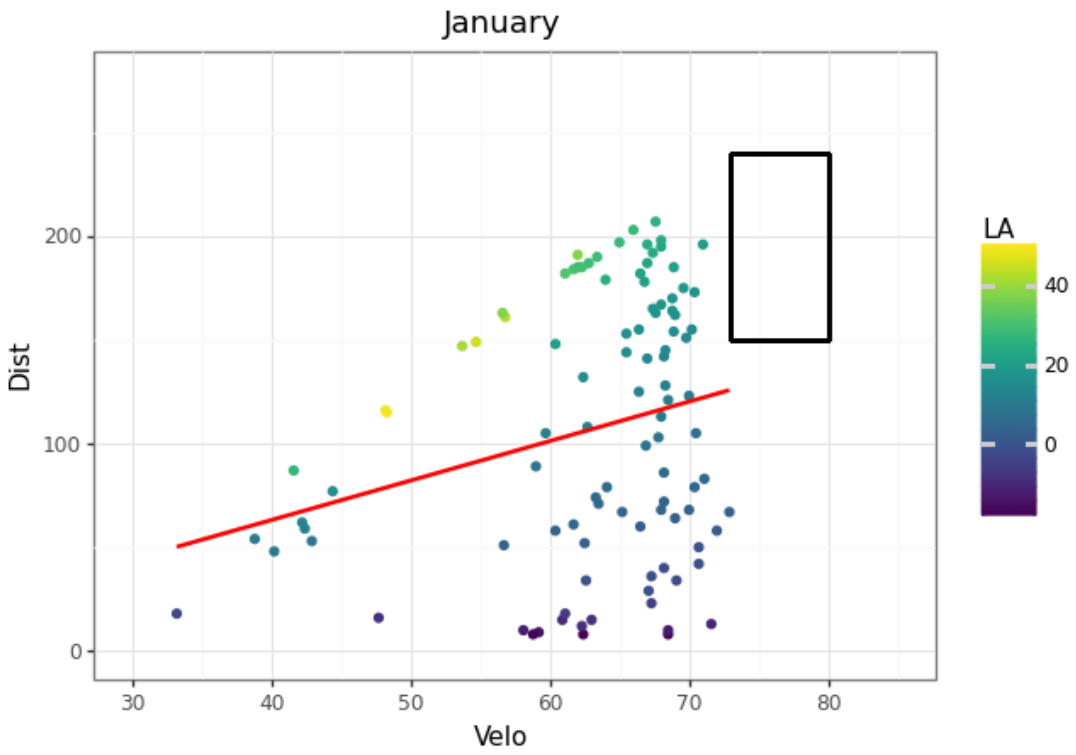
#Graph comparing Fang's exit velocity to distance, colored by launch angle

[5]:
```
ggplot(Sep, aes(x = "Velo", y = "Dist", color='LA')) + geom_point() +␣
↪theme_bw() + geom_smooth(method='lm',  se=False, color = "red") + xlim(30,␣
↪85) + ylim(0, 275) + ggtitle("September") + geom_rect(xmin = 73, xmax = 80, ␣
↪ ymin = 150, ymax = 240, alpha=0, color = "black")
```



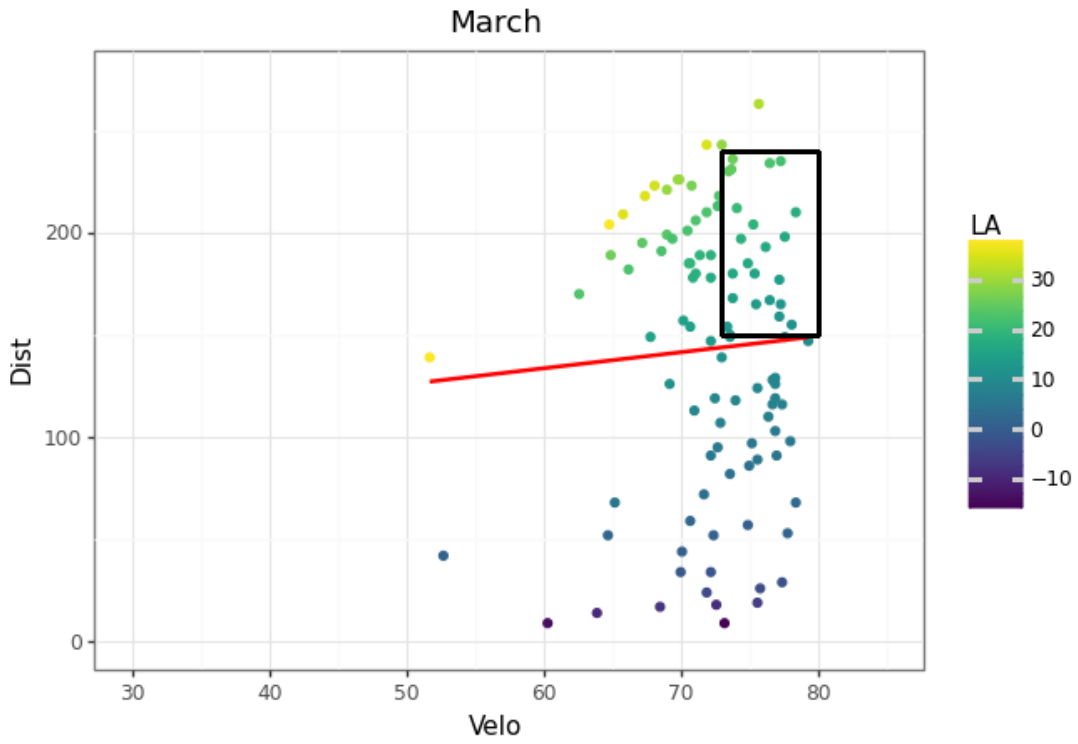
[5]: <ggplot: (8755054611426)>

[6]:
```
ggplot(Jan, aes(x = "Velo", y = "Dist", color='LA')) + geom_point() +␣
↪theme_bw() + geom_smooth(method='lm',  se=False, color = "red") + xlim(30,␣
↪85) + ylim(0, 275) + ggtitle("January") + geom_rect(xmin = 73, xmax = 80,  ␣
↪ymin = 150, ymax = 240, alpha=0, color = "black")
```



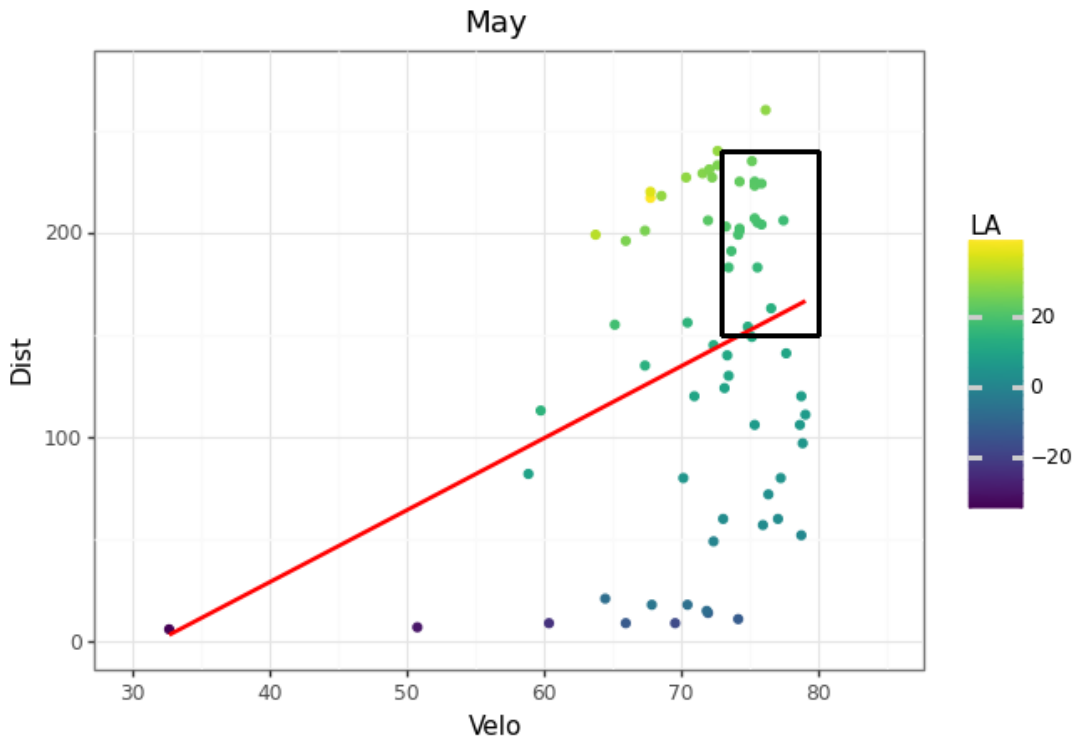
[6]: <ggplot: (8755054505849)>

[7]:
```
ggplot(Mar, aes(x = "Velo", y = "Dist", color='LA')) + geom_point() +␣
↪theme_bw() + geom_smooth(method='lm',  se=False, color = "red") + xlim(30,␣
↪85) + ylim(0, 275) + ggtitle("March") + geom_rect(xmin = 73, xmax = 80,  ␣
↪ymin = 150, ymax = 240, alpha=0, color = "black")
```



[7]: <ggplot: (8755054596187)>

[8]:
```
ggplot(May, aes(x = "Velo", y = "Dist", color='LA')) + geom_point() +␣
↪theme_bw() + geom_smooth(method='lm',  se=False, color = "red") + xlim(30,␣
↪85) + ylim(0, 275) + ggtitle("May") + geom_rect(xmin = 73, xmax = 80,   ymin␣
↪= 150, ymax = 240, alpha=0, color = "black")
```



[8]: <ggplot: (8755054380258)>

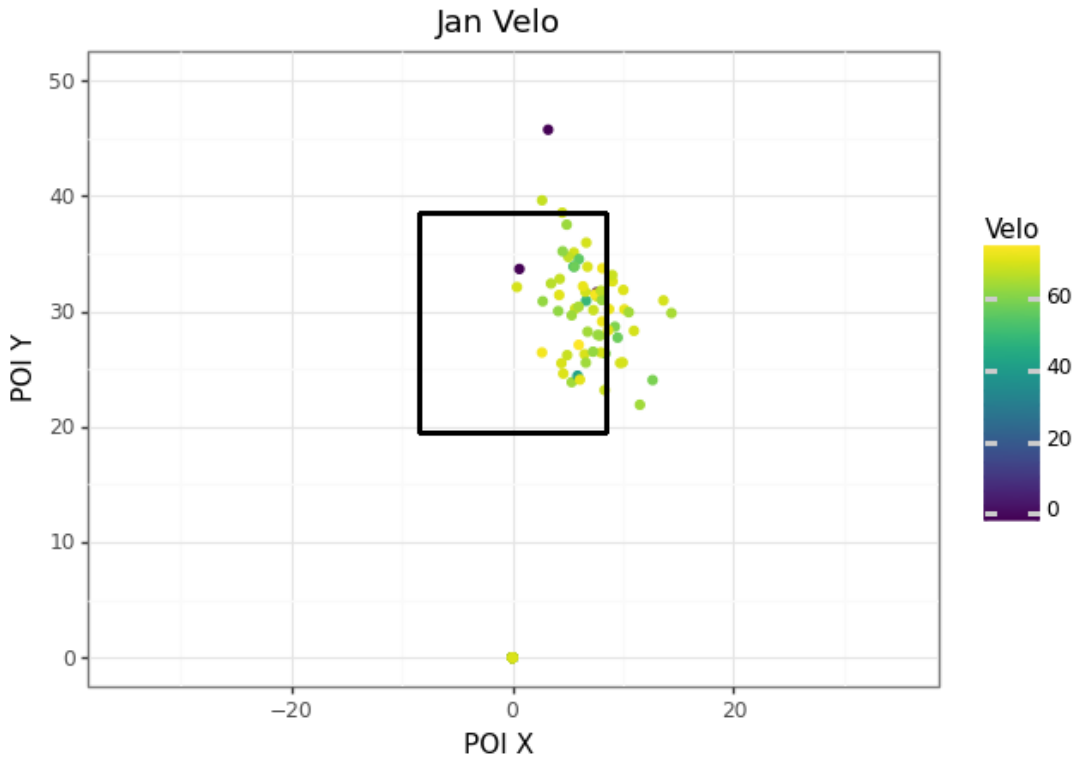
2 Results

In these graphs, I changed the color to represent the launch angle and created a rectangle in the graph which I call the "Goldilocks Zone," AKA the idea batted ball outcomes. Most of the batted balls that fall into that zone will end up as a solid base hit. I also only included the graphs that were of interest to myself, even though I did graph them all. Through this and looking at the data in Excel, I found that Fang's ideal launch angle is anywhere between 18 and 25 degrees.

As you can see from the September graph, Fang only had 1 batted ball inside of the Goldilocks Zone and as the months went on, there were more and more batted balls that ended up in the zone except for January. March and May were Fang's best months in terms of batted balls in the Goldilocks Zone. This outlier in January is where my curiousity peaked and I tried to find an explanation for why there was such as stark contrast between January and all of the other months.

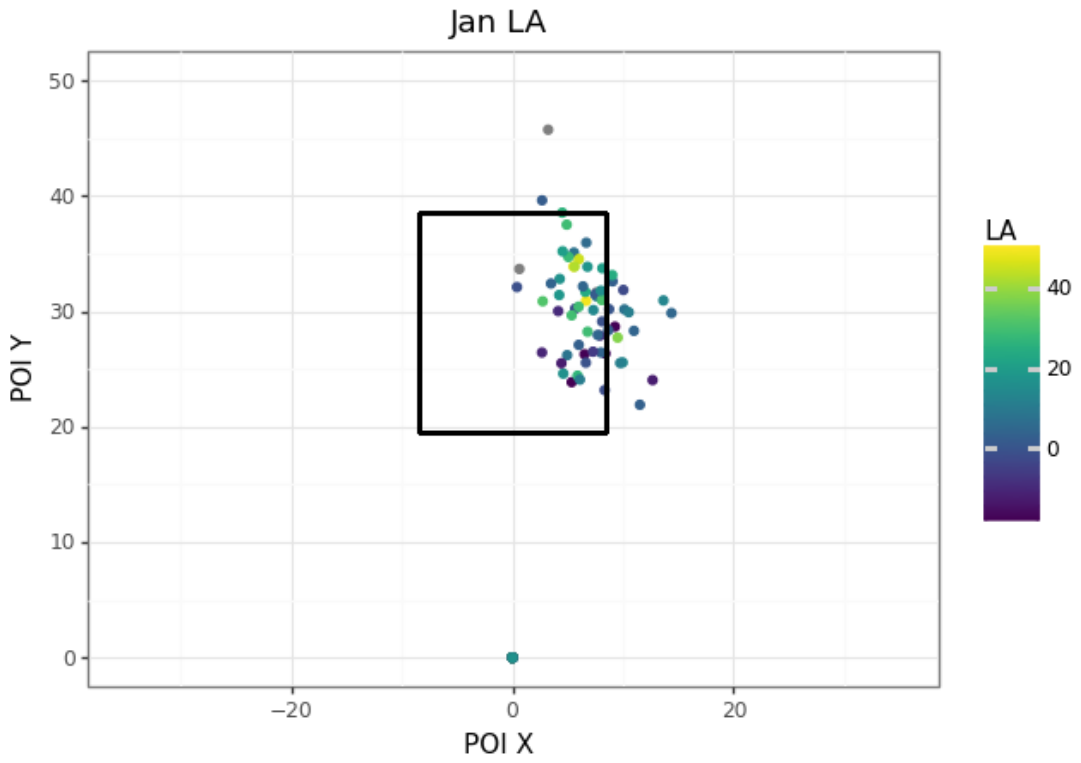
#Graph comparing the X and Y position of each pitch, colored by exit velocity

[9]:
```
ggplot(Jan, aes(x = "POI X", y = "POI Y", color='Velo')) + geom_point() +␣
↪theme_bw() + ggtitle("Jan Velo") + geom_rect(xmin = -8.5, xmax = 8.5,   ymin␣
↪= 19.6, ymax = 38.5, alpha=0 , color = "black") + xlim(-35, 35) + ylim(0, 50)
```



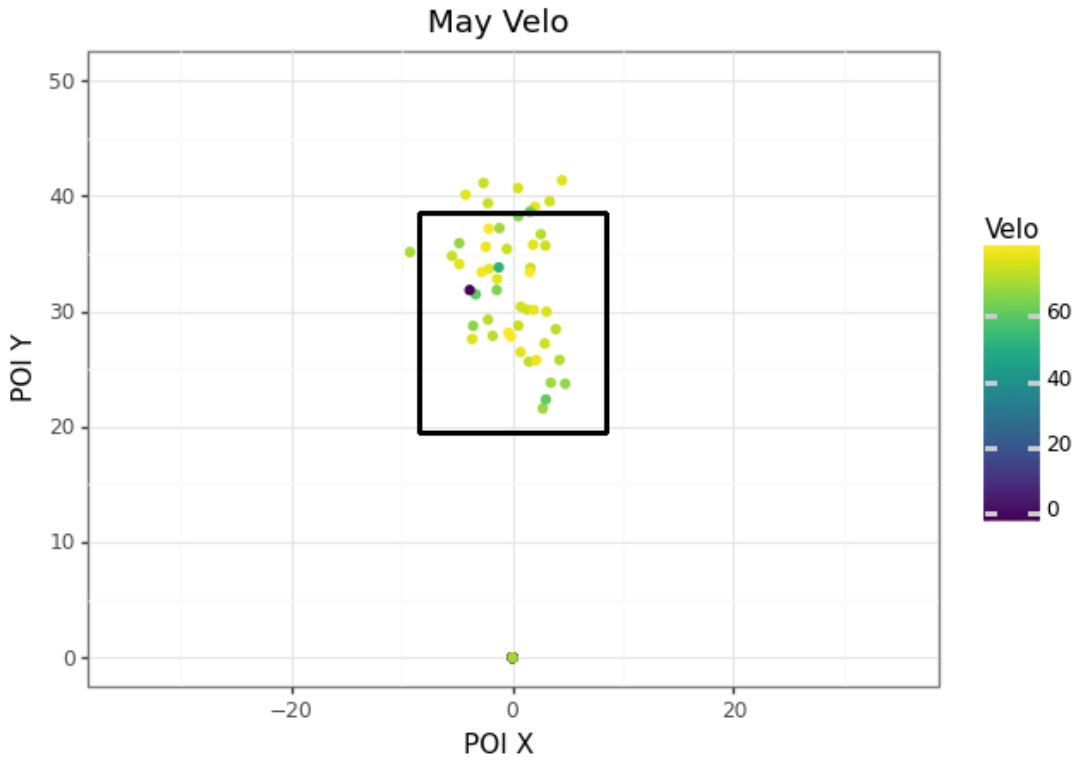
[9]: <ggplot: (8755054470286)>

[10]:
```
ggplot(Jan, aes(x = "POI X", y = "POI Y", color='LA')) + geom_point() +␣
↪theme_bw() + ggtitle("Jan LA") + geom_rect(xmin = -8.5, xmax = 8.5,   ymin =␣
↪19.6, ymax = 38.5, alpha=0 , color = "black") + xlim(-35, 35) + ylim(0, 50)
```



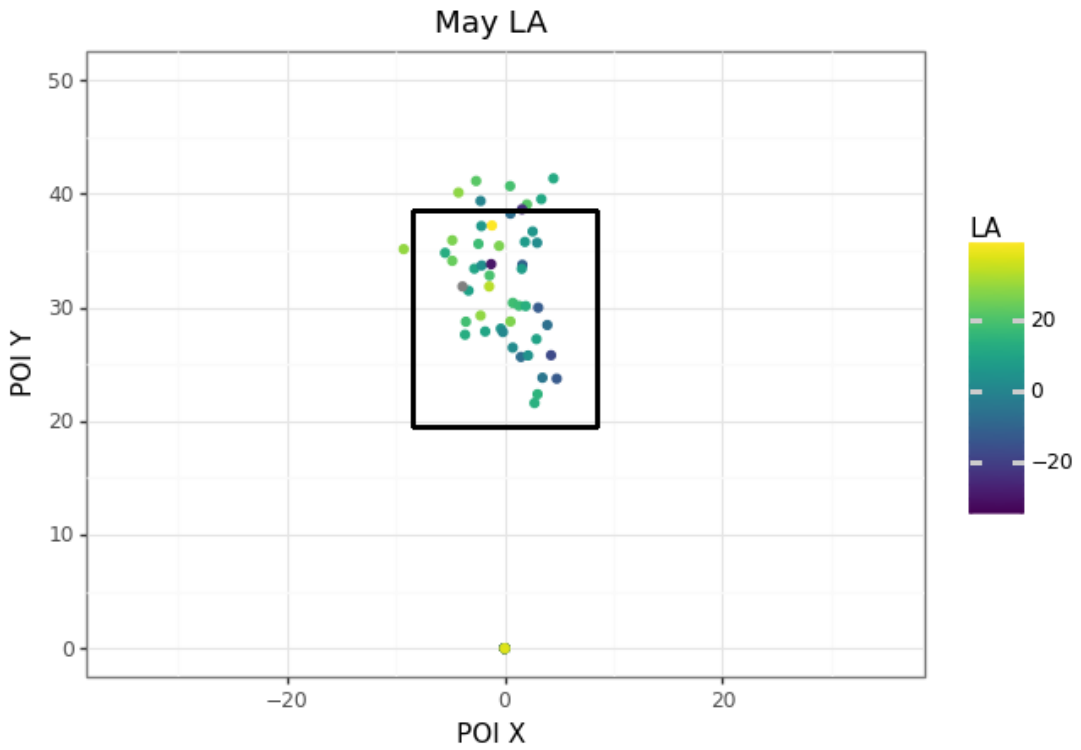
[10]: <ggplot: (8755054240669)>

[11]:
```
ggplot(May, aes(x = "POI X", y = "POI Y", color='Velo')) + geom_point() +␣
↪theme_bw() + ggtitle("May Velo") + geom_rect(xmin = -8.5, xmax = 8.5,   ymin␣
↪= 19.6, ymax = 38.5, alpha=0 , color = "black") + xlim(-35, 35) + ylim(0, 50)
```



[11]: <ggplot: (8755054169479)>

[12]:
```
ggplot(May, aes(x = "POI X", y = "POI Y", color='LA')) + geom_point() +␣
↪theme_bw() + ggtitle("May LA") + geom_rect(xmin = -8.5, xmax = 8.5,   ymin =␣
↪19.6, ymax = 38.5, alpha=0 , color = "black") + xlim(-35, 35) + ylim(0, 50)
```



[12]: <ggplot: (8755054265014)>

3 Strike Zone results

The conclusion that I came to as to why January was such an outlier is that a majority of the pitches in the January lesson were middle-away (in the middle to the outer half of the strike zone). I plotted the position of every pitch and created a strike zone for reference. Immediately, it was obvious that the pitches in the January lesson were away, leading me to color by velocity, then by launch angle.

In the January lesson, the exit velocities of the batted balls were sligtly lowered which is to be expected on pitches away. It is typical that hitters pull the ball with more authority than they do going to the opposite field. Since this seemed so normal, there had to be some factor that was hindering Fang's results, which I found in his launch angles. The batted balls in January had a signifiantly lower launch angle on average than the ones in May which explain absence of batted balls in the Goldilocks Zone.

4 Struggles on Away Pitches

Jan # of batted balls within ideal launch angle: 12 balls, 11.5% of pitches

Average Exit Velocity of batted balls within ideal launch: 67.1 mph

March # of batted balls within ideal launch angle: 22 balls, 21% of pitches

Average Exit Velocity of batted balls within ideal launch: 71.4 mph

May # of batted balls within ideal launch angle: 14 balls, 20.6% of pitches

Average Exit Velocity of batted balls within ideal launch: 74.5 mph

5 Conclusion

Fang struggles when it comes to hitting the outside pitch. After finishing a rough draft of this data, I was able to meet with Fang to explain my suspicions, to which he confessed that he feels that he does struggle with the away pitch. It is not only that he is struggling to hit the away pitch as he does hit the pitch sufficiently, but he needs to increase the launch angle of these batted balls as most of his hardest in terms of exit velocities would be catchable by the second baseman in a game scenario. I was able to conclude this by recognizing a dip in expected exit velocity numbers within a certain dataset and looking at different factors to try to explain this dip. I was able to find that the dip was caused by a high increase in pitches being thrown to the outside part of the plate and to confirm this dip I had to look at exit velocities and launch angles of each pitch to determine that a majority of hard contact didn't have a high enough launch angle to produce any ideal results.

[ ]:
```
# doesn't show this cells output when downloading PDF
!pip install gwpy &> /dev/null

# installing necessary files
!apt-get install texlive texlive-xetex texlive-latex-extra pandoc
!sudo apt-get update
!sudo apt-get install texlive-xetex texlive-fonts-recommended␣
 ↪texlive-plain-generic

# installing pypandoc
!pip install pypandoc

# connecting your google drive
from google.colab import drive
drive.mount('/content/drive')

# copying your file over. Change "Class6-Completed.ipynb" to whatever your file␣
 ↪is called (see top of notebook)
!cp "drive/My Drive/Colab Notebooks/Fang Analysis.ipynb" ./

# Again, replace "Class6-Completed.ipynb" to whatever your file is called (see␣
 ↪top of notebook)
!jupyter nbconvert --to PDF "Fang Analysis.ipynb"
```

```
Reading package lists… Done
Building dependency tree
Reading state information… Done
pandoc is already the newest version (2.5-3build2).
pandoc set to manually installed.
```