RIT

Mathematical Modeling II

Qualifying Exam Practicum

Mario L. Gutierrez Abed

March 28^{th}, 2021

R·I·T

Changes in the MLB

Summary

In this report we analyze the potential impact of introducing bigger bases in the MLB. In particular, we investigate whether the upgrade in base size from 15” to 18” will increase the number of bases successfully stolen, as well as the number of infield hits from groundballs and bunts. Our results indicate encouraging signs for hitters and baserunners, while not giving them an unfair advantage over the defense. The size upgrade is rather conservative, and so is the impact shown in our analysis. Nevertheless, the difference is not negligible; we can expect that, as more studies are undertaken and more data becomes available, there will almost certainly be more steal attempts and bunt attempts in the league. After all, we are in an era where managers do pay attention to the analytics and make in-game decisions based on data and scouting reports.

Introduction

In the modern era of the MLB almost every team relies on the “long ball” to win games; the more old school speed game has recently taken a backseat to the power game. Teams in the MLB now tend to stack up on sluggers that mostly swing for the fences, while the Tony Gwynn’s of today’s baseball are very few (if not nonexistent). As a result, managers tend to hold the baserunners in place unless, perhaps, a hitter with very good contact is at the plate. Partly in an effort to encourage managers to take more chances on greenlighting baserunners and focus more on putting the ball in play by creating more bunt plays and situational groundball hitting, the MLB is seeking to limit defensive shifts as well as introduce bigger bases. A modest base size bump from 15” to 18” should incentivize managers to take more chances with his baserunners/hitters, while not providing the latter with an unfair advantage over the defense. In this report we consider two case studies: base stealing from first to second base, and groundballs/bunts (with varying bat exit velocities) to third base. Although simple in nature, the results of our study should be indicative of the impact of introducing 18” bases in the league.

Methodology-Base Stealing

There are many factors that need to be considered when attempting to steal a base; e.g., i) how much contact and/or power the hitter currently at the plate has; ii) how good are the baserunner’s reflexes (i.e., how much of an aggressive lead the baserunner takes from reading the pitcher’s pickoff/windup moves); iii) how good is the arm of the catcher behind the plate; iv) how long does the pitcher currently on the mound takes to deliver the ball to the plate, etc. For instance, Noah Syndergaard from the NY Mets has a notoriously slow windup (average 1.65 seconds; sometimes even longer (Fig. 1)), while the rest of the league takes an average of 1.4 seconds ([NYT17]). In our investigation we choose to ignore factors that are hard to quantify (the baserunner’s reflexes, for instance), and we shall focus instead on the pitcher’s delivery to home plate, the catcher’s *pop time* (i.e., the time it takes a catcher to get the ball out of his glove and to second base on a steal attempt), and the baserunner’s speed. All of our results on steal attempts will be based on data from these three categories.

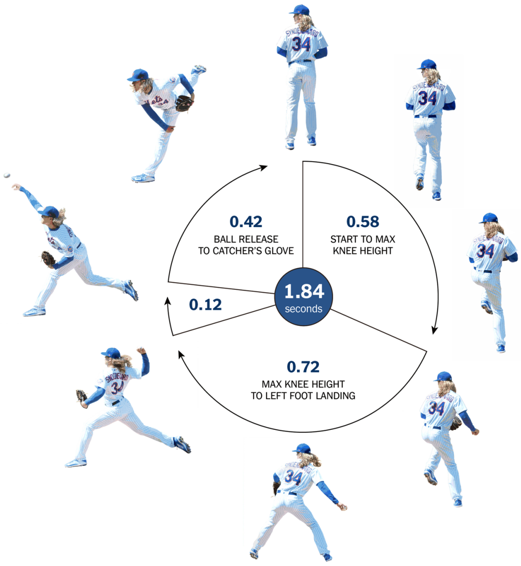

Figure 1: Noah Syndergaard's slow windup (Image from [NYT17]).

While it is very common to hear that, in baseball, bases are 90 ft apart, this is not entirely accurate for our purposes since we are considering the distance from edge to edge of the bases (and we certainly want to make every inch count!). For instance, when going from first to second base, 90 ft (1080") is actually the distance from the foul line to the *center* of second base ([MLB21]). In fact, first base is placed completely in fair territory (to the left of the foul line), so this alone saves the runner 15", while second base is centered on the infield diamond, so it saves the runner another 15"/2 = 7.5". Hence, in total, the baserunner only needs to cover 1080 − 15 − 7.5 = 1057.5". Of course, baserunners do not run perfectly straight lines, but we neglect this in our study. On the other hand, what we do not ignore is the lead that runners take when they get on base, since this is (very!) nonnegligible. As Figure 2 shows, even though there are players with more raw speed than Ichiro, he may very well have (had) an edge over others when it comes to stealing bases because of his ability to read the pitcher and take very aggressive leads.

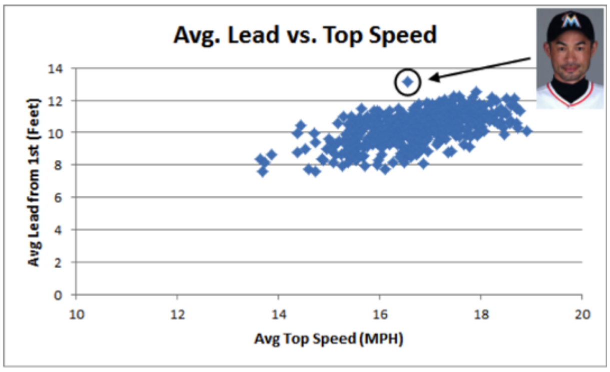

Figure 2: Average lead off of first base by MLB players in the 2015 season ([Tri15]). Ichiro is the oldest and riskiest!

From the figure we can infer a league average of $\sim 10\,\text{ft} = 120"$, so we take the final distance from first to second base to be 1057.5" − 120" = 937.5". In fact, since 120" is only an estimate based on the data shown in Fig. 2, let us round up and settle for a final distance $d^{(15")}_{1\text{B}\to 2\text{B}} = 938"$. A similar calculation for the new 18" bases yields 1080" − 18" − 18"/2 − 120" = 933"; thus we set $d^{(18")}_{1\text{B}\to 2\text{B}} = 933"$. Hence, the time t_{steal} it takes a baserunner with speed v_{runner} to reach second base on a steal attempt is

$$t_{\text{steal}} = \frac{d_{1\text{B}\to 2\text{B}}}{v_{\text{runner}}},$$

while the total time t_{defense} it takes the ball to leave the pitcher's hand to home plate and then from the catcher to second base is

$$t_{\text{defense}} = t_{\text{windup}} + t_{\text{pop}},$$

where t_{windup} is the time it takes the pitcher to deliver the ball to home plate and t_{pop} is the catcher's pop time, which we explained above. If $t_{\text{steal}} \leq t_{\text{defense}}$ we declare the base stolen; otherwise it is a failed steal attempt. [1] Data for v_{runner}, t_{windup}, and t_{pop} is taken from the tables in Fig. 3. I took a sample of ten players with varying skills from each category to create 10^3 different steal attempt scenarios. Data for pop time and baserunners' speed can be easily gathered from Statcast ([Sav21]); on the other hand, data for pitchers' delivery times are very hard to come by. I had to resort to [SL13], which only has data collected from a scouting report of the Baltimore Orioles' pitching roster. Syndergaard is the only exception on this list; his windup is notoriously slow and well documented ([NYT17]).

Pitcher	Delivery times (Avg in sec)
Dylan Bundy	1.36
Pedro Strop	1.39
Zach Britton	1.42
Jason Hammel	1.44
T.J. McFarland	1.45
Troy Patton	1.48
Mark Hendrickson	1.53
Michael Belfiore	1.55
Jair Jurrjiens	1.57
Noah Syndergaard	1.65

Catcher	Pop time (Avg in sec)
J.T. Realmuto	1.89
Willson Contreras	1.92
Martín Maldonado	1.96
Pedro Severino	1.99
Sandy León	2.02
Russell Martin	2.04
Austin Romine	2.06
Bobby Wilson	2.08
Tyler Flowers	2.12
Stephen Vogt	2.14

Baserunner	Sprint Speed (Avg in inch/sec)
Tim Locastro	368.4
Harrison Bader	354
Jon Berti	351.6
Randy Arozarena	345.6
Jose Altuve	340.8
Brett Gardner	338.4
Mike Tauchman	325.2
Vladimir Guerrero Jr.	303.6
Erik Kratz	285.6
Albert Pujols	264

Figure 3: Data extracted from [Sav21], [SL13], [NYT17]. I chose ten players with varying degrees of success from each category.

Assumptions made:

- Only the pitcher's delivery time, catcher's pop time, and the baserunner's speed are considered in the steal attempt; all others factors (such as the baserunner's reflexes) are ignored;
- Since we are ignoring individual baserunners' reflexes, we do not vary the baserunner's lead off first base on a case by case basis; instead we assume that all runners take the same lead ($\sim$ 120");
- We only consider steal attempts from first to second base (there is no loss of generality by testing our simulations for this particular case);
- Baserunners are assumed to run on a straight line;
- The catcher's pop time assumes the throw is right on target;
- The umpires are, for all intent and purposes, robots...we neglect human error in our studies.

Results- Steals

Figure 4 shows 1000 different scenarios, involving all ten players from each category from the tables in Fig. 3, for both the old (15") and newly proposed (18") base dimensions (Python code is provided in the appendices). As expected, the newly introduced base dimensions show an increase in successful steal attempts (932 successful steals for the 15" bases vs 939 for the 18" bases). It is somewhat reasonable that there are only 7 more successful attempts, since the size upgrade of the bases is only modest at best.

[1] The case $t_{\text{steal}} = t_{\text{defense}}$ is commonly known as a "bang-bang" play; in this case the call *should* favor the baserunner. Of course, umpires are not robots (this is, in fact, another proposed change to the MLB!), but we are neglecting human error in our investigation.

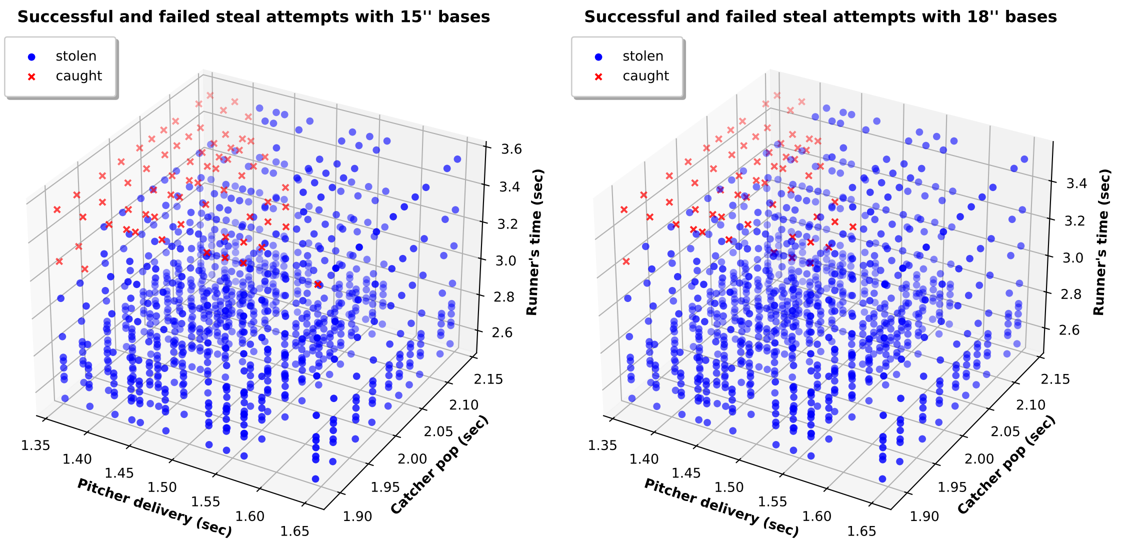

Figure 4: Results for 1000 different base-stealing scenarios, using data from the tables shown in Fig. 3. The left figure shows steal attempts with base size 15”, while on the right we have the expected results for the newly introduced 18” bases.

There is a couple of interesting observations we can infer from the results. As the figure shows, the more dominant factors are the baserunner’s speed and the pitcher’s windup time. For a pitcher with a fast windup, all but the most elite baserunners will get caught stealing, even with catchers with below-average pop time. On the other hand, for a slow pitcher like Syndergaard it almost doesn’t matter how good the catcher’s arm is or how slow the baserunner is; the base will most likely be stolen. The only runner from this data set that Syndergaard manages to catch on a steal attempt is Albert Pujols (the slowest baserunner of them all!), and even to accomplish this he required a catcher with an elite pop time (J. T. Realmuto). Moreover, with the base size upgrade we can see on the right figure that he cannot catch even Pujols! Syndergaard’s base-stealing woes are nicely illustrated in Fig. 5, taken also from [NYT17].

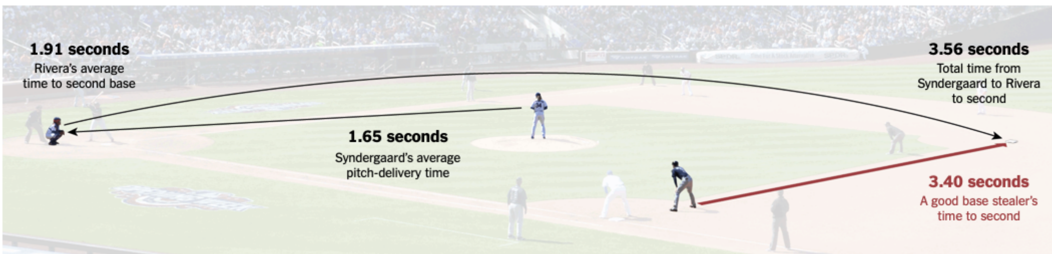

Figure 5: Image from [NYT17]; it shows that even while having a catcher with above-average pop time like René Rivera, a good baserunner should still be able to successfully steal a base against Syndergaard.

Methodology-Groundballs/Bunts

Now the distance the runner must cover from home plate to first base is $d^{(15”)}_{\mathrm{HP}\to 1\mathrm{B}} = 1072.5”$ $(1080 - 15/2)$, or $d^{(18”)}_{\mathrm{HP}\to 1\mathrm{B}} = 1071”$ $(1080 - 18/2)$ with the new dimensions. We are not making any distinction between left-handed and right-handed hitters; although it is true that, in reality, lefties have an edge over righties in the sense that they are positioned a bit closer to first base from their home stance, this will be ignored in our studies. The biggest hurdle we face now is that there is no data whatsoever…nothing/nada/zilch/niente …on arm strength of infielders. All the data in this category from Statcast (at least the data that is publicly available) is on catchers’ arms (both their pop time and their arm speed have readily available data), but nothing on the remainder of the infield. (I actually spent more time than I would like to admit trying to dig up some data somewhere in the internet, but I came up empty.) All I found was an arm strength “rating” for a few players on [Gra21], but even looking up their glossary there is no indication whatsoever of what they mean by “Arm Strength.” One thing is clear though: those numbers do not indicate arm speed.

Third baseman	Arm strength (inches per sec)
Gio Urshela	1526
Evan Longoria	1496
Yolmer Sanchez	1466
Todd Frazier	1431
Brock Holt	1382
Andres Gimenez	1357
Matt Carpenter	1322
Greg Garcia	1300
Brian Anderson	1243
Alex Bregman	1183

In the absence of data equivalent to pop time, arm speed is our best bet. Since we do not have data for infielders for the latter either, but we do for catchers, we will assign catcher arm speed values to infielders (specifically, to thirdbasemen). This is not an entirely outlandish proposition; it is quite realistic to expect that infielders (especially thirdbasemen and shortstops) have arm strength comparable to that of catchers, since they have to make throws across the infield that cover roughly the same distance. The table on the right shows ten thirdbasemen with catchers' armspeeds randomly asssigned to them (the catchers' armspeeds are taken from [Sav21]). The data is given in miles per hour (MPH), but converting is straightforward; for instance, to Evan Longoria we assigned a strong arm (Martín Maldonado; 85MPH):

$$\frac{85 \text{ mile}}{1 \text{ hour}} \times \frac{1 \text{ hour}}{60 \text{ minute}} \times \frac{1 \text{ minute}}{60 \text{ second}} \times \frac{63360 \text{ inch}}{1 \text{ mile}} = 1496 \, \frac{\text{inches}}{\text{second}}.$$

All the groundballs/bunts will be assumed to be hit directly at third base, and a shift is put in place so that the thirdbaseman is standing right on top of the base and he catches the ball there (in other words, the thirdbaseman doesn't have to move at all to catch the ball). Hence we shall consider a range of softly hit groundballs/bunts; otherwise the hitter would always be called out! Admittedly, this is a very idealized scenario and we could do better (e.g., we could consider a range of different groundball directions, or even be fancy and apply Monte Carlo ...), but due to time constraints we will restrict ourselves to this limited script. In the end, our sole objective is to determine whether there are different outcomes with the newly introduced base dimensions, and to that end even this simple scenario suffices (see results below). The baserunners' speeds used are the same as before, while the bunts/groundballs' speeds are based on a range of bat exit velocities, from 15MPH (264"/sec) to 60MPH (1056"/sec). [2] Hence, putting all together, we have the time t_{runner} it takes the hitter to go from the plate to first base,

$$t_{\text{runner}} = \frac{d_{\text{HP}\to 1\text{B}}}{v_{\text{runner}}},$$

as well as the total time t_{play} that the ball is in play before it reaches the firstbaseman's glove,

$$\begin{aligned} t_{\text{play}} &= t_{\text{ground}} + t_{\text{throw}} \\ &= \frac{d_{\text{HP}\to 3\text{B}}}{v_{\text{exit}}} + \frac{d_{3\text{B}\to 1\text{B}}}{v_{\text{throw}}}. \end{aligned}$$

Here v_{exit} is the bunt/groundball's exit velocity off the bat and v_{throw} is the (simulated) arm speed of the fielder. Similar to before, if $t_{\text{runner}} \leq t_{\text{play}}$ we have a Safe play; otherwise we have an Out. The distance $d_{\text{HP}\to 3\text{B}}$ is taken to be 1080" (the base dimension has no influence here; this is simply the distance from home plate straight to third base of an idealized groundball/bunt perfectly aligned with the third base foul line, which is still considered fair ball). Similarly, the distance $d_{3\text{B}\to 1\text{B}}$ that the ball must travel in the throw from third base to first base is taken to be 1527", per MLB specifications ([MLB21]). We ignore the firstbaseman stretch here, since we are already considering such an idealized scenario it would be moot to be too precise about the exact distance from the thirdbaseman to the firstbaseman's glove.

[2] We can think of the slowest balls in this range as the bunts, although in reality there are groundballs that are hit as softly as bunts (and in fact, there are bunts that come off as line drives with higher exit velocities than groundballs). It is for this reason that we consider both groundballs and bunts in the same case study.

Assumptions made:

- We only consider groundballs/bunts hit directly at third base. Furthermore, a defensive shift is set so that the thirdbaseman is standing right on top of the base and he catches the ball there;
- The firstbaseman's stretch is ignored;
- There is no distinction between right-handed and left-handed batters;
- Fielders (thirdbasemen) have arm strength comparable to catchers.

Results- Groundballs/Bunts

Figure 6 shows all the different scenarios played out, with both the old and new base sizes. The results are nearly identical, and very little difference can be inferred from the plots, but looking at the resulting data we find that two more baserunners reached first base safely with the 18" bases than with the 15" bases (330 for the former vs 328 for the latter). This modest contrast is to be expected, however, since this time the difference between the two cases is a meager 1.5 inches. Infield hits are not easy to come by, except for the slowest and better placed groundballs/bunts. In this simple case study the balls are hit directly at a fielder, so it is no surprise that the graph shows a hit only for the slowest groundballs/bunts ($t_{\text{ground}} \gtrsim 2$ sec). Nevertheless, this straightforward analysis suggests that, all things equal, the extra 1.5" provided by the new bases will result in (at least) a handful more plays going in the baserunner's favor.

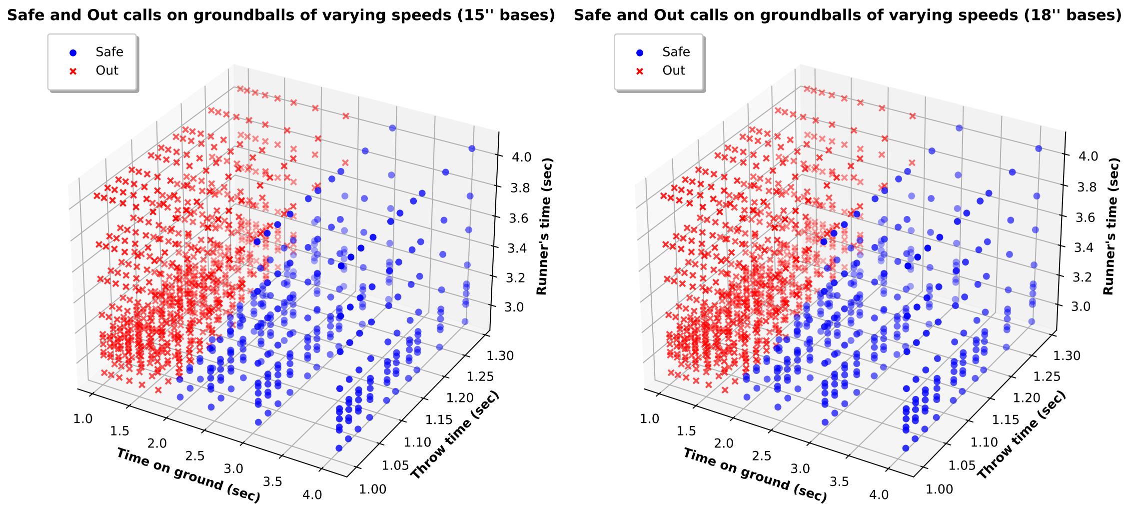

Figure 6: Results for 1000 different safe/out calls, based on a sample of ten different baserunner speeds, thirdbasemen arm strengths, and bat exit velocities common to groundballs and bunts. The left figure shows results with base size 15", while on the right we have the expected results for the newly introduced 18" bases. Only two more baserunners reached first base safely with the 18" bases than with the 15" bases.

Conclusions

We have considered two case studies (steal attempts and reaching first base on groundballs and bunts) to determine the potential impact of a size upgrade for bases in the MLB. Our results indicate that the newly introduced 18" bases do benefit baserunners, albeit slightly. There is just enough difference brought on by the new base dimensions to encourage baserunners to be more aggressive, while at the same time the differences are not game-breaking or unfair to the defense. There were 9 more stolen bases and 2 more players called Safe at first base on groundballs/bunts with the 18" bases than with the standard 15" size. While we made some overly simplifying assumptions in our study, the task of finding the impact (or lack thereof) of the new base dimensions has been accomplished. A more thorough investigation could use a larger (and more accurate) dataset, as well as perhaps use a Monte Carlo simulation to consider a wider, more complex range of play scenarios. For the base-stealing case study we could also consider individual baserunners' reflexes to create more personalized simulated data for each player on a case-by-case basis.

Appendix A. Python Code

The listing below shows the Python code for the base-stealing case. The code for the groundballs/bunts case is similar and can be provided upon request.

```
# Import libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

font = {'family' : 'serif',
        'weight' : 'normal',
        'size'   : 14}

#read data
v_runner  = pd.read_csv("~/$PATH/runners.csv", dtype=object, header = None)
v_runner  = pd.to_numeric(v_runner[1][1:]);

t_catcher = pd.read_csv("~/$PATH/poptime.csv", dtype=object, header = None)
t_catcher = pd.to_numeric(t_catcher[1][1:]);

t_pitcher = pd.read_csv("~/$PATH/pitchers.csv", dtype=object, header = None)
t_pitcher = pd.to_numeric(t_pitcher[1][1:]);


#distance from 1st to 2nd base (old and new)
d_old = 938;
d_new = 933;

#time for baserunner to go from 1st to 2nd base (old and new)
t_off_old = d_old/v_runner;
t_off_new = d_new/v_runner;

#initiliaze "caught stealing" and "stolen base" vectors
caught_old = [];
stolen_old = [];
caught_new = [];
stolen_new = [];

#Locate on the grid the succesful and failed steal attempts
for i in t_pitcher:
    for j in t_catcher:
        for k in t_off_old:
            if k <= i+j:
                stolen_old.append([i,j,k])
            else:
                caught_old.append([i,j,k])

for i in t_pitcher:
    for j in t_catcher:
        for k in t_off_new:
            if k <= i+j:
                stolen_new.append([i,j,k])
            else:
                caught_new.append([i,j,k])


#initialize coordinate vectors
x_stolen_old = [];
y_stolen_old = [];
z_stolen_old = [];

x_caught_old = [];
y_caught_old = [];
z_caught_old = [];

x_stolen_new = [];
y_stolen_new = [];
z_stolen_new = [];

```

```
x_caught_new = [];
y_caught_new = [];
z_caught_new = [];


for i in stolen_old:
    x_stolen_old.append(i[0])
    y_stolen_old.append(i[1])
    z_stolen_old.append(i[2])

for i in caught_old:
    x_caught_old.append(i[0])
    y_caught_old.append(i[1])
    z_caught_old.append(i[2])

for i in stolen_new:
    x_stolen_new.append(i[0])
    y_stolen_new.append(i[1])
    z_stolen_new.append(i[2])

for i in caught_new:
    x_caught_new.append(i[0])
    y_caught_new.append(i[1])
    z_caught_new.append(i[2])


# Plot for results of 15" bases
fig = plt.figure(figsize = (10, 7))
ax1  = plt.axes(projection ="3d")
ax1.scatter3D(x_stolen_old, y_stolen_old, z_stolen_old,
              marker = 'o', color = "blue", label='stolen')
ax1.scatter3D(x_caught_old, y_caught_old, z_caught_old,
              marker = 'x', color = "red", label='caught')

plt.legend(fancybox=True, framealpha=1, borderpad=1, shadow=True,  loc='upper left')

ax1.set_xlabel('Pitcher delivery (sec)', fontweight ='bold')
ax1.set_ylabel('Catcher pop (sec)', fontweight ='bold')
ax1.set_zlabel("Runner's time (sec)", fontweight ='bold')

plt.title("Successful and failed steal attempts with 15'' bases", fontweight ='bold')

# save plot
plt.savefig('../../Figures/baseball_SB_old.pdf', bbox_inches='tight')
plt.close()


# Plot for results of 18" bases
fig = plt.figure(figsize = (10, 7))
ax1  = plt.axes(projection ="3d")
ax1.scatter3D(x_stolen_new, y_stolen_new, z_stolen_new,
              marker = 'o', color = "blue", label='stolen')
ax1.scatter3D(x_caught_new, y_caught_new, z_caught_new,
              marker = 'x', color = "red", label='caught')

plt.legend(fancybox=True, framealpha=1, borderpad=1, shadow=True,  loc='upper left')

ax1.set_xlabel('Pitcher delivery (sec)', fontweight ='bold')
ax1.set_ylabel('Catcher pop (sec)', fontweight ='bold')
ax1.set_zlabel("Runner's time (sec)", fontweight ='bold')

plt.title("Successful and failed steal attempts with 18'' bases", fontweight ='bold')

# save plot
plt.savefig('../../Figures/baseball_SB_new.pdf', bbox_inches='tight')
plt.close()
```

Bibliography

[Gra21] Graphs, F. *fangraphs.com*. 2021. URL: https://www.fangraphs.com/leaders.aspx?pos=3b&stats=fld&lg=all&qual=y&type=2&season=2018&month=0&season1=2018&ind=0&team=0&rost=0&age=0&filter=&players=0&startdate=&enddate= (visited on 03/25/2021).

[MLB21] MLB. *Major League Baseball*. 2021. URL: https://www.mlb.com (visited on 03/25/2021).

[NYT17] NYT. *New York Times*. 2017. URL: https://www.nytimes.com/interactive/2017/04/14/sports/baseball/noah-syndergaard-stolen-bases.html (visited on 03/25/2021).

[Sav21] Savant. *Baseball Savant*. 2021. URL: https://baseballsavant.mlb.com (visited on 03/25/2021).

[SL13] Sports, B. and Life. *Delivery and Pop Time Significance*. 2013. URL: https://baltimoresportsandlife.com/delivery-time-and-pop-time-significance/ (visited on 03/26/2021).

[Tri15] Triangle, T. *Grantland*. 2015. URL: https://grantland.com/the-triangle/2015-mlb-statcast-baserunning-data-ichiro-suzuki/ (visited on 03/25/2021).