Jean (Cheng-Young) An

New York University

Independent Study – Final Project

Dr. Bri Newland & Jason Rosenfeld

20 May, 2019

Introduction and Overview

When PITCHf/x debuted in MLB stadiums in 2006, it was as if baseball analysts both within and outside the industry received a brand new shiny toy to play with. Pitch tracking technology opened up endless possibilities of new studies that would not have been possible prior to its existence. One of these studies is how a catcher's framing ability influences balls and strikes calls, and the new data allowed analysts to take what was long known as an art throughout baseball history, and turn it into a science.

In 2015, Major League Baseball started using the Trackman Doppler radar system as their new pitch tracking technology, and along with high-speed cameras that tracks other aspects of the game that radar technology is incapable of, this technology system was coined "Statcast." A team of data scientists and analysts were formed to work on maximizing the capabilities of this system, and most recently, they have published their catcher framing metrics, shortly after Steamer published their framing metric on FanGraphs (Cross).

When Jared Cross, developer of the Steamer system, published his framing metric, he conducted a study that compared his metric to the other well-known existing framing metrics in order to test its credibility. Thus, the student intends to conduct a similar study for the Statcast framing metric to gain a better understanding of this newly-published catcher evaluation tool. This would potentially determine which framing metric is more reliable when one intends to measure the value provided by each catcher in this particular skillset.

Lastly, it is worth noting that this study serves as the final project for the student's independent study course in the Spring 2019 semester. In this course, the student has completed both the readings and the exercises of *Analyzing Baseball Data with R, Second Edition* (Marchi et al.). Therefore, the student will be using R as the sole programming language when handling all data used in this study as a way to demonstrate his proficiency. Readers can find the code used for this study on the student's GitHub (ID: anchengyoung) in the repository *Statcast_Framing*.

Method

Data Collection

All data used for this study were downloaded directly from the Internet. Framing data as calculated and published by Statcast were downloaded from their website Baseball Savant, in the "Catcher Framing" section of their Statcast Leaderboards page. Framing data as calculated and published by Baseball Prospectus were downloaded from their website of the same name, in the "Catcher Stats - Seasonal Totals" section found in the Stats/Tools page. Framing data as calculated by Steamer and published by FanGraphs were downloaded from FanGraphs' Leaderboards page. All data were downloaded in the format of Comma-Separated Values (CSV) files.

Data Selection

As statistically-significant sample sizes are desired, there were some minor processes that were necessary when selecting the data. According to the description on Baseball Savant, the qualifier that is used on their website is that "a catcher must receive 6 called pitches per team game." Therefore, this was used as the qualifying threshold in this study. This threshold resulted in 56 qualified catchers in 2015, 63 qualified catchers in 2016, 62 qualified catchers in 2017, and 60 qualified catchers in 2018.

When selecting data from Baseball Prospectus, however, the same qualifier was not provided. Baseball Prospectus allowed users to set their own qualifiers based on CSAA_CHANCES, otherwise known as Framing Chances. For the 2018 season, using 2,060 Framing Chances as a threshold resulted in the exact same 60 catchers that qualified in the 2018 Statcast dataset; for the 2017 season, the threshold of 2,000 Framing Chances was used to obtain the 62 catchers; and for the 2016 season, the threshold used was 2,050 Framing Chances for the 63 catchers.

For the 2015 season, the threshold of 1,950 Framing Chances was necessary for the return to include all 56 catchers that qualified in the 2015 Statcast dataset. However, this threshold included a total of 60 catchers. The four catchers that met this 1,950 Framing Chances threshold but did not meet Statcast's qualifier were Sandy Leon, Tony Cruz, Andrew Susac, and Tuffy Gosewisch. They were later on removed from the 2015 Baseball Prospectus dataset.

Lastly, when selecting data from FanGraphs, the only qualifier provided was innings played on defense. The 250-inning threshold was used for all four seasons in order to include all catchers that did meet the aforementioned Statcast qualifier. This resulted in 63 catchers in 2015, 66 catchers in 2016, 63 catchers in 2017, and 64 catchers in 2018. All the extra catchers were later on removed from the datasets.

Data Organization

Once all the datasets were imported into RStudio, more steps were necessary to organize the data in a way that could be easily used for calculations in this study. First, the only identifier used in the Baseball Prospectus datasets is the full name of each catcher (e.g. Yadier Molina), stored in the `NAME` column. Since this is the only identifier available in the Baseball Prospectus datasets, Statcast and FanGraphs datasets were adjusted accordingly.

The main identifier used in the Statcast datasets is the column titled `fielder_2`, which stores the six-digit Statcast player ID corresponding to each catcher. Each catcher's name is separately stored in the `last_name` and `first_name` columns. Therefore, for the purpose of joining datasets from each source later on in this study, the `first_name` and `last_name` columns were combined into a `NAME` column to match with the Baseball Prospectus identifier.

As for FanGraphs datasets, in addition to the `playerid` column that stores their original numeric identifier, the full name of each catcher is also stored in the `Name` column. This column was also renamed to `NAME` in order to match with the datasets from the two other sources. Lastly, one minor data adjustment had to be made for the 2016 Baseball Prospectus dataset. One catcher, known as "Curt Casali" in both the Statcast and FanGraphs dataset, was named "Curtis Casali" in the Baseball Prospectus dataset. His first name was changed to "Curt."

Calculations

The next step of the study is the calculation. There are two major calculations conducted in this study. First, there is the year-to-year correlation calculation that is done separately for datasets from all three sources. This should lead to better understandings of the credibility of the framing runs metrics of each data source. Then, another source-to-source correlation analysis among datasets of these three sources is done, in order to have a better understanding of the relationship among them.

Year-to-Year Correlation

Assuming that framing is indeed a skill that the catcher possesses, conducting a year-to-year correlation analysis would provide a path to conclude whether or not the framing runs metric of each source successfully captures the essence of this skillset. There were 46 catchers that qualified in both 2015 and 2016, also 46 that qualified in both 2016 and 2017, and 44 catchers that met the requirements for 2017 and 2018, resulting in a total of 136 year-to-year pairs.

Source-to-Source Correlation

In addition to the individual year-to-year correlation of the framing metrics from each data source, the student would also like to understand how the metrics from each source compare to each other. The four datasets of each season from each source are bound together, resulting in a data frame of 241 rows, in order to evaluate the relationship of the framing metrics from each in a larger sample size.

This data frame includes the name of all catchers, the Statcast framing runs, the Baseball Prospectus framing runs, and the FanGraphs framing runs. The student created a function, `graph()`, that takes this data frame as an input, plus selection of two of the three sources; it produces a scatter plot with framing runs from one source on the x-axis and framing runs from the other source on the y-axis. A best-fit line is also included on the graph. This function would also work on the individual season datasets. Lastly, the correlation coefficients were calculated with a simple `summarize()` function.

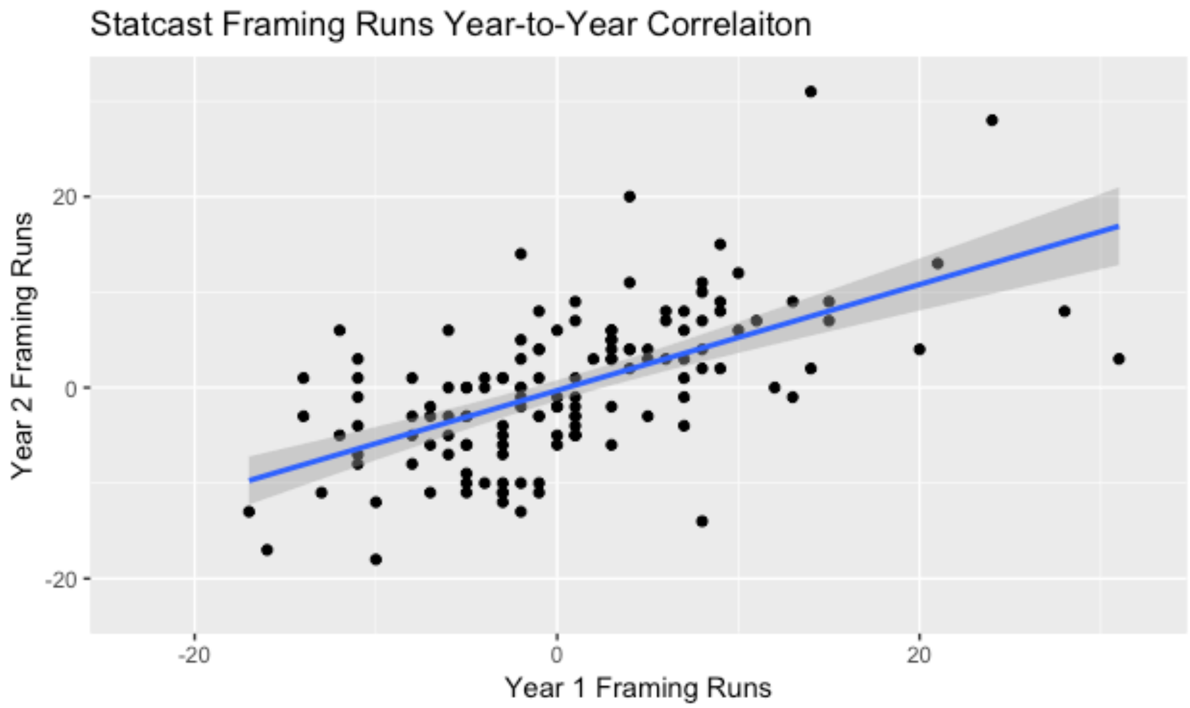
Results

Purpose of Study

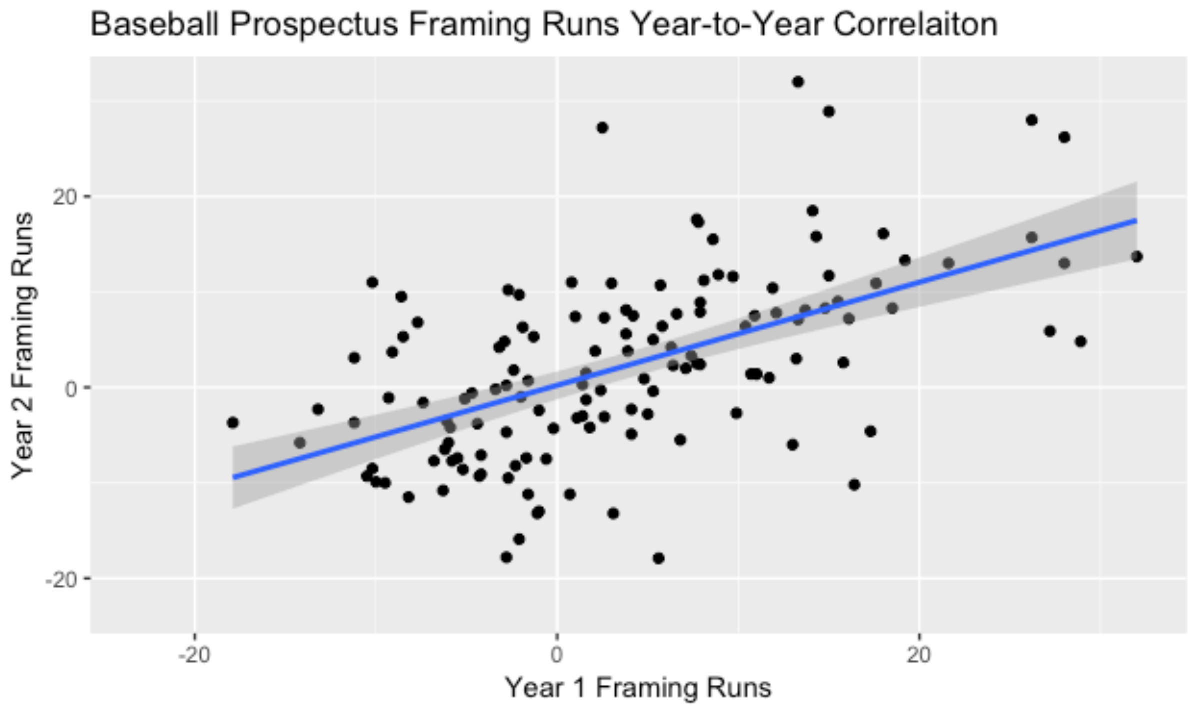
As the framing metrics as calculated and published by Statcast is the most newly released, publically-available version, it is to the student’s interest to understand how this version of framing metric compares and contrasts to the two other predominantly used framing metrics, one calculated and published by Baseball Prospectus, another calculated by Steamer and published by FanGraphs. This section will present the results of the calculations mentioned in the Method section and other findings.

Year-to-Year Correlation

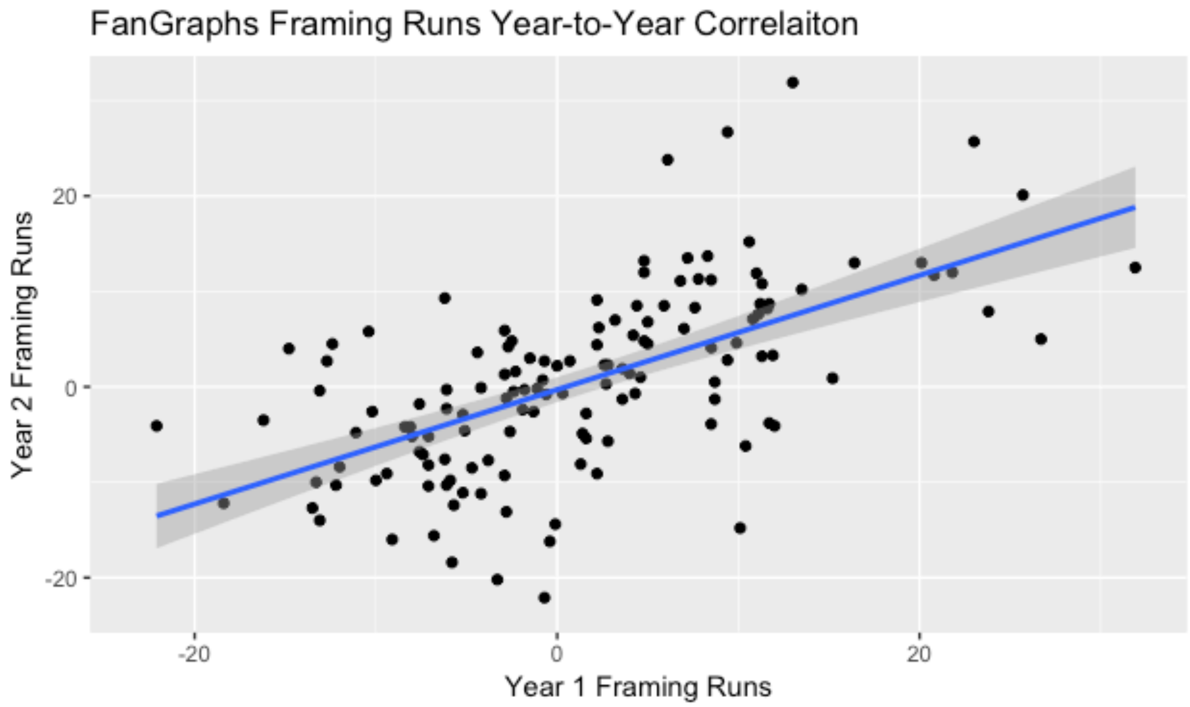
As mentioned, year-to-year correlation is an important aspect to understanding a metric that aims to capture the performance of players in terms of a skill. Here, the student will present the scatter plots of the year-to-year framing runs relationship of all three data sources, beginning with Statcast, then Baseball Prospectus, and ending with FanGraphs. The correlation coefficients will be presented altogether after the three graphs.



The graph above shows the relationship between catcher's Statcast framing runs from one year to the next. What can be observed from the graph is that the majority of the points lies between ±10 runs on both axes. The best-fit line, which goes through the origin, shows a clear upward trend, suggesting a positive relationship between the two axes. Lastly, one can see that the top-left and bottom-right corners have no data points present. This means that no catcher's framing, as suggested by the metric, suddenly went from very bad to very good, or vice versa, leading us to believe that the positive relationship is statistically significant.



The graph above shows the relationship between catcher's Baseball Prospectus framing runs from one year to the next. Similar to the previous graph, which charts the Statcast framing runs, we can see the majority of the points lies between ±10 runs on both axes. However, there also seems to be more points present in the > 10 runs area when compared to Statcast's graph. Lastly, the best-fit line, which also goes through the origin, and the lack of data points in the top-left and bottom-right corners, also suggest a statistically significant positive relationship.



The graph above shows the relationship between catcher's FanGraphs framing runs from one year to the next. The graph shows similar traits to the previous two graphs; best-fit line that goes through the origin and no data points in the top-left, bottom-right corners, once again suggesting a statistically-significant, positive relationship. It is worth noting that the majority of the data points now lie in a larger range, roughly between ±15 runs, and are not as condensed as the data points shown in the two previous graphs.

The following table shows the correlation coefficient of each graph:

Data Source	Statcast	Baseball Prospectus	FanGraphs
Correlation Coefficient	0.593	0.562	0.611

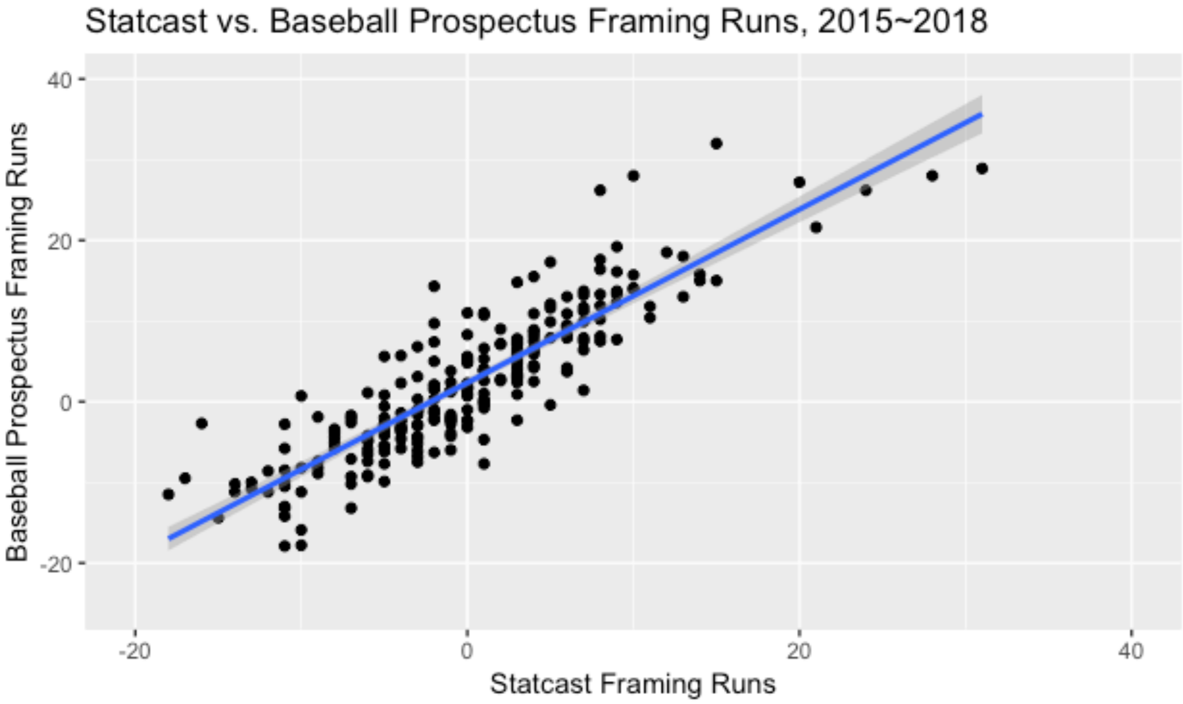
Interestingly, all three datasets returned similar year-to-year correlations for their respective framing metrics. The newly published Statcast framing runs, in the limited sample size of 136 year pairs, actually has a slightly higher correlation coefficient than Baseball Prospectus' framing runs, and is very closely behind that of FanGraphs. This suggests that all three data sources likely have a similar grasp on the quantifiable aspect of pitch framing.

In addition, while the correlation coefficients do not seem to be as high as one would want, one must remember that framing is not a one-dimensional skill that is all dependent on the catcher. The pitcher, the umpire, and even the hitter all have impacts on the results of each pitch. Moreover, as framing runs is a cumulative statistic, the total framing chances each catcher receives in different seasons may also cause the end result to differ. These influencers would later be discussed in the recommendations.

Source-to-Source Correlation

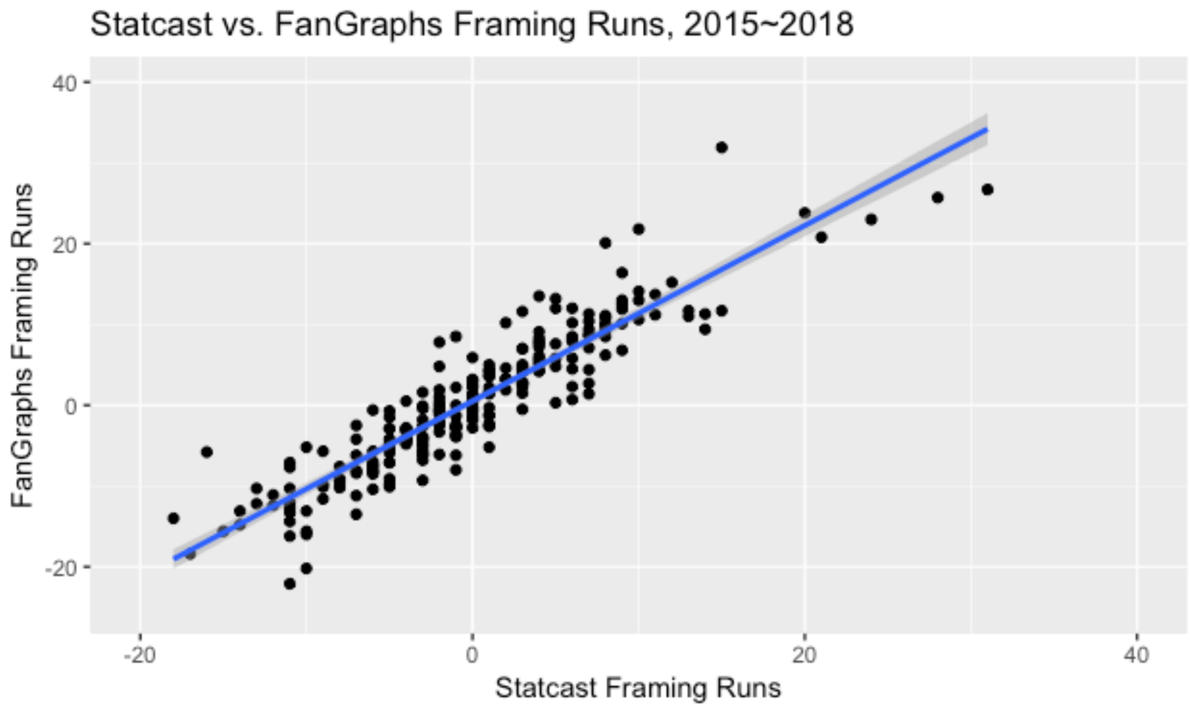
Individual year-to-year correlations give a good understanding of how well the framing metrics were constructed by each data source separately. However, it does not shed light on how these different metrics truly compare or contrast with each other. As mentioned in the introduction, when a new version of a metric – in this case the Statcast framing runs – is born, it is important to compare it with existing versions of the metric to test whether or not the new version represents an upgrade over its competitors.

Once again, as Statcast framing runs is the primary metric to be tested, the student begins with comparing Statcast framing runs to Baseball Prospectus framing runs. The resulting graph is as follows.



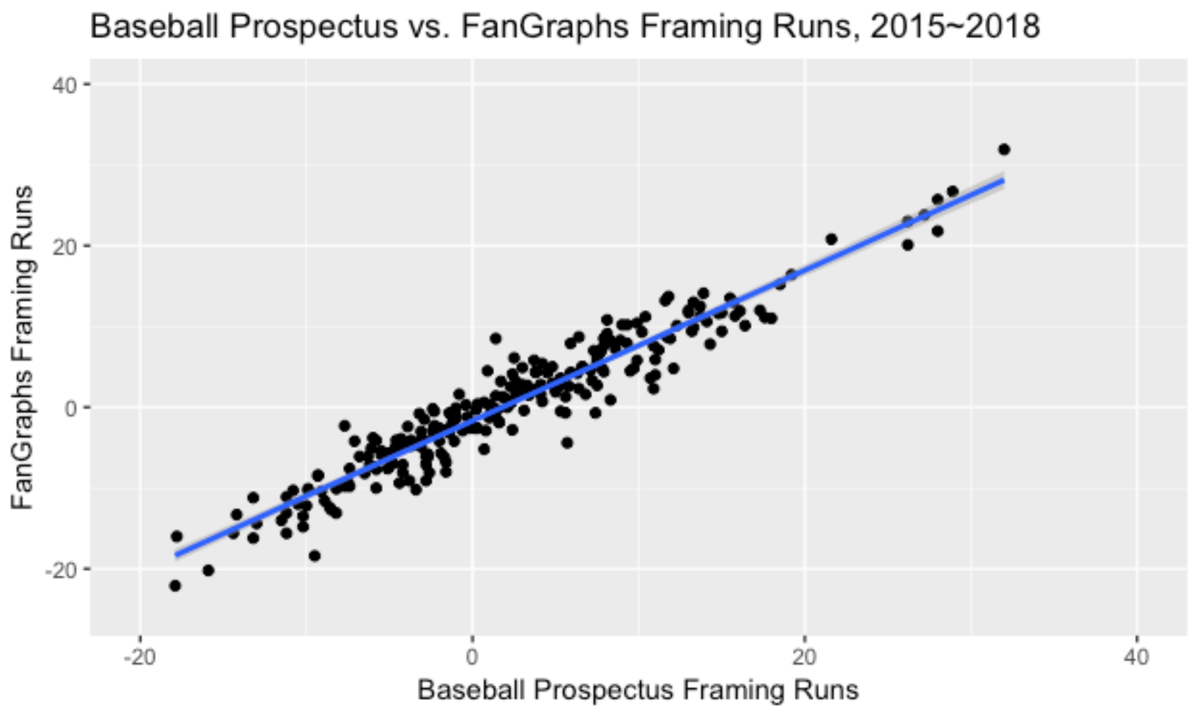
As one can see, the two framing runs metrics have a very strong positive relationship. However, it is worth noting that the best-fit line does not cross through the origin; instead, it is slightly above it. This suggests that, of these 241 catcher-seasons, Baseball Prospectus framing runs, on average, credited these catchers with a slightly higher contribution than what Statcast has credited them with. This would be later on discussed in the recommendations. Next, we compare Statcast framing runs with FanGraphs framing runs, with the graph as follows.

(graph on next page)



Similar to the previous graph, a strong positive relationship is observed. In contrast, the best-fit line does cross through the origin in this case, suggesting that as a whole, Statcast and FanGraphs are giving similar credits to the 241 catcher-seasons. Also, there does not seem to be as many data points that are scattered far away from the best-fit line. This would lead us to believe that Statcast framing runs is more correlated with FanGraphs framing runs compared to Baseball Prospectus framing runs. This is later on confirmed when the correlation coefficients are presented at the end of this section.

Lastly, even though Statcast framing runs is our primary metric to test, framing runs from the two other sources are also compared with each other and the resulting graph is as follows.



It is quite clear that this pair presents the highest positive relationship when compared to the two previous graphs. The data points are all much closer to the best-fit line. Moreover, as one might expect after observing the two previous graphs, this best-fit line has a y-intercept slightly towards the negative, suggesting that Baseball Prospectus framing runs has credited the 241 catcher-seasons slightly higher than FanGraphs framing runs has.

The following table shows the correlation coefficient between each data source:

Data Source	Statcast	Baseball Prospectus	FanGraphs
Statcast	1.000	0.879	0.914
Baseball Prospectus	0.879	1.000	0.958
FanGraphs	0.914	0.958	1.000

While all three pairs have a high correlation as one would expect, it is confirmed that Baseball Prospectus and FanGraphs framing runs indeed has the highest correlation coefficient, exceeding 0.95. On the other hand, Baseball Prospectus and Statcast framing runs has the lowest correlation coefficient, but still a respectable figure close to 0.88.

Conclusions and Recommendations

This concludes the study the student has conducted with regard to the newly-published Statcast framing runs metric. In this final section, the student would provide a few recommendations of future studies that could be extended in relation to this study. The student would also point of a few shortcomings of this study and elaborate on how improvements could be made in the future, especially when access to new data that are not currently available becomes public.

Recommendation #1

As discussed when sharing the correlation coefficient of the year-to-year correlation analysis, framing runs is a cumulative statistic. This means that the outcome is largely dependent on the chances each catcher gets to frame pitches (i.e. called pitches in certain areas around the strike zone). The student originally intended to attempt the year-to-year correlation on a per 100 pitches bases, but ultimately decided against doing so because of data discrepancy.

The datasets from Statcast provided a column titled `n_called_pitches`, which stores the number of called pitches each catcher receives in a defined "Shadow Zone" (Tangotiger). This number ranges from roughly 1,000 to 4,000 in our qualified catcher-seasons. Data from Baseball Prospectus, on the other hand, provided a column titled `Framing.Chances`, which according to the description in their glossary, includes all pitches that have a non-zero probability of being called a strike. This number ranges from roughly 2,000 to 8,000 in our qualified catcher-seasons.

Because of the differences in how these two sources determine what is a "framable" pitch, it became difficult to decide on how to adjust the framing runs from each source to a per 100 pitch bases. One could choose to divide the framing runs by their respective "framable" pitch totals, or one could choose either `n_called_pitches` or `Framing.Chances` as the common denominator. Future researchers who make this decision would be able to conduct a study to see whether or not the year-to-year correlations improve once this cumulative stat is essentially scaled to become a rate stat. It may also improve the correlation between the Statcast and Baseball Prospectus framing runs, which came out at roughly 0.88 in this study.

Recommendation #2

As noted in the Introduction and Overview, this study was inspired by the study Jared Cross conducted when publishing his Steamer framing runs on FanGraphs. In that study, Cross compared his metric to those of Baseball Prospectus, StatCorner and Sports Info Solutions. In this study, however, the student only compared the Statcast framing metric to those of Baseball Prospectus and FanGraphs (i.e. Steamer by Cross). This is because the student is not as familiar with framing metrics produced by StatCorner and SIS.

Thus, if future researchers wish to expand on this study, the inclusion of StatCorner and SIS may be something worth considering. However, in Cross' study, framing metrics published by StatCorner and SIS produced year-to-year correlation coefficients south of 0.7, while both FanGraphs and Baseball Prospectus returned coefficients above 0.7, suggesting that the latter two may be superior metrics. In such a case, it may not be necessary to include the lesser metrics (unless the intention is to prove that the Statcast framing metric belongs in the same tier as that of StatCorner and SIS).

Recommendation #3

When comparing the framing runs of Statcast with that of Baseball Prospectus, the student noted how the latter seemingly credited catchers with a higher framing runs contribution that both Statcast and FanGraphs. In fact, when one adds up the total framing runs of all the 241 catcher seasons in the qualified datasets, Statcast framing runs produced a total of +2 runs, FanGraphs framing runs sums up to +125.8 runs, and Baseball Prospectus framing runs comes out to an astounding total of +568.2 runs.

To put this in another perspective, the average catcher-season in our dataset is essentially at zero Statcast framing runs, which makes sense, because framing runs is a statistic that is normalized to zero being league average. FanGraphs framing runs has an average catcher-season in our dataset at roughly 0.5 runs, so it is within a reasonable margin of error. However, the average catcher-season in terms of Baseball Prospectus framing runs is almost 2.36 runs. The median values are 0, 0, and 1.5, respectively.

Of course, it is logical for the average catcher-season selected in this study to be above zero. Because the catchers who meet our qualifier are those that get the most playing time, these catchers should be better framers compared to those who do not qualify (i.e. framing ability may be part of how teams decide whether to play a catcher more than another). There are obviously other factors involved, but the student believes this is a logical explanation.

However, this explanation likely still does not justify the large difference observed in terms of the average Baseball Prospectus framing runs compared to the other two sources. Future researchers may be interested in finding out why Baseball Prospectus has returned a significantly higher framing runs total when selecting catchers using the Statcast qualifier. It is also possible that while Baseball Prospectus is returning a high value, Statcast is returning a low value, and the ideal outcome is actually somewhere in the middle.

Recommendation #4

The reader may have noticed through the reading of the graph that, while Baseball Prospectus and FanGraphs both publish their framing runs number to one decimal point, Statcast publishes their framing runs total in integer values. At first, the student thought this may be the reason that Statcast framing runs is unable to produce as high a correlation with the two other sources as that between Baseball Prospectus and FanGraphs.

However, when the student rounded both the Baseball Prospectus and FanGraphs framing runs to integer values and re-conducted the correlation analysis, the correlation coefficient still came out at 0.879 between Statcast and Baseball Prospectus, 0.914 between Statcast and FanGraphs, and 0.958 between FanGraphs and Baseball Prospectus. Changes were only observed beginning at the fourth decimal place of the correlation coefficient, meaning that the rounding error was likely negligible.

If one could obtain or calculate Statcast's framing runs down to the first decimal point or further, one could re-run the study to see if that improves any of the correlations calculated in this study. However, if rounding the numbers up to the integer level only produced such a minor and essentially negligible difference, the student would believe that expanding Statcast framing runs down to one decimal point would only bring about a similarly minimal effect.

Conclusion

The student conducted this study with the purpose of understanding how the new Statcast framing runs metric compares to two of the well-known framing runs metrics, one calculated and published by Baseball Prospectus, one calculated by Steamer and published by FanGraphs.

Using catchers that met the Statcast qualifier from the 2015 to 2018 seasons, which results in a total of 136 year pairs, the student first conducted year-to-year correlations for each metric, and the results showed that the Statcast framing runs returned a correlation coefficient sandwiched between that of Baseball Prospectus and FanGraphs. As all three coefficients were similar, it suggested that all three metrics have similar credibility and predictive strengths.

The student then used the entire dataset, which includes a total of 241 catcher-seasons, and looked at how the framing runs from each of the three sources compared with one another. The result suggested that Baseball Prospectus and FanGraphs framing runs were most alike in terms of assessment, while Statcast and Baseball Prospectus were least alike, yet still highly similar. The 0.958 correlation coefficient between Baseball Prospectus and FanGraphs is in line with the 0.96 correlation coefficient that Cross presented in his study.

Lastly, the student is also aware of the recent news, reported by Eno Sarris of The Athletic, that Major League Baseball is preparing to replace the Trackman radar pitch-tracking system with the Hawk-Eye optical technology beginning in the 2020 season. The article states that multiple sources within the industry suggested that lack of accuracy is the main reason that the MLB is moving on from a relatively young Trackman system. If pitch location data is indeed improved in accuracy in the future, it would lead to more accurate results in all of the framing runs metrics presented in this study.

References

Cross, Jared. "FanGraphs Pitch Framing." *FanGraphs Baseball*, FanGraphs, 20 March 2019, https://blogs.fangraphs.com/fangraphs-pitch-framing/

Marchi, Max, et al. *Analyzing Baseball Data with R, Second Edition*. CRC Press, 2019.

Sarris, Eno. "Sarris: MLB moving from Trackman to Hawk-Eye tracking system." *The Athletic*, The Athletic Media Company, 14 May 2019, https://theathletic.com/975903/

Tangotiger. "WOWY Framing, part 1 of N." Tangotiger Blog, Tangotiger.com, 22 March 2019, http://tangotiger.com/index.php/site/comments/wowy-framing-part-1-of-n